Single cell transcriptome analysis defines heterogeneity of the murine pancreatic ductal tree

Audrey M. Hendley[1,12], Arjun A. Rao[2,3,13], Laura Leonhardt[1,13], Sudipta Ashe[1,13], Jennifer A. Smith[1], Simone Giacometti[1], Xianlu L Peng[4,5], Honglin Jiang[6], David I. Berrios[1], Mathias Pawlak[7], Lucia Y. Li[1], Jonghyun Lee[1], Eric A. Collisson[6], Mark Anderson[1], Gabriela K. Fragiadakis[2,3,8], Jen Jen Yeh[4,5,9], Jimmie Ye Chun[10], Grace E. Kim[11], Valerie M. Weaver[12], and Matthias Hebrok[1]

Correspondence: Matthias.Hebrok@ucsf.edu

Keywords and phrases: scRNA-seq, *Gmnn*, *Spp1*, pancreatic duct ligation, duct heterogeneity

[1]Diabetes Center, University of California, San Francisco, San Francisco, California, USA.
[2]CoLabs, University of California, San Francisco, California, USA.
[3]Bakar ImmunoX Initiative, University of California, San Francisco, California, USA.
[4]Department of Pharmacology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.
[5]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.
[6]Division of Hematology and Oncology, Department of Medicine and Helen Diller Family Comprehensive Cancer Center, UCSF, San Francisco, California, USA.
[7]Evergrande Center for Immunologic Diseases, Harvard Medical School and Brigham and Women's Hospital, Boston, Massachusetts, USA.
[8]Department of Medicine, Division of Rheumatology, University of California, San Francisco, California, USA.
[9]Department of Surgery, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.
[10]Parker Institute for Cancer Immunotherapy, San Francisco, California, USA.
[11]Department of Pathology, University of California, San Francisco, San Francisco, California, USA.
[12]Center for Bioengineering and Tissue Regeneration, UCSF, San Francisco, California, USA.

[13]These authors contributed equally.

ABSTRACT

To study disease development, an inventory of an organ's cell types and understanding of physiologic function is paramount. Here, we performed single-cell RNA sequencing to examine heterogeneity of murine pancreatic duct cells, pancreatobiliary cells, and intrapancreatic bile duct cells. We describe an epithelial-mesenchymal transitory axis in our three pancreatic duct subpopulations and identify osteopontin as a regulator of this fate decision as well as human duct cell dedifferentiation. Our results further identify functional heterogeneity within pancreatic duct subpopulations by elucidating a role for geminin in accumulation of DNA damage in the setting of chronic pancreatitis. Our findings implicate diverse functional roles for subpopulations of pancreatic duct cells in maintenance of duct cell identity and disease progression and establish a comprehensive road map of murine pancreatic duct cell, pancreatobiliary cell, and intrapancreatic bile duct cell homeostasis.

IMPACT STATEMENT

Single cell RNA-sequencing defines heterogeneity within the pancreatic ductal tree, and follow-up functional analyses identify unique properties of subpopulations of duct cells including an epithelial-mesenchymal transcriptomic axis and roles in chronic pancreatic inflammation.

INTRODUCTION

Pancreatic duct cells, while a minority of the composition of the pancreas, play an integral role in secretion and transport of digestive fluid containing proenzymes synthesized by acinar cells, electrolytes, mucins, and bicarbonate. They can serve as a cell of origin for pancreatic ductal adenocarcinoma (PDA) (1, 2) and have been implicated in the pathophysiology of multiple other diseases including cystic fibrosis (3) and pancreatitis (4).

Heterogeneity of a cell type becomes increasingly important in the context of disease and regeneration since different subpopulations can be the driving forces behind pathogenesis. The function of exocrine pancreatic cells is required for survival, yet these cells exhibit limited regenerative capabilities in response to injury. Chronic pancreatitis (CP) is a risk factor for pancreatic cancer. The underlying mechanisms for PDA progression in CP patients are incompletely understood and are likely multifactorial, including both genetic and environmental insults (5). Studies have shown that cytokines and reactive oxygen species generated during chronic inflammation can cause DNA damage. It has been hypothesized that pancreatic cells might acquire DNA damage in the protooncogene *KRAS* or tumor suppressor genes *TP53* or *CDKN2A*, thereby accelerating malignant transformation (6, 7). Thus, it is imperative to understand the mechanisms by which DNA damage occurs in the setting of CP. Duct obstruction is one cause of CP, and the ability of ductal cells to acquire DNA damage in the setting of CP is incompletely understood.

In this report, we conducted single-cell RNA sequencing (scRNA-seq) on homeostatic murine pancreatic duct, intrapancreatic bile duct, and pancreatobiliary cells using a DBA^+ lectin sorting strategy, and present a high-resolution atlas of these murine duct cells. By extensively comparing our subpopulations to previously reported mouse and human pancreatic duct subpopulations (8-10), we both corroborate several previous findings and identify and validate novel duct cell heterogeneity with unique functional properties including roles for subpopulation markers in CP. Our findings suggest that multiple duct subpopulations retain progenitor capacity, which is influenced by expression of markers driving subpopulation identity.

RESULTS

scRNA-seq identifies multiple pancreas cell types with DBA lectin sorting

Previously reported subpopulations of murine pancreatic duct cells were identified by single cell analysis of pancreatic cells obtained using an islet isolation procedure; thus, exocrine duct cells were of low abundance (9). To circumvent this issue, we employed a DBA lectin sorting strategy that has been extensively used to isolate and characterize all murine pancreatic duct cell types (11, 12), to investigate murine duct heterogeneity. We isolated live DBA^+ cells from the pancreata of four adult female C57BL/6J littermates, and performed scRNA-seq on the pooled cells using the 10X Genomics platform (Figure 1A and S1A). After filtering out doublets and low-quality cells (defined by low transcript counts), our dataset contained 6813 cells. Clustering analysis identified 16 distinct cell populations with an average of 5345 transcripts per cell and 1908 genes per cell (Figure 1B and Figure 1-source data 1). Significantly differentially expressed genes (DEGs) when comparing a cluster to all other clusters are listed in Figure 1-source data 2. Annotation of these 16 clusters was accomplished by analysis of known markers (Figure 1B-D). Our dataset comprises 2 populations of ductal cells, a cluster of endothelial cells, one cluster of fibroblasts, and 12 immune cell clusters. As expected, murine endocrine and acinar cells are not present in our dataset because they are not DBA^+ cells. Gene and transcript counts for each cluster are shown in Figure 1-figure supplement 1B. We identified DBA^+Collagen I^+ fibroblasts and DBA^+CD45^+ immune cells by immufluorescence. $CD31^+$ endothelial cells are not DBA^+. Their presence in our dataset might be explained by the close

juxtaposition of pancreatic duct cells with endothelial cells throughout the murine pancreas (Figure 1-figure supplement 1C).

Subpopulations of ductal cells are characterized by unique gene signatures and regulation of pathways

To get a better understanding of duct cell heterogeneity, we generated an Uniform Manifold Approximation and Projection (UMAP) plot using all duct cells (clusters 0 and 8), which revealed six distinct ductal clusters. Annotation of each duct cluster was accomplished using DEGs, Ingenuity Pathways Analysis (IPA) and upstream regulator analysis, and marker assessment in murine and human pancreas (Figure 2A-D, Figure 1-figure supplement 1D-E, Schematic 1, and Figure 2-source data 1-3). Gene and transcript counts for each cluster are shown in Figure 1-figure supplement 1F and Figure 2-source data 4. We observed variable expression of known ductal markers within clusters. Notably, fewer murine duct cells express the transcription factor *Hnf1b* when compared to *Sox9*. This observation is in contrast to a previous report demonstrating a similar prevalence of adult murine $HNF1B^+$ and $SOX9^+$ duct cells, which might be explained by different ductal cell isolation methods (Figure 1-figure supplement 1G) (13).

Cluster 0 contains the most cells of all duct clusters in the dataset (Figure 2-source data 4). A gene that positively regulates Ras signaling *Mmd2*, the voltage-gated potassium channel protein encoded by *Kcne3*, as well as the ATP-binding cassette (ABC) transporter chloride channel protein encoded by *Cftr,* were significantly upregulated in cluster 0 when compared to all other ductal clusters (Figure 2C and Figure 2-source data 1). Notably, cluster 0 shows upregulation or activation of multiple genes whose alteration play important roles in the pathophysiology of human pancreatic diseases such as *CFTR* for hereditary chronic pancreatitis (14) and *TGFB2* and *CTNNB1* for pancreatic cancer (15-17) (Figure 2-source data 1,3).

To validate gene expression patterns and determine the location of cluster 0 cells within the hierarchical pancreatic ductal tree (18), we next examined expression of select significantly DEGs. *Gmnn*, an inhibitor of DNA replication, was expressed in both clusters 0 and 2, so we decided to examine histologically, and were surprised to find rare GMNN protein expression, which was in contrast to the widespread RNA expression depicted by the feature plot (Figure 2-figure supplement 1A). After examining more than 1500 main pancreatic duct cells from 5 donors, we were unable to find a GMNN positive cell, indicating very low or absent expression of GMNN in human main pancreatic ducts. *Spp1*, which encodes osteopontin, and *Wfdc3*, which are significantly DEGs in both clusters 0 and 2, show cytoplasmic protein expression in all mouse and human pancreatic duct types (Figure 2-figure supplement 1B-C and Supplementary file 1).

Cells in cluster 1 have significantly upregulated expression of the exosome biogenesis gene *Rab27b* as well as *Ppp1r1b* that encodes for a molecule with kinase and phosphatase inhibition activity (Figure 2A-C and Figure 2-source data 1). IPA results suggested an enrichment in molecules regulating Calcium Transport I (Figure 2D and Figure 2-source data 2). IPA upstream regulator analysis predicted an activated state for the transcriptional regulator *Smarca4* and the two growth factors TGFB1 and GDF2 (Figure 2-source data 3). Intracellular calcium signaling in pancreatic duct cells is an important regulator of homeostatic bicarbonate secretion (19). *PPP1R1B*, *SMARCA4*, and *TGFB1* have well described roles in the pathogenesis of pancreatic cancer (20-22). We observed expression of markers of cluster 1, *Anxa3* and *Pah*, which are also DEGs in cluster 4, to have cytoplasmic protein expression in all mouse and human pancreatic duct types (Figure 2-figure supplement 2A-B and Supplementary file 1). Co-staining of CFTR, a marker of cluster 0, and ANXA3 show both overlapping and non-overlapping patterns of expression in human intercalated ducts, validating the heterogeneity observed in our murine pancreatic duct dataset in human pancreatic duct cells (Figure 2-figure supplement 2C).

Cluster 2 is characterized by low level or lack of expression of multiple ductal cell markers (*Cftr*, *Kcne3*, *Sparc*, *Mmd2, Krt7*) found in other clusters (Figure 2B-C and Figure 1-figure supplement 1G). Cluster 2 has the lowest average expression of total genes and transcripts (Figure 1-figure supplement 1F and Figure 2-source data 4). We therefore posit that cluster 2 represents a stable, fairly transcriptionally and metabolically inactive duct cell subpopulation when compared to other duct clusters. Cluster 3 cells are located almost entirely within cluster 8 of the UMAP containing 16 DBA^+ clusters (Figure 1-figure supplement 1E). This, along with high expression of genes regulating cilia biogenesis (*Foxj1*, *Cfap44, Tuba1a*) led to the identification of cluster 3 as intrapancreatic bile duct cells (Figure 2A-C and Figure 2-source data 1). Expression of cilia biogenesis genes is more prominent in intrapancreatic bile duct cells when compared to pancreatic duct cells (Figure 2-figure supplement 2D, Figure 2-source data 1, and data not shown).

Cells in cluster 4 have significantly higher expression of *Tgfb3* and *Dclk1* when compared to all other ductal clusters (Figure 2C and Figure 2-source data 1). DCLK1 labels tuft cells which are present in normal murine intrapancreatic bile ducts, pancreatobiliary ductal epithelium (23), and rare normal murine pancreatic duct cells (24). YAP1, a transcriptional regulator essential for homeostasis of biliary duct cells (25), was predicted to be in an activated state by IPA upstream regulator analysis (Figure 2-source data 3). Cluster 4 also contained a small population (13 cells) of *Dmbt1* and *Ly6d*-expressing cells previously identified in extrahepatic biliary epithelium (25) (Figure 2-figure supplement 3A). These 13 cells appeared as a small population separate from other cells in cluster 4 in the UMAP (Figure 2A). Similar to the immunofluorescence (IF) validation reported for extrahepatic biliary epithelial cells (BECs) (25), our IF assessment of CXCL5, another maker of the *Dmbt1* and *Ly6d*-expressing subpopulation, showed a greater abundance of these cells than what would be expected given the number identified in the clustering analysis (13). It is possible that this cell type is sensitive to single cell dissociation. Cells in cluster 4 are juxtaposed to pancreatic duct cells (clusters 0, 1, and 2) in the UMAP, suggesting transcriptional commonalities with pancreatic duct cells. In addition, *Dmbt1* and *Ly6d*-expressing cells are present in cluster 4, suggesting a bile duct identity. Based on these shared features of bile and pancreas ducts, we postulate that cluster 4 contains pancreatobiliary duct cells.

Replicating duct cells are characterized by high expression of *Mki67*, *Cenpf*, and *Cenpe* and comprise 1.65% of all duct cells in our dataset (Figure 2A-C, Figure 2-figure supplement 2D, and Figure 2-source data 1). Consistent with previous reports (26, 27), pancreatic duct cells are fairly mitotically inactive.

Summarily, our high-resolution single cell analysis has identified the substructure of murine pancreatic duct cells and characterized pancreatobiliary and intrapancreatic bile duct cells.

Comparison of clusters defines heterogeneity within duct subpopulations

We next sought to determine the relationships between duct clusters by examining their similarities and differences. Dendrogram analysis, Pearson's correlation, and DEGs revealed close relationships between clusters 0 and 2 as well as clusters 1 and 4 (Figure 3A-B and Figure 3-source data 1). Comparison of clusters 0 and 2 showed only 9 significant DEGs, suggesting a shared core gene expression program (Figure 3C-D). Overrepresentation of molecules regulating the cell cycle was observed in cluster 0 when compared to cluster 2 (Figure 3E). The DEGs upregulated in cluster 0 promote duct cell function (*Cftr*, *Tuba1a*, *Kcne3*), suggesting that cluster 0 comprises workhorse pancreatic duct cells (28).

When comparing pancreatobiliary cells of cluster 4 to pancreatic duct cells in cluster 1, one of the most striking differences is the enrichment in expression of genes regulating assembly of cell junctions including tight junctions, epithelial adherens junction signaling,

regulation of actin-based motility by Rho, and actin cytoskeleton signaling. A strong network of stress fibers, comprised of actin filaments, myosin II, and other proteins, that function in bearing tension, supporting cellular structure, and force generation may be important for pancreatobiliary cell function and maintenance (Figure 3F-H and Figure 3-source data 2-3) (29, 30). Cluster 4: *Dmbt1*$^{+}$*Ly6d*$^{+}$ cells are characterized by strong upregulation of pathways regulating Xenobiotic metabolism when compared to all other cluster 4 cells suggesting a prominent role for these cells in the bile acid and xenobiotic system (BAXS) (Figure 3I-K, and Figure 3-source data 2-3) (31). Comparison of intrapancreatic bile duct cells and pancreatobiliary cells showed many unique features of these populations including upregulation of EIF2 signaling in pancreatobiliary cells and upregulation of coronavirus pathogenesis pathway in intrapancreatic bile duct cells (Figure 2-figure supplement 3B-D and Figure 2-figure supplement 3-source data 1-3).

Pancreatobiliary cells express a gene signature enriched in several targets of the Hippo signaling pathway YAP

Two subpopulations of adult murine hepatic homeostatic BECs, A and B, have been previously described (25). To determine if these subpopulations are present in intrapancreatic bile duct (cluster 3) and pancreatobiliary cells (cluster 4), we aligned our dataset with an adult hepatic murine BEC scRNA-seq dataset comprised of 2,344 homeostatic BECs (25). Intrapancreatic bile duct and pancreatobiliary cells aligned well with hepatic BECs, with no apparent batch effect (Figure 3-figure supplement 1A). Intrapancreatic bile duct cells primarily cluster together with hepatic BECs expressing subpopulation B genes, and pancreatobiliary cells primarily cluster together with hepatic BECs expressing subpopulation A genes (Figure 3-figure supplement 1B-G and Figure 3-figure supplement 1-source data 1-2). The subpopulation A expression signature contains many genes significantly enriched as YAP targets, a signature that has been previously proposed to reflect a dynamic BEC state as opposed to defining a unique cell type (25).

Ductal subpopulations are conserved and evident during pancreas development

To investigate whether pancreas ductal subpopulations become evident during development, we analyzed 10X Genomics single cell published datasets of epithelial-enriched pancreas cells at E12.5, E14.5, and E17.5 (32). We found distinct subpopulations of ductal cells that notably overlap in expression of key marker genes associated with adult pancreas ductal subpopulations (Figure 3-figure supplement 2A-L). As we expected, clear patterns of marker gene expression associated with adult clusters manifest at later stages of development (Figure 3-figure supplement 2D,H,L). Since the developmental biology samples were obtained from Swiss Webster mice, our results suggest the subpopulations of adult pancreas duct cells we describe in C57BL/6J mice are conserved.

DBA^{+} lectin murine pancreas sorting identifies previously missed ductal subpopulations

To determine the novelty of adult duct cell heterogeneity manifested using DBA^{+} lectin sorting of murine pancreas, we next compared our DBA^{+} murine pancreatic ductal clusters to previously reported subpopulations of mouse and human pancreas duct cells. Using inDrop and an islet isolation pancreas preparation, Baron *et al.* (2016) identified the substructure of mouse and human pancreatic duct cells (9). Two subpopulations of mouse pancreatic duct cells characterized by expression of *Muc1* and *Tff2* (subpopulation 1) and *Cftr* and *Plat* (subpopulation 2) were described. While *Cftr* expression is characteristic of our cluster 0 (Figure 2C), *Muc1*, *Tff2*, and *Plat* expression didn't typify any murine DBA^{+} pancreatic duct

subpopulation (Figure 3-figure supplement 2M). Two subpopulations were similarly described for human pancreas duct cells characterized by expression of 1) *TFF1*, *TFF2*, *MUC1, MUC20,* and *PLAT* and 2) *CFTR* and *CD44*. *Tff1* is not expressed in murine DBA$^+$ ductal cells (clusters 0-5). *Cd44* is significantly upregulated in pancreatobiliary cells, and *Muc20* as well as *Tff2* are significantly upregulated in 4:*Dmbt1*$^+$*Lyd6*$^+$ cells (Figure 2-source data 1, Figure 3-source data 1, and Figure 3-figure supplement 2M-N). Dominic Grün *et al*. (2016) previously reported 4 subpopulations of human pancreatic duct cells characterized by expression of *CEACAM6*, *FTH1*, *KRT19*, and *SPP1* using an islet isolation pancreas preparation and the CEL-seq protocol (10). While *Spp1* is significantly upregulated in DBA$^+$ pancreas duct clusters 0 and 2, *Fth1* doesn't characterize any murine DBA$^+$ pancreas duct population, and *Krt19* is significantly upregulated in pancreatobiliary cells (Figure 2-source data 1, Figure 1-figure supplement 1G, and Figure 3-figure supplement 2O). *CEACAM6* has no mouse homolog. The differences in pancreatic ductal subpopulation identification may be due to single cell methodology (inDrop, CEL-seq, and 10X Genomics), pancreas preparation method (islet isolation vs DBA$^+$ lectin sorting), differences in ductal cell numbers analyzed, or potential differences between mouse and human duct cells.

Six subpopulations of human pancreatic duct cells have been described using the 10X Genomics platform based on sorting for *BMPR1A*/*ALK3* (8). Using AddModuleScore in Seurat, we calculated a score comparing each of our murine duct clusters to the human ALK3$^+$ clusters (Figure 3-figure supplement 3A-F) (33). Murine pancreatic duct clusters 0-2 had the highest scores when compared to human ALK3$^+$ clusters 1 (OPN$^+$ Stress/harboring progenitor-like cells) and 2 (TFF1$^+$ activated/migrating progenitor cells). Murine pancreatobiliary cells (cluster 4) scored the highest when compared to the human ALK3$^+$ cluster 3 (AKAP12$^+$ small ducts). The human ALK3$^+$ cluster 4 (WSB1$^+$ centroacinar cells) didn't distinguishably overlap with any DBA$^+$ mouse pancreas ductal clusters. DBA is expressed in murine centroacinar/terminal ducts as early as three weeks of age (34), thus these cells would be expected to be present in our dataset (11). Examination of expression of centroacinar/terminal ductal cell markers *Hes1* (35), *Aldh1a1* (36), and *Aldh1b1* (37) in our dataset showed broad expression enriched in either clusters 0 and 2 (*Hes1* and *Aldh1b1*) or clusters 1 and 4 (*Aldh1a1*), rather than a distinct subpopulation as is seen in the ALK3$^+$ human single-cell RNA sequencing pancreatic duct dataset. *Aldh1a7* is negligibly expressed in murine duct clusters 0-5 (Figure 3-figure supplement 3G). Unlike in mouse DBA$^+$ pancreas ductal clusters, the human ALK3$^+$ dataset contains two ducto-acinar subpopulations characterized by expression of genes enriched in acinar cells. To assess the presence of ducto-acinar cells in adult murine pancreas, we performed immunolabeling for markers of the ALK3$^+$ human ducto-acinar clusters 5 (*CPA1*) and 6 (*AMY2A* and *AMY2B*). Although ducto-acinar cells, like centroacinar/terminal ductal cells, don't define a unique cluster in our DBA$^+$ murine duct subpopulations, we identified DBA$^+$CPA1$^+$ and DBA$^+$α-amylase$^+$ ducto-acinar cells in adult murine pancreas (Figure 3-figure supplement 3H). Taken together, these data suggest murine centroacinar/terminal ductal and ducto-acinar cells are largely transcriptionally homogenous with other murine duct cell types.

RaceID3/StemID2 suggest murine DBA$^+$ duct cluster 0 and 2 cells are the most progenitor-like

Given the close relationships observed between DBA$^+$ duct clusters 0 and 2 as well as 1 and 4, we next assessed differentiation potential using RaceID3/StemID2 to predict cell types, lineage trajectories, and stemness (38). Unsupervised clustering with RaceID3 generated 17 clusters. RaceID3 clusters with 10 cells or less were removed from subsequent analyses, and Seurat duct clusters 3 and 5 are not included in this analysis (Figure 4A-B). RaceID3 clusters with the highest StemID2 score correlate to cells present in Seurat duct clusters 0 and 2 (Figure 4C and Figure 5-figure supplement 1A-B). The variable StemID2 scores observed for cells within Seurat

duct clusters 0, 1, 2, and 4 suggest distinct stages of differentiation or maturation. Consistent with previous literature, the pancreatic ductal cell progenitor niche isn't restricted to a single cluster (8).

Pseudotime ordering identifies an epithelial-mesenchymal transition (EMT) axis in pancreatic duct cells

To further examine the lineage relationships among pancreas duct subpopulations, we ordered cells in pseudotime based on their transcriptional similarity (39). Monocle 3 analysis suggested DBA^+ duct clusters 3 and 5 were disconnected from the main pseudotime trajectory, so we focused our analysis on DBA^+ duct clusters 0, 1, 2, and 4 (Figure 5-figure supplement 1C). Because RaceID3/StemID2 analysis suggested Seurat clusters 0 and 2 have the highest StemID scores, we started the pseudotime ordering beginning with cluster 0 as Seurat clusters 0 and 2 are juxtaposed in the Monocle 3 clustering (Figure 4D-E and Figure 5-figure supplement 1D).

In Monocle 3 analysis, genes with similar patterns of expression that vary over time across the pseudotime trajectory are coalesced into modules (Figure 5A). We performed IPA and upstream regulator analysis, a pairwise comparison, comparing select clusters within a module to analyze the gene expression changes along the pseudotime trajectory (Figure 5B-D and Figure 5-source data 1-3). Examination of pathways deregulated in modules 4 and 14 showed a shift in the molecules driving the Xenobiotic Metabolism CAR Signaling Pathway. The Xenobiotic nuclear receptor CAR is an important sensor of physiologic toxins and plays a role in their removal (40). The genes highlighted in the Xenobiotic Metabolism CAR Signaling Pathway were *Aldh1b1*, *Aldh1l1*, *Gstt2*/*Gstt2b*, *Hs6st2*, and *Ugt2b7* for clusters 0 and 2 and *Aldh1a1*, *Fmo3*, *Gstm1*, and *Sod3* for cluster 1, suggesting that these clusters might respond differently when exposed to toxins or play heterogenous roles in endogenous toxin elimination (Figure 5B-C).

Regulation of the Epithelial Mesenchymal Transition By Growth Factors Pathway was upregulated in cluster 1 when compared to cluster 0 in Module 34. Molecules altered in this pathway play variable roles in promoting the epithelial or mesenchymal state and include *Fgf12*, *Fgfr2*, *Fgfr3*, *Pdgfc*, and *Smad3* (Figure 5D). When comparing clusters 0 and 1, examination of EMT markers *Vim* and *Cdh1* showed a stronger probability of expression of *Cdh1* in cluster 1 and a stronger probability of expression of *Vim* in cluster 0 (Figure 5E). Using IF, we detected $vimentin^+$, $SNAI1^+$, and $fibronectin^+$ ductal cells in both mouse and human pancreas, providing evidence for this epithelial-mesenchymal transitional axis (Figure 5F, Figure 5-figure supplement 1E, and data not shown).

Osteopontin is required for mature human pancreas duct cell identity

Our analysis thus far reveals multiple transcriptional programs expressed by murine pancreatic duct cells and predicts possible lineage relationships among them. Amidst the duct subpopulation markers, *Spp1* and *Anxa3* caught our eye due to their known roles in pancreatic cancer progression (41-43); however, their functions in normal pancreatic duct epithelium have not been fully explored. *Spp1*, a marker of clusters 0 and 2, has been shown by us and others to mark a pancreas duct cell type enriched in progenitor capacity (8, 44). *Anxa3*, a marker of clusters 1 and 4, inhibits phospholipase A2 and cleaves inositol 1,2-cyclic phosphate generating inositol 1-phosphate in a calcium dependent manner (45, 46). *Gmnn* expression is highly conserved and plays crucial roles in development biology, yet it's function in normal pancreatic duct cells is incompletely understood (47). *Gmnn*, a marker of cluster 0, acts to inhibit re-replication of DNA during DNA synthesis by inhibiting the prereplication complex (48, 49). Understanding the function of a gene in normal physiology is central to dissecting its role in

disease. To get a better understanding of the function of these subpopulation markers in normal human pancreatic duct cells, we next examined the consequences of their loss in the immortalized human pancreatic duct cell line HPDE6c7 (50). HPDE6c7 cells demonstrate several features of normal pancreatic duct epithelium including gene expression of *MUC1*, *CA2*, and *KRT19* and have been used in many investigations as an *in vitro* model of "near normal" human pancreatic duct cells (50-53). We generated and validated *SPP1*, *GMNN*, and *ANXA3* knockout HPDE6c7 lines using CRISPR/Cas9 (Figure 6A-C). Strong, consistent phenotypes were observed among different knockout lines for each gene despite some lines not demonstrating full loss of the protein (HPDE6c7 *ANXA3* gRNA2 and HPDE6c7 *SPP1* gRNAs 1-4). Cellular morphology was similar to the scrambled (scr) gRNA control (54) for every knockout line except HPDE6c7 *SPP1* gRNAs 1-4, which displayed a dramatic change in cellular morphology. HPDE6c7 *SPP1* knockout cells showed prominent filipodia and significantly increased proliferation when compared to the HPDE6c7 scr gRNA control, a phenotype suggestive of increased progenitor function (Figure 6D-E). The change in cellular morphology in HPDE6c7 *SPP1* knockout lines is accompanied by decreased duct function as measured by carbonic anhydrase activity (Figure 6F).

To assess the changes in HPDE6c7 *SPP1* knockout lines on a molecular scale, we performed bulk RNA-sequencing on all 4 HPDE6c7 *SPP1* knockout lines and the HPDE6c7 scr gRNA control. A significant increase in markers associated with epithelial-mesenchymal transition (EMT) (*VIM*, *ZEB1*, *TWIST1*, *MMP2*) was observed in HPDE6c7 *SPP1* knockout lines when compared to the control (Figure 6-source data 1-3, Figure 6G-H, and Figure 6-figure supplement 1A-C). Markers of mature duct cells (*HNF1B*, *SOX9*, *KRT19*) were significantly downregulated in HPDE6c7 *SPP1* knockout lines when compared to the control (Figure 6G, I-J and Figure 6-source data 1). Gene Set Enrichment Analysis (GSEA) suggested positive enrichment of pathways that regulate embryogenesis (Hox genes and Notch signaling) and cell cycle regulation in HPDE6c7 *SPP1* knockout lines when compared to the HPDE6c7 scr gRNA control (Figure 6-figure supplement 1D-F). Additionally, qPCR analysis demonstrated a significant increase in pancreatic progenitor markers (55, 56) in HPDE6c7 *SPP1* knockout lines when compared to the HPDE6c7 scr gRNA control, supporting the notion that loss of OPN leads to a more immature, progenitor-like state (Figure 6K). Taken together, these results define unique functional properties for markers that characterize murine DBA^+ pancreas duct cells and suggest that *SPP1* is an essential regulator of human pancreatic duct cell maturation and function.

Transdifferentiation of pancreatic duct cells to endocrine cells at early postnatal stages and in pancreatic injury models has been suggested by several studies (57, 58). To query whether HPDE6c7 *SPP1* knockout progenitor-like, dedifferentiated duct cells harbor the capacity to redifferentiate to endocrine cells *in vivo*, we injected HPDE6c7 *SPP1* knockout cell lines and HPDE6c7 scr gRNA control cells subcutaneously into NSG mice (Figure 7A). After 5 days post-injection, α-amylase$^+$ $CK19^+$ double positive cells were evident in HPDE6c7 scr gRNA control cells, but not in HPDE6c7 *SPP1* knockout cells (Figure 7B). This observation is consistent with the previously described ducto-acinar axis characteristic of human pancreatic duct cells (8). We observed Neurogenin-$3^+$$SOX9^+$ co-positive HPDE6c7 *SPP1* knockout cells, suggesting potential for differentiation towards the endocrine lineage (Figure 7C). We detected Synaptophysin$^+$ Glucagon$^+$ as well as Synaptophysin$^+$ C-peptide$^+$ double positive HPDE6c7 *SPP1* knockout cells (Figure 7D-E). Expression of endocrine markers in subcutaneously injected HPDE6c7 scr gRNA cells was not observed at Day 5 post-injection. A small subset of C-peptide$^+$ HPDE6c7 *SPP1* knockout cells express both PDX1 and NKX6.1 (Figure 7E-F). Together, these data point to previously unappreciated roles for *SPP1* in maintaining duct cell properties and preventing changes in cell identity.

Geminin safeguards against accumulation of DNA damage in mouse ductal cells in the setting of chronic pancreatitis

One marker of the workhorse population of pancreatic duct cells *Gmnn* has previously been associated with chronic inflammatory diseases such as asthma (59). We therefore queried its role in pancreas inflammatory disease. *Gmnn* binds to CDT1 and inhibits DNA replication during the S phase. Geminin is a crucial regulator of genomic stability; its inhibition in multiple cancer cell lines leads to DNA re-replication and aneuploidy (60, 61). To determine the requirement for *Gmnn* in normal homeostatic pancreatic ductal cells, we generated a conditional *Gmnn* floxed allele and crossed the mouse to the *Sox9-CreERT2* (62) and *Hnf1b*^{CreERT2} (63) transgenic lines (Figure 8-figure supplement 1). Adult mice, between the ages of 7-9 weeks, were injected with tamoxifen to ablate Geminin in mouse pancreatic duct cells. Tamoxifen injected *Sox9-CreERT2*$^{Tg/wt}$; *Geminin*$^{f/f}$, *Sox9-CreERT2* $^{Tg/wt}$; *Geminin*$^{f/wt}$, and *Hnf1b*$^{CreERT2\ Tg/wt}$; *Geminin*$^{f/f}$ mice displayed no histological abnormalities as assessed by hematoxylin and eosin (H&E) staining and no significant alterations in DNA damage as assessed by ATR and γ-H2AX IF up to 6 months post tamoxifen injection (data not shown). We were unsurprised by these findings, given the low proliferation rate of murine pancreatic duct cells suggested by our single cell data. Thus, Geminin may only be required in the context of pathologies characterized by increased proliferation in the pancreas such as pancreatitis or PDA (64).

We examined proliferation in human pancreas duct cells in CP patients (N=5 patients) and found a significant increase in geminin expression when compared to normal human pancreatic duct cells (N=10 donors) (Figure 8A-B). Pancreatic duct ligation (PDL), an experimental technique that recapitulates features of human gallstone pancreatitis, results in an increase in proliferation of rat pancreatic duct cells (65, 66). To investigate the role of Geminin in mouse pancreatic duct cells in the setting of CP, we performed PDL on *Sox9-CreERT2*$^{Tg/wt}$; *Geminin*$^{f/f}$, *Sox9-CreERT2*$^{Tg/wt}$; *Geminin*$^{f/wt}$, *Hnf1b*$^{CreERT2\ Tg/wt}$; *Geminin*$^{f/f}$ and littermate control mice (Figure 8C). As in the human setting, we also observed upregulation of Geminin in ductal epithelium in the control PDL mouse group (Figure 8D). Previously reported features of the PDL model were evident in our transgenic mice including replacement of parenchymal cells with adipose tissue, inflammation, and fibrosis (67, 68) (Figure 8-figure supplement 2A-B). Significant attenuation of Geminin expression was observed in *Sox9-CreERT2*$^{Tg/wt}$; *Geminin*$^{f/f}$, *Sox9-CreERT2*$^{Tg/wt}$; *Geminin*$^{f/wt}$, and *Hnf1b*$^{CreERT2\ Tg/wt}$; *Geminin*$^{f/f}$ mouse pancreatic duct cells when compared to controls (Figure 8D and Figure 8-figure supplement 3A). Homozygous *Gmnn* loss in SOX9^{+} pancreatic ductal cells promoted an acute increase in proliferation, as assessed by BrdU incorporation, at Day 7 which became insignificant at Day 30 (Figures S12B-E). No changes were observed in apoptosis for any model or time point when compared to controls as assessed by cleaved caspase-3 IF (data not shown). Examination of duct cell DNA damage by γ-H2AX IF showed significantly increased γ-H2AX foci in *Sox9-CreERT2*$^{Tg/wt}$; *Geminin*$^{f/f}$ mice at Day 7, an observation that was sustained at Day 30 (Figure 8E-H). Assessment of DNA damage in *Sox9-CreERT2*$^{Tg/wt}$; *Geminin*$^{f/f}$, *Sox9-CreERT2*$^{Tg/wt}$; *Geminin*$^{f/wt}$, and *Hnf1b*$^{CreERT2\ Tg/wt}$; *Geminin*$^{f/f}$ mice by ATR IF showed no significant changes (data not shown). The lack of phenotypes observed in the *Hnf1b*$^{CreERT2\ Tg/wt}$; *Geminin*$^{f/f}$ model may be due to differences in recombination induced by the *Sox9-CreERT2* and *Hnf1b*^{CreERT2} lines, since fewer cells of the pancreatic ductal epithelium express HNF1B (Figure S1G and Figure 8C). Taken together, these data suggest Geminin is an important regulator of genomic stability in pancreatic ductal cells in the setting of CP.

DISCUSSION

We present a single cell transcriptional blueprint of murine pancreatic duct cells, intrapancreatic bile duct cells, and pancreatobiliary cells. Notably, our single cell analysis indicated that

endothelial cells, fibroblasts, and immune cells are also obtained using the DBA^+ lectin sorting strategy (12), and suggests that a subsequent ductal purification step is required to obtain pure pancreatic duct cells using this protocol. A static transcriptional picture in time has highlighted a very dynamic view of pancreas duct cell heterogeneity. Our study provokes reinterpretation of several previously published lineage tracing reports using ductal-specific Cre mouse lines, and will help plan future lineage tracing studies.

Cluster 0 workhorse pancreatic duct cells comprise the largest pancreatic duct subpopulation identified. Although clusters 0 and 2 share many markers, we found compelling differences in metabolic states as manifested in part by an overall lower gene and transcript count for cluster 2. IPA suggested that subpopulations of pancreatic duct cells may use different predominant mechanisms for bicarbonate secretion such as CFTR (69) for cluster 0 and calcium signaling for cluster 1 (70). One notable difference between clusters 0 and 2 vs 1 is the molecules which regulate the Xenobiotic Metabolism CAR signaling pathway. We observed expression of several genes, whose alteration contributes to PDA progression including *Tgfb2* and *Ctnnb1* in cluster 0 and *Ppp1r1b, Smarca4,* and *Tgfb1* in cluster 1. IPA upstream regulator analysis of Monocle 3 Module 14 predicted significant inhibition of *Kras* in cluster 1 when compared to cluster 0. Additionally, IPA upstream regulator analysis comparing cluster 2 vs 0 in Module 19 predicted activation of *Myc* and *Mycn* in cluster 2. These genes play central roles in homeostasis of pancreatic duct cells, and it's possible that distinct ductal cell subpopulations which are actively expressing these pathways may have different predispositions to PDA with mutations in these genes, heterogeneity which may also contribute to development of different subtypes of PDA.

The role of *Spp1* in homeostatic pancreatic ductal cells has been elusive, since *Spp1* knockout mice have no apparent pancreatic duct phenotypes (44). We identified an EMT axis in pancreatic duct cells using Monocle 3 and validated this observation in mouse and human duct cells. *Spp1* is one gatekeeper of this epithelial to mesenchymal transitory duct phenotype as manifested by loss of ductal markers, reduced duct function, and upregulation of EMT genes in HPDE6c7 *SPP1* knockout cells when compared to controls. Clusters 0 and 2, characterized by strong expression of *Spp1*, show the highest StemID2 scores. *SPP1* knockout HPDE6c7 cells display prominent filipodia and the highest proliferative capacity of all markers examined when compared to controls. Taken together, these phenotypes along with upregulation of pathways regulating mammalian development (Notch signaling and Hox genes) manifested by GSEA suggest *SPP1* loss promotes human duct cell dedifferentiation.

During pancreas development, the multipotent epithelial progenitors become increasingly compartmentalized into tip and trunk progenitors that give rise to acinar and endocrine/ductal cells, respectively (71). Our data suggest that OPN-deficient HPDE6c7 cells dedifferentiate into a trunk, and not tip, progenitor-like cell and that redifferentiation of HPDE6c7 cells to a human pancreatic duct or acinar cell lineage isn't favored *in vivo* following OPN loss. These data underscore the requirement for OPN expression for mature human pancreas duct cell identity. It has been hypothesized that *SPP1*'s role in mature pancreatic duct cells is evident during pathogenesis. Several groups have already nicely shown that *Spp1* plays important roles in pancreatic pathologies including PDA (42, 72). In human pancreas duct cells, the subpopulation characterized by *SPP1* expression is described as "stress/harboring progenitor-like cells" (8). We observed significant deregulation of 14 cancer-related IPA pathways for which pathway directionality was known in HPDE6c7 *SPP1* knockout lines vs HPDE6c7 scr gRNA controls. 13/14 of these cancer-related pathways, including Pancreatic Adenocarcinoma Signaling, were in a direction suggestive that *SPP1* loss protects against tumor progression in human pancreatic duct cells. These findings are in agreement with published studies suggesting that *SPP1* loss ameliorates aggressiveness of pancreatic cancer cells (41, 42) and colon cancer cells (72, 73).

The requirement for Geminin in prevention of DNA re-replication initiation has been postulated to be when cells are stressed to divide quickly (74). We were unable to detect DNA damage with Geminin loss in homeostatic pancreatic duct cells, which may be due to the low proliferation rate of pancreatic duct cells and/or the presence of compensatory mechanisms with redundant function, such as ubiquitin-dependent degradation of CDT1 at the time of replication licensing (75-78). Compensatory mechanisms are not sufficient to rescue the effects of Geminin loss in pancreatic duct cells in the context of CP, the result of which is accumulation of sustained DNA damage evident by γ-H2AX, but not ATR labeling. It has been previously reported that ATR is activated in Geminin-depleted colon cancer cell lines (79). Activation of the ATR-CHK1 pathway isn't a major player in pancreatic duct cells in the setting of CP (80), suggesting different mechanisms participate in sensing Geminin depletion-induced DNA damage in different experimental systems and tissues. While our limited functional analyses of the *SPP1* and *Gmnn* mutant models provide important information regarding their function in pancreas duct cells, additional studies will be required to fully understand their roles in normal and disease pancreatic physiology.

MATERIALS AND METHODS

Preparation of pancreatic duct cells for single cell analysis

Pancreata from four 9 week old female C57BL/6J littermates (Jackson Labs, Stock 000664) were dissected, digested into single cells, and the DBA^+ fraction obtained as previously described (12). Subsequently, live DBA^+ cells were isolated for scRNA-seq by excluding propidium iodide (Thermo Fisher Scientific, P3566) positive single cells during FACS. scRNA-seq was performed by the Institute for Human Genetics Genomics Core Facility at University of California San Francisco (UCSF) using the 10X Genomics platform. Briefly, live, single, DBA^+ pancreatic cells were loaded onto the microfluidic chip to generate single cell GEMs (Gel Bead-In EMulsions). Following cell lysis and unique barcode labeling, the cDNA library of 18,624 live pancreatic cells was generated using the Chromium Single Cell 3' GEM, Library & Gel Bead Kit v2 (10X Genomics). The cDNA library was sequenced on one lane using an Illumina HiSeq 4000.

Single cell RNA-seq data processing

scRNA-seq data was generated on the 10X platform (10X Genomics, Pleasanton, CA) according to Single Cell 3' protocol (v2 Chemistry) recommended by the manufacturer (81). The Cell Ranger software pipeline (version 2.1.1) was used to demultiplex cellular barcodes, map reads to the genome and transcriptome using the STAR aligner, and produce a matrix of gene counts versus cells. Doublets were filtered by excluding cells having RNA counts > 30000 and mitochondrial genes percentage > 10% in addition to using Scrublet (82). The R package Seurat (83) was used to process the unique molecular identifier (UMI) count matrix and to perform data normalization (gene expression measurements for each cell were normalized by total expression, and log-transformed), dimensionality reduction, clustering, ductal cell isolation, and differential expression analysis. We identified three clusters enriched in genes from 2 different cell types including: 1) acinar and T cell, 2) acinar cell and macrophage and 3) acinar cell and duct cell. Because our dataset doesn't contain a population of acinar cells (they aren't DBA^+), doublet detector algorithms won't remove acinar cell doublets from our dataset. Based on this reasoning, we removed these clusters containing a high threshold level of expression of acinar cell genes.

Generation of *Gmnn* conditional floxed allele

The general strategy to achieve Cre recombinase-mediated conditional gene ablation was to flank exons 3 and 4 of *Mus musculus Gmnn* by loxP sites (Figure 8-figure supplement 1A). The

arms of homology for the targeting construct were amplified from BAC clone RP23-92G13 by PCR with high fidelity Taq polymerase. One primer contained a loxP site and a single SphI site which was used to verify the presence of the loxP site associated with it. Finally, the selectable cassette *CMV-hygro-TK* was incorporated into the targeting vector. The selectable marker itself was flanked by two additional loxP sites generating a targeting vector containing three loxP sites. Such a strategy allows the generation of ES cells with both a knockout allele and a conditional knockout allele after Cre mediated removal of the selection cassette *in vitro*. The targeting vector was sequenced to guarantee sequence fidelity of exons 3-4 and the proper unidirectional orientation of the three loxP sites. The complete left arm of homology was about 3200bp in length and the right arm of homology was 2100bp in length.

V6.5 ES cells were electroporated (25µF, 400V) with the three loxP sites-containing targeting construct, and hygromycin selection was performed to identify correctly targeted ES cells. Successfully targeted ES cells (3loxP) were identified with Southern blot (Figure 8-figure supplement 1B). These 3loxP ES cells were then electroporated with a Cre-expressing plasmid and counter-selected with ganciclovir. ES cells that contained either one loxP or two loxP sites, respectively, were identified by Southern blot (Figure 8-figure supplement 1C). An ES cell clone was chosen that carried the conditional knockout allele (two loxP sites flanking exons 3 and 4) and was used for blastocyst injections to generate chimeric founder mice. *Gmnn*$^{f/f}$ mice displayed normal litter sizes. For routine genotyping of *Gmnn*$^{f/f}$ mice, the primers GCCTCGAACTCAGAAATCCA (primer A) and AACACAAAATTTGGCCTGCT (primer B) were used. To identify the deleted allele by PCR, primer C (TAGCCCGGACTACACAGAGG) can be used with primer A.

Southern blot

For Southern blotting of genomic DNA, samples were digested with SphI or Bsu36I restriction enzymes for at least 4hrs and separated on an 0.8% agarose gel. The DNA was transferred to a Hybond-XL membrane (GE-Healthcare) in a custom transfer setup. Before assembly, the agarose gel was treated for 15min in depurination solution (21.5ml 37% HCl in 1L H_2O), briefly rinsed in H_2O and then soaked in denaturing solution (20g NaOH pellets, 87.6g NaCl in 1L H_2O) for 30min. After transfer, the DNA was crosslinked to the membrane with UV light. The PCR amplified external Southern blot probes were labeled with ^{32}P using the Prime-It II Random Primer Labeling kit from Stratagene. After hybridization of the probe and washing of the membrane, Kodak MS film was exposed to it and then developed.

Mice

NSG mice from Jackson Labs (Stock 005557) were used. The transgenic mouse strain *Sox9-CreERT2* was obtained from Jackson Labs (Stock 018829), and *Hnf1b*^{CreERT2} has been previously described (63). Mice were maintained on a mixed genetic background. To induce Cre recombination, mice were injected with 6.7mg tamoxifen (Actavis, NDC 0591-2473-30) via oral gavage three different days over the course of a week at 7-9 weeks of age. Pancreatic duct ligations were performed as previously described (84). BrdU (Sigma, B9285-1G) injections were performed 24 hours and 4 hours prior to dissection. 100,000 HPDE6c7 cells mixed 1:1 with media containing cells and matrigel (Corning, 356231) in a total volume of 100uL were injected subcutaneously into NSG mice. Mice were genotyped by PCR or Transnetyx. All animal studies were approved by the Institutional Animal Care and Use Committee at UCSF.

Histology/immunostaining

Tissues were fixed in Z-Fix (Anatech Ltd., 174), processed according to a standard protocol, and embedded in Paraplast Plus embedding agent for histology, with DMSO (VWR 15159-464). For immunostaining, paraffin sections were deparaffinized, rehydrated, and antigen retrieval was performed, for all antibodies except BrdU, with Antigen Retrieval Citra (Biogenex, HK086-

9K) using a heat-mediated microwave method. For immunostaining of BrdU, antigen retrieval was performed as previously described (85). For IHC, endogenous peroxidase activity was blocked by incubation with 3% hydrogen peroxide (Fisher Scientific, H325-100) following antigen retrieval. Primary antibodies were incubated overnight at 4°C. Secondary antibodies were used at 4ug/mL and incubated at room temperature for 1 hour (IHC) or 2 hours (IF). For IF, slides were mounted in ProLong Diamond Antifade Mountant with DAPI (ThermoFisher, P36962). For IHC, Vectastain Elite ABC kit (Vector Laboratories, PK-6100) and DAB Peroxidase (HRP) Substrate kit (Vector Laboratories, SK-4100) were used. Primary antibodies used in this study are listed in Supplementary file 2. Secondary antibodies used in this study were obtained from Invitrogen, Jackson ImmunoResearch, and Biotium.

Immunostaining of cluster markers as well as the types of ducts within the ductal hierarchy tree were reviewed and classified by a board-certified pathologist. For expression analysis of selected markers in murine and human tissues, images shown are representative of at least 3 different donors or 9 week-old C57BL/6J mice. For quantification of BrdU, cleaved caspase 3, Geminin, Ki67, γ-H2AX, and ATR, at least 60 cells from 3 different ducts were analyzed. Quantification of γ-H2AX foci included duct cells with zero foci. For quantification of c-peptide/PDX1/NKX6.1 triple positive cells, at least 3 images containing an average of 188 cells were counted from 3 biological replicates per cell line. Normal human tissue used in this study was obtained from research consented human cadaver donors through UCSF's Islet Production Core. Human pancreatic tissue specimens from five surgical resections from patients without pancreaticobiliary carcinoma or high grade pancreatic intraepithelial neoplasia were obtained. The pancreatic histologic section demonstrated chronic pancreatitis with loss of acinar parenchyma resulting in atrophic lobules along with variable fibrosis and chronic inflammation (most had no to sparse lymphocytic inflammation).

Immunocytochemistry

Cells were grown on coverslips in 6 well plates and fixed at RT for 15 minutes with 4% paraformaldehyde. Cells were permeabilized with permeabilization solution (0.1% w/v Saponin, 5% w/v BSA in PBS-/-). The primary antibody was incubated in permeabilization solution at 4°C overnight. After washing off unbound primary antibody with PBS-/-, the secondary antibody was incubated in permeabilization solution for 1 hour at RT. After washing off unbound secondary antibody with PBS-/-, cells were mounted using ProLong Diamond Antifade Mountant with DAPI (ThermoFisher, P36962).

Flow cytometry

For analysis of cell surface markers (EPCAM), cells were resuspended in FACS buffer (1% FBS + 2mM EDTA in PBS -Mg/-Ca), filtered in FACS tubes with a cell strainer cap, and spun at 1350 rpm for 3-5 min. The supernatant was discarded, and cells were resuspend in 100ul of directly conjugated primary antibody diluted in FACS buffer and stained for 60 min at RT. Stained cells were washed with 2ml FACS buffer and spun at 1350 rpm for 3-5 min. The supernatant was discarded and cells were resuspended in 250uL FACS buffer containing 0.5ug/mL DAPI immediately before analyzing on the flow cytometer.

For analysis of intracellular antigens, single cell suspensions of cell lines were prepared. Cells were washed with PBS -Mg/-Ca, resuspended in 250uL FACS buffer, and filtered in FACS tubes with a cell strainer cap. 2mL 1X permeabilization buffer (Affymetrix eBiosciences, 00-8333-56) was added to cells, and cells were subsequently spun at 1500 rpm for 5 minutes. Supernatant was removed, and 100uL primary antibody diluted in CAS-Block (Invitrogen, 00-8120) + 0.2% TritonX-100 was added to cells. Cells were stained overnight at 4°C. Subsequently, cells were washed with 3mL 1X permeabilization buffer and spun at 1500rpm for 5 minutes. The supernatant was discarded. If using directly conjugated primary antibodies, cells were resuspended in 250uL FACS buffer and analyzed on the flow cytometer. If using

unconjugated primary antibodies, 100uL secondary antibody diluted in CAS-Block + 0.2% Triton X-100 was added to cells, and cells were incubated at 4°C for 50 minutes. Subsequently, 3mL1X permeabilization buffer was added to the cells, and cells were spun at 1500 rpm for 5 minutes. The supernatant was discarded. Cells were resuspended in 250uL FACS buffer and analyzed on the flow cytometer.

RNA-seq and qPCR

RNA was isolated using the RNeasy Mini Kit (Qiagen, 74106) as per manufacturer's instructions. To obtain N=3 for the HPDE6c7 scr gRNA control, RNA was isolated on 3 different days of subsequent passages. For qPCR, cDNA was prepared using the SuperScript III First Strand synthesis kit (Thermo Fisher Scientific, 18080085) using 500 ng of RNA and random hexamers. qPCR was performed using FastStart Universal SYBR Green mix (Sigma, 4913914001). RNA expression of target genes was normalized to GAPDH. qPCR primer sequences are included in Supplementary file 4.

For RNA-seq, a stranded mRNA library prep was prepared using PolyA capture and paired-end sequencing was performed by Novogene. 40 million reads were sequenced for each sample. Quality of raw FASTQ sequences was assessed using FASTQC. To process RNA-Seq libraries, adaptor sequences were trimmed using Cutadapt version 1.14 (requiring a length greater than 10 nt after trimming) and quality-filtered by requiring all bases to have a minimum score of 20 (-m 20 -q 20). Only reads that passed the quality or length threshold on both strands were considered for mapping. Reads were aligned to the human genome GRCh38 (hg38) with the STAR Aligner (version 020201). Ensembl reference annotation version 89 was used to define gene models for mapping quantification. Uniquely mapped reads for each gene model were produced using STAR parameter "--quantMode GeneCounts." Differential expression analysis was performed in R using DESeq2 (v.1.16.0) with the default parameters, including the Cook's distance treatment to remove outliers. The RNA-seq and scRNA-seq datasets were deposited to GEO (GEO accession #GSE159343).

Cell culture assays

HPDE6c7 cells (RRID:CVCL_0P38) were authenticated by ATCC and tested negative for mycoplasma using a kit from InvivoGen (rep-pt1). HPDE6c7 cells (51) were cultured in DMEM (Life Technologies 11995073), 10% FBS (Corning, 35011CV), 1X Penicillin : Streptomycin solution (Corning, 30-002-CI). For cell counting, 25,000 cells were seeded in a sterile 6-well TC-treated plate (Corning, 353046). Values depicted for all cell culture experiments represent the average of at least 3 independent experiments.

For carbonic anhydrase activity assays, cell lysates were prepared using standard protocols and cell lysis buffer (Cell Signaling Technologies, 9803S) containing 100 mM PMSF, 1X cOmplete Protease Inhibitor Cocktail (Roche, 11697498001), and 1X PhosSTOP (Sigma Aldrich, 4906845001). Carbonic anhydrase activity was measured using the Carbonic Anhydrase Activity Assay Kit (Biovision, K472-100). For normalization, equal amounts of protein (10ug) per sample were used in the assay. Protein concentration was determined using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, 23225).

Generation of stable knockout HPDE6c7 cell lines

For generation of stable knockouts, gRNAs were cloned into eSPCas-LentiCRISPR v2 (Genscript). gRNA sequences are included in Supplementary file 3. Each gRNA-containing plasmid was incorporated into lentivirus. HPDE6c7 cells were transduced with these lentiviruses, and cells expressing the gRNA-containing plasmid were selected for with puromycin. All cell culture experiments were performed using bulk transduced HPDE6c7 cells.

Western blotting

Cell lysates were prepared using standard protocols and RIPA buffer (Thermo Fisher Scientific, 89901) containing 100 mM PMSF, 1X cOmplete Protease Inhibitor Cocktail (Roche, 11697498001), and 1X PhosSTOP (Sigma Aldrich, 4906845001). PVDF membranes were incubated with primary antibodies overnight at 4°C. After RT incubation with the appropriate HRP-conjugated secondary antibody for 1 hour, membranes were developed using SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Scientific, 34580).

Bioinformatics and statistical analysis

For cell culture studies, sample size was computed based on accepted scientific standards using a minimum of 2 CRISPR/Cas9-generated KO lines to control for off-target effects. Cell culture experiments were repeated a minimum of 3 independent times. For mouse experiments, sample size was computed based on the number of biological replicates required to obtain statistical significance. For *Gmnn* mouse model studies, mice with relevant genotypes were chosen randomly for PDL or control groups. We used $p \leq 0.05$ as a cutoff for DEG inclusion for IPA and IPA upstream regulator analysis. Due to low cell number and high similarity, some comparisons did not yield an acceptable number of statistically significant DEGs (≤25), and we used a relaxed $p \leq 0.1$ as a cutoff for these in order to identify more targets. GSEA was performed on the identified DEGs with the GSEA software (version 3.0) in the pre-ranked mode, with the Reactome pathway dataset (version 7.2). For analysis of published single cell developmental biology datasets, GSM3140915 (E12.5 SW), GSM3140916 (E14.5 SW), GSM3140917 (E17.5 1 SW), and GSM3140918 (E17.5 2 SW) were used. The two E17.5 datasets were from the same animal and were merged. Ductal clusters were identified by expression of marker genes *Sox9*, *Krt19*, and *Epcam*. Data are presented as mean ± SEM and were analyzed in GraphPad Prism or Microsoft Office Excel. Statistical significance was assumed at a p or q value of ≤ 0.05. P or q values were calculated with a t-test. For interpretation of statistical t-test results, * = p or q value ≤ 0.05, ** = p or q value ≤ 0.01, *** = p or q value ≤ 0.001, and **** = p or q value ≤ 0.0001. For all statistical analyses, outliers were identified and excluded using the Grubbs' outlier test (alpha = 0.05) or ROUT (Q=10%).

ACKNOWLEDGEMENTS

The authors thank Christina S. Chung, Debbie Ngow, and Melissa Campbell for their indispensable help with mouse colony maintenance, genotyping, and technical assistance. The authors wish to acknowledge Luc Baeyens and Michael S. German for helpful intellectual discussions. Graphical illustrations were developed in part by Jimmy Chen using BioRender.com. The *Gmnn* floxed mouse strain was generated in Rudolf Jaenisch's laboratory. This work was supported by a Richard G. Klein Fellowship in Pancreas Development, Regeneration or Cancer (to A.M.H.). A.M.H. was additionally supported by F32 CA221114 and a Hirshberg Foundation for Pancreatic Cancer Research Seed Grant. A.A.R. and G.K.F. were supported by the UCSF Bakar ImmunoX Initiative. E.A.C. was supported by R01 CA222862. Work in M.H.'s laboratory was supported by R01 CA172045 and a grant from the Parker Institute for Cancer Immunotherapy (PICI). M.H. owns stocks/stock options in Viacyte, Encellin, Thymmune, EndoCrine, and Minutia. He also serves as SAB member to Thymmune and Encellin and is co-Founder, SAB and Board member for EndoCrine and Minutia.

REFERENCES

1. J. M. Bailey *et al.,* p53 mutations cooperate with oncogenic Kras to promote adenocarcinoma from pancreatic ductal cells. *Oncogene* **35***,* 4282-4288 (2016).

2. A. Y. L. Lee *et al.*, Cell of origin affects tumour development and phenotype in pancreatic ductal adenocarcinoma. *Gut* **68**, 487-498 (2019).
3. M. Wilschanski, I. Novak, The cystic fibrosis of exocrine pancreas. *Cold Spring Harb Perspect Med* **3**, a009746 (2013).
4. M. V. Apte, J. S. Wilson, M. A. Korsten, Alcohol-related pancreatic damage: mechanisms and treatment. *Alcohol Health Res World* **21**, 13-20 (1997).
5. B. Etemad, D. C. Whitcomb, Chronic pancreatitis: diagnosis, classification, and new genetic developments. *Gastroenterology* **120**, 682-707 (2001).
6. D. Whitcomb, J. Greer, Germ-line mutations, pancreatic inflammation, and pancreatic cancer. *Clin Gastroenterol Hepatol* **7**, S29-34 (2009).
7. P. Dhar, S. Kalghatgi, V. Saraf, Pancreatic cancer in chronic pancreatitis. *Indian J Surg Oncol* **6**, 57-62 (2015).
8. M. M. F. Qadir *et al.*, Single-cell resolution analysis of the human pancreatic ductal progenitor cell niche. *Proc Natl Acad Sci U S A* **117**, 10876-10887 (2020).
9. M. Baron *et al.*, A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. *Cell Syst* **3**, 346-360 e344 (2016).
10. D. Grun *et al.*, De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell Stem Cell* **19**, 266-277 (2016).
11. R. L. Beer, M. J. Parsons, M. Rovira, Centroacinar cells: At the center of pancreas regeneration. *Dev Biol* **413**, 8-15 (2016).
12. M. Reichert *et al.*, Isolation, culture and genetic manipulation of mouse pancreatic ductal cells. *Nat Protoc* **8**, 1354-1365 (2013).
13. H. Rezanejad *et al.*, Heterogeneity of SOX9 and HNF1beta in Pancreatic Ducts Is Dynamic. *Stem Cell Reports* **10**, 725-738 (2018).
14. K. L. Raphael, F. F. Willingham, Hereditary pancreatitis: current perspectives. *Clin Exp Gastroenterol* **9**, 197-207 (2016).
15. W. Shen *et al.*, TGF-beta in pancreatic cancer initiation and progression: two sides of the same coin. *Cell Biosci* **7**, 39 (2017).
16. K. J. Gordon, M. Dong, E. M. Chislock, T. A. Fields, G. C. Blobe, Loss of type III transforming growth factor beta receptor expression increases motility and invasiveness associated with epithelial to mesenchymal transition during pancreatic cancer progression. *Carcinogenesis* **29**, 252-262 (2008).
17. P. W. Heiser *et al.*, Stabilization of beta-catenin induces pancreas tumor formation. *Gastroenterology* **135**, 1288-1300 (2008).
18. M. Reichert, A. K. Rustgi, Pancreatic ductal cells in development, regeneration, and neoplasia. *J Clin Invest* **121**, 4572-4578 (2011).
19. J. Maleth, P. Hegyi, Calcium signaling in pancreatic ductal epithelial cells: an old friend and a nasty enemy. *Cell Calcium* **55**, 337-345 (2014).
20. N. Roy *et al.*, Brg1 promotes both tumor-suppressive and oncogenic activities at distinct stages of pancreatic cancer formation. *Genes Dev* **29**, 658-671 (2015).
21. C. J. David *et al.*, TGF-beta Tumor Suppression through a Lethal EMT. *Cell* **164**, 1015-1030 (2016).
22. A. Tiwari *et al.*, Loss of HIF1A From Pancreatic Cancer Cells Increases Expression of PPP1R1B and Degradation of p53 to Promote Invasion and Metastasis. *Gastroenterology* 10.1053/j.gastro.2020.07.046 (2020).
23. K. E. Delgiorno *et al.*, Identification and manipulation of biliary metaplasia in pancreatic tumors. *Gastroenterology* **146**, 233-244 e235 (2014).
24. C. B. Westphalen *et al.*, Dclk1 Defines Quiescent Pancreatic Progenitors that Promote Injury-Induced Regeneration and Tumorigenesis. *Cell Stem Cell* **18**, 441-455 (2016).

25. B. J. Pepe-Mooney *et al.*, Single-Cell Analysis of the Liver Epithelium Reveals Dynamic Heterogeneity and an Essential Role for YAP in Homeostasis and Regeneration. *Cell Stem Cell* **25**, 23-38 e28 (2019).
26. A. S. Moin, P. C. Butler, A. E. Butler, Increased Proliferation of the Pancreatic Duct Gland Compartment in Type 1 Diabetes. *J Clin Endocrinol Metab* **102**, 200-209 (2017).
27. A. E. Butler *et al.*, Pancreatic duct replication is increased with obesity and type 2 diabetes in humans. *Diabetologia* **53**, 21-26 (2010).
28. M. Hayashi, I. Novak, Molecular basis of potassium channels in pancreatic duct epithelial cells. *Channels (Austin)* **7**, 432-441 (2013).
29. K. Burridge, E. S. Wittchen, The tension mounts: stress fibers as force-generating mechanotransducers. *J Cell Biol* **200**, 9-19 (2013).
30. S. Tojkander, G. Gateva, P. Lappalainen, Actin stress fibers--assembly, dynamics and biological roles. *J Cell Sci* **125**, 1855-1864 (2012).
31. W. Dubitzky, O. Wolkenhauer, K. H. Cho, H. Yokota, *Encyclopedia of systems biology* (Springer Reference, New York, 2013), pp. 4 volumes (xlvii, 2366 pages).
32. L. E. Byrnes *et al.*, Lineage dynamics of murine pancreatic development at single-cell resolution. *Nat Commun* **9**, 3922 (2018).
33. H. Alshetaiwi *et al.*, Defining the emergence of myeloid-derived suppressor cells in breast cancer using single-cell transcriptomics. *Sci Immunol* **5** (2020).
34. B. Z. Stanger *et al.*, Pten constrains centroacinar cell expansion and malignant transformation in the pancreas. *Cancer Cell* **8**, 185-195 (2005).
35. D. Kopinke *et al.*, Lineage tracing reveals the dynamic contribution of Hes1+ cells to the developing and adult pancreas. *Development* **138**, 431-441 (2011).
36. M. Rovira *et al.*, Isolation and characterization of centroacinar/terminal ductal progenitor cells in adult mouse pancreas. *Proc Natl Acad Sci U S A* **107**, 75-80 (2010).
37. E. Mameishvili *et al.*, Aldh1b1 expression defines progenitor cells in the adult pancreas and is required for Kras-induced pancreatic cancer. *Proc Natl Acad Sci U S A* **116**, 20679-20688 (2019).
38. J. S. Herman, Sagar, D. Grun, FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. *Nat Methods* **15**, 379-386 (2018).
39. J. Cao *et al.*, The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).
40. Y. E. Timsit, M. Negishi, CAR and PXR: the xenobiotic-sensing receptors. *Steroids* **72**, 231-246 (2007).
41. A. Kolb *et al.*, Osteopontin influences the invasiveness of pancreatic cancer cells and is increased in neoplastic and inflammatory conditions. *Cancer Biol Ther* **4**, 740-746 (2005).
42. C. R. Adams *et al.*, Transcriptional control of subtype switching ensures adaptation and growth of pancreatic cancer. *Elife* **8** (2019).
43. X. Wan, D. Guo, Q. Zhu, R. Qu, microRNA-382 suppresses the progression of pancreatic cancer through the PI3K/Akt signaling pathway by inhibition of Anxa3. *Am J Physiol Gastrointest Liver Physiol* **319**, G309-G322 (2020).
44. G. Kilic, J. Wang, B. Sosa-Pineda, Osteopontin is a novel marker of pancreatic ductal tissues and of undifferentiated pancreatic precursors in mice. *Dev Dyn* **235**, 1659-1667 (2006).
45. J. F. Tait, C. Smith, L. Xu, B. T. Cookson, Structure and polymorphisms of the human annexin III (ANX3) gene. *Genomics* **18**, 79-86 (1993).
46. V. Gerke, S. E. Moss, Annexins: from structure to function. *Physiol Rev* **82**, 331-371 (2002).
47. P. P. Kushwaha, K. C. Rapalli, S. Kumar, Geminin a multi task protein involved in cancer pathophysiology and developmental process: A review. *Biochimie* **131**, 115-127 (2016).

48. T. J. McGarry, M. W. Kirschner, Geminin, an inhibitor of DNA replication, is degraded during mitosis. *Cell* **93**, 1043-1053 (1998).
49. A. Ballabeni, R. Zamponi, J. K. Moore, K. Helin, M. W. Kirschner, Geminin deploys multiple mechanisms to regulate Cdt1 before cell division thus ensuring the proper execution of DNA replication. *Proc Natl Acad Sci U S A* **110**, E2848-2853 (2013).
50. H. Ouyang *et al.*, Immortal human pancreatic duct epithelial cell lines with near normal genotype and phenotype. *Am J Pathol* **157**, 1623-1631 (2000).
51. T. Furukawa *et al.*, Long-term culture and immortalization of epithelial cells from normal adult human pancreatic ducts transfected by the E6E7 gene of human papilloma virus 16. *Am J Pathol* **148**, 1763-1770 (1996).
52. J. Qian, J. Niu, M. Li, P. J. Chiao, M. S. Tsao, In vitro modeling of human pancreatic duct epithelial cell transformation defines gene expression changes induced by K-ras oncogenic activation in pancreatic carcinogenesis. *Cancer Res* **65**, 5045-5053 (2005).
53. J. Lee *et al.*, Reconstituting development of pancreatic intraepithelial neoplasia from primary human pancreas duct cells. *Nat Commun* **8**, 14686 (2017).
54. K. Suzuki *et al.*, In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144-149 (2016).
55. G. Gu *et al.*, Global expression analysis of gene regulatory pathways during endocrine pancreatic development. *Development* **131**, 165-179 (2004).
56. S. J. Willmann *et al.*, The global gene expression profile of the secondary transition during pancreatic development. *Mech Dev* **139**, 51-64 (2016).
57. S. Bonner-Weir *et al.*, Transdifferentiation of pancreatic ductal cells to endocrine beta-cells. *Biochem Soc Trans* **36**, 353-356 (2008).
58. H. S. Kim, M. K. Lee, beta-Cell regeneration through the transdifferentiation of pancreatic cells: Pancreatic progenitor cells in the pancreas. *J Diabetes Investig* **7**, 286-296 (2016).
59. N. Garbacki *et al.*, MicroRNAs profiling in murine models of acute and chronic asthma: a relationship with mRNAs targets. *PLoS One* **6**, e16509 (2011).
60. W. Zhu, M. L. Depamphilis, Selective killing of cancer cells by suppression of geminin activity. *Cancer Res* **69**, 4870-4877 (2009).
61. S. Saxena, A. Dutta, Geminin-Cdt1 balance is critical for genetic stability. *Mutat Res* **569**, 111-121 (2005).
62. J. L. Kopp *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 (2011).
63. M. Solar *et al.*, Pancreatic exocrine duct cells give rise to insulin-producing beta cells during embryogenesis but not after birth. *Dev Cell* **17**, 849-860 (2009).
64. M. R. Salabat *et al.*, Geminin is overexpressed in human pancreatic cancer and downregulated by the bioflavanoid apigenin in pancreatic cancer cell lines. *Mol Carcinog* **47**, 835-844 (2008).
65. S. Githens, The pancreatic duct cell: proliferative capabilities, specific characteristics, metaplasia, isolation, and culture. *J Pediatr Gastroenterol Nutr* **7**, 486-506 (1988).
66. N. I. Walker, A. W. Pound, An autoradiographic study of the cell proliferation during involution of the rat pancreas. *J Pathol* **139**, 407-418 (1983).
67. A. A. Aghdassi *et al.*, Animal models for investigating chronic pancreatitis. *Fibrogenesis Tissue Repair* **4**, 26 (2011).
68. C. Rastellini *et al.*, Induction of chronic pancreatitis by pancreatic duct ligation activates BMP2, apelin, and PTHrP expression in mice. *Am J Physiol Gastrointest Liver Physiol* **309**, G554-565 (2015).

69. H. Ishiguro *et al.*, CFTR functions as a bicarbonate channel in pancreatic duct cells. *J Gen Physiol* **133**, 315-326 (2009).
70. H. Ishiguro *et al.*, Physiology and pathophysiology of bicarbonate secretion by pancreatic duct epithelium. *Nagoya J Med Sci* **74**, 1-18 (2012).
71. Q. Zhou *et al.*, A multipotent progenitor domain guides pancreatic organogenesis. *Dev Cell* **13**, 103-114 (2007).
72. H. Zhao *et al.*, The role of osteopontin in the progression of solid organ tumour. *Cell Death Dis* **9**, 356 (2018).
73. R. Ishigamori *et al.*, Osteopontin Deficiency Suppresses Intestinal Tumor Development in Apc-Deficient Min Mice. *Int J Mol Sci* **18** (2017).
74. K. A. Barry, K. M. Schultz, C. J. Payne, T. J. McGarry, Geminin is required for mitotic proliferation of spermatogonia. *Dev Biol* **371**, 35-46 (2012).
75. E. E. Arias, J. C. Walter, Replication-dependent destruction of Cdt1 limits DNA replication to a single round per cell cycle in Xenopus egg extracts. *Genes Dev* **19**, 114-126 (2005).
76. S. L. Kerns, S. J. Torke, J. M. Benjamin, T. J. McGarry, Geminin prevents rereplication during xenopus development. *J Biol Chem* **282**, 5514-5521 (2007).
77. A. Li, J. J. Blow, Cdt1 downregulation by proteolysis and geminin inhibition prevents DNA re-replication in Xenopus. *EMBO J* **24**, 395-404 (2005).
78. D. Maiorano, L. Krasinska, M. Lutzmann, M. Mechali, Recombinant Cdt1 induces rereplication of G2 nuclei in Xenopus egg extracts. *Curr Biol* **15**, 146-153 (2005).
79. J. J. Lin, A. Dutta, ATR pathway is the primary pathway for activating G2/M checkpoint induction after re-replication. *J Biol Chem* **282**, 30357-30362 (2007).
80. J. Smith, L. M. Tho, N. Xu, D. A. Gillespie, The ATM-Chk2 and ATR-Chk1 pathways in DNA damage signaling and cancer. *Adv Cancer Res* **108**, 73-112 (2010).
81. G. X. Zheng *et al.*, Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049 (2017).
82. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* **8**, 281-291 e289 (2019).
83. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502 (2015).
84. S. De Groef *et al.*, Surgical Injury to the Mouse Pancreas through Ligation of the Pancreatic Duct as a Model for Endocrine and Exocrine Reprogramming and Proliferation. *J Vis Exp* 10.3791/52765, e52765 (2015).
85. S. Puri *et al.*, Replication confers beta cell immaturity. *Nat Commun* **9**, 485 (2018).

Main figure captions

Figure 1. Transcriptomic map of DBA$^+$ pancreatic cells. A) Schematic of experiment workflow. B) The UMAP depicts murine pancreatic DBA$^+$ cells obtained using the protocol. C) A matrix plot shows average expression of ductal cell markers in all clusters, identifying clusters 0 and 8 as ductal cells. D) Feature plots illustrate markers of various cell types including epithelial (*Epcam*), ductal (*Krt19* and *Sox9*), CD45$^+$ immune cells (*Ptprc*), endothelial cells (*Pecam1*), fibroblasts (*Col1a1*), endocrine cells (*Chga*), and acinar cells (*Pnliprp1*). We observed low level expression of acinar cell markers uniformly across all clusters that is likely contaminating acinar cell mRNA.

Figure 2. Transcriptomic map of DBA$^+$ pancreatic duct cells. A) UMAP depicts identity of clusters. B) The dot plot shows the top five significantly DEGs with the highest fold change for each cluster. C) Feature plots show expression of significantly DEGs for clusters 0, 1, 3, 4, and 5. Cluster 2 is characterized by lack of or low level expression of significantly DEGs found in other clusters. D) IPA results show the top 8 deregulated pathways when comparing a cluster to all other clusters. The ratio line indicates the fraction of molecules significantly altered out of all molecules that map to the canonical pathway from within the IPA database. A positive z-score represents upregulation, and a negative z-score indicates downregulation of a pathway in that cluster when compared to all other clusters. A gray bar depicts significant overrepresentation of a pathway, the direction of which cannot yet be determined.

Figure 3. Comparison of ductal clusters 0 vs 2, 4 vs 1, and 4 vs 4:*Dmbt1*$^+$*Ly6d*$^+$. A) The cluster dendrogram created using dims (used to define the cluster) shows the Euclidean relationships between clusters. The tree is calculated in the PCA space. The genes used to define the tree were set as the variable features of the object. B) Pearson's correlation calculated using average gene expression is depicted. C) Stacked violin plots show five DEGs sharing similar expression patterns in clusters 0 and 2. D) The dot plot shows all 9 DEGs found when comparing clusters 0 vs 2. E) The top 8 altered pathways from IPA comparing clusters 0 vs 2 are depicted. F) Stacked violin plots show five DEGs sharing similar expression patterns in clusters 4 and 1. G) The dot plot shows the top 20 DEGs ranked by fold change when comparing clusters 4 vs 1. H) The top 8 deregulated pathways from IPA comparing clusters 4 vs

1 are depicted. I) Stacked violin plots of five DEGs sharing similar expression patterns in clusters 4:*Dmbt1*$^+$*Ly6d*$^+$ and 4. J) The dot plot shows the top 20 DEGs ranked by fold change when comparing clusters 4:*Dmbt1*$^+$*Ly6d*$^+$ and 4. K) The top 8 changed pathways from IPA comparing clusters 4:*Dmbt1*$^+$*Ly6d*$^+$ and 4 are depicted.

Figure 4. RaceID3/StemID2 predict clusters 0 and 2 have the highest progenitor potential. A) The lineage tree inferred by StemID2 is shown in the RaceID3 clusters. Node color represents the level of transcriptome entropy, edge color describes level of significance, and edge width describes link score. B) Heat map depicts expression of top 5 DEGs in RaceID3 clusters with FDR < 0.01 and fc > 1.2. C) StemID2 scores for RaceID3 clusters are graphed. D) Monocle 3 clustering of murine DBA$^+$ duct clusters 0, 1, 2, and 4 are depicted. E) Each cell's relative pseudotime value is depicted that is a measurement of the distance between its position along the trajectory and the starting point (cluster 0).

Figure 5. Monocle 3 analysis reveals an epithelial mesenchymal axis in pancreatic duct cells. A) Expression changes of the modules generated by Monocle 3 analysis are shown for each cluster. B-D) Expression of modules 4, 14, and 34 along with select IPA results of the top 10 deregulated pathways are shown. Genes in parenthesis are altered in the pathway containing an asterisk in the bar. E) Stacked violin plots show expression of genes in the Regulation of the Epithelial Mesenchymal Transition By Growth Factors Pathway in DBA$^+$ duct clusters 0-5. F) IF depicts CK19$^+$ Vimentin$^+$ copositive pancreatic duct cells in mouse (yellow arrow) and human. The main pancreatic duct is shown for humans. Scale bars are 50uM.

Figure 6. *SPP1* loss promotes a progenitor-like state in human pancreatic duct cells. A-C) Western blot and quantification of western blot images shows expression of Annexin A3, Osteopontin, and Geminin in knockout HPDE6c7 lines and the control. D) Brightfield images show changes in cellular morphology of HPDE6c7 *SPP1* knockout lines. Yellow arrows point to filipodia. Scale bars are 100 µm. E) Cell counting demonstrates a significant increase in cell number at Day 6 in HPDE6c7 *SPP1* KO cells when compared to the HPDE6c7 scr gRNA control (p=0.0089 for HPDE6c7 *SPP1* gRNA1 and p=0.0042 for HPDE6c7 *SPP1* gRNA2). F) Significantly decreased carbonic anhydrase activity is observed in HPDE6c7 *SPP1* knockout lines when compared to the control. G) Relative fold changes calculated using RPM values of mesenchymal and duct markers are shown. Average RPM values for *SPP1* are 79.8 ± 43.2 (scr) and 2.8 ± 0.4 (KO), *HNF1B* are 418.5 ± 33.4 (scr) and 1 ± 0.3 (KO), *SOX9* are 1,555.2 ± 124.8 (scr) and 316 ± 38.7 (KO), *KRT19* are 16,789.5 ± 2,431.2 (scr) and 61 ± 60.5 (KO), *CLDN7* are 8,651.8 ± 923.2 (scr) and 112.8 ± 27.7 (KO), *CDH1* are 8,651.8 ± 923.2 (scr) and 112.8 ± 27.8 (KO), and *VIM* are 80.2 ± 29.1 (scr) and 6,879.8 ± 652.6 (KO). H) The top 14 deregulated pathways from IPA are shown comparing HPDE6c7 *SPP1* KO vs HPDE6c7 scr gRNA control. I) Immunocytochemistry (ICC) demonstrated reduced Osteopontin expression in HPDE6c7 gRNA2 and HPDE6c7 *SPP1* gRNA4 when compared to HPDE6c7 scr gRNA. Vimentin ICC depicts organized intermediate filaments in HPDE6c7 *SPP1* gRNA2 and HPDE6c7 *SPP1* gRNA4 while HPDE6c7 scr gRNA cells show diffuse, light labeling. Scale bar is 50 µm. J) CK19 ICC shows organized intermediate filaments in HPDE6c7 scr gRNA cells while HPDE6c7 *SPP1* gRNA2 and HPDE6c7 *SPP1* gRNA4 cells display punctate CK19 labeling, where present. Scale bar denotes 50 µm. K) qPCR results of pancreatic progenitor markers are shown for HPDE6c7 scr gRNA control and HPDE6c7 *SPP1* knockout lines.

Figure 7. HPDE *SPP1* knockout cells are capable of differentiating into cells with endocrine appearance, including cells exhibiting α- and β-like appearance, but not duct-like or acinar-like cells *in vivo*. A) Schematic of *in vivo* experiment. B) IF shows CK19$^+$ α-amylase$^+$ double positive HPDE6c7 scr gRNA cells. C) IF depicts NGN3$^+$ SOX9$^+$ double positive

HPDE6c7 *SPP1* knockout cells (yellow arrows). D) Synaptophysin$^+$ Glucagon$^+$ double positive cells (yellow arrows) are detected in HPDE6c7 *SPP1* knockout cells. E) C-peptide, synaptophysin, NKX6.1, and PDX1 expression are evident in HPDE6c7 *SPP1* knockout cells. C-peptide, synaptophysin, and PDX1 triple positive cells are highlighted with yellow arrows. F) The percentage of C-peptide$^+$, NKX6.1$^+$, and PDX1$^+$ triple positive cells for HPDE6c7 scr gRNA cells is 0, HPDE6c7 *SPP1* gRNA2 is 1.343 ± 0.2664, and HPDE6c7 *SPP1* gRNA 4 cells is 1.642 ± 0.1533. *All* scale bars in this figure are 50 µm.

Figure 8. Geminin is a regulator of genomic stability in mouse pancreatic duct cells during chronic pancreatitis. A-B) High magnification IF images and quantification show a significant increase in proliferation in pancreatic duct cells in CP patients when compared to normal human pancreatic duct cells. Yellow arrows point to Geminin$^+$ Ki67$^+$ copositive cells. C) A schematic of tamoxifen and BrdU administration is shown. The UMAP depicts the pancreatic cells (clusters 0-2) that were analyzed in this experiment. The Venn diagram shows the number of cells in clusters 0-2 that are SOX9$^+$, HNF1B$^+$, and SOX9$^+$HNF1B$^+$ copositive. D) Quantification of Geminin positive ductal cells at Day 7 in *Sox9-CreERT2*$^{Tg/wt}$; *Geminin*$^{f/f}$, *Sox9-CreERT2*$^{Tg/wt}$; *Geminin*$^{f/wt}$, *Hnf1b*$^{CreERT2\ Tg/wt}$; *Geminin*$^{f/f}$ and control mice is depicted. E-F) Representative IF images and quantification of γ-H2AX positive foci is shown at Day 7 in the PDL transgenic models. *G-H)* Representative IF images and quantification of γ-H2AX positive foci are shown at Day 30 in the *Sox9-CreERT2*$^{Tg/wt}$; *Geminin*$^{f/f}$ and control PDL models. All scale bars in this figure are 50 µm.

Schematic 1. A word cloud depicts the top DEGs in each duct cluster. The size of each gene is relative to the adjusted p value.

Supplemental figure captions

Figure 1-figure supplement 1. Features of DBA$^+$ (clusters 0-15) and ductal (clusters 0-5) cells. A) The sorting strategy for live, DBA$^+$ pancreatic cells is displayed. The graph on the left shows the cells plotted by forward and side scatter. In this graph, red cells are PI$^+$, purple cells are DBA$^+$PI$^-$, and blue and green cells are doublets. The graph on the right shows the gating strategy used to sort DBA$^+$PI$^-$ pancreatic cells. B) The number of genes and transcripts for each cell in clusters 0-15 are shown. C) Co-immunofluorescence labeling identifies CK19$^+$DBA$^+$ and Collagen I$^+$DBA$^+$ copositive pancreatic cells and CD45$^+$DBA$^+$ copositive pancreatic lymph node cells. Pancreatic CD31$^+$ cells are not DBA$^+$. A yellow arrow points to a copositive cell. Scale bars are 50 µm. D) The plot shows whether a ductal cell is in cluster 0 or 8 (from the 0-15 cluster dataset) in the 0-5 cluster UMAP. E) This plot depicts the location of the ductal clusters 0-5 in the 0-15 cluster UMAP. F) Violin plots show the number of genes and transcripts in each cell for ductal clusters 0-5. G) Feature plots depict expression of genes normally enriched in pancreatic duct cells.

Figure 2-figure supplement 1. IHC illustrates expression of markers in clusters 0 and 2 in the mouse and human ductal tree. A) Geminin is expressed in rare ductal cells and acinar cells in the mouse and human pancreas. Yellow arrows point to geminin positive cells. B) Osteopontin expression is observed in all duct cell types throughout the mouse and human ductal tree as well as in acinar cells. C) WFDC3 is expressed in all duct cell types in the mouse and human pancreas and in acinar cells. Red arrows point to the indicated duct type. Scale bars are 40 µm.

Figure 2-figure supplement 2. IHC and IF depict expression of markers in clusters 1, 3, 4, and 5 in mouse and human pancreas duct cells. A) Annexin A3 expression is observed in all

ductal cell types in the mouse and human pancreas. The left mouse IHC image under Large duct – Interlobular/Main duct shows Annexin A3 cytoplasmic expression in pancreatobiliary cells and cells within the Ampulla of Vater. Scale bars are 40 µm. B) PAH is expressed in all duct types throughout the mouse and human ductal tree as well as in acinar cells. Scale bars are 40 µm. C) Yellow arrows point to heterogeneous expression of ductal markers CFTR, Annexin A3, and CK19 in human pancreatic duct cells. Scale bars are 20 µm. D) Proliferating and acetylated alpha tubulin positive duct cells are observed in the intrapancreatic bile duct, peribiliary glands, and pancreatobiliary cells in mouse and human. Scale bars are 50 µm.

Figure 2-figure supplement 3. Characteristics of intrapancreatic bile duct and pancreatobiliary cells. A) CXCL5, another marker of the *Dmbt1*$^{+}$*Ly6d*$^{+}$ subpopulation, positive cells are located in murine and human intrapancreatic bile duct cells, peribiliary glands, and pancreatobiliary cells. Yellow arrows point to ductal cells displaying upregulated CXCL5. Scale bars are 50 µm. B) Stacked violin plots show expression of 5 genes which are upregulated in clusters 3 and 4. C) The dot plot shows the top 20 DEGs ranked by fold change when comparing clusters 3 vs 4.

Figure 3-figure supplement 1. Alignment to an adult murine hepatic biliary epithelial cell dataset. A) UMAP showing alignment of adult murine hepatic BECs (blue) to our murine intrapancreatic bile duct cells (red) and pancreatobiliary cells (green). B) Clustering of merged datasets defines 5 clusters. C) Intrapancreatic bile duct cells in DBA^{+} duct cluster 3 are primarily located within the merged clusters 0 and 1, and pancreatobiliary cells in DBA^{+} duct cluster 4 are primarily located within the merged clusters 1 and 2. The heatmap shows the percent of cells from our clusters 3 and 4 within each of the merged clusters 0-4. D) Feature plots depict the 75th percentile and higher of cells expressing the published gene signatures of hepatic BEC subpopulation A and B respectively. E) Cells in clusters 1, 2, and 4 have the strongest enrichment for subpopulation A genes, while cells in clusters 0, 1, and 3 have the strongest enrichment for subpopulation B genes in the merged dataset. F) Dual violin plots show expression of the ductal marker *Sox9* and the YAP1 targets *Cyr61*, *Ankrd1*, and *Gadd45b* in the merged clusters. G) Dot plot shows expression of hepatic BEC subpopulation A and B genes, analyzed by t-SNE, in Figure 1D of Pepe-Mooney *et al.* (2019) in our murine pancreas DBA^{+} duct clusters 0-5.

Figure 3-figure supplement 2. Analysis of pancreas duct cells during development. A, E, I) UMAPs depict ductal clusters evident at E12.5, E14.5, and E17.5, respectively. B, F, J) Cluster dendrograms created using dims (used to define the cluster) shows the Euclidean relationships between clusters at E12.5, E14.5, and E17.5, respectively. C, G, K) Dot plots show expression of the top 5 genes defining adult C57BL/6J duct subpopulations in developmental biology samples at E12.5, E14.5, E17.5, respectively. M-N) Feature plots of genes that characterize subpopulations of mouse and human duct cells in Baron *et al.* (2016). O) Feature plot showing *Fth1* expression, which typifies a human pancreas duct subpopulation in Grün *et al*. (2016).

Figure 3-figure supplement 3. Comparison of DBA^{+} lectin sorted mouse pancreas duct subpopulations to ALK3^{+} human pancreas duct subpopulations. A-E) Aggregated expression of control feature sets shown in panel A were subtracted from the average expression levels of DEGs for each cluster 0-5 on a single cell level to determine the AddModuleScore comparing each DBA^{+} pancreas ductal cluster to ALK3^{+} human pancreas clusters. Panel D shows the number of DEGs in murine DBA^{+} pancreas duct clusters 0-5 that have a human homolog and could be used in this comparison. F) *Bmpr1a* is expressed in a subset of murine DBA^{+} pancreas duct cells. G) Stacked violin plots depict expression of

centroacinar/terminal ductal cell markers *Hes1*, *Aldh1b1*, *Aldh1a1*, and *Aldh1a7* in DBA$^+$ pancreas duct clusters 0-5. H) Immunostaining identifies ducto-acinar cells in murine pancreas. Yellow arrows point to CPA1 or α-amylase positive murine ductal cells. Similar to other murine ductal cell markers (Figure S1G), DBA lectin also shows heterogenous expression in murine pancreatic duct cells. Blue arrows point to a DBA lectin negative duct cell. The scale bar is 50um.

Figure 5-figure supplement 1. RaceID3 clusters and Monocle 3 analysis. A) The location of DBA$^+$ duct cluster 0, 1, 2, and 4 cells in RaceID3 clusters are depicted in Seurat space. Clusters 3 and 5 were not included in this analysis and all clusters with <10 cells were also removed. All removed cells are depicted in the N/A square. B) RaceID3 cluster 10 cells, which have the highest StemID2 score, are depicted in the Monocle 3 UMAP. C) The Monocle 3 UMAP and trajectory are shown when DBA$^+$ duct clusters 0-5 are all included in the analysis. D) A violin plot showing the distribution of cells along the relative pseudotime axis, split by DBA$^+$ duct clusters, is shown. Cluster 4 appears to be the most differentiated from the inferred root, cluster 0. E) Co-IF depicts fibronectin, e-cadherin, and SNAI1 expression in human pancreatic ducts. A blue arrow points to a SNAI1 negative cell, and a yellow arrow points to 2 fibronectin positive cells. Scale bar is 50 µm.

Figure 6-figure supplement 1. Characterization of DBA$^+$ murine ductal markers. qPCR analysis of epithelial (A) and mesenchymal (B) markers in HPDE6c7 *SPP1* KO lines and the HPDE6c7 scr gRNA control are shown. C) Flow cytometry profiles of select markers in HPDE6c7 *SPP1* KO lines and the HPDE6c7 scr gRNA control are depicted. D-F) GSEA enrichments plots for select pathways are depicted.

Figure 8-figure supplement 1. Generation of 2loxP and 1loxP heterozygous ES cells for mouse *Geminin.* A) The targeting strategy to generate a conditional KO allele for *Geminin* in ES cells is shown. The exact distance between individual exons and their relative sizes is not shown. ES cells heterozygous for the 3loxP allele were obtained through homologous recombination. A Cre recombinase was used to generate ES cells harboring either the 2loxP allele or the 1loxP allele *in vitro.* A SphI restriction site was introduced with the leftmost loxP site to allow screening for its presence by Southern blot analysis. B) The 5' probe was used in conjunction with a SphI digest. Besides the wild-type allele, a fragment of about 2.5kb is expected for the 3loxP allele as indicated. Clone 17 and clone H18 tested positive. C) Clone H18 was chosen for Cre treatment *in vitro*. Using a Bsu36I digest and the 5' probe, the 2loxP allele displays a single fragment of the same size as the wild-type allele whereas the 1loxP allele produces a smaller fragment. Bsu36I restriction sites are omitted in the schematic shown in A) for clarity. The 3' probe clearly distinguishes between wild-type, 1loxP, 2loxP, and 3loxP alleles with a SphI digest. Clone 53 was identified as an ES cell clone heterozygous for the conditional 2loxP allele and used for blastocyst injection.

Figure 8-figure supplement 2. Histology of transgenic PDL models. A-B) Representative H&E images of PDL transgenic models at Day 7 and Day 30 are shown. Scale bars are 40 µm.

Figure 8-figure supplement 3. Geminin loss causes a transient proliferation response in *Sox9-CreERT2$^{Tg/wt}$*; *Geminin$^{f/f}$* mice. A) Representative IF images of the quantification shown in Figure 7D are depicted. Scale bar is 20 µm. B-C) IF images and quantification of BrdU positive pancreatic ductal cells is shown at Day 7 in the PDL transgenic models. Scale bars are 50 µm. D-E) IF images and quantification of BrdU positive pancreatic ductal cells is shown at Day 30 in the *Sox9-CreERT2$^{Tg/wt}$*; *Geminin$^{f/f}$* and control PDL models. Scale bars are 50 µm.

Supplementary file 1. Expression scoring of markers of subpopulations of pancreatic duct cells. This table depicts a summary of expression scoring of selected markers for subpopulations of pancreatic duct cells in mouse and human tissue. Homogeneous refers to an observed uniform expression level and pattern within a particular ductal cell type. Heterogeneous means that either the observed expression level or pattern varies among cells within a particular ductal cell type.

Supplementary file 2. The list of antibodies used in this study.

Supplementary file 3. The list of gRNA sequences used in this study.

Supplementary file 4. The list of qPCR primer sequences used in this study.

Figure 1-source data 1. Number of cells and average number of genes and transcripts in all DBA$^+$ clusters.

Figure 1-source data 2. DEGs in all DBA$^+$ clusters.

Figure 2-source data 1. DEGs in all DBA$^+$ duct clusters.

Figure 2-source data 2. IPA results for all DBA$^+$ duct clusters.

Figure 2-source data 3. IPA Upstream Regulator analysis results for all DBA$^+$ duct clusters.

Figure 2-source data 4. Number of cells and average number of genes and transcripts in all DBA$^+$ duct clusters.

Figure 3-source data 1. DEGs comparing duct cluster 0 vs. 2, duct cluster 4 vs. 1, duct cluster 4-Dbmt1$^+$Ly6d$^+$ vs 4.

Figure 3-source data 2. IPA results comparing duct cluster 0 vs. 2, duct cluster 4 vs. 1, duct cluster 4-Dbmt1$^+$Ly6d$^+$ vs 4.

Figure 3-source data 3. IPA Upstream Regulator analysis comparing duct cluster 0 vs. 2, duct cluster 4 vs. 1, duct cluster 4-Dbmt1$^+$Ly6d$^+$ vs 4.

Figure 2-figure supplement 3-source data 1. DEGs comparing duct cluster 3 vs. 4.

Figure 2-figure supplement 3-source data 2. IPA results comparing duct cluster 3 vs. 4.

Figure 2-figure supplement 3-source data 3. IPA Upstream Regulator analysis comparing duct cluster 3 vs. 4.

Figure 3-figure supplement 1-source data 1. Number of cells and average number of genes and transcripts for merged BEC – DBA$^+$ duct clusters 3 and 4 dataset.

Figure 3-figure supplement 1-source data 2. DEGs in merged BEC – DBA$^+$ duct clusters 3 and 4 dataset.

Figure 5-source data 1. IPA results comparing select modules in Monocle 3 analysis.

Figure 5-source data 2. IPA Upstream Regulator analysis comparing comparing select modules in Monocle 3 analysis.

Figure 5-source data 3. Log fold change analysis comparing select modules in Monocle 3 analysis.

Figure 6-source data 1. DEGs comparing HPDE6c7 *SPP1* KO vs. HPDE6c7 scr gRNA.

Figure 6-source data 2. IPA results comparing HPDE6c7 *SPP1* KO vs. HPDE6c7 scr gRNA.

Figure 6-source data 3. IPA Upstream Regulator analysis comparing HPDE6c7 *SPP1* KO vs. HPDE6c7 scr gRNA

A

1) Dissect pancreas from 4 C57BL/6J mice

2) Isolate live DBA+ pancreatic cells from pooled pancreata

3) Single cell GEM formation

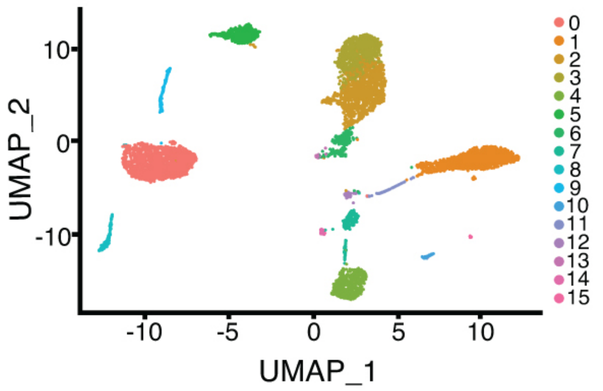
Gel beads

Single DBA+ cells

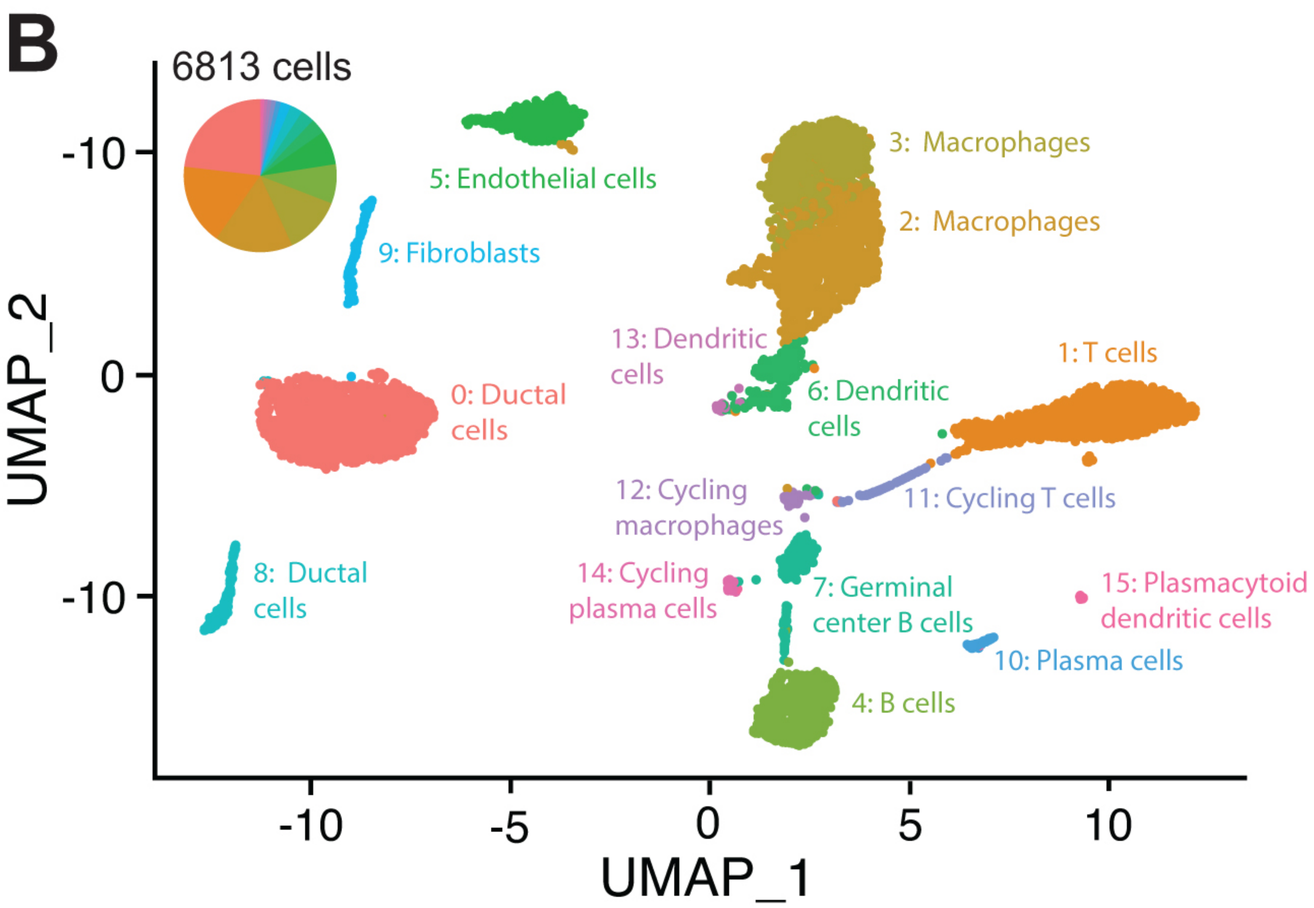
Oil

4) cDNA library preparation and scRNA-seq (10X Genomics)

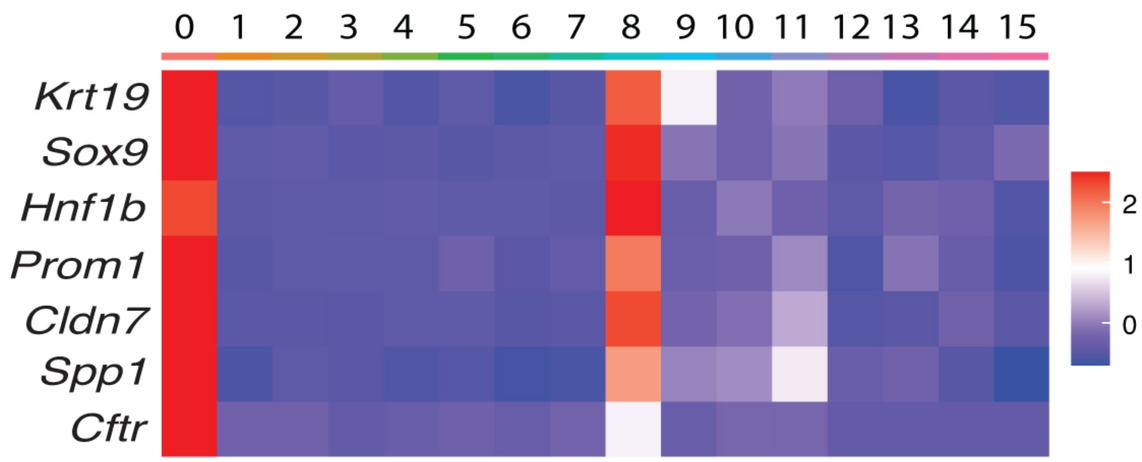
5) Analyze subpopulations of pancreatic and bile duct cells



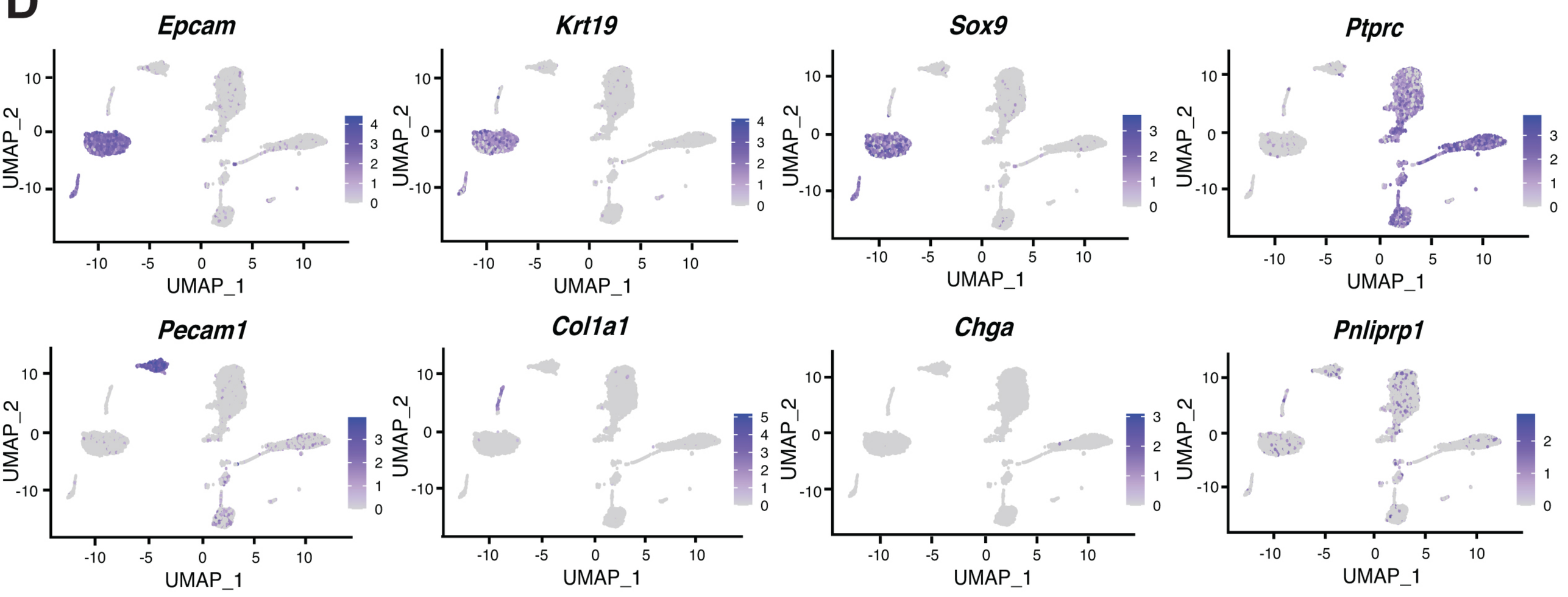
B

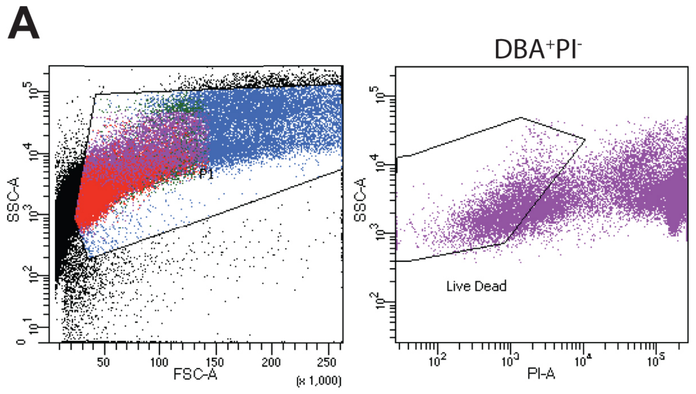


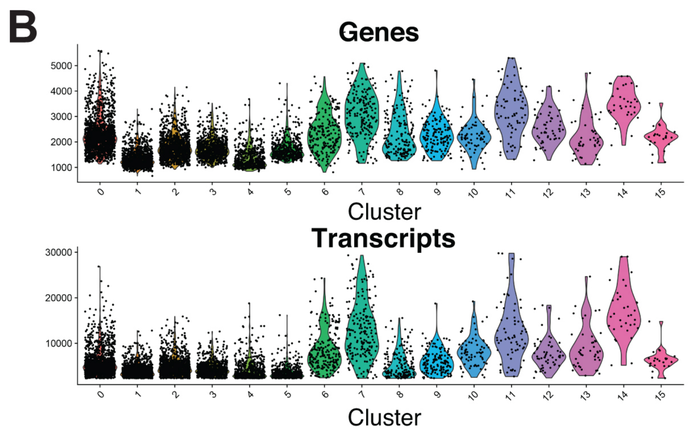
C

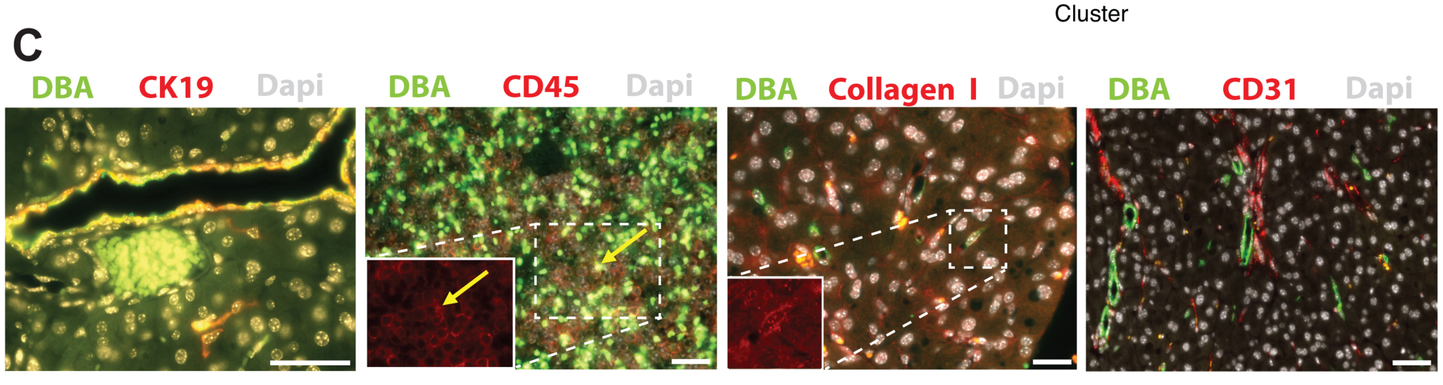


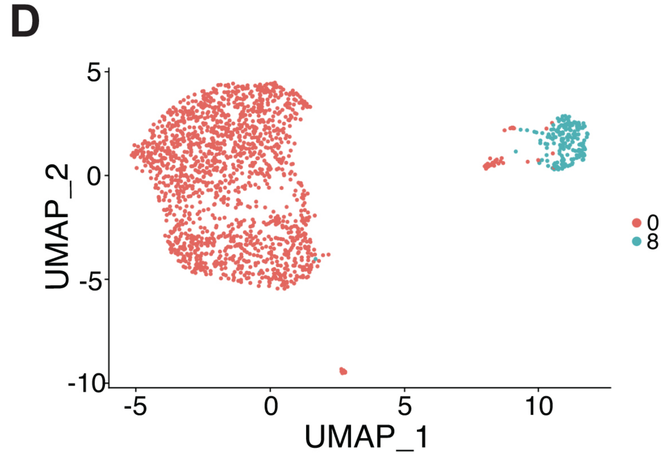
D

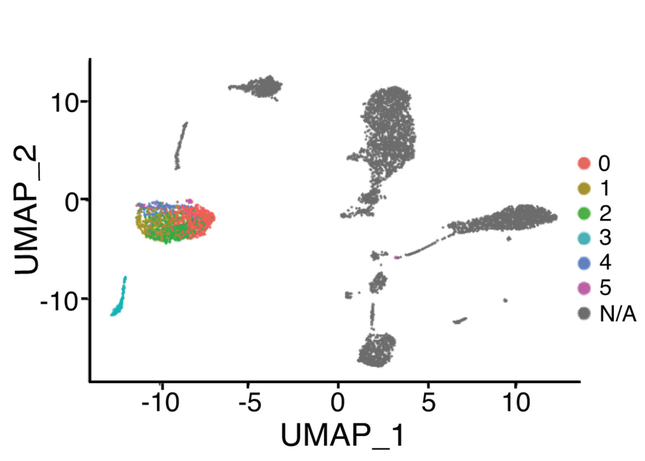


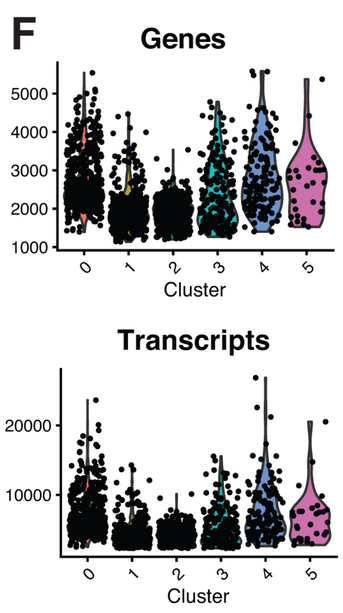

A
DBA+PI-
SSC-A
FSC-A
(x 1,000)
PI-A
Live Dead

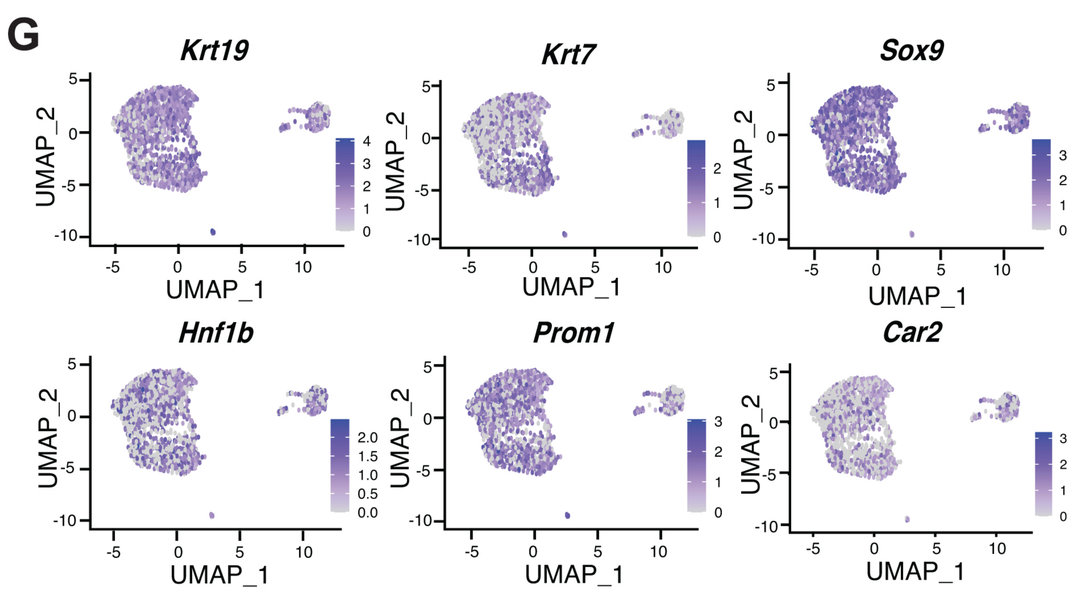

B
Genes
Cluster
Transcripts
Cluster


C
DBA CK19 Dapi
DBA CD45 Dapi
DBA Collagen I Dapi
DBA CD31 Dapi

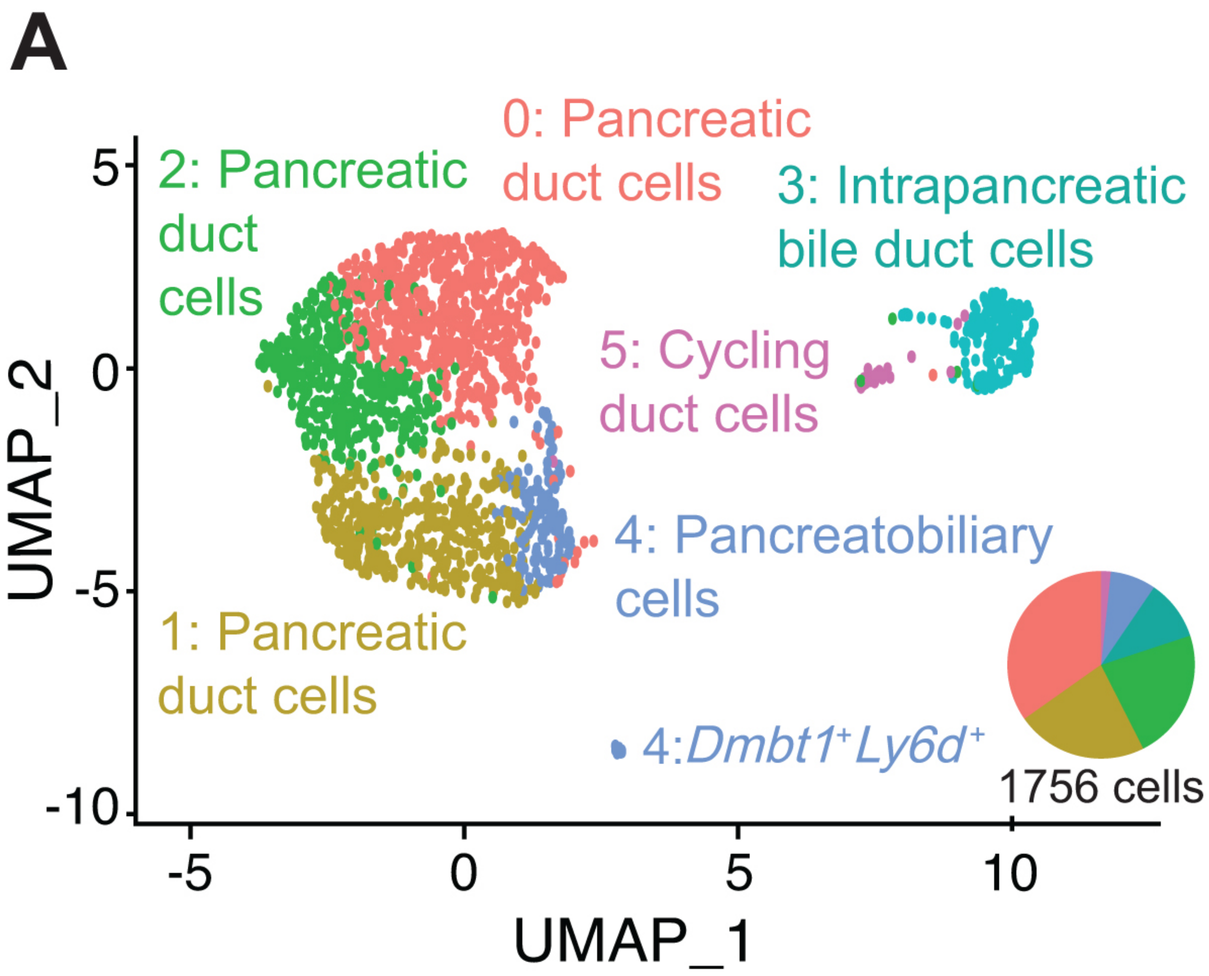

D
UMAP_2
UMAP_1
0
8


E
UMAP_2
UMAP_1
0
1
2
3
4
5
N/A

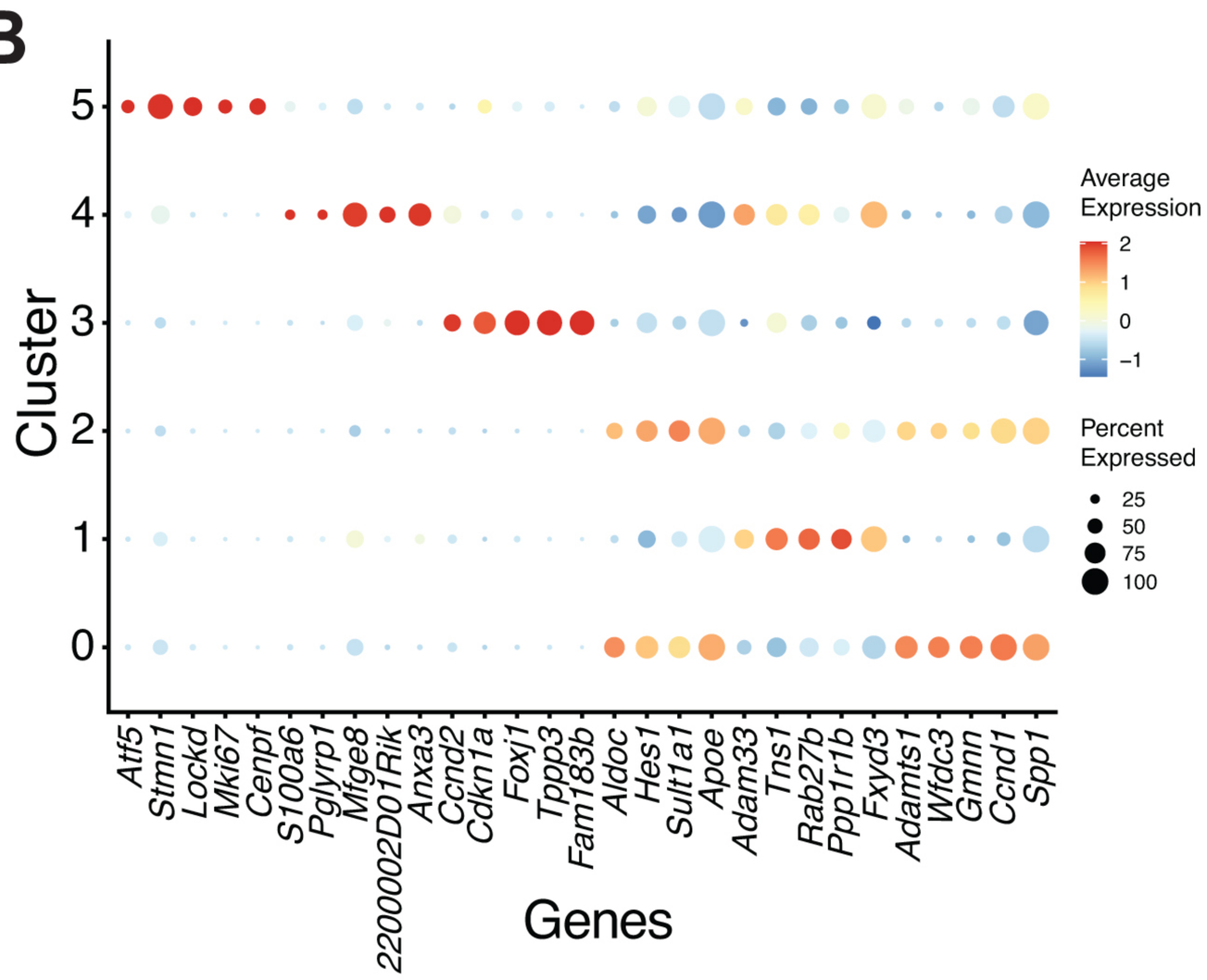

F
Genes
Cluster
Transcripts
Cluster


G
Krt19
Krt7
Sox9
Hnf1b
Prom1
Car2
UMAP_2
UMAP_1

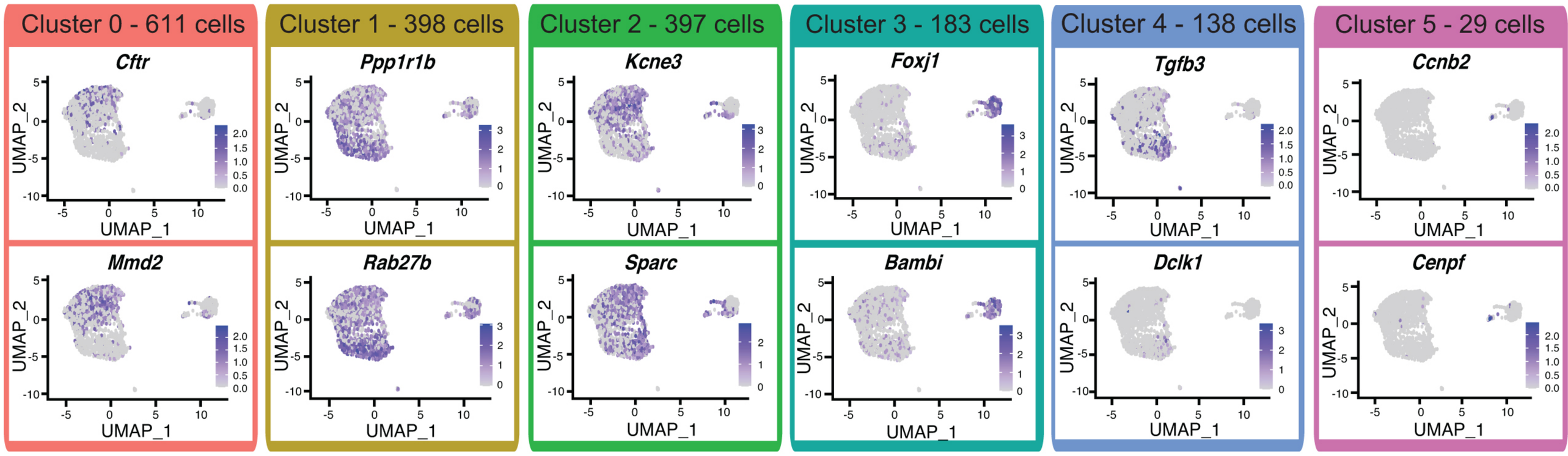
A


0: Pancreatic duct cells
2: Pancreatic duct cells
3: Intrapancreatic bile duct cells
5: Cycling duct cells
4: Pancreatobiliary cells
1: Pancreatic duct cells
4:Dmbt1+Ly6d+
1756 cells
UMAP_2
UMAP_1

B


Cluster
Genes
Average Expression
Percent Expressed
Atf5
Stmn1
Lockd
Mki67
Cenpf
S100a6
Pglyrp1
Mfge8
2200002D01Rik
Anxa3
Ccnd2
Cdkn1a
Foxj1
Tppp3
Fam183b
Aldoc
Hes1
Sult1a1
Apoe
Adam33
Tns1
Rab27b
Ppp1r1b
Fxyd3
Adamts1
Wfdc3
Gmnn
Ccnd1
Spp1

C


Cluster 0 - 611 cells
Cftr
Mmd2
Cluster 1 - 398 cells
Ppp1r1b
Rab27b
Cluster 2 - 397 cells
Kcne3
Sparc
Cluster 3 - 183 cells
Foxj1
Bambi
Cluster 4 - 138 cells
Tgfb3
Dclk1
Cluster 5 - 29 cells
Ccnb2
Cenpf
UMAP_1
UMAP_2

D


positive Z-score
Z-score = 0
negative Z-score
no activity pattern available
Ratio
Cluster 0
VDR/RXR Activation
Hepatic Fibrosis Signaling Pathway
Glycolysis I
Gluconeogenesis I
Colorectal Cancer Metastasis Signaling
Cysteine Biosynthesis/Homocysteine Degradation
HMGB1 Signaling
Glioblastoma Multiforme Signaling
Cluster 1
LPS/IL-1 Mediated Inhibition of RXR Function
Tyrosine Biosynthesis IV
Phenylalanine Degradation I (Aerobic)
IGF-1 Signaling
Calcium Transport I
Histamine Degradation
Oxidative Ethanol Degradation III
Hepatic Fibrosis / Hepatic Stellate Cell Activation
Cluster 2
VDR/RXR Activation
BEX2 Signaling Pathway
Cysteine Biosynthesis / Homocysteine Degradation
Arginine Degradation I (Arginase Pathway)
Arginine Biosynthesis IV
Proline Biosynthesis II (from Arginine)
Arginine Degradation VI (Arginase 2 Pathway)
Sucrose Degradation V (Mammalian)
Cluster 3
Aryl Hydrocarbon Receptor Signaling
Xenobiotic Metabolism AHR Signaling Pathway
p53 Signaling
GADD45 Signaling
Pyridoxal 5'-phosphate Salvage Pathway
BAG2 Signaling Pathway
NRF2-mediated Oxidative Stress Response
FXR/RXR Activation
Cluster 4
Hepatic Fibrosis / Hepatic Stellate Cell Activation
Leukocyte Extravasation Signaling
Death Receptor Signaling
Tight Junction Signaling
RhoGDI Signaling
Lymphotoxin β Receptor Signaling
Agranulocyte Adhesion and Diapedesis
TNFR2 Signaling
Cluster 5
EIF2 Signaling
Cell Cycle: G2/M DNA Damage Checkpoint Regulation
p53 Signaling
Regulation of eIF4 and p70S6K Signaling
Thioredoxin Pathway
Antigen Presentation Pathway
Aryl Hydrocarbon Receptor Signaling
Endocannabinoid Cancer Inhibition Pathway
-log(p-value)
Threshold

A Cluster 0

Small duct

Large duct - Interlobular/Main duct

Mouse

Intercalated duct

Centroacinar/ terminal ductal cell

Intralobular duct

Gmnn

UMAP_2

UMAP_1

Human

Centroacinar/ terminal ductal cell

Intercalated duct

Intralobular duct

Main pancreatic duct

Interlobular duct

Geminin

B Clusters 0+2

Small duct

Large duct - Interlobular/Main duct

Mouse

Centroacinar/terminal ductal cell

Intercalated duct

Intralobular duct

Spp1

UMAP_2

UMAP_1

Human

Intercalated duct

Centroacinar/ terminal ductal cell

Intralobular duct

Intercalated duct

Interlobular duct

Main pancreatic duct

Osteopontin

C Clusters 0+2

Small duct

Large duct - Interlobular/Main duct

Mouse

Centroacinar/terminal ductal cell

Intralobular duct

Intercalated duct

Centroacinar/terminal ductal cells

Intercalated duct

Wfdc3

UMAP_2

UMAP_1

Human

Intralobular duct

Intercalated duct

Centroacinar/terminal ductal cell

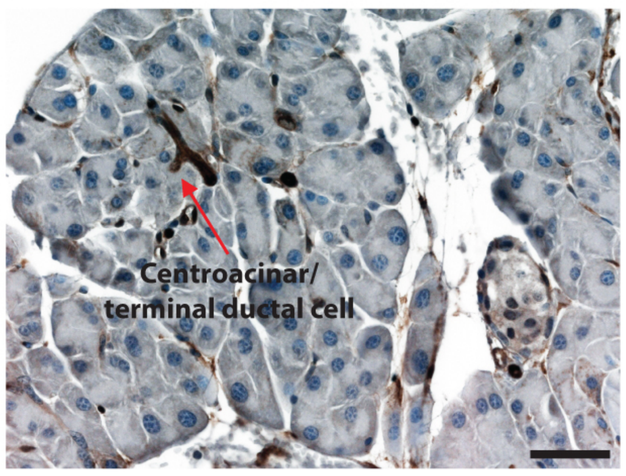
Interlobular duct

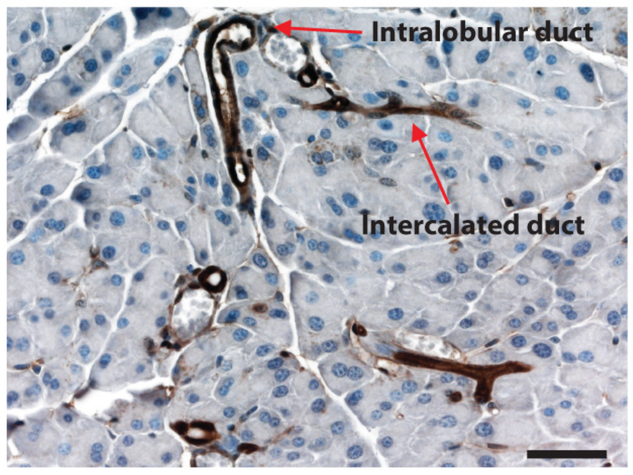
Main pancreatic duct

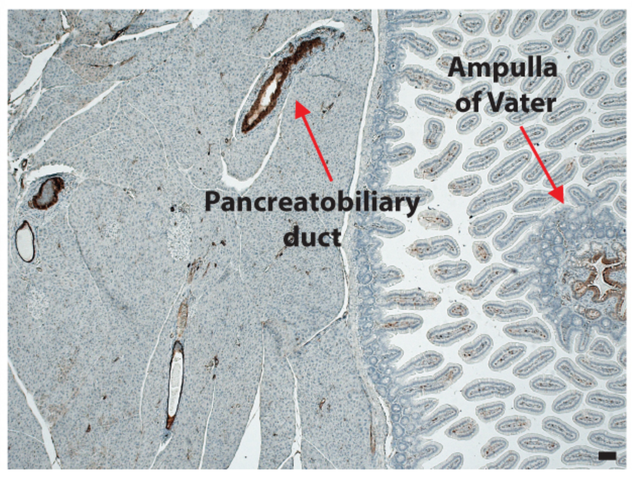
WFDC3

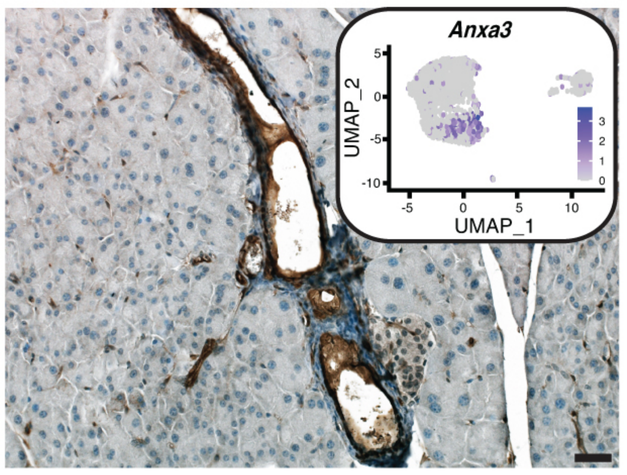
A Clusters 1+4 — Small duct | Large duct - Interlobular/Main duct

Mouse

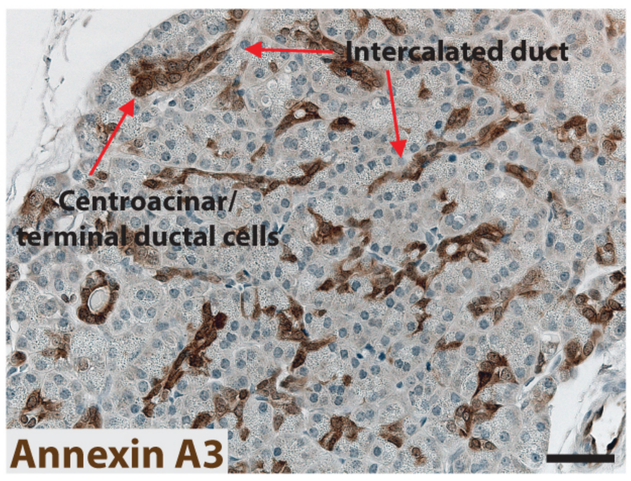


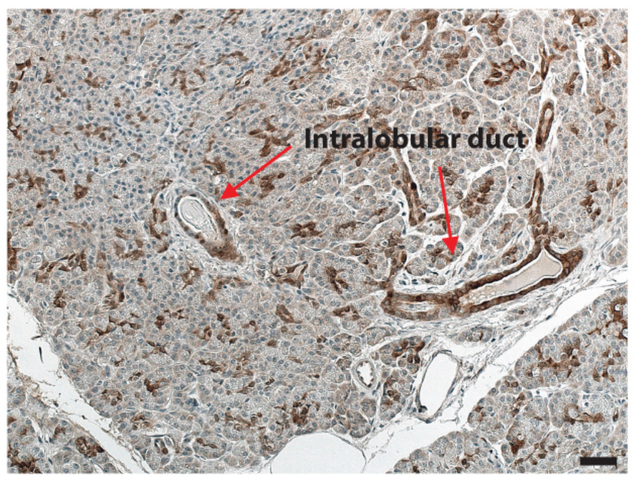


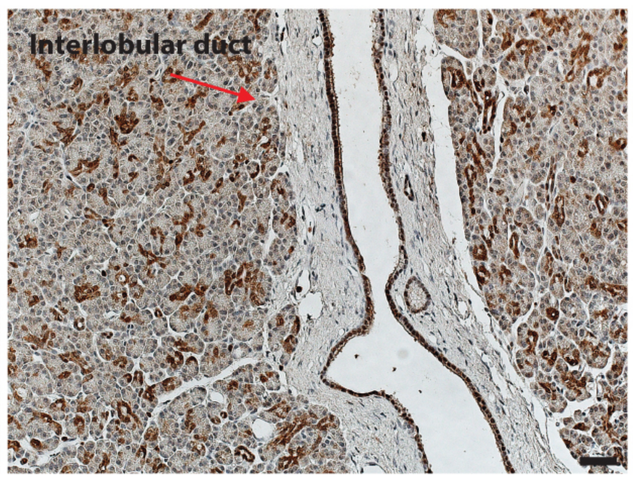


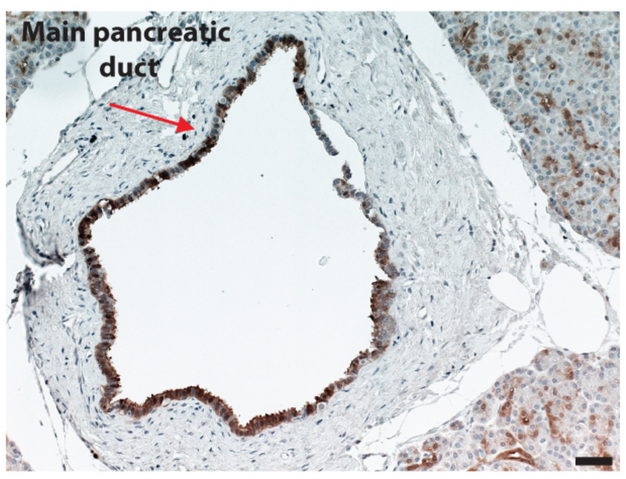


Human



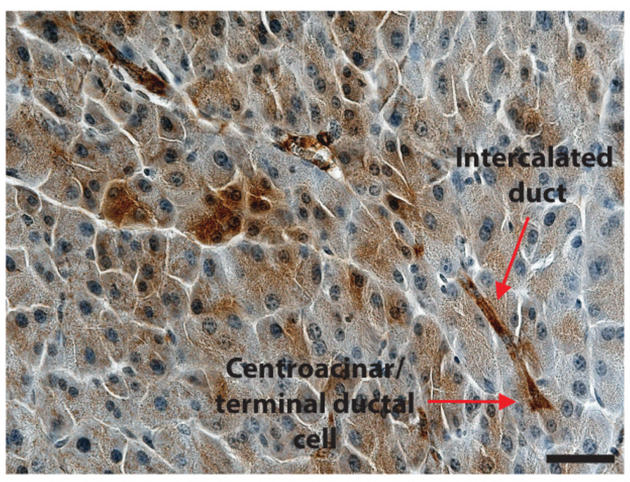


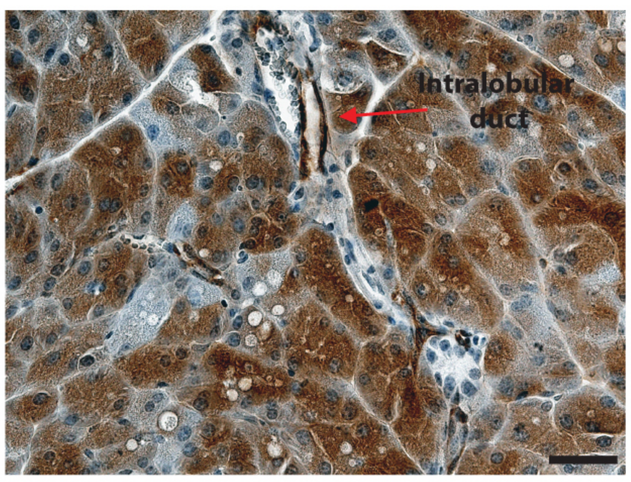


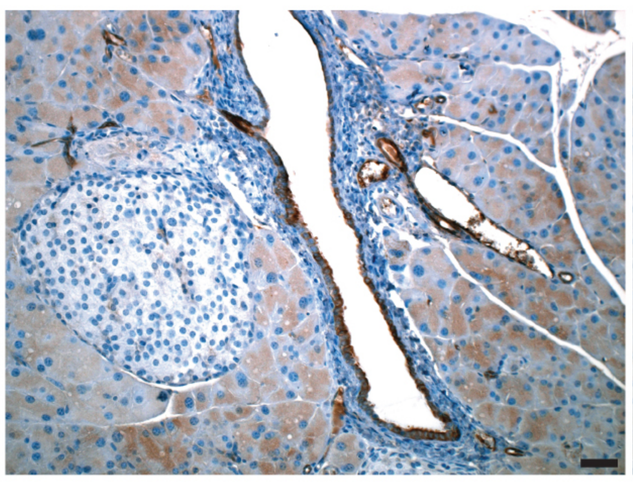


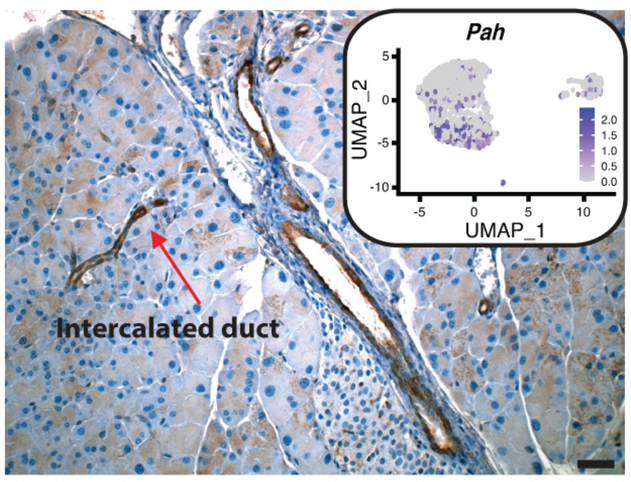
B Clusters 1+4 — Small duct | Large duct - Interlobular/Main duct

Mouse

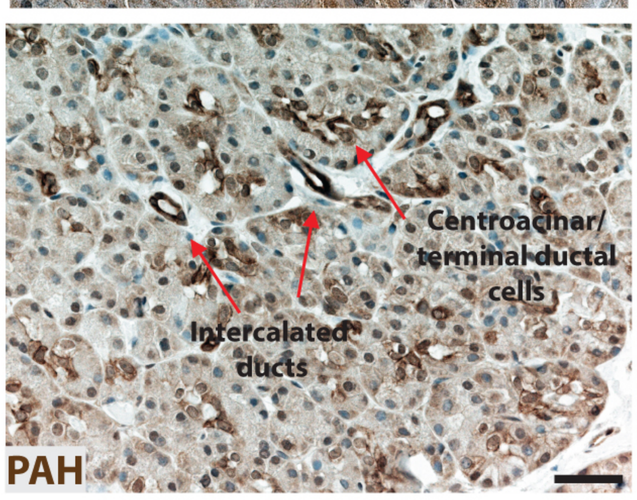


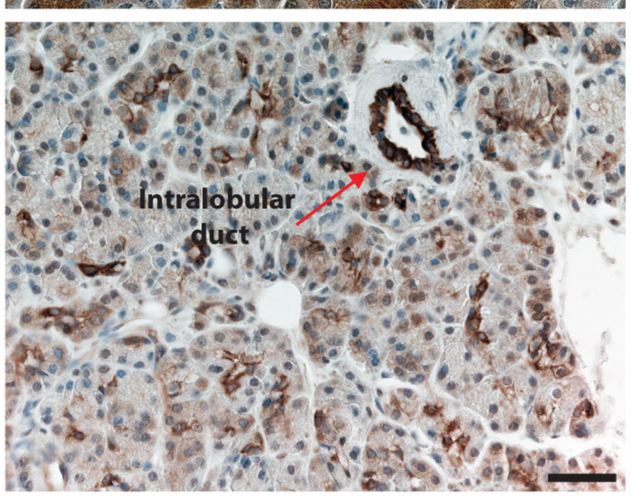


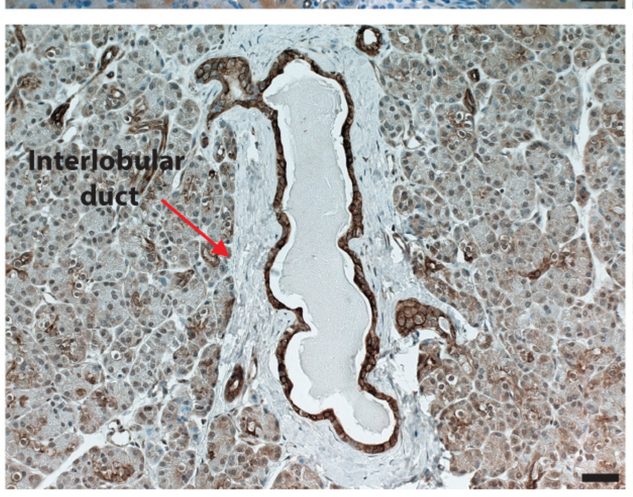


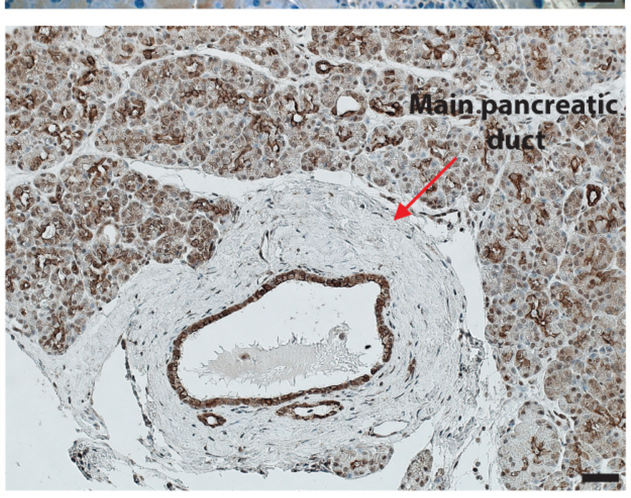


Human

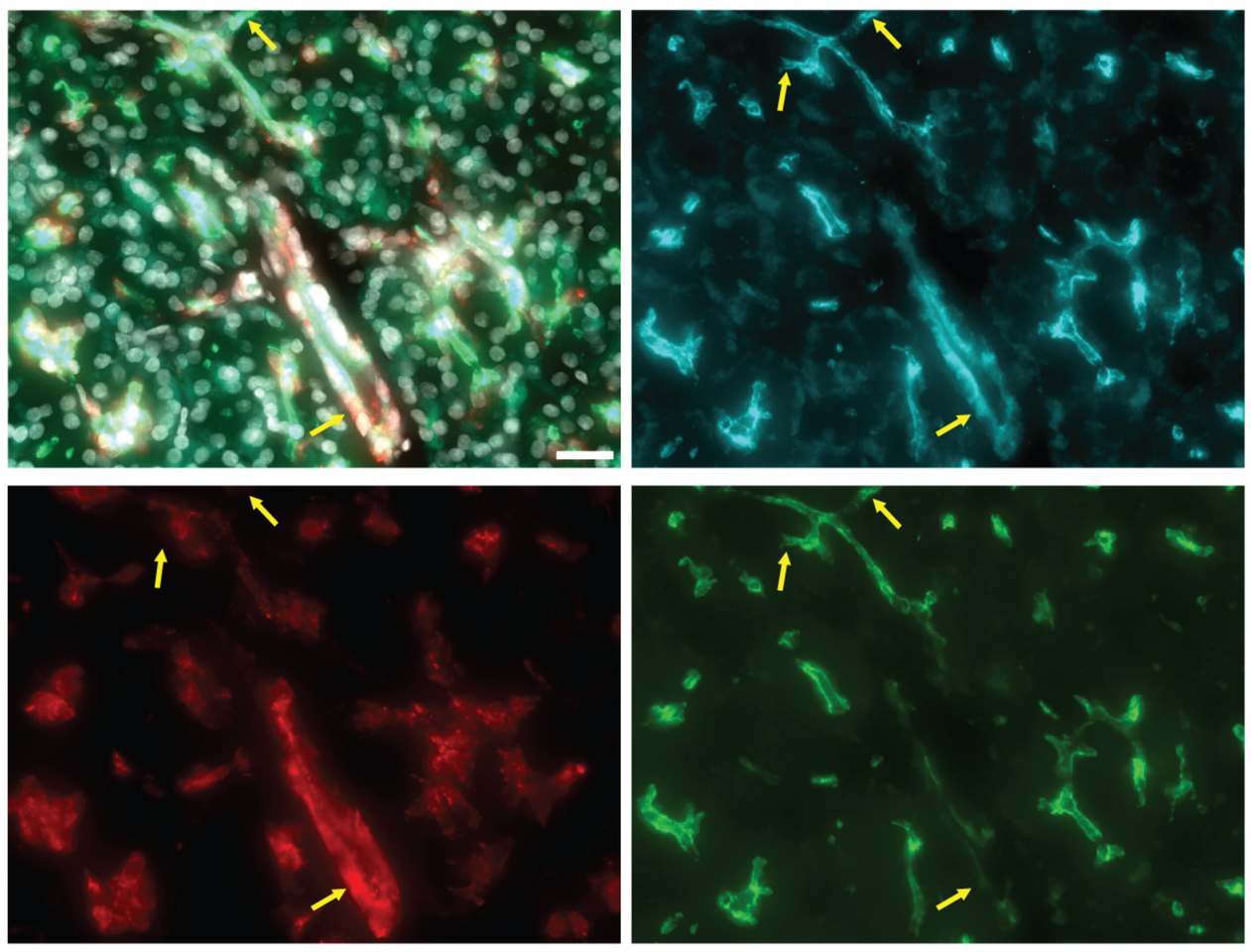






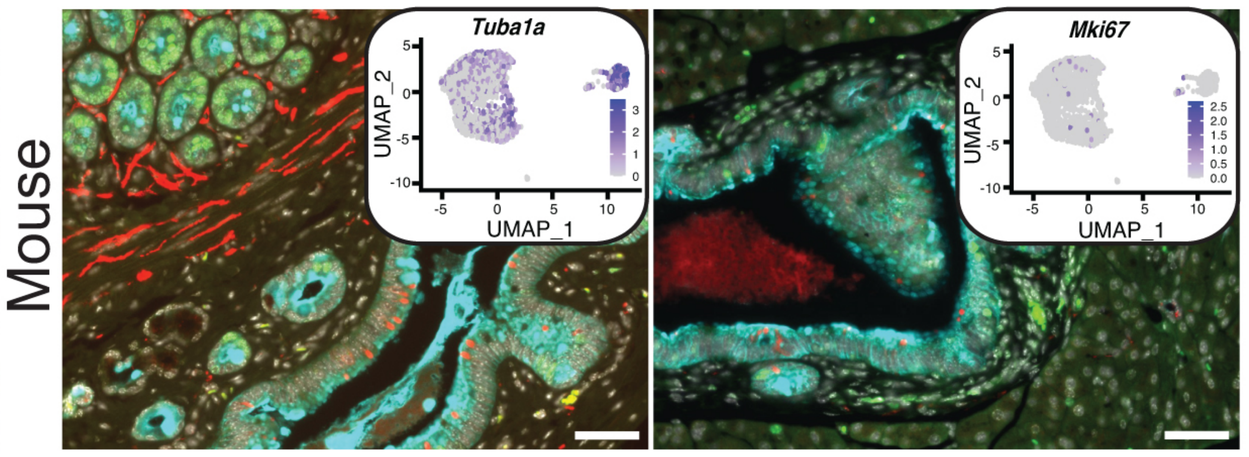


C

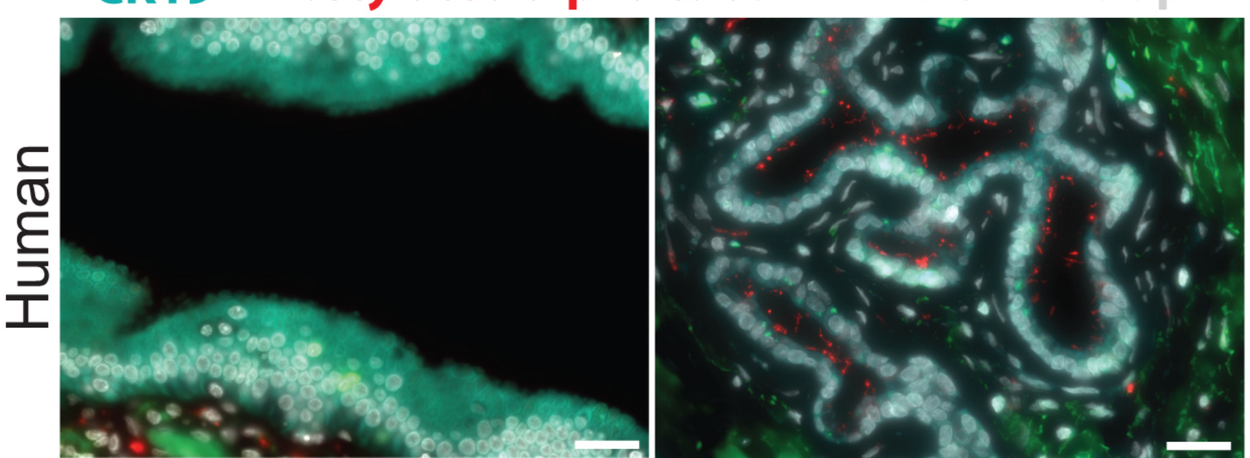


D Clusters 3+5

Pancreatobiliary duct | Intrapancreatic bile duct



Intrapancreatic bile duct | Peribiliary glands



A

Pancreatobiliary cells

Intrapancreatic bile duct cells

Mouse

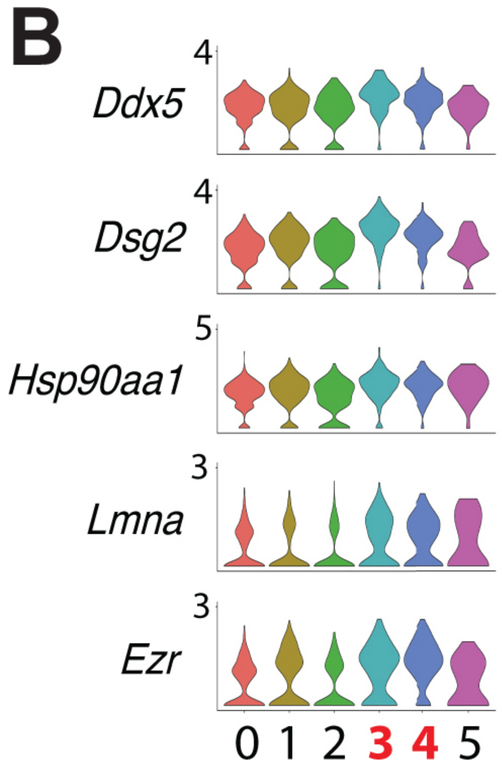
DBA
CXCL5

Human

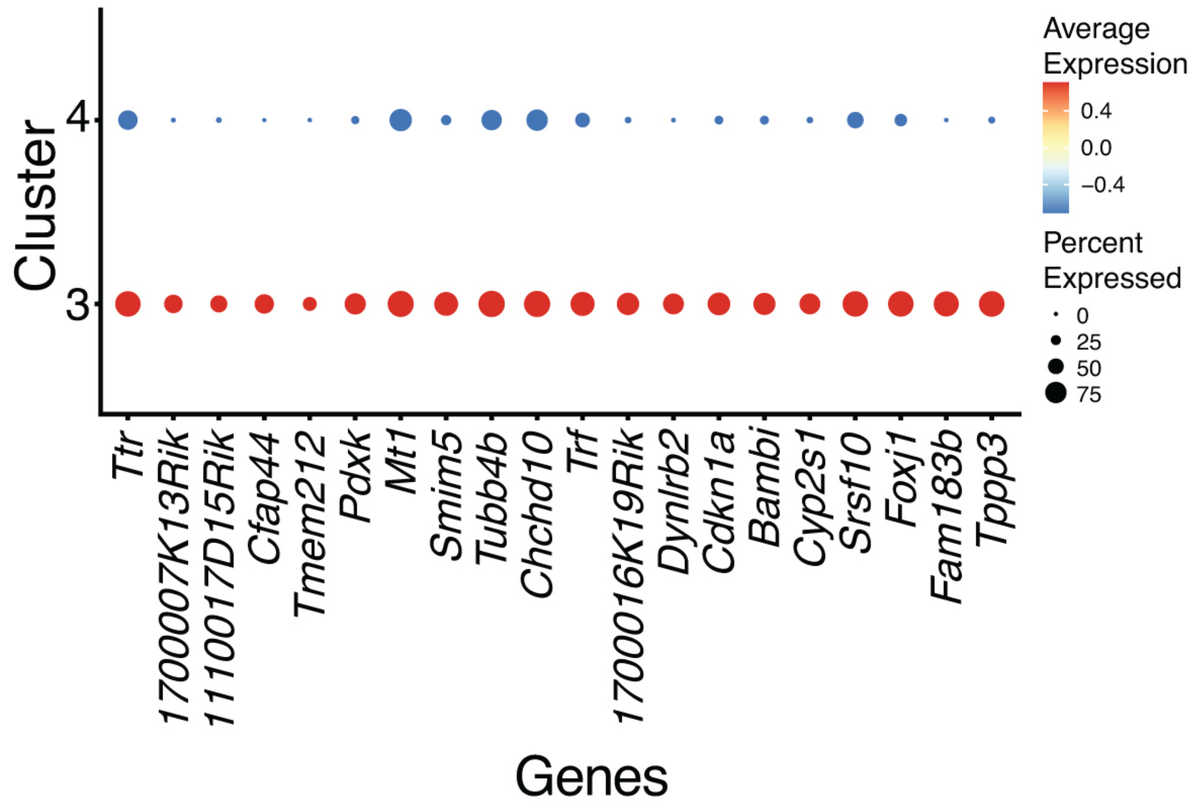
CK19
CXCL5

Cluster 3 vs 4

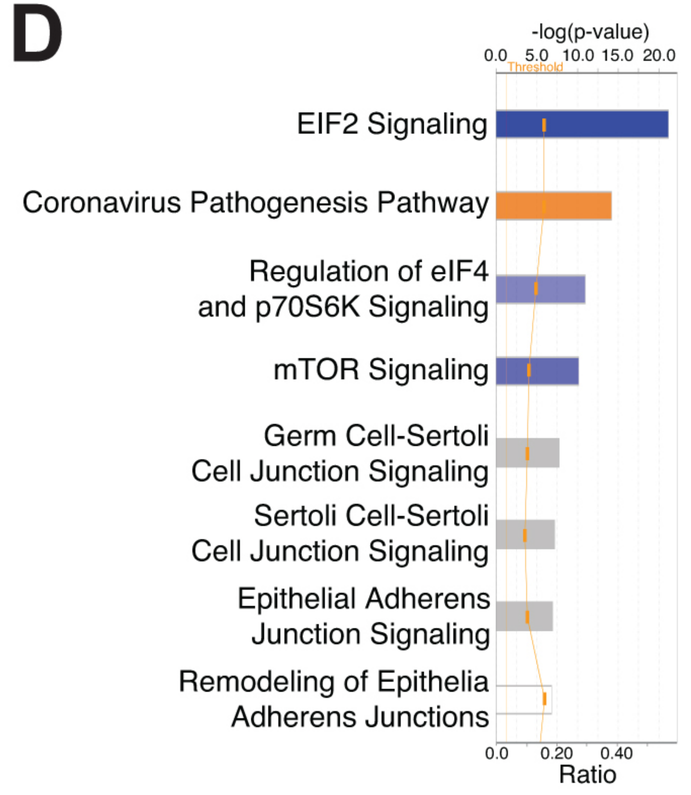
B

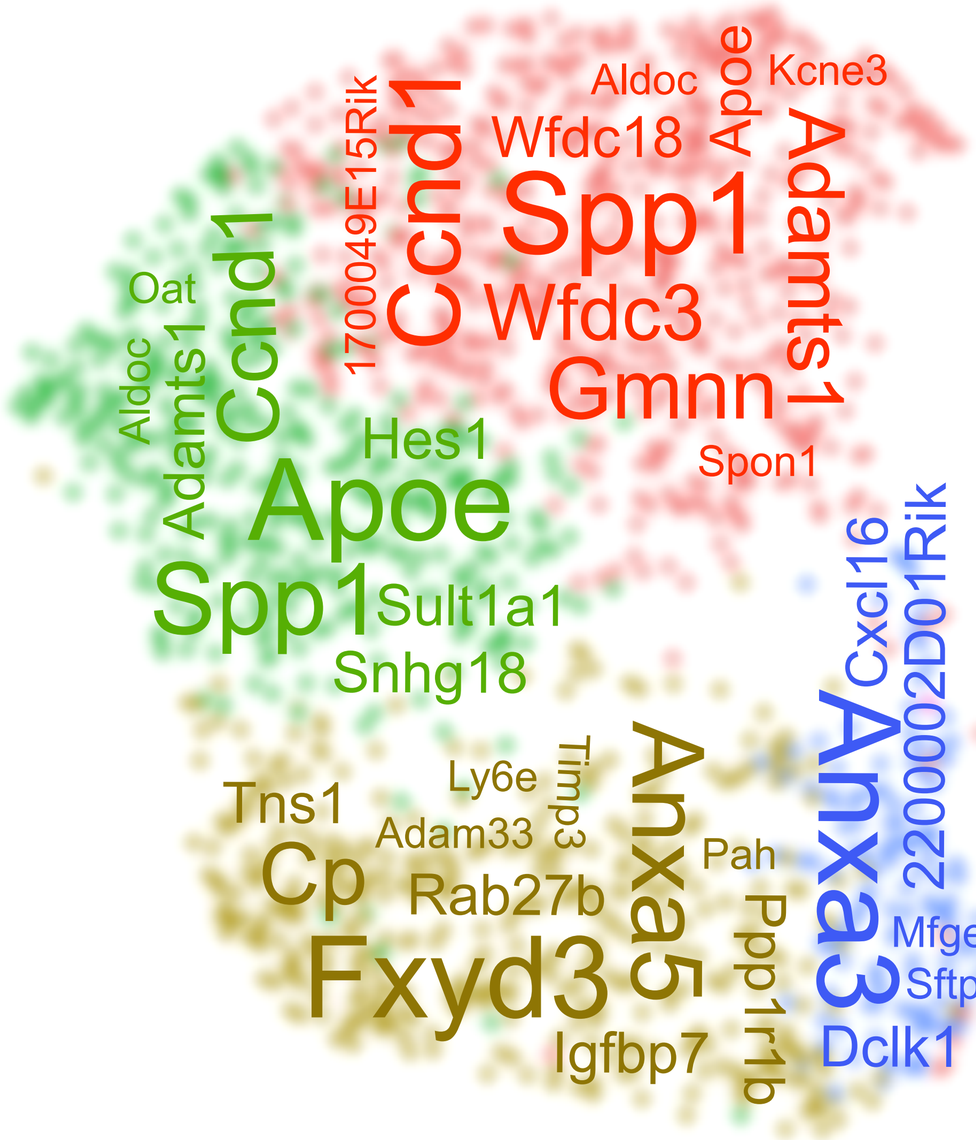


C

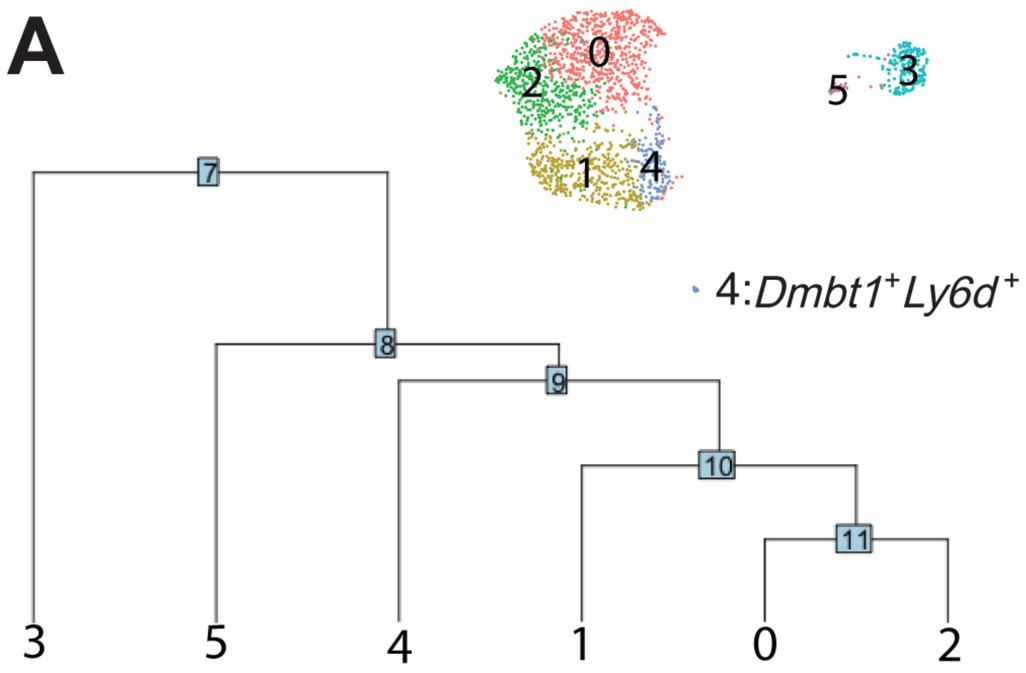


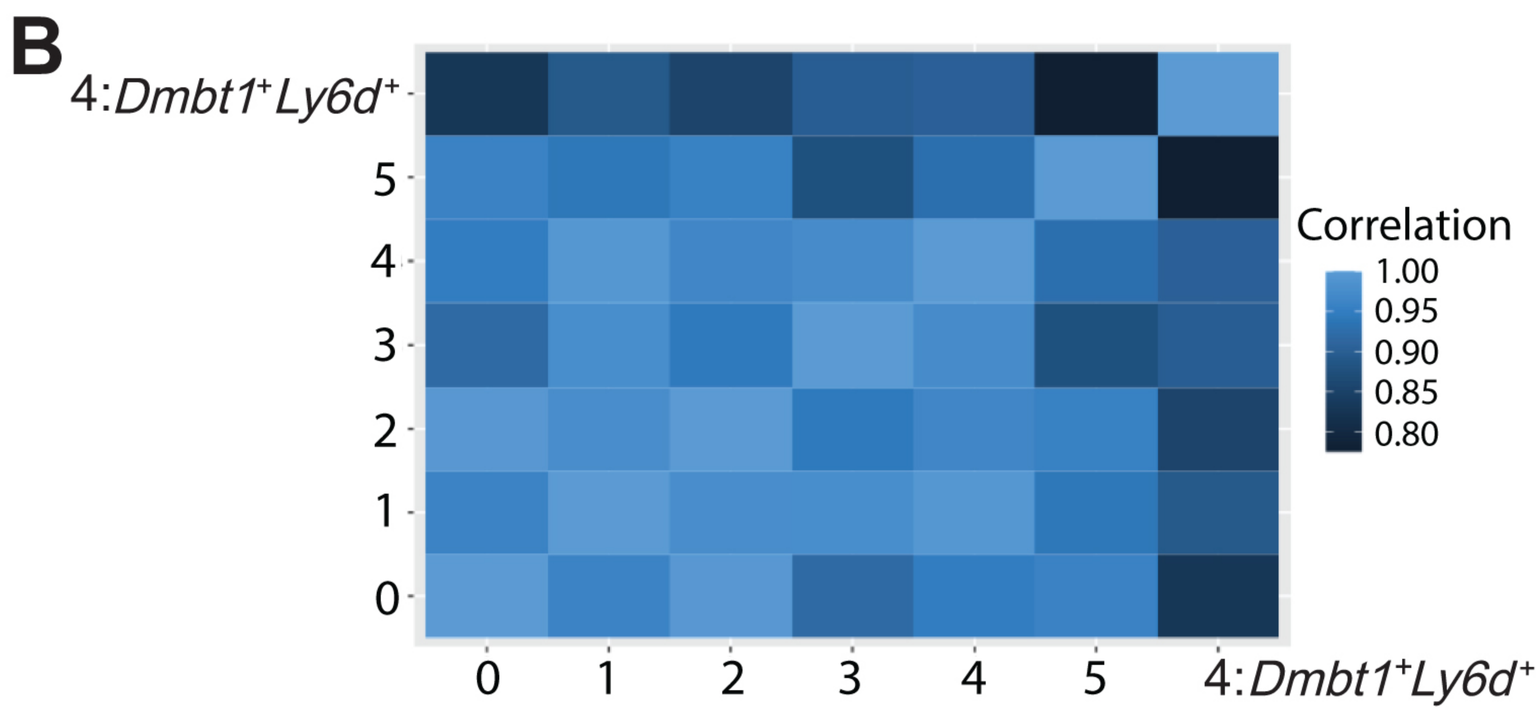
D



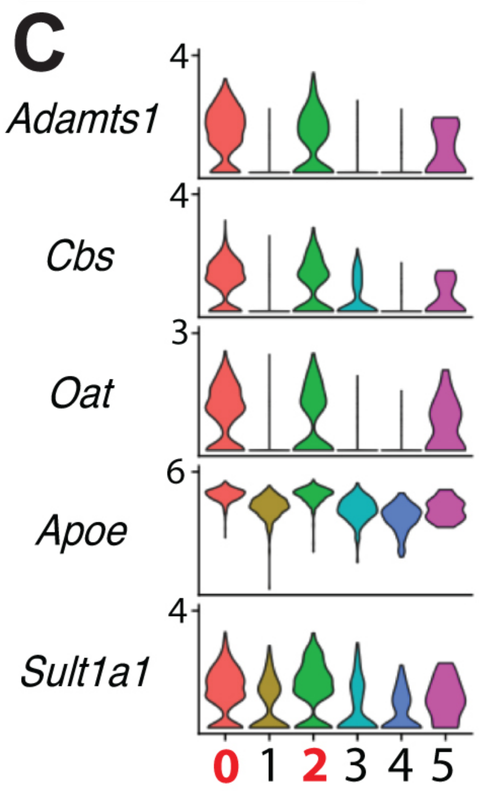

Aldoc
Apoe
Kcne3
Wfdc18
1700049E15Rik
Ccnd1
Spp1
Adamts1
Oat
Aldoc
Adamts1
Ccnd1
Wfdc3
Gmnn
Hes1
Spon1
Apoe
Cxcl16
2200002D01Rik
Spp1
Sult1a1
Snhg18
Anxa3
Ly6e
Timp3
Tns1
Adam33
Anxa5
Pah
Cp
Rab27b
Ppp1r1b
Mfge8
Fxyd3
Sftpd
Igfbp7
Dclk1

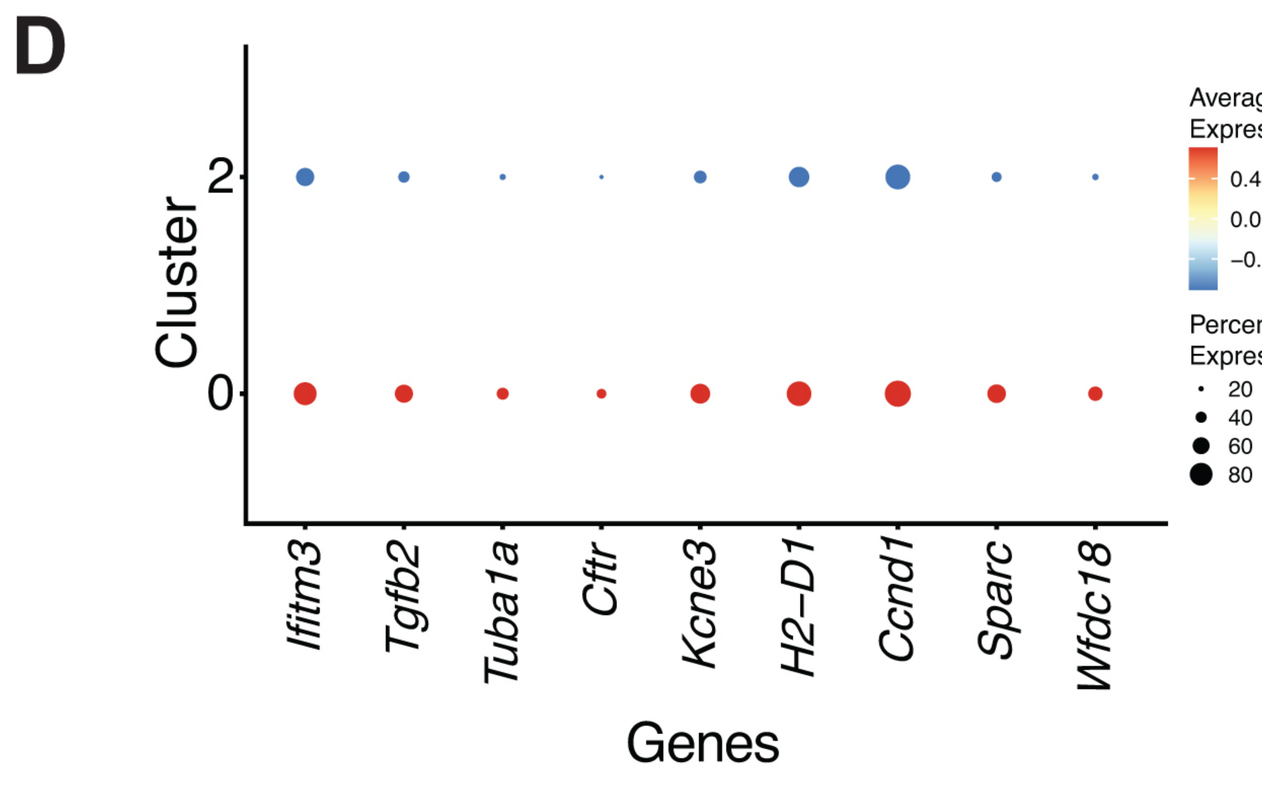

Cyp2s1
Dynlrb2
Fam183b
Tppp3
Cfap44
Cenpf
Cenpe

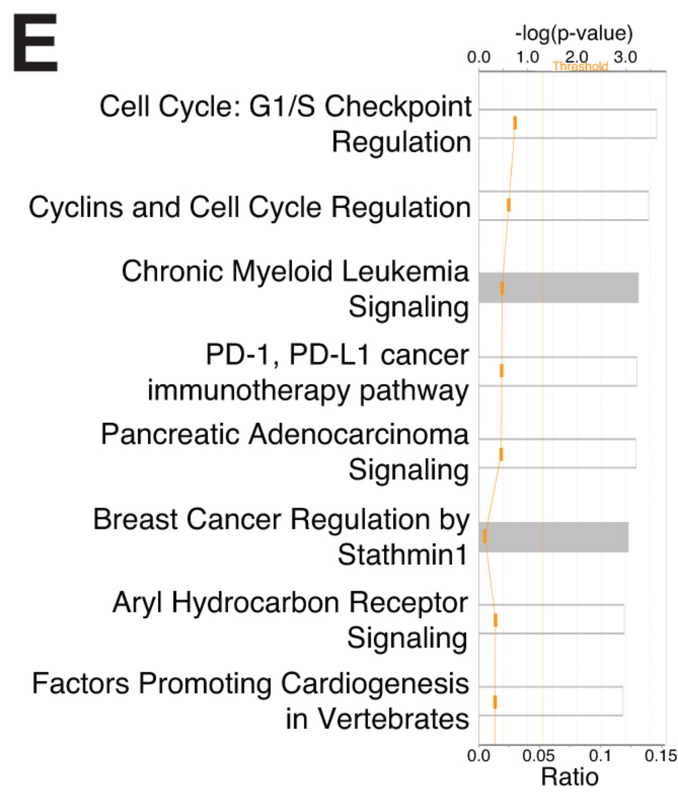




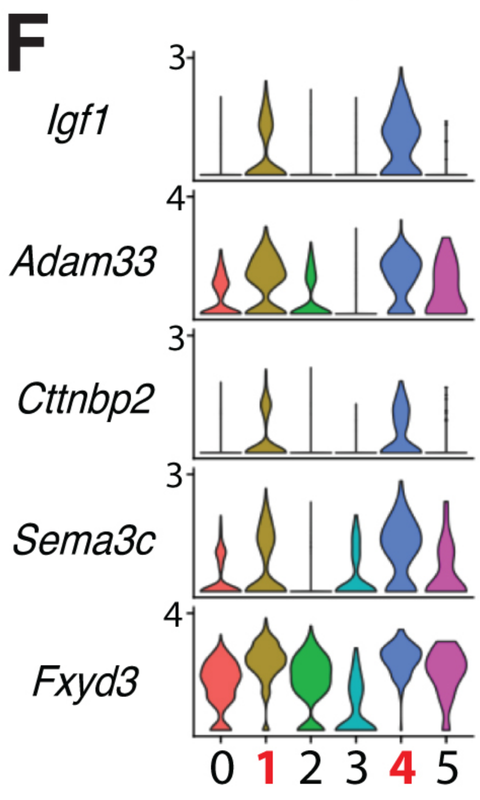
Cluster 0 vs 2

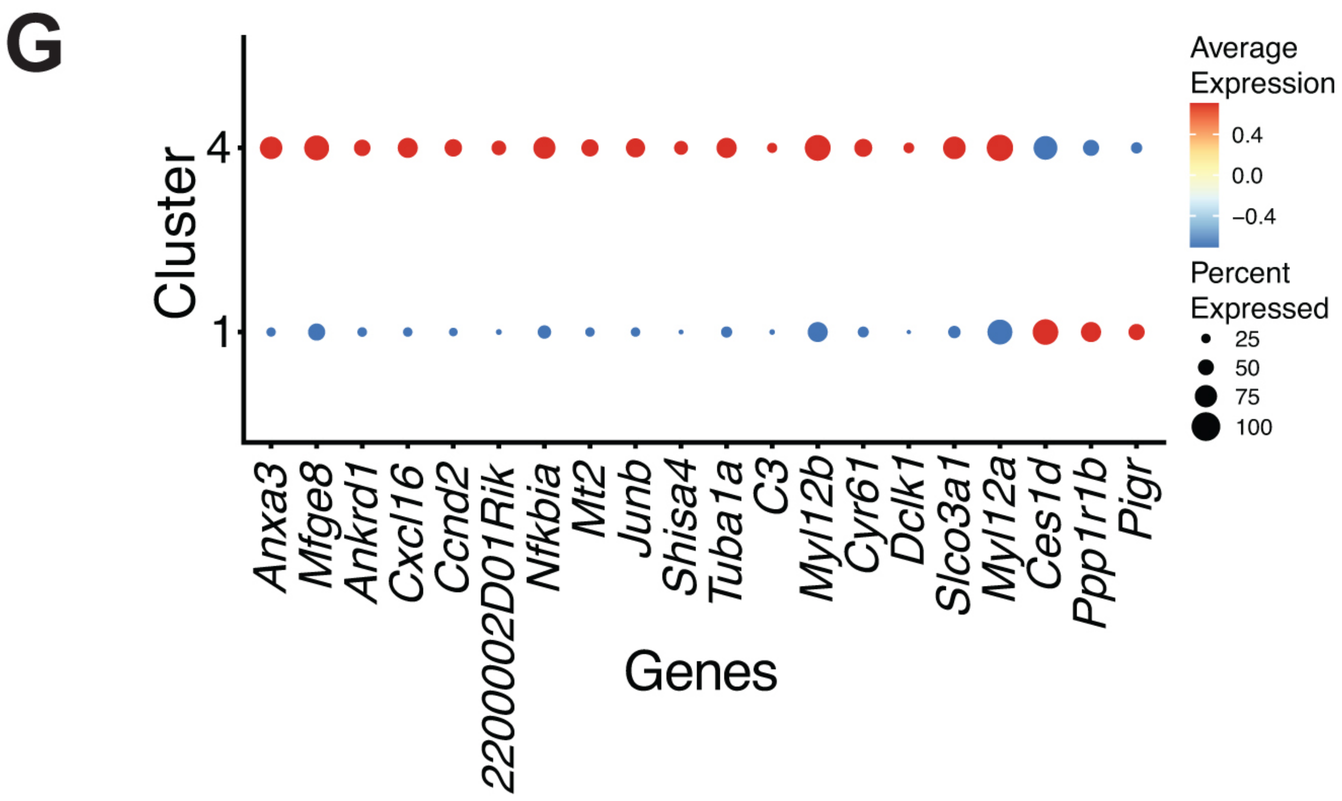


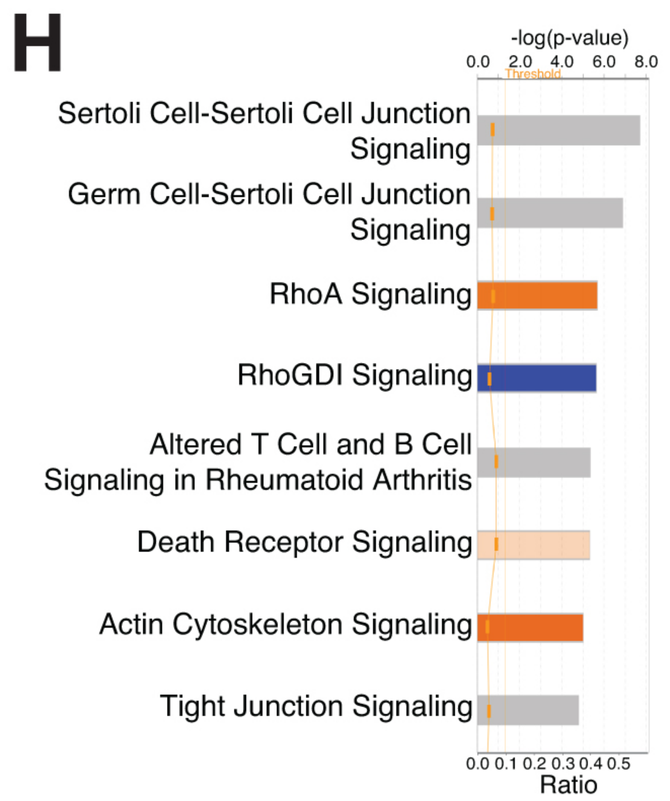




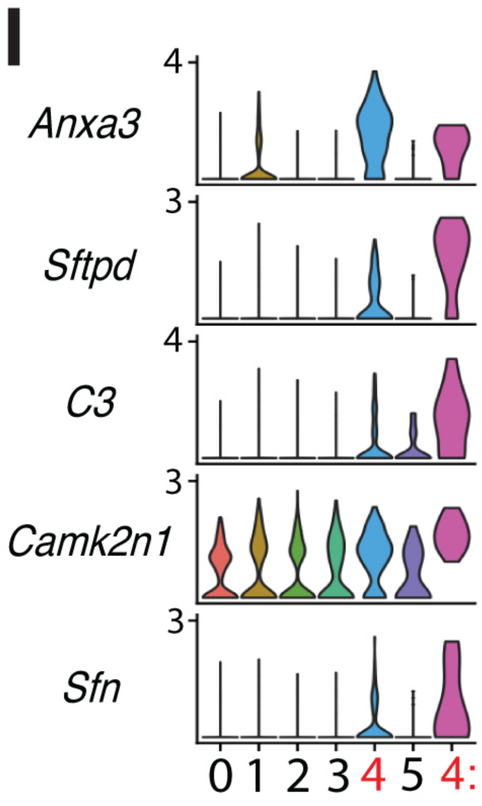
Cluster 4 vs 1

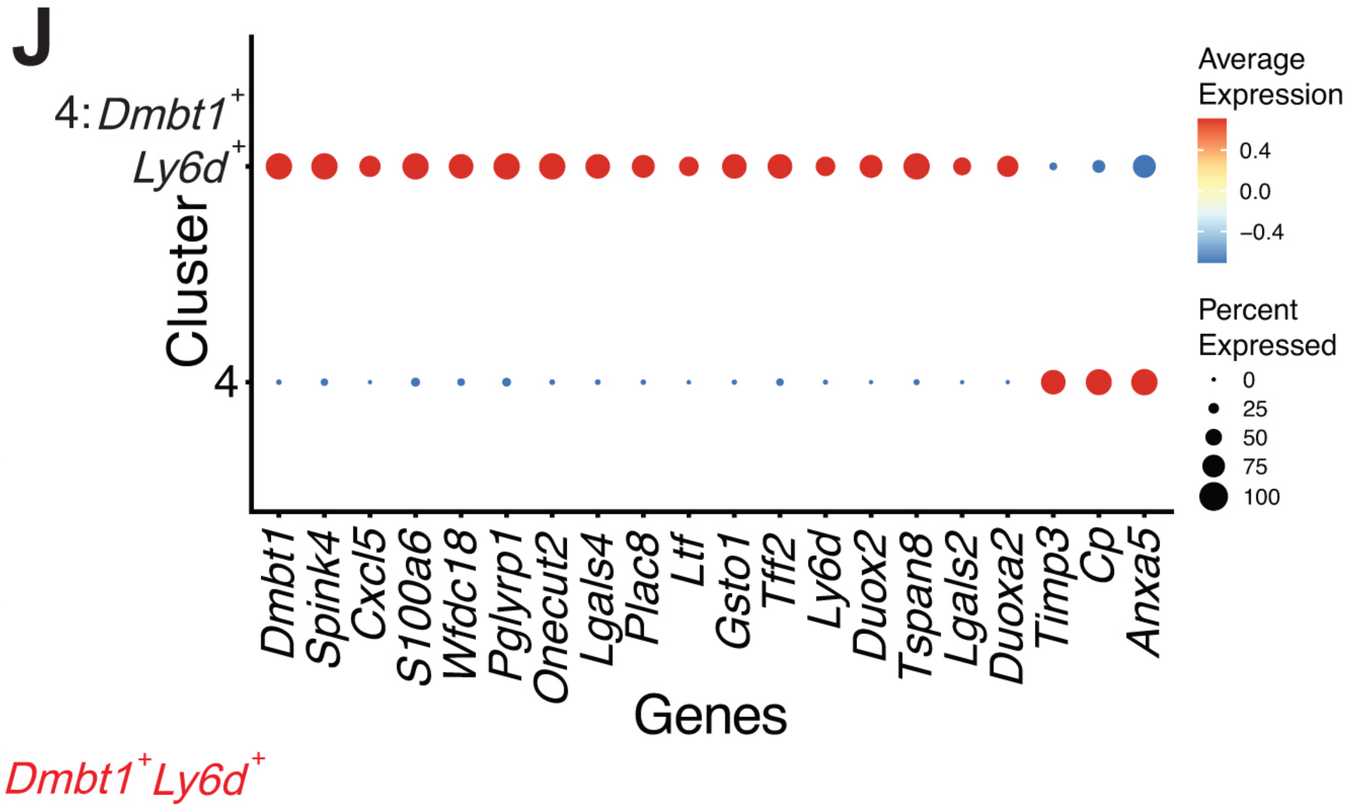


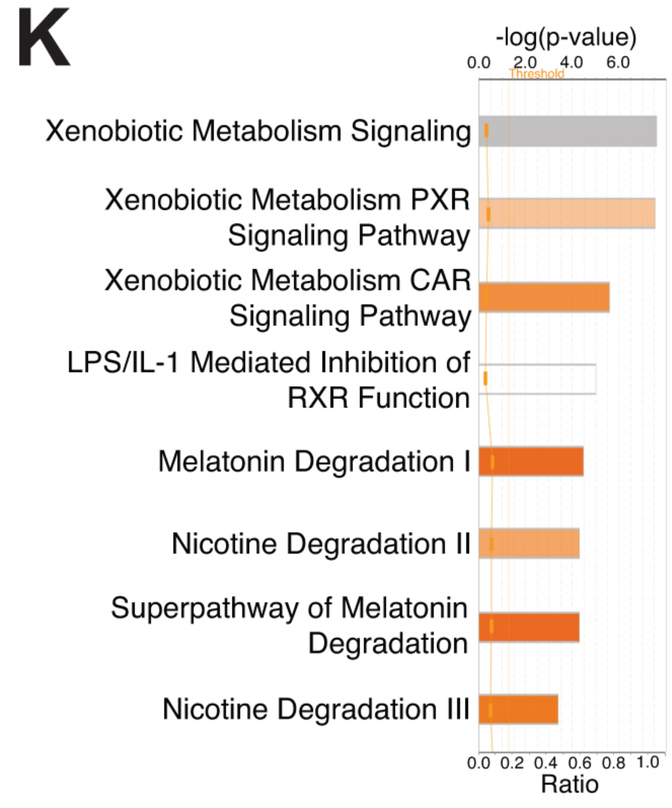


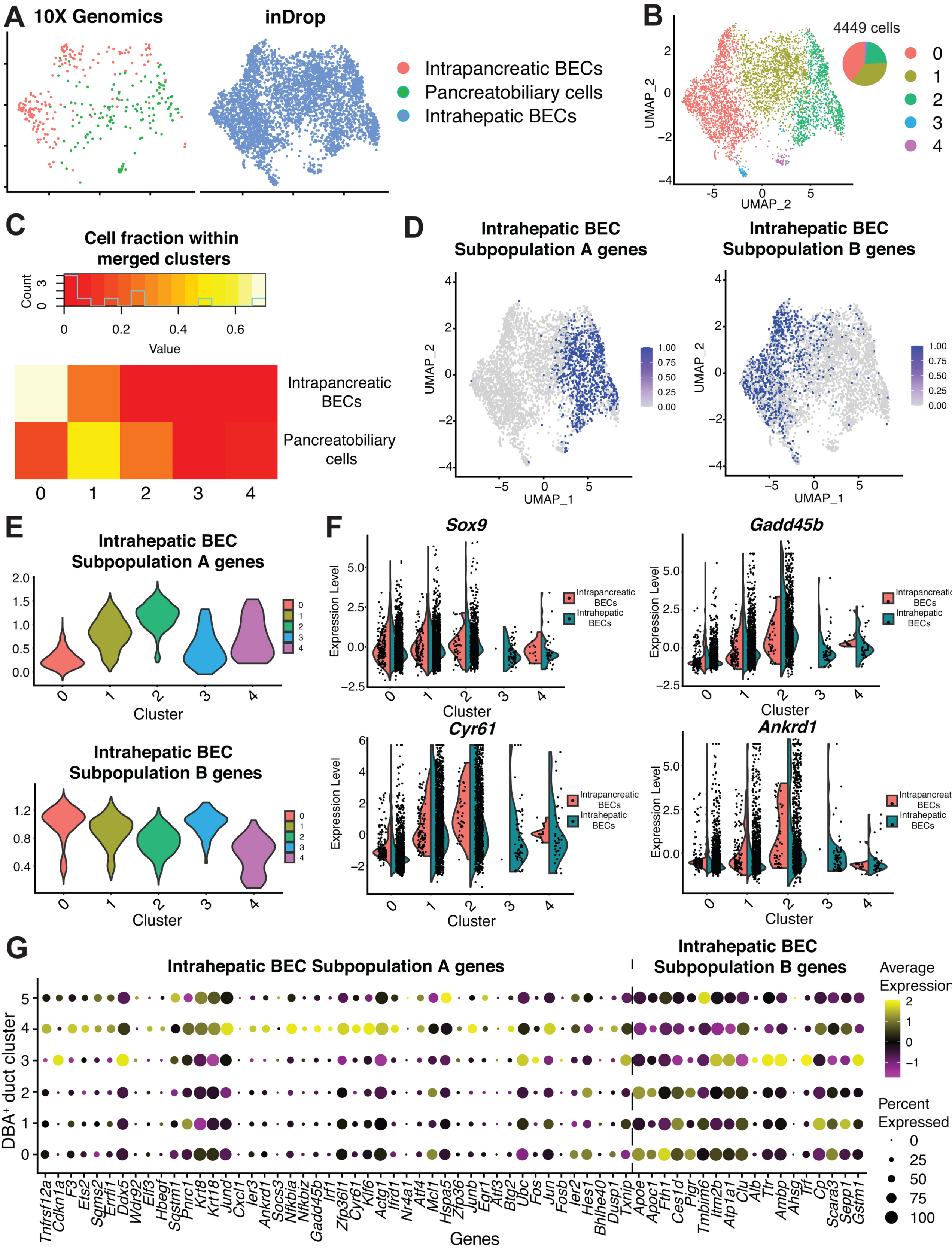


Cluster 4: *Dmbt1*$^{+}$*Ly6d*$^{+}$ vs Cluster 4

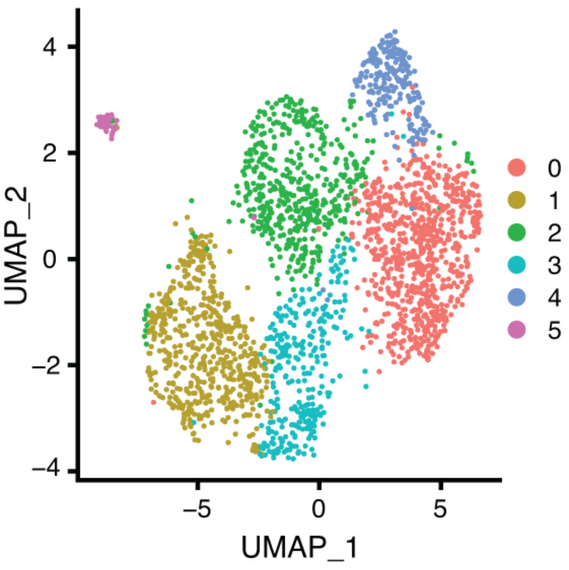




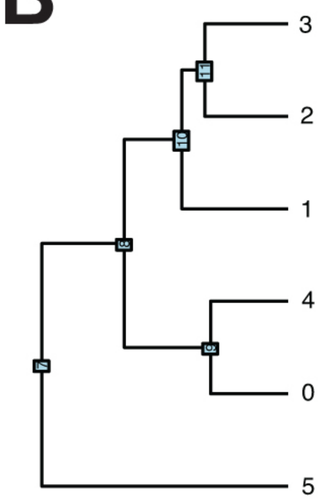



A
10X Genomics
inDrop
Intrapancreatic BECs
Pancreatobiliary cells
Intrahepatic BECs
B
4449 cells
0
1
2
3
4
UMAP_2
UMAP_2
C
Cell fraction within merged clusters
Count
Value
Intrapancreatic BECs
Pancreatobiliary cells
D
Intrahepatic BEC Subpopulation A genes
Intrahepatic BEC Subpopulation B genes
UMAP_1
E
Intrahepatic BEC Subpopulation A genes
Intrahepatic BEC Subpopulation B genes
Cluster
F
Sox9
Gadd45b
Cyr61
Ankrd1
Expression Level
Intrapancreatic BECs
Intrahepatic BECs
G
Intrahepatic BEC Subpopulation A genes
Intrahepatic BEC Subpopulation B genes
DBA+ duct cluster
Genes
Average Expression
Percent Expressed
Tnfrsf12a
Cdkn1a
F3
Ets2
Sqms2
Errfi1
Ddx5
Wdr92
Elf3
Hbegf
Sqstm1
Pnrc1
Krt8
Krt18
Jund
Cxcl1
Ier3
Ankrd1
Socs3
Nfkbia
Nfkbiz
Gadd45b
Irf1
Zfp36l1
Cyr61
Klf6
Actg1
Ifrd1
Nr4a1
Atf4
Mcl1
Hspa5
Zfp36
Junb
Egr1
Atf3
Btg2
Ubc
Fos
Jun
Fosb
Ier2
Hes1
Bhlhe40
Dusp1
Txnip
Apoe
Apoc1
Fth1
Ces1d
Pigr
Tmbim6
Itm2b
Atp1a1
Clu
Alb
Ttr
Ambp
Ahsg
Trf
Cp
Scara3
Sepp1
Gstm1

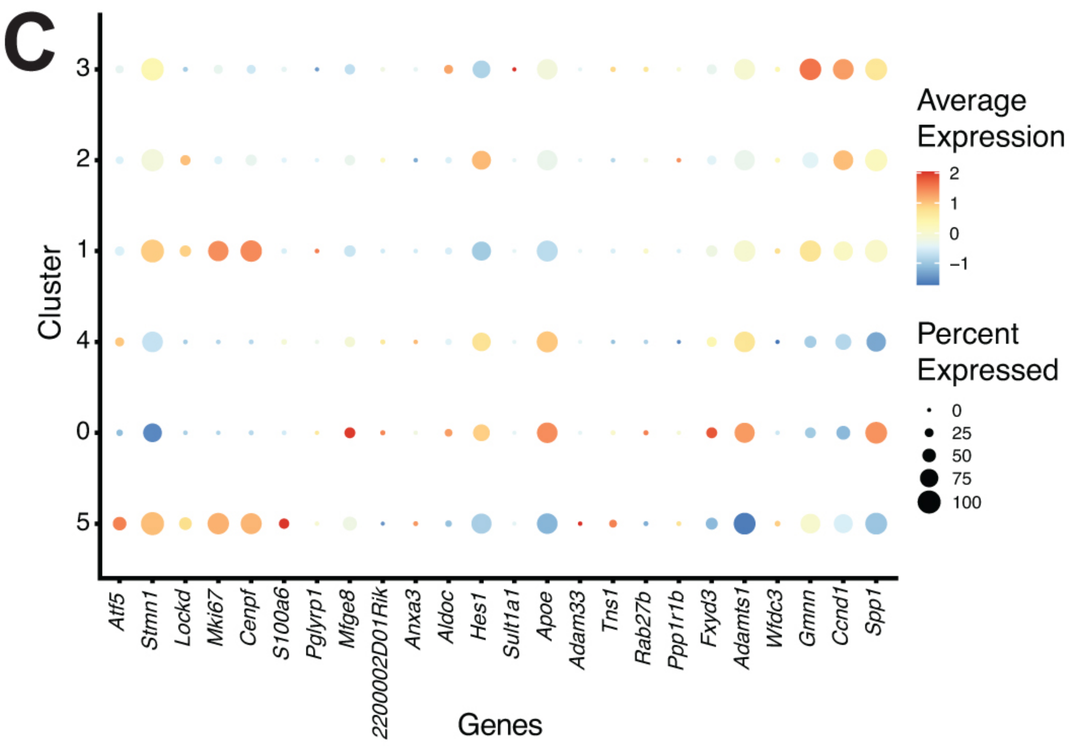
A E12.5



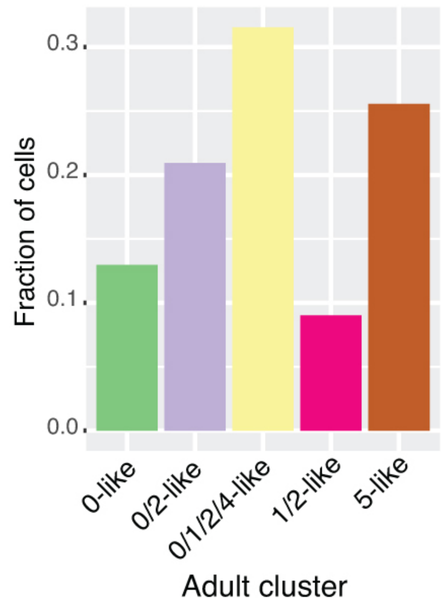
B



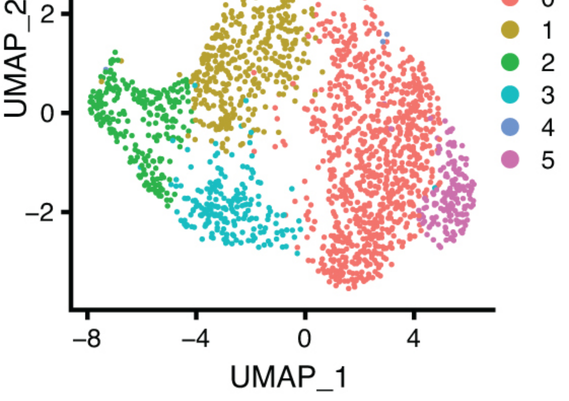
C



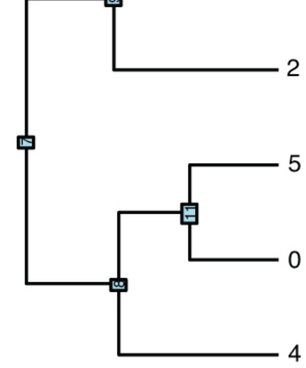
D



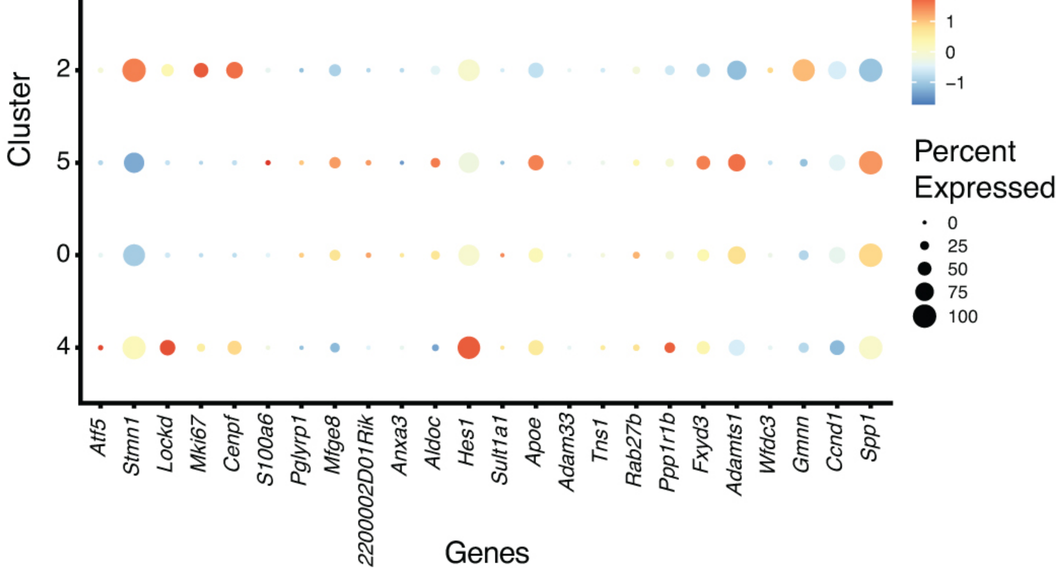
E E14.5



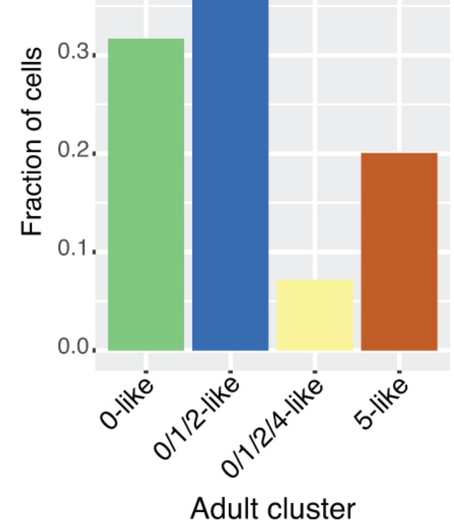
F



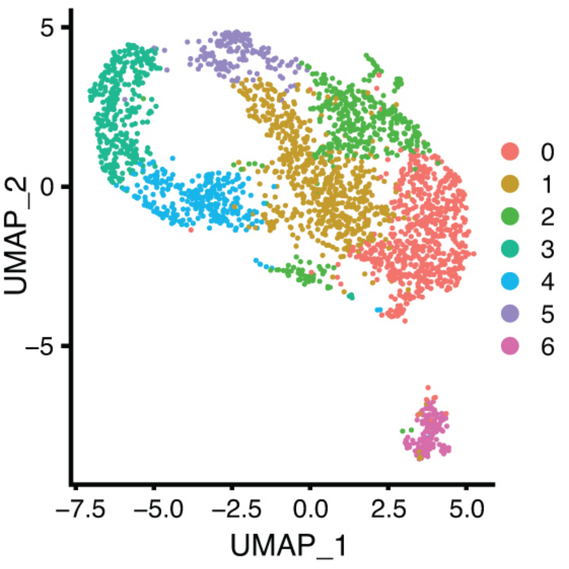
G



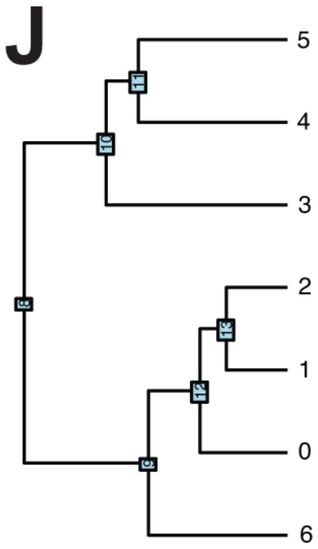
H



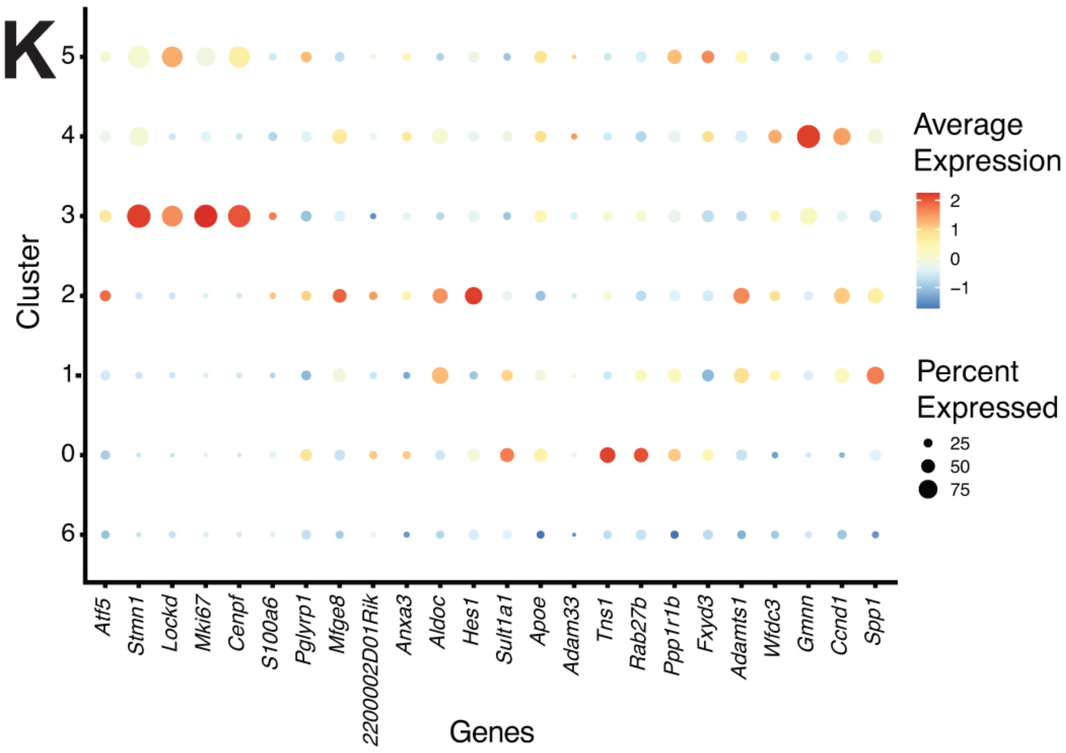
I E17.5



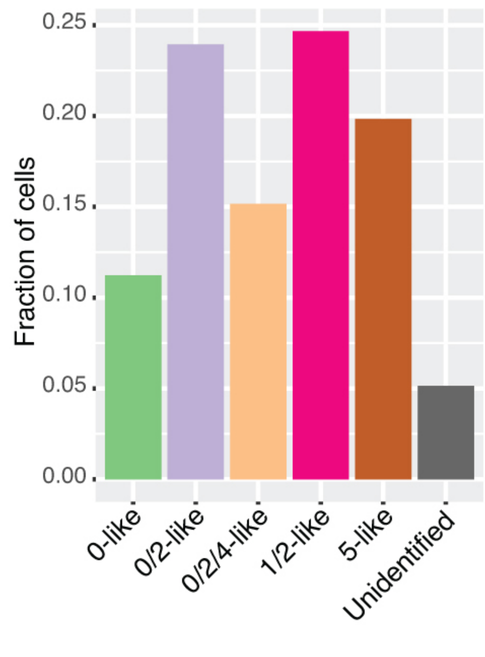
J



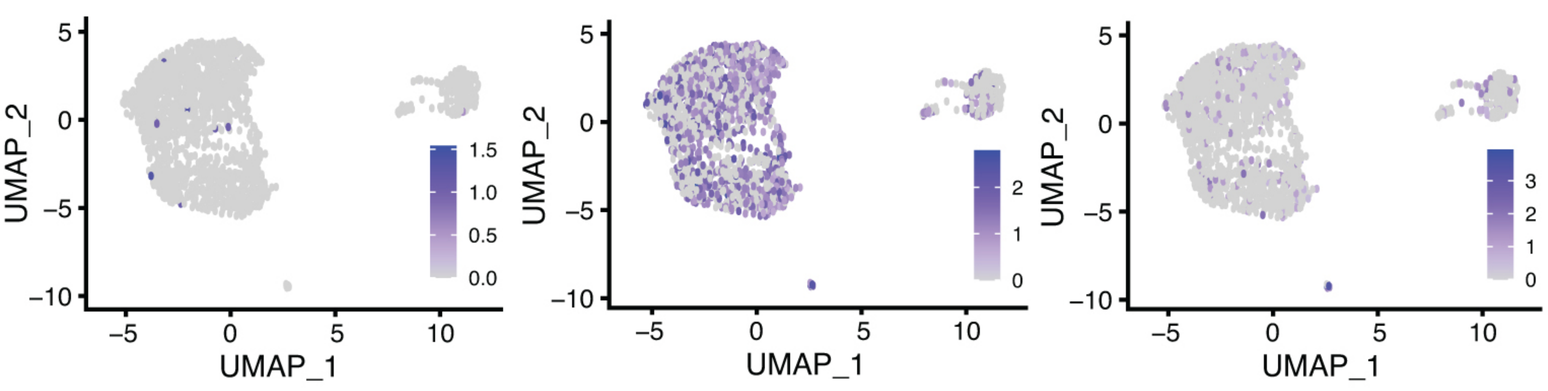
K



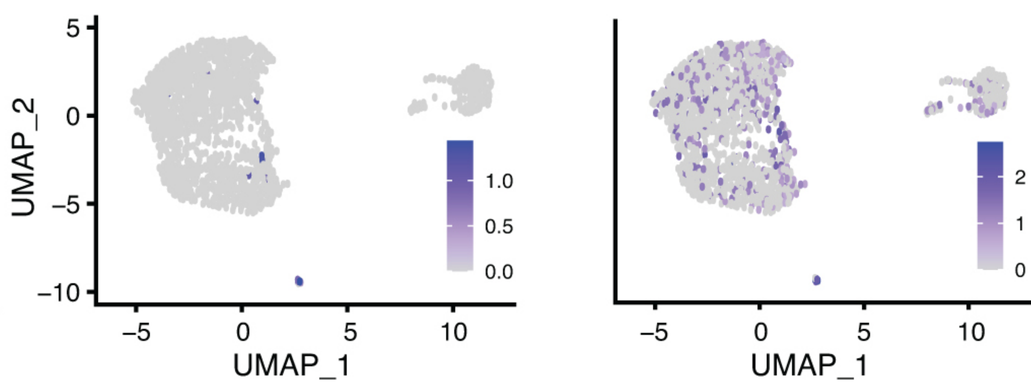
L



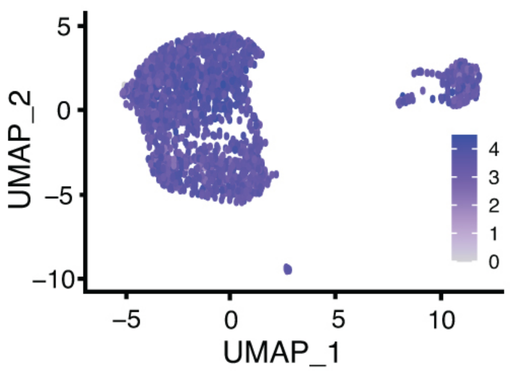
M



N



O

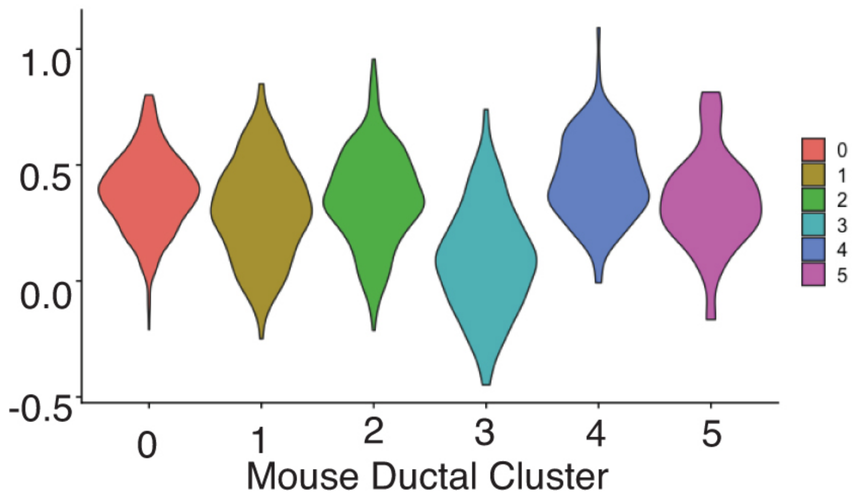


A

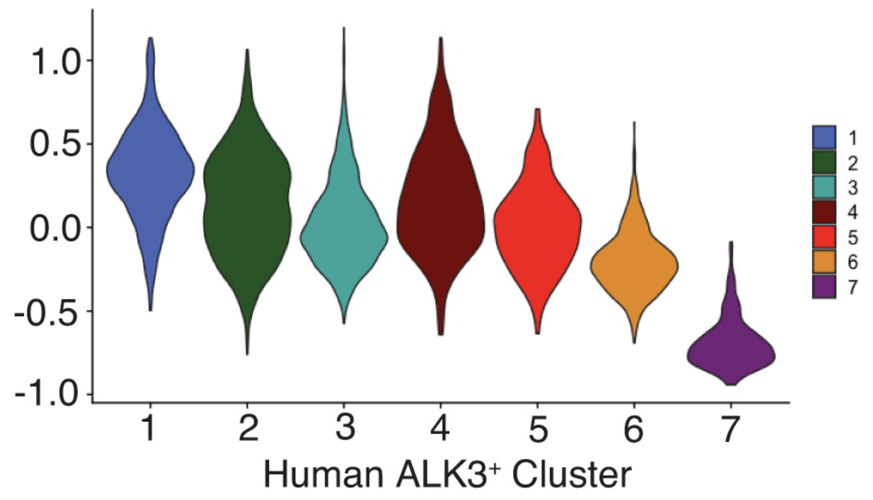
Control features

MGI.symbol	HGNC.symbol
Spp1	*SPP1*
Epcam	*EPCAM*
Cftr	*CFTR*
Car2	*CA2*
Sox9	*SOX9*
Ctnnd1	*CTNND1*
Anxa3	*ANXA3*
Muc1	*MUC1*
Cdh1	*CDH1*
Cldn7	*CLDN7*

B



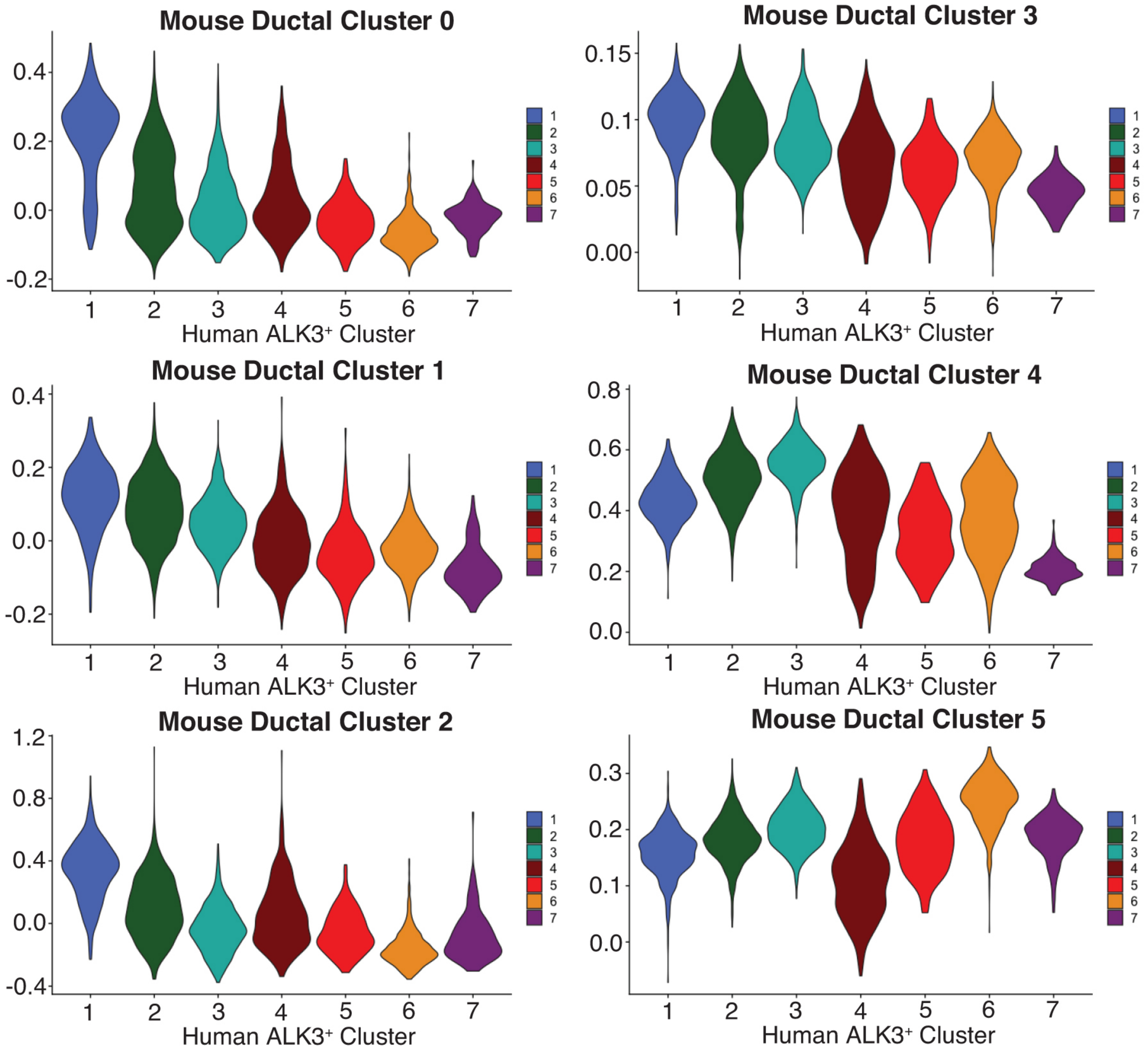
C



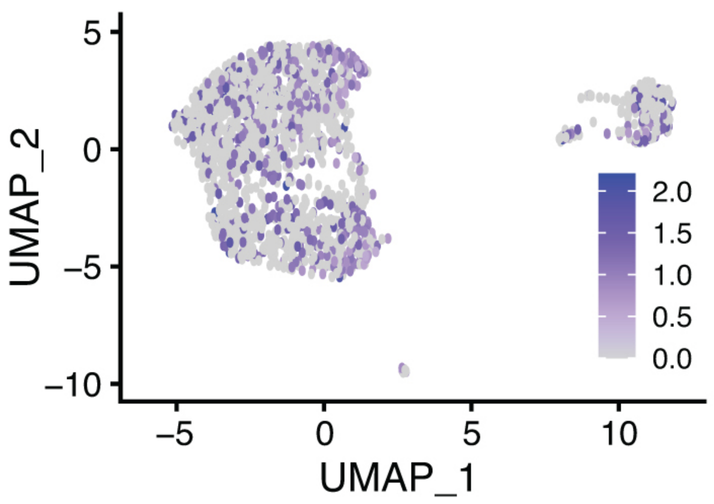
D

Mouse DBA+ Cluster	Mouse gene	Matched in Human
Cluster 0	53	53
Cluster 1	29	29
Cluster 2	11	11
Cluster 3	502	501
Cluster 4	160	159
Cluster 5	155	154

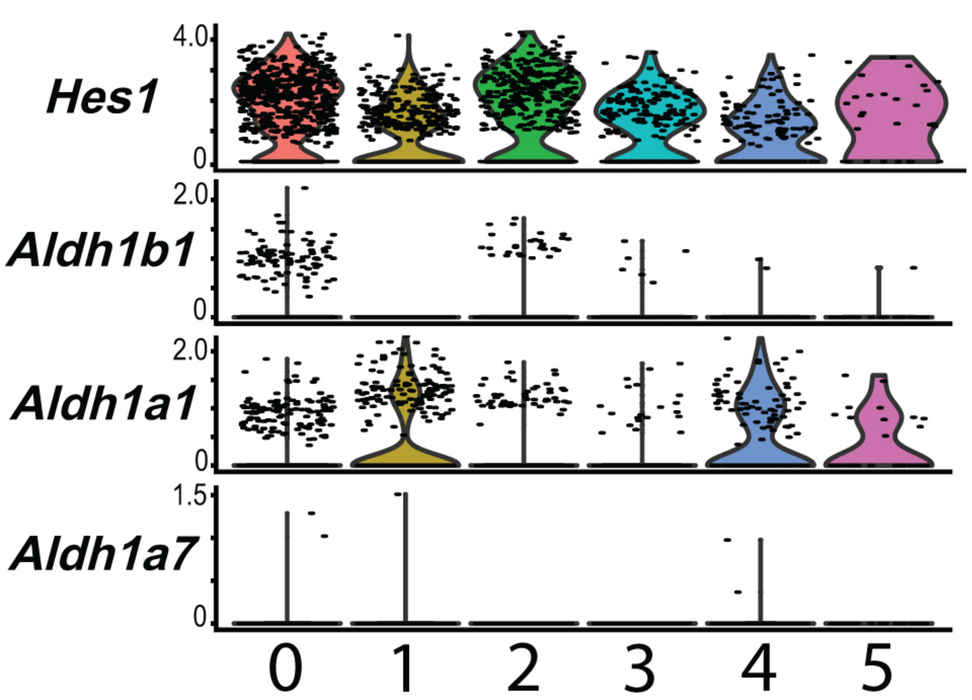
E



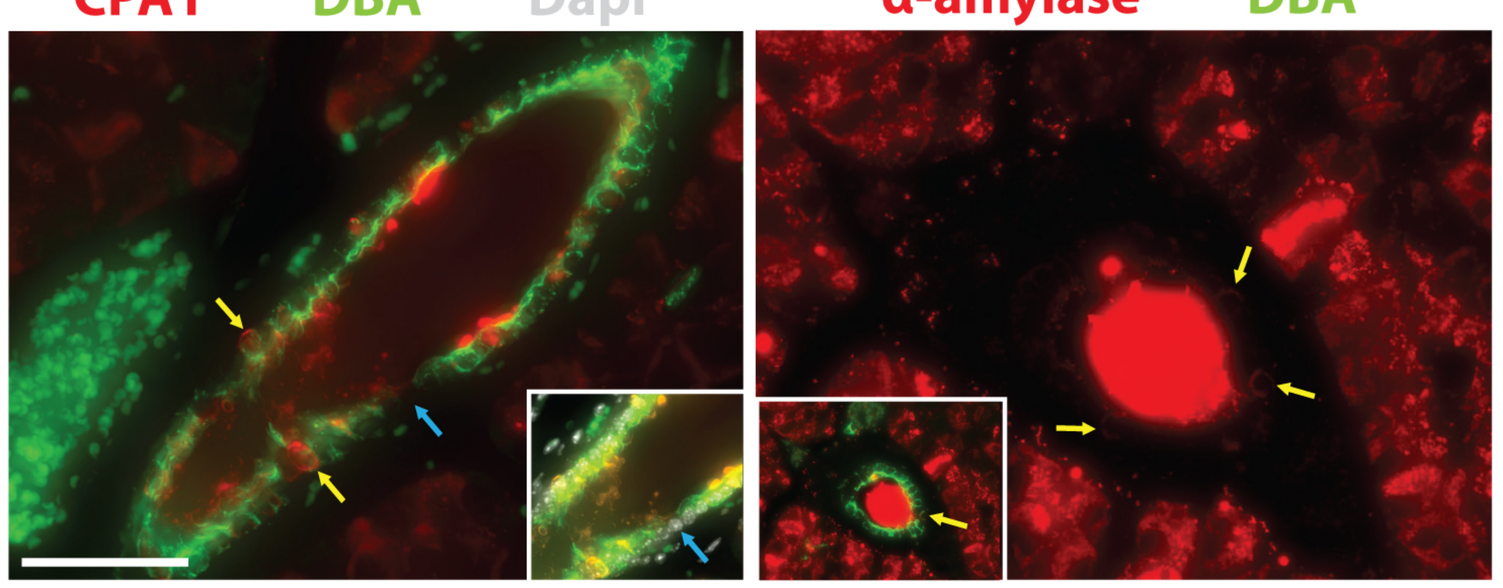
F

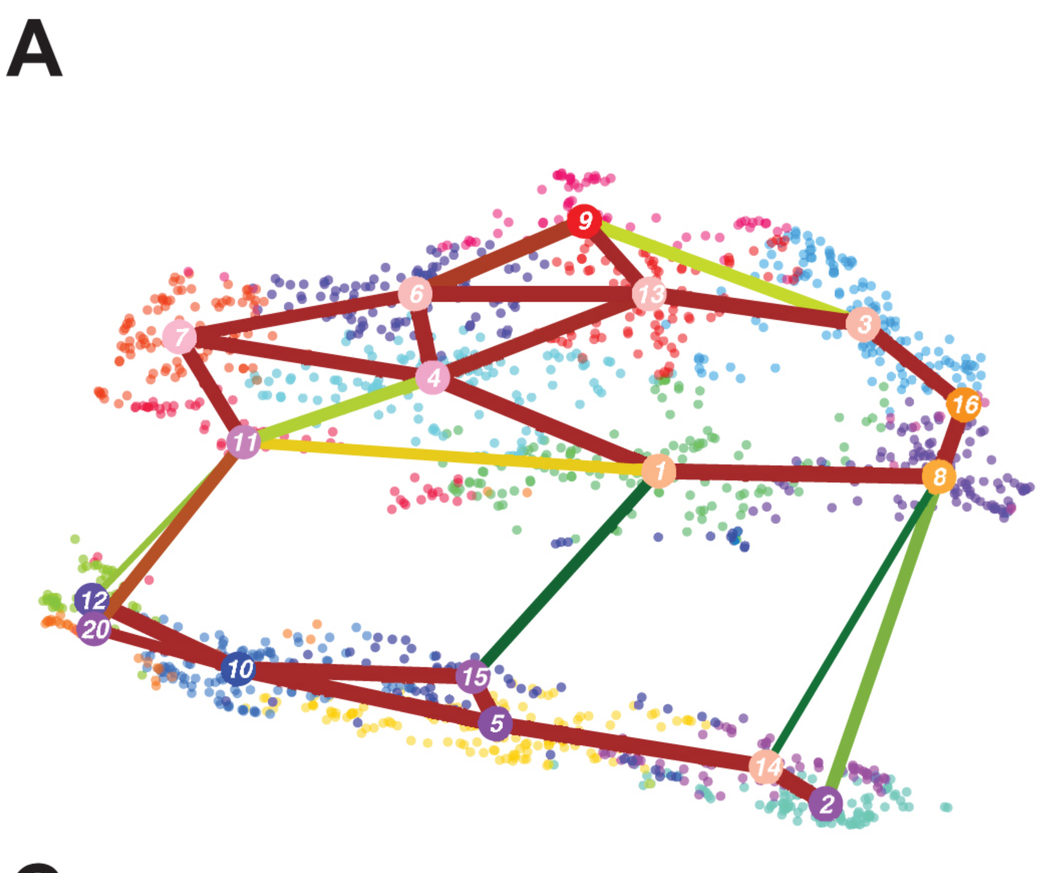


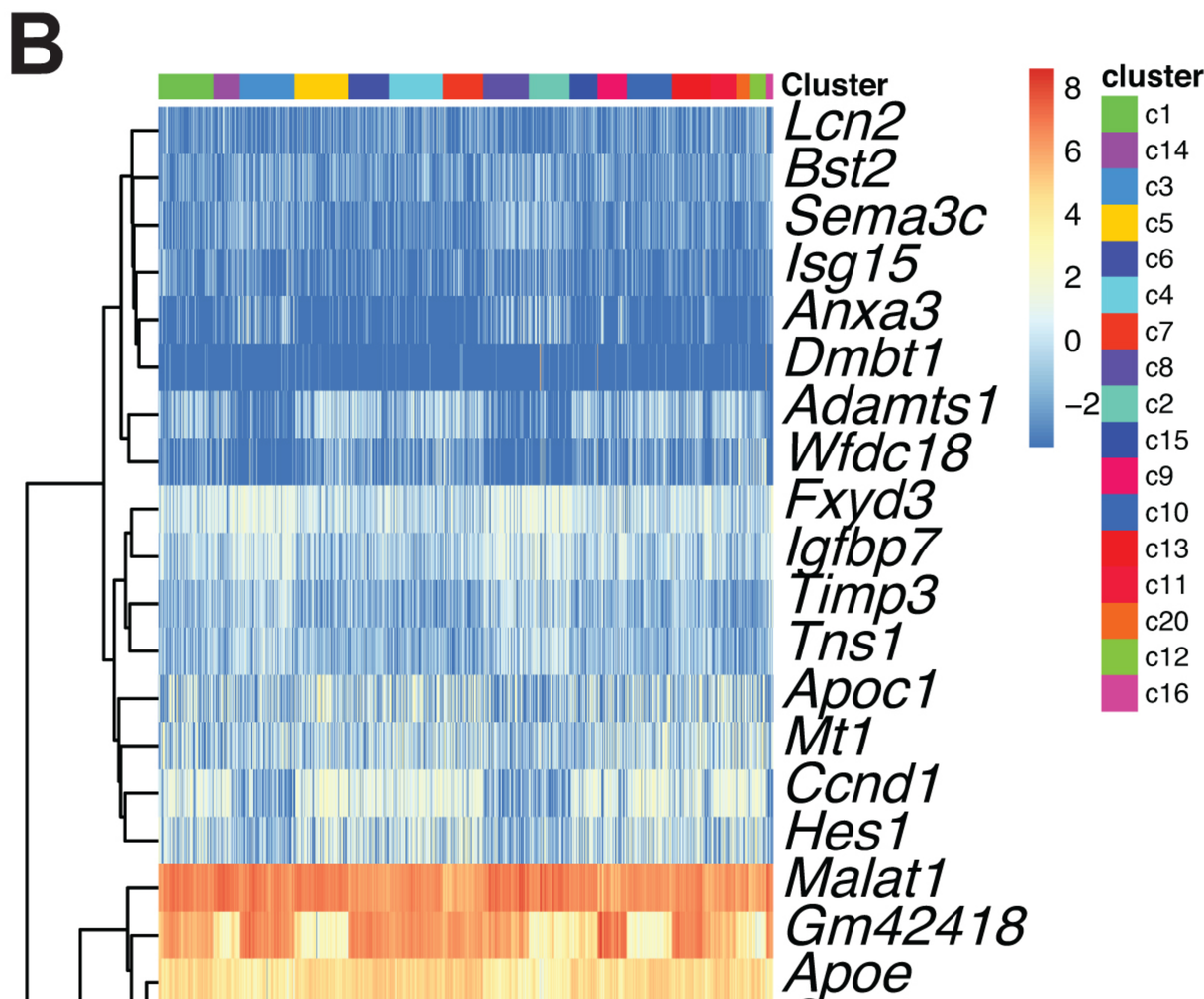
G

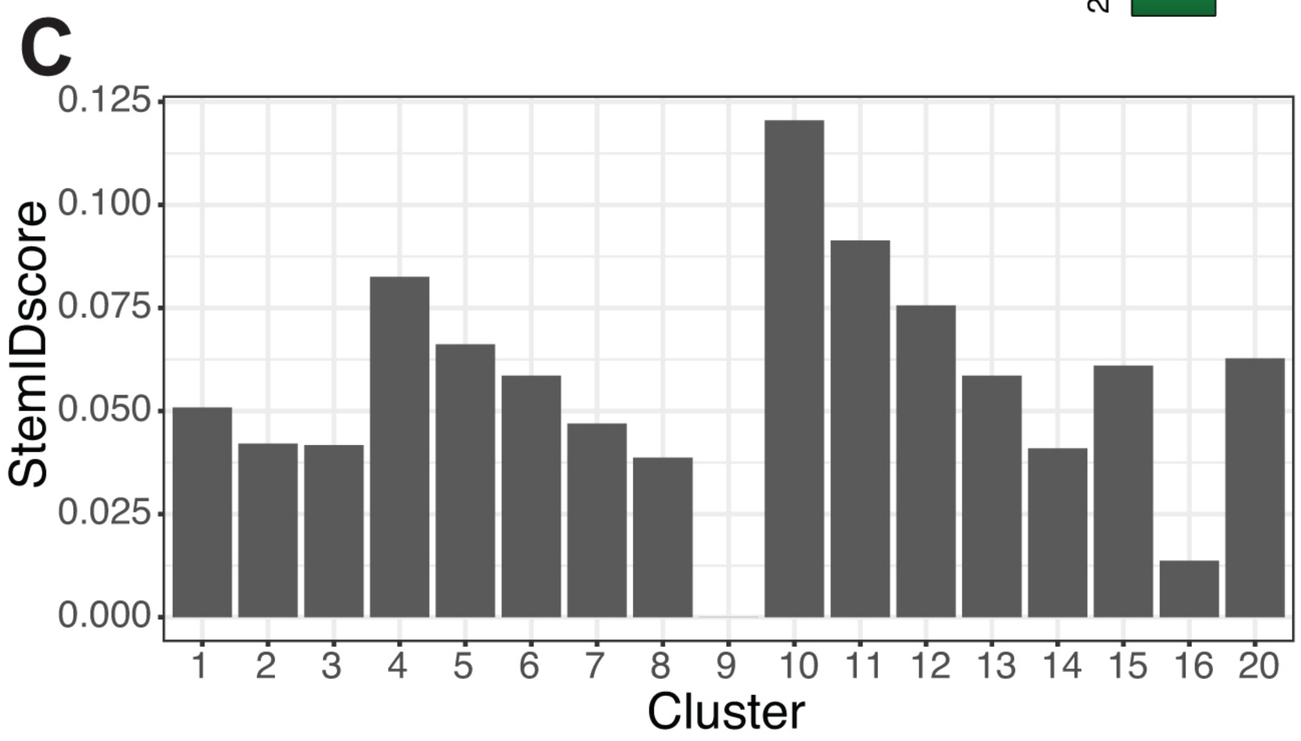


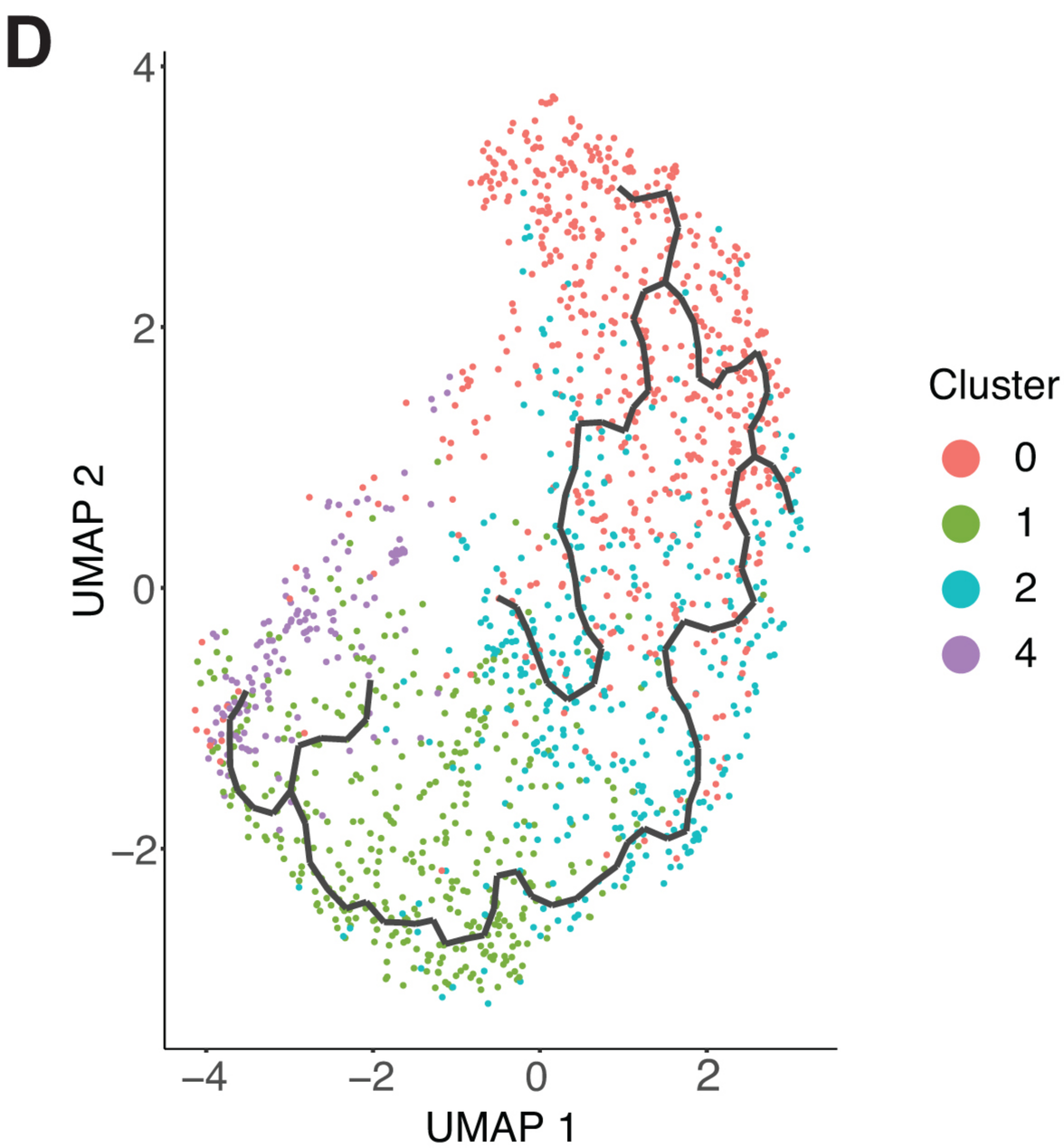
H

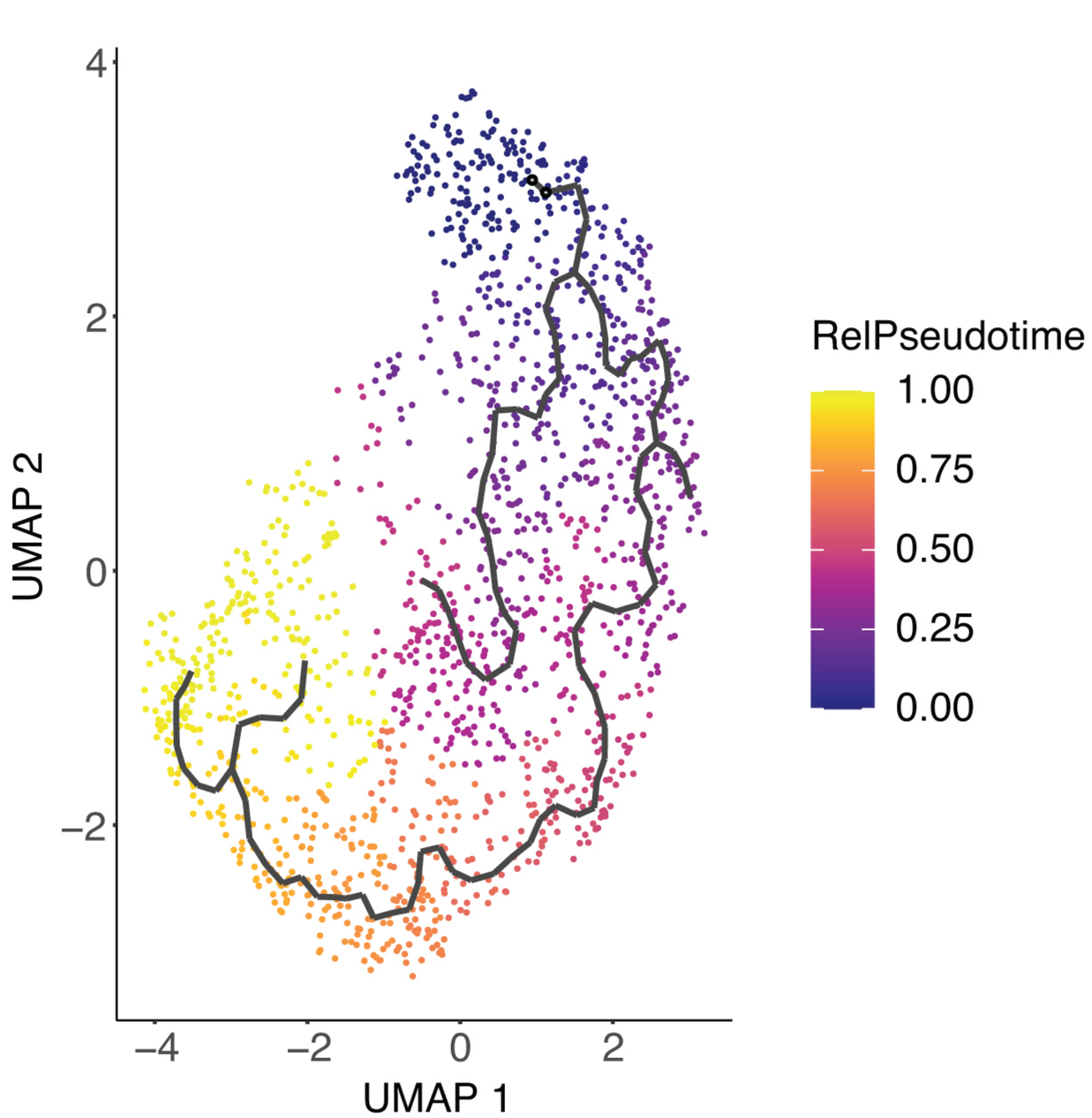



A
Entropy
0.710
0.700
0.690
-log10 p-value
5.0 4.5 4.0 3.5 3.0 2.5 2.0
9
6
13
3
7
4
16
11
1
8
12
20
10
15
5
14
2

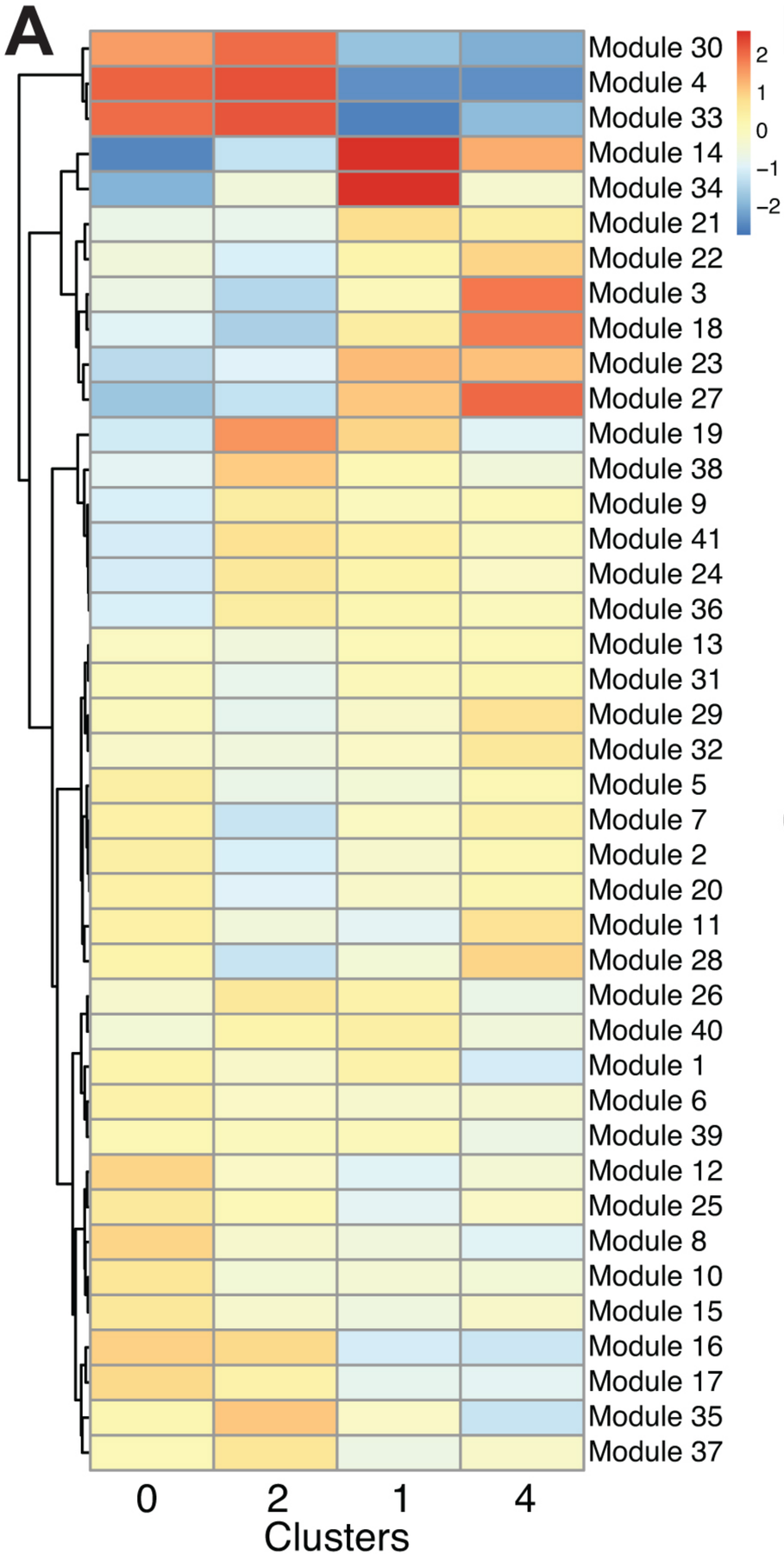

B
Cluster
Lcn2
Bst2
Sema3c
Isg15
Anxa3
Dmbt1
Adamts1
Wfdc18
Fxyd3
Igfbp7
Timp3
Tns1
Apoc1
Mt1
Ccnd1
Hes1
Malat1
Gm42418
Apoe
Spp1
Anxa5
Dbi
Fth1
AY036118
Actb
Rplp0
Ces1d
Cp
8
6
4
2
0
−2
cluster
c1
c14
c3
c5
c6
c4
c7
c8
c2
c15
c9
c10
c13
c11
c20
c12
c16


C
StemIDscore
0.125
0.100
0.075
0.050
0.025
0.000
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 20
Cluster

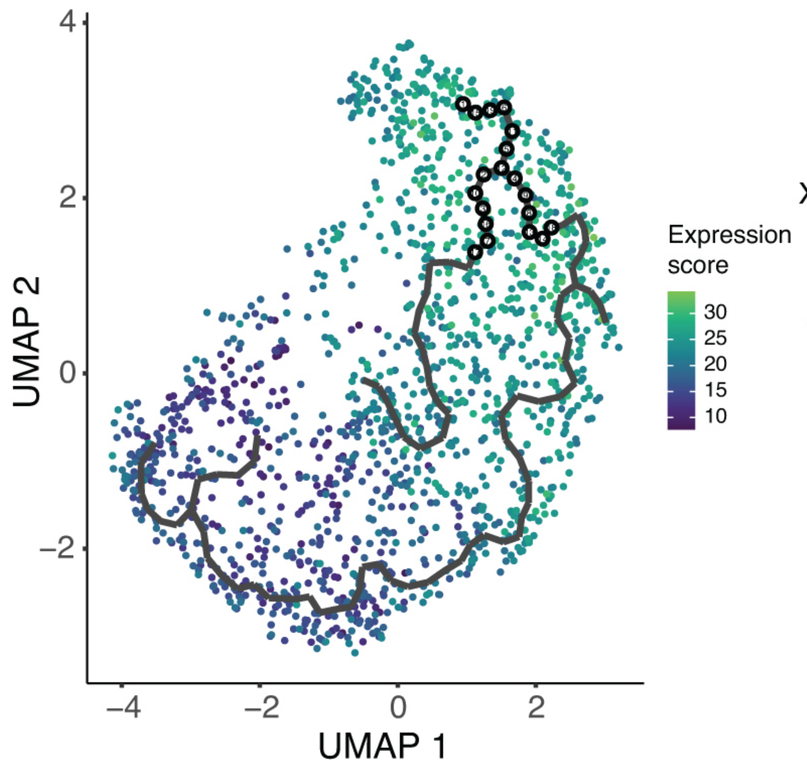

D
UMAP 2
UMAP 1
4
2
0
−2
−4 −2 0 2
Cluster
0
1
2
4

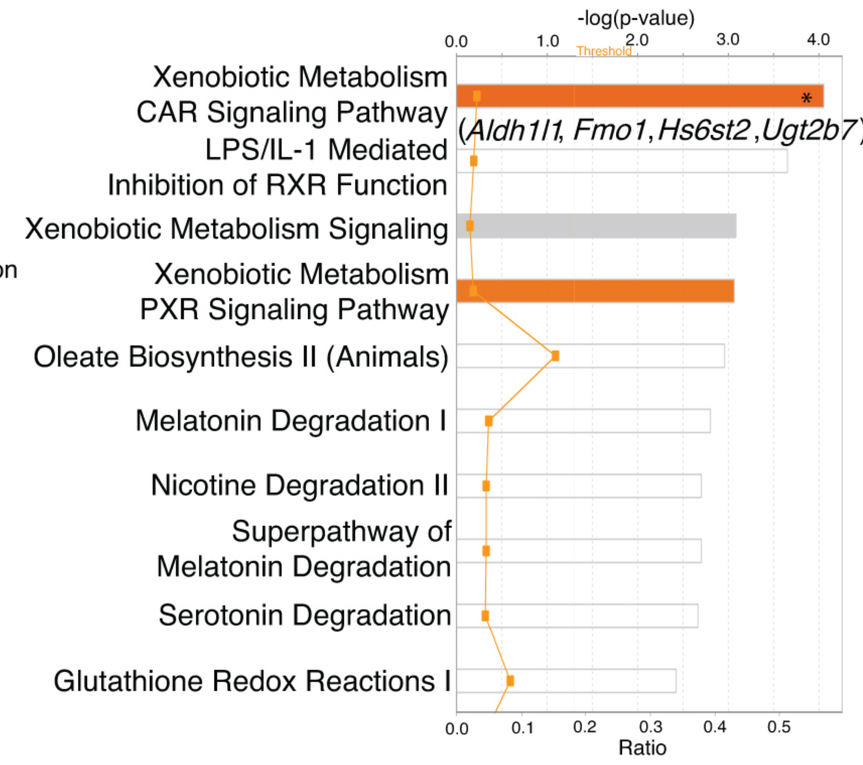

E
UMAP 2
UMAP 1
4
2
0
−2
−4 −2 0 2
RelPseudotime
1.00
0.75
0.50
0.25
0.00

A

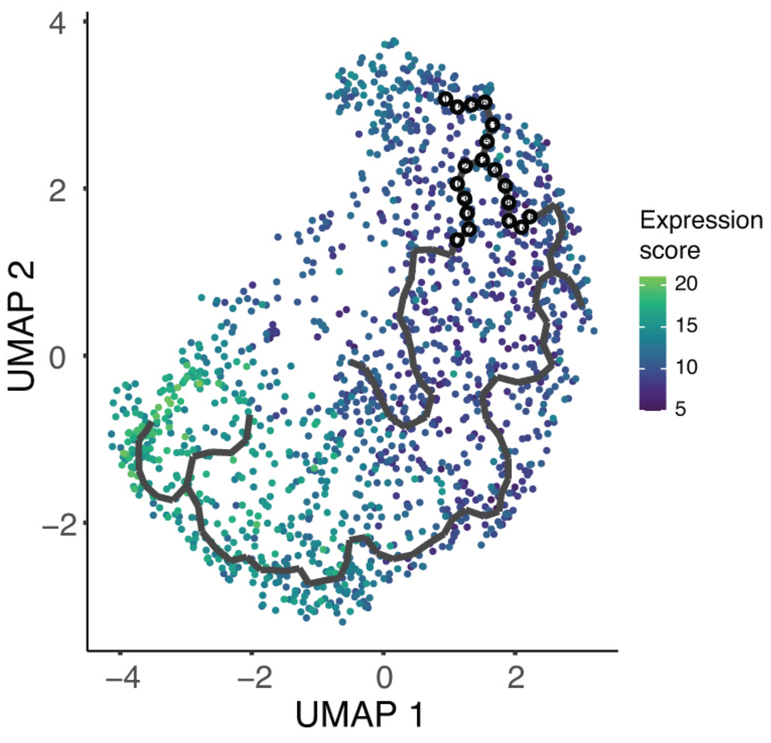

Module 30
Module 4
Module 33
Module 14
Module 34
Module 21
Module 22
Module 3
Module 18
Module 23
Module 27
Module 19
Module 38
Module 9
Module 41
Module 24
Module 36
Module 13
Module 31
Module 29
Module 32
Module 5
Module 7
Module 2
Module 20
Module 11
Module 28
Module 26
Module 40
Module 1
Module 6
Module 39
Module 12
Module 25
Module 8
Module 10
Module 15
Module 16
Module 17
Module 35
Module 37
2
1
0
−1
−2
0
2
1
4
Clusters

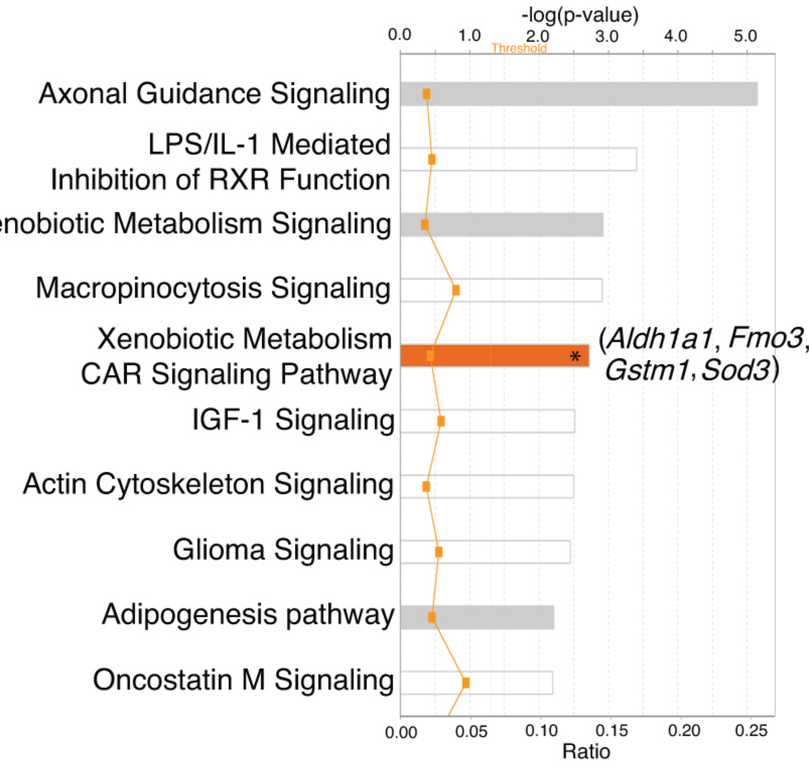
B


Module 4
UMAP 2
UMAP 1
Expression score
30
25
20
15
10

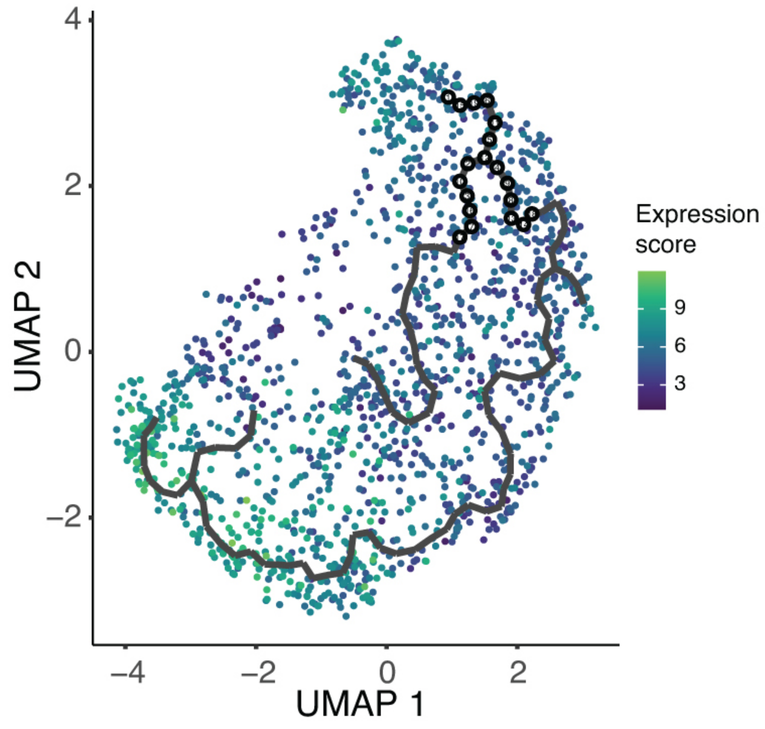

M4, Cluster 2 vs 1
-log(p-value)
Threshold
Xenobiotic Metabolism CAR Signaling Pathway
(Aldh1l1, Fmo1, Hs6st2, Ugt2b7)
LPS/IL-1 Mediated Inhibition of RXR Function
Xenobiotic Metabolism Signaling
Xenobiotic Metabolism PXR Signaling Pathway
Oleate Biosynthesis II (Animals)
Melatonin Degradation I
Nicotine Degradation II
Superpathway of Melatonin Degradation
Serotonin Degradation
Glutathione Redox Reactions I
Ratio

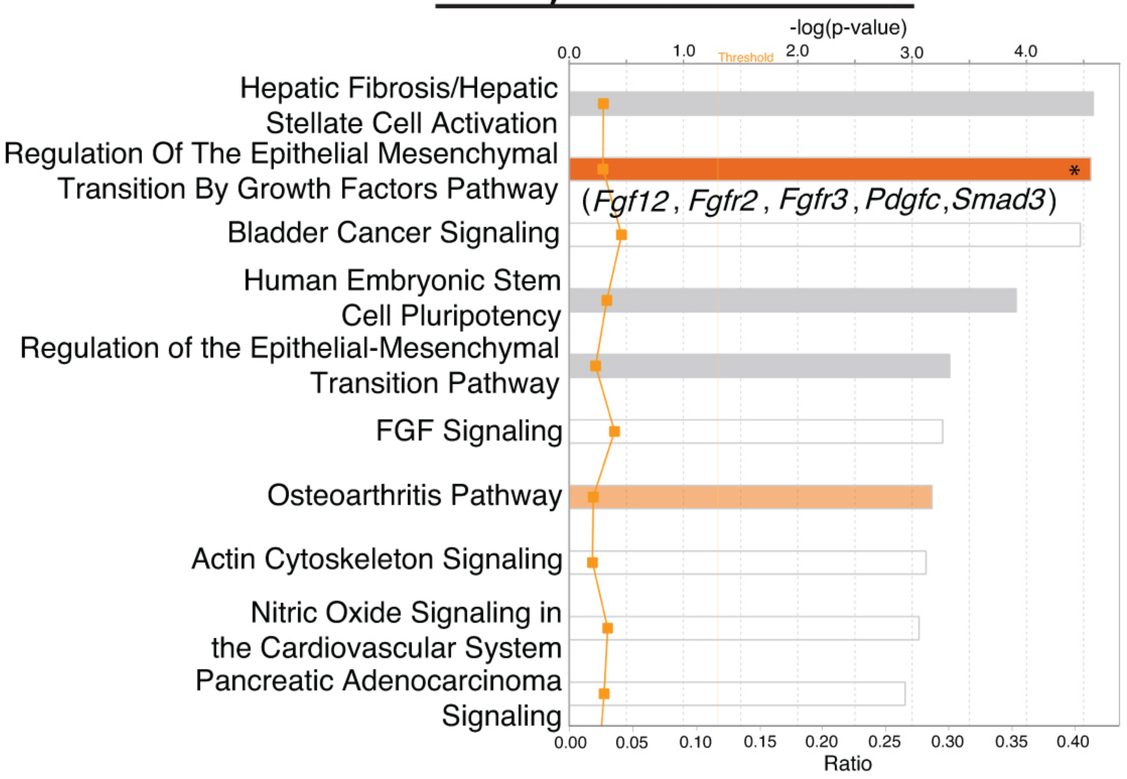
C


Module 14
UMAP 2
UMAP 1
Expression score
20
15
10
5

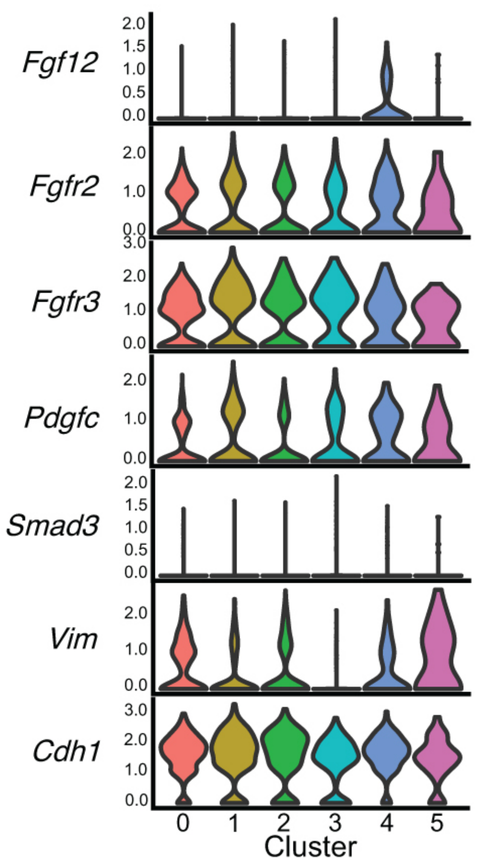

M14, Cluster 1 vs 0
-log(p-value)
Threshold
Axonal Guidance Signaling
LPS/IL-1 Mediated Inhibition of RXR Function
Xenobiotic Metabolism Signaling
Macropinocytosis Signaling
Xenobiotic Metabolism CAR Signaling Pathway
(Aldh1a1, Fmo3, Gstm1, Sod3)
IGF-1 Signaling
Actin Cytoskeleton Signaling
Glioma Signaling
Adipogenesis pathway
Oncostatin M Signaling
Ratio

D

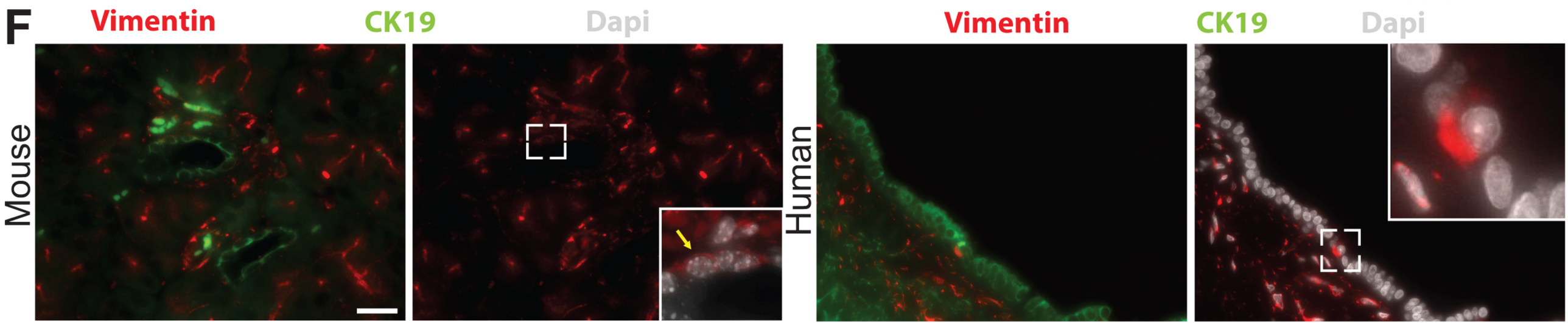

Module 34
UMAP 2
UMAP 1
Expression score
9
6
3


M34, Cluster 1 vs 0
-log(p-value)
Threshold
Hepatic Fibrosis/Hepatic Stellate Cell Activation
Regulation Of The Epithelial Mesenchymal Transition By Growth Factors Pathway
(Fgf12, Fgfr2, Fgfr3, Pdgfc, Smad3)
Bladder Cancer Signaling
Human Embryonic Stem Cell Pluripotency
Regulation of the Epithelial-Mesenchymal Transition Pathway
FGF Signaling
Osteoarthritis Pathway
Actin Cytoskeleton Signaling
Nitric Oxide Signaling in the Cardiovascular System
Pancreatic Adenocarcinoma Signaling
Ratio

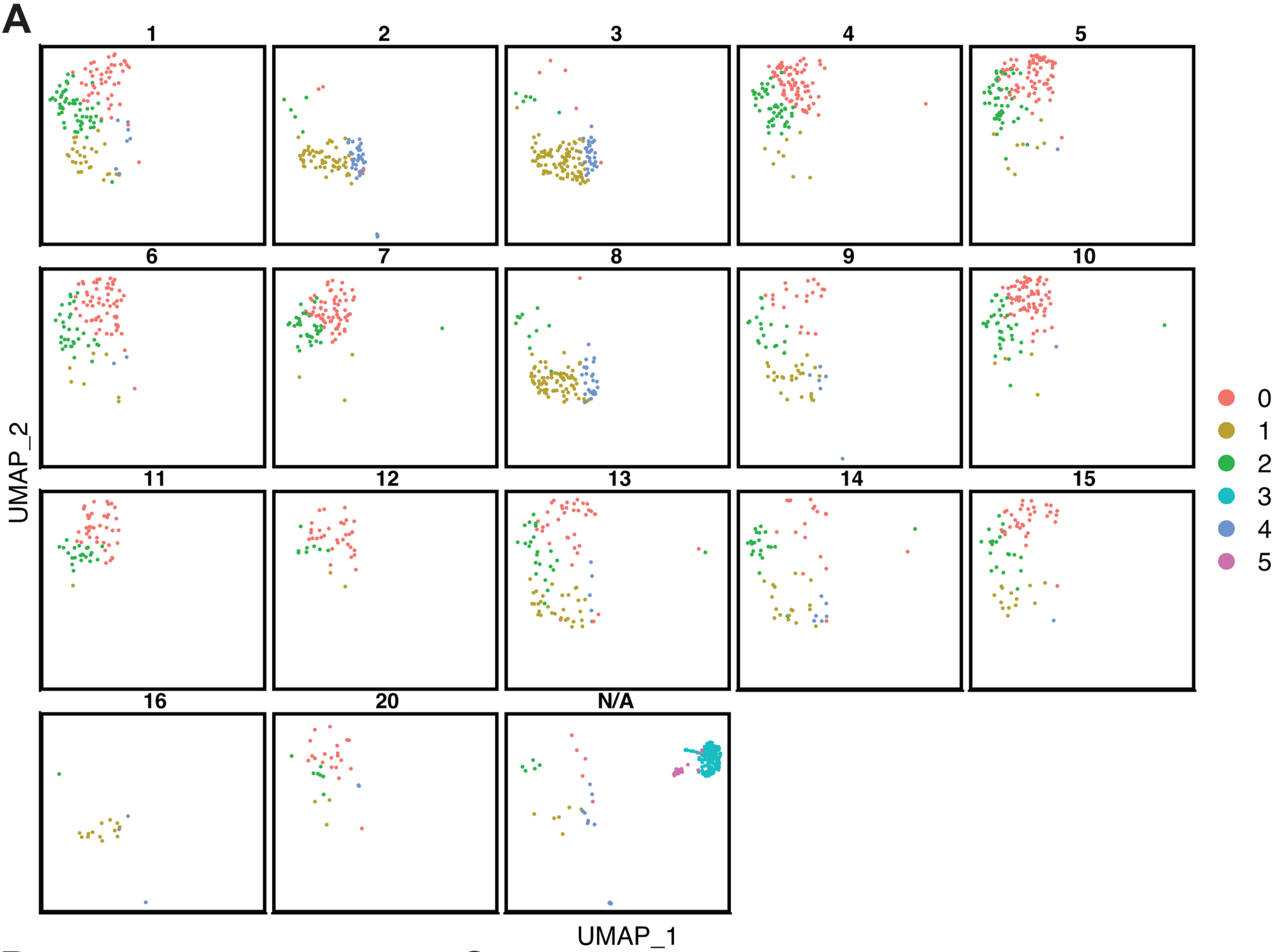
E


Fgf12
Fgfr2
Fgfr3
Pdgfc
Smad3
Vim
Cdh1
Cluster

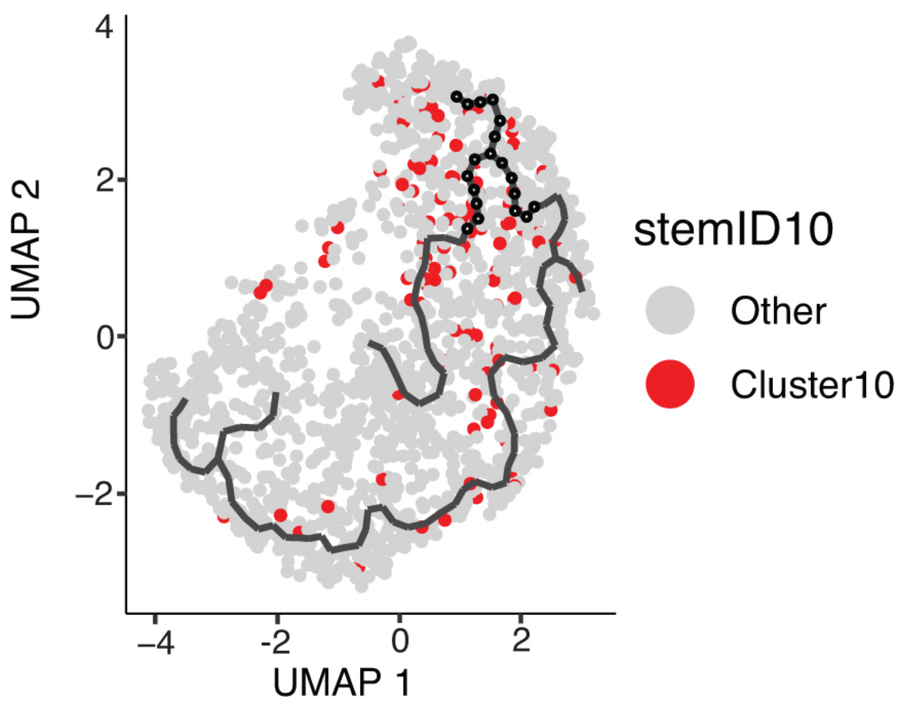
F


Vimentin
CK19
Dapi
Mouse
Vimentin
CK19
Dapi
Human

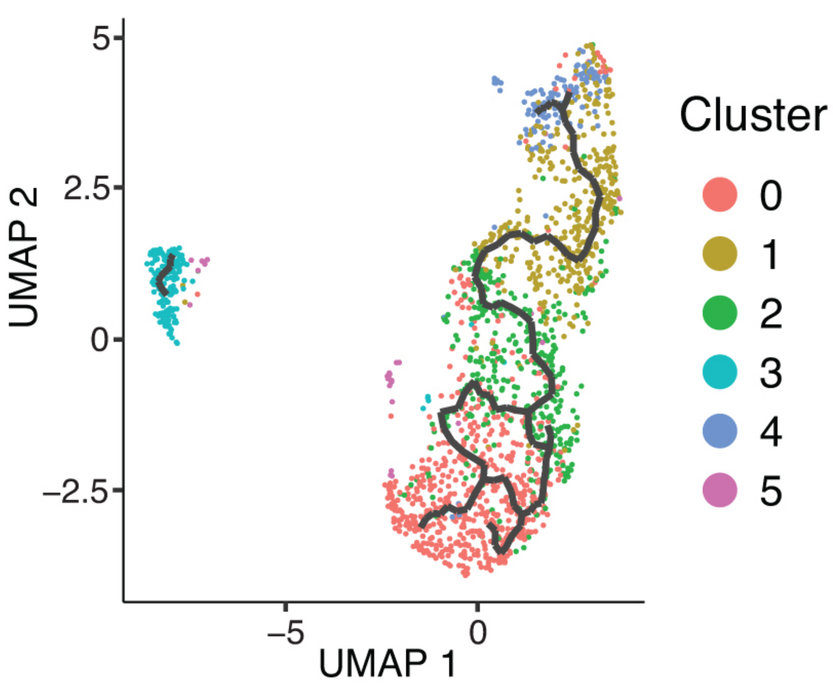
A


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
20
N/A
UMAP_2
UMAP_1
0
1
2
3
4
5

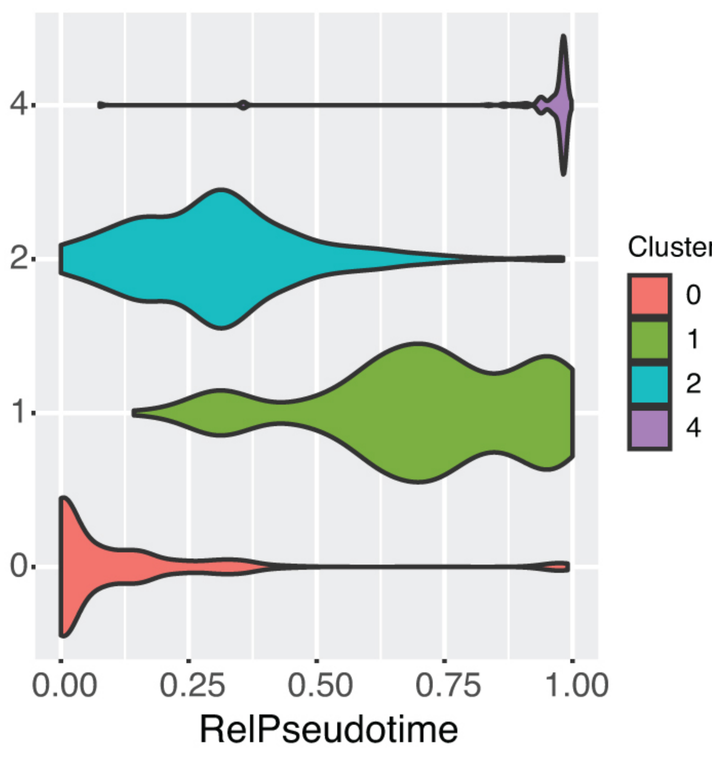
B


UMAP 2
UMAP 1
stemID10
Other
Cluster10

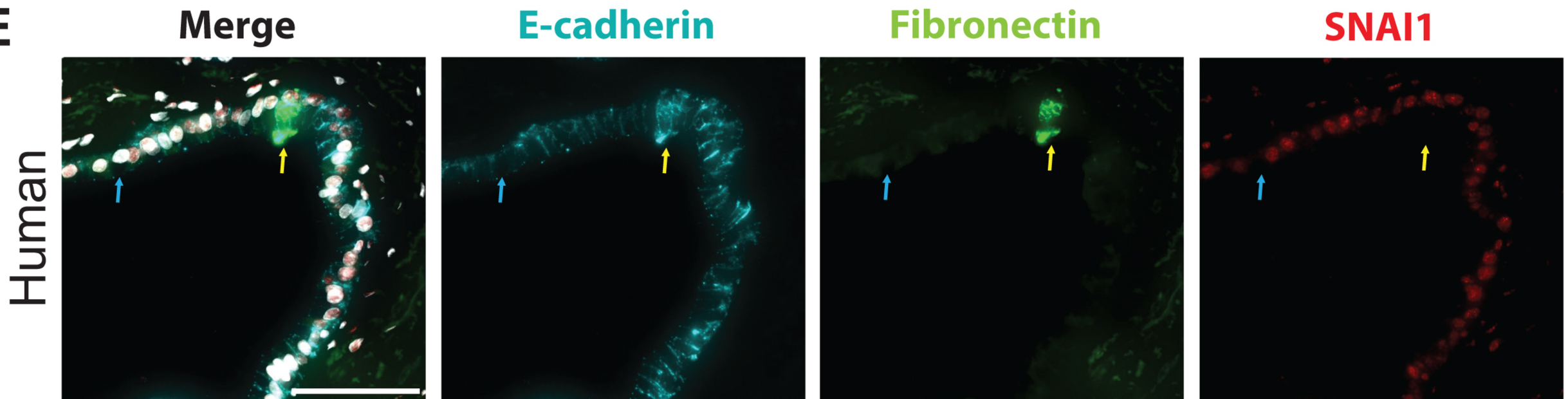
C

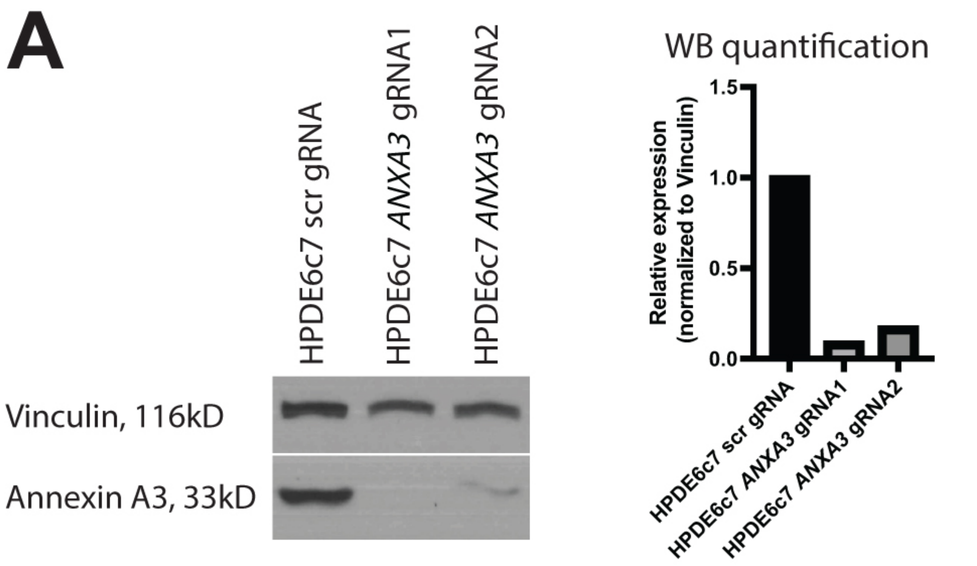

UMAP 2
UMAP 1
Cluster
0
1
2
3
4
5

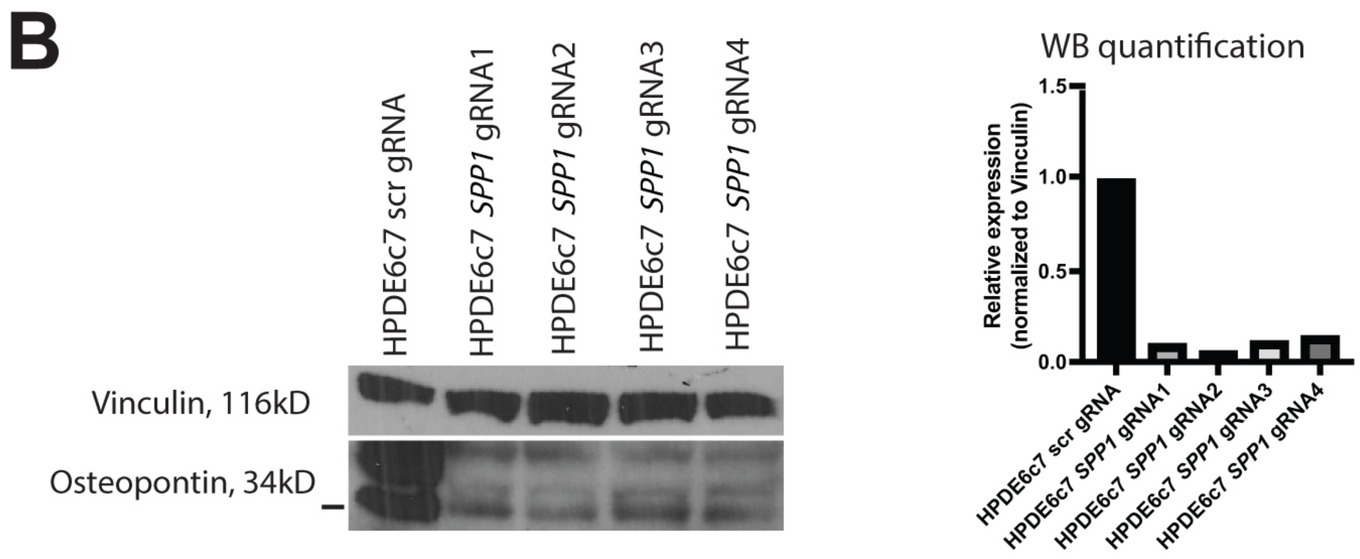
D

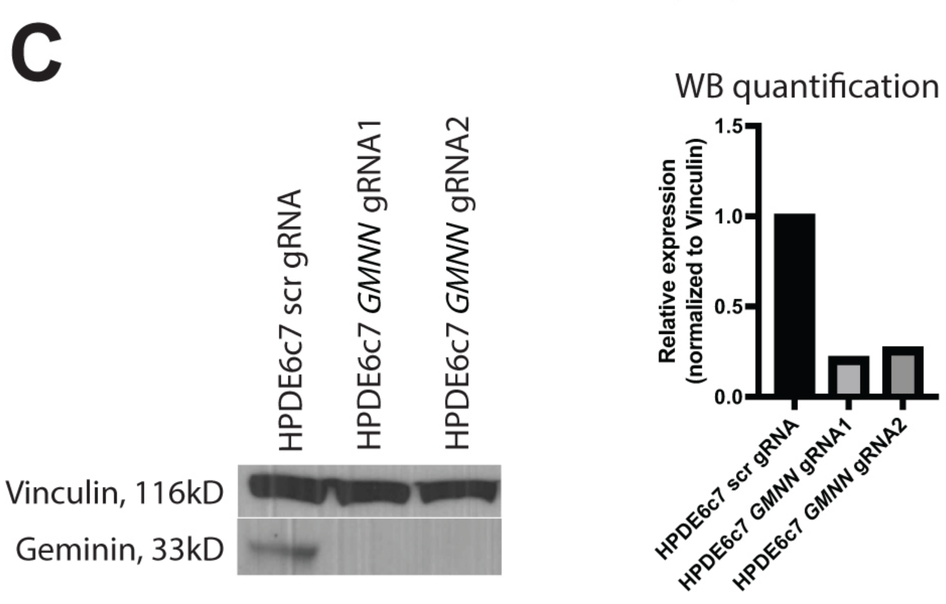

RelPseudotime
Cluster
0
1
2
4

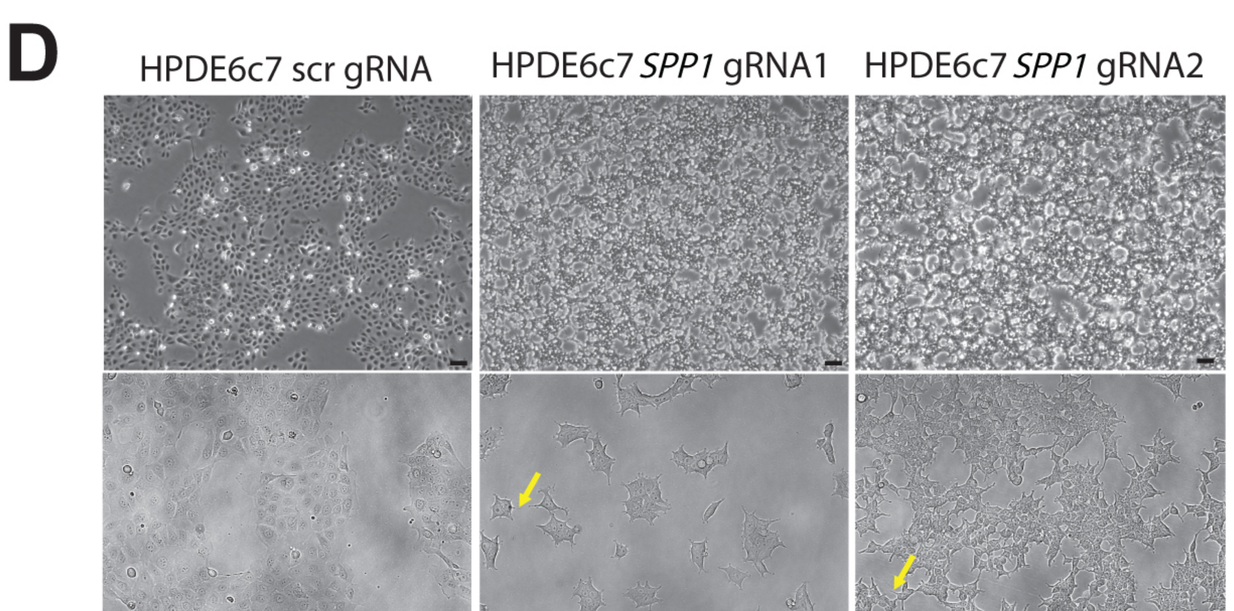
E

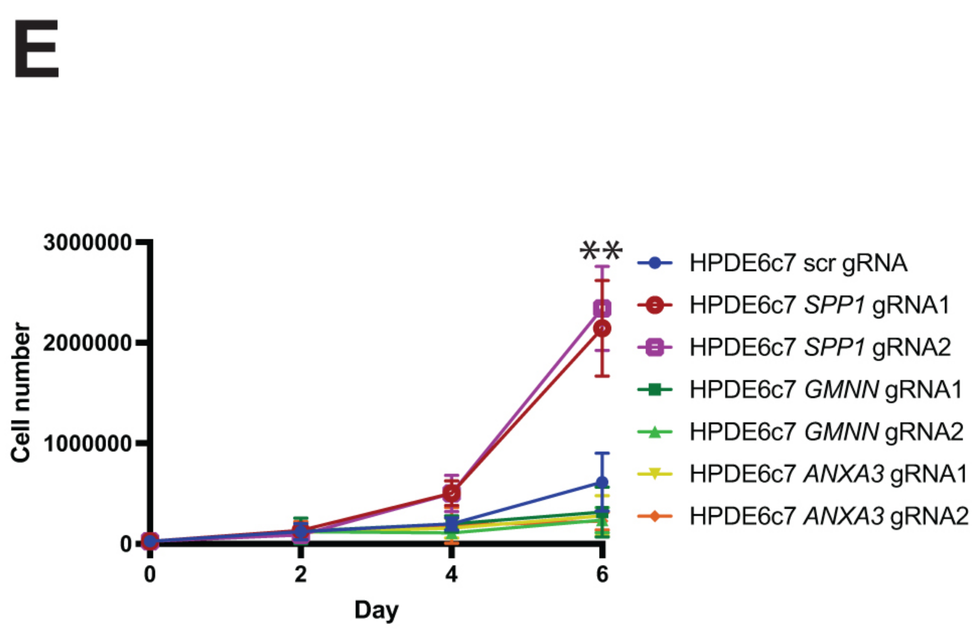

Merge
E-cadherin
Fibronectin
SNAI1
Human

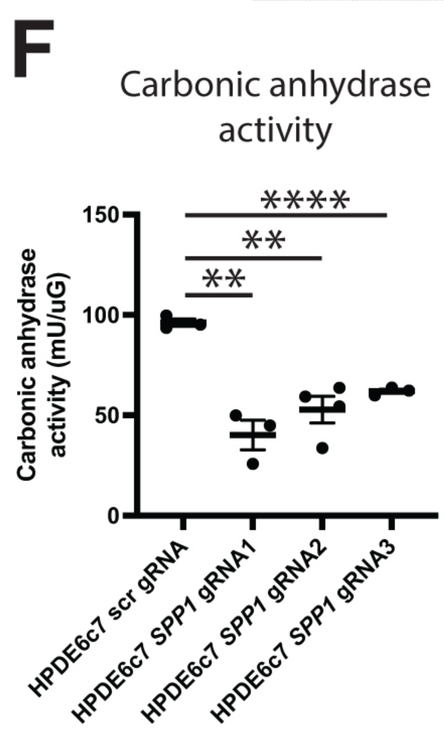

A
HPDE6c7 scr gRNA
HPDE6c7 ANXA3 gRNA1
HPDE6c7 ANXA3 gRNA2
Vinculin, 116kD
Annexin A3, 33kD
WB quantification
Relative expression (normalized to Vinculin)
1.5
1.0
0.5
0.0
HPDE6c7 scr gRNA
HPDE6c7 ANXA3 gRNA1
HPDE6c7 ANXA3 gRNA2

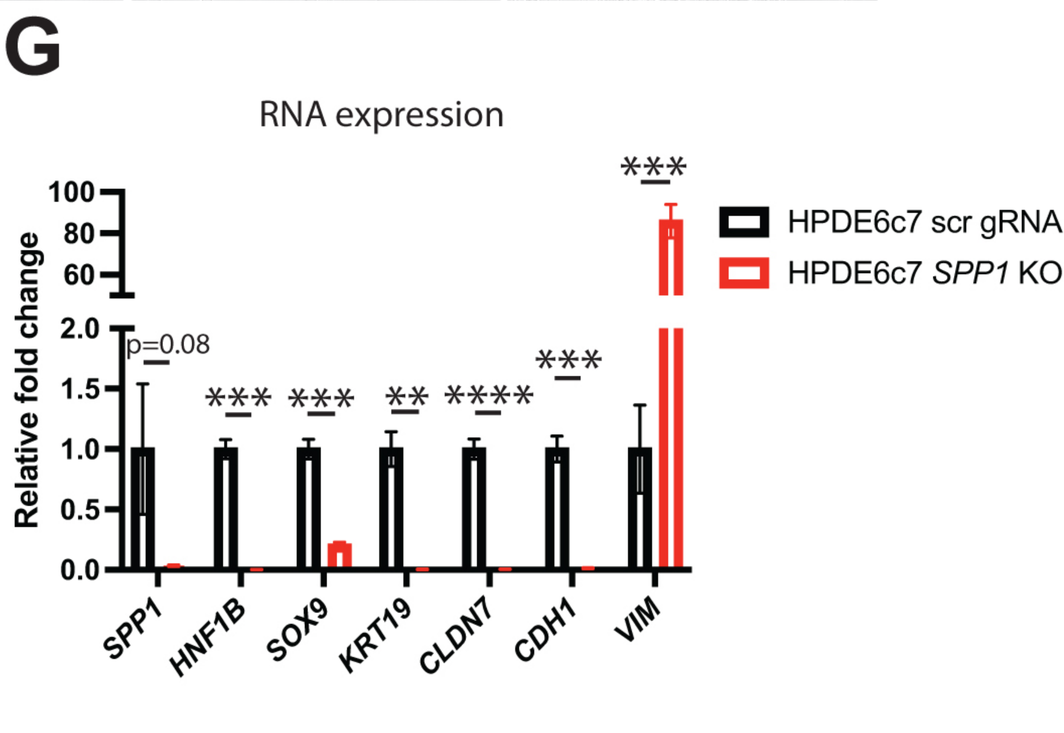

B
HPDE6c7 scr gRNA
HPDE6c7 SPP1 gRNA1
HPDE6c7 SPP1 gRNA2
HPDE6c7 SPP1 gRNA3
HPDE6c7 SPP1 gRNA4
Vinculin, 116kD
Osteopontin, 34kD
WB quantification
Relative expression (normalized to Vinculin)
1.5
1.0
0.5
0.0
HPDE6c7 scr gRNA
HPDE6c7 SPP1 gRNA1
HPDE6c7 SPP1 gRNA2
HPDE6c7 SPP1 gRNA3
HPDE6c7 SPP1 gRNA4

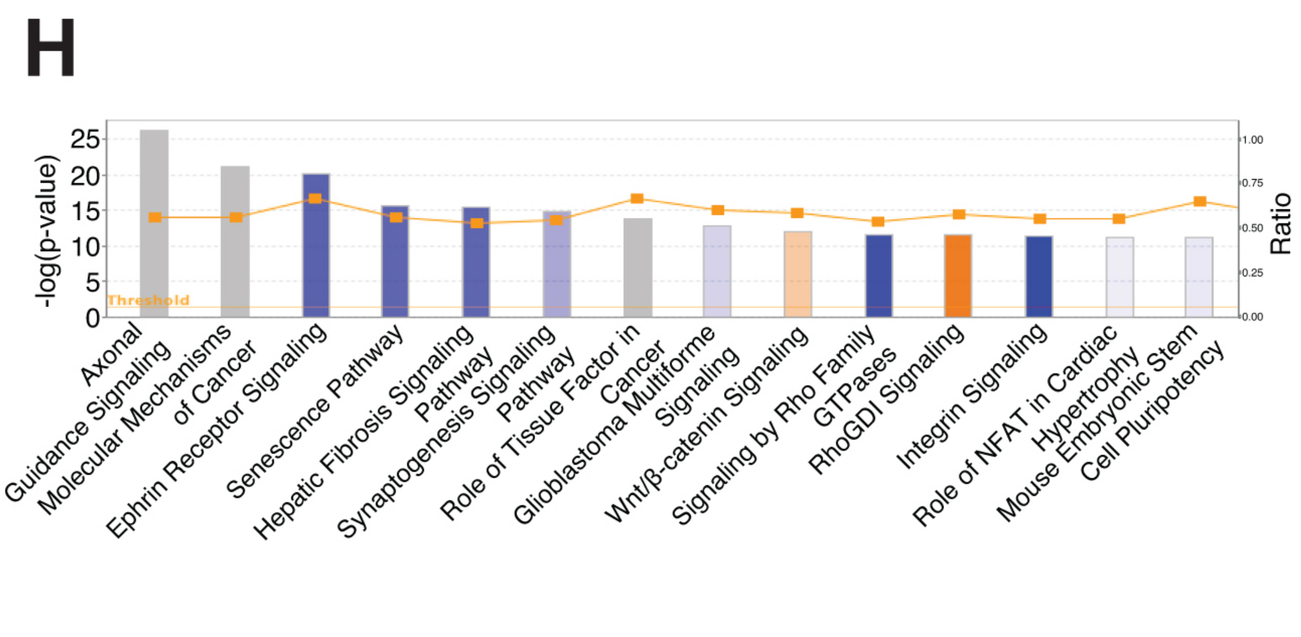

C
HPDE6c7 scr gRNA
HPDE6c7 GMNN gRNA1
HPDE6c7 GMNN gRNA2
Vinculin, 116kD
Geminin, 33kD
WB quantification
Relative expression (normalized to Vinculin)
1.5
1.0
0.5
0.0
HPDE6c7 scr gRNA
HPDE6c7 GMNN gRNA1
HPDE6c7 GMNN gRNA2

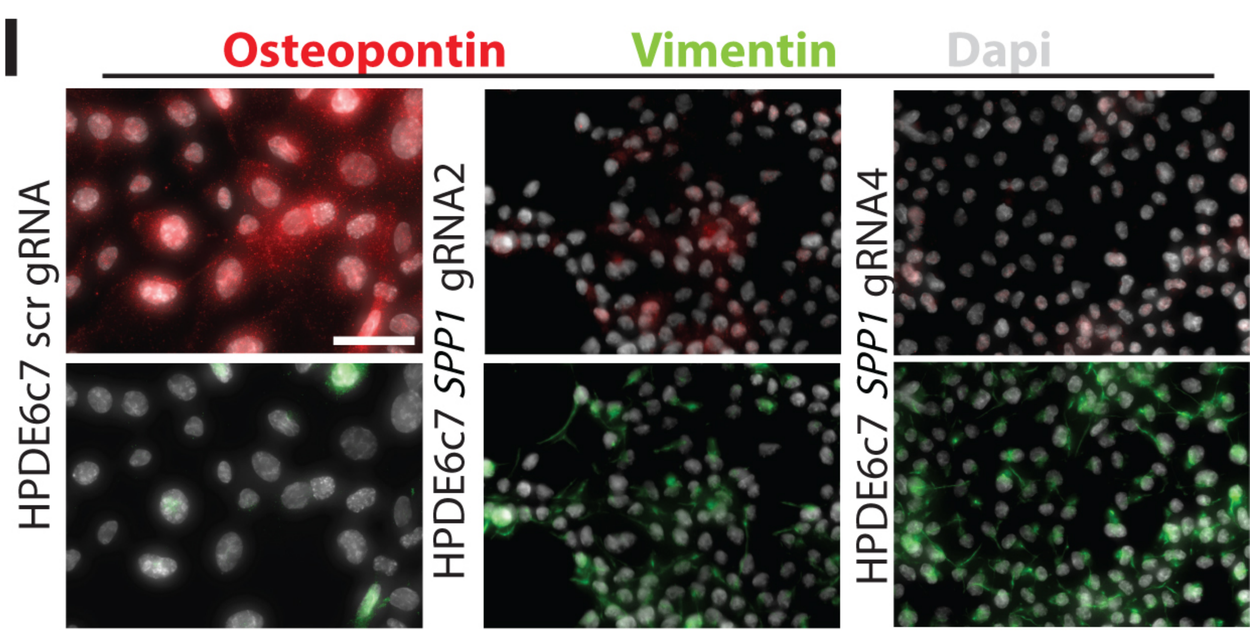

D
HPDE6c7 scr gRNA
HPDE6c7 SPP1 gRNA1
HPDE6c7 SPP1 gRNA2

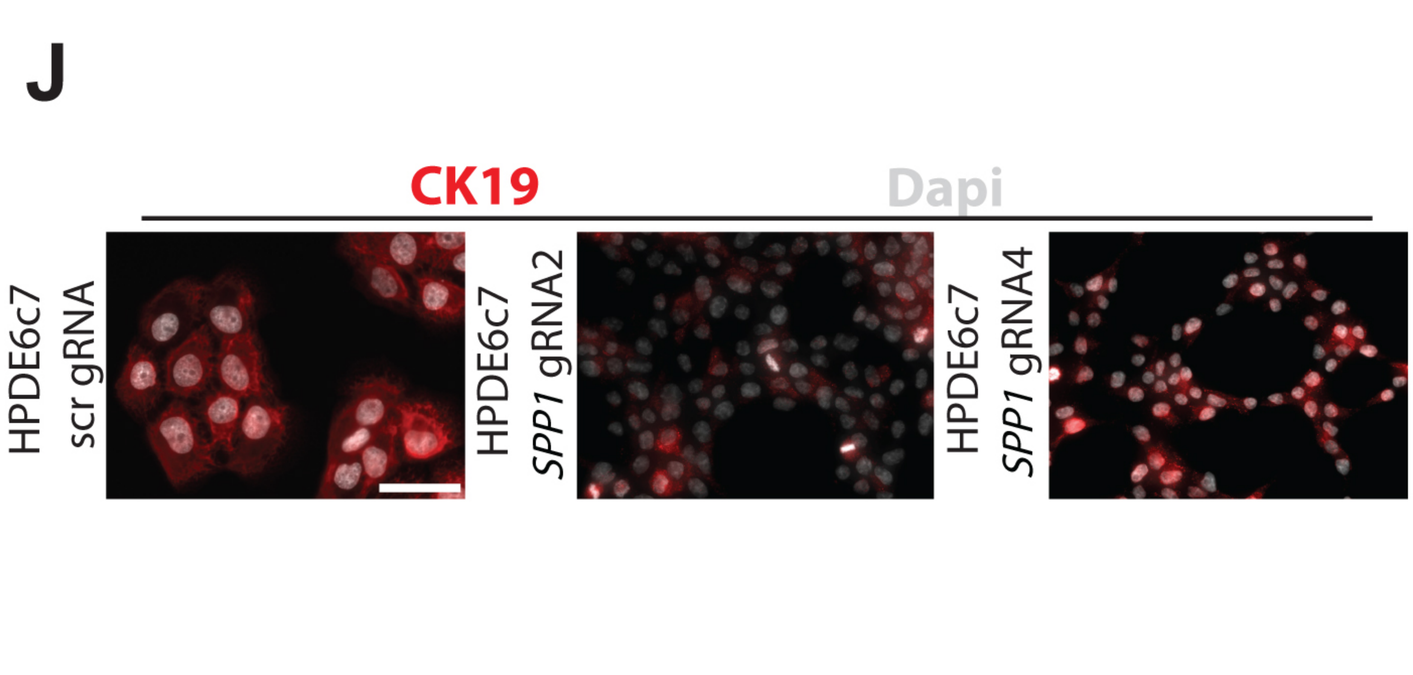

E
Cell number
3000000
2000000
1000000
0
0
2
4
6
Day
**
HPDE6c7 scr gRNA
HPDE6c7 SPP1 gRNA1
HPDE6c7 SPP1 gRNA2
HPDE6c7 GMNN gRNA1
HPDE6c7 GMNN gRNA2
HPDE6c7 ANXA3 gRNA1
HPDE6c7 ANXA3 gRNA2

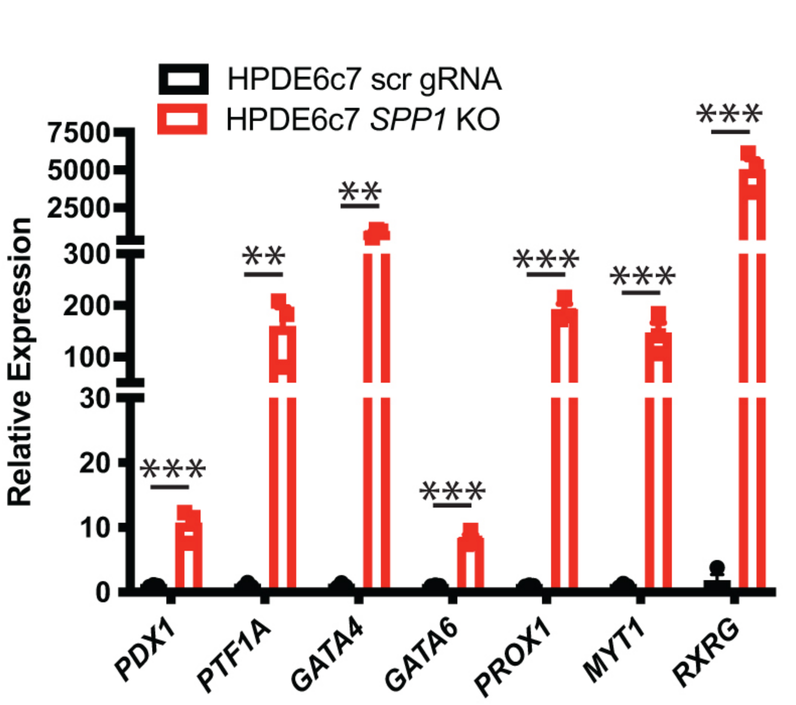

F
Carbonic anhydrase activity
Carbonic anhydrase activity (mU/uG)
150
100
50
0
****
**
**
HPDE6c7 scr gRNA
HPDE6c7 SPP1 gRNA1
HPDE6c7 SPP1 gRNA2
HPDE6c7 SPP1 gRNA3

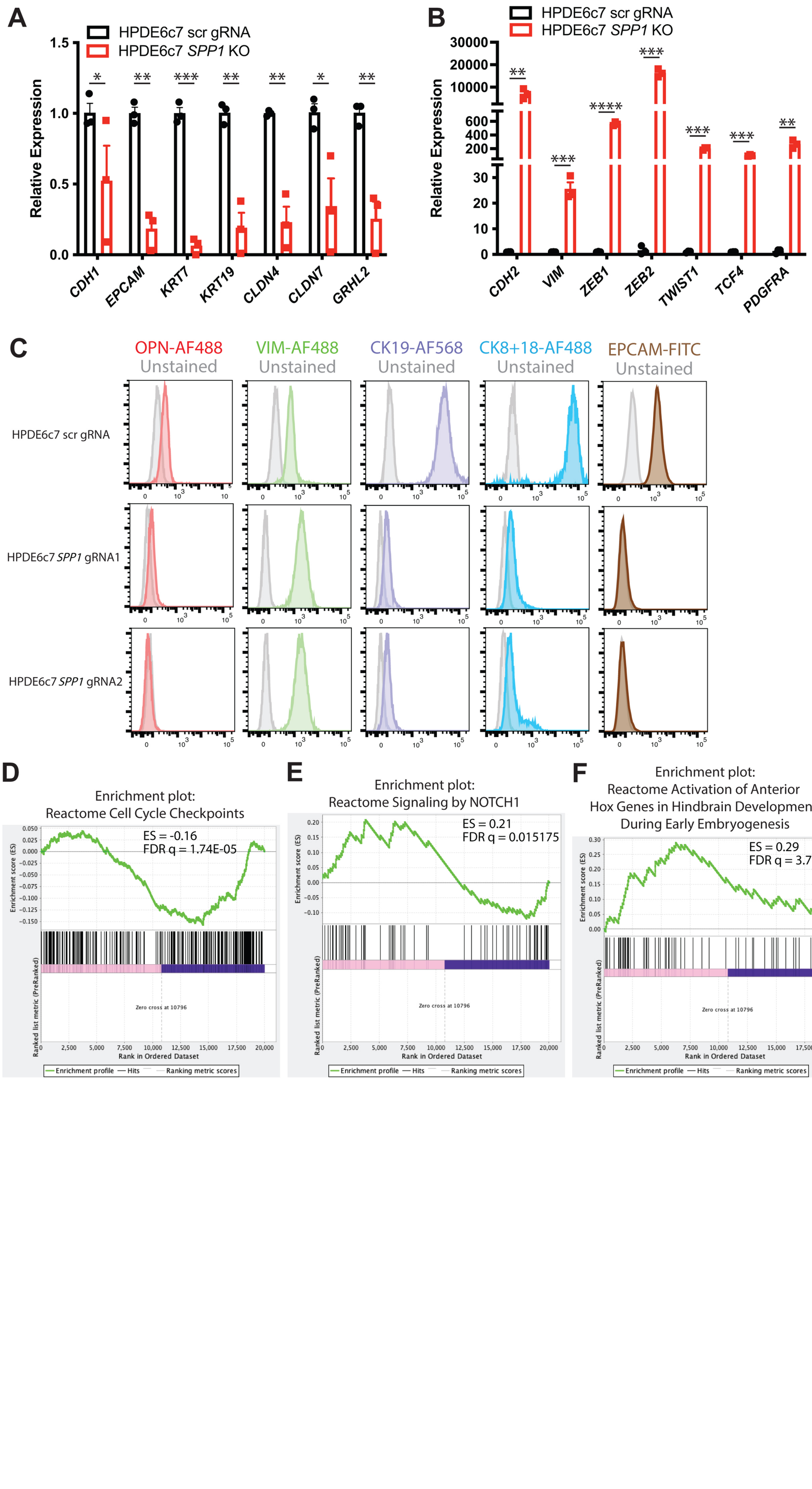

G
RNA expression
Relative fold change
100
80
60
2.0
1.5
1.0
0.5
0.0
p=0.08
***
***
**
****
***
***
SPP1
HNF1B
SOX9
KRT19
CLDN7
CDH1
VIM
HPDE6c7 scr gRNA
HPDE6c7 SPP1 KO

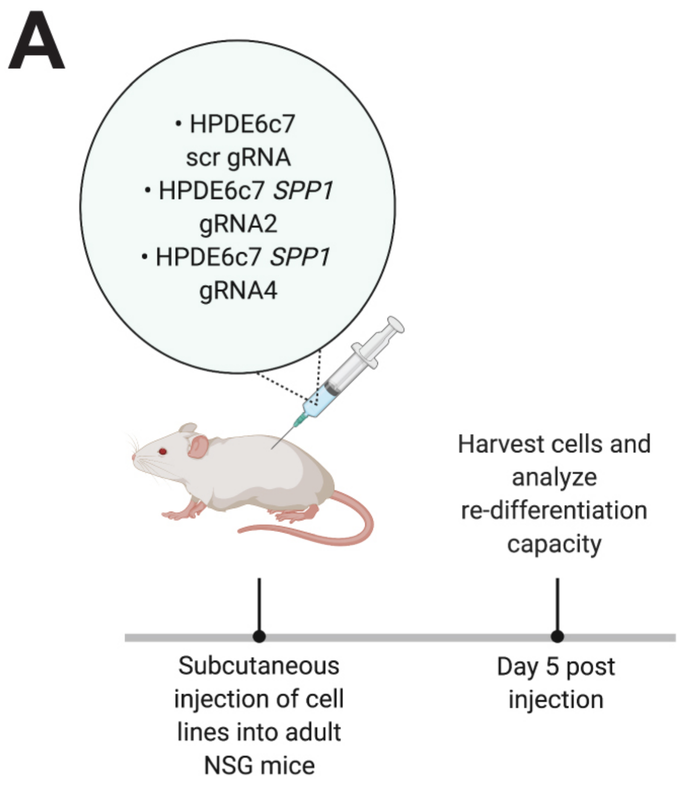

H
-log(p-value)
25
20
15
10
5
0
Threshold
Ratio
1.00
0.75
0.50
0.25
0.00
Axonal Guidance Signaling
Molecular Mechanisms of Cancer
Ephrin Receptor Signaling
Senescence Pathway
Hepatic Fibrosis Signaling Pathway
Synaptogenesis Signaling Pathway
Role of Tissue Factor in Cancer
Glioblastoma Multiforme Signaling
Wnt/β-catenin Signaling
Signaling by Rho Family GTPases
RhoGDI Signaling
Integrin Signaling
Role of NFAT in Cardiac Hypertrophy
Mouse Embryonic Stem Cell Pluripotency

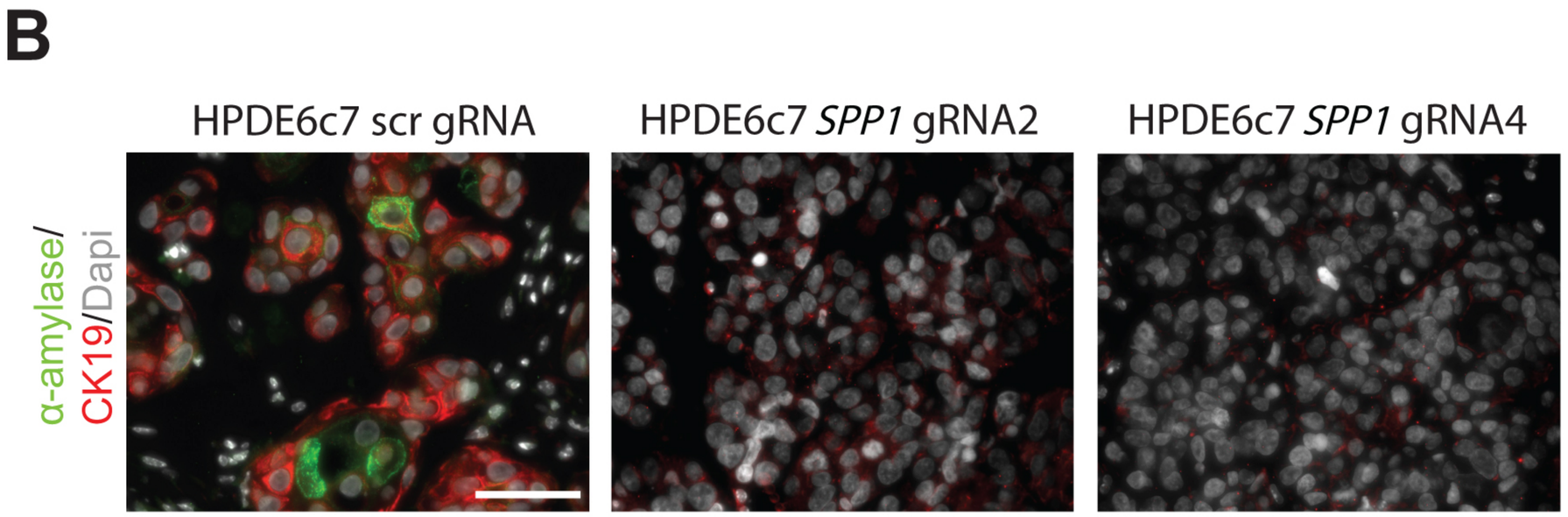

I
Osteopontin
Vimentin
Dapi
HPDE6c7 scr gRNA
HPDE6c7 SPP1 gRNA2
HPDE6c7 SPP1 gRNA4

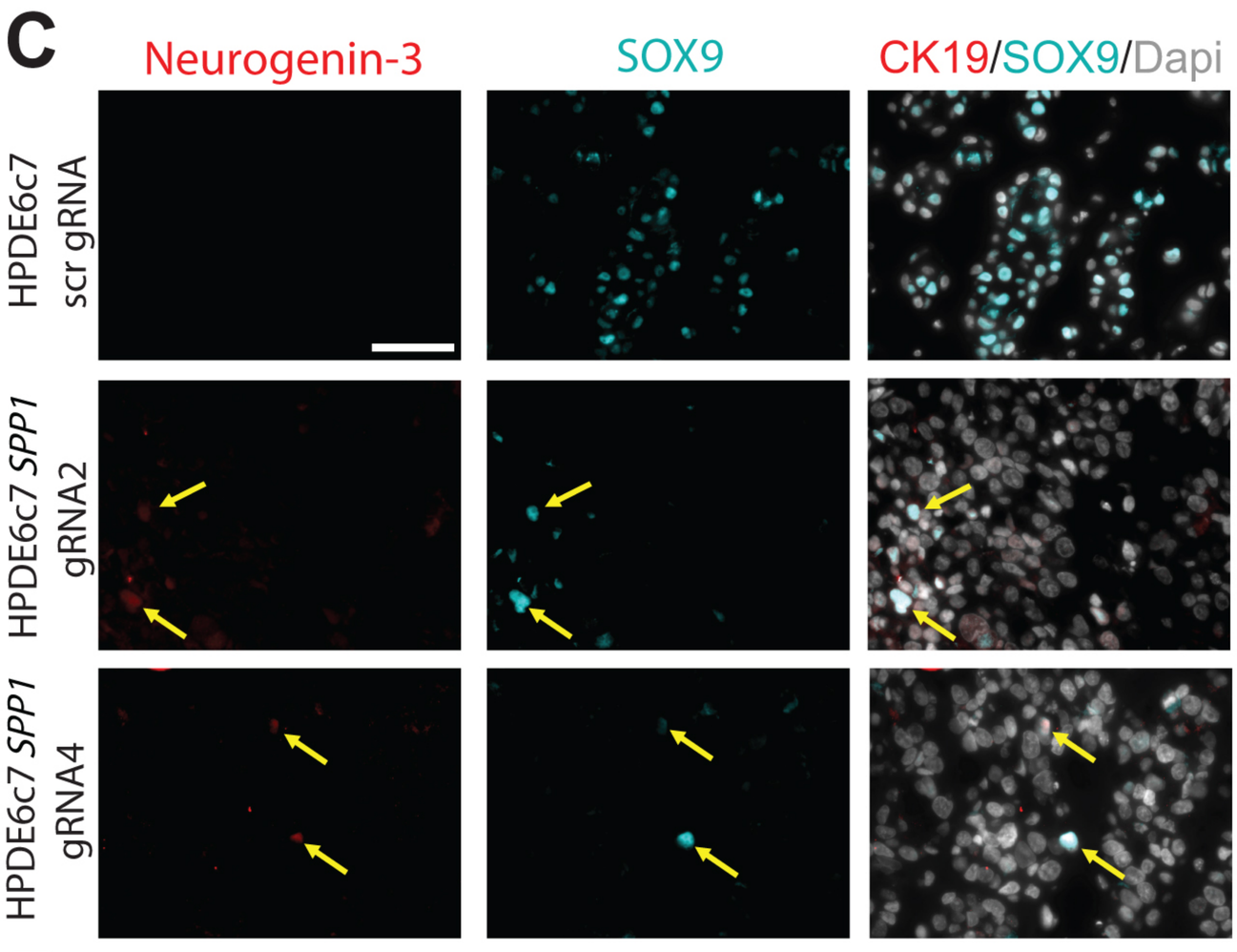

J
CK19
Dapi
HPDE6c7 scr gRNA
HPDE6c7 SPP1 gRNA2
HPDE6c7 SPP1 gRNA4

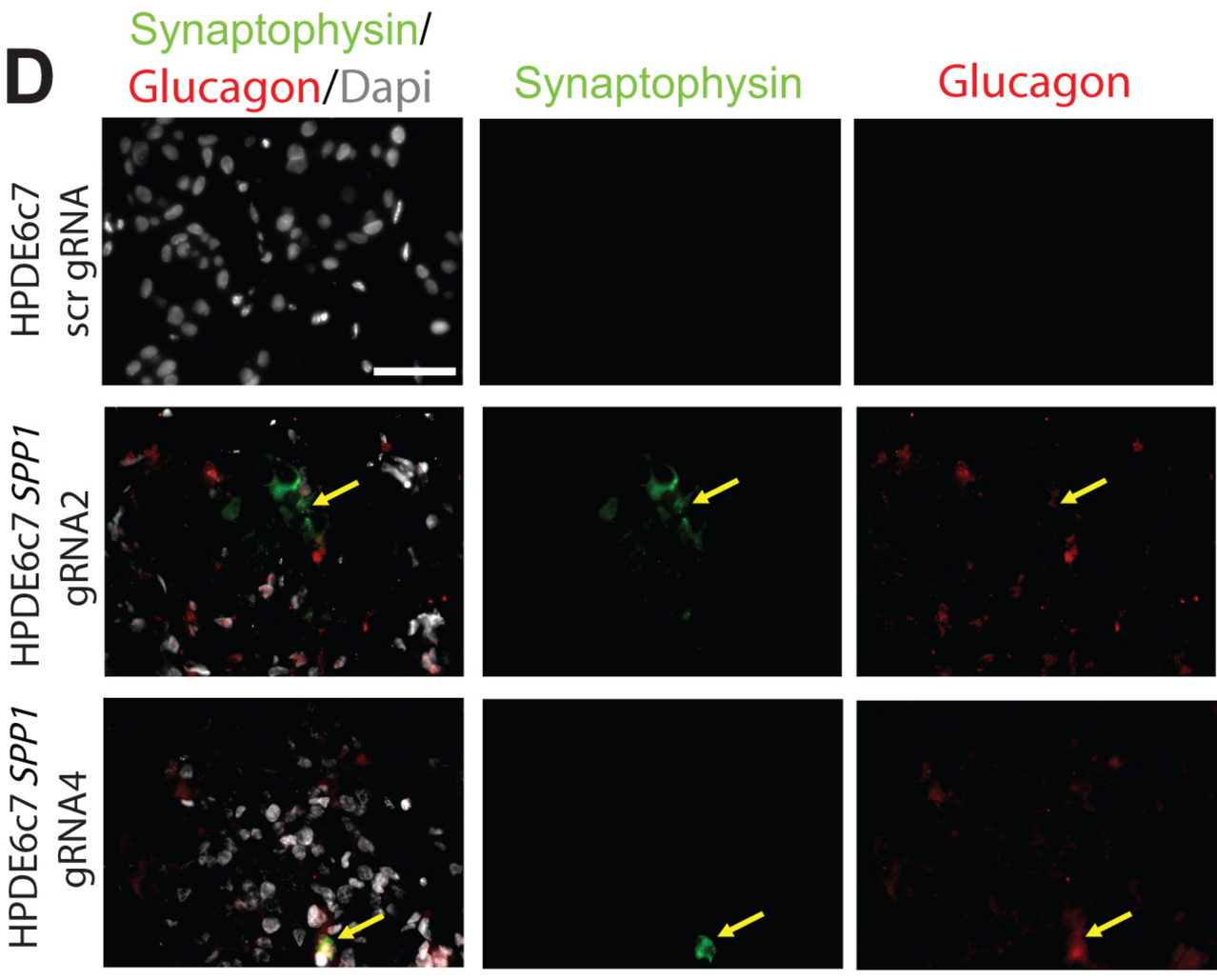

K
HPDE6c7 scr gRNA
HPDE6c7 SPP1 KO
Relative Expression
7500
5000
2500
300
200
100
30
20
10
0
***
**
**
***
***
***
PDX1
PTF1A
GATA4
GATA6
PROX1
MYT1
RXRG

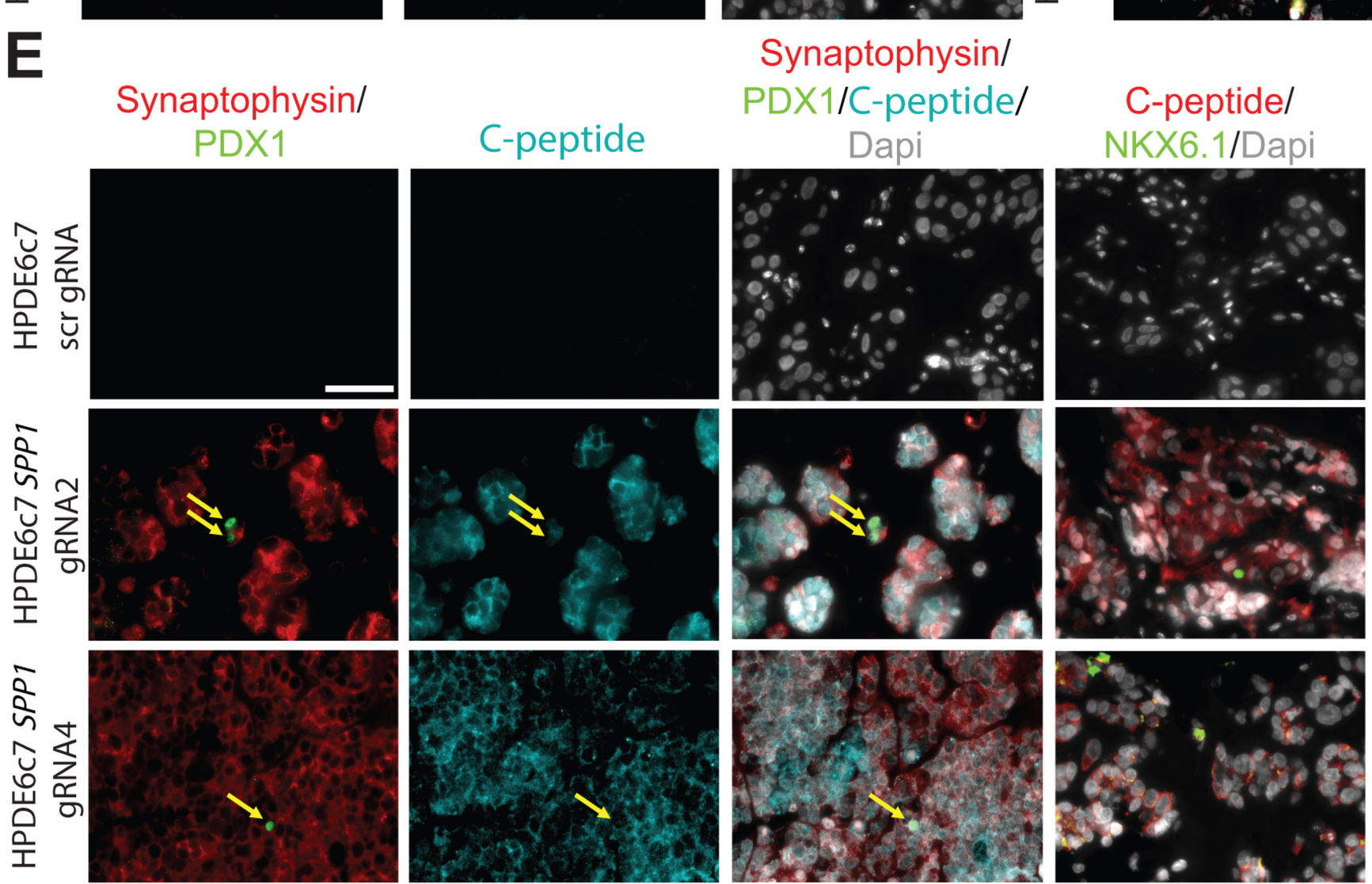

A
HPDE6c7 scr gRNA
HPDE6c7 SPP1 KO
Relative Expression
1.5
1.0
0.5
0.0
CDH1
EPCAM
KRT7
KRT19
CLDN4
CLDN7
GRHL2
B
HPDE6c7 scr gRNA
HPDE6c7 SPP1 KO
Relative Expression
30000
20000
10000
600
400
200
30
20
10
0
CDH2
VIM
ZEB1
ZEB2
TWIST1
TCF4
PDGFRA
C
OPN-AF488
Unstained
VIM-AF488
Unstained
CK19-AF568
Unstained
CK8+18-AF488
Unstained
EPCAM-FITC
Unstained
HPDE6c7 scr gRNA
HPDE6c7 SPP1 gRNA1
HPDE6c7 SPP1 gRNA2
D
Enrichment plot:
Reactome Cell Cycle Checkpoints
ES = -0.16
FDR q = 1.74E-05
Enrichment score (ES)
Ranked list metric (PreRanked)
Zero cross at 10796
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
E
Enrichment plot:
Reactome Signaling by NOTCH1
ES = 0.21
FDR q = 0.015175
Enrichment score (ES)
Ranked list metric (PreRanked)
Zero cross at 10796
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
F
Enrichment plot:
Reactome Activation of Anterior
Hox Genes in Hindbrain Development
During Early Embryogenesis
ES = 0.29
FDR q = 3.76E-04
Enrichment score (ES)
Ranked list metric (PreRanked)
Zero cross at 10796
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

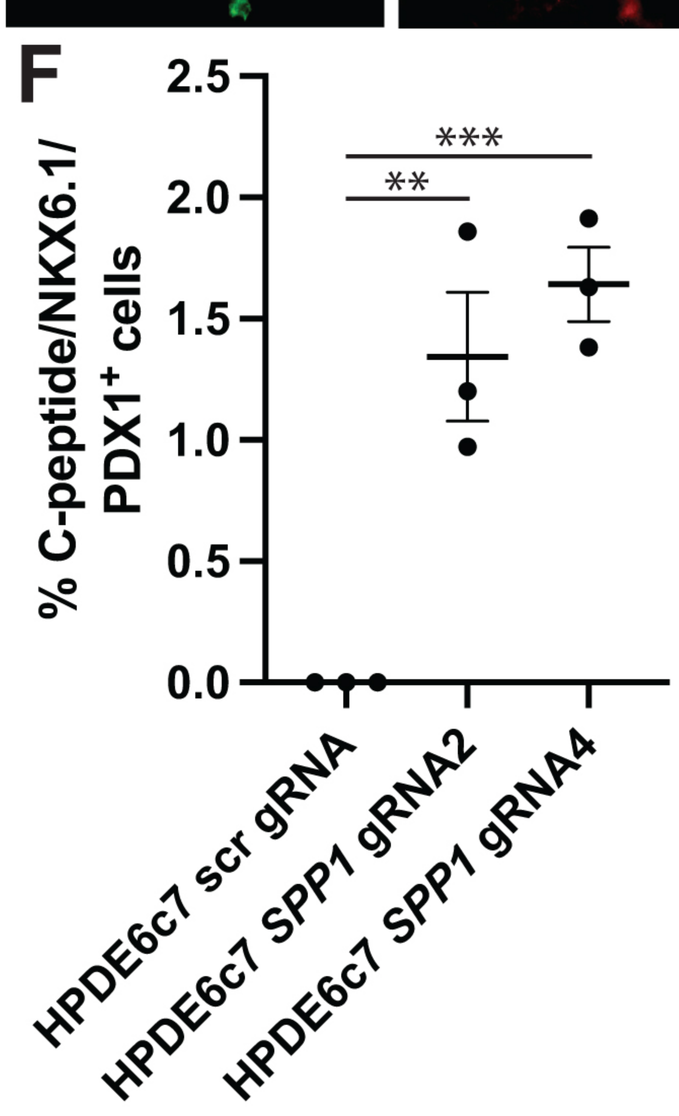

A
• HPDE6c7 scr gRNA
• HPDE6c7 SPP1 gRNA2
• HPDE6c7 SPP1 gRNA4
Harvest cells and analyze re-differentiation capacity
Subcutaneous injection of cell lines into adult NSG mice
Day 5 post injection

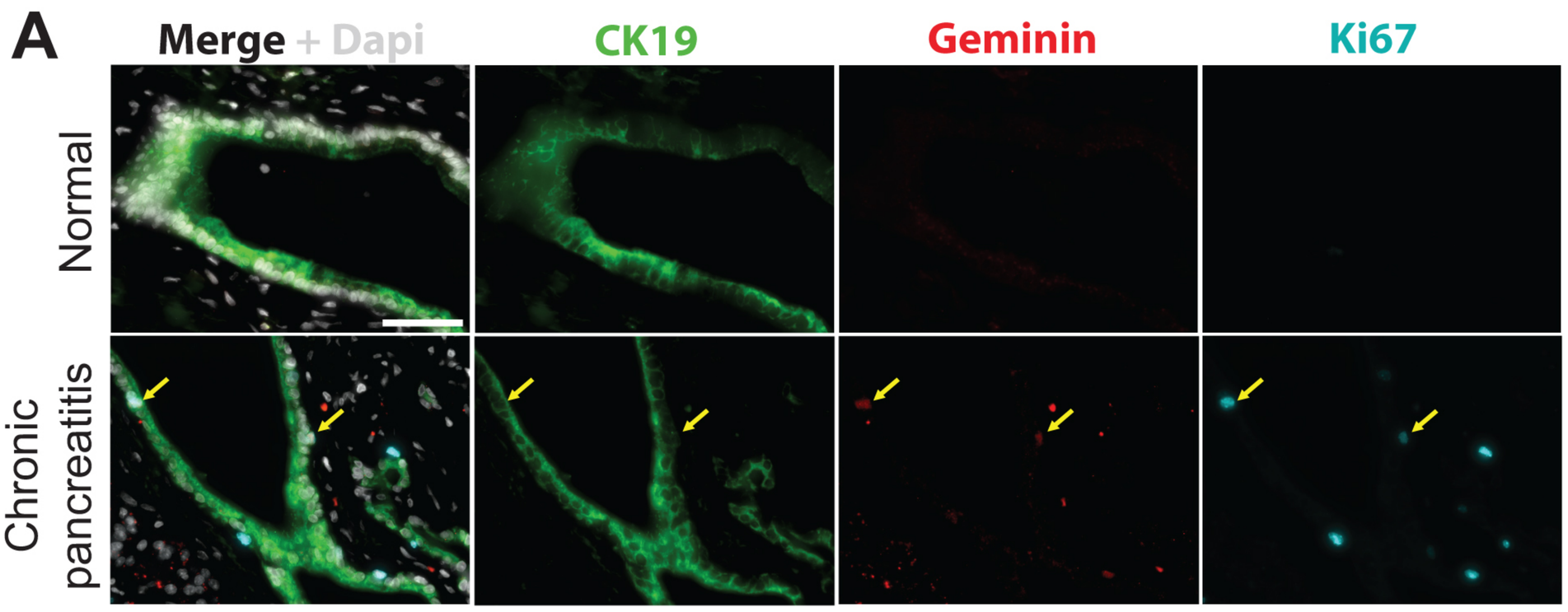

B
HPDE6c7 scr gRNA
HPDE6c7 SPP1 gRNA2
HPDE6c7 SPP1 gRNA4
α-amylase/ CK19/Dapi

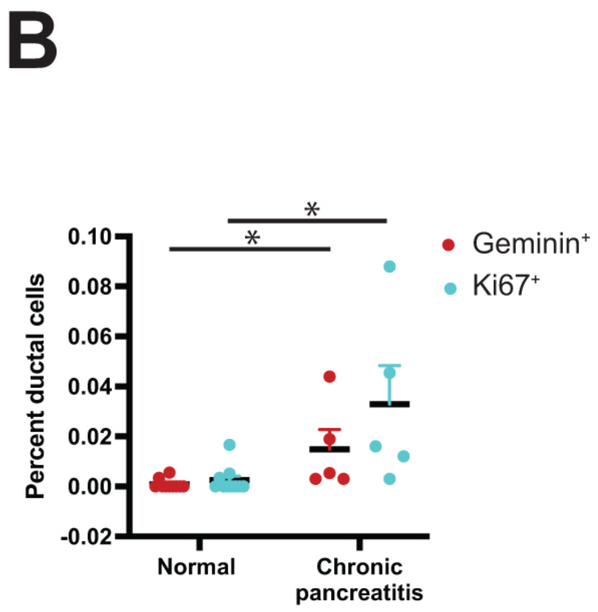

C
Neurogenin-3
SOX9
CK19/SOX9/Dapi
HPDE6c7 scr gRNA
HPDE6c7 SPP1 gRNA2
HPDE6c7 SPP1 gRNA4

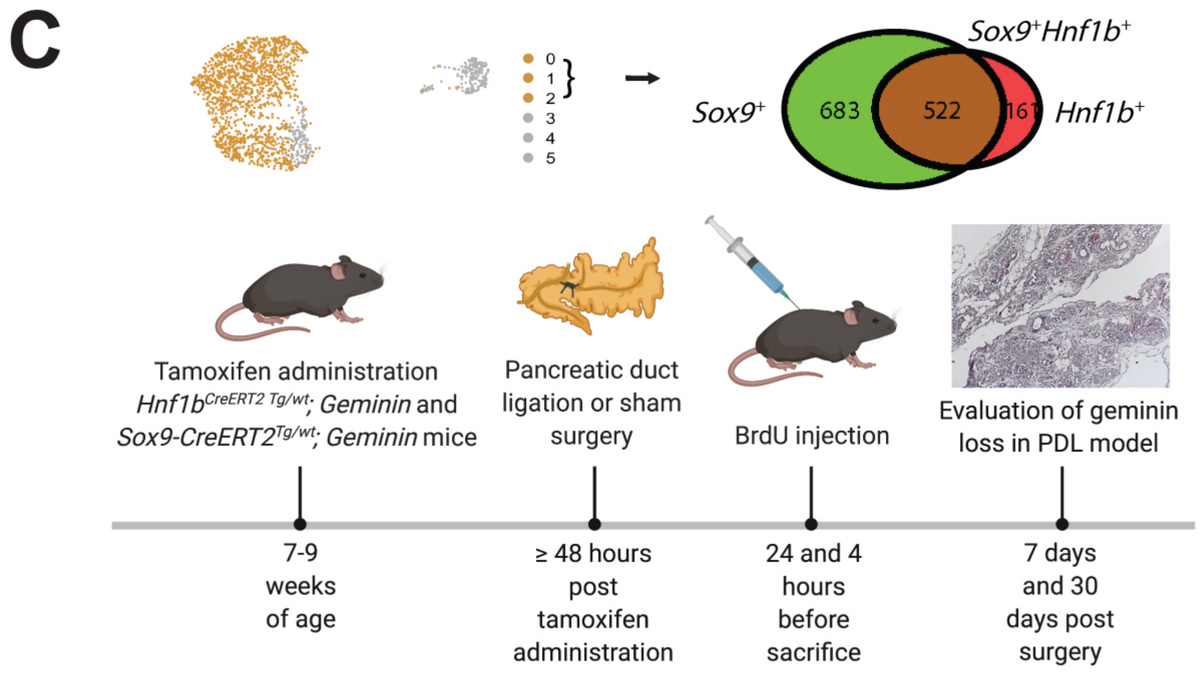

D
Synaptophysin/ Glucagon/Dapi
Synaptophysin
Glucagon
HPDE6c7 scr gRNA
HPDE6c7 SPP1 gRNA2
HPDE6c7 SPP1 gRNA4

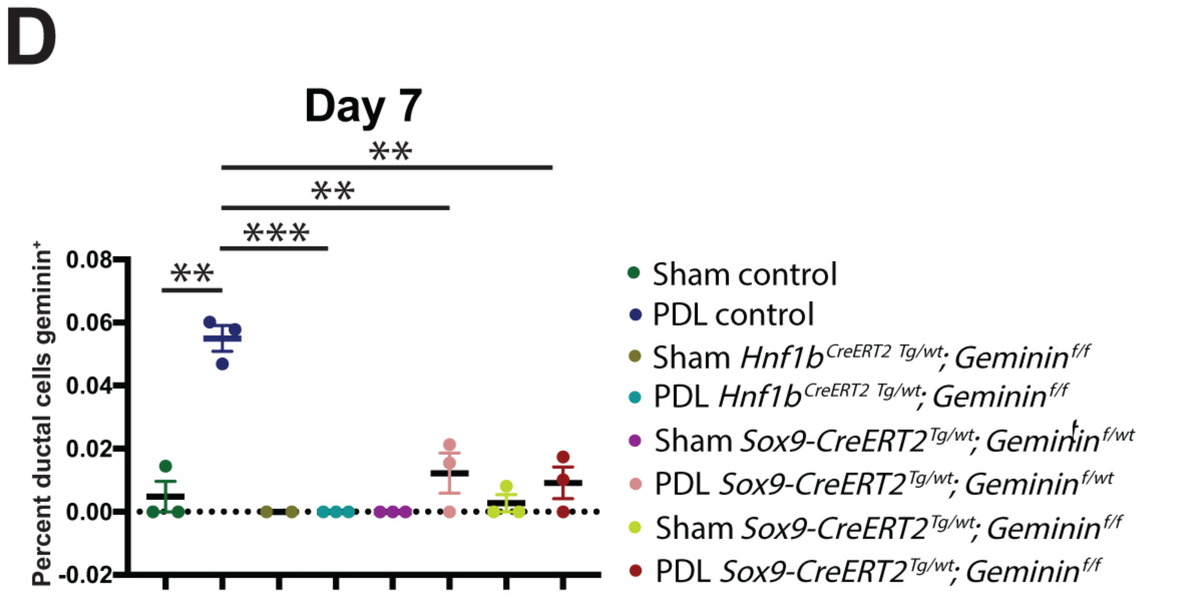

E
Synaptophysin/ PDX1
C-peptide
Synaptophysin/ PDX1/C-peptide/ Dapi
C-peptide/ NKX6.1/Dapi
HPDE6c7 scr gRNA
HPDE6c7 SPP1 gRNA2
HPDE6c7 SPP1 gRNA4

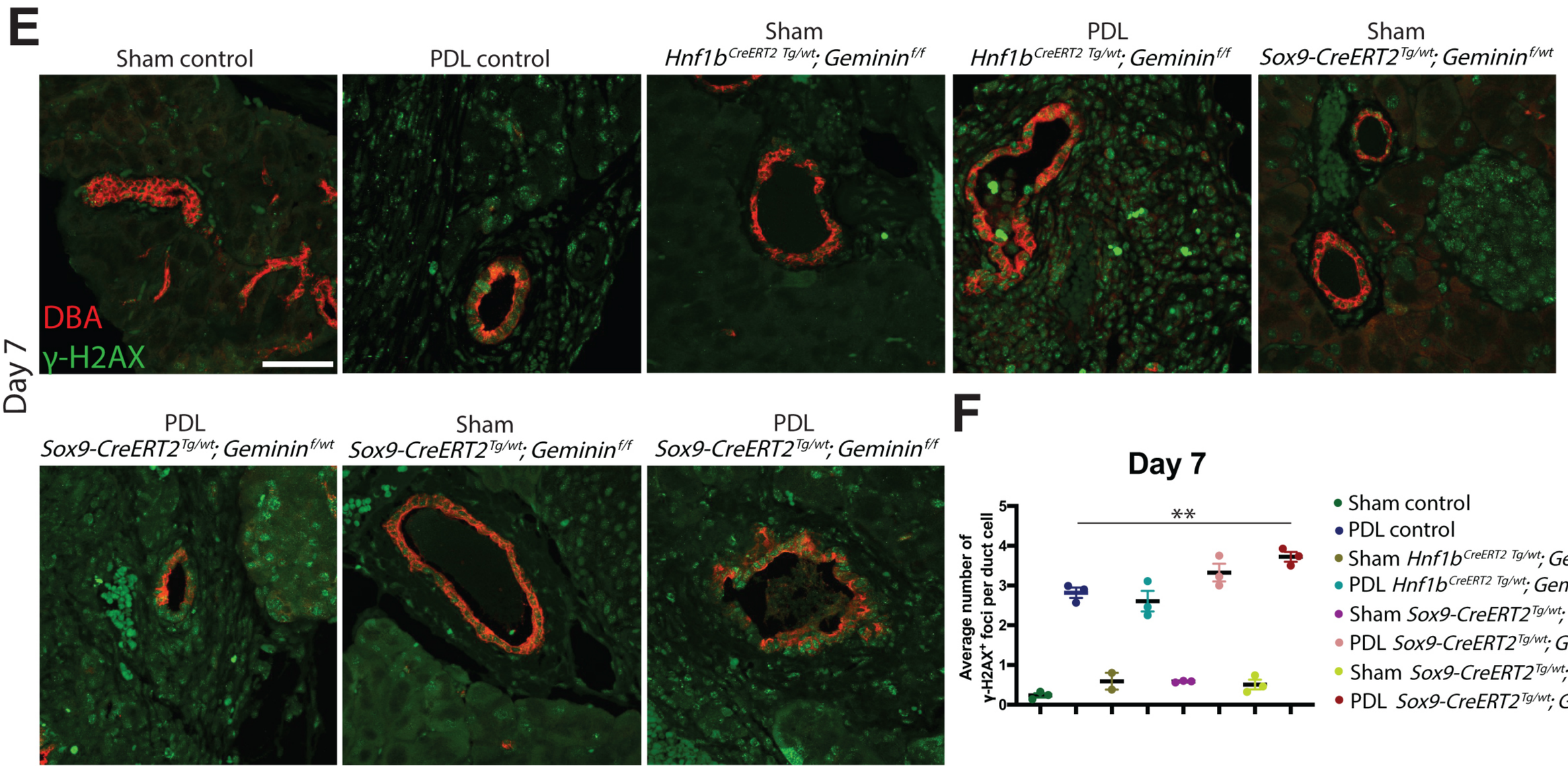

F
% C-peptide/NKX6.1/ PDX1+ cells
2.5
2.0
1.5
1.0
0.5
0.0
***
**
HPDE6c7 scr gRNA
HPDE6c7 SPP1 gRNA2
HPDE6c7 SPP1 gRNA4

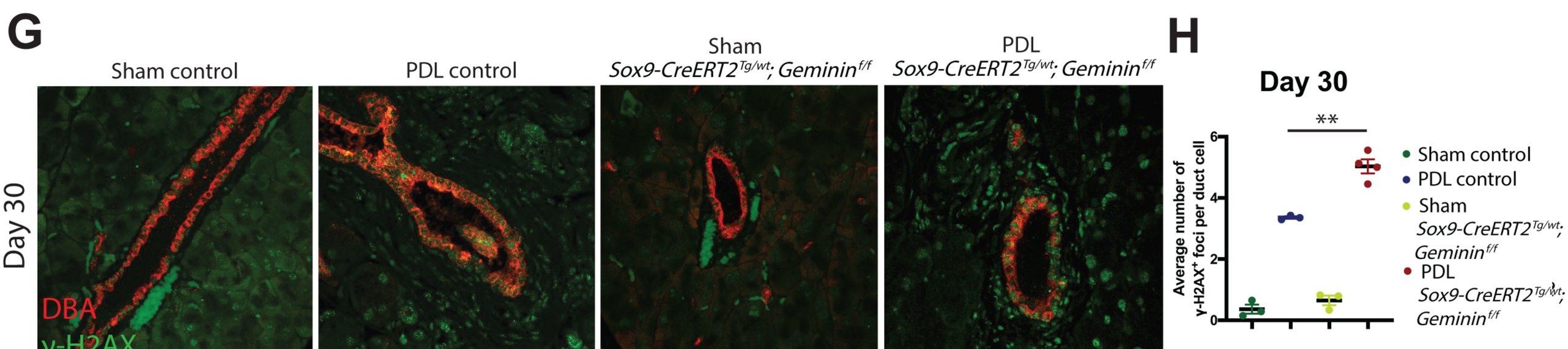

A
Merge + Dapi
CK19
Geminin
Ki67
Normal
Chronic pancreatitis


B
Percent ductal cells
0.10
0.08
0.06
0.04
0.02
0.00
-0.02
*
*
Normal
Chronic pancreatitis
Geminin+
Ki67+

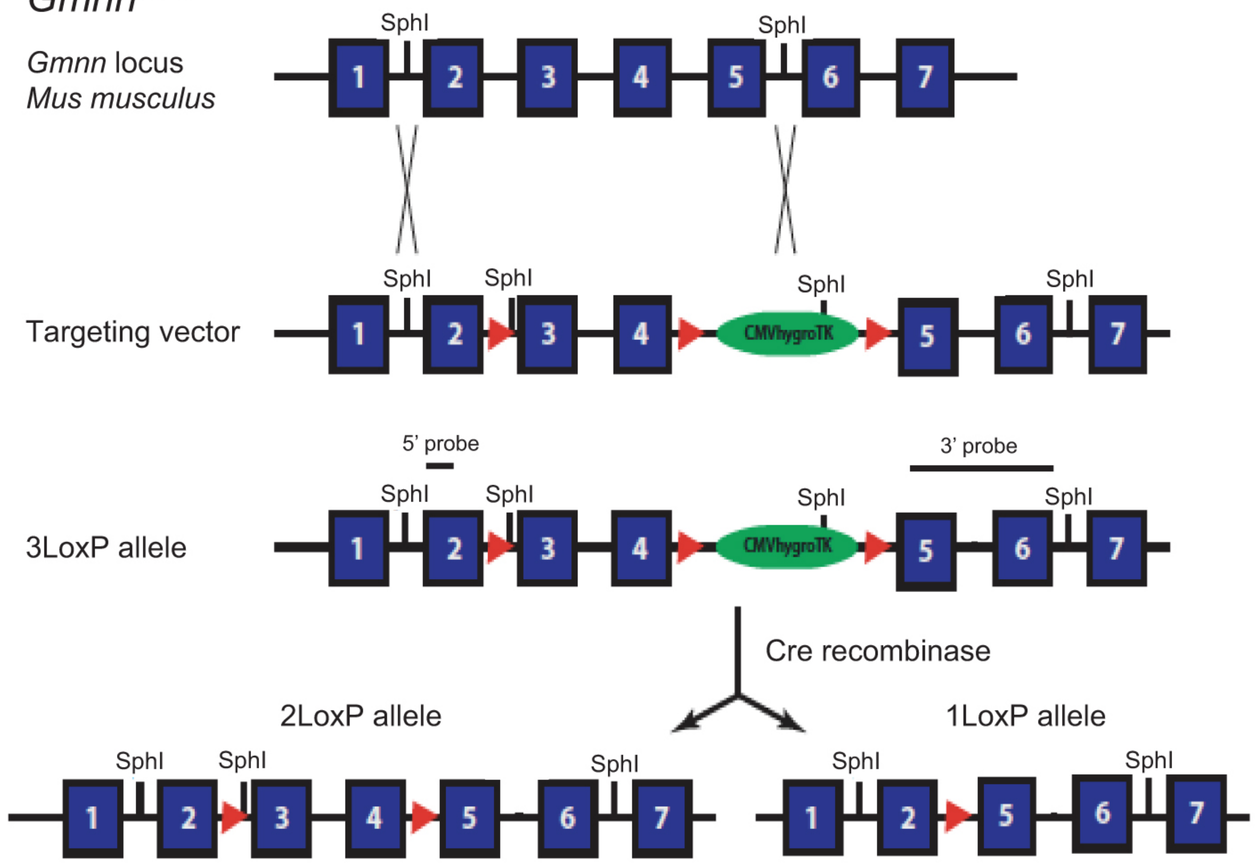

C
0
1
2
3
4
5
Sox9+Hnf1b+
Sox9+
683
522
161
Hnf1b+
Tamoxifen administration
Hnf1b CreERT2 Tg/wt; Geminin and
Sox9-CreERT2 Tg/wt; Geminin mice
Pancreatic duct ligation or sham surgery
BrdU injection
Evaluation of geminin loss in PDL model
7-9 weeks of age
≥ 48 hours post tamoxifen administration
24 and 4 hours before sacrifice
7 days and 30 days post surgery


D
Day 7
**
**
***
Percent ductal cells geminin+
0.08
0.06
0.04
0.02
0.00
-0.02
Sham control
PDL control
Sham Hnf1b CreERT2 Tg/wt; Geminin f/f
PDL Hnf1b CreERT2 Tg/wt; Geminin f/f
Sham Sox9-CreERT2 Tg/wt; Geminin f/wt
PDL Sox9-CreERT2 Tg/wt; Geminin f/wt
Sham Sox9-CreERT2 Tg/wt; Geminin f/f
PDL Sox9-CreERT2 Tg/wt; Geminin f/f

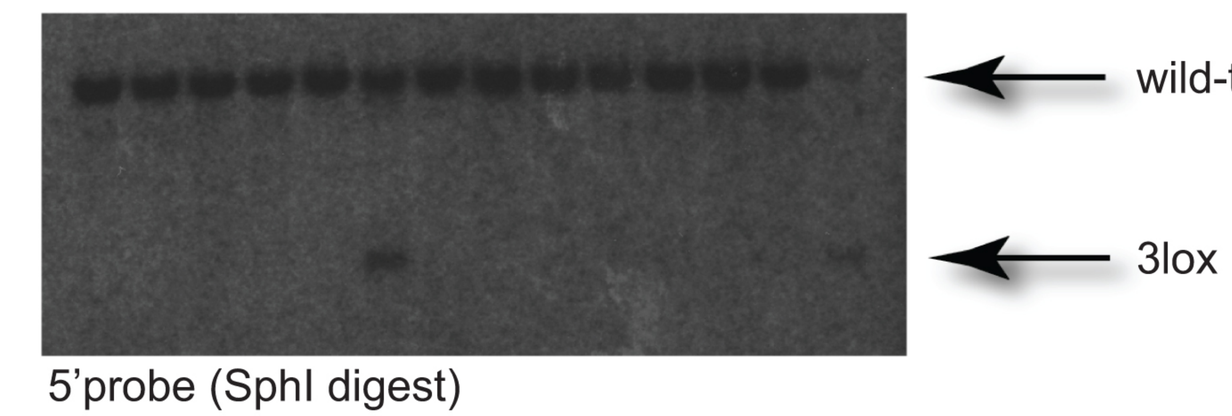

E
Sham control
PDL control
Sham Hnf1b CreERT2 Tg/wt; Geminin f/f
PDL Hnf1b CreERT2 Tg/wt; Geminin f/f
Sham Sox9-CreERT2 Tg/wt; Geminin f/wt
DBA
γ-H2AX
Day 7
PDL Sox9-CreERT2 Tg/wt; Geminin f/wt
Sham Sox9-CreERT2 Tg/wt; Geminin f/f
PDL Sox9-CreERT2 Tg/wt; Geminin f/f
F
Day 7
**
Average number of γ-H2AX+ foci per duct cell
5
4
3
2
1
0
Sham control
PDL control
Sham Hnf1b CreERT2 Tg/wt; Geminin f/f
PDL Hnf1b CreERT2 Tg/wt; Geminin f/f
Sham Sox9-CreERT2 Tg/wt; Geminin f/wt
PDL Sox9-CreERT2 Tg/wt; Geminin f/wt
Sham Sox9-CreERT2 Tg/wt; Geminin f/f
PDL Sox9-CreERT2 Tg/wt; Geminin f/f


G
Sham control
PDL control
Sham Sox9-CreERT2 Tg/wt; Geminin f/f
PDL Sox9-CreERT2 Tg/wt; Geminin f/f
DBA
γ-H2AX
Day 30
H
Day 30
**
Average number of γ-H2AX+ foci per duct cell
6
4
2
0
Sham control
PDL control
Sham Sox9-CreERT2 Tg/wt; Geminin f/f
PDL Sox9-CreERT2 Tg/wt; Geminin f/f

A

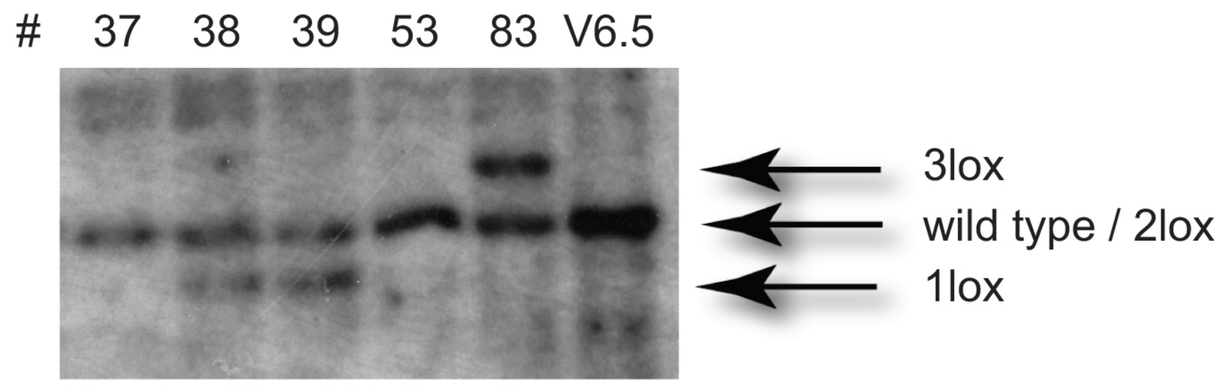


B Generation of 3lox ES cells

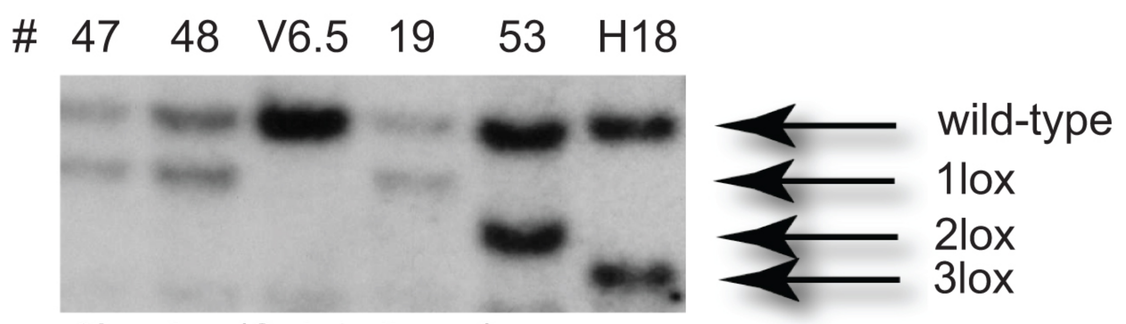


5'probe (SphI digest)

C Generation of 2lox and 1lox ES cells



5'probe (Bsu36I digest)



3'probe (Sph1 digest)

A

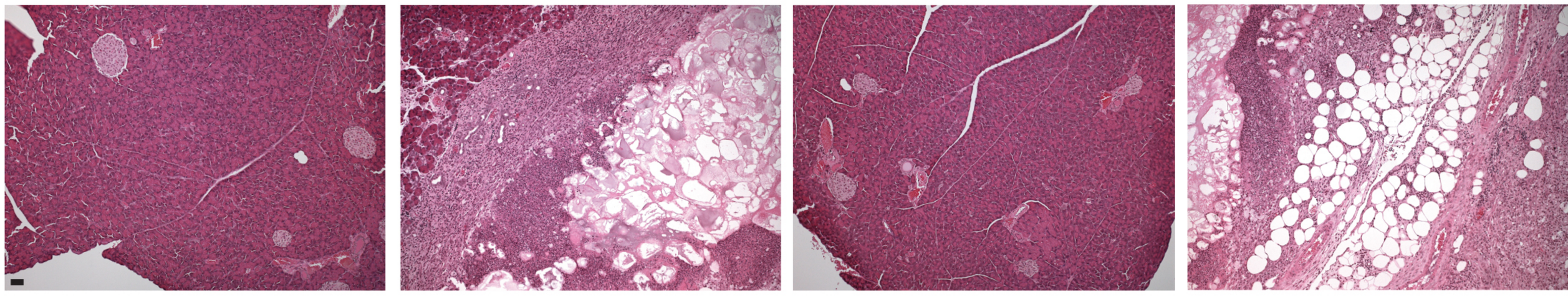
Day 7

Sham control

PDL control

Sham
Hnf1b$^{CreERT2\ Tg/wt}$; *Geminin*$^{f/f}$

PDL
Hnf1b$^{CreERT2\ Tg/wt}$; *Geminin*$^{f/f}$

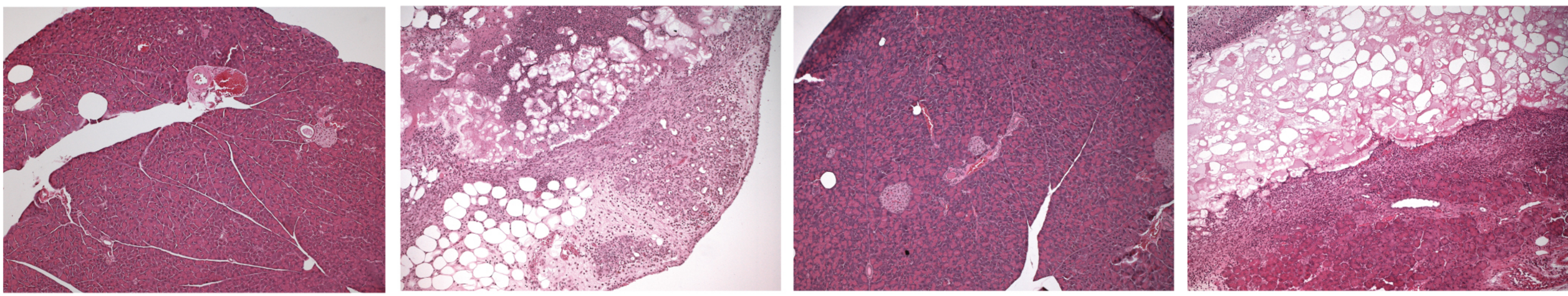


Sham
Sox9-CreERT2$^{Tg/wt}$; *Geminin*$^{f/wt}$

PDL
Sox9-CreERT2$^{Tg/wt}$; *Geminin*$^{f/wt}$

Sham
Sox9-CreERT2$^{Tg/wt}$; *Geminin*$^{f/f}$

PDL
Sox9-CreERT2$^{Tg/wt}$; *Geminin*$^{f/f}$

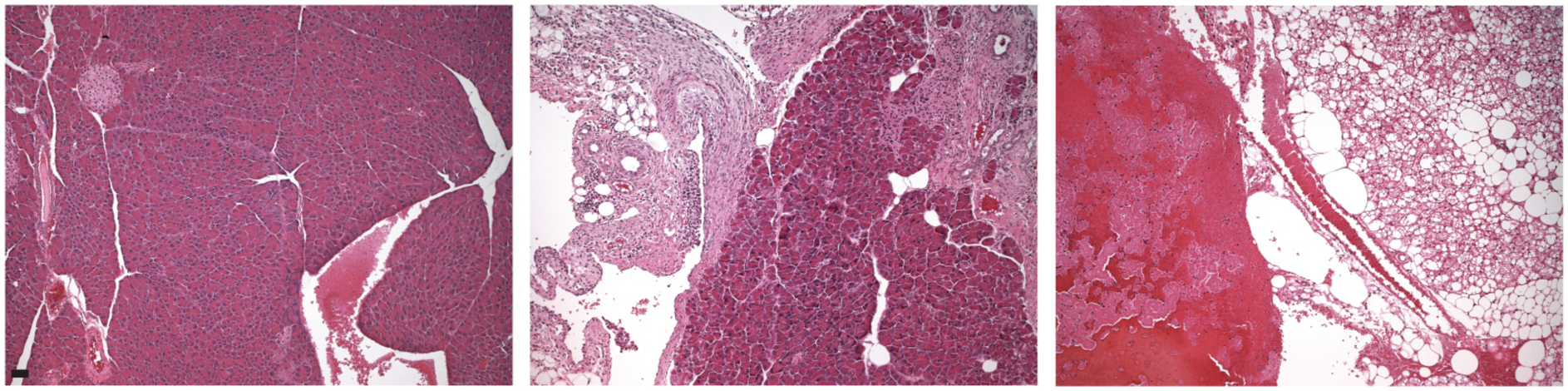


B

Day 30

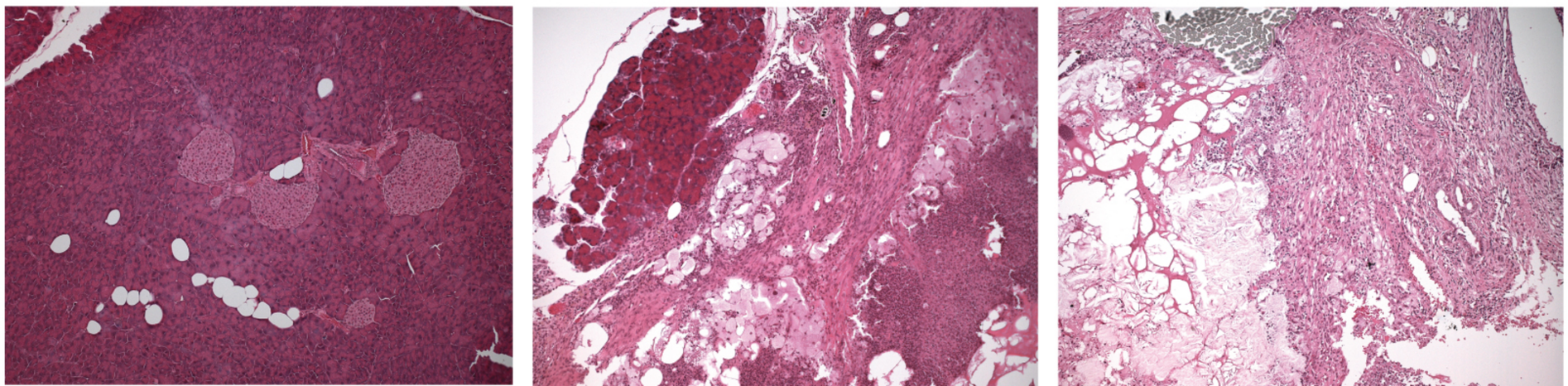
Sham control

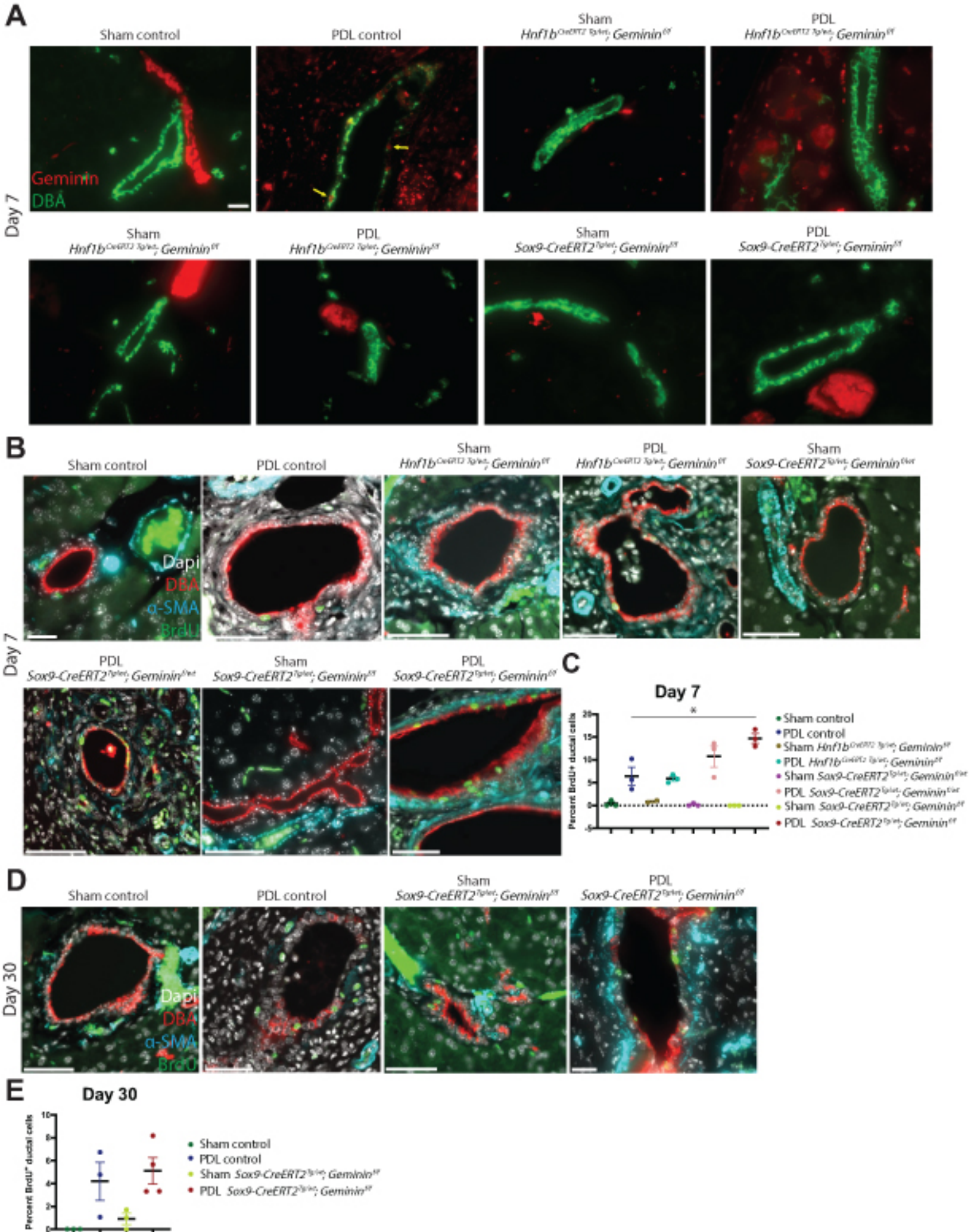
PDL control



Sham
Sox9-CreERT2$^{Tg/wt}$; *Geminin*$^{f/f}$

PDL
Sox9-CreERT2$^{Tg/wt}$; *Geminin*$^{f/f}$




A
Sham control
PDL control
Sham
Hnf1b CreERT2 Tg/wt; Geminin f/f
PDL
Hnf1b CreERT2 Tg/wt; Geminin f/f
Geminin
DBA
Day 7
Sham
Hnf1b CreERT2 Tg/wt; Geminin f/f
PDL
Hnf1b CreERT2 Tg/wt; Geminin f/f
Sham
Sox9-CreERT2 Tg/wt; Geminin f/f
PDL
Sox9-CreERT2 Tg/wt; Geminin f/f
B
Sham control
PDL control
Sham
Hnf1b CreERT2 Tg/wt; Geminin f/f
PDL
Hnf1b CreERT2 Tg/wt; Geminin f/f
Sham
Sox9-CreERT2 Tg/wt; Geminin f/wt
Dapi
DBA
α-SMA
BrdU
Day 7
PDL
Sox9-CreERT2 Tg/wt; Geminin f/wt
Sham
Sox9-CreERT2 Tg/wt; Geminin f/f
PDL
Sox9-CreERT2 Tg/wt; Geminin f/f
C
Day 7
*
Percent BrdU+ ductal cells
Sham control
PDL control
Sham Hnf1b CreERT2 Tg/wt; Geminin f/f
PDL Hnf1b CreERT2 Tg/wt; Geminin f/f
Sham Sox9-CreERT2 Tg/wt; Geminin f/wt
PDL Sox9-CreERT2 Tg/wt; Geminin f/wt
Sham Sox9-CreERT2 Tg/wt; Geminin f/f
PDL Sox9-CreERT2 Tg/wt; Geminin f/f
D
Sham control
PDL control
Sham
Sox9-CreERT2 Tg/wt; Geminin f/f
PDL
Sox9-CreERT2 Tg/wt; Geminin f/f
Dapi
DBA
α-SMA
BrdU
Day 30
E
Day 30
Percent BrdU+ ductal cells
Sham control
PDL control
Sham Sox9-CreERT2 Tg/wt; Geminin f/f
PDL Sox9-CreERT2 Tg/wt; Geminin f/f