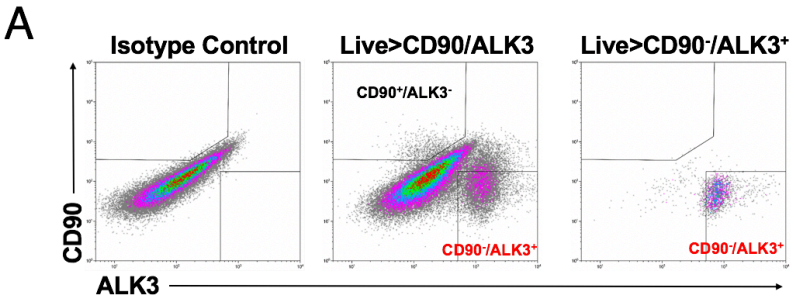
SI APPENDIX

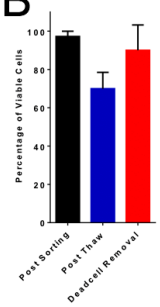
TITLE: Single cell resolution analysis of the human pancreatic ductal progenitor cell niche

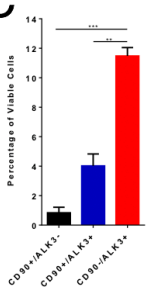
Qadir *et al.*

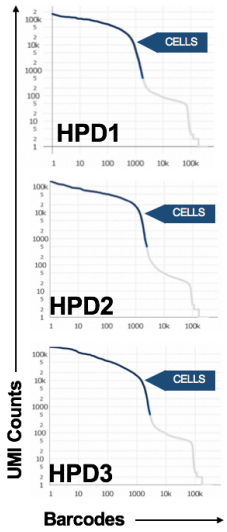
INDEX:

Figure S1



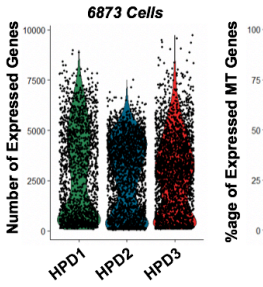


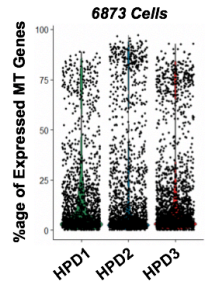


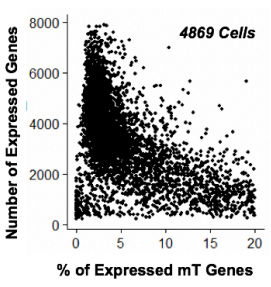


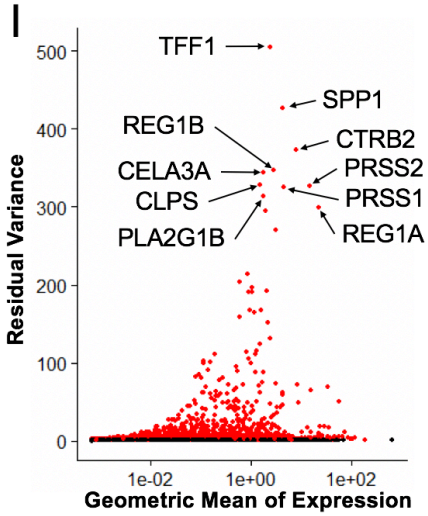
E

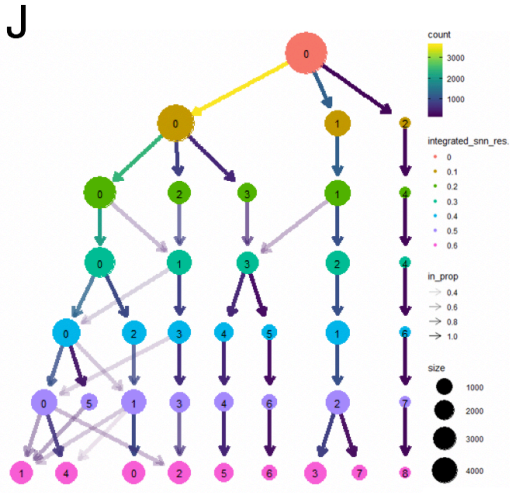
	HPD1	HPD2	HPD3
Sequencing Details (Mapped against GRCh38)			
Total number of Reads	228,375,538	287,545,039	321,425,551
Sequencing Saturation	73.1%	76.3%	76.3%
Valid Barcodes	97.7%	98.1%	98.1%
Mapping Details			
Reads Mapped confidently to Genome	90.1%	92.7%	90.8%
Reads Mapped confidently to Transcriptome	70.8%	73.5%	68.9%
Cell Details			
Estimated number of Cells	1,761	2,412	2,700
Fraction Reads in Cells	87.8%	95.9%	91.7%
Mean Reads per Cell	129,685	119,214	119,046
Median Genes per Cell	2,915	3,045	2,852
Total Genes Detected	20,290	20,513	21,063
Median UMI counts per Cell	10,315	13,677	10,722

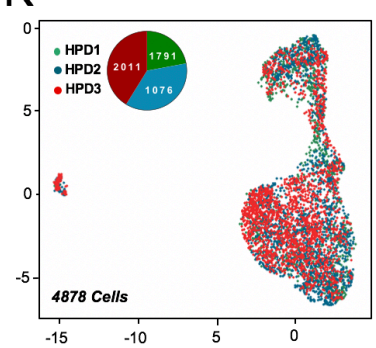












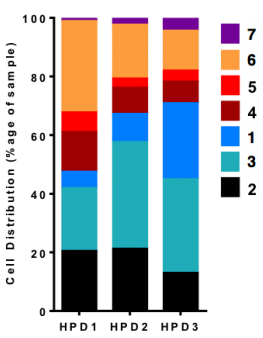


Figure S1. Cell sorting strategy and quality control of ALK3$^{bright+}$ sorted cells. (A) Representative FACS plots for live single exocrine cells stained for CD90 (mesenchymal) and ALK3 (pan-ductal) markers. CD90^{-}/ALK3$^{bright+}$ cells were sorted and cryopreserved for single cell library preparation. **(B)** Bar plot showing viability after 1) FACS sorting; 2) thawing; and 3) dead cell removal. **(C)** Bar plot showing the percentage of viable CD90^{+}/ALK3^{-}, CD90^{+}/ALK3^{+} and CD90^{-}/ALK3^{+} cells, in human exocrine tissue. **(D)** Cell calls generated with the Cell Ranger software, which is based on the number of UMIs/barcode (cell). The sharp drop denotes the UMI threshold for a barcode to be correctly categorized as a cell. **(E)** Quality control parameters and statistics for all single cell libraries sequenced with the 10X Genomics platform. Sequencing saturation: the fraction of library complexity that was sequenced. An inverse of this metric denotes the number of reads it would theoretically take to detect a new transcript. Valid barcodes: denotes percentage of cells with a barcode matching the correct synthetic barcode present on a gel-emulsion (GEM) bead. Mapped reads to transcriptome/genome: denotes the percentage of reads confidently mapping to a reference transcriptome and genome, respectively. Fraction reads in cells: denotes the percentage of total reads that contain a cell-specific barcode. Mean reads per cell: the average number of reads per barcode-specified cell. Median genes per cell: the median number of unique genes expressed in any given barcode-specified cell. **(F)** Total number of expressed genes in each cell, where each cell is represented by a single dot. **(G)** Bar plot showing the expression of mitochondrial genes as a percentage of total mRNA in each cell. Each dot represents a single cell. **(H)** During downstream thresholding and normalization, only a small subset of cells shown in **Figure S1*F*** and **Figure S1-*G*** are filtered for further analysis. This scatter plot shows only the selected cells. The plot shows the expression of mitochondrial genes as a percentage of total mRNA, compared to the total number of genes expressed in any given cell. Here, a single dot represents a single cell. **(I)** Scatterplot showing the dispersion of variable genes as an outcome of the integration analysis across cells used. Each dot represents a unique gene. The top 3,000 genes are shown in red, with the top 10 marked on the graphic. **(J)** A clustree plot created using Clustree v0.4.1 (1). Each level in this plot denotes a dimensional resolution used for clustering cells, where resolution increases moving down the plot. Lines denote contributions of cells from previous clusters, and the size of the circles represents the number of cells. The goal of this analysis is to select in an unbiased manner a resolution consistent with overall cluster stability, preventing 'over' or 'under' clustering of scRNAseq data. Based on this analysis, since cluster destabilization begins to occur at a resolution of 0.5 and 0.6, we selected a resolution of 0.4. **(K)** Uniform Manifold Approximation and Projection (UMAP) plot showing the contribution of each donor to the analyzed clusters. **(L)** Bar plot indicating the percentage distribution and contribution of various donors towards clusters. **S1*B*** and **S1*C***: **p<0.01, ***p<0.001.

Figure S2

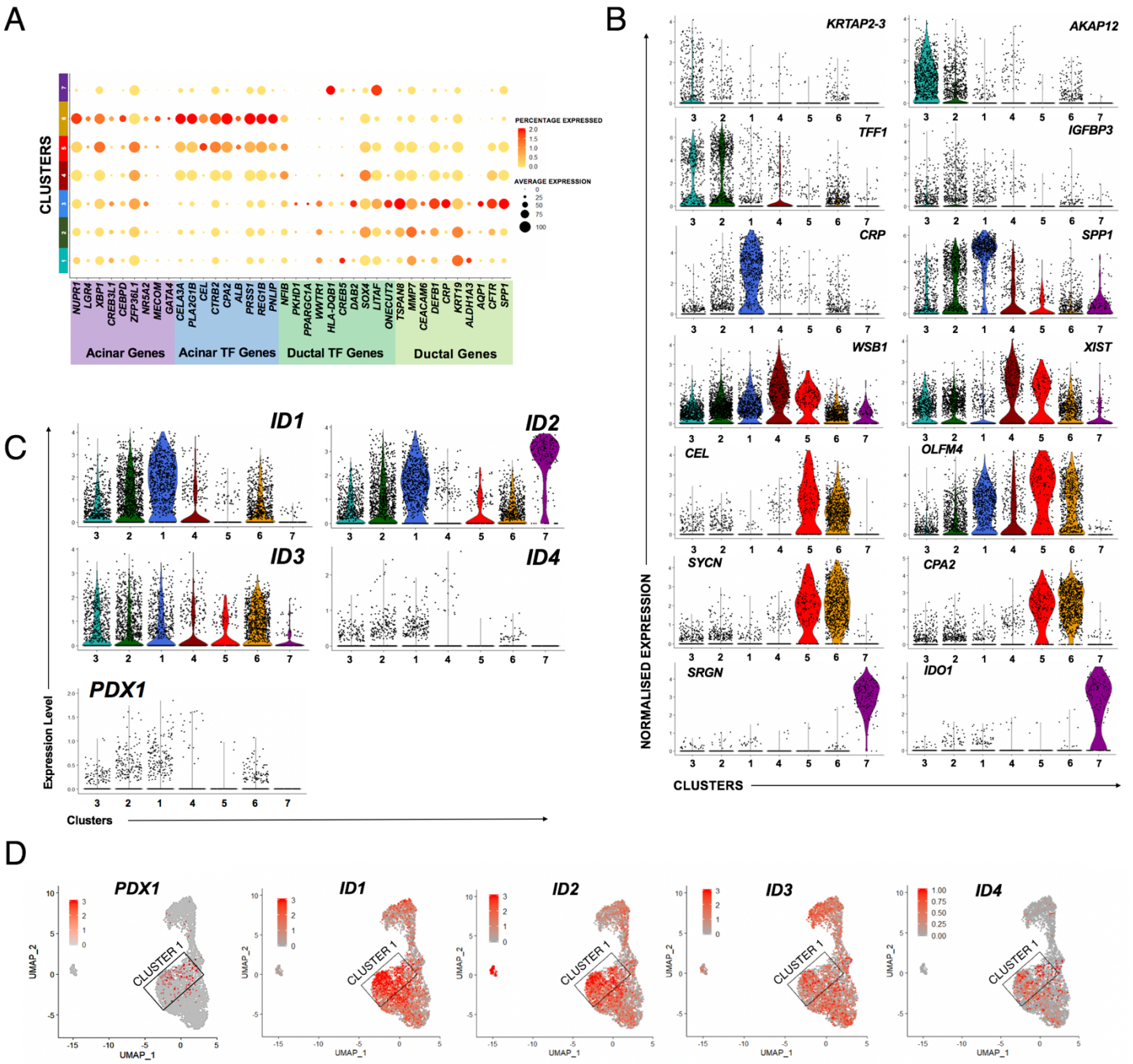


Figure S2. Pathway analysis yields unique populations of ductal and progenitor-like cells. **(A)** Dot plot representing the expression of known ductal and acinar gene transcription factors (2). **(B)** Violin plots showing probability distribution and expression of two selected differentially expressed genes/cluster as outlined in **Figure 2*A-C***. **(C)** Violin plots showing probability distribution and expression of *ID* genes and *PDX1* across all clusters. **(D)** UMAP plot showing gene expression for *ID* genes and *PDX1*.

Figure S3

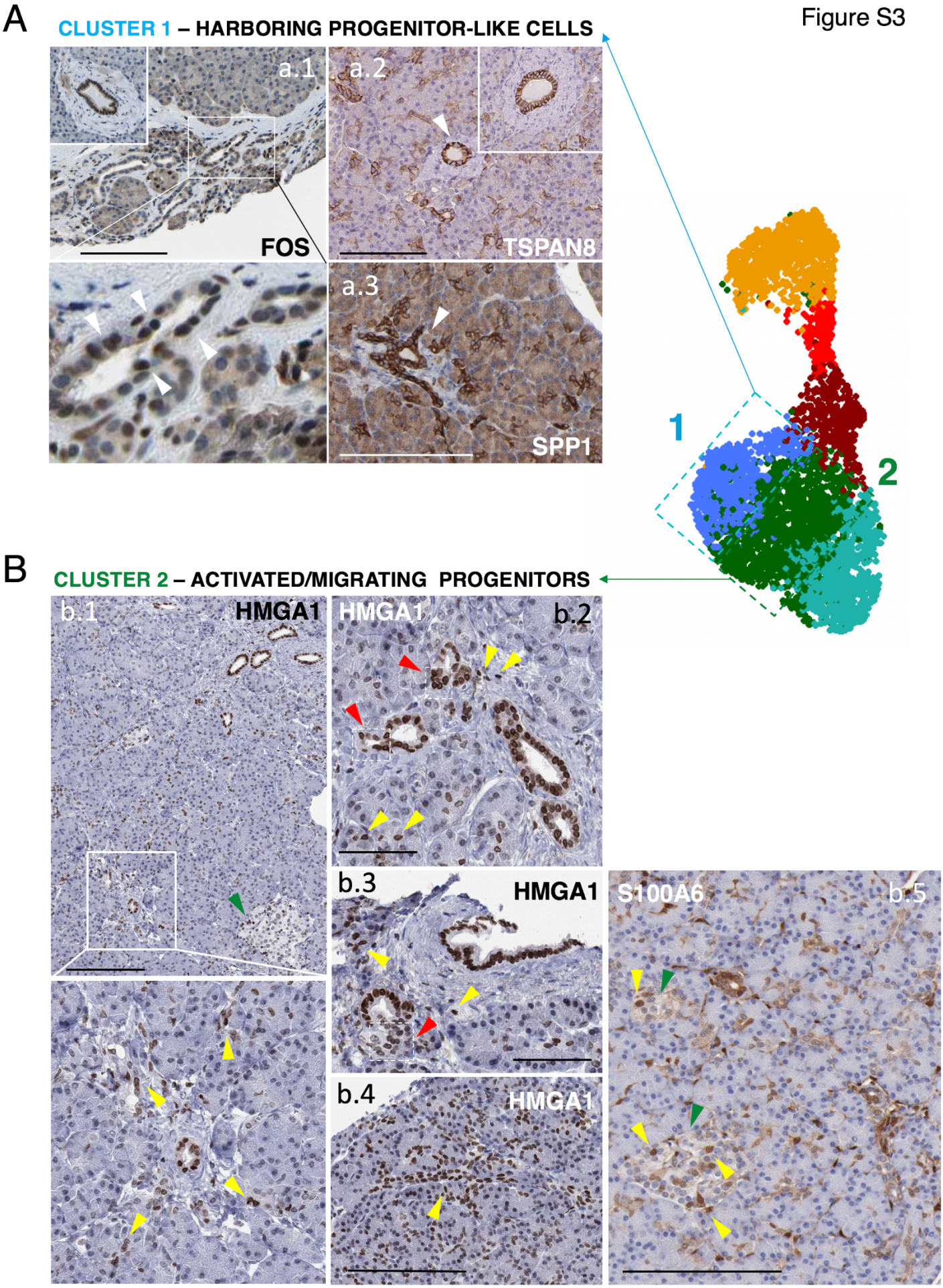

A
CLUSTER 1 – HARBORING PROGENITOR-LIKE CELLS
a.1
FOS
a.2
TSPAN8
a.3
SPP1
1
2
B
CLUSTER 2 – ACTIVATED/MIGRATING PROGENITORS
b.1
HMGA1
HMGA1
b.2
b.3
HMGA1
b.4
HMGA1
S100A6
b.5

Figure S3. Representative immunostaining of select differentially expressed genes for clusters 1 and 2. **(A)** Immunostaining of FOS (a.1 and detail), TSPAN8 (a.2) and SPP1 (a.3), representative DE genes for cluster 1. In general, staining patterns preferentially suggest major ductal/PDG locations (white arrows). PDGs exhibit an intermittent pattern of staining for FOS, with cells positive for this marker intercalated with negative ones (a.1 magnified panel). Insets for FOS and TSPAN8 show larger ducts with thick fibromuscular stroma. **(B)** Immunostaining of HMGA1 (b.1-b.4) and S100A6 (b.5), representative DE genes for cluster 2. As in cluster 1, staining is also typically strong in major ducts and PDGs, but cells seemingly detaching from them, suggesting potential migration, are seen throughout the tissue section. Yellow arrows in b.1 (magnified detail) and b.2-b.5 show single cells. Dotted white boxes and red arrows in b.2 and b.3 denote areas of potential 'breakage' from the tight epithelial structure of the ducts and PDGs. Green arrows indicate islets. Size bars: 100 μm for A (a.1-a.3) and B (b.1 and b.5). 50 μm for B (b.3 and b.4). Images were obtained from the Human Protein Atlas (www.proteinatlas.org). Specific links to individual markers are provided in the Resources Table.

Figure S4

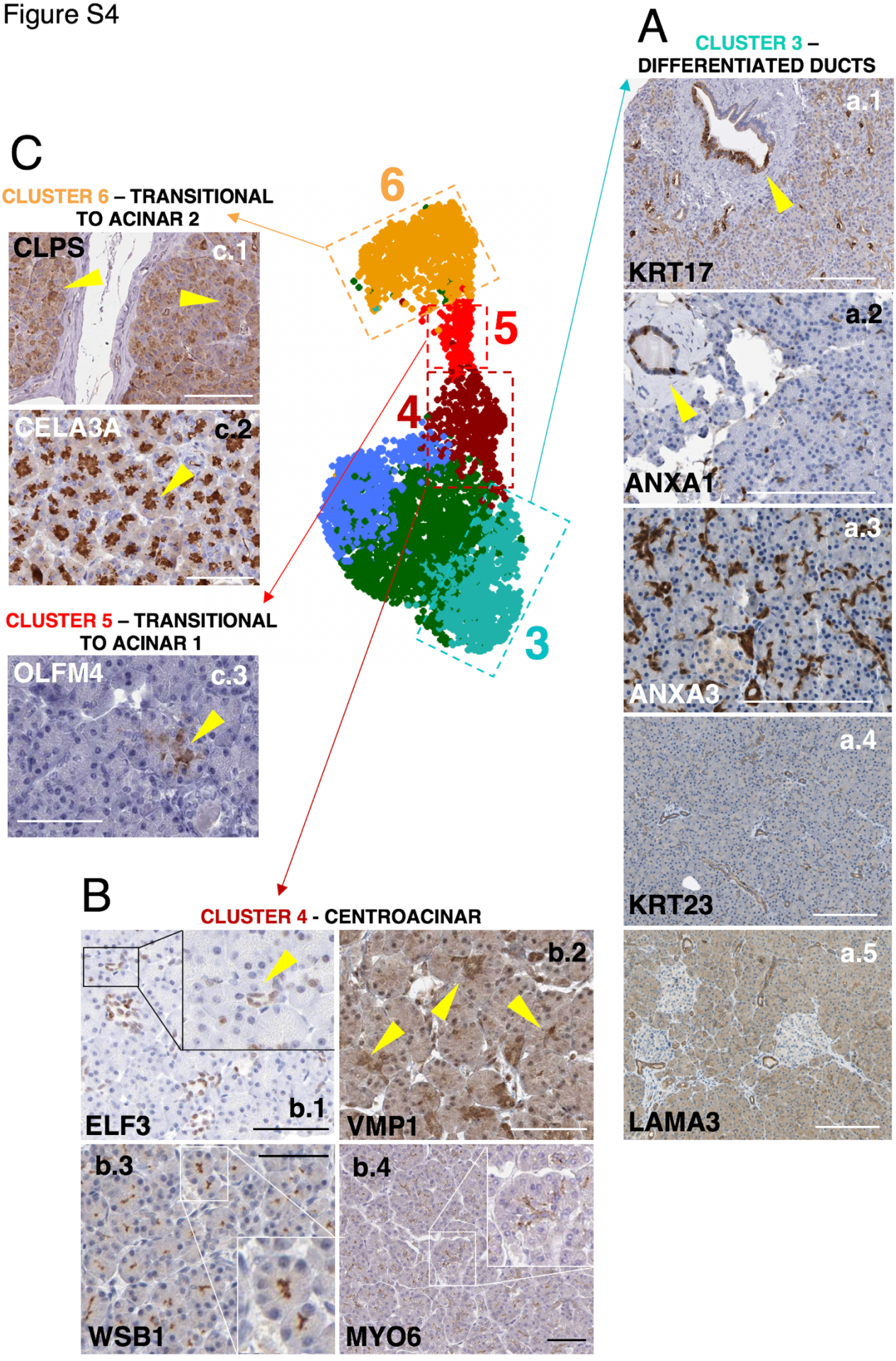

A
CLUSTER 3 – DIFFERENTIATED DUCTS
a.1
KRT17
a.2
ANXA1
a.3
ANXA3
a.4
KRT23
a.5
LAMA3
C
CLUSTER 6 – TRANSITIONAL TO ACINAR 2
CLPS
c.1
CELA3A
c.2
CLUSTER 5 – TRANSITIONAL TO ACINAR 1
OLFM4
c.3
6
5
4
3
B
CLUSTER 4 - CENTROACINAR
ELF3
b.1
b.2
VMP1
b.3
WSB1
b.4
MYO6

Figure S4. Representative immunostaining of select differentially expressed genes for clusters 3-6. **(A)** Immunostaining of KRT17 (a.1), ANXA1 (a.2), ANXA3 (a.3), KRT23 (a.4) and LAMA3 (a.5), representative DE genes for cluster 3. In general, staining patterns preferentially suggest minor ductal locations, with major ducts (a.1 and a.2, yellow arrows) often displaying intermittent signal. **(B)** Immunostaining of ELF3 (b.1), VMP1 (b.2), WSB1 (b.3) and MYO6 (b.4), representative DE genes for cluster 4. The signal is generally consistent with a preferential centroacinar location (b.1-b.4, yellow arrows), with ELF3 staining found in cells with the typical flattened shape of centroacinar cells (b.1, magnified inset, yellow arrow). **(C)** Immunostaining of CLPS (c.1), CELA3A (c.2), representative DE genes for cluster 6; and OLFM4 (c.3), representative DE gene for cluster 5. Both clusters exhibit "transitional to acinar" characteristics, with the selected DE genes showing a centroacinar rather than a pan-acinar pattern of staining. Size bars: 100 μm for A (a.1-a.5) and 50 μm for B (b.1.-b.4) and C (c.1-c.3). Images were obtained from the Human Protein Atlas (www.proteinatlas.org). Specific links to individual markers are provided in the Resources Table.

Figure S5

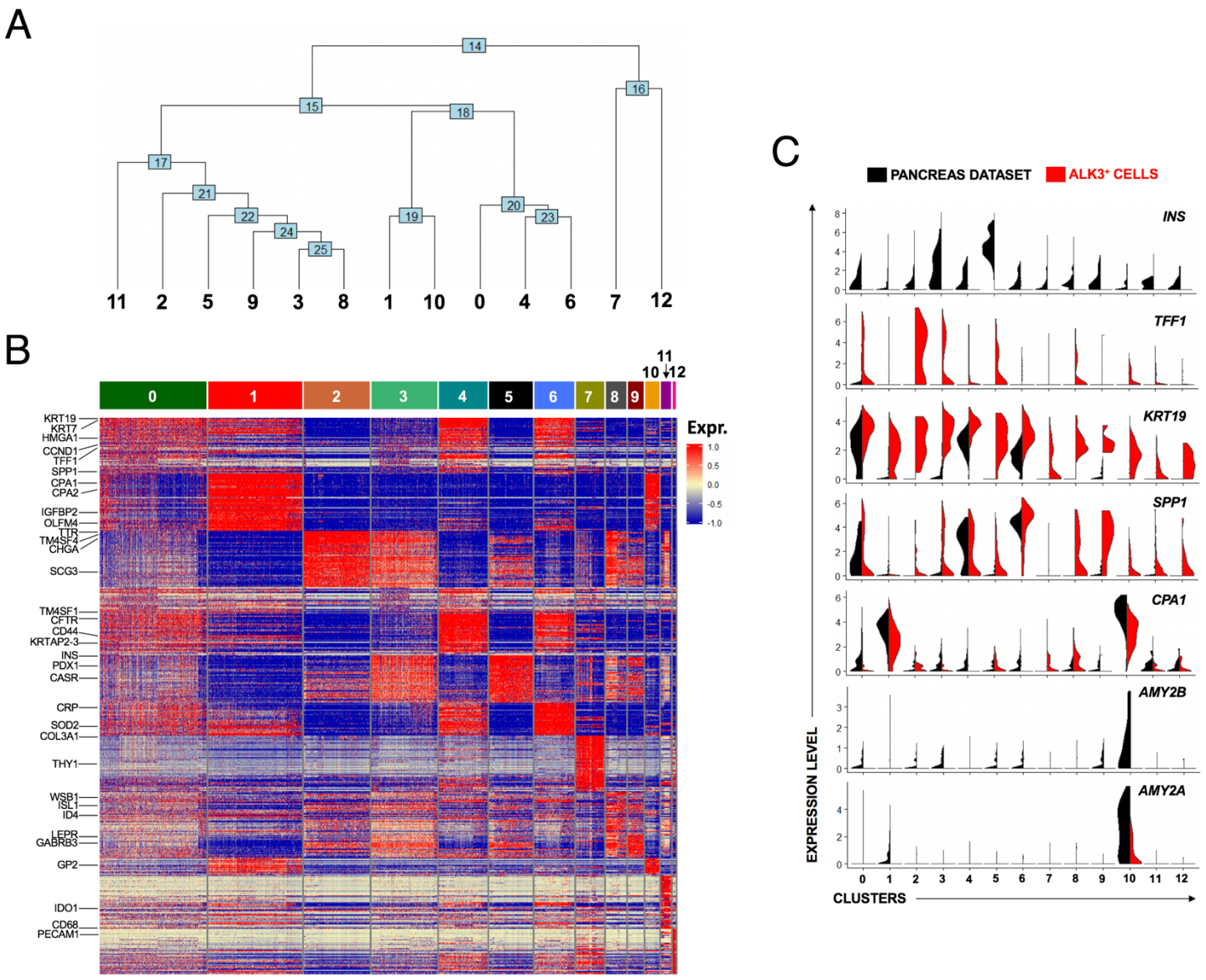


Figure S5. Unique ductal and progenitor-like cell cluster distribution is maintained after an integrated analysis against a human islet (whole-pancreas) single cell gene expression dataset. **(A)** Cluster dendrogram created using the {PlotClusterTree} function in {Seurat}. The dendrogram shows the Euclidean relationships between analyzed clusters. **(B)** Heatmap showing the top 50 most differentially expressed genes in each cluster of the integrated analysis, as shown in **Figure 4*A***. Select genes differentiating clusters are shown on the left. **(C)** Dual violin plots showing side by side gene expression distribution of cell clusters derived from Muraro-Grun (black) and the ALK3^{+} sorted single-cell transcriptome (red).

Figure S6

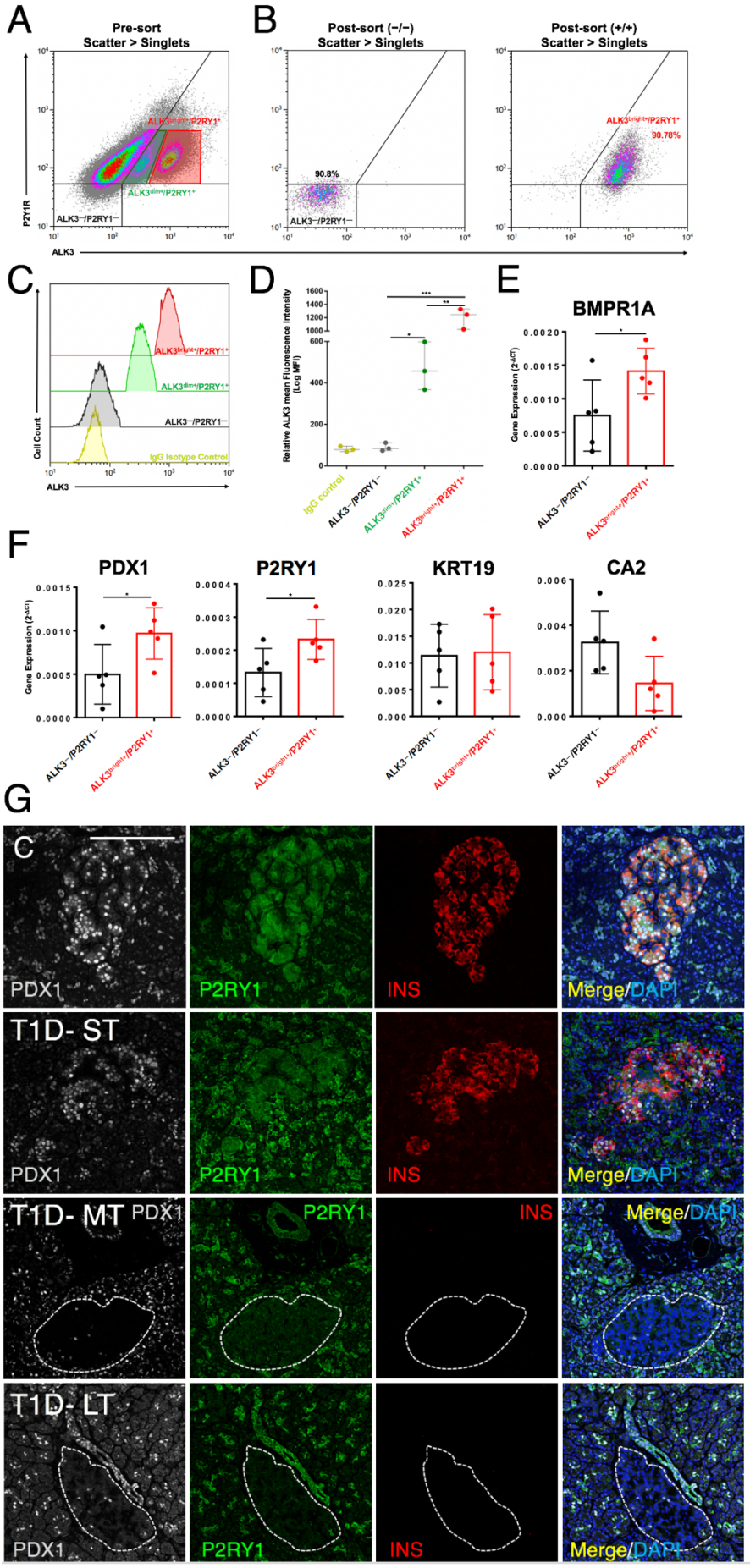

A
Pre-sort
Scatter > Singlets
P2Y1R
ALK3
ALK3−/P2RY1−
ALK3dim+/P2RY1+
ALK3bright+/P2RY1+
B
Post-sort (−/−)
Scatter > Singlets
90.8%
Post-sort (+/+)
Scatter > Singlets
90.78%
C
Cell Count
IgG Isotype Control
D
Relative ALK3 mean Fluorescence Intensity (Log MFI)
IgG control
E
BMPR1A
Gene Expression (2^ΔCT)
F
PDX1
P2RY1
KRT19
CA2
G
C
PDX1
P2RY1
INS
Merge/DAPI
T1D- ST
T1D- MT
T1D- LT

Figure S6. ALK3$^{bright+}$/P2RY1^{+} sorting enriches for PDX1^{+}/ALK3^{+}/CAII^{-} cells. (A) Representative FACS plot for the strategy employed to sort ALK3$^{bright+}$/P2RY1^{+} cells. The cells shown have been gated for live singlets. **(B)** Representative FACS plots of post-sorted ALK3^{-}/P2RY1^{-} cells (left panel) and ALK3$^{bright+}$/P2RY1^{+} cells (right panel). **(C)** Representative mean fluorescence intensity (MFI) histogram-projection plots showing the MFI of ALK3 expression across IgG control (yellow), ALK3^{-}/P2RY1^{-} (gray), ALK3^{dim+}/P2RY1^{+} (green) and ALK3$^{bright+}$/P2RY1^{+} (red) cells. **(D)** MFI intensity scatterplots for IgG control (yellow), ALK3^{-}/P2RY1^{-} (gray), ALK3^{dim+}/P2RY1^{+} (green) and ALK3$^{bright+}$/P2RY1^{+} (red) cells. **(E)** Normalized gene expression of BMPR1A (ALK3) in ALK3^{-}/P2RY1^{-} and ALK3$^{bright+}$/P2RY1^{+} cells. **(F)** Normalized gene expression of PDX1, P2RY1, KRT19 and CAII in ALK3^{-}/P2RY1^{-} and ALK3$^{bright+}$/P2RY1^{+} cells. **(G)** Representative confocal immunofluorescence imaging of a Z-stack maximal projection showing protein expression of: Pancreatic duodenal homeobox-1 (grey), Purinergic receptor Y1 (green), Insulin (red), and DAPI (blue). Panels show: C, non-diabetic human donors; T1D-ST, short-term; T1D-MT, medium-term; and T1D-LT, long-term type 1 diabetic human donors. Dashes outline islets. Size bars = 200 μm.

TABLE S1

Sample ID	Diabetes Duration (Years)	Auto Antibody Assay (RIA)	C-Peptide (ng/ml)	Age (Years)	Sex	Ethnicity	BMI
HPD1	0	N/A	N/A	47	F	Caucasian	24.1
HPD2	0	N/A	N/A	51	F	Caucasian	28.5
HPD3	0	N/A	N/A	35	M	Caucasian	24.7
HPD4	0	N/A	N/A	56	F	Caucasian	26.6
HPD5	0	N/A	N/A	49	M	Hispanic	25.9
HPD6	0	N/A	N/A	36	M	Caucasian	25.5
HPD7	20 (T2D)	N/A	N/A	69	F	Caucasian	35
HPD8	2 (T2D)	N/A	N/A	59	M	Caucasian	31.6
HPD9	5 (T2D)	N/A	N/A	61	M	African American	33
HPD10	2 (T2D)	N/A	N/A	52	M	Hispanic	30.9
nPOD6393	0	Neg	8.3	15.5	M	Hispanic	27.1
nPOD6389	0	Neg	7.22	18.6	M	Caucasian	20.9
nPOD6368	0	Neg	3.05	38.3	M	Caucasian	20.7
nPOD6380	0	Neg	0.22	11.6	F	African American	14.6
nPOD6371	2 (T1D)	GADA A+ , mlAA+ , IA-2A+ , ZnT8A+	0.11	12.5	F	Caucasian	16.6
nPOD6342	2 (T1D)	IA-2A+ , mlAA+	0.26	14	F	Caucasian	24.3
nPOD6265	8 (T1D)	GADA A+ , mlAA+	0.06	11	M	Caucasian	12.9
nPOD6207	10 (T1D)	IA-2A+ , ZnT8A+ , mlAA+	<0.05	16.7	F	African American	24.4
nPOD6045	8 (T1D)	ZnT8A+ , mlAA+	<0.05	18.8	M	Caucasian	23.1
nPOD6328	20 (T1D)	GADA A+ , mlAA+	<0.02	39	M	Hispanic	24
nPOD6327	57 (T1D)	mlAA+	<0.02	71.2	M	Hispanic	23.1
nPOD6205	33 (T1D)	mlAA+	0.14	40.9	F	Caucasian	22.6

Table S1. Demographics. A table of describing demographic information of all human donor pancreata used in this manuscript. Non-diabetes donor samples are shaded in green, type 1 diabetes donor samples in orange, and type 2 diabetes donor samples in blue.

B. SUPPLEMENTAL METHODS

Exocrine tissue isolation and tissue culture

Pancreata were isolated from deceased human donors and processed for islet isolation via the Ricordi method (1). Exocrine tissue, a by-product of islet isolation, was washed 5X using DPBS (Sigma Aldrich, St. Louis, MO, USA, Cat# D8537), supplemented with 100 µg/ml trypsin inhibitor isolated from *Glycine max* (Sigma Aldrich, St. Louis, MO, USA, Cat# T6522), 1X Penicillin/Streptomycin/Amphotericin-B solution (Sigma Aldrich, St. Louis, MO, USA, Cat# A5955) and 5% fetal bovine serum (Invitrogen, Carlsbad, CA, USA, Cat# 10082147). Exocrine clusters were allowed to settle by gravity for 5-10 min between each wash (i.e., not centrifuged). The tissue was then washed once with stabilization optimized media for exocrine pancreas (StOMP), which is composed of RPMI 1640 plus GlutaMAX (Invitrogen, Carlsbad, CA, USA, Cat# 61870036) and further supplemented with 10 mM HEPES buffer (Invitrogen, Carlsbad, CA, USA, Cat# 15630080), 100 µg/ml trypsin inhibitor (Sigma Aldrich, St. louis, MO, USA, Cat# T6522), 100 µg/ml Aprotinin (Sigma Aldrich, St. Louis, MO, Cat# A6106), 1X Penicillin/Streptomycin/Amphotericin-B solution (Sigma Aldrich, St. Louis, MO, USA, Cat# A5955), 1 mM sodium pyruvate (Invitrogen, Carlsbad, CA, USA, Cat# 11360070) and 10% fetal bovine serum (Invitrogen, Carlsbad, CA, USA, Cat# 10082147) or 1% B27-minus insulin (Invitrogen, Carlsbad, CA, USA, Cat# A1895601). Exocrine tissue was then suspended in 30 ml StOMP and plated in T175 suspension flasks (Sarstedt, Germany, Cat# 83.3912.502), at 400 µl tissue/flask. Post-isolation exocrine tissue was cultured overnight (<12hrs) and then prepared for cell sorting.

FACS and cryopreservation of ALK3$^{bright+}$ cells for scRNAseq analysis

Overnight (<12hrs) cultured exocrine tissue was dissociated into single cells by suspending it in Tryple-E Express Enzyme (1X) (Invitrogen, Carlsbad, CA, USA, Cat# 12604013) for 5-10 min. Enzyme activity was inhibited upon cell dispersion by StOMP at a 1:5 ratio to Tryple-E Express Enzyme. Cells were then passed through a 40 µm cell strainer (Corning, Corning, NY, USA, Cat#: 352340) to obtain single cells. These were then washed with StOMP 3x times. Cells were re-suspended in DPBS (Sigma Aldrich, St. Louis, MO, USA, Cat# D8537) and stained with a near-IR viability dye (Invitrogen, Carlsbad, CA, USA, Cat# L34975) according to the manufacturer's recommendations. $5x10^6$ live cells (dispersed in 5ml, viability assessed using Trypan blue and a hemocytometer) were then washed 3X with StOMP and the re-suspended in StOMP for counterstaining with anti-CD90/AF488 or anti-CD90/PE (1:100) and anti-ALK3/AF647 (1:200), for 45 min at 4°C. Stained cells were then washed 3X with StOMP, prior to re-suspending in StOMP at $1x10^6$ cells/ml of media for cell sorting. Cells were sorted after gating for the live CD90$^-$/ALK3$^{bright+}$ cell fraction on a MoFlo Astrios EQ cell sorter (BD, Franklin Lakes, NJ, USA), and analyzed using the Summit software v6.3 (Cytomation) and the Kaluza software v1.5a (BD, Franklin Lakes, NJ, USA). Cells were then washed 3x times with StOMP and re-suspended in CryoStor CS10 (Stemcell Technologies, Canada, Cat# 07930). Cryopreserved cells were then stored at -80°C until preparation for library construction.

FACS of P2RY1$^+$/ALK3$^{bright+}$ cells for transplantation

Exocrine tissue cultured in StOMP for 48 hours was then sorted according to the strategy previously utilized for ALK3$^{bright+}$ cells alone, with a small modification. In addition to anti-ALK3-AF647, cells were also incubated with anti-P2RY1-AF488 antibody. Cells were sorted after gating for the live P2RY1$^+$/ALK3$^{Bright+}$ cell fraction on a MoFlo Astrios EQ cell sorter (BD, Franklin Lakes, NJ, USA), and analyzed using the Summit software v6.3 (Cytomation) and the Kaluza software v1.5a (BD, Franklin Lakes, NJ, USA).

Animal Experimentation

All animal experiments were conducted under the supervision and oversight of the University of Miami Institutional Animal Care and Use Committee (IACUC) and Division of Veterinary Resources (DVR) at the University of Miami. *Foxn1^{nu}* (nude) mice (5-6 weeks old; Jackson Laboratories, Bar Harbor, ME, Cat# 002019) were housed in specific pathogen-free (SPF) conditions at the DVR's animal care facility. For all experiments, mice were acclimated for 7-10 days prior to any experimental intervention. They were maintained on a 12h light/dark cycle with *ad libitum* access to standard irradiated chow and filtered drinking water.

P2RY1$^+$/ALK3$^{bright+}$ transplantation into the kidney capsule of nu/nu mice

Sorted P2RY1$^+$/ALK3$^{bright+}$ cells were suspended in hESC grade Matrigel (Corning, Corning, NY, USA, Cat# 354277) and DPBS (Sigma Aldrich, St. louis, MO, USA, Cat# D8537) in a 1:1 ratio (care should be taken to thaw Matrigel on ice). Male and female immune deficient *Foxn1^{nu}* mice were used for experiments as outlined by the NIH (Notice number: NOT-OD-15-102). Mice were anaesthetized with inhalable isoflurane and received ~$5x10^5$ cells under the left kidney capsule.

Metabolic analysis of in vivo differentiation

All metabolic analyses were performed on healthy restrained, conscious mice on days indicated in the manuscript (see Figure 5B for an outline). Peripheral blood was obtained via saphenous vein, and tested for glucose using a handheld glucometer (Bayer, Whippany, NJ). For C-peptide assessments, saphenous vein blood samples were collected in heparinized microhematocrit tubes. Plasma was separated by centrifuging samples at 2,400 rpm for 10-15 mins. Collected plasma was frozen at -20°C and analyzed using a ultra-sensitive human C-peptide ELISA kit according to the manufacturer's recommendations (Mercodia, Cat#: 10-1141-01). Intraperitoneal glucose tolerance tests (IPGTTs) were performed at 29 days post transplantation. For IPGTTs, mice were fasted for 6 hours prior to

administration of a 2 g/kg BW glucose bolus. Blood was collected prior to the administration of the bolus and after 60 minutes of administration. Human C-peptide was calculated using an ultrasensitive ELISA according to the manufacturer's recommendations (Mercodia, Cat#: 10-1141-0). At the termination of the experiment, mice were sacrificed and transplanted kidneys extracted and preserved in 4% paraformaldehyde for histological immunofluorescence imaging analysis. Processing of kidney tissues were performed by an investigator blinded to the samples.

Immunofluorescence analysis of human, non-human primate tissue donors and mice

nPOD cases: Human pancreatic samples were obtained from the Network of Pancreatic Organ Donors with Diabetes (nPOD) biorepository (see resources table). nPOD fixates the different regions of the pancreas (head, body and tail) in 10% neutral buffered formalin, which are subsequently embedded in blocks of paraffin. From these blocks, slices of 5 µm are cut and placed on a glass slide. The analyzed T1D samples were selected from donors with short-term (0-5 years), medium-term (6-10 years) or long-term (>10 years) diabetes. The demographics of the donors are presented in **Table S1**. *DRI samples:* Human pancreatic samples for immunofluorescence analysis were also obtained from the DRI's cGMP facility (see key resource table). A biopsy from a donor with type 2 diabetes shown in this study was fixated in 10% neutral buffered formalin, and then embedded in paraffin blocks. From these blocks, slices of 5 µm were cut and placed on a glass slide.

Tissue sections were taken from the 4°C storage refrigerator and adjusted to room temperature (RT). Deparaffinization of the slides was performed using an Autostainer XL (Leica Biosystems, Wetzlar, Germany, Cat# ST5010) programmed for 3x 2 minutes dips in xylene, 2x 1 minute dips in absolute ethanol, 1x 30 seconds dip in 95% ethanol, 1x 45 seconds dip in 70% ethanol, and a final wash in dH_20. Antigen retrieval was performed by submerging the slides of pancreatic tissue in nuclear decloaker at pH 9.5 (BioCare Medical, Pacheco, CA, Cat# CB911M), or murine kidney tissue in antigen decloaker at pH 6.0 (BioCare Medical, Pacheco, CA, Cat# CB910M). The slides with the decloaker solutions were then placed inside a Decloaking Chamber Pro (BioCare Medical, Pacheco, CA, Cat# DC2012) for 10 minutes at 120 °C and then cooled down to RT. Slides were then washed 2x for 5 minutes in 1X phosphate buffered saline (PBS), pH 7.4 (Sigma Aldrich, St. Louis, MO, USA Cat# P3813), and 1x for 5 minutes with dH_2O. For permeabilization, we added 0.3% Triton (Sigma Aldrich, St. Louis, MO, USA Cat# T9284-500ml) to 1X PBS, pH 7.4. To apply this buffer, a hydrophobic edge was drawn on the glass slides around the tissue section using the ImmEdge Pen (Vector laboratories Burlingame, CA, Cat# H-4000). Two drops of permeabilization buffer were placed on the tissue section within the edge. After 30 minutes, the permeabilization buffer was replaced by blocking buffer containing [dH_2O, 5% normal donkey serum (Jackson Labs, Bar Harbor, ME, Cat# H-400), 0.1% Triton and 1X power block (Biogenex, San Ramon, CA, Cat# HK0855K)]. After 1 hour, an additional blocking step was done for 10 minutes with serum-free protein block (Dako now Agilent, Santa Clara, CA, Cat# X0909). Primary antibodies were dissolved in the blocking buffer and incubated overnight at 4°C (see key resource table). Primary antibodies were then removed, and sections washed 7x for 5 minutes with 1X PBS, pH 7.4/0.1% Triton. Secondary antibody solutions (1:400) were made with Alexa Fluor 488, 594 or 647 donkey anti-primary antibody species (see key resource table) and 1:400 4′, 6-diamidino-2-phenylindole (DAPI) (Thermo Fisher/Life Technologies, Waltham, MA, Cat# D1306) in blocking buffer. Samples were incubated with secondary antibodies for 90 minutes and then washed 7x for 5 minutes with 1X PBS, pH 7.4. To get the least amount of background fluorescence, an additional wash was performed overnight at 4°C with 1X PBS, pH 7.4. After the last washing step, a drop of Prolong™ Gold Antifade solution (Thermo Fisher, Waltham, MA, Cat# P36930) was added and the sections protected with Micro cover glass (VWR, Radnor, PA Cat # 48393-081) and sealed with clear nail polish. (INM, Anaheim, CA, Cat#S401025).

Sections from the pancreas head, body, and tail were imaged and analyzed. From each section, three images at 20x magnification, for a total of nine per donor (4 channels each), were taken using an ApoTome Axiovert 200M (Zeiss) fluorescent microscope. All images contained at least one large pancreatic duct. For quantification, only the cells of the large pancreatic ducts were counted. We used the cell counter plug-in from ImageJ/FIJI software (NIH) to count cells of specific phenotypes. From every image, DAPI-positive cells in the ducts were counted to determine the total number of cells, against which all the other markers were quantified. Statistical analyses were done using GraphPad Prism v8. We determined four ratios representing the most important populations. The first ratio was the amount of $PDX1^+$ cells versus the amount of $DAPI^+$ cells. The second was that of $ALK3^+/PDX1^+$ cells vs $PDX1^+$ cells. Third and fourth were the ratios of $CAII^+$ and $CAII^-$ cells out of $ALK3^+/PDX1^+$ cells, respectively. All ratios for each anatomical region were also compared separately for every duration of the disease. For pairwise group comparisons, we systematically used a non-parametric statistical test (unpaired, two-tailed Mann-Whitney test). The p value significance threshold was defined as 0.05. Finally, graphs were made with GraphPad Prism v8.

Dead cell removal

This was performed at Genewiz (South Plainfield, NJ, USA). Prior to library preparation, cryopreserved $ALK3^{bright+}$ cells were thawed at 37°C and enriched for live cells using a dead cell removal magnetic cell sorting kit (Miltenyi Biotec, Cat# 130-090-101) according to the manufacturer's recommendations. Viability after dead cell removal was 90±13.3%.

Single cell capture, library preparation and sequencing

This was performed at Genewiz (South Plainfield, NJ, USA). Sorted live $CD90^{-}/ALK3^{bright+}$ cells were captured for library preparation using the Chromium Single Cell 3' Reagent Version 2 Kit (10X Genomics) (2). For library preparation, 6873 (2291±481, per donor) cells from each sample were loaded into their respective 10X chip wells to produce Gel Bead-in-Emulsions (GEMs). GEMs were then subjected to reverse transcription to generate barcoded-RNA prior to cleanup and cDNA amplification. The sequencing libraries were evaluated for quality on the Agilent TapeStation (Agilent Technologies, Palo Alto, CA, USA), and quantified by using a Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA). Pooled libraries were quantified using qPCR (Applied Biosystems, Carlsbad, CA, USA) prior to loading onto an Illumina sequencing platform. Libraries were generated using the Chromium Single Cell 3' reagent Version 2 Kit. Each sample was then sequenced on 1 lane of a HiSeq2500 Sequencer (Illumina) set in Rapid Run Mode with paired-end-sequencing parameters selected: Read1, 98 cycles; Index1, 14 cycles; Index2, 7 cycles; and Read2, 10 cycles. Samples were sequences with a sequencing configuration compatible with recommended guidelines outlined by 10X Genomics.

Computational analysis

A complete compendium of the coding software, coding environment and data-analysis pipeline is available online as an R, Python and Linux script repository on Github: https://github.com/JDBLab/Pancreas_ductal_scRNAseq. These scripts replicate the data analysis outlined in this manuscript.

A. *Alignment and Generation of Filtered Gene counts Using CellRanger*

For all datasets, we utilized CellRanger v3.0.2 software using the {-mkfastq} command in order to de-multiplex FASTQ data. Reads were mapped and aligned to the human genome (10X genomics pre-built GRCh38 reference genome) with STAR (3). Subsequently, final digital gene expression matrices and c-loupe files were generated for downstream multimodal analysis.

B. *Multimodal single-cell Transcriptomic Analysis using Seurat*

i. Selecting appropriate scRNA-seq analysis tool

Prior to selecting a comprehensive single-cell RNA sequencing package for analysis, we went over multiple scRNAseq packages including CALISTA (4), SCRAN (5), Seurat (6-8) and Monocle (9). Although each of these analysis toolkits are excellent, taking into account analytical power, user friendly functioning, de-bugging support and system resource requirements, Seurat (https://CRAN.R-project.org/package=Seurat) was selected as our analytical package of choice (v3.1.1 released to CRAN as of 10/3/2019).

ii. R, RStudio and Seurat versions

Using R v3.5.3 64x (a software environment for statistical computing and graphics: https://cran.r-project.org/bin/windows/base/old/3.5.3/) and RStudio v1.2.1335 64x (an integrated development environment for R: https://www.rstudio.com/products/rstudio/download/), we deployed Seurat v3.1.1 (https://CRAN.R-project.org/package=Seurat) scripts to perform merging, thresholding, normalization, principal component analysis (linear dimensionality reduction), clustering analysis (non-linear multidimensional reduction), visualization and differential gene expression analysis.

iii. Merging $ALK3^{bright+}$ Datasets

All datasets were merged using the {merge} function in Seurat. Care was taken not to normalize datasets, as this would be performed downstream on the entire merged dataset. Seurat v3.0.0 allows for the addition of sample specific IDs to cellular barcodes, which was performed using the {add.cell.ids} function. This allows for easy sample based discrimination of cells during downstream analysis. The {tail}, {head} and {table} functions can be used to query datasets as well as metadata files.

iv. Dataset thresholding, normalization, cluster subsetting, analysis and data visualization.

Cells having total mitochondrial RNA contribution beyond 20% were eliminated from the analysis, along with cells expressing less than 200 or greater than 8000 total genes (doublet elimination). After this, we performed global-normalization (using the {SCTransform} function in Seurat), which is a novel modelling framework allowing for both gene expression normalization and gene expression variance stabilization. We have observed over Seurat's {NormalizeData} function (and other functions available for scRNAseq analysis), SCTransform to be far-superior in its ability to remove cell-cell variation due to confounding technical factors such as total detected RNA molecules. SCTransform is a mathematical model which derives Pearson's residuals from regularized negative bionomial regression values, calculated across gene expression count-data derived from each cell. This method removes technical effects, yet maintains biological heterogeneity. For a detailed understanding of the mathematical design of SCTransform please read Christoph Hafemeister and Rahul Satija's article (10). {SCTransform} also calculates and stores data pertaining to the 3000 most variable genes (SCTransform completely removes the necessity of performing {NormalizeData}, {ScaleData} and {FindVariableFeatures}). It is important to note ONLY these 3000 most variable genes were used in all downstream analysis, if a gene is not deemed 'variable' it does not meet the minimal criteria of detection for further analysis, and are therefore de-classified as noise. Since we did not want to give any one sample more weight over the other, we performed an unbiased cross-sample data-set integration as outlined and recommended by Seurat developers (https://satijalab.org/seurat/v3.1/integration.html, see SCTransform Seurat v3.1) to create an integrated dataset.

We then performed a PCA analysis on the integrated dataset. Using Seurat's non-linear dimensional reduction algorithm {RunUMAP}, we performed Uniform Manifold Approximation and Projection (UMAP) to cluster cells based on variable genes identified previously. It is important to note, non-linear dimensionality reduction via UMAP is far superior in comparison to t-SNE (t-Stochastic Neighbor Embedding), in fact use of the latter is now not recommended by most scRNAseq data-analysis pipeline developers. For more on this debate please see [https://towardsdatascience.com/how-exactly-umap-works-13e3040e1668 and (11)]. Differential gene expression analysis was performed using the {FindAllMarkers} function in Seurat. We used a Wilcoxon rank sum test to perform the analysis and confirmed with another algorithm: Model-based Analysis of Single-cell Transcriptomics (MAST) using the {FindAllMarkers} function on default settings. In our analysis:

1. A differentially expressed gene is defined as any gene which is expressed in at least 10% cells, has a p-value <0.001 and a 1.5x average fold change from all other clusters being tested.
2. Cell clusters were defined to have at least 20 differentially expressed genes for which at least 10 were unique when compared to other clusters.

Dimensional reduction was within recommended parameters outlined by linear dimensionality testing using the Seurat functions {RunPCA}, {JackStraw} and {PCElbowPlot}. Our analyzed cells contained some mesenchymal cells, based on clustering and differentially high expression of *COL1A1*. We subsequently sub-setted mesenchymal cells out of the analysis using Seurat's {SubsetData} function. The sub-setted Seurat object was used for all analysis and graphical representation throughout the manuscript. After initial sub-setting, a re-clustering and PCA analysis was performed to generate clusters for differential gene expression analysis as outlined above. In order to test for cluster separation, and to insure our dimensional reduction was within acceptable parameters, we performed cluster analysis using {clustree}. {clustree} is an algorithm which models multiple dimensionality reductions (dimensionality reductions control for the degree of clustering in the data) which allows for the user to visualize cluster stability as higher resolutions are used for dimensionality reduction. This mathematical modelling allows for optimal clustering of your data, where only stable clusters (maintaining their characteristics over a spectrum of resolutions) are used to cluster data. This allows for a more accurate representation of clusters, without over or under-clustering the data. For a detailed explanation on the mathematical modelling running {clustree} please see (12).

v. Pathway Analysis using EnrichR and GOPlot

Gene ontology analysis was performed using EnrichR (13, 14). GO Biological process output comma-delimited (CSV) files were saved, and all those pathways having an adjusted p-value <0.01 were selected in the analysis. We selected commonly reoccurring and highly ranked GO pathways, highly enriched in genes derived from the differential gene expression analysis. Bar plots of GO pathways were generated using Graphpad Prism v6.01, p-adjusted values were normalized using the following formula: -Log(p-value).

vi. Graphical visualization of scRNAseq data

Seurat has a comprehensive palette of integrated graphical visualization tools derived in part from the R function {ggplot2} (15). We used the {DimPlot} function to visualize UMAP plots, using metadata files to color-code various cells and variable genes to assist in cell clustering. Using the {VlnPlot} function, we generated Violin plots to look at gene expression across clusters, and {DoHeatmap} to generate heatmaps of top most differentially expressed genes.

vii. Merging datasets with the Human Pancreas Atlas

Since its conception, Seurat has evolved into a powerful scRNAseq integrative analysis tool which allows for the assembly of multiple scRNAseq datasets into an integrated reference dataset, against which query datasets can be mapped (to see Seurat's reference based guidelines please see: https://satijalab.org/seurat/v3.1/integration.html). This allows for the dissection of cellular relationships within the context of a comprehensive single-cell atlas. In order to achieve this for our data analysis, we utilized two datasets (GSE85241) (16) and (GSE81076) (17) of human pancreatic islet cells to generate a reference dataset of 3277 single cells. We performed standard data pre-processing: 1) thresholding, 2) normalization (using {SCTransform}), and 3) variable gene calculation (using {SCTransform}). Then, using the {SelectIntegrationFeatures} and the {PrepSCTIntegration} functions, we calculate integral features for downstream analysis. After finding integration features, we select our reference datasets and map our 3x query ALK3 datasets against them. Using the {FindIntegrationAnchors} and {IntegrateData} functions, we create an optimally integrated dataset. Downstream visualizations and the code used for them utilize {DimPlot}, {FeaturePlot}, {VlnPlot} and {DoHeatmap} functions as previously described in Seurat v3.1.1.

viii. Pseudotime calculations and Pseudotemporal Ordering using Monocle

We utilized Monocle v2.10.1 (9, 18, 19) in order to order cells in pseudotime based on their transcriptional similarities, in an unbiased manner. Preprocessed integrated Seurat v3.1.1 objects were integrated into Monocle v2.10.0 utilizing the expressionFamily = uninormal() command line, finally after pseudotime calculations were made for each cell, clusters (information derived from the Seurat object) were projected onto the minimum spanning tree upon cell ordering using the {plot_cell_trajectory} function.

ix. 3-Dimensional UMAP analysis and visualization

We designed a coding script which is compatible with Seurat v3.1.1 objects, capable of transforming 2D UMAP projections into 3D projections (20). Using the Seurat object, we re-ran the {RunUMAP} function but with a n.components = 3L command line to ensure x3 [["UMAP"]]@cell.embeddings to be extracted and utilized for projecting cells in 3D space. This is an extension of the Seurat package, utilizing the (21) engine (21). Since the tSNE calculations are in agreement even with increased dimensionality of the data analysis, 3D tSNE plots derived from our analysis provided further validation of the analysis performed, in addition to higher accuracy achieved in data representation (for a better understanding of 3D UMAP generation please see: https://github.com/Dragonmasterx87/Interactive-3D-Plotting-in-Seurat-3.0.0).

x. Statistical analysis

For graphing data into box plots and pie charts, GraphPad Prism v5 and v8 was utilized. All other graphing was performed using R and the integrated graphing functions {ggplot2} and {RShiny}. Following the Shapiro-Wilk normality test, statistical differences between groups were calculated by two-tailed paired t test or Wilcoxon signed-rank test, with 95% confidence intervals ($*p < 0.05$; $**p < 0.01$; $***p < 0.001$). Results are expressed as mean ± SD. A description of statistical tests and algorithms used for differential gene expression analysis and pathway analysis are provided within each of their descriptions.

RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-BMPR1A-AF647-Rabbit Polyclonal	Bioss	Cat# bs-1509R-A647; RRID: AB_2801504
Anti-BMPR1A-Mouse Polyclonal	LS Biosciences	Cat# LS-C191759; RRID: AB_2801506

Anti-BMPR1A-Rabbit Polyclonal	Abcam	Cat# ab38560; RRID: AB_722713
Anti-P2RY1-Rabbit Polyclonal	LS Biosciences	Cat# LS-A387; RRID: AB_592423
Anti-P2RY1-FITC-Rabbit Polyclonal	Biorbyt	Cat# orb16131; RRID: AB_1073728
Anti-PDX1-Goat Polyclonal	R&D Systems	Cat# AF2419; RRID: AB_355257
Anti-NKX6.1-Goat Polyclonal	R&D Systems	Cat# AF5857; RRID: AB_1857045
Anti-INSULIN-Guinea Pig Polyclonal	Dako/ Agilent	Cat# A0564; RRID: AB_10013624
FLEX Polyclonal Guinea Pig Anti-Insulin, Ready-to-Use antibody	Dako/ Agilent	Cat# IR00261-2, RRID: AB_2800361
Anti-Glucagon-Mouse Polyclonal	R&D Systems	Cat# MAB1249; RRID: AB_2107340
Anti-Glucagon-Rabbit Polyclonal	Dako/ Agilent	Cat# A0565; RRID: AB_10013726
Anti-KRT19-Rabbit Monoclonal	Abcam	Cat# ab52625; RRID: AB_2281020
Anti-KRT19-Mouse Polyclonal	Dako/ Agilent	Cat# M0888; RRID: AB_2234418
Anti-Somatostatin-Rat Polyclonal	Millipore	Cat# MAB354 RRID: AB_2255365
Polyclonal Rabbit Anti-Human Somatostatin antibody	Dako/ Agilent	Cat# A0566, RRID: AB_2688022
Anti-Alpha Amylase-Rabbit Polyclonal	Sigma Aldrich	Cat# A8273; RRID: AB_258380
Anti-Carbonic Anhydrase 2-Sheep Polyclonal	Thermofisher Scientific	Cat# PA5-33167; RRID: AB_2550596
Anti-Carbonic Anhydrase 2-Rabbit Polyclonal	Abcam	Cat# ab191343; RRID: AB_2801505
Rabbit Anti-Chromogranin A Polyclonal Antibody, Unconjugated	Abcam	Cat# ab15160 RRID: AB_301704
Anti-CD90/THY1-PE-Mouse-Monoclonal	R&D Systems	Cat# FAB2067P; RRID: AB_2271796
Anti-CD90/THY1- AF488-Mouse-Monoclonal	R&D Systems	Cat# FAB2067G; RRID: AB_2801507
Alexa Fluor® 594 AffiniPure Donkey Anti-Guinea Pig IgG (H+L)	Jackson Immuno Research laboratories, Inc.	Cat# 706-585-148; RRID: AB_2340474
Alexa Fluor® 647 AffiniPure Donkey Anti-Guinea Pig IgG (H+L)	Jackson Immuno Research laboratories, Inc.	Cat# 706-605-148; RRID: AB_2340476
Alexa Fluor Donkey polyclonal anti-rabbit IgG, Alexa Fluor 488	Thermofisher Scientific	Cat# A-21206; RRID: AB_2535792
Donkey Anti-Sheep IgG (H+L) Antibody, Alexa Fluor 488 Conjugated	Thermofisher Scientific	Cat# A-11015, RRID: AB_141362
Donkey Anti-Rabbit IgG (H+L) Polyclonal Antibody, Alexa Fluor 647 Conjugated	Thermofisher Scientific	Cat# A-31573, RRID: AB_2536183
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 594	Thermofisher Scientific	Cat# A-21203, RRID: AB_2535789

Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 647	Thermofisher Scientific	Cat# A-21447, RRID: AB_2535864
Donkey anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 594	Thermofisher Scientific	Cat# A-21207, RRID: AB_141637
Donkey anti-Goat IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	Thermofisher Scientific	Cat# A-11055 RRID: AB_2534102
Donkey anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 647	Thermofisher Scientific	Cat# A-31571 RRID: AB_162542
Biological Samples		
Human Pancreas Exocrine Tissue and Sections	cGMP Facility, Diabetes Research Institute, Miami, FL	http://www.diabetesresearch.org/cGMP-GTP-cell-processing
Human Pancreas Exocrine Tissue	Institute of Cellular Therapeutics, Pittsburgh, PA	https://www.ahn.org/research/patient-information/cellular-therapeutics
Human Pancreas Exocrine Tissue	Prodo Labs, Aliso Viejo, CA	https://prodolabs.com/
Human Pancreas Sections	nPOD, University of Florida, Gainesville, FL	https://www.jdrfnpod.org
Non-Human Primate Pancreas Tissue Biopsy	cGMP Facility, Diabetes Research Institute, Miami, FL	https://www.diabetesresearch.org/
Chemicals, Peptides, and Recombinant Proteins		
Trypsin inhibitor isolated from *Glycine max*	Sigma Aldrich	Cat# T6522
DPBS	Sigma Aldrich	Cat# D8537
Fetal bovine serum	Thermofisher Scientific	Cat# 10082147
RPMI 1640 containing GlutaMAX	Thermofisher Scientific	Cat# 61870036
HEPES buffer	Thermofisher Scientific	Cat# 15630080
Aprotinin	Sigma Aldrich	Cat# A6106; CAS# 9087-70-1
Sodium Pyruvate	Thermofisher Scientific	Cat# 11360070
B27-minus insulin	Thermofisher Scientific	Cat# A1895601
Tryple-E Express Enzyme	Thermofisher Scientific	Cat# 12604013
Corning® Matrigel® hESC-Qualified Matrix, *LDEV-free	Corning	Cat# 354277
CryoStor® CS10	StemCell Technologies	Cat# 07930
DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride)	Thermofisher Scientific	Cat# 1306
ProLong™ Gold Antifade Mountant	Thermofisher Scientific	Cat#: P36930
Nuclear Decloaker, 10X	BioCare Medical	Cat # CB911M
Antigen Decloaker, 10X	BioCare Medical	Cat # CB910M
Protein Block, Serum-Free	Dako/ Agilent	Cat # X090930-2
ImmEdge Pen	Vector Laboratories	Cat# H-4000, RRID: AB_2336517
Normal Donkey Serum	Jackson Immuno Research laboratories, Inc.	Cat# 017-000-121 RRID: AB_2337258

Out the Door-Top Coat	International Nail Manufacturers, (inm)	Cat#S401025
Critical Commercial Assays		
LIVE/DEAD™ Fixable Near-IR Dead Cell Stain Kit, for 633 or 635 nm excitation	Invitrogen	Cat# L34975
Contour Next Ez Blood Glucose Monitoring Kit	Bayer	Cat# 9628; UPC# 301939628014
Mercodia Ultrasensitive C-peptide ELISA	Mercodia	Cat# 10-1141-01
Dead-cell Removal kit	Miltenyi Biotec	Cat# 130-090-101
Chromium Single Cell 3' Reagent Kits v2	10X Genomics	Cat# CG00052
Deposited Data		
Raw and analyzed data	This paper	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131886
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/grc/human
A single-cell transcriptome atlas of the human pancreas (GSE81076)	Gene Expression Omnibus	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE81076
A single-cell transcriptome atlas of the human pancreas [CEL-seq2] (GSE85241)	Gene Expression Omnibus	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE85241
Experimental Models: Organisms/Strains		
Mouse: Foxn1^{nu} athymic nude Female NU/J, Wks. Age: 6 weeks Homozygous for Foxn1^{nu} Male NU/J, Wks. Age: 6 weeks us Homozygous for Foxn1^{nu} Mouse: Foxn1^{nu}	The Jackson Laboratory	Cat# 002019; MGI: 1856108; https://www.jax.org/strain/002019
Software and Algorithms		
Complete coding and analysis compendium	This paper	https://github.com/JDBLab/Pancreas_ductal_scRNAseq
R v3.5.3 and R v3.5.1 (64x bit, for Windows)	The R Consortium	https://cran.r-project.org/bin/windows/base/old/
RStudio v1.2.1335 (64x bit, for Windows)	The R Consortium	https://www.rstudio.com/products/rstudio/download/#download
Cellranger v3.0.2	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/overview/welcome
Seurat v3.1.1	(6, 7)	https://satijalab.org/seurat/
Monocle v2.10.0	(9, 19)	http://cole-trapnell-lab.github.io/monocle-release/
3D Plotting of scRNAseq data using Seurat objects v1.3.0	(20)	https://github.com/Dragonmasterx87/Interactive-3D-Plotting-in-Seurat-3.0.0
EnrichR	(13)	https://amp.pharm.mssm.edu/Enrichr/
Python v3.5.7	(22)	https://www.python.org/downloads/
Fiji ImageJ	(23)	https://fiji.sc/

GraphPad Prism v8	GraphPad	https://www.graphpad.com/scientific-software/prism/
Summit software v6.3	Beckman Coulter	https://www.beckman.com/flow-cytometry
Kaluza v1.5a and v2.1.1	Beckman Coulter	https://www.beckman.com/flow-cytometry/software/kaluza
CentOS 6.5 (64x bit, for Windows)	linux	http://vault.centos.org/6.5/isos/x86_64/
Windows 10 Professional (64x bit)	Microsoft	https://www.microsoft.com/en-us/p/windows-10-pro/df77x4d43rkt?activetab=pivot%3aoverviewtab
FileZilla v3.42.1 (For windows)	FileZilla	https://filezilla-project.org/
AxioVision v4.6 (32x bit, for Windows)	Zeiss	https://www.micro-shop.zeiss.com/en/us/system/software+axio+vision-axiovision+program-axiovision+software/10221/
Leica Application Suite (LAS) v5	Leica Microsystems	https://www.leica-microsystems.com/products/microscope-software/p/leica-application-suite/
Other		
Sequencing QC	This paper	https://github.com/JDBLab/Pancreas_ductal_scRNAseq/tree/master/sequencing_qc
Integrated alk3n3/Pancreas Single Cell Gene expression Atlas 3D tSNE Visualization web browser	This paper	https://github.com/JDBLab/Pancreas_ductal_scRNAseq/tree/master/3d_tsne_html
Differentially expressed genes	This paper	Supplementary File
Gene Ontology Pathway Analysis	This paper	Supplementary File
Donor Demographics	This paper	Table S1
Human Protein Atlas version 19	The Human Protein Atlas	www.proteinatlas.org
Figure S3 – FOS immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000170345-FOS/tissue/pancreas#img
Figure S3 – TSPAN8 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000127324-TSPAN8/tissue/pancreas#img
Figure S3 – SPP1 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000118785-SPP1/tissue/pancreas#img
Figure S3 – HMGA1 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000137309-HMGA1/tissue/pancreas#img
Figure S3 – S100A6 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000197956-S100A6/tissue/pancreas#img
Figure S4 – KRT17 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000128422-KRT17/tissue/pancreas#img

Figure S4 – ANXA1 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000135046-ANXA1/tissue/pancreas#img
Figure S4 – ANXA3 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000138772-ANXA3/tissue/pancreas#img
Figure S4 – KRT23 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000108244-KRT23/tissue/pancreas#img
Figure S4 – LAMA3 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000053747-LAMA3/tissue/pancreas#img
Figure S4 – ELF3 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000163435-ELF3/tissue/pancreas#img
Figure S4 – VMP1 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000062716-VMP1/tissue/pancreas#img
Figure S4 – WSB1 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000109046-WSB1/tissue/pancreas#img
Figure S4 – MYO6 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000196586-MYO6/tissue/pancreas#img
Figure S4 – CLPS immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000137392-CLPS/tissue/pancreas#img
Figure S4 – CELA3A immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000142789-CELA3A/tissue/pancreas#img
Figure S4 – OLFM4 immunostaining	The Human Protein Atlas	https://www.proteinatlas.org/ENSG00000102837-OLFM4/tissue/pancreas#img

References:

1. Ricordi C, Lacy PE, Finke EH, Olack BJ, & Scharp DW (1988) Automated method for isolation of human pancreatic islets. *Diabetes* 37(4):413-420.
2. Zheng GX*, et al.* (2017) Massively parallel digital transcriptional profiling of single cells. *Nature communications* 8:14049.
3. Dobin A*, et al.* (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* 29(1):15-21.
4. Papili Gao N, Hartmann T, Fang T, & Gunawan R (2019) CALISTA: Clustering And Lineage Inference in Single-Cell Transcriptional Analysis. *bioRxiv*:257550.
5. Lun AT, McCarthy DJ, & Marioni JC (2016) A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. *F1000Research* 5:2122.
6. Stuart T*, et al.* (2018) Comprehensive integration of single cell data. *bioRxiv*:460147.
7. Butler A, Hoffman P, Smibert P, Papalexi E, & Satija R (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature biotechnology* 36(5):411-420.
8. Macosko EZ*, et al.* (2015) Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161(5):1202-1214.
9. Trapnell C*, et al.* (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature biotechnology* 32(4):381-386.
10. Hafemeister C & Satija R (2019) Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *bioRxiv*:576827.

11. McInnes L, Healy J, & Melville J (2018) Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.
12. Zappia L & Oshlack A (2018) Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *GigaScience* 7(7):giy083.
13. Chen EY*, et al.* (2013) Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics* 14:128.
14. Kuleshov MV*, et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic acids research* 44(W1):W90-97.
15. Wickham H (2016) *ggplot2: elegant graphics for data analysis* (Springer).
16. Muraro MJ*, et al.* (2016) A single-cell transcriptome atlas of the human pancreas. *Cell systems* 3(4):385-394. e383.
17. Grun D*, et al.* (2016) De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. *Cell stem cell* 19(2):266-277.
18. Qiu X*, et al.* (2017) Single-cell mRNA quantification and differential analysis with Census. *Nature methods* 14(3):309-315.
19. Qiu X*, et al.* (2017) Reversed graph embedding resolves complex single-cell trajectories. *Nature methods* 14(10):979-982.
20. Qadir MMF, Sadiq S, & Dominguez-Bendala J (2019) 3D Plotting of scRNAseq data using Seurat objects.
21. Inc. PT (2015) Collaborative data science.
22. Sanner MF (1999) Python: a programming language for software integration and development. *J Mol Graph Model* 17(1):57-61.
23. Schindelin J*, et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nature methods* 9(7):676.

C. INDEX OF SUPPLEMENTAL FILES

a. Dataset S1. DE genes

b. Dataset S2. GO pathways

c. Movie S1. 3D plot showing alignment of the ALK3$^{bright+}$ clusters against a scRNAseq dataset of the whole pancreas, confirming the ducto-acinar transitional axis and suggesting potential association of the pro-ductal and endocrine compartments. 3D UMAP plot showing cellular distribution when the ALK3$^{bright+}$ dataset (n=3, 4,878 cells) is mapped against two combined islet datasets claimed to represent the whole pancreas (3,277 cells). Clustering of pancreatic subtypes reflects differential gene expression and was validated by visualizing gene expression in UMAP plots.