SUPPLEMENTARY MATERIALS

Supplemental Tables

Supplementary Table 1. List of antibodies used in this study.

Antigen	Host	Concentration	Manufacturer	Catalog #	Notes
Amylase	Mouse	1:100	Santa Cruz	sc-46657	
aPKC	Rabbit	1:100	Santa Cruz	sc-216	
B-catenin	Goat	1:100	Santa Cruz	sc-1496	
Cdc42	Rabbit	1:100	Selleckchem	F1079	used on cryosections
Clathrin	Rabbit	1:100	Cell Signaling	4796S	used on cryosections
Cleaved Caspase 3	Rabbit	1:100	Cell Signaling	9661S	
CPA1	Goat	1:100	R&D	AF2765	
Crb3	Rabbit	1:100	Sigma	HPA013835	used on cryosections
DBA	biotin	1:100	Vector labs	B-1035	
E-Cadherin	Rabbit	1:100	BiCell Scientific	00101	
E-Cadherin	Rat	1:50	ThermoFisher	13-1900	
E-Cadherin	Mouse	1:100	BD Transduction	610182	
Ezrin	Rabbit	1:100	Millipore	07-130	used on cryosections
Glucagon	Rabbit	1:100	Cell Signaling	2760S	
Glucagon	Mouse	1:100	Sigma	62654	
Golgi	Mouse	1:100	BD Transduction	610822	
Golgi	Rabbit	1:100	Abcam	Ab52649	
Insulin	Rabbit	1:100	Cell Signaling	3014S	
Laminin	Rabbit	1:100	Sigma	L9393	
Merlin	Mouse	1:50	Santa Cruz	sc-55575	
Merlin	Rabbit	1:100	Millipore Sigma	HPA003097	
Mucin1	Armenian Hamster	1:250	ThermoFisher	MA5-11202	
Ngn3	Mouse	1:100	DSHB	F25A1B3-s	
Par3	Rabbit	1:100	Millipore	07-330	
Par6b	Rabbit	1:100	Santa Cruz	Sc67392	
Pdx1	Goat	1:100	Santa Cruz	sc14664	
phospho Histone H3	Rabbit	1:100	Millipore	06-570	
pPI3K (p85 (y458)/p55 (y199))	Rabbit	1:100	Cell Signaling	4228S	used on cryosections
Rab11A	Rabbit	1:50	US Biologics	R0009-05	used on cryosections
Rab5	Rabbit	1:100	Cell Signaling	3547S	used on cryosections
RFP	Rabbit	1:100	Rockland	600-401-379	
Sox9	Rabbit	1:250	Millipore	AB5535	
YAP1	Mouse	1:100	Abcam	ab56701	
ZO-1	Mouse	1:100	Invitrogen	33-9100	
Anti rabbit 555	Donkey	1:500	Invitrogen	A31572	
Anti mouse 555	Donkey	1:500	Invitrogen	A31570	
Anti mouse 488	Donkey	1:500	Invitrogen	A21202	
Anti goat 647	Donkey	1:500	Invitrogen	A32849	
Anit hamster 647	Goat	1:500	Invitrogen	A78967	

Supplementary Table 2. qPCR primers

Gene Name	**Sequence**
Gapdh fwd	AAGGTCATCCCAGAGCTGAA
Gapdh rev	AGGAGACAACCTGGTCCTCA
Erbb4 fwd	TGGCATTCATGGAGACCCTT
Erbb4 rev	GGGGCCAAGACTGTATGTTC
Glis3 fwd	ATTCAGCTCCCATTTTCGCC
Glis3 rev	GTAACTGGGAGGAGGGGTTG
Shroom3 fwd	AGATGACTTTCCTCCGCCTC
Shroom3 rev	ATCACCCTGGGAATTCTGCA

Extended Methods

Image Quantification

Pancreatic lineage volume quantifications: Whole-mount immunofluorescence images of pancreata stained for MUC1 and lineage markers (endocrine, CPA1, or DBA) were acquired and analyzed using IMARIS (Bitplane). For each sample, a region of interest (ROI) encompassing the pancreatic epithelium was manually defined based on MUC1 or CDH1 signal. For endocrine and CPA1 quantifications, a three-dimensional MUC1 surface was generated within the ROI to define total luminal volume. For DBA quantifications, a CDH1 surface was generated to define total epithelial volume. Lineage marker-positive volumes were quantified by generating a second surface within the same ROI using a manually adjusted intensity threshold optimized for each sample to accurately delineate tissue morphology while minimizing background signal. Surface volume measurements were recorded and exported to Excel. Lineage-positive volumes were normalized to total luminal volume (MUC1) or total epithelial volume (CDH1) for each pancreas. Normalized values were analyzed and plotted using GraphPad Prism.

Plexus quantitative assessment: To gain quantitative insight into these defects, we utilized the Skeletonize plugin in ImageJ (Fiji) as previously described.32 Confocal image stacks were first processed with a Gaussian blur (sigma = 2.0) and then converted into binary masks using Huang thresholding. Due to differences in staining intensity between control and mutant samples, the threshold was manually adjusted for each image to accurately capture epithelial morphology while minimizing background signal. Skeletonization was performed using the "Skeletonize (2D/3D)" plugin. Network morphology was quantified using the "Analyze Skeleton" plugin to extract parameters including branch number, junctions, endpoints, and total branch length. Plexus junctions were defined as the sum of triple and quadruple branch points detected per pancreas. For P1 data, as a measure of cyst formation, we quantified the number of junction voxels was summed for each pancreas and divided by the total number of junctions.

Cell death: Images were analyzed as maximum-intensity projections CC3+ epithelial cells were defined as CDH1+ cells containing a discrete CC3 signal. CC3+ cells lacking CDH1 were excluded. Apoptotic indices were calculated by normalizing CC3+ cells to total CDH1+ cells. FOVs from *Nf2*^{PancKO} were selected such that each FOV contained regions of dysmorphic epithelium (cysts present, stratified epithelium).

Cell proliferation quantification: E14.5 and P1 sections were stained for CDH1, AMY, and phospho-histone H3 (pHH3). Maximum-intensity projections were analyzed. Proliferating epithelial cells (pHH3+) were normalized to total CDH1+ cells. FOVs from *Nf2*^{PancKO} were selected such that each FOV contained regions of dysmorphic epithelium

Adult islet characterization and surface area analysis: Adult sections stained for CDH1, glucagon, and insulin were imaged at 20× or 40× and analyzed as Z-stacks or maximum projections. Islets were manually classified as mixed, mantled, glucagon-only, or insulin-only. Islet surface area was measured in ImageJ using the freehand selection tool.

Adult islet cell quantification: Adult pancreatic sections were immunostained for E-cadherin, insulin, and either somatostatin (SS) or pancreatic polypeptide (PP) and imaged at 40× magnification. Individual islets within each field of view were analyzed as z-stacks in ImageJ. The number of SS+ or PP+ cells was quantified for each islet and normalized to the number of insulin-positive (Ins+) cells.

Amylase-positive epithelial cell quantification: Maximum-intensity projections were generated for all images prior to analysis. For each field of view (FOV), the total number of epithelial cells and AMY+

cells was quantified, and amylase epithelial density was calculated as the ratio of Amylase+ cells to total epithelial cells.

NGN3 quantification: E13.5 sections were stained for CDH1 and NGN3 and imaged at 40× magnification. NGN3+ cells were quantified and normalized to the total number of CDH1–positive epithelial cells per field of view.

Stratification Analysis (E14.5 pancreatic tissue): Light sheet microscopy data was utilized to assess stratification of E14.5 pancreatic tissue. A single optical section from the middle of the z stack was assessed, and the maximum number of cell layers between the basal (LAMA+) and apical (MUC1+) markers was quantified. Between 2-4 FOVs were analyzed per pancreata.

Stratification Analysis (Explant): Stratification in explants was performed as previously described.[1] In brief, two metrics were quantified – the number of cell layers from the edge of the pancreas to the nearest lumens, and the number of cell layers between lumens. A single optical section from the middle of each pancreas was analyzed, and ten measurements were taken per analysis.

Explant morphology: For explant area and clefting analysis, maximum-intensity projections were used. For area measurements, the epithelial region (myrTom or CHD1-positive signal) of each explant was manually outlined, and total area was quantified in FIJI (ImageJ). Clefts were defined as clear invaginations or indentations of the epithelial boundary extending inward from the outer epithelial contour, visible in maximum-intensity projections. The number of clefts was quantified per explant.

Number of microlumens: MUC1 staining was performed on E11.5 and 12.5 pancreata, and confocal z-stacks through the entire dorsal bud were obtained. The number of disconnected, isolated MUC1 puncta were manually counted by scrolling through the stack.

Mosaic and *Nf2*^{PancKO} cell fate analysis: Confocal z-stacks were obtained through thick vibratome sections for each genotype. Each myrTom positive cell was scored as either positive of CPA1, DBA, or neither based on overlap of myrTom signal and other markers. Cell counter plugin in FIJI was utilized to score each cell. Data was analyzed in GraphPad. For both mosaic and full deletion, heterozygotes (*ptf1aCreERT2,Nf2*$^{f/+}$, *myrTom* $^{f/+or\ ff}$ and *pdx1Cre, Nf2*$^{f/+}$, *myrTom* $^{f/+or\ ff}$, respectively) were utilized as controls.

Intracellular myrTom puncta quantification**:** Confocal z-stacks were acquired and converted into maximum intensity projections (MIPs) for analysis. Images were processed in FIJI (ImageJ) by applying background subtraction using a rolling ball radius of 20 pixels. Processed images were then thresholded using the Otsu algorithm to generate binary masks of intracellular myrTom signal, and watershedding was applied.
myrTom-positive puncta were quantified using the “Analyze Particles” function in FIJI, with consistent size and circularity (>0.25) parameters applied across all samples. The number of puncta per cell was recorded and used as a measure of intracellular myrTom accumulation. Quantitative data were exported to GraphPad Prism for statistical analysis.

Intracellular Rab5/Rab11 puncta quantification (cells)**:** Confocal z-stacks were acquired and converted into maximum intensity projections (MIPs) for analysis. Images were processed in FIJI (ImageJ) by applying background subtraction using a rolling ball radius of 10 pixels. Processed images were then thresholded using the Otsu algorithm to generate binary masks of intracellular Rab5/Rab11. Puncta were quantified using the “Analyze Particles” function in FIJI, with consistent size (>0.1um) parameters applied across all samples. The number of puncta per cell was recorded and exported to GraphPad Prism for statistical analysis.

Rab5/Rab11 total fluorescence intensity quantification (pancreatic tissue): A single optical section per z-stack, corresponding to the plane with maximal CDH1 signal, was selected for analysis. Individual $CDH1^+$ epithelial cells were manually outlined, and mean gray value of Rab5 or Rab11 fluorescence within each cell was measured in FIJI (ImageJ).

Rab5/Rab11 apical/basal intensity ratio (pancreatic tissue): A single optical section per z-stack corresponding to the plane with maximal CDH1 signal was selected for analysis. $CDH1^+$ epithelial cells lining $MUC1^+$ lumens, defined by a clearly identifiable apical MUC1-positive surface, were analyzed. The apical third of each cell was manually delineated, and mean gray value of the relevant channel was measured in apical and basal regions. The apical-to-basal intensity ratio was calculated for each cell.

Single-nucleus RNA-seq data analysis

Preprocessing and quality control
Raw gene-by-barcode count matrices generated by 10x Genomics Cell Ranger in HDF5 format were imported into R using DropletUtils. [2] Gene identifiers were mapped to unique feature names, and cell barcodes were assigned as column identifiers. Unless otherwise noted, preprocessing and downstream analyses were performed using Bioconductor workflows, including scuttle, scater, scran, and bluster.[3,4]

Sequencing reads were processed using Cell Ranger v5.0.1 and aligned to the mm10-2020-A mouse reference genome. Computational analyses were performed using R v4.5.3, Bioconductor v3.21, Python v3.10.18, Scanpy v1.11.5, and scVelo v0.3.3.

Droplet quality was assessed using barcode rank distributions with barcodeRanks. A lower UMI threshold of 1,024 was used to distinguish cell-containing droplets from background. Per-cell quality control metrics were computed using addPerCellQC, including total UMI counts, number of detected genes, and the proportion of mitochondrial transcripts. Mitochondrial genes were identified from Ensembl annotation release 101. Cells were flagged as outliers using median absolute deviation-based criteria for low library size, low feature detection, and high mitochondrial content, with a minimum difference threshold of 0.5 applied to mitochondrial proportions. Cells failing any criterion were excluded.

Putative doublets were identified using scDblFinder with cluster-based information.[5] In datasets with discrete doublet-enriched clusters, those clusters were removed. In other datasets, clusters with greater than 50% predicted doublets were excluded.

Normalization, feature selection, and dimensionality reduction
Normalization was performed using deconvolution-based size factor estimation in scran. [4] Cells were pre-clustered using quickCluster, size factors were estimated using computeSumFactors with a minimum mean expression threshold of 0.1, and log-transformed normalized expression values were calculated using logNormCounts.

Highly variable genes were identified using a Poisson-based variance model with modelGeneVarByPoisson. Either the top 10% of genes ranked by biological variance or a fixed set of 4,096 highly variable genes was retained, depending on the analysis. Principal component analysis was performed on highly variable genes using scater.[3]

Low-dimensional embeddings for visualization were generated using UMAP, [6] typically with 8 neighbors, 3 components, minimum distance of 0.0, and spread of 1.5. For some analyses, reference-based UMAP was used, in which the embedding was learned on a reference population and query cells were projected into the same coordinate space.

Clustering and cell type annotation

Cells were clustered using graph-based approaches. Shared nearest neighbor graphs were constructed from PCA or diffusion map embeddings, followed by community detection using Louvain or Walktrap algorithms. In selected analyses, density-based clustering with HDBSCAN was applied to UMAP embeddings.[7]

Clusters were annotated post hoc based on differential marker gene expression, established pancreatic lineage markers, and comparison with expected embryonic pancreatic cell states. Both full and abbreviated cell type labels were assigned and used for downstream analyses.

Diffusion map analysis and MAGIC-based imputation

Continuous transcriptional structure was analyzed using diffusion maps.[8,9] Diffusion maps were computed using a custom R wrapper, calculateDiffusion, implementing the Coifman diffusion map construction. Pairwise distances were computed in PCA space, converted to a locally scaled affinity kernel, sparsified by retaining extended nearest-neighbor neighborhoods, and normalized using anisotropic density correction. Diffusion coordinates were obtained from the generalized eigenvalue problem defined by the graph Laplacian and degree matrix.

For some analyses, calculateDiffusion also generated diffusion-smoothed expression values using an implementation conceptually based on MAGIC.[10] A Markov transition matrix was iteratively applied to log-transformed expression values until convergence criteria were met or a maximum of 8 diffusion steps was reached. Imputed values were rescaled by gene-specific 99th-percentile expression. Imputed expression values were used for visualization, AUCell gene set activity scoring, and SCENIC regulon activity analysis, where diffusion smoothing improved detection of coordinated low-abundance transcriptional programs. Differential expression testing, marker gene identification, trajectory inference, and trajectory-based differential expression analyses were performed exclusively using observed expression values.

Marker gene identification and gene set analysis

Cluster-specific marker genes were identified using EIGEN, an ensemble marker detection framework that combines Welch's t-tests, Wilcoxon rank-sum tests, binomial detection tests, and gene set–based enrichment statistics into a consensus ranking using a Borda-style vote aggregation strategy. [11] For computational efficiency, analyses were limited to a maximum of 8,192 cells when required. Gene set enrichment analysis was performed using MSigDB collections, including KEGG, Reactome, WikiPathways, BioCarta, transcription factor target sets from GTRD, and Gene Ontology biological process terms.[12,13] Gene sets containing fewer than 15 genes were excluded.

Per-cell gene set activity was quantified using AUCell applied to diffusion-imputed expression values. [14] Gene set activity scores were summarized at the cluster or cell type level using eigenGeneSets, which applies the EIGEN vote-based marker-ranking framework to AUCell score matrices rather than gene expression matrices. Regulatory network activity was assessed using pySCENIC. [14,15] Gene regulatory networks were inferred de novo using GRNBoost2, followed by motif enrichment–based pruning to identify regulons. Regulon activity was quantified on a per-cell basis using AUCell applied to diffusion-imputed expression values and subsequently summarized at the cluster and cell type levels.

Differential expression analysis

Differential expression analyses were performed using the MAST framework [16]. Genes expressed in fewer than 5% of cells within the tested population were excluded. For cell type-restricted analyses,

hurdle models included condition as the primary explanatory variable. For analyses performed across multiple cell types, models included terms for cell type, condition, and their interaction. Differential expression was assessed using Wald or likelihood ratio tests, and p-values were adjusted using the Benjamini-Hochberg false discovery rate procedure. Cell type-restricted analyses were performed by subsetting specific populations prior to model fitting.

RNA velocity analysis

RNA velocity was estimated independently for each dataset. Spliced and unspliced count matrices were generated from aligned BAM files using velocyto [17] and analyzed with scVelo [18]. Genes were filtered using a minimum shared-count threshold of 10. Transcriptional dynamics were inferred using the dynamical model implemented in scVelo. Velocity vectors and transition graphs were computed and projected onto low-dimensional embeddings, including diffusion map space.

Time-embedded diffusion integration across developmental stages

To integrate wild-type E11.5 and E14.5 nuclei across developmental time, a custom time-embedded diffusion framework was implemented in Python. Highly variable genes were selected using the Seurat v3 method with developmental stage treated as a batch variable, and PCA was performed in a shared feature space [19].

Pairwise distances were computed in PCA space. Within-stage distances were retained for each developmental stage, whereas between-stage distances were computed only for adjacent developmental timepoints, yielding a block-structured distance matrix that enforces temporal locality. Restricting inter-stage connections to adjacent developmental stages prevents biologically implausible shortcuts between distant developmental states while preserving continuous developmental trajectories and maintaining the temporal ordering of developmental progression. Distances were converted to an anisotropic diffusion kernel using locally adaptive bandwidths estimated from k-nearest neighbors, with k = 8 and epsilon = 1.0. The graph was sparsified using separate neighbor budgets for within-stage and between-stage edges, typically k_intra = 64 and k_inter = 32. After alpha-normalization to reduce sampling density bias, the resulting kernel was row-normalized to define a Markov diffusion operator.

Diffusion components were obtained from the generalized eigenvalue problem defined by the graph Laplacian and degree matrix. The resulting 50-component time-embedded diffusion map represents a temporally constrained developmental manifold that preserves local transcriptional similarity while preventing spurious long-range connections between non-adjacent stages.

Reference-based manifold projection

To compare mutant E14.5 nuclei with wild-type developmental structure, mutant cells were projected into a wild-type reference diffusion manifold using geometric harmonics [20]. Wild-type cells were used to define the reference manifold to avoid distortion of developmental trajectories by mutant-specific transcriptional programs.

A diffusion map was constructed using wild-type cells in PCA space with top_k = 32, alpha = 1.0, t = 1, and 50 nontrivial diffusion components. A ridge-regularized geometric harmonics model was then fit to map PCA coordinates to wild-type diffusion coordinates, using tau = 10^-3 and an automatically estimated kernel bandwidth. Model fit was assessed by reconstructing wild-type diffusion coordinates from the fitted geometric harmonics model and calculating root mean squared reconstruction error prior to projection of mutant cells. Mutant cells were then projected into the wild-type diffusion coordinate system as out-of-sample extensions.

Integration using mutual nearest neighbors

As a complementary integration strategy, datasets were integrated using mutual nearest neighbors with fastMNN from batchelor [21]. Integration was performed using 50 dimensions and k = 8 neighbors. The corrected embedding was used for UMAP visualization and clustering, providing an independent comparison to diffusion-based integration.

Trajectory inference and pseudotime analysis

Developmental trajectories were inferred using slingshot on diffusion-based embeddings.[22] Lineages were defined using cluster labels, with specified root and terminal states, and simultaneous principal curves were fit to model developmental trajectories.
Mutant cells were projected onto wild-type trajectories by nearest-point projection in diffusion space. In integrated analyses, generalized additive models were used to map diffusion coordinates to pseudotime, enabling consistent pseudotime assignment across conditions. Lineage assignment was quantified using both hard assignments based on minimum distance to lineage curves and soft probabilistic assignments derived from distance-based weights.

Trajectory-based differential expression and gene dynamics
Gene expression dynamics along developmental trajectories were analyzed using tradeSeq. [23] Generalized additive models were fit as smooth functions of pseudotime, typically using 6 knots. Genes associated with trajectory progression were identified using association tests, and differences between early and late trajectory states were assessed using start-versus-end tests.
Genes were grouped into modules based on shared expression dynamics using clusterExpressionPatterns, followed by hierarchical merging of related modules. Temporal activation patterns were further analyzed by identifying points of maximal expression change along pseudotime. Stage-specific comparisons were performed by identifying representative pseudotime positions, including density peaks within each stage, and testing for differential expression between these positions.

Fate probability and compositional analysis
Lineage commitment was quantified by calculating distances from cells to lineage curves in diffusion space. Probabilistic lineage assignments were derived using distance-based weighting, and entropy was used to quantify uncertainty in fate assignment.
Cell type composition was analyzed by aggregating nuclei across developmental stage, condition, and annotated cell type. Both absolute counts and relative abundances were calculated.

Code Availability
Custom software used for diffusion map analysis, time-embedded diffusion integration, geometric harmonics projection, EIGEN marker identification, and downstream analyses is available at: https://github.com/cpchaney/pdx1-cre_nf2

Supplemental Movie Legends.

Supplemental Movie 1: Control E14.5 pancreata, stained with CDH1 (cyan), MUC1 (yellow), and laminin (magenta). Movie shows sequential optical sections from z-stack acquired on a light sheet microscope.

Supplemental Movie 2: *Nf2*^{PancKO} E14.5 pancreata, stained with CDH1 (cyan), MUC1 (yellow), and laminin (magenta). Movie shows sequential optical sections from z-stack acquired on a light sheet microscope.

Supplemental Movie 3: Projection of control E14.5 pancreata, stained with CDH1 (gray), MUC1 (blue), and laminin (green). Movie begins on a single optical section through the light sheet, red outline indicates a cell as identified by the IMARIS cell function. Cell is in direct contact with MUC1+ lumen.

Supplemental Movie 4: Projection of *Nf2*^{PancKO} E14.5 pancreata, stained with CDH1 (gray), MUC1 (blue), and laminin (green). Movie begins on a single optical section through the light sheet, green and purple outline indicates cells as identified by the IMARIS cell function. Green cell does not in contact a lumen

Supplemental Movie 5: 3D projection of E14.5 control luminal plexus, marked by MUC1 WMIF.

Supplemental Movie 6: 3D projection of E14.5 *Nf2*^{PancKO} luminal plexus, marked by MUC1 WMIF.

Supplemental Movie 7: Live imaging of a control pancreatic explant expressing the *Crb3*^{GFP} reporter, acquired at 1 frame every 10 minutes. The movie is shown as a maximum intensity projection.

Supplemental Movie 8: Live imaging of a *Nf2*^{PancKO} pancreatic explant expressing the *Crb3*^{GFP} reporter, acquired at 1 frame every 10 minutes. The movie is shown as a maximum intensity projection.

Supplemental Movie 9: 3D projection of a *Nf2*^{PancKO} pancreatic explant cultured for 72 hours. *Nf2*^{pancKO} pancreata containing *myrTom* (magenta) and *Crb3*^{GFP} (cyan) reporters showing a large vesicle observed within an epithelial cell.

Supplemental Movie 10: 3D projection of e14.5 *Nf2*^{PancKO} pancreatic section (15μm), stained for CDH1 (green) and MUC1 (magenta). Note large intracellular vesicle.

Supplemental Movie 11: Live imaging of a control pancreatic explant expressing the *Crb3*^{GFP} and *myrTom* reporters, acquired at 1 frame every 25 seconds over a period of 10 minutes.

Supplemental Movie 12: Live imaging of a *Nf2*^{PancKO} pancreatic explant expressing the *Crb3*^{GFP} and *myrTom* reporters, acquired at 1 frame every 25 seconds over a period of 10 minutes.

Supplemental Movie 13: Live imaging of a control pancreatic explant expressing the *Crb3*^{GFP} and *myrTom* reporters, acquired at 1 frame every 25 seconds. Green boxes indicate mature lumens, and red boxes indicate nascent lumens.

Supplemental Movie 14: Live imaging of a control pancreatic explant expressing the *Crb3*^{GFP}, acquired at 1 frame every 25 seconds.

Supplemental Movie 15: Control P1 pancreata, stained with MUC1 (yellow). Movie shows sequential optical sections from z-stack.

Supplemental Movie 16: *Nf2*^{PancKO} P1 pancreata, stained with MUC1 (yellow). Movie shows sequential optical sections from z-stack.

Supplemental Movie 17: *Pdx1-Cre*;*Nf2*^{ff}; *Yap*$^{f/+}$;*Taz*$^{f/+}$ P1 pancreata, stained with MUC1 (yellow). Movie shows sequential optical sections from z-stack.

Supplemental Movie 18: *Pdx1-Cre*;*Nf2*^{ff}; *Yap*^{ff};*Taz*^{ff} P1 pancreata, stained with MUC1 (yellow). Movie shows sequential optical sections from z-stack.

Supplemental References

1. Villasenor, A., Chong, D.C., Henkemeyer, M. & Cleaver, O. Epithelial dynamics of pancreatic branching morphogenesis. *Development* **137**, 4295-4305 (2010).
2. Lun, A.T.L.*, et al.* EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. *Genome Biology* **20**, 1-9 (2019).
3. McCarthy, D.J., Campbell, K.R., Lun, A.T.L. & Wills, Q.F. Scater: Pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**, 1179-1186 (2017).
4. L. Lun, A.T., Bach, K. & Marioni, J.C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biology* **17**, 75 (2016).
5. Germain, P.L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M.D. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res* **10**, 979 (2021).
6. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for dimension reduction. (2018).
7. Campello, R.J.G.B., Moulavi, D. & Sander, J. Density-Based Clustering Based on Hierarchical Density Estimates. 160-172 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).
8. Coifman, R.R. & Lafon, S. Diffusion maps. *Applied and Computational Harmonic Analysis* **21**, 5-30 (2006).
9. Haghverdi, L., Büttner, M., Wolf, F.A., Buettner, F. & Theis, F.J. Diffusion pseudotime robustly reconstructs lineage branching. in *Nature Methods*, Vol. 13 845-848 (2016).
10. van Dijk, D., Sharma, R., Nainys, J. & et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716-729.e727 (2018).
11. CP, C., KA, D. & TJ, C. Integration of Multiple, Diverse Methods to Identify Biologically Significant Marker Genes. *Journal of molecular biology* **434**(2022).
12. Subramanian, A.*, et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**, 15545-15550 (2005).
13. Liberzon, A.*, et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739-1740 (2011).
14. Aibar, S., González-Blas, C.B., Moerman, T. & et al. SCENIC: Single-cell regulatory network inference and clustering. *Nature Methods* **14**, 1083-1086 (2017).
15. Van de Sande, B.*, et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* **15**, 2247-2276 (2020).
16. Finak, G.*, et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biology* **16**, 1-13 (2015).
17. La Manno, G.*, et al.* RNA velocity of single cells. in *Nature*, Vol. 560 494-498 (Springer US, 2018).
18. Bergen, V., Lange, M., Peidli, S., Wolf, F.A. & Theis, F.J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology* **38**, 1408-1414 (2020).
19. Stuart, T.*, et al.* Comprehensive integration of single cell data. in *bioRxiv* 460147 (2018).
20. Coifman, R.R.*, et al.* Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. in *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 102 7426-7431 (National Academy of Sciences, 2005).
21. Haghverdi, L., Lun, A.T.L., Morgan, M.D. & Marioni, J.C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. in *Nature Biotechnology*, Vol. 36 421-427 (Nature Publishing Group, 2018).
22. Street, K.*, et al.* Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. in *BMC Genomics*, Vol. 19 1-16 (BMC Genomics, 2018).

23. Van den Berge, K.*, et al.* Trajectory-based differential expression analysis for single-cell sequencing data. *Nature Communications* **11**, 1201 (2020).