Single nucleus RNA sequencing maps acinar cell states in a human pancreas cell atlas

Luca Tosti[1,2], Yan Hang[3], Timo Trefzer[1,2], Katja Steiger[4], Foo Wei Ten[1,2], Soeren Lukassen[1,2], Simone Ballke[4], Anja A. Kuehl[5], Simone Spieckermann[5], Rita Bottino[6], Wilko Weichert[4], Seung K. Kim[3,7,8], Roland Eils[1,2,9,]* and Christian Conrad[1,2,]*

[1]Berlin Institute of Health (BIH), Berlin, Germany. [2]Charité - Universita□tsmedizin Berlin, Digital Health Center, Berlin, Germany. [3]Department of Developmental Biology, Stanford University School of Medicine, Stanford, CA, USA. [4]Institute of Pathology, Technische Universität München, Munich, Germany. [5]iPATH.Berlin, Charité-Universitätsmedizin Berlin, Berlin, Germany. [6]Institute of Cellular Therapeutics, Allegheny Health Network, Pittsburgh, PA, USA. [7]Department of Medicine, Endocrinology and Oncology Divisions, Stanford University School of Medicine, Stanford, CA, USA. [8]Stanford Diabetes Research Center, Stanford University School of Medicine, Stanford, CA, USA. [9]Health Data Science Center, Faculty of Medicine, University of Heidelberg, Heidelberg, Germany. *Corresponding authors.

Abstract

The cellular heterogeneity of the human pancreas has not been previously characterized due to the presence of extreme digestive enzymatic activities, causing rapid degradation of cells and RNA upon resection. Therefore, previous cellular mapping studies based on gene expression were focused on pancreatic islets, leading to a vast underrepresentation of the exocrine compartment. By profiling the transcriptome of more than 110,000 cells from human donors, we created the first comprehensive pancreas cell atlas including all the tissue components. We unveiled the existence of four different acinar cell states and suggest a division of labor for enzyme production within the healthy exocrine pancreas, which has so far been considered a homogeneous tissue. This work provides a novel and rich resource for future investigations of the healthy and diseased pancreas.

Main text

Single-cell RNA sequencing (scRNA-seq) has tremendously expanded our understanding of heterogeneous human tissues and made the identification of novel functional cell types in the lung, brain and liver possible[1–5]. The development of single-nucleus RNA-seq (sNuc-seq) has further broadened its application to tissues which are difficult to dissociate or already archived, such as clinical samples[6]. Pancreatic exocrine tissues contain among the highest level of digestive enzymatic activities in the human body[7], hindering the preparation of undegraded RNA from this organ. Therefore, previous scRNA-seq studies of the human pancreas have been restricted to the islets of Langerhans (the endocrine part of the organ) in order to remove the exocrine compartment, namely the acinar and ductal cells responsible for the production and transport of digestive enzymes. Following their isolation, the endocrine islets were cultured *in*

vitro, enzymatically dissociated and processed on microfluidics devices before next-generation sequencing[8–14]. While this strategy proved to be successful in generating a draft of the endocrine human pancreas cell atlas, it has distinct disadvantages. For example, only a very small number of exocrine cells have been captured and their numbers are largely underrepresented relative to homeostatic physiological conditions (approximately 5% rather than 95%). Moreover, *in vitro* culture and dissociation steps are known to introduce technical artefacts in gene expression measurements[15]. In this work we opted to use flash-frozen tissue biopsies isolated from pancreata of six human donors followed by sNuc-seq (Fig. 1a), avoiding *in vitro* expansion and dissociation procedures, aiming to obtain an unbiased sampling of the organ.

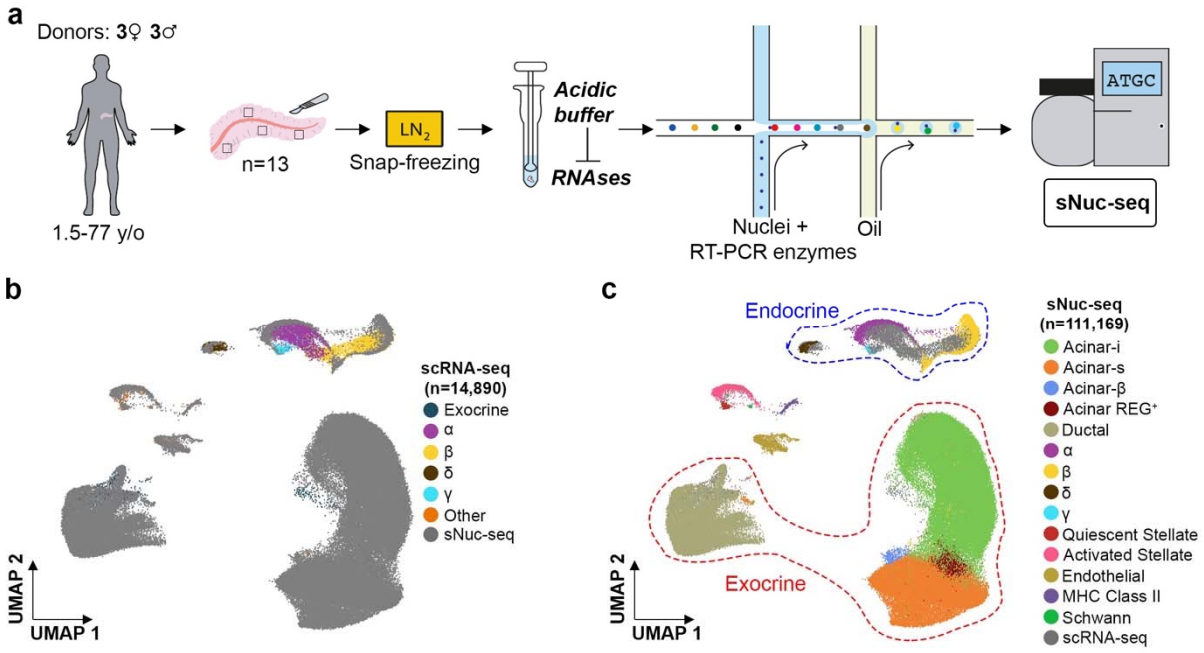


Fig. 1 | sNuc-Seq identifies cell types in the human healthy pancreas. a, Overview of the strategy used to perform sNuc-seq. **b,** Merging of sNuc-seq data generated in this study with previous scRNA-seq datasets[8–12] of the endocrine human pancreas, shown as clusters in a two-dimensional UMAP embedding. **c,** Major cell types identified from sNuc-Seq of the human pancreas shown as clusters in a two-dimensional UMAP embedding.

To isolate nuclei, we initially applied a protocol commonly used in sNuc-seq[16], but we were not able to recover intact RNA (Suppl. Fig. 1a-c). On the basis of distinct protocols described in the

19th century[17] and applied throughout the first decades of the 20th century[18–20], we optimized a citric acid-based buffer which enabled us to reduce RNA degradation during nuclei isolation and achieve a much higher yield of cDNA from human pancreatic samples (approximately 40-50 times higher than the standard protocol) (Suppl. Fig. 1d). We isolated nuclei from human pancreas biopsies collected from three male and three female neurologically deceased donors, spanning the age range from 1.5 to 77 (13 samples in total) (Fig. 1a and Suppl. Table 1), generating the largest human pancreas cell atlas dataset available to date. The average number of UMIs detected per nucleus was 3,597 and the average number of genes detected per nucleus was 1,190 (Suppl. Fig. 2a). As we sought to identify previously described pancreatic cell types, we applied canonical correlation analysis (CCA) to both reduce batch effects and integrate our data with previously annotated human pancreas scRNA-seq data[21]. Our results confirmed that the different sNuc-seq samples were homogeneously merged and fully integrated with scRNA-seq despite the use of different entities as starting material (nucleus versus whole cell) (Fig. 1b and Suppl. Fig. 3). We annotated the majority of the clusters based on previous studies, confirming that sNuc-seq enabled us to capture most of the previously reported human pancreatic cell types (Fig. 1b-c). Importantly, the proportion of cells identified with distinct technologies differ substantially since earlier scRNA-seq studies focused on the endocrine compartment of the pancreas while in our work the majority of the data is constituted by nuclei derived from acinar and ductal cells (Suppl. Fig. 4a), hence complementing and completing the previous scRNA-seq analyses performed in the healthy organ (approximately 10-fold increase in analyzed cells). One major group of clusters contained the different endocrine cells (approximately 4% of the total number of nuclei) and their identity was confirmed by the expression of known specific hormones, namely insulin (*INS*, β cells), glucagon (*GCG*, α cells), somatostatin (*SST*, δ cells) and pancreatic polypeptide (*PPY*, γ cells) (Fig. 2a). Smaller clusters included endothelial cells (1,6% of the total nuclei), characterized by the expression of *FLT1*, *PLVAP*, *VWF*, *CD36* and *SLCO2A1*

and antigen-presenting MHC class II (0,4% of the total nuclei, expressing *CD74*, *CD45*, *ZEB2*, *HLA-DRA*, *HLA-DRB1* and *HLA-DPA1*) (Fig. 2b).

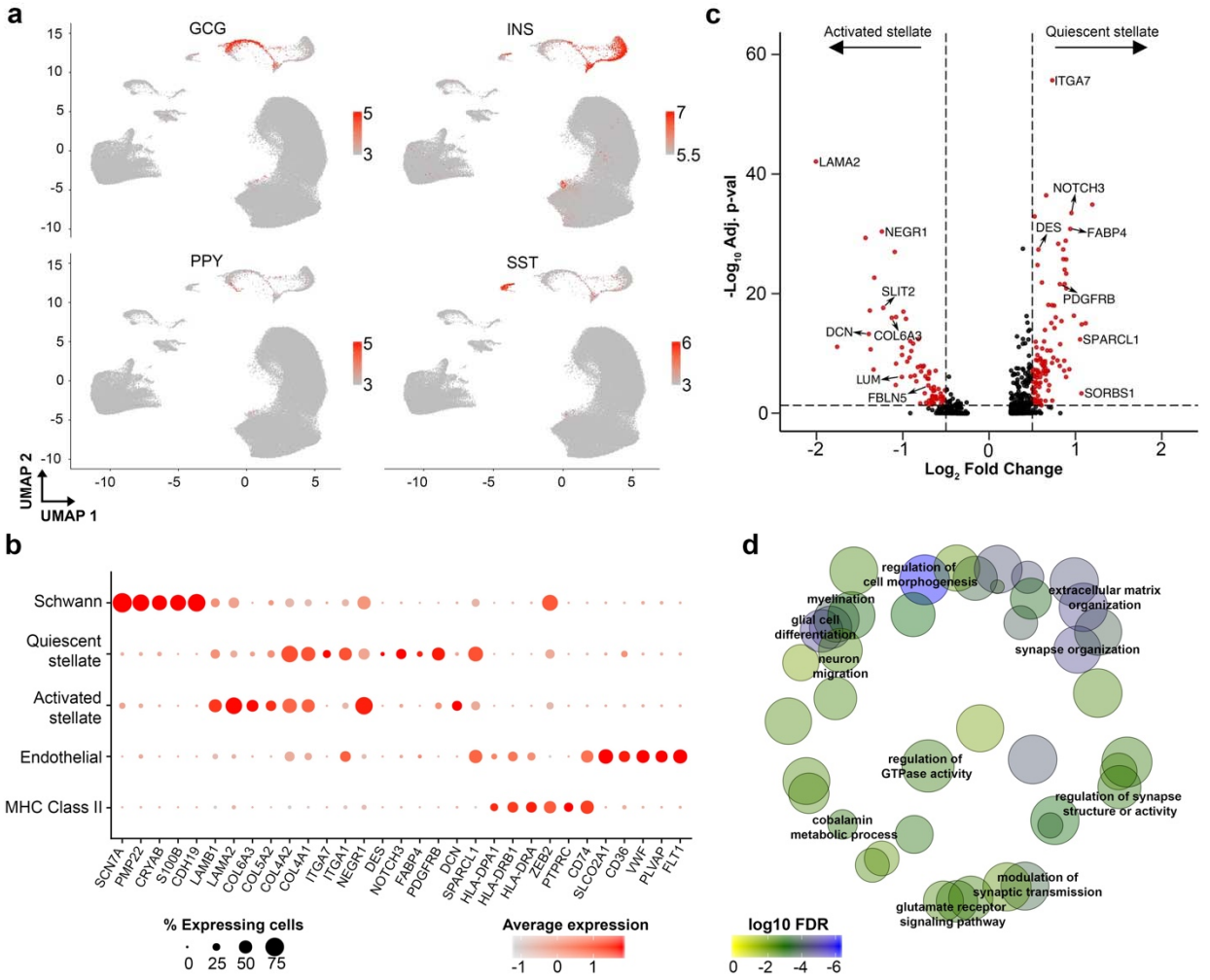


Fig. 2 | Characterization of other pancreatic cell types. a, UMAP plots showing the expression of the endocrine cell markers *GCG* (α cells), *INS* (β cells), *PPY* (γ cells) and *SST* (δ cells). **b,** Dotplot showing the expression of specific markers in Schwann, quiescent stellate, activated stellate, endothelial and MHC class II cells. **c,** Volcano plot showing differentially expressed genes between activated and quiescent stellate cells. Red dots represent genes with average log expression >0.5 and an adjusted p-value <0.05. **d,** KEGG pathway over-represented ontology terms enriched in Schwann cells. Colors indicate false discovery rate, while the size of each circle is proportional to the number of genes associated with the KEGG term.

Pancreatic stellate cells (PSCs) are known to play key roles in the normal physiology of the pancreas[22] and in diseases such as pancreatitis and pancreatic cancer[23]. We identified two distinct states of PSCs, activated (aPSCs) and quiescent pancreatic stellate cells (qPSCs), confirming that they are both present in the normal pancreas and that their activation is not an artefact of *in vitro* culturing of endocrine islets[10]. In particular, qPCSs expressed higher level of *SPARCL1*, similarly to the liver counterpart[24], *PDGFRB* and *FABP4*, likely involved in the retinoid-storing function of these cells[25] (Fig. 2b). Moreover, qPSCs were enriched in the intermediate filament protein desmin (*DES*) and integrin genes such as *ITGA1* and *ITGA7*, suggesting a structural role within the pancreas (Fig. 2b-c). When qPSCs become activated, they change their morphology towards a myofibroblast-like phenotype and are able to migrate and remodel the extracellular matrix (ECM)[22]. Both qPSCs and aPSCs express *COL4A1* and *COL4A2* (Fig. 2b), but aPSCs showed higher levels of other collagen genes such as *COL5A2*, *COL6A3* and components of the basement membrane such as laminin proteins *LAMA2* and *LAMB1* (Fig. 2b). Furthermore, we detected higher levels of *SLIT2*, *FBLN5* and *LUM*, known mediators of fibrogenesis and migration in hepatic stellate cells[26,27] (Fig. 2c). The smallest cluster of cells identified in this study is represented by Schwann cells (22 nuclei, constituting 0.02% of the total) expressing specific markers such as *CDH19*, *S100B*, *CRYAB*, *PMP22* and *SCN7A* (Fig. 2b). We did not detect genes associated with Schwann cell dedifferentiation and repression of myelin sheath formation, previously reported to be upregulated due to extraction and culture conditions[10]. Instead, gene over-representation analysis showed the enrichment of specific terms such as ‘myelination’, ‘synapse organization’ and ‘modulation of synaptic activity’ (Fig. 2d).

Our data closely reflect the composition of the human pancreas in its physiological status, therefore the majority of the analyzed nuclei belong to two main clusters representing the exocrine pancreas. Acinar cells, accounting for 80% of the nuclei, were identified based on the expression of digestive enzymes such as *AMY1A/B*, *CPA1/2*, *PRSS1* and transcription factors

(TFs) such as *RBPJL* or *FOXP2*, while ductal cells represented 15% of the total amount of cells and were identified based on the expression of *CFTR*, *ANXA4*, and *BICC1* (Fig. 3a). In agreement with previous scRNA-seq studies[9,11], ductal cells appeared homogeneous and we could not distinguish the $CD44^{+}/CFTR^{+}$ and the $MUC1^{+}$ subtypes previously described[10], possibly due to the lower sensitivity of sNuc-seq compared to scRNA-seq.

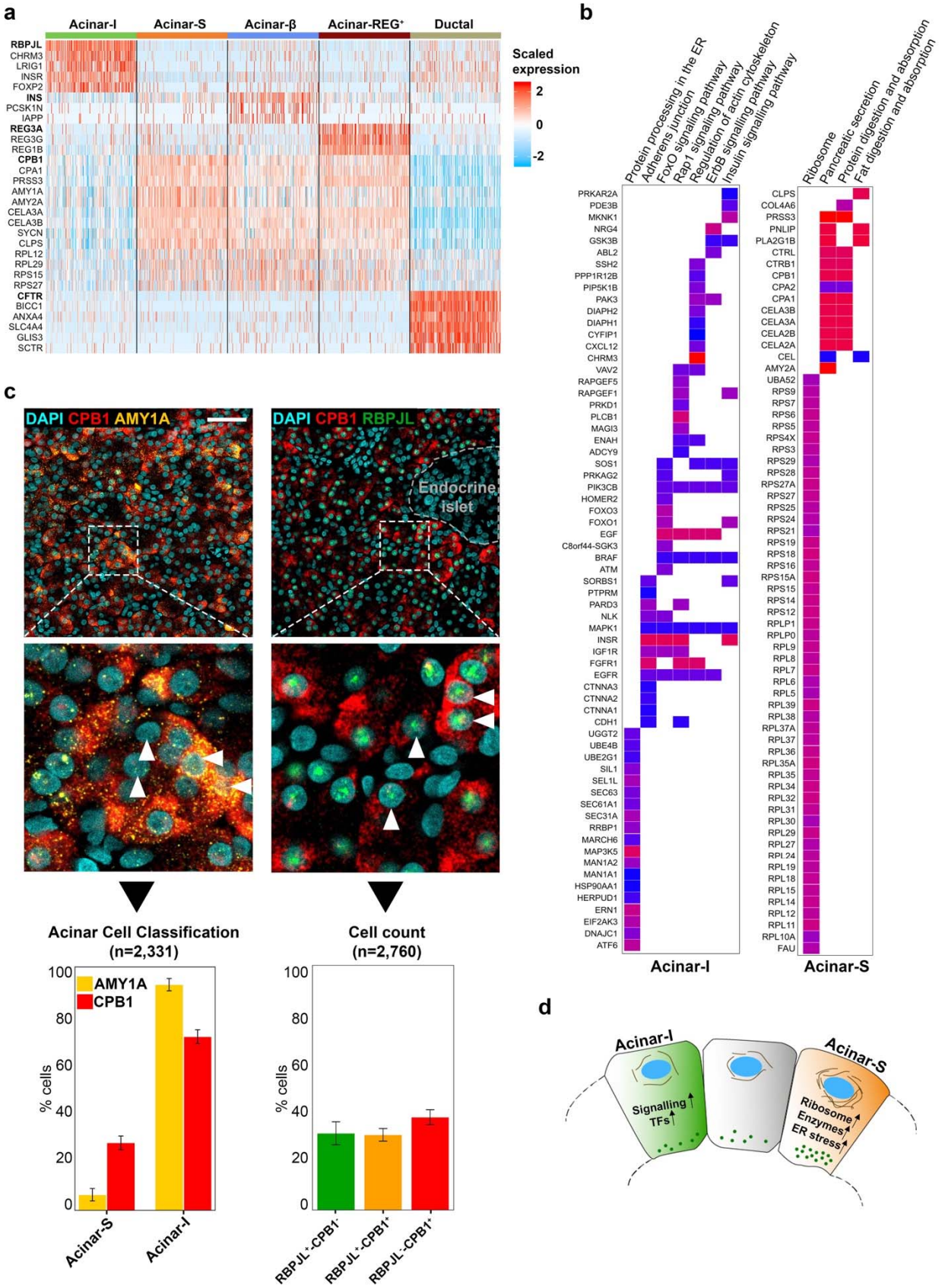

a
Acinar-I
Acinar-S
Acinar-β
Acinar-REG+
Ductal
RBPJL
CHRM3
LRIG1
INSR
FOXP2
INS
PCSK1N
IAPP
REG3A
REG3G
REG1B
CPB1
CPA1
PRSS3
AMY1A
AMY2A
CELA3A
CELA3B
SYCN
CLPS
RPL12
RPL29
RPS15
RPS27
CFTR
BICC1
ANXA4
SLC4A4
GLIS3
SCTR
Scaled expression
2
0
-2
b
Protein processing in the ER
Adherens junction
FoxO signaling pathway
Rap1 signaling pathway
Regulation of actin cytoskeleton
ErbB signaling pathway
Insulin signalling pathway
Ribosome
Pancreatic secretion
Protein digestion and absorption
Fat digestion and absorption
Acinar-I
Acinar-S
c
DAPI CPB1 AMY1A
DAPI CPB1 RBPJL
Endocrine islet
Acinar Cell Classification (n=2,331)
Cell count (n=2,760)
AMY1A
CPB1
% cells
Acinar-S
Acinar-I
RBPJL+CPB1+
RBPJL+CPB1-
RBPJL-CPB1+
d
Acinar-I
Acinar-S
Signalling TFs
Ribosome
Enzymes
ER stress

Fig. 3 | Characterization of acinar cells in the human exocrine pancreas. a, Heatmap of acinar and ductal cell specific genes. **b,** Over-representation analysis examining cellular pathways active in acinar-I and acinar-S cells. **c,** On the left, example image of RNA-FISH for *CPB1* and *AMY1A* and quantification of secretory (horizontal triangles) and idling (vertical triangles) acinar cells. For further details about the classification, see Materials and Methods and Suppl. Fig. 6. On the right, example image of RNA-FISH for *CPB1* and *RBPJL* and quantification of positive cells for each probe. Scale bar = 50 μm. **d,** A model of the human pancreatic acinus highlighting newly identified secretory and idling cells.

The unprecedented availability of sNuc-seq data enabled us to uncover a previously unknown degree of heterogeneity among acinar cells and to identify four distinct subtypes, including two which were not described before (Fig. 3a,d). The smallest population (called acinar-β) was characterized by the simultaneous expression of acinar markers and the insulin hormone. Acinar-β cells were previously identified in different organisms as well as in healthy and diabetic human patients but their functional role in pathological or healthy conditions has not been investigated[28–30] (Fig. 3a). A second small population of acinar cells (acinar-REG$^+$) expressed higher levels of members of regenerating (REG) protein family such as *REG3A* and *REG3B* (Fig. 3a). Acinar-REG$^+$ cells were reported in a previous scRNA-seq study[9] and represent an extremely interesting population of cells, considering the important role of REG protein upregulation in pancreatic lesions such as acinar-to-ductal metaplasia (ADM) and pancreatic intraepithelial neoplasia (PanIN)[31,32]. Strikingly, we detected two yet unknown subpopulations of acinar cells not previously identified in human scRNA-seq experiments. These two clusters had a similar number of UMI per nucleus, indicating a similar content of RNA molecules, but one population had a threefold higher number of expressed genes, denoting a more complex transcriptome (Suppl. Fig. 5a). To characterize these two populations, we identified differentially expressed genes and performed gene over-representation analysis. The subtype with the higher number of genes showed enrichment in terms associated with ‘adherent junctions’ and

‘regulation of actin cytoskeleton’, ‘Insulin signaling pathway’, ‘Rap1 signaling pathway’, and ‘protein processing in the ER’ (Fig. 3b). Therefore, these cells present typical acinar cell features, such as the ability to form adherent junctions necessary to maintain the integrity of the acini and their electrophysiological coupling, while also being able to respond to external stimuli, but they are less committed to the protein secretion task. We named these cells, ‘idling acinar’ cells (acinar-I). The other acinar cell cluster was characterized by strong expression of ribosomal genes, in agreement with the notion that acinar cells have the highest rate of protein synthesis in the human body[33], and digestive enzyme genes (Fig. 3b), hence we named them ‘secretory acinar’ cells (acinar-S). To validate the sNuc-seq findings, we performed RNA-FISH on the same samples used for nuclei isolation. Notably, RNA-FISH in the human healthy exocrine pancreas has not been previously reported since, as for sNuc-seq, elevated RNA degradation usually hinder these experiments (see Materials and Methods). Successful RNA-FISH experiments using probes for *CPB1* (Carboxypeptidase), *AMY1A* (Amylase) and *RBPJL* (a key transcription factor of acinar cells) were performed and downstream analyses confirmed the existence of distinct acinar cell states expressing different levels of these transcripts (Fig. 3c). In particular, mRNA of *CPB1* and *AMY1A* (both encoding for digestive enzymes) showed high heterogeneity across the tissue and we were able to distinguish the two classes representing acinar-I and -S cells (Fig. 3c). Interestingly, we detected *RBPJL*, a gene highly expressed in acinar-I cells (Fig. 3a), in the acinar-S cells (Fig. 3c). This discrepancy between RNA-FISH and sNuc-seq is likely due to the different sensitivity of the two technologies. It has been previously estimated that approximately 90% of the mRNA molecules in a pancreatic acinar cell encode for fewer than 30 proteins, namely the secretory enzymes[34,35]. Hence, the high amount of RNA encoding for ribosomal and digestive enzyme genes is likely to hamper the detection of lowly expressed genes via sNuc-seq in acinar-S cells, while RNA-FISH is able to capture these transcripts. Importantly, the nuclear localization of *RBPJL* is consistent with its lower expression

level compared to the *CPB1* mRNA (Fig. 3c). To further investigate the molecular features of and acinar-I and -S cells we applied SCENIC, a computational tool able to infer putative transcription factor-target regulatory networks (regulons) from single cell gene expression data[36]. Acinar-S cells showed activation of two regulons, *XBP1* and *ATF4*, both involved in the unfolded protein response (UPR) pathway[37], reflecting strongly increased protein production and endoplasmic reticulum (ER) stress. *UBTF*, a key regulator of ribosomal RNA transcription, is also active in acinar-S cells, reflecting their engagement in protein translation. On the other hand, acinar-I cells showed activation of regulons such as *GATA4*, *FOXP2* and *NR5A2*, known exocrine pancreas transcription factors and *CREB3L1/2*, likely involved in the basal secretory activity of the cells (Suppl. Fig. 5b). Thus, RNA-FISH and SCENIC analyses support the distinction of the two populations of acinar cells identified through sNuc-seq. Acinar-S cells sustain the elevated protein production task of the exocrine pancreas, while acinar-I cells appear less committed to this function although they express markers specific for acinar cell identity. We thus hypothesize that acinar-I cells represent a different stage of acinar cell activation that could be converted to acinar-S cells following hormonal or nutritional stimulation, resembling the different activation states recently identified in human and mouse β cells[38,39]. Moreover, acinar-I cells might constitute a source of cell replacement for acinar-S cells, which are subject to high level of ER stress and therefore more prone to cell death via apoptosis.

In this work we developed a new protocol for nuclei isolation that could be applied to other challenging tissues and to archived clinical samples. Moreover, we constructed the first comprehensive human pancreas cell atlas by single nuclei sequencing, providing a novel resource for the community of pancreas researchers. We found evidence of two major acinar cell populations (acinar-I and acinar-S), distinguished by differential expression of ribosomal and digestive enzyme genes (Fig. 3d). In rodents, acinar cells exploit polyploidy or multinucleation to deal with the required high biosynthetic activity, but this is not the case for the human

counterpart[40]. Hence, a different evolutionary strategy might be in place to allocate the protein production task to specific cells that, as a consequence, will be subject to ER stress and higher risk of apoptosis. Rodents are not appropriate as model organisms for human healthy exocrine pancreas due to structural, physiological and developmental differences[41,42] and *in vitro* systems (such as organoids) of mature human acinar cells have not yet been described. Further investigations in new models will need to clarify how the acinar heterogeneity is achieved and maintained during homeostasis, whether the acinar-I and acinar-S cells can interconvert under physiological conditions, and their potential role in pancreatic diseases such as pancreatitis and pancreatic ductal adenocarcinoma.

Methods

Human and pig pancreas samples

Samples from Stanford University were procured from non-diabetic cadaveric organ donors. All studies involving human specimens were conducted in accordance with Stanford University Institutional Review Board guidelines. Deidentified human pancreata were procured from previously healthy, non-diabetic donors with less than 12-hour cold ischemia time through the Center for Organ Recovery and Education (CORE, Pittsburgh, PA, USA) and National Diabetes Research Institute (NDRI, Philadelphia, PA, USA) as reported previously[43]. Tissue blocks of 2 cm × 1 cm × 0.2 cm were excised from 3-4 anatomic locations (i.e. head, body, mid-body, and tail) and then immediately transferred into liquid nitrogen to snap freeze. The frozen samples were shipped and stored at -80°C until they were used for nuclei isolation and sequencing. Samples from TUM were procured from non-diseased pancreatic tissue from patients undergoing partial pancreatectomy. Tissue blocks of 0.5 cm × 0.5 cm × 1 cm were collected immediately after removal of the pancreas, placed into cryo tubes and transferred into liquid nitrogen. The samples were stored at -196°C until they were used for sequencing. The study was approved by the hospital Ethics Committee (number 403/17S). To further dissect the preanalytical problems in procurement of pancreatic tissue for single-cell sequencing, we also sampled pancreatic tissue from healthy pigs sacrificed due to other reasons (approved by local authorities AZ .3-8-07, Regierung von Oberbayern, München, Germany) under completely standardized conditions. The pancreas was removed after the heartbeat had stopped and tissue blocks (0.5 cm × 0.5 cm × 1 cm) were sampled at different time points (15 min and 30 min cold ischemia time) and transferred into liquid nitrogen. The samples were stored according to the requirements of fixation solution/procedure. To check for morphological integrity of the tissue, a paraffin block

and hematoxylin-eosin stained slide was produced from each sampling site and evaluated by two experienced pathologists (S.B. and K.S.).

Nuclei isolation

To isolate nuclei from frozen tissue, we initially applied a protocol based on the use of dense sucrose solutions and detergents at slightly alkaline pH values[16]. However, the RNA extracted from isolated nuclei was highly degraded compared to the RNA in the original bulk tissue. Several modifications to the original protocol were tested on the basis of previously described protocols involving the use of DSP[44], methanol fixation[45,46], or the addition of further RNAse inhibitors such as ribonucleoside vanadyl complexes, but none of these approaches improved the quality of the RNA isolated from pancreatic nuclei. We optimized a citric acid-based buffer which enabled us to limit RNA degradation and successfully generate cDNA libraries from human pancreatic nuclei. Generally, snap-frozen pancreatic tissue samples were cut into pieces <0.3 cm and homogenized with one stroke of "loose" pestle in 1 mL citric-acid based buffer (Sucrose 0.25 M, Citric Acid 25 mM, Hoechst 33342 1 μg/mL) using a glass dounce tissue grinder. The tissue was incubated on ice for 5 minutes and then homogenized with 5-10 more strokes. After further 5 minutes of incubation, tissue was homogenized with 3-5 strokes using the "loose" pestle and then 5 more strokes using the "tight" pestle. Homogenate was filtered through a 35-μm cell strainer and centrifuged for 5 minutes at 500 x g at 4°C. Supernatant was carefully removed, nuclei were resuspended in 1 mL of citric acid buffer and the centrifugation step was repeated. Nuclei were then resuspended in 300 μL of cold resuspension buffer (KCl 25 mM, $MgCl_2$ 3 mM, Tris-buffer 50 mM, RNaseIn 0.4 U/μL, DTT 1mM, SuperaseIn 0.4 U/μL, Hoechst 33342 1 μg/mL). Nuclei were counted on a Countess II FL Automated Cell Counter, diluted to the desired concentration and immediately loaded on the 10X Chromium controller.

10X sample processing, library preparation and sequencing

Samples were prepared according to the 10x Genomics Single Cell 3′ v2 Reagent Kit user guide with small modifications. The nuclei were diluted using an appropriate volume of resuspension buffer without Hoechst (KCl 25 mM, $MgCl_2$ 3 mM, Tris-buffer 50 mM, RNaseIn 0.4 U/µL, DTT 1mM, SuperaseIn 0.4 U/µL) for a target capture of 10,000 nuclei. After droplet generation, samples were transferred onto a pre-chilled 96-well plate (Eppendorf), heat-sealed and reverse transcription was performed using a Bio-Rad C1000 Thermal Cycler. After the reverse transcription, cDNA was recovered using Recovery Agent followed by Silane DynaBead clean-up step. Purified cDNA was amplified for 15 cycles before bead cleanup using SPRIselect beads (Beckman). Samples were quantified using an Invitrogen Qubit 4 Fluorometer. cDNA libraries were prepared according to the Single Cell 3′ Reagent Kits v2 guide with appropriate choice of PCR cycle number based on the calculated cDNA concentration. Final libraries were sequenced with the NextSeq 500 system in high-output mode (paired-end, 75 bp).

Single-cell RNA sequencing data analysis

Alignment

Since unspliced mRNA is captured in sNuc-seq experiments, we included the intronic reads in the preprocessing steps (alignment and read quantification) to accurately measure gene expression levels. Therefore, raw sequencing data were aligned to the human reference transcriptome using zUMIs[47], software version 2.0.7 and a modified version of GRCh37/hg19 Reference - 2.1.0 provided by 10X, generating a gene expression matrix that was used for downstream analyses.

Quality control and downstream analyses

Quality control and filtering of the nuclei was performed for each sample using the Python package Scanpy version 1.4.1[48]. Nuclei with fewer than 800 or more than 15000 UMIs were discarded. Moreover, nuclei with less than 300 detected genes were removed from the dataset. Downstream analyses were performed using the R package Seurat version 3.0[21] and included also five previously published scRNA-seq datasets (GEO accession numbers GSE81076, GSE85241, GSE86469, GSE84133 and ArrayExpress accession number E-MTAB-5061). Each sNuc-seq dataset was scaled by library size and log-transformed (using a size factor of 10,000 molecules per cell). For each sample, the top 3,000 most variable genes were identified and the sNuc-seq and scRNA-seq datasets were integrated using the "FindIntegrationAnchors" and "IntegrateData" available in Seurat 3.0[21]. Data were scaled to unit variance and zero mean and the dimensionality of the data was reduced by principal component analysis (PCA) (30 components) and visualized with UMAP[49]. Clustering was performed using the Louvain algorithm on the 30 principal components (resolution = 1.0). Small clusters including Schwann cells and delta cells were manually assigned using the "CellSelector" function. Cluster-specific markers were identified with the "FindAllMarkers" function and clusters were assigned to known cell types on the basis of their specific markers (described in the main text). Clusters that appeared to correspond to the same cell types were merged. The density map in Suppl. Fig. 4b was calculated and plotted using the "embedding_density" function of Scanpy version 1.4.1.

Regulon - SCENIC analysis

SCENIC[36] is able to infer gene regulatory networks from single cell gene expression data through three main steps: (a) identification of co-expression modules between TF and putative targets; (b) within each co-expression module, derivation of direct TF-target gene interaction based on enrichment of TF motif in the promoter of target genes, as to generate "regulons"; (c) for each cell, the regulon activity score (RAS) is calculated. Due to the low scalability of

SCENIC (in the original R implementation) to large datasets, we downsampled our dataset to ~11,000 nuclei and the RAS was then projected onto the UMAP embedding.

Gene over-representation analysis

Symbol gene IDs were converted to Entrez gene IDs using the R package "annotables"[50]. The Gene Ontology over-representation analysis was performed using the "enrichGO" function of the clusterProfiler R package[51] (using adjusted p-value < 0.05 and average log(Fold Change) > 0.25) and ReviGO[52] was applied to summarize and visualize the results in Fig. 2d. The KEGG over-representation test was performed using the "enrichKEGG" function and the heatplot in Fig. 3b was generated using the "heatplot" function.

Histology and RNA-FISH

To perform RNA-FISH in the human pancreas, we used thin snap-frozen (2 mm) biopsies for formalin fixation and paraffin embedding, reasoning that the fixation of the tissue would be faster due to the thinness of the tissue, limiting the degradation processes. Therefore, human pancreatic snap-frozen samples were fixed in 10% formalin at 4°C for 14-16 hours and paraffin-embedded. Sections (4 μm) were cut from FFPE pancreatic tissue and processed for RNA *in situ* detection using the RNAscope Multiplex Fluorescent Reagent Kit v2 according to the manufacturer's instructions (Advanced Cell Diagnostics). RNAscope human probes used were: Hs-CPB1 (#569891-C3), Hs-RBPJL (#581131), Hs-AMY1A (#503551-C2). RNA-FISH images were acquired on a Leica SP8 confocal laser-scanning microscope equipped with a 40x/1.30 oil objective (Leica HC APO CS2).

RNA-FISH image analysis

Automated nuclei instance detection and segmentation were implemented and performed using a deep learning object detection and instance segmentation workflow based on the Mask R-CNN architecture[53]. The neural network was initialized using pre-trained models trained on the Microsoft COCO: Common Objects in Context dataset[54] and fine-tuned on curated datasets of nuclei images. Nuclei images on the DAPI channel were used as inputs for the neural network to produce segmentation for each individual nucleus. The nuclei sizes were calculated using these segmented nuclei masks, and objects <150 pixels were filtered out and excluded from subsequent analyses.

To perform transcript abundance analysis, the RNA-FISH channels were thresholded and binarized by computing the gray-level moments of the input images as implemented in Fiji. Transcript abundance was estimated by overlaying the nuclei masks on the thresholded probe channels and calculating the number of pixels within each mask. In order to account for transcript signals that are predominantly localized outside of the nuclei masks, we expanded the nuclei masks by morphological dilation (3 iterations using a 7x7 elliptical kernel) as implemented in OpenCV[55] prior to quantification. We then performed k-means clustering on the frequency distributions of pixel counts per cell (nucleus) to identify and separate the cells into population classes (e.g. high, low, and negative expression/abundance). A cluster number of 3 was selected for the FISH signals to better capture gradual differences between cells.

Data availability

Data generated during this study will be deposited in the Gene Expression Omnibus (GEO) with the accession code [GSEXXXXXX]. The human pancreas cell atlas can be interactively explored at [www.XXXXXX.de].

References

1. Montoro, D. T. et al. A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature 560, 319–324 (2018).

2. Plasschaert, L. W. et al. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. Nature 560, 377–381 (2018).

3. Islam, S. et al. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat. Methods 11, 163–166 (2014).

4. MacParland, S. A. et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nat. Commun. 9, 4383 (2018).

5. Aizarani, N. et al. A human liver cell atlas reveals heterogeneity and epithelial progenitors. Nature 572, 199–204 (2019).

6. Habib, N. et al. Massively parallel single-nucleus RNA-seq with DroNc-seq. Nat. Methods 14, 955–958 (2017).

7. Farrell, R. E. Chapter 7 - Resilient Ribonucleases. in RNA Methodologies (Fourth Edition) (ed. Farrell, R. E.) 155–172 (Academic Press, 2010).

8. Grün, D. et al. De Novo Prediction of Stem Cell Identity using Single-Cell Transcriptome Data. Cell Stem Cell 19, 266–277 (2016).

9. Muraro, M. J. et al. A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst 3, 385–394.e3 (2016).

10. Baron, M. et al. A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. Cell Syst 3, 346–360.e4 (2016).

11. Segerstolpe, Å. et al. Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. Cell Metab. 24, 593–607 (2016).

12. Lawlor, N. et al. Single-cell transcriptomes identify human islet cell signatures and reveal cell-type-specific expression changes in type 2 diabetes. Genome Res. 27, 208–222 (2017).

13. Wang, Y. J. et al. Single-Cell Transcriptomics of the Human Endocrine Pancreas. Diabetes 65, 3028–3038 (2016).

14. Enge, M. et al. Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. Cell 171, 321–330.e14 (2017).

15. van den Brink, S. C. et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nat. Methods 14, 935–936 (2017).

16. Grindberg, R. V. et al. RNA-sequencing from single nuclei. Proc. Natl. Acad. Sci. U. S. A. 110, 19802–19807 (2013).

17. Carpenter, W. B. & Smith, F. G. The microscope and its revelations. (Blanchard and Lea, 1856).

18. Crossmon, G. The isolation of muscle nuclei. Science 85, 250 (1937).

19. Dounce, A. L. Further studies on isolated cell nuclei of normal rat liver. J. Biol. Chem. 151, 221–233 (1943).

20. Birnie, G. D. Isolation of Nuclei from Animal Cells in Culture (Chapter 2). in Methods in Cell Biology (eds. Stein, G., Stein, J. & Kleinsmith, L. J.) 17, 13–26 (Academic Press, 1978).

21. Stuart, T. et al. Comprehensive Integration of Single-Cell Data. Cell 177, 1888–1902.e21

(2019).

22. Erkan, M. et al. StellaTUM: current consensus and discussion on pancreatic stellate cell research. Gut 61, 172–178 (2012).

23. Shi, Y. et al. Targeting LIF-mediated paracrine interaction for pancreatic cancer therapy and monitoring. Nature 569, 131–135 (2019).

24. Coll, M. et al. Integrative miRNA and Gene Expression Profiling Analysis of Human Quiescent Hepatic Stellate Cells. Sci. Rep. 5, 11549 (2015).

25. D'Ambrosio, D. N. et al. Distinct populations of hepatic stellate cells in the mouse liver have different capacities for retinoid and lipid storage. PLoS One 6, e24993 (2011).

26. Bracht, T. et al. Analysis of disease-associated protein expression using quantitative proteomics—fibulin-5 is expressed in association with hepatic fibrosis. J. Proteome Res. 14, 2278–2286 (2015).

27. Chang, J. et al. Activation of Slit2-Robo1 signaling promotes liver fibrosis. J. Hepatol. 63, 1413–1420 (2015).

28. Melmed, R. N., Benitez, C. J. & Holt, S. J. Intermediate cells of the pancreas. I. Ultrastructural characterization. J. Cell Sci. 11, 449–475 (1972).

29. Yu, L. et al. Insulin-producing acinar cells in adult human pancreas. Pancreas 43, 592–596 (2014).

30. Masini, M. et al. Co-localization of acinar markers and insulin in pancreatic cells of subjects with type 2 diabetes. PLoS One 12, e0179398 (2017).

31. Liu, X. et al. REG3A accelerates pancreatic cancer cell growth under IL-6-associated

inflammatory condition: Involvement of a REG3A-JAK2/STAT3 positive feedback loop. Cancer Lett. 362, 45–60 (2015).

32. Li, Q. et al. Reg proteins promote acinar-to-ductal metaplasia and act as novel diagnostic and prognostic markers in pancreatic ductal adenocarcinoma. Oncotarget 7, 77838–77853 (2016).

33. Kubisch, C. H. & Logsdon, C. D. Endoplasmic reticulum stress and the pancreatic acinar cell. Expert Rev. Gastroenterol. Hepatol. 2, 249–260 (2008).

34. Harding, J. D. et al. Changes in the frequency of specific transcripts during development of the pancreas. J. Biol. Chem. 252, 7391–7397 (1977).

35. Hoang, C. Q. et al. Transcriptional Maintenance of Pancreatic Acinar Identity, Differentiation, and Homeostasis by PTF1A. Mol. Cell. Biol. 36, 3033–3047 (2016).

36. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. Nat. Methods 14, 1083–1086 (2017).

37. Lee, A.-H., Iwakoshi, N. N. & Glimcher, L. H. XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response. Mol. Cell. Biol. 23, 7448–7459 (2003).

38. Xin, Y. et al. Pseudotime Ordering of Single Human β-Cells Reveals States of Insulin Production and Unfolded Protein Response. Diabetes 67, 1783–1794 (2018).

39. Farack, L. et al. Transcriptional Heterogeneity of Beta Cells in the Intact Pancreas. Dev. Cell 48, 115–125.e4 (2019).

40. Anzi, S. et al. Postnatal Exocrine Pancreas Growth by Cellular Hypertrophy Correlates with

a Shorter Lifespan in Mammals. Dev. Cell 45, 726–737.e3 (2018).

41. Case, R. M. Is the rat pancreas an appropriate model of the human pancreas? Pancreatology 6, 180–190 (2006).

42. Dolenšek, J., Rupnik, M. S. & Stožer, A. Structural similarities and differences between the human and the mouse pancreas. Islets 7, e1024405 (2015).

43. Goodyer, W. R. et al. Neonatal β cell development in mice and humans is regulated by calcineurin/NFAT. Dev. Cell 23, 21–34 (2012).

44. Attar, M. et al. A practical solution for preserving single cells for RNA sequencing. Sci. Rep. 8, 2151 (2018).

45. Alles, J. et al. Cell fixation and preservation for droplet-based single-cell transcriptomics. BMC Biol. 15, 44 (2017).

46. Chen, J. et al. PBMC fixation and processing for Chromium single-cell RNA sequencing. J. Transl. Med. 16, 198 (2018).

47. Parekh, S., Ziegenhain, C., Vieth, B., Enard, W. & Hellmann, I. zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. Gigascience 7, (2018).

48. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15 (2018).

49. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv [stat.ML] (2018).

50. Turner, S. annotables. (Github).

51. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16, 284–287 (2012).

52. Supek, F., Bošnjak, M., Škunca, N. & Šmuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One 6, e21800 (2011).

53. He, K., Gkioxari, G., Dollár, P. & Girshick, R. Mask R-CNN. in 2017 IEEE International Conference on Computer Vision (ICCV) 2980–2988 (2017).

54. Lin, T.-Y. et al. Microsoft COCO: Common Objects in Context. arXiv [cs.CV] (2014).

55. Bradski, G. The opencv library. Dr Dobb's J. Software Tools (2000).

Acknowledgements: The authors would like to thank organ donors and their families, David Ibberson (Heidelberg University) for NGS services, the Biomaterial Bank (MTBIO) of the Technical University Munich for support, Katharina Jechow (BIH, Berlin), Lorenz Chua (BIH, Berlin), Alison McGarvey (MDC, Berlin) for critically revising the manuscript and all the members of the Conrad lab for the constructive discussions. This study was funded by Human Cell Atlas (HCA) pilot studies of the Chan Zuckerberg initiative (Charité and Technische Universita□t Mu□nchen: CZI grant 2017-174170) and the European Marie-Skłodowska Curie Actions (EC no. 841755). This publication is part of the Human Cell Atlas - www.humancellatlas.org/publications.

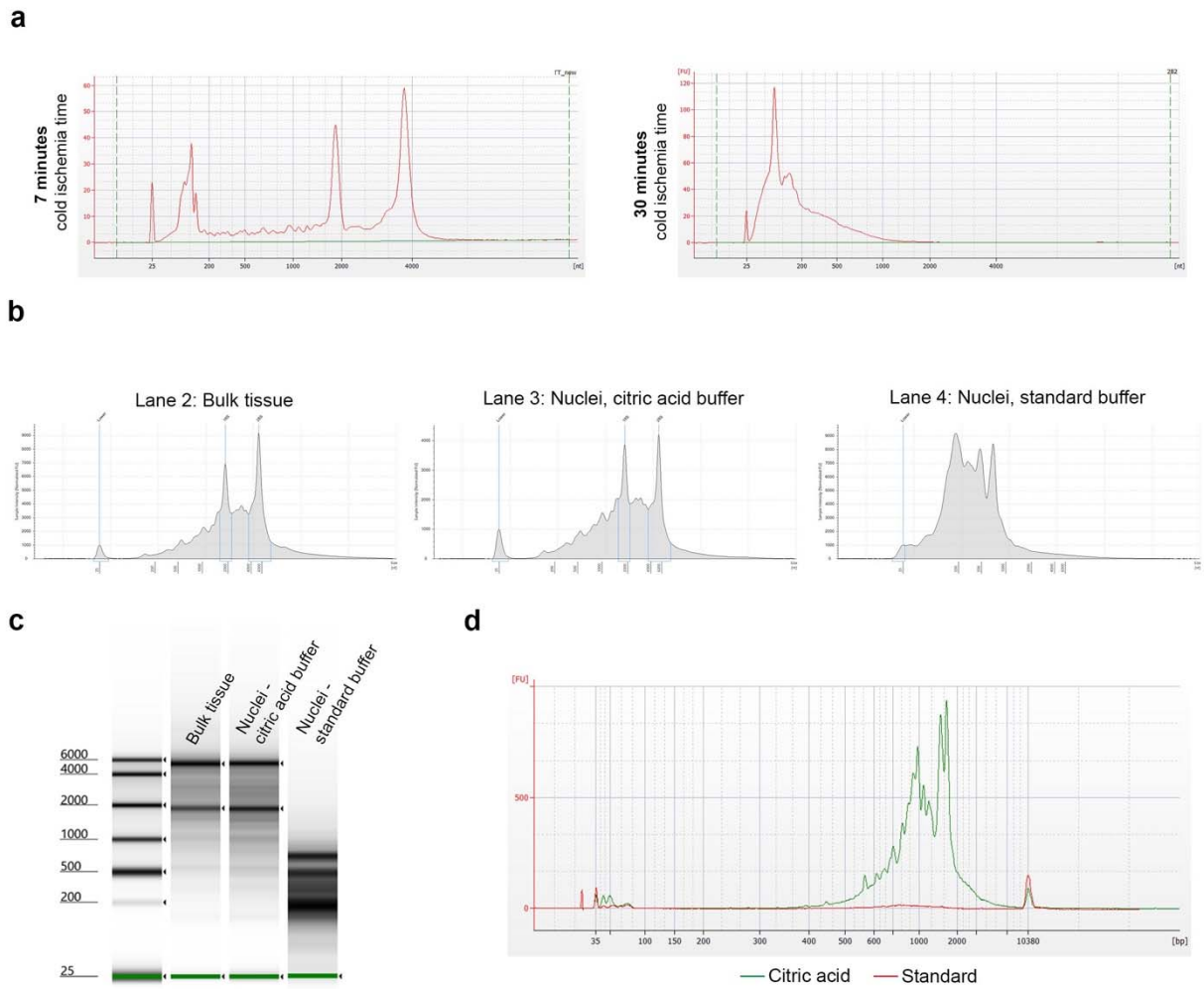
Author contributions: C.C., R.E., S.K.K. and W.W. conceived the study. L.T. optimized the citric acid-based protocol, performed sNuc-seq and RNA-FISH experiments, analyzed the sNuc-seq data and created the interactive app for data visualization. Y.H., R.B., K.S., S.B. procured human and pig samples. T.T. performed initial experiment in pig pancreas. A.A.K. and S.S.

provided support with histology experiments. F.W.T. analyzed RNA-FISH images and S.L. provided support with data analysis and interpretation. R.E. and C.C. supervised experiment design and data interpretation. L.T., C.C. and R.E. wrote the manuscript with input from all the coauthors.

Author information: Authors declare no competing interests. Correspondence and requests should be addressed to christian.conrad@bihealth.de or roland.eils@bihealth.de.

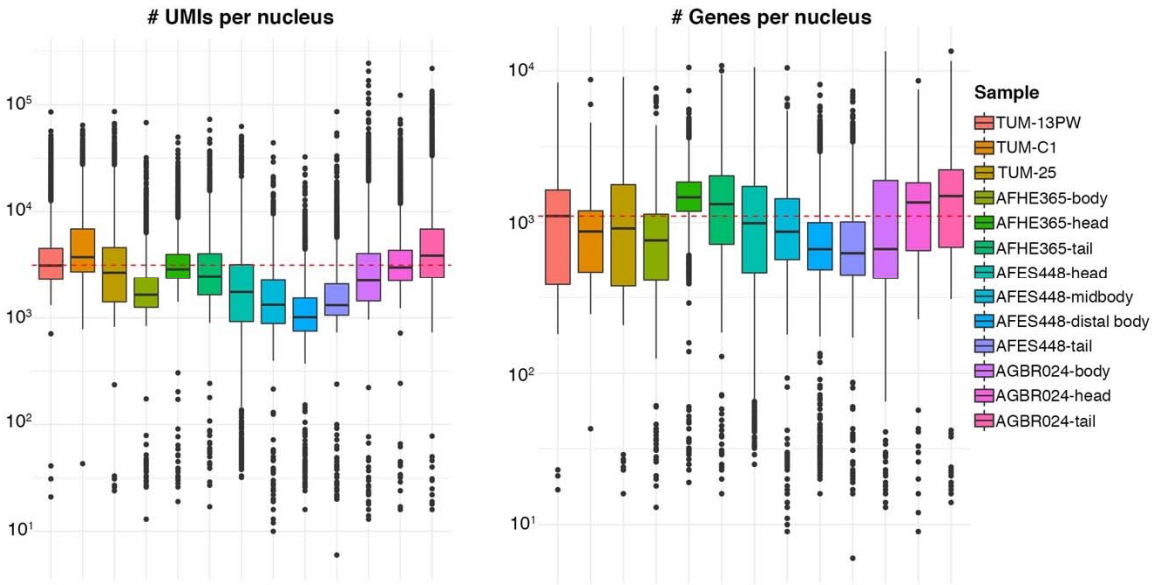
Supplementary Table 1. Human donor metadata

Sample ID	Sex	Age	Pancreatic disease	Diabetes	Procurement lab	Pancreas location
AFHE365-body	F	53	None	None	Stanford, USA	Body
AFHE365-head	F	53	None	None	Stanford, USA	Head
AFHE365-tail	F	53	None	None	Stanford, USA	Tail
AFES448-head	M	30	None	None	Stanford, USA	Head
AFES448-midbody	M	30	None	None	Stanford, USA	Mid-Body
AFES448-body	M	30	None	None	Stanford, USA	Distal Body
AFES448-tail	M	30	None	None	Stanford, USA	Tail
AGBR024-body	M	1.5	None	None	Stanford, USA	Body
AGBR024-head	M	1.5	None	None	Stanford, USA	Head
AGBR024-tail	M	1.5	None	None	Stanford, USA	Tail
TUM-13	F	46	Neuroendocrine Tumor	N/A	Munich, Germany	Tail
TUM-C1	M	77	Pancreatic Ductal Adenocarcinoma	N/A	Munich, Germany	Body
TUM-25	F	59	Mixed Muellerian Tumor	N/A	Munich, Germany	Body

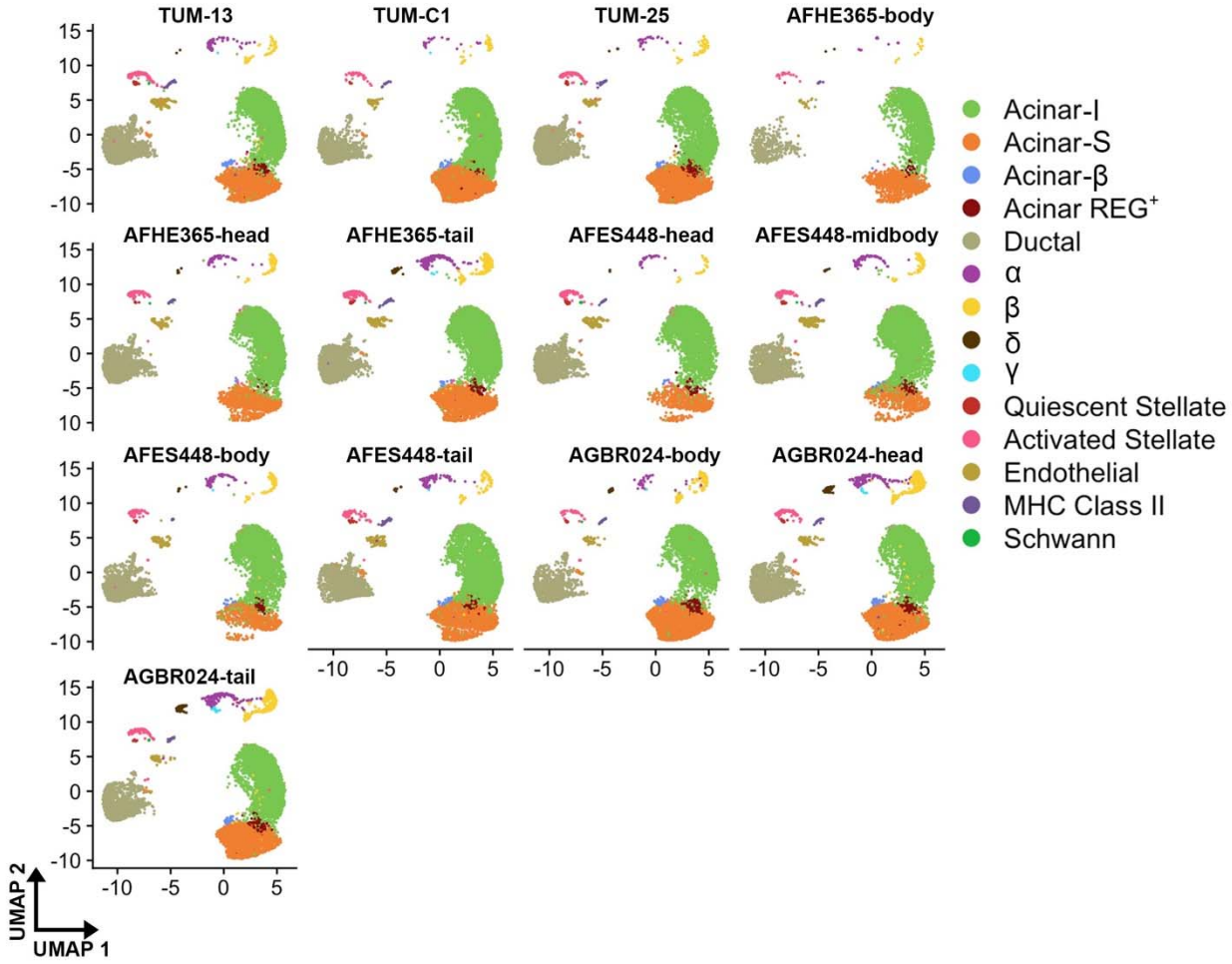


Suppl. Fig. 1 | Reduced RNA degradation using a citric acid buffer for nucleus isolation.

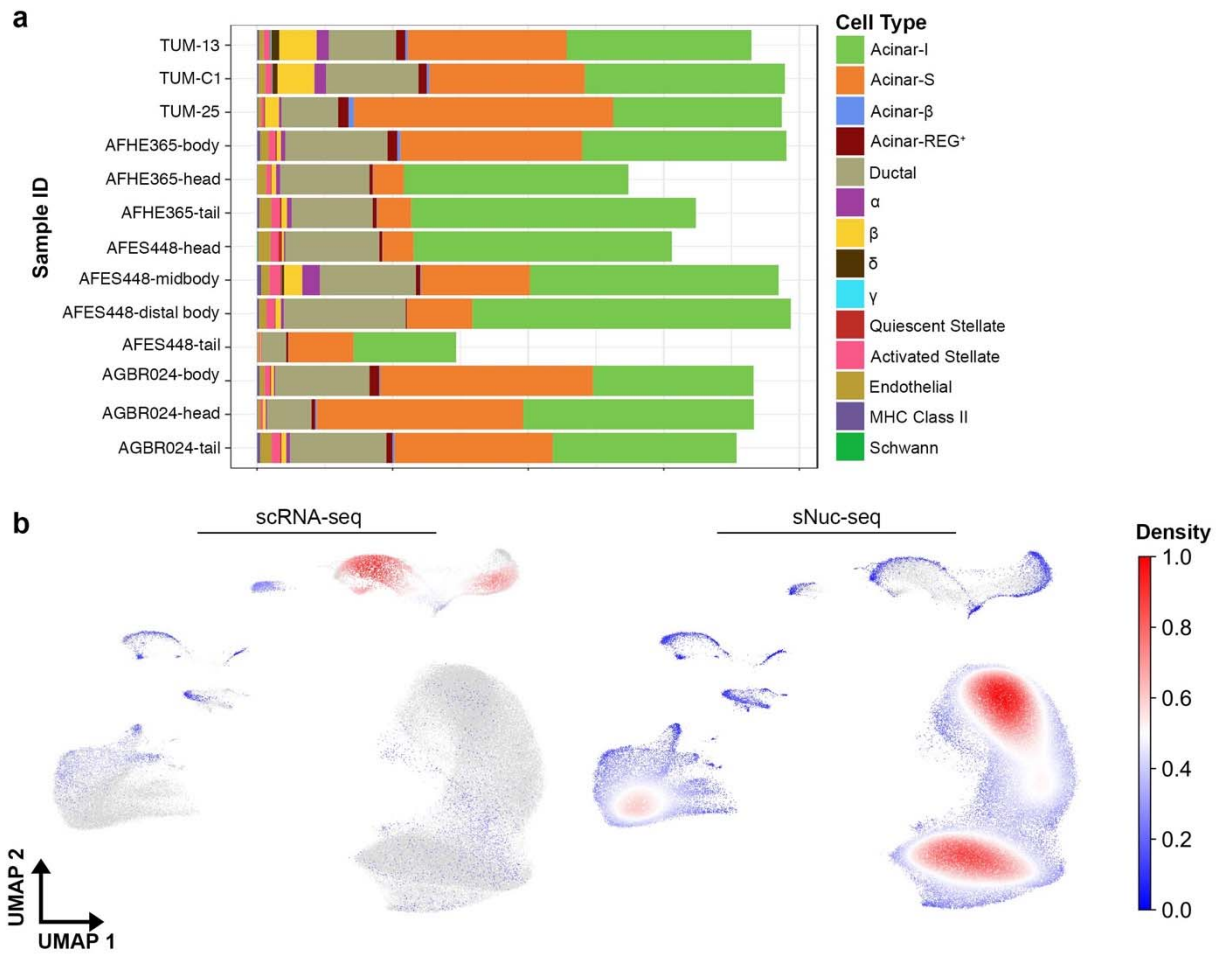
Initial experiments performed in pig pancreas showed that RNA degradation was limited when the total cold ischemia time was shorter than 10 minutes. Based on these results, pancreas biopsies from human donors were collected and immediately snap-frozen. **a**, Electropherogram of bulk RNA extracted from snap-frozen pig pancreatic tissue subject to either 7 or 30 minutes of total cold ischemia. **b**, Electropherograms of bulk RNA extracted from snap-frozen human pancreatic tissue (Bulk tissue). RNA was extracted from nuclei that were isolated from the same tissue as in lane 2 by using either a citric acid buffer or the standard buffer (lanes 3 and 4). **c**, Gel view of the same samples as in **b**. **d**, Yield of cDNA from a sample processed with either the standard or the citric acid-based protocol. The same number of nuclei and PCR cycles were used for both conditions.



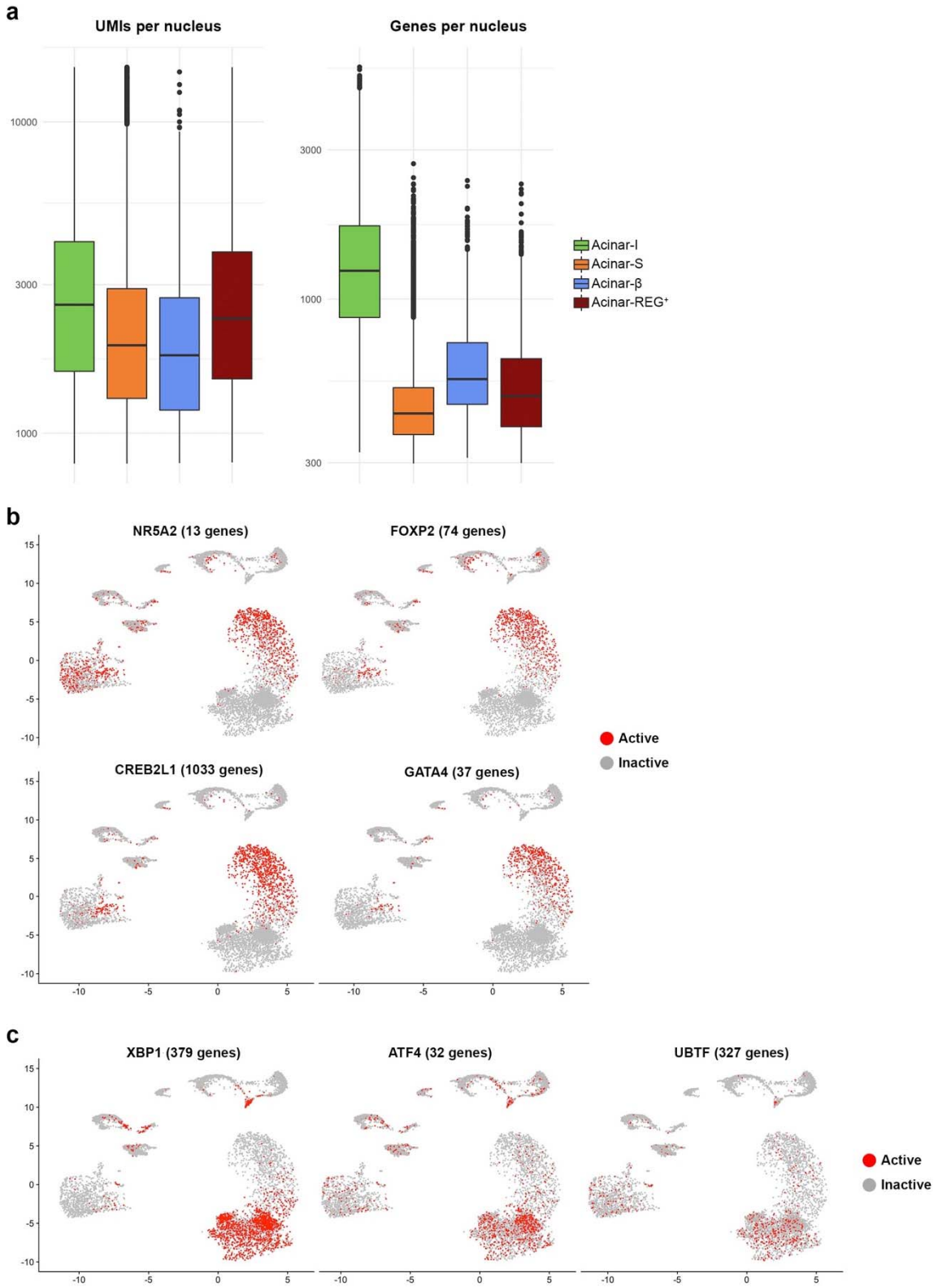
Suppl. Fig. 2 | Quality control of sNuc-seq data and comparison with published scRNA-seq studies. On the left, the boxplots show the distribution of Unique Molecular Identifiers (UMIs) per nucleus for each sample processed in this study. On the right, the boxplots show the distribution of genes per nucleus for each sample. The red dashed lines represent mean values (3,597 for UMIs and 1,190 for the genes).



Suppl. Fig. 3 | Merging of different sNuc-seq samples. sNuc-seq data were split by sample of origin and shown in a two-dimensional UMAP embedding.

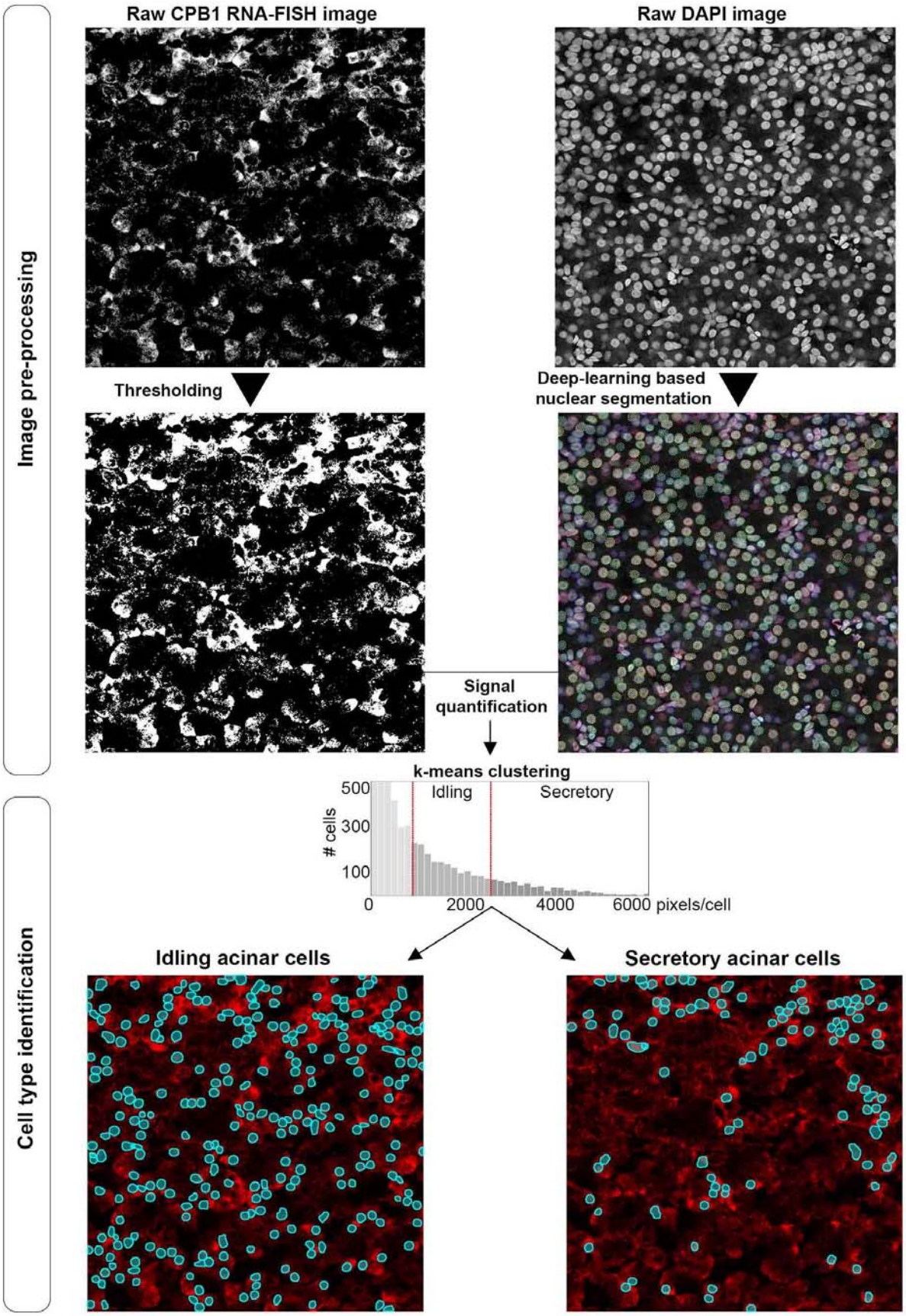


Suppl. Fig. 4 | Different proportion of cells detected by sNuc-seq and scRNA-seq. a, Barplots showing the proportion of cell types identified in each sNuc-seq sample. **b**, Gaussian kernel density estimation was used to calculate the density of cells and was represented in the UMAP embedding for the two distinct technologies, namely scRNA-seq and sNuc-seq. High density values indicate strong contribution of the cells to the overall dataset (i.e. exocrine cells have higher contribution in sNuc-seq and endocrine cells in scRNA-seq).



Suppl. Fig. 5 | SCENIC regulon analysis of acinar-I and acinar-S cells. a, Quantification of UMI per nucleus and genes per nucleus for the different acinar cell populations identified in this study. **b**, SCENIC regulons specifically

active in the acinar-I cells. **c**, SCENIC regulons specifically active in the acinar-S cells. Above each plot, the transcription factor and the number of downstream target genes are indicated.



Suppl. Fig. 6 | Overview of RNA-FISH quantification strategy. Raw RNA-FISH images were thresholded to

remove the background signal and a nuclear segmentation mask was generated by applying a deep-learning algorithm to the DAPI channel of the same image. Signal was quantified for each nucleus and the signal distribution was used to identify idling and secretory acinar cells as explaine