communications biology

Article

https://doi.org/10.1038/s42003-024-07193-3

Exploring human pancreatic organoid modelling through single-cell RNA sequencing analysis

Alessandro Cherubini[1] ✉, Francesco Rusconi[2], Roberta Piras[3], Kaja Nicole Wächtershäuser[4], Marta Dossena[2], Mario Barilani[2], Cecilia Mei[2], Lotta Hof[4], Valeria Sordi[5], Francesco Pampaloni[4], Vincenza Dolo[6], Lorenzo Piemonti[5,7] & Lorenza Lazzari[2] ✉

Human organoids have been proposed to be powerful tools mimicking the physiopathological processes of the organs of origin. Recently, human pancreatic organoids (hPOs) have gained increasing attention due to potential theragnostic and regenerative medicine applications. However, the cellular components of hPOs have not been defined precisely. In this work, we finely characterized these structures, focusing first on morphology and identity-defining molecular features under long-term culture conditions. Next, we focused our attention on hPOs cell type composition using single-cell RNA sequencing founding a complex heterogeneity in ductal components, ranging from progenitor components to terminally differentiated ducts. Furthermore, an extensive comparison of human pancreatic organoids with previously reported transcriptomics signature of human and mouse pancreatic ductal populations, confirmed the functional pancreatic duct subpopulation heterogeneity. Finally, we showed that pancreatic organoid cells follow a precise developmental trajectory and utilize diverse signalling mechanisms, including EGF and SPP1, to facilitate cell-cell communication and maturation. Together our results offer an in-depth description of human pancreatic organoids providing a strong foundation for future in vitro diagnostic and translational studies of pancreatic health and disease.

The pancreas is a complex glandular organ with endocrine and exocrine compartments that affects the functioning of the entire body. Ductal cells, forming about 30-40% of human pancreas[1], play an important role in the transport of digestive enzymes produced by acinar cells, secreting bicarbonate-rich alkaline aqueous solution and mucin[2]. Dysfunctions of ductal cell physiology have been implicated in the occurrence of several diseases including pancreatitis, cystic fibrosis and the rare Alagille syndrome, leading to a worsening of entire organ[3–5]. Moreover, ductal cell plasticity has been involved in pancreatic ductal adenocarcinoma progression[6,7], one of the most lethal types of human cancer[8].

The development of regenerative medicine for pancreatic disorders has been hindered due to the lack of an appropriate in vitro model for the understanding of mechanisms associated with the initiation and progression of the mentioned diseases. Recent technological breakthroughs enable the creation of 3D culture models, called organoids, generated from both primary and reprogrammed stem/progenitor cells. Due to their capacity to self-organize into minimal biological units, and the potential to recapitulate the functions and the complexity of the tissue of origin, they have emerged as powerful models for the study of human development and disease[9,10]. Although organoid formation and characterization from tissues of the liver[11], small intestine[12], prostate[13] and lung[14] have been studied extensively, pancreatic organoids (hPO) remain among the least investigated. A recent characterization of pluripotent stem cell (PSC)-derived pancreatic organoids by single-cell RNA sequencing (scRNA-seq) revealed that at the end of the differentiation this structure was composed by distinct ductal cell subpopulations lacking the acinar identity[15]. However, cell

[1]Precision Medicine Lab - Department of Transfusion Medicine, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. [2]Unit of Cell and Gene Therapies, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy. [3]Department of Radiation Oncology, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [4]Physical Biology Group, Buchmann Institute for Molecular Life Sciences (BMLS), Goethe Universität Frankfurt am Main, Frankfurt am Main, Germany. [5]IRCCS Ospedale San Raffaele, San Raffaele Diabetes Research Institute, Milan, Italy. [6]Department of Life, Health and Environmental Sciences, University of L'Aquila, L'Aquila, Italy. [7]Vita-Salute San Raffaele University, Milan, Italy.
✉e-mail: alessandro.cherubini@policlinico.mi.it; lorenza.lazzari@policlinico.mi.it

reprogramming is associated with several issues, such as the relative efficiency of the process, genetic aberration, and incomplete erasure of the epigenetic memory[16]. These issues thwarted the application of PSC-derived organoids in clinical research and regenerative medicine[17,18].

On the other hand, the ability to derive organoids from healthy and diseased primary progenitor cells, tissue-derived pancreatic organoids can be an unprecedented alternative source of functional cells for the investigation of human pancreas development, homoeostasis, and disease. We recently contributed to the field developing a scale-up strategy for the clinical translation of hPO production[19], but deep characterization of these structures is mandatory before their use as a reliable avatar to recreate physiological conditions that closely mimic the human three-dimensional pancreas.

In the present study, using single cell RNA-seq we generated a comprehensive molecular atlas, describing hPOs derived from human islet-depleted pancreatic tissue. This analysis shows that distinct ductal subpopulations, ranging from progenitor cells to mature duct, may coexist within the same sample. Consistent with previous literature suggesting that ductal progenitor cells are not restricted to a single cluster[20–22], we identified the existence of multiple duct subpopulations that retain progenitor capacity. Finally, our dataset was extensively compared with previously reported transcriptomic signature of human and mouse pancreatic ductal population[20,21,23], corroborating our findings.

Our gene signature of tissue-derived human pancreatic organoids is expected to be useful as guides for future clinical applications to better define the biology of the human pancreas in health and disease.

Results

Isolation and Genetic Stability in Culture of Human Pancreatic Organoids

Ductal fragments were isolated from healthy islet-depleted pancreatic tissue through mechanical dissociation[19]. When embedded in Matrigel and cultured with complete organoid medium to instruct signalling cues crucial for the growth of pancreatic epithelial cells (Fig. 1A), cells derived from ductal fragments established a mono-stratified pancreatic epithelium resembling well-defined apical–basal polarity with central hollow lumens, as shown by F-actin expression on the apical sides of the cells and E-cadherin expression at the lateral epithelial adherent junctions (Fig. 1B). This morphological organization was also illustrated by electron microscopic observation of lumens and basement membranes (Fig. 1C). After isolation, hPOs were split by mechanical disruption at 1:4 to 1:6 every 10–12 days, demonstrating the ability of the human pancreatic organoids to maintain their viability for more than 20 passages, overcoming the limitations of primary cell culture (Fig. 1D and Supplementary Fig. 1A).

Given that the hPOs could be expanded for more than 6 months, we compared the degree of senescence in them in early (3–5) and late (> 12) passages. To obtain direct evidence of cellular senescence, we determined the levels of senescence-associated β-galactosidase, which did not show differences in early and late passages (Fig. 1E). Then, we performed a polymerase chain reaction array designed to analyse a panel of genes related to senescence, which showed no global increase in the levels of these markers in long-term hPO cultures (Fig. 1F and Supplementary Fig. 1B). Consistent with these results, p21 expression did not increase significantly, and same gene expression and protein level of the cyclin-dependent kinase inhibitor $p16^{INK4a}$ were observed, confirming that the hPO proliferation ability was maintained through passages (Fig. 1G, Supplementary Fig. 1C). Cell cycle analysis confirmed this maintenance (Fig. 1H). Subsequently, we measured the levels of senescence-associated secretory phenotype (SASP), normally developed by senescent cells. We found a slight increase in interleukin (IL)-6 (fold change <2), and no significant increase in IL-8, or chemokine (C-X-C motif) ligand 1 (CXCL1/GROα), the most prominent cytokines in the SASP family, in long-term cultured hPOs (Supplementary Fig. 1D). Moreover, telomere length was preserved in hPOs cultured for 6 months (Supplementary Fig. 1E), suggesting that premature aging did not occur as expected in primary pancreatic cultures.

To exclude the dependence of the high hPO proliferation rate on genomic instability, we performed karyotype analysis, which confirmed that these organoids had a normal chromosome number after 4 months of in vitro expansion (Supplementary Fig. 1F). The genomic stability was also confirmed by nuclear foci of phosphorylated H2A.X staining, a specific molecular marker of DNA damage-induced cellular senescence (Supplementary Fig. 1G). In addition, a comparative genome hybridization array showed no considerable structural aberration in long-term cultured hPOs compared to freshly isolated cells (Supplementary Fig. 2). Taken together, these data suggest the ability of the human pancreatic organoids to support the long-term expansion, resembling in vivo pancreatic epithelium with no significant degenerative senescence, even during late passages.

Single-cell transcriptome analysis reveals expression of ductal and proliferation markers in human pancreatic organoids

To fully characterize the hPO cell composition at the transcriptional level, we employed single-cell RNA-seq technology using the 10× Genomic Chromium system on hPOs cells derived from three independent donors at passage 5 (P5). After the application of stringent quality control criteria and filtering according to multiple criteria (see Methods), the three datasets were integrated, yielding 3187 unique single-cell transcriptomes with an average of 2166 genes and 5733 transcripts per cell (Fig. 2A and Supplementary Fig. 3A). Interestingly, performing cell cycle analysis we identified that 1147 cells (~ 36% of total cells) were involved in S and G2/M phases (Fig. 2B). Clustering of cells revealed six transcriptional distinct clusters that were conserved across samples with similar variability in cell frequency (Fig. 2C, D, Supplementary Fig. 3B–D), supporting the reproducibility of our culture system.

A widespread expression of epithelial ductal markers such as epithelial cell adhesion molecule (*EpCAM*), cytokeratin 19 (*KRT19*), SRY-box transcription factor 9 (*SOX9)* and claudin2 (*CLDN2)* (Fig. 2E) was observed across the entire population. On the other hand, acinar and endocrine counterpart were not observed in the hPO dataset, as indicated by very low expression of the archetypal marker carboxypeptidase A1 (*CPA1*), glucagon (*GCG*), somatostatin (*SST*) and insulin (*INS*), while insulin pancreatic polypeptide (*PPY*) was not detected (Supplementary Fig. 3E). Finally, clusters 3 and 4 showed high expression levels of proliferation markers such as *CENPM*, *PCNA*, *MKI67* and *TOP2A* (Fig. 2F), confirming the proliferation state of these cells. GO analysis of genes differentially expressed by these clusters confirmed enrichment for terms such as DNA metabolic process and replication, mitotic spindle organization and chromosome segregation (Supplementary Fig. 3F,G).

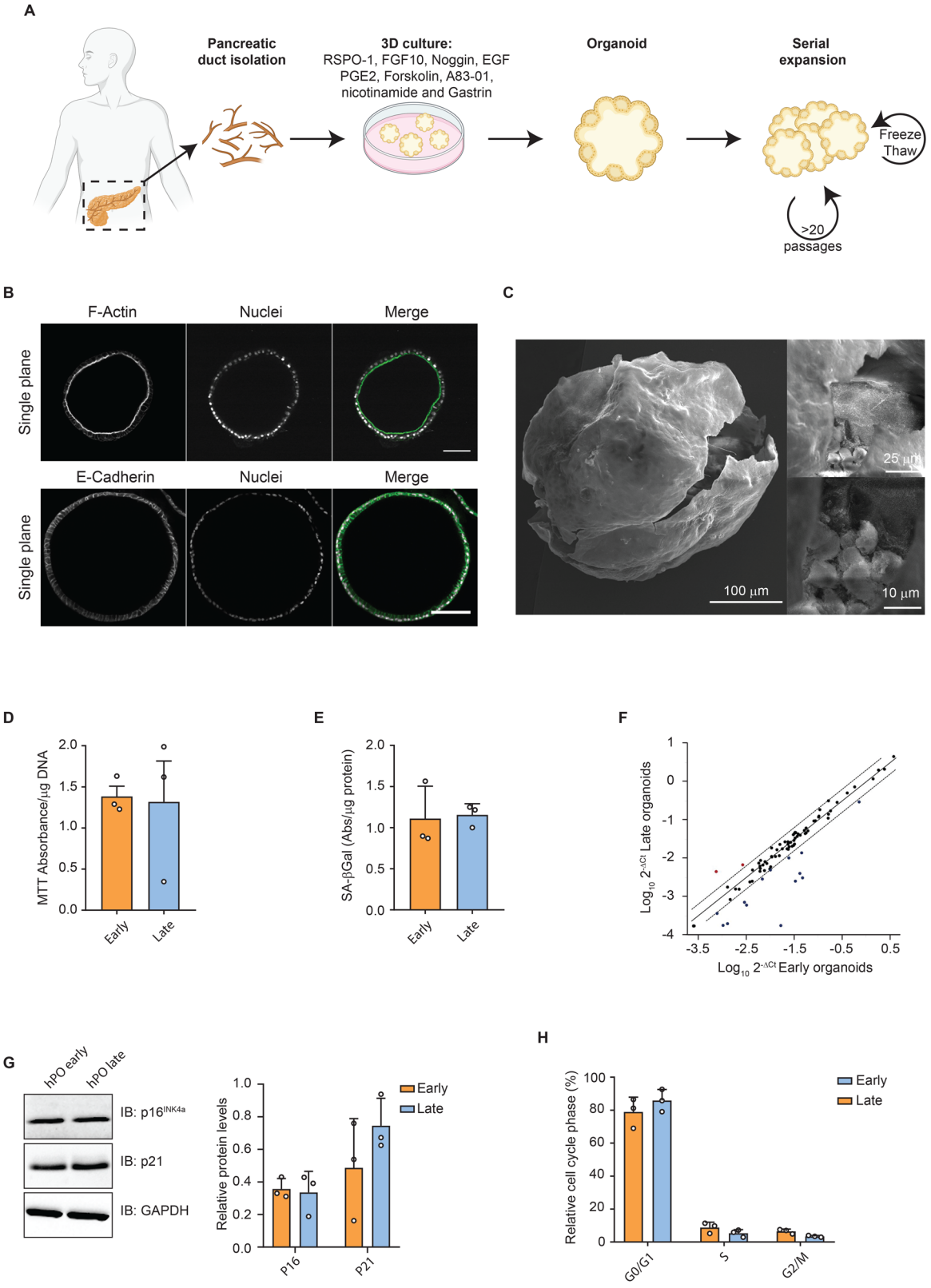
These findings were corroborated also at protein level by immunofluorescence staining for ductal and proliferation markers SOX9 and MKI67 respectively (Fig. 2G, H).

So, we can conclude that human pancreatic organoids displayed a duct-like phenotype, and only a cell subset proliferates, maintaining the organoid expansion in culture over 6 months as previously reported[24].

Human pancreatic organoids are characterized by distinct ductal subpopulations with unique gene signatures and regulation of pathways

To get a better understanding heterogeneity in pancreatic organoids, we re-clustered cells (derived from previous cluster 0, 1, 2 and 5), obtaining seven transcriptional distinct ductal clusters (Fig. 3A, B and Supplementary Fig. 4A). DEGs analysis revealed few genes differentially expressed between all clusters (Supplementary Data). However, dendrogram and Spearman correlation analysis showed a close relationship between clusters 0 and 1, as well as clusters 2 and 3 (Fig. 3C, D), suggesting a shared core gene expression programme between these clusters. On the other hand, cluster 4, 5 and 6 have their own identity diverging in separated branches (Fig. 3C).

Cluster 0 showed a significant upregulation of the mucin-related genes *MUC5AC*, *MUC5B*, *PHGR1*, *SPINK4* as well as Trefoil family factors -1, -2 and -3 (TFF1, TFF2 and TFF3) (Supplementary Fig. 4B and Supplementary



Data), and these features resulted more evident when compared to cluster 1 (Fig. 3E). Furthermore, GO analysis reveals terms related to mucin secretor-like cells, epithelial structure maintenance and secretory granule membrane (Fig. 3F). In line with the abundant secretion of TFFs protein, which is a feature of mucus-secreting cells of the gastrointestinal tract that promote epithelial repair by cell migration (supported, also, by gene ontology terms of epithelial structure maintenance) we identified cluster 0 as pancreatic duct stem cell niche.

Cluster 1 showed only six genes differentially expressed when compared to all the other clusters. Among them ductal genes *CLDN2* and *AQP3* have a significant high expression (Supplementary Fig. 4C). CLDN2 plays a crucial role in the tight junction compartment, forming an epithelial barrier

Fig. 1 | Primary human pancreatic organoids show no senescence-related phenotype in long-term culture. A Schematic representation of the human pancreatic organoids isolation protocol. Created in BioRender. Cherubini, A. (2024) BioRender.com/m35c022. **B** Confocal images of day-12 organoids immunostained for the apical marker F-actin (green), the adherent junction marker E-cadherin (green), and DAPI (blue). Scale bars, 200 µm. **C** Scanning electron microscopic image of a partially opened human pancreatic organoid showing its 3D architecture and basal and apical ultrastructure. Detailed views show the apical surfaces of secretory cells. **D** Cell viability of hPO early and late passages measured by MTT assay normalized on µg of DNA. Data are means ± SD ($n = 3$ biologically independent samples). **E** senescence-associated beta galactosidase (SA-βGal) activity measurement. Data are means ± SD ($n = 3$ biologically independent samples). **F** Scatterplots of the PCR array comparing gene expression data of hPO early and late passages. Dashed lines delimitate fold changes (FC) of genes up- (red) and downregulated (blue) in hPO late with respect to hPO early passage. **G** Western blots of $P16^{INK4a}$ and P21 in hPOs in early and late passages; GAPDH was used as a loading control. Signal quantification is reported. Data are means ± SDs ($n = 3$ biologically independent samples). **H** Cell-cycle profiles of hPOs in early and late passages. Data are means ± SD ($n = 3$ biologically independent samples).

to regulate physiologic secretion and preventing pathogens invasion[25,26]. Aquaglyceroporin 3 (*AQP3*) regulates water and small solutes transport across epithelial barrier[27], maintaining host homoeostasis. When compared to cluster 0, cluster 1 resulted enriched in gene related to immune response, such as *CXCL1* and GO analysis revealed terms related to an epithelial protection and maintenance (Fig. 3E, F) supporting the strong relationship between these two clusters.

Cells in cluster 2 have a significant upregulation in the expression of osteopontin/secreted phosphoprotein 1 (*SPP1*), compared with both the nearest cluster 3 and all the other clusters (Fig. 3G and Supplementary Fig. 4D). This gene was previously identified to pinpoint pancreatic ductal cells enriched by progenitor cells both in mouse and humans[21,28]. Cluster 2 had also high levels of Inhibitor of Differentiation gene family (*ID1-3*), defined as pancreatic progenitor cell markers[29]. ID's gene family inhibit the function of basic helix–loop–helix transcription factors playing a critical for pancreatic development[21,30]. Importantly, ID2 has been shown to prevent NeuroD expression promoting the exocrine epithelial maintenance upon injury[29]. Furthermore, GO analysis showed an enrichment of terms related to stemness such as glycolytic process, regulation of cell proliferation and positive regulation of cell differentiation (Fig. 3H).

The near cluster 3 is characterized by the highest expression of N-Myc downstream regulated gene 1 (*NDRG1*) gene in the entire dataset (Supplementary Fig. 4E). NDRG1 is a stress-activated epithelial protein implicated in cellular differentiation[31,32]. Furthermore, also the expression of *TFF1* and *S100A14*, genes associated with cell motility resulted relatively elevated (Fig. 3G and Supplementary Fig. 4E). Interestingly, cluster 3 showed also high levels of genes related to hypoxia response, such as EGLN3 and ERO1A (Supplementary Data). Indeed, low oxygen tension or hypoxia is a common feature of stem cell niche regulating the plasticity of progenitor cells[33]. These findings were corroborated also by GO analysis, where terms related to hypoxia response and cell proliferation and differentiations resulted enriched (Fig. 3H).

Cluster 4 is represented by cystic fibrosis transmembrane conductance regulator (*CFTR*) and carbonic anhydrase II (*CA2*) (Fig. 3I). These genes were previously identified as label of intercalated ductal cells[34–36]. Intercalated ducts are composed by small specialized epithelial cells connecting the secretory end-pieces to the second part of the duct system. Here, these cells, secrete an isotonic and highly alkaline fluid containing bicarbonate ions to collect digestive enzymes produced by acinar cells[36,37]. Furthermore, CA2, marker of centroacinar cells, catalyses the generation of $HCO3^-$ from CO2 to neutralize pancreatic enzymes entering in contact with ductal cells[35]. GO analysis showed terms related to normal ductal pancreatic functions, such as transport of different compounds and bicarbonate transport (Fig. 3J), supporting that cluster 4 possess a certain degree of ductal cell functions like to intercalated ductal/centroacinar cells.

Cells in cluster 5 have significantly upregulated expression of multiple cytokeratin, such as *KRT18*, *KRT8*, *KRT23* and *KRT7* (Fig. 3K and Supplementary Data), which are markers of terminally differentiated ducts[38,39]. Furthermore, cluster 5 showed a considerable expression of tubulin alpha 1a (*TUBA1A*) compared to all the other clusters (Fig. 3K). TUBA1A protein is the major component of the duct cilium, which plays an essential role in pancreatic tissue organization, as well as flow sensor activity[40,41]. GO analysis showed an upregulation of regulation of cell migration and cytoskeleton organization (Fig. 3L), supporting mature ductal nature of these cluster.

Finally, the small cluster 6 was composed only by 18 cells (~ 0.9%) expressing some ductal markers but could not be assigned to a specific identity (Supplementary Data).

The presence of specific cell type markers for each cluster and the intra-organoids heterogeneity were confirmed using immunofluorescence analysis of whole-mount pancreatic organoids (Fig. 4 and Supplementary Fig. 5).

Taken together, our high-resolution single-cell analysis characterized the substructure of human pancreatic organoids revealing distinct duct subpopulations harbouring progenitor cells able to sustain the structure in culture.

Comparison with human pancreas scRNA datasets confirms the presence of distinct duct subpopulations in human pancreatic organoids

To support the exocrine identity of hPOs, we integrated our organoid datasets (without cycling cells) with three scRNA-seq data sets from primary human pancreas tissue[23,42,43]. All pancreatic cell types were largely preserved after clustering and plotting of the combined datasets (Fig. 5A, Supplementary Fig. 6A). Although, cells derived from hPO dataset seems to re-clustered individually, indeed cluster 1 was composed principally by hPO cells (~ 93%, Fig. 5C and Supplementary Fig. 6B), dataset alignment showed a high overlap with the exocrine compartment (Fig. 5B). Interestingly, cluster colour coding of our dataset revealed that cluster 4 relapse between the ductal and acinar compartments, confirming its intercalated duct/centroacinar identity, while cluster 5, previously named as mature duct, entirely relapses in the ductal compartment (Supplementary Fig. 6C). Finally, no cells of hPO dataset correspond to endocrine or acinar cells (Supplementary Fig. 6D).

To confirm our ductal cell subpopulation analysis without data integration, we also used AddModuleScore in Seurat, comparing our hPO clusters to the previously reported ALK3+ clusters[21]. In this case, cluster 2 showed the highest score when compared to human ALK3+ clusters 1 (*SPP1*+ stress/harbouring progenitor-like cells), while cluster 3 showed the highest score when compared to ALK3+ cluster 2 (*TFF1*+ activated/migrating progenitor cells), sustaining the exocrine progenitor identity of cluster 3 and 4 (Fig. 5D and Supplementary Fig. 7A). Mature duct cells (cluster 5) showed the highest score when compared to ALK3+ cluster 3 (*AKAP12*+ small ducts) confirming the terminally differentiated duct nature of this cluster (Supplementary Fig. 7A). Intriguingly, ALK3+ cluster 5 identified as Ducto-Acinar showed a higher homology with hPO cluster 4, reinforcing its intercalated duct/centroacinar identity (Supplementary Fig. 7A). The human $ALK3^+$ cluster 4 (*OLFM4*+ transitional to acinar 1) and cluster 6 (*CPA2*+ transitional to acinar 2) showed the lowest homology score compared to our hPO clusters.

Finally, we compared each of murine duct clusters previously published[20] with our clusters. Murine pancreatic duct clusters 0 and 2, indicated as enriched in progenitor cells, showed the highest scores when compared to hPO cluster 2 (Supplementary Fig. 7B). Murine intercalated duct (cluster 1) showed the highest score when compared to the human hPO cluster 4, while murine cluster 3 and 4 (intrapancreatic bile duct and pancreatobiliary duct respectively) showed the highest score when compared to mature duct cluster 5 of human hPOs (Supplementary Fig. 7B).

Taken together, these results suggest that hPOs recapitulate distinct ductal subpopulations and harbour pancreatic ductal progenitor cells that could sustain hPO self-renewal.

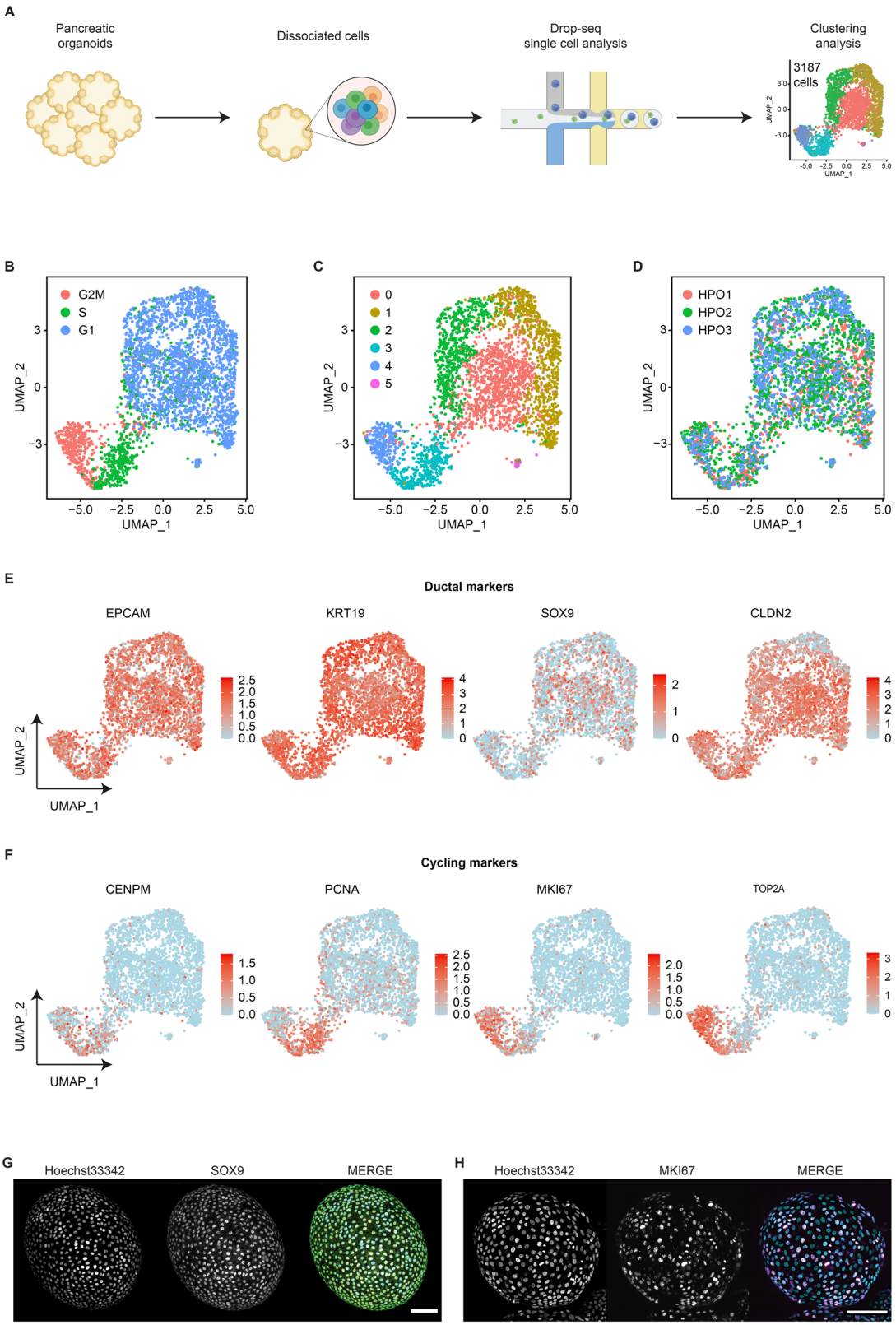


Fig. 2 | Primary human pancreatic organoids show a broad expression of ductal markers and a presence of a subset of cycling cells. A Schematic representation of the experimental workflow, depicting the collection and processing of human pancreatic organoids for scRNA-seq. Created in BioRender. Cherubini, A. (2024) BioRender.com/e66i954. **B** UMAP plot of expanded hPO from three distinct donors ($n = 3187$ cells) labelled by cycling state. **C** UMAP plot showing cell clusters of the primary human pancreatic organoid scRNA-seq data. **D** UMAP plot coloured on the basis of hPO line. **E** UMAP plots indicating the expression of representative ductal markers. **F** UMAP plots indicating the expression of representative markers associated with proliferation. **G** Confocal image of day-12 organoids immunostained for the ductal marker SOX9. Scale bar, 100 µm. **H** Confocal image of day-12 organoids immunostained for the proliferation marker MKI67. Scale bar, 100 µm.

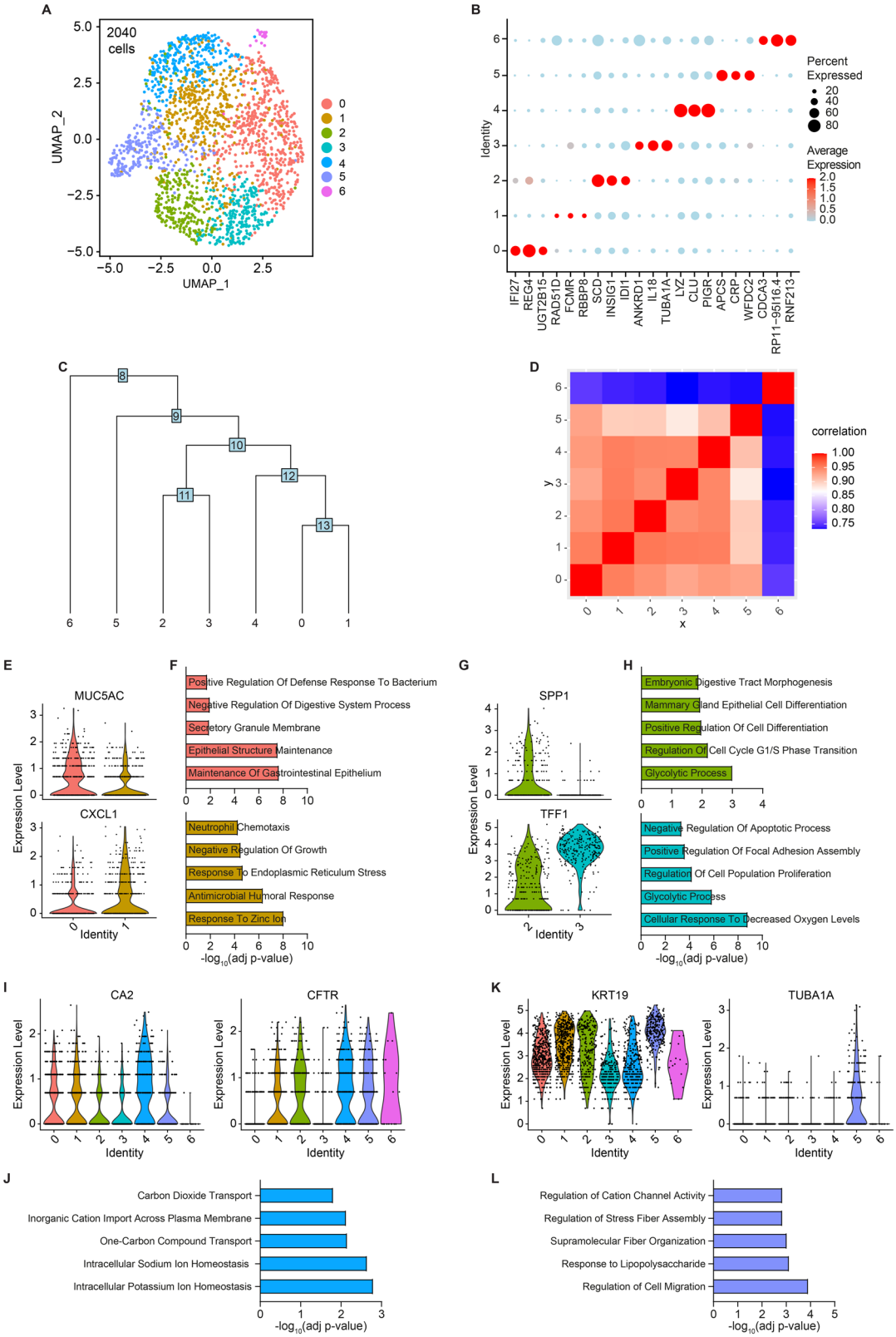


Fig. 3 | Single-cell RNA sequencing pinpoints a ductal heterogeneity in primary human pancreatic organoids. **A** UMAP plot depicting the re-clustering of organoids without cells in G2 phase. **B** Dot plot of the three dot marker genes for each cluster. Rows indicate clusters and columns indicate genes. **C** Cluster dendrogram created using PlotClusterTree function shows the Euclidean relationships between clusters. **D** Heatmap of the Spearman correlation matrix calculated using average gene expression for each cluster. **E** Stacked violin plots show two differentially expressed genes (*MUC5AC* and *CXCL1*) in cluster 0 vs. 1. **F** Five altered biological processes from Gene Ontology comparing cluster 0 vs. 1 are depicted. **G** Stacked violin plots show two differentially expressed genes (*SPP1* and *TFF1*) in cluster 2 vs. 3. **H** Five altered biological processes from Gene Ontology comparing cluster 2 vs. 3 are depicted. **I** Violin plots show two differentially expressed genes (*CA2* and *CFTR*) in cluster 4. **J** Five biological processes from Gene Ontology enriched in cluster 4. **K** Violin plots show two differentially expressed genes (*KRT19* and *TUBA1A*) in cluster 5. **L** Five biological processes from Gene Ontology enriched in cluster 5.

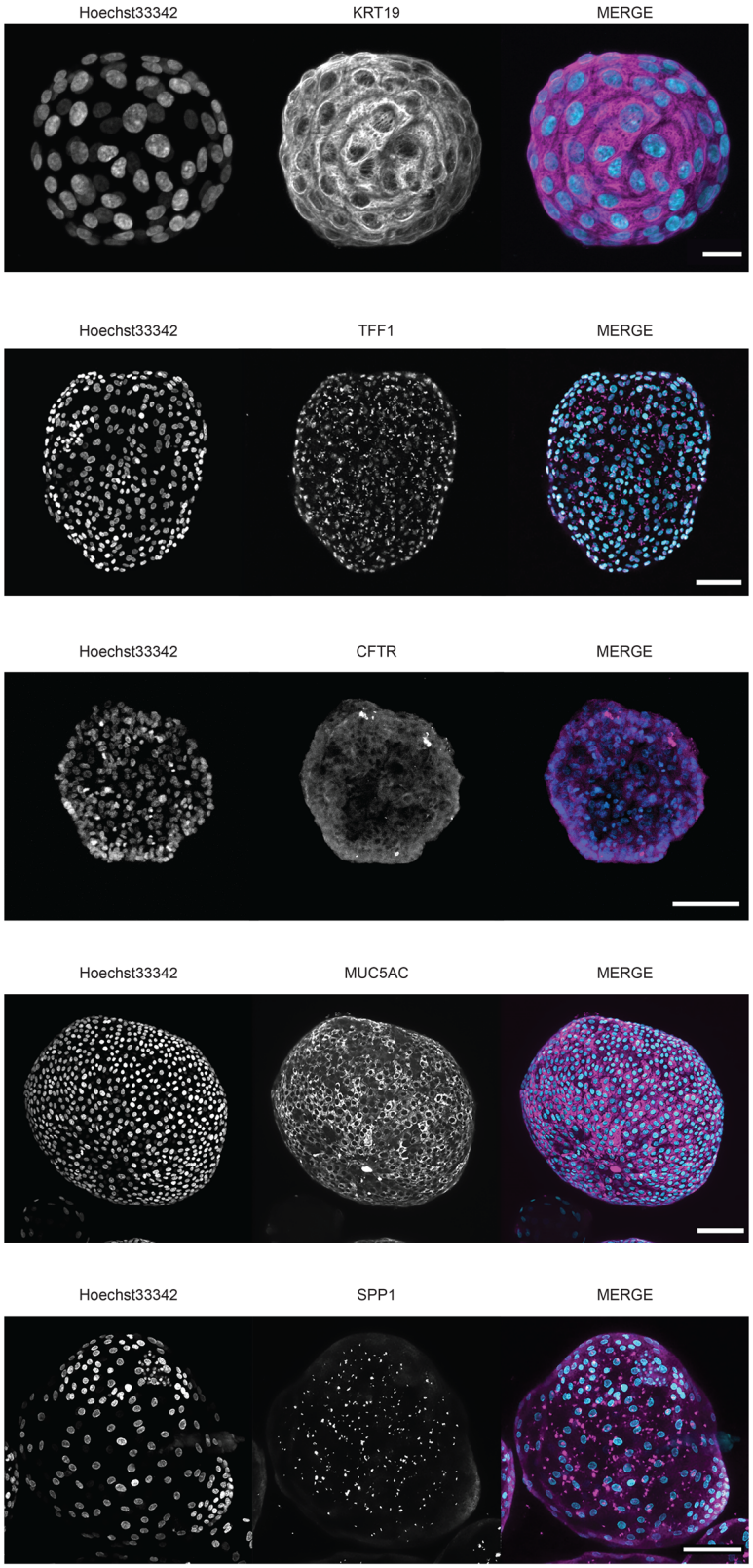


Fig. 4 | Ductal subtype-specific markers located within primary human pancreatic organoids. Confocal images of day-12 organoids immunostained for the ductal subtype markers (KRT19, TFF1, CFTR, MUC5AC and SPP1) identified by scRNA-seq analysis. Scale bar, 100 μm.

Single-cell trajectory analysis revealed a cellular heritage of pancreatic progenitor cells toward all hPO ductal subpopulations

Considering the high content of progenitor cells, we evaluated their differentiation potential using RaceID3/StemID2 algorithms able to identify cell lineage trees and predicts multipotent cell identities[44]. Among the 17 clusters generated by RaceID3 in an unsupervised manner (Fig. 6A), cluster 17 enriched in ID and TFF family genes (Supplementary Fig. 8A) showed the highest StemID2 score and correlated to cells present in Seurat duct clusters 2 and 3 (Fig. 6B and Supplementary Fig. 8B).

To infer the lineage relationships among hPO ductal subpopulations, we used Monocle3[45] to order the cell clusters along a continuous pseudo-temporal trajectory based on their transcriptional similarity. Monocle3 analysis showed that cluster 6 resulted disconnected from the main

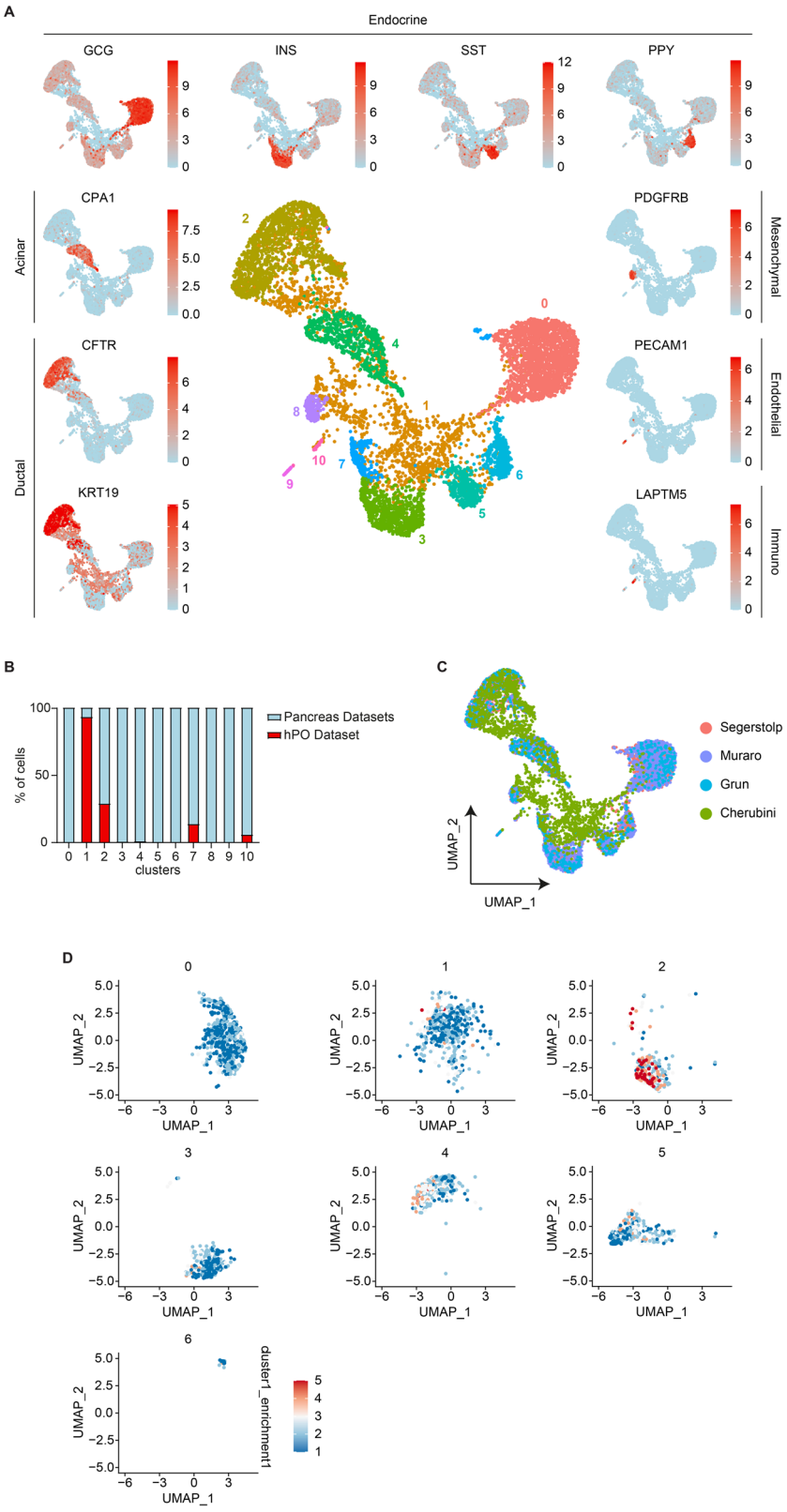


Fig. 5 | Alignment against human whole pancreas transcriptome datasets. A Integration of primary pancreatic organoids with three primary pancreas scRNA-seq data sets[23,42,43] into 11 Louvain clusters (middle plot). UMAP expression plots display cell type-specific marker genes for each cluster (surrounding plots). Scale bars for each marker denote average expression in gene counts. **B** UMAP plot showing distribution of pancreatic organoid cells (green colour) vs. the whole-pancreas datasets. **C** Stacked bar plots showing the distribution of cells derived from different sample origins in each cell cluster. **D** Enriched expression of the marker genes for the quiescent progenitor cells defined in Qadir et al.[21] in primary human pancreatic organoids clustering analysis.

pseudotime trajectory (Supplementary Fig. 9A), so we kept our analysis focusing on clusters 0, 1, 2, 3, 4 and 5. Clusters with highest StemID2 score were considered as root node of the trajectory and pseudotime path were generated (Fig. 6C, D and supplementary Fig. 9B). Interestingly, leucine-rich repeat-containing G-protein coupled receptor 4 (LGR4), a marker of stemness, resulted enriched at the beginning of the pseudotime, while markers of differentiation such as MUC5AC, CFTR and KRT23 increased with the pseudotime (Supplementary Fig. 9C).

We next explored genes with similar patterns of expression that vary across the pseudotime trajectory using Moran's I test for spatial autocorrelation. Hierarchical clustering analysis revealed that module 19 and 1 showed remarkably higher expression levels in cluster 2 and 3, respectively

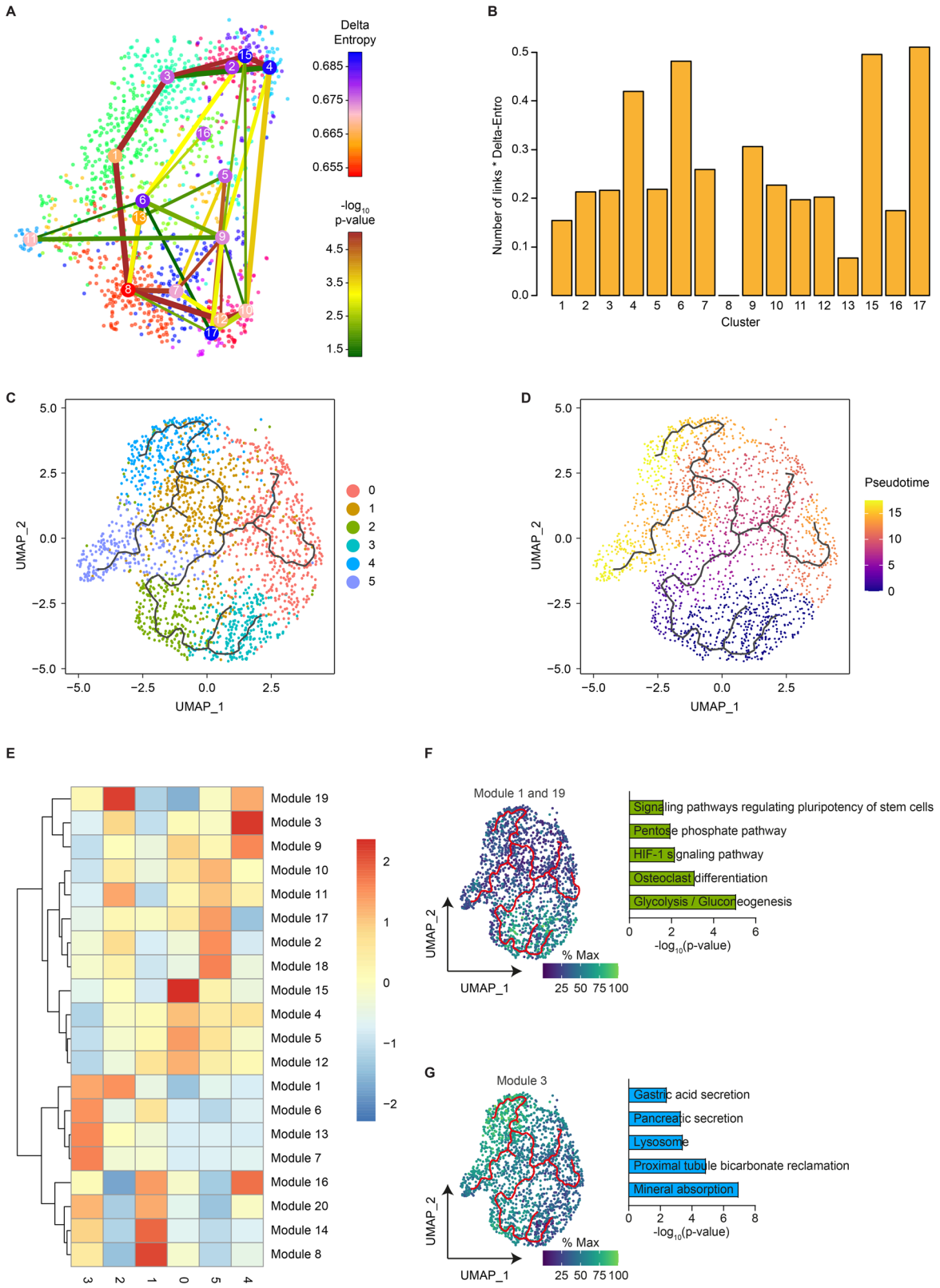


Fig. 6 | Primary human organoid progenitors show a transitional trajectory able to differentiate in all ductal cell subtypes. A Organoid lineage tree inferred by StemID2 is shown in the RaceID3 clusters. Only significant links are shown ($P < 0.01$). Node colour represents the level of transcriptome entropy, edge colour describes the level of significance, and edge width describes link score. **B** Barplot of StemID2 scores for RaceID3 clusters. Cluster 17, which shows highest expression of progenitor markers, such as SPP1 and TFF1, receives the highest StemID2 score. **C** Monocle 3 UMAP and trajectory of primary human pancreatic organoids duct clusters 0–5. **D** Each cell's relative pseudotime value is depicted that is a measurement of the distance between its position along the trajectory and the starting point (cluster 4). **E** Expression changes of the modules generated by Monocle 3 analysis are shown for each cluster. **F, G** Expression of modules 1, 19, and 13 along trajectory and relative top 5 deregulated gene ontology biological processes terms.

(Fig. 6E, F). GO analysis showed terms related to stem cell differentiation confirming that these two clusters might represent the progenitor cells of pancreatic organoids. Finally, enrichment of terms related to pancreatic secretion and cytoskeletal organization were upregulated in cluster 4 and 5 when compared to progenitor clusters in module 3 and modules 2/18 respectively (Fig. 6G and Supplementary Fig. 9D), confirming the distinct duct identity with specific physiological function of these two clusters.

Taken together, our results suggest that hPOs recapitulate pancreatic differentiation process similarly to that occurring in the in vivo environment[46,47], with the activation of multipotent progenitor cells after tissue injury for the maintenance of stem cell niche and regeneration of distinct ductal populations.

Estimation of cell-to-cell communication in human pancreatic organoids

It is generally well known that cells communicate with each other by sending and receiving signals within the environment. To characterize cell-to-cell communication between different cell type in human pancreatic organoids, we interrogated ligand–receptor (L–R) interactions using CellChat tool[48]. The results revealed a complex cell-to-cell interaction network within the ductal subpopulation (Fig. 7A), composed by 537 significant ligand-receptor pairs categorized into 17 major signalling networks. Based on permutation testing of randomized network connections, cluster 5 showed the largest number and weighting of significant inbound connections (Fig. 7B), presenting strong interactions with itself, but also with the progenitor clusters 2 and 3 (Fig. 7C). Furthermore, at the incoming end of signalling, cluster 5 was driven by communication patterns involved in adhesion molecule including laminin, cadherin and thrombospondin (Supplementary Fig. 10), reinforcing the mature ductal identity of these cluster.

Interestingly, cluster 3 acts as the major receivers from multiple cells in epidermal growth factor (EGF) signalling pathway (Fig. 7D, E and Supplementary Fig. 10), which was reported to play a pivotal role in activation and differentiation of foetal and adult pancreatic progenitors in mice and humans[49,50]. Network centrality analysis of the inferred EGF signalling pathway identified that cells related to mature duct (cluster 5) were the most abundant source of ligands (Fig. 7E). In addition, we identified that cluster 3 were the only influencer for EGF intercellular communications, while any cell type act as mediator (Fig. 7E), reinforcing the central role played by EGF signalling pathway in pancreatic progenitor cells. Among all known ligands of EGF signalling, CellChat analysis revealed the regulation of this pathway by ligand-receptor pair AREG-EGFR (Fig. 7F), that was noticeably expressed in epithelial cells.

Finally, cluster 2 resulted the main sender of SPP1 signalling pathway (Fig. 7G, H), and the main receiver of this ligand was cluster 4 and 5, two clusters characterized as terminally differentiated. Indeed, SPP1 has been previously shown to define duct cell type enriched in progenitor capacity[21,28], and it was released by these cells to support the human pancreatic duct cell maturation and function[20]. Furthermore, the signalling pathway was regulated by CD44 receptor (Fig. 7I), that it is well known to transduce inner cellular signals for duct cell differentiation and maturation.

These data suggested that cells in human pancreatic organoids coordinately operate to sustain the long-term proliferation of these structures, regulating the differentiation in distinct duct subpopulation.

Discussion

The human pancreas is a complex organ composed of both exocrine glands and endocrine cells playing diverse functions, including enzyme production for digestion of nutrients and hormone secretion to control the glucose homoeostasis. Pancreatic insufficiency, which can affect the functioning of human body, can be caused by a variety of clinical conditions, including pancreatitis, autoimmune disease, diabetes and pancreatic cancer. In the last years, advances in the organoid technology, allowing to create 3D models able to recapitulate heterogeneity, structure and functions of the primary tissue, have revolutionized human disease modelling by bridging the gap between traditional two-dimensional cell cultures and real tissues, potentially eliminating ethical and technical issues associated with animal models and minimizing animal suffering[51]. These 3D mimics offer several benefits, including a deeper understanding of drug mechanism of action, faster and cost-effective read-outs, reduce animal use, and improved patient response prediction. In the context of pancreas diseases, the aim of completely exploiting the potential of 3D models cannot be achieved without deeper knowledge of their cell composition and the degree to which they produce differentiated pancreatic cell types. Currently, tissue-derived hPOs have been described as homogeneous duct-like structures due to the remarkable KRT19 and SOX9 expression and the apical-basal polarity[24,52–54]. However, the adult pancreatic duct system resulted composed by different cell types with specific physiologic functions including mature intralobular duct, intercalated duct/centroacinar cells. In this work we propose a new characterization of pancreatic organoids generated from human primary progenitor cells. We presented a single-cell level transcriptomic dataset of tissue-derived human pancreatic organoids, unveiling that these structures are able to recapitulate all the ductal subpopulations panorama.

In the adult human pancreas, progenitor population has been predicted to reside in the exocrine compartment, in a niche among the ductal population called pancreatic duct gland (PDGs)[21,23,47]. PDGs are bud-like structures surrounding the pancreatic ductal tree that, similar to stem cells of the gastric and colonic crypts, can be activated upon pancreatic injury. PDG compartment possesses a specific architecture and gene expression consistent with a stem cell niche including markers such as trefoil-gene family (TFF1-3), and several mucins[21,47], that makes them distinct from pancreatic duct epithelium. Similar to gastrointestinal (GI) crypt/gland, PDG contain different cell populations, including adult quiescent and activated stem cells, as well as cells with a protection role of the stem cell niche[55].

Consistent with previous literature, the human pancreatic ductal progenitor cells were not restricted to a single cluster[21,22], here we describe two distinct clusters, 2 and 3, as pancreatic duct progenitors due to the enrichment in genes commonly associated with these cell types. More precisely, cluster 2 showed high expression of ID1-4 genes, which are master regulators of stem cell fate, differentiation, and proliferation during development and in adult life[30,56,57]. Interestingly, the marker SPP1 was also enriched in cluster 2 that has been shown to be expressed in proliferative pancreatic duct progenitors in mice[28,58], as well as in quiescent progenitor cells in humans[21,59]. On the other hand, cluster 3 resulted enriched in the expression of TFF genes, a family member of proteins that play an essential role in the migration of progenitor cells from the PDG compartment to the main duct[47]. Cluster 3 resulted also strongly enriched by the metastasis suppressor NDRG1, previously reported as inhibitor of stem cells in numerous cancers, including pancreatic cancer[60,61]. Furthermore, in line with the relatively hypoxic environment[62,63], cluster 3 showed a high expression of oxygen-related sensor genes that allow stem cells to maintain an undifferentiated status and a multilineage differentiation potential.

Finally, cluster 0 strongly enriched in mucin, as well as TFF family, may act as pancreatic stem cell niche to maintain duct homoeostasis. Indeed, mucus-secreting cells of the gastrointestinal tract and the pancreatic duct secrete TFFs abundantly onto the mucosal surface to promote epithelial repair by cell migration[47,64,65].

To clarified the maturation and lineage relationships between the hPO cells we used a pseudotemporal trajectory analysis. Based on the pseudotemporal ordering of cells proposed by the Monocle 3 software, together with knowledge of mouse and human pancreatic development, we can infer evidence of cell differentiation in the hPOs. Clusters previously identified as progenitor cells (cluster 2 and 3) were located at the beginning of the differentiation trajectory, with strong enrichment of genetic markers related to this population (e.g., ID1, and LGR4)[66]. As in response to pancreatic injury, the organoid progenitors began to express markers of activated and migrated stem cells and could be differentiated toward distinct exocrine cell types. This knowledge supports our hypothesis that pancreatic progenitors could be stimulated upon injury, becoming an activated and migrating stem cell population, represented in the hPO structures by the cluster 3 and,

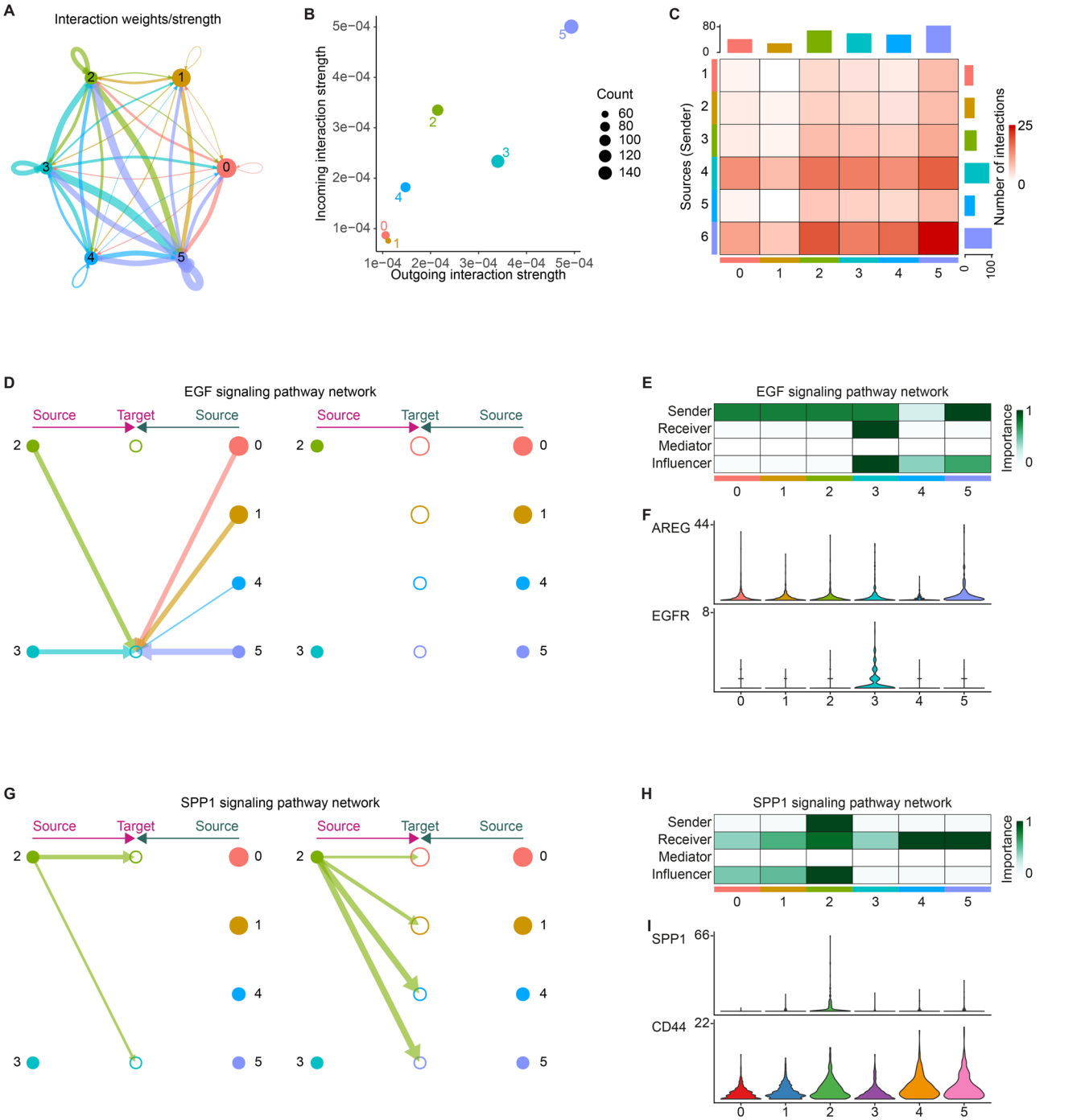


Fig. 7 | Estimation of cell–cell communication within primary human pancreatic organoids using CellChat. A Circle plots displaying putative ligand-receptor interactions, with the width of edges representing the strength of the communication. **B** Scatter plots comparing the outgoing and incoming interaction strength in the 2D space in primary human pancreatic organoids. **C** Heatmaps of the differential number of interactions showing the outgoing and incoming signalling change of each cluster. The top coloured bar plot represents the sum of each column of values displayed in the heatmap (incoming signalling). The right coloured bar plot represents the sum of each row of values (outgoing signalling). **D** Hierarchical plot shows the inferred intercellular communication network for EGF signalling pathway. Left and right panels highlight the autocrine and paracrine signalling respectively. **E** Network centrality scores of the EGF signalling pathway for each cluster. **F** Stacked violin plots showing the expression levels of ligand-receptor pair included in the EGF signalling pathway in the different clusters. **G** Hierarchical plot shows the inferred intercellular communication network for SPP1 signalling pathway. Left and right panels highlight the autocrine and paracrine signalling respectively. **H** Network centrality scores of the SPP1 signalling pathway for each cluster. **I** Stacked violin plots showing the expression levels of ligand–receptor pair included in the SPP1 signalling pathway in the different clusters.

walking through the pseudotemporal trajectory, activated progenitors could differentiate in distinct ductal subpopulations represented by mucine (mucus-secreting cells), CFTR and CA2 markers (intercalated duct/centroacinar), as well as several keratins at the end (mature duct).

Within a multicellular environment, cell-to-cell communication plays a fundamental role to regulate cell processes regulating tissue function, homoeostasis and pathophysiology. Although, cell signalling ultimately occurs at the protein levels, we infer intercellular relationship between cell

population in pancreatic organoids based on gene expression of ligands and their receptors and cofactors. Interestingly, our analysis identified cluster 2 as the main source of secreted phosphoprotein 1 (SPP1, also known as osteopontin). This finding is consistent with previously published papers in which SPP1 is reported to specifically mark pancreatic progenitors and proliferative ducts in humans and mice[21,22,28]. By analysing each ligand-receptor pair of SPP1 signalling between cluster 2 and the other clusters, we identified that SPP1-CD44 was the most elevated pathway sent from cluster 2 to cluster 4 (intercalated duct/centroacinar) and cluster 5 (mature duct but also regulate pancreatic ductal maturation[20].

In summary, our findings demonstrate that tissue-derived pancreatic organoids recapitulate distinct duct subpopulations, retaining the characteristics of the tissue of origin. We better define the progenitor population able to maintain the expansion of pancreatic organoids overtime, maintaining distinct ductal identity and genomic integrity.

These results shed new light on the knowledge of a complex and still uncharacterised 3D model, providing a strong foundation for future studies of pancreas development, as well as pancreatic modelling.

A limitation of our study is that a mixed of ductal cells from intercalated, interlobular, intralobular, terminal ductal and centroacinar cells, without a selection by picking, are included in the organoid cultures. Therefore, we cannot determine whether the organoid heterogeneity is determined by a single progenitor able to differentiate into all the ductal types or by different types of ducts from the beginning. However, considering the recent publications[53,67,68], we can consider that pancreatic organoids, when in culture, are originated only by a small portion of cells, presumably by progenitor cells, that are present along ductal structure in native tissue. Also, future functional experiment will be useful to understand how ductal cells within pancreatic organoids communicate with each other to support proliferation and differentiation, maintaining the heterogeneity for long time.

Materials and methods

Ethics

All experiments were performed according to the amended Declaration of Helsinki. Informed written consent was obtained from each patient and the use of human specimens was approved by the Ethical Committee of Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico (n° 1982, 14th January 2020) and by the Ethical Committee of IRCCS Ospedale San Raffaele (IPF002-2014, 6th June 2014). Human healthy pancreatic tissues were obtained from the Diabetes Research Institute, IRCCS Ospedale San Raffaele, Milan, Italy from multi-organ donors. Human islets were isolated as previously described[69] at the Pancreatic Islet Processing Unit, a National Transplant Center accredited facility (IT000679, https://webgate.ec.europa.eu/eucoding/reports/te/index.xhtml).

Human pancreatic organoids isolation and culture

Islet-depleted pancreatic tissue was mechanically dissociated by the GentleMACS™ Dissociator (Miltenyi, Bergisch Gladbach, Germany) using the software programme "m_spleen04" (4X). Whole pancreatic tissue fragments were embedded in Matrigel growth factor reduced. When Matrigel was solidified, culture medium was added. The human complete culture medium was composed of Advanced DMEM/F12 medium (Thermo Fisher Scientific, Waltham, MA) supplemented with 10 mM HEPES, 2 mM L-Glutamine (all Lonza, Basel, Switzerland), 1% B27 without vitamin A, 1% N2 supplement xeno-free (all Thermo Fisher Scientic), 1 mM N-acetyl-L-cysteine, 10 nM [Leu15]-Gastrin I Human, 10 mM Nicotinamide (all Sigma-Aldrich, St Louis, MO), 1 µg/ml human recombinant protein R-Spondin 1 (RSPO1), 0.1 µg/ml human Noggin recombinant protein, 50 ng/ml human EGF, 100 ng/ml FGF10, (all Peprotech, Cranbury, NJ), 10 µM Forskolin, 0.5 µM A83-01 and 3 µM Prostaglandin E2 (all Tocris, Bristol, UK). After isolation, medium was changed every 3-4 days and hPOs were split with 1:4–1:6 ratios every 10–12 days for at least 6 months. All the hPOs were routinely tested to exclude mycoplasma contamination.

Immunostaining

To perform immunofluorescent staining, hPOs were removed from Matrigel with Cell Recovery Solution (Corning), briefly washed with PBS and fixed with 4% PFA (Electron Microscopy Science, Hatfield, PA) in PBS for 40 min on ice. Next, hPOs were permeabilized with 0.3% Triton X-100 in PBS for 40 min at RT, washed again with PBS and incubated three times for 10 min with 100 mM glycine in PBS at RT. After another wash with PBS, they were blocked for 12 h at RT with blocking solution [0.1% BSA (Sigma-Aldrich), 0.2% Triton X-100 (Sigma-Aldrich), 0.05% Tween-20 (Sigma-Aldrich) and 10% goat serum (Thermo Fisher Scientific) in PBS]. The samples were incubated with the primary antibodies diluted in blocking solution for 24 h at 37 °C in a shaker. They were then washed carefully with PBS and incubated with the secondary antibodies and DAPI (Thermo Fisher; 1 µg/ml in PBS) for 2 h at 27 °C. The samples were briefly washed and transferred to a 96-well plate before capturing images with a microscope (Zeiss LSM780). All antibodies used in this study were reported in Supplementary Table 1.

RNA extraction and analysis

Total RNA from cells was isolated using TRIzol reagent (Thermo Fisher Scientific), following manufacturer's instruction. RNA concentration and purity were verified using a NanoDrop ND-100 spectrophotometer (Thermo Fisher Scientific).

For the quantitative real time PCR (RT-qPCR) assay, cDNA was synthesized from 500 ng of total RNA with SuperScript IV VILO (Thermo Fisher Scientific). The cDNA was diluted 10-fold and 1 µl used as template for RT-qPCR analysis using SYBR Select Master Mix (Thermo Fisher Scientific) on a CFX96 thermal cycler (Bio-Rad, Hercules, CA) following the manufacturer's instruction. The relative expression levels of the selected targets were determined using the $\Delta\Delta C_t$ method and normalized, where not differentially specified, to geometric mean of ACTB and TBP mRNA levels.

All primers used in this study, listed in Supplementary Table 2, were designed using Primer 3 software.

DNA extraction and analysis

To perform DNA isolation hPO cells were collected and washed with PBS and treated overnight with a Lysis Buffer (100 mM NaCl, 50 mM Tris–HCl, 250 mM EDTA, 10% SDS and 1 mg/ml proteinase K). After that, genomic DNA was extracted using a 1:1 phenol-chloroform buffer. Then Genomic material was precipitated adding 1/10 volume of 7.5 ammonium acetate and 2 volumes of absolute ethanol (EtOH). Finally, DNA was washed in a 70% EtOH solution and resuspended in DNase/RNase-Free Distilled Water.

To perform telomere length analysis, a real-time quantitative polymerase chain reaction was performed on a CFX96 thermal cycler (Bio-Rad) preparing one 96-well plate for telomere and one for single copy reference gene (36B4). The thermal profile was set up for optimal activity of SYBR Select Master Mix: enzyme activation at 98 °C (30 s), 46 cycles of 95 °C (5 s), 54 °C (2 min), and a final stage at 62 °C (5 s) for telomere assay plate; enzyme activation at 98 °C (30 s), 46 cycles of 95 °C (5 s), 58 °C (1 min), and a final stage at 62 °C (5 s) for reference gene assay plate.

All primers used in this study, listed in Supplementary Table 2, were designed using Primer 3 software.

Protein extraction and western blot analysis

hPOs were washed with cold PBS and centrifuged for 5 min at 300x g. Harvested cell pellets were lysed by the addition of 10X (v/v) ice-cold Lysis Buffer (10 mM Tris-HCl at pH 7.0, 100 mM NaCl, 30 mM Na Pyrophosphate, 50 mM NaF, 5 mM ZnCl2, 1% Triton X-100, 1X PIC) for 30 min at 4 °C. Lysates were clarified by centrifugation at 12,000x g at 4 °C for 10 min and supernatant was collected on ice. Protein concentration was determined using Pierce BCA protein assay kit (ThermoFisher Scientific) according to manufacturer's instruction.

For western blot analysis, 20 µg of protein samples were boiled and loaded onto a pre-cast 4–20% Tris-Glycine mini gels and run in Tris-glycine buffer (all Thermo Fisher Scientific). After electrophoresis, proteins were

transferred to a nitrocellulose membrane. Then, the membranes were blocked in Blocking Buffer (PBS-T containing 5% Blotting-Grade blocker), for 1 h at room temperature with constant agitation and incubated with primary antibody O/N at 4 °C in Blocking Buffer. The membranes were washed three times with PBS-T, followed by incubation with secondary antibody horseradish peroxidase-conjugated for 2 h at RT. ECL reagents (Merck KGaA, Darmstadt, Germany) was used to initiate the chemiluminescence of peroxidase. The chemiluminescence signal was captured using ChemiDoc system (Bio-Rad).

All antibodies used in this study were reported in Supplementary Table 1.

Karyotype assay

hPO fragments from confluent culture were split at 1:8 ratios and allowed to grow for 4 days and then processed for metaphase analysis. Conventional karyotyping was carried out by Q-banding technique according to standard laboratory protocols. Numerical and structural abnormalities were analysed at 400 banding level, according to the international system for human cytogenomic nomenclature and the European general guidelines and quality assurance for cytogenetics. The analysis included a minimum of 20 metaphase-arrested cells to exclude clonal rearrangements. The capture of metaphase images was performed by the automated sample analysis system IKAROS (MetaSystems).

Array-based Comparative Genome Hybridization (a-CGH)

Array-based CGH analysis was performed using a 60-mer oligonucleotide probes technology SurePrint G3 Human CGH 8x60K. Labelling, purification and hybridization of DNA samples were carried out according to manufacturer's protocol guidelines (Agilent Oligonucleotide a-CGH for Genomic DNA Analysis, version 7.5). Data were generated using Agilent Feature Extraction and analysed by CytoGenomics 4.0.3.12 using ADM-2 algorithm (Agilent Technologies, Santa Clara, CA). To improve the accuracy of the results the Diploid Peak Centralization algorithm was applied.

The aberration filter was set to detect a minimum number of 3 consecutive probes/region and the minimum absolute average Log Ratio (MAALR) was ± 0,25. A second analysis was run with a MAALR of ±0,15 and with a minimum number of 3 probes/region to detect low level mosaicism.

Copy number variations were not reported if they coincided with published DNA variants listed in the Database of Genomic Variants (http://projects.tcag.ca/variation/). Genomic coordinates are set according to the 37 build (March 2009) of the Human Genome Reference consortium (GRch37/hg19).

Senescence-associated β-galactosidase (SA-βGal) assay

β-Galactosidase staining was performed using the Senescence Cells Histochemical Staining Kit (Sigma-Aldrich) according to the manufacturer's instructions upon reaching confluency. Briefly, the hPOs were incubated with 1x Fixation Buffer for 7 min at RT. hPOs were then washed and incubated with staining mixture (containing X-gal) at 37 °C for 4 hours. hPOs were observed with Eclipse TS 100 and images were taken with Digital Sight DS- L2 (all Nikon, Tokyo, Japan) at final magnification of 4x. Stained hPOs were stored at 4 °C in glycerol 70%. For quantification, pancreatic organoids were lysed in RIPA buffer for 30 min, sonicated at high intensity (pulse of 30 s with 30 s intervals) and centrifuged at 10,000x g for 5 min. The supernatant was collected and the absorbance was measured on a plate reader at A_{460}. The signal was normalized on total protein content.

ELISA

Aliquots of conditioned culture media collected at early and late passages of hPO cultures were quantified, according to the manufacturer's instructions, with Human GRO-α/MGSA (CXCL1) Standard ABTS ELISA Development kit, Human IL-6 Pre-Coated ELISA kit and Human IL-8 Pre-Coated ELISA kit (all Peprotech) to detect senescence-associated proteins.

The specific protein concentration in culture media was measured in accordance with the standard guideline protocol supplied with ELISA kits. The background value of each growth factor analysed was subtracted. Absorbance was measured at 450 nm or 405 nm with a microplate photometric reader Tecan Genius Plus (Tecan, Männedorf, Switzerland).

Scanning Electron microscopy

Scanning electron microscopy (SEM) was carried out on hPOs at early passage ($P \leq 5$) which were left to adhere overnight on coverslips that were pre-coated with poly-lysine and fixed with 2% glutaraldehyde in PBS overnight. The coverslips were briefly rinsed with PBS and water and then dehydrated in ethanol solutions 25 – 40 – 50 – 60 – 70 – 90% in H_2O and three times 100%, for 15 min. each. For HMDS drying, the samples were immersed for 3 min. in 100% HMDS and then the excess HMDS was blotted away by filter paper. The samples were then transferred to a desiccator for 25 min to avoid water contamination, mounted on stubs, sputter-coated with chromium in a Quorumtech Q150TES Turbo chromium sputter, and scanned by Zeiss Gemini SEM 500 (Zeiss, Oberkochen, Germany).

Cell viability

Cell viability was assessed using the MTT assay (Sigma-Aldrich). After 10 days, culture medium was removed and hPOs were washed twice with PBS, and then incubated with complete medium containing 0.5 mg/ml of MTT solution for 4 h. After incubation, the medium was removed and 200 µl of EtOH were added to each well to solubilize the formazan crystals. Absorbance was measured at 570 nm using a microplate reader (Tecan).

Computational analysis

Generation of single cell suspensions and Drop-seq. Organoids generated from 3 distinct donors were harvested at day 12 of passage 5 by washing the well three times with PBS. For the dissociation of hPOs into single cells, organoids were incubated in TrypLE Select for 15 min at 37 °C with shaking. Then, cells were further dispersed by gentle pipetting, filtered through a 30 µm cell strainer, counted by hemocytometer and resuspended in PBS + 0.01% BSA. Single cell RNA-seq was performed using Drop-seq technology[70]. Drop-seq was performed as described in Makosko et al., except for the fact that flow rates of 1100 µL/hour were used for the cell suspension, and 13000 µL/h for the oil. Libraries were prepared from 600 pg of of cDNA using the Nextera XT DNA sample prep kit according manufacturer's instruction using a custom Read 1 primer (GCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC, IDT) and were sequenced on a NovaSeq 6000 machine (Illumina). Read 1 was 20 bp (bases 1–12 cell barcode, bases 13–20 UMI), read 2 (paired end) was 60 bp, and index primer was 8 bp.

Drop-seq read preprocessing and generation of digital expression matrix. Raw sequencing data were analysed using UMI tool (https://github.com/CGATOxford/UMI-tools/), a bioinformatic pipeline that includes: (i) read preprocessing (ii) alignment of reads to the human genome GRCh38 (hg38) using STAR v. 2.7.5a[71], (iii) grouping of mapped reads by each single cell barcode (iv) UMI-based determination of unique transcript numbers per each cell and (v) generation of a digital gene expression matrix.

Datasets quality control and integration. Gene expression count matrices were imported into the Seurat package (v4.3.0)[72] using R environment (v4.2.2), where further quality control was performed. First, we retained only genes detected in more than 3 cells, then dead or stressed cells, identified by a percentage of mitochondrial genes higher than 15%, cells with less than 200 or more than 5000 expressed genes were excluded. Afterwards, the 3 datasets were integrated performing an unbiased cross-sample data-set integration as outlined and recommended by Seurat developers. Briefly, the following Seurat functions were used: (1) SCTransform() for normalization, scaling and identification of top 3000 variable features, and to regress out for percentage of mitochondrial gene

expression; (2) SelectIntegrationFeatures() to choose the features to use when integrating multiple datasets. This function ranks features by the number of datasets they appear in, breaking ties by the median rank across datasets. It returns the highest features by this ranking; (3) PrepSCTIntegration() to prepare an object list that has been run through SCTransform() for integration; (4) FindIntegrationAnchors() to find the integration anchors; (5) IntegrateData() to perform dataset integration using a pre-computed anchorset.

Dimensionality reduction, clustering, and cell type annotation. The 3000 most variable features were used for the principal component (PC) analysis. The number of PCs use to construct an approximate nearest-neighbour graph using FindNeighbors() function have been choose based on ElbowPlot. Cell clustering was performed with FindClusters() function using the original Louvain algorithm with the resolution of 0.4 decided by Clustree to avoid over or under clusterization of the data[73]. For visualization, the dimensionality of the data was reduced using Uniform Manifold Approximation and Projection (UMAP) {https://doi.org/10.48550/arXiv.1802.03426}. Differential gene expression analysis was performed with FindAllMarkers() function using the non-parametric Wilcoxon Rank Sum statistics. Differential expression was defined if absolute logFC was >0.25 and adjusted *P*-value (Bonferroni) < 0.05. FindMarkers() function was used to compare two different cell clusters/groups. The same parameters above mentioned were applied also to this function. Cluster annotation was accomplished integrating the information available in three largest database of human cell population markers, Human Biomolecular Atlas Program (HuBMAP)[74], CellMarker[75] and PanglaoDB[76], by considering all the gene expression changes among the cell clusters (Supplementary Data).

Cell-cycle analysis. We assigned a cell cycle score on each cell according to its gene expression of G2/M and S phase markers[77] Based on this scoring system, we classified each cell in either G2M, S or G1 phase using the CellCycleScoring() function in Seurat. The cells at different cell cycle classifications were visualized in the UMAP map, and the expression of cell cycle genes were plotted out using FeaturePlot() function in Seurat.

Gene onthology and pathway analysis. Gene ontology analysis was performed using EnrichR[78,79]. We selected commonly reoccurring and highly ranked GO terms, highly enriched in genes having an adjusted *p*-value < 0.05.

Stemness calculation and pseudotemporal ordering. The lineage tree was calculated using the StemID2 algorithm included in the RaceID3 package[44]. The number of significant links and the transcriptome entropy (reflecting the uniformity of the transcriptome) were calculated for each cluster. The StemID score for each cluster was determined by multiplying the link number and Δentropy. Marker gene expression was determined for the RaceID clusters. The expression of these markers was employed to determine which Seurat cell cluster corresponded to each RaceID cluster using AddModuleScore() function.

Preprocessed integrated Seurat object were integrated into Monocle3 (v1.3.1)[45,80] to infer developmental trajectories of Seurat clusters in an unbiased manner. To order the cells according to pseudotime, the clusters with the highest StemID scores were chosen as the root nodes.

Integration dataset with primary pancreas atlas. To assess similarity between pancreatic organoid cell types and pancreas, we integrated our datasets with three human datasets from adult human pancreas (GSE8413324, GSM2230757 and GSM2230759). We performed standard data preprocessing: (1) thresholding, (2) normalization using SCTransform(), and (3) variable gene calculation using SCTransform(). Then, using SelectIntegrationFeatures() and the PrepSCTIntegration() functions, we calculate integral features for downstream analysis. Using the FindIntegrationAnchors() and IntegrateData() functions, we create an optimally integrated dataset. Then for calculating the neighbourhood graph, 10 PCAs (based on an ElbowPlot) were used to construct an approximate nearest-neighbour graph using FindNeighbors() function. Cell clustering was performed with FindClusters() function using the original Louvain algorithm with the resolution of 0.4.

Cell-cell communication analysis. Intercellular interactions were inferred in human pancreatic organoids with CellChat package (v1.6.1)[48]. following the official workflow and using standard parameters. The cell-cell communication analysis was based on expression of ligand-receptor pairs from the CellChat human database using all three subset types: secreted signalling, cell–cell contact and ECM–Receptor. Cluster 6 was excluded from the analysis due to low cell numbers. Cell-cell interaction analysis were performed with computeCommunProbPathway(), aggregateNet() and netAnalysis_computeCentrality(). The summed incoming and outgoing interactions strengths were visualized with netAnalysis_signalingRole_scatter().

Statistics and reproducibility. All the quantitative data are shown as means plus SD, as specified in each figure legend. No statistical method was used to predetermine sample size and all experiments were repeated at least three times with specific sample sizes reported in each figure legend. All statistical tests were performed using Prism 7 and statistical significance (p-value) assessed by unpaired two-tailed Student's *t*-test. *P* values < 0.05 were considered significant, while *P* values > 0.05 were not reported.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The source data underlying the graphs and differential gene expression analysis can be found in Supplementary Data. The original blot images are included in Supplementary Fig. 11. Sequencing data have been deposited in the Gene Expression Omnibus (GEO), accession no. GSE158035. Previously published scRNA-seq data reanalysed here are available at under GEO accession no. GSE81076/GSE131886. All script used for scRNA-seq analysis will be available upon request to corresponding author A.C.

Received: 12 April 2024; Accepted: 31 October 2024;
Published online: 18 November 2024

References

1. Bouwens, L. & Pipeleers, D. G. Extra-insular beta cells associated with ductules are frequent in adult human pancreas. *Diabetologia* **41**, 629–633 (1998).
2. Lee, M. G., Ohana, E., Park, H. W., Yang, D. & Muallem, S. Molecular mechanism of pancreatic and salivary gland fluid and HCO3 secretion. *Physiol. Rev.* **92**, 39–74 (2012).
3. Chong, S. K., Lindridge, J., Moniz, C. & Mowat, A. P. Exocrine pancreatic insufficiency in syndromic paucity of interlobular bile ducts. *J. Pediatr. Gastroenterol. Nutr.* **9**, 445–449 (1989).
4. Wilschanski, M. & Novak, I. The cystic fibrosis of exocrine pancreas. *Cold Spring Harb. Perspect. Med.* **3**, a009746 (2013).
5. Boxhoorn, L. et al. Acute pancreatitis. *Lancet Lond. Engl.* **396**, 726–734 (2020).
6. Ferreira, R. M. M. et al. Duct- and acinar-derived pancreatic ductal adenocarcinomas show distinct tumor progression and marker expression. *Cell Rep.* **21**, 966–978 (2017).
7. Lee, A. Y. L. et al. Cell of origin affects tumour development and phenotype in pancreatic ductal adenocarcinoma. *Gut* **68**, 487–498 (2019).
8. Kleeff, J. et al. Pancreatic cancer. *Nat. Rev. Dis. Prim.* **2**, 16022 (2016).

9. Clevers, H. Modeling development and disease with organoids. *Cell* **165**, 1586–1597 (2016).
10. Fatehullah, A., Tan, S. H. & Barker, N. Organoids as an in vitro model of human development and disease. *Nat. Cell Biol.* **18**, 246–254 (2016).
11. Huch, M. et al. Long-term culture of genome-stable bipotent stem cells from adult human liver. *Cell* **160**, 299–312 (2015).
12. Sato, T. Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. **459**, 5 (2009).
13. Gao, D. et al. Organoid cultures derived from patients with advanced prostate cancer. *Cell* **159**, 176–187 (2014).
14. Sachs, N. et al. Long-term expanding human airway organoids for disease modeling. *EMBO J.* **38**, e100300 (2019).
15. Wiedenmann, S. et al. Single-cell-resolved differentiation of human induced pluripotent stem cells into pancreatic duct-like organoids on a microwell chip. *Nat. Biomed. Eng.* **5**, 897–913 (2021).
16. Araki, R. et al. Genetic aberrations in iPSCs are introduced by a transient G1/S cell cycle checkpoint deficiency. *Nat. Commun.* **11**, 197 (2020).
17. Lund, R. J., Närvä, E. & Lahesmaa, R. Genetic and epigenetic stability of human pluripotent stem cells. *Nat. Rev. Genet.* **13**, 732–744 (2012).
18. Mummery, C. Induced pluripotent stem cells-a cautionary note. *N. Engl. J. Med.* **364**, 2160–2162 (2011).
19. Dossena, M. et al. Standardized GMP-compliant scalable production of human pancreas organoids. *Stem Cell Res. Ther.* **11**, 94 (2020).
20. Hendley, A. M. et al. Single-cell transcriptome analysis defines heterogeneity of the murine pancreatic ductal tree. *eLife* **10**, e67776 (2021).
21. Qadir, M. M. F. et al. Single-cell resolution analysis of the human pancreatic ductal progenitor cell niche. *Proc. Natl Acad. Sci. Usa.* **117**, 10876–10887 (2020).
22. Doke, M. et al. Dynamic scRNA-seq of live human pancreatic slices reveals functional endocrine cell neogenesis through an intermediate ducto-acinar stage. *Cell Metab.* **35**, 1944–1960.e7 (2023).
23. Grün, D. et al. De novo prediction of stem cell identity using single-cell transcriptome data. *Cell Stem Cell* **19**, 266–277 (2016).
24. Huch, M. et al. Unlimited in vitro expansion of adult bi-potent pancreas progenitors through the Lgr5/R-spondin axis. *EMBO J.* **32**, 2708–2721 (2013).
25. Venugopal, S., Anwer, S. & Szászi, K. Claudin-2: Roles beyond Permeability Functions. *Int. J. Mol. Sci.* **20**, 5655 (2019).
26. Kojima, T. et al. Tight junctions in human pancreatic duct epithelial cells. *Tissue Barriers* **1**, e24894 (2013).
27. Rojek, A., Praetorius, J., Frøkiaer, J., Nielsen, S. & Fenton, R. A. A current view of the mammalian aquaglyceroporins. *Annu. Rev. Physiol.* **70**, 301–327 (2008).
28. Kilic, G., Wang, J. & Sosa-Pineda, B. Osteopontin is a novel marker of pancreatic ductal tissues and of undifferentiated pancreatic precursors in mice. *Dev. Dyn. Publ. Am. Assoc. Anat.* **235**, 1659–1667 (2006).
29. Hua, H. et al. BMP4 regulates pancreatic progenitor cell expansion through Id2. *J. Biol. Chem.* **281**, 13574–13580 (2006).
30. Lasorella, A., Benezra, R. & Iavarone, A. The ID proteins: master regulators of cancer stem cells and tumour aggressiveness. *Nat. Rev. Cancer* **14**, 77–91 (2014).
31. Lachat, P. et al. Expression of NDRG1, a differentiation-related gene, in human tissues. *Histochem. Cell Biol.* **118**, 399–408 (2002).
32. Ellen, T. P., Ke, Q., Zhang, P. & Costa, M. NDRG1, a growth and cancer related gene: regulation of gene expression and function in normal and disease states. *Carcinogenesis* **29**, 2–8 (2008).
33. Keith, B. & Simon, M. C. Hypoxia-inducible factors, stem cells, and cancer. *Cell* **129**, 465–472 (2007).
34. Arsenijevic, T., Perret, J., Van Laethem, J.-L. & Delporte, C. Aquaporins involvement in pancreas physiology and in pancreatic diseases. *Int. J. Mol. Sci.* **20**, 5052 (2019).
35. Beer, R. L., Parsons, M. J. & Rovira, M. Centroacinar cells: at the center of pancreas regeneration. *Dev. Biol.* **413**, 8–15 (2016).
36. Saint-Criq, V. & Gray, M. A. Role of CFTR in epithelial physiology. *Cell. Mol. Life Sci. CMLS* **74**, 93–115 (2017).
37. Cleveland, M. H., Sawyer, J. M., Afelik, S., Jensen, J. & Leach, S. D. Exocrine ontogenies: on the development of pancreatic acinar, ductal and centroacinar cells. *Semin. Cell Dev. Biol.* **23**, 711–719 (2012).
38. Reichert, M. & Rustgi, A. K. Pancreatic ductal cells in development, regeneration, and neoplasia. *J. Clin. Invest.* **121**, 4572–4578 (2011).
39. Bouwens, L. Cytokeratins and cell differentiation in the pancreas. *J. Pathol.* **184**, 234–239 (1998).
40. dilorio, P., Rittenhouse, A. R., Bortell, R. & Jurczyk, A. Role of cilia in normal pancreas function and in diseased states. *Birth Defects Res. Part C. Embryo Today Rev.* **102**, 126–138 (2014).
41. Grapin-Botton, A. Ductal cells of the pancreas. *Int. J. Biochem. Cell Biol.* **37**, 504–510 (2005).
42. Muraro, M. J. et al. A single-cell transcriptome atlas of the human pancreas. *Cell Syst.* **3**, 385–394.e3 (2016).
43. Segerstolpe, Å. et al. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. *Cell Metab.* **24**, 593–607 (2016).
44. Herman, J. S., Sagar & Grün, D. FateID infers cell fate bias in multipotent progenitors from single-cell RNA-seq data. *Nat. Methods* **15**, 379–386 (2018).
45. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
46. Ioannou, M. et al. ALDH1B1 is a potential stem/progenitor marker for multiple pancreas progenitor pools. *Dev. Biol.* **374**, 153–163 (2013).
47. Yamaguchi, J. et al. Pancreatic duct glands (PDGs) are a progenitor compartment responsible for pancreatic ductal epithelial repair. *Stem Cell Res* **15**, 190–202 (2015).
48. Jin, S. et al. Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088 (2021).
49. Wedeken, L. et al. Adult murine pancreatic progenitors require epidermal growth factor and nicotinamide for self-renewal and differentiation in a serum- and conditioned medium-free culture. *Stem Cells Dev.* **26**, 599–607 (2017).
50. Bonfanti, P. et al. Ex vivo expansion and differentiation of human and mouse fetal pancreatic progenitors are modulated by epidermal growth factor. *Stem Cells Dev.* **24**, 1766–1778 (2015).
51. Kim, J., Koo, B.-K. & Knoblich, J. A. Human organoids: model systems for human biology and medicine. *Nat. Rev. Mol. Cell Biol.* **21**, 571–584 (2020).
52. Boj, S. F. et al. Organoid models of human and mouse ductal pancreatic cancer. *Cell* **160**, 324–338 (2015).
53. Loomans, C. J. M. et al. Expansion of adult human pancreatic tissue yields organoids harboring progenitor cells with endocrine differentiation potential. *Stem Cell Rep.* **10**, 712–724 (2018).
54. Georgakopoulos, N. et al. Long-term expansion, genomic stability and in vivo safety of adult human pancreas organoids. *BMC Dev. Biol.* **20**, 4 (2020).
55. Díaz-Flores, L. et al. Adult stem and transit-amplifying cell location. *Histol. Histopathol.* **21**, 995–1027 (2006).
56. Aloia, L., Gutierrez, A., Caballero, J. M. & Di Croce, L. Direct interaction between Id1 and Zrf1 controls neural differentiation of embryonic stem cells. *EMBO Rep.* **16**, 63–70 (2015).
57. Stankic, M. et al. TGF-β-Id1 signaling opposes Twist1 and promotes metastatic colonization via a mesenchymal-to-epithelial transition. *Cell Rep.* **5**, 1228–1242 (2013).
58. Sneddon, J. B. et al. Stem cell therapies for treating diabetes: progress and remaining challenges. *Cell Stem Cell* **22**, 810–823 (2018).
59. Brown, L. F. et al. Expression and distribution of osteopontin in human tissues: widespread association with luminal epithelial surfaces. *Mol. Biol. Cell* **3**, 1169–1180 (1992).

60. Re-Shaping the Pancreatic Cancer Tumor Microenvironment: A New Role for the Metastasis Suppressor NDRG1 - PubMed. https://pubmed.ncbi.nlm.nih.gov/37345116/.
61. Geleta, B. et al. Targeting Wnt/tenascin C-mediated cross talk between pancreatic cancer cells and stellate cells via activation of the metastasis suppressor NDRG1. *J. Biol. Chem.* **298**, 101608 (2022).
62. Huang, X., Trinh, T., Aljoufi, A. & Broxmeyer, H. E. Hypoxia signaling pathway in stem cell regulation: good and evil. *Curr. Stem Cell Rep.* **4**, 149–157 (2018).
63. Mohyeldin, A., Garzón-Muvdi, T. & Quiñones-Hinojosa, A. Oxygen in stem cell biology: a critical component of the stem cell niche. *Cell Stem Cell* **7**, 150–161 (2010).
64. Hoffmann, W. Trefoil factors TFF (trefoil factor family) peptide-triggered signals promoting mucosal restitution. *Cell. Mol. Life Sci. CMLS* **62**, 2932–2938 (2005).
65. Taupin, D. & Podolsky, D. K. Trefoil factors: initiators of mucosal healing. *Nat. Rev. Mol. Cell Biol.* **4**, 721–732 (2003).
66. Ramond, C. et al. Understanding human fetal pancreas development using subpopulation sorting, RNA sequencing and single-cell profiling. *Dev. Camb. Engl.* **145**, dev165480 (2018).
67. Balak, J. R. A., Juksar, J., Carlotti, F., Lo Nigro, A. & de Koning, E. J. P. Organoids from the human fetal and adult pancreas. *Curr. Diab. Rep.* **19**, 160 (2019).
68. Casamitjana, J., Espinet, E. & Rovira, M. Pancreatic organoids for regenerative medicine and cancer research. *Front. Cell Dev. Biol.* **10**, 886153 (2022).
69. Melzi, R. et al. Donor and isolation variables associated with human islet monocyte chemoattractant protein-1 release. *Transplantation* **78**, 1564–1567 (2004).
70. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
71. STAR: ultrafast universal RNA-seq aligner - PubMed. https://pubmed.ncbi.nlm.nih.gov/23104886/.
72. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
73. Zappia, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *GigaScience* **7**, giy083 (2018).
74. HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187–192 (2019).
75. Zhang, X. et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res*. **47**, D721–D728 (2019).
76. Franzén, O., Gan, L.-M. & Björkegren, J. L. M. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database J. Biol. Databases Curation* **2019**, baz046 (2019).
77. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
78. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. **44**, W90–W97 (2016).
79. Xie, Z. et al. Gene set knowledge discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).
80. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).

Acknowledgements

This work was funded by the EU Horizon2020 grant "LSFM4LIFE–Production and characterization of endocrine cells derived from human pancreas organoids for the cell-based therapy of Type 1 diabetes", project number 668350. This research was funded by Ricerca Corrente from the Italian Ministry of Health. LL thanks Carola Morell, Christopher Gribben and Meritxell Huch for their critical revision and fruitful advice.

Author contributions

AC and LL conceived, designed the experiments, and supervised the study. AC performed the bioinformatics analysis. AC, FR, RP, MD, CM, MB, KW and LH performed the experiments. AC and FR analysed the data. AC, FR and LL wrote the original draft of the manuscript. FP and VD supervised the microscopy analysis. VS and LP provided the sample collection and single cell RNA sequencing methodology. LL acquired the funding and administered the project. All authors read the manuscript and approved the final version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-07193-3.

Correspondence and requests for materials should be addressed to Alessandro Cherubini or Lorenza Lazzari.

Peer review information *Communications Biology* thanks Meritxell Rovira and Konstantinos Dimas for their contribution to the peer review of this work. Primary Handling Editors: Dr Alex Nord and Dr Ophelia Bu. A peer review file is available.