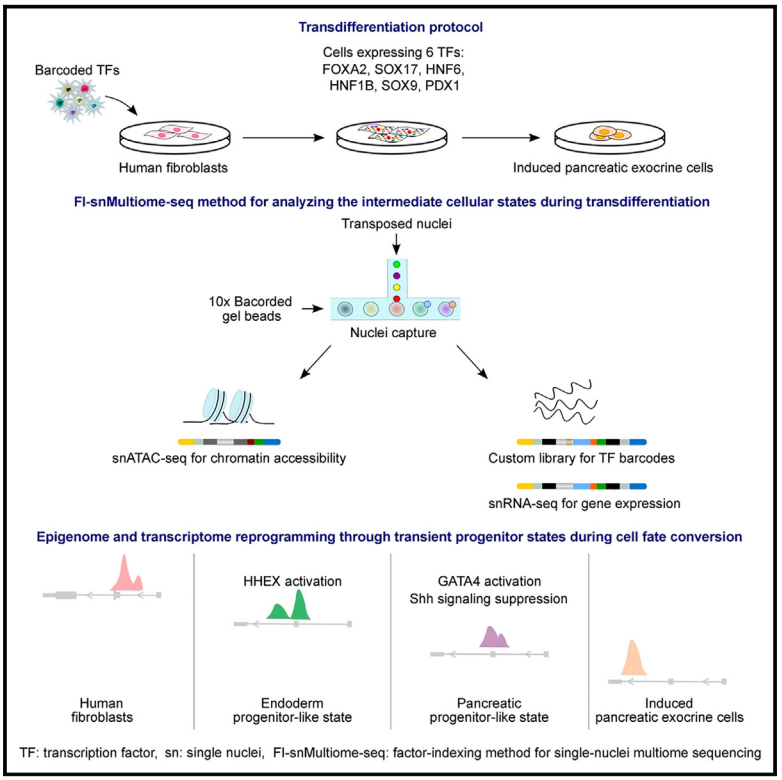
Article

Developmental Cell

Single-cell epigenome analysis identifies molecular events controlling direct conversion of human fibroblasts to pancreatic ductal-like cells

Graphical abstract



Highlights

- Induction of pancreatic cell fate from somatic cells by lineage-specific TFs
- Epigenome and transcriptome remodeling facilitates cell fate conversion
- Direct transdifferentiation occurs through transient progenitor cell states
- FI-snMultiome-seq dissects the role of individual TFs at single-cell resolution

Authors

Liangru Fei, Kaiyang Zhang, Nikita Poddar, Sampsa Hautaniemi, Biswajyoti Sahu

Correspondence

biswajyoti.sahu@helsinki.fi

In brief

Fei et al. report a combination of six transcription factors that are necessary and sufficient for converting human fibroblasts to pancreatic exocrine cells. Reprogramming factors mediate the cell fate switch in a highly ordered and temporal fashion in collaboration with endogenous transcription factors.

Fei et al., 2023, Developmental Cell *58*, 1701–1715
September 25, 2023
https://doi.org/10.1016/j.devcel.2023.08.023

CellPress

Developmental Cell

CellPress
OPEN ACCESS

Article

Single-cell epigenome analysis identifies molecular events controlling direct conversion of human fibroblasts to pancreatic ductal-like cells

Liangru Fei,[1] Kaiyang Zhang,[2,6] Nikita Poddar,[1,6] Sampsa Hautaniemi,[2] and Biswajyoti Sahu[1,3,4,5,7,*]
[1]Applied Tumor Genomics Program, Research Programs Unit, Faculty of Medicine, University of Helsinki, Haartmaninkatu 8, Helsinki 00014, Finland
[2]Research Program in Systems Oncology, Research Programs Unit, Faculty of Medicine, University of Helsinki, Haartmaninkatu 8, Helsinki 00014, Finland
[3]iCAN Digital Precision Cancer Medicine Flagship, University of Helsinki, Haartmaninkatu 8, Helsinki 00014, Finland
[4]Medicum, Faculty of Medicine, University of Helsinki, Haartmaninkatu 8, Helsinki 00014, Finland
[5]Centre for Molecular Medicine Norway, Faculty of Medicine, University of Oslo, Gaustadelléen 21, 0349 Oslo, Norway
[6]These authors contributed equally
[7]Lead contact
*Correspondence: biswajyoti.sahu@helsinki.fi
https://doi.org/10.1016/j.devcel.2023.08.023

SUMMARY

Cell fate can be reprogrammed by ectopic expression of lineage-specific transcription factors (TFs). However, the exact cell state transitions during transdifferentiation are still poorly understood. Here, we have generated pancreatic exocrine cells of ductal epithelial identity from human fibroblasts using a set of six TFs. We mapped the molecular determinants of lineage dynamics using a factor-indexing method based on single-nuclei multiome sequencing (FI-snMultiome-seq) that enables dissecting the role of each individual TF and pool of TFs in cell fate conversion. We show that transition from mesenchymal fibroblast identity to epithelial pancreatic exocrine fate involves two deterministic steps: an endodermal progenitor state defined by activation of HHEX with FOXA2 and SOX17 and a temporal GATA4 activation essential for the maintenance of pancreatic cell fate program. Collectively, our data suggest that transdifferentiation—although being considered a direct cell fate conversion method—occurs through transient progenitor states orchestrated by stepwise activation of distinct TFs.

INTRODUCTION

Direct cell fate conversion (transdifferentiation) is a process in which somatic cells are reprogrammed to defined cell types of another lineage without a pluripotent intermediate state using either ectopic expression of cell- and lineage-specific transcription factors (TFs), non-coding RNAs, or small molecules in a defined lineage-dependent media.[1] For example, fibroblasts have been directly converted to cells representing all three germ layers, including induced neuronal cells (iNCs) from the ectoderm,[2] cardiomyocytes from the mesoderm,[3] and hepatocytes from the endoderm.[4] Transdifferentiation is controlled by coordinated action of pioneer TFs such as FOXA or GATA factors and other lineage-specific TFs, resulting in global reprogramming of epigenetic landscape and gene expression.[5] Previous studies have demonstrated considerable cellular heterogeneity during lineage conversion,[6–8] but the exact path of cellular states through which the direct lineage conversion occurs is not fully understood.

Transdifferentiation approaches have been critical not only for understanding basic developmental mechanisms governing cell identity but also for innovative experimental strategies in disease modeling and potential therapeutic applications. These include, for example, conversion of astrocytes to dopaminergic neurons in Parkinson's disease[9] and glial cells into neurons after brain injury,[10] as well as combining direct transdifferentiation from fibroblasts to induced hepatocytes (iHeps) with controlled expression of cancer-specific oncogenes to model liver cancer development.[11] However, the lack of transdifferentiation factors for many human cell and tissue types hampers the use of this approach for understanding various human diseases. Pancreas is a complex organ that harbors multiple specialized cell types with endocrine (α, β, δ, γ, and ε cells) and exocrine (acinar and ductal cells) functions. Enormous efforts have gone into identifying the TFs required for the generation of endocrine β cells,[12–15] but the factors required for other defined cell types are currently unknown. However, the exocrine cells such as the ductal epithelial cells are important not only for the normal physiological function of the pancreas but also in various disease contexts.[16] For example, chronic pancreatitis and cystic fibrosis impair exocrine function, and highly lethal pancreatic cancer that

is often diagnosed at the terminal stages typically originates from the exocrine cells. Recent functional studies have implicated the lineage-determining TFs in tumorigenic processes.[11,17,18] Thus, it is pertinent to study the role of defined TFs in controlling pancreatic exocrine cell identity for a better understanding of pancreatic cancer as well as for potential regenerative medicine applications.

Here, we report the defined factors necessary and sufficient for a direct conversion of human fibroblasts to induced pancreatic exocrine cells (iPECs) of ductal epithelial identity. We have dissected the mechanistic role of individual TFs in the transdifferentiation process, identifying the critical factors that control the process in a temporally coordinated manner.

RESULTS

Generation of pancreatic exocrine cells by direct lineage conversion

To establish a direct cell fate conversion protocol for transdifferentiating human fibroblasts to iPECs in defined media (Figure 1A), we set out to identify the pool of TFs required for inducing pancreatic exocrine cell fate. We selected 14 candidate TFs (Figure S1A), eight of which were literature-curated based on their reported role in the maturation of exocrine cells during human pancreas development[21,22] and another eight predicted using the Mogrify computational framework.[23] Two TFs, FOXA2 and GATA4, were suggested by both approaches. The candidate TFs were cloned into a lentiviral expression vector and systematically studied in a series of reprogramming experiments by transducing 20 different TF combinations to human foreskin fibroblasts (HFFs; Figure S1B). Transduced cells were monitored for morphological changes (Figure S2) and for the expression of pancreas-related marker genes using quantitative real-time-PCR or RNA sequencing (RNA-seq) at different time points (Figure S1B). Briefly, after observing clear morphological changes in a pilot experiment (condition 1; Figures S1B and S1C), we compared the gene expression profiles of transduced cells with control fibroblasts and detected upregulation of pancreatic exocrine cell markers as well as enrichment of pancreas-related gene set among the differentially expresses genes (DEGs) (Figures S1D–S1F). We then designed two sub-pools for generating acinar- and ductal-like cells separately (conditions 2 and 3 with ten and nine TFs, respectively; Figure S1B). Briefly, factors implicated in the specification of endoderm and pancreatic progenitors (such as FOXA2 and GATA4)[21,22] were included in both conditions, whereas factors with previously reported roles in cell type specification, such as PTF1A for acinar cells[24] and HNF6 for ductal cells,[25] were restricted to these specific pools (see STAR Methods). Reprogramming experiments using these pools demonstrated the feasibility of producing iPECs since quantitative real-time-PCR and RNA-seq analyses showed a gradual increase in the expression of marker genes for pancreatic exocrine cells and downregulation of fibroblast-related genes (Figures S1G–S1J). The pools were further refined based on the previous literature about pancreas development,[26–30] resulting in eight and six TFs for acinar and ductal cells, respectively (conditions 5 and 7; Figure S1B; see STAR Methods). The cells transduced with acinar cell TFs (condition 5) did not develop epithelial-like morphology. However, the cells transduced with a pool of six TFs for ductal cells (condition 7), FOXA2, SOX17, PDX1, HNF1B, HNF6 (ONECUT1), and SOX9 (henceforth referred to as 6F), showed clear morphological changes during transdifferentiation with several clusters of epithelial-like cells appearing around 3 weeks of reprogramming (Figure 1B). Dissociating the cells and re-plating them on growth factor-reduced Matrigel-coated dishes resulted in further maturation toward epithelial cell-like morphology: cell clusters that were picked and re-plated at 4 weeks after induction could be expanded up to 8 weeks as the cells mature (Figure 1B). Other than the 6F-pool (condition 7), few clusters of epithelial-like cells were also observed in conditions 10 (6F without SOX17) and 12 (6F without SOX17 and SOX9; Figure S1B), and these cells also showed upregulation of ductal cell markers (Figures S3A–S3C). However, the 6F-pool enabled robust induction of reprogrammed cells with both sustainable epithelial-like morphology and elevated expression of pancreatic ductal cell markers and was thus used in the subsequent experiments.

For initial characterization of the identity of cells generated using the 6F combination, RNA-seq was performed from reprogrammed cells at 6 and 10 weeks after TF transduction using hTERT-immortalized cells for the later time point, 10 weeks^{TERT}, as the epithelial-like cell clusters could only be expanded until 8 weeks after induction. Global gene expression analysis identified a distinct transcriptional program within the reprogrammed cells compared with HFFs (Figures 1C and S3D). Pancreatic marker genes such as secreted phosphoprotein 1 (*SPP1*), prominin 1 (*PROM1*), and carbonic anhydrase 2 (*CA2*) were upregulated, and the genes highly expressed in fibroblasts were efficiently downregulated (Figures 1D and S3E), consistent with earlier reports of hepatic, cardiac, and neuronal conversion from fibroblasts.[3,31,32] Importantly, gene set enrichment analysis (GSEA) for DEGs between 6F-reprogrammed cells and control fibroblasts showed statistically significant positive enrichment specifically for pancreatic ductal cell gene signature among signatures for all major cell types of endodermal origin (Figures 1E and S3F) and among all pancreatic cell types (Figure S3G). Comparison of gene expression programs in iPECs to those in human pancreatic acinar[19] and ductal cells[20] further supported the observation that the iPECs resemble ductal cells rather than acinar cells (Figure 1F). These results indicate that the 6F combination induces cell fate conversion toward iPECs with pancreatic ductal cell identity.

Cell fate conversion results in iPECs with functional properties of pancreatic ductal cells

Functional properties of iPECs induced with 6F were tested in activity assays for three pancreatic enzymes: CA, which is the key enzyme expressed in pancreatic ducts, as well as amylase and trypsin, which are specific for acinar cells. Strong CA activity was detected in iPECs (Figure 1G), whereas amylase and trypsin activities were negligible. Of note, CA activity can originate from several CA family proteins, such as CA2 and CA12 enzymes in these cells, but only CA2 has been considered a ductal cell marker.[33] HFFs express *CA12* (Figure 1H) resulting in their basal CA activity (Figure 1G), whereas upregulated *CA2* expression along with increased CA activity detected in 6-week iPECs speaks for their ductal cell attributes (Figures 1G and 1H). The weaker CA activity in 10 weeks^{TERT} iPECs suggests that

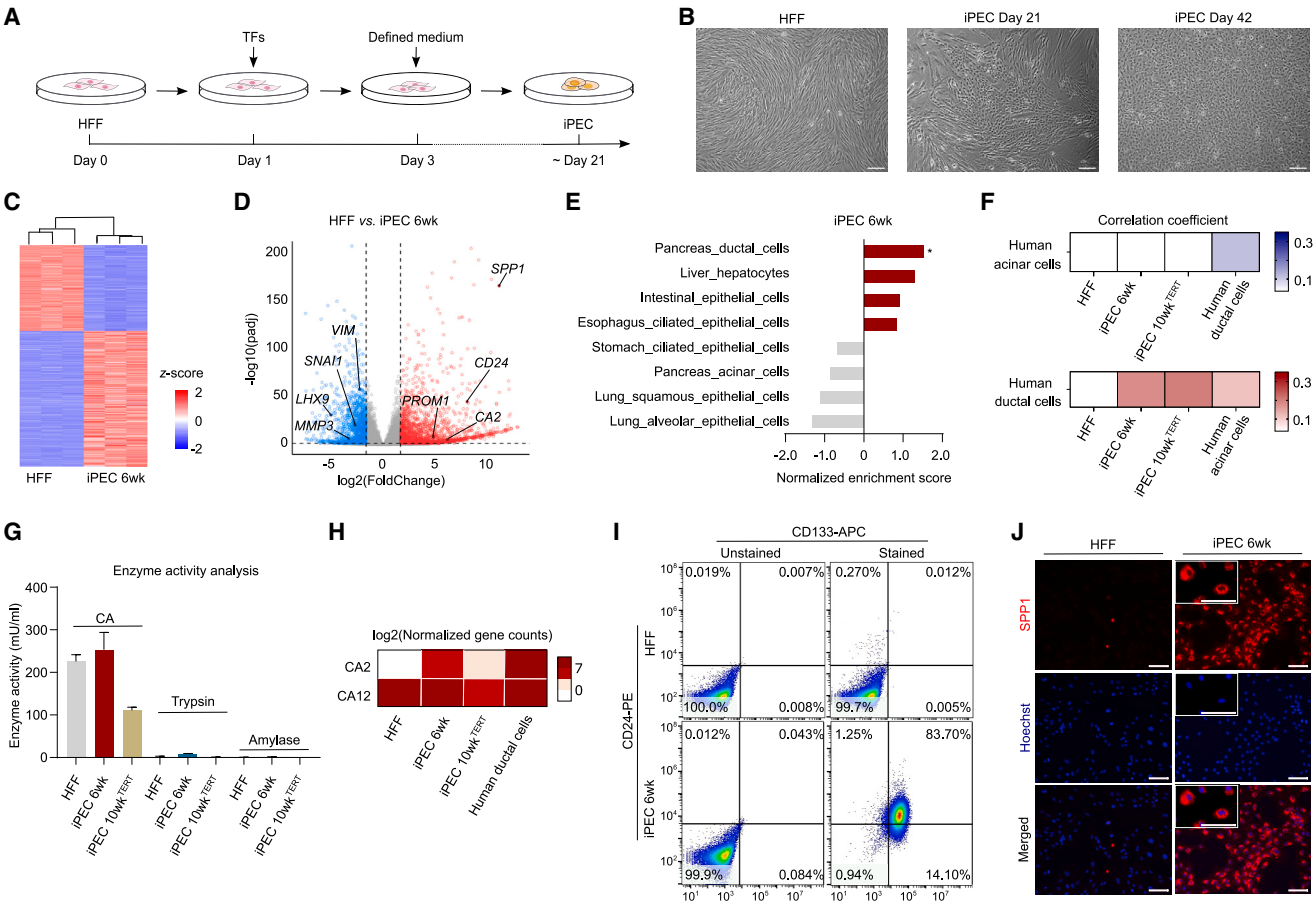


Figure 1. Direct conversion of human fibroblasts to pancreatic ductal-like cells

(A) Experimental design for generating iPECs. Human foreskin fibroblasts (HFFs) were transduced with different combinations of lentiviral TF constructs. The culture medium was changed to defined medium 2 days after transduction (see STAR Methods), and the cells were monitored for morphological changes.

(B) Brightfield images showing spindle-shaped HFFs and epithelial-like morphology of iPECs at days 21 and 42 after induction. Scale bars in all microscope images represent 100 μm.

(C) Hierarchical clustering of DEGs in HFFs vs. iPECs at 6 weeks ($|\log_2 FC| \geq 1.5$, p value < 0.05); *Z* score of normalized expression values is indicated using a color scale. All RNA-seq experiments in Figures 1, 2, and 3 were performed in three technical replicates.

(D) Volcano plot for DEGs in HFFs vs. iPECs at 6 weeks. Significant upregulated and downregulated DEGs in iPECs ($|\log_2 FC| \geq 1.5$, p value < 0.05; n = 3) marked in red and blue, respectively.

(E) Normalized enrichment score from GSEA for gene signatures of all major cell types of endodermal origin among DEGs in iPEC at 6 weeks vs. HFF (*p value < 0.01 and false discovery rate [FDR] < 5%).

(F) Pearson correlation coefficients for the top 10% most-variable genes from $\log_2$ normalized gene expression data of HFFs, iPECs and human acinar,[19] and ductal cells.[20]

(G) Measurement of enzyme activities (mean ± SEM; n = 6, including three biological replicates and two technical replicates for each biological replicate) for CA, trypsin, and amylase in the lysates of HFFs and iPECs at 6 and 10 weeks (10-week sample immortalized with hTERT).

(H) Heatmap for expression levels of *CA2* and *CA12* from RNA-seq data in HFFs, iPECs at 6 and 10 weeks, and human ductal cells.[20]

(I) Flow cytometry analysis for CD133 and CD24 to determine the proportion of ductal-like cells in the iPEC population at 6 weeks after induction. Of note, CD133 protein is encoded by the *PROM1* gene (see Figure 1D).

(J) Immunofluorescence staining for SPP1 and cellular DNA (Hoechst) to assess the epithelial phenotype of iPECs at 6 weeks.

See also Figures S1–S3.

immortalization may have affected their functionality. Flow cytometry analysis using cell surface markers for pancreatic ductal and exocrine cells, CD133 and CD24,[34] respectively, identified that over 80% of iPECs expressed both markers, demonstrating high purity of ductal cells in the isolated iPEC colonies (Figure 1I). Moreover, clear immunofluorescence signal was detected for SPP1, an essential regulator of human pancreatic duct cell maturation and epithelial phenotype[35] (Figure 1J). Taken together, these results speak for the functional ductal phenotype and efficient mesenchymal to epithelial conversion in the iPECs reprogrammed from human fibroblasts using the 6F pool.

iPEC reprogramming is controlled by temporal activation of specific TFs

To gain detailed insights into the transcriptional changes during direct lineage conversion between fibroblasts and iPECs, cells

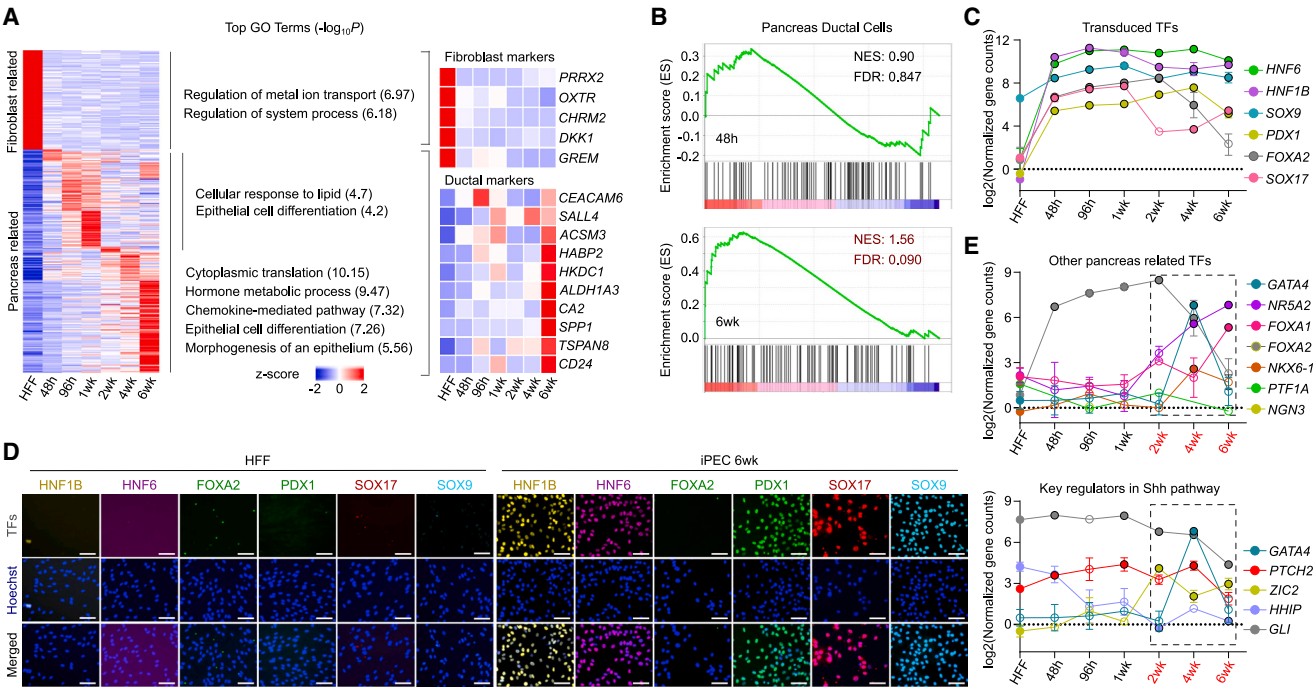


Figure 2. Gene expression dynamics during reprogramming

(A) Heatmap clustering for average normalized expression values for DEGs ($|\log_2 FC| \geq 1.5$, p value < 0.05; n = 3) between reprogrammed cells at indicated time points vs. HFF; DEGs detected in all time points are shown. GO enrichment analysis for biological processes was performed for the indicated clusters.

(B) GSEA for fetal pancreatic ductal cell signature in reprogrammed cells at 48 h and 6-week time points. NES, normalized enrichment score.

(C) Line plots showing the expression dynamics for the 6F transcripts at indicated time points.

(D) Immunofluorescence staining for the TFs from 6F pool in HFFs and iPEC at 6 weeks.

(E) Line plots showing gene expression dynamics of pancreas-related TFs.

In (C) and (D), the transcript and protein measurements do not differentiate between transduced and endogenously expressed genes.

In (C) and (E), normalized $\log_2$ transformed gene counts from RNA-seq data were plotted (mean ± SEM; n = 3). *GLI* represents the average expression levels of *GLI1*, *GLI2*, and *GLI3*. For statistical analysis, the expression of each TF at each time point was compared with HFFs using two-tailed Student's t test. Hollow circles, not significant; solid-colored dots with black circles, p value < 0.01 (n = 3).

See also Figure S3.

were collected for bulk RNA-seq analysis at six different time points from 48 h to 6 weeks after 6F transduction. Transcriptome-wide changes were analyzed using principal-component analysis (PCA) (Figure S3H), and pairwise DEG analysis was performed for reprogrammed cells at each time point against HFF. Unsupervised hierarchical clustering for the significant DEGs overlapping all pairwise comparisons identified two major clusters (Figure 2A), the fibroblast-related genes that were downregulated and the pancreas-related genes that showed dynamic expression changes during transdifferentiation process. Gene ontology (GO) terms related to epithelial cell differentiation were significantly enriched among the genes that were induced within the first week of reprogramming (Figure 2A), consistent with recent studies showing that mesenchymal-epithelial transition is an essential early step in somatic cell reprogramming.[36] Rapid downregulation of fibroblast-related genes already at 48 h time point and strong induction of ductal cell marker genes such as *SALL4*, *ACSM3*, *HABP2*, *CA2*, and *SPP1* during later stages of reprogramming (Figure 2A) indicate a successful cell fate switch between fibroblasts and iPECs and gradual maturation of the ductal cells. Moreover, pancreatic ductal cell signature was markedly enriched at 6-week time point but not yet at 48 h after 6F transduction (Figure 2B), and the CA activity along with *CA2* expression was higher at 6-week iPECs compared with those measured at 48 h (Figures S3I and S3J).

For a comprehensive understanding of the reprogramming process, we analyzed the expression patterns for the six reprogramming TFs as well as for endogenous pancreas-related TFs that were not part of the 6F pool. The expression levels were largely consistent for four reprogramming TFs, *HNF6*, *HNF1B*, *SOX9*, and *PDX1*, whereas *SOX17* and *FOXA2* were downregulated around 1–2 weeks of reprogramming (Figure 2C) despite being expressed from a constitutively active promoter. From week 4 onward, the expression of SOX17 increased again but FOXA2 remained low, as seen from mRNA expression (Figure 2C) and protein levels (Figure 2D). However, the expression of endogenous *FOXA1* increased, potentially compensating for the FOXA2 loss (Figure 2E, top). These results suggest that SOX17 and FOXA proteins play distinct roles during the early and late stages of cell fate conversion. Of the endogenous pancreas-related TFs, GATA4 expression showed a striking temporal control with strong induction of its mRNA at 4 weeks and downregulation around 6 weeks of transdifferentiation (Figure 2E, top). Previous studies have shown that GATA4 loss

results in conversion of pancreatic cells to alternate intestinal and gastric cell fates due to aberrant activation of sonic hedgehog (Shh) pathway in pancreatic progenitors[37] and that elevated levels of Shh signaling block pancreas formation.[38] Thus, we analyzed RNA-seq data for the expression of all canonical Shh signaling pathway genes from the Molecular Signatures Database (MSigDB; see STAR Methods) during reprogramming. ZIC2, a downstream target of GATA4,[39] has been earlier reported to enhance Shh signaling through nuclear retention of Gli1,[40] and PTCH2 is a negative regulator of Shh signaling.[41] Activation of *GATA4* at 4 weeks was concordant with downregulation of *ZIC2* and upregulation of *PTCH2* (Figure 2E, bottom), suggesting their role in suppressing the Shh pathway and inducing pancreatic progenitor cell identity before ductal cell maturation. Downregulation of GATA4 (Figure 2E, bottom) after pancreatic endoderm progenitor specification agrees with previously reported GATA4 expression in the pancreatic progenitor state that is subsequently restricted to acinar cells during pancreatic maturation.[26] These results give a mechanistic insight into how pancreatic identity is established and maintained through specific temporal suppression of Shh signaling.

Activation of other pancreas-related TFs such as *NR5A2* and *NKX6.1* was also observed at 4 weeks of reprogramming (Figure 2E, top). *NR5A2* expression remained strong also at 6 weeks, whereas *NKX6.1* expression decreased, commensurate with the previously reported role of NR5A2 in pancreatic exocrine cells[30] and NKX6.1 in β cells of the mature pancreas.[22] Notably, TFs related to acinar cell specification such as *PTF1A* and endocrine cell fate such as *NEUROG3* were not detected during the reprogramming (Figure 2E, top), indicating that the 6F pool mediates reprogramming of iPECs specifically toward pancreatic ductal cell identity.

Global chromatin remodeling during iPEC reprogramming

As seen from the RNA-seq analysis in Figure 2A, the epithelial cell differentiation program already starts within the first week of iPEC reprogramming. Thus, to delineate the early changes in chromatin state that facilitate these programs, we performed an assay for transposase-accessible chromatin sequencing (ATAC-seq) at the same time points that were used for RNA-seq within the first week of reprogramming. Consistent with the temporal changes in gene expression patterns, chromatin accessibility was altered along the reprogramming time course, and six broad clusters of differentially accessible regions were identified (Figure 3A). Chromatin immunoprecipitation sequencing (ChIP-seq) for histone 3 lysine 27 acetylation (H3K27ac) at the same time points showed that epigenome reprogramming corresponds to changes in the levels of activating chromatin marks at the same loci (Figure S3K). Motif enrichment analysis of ATAC-seq peaks in different clusters identified a shift in motif accessibility already around 96 h of reprogramming from TFs that control the somatic cell identity of fibroblasts (FOS/TEAD/RUNX)[42] to pancreas-related TFs corresponding to 6F pool such as HNF6, HNF1, SOX, and FOXA (Figure 3A). The same pattern was observed from differential motif analysis of accessible chromatin regions between 1-week reprogrammed cells and HFF (Figure 3B).

In addition to the motifs for the 6F reprogramming TFs, a notable increase in the predicted binding activity was observed for HHEX between 1-week reprogrammed cells and HFFs (Figure 3B). This was commensurate with a specific upregulation of *HHEX* expression at 96 h and 1-week time points (Figure 3C). Interestingly, HHEX is one of the earliest markers of the foregut progenitor cells that is required for the formation of pancreas[43,44] and the specification of hepatopancreatic ductal system.[45] TF footprinting analysis by TOBIAS framework[46] showed dynamic epigenome reprogramming of fibroblast chromatin during iPEC conversion with decreased chromatin accessibility for TFs that maintain the somatic cell identity of the fibroblasts (JUNB, FOSL1)[42] and increase for early endoderm specification TFs like SOX17, FOXA2, and HHEX, as well as pancreas-lineage determining TFs such as HNF6, HNF1B, and SOX9 (Figure 3D, left). The chromatin regions that showed increased accessibility for pancreatic development TFs were found to be near several pancreas-related TFs (such as *HNF1B*, *HNF6*, *GATA6*, and *NR5A2*), suggesting a regulatory loop in which the reprogramming TFs activate the expression of pancreas-related TFs, including the endogenous genes corresponding to the reprogramming factors such as *HNF1B* and *HNF6* (Figure 3D, right; Table S1). The changes in the chromatin state measured by ATAC-seq and H3K27ac ChIP-seq are commensurate with gene expression changes during reprogramming as seen from activation of pancreatic ductal marker genes such as *CD24*, *SPP1*, and *SALL4* and downregulation of fibroblast-related genes such as *MMP3* and *PRRX2* (Figure 3E, top). The motifs of somatic (TEAD, Bach)[42] and mesenchymal (Twist)[47] TFs were enriched at the regulatory regions of *MMP3* and *PRRX2* (Figure 3E, bottom; Table S3). For ductal cell markers, motifs for both reprogramming TFs (such as Hnf1 and ONECUT) as well as other TFs (such as Pparg and Nfe2l2) were enriched (Figure 3E, bottom), suggesting both direct and indirect mechanisms in the marker gene induction during cell fate conversion.

Taken together, these data suggest that transdifferentiation involves transient intermediate progenitor states during which the cells move toward a dedifferentiated state. Specific temporal activation and inactivation of reprogramming TFs control these distinct steps: (1) initiation of cell fate conversion by FOXA2 and SOX17, resulting in chromatin reprogramming and loss of the gatekeeper TFs (e.g., FOS/JUN/TEAD/RUNX) that maintain the fully differentiated somatic cell identities,[42] (2) cell fate specification and determination by HNF6 and HNF1B, and (3) cell fate maintenance and maturation by HNF6 and HNF1B together with SOX9 and pancreas-specific PDX1. This process also involves transient activation of distinct endogenous TF programs, namely HHEX at the endoderm progenitor state and GATA4 at the pancreatic progenitor state.

Factor-indexing approach for FI-snMultiome-seq to dissect the role of defined TFs in gene expression and chromatin accessibility

To analyze the role of each individual TF and different combinations of TFs in cell fate control at single-cell resolution and segregate the expression of exogenous and endogenous TFs, we developed a factor-indexing method based on single-nuclei multiome sequencing (FI-snMultiome-seq) method. It combines a TF-indexing (barcoding) strategy with the single-nuclei multiome sequencing platform from 10× Genomics for concomitant epigenome and gene expression profiling from the same

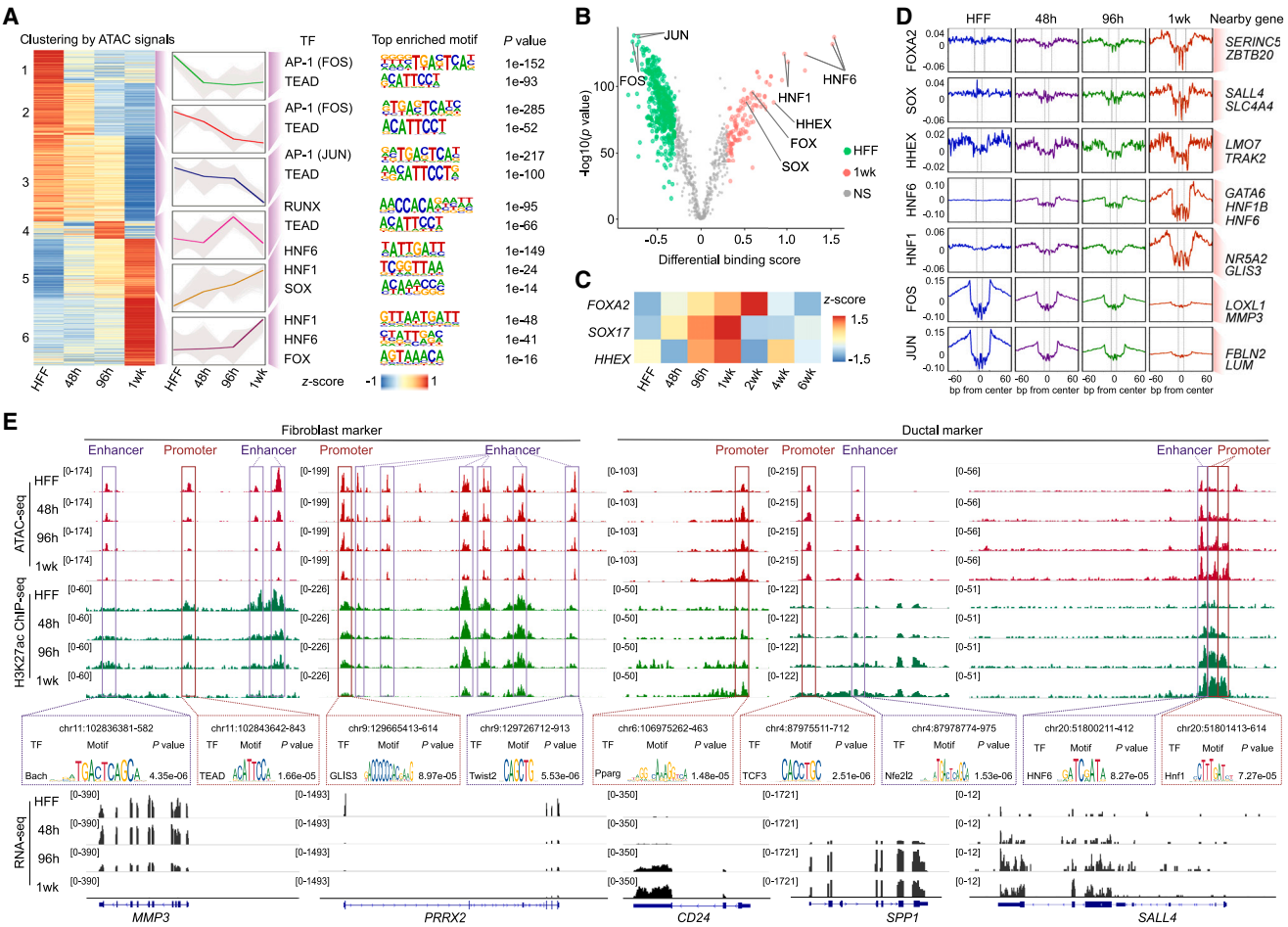


Figure 3. Chromatin accessibility dynamics during reprogramming

(A) Unsupervised clustering heatmap of top 50% master ATAC-seq peaks across HFFs and iPECs at 48 h, 96 h, and 1 week. Line plots show the trend of each cluster. Top representative motifs ranked by p value are shown. ATAC-seq was performed in two biological replicates.

(B) Volcano plot showing differential predicted binding activity of all investigated TF motifs between HFF (left) and 1-week reprogrammed cells (right). TFs with | differential binding score| > 0.3 are colored on both sides. From each group, the top 1,000 highest-ranking peak regions according to reads per kilobase per million mapped reads (RPKM) values were used.

(C) Heatmap showing the *Z* score of normalized expression values for *FOXA2*, *SOX17*, and *HHEX* from RNA-seq data during reprogramming.

(D) Aggregate accessibility profile of TF binding sites for selected differentially bound TFs shown in (B) from bias-corrected ATAC-seq footprints. Signal depletion indicates TF binding. Representative pancreas-related TFs are shown from overlap analysis between nearby genes within a 10-kb window from the TF motif center in 1-week sample and the gene sets GOBP_ENDODERM_DEVELOPMENT, GOBP_PANCREAS_DEVELOPMENT, and MURARO_PANCREAS_DUCTAL_CELL from GSEA database. Representative somatic TFs are shown from the overlapping genes between "nearby genes" in HFFs and the fibroblast marker genes from Table S2.

(E) Genome browser snapshots showing ATAC-seq, H3K27ac ChIP-seq, and RNA-seq signals at indicated time points for marker genes of fibroblasts (*MMP3, PRRX2*) and pancreatic exocrine cells (*CD24*, *SPP1*, *SALL4*). Representative TF motifs within ATAC-seq peaks at promoters (1 kb from transcription start site [TSS]) and enhancers (>1 kb from TSS) highlighted for each gene are shown. H3K27ac ChIP-seq was performed in two biological replicates.

See also Figure S3 and Tables S1, S2, and S3.

nuclei/cell using single-nuclear ATAC-seq (snATAC-seq) and single-nuclear RNA-seq (snRNA-seq), respectively. For this, we designed a lentiviral expression construct harboring unique sequence barcodes at the 3′ untranslated region (UTR) of the open reading frame (ORF) (Figure 4A). Each TF from the 6F pool was cloned into this vector, resulting in TF constructs with transcribed barcodes that can be detected from the snRNA-seq data (Figure 4A) after custom library preparation protocol. Importantly, the FI-snMultiome-seq approach is an improvement over the existing single-nuclei sequencing methods because the TF barcodes enable linking TF expression directly to changes in gene expression and chromatin accessibility in each individual cell expressing one or multiple TFs and pooling different experimental conditions into one Chromium multiome run, which is not possible with the currently available methods.

Barcoded TFs were transduced to HFFs either individually or as a 6F pool, and the cells were maintained in the defined reprogramming media. The cells were harvested for FI-snMultiome-seq at three time points, 48 h, 1 week, and 2 weeks after transduction to analyze the intermediate states through which individual cells

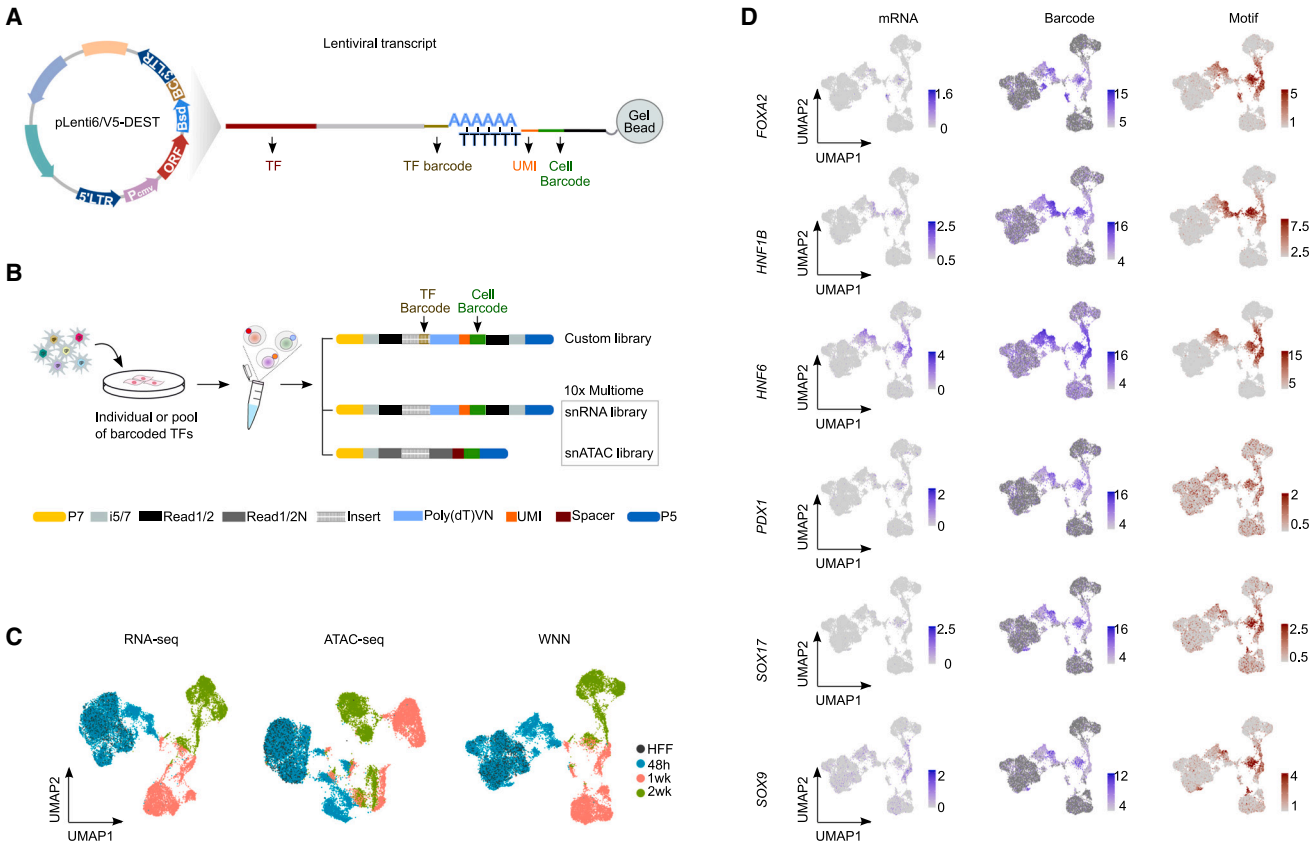


Figure 4. FI-snMultiome method for dissecting the role of defined TFs in reprogramming at single-cell resolution

(A) Schematic presentation of the lentiviral expression construct for barcoded TFs and the strategy for capturing the barcodes during 10× Multiome snRNA-seq workflow. UMI, unique molecular identifier.

(B) Reprogrammed cells transduced with six barcoded TFs individually or as a pool were harvested at different time points and multiplexed for the analysis of transcriptomic (snRNA-seq) and epigenetic (snATAC-seq) changes from the same cell. Custom TF barcode library was generated by an additional PCR after the pre-amplification step during the 10× Multiome workflow (see STAR Methods), enabling correlation of the TF barcodes with the 10× cell barcodes during downstream analysis.

(C) Uniform manifold approximation and projection (UMAP) plots of all cells from different time points (indicated with colors) based on gene expression (left) and chromatin accessibility (middle) separately and their integrated profiles using weighted nearest neighbor (WNN) analysis (right). HFFs transduced with green fluorescent protein (GFP) reporter gene were used as a control. For all FI-snMultiome-seq analyses, 5,399 individual cells from 1 week, 6,921 from 2 weeks, and 10,276 from 48 h and HFFs were used.

(D) UMAPs showing the transcript levels of endogenous TFs (left), exogenous TFs detected from their barcodes (middle), and motif accessibility analyzed from the snATAC-seq data (right).

progress during the early stages of reprogramming (Figure 4B). Sequencing library for the TF barcodes was prepared by an additional custom PCR step from the same pre-amplified material that is used for snRNA-seq and snATAC library preparation (see STAR Methods). This results in three distinct libraries—snRNA-seq, snATAC-seq, and custom barcode libraries—that all are marked by the cell barcodes introduced by the 10× multiome workflow (Figure 4B). Importantly, this strategy enables mapping each TF barcode to the cell barcodes and analyzing, which TFs are expressed in a cell and what are the corresponding chromatin and transcriptional states.

All cells harvested at different time points were analyzed together and projected into low-dimensional subspaces based on snRNA-seq for gene expression (Figure 4C, left), snATAC-seq for chromatin accessibility (Figure 4C, middle), or their combination (Figure 4C, right). The cells collected at 48 h still resemble HFF, but 1 and 2 weeks after TF transduction, they have clearly distinct epigenome and transcriptome profiles (Figure 4C). The FI-snMultiome-seq method with barcoded TF constructs has three major advantages. First, it has high sensitivity in detecting ectopically expressed TFs due to the barcoded vector (Figure 4D). The motifs of 6F were enriched within accessible chromatin in the cells where their barcodes were detected, suggesting activity of the corresponding TFs in the transduced cells (Figure 4D). Since the current 10× multiome workflow is based on poly(A) capture and 3′ RNA sequencing, it cannot capture transcripts originating from typical lentiviral expression vectors with >1 kb distance between poly(A) (located in the 3′ long terminal repeat [LTR]) and the ORF. In the FI-snMultiome-seq vector, the TF barcodes reside close to 3′ LTR (Figure 4A), enabling their efficient capture and optimal library size for Illumina sequencing. Second, FI-snMultiome-seq method allows dissecting the effect

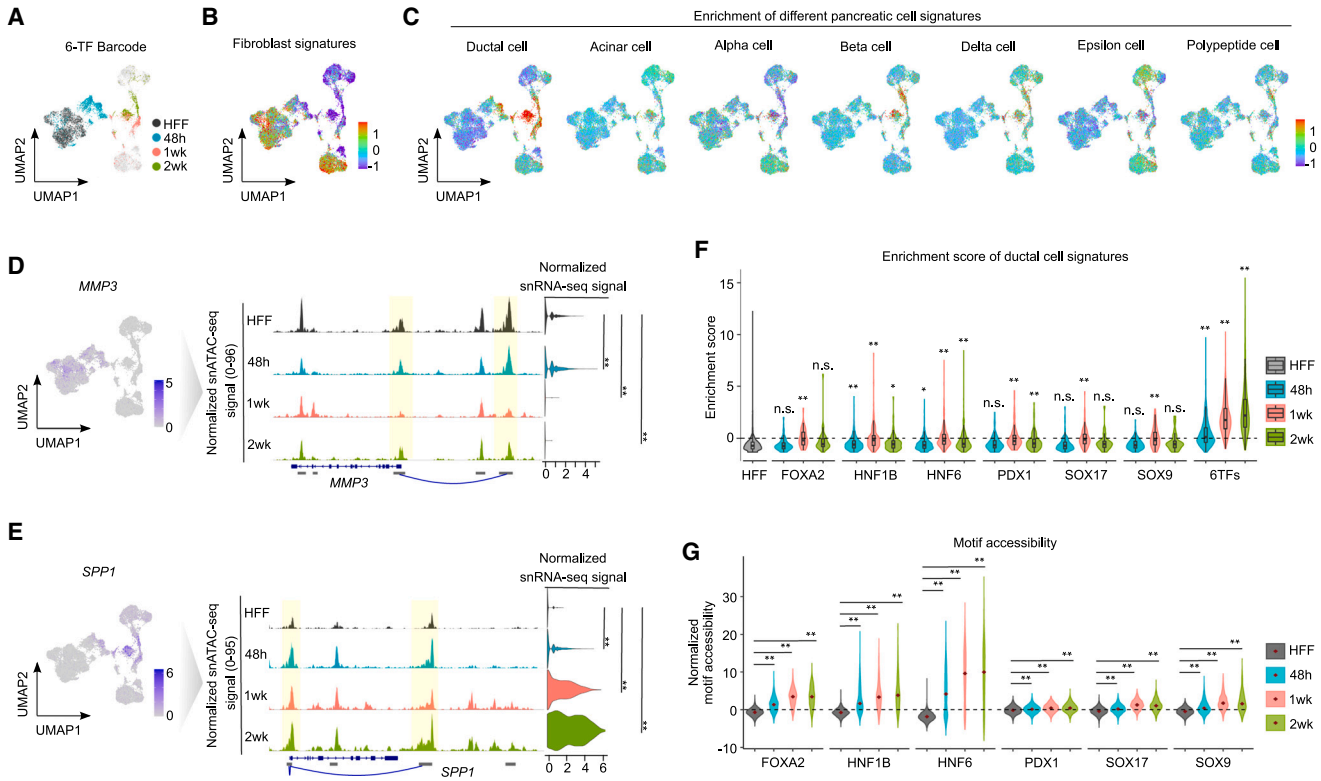


Figure 5. FI-snMultiome-seq identifies the specific enrichment of pancreatic ductal cell signatures in iPECs

(A) UMAP highlighting the control cells and the reprogrammed cells expressing the barcodes for 6F.

(B) UMAP showing the expression of fibroblast signatures among all the cells.

(C) UMAPs showing the expression of signatures for different pancreatic cell lineages.

(D and E) Coverage plots showing chromatin accessibility at genomic loci of ductal cell marker *SPP1* and fibroblast marker *MMP3* and their mRNA expression levels.

(F) Violin plot showing enrichment score of pancreatic ductal cell signatures in HFFs and in cells expressing individual TFs and the 6F pool. Data are represented as the median, top, and bottom quartiles.

(G) Violin plot showing the motif accessibility scores (*Z* scores) of individual TFs in the cells with 6F pool at three time points and in HFFs.

(D–G) Two-sided Wilcoxon rank-sum test, FDR from Benjamini-Hochberg (BH) adjusted p values; n.s., not significant; *FDR < 0.01. **FDR < 0.001.

See also Figures S4–S6.

of each transduced barcoded TF from the respective endogenous TF (Figure 4D) since the mRNA reads for each TF originate exclusively from the endogenous gene. Third, FI-snMultiome-seq enables analyzing only the cells that have been successfully transduced with TFs of interest and discarding all non-transduced cells from the analysis. Notably, the cell population that expresses all 6F barcodes (Figure 5A) shows downregulation of the fibroblast signature and enrichment of the ductal cell signature (Figures 5B and 5C), whereas the signal for other pancreatic cell types[48] such as acinar cells or endocrine cell types was weak (Figure 5C).

To further confirm the ductal cell identity of the reprogrammed cells, we analyzed publicly available human single-cell RNA-seq datasets for major pancreatic cell types[49] and compared their gene expression programs with the reprogrammed cells expressing 6F pool. This demonstrated that the reprogrammed cells resemble human ductal cells more than any other pancreatic cell type at all analyzed time points (Figure S4A). However, the time points analyzed here using the FI-snMultiome-seq method are relatively early in the reprogramming timeline (cf. Figure 2), and the cells transdifferentiating toward ductal cell identity may still retain some fibroblast attributes. We computed a probabilistic assignment for each cell to fibroblasts, major pancreatic cell types, and transient states between them. Consistent with the enrichment of ductal cell signature toward later time points (Figure 5C), also the proportion of cells in fibroblast-ductal transient state increased during reprogramming (Figures S4B and S4C). In conclusion, these results demonstrate that FI-snMultiome-seq method is a powerful tool for analyzing the molecular features during cell fate conversion at single-cell resolution.

FI-snMultiome-seq method identifies regulatory dynamics during cell fate conversion

The key feature of the FI-snMultiome-seq method is that it enables concomitant analysis of chromatin accessibility and gene expression from the same individual cells and correlates them to TF expression. Different time points collected for FI-snMultiome-seq analysis (48 h, 1 week, and 2 weeks after 6F transduction) allow temporal analysis of these regulatory

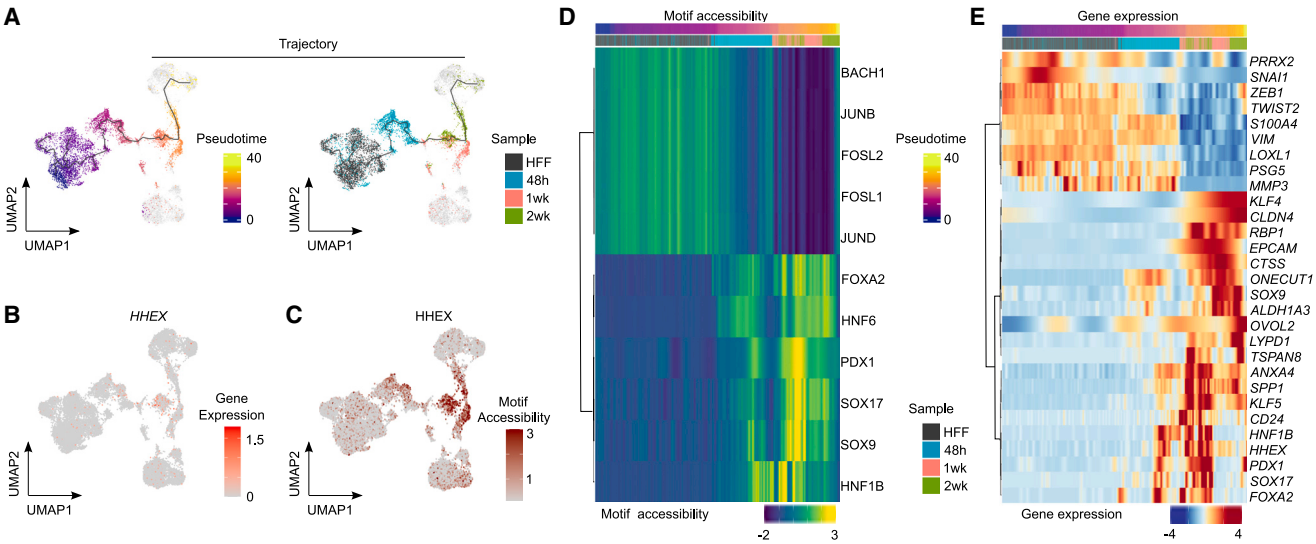


Figure 6. Reconstructing the reprogramming path from fibroblasts to pancreatic ductal-like cells using FI-snMultiome-seq data

(A) UMAP with cells colored by pseudotime (left) and experimental time points (right). The gray line corresponds to the principal graph learned by Monocle 3 (see STAR Methods).

(B) *HHEX* expression.

(C) Motif accessibility for HHEX.

(D) Heatmap showing smoothed pseudotime-dependent motif accessibility of representative motifs of somatic and 6F pancreatic TFs.

(E) Heatmap showing smoothened expression levels of representative marker genes for fibroblasts and pancreatic cells.

See also Figure S7.

changes during the early stages of reprogramming. Marker genes for fibroblasts and ductal cells were clearly expressed in different cell populations (Figures 5D, 5E, S4D, and S4E). Genomic loci of fibroblast-related genes (e.g., *MMP3*, *LOXL1*, and *FBLN2*) undergo dynamic reprogramming of chromatin accessibility already 48 h time point commensurate with down-regulation of their transcript expression (Figures 5D and S4F). By contrast, the changes in chromatin accessibility and transcript expression for ductal marker genes (e.g., *SPP1*, *HNF6*, and *HNF1B*) occur at 1-week time point (Figures 5E and S4G). These observations agree with the bulk RNA-seq/ATAC-seq analyses that showed dynamic epigenome reprogramming from 48 h onward, validating the population-level observations at single-cell resolution.

Next, we used the FI-snMultiome-seq data to address whether all six TFs from the 6F pool are necessary for pancreatic ductal cell identity since the factor-specific barcodes identify the combination of TFs expressed in each individual cell. Previous reports have suggested that ectopic expression of a single TF can induce cell fate conversion,[50,51] such as conversion of fibroblasts to neurons by *Ascl1*.[52] None of the individual TFs in the 6F pool induced the expression of pancreatic ductal cell signature (Figure 5F), indicating that a concerted action of multiple TFs is necessary for pancreatic exocrine cell fate conversion. Similarly, we analyzed the effect of all TF combinations from the 6F pool (five, four, three, and two TFs) on iPEC conversion, showing that the 6F pool was the most efficient in induction of ductal cell signatures (Figures S5A–S5D). Among all TF combinations, ductal cell signatures were more significantly enriched in the conditions involving HNF6 (Figures S6A–S6F), suggesting a critical role for HNF6 in cell fate conversion. Increased motif accessibility was observed for all six TFs from 48 h to 2-week time points compared with control fibroblasts (Figure 5G), indicating their functionality in the reprogramming process. Taken together, 6F pool is the minimum combination of TFs for efficient specification of ductal-like lineage from human fibroblasts.

Reconstructing the direct reprogramming path from fibroblasts to pancreatic ductal-like cells

To generate a high-resolution view of combined transcriptomic and epigenomic landscape in the heterogeneous cell population during transdifferentiation, we reconstructed the reprogramming path based on pseudo-temporal ordering of the cells from the FI-snMultiome-seq data. The *in silico* ordering produced a continuum of intermediate cellular states (Figure 6A, left) that correlated well with the true experimental timeline of reprogramming (Figure 6A, right). This enabled detailed temporal analysis of cell fate conversion at single-cell resolution, identifying specific expression of endoderm progenitor marker *HHEX* in the early stages of transdifferentiation (Figure 6B) commensurate with a robust increase in the HHEX motif accessibility (Figure 6C), confirming the observation from bulk RNA-seq/ATAC-seq analyses (cf. Figures 3B–3D). Importantly, upregulation of other definitive endoderm markers such as *KIT*, *KLF5*, and early endoderm regulon controlling TFs such as *YAP*[53–55] coincides with the *HHEX* induction (Figure S7A), supporting the notion that the direct cell fate conversion occurs through an intermediate progenitor state.

Motif enrichment analysis showed that motif accessibility increased markedly for pancreas-related TFs and decreased for fibroblast-related TFs along the pseudo-time course (Figures 6D and S7B). This enabled defining the role of individual

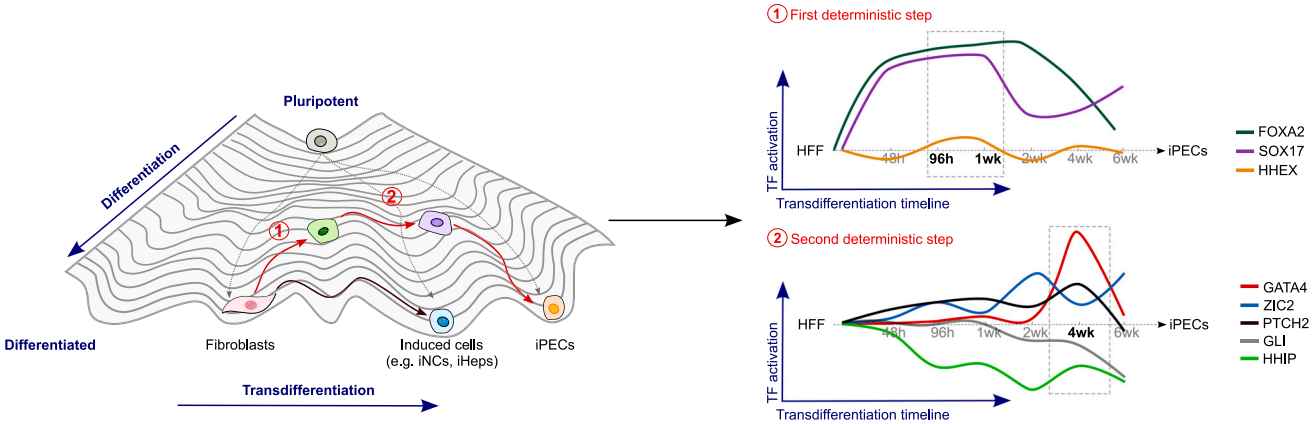


Figure 7. Schematic presentation of the transdifferentiation process from fibroblasts to iPECs in the context of Waddington's epigenetic landscape

Our results suggest that despite being considered direct cell fate conversion, transdifferentiating cells dedifferentiate toward distinct molecularly controlled intermediate progenitor states: (1) definitive endoderm-like progenitor state controlled by FOXA2 and SOX17 and marked by HHEX expression and (2) pancreatic progenitor state controlled by specific activation of GATA4. iNCs, induced neuron cells; iHeps, induced hepatocytes.

TFs from the 6F pool in the regulatory events controlling cell fate conversion. The first event in the reprogramming path is high FOXA2 motif accessibility (Figures 6D and S7B), indicating its role in chromatin reprogramming and enhancer priming required for the initiation of cell fate conversion. Second, increased motif accessibility for HNF6 (ONECUT1) follows FOXA2, staying high throughout reprogramming (Figures 6D and S7B), suggesting that HNF6 may act as the master TF for pancreatic exocrine cell fate specification (cf. Figure S6). Third, the motif accessibility is increased for HNF1B, followed by the remaining TFs from the 6F pool—PDX1, SOX17, and SOX9—suggesting their role in pancreatic cell maturation and differentiation (Figures 6D and S7B). Of note, the strength of the pseudotime analysis in detecting dynamic regulatory changes was highlighted by the substantial increase in the motif accessibility observed for PDX1 that was not clearly visible from the analysis based on the sample time points (cf. Figure 5G). Motif accessibility for the TFs that maintain the somatic cell identity, such as FOSL1 and JUNB,[42] decreased gradually along the pseudo-temporal axis (Figures 6D and S7B) along with downregulation of fibroblast-specific genes such as *VIM* and *MMP3* as well as TFs important for fibroblast identities such as *ZEB1*, *TWIST2*, and *SNAI1* (Figures 6E and S7C). The expression of genes reported in early pancreatic progenitors, such as *FOXA2*, *HHEX*, and *SOX17*, as well as pancreatic ductal cell and epithelial marker genes *SPP1*, *CD24*, and *EPCAM* increased along the pseudotime (Figures 6E and S7C), providing a time-resolved view of direct cell fate conversion from HFFs to pancreatic ductal epithelial cells.

Taken together, we have identified the minimum and sufficient combination of six TFs required for HFF conversion to iPECs of ductal epithelial cell type. Our results show that transdifferentiation—although being considered a direct cell fate conversion method—occurs through transient progenitor states orchestrated by stepwise activation of transduced and endogenous TFs: an endodermal progenitor state defined by activation of HHEX concurrently with FOXA2 and SOX17 and a pancreatic progenitor state marked by temporal activation of GATA4 (Figure 7), suggesting a paradigm shift in our understanding of transdifferentiation as a direct process in the context of Waddington's epigenetic landscape.

DISCUSSION

Transdifferentiation is a powerful method for converting differentiated cells to other specialized cell types through direct reprogramming. However, identifying TFs that are necessary and sufficient for cell fate conversion has been challenging, owing to the enormous compendia of cell and tissue types in human body. Computational frameworks based on gene regulatory network analysis[23,56] can predict the TFs required for cell fate conversion for broad tissue types, but experimental approaches are often required for determining the effective combinations of reprogramming factors for distinct cell types. Understanding the process of cellular differentiation is particularly challenging when the target cell types arise from the same multipotent progenitors and express similar TFs, such as definitive endoderm-derived tissues pancreas, liver, and intestine. For instance, a combination of FOXA1 and HNF4A can give rise to both liver- and intestine-like cells.[57] For more definitive hepatocyte identity, three TFs, FOXA3, HNF4A, and HNF1A, are required,[31] whereas combining KLF4/5 with FOXA1 and HNF4A results in stronger intestinal-like identity.[57] Here, we report the minimal and essential combination of six TFs, FOXA2, SOX17, PDX1, HNF1B, HNF6, and SOX9, for converting human fibroblasts to iPECs with epithelial-like morphology, and gene expression profile and functional characteristics similar to pancreatic ductal cells.

TF-mediated cell fate conversion experiments often suffer from low conversion efficiency and mixed cell populations. Therefore, identification of the "real" reprogrammed cells from the heterogeneous population is critical for their robust molecular analyses. Our FI-snMultiome-seq method with barcoded TF constructs enables characterization of the cells that express all stochastic combinations of transduced TFs, evaluation of the contribution of each individual TF to the acquired cell fate by comparing the cells

expressing sub-pools of reprogramming TFs, and segregation of TF expression from the lentiviral constructs and endogenous genes. Our results from conversion of fibroblasts to pancreatic ductal cells demonstrated that (1) direct cell fate conversion occurs through transient touch-and-go steps during which the cells dedifferentiate to distinct progenitor states, and (2) these steps are regulated by dynamic and temporally controlled action of both reprogramming TFs and endogenous tissue-specific TFs.

Previous developmental studies have implicated various TFs in pancreatic differentiation, such as FOXA2 and SOX17 in establishing definitive endoderm state[58,59] and HNF6 and HNF1B in exocrine ductal cell specification.[25,60] Interestingly, cell fate plasticity of progenitor populations was also recently reported during pancreatic development.[61] Here, our direct reprogramming model enabled determining the exact regulatory steps controlling pancreatic ductal cell identity. First, cell fate switch is initiated by FOXA2 and SOX17 along with activation of endogenous HHEX, mimicking the definitive endoderm progenitor state. Early activation of FOXA2 speaks for its pioneer factor activity necessary for cell fate conversion, consistent with its previously reported role during the development of endodermal organs.[62–64] Second, cell fate specification is controlled by HNF6 and HNF1B, involving activation of endogenous GATA4 that—consistent with previous reports[37]—specifically downregulates Shh pathway during pancreatic progenitor state. Finally, cell fate maintenance and maturation are controlled by HNF6, HNF1B, SOX9, and PDX1 commensurate with downregulation of the progenitor-state signals.

Dynamic expression and chromatin accessibility that we observed for the reprogramming factors suggest that although transduced as a pool and driven by a constitutively active promoter, they mediate the cell fate switch in a highly ordered and temporal fashion. During pluripotent stem cell reprogramming, sequential activation of pluripotency genes has been shown to be controlled by reorganization of genome topology,[65] suggesting a plausible mechanism for the regulation of transdifferentiation factors. Furthermore, our results demonstrate that despite being important regulators of iPEC reprogramming, ectopic expression of some TFs such as GATA4 was not necessary due to activation of respective endogenous genes at specific time points during iPEC reprogramming.

Taken together, we report here a protocol for generating pancreatic exocrine cells from human fibroblasts. This reprogramming system provides a platform for further studies of pancreatic malignancies and tissue engineering, suggesting the possibility of restoring impaired exocrine functions. In the broader context of regenerative medicine, our paradigm shift findings about the transient progenitor states during direct cell reprogramming can help in identifying proliferative cellular states that are necessary for developing feasible cell-based therapies for various human disorders. In conclusion, our results provide a temporally resolved map of the molecular events that determine the defined cellular states during direct cell fate conversion, opening avenues for further understanding of cell fate trajectories in light of these findings.

Limitations of the study

We have proposed that transdifferentiation occurs through transient progenitor states. Our findings are based on converting human fibroblasts to pancreatic exocrine cells, but further studies are required to determine whether similar mechanisms control transdifferentiation to other cell types. Another limitation of this study is the lifespan of iPECs. After 8 weeks in culture, their growth slows down, limiting their characterization and future applications. TERT immortalization of the iPECs resulted in cells with strong pancreatic ductal gene signatures but decreased enzyme activity. Thus, further investigation on extending the lifespan of iPECs without affecting their functional activity is required.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cells and cell line models
- METHOD DETAILS
  - Cloning of human ORFs and lentivirus production
  - Strategy for optimizing the combination of TFs for generating iPECs
  - Immunofluorescence staining
  - Flow cytometry analysis
  - Enzyme activity analysis
  - RNA isolation, quantitative real-time-PCR and bulk RNA-sequencing (RNA-seq)
  - Chromatin immunoprecipitation-sequencing (ChIP-seq)
  - Assay for Transposase-Accessible Chromatin-sequencing (ATAC-seq)
  - Factor barcoding design for FI-snMultiome-seq assay for snRNA-seq and snATAC-seq
  - FI-snMultiome-seq assay (snRNA-seq and snATAC-seq) from reprogramming cells
  - RNA-seq data processing
  - ATAC-seq and ChIP-seq data processing
  - ATAC-seq and ChIP-seq data clustering and motif searching
  - FI-snMultiome-seq (snRNA-seq and snATAC-seq) data processing
  - snATAC-seq TF motif analysis
  - Cell type score
  - Probabilistic cell type assignment
  - Trajectory and pseudotime analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.devcel.2023.08.023.

ACKNOWLEDGMENTS

We thank Professor Jussi Taipale, Professor Lauri Aaltonen, and Dr. Päivi Pihlajamaa for critical reading of the manuscript. We thank HiLIFE research infrastructures including the FIMM Single-Cell Analytics Unit, Biomedicum Flow

Cytometry Unit, Biomedicum Imaging Unit, Biomedicum Functional Genomics Unit (FuGU), and the FIMM NGS Genomics units at the University of Helsinki. We thank the Center for Scientific Computing (CSC), Finland for the computational infrastructure, Professor Lauri Aaltonen's laboratory facilities for genomics work, and Pinja Perkkiö and Kerim Yavuz for their technical assistance. This study was supported by the Academy of Finland (B.S.: 317807, 320114, 346065), Cancer Foundation Finland (B.S. and S.H.), Sigrid Jusélius Foundation (B.S. and S.H.), iCAN Digital Precision Cancer Medicine Flagship (B.S.; 320185), Jane and Aatos Erkko Foundation (B.S.), and European Union's Horizon 2020 Research and Innovation Programme under grant agreement no. 965193 (DECIDER; S.H.).

AUTHOR CONTRIBUTIONS

B.S. conceived and supervised the study. L.F. designed and performed all experiments. K.Z. analyzed the FI-snMultiome-seq data with the support from S.H. N.P. performed the bulk RNA-seq and ATAC-seq data analyses. B.S. and L.F. wrote the manuscript with contributions from all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: March 8, 2023
Revised: July 13, 2023
Accepted: August 16, 2023
Published: September 25, 2023

REFERENCES

1. Wang, H.F., Yang, Y.C., Liu, J.D., and Qian, L. (2021). Direct cell reprogramming: approaches, mechanisms and progress. Nat. Rev. Mol. Cell Biol. *22*, 410–424. https://doi.org/10.1038/s41580-021-00335-z.
2. Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. Nature *463*, 1035–1041. https://doi.org/10.1038/nature08797.
3. Ieda, M., Fu, J.D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell *142*, 375–386. https://doi.org/10.1016/j.cell.2010.07.002.
4. Sekiya, S., and Suzuki, A. (2011). Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. Nature *475*, 390–393. https://doi.org/10.1038/nature10263.
5. Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. Genes Dev. *28*, 2679–2692. https://doi.org/10.1101/gad.253443.114.
6. Biddy, B.A., Kong, W., Kamimoto, K., Guo, C., Waye, S.E., Sun, T., and Morris, S.A. (2018). Single-cell mapping of lineage and identity in direct reprogramming. Nature *564*, 219–224. https://doi.org/10.1038/s41586-018-0744-4.
7. Francesconi, M., Di Stefano, B., Berenguer, C., de Andrés-Aguayo, L., Plana-Carmona, M., Mendez-Lago, M., Guillaumet-Adkins, A., Rodriguez-Esteban, G., Gut, M., Gut, I.G., et al. (2019). Single cell RNA-seq identifies the origins of heterogeneity in efficient cell transdifferentiation and reprogramming. eLife *8*, e41627. https://doi.org/10.7554/eLife.41627.
8. Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., Gould, J., Liu, S., Lin, S., Berube, P., et al. (2019). Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. Cell *176*, 1517. https://doi.org/10.1016/j.cell.2019.02.026.
9. Qian, H., Kang, X.J., Hu, J., Zhang, D.Y., Liang, Z.Y., Meng, F., Zhang, X., Xue, Y.C., Maimon, R., Dowdy, S.F., et al. (2020). Reversing a model of Parkinson's disease with in situ converted nigral neurons (vol 53, pg 863, 2020). Nature *584*, E17. https://doi.org/10.1038/s41586-020-2583-3.
10. Guo, Z.Y., Zhang, L., Wu, Z., Chen, Y.C., Wang, F., and Chen, G. (2014). In vivo direct reprogramming of reactive glial cells into functional neurons after brain injury and in an Alzheimer's disease model. Cell Stem Cell *14*, 188–202. https://doi.org/10.1016/j.stem.2013.12.001.
11. Sahu, B., Pihlajamaa, P., Zhang, K.Y., Palin, K., Ahonen, S., Cervera, A., Ristimäki, A., Aaltonen, L.A., Hautaniemi, S., and Taipale, J. (2021). Human cell transformation by combined lineage conversion and oncogene expression. Oncogene *40*, 5533–5547. https://doi.org/10.1038/s41388-021-01940-0.
12. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., and Melton, D.A. (2008). In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. Nature *455*, 627–632. https://doi.org/10.1038/nature07314.
13. Lee, J., Sugiyama, T., Liu, Y.H., Wang, J., Gu, X., Lei, J., Markmann, J.F., Miyazaki, S., Miyazaki, J.I., Szot, G.L., et al. (2013). Expansion and conversion of human pancreatic ductal cells into insulin-secreting endocrine cells. eLife *2*, e00940.
14. Chen, Y.J., Finkbeiner, S.R., Weinblatt, D., Emmett, M.J., Tameire, F., Yousefi, M., Yang, C.H., Maehr, R., Zhou, Q., Shemer, R., et al. (2014). De novo formation of insulin-producing "neo-beta cell islets" from intestinal crypts. Cell Rep. *6*, 1046–1058. https://doi.org/10.1016/j.celrep.2014.02.013.
15. Ariyachet, C., Tovaglieri, A., Xiang, G.J., Lu, J.Q., Shah, M.S., Richmond, C.A., Verbeke, C., Melton, D.A., Stanger, B.Z., Mooney, D., et al. (2016). Reprogrammed Stomach Tissue as a Renewable Source of Functional beta Cells for blood glucose Regulation. Cell Stem Cell *18*, 410–421. https://doi.org/10.1016/j.stem.2016.01.003.
16. Ellis, C., Ramzy, A., and Kieffer, T.J. (2017). Regenerative medicine and cell-based approaches to restore pancreatic function. Nat. Rev. Gastroenterol. Hepatol. *14*, 612–628. https://doi.org/10.1038/nrgastro.2017.93.
17. Patel, S.A., Hirosue, S., Rodrigues, P., Vojtasova, E., Richardson, E.K., Ge, J.F., Syafruddin, S.E., Speed, A., Papachristou, E.K., Baker, D., et al. (2022). The renal lineage factor PAX8 controls oncogenic signalling in kidney cancer. Nature *606*, 999–1006. https://doi.org/10.1038/s41586-022-04809-8.
18. Baggiolini, A., Callahan, S.J., Montal, E., Weiss, J.M., Trieu, T., Tagore, M.M., Tischfield, S.E., Walsh, R.M., Suresh, S., Fan, Y.J., et al. (2021). Developmental chromatin programs determine oncogenic competence in melanoma. Science *373*, 1104. https://doi.org/10.1126/science.abc1048.
19. Perkins, C.M., Jiang, J.M., Hakimjavadi, H., Bray, J.K., Gosling, A., da Silva, L., Bulut, G., Ali, J., Setiawan, W., Campbell-Thompson, M., et al. (2022). Transcriptional profile of human pancreatic acinar ductal metaplasia. Cancer Res. *82*, 532–543.
20. Ayars, M., O'Sullivan, E., Macgregor-Das, A., Shindo, K., Kim, H., Borges, M., Yu, J., Hruban, R.H., and Goggins, M. (2017). IL2RG, identified as overexpressed by RNA-seq profiling of pancreatic intraepithelial neoplasia, mediates pancreatic cancer growth. Oncotarget *8*, 83370–83383. https://doi.org/10.18632/oncotarget.19848.
21. Jennings, R.E., Berry, A.A., Strutt, J.P., Gerrard, D.T., and Hanley, N.A. (2015). Human pancreas development. Development *142*, 3126–3137. https://doi.org/10.1242/dev.120063.
22. Petersen, M.B.K., Gonçalves, C.A.C., Kim, Y.H., and Grapin-Botton, A. (2018). Recapitulating and deciphering human pancreas development from human pluripotent stem cells in a dish. Curr. Top. Dev. Biol. *129*, 143–190. https://doi.org/10.1016/bs.ctdb.2018.02.009.
23. Rackham, O.J., Firas, J., Fang, H., Oates, M.E., Holmes, M.L., Knaupp, A.S., FANTOM Consortium, Suzuki, H., Nefzger, C.M., Daub, C.O., et al. (2016). A predictive computational framework for direct reprogramming between human cell types. Nat. Genet. *48*, 331–335. https://doi.org/10.1038/ng.3487.
24. Hoang, C.Q., Hale, M.A., Azevedo-Pouly, A.C., Elsässer, H.P., Deering, T.G., Willet, S.G., Pan, F.C., Magnuson, M.A., Wright, C.V., Swift, G.H., et al. (2016). Transcriptional maintenance of pancreatic acinar identity,

differentiation, and homeostasis by PTF1A. Mol. Cell. Biol. *36*, 3033–3047. https://doi.org/10.1128/MCB.00358-16.

25. Pierreux, C.E., Poll, A.V., Kemp, C.R., Clotman, F., Maestro, M.A., Cordi, S., Ferrer, J., Leyns, L., Rousseau, G.G., and Lemaigre, F.P. (2006). The transcription factor hepatocyte nuclear factor-6 controls the development of pancreatic ducts in the mouse. Gastroenterology *130*, 532–541. https://doi.org/10.1053/j.gastro.2005.12.005.
26. Villamayor, L., Cano, D.A., and Rojas, A. (2020). GATA factors in pancreas development and disease. IUBMB Life *72*, 80–88. https://doi.org/10.1002/iub.2170.
27. Solar, M., Cardalda, C., Houbracken, I., Martín, M., Maestro, M.A., De Medts, N., Xu, X.B., Grau, V., Heimberg, H., Bouwens, L., et al. (2009). Pancreatic Exocrine Duct Cells Give Rise to insulin-Producing beta Cells during Embryogenesis but Not after Birth. Dev. Cell *17*, 849–860.
28. Schaffer, A.E., Freude, K.K., Nelson, S.B., and Sander, M. (2010). Nkx6 transcription factors and Ptf1a function as antagonistic lineage determinants in multipotent pancreatic progenitors. Dev. Cell *18*, 1022–1029.
29. Shih, H.P., Kopp, J.L., Sandhu, M., Dubois, C.L., Seymour, P.A., Grapin-Botton, A., and Sander, M. (2012). A Notch-dependent molecular circuitry initiates pancreatic endocrine and ductal cell differentiation. Development *139*, 2488–2499.
30. Hale, M.A., Swift, G.H., Hoang, C.Q., Deering, T.G., Masui, T., Lee, Y.K., Xue, J., and MacDonald, R.J. (2014). The nuclear hormone receptor family member NR5A2 controls aspects of multipotent progenitor cell formation and acinar differentiation during pancreatic organogenesis. Development *141*, 3123–3133.
31. Huang, P.Y., Zhang, L.D., Gao, Y.M., He, Z.Y., Yao, D., Wu, Z.T., Cen, J., Chen, X.T., Liu, C.C., Hu, Y.P., et al. (2014). Direct reprogramming of human fibroblasts to functional and expandable hepatocytes. Cell Stem Cell *14*, 370–384. https://doi.org/10.1016/j.stem.2014.01.003.
32. Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A.M., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M., et al. (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature *534*, 391–395. https://doi.org/10.1038/nature18323.
33. Kumpulainen, T., and Jalovaara, P. (1981). Immunohistochemical localization of carbonic anhydrase isoenzymes in the human pancreas. Gastroenterology *80*, 796–799.
34. Muraro, M.J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning, E.J.P., et al. (2016). A single-cell transcriptome atlas of the human pancreas. Cell Syst. *3*, 385–394.e3. https://doi.org/10.1016/j.cels.2016.09.002.
35. Hendley, A.M., Rao, A.A., Leonhardt, L., Ashe, S., Smith, J.A., Giacometti, S., Peng, X.L., Jiang, H.L., Berrios, D.I., Pawlak, M., et al. (2021). Single-cell transcriptome analysis defines heterogeneity of the murine pancreatic ductal tree. eLife *10*, e67776.
36. Pei, D.Q., Shu, X.D., Gassama-Diagne, A., and Thiery, J.P. (2019). Mesenchymal-epithelial transition in development and reprogramming. Nat. Cell Biol. *21*, 44–53. https://doi.org/10.1038/s41556-018-0195-z.
37. Xuan, S.H., and Sussel, L. (2016). GATA4 and GATA6 regulate pancreatic endoderm identity through inhibition of hedgehog signaling. Development *143*, 780–786. https://doi.org/10.1242/dev.127217.
38. Hebrok, M., Kim, S.K., and Melton, D.A. (1998). Notochord repression of endodermal Sonic hedgehog permits pancreas development. Genes Dev. *12*, 1705–1713. https://doi.org/10.1101/gad.12.11.1705.
39. Rouillard, A.D., Gundersen, G.W., Fernandez, N.F., Wang, Z.C., Monteiro, C.D., McDermott, M.G., and Ma'ayan, A. (2016). The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database-Oxford *2016*, baw100. https://doi.org/10.1093/database/baw100.
40. Chan, D.W., Liu, V.W.S., Leung, L.Y., Yao, K.M., Chan, K.K.L., Cheung, A.N.Y., and Ngan, H.Y.S. (2011). Zic2 synergistically enhances Hedgehog signalling through nuclear retention of Gli1 in cervical cancer cells. J. Pathol. *225*, 525–534. https://doi.org/10.1002/path.2901.
41. Alfaro, A.C., Roberts, B., Kwong, L., Bijlsma, M.F., and Roelink, H. (2014). Ptch2 mediates the Shh response in Ptch1(-/-) cells. Development *141*, 3331–3339. https://doi.org/10.1242/dev.110056.
42. Li, D.W., Shu, X.D., Zhu, P., and Pei, D.Q. (2021). Chromatin accessibility dynamics during cell fate reprogramming. EMBO Rep. *22*, e51644. https://doi.org/10.15252/embr.202051644.
43. Bort, R., Martinez-Barbera, J.P., Beddington, R.S.P., and Zaret, K.S. (2004). Hex homeobox gene-dependent tissue positioning is required for organogenesis of the ventral pancreas. Development *131*, 797–806. https://doi.org/10.1242/dev.00965.
44. Zhao, H., Han, D.D., Dawid, I.B., Pieler, T., and Chen, Y.L. (2012). Homeoprotein hhex-induced conversion of intestinal to ventral pancreatic precursors results in the formation of giant pancreata in Xenopus embryos. Proc. Natl. Acad. Sci. USA *109*, 8594–8599. https://doi.org/10.1073/pnas.1206547109.
45. Villasenor, A., Gauvrit, S., Collins, M.M., Maischein, H.M., and Stainier, D.Y.R. (2020). Hhex regulates the specification and growth of the hepato-pancreatic ductal system. Dev. Biol. *458*, 228–236. https://doi.org/10.1016/j.ydbio.2019.10.021.
46. Bentsen, M., Goymann, P., Schultheis, H., Klee, K., Petrova, A., Wiegandt, R., Fust, A., Preussner, J., Kuenne, C., Braun, T., et al. (2020). ATAC-seq footprinting unravels kinetics of transcription factor binding during zygotic genome activation. Nat. Commun. *11*, 4267. https://doi.org/10.1038/s41467-020-18035-1.
47. Isenmann, S., Arthur, A., Zannettino, A.C., Turner, J.L., Shi, S., Glackin, C.A., and Gronthos, S. (2009). TWIST family of basic helix-loop-helix transcription factors mediate human mesenchymal stem cell growth and commitment. Stem Cells *27*, 2457–2468. https://doi.org/10.1002/stem.181.
48. Ianevski, A., Giri, A.K., and Aittokallio, T. (2022). Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. Nat. Commun. *13*, 1246. https://doi.org/10.1038/s41467-022-28803-w.
49. Segerstolpe, Å., Palasantza, A., Eliasson, P., Andersson, E.M., Andréasson, A.C., Sun, X., Picelli, S., Sabirsh, A., Clausen, M., Bjursell, M.K., et al. (2016). Single-cell transcriptome profiling of human pancreatic islets in health and Type 2 diabetes. Cell Metab. *24*, 593–607. https://doi.org/10.1016/j.cmet.2016.08.020.
50. Chanda, S., Ang, C.E., Davila, J., Pak, C., Mall, M., Lee, Q.Y., Ahlenius, H., Jung, S.W., Südhof, T.C., and Wernig, M. (2014). Generation of induced neuronal cells by the single reprogramming factor ASCL1. Stem Cell Rep. *3*, 282–296. https://doi.org/10.1016/j.stemcr.2014.05.020.
51. Ng, A.H.M., Khoshakhlagh, P., Arias, J.E.R., Pasquini, G., Wang, K., Swiersy, A., Shipman, S.L., Appleton, E., Kiaee, K., Kohman, R.E., et al. (2021). A comprehensive library of human transcription factors for cell fate engineering. Nat. Biotechnol. *39*, 510–519. https://doi.org/10.1038/s41587-020-0742-6.
52. Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., et al. (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell *155*, 621–635. https://doi.org/10.1016/j.cell.2013.09.028.
53. Moore-Scott, B.A., Opoka, R., Lin, S.C.J., Kordich, J.J., and Wells, J.M. (2007). Identification of molecular markers that are expressed in discrete anterior-posterior domains of the endoderm from the gastrula stage to mid-gestation. Dev. Dyn. *236*, 1997–2003. https://doi.org/10.1002/dvdy.21204.
54. Nostro, M.C., Sarangi, F., Ogawa, S., Holtzinger, A., Corneo, B., Li, X.L., Micallef, S.J., Park, I.H., Basford, C., Wheeler, M.B., et al. (2011). Stage-specific signaling through TGF beta family members and WNT regulates patterning and pancreatic specification of human pluripotent stem cells. Development *138*, 861–871. https://doi.org/10.1242/dev.055236.
55. Peng, G.D., Suo, S.B., Cui, G.Z., Yu, F., Wang, R., Chen, J., Chen, S.R., Liu, Z.W., Chen, G.Y., Qian, Y., et al. (2020). Molecular architecture of lineage allocation and tissue organization in early mouse embryo (vol 572, pg 528, 2019). Nature *586*, E7. https://doi.org/10.1038/s41586-020-2755-1.

56. Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E.L., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. Cell *158*, 903–915. https://doi.org/10.1016/j.cell.2014.07.020.

57. Morris, S.A., Cahan, P., Li, H., Zhao, A.M., San Roman, A.K.S., Shivdasani, R.A., Collins, J.J., and Daley, G.Q. (2014). Dissecting engineered cell types and enhancing cell fate conversion via CellNet. Cell *158*, 889–902. https://doi.org/10.1016/j.cell.2014.07.021.

58. Burtscher, I., and Lickert, H. (2009). Foxa2 regulates polarity and epithelialization in the endoderm germ layer of the mouse embryo. Development *136*, 1029–1038. https://doi.org/10.1242/dev.028415.

59. Viotti, M., Nowotschin, S., and Hadjantonakis, A.K. (2014). SOX17 links gut endoderm morphogenesis and germ layer segregation. Nat. Cell Biol. *16*, 1146–1156. https://doi.org/10.1038/ncb3070.

60. Reichert, M., and Rustgi, A.K. (2011). Pancreatic ductal cells in development, regeneration, and neoplasia. J. Clin. Invest. *121*, 4572–4578. https://doi.org/10.1172/JCI57131.

61. Willnow, D., Benary, U., Margineanu, A., Vignola, M.L., Konrath, F., Pongrac, I.M., Karimaddini, Z., Vigilante, A., Wolf, J., and Spagnoli, F.M. (2021). Quantitative lineage analysis identifies a hepato-pancreato-biliary progenitor niche. Nature *597*, 87–91. https://doi.org/10.1038/s41586-021-03844-1.

62. Lee, K., Cho, H., Rickert, R.W., Li, Q.V., Pulecio, J., Leslie, C.S., and Huangfu, D.W. (2019). FOXA2 is required for enhancer priming during pancreatic differentiation. Cell Rep. *28*, 382–393.e7. https://doi.org/10.1016/j.celrep.2019.06.034.

63. Wang, A., Yue, F., Li, Y., Xie, R.Y., Harper, T., Patel, N.A., Muth, K., Palmer, J., Qiu, Y.J., Wang, J.Z., et al. (2015). Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. Cell Stem Cell *16*, 386–399. https://doi.org/10.1016/j.stem.2015.02.013.

64. Geusz, R.J., Wang, A., Lam, D.K., Vinckier, N.K., Alysandratos, K.D., Roberts, D.A., Wang, J.Z., Kefalopoulou, S., Ramirez, A., Qiu, Y.J., et al. (2021). Sequence logic at enhancers governs a dual mechanism of endodermal organ fate induction by FOXA pioneer factors. Nat. Commun. *12*, 6636. https://doi.org/10.1038/s41467-021-26950-0.

65. Stadhouders, R., Vidal, E., Serra, F., Di Stefano, B., Le Dily, F., Quilez, J., Gomez, A., Collombet, S., Berenguer, C., Cuartero, Y., et al. (2018). Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat. Genet. *50*, 238–249. https://doi.org/10.1038/s41588-017-0030-7.

66. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

67. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

68. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169. https://doi.org/10.1093/bioinformatics/btu638.

69. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

70. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA *102*, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

71. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. *10*, 1523. https://doi.org/10.1038/s41467-019-09234-6.

72. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/Nmeth.1923.

73. Wang, S., Sun, H., Ma, J., Zang, C., Wang, C., Wang, J., Tang, Q., Meyer, C.A., Zhang, Y., and Liu, X.S. (2013). Target analysis by integration of transcriptome and ChIP-seq data with BETA. Nat Protoc. *8*, 2502–2515. https://doi.org/10.1038/nprot.2013.150.

74. Fornes, O., Castro-Mondragon, J.A., Khan, A., van der Lee, R., Zhang, X., Richmond, P.A., Modi, B.P., Correard, S., Gheorghe, M., Baranašić, D., et al. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. Nucleic Acids Res *48*, D87–D92. https://doi.org/10.1093/nar/gkz1001.

75. Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y.M., Albu, M., Chen, X.T., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell. *175*, 598–599. https://doi.org/10.1016/j.cell.2018.09.045.

76. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137. 10.13710.1186/gb-2008-9-9-r137.

77. Ramírez, F., Dündar, F., Diehl, S., Grüning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. *42*, W187–W191. https://doi.org/10.1093/nar/gku365.

78. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192. https://doi.org/10.1093/bib/bbs017.

79. Qiu, X., Feit, A.S., Feiglin, A., Xie, Y., Kesten, N., Taing, L., Perkins, J., Gu, S., Li, Y., Cejas, P., et al. (2021). CoBRA: containerized bioinformatics workflow for reproducible ChIP/ATAC-seq analysis. Genomics Proteomics Bioinformatics *19*, 652–661. https://doi.org/10.1016/j.gpb.2020.11.007.

80. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

81. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

82. Hao, Y.H., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S.W., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

83. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2022). Author Correction: single-cell chromatin state analysis with Signac. Nat. Methods *19*, 257. https://doi.org/10.1038/s41592-022-01393-7.

84. Schep, A.N., Wu, B.J., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods *14*, 975–978. https://doi.org/10.1038/Nmeth.4401.

85. Gayoso, A., Lopez, R., Xing, G., Boyeau, P., Valiollah Pour Amiri, V., Hong, J., Wu, K., Jayasuriya, M., Mehlman, E., Langevin, M., et al. (2022). A Python library for probabilistic analysis of single-cell omics data. Nat. Biotechnol. *40*, 163–166. https://doi.org/10.1038/s41587-021-01206-w.

86. Cao, J.Y., Spielmann, M., Qiu, X.J., Huang, X.F., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.

87. Vanheer L., Fantuzzi F., To S.K., Schiavo A.A., Haele M.V., Haesen T., Yi X., Janiszewski A., Chappell J., Rihoux A., et al. Predicting the key regulators of cell identity in human adult pancreas. 2022. doi:10.1101/2020.09.23.310094.

88. Nissim, S., Weeks, O., Talbot, J.C., Hedgepeth, J.W., Wucherpfennig, J., Schatzman-Bone, S., Swinburne, I., Cortes, M., Alexa, K., Megason, S., et al. (2016). Iterative use of nuclear receptor Nr5a2 regulates multiple stages of liver and pancreas development. Dev. Biol. *418*, 108–123. https://doi.org/10.1016/j.ydbio.2016.07.019.

89. Sahu, B., Laakso, M., Ovaska, K., Mirtti, T., Lundin, J., Rannikko, A., Sankila, A., Turunen, J.P., Lundin, M., Konsti, J., et al. (2011). Dual role of FoxA1 in androgen receptor binding to chromatin, androgen signalling and prostate cancer. EMBO J *30*, 3962–3976. https://doi.org/10.1038/emboj.2011.328.

90. Buenrostro J.D., Wu B., Chang H.Y., Greenleaf W.J. ATAC-seq: A method for assaying chromatin accessibility genome-wide. Curr. Protoc. Mol. Biol. 2015;109. doi:10.1002/0471142727.mb2129s109. 21.29.1–21.29.9.

91. Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962. https://doi.org/10.1038/Nmeth.4396.

92. Sahu, B., Hartonen, T., Pihlajamaa, P., Wei, B., Dave, K., Zhu, F.J., Kaasinen, E., Lidschreiber, K., Lidschreiber, M., Daub, C.O., et al. (2022). Sequence determinants of human gene regulatory elements. Nat. Genet. *54*, 283–294. https://doi.org/10.1038/s41588-021-01009-4.

93. Muhl, L., Genové, G., Leptidis, S., Liu, J.P., He, L.Q., Mocci, G., Sun, Y., Gustafsson, S., Buyandelger, B., Chivukula, I.V., et al. (2020). Single-cell analysis uncovers fibroblast heterogeneity and criteria for fibroblast and mural cell identification and discrimination. Nat. Commun. *11*, 3953. https://doi.org/10.1038/s41467-020-17740-1.

94. Zhang, A.W., O'Flanagan, C., Chavez, E.A., Lim, J.L.P., Ceglia, N., McPherson, A., Wiens, M., Walters, P., Chan, T., Hewitson, B., et al. (2019). Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. Nat. Methods *16*, 1007–1015. https://doi.org/10.1038/s41592-019-0529-1.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat Anti-FOXA2	R&D	AF2400; RRID: AB_2294104
Goat Anti-SOX17	R&D	AF1924; RRID: AB_355060
Mouse Anti-PDX-1 (B-11)	Santa Cruz Biotechnology	sc-390792; RRID: AB_2938928
Mouse Anti-HNF-6 (G-10)	Santa Cruz Biotechnology	sc-376167, RRID: AB_10989089
Rabbit Anti-SOX9	Millipore	AB5535, RRID: AB_2239761
Rabbit Anti-HNF1B	Sigma-Aldrich	HPA002083; RRID: AB_1080232
Rabbit Anti-Osteopontin	Proteintech	22952-1-AP, RRID: AB_2783651
Donkey Anti-Goat (Alexa Fluor 488) secondary antibody	Abcam	ab150129, RRID: AB_2687506
Goat Anti-Mouse (Alexa Fluor 488) secondary antibody	Thermo Fisher Scientific	A-11029, RRID: AB_2534088
Goat Anti-Rabbit (Alexa Fluor 594) secondary antibody	Thermo Fisher Scientific	A-11012, RRID: AB_2534079
Goat Anti-Rabbit (Alexa Fluor 488) secondary antibody	Thermo Fisher Scientific	A-11034, RRID: AB_2576217
PE Anti-Human CD24	BioLegend	311105, RRID: AB_314854
APC Anti-Human CD133/2	Miltenyi Biotec	130-113-746, RRID: AB_2726285
Rabbit Anti-H3K27ac	Diagenode	C15410196, RRID: AB_2637079
Normal rabbit IgG	Santa Cruz Biotechnology	sc-2027, RRID: AB_737197
Bacterial and virus strains		
One Shot™ Stbl3™ Chemically Competent E. coli	Thermo Fisher Scientific	C737303
One Shot™ ccdB Survival™ 2 T1R Competent Cells	Thermo Fisher Scientific	A10460
Chemicals, peptides, and recombinant proteins		
Ascorbic Acid	Sigma-Aldrich	A1300000
Retinoic Acid	Sigma-Aldrich	R2625
PD0325901	Sigma-Aldrich	PZ0162
CHIR99021	StemMACS	130-106-539
LDN193189	StemMACS	130-103-925
Activin A	Peprotech	AF-120-14E
FGF7	Peprotech	AF-100-19
Critical commercial assays		
Carbonic Anhydrase Activity Assay Kit	Biovision	K472
Trypsin Activity Colorimetric Assay Kit	Sigma-Aldrich	MAK290
Amylase Activity Assay kit	Sigma-Aldrich	MAK009
KAPA mRNA HyperPrep Kit	Roche	KR1352
Illumina Tagment DNA Enzyme and Buffer Small Kit	Illumina	20034197
Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Bundle	10xGenomics	PN-1000285
Deposited data		
Raw data	This paper	GEO: GSE216859
Bulk RNA-seq data of human acinar cells	Perkins et al.[19]	GEO: GSE179248
Bulk RNA-seq data of human ductal cells	Ayars et al.[20]	GEO: GSE96784
Single-cell RNA-seq data of human pancreatic cells	Segerstolpe et al.[49]	ArrayExpress: E-MTAB-5060
Human reference genome NCBI build 38, GRCh38	Genome Reference Consortium	https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405
ENCODE blacklisted genomic regions for hg38	ENCODE	ENCODE: ENCFF356LFX
Experimental models: Cell lines		
HFF	ATCC	CRL-2429

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Oligonucleotides		
Primer: *GAPDH* Forward: GGCCTCCAAGGAGTAAGACC	This paper	N/A
Primer: *GAPDH* Reverse: AGGGGAGATTCAGTGTGGTG	This paper	N/A
Primer: *CA2* Forward: GACCCCTGGATGGCACTTAC	This paper	N/A
Primer: *CA2* Reverse: AGTGAACCAAGTGAAGTTCTGC	This paper	N/A
Primer: *CFTR* Forward: GGCTCCTTGGAAACACTCCTC	This paper	N/A
Primer: *CFTR* Reverse: CAAAGTGTCGGCTACTCCCA	This paper	N/A
Primer: *CPA1* Forward: CTTCACGCACAGCACGAATC	This paper	N/A
Primer: *CPA1* Reverse: GGACTTGACCTCCACTTCGG	This paper	N/A
Primer: *PRSS1* Forward: ACTGCTACAAGTCCCGCATC	This paper	N/A
Primer: *PRSS1* Reverse: TTCCTGTCGTATTGGGGGTG	This paper	N/A
Recombinant DNA		
pLV-eGFP	Addgene	36083
psPAX2	Addgene	12260
pMD2.G	Addgene	12259
pLenti6/V5-DEST™	Thermo Fisher Scientific	V49610
pLenti6/V5-FOXA2	This paper	N/A
pLenti6/V5-GATA4	This paper	N/A
pLenti6/V5-GATA6	This paper	N/A
pLenti6/V5-HEYL	This paper	N/A
pLenti6/V5-HNF1B	This paper	N/A
pLenti6/V5-HNF6	This paper	N/A
pLenti6/V5-NKX6.1	This paper	N/A
pLenti6/V5-NR5A2	This paper	N/A
pLenti6/V5-PDX1	This paper	N/A
pLenti6/V5-PTF1A	This paper	N/A
pLenti6/V5-RBPJL	This paper	N/A
pLenti6/V5-SOX9	This paper	N/A
pLenti6/V5-SOX17	This paper	N/A
pLenti6/V5-XBP1	This paper	N/A
pLenti6/V5-Barcode-FOXA2	This paper	N/A
pLenti6/V5-Barcode-HNF1B	This paper	N/A
pLenti6/V5-Barcode-HNF6	This paper	N/A
pLenti6/V5-Barcode-PDX1	This paper	N/A
pLenti6/V5-Barcode-SOX9	This paper	N/A
pLenti6/V5-Barcode-SOX17	This paper	N/A
Software and algorithms		
Fiji (v1.53)	NIH	https://imagej.net/software/fiji/
FlowJo (v10.6.1)	BD	https://www.flowjo.com/
GraphPad Prism (8.3.0)	GraphPad	https://www.graphpad.com/
FastQC	Babraham Bioinformatics	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
STAR aligner (v2.7.5c)	Dobin et al.[66]	https://github.com/alexdobin/STAR
Samtools (v1.9)	Li et al.[67]	http://samtools.sourceforge.net/
HTSeq-count (v0.11.2)	Anders et al.[68]	https://github.com/htseq/htseq
DESeq2 (v1.30.1)	Love et al.[69]	https://github.com/mikelove/DESeq2
Gene Set Enrichment Analysis (GSEA) (v4.2.3)	Subramanian et al.[70]	https://github.com/GSEA-MSigDB/gseapreranked-gpmodule
Metascape	Zhou et al.[71]	https://metascape.org/gp/index.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bowtie2 (v2.2.5)	Langmead et al.[72]	https://github.com/BenLangmead/bowtie2
Picard MarkDuplicates	N/A	http://broadinstitute.github.io/picard/
BETA	Wang et al.[73]	http://cistrome.org/BETA/index.html
JASPAR 2020	Fornes et al.[74]	https://jaspar.genereg.net/
The Human Transcription Factors	Lambert et al.[75]	http://humantfs.ccbr.utoronto.ca/
MACS2 (v2.1.1, v2.2.7.1)	Zhang et al.[76]	http://github.com/taoliu/MACS/
deeptools (v3.1.3)	Ramirez et al.[77]	https://github.com/deeptools/deepTools
Integrative Genomics Viewer (IGV) (v2.5)	Thorvaldsdóttir et al.[78]	https://software.broadinstitute.org/software/igv/
CoBRA (v2.0)	Qiu et al.[79]	https://bitbucket.org/cfce/cobra
HOMER (v4.10.4)	Heinz et al.[80]	http://homer.ucsd.edu/homer/motif/
BEDTools (v2.30.0)	Quinlan et al.[81]	http://bedtools.readthedocs.org/
TOBIAS (v0.13.3)	Bentsen et al.[46]	https://github.com/loosolab/TOBIAS
R (v4.0.5, v4.2.1)	R Core Team. R Foundation for Statistical Computing. Vienna, Austria.	https://www.r-project.org/
Seurat (v4.1.1)	Hao et al.[82]	https://satijalab.org/seurat/
Signac (v1.7.0)	Stuart et al.[83]	https://stuartlab.org/signac/
chromVAR (v1.18.0)	Schep et al.[84]	https://github.com/GreenleafLab/chromVAR
ScType (v1.0)	Ianevski et al.[48]	https://github.com/IanevskiAleksandr/sc-type
SeuratWrappers (v0.3.0)	Hao et al.[82]	https://github.com/satijalab/seurat-wrappers
scvi-tools (v1.0.0)	Gayoso et al.[85]	https://scvi-tools.org/
Monocle3 (v1.0.0)	Cao et al.[86]	https://cole-trapnell-lab.github.io/monocle3/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact Biswajyoti Sahu (biswajyoti.sahu@helsinki.fi, biswajyoti.sahu@ncmm.uio.no).

Materials availability

Plasmids generated in this study are available upon request from the lead contact with a completed Materials Transfer Agreement.

Data and code availability

All sequencing data generated in this study are available under GEO: GSE216859. Bulk RNA-seq data of human acinar and ductal cells were downloaded from GEO: GSE179248 and GEO: GSE96784, respectively. Single-cell RNA-seq data of human pancreatic acinar, ductal, alpha, beta, delta and gamma cells were downloaded from ArrayExpress: E-MTAB-5060. Blacklisted genomic regions for hg38 were downloaded from ENCODE: ENCFF356LFX. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cells and cell line models

Human foreskin fibroblasts (HFF; ATCC, #CRL-2429) were cultured in fibroblast medium (DMEM supplemented with 10% FBS, 2 mM L-Glutamine, 1 mM Sodium pyruvate and 1% penicillin-streptomycin). Early-passage HFFs were plated on matrigel-coated (Corning, #356230) dish on day 0 and transduced with constructs for TF expression (MOI = 1 for SOX17, FOXA2 and PDX1; MOI = 2 for HNF1B, HNF6 and SOX9) with 8 μg/ml polybrene on day 1. The medium was changed to fresh fibroblast medium on day 2. The basal medium from day 3 to day 9 was DMEM/F12 (Gibco, #12634028) with 50 μg/ml ascorbic acid (Sigma-Aldrich, #A1300000), 1% B27 (Gibco, #17504044), 1% KnockOut serum replacement (Gibco, #10828010), 1X GlutaMAX (Gibco, #35050038), 1% penicillin-streptomycin (Gibco, #15140122), 25 mM HEPES (Gibco, #15630080), 0.1 mM 2-mercaptoethanol (Gibco, #21985023), 0.5X ITS-X (Gibco, #51500056). On day 3, cells were cultured in basal medium supplemented with 100 ng/ml Activin A (Peprotech, #AF-120-14E), 1 μM CHIR99021 (StemMACS, #130-106-539) and 50 ng/ml FGF7 (Peprotech, #AF-100-19). After two days, the medium was changed to basal medium supplemented with 100 ng/ml Activin A, and 50 ng/ml FGF7. On day 6, the medium was changed to basal medium supplemented with 2 μM retinoic acid (Sigma-Aldrich, #R2625), 500 nM PD0325901 (Sigma-Aldrich, #PZ0162), 200 nM

LDN193189 (StemMACS, #130-103-925). After one day, the medium was changed to basal medium with 100 ng/ml Activin A and 500 nM PD0325901. From day 10 onwards, the cells were maintained in serum-free human epithelial cell medium (Cell Biologics, #PB-H-6621) supplemented with 5% KnockOut serum replacement and 50 μg/ml ascorbic acid. During days 21–28, iPEC colonies were picked under microscope in a sterile hood, disassociated using Accutase (Gibco, #A1110501) and replated on matrigel-coated dish. Lentiviral expression construct for hTERT (MOI=0.5) was used for the immortalization of iPECs. The images of reprogramming cells were taken with ZEISS Axio Vert A1 microscope. All cell images throughout this manuscript were analyzed using Fiji (v1.53).

METHOD DETAILS

Cloning of human ORFs and lentivirus production

The full-length ORFs of all individual TFs were obtained from GenScript and cloned into lentiviral destination vector pLenti6/V5-DEST™ (Thermo Fisher Scientific, #V49610) using the Gateway™ LR Clonase™ II (Thermo Fisher Scientific, #11791020). The recombinant vectors were transformed into Stbl3™ competent cells (Thermo Fisher Scientific, #C737303) followed by single-colony screening. The plasmid pLV-eGFP (#36083) was obtained from Addgene. For virus production, each TF expression plasmid was co-transfected with the packaging plasmids psPAX2 (Addgene #12260) and pMD2.G (Addgene, #12259) in 4:3:1 ratio into 293FT cells (Thermo Fisher Scientific, #R70007) with Lipofectamine 2000 (Thermo Fisher Scientific, #11668019). The culture medium was replenished on the following day and supernatant containing the viral particles was collected after 48 hours, filtered with 0.45 μm filters (Merck, #SE1M003M00), and concentrated using Lenti-X concentrator (Clontech, #631232), tittered using p24 assay and stored as single-use aliquots in -80 °C.

Strategy for optimizing the combination of TFs for generating iPECs

In the first pilot experiment, HFFs were transduced with 13 TFs along with GFP (MOI=1; condition 1; Figure S1B) to induce pancreatic exocrine cell fate. The observed morphological and gene expression changes (cf. Figures S1C–S1F) suggested that some combination(s) of these candidate TFs can induce pancreatic identity, but the progress of cell fate conversion was slow and the number of cells showing morphological changes was small. We reasoned that constitutive expression of too many developmental TFs, or a potential lack of an essential TF may hamper efficient conversion. Therefore, we set out to optimize the reprogramming conditions, specifically by analyzing 1) if any TFs are redundant in the 13-TF pool, 2) if any essential TFs are still missing, and 3) by designing different reprogramming conditions for generating acinar- and ductal-like cells. Of note, in the initial experiments with large pools of TFs, MOI=1 was used for individual TFs to avoid lentivirus-related toxicity to the transduced cells. In the subsequent experiments with TF pools comprising of fewer TFs, higher MOI values were tested to increase the expression levels of certain TFs (as detailed in Figure S1B). Based on flow cytometry analysis of GFP-positive HFFs, transduction with MOI=1 resulted in >85% and with MOI=10 around 95% infection efficiency.

To optimize the TF pool for reprogramming experiments, we first excluded the two TFs (XBP1, GATA6) that were already expressed in HFFs and whose expression was not markedly induced at the two-week timepoint in the 13-TF sample. Based on the RNA-seq data, four of the TFs had basal expression in HFFs [mean $\log_2$(normalized gene count) from three replicates for XBP1, GATA6, SOX9, and HEYL being 10.9239, 7.4991, 6.5852, and 4.7049, respectively]. Further analysis of their expression in the 13-TF sample with mean $\log_2$(normalized gene count) of 11.2108, 9.2251, 9.0446, 10.8986, respectively, showed only weak induction for XBP1 and GATA6, suggesting their redundancy in the 13-TF pool. Moreover, GATA6 has been shown to be functionally redundant with GATA4 during pancreas development,[26] further supporting its exclusion from the reprogramming pool. Next, we designed two sub-pools for generating acinar- and ductal-like cells separately based on previous information about human pancreas development. Specifically, FOXA2, GATA4, NKX6.1 PDX1, SOX9 and SOX17, which have been reported to be involved in the specification of pancreatic endoderm and progenitors,[21,22] are included in both conditions. The master regulator of pancreatic acinar differentiation PTF1A and its co-operator RBPJL are included in the pool for generating acinar-like cells,[24] as well as HEYL that was predicted to be mainly activated in human adult acinar cells.[87] HNF6 is restrained to the condition for ductal-like cells due to its predominant role in the development of pancreatic ducts[25] and much higher expression levels in pancreatic ducts compared to acinus based on RNA-seq data of human acinar and ductal cells.[19,20] Nr5a2 is required for exocrine pancreas formation during hepatopancreas progenitor specification,[88] thus it was included in both conditions. For the ductal cell pool, we also included HNF1B that was not part of the initial 13-TF pool, due to its previously reported role in ductal cells of postnatal and adult pancreas.[27] In total, 14 candidate TFs were included in the two reprogramming conditions, resulting in a 10-TF pool for acinar cells and a 9-TF pool for ductal cells (conditions 2 and 3, respectively; Figure S1B)

After transducing HFFs with the TF pools for acinar and ductal cells (conditions 2 and 3), we observed reduced expression of fibroblast-related genes and gradually increasing expression of acinar and ductal cell markers, respectively, from quantitative real-time-PCR and RNA-seq analyses (cf. Figures S1G–S1J). In condition 2 for acinar cells, HFFs gradually lost their normal morphology after induction but we did not observe clear epithelial-like morphology during reprogramming. In condition 3 for ductal cells, some epithelial-like cells appeared at 21 days after induction, but these cells did not expand after passaging, potentially due to the small cell number. To increase the conversion efficiency, we further refined the TF pools based on previous literature.

First, we excluded NKX6.1 and SOX9 from condition 2, resulting in an 8-TF pool for generating acinar-like cells (condition 5; Figure S1B). The exclusion was justified by the fact that both factors are expressed in pancreatic progenitor cells, but NKX6.1 is restricted to endocrine beta cells during pancreas maturation,[22] whereas the cells that retain SOX9 expression undergo ductal

fate specification.[29] Furthermore, the cross-antagonism between Nkx6.1 and Ptf1a in pancreatic multipotent progenitors balance the endocrine and acinar cell neogenesis during normal development.[28] However, in the reprogramming experiments using the condition 5 for acinar cells, we still did not observe clear epithelial morphology, and thus the experiments towards the acinar fate were not continued.

Second, to refine the TF pool for the ductal fate, we excluded GATA4, NR5A2 and NKX6.1 from condition 3 based on their previously reported roles in pancreatic development, resulting in a 6F pool (condition 7; Figure S1B). In mature pancreas, GATA4 is largely exclusive to acinar cells[26] and NKX6.1 to beta cells; NR5A2 is required for the development of exocrine pancreas but plays a more important role in acinar formation.[30] Reprogramming experiments identified that the 6F pool is able to generate pancreatic exocrine cells with phenotypic and functional characteristics of ductal epithelial cells (*cf.* Figures 1 and 2). Other sub-combinations of candidate TFs were also tested, and the cells were characterized by their morphology (Figure S2) and by bulk RNA-seq and/or FlsnMultiome-seq (*cf.* Figures S1B and S5A–S5D). Taken together, the combination of six TFs, FOXA2, SOX17, PDX1, HNF1B, HNF6 (ONECUT1) and SOX9, is necessary and sufficient for efficiently inducing pancreatic ductal cell fate from human fibroblasts.

Immunofluorescence staining

Cells were fixed with 100% ice-cold methanol for 5 min at -20 °C and washed three times with PBS. After fixation, cells were permeabilized with 0.2% Triton X-100 for 15 min at room temperature (RT) and then washed three times with PBS. After permeabilization, cells were blocked by 5% BSA in PBS for 1 h at RT and then incubated with primary antibodies at 4 °C overnight. For the secondary antibody staining, cells were incubated with appropriate fluorescence-conjugated secondary antibody for 1 h at RT in dark. Nuclei were stained with Hoechst (Thermo Fisher Scientific, #62249). Primary and secondary antibodies were diluted in PBS containing 3% BSA. The antibodies used were: FOXA2, 20 μg/ml (R&D, #AF2400); SOX17, 15 μg/ml (R&D, #AF1924); PDX1, 1:50 (Santa Cruz Biotechnology, #sc-390792); HNF1B, 1:200 (Sigma-Aldrich, #HPA002083); HNF6, 1:50 (Santa Cruz Biotechnology, #sc-376167); SOX9, 1:1000 (Millipore, #AB5535); Osteopontin (SPP1), 1:500 (Proteintech, #22952-1AP); donkey anti-goat (Alexa Fluor 488), 1:1000 (Abcam, #ab150129); goat anti-mouse (Alexa Fluor 488), 1:1000 (Thermo Fisher Scientific, #A-11029); goat anti-rabbit (Alexa Fluor 594), 1:1000 (Thermo Fisher Scientific, #A-11012); goat anti-rabbit (Alexa Fluor 488), 1:1000 (Thermo Fisher Scientific, #A-11034). The images were taken with Nikon Eclipse Ti-E (with Primo) inverted microscope and analyzed using Fiji (v1.53).

Flow cytometry analysis

Cells were harvested and centrifuged at 300g for 10 minutes. Cell pellet was resuspended in cell staining buffer containing 0.5% BSA and 2 mM EDTA. FcR blocking reagent (Miltenyi Biotec, #130-059-901) was used to reduce non-specific staining. Cells were then incubated with fluorescence-conjugated primary antibodies for 10 min in the dark at 4 °C. Cells were then washed twice in 1.5 mL of cell staining buffer and centrifuged at 300g for 10 minutes. Cell pellet was resuspended in 0.4 ml cell staining buffer for analysis. The antibodies used and their dilutions were: CD24-PE, 1:20 (BioLegend, #311105); CD133/2-APC, 1:50 (Miltenyi Biotec, #130-113-746). Cells were sorted using BD Accuri C6 flow cytometer and the data were analyzed using FlowJo (v10.6.1).

Enzyme activity analysis

Cells were harvested at different time points as indicated in the figures, washed with cold PBS and centrifuged for 3 minutes at 4 °C at 500g. Cell pellet was resuspended in 4X volumes of lysis buffer (150 mM NaCl, 10 mM Tris, 1% Triton X-100, pH = 7.2) and homogenized by vortex in cold room. Cell lysate was centrifuged for 2 minutes at 4 °C at 18,000g to remove insoluble material. Supernatants were then transferred to a clean tube and enzyme activities were analyzed using Carbonic Anhydrase Activity Assay Kit (Biovision, #K472), Amylase Activity Assay kit (Sigma-Aldrich, # MAK009) and Trypsin Activity Colorimetric Assay Kit (Sigma-Aldrich, #MAK290), following the instructions by the manufacturer. Optical density (OD) at 450 nm was read by FLUOstar Omega and the data were plotted using GraphPad Prism (v8.3.0).

RNA isolation, quantitative real-time-PCR and bulk RNA-sequencing (RNA-seq)

iPECs were harvested for RNA isolation at 48 h, 96 h, one week, two weeks, four weeks, six weeks and 10 weeks after induction along with control HFFs. Total RNA was isolated using RNeasy Plus Mini Kit (Qiagen, #74134). For quantitative real-time-PCR, cDNA synthesis from at least three biological replicates was performed using the PrimeScript™ RT Master Mix (TaKaRa, #RR036A) and real-time PCR was performed using SYBR Green I Master (Roche, #04707516001). The transcript levels of the target genes were normalized to GAPDH mRNA levels. Bulk RNA-seq libraries were prepared from 500 ng of total RNA for each sample using KAPA mRNA HyperPrep Kit for Illumina (Roche, #KR1352) following manufacturer's instruction. Final libraries were quantified using Qubit HS Assay kit (Thermo Fisher Scientific, #15850210) and Tapestation High Sensitivity D5000 Reagents (Agilent, #5067-5593) for concentration and size distribution, respectively, then paired-end sequenced on Illumina Novaseq 6000.

Chromatin immunoprecipitation-sequencing (ChIP-seq)

HFFs transduced only with GFP reporter construct and the iPECs at 48 h, 96 h and 1 week after 6F transduction were collected for ChIP-seq analyses. ChIP assays were performed as previously described.[89] Briefly, cells were fixed in 1% formaldehyde for 10 min at RT. Glycine was added to a final concentration of 0.125 M and incubated for 5 min to quench the reaction. Cells were then washed twice with ice-cold PBS and collected for lysis in RIPA buffer with protease inhibitors. Cross-linked chromatin was sonicated to an average fragment size of 200-500 bp, and immunoprecipitated with either H3K27ac antibody (Diagenode, #C15410196) or normal

rabbit IgG (Santa Cruz Biotechnology, #sc-2027), 2 μg of antibody per reaction. ChIP-seq libraries were prepared according to Illumina's instructions and single-end sequenced on Illumina Novaseq 6000.

Assay for Transposase-Accessible Chromatin-sequencing (ATAC-seq)

HFFs transduced only with GFP reporter construct and the iPECs at 48 h, 96 h and one week after 6F transduction were collected for analysis. ATAC-seq libraries were prepared from 75,000 cells as previously described.[90,91] Cells were washed with ice-cold PBS and resuspended in 50 μl of lysis buffer and incubated for 3 min on ice. The nuclei were isolated and transposed with Tn5 transposase in 2X tagmentation buffer (Illumina, #20034197) and incubated at 37 °C on thermomixer for 30 min at 1,000 rpm. The reaction was purified using MinElute PCR Purification Kit (Qiagen, #28004) and eluted in nuclease-free water. The samples were amplified for 8-11 total cycles and purified with AMPure beads (Agencourt, #A63881). Libraries were paired-end sequenced on Illumina Novaseq 6000.

Factor barcoding design for FI-snMultiome-seq assay for snRNA-seq and snATAC-seq

A 20-bp random oligo (N20) with NheI site was introduced into the lentiviral expression vector pLenti6/V5-DEST™ downstream of the ORF and 78 bp upstream of its 3' LTR region by PCR. The PCR generates a unique indexing barcode for every single molecule and this unique barcode will be transcribed together with the respective ORF cloned into the expression vector allowing us to uniquely index and barcode each factor. Since this indexing barcode is close to the poly(A) tail of the transcripts, a large fraction of the barcodes is captured during polyA enrichment and 3'RNA-seq library preparation protocol used in 10x multiome assay kit (10x Genomics). For the barcoding PCR, 10 ng of pLenti6/V5-DEST™ vector was used as template per reaction with Q5® High-Fidelity DNA Polymerase (NEB, #M0491S). The PCR program used was: 98 °C for 30 s; 30 cycles of (98 °C for 10 s; 72 °C for 5 min); 72 °C for 10 min. The full-length primers used for creating the barcoding pLenti6/V5-DEST™ construct were Barcode_Temp_NheI_F: CATGCTAGCNNNNNNNNNNNNNNNNNNNNCGAGCTCGGTACCTTTAAGACC and Barcode_Temp_NheI_R: CATGCTAGCTTGT GCTTAGCCCTCCCACAC. The PCR product was digested by NheI (Thermo Fisher Scientific, #FD0974) and DpnI (Thermo Fisher Scientific, #FD1703) to create sticky ends for the following ligation and to get rid of pLenti6/V5-DEST™ template, respectively. The linear barcoded pLenti6/V5-DEST™ vector was ligated using T4 DNA Ligase (Thermo Scientific, #EL0011) and transformed into ccdB Survival™ 2T1^{R} (Thermo Fisher Scientific, #A10460) competent cells followed by single-colony screening. The pLenti6/V5-DEST™ vectors with unique barcode were verified by Sanger sequencing. ORFs for individual TFs were cloned into the barcoded pLenti6/V5-DEST™ vectors using the Gateway recombination system as described earlier. The primers used for Sanger sequencing were: Sseq_F: GGAAATGAGAACAGGGGCATCTTG and Sseq_R: GCTGCAATAAACAAGTTCCTCTCAC. The barcodes for each TF are as shown in Table S4.

FI-snMultiome-seq assay (snRNA-seq and snATAC-seq) from reprogramming cells

HFFs transduced only with GFP reporter construct and the iPECs at 48 h, one week and two weeks after induction were collected for analysis. Nuclei isolation for snRNA-seq and snATAC-seq was optimized and performed following the demonstrated protocol CG000365 from 10x Genomics. HFFs transduced with barcoded lentiviral expression constructs for 6F reprogramming TFs, both as individual TFs and as a pool of six TFs, were harvested at different time points after induction. For each time point, equal number of nuclei from each condition were pooled for library preparation. The libraries were prepared by the FIMM single-cell core facility (University of Helsinki) following the 10x Genomics user guide CG000338 for Single Cell Multiome ATAC + Gene Expression Reagent Bundle (10xGenomics, #PN-1000285). An aliquot of the pre-amplified cDNA was used for preparing a custom library for detecting the TF barcodes. This strategy enables correlating each TF barcode to the cell barcodes introduced by the 10x multiome workflow. Custom barcode libraries were prepared using pre-amplified cDNA from Step 4 in protocol CG000338 by a two-step PCR process. The first PCR step was to amplify TF barcodes from cDNA and to add Trueseq_Read 2 sequence to 5' end of cDNA. For this, 3 μl of cDNA from each sample was used as template per reaction with KAPA HiFi HotStart ReadyMix (Kapa Biosystems, #KK2602) with primers Truseq_Read2_Vec_Amp_F: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGGGCTAAGCACAAGCTAG*C and Truseq_Read1_Amp_R: ACACTCTTTCCCTACACGACGCTCTTCCGATC*T. The PCR program used was 95 °C for 3 min; 22 cycles of (98 °C for 20 s; 65 °C for 15 s; and 72 °C for 10 s). Reactions were purified with AMPure XP beads at a 0.8x ratio and quantified by Qubit HS Assay kit. The second PCR was performed from 65 ng of the purified product per reaction from the first PCR. Index primers from Dual Index Kit TT Set A (10x Genomics, #PN-1000215) were used to attach Illumina adapters and indices to the samples. The PCR program used was: 95 °C for 3 min; 4 cycles of (98 °C for 20 s; 65 °C for 15 s; 72 °C for 10 s). The reactions were purified with AMPure beads at a 0.8x ratio. The purified custom libraries were quantified and pooled together with gene expression (GEX) libraries for sequencing. The required PCR cycles for each step were tested to generate libraries that are minimally PCR amplified. The sequencing run parameters for the custom barcode libraries and GEX libraries were R1:28, i7-index: 10, i5-index: 10, R2:90 and for ATAC libraries R1:50, i7-index:8, i5-index:24, R2:49 on Illumina Novaseq 6000.

RNA-seq data processing

The sample quality of the RNA-seq datasets were evaluated by FastQC analysis (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequence reads were aligned to the human reference genome (UCSC GRCh38/hg38) using STAR aligner (v2.7.5c) with default parameters.[66] Samtools (v1.9)[67] was used to sort the bam files. Gene counts were quantified using HTSeq-count (v0.11.2).[68] Genes with low counts having an expression of less than 10 across all samples were filtered out and the remaining genes were normalized using the DESeq2 (v1.30.1) pipeline.[69] To identify the differentially expressed genes (DEGs) between two

conditions, a threshold of |log2FoldChange| > 1.5 and adjusted p-value < 0.05 was applied. Pathway enrichment analysis was conducted with a preranked gene list of all DEGs based on the sign of fold change and p-value using Gene Set Enrichment Analysis (GSEA) Preranked tool (v4.2.3).[70] PCA and volcano plots were generated using ggplot2 (v3.3.5) in R (v4.0.5). Heatmaps with hierarchical clustering using average linkage and euclidean as distance metric was created with pheatmap (v1.0.12). We overlapped the DEGs from all the time points in 6F pool to visualize the transient change in their expression pattern. The integrated average expression of the biological replicates, scaled and centered by row were plotted in the heatmap. Genes from the cluster were used to perform enrichment analysis for Gene Ontology terms by Metascape web-based platform.[71] To fetch the genes involved in Shh signaling pathway, we combined the gene sets of KEGG, Reactome, Pathway Interaction Database and WikiPathways from Molecular Signatures Database (MSigDB) and created a master set of 203 genes. We then analyzed the expression of these genes from our RNA-seq data at two, four and six weeks of reprogramming to generate a line plot. To calculate the pairwise correlation coefficient between the ductal, acinar, iPECs and HFF based on the normalized gene counts, the Pearson method was used with 'cor' function in R.

ATAC-seq and ChIP-seq data processing

ATAC-seq and ChIP-seq data processing was performed as previously described.[92] First, paired-end reads from all ATAC-seq timepoints were down-sampled to the similar sequencing read depth (36 and 32 million reads per sample for replicate 1 and 2, respectively). The quality metrics of the ATAC-seq and single-end reads from ChIP-seq fastq files were checked with FastQC. The reads were mapped to the human reference genome (UCSC GRCh38/hg38) by Bowtie2 (v2.2.5) (UCSC GRCh38/hg38) with option '–very-sensitive'.[72] The ATAC-seq reads that mapped to mitochondrial DNA were discarded. Duplicates were marked using Picard MarkDuplicates (http://broadinstitute.github.io/picard/) and removed using samtools (v1.9) view with option '-F 1024'. ATAC-seq and ChIP-seq reads were subsequently filtered for alignment quality of less than 10 and 20 respectively. The peaks were called using MACS2 (2.1.1) with the parameter '–keep-dup all' and the blacklisted regions downloaded from ENCODE were removed.[76] To generate the bigwig signal tracks from the aligned bam files, deeptools (3.1.3) bamCoverage was used with options '—normalizeUsing RPKM –binSize 10'.[77] Along with the biological replicates, we also generated i) pooled replicates by combining the reads of biological replicates and ii) pseudoreplicates which was made by randomly shuffling the pooled replicates and equal splitting. A similar peak calling strategy was used for these two methods as well and an overlap of peaks from all the methods was considered as the high confidence peak set for each time point. The peaks were further filtered using q value < 0.0001. The signal tracks from the pooled replicates were used for visualization in Integrative Genomics Viewer (IGV) genome browser (v2.5).[78]

ATAC-seq and ChIP-seq data clustering and motif searching

In order to observe the changes in chromatin accessibility across the time points, we employed CoBRA (v2.0) to pooled replicates of time-series data from ATAC-seq.[79] Briefly, the pipeline generates a master set of peaks from all samples and calculates the read density across those regions. Quantile normalization was used to normalize the read count matrix. The top 50% peaks were considered for the downstream unsupervised analysis and the rest were filtered out based on low reads per kilobase per million mapped reads (RPKM) values across multiple samples. k-means (k = 6) clustering was performed in the resulting peak set to highlight the substantial disparities in open chromatin profiles. Each cluster was paired with timepoint-specific line plots with a mean trend line using ggplot2 to visualize cluster trends. Peaks from each cluster were used to search for potential known and *de novo* motifs by using HOMER (v4.10.4) with the option '-size 100'.[80] A similar approach was used for the H3K27ac ChIP-seq data to generate the master peak set. BEDTools (v2.30.0) intersect was implemented to investigate the overlap between H3K27ac peak set and ATAC-seq peaks from each of the clusters.[81] The normalized read counts of the overlapped peak set were calculated using CoBRA and plotted in R with k-means 6 clustering. Transcription factor Occupancy prediction By Investigation of ATAC-seq Signal (TOBIAS v0.13.3) was used to perform TF footprinting analysis in open chromatin regions.[46] stat, ATACorrect was used to perform bias correction of the reads in open chromatin by shifting +4 bp and −5 bp on positive and negative strands, respectively. Then we used ScoreBigwig to calculate footprint scores from the corrected cutsites with default parameters. Finally, we used BINDetect to find transcription factor motifs that were differentially footprinted at each time point using motifs from 'The Human Transcription Factors' (http://humantfs.ccbr.utoronto.ca/) database.[75] The aggregate footprint across the transcription factor binding sites of the differentially footprinted motifs were plotted using TOBIAS PlotAggregate. In addition, we looked for the nearby genes (±10kb) of the bound sites of the representative genes from our TOBIAS footprinting analysis by Binding and Expression Target Analysis (BETA).[73] We searched for all motifs from the JASPAR[74] database around the summit (±100bp) of the ATAC peaks of representative genes using Find Individual Motif Occurences (FIMO) from the MEME Suite.

FI-snMultiome-seq (snRNA-seq and snATAC-seq) data processing

Raw sequencing data were processed using the Cell Ranger ARC pipeline (v2.0.1) with the GRCh38 reference (refdata-cellranger-arc-GRCh38-2020-A-2.0.0) for demultiplexing, alignment, barcode and feature counting to generate both ATAC and GEX feature-barcode matrices, which were loaded into Seurat (v4.1.1)[82] and Signac (v1.7.0)[83] for further analyses. We filtered out cells with less than 1,000 RNA counts, 1,000 ATAC fragments or more than 100,000 RNA counts, 500,000 ATAC fragments, 30% mitochondrial-derived RNA counts. For scRNA-seq, the raw counts were normalized, scaled with the percentage of mitochondrial-derived counts and cell cycle scores regressed out using SCTransform. The top 3,000 highly variable features were used for principal component analysis (PCA), and the top 35 PCs were used for UMAP. For scATAC-seq, we called the peaks using MACS2 (v2.2.7.1).[76] The

peaks on nonstandard chromosomes and in GRCh38 genomic blacklist regions were removed. We performed the frequency inverse document frequency (TF-IDF) normalization using RunTFIDF, identified the top features using FindTopFeatures with min.cutoff = 'q5', performed latent semantic indexing (LSI) reduction using RunSVD, and the 2-35 dimensions used for calculating UMAP. We then constructed the WNN graph using FindMultiModalNeighbors with the 1-35 PCs from the scRNA-seq data and the 2:35 LSI dimensions from the scATAC-seq data for the joint UMAP visualization. Cells from different timepoints were analyzed in separate 10x Chromium runs but control HFFs transduced with GFP were pooled together with the 48h timepoint sample. During the analysis, the cells were negative to all TF barcodes were assigned as a control.

snATAC-seq TF motif analysis

The same human motif position frequency matrices[75] were used for snATAC-seq motif analysis that were used in the bulk analysis. Per-cell motif activity scores were computed using chromVAR (v1.18.0)[84] and the UCSC hg38 genome (BSgenome.Hsapiens.UCSC.hg38) with the Signac RunChromVAR wrapper.

Cell type score

The cell type scores were computed using ScType (v1.0)[48] with cell-type-specific markers. The markers for pancreatic cell types were obtained from the ScType in-build database and the fibroblast markers were collected from literature.[93] All the markers used are listed in Table S2.

Probabilistic cell type assignment

Cell type assignment were computed for HFFs and reprogrammed cells expressing 6F_pool using the reimplementation of the CellAssign model[94] in the scvi-tools library.[85] We randomly downsampled 20 markers for fibroblast and major pancreatic cell type including acinar, alpha, beta, delta, ductal and gamma cells from the ScType in-build database, respectively. These marker genes are shown in Table S5. To quantify the reprogrammed cells at transient state and expressing marker genes for any two cell types, we added those cells with mixed identities as separate ''cell types'' to the input marker gene matrix. Default parameters were used and CellAssign returned the assignment probabilities to each cell type for each cell. Cells were labelled by the cell type with the maximum assignment probability.

Trajectory and pseudotime analysis

The trajectory graph for control cells and the cells with six TFs was constructed using Monocle3 (v1.0.0).[86] The expression data and the joint UMAP projection was loaded into monocle3 using the as.cell_data_set function from SeuratWrappers (v0.3.0). The cells were clustered using the leiden method with the number of nearest neighbors set to 50, and the trajectory graph was learned using the learn_graph function with default parameters. Pseudotime was assigned using the order_cells function with setting the control cells as the starting points, and graph_test was used to identify genes and motifs that vary over pseudotime.

QUANTIFICATION AND STATISTICAL ANALYSIS

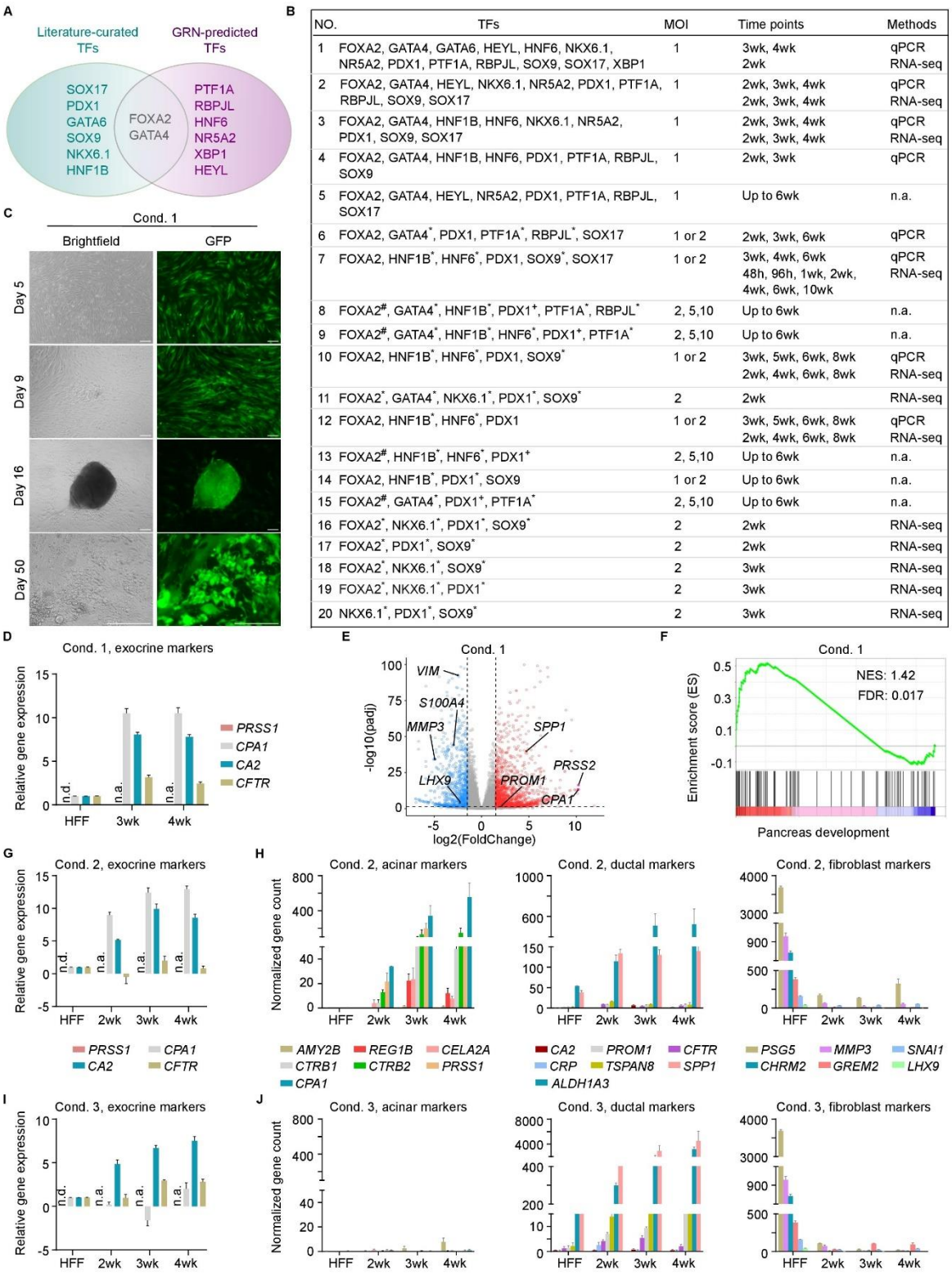
Statistical analyses were performed using built-in tools in GraphPad8 or R. Two-tailed Student's t test or Wilcoxon test was used for statistical comparisons between the groups. Statistical parameters and the test used for each experiment are indicated in the figure legends.

Developmental Cell, Volume *58*

Supplemental information

Single-cell epigenome analysis identifies molecular events controlling direct conversion of human fibroblasts to pancreatic ductal-like cells

Liangru Fei, Kaiyang Zhang, Nikita Poddar, Sampsa Hautaniemi, and Biswajyoti Sahu



NO.	TFs	MOI	Time points	Methods
1	FOXA2, GATA4, GATA6, HEYL, HNF6, NKX6.1, NR5A2, PDX1, PTF1A, RBPJL, SOX9, SOX17, XBP1	1	3wk, 4wk 2wk	qPCR RNA-seq
2	FOXA2, GATA4, HEYL, NKX6.1, NR5A2, PDX1, PTF1A, RBPJL, SOX9, SOX17	1	2wk, 3wk, 4wk 2wk, 3wk, 4wk	qPCR RNA-seq
3	FOXA2, GATA4, HNF1B, HNF6, NKX6.1, NR5A2, PDX1, SOX9, SOX17	1	2wk, 3wk, 4wk 2wk, 3wk, 4wk	qPCR RNA-seq
4	FOXA2, GATA4, HNF1B, HNF6, PDX1, PTF1A, RBPJL, SOX9	1	2wk, 3wk	qPCR
5	FOXA2, GATA4, HEYL, NR5A2, PDX1, PTF1A, RBPJL, SOX17	1	Up to 6wk	n.a.
6	FOXA2, GATA4*, PDX1, PTF1A*, RBPJL*, SOX17	1 or 2	2wk, 3wk, 6wk	qPCR
7	FOXA2, HNF1B*, HNF6*, PDX1, SOX9*, SOX17	1 or 2	3wk, 4wk, 6wk 48h, 96h, 1wk, 2wk, 4wk, 6wk, 10wk	qPCR RNA-seq
8	FOXA2#, GATA4*, HNF1B*, PDX1+, PTF1A*, RBPJL*	2, 5,10	Up to 6wk	n.a.
9	FOXA2#, GATA4*, HNF1B*, HNF6*, PDX1+, PTF1A*	2, 5,10	Up to 6wk	n.a.
10	FOXA2, HNF1B*, HNF6*, PDX1, SOX9*	1 or 2	3wk, 5wk, 6wk, 8wk 2wk, 4wk, 6wk, 8wk	qPCR RNA-seq
11	FOXA2*, GATA4*, NKX6.1*, PDX1*, SOX9*	2	2wk	RNA-seq
12	FOXA2, HNF1B*, HNF6*, PDX1	1 or 2	3wk, 5wk, 6wk, 8wk 2wk, 4wk, 6wk, 8wk	qPCR RNA-seq
13	FOXA2#, HNF1B*, HNF6*, PDX1+	2, 5,10	Up to 6wk	n.a.
14	FOXA2, HNF1B*, PDX1*, SOX9	1 or 2	Up to 6wk	n.a.
15	FOXA2#, GATA4*, PDX1+, PTF1A*	2, 5,10	Up to 6wk	n.a.
16	FOXA2*, NKX6.1*, PDX1*, SOX9*	2	2wk	RNA-seq
17	FOXA2*, PDX1*, SOX9*	2	2wk	RNA-seq
18	FOXA2*, NKX6.1*, SOX9*	2	3wk	RNA-seq
19	FOXA2*, NKX6.1*, PDX1*	2	3wk	RNA-seq
20	NKX6.1*, PDX1*, SOX9*	2	3wk	RNA-seq

Figure S1. Identification of TFs for generating pancreatic exocrine cells, related to Figure 1

(A) A pool of 14 candidate TFs for the induction of exocrine pancreas cells from human fibroblasts tested in reprogramming experiments. Eight TFs were selected from curation of developmental studies of human pancreas [S1, S2] and another eight were predicted by gene regulatory network (GRN) algorithm Mogrify [S3].

(B) A list of all TF combinations that were tested for generating iPECs from human fibroblasts. Transduced cells were monitored for morphological changes, and the conditions that showed altered morphology were characterized for the expression of pancreatic maker genes using qRT-PCR and/or RNA-seq. MOI, multiplicity of infection. Each TF was transduced with MOI=1 unless specified with the following markings: * for MOI=2, $^{+}$ for MOI=5, and $^{\#}$ for MOI=10.

(C) Morphology of cells transduced with a 13-TF pool (condition 1) at indicated timepoints after induction. Fluorescent signal was detected from GFP that was transduced to the cells together with the 13-TF pool.

(D) The expression of pancreatic exocrine marker genes induced by the 13-TF pool (condition 1) measured by qRT-PCR.

(E) Volcano plot for DEGs from RNA-seq showing down-regulation of fibroblast-specific genes (left) and upregulation of exocrine pancreas-specific genes in iPECs (right) at 14 days after induction with condition 1. Significant DEGs ($|\log2FC| \geq 1.5$, p-value < 0.05, n=3) were colored on both sides (down-regulated significant DEGs in blue; up-regulated significant DEGs in red).

(F) GSEA from RNA-seq data for control fibroblasts and cells transduced with 13-TFs (condition 1) at 14 days after induction showing the enrichment of the genes involved in pancreas development.

(G) The expression of pancreatic exocrine marker genes induced by the 10-TF pool (condition 2) measured by qRT-PCR.

(H) Normalized gene counts of acinar, ductal marker genes and fibroblast-related genes during reprogramming using condition 2 measured by RNA-seq.

(I) The expression of pancreatic exocrine marker genes induced by the 9-TF pool (condition 3) measured by qRT-PCR.

(J) Normalized gene counts of acinar, ductal marker genes and fibroblast-related genes during reprogramming using condition 3 measured by RNA-seq.

Data in (D), (G) and (I) were normalized to the expression levels in HFFs, and were represented as the mean ± SEM, n=8, including four biological replicates and two technical replicates for each biological replicate. Data in (H) and (J) were represented as the mean ± SEM, n=3 technical replicates. Cond, condition. n.d., not detectable. n.a., not applicable.

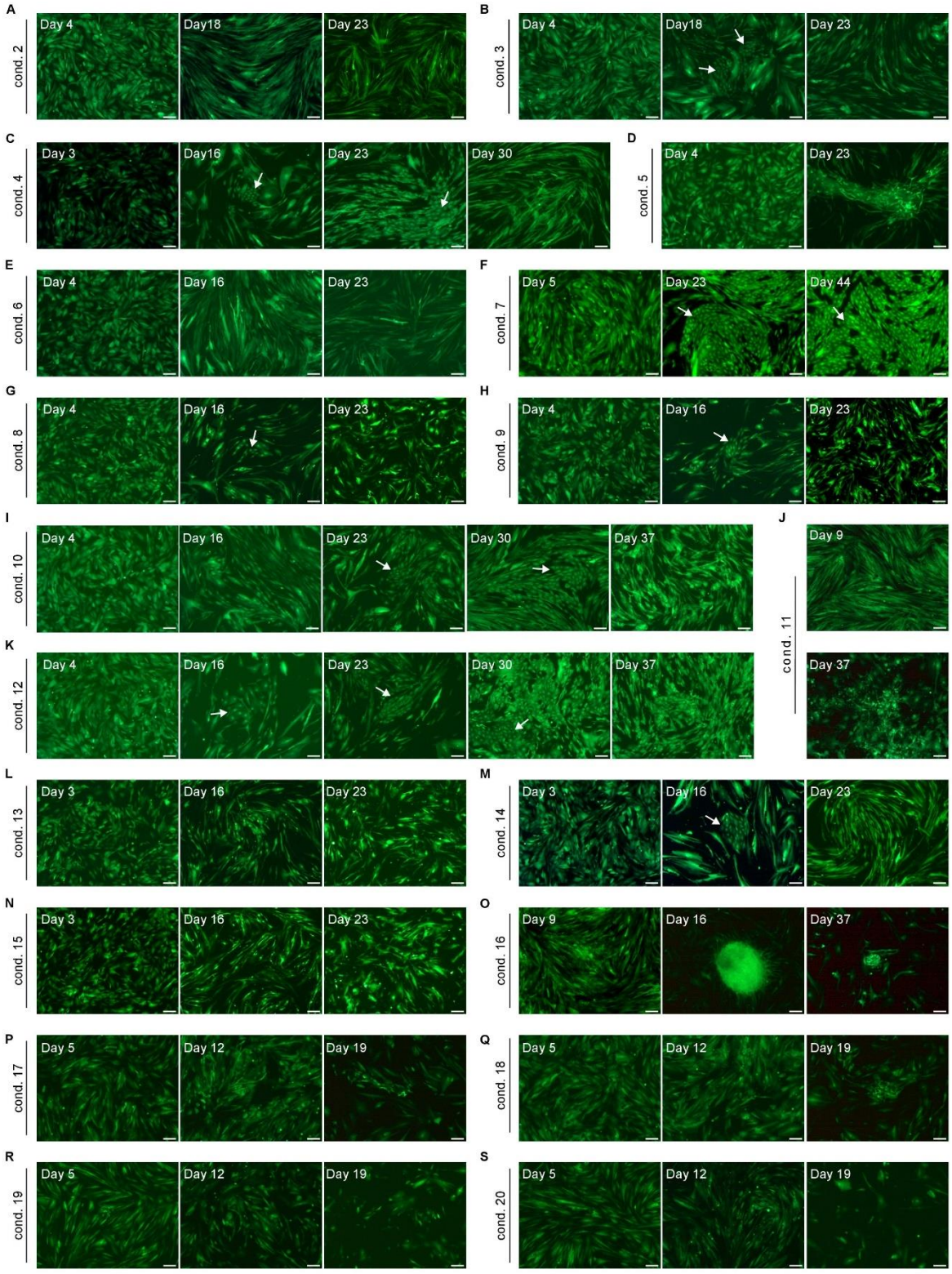


Figure S2. Morphological changes of reprogrammed cells during cell fate conversion in all tested conditions, related to Figure 1

(A-S) Morphology of cells transduced with different combinations of TFs at indicated timepoints after induction. Fluorescent signal was detected from lentiviral GFP construct that was transduced to the cells together with reprogramming TFs. TF combinations used in each condition (cond.) are listed in **Figure S1B**.

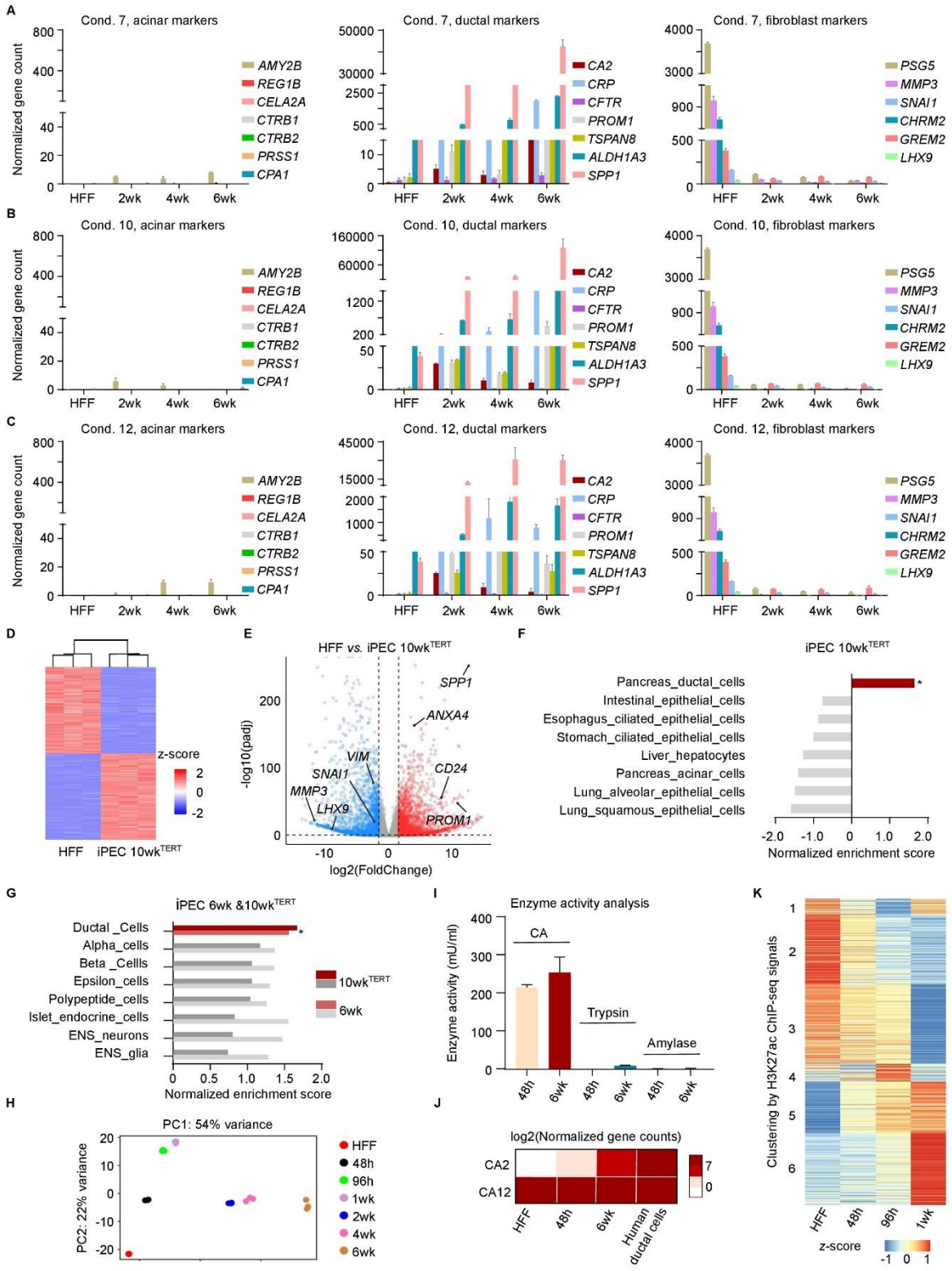


Figure S3. Epigenome and transcriptome reprogramming during cell fate conversion, related to Figures 1-3

(A-C) Normalized gene counts of acinar and ductal marker genes and fibroblast-related genes during reprogramming using conditions 7, 10 and 12 measured by RNA-seq. Data were represented as the mean ± SEM, n=3 technical replicates.

(D) Hierarchical clustering of differentially expressed genes (DEG) in HFFs vs. iPECs at $10wk^{TERT}$ ($|log2FC| \geq 1.5$, *p*-value < 0.05, n=3); *z*-score of normalized expression values is indicated using a color scale.

(E) Volcano plot for DEGs from RNA-seq data showing down-regulation of fibroblast-specific genes (left) and upregulation of exocrine pancreas-specific genes in iPECs (right) at $10wk^{TERT}$ after induction. Significant DEGs ($|log2FC| \geq 1.5$, p-value < 0.05, n=3) were colored on both sides (down-regulated significant DEGs in blue; up-regulated significant DEGs in red).

(F) Normalized enrichment score from GSEA for gene signatures of all major cell types of endodermal origin among DEGs in iPEC vs. HFF showing significantly enriched fetal pancreatic ductal cell signatures in iPECs at $10wk^{TERT}$ after induction.

(G) Normalized enrichment scores from GSEA for signatures of different pancreatic cell types within upregulated genes in iPECs at 6wk and $10wk^{TERT}$. (*, *p*-value < 0.01 and FDR < 5%).

(H) Principal component analysis (PCA) of normalized RNA-seq read counts for HFFs and the reprogrammed cells at indicated timepoints.

(I) Measurement of enzyme activities for CA, trypsin and amylase in the lysates of iPECs at 48 h and six weeks after induction. Data are represented as the mean ± SEM, n = 6, including three biological replicates and two technical replicates for each biological replicate.

(J) Heatmap showing the gene expression levels of *CA2* and *CA12* in reprogrammed cells at 48 h and six weeks after induction. HFFs and human ductal cells [S4] were used as control.

(K) Heatmap clustering of H3K27ac ChIP-seq signal at the sites corresponding to the six ATAC-seq clusters in **Figure 3A** showing the temporal dynamic epigenetic landscape.

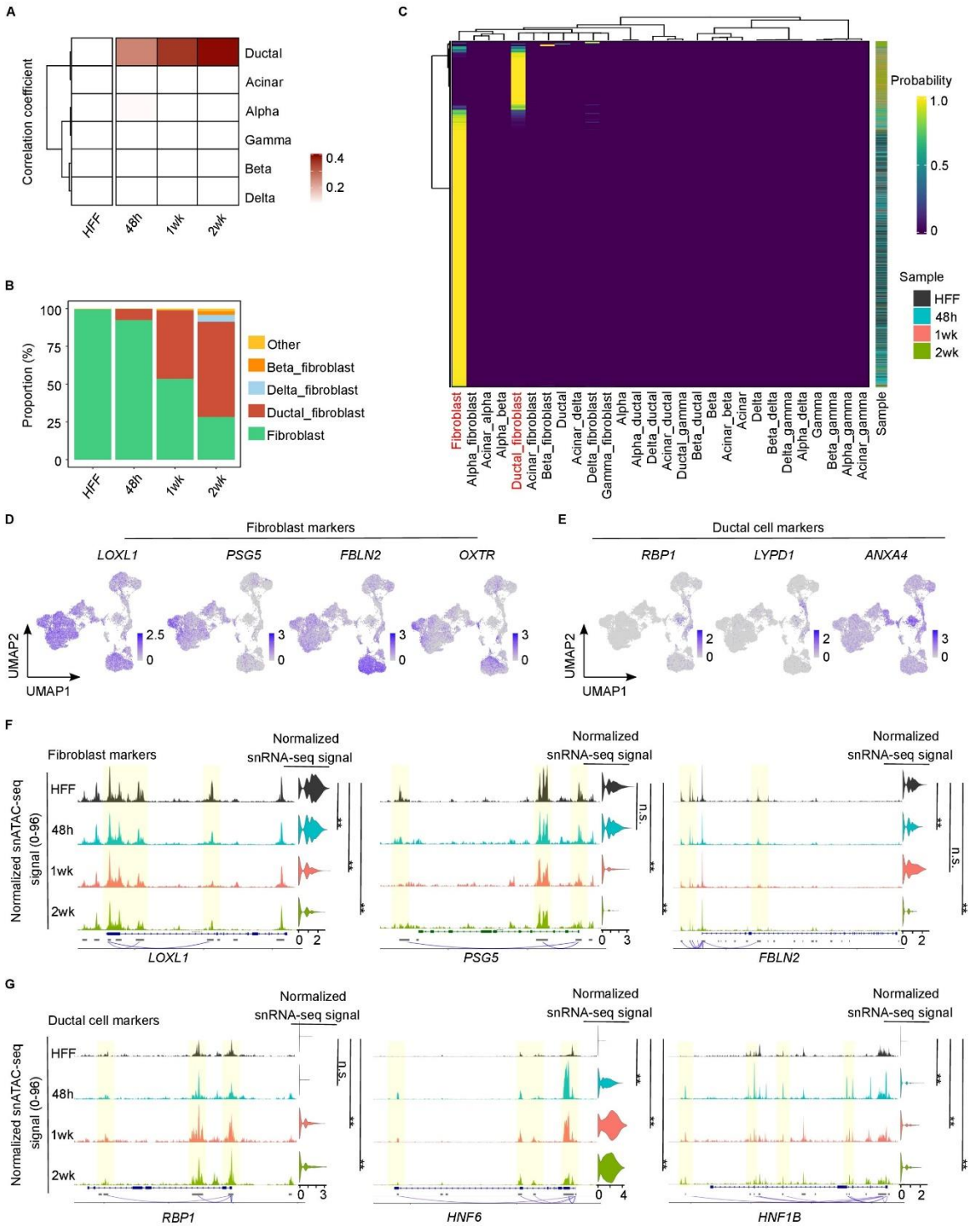


Figure S4. Analysis of gene signatures for major pancreatic cell types and fibroblasts in reprogrammed cells from FI-snMultiome-seq data, related to Figure 5

(A) Heatmap showing the Pearson correlation coefficients between HFF, reprogrammed cells expressing 6F pool and human pancreatic cell types [S5] based on the cell type marker genes from the ScType database.

(B) Proportions of cells expressing different hybrid gene signatures analyzed from HFF cells and from cells expressing 6F pool at each timepoint. Cell types with proportions < 1% were shown as “others”.
(C) Heatmap showing probability of assignment for each cell (row) to each cell type (column).
(D) and (E) UMAPs showing the expression of representative marker genes of ductal cells and fibroblasts.
(F) and (G) Coverage plots showing chromatin accessibility at the ductal cell and fibroblast marker gene loci in HFFs and in the cells transduced with the 6F pool, as well as the mRNA expression levels of the respective genes. Two-sided Wilcoxon rank-sum test was used in (F) and (G). FDR from Benjamini–Hochberg (BH) adjusted *P*-values; n.s., not significant; *, FDR < 0.01. **, FDR < 0.001.

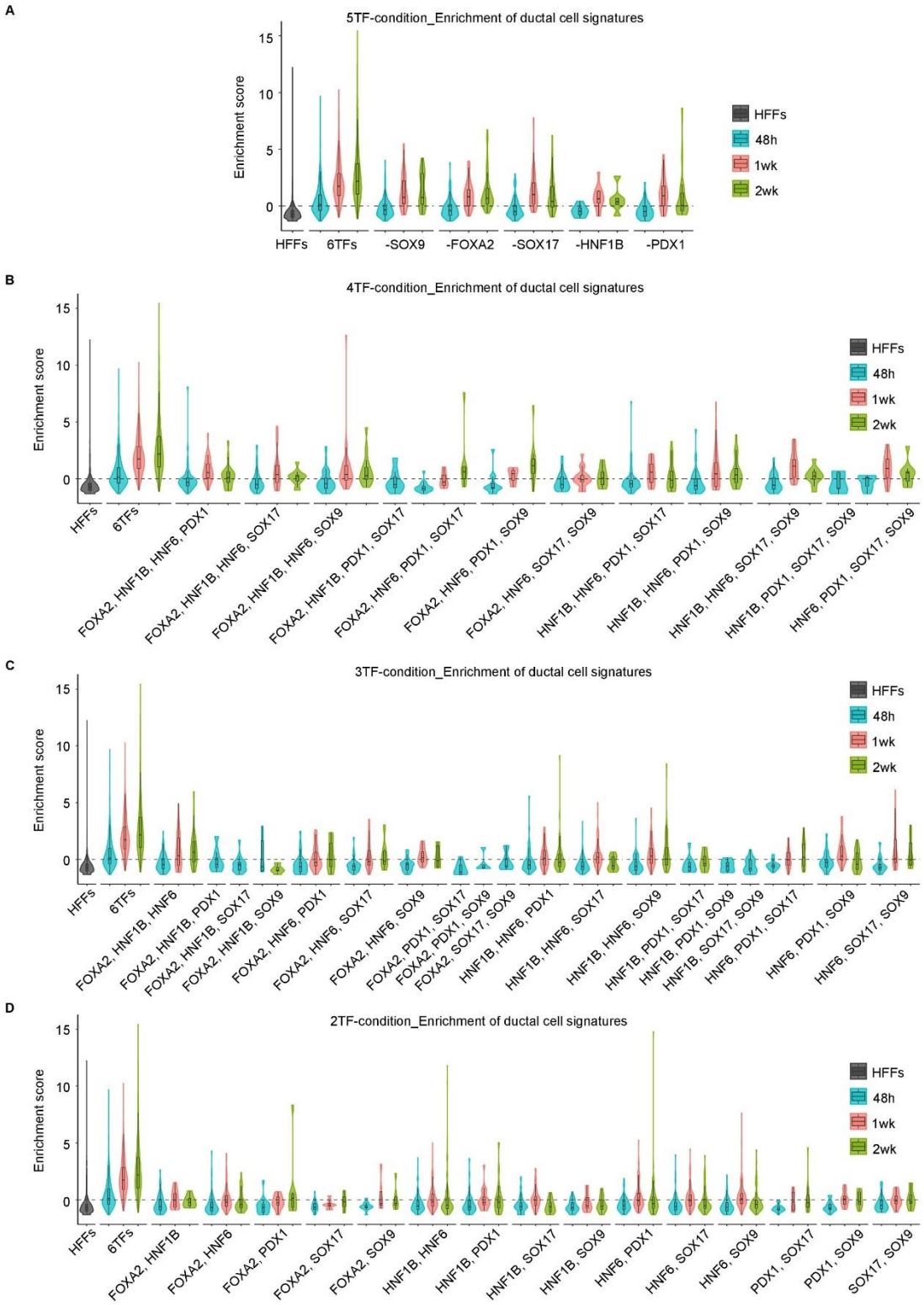


Figure S5. Dissection of minimal combination of TFs required for generating pancreatic ductal cells from FI-snMultiome-seq data, related to Figure 5

(A-D) Violin plots showing the enrichment score of ductal cell signatures in the reprogrammed cells expressing five TFs (A), four TFs (B), three TFs (C) and two TFs (D) of the 6F pool at indicated timepoints. The conditions with more than five cells were plotted. Data are represented as the median, top and bottom quartiles.

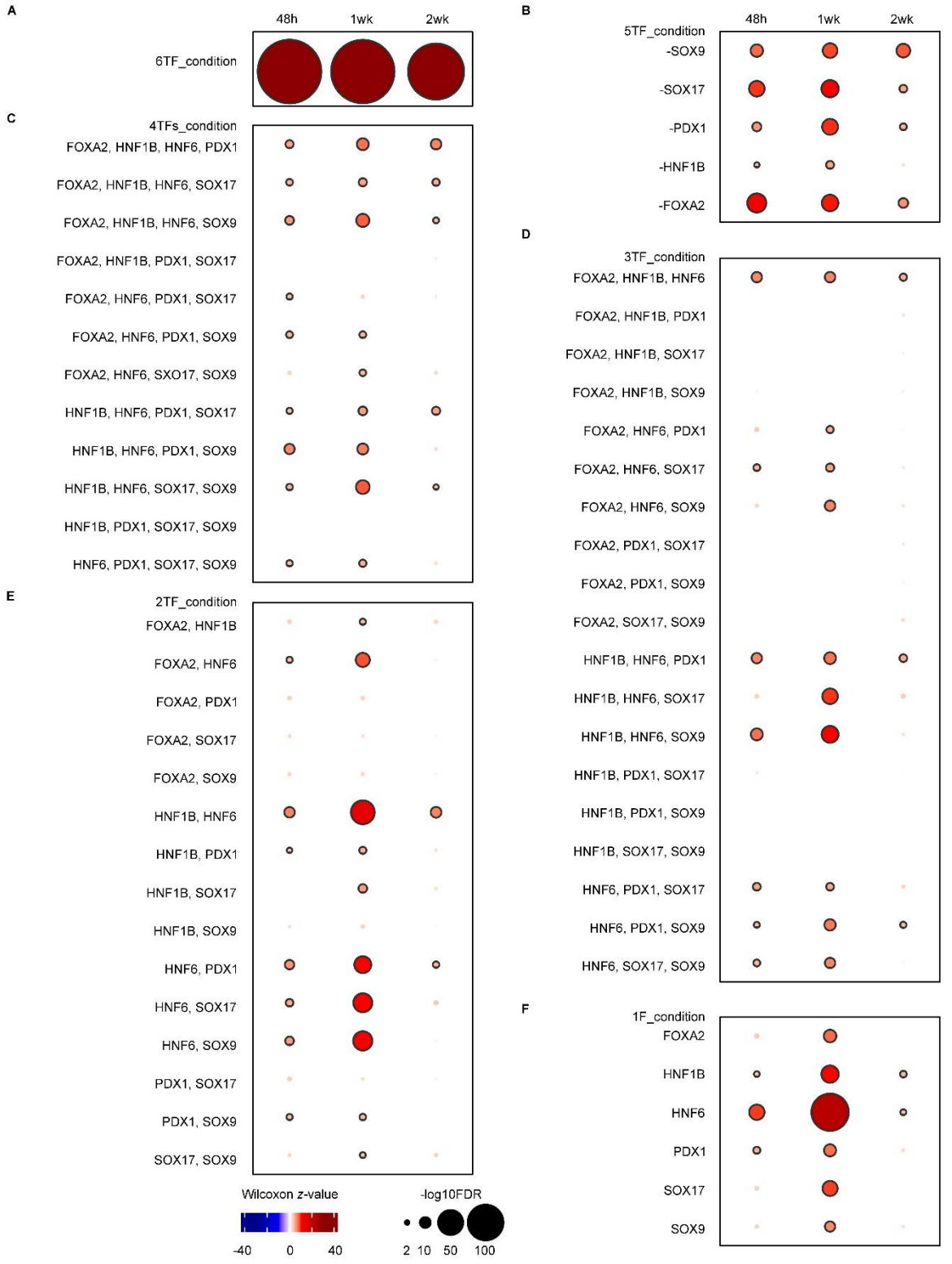


Figure S6. Comparison of each combination of TFs in generating pancreatic ductal cells from FI-snMultiome-seq data, related to Figure 5

(A-F) Dot plots showing the enrichment of ductal cell signatures in the reprogrammed cells expressing all six reprogramming TFs (A), and different combinations of five (B), four (C), three (D), two (E) and

one TFs (F) compared to HFFs. The conditions for which more than five cells were recovered from the FI-snMultiome-seq dataset are plotted. The color gradient indicates the Wilcoxon *z*-value and the size indicate Benjamini–Hochberg (BH) adjusted -log10 (p-value).

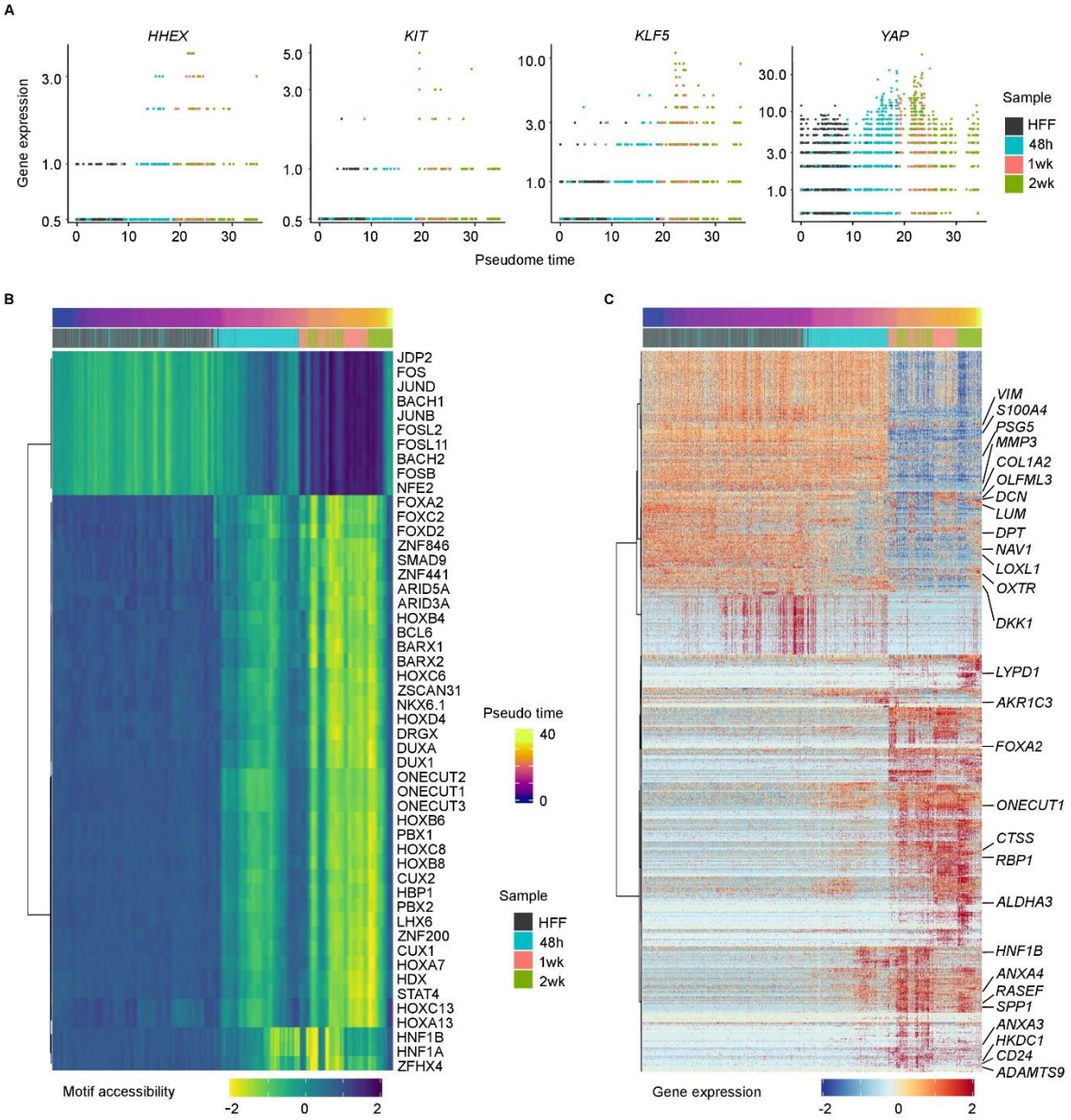


Figure S7. Pseudotime analysis for TF motif accessibility and marker gene expression during transdifferentiation, related to Figure 6

(A) Gene expression dynamics of example endodermal markers along pseudotime.

(B) Unbiased analysis of motif accessibility for top 50 TFs that show differential motif accessible along pseudotime (Moran's I > 0.15, FDR < 0.01).

(C) Unbiased analysis of expression levels of top1000 genes that are differentially expressed along pseudotime (Moran's I > 0.15, FDR < 0.01) showing transcriptional shift from fibroblast-related genes towards pancreatic identity in cells ordered along the pseudotime.

Reference

S1 Jennings, R. E. *et al.* (2015). Human pancreas development. Development *142*, 3126-3137. 10.1242/dev.120063.

S2 Petersen, M. B. K., Goncalves, C. A. C., Kim, Y. H. & Grapin-Botton, A. (2018). Recapitulating and Deciphering Human Pancreas Development From Human Pluripotent Stem Cells in a Dish. Curr Top Dev Biol *129*, 143-190. 10.1016/bs.ctdb.2018.02.009.

S3 Rackham, O. J. L. *et al.* (2016). A predictive computational framework for direct reprogramming between human cell types. Nat Genet *48*, 331-335. 10.1038/ng.3487.

S4 Ayars, M., E. O'Sullivan, A. Macgregor-Das, K. Shindo, H. Kim, M. Borges, J. Yu, R.H. Hruban, and M. Goggins (2017). IL2RG, identified as overexpressed by RNA-seq profiling of pancreatic intraepithelial neoplasia, mediates pancreatic cancer growth. Oncotarget 8, 83370-83383.10.18632/oncotarget.19848.

S5 Segerstolpe, A., A. Palasantza, P. Eliasson, E.M. Andersson, A.C. Andreasson, X. Sun, S. Picelli, A. Sabirsh, M. Clausen, M.K. Bjursell, et al. (2016). Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. Cell Metab 24, 593-607. 10.1016/j.cmet.2016.08.020.