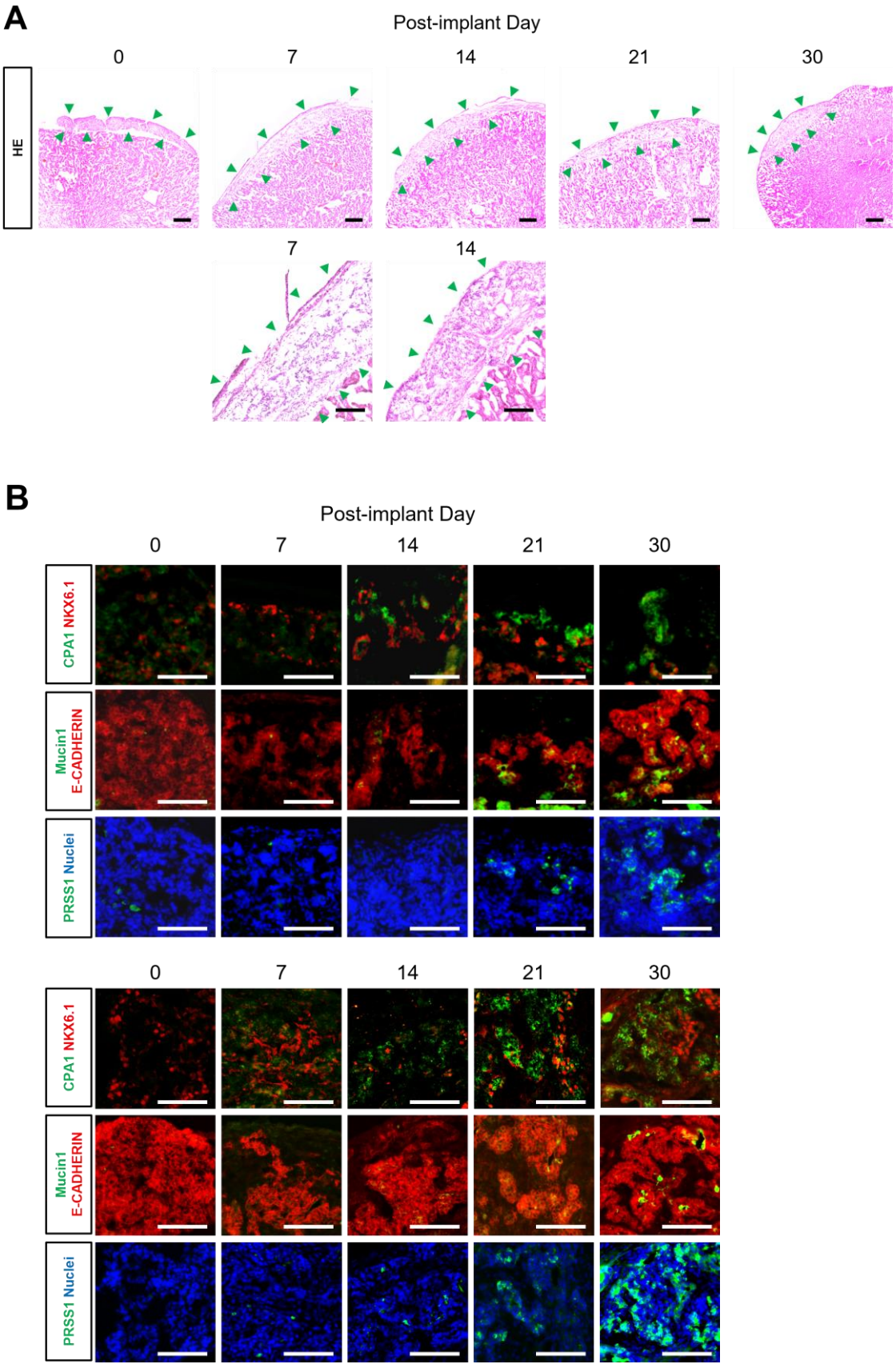
Supplementary Information

Mechanistic elucidation of human pancreatic acinar development using single-cell transcriptome analysis on a human iPSC differentiation model

Atsushi Mima[1,2], Azuma Kimura[1,3], Ryo Ito[1], Yu Hatano[1], Hiraku Tsujimoto[1,3], Shin-Ichi Mae[1], Junko Yamane[1], Wataru Fujibuchi[1], Norimitsu Uza[2], Taro Toyoda[1], Hiroshi Seno[2] & Kenji Osafune[1]

1. Center for iPS Cell Research and Application (CiRA), Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan.
2. Department of Gastroenterology and Hepatology, Kyoto University, 54 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan.
3. Rege Nephro Co., Ltd., Med-Pharm Collaboration Building, Kyoto University, 46-29 Yoshidashimoadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan.

*Correspondence should be addressed to T.T. (e-mail address: t.toyoda@cira.kyoto-u.ac.jp, fax number: +81-75-366-7023) and K.O. (e-mail address: osafu@cira.kyoto-u.ac.jp, fax number: +81-75-366-7077).


A
Post-implant Day
0
7
14
21
30
HE
7
14
B
Post-implant Day
0
7
14
21
30
CPA1 NKX6.1
Mucin1 E-CADHERIN
PRSS1 Nuclei
0
7
14
21
30
CPA1 NKX6.1
Mucin1 E-CADHERIN
PRSS1 Nuclei

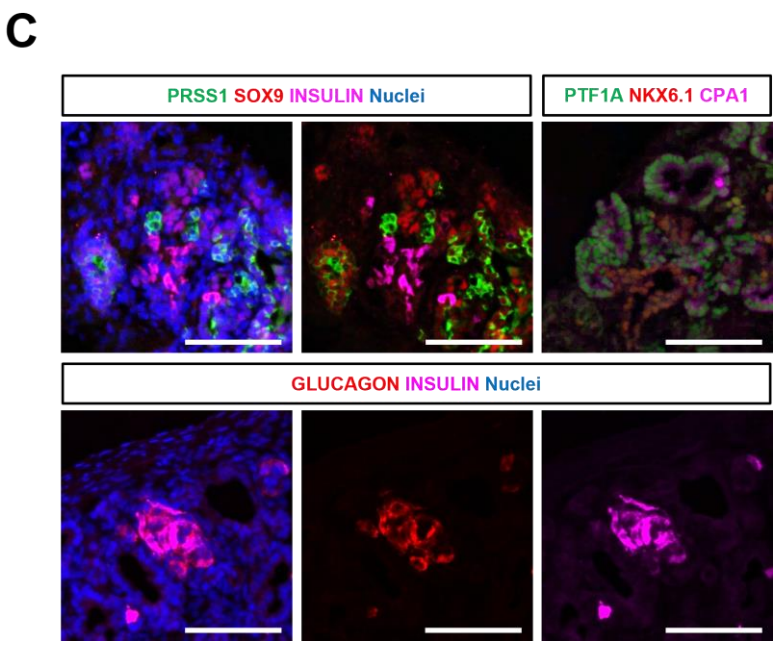
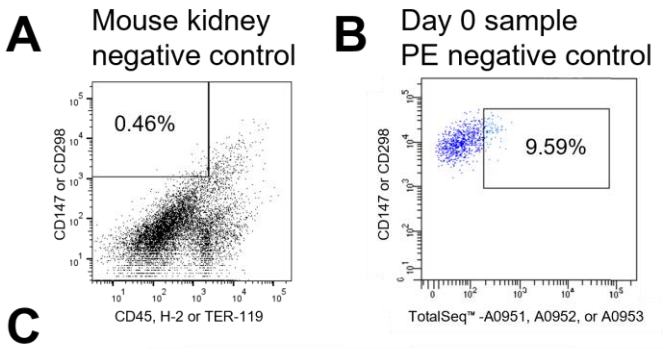


Figure S1. hiPSC-derived pancreatic endoderm cells differentiate into both pancreatic exocrine and endocrine tissues *in vivo*.

(A) Hematoxylin-Eosin (HE) staining of grafts on days 0, 7, 14, 21, and 30 after implantation (upper panels) and magnified images of grafts on days 7 and 14 (lower panels). Green arrowheads indicate grafts. Note that the graft on day 0 immediately after implantation was not attached to the host mouse kidney. Representative graft images from three independent experiments (n=3) are shown. (B) Immunostaining of grafts on the indicated days after implantation in the two different series of experiments from Fig. 1E for CPA1 (green) and NKX6.1 (red; upper panels), Mucin 1 (green) and E-CADHERIN (red; middle panels), and PRSS1 (green) and nuclei (blue; lower panels). (C) Immunostaining of grafts on day 30 for PRSS1 (green), SOX9 (red), INSULIN (magenta), and nuclei (blue; upper left and middle panels); PTF1A (green), NKX6.1 (red), and CPA1 (magenta; upper right panel); and GLUCAGON (red), INSULIN (magenta), and nuclei (blue; lower panels). Scale bars: 100 μm.


A
Mouse kidney
negative control
0.46%
CD147 or CD298
CD45, H-2 or TER-119
B
Day 0 sample
PE negative control
9.59%
CD147 or CD298
TotalSeq™ -A0951, A0952, or A0953

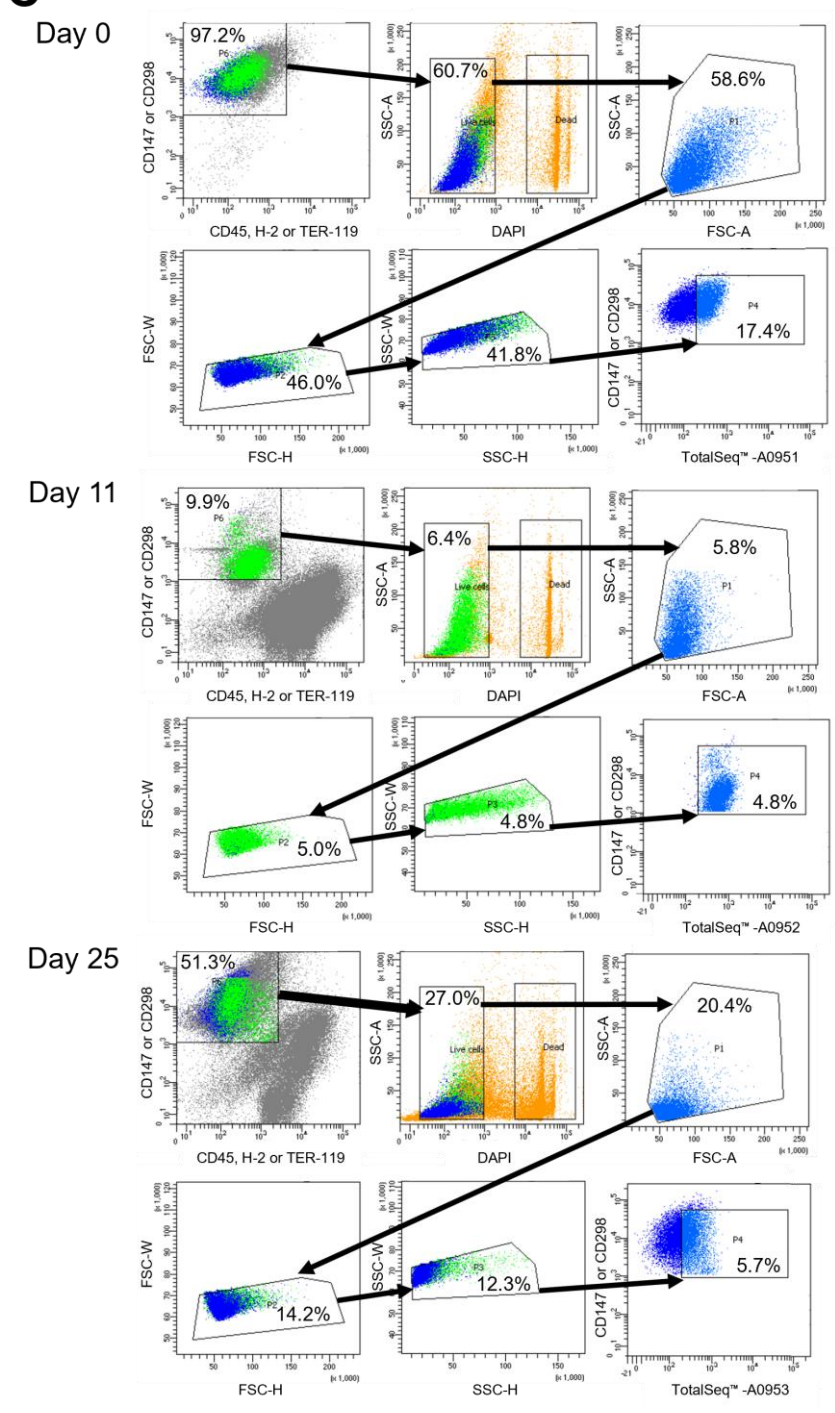

C
Day 0
97.2%
CD147 or CD298
CD45, H-2 or TER-119
60.7%
SSC-A
DAPI
58.6%
FSC-A
46.0%
FSC-W
FSC-H
41.8%
SSC-W
SSC-H
17.4%
CD147 or CD298
TotalSeq™ -A0951
Day 11
9.9%
CD147 or CD298
CD45, H-2 or TER-119
6.4%
SSC-A
DAPI
5.8%
FSC-A
5.0%
FSC-W
FSC-H
4.8%
SSC-W
SSC-H
4.8%
CD147 or CD298
TotalSeq™ -A0952
Day 25
51.3%
CD147 or CD298
CD45, H-2 or TER-119
27.0%
SSC-A
DAPI
20.4%
FSC-A
14.2%
FSC-W
FSC-H
12.3%
SSC-W
SSC-H
5.7%
CD147 or CD298
TotalSeq™ -A0953

Figure S2. Sorting strategy for flow cytometry

(A) To isolate human cells using a combination of FITC-conjugated mouse-specific CD45, H-2, and TER-119 antibodies and APC-conjugated human-specific CD147 and CD298 antibodies, gating was set such that mouse cells had a positive fraction of less than 1% based on a negative control experiment using only NOD/SCID mouse kidney cells. (B) Staining using TotalSeq™-A0951 PE-Streptavidin, TotalSeq™-A0952 PE-Streptavidin, and TotalSeq™-A0953 PE-Streptavidin with EZ-Link™ Sulfo-NHS-LC-Biotin was used to identify cells on days 0, 11, and 25, respectively. To isolate PE-Streptavidin-Biotin-tagged cells positive for TotalSeq™-A0951, A0952, or A0953, gating was set such that non-fluorescent cells had a positive fraction of less than 10% based on a negative control experiment using unstained hiPSC-derived pancreatic endoderm cells. (C) First, human cells positive for either human CD147 or CD298 but not mouse CD45, H-2, or TER-119 were sorted from mixtures of human and mouse cells based on a combination of FITC and APC fluorescence intensity. Blood cells were also sorted and removed by mouse-specific CD45, H-2, and TER-119 antibodies. Second, $DAPI^-$ live cells were sorted from human cells, with broken cells and doublets removed using a combination of FSC and SSC gating. Third, only cells positive for TotalSeq™-A0951, A0952, or A0953 were sorted based on PE fluorescence intensity. The percentages listed in each gate indicate the proportion of all cells before sorting on days 0, 11, and 25. In the left and middle panels, green and blue dots indicate TotalSeq™-A0951, A0952, or A0953 $(PE)^+$ and PE^- cells, corresponding to the light blue and dark blue dots in the lower right panels, respectively.

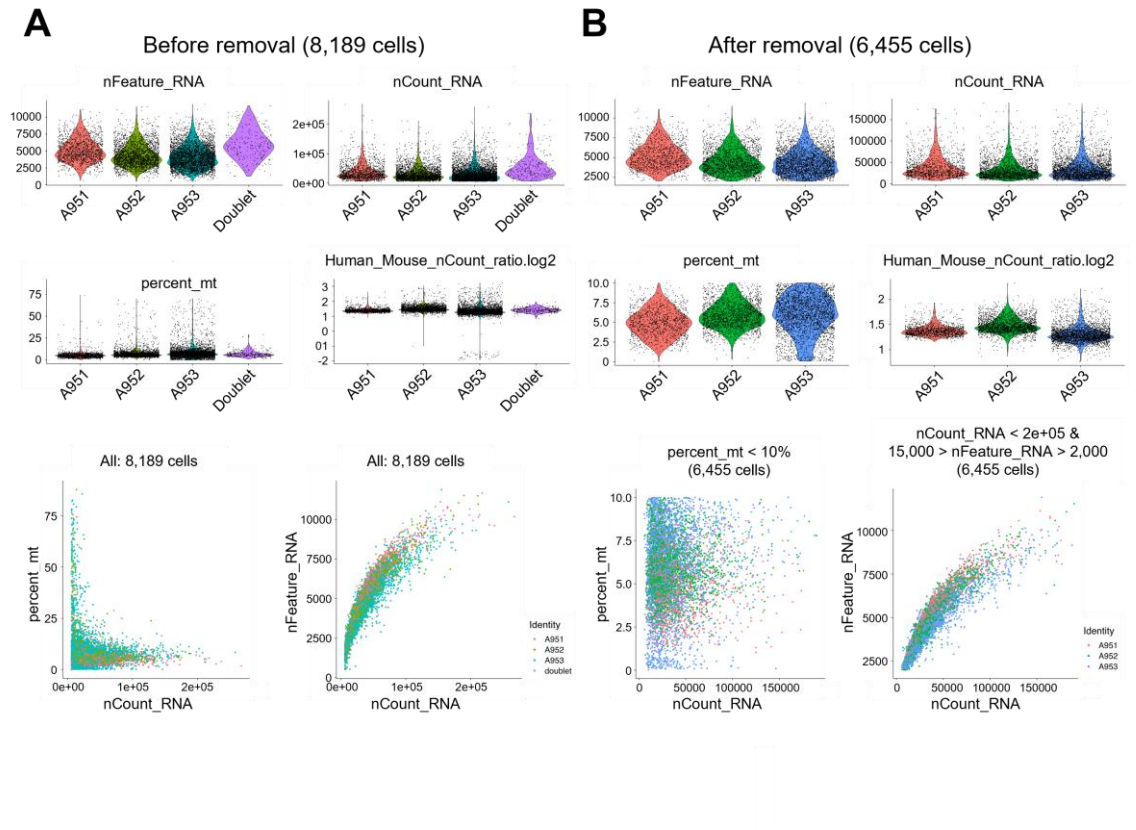


Figure S3. Quality control of scRNA-seq

(A, B) Violin plots of the number of genes (nFeature_RNA; upper left panels) and RNA count (nCount_RNA; upper right panels) detected in each cell and fraction of RNA count of mitochondrial genes (percent.mt; lower left panels) and Human_Mouse_nCount_ratio.log2 (lower right panels) before (A) and after excluding cells not suitable for analysis (B). The exclusion criteria were nCount_RNA ≧20,000, nFeature_RNA ≦ 2,000 or ≧ 15,000, percent.mt ≧ 10%, and Human_Mouse_nCount_ratio.log2 ≦0.6. All 8,189 cells derived from 96 pancreatic endoderm aggregates on day 0, as well as 8 and 5 graft samples isolated on days 11 and 25, respectively, were pooled upon isolation.

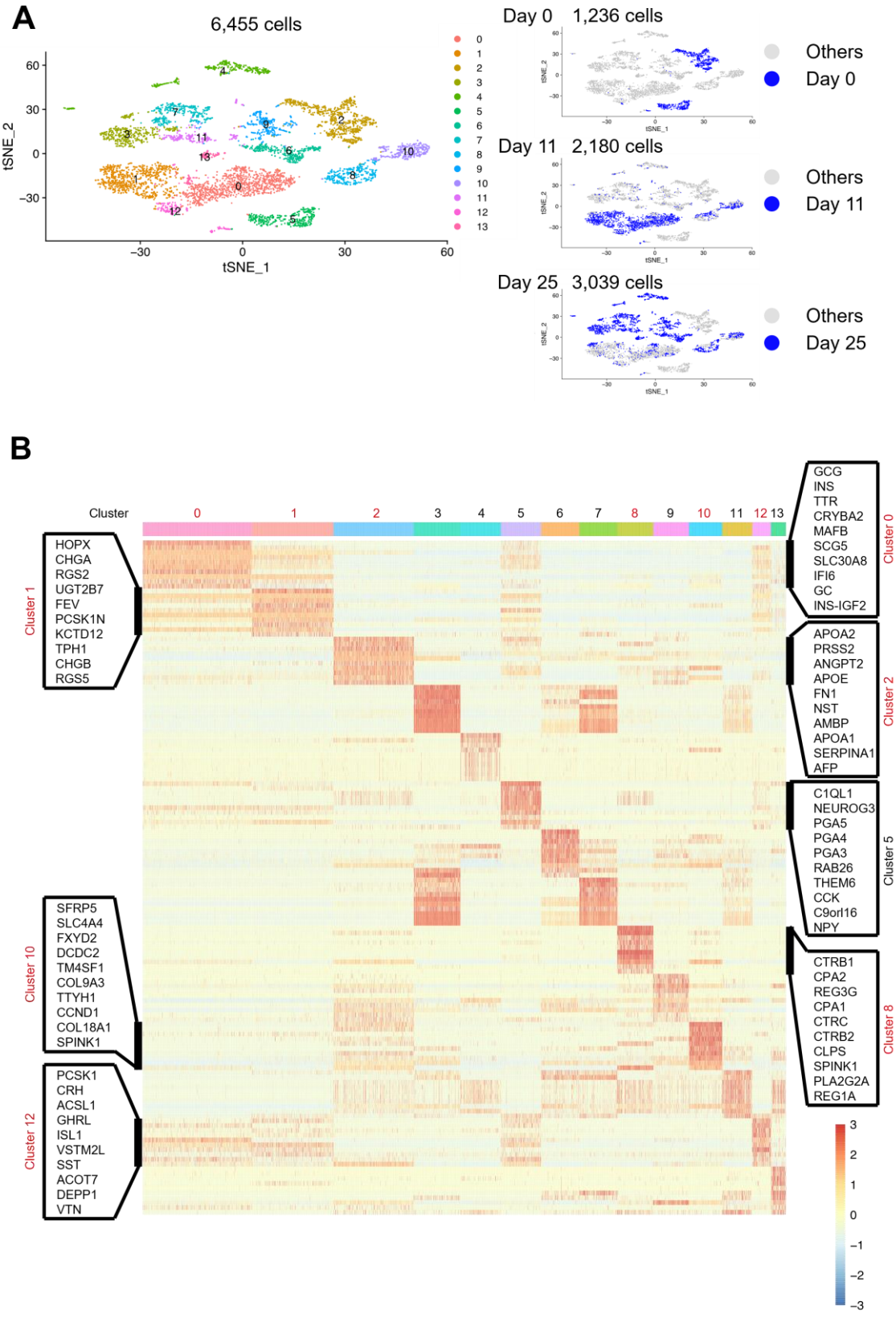
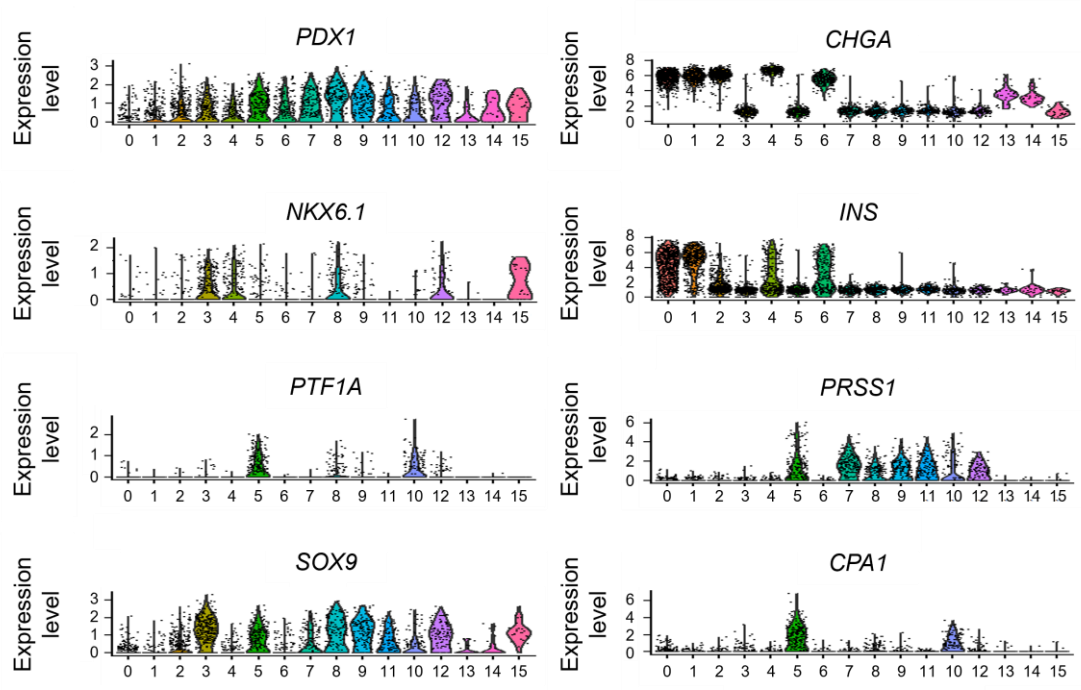

A
6,455 cells
tSNE_2
tSNE_1
Day 0 1,236 cells
Others
Day 0
Day 11 2,180 cells
Others
Day 11
Day 25 3,039 cells
Others
Day 25
B
Cluster
HOPX
CHGA
RGS2
UGT2B7
FEV
PCSK1N
KCTD12
TPH1
CHGB
RGS5
Cluster 1
SFRP5
SLC4A4
FXYD2
DCDC2
TM4SF1
COL9A3
TTYH1
CCND1
COL18A1
SPINK1
Cluster 10
PCSK1
CRH
ACSL1
GHRL
ISL1
VSTM2L
SST
ACOT7
DEPP1
VTN
Cluster 12
GCG
INS
TTR
CRYBA2
MAFB
SCG5
SLC30A8
IFI6
GC
INS-IGF2
Cluster 0
APOA2
PRSS2
ANGPT2
APOE
FN1
NST
AMBP
APOA1
SERPINA1
AFP
Cluster 2
C1QL1
NEUROG3
PGA5
PGA4
PGA3
RAB26
THEM6
CCK
C9orl16
NPY
Cluster 5
CTRB1
CPA2
REG3G
CPA1
CTRC
CTRB2
CLPS
SPINK1
PLA2G2A
REG1A
Cluster 8

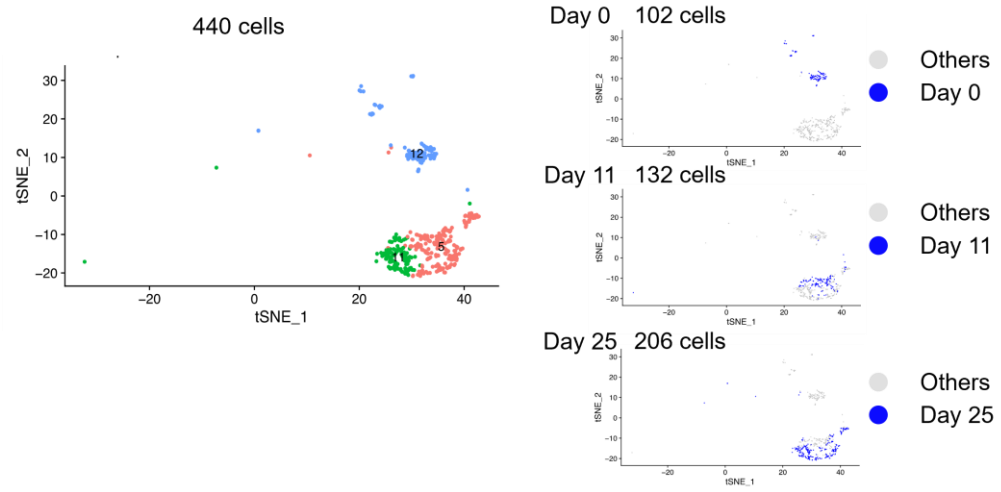
Figure S4. Transition dynamics of differentiation states of all human cells in hiPSC-derived pancreatic endoderm and grafts revealed by scRNA-seq.

(A) Clustering of a total of 6,455 cells consisting of all the obtained human cells after QC shown in Fig. S3 (left panel) and distribution of cells on the indicated days (right panels) (n=96 for pancreatic endoderm aggregate, n=8 for day 11 grafts and n=5 for day 25 grafts). All pancreatic endoderm aggregates and graft samples collected on days 11and 25 were pooled upon isolation. (B) A heatmap showing the differentially expressed genes (DEGs; $|\log2FC| \geq 0.25$, P value< 0.01). We annotated clusters 0, 1, and 12 as pancreatic endocrine cells, cluster 2 as pancreatic precursor cells, clusters 8 and 10 as pancreatic exocrine cells, cluster 5 as duodenal cells, and clusters 3, 7, 11, and 13 as mesenchymal cells, but clusters 4, 6, and 9 could not be annotated. All cluster-specific markers are listed in Supplementary Table 1. Note that by using larger-sized gene names, clusters 0, 1, and 12 (endocrine cells), 2 (pancreatic precursor cells), 5 (duodenal cells), and 8 and 10 (exocrine cells) are zoomed in. The pancreas-specific clusters are shown in red.

A


PDX1
CHGA
NKX6.1
INS
PTF1A
PRSS1
SOX9
CPA1
Expression level
0 1 2 3 4 5 6 7 8 9 11 10 12 13 14 15

B


440 cells
tSNE_1
tSNE_2
Day 0 102 cells
Day 11 132 cells
Day 25 206 cells
Others
Day 0
Day 11
Day 25

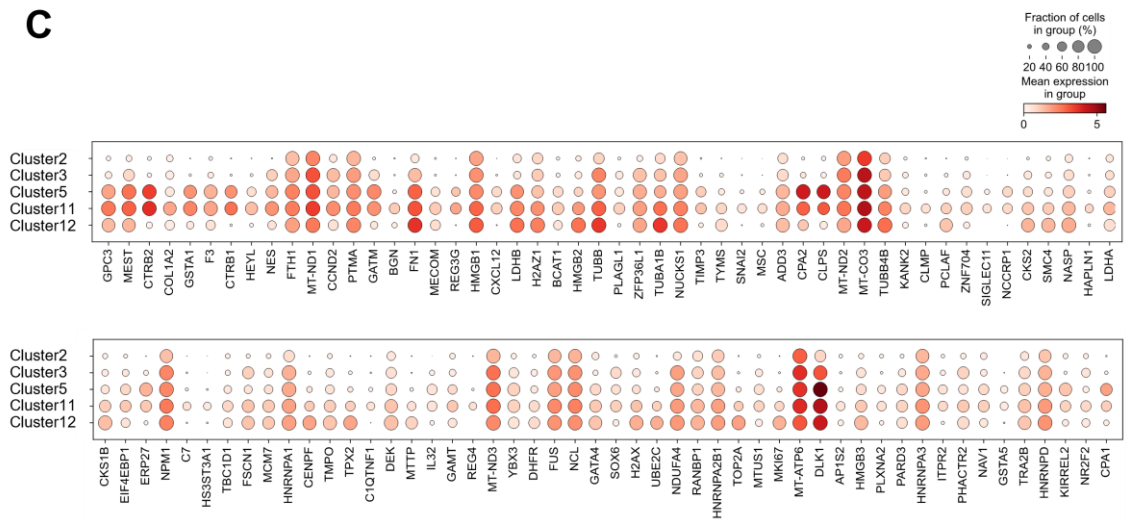


Figure S5. Transition dynamics of differentiation states after reclustering revealed by scRNA-seq.

(A) Distribution of cells expressing marker genes for pancreatic endoderm cells (*PDX1*, *NKX6.1*, *PTF1A*, and *SOX9*), pancreatic endocrine cells (*CHGA* and *INS*), and pancreatic acinar cells (*PRSS1* and *CPA1*), depicted as violin plots. (B) Clustering of 440 cells after narrowing down the clusters to 5, 11, and 12 (left panel) and distribution of cells on the indicated days (right panels). (C) Dot plots of DEGs (|log2FC|> 1, P value< 0.01) showing the top 100 genes whose expression levels were significantly higher in acinar progenitor cells (cluster 11) than in other clusters, pancreatic endoderm cells (cluster 12), acinar cells (cluster 5), endocrine cells (cluster 2), and duct cells (cluster 3).

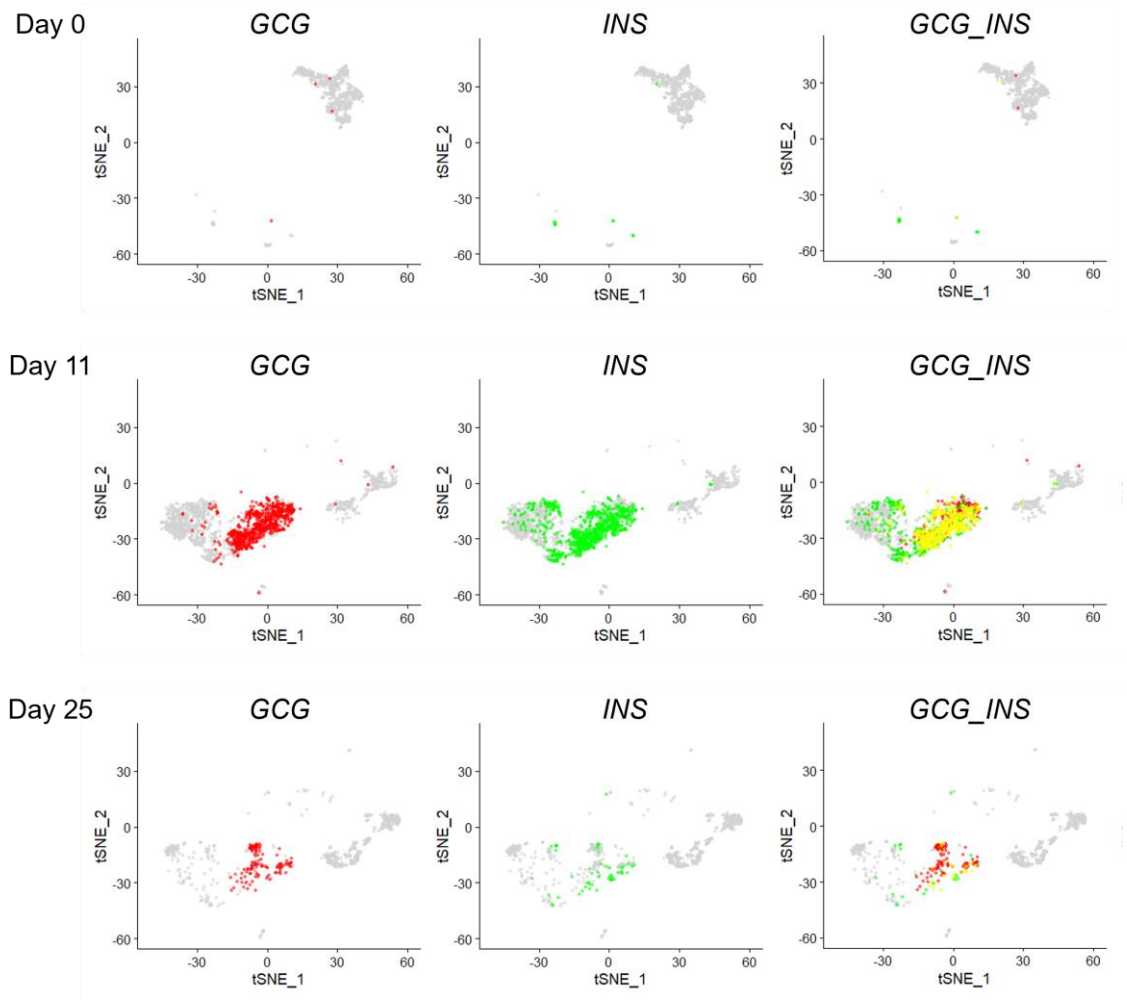
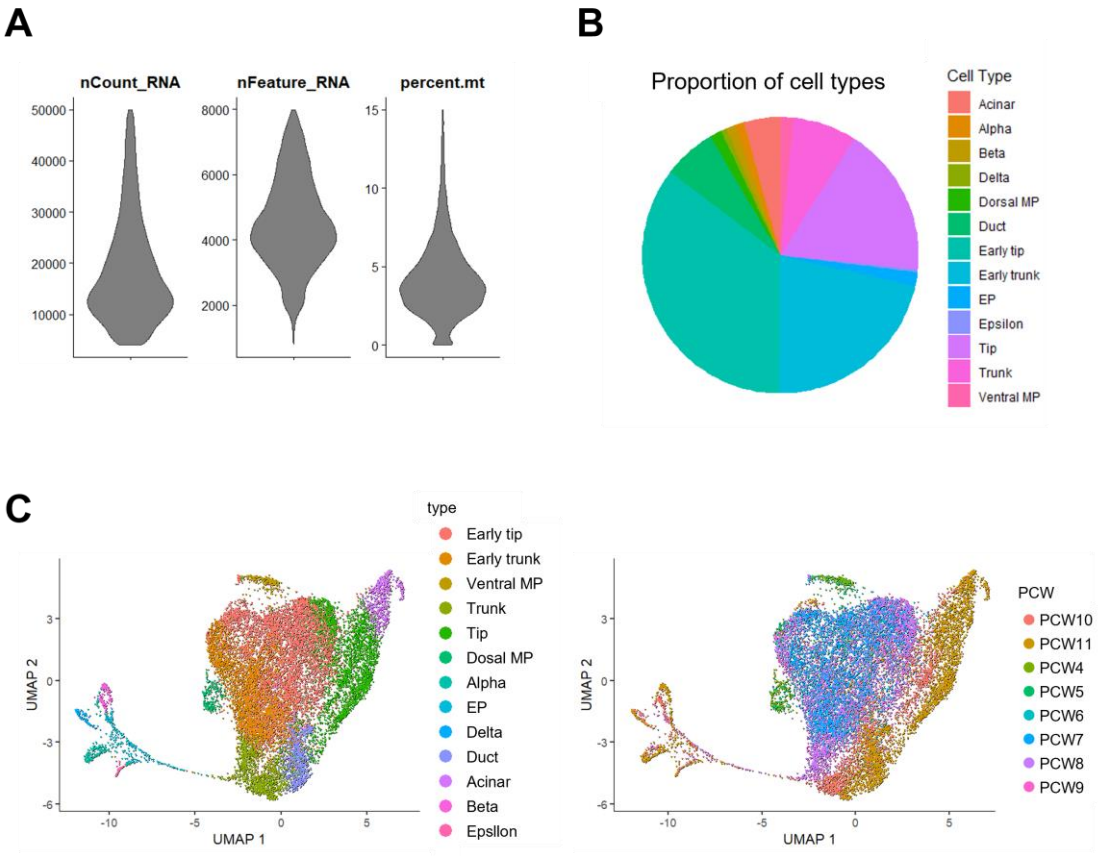
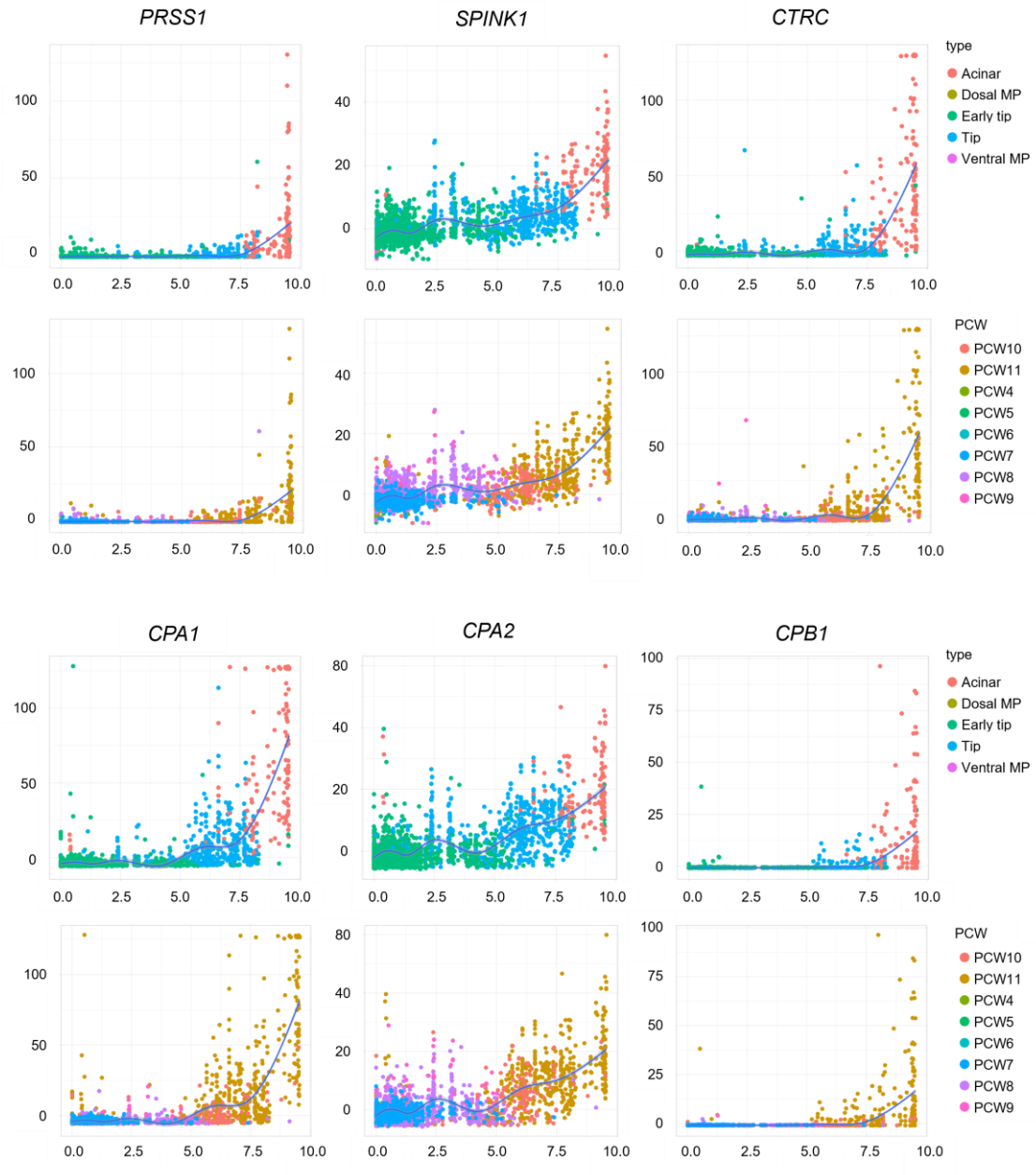


Figure S6. Cell transition from poly- to mono-hormonal over time

Feature plots showing the expression levels of *GLUCAGON* (*GCG*; left panels) and *INSULIN* (*INS*; middle panels) and co-expression levels of *GCG* and *INS* (right panels) among the cells of grafts on days 0 (upper panels), 11 (middle panels), and 25 (lower panels) after implantation.


A
nCount_RNA
nFeature_RNA
percent.mt
50000
40000
30000
20000
10000
8000
6000
4000
2000
15
10
5
0
B
Proportion of cell types
Cell Type
Acinar
Alpha
Beta
Delta
Dorsal MP
Duct
Early tip
Early trunk
EP
Epsilon
Tip
Trunk
Ventral MP
C
type
Early tip
Early trunk
Ventral MP
Trunk
Tip
Dosal MP
Alpha
EP
Delta
Duct
Acinar
Beta
Epsllon
UMAP 2
UMAP 1
PCW
PCW10
PCW11
PCW4
PCW5
PCW6
PCW7
PCW8
PCW9

D


PRSS1
SPINK1
CTRC
type
Acinar
Dosal MP
Early tip
Tip
Ventral MP
PCW
PCW10
PCW11
PCW4
PCW5
PCW6
PCW7
PCW8
PCW9
CPA1
CPA2
CPB1

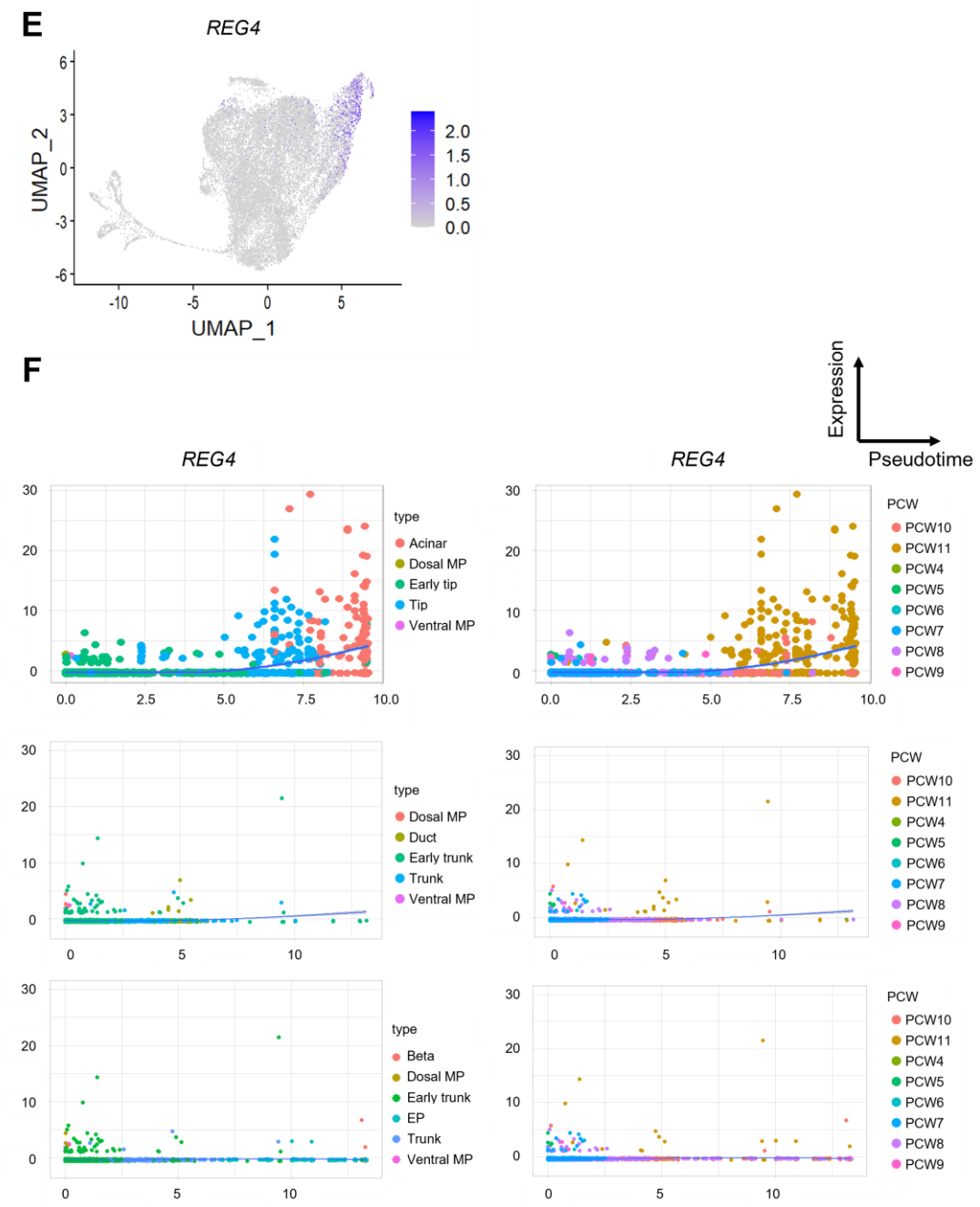


Figure S7. Reanalysis of a previously published human fetal pancreas dataset in the reference [48]

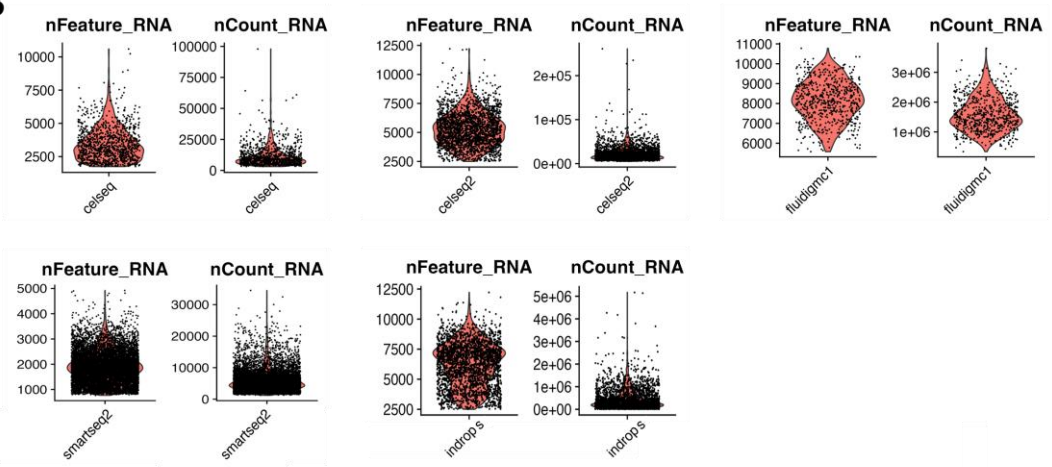
(A) QC metrics, nCount_RNA (left panel), nFeature_RNA (middle panel) and percent.mt (right panel). (B) Pie chart showing cellular type composition in the scRNA-seq data. (C) UMAP plots of all pancreatic cells from post-conception weeks (PCWs) 4 to 11 colored by cell types (left panel) and PCWs (right panel). (D) Scatter plots colored by cell types

(upper panels) and PCWs (lower panels) showing the gene expression levels of the indicated acinar markers in acinar-related cells (ventral and dorsal multipotent progenitors, early tip, tip, and acinar cells) across pseudotime. Lines indicate the fitted generalized additive models. (E) Feature plot showing the expression of *REG4*. (F) Scatter plots colored by cell types (left panels) and PCWs (right panels) showing *REG4* expression levels in acinar-related cells (ventral and dorsal multipotent progenitors, early tip, tip, and acinar cells; upper panels), duct-related cells (ventral and dorsal multipotent progenitors, early trunk, trunk, and duct cells; middle panels), and endocrine-related cells (ventral and dorsal multipotent progenitors, early trunk, trunk, endocrine progenitors, and beta cells; lower panels) across pseudotime. Lines indicate the fitted generalized additive models.

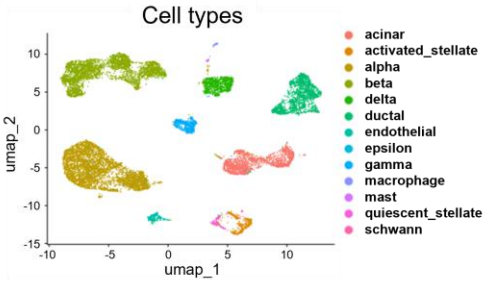
A

Dataset	Technology	Number of cells	Human doner
GSE81076	CEL-Seq	1004	5 deceased organ donors with and without T2D
GSE85241	CEL-Seq2	2285	4 deceased organ donors
GSE86469	Fluidigm C1	638	8 deceased organ donors with (3) and without (5) T2D
E-MTAB-5061	Smart-Seq2	8569	10 deceased organ donors with (4) and without (6) T2D
GSE84133	inDrops	2394	4 deceased organ donors

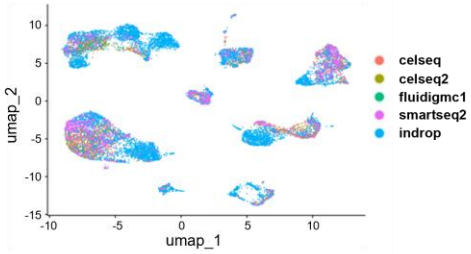
B



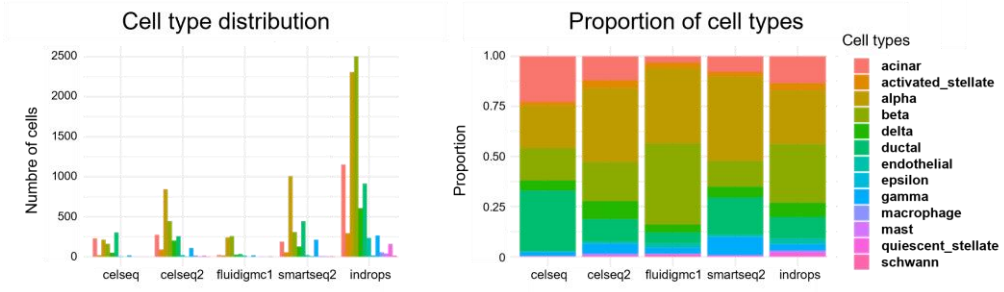
C



D



E



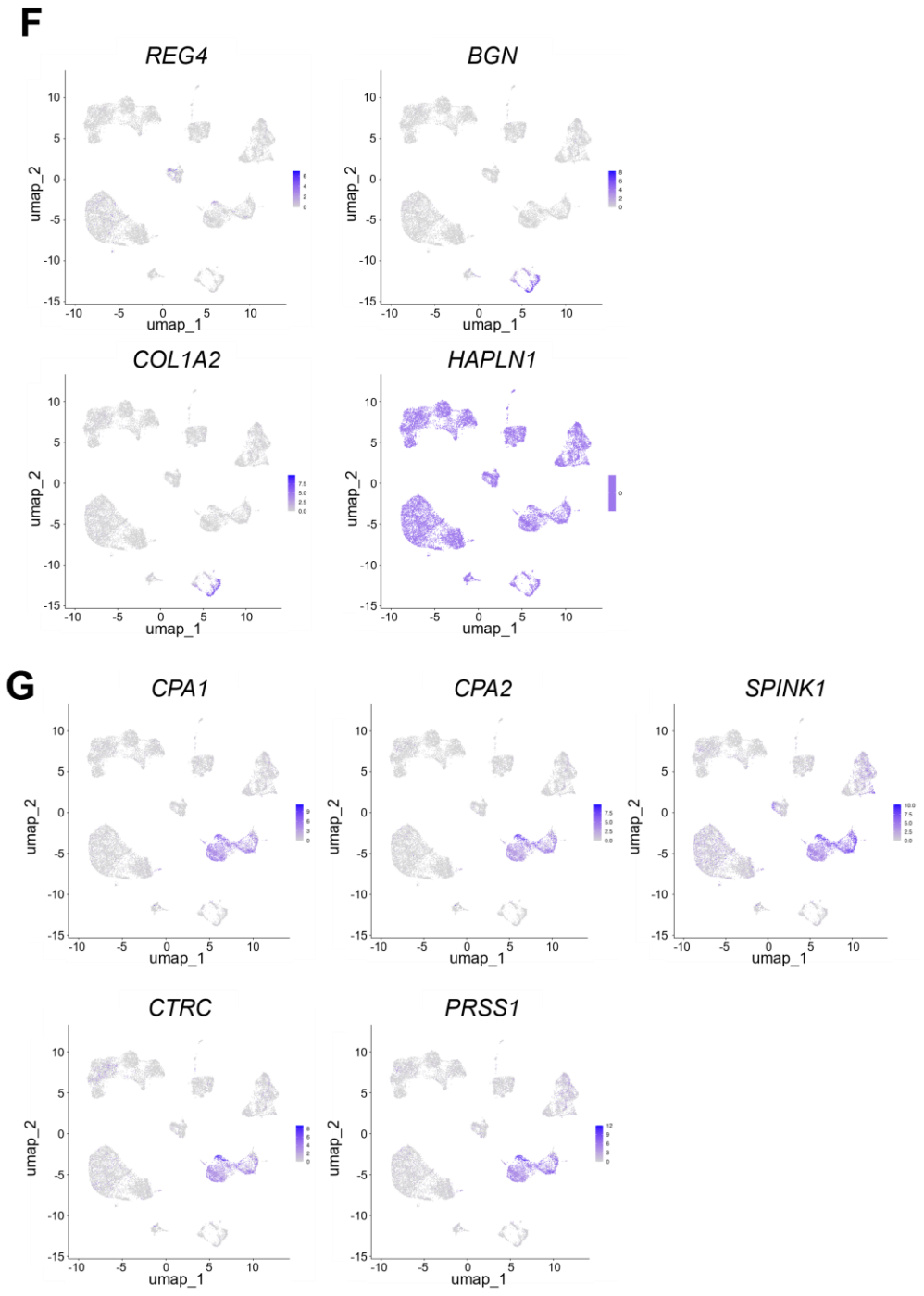


Figure S8. Reanalysis of previously published human adult pancreas datasets, originally derived from five datasets (GSE81076, GSE85241, GSE86469, E-MTAB-5061, and GSE84133) in the references [50-54]

(A) Sample information of the original data sets (GSE81076, GSE85241, GSE86469, E-MTAB-5061, and GSE84133; 50-54). (B) QC metrics of each dataset, nFeature_RNA (left panels) and nCount_RNA (right panels). (C) Visualization of each cell cluster with UMAP. (D) Evaluation of unbiased cell types by single-cell technologies. (E) Evaluation of cell type distribution (left panel), and proportion of cell types by single-cell technologies (right panel). (F, G) Gene expression pattern across cell types for candidate acinar progenitor cell markers, *REG4*, *BGN*, *COL1A2*, and *HAPLN1* (F), and acinar cell markers, *CPA1*, *CPA2*, *SPINK1*, *CTRC*, and *PRSS1* (G).

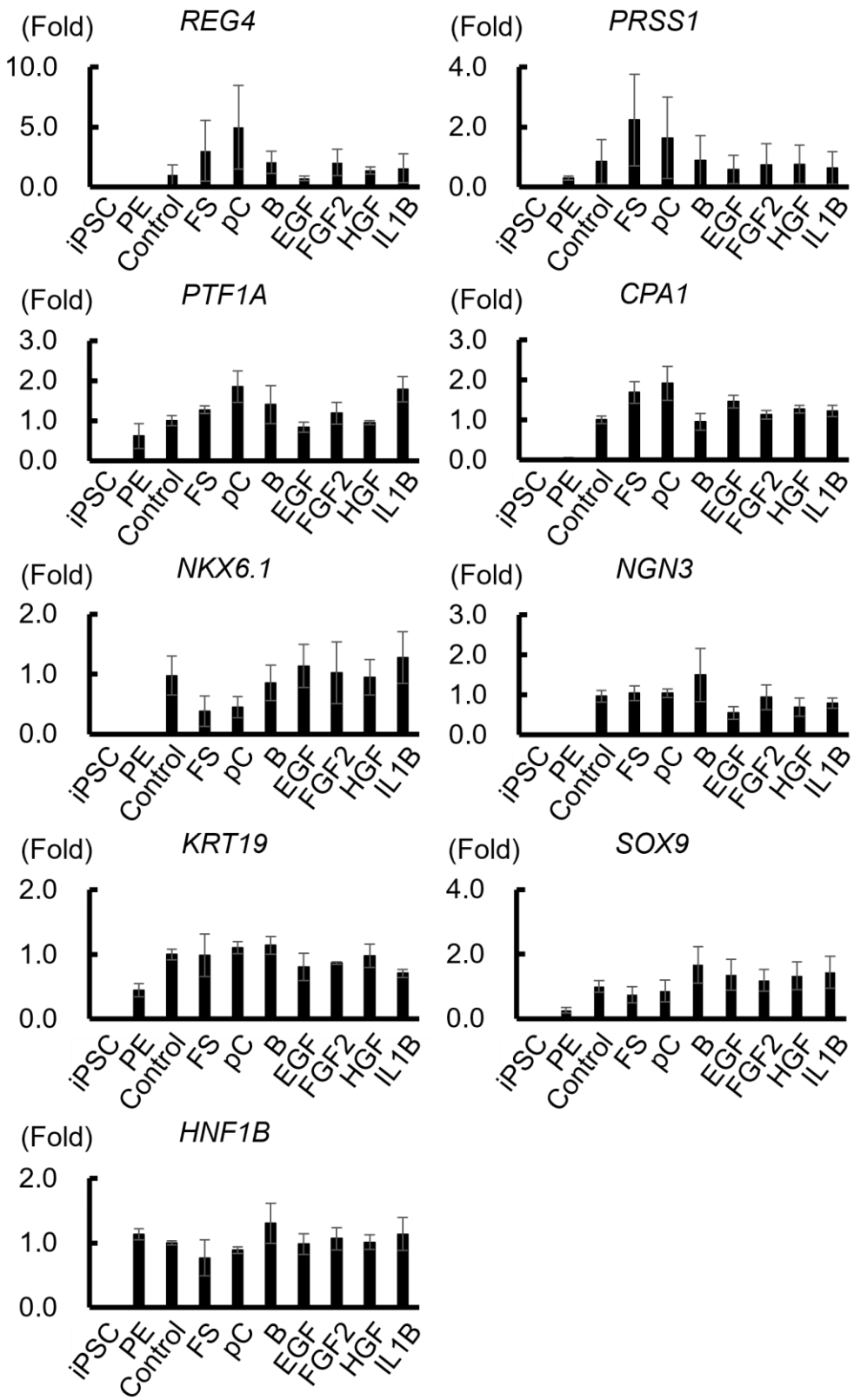

(Fold)
REG4
10.0
5.0
0.0
iPSC
PE
Control
FS
pC
B
EGF
FGF2
HGF
IL1B
(Fold)
PRSS1
4.0
2.0
0.0
(Fold)
PTF1A
3.0
2.0
1.0
0.0
(Fold)
CPA1
3.0
2.0
1.0
0.0
(Fold)
NKX6.1
2.0
1.0
0.0
(Fold)
NGN3
3.0
2.0
1.0
0.0
(Fold)
KRT19
2.0
1.0
0.0
(Fold)
SOX9
4.0
2.0
0.0
(Fold)
HNF1B
2.0
1.0
0.0

Figure S9. Marker gene expression for three pancreatic lineages induced by treatment of hiPSC-derived pancreatic endoderm with candidate factors.

A comparison of the effects of 7 candidate factors on the expression of pancreatic acinar lineage markers, *REG4*, *PRSS1*, *PTF1A*, and *CPA1*, endocrine markers, *NKX6.1* and *NGN3*, and duct markers, *KRT19*, *SOX9*, and *HNF1B*, as evaluated by qRT-PCR. The treatment started on Stage 4 Day 6 (total day 17) and continued for 8 days, ending on total day 25. The following factors and concentrations were examined: 20 nM forskolin (FS); 100 µM pCPT-cAMP (pC); 0.02 nM bexarotene (B); 100 ng/mL EGF; 10 ng/mL FGF2; 100 ng/mL HGF; and 2 ng/mL IL-1B. Data from three independent experiments are presented as mean ± SD (n = 3). We normalized the expression values against the house keeping gene *GAPDH* and then against the untreated control (Control). PE, pancreatic endoderm.

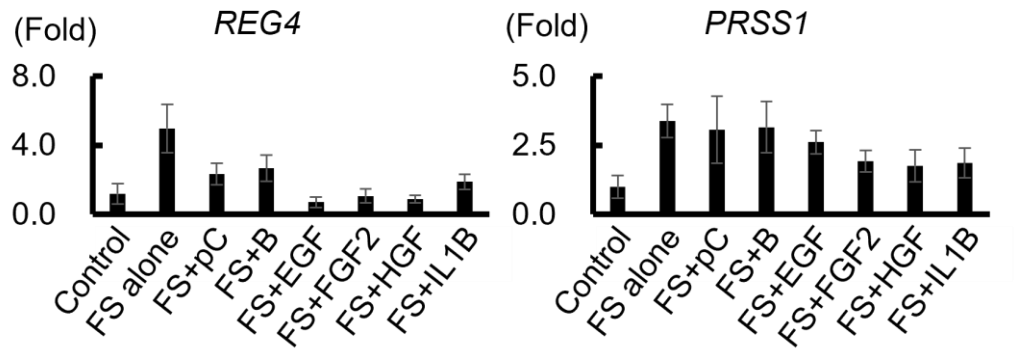


Figure S10. Expression of *REG4* and *PRSS1* induced by combinatorial treatments of hiPSC-derived pancreatic endoderm with forskolin and each of six candidate factors.

A comparison of the effects of combinatorial treatments with forskolin (FS) and each of six candidate factors, EGF, IL-1B, HGF, pCPT-cAMP (pC), FGF2, and bexarotene (B) on the expression of *REG4* and *PRSS1* as evaluated by qRT-PCR. The treatment started on Stage 4 Day 6 (total day 17) and continued for 8 days, ending on total day 25. The same factors and concentrations as used in Fig. S9 were examined. Data from three independent experiments are presented as mean ± SD (n = 3). We normalized the expression values against *GAPDH* and then against the untreated control (Control).

Supplementary Table 5. Antibodies used in this study

Antigen	Host Species	Company	Catalog Number	Dilution
PDX1	Goat	R&D	AF2419	1:200
NKX6.1	Mouse	University of Iowa	F55A12	1:40
PTF1A	Armenian hamster	Kind gift from Dr. Yuichi Ono, KAN Institute		1:2
CPA1	Rabbit	AbD Serotec	AHP2054	1:100
PRSS1	Sheep	R&D	AF3586	1:200
Mucin1	Rabbit	Abcam	ab15481	1:300
E-CADHERIN	Mouse	BD Biosciences	610181	1:1000
SOX9	Mouse	Abnova	H00006668-M04	1:200
INSULIN	Guinea pig	DAKO	A0564	1:50
GULCAGON	Mouse	SIGMA	G2654	1:100
CD45	Rat	BioLegend	147709	1:40
H-2	Rat	BioLegend	125507	1:40
TER-119	Rat	BioLegend	116205	1:40
CD298	Mouse	BioLegend	341706	1:40
CD147	Mouse	BioLegend	306213	1:40
Alexa Fluor 488-anti-goat	donkey	Thermo Fisher Scientific	A-11055	1:500
Alexa Fluor 488-anti-sheep	donkey	Thermo Fisher Scientific	A11015	1:500
Alexa Fluor 488-anti-rabbit	donkey	Thermo Fisher Scientific	A21206	1:500
PE- anti-armenian hamster	goat	Thermo Fisher Scientific	PA1-32045	1:500
Alexa Fluor 546-anti-mouse	donkey	Thermo Fisher Scientific	A-10036	1:500
Alexa Fluor 546-anti-rabbit	donkey	Thermo Fisher Scientific	A10040	1:500
Alexa Fluor 647-anti-guinea pig	donkey	Jackson ImmunoResarch	706-605-148	1:500

Alexa Fluor 647-anti-rabbit	donkey	Thermo Fisher Scientific	A31573	1:500
TotalSeq™-A0951 PE-Streptavidin		BioLegend	405251	1:500
TotalSeq™-A0952 PE-Streptavidin		BioLegend	405253	1:500
TotalSeq™-A0953 PE-Streptavidin		BioLegend	405255	1:500

Supplementary Table 6. Primers used in this study

Primer	Forward	Reverse
GAPDH	GAAGGTGAAGGTCGGAGTC	GAAGATGGTGATGGGATTTC
REG4	ATCAGAGAAGCCAGCCGATA	TGGAGCTCATCTCAGCACAG
PRSS1	AGCCAGGCTAAGTGTGAAGC	CTGGGGGCTTTAGCTATTGG
PTF1A	CCCCAGCGACCCTGATTA	GGACACAAACTCAAATGGTGG
CPA1	ACTACGCCACCTACCACACC	CAGATGGCTGGACGCTTACT
SPINK1	CCTTGGCCCTGTTGAGTCTA	GGTGACCTGATGGGATTTCA
CTRC	TGTGGCGGGACTTTGATT	GGTGTCCACACCCACAAA
NGN3	ACCCCATTCTCTCTTCTTTTCTCCT	GAGCCGTCATCCTTTCTACCG
NKX6.1	ATTCGTTGGGGATGACAGAG	TGGGATCCAGAGGCTTATTG
KRT19	CTACAGCCACTACTACACGAC	CAGAGCCTGTTCCGTCTCAAA
HNF1B	TCACAGATACCAGCAGCATCAGT	GGGCATCACCAGGCTTGTA
SOX9	GCTCTGGAGACTTCTGAACGA	CCGTTCTTCACCGACTTCCT