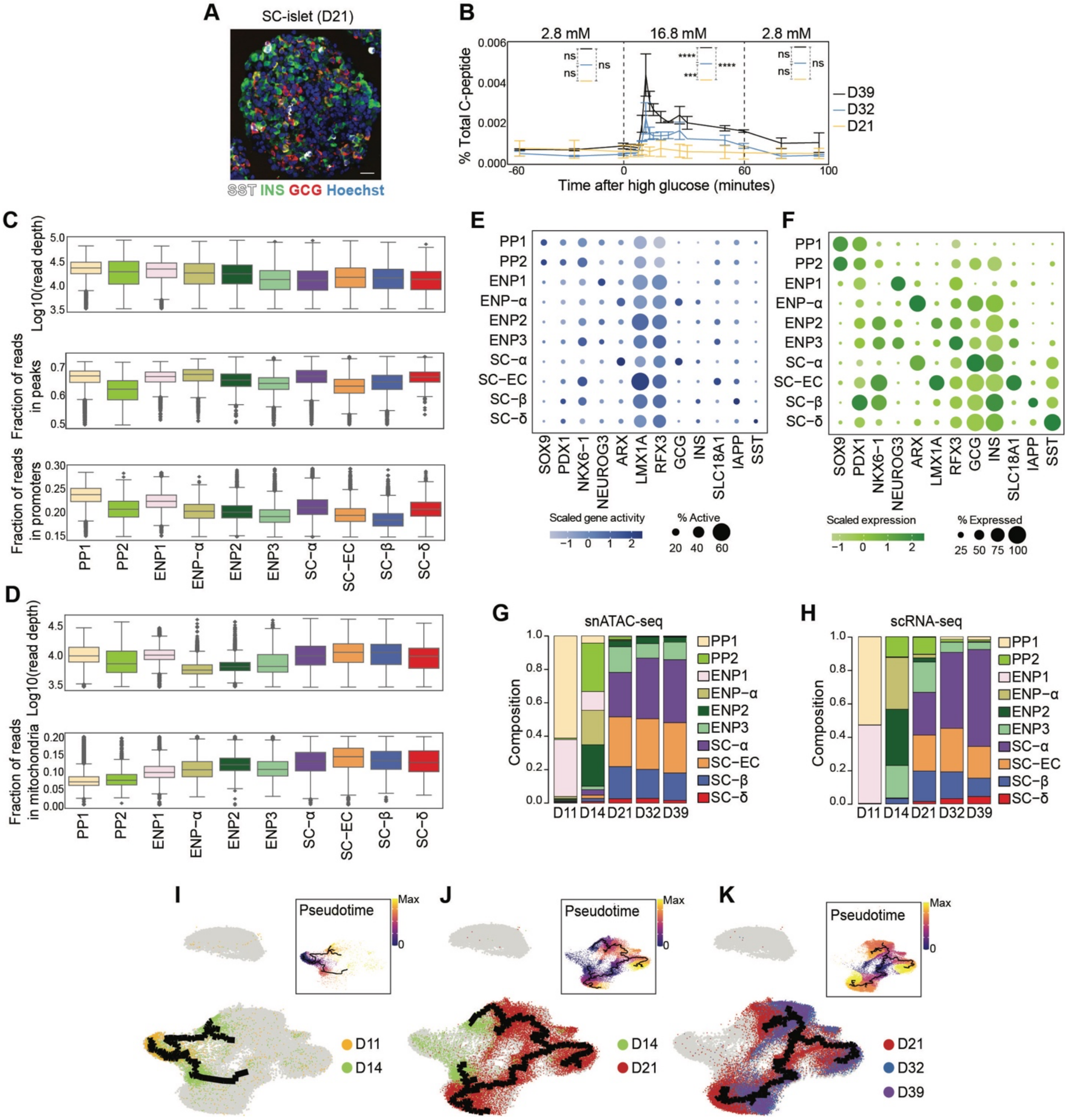
A
SC-islet (D21)
SST INS GCG Hoechst
B
2.8 mM
16.8 mM
2.8 mM
% Total C-peptide
0.006
0.004
0.002
0.000
-60
0
60
100
Time after high glucose (minutes)
ns
ns
ns
****
***
****
ns
ns
ns
D39
D32
D21
C
Log10(read depth)
5.0
4.5
4.0
3.5
Fraction of reads in peaks
0.7
0.6
0.5
Fraction of reads in promoters
0.30
0.25
0.20
0.15
PP1
PP2
ENP1
ENP−α
ENP2
ENP3
SC−α
SC−EC
SC−β
SC−δ
D
Log10(read depth)
4.5
4.0
3.5
Fraction of reads in mitochondria
0.20
0.15
0.10
0.05
0.00
PP1
PP2
ENP1
ENP−α
ENP2
ENP3
SC−α
SC−EC
SC−β
SC−δ
E
PP1
PP2
ENP1
ENP−α
ENP2
ENP3
SC−α
SC−EC
SC−β
SC−δ
SOX9
PDX1
NKX6−1
NEUROG3
ARX
LMX1A
RFX3
GCG
INS
SLC18A1
IAPP
SST
Scaled gene activity
−1
0
1
2
% Active
20
40
60
F
PP1
PP2
ENP1
ENP−α
ENP2
ENP3
SC−α
SC−EC
SC−β
SC−δ
SOX9
PDX1
NKX6−1
NEUROG3
ARX
LMX1A
RFX3
GCG
INS
SLC18A1
IAPP
SST
Scaled expression
−1
0
1
2
% Expressed
25
50
75
100
G
snATAC-seq
Composition
1.0
0.8
0.6
0.4
0.2
0.0
D11
D14
D21
D32
D39
PP1
PP2
ENP1
ENP−α
ENP2
ENP3
SC−α
SC−EC
SC−β
SC−δ
H
scRNA-seq
Composition
1.0
0.8
0.6
0.4
0.2
0.0
D11
D14
D21
D32
D39
PP1
PP2
ENP1
ENP−α
ENP2
ENP3
SC−α
SC−EC
SC−β
SC−δ
I
Pseudotime
Max
0
D11
D14
J
Pseudotime
Max
0
D14
D21
K
Pseudotime
Max
0
D21
D32
D39

Figure S1. Quality control of snATAC-seq and scRNA-seq data analysis - related to Figure 1.

(A) Representative immunofluorescent images for somatostatin (SST), insulin (INS) and glucagon (GCG) at D21 of SC-islet differentiation. Nuclei were labeled with Hoechst. Scale bar, 20 μm.

(B) Human C-peptide secretion by SC-islets at D21, D32 and D39 during perifusion with the indicated glucose (Glc) concentrations (in mM). Data are shown as mean ± S.D. at the indicated time points. ns, not significant, ****P* < 0.001, *****P* < 0.0001, using two-way ANOVA followed by Tukey's multiple comparisons test for different days of differentiation in each time block.

(C, D) Box plots of quality control matrices for snATAC-seq (C) and scRNA-seq (D) in each cluster in Figure 1b, showing that these metrics do not drive single-cell grouping in UMAP space.

(E, F) Dot plots showing scaled average gene activity (E) or gene expression (F). The color of each dot represents the average gene activity (E) or expression level (F) and the size of each dot the percentage of positive cells for each gene.

(G, H) Relative abundance of each cell type based on snATAC-seq (G) and scRNA-seq (H) UMAP annotations in Figure 1B. Columns represent SC-islet differentiation time points.

(I-K) Trajectory analysis based on chromatin accessibility, showing trajectories from D11 and D14 (I), D14 and D21 (J), and D21 and D32/39 (K) data with ENP1, ENP2, and ENP3 set as the root, respectively. Cells were color-coded by either time point of collection or pseudotime values (insets). PP1 and PP2 cells were excluded from the analysis.

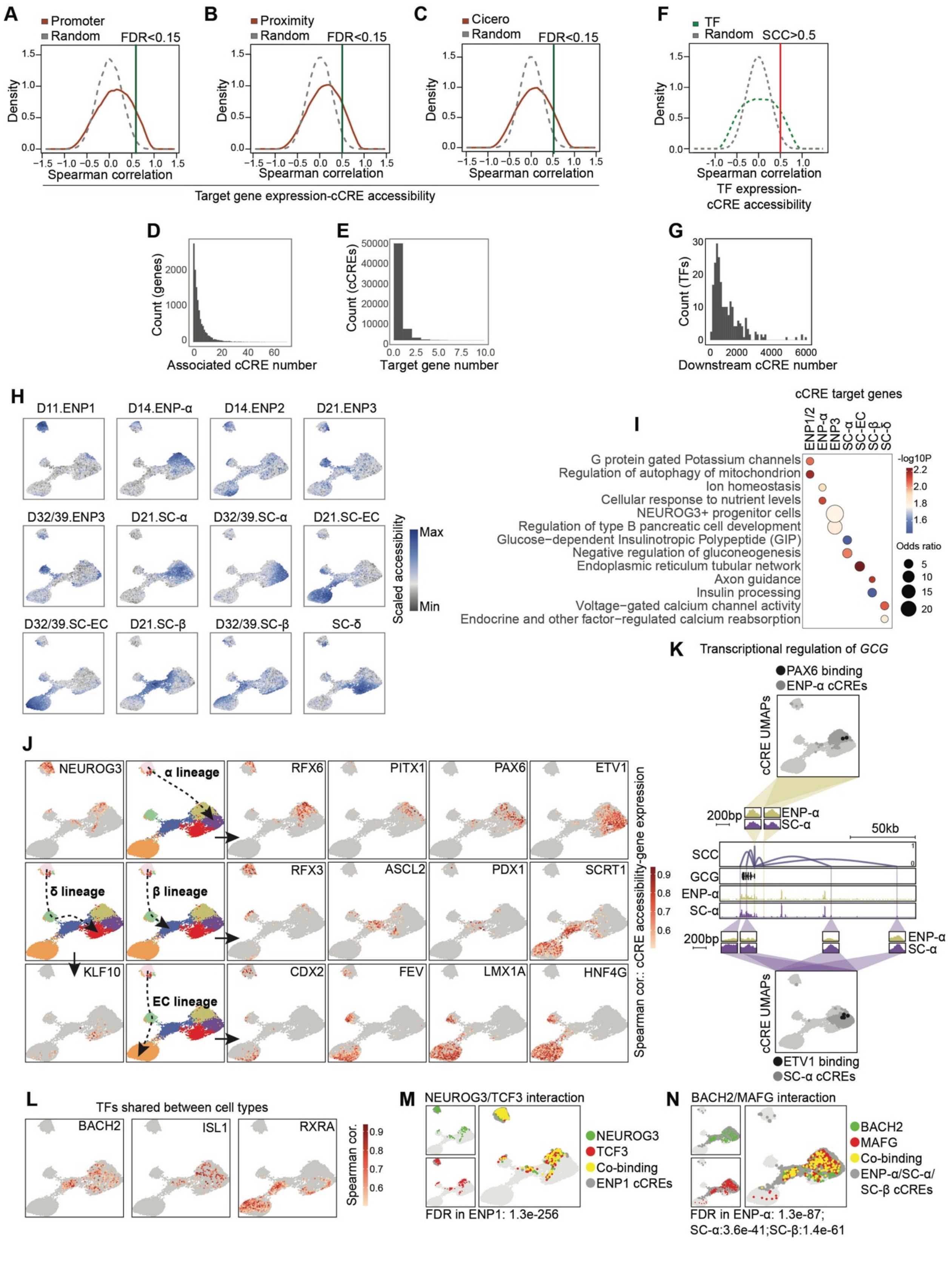
A
Promoter
Random
FDR<0.15
Density
Spearman correlation
B
Proximity
Random
FDR<0.15
Density
Spearman correlation
C
Cicero
Random
FDR<0.15
Density
Spearman correlation
Target gene expression-cCRE accessibility
F
TF
Random
SCC>0.5
Density
Spearman correlation
TF expression-
cCRE accessibility
D
Count (genes)
Associated cCRE number
E
Count (cCREs)
Target gene number
G
Count (TFs)
Downstream cCRE number
H
D11.ENP1
D14.ENP-α
D14.ENP2
D21.ENP3
D32/39.ENP3
D21.SC-α
D32/39.SC-α
D21.SC-EC
D32/39.SC-EC
D21.SC-β
D32/39.SC-β
SC-δ
Scaled accessibility
Max
Min
I
cCRE target genes
ENP1/2
ENP-α
ENP3
SC-α
SC-EC
SC-β
SC-δ
G protein gated Potassium channels
Regulation of autophagy of mitochondrion
Ion homeostasis
Cellular response to nutrient levels
NEUROG3+ progenitor cells
Regulation of type B pancreatic cell development
Glucose-dependent Insulinotropic Polypeptide (GIP)
Negative regulation of gluconeogenesis
Endoplasmic reticulum tubular network
Axon guidance
Insulin processing
Voltage-gated calcium channel activity
Endocrine and other factor-regulated calcium reabsorption
-log10P
2.2
2.0
1.8
1.6
Odds ratio
5
10
15
20
J
NEUROG3
α lineage
RFX6
PITX1
PAX6
ETV1
δ lineage
β lineage
RFX3
ASCL2
PDX1
SCRT1
KLF10
EC lineage
CDX2
FEV
LMX1A
HNF4G
Spearman cor.: cCRE accessibility-gene expression
0.9
0.8
0.7
0.6
K
Transcriptional regulation of GCG
PAX6 binding
ENP-α cCREs
cCRE UMAPs
200bp
ENP-α
SC-α
50kb
SCC
GCG
ENP-α
SC-α
200bp
ENP-α
SC-α
cCRE UMAPs
ETV1 binding
SC-α cCREs
L
TFs shared between cell types
BACH2
ISL1
RXRA
Spearman cor.
0.9
0.8
0.7
0.6
M
NEUROG3/TCF3 interaction
NEUROG3
TCF3
Co-binding
ENP1 cCREs
FDR in ENP1: 1.3e-256
N
BACH2/MAFG interaction
BACH2
MAFG
Co-binding
ENP-α/SC-α/
SC-β cCREs
FDR in ENP-α: 1.3e-87;
SC-α:3.6e-41;SC-β:1.4e-61

Figure S2. Building a gene regulatory network of stem cell islet development - related to Figure 2.

(A-C) Identification of putative cCRE-gene pairs. cCREs were assigned to a gene if they are located (A) in promoter (± 500 bp) of target gene; (B) in proximity (± 50 kb) of the gene promoter; or (C) show co-accessibility with the gene promoter by Cicero analysis. A total of 65,479 positively correlated cCRE-gene pairs were identified using an empirically defined significance threshold of FDR<0.15.

(D, E) Histogram showing characteristics of cCRE-gene pairs. Each target gene in the GRN is regulated by a mean of 5.3 cCREs (D) and each cCRE regulates a mean of 1.2 target genes (E).

(F) Identification of putative TF-cCRE pairs. Putative TF-cCRE pairs were defined using motifmatchr. A total of 280,019 significantly correlated TF-cCRE pairs were identified using an empirically defined threshold of a spearman correlation coefficient (SCC)>0.5.

(G) Histogram showing characteristics of TF-cCRE pairs. Each TF has a mean of 1,053 predicted binding sites.

(H) UMAP projections of scaled chromatin accessibility of all cCREs in each aggregated pseudo-cell. Pseudo-cells were aggregated based on time point of differentiation and assigned cell type identity.

(I) Enriched gene ontology terms/pathways among target genes associated with each cCRE module. Significance (-log10 p-value) and odds ratio of the enrichments are represented by color and dot size, respectively.

(J) UMAP projections of TF-cCRE correlations for select cell type-enriched TFs. Spearman correlation coefficients between TF expression and cCRE accessibility are shown.

(K) UMAP locations and genome browser snapshots of predicted PAX6- or ETV1-bound cCREs at the *GCG* gene locus. Dark grey dots indicate cCRE modules with predicted TF interactions. Genome browser tracks show aggregated ATAC reads in ENP-α and SC-α-cells. All tracks are scaled to uniform $1x10^6$ read depth. SCC, spearman correlation coefficients for cCRE accessibility and target gene expression.

(L) UMAP projections of TF-cCRE correlations for select TFs shared between different cCRE module. Spearman correlation coefficients between TF expression and cCRE accessibility are shown.

(M, N) UMAP projections of predicted TF-TF interactions. Green dots, cCREs bound by background TF; red dots, cCREs bound by test TF; yellow dots, cCREs co-bound by both TFs; dark grey dots, cCRE module(s) with predicted TF interaction(s).

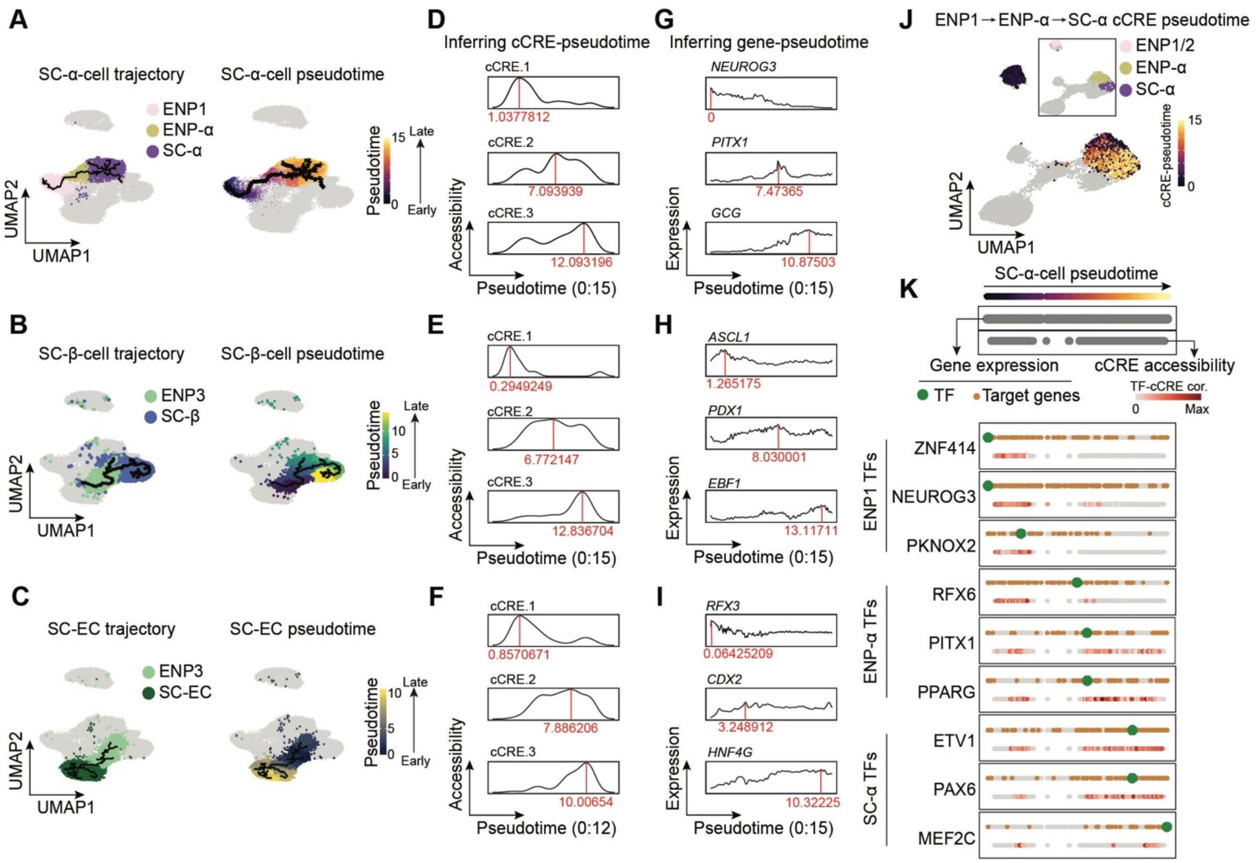

Figure S3. Pseudotime ordering of lineage-specific transcriptional programs - related to Figure 3.

(A-C) Feature plots showing cell cluster identity (left) and pseudotime values (right) on snATAC-seq UMAPs. (A) SC-α-cell trajectory (ENP1/ENP-α/SC-α-cells), (B) SC-β-cell trajectory (ENP3/SC-β-cells) and (C) SC-EC trajectory (ENP3/SC-ECs) are shown.

(D-F) Representative illustration of inference of cCRE accessibility in pseudotime on SC-α-cell (D), SC-β-cell (E) and SC-EC (F) trajectories. Values in red represent cCRE pseudotime values based on maximal cCRE accessibility along the trajectory.

(G-I) Representative illustration of inference of gene expression in pseudotime on SC-α-cell (G), SC-β-cell (H) and SC-EC (I) trajectories. Values in red represent gene pseudotime values based on maximal gene expression along the trajectory.

(J) UMAP projections of cCRE pseudotime on SC-α-cell lineage trajectory. Insets show cell type annotations of cCRE modules.

(K) Pseudotime ordering of transcriptional programs along SC-α-cell lineage trajectory from ENP1 and ENP-α progenitors. Gene expression and cCRE accessibility were assigned pseudotime values and plotted in two separate dotted lines (genes, top; cCREs, bottom). For each shown TF, the TF (green), TF-bound cCREs (colored based TF-cCRE correlations) and target genes (brown) are shown.

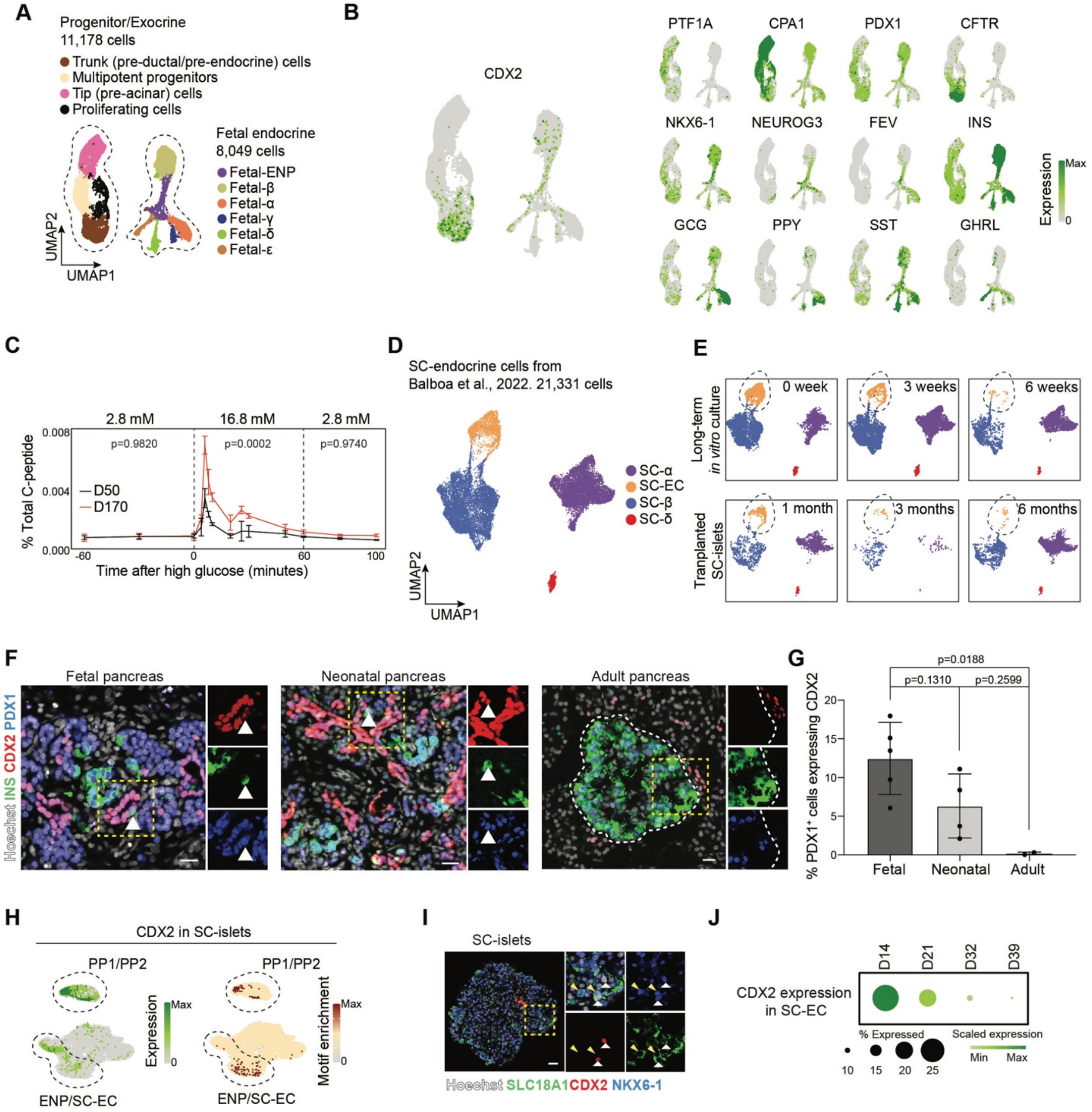
A
Progenitor/Exocrine
11,178 cells
● Trunk (pre-ductal/pre-endocrine) cells
● Multipotent progenitors
● Tip (pre-acinar) cells
● Proliferating cells
Fetal endocrine
8,049 cells
● Fetal-ENP
● Fetal-β
● Fetal-α
● Fetal-γ
● Fetal-δ
● Fetal-ε
UMAP2
UMAP1
B
CDX2
PTF1A
CPA1
PDX1
CFTR
NKX6-1
NEUROG3
FEV
INS
GCG
PPY
SST
GHRL
Expression
Max
0
C
2.8 mM
16.8 mM
2.8 mM
p=0.9820
p=0.0002
p=0.9740
% Total C-peptide
0.008
0.004
0.000
D50
D170
-60
0
60
100
Time after high glucose (minutes)
D
SC-endocrine cells from
Balboa et al., 2022. 21,331 cells
● SC-α
● SC-EC
● SC-β
● SC-δ
UMAP2
UMAP1
E
Long-term
in vitro culture
0 week
3 weeks
6 weeks
Tranplanted
SC-islets
1 month
3 months
6 months
F
Fetal pancreas
Neonatal pancreas
Adult pancreas
Hoechst INS CDX2 PDX1
G
p=0.0188
p=0.1310
p=0.2599
% PDX1+ cells expressing CDX2
20
15
10
5
0
Fetal
Neonatal
Adult
H
CDX2 in SC-islets
PP1/PP2
PP1/PP2
ENP/SC-EC
ENP/SC-EC
Expression
Max
0
Motif enrichment
Max
0
I
SC-islets
Hoechst SLC18A1 CDX2 NKX6-1
J
D14
D21
D32
D39
CDX2 expression
in SC-EC
% Expressed
10
15
20
25
Scaled expression
Min
Max

Figure S4. Identification of CDX2-expressing cells in human fetal pancreas - related to Figure 4.

(A) UMAP embedding of transcriptome data from fetal human pancreas. Cluster identities were defined by expression of marker genes.

(B) Expression of *CDX2* and cell type marker genes in fetal human pancreas.

(C) Human C-peptide secretion by SC-islets at D50 and D170 during perifusion with the indicated glucose concentrations (in mM). Data are shown as mean ± S.D. at the indicated time points. P-values were calculated using two-way ANOVA followed by Šidák-Holm's multiple comparisons test for different days of differentiation in each time block.

(D) UMAP embedding of single cell transcriptome data from [7]. Clusters were defined based on marker gene expression.

(E) Split UMAPs showing localization of each cell type at indicated time points after the beginning of stage 7 (top) or transplantation (bottom).

(F) Representative immunofluorescent images for CDX2, PDX1, and insulin (INS) on human pancreas at 10 weeks post conception (wpc), postnatal day 1, and 52 years of age. Arrowheads in insets indicate CDX2 and INS co-positive cells. Nuclei were labeled with Hoechst. Scale bar, 20 µm.

(G) Quantification of PDX1$^+$ cells expressing CDX2 in fetal (10-20 wpc, n = 5), neonatal (1 day to 3.7 months postnatally, n = 4), and adult (20-52 years, n = 2) human pancreas. Data are shown as mean ± S.D. P-values were calculated using Tukey's multiple comparisons test after one-way ANOVA.

(H) *CDX2* expression (left) and CDX2 motif enrichment (right) in cell types during SC-islet differentiation. Dash circles highlight populations with high *CDX2* expression and transcriptional activity.

(I) Representative immunofluorescent image for CDX2, NKX6-1, and SLC18A1 on SC-islets at D21. Arrowheads in insets indicate CDX2, NKX6-1 and SLC18A1 co-positive cells. Nuclei were labeled with Hoechst. Scale bar, 20 µm.

(J) Dot plots showing average CDX2 expression in SC-ECs at different time points of differentiation. The color of each dot represents the expression level and the size of each dot the percentage of positive cells for each gene.

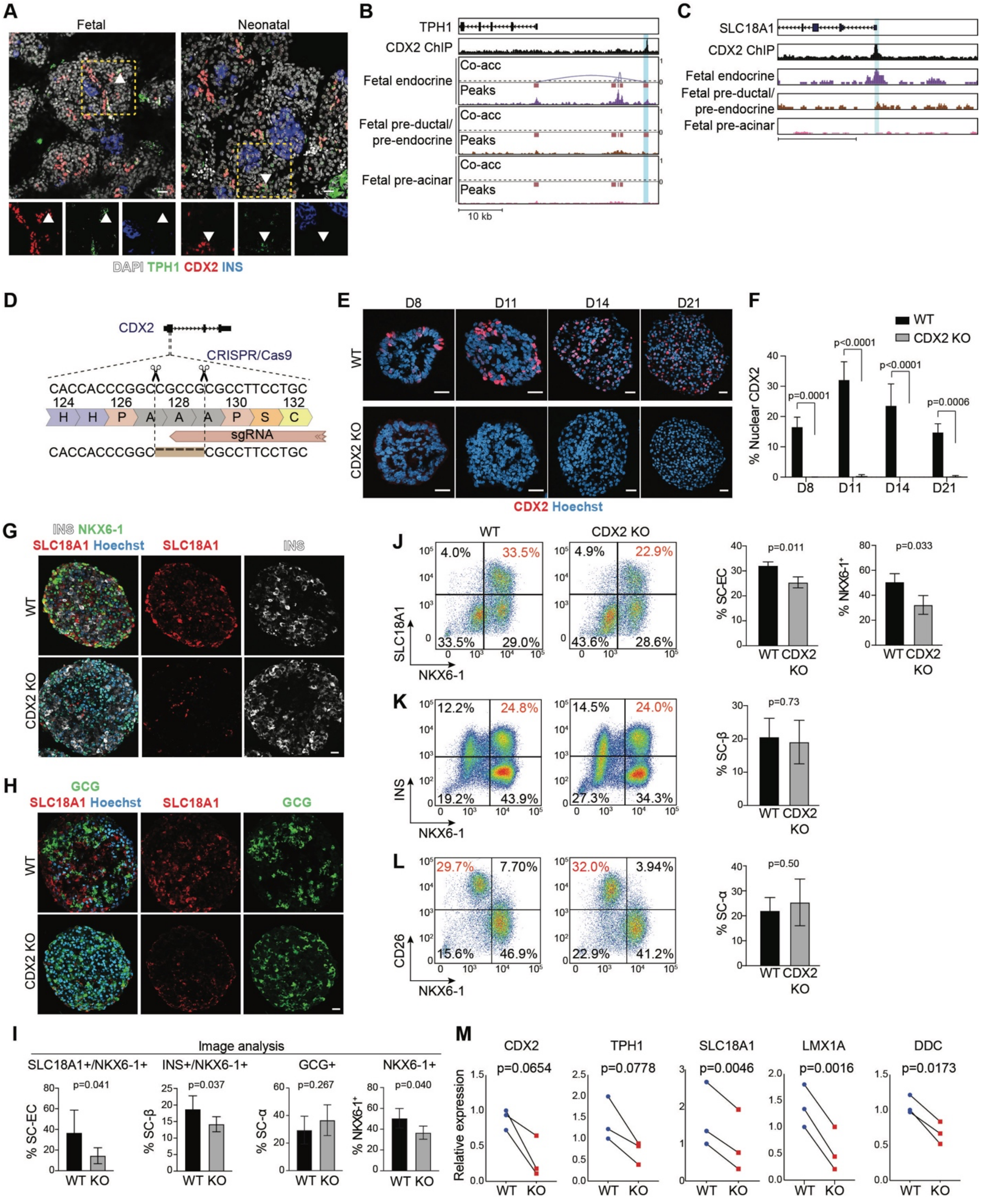
A
Fetal
Neonatal
DAPI TPH1 CDX2 INS
B
TPH1
CDX2 ChIP
Fetal endocrine
Co-acc
Peaks
Fetal pre-ductal/ pre-endocrine
Co-acc
Peaks
Fetal pre-acinar
Co-acc
Peaks
1
0
10 kb
C
SLC18A1
CDX2 ChIP
Fetal endocrine
Fetal pre-ductal/ pre-endocrine
Fetal pre-acinar
D
CDX2
CRISPR/Cas9
CACCACCCGGCCGCCGCGCCTTCCTGC
124
126
128
130
132
H H P A A A P S C
sgRNA
CACCACCCGGC CGCCTTCCTGC
E
D8
D11
D14
D21
WT
CDX2 KO
CDX2 Hoechst
F
WT
CDX2 KO
% Nuclear CDX2
40
30
20
10
0
p=0.0001
p<0.0001
p<0.0001
p=0.0006
D8
D11
D14
D21
G
INS NKX6-1
SLC18A1 Hoechst
SLC18A1
INS
WT
CDX2 KO
H
GCG
SLC18A1 Hoechst
SLC18A1
GCG
WT
CDX2 KO
J
WT
CDX2 KO
4.0%
33.5%
33.5%
29.0%
4.9%
22.9%
43.6%
28.6%
SLC18A1
NKX6-1
p=0.011
% SC-EC
p=0.033
% NKX6-1+
WT
CDX2 KO
K
12.2%
24.8%
19.2%
43.9%
14.5%
24.0%
27.3%
34.3%
INS
NKX6-1
p=0.73
% SC-β
WT
CDX2 KO
L
29.7%
7.70%
15.6%
46.9%
32.0%
3.94%
22.9%
41.2%
CD26
NKX6-1
p=0.50
% SC-α
WT
CDX2 KO
I
Image analysis
SLC18A1+/NKX6-1+
INS+/NKX6-1+
GCG+
NKX6-1+
p=0.041
% SC-EC
p=0.037
% SC-β
p=0.267
% SC-α
p=0.040
% NKX6-1+
WT KO
M
CDX2
p=0.0654
TPH1
p=0.0778
SLC18A1
p=0.0046
LMX1A
p=0.0016
DDC
p=0.0173
Relative expression
WT KO

Figure S5. *CDX2* inactivation in human pluripotent stem cells - related to Figure 5.

(A) Representative immunofluorescent images for TPH1, CDX2, and insulin (INS) on human pancreas at indicated development stages. Nuclei were labeled with DAPI. Arrowheads in insets indicate CDX2 and TPH1 co-positive cells. Scale bar, 20 μm.

(B,C) Genome browser tracks showing CDX2 ChIP-seq reads in SC-islets and aggregated ATAC reads in human fetal pancreatic endocrine, ductal and acinar cells at *TPH1* (B) and *SLC18A1* (C) gene loci. CDX2-bound cCREs are highlighted. All tracks are scaled to uniform 1x10^6 read depth. Co-acc, co-accessibility scores between distal cCRE and transcriptional start site of *TPH1*.

(D) Schematic showing generation of *CDX2* knockout (KO) H1 hPSCs. A 5-base-pair deletion was introduced into the first exon of *CDX2*, leading to a frameshift and premature translation termination.

(E) Representative immunofluorescent images for CDX2 from wild type (WT) and *CDX2* KO cell aggregates at different days (D) of SC-islet differentiation. Nuclei were labeled with Hoechst. Scale bar, 20 μm.

(F) Quantification of CDX2$^+$ cells in WT and *CDX2* KO cell aggregates at different days of SC-islet differentiation. Data are shown as mean ± S.D. P-values were calculated using two-way ANOVA followed by Šidák-Holm's multiple comparisons test for different days of differentiation.

(G) Representative immunofluorescent images for insulin (INS), NKX6-1 and SLC18A1 on WT and *CDX2* KO SC-islets at D21. Nuclei were labeled with Hoechst. Scale bar, 20 μm.

(H) Representative immunofluorescent images for glucagon (GCG) and SLC18A1 on WT and *CDX2* KO SC-islets at D21. Nuclei were labeled with Hoechst. Scale bar, 20 μm.

(I) Quantification of the percentage of SC-ECs (SLC18A1$^+$/NKX6-1$^+$), SC-β-cells (INS$^+$/NKX6-1$^+$), SC-α-cells (GCG$^+$) and total NKX6-1$^+$ cells in SC-islets based on immunofluorescent staining in (G) and (H). Data are shown as mean ± S.D. (n = 3 independent differentiations). P-values were calculated by unpaired two-tailed t-test.

(J-L) Representative flow cytometry plots (left, percentage of population of interest in red) and quantifications (right) of SC-ECs (NKX6-1+/SLC18A1+), total NKX6-1$^+$ cells (J), SC-β-cells (NKX6-1+/INS+, K), and SC-α-cells (NKX6-1-/CD26+, L) in WT and CDX2 KO SC-islets at D21. Data are shown as mean ± S.D. (n = 3 independent differentiations). P-values were calculated by unpaired two-tailed t-test.

(M) qPCR analysis of *CDX2* and serotonin synthesis genes in WT and *CDX2* KO SC-islets at D21. n = 100 SC-islets per group from 3 independent differentiations. P-values were calculated by two-tailed paired t-test.

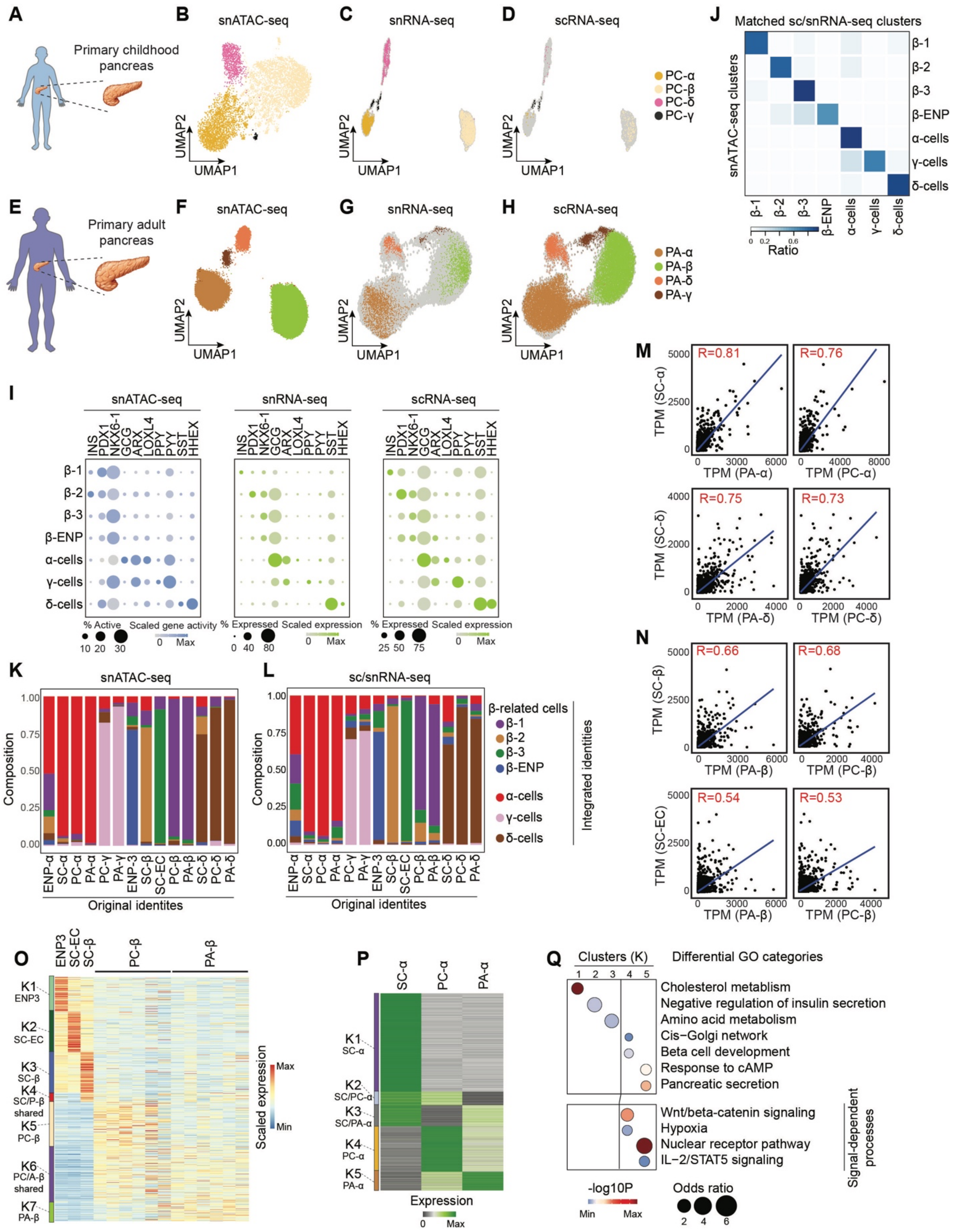
A
Primary childhood pancreas
B
snATAC-seq
C
snRNA-seq
D
scRNA-seq
UMAP1
UMAP2
PC-α
PC-β
PC-δ
PC-γ
J
Matched sc/snRNA-seq clusters
snATAC-seq clusters
β-1
β-2
β-3
β-ENP
α-cells
γ-cells
δ-cells
0 0.2 0.6
Ratio
E
Primary adult pancreas
F
snATAC-seq
G
snRNA-seq
H
scRNA-seq
PA-α
PA-β
PA-δ
PA-γ
I
snATAC-seq
snRNA-seq
scRNA-seq
INS
PDX1
NKX6-1
GCG
ARX
LOXL4
PPY
PYY
SST
HHEX
β-1
β-2
β-3
β-ENP
α-cells
γ-cells
δ-cells
% Active
10 20 30
Scaled gene activity
0 Max
% Expressed
0 40 80
Scaled expression
0 Max
% Expressed
25 50 75
Scaled expression
0 Max
M
R=0.81
R=0.76
TPM (SC-α)
TPM (PA-α)
TPM (PC-α)
R=0.75
R=0.73
TPM (SC-δ)
TPM (PA-δ)
TPM (PC-δ)
N
R=0.66
R=0.68
TPM (SC-β)
TPM (PA-β)
TPM (PC-β)
R=0.54
R=0.53
TPM (SC-EC)
TPM (PA-β)
TPM (PC-β)
K
snATAC-seq
Composition
1.00
0.75
0.50
0.25
0.00
ENP-α
SC-α
PC-α
PA-α
PC-γ
PA-γ
ENP-3
SC-β
SC-EC
PC-β
PA-β
SC-δ
PC-δ
PA-δ
Original identites
L
sc/snRNA-seq
Original identites
β-related cells
β-1
β-2
β-3
β-ENP
α-cells
γ-cells
δ-cells
Integrated identities
O
ENP3
SC-EC
SC-β
PC-β
PA-β
K1 ENP3
K2 SC-EC
K3 SC-β
K4 SC/P-β shared
K5 PC-β
K6 PC/A-β shared
K7 PA-β
Scaled expression
Max
Min
P
SC-α
PC-α
PA-α
K1 SC-α
K2 SC/PC-α
K3 SC/PA-α
K4 PC-α
K5 PA-α
Expression
0 Max
Q
Clusters (K)
1 2 3 4 5
Differential GO categories
Cholesterol metablism
Negative regulation of insulin secretion
Amino acid metabolism
Cis−Golgi network
Beta cell development
Response to cAMP
Pancreatic secretion
Wnt/beta-catenin signaling
Hypoxia
Nuclear receptor pathway
IL−2/STAT5 signaling
Signal-dependent processes
-log10P
Min Max
Odds ratio
2 4 6

Figure S6. Integration of chromatin accessibility and transcriptome data from stem cell islets and primary pancreas - related to Figure 6.

(A) Schematic showing source of cells used in single cell analyses of primary childhood human pancreas.

(B-D) UMAP embedding of chromatin accessibility (B) and transcriptome (C and D) data from isolated nuclei (C) or whole cells (D) from childhood pancreas. Cluster identities were defined by promoter accessibility (snATAC-seq) or expression (sc/snRNA-seq) of marker genes.

(E) Schematic showing source of cells used in single cell analyses of primary adult human pancreas.

(F-H) UMAP embedding of chromatin accessibility (F) and transcriptome (G and H) data from isolated nuclei (G) or whole cells (H) from adult pancreas. Cluster identities were defined by promoter accessibility (snATAC-seq) or expression (sc/snRNA-seq) of marker genes.

(I) Dot plots showing scaled average gene activity (left) or gene expression (middle and right). The color of each dot represents the average gene activity or expression level and the size of each dot the percentage of positive cells for each gene.

(J) Heatmap showing ratio of cells with identities in sc/snRNA-seq (column) data matching identities in snATAC-seq (row) data.

(K, L) Comparison of integrated cell type annotations based on snATAC-seq (K) and scRNA-seq (L) UMAP in Figure 6B,C to cell type annotations before data integration.

(M,N) Scatter plots showing normalized gene expression (TPM) in indicated cell types. Dots denote individual genes. R scores indicate Spearman correlation between two cell types.

(O) K-means clustering of genes with variable expression across β-related cell types (ENP3, SC-ECs, SC-β-cells, PC-β-cells, and PA-β-cells). Clusters were annotated and color-coded based on gene expression patterns. PC- and PA-β-cells from individual donor were plotted.

(P) K-means clustering of genes with variable expression across α-related cell types (SC-α-cells, PC-α-cells, and PA-α-cells). Clusters were annotated and color-coded based on gene expression patterns.

(Q) Enriched gene ontology terms/pathways in each cluster. Significance (-log10 p-value) and odds ratio of the enrichments are represented by color and dot size, respectively.

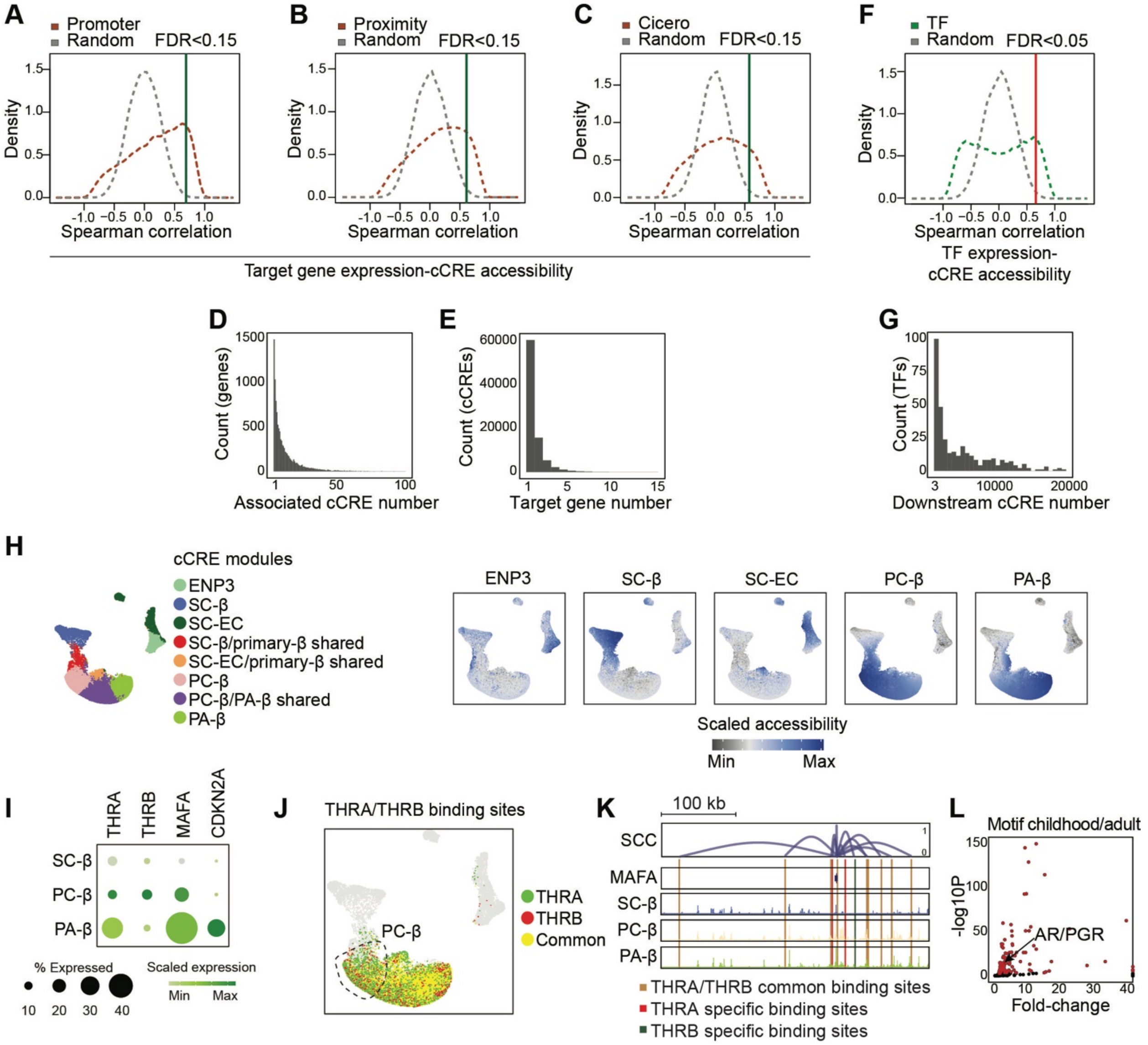
A
Promoter
Random
FDR<0.15
Density
Spearman correlation
B
Proximity
Random
FDR<0.15
Density
Spearman correlation
C
Cicero
Random
FDR<0.15
Density
Spearman correlation
Target gene expression-cCRE accessibility
F
TF
Random
FDR<0.05
Density
Spearman correlation
TF expression-
cCRE accessibility
D
Count (genes)
Associated cCRE number
E
Count (cCREs)
Target gene number
G
Count (TFs)
Downstream cCRE number
H
cCRE modules
ENP3
SC-β
SC-EC
SC-β/primary-β shared
SC-EC/primary-β shared
PC-β
PC-β/PA-β shared
PA-β
ENP3
SC-β
SC-EC
PC-β
PA-β
Scaled accessibility
Min
Max
I
THRA
THRB
MAFA
CDKN2A
SC-β
PC-β
PA-β
% Expressed
10
20
30
40
Scaled expression
Min
Max
J
THRA/THRB binding sites
PC-β
THRA
THRB
Common
K
100 kb
SCC
MAFA
SC-β
PC-β
PA-β
THRA/THRB common binding sites
THRA specific binding sites
THRB specific binding sites
L
Motif childhood/adult
-log10P
AR/PGR
Fold-change

Figure S7. Building a gene regulatory network of β-cell maturation - related to Figure 7.

(A-C) Identification of putative cCRE-gene pairs. cCREs were assigned to a gene if they are located (A) in promoter (± 500 bp) of target gene; (B) in proximity (± 50 kb) of the gene promoter; or (C) show co-accessibility with the gene promoter by Cicero analysis. A total of 167,618 positively correlated cCRE-gene pairs were identified using an empirically defined significance threshold of FDR<0.15.

(D, E) Histogram showing characteristics of cCRE-gene pairs. Each target gene in the GRN is regulated by a mean of 13.6 cCREs (D) and each cCRE regulates a mean of 1.55 target genes (E).

(F) Identification of putative TF-cCRE pairs. Putative TF-cCRE pairs were defined using motifmatchr. A total of 1,017,318 significantly correlated TF-cCRE pairs were identified using an empirically defined threshold of FDR<0.05.

(G) Histogram showing characteristics of TF-cCRE pairs. Each TF has a mean of 2,698 predicted binding sites.

(H) UMAP projections of scaled chromatin accessibility of all cCREs in each aggregated pseudo-cell. Pseudo-cells were aggregated based on cell type identities of primary cells and stem cell-derived populations before data integration.

(I) Dot plots showing average *THRA, THRB, MAFA* and *CDKN2A* expression SC-, primary childhood (PC)- and primary adult (PA)-β-cells. The color of each dot represents the expression level and the size of each dot the percentage of positive cells for each gene.

(J) UMAP projections of predicted THRA and THRB binding sites. Green dots, cCREs bound by THRA; red dots, cCREs bound by THRB; yellow dots, cCREs co-bound by both TFs.

(K) Genome browser tracks show aggregated ATAC reads in SC-, PC- and PA-β-cells. All tracks are scaled to uniform 1x10^6 read depth. SCC, spearman correlation coefficients for cCRE accessibility and target gene expression.

(L) Scatter plot showing fold-change and significance (-log10 p-value) of TF motif enrichment at chromatin sites with increased H3K27ac signals in childhood (<9 years) β-cells against a background of sites with increased H3K27ac signal in adult β-cells. Dots denote individual TF motifs.

Oligonucleotide names	**Sequence (5’ to 3’)**
sgRNA oligos	
sgRNA targeting human CDX2	GAGAAGCGCAGGAAGGCGCGG
PCR genotyping primers for CDX2 locus	
gt-CDX2-forward	TCACGGCCTCAACGGTGGCT
gt-CDX2-reverse	GCCTTCCCAAGCACCCTCCGAA
qPCR primers	
LMX1A-f	AGTGTCTACAGCTCAGATCCC
LMX1A-r	GGTCATGGAAAAGGGGCTCG
DDC-f	TGGGGACCACAACATGCTG
DDC-r	TCAGGGCAGATGAATGCACTG
CDX2-f	GACGTGAGCATGTACCCTAGC
CDX2-r	GCGTAGCCATTCCAGTCCT
TPH1-f	ACGTCGAAAGTATTTTGCGGA
TPH1-r	ACGGTTCCCCAGGTCTTAATC
SLC18A1-f	AAGGCTGTGATGCAACTTCTG
SLC18A1-r	CGGGCCACAAAGAGTAGAGTAT

Table S7. Oligonucleotides used in this study, related to STAR Methods Key Resource Table (Oligonucleotides section)