Supplemental information

I. Supplemental Data

Supplemental Figures

Figure S1 - Related to Figure 1

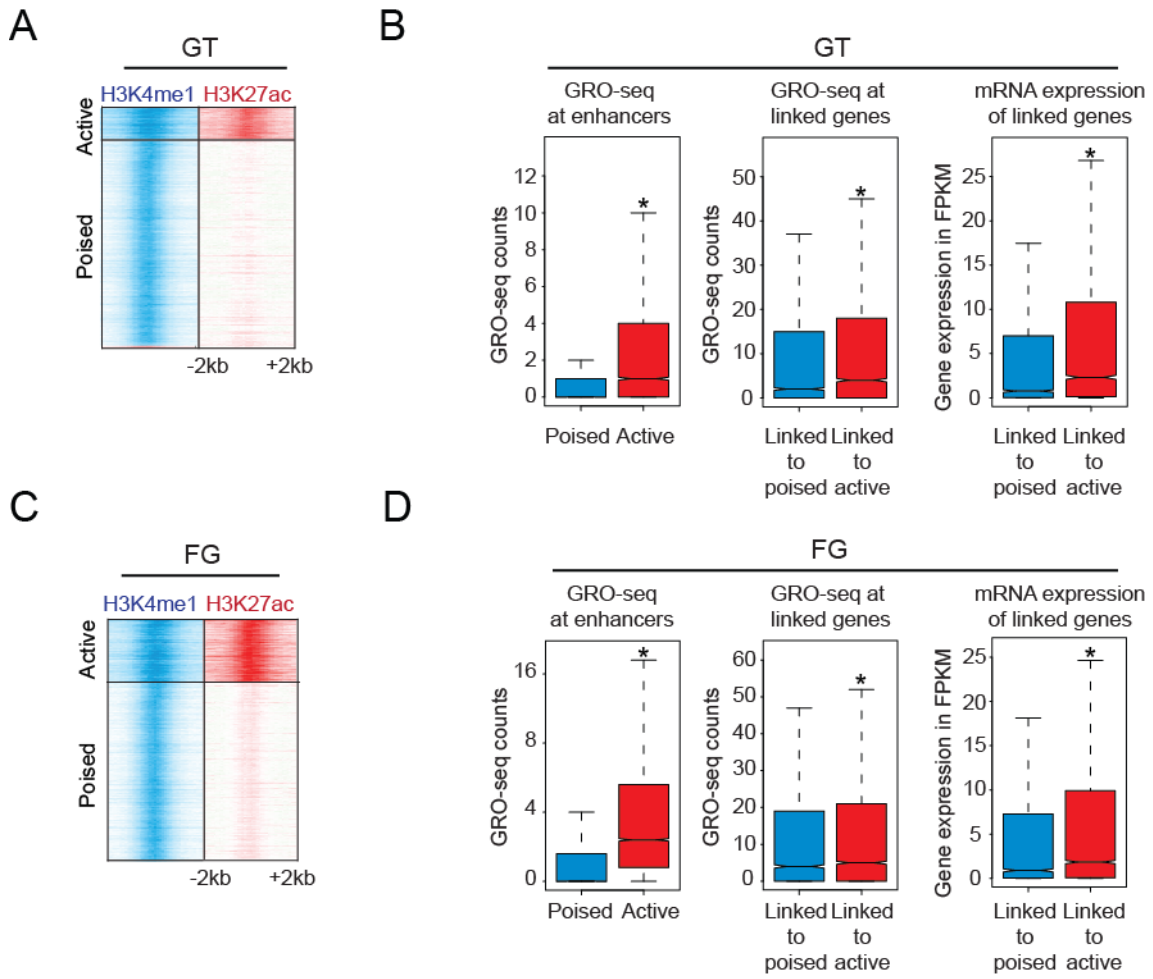

Figure S1 - Related to Figure 1. Global identification and characterization of poised and active enhancers in lineage intermediates during pancreatic differentiation of human embryonic stem cells. (**A, C**) Density of ChIP-seq reads for H3K4me1 and H3K27ac relative to midpoint at putative poised and active enhancers at the primitive gut tube (**A**; GT) and posterior foregut (**C**; FG) stage. (**B, D**) Box plots of GRO-seq counts at poised (H3K4me1 only) and active (H3K4me1 and H3K27ac) enhancers, GRO-seq counts at linked genes of poised and

active enhancers, and mRNA expression in FPKM at linked genes of poised and active enhancers at the GT (**B**) and FG (**D**) stage. * P-values < 2.2e-16, wilcoxon test.

Figure S2 - Related to Figure 2

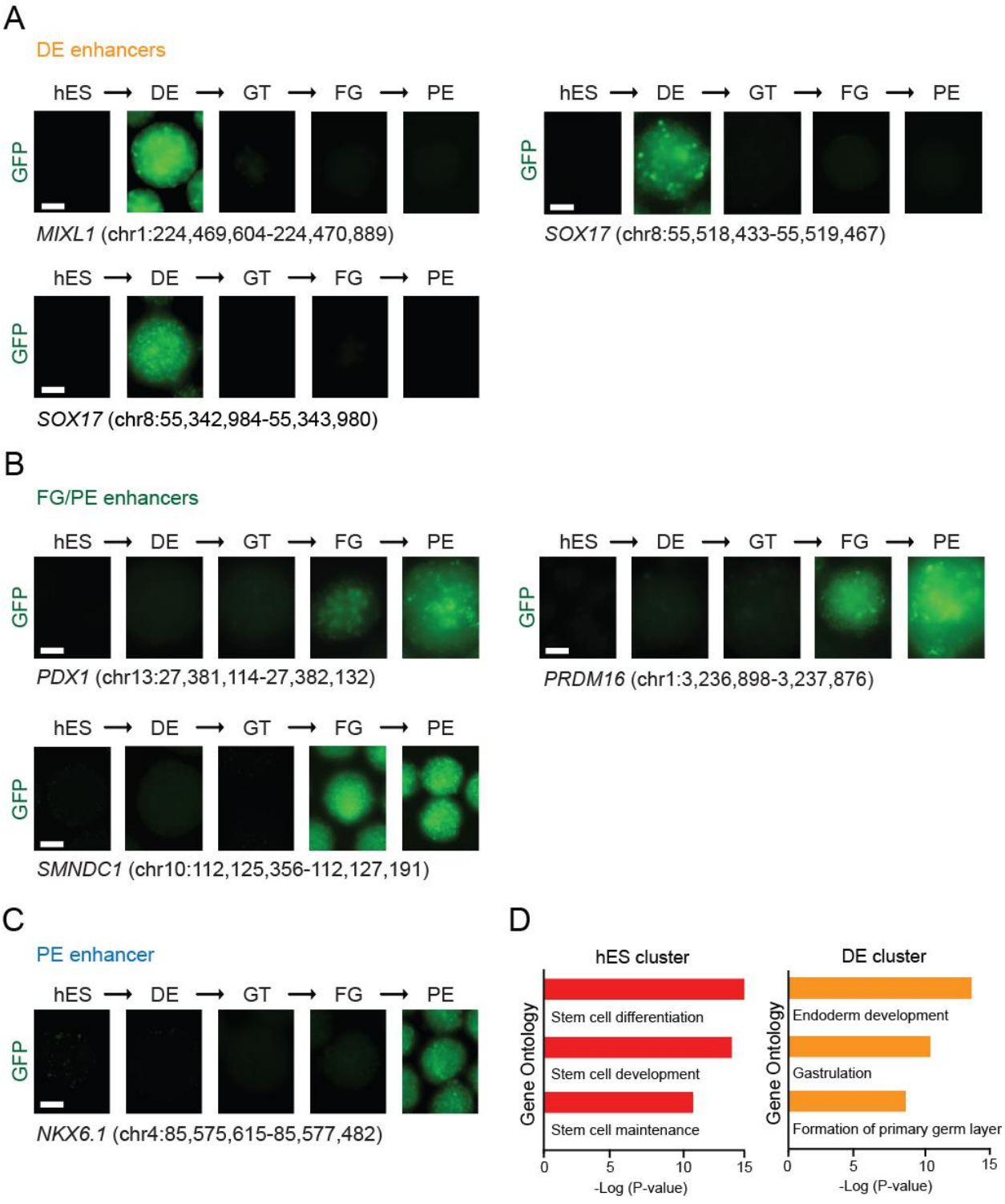

Figure S2 - Related to Figure 2. Enhancer activity coincides with H3K27ac deposition. (**A-C**) *In vitro* GFP reporter assays for representative DE-specific (**A**) FG/PE-specific (**B**) and PE-specific (**C**) enhancers. GFP images of cell aggregates during pancreatic differentiation are shown. Scale bars, 100μm. (**D**) Enriched Gene Ontology terms for hES cell-specific and DE-specific enhancers. hES, human embryonic stem cells; DE, definitive endoderm; GT, primitive gut tube; FG, posterior foregut; PE, pancreatic endoderm.

Figure S3 - Related to Figure 3

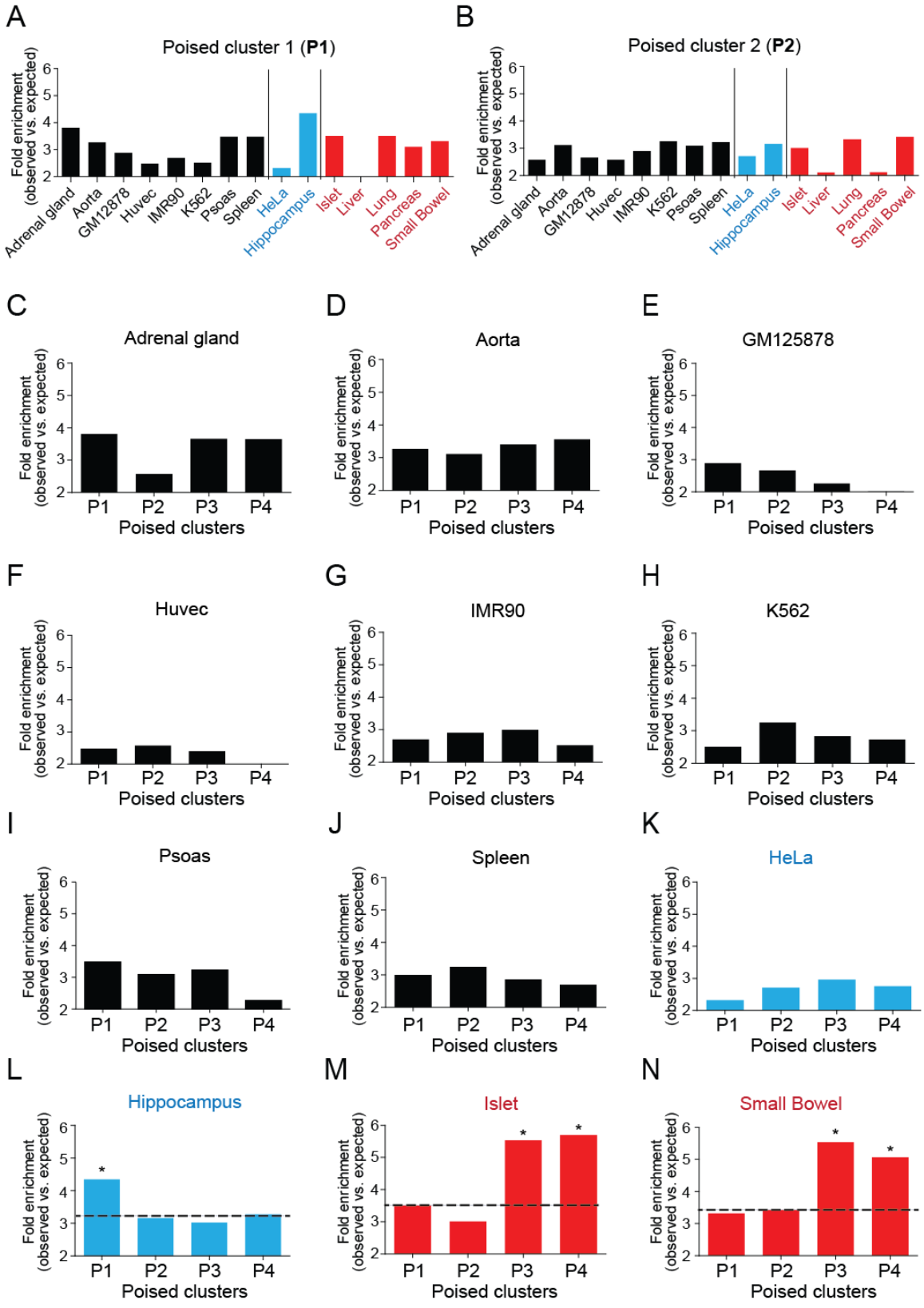

Figure S3 - Related to Figure 3. Analysis of active enhancers identified from the ENCODE and Roadmap Epigenomics projects. (**A, B**) Enrichment of tissue- or cell type-specific H3K27ac peaks in poised enhancer clusters P1 (**A**) and P2 (**B**). Poised enhancer clusters P1-P4 are shown in Figure 2A. (**C-N**) Enrichment of tissue- or cell type-specific H3K27ac peaks in poised enhancer clusters P1, P2, P3 and P4. * P-values < 1e-06, chi-square test.

Figure S4 - Related to Figure 4

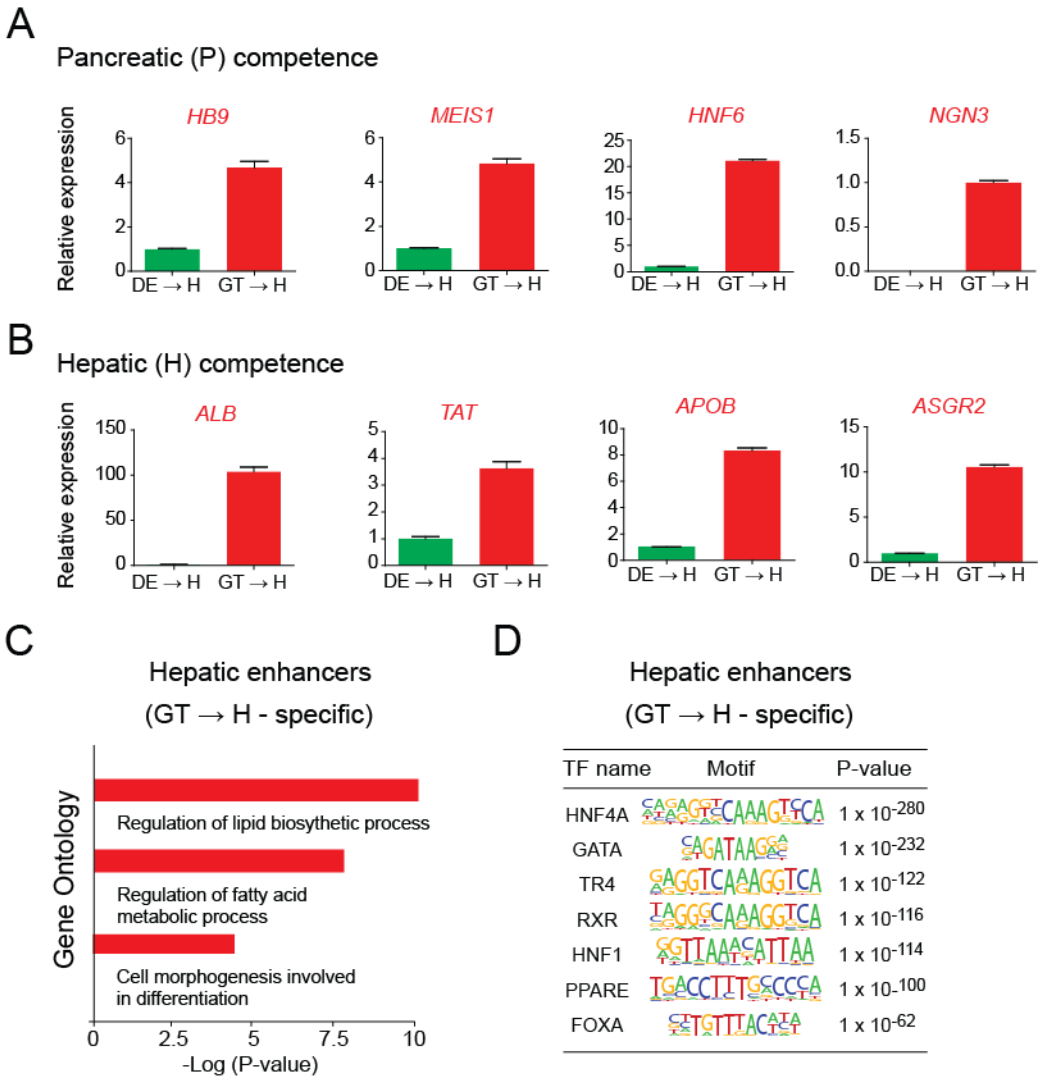

Figure S4 - Related to Figure 4. Pancreatic and hepatic competence is acquired at the gut tube stage. (**A**, **B**) Quantitative reverse transcription PCR analysis for the early pancreas (P) markers *HB9*, *MEIS1*, *HNF6*, and *NGN3* in DE and GT cells treated with retinoic acid, cyclopamine and noggin for 2 days (**A**), and the early liver markers *ALB*, *TAT*, *APOB*, and *ASGR2* in DE and GT cells treated with fibroblast growth factor (FGF) and bone morphogenetic protein (BMP) for 3 days (**B**). Data are shown as average ± S.E.M. (**C**) Enriched Gene Ontology terms in hepatic (H) enhancers (specific H3K27ac peaks in GT cells treated with BMP and FGF

for 3 days; GT-> H cells). (**D**) Enriched transcription factor (TF) binding motifs and associated P-values at hepatic enhancers (GT->H cell-specific H3K27ac peaks). DE, definitive endoderm; GT, primitive gut tube.

Figure S5 - Related to Figure 5

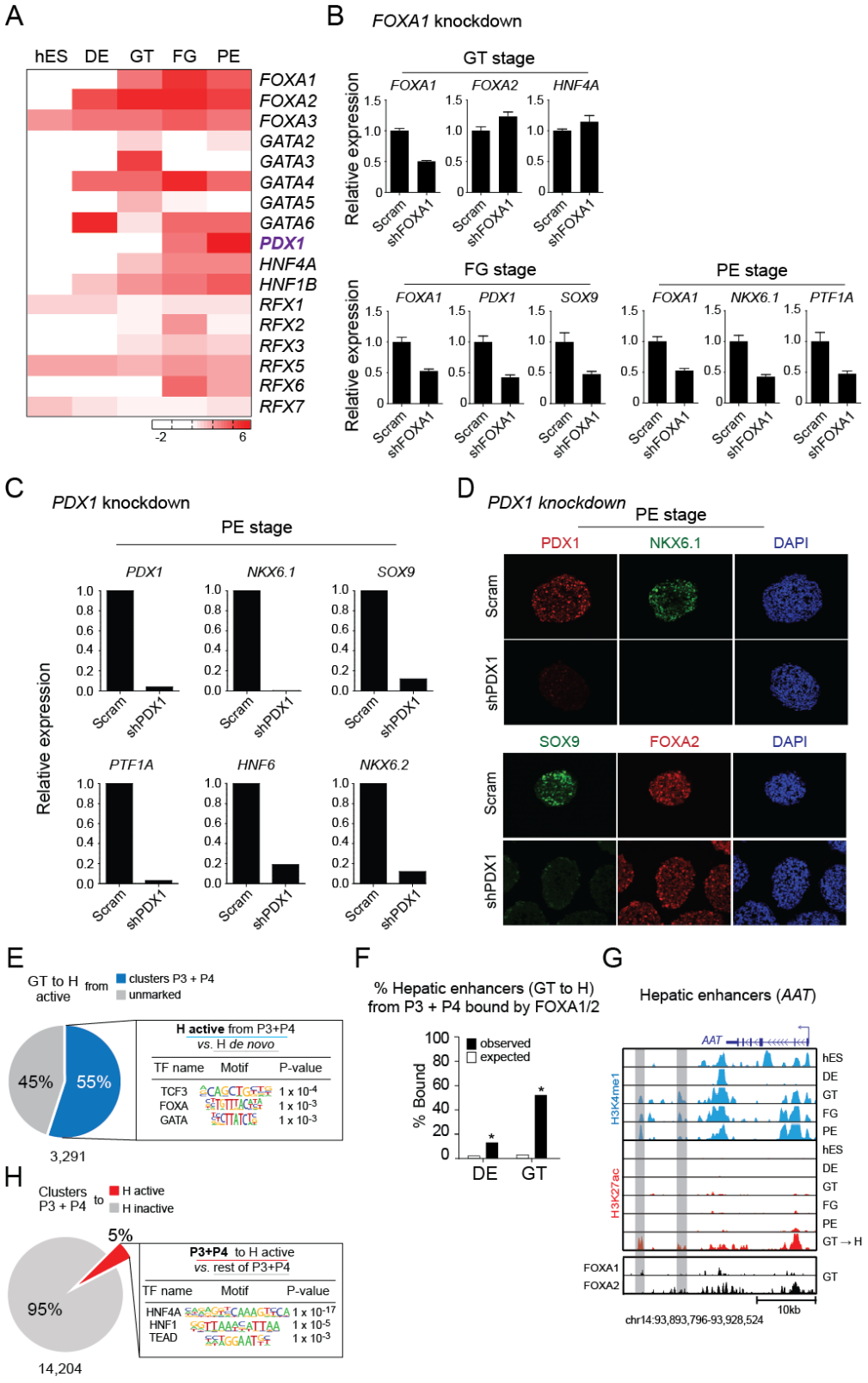

Figure S5 - Related to Figure 5. *FOXA1* and *PDX1* are independently required for pancreatic gene activation and FOXA recruitment precedes enhancer activation. (**A**) Heatmap showing mRNA expression levels, measured in FPKM, of transcription factors (TFs) known to recognize motifs enriched at pancreatic enhancers (FG/PE- and PE-specific clusters). Only TFs with an FPKM > 1.0 are shown. (**B**) Quantitative reverse transcription PCR analysis of the gut tube markers *FOXA1*, *FOXA2* and *HNF4A*, the early pancreatic markers *PDX1* and *SOX9*, and later pancreatic markers *NKX6.1* and *PTF1A* in scrambled control (Scram) and *FOXA1* knockdown (shFOXA1) cells differentiated to the GT, FG or PE stage. Data are shown as average ± S.E.M. (**C**) Quantitative reverse transcription PCR analysis of the pancreatic marker *PDX1*, *NKX6.1*, *SOX9*, *PTF1A*, *HNF6*, and *NKX6.2* in scrambled control (Scram) and *PDX1* knockdown (shPDX1) cells differentiated to PE. Data are shown as average ± S.E.M. (**D**) Immunofluorescence staining for PDX1, NKX6.1, SOX9, and FOXA2 in scrambled control (Scram) and *PDX1* knockdown (shPDX1) cells differentiated to PE. (**E**) Enriched TF binding motifs with associated P-values for the 55% of active hepatic enhancers (specific H3K27ac peaks in GT cells treated with BMP and FGF for 3 days; GT to H cells) that overlap with poised enhancers from clusters P3 and P4 *versus* the remaining 45% of active hepatic enhancers. (**F**) Percentage of hepatic enhancers (GT to H-specific) from poised enhancer clusters P3 and P4 *versus* random genomic regions bound by FOXA1 or FOXA2 at the DE or GT stage. * P-values < 2.2e-16, chi-square test. (**G**) H3K27ac, H3K4me1, FOXA1, and FOXA2 ChIP-seq profiles of enhancers near the hepatic gene *alpha 1-antitrypsin* (*AAT*). (**H**) Enriched TF binding motifs with associated P-values for the 5% of poised enhancers in enhancer clusters P3 and P4 that acquire H3K27ac in hepatic cells (GT to H cells) *versus* the remaining 95% of poised enhancers in

clusters P3 and P4. hES, human embryonic stem cells; DE, definitive endoderm; GT, primitive gut tube; FG, posterior foregut; PE, pancreatic endoderm.

Figure S6 - Related to Figure 6

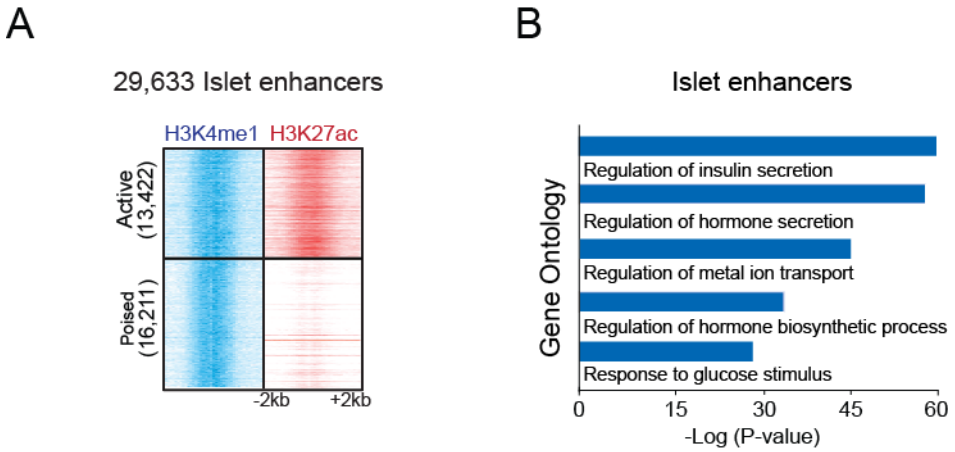

Figure S6 - Related to Figure 6. Identification and characterization of islet enhancers. (**A**) Density of ChIP-seq reads for H3K4me1 and H3K27ac relative to midpoint at putative poised and active enhancers in islets. (**B**) Gene Ontology terms enriched in active enhancers in islets.

Supplemental Tables

Table S1 - Related to Figure 1. Putative enhancers identified during pancreatic lineage progression. (**A**) hES cell poised enhancers. (**B**) hES cell active enhancers. (**C**) DE poised enhancers. (**D**) DE active enhancers. (**E**) GT poised enhancers. (**F**) GT active enhancers. (**G**) FG poised enhancers. (**H**) FG active enhancers. (**I**) PE poised enhancers. (**J**) PE active enhancers.

Table S2 - Related to Figure 3. Gene Ontology terms for cell-type specific H3K27ac peaks from Roadmap Epigenomics project and ENCODE data. (**A**) Adrenal gland. (**B**) Aorta. (**C**) GM12878 cells. (**D**) Huvec cells. (**E**) IMR90 cells. (**F**) K562 cells. (**G**) Psoas. (**H**) Spleen. (**I**) HeLa cells. (**J**) Hippocampus. (**K**) Islet. (**L**) Liver. (**M**) Lung. (**N**) Pancreas. (**O**) Small bowel.

Table S3 - Related to Figure 4. Putative enhancers specific to hepatic cells.

Table S4 - Related to Figure 5. PDX1 ChIP-seq peaks.

Table S5 - Related to Figures 5 and 6. FOXA1 ChIP-seq peaks in (A) GT, **(B)** FG and **(C)** PE.

Table S6 - Related to Figures 5 and 6. FOXA2 ChIP-seq peaks in (A) DE, **(B)** GT, **(C)** FG, and **(D)** PE**.**

Table S7 - Related to Figure 6. Putative enhancers identified in primary human islets. (**A**) active and (**B**) poised.

II. Supplemental Experimental Procedures

hES Cell Culture

CyT49 human embryonic stem cells (NIH registration number: 0041) were maintained as follows. Propagation of hESC was carried out by passing cells every 3 to 4 days using Accutase™ (Innovative Cell Technologies) for enzymatic cell dissociation and with 10% (vol/vol) human AB serum (Valley Biomedical) included in the hESC media the day of passage. Seeding of hESC into tissue culture flasks was done at a density of 50,000 cells/cm2 for 3 days growth. hESC media was comprised of DMEM/F12 (Mediatech or Life Technologies) supplemented with 10% (vol/vol) KnockOut™ Serum Replacement XenoFree (Life Technologies), 0.1 mM MEM non-essential amino acids (Life Technologies), 1X GlutaMAX™ I (Life Technologies), 1% (vol/vol) penicillin/streptomycin (Life Technologies), 0.1mM 2-mercaptoethanol (Life Technologies), 10 ng/ml Activin A (R&D Systems), and 10 ng/ml Heregulin-β1 (PeproTech).

Pancreatic Differentiation

Pancreatic differentiation was performed as previously described (Schulz et al., 2012). Briefly, we employed a suspension-based format using rotational culture. Aggregates of undifferentiated hESCs were generated by re-suspending dissociated cells in hESC media at 1×10^6 cells/mL and culturing them overnight in six-well ultra-low attachment plates (Costar) with 5.5ml per well on an orbital rotator (Innova2000, New Brunswick Scientific) at 95 rpm. The following day undifferentiated aggregates were washed in RPMI media and then differentiated using a multi-step protocol with daily media feeding except on day 10 and continued orbital rotation at either

95 rpm or at 105 rpm on days 4 to 8. In addition to GlutaMAX™ and penicillin/streptomycin, RPMI media (Mediatech) was supplemented with 0.2% (vol/vol) FBS (HyClone) and DMEM High Glucose media (HyClone) was supplemented with 0.5X B-27® supplement (Life Technologies). Human activin, mouse Wnt3a, human KGF, human Noggin, and human EGF were purchased from R&D systems. Other media components included TGFb R1 kinase inhibitor IV (EMD Bioscience), KAAD-Cyclopamine (Toronto Research Chemicals), the retinoid analog TTNPB (Sigma Aldrich), and Insulin-Transferrin-Selenium (ITS; Life Technologies).

Day 0 – Day 1: RPMI/FBS, 100ng/mL Activin, 50ng/mL mouse Wnt3a, 1:5000 ITS

Day 1 – Day 2: RPMI/FBS, 100ng/mL Activin, 1:5000 ITS

Day 2 – Day 3: RPMI/FBS, 2.5 mM TGFb R1 kinase inhibitor IV, 25ng/mL KGF, 1:1000 ITS

Day 3 – Day 5: RPMI/FBS, 25ng/mL KGF, 1:1000 ITS

Day 5 – Day 8: DMEM/B27, 3nM TTNPB (RA agonist), 0.25mM Cyclopamine, 50ng/mL Noggin

Day 8 – Day 12: DMEM/B27, 50ng/mL KGF, 50ng/mL EGF

Hepatic and Lung Induction

Hepatic and lung induction of hESC-derived endodermal intermediates was performed employing the same culture conditions as described above for pancreatic differentiation with minor modifications. For hepatic induction, each developmental intermediate was treated with 50ng/ml BMP4 (Millipore) and 10ng/ml FGF2 (Millipore) in RPMI media (Mediatech) supplemented with 0.2% (vol/vol) FBS (HyClone) for 3 days with daily feeding. For lung

induction, each developmental intermediate was treated for 3 days with a previously described growth factor combination for lung induction (Huang et al., 2014). Specifically, cells were treated with 3µM CHIR99021 (Stemgent), 10ng/ml human FGF10 (R&D), 10ng/ml human KGF ng/ml (R&D), 10ng/ml BMP4 (Millipore), 20ng/ml human EGF (R&D), and 0.05µM all-trans retinoic acid (Sigma) in RPMI media (Mediatech) supplemented with 0.2% (vol/vol) FBS (HyClone) with daily feeding.

Human Islets

Human islets were obtained from the integrated Islet Distribution Program (IIDP), donor number 790. Islets were further enriched using zinc-dithizone.

ChIP-seq and Data Analysis

ChIP-seq was performed as previously described (Hawkins et al., 2010) using 100 µg of chromatin from each cell stage. 5 µg of antibody (H3K4me1, Abcam ab8895; H3K27ac, Activemotif 39133; FOXA1, Abcam ab5089; FOXA2, Santa Cruz, sc-6554) or 15µl (PDX1, BCBC goat anti-PDX1) were used for each respective ChIP-seq assay. ChIP and input libraries were prepared as previously described following Illumina protocols (Hawkins et al., 2010). Previously published ChIP-seq data for H3K4me3 and H3K4me1 (Bhandare et al., 2010; Xie et al., 2013a) were realigned and analyzed in a similar fashion as obtained H3K4me1 and H3K27ac data. Each ChIP-seq analysis was performed in at least two biological replicates. The correlation between biological replicates are as follows:

Pearson correlation of biological replicates

Sample	H3K27ac	H3K4me1
hES	0.93	0.91
DE	0.81	0.93
GT	0.69	0.91
FG	0.87	0.91
PE	0.85	0.93

Tissue and cell type-specific enhancers in Figure 3E were determined based on H3K27ac ChIP-seq data from both the Roadmap Epigenomics and ENCODE projects. Potential enhancers were defined as H3K27ac peaks that are > 3kb away from known TSSs and H3K4me3 peaks. In addition, enhancers within 1kb of each other were merged. Tissue-specific enhancers were defined using a similar strategy as described (Xie et al., 2013b). Briefly, tissue- or cell type-specific enhancers were determined based on the following filtering criteria: H3K27ac must be enriched in only one tissue (quantile normalized RPKM >1).

Enhancers specific to hepatic cells in Figure 4F were defined in a similar fashion. GT stage cells were cultured in 50ng/ml BMP4 and 10ng/ml FGF2 for 3 days (see conditions for hepatic induction) and analyzed by H3K27ac ChIP-seq. Potential enhancers were defined as described above. Hepatic-specific enhancers were then determined by filtering against H3K27ac ChIP-seq data from hES, DE, GT, FG and PE stage cells, using the criteria described above (H3K27ac must be enriched in hepatic cells only; quantile normalized RPKM >1).

GRO-seq and Data Analysis

Global run-on, library preparation and analysis were performed as described (Core et al., 2008; Jin et al., 2013; Wang et al., 2011). 10 million cells from each cell population were used to harvest nuclei. Run-on assays were performed for 5 min at 30 °C. Newly transcribed RNA was labeled with BrdU and then extracted with TRIzol LS reagent (Invitrogen) according to the manufacturer's instructions. The resulting nascent RNA was pulled down using anti-BrdU agarose beads (Santa Cruz Biotech). Reverse transcription was performed using oNTI223 primer (5′-pGATCGTCGGACTGTAGAACTCT;CAAGCAGAAGACGGCATACGATTTTTTTTTTTTT TTTTTTTVN-3′) where the p indicates 5′ phosphorylation, ';' indicates the abasic dSpacer furan and VN indicates degenerate nucleotides. First-strand single-strand cDNA was then circularized (CircLigase, Epicentre) and re-linearized using Ape1 (NEB). Finally, the re-linearized product (~120–300 bp) was recovered and PCR amplified using primers oNTI200 (5′-CAAGCAGAAGACGGCATA-3′) and oNTI201 (5′-AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGACG-3′) to generate the DNA library. Sequencing data were analyzed as previously described (Jin et al., 2013). Note the GRO-seq reads were iteratively aligned to hg18 using Bowtie: the first 36 bases were used in the first round of alignment; and the unmapped reads were then mapped in two additional iterations using the first 30 or 25 bases.

To measure the GRO-seq signal at predicted enhancers, we developed an in-house computational pipeline. For the intergenic enhancers, we use the average of sense and anti-sense GRO-seq around the enhancers. For genic enhancers, we only used the anti-sense reads, with an effort to distinguish the enhancer RNA and RNA. To assess GRO-seq signal at target gene

promoters, we counted GRO-Seq reads for upstream 1kb windows from assigned target transcriptional start sites.

RNA-seq Analysis

FASTQ files for RNA-seq data (Xie et al., 2013a) were mapped to the human genome (hg18) by TopHat software version (Trapnell et al., 2009). We assigned expression value (FPKM) for each RefSeq gene with Cufflinks (Trapnell and Salzberg, 2009). To normalize gene expression levels between different samples, we used the quantile normalization function in R.

Assignment of Enhancer Target Genes

Putative enhancers were assigned to target genes based on overlap of the predicted enhancers with a published map of distal DNaseI hypersensitivity (DHS)-to-promoter connections covering 578,905 distal regulatory elements across 79 cell types (Thurman et al., 2012).

Gene Ontology and MOTIF analysis

Gene Ontology analysis was performed with the GREAT tool (http://bejerano.stanford.edu/great/public/html/; (McLean et al., 2010), using the parameters basal plus extension, 5kb upstream, 1kb downstream and plus distal 200kb regions. Motif analysis for enhancer regions was performed using Homer software (Heinz et al., 2010) with default parameters.

Enhancer-Driven GFP-Reporter Assay

Candidate enhancers were amplified from human genomic DNA by PCR and cloned into the GFP reporter vector pSinTK (Rada-Iglesias et al., 2011), using the polymerase incomplete primer extension (PIPE) cloning method (Klock et al., 2008). Lentiviruses were constructed by co-transfecting the pSinTK with pCMV R8.74 and pMD.G helper plasmids into HEK293T cells. Viral supernatant was collected and concentrated by ultracentrifugation for 2 hours at 19,400 RPM using an Optima L-80 XP Ultracentrifuge (Beckman Coulter).

Undifferentiated CyT49 hES cells were transduced with the reporter virus and maintained as described above with the addition of 300µg/ml Geneticin (G418 antibiotic) for selection of transduced cells. After one to two weeks of antibiotic selection, the cells were differentiated as described above. At the appropriate stages of differentiation, aggregates were collected for live cell imaging. For cell imaging, 40 µl of cell aggregates were washed in PBS, placed in an optically clear glass bottom dish (MatTek) and imaged at 20x magnification. Images from each time point were acquired using the identical exposure time with a Zeiss Axio-Observer-Z1 microscope and a Zeiss AxioCam digital camera.

Candidate Enhancers

Enhancer name	Coordinates
PDX1-1	chr13:27,381,114-27,382,132
PDX1-2	chr13:27,383,262-27,384,342
SOX17-1	chr8:55,342,984-55,343,980
SOX17-2	chr8:55,518,433-55,519,467
PRDM16	chr1:3,236,898-3,237,876

MIXL1	chr1:224,469,604-224,470,889
NKX6.1	chr4:85,575,615-85,577,482
SMDC1	chr10:112,125,356-112,127,191

Primers Used for Enhancer Cloning

Primer name	Sequence
Pdx1-1 F	5'-GAGACTAGCCTCGAGTTACATGAGAATGAGGAACAGGTCA
Pdx1-1 R	5'-GGCACGCGTTTCGAACTTTCACTTTACTCTATGGGCTCAC
Pdx1-2 F	5'-GAGACTAGCCTCGAGCCAGGCTCAGCTAATTTTTGTATTT
Pdx1-2 R	5'-GGCACGCGTTTCGAAGTATTAAGGGAGGATCAGCATGATC
Sox17-1 F	5'-GAGACTAGCCTCGAGACCTCACTCTATCACCTCCACAAGG
Sox17-2-R	5'-GGCACGCGTTTCGAATGGATTTCTCTAAAGGCAGAGCCAG
Sox17-2 F	5'-GAGACTAGCCTCGAGCCTGTGAACTGTTTGTATCCCACCT
Sox17-2 R	5'-GGCACGCGTTTCGAAGGTGTTACCTGTGTTGTTCATTGCT
Prdm16 F	5'-GAGACTAGCCTCGAGCCATTCTCTTCTGCTCTTCTTCCCA
Prdm16 R	5'-GGCACGCGTTTCGAACAACAAACGTGCTTTGAACCCCTCT
Mixl-F	5'-GAGACTAGCCTCGAGAGGCCTGGCACATTCTGCATTTGTA
Mixl-R	5'-GGCACGCGTTTCGAAACACAAGGTAAACAGACTGCTGACG
Nkx6.1-F	5'-GAGACTAGCCTCGAGACTAAAACCCCTTGGATGAGACCA
Nkx6.1-R	5'-GGCACGCGTTTCGAAAAACAAACAAAACCCCAGGACACT
SMDC-F	5'-GAGACTAGCCTCGAGGTTCCCGTAATCGGTGACTCTGTG
SMDC-R	5'-GGCACGCGTTTCGAACCCCACTCAAGGGTCCCTCTGAAG

PDX1 and *FOXA1* Knockdown

PDX1 and *FOXA1* knockdowns were performed using lentiviral vectors containing shRNA targeting *PDX1, FOXA1,* or a scrambled control. *PDX1* shRNA oligos were engineered using the Whitehead Institute siRNA design algorithm (http://sirna.wi.mit.edu/) and ligated into pLL3.7 as previously described (Rubinson et al., 2003), while *FOXA1* shRNA oligos were generated using designs from the Broad Institute's RNAi Consortium (http://www.broadinstitute.org/rnai/trc) and ligated into pLKO.1 (Addgene, Inc.). Lentivirus for both *PDX1* and *FOXA1* was generated in an identical manner as described for the enhancer-GFP reporter construct viruses. Undifferentiated CyT49 hESCs were infected with either *PDX1* shRNA (shPDX1), a cocktail of eight *FOXA1* shRNAs (shFOXA1.1 - 8), or a scrambled control (Scram) virus, then differentiated to the GT, FG and PE stages and harvested for analysis. shRNA oligos used are as follows:

shPDX1 sense: (5'-
TGGAGTTCCTATTCAACAAGTTCAAGAGACTTGTTGAATAGGAACTCCTTTTTTC-3')
shPDX1 antisense: (5'-
TCGAGAAAAAAGGAGTTCCTATTCAACAAGTCTCTTGAACTTGTTGAATAGGAACTC
A-3')
Scrambled sense: (5'-
TGAACAAGATGAAGAGCACCTTCAAGAGAGGTGCTCTTCATCTTGTTCTTTTTTC-3')
Scrambled antisense: 5'-
TCGAGAAAAAAGAACAAGATGAAGAGCACCTCTCTTGAAGGTGCTCTTCATCTTGTC
A-3')

shFOXA1.1 sense: (5’-

CCGGGCGTACTACCAAGGTGTGTATCTCGAGATACACACCTTGGTAGTACGCTTTTT

G-3’)

shFOXA1.1 antisense: (5’-

AATTCAAAAAGCGTACTACCAAGGTGTGTATCTCGAGATACACACCTTGGTAGTACG

C-3’)

shFOXA1.2 sense: (5’-

CCGGGCAGCATAAGCTGGACTTCAACTCGAGTTGAAGTCCAGCTTATGCTGCTTTTT

G-3’)

shFOXA1.2 antisense: (5’-

AATTCAAAAAGCAGCATAAGCTGGACTTCAACTCGAGTTGAAGTCCAGCTTATGCTG

C-3’)

shFOXA1-3 sense: (5’-

CCGGGCGAAGTTTAATGATCCACAACTCGAGTTGTGGATCATTAAACTTCGCTTTTT

G-3’)

shFOXA1.3 antisense: (5’-

AATTCAAAAAGCGAAGTTTAATGATCCACAACTCGAGTTGTGGATCATTAAACTTCG

C-3’)

shFOXA1.4 sense: (5’-

CCGGATACGAACAGGCACTGCAATACTCGAGTATTGCAGTGCCTGTTCGTATTTTTT

G-3’)

shFOXA1.4 antisense: (5’-

AATTCAAAAAATACGAACAGGCACTGCAATACTCGAGTATTGCAGTGCCTGTTCGTA

T-3’)

shFOXA1.5 sense: (5’-

CCGGGTATTCCAGACCCGTCCTAAACTCGAGTTTAGGACGGGTCTGGAATACTTTTT

G-3’)

shFOXA1.5 antisense: (5’-

AATTCAAAAAGTATTCCAGACCCGTCCTAAACTCGAGTTTAGGACGGGTCTGGAATA

C-3’)

shFOXA1.6 sense: (5’-

CCGGCAAACCGTCAACAGCATAATACTCGAGTATTATGCTGTTGACGGTTTGTTTTT

G-3’)

shFOXA1.6 antisense: (5’-

AATTCAAAAACAAACCGTCAACAGCATAATACTCGAGTATTATGCTGTTGACGGTTT

G-3’)

shFOXA1.7 sense: (5’-

CCGGTCTAGTTTGTGGAGGGTTATTCTCGAGAATAACCCTCCACAAACTAGATTTTT

G-3’)

shFOXA1.7 antisense: (5’-

AATTCAAAAATCTAGTTTGTGGAGGGTTATTCTCGAGAATAACCCTCCACAAACTAG

A-3’)

shFOXA1-8 sense: (5'-
CCGGGAACACCTACATGACCATGAACTCGAGTTCATGGTCATGTAGGTGTTCTTTTT
G-3')
shFOXA1.8 antisense: (5'-
AATTCAAAAAGAACACCTACATGACCATGAACTCGAGTTCATGGTCATGTAGGTGTT
C-3')

Immunocytochemical Analysis

hESC-derived cells were washed twice before fixation with 4% paraformaldehyde in PBS for 30 min at room temperature or overnight at 4°C. Samples were subsequently washed three times with PBS and then incubated in 30% sucrose at 4 °C overnight. Cells were mounted with Optimal Cutting Temperature Compound (Tissue-Tek) to prepare frozen blocks and were sectioned at 10 µm. Sections were washed with PBS for 5 min, and then permeabilized with PBSTr (PBS/0.15% (vol/vol) Triton X-100 (Sigma)) for 1 h. Cells were blocked with 10% normal donkey serum (Jackson Immuno Research Laboratories)/PBST (PBS/0.1% (vol/vol) Tween20 (Fisher Scientific)) for 1 h at room temperature. Primary and secondary antibodies were diluted in 1% normal donkey serum/PBST. Primary antibodies were incubated at 4°C overnight and secondary antibodies for 1 h at room temperature. The following primary antibodies were used: guinea pig anti-PDX1 (gift from Christopher Wright, 1:1000), mouse anti-NKX6.1 (BCBC, 1:300), rabbit anti-SOX9 (Millipore, 1:1000), goat anti-FOXA2 (Santa Cruz, 1:1000), rabbit anti-AFP (Dako, 1:1000), rabbit anti-Transferrin (Santa Cruz, 1:1000), rabbit anti-AAT (Sigma, 1:1000), mouse anti-NKX2.1 (TTF-1) (Santa Cruz, 1:100), rabbit anti-SOX2 (Abcam, 1:1000), and rabbit anti-PROX1 (Angiobio, 1:1000). Secondary antibodies were Cy3-,

Alex488-conjugated donkey antibodies against guinea pig, rabbit, mouse, and goat (Jackson Immuno Research Laboratories). Images were acquired on a Zeiss Axio-Observer-Z1 microscope with a Zeiss AxioCam digital camera and figures prepared with Adobe Photoshop/Illustrator CS4.

Reverse Transcription and Quantitative PCR

Total RNA was isolated from hESC-derived cellular aggregates with the RNAeasy Micro kit (Qiagen) with subsequent cDNA synthesis using the SuperScript III First-Strand Synthesis SuperMix (Life Technologies). PCR reactions were run in triplicate with 10 ng cDNA per reaction using CFX96 real-time system (BioRad) in which the housekeeping gene, glucoronidase beta (GUSB) was used as endogenous control. Primer sequences used in this study are listed below.

Primers Used in qRT-PCR Analysis

Primer name	Sequence
PDX1 forward	5'-AAGTCTACCAAAGCTCACGCG
PDX1 reverse	5'-GTAGGCGCCGCCTGC
SOX9 forward	5'-AGTACCCGCACTTGCACAAC
SOX9 reverse	5'-ACTTGTAATCCGGGTGGTCCTT
HNF6 forward	5'-CGCTCCGCTTAGCAGCAT
HNF6 reverse	5'-GTGTTGCCTCTATCCTTCCCAT
PTF1A forward	5'-GAAGGTCATCATCTGCCATCG
PTF1A reverse	5'-GGCCATAATCAGGGTCGCT

NKX6.1 forward	5'-CTGGCCTGTACCCCTCATCA
NKX6.1 reverse	5'-CTTCCCGTCTTTGTCCAACAA
NKX6.2 forward	5’-TCTGGAACCAGACCTTCAAC
NKX6.2 reverse	5’-AGATCTTCGCGCTGGAAGAAAA
AFP forward	5’-CGTTCTTGAACAAACTGGGCAAA
AFP reverse	5’-ACTGAATCCAGAACACTGCATAG
ALB forward	5'-TTTATGCCCCGGAACTCCTTT
ALB reverse	5'-ACAGGCAGGCAGCTTTATCAG
TAT forward	5’-TACAGACCCTGAAGTTACCCA
TAT reverse	5’-TAAGAAGCAATCTCCTCCCGA
Transferrin forward	5’-TGCAATCAAGGTAGGAGGCTT
Transferrin reverse	5’-AAACTGTGAGATGGTGTGCAG
AAT forward	5’-CGCATTGCTATCGAGGTTTCTTA
AAT reverse	5’-CACCAACTTTAGGGACTCTGC
APOB forward	5’ TCCCCAATGGTCCCTTCTC
APOB reverse	5’- CGTAGTACTCGCTCTGGTAATCTCC
ASGR2 forward	5’-GTGCGGTGGCTAAAGAATGAT
ASGR2 reverse	5’-ATTCGCAGTCGTGAACCATATT
SOX2 forward	5’-GGCAGCTACAGCATGATGCAGGAGC
SOX2 reverse	5’-CTGGTCATGGAGTTGTACTGCAGG
NKX2.1 forward	5’-CGGCATGAACATGAGCGGCAT
NKX2.1 reverse	5’-GCCGACAGGTACTTCTGTTGCTTG
HB9 forward	5’-CACCGCGGGCATGATC

HB9 reverse	5’-CACCGCGGGCATGATC
MEIS1 forward	5’-CAAGCCATACAAGTATTAAGGTTTCA
MEIS1 reverse	5’-ATCACCAAATCGATAGGCATT
NGN3 forward	5’-GCTCATCGCTCTCTATTCTTTTGC
NGN3 reverse	5’-GGTTGAGGCGTCATCCTTTCT
FOXA1 forward	5’-CTGTGGATGGTTGTATTGGGCAG
FOXA1 reverse	5’-GCTCGTAGTCATGGTGTTCA
FOXA2 forward	5’-GGGAGCGGTGAAGATGGA
FOXA2 reverse	5’-TCATGTTGCTCACGGAGGAGTA
HNF4A forward	5’-TTGACCTTCGAGTGCTGATCC
HNF4A reverse	5’-CACGGGCAAACACTACGGT
PROX1 forward	5’-AACATGCACTACAATAAAGCAAATGAC
PROX1 reverse	5’-CAGGAATCTCTCTGGAACCTCAAA

ChIP-qPCR

Chromatin immunoprecipitation for H3K4me1 in Figure 5F was performed using the same antibody and method as described for ChIP-seq experiments. ChIP experiments were quantified using iQ SYBR green Supermix (BioRad) and the CFX96 real-time system (BioRad). Enrichment was calculated as percentage of input. The PCR primers used in the ChIP-qPCR analysis were as follows:

Primer Name	Sequence
PDX1-1 forward	5’-GACTGCTCGGATTTTGGTGT

PDX1-1 reverse	5’-AAGGTACGGCTGACTCTCCA
PDX1-2 forward	5’-GGGAGTTTGTCAGACCAGGA
PDX1-2 reverse	5’-GGCAGCGAGAAAACACTAGG
PDX1-3 forward	5’-GGGCAAATTGGAGCAGTAAA
PDX1-3 reverse	5’-ATTATGCACACACCCACCAA
PTF1A forward	5’-AACAGCTCAAGAGGACAGGT
PTF1A reverse	5’-TAAAACAGACACGGCAACCC
SOX9-1 forward	5’-AGGGGCCAGGTCCTTATCTA
SOX9-1 reverse	5’-TCAGGCAACTTCCCAAGTCT
SOX9-2 forward	5’-GAGAGCTGTGGGACAGGAAG
SOX9-2 reverse	5’-GCCTGCATCCAAGAAAACAT
NEG forward	5’-TCAAAGGCTCATCTTTGCAG
NEG reverse	5’-AAAGCTGGACTGGTGAATGC

III. Supplemental References

Bhandare, R., Schug, J., Le Lay, J., Fox, A., Smirnova, O., Liu, C., Naji, A., and Kaestner, K.H. (2010). Genome-wide analysis of histone modifications in human pancreatic islets. Genome Res *20*, 428-433.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science (New York, NY *322*, 1845-1848.

Hawkins, R.D., Hon, G.C., Lee, L.K., Ngo, Q., Lister, R., Pelizzola, M., Edsall, L.E., Kuan, S., Luu, Y., Klugman, S., *et al.* (2010). Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell *6*, 479-491.

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.

Huang, S.X., Islam, M.N., O'Neill, J., Hu, Z., Yang, Y.G., Chen, Y.W., Mumau, M., Green, M.D., Vunjak-Novakovic, G., Bhattacharya, J., *et al.* (2014). Efficient generation of lung and airway epithelial cells from human pluripotent stem cells. Nat Biotechnol *32*, 84-91.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290-294.

Klock, H.E., Koesema, E.J., Knuth, M.W., and Lesley, S.A. (2008). Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts. Proteins *71*, 982-994.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature biotechnology *28*, 495-501.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279-283.

Rubinson, D.A., Dillon, C.P., Kwiatkowski, A.V., Sievers, C., Yang, L., Kopinja, J., Rooney, D.L., Zhang, M., Ihrig, M.M., McManus, M.T., *et al.* (2003). A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. Nature genetics *33*, 401-406.

Schulz, T.C., Young, H.Y., Agulnick, A.D., Babin, M.J., Baetge, E.E., Bang, A.G., Bhoumik, A., Cepa, I., Cesario, R.M., Haakmeester, C., *et al.* (2012). A Scalable System for Production of Functional Pancreatic Progenitors from Human Embryonic Stem Cells. PLoS ONE *7*, e37004.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., *et al.* (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75-82.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105-1111.

Trapnell, C., and Salzberg, S.L. (2009). How to map billions of short reads onto genomes. Nat Biotechnol *27*, 455-457.

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., *et al.* (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature *474*, 390-394.

Xie, R., Everett, Logan J., Lim, H.-W., Patel, Nisha A., Schug, J., Kroon, E., Kelly, Olivia G., Wang, A., D'Amour, Kevin A., Robins, Allan J., *et al.* (2013a). Dynamic Chromatin Remodeling Mediated by Polycomb Proteins Orchestrates Pancreatic Differentiation of Human Embryonic Stem Cells. Cell Stem Cell *12*, 224-237.

Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., *et al.* (2013b). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. Cell *153*, 1134-1148.