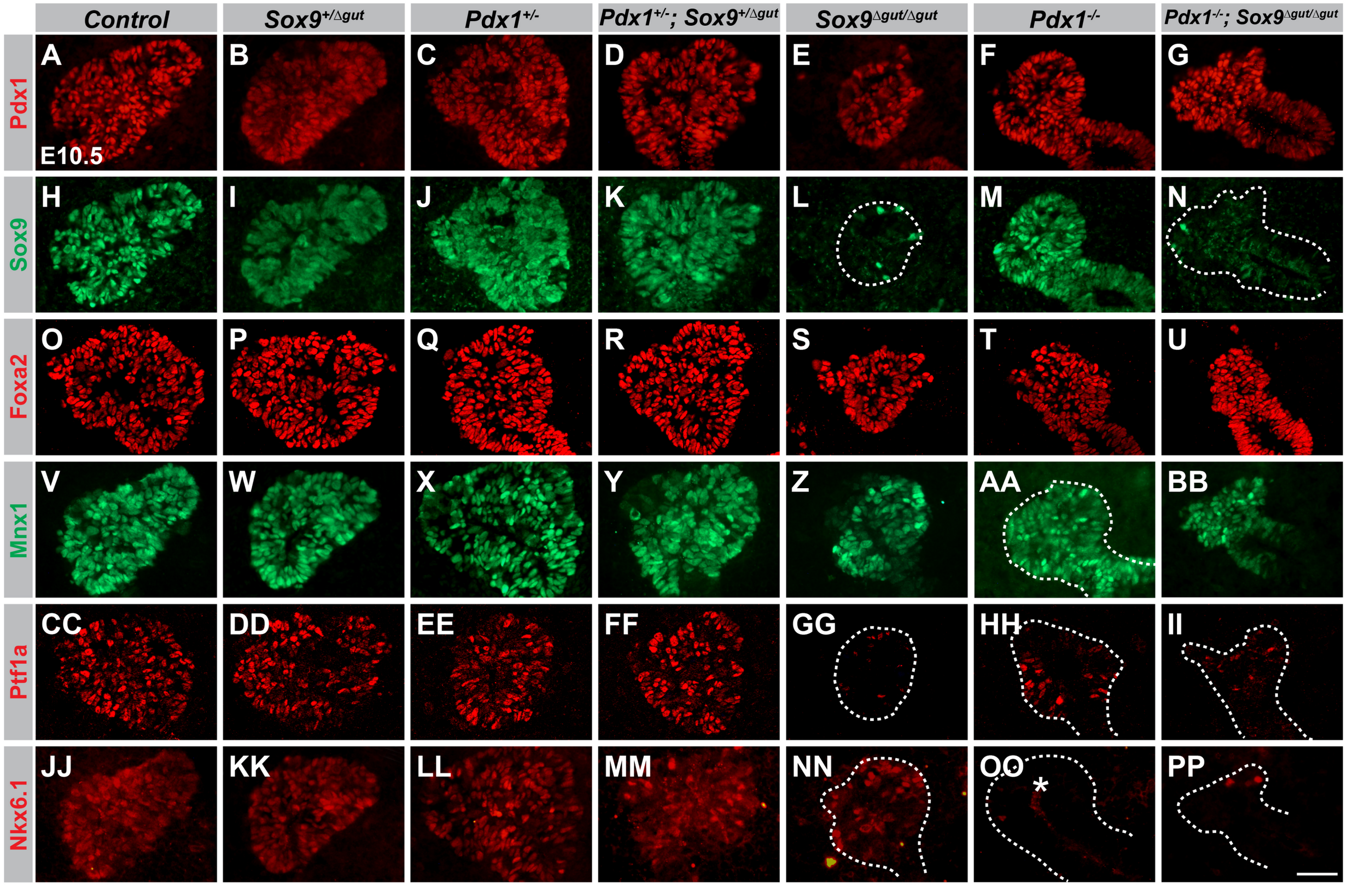
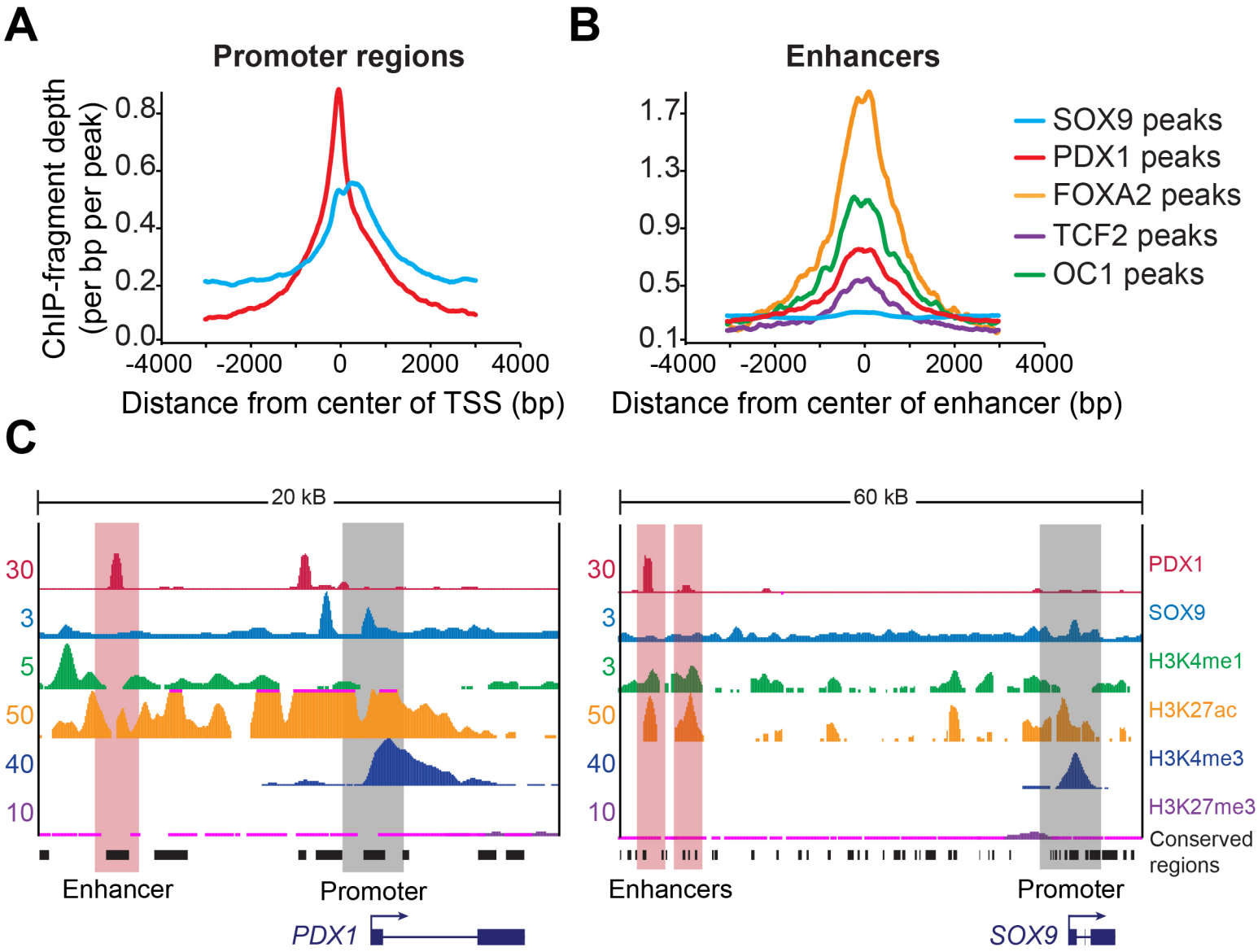
Shih_Figure S1. Related to Figure 3



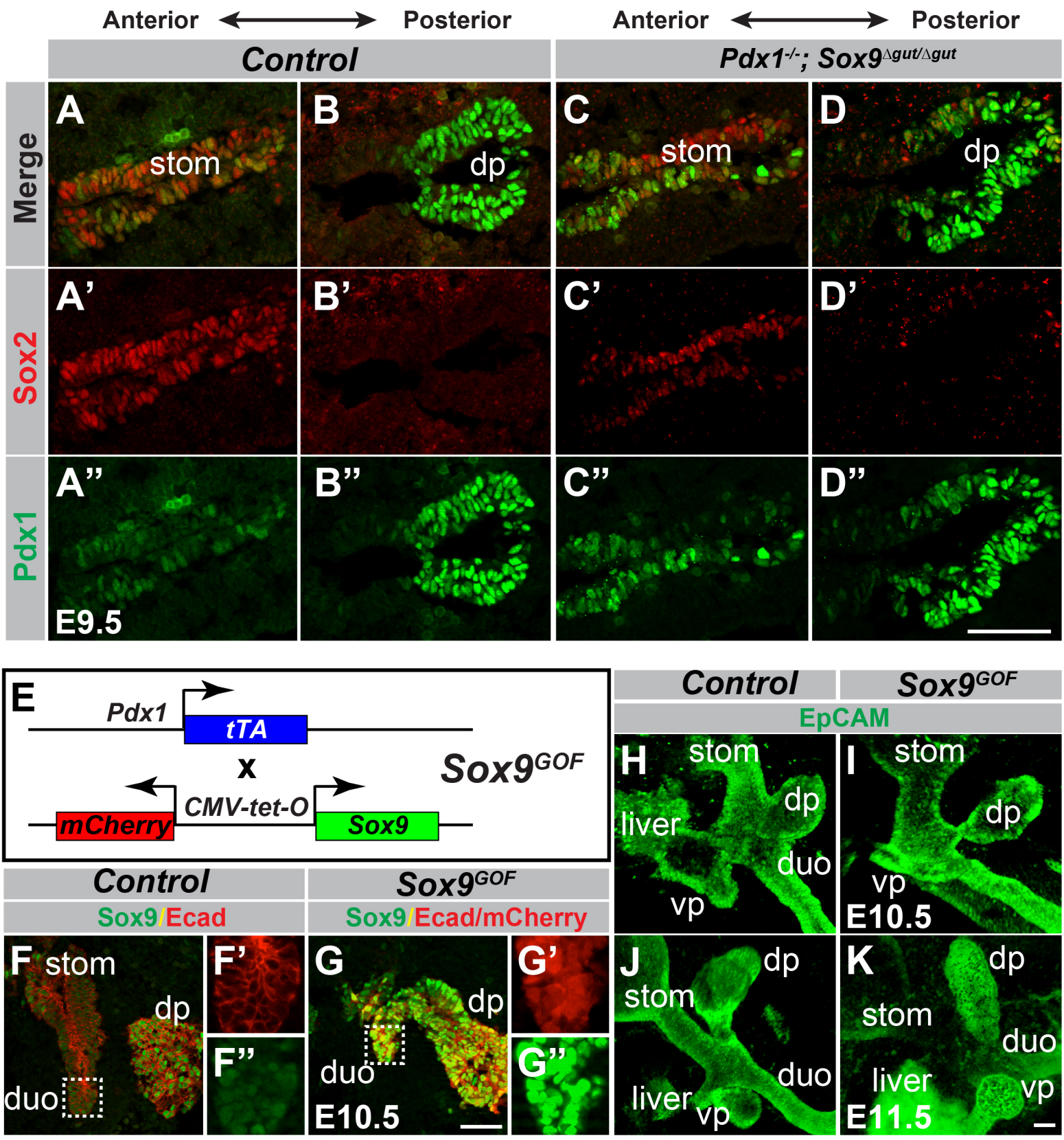
Supplemental Figure S1, Related to Figure 3. Combined loss of *Pdx1* and *Sox9* abrogates the early pancreatic program. (**A-PP**) Immunofluorescence analysis for Pdx1 (**A-G**), Sox9 (**H-N**), Foxa2 (**O-U**), Mnx1 (**V-BB**), Ptf1a (**CC-II**), and Nkx6.1 (**JJ-PP**) on embryonic day (E) 10.5 embryos carrying various combinations of *Pdx1* and *Sox9* mutant alleles. Foxa2 and Mnx1 are expressed in a Pdx1- and Sox9-independent manner, whereas expression of Ptf1a and Nkx6.1 is regulated by Pdx1 and Sox9. Where necessary, dorsal pancreas or the entire foregut region is demarcated by a dashed line. Non-specific signal for Nkx6.1 is evident in *Pdx1*$^{-/-}$ dorsal pancreas lumen (**OO**, asterisk) due to antibody trapping. Scale bar = 50 μm.

Shih_Figure S2. Related to Figure 4.



Supplemental Figure S2, Related to Figure 4. SOX9 occupies promoters whereas PDX1 occupies enhancers together with other pancreatic transcription factors. (**A-B**) Histogram of SOX9 and PDX1 binding peaks around transcriptional start sites (TSSs; promoter regions) in human embryonic stem cell (hESC)-derived pancreatic progenitors (**A**). Histogram of SOX9, PDX1, FOXA2, TCF2, and ONECUT-1 (OC1) binding peaks at enhancers in hESC-derived pancreatic progenitors (enhancers were defined by presence of H3K4me1 and H3K27ac, and absence of H3K4me3; **B**). (**C**) ChIP-seq binding profiles (reads per million) for PDX1, SOX9 and histone modifications (H3K4me1, H3K27ac, H3K4me3, H3K27me3) at the *PDX1* and *SOX9* loci. Black boxes indicate conserved regions in mice. kB, kilobases.

Shih_Figure S3. Related to Figure 6.



Supplemental Figure S3, Related to Figure 6. Lack of ectopic Sox2 expression following combined *Pdx1* and *Sox9* deletion and unperturbed foregut morphogenesis after forced expression of *Sox9* in the Pdx1^{+} domain. (**A-D**) Immunofluorescence analysis for Pdx1 and Sox2 on *Pdx1* and *Sox9* double mutant (**C,D**) and control (**A,B**) embryos at embryonic day (E) 9.5. **A-A"** and **C-C"** represent sections more anterior to those shown in **B-B"** and **D-D"**. The pancreatic area is shown in **B-B"** and **D-D"**. (**E**) Schematic of the experimental strategy: *Pdx1*^{tTA} mice were crossed with *Rosa26*$^{mCherry\text{-}tetO\text{-}Sox9}$ mice to generate *Sox9*^{GOF} embryos. (**F,G**) In the absence of doxycycline, both mCherry and Sox9 are strongly expressed in the Pdx1^{+} pancreatic and duodenal domain at E10.5. Duodenal Sox9 expression is notably increased in *Sox9*^{GOF} mice compared to endogenous Sox9 expression levels in control littermates. Fields demarcated by dashed boxes in **F** and **G** are shown at higher magnification in **F',F''** and **G',G"**, respectively. (**H-K**) 2D projections of 3D *z*-stacks of developing foregut regions in control and *Sox9*^{GOF} embryos at E10.5 (**H,I**) and E11.5 (**J,K**) following whole-mount immunofluorescence staining for EpCAM. Gross gut morphology is unaffected in *Sox9*^{GOF} embryos. dp, dorsal pancreas; vp, ventral pancreas; duo, duodenum; stom, stomach. Scale bar = 50 μm.

Supplemental Movie S1, Related to Figure 2. 3D projections of the embryonic gut tube stained for Cdx2, Sox9, and Pdx1. Whole mount immunofluorescence staining for Cdx2 (blue), Sox9 (red) and Pdx1 (green) at embryonic day 8.75 reveals that Cdx2 is largely absent from the pancreatic domain.

Table S1, Related to Figure 4
Table S2, Related to Figure 5
Table S3, Related to Figure 5
Table S4, Related to Figure 5
Table S5, Related to Figure 5
Table S6, Related to Supplemental Experimental Procedures

Supplemental Experimental Procedures

Immunofluorescence analysis

Dissected E8.75-E10.5 embryos were fixed in 4% paraformaldehyde (PFA) in PBS, equilibrated in 30% sucrose in PBS, cryoembedded in Tissue-Tek OCT (Sakura Finetek USA, Torrance, CA, USA) then sectioned at 10 μm. For immunofluorescence analysis, antigen retrieval was conducted in pH 6.0 citrate buffer followed by permeabilization in 0.15% Triton X-100 in PBS. Sections were blocked in 1% normal donkey serum in PBS with 0.1% Tween-20 then incubated overnight at 4°C with primary antibodies diluted in the same buffer; detection with secondary antibodies was conducted by a 1.5 h incubation at room temperature. Where necessary, nuclei were counterstained with Hoechst 33342 (Invitrogen) at 10 μg/ml.

Images were captured on Zeiss Axioplan 2 or Axio Observer Z1 microscopes running Zeiss AxioVision 3.1 or 4.8 (both Carl Zeiss, Thornwood, NY, USA) respectively, or on a Leica TCS SP8 confocal microscope running Leica LAS AF v.3.3.0

(Leica Microsystems, Wetzlar, Germany). Figures were prepared with Adobe Photoshop/Illustrator CS5.5 and CS6 (Adobe Systems, San Jose, CA, USA).

Whole-mount immunofluorescence analysis of embryos was performed as previously described (Ahnfelt-Ronne et al., 2007). Briefly, primary and secondary antibodies were used at the dilutions noted in **Supplemental Table 6**. Following dehydration to methanol and clearing in BABB (one part benzyl alcohol to two parts benzyl benzoate), *z*-stacks were captured on a Zeiss LSM710 confocal microscope driven by Zeiss Zen software, pseudocolored, and projected to 3D using Imaris x64 7.1.1 or Amira 3D 6.0.

X-Gal histochemistry

Whole-mount X-Gal staining of whole embryos was performed as previously described (Seymour et al., 2004). Embryos were dehydrated to methanol and cleared in BABB following staining. Images were acquired on a Zeiss Stemi 2000C with a Zeiss AxioCam digital camera driven by Zeiss AxioVision 3.1.

hESC culture and human fetal pancreas

CyT49 hESCs were maintained and differentiated to the pancreatic progenitor cell stage as previously described (Schulz et al., 2012; Xie et al., 2013). Differentiation of hESCs into hepatic progenitors was performed employing the same culture conditions as described for pancreatic differentiation with minor modifications: at the definitive endoderm stage (day 2), cell aggregates were treated for six days with 50 ng/ml BMP4 (Millipore) and 10 ng/ml FGF2 (Millipore) in RPMI media (Mediatech) supplemented with 0.2% (vol/vol) FBS (HyClone) with daily feeding.

hESC research was approved by the University of California San Diego Institutional Review Board and Embryonic Stem Cell Research Oversight Committee. Microdissected human fetal pancreata at day 54 to 59 of gestation were obtained from the University of Washington Birth Defects Research Laboratory.

RNA-seq sample preparation and analysis

RNA-seq data sets for all hESC-derived pancreatic cell populations, their developmental precursors and primary human islets have been described (Xie et al., 2013). For hESC-derived hepatic progenitors and human fetal pancreata (three pancreata were pooled at days 54, 57, and 59 of gestation), strand-specific RNA-seq libraries were prepared as previously described (Parkhomchuk et al., 2009), with minor modifications. Briefly, cells/tissues were lysed in Trizol (Life Technologies) for extraction of total RNA. Residual contaminating genomic DNA was removed using the Turbo DNase kit (Ambion). mRNA was isolated from 2 μg of DNA-free total RNA using the Dynabeads mRNA Purification kit (Life Technologies). Following purification, the mRNA was primed with Olig(dT)s and random hexamers and then reverse-transcribed to first-strand cDNA. Residual dNTPs were removed using Illustra MicroSpin G-25 columns (GE Healthcare). In the second-strand synthesis reaction, dUTPs were used instead of dTTPs. The double-strand cDNA was fragmented using a Bioruptor Sonicator (60 cycles of 30 sec on and off). After end-repair and adenine base addition, the cleaved double-strand cDNA fragments were ligated to Pair-end Adaptor Oligo Mix (Illumina) and size-fractionated on a 2% agarose gel. cDNA fragments of 200±25 bp were recovered and incubated with uracil-N-glycosylase (UNG) to digest the second-strand cDNA. Purified

single-strand cDNA was then used as template for 15 cycles of amplification using pair-end PCR primers (Illumina). The amplified products were separated on a 2% agarose gel and a band between 225-275 bp was excised.

For each sample, sequence reads were aligned to the transcriptome using RUM, and a “Feature Quantification” (FQ) value was computed for each Refseq mRNA transcript, where each FQ value = the number of reads overlapping each transcript per million reads sequenced, per kb of transcript length. In accordance with recommendations from ENCODE and the BCBC, these experiments were performed on two independent biological replicates. The FQ values for each pair of sample replicates showed high correlation, and were therefore averaged together before subsequent analysis. RPMK values were determined as described (Xie et al., 2013). A gene was considered “expressed” in hESC-derived pancreatic progenitors, if RPKM values were ≥ 0.1. Genes with an RPKM of < 0.1 were considered "not expressed".

ChIP-seq for histone modifications and enhancer predictions

ChIP-seq of histone modifications was performed as previously described (Hawkins et al., 2010). All the sequencing experiments were performed using Illumina Hi-Seq 2000 instruments. Each read was aligned to the human genome build hg18 with Bowtie (Langmead et al., 2009). We used the first 36 bp for the alignment and only kept reads with up to two mismatches. Duplicated reads from the same library were removed. Data sets from highly correlated biological replicates were pooled for subsequent analysis. MACS (Zhang et al., 2008) was used for peak calling. Peaks were further filtered as described (Shen et al., 2012).

Enhancers were predicted as described, using H3K4me1, H3K4me3, and H3K27ac ChIP-seq profiling (Rajagopal et al., 2013). We first divided the human genome into 100 bp bins and counted the number of reads that fell within each bin. Then the tag counts in each bin were normalized against the total number of reads and input as described (Shen et al., 2012). The normalized signals for each mark were merged as one input file for the enhancer prediction pipeline. To compute the FDR, we first shuffled the rows and columns of the input data. Second, we ran the enhancer prediction pipeline on this simulated data. The FDR was computed as the ratio of the number of predicted enhancers from simulated data over the real data. We required that predicted enhancers have an FDR of < 2% and are at least 3 kb away from a known transcriptional start site.

SOX9 and PDX1 ChIP-seq sample preparation

Chromatin immunoprecipitations were performed as previously described (Bhandare et al., 2010). Briefly, samples were crosslinked in 1.1% formaldehyde/PBS for 15 min at room temperature and then quenched with 0.125 M glycine/PBS. Samples were subsequently washed twice with PBS and then lysed in 1% SDS. For sonication, lysates were sonicated with a Bioruptor Sonicator six times for 5 min each with a 30 sec on and off cycle, resulting in 200-500 bp chromatin fragments. Sheared chromatin was incubated overnight at 4°C with 5 µg rabbit anti-SOX9 antibody (Millipore, AB5535; Lot number 2262679) or 15 µl goat anti-PDX1 antiserum (BCBC). Chromatin and antibody complex was incubated with 12.5 µl of Dynabeads protein A plus 12.5 µl of Dynabeads protein G (Life Technologies) for 4 h at 4°C. Immunoprecipitated complexes were further eluted, reverse crosslinked, and subjected to library preparation.

ChIP-seq libraries were prepared as per Illumina's instructions (http://www.illumina.com). For input library preparation, 50 ng of input DNA from each sample was used. After adaptor ligation, DNA fragments were size-fractionated by gel electrophoresis and excised at 200±25 bp. Following gel purification, DNA fragments were amplified with 18 PCR cycles and purified using a MiniElute PCR Purification kit (Qiagen). 10 nM purified DNA was loaded on the flow cell, and sequencing was performed on an Illumina/Solexa Genome Analyzer II in accordance with the manufacturer's protocols.

mRNA expression profiling using microarrays

Total RNA was isolated from microdissected E12.5 pancreatic epithelia of *Pdx1*$^{+/-}$, *Sox9*$^{fl/+}$*;Foxa3-Cre* and *Pdx1*$^{+/-}$*;Sox9*$^{fl/+}$*;Foxa3-Cre* littermates. A total of twelve pancreatic epithelia were isolated per genotype. Each individual RNA sample was prepared from four pancreata as per the manufacturer's instructions (Micro RNA isolation kit, Qiagen). RNA quality was assessed with the Agilent 2100 Bioanalyzer (Agilent Technologies). Approximately 250 ng of total RNA was amplified and labeled with Cy3 using the QuickAmp Labeling Kit (Agilent Technologies). Four pancreatic epithelia per genotype were pooled for three biological replicates to hybridize to Agilent Whole Mouse Genome Oligo Microarray G4122A chips (Agilent Technologies, Palo Alto, CA, USA). The gene expression data were analyzed by the statistical tool corgon as previously described (Glatt et al., 2005; Sasik et al., 2002).

Supplemental References

Ahnfelt-Ronne, J., Jorgensen, M.C., Hald, J., Madsen, O.D., Serup, P., and Hecksher-Sorensen, J. (2007). An improved method for three-dimensional reconstruction of protein expression patterns in intact mouse and chicken embryos and organs. J Histochem Cytochem *55*, 925-930.

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 research *20*, 428-433.

Glatt, S.J., Everall, I.P., Kremen, W.S., Corbeil, J., Sasik, R., Khanlou, N., Han, M., Liew, C.C., and Tsuang, M.T. (2005). Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. Proceedings of the National Academy of Sciences of the United States of America *102*, 15533-15538.

Hawkins, R.D., Hon, G.C., and Ren, B. (2010). Next-generation genomics: an integrative approach. Nature reviews Genetics *11*, 476-486.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., and Soldatov, A. (2009). Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic acids research *37*, e123.

Rajagopal, N., Xie, W., Li, Y., Wagner, U., Wang, W., Stamatoyannopoulos, J., Ernst, J., Kellis, M., and Ren, B. (2013). RFECS: a random-forest based algorithm for

enhancer identification from chromatin state. PLoS computational biology *9*, e1002968.

Sasik, R., Calvo, E., and Corbeil, J. (2002). Statistical analysis of high-density oligonucleotide arrays: a multiplicative noise model. Bioinformatics *18*, 1633-1640.

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.

Seymour, P.A., Bennett, W.R., and Slack, J.M. (2004). Fission of pancreatic islets during postnatal growth of the mouse. J Anat *204*, 103-116.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V.*, et al.* (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116-120.

Xie, R., Everett, L.J., Lim, H.W., Patel, N.A., Schug, J., Kroon, E., Kelly, O.G., Wang, A., D'Amour, K.A., Robins, A.J.*, et al.* (2013). Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. Cell Stem Cell *12*, 224-237.

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 biology *9*, R137.