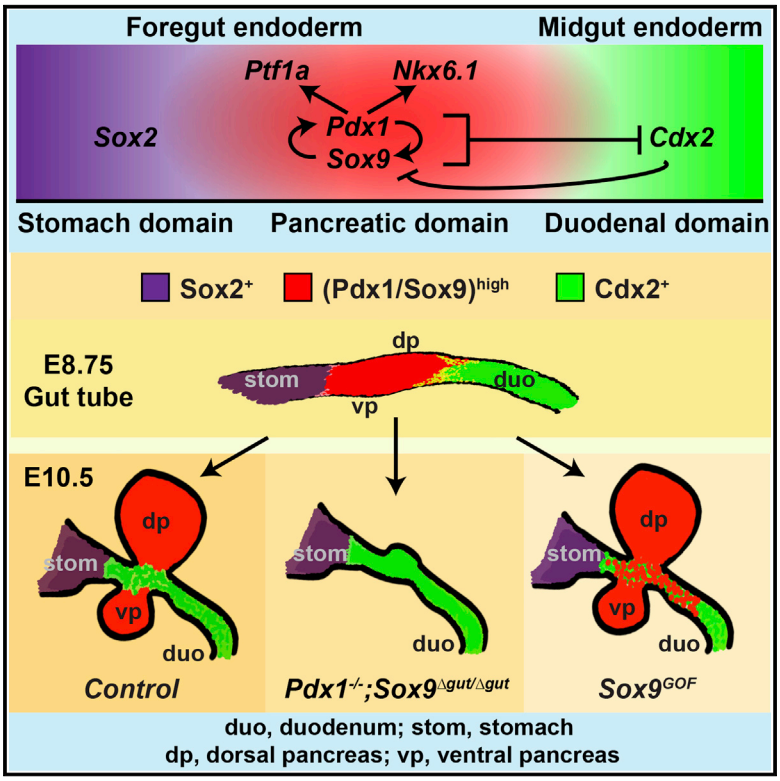
Article

Cell Reports

# A Gene Regulatory Network Cooperatively Controlled by Pdx1 and Sox9 Governs Lineage Allocation of Foregut Progenitor Cells

## Graphical Abstract




## Authors

Hung Ping Shih, Philip A. Seymour, Nisha A. Patel, ..., Gene W. Yeo, Mark A. Magnuson, Maike Sander

## Correspondence

masander@ucsd.edu

## In Brief

Shih et al. identify a positive cross-regulatory Pdx1-Sox9 loop that promotes expression of the pancreas-specific factors Ptf1a and Nkx6.1 while repressing intestinal cell fate determinants, including Cdx2, favoring adoption of a pancreatic fate.

## Highlights

- Genetic studies show Pdx1 and Sox9 cooperatively specify the pancreatic lineage
- Pdx1+Sox9 co-occupy regulatory sequences of pancreatic and intestinal genes
- Pdx1+Sox9 cooperatively repress intestinal cell fate determinants such as Cdx2
- Pdx1+Sox9 are necessary and sufficient to repress the intestinal fate choice

## Accession Numbers

GSE61945
GSE61946
GSE61947
GSE62023

Shih et al., 2015, Cell Reports *13*, 326–336
October 13, 2015 ©2015 The Authors
http://dx.doi.org/10.1016/j.celrep.2015.08.082

OPEN ACCESS
CellPress

Cell Reports
Article

# A Gene Regulatory Network Cooperatively Controlled by Pdx1 and Sox9 Governs Lineage Allocation of Foregut Progenitor Cells

Hung Ping Shih,[1,5,7] Philip A. Seymour,[1,4,7] Nisha A. Patel,[1] Ruiyu Xie,[1,6] Allen Wang,[1] Patrick P. Liu,[2] Gene W. Yeo,[2] Mark A. Magnuson,[3] and Maike Sander[1,*]

[1]Departments of Pediatrics and Cellular & Molecular Medicine, Pediatric Diabetes Research Center, University of California San Diego, La Jolla, CA 92093, USA
[2]Department of Cellular & Molecular Medicine, University of California San Diego, La Jolla, CA 92093, USA
[3]Center for Stem Cell Biology, Vanderbilt University, 2213 Garland Avenue, Nashville, TN 37232, USA
[4]DanStem, University of Copenhagen, Blegdamsvej 3B, DK-2200, Copenhagen N, Denmark
[5]Present address: Department of Translational Research and Cellular Therapeutics, City of Hope, Duarte, CA 91010, USA
[6]Present address: Faculty of Health Sciences, University of Macau, Taipa, Macau, China
[7]Co-first author
*Correspondence: masander@ucsd.edu
http://dx.doi.org/10.1016/j.celrep.2015.08.082


## SUMMARY

**The generation of pancreas, liver, and intestine from a common pool of progenitors in the foregut endoderm requires the establishment of organ boundaries. How dorsal foregut progenitors activate pancreatic genes and evade the intestinal lineage choice remains unclear. Here, we identify Pdx1 and Sox9 as cooperative inducers of a gene regulatory network that distinguishes the pancreatic from the intestinal lineage. Genetic studies demonstrate dual and cooperative functions for Pdx1 and Sox9 in pancreatic lineage induction and repression of the intestinal lineage choice. Pdx1 and Sox9 bind to regulatory sequences near pancreatic and intestinal differentiation genes and jointly regulate their expression, revealing direct cooperative roles for Pdx1 and Sox9 in gene activation and repression. Our study identifies Pdx1 and Sox9 as important regulators of a transcription factor network that initiates pancreatic fate and sheds light on the gene regulatory circuitry that governs the development of distinct organs from multi-lineage-competent foregut progenitors.**

## INTRODUCTION

During mammalian development, naive endodermal progenitors are directed toward different organ fates, including lung, pancreas, liver, and intestine. At developmental junctures, multipotent progenitors must be allocated to different lineages, exemplified by progenitors in the foregut endoderm, which give rise to pancreas, stomach, duodenum, liver, and the hepatobiliary system. Organ lineage choices are initiated by cross-repressive interactions between transcription factors (TFs) driving alternative lineage programs, followed by feed-forward induction of additional TFs to further execute the differentiation process (Holmberg and Perlmann, 2012). A large body of work has identified numerous TFs that are required for the early development of individual organs, in particular, the pancreas and liver (Seymour and Sander, 2011; Zaret, 2008). Despite these significant advances, it is still poorly understood which regulatory networks induce specific organ fates and how organ boundaries are established in the foregut endoderm. Identifying the mechanisms responsible for specifying individual organ fates is important for devising cell reprogramming strategies, which are still lacking for ex vivo production of pancreatic cells.

The pancreas arises as two buds on opposing sides of the gut tube at the boundary between the stomach and duodenum, the most rostral portion of the intestine (Shih et al., 2013). The anatomical location of the pancreas implies that an organ boundary must be established that distinguishes pancreatic from stomach and intestinal progenitors. The TF Cdx2 is exclusively expressed in intestinal epithelial cells, spanning the length of the alimentary tract from the proximal duodenum to the distal rectum. Cdx2 is essential for intestinal development and induces intestinal epithelial differentiation by activating the transcription of intestine-specific genes, such as MUC2, sucrase, and carbonic anhydrase I (Gao et al., 2009; Verzi et al., 2011). However, the mechanisms preventing expansion of the Cdx2 expression domain beyond the duodenal boundary in the foregut endoderm remain undefined.

The TFs Pdx1, Foxa2, Mnx1 (Hb9), Onecut-1 (Hnf6), Prox1, Tcf2, Gata4/Gata6, Sox9, and Ptf1a each play an important role in early pancreas development, yet deletion of no single factor alone is sufficient to abrogate pancreatic lineage induction (Carrasco et al., 2012; Harrison et al., 1999; Haumaitre et al., 2005; Jacquemin et al., 2000; Kawaguchi et al., 2002; Lee et al., 2005; Offield et al., 1996; Seymour et al., 2007; Wang et al., 2005; Xuan et al., 2012). These observations imply either

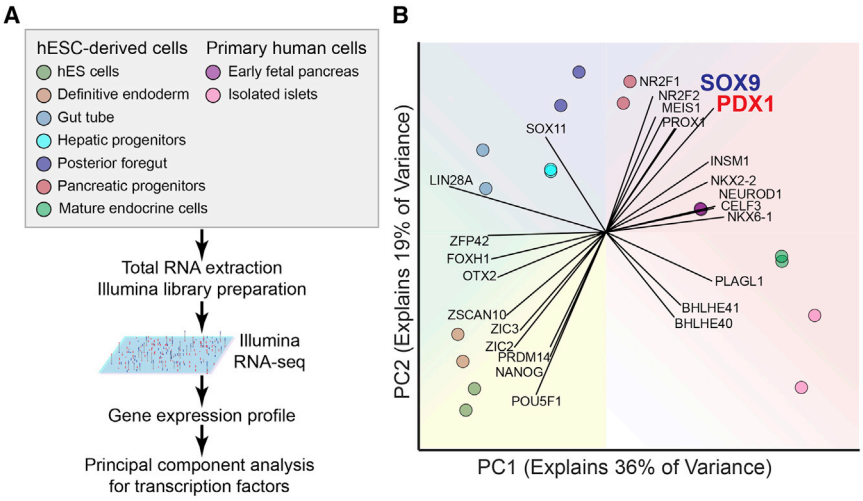



**Figure 1. Principal Component Analysis for Expression of Transcription Factors in Endodermal Cell Populations**

(A) Experimental strategy for principal component analysis of transcription factors in various endodermal cell populations.

(B) Principal component (PC) analysis of the expression values (RPKM) characterizing the variance explained by transcription factors expressed in human embryonic stem cell (hESC)-derived populations and primary human cells. Each vector emanating from the origin represents an individual gene. Each dot represents a sample, and each color represents the type of sample.

that the inducer of the pancreatic fate remains to be identified or that the pancreatic fate is specified through a cooperative mechanism involving multiple TFs.

Combining genetic, cistrome, and transcriptome analysis, we here identify the TFs Pdx1 and Sox9 as cooperative inducers of the pancreatic lineage. The combined inactivation of *Pdx1* and *Sox9* leads to an intestinal fate conversion of the pre-pancreatic domain, illustrated by expansion of the field of Cdx2 expression. Conversely, ectopic expression of Sox9 in intestinal progenitors is sufficient to induce Pdx1 and repress Cdx2. At a mechanistic level, we show that Pdx1 and Sox9 function as direct and cooperative activators of pancreatic genes and repressors of intestinal lineage regulators. Together, these findings shed light on the transcriptional mechanisms that induce the pancreatic fate and establish the pancreatic-to-intestinal organ boundary.

## RESULTS

### Pdx1 and Sox9 Cooperatively Induce the Pancreatic Lineage Program

To identify TFs most closely associated with pancreatic lineage induction, we compared expression levels of TFs represented in the RNA-seq data from pancreatic progenitor cells and closely related endodermal cell populations. These comprised human embryonic stem cell (hESC)-derived definitive endoderm, gut tube progenitors, posterior foregut, pancreatic progenitors, hepatic progenitors, and endocrine cells as well as primary human fetal pancreatic anlagen and primary cadaver pancreatic islets (Figure 1A). Principal component analysis of TF expression data clustered the different cell populations by developmental proximity, effectively reconstructing the dynamics of endodermal development and underscoring the importance of TF levels in successfully delineating these cell types (Figure 1B). Two TFs, PDX1 and SOX9, most strongly distinguished pancreatic progenitors from other cell populations (Figure 1B), suggesting possible cooperative roles for PDX1 and SOX9 in pancreatic lineage specification.

First, to define the domains of Pdx1 and Sox9 expression during pancreatic specification, we performed co-immunofluorescence staining for Pdx1 and Sox9 together with the anterior foregut marker Sox2 or the mid/hindgut marker Cdx2, respectively, at embryonic day (E) 8.75 (15–17 somites). The $Sox2^+$ domain, from which the stomach develops (McCracken et al., 2014; Sherwood et al., 2009), formed a boundary with both the $Pdx1^+$ and $Sox9^+$ domains (Figures 2A–2A″). Very few cells co-expressing Sox2, Pdx1, and Sox9 were observed at this boundary (Figures 2A–2A″). Cells in the presumptive proximal duodenum expressed high levels of Cdx2 and Sox9 (Figures 2B–2B″). In contrast to Sox9, which spanned the proximal duodenal and pre-pancreatic domains, Pdx1 was restricted to the pre-pancreatic domain (Figures 2B and 2B′). At the boundary between the duodenal and pre-pancreatic domains, we observed a transition from a $Cdx2^{high}$ to a $Cdx2^{low}$ state (Figures 2B and 2B″, dashed line; Movie S1). Consistent with previous studies (McCracken et al., 2014), Cdx2 was largely absent from the pancreatic buds (Figure 2C), showing that Cdx2 is gradually excluded from the pancreatic domain.

To determine the fate of *Sox9*- or *Pdx1*-expressing cells in the foregut endoderm, we performed lineage tracing in embryos carrying the *Rosa26*$^{mTomato/mGFP}$ (*R26*$^{mT/mG}$) reporter allele and an inducible form of Cre-recombinase, *CreER*, driven by either *Sox9* or *Pdx1* regulatory sequences. In these mice, tamoxifen administration to pregnant dams turns off constitutive expression of membrane-targeted Tomato (mT) and induces heritable expression of membrane-targeted GFP (mGFP), permitting recombined cells and their progeny to be traced by mGFP labeling. Tamoxifen administration at E8.0 resulted in labeling of the pancreatic epithelium in *R26*$^{mT/mG}$*;Pdx1-CreER* (Figure 2D) and *R26*$^{mT/mG}$*;Sox9-CreER* (Figure 2E) embryos at E10.5. Consistent with the incomplete segregation of the $Cdx2^+$ and $Pdx1^+$/$Sox9^+$ domains at E8.75 (Figures 2B–2B″ and 2C), mGFP labeling was also observed in scattered $Cdx2^+$ cells of the proximal duodenum (Figures 2D and 2E). $mGFP^+$ cells in the $Sox2^+$ gastric region were extremely rare (data not shown). Together, these findings indicate that the pancreatic-to-stomach boundary is largely established by E8.75, whereas the pancreatic and duodenal domains separate gradually between E8.75 and E10.5.

Previous studies have shown that pancreatic outgrowth and induction of a subset of early pancreatic markers still occur in *Pdx1* null mutants (Offield et al., 1996). Similarly, after conditional

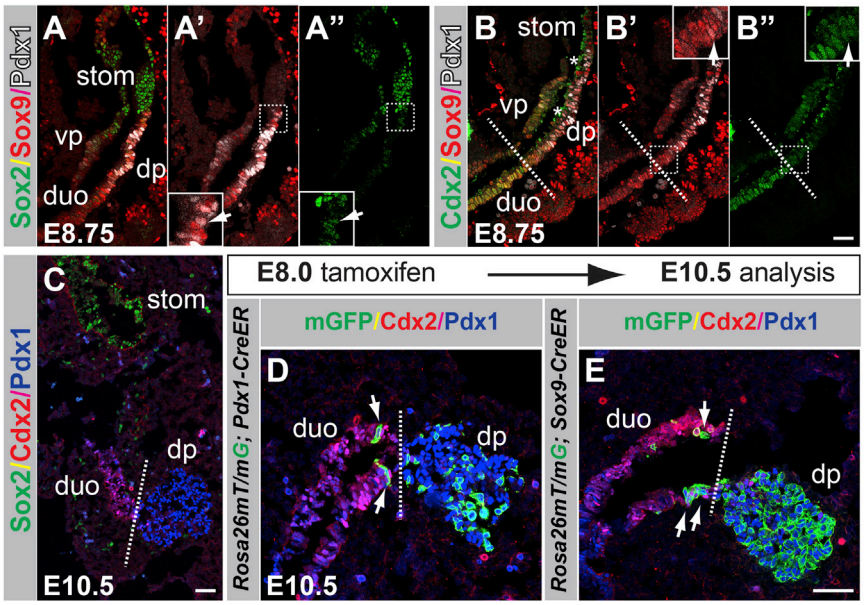



**Figure 2. Pdx1 and Sox9 Are Co-expressed in the Pancreatic Domain in the Foregut Endoderm**

(A–B″) Immunofluorescence staining for Sox2, Sox9, and Pdx1 (A–A″) and Cdx2, Sox9, and Pdx1 (B–B″) on embryonic sections at embryonic day (E) 8.75. The arrows in (A′) and (A″) and (B′) and (B″) indicate Pdx1$^{+}$/Sox9$^{+}$ cells co-expressing Sox2 and Cdx2, respectively. The dashed line in (B)–(B″) demarcates the transition from the presumptive duodenal to the pre-pancreatic region. Fields demarcated by white dashed boxes in (A′), (A″), (B′), and (B″) are shown at higher magnification in the same panels. Non-specific signal for Cdx2 is evident in the foregut lumen (B and B″, asterisks) due to antibody trapping.

(C) Immunofluorescence staining for Cdx2, Sox2, and Pdx1 at E10.5.

(D and E) Dams carrying *R26*$^{mT/mG}$ embryos expressing CreER driven by either the *Pdx1* or *Sox9* regulatory sequences were injected with tamoxifen at E8.0, embryos sectioned at E10.5, and immunofluorescence staining performed for Cdx2, Pdx1, and GFP. Recombined, membrane-targeted GFP$^{+}$ (mGFP$^{+}$) cells trace to the pancreatic epithelium; scattered labeled cells are also detectable in the proximal duodenum in *R26*$^{mT/mG}$*;Pdx1-CreER* (D) and *R26*$^{mT/mG}$*;Sox9-CreER* (E) embryos. dp, dorsal pancreas; vp, ventral pancreas; duo, duodenum; stom, stomach. Scale bars represent 50 μm.

*Sox9* inactivation with a *Pdx1-Cre* transgene pancreatic buds evaginate (Seymour et al., 2007, 2012). However, since *Pdx1-Cre* deletes *Sox9* after the pancreatic program has been initiated, it remains unclear whether Sox9 is necessary to initiate the pancreatic program. To determine whether *Sox9* is required for pancreatic specification, we generated global *Sox9* null mutant embryos (Figures 3A and 3C). While hypoplastic, dorsal and ventral pancreatic rudiments arise in *Sox9* null embryos (Figures 3B, 3B′, 3D, and 3D′), showing that *Sox9* is dispensable for pancreatic fate assignment and outgrowth of the pancreatic buds. Notably, although Pdx1 staining intensity is reduced, Pdx1 is expressed in both dorsal and ventral pancreatic buds of *Sox9*$^{-/-}$ embryos (Figures 3B, 3B′, 3D, and 3D′), showing that Sox9 is dispensable for Pdx1 induction. Similarly, we have previously found Sox9 to be expressed in Pdx1-deficient dorsal pancreatic progenitors at E10.5 (Seymour et al., 2012). Thus, neither Pdx1 nor Sox9 is required for pancreas specification or induction of the other's expression.

Based on their early expression in pre-pancreatic cells, we postulated that Sox9 and Pdx1 might function together and induce the pancreatic lineage in a cooperative manner. To test this, we generated mice lacking various combinations of either one or two alleles of *Pdx1*, *Sox9*, or both. Since early embryonic lethality of *Sox9* null embryos precluded the analysis of compound mutants beyond E11.5 (Akiyama et al., 2004), we employed a conditional *Sox9* ablation strategy, using the *Foxa3-Cre* transgenic line (Lee et al., 2005), which ablates *Sox9* efficiently in the gut tube by E9.5 (*Sox9*$^{\Delta gut}$) (Figures 3E–3H″).

We next generated compound mutants carrying various combinations of the *Pdx1* null (*Pdx1*$^{LacZko}$) and *Sox9*$^{\Delta gut}$ alleles and visualized the dorsal and ventral pancreatic buds, antral stomach, and duodenum by X-Gal staining for β-galactosidase (β-gal) expressed from the *Pdx1*$^{LacZko}$ allele (Figures 3I and 3J). With progressive loss of *Sox9* gene dosage (*Sox9*$^{+/+}$ > *Sox9*$^{+/\Delta gut}$ > *Sox9*$^{\Delta gut/\Delta gut}$) on the *Pdx1*-heterozygous mutant background, the pancreatic buds became increasingly hypoplastic (Figures 3I–3N). In E12.5 *Pdx1*$^{+/-}$*;Sox9*$^{\Delta gut/\Delta gut}$ embryos, the dorsal pancreas was reduced to a severely hypoplastic remnant, and the ventral pancreatic bud was undetectable (Figure 3N; absent ventral pancreas denoted by asterisk). Notably, the size of the ventral pancreatic bud was significantly reduced in compound-heterozygous mutants (Figures 3K, 3L, and 3U), which contrasted with the normal bud size seen in embryos deficient for a single copy of either *Pdx1* (Figures 3I and 3J) or *Sox9* (Seymour et al., 2008). This phenotype in compound-heterozygous mutants demonstrates genetic interaction between *Pdx1* and *Sox9*. The dorsal pancreas remnant (the ventral pancreas is undetectable in *Pdx1*$^{-/-}$ embryos; Figure 3P, asterisk) became increasingly smaller with decreasing *Sox9* gene dosage on a *Pdx1* null background (Figures 3O–3T) and was morphologically almost indiscernible in compound-homozygous *Pdx1*$^{-/-}$*;Sox9*$^{\Delta gut/\Delta gut}$ mutants (Figures 3S–3U). Combined, these genetic findings demonstrate cooperative functions of *Pdx1* and *Sox9* in early pancreas development.

To determine whether deletion of *Pdx1* and *Sox9* perturbs induction of the pancreatic program, we next analyzed the expression of early pancreatic markers in *Pdx1;Sox9* compound mutants. Confirming previous findings (Seymour et al., 2007, 2012), Sox9 expression was maintained in pancreatic rudiments of *Pdx1*$^{-/-}$ embryos at E10.5, and conversely, Pdx1 was also expressed in *Sox9*$^{\Delta gut/\Delta gut}$ mutants (Figures S1A–S1N; note that the truncated Pdx1 protein expressed from the *Pdx1* null allele is detected by the anti-Pdx1 antibody used). Immunofluorescence staining for Foxa2, Mnx1, Onecut-1, Tcf2, Gata4, and

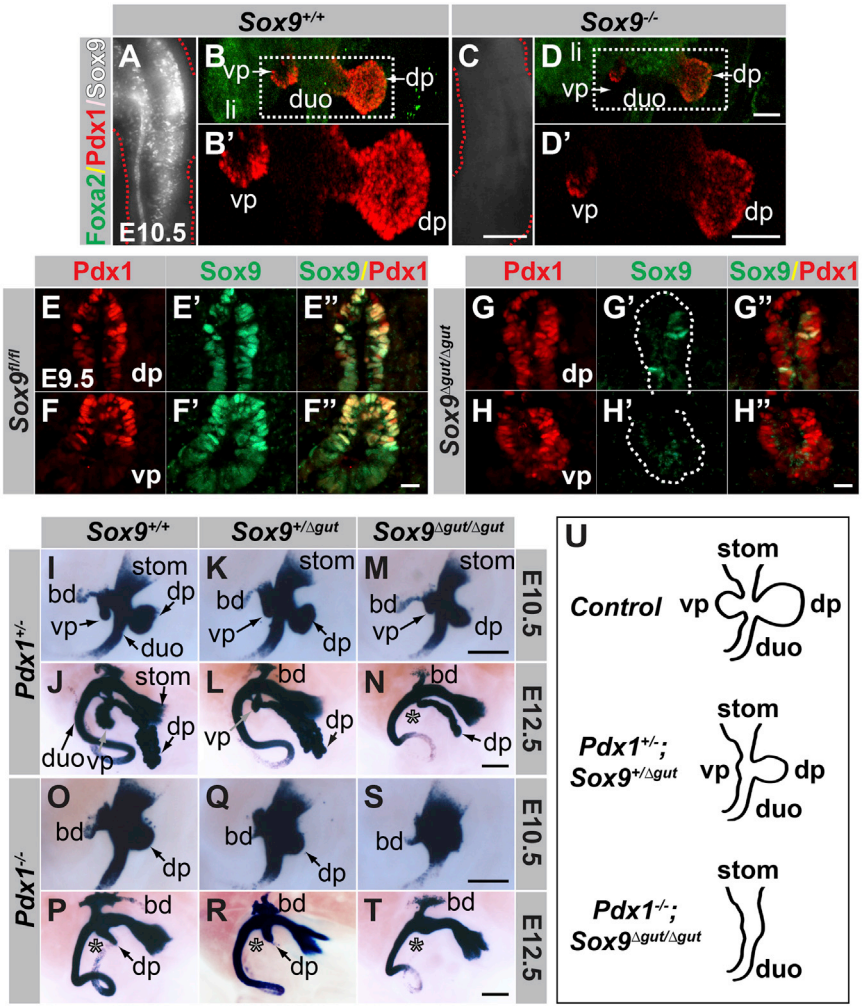



**Figure 3. Sox9 and Pdx1 Cooperatively Specify the Pancreatic Lineage**

(A and C) Confirmation of global *Sox9* deletion by whole mount immunofluorescence staining for Sox9 of tail tips from control (A) and *Sox9* null (C) embryos at E10.5.

(B and D) 2D projections of 3D Imaris-reconstructed z stacks through trunks of control (B) and *Sox9* null (D) embryos after whole mount immunofluorescence staining for Foxa2 and Pdx1. Although smaller, dorsal and ventral pancreatic buds are present in E10.5 *Sox9* null embryos (D and D′). Fields demarcated by white dashed boxes in (B) and (D) are shown at higher magnification in (B′) and (D′), respectively. Only single-channel Pdx1 signal is shown in (B′) and (D′).

(E–H″) Immunofluorescence staining of sections through the pancreatic region of control $Sox9^{fl/fl}$ (E–F″) and $Sox9^{fl/fl}$;*Foxa3-Cre* ($Sox9^{\Delta gut/\Delta gut}$; G–H″) embryos at E9.5. *Sox9* is efficiently deleted in dorsal (G′ and G″) and ventral (H′ and H″) pancreatic buds of $Sox9^{\Delta gut/\Delta gut}$ embryos. Dashed line in (G′) and (H′) demarcates the $Pdx1^+$ domain.

(I–T) X-Gal staining for β-galactosidase expressed from the $Pdx1^{LacZko}$ allele in E10.5 and E12.5 embryos carrying combinations of mutant alleles for *Pdx1* and *Sox9*. With increasing loss of *Sox9* dosage on either *Pdx1*-heterozygous (I–N) or *Pdx1* null (O–T) backgrounds, dorsal and ventral pancreatic buds become increasingly hypoplastic. In $Pdx1^{-/-}$;$Sox9^{\Delta gut/\Delta gut}$ embryos (S and T), pancreatic buds are not discernible. Note the reduced ventral pancreas in E12.5 compound heterozygous mutants (L). Asterisks denote absence of ventral pancreas.

(U) With decreasing dosage of functional *Pdx1* and *Sox9* alleles, pancreatic morphogenesis becomes increasingly perturbed.

dp, dorsal pancreas; vp, ventral pancreas; duo, duodenum; stom, stomach; li, liver; bd, bile duct. Scale bars represent 50 μm (E–H″), 70 μm (B, B′, D, and D′), 200 μm (A and C), and 250 μm (I–T).

Prox1 further revealed maintenance of their expression in embryos lacking either *Pdx1*, *Sox9*, or both (Figures S1O–S1BB and data not shown).

In contrast, expression of the pancreas-specific TF Ptf1a was drastically reduced in $Sox9^{\Delta gut/\Delta gut}$ and $Pdx1^{-/-}$;$Sox9^{\Delta gut/\Delta gut}$ embryos (Figures S1CC–S1II), showing that Ptf1a expression is Sox9-dependent. Albeit to a lesser extent, Ptf1a expression was also diminished in $Pdx1^{-/-}$ embryos (Figure S1HH). Like Ptf1a, the TF Nkx6.1 is pancreas-specific and, together with Ptf1a, governs the endocrine versus acinar cell fate choice (Schaffer et al., 2010). Nkx6.1 was not detected in $Pdx1^{-/-}$ and $Pdx1^{-/-}$;$Sox9^{\Delta gut/\Delta gut}$ embryos and was reduced in $Sox9^{\Delta gut/\Delta gut}$ embryos (Figures S1NN–S1PP). This confirms earlier findings in $Pdx1^{-/-}$ embryos (Pedersen et al., 2005) and suggests that Pdx1 is dominant over Sox9 in regulating Nkx6.1 expression. Together, our findings show that expression of the pancreas-restricted TFs Ptf1a and Nkx6.1 is under the control of Pdx1 and Sox9, whereas the expression of Foxa2, Mnx1, Onecut-1, Tcf2, Gata4, and Prox1 is Pdx1- and Sox9-independent.

## PDX1 and SOX9 Co-regulate Intestinal Cell Fate Determinants

To define the mechanistic basis of the observed cooperativity between Pdx1 and Sox9 in specifying the pancreatic fate, we mapped where PDX1 and SOX9 bind in the genome to explore synergy at the level of gene regulation. As the number of pancreatic progenitors in early mouse embryos is extremely limited, we generated pancreatic progenitors from hESCs (Xie et al., 2013) and performed chromatin immunoprecipitation and sequencing (ChIP-seq) analysis for PDX1 and SOX9. We mapped 55,481 unique binding peaks for PDX1 and 9,767 unique peaks for SOX9 (Figure 4A). PDX1 and SOX9 peaks exhibited surprisingly limited overlap (Figure 4B), which was unexpected given that lineage-determining TFs generally bind to *cis*-regulatory elements, in particular enhancers, as a collective unit (Spitz and Furlong, 2012). To understand the basis for the limited overlap in PDX1 and SOX9 binding sites, we analyzed PDX1 and SOX9 occupancy specifically at promoters and enhancers, using chromatin maps we recently generated based on histone modifications (Wang et al., 2015). This analysis revealed recruitment of

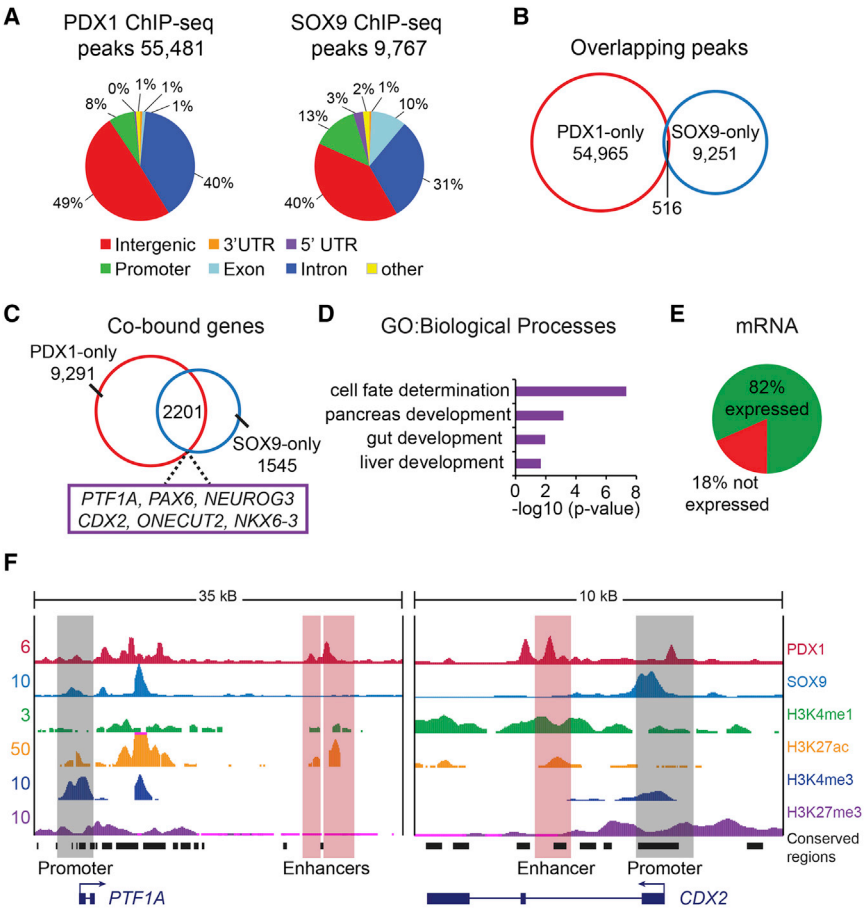



**Figure 4. PDX1 and SOX9 Co-occupy Pancreatic and Intestinal Genes**

(A) Genome-wide distribution of PDX1 and SOX9 binding peaks within the human genome from ChIP-seq analysis of hESC-derived pancreatic progenitors.

(B) Venn diagram of the overlap between PDX1 binding peaks and SOX9 binding peaks (minimum of 1-bp overlap).

(C) Venn diagram of the overlap between genes bound by PDX1 and SOX9, showing 2,201 genes to be co-bound by PDX1 and SOX9 (hypergeometric analysis: p value = $4.3\times10^{-9}$).

(D) Gene ontology (GO) analysis of PDX1 and SOX9 co-bound genes (defined as PDX1 and SOX9 binding at enhancers and/or promoters within a 200-kb window).

(E) Analysis of co-bound genes revealed that 82% of the co-bound genes are expressed, and 18% are not expressed in hESC-derived pancreatic progenitors.

(F) ChIP-seq binding profiles (reads per million) for PDX1, SOX9, and histone modifications (H3K4me1, H3K27ac, H3K4me3, and H3K27me3) at the *PTF1A* and *CDX2* loci in hESC-derived pancreatic progenitors. Enhancers were identified based on presence of H3K27ac and H3K4me1 and absence of H3K3me3. Black boxes indicate conserved regions in mice. kB, kilobases.

both PDX1 and SOX9 to promoters, albeit to not entirely overlapping sites (Figure S2A). Strikingly, and in stark contrast to PDX1, there was little recruitment of SOX9 to enhancers (Figure S2B). Other TFs with roles in early pancreatic development, such as FOXA2, ONECUT-1, and TCF2, occupied enhancers together with PDX1 (Figure S2B), consistent with TFs forming regulatory collectives at transcriptional enhancers (Calo and Wysocka, 2013). Together, these findings show that SOX9 is predominantly recruited to promoter regions, while PDX1 and other early pancreatic TFs co-occupy enhancers.

To relate PDX1 and SOX9 binding patterns to gene regulatory functions, we used the Genomic Regions Enrichment of Annotations Tool (GREAT) to predict putative target genes of PDX1-bound enhancers and then cataloged genes with binding peaks for PDX1 and SOX9 around transcriptional start sites and/or at PDX1-bound enhancers. This analysis identified 2,201 PDX1 and SOX9 co-bound genes (Figure 4C; Table S1). Consistent with the cooperative role of Pdx1 and Sox9 in pancreatic fate determination, regulators of pancreatic development are PDX1 and SOX9 co-bound, exemplified by the TFs *PTF1A*, *PAX6*, and *NEUROG3* (Figures 4C and 4F). Interestingly, PDX1 and SOX9 co-bound genes were enriched for Gene Ontology (GO) categories associated with cell developmental processes, including gut and liver development (Figure 4D). Occupancy of hepatic genes by PDX1 and SOX9 provides a possible explanation for why hepatic genes are ectopically expressed in Pdx1- and Sox9-deficient pancreatic buds (Seymour et al., 2012). PDX1 and SOX9 co-bound genes included several intestinal cell-fate-determining TFs, such as *CDX2*, *ONECUT-2*, and *NKX6-3* (Figures 4C and 4F) (Dusing et al., 2010; Nelson et al., 2005; Pedersen et al., 2005), suggesting a possible role for SOX9 and PDX1 in regulating these genes at the lineage bifurcation of pancreas and gut. Eighteen percent of all PDX1 and SOX9 co-bound genes were not expressed in pancreatic progenitors (Figure 4E), indicating that PDX1 and SOX9 could play a role in gene silencing. Combined, these results suggest cooperative roles for SOX9 and PDX1 in the regulation of pancreatic and intestinal genes.

Based on these findings, we predicted that decreased Pdx1 and Sox9 levels would induce ectopic activation of intestinal genes in the pancreatic domain. To test this, we identified co-regulated genes of both factors through transcriptional profiling of pancreatic progenitors from embryos with reduced *Pdx1* and *Sox9* gene dosage. Given that (1) both pancreatic buds are virtually absent in *Pdx1;Sox9* double-homozygous mutants and (2) evidence of genetic interaction in compound *Pdx1;Sox9* heterozygous mutants, we reasoned that mRNA profiling of pancreata from compound *Pdx1;Sox9* heterozygous mutants versus either single-heterozygous mutant could identify co-regulated genes. Hence, we performed cDNA microarray profiling of dorsal pancreatic epithelia from $Pdx1^{+/-}$, $Pdx1^{+/-};Sox9^{+/\Delta gut}$, and $Sox9^{+/\Delta gut}$ littermates at E12.5 when the epithelium is still predominantly composed of undifferentiated progenitor cells (Figure 5A).

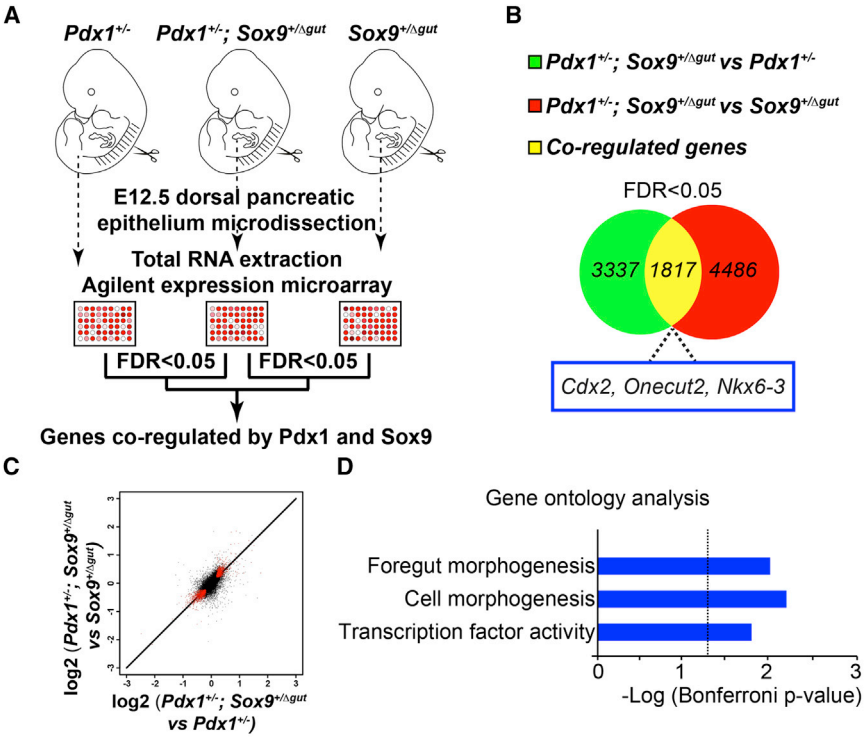



**Figure 5. *Pdx1* and *Sox9* Cooperatively Silence Genes Encoding Intestinal Cell Fate Regulators**

(A) Illustration of the experimental strategy for gene expression microarray analysis. The mRNA profiles of E12.5 pancreata (n = 12 per genotype) from (1) $Pdx1^{+/-}$ versus $Pdx1^{+/-};Sox9^{+/\Delta gut}$ and (2) $Sox9^{+/\Delta gut}$ versus $Pdx1^{+/-};Sox9^{+/\Delta gut}$ littermates were compared.

(B) A total of 3,337 and 4,486 genes were differentially expressed between (1) and (2), respectively. A total of 1,817 genes were common to both sets of significantly regulated genes (FDR < 0.05) with the same sign of change (i.e., upregulated or downregulated).

(C) Pdx1- and Sox9-co-regulated genes were identified by cross-comparing mRNA profiles of E12.5 pancreata (n = 12 per genotype) from (1) $Pdx1^{+/-}$ versus $Pdx1^{+/-};Sox9^{+/\Delta gut}$ and (2) $Sox9^{+/\Delta gut}$ versus $Pdx1^{+/-};Sox9^{+/\Delta gut}$ littermates. A total of 1,817 genes (denoted by red pixels) were common to both sets of significantly regulated genes (FDR < 0.05) with the same sign of change.

(D) Gene ontology analysis of the 1,817 Pdx1- and Sox9-co-regulated genes.

(E) The top 20 Pdx1- and Sox9-co-repressed genes with the highest fold change.

**E**

**Pdx1- and Sox9-co-repressed genes**
(**Up-regulated** genes in $Pdx1^{+/-};Sox9^{+/\Delta gut}$)

| Gene Symbol | FDR* | FDR** | Gene Symbol | FDR* | FDR** |
|---|---|---|---|---|---|
| *Prap1* | 1.8E-07 | 1.2E-10 | *Lars2* | 5.6E-03 | 3.7E-02 |
| *Sprr2f* | 7.0E-04 | 4.8E-03 | *Itih2* | 6.4E-06 | 6.5E-08 |
| *D14Ertd449e* | 3.2E-05 | 1.1E-04 | *F12* | 2.6E-07 | 1.7E-05 |
| *Alb1* | 1.8E-07 | 1.4E-07 | *C730009D12* | 1.2E-03 | 1.0E-07 |
| *Aqp8* | 9.7E-05 | 6.1E-05 | *Centd1* | 6.5E-03 | 2.2E-04 |
| *Cdx2* | 9.4E-04 | 1.5E-03 | *Myh3* | 7.2E-03 | 6.4E-05 |
| *Cfb* | 4.3E-07 | 8.6E-06 | *C330003B14* | 8.0E-05 | 1.9E-03 |
| *Onecut2* | 2.4E-05 | 3.6E-09 | *Nkx6-3* | 1.9E-02 | 2.3E-05 |
| *Akr1b7* | 4.7E-06 | 6.5E-08 | *Tll2* | 2.8E-03 | 6.0E-04 |
| *Prdm16* | 2.0E-03 | 1.3E-08 | *Aldob* | 2.3E-06 | 3.8E-05 |

* $Pdx1^{+/-};Sox9^{+/\Delta gut}$ vs $Pdx1^{+/-}$ ** $Pdx1^{+/-};Sox9^{+/\Delta gut}$ vs $Sox9^{+/\Delta gut}$

Comparison of gene expression profiles revealed significant differences in the expression of 3,337 genes (false discovery rate [FDR] < 0.05) between $Pdx1^{+/-};Sox9^{+/\Delta gut}$ and $Pdx1^{+/-}$ pancreatic epithelia and 4,486 genes (FDR < 0.05) between $Pdx1^{+/-};Sox9^{+/\Delta gut}$ and $Sox9^{+/\Delta gut}$ epithelia (Figure 5B; Tables S2 and S3). We then performed a cross-comparison of these two datasets in order to identify Pdx1- and Sox9-co-regulated genes. A total of 1,817 genes were common to both sets of significantly regulated genes with the same sign of change (i.e., upregulated or downregulated) (Figures 5B and 5C: co-regulated genes are denoted by red pixels in Figure 5C; Table S4) and associated with the GO term foregut morphogenesis (Figure 5D; Table S5). Intriguingly, among the top 20 Pdx1- and Sox9-co-repressed genes with the highest fold change were several genes encoding intestinal cell fate regulators, including *Cdx2*, *Onecut-2*, and *Nkx6.3* (Figure 5E), which also showed co-recruitment of PDX1 and SOX9 to their regulatory regions (Figures 4C and 4F; Table S1). These intestinal markers were all upregulated in pancreatic epithelia from compound *Pdx1;Sox9* heterozygous mutants, suggesting a synergistic and direct role for Pdx1 and Sox9 in repressing genes encoding intestinal lineage regulators.

## *Pdx1* and *Sox9* Jointly Control the Pancreatic versus Intestinal Cell Fate Choice

To determine whether Pdx1 and Sox9 indeed control the fate decision between pancreas and intestine, we analyzed the expression of the intestinal marker Cdx2 in the pancreatic region of embryos carrying various combinations of the *Pdx1* null and $Sox9^{\Delta gut}$ alleles. In control embryos at E10.5, cells of the dorsal pancreatic bud can be identified by high levels of Pdx1 expression, whereas prospective duodenal cells express the intestinal marker Cdx2 (Figures 6A–6A″ and 6P). At the duodenal-pancreatic junction, the $Pdx1^{high}$ domain forms a boundary with the $Cdx2^{+}$ domain; only a few $Pdx1^{high}$ cells express Cdx2 (Figures 6A–6A″ and 6P; note, duodenal precursors express low levels of Pdx1; Fukuda et al., 2006). As in control embryos, the $Pdx1^{high}$ and $Cdx2^{+}$ domains were distinct in embryos deficient for a single copy of either *Pdx1* or *Sox9*, compound *Pdx1;Sox9* heterozygous mutant embryos, and *Pdx1* or *Sox9* single-homozygous mutants (Figures 6B–6F″). In stark contrast, immunofluorescence staining for the truncated Pdx1 protein and Cdx2 in embryos with a combined homozygous deletion of *Pdx1* and *Sox9*

revealed extensive overlap between the Cdx2$^+$ and Pdx1$^+$ domains (Figures 6G–6G″, and 6P). Thus, the presence of either Pdx1 or Sox9 is sufficient to repress the intestinal marker Cdx2 in the pancreatic domain, whereas loss of both Pdx1 and Sox9 results in ectopic Cdx2 expression. In contrast, combined *Pdx1* and *Sox9* deletion did not result in ectopic expression of the stomach marker Sox2 in the Pdx1$^+$ domain (Figures S3A–S3D″), showing that Pdx1 and Sox9 cooperatively repress intestinal, but not anterior, foregut markers.

To directly test whether Pdx1 and Sox9 are sufficient to repress the intestinal fate in vivo, we forcibly expressed *Sox9* in *Pdx1*-expressing foregut progenitor cells, using a *Pdx1*-driven tetracycline transactivator mouse (*Pdx1*$^{tTA}$) and a single copy, tetracycline-regulated *Sox9* transgene (*mCherry-tetO-Sox9*) inserted into the disabled *Rosa26* locus (*Rosa26*$^{mCherry\text{-}tetO\text{-}Sox9}$) (Figure S3E). In this system, *Sox9* and the *mCherry* reporter gene are expressed in the Pdx1$^+$ domain in the absence of doxycycline; administration of doxycycline suppresses transgene expression. In *Pdx1*$^{tTA}$;*Rosa26*$^{mCherry\text{-}tetO\text{-}Sox9}$ (*Sox9*$^{GOF}$) embryos never exposed to doxycycline, Sox9 expression was enforced in Pdx1$^+$ cells of the pancreatic buds, antral stomach, and duodenum (Figures S3F–S3G″). In control embryos, Sox9 is detectable in the antral stomach and duodenum, but at much lower levels than in the pancreas (Figures S3F–S3F″). Formation of the pancreatic buds and gross gut morphology in *Sox9*$^{GOF}$ embryos were comparable to controls (Figures S3H–S3K).

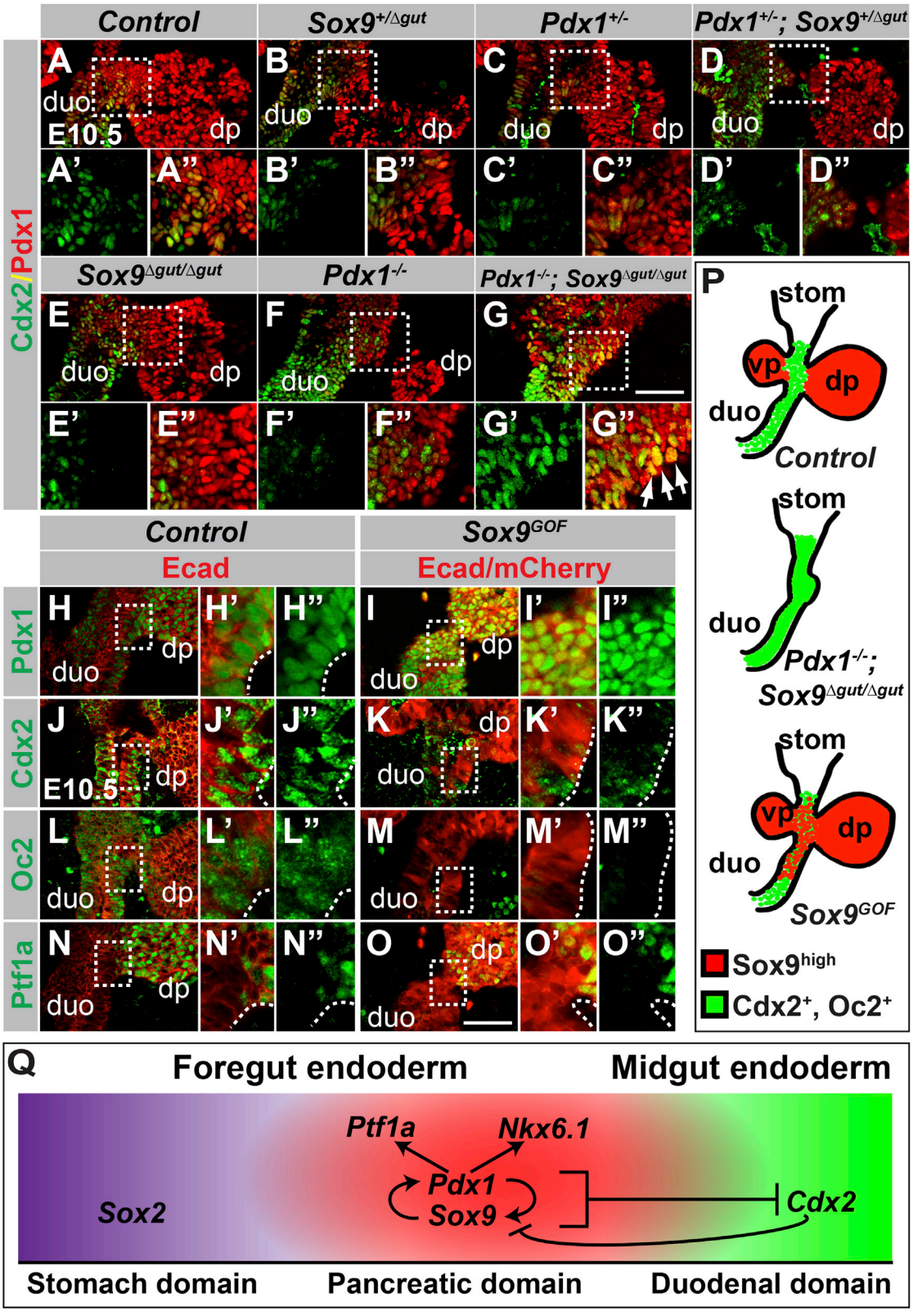
Consistent with previous observations that Sox9 reinforces Pdx1 expression (Dubois et al., 2011; Seymour et al., 2012), ectopic Sox9 expression resulted in increased Pdx1 staining intensity in the duodenal domain (Figures 6H–6I″), thus creating an extra-pancreatic Sox9$^{high}$/Pdx1$^{high}$ domain. In this domain, we observed reduced expression of the intestinal markers Cdx2 and Onecut-2, showing that the concerted activities of Pdx1 and Sox9 are sufficient to repress intestinal cell fate determinants (Figures 6J–6M″ and 6P). Notably, despite induction of a Pdx1$^{high}$ state and repression of intestinal markers in *Sox9*$^{GOF}$ embryos, Sox9 overexpression failed to induce Ptf1a in intestinal progenitors (Figures 6N–6O″). Previous work has shown that Ptf1a misexpression in the gut tube induces ectopic pancreas formation (Willet et al., 2014). Consistent with the lack of Ptf1a induction, an ectopic pancreatic bud was not observed in *Sox9*$^{GOF}$ embryos (Figures 6N–6O″). Combined, these results show that a Sox9$^{high}$/Pdx1$^{high}$ state prevents foregut endoderm progenitor cells from adopting intestinal lineage identity.

## DISCUSSION

In this study, we uncover a cooperative role for Pdx1 and Sox9 in governing the lineage choice between pancreas and intestine. Our data suggest a model whereby Pdx1 and Sox9 establish pancreatic lineage identity by excluding intestinal lineage-restricted TFs, such as Cdx2, from foregut endoderm progenitor cells (Figure 6Q). Our work further shows that the concerted activities of Pdx1 and Sox9 induce pancreatic differentiation programs through regulation of the pancreas-specific TFs Ptf1a and Nkx6.1. Interestingly, although the TFs Foxa2, Mnx1, Onecut-1, Tcf2, Gata4, and Prox1 are also important in early pancreas development (Seymour and Sander, 2011; Shih et al., 2013), their expression was not affected by combined *Pdx1* and *Sox9* deletion. These findings suggest that Sox9 and Pdx1 together are essential for driving pancreatic gene expression. The pancreatic program is reinforced by both positive autoregulation of Pdx1 (Marshak et al., 2000) and Sox9 (Lynn et al., 2007; Mead et al., 2013) and a positive cross-regulatory loop between Pdx1 and Sox9 (Dubois et al., 2011; Seymour et al., 2012). The mutual reinforcement of expression between Pdx1 and Sox9 appears to be direct, as PDX1 occupied *SOX9* regulatory sequences and vice versa (Figure S2C). Early pancreatic TFs induce a Notch$^{high}$ state that is important for maintaining the pancreatic state (Ahnfelt-Rønne et al., 2012; Jensen et al., 2000). For example, Sox9 and Ptf1a both promote expression of the Notch effector Hes1 in the early pancreas, and Hes1 in turn reinforces Ptf1a expression (Ahnfelt-Rønne et al., 2012).

Previous studies have shown that a subset of normally pancreas-fated cells adopt intestinal identity in *Ptf1a* null mutant mice (Kawaguchi et al., 2002). This invokes the question of how *Pdx1*, *Sox9*, and *Ptf1a* contribute to the gene regulatory network that establishes pancreatic identity and prevents foregut progenitors from becoming intestinal cells. Together with published observations, findings reported here identify Sox9 and Pdx1 as lying upstream of Ptf1a in the transcriptional regulatory cascade effecting pancreas induction (Figure 6Q). Several observations support this conclusion. First, combined deletion of *Pdx1* and *Ptf1a* phenocopies the effects of *Pdx1* deletion, arguing that Pdx1 is required prior to Ptf1a in pancreatic specification (Burlison et al., 2008). Second, we show that Ptf1a is not expressed in the absence of Sox9 (Figure S1GG), whereas Sox9 and Pdx1 induction do not depend on Ptf1a (Seymour et al., 2012). We note that Sox9 regulates Ptf1a only during pancreas specification, but not later in pancreas development, when the Sox9 and Ptf1a expression domains are distinct (Shih et al., 2012).

It is important to consider that after combined inactivation of *Pdx1* and *Ptf1a* in mice or *Xenopus*, the dorsal pancreatic bud still forms and early pancreatic genes are activated (Afelik et al., 2006; Burlison et al., 2008). Furthermore, we found that despite intestinal fate conversion of some Ptf1a-deficient cells (Kawaguchi et al., 2002), Cdx2 remains excluded from the pancreatic domain in *Ptf1a* null mutants (data not shown). These findings suggest that the pancreatic-to-intestinal boundary is still established in the absence of Pdx1 and Ptf1a. In contrast, we show that combined deletion of *Sox9* and *Pdx1* leads to misspecification of progenitors in the foregut endoderm, converting the pancreatic domain into a Cdx2-expressing intestinal domain (Figure 6G). Moreover, ectopic expression of Sox9 in duodenal precursors was sufficient to induce Pdx1 and repress Cdx2 (Figures 6I and 6K). These findings identify Sox9 as a critical early component of the gene regulatory network that governs both the activation of pancreatic genes and the repression of intestinal genes. Consistent with this notion, we found that SOX9 occupies genomic regions near genes required for early pancreatic development (i.e., *PTF1A*) as well as intestinal development (i.e., *CDX2*). Mechanistically, our data imply that Sox9 can function as either a transcriptional activator or repressor. Such a dual role for Sox9 is consistent with its ability to recruit both transcriptional




*(legend on next page)*

coactivators and corepressors (Lee et al., 2012; Leung et al., 2011).

Of interest is our finding that SOX9 and PDX1 bind to distinct *cis*-regulatory elements within the genome. While PDX1, FOXA2, ONECUT-1, and TCF2 collectively occupy enhancers, SOX9 was predominantly detected in promoter regions, suggesting a unique role for SOX9 in the regulation of gene expression. This observation could be relevant to gene regulatory mechanisms in multiple contexts, as Sox9 controls cell lineage decisions in several tissues, including gonad, lung, and kidney (Reginensi et al., 2011; Rockich et al., 2013; Sekido and Lovell-Badge, 2008). A future direction will be to test whether promoter-specific recruitment of Sox9 is also seen in other tissues and to determine how Sox9 deposition at promoters evokes cooperative effects with tissue-specific TFs bound to enhancers.

## EXPERIMENTAL PROCEDURES

### Mouse Strains

All animal experiments described herein were approved by the University of California San Diego Institutional Animal Care and Use Committees. The following mouse strains have been previously described: *Sox9*$^{flox}$ (Kist et al., 2002), *Pdx1*$^{LacZko}$ (herein designated *Pdx1*$^{-}$) (Offield et al., 1996), *Foxa3-Cre* (Lee et al., 2005), *Sox9-CreER* (Kopp et al., 2011), *Pdx1-CreER* (Gu et al., 2002), *Prm1-Cre* (O'Gorman et al., 1997), *Zp3-Cre* (de Vries et al., 2000), *Pdx1*$^{tTA}$ (Holland et al., 2002), and *R26*$^{mT/mG}$ (Muzumdar et al., 2007). To generate *Sox9* null mice, germline recombination of the *Sox9-flox* allele was employed as previously described (Akiyama et al., 2004). Briefly, *Sox9*$^{fl/+}$ mice were bred to carry either the oocyte-specific *Zp3-Cre* (de Vries et al., 2000) or the spermatid-specific *Prm1-Cre* (O'Gorman et al., 1997) transgenes. One *Sox9* allele was deleted in the oocytes or spermatids of *Zp3-Cre;* or *Prm1-Cre;Sox9*$^{fl/+}$ mice, respectively; these mice were then crossed to obtain *Sox9* null embryos. To generate *Rosa26*$^{mCherry\text{-}tetO\text{-}Sox9}$ mice, mouse *Sox9* coding sequences with MluI and NheI restriction sites on the 5′ and 3′ ends were generated from E15.5 pancreas by linker-primer PCR. The PCR product was then cloned into MluI and NheI sites of pBR322-hygro-ptight-mcherry, screened for orientation, and confirmed for bidirectionality (primers: Sox9-F MluI, 5′-tcacgcgtATGAATCTCCTGGACCCCTT-3′; Sox9-R NheI, 5′-ggctagcTCAGGGTCTGGTGAGCTGTGT-3′). The bidirectional *mCherry-tetO-Sox9* gene was inserted as a single copy transgene into a functionally disabled *Rosa26* gene locus using recombinase-mediated cassette exchange as previously described (Chen et al., 2011; Long et al., 2004). Mice bearing the *Rosa26*$^{mCherry\text{-}tetO\text{-}Sox9}$ allele were obtained after blastocyst microinjections, chimera matings, and FlpE-mediated removal of an FRT-flanked hygromycin resistance cassette.

A single dose of 2 mg/40 g body weight tamoxifen (Sigma) dissolved at 10 mg/ml in corn oil was administered by intraperitoneal injection. For each experiment, a minimum of three embryos per genotype was analyzed. Midday on the day of vaginal plug appearance was considered E0.5.

### Analysis of ChIP-Seq Data

Raw Illumina sequencing reads were mapped to reference human genomic database (version hg18) using Bowtie (version 1.1.0, http://bowtie-bio.sourceforge.net/index.shtml) to generate sam files. Sam files were subsequently converted to tag directories using HOMER (http://homer.salk.edu/homer/ngs/index.html). The ChIP-seq peak, peak distribution, and gene annotations were also annotated by HOMER analysis. Input sequencing data were used to normalize background reads for peak calling. Overlapping peaks were determined using the table browser function on the University of California Santa Cruz (UCSC) Genome Browser website, with minimum of 1-bp overlap. A 200-kb window was used to identify genes associated with the peaks.

Transcription factor binding to a promoter was determined by presence of a ChIP-seq peak within 20-kb upstream and 5-kb downstream of a transcriptional start site (TSS) of an annotated gene. Transcription factor binding to an enhancer was determined based on a minimum of 1-bp overlap between a transcription factor ChIP-seq peak and a predicted enhancer (defined as ± 500 bp from the center of the enhancer using the enhancer prediction tool; Rajagopal et al., 2013). We assigned PDX1-bound enhancers to nearest genes using GREAT (version 2.0, http://bejerano.stanford.edu/great/public/html/) with a basal plus 200-kb extension rule setting. In Figures 4C–4E, PDX1-bound genes were defined as genes with PDX1 binding at either promoters or enhancers corresponding to the gene. Since SOX9 did not exhibit significant enrichment at enhancers, SOX9-bound genes were defined as genes with SOX9 binding at promoters. Conserved regions were identified using the vista point tool comparing human to mouse (Frazer et al., 2004).

Gene ontology analysis was performed using the web tool DAVID Functional Annotation Bioinformatics Database (http://david.abcc.ncifcrf.gov/home.jsp) (Huang et al., 2009). The complete set of all RefSeq genes was used as a background.

ChIP-seq data for FOXA2, TCF2, and ONECUT-1 in hESC-derived pancreatic progenitors have been previously described (Weedon et al., 2014).

### Principal Component Analysis

The quality of the RNA sequencing data was analyzed using the FastQC v0.10.1 software. Once the samples passed quality control, they were aligned to the hg19 genome using RNA-Star 2.3.0e, with the parameters set to default. After alignment, Sailfish 0.6.3 and Cufflinks 2.2.0 were used to determine gene expression values. Datasets incorporating multivariate sequencing information (commonly gene expression values or splicing scores) were analyzed via the dimensionality reduction method principal component analysis (PCA) with the intention of uncovering features of the data that can explain variation within the dataset and as a visual summary of the sample data. The data were stored in pandas dataframes (pandas Python package v0.14.1) and visualized using Matplotlib v0.13.

A detailed description of all methods is available in the Supplemental Experimental Procedures.

**Figure 6. Pdx1 and Sox9 Are Necessary and Sufficient to Repress the Intestinal Lineage Choice**

(A–G) Immunofluorescence analysis for Pdx1 and Cdx2 on E10.5 embryos carrying various combinations of *Pdx1* and *Sox9* mutant alleles. In compound *Pdx1;Sox9* heterozygous mutant or *Pdx1* or *Sox9* single-homozygous mutant embryos, Cdx2 expression is restricted to duodenal precursors and excluded from the Pdx1$^{high}$ dorsal pancreas (A–F). In *Pdx1*$^{-/-}$*;Sox9*$^{\Delta gut/\Delta gut}$ embryos, a duodenal-pancreatic junction is not discernible, and Pdx1 and Cdx2 are co-expressed in a broad domain (arrows in G″).

(H–O) Immunofluorescence staining of sections from *Sox9*$^{GOF}$ and control littermates shows repression of the intestinal markers Cdx2 (J and K) and Onecut-2 (Oc2; L and M) in mCherry$^{+}$ duodenal precursors in *Sox9*$^{GOF}$ mice. Pdx1 is upregulated (H and I), but Ptf1a is not induced (N and O) in duodenal precursors in *Sox9*$^{GOF}$ embryos.

Fields demarcated by dashed boxes in (A)–(O) are shown at higher magnification in (A′)–(O″).

(P) Summary of the phenotypes observed after combined *Pdx1* and *Sox9* deletion or *Sox9* overexpression.

(Q) Graphical model summary. Our data support a model whereby Pdx1 and Sox9 cooperatively specify the pancreatic lineage by inducing the pancreatic transcription factors Nkx6.1 and Ptf1a and repressing the duodenal transcription factor *Cdx2*. A positive regulatory loop between Pdx1 and Sox9 maintains the pancreatic fate choice. Repression of Sox9 by Cdx2 creates bistability of the fate choice (Gao et al., 2009).

dp, dorsal pancreatic bud; vp, ventral pancreatic bud; duo, duodenum; stom, stomach. Scale bar represents 50 μm.

## ACCESSION NUMBERS

The NCBI GEO (http://www.ncbi.nlm.nih.gov/geo/) accession numbers for next generation sequencing data reported in this paper are GSE61945: human fetal pancreas transcriptome analysis; GSE61946: hESC-derived liver progenitor cell transcriptome analysis; GSE61947: SOX9 cistrome analysis in hESC-derived pancreatic progenitors. The accession number for the MIAME-Compliant Microarray Data set reported in this paper is NCBI GEO: GSE62023: identification of Sox9/Pdx1-coregulated genes during pancreas organogenesis.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, six tables, and one movie and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.08.082.

## AUTHOR CONTRIBUTIONS

H.P.S., P.A.S., and M.S. conceived the project. H.P.S., P.A.S., and M.S. designed the experiments and analyzed the data. H.P.S. and P.A.S. performed all analyses of mouse genetic models. R.X. and A.W. performed ChIP-seq experiments. N.A.P. and A.W. analyzed ChIP-seq data. P.P.L. and G.W.Y. performed PCA. M.A.M. designed and generated the *Rosa26*$^{mCherry\text{-}tetO\text{-}Sox9}$ mouse strain. H.P.S., P.A.S., and M.S. wrote the manuscript.

## ACKNOWLEDGMENTS

We are grateful to G. Scherer, K. Kaestner, R. MacDonald, D. Melton, and C. Wright for mouse strains, and R. Behringer and H. Chang for *Sox9* null embryos. We also thank M. Wegner, C. Wright, B. Bréant, J. Kehrl, and F. Lemaigre for antibodies, M. Jørgensen for expert assistance with confocal imaging, B. Armstrong for expert assistance with 3D image rendering, and P. Serup for support. We acknowledge the Core Facility for Integrated Microscopy, Faculty of Health and Medical Sciences, University of Copenhagen. We acknowledge the support of the UC San Diego Biogem Core Facility for microarray analyses, the University of Pennsylvania Functional Genomics Core for ChIP-seq analysis, Q. Zhang and the Vanderbilt Transgenic Mouse/ESC Shared Resource for deriving *Rosa26*$^{mCherry\text{-}tetO\text{-}Sox9}$ mice, and the Washington Birth Defects Research Laboratory for human fetal pancreas. We thank members of the Sander laboratory for constructive comments on the manuscript. We apologize to our colleagues whose references were omitted because of space constraints. This work was supported by NIH/NIDDK awards to M.S. (DK078803, DK68471, and DK089567) and M.A.M. (DK89523), JDRF postdoctoral fellowships to P.A.S. (3-2004-608), H.P.S. (3-2009-161), A.W. (3-2012-177), and the California Institute for Regenerative Medicine training grant TG2-01154 to N.A.P and R.X.

Received: February 2, 2015
Revised: July 10, 2015
Accepted: August 29, 2015
Published: October 1, 2015

## REFERENCES

Afelik, S., Chen, Y., and Pieler, T. (2006). Combined ectopic expression of Pdx1 and Ptf1a/p48 results in the stable conversion of posterior endoderm into endocrine and exocrine pancreatic tissue. Genes Dev. *20*, 1441–1446.

Ahnfelt-Rønne, J., Jørgensen, M.C., Klinck, R., Jensen, J.N., Füchtbauer, E.M., Deering, T., MacDonald, R.J., Wright, C.V., Madsen, O.D., and Serup, P. (2012). Ptf1a-mediated control of Dll1 reveals an alternative to the lateral inhibition mechanism. Development *139*, 33–45.

Akiyama, H., Chaboissier, M.C., Behringer, R.R., Rowitch, D.H., Schedl, A., Epstein, J.A., and de Crombrugghe, B. (2004). Essential role of Sox9 in the pathway that controls formation of cardiac valves and septa. Proc. Natl. Acad. Sci. USA *101*, 6502–6507.

Burlison, J.S., Long, Q., Fujitani, Y., Wright, C.V., and Magnuson, M.A. (2008). Pdx-1 and Ptf1a concurrently determine fate specification of pancreatic multipotent progenitor cells. Dev. Biol. *316*, 74–86.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol. Cell *49*, 825–837.

Carrasco, M., Delgado, I., Soria, B., Martín, F., and Rojas, A. (2012). GATA4 and GATA6 control mouse pancreas organogenesis. J. Clin. Invest. *122*, 3504–3515.

Chen, S.X., Osipovich, A.B., Ustione, A., Potter, L.A., Hipkens, S., Gangula, R., Yuan, W., Piston, D.W., and Magnuson, M.A. (2011). Quantification of factors influencing fluorescent protein expression using RMCE to generate an allelic series in the ROSA26 locus in mice. Dis. Model. Mech. *4*, 537–547.

de Vries, W.N., Binns, L.T., Fancher, K.S., Dean, J., Moore, R., Kemler, R., and Knowles, B.B. (2000). Expression of Cre recombinase in mouse oocytes: a means to study maternal effect genes. Genesis *26*, 110–112.

Dubois, C.L., Shih, H.P., Seymour, P.A., Patel, N.A., Behrmann, J.M., Ngo, V., and Sander, M. (2011). Sox9-haploinsufficiency causes glucose intolerance in mice. PLoS ONE *6*, e23131.

Dusing, M.R., Maier, E.A., Aronow, B.J., and Wiginton, D.A. (2010). Onecut-2 knockout mice fail to thrive during early postnatal period and have altered patterns of gene expression in small intestine. Physiol. Genomics *42*, 115–125.

Frazer, K.A., Pachter, L., Poliakov, A., Rubin, E.M., and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. Nucleic Acids Res. *32*, W273–279.

Fukuda, A., Kawaguchi, Y., Furuyama, K., Kodama, S., Horiguchi, M., Kuhara, T., Koizumi, M., Boyer, D.F., Fujimoto, K., Doi, R., et al. (2006). Ectopic pancreas formation in Hes1 -knockout mice reveals plasticity of endodermal progenitors of the gut, bile duct, and pancreas. J. Clin. Invest. *116*, 1484–1493.

Gao, N., White, P., and Kaestner, K.H. (2009). Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. Dev. Cell *16*, 588–599.

Gu, G., Dubauskaite, J., and Melton, D.A. (2002). Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. Development *129*, 2447–2457.

Harrison, K.A., Thaler, J., Pfaff, S.L., Gu, H., and Kehrl, J.H. (1999). Pancreas dorsal lobe agenesis and abnormal islets of Langerhans in Hlxb9-deficient mice. Nat. Genet. *23*, 71–75.

Haumaitre, C., Barbacci, E., Jenny, M., Ott, M.O., Gradwohl, G., and Cereghini, S. (2005). Lack of TCF2/vHNF1 in mice leads to pancreas agenesis. Proc. Natl. Acad. Sci. USA *102*, 1490–1495.

Holland, A.M., Hale, M.A., Kagami, H., Hammer, R.E., and MacDonald, R.J. (2002). Experimental control of pancreatic development and maintenance. Proc. Natl. Acad. Sci. USA *99*, 12236–12241.

Holmberg, J., and Perlmann, T. (2012). Maintaining differentiated cellular identity. Nat. Rev. Genet. *13*, 429–439.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Jacquemin, P., Durviaux, S.M., Jensen, J., Godfraind, C., Gradwohl, G., Guillemot, F., Madsen, O.D., Carmeliet, P., Dewerchin, M., Collen, D., et al. (2000). Transcription factor hepatocyte nuclear factor 6 regulates pancreatic endocrine cell differentiation and controls expression of the proendocrine gene ngn3. Mol. Cell. Biol. *20*, 4445–4454.

Jensen, J., Pedersen, E.E., Galante, P., Hald, J., Heller, R.S., Ishibashi, M., Kageyama, R., Guillemot, F., Serup, P., and Madsen, O.D. (2000). Control of endodermal endocrine development by Hes-1. Nat. Genet. *24*, 36–44.

Kawaguchi, Y., Cooper, B., Gannon, M., Ray, M., MacDonald, R.J., and Wright, C.V. (2002). The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors. Nat. Genet. *32*, 128–134.

Kist, R., Schrewe, H., Balling, R., and Scherer, G. (2002). Conditional inactivation of Sox9: a mouse model for campomelic dysplasia. Genesis *32*, 121–123.

Kopp, J.L., Dubois, C.L., Schaffer, A.E., Hao, E., Shih, H.-P., Seymour, P.A., Ma, J., and Sander, M. (2011). Sox9+ ductal cells are multipotent progenitors

throughout development but do not produce new endocrine cells in the normal or injured adult pancreas. Development *138*, 653–665.

Lee, C.S., Sund, N.J., Behr, R., Herrera, P.L., and Kaestner, K.H. (2005). Foxa2 is required for the differentiation of pancreatic alpha-cells. Dev. Biol. *278*, 484–495.

Lee, P.C., Taylor-Jaffe, K.M., Nordin, K.M., Prasad, M.S., Lander, R.M., and LaBonne, C. (2012). SUMOylated SoxE factors recruit Grg4 and function as transcriptional repressors in the neural crest. J. Cell Biol. *198*, 799–813.

Leung, V.Y., Gao, B., Leung, K.K., Melhado, I.G., Wynn, S.L., Au, T.Y., Dung, N.W., Lau, J.Y., Mak, A.C., Chan, D., and Cheah, K.S. (2011). SOX9 governs differentiation stage-specific gene expression in growth plate chondrocytes via direct concomitant transactivation and repression. PLoS Genet. *7*, e1002356.

Long, Q., Shelton, K.D., Lindner, J., Jones, J.R., and Magnuson, M.A. (2004). Efficient DNA cassette exchange in mouse embryonic stem cells by staggered positive-negative selection. Genesis *39*, 256–262.

Lynn, F.C., Smith, S.B., Wilson, M.E., Yang, K.Y., Nekrep, N., and German, M.S. (2007). Sox9 coordinates a transcriptional network in pancreatic progenitor cells. Proc. Natl. Acad. Sci. USA *104*, 10500–10505.

Marshak, S., Benshushan, E., Shoshkes, M., Havin, L., Cerasi, E., and Melloul, D. (2000). Functional conservation of regulatory elements in the pdx-1 gene: PDX-1 and hepatocyte nuclear factor 3beta transcription factors mediate beta-cell-specific expression. Mol. Cell. Biol. *20*, 7583–7590.

McCracken, K.W., Catá, E.M., Crawford, C.M., Sinagoga, K.L., Schumacher, M., Rockich, B.E., Tsai, Y.H., Mayhew, C.N., Spence, J.R., Zavros, Y., and Wells, J.M. (2014). Modelling human development and disease in pluripotent stem-cell-derived gastric organoids. Nature *516*, 400–404.

Mead, T.J., Wang, Q., Bhattaram, P., Dy, P., Afelik, S., Jensen, J., and Lefebvre, V. (2013). A far-upstream (-70 kb) enhancer mediates Sox9 auto-regulation in somatic tissues during development and adult regeneration. Nucleic Acids Res. *41*, 4459–4469.

Muzumdar, M.D., Tasic, B., Miyamichi, K., Li, L., and Luo, L. (2007). A global double-fluorescent Cre reporter mouse. Genesis *45*, 593–605.

Nelson, S.B., Janiesch, C., and Sander, M. (2005). Expression of Nkx6 genes in the hindbrain and gut of the developing mouse. J. Histochem. Cytochem. *53*, 787–790.

O'Gorman, S., Dagenais, N.A., Qian, M., and Marchuk, Y. (1997). Protamine-Cre recombinase transgenes efficiently recombine target sequences in the male germ line of mice, but not in embryonic stem cells. Proc. Natl. Acad. Sci. USA *94*, 14602–14607.

Offield, M.F., Jetton, T.L., Labosky, P.A., Ray, M., Stein, R.W., Magnuson, M.A., Hogan, B.L., and Wright, C.V. (1996). PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. Development *122*, 983–995.

Pedersen, J.K., Nelson, S.B., Jorgensen, M.C., Henseleit, K.D., Fujitani, Y., Wright, C.V., Sander, M., and Serup, P.; Beta Cell Biology Consortium (2005). Endodermal expression of Nkx6 genes depends differentially on Pdx1. Dev. Biol. *288*, 487–501.

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 Comput. Biol. *9*, e1002968.

Reginensi, A., Clarkson, M., Neirijnck, Y., Lu, B., Ohyama, T., Groves, A.K., Sock, E., Wegner, M., Costantini, F., Chaboissier, M.C., and Schedl, A. (2011). SOX9 controls epithelial branching by activating RET effector genes during kidney development. Hum. Mol. Genet. *20*, 1143–1153.

Rockich, B.E., Hrycaj, S.M., Shih, H.P., Nagy, M.S., Ferguson, M.A., Kopp, J.L., Sander, M., Wellik, D.M., and Spence, J.R. (2013). Sox9 plays multiple roles in the lung epithelium during branching morphogenesis. Proc. Natl. Acad. Sci. USA *110*, E4456–4464.

Schaffer, A.E., Freude, K.K., Nelson, S.B., and Sander, M. (2010). Nkx6 transcription factors and Ptf1a function as antagonistic lineage determinants in multipotent pancreatic progenitors. Dev. Cell *18*, 1022–1029.

Sekido, R., and Lovell-Badge, R. (2008). Sex determination involves synergistic action of SRY and SF1 on a specific Sox9 enhancer. Nature *453*, 930–934.

Seymour, P.A., and Sander, M. (2011). Historical perspective: beginnings of the beta-cell: current perspectives in beta-cell development. Diabetes *60*, 364–376.

Seymour, P.A., Freude, K.K., Tran, M.N., Mayes, E.E., Jensen, J., Kist, R., Scherer, G., and Sander, M. (2007). SOX9 is required for maintenance of the pancreatic progenitor cell pool. Proc. Natl. Acad. Sci. USA *104*, 1865–1870.

Seymour, P.A., Freude, K.K., Dubois, C.L., Shih, H.P., Patel, N.A., and Sander, M. (2008). A dosage-dependent requirement for Sox9 in pancreatic endocrine cell formation. Dev. Biol. *323*, 19–30.

Seymour, P.A., Shih, H.P., Patel, N.A., Freude, K.K., Xie, R., Lim, C.J., and Sander, M. (2012). A Sox9/Fgf feed-forward loop maintains pancreatic organ identity. Development *139*, 3363–3372.

Sherwood, R.I., Chen, T.Y., and Melton, D.A. (2009). Transcriptional dynamics of endodermal organ formation. Dev. Dyn. *238*, 29–42.

Shih, H.P., Kopp, J.L., Sandhu, M., Dubois, C.L., Seymour, P.A., Grapin-Botton, A., and Sander, M. (2012). A Notch-dependent molecular circuitry initiates pancreatic endocrine and ductal cell differentiation. Development *139*, 2488–2499.

Shih, H.P., Wang, A., and Sander, M. (2013). Pancreas organogenesis: from lineage determination to morphogenesis. Annu. Rev. Cell Dev. Biol. *29*, 81–105.

Spitz, F., and Furlong, E.E. (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Verzi, M.P., Shin, H., Ho, L.L., Liu, X.S., and Shivdasani, R.A. (2011). Essential and redundant functions of caudal family proteins in activating adult intestinal genes. Mol. Cell. Biol. *31*, 2026–2039.

Wang, J., Kilic, G., Aydin, M., Burke, Z., Oliver, G., and Sosa-Pineda, B. (2005). Prox1 activity controls pancreas morphogenesis and participates in the production of "secondary transition" pancreatic endocrine cells. Dev. Biol. *286*, 182–194.

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

Weedon, M.N., Cebola, I., Patch, A.M., Flanagan, S.E., De Franco, E., Caswell, R., Rodríguez-Seguí, S.A., Shaw-Smith, C., Cho, C.H., Lango Allen, H., et al.; International Pancreatic Agenesis Consortium (2014). Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. Nat. Genet. *46*, 61–64.

Willet, S.G., Hale, M.A., Grapin-Botton, A., Magnuson, M.A., MacDonald, R.J., and Wright, C.V. (2014). Dominant and context-specific control of endodermal organ allocation by Ptf1a. Development *141*, 4385–4394.

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.

Xuan, S., Borok, M.J., Decker, K.J., Battle, M.A., Duncan, S.A., Hale, M.A., Macdonald, R.J., and Sussel, L. (2012). Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic agenesis. J. Clin. Invest. *122*, 3516–3528.

Zaret, K.S. (2008). Genetic programming of liver and pancreas progenitors: lessons for stem-cell differentiation. Nat. Rev. Genet. *9*, 329–340.

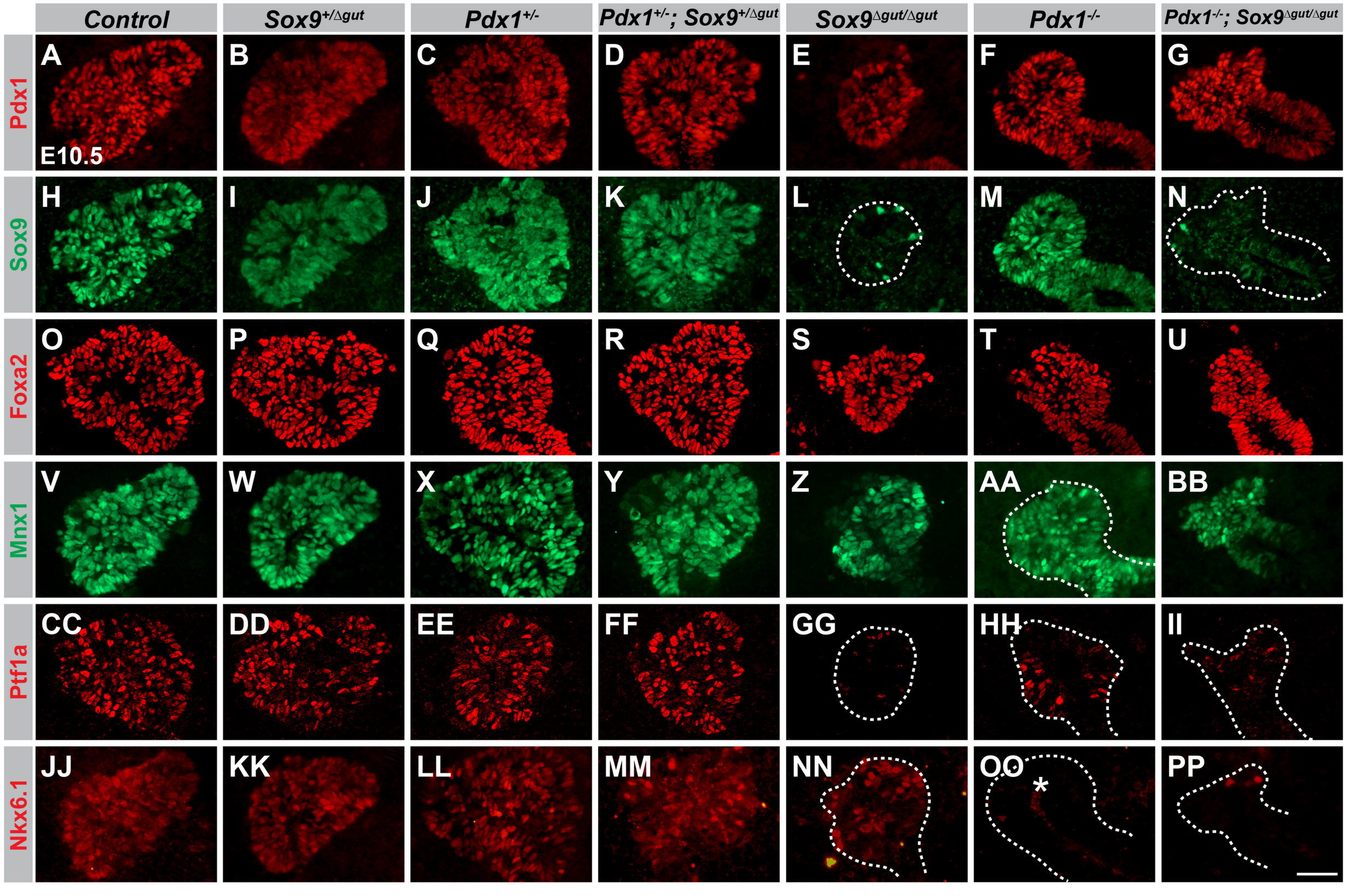
**Cell Reports**

**Supplemental Information**

# A Gene Regulatory Network Cooperatively Controlled by Pdx1 and Sox9 Governs Lineage Allocation of Foregut Progenitor Cells

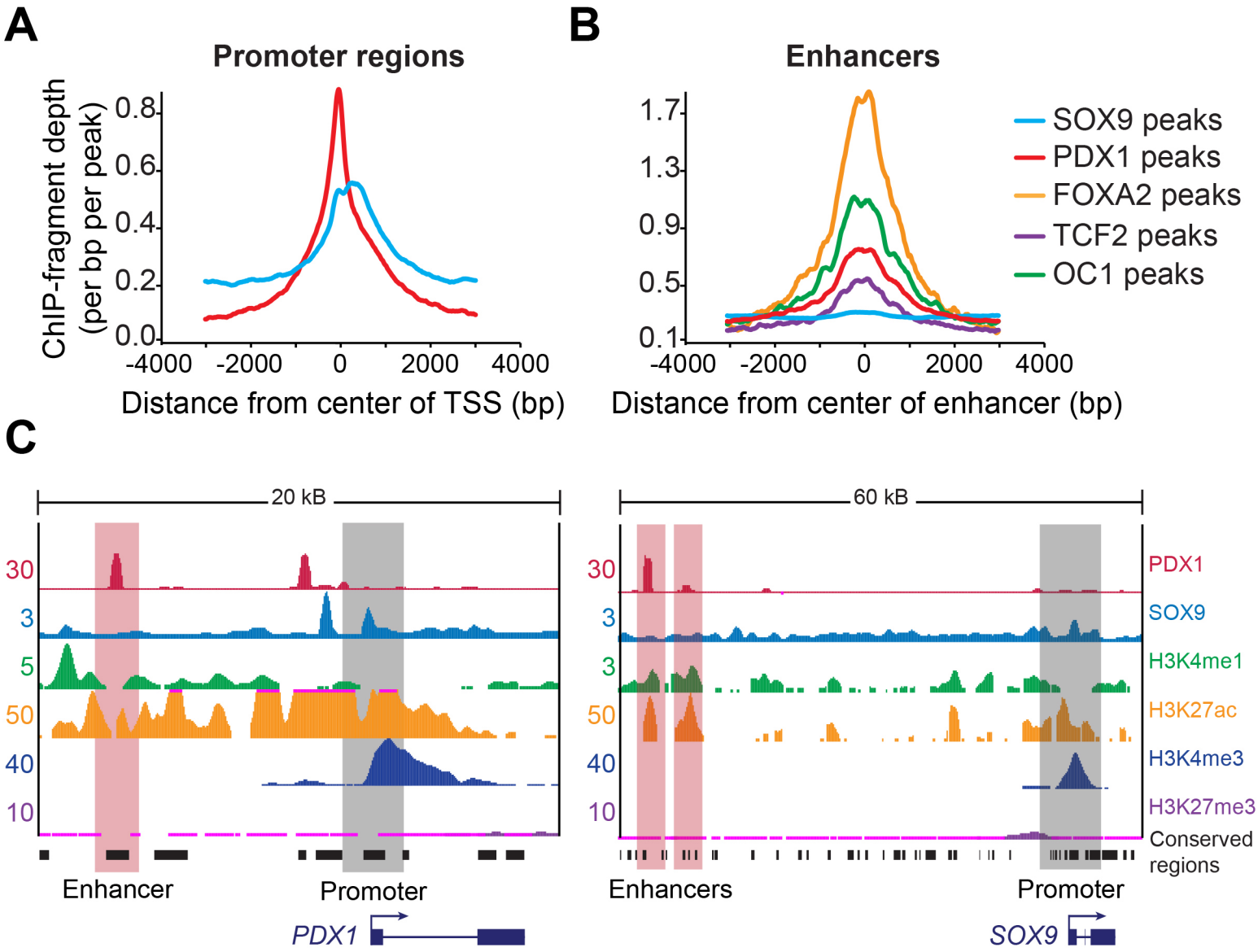
**Hung Ping Shih, Philip A. Seymour, Nisha A. Patel, Ruiyu Xie, Allen Wang, Patrick P. Liu, Gene W. Yeo, Mark A. Magnuson, and Maike Sander**

# Shih_Figure S1. Related to Figure 3



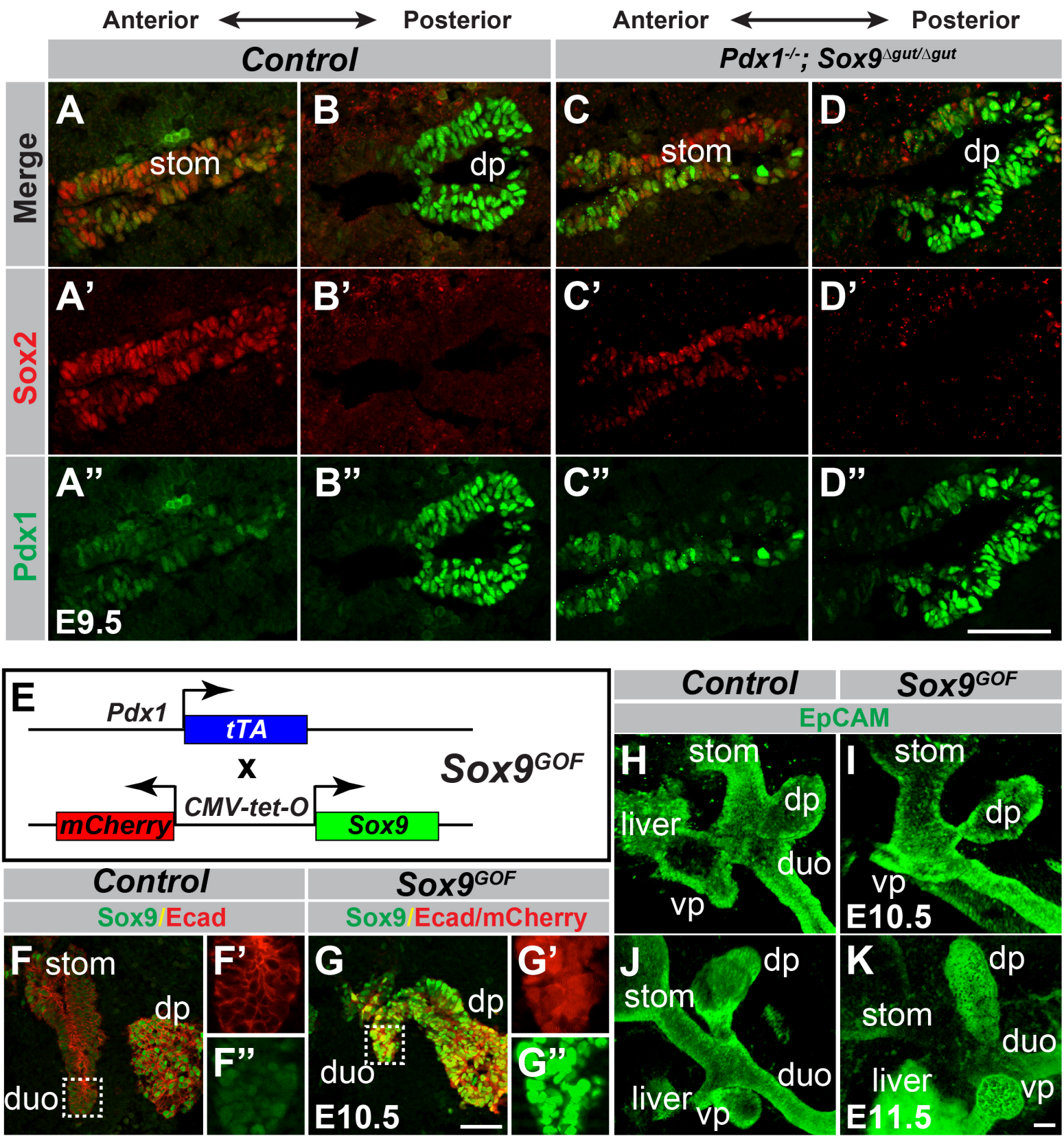
**Supplemental Figure S1, Related to Figure 3. Combined loss of *Pdx1* and *Sox9* abrogates the early pancreatic program.** (**A-PP**) Immunofluorescence analysis for Pdx1 (**A-G**), Sox9 (**H-N**), Foxa2 (**O-U**), Mnx1 (**V-BB**), Ptf1a (**CC-II**), and Nkx6.1 (**JJ-PP**) on embryonic day (E) 10.5 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.