Stem Cell Reports
Report
ISSCR
OPEN ACCESS

Loss of TBX3 enhances pancreatic progenitor generation from human pluripotent stem cells

Somdutta Mukherjee,[1,2] Deborah L. French,[2,3] and Paul Gadue[2,3,*]
[1]Graduate Program in Cell and Molecular Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
[2]Center for Cellular and Molecular Therapeutics, The Children's Hospital of Philadelphia, Philadelphia, PA, USA
[3]Department of Pathology and Laboratory Medicine, The Children's Hospital of Philadelphia and the Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
*Correspondence: gaduep@chop.edu
https://doi.org/10.1016/j.stemcr.2021.09.004

SUMMARY

Tbx3 has been identified as a regulator of liver development in the mouse, but its function in human liver development remains unknown. *TBX3* mutant human pluripotent stem cell (PSC) lines were generated using CRISPR/Cas9 genome editing. TBX3 loss led to impaired liver differentiation and an upregulation of pancreatic gene expression, including *PDX1*, during a hepatocyte differentiation protocol. Other pancreatic genes, including *NEUROG3* and *NKX2.2*, displayed more open chromatin in the *TBX3* mutant hepatoblasts. Using a pancreatic differentiation protocol, cells lacking *TBX3* generated more pancreatic progenitors and had an enhanced pancreatic gene expression signature at the expense of hepatic gene expression. These data highlight a potential role of TBX3 in regulating hepatic and pancreatic domains during foregut patterning, with implications for enhancing the generation of pancreatic progenitors from PSCs.

INTRODUCTION

The liver and pancreas play vital roles in metabolism and digestion. Both organs arise from the posterior foregut region of the developing gut tube (Deutsch et al., 2001). The gut tube is characterized by an expression pattern of different transcription factors that specify distinct domains for each of the endodermal organs. While the hepatic and pancreatic domains of the gut tube lie in close proximity to each other, they are specified by different signals. Fibroblast growth factor (FGF) from the cardiac mesoderm (Jung et al., 1999) and bone morphogenetic protein 4 from the septum transversum specify the hepatic endoderm (Rossi et al., 2001). Activin and FGF2 signals from the notochord, and retinoic acid from the lateral plate mesoderm specify the dorsal pancreatic endoderm (Hebrok et al., 1998).

T-box transcription factor 3 (TBX3) is involved in development of a number of model systems. Tbx3 drives mesendoderm in *Xenopus* embryos and mouse pluripotent stem cells (PSCs) (Weidgang et al., 2013), and regulates limb and heart development (Gibson-Brown et al., 1998; Singh et al., 2012). TBX3 is a member of the T-box gene family that acts primarily as a transcriptional repressor (Carlson et al., 2001). In humans, heterozygous mutations in *TBX3* result in Ulnar-Mammary Syndrome, a disorder causing defects in limb, mammary, and apocrine gland development (Bamshad et al., 1997). *Tbx3* has also been implicated in liver development. *Tbx3* is expressed in hepatoblasts, bipotential progenitors that give rise to hepatocytes and cholangiocytes, which comprise the liver bud. *Tbx3*$^{-/-}$ mice have small liver buds that fail to expand and mature (Lüdtke et al., 2009). *Tbx3* has been detected in the mouse pancreas, both during development and in the adult, but its function remains unknown (Begum and Papaioannou, 2011). The role of TBX3 in human liver and pancreas development remains unclear.

PSCs, such as human embryonic stem cells (ESC) and induced PSCs (iPSCs) can give rise to all cell types, and are a model system for studying human development, physiology, and disease. PSCs are easily manipulated using CRISPR/Cas9 genome editing technology, making them valuable for studying the role of specific genes in development and disease (Maguire et al., 2019). Protocols to differentiate PSCs to both hepatocytes (Ogawa et al., 2013; Si-Tayeb et al., 2010) and pancreatic β-cells (D'Amour et al., 2006; Nostro et al., 2011; Rezania et al., 2014) have improved drastically over time. However, it is still difficult to generate functionally mature terminal cell types *in vitro*.

Here we use PSCs to study the role of TBX3 in human liver and pancreas development. *TBX3* mutant PSC lines were generated and differentiated to hepatocytes and pancreatic progenitors. The loss of *TBX3* caused a defect in hepatocyte differentiation. Interestingly, the loss of *TBX3* resulted in the expression of Pancreatic and Duodenal Homeobox 1 (*PDX1*), a master regulator for pancreas development (Ahlgren et al., 1996). PSCs lacking *TBX3* differentiated more efficiently to pancreatic progenitors than wild-type PSCs. These data suggest that TBX3 may regulate liver development through suppression of pancreatic genes. A better understanding of how TBX3 regulates pancreatic precursor efficiency may provide a potential avenue to improve generation of *in vitro*–derived pancreatic β-cells for use in therapeutic contexts.

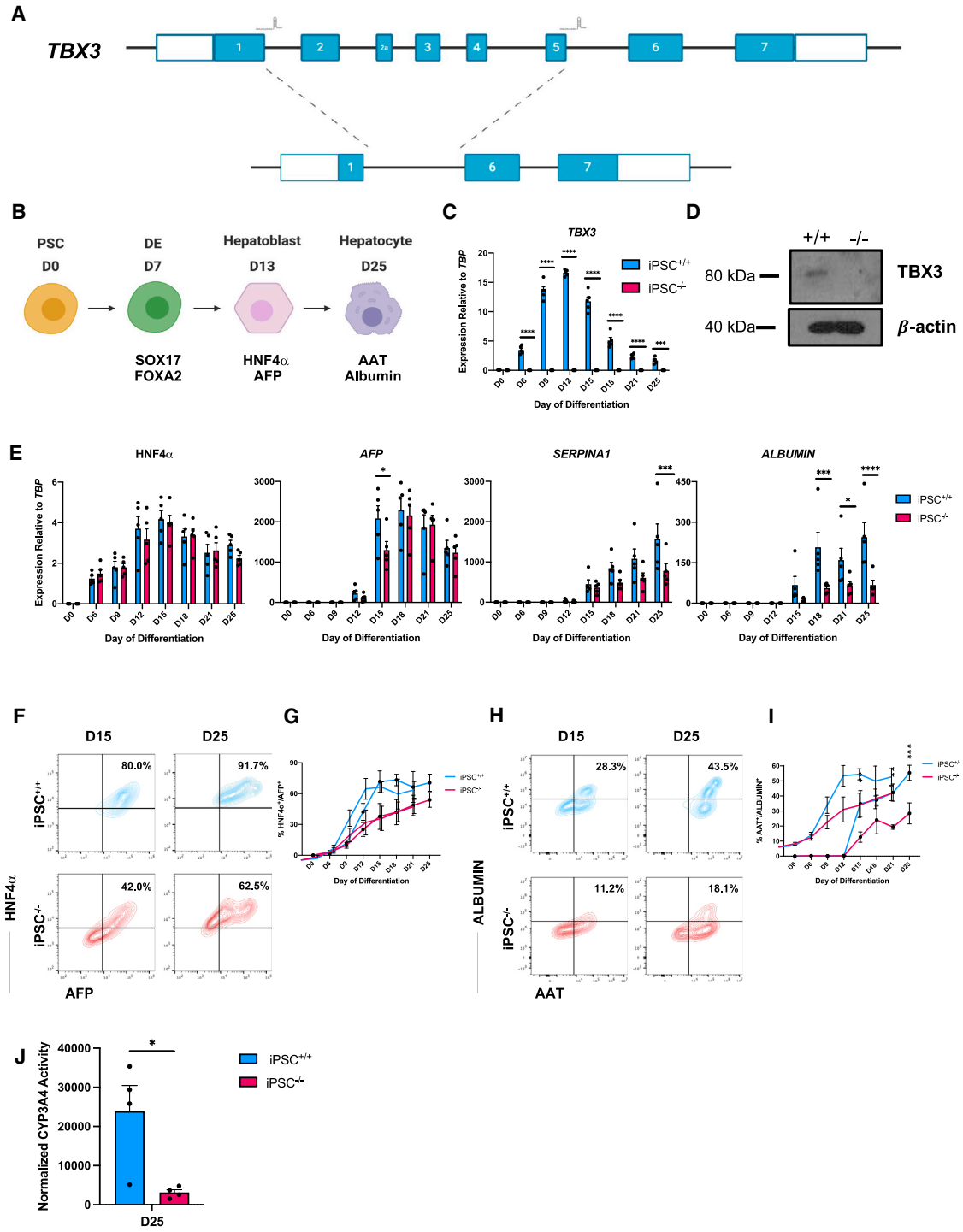


Figure 1. Loss of *TBX3* impairs hepatocyte differentiation

(A) Schematic of endogenous *TBX3* locus with sites of gRNAs indicated above exons. Schematic underneath represents *TBX3* locus after 6.9-kb deletion.

(B) Schematic of directed differentiation protocol of PSCs to hepatocytes.

(C) Time course of *TBX3* expression during hepatocyte differentiation by qRT-PCR (n = 5 separate experiments per time point, per cell line).

(D) Western blot of TBX3 protein in day 12 iPSC$^{+/+}$ and iPSC$^{-/-}$ hepatoblasts.

(legend continued on next page)

RESULTS AND DISCUSSION

Loss of *TBX3* impairs hepatocyte differentiation in human PSCs

To study the role of *TBX3* in human liver development, a *TBX3* mutant line was generated in the CHOPi004-A iPSC background (Mukherjee et al., 2020) using CRISPR/Cas9 genome editing technology. Two guide RNAs were designed to generate a 6.9-kb deletion that includes the transcriptional start site and the entire DNA binding domain, generating a nonfunctional protein (Figure 1A). The deletion in both alleles was verified by sequencing and confirmed to have a normal karyotype (Figures S1A and S1B). The genome-edited line is termed iPSC$^{-/-}$, and the unedited line, designated as iPSC$^{+/+}$, was used as an isogenic control (Table S1). To confirm the phenotype in a second genetic background, we generated *TBX3* mutations in the Mel1 ESC line (Micallef et al., 2012) using the same strategy (Figures S1C and S1D), referred to as ESC$^{-/-}$ (Table S1).

While *TBX3* is expressed in mouse ESCs, it is not expressed in human PSCs (Esmailpour and Huang, 2012), therefore control and mutant lines were differentiated to hepatocytes (Ogawa et al., 2013) to verify loss of *TBX3* (Figure 1B). In iPSC$^{+/+}$ cells, *TBX3* expression was induced at the definitive endoderm stage of differentiation and peaked at the hepatoblast stage (day 12). *TBX3* was not expressed in the iPSC$^{-/-}$ line during differentiation at the mRNA or protein level (Figures 1C and 1D).

Liver marker expression was measured to determine the impact of *TBX3* loss on human liver development. There was no difference in levels of hepatoblast marker *HNF4α* at the mRNA level (Figure 1E), although expression at the protein level was decreased, along with delayed expression of *α-fetoprotein (AFP)* in the iPSC$^{-/-}$ line. *Tbx3*$^{-/-}$ mice develop a normal liver bud but have a small liver due to impaired migration out of the liver bud (Lüdtke et al., 2009), suggesting that Tbx3 is not required for hepatoblast specification, but rather for maturation. This explains the subtle effect observed on hepatoblast marker expression in the differentiation of the iPSC$^{-/-}$ line. A more severe defect was observed later in the differentiation, as expression of hepatocyte markers *SERPINA1*, which encodes the α1-antitrypsin (AAT) protein, and *ALBUMIN* were significantly reduced in iPSC$^{-/-}$ cells compared with iPSC$^{+/+}$ cells (Figure 1E). At the protein level, there were fewer hepatocyte nuclear factor 4 (HNF4)α$^+$/AFP$^+$ cells (Figures 1F and 1G) and AAT$^+$/ALBUMIN$^+$ cells (Figures 1H and 1I) in the iPSC$^{-/-}$ line compared with the iPSC$^{+/+}$ line. iPSC$^{-/-}$ hepatocytes had significantly reduced CYP3A4 activity compared with iPSC$^{+/+}$ hepatocytes, indicating that the TBX3 is also needed for hepatocyte functionality (Figure 1J). Similar defects were seen in the ESC$^{-/-}$ line (Figure S2) which were slightly more severe, including a greater downregulation of *HNF4α*, which may indicate developmental abnormalities are occurring even at the hepatoblast stage. These results indicate that the loss of *TBX3* impairs the ability of PSCs to differentiate to hepatocytes, suggesting that TBX3 is important in both mouse and human liver development.

Pancreas-specific genes are upregulated in *TBX3* mutant hepatoblasts

The developing gut tube displays a well-established expression pattern of transcription factors that specify different organ domains (Zorn and Wells, 2009). To determine if the loss of TBX3 impacts gut tube patterning, the expression of the anterior gut tube marker *SOX2*, posterior gut tube marker *CDX2*, and pancreatic master regulator *PDX1* was examined during hepatocyte differentiation. There was no impact on *SOX2* expression, and a minor upregulation of *CDX2* in the iPSC$^{-/-}$ line. However, *PDX1* was expressed in the iPSC$^{-/-}$ line (Figure 2A), which was surprising, as *PDX1* is not expressed during liver differentiation. Because PDX1 protein levels are tightly regulated by ubiquitination and proteasomal degradation (Claiborn et al., 2010), the proteasome inhibitor MG132 (1 μM) was added for 8 h on day 15 when *PDX1* expression was highest. In the presence of MG132, iPSC$^{-/-}$ cells had higher PDX1 protein levels than without MG132, while the iPSC$^{+/+}$ line displayed no PDX1 protein expression with or without MG132 (Figure 2B). This indicates that the loss of *TBX3* allows for PDX1 protein expression during the hepatocyte differentiation. In addition, expression of *ISL1* (Ahlgren et al., 1997) and Motor Neuron and Pancreas Homeobox 1 (*MNX1*) (Li et al., 1999), genes involved in

(E) Time course of hepatoblast (*HNF4α* and *AFP*) and hepatocyte (*SERPINA1* and *Albumin*) markers during hepatocyte differentiation by qRT-PCR (n = 5 per time point, per cell line).
(F) Representative example of HNF4α and AFP expression at day 15 and day 25 by flow cytometry.
(G) Time course of percentage HNF4α$^+$/AFP$^+$ cells by flow cytometry (n = 5 per time point, per cell line).
(H) Representative example of AAT and ALBUMIN expression at day 15 and day 25 by flow cytometry.
(I) Time course of percentage of AAT$^+$/ALBUMIN$^+$ cells in iPSC$^{+/+}$ and iPSC$^{-/-}$ lines by flow cytometry (n = 5 per time point, per cell line).
(J) Rifampicin-induced CYP3A4 activity in iPSC$^{+/+}$ and iPSC$^{-/-}$ hepatocytes at day 25 (n = 4 per cell line).
For all statistical analysis, *p < 0.05, **p < 0.01, ***p < 0.001, and ****p ≤ 0.0001.

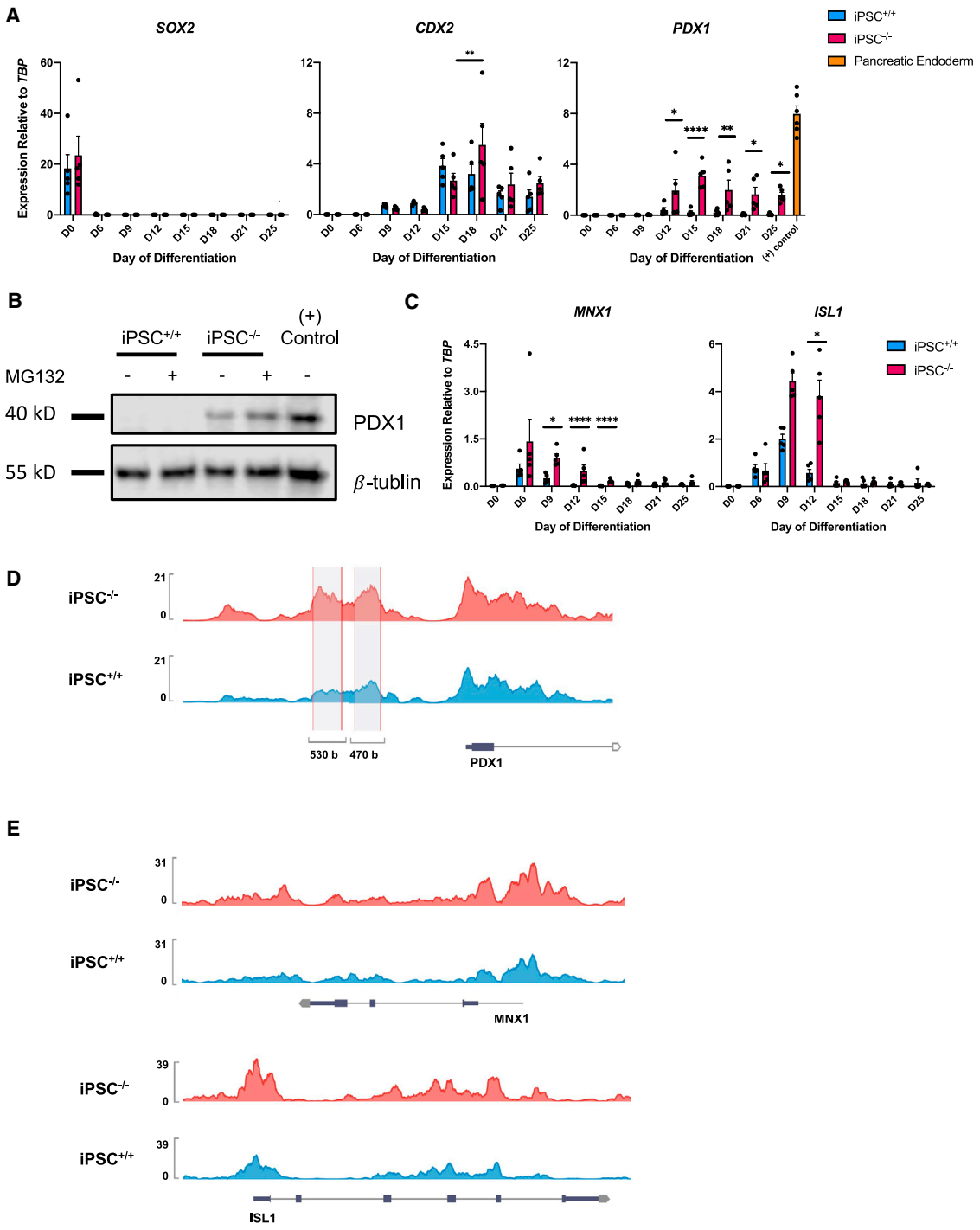


Figure 2. Pancreatic genes are expressed in the iPSC$^{-/-}$ line during hepatocyte differentiation

(A) Time course of *SOX2*, *PDX1*, and *CDX2* during hepatocyte differentiation by qRT-PCR (n = 5 per time point, per cell line). PDX1 expression in iPSC$^{+/+}$ differentiated to pancreatic endoderm for (+) control (n = 6).

(B) Western blot of PDX1 protein in day 15 iPSC$^{+/+}$ and iPSC$^{-/-}$ immature hepatocytes with or without MG132 and pancreatic endoderm for (+) control.

(C) Time course of early pancreatic markers *ISL1* and *MNX1* during hepatocyte differentiation by qRT-PCR (n = 5 per time point, per cell line).

(legend continued on next page)

early pancreas development, was also increased in the iPSC$^{-/-}$ line (Figure 2C), suggesting a loss of repression of pancreatic genes. Similar results were seen the ESC$^{-/-}$ line (Figures S3A–S3C).

PDX1 drives PSCs to a pancreatic fate during differentiation, in part by repression of hepatic genes, including TBX3 (Teo et al., 2015). The upregulation of *PDX1*, *ISL1*, and *MNX1* in the iPSC$^{-/-}$ line suggests that TBX3 may be suppressing pancreatic genes to drive hepatocyte differentiation. To understand how the loss of TBX3 impacts chromatin accessibility, we performed assay for transposase-accessible chromatin sequencing (ATAC-seq) in iPSC$^{+/+}$ and iPSC$^{-/-}$ day 12 hepatoblasts. The *PDX1* locus is generally accessible in both control and mutant hepatoblasts, but the promoter region of *PDX1* is more accessible in iPSC$^{-/-}$ than in iPSC$^{+/+}$ hepatoblasts (Figure 2D). This suggests that TBX3 may inhibit *PDX1* expression via regulation of the *PDX1* promoter, directly or indirectly, to prevent transcription in wild-type hepatoblasts. In addition, the promoter regions of both *MNX1* and *ISL1* are more accessible in iPSC$^{-/-}$ hepatoblasts than in iPSC$^{+/+}$ hepatoblasts (Figure 2E). The loss of *TBX3* results in inappropriate expression of pancreatic genes during hepatocyte differentiation, suggesting that TBX3 may be important to maintain lineage fidelity of the hepatic domain during foregut patterning.

Loss of *TBX3* enhances pancreatic progenitor generation from PSCs

Considering that *TBX3* loss led to de-repression of pancreatic genes during liver differentiation, we tested whether loss of *TBX3* enhanced pancreatic differentiation. The iPSC$^{+/+}$ and iPSC$^{-/-}$ lines were differentiated to pancreatic progenitors (Rezania et al., 2014) (Figure 3A). The iPSC$^{-/-}$ line generated a higher percentage of the PDX1^{+}/NKX6.1^{+} pancreatic progenitor 2 (PP2) population compared with iPSC$^{+/+}$ cells (Figures 3B and 3C). *PDX1* mRNA levels trend higher in the iPSC$^{-/-}$ line at the pancreatic progenitor 1 stage (PP1, day 8) of the pancreas differentiation but do not reach statistical significance, and there was no difference at the PP2 stage (day 11). This is likely because the pancreatic differentiation protocol has factors that drive cells toward a pancreatic identity, and virtually all cells already express PDX1 (Figures 3C and 3D). However, *NKX6.1* expression was higher in iPSC$^{-/-}$ PP2 cells compared with iPSC$^{+/+}$ PP2 cells. *NKX6.1* is crucial for pancreatic progenitor identity and required for further differentiation to β cells (Rezania et al., 2013). In addition, expression of *ISL1* was markedly increased, while *MNX1* trended higher in iPSC$^{-/-}$ cells compared with iPSC$^{+/+}$ cells (Figure 3D). In the ESC$^{-/-}$ line, MNX1 expression was significantly increased (Figure S3F). We also examined hepatic gene expression during the pancreatic differentiation. Levels of hepatic genes were lower in the iPSC$^{-/-}$ line compared with the iPSC$^{+/+}$ line (Figure 3E), suggesting the loss of *TBX3* generates a purer PP2 population. Similar results were seen in the ESC$^{-/-}$ line (Figures S3D–S3H). These data suggest that a lack of TBX3 improves differentiation efficiency and purity, as the iPSC$^{-/-}$ line generated more PP2 cells with increased expression of pancreas genes and decreased expression of hepatic genes, compared with the iPSC$^{+/+}$ line.

iPSC$^{-/-}$ cells are enriched for a pancreatic gene signature

To investigate the effect of the loss of *TBX3* on pancreatic differentiation in greater detail, genome-wide gene expression was examined using RNA sequencing (RNA-seq) in iPSC$^{-/-}$ versus iPSC$^{+/+}$ lines at days 6 and 8 of pancreas differentiation (Figure 4A). We chose day 6 as *TBX3* expression peaked at this point in the differentiation, and was comparable to *TBX3* levels at day 12 during the hepatocyte differentiation (Figure 4B). This stage of differentiation is representative of the gut tube (GT) endoderm when endodermal cells are patterned to foregut. We chose day 8 PP1 cells, as this is when cells begin expressing *PDX1*. At both the GT and the PP1 stages, pancreatic genes were upregulated while hepatic genes were downregulated in iPSC$^{-/-}$ versus iPSC$^{+/+}$ (Figures S4A and 4C). Pathway analysis showed that genes involved in metabolic pathways were downregulated in iPSC$^{-/-}$ GT cells compared with iPSC$^{+/+}$ GT cells (Table S4A), and pathways relating to pancreatic development and function were upregulated in iPSC$^{-/-}$ PP1 (Table S4B). To determine if the gene expression signature of iPSC$^{-/-}$ cells in our *in vitro* differentiation was indicative of pancreas commitment at the expense of liver fate, we used gene set enrichment analysis with a previously published gene set from primary human embryos, comparing the dorsal pancreatic buds with hepatic cords (Jennings et al., 2017). Hepatic cord–specific genes were enriched in iPSC$^{+/+}$ versus iPSC$^{-/-}$ GT cells, and pancreatic bud–specific genes were significantly enriched in the iPSC$^{-/-}$ PP1 population (Figures S4B and 4D). These data further confirm that the loss of *TBX3* helped drive cells toward a pancreatic fate and away from a hepatic fate.

(D) Representative example of tracks showing increased accessibility at the *PDX1* promoter in iPSC$^{-/-}$ day 12 hepatoblasts.
(E) Representative example of tracks showing differential accessibility at the *MNX1* and *ISL1* loci in day 12 hepatoblasts. The y axis represents peak height.
For all statistical analysis, *p < 0.05, **p < 0.01, ***p < 0.001, and ****p $\leq$ 0.0001.

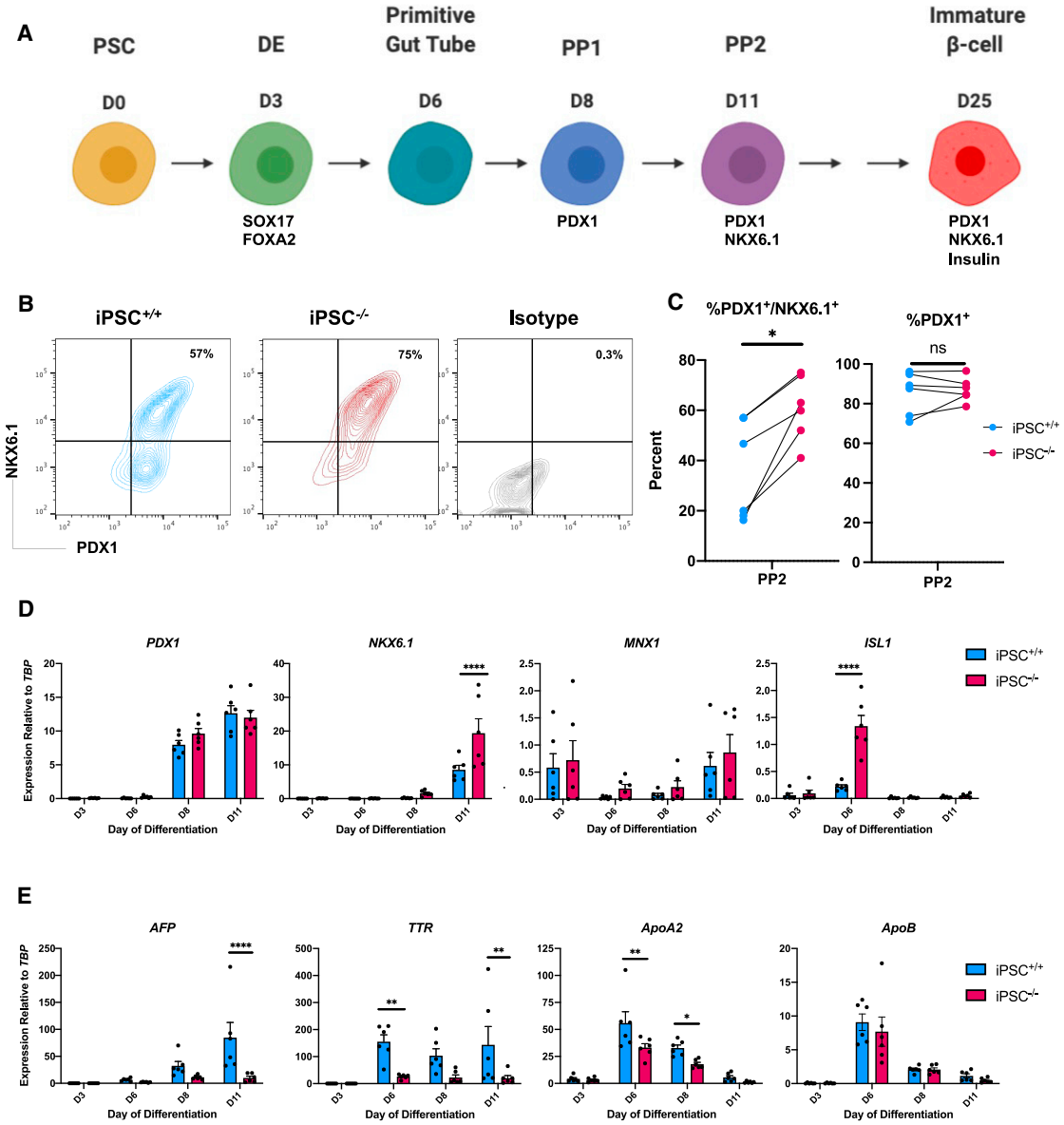


Figure 3. Loss of *TBX3* enhances pancreatic progenitor generation from PSCs

(A) Schematic of directed differentiation protocol of PSCs to immature β cells.

(B) Representative example of PDX1 and NKX6.1 expression in PP2 cells by flow cytometry.

(C) Quantification of the percentage of $PDX1^{+}/NKX6.1^{+}$ and $PDX1^{+}$ PP2 cells (n = 6 per cell line).

(D) Time course of *PDX1*, *NKX6.1*, *ISL1*, and *MNX1* expression during pancreatic differentiation by qRT-PCR (n = 6 per time point, per cell line).

(E) Time course of *AFP*, *TTR*, *ApoA2*, and *ApoB* expression during pancreatic differentiation by qRT-PCR (n = 6 per time, point per cell line).

For all statistical analysis, *p < 0.05, **p < 0.01, ***p < 0.001, and ****p $\leq$ 0.0001.

Additional characterization was performed using RNA-seq analysis of $iPSC^{-/-}$ versus $iPSC^{+/+}$ hepatoblasts (Figure 4A). *PROX1*, which is required for hepatoblast migration and is downregulated in $Tbx3^{-/-}$ mice (Lüdtke et al., 2009), and *CDH2*, encoding mesenchymal marker N-cadherin, were downregulated in $iPSC^{-/-}$ hepatoblasts, suggesting a migratory defect. Consistent with our prior gene expression findings, *PDX1*, *ISL1*, and *MNX1* were upregulated in $iPSC^{-/-}$ hepatoblasts (Figures S4C). To better understand the global changes in gene expression due to the loss of *TBX3*, a covariate analysis was performed to identify commonly dysregulated genes in $iPSC^{-/-}$ cells at

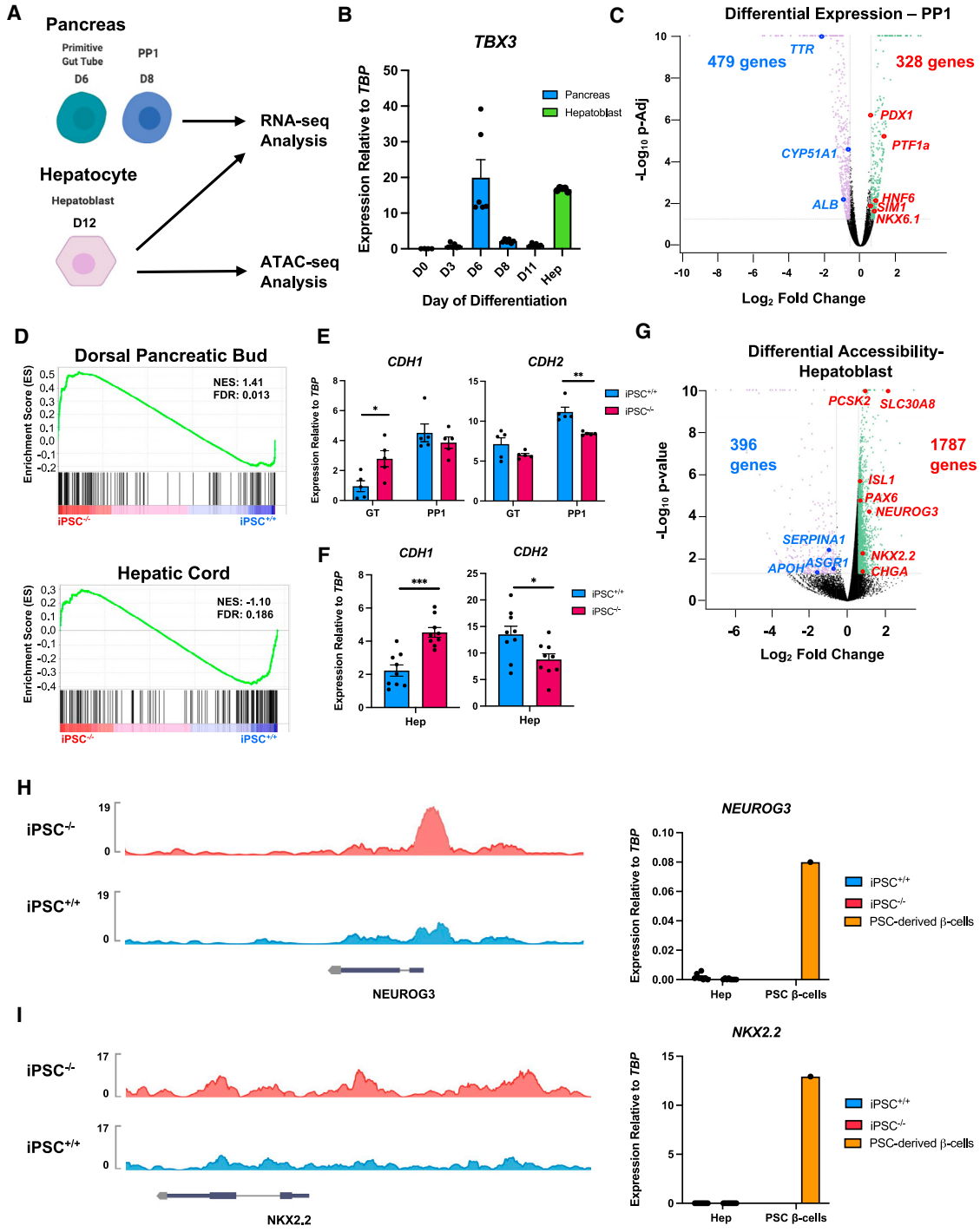


Figure 4. iPSC$^{-/-}$ cells are enriched for a pancreatic gene signature

(A) Schematic of stages from PSC differentiations collected for sequencing analysis.

(B) Time course of *TBX3* expression by qRT-PCR during pancreatic differentiation (n = 6) and hepatoblasts (n = 5) for comparison.

(C) Volcano plot of up- and downregulated genes in iPSC$^{-/-}$ versus iPSC$^{+/+}$ pancreatic progenitor 1 (PP1) cells. p-Adj = 0.05, fold change: $\geq$ 1.5 and $\leq$ −1.5.

(D) Gene set enrichment analysis comparing normalized gene expression of samples in (C) to genes enriched in human fetal dorsal pancreatic bud and hepatic cord.

(E) Expression of *CDH1* and *CDH2* in GT and PP1 cells from pancreatic differentiation by qRT-PCR (n = 5 per time point, per cell line).

(legend continued on next page)

three developmental stages: GT and PP1 cells from the pancreatic differentiation, and hepatoblasts from the hepatocyte differentiation (Figures 4A and S4D). The covariate analysis showed 1,398 genes were similarly up- or downregulated in all three developmental stages (Figure S4E and Table S4C). Several pathways, including epithelial-mesenchymal transition (EMT), were dysregulated upon loss of *TBX3*, regardless of cell type (Table S4D), suggesting a common role of TBX3 in pancreas and liver development. At the GT and hepatoblast stages, *CDH1*, encoding epithelial marker E-cadherin, expression was higher in iPSC$^{-/-}$ cells than iPSC$^{+/+}$ cells, while there was no significant difference at the PP1 stage. *CDH2* expression was lower in iPSC$^{-/-}$ cells than iPSC$^{+/+}$ cells at all three stages (Figures 4E and 4F). E-cadherin protein levels were higher in iPSC$^{-/-}$ cells than in iPSC$^{+/+}$ cells, while the effect on N-cadherin protein was subtle (Figures S4F and S4G). Taken together, these data suggest that EMT may be dysregulated in liver and pancreas cells with the loss of *TBX3*. Although this dysregulation does not appear to have clear implications in impacting a liver versus pancreas fate, it may have roles at other developmental stages.

TBX3 promotes invasiveness of cancer cells through direct regulation of EMT genes *SLUG* (Krstic et al., 2019) and E-cadherin (Rodriguez et al., 2008). During liver bud expansion, hepatoblasts undergo EMT and delaminate into the surrounding mesenchyme. *Tbx3*$^{-/-}$ mice maintain E-cadherin expression, indicating a failure in EMT (Lüdtke et al., 2009). In pancreas development, endocrine progenitors undergo EMT and delaminate from the branching epithelium to form pancreatic islets (Gouzi et al., 2011). Although *Tbx3* expression has been observed in the developing and adult mouse pancreas, its role is unclear. Based on the role of TBX3 in regulating EMT in other biological contexts, it may function similarly in delamination during islet formation. Cytoskeletal structure influences delamination (Kesavan et al., 2014) and differentiation (Mamidi et al., 2018) in the mouse, and enhances PP2 generation from PSCs (Hogrebe et al., 2020). We show that loss of *TBX3* enhances PP2 generation, highlighting possible links among TBX3, cytoskeletal state, and differentiation.

A more detailed analysis of the ATAC-seq dataset from iPSC$^{-/-}$ versus iPSC$^{+/+}$ hepatoblasts was performed and compared with the RNA-seq datasets; 1,787 genes had more accessible chromatin in iPSC$^{-/-}$ hepatoblasts, while 396 genes had less accessible chromatin (Figure 4G), highlighting the role of TBX3 as a transcriptional repressor. Upregulated genes from the RNA-seq covariate analysis largely overlap with genes identified by ATAC-seq as more open (Figures S4H and S4I). Pathway analysis of the ATAC-seq dataset showed that genes involved the same pathways such as EMT identified in the RNA-seq covariate analysis are differentially accessible, demonstrating good correlation between accessibility and gene expression (Table S4E). The top hit in the pathway analysis was upregulation of Hallmark Pancreas Beta Cells, confirming de-repression of the pancreatic program.

Interestingly, some genes involved in later endocrine cell development and maturation, such as *NEUROG3* and *NKX2.2,* are more accessible in iPSC$^{-/-}$ than in iPSC$^{+/+}$ hepatoblasts. Although they are more accessible, *NEUROG3* and *NKX2.2* are not expressed in iPSC$^{-/-}$ hepatoblasts (Figures 4H and 4I). These data suggest that TBX3 not only represses genes involved in early pancreatic development such as *PDX1*, but also is involved in keeping genes expressed later in pancreatic endocrine cell development in a less accessible state. Further studies on how the loss of *TBX3* affects later pancreatic differentiation can enhance our understanding of endocrine cell development and improve our methodology for generating these cells from PSCs.

This study provides insight into the biology of patterning of the hepatic and pancreatic domains during human foregut development. We established that TBX3 is critical for proper hepatocyte development and maturation in humans, similar to the role of Tbx3 in mouse liver development. We also demonstrate that loss of *TBX3* increases the number of pancreatic progenitors derived from PSCs. These results contribute to our knowledge of early endoderm patterning and provide a potential route to improve the generation of *in vitro*–derived pancreatic progenitors, which can be further differentiated into pancreatic cell types.

EXPERIMENTAL PROCEDURES

PSC lines

The CHOPi004-A iPSC line was made by the Human Pluripotent Stem Cell Core at the Children's Hospital of Philadelphia (Mukherjee et al., 2020). The Mel1 ESC line with a GFP reporter in the *Insulin* locus, and reverse tetracycline transactivator construct in the

(F) Expression of *CDH1* and *CDH2* in hepatoblasts from hepatic differentiation by qRT-PCR (n = 9 per time point, per cell line).
(G) Volcano plot of differential accessibility of genes identified by ATAC-seq analysis in iPSC$^{-/-}$ versus iPSC$^{+/+}$ hepatoblasts. p value = 0.05, fold change: $\geq$ 1.5 and $\leq$ −1.5.
(H) Representative example of tracks showing differential accessibility at the *NEUROG3* and *NKX2.2* loci. The y axis represents peak height.
(I) Expression of *NEUROG3* and *NKX2.2* in hepatoblasts (n = 8 per cell line) from hepatic differentiation and PSC-derived β cells (n = 1) by qRT-PCR.
For all statistical analysis, *p < 0.05, **p < 0.01, ***p < 0.001, and ****p $\leq$ 0.0001.

AAVS1 locus was used as a second genetic background (Micallef et al., 2012; Tiyaboonchai et al., 2017).

Genome editing using CRISPR/Cas9

TBX3 mutant PSC lines were generated, as described previously (Maguire et al., 2019), using two guide RNAs (gRNAs) to create a deletion in the endogenous *TBX3* locus. *TBX3* gRNA1, 5′-GGAG TGGATGAGCCTCTCC-3′, targeted the transcription site and *TBX3* gRNA2, 5′-CAAAGGTAAACCATGTCAC-3′, targeted the boundary of exon 5 and intron 4 of the *TBX3* locus. Single colonies were screened for the deletion by PCR and topoisomerase-based cloning, and sequenced for confirmation.

Hepatocyte differentiation

PSCs were differentiated to hepatocytes, as previously described (Ogawa et al., 2013), with modifications. PSCs were split on to 1:3 diluted Matrigel (Corning)-coated 6-well plates, and the differentiation was started when cells reached 80% to 90% confluency. Cells were differentiated until the hepatocyte maturation B stage (day 25). Detailed modifications can be found in the supplemental experimental procedures.

Pancreatic β-cell differentiation

PSCs were differentiated to the PP2 stage, as previously described (Rezania et al., 2014), with modifications. PSCs were split onto 1:30 Matrigel-coated 6-well plates, and differentiated as a monolayer until the end of stage 4 (day 11). Detailed modifications can be found in the supplemental experimental procedures.

Genomic analysis

For RNA-seq analysis, iPSC$^{+/+}$ and iPSC$^{-/-}$ lines were harvested on day 12 of the hepatocyte differentiation protocol, and days 6 and 8 of the pancreatic differentiation protocol. Samples were collected from three biological replicates for each time point and for each genotype. RNA was extracted as described in the supplemental experimental procedures, and sent for sequencing. For ATAC-seq analysis, iPSC$^{+/+}$ and iPSC$^{-/-}$ cells were harvested on day 12 of the hepatocyte differentiation protocol. Samples were collected from two biological replicates for each genotype and sent to Genewiz for tagmentation and sequencing. Data were analyzed by Rosalind (https://rosalind.onramp.bio/), with a HyperScale architecture developed by OnRamp BioInformatics, Inc. (San Diego, CA). A gene list comparing gene expression in dorsal pancreatic bud and hepatic cord tissues dissected from human embryos (Jennings et al., 2017) was used for the gene set enrichment analysis.

Data and code availability

The RNA-seq and ATAC-seq data reported in this study can be found using the GEO accession number GEO: GSE180528.

Statistical analysis

Results from multiple experiments are expressed as the mean ± SEM. An unpaired two-tailed Student's t test for groups with equal variance was performed to determine p values. All statistical analyses were performed on Prism version 8.4.3 for Mac (GraphPad Software). In the figures, *p < 0.05, **p < 0.01, ***p < 0.001, and ****p ≤ 0.0001, and n denotes individual experiments.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stemcr.2021.09.004.

AUTHOR CONTRIBUTIONS

SM designed and performed all experiments described in this study, interpreted data, and prepared the manuscript and figures. DLF and PG provided overall scientific guidance and manuscript preparation, review, and editing.

CONFLICTS OF INTEREST

The authors declare no competing interests.

ACKNOWLEDGMENTS

All illustrations in the figures were created with BioRender.com. This study was supported by NIH grants R01 DK118155 and R01 DK123162.

Received: October 12, 2020
Revised: September 13, 2021
Accepted: September 14, 2021
Published: October 14, 2021

REFERENCES

Ahlgren, U., Jonsson, J., and Edlund, H. (1996). The morphogenesis of the pancreatic mesenchyme is uncoupled from that of the pancreatic epithelium in IPF1/PDX1-deficient mice. Development *122*, 1409–1416.

Ahlgren, U., Pfaff, S.L., Jessell, T.M., Edlund, T., and Edlund, H. (1997). Independent requirement for ISL1 in formation of pancreatic mesenchyme and islet cells. Nature *385*, 257–260.

Bamshad, M., Lin, R.C., Law, D.J., Watkins, W.C., Krakowiak, P.A., Moore, M.E., Franceschini, P., Lala, R., Holmes, L.B., Gebuhr, T.C., et al. (1997). Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. Nat. Genet. *16*, 311–315.

Begum, S., and Papaioannou, V.E. (2011). Dynamic expression of Tbx2 and Tbx3 in developing mouse pancreas. Gene Expr. Patterns *11*, 476–483.

Carlson, H., Ota, S., Campbell, C.E., and Hurlin, P.J. (2001). A dominant repression domain in Tbx3 mediates transcriptional repression and cell immortalization: relevance to mutations in Tbx3 that cause ulnar-mammary syndrome. Hum. Mol. Genet. *10*, 2403–2413.

Claiborn, K.C., Sachdeva, M.M., Cannon, C.E., Groff, D.N., Singer, J.D., and Stoffers, D.A. (2010). Pcif1 modulates Pdx1 protein stability and pancreatic β cell function and survival in mice. J. Clin. Invest. *120*, 3713–3721.

D'Amour, K.A., Bang, A.G., Eliazer, S., Kelly, O.G., Agulnick, A.D., Smart, N.G., Moorman, M.A., Kroon, E., Carpenter, M.K., and Baetge, E.E. (2006). Production of pancreatic hormone-expressing

endocrine cells from human embryonic stem cells. Nat. Biotechnol. *24*, 1392–1401.

Deutsch, G., Jung, J., Zheng, M., Lóra, J., and Zaret, K.S. (2001). A bipotential precursor population for pancreas and liver within the embryonic endoderm. Development *128*, 871–881.

Esmailpour, T., and Huang, T. (2012). TBX3 promotes human embryonic stem cell proliferation and neuroepithelial differentiation in a differentiation stage-dependent manner. Stem Cells *30*, 2152–2163.

Gibson-Brown, J.J., Agulnik, S.I., Silver, L.M., Niswander, L., and Papaioannou, V.E. (1998). Involvement of T-box genes Tbx2-Tbx5 in vertebrate limb specification and development. Development *125*, 2499–2509.

Gouzi, M., Kim, Y.H., Katsumoto, K., Johansson, K., and Grapin-Botton, A. (2011). Neurogenin3 initiates stepwise delamination of differentiating endocrine cells during pancreas development. Dev. Dyn. *240*, 589–604.

Hebrok, M., Kim, S.K., and Melton, D.A. (1998). Notochord repression of endodermal sonic hedgehog permits pancreas development. Genes Dev. *12*, 1705–1713.

Hogrebe, N.J., Augsornworawat, P., Maxwell, K.G., Velazco-Cruz, L., and Millman, J.R. (2020). Targeting the cytoskeleton to direct pancreatic differentiation of human pluripotent stem cells. Nat. Biotechnol. *38*, 460–470.

Jennings, R.E., Berry, A.A., Gerrard, D.T., Wearne, S.J., Strutt, J., Withey, S., Chhatriwala, M., Piper Hanley, K., Vallier, L., Bobola, N., et al. (2017). Laser capture and deep sequencing reveals the transcriptomic programmes regulating the onset of pancreas and liver differentiation in human embryos. Stem Cell Reports *9*, 1387–1394.

Jung, J., Zheng, M., Goldfarb, M., and Zaret, K.S. (1999). Initiation of Mammalian liver development from endoderm by fibroblast Growth factors. Science *284*, 1998–2003.

Kesavan, G., Lieven, O., Mamidi, A., Öhlin, Z.L., Johansson, J.K., Li, W.C., Lommel, S., Greiner, T.U., and Semb, H. (2014). Cdc42/N-WASP signaling links actin dynamics to pancreatic β cell delamination and differentiation. Development *141*, 685–696.

Krstic, M., Kolendowski, B., Cecchini, M.J., Postenka, C.O., Hassan, H.M., Andrews, J., MacMillan, C.D., Williams, K.C., Leong, H.S., Brackstone, M., et al. (2019). TBX3 promotes progression of pre-invasive breast cancer cells by inducing EMT and directly up-regulating SLUG. J. Pathol. *248*, 191–203.

Li, H., Arber, S., Jessell, T.M., and Edlund, H. (1999). Selective agenesis of the dorsal pancreas in mice lacking homeobox gene Hlxb9. Nat. Genet. *23*, 67–70.

Lüdtke, T.H.W., Christoffels, V.M., Petry, M., and Kispert, A. (2009). Tbx3 promotes liver bud expansion during mouse development by suppression of cholangiocyte differentiation. Hepatology *49*, 969–978.

Maguire, J.A., Cardenas-Diaz, F.L., Gadue, P., and French, D.L. (2019). Highly efficient CRISPR-Cas9-mediated genome editing in human pluripotent stem cells. Curr. Protoc. Stem Cell Biol. *48*, 1–14.

Mamidi, A., Prawiro, C., Seymour, P.A., de Lichtenberg, K.H., Jackson, A., Serup, P., and Semb, H. (2018). Mechanosignalling via integrins directs fate decisions of pancreatic progenitors. Nature *564*, 114–118.

Micallef, S.J., Li, X., Schiesser, J.V., Hirst, C.E., Yu, Q.C., Lim, S.M., Nostro, M.C., Elliott, D.A., Sarangi, F., Harrison, L.C., et al. (2012). INSGFP/w human embryonic stem cells facilitate isolation of in vitro derived insulin-producing cells. Diabetologia *55*, 694–706.

Mukherjee, S., Gagne, A.L., Maguire, J.A., Jobaliya, C.D., Mills, J.A., Gadue, P., and French, D.L. (2020). Generation of human control iPSC line CHOPi004-A from juvenile foreskin fibroblast cells. Stem Cell Res. *49*, 102084.

Nostro, M.C., Sarangi, F., Ogawa, S., Holtzinger, A., Corneo, B., Li, X., Micallef, S.J., Park, I.-H., Basford, C., Wheeler, M.B., et al. (2011). Stage-specific signaling through TGF family members and WNT regulates patterning and pancreatic specification of human pluripotent stem cells. Development *138*, 1445.

Ogawa, S., Surapisitchat, J., Virtanen, C., Ogawa, M., Niapour, M., Sugamori, K.S., Wang, S., Tamblyn, L., Guillemette, C., Hoffmann, E., et al. (2013). Three-dimensional culture and cAMP signaling promote the maturation of human pluripotent stem cell-derived hepatocytes. Development *140*, 3285–3296.

Rezania, A., Bruin, J.E., Xu, J., Narayan, K., Fox, J.K., O'Neil, J.J., and Kieffer, T.J. (2013). Enrichment of human embryonic stem cell-derived NKX6.1-Expressing pancreatic progenitor cells accelerates the maturation of insulin-secreting cells in vivo. Stem Cells *31*, 2432–2442.

Rezania, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T., et al. (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. Nat. Biotechnol. *32*, 1121–1133.

Rodriguez, M., Aladowicz, E., Lanfrancone, L., and Goding, C.R. (2008). Tbx3 represses E-cadherin expression and enhances melanoma invasiveness. Cancer Res. *68*, 7872–7881.

Rossi, J.M., Dunn, N.R., Hogan, B.L.M., and Zaret, K.S. (2001). Distinct mesodermal signals, including BMPs from the septum, transversum mesenchyme, are required in combination for hepatogenesis from the endoderm. Genes Dev. *15*, 1998–2009.

Si-Tayeb, K., Noto, F.K., Nagaoka, M., Li, J., Battle, M.A., Duris, C., North, P.E., Dalton, S., and Duncan, S.A. (2010). Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. Hepatology *51*, 297–305.

Singh, R., Hoogaars, W.M., Barnett, P., Grieskamp, T., Sameer Rana, M., Buermans, H., Farin, H.F., Petry, M., Heallen, T., Martin, J.F., et al. (2012). Tbx2 and Tbx3 induce atrioventricular myocardial development and endocardial cushion formation. Cell. Mol. Life Sci. *69*, 1377–1389.

Teo, A.K.K., Tsuneyoshi, N., Hoon, S., Tan, E.K., Stanton, L.W., Wright, C.V.E., and Dunn, N.R. (2015). PDX1 binds and represes hepatic genes to ensure robust pancreatic commitment in differentiating human embryonic stem cells. Stem Cell Reports *4*, 578–590.

Tiyaboonchai, A., Cardenas-Diaz, F.L., Ying, L., Maguire, J.A., Sim, X., Jobaliya, C., Gagne, A.L., Kishore, S., Stanescu, D.E., Hughes, N., et al. (2017). GATA6 plays an important role in the induction of human definitive endoderm, development of the pancreas,

and functionality of pancreatic β cells. Stem Cell Reports *8*, 589–604.

Weidgang, C.E., Russell, R., Tata, P.R., Kühl, S.J., Illing, A., Müller, M., Lin, Q., Brunner, C., Boeckers, T.M., Bauer, K., et al. (2013). TBX3 directs cell-fate decision toward mesendoderm. Stem Cell Reports *1*, 248–265.

Zorn, A.M., and Wells, J.M. (2009). Vertebrate endoderm development and organ formation. Annu. Rev. Cell Dev. Biol. *25*, 221–251.

Stem Cell Reports, Volume *16*

Supplemental Information

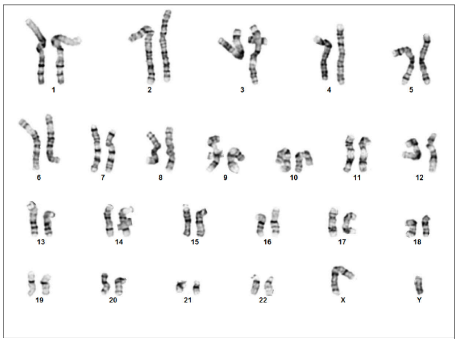
Loss of TBX3 enhances pancreatic progenitor generation from human pluripotent stem cells

Somdutta Mukherjee, Deborah L. French, and Paul Gadue

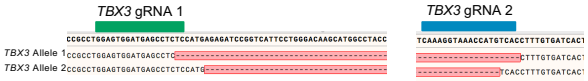
Figure S1: Characterization of *TBX3* knockout PSC lines supporting all figures

A

B



C



D

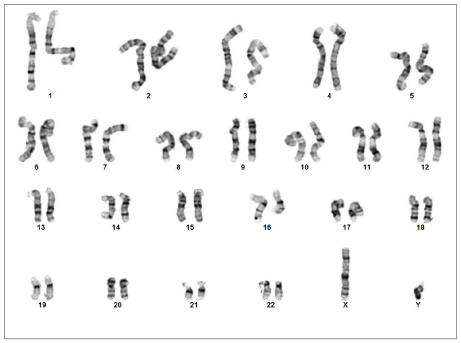


Figure S2: Loss of *TBX3* impairs hepatocyte differentiation supporting Figure 1

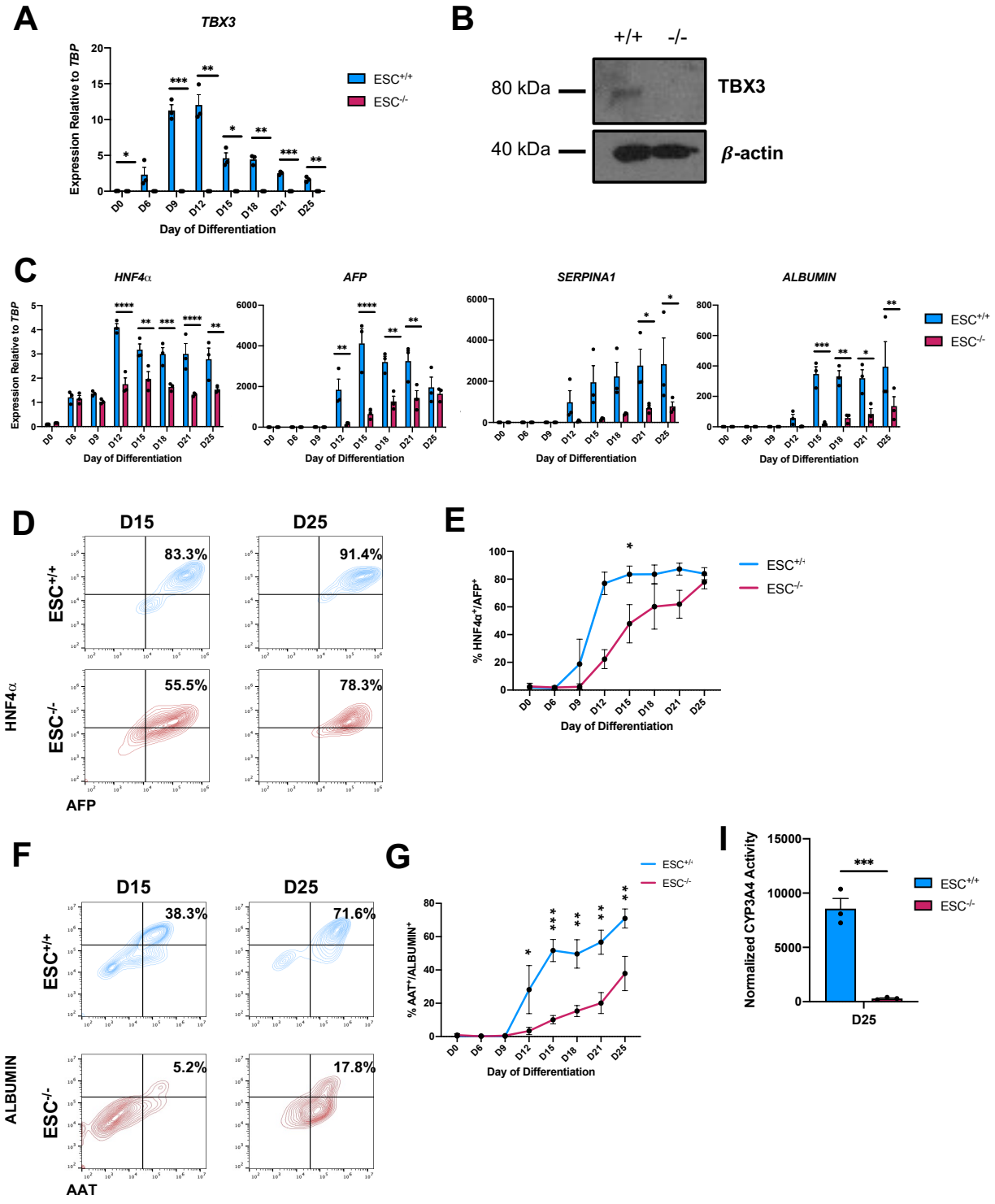


Figure S3: Loss of *TBX3* enhances pancreatic differentiation supporting Figure 2 and Figure 3

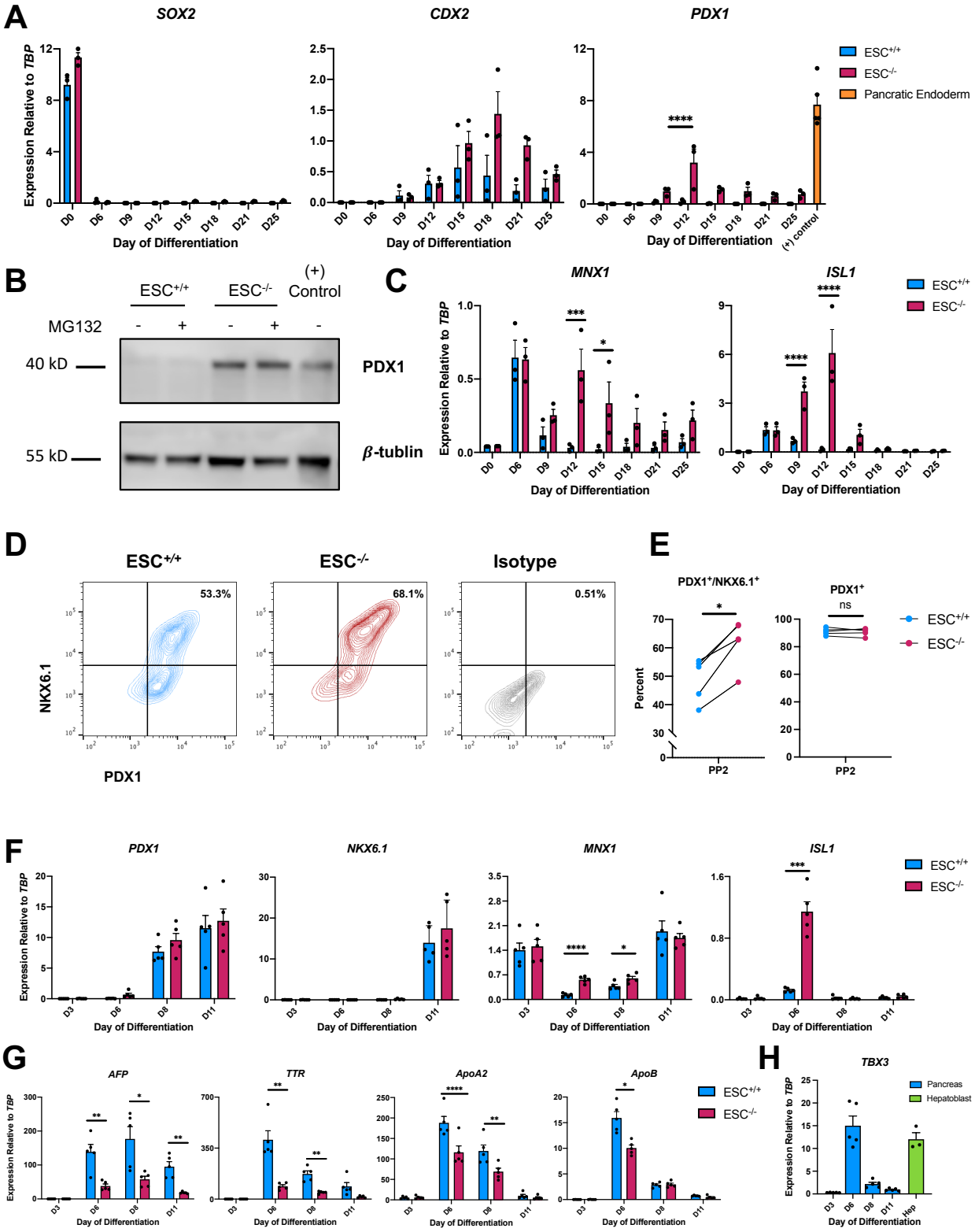


Figure S4: iPSC$^{+/+}$ cells are enriched for hepatic gene signature supporting Figure 4

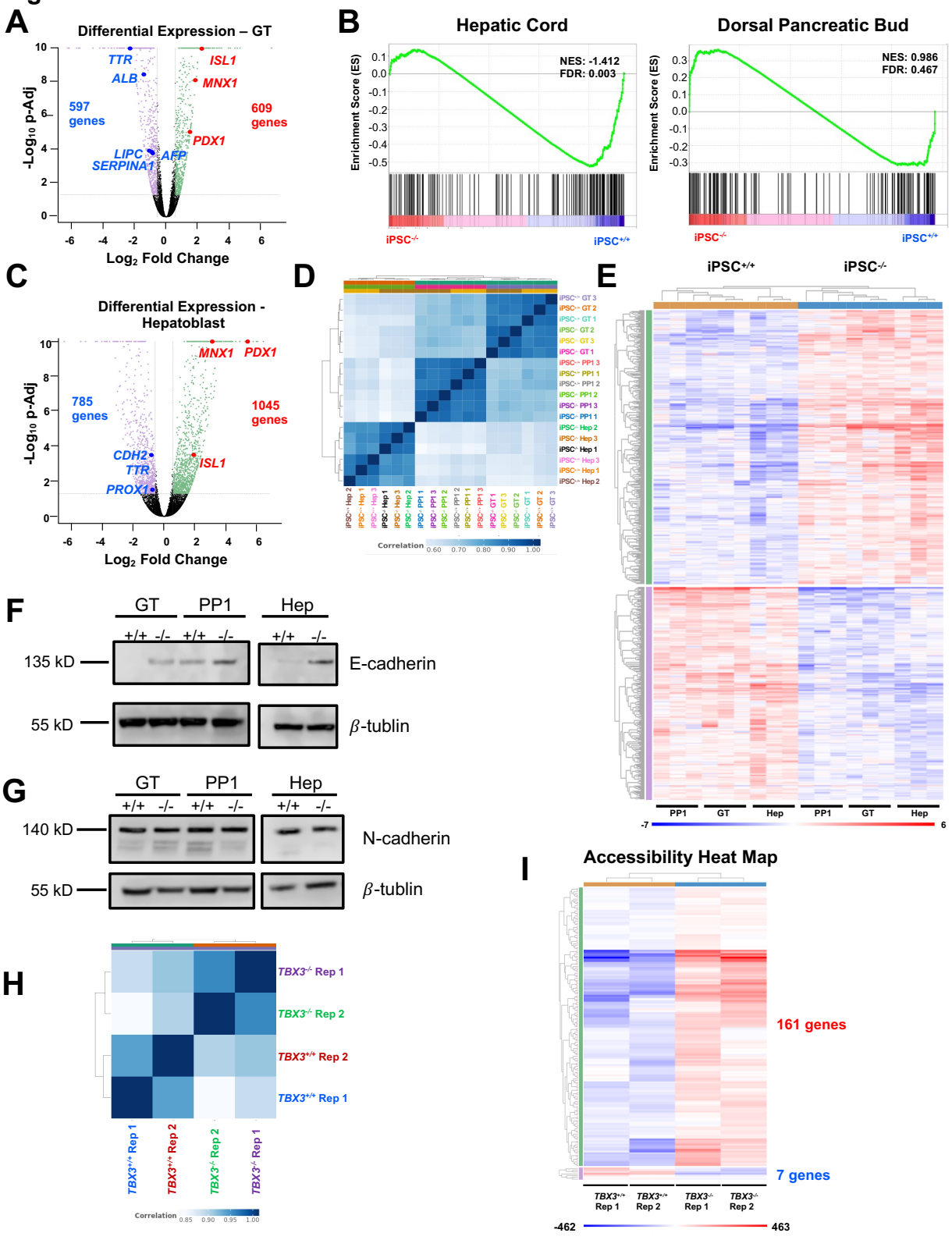


Figure S1. Characterization of *TBX3* knockout PSC lines all figures
(A) Sequencing of the region to verify deletion in both alleles of the *TBX3* locus in the iPSC$^{-/-}$ line. Bars represent locations of *TBX3* gRNA1 and *TBX3* gRNA2.
(B) Karyotype of iPSC$^{-/-}$ line.
(C) Sequencing of the region of the gRNA sequences to verify deletion in both alleles of endogenous *TBX3* locus in the ESC$^{-/-}$ line. Bars represent locations of *TBX3* gRNA1 and *TBX3* gRNA2.
(D) Karyotype of ESC$^{-/-}$ line.

Figure S2. Loss of *TBX3* impairs hepatocyte differentiation supporting Figure 1
(A) Time-course of *TBX3* expression during hepatocyte differentiation by qRT-PCR. (n =3 per time point, per cell line).
(B) Western blot of TBX3 protein in day 12 ESC$^{+/+}$ and ESC$^{-/-}$ hepatoblasts.
(C) Time-course of hepatoblast (*HNF4*α and *AFP*) and hepatocyte (*SERPINA1* and *ALBUMIN*) markers during hepatocyte differentiation by qRT-PCR (n = 3 separate experiments per time, point per cell line).
(D) Representative example of HNF4α and AFP expression at day 15 and Day 25 by flow cytometry.
(E) Time-course of percentage of HNF4α^+/AFP$^+$ cells by flow cytometry (n=3 per time point, per cell line).
(F) Representative example of AAT ad ALBUMIN expression at day 15 and Day 25 by flow cytometry.
(G) Time-course of percentage of AAT$^+$/ALBUMIN$^+$ cells by flow cytometry (n=3 per time point, per cell line).
(J) Rifampicin-induced CYP3A4 activity in ESC$^{+/+}$ and ESC$^{-/-}$ hepatocytes at day 25 (n = 3 per cell line). For all statistical analysis, *p<0.05, **p<0.01, ***p<0.001, and **** p≤ 0.0001.

Figure S3. Loss of *TBX3* enhances pancreatic differentiation supporting Figure 2 and Figure 3
(A) Time-course of anterior gut tube (*SOX2*), pancreatic endoderm (*PDX1*), and posterior gut tube (*CDX2*) markers during hepatocyte differentiation by qRT-PCR (n = 3 per time point, per cell line). PDX1 expression in ESC$^{+/+}$ differentiated to pancreatic endoderm for (+) control (n = 5).
(B) Western blot of PDX1 protein in day 15 ESC$^{+/+}$ and ESC$^{-/-}$ immature hepatocytes with or without MG132 and pancreatic endoderm for (+) control.
(C) Time-course of early pancreatic markers *ISL1* and *MNX1* during hepatocyte differentiation by qRT-PCR (n = 3 per time point, per cell line).
(D) Representative example of PDX1 and NKX6.1 expression in PP2 cells from pancreas differentiation.
(E) Quantification of the percentage of PDX1$^+$/NKX6.1$^+$ and PDX1$^+$ PP2 cells (n = 5 per cell line).
(F) Time-course of *PDX1*, *NKX6.1*, *ISL1*, and *MNX1* during pancreatic differentiation by qRT-PCR (n = 5 per time point, per cell line).
(G) Time-course of hepatoblast markers *AFP, TTR*, *ApoA2*, and *ApoB* during pancreatic differentiation by qRT-PCR (n = 5 per time point, per cell line).
(H) Time-course of *TBX3* expression during pancreatic differentiation (n = 5) by qRT-PCR and day 12 hepatoblasts (n = 3) for comparison. For all statistical analysis, *p<0.05, **p<0.01, ***p<0.001, and **** p≤ 0.0001.

Figure S4. iPSC$^{+/+}$ cells are enriched for hepatic gene signature supporting Figure 4
(A) Volcano plot of up- and downregulated genes in iPSC$^{-/-}$ versus iPSC$^{+/+}$ GT cells. p-Adj = 0.05, fold change: ≥ 1.5 and ≤ -1.5.
(B) GSEA analysis comparing normalized gene expression of samples examined in (A) to genes enriched in human fetal hepatic cords and dorsal pancreatic bud.
(C) Volcano plot of up- and downregulated genes in iPSC$^{-/-}$ versus iPSC$^{+/+}$ hepatoblast cells. p-Adj = 0.05, fold change: ≥ 1.5 and ≤ -1.5.
(D) Sample correlation of heatmap correlating RNA-seq samples of same replicate group, developmental stage, tissue type, and genotype.
(E) Heat map of genes identified by covariate analysis that are commonly up- and downregulated in iPSC$^{-/-}$ versus iPSC$^{+/+}$ cells at different developmental stages and tissues types.
(F) Western blot of E-cadherin protein in iPSC$^{+/+}$ and iPSC$^{-/-}$ cells.
(G) Western blot of N-cadherin protein in iPSC$^{+/+}$ and iPSC$^{-/-}$ cells.
(H) Sample correlation of heatmap correlating ATAC-seq samples of same replicate group and genotype.

(I) Upregulated genes from (E) were analyzed for differential accessibility by ATAC-seq. Heat map showing differentially accessible genes only (161 genes were more accessible while only 7 were less accessible).
GT = day 6 gut tube, PP1 = day 8 pancreatic progenitor 1, and Hep = day 12 hepatoblast

Table S1: *TBX3* mutant cells lines supporting all figures

Name	Genetic Background	Allele 1	Allele 2
$iPSC^{+/+}$	CHOPi004-A	Wild type	Wild type
$iPSC^{-/-}$	CHOPi004-A	6.9 Kb deletion	6.9 Kb deletion
$ESC^{+/+}$	Mel1	Wild type	Wild type
$ESC^{-/-}$	Mel1	6.9 Kb deletion	6.9 Kb deletion

Table S2: Primary and secondary antibodies supporting Figures 1-3

Antibody	Species	Dilution and Application	Source and Catalog Number
TBX3	Goat Polyclonal IgG	1:200 for Western Blot	Santa Cruz #sc-17871
β-actin	Mouse	1:2000 for Western Blot	Sigma #A1978
HNF4a	Rabbit Monoclonal IgG	1:100 for flow cytometry	Cell signaling #3113
AFP	Mouse Monoclonal IgG1	1:400 for flow cytometry	R&D Systems #MAB1368
AAT	Mouse Monoclonal IgG1	1:100 for flow cytometry	Santa Cruz #sc-59438
Albumin	Rabbit Polyclonal IgG	1:1000 for flow cytometry	DakoCytomation #A0001
PDX1-biotinylated	Goat Polyclonal IgG	1:50 for flow cytometry 1:500 for Western blot	R&D Systems #BAF2419
β-tubulin	Rabbit	1:1000 for Western blot	
NKX6.1	Mouse Monoclonal IgG1	1:250 for flow cytometry	DSHB #F55A10
E-Cadherin	Mouse	1:1000 for Western blot	
N-Cadherin	Rabbit	1:1000 for Western blot	
Mouse IgG1 - Alexa488	Goat	1:400 for flow cytometry	Jackson Immunoresearch # 115-545-205
Rabbit IgG - Alexa 647	Goat	1:400 for flow cytometry	Jackson Immunoresearch #111-605-144
Mouse IgG1 - 647	Donkey	1:400 for flow cytometry	Jackson Immunoresearch # 715-606-151
Streptavidin Pacific Blue		1:400 for flow cytometry	Invitrogen #S-11222
Goat-HRP	Rabbit	1:1000 for Western Blot	R&D Systems #HAF017
Mouse-HRP	Goat	1:5000 for Western Blot	Biorad #170-6516
Rabbit-HRP	Goat	1:5000 for Western Blot	Biorad #170-6515

Table of primary and secondary antibodies used for flow cytometry and western blot.

Table S3: Table of forward and reverse primers used for qRT-PCR supporting Figures 1-4.

Gene	Forward Primer	Reverse Primer
TBX3	5'-TGAGATGTTCTGGGCTGG-3'	5'-CTTACCAGCCACCATCCA-3'
AFP	5'-GTTTGTTCAAGAAGCCACTTAC-3'	5'-CACCCTGAAGACTGTTCATC-3'
HNF4α	5-TCCAACCCAACCTCATCCTCCTTCTT-3'	5'TCCTCTCCACTCCAAGTTCCTGTT-3'
SERPINA1	5' - AGGGCCTGAAGCTAGTGGATAGT-3'	5'-TCTGTTTCTTGGCCTCTTCGGTGT-3'
ALBUMIN	5'-GTGAAACACAAGCCCAAGGCAACA-3'	5'-TCCTCGGCAAAGCAGGTCTC-3'
SOX2	5'-CCATCACCCACAGCAAAT-3'	5'-AGTCCAGGATCTCTCTCATAA-3'
PDX1	5'-GGAGCTGGCTGTCATGTTG-3'	5'-CACTTCATGCGGCGGTTT-3'
CDX2	5'-AAGGACGTGAGCATGTACCCTAGC-3'	5'-CACGTGGTAACCGCCGTAGTC-3'
ISL1	5--CAGAAGGAGGACCGGGCTCTAAT-3'	5'-GACTGGCTACCATGCTGTTAGGTGTAT-3'
MNX1	5'-AGAAGGCGGAAACCCACAGTGTAA-3'	5'-CCCAGAGACGTAAGCATAAACCCT-3'
NKX6.1	5'-AAGAAGCACGCTGCCGAGATG-3'	5'-CCGAGTTGGGATCCAGAGGCTTATT-3'
TTR	5'-ATGGGCTCACAACTGAGGAGGAAT-3'	5'-AGATGCCAAGTGCCTTCCAGTAAGA-3'
APOA2	5'-ATGTGTGGAGAGCCTGGTTTCTCA-3'	5'-AAGCTCTGGGCTCTTGACCTTCT-3'
APOB	5'-ACTCACATCCTCCAGTGGCTGAAA-3'	5'-CGCTGATCCCTCGCCATGTT-3'
CDH1	5'-TTCCCTCGACACCCGATTCAAAGT -3'	5'-GAGTCCCAGGCGTAGACCAAGAAAT-3'
CDH2	5'-TGAGAGCAGTGAGCCTGCAGATTT-3'	5'-TGAGAGCAGTGAGCCTGCAGATTT-3'
NEUROG3	5'GGAGTCGGCGAAAGAAG-3'	5'-ATGTAGTTGTGGGCGAAG-3'
NKX2.2	5'-GAATGTTTGCGCAGCTTCGCTTCT-3'	5'-CCACTTGCTTTAGAAGACGGCTGA-3'
TBP	5'- TTGCTGAGAAGAGTGTGCTGGAGATG-3	5'-CGTAAGGTGGCAGGCTGTTGTT-3'

Supplemental Experimental Procedures

PSC lines

PSCs were cultured in an environment of 5% CO_2, 5% O_2, and 90% N_2 on 0.1% gelatin (Sigma) and irradiated mouse embryonic fibroblasts. They were grown in human embryonic stem cell (hES) medium consisting of DMEM/F12 (Corning) with 15% Knockout Serum Replacement (Gibco), 2mM L-glutamine (Corning), 1x Non-Essential Amino Acids (Gibco), 1x Penicillin/Streptomycin (Corning), 0.1 mM 2 β-mercaptoethanol (Gibco), and 10 ng/mL bFGF (R&D Systems). Medium was changed daily. PSCs were grown to 80% confluency and split at a 1:12 ratio using TryplE (Gibco) dissociation regent. Cells were replated in hES medium with 5μM ROCK inhibitor Y-27632 dihydrochloride (Tocris).

Hepatocyte Differentiation

The hepatocyte differentiation previously described (Ogawa et al., 2013) was modified as follows. PSCs were split onto 1:3 Matrigel (Corning) coated 6-well plates and cultured until they reached 80-90% confluency before starting Day 0 of the differentiation. Cells were differentiated as a monolayer in 5% O_2, 5% CO_2, and 90% N_2. Cells were cultured for 24 hours (Day 0) in RPMI medium (Corning) with 2mM L-glutamine, 50 μg/ml stabilized ascorbic acid (SAA) (Wako), 4.5×10^{-4} M MTG (Sigma), 100ng/mL Activin A (R&D Systems), and 2μM CHIR (Tocris). Cells were then cultured for 24 hours in RPMI medium with 2mM L-glutamine, 50 μg/ml SAA, 4.5×10^{-4} M MTG, 100ng/mL Activin A, and 5 ng/mL bFGF (R&D Systems). From Days 2-6, cells were cultured serum free differentiation (SFD) medium, 2mM L-glutamine, 50 μg/ml SAA, 4.5×10^{-4} M MTG, 100ng/mL Activin A, and 5 ng/mL bFGF. SFD medium consists of Iscove's DMEM (Corning) supplemented with 25% Ham's/F12 (Corning), 0.5% N2 supplement (Gibco), 1% B27 without retinoic acid (RA) supplement (Gibco), and 1% bovine serum albumin (BSA) (Sigma). At Day 6, purity of the definitive endoderm (DE) was assessed by flow cytometry with DE being 95-98% $CXCR4^+/C\text{-}kit^+$. Cells were then dissociated using TryplE for 2 minutes and replated with Y-27632 dihydrochloride onto 1:3 Matrigel coated 6-well plates at a density of 3.5×10^5 cells per well. On Day 7, the medium was replaced with hepatoblast specification consisting of H16 medium supplemented with supplemented 2mM L-glutamine, 50 μg/ml SAA, 4.5×10^{-4} M MTG, 40 ng/mL bFGF, and 50 ng/mL BMP4 (R&D Systems). H16 base medium consists of DMEM Low glucose (Gibco) supplemented with 25% Ham's/F12, 1% B27 with RA (Gibco), and 1% BSA. Medium was changed every other day. On Day 13 medium was replaced with hepatocyte maturation medium A consisting of H16 medium supplemented with 2mM L-glutamine, 50 μg/ml SAA, 4.5×10^{-4} M MTG, 20 ng/mL Hepatocyte Growth Factor (HGF) (R&D Systems), 20 ng/mL Oncostatin-M (OSM) (R&D Systems), and 40 ng/ml dexamethasone (Dex) (Sigma). Medium was changed every other day. On Day 21, cells were transferred to an environment of 20% O_2 and 5% CO_2, and 90% N_2. Medium was replaced with hepatoblast maturation medium B consisting of H21 medium supplemented with 2mM L-glutamine, 50 μg/ml SAA, 4.5×10^{-4} M MTG, 20 ng/mL Hepatocyte Growth Factor (HGF) (R&D Systems), 20 ng/mL Oncostatin-M (OSM) (R&D Systems), and 40 ng/ml dexamethasone (Dex) (Sigma). H21 medium is the same as H16 except using DMEM high glucose (Gibco) as the base medium. Medium was changed every other day until Day 25. Cells were harvested on Day 0, Day 6, Day 9, Day 12, Day 15, Day 18, Day 21 and Day 25 for analysis.

Cytochrome P450 activity assay

The cytochrome p450 activity assay was performed using P450-GloTM CYP3A4 Luciferin-IPA kit (Promega, V9002). On Day 22 of the hepatocyte differentiation, cells were cultured with 25 μM Rifampicin (Sigma) or with DMSO (Tocris) as a vehicle control. To confirm that the induced activity was specific for CYP3A4 enzyme, inhibition controls included the selective inhibitor ketoconazole at 1 μM (Sigma) in the presence of 25 μM Rifampicin. Net signal was calculated by subtracting background luminescence values (no-cell control) from Rifampicin or DMSO values. Luminescence values were normalized for $5x10^5$ cells to account for differences in cell numbers between wells.

Pancreatic β-cell Differentiation

The pancreatic β-cell differentiation previously described (Rezania et al., 2014) was modified as follows. PSCs were split onto 1:30 Matrigel coated 6-well plates and cultured until they reached 80-90% confluency before starting Day 0 of the differentiation. Cells were grown in 20% O_2 and 5% CO2, and 90% N_2, and differentiated as a monolayer. Cells were cultured for 24 hours in RPMI medium with 100ng/mL Activin A, and 3μM CHIR (Day 0). Cells were cultured for another 24 hours (Day 1) in RPMI medium with 100 ng/mL

Activin A, 0.3 μM CHIR, and 5 ng/mL bFGF. On Day 2, cells were differentiated to definitive endoderm (DE) with serum free differentiation (SFD) medium containing 100 ng/mL Activin A. At Day 3, purity of the DE was assessed as described above. From Day 3-5, cells were cultured in DMEM/F12 medium with 1% fetal bovine serum (FBS) (Gibco), 50μg/ml stabilized ascorbic acid (SAA), 1.25 mM IWP2 (Tocris), and 50 ng/ml FGF7 (R&D Systems). Medium was changed every day. On Day 6 and 7, cells were cultured in DMEM high glucose medium with 1% B27 without RA, 1X Glutamax (Gibco), 50 μg/ml SAA, 0.5% ITS-X (Gibco), 50 ng/mL FGF7, 0.5μM SANT1 (Sigma), 1μM RA (Sigma), 100nM LDN (Tocris), and 500nM Phorbol 12-myristate 13-acetate (Tocris). Medium was changed daily. From Day 8-10, cells were cultured in DMEM high glucose medium with 1% B27 without RA, 1X Glutamax, 50 μg/ml SAA, 0.5% ITS-X, 2 ng/mL FGF7, 0.5μM SANT1, 0.1μM RA, 100nM LDN, and 250nM Phorbol 12-myristate 13-acetate. Medium was changed daily. Cells were harvested on Day 0, Day 3, Day 6, Day 8, and Day 11 for analysis.

RNA-seq Genomic Analysis

Cells were harvested on day 12 of the hepatocyte differentiation by dissociation with 0.25% Trypsin/EDTA (Gibco) for 4 minutes. RNA was extracted as described below and sent to Genewiz (South Plainfield, NJ) for library preparation and sequencing. Cells were harvested on day 6 and day 8 of the pancreas differentiation as described above. RNA was extracted and sent to the Center for Applied Genomics Biorepository Core at the Children's Hospital of Philadelphia for library preparation and sequencing. Data was analyzed by ROSALIND (https://rosalind.onramp.bio/), with a HyperScale architecture developed by ROSALIND, Inc. (San Diego, CA). Reads were trimmed using cutadapt (Martin, 2011). Quality scores were assessed using FastQC[2]. Reads were aligned to the Homo sapiens genome build hg19 using STAR (Dobin et al., 2013). Individual sample reads were quantified using HTseq (Anders et al., 2015) and normalized via Relative Log Expression (RLE) using DESeq2 R library (Love et al., 2014). Read Distribution percentages, violin plots, identity heatmaps, and sample MDS plots were generated as part of the QC step using RSeQC (Wang et al., 2012). Deseq2 was also used to calculate fold changes and p-values and perform optional covariate correction. Clustering of genes for the final heatmap of differentially expressed genes was done using the PAM (Partitioning Around Medoids) method using the fpc R library. Hypergeometric distribution was used to analyze the enrichment of pathways, gene ontology, domain structure, and other ontologies. The topGO R library was used to determine local similarities and dependencies between GO terms in order to perform Elim pruning corrections. Several database sources were referenced for enrichment analysis, including Interpro (Mitchell et al., 2019), NCBI (Geer et al., 2009), KEGG (Kanehisa and Goto, 2000; Kanehisa et al., 2017, 2019), MSigDB (Liberzon et al., 2011; Subramanian et al., 2005), REACTOME (Fabregat et al., 2018), WikiPathways (Slenter et al., 2018). Enrichment was calculated relative to a set of background genes relevant for the experiment. Functional enrichment analysis of pathways, gene ontology, domain structure and other ontologies was performed using HOMER (Heinz et al., 2010). Gene set enrichment analysis was performed (Mootha et al., 2003; Subramanian et al., 2005) using a gene list comparing gene expression in dorsal pancreatic bud and hepatic cord tissues dissected from human embryos using laser capture technology (Jennings et al., 2017). The gene list was sorted by a p-value of <0.05, then by fold change. The top 200 upregulated genes were the most enriched in the dorsal pancreatic bud, and the top 200 downregulated genes were most enriched in the hepatic cord. GSEA was performed on these subsets using the "gene_set" permutation and an FDR cutoff of 5%.

ATAC-seq Genomic Analysis

Cells were harvested on day 12 of the hepatocyte differentiation by dissociation with 0.25% Trypsin/EDTA (Gibco) for 4 minutes. $1x10^6$ cells were frozen in 90% FBS (Gibco) and 10% DMSO (Tocris), and sent to Genewiz (South Plainfield, NJ) for tagmentation, library preparation, and sequencing. Data was analyzed by ROSALIND (https://rosalind.onramp.bio/), with a HyperScale architecture developed by ROSALIND, Inc. (San Diego, CA). Reads were trimmed using cutadapt (Martin, 2011). Quality scores were assessed using FastQC. Reads were aligned to the Homo sapiens genome build hg19 using bowtie2 (Langmead and Salzberg, 2012). Per-sample quality assessment plots were generated with HOMER (Heinz et al., 2010) and Mosaics (Kuan et al., 2011). Peaks were called using MACS2 (Zhang et al., 2008). Peak overlaps and differential accessibility were calculated using the DiffBind R library (Ross-Innes et al., 2012). Differential accessibility was calculated at gene promoter sites. Read distribution percentages, identity heatmaps, and FRiP plots were generated as part of the QC step using ChIPQC R library (Carroll and Stark, 2014) and

HOMER. HOMER was also used to generate known and de novo motifs and perform functional enrichment analysis of pathways, gene ontology, domain structure and other ontologies.

RNA extraction, cDNA synthesis, and qRT-PCR

Cells were harvested by dissociation with 0.25% Trypsin/EDTA for 4 minutes. Cellular RNA was isolated using the PureLink RNA Micro Scale Kit (Invitrogen) following the manufacturer's protocol. Random hexamers (Invitrogen) were used with the SuperScript III Reverse Transcriptase System (Invitrogen) to synthesize cDNA from 500 ng of extracted RNA. qRT-PCR reactions were done in triplicate on a Roche LightCycler 480 II using SYBR Select Master Mix (Applied Biosystems). Serial dilutions of H9 embryonic stem cell genomic DNA were used to generate a standard curve, and *TBP* (Veazey and Golding, 2011) was used as a house keeping gene to determine relative gene expression levels. The primers that were used in this study can be found in Table S3.

Western Blot

Cells were harvested with 1.5x Laemmli buffer (75 mM Tris-HCl3, 15% glycerol, 3% SDS, 3.75 mM EDTA, and 200 mM NaF)**.** Cell lysates were boiled at 95^{o}C for 20 minutes and protein was quantified using Pierce BCA Protein Assay Kit (Thermo Fisher Scientific). Samples were aliquoted and diluted with 4x Laemmli buffer. 20 µg of protein from each sample were run on a 4%-12% Bis-Tris SDS-polyacrylamide gel (Invitrogen) and transferred to a 0.45 µm pore size nitrocellulose membrane (Thermo Fisher Scientific). The membrane was stained with Ponceau S (Sigma-Aldrich) to ensure successful transfer. The membrane was washed with 1X TBS (Bio-Rad) with 0.1% Tween-20 (Sigma) (TBST), then blocked in 5% nonfat dry milk in TBST for 1 hour at room temperature. The membrane was probed with primary antibody diluted in 5% nonfat dry milk in TBST overnight at 4°C. The membrane was washed with TBST and placed in a horseradish peroxidase conjugated secondary antibody diluted in 5% nonfat dry milk in 1X TBST for 1 hour at room temperature. Pierce ECL Western Blotting Substrate (Thermo Fisher Scientific) was added to the membrane and exposed onto HyBlot CL autoradiography film (Denville Scientific) for visualization the TBX3 blots. Super Signal West Femto Maximum Sensitivity Substrate (Thermo Fisher Scientific) was added to the membrane and visualized using a C-DiGit Blot Scanner and ImageStudioDigits Software Version 5.2.5 (LiCor) for the PDX1, E-Cadherin, and N-Cadherin blots. A list of antibodies used can be found in Table S2.

Flow Cytometry

Cells were harvested by dissociation with 0.25% Trypsin/EDTA for 3 to 5 minutes. For intracellular staining, cells were fixed with 1.6% paraformaldehyde (Electron Microscopy Science) for 30 minutes at 37°C. Cells were washed in 1X PBS (Corning), then permeabilized and stained in 1X saponin buffer (Biolegend). Primary and secondary antibodies were diluted to the appropriate concentrations in saponin buffer and cells were stained for thirty minutes each at room temperature. Following staining, cells were washed in saponin, and resuspended in FACS buffer (1X PBS (Corning) with 0.1% BSA (Sigma) and 0.1% sodium azide (Sigma)). For extracellular staining, conjugated primary antibodies were diluted to the appropriate concentration in FACS buffer and cells were stained for fifteen minutes at room temperate. Following the staining, cells were washed and resuspended in FACS buffer. All samples were analyzed on a CytoFLEX V2-Br-R2 flow cytometer (Beckman Coulter Life Sciences) and FlowJo Version 10.6.2 (Beckton Dickenson) software program. Cell sorting was carried out on a single cell suspension using BD FACSAria II (Becton Dickinson). A list of antibodies used can be found in Table S2.

Supplemental References

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Carroll, T.S., and Stark, R. (2014). Assessing ChIP-seq sample quality with ChIPQC. 1–21.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B.,

Korninger, F., May, B., et al. (2018). The Reactome Pathway Knowledgebase. Nucleic Acids Res. *46*, D649–D655.
Geer, L.Y., Marchler-Bauer, A., Geer, R.C., Han, L., He, J., He, S., Liu, C., Shi, W., and Bryant, S.H. (2009). The NCBI BioSystems database. Nucleic Acids Res. *38*, 492–496.
Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.
Jennings, R.E., Berry, A.A., Gerrard, D.T., Wearne, S.J., Strutt, J., Withey, S., Chhatriwala, M., Piper Hanley, K., Vallier, L., Bobola, N., et al. (2017). Laser Capture and Deep Sequencing Reveals the Transcriptomic Programmes Regulating the Onset of Pancreas and Liver Differentiation in Human Embryos. Stem Cell Reports *9*, 1387–1394.
Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. *28*, 27–30.
Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res. *45*, D353–D361.
Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., and Tanabe, M. (2019). New approach for understanding genome variations in KEGG. Nucleic Acids Res. *47*, D590–D595.
Kuan, P.F., Chung, D., Pan, G., Thomson, J.A., Stewart, R., and Keleş, S. (2011). A statistical framework for the analysis of ChIP-Seq data. J. Am. Stat. Assoc. *106*, 891–903.
Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.
Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740.
Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 1–21.
Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal *17*, 10–12.
Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., et al. (2019). InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. *47*, D351–D360.
Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E., et al. (2003). PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat. Genet. *34*, 267–273.
Ogawa, S., Surapisitchat, J., Virtanen, C., Ogawa, M., Niapour, M., Sugamori, K.S., Wang, S., Tamblyn, L., Guillemette, C., Hoffmann, E., et al. (2013). Three-dimensional culture and cAMP signaling promote the maturation of human pluripotent stem cell-derived hepatocytes. Development *140*, 3285–3296.
Rezania, A., Bruin, J.E., Arora, P., Rubin, A., Batushansky, I., Asadi, A., O'Dwyer, S., Quiskamp, N., Mojibian, M., Albrecht, T., et al. (2014). Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells. Nat. Biotechnol. *32*, 1121–1133.
Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389–393.
Slenter, D.N., Kutmon, M., Hanspers, K., Riutta, A., Windsor, J., Nunes, N., Mélius, J., Cirillo, E., Coort, S.L., DIgles, D., et al. (2018). WikiPathways: A multifaceted pathway database bridging metabolomics to other omics research. Nucleic Acids Res. *46*, D661–D667.
Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. *102*, 15545–15550.
Veazey, K.J., and Golding, M.C. (2011). Selection of stable reference genes for quantitative RT-PCR comparisons of mouse embryonic and Extra-Embryonic stem cells. PLoS One *6*.
Wang, L., Wang, S., and Li, W. (2012). RSeQC: Quality control of RNA-seq experiments. Bioinformatics *28*, 2184–2185.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.