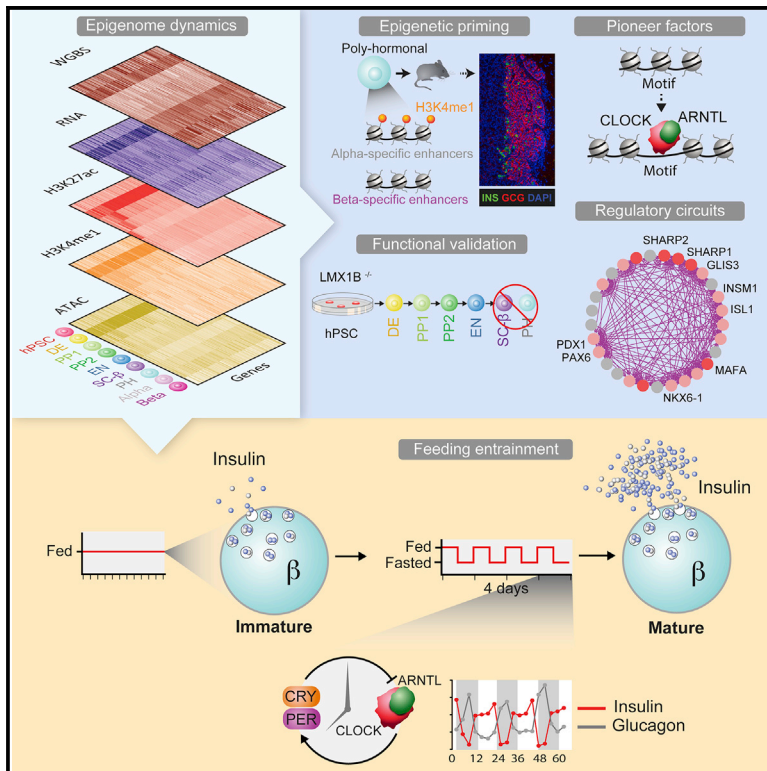


# Cell Stem Cell

## Circadian Entrainment Triggers Maturation of Human *In Vitro* Islets

### Graphical Abstract



### Authors

Juan R. Alvarez-Dominguez,  
Julie Donaghey, Niloofar Rasouli, ...,  
Juerg R. Straubhaar,  
Alexander Meissner, Douglas A. Melton

### Correspondence

dmelton@harvard.edu

### In Brief

Epigenomic studies reveal mechanisms driving differentiation of human stem cells into pancreatic islets and uncover a role for circadian rhythms in their functional maturation. Circadian-clock-entrained islets gain metabolic oscillations and stable chromatin changes that trigger cyclic insulin responses with a raised glucose threshold and function within days of transplant.

### Highlights

- Epigenome dynamics of human-stem-cell-derived islet differentiation and maturation
- Pioneer factors coordinate pervasive enhancer priming to steer endocrine cell fates
- Core regulatory circuits identify LMX1B as critical for endocrine cell generation
- Circadian rhythms trigger islet maturation via clock-controlled metabolic cycles

# Circadian Entrainment Triggers Maturation of Human *In Vitro* Islets

Juan R. Alvarez-Dominguez,<sup>1</sup> Julie Donaghey,<sup>1</sup> Niloofar Rasouli,<sup>1</sup> Jennifer H.R. Kenty,<sup>1</sup> Aharon Helman,<sup>1</sup> Jocelyn Charlton,<sup>1,2</sup> Juerg R. Straubhaar,<sup>1</sup> Alexander Meissner,<sup>1,2</sup> and Douglas A. Melton<sup>1,3,\*</sup>

<sup>1</sup>Department of Stem Cell and Regenerative Biology, Harvard Stem Cell Institute, Harvard University, Cambridge, MA 02138, USA

<sup>2</sup>Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Berlin 14195, Germany

<sup>3</sup>Lead Contact

\*Correspondence: [dmelton@harvard.edu](mailto:dmelton@harvard.edu)

<https://doi.org/10.1016/j.stem.2019.11.011>

## SUMMARY

Stem-cell-derived tissues could transform disease research and therapy, yet most methods generate functionally immature products. We investigate how human pluripotent stem cells (hPSCs) differentiate into pancreatic islets *in vitro* by profiling DNA methylation, chromatin accessibility, and histone modification changes. We find that enhancer potential is reset upon lineage commitment and show how pervasive epigenetic priming steers endocrine cell fates. Modeling islet differentiation and maturation regulatory circuits reveals genes critical for generating endocrine cells and identifies circadian control as limiting for *in vitro* islet function. Entrainment to circadian feeding/fasting cycles triggers islet metabolic maturation by inducing cyclic synthesis of energy metabolism and insulin secretion effectors, including antiphase insulin and glucagon pulses. Following entrainment, hPSC-derived islets gain persistent chromatin changes and rhythmic insulin responses with a raised glucose threshold, a hallmark of functional maturity, and function within days of transplantation. Thus, hPSC-derived tissues are amenable to functional improvement by circadian modulation.

## INTRODUCTION

Generating functional stem-cell-derived tissues is a major goal of regenerative medicine, yet current strategies yield products that often fail to recapitulate endogenous cellular function (Sances et al., 2016; Sneddon et al., 2018; Yang et al., 2014). Our group and others advanced methods to differentiate human pluripotent stem cells (hPSCs) into pancreatic islet cells (Pagliuca et al., 2014; Rezaei et al., 2014; Russ et al., 2015). These comprise SC- $\beta$  (hPSC-derived  $\beta$ ) cells, among other cell types (Pagliuca et al., 2014; Sharon et al., 2019; Veres et al., 2019), including insulin<sup>+</sup> glucagon<sup>+</sup> polyhormonal cells (PH). Better control over deriving all islet cell types is limited by incomplete understanding of the mechanisms driving endocrine lineage specification.

hPSC-derived islets (SC-islets) show glucose-responsive insulin release and cure diabetic rodents, yet they lack the full

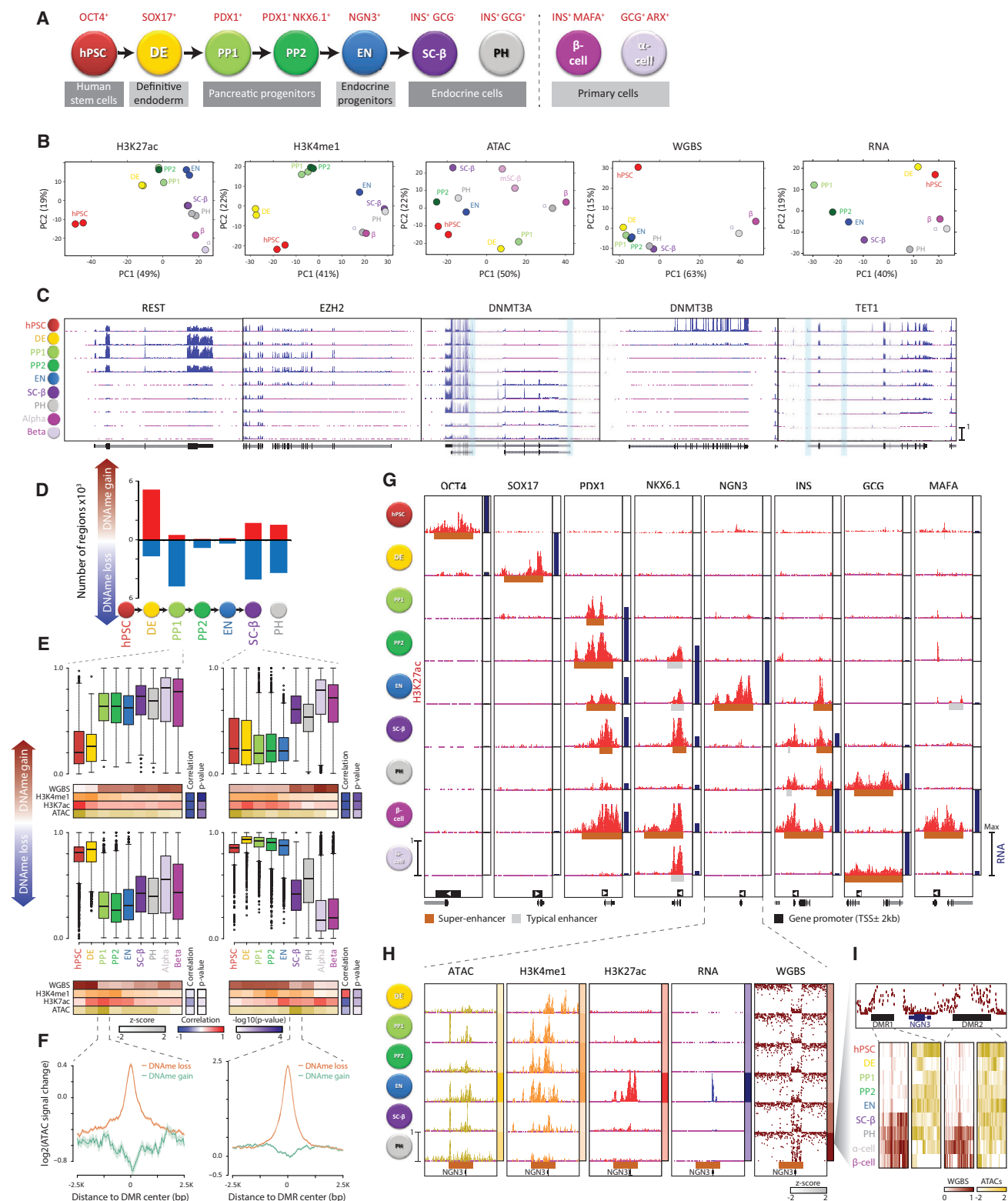
magnitude of responsiveness shown by mature islets *in vivo*. Mature responsiveness develops postnatally as insulin secretion capacity and the glucose threshold for secretion increase (Aguayo-Mazzucato et al., 2006; Blum et al., 2012; Stolovich-Rain et al., 2015). While factors promoting islet maturation have been described (e.g., MAFA and NEUROD1), the underlying mechanisms remain unclear (Liu and Hebrock, 2017).

Here, we use epigenome analyses to better understand mechanisms driving human islet differentiation and maturation. We define regulatory landscapes, the putative pioneer factors that establish them, and their state dynamics throughout islet development. We find that enhancer turnover occurs mainly upon lineage commitment, which is foreshadowed by widespread yet dynamic enhancer priming. Accordingly, we show that priming of  $\alpha$ -cell-specific enhancers steers PH toward an  $\alpha$ -cell fate. Modeling core regulatory circuits (CRCs) across islet differentiation and maturation stages captures both known and unexpected regulators such as LMX1B, which we validate as critical for generating endocrine cells. Finally, contrasting regulatory landscapes in SC- $\beta$  and mature  $\beta$  cells reveals a role for circadian rhythms in fostering mature glucose responsiveness. Entrainment to daily feeding/fasting cycles activates islet clocks and elicits rhythmic transcription of energy metabolism and insulin synthesis/transport/release genes, triggering metabolic rhythms and cyclic insulin responses with a higher glucose threshold, thus recapitulating an aspect of postnatal maturation. Clock-entrained SC-islets gain stable epigenetic changes at genes enabling mature insulin responses and sustain function from as early as 3 days after *in vivo* transplantation. These data, available via an online resource ([http://meltonlab.rc.fas.harvard.edu/data/pancreatic\\_enhancers/](http://meltonlab.rc.fas.harvard.edu/data/pancreatic_enhancers/)), reveal mechanisms controlling human islet development and illustrate how clock modulation can be harnessed to further functional maturation of a stem-cell-derived product.

## RESULTS

### Epigenome Dynamics during Islet Lineage Progression

We exploited the stepwise differentiation of islets from  $\sim 10^8$  hPSCs (Millman et al., 2016; Pagliuca et al., 2014) to purify lineage intermediates in numbers not available through human biopsy samples (Figure 1A). These included early and late pancreatic progenitors (PP1 and PP2) and endocrine progenitors (ENs), isolated based on known markers, and we found that DPP4/CD26 labels glucagon<sup>+</sup> cells, allowing sorting live SC- $\beta$



**Figure 1. Epigenome Dynamics during Human Islet Differentiation**

(A) Stages of directed differentiation from human pluripotent stem cells (hPSCs) to hormone-producing islet cell types.

(B) Stepwise developmental trajectory of hPSC-derived islet cells reconstructed by principal component (PC) analysis of ATAC, H3K4me1, H3K27ac, RNA, and WGBS landscapes. Stage labels as in (A). The fraction of sample variance explained is indicated for each component.

(legend continued on next page)

from PH. hPSC-derived and primary  $\alpha/\beta$  cells were then subjected to whole-genome bisulfite sequencing (WGBS), assay for transposase-accessible chromatin by sequencing (ATAC-seq), chromatin immunoprecipitation sequencing for two histone marks (H3K27ac and H3K4me1), and directional total RNA sequencing (RNA-seq). In total, we generated ~200 datasets with ~12 billion reads aligned to the human genome (Table S1). Replicate epigenome and transcriptome measurements showed high consistency and independently recapitulate the stepwise development of the islet lineage from hPSCs (Figure 1B).

Epigenome and total RNA maps reveal major switches upon endocrine commitment. Profiling enhancer/promoter histone marks H3K27ac and H3K4me1 (Creighton et al., 2010; Heintzman et al., 2009; Rada-Iglesias et al., 2011) locates >360,000 enriched sites (Figure S1A). Despite consistent genome coverage and marking levels across differentiation stages, >75% of H3K27ac/H3K4me1 sites are detected in at most 3 stages (Figures S1B–S1E). These dynamic chromatin sites are mostly distal to promoters and are overrepresented in EN and  $\alpha/\beta$  cells, suggesting widespread remodeling of enhancers. Moreover, total RNA maps detect, among ~2,000 genes with differential isoform usage (Table S2), an endocrine switch toward *DNMT3A* and *TET1* isoforms generated from alternative promoters (Figure 1C) that confer distinct subnuclear localization/biochemical activities (Chen et al., 2002; Gu et al., 2018; Zhang et al., 2016).

DNA methylation dynamics further reveal substantial endocrine remodeling. We detect >125,000 differentially methylated regions across stage transitions, comprising distal intergenic sites with focal H3K27ac/H3K4me1 (Figures 1D and S1G; Table S3). Hypermethylation is highest upon pluripotency exit and reflects stable chromatin silencing (Figures 1D and 1E, top). Hypomethylation, by contrast, reflects stage-specific chromatin activation and is most prevalent upon pancreatic/endocrine specification (Figures 1D and 1E, bottom). Indeed, hypomethylation co-occurs with focal H3K27ac/H3K4me1 and ATAC gain (Figure 1F), shows greater stage specificity than hypermethylation (Figure S1H), and occurs at sites enriched for binding of factors that regulate the respective stage (Figure S1E). Critically, hyper-/hypomethylation concurs with silencing/activation of the nearest gene (Figures 1I and S1I). These data chart epigenome dynamics throughout islet lineage progression, linking them to the control of stage-specific gene expression.

## Widespread Enhancer Resetting upon Lineage Commitment

We used H3K27ac enrichment outside promoters to define >34,000 putative enhancer domains active during islet differentiation (Figures 1G and S1J; Table S4). These are broad H3K27ac regions that often overlap the gene with which they were associated (Figures S1K–S1M) and show striking stage specificity. For example, the *SOX17* domain is specific to endoderm, the *NGN3* domain is specific to EN, and the *MAFA* domain is specific to mature  $\beta$  cells (Figure 1G). In all, 85% of enhancer domains change activity across stage transitions, eliciting changes in DNA openness, H3K4me1 marking, and RNA production (Figure 2A). Notably, differential enhancer activity concurs with differential expression of the overlapping/nearest gene (Figure 2B).

Enhancer establishment, marked by H3K4me1 deposition, co-occurs with focal DNA opening and methylation loss (Figures S2A and S2B), consistent with targeting by sequence-specific transcription factors (TFs) (Felsenfeld et al., 1996; Stadler et al., 2011). A stepwise acquisition of these marks can be seen at key lineage-specifying TF loci. For instance, *NGN3* distal enhancers gain ATAC and H3K4me1 marking in PP1, prior to H3K27ac gain and *NGN3* expression in EN (Figure 1H). Similar conclusions can be drawn for master regulators of other differentiation stages (*SOX17*, *PDX1*, and *NKX6.1*). Thus, stepwise epigenetic changes can forecast gene programs that drive developmental transitions (see below).

Clustering enhancers by epigenome/RNA dynamics (Figures 2C–2E) reveals 10 groups: I–II share enhancers formed before pancreatic specification, III–IV cluster enhancers formed in committed pancreatic progenitors, and V–VIII comprise enhancers formed upon endocrine specification and activated in progenitor/terminally differentiated endocrine cells. We also found sets of enhancers (group IX) that are lost from hPSCs and re-form upon islet cell maturation, including loci specifically active in  $\beta$  cells (e.g., *MAFA*, *UCN3*, and *SIX3*) (Figure S2C).

Dissecting enhancer turnover shows that only 10% of islet lineage enhancers are preset in hPSCs. We investigated at which stage enhancers are gained/lost between hPSC and differentiated progeny (Figures 2F and S2D). Strikingly, most (~60%) enhancers gained in PP1 or EN form in the first lineage progenitor, with reduced formation upon subsequent differentiation. Similarly, enhancer loss and activation/deactivation are highest at the pancreatic/endocrine lineage branching

(C) Dynamic expression and isoform usage patterns of key epigenome regulators. Tracks display normalized RNA. Gene models are shown in black below. Alternative *DNMT3A* and *TET1* promoters are highlighted.

(D) DNA methylation dynamics during directed differentiation. Regions showing DNA methylation (DNAm) gain or loss are highest upon endoderm or pancreatic/endocrine specification, respectively.

(E) Stable epigenetic silencing upon DNAm gain (top) versus activation upon DNAm loss (bottom). Boxplots show distribution of DNAm levels at all stages for regions differentially methylated during PP1 (left) and SC- $\beta$  (right) specification. Heatmaps below display their median WGBS, H3K4me1, H3K27ac, and ATAC levels at each stage, with correlations between WGBS and H3K4me1/H3K27ac/ATAC patterns and their significance quantified to the right.

(F) DNAm gain/loss coincides with focal ATAC loss/gain. Shown is the ATAC signal change around differentially methylated regions from definitive endoderm (DE) to PP1 (left) and from EN to SC- $\beta$  (right).

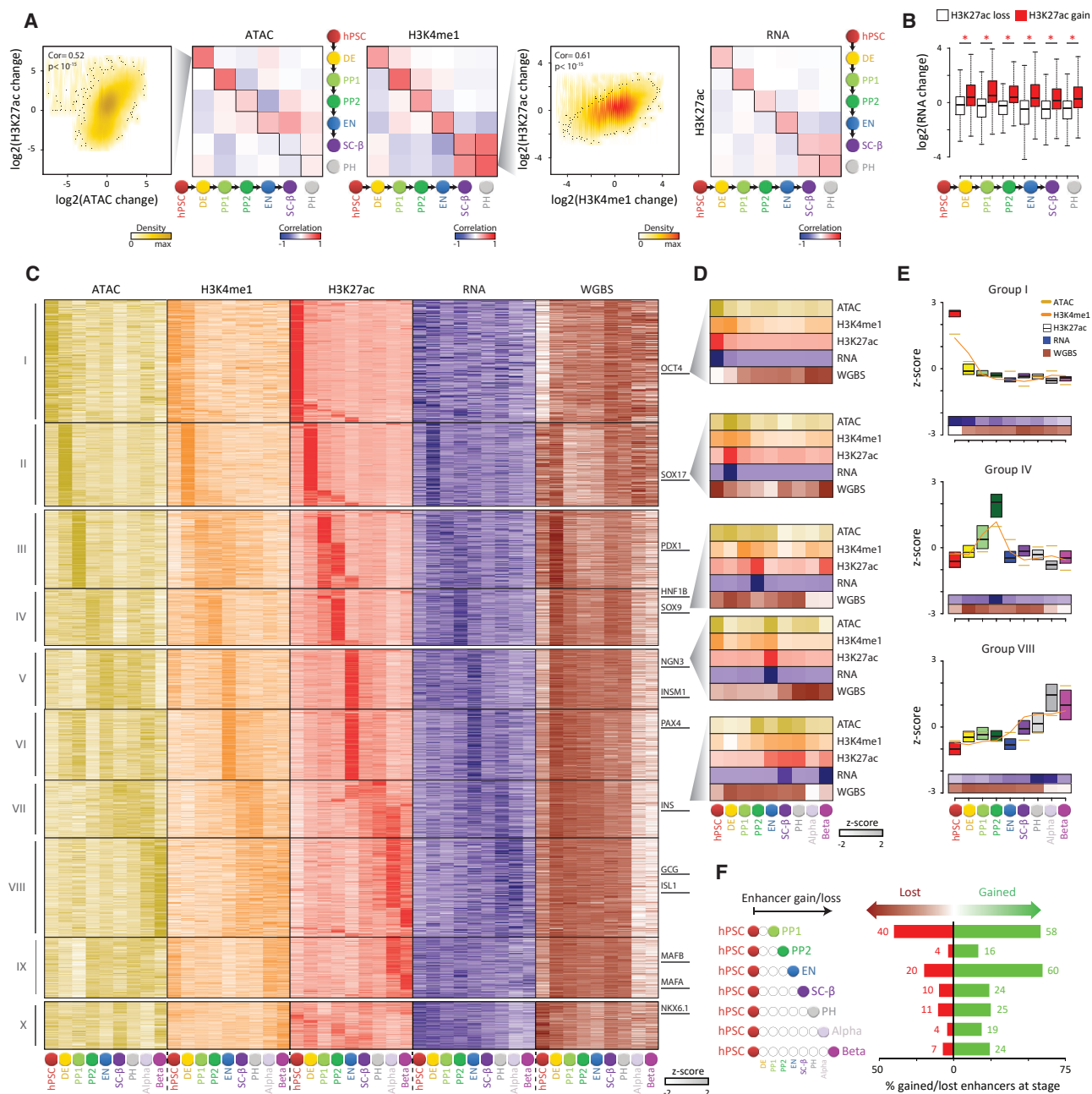
(G) Stage-specific H3K27ac and RNA dynamics across islet development. Tracks show normalized H3K27ac signal over a region 1.5 times greater than the superenhancer domain at the stage marked by genes highlighted in (A), with relative gene expression quantified to the right.

(H) Coordinated *NGN3* ATAC, H3K4me1, H3K27ac, RNA, and WGBS dynamics. Tracks as in (G); heatmaps to the right display relative signal over the domain shown below.

(I) Concerted chromatin closing and DNAm gain within *NGN3*-flanking DMRs. Heatmaps show WGBS/ATAC levels at individual CpGs.

See also Figure S1 and Tables S1–S3.





**Figure 2. Enhancer Transitions during Islet Lineage Progression**

(A and B) Synchronous changes in chromatin accessibility, modification, and transcription during directed differentiation. Correlations between H3K27ac and ATAC/H3K4me1/RNA changes across all enhancer domains for each pair of differentiation stages shown in (A). Concordance between enhancer H3K27ac loss/gain and differential expression of the overlapping/nearest gene across subsequent developmental transitions shown in (B);  $p < 10^{-24}$  (Wilcoxon test).

(C–E) Transitioning enhancer states throughout the islet lineage. Clustering of enhancer domains based on their ATAC, H3K4me1, H3K27ac, RNA, and WGBS profiles (C). Profiles for domains associated with the genes highlighted to the right shown in (D). Profiles for select enhancer clusters representing hPSC (I), pancreatic progenitors (IV), and hormone-producing islet cells (VIII) are summarized in (E) by their median values (ATAC and H3K4me1), boxplots (H3K27ac), and heatmaps (RNA and WGBS).

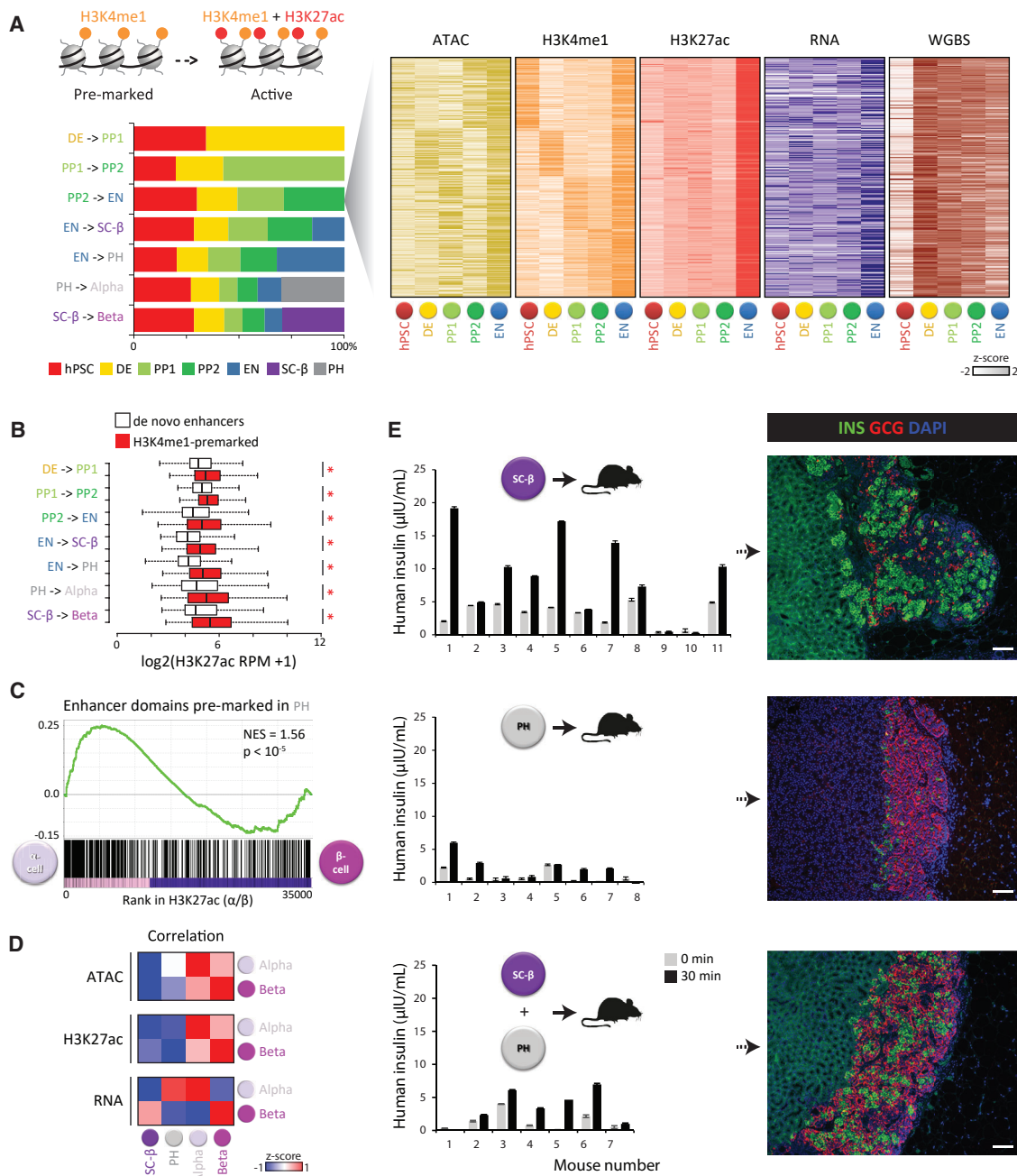
(F) Widespread enhancer turnover at lineage branching points. For enhancers gained/lost between hPSCs and each stage, the fraction specifically gained/lost at that stage is shown.

See also Figure S2 and Table S4.

points (Figure S2E). These data define enhancer dynamics from pluripotent to postmitotic islet cells, revealing that enhancer potential is largely reset upon lineage commitment.

### Epigenetic Priming Predicts Lineage Potential

The competence to execute specific cell fates following inductive cues can be linked to a gain of H3K4me1 before H3K27ac



**Figure 3. Extensive Epigenetic Priming Predicts Lineage Potential**

(A) Substantial turnover of enhancer H3K4me1 pre-marking. Left: for enhancers activated at each stage transition that were previously H3K4me1-marked, the fraction whose maximal H3K4me1 deposition occurred at each previous stage is shown. Right: epigenome/RNA dynamics for 2,097 enhancers activated in EN that were H3K4me1 marked in previous stages.

(B) H3K4me1-premarked enhancers gain higher H3K27ac levels than *de novo* enhancers across stage transitions; \* $p < 10^{-8}$  (Wilcoxon test).

(C) Selective priming of  $\alpha$ -specific enhancers in PHs. Enrichment analysis of enhancers inactive but H3K4me1-marked in PHs by their H3K27ac level in  $\alpha$  versus  $\beta$  cells. NES, normalized enrichment score.

(D) PHs resemble  $\alpha$  over  $\beta$  cells. Correlations among ATAC (top), H3K27ac (middle), and RNA (bottom) levels across all enhancer domains between SC- $\beta$ /PH and primary  $\alpha$ / $\beta$  cells.

(E) PHs resolve toward  $\alpha$  cells *in vivo*. Left: fasted immunocompromised mice transplanted with purified SC- $\beta$ , PH, or unsorted re-aggregated cells were assayed for serum human insulin before/30 min after a glucose injection 4–6 weeks post-transplantation. Data are mean  $\pm$  SEM of 3 replicate measurements. Right: hormonal resolution evidenced by staining retrieved grafts for insulin (green)/glucagon (red). Scale bar, 100  $\mu$ m.

See also Figure S3.

deposition at lineage-specific enhancers (Creyghton et al., 2010; Wamstad et al., 2012; Zhang et al., 2012). We find that only 10%–20% of enhancers activated across stage transitions gain H3K4me1 in the preceding stage, enriching for critical effectors of the subsequent one (Figures 2D and S3A). However, up to 60% of newly activated enhancers are H3K4me1 marked in earlier stages and show dynamic H3K4me1 gain/loss prior to activation (Figures 3A and S3C), revealing H3K4me1-premarking to be pervasive but substantially plastic. Importantly, H3K27ac and RNA levels are higher at H3K4me1-premarked versus *de novo* enhancers (Figures 3B and S3B), compatible with a transcriptionally primed epigenetic state.

We investigated to which extent enhancer priming reflects lineage potential. Primed-to-active enhancer transitions between EN and  $\alpha/\beta$  are comparable, supporting EN multipotency, whereas SC- $\beta$  priming is biased toward  $\beta$ -specific enhancers (Figure S3D). Intriguingly, PHs are biased toward  $\alpha$ -specific enhancers (Figure 3C), suggesting predisposition to become  $\alpha$  cells, which they resemble more closely than  $\beta$  cells (Veres et al., 2019) (Figure 3D). To test this, we capitalized on our ability to purify PH and SC- $\beta$  and transplanted them alone or in combination under the kidney capsule of immunocompromised mice (Figures 3E and S3E). Unlike mice receiving SC- $\beta$  cells, those receiving comparable numbers of PH could not sustain robust glucose-stimulated insulin secretion (GSIS). Kidney grafts were then retrieved and stained for insulin/glucagon. Unlike SC- $\beta$  grafts, which contain mainly insulin<sup>+</sup> monohormonal cells, PH grafts contain mainly insulin<sup>+</sup> glucagon<sup>+</sup> cells (Figures 3E and S3F), demonstrating  $\alpha$ -cell resolution. Thus, an epigenetic program steers PH toward an  $\alpha$ -cell fate.

### Pioneer Factors during Islet Differentiation and Maturation

Developmental competence is endowed by chromatin-opening pioneer factors (Iwafuchi-Doi and Zaret, 2014). We find that newly opened chromatin sites during islet differentiation enrich for binding sites of known pioneer TFs (Figures S4A and S4B) and show focal demethylation (Figures S4C and S4D), in line with pioneer activity. We identify pioneer FOXA/GATA and HNF/MEIS factors in PP1 and PP2, RFX/NEUROD in EN, and MAFA factors in islet cells, along with unexpected TFs such as NHLH1/2 in EN and IRF1, ELF3, and AP-1 components (FOS, FOSL2, and JUNB) in  $\beta$  cells (Figure 4A; Table S5). Stage-specific expression patterns (Figure S4E) support roles for these factors in directing islet lineage progression.

To predict maturation pioneers, we surveyed DNA openness over a 4-week *in vitro* time course during which SC- $\beta$  cells gain glucose responsiveness (Figures 4B and S4F). Robust function (GSIS >2  $\mu$ IU/10<sup>3</sup> cells and >1.5-fold stimulation) was seen by week 3, reflecting gains in both insulin content and secretory capacity. With this transition, we detect newly opened sites overlapping/near insulin as well as genes regulating its processing and secretion (e.g., *CADPS*, *SYT4*, and *STX2*), in SC- $\beta$  cells, but not PHs (Figures 4C and S4G). Surprisingly, these sites feature the core circadian clock activators (CLOCK and ARNTL/BMAL1) among the top-enriched TF binding motifs (Figure 4D; Table S5; see below). Conversely, known controllers MAFA and NEUROD1 only appear by week 4, suggesting an incomplete maturation process.

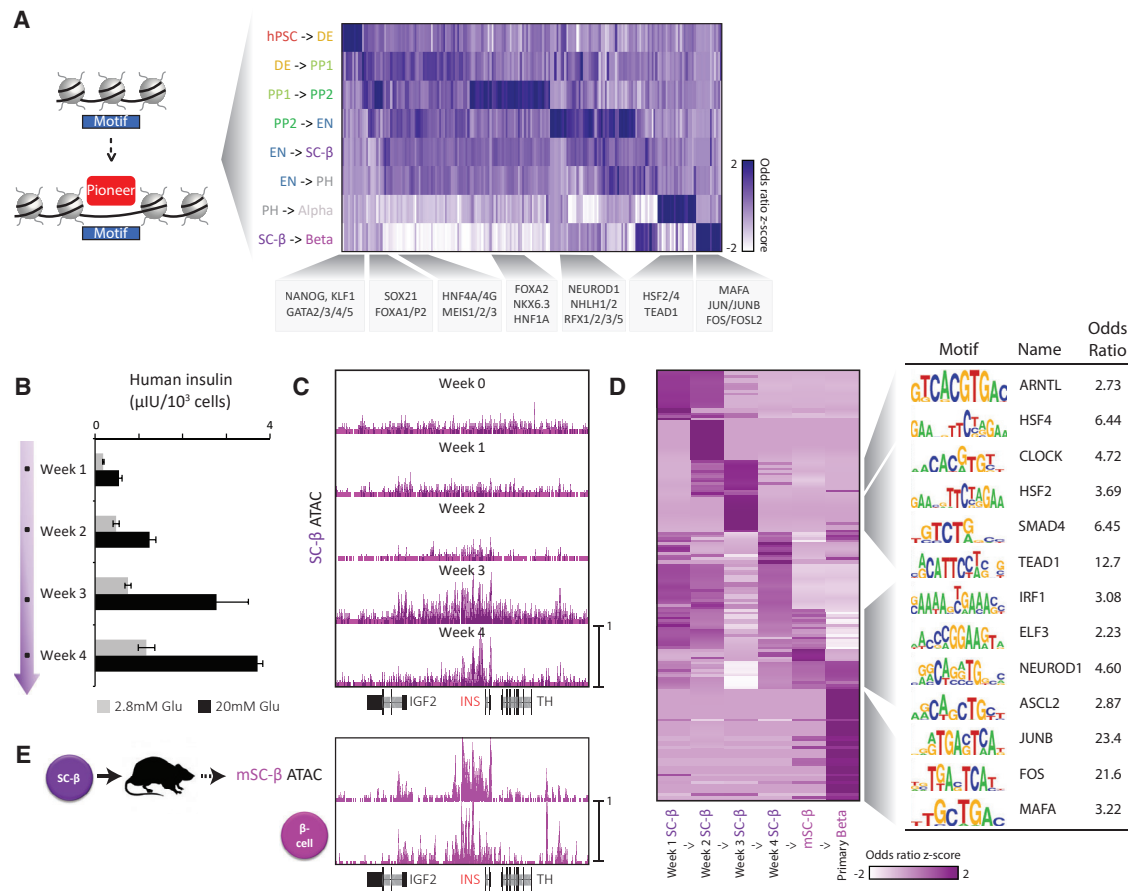
To examine changes upon further maturation *in vivo*, we transplanted SC- $\beta$  into immunocompromised rats, which enabled recovery of enough cells for ATAC-seq (Figures 4E and S4H). Engrafted SC- $\beta$  cells more closely resemble  $\beta$  cells (Figure 1B), gaining >40,000 open chromatin sites, overrepresented at TF (e.g., *ISL1* and *PAX6*) and secretory genes (e.g., *UCN3* and *SLC30A8*), which are often shared by both (Figure S4I; Table S5). These sites enrich for MAFA/NEUROD1 binding motifs, among other unexpected TFs (e.g., IRF1 and ELF3) (Figure 4D). These data reveal putative pioneer TFs throughout islet development, highlighting potential regulators of the onset and refinement of  $\beta$  cell function.

### Islet Differentiation and Maturation Regulatory Circuits

Stable cell states are reportedly set by TFs via autoregulatory loops involving joint formation of extended or superenhancers (SEs) (Gaulton et al., 2010; Parker et al., 2013; Whyte et al., 2013). A cell's CRC can thus be modeled by finding SE-driven TFs in interconnected autoregulatory loops (Boyer et al., 2005; Lin et al., 2016; Saint-André et al., 2016). Using this logic, we generated CRC models that effectively capture master pancreatic/endocrine regulators (Figure 5; Table S6). For example, PP2 CRCs contain PDX1, NKX6.1, SOX9, and ONECUT1/HNF6, while  $\beta$  CRCs contain MAFA, PAX6, and ISL1. Binding among CRC factors is supported by immunoprecipitation studies (not shown), and depleting a given TF triggers selective depletion of the others in its CRC (Figure S5A), verifying strong interconnectivity.

How are CRCs remodeled during development? We reasoned that lineage-inductive cues could trigger activation of new CRC-forming TF sets that in turn reprogram cell state. Indeed, lineage-specifying CRC factors show high stage specificity (Figure 5C). For example, endoderm-specifying (GSC, EOMES, and SOX17) and pancreatic-specifying (PDX1, SOX9, and ONECUT1) TFs are highly endoderm and PP1 specific, respectively, verifying earlier studies (Oliver-Krasinski and Stoffers, 2008). This analysis also points to unanticipated TFs such as LMX1B (specific to EN) and the circadian factors BHLHE40/SHARP2, BHLHE41/SHARP1 (specific to  $\beta$  cells), and HLF (specific to  $\alpha$  cells). These show high interconnectivity (Figure S5C), including with key TFs that partake in CRCs at multiple stages (PDX1, NKX6.1, and PAX6) (Figure 5D), suggesting that stage-specific TFs rewire otherwise stable CRCs during development.

We tested whether LMX1B regulates the EN state. Depleting LMX1B during *in vitro* endocrine induction limited the pool of EN progenitors and hampered endocrine gene activation (Figures S5D and S5E). We thus disrupted LMX1B using an inducible CRISPR/Cas9 system (González et al., 2014) in independent hPSC lines (Figures 5E and S5F). LMX1B<sup>-/-</sup> lines differentiated normally up to EN progenitor formation, yet subsequent SC- $\beta$  and PH generation was drastically impaired, with median <1% SC- $\beta$  cells formed (Figures 5E and S5G). RNA profiling consistently reveals hampered induction of EN CRC TFs (PAX4, INSM1, and MAFA), pan-endocrine TFs (NEUROD1 and PAX6), secretory proteins (CHGA and SCG5), and hormones (insulin, glucagon, and somatostatin) (Figures 5F and S5H). Conversely, TFs (ONECUT/HES) and signaling pathways (Wnt, Notch, and Bmp) selective for pancreatic progenitors (Sharon et al., 2019) remain active in LMX1B<sup>-/-</sup> cells, indicating



**Figure 4. Pioneer Factors during Islet Differentiation and Maturation**

(A) Recognition motifs within newly opened chromatin regions (left diagram) enrich for known and unexpected pioneer TFs during islet development (right). Heatmap shows relative motif enrichment for all expressed TFs across developmental transitions.

(B) Refinement of glucose-stimulated insulin secretion (GSIS) function during extended culture of terminal-stage SC-β preparations. Data are mean ± SEM of N = 3 preparations with n = 3 replicate measurements each.

(C) Concomitant opening of DNA sites around the insulin locus as SC-β cells gain robust GSIS. Tracks display normalized ATAC signal from SC-β cells.

(D) Putative pioneer factors during β cell maturation. Left: TF motif enrichments as in (B) across subsequent *in vitro/in vivo* SC-β maturation stages, as well as between SC-β and primary β cells. Right: select TF motifs among the most enriched during key maturation stage transitions.

(E) Sustained insulin openness in SC-β transplanted into immunocompromised rats. Tracks as in (E) display ATAC signal from engrafted SC-β/primary β cells. See also Figure S4 and Table S5.

a differentiation block. Thus, LMX1B is critical for *in vitro* endocrine cell differentiation.

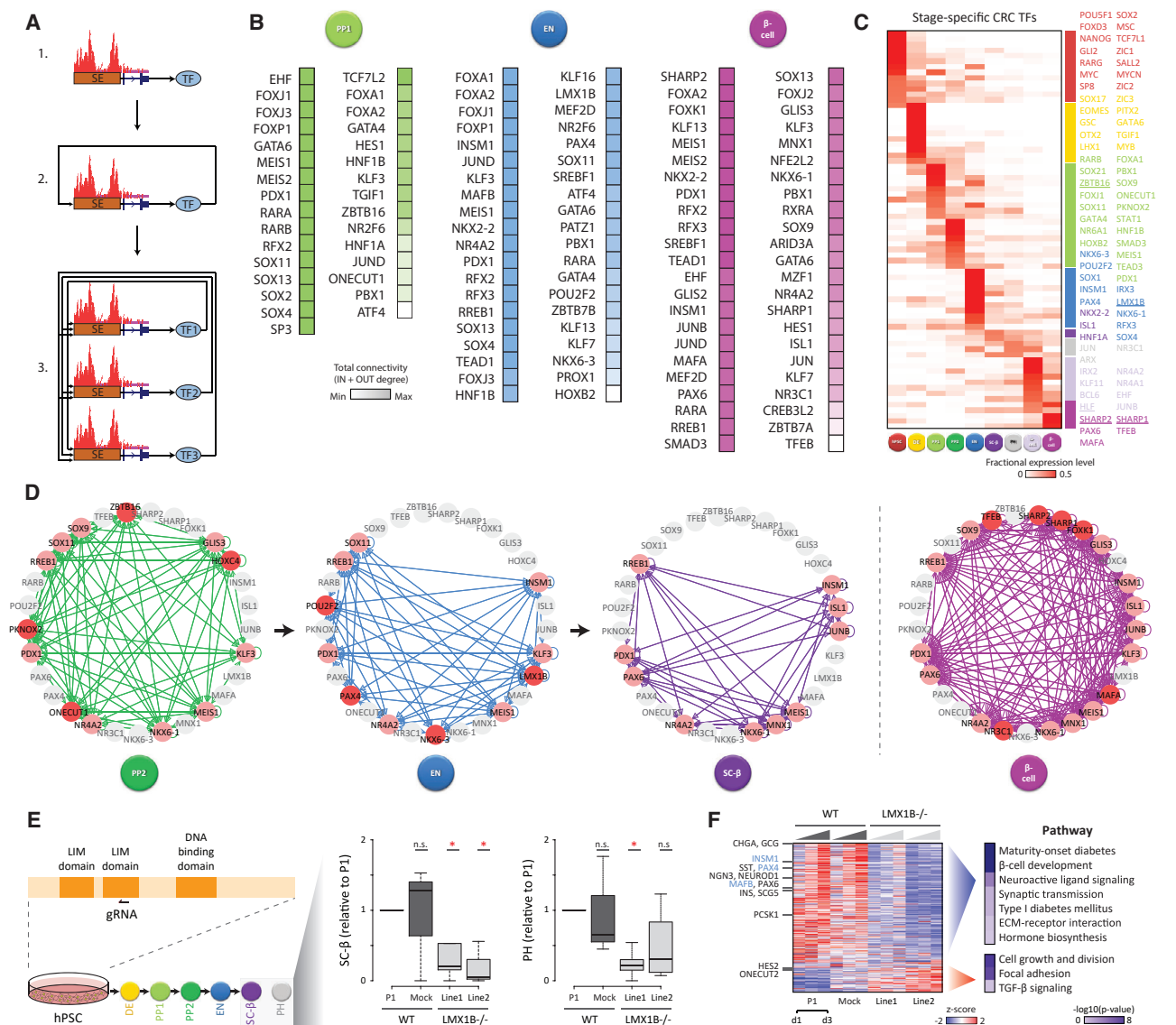
### Circadian Rhythms Trigger an Islet Maturation Step

Circadian clocks are vital for attuning insulin secretion to daily feeding/fasting cycles (Bass and Takahashi, 2010; Gamble et al., 2014). Clock controllers are linked to islet maturation by their chromatin and expression patterns (Figures S4E and S5B), footprint at regulatory sites gained as SC-β cells raise insulin production/secretion (Figure 4D), and interconnections with maturity-linked factors in the β cell CRC (Figure 5D). We thus asked if clock entrainment can foster *in vitro* islet maturation. First, we verified that insulin responses of cadaveric and SC-islets can be entrained by a single metabolic shock (Figure S6A), as seen in rodents (Perelis et al., 2015). After recovery from shock, cultures grown in constant medium show time-of-day variation in insulin content and stimulated secretion, with peak production/

secretion 12 h post-recovery. Fold-stimulation is modestly increased relative to parallel mock-treated cultures at this time (Figure S6B). Entrainment to daily feeding/fasting rhythms using various stimuli (glucose, arginine, forskolin, and insulin), however, can boost insulin stimulation capacity by as much as 6-fold (Figures 6A, S6C, and S6D). Surprisingly, this reflects lower responses to basal (2.8 mM) glucose (Figure 6B) without significant change in overall responses to stimulatory (20 mM glucose/30 mM KCL) conditions or overall insulin content (Figure S6E). Enhanced responsiveness reflects new cellular states, as individual circadian-entrained islet cells flux significantly more calcium than most (~75%) mock-treated counterparts (Figure 6C).

We then tested our optimized entrainment method in SC-islets. The results in Figures 6D and S6F show cyclic insulin production/secretion, sustaining rhythmic GSIS for 72 h post-entrainment. As with natural islets, we detect lower basal insulin secretion and enhanced glucose responsiveness, which





**Figure 5. CRCs Identify LMX1B as Critical for Endocrine Cell Generation**

(A and B) Modeling core transcriptional regulatory circuits (CRCs) in the human islet lineage. 1. Superenhancer (SE)-driven TFs are identified. 2. TFs predicted to bind their own SE are detected. 3. CRCs are defined as sets of autoregulated TFs predicted to bind each other's SE, forming fully interconnected autoregulatory loops (A). The union set of TFs from all possible models for PP1, EN, and  $\beta$  are listed in (B), ranked by their interconnectivity.

(C) Lineage-specifying CRC TFs are distinguished by their stage-specific expression. Newly implicated TFs are underlined to the right.

(D) Rewiring of CRCs by stage-specific TFs. Nodes represent stage-specific CRC TFs from (C) that partake (pink) or not (gray) in PP2, EN, SC- $\beta$ , or  $\beta$  CRCs. Edges represent predicted transcriptional regulatory relationships between TFs within the same CRC. TFs unique to each stage are highlighted in red.

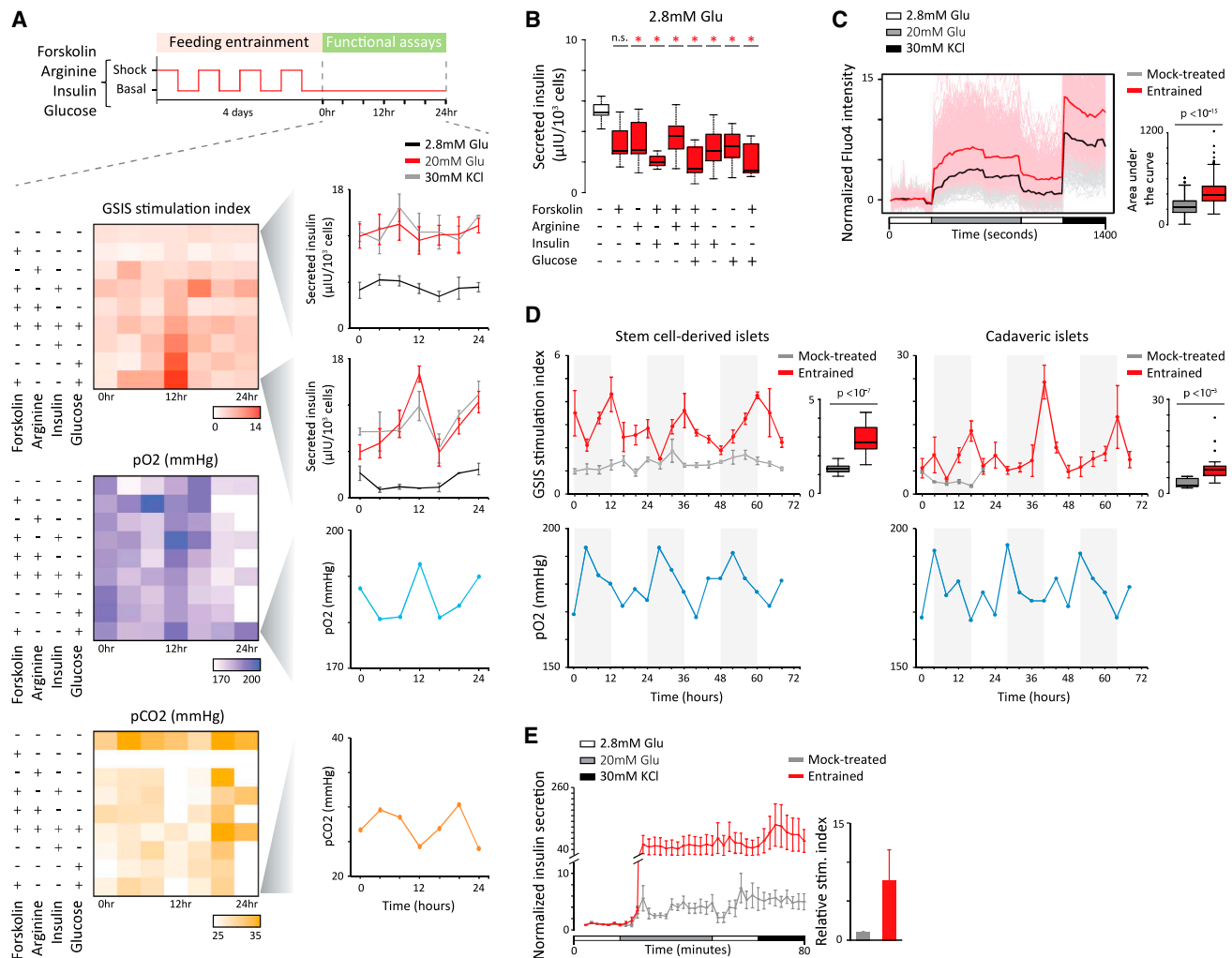
(E) LMX1B regulates *in vitro* endocrine cell generation. CRISPR/Cas9 guide RNA (gRNA)-mediated *LMX1B* disruption in hPSC followed by directed differentiation (left) significantly ( $p < 0.05$ , t test) impairs SC- $\beta$  (middle) and PH (right) generation relative to parental and mock-targeted control lines. Results were normalized to the parental control from the same experiment and combined for statistical analysis. Data from N = 4–5 differentiation experiments with n = 1–2 biological replicates each.

(F) LMX1B disruption blocks endocrine gene induction. Left: differentially expressed ( $p < 0.05$ , DESeq) genes in independent *LMX1B*<sup>-/-</sup> versus parental lines profiled from day 1 to day 3 of *in vitro* endocrine differentiation. EN CRC TFs are highlighted in blue, among other TFs, secretory proteins, and hormones. Right: pathways enriched among differentially regulated genes.

See also Figure S5 and Table S6.

can reach 8-fold enhancement at peak secretion compared to mock-treated counterparts (Figures 6D, 6E, and S6G). Interestingly,  $O_2/CO_2$  levels in the medium of unstimulated entrained cultures are also rhythmic (Figures 6A and 6D), linking GSIS oscillations to metabolic rhythms. Of note, cadaveric/SC-

islets lacking GSIS function do not respond to entrainment (not shown), in line with roles in tuning, but not eliciting, glucose responses. Thus, circadian entrainment can improve *in vitro* islet function by raising the glucose threshold for insulin secretion, a hallmark of functional maturity (Blum et al., 2012).



**Figure 6. Circadian Rhythms Trigger an Islet Maturation Step**

(A) Human islet glucose-stimulated insulin secretion (GSIS) and respiration rhythms induced by circadian entrainment with various stimuli. Schematic: timeline for shock/recovery cycles conducted 4 times over 4 days, followed by functional assays over 24 h. Panels: GSIS stimulation indexes (top), partial O<sub>2</sub> pressure in the medium of unstimulated cultures (middle), and partial CO<sub>2</sub> pressure (bottom), with patterns for mock-treated versus entrained cultures shown to the right. Data are mean ± SEM of n = 3 replicate measurements.

(B) Diminished insulin response to basal glucose levels in clock-entrained islet cultures. Boxplots summarize insulin secretion across 24 h for the indicated conditions; \*p < 0.01 relative to mock-treated (t test).

(C) Circadian entrainment enhances islet glucose responsiveness. Calcium influx in mock-treated versus entrained islets during the indicated incubations, detected using Fluo4-AM staining. Shown are population- and single cell-level cytosolic Ca<sup>2+</sup> traces, summarized to the right. Traces show n = 725 cells from N = 14 islets sampled from both cultures at the 12-h circadian time point, each normalized to the mean of the first incubation.

(D) GSIS and respiration rhythms in clock-entrained cadaveric/SC-islet cultures sampled across 72 h. Top: GSIS stimulation indexes for mock-treated versus entrained SC-islets (left) and cadaveric islets (right), summarized to the right of each. Data are mean ± SEM of n = 3 replicate measurements. Bottom: partial O<sub>2</sub> pressure in the medium of unstimulated entrained SC-islets (left) and cadaveric islets (right).

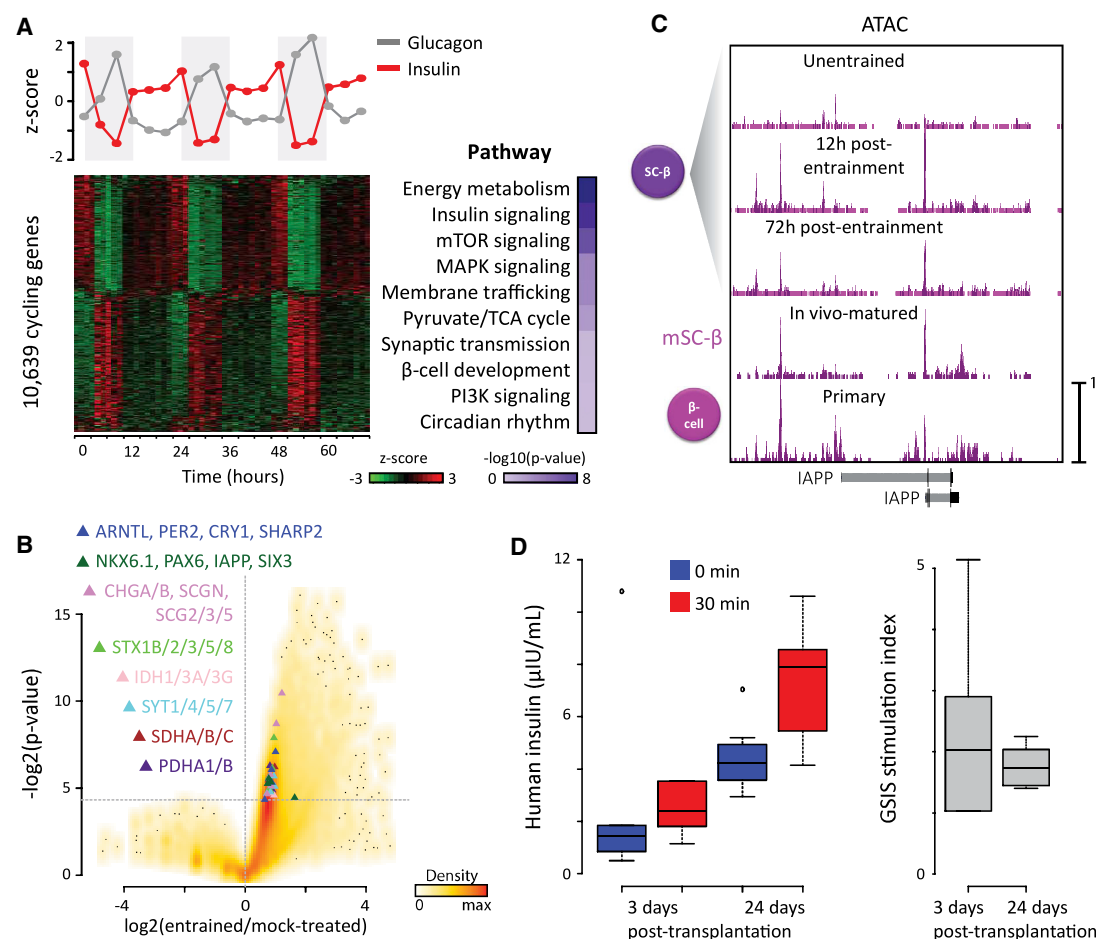
(E) Circadian entrainment enhances SC-islet glucose responsiveness. Insulin secretion from mock-treated versus entrained SC-islets during the indicated incubations, with relative stimulation indexes quantified to the right. Data are mean ± SEM of N = 3 cultures sampled at the 12-h circadian time point with n = 2 replicate measurements each, normalized to mean of the first incubation.

See also Figure S6.

### Genomic Changes and *In Vivo* Function of Clock-Entrained SC-Islets

How do circadian rhythms promote *in vitro* islet maturation? RNA sequencing identifies energy metabolism enzymes among >10,000 genes (31% of all protein- and non-coding genes detected) that oscillate specifically upon clock entrainment (Figures 7A and S7A–S7C; Table S7). These include

glucose import/metabolism (GCK and PDHA1/B/X) and tricarboxylic acid (TCA) cycle enzymes, as well as electron transport chain and ATP synthase components. Consistent with their cycling, the mTOR/phosphatidylinositol 3-kinase (PI3K)/AKT and MAPK signaling factors are overrepresented among rhythmic RNAs. These data link islet metabolic rhythms to the pulsatile expression of energy metabolism genes.



**Figure 7. Clock-Entrained SC Islets Gain Durable Rhythmicity and Function 3 Days after Transplant**

(A) Rhythmic RNA expression in SC-islet cultures sampled for 72 h following circadian entrainment. Top: antiphasic insulin and glucose expression; data are mean of  $N = 3$  cultures. Bottom left: RNA expression of cycling genes ( $p < 0.05$ , harmonic regression test for rhythmicity); data pooled from  $N = 3$  cultures. Bottom right: pathways enriched in the cycling gene set.

(B) Upregulation of core clock TFs, maturity-linked factors, and machinery involved in energy metabolism and insulin secretion in SC-islet cultures following circadian entrainment. Volcano plot summarizes magnitude and statistical significance of expression changes. Data as in (A).

(C) Newly opened DNA sites around *IAPP* persist in SC-islets after circadian entrainment. Tracks display normalized ATAC signal at the indicated stages.

(D) Clock-entrained SC-islets function within 3 days of transplant. Fasted immunocompromised mice transplanted with SC-islets following circadian entrainment were assayed for serum human insulin before/30 min after a glucose injection 3- and 24-days post-transplantation. Glucose-stimulated insulin secretion (GSIS) stimulation indexes are summarized to the right. Data from  $N = 10$  mice with  $n = 2$  replicate measurements each.

See also Figure S7 and Table S7.

We also detect antiphasic insulin and glucagon expression (Figure 7A), reminiscent of priming of insulin/glucagon secretion according to diurnal/nocturnal physiological need (Bass and Takahashi, 2010; Gamble et al., 2014). Indeed, we observe widespread cycling of factors mediating insulin metabolism (*PCSK1* and *IDE*), signaling, and transport/secretion (Figure S7D). Surprisingly, the circadian transcriptome includes *PDX1*, *NKX6.1*, and maturity-linked factors (*NEUROD1* and *IAPP*), all of which show high-amplitude oscillations (Figure S7E). These data provide a molecular basis for circadian variation in insulin responses via the rhythmic control of its synthesis, transport, and release.

Clock-controlled oscillations may already exist without entrainment but remain hidden from bulk measurements due to lack of synchronicity. However, single-cell RNA-seq from

unentrained SC- $\beta$  cultures (Veres et al., 2019) (Figure S7F) shows that unlike TFs (e.g., *PDX1* and *NKX6.1*) detected consistently ( $>60\%$  of cells) and robustly (mean counts per million mapped reads (CPM)  $>2$ ), core clock controllers (*ARNTL*, *PER1/2*, and *CRY1/2*) are expressed rarely ( $<15\%$  of cells) and lowly (mean CPM  $\sim 1$ ). By contrast, their expression is induced upon entrainment (Figure 7B), indicating that feeding/fasting cycles not only synchronize but also activate islet clocks.

Circadian entrainment also induces maturity-linked factors (*PAX6*, *IAPP*, and *SIX3*) along with machinery involved in energy metabolism (pyruvate/TCA cycle enzymes) and insulin secretion (synaptotagmins, syntaxins, and secretogranins) (Figure 7B). Higher/rhythmic expression of these effectors is consistent with greater/pulsatile GSIS responses that persist 72 h post-entrainment (Figure 6D). To study the basis for such persistence,

we used ATAC-seq to detect ~20,000 newly opened chromatin sites 12 h following entrainment, with ~18,500 remaining after 72 h (Figures 7C and S7G). These sites are indeed overrepresented at TF genes regulating the mature  $\beta$  cell phenotype (*MAFA*, *NEUROD1*, and *PAX6*) and at direct GSIS effectors (*G6PC2/3*, *SLC30A8*, and *RAB11FIP2*), suggesting that circadian control of genes enabling mature GSIS function persists as a result of stable chromatin changes.

To test whether maturation *in vitro* leads to better function *in vivo*, we transplanted clock-entrained SC-islets under the kidney capsule of immunocompromised mice (Figure 7D). Remarkably, entrained SC-islets show GSIS  $>2 \mu\text{U/ml}$  and  $>1.5$ -fold stimulation as early as 3 days post-transplantation and maintain this function 24 days after. Circadian entrainment can thus trigger stable changes reminiscent of postnatal maturation to foster robust GSIS, revealing a mechanistic aspect of the mature islet phenotype.

## DISCUSSION

The promise of *in-vitro*-derived tissues for regenerative medicine benefits from a better understanding of mechanisms governing their development and function. Prior SC-islet studies described epigenetic processes influencing pancreatic specification (Cebola et al., 2015; Lee et al., 2019; Wang et al., 2015; Xie et al., 2013; Xu et al., 2014). Extending these to specification of endocrine cell types has been hampered by their inefficient/incomplete derivation, which yields highly heterogeneous preparations (Pagliuca et al., 2014; Rezanian et al., 2014; Russ et al., 2015). Here, we circumvent these challenges using cell purification strategies that enable deep epigenome profiling throughout the specification, terminal differentiation, and functional maturation of  $\alpha$ - and  $\beta$ -like cells. Our data chart an epigenomic roadmap of *in vitro* islet differentiation, revealing a chromatin-level determination of endocrine cell fates and function.

How epigenome changes coordinate genetic circuits to steer cell fate is little understood. We find that lineage decisions involve pervasive enhancer remodeling and identify widespread chromatin transitions that foreshadow fate choices. The results show that enhancer repertoires are reset during differentiation rather than preset (Choukrallah et al., 2015; Lara-Astiaso et al., 2014; Luyten et al., 2014). Two major waves of enhancer gain/loss mark the pancreatic/ endocrine commitment points, entailing concurrent turnover of DNA accessibility, methylation, and histone marking. These changes gate access of stage-specific TFs to their target sites, linking chromatin transitions to the rewiring of transcriptional circuits that control cell fates.

We also detect extensive H3K4me1 premarking between progenitor and differentiated cell states and suggest it fosters transcriptionally permissive enhancer states. Such epigenetic priming may promote robust, concerted activation of lineage-specifying factors. Indeed, we find that insulin<sup>+</sup> glucagon<sup>+</sup> cells primed for activating  $\alpha$ - versus  $\beta$ -selective enhancers resolve toward an  $\alpha$ -cell fate. Preparations containing such polyhormonal cells reportedly yield mainly insulin<sup>+</sup>  $\alpha$ -like cells upon engraftment/extended culture (Rezanian et al., 2011; Veres et al., 2019), but these could in principle derive from other precursors in the preparations. Our transplant studies resolve this controversy by directly tracking the fate of purified PHs.

Cell fate programming has been linked to CRCs formed by super-enhancer-driven TFs (Lin et al., 2016; Saint-André et al., 2016). How these circuits are rewired to program sequential cell fates during development is unclear. We find that islet lineage CRCs are dynamically reconfigured by stage-specific gain/loss of SE-driven TFs (Adam et al., 2015; Goode et al., 2016). One such factor, LMX1B, is indeed critical for *in vitro* islet cell derivation. Interestingly, LMX1B controls development of dopaminergic/serotonergic neurons downstream of NKX2.2 (Cheng et al., 2003; Smidt et al., 2000) and can transactivate insulin ectopically (German et al., 1992), suggesting a related pathway shapes SC-islet development.

Finally, we show that the key function of SC-islets, glucose responsiveness, can be improved by circadian modulation. Mature responsiveness develops between birth and weaning (Aguayo-Mazzucato et al., 2006; Blum et al., 2012; Stolovich-Rain et al., 2015), along with the onset of circadian behavioral (sleep and feeding) cycles that induce islet clocks (Rakshit et al., 2018). Interestingly, embryonic or adult  $\beta$  cell-specific *Arntl/Bmal1* deletion in rodents yields islets with diminished insulin secretory capacity, reminiscent of neonatal/immature islets (Marcheva et al., 2010; Perelis et al., 2015; Rakshit et al., 2018). Our gain-of-function studies further dissect how clocks modulate islet function in a controlled, human setting. We show that re-creating fasting/feeding cycles activates islet clocks, inducing rhythmic transcription of genes controlling energy metabolism and GSIS function. Accordingly, clock-entrainment drives antiphasic waves of insulin and glucagon expression, suggesting that the capacity to anticipate diurnal changes in insulin demand can be recreated *in vitro* via organ-autonomous rhythms. These changes coordinate metabolic and insulin cycles, fostering an increased GSIS threshold in SC-islets that otherwise expand basal insulin secretion as they expand their production/secretory capacity. The same mechanism operates in primary islets, recapitulating an important aspect of *in vivo* maturation. Importantly, this functional step persists 72 h after entrainment has ceased, likely due to stable epigenetic changes at gene loci enabling mature GSIS, and renders SC-islets functional within days of transplant.

Overall, this study provides insights that underpin human islet development and function. These insights may be valuable in improving  $\beta$  cell programming strategies and understanding how disrupted genetic circuits contribute to metabolic diseases, including diabetes. An intriguing possibility is that circadian entrainment may be harnessed to further functional maturation of other stem-cell-derived products (Sances et al., 2016; Sneddon et al., 2018; Yang et al., 2014), consistent with the ability of clock controllers to bind distinct targets in distinct tissues (Koike et al., 2012; Perelis et al., 2015). Thus, our general approach may inform attempts to control the fate and function of any human cell type.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY



## ● EXPERIMENTAL MODEL AND SUBJECT DETAILS

- Human cell lines
- Primary human samples
- Rodent strains

## ● METHOD DETAILS

- Cell culture adaptation
- Cell differentiation
- Cell purification
- ChIP-seq
- ATAC-seq
- WGBS
- RNA-seq
- ChIP-seq mapping
- ATAC-seq mapping
- WGBS mapping
- RNA-seq mapping
- Coverage quantification
- Enhancer and super-enhancer identification
- Enhancer state dynamics
- Clustering analyses
- Transplantation studies
- Glucose-stimulated insulin secretion assays
- Immunohistochemistry
- Putative pioneer factor identification
- Core regulatory circuits
- shRNA studies
- NanoString and qPCR studies
- CRISPR/Cas9-mediated genome editing
- *In vitro* circadian entrainment
- Calcium influx assays
- RNA rhythmicity and differential expression
- Region-based annotations
- Gene set and pathway enrichment analysis
- Motif enrichment analysis
- Additional bioinformatics methods

## ● QUANTIFICATION AND STATISTICAL ANALYSIS

## ● DATA AND CODE AVAILABILITY

## ● ADDITIONAL RESOURCES

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.stem.2019.11.011>.

## ACKNOWLEDGMENTS

We thank Deanne Watson, Dani Swain, Jeff Davis, Ramona Pop, and Samantha Collins for reagents and assistance with experiments; the HSCRB histology and flow cytometry, BPF next-gen sequencing, and Bauer flow cytometry and sequencing core facilities at Harvard University for technical support and critical discussions; and N. Slavov, A. Yuan, K. Villa, and members of the Melton and Meissner laboratories for critical feedback on this manuscript. J.R.A.-D. is a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation. D.A.M. is an investigator of the Howard Hughes Medical Institute. This work was supported by the Max Planck Society and NIH grants DP3K111898 and P01GM099117 (A.M.) and by grants from the Juvenile Diabetes Research Foundation, Helmsley Charitable Trust, and JPB Foundation.

## AUTHOR CONTRIBUTIONS

J.R.A.-D., J.D., N.R., J.H.R.K., and A.H. performed experiments. J.R.A.-D., J.D., J.C., and J.R.S. conducted bioinformatics analyses. J.R.A.-D., J.D.,

J.H.R.K., A.M., and D.A.M. designed the research, interpreted results, and wrote the manuscript.

## DECLARATION OF INTERESTS

D.A.M. serves on the scientific advisory board of Semma/Vertex. The remaining authors declare no competing interests.

Received: July 5, 2019

Revised: October 7, 2019

Accepted: November 19, 2019

Published: December 12, 2019

## SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Gutiérrez et al. (2017); Teo et al. (2011); and Wang et al. (2012b).

## REFERENCES

- Adam, R.C., Yang, H., Rockowitz, S., Larsen, S.B., Nikolova, M., Oristian, D.S., Polak, L., Kadaja, M., Asare, A., Zheng, D., and Fuchs, E. (2015). Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. *Nature* 521, 366–370.
- Aguayo-Mazzucato, C., Sanchez-Soto, C., Godínez-Puig, V., Gutiérrez-Ospina, G., and Hiriart, M. (2006). Restructuring of pancreatic islets and insulin secretion in a postnatal critical window. *PLoS ONE* 1, e35.
- Alvarez-Dominguez, J.R., Knoll, M., Gromatzky, A.A., and Lodish, H.F. (2017a). The super-enhancer-derived alncRNA-EC7/Bloodline potentiates red blood cell development in trans. *Cell Rep.* 19, 2503–2514.
- Alvarez-Dominguez, J.R., Zhang, X., and Hu, W. (2017b). Widespread and dynamic translational control of red blood cell development. *Blood* 129, 619–629.
- Ameri, J., Borup, R., Prawiro, C., Ramond, C., Schachter, K.A., Scharfmann, R., and Semb, H. (2017). Efficient generation of glucose-responsive beta cells from isolated GP2<sup>+</sup> human pancreatic progenitors. *Cell Rep.* 19, 36–49.
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* 11, R106.
- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.* 25, 25–29.
- Bass, J., and Takahashi, J.S. (2010). Circadian integration of metabolism and energetics. *Science* 330, 1349–1354.
- Bastian, M., Heymann, S., Jacomy, M.. Gephi: An Open Source Software for Exploring and Manipulating Networks. International AAAI Conference on Web and Social Media. 2009. Available at: <https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154>. Date accessed: 05 Dec. 2019.
- Blodgett, D.M., Nowosielska, A., Afik, S., Pechhold, S., Cura, A.J., Kennedy, N.J., Kim, S., Kucukural, A., Davis, R.J., Kent, S.C., et al. (2015). Novel Observations From Next-Generation RNA Sequencing of Highly Purified Human Adult and Fetal Islet Cell Subsets. *Diabetes* 64, 3172–3181.
- Blum, B., Hrvatin, S., Schuetz, C., Bonal, C., Rezanian, A., and Melton, D.A. (2012). Functional beta-cell maturation is marked by an increased glucose threshold and by expression of urocortin 3. *Nat. Biotechnol.* 30, 261–264.
- Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122, 947–956.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218.

- Cebola, I., Rodríguez-Seguí, S.A., Cho, C.H., Bessa, J., Rovira, M., Luengo, M., Chhatrivala, M., Berry, A., Ponsa-Cobas, J., Maestro, M.A., et al. (2015). TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors. *Nat. Cell Biol.* 17, 615–626.
- Chen, T., Ueda, Y., Xie, S., and Li, E. (2002). A novel Dnmt3a isoform produced from an alternative promoter localizes to euchromatin and its expression correlates with active de novo methylation. *J. Biol. Chem.* 277, 38746–38754.
- Cheng, L., Chen, C.L., Luo, P., Tan, M., Qiu, M., Johnson, R., and Ma, Q. (2003). Lmx1b, Pet-1, and Nkx2.2 coordinately specify serotonergic neurotransmitter phenotype. *J. Neurosci.* 23, 9961–9967.
- Choukallah, M.A., Song, S., Rolink, A.G., Burger, L., and Matthias, P. (2015). Enhancer repertoires are reshaped independently of early priming and heterochromatin dynamics during B cell differentiation. *Nat. Commun.* 6, 8324.
- Cogger, K.F., Sinha, A., Sarangi, F., McGaugh, E.C., Saunders, D., Dorrell, C., Mejia-Guerrero, S., Aghazadeh, Y., Rourke, J.L., Sreanor, R.A., et al. (2017). Glycoprotein 2 is a specific cell surface marker of human pancreatic progenitors. *Nat. Commun.* 8, 331.
- Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. USA* 107, 21931–21936.
- Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34, 184–191.
- Donaghey, J., Thakurela, S., Charlton, J., Chen, J.S., Smith, Z.D., Gu, H., Pop, R., Clement, K., Stamenova, E.K., Karnik, R., et al. (2018). Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat. Genet.* 50, 250–258.
- Felsenfeld, G., Boyes, J., Chung, J., Clark, D., and Studitsky, V. (1996). Chromatin structure and gene expression. *Proc. Natl. Acad. Sci. USA* 93, 9384–9388.
- Feng, H., Conneely, K.N., and Wu, H. (2014). A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res.* 42, e69.
- Gamble, K.L., Berry, R., Frank, S.J., and Young, M.E. (2014). Circadian clock control of endocrine factors. *Nat. Rev. Endocrinol.* 10, 466–475.
- Gaulton, K.J., Nammo, T., Pasquali, L., Simon, J.M., Giresi, P.G., Fogarty, M.P., Panhuis, T.M., Mieczkowski, P., Secchi, A., Bosco, D., et al. (2010). A map of open chromatin in human pancreatic islets. *Nat. Genet.* 42, 255–259.
- German, M.S., Wang, J., Chadwick, R.B., and Rutter, W.J. (1992). Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: building a functional insulin minienhancer complex. *Genes Dev.* 6, 2165–2176.
- Gifford, C.A., Ziller, M.J., Gu, H., Trapnell, C., Donaghey, J., Tsankov, A., Shalek, A.K., Kelley, D.R., Shishkin, A.A., Issner, R., et al. (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* 153, 1149–1163.
- González, F., Zhu, Z., Shi, Z.D., Lelli, K., Verma, N., Li, Q.V., and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* 15, 215–226.
- Goode, D.K., Obier, N., Vijayabaskar, M.S., Lie-A-Ling, M., Lilly, A.J., Hannah, R., Lichtinger, M., Batta, K., Florkowska, M., Patel, R., et al. (2016). Dynamic gene regulatory networks drive hematopoietic specification and differentiation. *Dev. Cell* 36, 572–587.
- Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018.
- Gu, T., Lin, X., Cullen, S.M., Luo, M., Jeong, M., Estecio, M., Shen, J., Hardikar, S., Sun, D., Su, J., et al. (2018). DNMT3A and TET1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. *Genome Biol.* 19, 88.
- Gutiérrez, G.D., Bender, A.S., Cirulli, V., Mastracci, T.L., Kelly, S.M., Tsigos, A., Kaestner, K.H., and Sussel, L. (2017). Pancreatic  $\beta$  cell identity requires continual repression of non- $\beta$  cell programs. *J. Clin. Invest.* 127, 244–259.
- Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res.* 22, 1760–1774.
- Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–112.
- Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. *Genes Dev.* 28, 2679–2692.
- Jindal, R.M., Taylor, R.P., Gray, D.W., Esmeraldo, R., and Morris, P.J. (1992). A new method for quantification of islets by measurement of zinc content. *Diabetes* 41, 1056–1062.
- Jolma, A., Kivioja, T., Toivonen, J., Cheng, L., Wei, G., Enge, M., Taipale, M., Vaquerizas, J.M., Yan, J., Sillanpää, M.J., et al. (2010). Multiplexed massively parallel SELEX for characterization of human transcription factor binding specificities. *Genome Res.* 20, 861–873.
- Kenty, J.H., and Melton, D.A. (2015). Testing pancreatic islet function at the single cell level by calcium influx with associated marker expression. *PLoS ONE* 10, e0122044.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36.
- Koike, N., Yoo, S.H., Huang, H.C., Kumar, V., Lee, C., Kim, T.K., and Takahashi, J.S. (2012). Transcriptional architecture and chromatin landscape of the core circadian clock in mammals. *Science* 338, 349–354.
- Lamparter, D., Marbach, D., Rueedi, R., Bergmann, S., and Kutalik, Z. (2017). Genome-wide association between transcription factor expression and chromatin accessibility reveals regulators of chromatin accessibility. *PLoS Comput. Biol.* 13, e1005311.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Immunogenetics. Chromatin state dynamics during blood formation. *Science* 345, 943–949.
- Lee, K., Cho, H., Rickert, R.W., Li, Q.V., Pulecio, J., Leslie, C.S., and Huangfu, D. (2019). FOXA2 is required for enhancer priming during pancreatic differentiation. *Cell Rep.* 28, 382–393.e387.
- 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.
- Lin, C.Y., Erkek, S., Tong, Y., Yin, L., Federation, A.J., Zapotka, M., Haldipur, P., Kawauchi, D., Risch, T., Warnatz, H.J., et al. (2016). Active medulloblastoma enhancers reveal subgroup-specific cellular origins. *Nature* 530, 57–62.
- Liu, J.S., and Hebrok, M. (2017). All mixed up: defining roles for  $\beta$ -cell subtypes in mature islets. *Genes Dev.* 31, 228–240.
- Liu, H., Yang, H., Zhu, D., Sui, X., Li, J., Liang, Z., Xu, L., Chen, Z., Yao, A., Zhang, L., et al. (2014). Systematically labeling developmental stage-specific genes for the study of pancreatic  $\beta$ -cell differentiation from human embryonic stem cells. *Cell Res.* 24, 1181–1200.
- LLoyd, S.P. (1982). Least squares quantization in PCM. *IEEE Trans. Inf. Theory* 28, 128–137.
- Lück, S., Thurlay, K., Thaben, P.F., and Westermark, P.O. (2014). Rhythmic degradation explains and unifies circadian transcriptome and proteome data. *Cell Rep.* 9, 741–751.
- Luyten, A., Zang, C., Liu, X.S., and Shivdasani, R.A. (2014). Active enhancers are delineated de novo during hematopoiesis, with limited lineage fidelity among specified primary blood cells. *Genes Dev.* 28, 1827–1839.
- Marcheva, B., Ramsey, K.M., Buhr, E.D., Kobayashi, Y., Su, H., Ko, C.H., Ivanova, G., Omura, C., Mo, S., Vitaterna, M.H., et al. (2010). Disruption of the clock components CLOCK and BMAL1 leads to hypoinsulinaemia and diabetes. *Nature* 466, 627–631.

- Mashal, R.D., Koontz, J., and Sklar, J. (1995). Detection of mutations by cleavage of DNA heteroduplexes with bacteriophage resolvases. *Nat. Genet.* 9, 177–183.
- McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* 28, 495–501.
- Mentlein, R., Gallwitz, B., and Schmidt, W.E. (1993). Dipeptidyl-peptidase IV hydrolyses gastric inhibitory polypeptide, glucagon-like peptide-1(7–36) amide, peptide histidine methionine and is responsible for their degradation in human serum. *Eur. J. Biochem.* 214, 829–835.
- Millman, J.R., Xie, C., Van Dervort, A., Gürtler, M., Pagliuca, F.W., and Melton, D.A. (2016). Generation of stem cell-derived  $\beta$ -cells from patients with type 1 diabetes. *Nat. Commun.* 7, 11463.
- Muraro, M.J., Dharmadhikari, G., Grun, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning, E.J., et al. (2016). A single-cell transcriptome atlas of the human pancreas. *Cell Syst.* 3, 385–394.e383.
- Oliver-Krasinski, J.M., and Stoffers, D.A. (2008). On the origin of the beta cell. *Genes Dev.* 22, 1998–2021.
- Pagliuca, F.W., Millman, J.R., Gürtler, M., Segel, M., Van Dervort, A., Ryu, J.H., Peterson, Q.P., Greiner, D., and Melton, D.A. (2014). Generation of functional human pancreatic  $\beta$  cells in vitro. *Cell* 159, 428–439.
- Parker, S.C., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Black, B.L., et al.; NISC Comparative Sequencing Program; National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Authors; NISC Comparative Sequencing Program Authors (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. USA* 110, 17921–17926.
- Pasquali, L., Gaulton, K.J., Rodriguez-Segui, S.A., Mularoni, L., Miguel-Escalada, I., Akerman, I., Tena, J.J., Moran, I., Gomez-Marin, C., van de Bunt, M., et al. (2014). Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* 46, 136–143.
- Perelis, M., Marcheva, B., Ramsey, K.M., Schipma, M.J., Hutchison, A.L., Taguchi, A., Peek, C.B., Hong, H., Huang, W., Omura, C., et al. (2015). Pancreatic  $\beta$  cell enhancers regulate rhythmic transcription of genes controlling insulin secretion. *Science* 350, aac4250.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–283.
- Rakshit, K., Qian, J., Gaonkar, K.S., Dhawan, S., Colwell, C.S., and Matveyenko, A.V. (2018). Postnatal ontogenesis of the islet circadian clock plays a contributory role in  $\beta$ -cell maturation process. *Diabetes* 67, 911–922.
- Rezania, A., Riedel, M.J., Wideman, R.D., Karanu, F., Ao, Z., Warnock, G.L., and Kieffer, T.J. (2011). Production of functional glucagon-secreting  $\alpha$ -cells from human embryonic stem cells. *Diabetes* 60, 239–247.
- 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.
- Russ, H.A., Parent, A.V., Ringler, J.J., Hennings, T.G., Nair, G.G., Shveygert, M., Guo, T., Puri, S., Haataja, L., Cirulli, V., et al. (2015). Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro. *EMBO J.* 34, 1759–1772.
- Saint-André, V., Federation, A.J., Lin, C.Y., Abraham, B.J., Reddy, J., Lee, T.I., Bradner, J.E., and Young, R.A. (2016). Models of human core transcriptional regulatory circuitries. *Genome Res.* 26, 385–396.
- Sances, S., Bruijn, L.I., Chandran, S., Eggan, K., Ho, R., Klim, J.R., Livesey, M.R., Lowry, E., Macklis, J.D., Rushton, D., et al. (2016). Modeling ALS with motor neurons derived from human induced pluripotent stem cells. *Nat. Neurosci.* 19, 542–553.
- Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675.
- Sharon, N., Vanderhoof, J., Straubhaar, J., Mueller, J., Chawla, R., Zhou, Q., Engquist, E.N., Trapnell, C., Gifford, D.K., and Melton, D.A. (2019). Wnt signaling separates the progenitor and endocrine compartments during pancreas development. *Cell Rep.* 27, 2281–2291.e2285.
- Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* 15, 284.
- Sherwood, R.I., Hashimoto, T., O'Donnell, C.W., Lewis, S., Barkal, A.A., van Hoff, J.P., Karun, V., Jaakkola, T., and Gifford, D.K. (2014). Discovery of directional and non-directional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* 32, 171–178.
- Shi, Z.D., Lee, K., Yang, D., Amin, S., Verma, N., Li, Q.V., Zhu, Z., Soh, C.L., Kumar, R., Evans, T., et al. (2017). Genome editing in hPSCs reveals GATA6 haploinsufficiency and a genetic interaction with GATA4 in human pancreatic development. *Cell Stem Cell* 20, 675–688.e676.
- Smidt, M.P., Asbreuk, C.H., Cox, J.J., Chen, H., Johnson, R.L., and Burbach, J.P. (2000). A second independent pathway for development of mesencephalic dopaminergic neurons requires Lmx1b. *Nat. Neurosci.* 3, 337–341.
- Sneddon, J.B., Tang, Q., Stock, P., Bluestone, J.A., Roy, S., Desai, T., and Hebrok, M. (2018). Stem cell therapies for treating diabetes: progress and remaining challenges. *Cell Stem Cell* 22, 810–823.
- Soh, C.L., and Huangfu, D. (2017). CRISPR/Cas9-mediated mutagenesis of human pluripotent stem cells in defined Xeno-free E8 medium. *Methods Mol. Biol.* 1498, 57–78.
- Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490–495.
- Stolovich-Rain, M., Enk, J., Vikesa, J., Nielsen, F.C., Saada, A., Glaser, B., and Dor, Y. (2015). Weaning triggers a maturation step of pancreatic  $\beta$  cells. *Dev. Cell* 32, 535–545.
- 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., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.
- Teo, A.K., Arnold, S.J., Trotter, M.W., Brown, S., Ang, L.T., Chng, Z., Robertson, E.J., Dunn, N.R., and Vallier, L. (2011). Pluripotency factors regulate definitive endoderm specification through eomesodermin. *Genes Dev.* 25, 238–250.
- Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31, 46–53.
- Tsankov, A.M., Gu, H., Akopian, V., Ziller, M.J., Donaghey, J., Amit, I., Gnirke, A., and Meissner, A. (2015). Transcription factor binding dynamics during human ES cell differentiation. *Nature* 518, 344–349.
- Veres, A., Faust, A.L., Bushnell, H.L., Engquist, E.N., Kenty, J.H., Harb, G., Poh, Y.C., Sintov, E., Gürtler, M., Pagliuca, F.W., et al. (2019). Charting cellular identity during human in vitro  $\beta$ -cell differentiation. *Nature* 569, 368–373.
- Wamstad, J.A., Alexander, J.M., Truty, R.M., Shrikumar, A., Li, F., Eilertson, K.E., Ding, H., Wylie, J.N., Pico, A.R., Capra, J.A., et al. (2012). Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* 151, 206–220.
- Wang, L., Wang, S., and Li, W. (2012a). RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 28, 2184–2185.
- Wang, Z., Oron, E., Nelson, B., Razis, S., and Ivanova, N. (2012b). Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell* 10, 440–454.
- 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.

- Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K., et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* 158, 1431–1443.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319.
- Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPPING program. *BMC Bioinformatics* 10, 232.
- 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.
- Xu, C.R., Li, L.C., Donahue, G., Ying, L., Zhang, Y.W., Gadue, P., and Zaret, K.S. (2014). Dynamics of genomic H3K27me3 domains and role of EZH2 during pancreatic endocrine specification. *EMBO J.* 33, 2157–2170.
- Yang, X., Pabon, L., and Murry, C.E. (2014). Engineering adolescence: maturation of human pluripotent stem cell-derived cardiomyocytes. *Circ. Res.* 114, 511–523.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-seq (MACS). *Genome Biol.* 9, R137.
- Zhang, J.A., Mortazavi, A., Williams, B.A., Wold, B.J., and Rothenberg, E.V. (2012). Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* 149, 467–482.
- Zhang, W., Xia, W., Wang, Q., Towers, A.J., Chen, J., Gao, R., Zhang, Y., Yen, C.A., Lee, A.Y., Li, Y., et al. (2016). Isoform switch of TET1 regulates DNA demethylation and mouse development. *Mol. Cell* 64, 1062–1073.
- Ziller, M.J., Gu, H., Muller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500, 477–481.



## STAR★METHODS

### KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Goat anti SOX17	R&D	Cat# AF1924, RRID:AB_355060
Goat anti PDX1	R&D	Cat# AF2419, RRID:AB_355257
Mouse anti NKX6.1	DSHB	Cat# F55A12, RRID:AB_532379
Sheep anti NGN3	R&D systems	Cat# AF3444, RRID:AB_2149527
Rat anti C-PEPTIDE	DSHB	Cat# GN-ID4, RRID:AB_2255626
Goat anti GLP2	Santa Cruz Biotechnology	Cat# sc-7781, RRID:AB_2107346
Mouse anti GP2	MBL International	Cat# D277-5, RRID:AB_11160953
Mouse anti SUSD2	Milentyi Biotec	Cat# 130-106-401, RRID:AB_2653618
Mouse anti DPP4	Milentyi Biotec	Cat# 130-093-441, RRID:AB_1103210
Mouse anti HPI2	Novus Biologicals	NBP1-18946AF488, RRID:AB_2818946
Mouse anti TM4SF4	Novus Biologicals	Cat# FAB7998R, RRID:AB_2818948
Rabbit anti H3K27Ac	Active Motif	Cat# 39133, RRID:AB_2561016
Mouse anti H3K4me1	Millipore	Cat# 17-614, RRID:AB_11212770
<b>Biological Samples</b>		
Human islets	Prodo Laboratories	N/A
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
TSQ	Thermo Fisher Scientific	Cat# M688
Forskolin	Stemgent	Cat# 04-0025
Insulin	Sigma	Cat# I9278
Fluo4-AM	Life Technologies	Cat# F-14217
<b>Critical Commercial Assays</b>		
NEBNext Ultra II DNA Library Prep Kit	New England Biolabs	Cat# E7103
Agencourt AMPure	Beckman Coulter	Cat# A63881
EZ DNA Methylation-Gold kit	Zymo Research	Cat# D5005
Accel-NGS Methyl-Seq DNA library kit	Swift Biosciences	Cat# 30024
Stranded RNA-seq Library Preparation kit	KAPA Biosystems	Cat# KR0934
SuperScript III first-strand synthesis kit	Life Technologies	Cat# 18080051
MEGAshortscript T7 Transcription kit	Thermo Fisher Scientific	Cat# AM1354
MEGAclear Transcription Clean-Up kit	Thermo Fisher Scientific	Cat# AM1908
<b>Deposited Data</b>		
Raw and analyzed sequencing data	This paper	GEO: GSE139817
<b>Experimental Models: Cell Lines</b>		
Human: HUES 8 hESC line	HSCI	NIHhESC-09-0021
Human: HUES 8 iCas9 hESC line	<a href="#">González et al., 2014</a>	Derived from the parental line (NIHhESC-09-0021)
Human: H1 iCas9 hESC line	<a href="#">Shi et al., 2017</a>	Derived from the parental line (NIHhESC-10-0043)
<b>Experimental Models: Organisms/Strains</b>		
Mouse: CB17.Cg-Prkdc <sup>scid</sup> Lysf <sup>bg-j</sup> /Crl	Charles River	Strain Code 250
Rat: Crl:NIH-Foxn1 <sup>mu</sup>	Charles River	Strain Code 316
<b>Oligonucleotides</b>		
LMX1B sgRNA TGTGAACGGCAGCTACGCAA	This paper	N/A
LMX1B sgRNA flanking PCR primer F CCATGGAAGAGTCTGGAGCA	This paper	N/A

(Continued on next page)

## Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
LMX1B sgRNA flanking PCR primer R GCACACAAGCACAATCCACA	This paper	N/A
LMX1B qPCR primer F GGTCCAGGTCTGGTTTCAGAAC	This paper	N/A
LMX1B qPCR primer R GTGTAGGAAGCCATCATGCCCT	This paper	N/A
Recombinant DNA		
GIPZ LMX1B lentiviral shRNA1	Dharmacon	RHS4430-200298213
GIPZ LMX1B lentiviral shRNA2	Dharmacon	RHS4430-200296395
GIPZ LMX1B lentiviral shRNA3	Dharmacon	RHS4430-200296748
GIPZ LMX1B lentiviral shRNA4	Dharmacon	RHS4430-200294187
GIPZ Lentiviral Empty Vector shRNA Control	Dharmacon	RHS4349
GIPZ Non-silencing Lentiviral shRNA Control	Dharmacon	RHS4346
GIPZ GAPDH Lentiviral shRNA Positive Control	Dharmacon	RHS4371
GIPZ EG5 Lentiviral shRNA Positive Control	Dharmacon	RHS4480
Software and Algorithms		
Bowtie2	Langmead and Salzberg, 2012	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
MACS2	Zhang et al., 2008	<a href="https://github.com/taoliu/MACS">https://github.com/taoliu/MACS</a>
BEDTools	Quinlan and Hall, 2010	<a href="https://bedtools.readthedocs.io/en/latest/">https://bedtools.readthedocs.io/en/latest/</a>
BMap	Xi and Li, 2009	<a href="https://sites.google.com/a/brown.edu/bioinformatics-in-biomed/bsmap-for-methylation">https://sites.google.com/a/brown.edu/bioinformatics-in-biomed/bsmap-for-methylation</a>
TopHat2	Kim et al., 2013	<a href="https://ccb.jhu.edu/software/tophat/index.shtml">https://ccb.jhu.edu/software/tophat/index.shtml</a>
R v.3.3.2	R Core Team, 2016	<a href="http://www.Rproject.org/">http://www.Rproject.org/</a>
nSolver 3.0	NanoString Technologies	<a href="https://www.nanostring.com/products/analysis-software/nsolver">https://www.nanostring.com/products/analysis-software/nsolver</a>
Other		
Online resource	This paper	<a href="http://meltonlab.rc.fas.harvard.edu/data/pancreatic_enhancers/">http://meltonlab.rc.fas.harvard.edu/data/pancreatic_enhancers/</a>

## LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to the Lead Contact, Douglas A. Melton ([dmelton@harvard.edu](mailto:dmelton@harvard.edu)). All unique/stable reagents generated in this study are available from the Lead Contact with a completed Materials Transfer Agreement.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Human cell lines

The HUES8 (NIH hESC registry #09-0021; male) line was used for directed differentiation, and lines derived from HUES8 and H1 (NIH hESC registry #10-0043; male) were used for CRISPR/Cas9 genome editing followed by directed differentiation. Undifferentiated cells were maintained as aggregates in supplemented mTeSR1 medium (StemCell Technologies) using spinner flasks (Corning) set at a 70rpm rotation rate in a 37°C 5% CO<sub>2</sub> incubator.

### Primary human samples

Primary human adult islets were obtained from Prodo laboratories. The cadaveric donors had no history of diabetes and spanned a range of ages and health parameters:

- Donor 1: 32-year-old male, 6'1," 190 lbs., with a BMI of 25.1 and HbA1c 4.8%
- Donor 2: 43-year-old female, 5'0," 115 lbs., with a BMI of 22.4 and HbA1c 5.2%
- Donor 3: 22-year-old male, 6'1," 247 lbs., with a BMI of 32.7 and HbA1c 5.9%
- Donor 4: 24-year-old male, 6', 250 lbs., with a BMI of 34.6
- Donor 5: 38-year-old female, 5'9," 233 lbs., with a BMI of 34 and HbA1c 5.0%

Donor 6: 59-year-old male, 5'6," 161 lbs, with a BMI of 25.9 and HbA1c 5.6%  
 Donor 7: 64-year-old male, 5'10," 150 lbs, with a BMI of 21.4 and HbA1c 5.1%  
 Donor 8: 23-year-old male, 5'10," 175 lbs, with a BMI of 25.1 and HbA1c 5.6%  
 Donor 9: 62-year-old female, 5'8," 106 lbs, with a BMI of 16.1 and HbA1c 5.3%  
 Donor 10: 55-year-old male, 77," 230 lbs, with a BMI of 27.1 and HbA1c 5.8%  
 Donor 11: 49-year-old male, with a BMI of 26.0 and HbA1c 5.6%  
 Donor 12: 49-year-old female, 5'10," 161 lbs, with a BMI of 23 and HbA1c 5.7%  
 Donor 13: 27-year-old female, 5'1," 147 lbs, with a BMI of 27.7 and HbA1c 5.0%  
 Donor 14: 33-year-old male, 70," 215 lbs, with a BMI of 30.8 and HbA1c 5.7%  
 Donor 15: 27-year-old female, 5'1," 147 lbs, with a BMI of 27.7 and HbA1c 5.0%  
 Donor 16: 46-year-old female, 5'5," 121 lbs, with a BMI of 20.2 and HbA1c 5.0%  
 Donor 17: 53-year-old female, 162 cm, 69.4 kgs, with a BMI of 26.3 and HbA1c 5.4%  
 Donor 18: 46-year-old female, 67," 123 lbs, with a BMI of 19.0 and HbA1c 5.8%  
 Donor 19: 61-year-old male, 66," 179 lbs, with a BMI of 28.8 and HbA1c 5.1%  
 Donor 20: 56-year-old male, 67," 88 lbs, with a BMI of 13.7 and HbA1c 5.0%

Donor anonymity was preserved, and donors 1-10 were pooled for ChIP-seq experiments; donors 11 and 12 were used for ATAC-seq and WGBS experiments, and 13-20 for circadian entrainment and RNA-seq experiments. Islets were cultured in supplemented CRML-1066 medium +10% FBS in low-attachment plates (Corning) in a 37°C 5% CO<sub>2</sub> incubator.

All methods involving human cells were approved by the Harvard University IRB and ESCRO committees.

### Rodent strains

Male immunodeficient SCID-Beige mice and Rowett Nude rats, aged 8-10 weeks, were obtained from Charles River Laboratories and were used for transplantation studies in accordance with the NIH Guide for the Care and Use of Laboratory Animals recommendations. All animals were handled according to the Harvard University Institutional Animal Care and Use Committee, approved by the Committee on the Use of Animals in Research and Teaching of Harvard University Faculty of Arts & Sciences, which is AAALAC International accredited, has a PHS Assurance on file with the NIH Office of Laboratory Animal Welfare (A3593-01), and is registered with the USDA (14-R-0128). Animals were maintained on regular chow (Prolab Isopro RMH 3000, LabDiet) and housed in groups before transplant and individually after transplant, with a 12hr light/dark cycle with continuous access to food and water.

## METHOD DETAILS

### Cell culture adaptation

CRISPR/Cas9-edited lines were adapted to grow in suspension culture using mTeSR 3D medium (StemCell Technologies) as instructed by the manufacturer with minor modifications. Briefly, cells were expanded in plates, incubated in Gentle Cell Dissociation Reagent (GCDR; StemCell Technologies) for 4min, gently scraped, and resuspended in mTeSR 3D+10μM ROCK Inhibitor Y27632 at 1x10<sup>6</sup> cells/ml. Suspended cells were fed with mTeSR 3D Feed medium 24h and 48h later, then half of their medium was replaced with fresh mTeSR 3D 72h and 96h later, followed by medium replacement with mTeSR1 120h later. Cell clusters were fed every other day thereafter and monitored for size, and upon reaching 200-300μm in diameter, were passed through a 300 μm cell strainer and passaged as clumps by incubating in GCDR at 37°C for 6min, straining through a 37 μm strainer, and resuspending at 1x10<sup>6</sup> cells/mL in mTeSR 3D+10μM Y27632. This procedure was repeated for another passage, and for the third passage clumps were directly passaged into mTeSR1 +10μM Y27632, with half their medium replaced with fresh mTeSR1 24h later. Clusters were fed every other day thereafter and further passaged at 0.6-0.8 x10<sup>6</sup> cells/ml into mTeSR1 whenever reaching 200-300μm. Suspension cultures doubling in cell numbers within 3-4 days were then considered adapted and ready for differentiation.

### Cell differentiation

Directed hPSC differentiation into islet cells was conducted as described previously (Pagliuca et al., 2014) with the modifications described in (Millman et al., 2016). Briefly, 150million cells were seeded in 300ml mTeSR1 +10μM ROCK Inhibitor Y27632, fed with mTeSR1 48h later, and 72h later stepwise differentiation stages were induced by the following treatments:

Stage 1: 24h in S1 medium +100ng/ml ActivinA +14μg/ml CHIR99021, followed by 48h in the same medium without CHIR99021.  
 Stage 2: 72h in S2 medium +50ng/ml KGF.  
 Stage 3: 24h in S3 medium +50ng/ml KGF +0.25μM Sant1 +2μM Retinoic acid (RA) +500nM PDBU +10μM Y27632 +200nM LDN193189, followed by 24h in the same medium without LDN193189.  
 Stage 4: 6 days in S3 medium +50ng/ml KGF +0.25μM Sant1 +0.1μM RA +10μM Y27632 +5ng/ml ActivinA.  
 Stage 5: 4 days in BE5 medium +0.25μM Sant1 +0.1μM RA +1μM XXI +10μM Alk5i II +1μM T3 +20ng/ml Betacellulin +10μM Y27632, followed by 3 days in BE5 +25nM RA +1μM XXI +10μM Alk5i II +1μM T3 +20ng/ml Betacellulin +10μM Y27632.  
 Stage 6: 7-30 days in supplemented CRML-1066 medium +10% defined fetal bovine serum (FBS, Hyclone) +10μM Alk5i II +1μM T3.

Factors and media are as described in (Pagliuca et al., 2014).

### Cell purification

Cell clusters sampled upon completion of each differentiation stage or primary islets were washed with PBS, incubated in PBS +50% Accutase (StemCell Technologies) for 7min, and dissociated mechanically by petting up and down in PBS +10 $\mu$ M Y27632. For fixed-cell purification, cells were cross-linked in PBS +1% paraformaldehyde for 10min, quenched with 125mM glycine for 5min, and washed with PBS. Next, cells were blocked in PBS +5% donkey serum (Jackson ImmunoResearch) +0.1% saponin (Sigma) for 20min and subsequently stained with primary antibodies in blocking buffer for 1hr and fluorophore-conjugated secondary antibodies in blocking buffer for 30min. Stained cells were washed with and resuspended in PBS +5% donkey serum. For live-cell purification, single-cell suspensions were stained in PBS + primary fluorophore-conjugated antibodies for 30min on ice, washed with and resuspended in PBS. Stained fixed/live cells were filtered through a 40 $\mu$ m nylon mesh into flow cytometry tubes (BD Falcon) and sorted using MoFlo flow cytometers (Beckman Coulter) into PBS +1% BSA (Sigma) on ice. The cell populations as defined by antibody combinations were:

DE: SOX17+ (R&D Systems; AF1924)

PP1: PDX1+ (R&D Systems; AF2419)

PP2: PDX1+ NKX6+ (DSHB; F55A12) for fixed cells, and GP2+ (MBL International; D277-5) (Ameri et al., 2017; Cogger et al., 2017) for live cells.

EN: NGN3+ (R&D Systems; AF3444) for fixed cells, and SUSD2+ (Miltenyi Biotec; 130-106-401) (Liu et al., 2014) for live cells.

SC- $\beta$ : C-peptide+ (DSHB; GN-ID4) Glucagon-like peptide-2 (GLP2)- (Santa Cruz; sc-7781) for fixed cells, and TSQ+ (Thermo Fisher Scientific; M688) DPP4- (Miltenyi Biotec; 130-093-441) for live cells.

PH: C-peptide+ GLP2+ for fixed cells, and TSQ+ DPP4+ for live cells.

Primary  $\beta$ : C-peptide+ GLP2- for fixed cells, and TSQ+ HPi2+ (Novus Biologicals; NBP1-18946AF488) TM4SF4- (Novus Biologicals; FAB7998R) (Muraro et al., 2016) for live cells.

Primary  $\alpha$ : C-peptide+ GLP2+ for fixed cells, and TSQ+ HPi2+ (Novus Biologicals; NBP1-18946AF488) TM4SF4+ (Novus Biologicals; FAB7998R) (Muraro et al., 2016) for live cells.

Live cells were defined as Propidium Iodide (PI)-negative cells. Sorting strategies for each subpopulation are summarized in Figure S1.

To sort live SC- $\beta$  from PH, we devised a strategy that combines endocrine cell enrichment by the TSQ zinc dye (Jindal et al., 1992) with detection of DPP4/CD26, which inactivates glucagon-like peptide-1 (GLP1) (Mentlein et al., 1993) and specifically labels glucagon-expressing islet cells (Blodgett et al., 2015). SC- $\beta$  and PH were enriched in the TSQ<sup>+</sup> DPP4<sup>-</sup> and TSQ<sup>+</sup> DPP4<sup>+</sup> populations with > 90% and > 85% purity, respectively, as assayed by intracellular staining.

### ChIP-seq

FACS-purified fixed cells (typically ~1million) were pelleted and flash-frozen. ChIP-seq was conducted as described in (Gifford et al., 2013) with minor modifications. Briefly, cell pellets were thawed on ice for 30min, incubated in lysis buffer (0.5% NP-40 +85mM KCl +20mM Tris-HCl pH8.0 +protease inhibitor) for 10min on ice, and nuclei were pelleted and incubated in lysis buffer (1% NP-40 +0.5% sodium deoxycholate +0.1% SDS +10mM Tris-HCl pH7.5 +protease inhibitor) for 10min on ice. Chromatin was then sheared with a Branson Sonifier (model S-450D) at 4°C and incubated with 1 $\mu$ g/million cells H3K27Ac (Active Motif; 39133) or H3K4me1 (Millipore, 17-614) antibody overnight at 4°C. Next, antibody-protein complexes were isolated by incubation with Protein A/G beads (Life Technologies; 100-02D/100-07D) for 2h at 4°C. Samples were then sequentially washed twice with low-salt buffer (0.1% SDS +1% Triton X-100 +2mM EDTA +20mM Tris-HCl pH8.1 +150mM NaCl), twice with high-salt buffer (0.1% SDS +1% Triton X-100 +2mM EDTA +20mM Tris), twice with LiCl buffer (0.25M LiCl +1% NP-40 +1% deoxycholate +1mM EDTA +10mM Tris-HCl pH8.1), twice with TE (10mM Tris-HCl pH8.0 +1mM EDTA), and finally eluted in freshly-prepared elution buffer (1% SDS +0.1M NaHCO<sub>3</sub>) at 65°C for 30min. Eluates were then treated with reverse crosslinking salt mixture (250mM Tris-HCl pH6.5 +62.5mM EDTA pH8.0 +1.25M NaCl +5mg/ml Proteinase K + 62.5ng/ $\mu$ l RNase A) overnight at 65°C. DNA was then purified using AMPure XP magnetic beads (Beckman Coulter), and sequencing libraries were generated using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs; E7103), pooled, and sequenced on a HiSeq 2500 instrument (Illumina).

### ATAC-seq

FACS-purified live cells (typically ~50,000) were pelleted and tagmentation was performed at 37°C for 30min as described in (Buenrostro et al., 2013). Briefly, cell pellets were resuspended in buffer containing Tn5 Transposase (Nextera DNA Sample Preparation kit; Illumina), incubated at 37°C for 30min, and DNA was isolated using the MinElute PCR purification kit (QIAGEN), PCR-amplified for 10 cycles, and purified using AMPure XP magnetic beads (Beckman Coulter). Double-sided AMPure cleanup to remove high-molecular-weight fragments was conducted by incubation of double-concentrated AMPure beads added at 0.55x volume to the PCR reactions, followed by cleanup with a 1x AMPure volume. Sequencing libraries were pooled and sequenced on a HiSeq 2500 instrument.



### WGBS

FACS-purified live cells (typically ~500,000) were pelleted and WGBS was performed as described in (Donaghey et al., 2018). Briefly, DNA was fragmented with a Covaris S2 sonicator for 6min, purified using the DNA Clean and Concentrator kit (Zymo Research), and bisulfite conversion was performed using the EZ DNA Methylation-Gold kit (Zymo Research). Sequencing libraries were generated using the Accel-NGS Methyl-Seq DNA library kit (Swift Biosciences), pooled, and sequenced on a HiSeq 4000 instrument.

### RNA-seq

FACS-purified live cells (typically ~500,000) or whole islet/SC-islet preparations (typically ~1million cells) were pelleted and RNA-seq was performed from total RNA depleted for rRNA as described in (Alvarez-Dominguez et al., 2017a). Directional cDNA libraries were prepared using a stranded RNA-seq library preparation kit (KAPA Biosystems), pooled, and sequenced on a HiSeq 2500 instrument. Processed single-cell RNA sequencing counts and cell-type metadata from (Veres et al., 2019) (GEO accession number GSE114412) were used to study relative and fractional expression of select TFs during extended culture of terminal-stage SC- $\beta$  (Figure S7F)

### ChIP-seq mapping

Sequencing reads were mapped to the human hg19 reference genome assembly using bowtie2 (Langmead and Salzberg, 2012) with default parameters. Peak calling was performed using MACS2 (Zhang et al., 2008) with default parameters and “-broad-qvalue 0.01” to composite broad regions of read enrichment over background with a minimum FDR q-value cutoff of 0.01, using whole-cell extract from each stage as the background control. H3K4me1 and H3K27ac peaks across all cell types were concatenated and merged whenever overlapping by BEDTools (Quinlan and Hall, 2010) to obtain a unified catalog of H3K27ac/H3K4me1 regions (Figure S1).

### ATAC-seq mapping

Sequencing reads were mapped to the human hg19 reference genome assembly using bowtie2 (Langmead and Salzberg, 2012) with default parameters. Peak calling was performed using MACS2 (Zhang et al., 2008) with default parameters, “-nomodel-shift 37-ext-size 73” to bypass building a ChIP-based shifting model and instead shift read 5' ends and extend in the 5' -> 3' direction by a half-nucleosome size (73bp), and “-broad-qvalue 0.01” to composite broad regions of read enrichment with a minimum FDR q-value cutoff of 0.01.

### WGBS mapping

Sequencing reads were mapped to the human hg19 reference genome assembly using BSMAP (Xi and Li, 2009) in bisulfite mode with default parameters. CpG methylation was called as described in (Ziller et al., 2013), excluding low-quality, duplicate, and > 10%-mis-matched reads. Only CpGs with > 5X coverage, which totaled 23-27million per cell stage, were considered for downstream analyses. Differentially methylated regions (DMRs) were determined using the DSS R package (Feng et al., 2014) with difference > 0.2,  $p < 0.05$ , minimum CpGs = 4, and merge regions if closer than 500bp. DMR methylation specificity was quantified as described in (Ziller et al., 2013), based on the Jensen-Shannon divergence from the extreme case in which a region is completely methylated in only one sample and unmethylated in all others or vice versa.

### RNA-seq mapping

Sequencing reads were mapped to the human hg19 reference genome assembly using TopHat2 (Kim et al., 2013) with default and “-min-anchor 5” parameters. Differential mRNA/lncRNA expression based on GENCODE v19 annotations (Harrow et al., 2012) was determined using HTSeq (Anders et al., 2015) and DESeq (Anders and Huber, 2010) as described (Alvarez-Dominguez et al., 2017a). Expression counts were normalized as counts per million mapped reads (CPM) and only genes with CPM > 1 were considered expressed.

Differential splicing/promoter/CDS usage was determined using Cuffdiff2 (Trapnell et al., 2013) with default parameters. For each mRNA/lncRNA differentially expressed across developmental stages, we identified the closest non-overlapping mRNA lying within 1MB using BEDTools, and calculated the correlation between their expression provided the mRNA neighbor was expressed in at least one stage. Randomly-shuffled expression of the same mRNA neighbor was used to control for fortuitous correlation.

### Coverage quantification

ChIP-seq/ATAC-seq/RNA-seq coverage across a region of interest (e.g., DMRs, enhancers) was summarized as the cumulative read pileup across the region, counted using HTSeq (Anders et al., 2015) in “-m union” mode, and normalized as counts per million mapped reads (CPM). WGBS coverage across a region of interest was summarized as the mean CpG methylation across the region; this yielded intermediate methylation levels within active chromatin regions, since hypomethylated CpGs within active chromatin regions are sparse (Ziller et al., 2013).

### Enhancer and super-enhancer identification

Enhancers were defined as H3K27ac-enriched regions identified by MACS2 that do not overlap gene promoter-proximal (TSS  $\pm$  2kb) regions. RefSeq annotations obtained from the UCSC genome browser were used to extract TSS  $\pm$  2kb regions, which were then intersected with H3K27ac broad regions using BEDTools.

Enhancer domains were defined as described in (Whyte et al., 2013), by linking enhancer sites within 12.5kb of each other and ranking the resulting domains by increasing H3K27ac enrichment. The point for which a line with slope of 1 is tangent to the curve of region H3K27ac enrichment versus region ranking was then used to distinguish super-enhancer versus typical enhancer domains. H3K27ac enrichment was calculated based on the normalized, background-subtracted H3K27ac coverage within stitched-enhancer domains. Domains were assigned to the overlapping/nearest expressed (RNA-seq CPM > 1) gene, determined using BEDTools.

A unified enhancer catalog was compiled by merging enhancer domains from all cell stages using BEDTools. ATAC, H3K4me1, H3K27ac, RNA, and WGBS coverage were quantified at each stage for each enhancer in the catalog, and only enhancers with a cumulative coverage across all stages > 0 for each of the datasets were used for downstream analyses.

### Enhancer state dynamics

Enhancer state was defined based on H3K4me1 enrichment. For each cell type, enhancer domains where statistically significant H3K4me1 enrichment over background was identified by MACS2 were considered H3K4me1-marked in that cell type. Enhancer activity was analogously defined based on H3K27ac enrichment.

Enhancer gain/loss was determined based on gain/loss of H3K4me1 enrichment supported by significant gain/loss of H3K4me1 coverage as determined by DESeq. Enhancer activation/deactivation was analogously determined based on differential H3K27ac marking.

To investigate enhancer turnover along the differentiation path, we determined, for enhancers gained/lost and activated/deactivated at each differentiation stage, the stage with the maximal change in H3K4me1 and H3K27ac levels, respectively, as shown in Figures S7D and S7E.

Enhancers activated between two differentiation stages that were H3K4me1-marked in the first stage were deemed to undergo a primed-to-active transition. To investigate turnover of enhancer priming along the differentiation path, we determined, for enhancers activated at each stage that were H3K4me1-marked in some previous stage, the stage with the maximal change in H3K4me1 levels, as shown in Figure 2A.

Alternative approaches for defining H3K4me1- and H3K27ac-marking (e.g., by using arbitrary read coverage cutoffs) altered the ratio of enhancers gained versus lost, active versus inactive, and H3K4me1-premarked versus previously unmarked within each differentiation stage, but did not change the relative trends across stages and thus did not change our conclusions.

### Clustering analyses

Correlation studies (e.g., pairwise enhancer catalog H3K4me1 correlations between all cell types and replicates) were generated in R v.3.3.2 (<http://www.Rproject.org/cor> function) with default parameters and “method=“spearman”” to use the Spearman’s rank-order implementation.

Principal component analysis was implemented in R (prcomp function) with default parameters and “scale=T” to scale the data to have unit variance.

K-means clustering analyses (e.g., enhancer coverage, pioneer TF enrichment) were conducted in R (kmeans function) with default parameters and “algorithm=“Lloyd,” nstart = 100, iter.max = 100000” to use the Lloyd-Forgy implementation (LLoyd, 1982) with 100 random starts and maximum 100000 iterations. Data was standardized prior to clustering by using z-scores to compute the number of standard deviations from the mean across conditions for each data point. To determine the appropriate number of clusters (K), we plotted the cumulative within-cluster sum of squared error (WSSE) for a sequence of cluster solutions (K = 1 to K = 30), and defined the point at which the reduction in WSSE slows dramatically as the optimal solution.

### Transplantation studies

Transplantations into immunodeficient SCID-Beige mice (Jackson Laboratory) or Rowett Nude rats (Charles River) were conducted as described in (Pagliuca et al., 2014) with minor modifications: sorted and unsorted cells were re-aggregated in CRML-1066 medium +10% FBS over 2-4 days with feeding every second day. The re-aggregated cell clusters were then resuspended in RPMI-1640 medium (Life Technologies; 11875-093) and kept on ice until loading into a catheter for cell delivery under the mouse kidney capsule.

Kidneys containing the grafts were dissected from freshly euthanized transplanted mice, fixed in PBS+4% paraformaldehyde overnight, embedded in paraffin, and sectioned at 100μm for histological analysis.

For ATAC-seq, grafts were washed in saline and dissociated mechanically by trituration with scissors in PBS +50% Accutase for 7min followed by petting up and down with a 16.5gauge needle. Single-cell suspensions were washed with and resuspended in PBS, stained with TSQ at 37°C for 10min, filtered through a 40μm nylon mesh into flow cytometry tubes (BD Falcon), and PI- TSQ+ cells were sorted using MoFlo flow cytometers (Beckman Coulter) into PBS +1% BSA (Sigma) on ice.

### Glucose-stimulated insulin secretion assays

Cell clusters sampled from cadaveric or hPSC-derived islet preparations were divided into four parts for triplicate GSIS assays and insulin content determination. Clusters were washed twice with Krebs buffer containing 2.8mM glucose and loaded into 24-well plate inserts (Millicell Cell Culture Insert; PIXP01250), followed by pre-incubation in Krebs buffer containing 2.8mM glucose for 1h to remove residual insulin. Subsequently, clusters were washed with Krebs buffer containing 2.8mM glucose and sequentially challenged with Krebs buffer containing 2.8mM, 20mM, 2.8mM, and 20mM glucose, with a 1h incubation for each concentration.

and an additional wash between 20mM and 2.8mM to remove residual glucose, followed by incubation with 2.8mM glucose + 30mM KCl for 1h. Clusters were then dissociated using TrypLE Express (Life Technologies) and cells were counted by an automated Vi-Cell (Beckman Coulter). All incubations were carried at 37°C, with supernatant samples collected at the end of each incubation.

Detection of serum human insulin following a glucose challenge in immunocompromised transplanted animals was conducted as described in (Pagliuca et al., 2014): animals were fasted for 16h overnight, and the glucose challenge was performed by intraperitoneal injection of 2g D-(+)-glucose/1 kg body weight. Serum was collected both before and 30min after injection through mandibular bleeding using a lancet (Feather; 2017-01). Serum was separated out using Microvettes (Sarstedt; 16.443.100) and stored at –80°C until ELISA analysis.

GSIS dynamics of handpicked cadaveric/SC-islets were assayed in triplicate as described in (Blum et al., 2012) on an automated Perfusion System (BioRep): chambers were first perfused with Krebs buffer containing 2.8mM glucose for 1h at a flow rate of 100ul/min and sequentially perfused with 2.8mM, 20mM, and 2.8mM glucose for 15min, 30min, and 15min, respectively, followed by perfusion with 2.8mM glucose + 30mM KCL for 15min.

Insulin levels in supernatant/serum samples containing secreted insulin were processed using a Human Ultrasensitive Insulin ELISA kit (ALPCO Diagnostics; 80-INSHUU-E10) according to the manufacturer's instructions.

### Immunohistochemistry

Immunohistochemistry of kidney graft sections was performed as described in (Pagliuca et al., 2014): sections were subjected to deparaffinization using Histoclear (Thermo Fisher Scientific; C78-2-G) and rehydrated, then emerged in 0.1M EDTA (Ambion; AM9261) for antigen retrieval, and placed in a pressure cooker (Proteogenix; 2100 Retriever) for 2h. Slides were blocked in PBS +0.1% Triton X-100 (VWR; EM-9400) +5% donkey serum (Jackson ImmunoResearch; 017-000-121) for 1h, followed by incubation in blocking solution with primary antibodies overnight at 4°C. The next day, cells were washed twice in PBS, followed by secondary antibody incubation for 2h protected from light. For imaging, secondary antibodies were washed twice with PBS, and slides were mounted in Vectashield mounting medium with DAPI (Vector Laboratories; H-1200), covered with coverslips and sealed with nail polish. Images were taken using a Zeiss LSM 510 microscope (Carl Zeiss).

### Putative pioneer factor identification

To study pioneer factors across developmental transitions, sites of newly-gained ATAC signal were identified by running MACS2 as detailed above but using a given stage as the treatment and the known/inferred previous stage as the background control:

Treatment = DE, control = hPSC  
 Treatment = PP1, control = DE  
 Treatment = PP2, control = PP1  
 Treatment = EN, control = PP2  
 Treatment = SC-β, control = EN  
 Treatment = PH, control = EN  
 Treatment = α-cell, control = PH  
 Treatment = β-cell, control = SC-β

To assess the pioneer ability of TFs highlighted by motif enrichment analysis as detailed below, we downloaded pioneer index and chromatin log odds scores from (Sherwood et al., 2014), as well as chromatin accessibility regulator ranks from (Lamparter et al., 2017). Given motif redundancy between TFs of the same evolutionarily related subfamily (Jolma et al., 2010), we assigned the same motif score to no more than one TF subfamily member (e.g., the RFX subfamily score was only assigned to RFX1).

To examine DNase dynamics at the binding sites of known pioneer TFs highlighted by our analysis, we downloaded binding peaks determined by ChIP-seq for FOXA1/FOXA2/GATA4/GATA6 in DE cells (Tsankov et al., 2015) and quantified the relative change in mean methylation between the hPSC and DE ± 2.5kb from the center of each TF peak using a fixed number of equal-sized bins.

### Core regulatory circuits

Core regulatory circuits (CRCs), defined as groups of transcription factors (TFs) encoded by genes associated with super-enhancers that bind the SE associated with their own and each other's SE to form fully interconnected autoregulatory loops, were identified using CRCmapper.py as described in (Saint-André et al., 2016) with minor modifications. Briefly, the set of expressed genes that encode TFs overlapping/proximal to SEs was first identified, then the set of SE-assigned TF genes whose set of SE constituents contains at least 3 DNA sequence motif instances for their own protein products, as predicted by FIMO (Grant et al., 2011) based on a database of TF recognition motifs (Saint-André et al., 2016), was determined, and lastly groups of such autoregulated SE-driven TFs that are similarly predicted to bind each other's SE were identified recursively.

We verified that motif-based binding predictions are effectively captured by ChIP-seq for OCT4/SOX2/NANOG/MYC in hPSC and SOX17/FOXA2/OTX2/EOMES in DE (Tsankov et al., 2015), as well as for PDX1/ONECUT1/FOXA1/FOXA2 in PP2 (Cebola et al., 2015) (Wang et al., 2015) and FOXA2/NKX2-2/NKX6-1/PDX1 in cadaveric islets (Pasquali et al., 2014). Many possible CRCs are identified for each differentiation stage (Table S5), and the union set of TFs from all possible CRCs at each stage are displayed in Figure 5B. Of note, TFs without a known recognition motif (e.g., NGN3) are accordingly missing in CRC models.

Testing different parameters for CRC identification (e.g., restricting motif analysis to ATAC-seq peaks within SEs instead of SE constituents, using a different TF motif database (Weirauch et al., 2014), changing the criteria for identifying expressed genes, and changing SE-gene association rules) led to highly similar lists that capture known master regulators of each pancreatic differentiation stage and thus did not change our conclusions.

The expression specificity of CRC TFs was scored by the fraction of the TF's cumulative expression across all stages represented in that stage (i.e., its fractional expression level). Median specificity scores ranged from 0.06 to 0.10 across differentiation stages. We selected an empirical specificity cutoff of 0.30 that optimally distinguished between known stage-specific factors and known uniformly expressed ones.

To quantify regulatory connectivity, for each SE-driven TF we defined the inward binding of other TFs to its SE as its regulatory IN degree, and the outward binding of the TF to other TFs' SEs as its regulatory OUT degree. Binding was predicted as before via TF recognition motif scanning within SE domain enhancer constituents using FIMO as detailed below, and total connectivity was calculated as the sum of the regulatory IN and OUT degrees.

### shRNA studies

Lentiviral vectors carrying independent *LMX1B*-targeting or control shRNAs (GE Healthcare Dharmacon, Inc.) were transfected into mitotic LentiX-293T (Takara Bio; 632180) cells maintained in DMEM medium (Life Technologies) using TransIT transfection reagent (Mirus). Viruses were concentrated 48h and 72h post-transfection by precipitation with PEG-it virus precipitation solution (System Biosciences) overnight at 4°C followed by centrifugation at 3000 g for 30min at 4°C, and stored at –80°C until infection.

For lentiviral infection, cell clusters sampled from suspension cultures the day of EN stage induction were dissociated by incubation in PBS +50% Accutase (StemCell Technologies) for 7min, followed by mechanical dissociation by petting up and down in PBS +10 $\mu$ M Y27632. Cells were resuspended at a density of 5–10million cells/mL in EN-stage day 1 medium (BE5 medium +0.25 $\mu$ M Sant1 +0.1 $\mu$ M Retinoic acid (RA) +1 $\mu$ M XXI +10 $\mu$ M Alk5i II +1 $\mu$ M T3 +20ng/ml Betacellulin +10 $\mu$ M Y27632), and Polybrene infection reagent (Santa Cruz) was added at 8 $\mu$ g/mL. Cells were then plated into 6-well plates pre-coated with BD Matrigel Matrix High Concentration (BD Biosciences) and viruses were added dropwise to each well. Cells were incubated at 37°C and fed daily.

For infected cell collection, the day of change to EN-stage day 5 medium cells were dissociated by incubation with TrypLE Express (GIBCO) for 7min at 37°C and quenched in BE5 medium +10 $\mu$ M Y27632. For RNA studies, single-cell suspensions were filtered through a 40 $\mu$ m nylon mesh into flow cytometry tubes (BD Falcon), and live infected (PI- GFP+) cells were sorted using MoFlo flow cytometers (Beckman Coulter) into PBS +1% BSA (Sigma) on ice. For FACS immunostaining studies, single-cell suspensions were fixed in PBS +4% paraformaldehyde for 10min, washed with PBS, and blocked in PBS +5% donkey serum (Jackson ImmunoResearch) +0.1% saponin (Sigma) for 20min. Cells were then stained in blocking buffer +NGN3 primary antibody for 1hr and in blocking buffer +fluorophore-conjugated secondary antibody for 30min, then washed with and resuspended in PBS +5% donkey serum. Stained cells were filtered through a 40 $\mu$ m nylon mesh into flow cytometry tubes (BD Falcon), and infected (GFP+) cells were analyzed for NGN3 levels using MoFlo flow cytometers (Beckman Coulter).

### NanoString and qPCR studies

Total RNA from sorted cells was isolated using a miRNeasy Mini Kit (QIAGEN), and RNA was stored at –80°C until downstream studies. For Nanostring profiling, 100–300ng RNA was hybridized to a custom nCounter XT probe set, processed and imaged using the Nanostring prep station and nCounter (NanoString Technologies), and analyzed using the Nanostring nSolver software with default parameters and with the geometric mean expression of five housekeeping genes (RPL15, RPL19, UBE2D3, ITCH, and TCEB1) as an internal normalization control. Only samples with enough readout complexity (normal distribution of expression estimates) were quantified.

For real-time quantitative PCR, cDNA was synthesized with a SuperScriptIII first-strand synthesis kit (Life Technologies) using random hexamers (Thermo Fisher Scientific). PowerUp SYBR Green-based real-time PCR (Life Technologies) was performed in a 7900HT Fast Real-time PCR System (Applied Biosystems) with 18S rRNA as an internal normalization control.

### CRISPR/Cas9-mediated genome editing

To disrupt the *LMX1B* locus we used inducible CRISPR/Cas9 hPSC platforms (González et al., 2014; Shi et al., 2017), wherein doxycycline-inducible Cas9 expression from the AAVS1/PPP1R12C locus has been engineered. Single chimeric guide RNAs (gRNAs) were designed using the GPP sgRNA Designer (Doench et al., 2016) and generated by *in vitro* transcription using a MEGAshortscript T7 Transcription kit (Thermo Fisher Scientific) from a PCR-amplified 120nt template comprising a 20nt T7 promoter sequence, the variable 20nt sgRNA recognition sequence, and the invariable chimeric 80nt sgRNA sequence (Soh and Huangfu, 2017). The resulting gRNAs were purified using a MEGAclear Transcription Clean-Up kit (Thermo Fisher Scientific), diluted to 200–800 ng/ $\mu$ l, and stored at –80°C until further use.

gRNA transfection was performed as described in (Soh and Huangfu, 2017). Briefly, cells were pre-treated with 2  $\mu$ g/ml +10 $\mu$ M ROCK Inhibitor Y27632 for 24h, re-plated onto 24-well plates, and transfected with 300ng sgRNA using Lipofectamine RNAiMAX (Life Technologies). 3–4 days following transfection, 1–2x10<sup>6</sup> cells were re-plated onto 10cm plates to allow single-cell colonies to grow, and genomic DNA was extracted from the rest using a Quick-DNA miniprep Kit (Zymo Research Corp.), PCR-amplified with primers flanking the gRNA target sites, and used for T7 Endonuclease I assays (Mashal et al., 1995) to verify and quantify insertion/deletion (indel) efficiency (IE). Then, either 48 (for IE > 20%) or 96 (for 5% < IE < 20%) ~2mm single-cell colonies were



picked, grown in 48/96-well plates, and assessed for DNA edits by Sanger sequencing, and those with verified homozygous indels were expanded in culture for further use.

### **In vitro circadian entrainment**

Cadaveric islets (typically  $5 \times 10^3$  IEQ per condition) and SC-islets (typically  $5 \times 10^6$  cells per condition) cultured in ABLE 30ml disposable bioreactors (Reprocell; ABBVS03A) underwent the following treatments (individually or in combination) for 1–4 days: 10uM Forskolin (Stemgent; 04-0025) for 1h followed by 23h recovery; 20mM of Glucose (Sigma; G7528) for 12h followed by 12h recovery; 37.5mg/L Insulin (Sigma; I9278) for 12h followed by 12h recovery; 210mg/L Arginine (Sigma; A6969) for 12h followed by 12h recovery. Stringent washes were implemented before change to media without factors to remove residual factors. The same wash/incubation times were used for mock-treated cultures. At the end of shock/ recovery cycles, cultures were kept in constant media and samples were taken every 4h for RNA profiling and insulin secretion/calcium influx assays. Supernatant samples were also taken and profiled for dissolved gases using a Stat Profile Prime instrument (Nova Biomedical).

### **Calcium influx assays**

Calcium influx dynamics of cadaveric or hPSC-derived islet preparations in response to a glucose challenge were assayed as described in (Kenty and Melton, 2015; Pagliuca et al., 2014): cell cluster samples embedded in hESC-qualified Matrigel (VWR; 47743-722) were stained with Fluo4-AM (Life Technologies; F-14217) for 45min, incubated for 15min without the dye, and imaged on an AxioZoom V16 microscope (Carl Zeiss) during sequential static incubations with Krebs buffer containing 2.8mM, 20mM, and 2.8mM glucose for 5min, 10min, and 5min, respectively, followed by incubation with 2.8mM glucose + 30mM KCL for 5min. Images were acquired every 17 s throughout incubations, and the mean fluorescence intensity of each cluster across the time series was quantified using ImageJ (Schneider et al., 2012).

### **RNA rhythmicity and differential expression**

RNA-seq data from circadian-entrained cadaveric/SC-islet cultures were normalized across libraries using the DESeq2 scaling function to estimate library size factors. Only genes with CPM > 1 were across all samples were retained for downstream analyses. Rhythmicity was evaluated using harmonic regression by fitting a truncated Fourier series to z-score standardized gene expression time courses with the HarmonicRegression R library (Lück et al., 2014), which uses a Gaussian error assumption to calculate a p value for each gene.

An empirical alternative for establishing statistical significance, implemented by comparing rhythmicity F-statistics under the Gaussian error assumption for the original versus 100 randomly-shuffled expression profiles for each gene, led to comparable/greater number of cycling genes enriched for the same ontology terms and thus did not change our conclusions.

We compared gene expression distributions (regardless of rhythmicity) between mock- and circadian-entrained samples using a two-sided t test to determine confidence on the measured difference in their means, and considered genes with  $p < 0.05$  in at least one sample comparison to be differentially expressed.

### **Region-based annotations**

Genomic regions of interest (e.g., DMRs) were analyzed for associated gene ontology (Ashburner et al., 2000), and pathway annotations from the Molecular signatures database (Liberzon et al., 2011) with GREAT (McLean et al., 2010) using default settings. Only annotations meeting a > 2 region-based fold enrichment and an FDR q-value < 0.05 by both a binomial test over genomic regions and a hypergeometric test over associated genes were considered significantly enriched.

To study annotations associated with enhancers gained across developmental transitions, we first identified *de novo* H3K27ac peaks by running MACS2 as detailed for putative pioneer factor identification above, i.e., using a given stage as the treatment and the known/inferred previous stage as the background control.

To investigate whether a defined set of regions (e.g., H3K4me1-premarked enhancers) shows statistically significant, concordant differences between two states of interest (e.g., H3K27ac levels in  $\alpha$  versus  $\beta$  cells) we used GSEA (Subramanian et al., 2005) with default parameters and “-metric log2\_Ratio\_of\_Classes.”

To assess the genomic distribution of regions of interest (e.g., H3K4me1/H3K27ac peaks, DMRs), a custom script was used to overlap their coordinates with a database of genic/intergenic regions (Alvarez-Dominguez et al., 2017a)

### **Gene set and pathway enrichment analysis**

Gene lists ranked by a feature of interest (e.g., expression change,  $-\log_{10}$  p value of rhythmicity) were analyzed for enrichment of genes grouped by biological process ontology or by curated annotations from the Molecular signatures database with GSEA using default parameters and “-metric log2\_Ratio\_of\_Classes.” Pathway annotations from the Kyoto Encyclopedia of Genes and Genomes and the Reactome database were extracted to investigate overrepresentation of pre-defined pathways.

### **Motif enrichment analysis**

Genomic sequences from regions of interest (e.g., DMRs, *de novo* ATAC peaks) were searched for matches to a database of TF recognition sites (Saint-André et al., 2016) for TFs expressed in the relevant cell type using FIMO (Grant et al., 2011) as described in (Alvarez-Dominguez et al., 2017b) with minor modifications: a Markov model of sequence nucleotide composition was used as

the background model for motif matching (to normalize for biased distribution of individual letters in the examined sequences), and motifs with an odds ratio  $> 2$  and q-value  $< 0.05$  (Fisher's exact test) relative to 10 randomly-shuffled controls were considered significantly enriched.

### Additional bioinformatics methods

All sequencing reads were quality-checked with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Genome-wide read density maps were generated by MACS2 using the “-bdg” option, normalized by RSeQC (Wang et al., 2012a) using the “normalize\_bigwig.py” function, and visualized using BEDTools and the UCSC genome browser. Signal coverage and signal change surrounding regions of interest (e.g., DMRs, enhancer sites) were visualized using the ngs.plot R package (Shen et al., 2014). Data heatmaps were generated using the heatmap.2 function of the gplots R package (<http://cran.r-project.org/web/packages/gplots/index.html>). Network diagrams were generated using Gephi (Bastian et al., 2009).

### QUANTIFICATION AND STATISTICAL ANALYSIS

No statistical methods were used to predetermine sample size or remove outliers. The statistical difference between two sets of paired count data (e.g., motif matches in test versus randomly-shuffled sequences) was assessed by a Fisher's exact test using the fisher.test R implementation with default parameters. For unpaired data, a Shapiro-Wilk normality test was first performed using the shapiro.test R implementation with default parameters; for normally distributed data (e.g., LMX1B levels in test versus control treatments) we then used a two-sided t test (t.test R implementation with default parameters) to assess confidence on the measured difference of their mean values. For unpaired data that don't follow a normal distribution, we used a non-parametric Wilcoxon rank sum test to determine if they belong to the same distribution.

### DATA AND CODE AVAILABILITY

The ChIP-seq, ATAC-seq, WGBS, and RNA-seq data generated during this study are available at NCBI GEO accession number: GSE139817.

### ADDITIONAL RESOURCES

Enhancer annotations and genome-wide visualizations of sequencing datasets are available at [http://meltonlab.rc.fas.harvard.edu/data/pancreatic\\_enhancers/](http://meltonlab.rc.fas.harvard.edu/data/pancreatic_enhancers/)