HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2019 July 17.

Published in final edited form as:

Cell Rep. 2019 July 09; 28(2): 382–393.e7. doi:10.1016/j.celrep.2019.06.034.

FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation

Kihyun Lee[1,2,5,6], **Hyunwoo Cho**[2,3,5], **Robert W. Rickert**[1], **Qing V. Li**[1,4], **Julian Pulecio**[1], **Christina S. Leslie**[3,*], and **Danwei Huangfu**[1,7,*]

[1]Developmental Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

[2]Weill Graduate School of Medical Sciences at Cornell University, New York, NY 10065, USA

[3]Computational Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

[4]Louis V. Gerstner Jr. Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

[5]These authors contributed equally

[6]Present address: Gladstone Institutes, San Francisco, CA 94158, USA

[7]Lead Contact

SUMMARY

Transcriptional regulatory mechanisms of lineage priming in embryonic development are largely uncharacterized because of the difficulty of isolating transient progenitor populations. Directed differentiation of human pluripotent stem cells (hPSCs) combined with gene editing provides a powerful system to define precise temporal gene requirements for progressive chromatin changes during cell fate transitions. Here, we map the dynamic chromatin landscape associated with sequential stages of pancreatic differentiation from hPSCs. Our analysis of chromatin accessibility dynamics led us to uncover a requirement for *FOXA2*, known as a pioneer factor, in human pancreas specification not previously shown from mouse knockout studies. *FOXA2* knockout hPSCs formed reduced numbers of pancreatic progenitors accompanied by impaired recruitment of GATA6 to pancreatic enhancers. Furthermore, FOXA2 is required for proper chromatin remodeling and H3K4me1 deposition during enhancer priming. This work highlights the power of

*Correspondence: cleslie@cbio.mskcc.org (C.S.L.), huangfud@mskcc.org (D.H.).

AUTHOR CONTRIBUTIONS

K.L. conceived the initial ideas and performed most experiments. K.L. and D.H. wrote the manuscript. H.C. carried out the bioinformatic analyses of ATAC-seq, RNA-seq, and ChIP-seq data. R.W.R., J.P., and Q.V.L. performed other necessary experiments. D.H. and C.S.L. supervised the work. All authors provided intellectual input and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

SUPPLEMENTAL INFORMATION

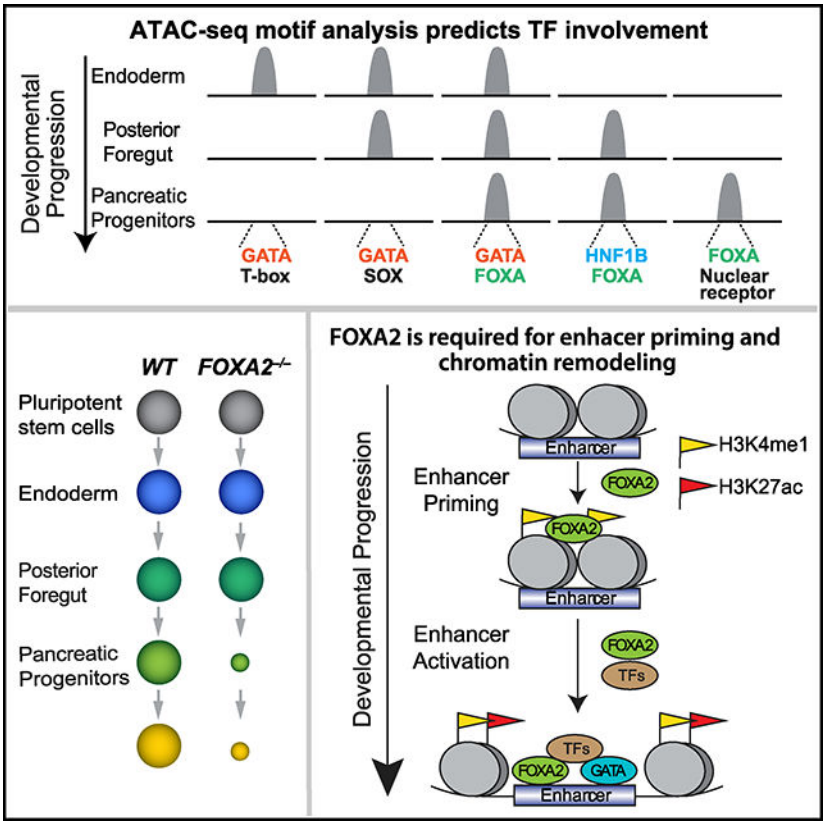
Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.06.034.

combining hPSC differentiation, genome editing, and computational genomics for discovering transcriptional mechanisms during development.

In Brief

Lee et al. use ATAC-seq to identify key transcriptional factors involved in human pancreatic differentiation. *FOXA2* knockout human pluripotent stem cells showed impaired differentiation to pancreatic progenitors, a phenotype not observed in *Foxa2* conditional knockout mice. Furthermore, *FOXA2* is required for proper chromatin remodeling and H3K4me1 deposition during enhancer priming.

Graphical Abstract



INTRODUCTION

Precise spatial and temporal control of gene expression during development requires coordinated binding of transcription factors (TFs) to enhancers (Calo and Wysocka, 2013). The prevailing model suggests that enhancers are activated sequentially with correlated histone modifications, most prominently histone H3 lysine 4 monomethylation (H3K4me1) and histone H3 lysine 27 acetylation (H3K27ac) (Heintzman et al., 2007; Zentner et al., 2011). During development, lineage-specific enhancers are thought to transition through a primed state before activation. Both states are typically marked by H3K4me1, while the active state is specifically enriched in H3K27ac (Creyghton et al., 2010; Rada-Iglesias et al.,

2011). Lineage-specific TFs have been shown to activate enhancers for cell type-specific induction of gene expression (Spitz and Furlong, 2012), but a specific requirement of TFs for H3K4me1 deposition during lineage priming has not been demonstrated. Thus, the exact regulation of the successive priming and activation of enhancers remains elusive.

Stepwise differentiation of human pluripotent stem cells (hPSCs) offers a powerful approach to capture consecutive molecular events during lineage transition. Several groups, including our own, have used hPSC-directed differentiation and gene editing to study how TFs regulate human pancreatic differentiation (Amin et al., 2018; Gage et al., 2015; McGrath et al., 2015; Shi et al., 2017; Tiyaboonchai et al., 2017; Zhu et al., 2016). For instance, we uncovered a haploinsufficient requirement for *GATA6* in human pancreatic differentiation consistent with the diabetes symptoms found in patients (Shi et al., 2017). We have also applied genome-wide CRISPR/Cas9-mediated loss-of-function screening in hPSCs and uncovered a previously unrecognized role of the Jun N-terminal kinase (JNK)-JUN family genes as key barriers of hPSC differentiation to definitive endoderm (DE), the first step toward pancreatic lineage specification (Li et al., 2019). Here, we use open chromatin profiling with Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) followed by motif analysis to identify TFs involved in developmental lineage priming (Buenrostro et al., 2013; Liu et al., 2016; Tsompana and Buck, 2014). We identified prominent enrichment of FOXA binding sites at open chromatin regions associated with the initiation of pancreatic differentiation, suggesting an inductive role for one or multiple FOXA family members: FOXA1, FOXA2, and FOXA3 (Hannenhalli and Kaestner, 2009). FOXA1 and FOXA2 are thought to act as pioneer factors that can bind to nucleosomal DNA and facilitate enhancer activation (Iwafuchi-Doi and Zaret, 2014).Previous studies have shown the recruitment of FOXA1 and FOXA2 to primed enhancers marked with H3K4me1, but *FOXA1* knockdown did not affect H3K4me1 levels (Lupien et al., 2008; Wang et al., 2015). The role of *FOXA½* has not been determined during human pancreas development because of a lack of knockout (KO) studies. Here, we have generated individual *FOXA1* and *FOXA2* KO hPSC lines using the CRISPR/Cas9 system and report that *FOXA2* is required for proper differentiation of hPSCs into pancreatic progenitors. We show that FOXA2 is required for proper chromatin remodeling, H3K4me1 deposition before enhancer activation, and recruitment of GATA6 to these enhancers. We conclude that human pancreatic development requires coordinated enhancer priming and activation, which are mediated at least partly by FOXA2, providing a mechanism for how pioneer factors can influence cell type-specific sequential chromatin modifications to establish lineage-specific transcriptional programs.

RESULTS

Distinct TF Motif Patterns during Successive Stage Transitions

Using a well-established differentiation protocol (Pagliuca et al., 2014; Rezania et al., 2014; Shi et al., 2017; Zhu et al., 2016), hPSCs can be differentiated into relatively homogeneous populations of DE, posterior foregut (FG), and primary pancreatic progenitor (PP1) cells (Figure 1A). This system provides a window into the sequential differentiation events leading to the emergence of the pancreatic primordium in humans. ATAC-seq analysis of

HUES8 human embryonic stem cells (hESCs) at the embryonic stem (ES), DE, FG, and PP1 stages revealed dynamic changes in the chromatin landscape during pancreatic differentiation (Figure 1B). To predict how TF occupancy changes during pancreatic differentiation, we first identified 125 TF motifs from the Catalog of Inferred Sequence Binding Preferences (CIS-BP) database (Weirauch et al., 2014) that correspond to TFs that are expressed during pancreatic differentiation based on RNA sequencing (RNA-seq) data (≥ 100 tags per million [TPM] for at least one of the four stages). We then used the Kolmogorov-Smirnov (KS) test to examine the association of TF motifs with opened or closed ATAC-seq peaks during successive lineage transitions in comparison to the total atlas, which was a collection of reproducible ATAC-seq peaks from all stages (STAR Methods; Table S1). These analyses revealed several distinct temporal patterns of TF motif association across pancreatic differentiation (Figure 1C). GATA motifs were associated with opened sites during the ES-to-DE stage transition but associated with closed sites during the DE-to-FG and FG-to-PP1 transitions, the HNF1B motif was associated with opened sites at the ES-to-DE and DE-to-FG stage transitions, and FOXA motifs were associated with opened sites during all transitions but most prominently during the FG-to-PP1 transition (Figure 1C). Changes in the proportion of opened ATAC-seq peaks containing GATA6, HNF1B, or FOXA2 motifs during the successive transitions suggested corresponding temporal requirements for these TFs during hPSC differentiation to the DE, FG, and PP1 stages (Figure 1D). In support of this model, we and others have found that *GATA6* is required for efficient human DE differentiation (Chia et al., 2019; Fisher et al., 2017; Shi et al., 2017; Tiyaboonchai et al., 2017). Similar motif enrichments were confirmed by analysis with the Hypergeometric Optimization of Motif Enrichment (HOMER) algorithm (Heinz et al., 2010) (Figure 1E).

Association of FOXA Motifs with Pancreatic Specification

To better identify TFs that drive the induction of the pancreatic fate, we focused on ATAC-seq peaks accessible at the PP1 stage. Some sites gained chromatin accessibility specifically at the PP1 stage, whereas others were opened at the earlier DE or FG stage and remained open until the PP1 stage (Figure 2A). Thus, we grouped ATAC-seq peaks based on their patterns of accessibility across stages (Figure 2A; Table S2). These stage-specific groups had similar proportions of introns, promoters, exons, and intergenic regions (Figure S1A). This categorization showed that ~30% of the PP1-accessible peaks were acquired at the PP1 stage (the PP1-specific group), and around half of the PP1-accessible peaks first became accessible at the DE stage (the DE-PP1 group) (Figure 2B).

Next, we investigated the association of TF motifs and stage-specific groups using the KS test and the HOMER algorithm (Figure 2C; Figures S1B and S1C; Table S3). More than 50% of sites that gained accessibility specifically at the PP1 stage contained FOXA motifs (odds ratio > 2.03) (Figure 2C; Figure S1B). We confirmed that all DE-accessible groups (DE specific, DE-FG, and DE-PP1) had enriched GATA motifs (Figures S1B and S1C). We used the HOMER algorithm to identify which TFs may distinguish the DE-PP1 group from the other two DE-accessible groups (Figure 2D). FOXA motifs were strongly associated with the DE-PP1 group, which gained accessibility at the DE stage and remained accessible through the FG and PP1 stages (Figure 2D). By performing chromatin immunoprecipitation

sequencing (ChIP-seq) analyses, we found that the binding of FOXA2, GATA6, GATA4, and HNF1B corresponded well with stage-specific ATAC-seq signals and motif association (Figure 2A). PDX1, a key pancreatic lineage TF (Jonsson et al., 1994; Stoffers et al., 1997), also showed significantly enriched binding in all PP1-accessible groups (DE-PP1, FG-PP1, and PP1 specific). When examining the PP1-specific ATAC-seq group, which gained chromatin accessibility specifically at the PP1 stage, we observed FOXA2 binding at the DE and FG stages before the PP1 stage (Figure 2A). These results raised the possibility that FOXA2 participates in the initial opening of chromatin and enhancer priming and facilitates the activation of pancreatic enhancers through potential cooperation with GATA6, HNF1B, and PDX1. Altogether, analyses of chromatin accessibility changes in successive developmental stages revealed stage-specific involvement of TFs during lineage specification (Figure 2E).

Requirements of *FOXA2* in Pancreatic Progenitor Specification

Deleting individual *Foxa* genes in mice does not affect the induction of the pancreatic program, but deleting both *Foxa1* and *Foxa2* results in pancreatic hypoplasia (Gao et al., 2008; Kaestner et al., 1998, 1999; Lee et al., 2005). We have previously observed that the human pancreatic differentiation system is more sensitive to the reduced gene dosage of GATA6 (Shi et al., 2017), so we speculated that the human system could also be more sensitive to the loss of *FOXA1* or *FOXA2*. Supporting a potential role of *FOXA½*, both genes were upregulated during DE and pancreatic differentiation (Figures S2A and S2B). We then used the iCRISPR platform (González et al., 2014) to generate clonal KO lines with either *FOXA1* or *FOXA2* deleted from the H1 hESC background (Figure S2C; Table S4), and we confirmed the loss of FOXA1 or FOXA2 proteins by western blotting (Figure S2D). Pancreatic-directed differentiation showed that *FOXA1* KO hESCs behave like wild-type (WT) cells, whereas *FOXA2* KO hESCs formed significantly reduced numbers of $PDX1^+$ pancreatic progenitors (Figure S2E).

To confirm that the requirement of *FOXA2* is independent of the genetic background, we generated additional *FOXA2* KO lines in the HUES8 hESC background (Figure 3A; Figure S2C; Table S4). Focusing on three lines with biallelic frameshift mutations, we confirmed the loss of FOXA2 protein by western blotting (Figure 3B). The effects of *FOXA2* deletion were determined by stepwise differentiation of *FOXA2* KO and isogenic WT control hPSCs, followed by flow cytometric analysis for expression of lineage-specific TFs (Figure 3C). *FOXA2* KO hPSCs formed $SOX17^+$ DE cells and $HNF1B^+$ FG cells with efficiencies comparable to those of WT hPSCs (Figure 3C; Figure S2F). However, at the PP1 stage, significantly fewer $PDX1^+$ pancreatic progenitors were formed from *FOXA2* KO hPSCs. As a consequence, the subsequent transition to the mature $PDX1^+$ $NKX6–1^+$ pancreatic progenitor (PP2) stage was also affected (Figures 3C–3E). The expression levels of *FOXA1* and *FOXA3* were decreased in *FOXA2* KO cells at the DE and FG stages; some increase in *FOXA1* expression was observed at the PP1 stage, which could have compensatory effects (Figure S2G).

To characterize global transcriptomic changes, we performed RNA-seq on HUES8 WT and *FOXA2* KO cells at all differentiation stages. We used a two-factor generalized linear model

(GLM) to identify genotype-dependent differentially expressed genes while controlling for the effect of differentiation stages. Here, 890 genes showed significantly different expression in at least one stage-matched comparison between WT and *FOXA2* KO cells (Table S5). Principal component analysis (PCA) based on the 890 differentially expressed genes showed that *FOXA2* KO cells began to differ substantially from the WT cells at the PP1 stage (Figure 3F). Unsupervised hierarchical clustering showed that *FOXA2* KO cells at the PP1 stage were grouped more closely with FG samples than with WT PP1 samples (Figure 3G). These results indicate that *FOXA2* deletion impedes the transition of the transcriptional program from the FG to the PP1 stage during pancreatic differentiation.

Requirements of *FOXA2* for Chromatin Remodeling during Pancreatic Differentiation

We examined the effects of *FOXA2* deletion on chromatin accessibility dynamics during pancreatic differentiation. The lineage transition from the ES to the PP1 stage was accompanied by global changes in chromatin accessibility, with 19,363 and 21,925 loci gaining and losing accessibility, respectively. Significant divergence was observed with *FOXA2* KO cells from the WT cells starting from the FG stage (Figure 4A). Examination of the top 3,000 variable peaks between *FOXA2* KO and WT cells revealed a similar divergence starting at the FG stage in a PCA plot (Figure S3A). Stage-specific comparison of chromatin accessibility in *FOXA2* KO versus WT cells revealed an overall decrease in the chromatin accessibility at the DE, FG, and PP1 stages in *FOXA2* KO cells, with many sites becoming less accessible and only a small number of sites gaining accessibility (Figure S3B; Table S6). The strongest *FOXA2* KO effect was observed at the FG stage that preceded the PP1 stage that showed overt differentiation and transcriptomic changes. The less accessible peaks in *FOXA2* KO were primarily located at intronic or intergenic loci in the genome, suggesting that the loss of *FOXA2* predominantly affected enhancers (Figure S3C). Regions with decreased accessibility in *FOXA2* KO cells showed strong FOXA2 binding in WT cells (Figure 4B), supporting a direct role for FOXA2 binding in establishing chromatin accessibility at these sites.

We have previously demonstrated a dosage-dependent requirement for *GATA6* in pancreatic progenitor differentiation (Shi et al., 2017). Because we observed a significant overlap between the genomic occupancy of GATA6 and that of FOXA2 (Figures 2A and 4C; Figure S3D), we examined the effect of *FOXA2* deletion on GATA6 binding, which could account for the pancreatic phenotype observed in *FOXA2* KO cells. *FOXA2* KO cells showed no significant change in *GATA6* expression at the DE and FG stages (Figure S3E), but they showed significantly reduced GATA6 binding in sites with decreased ATAC-seq signals (Figures 4B and 4D). For instance, PP1-specific ATAC-seq peaks upstream of the *PDX1* locus were bound by FOXA2 and GATA6 in WT cells at the DE, FG, and PP1 stages; however, GATA6 binding was significantly reduced in *FOXA2* KO cells at all three stages (Figure S4A). Similarly, FOXA2 and GATA6 co-occupancy was observed at the *SOX9* locus, in which chromatin accessibility and GATA6 binding were both decreased in *FOXA2* KO cells (Figure S4B). Therefore, in the absence of FOXA2 binding, chromatin accessibility is not appropriately established for pancreatic lineage specification, and GATA6 (and potentially other lineage-specific TFs) cannot bind properly to promote the PP1 transcriptional program.

We examined stage-specific ATAC-seq groups defined in WT cells (Figure 2A) to investigate the effect of *FOXA2* KO on chromatin architecture, using ATAC-seq data to infer nucleosome positions in open chromatin regions during successive fate transitions (Buenrostro et al., 2013). We first confirmed that the *FOXA2* deletion had the most significant effect at the FG stage (Figure 5A; Figures S5A and S5B). We next focused on the PP1-accessible groups (DE-PP1, FG-PP1, and PP1 specific) and computed a 2-dimensional visualization of ATAC-seq data using insert size density plot (V plot). This allowed us to distinguish nucleosome-free regions (NFRs) from mono-nucleosome- and dinucleosome-associated regions (Figure 5B). In WT cells, the nucleosome fingerprint pattern emerged in the V plots centered on ATAC-seq peaks in a differentiation stage-specific manner (Figure 5B). In *FOXA2* KO cells, signals from both NFRs and nucleosome-associated regions (NARs) were diminished at the FG stage for all three groups (Figure 5B). At the PP1 stage, the effect of *FOXA2* deletion was more evident in NARs (Figure 5B). We aggregated ATAC-seq inserts based on the distance to the nearest FOXA2 ChIP-seq peaks in WT cells and confirmed similar effects of *FOXA2* deletion on NFRs and NARs during pancreatic differentiation (Figure S5C). The same KO effects can be visualized in tornado plots showing NFR (<150 bp) and NAR (≥150 bp) ATAC-seq fragments separately (Figure 5C). Therefore, FOXA2 is required for proper remodeling of the chromatin architecture during pancreatic lineage specification.

FOXA2 Is Required for the Establishment of the Primed Enhancer State

Previous studies showed that FOXA1 binds to primed enhancers marked by H3K4me1 before the acquisition of the H3K27ac activation mark (Wang et al., 2015). However, it is not known whether FOXA factors (or other TFs) are required for H3K4me1 deposition before enhancer activation. We performed H3K27ac and H3K4me1 ChIP-seq analyses in WT cells and found that H3K27ac and ATAC-seq signals correlated well in all stage-specific groups (Figures 2A and 6A). The PP1-specific group showed H3K4me1 signals at the FG and PP1 stages, but ATAC-seq and H3K27ac signals only became apparent at the PP1 stage, demonstrating successive enhancer priming and activation (Figures 2A and 6A). *FOXA2* KO cells showed reduced H3K27ac and H3K4me1 signals in all stage-specific groups (Figure 6A; Figure S6A). Most notably, in the PP1-specific group, the deposition of H3K4me1 was affected in KO cells at the DE and FG stages before overt enhancer activation in WT cells at the PP1 stage.

To specifically investigate enhancer priming, we focused on the PP1-specific group and identified primed PP1 peaks by excluding peaks that showed H3K27ac signals at the DE and FG stages (Figure 6B; Table S7). We observed FOXA2 binding and H3K4me1 signals at the DE and FG stages, which preceded the acquisition of H3K27ac active enhancer marks at the PP1 stage, suggesting a role of FOXA2 for enhancer priming (Figure 6B). *FOXA2* deletion caused a decrease in H3K4me1 levels at the DE and FG stages, before it affected both H3K4me1 and H3K27ac signals at the PP1 stage (Figure 6B). Furthermore, genes associated with primed PP1 peaks showed significantly increased expression levels at the PP1 and PP2 stages compared with the DE and FG stages, and the expression levels at the PP1 and PP2 stages were significantly reduced in *FOXA2* KO cells (Figure 6C). These findings support a

specific requirement of FOXA2 for enhancer priming and subsequent transcriptional activation.

To investigate the direct effects of FOXA2 on H3K4me1 deposition, we isolated FOXA2-bound primed PP1 peaks (DE, n = 494; FG, n = 593; PP1, n = 993) and compared the H3K4me1 signal in WT and *FOXA2* KO cells (Figure 7A). A significant reduction of H3K4me1 signals was observed during the priming stages (DE and FG) and activation stages (PP1) in FOXA2-bound primed PP1 peaks (Figure 7A). Furthermore, regions with greater FOXA2 binding in WT cells correlated significantly with decreased H3K4me1 levels after *FOXA2* deletion (Figure 7B). For example, at the *PDX1* locus, FOXA2 binding and primed H3K4me1 peaks were observed in WT cells at the DE and FG stages, and H3K4me1 levels were significantly decreased in *FOXA2* KO cells (Figure S6B). Altogether, our findings support a model in which FOXA2 is directly required for enhancer priming, which then reconfigures the chromatin for the recruitment of additional TFs such as GATA6 to activate the transcriptional program necessary for human pancreatic differentiation (Figure 7C).

DISCUSSION

Modeling human development through hPSC-directed differentiation offers an approach for studying disease phenotypes that are not evident in animal models (Tabar and Studer, 2014; Zhu and Huangfu, 2013). In our previous work, we discovered haploinsufficient requirements for *GATA6* and *GATA4* in human pancreatic progenitor differentiation (Shi et al., 2017), which is consistent with diabetic phenotypes of human patients (Allen et al., 2011; Shaw-Smith et al., 2014). However, this human disease phenotype is not well recapitulated in murine models. Instead, simultaneous deletion of both *Gata6* and *Gata4* is needed in mice to reproduce the pancreatic defects in human patients (Carrasco et al., 2012; Xuan et al., 2012). Similar to the *Gata* factors, mouse conditional KO of *Foxa2* in endoderm (using *Foxa3-Cre*) (Lee et al., 2005) or pancreatic progenitors (using *Pdx1-Cre*) (Gao et al., 2008) does not affect the induction of *Pdx1* expression or the onset of pancreatic development. Pancreatic hypoplasia is observed only when both *Foxa1* and *Foxa2* are deleted from mice (Gao et al., 2008; Lee et al., 2005). Thus, the *FOXA* factors exhibit species-specific gene dosage requirements similar to the *GATA* factors. *Gata4/6* was reported to regulate pancreatic differentiation through inhibition of Sonic Hedgehog (Shh) signaling in mice (Xuan and Sussel, 2016). We also observed upregulation of *GLI1*, *GLI2*, and *GLI3* expression levels in *FOXA2* KO cells at the pancreatic progenitor stages (Figure S3E), suggesting that FOXA2 negatively regulates Shh signaling during pancreatic differentiation. FOXA and GATA factors likely cooperate during pancreatic differentiation based on the significant FOXA2 and GATA6 co-occupancy and the reduction of GATA6 binding upon *FOXA2* deletion.

Our analyses of the chromatin accessibility landscape led us to uncover a requirement of *FOXA2* in human pancreas specification. It is likely that FOXA2, through its involvement in enhancer priming, also prepares endoderm cells for differentiation to additional endoderm-derived lineages, as shown for the hepatic lineage in an independent study (Genga et al., 2019). This requirement is consistent with FOXA2 playing important roles in chromatin

remodeling and enhancer activation through the recruitment of lineage-specific TFs (Iwafuchi-Doi and Zaret, 2014). Foxa2 can bind nucleosome-bound DNA (Cirillo et al., 1998, 2002; Sekiya et al., 2009) and displace nucleosomes during mouse ESC differentiation toward endoderm (Li et al., 2012). However, it is unknown what role, if any, FOXA2 plays in the successive priming of lineage-specific enhancers during differentiation. A previous study found FOXA½ enrichment at H3K4me1+ primed enhancers during pancreatic differentiation but did not establish a requirement for *FOXA1* or *FOXA2* in H3K4me1 deposition or pancreatic lineage specification (Wang et al., 2015). *FOXA1* knockdown did not affect H3K4me1 levels, but it is unclear whether this is because of an incomplete knockdown (Wang et al., 2015). Here, we show by genetic KO experiments that *FOXA2*, but not *FOXA1*, is the primary FOXA factor required for specifying sufficient numbers of pancreatic progenitor cells. It is possible that *FOXA1* and *FOXA2* have overlapping functions that could be revealed by deleting both *FOXA1* and *FOXA2* in hPSCs.

FOXA2 is required for enhancer priming through H3K4me1 deposition during pancreatic differentiation. This requirement has not been previously shown for TFs in developmental contexts. Previous work on potential roles of TFs in enhancer priming has been conducted in the contexts of overexpression studies during reprogramming or studies of immortalized cancer cell lines. Overexpression of *FOXA2* in human fibroblasts increases H3K4me1 levels (Donaghey et al., 2018), and overexpression of another TF, *Pax7*, in mouse cancer cells facilitates H3K4me1 deposition by recruiting mixed lineage leukemia (MLL) complex Ash2l histone methyltransferase (Mayran et al., 2018). *FOXA1* knockdown in human breast cancer cells has also been shown to affect the recruitment of histone methyltransferase KMT2C and consequently the maintenance of H3K4me1 levels by ChIP-qPCR at a few loci (Jozwik et al., 2016). Through stepwise hPSC differentiation, we were able to track sequential epigenomic events during developmental fate transitions, which allowed us to identify a specific requirement of *FOXA2* in the establishment of a primed enhancer state and chromatin architecture during developmental lineage specification. This requirement may rely on the unique ability of pioneer TFs to bind nucleosomal DNA and subsequently reshape the epigenome to activate a lineage-specific transcriptional program during development.

STAR★METHODS

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for reagents may be directed to and will be fulfilled by the Lead Contact, Dr. Danwei Huangfu (huangfud@mskcc.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human embryonic stem cell lines HUES8 (NIHhESC-09–0021, male) and H1 (NIHhESC-10–0043, male) were authenticated by the standard short tandem repeat (STR) profiling using the MSK Integrated Genomics Operation core facility. Experiments involving hPSCs were conducted per NIH guidelines and approved by the Tri-SCI Embryonic Stem Cell Research Oversight Committee. Undifferentiated hPSCs were cultured in complete E8 culture medium (Thermo Fisher Scientific, A1517001) at 37°C with 5% CO_2. The E8

medium was changed every day, and cells were passaged every 4 days using 0.5 mM EDTA (KD Medical, RGE-3130) and 5 μM Rho-associated protein kinase (ROCK) inhibitor Y-27632 (Selleck Chemicals, S1049). Plates coated with vitronectin (VTN, Thermo Fisher Scientific, A14700) were used for passage. Mycoplasma test was regularly performed by the MSKCC Antibody & Bioresource Core Facility.

METHOD DETAILS

Generation and expansion of hPSC Mutant Lines—The iCRISPR platform in HUES8 or H1 background was used to generate mutant lines in this study (González et al., 2014; Shi et al., 2017). A day before guide RNA (gRNA) transfection, 2 μg/ml of doxycycline (Sigma, D9891) and 5 μM of Y-27632 were added to iCas9 hPSCs culture medium. On the following day, cells were passaged using TrypLE select (Thermo Fisher Scientific, 12563–029) with a 1:3 – 1:6 ratio on VTN-coated plates in the presence of 2 μg/ml of doxycycline and 5 μM of Y-27632. Lipofectamine RNAiMAX (Thermo Fisher Scientific, 13778–150) was used for transfection following manufacturer's instructions. gRNAs targeting *FOXA* factors were diluted in Opti-MEM (Thermo Fisher Scientific, 31985070). The sequences of gRNAs are listed in Table S4. Lipofectamine RNAiMAX was also diluted in Opti-MEM and then mixed with diluted gRNAs. After incubating 5 min at room temperature (RT), the gRNA-lipid complex containing 10 nM of gRNAs was added to cells.

After transfection, cells were cultured in the presence of 2 μg/ml of doxycycline and 5 μM of Y-27632 for one more day. Two days after the transfection, targeted cells were passaged using TrypLE select and plated at clonal density (2,000 cells per 10 cm dish). The remaining cells were collected, and genomic DNA was extracted using the DNeasy kit (QIAGEN, 69506) for T7 endonuclease I assay to measure CRISPR mutagenesis efficiency. 10 days later, colonies derived from single cells were picked by mechanical dissociation into individual wells of 96-well plates pre-coated with VTN. Genomic DNA was harvested from each colony using lysis buffer of 1 × PCR buffer (Sigma, P2192) and 1 μg/ml of proteinase K (Fisher Scientific, BP1700–100). After incubating overnight at 55°C, lysis was inactivated by incubating at 100°C for 10 min. Herculase II Fusion DNA Polymerase (Agilent Technologies, 600679) was used to perform PCR and then targeted clones were identified by Sanger sequencing. Primers used for amplification and sequencing are listed in Table S4. Western blotting was performed during pancreatic differentiation to confirm the loss of FOXA1 or FOXA2 protein. Verified clonal lines were stored for further analysis. WT lines from the same targeting experiment were also expanded and used for pancreatic differentiation. During pancreatic differentiation, FOXA1 or FOXA2 protein expression was re-analyzed by flow cytometry to confirm the correct genotypes. *FOXA2* mutant cells in HUES8 background were used for RNA-seq, ATAC-seq and ChIP-seq experiments and *FOXA1* or *FOXA2* mutant cells in H1 background were analyzed in Figure S2.

hPSC Pancreatic Lineage Differentiation—All differentiation experiments were repeated at least 3 times independently with 3 individual lines of the same genotype. Isogenic WT lines derived from the same targeting experiments were used for each differentiation experiment together with KO lines. Pancreatic differentiation was performed

as previously described (Pagliuca et al., 2014; Rezania et al., 2014; Shi et al., 2017; Zhu et al., 2016). Briefly, hPSCs were passaged at 65k - 100k cells/cm^2 density. When cells reached 80 – 90% confluence, pancreatic differentiation was initiated (d0). Definitive endoderm (DE) stage was induced by 100 ng/ml of Activin A (PeproTech, 120–14E) for 3 days, 5 μM GSK-3 inhibitor, CHIR-99021 (Stemgent, 04–0004) for the first day, and 0.5 μM CHIR-99021 for the second day. Following treatment of 0.25 mM of L-Ascorbic acid (Sigma-Aldrich, A4544) and 50 ng/ml of FGF7 (R&D, 251-KG) for 2 days resulted in Foregut (FG) stage. PDX1$^+$ early pancreatic progenitor 1 (PP1) cells were generated by adding 0.25 mM of L-Ascorbic acid, 50 ng/ml of FGF7, 250 nM of the hedgehog inhibitor, SANT-1 (Sigma, S4572), 1 μM of retinoic acid (Sigma, R2625), 100 nM of the BMP inhibitor, LDN-193189 (Stemgent, 04–0019), and 200 nM of PKC activator, TPB (EMD Millipore, 565740) for 2 days. Subsequently, PDX1$^+$NKX6–1$^+$ pancreatic progenitor (PP2) were generated by adding 0.25 mM of L-Ascorbic acid, 2 ng/ml of FGF7, 250 nM of SANT-1, 0.1 μM of retinoic acid, 200 nM of LDN, and 100 nM of TPB for 3 days. The differentiation efficiency was analyzed by measuring lineage marker expression using flow cytometry and immunostaining; SOX17 expression on day 3 for the DE stage, HNF1B expression on day 5 for the FG stage, PDX1 expression on day 7 for the PP1 stage, and PDX1 and NKX6–1 expression on day 10 for the PP2 stage.

Immunofluorescence Staining—Differentiated cells on tissue culture plates were fixed with 4% paraformaldehyde in phosphate-buffered saline (PBS) for 10 min at RT followed by washing with PBS. Permeabilization was performed with PBS with 0.2% Triton (PBST) by incubating for 30 min at RT. 5% donkey serum in PBST was used as a blocking solution and antibody staining solution. After blocking for 30 min at RT, blocking solution containing primary antibodies was added. The following primary antibodies and dilutions were used: goat anti-PDX1 (R&D, AF2419, 0.4 μg/ml), mouse anti-NKX6–1 (DSHB F55A12-c, 1:500). Primary antibodies in blocking solution were incubated overnight at 4°C. After washing with PBST, Alexa Fluor secondary antibodies (Molecular Probes, 1:500) were incubated at RT for 1 hour. DAPI (Sigma, 32670–5MG-F, 0.2 μg/ml) was used to stain cell nuclei. Zeiss Axio Observer microscope was used for imaging. Antibodies for this study are summarized in Table S8.

Flow Cytometry—To dissociate cells to single cells, TrypLE Select was used and cells were collected with ice-cold FACS buffer (2% FBS in PBS). LIVE-DEAD Fixable Violet Dead Cell Stain (Thermo Fisher Scientific, L34955, 1:1,000) was used to discriminate dead cells from live cells. After incubating for 10 min at RT with LIVE/DEAD stain and surface markers including mouse CXCR4-APC (R&D, FAB170A, 1:25), cells were washed with FACS buffer. The Foxp3 Staining Buffer Set (eBioscience, 00–5523-00) was used to perform intracellular staining to measure the expression level of nuclear transcriptional factors following the manufacturer's instructions. Permeabilization/fixation was performed at RT for 1 hour. Antibody staining was performed in permeabilization buffer with the following antibodies: rat OCT¾-eFluor660 (eBioscience, 50–5841-82, 1:50), rat SOX2-Alexa Fluor 488 (eBioscience, 53–9811-82, 1:50), mouse SOX2-PE (Biolegend, 656103, 1:50), mouse anti-SOX17-PE (BD Biosciences, 561591, 1:50), goat anti-FOXA2-Alexa Fluor 488 (R&D, IC2400G, 1:50), mouse anti-HNF1B (Santa Cruz, sc-130407, 1:100), goat

anti-PDX1 (R&D, AF2419, 1:250), mouse anti-NKX6–1 (DSHB, F55A12-c, 1:250) and Alexa Fluor secondary antibodies (Molecular Probes, 1:500). Antibodies for this study are summarized in Table S8.

Flow cytometry data was analyzed by FlowJo V9. From FSC-A/SSC-A plot, P1 gate was drawn by excluding debris. From P1 population, singlet was isolated sequentially by excluding doublets from FSC-W/FSC-H (P2) and SSC-W/SSC-H plots (P3). From P3 population, the negative population from Live/Dead staining was selected as Live cells (P4). From Live cells, the frequency of cells with markers was measured. An example of the flow gating strategy is shown in Figure S1D.

Western Blotting—Cells were harvested and lysed with cell lysis buffer (Cell Signaling Technology, 9803) containing protease inhibitor (Roche, 05892791001) and 1 mM PMSF (MP Biomedicals, ICN19538105). Sample preparation follows NuPAGE Novex protocol. 10% Bis-Tris Gel (Novex, NP0301BOX) was used and transferred to Nitrocellulose Pre-Cut Blotting Membranes (Novex, LC2001). Blocking was performed with 5% milk in Tris-based saline with 0.1% Tween 20 (TBST) for 30 min at RT. After blocking, primary antibodies were diluted in blocking solution and incubated overnight at 4°C. The following antibodies were used with noted dilution ratios: goat anti-FOXA2 (R&D, AF2400, 1:1000), rabbit anti-FOXA2 (Cell Signaling Technology, 3143S, 1:1000), rabbit anti-FOXA1 (Cell Signaling Technology, 58613S, 1:1000), mouse anti-ACTB (Cell Signaling Technology, 3700S, 1:10,000). After washing with TBST, HRP conjugated secondary antibodies in blocking solution were incubated at RT for 1 hour and the gel was visualized using ECL western blotting detection reagent (Amersham, RPN2236). Antibodies for this study are summarized in Table S8.

RNA isolation for quantitative RT-PCR and RNA-seq sample preparation—Cells were lysed in TRIzol (Thermo Fisher Scientific, 15596018) and total RNA was isolated using the RNeasy Plus Mini Kit (QIAGEN, 74134). High Capacity cDNA Reverse Transcription Kit was used to generate cDNA (Applied Biosystems, 4368814). Quantitative real-time PCR was performed in triplicate using ABsolute Blue QPCR SYBR Green Mix with low ROX (Thermo Fisher Scientific, AB4322B) on the ABI PRISM 7500 Real-Time PCR System (Applied Biosystems). Primers used for qRT-PCR are listed in Table S4.

The quality of RNA was assessed using TapeStation 2200, Agilent Technologies, and 1,000ng of total RNA were used for cDNA Library generation with QuantSeq 3′mRNA-Seq Library Prep Kit FWD (Lexogen, Vienna Austria). According to manufacturer's recommendations, a common set of external RNA controls were used (ERCC RNA Spike-In mix, ThermoFisher Scientific, 4456740). Samples were submitted to New York Genome Center for SE50 sequencing using a HiSeq 2000. Three independent clonal lines per genotype were used for RNA-seq analysis.

Sample Preparation for ATAC-seq

Cell lysis: After washing with 1 mL of cold PBS, ~50,000 cells were washed with 1 mL of ice-cold ATAC Buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$), and then cells were lysed with 50 μL of ATAC-Lysis Buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM

$MgCl_2$, 0.1% NP-40 or IGEPAL-Ca630). After 2 min incubation on ice with ATAC-lysis buffer, cells were diluted with 1 mL of ice-cold ATAC Buffer, and pellets were collected by centrifugation at 1,500 rpm for 10 min at 4°C. Two independent clonal lines per genotype were used for ATAC-seq analysis.

Tagmentation: Pellets were resuspended in 22.5 μL of residual supernatant and transferred to a PCR tube. 2.5 μL of Tagmentation Enzyme (transposase) and 25 μL of Tagmentation Buffer (Illumina Nextera DNA Sample Preparation Kit) were added to each tube and incubated at 37°C for 30 min. SDS was added to samples to a final concentration of 0.2% and incubated for 5 min at RT. 2X Agencourt AMPure XP beads (Beckman Coulter A63881) was used for purification and elution was performed with 50 μL of 0.1 × TE buffer.

Barcoding and Library Preparation: NEBNext Q5 Hot Start HiFi PCR Master Mix (NEB, M0543L) and Nextera primers (Buenrostro et al., 2015) were used for library preparation. 55 μL of PCR master mix and 5 μL of primer mix were used for 50 μL of samples. Every sample had 25 μM of final concentration by mixing the universal primer Ad1 and the unique index primer 2.X (ex. 2.1, 2.2, 2.3, etc.). The following condition was used for amplification: 65°C, 5 minutes; 98°C, 30 s; 98°C, 10 s, 65°C 30 s, 12 cycles; 4°C indefinitely. 1.5X AMPure XP beads were used for sample purification. The sample concentration was checked using PicoGreen and median fragment size was checked using the Agilent D1000 screentape on the Agilent Technologies 2200 TapeStation. The library was sent to New York Genome Center or MSKCC Integrated Genomics Operation (IGO) center for PE50 sequencing using a HiSeq 2500.

Sample Preparation for ChIP-Seq

Fixation: Crosslinking was performed with 1% formaldehyde (Sigma, F1635) at 37°C for 10 min. A final concentration of 125 mM glycine was added for 5 minutes at RT to quench formaldehyde crosslinking. Cells were washed with ice-cold PBS, harvested on ice and then transferred to 50 mL falcon tubes. After centrifugation at 4,000 rpm for 5 min at 4°C, cell pellets were transferred to Eppendorf tubes and frozen using liquid nitrogen. The cell pellets were stored at −80°C for further analysis. Two independent clonal lines per genotype were used for ChIP-seq analysis.

Sonication: ~50 million cells were lysed with 1ml of SDS lysis buffer (1% SDS, 10mM EDTA, 50mM Tris-HCl pH8) in the presence of protease inhibitor (Roche, #11836170001) and phosphatase inhibitor (Thermo #78427). After 10 min incubation on ice, cells were disrupted using the Branson Sonifier 250 (Branson Ultrasonics 101063588) with a 20% amplitude setting for 5 min 30 s with 10 s of on and off intervals. After sonication, 1 mL of ChIP dilution buffer (0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl pH8, 167 mM NaCl) with proteinase and phosphatase inhibitor was added to prevent SDS precipitation. After centrifugation at 14,000 rpm for 10 min at 4°C, 2 mL of supernatant were transferred to a 15 mL falcon tube and diluted with 8 mL of ChIP dilution buffer containing proteinase and phosphatase inhibitor.

Pre-clearing and antibody incubation: 50 μL of Dynabeads (Thermo Fisher Scientific 10009D), which were previously blocked with 1% BSA, were added to samples and incubated at 4°C with rotation for 1 hour. After pre-clearing, Dynabeads beads were removed by spinning down at 4,000 rpm on 4°C for 5 min. The supernatant was transferred to new 15 mL falcon tubes and 200 μL out of 10 mL volume was collected as 2% input. Antibodies were added to the pre-cleared samples for overnight incubation at 4°C with rotation. The following antibodies were used: rabbit anti-H3K4me1 (Abcam, ab8895, 5 μg), rabbit anti-H3K27ac (active motif, 39133, 5 μg), goat anti-FOXA2 (R&D, AF2400, 10 μg), rabbit anti-GATA6 (Cell Signaling Technology, 5851, 10 μg), goat anti-GATA4 (R&D, AF2606, 10 μg), rabbit anti-HNF1B (Santa Cruz, sc-22840x, 10 μg), goat anti-PDX1 (R&D, AF2419, 10 μg). 200 μL of Dynabeads, which were previously blocked with 1% BSA, were incubated at 4°C for 6 hours with rotation. After bead incubation, chromatin-bound beads were collected by centrifugation at 3,000 rpm for 5 min at 4°C.

Bead washing and de-crosslinking: Beads were washed for 5 min each at 4°C with rotations in the following order: one time with low salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH8, 150 mM NaCl), twice with high salt buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH8, 500 mM NaCl) and twice with TE buffer (10 mM Tris-HCl pH8, 1 mM EDTA). After washing, beads were resuspended in 250 μL of elution buffer (1% SDS, 0.1 M NaHCO3) and incubated at 60°C for 30 min using a thermomixer (850 rpm). After separating beads, the supernatant was transferred to new tubes and 5 M NaCl was added for overnight de-crosslinking at 65°C. On the following day, 10 μL of 0.5 M EDTA, 20 μL of 1 M Tris-HCl pH6.5 and 1 μL of proteinase K (20 mg/ml, NEB, P8107S) were added and incubated at 45°C for 1 hour for de-crosslinking.

Library preparation: ChIP-DNA was isolated by using the QIAquick PCR purification kit (QIAGEN, 28104) and NEBNext® ChIP-seq Library Prep Master Mix Set for Illumina ® (NEB, E6240L) was used for library preparation. Agilent Technologies 2200 TapeStation was used to determine fragment size of DNA and the concentration of DNA was measured by PicoGreen (Thermo Fisher Scientific, P7589). Samples were pooled and submitted to New York Genome Center or MSKCC Integrated Genomics Operation (IGO) center for SE50 sequencing using a HiSeq 2500.

QUANTIFICATION AND STATISTICAL ANALYSIS

Gene expression profiling by RNA-seq—Quality control of sequenced reads was performed by FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and adaptor filtration was performed by Trimmomatic version 0.36. The filtered reads were aligned to the hg19 reference genome and GENCODE version 19 gene annotation (Harrow et al., 2012), using STAR aligner version 2.5.3a (Dobin et al., 2013). Expression level of exonic regions was quantified using htseq-count (Anders et al., 2015). Only the genes with the normalized expression above 1 tag per million (TPM) in at least one cell type were selected to gain specificity of the differential analysis.

The gene expression was analyzed with the generalized linear model involving two factors of (1) the developmental stage and (2) the genotype of *FOXA2*. Genes with *FOXA2*

knockout effects after controlling for the effect of successive development stages were collected by four successive pairwise comparisons. The likelihood ratio test was done to assess the additive effect of the *FOXA2* genotype under an FDR cutoff of 0.001. A total of 890 genes was defined from the union of four comparisons, to be used in the principal component analysis and the cluster analysis in Figure 2.

The count data were transformed by DESeq2′s variance-stabilizing transformation, quantile normalization, and Z-score transformation for use in the cluster analysis. The dendrograms were drawn using pheatmap (https://cran.r-project.org/web/packages/pheatmap/pheatmap.pdf) with the average linkage for columns. The order of genes was defined by the clustering of each stage-specific group.

Chromatin landscape profiling by ATAC-seq—Quality control of sequenced reads was performed by FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and adaptor filtration was performed by Trimmomatic version 0.36, with the Nextera paired-end adaptor annotation provided in the software. The filtered reads were aligned to the hg19 reference genome using bowtie2 version 2.3.3.1 (Langmead and Salzberg, 2012) with parameters -N 1–mm. Ends of the aligned read were shifted to remove the artifact of Tn5 transposase, as described in (Buenrostro et al., 2013; van der Veeken et al., 2016)

Macs2 version 2.1.1.20160309 (Zhang et al., 2008) was used for removing duplicate reads and calling peaks. In the duplicate removal step, macs2 filterdup–keep-dup auto was used. Subsequently, in the peak calling step, all samples in the given comparison set were used in the treatment input, and the candidate peaks were found using a permissive threshold, with parameter options–shift −100–extsize 200 -g hs -p 1e-2–to-large–keep-dup all, effectively lengthening the aligned reads by 100 bp in each side. The coverage signal of each cell type was generated by merging all replicates and running macs2 callpeaks with the parameters -B–SPMR and converting the treatment pileup file to bigWig format.

Peaks were filtered using Irreproducible Discovery Rate (IDR) version 2.0.3 (Li et al., 2011) using the two replicates of each cell type with the threshold of 0.01. Filtered peaks showing reproducibility in any cell type were combined to create the atlas used in subsequent analyses. Differentially accessible peaks in the atlas were called by DESeq2, with FDR threshold of 0.01 to define stage-specific peaks, and 0.05 for the same-stage comparison between WT and *FOXA2* KO. HOMER findMotifsGenome.pl (Heinz et al., 2010) was used to investigate the motif enrichment of stage-specific groups compared with the total atlas using default parameters (Heinz et al., 2010). To examine which motifs were enriched among the DE-accessible groups (DE specific, DE–FG, and DE–PP1), we combined all three groups as the background and compared each group to this background.

Definition and visualization of ATAC-seq stage-specific groups—To define stage-specific groups of peaks in ATAC-seq as shown in Figure 2A, we implemented a generalized linear model using DESeq2 to identify peaks following specified up/down patterns across cell types. In this way, we found peaks whose accessibility in any cell type belonging to the “up” set showed greater magnitude than any of those in the “down” set, with the Benjamini-Hochberg corrected *P* value less than the cutoff of 0.01. For example, the DE specific group

had only DE as up and ES, FG, and PP1 as down, whereas the DE-PP1 group had DE, FG, and PP1 as up and ES as down.

Signal tracks of each cell type in bigWig file format were used to generate the visualization of ATAC-seq profile in stage-specific groups in Figure 2. Their units were TPM from the 2 replicates combined, using MACS2 parameters of -B–SPMR. For each peak, the corresponding bigWig file was read for the genomic region within 3 kb from its summit. Finally, in representing the visualization with the grids of tornado plot, the maximum color was defined as the 99th percentile of the bigWig file readout for the assay of interest, so as to mitigate the effect of highly accessible outliers and show the contrast more effectively in the low-accessibility regions. The color scale was defined proportional to the 99th percentile, and values above it were capped to be the same as the maximum.

Motif enrichment by Kolmogorov-Smirnov test—ATAC-seq peaks in the atlas were associated with TF motifs in the CIS-BP database (Weirauch et al., 2014) using FIMO (Grant et al., 2011) of MEME suite (Bailey et al., 2009), under the *P value* cutoff of 1e-4. We chose 125 TFs that showed expression above 100 TPM in at least one cell type in the 3′-seq RNA-seq samples. CTCF, a DNA-binding protein associated with 3D chromatin structure, and DNA methylation-related enzyme, such as DNMT1 and TET1, were excluded.

During the transition between differentiation stages, the shift in the cumulative distribution of chromatin accessibility changes in WT and *FOXA2* KO was compared between the subset of the atlas containing the TF and the total atlas, which was a collection of reproducible ATAC-seq peaks from all stages. The comparison was measured by a one-sided Kolmogorov-Smirnov (KS) test in either direction.

Normalized counts were obtained using DEseq2′s variance-stabilizing transformation and averaging over replicates. The shift in the cumulative distribution of normalized counts was compared between the subset of the atlas containing the TF motif and the total atlas, as shown in Figure 2C inset. The comparison was made by a one-sided KS test.

Motif enrichment by hypergeometric test—For ATAC-seq peaks in stage-specific groups, the proportion of peaks containing a transcription factor motif in each group (foreground ratio) was compared with that in the entire atlas (background ratio) using the hypergeometric test. We used the binomial Z-score to visualize the motif enrichment. The expected count was calculated from the size of the foreground group and the background ratio, and the standard deviation was estimated based on the binomial distribution. Then the observed count was transformed to a Z-score, showing the number of standard deviations away from the expected count.

2-dimensional visualization of paired-end ATAC-seq data (V-plot)—For ATAC-seq reads in each cell type mapped with proper pair orientation, the reads at the outer end were trimmed by 9 bp, the length of base insertion by Tn5 transposase (Buenrostro et al., 2013). We calculated their insert size as the distance between the leftmost and the rightmost mapped position. We also defined the representative location of the read pair as the midpoint of the leftmost and the rightmost mapped position. For each ATAC-seq peak grouped by

stage-specific groups, we counted the read pairs whose representative position was within 1 kb base pairs of the peak summit. The distance from the summit to read was aggregated over each stage-specific group. The 2-dimensional kernel density for each 1 bp × 1 bp area was estimated using the MASS R package and default bandwidth (Venables et al., 2002). The density plot (V-plot) of insert size against the relative position was drawn for stage-specific groups of WT and *FOXA2* KO. For proper comparison, the density was normalized, proportional to the number of total reads in the summit ± 500 bp window and inversely proportional to the size of stage-specific groups.

V-plot centered by the distance to FOXA2 ChIP peaks—While aggregating the ATAC-seq read pairs mapped to stage-specific peak regions, the horizontal axis for the V-plot was defined as the distance from the midpoint to the nearest peak summit in the FOXA2 ChIP-seq of the respective cell type. The density plot was shown only for the regions of FOXA2 ChIP summit ± 500 bp. The density was normalized similarly to the V-plot centered by ATAC-seq signals.

Separate visualization of nucleosome-associated and nucleosome-free ATAC-seq reads—To distinguish ATAC-seq read pairs mapped to the open chromatin from those mapped next to the positioned nucleosome, we divided the reads mapped with proper pair orientation with the insert size cutoff of 150 bp. Pairs with insert size greater than or equal to 10 bp and less than 150 bp were labeled as nucleosome-free regions, and pairs with insert size greater than or equal to 150 bp as nucleosome-associated regions. The reads were visualized by the tornado plot on the domain of summit ± 1 kb window.

Expression of primed peak-associated genes (beeswarm plot)—Using the table of gene annotation (Table S7), the total of 686 genes was found from 1136 primed PP1 peaks. The normalized count on each cell type was transformed by DESeq2′s variance-stabilizing transformation and the primed PP1 genes were selected. The subset was then averaged over replicates and scaled by each row to produce Z-scores. The R package beeswarm was used to plot the data points in the cell type, which were overlaid on the boxplot. A one-sided unpaired Wilcoxon rank sum test was used to compare the expression profile of cell types.

Transcription factor ChIP-seq peak calling and analysis for the intersection—To find the number of FOXA2 and GATA6 ChIP-seq peaks co-occurring at the same location, an unbiased and lenient peak calling was done on FOXA2 and GATA6 ChIP-seq samples of all replicates in the same cell type, using macs2 callpeak and the default method to estimate the fragment length. For each cell type, the ChIP-seq atlas was defined where the IDR of either FOXA2 or GATA6 was less than 0.05. Then the coverage of ChIP-seq on peaks was quantified by RPKM, with the library size defined as the number of reads mapped to any genomic location within 3 kb of summits in the ChIP-seq atlas. Finally, the number of peaks was counted where the average level was above 10 RPKM for FOXA2 or GATA6 for each cell type.

Quantification and statistical analysis of flow cytometry—Quantification of flow cytometry data was presented as mean ± SD. Data from three independent clonal lines of the

same genotype were combined for calculating the significance of the differences between different genotypes. Unpaired two-tailed Student's t test was used to compare two groups with Prism 6. $p < 0.05$, 0.01, 0.001 and 0.0001 are indicated with 1, 2, 3 and 4 asterisks, respectively.

DATA AND CODE AVAILABILITY

Sequencing datasets generated in this paper are available under NCBI GEO: GSE114102 and GSE109524 (GSM3381978 and GSM3381980). Code used to perform the analyses in this study is available from the corresponding authors upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Matthew D. Witkin for assistance with ATAC-seq and ChIP-seq experiments; E. Apostolou and T. Vierbuchen for critical comments on the manuscript; and T. Evans, L. Studer, and members of D.H.'s laboratory for insightful discussions. This study was supported in part by the NIH/NIDDK (R01DK096239), the New York State Stem Cell Science (NYSTEM C029567, C029156, and C32593GG), and the MSKCC Cancer Center Support Grant (P30CA008748). K.L. was supported by the Korean Government Scholarship Program for Study Overseas and the Mogam Science Scholarship Foundation. The monoclonal antibody NKX6-1 (DSHB F55A12-c) developed by Hagedorn Research Institute was obtained from the Developmental Studies Hybridoma Bank, created by the NICHD of the NIH, and maintained at the University of Iowa, Department of Biology, Iowa City, IA 52242, USA.

REFERENCES

Allen HL, Flanagan SE, Shaw-Smith C, De Franco E, Akerman I, Caswell R, Ferrer J, Hattersley AT, and Ellard S; International Pancreatic Agenesis Consortium (2011). GATA6 haploinsufficiency causes pancreatic agenesis in humans. Nat. Genet 44, 20–22. [PubMed: 22158542]

Amin S, Cook B, Zhou T, Ghazizadeh Z, Lis R, Zhang T, Khalaj M, Crespo M, Perera M, Xiang JZ, et al. (2018). Discovery of a drug candidate for GLIS3-associated diabetes. Nat. Commun 9, 2681. [PubMed: 29992946]

Anders S, Pyl PT, and Huber W (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. [PubMed: 25260700]

Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, and Noble WS (2009). MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 37, W202–W208. [PubMed: 19458158]

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, and Greenleaf WJ (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. [PubMed: 24097267]

Buenrostro JD, Wu B, Chang HY, and Greenleaf WJ (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. Curr. Protoc. Mol. Biol 109, 21.29.1–21.29.9.

Calo E, and Wysocka J (2013). Modification of enhancer chromatin: what, how, and why? Mol. Cell 49, 825–837. [PubMed: 23473601]

Carrasco M, Delgado I, Soria B, Martín F, and Rojas A (2012). GATA4 and GATA6 control mouse pancreas organogenesis. J. Clin. Invest 122, 3504–3515. [PubMed: 23006330]

Chia CY, Madrigal P, Denil SLIJ, Martinez I, Garcia-Bernardo J, El-Khairi R, Chhatriwala M, Shepherd MH, Hattersley AT, Dunn NR, and Vallier L (2019). GATA6 Cooperates with EOMES/SMAD2/3 to Deploy the Gene Regulatory Network Governing Human Definitive Endoderm and Pancreas Formation. Stem Cell Reports 12, 57–70. [PubMed: 30629940]

Cirillo LA, McPherson CE, Bossard P, Stevens K, Cherian S, Shim EY, Clark KL, Burley SK, and Zaret KS (1998). Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J. 17, 244–254. [PubMed: 9427758]

Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, and Zaret KS (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell 9, 279–289. [PubMed: 11864602]

Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA 107, 21931–21936. [PubMed: 21106759]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]

Donaghey J, Thakurela S, Charlton J, Chen JS, Smith ZD, Gu H, Pop R, Clement K, Stamenova EK, Karnik R, et al. (2018). Genetic determinants and epigenetic effects of pioneer-factor occupancy. Nat. Genet 50, 250–258. [PubMed: 29358654]

Fisher JB, Pulakanti K, Rao S, and Duncan SA (2017). GATA6 is essential for endoderm formation from human pluripotent stem cells. Biol. Open 6, 1084–1095. [PubMed: 28606935]

Gage BK, Asadi A, Baker RK, Webber TD, Wang R, Itoh M, Hayashi M, Miyata R, Akashi T, and Kieffer TJ (2015). The Role of ARX in Human Pancreatic Endocrine Specification. PLoS ONE 10, e0144100. [PubMed: 26633894]

Gao N, LeLay J, Vatamaniuk MZ, Rieck S, Friedman JR, and Kaestner KH (2008). Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. Genes Dev. 22, 3435–3448. [PubMed: 19141476]

Genga RMJ, Kernfeld EM, Parsi KM, Parsons TJ, Ziller MJ, and Maehr R (2019). Single-Cell RNA-Sequencing-Based CRISPRi Screening Resolves Molecular Drivers of Early Human Endoderm Development. Cell Rep 27, 708–718. [PubMed: 30995470]

González F, Zhu Z, Shi ZD, Lelli K, Verma N, Li QV, 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. [PubMed: 24931489]

Grant CE, Bailey TL, and Noble WS (2011). FIMO: scanning for occurrences of a given motif. Bioinformatics 27, 1017–1018. [PubMed: 21330290]

Hannenhalli S, and Kaestner KH (2009). The evolution of Fox genes and their role in development and disease. Nat. Rev. Genet 10, 233–240. [PubMed: 19274050]

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. 22, 1760–1774. [PubMed: 22955987]

Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet 39, 311–318. [PubMed: 17277777]

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, and Glass CK (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589. [PubMed: 20513432]

Iwafuchi-Doi M, and Zaret KS (2014). Pioneer transcription factors in cell reprogramming. Genes Dev. 28, 2679–2692. [PubMed: 25512556]

Jonsson J, Carlsson L, Edlund T, and Edlund H (1994). Insulin-promoter-factor 1 is required for pancreas development in mice. Nature 371, 606–609. [PubMed: 7935793]

Jozwik KM, Chernukhin I, Serandour AA, Nagarajan S, and Carroll JS (2016). FOXA1 Directs H3K4 Monomethylation at Enhancers via Recruitment of the Methyltransferase MLL3. Cell Rep. 17, 2715–2723. [PubMed: 27926873]

Kaestner KH, Hiemisch H, and Schütz G (1998). Targeted disruption of the gene encoding hepatocyte nuclear factor 3gamma results in reduced transcription of hepatocyte-specific genes. Mol. Cell. Biol 18, 4245–4251. [PubMed: 9632808]

Kaestner KH, Katz J, Liu Y, Drucker DJ, and Schütz G (1999). Inactivation of the winged helix transcription factor HNF3alpha affects glucose homeostasis and islet glucagon gene expression *in vivo*. Genes Dev. 13, 495–504. [PubMed: 10049364]

Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. [PubMed: 22388286]

Lee CS, Sund NJ, Behr R, Herrera PL, and Kaestner KH (2005). Foxa2 is required for the differentiation of pancreatic alpha-cells. Dev. Biol 278, 484–495. [PubMed: 15680365]

Li Q, Brown JB, Huang H, and Bickel PJ (2011). Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat 5, 1752–1779.

Li Z, Gadue P, Chen K, Jiao Y, Tuteja G, Schug J, Li W, and Kaestner KH (2012). Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. Cell 151, 1608–1616. [PubMed: 23260146]

Li QV, Dixon G, Verma N, Rosen BP, Gordillo M, Luo R, Xu C, Wang Q, Soh C-L, Yang D, et al. (2019). Genome-scale screens identify JNK-JUN signaling as a barrier for pluripotency exit and endoderm differentiation. Nat. Genet 51, 999–1010. [PubMed: 31110351]

Liu L, Zhao W, and Zhou X (2016). Modeling co-occupancy of transcription factors using chromatin features. Nucleic Acids Res. 44, e49. [PubMed: 26590261]

Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, and Brown M (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell 132, 958–970. [PubMed: 18358809]

Mayran A, Khetchoumian K, Hariri F, Pastinen T, Gauthier Y, Balsalobre A, and Drouin J (2018). Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. Nat. Genet 50, 259–269. [PubMed: 29358650]

McGrath PS, Watson CL, Ingram C, Helmrath MA, and Wells JM (2015). The Basic Helix-Loop-Helix Transcription Factor NEUROG3 Is Required for Development of the Human Endocrine Pancreas. Diabetes 64, 2497–2505. [PubMed: 25650326]

Pagliuca FW, Millman JR, Gürtler M, Segel M, Van Dervort A, Ryu JH, Peterson QP, Greiner D, and Melton DA (2014). Generation of functional human pancreatic β cells in vitro. Cell 159, 428–439. [PubMed: 25303535]

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, and Wysocka J (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470, 279–283. [PubMed: 21160473]

Rezania A, Bruin JE, 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. [PubMed: 25211370]

Sekiya T, Muthurajan UM, Luger K, Tulin AV, and Zaret KS (2009). Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. Genes Dev. 23, 804–809. [PubMed: 19339686]

Shaw-Smith C, De Franco E, Lango Allen H, Batlle M, Flanagan SE, Borowiec M, Taplin CE, van Alfen-van der Velden J, Cruz-Rojo J, Perez de Nanclares G, et al. (2014). GATA4 mutations are a cause of neonatal and childhood-onset diabetes. Diabetes 63, 2888–2894. [PubMed: 24696446]

Shi ZD, Lee K, Yang D, Amin S, Verma N, Li QV, Zhu Z, Soh CL, 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. [PubMed: 28196600]

Spitz F, and Furlong EE (2012). Transcription factors: from enhancer binding to developmental control. Nat. Rev. Genet 13, 613–626. [PubMed: 22868264]

Stoffers DA, Zinkin NT, Stanojevic V, Clarke WL, and Habener JF (1997). Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence. Nat. Genet 15, 106–110. [PubMed: 8988180]

Tabar V, and Studer L (2014). Pluripotent stem cells in regenerative medicine: challenges and recent progress. Nat. Rev. Genet 15, 82–92. [PubMed: 24434846]

Tiyaboonchai A, Cardenas-Diaz FL, Ying L, Maguire JA, Sim X, Jobaliya C, Gagne AL, Kishore S, Stanescu DE, Hughes N, et al. (2017). GATA6 Plays an Important Role in the Induction of Human

Definitive Endoderm, Development of the Pancreas, and Functionality of Pancreatic β Cells. Stem Cell Reports 8, 589–604. [PubMed: 28196690]

Tsompana M, and Buck MJ (2014). Chromatin accessibility: a window into the genome. Epigenetics Chromatin 7, 33. [PubMed: 25473421]

van der Veeken J, Gonzalez AJ, Cho H, Arvey A, Hemmers S, Leslie CS, and Rudensky AY (2016). Memory of Inflammation in Regulatory T Cells. Cell 166, 977–990. [PubMed: 27499023]

Venables WN, Ripley BD, and Venables WN (2002). Modern Applied Statistics with S, Fourth Edition (Springer).

Wang A, Yue F, Li Y, Xie R, Harper T, Patel NA, 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. [PubMed: 25842977]

Weirauch MT, Yang A, Albu M, Cote AG, Montenegro-Montero A, Drewe P, Najafabadi HS, Lambert SA, Mann I, Cook K, et al. (2014). Determination and inference of eukaryotic transcription factor sequence specificity. Cell 158, 1431–1443. [PubMed: 25215497]

Xuan S, and Sussel L (2016). GATA4 and GATA6 regulate pancreatic endoderm identity through inhibition of hedgehog signaling. Development 143, 780–786. [PubMed: 26932670]

Xuan S, Borok MJ, Decker KJ, Battle MA, Duncan SA, Hale MA, Macdonald RJ, and Sussel L (2012). Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic agenesis. J. Clin. Invest 122, 3516–3528. [PubMed: 23006325]

Zentner GE, Tesar PJ, and Scacheri PC (2011). Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. Genome Res. 21, 1273–1283. [PubMed: 21632746]

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, and Liu XS (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137. [PubMed: 18798982]

Zhu Z, and Huangfu D (2013). Human pluripotent stem cells: an emerging model in developmental biology. Development 140, 705–717. [PubMed: 23362344]

Zhu Z, Li QV, Lee K, Rosen BP, González F, Soh CL, and Huangfu D (2016). Genome Editing of Lineage Determinants in Human Pluripotent Stem Cells Reveals Mechanisms of Pancreatic Development and Diabetes. Cell Stem Cell 18, 755–768. [PubMed: 27133796]

Highlights

- ATAC-seq combined with motif analysis predicts key transcription factor involvement
- FOXA2 is necessary for efficient formation of human pancreatic progenitor cells
- Remodeling of chromatin architecture during pancreatic differentiation requires FOXA2
- FOXA2 is required for H3K4me1 deposition before enhancer activation

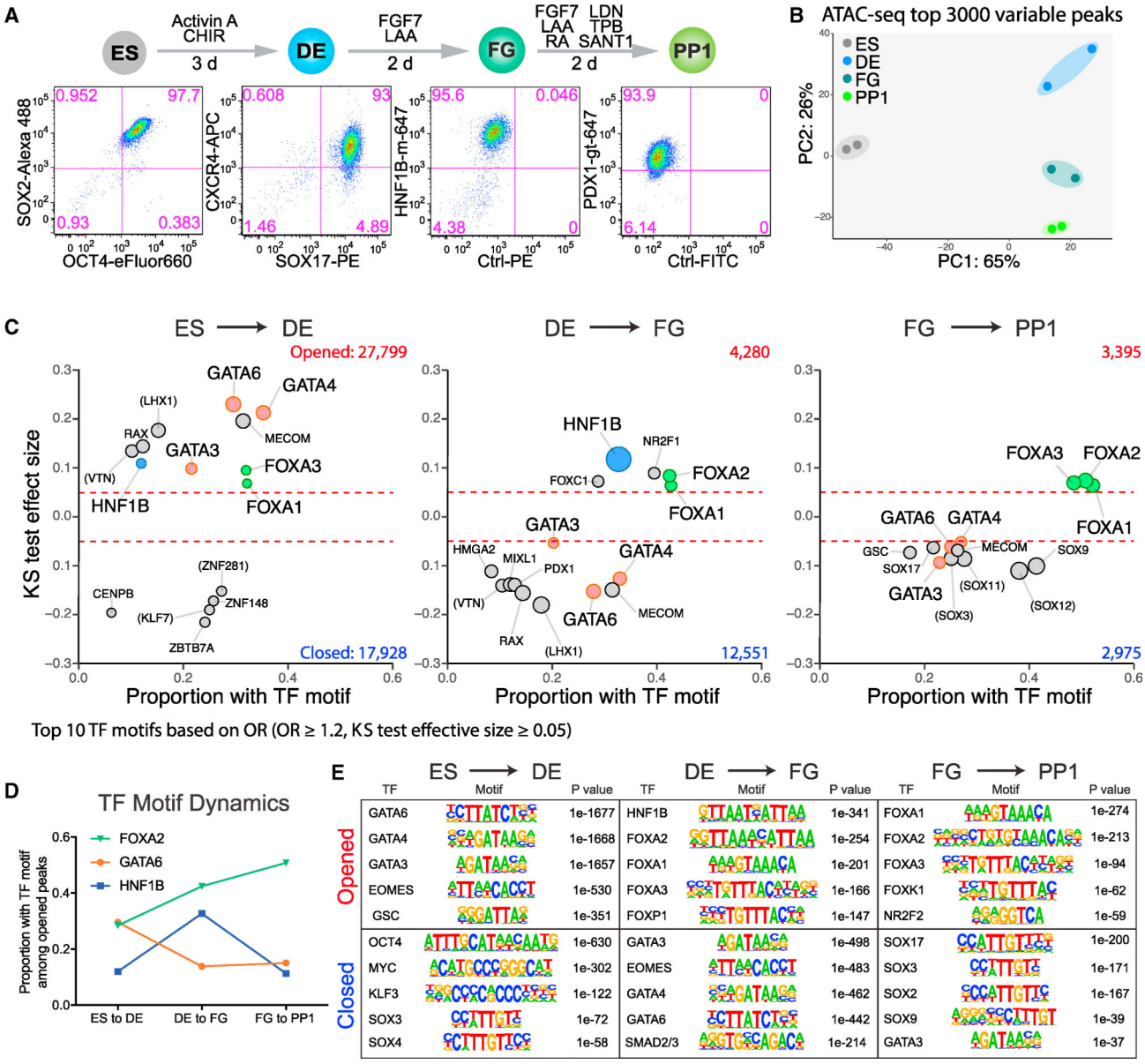


Figure 1. Different Stages during Pancreatic Differentiation Are Associated with Distinct TF Motifs

(A) Pancreatic differentiation efficiency was verified by examining lineage markers using flow cytometry. ES, embryonic stem cells; DE, definitive endoderm; FG, posterior foregut; PP1, primary pancreatic progenitor; d, day(s). See STAR Methods for a detailed differentiation protocol.

(B) Principal-component analysis (PCA) of ATAC sequencing (ATAC-seq) for the top 3,000 variable peaks.

(C) TF motif enrichment during stage transitions. The numbers of opened and closed peaks during stage transitions are indicated. Opened or closed peaks between successive stages were compared with the total atlas to examine the TF motif enrichments using the one-sided Kolmogorov-Smirnov (KS) test. The KS test effect size is shown on the y axis, and the proportion of peaks associated with the TF motif is plotted on the x axis. The size of each

circle represents the odds ratio (OR), which was defined as the frequency of the TF in an opened or closed group divided by its frequency in the entire atlas. TF motifs with a KS test effect size ≥ 0.05 (indicated by the dashed lines) and OR ≥ 1.2 are shown; when there were more than 10 such motifs, only the top 10 are shown. TF motifs derived from species other than *Homo sapiens* are marked with brackets. GATA motifs (GATA2, GATA3, GATA4, and GATA6) are marked in red, the HNF1B motif is in blue, and FOXA motifs (FOXA1, FOXA2, and FOXA3) are in green. See also Table S1.

(D) Proportion of ATAC-seq peaks with the GATA6, FOXA2, and HNF1B motif are shown.

(E) HOMER motif analysis was performed during the stage transition. The top 5 known motifs are shown after removing redundant motifs.

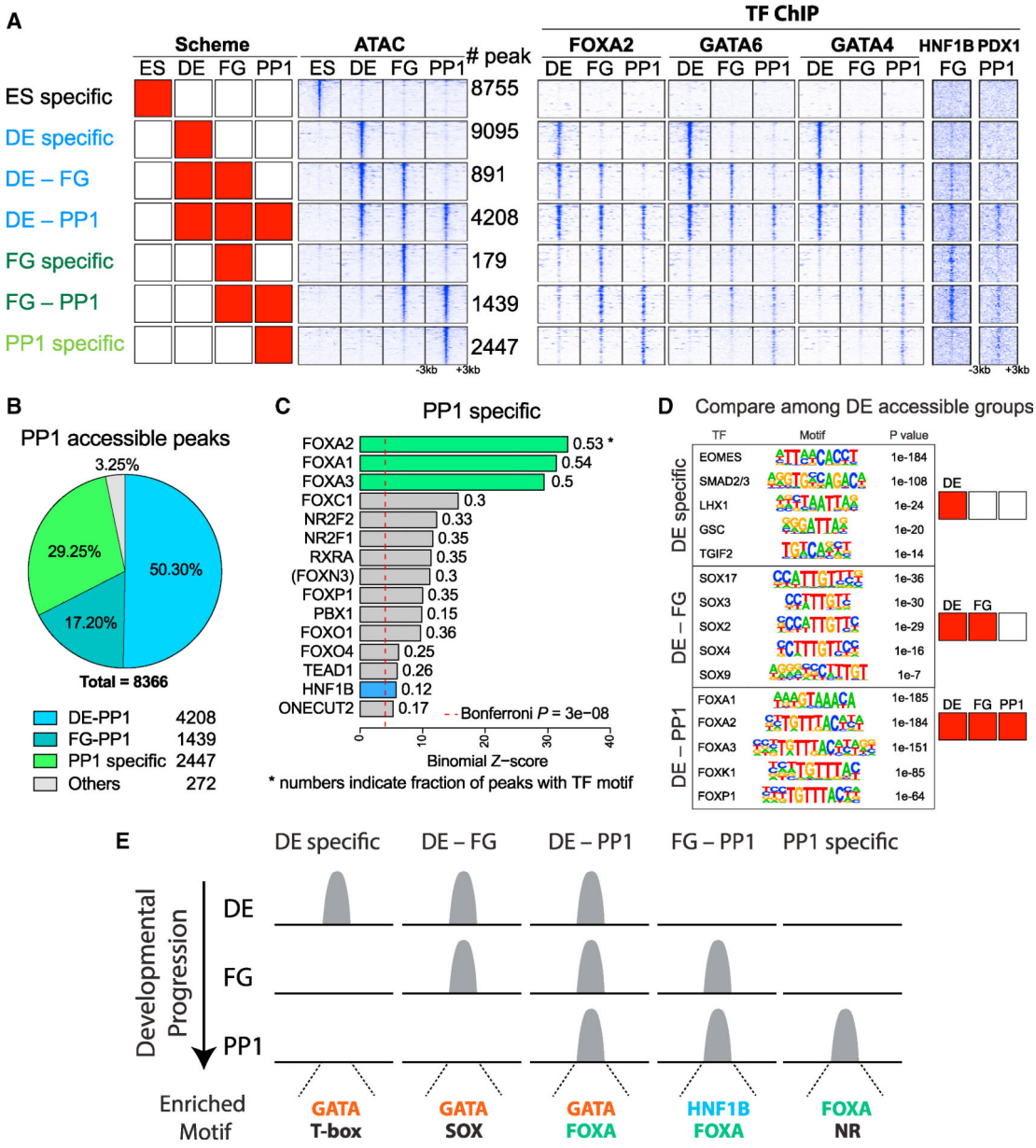


Figure 2. Pancreatic Progenitor Stage-Specific Accessible Chromatin Regions Are Enriched with FOXA Motifs

(A) ATAC-seq stage-specific groups; the chromatin accessibility was compared among 4 stages (ES, DE, FG, and PP1), and stage-specific groups were defined based on their patterns of chromatin status across stages. The number of peaks for each group are annotated. ChIP-seq profiles are shown in the same order as ATAC-seq peaks. See also Table S2.

(B) Composition of accessible peaks at the PP1 stage. The number of peaks for each group is annotated.

(C) TF motif enrichment in the PP1-specific group. The hypergeometric test was used to compare the enrichment of proportions of TF motifs for the PP1-specific group (foreground ratio) versus those for the total atlas (background ratio). The number of peaks containing TF

motifs in the PP1-specific group and the total atlas, as well as the size of the PP1-specific group and the total atlas, are supplied. The 15 most highly enriched TFs are shown with the foreground ratio to the right; the horizontal axis shows the binomial *Z* score, representing the number of SDs between the observed count of PP1-specific peaks containing a TF motif and the expected count based on the background ratio. The Bonferroni-corrected hypergeometric p value of the bottom-most TF is shown with the red vertical line. TF motifs derived from species other than *Homo sapiens* are marked with brackets. GATA motifs (GATA2, GATA3, GATA4, and GATA6) are marked in red, the HNF1B motif is in blue, and FOXA motifs (FOXA1, FOXA2, and FOXA3) are in green. See Figure S1B and Table S3 regarding other stage-specific groups.

(D) TF motif enrichment results using the HOMER algorithm among DE-containing groups; DE specific, DE-FG, and DE-PP1. The top 5 known motifs are shown after removing redundant motifs.

(E) Summary of motif enrichment for each stage-specific group.

See also Figure S1.

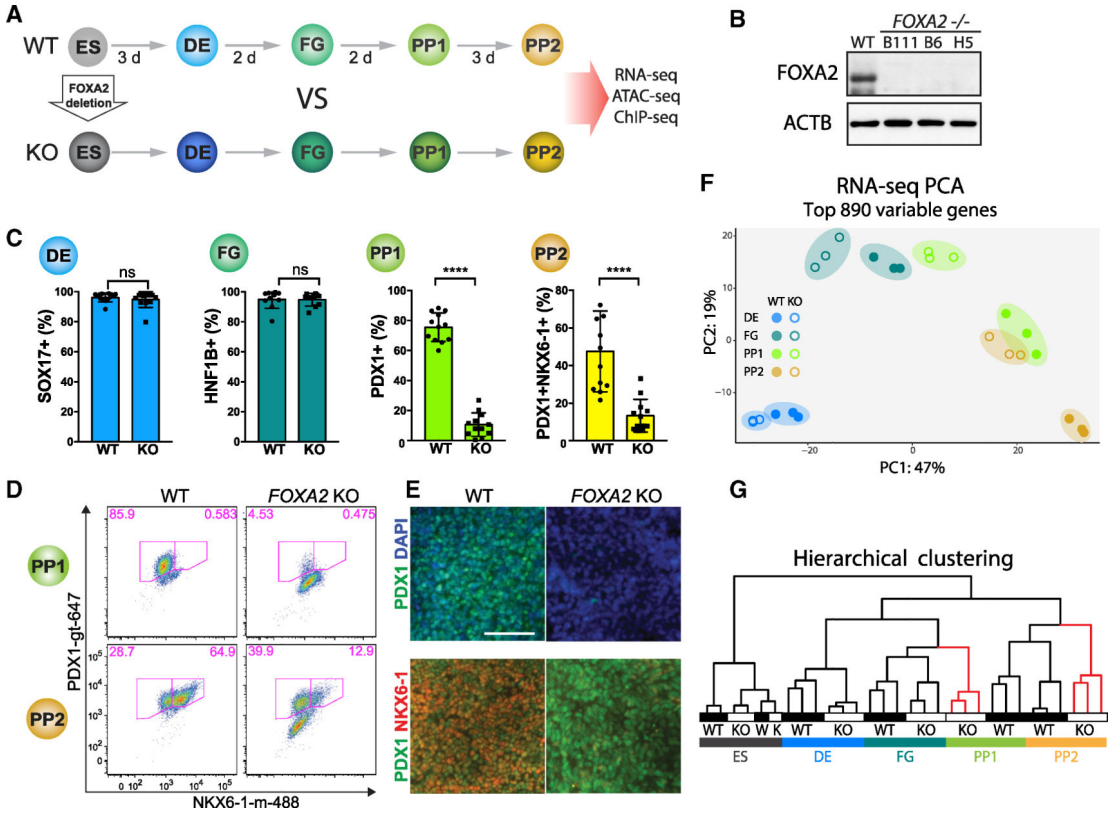


Figure 3. Requirements for *FOXA2* in Pancreatic Progenitor Cell Specification

(A) HUES8 *FOXA2* KO lines were generated at the ES stage and analyzed by RNA-seq, ATAC-seq, and ChIP-seq throughout pancreatic differentiation. d, day(s). See also Table S4.

(B) Loss of protein from HUES8 *FOXA2* KO cells was verified by western blotting at the FG stage.

(C) Expression of stage-specific lineage markers in HUES8 *FOXA2* KO was examined by flow cytometry. SOX17 expression at the DE stage, HNF1B expression at the FG stage, and PDX1 and NKX6–1 expression at the PP1 and PP2 stages are shown. The differentiation experiments were repeated four times with three independent clonal lines per genotype. Results from clonal lines of the same genotype were combined (n = 12), and results from each experiment are shown, together with mean ± SD. Statistical analysis was performed by unpaired two-tailed Student's t test. $^{****}p < 0.0001$; ns, not significant ($p \geq 0.05$). Results from individual lines are plotted in Figure S2D.

(D) Representative plot of flow cytometry are shown for $PDX1^{+}$ $NKX6–1^{+}$ expressions in WT and *FOXA2* KO at the PP1 and PP2 stages.

(E) $PDX1^{+}$ $NKX6–1^{+}$ expressions at the PP1 and PP2 stages were examined by immunostaining. Scale bar indicates 100 μm.

(F) Principal component analysis (PCA) of the top 890 variable genes between WT and *FOXA2* KO when controlling for the stage effect by two-factor modeling. n = 3, independent clonal lines. See also Table S5.

(G) Hierarchical clustering of the top 890 variable genes from (F).

See also Figure S2.

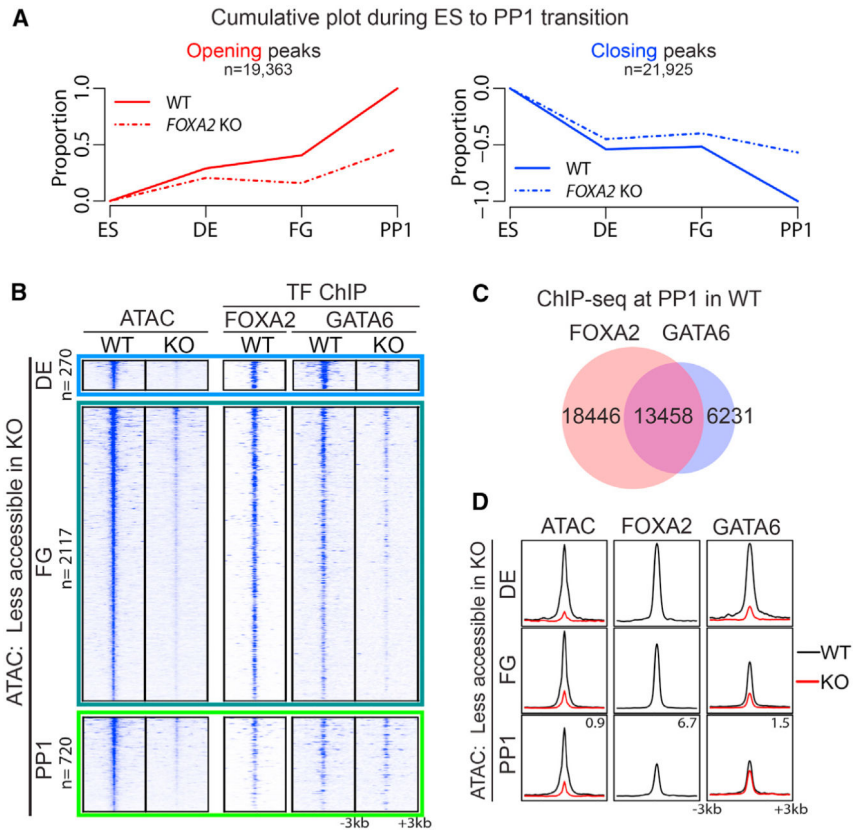


Figure 4. Pancreatic TF Bindings Are Enriched in Less Accessible Regions in *FOXA2* KO

(A) Cumulative plot of opening (19,363) and closing (21,925) peaks from the ES-to-PP1 transition.

(B) Tornado plot of less accessible peaks in *FOXA2* KO shown for the DE, FG, and PP1 stages. FOXA2 and GATA6 ChIP-seq in less accessible peaks is shown in the same order as ATAC-seq. See also Table S6.

(C) Venn diagram showing the overlap between FOXA2 and GATA6 binding sites at the PP1 stage in WT cells.

(D) Average of normalized counts visualized as a metapeak. Metapeaks show FOXA2 binding in WT cells and GATA6 binding in WT and *FOXA2* KO cells at less accessible peaks. The maximum value of each y axis is annotated in TPM.

See also Figures S3 and S4.

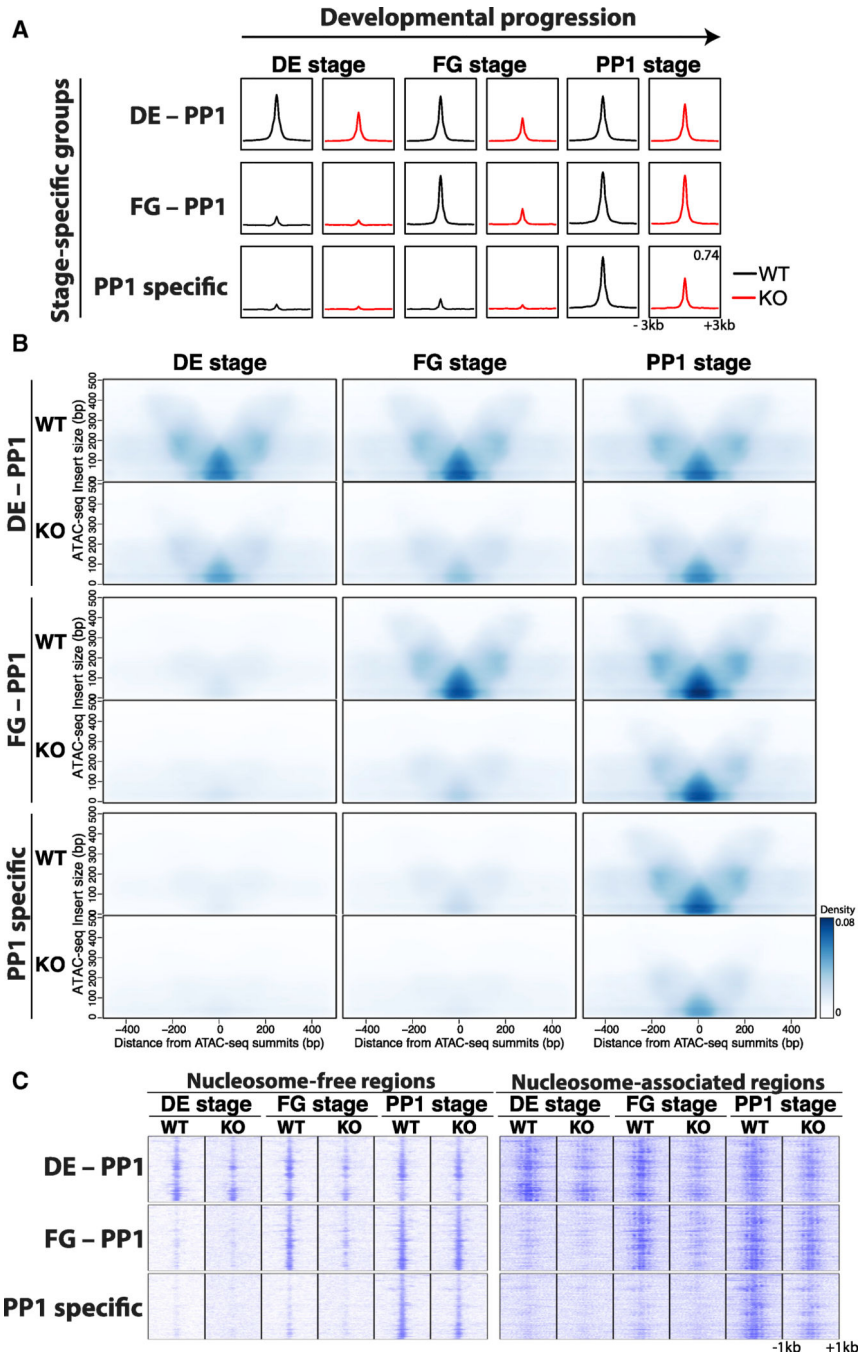


Figure 5. FOXA2 Is Required to Acquire Proper Chromatin Accessibility during Pancreatic Differentiation

(A) *FOXA2* KO effects on ATAC-seq stage-specific groups are shown. The average of normalized counts is visualized as a metapeak, and the maximum value of y axis is indicated in the plot as TPM. The WT signal is marked in black and the *FOXA2* KO signal is in red in the tornado plots.

(B) ATAC-seq insert size of WT and *FOXA2* KO for DE-PP1-, FG-PP1-, and PP1-specific groups is visualized. The density of insert size is shown by aggregating based on the distance from the ATAC-seq peak summit.

(C) ATAC-seq read pairs are visualized as tornado plots, divided by the insert size: nucleosome-free regions < 150 bp and nucleosome-associated regions ≥ 150 bp. See also Figure S5.

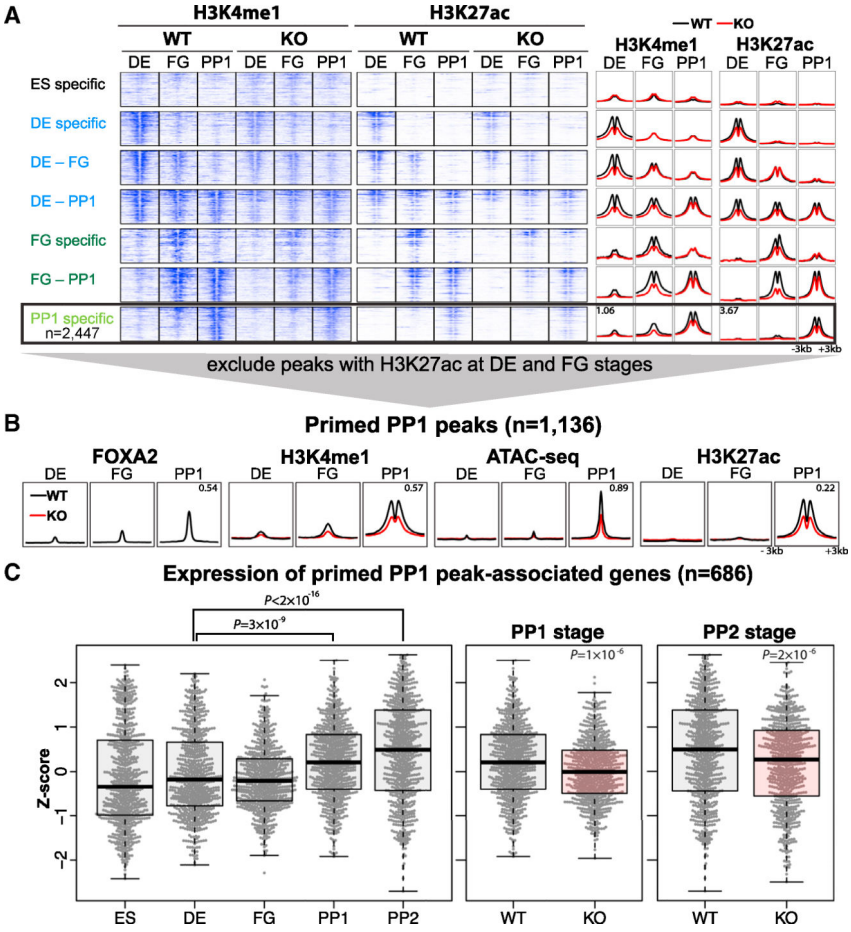


Figure 6. H3K4me1 Deposition during Pancreatic Differentiation

(A) Tornado plot and metapeak of the ATAC-seq stage-specific group show the profile of H3K4me1 and H3K27ac histone modifications in WT and *FOXA2* KO cells. The maximum value (TPM) of the y axis is indicated in the plot. The WT signal is marked in black and the *FOXA2* KO signal is in red in the tornado plots.

(B) Primed PP1-specific ATAC-seq group was defined by excluding peaks that contained the H3K27ac signal at the DE or the FG stage (n = 1,136). See also Table S7. Metapeak analysis shows FOXA2, H3K4me1, ATAC-seq, and the H3K27ac signal of the primed PP1-specific group. WT is marked in black, and *FOXA2* KO is in red. The maximum value (TPM) of the y axis is annotated in the plot.

(C) Beeswarm plot to show the expression of 686 genes associated with the primed ATAC-seq group. *Z* scores of normalized counts are plotted, with each cell type averaged over replicates. Five WT samples, PP1 samples, and PP2 samples are grouped by themselves for comparison. See Figure S6C for the beeswarm plot of all cell types. The graph shows p values for significant increases in gene expression levels when comparing PP1 versus DE and PP2 versus DE using one-sided unpaired Wilcoxon rank-sum tests. There are also significant increases when comparing PP1 versus FG ($p = 4.639 \times 10^{-18}$), PP2 versus FG ($p = 1.603 \times 10^{-27}$), and PP2 versus PP1 ($p = 6.329 \times 10^{-6}$).

See also Figure S6.

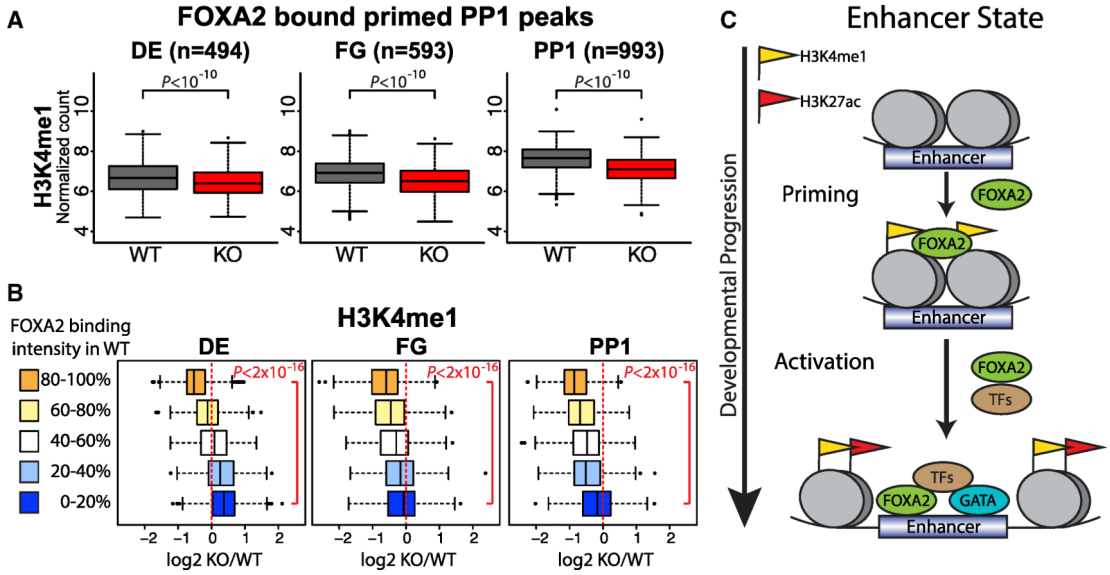


Figure 7. FOXA2 Is Required for the Establishment of Primed Enhancer Mark H3K4me1 Deposition

(A) Comparison of normalized levels of H3K4me1 at the primed PP1-specific groups. FOXA2 binding regions were defined by the location of reproducible FOXA2 ChIP-seq peaks (irreproducible discovery rate [IDR] ≤ 0.01). The values of histone modification in WT and *FOXA2* KO cells were compared by a one-sided unpaired Wilcoxon signed rank test with Bonferroni correction. The center line in the boxplots shows the median; the box limits are the upper and the lower quartiles, respectively; and whiskers are defined as 1.5 times the interquartile range above and below the box limits. All outliers outside the whiskers are shown as points.

(B) Fold change of the H3K4me1 signal in *FOXA2* KO compared with WT for the primed PP1-specific group. Peaks were grouped by signal strength quintile of the reproducible FOXA2 ChIP-seq peaks at each stage. At all 3 stages, stronger FOXA2 binding showed a more negative shift in the H3K4me1 signal. The H3K4me1 signals at the top 20% and bottom 20% of FOXA2 peaks were compared at each stage using a one-sided unpaired Wilcoxon signed rank test (p values as annotated).

(C) Model of the role of FOXA2 in enhancer priming during pancreatic differentiation. During enhancer priming, FOXA2 recruitment facilitates H3K4me1 deposition. This is followed by recruitment of additional lineage TFs, such as GATA factors (GATA6 and GATA4), and acquisition of the H3K27ac activation mark.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
OCT3/4-eFluor660	eBioscience	50–5841-82; RRID: AB_11218890
CXCR4-APC	R&D Systems	FAB170A: RRID: AB_357073
SOX2 Alexa Fluor 488	eBioscience	53–9811-82; RRID: AB_2574479
SOX2-PE	BioLegend	656104; RRID: AB_2562853
SOX17-PE	BD Biosciences	561591; RRID: AB_10717121
HNF1B	Santa Cruz	sc-130407; RRID: AB_2248215
FOXA2 Alexa Fluor 488	R&D Systems	IC2400G; RRID: AB_2801552
PDX1	R&D Systems	AF2419; RRID: AB_355257
NKX6.1	DSHB	F55A12-c; RRID: AB_532379
FOXA2	R&D Systems	AF2400; RRID: AB_2294104
FOXA2	Cell Signaling Technology	3143S; RRID: AB_2104878
FOXA1	Cell Signaling Technology	58613S; RRID: AB_2799548
ACTB	Cell Signaling Technology	3700S; RRID: AB_2242334
H3K4me1	Abcam	ab8895; RRID: AB_306847
H3K27ac	Active Motif	39133; RRID: AB_2561016
HNF1B	Santa Cruz	sc-22840x; RRID: AB_2279595
GATA4	R&D Systems	AF2606; RRID: AB_2232177
GATA6	Cell Signaling Technology	5851S; RRID: AB_10705521
Chemicals, Peptides, and Recombinant Proteins		
Complete Essential 8 (E8) medium	Thermo Fisher Scientific	A1517001
Vitronectin (VTN-N)	Thermo Fisher Scientific	A14700
EDTA	KD Medical	RGE-3130
TrypLE Select 1X	Thermo Fisher Scientific	12563–029
Y-27632 dihydrochloride	Selleck Chemicals	S1049
Doxycycline	Sigma	D9891
Lipofectamine RNAiMAX Transfection Reagent	Thermo Fisher Scientific	13778150
Opti-MEM	Thermo Fisher Scientific	31985070
Proteinase K	Fisher Scientific	BP1700–100
10X PCR buffer	Sigma	P2192
BIO-acetoxime	Tocris	3874
Advanced RPMI	Thermo Fisher Scientific	12633020
MCDB 131	Thermo Fisher Scientific	10372019
Activin A	PeproTech	120–14E
CHIR-99021	Stemgent	04–0004
L-Ascorbic acid (vitamin C)	Sigma-Aldrich	A4544

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FGF7 (KGF)	PeproTech	100–19
SANT-1	Sigma	S4572
Retinoic acid	Sigma	R2625
LDN-193189	Reprocell	04–0074
TPB	EMD Millipore	565740
ITS-X (100X)	Thermo Fisher Scientific	51500–056
TRIzol	Thermo Fisher Scientific	15596018
Critical Commercial Assays		
Herculase II Fusion DNA Polymerase	Agilent Technologies	600679
DNeasy kit	QIAGEN	69506
miRNeasy Mini Kit	QIAGEN	217004
Deposited Data		
Raw and processed data	This paper	GSE114102
Experimental Models: Cell Lines		
HUES8 hESC line	HSCI iPS Core, Harvard University	NIHhESC-09–0021; RRID: CVCL_B207
H1 hESC line	WiCell Research Institute	NIHhESC-10–0043; RRID: CVCL_9771
HUES8 iCas9 line	González et al., 2014	Derived from the parental line (NIHhESC-09–0021); RRID: CVCL_VR10
H1 iCas9 line	Shi et al., 2017	Derived from the parental line (NIHhESC-10–0043); RRID: CVCL_WS14
Sequence-Based Reagents		
See Table S4 for CRISPR targeting related sequences, and primer sequences for PCR genotyping and RT-qPCR.	This paper	N/A
Software and Algorithms		
PRISM v6 graphing and statistical software	GraphPad Software	https://www.graphpad.com/; RRID: SCR_002798
Flowjo-v9	FlowJo LLC	https://www.flowjo.com/; RRID: SCR_008520
FastQC	v0.11.4	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/; RRID: SCR_014583
Trimmomatic	v0.36	http://www.usadellab.org/cms/?page=trimmomatic; RRID: SCR_011848
STAR	v2.5.3a	https://github.com/alexdobin/STAR; RRID: SCR_015899
HTSeq	v0.8.0	https://github.com/simon-anders/htseq; RRID: SCR_005514
R	v3.5.0	https://cran.r-project.org; RRID: SCR_003005
DESeq2	v1.16.1	https://bioconductor.org/packages/release/bioc/html/DESeq2.html; RRID: SCR_015687
pheatmap	v1.0.8	https://github.com/raivokolde/pheatmap; RRID: SCR_016418
Bowtie 2	v2.3.3.1	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml; RRID: SCR_005476
MACS2	v2.1.1.20160309	https://github.com/taoliu/MACS; RRID: SCR_013291
IDR	v2.0.3	https://github.com/nboley/idr; RRID: SCR_017237
HOMER	v4.9	http://homer.ucsd.edu/homer/index.html; RRID: SCR_010881
CIS-BP	v1.0.2	http://cisbp.ccbr.utoronto.ca; RRID: SCR_017236

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MEME suite	v4.12.0	http://meme-suite.org; RRID: SCR_001783
Other		
Raw data	Li et al., 2019	GSE109524 (GSM3381978 and GSM3381980)