Published in final edited form as:
Cell Rep. 2019 April 16; 27(3): 708–718.e10. doi:10.1016/j.celrep.2019.03.076.

Single-Cell RNA-Sequencing-Based CRISPRi Screening Resolves Molecular Drivers of Early Human Endoderm Development

Ryan M.J. Genga[1,3], **Eric M. Kernfeld**[1,3], **Krishna M. Parsi**[1], **Teagan J. Parsons**[1], **Michael J. Ziller**[2], and **René Maehr**[1,4,*]

[1]Program in Molecular Medicine, Diabetes Center of Excellence, University of Massachusetts Medical School, Worcester, MA 01605, USA

[2]Department of Translational Psychiatry, Max Planck Institute of Psychiatry, 80804 Munich, Germany

[3]These authors contributed equally

[4]Lead Contact

SUMMARY

Studies in vertebrates have outlined conserved molecular control of definitive endoderm (END) development. However, recent work also shows that key molecular aspects of human END regulation differ even from rodents. Differentiation of human embryonic stem cells (ESCs) to END offers a tractable system to study the molecular basis of normal and defective human-specific END development. Here, we interrogated dynamics in chromatin accessibility during differentiation of ESCs to END, predicting DNA-binding proteins that may drive this cell fate transition. We then combined single-cell RNA-seq with parallel CRISPR perturbations to comprehensively define the loss-of-function phenotype of those factors in END development. Following a few candidates, we revealed distinct impairments in the differentiation trajectories for mediators of TGFβ signaling and expose a role for the *FOXA2* transcription factor in priming human END competence for human fore-gut and hepatic END specification. Together, this single-cell functional genomics study provides high-resolution insight on human END development.

Graphical Abstract

[*]Correspondence: rene.maehr@umassmed.edu.

AUTHOR CONTRIBUTIONS

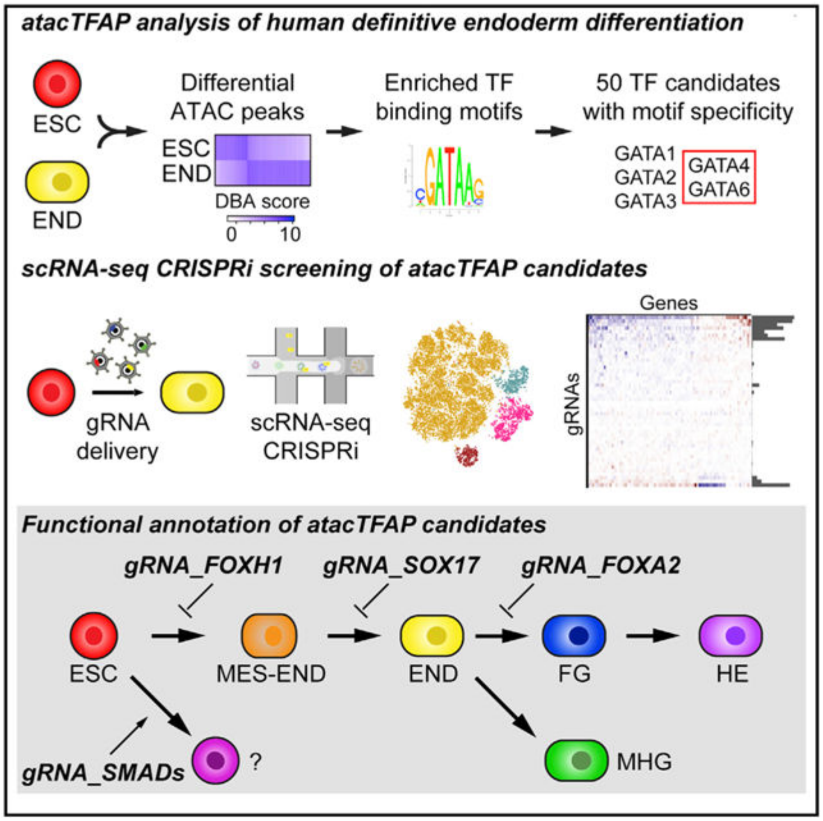
R.M. designed the study and performed data analysis. R.M.J.G. generated RNA-seq, ATAC-seq, and scRNA-seq datasets and performed the molecular biology, functional follow-up, and data analysis. E.M.K. performed analysis of scRNA-seq CRISPRi and Quant-seq datasets. K.M.P. generated the H1-AAVS1-TetOn-dCas9-KRAB hESCs and generated ATAC-seq and ChIP-seq datasets. T.J.P. performed gRNA cloning and generated lentivirus. M.J.Z. performed chromatin, RNA-seq, and atacTFAP analysis and designed its data analysis strategies. R.M.J.G., E.M.K., and R.M. wrote the manuscript with input from all authors.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.celrep.2019.03.076.

DECLARATION OF INTERESTS

The authors declare no competing interests.


atacTFAP analysis of human definitive endoderm differentiation
ESC
END
Differential ATAC peaks
ESC
END
DBA score
0
10
Enriched TF binding motifs
50 TF candidates with motif specificity
GATA1
GATA2
GATA3
GATA4
GATA6
scRNA-seq CRISPRi screening of atacTFAP candidates
gRNA delivery
scRNA-seq CRISPRi
Genes
gRNAs
Functional annotation of atacTFAP candidates
gRNA_FOXH1
gRNA_SOX17
gRNA_FOXA2
ESC
MES-END
END
FG
HE
gRNA_SMADs
?
MHG

In Brief

Genga et al. utilize a single-cell RNA-sequencing-based CRISPR interference approach to screen transcription factors predicted to have a role in human definitive endoderm differentiation. The perturbation screen identifies an important role of TGFβ signaling-related factors. Follow-up of *FOXA2* reveals genome-wide molecular changes and altered differentiation competency in endoderm.

INTRODUCTION

Human embryonic stem cell (ESC) differentiation strategies to generate definitive endoderm (END) allow for interrogation of differentiation-associated signaling requirements and chromatin states (D’Amour et al., 2005; Gifford et al., 2013; Loh et al., 2014). While various transcription factors (TFs) have been evaluated for their role in vertebrate END formation (Zorn and Wells, 2009), there are notable species differences in TF requirements (Shi et al., 2017; Tiyaboonchai et al., 2017; Zhu and Huangfu, 2013). For example, recent loss-of-function analyses revealed important roles of TFs, including *GATA6* and *KLF8*, specifically in human END (Allison et al., 2018; Chu et al., 2016; Tiyaboonchai et al., 2017), highlighting the need for increased throughput in functional analyses of TF dependencies in human.

CRISPR interference (CRISPRi) systems can effectively disrupt gene function in human pluripotent stem cells (Kearns et al., 2014; Mandegar et al., 2016), with low off-target effects

(Gilbert et al., 2014). CRISPRi can also be combined with droplet-based, single-cell RNA-sequencing (scRNA-seq) read-outs (Adamson et al., 2016; Xie et al., 2017), allowing functional analysis of molecular pathways guiding differentiation while balancing resolution and throughput. Here, we predicted candidate molecular drivers of END differentiation (END-Diff) by computationally integrating dynamics of chromatin accessibility and transcriptome-wide changes. To delineate candidate roles, we conducted a parallel scRNA-seq CRISPRi screen to perturb the factors during END-Diff. We uncover distinct blocks in early human END development mediated by loss of TFs involved in transforming growth factor b (TGFβ) signaling, while perturbation of the TF *FOXA2* results in an altered differentiation competency at later stages.

RESULTS

Chromatin Accessibility and Transcriptome Dynamics of END-Diff

Using an efficient ESC differentiation platform (Figures S1A and S1B), we compare ESC and END by RNA-seq and assay for transposase-accessible chromatin using sequencing (ATAC-seq) (Figure 1A) revealing 2,905 differentially expressed transcripts (Figure S1C; Table S2; false-discovery rate [FDR] <0.01; log fold change ≥ 1.0) and differential chromatin accessibility at 34,025 sites (Figures 1B and S1D; Table S2; FDR < 0.05; log fold change ≥ 1.0), respectively. Analysis by ATAC-seq transcription factor activity prediction (atacTFAP) of ESC, END, and pancreatic beta cells was applied to reveal putative molecular drivers of END-Diff. While many of the predicted DNA-binding proteins have been associated with mesendoderm and END formation (e.g., *GATA4*, *GATA6*, *GSC*, *SOX17*, *FOXH1*, *FOXA2*, *OTX2*, *EOMES*, *SMAD2*, *SMAD4*, and *MIXL1*) (Zorn and Wells, 2009), other candidates have not been directly implicated in early END-Diff (e.g., *ZBTB33*, *ESRRA*, *ZNF410*, and *E2F6*). In total, 50 TF candidates were selected for functional follow-up, covering a spectrum of potential facilitators and repressors of human END-Diff. Among the 50 candidates, two factors were included that are implicated in structural aspects of chromatin organization (*CTCF* and *YY1*) with described roles in the exit of pluripotency (Balakrishnan et al., 2012; Weintraub et al., 2017).

Single-Cell CRISPRi Screening Reveals Candidate Regulators of END-Diff

We utilized a lentiviral guide RNA (gRNA) delivery system (Datlinger et al., 2017) together with a gene-targeted H1-*AAVS1*-TetOn-dCas9-KRAB ESC line (Figures S1E–S1H) to assay the transcriptomic effects of atacTFAP candidate repression on pooled human END-Diff at single-cell resolution while simultaneously identifying the gRNA delivered to each cell. Analysis was performed at the END time point (n = 2 biological replicates) via droplet-based capture and profiling of individual cells by scRNA-seq (Figure 1D).

After cell filtering and quality control (Figures S2A–S2J; Table S3), unsupervised analysis yields four clusters as visualized by t-distributed stochastic neighbor embedding (tSNE) (Figure 1E). The three smaller clusters contain 1,627 (cluster 1), 1,015 (cluster 2), and 714 (cluster 3) cells. The largest cluster (cluster 0) contains 12,754 cells and captures most control scramble gRNAs (Figure 1F; $p < 2.2E-16$). Ranking of transcripts differentially expressed between clusters (Figure 1G; $q < 0.05$, fold change [FC] > 1.5) places END-associated transcripts *LEFTY1*, *LEFTY2*, *CXCR4*, and *SOX17* in the top 25 transcripts for cluster 0 (Table S3).

Cluster characterization via Enrichr (Chen et al., 2013; Kuleshov et al., 2016) links clusters 0 and 3 to END formation, cluster 1 to SOX2 and NANOG binding, and cluster 2 to FOSL2 binding. The END-associated transcript *SOX17* is expressed in clusters 0 and 2, while the pluripotency-associated transcript *POU5F1* is expressed mostly in cells of cluster 1 (Figure 1H). *MIXL1* is expressed in all clusters except cluster 2, and the BMP target gene, *ID1*, is highly expressed in cluster 2 (Figure 1H). The END annotation in cluster 3 is driven by mesendodermal genes *MIXL1*, *LHX1*, and *NODAL* rather than END hallmarks such as *SOX17*. This analysis suggests that the four clusters represent entirely different cellular states, rather than subtle transcriptomic differences. We predicted these differences were driven by effects of atacTFAP candidate perturbation.

Targeting of the TGFβ Pathway Affects Differentiation in a Target-Specific Manner

Since analysis of the scRNA-seq CRISPRi END libraries reveals distinct clusters, we sought to identify the gRNA-targeted TFs driving the transcriptomic changes. Further investigation of clusters 1, 2, and 3 reveals significant enrichment of gRNAs specific to TFs known to function within the TGFβ signaling pathway, including *FOXH1*, *SMAD2*, and *SMAD4* (Massagué, 2012), or to be regulated by TGFβ signaling (*SOX17*) (Alexander and Stainier, 1999) (Figures 2A and 2B; Table S3). *FOXH1*-specific gRNAs are significantly enriched in cluster 1, *SMAD2*- and *SMAD4*-specific gRNAs are significantly enriched in cluster 2, and *SOX17*-specific gRNAs are enriched in cluster 3 (Figure 2B; Table S3).

To see whether any cluster could be interpreted as a differentiation block, we characterized a bulk RNA-seq END-Diff time course of control ESCs (Figure 2C; transcript list curated from literature; Chu et al., 2016; Loh et al., 2014). Expression of pluripotency-associated transcripts (e.g., *POU5F1*, *SOX2*, *NANOG*) decreases over time, while expression of END-associated transcripts increases (e.g., *SOX17*, *FOXA2*, *GATA4*, *GATA6*). Expression of mesendoderm-associated transcripts (e.g., *MIXL1*, *EOMES*, *T*) is high at day 1 of differentiation and decreases or is maintained through differentiation (Figure 2C).

Cluster 0 is enriched for scramble-gRNAs and expresses high levels of known END transcripts, including *SOX17* (Figures 1H and 2C). The expression of pluripotency markers is low and overall gene expression is similar to day 3 of the time course (Figure 2C). Cluster 1 expresses the highest levels of ESC markers, including *POU5F1* (Figures 1H and 2C). Together with low expression of END markers, cluster 1 is most comparable to day 0 of the time course (Figure 2C). In cluster 2, there is relatively low expression of both ESC markers and mesendoderm markers (Figure 2C). Some END markers are expressed, including *SOX17* and *FOXA2*, while others are low, such as *GATA6*, *NODAL*, and *HHEX* (Figure 2C). Notably, expression of *ID1* is high in cluster 2 (Figures 1H and 2C), consistent with the finding that modulation of TGFβ signaling through *Smad2/3* gene knockouts in mouse ESCs results in increased *ID* transcript expression (Senft et al., 2018). Overall gene expression of cluster 2 is not comparable to any day within the time course. Cluster 3 expresses the highest levels of the mesendoderm markers *MIXL1*, *EOMES*, and *T*, while expression of END markers is either low (e.g., *FOXA2*, *CXCR4*, *GATA4*) or high (e.g., *GATA6*, *NODAL*, *HHEX*; Figure 2C). Comparison with the time course suggests that *SOX17* repression captures cells in a mesendoderm-like state (Figure 2C).

Concordant with a critical role of TGFβ in pluripotency and END-Diff (Avery et al., 2010; Sakaki-Yumoto et al., 2013; Shen, 2007; Wei and Wang, 2018), analysis of clusters 1, 2, and 3 reveals significant enrichment of perturbations linked to the TGFβ signaling pathway. Repression of *FOXH1* and *SOX17* halts differentiation at the pluripotent-or the mesendoderm-like state, respectively, and targeting of *SMAD2* and *SMAD4* most likely results in an alternative cellular state that is neither ESC-like or END-like, supporting a model of target-specific differentiation blocks rather than a generic block of END-Diff (Figure 2D).

Distinct Transcriptomic Signatures Are Observed among END-Like Cells

The majority of candidate-specific gRNA-containing cells reside in cluster 0, which contains most of the scramble-gRNA-expressing cells and exhibits an overall END-like expression signature (Figures 1F, 2B, and 2C). To identify more subtle effects within cluster 0, we employed the MIMOSCA modeling framework (Dixit et al., 2016), extracting common effects from its output via sparse principal-component analysis (PCA) (Zou et al., 2006). This approach reveals a gene set affected across several screen candidates (Figure 3A). Although sparse PCA does not involve manual selection of genes, it highlights key transcripts of the transition from pluripotency to END, including *LEFTY1*, *LEFTY2*, *FOXA2*, *CXCR4*, and *POU5F1*. As the permutation scheme of MIMOSCA does not account for batch effects within replicates, we also formally tested all genes against all perturbations via MAST (Finak et al., 2015). Cross-referencing the results shows that most major perturbations (>30 genes; FDR < 0.05) have directional effects on the END development gene set, with *CTCF* as the only exception. This demonstrates that atacTFAP analysis enriches for candidates important for END development. Furthermore, repression of TF candidates can lead to END-like states that are characterized by subtle changes in developmentally relevant transcripts.

Within cluster 0 only, targeting of *FOXA2* results in the largest change, affecting 266 transcripts (Figure 3A). When *FOXA2*-gRNA cells were differentiated independently in comparison to scramble-gRNA cells, SOX17 expression levels were similar (Figures 3B and 3C), suggesting differentiation of END despite loss of FOXA2 in most cells. However, decreases in other END markers, including *CXCR4*, *HHEX*, *LEFTY1*, *LEFTY2*, *MIXL1*, *CER1*, *GSC*, and *OTX2* (Table S4), revealed that the END state might differ. Since *FOXA2* targeting had the largest effect size together with a seemingly preserved END state, we selected this candidate for follow-up.

Loss of *FOXA2* Results in Genome-Wide Chromatin Changes in Human END

To better understand the genome-wide role of *FOXA2* in human END, we generated FOXA2-chromatin immunoprecipitation sequencing (ChIP-seq) data in END, as well as H3K27ac ChIP-seq and ATAC-seq in ESCs, scramble-gRNA END, and *FOXA2*-gRNA END (Figure 3D; Table S4). To gain insight into the developmental role of *FOXA2*, we focused on peaks that were significantly altered in both END-Diff and *FOXA2* knockdown (n = 1,814 for H3K27ac and n = 1,277 for ATAC-seq; Figure 3D). We clustered peaks and used genomic regions enrichment of annotations tool (GREAT) (McLean et al., 2010) to suggest possible functions (Table S4). GREAT analysis of H3K27ac clusters 2 and 4, which increase during normal differentiation and decrease upon *FOXA2* knockdown, revealed significant annotations for foregut epithelial, pancreatic, and hindgut development (Table S4), suggesting that regions activated by *FOXA2* are relevant to END derivatives. GREAT analysis of all dynamic and *FOXA2*-dependent ATAC-seq peaks (Figure 3D, right) show they are disproportionately near genes related to mouse pancreas hypoplasia and lung epithelium differentiation. They are also close to genes expressed in mouse END derivatives, including pharynx, thyroid, lung, trachea, esophagus, liver, midgut, and hindgut. Individual clusters display much lower enrichment signal, but when the analysis is restricted to regions open in control END (clusters 1 and 2), most END annotations persist, consistent with a model in which *FOXA2* establishes chromatin states relevant to differentiation toward more mature lineages.

When surveying foregut-associated loci, we noticed loss of H3K27ac and ATAC-seq signal at *TTR* and *HHEX* when comparing *FOXA2*-versus scramble-gRNA END (Figure 3E). *HHEX* also gained H3K27me3 signal in the *FOXA2* knockdown condition. Interestingly, the putative regulatory region that is impacted by *FOXA2* repression in the human *HHEX* locus overlaps with a conserved foregut-specific *cis*-regulatory element described in frogs (Rodriguez et al., 2001). Taken together, loss of *FOXA2* results in an altered chromatin landscape following END-Diff, and further analysis reveals that putative regulatory regions are altered at specific foregut-associated loci.

Loss of *FOXA2* Impairs Differentiation to Foregut END and Subsequent Hepatic END, while Mid-Hindgut END Differentiation Is Unaffected

Given the effects on chromatin state upon targeting *FOXA2*, we hypothesized that one role of *FOXA2* during human END-Diff is to shape the chromatin landscape to control differentiation competency. To test this, we adopted stepwise differentiation protocols to guide *FOXA2*-gRNA END to foregut END (Hannan et al., 2013) or mid-hindgut END (Múnera and Wells, 2017) in comparison to scramble-gRNA END (Figure 4A). Foregut transcript expression is lower following foregut END differentiation in the *FOXA2*-gRNA condition (Figure 4B). However, following mid-hindgut END differentiation, mid-hindgut transcript expression is comparable or higher in *FOXA2*-gRNA cells (Figure 4C), and both conditions result in generation of CDX2$^+$/SOX17$^+$ mid-hind-gut at high efficiency (Figure S3A). Taken together, these comparisons suggest that *FOXA2*-gRNA END may be less able to differentiate into foregut END while remaining competent to differentiate into mid-hindgut END.

To further test the ability of *FOXA2*-gRNA END cells to differentiate toward derivatives of the foregut, we extended the stepwise differentiation protocol to guide END through foregut END to hepatic END (Figure 4D) (Hannan et al., 2013). Liver progenitor transcripts are lower in the *FOXA2*-gRNA condition (Figure 4E). Furthermore, decreased expression of FOXA2, HNF4A, HHEX, and PROX1 was observed specifically in the *FOXA2*-gRNA condition (Figures 4F and S3B), substantiating a role for *FOXA2* in hepatic END differentiation. Although this suggests cells have differentiated to another END cell type, a panel of qPCR results for other END lineage markers shows little difference except for *SOX17* (Figure S3C).

To discern whether the observed effects are cell autonomous, we analyzed hepatic END that was generated in a coculture differentiation of *FOXA2*-gRNA and scramble-gRNA containing cells by scRNA-seq (Figure 4G). After initial quality control (Figures S3D–S3F), analysis yields three main clusters (Figure 4H). Clusters A (564 cells) and B (135 cells) express hepatic END transcripts (Hannan et al., 2013) (Figures 4I, 4J, and S3G; Table S5). In cluster C (204 cells), *FOXA2* expression is low and numerous liver-associated transcripts are downregulated (Figures 4I, 4J, and S3G; Table S5). Annotation of cluster C by Enrichr highlights similarity to a SOX17 overexpression signature, and concordant with qPCR data (Figure S3C), *SOX17* is more highly expressed (ln FC, 0.13; q < 10–7; Table S5). In cluster C, 92% of cells express *FOXA2*-gRNA, compared with 1.6% in cluster A and 4.4% in cluster B. Clusters A and B are enriched for scramble-gRNA (Figure 4K). Global gene expression analysis of *FOXA2*-gRNA cells reveals decreases in liver-associated transcripts in comparison to scramble-gRNA (Figure 4L; Table S5), supporting a role of *FOXA2* in early hepatic differentiation. These analyses suggest that altered END-like states could affect cellular competency for differentiation of END to more mature foregut cell types.

DISCUSSION

We surmised that perturbation effects can manifest in different ways: lethal effects would lower abundance; differentiation blocks or alteration of cell identity would give rise to outlying clusters; and smaller effects would appear in differential expression testing. Of the candidates we screened, the top 10 candidates by atacTFAP score include both *FOXA2* and *EOMES*. *FOXA2* repression has a large effect on overall expression, and *EOMES*-gRNA-containing cells are depleted in END. In addition, gRNAs targeting 5 of the top 10 atacTFAP candidates differ significantly (FDR < 0.1; differential cluster probability > 0.03) from the scramble gRNAs in terms of their cluster assignments (*SOX17*, *SMAD2*, *SMAD4*, *GATA6*, and *FOXH1*). This enrichment of END-relevant genes demonstrates that motif analysis of ATAC-seq data can predict master regulators of differentiation.

Furthermore, the differential cluster assignment results are consistent with studies in vertebrate models that demonstrate a critical role of TGFβ signaling during vertebrate END development (Chu et al., 2004; Heyer et al., 1999; Hoodless et al., 2001; Vincent et al., 2003; Yamamoto et al., 2001; Conlon et al., 1994; Kanai-Azuma et al., 2002; Tremblay et al., 2000) and destabilization of the pluripotent state by *SMAD2/4* downregulation in human cells (Avery et al., 2010; Sakaki-Yumoto et al., 2013). The use of scRNA-seq allows for thorough profiling of these knockdowns. As a technical note, future studies may achieve even higher resolution by using a higher proportion of non-targeting controls.

In addition to cluster assignment effects, we revealed another class of atacTFAP candidates. These caused more subtle transcriptomic alterations that likely would have remained undetected in screens based on currently available transgenic reporters of the END lineage, with an example being FOXA2. Study of *Foxa2* has been hampered by lethality due to defects in axial mesoderm development and gastrulation in mice (Ang and Rossant, 1994; Dufort et al., 1998; Weinstein et al., 1994), and conditional knockouts of *Foxa2* have focused on later stages of gut tube development, some of which required dual inactivation of *Foxa2* and *Foxa1* to manifest developmental phenotypes (Lee et al., 2005a, 2005b; Wan et al., 2005). Although FOXA2 associates with poised enhancers (Wang et al., 2015) and exhibits pioneer factor activity (Donaghey et al., 2018; Iwafuchi-Doi et al., 2016; Li et al., 2012), loss-of-function analyses provide more direct evidence of its role in differentiation. Whether pioneer factor activity or another molecular mechanism underlies control of cellular competency by *FOXA2* remains to be addressed. Our highly controlled *in vitro* loss-of-function screen reveals cell-autonomous effects with correct species and factor specificity. Importantly, the ability to differentiate into CDX2^{+} mid-hindgut END progenitors remains intact, consistent with effects suggested in *Foxa2*-knockout mouse studies (McKnight et al., 2010).

STAR★METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, René Maehr (rene.maehr@umassmed.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Culture—H1 hESCs (WiCell, WA-01) were maintained in mTeSR1 (StemCell Technologies, Inc., 05850) on hES-qualified Matrigel (Corning, 354277) coated plates. Cells were fed daily and split with ReLeSR (StemCell Technologies, Inc., 05872) every 4–5 days in mTeSR1. H1 hESCs were used to generate the targeted H1-AAVS1-TetOn-dCas9-KRAB hES cell line used in this study. H1-AAVS1-TetOn-dCas9-KRAB hESCs were maintained similarly to H1 hESCs.

HEK293T/17 cells (ATCC, CRL-11268) were maintained in DMEM (Thermo, 11965) supplemented with 10% FBS (Thermo, 10437), 1X GlutaMAX (Thermo, 35050), 1X non-essential amino acids (NEAA; Thermo, 11140), and 1X Penicillin-Streptomycin (Corning, 30–002-CI). Cells were split with 0.25% Trypsin (Thermo, 25200) every 3–5 days.

METHOD DETAILS

Generation of pAAVS1-TetOn-dCas9-KRAB targeting plasmid—The pAAVS1-TetOn-dCas9-KRAB plasmid was generated through Gateway cloning of pAAVS1-TetOn-Dest plasmid (the destination cassette derived from pHAGE-EF1a-DEST-HA-PGK-Puro is cloned in to PacI and NotI sites of pAAVS1-NDi-CRISPRi [Addgene, 73497] (Mandegar et al., 2016) and pENTR2B-dCas9-KRAB plasmid). The dCas9-KRAB fusion was assembled by combining the dCas9 from [Addgene, 60903] (Tanenbaum et al., 2014) with a KRAB repressor found in [Addgene, 50919] (Kearns et al., 2014). The pAAVS1-TetOn-dCas9-KRAB and the pENTR2B-dCas9-KRAB plasmids are available at Addgene (Addgene #115545 and 115547 respectively).

Generation of H1-AAVS1-TetOn-dCas9-KRAB hESCs—H1 hESCs were dissociated to single cells with TrypLE Express (Thermo, 12604013) and approximately 1 × 10^6 cells were washed with PBS at 500 × g for 5 min at RT. Transfection was performed using 4D-Nucleofector with Amax P3 primer cell 4D nucleofector X kit (V4XP-3024) as per the manufacturer's recommendations. 6 μg of target plasmid containing TetOn-dCas9-KRAB and 2 μg of AAVS1 ZnF nucleases were used for the targeting. Nucleofected H1 hESCs were distributed into 6 well plates containing mTeSR1 and 10 μM Y-27632. After 24 hours, cells were treated with 50 ng/ml neomycin for 12 days. After selection, individual colonies were picked into 24 well plates and PCR genotyped for dCas9-KRAB, an AAVS1 wild-type allele, and an AAVS1 targeted allele using the primers and PCR conditions listed in Table S1.

gRNA cloning—As maximal CRISPRi-mediated repression has been demonstrated using gRNAs proximal to the transcriptional start sites of genes (Gilbert et al., 2014; Horlbeck et al., 2016; Mandegar et al., 2016), we chose to employ target-specific gRNA sequences from an improved, predictive CRISPRi library (Horlbeck et al., 2016) in conjunction with 10 control scramble gRNA sequences (3 gRNA per candidate, 160 gRNA total). CRISPRi gRNA sequences (Table S1) from the human CRISPRi v2 (hCRISPRi-v2) library (Horlbeck et al., 2016) were cloned into the CROPseq-Guide-Puro gRNA backbone (Addgene, 86708) as described in Datlinger et al. (2017). Briefly, CROPseq-Guide-Puro backbone was digested with BsmBI to remove the filler sequence necessary for cloning. gRNA sequences with 5′ and 3′ arms of homology were ordered from Thermo Scientific and cloned individually into the digested plasmid using NEBuilder HiFi DNA Assembly Master Mix (NEB, E2621). Final individually cloned gRNA constructs were transformed into chemically competent cells and prepped using a QIAprep Spin Miniprep Kit (QIAGEN, 27106) to generate DNA for downstream lentivirus production.

Lentivirus production—Lentivirus was produced in HEK293T/17 cells (ATCC, CRL-11268). Briefly, sgRNA coding plasmids were transfected with lentiviral packaging plasmids pHDM-G (DNASU #235), pHDM-Hgpm2 (DNASU #236), pHDM-tat1b (DNASU #237), and pRC/CMV-rev1b (DNASU #246) using TransIT-293 transfection reagent (Mirus, 2700) in Opti-MEM (GIBCO, 31985) according to the manufacturer's instructions. Virus was harvested in mTeSR1 48 hours after transfection.

Transduction of H1-AAVS1-TetOn-dCas9-KRAB hESCs with gRNA lentivirus—H1-AAVS1-TetOn-dCas9-KRAB hESCs were split with TrypLE Express and incubated with sgRNA lentivirus for 3 hours in a low attachment plate. Transduced cells were then plated onto hESC-qualified Matrigel-coated plates in mTeSR1 supplemented with 10 μM Y-27632. Beginning 2 days after transduction, cells were treated with 1 μg/ml puromycin (Invitrogen, A1113803) for 72 hours in order to select for cells that were transduced with sgRNAs. For END experiments, selected cells were treated with 500 ng/mL doxycycline (Sigma, D9891) prior to differentiation.

For the END scRNA-seq CRISPRi screen of atacTFAP candidate TFs, 3 gRNAs targeting each of the top 50 candidates and 10 control scramble gRNAs were used per replicate. In order to maximize individual gRNA delivery, and to avoid confounding effects of multiple different gRNA species within one cell, gRNA transductions were performed as an array prior to pooling, which has proven effective for RNAi-based screening (Chen et al., 2014; Crotty and Pipkin, 2015). Following puromycin selection, equivalent cell numbers from each gRNA transduction were pooled. After pooling, expression of dCas9-KRAB was induced in the pooled culture with 500 ng/mL doxycycline for 48 hours before starting the differentiation. The analysis was performed at the END time-point in 2 biological replicates to ensure consistency within the results (160 gRNA conditions per replicate, 320 total individual transductions).

For the hepatic endoderm scRNA-seq CRISPRi experiment, H1-AAVS1-TetOn-dCas9-KRAB hESCs were individually transduced with scramble-gRNA1 or *FOXA2*-gRNA2 lentivirus. Following puromycin selection, equivalent cell numbers from each gRNA transduction were pooled and maintained in the presence of 500 ng/mL doxycycline for 48 hours prior to the start of differentiation. Cells were co-cultured throughout the differentiation to hepatic endoderm in the presence of 500 ng/mL.

Definitive endoderm differentiation—H1-AAVS1-TetOn-dCas9-KRAB hESCs were split to single cells with TrypLE Express (Thermo, 12604). Cells were resuspended in mTeSR1 supplemented with 10 μM Y27632 (Tocris, 1254) and 500 ng/mL doxycycline. 2 × 10^6 cells were plated into each well of a 6-well plate pre-coated with Growth Factor Reduced Matrigel (Corning, 356231). On Day 1, cells were fed with mTeSR1. On Day 2, cells were fed with RPMI1640 (Thermo, 21870) supplemented with 0.2% Hyclone FBS (GE Healthcare, SH30070.03), 100 ng/mL Activin A (R&D Systems, 338-AC-01M), 3 μM CHIR 99021 (Tocris, 4423), and 50 nM PI 103 (Tocris, 2930). On Days 3 and 4, cells were fed with RPMI1640 supplemented with 0.2% Hyclone FBS, 100 ng/mL Activin A, and 250 nM LDN-193189 (Tocris, 6053). Media was changed every 24 hours. For perturbation experiments and CRISPRi screening, the media was supplemented with 500 ng/mL doxycycline daily.

Foregut endoderm differentiation—Differentiation to foregut endoderm was performed as described in Hannan et al. (2013) with modifications. H1-AAVS1-TetOn-dCas9-KRAB hESCs were differentiated to END using the STEMdiff Definitive Endoderm Kit (StemCell Technologies, Inc., 05110) following the manufacturer’s instructions with the addition of 500 ng/mL doxycycline daily. END was split to single cells with TrypLE Express and resuspended in RPMI-1640 supplemented with 1X B27 minus vitamin A (B27-RPMI; Thermo, 12587010), 10 mM Y27632, 50 ng/mL Activin A, and 500 ng/mL doxycycline. 1 × 10^6 cells were plated into each well of a 6-well plate pre-coated with 804G conditioned-medium. Cells were fed for an additional 2 days with B27-RPMI supplemented with 50 ng/mL Activin and 500 ng/mL doxycycline.

Mid-hindgut endoderm differentiation—Differentiation to mid-hindgut endoderm was performed as described in Múnera and Wells (2017) (human intestinal organoid protocol) with the following modifications. H1-AAVS1-TetOn-dCas9-KRAB hESCs were differentiated to END using the STEMdiff Definitive Endoderm Kit following the manufacturer’s instructions with the addition of 500 ng/mL doxycycline daily. END was split to single cells with TrypLE Express and resuspended in RPMI-1640 supplemented with

2% Hyclone FBS (2%-RPMI; GE Healthcare, SH30070.03), 10 µM Y27632, 3 µM CHIR 99021, 500 ng/mL FGF4 (R&D Systems, 235-F4–025), and 500 ng/mL doxycycline. 2 × 106 cells were plated into each well of a 6-well plate pre-coated with 804G conditioned-medium. Cells were fed for an additional 3 days with 2%-RPMI supplemented with 3 µM CHIR 99021, 500 ng/mL FGF4, and 500 ng/mL doxycycline.

Hepatic endoderm differentiation—Differentiation to hepatic endoderm was performed as described in Hannan et al. (2013) with modifications. H1-AAVS1-TetOn-dCas9-KRAB hESCs were first differentiated to foregut endoderm as above. Then cells were fed for 4 days with B27-RPMI supplemented with 20 ng/mL BMP4 (R&D Systems, 314-BP-050), 10 ng/mL FGF10 (R&D Systems, 345-FG-025), and 500 ng/mL doxycycline.

Western blotting—H1-AAVS1-TetOn-dCas9-KRAB hESCs were treated with and without 500 ng/mL doxycycline for 48 hours. Cells were split with TrypLE and pelleted at 300 x g for 5 minutes at RT. The cell pellet was lysed with RIPA buffer (10 mM Tris-Cl pH 8, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 150 mM NaCl, and 1X protease inhibitors) on ice for 20–30 minutes and sonicated for a few pulses to shear the non-soluble chromatin. Samples were spun at 13,000 rpm for 15 minutes at 4°C and the supernatant was collected as total protein extract in a separate tube. Samples were run on precast SDS-PAGE gels (Biorad, 456–1084) and after transfer onto PVDF membrane, stained with anti-HA (1:1000; Cell Signaling Technologies, 3724) and anti-GAPDH (1:1000; R&D Systems, 2275-PC-100) antibodies.

Immunofluorescence—Cells were fixed with 10% formalin (Fisher, 032–060) for 30 min at room temperature. Fixed cells were incubated with 5% donkey serum (Lampire, 7332100) in PBS supplemented with 0.2% Triton X-100 (PBST; Sigma, X100) for 45 minutes at room temperature. Incubation with primary antibodies was performed overnight at 4C. Cells were washed 3 times with PBST for 5 minutes. Cells were incubated with Alexa Fluor-conjugated secondary antibodies for 2 hours at room temperature in the dark. Cells were washed 3 times with PBST for 5 minutes. Hoechst (Thermo, H3570) was used for nuclei staining. Fluorescent images were obtained on a Nikon Eclipse Ti microscope. Primary antibodies used in this study include: SOX17 antibody (1:300; R&D Systems, AF1924), OCT3/4 antibody (1:100; Santa Cruz Biotechnology, sc5279), FOXA2 antibody (1:300; Millipore EMD, 07–633), CDX2 antibody (1:300; BioGenex, MU392A-UC), HNF4A antibody (1:500; Abcam, ab41898), HHEX antibody (1:500; R&D Systems, MAB83771), TBX3 antibody (1:300; Santa Cruz Biotechnology, sc17871), PROX1 antibody (1:300; R&D Systems, AF2727), EPCAM antibody (1:1000; Biolegend, 324202).

For quantification of END efficiency, both ESC and END were stained with SOX17 and OCT4 antibodies and Hoechst. Total cells per well were counted based on Hoechst staining. ESC cells were counted as OCT4+, END cells were counted as SOX17+, and if a cell did not stain for either, it was labeled as “Other.” The percentages of OCT4+ and SOX17+ cells were calculated for both ESC and END. For ESC quantification, 3 random snapshots from 3 independent wells were used. For END quantification, 3 random snapshots from 3 independent differentiations were used.

Quantitative PCR analysis and bulk cell RNA-sequencing—Total RNA was isolated using Trizol Reagent (Invitrogen, 15596–018) according to the manufacturer's instructions. For quantitative PCR analysis, 1 μg of total RNA was reverse-transcribed using SuperScript III First-Strand Synthesis System (Thermo, 18080051). Resulting cDNA was utilized in qPCR reactions using specific primers (Table S1) in KAPA SYBR Fast Master Mix (Kapa Biosystems, KK4600). Relative transcript expression was calculated using the ΔΔCT method; all transcripts were normalized to *ACTB*. For qPCR primer sequences, see Table S1.

Bulk RNA-sequencing library preparation of H1 hESC and differentiated hEND was performed following a published protocol (Zhang et al., 2012). Briefly, 4 mg of total RNA was depleted of ribosomal RNA using Ribo-Zero rRNA Removal Kit (Epicenter, MRZH116) and then converted to cDNA. Final sequencing libraries were generated using dUTP (Thermo, R0133) incorporation and uracil-N-glycosylase (NEB, M0280) treatment to generate strand-specific RNA-seq libraries. The final libraries were sequenced at paired-end on a Hi-Seq2000 platform. Bulk RNA-sequencing library preparation of the H1-AAVS1-TetOn-dCas9-KRAB hESC differentiation time-course to hEND was performed using the QuantSeq 3′ mRNA-Seq Library Prep Kit FWD for Illumina (Lexogen, 015.96). Cells were differentiated in the presence of 500 ng/mL doxycycline and taken for analysis every 24 hours. The final libraries were sequenced (single-end, 75) on a NextSeq500 platform.

Assay for Transposase-Accessible Chromatin (ATAC) sequencing—Samples were processed according to Buenrostro et al. (2013) with slight modifications. Briefly, approximately 50,000 cells were pelleted at 750 x g for 15 minutes at 4°C. After the addition of lysis buffer (10 mM Tris-Cl pH 7.4, 10 mM NaCl, 3 mM MgCl2, and 0.1% IGEPAl CA-630), nuclei were pelleted at 750 x g for 15 minutes at 4°C and the supernatant was discarded. Transposase reaction was performed at 37°C for 30 minutes. The purified, tagmented DNA was amplified for 9–10 cycles and size selected using AMPure XP beads (Beckman Coulter, A63881) as follows: 0.55X volume of 2X concentrated AMPure beads was added to the amplified DNA and incubated for 5 minutes at room temperature. The supernatant, containing low-molecular weight DNA, was collected in a separate tube by removing the beads using a magnetic rack. 1X volume of AMPure beads was added to the supernatant and incubated for 5 min at room temperature. The beads were washed twice with 75% ethanol and then air-dried. The final purified library was eluted in EB buffer and sequenced (paired-end, 75–75) on a NextSeq500 or Hi-Seq2000 platform.

Chromatin Immunoprecipitation (ChIP) sequencing—Approximately 3–5 million cells were used for each histone ChIP-seq experiment. Briefly, cells were cross-linked with 1% formaldehyde for 10 minutes followed by quenching with 125 mM glycine for 4–5 minutes at room temperature. The cell pellet was lysed in cell lysis buffer (20 mM Tris-HCl pH 8, 85 mM KCl, 0.5% NP-40) supplemented with 1X protease inhibitors (Roche, 11836170001) on ice for 20 minutes then spun at 5000 rpm for 10 minutes. The nuclear pellet was resuspended in sonication buffer (10 mM Tris pH 7.5, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, and 1X protease inhibitors) and incubated for 10 minutes at 4°C. In order to achieve a 200–700 bp DNA fragmentation range, nuclei were sonicated using a

Bronson sonifier (model 250) with the following conditions: amplitude = 15%, time interval = 3min (total of 8–12 minutes) and pulse ON/OFF = 0.7 s/1.3 s. Chromatin was pre-cleared with Dynabeads Protein A (Invitrogen, 10002D) for 1 hour and incubated with antibody on a rotating wheel overnight at 4°C. Antibodies included: anti-H3K27ac (5 μg; Diagenode, C15410196), anti-H3K27me3 (5 μg; Millipore, 07–449), and anti-FOXA2 (8 μg; Millipore, 07–633). On the following day, 30–40 ml of Dynabeads Protein A was added to chromatin for 2–3 hours. The captured immuno-complexes were washed as follows – 1x in low-salt buffer, 1x in high-salt buffer, 1x in LiCl salt buffer, and 1x in TE. The immuno-complexes were eluted in ChIP-DNA elution buffer (10 mM Tris-HCl pH 8, 100 mM NaCl, 20 mM EDTA, and 1% SDS) for 20 minutes. The eluted ChIP-DNA was reverse cross-linked overnight at 65°C, followed by proteinase K (Thermo, 25530049) treatment, RNase A (Thermo, ENO531) treatment, and Phenol:Chloroform:Isoamyl alcohol extraction. The Illumina library construction steps were carried out with 5–10 ng of purified DNA. During library construction, purification was performed after every step using QIAquick PCR purification kit (QIAGEN, 28104) or QIAquick gel extraction kit (QIAGEN, 28706). The library reaction steps were as follows: end-repair, 3′ end A-base addition, adaptor ligation, and PCR amplification. The amplified libraries were size-selected for 200–450 bp on a 2% agarose E-gel (Thermo, G402002) and sequenced (single-end, 75) on a NextSeq500 or Hi-Seq2000 platform.

Perturbed definitive endoderm 10X Genomics library preparation—~8000 cells were captured per replicate on a 10X Chromium device using a 10X V2 Single Cell 3′ Solution kit (10X Genomics). All protocols were performed following the manufacturer's instructions. Final sequencing libraries were analyzed on a high sensitivity DNA fragment analyzer chip (Advanced Analytical) to determine the average base pair size and final library concentrations were determined with a Qubit High Sensitivity DNA assay kit (Thermo, Q32854). 10X genomics libraries were sequenced at paired-end (26–50) on a Nextseq500 using a Nextseq500/550 High Output v2 75-cycle kit (Illumina, FC-404–2005).

Perturbed hepatic endoderm Drop-seq library preparation—Drop-seq was performed following the Drop-seq Laboratory Protocol version 3.1 (http://mccarrolllab.org/dropseq). Briefly, single cell suspensions were resuspended at 1 × 10^5 cells/mL in PBS + 0.01% BSA (Sigma, A8412). The diluted cell suspension, barcoded Oligo-dT beads (Chemgenes, MACOSKO-2011–10), and droplet generation oil (Biorad, 1864006) was run through a PDMS co-flow microfluidic droplet generation device (Nanoshift, custom built based on the datafile 1 from Macosko et al. (2015) at flow rates of 4,000 μL per hour, 4,000 μL per hour, and 15,000 μL per hour, respectively. Droplet breakage, bead isolation, and cDNA synthesis were performed as described (Kernfeld et al., 2018; Macosko et al., 2015). cDNA libraries were tagmented with Nextera XT DNA Library Preparation Kit (Illumina, FC-131–1024) and sequencing libraries were amplified and individually barcoded. Agencourt AMPure XP beads (Beckman Coulter, A63881) were used for purification of cDNA and sequencing libraries according to the manufacturer's instructions. Final sequencing libraries were analyzed on a high sensitivity DNA fragment analyzer chip to determine the average base-pair size and final library concentrations were determined with a Qubit High Sensitivity DNA assay kit (Invitrogen, Q32854). Drop-seq libraries were sequenced at paired-end (20–50) on a Nextseq500 using a Nextseq500/550 High Output v2 75-cycle kit (Illumina, FC-404–2005).

gRNA amplification from 10X single-cell RNA-seq libraries—In order to increase resolution of gRNA assignments to individual cells, gRNA amplification off the U6 promoter was performed, which maintained UMI and cell barcode information following sequencing. 10 ng of final 10X genomics single-cell RNA-seq library was used in two subsequent PCR reactions using HiFi HotStart ReadyMix (Kapa Biosystems, KK2600) in order to amplify gRNA sequences and add on sequencing adaptors and multiplexing indices. Final gRNA amplification libraries were sequenced at paired-end (26–50) on a Nextseq500 using a Nextseq500/550 High Output v2 75-cycle kit (Illumina, FC-404–2005). For primer sequences and PCR conditions, see Table S1.

Genomic DNA (gDNA) sequencing from transduced cells for 10X library quality control—gDNA was isolated from both scRNA-seq CRISPRi replicates prior to doxycycline treatment (pre-dox, before differentiation) and following differentiation in the presence of doxycycline. gDNA samples were amplified and sequenced for library quality control. Briefly, gDNA isolation was performed using a DNeasy kit (QIAGEN, 69506). gRNA sequences were amplified with Q5 Hot Start High-Fidelity 2x Master Mix (NEB, M0494) and specific primers as described (Datlinger et al., 2017). Final amplified libraries were sequenced (single-end, 75) on a NextSeq500. Reads were aligned and exact matches were quantified using ScreenProcessing. (Version numbers are unavailable, but ScreenProcessing used the Git version control system, and we installed code from commit 50628c7).

QUANTIFICATION AND STATISTICAL ANALYSIS

Next-generation sequencing library data processing—For processing of bulk RNA-seq data utilized in atacTFAP analysis, raw reads were trimmed 5bp from the head orientation and to 55 bp at the tail. Subsequently, reads were aligned with salmon version 0.8.2 to human reference transcriptome version GRCH37 using default parameters setting the flags -l A-posBias-gcBias-seqBias. Transcript and gene expression levels were then quantified using the salmon output and the txImport (Soneson et al., 2015) R package. Subsequently, we removed all genes with less than 3 (20th percentile) or more than 27337.16 (99.5th percetnile) reads across the dataset. Differential expression analysis was performed using DESeq2 (Love et al., 2014). Genes with a minimal expression level of 5 RPKM in at least 2 samples and a FDR threshold of 0.01 were considered as differentially expressed.

For processing of ATAC-seq data, reads were aligned to the human reference genome hg19 using Bowtie 2 (Langmead and Salzberg, 2012) version 2.3.2 using default parameters and filtering duplicate reads. MACS2 in combination with the IDR framework was used for peak calling and detection of regions of genomic enrichment using two replicates per condition with an IDR cutoff of 0.1. (Modified from https://informatics.fas.harvard.edu/, Harvard University)

For processing of ChIP-seq data, reads were aligned to the human reference genome hg19 using Bowtie 2 (Langmead and Salzberg, 2012) version 2.3.2 using default parameters.

Duplicate reads were filtered out and reads were extended to 200 bp. For peak calling, MACS2 in combination with the IDR framework was utilized with two replicates per condition and an IDR cutoff of 0.1.

For differential ATAC-Seq or ChIP-Seq analysis, we used the R package diffbind in combination with DESeq2 for all IDR based peak sets, requiring no overlap of peaks across conditions and using a DBA score based quantification.

For RNA-seq, ATAC-seq, and ChIP-seq data visualization, IGV tools (Thorvaldsdóttir et al., 2013) was used to generate .tdf files. Quant-seq data were aligned using HISAT2 (Kim et al., 2015) v2.0.5 with hg19 as a reference and parameters “-p 12–rna-strand-ness F.” Quantification used ESAT (Derr et al., 2016) v0.1 with parameters “-wLen 100 -wOlap 50 -wExt 1000 -sigTest 0.01 -multimap normal.” For ESAT quantification, RefSeq transcript annotations were downloaded from the UCSC table browser with the following specifications: clade: Mammal; genome: Human; assembly: Feb. 2009 (GRCh37/hg19); group: Genes and Gene Predictions; track: NCBI RefSeq; table: UCSC RefSeq (refGene); region: genome; output format: all fields from selected table. A negative binomial likelihood ratio test was carried out using DESeq2 (Love et al., 2014) with full model having a separate fixed effect for each day, a constant null model, and fitType = “local.” For heatmapping, expression was normalized to counts per million and transformed as $X \Rightarrow log_2(X + 1)$. Rows were standardized and averaged by time point.

ATAC-seq Transcription Factor Activity Prediction (atacTFAP) analysis—In order to identify TFs that are likely relevant for the biology and fate of a particular cellular state, we performed regression analysis on the ATAC-seq signal across the union set of all putative gene regulatory elements (GREs) in hESC, END, and beta cells (GEO: GSM1978246, GSM1978247; mature endodermal cell population control) using predicted TF binding sites within each GRE as regressors (Ziller et al., 2015).

More specifically, we performed TF motif matching using PWM matrices obtained from JASPAR and HOCOMOCO employing the PWM matching tool FIMO (Grant et al., 2011) version 4.10.2 on the union peak set of hESC, END, and beta cells. In order to normalize the distribution of predicted TFBS sites across peaks, we standardized the peak length of all ATAC peaks to 600 bp, extending 300 bp in each direction from the peak center and used these regions for PWM matching. Subsequently, we only retained motif matches below a significance of 5e-04 and log10-transformed the values multiplied by −1.

In addition, we computed the RPKM values across the union ATAC-seq peak set using the effective library sizes defined as the total number of reads in peaks of the union peak set in each library as scaling factor with two replicates per condition. Lastly, we performed quantile normalization across all replicates and subsequently averaged the quantile normalized signal per condition, giving rise to the final dependent variable matrix. We then performed sparse partial least square regression with the R package spls (Chun and Keleş, 2010) and identified the unknown parameters K and eta by 5 fold cross-validation using a grid search across K = 2–20 and eta = 0.1–0.9 keeping kappa fixed at 0.5 with the algorithm set to pls2 and fit = simpls. This identified K = 16 and eta = 0.1 as the parameters with minimum cross-validation error. Finally, we used the difference in the estimated beta coefficients for each TF motif between hESC and END as a measure for the relative importance of this factor for the biology/establishment/maintenance of the hESC or END fate and defined this value as the atacTFAP score (Figure 1C; Table S2). We then further filtered the results for TFs that were not expressed at least 5 RPKM in hESC or END and determined the log2 fold change between the two conditions as the RNAdiff score (Figure 1C; Table S2). We then removed duplicate TFs (e.g., TFs with multiple motifs), retaining only the TF/motif with the highest atacTFAP score and retained only one motif per TF. From this list, we selected 50 factors (Figure 1C; Table S2).

Quantification of scRNA-seq CRISPRi data—Reads from sequencing of scRNA-seq CRISPRi results, including gRNA sequences, were generated, aligned and quantified using 10X CellRanger version 2.1.0. The reference genome was hg19, augmented with the 4,542 base pair dCas9-KRAB fusion sequence and the gRNA RNA sequences with homology arms. Each additional sequence was included on a separate chromosome. Each gRNA sequence included the 20 unique base pairs as well as 200 bp on either side. For purposes of tagging, 50 bp on either side of the unique gRNA sequence was marked as “exonic.” The entire dCas9-KRAB sequence was marked as exonic.

gRNA amplification runs were processed separately from the full data. For the gRNA amplification runs, cell calling was ignored, and barcodes were instead carried over from full data processing. Full data and gRNA amplification data were then merged within replicates. To avoid double-counting individual UMIs, gRNA counts from full sequencing runs were erased and replaced with gRNA counts from gRNA amplification sequencing. 17,234 cells were reported, with the median cell having 17,231 UMIs.

Quality control analysis of definitive endoderm scRNA-seq CRISPRi data—Analysis of count matrices from scRNA-seq CRISPRi data was carried out using R 3.4.3. Doublets were depleted by modeling the amount of the highest expressed and second highest expressed gRNAs. Calling the highest gRNA count x and the second highest y, quantile regressions of y on x were fitted targeting the 50th and 99th percentile to model single cells and doublets respectively. Intercepts were fixed at 0 and the 99th percentile (doublet) model attained a slope near 1, indicating roughly equal amounts for the top two gRNAs. Log10 likelihood ratios (log10 LRs) were computed for each cell, assuming Poisson-distributed counts around the regression estimate, and cells with log10 LR above 0.2 were excluded (730 cells). Quantile regressions of y on x were fitted via the quantreg package version 5.35.

Each cell was then assigned to its highest expressed gRNA. After doublet depletion and gRNA assignment, each scramble gRNA was tested against the other 9 scrambles within the DE samples using MAST (Finak et al., 2015) version 1.4.1 with a fixed effect for replicate. All transcripts were tested, including dCas9-KRAB. FDR adjustment was applied to each gRNA separately.

Cells excluded from scRNA-seq CRISPRi screen data—In each replicate, genes appearing in only one cell were excluded. Cells with doublet-modeling log10 LR above 0.2 were excluded (730 cells). Cells were excluded if no gRNAs were detected (29 cells). Cells assigned to scramble gRNA #5 were excluded based on the high number of differentially expressed transcripts (88 cells). This left 548 negative control cells (3.4%), which is lower than expected but still within the range used by similar studies (Dixit et al., 2016). Ten gRNAs yielded exactly 0 counts in the END gDNA ScreenProcessing results, and upon visualization of the RNA alignments, these were found to contain possible mutations or targeting errors (they were: ARNT_gRNA2, ATF3_gRNA3, CREB3_gRNA2, FOXA2_gRNA3, FOXA3_gRNA1, GATA4_gRNA1, GATA6_gRNA2, JUND_gRNA2, TGIF2_gRNA1, ZNF263_gRNA2). Cells assigned to these gRNAs were excluded from downstream analysis (634 cells). After all exclusions, 16,110 cells remained.

Unsupervised analysis of definitive endoderm scRNA-seq CRISPRi data—Normalization and unsupervised analysis were carried out using Seurat (Satija et al., 2015) version 2.3.0. Expression values were converted to transcripts per 10,000 and log-transformed via X - > ln(1+X). 1,062 variable genes were selected using a dispersion measure based on the mean and coefficient of variation (CV) for each gene. Specifically, genes were binned by expression and a local median was computed for each bin. Dispersion was measured as the ratio of the CV to the local median CV, and genes with dispersion > 0.5 were retained. gRNAs and dCas9-KRAB were excluded, and so were genes with log normalized expression outside the interval[0.0125, 3].

Each gene's log normalized expression was replaced with scaled residuals from a regression on the total number of UMIs, with the regression fitted to one datum per cell. PCA was computed on the results, and the top 15 PCs were used as input for Barnes-Hut t-stochastic neighbor embedding (Maaten and Hinton, 2008) and a Louvain algorithm variant (Waltman and van Eck, 2013). In the Louvain algorithm, the resolution parameter was set to 0.075.

Characterization of definitive endoderm scRNA-seq CRISPRi cell clusters—P values for gRNA enrichment in each cluster were calculated by applying Fisher's exact test to 4×2 tables, where each entry j, 1 of the table contains the number of cells in cluster j where the gene was targeted and entry j, 2 of the table contains the number of cells in cluster j with scramble guides. P values were FDR-adjusted across all targets by the method of Benjamini and Hochberg. Scramble enrichment in cluster 0 was tested similarly for significance, using a null distribution with the same total number of scrambles distributed to so as to form the same percentage of each cluster.

Characterization of gRNA effects within Cluster 0—gRNA effects within the main cluster were estimated via MIMOSCA (Dixit et al., 2016) within Python 2.7.13. (Version numbers for MIMOSCA were unavailable, but the software is tracked using the version control system Git, and we used code from commit 27199eb.) Raw counts were exported from R following quality control and unsupervised analysis. Expression data were normalized and log-transformed within MIMOSCA, and genes were standardized. MIMOSCA was run with a linear model using indicators for the interaction between replicate and gRNA. Scramble gRNAs did not receive separate indicators, so the resulting coefficients represent log fold change for each gRNA over the combined scrambles. MIMOSCA's built-in EM-like correction of gRNA assignments was applied. Genes with correlation across replicates less than 0.25 were excluded from downstream analysis.

To find common patterns among the MIMOSCA coefficients for each target, guides and replicates were combined via averaging. Sparse PCA was run using the R package elasticnet version 1.1 (regularization parameter of 20), with the sparsity penalty on the gRNAs (Zou et al., 2006). Only one component was retained, as subsequent components were concentrated within a single target. Genes for the heatmap were selected for having high loadings on PC1 (any gene whose loading exceeded 1/3 of the in absolute value).

gRNA effects were also estimated by comparing cells assigned to each gRNA against cells assigned to scramble gRNAs. Testing used the MAST package (Finak et al., 2015) (v1.4.1) with a fixed effect for the replicate. Tests were applied to the top 1000 genes according to absolute log fold change, and P values were corrected for multiple testing using the method of Benjamini and Hochberg. To correct for the effective number of tests (i.e., the number of genes G), the remaining G-1000 p values were conservatively assumed to be 1.

Quality control of gRNA effects—To assess consistency of cluster assignments for different guides targeting the same locus, we computed two quantities across all pairs of guides: the dissimilarity between assignments and the effect size. Dissimilarity was measured via the total variation norm between Pr(Cluster j guide 1) and Pr(Cluster | guide 2). Effect size was measured as the maximum over i of Pr(cell not in 0 | guide i). Then, smooth curves were computed to estimate dissimilarity given effect size. This was done separately for pairs of guides sharing a target and pairs not sharing a target. Smooth curves are penalized cubic splines fit via least-squares.

To assess consistency of differential expression for different guides targeting the same locus, we computed two quantities across all pairs of guides: the similarity in effects and the effect size. Similarity was measured via the percent of shared differentially expressed genes. Effect size was measured as the total number of differentially expressed genes. If two guides affected the same gene in opposite directions, this counted twice toward the effect size and not toward the shared percentage. This was done separately for pairs of guides sharing a target and pairs not sharing a target. Smooth curves are penalized cubic splines fit with a quasi-Poisson response distribution.

Despite low statistical power for measuring individual transcripts in scRNA-seq, guides display an overall trend toward knockdown (Figure S2G), with 75% of gRNAs knocking down their target (uncorrected $p < 0.05$, log $FC < 0$; Table S3), which is comparable to the results previously described (Dixit et al., 2016).

Quantification of hepatic endoderm scRNA-seq CRISPRi data—Reads were aligned and quantified using the Drop-seq tools (Macosko et al., 2015) version 1.0 and the STAR aligner v2.4.2 (Dobin et al., 2013). The reference genome was the same as for the END scRNA-seq CRISPRi data. During processing, cells were filtered to have at least 1,000 genes. This yielded 1,165 cells.

Analysis of counts from hepatic endoderm scRNA-seq CRISPRi data—Analysis of hepatic endoderm scRNA-seq CRISPRi data was carried out using R 3.4.3. Cells were removed if no gRNAs were detected (216 cells). Barcodes with multiple gRNAs (presumed doublets) were removed as in the END screening pool (46 cells removed).

Normalization and unsupervised analysis were carried out using Seurat version 2.3.0 (Butler et al., 2018). Expression values were converted to transcripts per 10,000 and log-transformed via X - > ln(1+X). Highly variable genes were selected using the same expression criteria and dispersion measure as above, but requiring dispersion > = 1.5. The resulting list contained 631 genes. Each gene's log normalized expression was replaced with scaled residuals from a regression on the total number of UMIs, with the regression fitted to one datum per cell. PCA was computed on the results, and the top 8 PCs were used as input for Barnes-Hut t-stochastic neighbor embedding (Maaten and Hinton, 2008) and a Louvain algorithm variant (Waltman and van Eck, 2013). In the Louvain algorithm, the resolution parameter was set to 0.2.

gRNA effects were estimated by comparing all *FOXA2*-gRNA cells with all scramble-gRNA cells. Testing used the MAST package (Finak et al., 2015). P values were corrected for multiple testing using the method of Benjamini and Hochberg, and genes were reported in the supplement as long as their q-values fell below 0.1. Cluster markers were computed similarly.

DATA AND SOFTWARE AVAILABILITY

The accession number for all NGS datasets generated in this paper is GEO: GSE127202. Code used for analysis will be released at https://github.com/maehrlab prior to publication.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank members of the Maehr lab and T. Fazzio for discussions, A. Grella and A. Derr for technical help, B. Conklin for sharing reagents, and the UMass Medical School Molecular Biology core for technical support. We are grateful to P. Zamore and J. Luban for sequencer access. R.M. is supported by NIH grants (DP3DK111898 and R01AI132963) and The Glass Charitable Foundation. M.J.Z. is supported by BMBF e:Med Program 01ZX1504 and the Max Planck Society.

REFERENCES

Adamson B, Norman TM, Jost M, Cho MY, Nuñez JK, Chen Y, Villalta JE, Gilbert LA, Horlbeck MA, Hein MY, et al. (2016). A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. Cell 167, 1867–1882.e21. [PubMed: 27984733]

Alexander J, and Stainier DY (1999). A molecular pathway leading to endoderm formation in zebrafish. Curr. Biol 9, 1147–1157. [PubMed: 10531029]

Allison TF, Smith AJH, Anastassiadis K, Sloane-Stanley J, Biga V, Stavish D, Hackland J, Sabri S, Langerman J, Jones M, et al. (2018). Identification and single-cell functional characterization of an endodermally biased pluripotent substate in human embryonic stem cells. Stem Cell Reports 10, 1895–1907. [PubMed: 29779895]

Ang SL, and Rossant J (1994). HNF-3 beta is essential for node and noto-chord formation in mouse development. Cell 78, 561–574. [PubMed: 8069909]

Avery S, Zafarana G, Gokhale PJ, and Andrews PW (2010). The role of SMAD4 in human embryonic stem cell self-renewal and stem cell fate. Stem Cells 28, 863–873. [PubMed: 20235236]

Balakrishnan SK, Witcher M, Berggren TW, and Emerson BM (2012). Functional and molecular characterization of the role of CTCF in human embryonic stem cell biology. PLoS ONE 7, e42424. [PubMed: 22879976]

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]

Butler A, Hoffman P, Smibert P, Papalexi E, and Satija R (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol 36, 411–420. [PubMed: 29608179]

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, and Ma'ayan A (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128. [PubMed: 23586463]

Chen R, Bélanger S, Frederick MA, Li B, Johnston RJ, Xiao N, Liu Y-C, Sharma S, Peters B, Rao A, et al. (2014). In vivo RNA interference screens identify regulators of antiviral CD4$^+$ and CD8$^+$ T cell differentiation. Immunity 41, 325–338. [PubMed: 25148027]

Chu GC, Dunn NR, Anderson DC, Oxburgh L, and Robertson EJ (2004). Differential requirements for Smad4 in TGFbeta-dependent patterning of the early mouse embryo. Development 131, 3501–3512. [PubMed: 15215210]

Chu L-F, Leng N, Zhang J, Hou Z, Mamott D, Vereide DT, Choi J, Kendziorski C, Stewart R, and Thomson JA (2016). Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. Genome Biol. 17, 173. [PubMed: 27534536]

Chun H, and Keleş S (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. J. R. Stat. Soc. Series B Stat. Methodol 72, 3–25. [PubMed: 20107611]

Conlon FL, Lyons KM, Takaesu N, Barth KS, Kispert A, Herrmann B, and Robertson EJ (1994). A primary requirement for Nodal in the formation and maintenance of the primitive streak in the mouse. Development 120, 1919–1928. [PubMed: 7924997]

Crotty S, and Pipkin ME (2015). In vivo RNAi screens: concepts and applications. Trends Immunol. 36, 315–322. [PubMed: 25937561]

D'Amour KA, Agulnick AD, Eliazer S, Kelly OG, Kroon E, and Baetge EE (2005). Efficient differentiation of human embryonic stem cells to definitive endoderm. Nat. Biotechnol 23, 1534–1541. [PubMed: 16258519]

Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, Schuster LC, Kuchler A, Alpar D, and Bock C (2017). Pooled CRISPR screening with single-cell transcriptome readout. Nat. Methods 14, 297–301. [PubMed: 28099430]

Derr A, Yang C, Zilionis R, Sergushichev A, Blodgett DM, Redick S, Bortell R, Luban J, Harlan DM, Kadener S, et al. (2016). End sequence analysis toolkit (ESAT) expands the extractable information from single-cell RNA-seq data. Genome Res. 26, 1397–1410. [PubMed: 27470110]

Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, et al. (2016). Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. Cell 167, 1853–1866.e17. [PubMed: 27984732]

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]

Dufort D, Schwartz L, Harpal K, and Rossant J (1998). The transcription factor HNF3beta is required in visceral endoderm for normal primitive streak morphogenesis. Development 125, 3015–3025. [PubMed: 9671576]

Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK, Slichter CK, Miller HW, McElrath MJ, Prlic M, et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. 16, 278. [PubMed: 26653891]

Gifford CA, Ziller MJ, Gu H, Trapnell C, Donaghey J, Tsankov A, Shalek AK, Kelley DR, Shishkin AA, Issner R, et al. (2013). Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. Cell 153, 1149–1163. [PubMed: 23664763]

Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, et al. (2014). Genome-scale CRISPR-mediated control of gene repression and activation. Cell 159, 647–661. [PubMed: 25307932]

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

Hannan NRF, Segeritz C-P, Touboul T, and Vallier L (2013). Production of hepatocyte-like cells from human pluripotent stem cells. Nat. Protoc 8, 430–437. [PubMed: 23424751]

Heyer J, Escalante-Alcalde D, Lia M, Boettinger E, Edelmann W, Stewart CL, and Kucherlapati R (1999). Postgastrulation Smad2-deficient embryos show defects in embryo turning and anterior morphogenesis. Proc. Natl. Acad. Sci. USA 96, 12595–12600. [PubMed: 10535967]

Hoodless PA, Pye M, Chazaud C, Labbé E, Attisano L, Rossant J, and Wrana JL (2001). FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. Genes Dev. 15, 1257–1271. [PubMed: 11358869]

Horlbeck MA, Gilbert LA, Villalta JE, Adamson B, Pak RA, Chen Y, Fields AP, Park CY, Corn JE, Kampmann M, and Weissman JS (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. eLife 5, 914.

Iwafuchi-Doi M, Donahue G, Kakumanu A, Watts JA, Mahony S, Pugh BF, Lee D, Kaestner KH, and Zaret KS (2016). The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. Mol. Cell 62, 79–91. [PubMed: 27058788]

Kanai-Azuma M, Kanai Y, Gad JM, Tajima Y, Taya C, Kurohmaru M, Sanai Y, Yonekawa H, Yazaki K, Tam PPL, and Hayashi Y (2002). Depletion of definitive gut endoderm in Sox17-null mutant mice. Development 129, 2367–2379. [PubMed: 11973269]

Kearns NA, Genga RMJ, Enuameh MS, Garber M, Wolfe SA, and Maehr R (2014). Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development 141, 219–223. [PubMed: 24346702]

Kernfeld EM, Genga RMJ, Neherin K, Magaletta ME, Xu P, and Maehr R (2018). A single-cell transcriptomic atlas of thymus organogenesis resolves cell types and developmental maturation. Immunity 48, 1258– 1270.e6. [PubMed: 29884461]

Kim D, Langmead B, and Salzberg SL (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360. [PubMed: 25751142]

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44 (W1), W90–W97. [PubMed: 27141961]

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

Lee CS, Friedman JR, Fulmer JT, and Kaestner KH (2005a). The initiation of liver development is dependent on Foxa transcription factors. Nature 435, 944–947. [PubMed: 15959514]

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

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]

Loh KM, Ang LT, Zhang J, Kumar V, Ang J, Auyeong JQ, Lee KL, Choo SH, Lim CYY, Nichane M, et al. (2014). Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. Cell Stem Cell 14, 237–252. [PubMed: 24412311]

Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. [PubMed: 25516281]

Maaten LVD, and Hinton G (2008). Visualizing data using t-SNE. J. Mach. Learn. Res 9, 2579–2605.

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161, 1202–1214. [PubMed: 26000488]

Mandegar MA, Huebsch N, Frolov EB, Shin E, Truong A, Olvera MP, Chan AH, Miyaoka Y, Holmes K, Spencer CI, et al. (2016). CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. Cell Stem Cell 18, 541–553. [PubMed: 26971820]

Massagué J (2012). TGFβ signalling in context. Nat. Rev. Mol. Cell Biol 13, 616–630. [PubMed: 22992590]

McKnight KD, Hou J, and Hoodless PA (2010). Foxh1 and Foxa2 are not required for formation of the midgut and hindgut definitive endoderm. Dev. Biol 337, 471–481. [PubMed: 19896480]

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, and Bejerano G (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol 28, 495–501. [PubMed: 20436461]

Múnera JO, and Wells JM (2017). Generation of gastrointestinal organoids from human pluripotent stem cells. Methods Mol. Biol 1597, 167–177. [PubMed: 28361317]

Rodriguez TA, Casey ES, Harland RM, Smith JC, and Beddington RS (2001). Distinct enhancer elements control Hex expression during gastrulation and early organogenesis. Dev. Biol 234, 304–316. [PubMed: 11397001]

Sakaki-Yumoto M, Liu J, Ramalho-Santos M, Yoshida N, and Derynck R (2013). Smad2 is essential for maintenance of the human and mouse primed pluripotent stem cell state. J. Biol. Chem 288, 18546–18560. [PubMed: 23649632]

Satija R, Farrell JA, Gennert D, Schier AF, and Regev A (2015). Spatial reconstruction of single-cell gene expression data. Nat. Biotechnol 33, 495–502. [PubMed: 25867923]

Senft AD, Costello I, King HW, Mould AW, Bikoff EK, and Robertson EJ (2018). Combinatorial Smad2/3 activities downstream of Nodal signaling maintain embryonic/extra-embryonic cell identities during lineage priming. Cell Rep. 24, 1977–1985.e7. [PubMed: 30134160]

Shen MM (2007). Nodal signaling: developmental roles and regulation. Development 134, 1023–1034. [PubMed: 17287255]

Shi Z-D, Lee K, Yang D, Amin S, Verma N, Li QV, 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.e6. [PubMed: 28196600]

Soneson C, Love MI, and Robinson MD (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 4, 1521. [PubMed: 26925227]

Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, and Vale RD (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell 159, 635–646. [PubMed: 25307933]

Thorvaldsdóttir H, Robinson JT, and Mesirov JP (2013). Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform 14, 178–192. [PubMed: 22517427]

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]

Tremblay KD, Hoodless PA, Bikoff EK, and Robertson EJ (2000). Formation of the definitive endoderm in mouse is a Smad2-dependent process. Development 127, 3079–3090. [PubMed: 10862745]

Vincent SD, Dunn NR, Hayashi S, Norris DP, and Robertson EJ (2003). Cell fate decisions within the mouse organizer are governed by graded Nodal signals. Genes Dev. 17, 1646–1662. [PubMed: 12842913]

Waltman L, and van Eck NJ (2013). A smart local moving algorithm for large-scale modularity-based community detection. Eur. Phys. J. B 86, 471.

Wan H, Dingle S, Xu Y, Besnard V, Kaestner KH, Ang S-L, Wert S, Stahlman MT, and Whitsett JA (2005). Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. J. Biol. Chem 280, 13809–13816. [PubMed: 15668254]

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]

Wei S, and Wang Q (2018). Molecular regulation of Nodal signaling during mesendoderm formation. Acta Biochim. Biophys. Sin. (Shanghai) 50, 74–81. [PubMed: 29206913]

Weinstein DC, Ruiz i Altaba A, Chen WS, Hoodless P, Prezioso VR, Jessell TM, and Darnell JE Jr. (1994). The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. Cell 78, 575–588. [PubMed: 8069910]

Weintraub AS, Li CH, Zamudio AV, Sigova AA, Hannett NM, Day DS, Abraham BJ, Cohen MA, Nabet B, Buckley DL, et al. (2017). YY1 is a structural regulator of enhancer-promoter loops. Cell 171, 1573–1588.e28. [PubMed: 29224777]

Xie S, Duan J, Li B, Zhou P, and Hon GC (2017). Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. Mol. Cell 66, 285–299.e5. [PubMed: 28416141]

Yamamoto M, Meno C, Sakai Y, Shiratori H, Mochida K, Ikawa Y, Saijoh Y, and Hamada H (2001). The transcription factor FoxH1 (FAST) mediates Nodal signaling during anterior-posterior patterning and node formation in the mouse. Genes Dev. 15, 1242–1256. [PubMed: 11358868]

Zhang Z, Theurkauf WE, Weng Z, and Zamore PD (2012). Strand-specific libraries for high throughput RNA sequencing (RNA-seq) prepared without poly(A) selection. Silence 3, 9. [PubMed: 23273270]

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

Ziller MJ, Edri R, Yaffe Y, Donaghey J, Pop R, Mallard W, Issner R, Gifford CA, Goren A, Xing J, et al. (2015). Dissecting neural differentiation regulatory networks through epigenetic footprinting. Nature 518, 355–359. [PubMed: 25533951]

Zorn AM, and Wells JM (2009). Vertebrate endoderm development and organ formation. Annu. Rev. Cell Dev. Biol 25, 221–251. [PubMed: 19575677]

Zou H, Hastie T, and Tibshirani R (2006). Sparse principal component analysis. J. Comput. Graph. Stat 15, 265–286.

Highlights

- atacTFAP analysis predicts key factors during human ESC differentiation to endoderm
- scRNA-seq CRISPRi identifies factors important for human endoderm differentiation
- Targeting of the TGFβ pathway affects differentiation in a target-specific manner
- *FOXA2* knockdown causes genome-wide changes and impairs differentiation

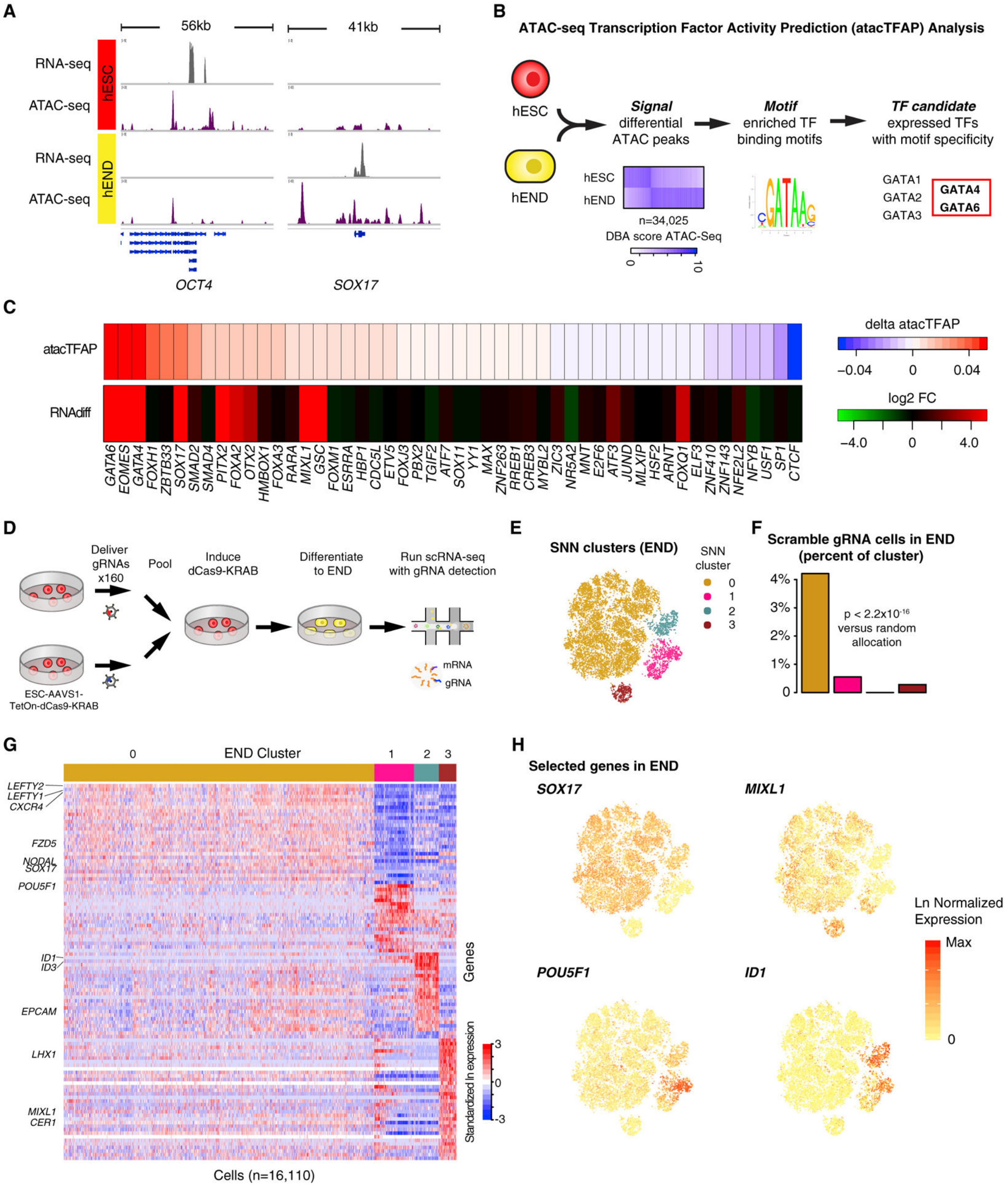


Figure 1. scRNA-Seq CRISPRi Screen Identifies Molecular Drivers of Human END-Diff

(A) Representative integrative genomics viewer (IGV) tracks at the *OCT4* and *SOX17* loci. RNA-seq and ATAC-seq datasets for H1 ESC or END highlight dynamic transcriptome and chromatin changes.

(B) Schematic of the atacTFAP analysis demonstrating how H1 ESC and END ATAC-seq and RNA-seq data (n = 2 biological replicates) are integrated to predict TF candidates during differentiation. Criteria for ATAC-seq peak analysis are FDR < 0.05 and log fold change ≥ 1.0.

(C) 50 TF candidates ordered by atacTFAP score (top) and differential transcript expression (RNAdiff) between ESC and END (bottom).

(D) Schematic of the scRNA-seq CRISPRi screening experiment during END-Diff. Expression of dCas9-KRAB is induced (via the addition of doxycycline) only after cells are pooled.
(E) tSNE and cluster assignments resulting from scRNA-seq CRISPRi experiment ($n = 2$ biological replicates).
(F) For each cluster, proportion of cells assigned to scramble gRNAs ($p < 2.2E\text{-}16$ versus random allocation; hypergeometric test).
(G) Heatmap of all 16,110 cells passing screen quality control. Genes shown are a subset of cluster markers with $q < 0.05$, $FC > 1.5$ in either direction, and detection in at least 10% of cells in some cluster.
(H) Feature plots selected from among top marker transcripts in each cluster. See also Figures S1 and S2, and Tables S1, S2, and S3.

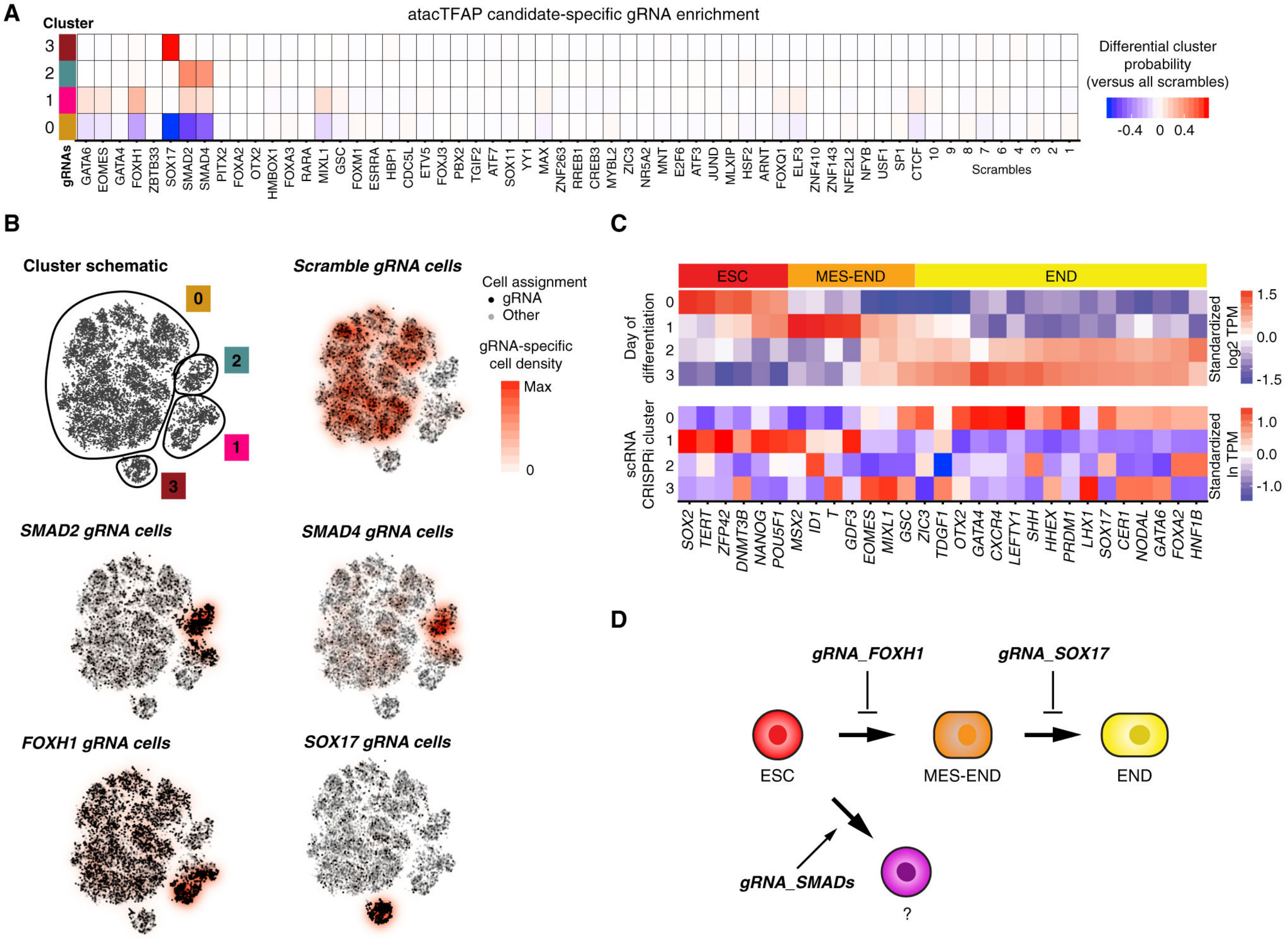


Figure 2. Dissecting the Role of TGFβ Mediators during END-Diff at the Single-Cell Level

(A) Cluster enrichment of gRNAs for each TF target. Heatmap shows Pr(cluster | gRNA) minus Pr(cluster | scramble).

(B) tSNE of cluster assignments and feature plots of gRNAs targeting specific TGFβ mediator genes. Each dot represents one cell.

(C) Staging of END-Diff via Quant-seq (top; n = 2 biological replicates) compared with scRNA-seq CRISPRi cluster characteristics (bottom).

(D) Model of the effects of TGFβ mediator perturbations on human END-Diff. See also Tables S1 and S3.

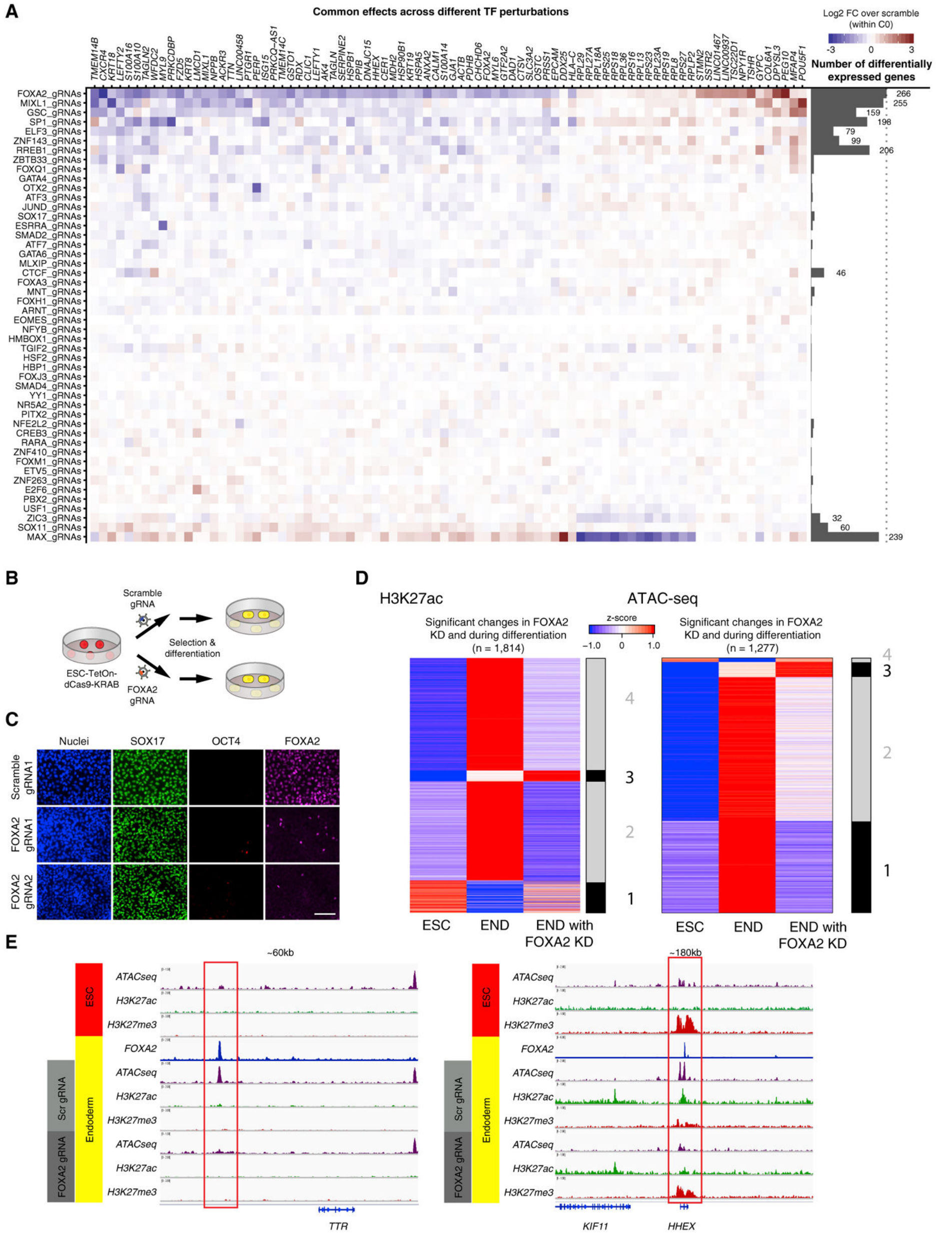


Figure 3. Loss of *FOXA2* Expression Results in Distinct Transcriptomic and Chromatin Changes within END

(A) A gene module affected across multiple perturbations. Within the main cluster only, log fold changes (color intensity) were estimated via MIMOSCA. Genes were selected by running sparse PCA and thresholding the top component. Bars to the right show the number of differentially expressed genes by target (MAST $q < 0.05$; cluster 0 only).

(B) Schematic of *FOXA2* perturbation in conjunction with scramble-gRNA control conditions during differentiation.

(C) Immunofluorescence analysis for SOX17, OCT4, and FOXA2 of scramble and *FOXA2* perturbed END. Nuclei are counterstained with Hoechst. Scale bar: 100 μm.

(D) Summary of dynamic changes (FDR < 0.05, log fold change ≥ 1.0) in H3K27ac (left) and chromatin accessibility as measured by ATAC-seq (right) between ESC, scramble-gRNA control END, and *FOXA2* perturbed END (n = 2 biological replicates).

(E) ATAC-seq, histone ChIP-seq, and TF ChIP-seq data at the *TTR* and *HHEX* loci showing decreased gene activity in *FOXA2*-gRNA containing END at foregut-associated genes that are potentially regulated by *FOXA2*. See also Tables S1 and S4.

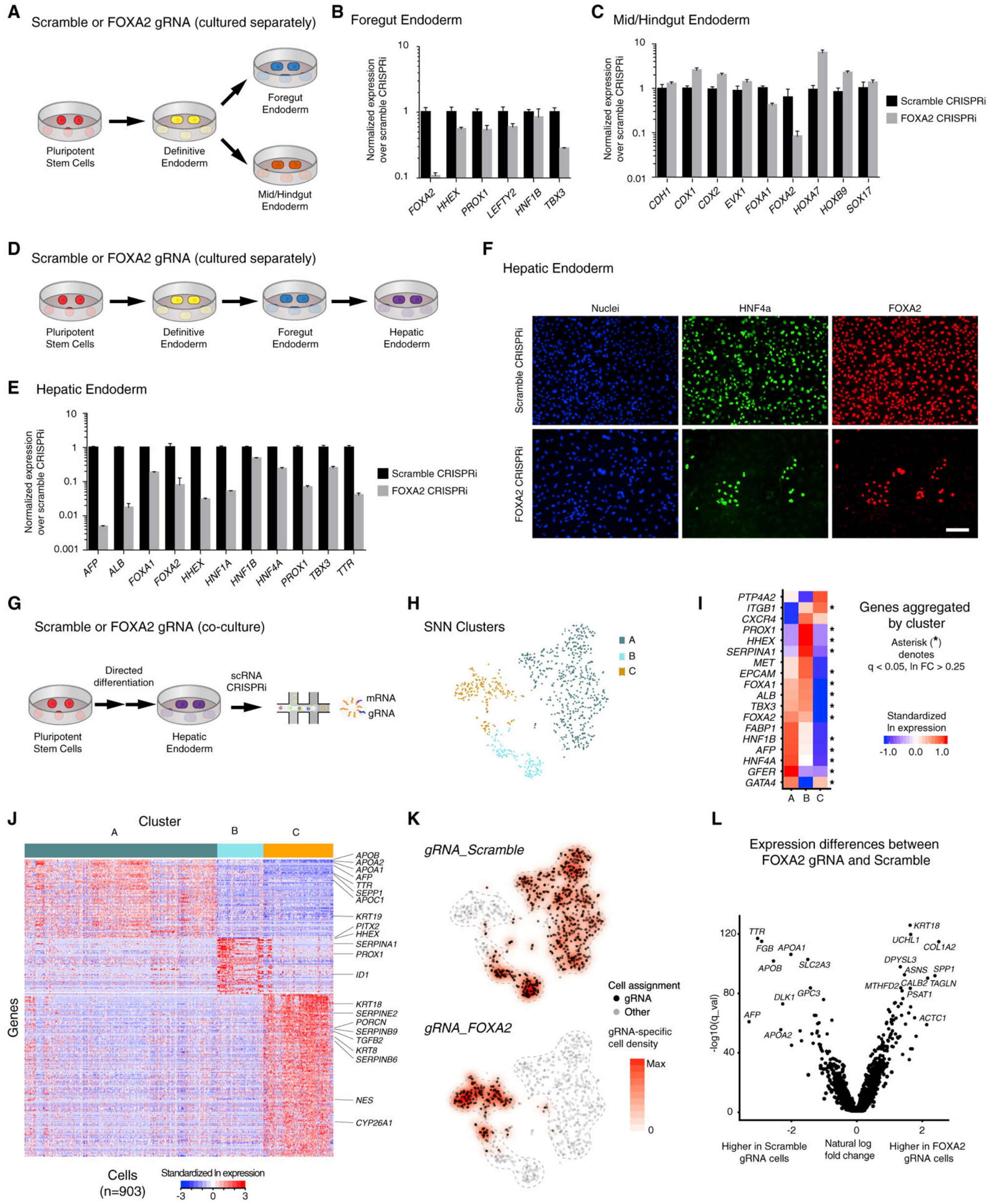


Figure 4. Loss of *FOXA2* Impairs Differentiation to Foregut END and Subsequent Hepatic END, while Mid-Hindgut END Differentiation Is Unaffected

(A) Schematic of perturbation experiments for assessment of effect of *FOXA2* repression on competency toward foregut and mid-hindgut END.

(B and C) qPCR analysis of transcripts in scramble or *FOXA2* perturbed foregut END (B) and mid-hindgut END (C). Relative transcript expression was calculated using the ΔΔCT method; all transcripts were normalized to *ACTB*. Error bars correspond to SD; n = 3 biological replicates.

(D) Schematic of perturbation experiments during human hepatic END differentiation.

(E) qPCR analysis of transcripts in scramble or *FOXA2* perturbed hepatic END. Relative transcript expression was calculated using the ΔΔCT method; all transcripts were normalized to *ACTB*. Error bars correspond to SD; n = 3 biological replicates.

(F) Immunofluorescence analysis for HNF4a and FOXA2 of scramble and *FOXA2* perturbed hepatic END cells. Nuclei are counterstained with Hoechst. Scale bar: 100 μm.

(G) Schematic of the scRNA-seq CRISPRi experiment during hepatic END differentiation. Scramble- and *FOXA2*-gRNA cells were co-cultured throughout differentiation.

(H) Unsupervised cluster assignments visualized via tSNE. Each dot represents one cell (n = 2 biological replicates).

(I) Heatmap of cluster average expression of selected hepatic END-associated transcripts. Asterisk denotes MAST $q < 0.05$; ln FC > 0.25.

(J) Cellwise heatmap (n = 903 cells) containing all cluster markers with MAST $q < 0.05$, FC > 1.5 in either direction, and detection rate > 10% in at least one cluster. Each transcript is standardized to have mean 0 and variance 1.

(K) Feature plots of scramble- and *FOXA2*-gRNA.

(L) Expression differences from single-cell data between *FOXA2*-gRNA versus scramble-gRNA containing hepatic END cells. All transcripts are displayed. See also Figure S3 and Table S5.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
SOX17 antibody	R&D Systems	Cat# AF1924; RRID: AB_355060
OCT3/4 antibody	Santa Cruz Biotechnology	Cat# sc-5279; RRID: AB_628051
FOXA2 antibody	EMD Millipore	Cat# 07–633; RRID: AB_390153
HNF4A antibody	Abeam	Cat# ab41898; RRID: AB_732976
HHEX antibody	R&D Systems	Cat# MAB83771; RRID: Not found
TBX3 antibody	Santa Cruz Biotechnology	Cat# sc17871; RRID: AB_661666
PROX1 antibody	R&D Systems	Cat# AF2727; RRID: AB_2170716
CDX2 antibody	BioGenex	Cat# MU392A-UC; RRID: AB_2650531
EPCAM antibody	Biolegend	Cat# 324202; RRID: AB_756076
HA antibody	Cell Signaling Technologies	Cat# 3724; RRID: AB_1549585
GAPDH antibody	R&D Systems	Cat# 2275-PC-100; RRID: AB_2107456
H3K27ac antibody	Diagenode	Cat# C15410196; RRID: AB_2637079
H3K27me3 antibody	Millipore	Cat# 07–449; RRID: AB_310624
Goat anti-Rabbit IgG (H+L), HRP	Thermo	Cat# A16104; RRID: AB_2534776
Donkey anti-goat Alexa 488	Thermo Scientific	Cat# A11055; RRID: AB_2534102
Donkey anti-mouse Alexa 488	Thermo Scientific	Cat# A21202; RRID: ABJ41607
Donkey anti-rabbit Alexa 488	Thermo Scientific	Cat# A21206; RRID: AB_2535792
Donkey anti-goat Alexa 594	Thermo Scientific	Cat# A11058; RRID: AB_2534105
Donkey anti-mouse Alexa 594	Thermo Scientific	Cat# A21203; RRID: AB_2535789
Donkey anti-rabbit Alexa 594	Thermo Scientific	Cat# A21207; RRID: ABJ41637
Donkey anti-rabbit Alexa 647	Thermo Scientific	Cat# A31573; RRID: AB_2536183
Chemicals, Peptides, and Recombinant Proteins		
Activin A	R&D Systems	Cat# 338-AC-01M
CHIR99021	Tocris	Cat# 4423
LDN193189	Tocris	Cat# 6053
PI-103	Tocris	Cat# 2930
FGF4	R&D Systems	Cat# 235-F4–025
FGF10	R&D Systems	Cat# 345-FG-025

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BMP4	R&D Systems	Cat# 314-BP-050
Doxycyoline hyclate	Sigma	Cat# D9891
Y27632	Tocris	Cat# 1254
Puromycin	Thermo	Cat# A11138
B27 minus vitamin A	Thermo	Cat# 12587010
Growth Factor Reduced Matrigel	Corning	Cat# 356231
Hoechst dye	Invitrogen	Cat# H3570
mTeSRI	StemCell Technologies, Inc.	Cat# 05850
hES-qualified matrigel	Corning	Cat# 354277
ReLeSR	StemCell Technologies, Inc.	Cat# 05872
TrypLE Express Enzyme (1X), no phenol red	Thermo	Cat# 12604
Penicillin and streptomycin	Corning	Cat# 30–002-CI
Non-essential amino acids	Thermo	Cat# 11140
GlutaMAX	Thermo	Cat# 35050
DMEM	Thermo	Cat# 11965
OptiMEM	Thermo	Cat# 31985
RPMI1640	Thermo	Cat# 21870
FBS	GE Healthcare	Cat# SH30070.03
10% formalin	Fisher	Cat# 032–060
PBS	Sigma	Cat# S0389
Donkey serum	Lampire	Cat# 7332100
Triton X-100	Sigma	Cat# X100
Trypsin-EDTA	GIBCO	Cat# 25200
7-AAD Viability Staining Solution	BioLegend	Cat# 420403
dUTP	Thermo	Cat# R0133
Uracil-DNA Glycosylase	NEB	Cat# M0280
Dynabeads, Protein A	Thermo	Cat# 10002D
Bovine Serum Albumin solution	Sigma	Cat# A8412
Droplet generation oil	Biorad	Cat# 1864006
Barcoded Oligo-dT beads	Chemgenes	Cat# MACOSKO-2011–10
Agencourt AMPure XP beads	Beckman Coulter	Cat# A63881

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical Commercial Assays		
DNeasy Blood and Tissue Kit	QIAGEN	Cat# 69504
QIAprep Spin Miniprep Kit	QIAGEN	Cat# 27106
STEMdiff Definitive Endoderm Kit	StemCell Technologies, Inc.	Cat# 05110
GoTaq DNA Polymerase	Promega	Cat# M3001
HiFi HotStart ReadyMix PCR Kit	Kapa Biosystems	Cat# KK2600
NEBuilder HiFi DNA Assembly Master Mix	NEB	Cat# E2621
TranslT-293 transfection reagent	Mirus	Cat# 2700
Ribo-Zero rRNA Removal Kit	Epicenter	Cat# MRZH116
Qubit High Sensitivity DNA assay kit	Invitrogen	Cat# Q32854
Nextera DNA Library Prep Kit	Illumina	Cat# 15028212
Nextera XT DNA Library Preparation Kit	Illumina	Cat# FC1311024
EZ DNA Methylation-Gold Kit	Zymo	Cat# D5005
Accel-NGS Methyl-Seq DNA Library Kit	Swift Biosciences	Cat# 30024
Chromium Single Cell 3′ Library & Gel Bead Kit v2	10XGenomics	Cat# 120237
Chromium Single Cell A Chip Kit	10XGenomics	Cat# 120236
QuantSeq 3′ mRNA-Seq Library Prep Kit FWD for Illumina	Lexogen	Cat# 015.96
Superscript III First-Strand Synthesis System	Thermo	Cat# 18080051
SYBR FAST qPCR Master Mix (2X) Universal	Kapa Biosystems	Cat# KK4600
P3 Primary Cell 4D-Nucleofector Kit L	Lonza	Cat# V4XP-3024
Nextseq500/550 High Output v2 75-cycle kit	Illumina	Cat# FC-404–2005
Deposited Data		
Raw and processed data	GEO:GSE127202	GEO link
Experimental Models: Cell Lines		
H1 hESCs	WiCell	Cat# WA01; RRID: CVCL_9771
H1-AAVS1-TetOn-dCas9-KRAB hESCs	This paper	N/A
HEK293T/17	ATCC	Cat# CRL-11268; RRID: CVCL_1926
Oligonucleotides		
AAVS1-WT-Fwd: CGGTTAATGTGGCTCTGGTT	This paper	N/A
AAVS1 -WT-Rev: AGGATCCTCTCTGGCTCCAT	This paper	N/A
AAVS1 -Targeted-Fwd: TCGACTTCCCCTCTTCCGATG	This paper	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
AAVS1-Targeted-Rev: GAGCCTAGGGCCGGGATTCTC	This paper	N/A
dCas9-KRAB-Fwd:GGCCGAGAATATCATCCACCTG	This paper	N/A
dCas9-KRAB-Rev: CTGGTATCCGAGACTGACGAG	This paper	N/A
U6-sgRNA-Fwd:GTCTCGTGGGCTCGGAGATGTGTAT AAGAGACAGTGTGGAAAGGACGAAACACC	This paper	N/A
Illumina-PE-P5-Rev: AATGATACGGCGACCACCGAGAT CTACACTCTTTCCCTACACGACGCTCTTCCGAT*C*T	This paper	N/A
N707-lndex:CAAGCAGAAGACGGCATACGAGATGTA GAGAGGTCTCGTGGGCTCGG	This paper	N/A
N708-lndex:CAAGCAGAAGACGGCATACGAGAT CCTCTCTGGTCTCGTGGGCTCGG	This paper	N/A
Recombinant DNA		
CROPseq-Guide-Puro	Addgene	Cat# 86708
pAAVS1-NDi-CRISPRi	Provided by B. Conklin	N/A
pHRdSV40-dCas9-10xGCN4_v4-P2A-BFP	Addgene	Cat# 60903
pAAVSI -TetOn-dCas9-KRAB	Addgene	Cat# 115545
pENTR2B-dCas9-KRAB	Addgene	Cat# 115547
pHAGE EF1a dCas9-KRAB	Addgene	Cat# 50919
Software and Algorithms		
10XCellranger	10XGenomics	V2.1.0
STAR aligner	https://github.com/alexdobin/STAR	V2.4.2
Picard tools	http://broadinstitute.github.io/picard/	V1.96
Samtools	http://www.htslib.org/	v1.3
Drop-seq tools	http://mccarrolllab.org/dropseq	v1.0
hisat2	https://ccb.jhu.edu/software/hisat2/index.shtml	V2.0.5
ESAT	https://www.umassmed.edu/garberlab/software/esat/	v0.1
R	R Foundation for Statistical Computing, Vienna, Austria	V3.4.3
Seurat	https://satijalab.org/seurat/	V2.3.3
MAST	https://www.bioconductor.org/packages/release/bioc/html/MAST.html	v1.8.2
Thymusatlastools2 and custom scripts	To be released before publication; https://github.com/maehrlab	v0.0.0.9000

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Seriation	https://cran.r-project.org/web/packages/seriation/index.html	v1.2–3
elasticnet	http://users.stat.umn.edu/~zouxx019/software.html#elasticnet	1.1
Python	www.nature.com/ https://conda.io/en/latest/user-guide/install/download.html	V2.7.13
MIMOSCA	https://github.com/asncd/MIMOSCA	Not formally versioned; Git commit is27199eb
ScreenProcessing	https://github.com/mhorlbeck/ScreenProcessing	Not formally versioned; Git commit is 50628c7
quantreg	https://cran.r-project.org/web/packages/quantreg/index.html	v5.38
Other		
Nikon Eclipse Ti Laser-scanning fluorescence microscope	N/A	N/A
Nikon SMZ1500 with Nikon Intensilight Epi-fluorescence Illuminator.	N/A	N/A
PDMS co-flow microfluidic droplet generation device	Nanoshift	custom built based on datafile 1 from Macosko et al. (2015)
Fragment analyzer high sensitivity DNA chip	Advanced Analytical	N/A
Nextseq500	Illumina	N/A
10X Genomics Chromium Controller	10X Genomics	N/A