Title: Chromatin regulatory dynamics of early development and regional specification in a directed differentiation model of the human small intestine

Authors: Yu-Han Hung[1], Sha Huang[2,3], Michael K. Dame[3], Jason R. Spence[2,3], Praveen Sethupathy[1]

[1]Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA
[2]Department of Cell and Developmental Biology, University of Michigan, Ann Arbor, MI, USA
[3]Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, MI, USA

Correspondence: Praveen Sethupathy, Address: Department of Biomedical Sciences, Cornell University, Ithaca, NY, USA. E-mail: pr46@cornell.edu, Phone: 607-253-4375

Grant Support: ADA Pathway to Stop Diabetes Research Accelerator (1-16-ACE-47 to P.S.); Empire State Stem Cell Fund (C30293GG to Y.-H. H.); Intestinal Stem Cell Consortium (U01DK103141 to J.R.S.); a collaborative research project funded by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and the National Institute of Allergy and Infectious Diseases (NIAID), the NIAID Novel Alternative Model Systems for Enteric Diseases (NAMSED) consortium (U19AI116482 to J.R.S.); the support from the University of Michigan Center for Gastrointestinal Research (UMCGR) (NIDDK 5P30DK034933).

ABSTRACT
The appropriate development of the small intestine (SI) is critical for efficient nutrient absorption and barrier function after birth. Most of the molecular features and regional patterning of the SI are programmed very early in prenatal development. However, the chromatin regulatory dynamics that underpin early SI development in humans is largely unknown. To fill this knowledge void, we apply a cutting-edge genomic technology to a state-of-the-art model of human SI development. Specifically, we leverage chromatin run-on sequencing (ChRO-seq) to define the landscape of active transcriptional regulatory elements across early stages of directed differentiation of human pluripotent stem cells (hPSCs) into SI organoids. Through comprehensive bioinformatic analysis of the data we provide the first-ever view of the changing chromatin regulatory landscape and define stage-specific key enhancer hotspots during human SI development. We also identify candidate transcription factors and their cistromes that are associated with the acquisition of SI identity and the initiation of regional patterning. This work offers a rich resource for studying transcriptional regulation of early human SI development.

KEY WORDS: small intestine; development; chromatin run-on sequencing (ChRO-seq); enhancer; gene regulation

INTRODUCTION

The embryonic development of the small intestine (SI) is critical for a fetus to thrive and grow. The genetic programming of SI cell identity and regional specification during early development is fundamental to ensure proper SI morphogenesis and maturation with specialized functions, including nutrient digestion and absorption, energy balance, and pathogen defense. Several seminal studies have identified important molecular regulators associated with gut development[1,2], and more recent studies have leveraged advanced genomic technologies (e.g., single cell RNA-seq and ATAC-seq) to provide insights at a more granular level into gut development[3-7]; however, this work relies almost exclusively on animal models. Studies of human SI development have been few, due in large part to limited access to primary human fetal tissues. Early time points of gut development in the human are essentially unexplored. In this study, we sought to fill this important knowledge gap by profiling for the first time the chromatin regulatory dynamics and transcriptional programs that are associated with early development of the human SI.

Robust temporal and spatial regulation of gene expression is fundamental to all biological processes including development[8,9]. Transcriptional programs are precisely controlled by promoters and distal *cis*-regulatory regions known as enhancers. Enhancers harbor binding sites for transcription factors (TFs), activate long-range gene transcription, and are often cell-type specific[10,11]. RNA polymerases are known to be recruited to active enhancers, generating divergent short transcripts (also known as enhancer RNA or eRNAs)[12,13]. Recently, an approach called chromatin run-on sequencing (ChRO-seq)[14] was developed for genome-wide identification of promoters, active enhancers, and actively transcribed gene bodies in a single assay. ChRO-seq represents the newest generation of nascent RNA sequencing technologies, and overcomes several limitations of previous versions including global run-on sequencing (GRO-seq)[15] and precision run-on sequencing (PRO-seq)[16]. ChRO-seq was very recently successfully applied to archived solid tumor tissues to define the distinct regulatory landscapes across different subtypes of glioblastoma multiforme[14].

Advances in the directed differentiation of human pluripotent stem cells (hPSCs), including human embryonic stem cells (hESCs), provide a powerful strategy for studying the early development of human SI[17,18]. Human SI organoids generated from hPSCs recapitulate many aspects of *in vivo* SI development[19,20], including embryonic patterning into different regions of the GI tract[21], and ultimately exhibit molecular, structural, and functional features similar to those of the human fetal SI[22]. While hPSC differentiation into different regions of the SI has been described, the molecular mechanisms governing this pattering is unclear. To reveal temporal dynamics of the chromatin state during early human SI development and differentiation, we performed ChRO-seq in the early stages of the hPSC-derived SI model, including hESC, definitive endoderm (DE), duodenal spheroid (primitive SI with proximal regional identity), and ileal spheroid (primitive SI with distal regional identity). This study provides the first-ever view of the changing chromatin regulatory landscape and defines stage-specific key enhancer hotspots during human SI development. We also identify candidate key transcription factors and their cistromes that are associated with the acquisition of SI identity and the initiation of regional patterning. Moreover, we offer an unprecedented resource for the research community to develop and test targeted hypotheses about key regulatory hotspots and molecular drivers of early events of human SI development.

RESULTS

Generation of SI spheroids by directed differentiation of hPSCs

The directed differentiation of hPSCs (here we used H9 hESCs) into SI organoids was carried out as previously described[19,21] **(Figure 1A)**. For genomic analysis, we included four early stages of the hPSC-derived SI model: hESC, DE, duodenum (Duo) spheroid, and ileal (Ile) spheroid[21]. These region-patterned spheroids, which represent the primitive fetal SI, are comprised of mainly stem/progenitor cells[23] that can give rise to different mature lineages after prolonged organoid culture, or following transplantation into a

murine host[24]. We performed ChRO-seq, together with deep RNA-seq, on these four stages in order to define the chromatin and gene expression dynamics in the key events of human SI development: DE formation (DE vs. hESC), SI identity acquisition (Duo spheroid vs. DE), and regional specification (Ile vs. Duo spheroid) **(Figure 1A)**. ChRO-seq enables the quantification of genome-wide nascent transcriptional activity at gene bodies, promoters, and enhancers, whereas RNA-seq enables only determination of steady state levels of gene expression **(Figure 1A)**.

ChRO-seq reveals temporal dynamics of nascent transcription at gene loci during early development of human SI

To assess nascent transcription of genes, we first analyzed ChRO-seq signals within the body of annotated genes. Hierarchical clustering analysis and principal component analysis (PCA) of the gene transcriptional profiles is sufficient to cleanly stratify different stages of the hPSC-derived SI model **(Figure 1B and C)**. The ChRO-seq and RNA-seq signal for genes encoding transcription factors (TFs) that are well-established protein markers of specific stages indicate the appropriate specificity **(Figure 1D, Supplementary Figure 1A)**. Moreover, global comparison of the ChRO-seq and RNA-seq profiles across all stages **(Figure 1E)** and within each stage **(Supplementary Figure 1B)** reveal a robust correlation, indicating that nascent transcription profiles globally reflect steady state expression profiles in this model system (albeit not perfectly, as expected due to other layers of gene regulation such as those at the post-transcriptional level).

Next we sought to identify from the ChRO-seq data differentially transcribed genes during DE formation, SI identity acquisition, and regional specification to the Ile region (DESeq2: padj < 0.2, $p < 0.05$, $\log_2$ fold change > 1, average TPM > 25) **(Figure 1F-L)**. During the transition from hESC to DE, 1886 genes (1328 up; 558 down) are altered **(Figure 1F)** and the Gene Ontology (GO) term analysis reveals that the up-transcribed genes in DE are enriched in pathways that drive endoderm formation **(Figure 1G)**. Upon the formation of Duo spheroids from DE, 4024 genes (2546 up; 1478 down) are altered and the up-transcribed genes are enriched in glucose metabolism processes **(Figure 1H and I)**. In the comparison of Ile relative to Duo spheroids, there are as expected fewer genes that are significantly altered (total 362 genes; 182 up; 180 down) **(Figure 1J)**. Although the GO term enrichment analysis of up-transcribed genes in Ile spheroids reveals only general transcriptional processes **(Figure 1K**), and although canonical WNT targets (*CTNNB1* and *AXIN2*) are not altered **(Supplementary Figure 1C)**, pathway enrichment analyses with the BioPlanet 2019 and NCI-Nature 2016 databases show a significant over-representation in the AP-1 TF network **(Figure 1L**), which likely reflects an increase in non-canonical WNT signaling. The pathway analyses of down-transcribed genes in all relevant comparisons are shown in **Supplementary Figure 1**. Also, the results of all of the same analyses carried out on RNA-seq data are shown in **Supplementary Figure 2**.

Identification of stage-specific markers at the levels of both transcription and steady state expression during early development of human SI

First we demonstrated from the ChRO-seq data that markers of the esophagus, lung, stomach, liver, pancreas, and colon lineages are transcribed at low levels, as expected, in the Duo and Ile spheroids **(Supplementary Figure 3A and 3B)** compared to *CDX2* **(Supplementary Figure 1A**), which is a known marker of SI. Next we determined that previously reported markers of human fetal duodenum and ileum[21] **(Supplementary Figure 3C)** are not sufficient to mark SI spheroids but rather only mature SI organoids after culture for 28 days **(Supplementary Figure 3D)**. This likely reflects that fact that spheroids, which are newly differentiated, represent the early gut lineage and have not been given time *in vitro* to mature into SI organoids, which are typically cultured for ~1 month prior to analysis. Given the lack of known markers distinguishing Duo from Ile spheroids, we next sought to fill this knowledge void.

To identify genes that label a specific stage during early human SI development at the levels of both transcription and steady-state expression, we performed integrative analysis of the ChRO-seq and RNA-seq data (see Methods). The results of hierarchical clustering analysis of those genes actively transcribed (TPM > 50) in at least one stage are shown in **Figure 2A**. We identified genes that are significantly elevated

according to both transcription and expression in each stage relative to all other stages: hESC (n=239), DE (n=149), SI spheroid irrespective of regional identity (n=88), Duo spheroid (n=9), and Ile spheroid (n=40) **(Figure 2B; Supplementary Data 2)**. The stage-specific genes we identified include well-established markers such as *POU5F1* and *SOX2* for hESCs **(Figure 2C)**, *NODAL*, *SOX17*, *GATA6*, *EOMES* and *LEFTY2* for DE **(Figure 2D)**, and *CDX2* for SI spheroid irrespective of regional identity **(Figure 2E)**. We defined many novel stage-specific markers including *HES4* and *HOXB* family members for SI spheroids irrespective of regional identity, *FOXJ1* for Duo spheroids, and *FJX1* for Ile spheroids. We also identified several long, non-coding RNA (lncRNA) and anti-sense transcript markers: *LINC00678* for hESCs (**Figure 2C**), *LINC00543* for SI spheroids irrespective of regional identity **(Figure 2E)**, and *EVX1-AS* for Ile spheroids (**Figure 2G**).

Active transcriptional regulatory element landscapes reveal chromatin re-wiring during early development of human SI

Active transcriptional regulatory elements (TREs), including promoters and enhancers, are identified in ChRO-seq data by the hallmark feature of short bidirectional transcription **(Figure 3A)**. To identify active TREs across the entire genome in each stage, we employed dREG[25], which was developed specifically for this purpose. Using this method, we identified a total of 125,863 active TREs across all four stages included in this study. The length distribution of these active TREs is consistent with what has been reported previously[26] **(Figure 3B)**. We found that, as expected, the vast majority of the active TREs are located in intergenic regions, introns, and annotated transcription start sites (TSS) **(Figure 3C)**. PCA showed that active TRE profiles are sufficient to cleanly stratify samples based on developmental stage and SI regional identity **(Supplementary Figure 4A)**. We next categorized the identified TREs into 49,855 proximal and 76,008 distal TREs, which from here on in we refer to as promoters and enhancers, respectively **(Figure 3A)**. Then we carried out unsupervised hierarchical clustering analyses using profiles of promoters plus enhancers, promoters only, and enhancers only **(Figure 3D-F)**. We observed that enhancer profiles stratify different stages more accurately than promoters or all TREs **(Figure 3D-F, Supplementary Figure 4B-D),** consistent with the notation that enhancer signature is the most cell-type specific[8,10].

Identification of genes associated with stage-specific TREs during the process of SI identity acquisition and regional patterning

Next we defined stage-specific TREs, determined the density of stage-specific TREs for every transcribed gene in each stage, and identified genes associated with the most stage-specific TREs **(Figure 4A; see Methods)**. We identified a total of 4166 Duo-specific and 540 Ile-specific active TREs **(Figure 4B and F)**. We confirmed that, as expected, genes associated with a greater number of Duo-specific TREs also exhibit greater increases in transcription (ChRO-seq signal) in Duo spheroids relative to DE (**Figure 4C**). Similarly, genes associated with a greater number of Ile-specific TREs also exhibit greater increases in transcription (ChRO-seq signal) in Ile spheroids relative to Duo **(Figure 4G)**. Of the 3317 genes highly transcribed in Duo, 195 are associated with at least one Duo-specific TRE, are significantly up-transcribed (ChRO-seq) relative to DE, and significantly increased in steady-state expression (RNA-seq) relative to DE **(Figure 4D)**. As expected, *CDX2* is one of the top genes ranked by the density of Duo-specific TREs **(Figure 4 E)**. Notably, most of the other top ranked genes are HOXB family members **(Figure 4E)**. Of the 3418 genes highly transcribed in Ile, only 48 are associated with at least one Ile-specific TRE, are significantly up-transcribed (ChRO-seq) relative to Duo, and significantly increased in steady-state expression (RNA-seq) relative to Duo **(Figure 4H)**. The genes associated with the greatest number of Ile-specific TREs include members of the *HOXA*, *C* and *D* families, factors involved in canonical or non-canonical WNT signaling *(JUND*, *FJX1* and *CSRNP1*), as well as *HOTTIP* **(Figure 4I)**. Several of these genes were identified as Ile spheroid-specific markers **(Figure 2G)**. The results of similar analyses focused on hESCs in comparison to DE, DE in comparison to hESCs, and Ile in comparison to DE are shown in **Supplementary Figure 5B-M**.

The same analyses as described above using the active enhancers only reveals the same temporal dynamics of the HOX cluster genes **(Supplementary Fig. 6)**. Indeed, we detected high transcriptional activity (ChRO-seq) in Duo spheroids and high density of nearby Duo-specific TREs for HOXB genes **(Figure 5A and B)**. The increase in transcription of each HOXB family member in Duo relative to DE correlates with the density of Duo-specific TREs **(Figure 5C)**. Also, we observed dramatic increases in transcriptional activity (ChRO-seq) in Ile spheroids and high density of Ile-specific TREs for HOXA, C, and D family members **(Figure 5D and E)**. The increase in transcription of each HOXA/C/D family member in Ile relative to Duo tracks closely with the density of nearby Ile-specific TREs **(Figure 5F)**. Together, these observations indicate an initial activation of HOXB during SI identity acquisition, likely driven by the identified nearby Duo-specific TREs, followed by the activation of the other HOX clusters during ileal specification, likely driven by the identified nearby Ile-specific TREs.

Identification of stage-specific enhancer hotspots associated with SI identify acquisition and ileal regional patterning

It has been shown in previous studies that dense clusters of highly active enhancers (which we refer to as 'hotspots') occur nearby to genes that are especially critical for defining cell identity and status[27-29]. We sought to define for the first time the changing landscape of enhancer hotspots in a model of human SI development. To accomplish this, we adapted a previously described methodology[27,28], which requires several different histone modification ChIP-seq datasets, to work with ChRO-seq data instead **(see Methods; Supplementary Data 3)**.

To identify enhancer hotspots and the candidate genes which they may regulate in SI identity acquisition **(Figure 6A)**, we first analyzed Duo-specific distal TREs to define 'stitched enhancers' specific to Duo spheroids relative to DE. Among the 2389 stitched enhancers, we identified 145 that exhibit strong enough transcriptional activity to be designated as Duo-specific 'enhancer hotspots' (**Figure 6B)**. The distribution of the distance between every gene and its nearest Duo-specific enhancer hotspot is shown in **Figure 6C**. We assigned every stitched enhancer in Duo spheroids (non-hotspots and hotspots) to the nearest gene and found that the overall increase in transcription is significantly greater for the set of genes associated with Duo-specific hotspots compared to those associated with stitched enhancers that are not hotspots **(Figure 6D)**. This finding is consistent with the notion that hotspots exert particularly strong effects on transcription. Among the 110 genes linked to Duo-specific enhancer hotspots, 30 exhibit highly significant increases in both transcription (ChRO-seq) and steady state expression (RNA-seq) during the transition from DE to Duo spheroid **(Figure 6E and F)**. Many of these 30 (e.g., *CDX2*, *FOXA1*, *HOXB* family members, and *LINC00543*) were defined earlier as SI or Duo-specific markers **(Figure 2)**.

We next performed a similar analysis to identify enhancer hotspots and nearby genes associated with ileal regional specification. Among 310 stitched enhancers formed by Ile-specific enhancers (relative to Duo spheroids), we defined 29 Ile-specific enhancer hotspots **(Figure 6G)**. The distribution of the distance between every gene and its nearest Ile-specific enhancer hotspot is shown in **Figure 6H**. Similar to the observation made for SI fate acquisition, the genes nearest to Ile-specific enhancer hotspots exhibit a significantly greater increase in transcription compared to those linked to non-hotspot stitched enhancers **(Figure 6I)**. Among the 31 genes linked to Ile-specific enhancer hotspots, 12 exhibit highly significant increases in both transcription (ChRO-seq) and steady state expression (RNA-seq) in Ile relative to Duo spheroids **(Figure 6J)**. These genes include *HAND2*, *HOXC/D* family members, *FJX1*, *DLX5*, and *PITX1* **(Figure 6K)**, many of which were also identified earlier as Ile-specific markers **(Figure 2)**. Similar analyses were performed for the hESC and DE stages and the results are summarized in **Supplementary Figure 7**.

Identification of candidate TFs and cistromes associated with SI identity acquisition and ileal regional patterning

To discover putative key transcription factor (TF) drivers and their cistromes (the genome-wide set of TF targets) along the early developmental axis of the hPSC-derived SI model, we developed and executed the following bioinformatic pipeline **(Figure 7A, see Methods)**: (1) Motif enrichment analysis was performed by HOMER in stage-specific TREs **(Supplementary Data 4)**, (2) The enriched binding motifs were filtered further based on the RNA-seq expression levels of the cognate TFs, and (3) The cistromes with different candidate TFs of interest were defined and compared across different TFs.

We first sought to identify candidate TF-motif modules that are associated with SI identity acquisition. Among the 4166 Duo-specific TREs, we detected significant enrichment for the binding motifs of CDX4, CDX2, and many forkhead box TFs including FOXA1 and FOXM1 **(Figure 7B; Supplementary Figure 9G)**. The motif enrichment analyses using only Duo-specific enhancers generated similar results **(Supplementary Figure 8)**. We chose the top ranked motifs associated with CDX4, CDX2, FOXA1 and EP300 for cistrome analysis **(Figure 7C; Supplementary Figure 10)**, which led to several notable observations. First, *FOXA1* and *CDX2* are associated with Duo-specific TREs that harbor FOXA1 and CDX2 binding sites, strongly suggesting that the transcription of these two TFs may be activated and maintained in part through an auto-regulatory mechanism. Second, we found that although the genes associated with Duo-specific TREs that contain CDX2 or CDX4 binding sites are largely overlapping as expected, there are some genes uniquely associated with only one or the other, indicating that there may be some distinct functional roles in SI fate acquisition. Third, many of the genes we had identified as markers of SI spheroids (and even specifically Duo-spheroids), including *HOXB* genes, *FOXJ1*, *WFDC2*, and *LINC00543*, are indeed associated with one or more of the TFs that are associated with SI fate acquisition.

We next performed a similar analysis for ileal regional specification. Among the 540 Ile-specific TREs, we detected significant enrichment for several key TF families **(Figure 7B; Supplementary Figure 9G and 10)**. We observed an enrichment for motifs of HOXD11 and PITX1, which themselves are associated with Ile-specific enhancer hotspots **(Figure 6)**, suggesting that these two factors are associated with Ile-specific chromatin re-wiring both upstream and downstream of their activity. Surprisingly, we also detected significant enrichment for motifs of CDX2 and CDX4 in Ile-specific TREs; suggesting that these two factors have stage-specific binding sites and therefore may have different cistromes for controlling SI fate acquisition and ileal regional specification. Finally, and perhaps most notably, we identified very strong enrichment in Ile-specific TREs of several AP-1 related factors (JUN, JUNB, JUND and FOSL2), which underscores a point made earlier in the GO analysis that the AP-1 transcriptional network may be critical for ileal regional specification **(Figure 1L)**. The motif enrichment analyses using only Ile-specific enhancers generated similar results **(Supplementary Figure 8)**. We next chose the motifs of CDX2, AP-1 factors, HOXD11 and PITX1 for further cistrome analysis **(Figure 7E)**, which yielded several findings. First, many of the genes we had identified as markers of Ile spheroids, including *HAND2*, *DXL5*, *FJX1*, and *HOXA/C/D* family members are indeed associated with one or more of the TFs that we have identified as important in the Ile-specific cistrome. Second, the members of the HOXA, C, and D families are associated with overlapping but distinct TF regulators. For example, while *HOXD9-11* are strongly associated with all four major TFs including AP-1, the *HOXA* genes are associated with all of the TFs except AP-1. Similar analyses were performed for the hESC and DE stages and the results are summarized in **Supplementary Figure 9 and 10**.

Overall, based on the chromatin regulatory dynamics defined by the analyses in this study, we have provided a working model of transcriptional programming that may drive SI fate acquisition and ileal regional patterning **(Figure 8)**.

DISCUSSION

By leveraging the recently developed ChRO-seq technology, we generated for the first time ever comprehensive chromatin regulatory landscapes across the beginning stages of directed differentiation from

hPSC to SI organoids. Specifically, we defined: (1) marker genes that label specific stages along the developmental axis of the human SI, (2) the map of active regulatory elements (promoters and enhancers) as well as enhancer hotspots associated with SI identity acquisition and ileal regional patterning, and (3) candidate key TF drivers and their cistromes relevant to these critical developmental events.

The markers that we identified as labels of SI fate acquisition and Duo or Ile regional identity provide the first glimpse of genes that are involved in early SI development and patterning. Several markers (candidate regulators) of SI spheroids drew our attention. *HES4* is identified for the first time in the context of gut development; one reason for this is that it is absent in the genome of the mouse[30], the model organism most used previously to study gut development. FOXA1, which is involved in endoderm formation in mice, was identified as a marker of SI spheroids and among the genes strongly associated with active enhancers that emerge only during SI identity acquisition. In fact, a recent study linked dysregulation of intestinal *FOXA1* with necrotizing enterocolitis in infants[31], which further supports its functional relevance in human SI development and thereby warrants future investigation. Our finding of *WFDC2* as a marker of SI spheroids is consistent with the enrichment of *Wfdc2* in SI spheroids derived from the murine fetal progenitor population compared to SI organoids derived from murine adult stem cells[32]. Notably, some of the cells in this model are committed to the mesoderm lineage **(Supplementary Figure 1A)**, consistent with previous observations[18,19]. Our marker analysis highlights genes that are traditionally thought to be mesenchymal (e.g., *WNT5A*, *DLX5*, *HAND2*) and have been shown to be critical for early gut development in mice or to be involved in regional patterning in other tissue types. Future experiments are required to confirm the cell type in which these candidate regulators are expressed and have functional roles in human SI development. Future experiments are also required to assess sufficiency and necessity of the stage-specific enhancers and enhancer hotspots identified in this study for controlling transcriptional networks relevant to human SI development and patterning.

Our ChRO-seq analysis highlights the spatiotemporal dynamics of HOX clusters during early development of human SI, unveiling the complexity of HOX biology in this context. We found that the HOXB cluster is associated with the acquisition of the SI fate whereas the other HOX clusters (HOXA, C and D) are associated with the initiation of ileal regional patterning. In fact, our study in the human SI model, together with a recent single cell RNA-seq study in E8.75 mouse gut[3], does not suggest the typical spatial collinearity of HOX coding observed in various organ types[33,34]. Importantly, we were able to identify known and novel TF drivers of HOX genes in the context of human SI development **(Figure 8)**. Our observation that CDX2 likely targets and regulates the HOXB cluster genes has been validated by a very recent CDX2 ChIP-seq study of SI-patterned endoderm monolayer derived from hPSCs[5], demonstrating that ChRO-seq is a very sensitive and robust tool to identify key TFs and their target genes in any particular biological context. Also, our finding supports the finding that ectopic co-expression of *CDX2* and *HOXB* genes in gastric intestinal metaplasia and Barrett esophagus[35-37] leads to the cells acquiring an intestine-like identity. Interestingly, we also found that CDX2 is predicted to regulate HOXA, C and D clusters only during ileal regional specification, which is a new discovery that merits further investigation.

We identified several candidate TF drivers, including PITX1 and AP-1 family members, associated with ileal specification. A recent scRNA-seq study of primary human fetal gut tissues showed that the regulatory network of JUN, JUNB and FOS are more enriched in an early stage of SI development compared to a late stage[4]; however, their roles in early SI development remained unknown. Our findings suggest that AP-1 members likely contribute to early SI development by establishing ileal identity. Notably, multiple AP-1 members are expressed in the H4 cell line established from human fetal ileum[38] and *JunD* is found to be enriched in ileal compared to duodenal enterocytes of adult mice[39], both of which indicate the relevance of AP-1 TFs to ileal properties. Interestingly, AP-1 factors are known to be involved in non-canonical WNT signaling; FJX1 (a marker of Ile identity in our study) functions through the Wnt/planar cell polarity (PCP) pathway to determine anterior-posterior axis in *Drosophila*[40,41] and is expressed in the epithelium of developing gut in mice[42]. Whether these genes/TF networks enriched in Ile spheroids function through

non-canonical WNT pathways and how they interplay with canonical WNT signaling during ileal regional patterning warrant detailed future investigation.

The use of the state-of-art *in vitro* model of human fetal SI is a critical feature of this study given the lack of other existing platforms to study early organogenesis of human SI and the growing appreciation for cross-species differences in developmental processes in vertebrates at the molecular level[43-46]. Furthermore, the genome-wide characterization of the chromatin regulatory landscape by ChRO-seq in this model system generates valuable translational knowledge for a better understanding of initial transcriptional programming of human SI development. This analysis pipeline can also be applied to other hPSC-derived organoid systems to gain insights into dynamic transcriptional programming and chromatin status during early stages of human organogenesis. Most importantly, the identification of the candidate drivers in the present study may ultimately improve methods of generating therapeutic replacement SI as well as molecular therapies for children and possibly adult patients.

ACKNOWLEDGMENTS

We would like to thank members of the Sethupathy laboratory, most notably Dr. Michael Shanahan, Dr. Ajeet Singh, and Dr. Matt Kanke, for helpful comments and feedback on the study and manuscript. We specially thank Dr. Matt Kanke and Dr. Tim Dinh in the Sethupathy laboratory for providing technical consultation on bioinformatics, and Ramja Sritharan in the Sethupathy laboratory for offering technical consultation on the ChRO-seq protocol. Additionally, we are grateful to the Danko laboratory, specifically Ed Rice and Dr. Charles Danko, for helpful conversations on improving implementation of the ChRO-seq protocol. Finally, we also thank Dr. Jen Grenier and the Cornell Transcriptional Regulation & Expression Facility for RNA sequencing support; Michael Dellheim of the University of Michigan BRCF Flow Cytometry Core; Gina Newsome, Erika Katz, Maliha Berner, and Angeline Wu of the Michigan Medicine Translational Tissue Modeling Laboratory; and a University of Michigan funded initiative (Center for Gastrointestinal Research, Office of the Dean, Comprehensive Cancer Center, Departments of Pathology, Pharmacology, and Internal Medicine) with support by the Endowment for Basic Sciences.

We also thank the following funding supports: ADA Pathway to Stop Diabetes Research Accelerator (1-16-ACE-47 to P.S.); Empire State Stem Cell Fund (C30293GG to Y.-H. H.); Intestinal Stem Cell Consortium (U01DK103141 to J.R.S.); a collaborative research project funded by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and the National Institute of Allergy and Infectious Diseases (NIAID), the NIAID Novel Alternative Model Systems for Enteric Diseases (NAMSED) consortium (U19AI116482 to J.R.S.); the support from the University of Michigan Center for Gastrointestinal Research (UMCGR) (NIDDK 5P30DK034933).

AUTHOR CONTRIBUTIONS

Conceptualization, Y.-H.H., P.S.; Cell culture, S.H.; Cell sorting, M.K.D.; Chromatin isolation and library preparation for sequencing studies, Y.-H.H.; Bioinformatic analyses and data curation, Y.-H.H.; Writing (original draft), Y.-H.H., P.S.; Review and editing, Y.-H.H., J.R.S., and P.S.; Supervision, P.S.; Funding acquisition, Y.-H.H., J.R.S., and P.S.

CONFLICTS OF INTEREST

The authors disclose no conflicts.

FIGURE LEGENDS

Figure 1. ChRO-seq defines patterns of nascent transcription of gene bodies during early stages of directed differentiation of human pluripotent stem cells to small intestinal organoids. (A) Overview of the early stages in the hPSC-HIO model and the genomics approaches (ChRO-seq and RNA-seq) used

in the study. (B) Hierarchical clustering analysis of gene transcription profiles (ChRO-seq signal across gene bodies) across multiple stages of development. (C) PCA of gene transcription profiles (ChRO-seq signal across gene bodies) across multiple stages of development. (D) Normalized ChRO-seq signal of genes encoding important TFs during human SI development. (E) Genome-wide correlation of transcribed levels (ChRO-seq) and expressed levels (RNA-seq) of genes. Transcription (ChRO-seq) and expression (RNA-seq) levels of genes are averaged across all stages. No transcription and expression thresholds are used (see also Supplementary Figure 1). (F, H, J) Volcano plots showing differentially transcribed genes in the indicated comparisons. Numbers in red and blue are for up- and down-transcribed genes in the comparison (average TPM > 25, $\log_2$ fold change of transcription > 1, padj < 0.2 and $p < 0.05$ by Wald test; DESeq2). (G, I, K) Pathway enrichment analyses of up-transcribed genes in the indicated comparisons (GO Biological Process 2018) (also see Supplementary Figure 1). (L) Additional pathway enrichment analyses suggest enrichment of AP-1 TF network in genes up-transcribed in Ile compared to Duo spheroids. hESC, n= 3; DE, n = 4; Duo spheroid (Duo), n = 3; Ile spheroid (Ile), n = 3.

Figure 2. Identification of stage-specific markers that label specific stages at both transcribed and expressed levels during early development of the human SI. (A) Hierarchical clustering analysis of genes with TPM > 50 at least in one stage. Color denotes variation across stages (normalized by z-score). (B) Identification of stage-specific markers at both transcribed (ChRO-seq) and expressed (RNA-seq) levels. The total number of markers (center of the donut plot) and the proportion of marker types are shown for each stage (ChRO-seq: TPM > 50 padj < 0.2, $p < 0.05$, fold change > 1.5 by DESeq2 in the stage of interest compared to all the other stages; RNA-seq: base mean > 100, padj < 0.2, $p < 0.05$, fold change > 1.5 by DESeq2 in the stage of interest compared to all the other stages). Also see Supplementary Table 1. (C) Top 50 most variable markers specific to hESC stage. (D) Top 50 most variable markers specific to DE stage. (E) Top 50 markers specific to SI spheroids regardless of region identity. (F) Markers specific to Duo spheroids. (G) Top 50 most variable markers specific to Ile spheroids. In (C)-(F), color denotes variation across stages (normalized by z-score). ChRO-seq study: hESC, n = 3; DE, n = 4; Duo spheroid (Duo), n = 3; Ile spheroid (Ile), n = 3. RNA-seq study: hESC, n = 2; DE, n = 3; Duo spheroid (Duo), n = 6; Ile spheroid (Ile), n = 4.

Figure 3. ChRO-seq defines profiles of active transcriptional regulatory elements (TREs) during early development of the human SI. (A) Schematic for definition and categorization of active TREs. Proximal TREs are considered promoters and distal TREs are considered enhancers. (B) Size distribution of TREs identified across all the samples. (C) Annotations of TRE locations across all the samples. (D-F) Hierarchical clustering analysis of profiles of overall TRE (D), proximal TRE only (E), and distal TRE only (F) across all four stages. Also see Supplementary Figure 4. hESC, n = 3; DE, n = 4; Duo spheroid (Duo), n = 3; Ile spheroid (Ile), n = 3.

Figure 4. Identification of genes associated with stage-specific TREs during SI identity acquisition and Ile specification. (A) Definition of stage-specific TREs ($\log_2$ fold change > 2.5, padj < 0.05 by DESeq2) followed by TRE density calculation (100 kb up and downstream of TSSs) for every gene in the genome. (B) Venn diagram showing stage-specific and shared TREs between DE and Duo spheroids. Duo-specific TREs (n = 4166) represent TREs emerging during SI identity acquisition. (C) Cumulative distribution of Duo-transcribed genes (TPM > 50) grouped into three different density categories of Duo-specific TREs. (D) Identification of genes associated with Duo-specific TREs (n = 195). (E) Bar graph showing genes associated with Duo-specific TREs (n = 195; left panel). Top 20 genes based on TRE density are highlighted (right panel). (F) Venn diagram showing stage-specific and shared TREs between Ile and Duo spheroids. Ile-specific TREs (n = 540) represent TREs emerging during Ile regional specification. (G) Cumulative distribution of Ile-transcribed genes (TPM > 50) grouped into three different density categories of Ile-specific TRE. (H) Identification of genes associated with Ile-specific TREs (n = 48). (I) Bar graph showing genes associated with Ile-specific TREs (n = 48). ChRO-seq study: DE, n = 4; Duo spheroid (Duo), n = 3; Ile spheroid (Ile), n = 3. RNA-seq study: DE, n = 3; Duo spheroid (Duo), n = 6; Ile spheroid (Ile), n = 4.

Figure 5. TRE density analysis reveals distinct HOX clusters associated with SI identity acquisition and Ile regional specification. (A) ChRO-seq transcription levels of HOX cluster genes in Duo spheroids. Color shade of each bar shows the density of Duo-specific TREs (defined in Figure 4B). The dashed line indicates transcribed level of TPM = 50. (B) Normalized ChRO-seq signals in the HOXB cluster in DE and Duo spheroids and the Duo-specific TREs identified around the region. (C) Lollipop graph showing ChRO-seq fold change of HOX genes in Duo spheroids (relative to DE) with color shade indicating the density of Duo-specific TREs (relative to DE). Only genes with TPM > 50 in Duo spheroids are shown. (D) ChRO-seq transcription levels of HOX cluster genes in Ile spheroids. Color shade of each bar shows the density of Ile-specific TREs (defined in Figure 4F). The dashed line indicates transcribed level (TPM > 50). (E) Normalized ChRO-seq signals in the HOXD cluster in Duo and Ile spheroids and the Ile-specific TREs identified around the region. (F) Lollipop graph showing ChRO-seq fold change of HOX genes in Ile spheroids (relative to Duo) with color shade showing the density of Ile-specific TREs (relative to Duo). Only genes with TPM > 50 in Ile spheroids are shown. DE, n = 4; Duo spheroid (Duo), n = 3; Ile spheroid (Ile), n = 3.

Figure 6. Identification of stage-specific enhancer hotspots involved in SI identity acquisition and Ile regional patterning. (A) Enhancer hotspots are defined as dense clusters of enhancers (< 12.5 kb between one and another) with high transcriptional activity. (B) TRE clusters formed by Duo-specific enhancers (relative to DE) are ranked by ChRO-seq transcriptional activity. The clusters with the highest transcription activity are defined as Duo-specific enhancer hotspots (n = 145; blue) and the rest are defined as Duo-specific stitched enhancers (n = 2244; black). (C) Density plot showing distance between the TSS of genes transcribed in Duo (TPM > 50) and their closest Duo-specific stitched enhancers or enhancer hotspots. Dashed line indicates 500 kb. (D) ChRO-seq fold change of genes associated with Duo-specific stitched enhancers or enhancer hotspots (Student's t-test). (E) Identification of genes associated with Duo-specific enhancer hotspots. (F) The Duo-specific enhancer hotspots and their relative position to the closest genes (n = 30) are shown. Negative and positive position indicate upstream and downstream of the TSS, respectively. Dot size denotes transcriptional activity of a given Duo-specific enhancer hotspot. (G) TRE clusters formed by Ile-specific enhancers (relative to Duo) are ranked by ChRO-seq transcriptional activity. The clusters with the highest transcriptional activity are defined as Ile-specific enhancer hotspots (n = 29; yellow) and the rest are defined as Ile-specific stitched enhancers (n = 276; black). (H) Density plot showing distance between the TSS of genes transcribed in Ile (TPM > 50) and their closest Ile-specific stitched enhancers or enhancer hotspots. Dashed line indicates 500 kb. (I) ChRO-seq fold change of genes associated with Ile-specific stitched enhancers or enhancer hotspots (Student's t-test). (J) Identification of genes associated with Ile-specific enhancer hotspots. (K) The Ile-specific enhancer hotspots and their relative position to the closest genes (n = 30) are shown. Negative and positive position indicate upstream and downstream of the TSS, respectively. Dot size denotes transcriptional activity of a given Ile-specific enhancer hotspot. ChRO-seq study: DE, n = 4; Duo spheroid (Duo), n = 3; Ile spheroid (Ile), n = 3. RNA-seq study: DE, n = 3; Duo spheroid (Duo), n = 6; Ile spheroid (Ile), n = 4.

Figure 7. Identification of TFs and their candidate target genes associated with SI identity acquisition and ileal specification. (A) Overview of our approach for identifying candidate stage-specific TF-TRE modules and their associated genes. (B) Binding motifs enriched in Duo-specific TREs (n = 4166) relative to non-Duo-specific TREs (n = 87784) as determined by HOMER (filter criteria: % target sequence > 5 , q-value < 0.05, fold of enrichment > 1.3). Only TFs for which the RNA-seq base mean > 100 in Duo spheroids are shown. (C) Identification of genes associated with active binding motifs of CDX2, CDX4, FOXA1, and EP300 in Duo spheroids relevant to DE. The Venn diagram shows the number of unique or shared genes likely regulated by these TFs. (D) Binding motifs enriched in Ile-specific TREs (n = 540) against non-Ile-specific TREs (n = 76517) as determined by HOMER (filter criteria: % target sequence > 5 , q-value < 0.05, fold of enrichment > 1.3). Only TFs for which the RNA-seq base mean > 100 in Ile spheroids are shown. (E) Identification of genes associated with active binding motif of CDX2, HOXD11, AP-1 (JUND, JUN, JUNB, FOSL2), and PITX1 in Ile relevant to Duo spheroids. The Venn diagram shows

the number of unique or shared genes likely regulated by these TFs. ChRO-seq study: DE, n = 4; Duo spheroid (Duo), n = 3. RNA-seq study: DE, n = 3; Duo spheroid (Duo), n = 6.

Figure 8. Proposed model of transcriptional programming of human SI during development. (A) Early SI development. (B) Ileal regional specification. The arrow types indicate specific regulatory relationships identified through ChRO-seq analyses.

METHODS

Directed differentiation of hESCs. Differentiation of H9 hESCs and organoids was performed as previously published, with minor modifications[20,55]. Briefly, the endoderm was generated by treatment of activin A (100 ng/ml) for 3 consecutive days in Roswell Park Memorial Institute 1640 (RPMI-1640) media supplemented with 0% (v/v), 0.2% (v/v) and 2.0% (v/v) Hyclone defined fetal bovine serum (dFBS). The endoderm cultures then received daily treatments of FGF4 (500 ng/mL) and CHIR99021 (2 μM) for next 10 days. The intestinal spheroids representing fetal duodenum and ileum were collected on day 5 and day 10, respectively[21]. Subsets of spheroids collected at day 5 and day 10 were then cultured in Matrigel with the previously defined intestine growth media[21] for 28 days in order to mature into organoids. The resulting organoids were then prepared for cell sorting to purify EPCAM+ cell population (epithelial fraction) and the sorted cells were processed for RNA-seq library preparation. The cells generated at the stages of hESC, endoderm, duodenal spheroids and ileal spheroids were subject for ChRO-seq and RNA-seq library preparation.

Fluorescence activated cell sorting (FACS). Methods for organoid dissociation into single cells followed by selection of the epithelial component with FACS, were based on previously described procedures[56]. All solutions, including overnight pretreatment of the organoid cultures, contained 10μM Y27632 (Tocris). Matrigel was digested for 30 minutes with cold 4mM EDTA-DPBS and organoids were washed 4X with cold DPBS. Structures were enzymatically dissociated into single cells using the Tumor Dissociation Kit (human) (Miltenyi Biotec) with a gentleMACS™ Octo Dissociator (with heaters; Miltenyi Biotec) for 50 minutes at 37°C. The cell suspension was then washed with 0.5% BSA-2mM EDTA-DPBS over a succession of cell strainers, 100μm, 40μm (Corning) and 20μm (CellTrics) and centrifuged for 5 minutes at 500xg. Cells were labeled with EpCAM phycoerythrin (PE)-conjugated antibody (BioLegend) and an EpCAM isotype-PE control (BioLegend), and were sorted in 0.1% BSA 2mM EDTA-DPBS on a MoFlo Astrios 1 (Beckman Coulter; Brea, California) instrument at the University of Michigan BRCF Flow Cytometry Core facility. Events were first selected with light-scatter and doublet discrimination gating, followed by exclusion of dead cells using 1μM DAPI dilactate (Molecular Probes). EpCAM-PE(+)/DAPI(-) cells were sorted into cold Advanced DMEM/F12 (Invitrogen). Collected cells were reanalyzed for a purity-check and showed greater than 89% viable and 98% EpCAM-PE(+) events. Cells were pelleted at 500xg for 5 minutes and flash-frozen for subsequent RNA isolation.

Chromatin isolation. The chromatin isolation for ChRO-seq library preparation was performed as previously described[14,57]. Briefly, chromatin was isolated from the cells with 1× NUN buffer [0.3 M NaCl, 1 M Urea, 1% NP-40 (w/v), 20 mM HEPES, pH 7.5, 7.5 mM MgCl2, 0.2 mM EDTA, 1 mM DTT, 20 units per mL SUPERase In RNase Inhibitor (Life Technologies, AM2694), 1× Protease Inhibitor Cocktail (Roche, 11 873 580 001)] and incubation at 12 °C on a ThermoMixer for 30 min. Samples were centrifuged at 12,500xg for 30 min at 4 °C. The chromatin pellet was washed 3 times with 1 mL 50 mM Tris-HCl, pH 7.5, supplemented with 40 units per ml RNase inhibitor. Chromatin storage buffer (50 mM Tris-HCl, pH 8.0, 25% glycerol (v/v), 5 mM Mg(CH3COO)2, 0.1 mM EDTA, 5 mM DTT, 40 units per ml RNase inhibitor) was added to each sample. The samples were loaded into a Bioruptor and sonicated to get the chromatin into suspension. Samples were stored at −80 °C before proceeding to ChRO-seq library preparation.

ChRO-seq library and sequencing. After chromatin isolation, the ChRO-seq library preparation closely followed the protocol described previously[58]. Brifely, chromatin from at least 1×10^6 cells per sample in chromatin storage buffer was mixed with an equal volume of 2×chromatin run-on buffer [10 mM Tris-HCl, pH 8.0, 5 mM MgCl2,1 mM DTT, 300 mM KCl, 400 µM ATP (NEB, N0450S), 40 µM Biotin-11-CTP (Perkin Elmer, NEL542001EA), 400 µM GTP (NEB, N0450S), 40 µM Biotin-11-UTP (Perkin Elmer, NEL543001EA), 0.8 units per µL RNase inhibitor, 1% (w/v) Sarkosyl (Fisher Scientific, AC612075000)]. The run-on reaction was incubated at 37 °C for 5 min. The reaction was stopped by adding Trizol LS (Life Technologies, 10296-010) and pelleted with GlycoBlue (Ambion, AM9515) to visualize the RNA pellet. The RNA pellet was resuspended in diethylpyrocarbonate (DEPC)-treated water and heat denatured at 65 °C for 40 s. In the present study, base hydrolysis of RNA was performed by incubating RNA with 0.2N NaOH on ice for 4 min. Nascent RNA was purified by binding streptavidin beads (NEB, S1421S) before and in between the following procedures: (1) 3′ adapter ligation with T4 RNA Ligase 1 (NEB, M0204L), (2) 5′ de-capping with RNA 5′ pyrophosphohydrolase (RppH, NEB, M0356S), (3) 5′ end phosphorylation using T4 polynucleotide kinase (NEB, M0201L), (4) 5′ adapter ligation with T4 RNA Ligase 1 (NEB, M0204L). The resulting RNA fragments were used for a reverse transcription reaction using Superscript III Reverse Transcriptase (Life Technologies, 18080-044) to generate cDNA. cDNA was then amplified using Q5 High-Fidelity DNA Polymerase (NEB, M0491L) to generate the ChRO-seq libraries. Libraries were sequenced (single-end 75x) using the NextSeq500 high-throughput sequencing system (Illumina) at the Cornell University Biotechnology Resource Center. **Supplementary Data 1** provides the mapping statistics of the ChRO-seq experiments.

Total RNA isolation, mRNA-seq library and sequencing. Total RNA was isolated using the Total Purification kit (Norgen Biotek, Thorold, ON, Canada). High Capacity RNA to cDNA kit (Life Technologies, Grand Island, NY) was used for reverse transcription of RNA. Libraries were generated using the NEBNext Ultra II Directional Library Prep Kit (New England Biolabs, Ipswich, MA) and subjected to sequencing (single-end 92x) on the NextSeq500 platform (Illumina) at the Cornell University Biotechnology Resource Center. At least 80M reads per sample were acquired.

Mapping sequencing reads. In the ChRO-seq studies, the publicly available pipeline[25] was used to align ChRO-seq reads. Since the libraries were prepared using adapters that contained a molecule-specific unique identifier (first 6 bp sequenced), the PCR duplicates were first removed using PRINSEQ lite. Adapters were trimmed from the 3′ end of remaining reads using cutadapt with a 10% error rate. Reads were mapped with the Burrows-Wheeler Aligner (BWA) to the human reference genome hg38 plus a single copy of the Pol I ribosomal RNA transcription unit (GenBank U13369.1). The location of active RNA polymerase was represented by a single base that denotes the 3′ end of the nascent RNA, which corresponds to the position on the 5′ end of each sequenced read. Mapped reads were converted to bigwig format using BedTools and the bedGraphToBigWig program in the Kent Source software package. ChRO-seq signal for visualization purpose was normalized by total bigwig signal (wigsum) of 10,000,000. The transcription (ChRO-seq) levels of genes were normalized by the length of gene bodies to transcripts per million (TPM).

In the RNA-seq studies, reads were mapped to human genome hg38 using STAR (v2.5.3a)[59] and transcript quantification was performed using Salmon (v0.6.0)[60] with GENCODE release 25 transcript annotations. The expression (RNA-seq) levels of genes were normalized using DESeq2[61]. All the samples except an Ile HIO sample had > 80% mapping rates. Although the Ile HIO sample had an unfavorable mapping rate, it was able to show elevated levels of Ile-associated regional markers compared to the Duo HIO sample **(Supplementary Figure 2)**.

Differential expression and pathway analyses. The differential analysis of gene bodies and TRE regions in ChRO-seq data was performed using DESeq2 package[61]. The differential analysis of RNA-seq was performed using DESeq2 package[61]. For all the analyses except stage-specific marker analysis (see below), the normalized levels of transcription or expression, foldchange and the statistic filtering were based on the

DESeq2 analysis including only the two stages in a comparison. The pathway enrichment analyses with subsets of genes were assessed using Enrichr[62].

ChRO-seq quantification of gene transcription activity. The quantification of transcription activity of genes was determined with exclusion of reads within 500 base downstream of transcription start site (TSS) to avoid bias generated by the RNA polymerase pausing at the promoters. Genes with gen body < 1000 base were excluded from all the gene body related analysis, given that genes with short gene bodies are likely biased TPM quantification when excluding the pause peak. ChRO-seq generates reads from both strands of chromatin. Only the reads from the corresponding strand of annotated genes were counted for quantification of transcription level.

Identification of faithful stage-specific markers. To identify genes that label stage of hESC, DE, Duo spheroids or Ile spheroids in ChRO-seq study, genes which are actively transcribed in the stage of interest (TPM > 50) are filtered with padj < 0.2, p < 0.05, fold change > 1.5 compared to all the other stages (DESeq2). To identify genes that label hESC, DE, Duo spheroids or Ile spheroids in RNA-seq study, genes which are expressed in the stage of interest (base mean > 100) are filtered with padj < 0.2, p < 0.05, fold change > 1.5 compared to all the other stages (DESeq2). To identify genes that label stage of SI spheroids irrespective of regional identities, genes which are actively transcribed and expressed in both Duo and Ile spheroids (TPM > 50 and base mean > 100) are filtered with padj < 0.2, p < 0.05, fold change > 1.5 compared to all the other stages and fold change < 1.5 between Duo and Ile spheroids (DESeq2). For each of the stage, the fateful markers at both transcription and expression level were identified by intersecting ChRO-seq and RNA-seq gene lists generated by the criteria mentioned above. The genes annotated under pseudogene category were excluded from this analysis.

TRE identification, annotation and categorization. The active transcriptional regulatory elements (TREs) were identified by dREG tool[25]. The annotation of the identified TREs was defined using *annotatePeaks.pl* function (genome = hg38) of HOMER package[54]. To assign active TREs as promoters, the TREs with at least 1 base overlapping with the window of -1000 base and +200 base of annotated TSSs were defined as proximal TREs, or promoters. The rest of the active TREs were defined as distal TREs, or enhancers.

Defining stage-specific TREs. The stage-specific TREs were defined as TREs of which the ChRO-seq intensity is significantly higher in a stage of interest relative to a comparative stage (padj < 0.05 and $\log_2$fold change > 2.5 by DESeq2)[61]. The ChRO-seq intensity was determined by the sum of the un-normalized ChRO-seq reads from both strands within a TRE region.

Stage-specific TRE density analysis. The density of stage-specific TREs was determined by the number of TREs present within the window of +100 kb and -100 kb around the TSS for all the genes which are actively transcribed in the stage of interest (TPM > 50 in the ChRO-seq study). The genes which are associated with stage-specific TREs are defined using the following criteria: (1) genes have density of stage-specific TRE > 0, (2) genes are actively transcribed (TPM > 50 in the ChRO-seq) and expressed (base mean > 100 in the RNA-seq) in the stage of interest, and (3) the genes are significantly uptranscribed and upregulated (padj < 0.2, p < 0.05, fold change > 1.5 in both ChRO-seq and RNA-seq by DESeq2) in the stage of interest relative to a comparison stage.

***Identification of de novo stage-specific enhancer hotspots* and the associated genes.** The stage-specific active distal TREs (enhancers) were used in the enhancer hotspot analysis. The enhancer hotspots in this study were identified by the criteria similar (with slight modifications) to the studies describing 'super-enhancers'[27,28]. Briefly, the stage-specific enhancers in proximity of distance (< 12.5kb) were stitched. For each of the stage-specific stitched enhancers, the transcription activity was determined by the sum of un-normalized ChRO-seq signals from both strands of each individual enhancer. To further identify stage-specific enhancer hotspots, a tangent line was applied the stage-specific stitched enhancers and they were

ranked based on their transcriptional activities in a plot. The ones above the tangent line in the analysis were defined as stage-specific enhancer hotpots. To identify the genes which are associated with stage-specific stitched enhancers, a given stage-specific stitched enhancer is assigned to the gene of which the transcription in the gene body is active (TPM > 50 in the matching stage) and the TSS is closest to the border of the enhancer region.

Transcription factor binding motif enrichment analysis. HOMER tool[54] was used to determine enrichment of sites corresponding to known motifs with stage-specific TREs (relative to a comparative stage). More specifically, we used function *findMotifsGenome.pl* (genome = hg38 and size = given) and the TREs which are shared or unique to the comparative stage were used as background.

Transcription factor cistrome analysis. For a TF of interest, the binding motifs present in the stage-specific TREs were identified and the density of the motif was determined by the number of motifs within the window of +100 kb and -100 kb around the TSS for all the genes which are actively transcribed in the stage of interest (TPM > 50). The cistrome of a TF is defined using the following criteria: (1) genes have motif density > 0, (2) genes are actively transcribed (TPM > 50 in the ChRO-seq) and expressed (base mean > 100 in the RNA-seq) in the stage of interest, and (3) the genes are significantly uptranscribed and upregulated (padj < 0.2, p < 0.05, fold change > 1.5 in both ChRO-seq and RNA-seq by DESeq2) in the stage of interest relative to a comparison stage.

Statistics. All the padj and p values presented in this study were determined by Wald test (DESeq2), unless otherwise specifically noted.

Data availability. Raw and processed data generated in the sequencing studies will be made available upon publication.

REFERENCES

1 Sheaffer, K. L. & Kaestner, K. H. Transcriptional networks in liver and intestinal development. *Cold Spring Harb Perspect Biol* **4**, a008284, doi:10.1101/cshperspect.a008284 (2012).

2 Sherwood, R. I., Chen, T. Y. & Melton, D. A. Transcriptional dynamics of endodermal organ formation. *Dev Dyn* **238**, 29-42, doi:10.1002/dvdy.21810 (2009).

3 Nowotschin, S. *et al.* The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* **569**, 361-367, doi:10.1038/s41586-019-1127-1 (2019).

4 Gao, S. *et al.* Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing. *Nat Cell Biol* **20**, 721-734, doi:10.1038/s41556-018-0105-4 (2018).

5 Kumar, N. *et al.* The lineage-specific transcription factor CDX2 navigates dynamic chromatin to control distinct stages of intestine development. *Development* **146**, doi:10.1242/dev.172189 (2019).

6 Dong, J. *et al.* Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. *Genome Biol* **19**, 31, doi:10.1186/s13059-018-1416-2 (2018).

7 Banerjee, K. K. *et al.* Enhancer, transcriptional, and cell fate plasticity precedes intestinal determination during endoderm development. *Genes Dev* **32**, 1430-1442, doi:10.1101/gad.318832.118 (2018).

8 Nord, A. S. *et al.* Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell* **155**, 1521-1531, doi:10.1016/j.cell.2013.11.033 (2013).

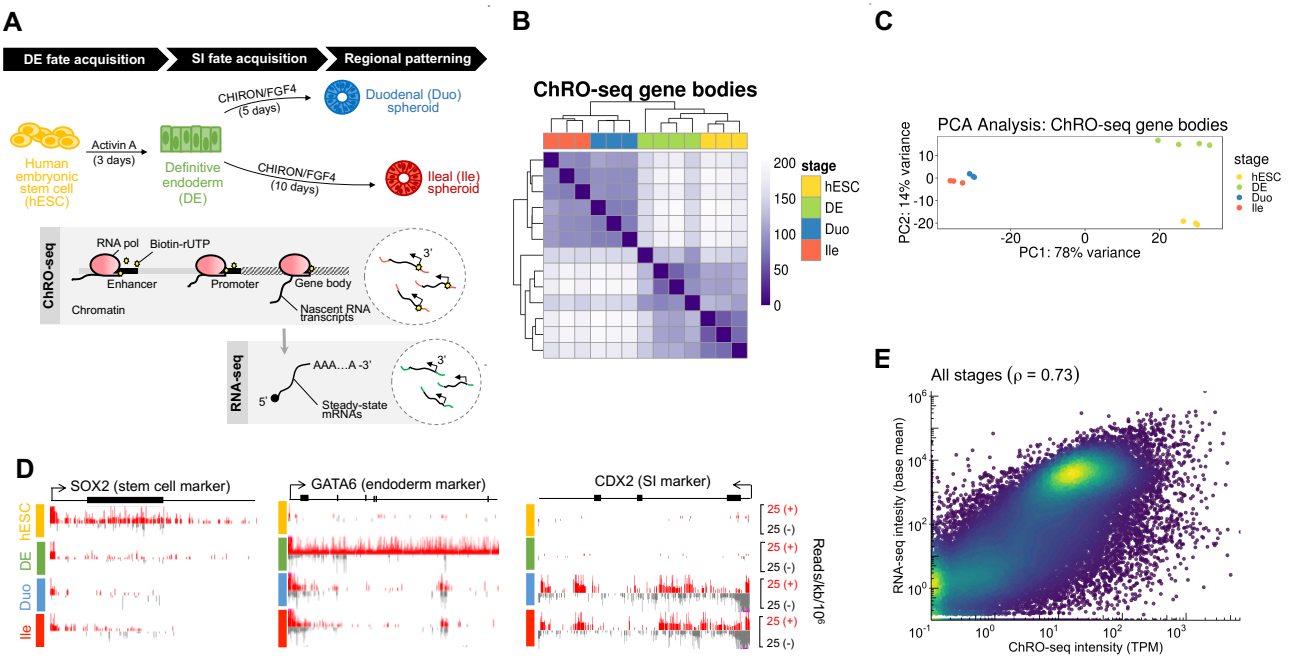
9 Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134-1148, doi:10.1016/j.cell.2013.04.022 (2013).

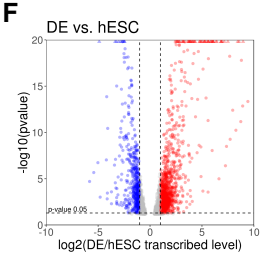
10 Heintzman, N. D. *et al.* Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108-112, doi:10.1038/nature07829 (2009).
11 Long, H. K., Prescott, S. L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. *Cell* **167**, 1170-1187, doi:10.1016/j.cell.2016.09.018 (2016).
12 Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-187, doi:10.1038/nature09033 (2010).
13 Koch, F. *et al.* Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol* **18**, 956-963, doi:10.1038/nsmb.2085 (2011).
14 Chu, T. *et al.* Chromatin run-on and sequencing maps the transcriptional regulatory landscape of glioblastoma multiforme. *Nat Genet* **50**, 1553-1564, doi:10.1038/s41588-018-0244-3 (2018).
15 Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845-1848, doi:10.1126/science.1162228 (2008).
16 Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950-953, doi:10.1126/science.1229386 (2013).
17 Wells, J. M. & Spence, J. R. How to make an intestine. *Development* **141**, 752-760, doi:10.1242/dev.097386 (2014).
18 Sinagoga, K. L. & Wells, J. M. Generating human intestinal tissues from pluripotent stem cells to study development and disease. *EMBO J* **34**, 1149-1163, doi:10.15252/embj.201490686 (2015).
19 Spence, J. R. *et al.* Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. *Nature* **470**, 105-109, doi:10.1038/nature09691 (2011).
20 McCracken, K. W., Howell, J. C., Wells, J. M. & Spence, J. R. Generating human intestinal tissue from pluripotent stem cells in vitro. *Nat Protoc* **6**, 1920-1928, doi:10.1038/nprot.2011.410 (2011).
21 Tsai, Y. H. *et al.* In vitro patterning of pluripotent stem cell-derived intestine recapitulates in vivo human development. *Development* **144**, 1045-1055, doi:10.1242/dev.138453 (2017).
22 Watson, C. L. *et al.* An in vivo model of human small intestine using pluripotent stem cells. *Nat Med* **20**, 1310-1314, doi:10.1038/nm.3737 (2014).
23 Hill, D. R. *et al.* Bacterial colonization stimulates a complex physiological response in the immature human intestinal epithelium. *Elife* **6**, doi:10.7554/eLife.29132 (2017).
24 Finkbeiner, S. R. *et al.* Transcriptome-wide Analysis Reveals Hallmarks of Human Intestine Development and Maturation In Vitro and In Vivo. *Stem Cell Reports*, doi:10.1016/j.stemcr.2015.04.010 (2015).
25 Chu, T., Wang, Z., Chou, S. P. & Danko, C. G. Discovering Transcriptional Regulatory Elements From Run-On and Sequencing Data Using the Web-Based dREG Gateway. *Curr Protoc Bioinformatics* **66**, e70, doi:10.1002/cpbi.70 (2019).
26 Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455-461, doi:10.1038/nature12787 (2014).
27 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).
28 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).
29 Boyd, M. *et al.* Characterization of the enhancer and promoter landscape of inflammatory bowel disease from human colon biopsies. *Nat Commun* **9**, 1661, doi:10.1038/s41467-018-03766-z (2018).
30 Kageyama, R., Ohtsuka, T. & Kobayashi, T. The Hes gene family: repressors and oscillators that orchestrate embryogenesis. *Development* **134**, 1243-1251, doi:10.1242/dev.000786 (2007).
31 Wu, Y. Z. *et al.* Dysregulation of miR-431 and target gene FOXA1 in intestinal tissues of infants with necrotizing enterocolitis. *FASEB J* **33**, 5143-5152, doi:10.1096/fj.201801470R (2019).

32 Mustata, R. C. *et al.* Identification of Lgr5-independent spheroid-generating progenitors of the mouse fetal intestinal epithelium. *Cell Rep* **5**, 421-432, doi:10.1016/j.celrep.2013.09.005 (2013).
33 Mallo, M., Wellik, D. M. & Deschamps, J. Hox genes and regional patterning of the vertebrate body plan. *Dev Biol* **344**, 7-15, doi:10.1016/j.ydbio.2010.04.024 (2010).
34 Deschamps, J. & van Nes, J. Developmental regulation of the Hox genes during axial morphogenesis in the mouse. *Development* **132**, 2931-2942, doi:10.1242/dev.01897 (2005).
35 Stairs, D. B., Kong, J. & Lynch, J. P. Cdx genes, inflammation, and the pathogenesis of intestinal metaplasia. *Prog Mol Biol Transl Sci* **96**, 231-270, doi:10.1016/B978-0-12-381280-3.00010-5 (2010).
36 Tu, W. *et al.* Overexpression of HOXB7 is associated with a poor prognosis in patients with gastric cancer. *Oncol Lett* **10**, 2967-2973, doi:10.3892/ol.2015.3630 (2015).
37 di Pietro, M. *et al.* Evidence for a functional role of epigenetically regulated midcluster HOXB genes in the development of Barrett esophagus. *Proc Natl Acad Sci U S A* **109**, 9077-9082, doi:10.1073/pnas.1116933109 (2012).
38 Cahill, C. M. *et al.* Differential Expression of the Activator Protein 1 Transcription Factor Regulates Interleukin-1ss Induction of Interleukin 6 in the Developing Enterocyte. *PLoS One* **11**, e0145184, doi:10.1371/journal.pone.0145184 (2016).
39 Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551**, 333-339, doi:10.1038/nature24489 (2017).
40 Villano, J. L. & Katz, F. N. four-jointed is required for intermediate growth in the proximal-distal axis in Drosophila. *Development* **121**, 2767-2777 (1995).
41 Zeidler, M. P., Perrimon, N. & Strutt, D. I. The four-jointed gene is required in the Drosophila eye for ommatidial polarity specification. *Curr Biol* **9**, 1363-1372, doi:10.1016/s0960-9822(00)80081-0 (1999).
42 Rock, R., Heinrich, A. C., Schumacher, N. & Gessler, M. Fjx1: a notch-inducible secreted ligand with specific binding sites in developing mouse embryos and adult brain. *Dev Dyn* **234**, 602-612, doi:10.1002/dvdy.20553 (2005).
43 Allison, T. F. *et al.* Identification and Single-Cell Functional Characterization of an Endodermally Biased Pluripotent Substate in Human Embryonic Stem Cells. *Stem Cell Reports* **10**, 1895-1907, doi:10.1016/j.stemcr.2018.04.015 (2018).
44 Chu, L. F. *et al.* Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol* **17**, 173, doi:10.1186/s13059-016-1033-x (2016).
45 Tiyaboonchai, A. *et al.* GATA6 Plays an Important Role in the Induction of Human Definitive Endoderm, Development of the Pancreas, and Functionality of Pancreatic beta Cells. *Stem Cell Reports* **8**, 589-604, doi:10.1016/j.stemcr.2016.12.026 (2017).
46 Cardoso-Moreira, M. *et al.* Gene expression across mammalian organ development. *Nature* **571**, 505-509, doi:10.1038/s41586-019-1338-5 (2019).
47 Munera, J. O. *et al.* Differentiation of Human Pluripotent Stem Cells into Colonic Organoids via Transient Activation of BMP Signaling. *Cell Stem Cell* **21**, 51-64 e56, doi:10.1016/j.stem.2017.05.020 (2017).
48 Trisno, S. L. *et al.* Esophageal Organoids from Human Pluripotent Stem Cells Delineate Sox2 Functions during Esophageal Specification. *Cell Stem Cell* **23**, 501-515 e507, doi:10.1016/j.stem.2018.08.008 (2018).
49 Dye, B. R. *et al.* In vitro generation of human pluripotent stem cell derived lung organoids. *Elife* **4**, doi:10.7554/eLife.05098 (2015).
50 Mun, S. J. *et al.* Generation of expandable human pluripotent stem cell-derived hepatocyte-like liver organoids. *J Hepatol* **71**, 970-985, doi:10.1016/j.jhep.2019.06.030 (2019).
51 D'Amour, K. A. *et al.* Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nat Biotechnol* **24**, 1392-1401, doi:10.1038/nbt1259 (2006).

52 McCracken, K. W. *et al.* Modelling human development and disease in pluripotent stem-cell-derived gastric organoids. *Nature* **516**, 400-404, doi:10.1038/nature13863 (2014).
53 Xie, R. *et al.* Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. *Cell Stem Cell* **12**, 224-237, doi:10.1016/j.stem.2012.11.023 (2013).
54 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
55 Finkbeiner, S. R. *et al.* Generation of tissue-engineered small intestine using embryonic stem cell-derived human intestinal organoids. *Biol Open* **4**, 1462-1472, doi:10.1242/bio.013235 (2015).
56 Dame, M. K. *et al.* Identification, isolation and characterization of human LGR5-positive colon adenoma cells. *Development* **145**, doi:10.1242/dev.153049 (2018).
57 Wuarin, J. & Schibler, U. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol Cell Biol* **14**, 7219-7225, doi:10.1128/mcb.14.11.7219 (1994).
58 Mahat, D. B. *et al.* Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* **11**, 1455-1476, doi:10.1038/nprot.2016.086 (2016).
59 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
60 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417-419, doi:10.1038/nmeth.4197 (2017).
61 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
62 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).

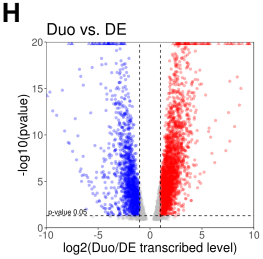
Figure 1.





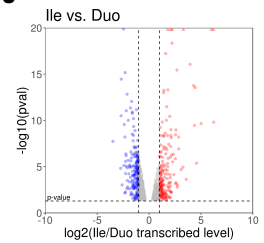
G

GO Term of up-transcribed genes in DE vs. hESC	*P* value
Endoderm development	1.572e-6
Endoderm formation	1.83e-5
Regulation of SMAD protein phosphorylation	2.94e-5
BMP signaling pathway	8.463e-5
Cellular response to BMP stimulus	2.015e-4



I

GO Term of up-transcribed genes in Duo vs. DE	*P* value
Canonical glycolysis	5.425e-38
Glycolysis through glucose-6-phosphate	7.596e-38
Glucose catabolism to pyruvate	1.117e-36
Protein import into peroxisome matrix	2.684e-35
Intracellular estrogen receptor signaling	8.402e-34



K

GO Term of up-transcribed genes in Ile vs. Duo	*P* value
Regulation of transcription from RNA pol II promoter	2.822e-11
Negative regulation of transcription from RNA pol II promoter	1.183e-8
Positive regulation of DNA-templated transcription	1.124e-7
Positive regulation of transcription From RNA pol II promoter	2.519e-7
Negative regulation of erythrocyte differentiation	1.477e-5

L

BioPlanet 2019	*P* value
* AP-1 TF network	7.928e-10
BDNF signaling pathway	8.842e-7
Telomerase regulation	3.396e-5
Nrf2 Oxidative stress response	3.594e-5
FGF1	8.662e-5

NCI-Nature 2016	*P* value
* AP-1 TF network	6.850e-10
Telomerase regulation	3.396e-5
HIF-1a TF network	3.454e-4
FOXM1 TF network	4.792e-4
Delta-Np63 transcriptional targets	7.540e-4

Figure 2.

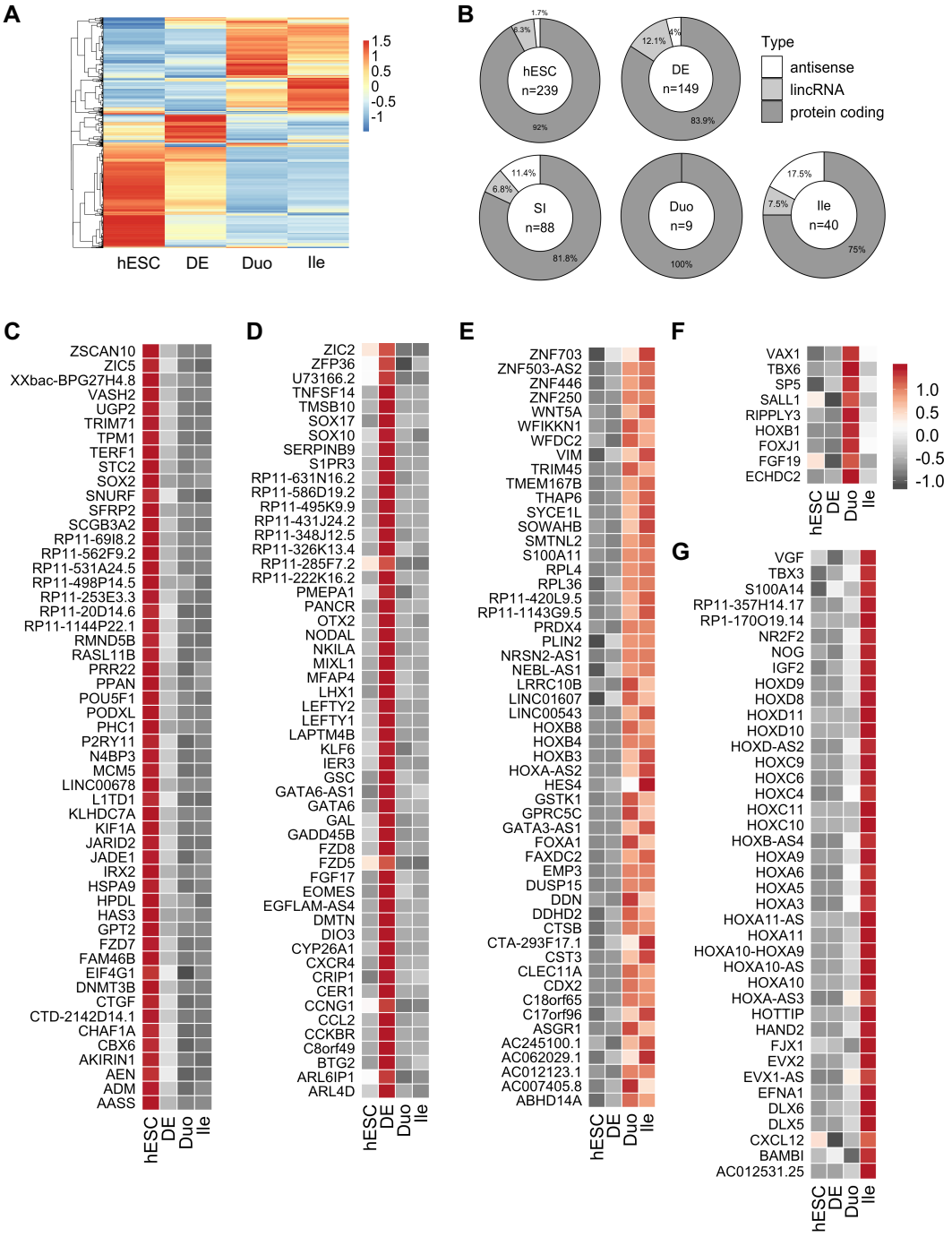

A
1.5
1
0.5
0
-0.5
-1
hESC
DE
Duo
Ile
B
hESC
n=239
1.7%
6.3%
92%
DE
n=149
12.1%
4%
83.9%
Type
antisense
lincRNA
protein coding
SI
n=88
11.4%
6.8%
81.8%
Duo
n=9
100%
Ile
n=40
17.5%
7.5%
75%
C
ZSCAN10
ZIC5
XXbac-BPG27H4.8
VASH2
UGP2
TRIM71
TPM1
TERF1
STC2
SOX2
SNURF
SFRP2
SCGB3A2
RP11-69I8.2
RP11-562F9.2
RP11-531A24.5
RP11-498P14.5
RP11-253E3.3
RP11-20D14.6
RP11-1144P22.1
RMND5B
RASL11B
PRR22
PPAN
POU5F1
PODXL
PHC1
P2RY11
N4BP3
MCM5
LINC00678
L1TD1
KLHDC7A
KIF1A
JARID2
JADE1
IRX2
HSPA9
HPDL
HAS3
GPT2
FZD7
FAM46B
EIF4G1
DNMT3B
CTGF
CTD-2142D14.1
CHAF1A
CBX6
AKIRIN1
AEN
ADM
AASS
hESC
DE
Duo
Ile
D
ZIC2
ZFP36
U73166.2
TNFSF14
TMSB10
SOX17
SOX10
SERPINB9
S1PR3
RP11-631N16.2
RP11-586D19.2
RP11-495K9.9
RP11-431J24.2
RP11-348J12.5
RP11-326K13.4
RP11-285F7.2
RP11-222K16.2
PMEPA1
PANCR
OTX2
NODAL
NKILA
MIXL1
MFAP4
LHX1
LEFTY2
LEFTY1
LAPTM4B
KLF6
IER3
GSC
GATA6-AS1
GATA6
GAL
GADD45B
FZD8
FZD5
FGF17
EOMES
EGFLAM-AS4
DMTN
DIO3
CYP26A1
CXCR4
CRIP1
CER1
CCNG1
CCL2
CCKBR
C8orf49
BTG2
ARL6IP1
ARL4D
hESC
DE
Duo
Ile
E
ZNF703
ZNF503-AS2
ZNF446
ZNF250
WNT5A
WFIKKN1
WFDC2
VIM
TRIM45
TMEM167B
THAP6
SYCE1L
SOWAHB
SMTNL2
S100A11
RPL4
RPL36
RP11-420L9.5
RP11-1143G9.5
PRDX4
PLIN2
NRSN2-AS1
NEBL-AS1
LRRC10B
LINC01607
LINC00543
HOXB8
HOXB4
HOXB3
HOXA-AS2
HES4
GSTK1
GPRC5C
GATA3-AS1
FOXA1
FAXDC2
EMP3
DUSP15
DDN
DDHD2
CTSB
CTA-293F17.1
CST3
CLEC11A
CDX2
C18orf65
C17orf96
ASGR1
AC245100.1
AC062029.1
AC012123.1
AC007405.8
ABHD14A
hESC
DE
Duo
Ile
F
VAX1
TBX6
SP5
SALL1
RIPPLY3
HOXB1
FOXJ1
FGF19
ECHDC2
1.0
0.5
0.0
-0.5
-1.0
hESC
DE
Duo
Ile
G
VGF
TBX3
S100A14
RP11-357H14.17
RP1-170O19.14
NR2F2
NOG
IGF2
HOXD9
HOXD8
HOXD11
HOXD10
HOXD-AS2
HOXC9
HOXC6
HOXC4
HOXC11
HOXC10
HOXB-AS4
HOXA9
HOXA6
HOXA5
HOXA3
HOXA11-AS
HOXA11
HOXA10-HOXA9
HOXA10-AS
HOXA10
HOXA-AS3
HOTTIP
HAND2
FJX1
EVX2
EVX1-AS
EFNA1
DLX6
DLX5
CXCL12
BAMBI
AC012531.25
hESC
DE
Duo
Ile

Figure 3.

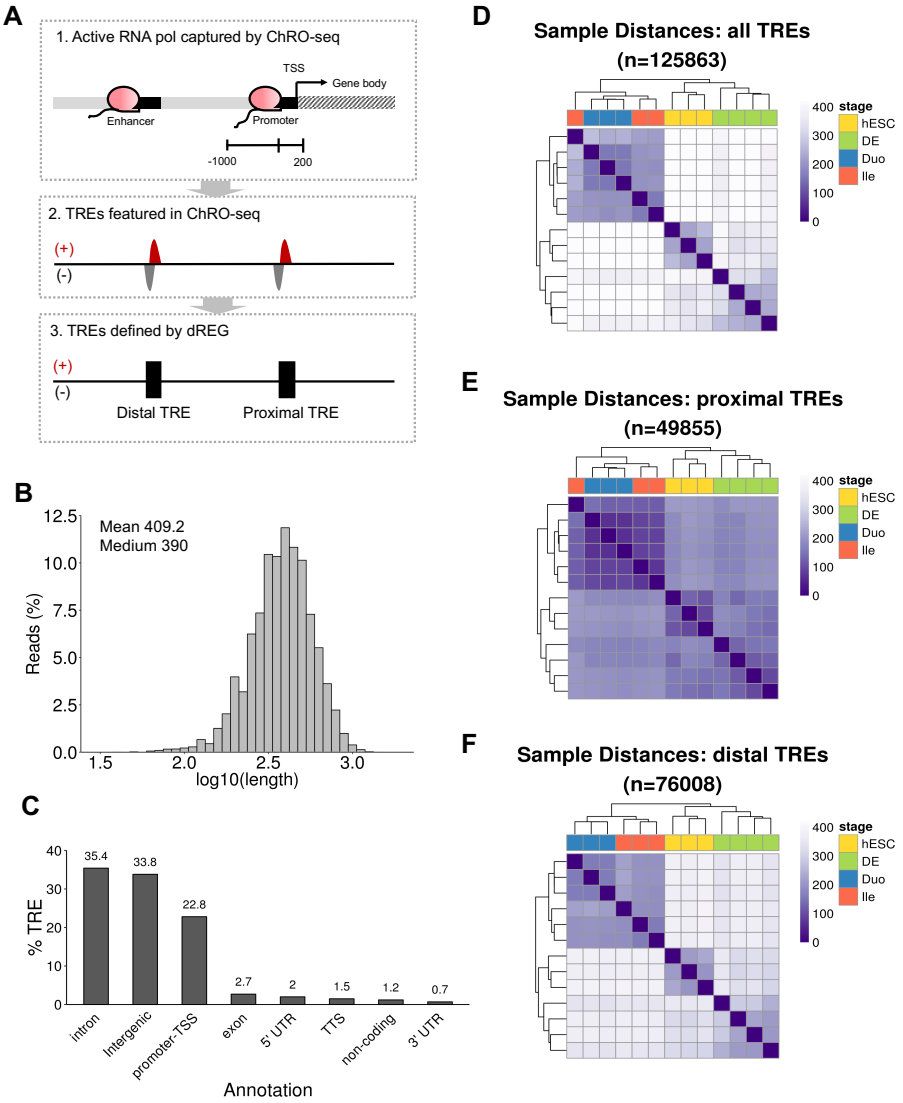

A
1. Active RNA pol captured by ChRO-seq
TSS
Gene body
Enhancer
Promoter
-1000
200
2. TREs featured in ChRO-seq
(+)
(-)
3. TREs defined by dREG
(+)
(-)
Distal TRE
Proximal TRE
B
Mean 409.2
Medium 390
Reads (%)
log10(length)
C
% TRE
35.4
33.8
22.8
2.7
2
1.5
1.2
0.7
intron
Intergenic
promoter-TSS
exon
5' UTR
TTS
non-coding
3' UTR
Annotation
D
Sample Distances: all TREs
(n=125863)
E
Sample Distances: proximal TREs
(n=49855)
F
Sample Distances: distal TREs
(n=76008)
stage
hESC
DE
Duo
Ile

Figure 4.

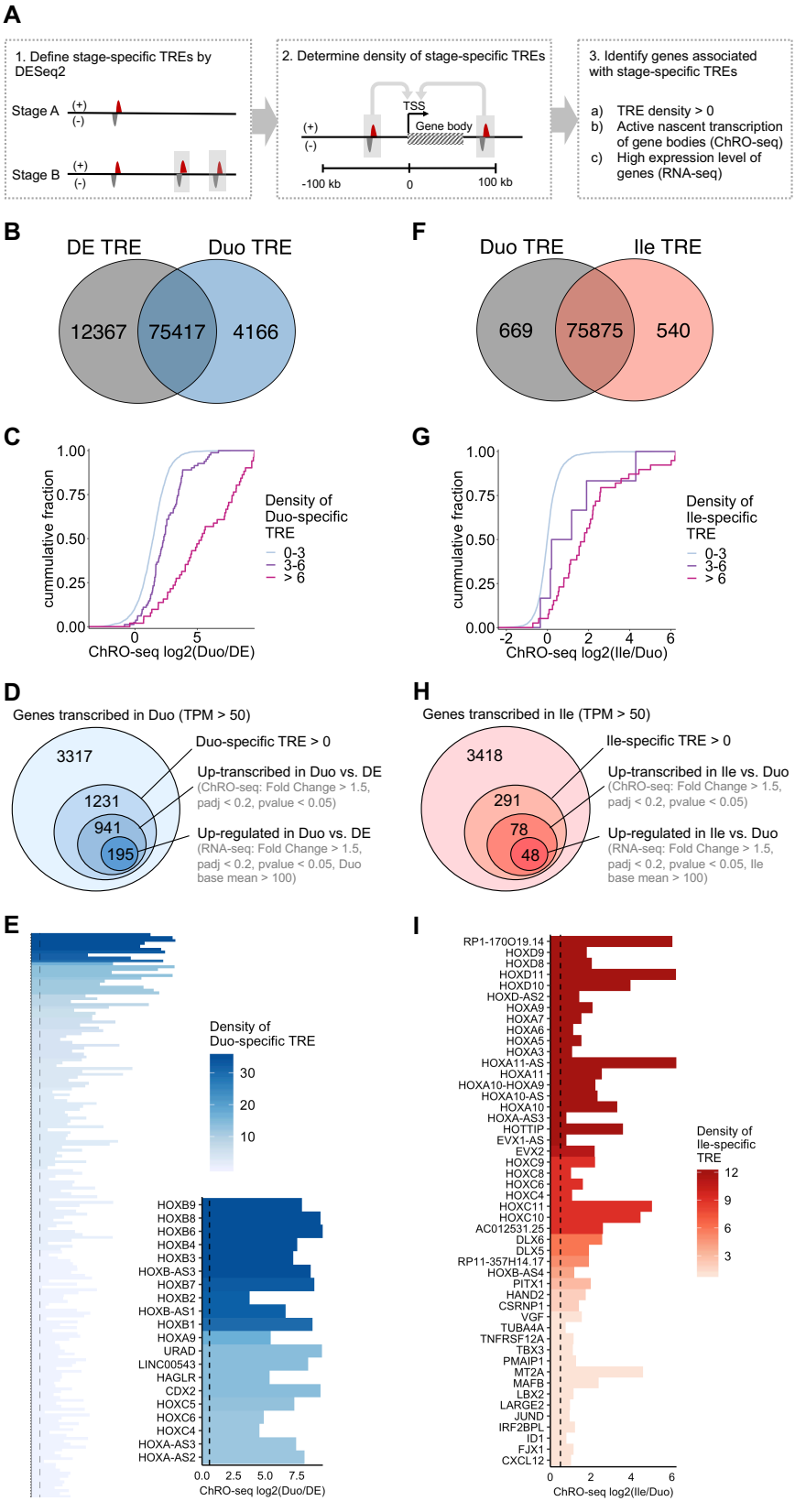

A
1. Define stage-specific TREs by DESeq2
Stage A
Stage B
(+)
(-)
2. Determine density of stage-specific TREs
TSS
Gene body
-100 kb
0
100 kb
3. Identify genes associated with stage-specific TREs
a) TRE density > 0
b) Active nascent transcription of gene bodies (ChRO-seq)
c) High expression level of genes (RNA-seq)
B
DE TRE
Duo TRE
12367
75417
4166
F
Duo TRE
Ile TRE
669
75875
540
C
cummulative fraction
ChRO-seq log2(Duo/DE)
Density of Duo-specific TRE
0-3
3-6
> 6
G
ChRO-seq log2(Ile/Duo)
Density of Ile-specific TRE
D
Genes transcribed in Duo (TPM > 50)
3317
1231
941
195
Duo-specific TRE > 0
Up-transcribed in Duo vs. DE
(ChRO-seq: Fold Change > 1.5, padj < 0.2, pvalue < 0.05)
Up-regulated in Duo vs. DE
(RNA-seq: Fold Change > 1.5, padj < 0.2, pvalue < 0.05, Duo base mean > 100)
H
Genes transcribed in Ile (TPM > 50)
3418
291
78
48
Ile-specific TRE > 0
Up-transcribed in Ile vs. Duo
(ChRO-seq: Fold Change > 1.5, padj < 0.2, pvalue < 0.05)
Up-regulated in Ile vs. Duo
(RNA-seq: Fold Change > 1.5, padj < 0.2, pvalue < 0.05, Ile base mean > 100)
E
Density of Duo-specific TRE
HOXB9
HOXB8
HOXB6
HOXB4
HOXB3
HOXB-AS3
HOXB7
HOXB2
HOXB-AS1
HOXB1
HOXA9
URAD
LINC00543
HAGLR
CDX2
HOXC5
HOXC6
HOXC4
HOXA-AS3
HOXA-AS2
ChRO-seq log2(Duo/DE)
I
RP1-170O19.14
HOXD9
HOXD8
HOXD11
HOXD10
HOXD-AS2
HOXA9
HOXA7
HOXA6
HOXA5
HOXA3
HOXA11-AS
HOXA11
HOXA10-HOXA9
HOXA10-AS
HOXA10
HOXA-AS3
HOTTIP
EVX1-AS
EVX2
HOXC9
HOXC8
HOXC6
HOXC4
HOXC11
HOXC10
AC012531.25
DLX6
DLX5
RP11-357H14.17
HOXB-AS4
PITX1
HAND2
CSRNP1
VGF
TUBA4A
TNFRSF12A
TBX3
PMAIP1
MT2A
MAFB
LBX2
LARGE2
JUND
IRF2BPL
ID1
FJX1
CXCL12
Density of Ile-specific TRE
ChRO-seq log2(Ile/Duo)

Figure 5.

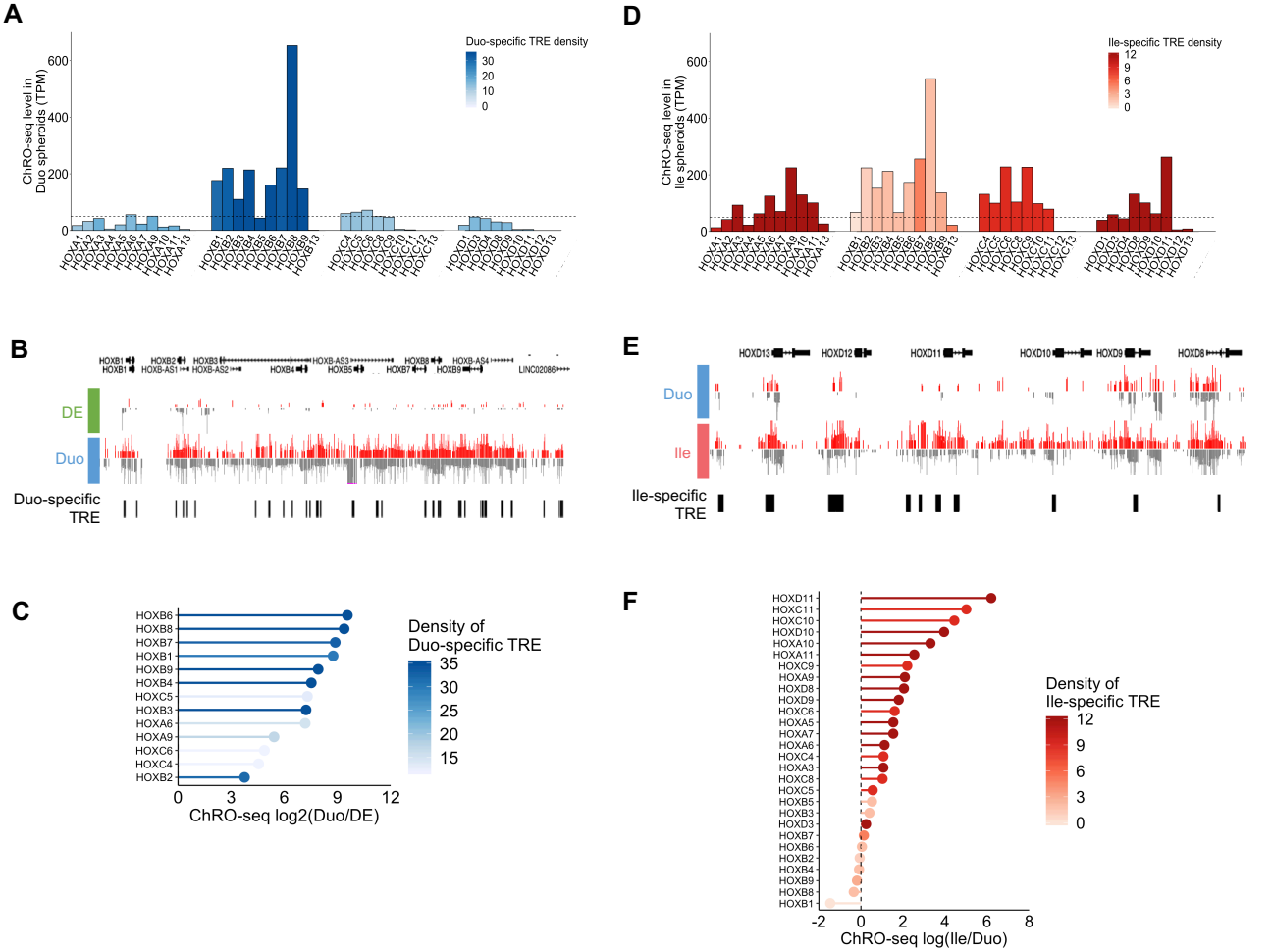

A
ChRO-seq level in Duo spheroids (TPM)
Duo-specific TRE density
30
20
10
0
D
ChRO-seq level in Ile spheroids (TPM)
Ile-specific TRE density
12
9
6
3
0
B
DE
Duo
Duo-specific TRE
E
Duo
Ile
Ile-specific TRE
C
Density of Duo-specific TRE
ChRO-seq log2(Duo/DE)
F
Density of Ile-specific TRE
ChRO-seq log(Ile/Duo)

Figure 6.

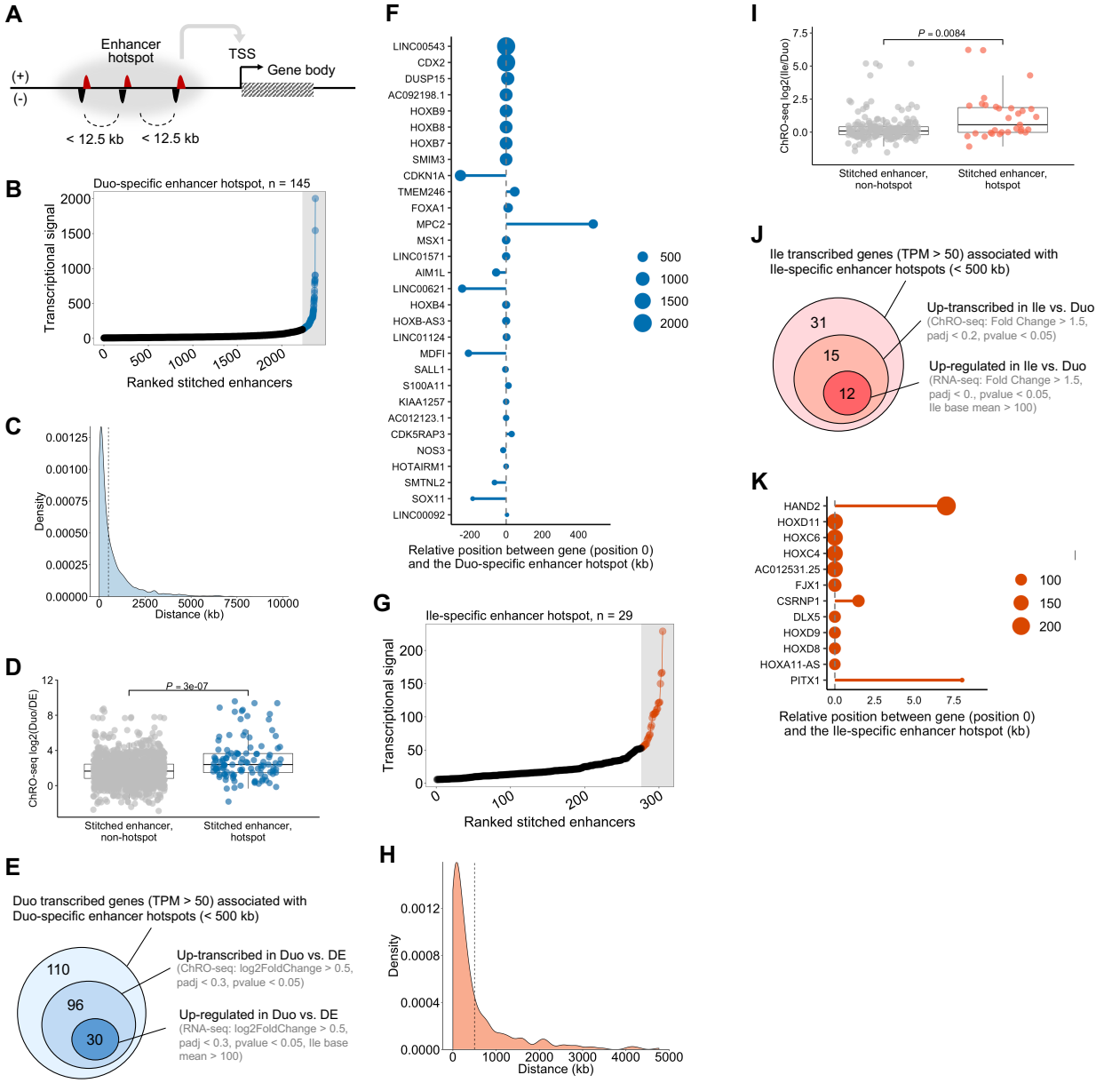

A
Enhancer hotspot
TSS
Gene body
(+)
(-)
< 12.5 kb
< 12.5 kb
B
Duo-specific enhancer hotspot, n = 145
Transcriptional signal
Ranked stitched enhancers
C
Density
Distance (kb)
D
P = 3e-07
ChRO-seq log2(Duo/DE)
Stitched enhancer, non-hotspot
Stitched enhancer, hotspot
E
Duo transcribed genes (TPM > 50) associated with Duo-specific enhancer hotspots (< 500 kb)
110
96
30
Up-transcribed in Duo vs. DE
(ChRO-seq: log2FoldChange > 0.5, padj < 0.3, pvalue < 0.05)
Up-regulated in Duo vs. DE
(RNA-seq: log2FoldChange > 0.5, padj < 0.3, pvalue < 0.05, Ile base mean > 100)
F
LINC00543
CDX2
DUSP15
AC092198.1
HOXB9
HOXB8
HOXB7
SMIM3
CDKN1A
TMEM246
FOXA1
MPC2
MSX1
LINC01571
AIM1L
LINC00621
HOXB4
HOXB-AS3
LINC01124
MDFI
SALL1
S100A11
KIAA1257
AC012123.1
CDK5RAP3
NOS3
HOTAIRM1
SMTNL2
SOX11
LINC00092
500
1000
1500
2000
Relative position between gene (position 0) and the Duo-specific enhancer hotspot (kb)
G
Ile-specific enhancer hotspot, n = 29
Transcriptional signal
Ranked stitched enhancers
H
Density
Distance (kb)
I
P = 0.0084
ChRO-seq log2(Ile/Duo)
Stitched enhancer, non-hotspot
Stitched enhancer, hotspot
J
Ile transcribed genes (TPM > 50) associated with Ile-specific enhancer hotspots (< 500 kb)
31
15
12
Up-transcribed in Ile vs. Duo
(ChRO-seq: Fold Change > 1.5, padj < 0.2, pvalue < 0.05)
Up-regulated in Ile vs. Duo
(RNA-seq: Fold Change > 1.5, padj < 0., pvalue < 0.05, Ile base mean > 100)
K
HAND2
HOXD11
HOXC6
HOXC4
AC012531.25
FJX1
CSRNP1
DLX5
HOXD9
HOXD8
HOXA11-AS
PITX1
100
150
200
Relative position between gene (position 0) and the Ile-specific enhancer hotspot (kb)

Figure 7.

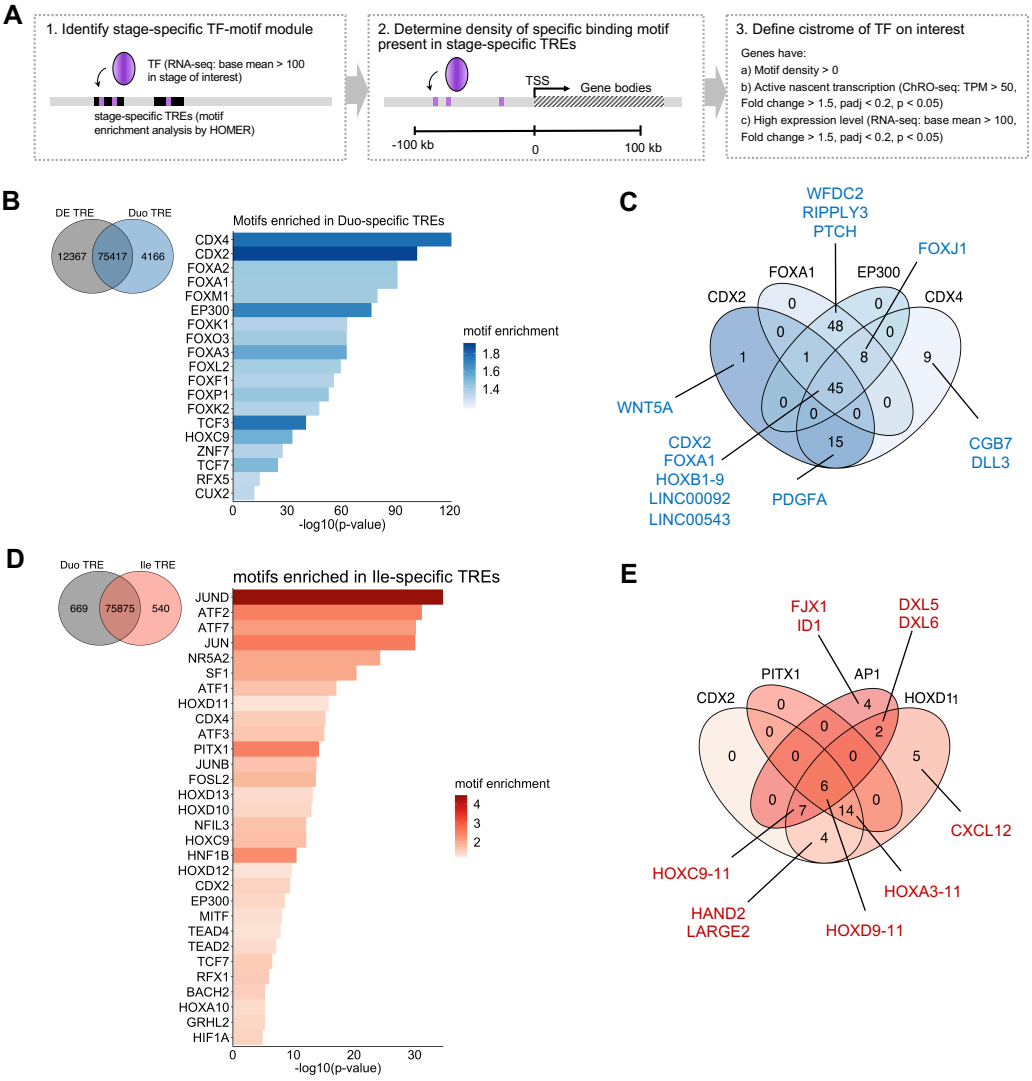
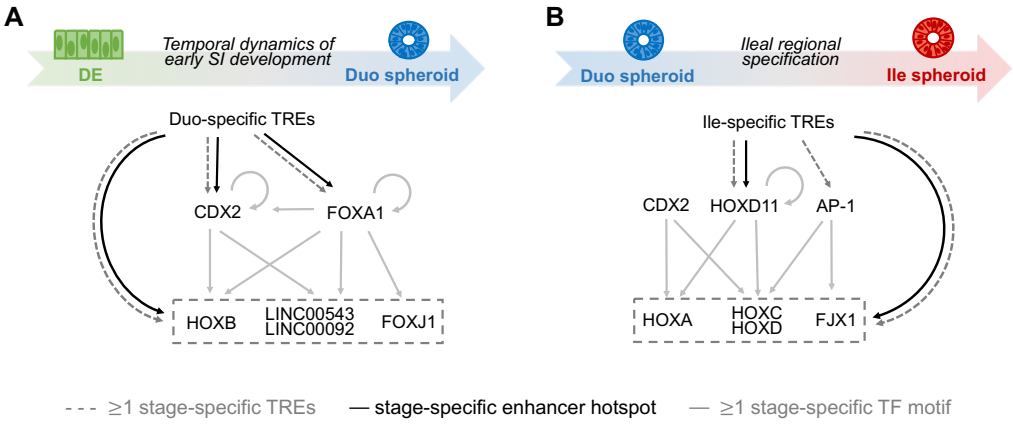

A
1. Identify stage-specific TF-motif module
TF (RNA-seq: base mean > 100 in stage of interest)
stage-specific TREs (motif enrichment analysis by HOMER)
2. Determine density of specific binding motif present in stage-specific TREs
TSS
Gene bodies
-100 kb
0
100 kb
3. Define cistrome of TF on interest
Genes have:
a) Motif density > 0
b) Active nascent transcription (ChRO-seq: TPM > 50, Fold change > 1.5, padj < 0.2, p < 0.05)
c) High expression level (RNA-seq: base mean > 100, Fold change > 1.5, padj < 0.2, p < 0.05)
B
DE TRE
Duo TRE
12367
75417
4166
Motifs enriched in Duo-specific TREs
CDX4
CDX2
FOXA2
FOXA1
FOXM1
EP300
FOXK1
FOXO3
FOXA3
FOXL2
FOXF1
FOXP1
FOXK2
TCF3
HOXC9
ZNF7
TCF7
RFX5
CUX2
motif enrichment
1.8
1.6
1.4
0
30
60
90
120
-log10(p-value)
C
WFDC2
RIPPLY3
PTCH
FOXJ1
FOXA1
EP300
CDX2
CDX4
WNT5A
CDX2
FOXA1
HOXB1-9
LINC00092
LINC00543
PDGFA
CGB7
DLL3
D
Duo TRE
Ile TRE
669
75875
540
motifs enriched in Ile-specific TREs
JUND
ATF2
ATF7
JUN
NR5A2
SF1
ATF1
HOXD11
CDX4
ATF3
PITX1
JUNB
FOSL2
HOXD13
HOXD10
NFIL3
HOXC9
HNF1B
HOXD12
CDX2
EP300
MITF
TEAD4
TEAD2
TCF7
RFX1
BACH2
HOXA10
GRHL2
HIF1A
motif enrichment
4
3
2
0
10
20
30
-log10(p-value)
E
FJX1
ID1
DXL5
DXL6
PITX1
AP1
CDX2
HOXD11
CXCL12
HOXC9-11
HOXA3-11
HAND2
LARGE2
HOXD9-11

Figure 8.


A
DE
Temporal dynamics of early SI development
Duo spheroid
Duo-specific TREs
CDX2
FOXA1
HOXB
LINC00543
LINC00092
FOXJ1
B
Duo spheroid
Ileal regional specification
Ile spheroid
Ile-specific TREs
CDX2
HOXD11
AP-1
HOXA
HOXC
HOXD
FJX1
- - - ≥1 stage-specific TREs
— stage-specific enhancer hotspot
— ≥1 stage-specific TF motif