

Chromatin regulatory dynamics of early human small intestinal development using a directed differentiation model

Yu-Han Hung¹, Sha Huang^{2,3}, Michael K. Dame³, Qianhui Yu⁴, Qing C. Yu⁵, Yi A. Zeng⁵, J. Gray Camp^{4,6}, Jason R. Spence^{2,3} and Praveen Sethupathy^{1,*}

¹Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA, ²Department of Cell and Developmental Biology, University of Michigan, Ann Arbor, MI 48109, USA, ³Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, MI 48109, USA, ⁴Institute of Molecular and Clinical Ophthalmology Basel, Basel 4056, Switzerland, ⁵State Key Laboratory of Cell Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, 200031, China and ⁶Department of Ophthalmology, University of Basel, Basel 4001, Switzerland

Received October 12, 2020; Revised November 20, 2020; Editorial Decision November 24, 2020; Accepted December 01, 2020

ABSTRACT

The establishment of the small intestinal (SI) lineage during human embryogenesis ensures functional integrity of the intestine after birth. The chromatin dynamics that drive SI lineage formation and regional patterning in humans are essentially unknown. To fill this knowledge void, we apply a cutting-edge genomic technology to a state-of-the-art human model of early SI development. Specifically, we leverage chromatin run-on sequencing (ChRO-seq) to define the landscape of active promoters, enhancers and gene bodies across distinct stages of directed differentiation of human pluripotent stem cells into SI spheroids with regional specification. Through comprehensive ChRO-seq analysis we identify candidate stage-specific chromatin activity states, novel markers and enhancer hotspots during the directed differentiation. Moreover, we propose a detailed transcriptional network associated with SI lineage formation or regional patterning. Our ChRO-seq analyses uncover a previously undescribed pattern of enhancer activity and transcription at HOX gene loci underlying SI regional patterning. We also validated this unique HOX dynamics by the analysis of single cell RNA-seq data from human fetal SI. Overall, the results lead to a new proposed working model for the regulatory underpinnings of human SI development, thereby adding a novel dimension to the literature that has relied almost exclusively on non-human models.

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 (1–5), and more recent studies have leveraged advanced genomic technologies (e.g. single cell RNA-seq and bulk ATAC-seq) to provide insights at a more granular level into gut development (6–11); however, this work relies almost exclusively on animal models. Studies of human SI development have been few (7,12,13), due in large part to limited access to primary human fetal tissues. Initial human SI lineage formation from the endodermal germ layer, which occurs much prior to the developmental time points at which human fetal tissues are generally available, is essentially uncharacterized. 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 human SI lineage formation and initial regional patterning.

Robust temporal and spatial regulation of gene expression is fundamental to all biological processes including development (14,15). 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 (16,17). RNA polymerases are known to be recruited to active enhancers,

*To whom correspondence should be addressed. Tel: +1 607 253 4375; Email: pr46@cornell.edu

generating divergent short transcripts (also known as enhancer RNA or eRNAs) (18,19). Recently, an approach called chromatin run-on sequencing (ChRO-seq) (20) 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) (21) and precision run-on sequencing (PRO-seq) (22). ChRO-seq was very recently successfully applied to archived solid tumor tissues (20,23).

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 developmental events of human SI that would be nearly impossible otherwise (24,25). The multi-step process of generating human ‘gut-in-a-dish’ from hPSCs recapitulates many aspects of *in vivo* SI development (26–28), including induction of the endodermal fate, formation of SI lineage (SI spheroids) with regional identities, and maturation into three-dimensional human intestinal organoids (HIOs) that exhibit molecular, structural and functional features similar to those of the human fetal SI (13,29). While the protocol for directed differentiation of hPSCs into SI spheroids has been established, the molecular mechanisms governing this process remains minimally characterized. To reveal temporal dynamics of the chromatin state during early human SI development and regional patterning, we performed ChRO-seq across the following stages: hPSC, definitive endoderm (DE), duodenal (proximal SI) spheroid and ileal (distal SI) spheroid. This study provides the first-ever view of the changing chromatin regulatory landscape, defines stage-specific enhancer hotspots and identifies key TF networks that underlie the acquisition of SI identity and the initiation of ileal regional patterning. Moreover, we uncover previously undescribed dynamics at *HOX* gene loci associated with these events. Additional analysis of single cell RNA-seq (scRNA-seq) of human fetal SI tissue underscores the unique *HOX* gene patterns, although these results are subject to the caveat that these tissues reflect developmental time points much later than what we have the opportunity to analyze in the directed differentiation model. Overall, this study offers an unprecedented resource that serves as a springboard from which the research community can develop and test targeted hypotheses about key regulatory hotspots and molecular drivers of early events of human SI development.

MATERIALS AND METHODS

Directed differentiation of hESCs

Differentiation of H9 hESCs and organoids was performed as previously published, with minor modifications (27,30). 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 μ M) for next 10 days. The intestinal

spheroids representing fetal duodenum and ileum were collected on day 5 and day 10, respectively (28). Subsets of spheroids collected at day 5 and day 10 were then cultured in Matrigel with the previously defined intestine growth media (28) 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 fluorescence activated cell sorting, were based on previously described procedures (31). All solutions, including overnight pretreatment of the organoid cultures, contained 10 μ M Y27632 (Tocris). Matrigel was digested for 30 min with cold 4 mM ethylenediaminetetraacetic acid (EDTA)-DPBS and organoids were washed 4 \times 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 min at 37°C. The cell suspension was then washed with 0.5% bovine serum albumin (BSA)-2 mM EDTA-DPBS over a succession of cell strainers, 100 μ m, 40 μ m (Corning) and 20 μ m (CellTrics) and centrifuged for 5 min at 500 \times g. Cells were labeled with EpCAM phycoerythrin (PE)-conjugated antibody (BioLegend) and an EpCAM isotype-PE control (BioLegend), and were sorted in 0.1% BSA 2 mM EDTA-DPBS on a MoFlo Astrios 1 (Beckman Coulter; Brea, CA, USA) 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 1 μ M 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 500 \times g for 5 min and flash-frozen for subsequent RNA isolation.

Procr-mGFP-2A-LacZ mouse line and staining

The Procr-mGFP-2A-LacZ mouse line was generated by knocking in a cassette of mGFP-2A-LacZ behind start codon of the *Procr* gene (32). To perform X-gal staining, embryos were isolated and washed in cold PBS followed by incubation in ice-cold fixative (30 min up to E10.5, or 50 mins for E11.5–E13.5) on a rocking platform. Fixative contains 37% formaldehyde, 25% glutaraldehyde, 10% NP-40 all dissolved in PBS. The fixative was removed and the whole embryo washed twice in PBS for 20 min at room temperature on a rocking platform. The β -galactosidase substrate (5 mM K3Fe(CN)₆, 5 mM K4Fe(CN)₆·3H₂O, 2 mM MgCl₂, 0.02% NP40, 0.1% sodium deoxycholate and 1 mg ml^{–1} X-gal in PBS) was then added and the tissues incubated in the dark overnight at room temperature.

The substrate was removed and the tissues washed twice in PBS for 20 min at room temperature on a rocking platform. Embryos were serially dehydrated using glycerol, and was stored in 80% glycerol at 4°C. Whole-embryo images were captured using Olympus SZX16.

Chromatin isolation

The chromatin isolation for ChRO-seq library preparation was performed as previously described (20,33). 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 12 500 × *g* for 30 min at 4°C. The chromatin pellet was washed three 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 (20,34). Briefly, chromatin from at least 1 × 10⁶ 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 adenosine triphosphate (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: (i) 3' adapter ligation with T4 RNA Ligase 1 (NEB, M0204L), (ii) 5' de-capping with RNA 5' pyrophosphohydrolase (RppH, NEB, M0356S), (iii) 5' end phosphorylation using T4 polynucleotide kinase (NEB, M0201L), (iv) 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 (5' single end; single-end 75×) 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, USA) was used for reverse transcription of RNA. Libraries were generated using the NEBNext Ultra II Directional Library Prep Kit (New England Biolabs, Ipswich, MA, USA) and subjected to sequencing (single-end 92×) 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 (35) 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. For visualization purpose, bigwig files from identical stages were merged and normalized to a total signal of 1 × 10⁶. In the RNA-seq studies, reads were mapped to human genome hg38 using STAR (v2.5.3a) (36) and transcript quantification was performed using Salmon (v0.6.0) (37) with GENCODE release 25 transcript annotations. The expression (RNA-seq) levels of genes were normalized using DESeq2 (38). 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 S3A).

ChRO-seq quantification of gene loci and promoter transcription activity

Gene definitions were obtained from GENCODE v25 annotations. To quantify transcription activity of gene loci, ChRO-seq signals present in annotated gene bodies were used 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 gene body <1000 base were excluded from all the gene body related analysis, given that genes with short gene bodies are likely biased when excluding the pause peak. The ChRO-seq reads were normalized by the length of gene bodies to transcripts per million (TPM). To determine promoter activity

of genes, stranded ChRO-seq signals at the promoter proximal region (500 bps upstream and 200 bps downstream of TSS) were used. The promoter regions that have significant changing patterns were defined using the following criteria: average TPM > 5 and padj < 0.05 across all the stages by likelihood ratio test (DESeq2). The promoter regions that have significant changing patterns and are annotated as ‘protein-coding’ gene type were subject to clustering analysis (‘degPatterns’ in R platform).

Differential expression and pathway analyses of genes

The differential analysis of gene bodies in ChRO-seq data was performed using DESeq2 package (38). For all the analyses except stage-specific marker analysis, 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 (39).

Identification of stage-specific marker genes

First, genes that have ChRO-seq signal in gene bodies (excluding the first 500 bps downstream of the TSS) greater than 50 TPM in the stage of interest are identified and filtered according to padj < 0.2, $P < 0.05$, fold change > 1.5 compared to all the other stages (Wald test; DESeq2). Second, genes that have RNA-seq signal > 100 base mean units are identified and filtered according to padj < 0.2, $P < 0.05$, fold change > 1.5 compared to all the other stages (Wald test; DESeq2). Third, the output from these two analyses are intersected to arrive at final gene lists for stage-specific markers. To identify genes that label SI spheroids irrespective of regional identity, analyses were done using the same criteria except the fold change criterion. The genes that have short gene bodies (<1000 bp) or annotated under the pseudogene category are excluded from this analysis.

TRE identification, annotation, categorization and comparison with other datasets

The active transcriptional regulatory elements (TREs) were identified by dREG tool (35). The annotation of the identified TREs was defined using *annotatePeaks.pl* function (genome = hg38) of HOMER package (40) based on GENCODE v25 annotations. To compare the TRE landscape of SI spheroids with developing and mature SI in humans, DNase-seq datasets from the Roadmap Epigenomics Project were used. Specifically, first, the files E085-DNase.hotspot.fdr0.01.peaks.v2.bed (fetal SI) and E109-DNase.hotspot.fdr0.01.peaks.v2.bed (adult SI) were downloaded in hg19, converted to hg38 using the UCSC liftOver tool and used to define DNase-based open chromatin regions unique to fetal and adult SI using the bedtools intersect function (at least 1 base overlapping). The fetal and adult SI-specific DNase open chromatin regions were further intersected with TREs present in SI spheroids using bedtools intersect function (at least 1 base overlapping). 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.

Stage-specific TRE density analysis

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 log₂fold change > 2.5 by DESeq2) (38). The ChRO-seq intensity was determined by the sum of the unnormalized ChRO-seq reads from both strands within a TRE region. The density of stage-specific TREs was determined by the number of TREs present within the window of +100 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: (i) genes have density of stage-specific TRE > 0, (ii) 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 (iii) 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’ (41,42). Briefly, the stage-specific enhancers in proximity of distance (<12.5 kb) 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 hotspots. 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 hotspot region. To assess overlap between the enhancer hotspots defined in our study with regions of H3K27ac enrichment, we downloaded datasets of H3K27ac peaks from primary human fetal SI (E085-H3K27ac.gappedPeak) from the Roadmap Epigenomics Project. The peak coordinates were first converted from hg19 to hg38 using the UCSC liftOver tool, and then intersected with the enhancer hotspot regions using the bedtools intersect function.

Transcription factor binding motif enrichment analysis

HOMER tool (40) 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 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: (i) genes have motif density > 0, (ii) genes are actively transcribed (TPM > 50 in the ChRO-seq) and expressed (base mean > 500 in the RNA-seq) in the stage of interest, and (iii) 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.

Single cell RNA-seq analysis

The scRNA-seq data of primary human fetal tissues was generated in recent studies (13,43). Briefly, tissue samples were processed into single cell suspension detailed in (13,43). As described in (13), the scRNA-seq dataset of human developing gastrointestinal organs and multiple gestational time points were generated by 10× Genomics platform and mapped to hg19 using default alignment parameters provided by Cell Ranger. Seurat (v3.1) package (44) was applied to perform the downstream analysis. Cells with >5% mitochondrial transcript proportion and fewer than 1000 unique detected genes were excluded. Ribosomal genes, mitochondrial genes and genes located on sex chromosomes were also removed before data integration into a Seurat object. The cell type annotation of the dataset was pre-defined and detailed in (13). In the present study, we focused on the data of matched human fetal duodenum and ileum samples (week 11, 14 and 19 of gestation).

Statistics

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

RESULTS

ChRO-seq reveals temporal dynamics of nascent transcription at gene loci in the directed differentiation model of human developing SI

The directed differentiation of hPSCs (here we used H9 hESCs) into SI spheroids with regional specification was carried out as previously described (26,28) (Figure 1A). The patterned SI spheroids represent the primitive form of the SI comprised of mainly stem/progenitor cells (45), which can give rise to different mature epithelial cell types after prolonged organoid culture, or following transplantation into a murine host (46). In this study, we included the four stages of this model: hESC, DE, duodenum (Duo) spheroid

and ileal (Ile) spheroid (Figure 1A). We carried out ChRO-seq in order to characterize the dynamics of genome-wide nascent transcriptional activity toward the goal of defining the changing landscape of promoters, enhancers and active gene bodies across the following sequential events: DE fate induction (from hESC to DE), SI lineage formation (from DE to Duo spheroid), and initial ileal regional patterning (from Duo to Ile spheroid) (Figure 1A).

To assess nascent transcription of genes, we first analyzed ChRO-seq signal within the body of annotated genes. ChRO-seq data in the first 500 bp downstream of the TSS, was excluded to remove signal contributed by RNA polymerase pausing. Hierarchical clustering analysis of the gene transcriptional profiles demonstrates clean stratification of the different stages (Figure 1B). The signal at gene bodies encoding TFs that are well-established markers of particular stages (*SOX2*, *GATA6* and *CDX2*) indicate the expected specificity (Figure 1C–D). Also, the very minimal expression of the mesenchymal marker *FOXF1* (Figure 1C–D), as well as markers of other organ lineages (Supplementary Figure S1A and B), demonstrate that the Duo and Ile spheroids are indeed dominated by early-stage SI epithelial cells. ChRO-seq can also be used to assess transcriptional activity of short, non-coding regions (e.g. microRNAs). We observed robust ChRO-seq signal within the miR-302–607 cluster, previously reported to be enriched in embryonic stem cells (47), in only the stages of hESC and DE and not in Duo or Ile spheroid (Supplementary Figure S1C). At other microRNA loci, such as miR-10a, we detected an enrichment of ChRO-seq signal in SI spheroids relative to hESC or DE (Supplementary Figure S1D).

Next we sought to identify from the ChRO-seq data differentially transcribed genes during each stage transition of this model (\log_2 fold change > 1, average TPM > 25, padj < 0.2, $P < 0.05$ by Wald test; DESeq2) (Figure 1E). The analyses identified 1886 differentially transcribed genes (1328 up; 558 down) during DE formation (DE versus hESC), 4024 differentially transcribed genes (2546 up; 1478 down) during SI lineage formation (Duo spheroids versus DE) and 362 differentially transcribed genes (182 up; 180 down) during ileal patterning event (Ile versus Duo spheroids) (Figure 1E). We also performed RNA-seq on the same stages to define differentially expressed genes and the Gene Ontology (GO) term enrichment analysis between differentially transcribed and differentially expressed genes are in general agreement (Supplementary Figure S2).

The same ChRO-seq data also affords the unique opportunity to evaluate promoter activity. Therefore, we next developed an approach to analyze promoter dynamics across all of the stages (Supplementary Figure S1E). We identified 9550 genes, the majority of which are protein-coding genes ($n = 7651$), which exhibit significantly changing patterns of promoter activity across different stages (padj < 0.05 by likelihood ratio test; DESeq2) (Supplementary Figure S1E). Notably, a subset of these protein-coding genes has low promoter activity in both hESC and DE but high promoter activity in both Duo and Ile spheroids ($n = 2992$). These include *BEX5* and *PROCR* (Figure 1F), which were recently observed to have robust expression in primary human fetal SI at early developmental time points (7,12). Analysis of the matched RNA-seq data, which showed a dramatic rise

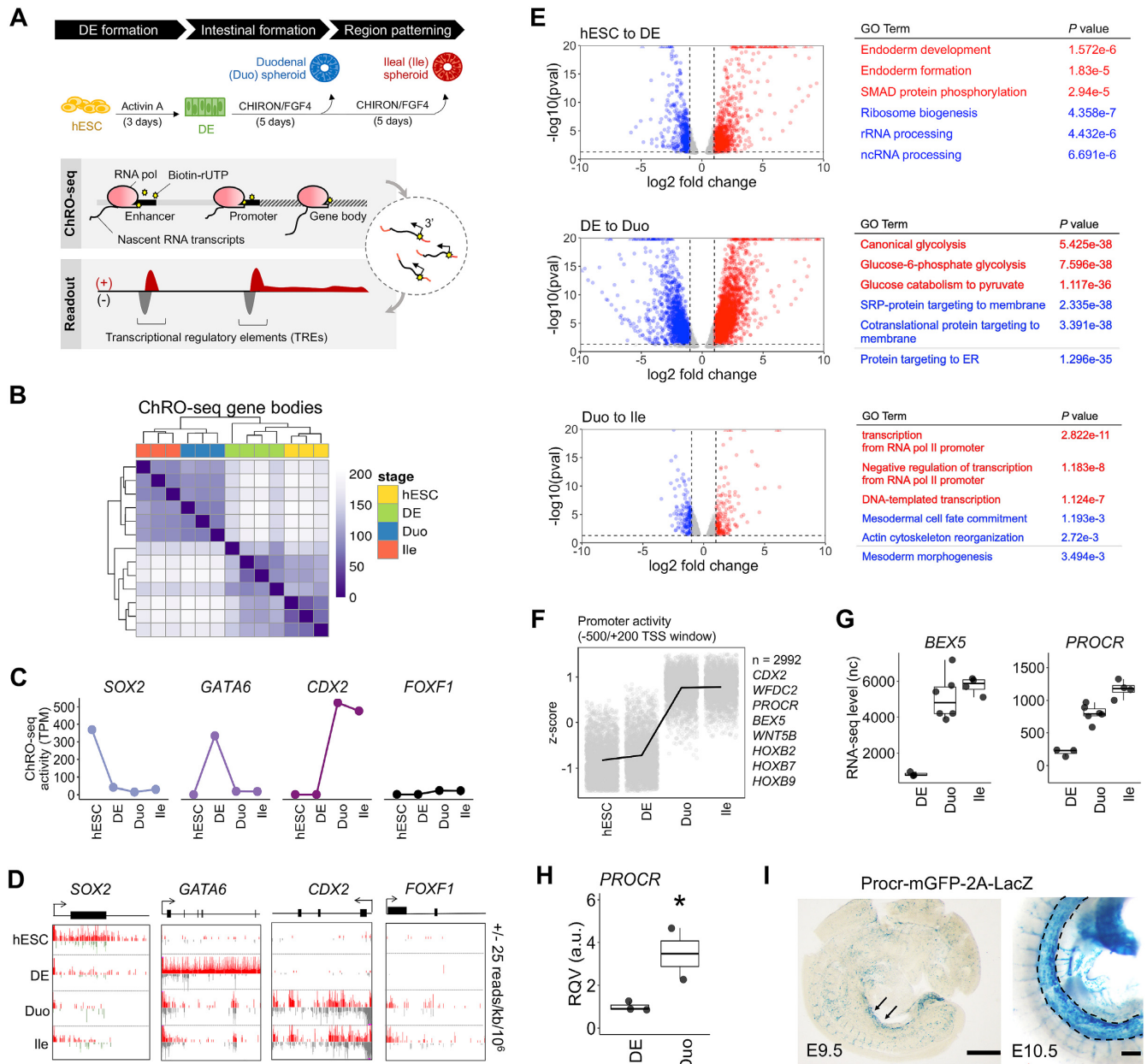


Figure 1. ChRO-seq defines dynamics of nascent transcription at promoters and gene loci during directed differentiation from hPSCs to small intestinal (SI) spheroids with regional specification. **(A)** Schematic diagram of the distinct stages of directed differentiation and the genome-scale approach of ChRO-seq. **(B)** Hierarchical clustering analysis of gene transcription profiles across all stages of the model. **(C)** Transcriptional activity (ChRO-seq signal) of the genes *SOX2*, *GATA6*, *CDX2* and *FOXF1* across all stages of the model. **(D)** Normalized ChRO-seq signal at the gene bodies of *SOX2*, *GATA6*, *CDX2* and *FOXF1* across all stages of the model. Scale (± 25 reads/kb/ 10^6) is fixed across stages. **(E)** Volcano plots showing differentially transcribed genes in the indicated comparisons (average TPM > 25, \log_2 fold change of transcription > 1, padj < 0.2 and $P < 0.05$ by Wald test; DESeq2). Results of pathway enrichment analyses of up-transcribed (red) and down-transcribed (blue) genes in the indicated comparisons (GO Biological Process 2018) are also shown. **(F)** Identification of protein-coding genes that have high promoter activity in SI spheroids ($n = 2992$) through application of the likelihood ratio test across all stages. **(G)** Expression levels (RNA-seq) of *BEX5* and *PROCR* in stages of DE and SI spheroids. **(H)** RT-qPCR of *PROCR* in an independent batch of samples (DE = 3; Duo spheroid = 2). * $P < 0.05$ by one-tailed t -test (validation of sequencing data). **(I)** X-gal staining for β -galactosidase activity of Procr-mGFP-2A-LacZ mouse reporter line showing a robust *Procr* expression in midgut at E9.5 (left; arrowhead) and E10.5 (right; dashed outline). Scale bar = 500 μ m. 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, $n = 6$; Ile, $n = 4$. TPM, transcripts per million. nc, normalized counts. RQV, relative quantitative value.

in *BEX5* and *PROCR* mRNA levels during formation of SI spheroids from DE (Figure 1G), confirm the findings from the promoter analysis. We also validated the upregulation of *PROCR* in SI spheroids by RT-qPCR with an independent batch of samples (Figure 1H) and demonstrated by *in situ* hybridization staining a robust expression of *Procr* in the midgut region of *Procr*-mGFP-2A-LacZ mice (32) at E9.5 and E10.5 (Figure 1I), indicating the conserved early presence of *PROCR* in the developing SI. Overall, these findings demonstrate that hPSC-derived SI spheroids recapitulate features of early developing SI and therefore represent the state-of-the-art model for further study of human SI lineage formation.

Identification of transcriptional markers of specific stages in the directed differentiation model of human developing SI

ChRO-seq signal at gene bodies (TPM > 50 in at least one stage) reveals the changing patterns of transcription across stages (Figure 2A). We sought to identify transcriptional markers of each stage in this model. It is important to note that previously validated SI regional markers using human fetal duodenum and ileum ranging 14–19 weeks of gestation (28) are not sufficient to distinguish Duo from Ile spheroids, although they are sufficient to distinguish Duo from Ile HIOs (derived after culturing the spheroids in Matrigel for 28 days) (Supplementary Figure S3A). This likely reflects the fact that spheroids, which are newly differentiated, represent the early gut lineage whereas the HIOs represent a more mature state. Here, by leveraging ChRO-seq signal at gene bodies as well as RNA-seq data, we were able to define gene sets that serve as early markers of SI lineage formation and ileal regional patterning at the levels of both transcription and steady-state expression.

We first demonstrated that ChRO-seq and RNA-seq signal for genes are reasonably well-correlated (Pearson correlation coefficient = 0.73) (Figure 2B and Supplementary Figure S3B–E), 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). To identify genes that label a specific stage at the levels of both transcription and steady-state expression, we developed a bioinformatic pipeline to integrate ChRO-seq and RNA-seq data ('Materials and Methods' section) and performed this integrative analysis with genes that are actively transcribed (TPM > 50) in at least one stage. We identified genes that are significantly elevated according to both transcription and steady-state 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 2C and Supplementary Data 2). The stage-specific genes we identified include well-established markers, such as *POU5F1* and *SOX2* for hESCs (Figure 2D) and *NODAL*, *SOX17*, *EOMES* and *LEFTY2* for DE (Figure 2E). Notably, while these known endoderm markers are highly transcribed in the stage of DE, markers of mesendoderm (e.g. *TBXT* and *PAX6*) are lowly transcribed in the DE stage (< 5 TPM; data not shown), suggesting that the DE cells used in this study are indeed highly enriched in cells with DE

identity, instead of mesendoderm cells that still have dual potential for endoderm and mesoderm. Furthermore, consistent with the literature, *CDX2* was identified to faithfully label SI spheroids irrespective of regional identity (Figure 2F). We defined additional stage-specific markers, many of which are previously undescribed: *HES4*, *FOXA1*, *WFDC2* and several *HOXB* family members for SI spheroids irrespective of regional identity; *SP5* and *SALL1* for Duo spheroids; and *FJX1*, *IGF2*, as well as several *HOXA/C/D* members for Ile spheroids (Figure 2F and G). We also identified several long, non-coding RNA (lncRNA) and antisense transcript markers: *LINC00543* for SI spheroids irrespective of regional identity, and *EVX1-AS* for Ile spheroids (Figure 2F and G). Notably, while many genes were defined as markers of SI spheroids irrespective of regional identity, the analysis identified much fewer genes as Duo-specific markers compared than Ile-specific markers (Figure 2C), which may suggest that Ile regional specification is mainly driven by gain of additional driver genes rather than loss of genes contributing to Duo regional identity.

Transcriptional regulatory element profiles reveal notable rewiring of chromatin activity during directed differentiation

Active TREs, including promoters and enhancers, are identified in ChRO-seq data by the hallmark feature of short bi-directional transcription (Figure 3A). To identify active TREs across the entire genome, we employed dREG (35), which was developed specifically for this purpose. Using this tool, we identified a total of 125863 active TREs across all four stages included in this study (Figure 3B). The length distribution of these active TREs is consistent with what has been reported previously (48) (Supplementary Figure S4A) and the vast majority of the active TREs, as expected, are located in intergenic regions, introns, and annotated TSSs (Supplementary Figure S4B). To specifically evaluate whether the TRE landscape of SI spheroids is more similar to developing as opposed to mature SI in humans, we leveraged data on DNase-accessible regions in both fetal and adult human SI available from the Roadmap Epigenomics Project. We first defined DNase-accessible regions unique to fetal or adult tissue ('Materials and Methods' section) and then intersected those with active TREs present in Duo and/or Ile spheroids. We found that 12541 out of 37892 (~33%) open chromatin loci specific to fetal human SI overlap with TREs present in Duo and/or Ile spheroids, whereas only 7504 out of 45242 (~16%) open chromatin loci specific to adult human SI overlap with TREs present in Duo and/or Ile spheroids (Figure 3C). To provide a specific example of a TRE defined in this model that captures an important regulatory region associated with human fetal SI, we turned to the very recently identified intestinal critical region (ICR). The ICR represents an important regulatory element of the new gene *PERCC1*, which exhibits transient expression in the SI during development and has been linked to congenital diarrheal disorders in humans (49). We identified a TRE that is highly overlapping with the ICR and active only in SI spheroids, especially in Duo spheroids (Figure 3D and E). Taken together, these observations demonstrate that TREs in hPSC-derived SI spheroids

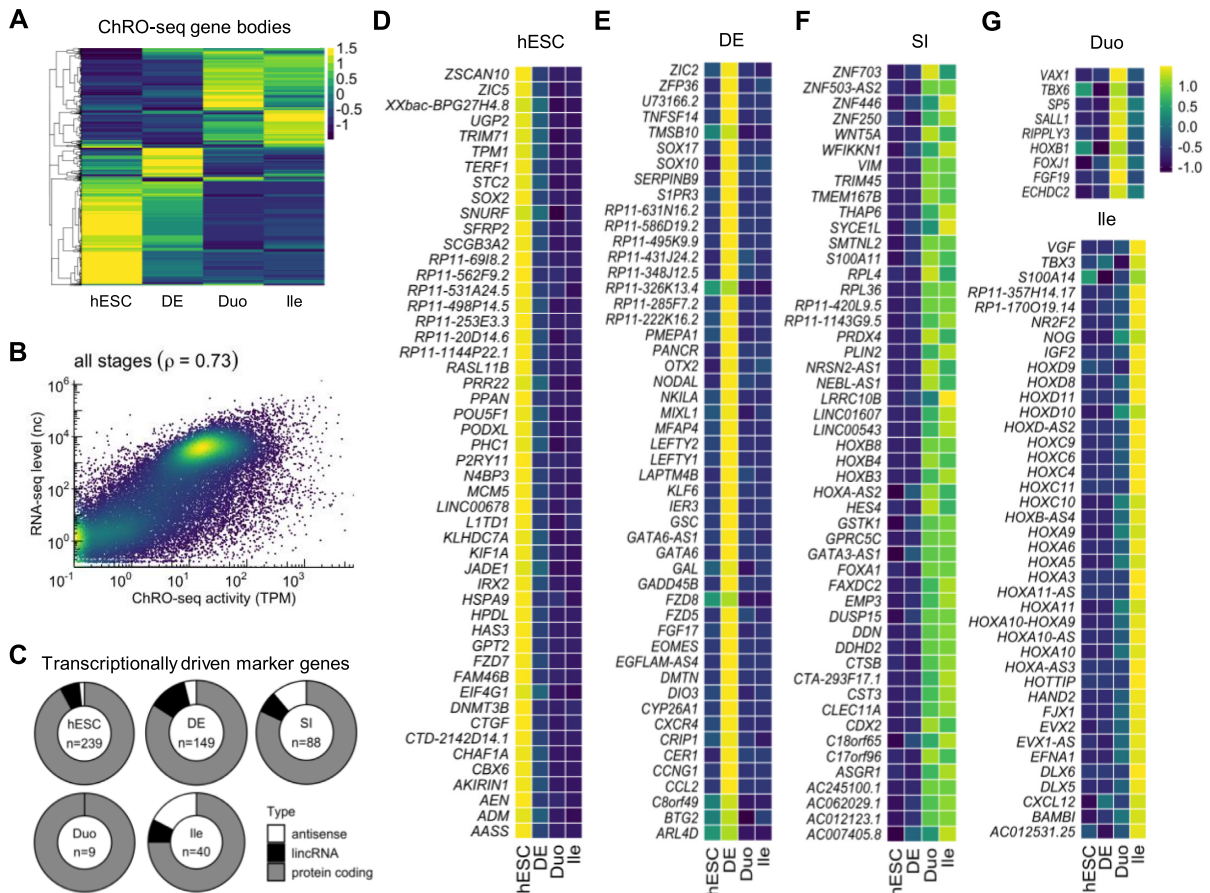


Figure 2. Identification of transcriptional markers of specific stages in the directed differentiation model of human developing SI. (A) Hierarchical clustering analysis of transcribed genes (TPM > 50 at least in one stage) (z -score). (B) Genome-wide correlation of transcribed levels (ChRO-seq) and expressed levels (RNA-seq) of genes. Transcription and expression levels of genes were averaged across all stages. No transcription and expression thresholds were used (see also Supplementary Figure S3B–E). (C) Identification of markers that label specific stages at the levels of both transcription and steady-state expression. The total number of markers (center of the donut plot) and the proportion of marker types are shown for each stage. (D–G) Marker genes of hESC (D), DE (E), SI spheroids regardless of regional identity (F), Duo and Ile (G). In panel (D–F), only the top 50 most variable genes across stages are shown (see Supplementary Data 2 for full list). Heatmaps denote variation across stages (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, $n = 6$; Ile, $n = 4$.

are able to capture important features of gene regulatory landscapes that are associated with the developing human SI.

We next categorized the identified active TREs into 49855 proximal and 76008 distal TREs, which from here on in we refer to as promoters and enhancers, respectively (Figure 3A). Unsupervised hierarchical clustering analyses reveal that enhancer profiles stratify different stages more accurately and clearly than promoters (Figure 3F and G), consistent with the notion that enhancer signature is the most cell-type specific (14,16). To further elucidate the dynamics of active TRE profiles across the directed differentiation process, we identified active TREs that are unchanged, gained (enhanced), or lost (suppressed) during the transition to DE (Figure 3H), SI lineage formation (Figure 3I), and ileal patterning (Figure 3J) ($\log_2FC > 2.5$, $p_{adj} < 0.05$ by Wald test; DESeq2). We found that a much greater proportion of enhancers, relative to promoters, exhibit significantly altered levels of activity (either gain or loss) across all stage transitions. In contrast, enhancers and promoters were roughly equally represented among those TREs that

were unchanged in activity across stage transitions (Figure 3H–J).

Identification of genes associated with high a density of nearby enhancers that emerge during SI lineage formation or ileal patterning

Next we developed an analysis pipeline to further investigate the enhancers and associated genes that emerge during different stages of the directed differentiation (Figure 4A). We focused on ‘Duo-specific enhancers’ and ‘Ile-specific enhancers’, which are the enhancers that gain activity during Duo and Ile spheroid formation, respectively (Figure 3I and J). We then determined the density of stage-specific enhancers for every actively transcribed gene (TPM > 50 in the stage of interest) by counting stage-specific enhancers within a 200 kb window centered on the TSS (Figure 4A). The genes that are associated with the stage-specific enhancers and also have significant activation in both ChRO-seq and RNA-seq are eventually defined as stage-specific enhancer linked genes (Figure 4A).

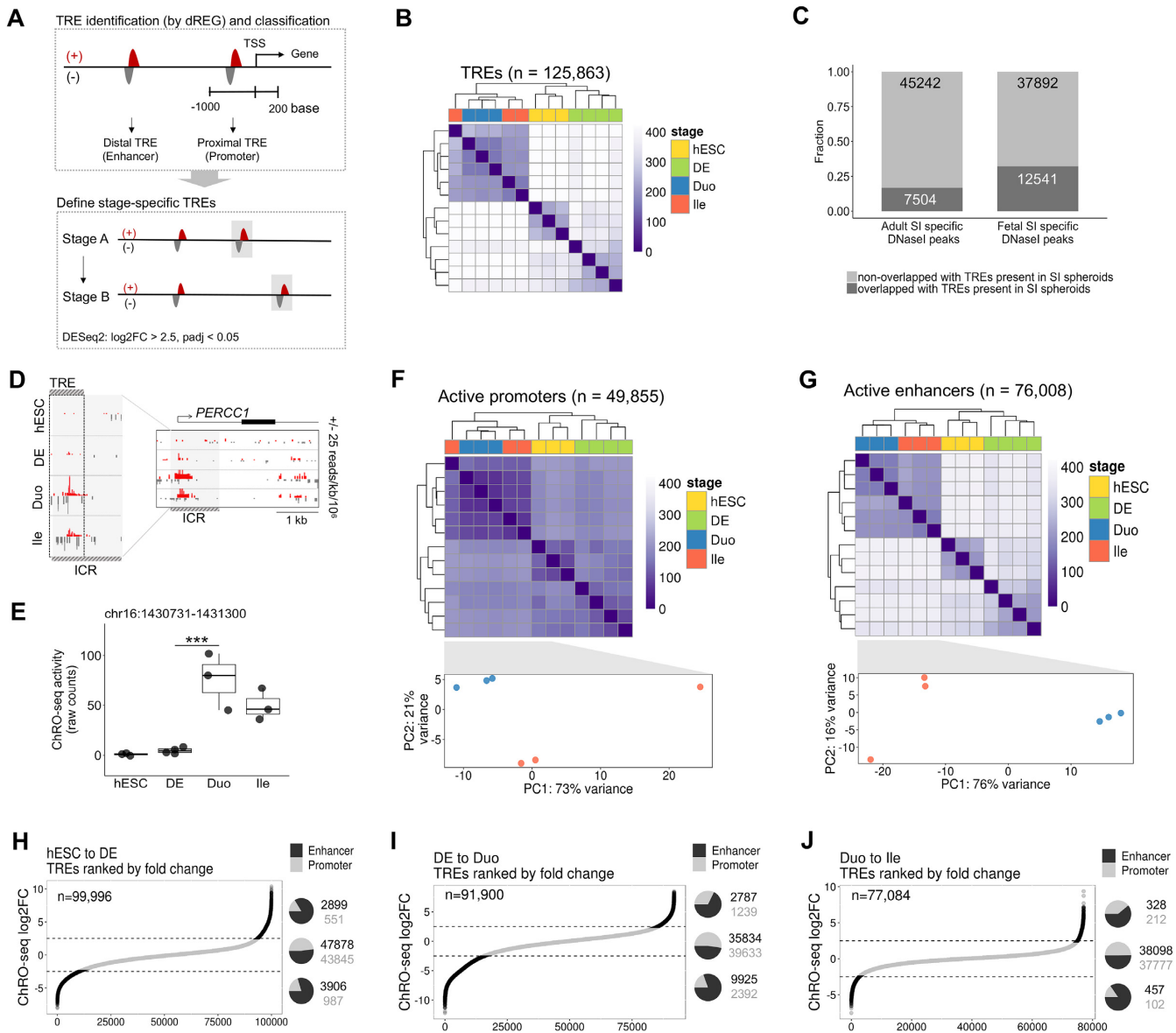


Figure 3. TRE profiles reveal notable re-wiring of chromatin activity in the directed differentiation model of human developing SI. (A) Schematic for identification and categorization of active TREs, followed by definition of TREs with altered activity during stage transitions of the model. (B) Hierarchical clustering analysis of profiles of all active TREs. (C) Fraction of fetal and adult SI-specific open chromatin regions (defined according to DNaseI-seq data from the Roadmap Epigenomics Project) overlapping with active TREs present in Duo and/or Ile spheroids. (D) ChRO-seq signal around the ICR (49). An active TRE identified in our study (chr16: 1430731–1431300; hg38) largely overlaps with the ICR and exhibits strong ChRO-seq signals in SI spheroids. Scales in right panel (+25/–25) and left panel (+166/–15; the maximum range) are fixed across different stages. (E) The transcriptional activity of the TRE associated with the ICR (chr16: 1430731–1431300; hg38) across all stages of the model (***) $P < 0.001$ by Wald test; DESeq2). (F) Hierarchical clustering (upper) and PCA (lower) analysis of promoter profiles in the indicated stage. (G) Hierarchical clustering (upper) and PCA (lower) analysis of enhancer profiles in the indicated stage. (H) Active TREs present in hESC and DE ($n = 99996$) and ranked based on the log₂ fold change in activity during the transition to DE. (I) Active TREs present in DE and Duo spheroid ($n = 91900$) and ranked based on the log₂ fold change in activity during Duo spheroid formation. (J) Active TREs present in Duo and Ile spheroid ($n = 77084$) and ranked based on the log₂ fold change in activity during Ile spheroid formation. In (H–J), the proportion and the number of promoters and enhancers that are gained (pie chart on the top), unchanged (pie chart in middle) or lost (pie chart at the bottom) are indicated in each stage transition. hESC, $n = 3$; DE, $n = 4$; Duo spheroid (Duo), $n = 3$; Ile spheroid (Ile), $n = 3$.

As defined in Figure 3, we identified a total of 2787 Duo-specific and 328 Ile-specific enhancers (Figure 4B and C). We confirmed that genes associated with a greater number of Duo-specific enhancers also exhibit greater increases in transcription in Duo spheroids relative to DE (Figure 4D and E). Similarly, genes associated with a greater number of Ile-specific enhancers also exhibit greater increases in tran-

scription in Ile relative to Duo spheroids (Figure 4I and J). The results of similar analyses focused on hESC and DE are shown in Supplementary Figure S5. These observations support the model in which the level of transcriptional activation during each stage transition is strongly associated with the total number of emerging TREs around TSSs.

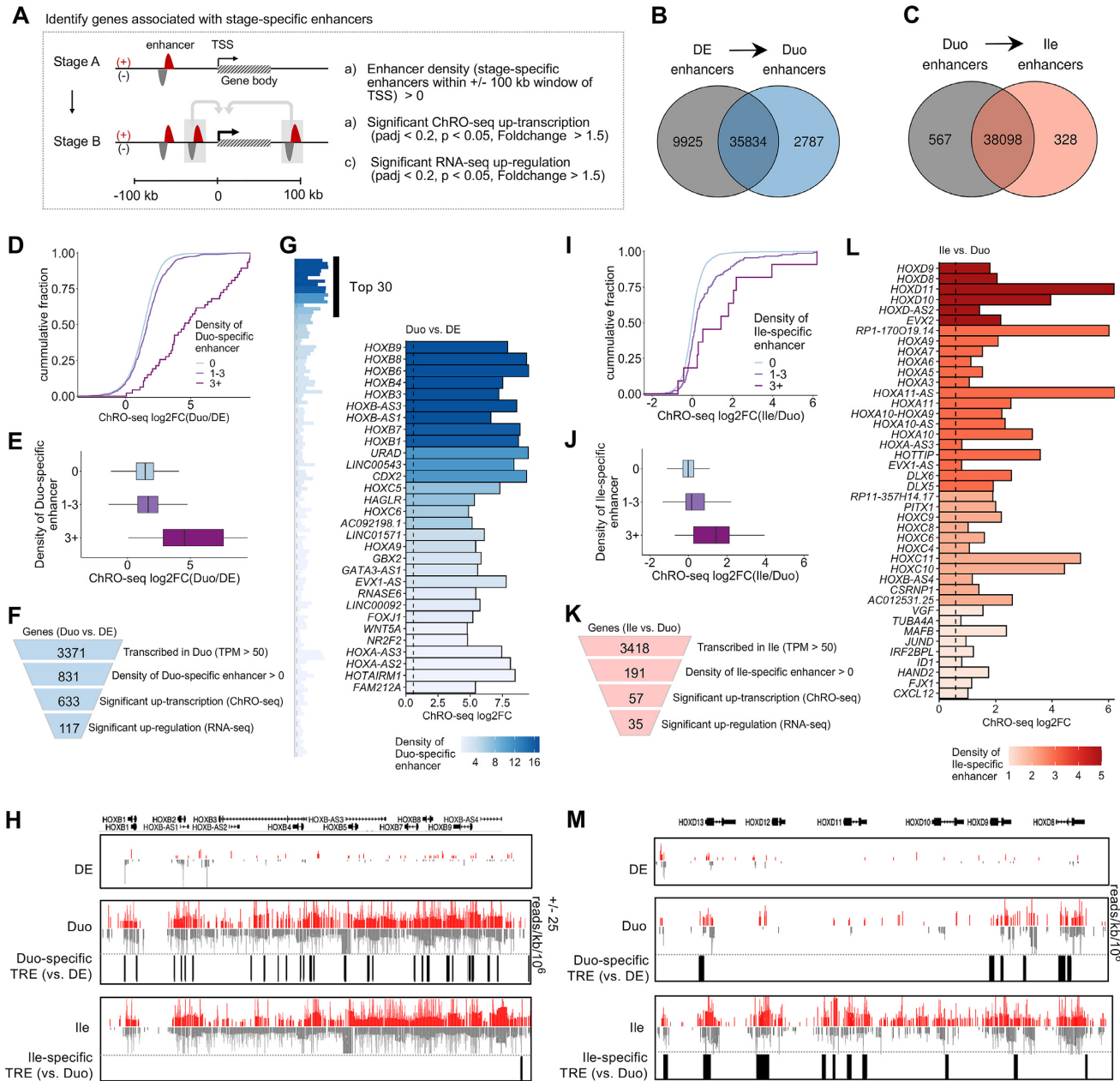


Figure 4. Identification of genes associated with a high density of nearby enhancers that emerge during SI lineage formation or ileal patterning. (A) The strategy for identifying genes associated with stage-specific enhancers. (B) Venn diagram showing stage-specific and shared enhancers between DE and Duo spheroids. Duo-specific enhancers ($n = 2787$) represent enhancers emerging during SI lineage formation. (C) Venn diagram showing stage-specific and shared enhancers between Duo and Ile spheroids in the event of ileal patterning. Ile-specific enhancers ($n = 328$) represent enhancers emerging during ileal regional patterning. (D and E) Cumulative distribution (top) and boxplot (bottom) of ChRO-seq fold change in genes grouped into three different categories of Duo-specific enhancer density during the transition from DE to Duo spheroids. (F) Identification of genes associated with Duo-specific enhancers ($n = 117$). (G) Bar graph showing genes associated with Duo-specific enhancers (left panel). Top 30 genes based on enhancer density are highlighted (right panel). (H) Normalized ChRO-seq signal around the *HOXB* cluster in the stages of DE, Duo and Ile spheroids. Duo-specific TREs are marked. Scale (+25/-25) is fixed across different stages. (I and J) Cumulative distribution (top) and boxplot (bottom) of ChRO-seq fold change in genes grouped into three different categories of Ile-specific enhancer density during the formation Ile spheroids. (K) Identification of genes associated with Ile-specific enhancers ($n = 35$). (L) Bar graph showing genes associated with Ile-specific enhancers. (M) Normalized ChRO-seq signal around the *HOXD* cluster in the stages of DE, Duo and Ile spheroids. Ile-specific TREs are marked. Scale (+25/-25) is fixed across different stages. ChRO-seq study: DE, $n = 4$; Duo spheroid (Duo), $n = 3$; Ile spheroid (Ile), $n = 3$. RNA-seq study: DE, $n = 3$; Duo, $n = 6$; Ile, $n = 4$.

Regarding the transition from DE to Duo spheroids, we identified 117 genes that are associated with at least one Duo-specific enhancer and are significantly increased in both transcription (ChRO-seq) and steady-state expression (RNA-seq) (Figure 4F). Among these, as expected, *CDX2* is one of the top genes ranked by the density of Duo-specific enhancers (Figure 4G). Notably, even more highly ranked are the *HOXB* family members (Figure 4H). Regarding ileal regional patterning, we identified 35 genes that are associated with at least one Ile-specific enhancer and are significantly increased in both transcription and steady state-expression in Ile relevant to Duo spheroids (Figure 4K). Among these, the genes associated with the greatest number of Ile-specific enhancers include members of the *HOXA/C/D* clusters, *FJX1*, *CSRNP1* and *HOTTIP* (Figure 4L and M), several of which were also identified as Ile spheroid-specific marker genes (Figure 2G). These analyses together reveal previously undescribed temporal dynamics of the *HOX* cluster genes during the events of SI lineage formation followed by ileal regional patterning: first, the *HOXB* cluster is activated during Duo spheroid formation (likely by nearby Duo-specific enhancers), then subsequently the other three *HOX* clusters are activated during ileal patterning (likely by nearby Ile-specific enhancers) (Figure 4H and M).

Identification of genes associated with enhancer ‘hotspots’ that emerge SI lineage formation or ileal 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 (41–42,50). We sought to define for the first time the changing landscape of enhancer hotspots in this human model of SI development. To accomplish this, we adapted a previously described methodology (41,42), which requires ChIP-seq based datasets, to work with ChRO-seq data instead (Figure 5A). By implementing this analysis pipeline with the Duo- and Ile-specific enhancers (defined in Figure 4B and C, respectively), we were able to identify enhancer hotspots and the associated genes relevant to SI lineage formation and ileal patterning, respectively.

To identify enhancer hotspots that emerge during SI lineage formation and ileal patterning, we clustered Duo- and Ile-specific enhancers, respectively, into ‘stitched enhancers’ (Supplementary Data 3). Among the 2389 stitched enhancers formed by Duo-specific enhancers (relative to DE), we identified 145 that exhibit strong enough transcriptional activity to be designated as Duo-specific ‘enhancer hotspots’ (Figure 5B). Among the 305 stitched enhancers formed by Ile-specific enhancers (relative to Duo spheroids), we defined 29 Ile-specific enhancer hotspots (Figure 5C). Similar analyses were performed with hESC- and DE-specific enhancers (Supplementary Figure S6A and E). Given that H3K27ac ChIP-seq signal has been successfully leveraged in previous studies to identify enhancer cluster regions (42), we compared the stage-specific enhancer hotspots with a published database of H3K27ac peaks in primary human fetal SI (from the Roadmap Epigenomics Project) (Figure 5D). We observed that nearly 50%

of Duo- and 55% of Ile-specific enhancer hotspots substantially overlap (at least half of the length) with regions of H3K27ac enrichment, whereas this is the case for only 21% of hESC- or 26% of DE-specific enhancer hotspots. (Figure 5D). This suggests that Duo- and Ile-specific enhancer hotspots defined in the hPSC-HIO model resemble, though of course are not identical to, active enhancer regions in primary human fetal SI. Moreover, we found that the active enhancers defined by our ChRO-seq analysis provide higher resolution in defining the boundaries of each individual enhancer within a given enhancer hotspot region, compared to the narrow, stringent peaks defined by H3K27ac activity (Supplementary Figure S7A and B).

Next, we performed an analysis to identify the genes likely associated with Duo- and Ile-specific enhancer hotspots. We assigned each of the Duo-specific stitched enhancers (including non-hotspots and hotspots) to its nearest active gene (TPM > 50 in ChRO-seq in Duo spheroids) within a 100 kb window on either end of the boundaries of the stitched enhancer. We found that the overall increase in gene body transcription is significantly greater for the set of genes associated with Duo-specific enhancer hotspots compared to those associated with stitched enhancers that are not hotspots (Figure 5E). We also assigned each of the Ile-specific stitched enhancers (including non-hotspots and hotspots) to its nearest active gene (TPM > 50 in ChRO-seq in Ile spheroids) using the same criteria. Similar to the observation made for the transition from DE to Duo spheroids, the genes nearest to Ile-specific enhancer hotspots exhibit a significantly greater increase in transcription compared to those associated with non-hotspot stitched enhancers (Figure 5F).

These findings are consistent with the notion that enhancer hotspots exert particularly strong effects on transcription. Among the 73 genes nearest to Duo-specific enhancer hotspots, 25 exhibit highly significant increases in both transcription (ChRO-seq) and steady state expression (RNA-seq) in Duo spheroids relative to DE (Figure 5D and E). Many of these were defined earlier as SI or Duo-specific markers (e.g. *CDX2*, *FOXA1*, *HOXB* cluster, *SALL1* and *LINC00543*) (Figure 2; Supplementary Figure S7A). Among the 22 genes nearest 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 5H). These genes include *HAND2*, *HOXC/D* family members, *FJX1*, *DLX5*, *CSRNP1* and *PITX1* (Figure 5I and Supplementary Figure S7B), 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 S6.

Identification of candidate TF drivers and their networks relevant to SI lineage formation or ileal patterning

To discover putative key TF drivers and their candidate cistromes (genome-wide set of TF targets) in this model, we first used the tool HOMER to determine motifs enriched in each set of stage-specific enhancers and then identified genes associated enhancers that harbor specific motifs of interest (Figure 6A). Expectedly, HOMER analyses

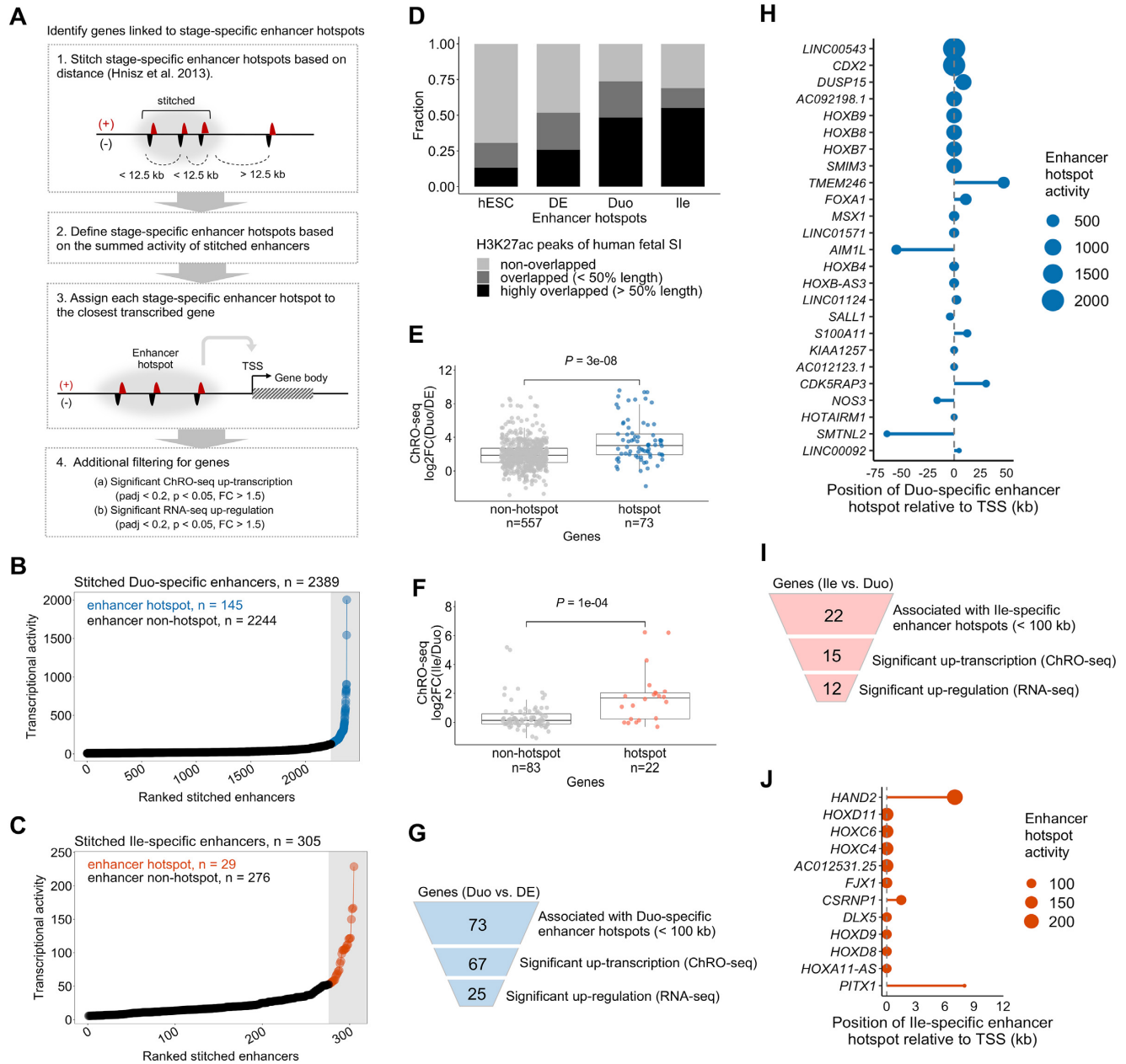


Figure 5. Identification of genes associated with enhancer ‘hotspots’ that emerge during SI lineage formation or ileal patterning. (A) The strategy for identification of stage-specific enhancer hotspots and their associated genes. (B) Duo-specific stitched enhancers are ranked by transcriptional activity (ChRO-seq signal). The stitched enhancers with the highest transcriptional activity are defined as Duo-specific enhancer hotspots ($n = 145$; blue) and the rest are enhancer non-hotspots ($n = 2244$; black). (C) Ile-specific stitched enhancers are ranked by transcriptional activity (ChRO-seq signal). The stitched enhancers with the highest transcriptional activity are defined as Ile-specific enhancer hotspots ($n = 29$; red) and the rest are enhancer non-hotspots ($n = 276$; black). (D) The fraction of stage-specific enhancer hotspots regions (defined by ChRO-seq) that are non-overlapping, moderately overlapping ($\leq 50\%$ length of enhancer hotspots), or highly overlapping ($> 50\%$ length of enhancer hotspots) with H3K27ac peaks from primary human fetal SI (the Roadmap Epigenomics Project) is shown. (E) ChRO-seq fold change of genes associated with Duo-specific stitched enhancers, non-hotspots versus hotspots (Wilcoxon rank sum test). (F) ChRO-seq fold change of genes associated with Ile-specific stitched enhancers, non-hotspots versus hotspots (Wilcoxon rank sum test). (G) Identification of genes associated with Duo-specific enhancer hotspots. (H) Genes associated with Duo-specific enhancer hotspots ($n = 25$). Relative position between enhancer hotspots and TSSs of the associated genes are shown. Dot size denotes transcriptional activity of a given DE-specific enhancer hotspot. (I) Identification of genes associated with Ile-specific enhancer hotspots. (J) Genes associated with Ile-specific enhancer hotspots ($n = 12$). Relative position between enhancer hotspots and TSSs of the associated genes are shown. Dot size denotes transcriptional activity of a given Duo-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, $n = 6$; Ile, $n = 4$.

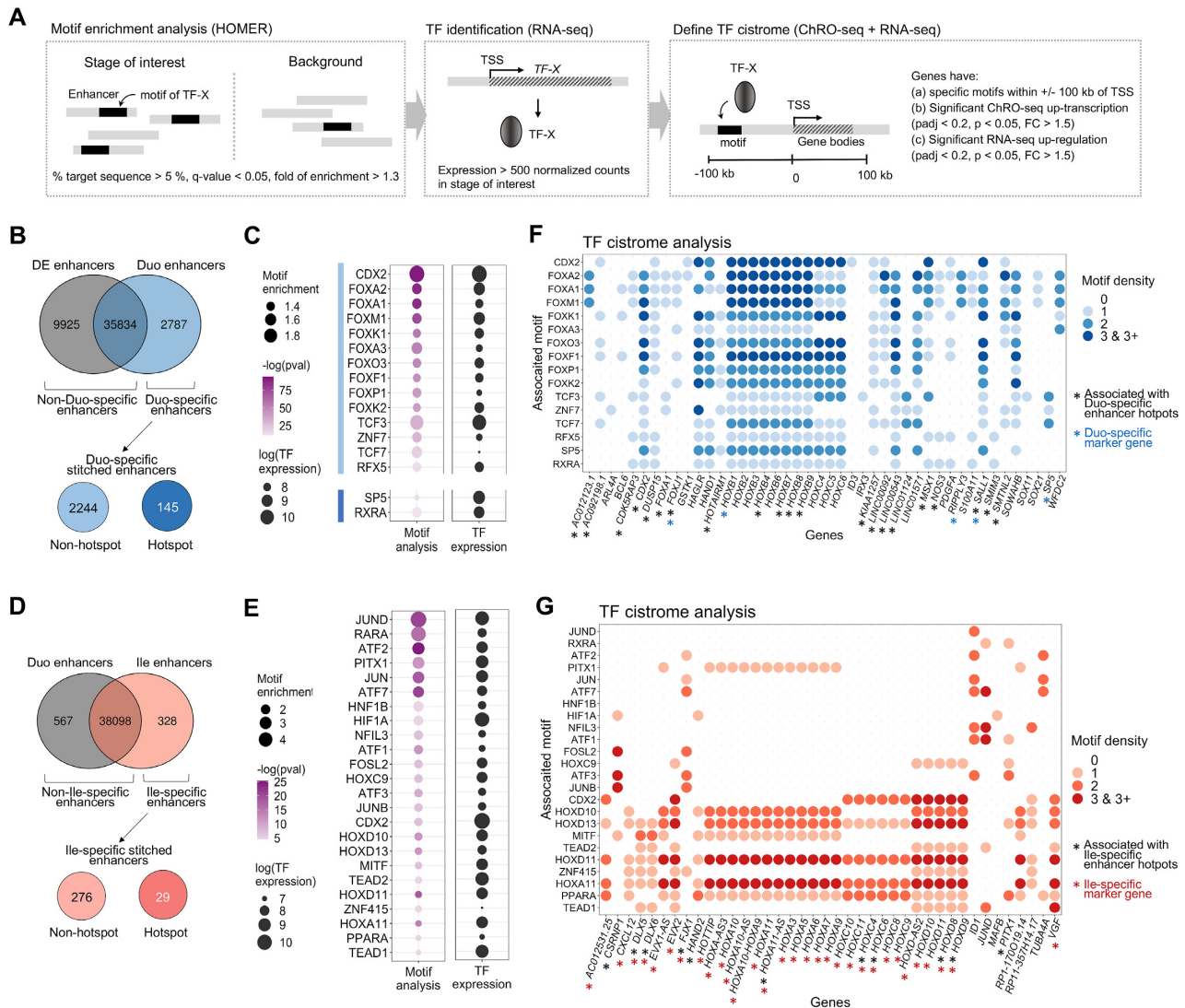


Figure 6. Identification of candidate TF drivers and their networks relevant to SI lineage formation or ileal patterning. (A) The strategy for identifying candidate TFs and their cisomes associated with each stage transition of the model. (B) TF motif enrichment analyses were performed in Duo-specific enhancers ($n = 2787$) relative to non-Duo-specific enhancers ($n = 45759$) as well as in Duo-specific enhancer hotspots ($n = 145$) relative to non-hotspots ($n = 2244$). (C) Motifs significantly enriched in Duo-specific enhancers (highlighted by light blue bar) or in Duo-specific enhancer hotspots (highlighted by dark blue bar). The steady-state expression (RNA-seq) of the corresponding TFs are also shown. (D) TF motif enrichment analyses were performed in Ile-specific enhancers ($n = 328$) relative to non-Ile-specific enhancers ($n = 38674$) as well as in Ile-specific enhancer hotspots ($n = 29$) relative to non-hotspots ($n = 276$). (E) Motifs significantly enriched in Ile-specific enhancers. The steady-state expression (RNA-seq) of the corresponding TFs are also shown. (F) Bubble plot showing genes associated with active binding motifs of TFs that exhibit an overall enrichment of binding motifs in Duo-specific enhancers (defined in C). See also Supplementary Figure S6E for full gene list. (G) Bubble plot showing genes associated with active binding motifs of TFs that exhibit an overall enrichment of binding motifs in Ile-specific enhancers (defined in E). ChRO-seq study: DE, $n = 4$; Duo spheroid (Duo), $n = 3$; Ile spheroid (Ile), $n = 3$. RNA-seq study: DE, $n = 3$; Duo, $n = 6$; Ile, $n = 4$.

of hESC- and DE-specific enhancers revealed binding sites of POU5F1 and SMAD2 to be the most enriched motifs, respectively (Supplementary Figure S8A–D). We next performed motif enrichment analyses on Duo- (Figure 6B) and Ile-specific enhancers (Figure 6D).

As expected, among the 2787 Duo-specific enhancers relative to non-Duo specific enhancers (Figure 6B), we detected significant enrichment for the binding motifs of CDX2 and many forkhead box TFs (Figure 6C). We also performed the same analysis with Duo-specific enhancer hotspots relative to non-hotspots (Figure 6B) and observed significant enrichment for binding motifs of SP5 and RXR

(Figure 6C), the former of which is defined as a Duo-specific gene marker (Figure 2G), but neither of which have been associated previously with early SI development. Genes that are associated with Duo-specific enhancers harboring motifs of one or more of these TFs and that are significantly increased in transcription (ChRO-seq) and steady-state expression (RNA-seq) are shown in Figure 6F (see full gene list in Supplementary Figure S8E). The TF cisrome analysis suggests that CDX2 is a prominent transcriptional activator of the *HOXB* cluster (Figure 6F), which is consistent with previous observations (8). Moreover, this analysis reveals candidate TF activators of the gene *CDX2*. While

CDX2 locus has a *CDX2* motif in its promoter region in Duo-spheroids (data not shown) and is thus likely subject to autoregulation as previously described (51), Duo-specific enhancers that are associated with the *CDX2* locus do not harbor any binding motifs for *CDX2* and instead have binding motifs for other TFs such as *FOXA1* and *FOXO3* (Figure 6F). More importantly, through this analysis we were able to identify candidate TF drivers of SI/Duo spheroid marker genes that do not appear to be a part of the *CDX2* cistrome. For example, *WFDC2* and *SALL1* (defined earlier as a marker gene of SI and Duo spheroids respectively) are associated with *FOXA1* and *FOXO3* motifs; *SP5* (defined earlier as a marker gene of Duo spheroid) is associated with the *TCF3* motif (Figure 6F).

We next performed a similar analysis for ileal regional patterning. Among the 328 Ile-specific enhancers relative to non-Ile specific enhancers (Figure 6D), we detected significant enrichment for the binding motifs of several key TF families including *CDX2*, *HOXC9*, *HOXA11* and *PITX1* (Figure 6E), the last of which is itself associated with Ile-specific enhancer hotspots (Figure 5I). We also performed the same analysis with Ile-specific enhancer hotspots relative to non-hotspots but did not observe significant enrichment for binding motifs (data not shown). Genes that are associated with Ile-specific enhancers harboring motifs of one or more of these TFs and that are significantly increased in transcription (ChRO-seq) and steady-state expression (RNA-seq) are shown in Figure 6G. This analysis led to several new observations. Firstly, *CDX2* likely plays a role in both Duo spheroid formation and ileal regional patterning, but apparently through distinct cistromes (Figure 6F and G). Notably, the activation of *HOXD* and *HOXC* (especially *HOXC8* and *9*) loci is associated with Ile-specific TREs harboring the *CDX2* motif (Figure 6E). While the *HOXB* cluster is induced in Duo spheroids, it remains highly transcribed in Ile spheroids, likely due to the fact that the nearby TREs with *CDX2* motifs that are activated during Duo spheroid formation are retained during ileal regional patterning (Figure 4H, and M). Secondly, we found that the activation of *HOXA*, *C* and *D* families are associated with overlapping but distinct TF regulators. Specifically, while *HOXA11* seems to act on all four *HOX* clusters, *PITX1* is associated only with the *HOXA* cluster and *HOXC9* is preferentially associated with the *HOXD* cluster (Figure 6G). Finally, this analysis also reveals other TF cistrome networks that do not target *HOX* genes during ileal patterning (Figure 6G). For example, *CSRNP1*, *FJX1* and *IDI* loci are associated with motifs of AP-1 (e.g. *JUNB* and *FOSL2*) and/or ATF factors (Figure 6G).

The *HOX* cluster dynamics observed in the directed differentiation model of SI is present in the early developmental stages of primary human fetal SI tissue

Across various analyses in this study we repeatedly observed a previously undescribed dynamics of *HOX* gene activation across SI lineage formation to ileal patterning. Notably, ChRO-seq data reveal strong transcriptional activation of the *HOXB* cluster (particularly *HOXB1*) upon acquisition of SI identity (Duo spheroids) and the sequential transcriptional activation of members of *HOXA*, *C*, *D*

clusters during ileal regional patterning (Ile spheroids) (Figure 7A), which as we showed previously is also reflected in the pattern of emergence of nearby enhancers and enhancer hotspots (Figures 4 and 5). This unique *HOX* cluster dynamics is largely reflected at the level of steady-state expression (RNA-seq) also (Figure 7B), except for the 3'-most *HOXA* genes (e.g. *HOXA1*, *A2*) and 5'-end *HOXB* genes (e.g. *HOXB7*, *B9*) (Figure 7A and B), which may be subject to post-transcriptional regulation.

To determine whether the differences in *HOX* gene expression between early Duo and Ile extends beyond this *in vitro* directed differentiation model, we mined the single cell RNA-seq (scRNA-seq) dataset of primary human Duo and Ile tissue at very early stages of fetal development (13,43). Specifically, we focused on matched Duo and Ile fetal tissue at week 11, 14 and 19 of gestation. It is important to note that these developmental time points, though considered as early/mid gestation stage, represent time points much later than what is reflected in our much earlier stage SI spheroids and therefore exhibited weaker overall expression of *HOX* genes. Since the SI spheroids generated from the directed differentiation method are mainly comprised of epithelial cells but also contain some mesenchymal cells, we analyzed both cell types from the scRNA-seq data, together and separately, in order to examine the *HOX* dynamics. At week 11 of gestation, we found that the *HOXB* cluster is active in both human fetal Duo and Ile cells (epithelium plus mesenchyme), with *HOXB1* particularly enriched in Duo cells and 5' *HOX* genes of all four clusters enriched in Ile cells (Figure 7C). This distinct *HOX* pattern between primary human fetal Duo and Ile tissues is consistent with our observations in the Duo and Ile spheroids generated from the directed differentiation method. Interestingly, this signal is present only at week 11 and dissipates as the developmental process proceeds (week 14 and 19) (Figure 7C). The scRNA-seq analysis also offers the new insight that the enrichment of *HOXB1* in primary fetal Duo tissue is mainly driven by the mesenchymal fraction, whereas the activation of 5' *HOX* genes of all four clusters in Ile tissue occurs in both epithelial and mesenchymal cells (Figure 7C). Overall, this analysis suggests that the Duo and Ile spheroids generated from the directed differentiation model appear to capture the region-specific *HOX* features that are transiently present in the developing human SI.

Beyond the finding of the previously undescribed temporal *HOX* cluster patterns highlighted here, the ChRO-seq analyses in this study provide a comprehensive characterization of chromatin regulatory dynamics, enhancer activity profiles, and transcriptional programs that are associated with human SI lineage formation and regional patterning. A working model is provided in Figure 7D.

DISCUSSION

ChRO-seq is capable of high-resolution definition (approximately single nucleotide resolution) of regulatory elements, which offers the opportunity to more precisely define the boundary of promoters and enhancers. ChRO-seq also captures active regulatory elements, as opposed to open chromatin regions which represent candidate regulatory elements. By leveraging ChRO-seq, we generated for the first

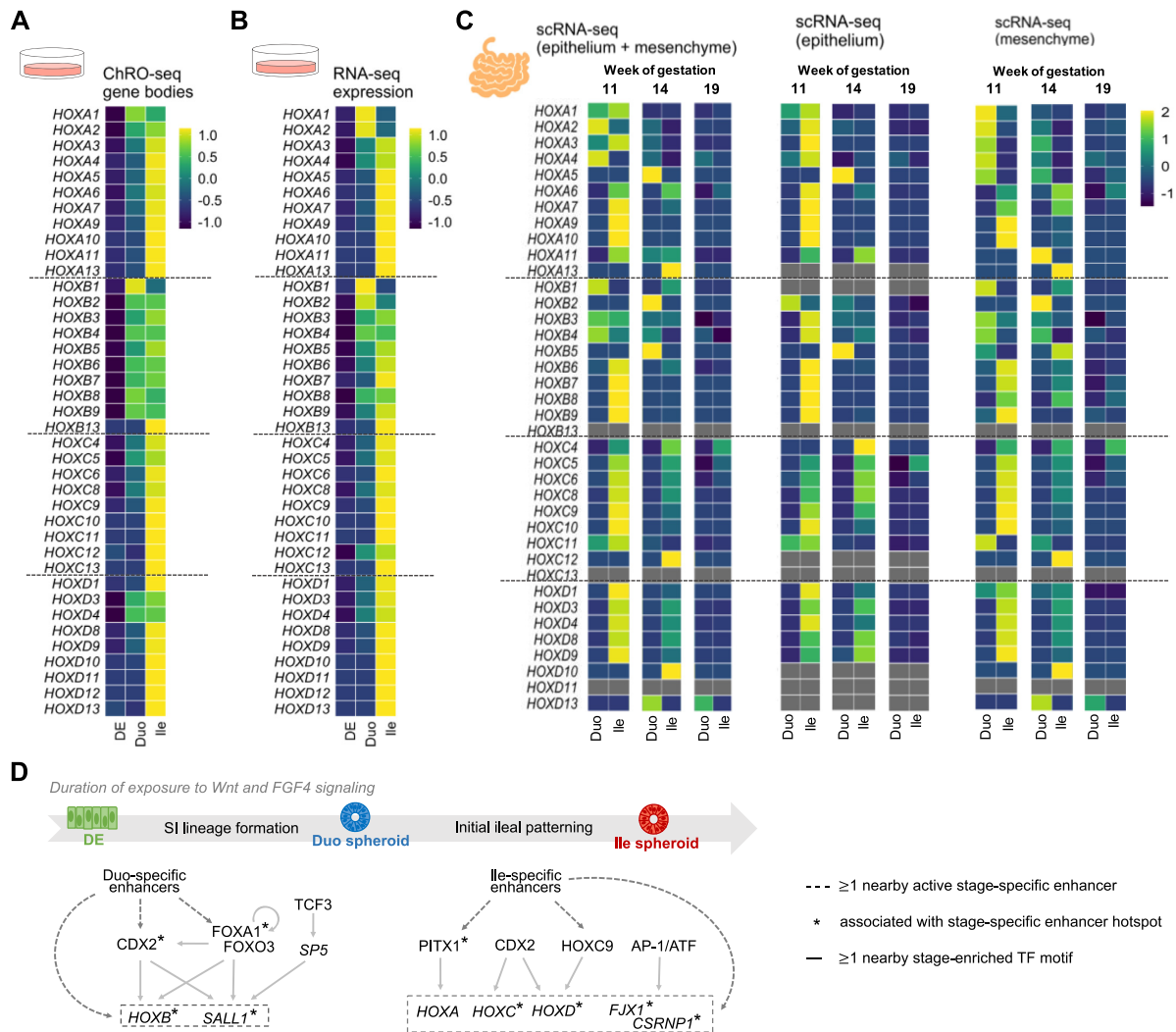


Figure 7. The HOX cluster dynamics observed in the directed differentiation model of SI is present in the early developmental stages of primary human fetal SI tissue. (A) Heatmap showing the changing transcriptional pattern of HOX genes (ChRO-seq) across the stages of DE, Duo spheroids and Ile spheroids *in vitro*. (B) Heatmap showing changing steady-state expression pattern of HOX genes (RNA-seq) across stages of DE, Duo spheroids and Ile spheroid *in vitro*. (C) Heatmap showing changing expression pattern of HOX genes (scRNA-seq) in primary human fetal Duo and Ile tissues at week 11, 14 and 19 of gestation. The expression level of each HOX gene is the average from the fraction that has both epithelial and mesenchymal cell types (left), mesenchymal cell only (middle), or epithelial cell only (right). (D) Proposed model of transcriptional programming in SI lineage formation and ileal regional patterning during human SI development. The genes activated in the relevant stages and the types of regulatory relationships identified through ChRO-seq analyses are indicated. ChRO-seq study: DE, $n = 4$; Duo spheroid (Duo), $n = 3$; Ile spheroid (Ile), $n = 3$. RNA-seq study: DE, $n = 3$; Duo, $n = 6$; Ile, $n = 4$. scRNA-seq data: matched Duo and Ile tissues from one human fetal subject per gestational time point.

time ever comprehensive chromatin regulatory landscapes across the different stages of directed differentiation from hPSC to SI spheroids with regional specification. Specifically, we defined: (i) early marker genes that label SI lineage formation and regional specification, (ii) the map of active regulatory elements (promoters and enhancers) as well as enhancer hotspots associated with SI lineage formation and ileal patterning and (iii) candidate key TF drivers and their putative cisromes relevant to these critical developmental events. However, in these analyses we also observed cases where genes near an emerging enhancer or enhancer hotspot do not exhibit changes in transcriptional activity and expression levels, which may indicate the presence of very long-range chromatin interactions beyond the

conservative distance criterion used in this study (100 kb window from TSSs). Recent seminal studies leveraging Hi-C technology observed that chromatin interactions are often in the range of 200 kb but could be up to 500 kb or more in the human genome (52,53). Nonetheless, the findings presented in our work motivate future experiments to investigate the functional role of candidate regulators and assess the sufficiency and necessity of the associated enhancers and enhancer hotspots for controlling transcriptional networks relevant to early human SI development and patterning.

We identified for the first time *HES4* as marker gene of SI spheroids irrespective of regional identity in the context of gut development; one likely reason for this is that it is

absent in the genome of the mouse (54), the model organism most used previously to study gut development. Moreover, the marker genes that we found to specifically label Duo identity include *SP5* and *SALL1*, the latter of which has been linked to Townes-Brocks syndrome with developmental malformation of multiple organs including the limb, kidney and gastrointestinal tract (55,56). While *SP5*, *SALL1* and other marker genes of Duo identity were reported to enhance Wnt signaling (57,58), we identified a distinct set of marker genes that specifically label Ile identity (e.g. *HAND2*, *FJX1* and *CSRNP1*) that are also known to be downstream of Wnt signaling. *Fjx1* has been shown to function through the Wnt/planar cell polarity pathway to determine anterior-posterior axis in *Drosophila* (59,60) and is expressed in the epithelium of murine developing gut (61). *Csrnp1* is expressed in neural crest progenitors in a Wnt1/ β -catenin-dependent manner and acts as a critical effector of Wnt signaling for driving neural crest formation (62). Together, these observations indicate that different sets of Wnt signaling effector genes are likely induced upon different durations of exposure to the Wnt-activating supplements during directed differentiation. Future investigation of the functional roles of these genes in the context of early gut development is warranted. The present study also uncovers the associated enhancers (or enhancer hotspots) and TF motifs of these marker genes, which may elucidate the potential mechanisms through which these genes are regulated during the establishment of SI identity and regional specification.

Although *HOX* genes have been reported to play important roles in proximal-distal axis patterning in the gut through a so-called ‘spatial collinearity’ mechanism, our observations in the directed differentiation model of human SI development suggest even greater complexity to *HOX* genes dynamics during small intestinal formation and regionalization. The spatial collinearity of *HOX* genes describes the correspondence between *HOX* paralogues along the chromosome (3' to 5') and their expression pattern along the anterior-posterior axis. Indeed, consistent with early literature in the mouse developing gut (63), our data shows higher expression of 3' *HOX* genes in Duo spheroids and higher expression of 5' *HOX* genes in Ile spheroids. Notably, though, our data additionally reveals that the formation of Duo and Ile spheroids are associated with the activation of distinct families of *HOX* genes. Specifically, we found that the activation of the *HOXB* cluster (especially *HOXB1-4*) first occurs during the formation of Duo spheroids and the other *HOX* clusters (*HOXA*, *C* and *D*) are activated subsequently during the formation of Ile spheroids. Through mining of a novel scRNA-seq dataset (13,43), we showed that this distinct *HOX* cluster behavior between Duo and Ile regions is present in primary human fetal SI tissue. Therefore, this distinct *HOX* cluster behavior we observed *in vitro* likely resembles SI regional specification *in vivo* in humans. Notably, a recent scRNA-seq study done in the digestive tract of E8.75 mouse embryos showed strong expression of *Hoxb1* and *Hoxb2* around the proximal midgut region (from where the duodenum is derived) in a pseudo anterior-posterior space (6). The same mouse study also indicated higher expression of several *Hoxc* and *Hoxd* genes in the distal digestive tract. Furthermore, the mouse model lacking

Hoxd3 was reported to have severe narrowing of the proximal colon and retention of gut luminal content in the terminal region of the SI (ileum), providing *in vivo* evidence for the role of a 3' *Hoxd* gene in regulating the development of the distal intestinal region. Using hPSC-derived SI spheroids, as well as in human primary fetal tissues, our study is able to show activation of distinct *HOX* clusters during early SI formation and regional specification. More importantly, through ChRO-seq analyses, we also provided additional insights into *HOX* biology by elucidating candidate TF drivers of specific *HOX* genes (and associated enhancers and enhancer hotspots), which merits detailed investigation.

Our ChRO-seq analysis mapped the key transcriptional networks associated with SI lineage formation or regional patterning. For example, we found that binding sites of FOXA1 and FOXO3 are enriched in active enhancers that emerge only during SI identity acquisition, including those nearby *CDX2*, which encodes the well-known master TF regulator of SI development. Notably, the *FOXA1* gene locus is actively up-transcribed during the transition from definitive endoderm to SI spheroids and is associated with an enhancer hotspot that emerges during this process. FOXA1 has been reported to regulate allocation of murine SI secretory lineage (64) and proper gene expression in murine intestinal epithelium (65). Recently, dysregulation of intestinal *FOXA1* also has been linked to necrotizing enterocolitis in infants (66). The observations made in this study as well as previous reports together underscores the functional importance of FOXA1 in human gut development. Also, our analyses suggest that CDX2 is involved in not only in initial SI lineage establishment but also the subsequent event of ileal specification, but evidently through distinct cisromes. One specific example is the association between the *HOXC9* locus and the nearby Ile-specific active enhancers containing CDX2 binding motifs, which is consistent with a previous study showing a remarkable down-regulation of *Hoxc9* in distal but not in proximal intestine of *Cdx2* null mice (2). Recent evidence also demonstrates that CDX2 regulates developing and fully mature gut through targeting distinct gene sets in mice (8) and that CDX2 is involved in patterning both intestinal epithelial and mesenchymal cells in the hPSC-based directed differentiation model (13). Our ChRO-seq findings pertaining to CDX2 may further elucidate how CDX2 functions through a stage- or cell type-dependent manner during early human gut development. Beyond CDX2, other candidate TFs associated with the establishment of ileal identity are worthy of future investigation. One notable example is PITX1, the ortholog of which has been reported previously in mouse models to be expressed in the distal SI region and to be involved in cecum budding during murine development (67). However, PITX1 has not been previously appreciated in the context of ileal regional specification.

The use of the state-of-art *in vitro* model of human SI development is a critical feature of this study given the lack of other existing platforms or resources that are readily available for the study of very early developmental time points pertaining to SI lineage formation and regional patterning. It is also important given the growing appreciation for cross-species differences in developmental processes at

the molecular level (68–71). 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 the transcriptional programming underlying early SI development in humans and motivates future functional studies of candidate TFs and enhancers in driving human SI development. Our 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.

DATA AVAILABILITY

Raw sequencing data of ChRO-seq and bulk RNA-seq are available through GEO accession GSE142316. Raw single cell RNA-seq data of human fetal duodenum is available through ArrayExpress (E-MTAB-9489) (13,43). Raw single cell RNA-seq data of human fetal ileum is available through ArrayExpress (MTAB-9906).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

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 Tim Dinh in the Sethupathy laboratory for help in developing the bioinformatic analysis pipeline of ChRO-seq data and providing technical consultation on bioinformatics, Dr Matt Kanke for providing additional 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.

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.; Resources and experiments of *Procr* reporter mice, Q.C.Y. and Y.A.Z.; Resources of scRNA-seq data, Q.Y., J.G.C. and

J.R.S.; 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.

FUNDING

American Diabetes Association [1-16-ACE-47 to P.S.]; Empire State Stem Cell Fund, funded by New York State Department of Health [C30293GG to Y.-H. H.]; Cornell Stem Cell Program Seed Grant, funded by New York State Department of Health [4353504 to P.S.]; Intestinal Stem Cell Consortium, a collaborative project funded by National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) and National Institute of Allergy and Infectious Diseases (NIAID) [U01DK103141 to J.R.S.]; Novel Alternative Model Systems for Enteric Diseases (NAMSED) Consortium, funded by NIAID [U19AI116482 to J.R.S.]; University of Michigan Center for Gastrointestinal Research (UMCGR) [NIDDK 5P30DK034933 to J.R.S.]; Chan Zuckerberg Initiative DAF, an advised fund of Silicon Valley Community Foundation [CZF2019-002440 to J.G.C.]; European Research Council [Anthropoid-803441 to J.G.C.]. Funding for open access charge: American Diabetes Association [1-16-ACE-47].

Conflict of interest statement. None declared.

REFERENCES

- Sheaffer, K.L. and Kaestner, K.H. (2012) Transcriptional networks in liver and intestinal development. *Cold Spring Harb. Perspect. Biol.*, **4**, a008284.
- Gao, N., White, P. and Kaestner, K.H. (2009) Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. *Dev. Cell*, **16**, 588–599.
- Thompson, C.A., Wojta, K., Pulakanti, K., Rao, S., Dawson, P. and Battle, M.A. (2017) GATA4 is sufficient to establish jejunal versus ileal identity in the small intestine. *Cell Mol. Gastroenterol. Hepatol.*, **3**, 422–446.
- Thompson, C.A., DeLaForest, A. and Battle, M.A. (2018) Patterning the gastrointestinal epithelium to confer regional-specific functions. *Dev. Biol.*, **435**, 97–108.
- Verzi, M.P., Hatzis, P., Sulahian, R., Philips, J., Schuijers, J., Shin, H., Freed, E., Lynch, J.P., Dang, D.T., Brown, M. *et al.* (2010) TCF4 and CDX2, major transcription factors for intestinal function, converge on the same cis-regulatory regions. *Proc. Natl. Acad. Sci. U.S.A.*, **107**, 15157–15162.
- Nowotschin, S., Setty, M., Kuo, Y.Y., Liu, V., Garg, V., Sharma, R., Simon, C.S., Saiz, N., Gardner, R., Boutet, S.C. *et al.* (2019) The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature*, **569**, 361–367.
- Gao, S., Yan, L., Wang, R., Li, J., Yong, J., Zhou, X., Wei, Y., Wu, X., Wang, X., Fan, X. *et al.* (2018) Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing. *Nat. Cell Biol.*, **20**, 721–734.
- Kumar, N., Tsai, Y.H., Chen, L., Zhou, A., Banerjee, K.K., Saxena, M., Huang, S., Toke, N.H., Xing, J., Shivdasani, R.A. *et al.* (2019) The lineage-specific transcription factor CDX2 navigates dynamic chromatin to control distinct stages of intestine development. *Development*, **146**, dev172189.
- Dong, J., Hu, Y., Fan, X., Wu, X., Mao, Y., Hu, B., Guo, H., Wen, L. and Tang, F. (2018) Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis. *Genome Biol.*, **19**, 31.
- Banerjee, K.K., Saxena, M., Kumar, N., Chen, L., Cavazza, A., Toke, N.H., O'Neill, N.K., Madha, S., Jadhav, U., Verzi, M.P. *et al.* (2018) Enhancer, transcriptional, and cell fate plasticity precedes intestinal determination during endoderm development. *Genes Dev.*, **32**, 1430–1442.

11. Chen, L., Toke, N.H., Luo, S., Vasoya, R.P., Fullem, R.L., Parthasarathy, A., Perekatt, A.O. and Verzi, M.P. (2019) A reinforcing HNF4-SMAD4 feed-forward module stabilizes enterocyte identity. *Nat. Genet.*, **51**, 777–785.
12. Elmentaite, R., Ross, A.D.B., Roberts, K., James, K.R., Ortmann, D., Gomes, T., Nayak, K., Tuck, L., Pritchard, S., Bayraktar, O.A. *et al.* (2020) Single-Cell Sequencing of Developing Human Gut Reveals Transcriptional Links to Childhood Crohn's Disease. *Dev Cell*, **S1534-5807**, 30886–30888.
13. Holloway, E.M., Czerwinski, M., Tsai, Y.H., Wu, J.H., Wu, A., Childs, C.J., Walton, K.D., Sweet, C.W., Yu, Q., Glass, I. *et al.* (2020) Mapping Development of the Human Intestinal Niche at Single-Cell Resolution. *Cell Stem Cell*, **S1934-5909**, 30546–30544.
14. Nord, A.S., Blow, M.J., Attanasio, C., Akiyama, J.A., Holt, A., Hosseini, R., Phouanavong, S., Plajzer-Frick, I., Shoukry, M., Afzal, V. *et al.* (2013) Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. *Cell*, **155**, 1521–1531.
15. Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D. *et al.* (2013) Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell*, **153**, 1134–1148.
16. Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W. *et al.* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature*, **459**, 108–112.
17. Long, H.K., Prescott, S.L. and Wysocka, J. (2016) Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell*, **167**, 1170–1187.
18. Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S. *et al.* (2010) Widespread transcription at neuronal activity-regulated enhancers. *Nature*, **465**, 182–187.
19. Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, A.L., Heidemann, M., Hintermair, C. *et al.* (2011) Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat. Struct. Mol. Biol.*, **18**, 956–963.
20. Chu, T., Rice, E.J., Booth, G.T., Salamanca, H.H., Wang, Z., Core, L.J., Longo, S.L., Corona, R.J., Chin, L.S., Lis, J.T. *et al.* (2018) Chromatin run-on and sequencing maps the transcriptional regulatory landscape of glioblastoma multiforme. *Nat. Genet.*, **50**, 1553–1564.
21. Core, L.J., Waterfall, J.J. and Lis, J.T. (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science*, **322**, 1845–1848.
22. Kwak, H., Fuda, N.J., Core, L.J. and Lis, J.T. (2013) Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science*, **339**, 950–953.
23. Dinh, T.A., Sritharan, R., Smith, F.D., Francisco, A.B., Ma, R.K., Bunaciu, R.P., Kanke, M., Danko, C.G., Massa, A.P., Scott, J.D. *et al.* (2020) Hotspots of aberrant enhancer activity in fibrolamellar carcinoma reveal candidate oncogenic pathways and therapeutic vulnerabilities. *Cell Rep.*, **31**, 107509.
24. Wells, J.M. and Spence, J.R. (2014) How to make an intestine. *Development*, **141**, 752–760.
25. Sinagoga, K.L. and Wells, J.M. (2015) Generating human intestinal tissues from pluripotent stem cells to study development and disease. *EMBO J.*, **34**, 1149–1163.
26. Spence, J.R., Mayhew, C.N., Rankin, S.A., Kuhar, M.F., Vallance, J.E., Tolle, K., Hoskins, E.E., Kalinichenko, V.V., Wells, S.I., Zorn, A.M. *et al.* (2011) Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. *Nature*, **470**, 105–109.
27. McCracken, K.W., Howell, J.C., Wells, J.M. and Spence, J.R. (2011) Generating human intestinal tissue from pluripotent stem cells in vitro. *Nat. Protoc.*, **6**, 1920–1928.
28. Tsai, Y.H., Nattiv, R., Dedhia, P.H., Nagy, M.S., Chin, A.M., Thomson, M., Klein, O.D. and Spence, J.R. (2017) In vitro patterning of pluripotent stem cell-derived intestine recapitulates in vivo human development. *Development*, **144**, 1045–1055.
29. Watson, C.L., Mahe, M.M., Munera, J., Howell, J.C., Sundaram, N., Poling, H.M., Schweitzer, J.I., Vallance, J.E., Mayhew, C.N., Sun, Y. *et al.* (2014) An in vivo model of human small intestine using pluripotent stem cells. *Nat. Med.*, **20**, 1310–1314.
30. Finkbeiner, S.R., Freeman, J.J., Wieck, M.M., El-Nachef, W., Althaim, C.H., Tsai, Y.H., Huang, S., Dyal, R., White, E.S., Grikscheit, T.C. *et al.* (2015) Generation of tissue-engineered small intestine using embryonic stem cell-derived human intestinal organoids. *Biol. Open*, **4**, 1462–1472.
31. Dame, M.K., Attili, D., McClintock, S.D., Dedhia, P.H., Ouillette, P., Hardt, O., Chin, A.M., Xue, X., Laliberte, J., Katz, E.L. *et al.* (2018) Identification, isolation and characterization of human LGR5-positive colon adenoma cells. *Development*, **145**, dev153049.
32. Wang, D., Wang, J., Bai, L., Pan, H., Feng, H., Clevers, H. and Zeng, Y.A. (2020) Long-term expansion of pancreatic islet organoids from resident Procr(+) progenitors. *Cell*, **180**, 1198–1211.
33. Wuari, J. and Schibler, U. (1994) Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol. Cell Biol.*, **14**, 7219–7225.
34. Mahat, D.B., Kwak, H., Booth, G.T., Jonkers, I.H., Danko, C.G., Patel, R.K., Waters, C.T., Munson, K., Core, L.J. and Lis, J.T. (2016) Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc.*, **11**, 1455–1476.
35. Chu, T., Wang, Z., Chou, S.P. and Danko, C.G. (2019) Discovering transcriptional regulatory elements from run-on and sequencing data using the web-based dREG gateway. *Curr. Protoc. Bioinform.*, **66**, e70.
36. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M. and Gingeras, T.R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, **29**, 15–21.
37. Patro, R., Duggal, G., Love, M.I., Irizarry, R.A. and Kingsford, C. (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods*, **14**, 417–419.
38. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, 550.
39. Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A. *et al.* (2016) Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.*, **44**, W90–W97.
40. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H. and Glass, C.K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell*, **38**, 576–589.
41. Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I. and Young, R.A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, **153**, 307–319.
42. Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre, V., Sigova, A.A., Hoke, H.A. and Young, R.A. (2013) Super-enhancers in the control of cell identity and disease. *Cell*, **155**, 934–947.
43. Holloway, E.M., Wu, J.H., Czerwinski, M., Sweet, C.W., Wu, A., Tsai, Y.H., Huang, S., Stoddard, A.E., Capeling, M.M., Glass, I. *et al.* (2020) Differentiation of human intestinal organoids with endogenous vascular endothelial cells. *Dev. Cell*, **54**, 516–528.
44. 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.
45. Hill, D.R., Huang, S., Nagy, M.S., Yadagiri, V.K., Fields, C., Mukherjee, D., Bons, B., Dedhia, P.H., Chin, A.M., Tsai, Y.H. *et al.* (2017) Bacterial colonization stimulates a complex physiological response in the immature human intestinal epithelium. *Elife*, **6**, e29132.
46. Finkbeiner, S.R., Hill, D.R., Althaim, C.H., Dedhia, P.H., Taylor, M.J., Tsai, Y.H., Chin, A.M., Mahe, M.M., Watson, C.L., Freeman, J.J. *et al.* (2015) Transcriptome-wide analysis reveals hallmarks of human intestine development and maturation in vitro and in vivo. *Stem Cell Rep.*, **4**, 1140–1155.
47. Suh, M.R., Lee, Y., Kim, J.Y., Kim, S.K., Moon, S.H., Lee, J.Y., Cha, K.Y., Chung, H.M., Yoon, H.S., Moon, S.Y. *et al.* (2004) Human embryonic stem cells express a unique set of microRNAs. *Dev. Biol.*, **270**, 488–498.
48. Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T. *et al.* (2014) An atlas of active enhancers across human cell types and tissues. *Nature*, **507**, 455–461.

49. Oz-Levi, D., Olender, T., Bar-Joseph, I., Zhu, Y., Marek-Yagel, D., Barozzi, I., Osterwalder, M., Alkelai, A., Ruzzo, E.K., Han, Y. *et al.* (2019) Noncoding deletions reveal a gene that is critical for intestinal function. *Nature*, **571**, 107–111.
50. Boyd, M., Thodberg, M., Vitezic, M., Bornholdt, J., Vitting-Seerup, K., Chen, Y., Coskun, M., Li, Y., Lo, B.Z.S., Klausen, P. *et al.* (2018) Characterization of the enhancer and promoter landscape of inflammatory bowel disease from human colon biopsies. *Nat. Commun.*, **9**, 1661.
51. Barros, R., da Costa, L.T., Pinto-de-Sousa, J., Duluc, I., Freund, J.N., David, L. and Almeida, R. (2011) CDX2 autoregulation in human intestinal metaplasia of the stomach: impact on the stability of the phenotype. *Gut*, **60**, 290–298.
52. Jung, I., Schmitt, A.D., Diao, Y., Lee, A.J., Liu, T., Yang, D., Tan, C., Eom, J., Chan, M., Chee, S. *et al.* (2019) A compendium of promoter-centered long-range chromatin interactions in the human genome. *Nat. Genet.*, **51**, 1442–1449.
53. Schmitt, A.D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C.L., Li, Y., Lin, S., Lin, Y., Barr, C.L. *et al.* (2016) A compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep.*, **17**, 2042–2059.
54. Kageyama, R., Ohtsuka, T. and Kobayashi, T. (2007) The Hes gene family: repressors and oscillators that orchestrate embryogenesis. *Development*, **134**, 1243–1251.
55. Kohlhasse, J., Wischermann, A., Reichenbach, H., Froster, U. and Engel, W. (1998) Mutations in the SALL1 putative transcription factor gene cause Townes-Brocks syndrome. *Nat. Genet.*, **18**, 81–83.
56. Kiefer, S.M., Ohlemiller, K.K., Yang, J., McDill, B.W., Kohlhasse, J. and Rauchman, M. (2003) Expression of a truncated Sall1 transcriptional repressor is responsible for Townes-Brocks syndrome birth defects. *Hum. Mol. Genet.*, **12**, 2221–2227.
57. Kennedy, M.W., Chalamalasetty, R.B., Thomas, S., Garriock, R.J., Jailwala, P. and Yamaguchi, T.P. (2016) Sp5 and Sp8 recruit beta-catenin and Tcf1-Lef1 to select enhancers to activate Wnt target gene transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 3545–3550.
58. Sato, A., Kishida, S., Tanaka, T., Kikuchi, A., Kodama, T., Asashima, M. and Nishinakamura, R. (2004) Sall1, a causative gene for Townes-Brocks syndrome, enhances the canonical Wnt signaling by localizing to heterochromatin. *Biochem. Biophys. Res. Commun.*, **319**, 103–113.
59. Villano, J.L. and Katz, F.N. (1995) four-jointed is required for intermediate growth in the proximal-distal axis in *Drosophila*. *Development*, **121**, 2767–2777.
60. Zeidler, M.P., Perrimon, N. and Strutt, D.I. (1999) The four-jointed gene is required in the *Drosophila* eye for ommatidial polarity specification. *Curr. Biol.*, **9**, 1363–1372.
61. Rock, R., Heinrich, A.C., Schumacher, N. and Gessler, M. (2005) Fjx1: a notch-inducible secreted ligand with specific binding sites in developing mouse embryos and adult brain. *Dev. Dyn.*, **234**, 602–612.
62. Simoes-Costa, M., Stone, M. and Bronner, M.E. (2015) Axud1 integrates Wnt signaling and transcriptional inputs to drive neural crest formation. *Dev. Cell*, **34**, 544–554.
63. Kawazoe, Y., Sekimoto, T., Araki, M., Takagi, K., Araki, K. and Yamamura, K. (2002) Region-specific gastrointestinal Hox code during murine embryonal gut development. *Dev. Growth Differ.*, **44**, 77–84.
64. Ye, D.Z. and Kaestner, K.H. (2009) Foxa1 and Foxa2 control the differentiation of goblet and enteroendocrine L- and D-cells in mice. *Gastroenterology*, **137**, 2052–2062.
65. Kerschner, J.L., Gosalia, N., Leir, S.H. and Harris, A. (2014) Chromatin remodeling mediated by the FOXA1/A2 transcription factors activates CFTR expression in intestinal epithelial cells. *Epigenetics*, **9**, 557–565.
66. Wu, Y.Z., Chan, K.Y.Y., Leung, K.T., Lam, H.S., Tam, Y.H., Lee, K.H., Li, K. and Ng, P.C. (2019) Dysregulation of miR-431 and target gene FOXA1 in intestinal tissues of infants with necrotizing enterocolitis. *FASEB J.*, **33**, 5143–5152.
67. Matsuyama, M., Aizawa, S. and Shimono, A. (2009) Sfrp controls apicobasal polarity and oriented cell division in developing gut epithelium. *PLoS Genet.*, **5**, e1000427.
68. Allison, T.F., Smith, A.J.H., Anastasiadis, 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 Rep.*, **10**, 1895–1907.
69. Chu, L.F., Leng, N., Zhang, J., Hou, Z., Mamott, D., Vereide, D.T., Choi, J., Kendziorski, C., Stewart, R. and Thomson, J.A. (2016) Single-cell RNA-seq reveals novel regulators of human embryonic stem cell differentiation to definitive endoderm. *Genome Biol.*, **17**, 173.
70. Tiyaboonchai, A., Cardenas-Diaz, F.L., Ying, L., Maguire, J.A., Sim, X., Jobaliya, C., Gagne, A.L., Kishore, S., Stanescu, D.E., Hughes, N. *et al.* (2017) GATA6 plays an important role in the induction of human definitive endoderm, development of the pancreas, and functionality of pancreatic beta cells. *Stem Cell Rep.*, **8**, 589–604.
71. Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y., Liechti, A., Ascencio, K., Rummel, C., Ovchinnikova, S. *et al.* (2019) Gene expression across mammalian organ development. *Nature*, **571**, 505–509.