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

Authors: Yu-Han Hung[1], Sha Huang[2,3], Michael K. Dame[3], Qianhui Yu[4], Qing Cissy Yu[5], Yi Arial Zeng[5], J. Gray Camp[4, 6, 7], Jason R. Spence[2,3], Praveen Sethupathy[1]

[1]Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY, USA

[2]Department of Cell and Developmental Biology, University of Michigan, Ann Arbor, MI, USA

[3]Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, MI, USA

[4]Institute of Molecular and Clinical Ophthalmology, Basel, Switzerland

[5]State 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

[6]University of Basel, Basel, Switzerland

[7]Department of Ophthalmology, University of Basel, Switzerland

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

Abstract

Background: The establishment of the small intestinal (SI) lineage during human embryogenesis is critical for the proper development of neonatal gut functions, including nutrient absorption and immune defense. The chromatin dynamics and regulatory networks that drive human SI lineage formation and regional patterning 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, super enhancers, and gene bodies across distinct stages of directed differentiation of human pluripotent stem cells (hPSCs) into SI spheroids with regional specification. **Results:** Through comprehensive ChRO-seq analysis we identify candidate stage-specific chromatin activity states, novel markers, and enhancer hotspots during the directed differentiation process. Moreover, we propose a detailed transcriptional network associated with SI lineage formation or initial regional patterning. Among our findings is a unique pattern of enhancer activity and transcription at HOX gene loci that is previously undescribed. Analysis of single cell RNA-seq data from human fetal SI at early developmental time points shed further light on the unique HOX gene temporal dynamics that underlies SI regional patterning. **Conclusions:** Overall, the results lead to a new proposed working model for the regulatory underpinnings of human SI lineage formation and regional patterning, thereby adding a novel dimension to the literature that has thus far relied almost exclusively on non-human models.

Key words: small intestine; human development; chromatin run-on sequencing (ChRO-seq); enhancer; transcriptional regulatory elements; directed differentiation; HOX genes

52 **Introduction**

53

54 The embryonic development of the small intestine (SI) is critical for a fetus to thrive and grow. The genetic
55 programming of SI cell identity and regional specification during early development is fundamental to
56 ensure proper SI morphogenesis and maturation with specialized functions, including nutrient digestion and
57 absorption, energy balance, and pathogen defense. Several seminal studies have identified important
58 molecular regulators associated with gut development [1-5], and more recent studies have leveraged
59 advanced genomic technologies (e.g., single cell RNA-seq and bulk ATAC-seq) to provide insights at a
60 more granular level into gut development [6-11]; however, this work relies almost exclusively on animal
61 models. Studies of human SI development have been few [7, 12, 13], due in large part to limited access to
62 primary human fetal tissues. Initial human SI lineage formation from the endodermal germ layer, which
63 occurs much prior to the developmental time points at which human fetal tissues are generally available, is
64 essentially uncharacterized. In this study, we sought to fill this important knowledge gap by profiling for
65 the first time the chromatin regulatory dynamics and transcriptional programs that are associated with
66 human SI lineage formation and initial regional patterning.

67

68 Robust temporal and spatial regulation of gene expression is fundamental to all biological processes
69 including development [14, 15]. Transcriptional programs are precisely controlled by promoters and distal
70 *cis*-regulatory regions known as enhancers. Enhancers harbor binding sites for transcription factors (TFs),
71 activate long-range gene transcription, and are often cell-type specific [16, 17]. RNA polymerases are
72 known to be recruited to active enhancers, generating divergent short transcripts (also known as enhancer
73 RNA or eRNAs) [18, 19]. Recently, an approach called chromatin run-on sequencing (ChRO-seq) [20]
74 was developed for genome-wide identification of promoters, active enhancers, and actively transcribed
75 gene bodies in a single assay. ChRO-seq represents the newest generation of nascent RNA sequencing
76 technologies, and overcomes several limitations of previous versions including global run-on sequencing

77 (GRO-seq) [21] and precision run-on sequencing (PRO-seq) [22]. ChRO-seq was very recently
78 successfully applied to archived solid tumor tissues [20, 23].
79
80 Advances in the directed differentiation of human pluripotent stem cells (hPSCs), including human
81 embryonic stem cells (hESCs), provide a powerful strategy for studying the early developmental events of
82 human SI that would be nearly impossible otherwise [24, 25]. The multi-step process of generating human
83 "gut-in-a-dish" from hPSCs recapitulates many aspects of *in vivo* SI development [26-28], including
84 induction of the endodermal fate, formation of SI lineage (SI spheroids) with regional identities, and
85 maturation into 3-dimensional human intestinal organoids (HIOs) that exhibit molecular, structural, and
86 functional features similar to those of the human fetal SI [13, 29]. While the protocol for directed
87 differentiation of hPSCs into SI spheroids has been established, the molecular mechanisms governing this
88 process remains minimally characterized. To reveal temporal dynamics of the chromatin state during early
89 human SI development and regional patterning, we performed ChRO-seq across the following stages: hPSC,
90 definitive endoderm (DE), duodenal (proximal SI) spheroid, and ileal (distal SI) spheroid. This study
91 provides the first-ever view of the changing chromatin regulatory landscape, defines stage-specific
92 enhancer hotspots, and identifies key TF networks that underlie the acquisition of SI identity and the
93 initiation of ileal regional patterning. Moreover, we uncover previously undescribed dynamics at HOX
94 gene loci associated with these events. Additional analysis of single cell RNA-seq (scRNA-seq) of human
95 fetal SI tissue underscores the unique HOX gene patterns, although these results are subject to the caveat
96 that these tissues reflect developmental time points much later than what we have the opportunity to analyze
97 in the directed differentiation model. Overall, this study offers an unprecedented resource that serves as a
98 springboard from which the research community can develop and test targeted hypotheses about key
99 regulatory hotspots and molecular drivers of early events of human SI development.
100
101 **Results**
102

103 **ChRO-seq reveals temporal dynamics of nascent transcription at gene loci in the directed**
104 **differentiation model of human developing SI**
105
106 The directed differentiation of hPSCs (here we used H9 hESCs) into SI spheroids with regional
107 specification was carried out as previously described [26, 28] **(Figure 1A)**. The patterned SI spheroids
108 represent the primitive form of the SI comprised of mainly stem/progenitor cells [30], which can give rise
109 to different mature epithelial cell types after prolonged organoid culture, or following transplantation into
110 a murine host [31]. In this study, we included the four stages of this model: hESC, DE, duodenum (Duo)
111 spheroid, and ileal (Ile) spheroid **(Figure 1A)**. We carried out ChRO-seq in order to characterize the
112 dynamics of genome-wide nascent transcriptional activity toward the goal of defining the changing
113 landscape of promoters, enhancers, and active gene bodies across the following sequential events: DE fate
114 induction (from hESC to DE), SI lineage formation (from DE to Duo spheroid), and initial ileal regional
115 patterning (from Duo to Ile spheroid) **(Figure 1A)**.
116
117 To assess nascent transcription of genes, we first analyzed ChRO-seq signal within the body of annotated
118 genes. ChRO-seq data in the first 500 bp downstream of the transcription start site, or TSS, was excluded
119 to remove signal contributed by RNA polymerase pausing. Hierarchical clustering analysis of the gene
120 transcriptional profiles demonstrates clean stratification of the different stages **(Figure 1B)**. The signal at
121 gene bodies encoding TFs that are well-established markers of particular stages (*SOX2*, *GATA6*, and *CDX2*)
122 indicate the expected specificity **(Figure 1C-D)**. Also, the very minimal expression of the mesenchymal
123 marker *FOXF1* **(Figure 1C-D),** as well as markers of other organ lineages **(Supplementary Figure 3B),**
124 demonstrate that the Duo and Ile spheroids are indeed dominated by early-stage SI epithelial cells. Next
125 we sought to identify from the ChRO-seq data differentially transcribed genes during each stage transition
126 of this model ($\log_2$ fold change > 1, average TPM > 25, padj < 0.2, $p < 0.05$ by Wald test; DESeq2) **(Figure**
127 **1E)**. The analyses identified 1886 differentially transcribed genes (1328 up; 558 down) during DE
128 formation (DE vs. hESC), 4024 differentially transcribed genes (2546 up; 1478 down) during SI lineage

129 formation (Duo spheroids vs. DE) and 362 differentially transcribed genes (182 up; 180 down) during ileal
130 patterning event (Ile vs. Duo spheroids) **(Figure 1E)**. We also performed RNA-seq on the same stages to
131 define differentially expressed genes and the Gene Ontology (GO) term enrichment analysis between
132 differentially transcribed and differentially expressed genes are in general agreement **(Supplementary**
133 **Figure 2)**.
134
135 The same ChRO-seq data also affords the unique opportunity to evaluate promoter activity. Therefore, we
136 next developed an approach to analyze promoter dynamics across all of the stages **(Supplementary Figure**
137 **1C)**. We identified 9550 genes, the majority of which are protein-coding genes (n = 7651)**,** which exhibit
138 significantly changing patterns of promoter activity across different stages (padj < 0.05 by likelihood ratio
139 test; DESeq2) **(Supplementary Figure 1C)**. Notably, a subset of these protein-coding genes has low
140 promoter activity in both hESC and DE but high promoter activity in both Duo and Ile spheroids (n=2992).
141 These include *BEX5* and *PROCR* **(Figure 1F)**, which were recently observed to have robust expression in
142 primary human fetal SI at early developmental time points [7, 12]. Analysis of the matched RNA-seq data,
143 which showed a dramatic rise in *BEX5* and *PROCR* mRNA levels during formation of SI spheroids from
144 DE **(Figure 1G)**, confirm the findings from the promoter analysis. We also validated the upregulation of
145 *PROCR* in SI spheroids by RT-qPCR with an independent batch of samples **(Figure 1H)** and demonstrated
146 by *in situ* hybridization staining a robust expression of *Procr* in the midgut region of *Procr*-mGFP-2A-
147 LacZ mice [32] at E9.5 and E10.5 (**Figure 2I**), indicating the conserved early presence of *PROCR* in the
148 developing SI. Overall, these findings demonstrate that hPSC-derived SI spheroids recapitulate features of
149 early developing SI and therefore represent the state-of-the-art model for further study of human SI lineage
150 formation.
151
152 **Identification of transcriptional markers of specific stages in the directed differentiation model of**
153 **human developing SI**
154

155 ChRO-seq signal at gene bodies (TPM > 50 in at least one stage) reveals the changing patterns of
156 transcription across stages **(Figure 2A)**. We sought to identify transcriptional markers of each stage in
157 this model. It is important to note that previously validated SI regional markers using human fetal
158 duodenum and ileum ranging 14-19 weeks of gestation [28] are not sufficient to distinguish Duo from Ile
159 spheroids, although they are sufficient to distinguish Duo from Ile HIOs (derived after culturing the
160 spheroids in Matrigel for 28 days) **(Supplementary Figure 3A)**. This likely reflects the fact that
161 spheroids, which are newly differentiated, represent the early gut lineage whereas the HIOs represent a
162 more mature state. Here, by leveraging ChRO-seq signal at gene bodies as well as RNA-seq data, we
163 were able to define gene sets that serve as early markers of SI lineage formation and ileal regional
164 patterning at the levels of both transcription and steady-state expression.
165
166 We first demonstrated that ChRO-seq and RNA-seq signal for genes are reasonably well-correlated
167 (Pearson correlation coefficient = 0.73) **(Figure 2B; Supplementary Figure 3B-E)**, indicating that nascent
168 transcription profiles globally reflect steady state expression profiles in this model system (albeit not
169 perfectly, as expected due to other layers of gene regulation such as those at the post-transcriptional level).
170 To identify genes that label a specific stage at the levels of both transcription and steady-state expression,
171 we developed a bioinformatic pipeline to integrate ChRO-seq and RNA-seq data (Methods) and performed
172 this integrative analysis with genes that are actively transcribed (TPM > 50) in at least one stage. We
173 identified genes that are significantly elevated according to both transcription and steady-state expression
174 in each stage relative to all other stages: hESC (n=239), DE (n=149), SI spheroid irrespective of regional
175 identity (n=88), Duo spheroid (n=9), and Ile spheroid (n=40) **(Figure 2C; Supplementary Data 2)**. The
176 stage-specific genes we identified include well-established markers such as *POU5F1* and *SOX2* for hESCs
177 **(Figure 2D)**; *NODAL*, *SOX17*, *EOMES* and *LEFTY2* for DE **(Figure 2E)**; and *CDX2* for SI spheroids
178 irrespective of regional identity **(Figure 2F)**. We defined additional-stage-specific markers, many of which
179 are previously undescribed: *HES4*, *FOXA1*, *WFDC2*, and several *HOXB* family members for SI spheroids
180 irrespective of regional identity; *SP5* and *SALL1* for Duo spheroids; and *FJX1*, *IGF2*, as well as several

181 HOXA/C/D members for Ile spheroids **(Figure 2F)**. We also identified several long, non-coding RNA
182 (lncRNA) and anti-sense transcript markers: *LINC00543* for SI spheroids irrespective of regional identity,
183 and *EVX1-AS* for Ile spheroids **(Figure 2C-F)**. Notably, while many genes were defined as markers of SI
184 spheroids irrespective of regional identity, the analysis identified much fewer genes as Duo-specific
185 markers compared than Ile-specific markers **(Figure 2C)**, which may suggest that Ile regional specification
186 is mainly driven by gain of additional driver genes rather than loss of genes contributing to Duo regional
187 identity.
188
189 **Transcriptional regulatory element profiles reveal notable re-wiring of chromatin activity during**
190 **directed differentiation**
191
192 Active transcriptional regulatory elements (TREs), including promoters and enhancers, are identified in
193 ChRO-seq data by the hallmark feature of short bi-directional transcription **(Figure 3A)**. To identify active
194 TREs across the entire genome, we employed dREG [33], which was developed specifically for this purpose.
195 Using this tool, we identified a total of 125,863 active TREs across all four stages included in this study
196 **(Figure 3B)**. The length distribution of these active TREs is consistent with what has been reported
197 previously [34] and the vast majority of the active TREs, as expected, are located in intergenic regions,
198 introns, and annotated TSSs **(Figure 3C-D)**. To specifically evaluate whether the TRE landscape of SI
199 spheroids is more similar to developing as opposed to mature SI in humans, we leveraged data on DNase-
200 accessible regions in both fetal and adult human SI available from the Roadmap Epigenomics Project. We
201 first defined DNase-accessible regions unique to fetal or adult tissue (Methods) and then intersected those
202 with active TREs present in Duo and/or Ile spheroids. We found that 12,541 out of 37, 892 (~33%) open
203 chromatin loci specific to fetal human SI overlap with TREs present in Duo and/or Ile spheroids, whereas
204 only 7,504 out of 45,242 (~16%) open chromatin loci specific to adult human SI overlap with TREs present
205 in Duo and/or Ile spheroids. To provide a specific example of a TRE defined in this model that captures
206 an important regulatory region associated with human fetal SI, we turned to the very recently identified

207 intestinal critical region (ICR). The ICR represents an important regulatory element of the new gene
208 *PERCC1*, which exhibits transient expression in the SI during development and has been linked to
209 congenital diarrheal disorders in humans [35]. We identified a TRE that is highly overlapping with the ICR
210 and active only in SI spheroids, especially in Duo spheroids **(Figure 3F-G)**. Taken together, these
211 observations demonstrate that TREs in hPSC-derived SI spheroids are able to capture important features of
212 gene regulatory landscapes that are associated with the developing human SI.
213
214 We next categorized the identified active TREs into 49,855 proximal and 76,008 distal TREs, which from
215 here on in we refer to as promoters and enhancers, respectively **(Figure 3A)**. Unsupervised hierarchical
216 clustering analyses reveal that enhancer profiles stratify different stages more accurately and clearly than
217 promoters **(Figure 3H-I)**, consistent with the notion that enhancer signature is the most cell-type specific
218 [14, 16]. To further elucidate the dynamics of active TRE profiles across the directed differentiation process,
219 we identified active TREs that are unchanged, gained (enhanced), or lost (suppressed) during the transition
220 to DE **(Figure 3J-K)**, SI lineage formation **(Figure 3L-M),** and ileal patterning **(Figure 3N-O)** (log2FC >
221 2.5, padj < 0.05 by Wald test; DESeq2). We found that a much greater proportion of enhancers, relative to
222 promoters, exhibit significantly altered levels of activity (either gain or loss) across all stage transitions. In
223 contrast, enhancers and promoters were roughly equally represented among those TREs that were
224 unchanged in activity across stage transitions **(Figure 3K, M, O)**. This finding underscores the importance
225 of enhancer dynamics during SI fate acquisition.
226
227 **Identification of genes associated with high a density of nearby enhancers that emerge during SI**
228 **lineage formation or ileal patterning**
229
230 Next we developed an analysis pipeline to further investigate the enhancers and associated genes that
231 emerge during different stages of the directed differentiation **(Figure 4A)**. We focused on 'Duo-specific
232 enhancers' and 'Ile-specific enhancers', which are the enhancers that gain activity during Duo and Ile

233 spheroid formation, respectively (**Figure 3M, O**). We then determined the density of stage-specific
234 enhancers for every actively transcribed gene (TPM > 50 in the stage of interest) by counting stage-specific
235 enhancers within a 200 kb window centered on the TSS **(Figure 4A)**. The genes that are associated with
236 the stage-specific enhancers and also have significant activation in both ChRO-seq and RNA-seq are
237 eventually defined as stage-specific enhancer linked genes **(Figure 4A)**.
238
239 As defined in **Figure 3**, we identified a total of 2,787 Duo-specific and 328 Ile-specific enhancers **(Figure**
240 **4B-C)**. We confirmed that genes associated with a greater number of Duo-specific enhancers also exhibit
241 greater increases in transcription in Duo spheroids relative to DE **(Figure 4D-E)**. Similarly, genes
242 associated with a greater number of Ile-specific enhancers also exhibit greater increases in transcription in
243 Ile relative to Duo spheroids **(Figure 4I-J)**. The results of similar analyses focused on hESC and DE are
244 shown in **Supplementary Figure 4**. These observations support the model in which the level of
245 transcriptional activation during each stage transition is strongly associated with the total number of
246 emerging TREs around TSSs.
247
248 Regarding the transition from DE to Duo spheroids, we identified 117 genes that are associated with at least
249 one Duo-specific enhancer and are significantly increased in both transcription (ChRO-seq) and steady-
250 state expression (RNA-seq) **(Figure 4F)**. Among these, as expected, *CDX2* is one of the top genes ranked
251 by the density of Duo-specific enhancers **(Figure 4G)**. Notably, even more highly ranked are the HOXB
252 family members **(Figure 4H)**. Regarding ileal regional patterning, we identified 35 genes that are
253 associated with at least one Ile-specific enhancer and are significantly increased in both transcription and
254 steady state-expression in Ile relevant to Duo spheroids **(Figure 4K)**. Among these, the genes associated
255 with the greatest number of Ile-specific enhancers include members of the *HOXA*/*C*/*D* clusters, *FJX1*
256 *CSRNP1*, and *HOTTIP* **(Figure 4L-M)**, several of which were also identified as Ile spheroid-specific
257 marker genes **(Figure 2G)**. These analyses together reveal previously undescribed temporal dynamics of
258 the HOX cluster genes during the events of SI lineage formation followed by ileal regional patterning: first,

259 the HOXB cluster is activated during Duo spheroid formation (likely by nearby Duo-specific enhancers),
260 then subsequently the other three HOX clusters are activated during ileal patterning (likely by nearby Ile-
261 specific enhancers) **(Figure 4H, M)**.
262
263 **Identification of genes associated with enhancer 'hotspots' that emerge SI lineage formation or ileal**
264 **patterning**
265
266 It has been shown in previous studies that dense clusters of highly active enhancers (which we refer to as
267 'hotspots') occur nearby to genes that are especially critical for defining cell identity and status [36-38].
268 We sought to define for the first time the changing landscape of enhancer hotspots in this human model of
269 SI development. To accomplish this, we adapted a previously described methodology [36, 37], which
270 requires ChIP-seq based datasets, to work with ChRO-seq data instead **(Figure 5A)**. By implementing this
271 analysis pipeline with the Duo- and Ile-specific enhancers (defined in **Figure 4B** and **C** respectively), we
272 were able to identify enhancer hotspots formed during SI lineage formation and ileal patterning,
273 respectively.
274
275 To identify enhancer hotspots and the associated candidate genes that may be critical for SI lineage
276 formation, we first clustered Duo-specific enhancers into 'stitched enhancers' **(Supplementary Data 3)**.
277 Among the 2389 stitched enhancers, we identified 145 that exhibit strong enough transcriptional activity to
278 be designated as Duo-specific 'enhancer hotspots' **(Figure 5B)**. We assigned each of the Duo-specific
279 stitched enhancers (including non-hotspots and hotspots) to its nearest active gene (TPM > 50 in ChRO-
280 seq) within a 100 kb window on either end of the boundaries of the stitched enhancer. We found that the
281 overall increase in gene body transcription is significantly greater for the set of genes associated with Duo-
282 specific enhancer hotspots compared to those associated with stitched enhancers that are not hotspots
283 **(Figure 5C)**. This finding is consistent with the notion that hotspots exert particularly strong effects on
284 transcription. Among the 73 genes nearest to Duo-specific enhancer hotspots, 25 exhibit highly significant

285 increases in both transcription (ChRO-seq) and steady state expression (RNA-seq) in Duo spheroids relative
286 to DE **(Figure 5D-E)**. Many of these were defined earlier as SI or Duo-specific markers (e.g., *CDX2*,
287 *FOXA1*, *HOXB* cluster, *SALL1* and *LINC00543*) **(Figure 2)**.
288
289 We next performed a similar analysis to identify enhancer hotspots and nearby genes associated with ileal
290 regional specification. Among 305 stitched enhancers formed by Ile-specific enhancers (relative to Duo
291 spheroids) **(Supplementary Data 3)**, we defined 29 Ile-specific enhancer hotspots **(Figure 5F)**. We
292 assigned each of the Ile-specific stitched enhancers (including non-hotspots and hotspots) to its nearest
293 active gene (TPM > 50 in ChRO-seq) within a 100 kb window on either end of the boundaries of the stitched
294 enhancer. Similar to the observation made for the transition from DE to Duo spheroids, the genes nearest
295 to Ile-specific enhancer hotspots exhibit a significantly greater increase in transcription compared to those
296 associated with non-hotspot stitched enhancers **(Figure 5G)**. Among the 22 genes nearest to Ile-specific
297 enhancer hotspots, 12 exhibit highly significant increases in both transcription (ChRO-seq) and steady-state
298 expression (RNA-seq) in Ile relative to Duo spheroids **(Figure 5H)**. These genes include *HAND2*, *HOXC/D*
299 family members, *FJX1*, *DLX5*, *CSRNP1* and *PITX1* **(Figure 5I)**, many of which were also identified earlier
300 as Ile-specific markers **(Figure 2)**. Similar analyses were performed for the hESC and DE stages and the
301 results are summarized in **Supplementary Figure 5**.
302
303 **Identification of candidate TF drivers and their networks relevant to SI lineage formation or ileal**
304 **patterning**
305
306 To discover putative key TF drivers and their cistromes (genome-wide set of TF targets) in this model, we
307 first used the tool HOMER to determine motifs enriched in each set of stage-specific enhancers and then
308 identified genes associated enhancers that harbor specific motifs of interest **(Figure 6A)**. Expectedly,
309 HOMER analyses of hESC- and DE-specific enhancers revealed binding sites of POU5F1 and SMAD2 to

310 be the most enriched motifs, respectively **(Supplementary Figure 6A-D)**. We next performed motif
311 enrichment analyses on Duo- **(Figure 6B)** and Ile-specific enhancers **(Figure 6D)**.
312
313 As expected, among the 2,787 Duo-specific enhancers relative to non-Duo specific enhancers **(Figure 6B)**,
314 we detected significant enrichment for the binding motifs of CDX2 and many forkhead box TFs **(Figure**
315 **6C)**. We also performed the same analysis with Duo-specific enhancer hotspots relative to non-hotspots
316 **(Figure 6B)** and observed significant enrichment for binding motifs of SP5 and RXR **(Figure 6C)**, the
317 former of which is defined as a Duo-specific gene marker **(Figure 2G)**, but neither of which have been
318 associated previously with early SI development. Genes that are associated with Duo-specific enhancers
319 harboring motifs of one or more of these TFs and that are significantly increased in transcription (ChRO-
320 seq) and steady-state expression (RNA-seq) are shown in **Figure 6F (see full gene list in Supplementary**
321 **Figure 6E)**. The TF cistrome analysis suggests that CDX2 is a prominent transcriptional activator of the
322 *HOXB* cluster **(Figure 6F)**, which is consistent with previous observations [8]. Moreover, this analysis
323 reveals candidate TF activators of the gene *CDX2*. While *CDX2* locus has a CDX2 motif in its promoter
324 region in Duo-spheroids (data not shown) and is thus likely subject to autoregulation as previously
325 described [39], Duo-specific enhancers that are associated with the *CDX2* locus do not harbor any binding
326 motifs for CDX2 and instead have binding motifs for other TFs such as FOXA1 and FOXO3 **(Figure 6F)**.
327 More importantly, through this analysis we were able to identify candidate TF drivers of SI/Duo spheroid
328 marker genes that do not appear to be a part of the CDX2 cistrome. For example, *WFDC2* and *SALL1*
329 (defined earlier as a marker gene of SI and Duo spheroids respectively) are associated with FOXA1 and
330 FOXO3 motifs; *SP5* (defined earlier as a marker gene of Duo spheroid) is associated with the TCF3 motif
331 **(Figure 6D)**.
332
333 We next performed a similar analysis for ileal regional patterning. Among the 328 Ile-specific enhancers
334 relative to non-Ile specific enhancers **(Figure 6D)**, we detected significant enrichment for the binding
335 motifs of several key TF families including CDX2, HOXC9, HOXA11 and PITX1 **(Figure 6E)**, the last of

336 which is itself associated with Ile-specific enhancer hotspots **(Figure 5I)**. We also performed the same
337 analysis with Ile-specific enhancer hotspots relative to non-hotspots but did not observed significant
338 enrichment for binding motifs (data not shown). Genes that are associated with Ile-specific enhancers
339 harboring motifs of one or more of these TFs and that are significantly increased in transcription (ChRO-
340 seq) and steady-state expression (RNA-seq) are shown in **Figure 6G**. This analysis led to several new
341 observations. Firstly, CDX2 likely plays a role in both Duo spheroid formation and ileal regional patterning,
342 but apparently through distinct cistromes **(Figure 6F and G)**. Notably, the activation of *HOXD* and *HOXC*
343 (especially *HOXC8* and *9*) loci is associated with Ile-specific TREs harboring the CDX2 motif **(Figure 6E)**.
344 While the HOXB cluster is induced in Duo spheroids, it remains highly transcribed in Ile spheroids, likely
345 due to the fact that the nearby TREs with CDX2 motifs that are activated during Duo spheroid formation
346 are retained during ileal regional patterning **(Figure 4H, M)**. Secondly, we found that the activation of
347 HOXA, C, and D families are associated with overlapping but distinct TF regulators. Specifically, while
348 HOXA11 seems to act on all four *HOX* clusters, PITX1 is associated only with the *HOXA* cluster and
349 HOXC9 is preferentially associated with the *HOXD* cluster **(Figure 6G)**. Finally, this analysis also reveals
350 other TF cistrome networks that do not target *HOX* genes during ileal patterning **(Figure 6G)**. For example,
351 *CSRNP1*, *FJX1* and *ID1* loci are associated with motifs of AP-1 (e.g. JUNB and FOSL2) and/or ATF factors.
352 **(Figure 6G)**.
353
354 **The HOX cluster dynamics observed in the directed differentiation model of SI is present in the early**
355 **developmental stages of primary human fetal SI tissue**
356
357 Across various analyses in this study we repeatedly observed a previously undescribed dynamics of HOX
358 gene activation across SI lineage formation to ileal patterning. Notably, ChRO-seq data reveals strong
359 transcriptional activation of the *HOXB* cluster (particularly HOXB1) upon acquisition of SI identity (Duo
360 spheroids) and the sequential transcriptional activation of members of *HOXA*, *C*, *D* clusters during ileal
361 regional patterning (Ile spheroids) **(Figure 7A)**, which as we showed previously is also reflected in the

362 pattern of emergence of nearby enhancers and enhancer hotspots **(Figure 4,5)**. This unique HOX cluster
363 dynamics is largely reflected at the level of steady-state expression (RNA-seq) also **(Figure 7B)**, except for
364 the 3'-most *HOXA* genes (e.g. *HOXA1*, *A2*) and 5'-end HOXB genes (e.g. *HOXB7*, *B9*) **(Figure 7A-B)**,
365 which may be subject to post-transcriptional regulation.
366
367 To determine whether the differences in *HOX* gene expression between early Duo and Ile extends beyond
368 this *in vitro* directed differentiation model, we mined the single cell RNA-seq (scRNA-seq) dataset of
369 primary human Duo and Ile tissue at very early stages of fetal development [13, 40, 41]. Specifically, we
370 focused on matched Duo and Ile fetal tissue at week 11, 14 and 19 of gestation. It is important to note that
371 these developmental time points, though considered as early/mid gestation stage, represent time points
372 much later than what is reflected in our much earlier stage SI spheroids, and therefore exhibited weaker
373 overall expression of *HOX* genes. Since the SI spheroids generated from the directed differentiation method
374 are mainly comprised of epithelial cells but also contain some mesenchymal cells, we analyzed both cell
375 types from the scRNA-seq data, together and separately, in order to examine the *HOX* dynamics. At week
376 11 of gestation, we found that the *HOXB* cluster is active in both human fetal Duo and Ile cells (epithelium
377 plus mesenchyme), with *HOXB1* particularly enriched in Duo cells and 5' *HOX* genes of all four clusters
378 enriched in Ile cells **(Figure 7C)**. This distinct HOX pattern between primary human fetal Duo and Ile
379 tissues is consistent with our observations in the Duo and Ile spheroids generated from the directed
380 differentiation method. Interestingly, this signal is present only at week 11 and dissipates as the
381 developmental process proceeds (week 14 and 19) **(Figure 7C)**. The scRNA-seq analysis also offers the
382 new insight that the enrichment of *HOXB1* in primary fetal Duo tissue is mainly driven by the mesenchymal
383 fraction, whereas the activation of 5' HOX genes of all four clusters in Ile tissue occurs in both epithelial
384 and mesenchymal cells **(Figure 7C)**. Overall, this analysis suggests that the Duo and Ile spheroids
385 generated from the directed differentiation model appear to capture the region-specific HOX features that
386 are transiently present in the developing human SI .
387

388 Beyond the finding of the previously undescribed temporal HOX cluster patterns highlighted here, the
389 ChRO-seq analyses in this study provide a comprehensive characterization of chromatin regulatory
390 dynamics, enhancer activity profiles, and transcriptional programs that are associated with human SI lineage
391 formation and regional patterning. A working model is provided in **Figure 7E**.
392
393 **Discussion**
394
395 By leveraging the recently developed ChRO-seq technology, we generated for the first time ever
396 comprehensive chromatin regulatory landscapes across the different stages of directed differentiation from
397 hPSC to SI spheroids with regional specification. Specifically, we defined: (1) early marker genes that
398 label SI lineage formation and regional specification, (2) the map of active regulatory elements (promoters
399 and enhancers) as well as enhancer hotspots associated with SI lineage formation and ileal patterning, and
400 (3) candidate key TF drivers and their cistromes relevant to these critical developmental events. These
401 findings motivate future experiments to investigate the functional role of candidate regulators and assess
402 the sufficiency and necessity of the associated enhancers and enhancer hotspots for controlling
403 transcriptional networks relevant to early human SI development and patterning.
404
405 We identified for the first time *HES4* as marker gene of SI spheroids irrespective of regional identity in the
406 context of gut development; one likely reason for this is that it is absent in the genome of the mouse [42],
407 the model organism most used previously to study gut development. Moreover, the marker genes that we
408 found to specifically label Duo identity include *SP5* and *SALL1*, the latter of which has been linked to
409 Townes-Brocks syndrome with developmental malformation of multiple organs including the limb, kidney
410 and gastrointestinal tract [43, 44]. While *SP5*, *SALL1* and other marker genes of Duo identity were reported
411 to enhance Wnt signaling [45, 46], we identified a distinct set of marker genes that specifically label Ile
412 identity (e.g. *HAND2*, *FJX1* and *CSRNP1*) that are also known to be downstream of Wnt signaling. *Fjx1*

413 has been shown to function through the Wnt/planar cell polarity (PCP) pathway to determine anterior-
414 posterior axis in *Drosophila* [47, 48] and is expressed in the epithelium of murine developing gut [49].
415 *Csrnp1* is expressed in neural crest progenitors in a Wnt1/β-catenin-dependent manner and acts as a critical
416 effector of Wnt signaling for driving neural crest formation [50]. Together, these observations indicate that
417 different sets of Wnt signaling effector genes are likely induced upon different durations of exposure to the
418 Wnt-activating supplements during directed differentiation. Future investigation of the functional roles of
419 these genes in the context of early gut development is warranted. The present study also uncovers the
420 associated enhancers (or enhancer hotspots) and TF motifs of these marker genes, which may elucidate the
421 potential mechanisms through which these genes are regulated during the establishment of SI identity and
422 regional specification.
423
424 Although HOX genes have been reported to play important roles in proximal-distal axis patterning in the
425 gut, our study uncovers previously undescribed dynamics of HOX cluster genes in the directed
426 differentiation model of human SI development, unveiling the complexity of HOX biology in the control
427 of small intestinal regionalization. The early literature has observed that 3' HOX genes are relevant to
428 proximal gut and 5' HOX genes are restricted to distal gut (known as 'spatial collinearity' of HOX genes)
429 [51, 52]; however, our study shows that HOX genes do not simply follow this collinear pattern. We found
430 that the activation of the HOXB cluster first occurs during the formation of Duo spheroids and the other
431 HOX clusters (HOXA, C and D) are activated subsequently during the formation of Ile spheroids. Through
432 mining of a novel scRNA-seq dataset [13, 40, 41], we also showed that this distinct HOX behavior between
433 Duo and Ile regions is present in primary human fetal SI tissue. Through ChRO-seq analyses, we also
434 provided additional insights into HOX biology by elucidating candidate TF drivers and associated
435 enhancers and enhancer hotspots, which merits detailed investigation.
436

437 Our ChRO-seq analysis mapped the key transcriptional networks associated with SI lineage formation or
438 regional patterning. For example, we found that binding sites of FOXA1 and FOXO3 are enriched in active
439 enhancers that emerge only during SI identity acquisition, including those nearby *CDX2*, which encodes
440 the well-known master TF regulator of SI development. Notably, the *FOXA1* gene locus is actively up-
441 transcribed during the transition from definitive endoderm to SI spheroids and is associated with an
442 enhancer hotspot that emerges during this process. FOXA1 has been reported to regulate allocation of
443 murine SI secretory lineage [53] and proper gene expression in murine intestinal epithelium [54]. Recently,
444 dysregulation of intestinal *FOXA1* also has been linked to necrotizing enterocolitis in infants [55]. The
445 observations made in this study as well as previous reports together underscores the functional importance
446 of FOXA1 in human gut development. Also, our TF cistrome analyses indicate that CDX2 is involved in
447 not only in initial SI lineage establishment but also the subsequent event of ileal specification, but evidently
448 through distinct cistromes. One specific example is the association between the *HOXC9* locus and the
449 nearby Ile-specific active enhancers containing CDX2 binding motifs, which is consistent with a previous
450 study showing a remarkable downregulation of *Hoxc9* in distal but not in proximal intestine of *Cdx2* null
451 mice [2]. Recent evidence also demonstrates that CDX2 regulates developing and fully mature gut through
452 targeting distinct gene sets in mice [8] and that CDX2 is involved in patterning both intestinal epithelial
453 and mesenchymal cells in the hPSC-based directed differentiation model [13]. Our ChRO-seq findings
454 pertaining to CDX2 may further elucidate how CDX2 functions through a stage- or cell type-dependent
455 manner during early human gut development. Beyond CDX2, other candidate TFs associated with the
456 establishment of ileal identity are worthy of future investigation. One notable example is PITX1, the
457 ortholog of which has been reported previously in mouse models to be expressed in the distal SI region and
458 to be involved in cecum budding during murine development [56]. However, PITX1 has not been
459 previously appreciated in the context of ileal regional specification.
460
461 The use of the state-of-art *in vitro* model of human SI development is a critical feature of this study given
462 the lack of other existing platforms or resources that are readily available for the study of very early

463 developmental time points pertaining to SI lineage formation and regional patterning. It is also important
464 given the growing appreciation for cross-species differences in developmental processes at the molecular
465 level [57-60]. Furthermore, the genome-wide characterization of the chromatin regulatory landscape by
466 ChRO-seq in this model system generates valuable translational knowledge for a better understanding of
467 the transcriptional programming underlying early SI development in humans. Our analysis pipeline can
468 also be applied to other hPSC-derived organoid systems to gain insights into dynamic transcriptional
469 programming and chromatin status during early stages of human organogenesis. Most importantly, the
470 identification of the candidate drivers in the present study may ultimately improve methods of generating
471 therapeutic replacement SI as well as molecular therapies for children and possibly adult patients.

472

473

474 **Methods**

475

476 ***Directed differentiation of hESCs***. Differentiation of H9 hESCs and organoids was performed as
477 previously published, with minor modifications [27, 61]. Briefly, the endoderm was generated by treatment
478 of activin A (100 ng/ml) for 3 consecutive days in Roswell Park Memorial Institute 1640 (RPMI-1640)
479 media supplemented with 0% (v/v), 0.2% (v/v) and 2.0% (v/v) Hyclone defined fetal bovine serum (dFBS).
480 The endoderm cultures then received daily treatments of FGF4 (500 ng/mL) and CHIR99021 (2 μM) for
481 next 10 days. The intestinal spheroids representing fetal duodenum and ileum were collected on day 5 and
482 day 10, respectively [28]. Subsets of spheroids collected at day 5 and day 10 were then cultured in Matrigel
483 with the previously defined intestine growth media[28] for 28 days in order to mature into organoids. The
484 resulting organoids were then prepared for cell sorting to purify EPCAM+ cell population (epithelial
485 fraction) and the sorted cells were processed for RNA-seq library preparation. The cells generated at the
486 stages of hESC, endoderm, duodenal spheroids and ileal spheroids were subject for ChRO-seq and RNA-
487 seq library preparation.

488

489 ***Fluorescence activated cell sorting (FACS).*** Methods for organoid dissociation into single cells followed
490 by selection of the epithelial component with FACS, were based on previously described procedures[62].
491 All solutions, including overnight pretreatment of the organoid cultures, contained 10μM Y27632 (Tocris).
492 Matrigel was digested for 30 minutes with cold 4mM EDTA-DPBS and organoids were washed 4X with
493 cold DPBS. Structures were enzymatically dissociated into single cells using the Tumor Dissociation Kit
494 (human) (Miltenyi Biotec) with a gentleMACS™ Octo Dissociator (with heaters; Miltenyi Biotec) for 50
495 minutes at 37°C. The cell suspension was then washed with 0.5% BSA-2mM EDTA-DPBS over a
496 succession of cell strainers, 100μm, 40μm (Corning) and 20μm (CellTrics) and centrifuged for 5 minutes
497 at 500xg. Cells were labeled with EpCAM phycoerythrin (PE)-conjugated antibody (BioLegend) and an
498 EpCAM isotype-PE control (BioLegend), and were sorted in 0.1% BSA 2mM EDTA-DPBS on a MoFlo
499 Astrios 1 (Beckman Coulter; Brea, California) instrument at the University of Michigan BRCF Flow
500 Cytometry Core facility. Events were first selected with light-scatter and doublet discrimination gating,
501 followed by exclusion of dead cells using 1μM DAPI dilactate (Molecular Probes). EpCAM-PE(+)/DAPI(-)
502 cells were sorted into cold Advanced DMEM/F12 (Invitrogen). Collected cells were reanalyzed for a
503 purity-check and showed greater than 89% viable and 98% EpCAM-PE(+) events. Cells were pelleted at
504 500xg for 5 minutes and flash-frozen for subsequent RNA isolation.

505

506 ***Procr-mGFP-2A-LacZ mouse line and staining.*** The Procr-mGFP-2A-LacZ mouse line was generated
507 by knocking in a cassette of mGFP-2A-LacZ behind start codon of the *Procr* gene [32]. To perform X-gal
508 staining, embryos were isolated and washed in cold PBS followed by incubation in ice-cold fixative (30
509 mins up to E10.5, or 50 mins for E11.5-E13.5) on a rocking platform. Fixative contains 37% formaldehyde,
510 25% glutaraldehyde, 10% NP-40 all dissolved in PBS. The fixative was removed and the whole embryo
511 washed twice in PBS for 20 mins at room temperature on a rocking platform. The β-galactosidase substrate
512 (5 mM K3FE(CN)6, 5 mM K4Fe(CN)6·3H2O, 2 mM MgCl2, 0.02% NP40, 0.1% sodium deoxycholate
513 and 1 mg ml–1 X-gal in PBS) was then added and the tissues incubated in the dark overnight at room

514 temperature. The substrate was removed and the tissues washed twice in PBS for 20 min at room
515 temperature on a rocking platform. Embryos were serially dehydrated using glycerol, and was stored in
516 80% glycerol at 4°C. Whole-embryo images were captured using Olympus SZX16.
517
518 ***Chromatin isolation.*** The chromatin isolation for ChRO-seq library preparation was performed as
519 previously described [20, 63]. Briefly, chromatin was isolated from the cells with 1× NUN buffer [0.3 M
520 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,
521 20 units per mL SUPERase In RNase Inhibitor (Life Technologies, AM2694), 1× Protease Inhibitor
522 Cocktail (Roche, 11 873 580 001)] and incubation at 12 °C on a ThermoMixer for 30 min. Samples were
523 centrifuged at 12,500xg for 30 min at 4 °C. The chromatin pellet was washed 3 times with 1 mL 50 mM
524 Tris-HCl, pH 7.5, supplemented with 40 units per ml RNase inhibitor. Chromatin storage buffer (50 mM
525 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
526 RNase inhibitor) was added to each sample. The samples were loaded into a Bioruptor and sonicated to get
527 the chromatin into suspension. Samples were stored at −80 °C before proceeding to ChRO-seq library
528 preparation.
529
530 ***ChRO-seq library and sequencing.*** After chromatin isolation, the ChRO-seq library preparation closely
531 followed the protocol described previously [20, 64]. Briefly, chromatin from at least 1×10^6 cells per
532 sample in chromatin storage buffer was mixed with an equal volume of 2×chromatin run-on buffer [10 mM
533 Tris-HCl, pH 8.0, 5 mM MgCl2,1 mM DTT, 300 mM KCl, 400 μM ATP (NEB, N0450S), 40 μM Biotin-
534 11-CTP (Perkin Elmer, NEL542001EA), 400 μM GTP (NEB, N0450S), 40 μM Biotin-11-UTP (Perkin
535 Elmer, NEL543001EA), 0.8 units per μL RNase inhibitor, 1% (w/v) Sarkosyl (Fisher Scientific,
536 AC612075000)]. The run-on reaction was incubated at 37 °C for 5 min. The reaction was stopped by
537 adding Trizol LS (Life Technologies, 10296-010) and pelleted with GlycoBlue (Ambion, AM9515) to
538 visualize the RNA pellet. The RNA pellet was resuspended in diethylpyrocarbonate (DEPC)-treated water
539 and heat denatured at 65 °C for 40 s. In the present study, base hydrolysis of RNA was performed by

540 incubating RNA with 0.2N NaOH on ice for 4 min. Nascent RNA was purified by binding streptavidin
541 beads (NEB, S1421S) before and in between the following procedures: (1) 3′ adapter ligation with T4 RNA
542 Ligase 1 (NEB, M0204L), (2) 5′ de-capping with RNA 5′ pyrophosphohydrolase (RppH, NEB, M0356S),
543 (3) 5′ end phosphorylation using T4 polynucleotide kinase (NEB, M0201L), (4) 5′ adapter ligation with T4
544 RNA Ligase 1 (NEB, M0204L). The resulting RNA fragments were used for a reverse transcription
545 reaction using Superscript III Reverse Transcriptase (Life Technologies, 18080-044) to generate cDNA.
546 cDNA was then amplified using Q5 High-Fidelity DNA Polymerase (NEB, M0491L) to generate the
547 ChRO-seq libraries. Libraries were sequenced (5' single end; single-end 75x) using the NextSeq500 high-
548 throughput sequencing system (Illumina) at the Cornell University Biotechnology Resource Center.
549 **Supplementary Data 1** provides the mapping statistics of the ChRO-seq experiments.
550
551 ***Total RNA isolation, mRNA-seq library and sequencing***. Total RNA was isolated using the Total
552 Purification kit (Norgen Biotek, Thorold, ON, Canada). High Capacity RNA to cDNA kit (Life
553 Technologies, Grand Island, NY) was used for reverse transcription of RNA. Libraries were generated
554 using the NEBNext Ultra II Directional Library Prep Kit (New England Biolabs, Ipswich, MA) and
555 subjected to sequencing (single-end 92x) on the NextSeq500 platform (Illumina) at the Cornell University
556 Biotechnology Resource Center. At least 80M reads per sample were acquired.
557
558 ***Mapping sequencing reads.*** In the ChRO-seq studies, the publicly available pipeline [33] was used to align
559 ChRO-seq reads. Since the libraries were prepared using adapters that contained a molecule-specific unique
560 identifier (first 6 bp sequenced), the PCR duplicates were first removed using PRINSEQ lite. Adapters
561 were trimmed from the 3′ end of remaining reads using cutadapt with a 10% error rate. Reads were mapped
562 with the Burrows-Wheeler Aligner (BWA) to the human reference genome hg38 plus a single copy of the
563 Pol I ribosomal RNA transcription unit (GenBank U13369.1). The location of active RNA polymerase was
564 represented by a single base that denotes the 3′ end of the nascent RNA, which corresponds to the position
565 on the 5′ end of each sequenced read. Mapped reads were converted to bigwig format using BedTools and

566 the bedGraphToBigWig program in the Kent Source software package. For visualization purpose, bigwig
567 files from identical stages were merged and normalized to a total signal of $1x10^6$. In the RNA-seq studies,
568 reads were mapped to human genome hg38 using STAR (v2.5.3a) [65] and transcript quantification was
569 performed using Salmon (v0.6.0) [66] with GENCODE release 25 transcript annotations. The expression
570 (RNA-seq) levels of genes were normalized using DESeq2 [67]. All the samples except an Ile HIO sample
571 had < 80% mapping rates. Although the Ile HIO sample had an unfavorable mapping rate, it was able to
572 show elevated levels of Ile-associated regional markers compared to the Duo HIO sample **(Supplementary**
573 **Figure 3A)**.
574
575 ***ChRO-seq quantification of gene loci and promoter transcription activity.*** Gene definitions were obtained
576 from GENCODE v25 annotations. To quantify transcription activity of gene loci, ChRO-seq signals
577 present in annotated gene bodies were used with exclusion of reads within 500 base downstream of
578 transcription start site (TSS) to avoid bias generated by the RNA polymerase pausing at the promoters.
579 Genes with gen body < 1000 base were excluded from all the gene body related analysis, given that genes
580 with short gene bodies are likely biased when excluding the pause peak. The ChRO-seq reads were
581 normalized by the length of gene bodies to transcripts per million (TPM). To determine promoter activity
582 of genes, stranded ChRO-seq signals at the promoter proximal region (500 bps upstream and 200 bps
583 downstream of TSS) were used. The promoter regions that have significant changing patterns were defined
584 using the following criteria: average TPM > 5 and padj < 0.05 across all the stages by likelihood ratio test
585 (DESeq2). The promoter regions that have significant changing patterns and are annotated as 'protein-
586 coding' gene type were further subject to clustering analysis ('degPatterns' in R platform) .
587
588 ***Differential expression and pathway analyses of genes.*** The differential analysis of gene bodies in ChRO-
589 seq data was performed using DESeq2 package [67]. For all the analyses except stage-specific marker
590 analysis, the normalized levels of transcription or expression, foldchange and the statistic filtering were

591 based on the DESeq2 analysis including only the two stages in a comparison. The pathway enrichment
592 analyses with subsets of genes were assessed using Enrichr [68].
593
594 ***Identification of stage-specific marker genes.*** Firstly, genes that have ChRO-seq signal in gene bodies
595 (excluding the first 500 bps downstream of the TSS) greater than 50 TPM in the stage of interest are
596 identified and filtered according to padj < 0.2, p < 0.05, fold change > 1.5 compared to all the other stages
597 (Wald test; DESeq2). Secondly, genes that have RNA-seq signal greater than 100 base mean units are
598 identified and filtered according to padj < 0.2, p < 0.05, fold change > 1.5 compared to all the other stages
599 (Wald test; DESeq2). Thirdly, the output from these two analyses are intersected to arrive at final gene
600 lists for stage-specific markers. To identify genes that label SI spheroids irrespective of regional identity,
601 analyses were done using the same criteria except the fold change criterion. The genes that have short gene
602 bodies (< 1000 bp) or annotated under the pseudogene category are excluded from this analysis.
603
604 ***TRE identification, annotation, categorization and comparison with other datasets.*** The active
605 transcriptional regulatory elements (TREs) were identified by dREG tool [33]. The annotation of the
606 identified TREs was defined using *annotatePeaks.pl* function (genome = hg38) of HOMER package [69]
607 based on GENCODE v25 annotations. To compare the TRE landscape of SI spheroids with developing
608 and mature SI in humans, DNase-seq datasets from the Roadmap Epigenomics Project were used.
609 Specifically, first, the files E085-DNase.hotspot.fdr0.01.peaks.v2.bed (fetal SI) and E109-
610 DNase.hotspot.fdr0.01.peaks.v2.bed (adult SI) were downloaded in hg19, converted to hg38 using the
611 USCS liftOver tool, and used to define DNase-based open chromatin regions unique to fetal and adult SI
612 using the bedtools intersect function (at least 1 base overlapping). The fetal and adult SI-specific DNase
613 open chromatin regions were further intersected with TREs present in SI spheroids using bedtools intersect
614 function (at least 1 base overlapping). To assign active TREs as promoters, the TREs with at least 1 base
615 overlapping with the window of -1000 base and +200 base of annotated TSSs were defined as proximal
616 TREs, or promoters. The rest of the active TREs were defined as distal TREs, or enhancers.

617

618 ***Stage-specific TRE density analysis.*** The stage-specific TREs were defined as TREs of which the ChRO-
619 seq intensity is significantly higher in a stage of interest relative to a comparative stage (padj < 0.05 and
620 $\log_2$fold change > 2.5 by DESeq2) [67]. The ChRO-seq intensity was determined by the sum of the un-
621 normalized ChRO-seq reads from both strands within a TRE region. The density of stage-specific TREs
622 was determined by the number of TREs present within the window of +100 kb and -100 kb around the TSS
623 for all the genes which are actively transcribed in the stage of interest (TPM > 50 in the ChRO-seq study).
624 The genes which are associated with stage-specific TREs are defined using the following criteria: (1) genes
625 have density of stage-specific TRE > 0, (2) genes are actively transcribed (TPM > 50 in the ChRO-seq) and
626 expressed (base mean > 100 in the RNA-seq) in the stage of interest, and (3) the genes are significantly
627 uptranscribed and upregulated (padj < 0.2, p < 0.05, fold change > 1.5 in both ChRO-seq and RNA-seq by
628 DESeq2) in the stage of interest relative to a comparison stage.

629

630 ***Identification of de novo stage-specific enhancer hotspots and the associated genes.*** The stage-specific
631 active distal TREs (enhancers) were used in the enhancer hotspot analysis. The enhancer hotspots in this
632 study were identified by the criteria similar (with slight modifications) to the studies describing 'super-
633 enhancers'[36, 37]. Briefly, the stage-specific enhancers in proximity of distance (< 12.5kb) were stitched.
634 For each of the stage-specific stitched enhancers, the transcription activity was determined by the sum of
635 un-normalized ChRO-seq signals from both strands of each individual enhancer. To further identify stage-
636 specific enhancer hotspots, a tangent line was applied the stage-specific stitched enhancers and they were
637 ranked based on their transcriptional activities in a plot. The ones above the tangent line in the analysis
638 were defined as stage-specific enhancer hotpots. To identify the genes which are associated with stage-
639 specific stitched enhancers, a given stage-specific stitched enhancer is assigned to the gene of which the
640 transcription in the gene body is active (TPM > 50 in the matching stage) and the TSS is closest to the
641 border of the enhancer hotspot region.

642

643 ***Transcription factor binding motif enrichment analysis.*** HOMER tool [69] was used to determine
644 enrichment of sites corresponding to known motifs with stage-specific TREs (relative to a comparative
645 stage). More specifically, we used function *findMotifsGenome.pl* (genome = hg38 and size = given) and
646 the TREs which are shared or unique to the comparative stage were used as background.
647
648 ***Transcription factor cistrome analysis.*** For a TF of interest, the binding motifs present in the stage-specific
649 TREs were identified and the density of the motif was determined by the number of motifs within the
650 window of +100 kb and -100 kb around the TSS for all the genes which are actively transcribed in the stage
651 of interest (TPM > 50). The cistrome of a TF is defined using the following criteria: (1) genes have motif
652 density > 0, (2) genes are actively transcribed (TPM > 50 in the ChRO-seq) and expressed (base mean >
653 500 in the RNA-seq) in the stage of interest, and (3) the genes are significantly uptranscribed and
654 upregulated (padj < 0.2, p < 0.05, fold change > 1.5 in both ChRO-seq and RNA-seq by DESeq2) in the
655 stage of interest relative to a comparison stage.
656
657 ***Single cell RNA-seq data analysis.*** The scRNA-seq analysis of primary human fetal SI tissues was curated
658 using recent, pre-existing datasets [13, 40, 41]. Briefly, the scRNA-seq dataset of human developing SI
659 across different regions and multiple gestational time points were generated by 10x Genomics platform and
660 mapped to hg19 using default alignment parameters provided by Cell Ranger. Seurat (v3.1) package [70]
661 was applied to the downstream analysis. Cells with greater than 5% mitochondrial transcript proportion
662 and fewer than 1000 unique detected genes were excluded. Ribosomal genes, mitochondrial genes, and
663 genes located on sex chromosomes were also removed before data integration into a Seurat object. The
664 cell type annotation of the dataset was pre-defined and detailed in [13]. In the present study, we focused
665 on the data of matched human fetal duodenum and ileum samples (week 11, 14 and 19 of gestation).
666
667 ***Statistics.*** All the padj and p-values presented in this study were determined by Wald test (DESeq2), unless
668 otherwise specifically noted.

669

670 **Availability of data, materials and computation scripts**

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

672

673 **Funding**

674 ADA Pathway to Stop Diabetes Research Accelerator (1-16-ACE-47 to P.S.); Empire State Stem Cell Fund
675 (C30293GG to Y.-H. H.); Intestinal Stem Cell Consortium (U01DK103141 to J.R.S.); a collaborative
676 research project funded by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)
677 and the National Institute of Allergy and Infectious Diseases (NIAID), the NIAID Novel Alternative Model
678 Systems for Enteric Diseases (NAMSED) consortium (U19AI116482 to J.R.S.); the support from the
679 University of Michigan Center for Gastrointestinal Research (UMCGR) (NIDDK 5P30DK034933 to
680 J.R.S.); the Chan Zuckerberg Initiative DAF (CZF2019-002440 to J.G.C.); the support an advised fund of
681 Silicon Valley Community Foundation (J.G.C.); the European Research Council (Anthropoid-803441 to
682 J.G.C.).

683

684 **Author Contributions**

685 Conceptualization, Y.-H.H., P.S.; Cell culture, S.H.; Cell sorting, M.K.D.; Chromatin isolation and
686 library preparation for sequencing studies, Y.-H.H.; Bioinformatic analyses and data curation, Y.-
687 H.H.; Resources and experiments of *Procr* reporter mice, Q.C.Y. and Y.A.Z.; Resources of scRNA-
688 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.,
689 J.R.S., and P.S.; Supervision, P.S.; Funding acquisition, Y.-H.H., J.R.S., and P.S.

690

691 **Acknowledgments**

692 We would like to thank members of the Sethupathy laboratory, most notably Dr. Michael Shanahan, Dr.
693 Ajeet Singh, and Dr. Matt Kanke, for helpful comments and feedback on the study and manuscript. We
694 specially thank Dr. Tim Dinh in the Sethupathy laboratory for help in developing the bioinformatic analysis

695 pipeline of ChRO-seq data and providing technical consultation on bioinformatics, Dr. Matt Kanke for
696 providing additional technical consultation on bioinformatics, and Ramja Sritharan in the Sethupathy
697 laboratory for offering technical consultation on the ChRO-seq protocol. Additionally, we are grateful to
698 the Danko laboratory, specifically Ed Rice and Dr. Charles Danko, for helpful conversations on improving
699 implementation of the ChRO-seq protocol. Finally, we also thank Dr. Jen Grenier and the Cornell
700 Transcriptional Regulation & Expression Facility for RNA sequencing support; Michael Dellheim of the
701 University of Michigan BRCF Flow Cytometry Core; Gina Newsome, Erika Katz, Maliha Berner, and
702 Angeline Wu of the Michigan Medicine Translational Tissue Modeling Laboratory; and a University of
703 Michigan funded initiative (Center for Gastrointestinal Research, Office of the Dean, Comprehensive
704 Cancer Center, Departments of Pathology, Pharmacology, and Internal Medicine) with support by the
705 Endowment for Basic Sciences.

706

707 **Conflicts of interest**

708 The authors disclose no conflicts.

709

710 **Figure Legends**

711 **Figure 1. ChRO-seq defines dynamics of nascent transcription at promoters and gene loci during**
712 **directed differentiation from human pluripotent stem cells (hPSCs) to small intestinal (SI) spheroids**
713 **with regional specification.** (A) Schematic diagram of the distinct stages of directed differentiation and
714 the genome-scale approach of ChRO-seq. (B) Hierarchical clustering analysis of gene transcription profiles
715 across all stages of the model. (C) Transcriptional activity (ChRO-seq signal) of the genes *SOX2*, *GATA6*,
716 *CDX2*, and *FOXF1* across all stages of the model. (D) Normalized ChRO-seq signal at the gene bodies of
717 *SOX2*, *GATA6*, *CDX2* and *FOXF1* across all stages of the model. Scale (+/- 25) is fixed across stages. (E)
718 Volcano plots showing differentially transcribed genes in the indicated comparisons (average TPM > 25,
719 $\log_2$ fold change of transcription > 1, padj < 0.2 and p < 0.05 by Wald test; DESeq2). Results of pathway

720 enrichment analyses of up-transcribed (red) and down-transcribed (blue) genes in the indicated comparisons
721 (GO Biological Process 2018) are also shown. (F) Identification of protein-coding genes that have high
722 promoter activity in SI spheroids (n = 2,992) through application of the likelihood ratio test across all stages.
723 (G) Expression levels (RNA-seq) of *BEX5* and *PROCR* in stages of DE and SI spheroids. (H) RT-qPCR
724 of *PROCR* in an independent batch of samples (DE = 3; Duo spheroid = 2). *P < 0.05 by one-tailed t test
725 (validation of sequencing data). (I) X-gal staining for β-galactosidase activity of Procr-mGFP-2A-LacZ
726 mouse reporter line showing a robust *Procr* expression in midgut at E9.5 (left; arrowhead) and E10.5 (right;
727 dashed outline). Scale bar = 500 μm. ChRO-seq study: hESC, n = 3; DE, n = 4; Duo spheroid (Duo), n =
728 3; Ile spheroid (Ile), n = 3. RNA-seq study: hESC, n = 2; DE, n = 3; Duo, n = 6; Ile, n = 4. TPM, transcripts
729 per million. nc, normalized counts. RQV, relative quantitative value.
730

731 **Figure 2. Identification of transcriptional markers of specific stages in the directed differentiation**
732 **model of human developing SI.** (A) Hierarchical clustering analysis of transcribed genes (TPM > 50 at
733 least in one stage) (z-score). (B) Genome-wide correlation of transcribed levels (ChRO-seq) and
734 expressed levels (RNA-seq) of genes. Transcription and expression levels of genes were averaged across
735 all stages. No transcription and expression thresholds were used (see also Supplementary **Figure 3B-E**).
736 (C) Identification of markers that label specific stages at the levels of both transcription and steady-state
737 expression. The total number of markers (center of the donut plot) and the proportion of marker types are
738 shown for each stage. (D-G) Marker genes of hESC (D), DE (E), SI spheroids regardless of regional
739 identity (F), Duo and Ile (G). In panel (D-F), only the top 50 most variable genes across stages are shown
740 (see **Supplementary Data 2** for full list). Heatmaps denote variation across stages (z-score). ChRO-seq
741 study: hESC, n = 3; DE, n = 4; Duo spheroid (Duo), n = 3; Ile spheroid (Ile), n = 3. RNA-seq study:
742 hESC, n = 2; DE, n = 3; Duo, n = 6; Ile, n = 4.
743

744 **Figure 3. Transcriptional regulatory element profiles reveal notable re-wiring of chromatin activity**
745 **in the directed differentiation model of human developing SI.** (A) Schematic for identification and
746 categorization of active TREs, followed by definition of TREs with altered activity during stage
747 transitions of the model. (B) Hierarchical clustering analysis of profiles of all active TREs. (C) Size
748 distribution of active TREs. (D) Genomic locations of active TREs (hg38 build). (E) Percentage of fetal
749 and adult SI-specific open chromatin regions (defined according to DNaseI-seq data from the Roadmap
750 Epigenomics Project) overlapping with active TREs present in Duo and/or Ile spheroids. (F) ChRO-seq
751 signal around the Intestinal Critical Region (ICR) [35]. An active TRE identified in our study (chr16:
752 1430731-1431300; hg38) largely overlaps with the ICR and exhibits strong ChRO-seq signals in SI
753 spheroids. Scales in right panel (+25/-25) and left panel (+166/-15; the maximum range) are fixed across
754 different stages. (G) The transcriptional activity of the TRE associated with the ICR (chr16: 1430731-
755 1431300; hg38) across all stages of the model (*** $P < 0.001$ by Wald test; DESeq2). (H) Hierarchical
756 clustering (upper) and PCA (lower) analysis of promoter profiles in the indicated stage. (I) (H)
757 Hierarchical clustering (upper) and PCA (lower) analysis of enhancer profiles in the indicated stage. (J)
758 Active TREs present in hESC and DE (n = 99,996) and ranked based on the log2 fold change in activity
759 during the transition to DE. (K) Numbers of promoters and enhancers that are unchanged, gained (up), or
760 lost (down) during the transition from hESC to DE. (L) Active TREs present in DE and Duo spheroid (n
761 = 91,900) and ranked based on the log2 fold change in activity during Duo spheroid formation. (M)
762 Numbers of promoters and enhancers that are unchanged, gained (up), or lost (down) during the transition
763 from DE to Duo spheroid. (N) Active TREs present in Duo and Ile spheroid (n = 77,084) and ranked
764 based on the log2 fold change in activity during Ile spheroid formation. (O) Numbers of promoters and
765 enhancers that are unchanged, gained (up), or lost (down) in Ile compared to Duo spheroids. hESC, n =
766 3; DE, n = 4; Duo spheroid (Duo), n = 3; Ile spheroid (Ile), n = 3.
767

768 **Figure 4. Identification of genes associated with a high density of nearby enhancers that emerge**
769 **during SI lineage formation or ileal patterning.** (A) The strategy for identifying genes associated with
770 stage-specific enhancers. (B) Venn diagram showing stage-specific and shared enhancers between DE
771 and Duo spheroids. Duo-specific enhancers (n = 2787) represent enhancers emerging during SI lineage
772 formation. (C) Venn diagram showing stage-specific and shared enhancers between Duo and Ile
773 spheroids in the event of ileal patterning. Ile-specific enhancers (n = 328) represent enhancers emerging
774 during ileal regional patterning. (D) Cumulative distribution of ChRO-seq fold change in genes grouped
775 into three different categories of Duo-specific enhancer density during the transition to Duo spheroids.
776 (E) Boxplot of ChRO-seq fold change in genes grouped into three different categories of Duo-specific
777 enhancer density during the transition to Duo spheroids. (F) Identification of genes associated with Duo-
778 specific enhancers (n = 117). (G) Bar graph showing genes associated with Duo-specific enhancers (left
779 panel). Top 30 genes based on enhancer density are highlighted (right panel). (H) Normalized ChRO-seq
780 signal around the *HOXB* cluster in the stages of DE, Duo, and Ile spheroids. Duo-specific TREs are
781 marked. (I) Cumulative distribution of ChRO-seq fold change in genes grouped into three different
782 categories of Ile-specific enhancer density during the transition to Ile spheroids. (J) Boxplot of ChRO-seq
783 fold change in genes grouped into three different categories of Ile-specific enhancer density during the
784 transition to Ile spheroids. (L) Identification of genes associated with Ile-specific enhancers (n = 35). (L)
785 Bar graph showing genes associated with Ile-specific enhancers. (M) Normalized ChRO-seq signal
786 around the *HOXD* cluster in the stages of DE, Duo, and Ile spheroids. Ile-specific TREs are marked.
787 ChRO-seq study: DE, n = 4; Duo spheroid (Duo), n = 3; Ile spheroid (Ile), n = 3. RNA-seq study: DE, n
788 = 3; Duo, n = 6; Ile, n = 4.
789

790 **Figure 5. Identification of genes associated with enhancer 'hotspots' that emerge during SI lineage**
791 **formation or ileal patterning.** (A) The strategy for identification of stage-specific enhancer hotspots
792 and their associated genes. (B) Duo-specific stitched enhancers are ranked by transcriptional activity

793 (ChRO-seq signal). The stitched enhancers with the highest transcriptional activity are defined as Duo-
794 specific enhancer hotspots (n = 145; blue) and the rest are enhancer non-hotspots (n = 2244; black). (C)
795 ChRO-seq fold change of genes associated with Duo-specific stitched enhancers, non-hotspots vs.
796 hotspots (Wilcoxon rank sum test). (D) Identification of genes associated with Duo-specific enhancer
797 hotspots. (E) Genes associated with Duo-specific enhancer hotspots (n=25). Relative position between
798 enhancer hotspots and transcription start sites (TSSs) of the associated genes are shown. Dot size denotes
799 transcriptional activity of a given DE-specific enhancer hotspot. (F) Ile-specific stitched enhancers are
800 ranked by transcriptional activity (ChRO-seq signal). The stitched enhancers with the highest
801 transcriptional activity are defined as Ile-specific enhancer hotspots (n = 29; red) and the rest are enhancer
802 non-hotspots (n = 276; black). (G) ChRO-seq fold change of genes associated with Ile-specific stitched
803 enhancers, non-hotspots vs. hotspots (Wilcoxon rank sum test). (H) Identification of genes associated
804 with Ile-specific enhancer hotspots. (I) Genes associated with Ile-specific enhancer hotspots (n=12).
805 Relative position between enhancer hotspots and TSSs of the associated genes are shown. Dot size
806 denotes transcriptional activity of a given Duo-specific enhancer hotspot. ChRO-seq study: DE, n = 4;
807 Duo spheroid (Duo), n = 3; Ile spheroid (Ile), n = 3. RNA-seq study: DE, n = 3; Duo, n = 6; Ile, n = 4.
808

809 **Figure 6. Identification of candidate TF drivers and their networks relevant to SI lineage formation**
810 **or ileal patterning.** (A) The strategy for identifying candidate TFs and their cistromes associated with
811 each stage transition of the model. (B) TF motif enrichment analyses were performed in Duo-specific
812 enhancers (n = 2787) relative to non-Duo-specific enhancers (n = 45,759) as well as in Duo-specific
813 enhancer hotspots (n = 145) relative to non-hotspots (n = 2,244). (C) Motifs significantly enriched in
814 Duo-specific enhancers (highlighted by light blue bar) or in Duo-specific enhancer hotspots (highlighted
815 by dark blue bar). The steady-state expression (RNA-seq) of the corresponding TFs are also shown. (D)
816 TF motif enrichment analyses were performed in Ile-specific enhancers (n = 328) relative to non-Ile-
817 specific enhancers (n = 38,674) as well as in Ile-specific enhancer hotspots (n = 29) relative to non-

818 hotspots (n = 276). (E) Motifs significantly enriched in Ile-specific enhancers. The steady-state
819 expression (RNA-seq) of the corresponding TFs are also shown. (F) Bubble plot showing genes
820 associated with active binding motifs of TFs that exhibit an overall enrichment of binding motifs in Duo-
821 specific enhancers (defined in C). See also **Supplementary Figure 6E** for full gene list. (G) Bubble plot
822 showing genes associated with active binding motifs of TFs that exhibit an overall enrichment of binding
823 motifs in Ile-specific enhancers (defined in E). ChRO-seq study: DE, n = 4; Duo spheroid (Duo), n = 3;
824 Ile spheroid (Ile), n = 3. RNA-seq study: DE, n = 3; Duo, n = 6; Ile, n = 4.
825

826 **Figure 7. The HOX cluster dynamics observed in the directed differentiation model of SI is present**
827 **in the early developmental stages of primary human fetal SI tissue.** (A) Heatmap showing the changing
828 transcriptional pattern of HOX genes (ChRO-seq) across the stages of DE, Duo spheroids, and Ile spheroids
829 *in vitro*. (B) Heatmap showing changing steady-state expression pattern of HOX genes (RNA-seq) across
830 stages of DE, Duo spheroids, and Ile spheroid *in vitro*. (C) Heatmap showing changing expression pattern
831 of HOX genes (scRNA-seq) in primary human fetal Duo and Ile tissues at week 11, 14 and 19 of gestation.
832 The expressional level of each HOX gene is the average from the fraction that has both epithelial and
833 mesenchymal cell types (left), mesenchymal cell only (middle), or epithelial cell only (right). (D) Proposed
834 model of transcriptional programming in SI lineage formation and ileal regional patterning during human
835 SI development. The genes activated in the relevant stages and the types of regulatory relationships
836 identified through ChRO-seq analyses are indicated. ChRO-seq study: DE, n = 4; Duo spheroid (Duo), n
837 = 3; Ile spheroid (Ile), n = 3. RNA-seq study: DE, n = 3; Duo, n = 6; Ile, n = 4. scRNA-seq data: matched
838 Duo and Ile tissues from 1 human fetal subject per gestational time point.
839

840 **Supplementary Figure 1. ChRO-seq profiling shows specificity of transcriptional markers in the**
841 **directed differentiation model of human developing SI.** (A) TFs associated with different
842 gastrointestinal organoids that are generated by the hPSC directed differentiation methods[45-51]. (B)

843 Transcriptional activity (ChRO-seq signal) of genes encoding the TFs shown in (A) across different
844 stages. (C) Characterization of genes with distinct patterns of promoter activity across different stages
845 based on the likelihood ratio test (DESeq2). The protein-coding subset of genes (n = 7,651) were
846 clustered into different groups based on the changing patterns of promoter activity. ChRO-seq study:
847 hESC, n = 3; DE, n = 4; Duo spheroid (Duo), n = 3; Ile spheroid (Ile), n = 3. RNA-seq study: hESC, n =
848 2; DE, n = 3; Duo, n = 6; Ile, n = 4. TPM, transcripts per million.
849

850 **Supplementary Figure 2. Steady-state gene expression profiles across all stages in the directed**
851 **differentiation model of the developing human SI.** (A) Hierarchical clustering analysis of RNA-seq
852 expression profiles across stages. Color shade denotes sample to sample distance. (B) PCA of profiles of
853 RNA-seq across stages. (C) RNA-seq expression levels of *SOX2*, *GATA6*, *SOX17*, *CDX2* and *FOXF1*
854 across stages. (D, F, H) Volcano plot showing differentially expressed genes in the indicated comparison.
855 Numbers in red and blue are number of up- and down-regulated genes in the comparison (base mean > 100,
856 $\log_2$ fold change of transcription > 1, padj < 0.2 and p < 0.05 by Wald test; DESeq2). (E, G, I) Pathway
857 enrichment analyses of up- and down-regulated genes in the indicated comparisons (GO term = GO
858 Biological Process 2018). hESC, n= 2; DE, n= 3; Duo spheroid (Duo), n = 6; Ile spheroid (Ile), n = 4. nc,
859 normalized counts.
860

861 **Supplementary Figure 3. Integrative analysis of ChRO-seq and RNA-seq defines marker genes in**
862 **the directed differentiation model of the developing human SI.** (A) RNA-seq expression of genes
863 associated with SI regional identity in spheroids and in HIOs with either Duo or Ile identity. Genes
864 associated with SI regional identity in primary human fetal gut as well as in transplanted HIOs in Tsai *et*
865 *al.*, 2017[21]. Duo and Ile HIOs were generated by culturing Duo and Ile spheroids in 3D Matrigel for 28
866 days and purified by EPCAM+ (epithelial marker) sorting. The HIO samples were from the same batch
867 used for ChRO-seq and RNA-seq of Duo and Ile spheroids in this study. (B-E) Genome wide correlation

868 of transcribed levels (ChRO-seq) and expressed levels (RNA-seq) in the indicated stages. No expression
869 or transcription level thresholds were used for genes to be included in the analysis. ChRO-seq study:
870 hESC, n= 3; DE, n= 4; Duo spheroid (Duo), n = 3; Ile spheroid (Ile), n = 3. RNA-seq study: hESC, n= 2;
871 DE, n= 3; Duo, n = 6; Ile, n = 4; Duo HIO, n = 1; Ile HIO, n = 1.
872

873 **Supplementary Figure 4. Identification of genes associated with stage-specific enhancers relevant to**
874 **the stem cell and definitive endoderm stages.** (A) Venn diagram showing stage-specific and shared
875 enhancers between hESC and DE. hESC-specific enhancers (n = 3906) were defined in the comparison
876 with DE. (B) Cumulative distribution of ChRO-seq fold change in the transcriptional activity of genes
877 grouped into three different categories of hESC-specific enhancer density in hESC vs. DE. (C) Boxplot
878 of ChRO-seq fold change in the transcriptional activity of genes grouped into three different categories of
879 hESC-specific enhancer density in hESC vs. DE. (D) Identification of genes associated with hESC-
880 specific enhancers (n = 239). (E) Bar graph showing genes associated with hESC-specific enhancers (left
881 panel). Top 30 genes based on enhancer density are highlighted (right panel). (F) Venn diagram showing
882 stage-specific and shared enhancers between hESC and DE. DE-specific enhancers (n = 2899) were
883 defined in the comparison with hESC. (G) Cumulative distribution of ChRO-seq fold change in
884 transcriptional activity of genes grouped into three different categories of DE-specific enhancer density in
885 DE vs. hESC. (H) Boxplot of ChRO-seq fold change in transcriptional activity of genes grouped into
886 three different categories of DE-specific enhancer density in DE vs. hESC. (I) Identification of genes
887 associated with DE-specific enhancers (n = 234). (J) Bar graph showing genes associated with DE-
888 specific enhancers (left panel). Top 30 genes based on enhancer density are highlighted (right panel).
889 ChRO-seq study: hESC, n= 3; DE, n= 4. RNA-seq study: hESC, n= 2; DE, n= 3.
890

891 **Supplementary Figure 5. Identification of hESC- and DE-specific enhancer hotspots and associated**
892 **genes.** (A) hESC-specific stitched enhancers are ranked by transcriptional activity (ChRO-seq signal).

893 The stitched enhancers with the highest transcription activity are defined as hESC-specific enhancer
894 hotspots (n = 211; yellow) and the rest are enhancers non-hotspots (n = 2881; black). (B) ChRO-seq fold
895 change in the transcriptional activity of genes associated with hESC-specific stitched enhancers, non-
896 hotspots vs. hotspots (Wilcoxon test). (C) Identification of genes associated with hESC-specific enhancer
897 hotspots. (D) Genes associated with hESC-specific enhancer hotspots (n=59). Relative position between
898 enhancer hotspots and TSSs of the associated genes are shown. Dot size denotes transcriptional activity
899 of a given DE-specific enhancer hotspot. (E) DE-specific stitched enhancers are ranked by transcriptional
900 activity (ChRO-seq signal). The stitched enhancers with the highest transcription activity are defined as
901 DE-specific enhancer hotspots (n = 147; green) and the rest are enhancer non-hotspots (n = 1987; black).
902 (F) ChRO-seq fold change in the transcriptional activity of genes associated with DE-specific stitched
903 enhancers, non-hotspots vs. hotspots (Wilcoxon test). (G) Identification of genes associated with DE-
904 specific enhancer hotspots. (H) Genes associated with Ile-specific enhancer hotspots (n=46). Relative
905 position between enhancer hotspots and TSSs of the associated genes are shown. Dot size denotes
906 transcriptional activity of a given DE-specific enhancer hotspot. ChRO-seq study: hESC, n= 2; DE, n= 4.
907 RNA-seq study: hESC, n=2; DE, n= 3.
908

909 **Supplementary Figure 6. TF binding motif enrichment analysis of TREs specific to hESC or DE.** (A)
910 TF motif enrichment analyses were performed in hESC-specific enhancers (n = 3906) relative to non-hESC-
911 specific TREs (n = 50,777) as well as in hESC-specific enhancer hotspots (n = 211) relative to non-hotspots
912 (n = 2,881). (B) Motifs significantly enriched in hESC-specific enhancers (highlighted by light orange bar)
913 or in hESC-specific enhancer hotspots (highlighted by dark orange bar). The steady-state expression (RNA-
914 seq) of the corresponding TFs are also shown. (C) TF motif enrichment analyses were performed in DE-
915 specific enhancers (n = 2899) relative to non-DE-specific TREs (n = 51,784) as well as in DE-specific
916 enhancer hotspots (n = 147) relative to non-hotspots (n = 1987). (D) Motifs significantly enriched in DE-
917 specific enhancers (highlighted by light green bar) or in DE-specific enhancer hotspots (highlighted by dark

green bar). (E) Bubble plot showing full list of genes associated with active binding motifs of TFs that exhibit an overall enrichment of binding motifs in Duo-specific enhancers (defined in **Figure 6C**). ChRO-seq study: hESC, n= 3; DE, n = 4; Duo spheroid, n = 3.

References

1. Sheaffer KL, Kaestner KH: **Transcriptional networks in liver and intestinal development.** *Cold Spring Harb Perspect Biol* 2012, **4:**a008284.
2. Gao N, White P, Kaestner KH: **Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2.** *Dev Cell* 2009, **16:**588-599.
3. Thompson CA, Wojta K, Pulakanti K, Rao S, Dawson P, Battle MA: **GATA4 Is Sufficient to Establish Jejunal Versus Ileal Identity in the Small Intestine.** *Cell Mol Gastroenterol Hepatol* 2017, **3:**422-446.
4. Thompson CA, DeLaForest A, Battle MA: **Patterning the gastrointestinal epithelium to confer regional-specific functions.** *Dev Biol* 2018, **435:**97-108.
5. Verzi MP, Hatzis P, Sulahian R, Philips J, Schuijers J, Shin H, Freed E, Lynch JP, Dang DT, Brown M, et al: **TCF4 and CDX2, major transcription factors for intestinal function, converge on the same cis-regulatory regions.** *Proc Natl Acad Sci U S A* 2010, **107:**15157-15162.
6. Nowotschin S, Setty M, Kuo YY, Liu V, Garg V, Sharma R, Simon CS, Saiz N, Gardner R, Boutet SC, et al: **The emergent landscape of the mouse gut endoderm at single-cell resolution.** *Nature* 2019, **569:**361-367.
7. Gao S, Yan L, Wang R, Li J, Yong J, Zhou X, Wei Y, Wu X, Wang X, Fan X, et al: **Tracing the temporal-spatial transcriptome landscapes of the human fetal digestive tract using single-cell RNA-sequencing.** *Nat Cell Biol* 2018, **20:**721-734.
8. Kumar N, Tsai YH, Chen L, Zhou A, Banerjee KK, Saxena M, Huang S, Toke NH, Xing J, Shivdasani RA, et al: **The lineage-specific transcription factor CDX2 navigates dynamic chromatin to control distinct stages of intestine development.** *Development* 2019, **146**.
9. Dong J, Hu Y, Fan X, Wu X, Mao Y, Hu B, Guo H, Wen L, Tang F: **Single-cell RNA-seq analysis unveils a prevalent epithelial/mesenchymal hybrid state during mouse organogenesis.** *Genome Biol* 2018, **19:**31.
10. Banerjee KK, Saxena M, Kumar N, Chen L, Cavazza A, Toke NH, O'Neill NK, Madha S, Jadhav U, Verzi MP, Shivdasani RA: **Enhancer, transcriptional, and cell fate plasticity precedes intestinal determination during endoderm development.** *Genes Dev* 2018, **32:**1430-1442.
11. Chen L, Toke NH, Luo S, Vasoya RP, Fullem RL, Parthasarathy A, Perekatt AO, Verzi MP: **A reinforcing HNF4-SMAD4 feed-forward module stabilizes enterocyte identity.** *Nat Genet* 2019, **51:**777-785.
12. Elmentaite R, Alexander Ross, Kylie R. James, Daniel Ortmann, omas Gomes, Kenny Roberts, Komal Nayak, Liz Tuck, Omer Ali Bayraktar, Robert Heuschkel, et al: **Single-cell sequencing of developing human gut reveals transcriptional links to childhood Crohn's disease.** *bioRxiv* 2020.

961 13. Yu Q, Kilik U, Holloway E, Yu-Hwai Tsai, Angeline Wu, Joshua H. Wu, Michael
962 Czerwinski, Charlie Childs, Zhisong He, Ian A. Glass, et al: **An organoid and multi-**
963 **organ developmental cell atlas reveals multilineage fate specification in the human**
964 **intestine.** *bioRxiv* 2020.
965 14. Nord AS, Blow MJ, Attanasio C, Akiyama JA, Holt A, Hosseini R, Phouanenavong S,
966 Plajzer-Frick I, Shoukry M, Afzal V, et al: **Rapid and pervasive changes in genome-**
967 **wide enhancer usage during mammalian development.** *Cell* 2013, **155:**1521-1531.
968 15. Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S,
969 Hawkins RD, Leung D, et al: **Epigenomic analysis of multilineage differentiation of**
970 **human embryonic stem cells.** *Cell* 2013, **153:**1134-1148.
971 16. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK,
972 Stuart RK, Ching CW, et al: **Histone modifications at human enhancers reflect global**
973 **cell-type-specific gene expression.** *Nature* 2009, **459:**108-112.
974 17. Long HK, Prescott SL, Wysocka J: **Ever-Changing Landscapes: Transcriptional**
975 **Enhancers in Development and Evolution.** *Cell* 2016, **167:**1170-1187.
976 18. Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptewicz M,
977 Barbara-Haley K, Kuersten S, et al: **Widespread transcription at neuronal activity-**
978 **regulated enhancers.** *Nature* 2010, **465:**182-187.
979 19. Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, Spicuglia S, de la
980 Chapelle AL, Heidemann M, Hintermair C, et al: **Transcription initiation platforms**
981 **and GTF recruitment at tissue-specific enhancers and promoters.** *Nat Struct Mol Biol*
982 2011, **18:**956-963.
983 20. Chu T, Rice EJ, Booth GT, Salamanca HH, Wang Z, Core LJ, Longo SL, Corona RJ,
984 Chin LS, Lis JT, et al: **Chromatin run-on and sequencing maps the transcriptional**
985 **regulatory landscape of glioblastoma multiforme.** *Nat Genet* 2018, **50:**1553-1564.
986 21. Core LJ, Waterfall JJ, Lis JT: **Nascent RNA sequencing reveals widespread pausing**
987 **and divergent initiation at human promoters.** *Science* 2008, **322:**1845-1848.
988 22. Kwak H, Fuda NJ, Core LJ, Lis JT: **Precise maps of RNA polymerase reveal how**
989 **promoters direct initiation and pausing.** *Science* 2013, **339:**950-953.
990 23. Dinh TA, Sritharan R, Smith FD, Francisco AB, Ma RK, Bunaciu RP, Kanke M, Danko
991 CG, Massa AP, Scott JD, Sethupathy P: **Hotspots of Aberrant Enhancer Activity in**
992 **Fibrolamellar Carcinoma Reveal Candidate Oncogenic Pathways and Therapeutic**
993 **Vulnerabilities.** *Cell Rep* 2020, **31:**107509.
994 24. Wells JM, Spence JR: **How to make an intestine.** *Development* 2014, **141:**752-760.
995 25. Sinagoga KL, Wells JM: **Generating human intestinal tissues from pluripotent stem**
996 **cells to study development and disease.** *EMBO J* 2015, **34:**1149-1163.
997 26. Spence JR, Mayhew CN, Rankin SA, Kuhar MF, Vallance JE, Tolle K, Hoskins EE,
998 Kalinichenko VV, Wells SI, Zorn AM, et al: **Directed differentiation of human**
999 **pluripotent stem cells into intestinal tissue in vitro.** *Nature* 2011, **470:**105-109.
1000 27. McCracken KW, Howell JC, Wells JM, Spence JR: **Generating human intestinal tissue**
1001 **from pluripotent stem cells in vitro.** *Nat Protoc* 2011, **6:**1920-1928.
1002 28. Tsai YH, Nattiv R, Dedhia PH, Nagy MS, Chin AM, Thomson M, Klein OD, Spence JR:
1003 **In vitro patterning of pluripotent stem cell-derived intestine recapitulates in vivo**
1004 **human development.** *Development* 2017, **144:**1045-1055.

1005 29. Watson CL, Mahe MM, Munera J, Howell JC, Sundaram N, Poling HM, Schweitzer JI,
1006 Vallance JE, Mayhew CN, Sun Y, et al: **An in vivo model of human small intestine**
1007 **using pluripotent stem cells.** *Nat Med* 2014, **20:**1310-1314.
1008 30. Hill DR, Huang S, Nagy MS, Yadagiri VK, Fields C, Mukherjee D, Bons B, Dedhia PH,
1009 Chin AM, Tsai YH, et al: **Bacterial colonization stimulates a complex physiological**
1010 **response in the immature human intestinal epithelium.** *Elife* 2017, **6**.
1011 31. Finkbeiner SR, Hill DR, Altheim CH, Dedhia PH, Taylor MJ, Tsai YH, Chin AM, Mahe
1012 MM, Watson CL, Freeman JJ, et al: **Transcriptome-wide Analysis Reveals Hallmarks**
1013 **of Human Intestine Development and Maturation In Vitro and In Vivo.** *Stem Cell*
1014 *Reports* 2015.
1015 32. Wang D, Wang J, Bai L, Pan H, Feng H, Clevers H, Zeng YA: **Long-Term Expansion**
1016 **of Pancreatic Islet Organoids from Resident Procr(+) Progenitors.** *Cell* 2020,
1017 **180:**1198-1211 e1119.
1018 33. Chu T, Wang Z, Chou SP, Danko CG: **Discovering Transcriptional Regulatory**
1019 **Elements From Run-On and Sequencing Data Using the Web-Based dREG**
1020 **Gateway.** *Curr Protoc Bioinformatics* 2019, **66:**e70.
1021 34. Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y,
1022 Zhao X, Schmidl C, Suzuki T, et al: **An atlas of active enhancers across human cell**
1023 **types and tissues.** *Nature* 2014, **507:**455-461.
1024 35. Oz-Levi D, Olender T, Bar-Joseph I, Zhu Y, Marek-Yagel D, Barozzi I, Osterwalder M,
1025 Alkelai A, Ruzzo EK, Han Y, et al: **Noncoding deletions reveal a gene that is critical**
1026 **for intestinal function.** *Nature* 2019, **571:**107-111.
1027 36. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI,
1028 Young RA: **Master transcription factors and mediator establish super-enhancers at**
1029 **key cell identity genes.** *Cell* 2013, **153:**307-319.
1030 37. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, Hoke HA, Young RA:
1031 **Super-enhancers in the control of cell identity and disease.** *Cell* 2013, **155:**934-947.
1032 38. Boyd M, Thodberg M, Vitezic M, Bornholdt J, Vitting-Seerup K, Chen Y, Coskun M, Li
1033 Y, Lo BZS, Klausen P, et al: **Characterization of the enhancer and promoter**
1034 **landscape of inflammatory bowel disease from human colon biopsies.** *Nat Commun*
1035 2018, **9:**1661.
1036 39. Barros R, da Costa LT, Pinto-de-Sousa J, Duluc I, Freund JN, David L, Almeida R:
1037 **CDX2 autoregulation in human intestinal metaplasia of the stomach: impact on the**
1038 **stability of the phenotype.** *Gut* 2011, **60:**290-298.
1039 40. Czerwinski M, Holloway E, Tsai Y-H, Angeline Wu, Qianhui Yu, Josh Wu, Katherine D.
1040 Walton, Caden Sweet, Charlie Childs, Ian Glass, et al: **In vitro and in vivo development**
1041 **of the human intestinal niche at single cell resolution.** *bioRxiv* 2020.
1042 41. Holloway E, Joshua H. Wu, Michael Czerwinkski, Caden W. Sweet, Angeline Wu, Yu-
1043 Hwai Tsai, Sha Huang, Amy E. Stoddard, Meghan M. Capeling, Ian Glass, Spence JR:
1044 **Differentiation of human intestinal organoids with endogenous vascular endothelial**
1045 **cells.** *bioRxiv* 2020.
1046 42. Kageyama R, Ohtsuka T, Kobayashi T: **The Hes gene family: repressors and**
1047 **oscillators that orchestrate embryogenesis.** *Development* 2007, **134:**1243-1251.
1048 43. Kohlhase J, Wischermann A, Reichenbach H, Froster U, Engel W: **Mutations in the**
1049 **SALL1 putative transcription factor gene cause Townes-Brocks syndrome.** *Nat*
1050 *Genet* 1998, **18:**81-83.

1051 44. Kiefer SM, Ohlemiller KK, Yang J, McDill BW, Kohlhase J, Rauchman M: **Expression**
1052 **of a truncated Sall1 transcriptional repressor is responsible for Townes-Brocks**
1053 **syndrome birth defects.** *Hum Mol Genet* 2003, **12:**2221-2227.
1054 45. Kennedy MW, Chalamalasetty RB, Thomas S, Garriock RJ, Jailwala P, Yamaguchi TP:
1055 **Sp5 and Sp8 recruit beta-catenin and Tcf1-Lef1 to select enhancers to activate Wnt**
1056 **target gene transcription.** *Proc Natl Acad Sci U S A* 2016, **113:**3545-3550.
1057 46. Sato A, Kishida S, Tanaka T, Kikuchi A, Kodama T, Asashima M, Nishinakamura R:
1058 **Sall1, a causative gene for Townes-Brocks syndrome, enhances the canonical Wnt**
1059 **signaling by localizing to heterochromatin.** *Biochem Biophys Res Commun* 2004,
1060 **319:**103-113.
1061 47. Villano JL, Katz FN: **four-jointed is required for intermediate growth in the**
1062 **proximal-distal axis in Drosophila.** *Development* 1995, **121:**2767-2777.
1063 48. Zeidler MP, Perrimon N, Strutt DI: **The four-jointed gene is required in the**
1064 **Drosophila eye for ommatidial polarity specification.** *Curr Biol* 1999, **9:**1363-1372.
1065 49. Rock R, Heinrich AC, Schumacher N, Gessler M: **Fjx1: a notch-inducible secreted**
1066 **ligand with specific binding sites in developing mouse embryos and adult brain.** *Dev*
1067 *Dyn* 2005, **234:**602-612.
1068 50. Simoes-Costa M, Stone M, Bronner ME: **Axud1 Integrates Wnt Signaling and**
1069 **Transcriptional Inputs to Drive Neural Crest Formation.** *Dev Cell* 2015, **34:**544-554.
1070 51. Kawazoe Y, Sekimoto T, Araki M, Takagi K, Araki K, Yamamura K: **Region-specific**
1071 **gastrointestinal Hox code during murine embryonal gut development.** *Dev Growth*
1072 *Differ* 2002, **44:**77-84.
1073 52. Sekimoto T, Yoshinobu K, Yoshida M, Kuratani S, Fujimoto S, Araki M, Tajima N,
1074 Araki K, Yamamura K: **Region-specific expression of murine Hox genes implies the**
1075 **Hox code-mediated patterning of the digestive tract.** *Genes Cells* 1998, **3:**51-64.
1076 53. Ye DZ, Kaestner KH: **Foxa1 and Foxa2 control the differentiation of goblet and**
1077 **enteroendocrine L- and D-cells in mice.** *Gastroenterology* 2009, **137:**2052-2062.
1078 54. Kerschner JL, Gosalia N, Leir SH, Harris A: **Chromatin remodeling mediated by the**
1079 **FOXA1/A2 transcription factors activates CFTR expression in intestinal epithelial**
1080 **cells.** *Epigenetics* 2014, **9:**557-565.
1081 55. Wu YZ, Chan KYY, Leung KT, Lam HS, Tam YH, Lee KH, Li K, Ng PC:
1082 **Dysregulation of miR-431 and target gene FOXA1 in intestinal tissues of infants**
1083 **with necrotizing enterocolitis.** *FASEB J* 2019, **33:**5143-5152.
1084 56. Matsuyama M, Aizawa S, Shimono A: **Sfrp controls apicobasal polarity and oriented**
1085 **cell division in developing gut epithelium.** *PLoS Genet* 2009, **5:**e1000427.
1086 57. Allison TF, Smith AJH, Anastassiadis K, Sloane-Stanley J, Biga V, Stavish D, Hackland
1087 J, Sabri S, Langerman J, Jones M, et al: **Identification and Single-Cell Functional**
1088 **Characterization of an Endodermally Biased Pluripotent Substate in Human**
1089 **Embryonic Stem Cells.** *Stem Cell Reports* 2018, **10:**1895-1907.
1090 58. Chu LF, Leng N, Zhang J, Hou Z, Mamott D, Vereide DT, Choi J, Kendziorski C,
1091 Stewart R, Thomson JA: **Single-cell RNA-seq reveals novel regulators of human**
1092 **embryonic stem cell differentiation to definitive endoderm.** *Genome Biol* 2016,
1093 **17:**173.
1094 59. Tiyaboonchai A, Cardenas-Diaz FL, Ying L, Maguire JA, Sim X, Jobaliya C, Gagne AL,
1095 Kishore S, Stanescu DE, Hughes N, et al: **GATA6 Plays an Important Role in the**

1096 **Induction of Human Definitive Endoderm, Development of the Pancreas, and**
1097 **Functionality of Pancreatic beta Cells.** *Stem Cell Reports* 2017, **8:**589-604.
1098 60. Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y, Liechti A,
1099 Ascencao K, Rummel C, Ovchinnikova S, et al: **Gene expression across mammalian**
1100 **organ development.** *Nature* 2019, **571:**505-509.
1101 61. Finkbeiner SR, Freeman JJ, Wieck MM, El-Nachef W, Altheim CH, Tsai YH, Huang S,
1102 Dyal R, White ES, Grikscheit TC, et al: **Generation of tissue-engineered small**
1103 **intestine using embryonic stem cell-derived human intestinal organoids.** *Biol Open*
1104 2015, **4:**1462-1472.
1105 62. Dame MK, Attili D, McClintock SD, Dedhia PH, Ouillette P, Hardt O, Chin AM, Xue X,
1106 Laliberte J, Katz EL, et al: **Identification, isolation and characterization of human**
1107 **LGR5-positive colon adenoma cells.** *Development* 2018, **145**.
1108 63. Wuarin J, Schibler U: **Physical isolation of nascent RNA chains transcribed by RNA**
1109 **polymerase II: evidence for cotranscriptional splicing.** *Mol Cell Biol* 1994, **14:**7219-
1110 7225.
1111 64. Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, Waters CT, Munson
1112 K, Core LJ, Lis JT: **Base-pair-resolution genome-wide mapping of active RNA**
1113 **polymerases using precision nuclear run-on (PRO-seq).** *Nat Protoc* 2016, **11:**1455-
1114 1476.
1115 65. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M,
1116 Gingeras TR: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics* 2013,
1117 **29:**15-21.
1118 66. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C: **Salmon provides fast and bias-**
1119 **aware quantification of transcript expression.** *Nat Methods* 2017, **14:**417-419.
1120 67. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion**
1121 **for RNA-seq data with DESeq2.** *Genome Biol* 2014, **15:**550.
1122 68. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S,
1123 Jenkins SL, Jagodnik KM, Lachmann A, et al: **Enrichr: a comprehensive gene set**
1124 **enrichment analysis web server 2016 update.** *Nucleic Acids Res* 2016, **44:**W90-97.
1125 69. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh
1126 H, Glass CK: **Simple combinations of lineage-determining transcription factors**
1127 **prime cis-regulatory elements required for macrophage and B cell identities.** *Mol*
1128 *Cell* 2010, **38:**576-589.
1129 70. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R: **Integrating single-cell**
1130 **transcriptomic data across different conditions, technologies, and species.** *Nat*
1131 *Biotechnol* 2018, **36:**411-420.
1132
1133
1134

Figure 1.

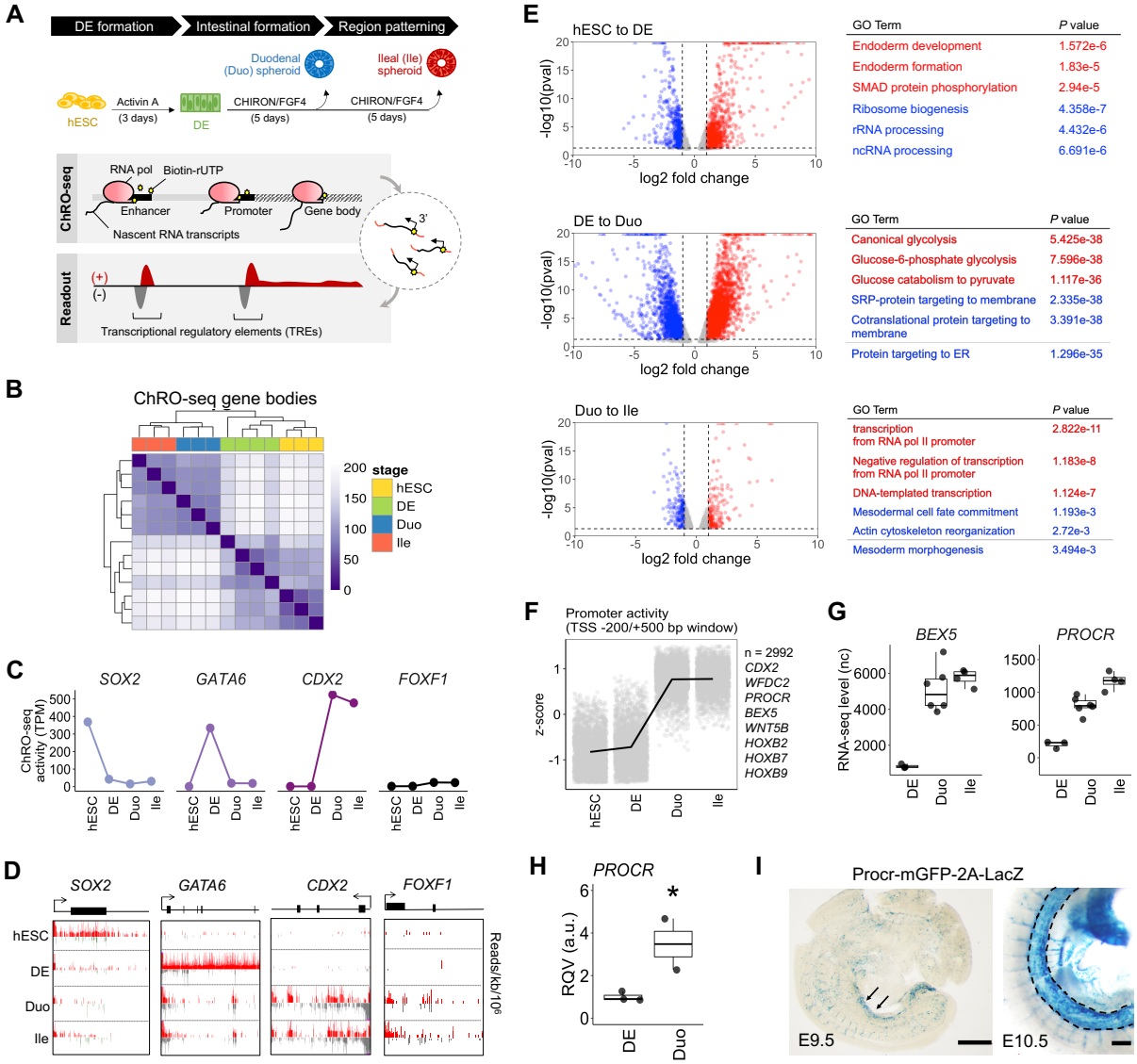

A
DE formation
Intestinal formation
Region patterning
Duodenal (Duo) spheroid
Ileal (Ile) spheroid
hESC
Activin A (3 days)
DE
CHIRON/FGF4 (5 days)
CHIRON/FGF4 (5 days)
ChRO-seq
RNA pol
Biotin-rUTP
Enhancer
Promoter
Gene body
Nascent RNA transcripts
3'
Readout
(+)
(-)
Transcriptional regulatory elements (TREs)
B
ChRO-seq gene bodies
stage
hESC
DE
Duo
Ile
C
SOX2
GATA6
CDX2
FOXF1
ChRO-seq activity (TPM)
D
Reads/kb/10^6
E
hESC to DE
DE to Duo
Duo to Ile
-log10(pval)
log2 fold change
GO Term
P value
Endoderm development 1.572e-6
Endoderm formation 1.83e-5
SMAD protein phosphorylation 2.94e-5
Ribosome biogenesis 4.358e-7
rRNA processing 4.432e-6
ncRNA processing 6.691e-6
Canonical glycolysis 5.425e-38
Glucose-6-phosphate glycolysis 7.596e-38
Glucose catabolism to pyruvate 1.117e-36
SRP-protein targeting to membrane 2.335e-38
Cotranslational protein targeting to membrane 3.391e-38
Protein targeting to ER 1.296e-35
transcription from RNA pol II promoter 2.822e-11
Negative regulation of transcription from RNA pol II promoter 1.183e-8
DNA-templated transcription 1.124e-7
Mesodermal cell fate commitment 1.193e-3
Actin cytoskeleton reorganization 2.72e-3
Mesoderm morphogenesis 3.494e-3
F
Promoter activity (TSS -200/+500 bp window)
z-score
n = 2992
CDX2
WFDC2
PROCR
BEX5
WNT5B
HOXB2
HOXB7
HOXB9
G
BEX5
PROCR
RNA-seq level (nc)
H
PROCR
RQV (a.u.)
I
Procr-mGFP-2A-LacZ
E9.5
E10.5

Figure 2.

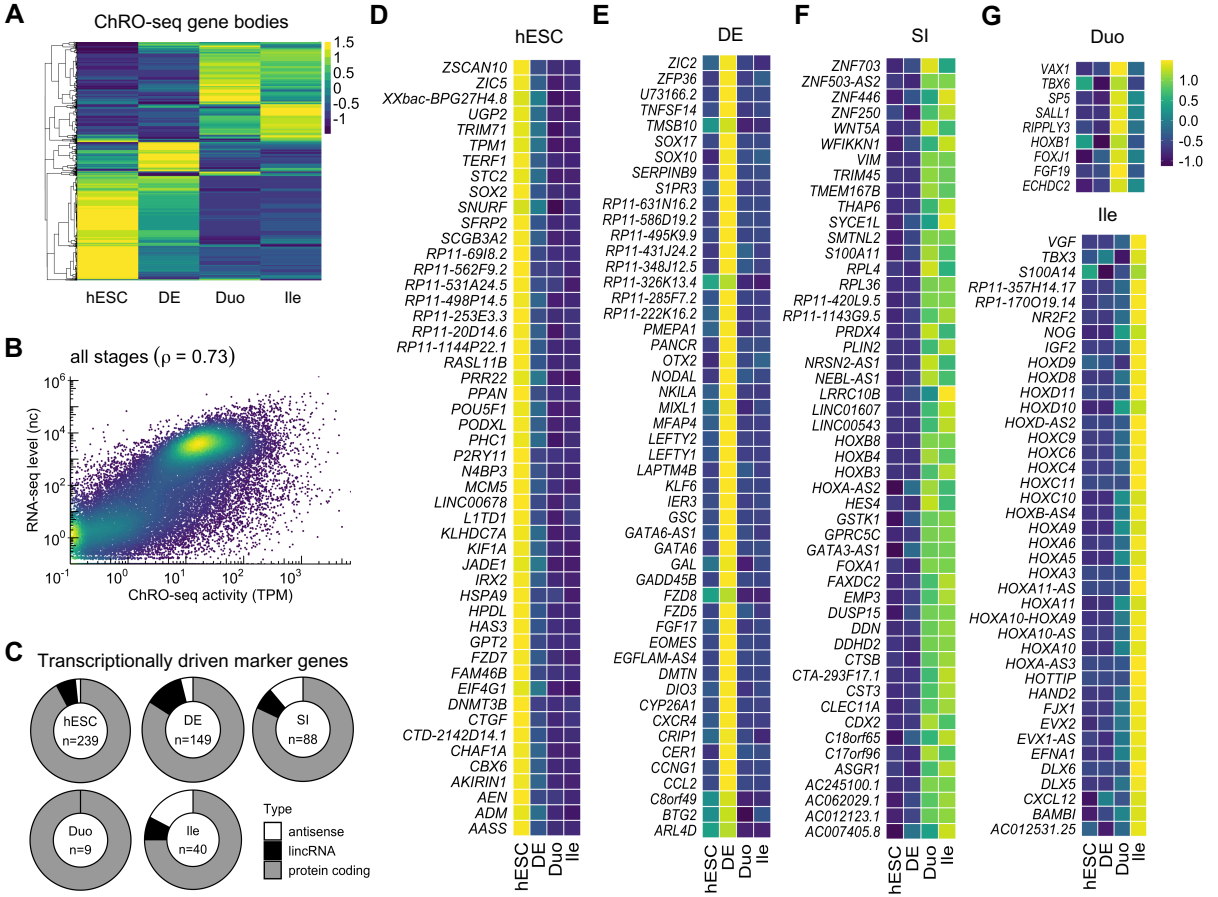

A
ChRO-seq gene bodies
1.5
1
0.5
0
-0.5
-1
hESC
DE
Duo
Ile
B
all stages (ρ = 0.73)
RNA-seq level (nc)
ChRO-seq activity (TPM)
C
Transcriptionally driven marker genes
hESC n=239
DE n=149
SI n=88
Duo n=9
Ile n=40
Type
antisense
lincRNA
protein coding
D
hESC
ZSCAN10
ZIC5
XXbac-BPG27H4.8
UGP2
TRIM71
TPM1
TERF1
STC2
SOX2
SNURF
SFRP2
SCGB3A2
RP11-69I8.2
RP11-562F9.2
RP11-531A24.5
RP11-498P14.5
RP11-253E3.3
RP11-20D14.6
RP11-1144P22.1
RASL11B
PRR22
PPAN
POU5F1
PODXL
PHC1
P2RY11
N4BP3
MCM5
LINC00678
L1TD1
KLHDC7A
KIF1A
JADE1
IRX2
HSPA9
HPDL
HAS3
GPT2
FZD7
FAM46B
EIF4G1
DNMT3B
CTGF
CTD-2142D14.1
CHAF1A
CBX6
AKIRIN1
AEN
ADM
AASS
E
DE
ZIC2
ZFP36
U73166.2
TNFSF14
TMSB10
SOX17
SOX10
SERPINB9
S1PR3
RP11-631N16.2
RP11-586D19.2
RP11-495K9.9
RP11-431J24.2
RP11-348J12.5
RP11-326K13.4
RP11-285F7.2
RP11-222K16.2
PMEPA1
PANCR
OTX2
NODAL
NKILA
MIXL1
MFAP4
LEFTY2
LEFTY1
LAPTM4B
KLF6
IER3
GSC
GATA6-AS1
GATA6
GAL
GADD45B
FZD8
FZD5
FGF17
EOMES
EGFLAM-AS4
DMTN
DIO3
CYP26A1
CXCR4
CRIP1
CER1
CCNG1
CCL2
C8orf49
BTG2
ARL4D
F
SI
ZNF703
ZNF503-AS2
ZNF446
ZNF250
WNT5A
WFIKKN1
VIM
TRIM45
TMEM167B
THAP6
SYCE1L
SMTNL2
S100A11
RPL4
RPL36
RP11-420L9.5
RP11-1143G9.5
PRDX4
PLIN2
NRSN2-AS1
NEBL-AS1
LRRC10B
LINC01607
LINC00543
HOXB8
HOXB4
HOXB3
HOXA-AS2
HES4
GSTK1
GPRC5C
GATA3-AS1
FOXA1
FAXDC2
EMP3
DUSP15
DDN
DDHD2
CTSB
CTA-293F17.1
CST3
CLEC11A
CDX2
C18orf65
C17orf96
ASGR1
AC245100.1
AC062029.1
AC012123.1
AC007405.8
G
Duo
VAX1
TBX6
SP5
SALL1
RIPPLY3
HOXB1
FOXJ1
FGF19
ECHDC2
1.0
0.5
0.0
-0.5
-1.0
Ile
VGF
TBX3
S100A14
RP11-357H14.17
RP1-170O19.14
NR2F2
NOG
IGF2
HOXD9
HOXD8
HOXD11
HOXD10
HOXD-AS2
HOXC9
HOXC6
HOXC4
HOXC11
HOXC10
HOXB-AS4
HOXA9
HOXA6
HOXA5
HOXA3
HOXA11-AS
HOXA11
HOXA10-HOXA9
HOXA10-AS
HOXA10
HOXA-AS3
HOTTIP
HAND2
FJX1
EVX2
EVX1-AS
EFNA1
DLX6
DLX5
CXCL12
BAMBI
AC012531.25
hESC
DE
Duo
Ile

Figure 3.

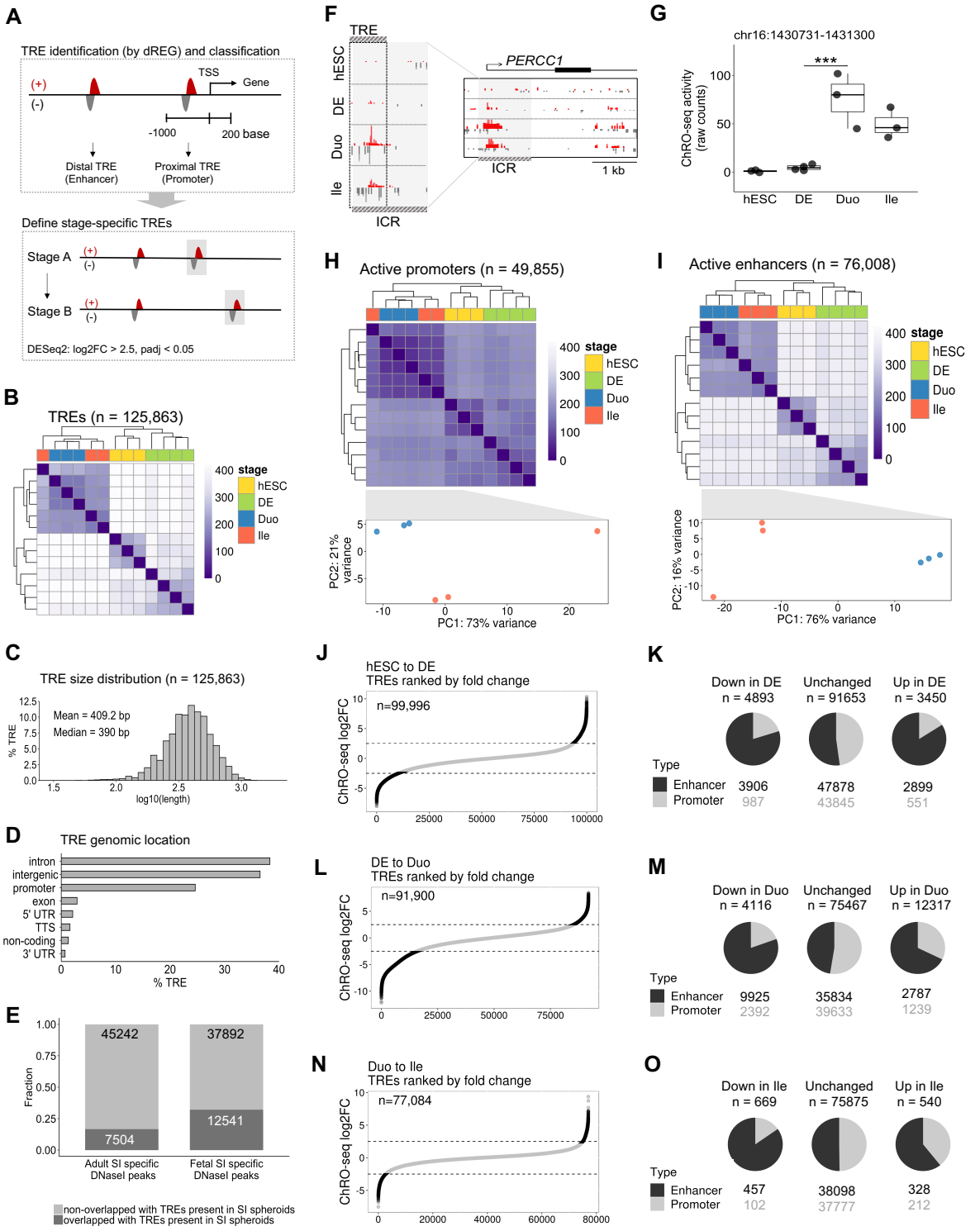

A
TRE identification (by dREG) and classification
TSS
Gene
(+)
(-)
-1000
200 base
Distal TRE (Enhancer)
Proximal TRE (Promoter)
Define stage-specific TREs
Stage A
Stage B
DESeq2: log2FC > 2.5, padj < 0.05
B
TREs (n = 125,863)
stage
hESC
DE
Duo
Ile
C
TRE size distribution (n = 125,863)
Mean = 409.2 bp
Median = 390 bp
% TRE
log10(length)
D
TRE genomic location
intron
intergenic
promoter
exon
5' UTR
TTS
non-coding
3' UTR
% TRE
E
Fraction
45242
7504
37892
12541
Adult SI specific DNaseI peaks
Fetal SI specific DNaseI peaks
non-overlapped with TREs present in SI spheroids
overlapped with TREs present in SI spheroids
F
TRE
hESC
DE
Duo
Ile
ICR
PERCC1
1 kb
G
chr16:1430731-1431300
ChRO-seq activity (raw counts)
***
hESC
DE
Duo
Ile
H
Active promoters (n = 49,855)
PC2: 21% variance
PC1: 73% variance
I
Active enhancers (n = 76,008)
PC2: 16% variance
PC1: 76% variance
J
hESC to DE
TREs ranked by fold change
n=99,996
ChRO-seq log2FC
K
Down in DE n = 4893
Unchanged n = 91653
Up in DE n = 3450
Type
Enhancer 3906 47878 2899
Promoter 987 43845 551
L
DE to Duo
TREs ranked by fold change
n=91,900
ChRO-seq log2FC
M
Down in Duo n = 4116
Unchanged n = 75467
Up in Duo n = 12317
Type
Enhancer 9925 35834 2787
Promoter 2392 39633 1239
N
Duo to Ile
TREs ranked by fold change
n=77,084
ChRO-seq log2FC
O
Down in Ile n = 669
Unchanged n = 75875
Up in Ile n = 540
Type
Enhancer 457 38098 328
Promoter 102 37777 212

Figure 4.

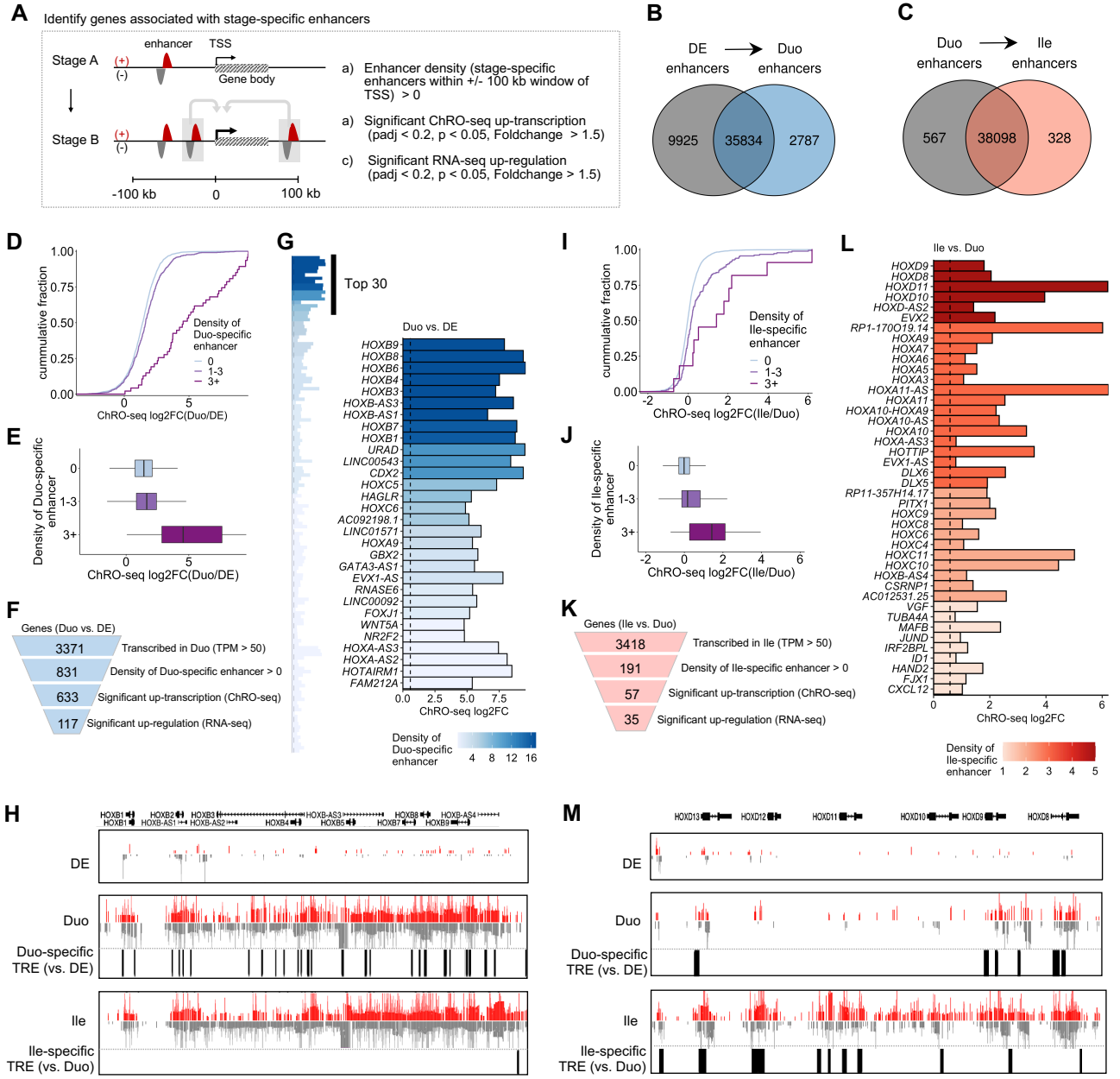

A
Identify genes associated with stage-specific enhancers
enhancer
TSS
Stage A
Gene body
Stage B
-100 kb
0
100 kb
a) Enhancer density (stage-specific enhancers within +/- 100 kb window of TSS) > 0
a) Significant ChRO-seq up-transcription (padj < 0.2, p < 0.05, Foldchange > 1.5)
c) Significant RNA-seq up-regulation (padj < 0.2, p < 0.05, Foldchange > 1.5)
B
DE enhancers
Duo enhancers
9925
35834
2787
C
Duo enhancers
Ile enhancers
567
38098
328
D
cummulative fraction
Density of Duo-specific enhancer
ChRO-seq log2FC(Duo/DE)
E
Density of Duo-specific enhancer
ChRO-seq log2FC(Duo/DE)
F
Genes (Duo vs. DE)
3371 Transcribed in Duo (TPM > 50)
831 Density of Duo-specific enhancer > 0
633 Significant up-transcription (ChRO-seq)
117 Significant up-regulation (RNA-seq)
G
Top 30
Duo vs. DE
HOXB9
HOXB8
HOXB6
HOXB4
HOXB3
HOXB-AS3
HOXB-AS1
HOXB7
HOXB1
URAD
LINC00543
CDX2
HOXC5
HAGLR
HOXC6
AC092198.1
LINC01571
HOXA9
GBX2
GATA3-AS1
EVX1-AS
RNASE6
LINC00092
FOXJ1
WNT5A
NR2F2
HOXA-AS3
HOXA-AS2
HOTAIRM1
FAM212A
ChRO-seq log2FC
Density of Duo-specific enhancer
I
cummulative fraction
Density of Ile-specific enhancer
ChRO-seq log2FC(Ile/Duo)
J
Density of Ile-specific enhancer
ChRO-seq log2FC(Ile/Duo)
K
Genes (Ile vs. Duo)
3418 Transcribed in Ile (TPM > 50)
191 Density of Ile-specific enhancer > 0
57 Significant up-transcription (ChRO-seq)
35 Significant up-regulation (RNA-seq)
L
Ile vs. Duo
HOXD9
HOXD8
HOXD11
HOXD10
HOXD-AS2
EVX2
RP1-170O19.14
HOXA9
HOXA7
HOXA6
HOXA5
HOXA3
HOXA11-AS
HOXA11
HOXA10-HOXA9
HOXA10-AS
HOXA10
HOXA-AS3
HOTTIP
EVX1-AS
DLX6
DLX5
RP11-357H14.17
PITX1
HOXC9
HOXC8
HOXC6
HOXC4
HOXC11
HOXC10
HOXB-AS4
CSRNP1
AC012531.25
VGF
TUBA4A
MAFB
JUND
IRF2BPL
ID1
HAND2
FJX1
CXCL12
ChRO-seq log2FC
Density of Ile-specific enhancer
H
HOXB1
HOXB2
HOXB3
HOXB-AS1
HOXB-AS2
HOXB4
HOXB5
HOXB-AS3
HOXB7
HOXB8
HOXB9
HOXB-AS4
DE
Duo
Duo-specific TRE (vs. DE)
Ile
Ile-specific TRE (vs. Duo)
M
HOXD13
HOXD12
HOXD11
HOXD10
HOXD9
HOXD8
DE
Duo
Duo-specific TRE (vs. DE)
Ile
Ile-specific TRE (vs. Duo)

Figure 5.

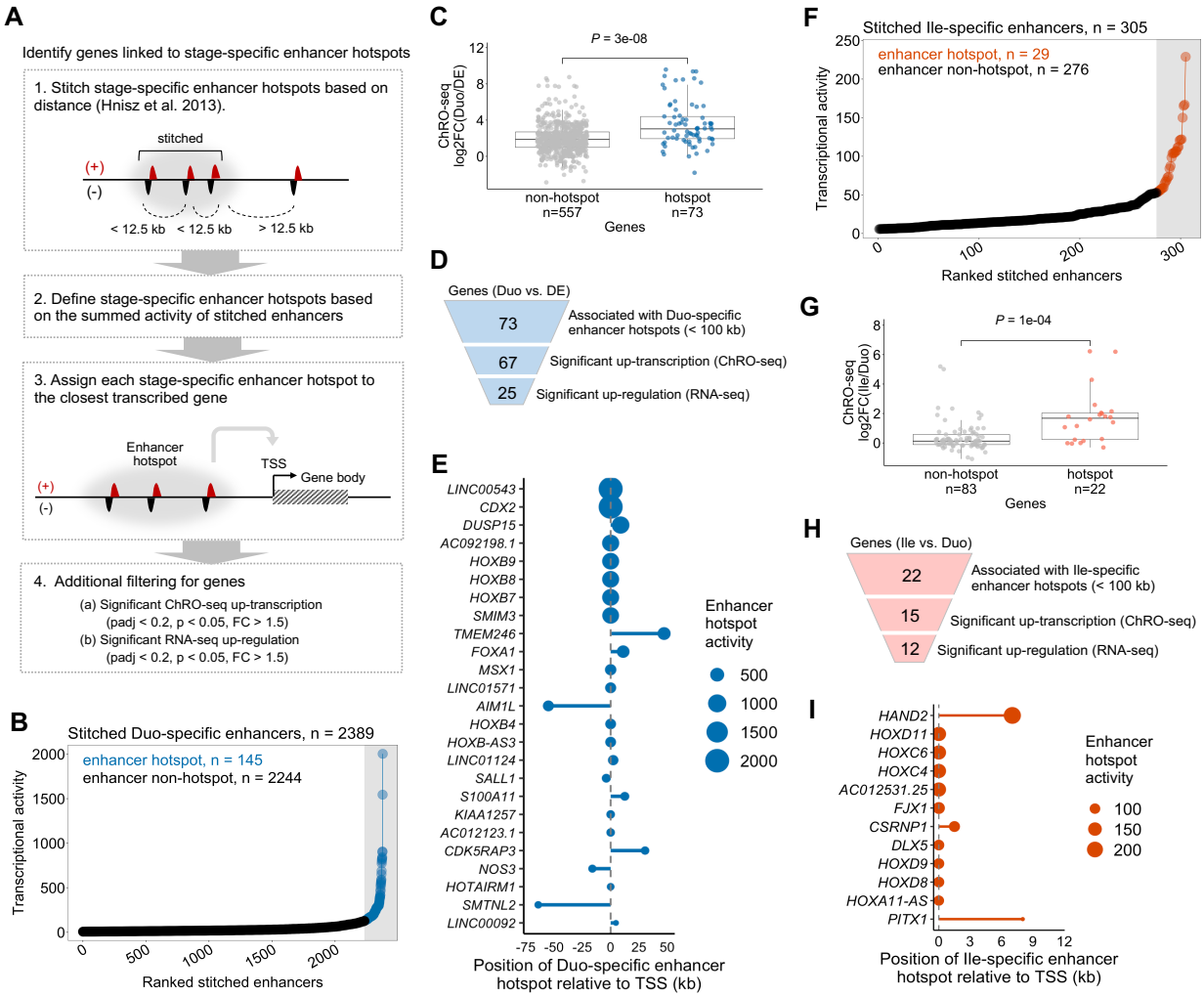

A
Identify genes linked to stage-specific enhancer hotspots
1. Stitch stage-specific enhancer hotspots based on distance (Hnisz et al. 2013).
stitched
(+)
(-)
< 12.5 kb
< 12.5 kb
> 12.5 kb
2. Define stage-specific enhancer hotspots based on the summed activity of stitched enhancers
3. Assign each stage-specific enhancer hotspot to the closest transcribed gene
Enhancer hotspot
TSS
Gene body
4. Additional filtering for genes
(a) Significant ChRO-seq up-transcription
(padj < 0.2, p < 0.05, FC > 1.5)
(b) Significant RNA-seq up-regulation
(padj < 0.2, p < 0.05, FC > 1.5)
B
Stitched Duo-specific enhancers, n = 2389
enhancer hotspot, n = 145
enhancer non-hotspot, n = 2244
Transcriptional activity
Ranked stitched enhancers
C
P = 3e-08
ChRO-seq log2FC(Duo/DE)
non-hotspot n=557
hotspot n=73
Genes
D
Genes (Duo vs. DE)
73 Associated with Duo-specific enhancer hotspots (< 100 kb)
67 Significant up-transcription (ChRO-seq)
25 Significant up-regulation (RNA-seq)
E
LINC00543
CDX2
DUSP15
AC092198.1
HOXB9
HOXB8
HOXB7
SMIM3
TMEM246
FOXA1
MSX1
LINC01571
AIM1L
HOXB4
HOXB-AS3
LINC01124
SALL1
S100A11
KIAA1257
AC012123.1
CDK5RAP3
NOS3
HOTAIRM1
SMTNL2
LINC00092
Enhancer hotspot activity
500
1000
1500
2000
Position of Duo-specific enhancer hotspot relative to TSS (kb)
F
Stitched Ile-specific enhancers, n = 305
enhancer hotspot, n = 29
enhancer non-hotspot, n = 276
Transcriptional activity
Ranked stitched enhancers
G
P = 1e-04
ChRO-seq log2FC(Ile/Duo)
non-hotspot n=83
hotspot n=22
Genes
H
Genes (Ile vs. Duo)
22 Associated with Ile-specific enhancer hotspots (< 100 kb)
15 Significant up-transcription (ChRO-seq)
12 Significant up-regulation (RNA-seq)
I
HAND2
HOXD11
HOXC6
HOXC4
AC012531.25
FJX1
CSRNP1
DLX5
HOXD9
HOXD8
HOXA11-AS
PITX1
Enhancer hotspot activity
100
150
200
Position of Ile-specific enhancer hotspot relative to TSS (kb)

Figure 6.

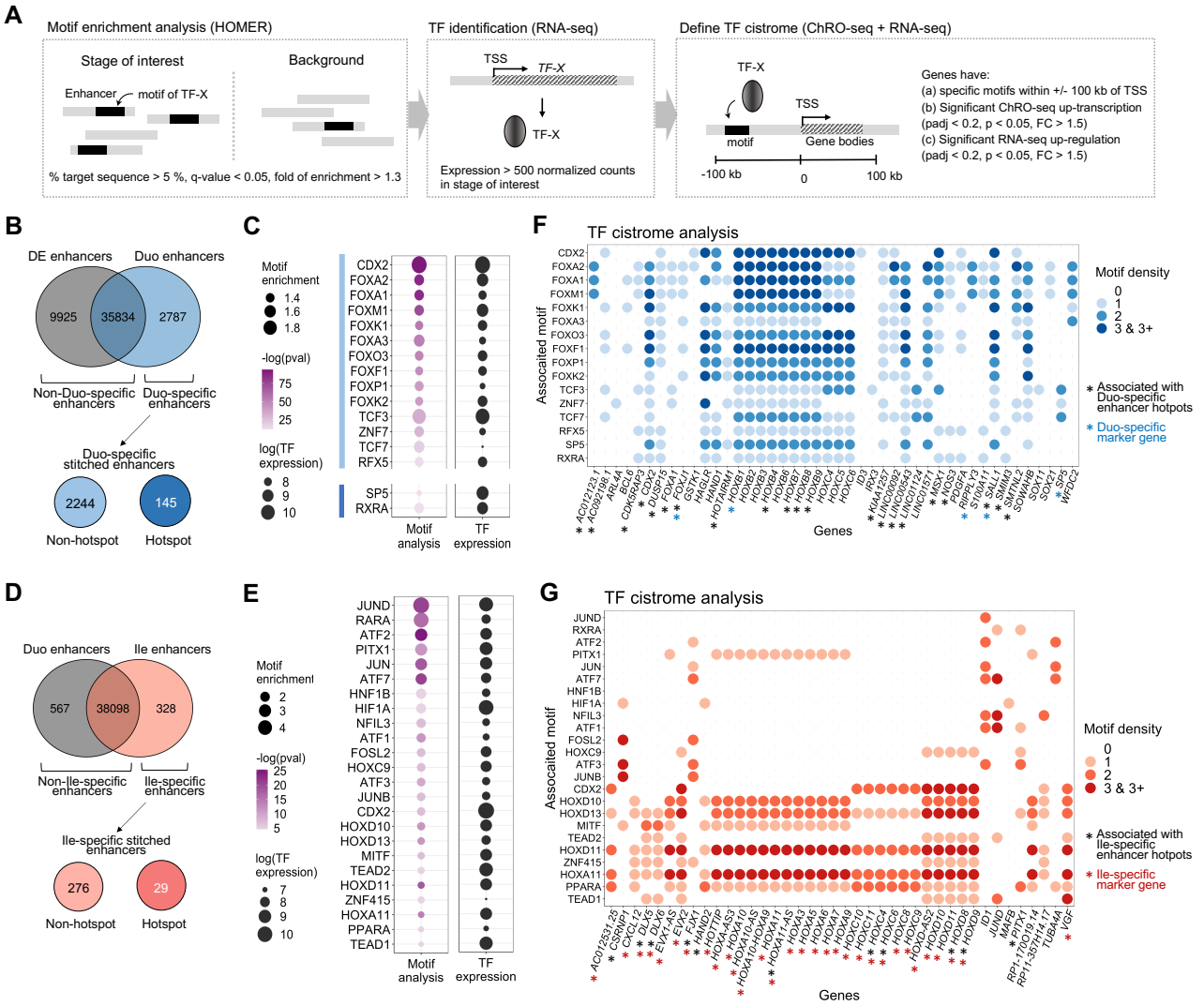
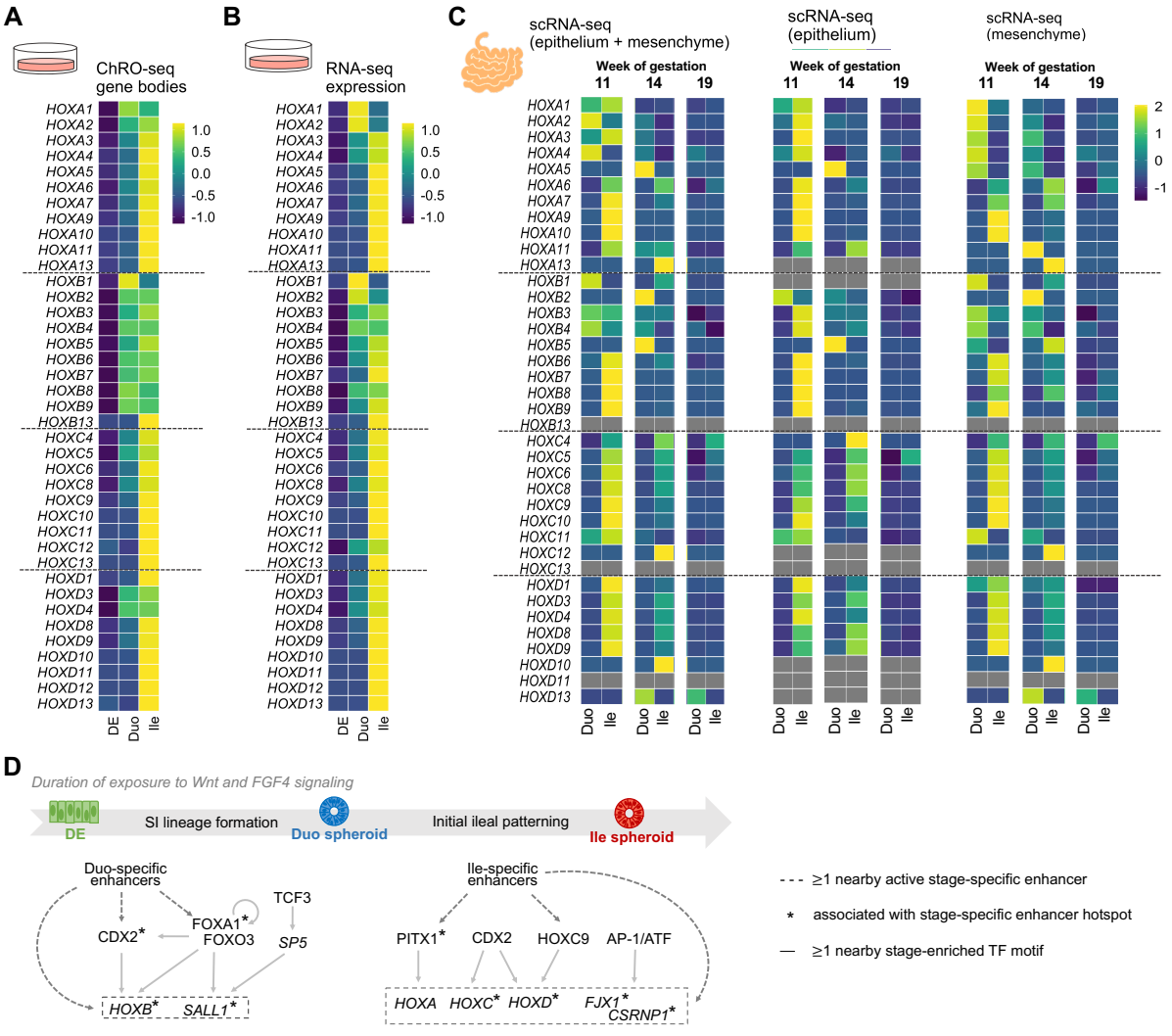
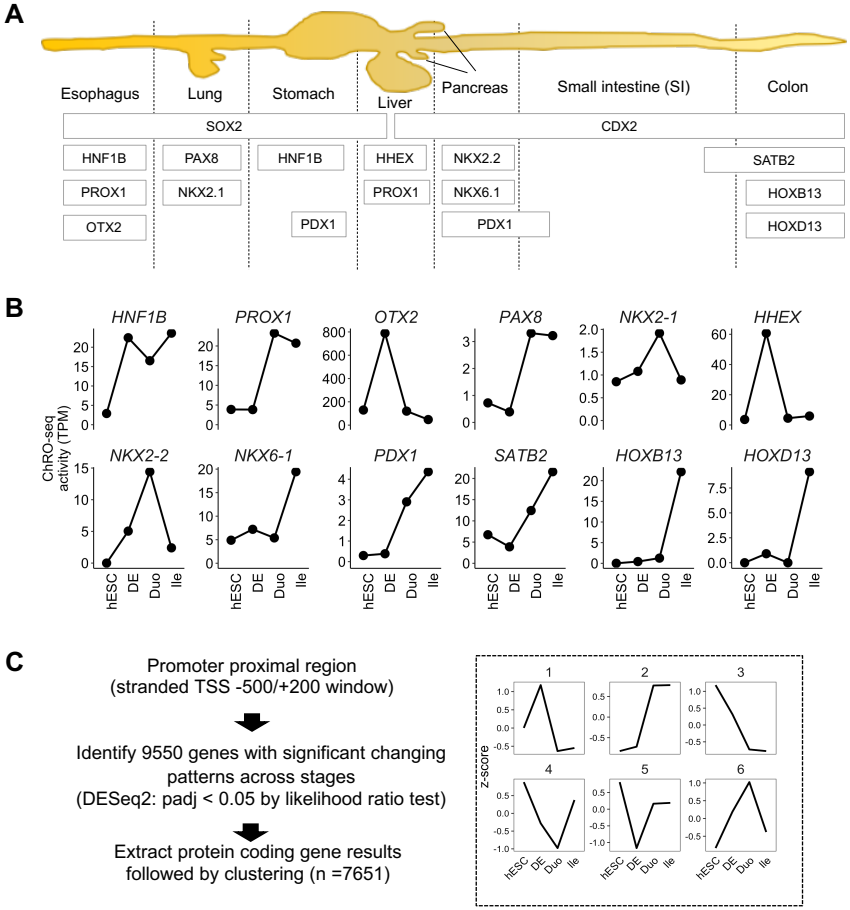

A
Motif enrichment analysis (HOMER)
Stage of interest
Background
Enhancer
motif of TF-X
% target sequence > 5 %, q-value < 0.05, fold of enrichment > 1.3
TF identification (RNA-seq)
TSS
TF-X
Expression > 500 normalized counts in stage of interest
Define TF cistrome (ChRO-seq + RNA-seq)
motif
Gene bodies
-100 kb
0
100 kb
Genes have:
(a) specific motifs within +/- 100 kb of TSS
(b) Significant ChRO-seq up-transcription (padj < 0.2, p < 0.05, FC > 1.5)
(c) Significant RNA-seq up-regulation (padj < 0.2, p < 0.05, FC > 1.5)
B
DE enhancers
Duo enhancers
9925
35834
2787
Non-Duo-specific enhancers
Duo-specific enhancers
Duo-specific stitched enhancers
2244
145
Non-hotspot
Hotspot
C
Motif enrichment
1.4
1.6
1.8
-log(pval)
75
50
25
log(TF expression)
8
9
10
CDX2
FOXA2
FOXA1
FOXM1
FOXK1
FOXA3
FOXO3
FOXF1
FOXP1
FOXK2
TCF3
ZNF7
TCF7
RFX5
SP5
RXRA
Motif analysis
TF expression
F
TF cistrome analysis
Assocaited motif
Motif density
0
1
2
3 & 3+
Associated with Duo-specific enhancer hotpots
Duo-specific marker gene
Genes
D
Duo enhancers
Ile enhancers
567
38098
328
Non-Ile-specific enhancers
Ile-specific enhancers
Ile-specific stitched enhancers
276
29
Non-hotspot
Hotspot
E
Motif enrichmenl
2
3
4
-log(pval)
25
20
15
10
5
log(TF expression)
7
8
9
10
JUND
RARA
ATF2
PITX1
JUN
ATF7
HNF1B
HIF1A
NFIL3
ATF1
FOSL2
HOXC9
ATF3
JUNB
CDX2
HOXD10
HOXD13
MITF
TEAD2
HOXD11
ZNF415
HOXA11
PPARA
TEAD1
Motif analysis
TF expression
G
TF cistrome analysis
Assocaited motif
Motif density
0
1
2
3 & 3+
Associated with Ile-specific enhancer hotpots
Ile-specific marker gene
Genes

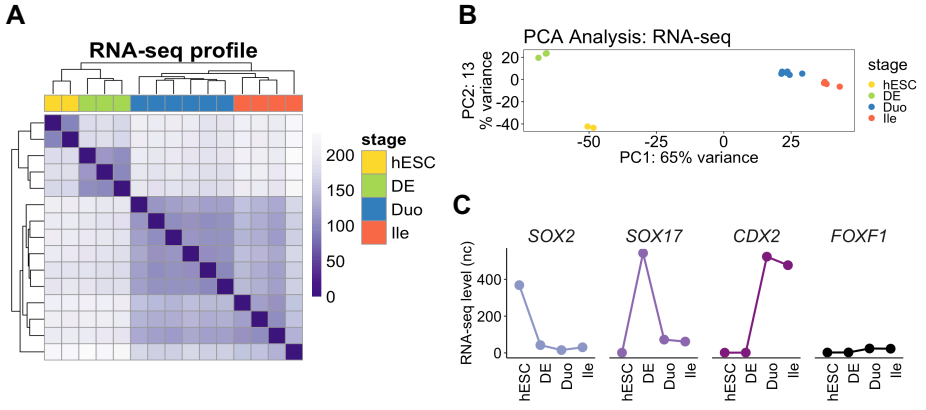
Figure 7.

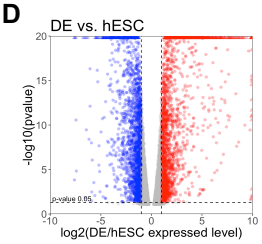

A
ChRO-seq
gene bodies
B
RNA-seq
expression
C
scRNA-seq
(epithelium + mesenchyme)
scRNA-seq
(epithelium)
scRNA-seq
(mesenchyme)
Week of gestation
11
14
19
HOXA1
HOXA2
HOXA3
HOXA4
HOXA5
HOXA6
HOXA7
HOXA9
HOXA10
HOXA11
HOXA13
HOXB1
HOXB2
HOXB3
HOXB4
HOXB5
HOXB6
HOXB7
HOXB8
HOXB9
HOXB13
HOXC4
HOXC5
HOXC6
HOXC8
HOXC9
HOXC10
HOXC11
HOXC12
HOXC13
HOXD1
HOXD3
HOXD4
HOXD8
HOXD9
HOXD10
HOXD11
HOXD12
HOXD13
DE
Duo
Ile
D
Duration of exposure to Wnt and FGF4 signaling
DE
SI lineage formation
Duo spheroid
Initial ileal patterning
Ile spheroid
Duo-specific enhancers
CDX2*
FOXA1*
FOXO3
TCF3
SP5
HOXB*
SALL1*
Ile-specific enhancers
PITX1*
CDX2
HOXC9
AP-1/ATF
HOXA
HOXC*
HOXD*
FJX1*
CSRNP1*
≥1 nearby active stage-specific enhancer
* associated with stage-specific enhancer hotspot
≥1 nearby stage-enriched TF motif

Supplementary Figure 1.



Supplementary Figure 2.

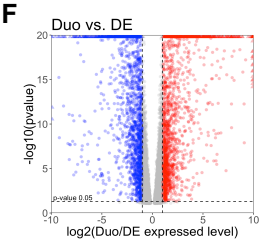




E

GO Term of down-regulated genes in DE vs. hESC	*P* value
Ribosome biogenesis	1.128e-9
rRNA processing	1.448e-8
Carboxylic acid transport	1.804e-8
rRNA metabolic process	6.399e-7
MAPK cascade	9.675e-7

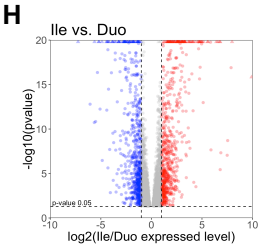
GO Term of up-regulated genes in DE vs. hESC	*P* value
Endoderm formation	1.435e-8
Axon guidance	2.444e-5
ECM organization	5.785e-5
Endoderm development	7.338e-5
Negative chemotaxis	7.637e-5



G

GO Term of down-regulated genes in Duo vs. DE	*P* value
Axon guidance	4.775e-8
Axonogenesis	3.226e-6
Apoptosis regulation	4.935e-6
Endoderm formation	1.035e-5
MAPK cascade	6.039e-5

GO Term of up-regulated genes in Duo vs. DE	*P* value
ECM organization	2.958e-18
Kidney development	8.395e-12
Skeletal system development	1.951e-11
Renal system development	5.335e-10
Ureteric bud development	9.528e-10

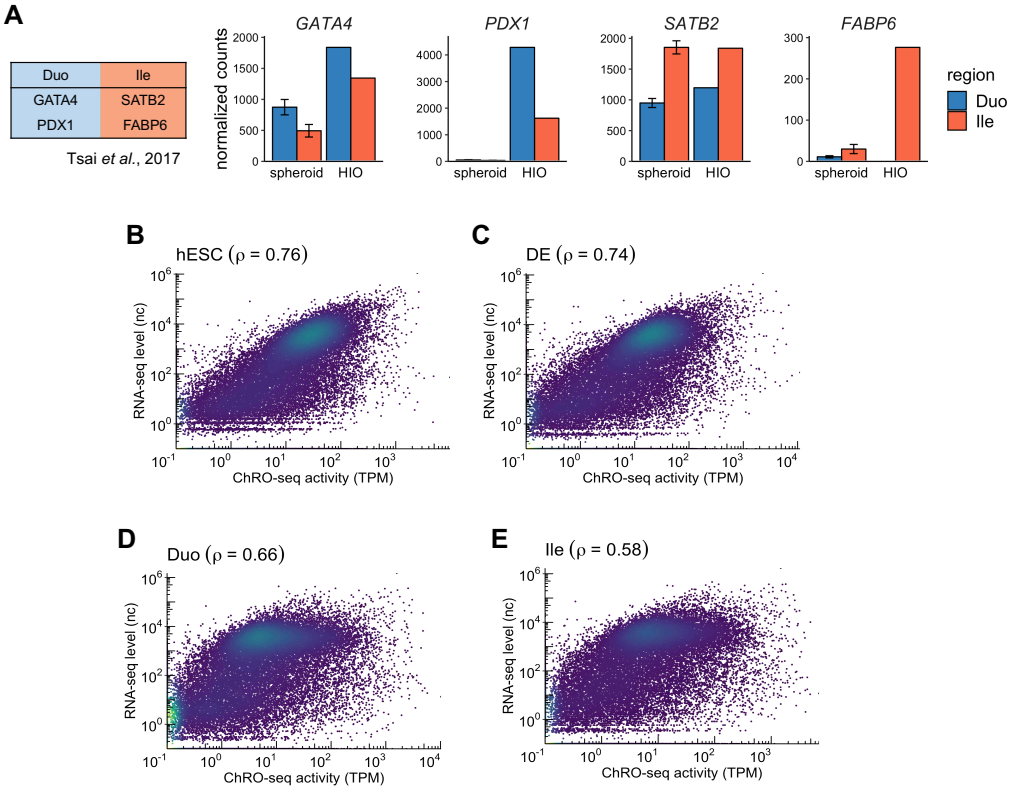


I

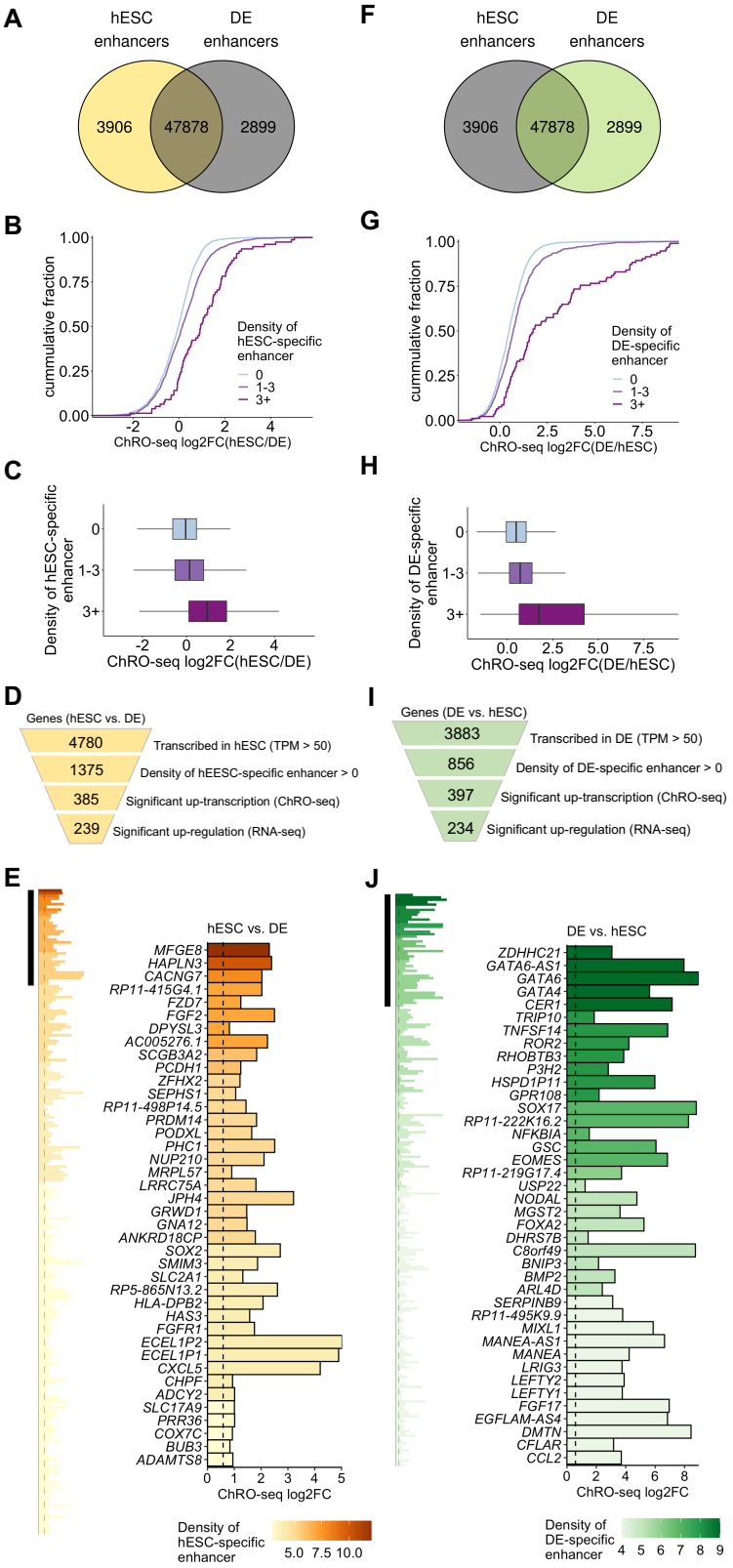
GO Term of down-regulated genes in Ile vs. Duo	*P* value
Heart development	1.876e-9
Mesoderm development	1.297e-7
Kidney development	6.208e-7
Circulatory system development	1.831e-6
Renal system development	2.304e-6

GO Term of up-regulated genes in Ile vs. Duo	*P* value
Positive regulation of transcription from RNA pol II promoter	1.042e-12
Positive regulation of DNA-templated transcription	1.388e-12
Regulation of transcription from RNA pol II promoter	1.385e-11
Cell proliferation	1.894e-11
Cell differentiation	2.646e-11

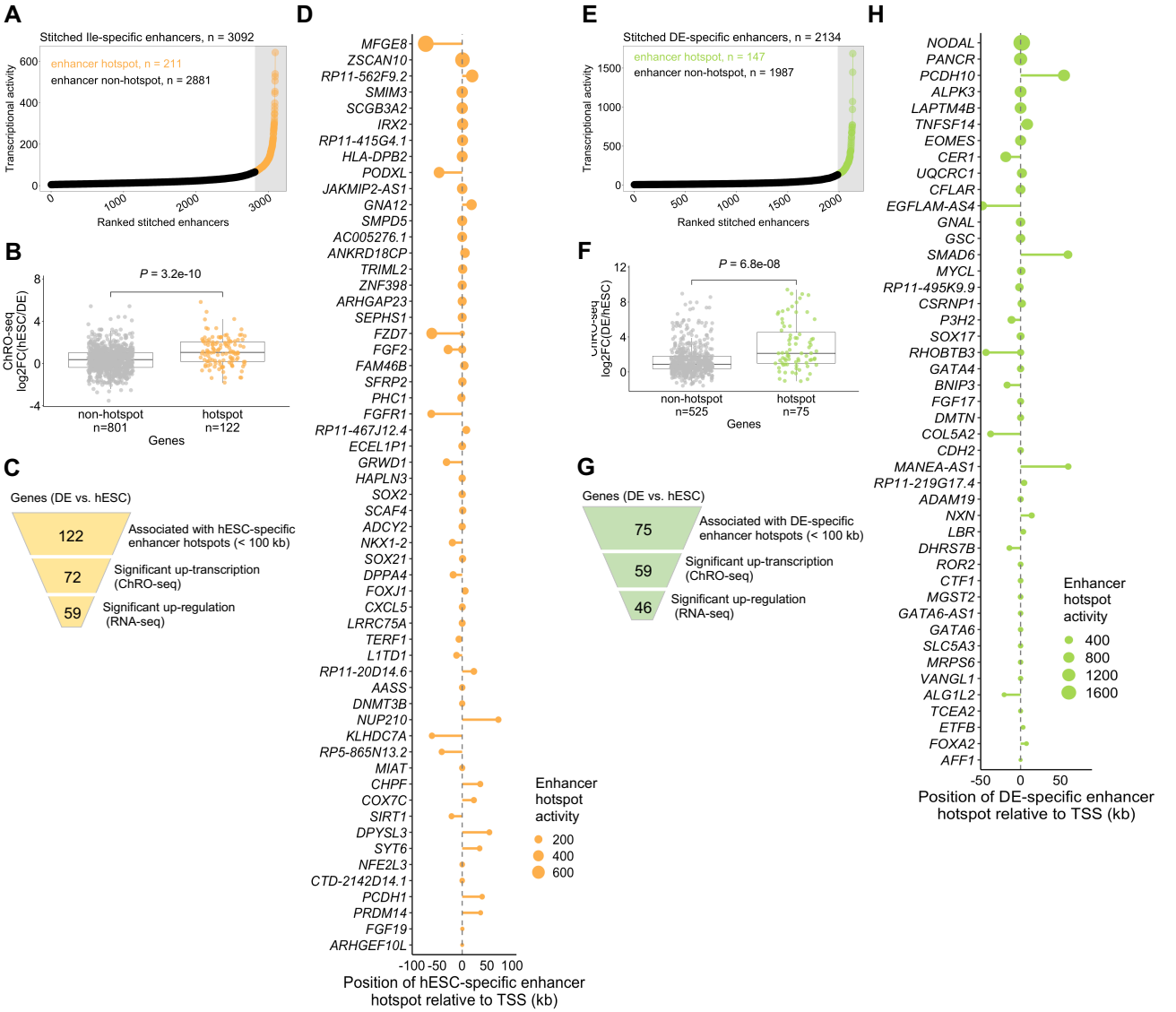
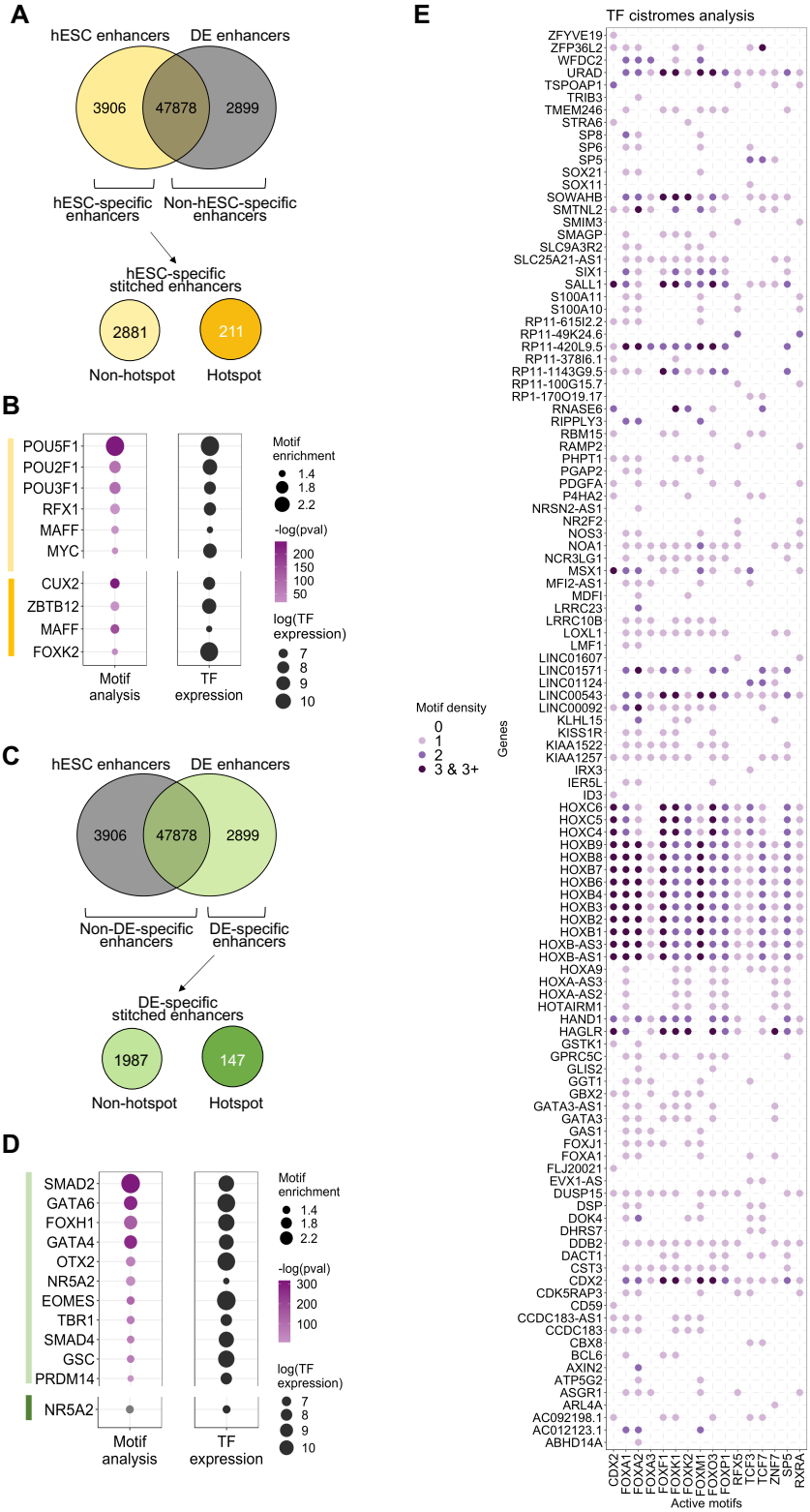
Supplementary Figure 3.



Supplementary Figure 4.



Supplementary Figure 5.



Supplementary Figure 6.