3D Chromatin Architecture Remodeling during Human Cardiomyocyte Differentiation Reveals A Role Of HERV-H In Demarcating Chromatin Domains

Authors:

Yanxiao Zhang[1*], Ting Li[2*], Sebastian Preissl[1,3*], Jonathan Grinstein[2], Elie N. Farah[2], Eugin Destici[2], Ah Young Lee[1], Sora Chee[1], Yunjiang Qiu[1], Kaiyue Ma[1], Zhen Ye[1], Rong Hu[1], Quan Zhu[1], Hui Huang[1], Rongxin Fang[1], Leqian Yu[4,5], Juan Carlos Izpisua Belmonte[6], Jun Wu[4,5], Sylvia M. Evans[2,7], Neil C. Chi[2,8,10], Bing Ren[1,3,8,9,10]

*these authors contributed equally

Affiliations:

[1]Ludwig Institute for Cancer Research, La Jolla, CA 92093, USA.

[2]Department of Medicine, Division of Cardiology, University of California, San Diego, La Jolla, California 92093, USA.

[3]Department of Cellular and Molecular Medicine, Center for Epigenomics, University of California, San Diego, School of Medicine, La Jolla, CA 92093, USA.

[4]Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

[5]Hamon Center for Regenerative Science and Medicine, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA

[6]The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, San Diego, CA 92186-5800, USA.

[7]Department of Pharmacology and Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, CA 92093, USA.

[8]School of Medicine, Institute of Genomic Medicine, University of California, San Diego, La Jolla, CA 92093, USA

[9]School of Medicine, Moores Cancer Center, University of California, San Diego, La Jolla, CA 92093, USA.
[10]Correspondence: Neil Chi (nchi@ucsd.edu) and Bing Ren (biren@ucsd.edu)

Abstract

Dynamic restructuring of chromatin architecture has been implicated in cell-type specific gene regulatory programs; yet, how chromatin remodels during lineage specification remains to be elucidated. Through interrogating chromatin reorganization during human cardiomyocyte differentiation, we uncover dynamic chromatin interactions between genes and distal regulatory elements harboring noncoding variants associated with adult and congenital heart diseases. Unexpectedly, we also discover a new class of human pluripotent stem cell (PSC)-specific topologically associating domains (TAD) that are created by the actively transcribed endogenous retrotransposon HERV-H. Deletion or silencing of specific HERV-H elements eliminates corresponding TAD boundaries, while *de novo* insertion of HERV-H can introduce new chromatin domain boundaries in human PSCs. Furthermore, comparative analysis of chromatin architecture in other species that lack HERV-H sequences supports a role for actively transcribed HERV-H in demarcating human PSC-specific TADs. The biological role of HERV-H is further underscored by the observation that deletion of a specific HERV-H reduces transcription of genes upstream and facilitates cell differentiation. Overall, our results highlight a previously unrecognized role for retrotransposons in restructuring genome architecture in the human genome and delineate dynamic gene regulatory networks during cardiomyocyte development that inform how non-coding genetic variants contribute to human heart diseases.

MAIN TEXT

The three-dimensional organization of chromosomes enables long-range communications between enhancers and promoters that are critical for building complex gene regulatory networks in multicellular species[1,2]. In somatic cells, interphase chromosomes occupy separate nuclear spaces known as chromosome territories[3]. Each chromosome is organized into a dynamic but non-random hierarchical structure characterized by stretches of transcriptionally active, megabase-long compartments that are interspersed with stretches of transcriptionally inactive compartments[4]. These compartments can be further partitioned into topologically associating domains (TADs), which exhibit high levels of intra-domain interactions and relatively low levels of inter-domain interactions[5-7]. The boundaries of TADs are generally conserved between cell types and closely related species[5,8], and can restrict *cis*-regulatory element interactions with target promoters[9]; therefore when TADs are disrupted, it can lead to significant change of transcriptional landscape and cause diseases[10-13]. At or within TAD boundaries, long-range interactions such as chromatin loops are frequently observed to link regulatory elements, including binding sites for the CCCTC-binding factor (CTCF), promoters and enhancers[14,15].

There is mounting evidence that a large fraction of TADs and chromatin loops are shaped by cohesin mediated loop extrusion[16-18], with TAD borders constrained by the CTCF[19,20]. However, not all TAD boundaries contain CTCF binding sites, and many CTCF binding sites do not localize on or near TAD boundaries[5], suggesting that their underlying relationship is more complex and raising the question that additional mechanisms for TAD formation may also exist. On the other hand, CTCF binding sites are dynamically evolving and underlie the evolution of genome architecture[21]. Interestingly, a subset of species-specific CTCF binding sequences are rapidly dispersed by transposable elements, especially SINE elements[22]. This leads to an important question that has not been thoroughly investigated: how much a role do transposable elements play in reshaping the genome architecture during evolution?

Genome architecture is not only reshaped during evolution, but also reconfigured in a

87 cell-type specific manner during stem cell differentiation and somatic cell
88 reprogramming[8,23-26]. Failure of proper restructuring of chromatin has potentially
89 detrimental effects and might lead to developmental defects[10]. One of the most common
90 birth defects is congenital heart disease (CHD) which impacts 1% of newborns[27].
91 Mutations in chromatin modifiers, remodelers as well as transcription factors have been
92 linked to CHD[28]. In addition, it was reported that cardiomyocyte-specific deletion of
93 CTCF was sufficient to induce heart failure in mice[29], highlighting the importance of
94 chromatin architecture in cardiac development. Two recent studies showed promoter
95 centric chromatin interaction maps of in vitro differentiated cardiomyocytes[30,31].
96 However, these studies did not capture the dynamics during the differentiation process.
97
98 Here, using high-resolution *in situ* high-throughput chromosome conformation capture
99 Hi-C (*in situ* Hi-C)[14,15], we interrogated the reorganization of hPSC chromatin
100 architecture during key developmental stages of human cardiomyocyte differentiation.
101 We report the finding of stage-specific long-range chromatin interactions that are
102 associated with temporal regulation of gene expression and predict candidate genes for
103 non-coding DNA variants associated with cardiac traits/diseases. Moreover, we reveal
104 a class of primate-specific endogenous retrotransposon, HERV-H, that is involved in
105 establishing boundaries of TADs *de novo* in hPSCs during primate evolution.
106 Importantly, we show that HERV-H demarcates TAD boundaries, which are dependent
107 on its transcription, as targeted silencing of HERV-H can weaken or eliminate the
108 insulation of TAD borders. Finally, comparative analysis of chromatin architecture of
109 PSC in different primate and non-primate species, as well as creation of new hPSC
110 TADs via *de novo* insertions of HERV-H further underscore this conclusion. Taken
111 together, our study emphasizes an important role for ERV family retrotransposon
112 elements in shaping chromatin organization in specific cell types, and provides a
113 valuable resource for the study of human cardiac development and disease.
114

115 **RESULTS**

116 **Dynamic chromatin organization during cardiomyocyte differentiation**

117 To study the dynamics of genome organization during cardiomyocyte development, we

118 utilized a transgenic human embryonic stem cell (hESC) H9 *MYL2:H2B-GFP* ventricular
119 cardiomyocyte reporter line, which can be used to purify ventricular cardiomyocytes[32]
120 (**Extended Data Fig. 1a**). We generated and collected samples in biological duplicates
121 at the following six critical time points of cardiomyocyte differentiation: hESCs (Day 0),
122 mesodermal cells (Day 2), cardiac mesodermal cells (Day 5), cardiac progenitors (Day
123 7), primitive cardiomyocytes (Day 15) and ventricular cardiomyocytes (Day 80) (**Fig.
124 1a**). Flow cytometry confirmed that differentiation efficiency for D2 – D15 time point
125 samples was at least ~80-90%, and *MYL2:H2B-GFP*+ ventricular cardiomyocytes were
126 sorted at day 80 (**Extended Data Fig. 1b**). *In situ* Hi-C[4] experiments were performed,
127 and on average three billion raw read pairs and one billion unique long-range cis
128 contacts (minimum distance 10 kb, **Extended Data Table 1**) were generated for each
129 time point. Close correlation of Hi-C maps between biological replicates verified that
130 these maps were highly reproducible at multiple scales (**Extended Data Fig. 2)**.
131 Complementing these Hi-C contact maps, ATAC-seq, ChIP-seq and RNA-seq
132 experiments were performed on these samples to interrogate chromatin accessibility,
133 identify CTCF occupancy sites and key histone modifications, and profile
134 transcriptomes, respectively (**Extended Data Fig. 3**).
135
136 Analyses of these genome-wide datasets revealed extensive reorganization of
137 chromatin architecture at all structural levels during hESC differentiation, which was
138 also accompanied by corresponding epigenomic changes (**Fig. 1a, b, c**). Across the
139 entire genome, the abundance of short-range chromatin interactions decreased while
140 the number of long-range chromatin interactions (above 500 kilobases) increased
141 (**Extended Data Fig. 4a**). Identification of active (A) and inactive (B) compartments at
142 all stages using principal component analysis showed that 81.6% of the genome
143 maintained the same compartment state throughout cardiomyocyte differentiation (**Fig.
144 1d**). Among the 18.4% of the genome that switched compartments, a similar proportion
145 of the genome switched from A to B and from B to A at each stage transition (**Extended
146 Data Fig. 4b**), and the majority (73%) only switched once, suggesting linear nuclear
147 architectural changes during lineage commitment (**Fig. 1d**). Hierarchical clustering of
148 genomic bins that switched compartments showed stage specific dynamics (**Fig. 1d**),

149 which positively correlated with gene expression (**Extended Data Fig. 4c**) as reported
150 previously[8]. For instance, gene loci of pluripotency factors including *SOX2* moved to
151 the inactive compartment when hESCs differentiated into mesodermal cells, whereas
152 genes involved in cardiomyocyte specification such as the transcription factor *HAND2*
153 and the cardiomyocyte-specific calcium channel *RYR2* displayed B to A transitions at
154 day 5 and day 15, respectively (**Fig. 1d; Extended Data Fig. 4d**).
155

156 Transcription of HERV-H is associated with TAD boundary formation in 157 human ES cells

158 Consistent with recent reports describing loss of TADs during the differentiation of
159 embryonic stem cells[23,33], we also observed that the number of TADs decreased as
160 hESCs differentiated into ventricular cardiomyocytes using multiple TAD-calling
161 algorithms[5,14,34] (**Extended Data Fig. 5a, b, c**). Clustering of stage-specific TAD
162 boundaries using Directionality Index (DI) delta scores, which estimates the strength of
163 the TAD boundaries, further showed that the majority of stage-specific TAD boundaries
164 were lost or weakened during differentiation and the effect was most pronounced at
165 early mesoderm specification (D0-D2), and ventricular cardiomyocyte
166 differentiation/maturation (D15-D80) (**Fig. 1e**). To investigate the mechanism of TAD
167 dynamics during cardiomyocyte differentiation, we examined the sequence features of
168 TAD boundaries that were lost in the course of differentiation with a particular focus on
169 D0 hESC-specific TAD boundaries [ESC(+)] and the boundaries lost during D15-D80
170 ventricular cardiomyocyte transition [vCM(-)] (**Extended Data Table 2**). While CTCF
171 has been shown to play a critical role in defining TAD boundaries[5,14,20], we observed no
172 significant difference in CTCF occupancy across all stages examined at ESC(+), vCM(-)
173 and stable TAD boundaries (**Extended Data Fig. 5d**). Thus, we explored whether other
174 potential mechanisms might be involved in the formation of these TAD boundaries.
175 Interestingly, we discovered that eight classes of repeat elements were over-
176 represented specifically at ESC(+) TAD boundaries when compared to stable TAD
177 boundaries (p-value < 0.01 and fold change > 2), whereas no enrichment of these
178 repeats was observed in vCM(-) TAD boundaries (**Extended Data Fig. 6a**).
179 Additionally, the ESC(+) TAD boundaries were associated with hESC-specific H3K27ac

180 signal and mRNA transcription (**Extended Data Fig. 5e, f**), which was particularly
181 enriched for HERV-H transcription (**Fig. 2a and Extended Data Fig. 6b**).
182
183 HERV-H is a class of primate-specific endogenous retrotransposons. It is
184 transcriptionally active in both human preimplantation embryos and pluripotent stem
185 cells[35,36], and plays a critical role in pluripotency, stem cell maintenance and somatic
186 cell reprograming[37-41]. Although more than 1000 copies of HERV-H sequences exist in
187 the human genome and are expressed at different levels in hESCs (**Fig. 2b)**[37], we
188 discovered that the top 50 highly transcribed HERV-H loci were predominantly located
189 at TAD boundaries in hESCs (**Fig. 2c, Extended Data. Fig 6c)**. Notably, transcriptional
190 silencing of these HERV-Hs was observed by day 2 in hESCs differentiating into
191 mesodermal cells and coincided with the loss of the corresponding TAD boundaries
192 (**Fig. 2d;** specific loci in **Fig2e, f and Extended Data Fig 6d, e, f**). As controls, genes
193 with similar transcriptional levels to these top 50 HERV-H loci did not display any
194 boundary activity as evidenced by the DI scores (**Extended Data Fig. 7a**). This
195 correlation was present in not only H9 hESCs but also other hPSCs including H1
196 hESCs[8,42] and human induced PSCs (iPSC)[43]; however, differentiated cells did not
197 exhibit these HERV-H associated TAD boundaries (**Extended Data Fig. 7b**).
198
199 Because HERV-Hs might contribute to TAD boundary formation by recruiting DNA
200 binding proteins, we examined publicly available ChIP-seq data of histone modifications
201 and transcription factors (TFs) in hESCs. We identified several modifications and
202 factors enriched at sites for the top 50 transcribed HERV-Hs when compared to those
203 for the lower-ranking HERV-Hs. They included POL2 transcriptional machinery
204 (POLR2A and POLR2AphosphoS5), active histone marks (H3K4me3, H3K4me2 and
205 H3K27ac), ubiquitously active TFs (JUND and SP1), pluripotency factor NANOG[39] and
206 all three subunits of the cohesin complex (**Extended Data Fig. 7c**). To further
207 investigate the role of cohesin in HERV-H-mediated TAD boundary formation, we
208 performed ChIP-seq in hESCs for the cohesin complex subunit SMC3. Analyzing this
209 ChIP-seq data along with RNA-seq as well as CTCF and POLR2A ChIP-seq data,
210 revealed broad enrichment of cohesin, CTCF and POLR2A specifically at the 3' end of

211 these highly transcribed HERV-H sequences (**Fig. 2g**). In contrast to these findings, the
212 enrichment of these factors was not similarly observed at sites of lower ranked HERV-
213 Hs (**Fig. 2g**), suggesting that these factors may be involved in mediating the formation
214 of TAD boundaries by HERV-Hs. However, we cannot completely rule out that CTCF
215 may also participate in defining these HERV-H-associated TAD boundaries, as weak
216 but reproducible enrichment of CTCF occupancy was observed in these regions despite
217 the absence of CTCF binding motifs (**Fig. 2g**). Overall, these data support that HERV-
218 Hs form TAD boundaries by accumulating cohesin complexes at its 3' end, and cohesin
219 is likely positioned by the transcribing POL2 complex[44,45].
220

221 **Deletion or silencing of HERV-H elements abolishes the corresponding**
222 **TAD boundaries**

223 To determine whether HERV-H sequences are indeed required for TAD boundary
224 formation, we deleted two HERV-H elements located at two of the strongest hESC TAD
225 boundaries, using CRISPR-Cas9 genome editing tools (termed as HERV-H1-KO and
226 HERV-H2-KO, respectively). In both cases, deletion of these HERV-Hs resulted in
227 elimination of TAD boundaries, as evidenced by the merging of the two TADs bordering
228 each HERV-H element (**Fig. 3a**), and reduced expression of genes in the TAD domain
229 immediately 5' upstream of the HERV-H sequence (**Fig. 3a**). These findings thus reveal
230 that HERV-H sequences are indispensable for the TAD boundaries formed at these
231 regions and furthermore the presence of HERV-H increases the overall transcriptional
232 activity of the neighboring genes in the same TAD domain, supporting that HERV-Hs
233 may serve as enhancers in hESCs[46,47]. Notably, this activating effect is only seen in the
234 TAD domain on the 5' end of HERV-H and not the 3' end, suggesting that boundaries
235 present on the 3' end of HERV-Hs, may insulate the effect of enhancers. Supporting
236 this possibility, additional analyses of all genes located within 500 kb of boundary-
237 associated HERV-Hs showed that only the genes in regions 5' of HERV-H were up-
238 regulated in hESCs compared to differentiated cells (**Fig. 3b**).
239

240 Further examination of the transcriptional profiles of both HERV-H knockouts revealed
241 that expression of 353 and 144 genes were significantly changed (**Extended Data Fig.**

242 **8a, b, Extended Data Table 3**), and the fold changes of HERV-H1-KO and HERV-H2-
243 KO were highly concordant (**Extended Data Fig. 8c**, Pearson r=0.5, p-value< 1e-15)
244 despite both engineered cells maintaining pluripotency. Of the 43 genes down-
245 regulated in both KOs, 10 (23%) were located within 20 kb of other HERV-H sequences
246 (**Extended Data Fig. 8d**), including two well studied chimeric transcripts *SCGB3A2*[48]
247 and *LINC00458*[49], which are known to be regulated by HERV-Hs (**Extended Data Fig.**
248 **8e, f**). Thus, these data suggest that deletion of individual HERV-H sequences could
249 lead to down-regulation of HERV-H related transcription at other loci. In addition to
250 these findings, we also discovered that the human-specific long-coding RNA (lncRNA),
251 *Heart Brake* (*HBL1*), which suppresses cardiomyocyte differentiation[50], was also
252 significantly reduced in the HERV-H1-KO (**Extended Data Fig. 8f)**. To further explore
253 whether these changes in transcriptional landscape in the HERV-H1-KO promotes
254 cardiomyocyte differentiation, we examined the cardiomyocyte differentiation efficiency
255 of HERV-H1-KO hESCs compared to control hESCs under suboptimal cardiomyocyte
256 differentiation conditions. While these conditions led to ~40% cardiomyocyte
257 differentiation efficiency at day 15 in control hESCs, we discovered that HERV-H1-KO
258 hESCs exhibited >80% cardiomyocyte differentiation efficiency (**Fig. 3c**). Altogether,
259 these results suggest that removal of one highly expressed TAD-associated HERV-H
260 sequence might facilitate the exit from pluripotency, including the cardiac lineage
261 examined in this study.
262
263 Having established that HERV-H DNA sequences are required for TAD boundary
264 formation, we investigated whether its transcriptional activity is also essential. To
265 address this question, we blocked the transcription of the aforementioned HERV-H loci
266 without altering the HERV-H DNA sequence by employing a CRISPR-dCas9-KRAB
267 system with sgRNAs that targeted the promoters (5'LTR7) of these HERV-Hs (**Fig. 3d**).
268 Each sgRNA specifically reduced the targeted HERV-H expression by over 70%, while
269 not affecting the expression of other HERV-Hs (**Fig. 3d**). Accompanying the reduction
270 of HERV-H, the TAD border was greatly weakened, evidenced by the decreased
271 directionality index (**Fig. 3e**). This result indicates that HERV-H's high transcription is
272 required for creating these groups of TAD boundaries in hESC.

HERV-H introduces *de novo* TAD boundaries

It is estimated that HERV-H integrated into the primate lineage 30–40 million years ago (MYA) at the time of divergence of Old and New World monkeys[51,52], with the largest expansion (LTR7 and LTR7B) occurring in the Old World monkey lineage, including the Great apes[51]. Consequently, the Great Apes, including humans, harbors over 1000 copies of HERV-H, whereas New World monkey has 50-100 copies and none in non-primates. We hypothesize that HERV-H might have introduced *de novo* chromatin domains during primate evolution. To test this hypothesis, we performed Hi-C to interrogate the chromatin architecture of chimpanzee iPSC, marmoset (New World monkey) iPSC and mouse ESC, and compared them to that of the human ESC. While the top 50 transcribing HERV-H loci showed contact insulation in human ESC, their syntenic regions in the marmoset iPSC and mouse ESC did not exhibit such insulation (**Fig. 4a**). Surprisingly, although chimpanzee has similar copies of HERV-Hs as human, we did not observe any TAD boundaries in the syntenic regions (**Fig. 4a**). One reason for the lack of boundaries in chimpanzee could be the significantly lower expression of HERV-Hs in chimpanzee iPSC under our culturing conditions (**Extended Data. Fig. 9a**). Taken together, our data suggests that HERV-H introduces *de novo* TAD boundaries in PSCs during primate evolution, in a transcription-dependent manner. Furthermore, additional sequence analysis of the HERV-H LTR sequence revealed that the TAD-forming HERV-Hs were predominantly flanked by a 450bp subtype of LTR7 at both ends (**Extended Data Fig. 9b**, **c**). And those HERV-Hs showed less sequence divergence between their 5' LTRs and 3' LTRs (**Extended Data Fig. 9d**), suggesting that they likely were inserted more recently than other HERV-Hs that do not form TAD boundaries, or were impacted by gene conversion or selection[53].

Finally, to investigate whether the *de novo* insertion of HERV-Hs can create new TAD boundaries, we engineered two human hESC lines in which an 8kb sequence from the HERV-H2 locus (with 1kb flanking HERV-H from each side) was cloned and randomly inserted it into the H9 hESC genome using a piggybac transposon (**Fig. 4b**). From sequence analyses, we discovered that this HERV-H2 construct was inserted into 52

new genomic loci (**Extended Data Fig. 9e,f**), of which six exhibited a significant increase in the local contact insulation at the place of insertion, as evidenced by the change in Hi-C contact matrix and DI score (One example in **Fig. 4c**). This data demonstrated that HERV-H insertion can create chromatin boundaries *de novo*, however, such ability is likely also influenced by the native genomic contexts.

Stage-specific chromatin loops form interaction networks around cardiac transcription factors

From our extensive interrogation of the genome during hESC cardiomyocyte differentiation, we also identified 14,216 chromatin loops spanning a median genomic distance of 180 kb at 10 kb resolution (**Extended Data Table 4**). More than 70% of the loops corresponded to either promoter-to-promoter or promoter-to-enhancer interactions (**Extended Data Fig. 10a**). Consistent with chromatin loops detected in different cell types or model systems[14,15], the loop anchors were enriched for accessible chromatin, histone marks, transcriptional start sites and CTCF binding sites (**Extended Data Fig. 10b**). More importantly, 23% (3,251) of these chromatin loops exhibited significant changes in interaction strength during cardiomyocyte differentiation, which illustrates a dramatic reconfiguration of the gene regulatory network (**Fig. 5a, Extended Data Fig. 10c**). Gene transcription and chromatin marks at anchors of stage-specific chromatin loops were correlated to a significantly higher degree than that of anchors of static chromatin loops and distance-matched control (**Fig. 5e**). Using K-means clustering, we defined five distinct clusters of stage-specific chromatin loops (**Fig. 5b**), each accompanied by corresponding epigenomic changes (**Extended Data Figure 10d**). Remarkably, the gene sets enriched for each chromatin loop cluster recapitulated the developmental processes from early differentiation, to heart development, and culminated in cardiac muscle development (**Fig. 5c**). Chromatin loops gained during cardiomyocyte differentiation included promoters of many well studied heart transcription factors, such as *GATA6*, *TBX5*, *NKX2-5*, *TBX3* and *MEF2A*[54]. Analysis of ATAC-seq peaks at loop anchors also revealed distinct patterns of transcription factor binding throughout differentiation (**Fig. 5d**). For instance, enrichment for GATA motifs emerged at day 2, and T-BOX motifs at day 5, whereas MEF2 motifs were only highly

335 enriched at day 80. These findings suggest a process involving pioneering factor
336 activity of GATA family members such as GATA4 for the initial rewiring of chromatin
337 loops to activate relevant early cardiac developmental gene programs, whereas MEF
338 family members such as MEF2C may play a role in the maturation of cardiomyocytes.
339 Finally, loops formed at later stages tended to span longer genomic distances, to have a
340 higher percentage of convergent CTCF-CTCF motifs, and were more likely to occur at
341 TAD boundaries (**Extended Data Fig. 10e, f, g**). These observations further support
342 the global trend of increased long-distance interactions during cardiomyocyte
343 differentiation (**Extended Data Fig. 4a**).
344
345 Integrative analysis of the chromatin architecture and ChIP-seq data revealed two major
346 types of loops: 1) loops enriched for the polycomb mark H3K27me3 (**Extended Data**
347 **Fig 10d, 11a**); and 2) loops enriched for H3K27ac (**Extended Data Fig 10d, 11b**).
348 H3K27me3-associated, polycomb dependent loop interactions, which have been
349 described in *Drosophila* and mouse ESC[23,55,56], were prominent in hESCs, spanned
350 very long genomic distances and were abruptly lost at the mesoderm stage (**Extended**
351 **Data Fig 10d, 11a**). Notably, such loops frequently involved promoters of transcription
352 factors, which were not expressed at the hESC stage (Fold enrichment=9.7, adjusted p-
353 value=1e-70, hypergeometric test), including the early cardiac transcription factor
354 *HAND2*. Further analyses of the *HAND2* locus revealed a multistep assembly of loops
355 formed between the *HAND2* promoter and several distal elements (**Fig. 1b**). In hESCs
356 and mesodermal cells, which do not express *HAND2*, the *HAND2* promoter formed an
357 H3K27me3 associated loop with the promoter region of the lncRNA *LINC02268*.
358 However, by day 5, this loop was eliminated and two new loop interactions were
359 established between the *HAND2* promoter and two distal regions marked by H3K27ac
360 and H3K4me1. This chromatin remodeling coincided with the onset of *HAND2*
361 expression in cardiac mesodermal cells (D5) and a subsequent 4-fold increase in
362 *HAND2* expression in primitive cardiomyocytes (D15) and ventricular cardiomyocytes
363 (D80); however, lncRNA *LINC02268* remained transcriptionally repressed (**Fig. 1b, c**).
364 Based on this multi-loop chromatin remodeling at the *HAND2* locus, we further
365 examined whether multi-loop formation was a general feature to tightly control

366 expression levels of transcription factors whose function may impact cell fate more than
367 that of other genes. To this end, we calculated the degree of interaction for each loop
368 anchor (**Fig. 5f**), and observed that 45% of the loop anchors interacted with at least two
369 other anchors (**Fig. 5g**), suggesting that chromatin loops frequently form network hubs.
370 In addition to *HAND2*, many other well-known cardiac TFs, including *TBX5*, *GATA4*,
371 *TBX20*, *MEIS1/2*, also interacted with four or more *cis*-elements and formed similar
372 network hubs (**Extended Data Fig 11c, d, e,** and **Extended Data Table. 5**). Supporting
373 these findings, the network hubs with more chromatin interactions were more likely to
374 include transcription start sites (TSSs) of TFs (**Fig. 5h**), and less likely to involve TSSs
375 of house-keeping genes (**Fig. 5i**). Finally, those network hubs with more interactions
376 frequently contained more CTCF peaks (**Extended Data. Fig. 11f**), suggesting that
377 these interactions were likely facilitated by CTCF.
378

379 **Chromatin loops predict cognate target genes for cardiac GWAS SNPs**

380 We next used the chromatin interaction maps to link non-coding genetic variants from
381 genome-wide association studies (GWAS) to potential target genes, and found a strong
382 enrichment of single nucleotide polymorphisms (SNPs) associated with cardiac traits in
383 the chromatin loop anchors (**Fig. 6a**). The list of predicted candidate genes for these
384 SNPs predominantly included genes involved in cardiac function (**Extended Data Table**
385 **6**) and cardiac ion channel genes (**Fig. 6b**). One notable cardiac GWAS example was
386 rs6781009, which was found in strong linkage disequilibrium (LD) with lead SNPs of
387 four cardiac electrophysiologic GWAS traits (QRS complex, QRS duration, QT interval
388 and electrocardiographic traits). Using our chromatin interaction maps, we found that
389 this SNP was located in a stage-specific D80 enhancer that interacted with the promoter
390 of the cardiac sodium channel gene *SCN5A* (**Extended Data Fig. 12a**), supporting
391 similar findings recently observed by circular chromosome conformation capture (4C)
392 studies in human cardiac tissue and functional studies in mouse hearts[57]. Another
393 cardiac GWAS example was rs17608766, the lead SNP for six cardiac traits including
394 aortic root size, blood pressure, pulse pressure, QRS complex, QRS duration and
395 systolic blood pressure. Our chromatin interaction maps revealed that this SNP
396 overlapped a D2 mesoderm specific enhancer that interacted ~110 kb away with the

397 promoter of *WNT3* (**Fig. 6c**), which is known to be involved in cardiac mesoderm
398 induction[58,59] and thus may explain the broad cardiac traits potentially affected by this
399 SNP. Because this SNP is within a highly conserved genomic region (**Fig. 6d**) and
400 predicted to alter a binding motif for the KLF4 transcription factor (**Extended Data Fig.**
401 **12b),** which has been shown to regulate cardiac function in mice[60], we deleted the
402 putative enhancer containing this SNP in hESCs using CRISPR-mediated genome
403 editing. This deletion consequently resulted in a ~20-50% reduction in *WNT3*
404 expression in mesodermal cells at day 2 (**Fig. 6e)**, but had no effect on expression of
405 the nearest gene *GOSR2* (**Extended Data Fig. 12c**), which has previously been
406 annotated to rs17608766. Finally, in addition to investigating SNPs associated with
407 cardiac functional traits, we also examined long-range acting SNPs for cardiac
408 developmental traits (i.e. congenital heart disease). Interrogating rs870142, a SNP
409 associated with atrial septal defect[61], revealed that it was located in an active D2
410 mesodermal enhancer within the intron of *STX18-AS1*, which formed dynamic long-
411 range interactions (~200 kb) with the promoter of *MSX1* (**Fig. 6f, g**), a developmental
412 TF reported to interact with *TBX5* during atrial septal development[62,63]. Deletion of this
413 enhancer containing rs870142 using CRISPR/Cas9-mediated genome editing resulted
414 in a 50% reduction in *MSX1* expression (**Fig. 6h**), whereas the expression of the
415 nearest gene *STX18* remained unchanged (**Extended Data Fig. 12d)**. Thus, these
416 findings support that our chromatin interaction maps can be used to discover candidate
417 genes affected by genetic disease or trait variants located in distal regulatory elements
418 controlling gene expression.
419

420 # DISCUSSION
421

422 Evolution of multi-cellular organisms is driven, in large part, by the invention of new
423 gene regulatory circuits responsible for the fitness traits of each species. A long-
424 standing theory holds that retroviruses may play an important role in the evolution of
425 gene regulatory logic[64,65]. Over the years, many classes of endogenous retroviral
426 elements have been found to recruit transcription factors to regulate nearby genes in a
427 cell-type specific manner, or initiate transcription of non-coding RNAs with important

428 regulatory functions[46,66]. Importantly, the expansion of SINE elements in rodents, dogs
429 and opossums have been attributed to the rewiring of gene regulatory networks in these
430 mammals, by expanding the repertoires in each species of CTCF[22], a DNA binding
431 protein with a critical role in chromatin organization. Surprisingly, no evidence has been
432 found so far for repeat-driven expansion of CTCF binding in the primate genomes[22],
433 raising the question whether retrotransposon-driven chromatin re-organization indeed is
434 a general strategy of evolution that applies also to the primate lineage. Moreover,
435 retrotransposon elements have been implicated in shaping chromatin boundaries[67],
436 however, this notion also remains to be experimentally investigated.
437
438 Here, we provide multiple lines of evidence to support that the primate-specific retroviral
439 elements HERV-H can delineate TAD boundaries in the human PSCs. Previous studies
440 suggest that HERV-Hs integrated into the human genome during primate evolution to
441 regulate human-specific pluripotency through creating novel chimeric transcripts
442 (*ESRG*[37], *linc-ROR*[41,68] and *LINC00458*[49]) and providing potential binding sites to recruit
443 pluripotency factors (*NANOG*, *SOX2* and *POU5F1*)[38]. Our findings reveal that HERV-H
444 sequences might affect gene regulatory program through yet a third mechanism, by
445 shaping the chromatin architecture. Given these HERV-H-mediated chromatin
446 remodeling findings, future investigations of other ERV families of repeats and/or other
447 families of repetitive elements are warranted to illuminate whether other distinct classes
448 of genomic repeats may also exhibit the capacity to shape chromatin architecture in a
449 broad range of cell types including PSCs.
450
451 Through correlation analysis and CRISPRi intervention, we reveal that HERV-H's ability
452 to form TAD boundaries is dependent on its transcriptional activity. In addition, we show
453 that Polymerase II and cohesin complex accumulates at the 3' terminal of highly
454 transcribing HERV-Hs. Cohesin complex is involved in forming long-range chromatin
455 loops at CTCF binding sites, through an ATP-dependent loop-extrusion mechanism[16,17].
456 In mammalian cells, most TAD borders are occupied by CTCF and cohesin complex.
457 However, in the human PSC, the HERV-H-associated TAD borders lack canonical
458 CTCF binding motifs and sharp CTCF ChIP-seq peaks, but still retain cohesin

459 complexes. Taken together with the CRISPRi experiment, our findings support that the
460 cohesin complexes are likely positioned by Polymerase II, rather than CTCF. This
461 model is consistent with and provide additional evidence to recent reports of the role of
462 transcription and polymerase movement to position cohesin complex and thus shaping
463 chromatin structure[44,45].
464
465 In addition to the HERV-H discovery, we have interrogated the dynamic chromatin
466 structure re-organization during cardiomyocyte differentiation, cataloged chromatin
467 loops and annotated potential target genes for GWAS variants from cardiac studies.
468 More importantly, we functionally validated our predicted enhancer/SNP to promoter
469 maps by CRISPR mediated deletion, and further showed that WNT3 and MSX1 are
470 likely linked to congenital heart diseases and other cardiac phenotypes. This highlights
471 the utility of the presented chromatin interaction maps to associate non-coding
472 regulatory elements to target genes. Overall, these detailed chromatin accessibility and
473 interaction maps will be a valuable genomic resource for not only providing novel
474 insights into cardiomyocyte development but also identifying important genetic variants
475 associated with cardiac traits including congenital heart disease.
476

Acknowledgements

We thank Samantha Kuan and Dr. Bin Li for sequencing and bioinformatics support. We thank Eric van Nostrand for RNA extraction. We would like to thank M. Daadi for providing marmoset iPSCs and F. Gage for providing chimpanzee iPSCs. This project is supported by funding from the Ludwig Institute for Cancer Research (to B.R.), the NIH (1UM1HL128773 to S.M.E., N.C.C., E.D. and B.R.). J.W. is the Virginia Murchison Linthicum Scholar in Medical Research. S.P. was supported by a postdoctoral fellowship from the Deutsche Forschungsgemeinschaft (DFG, PR 1668/1-1). E.N.F. was supported by NIH pre-doctoral training grant (5T32HL007444-35).

Author contributions

N.C.C. and B.R. designed and supervised the experiments, analysis, and data interpretation. Y.Z. implemented the analysis pipeline and analyzed all sequencing datasets, interpreted the results, designed the experiments for HERV-H and chromatin loop functional studies. T.L. generated the CRISPR-cas9 edited cell lines for HERV-H and chromatin loop functional studies, performed differentiation and qPCR of the corresponding cell lines. S.P. initiated the project, performed the Hi-C and ATAC-seq experiments, and helped with interpretation of the results. J.G. and E.N.F performed cell culture, differentiation and collected cells for Hi-C, ChIP-seq, ATAC-seq and RNA-seq assays. E.D. contributed to analysis and interpretation of the ChIP-seq data. ChIP-seq experiments were performed by A.Y.L. (histone modifications), S.C. (CTCF), Q.Z. and H.H. (SMC3). Y.Q. and R.F. helped with the analysis of Hi-C datasets. K.M. helped with the genome-editing experiments. L. Y., J. C. I. B and J. W. cultured and prepared non-human primate iPSCs for sequencing. Z.Y. performed the RNA-seq experiments. R.H. performed the Hi-C experiments for HERV-H knock-out, CRISPRi, HERV-H knock-in and primate iPSC cell lines. S.M.E. helped with interpretation of the results. Y.Z., T.L., S.P., N.C.C., and B.R. wrote the manuscript with input from all authors.

Author Information

Sequencing data from this study have been deposited in the Gene Expression Omnibus under the accession number GSE116862. The authors declare no competing financial

interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.C.C. (nchi@ucsd.edu) and B.R. (biren@ucsd.edu).

512

513 # Figure Legends

514 **Figure 1. Global chromatin architecture reorganization during human ventricular**
515 **cardiomyocyte differentiation.** (**a & b**) Hi-C contact matrices for each stage of
516 cardiomyocyte differentiation at mega-base resolution and at higher resolution at the
517 *HAND2* locus with corresponding epigenomic profiles. (**c**) Schematic illustration of the
518 multi-step decommissioning of inhibitory loop and assembly of subsequent activating
519 loops anchored at *HAND2* promoter. (**d**) A heatmap showing hierarchical clustering of
520 dynamic chromatin compartments during cardiomyocyte differentiation. The pseudo
521 color reflects the PC1 values (compartment A/B) of compartment bins. Negative PC1
522 value stands for compartment B and positive for compartment A. Representative genes
523 located in corresponding compartment bins are annotated to the right of the heatmap.
524 (**e**) A heatmap showing the DI delta scores for the stage-specific TAD boundaries,
525 ordered by the presence of TADs at six stages. vCM(-) stands for TAD boundaries lost
526 in purified ventricular cardiomyocytes at D80, ESC(+) for hESC-specific TAD
527 boundaries.

528

529 **Figure 2. Transcriptionally active HERV-H forms human ESC-specific TAD**
530 **boundaries.** (**a**) Aggregate RNA-seq expression profile (RPKM normalized) at ESC(+)
531 TAD boundaries that overlap HERV-H element. (**b**) Scatterplot shows the expression
532 levels (RPKM) across different HERV-H loci (ordered by expression levels from high to
533 low in ESC) at D0, D2 and D5 stages of differentiation. (**c**) Heatmap of aggregate Hi-C
534 contact matrix [log2(observed/expected)] within 200 kb of the top 50, 51-100 and 101-
535 150 ranked HERV-Hs, at D0. (**d**) Heatmap of the aggregated Hi-C matrix
536 [log2(observed/expected)] within 200 kb of the top 50 HERV-Hs, at D0, D2 and D5. (**e,f**)
537 Representative Hi-C interaction matrices of two HERV-H loci located at ESC(+) TAD
538 boundaries at D0, D2 and D5 (top) are shown as heatmaps along with genome browser
539 tracks of POLR2A, SMC3, CTCF, H3K27ac ChIP-seq and RNA-seq data of the
540 expanded genomic region containing the TAD boundary (arrow). (**g**) Aggregated
541 genomic profiles of RNA-seq, POLR2A, SMC3 and CTCF ChIP-seq around top 50

542 HERV-Hs located on the ESC(+) TAD boundaries (red) and lower ranked HERV-Hs
543 (grey).
544
545 **Figure 3. Deletion or silencing of two HERV-H sequences leads to merging of**
546 **TADs in hESC**. (**a**) Hi-C interaction matrices of the wild-type (WT) and mutant hESC
547 lines (HERV-H1-KO and HERV-H2-KO*)* are shown, along with DI scores, and fold
548 changes of gene expression at the HERV-H1 and HERV-H2 loci. The loss of TAD
549 boundary in the mutant cells is accompanied with decrease of RNA expression 5'
550 terminus to the HERV-H sequences. (**b**) Boxplots show expression levels (RPKMs) of
551 genes whose TSSs are located from -500 kb to the 5' LTR and from 3' LTR to +500 kb
552 of HERV-Hs of boundary-associated HERV-Hs. (**c**) Line chart (mean ± standard error,
553 N=3 cardiomyocyte differentiations) shows percentage of TNNT2 positive cells during
554 cardiomyocyte differentiation of WT and HERV-H1-KO hESC lines (two HERV-H1-KO
555 clones analyzed). HERV-H1-KO hESCs display increased cardiomyocyte differentiation
556 efficiency compared to WT hESCs. (**d**) Design of the CRISPR-dCas9-KRAB system to
557 silence HERV-H expression (top) and gene expression values of HERV-H1 and HERV-
558 H2 in WT and CRISPR-i targeted hESCs (bottom). (**e**) Hi-C interaction matrices of the
559 CRISPR-i targeted hESC lines (sgHERV-H1 and sgHERV-H2*)* are shown, along with DI
560 scores at the HERV-H1 and HERV-H2 loci.
561
562 **Figure 4. HERV-H introduces *de novo* TAD boundaries during primate evolution**
563 **and in engineered hESC.** (**a**) Simplified tree of primate evolution with the copies of
564 HERV-H annotated (left) and Hi-C interaction matrices of human ESC, chimpanzee
565 iPSC, marmoset iPSC and mouse ESC are shown, along with DI scores at the syntetic
566 regions to human HERV-H1 locus, HERV-H2 locus, and all top 50 transcribing HERV-
567 Hs. (**b**) Design of the piggybac vector to "transpose" HERV-H to random genomic
568 locations in the hESC (HERV-H2-KO line). (**c**) Hi-C interaction matrices of the parental
569 cell line (HERV-H2-KO) and the cell line with HERV-H insertions (HERV-H-ins.clone1)
570 are shown, along with DI scores at the locus of one HERV-H insertion.
571

572

573 **Figure 5. Dynamics of chromatin loops during human cardiomyocyte**
574 **differentiation.** (**a**) Bar-chart shows the percentages of static or dynamic chromatin
575 loops during cardiomyocyte differentiation. (**b**) Heatmap shows K-means clustering of
576 the stage-specific chromatin loops. Pseudo color reflects the normalized contact
577 frequencies between the loop anchors for each stage-specific loop. (**c**) Enriched Gene
578 Ontology (GO) terms for genes located at the loop anchors of each cluster in (**b**) along
579 with their adjusted p-values. (**d**) Over-represented transcription factor binding motifs for
580 distal open chromatin regions located at the loop anchors of each cluster in (**b**) and their
581 p-values. CTCF motif enrichment was not shown but observed in all clusters. (**e**)
582 Boxplots of Pearson correlation coefficients of histone marks of distal enhancers and
583 expression levels of their target genes predicted based on the chromatin loops.
584 Distance-matched controls are "mirror" loops that connect to genomic regions from the
585 opposite direction with the same distances. P-values are from Wilcoxon Rank Sum test.
586 (**f**) Diagram illustrates how the degree of chromatin interactions for each loop anchor is
587 determined. (**g**) Histogram of degrees of network connections for all loop anchors. (**h**)
588 Fraction of chromatin interaction network hubs containing at least one transcription start
589 site (TSS) of a transcription factor (TF) plotted against the degree of connectivity. (**i**)
590 Fraction of network hubs containing at least one TSS of a house-keeping gene (HKG)
591 plotted against the degree of connectivity.

592

593 **Figure 6. Identification of target gene for heart-related GWAS SNPs using**
594 **chromatin loops detected during cardiomyocyte differentiation.** (**a**) Bar graph
595 shows the enrichment of GWAS SNPs associated with cardiac traits at the chromatin
596 loop anchors detected during the cardiomyocyte differentiation. (**b**) Enriched GO terms
597 for the genes predicted to be linked to SNPs from cardiac GWAS studies along with P
598 values. Red dashed line indicates P value cutoff of 0.05. (**c**) Hi-C contact matrices, and
599 genome browser tracks of various histone modification ChIP-seq and RNA-seq data at
600 the *WNT3* locus. A cardiac GWAS SNP rs17608766 is marked by a red line. (**d**)
601 Expanded view of the genomic region containing rs17608766. The putative enhancer
602 containing rs17608766, which located in an open chromatin region flanked by H3K27ac

603 and H3K4me1 peaks, is targeted for CRISPR/Cas9-mediated deletion. (**e**) Bar charts
604 show the normalized expression levels of *WNT3* in three CRISPR genome edited
605 clones at D2 (mean ± standard error, N =5). Deleting the putative enhancer containing
606 rs17608766 specifically reduces *WNT3* expression. P-values are from one-sided paired
607 t-test. (**f**) Hi-C contact matrices, epigenomic profiles at *MSX1* locus. A cardiac GWAS
608 SNP rs870142 is marked by red line. (**g**) Expanded view of the genomic region
609 containing rs870142. CRISPR mediated deletion of the putative enhancer is marked.
610 (**h**) Bar charts show that deleting the putative enhancer containing rs870142 specifically
611 reduces *MSX1* expression in three independent clones at D2 (mean ± standard error, N
612 = 9). P-values are from one-sided paired t-test.
613

614

615 # Extended Data Figure Legends.

616

617 **Extended Data Figure 1. Imaging and Flow cytometry data for cardiomyocyte**
618 **differentiation.** (**a**) Immunostaining for MYL2 (white) shows exclusive expression in
619 *MYL2*-H2B-GFP + hESC-derived cardiomyocytes after 80 days of cardiac
620 differentiation. *MYL2*-H2B-GFP + cells were positive for cardiac troponin T
621 (cTNT) (red), but not all cTnT+ cells were MYL2+ cardiomyocytes. DNA was
622 stained with DAPI (blue). Images are representative of a minimum of three
623 independent experiments. (**b**) Flow cytometric quantification of distinct cell-
624 surface markers for D2-D15 time points: T (D2), KDR/PDGFRα (D5), TNNT2
625 (D15). H2B-GFP + ventricular cardiomyocytes were sorted at day 80. Numbers
626 represent the respective percentage of cells.

627

628 **Extended Data Figure2. The Hi-C data are highly reproducible. (a**) Smoothed
629 scatterplot of the Hi-C contacts between biological replicates for each stage. bin size =
630 100 kb. (**b**) Hierarchical clustering of Hi-C contact matrices based on 1-PCC (Pearson
631 correlation coefficients) of the contacts between all samples, bin size = 100 kb. (**c**)
632 Hierarchical clustering of the compartment A/B scores (PC1 values) between samples.
633 (**d**) Hierarchical clustering of the insulation scores between samples.

634

635 **Extended Data Figure 3. Quality metrics of ChIP-seq, RNA-seq and ATAC-seq**
636 **data used in the present study.** (**a**) Boxplot showing the number of reads for each
637 sample. (**b**) Boxplot showing the percentage of alignments to human genome for each
638 sample. (**c**) Boxplot showing the percentage of potential duplicated reads for each
639 sample. (**d**) Boxplot showing the number of peaks called for each ChIP-seq and ATAC-
640 seq sample. (**e**) Barplots showing the relative expression levels (by RNA-seq) of
641 representative genes during cardiomyocyte differentiation.

642

643 **Extended Data Figure 4. Global changes in chromatin organization during**
644 **cardiomyocyte differentiation.** (**a**) Probability of Hi-C interaction versus interacting
645 distance (log2) during cardiomyocyte differentiation. (**b**) Barplot shows the percentages
646 of compartment switches for each stage transition. (**c**) Histogram of the Pearson
647 correlation coefficients between gene expression and compartment A/B (PC1) value. (**d**)
648 Genome browser view of *SOX2*, *HAND2* and *RYR2* loci, showing PC1 (Blue/red), RNA-
649 seq (Black) and H3K27ac (Blue) signals.
650
651 **Extended Data Figure 5. Dynamics of TADs during cardiomyocyte differentiation.**
652 (**a**) Number of TADs as defined by a domain caller algorithm first reported by Dixon et
653 al. 2012[5]. (**b**) Number of TADs (or loop domains) found by Arrowhead algorithm[14]. (**c**)
654 The number of TADs found by Insulation score[34]. (**d**) Fraction of TAD boundaries that
655 contain CTCF ChIP-seq peaks at each stage of cardiomyocyte differentiation. “stable”
656 group stands for TADs that are present at all stages. (**e**) Boxplot showing the dynamics
657 of H3K27ac signal at ESC(+) TADs. (**f**) Boxplot showing the dynamics of RNA-seq
658 signal at ESC(+) TADs.
659
660 **Extended Data Figure 6. HERV-H elements are enriched at ESC(+) TAD**
661 **boundaries. (a)** Scatter plot shows the fold enrichment and -log10(p-values) for various
662 repeat element classes at ESC(+) and vCM(-) TAD boundaries relative to the rest of the
663 TAD boundaries. P-values are from two-sided proportion test. (**b**) Aggregate RNA-seq
664 expression profile (RPKM normalized) at ESC(+) TAD boundaries that overlap each of
665 the eight repeat elements (HERVH-int and LTR7 are combined as they both belong to
666 HERV-H). (**c**) Heatmap of DI scores within 70 kb of the top 50 most highly-expressed
667 HERV-Hs. (**d,e,f**) Hi-C heatmap, DI, and epigenomic profiles for three ESC(+) TAD
668 boundaries harboring HERV-H/LTR7 sequences.
669
670 **Extended Data Figure 7. Further characterization of the HERV-H-associated TAD**
671 **boundaries.** (**a**) DI score profiles around the TSSs of genes expressed at similar levels
672 to the top 50 transcribed HERV-Hs. (**b**) Aggregated DI score profiles of the TAD-
673 boundary associated with HERV-Hs in multiple H1 ESC derived lineages and iPSCs.

674 ESC: embryonic stem cell; MES: mesendoderm. MSC: mesenchymal stem cell; NPC:
675 neural stem cell; TRO: trophoblast-like cell; iPSC: induced pluripotent stem cell.
676 Interestingly, in human mesendoderm cells (an early human embryonic state that gives
677 rise to mesoderm and endoderm cells), both the expression of these HERV-Hs and their
678 corresponding TAD boundary strength were approximately half the levels compared to
679 those observed in hESCs. There might be some un-differentiated cells in this
680 population. (**c**) Barplot shows the ChIP-seq signal fold enrichment of the top 50 HERV-
681 Hs comparing to HERV-Hs ranking 51-300.
682
683 **Extended Data Figure 8. Changes in gene expression in HERV-H knockout**
684 **hESCs.** (**a**) MA plot showing average gene expression levels and fold changes of each
685 gene in HERV-H1-KO and wild-type (WT). (**b**) Same as (**a**) but for HERV-H2-KO. (**c**)
686 Scatterplot shows the changes in gene expression in HERV-H1-KO and HERV-H2-KO
687 cells over WT cells. The red dots are genes significantly changed in both mutant cell
688 lines. The numbers at the corners show the number of significantly changed genes in
689 each Quadrant. (**d**) Barplot showing the number of significantly changed genes located
690 within 20 kb of the HERV-H sequences. Genes down-regulated in both HERV-H
691 knockouts were more likely to be within 20 kb of HERV-H sequences. (**e**, **f**) RNA-seq
692 profile of wild-type (*WT*) and HERV-H1-KO and HERV-H2-KO lines at the *SCGB3A2*
693 and *LINC00458/HBL1* gene loci.
694
695 **Extended Data Figure 9. Analysis of HERV-H LTR sequences and HERV-H**
696 **insertion.** (**a**) Scatterplot shows the expression levels (RPKM) across different HERV-H
697 loci in chimpanzee iPSC (ordered by expression levels from high to low). The
698 expression of HERV-Hs is at least 10 fold less than their human counterparts. (**b**) Bar
699 graph shows the percentage of HERV-Hs flanked by various types of LTRs. HERV-Hs
700 are ranked by their expression levels in the hESCs, and grouped by bins of 50. (**c**)
701 Violin plot shows the length of the flanking LTRs. HERV-Hs ranked and binned as
702 described in (**b**). (**d**) Boxplot shows the sequence divergence of the 5' LTR and 3' LTR
703 for each bin of HERV-Hs. HERV-Hs ranked and binned same as (**b**). (**e**) The insertional
704 profile of HERV-H in the HERV-H-ins.clone1 mutant line. The y-axis shows the pileup of

705 Hi-C read pairs with one end mapped to the HERV-H2 sequence. Based on proximity
706 ligation principle, loci with high pileup should harbor HERV-H2 insertion. (**f**) Same as (**e**)
707 but for HERV-H-ins.clone2 mutant line.
708
709 **Extended Data Figure 10. Relationships between chromatin loops and histone**
710 **modification during cardiomyocyte differentiation.** (**a**) Barplot showing fraction of
711 chromatin loops in each category of interactions. E-E: enhancer to enhancer; E-P:
712 enhancer to promoter; P-P: promoter to promoter. (**b**) Barplot showing fraction of
713 chromatin loop anchors or genome-wide background containing each of the epigenomic
714 element. (**c**) Representative data for Hi-C contact matrices of stage-specific and static
715 chromatin loop interactions (**d**) Heatmap shows the clustering of stage-specific
716 chromatin loops and histone modification states at the corresponding loop anchors. (**e**)
717 Violin plot of the genomic distances for each class of chromatin loops in log scale.
718 Symbol S: static loop. 1-5: Loop clusters 1-5 in **Fig.4b** and **Ext. Fig. 10d**. (**f**) Barplot for
719 fraction of loops harboring different type of CTCF orientations. No-CTCF: at least one
720 anchor does not contain CTCF peak; unknown: at least one anchor does not contain a
721 matched CTCF motif; incorrect: CTCF motifs at both anchors do not face inward;
722 inward: CTCF motifs at both anchors face inward. (**g**) Barplot for fraction of loops in
723 each category related to TAD boundaries. Inter-TAD: loop is outside of all TADs; TAD-
724 corner: loop is within 20 kb of any TAD corner; intra-TAD: loop is inside of TAD.
725
726 **Extended Data Figure 11. Chromatin interaction hubs around genes encoding**
727 **cardiac transcription factors.** (**a**) Example of a H3K27me3-associated inhibitory loop
728 connecting *MSX1* and *HMX1* promoters. (**b**) Example of several stage-specific and
729 H3K27ac associated activating loops at *HAND1* locus. (**c,d,e**) Hi-C contact matrices
730 (top), loop arcs (middle), and epigenomic profiles (bottom) for *GATA4* (**c**), *MEIS1* (**d**)
731 and *TBX5* (**e**). (**f**) Fraction of network hubs containing CTCF peaks plotted against the
732 degree of connectivity as illustrated by the number of CTCF peaks at each anchor
733 connectivity.
734

735 **Extended Data Figure 12. Validation of target gene predictions for heart-related**
736 **GWAS SNPs.** (**a**) Hi-C contact matrices, epigenomic profiles (ATAC-seq, H3K4me1,
737 H3K27ac and RNA-seq) and cardiac GWAS SNP rs6781009 at *SCN5A* locus. (**b**)
738 rs17608766 is predicted to alter KLF4 motif in the putative enhancer. (**c**) Normalized
739 expression values of *GOSR2* in three CRISPR-genome edited clones shows that
740 deleting the putative enhancer containing rs17608766 does not change *GOSR2*
741 expression (mean ± standard error, N=5). (**d**) Normalized expression values of *STX18*
742 in three CRISPR-genome edited clones shows that deleting the putative enhancer
743 containing rs870142 does not change *STX18* expression (mean ± standard error, N=7).
744
745
746 **Extended Data Table 1. Summary statistics for Hi-C data**
747 **Extended Data Table 2. List of stage-specific TAD boundaries.**
748 **Extended Data Table 3. List of differentially expressed genes in HERV-H1-KO and**
749 **HERV-H2-KO.**
750 **Extended Data Table 4. List of chromatin loops.**
751 **Extended Data Table 5. List of genes located on the network hubs.**
752 **Extended Data Table 6. List of target genes assigned to cardiac GWAS SNPs with**
753 **chromatin loops.**
754 **Extended Data Table 7. List of antibodies used for ChIP-seq.**
755 **Extended Data Table 8. List of primers and cell lines used in this study.**
756
757
758
759 **Supplementary Materials:**
760 Materials and Methods
761 Figures S1-S12
762 Tables S1-S8
763
764

References

1 Dekker, J. & Mirny, L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110-1121, doi:10.1016/j.cell.2016.02.007 (2016).

2 Yu, M. & Ren, B. The Three-Dimensional Organization of Mammalian Genomes. *Annu Rev Cell Dev Biol* **33**, 265-289, doi:10.1146/annurev-cellbio-100616-060531 (2017).

3 Meaburn, K. J. & Misteli, T. Cell biology: chromosome territories. *Nature* **445**, 379-781, doi:10.1038/445379a (2007).

4 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293, doi:10.1126/science.1181369 (2009).

5 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).

6 Sexton, T. *et al.* Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* **148**, 458-472, doi:10.1016/j.cell.2012.01.010 (2012).

7 Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385, doi:10.1038/nature11049 (2012).

8 Dixon, J. R. *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331-336, doi:10.1038/nature14222 (2015).

9 Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: The Unit of Chromosome Organization. *Mol Cell* **62**, 668-680, doi:10.1016/j.molcel.2016.05.018 (2016).

10 Lupianez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025, doi:10.1016/j.cell.2015.04.004 (2015).

11 Lupianez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet* **32**, 225-237, doi:10.1016/j.tig.2016.01.003 (2016).

12 Ibn-Salem, J. *et al.* Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome biology* **15**, 423, doi:10.1186/s13059-014-0423-1 (2014).

13 Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265-269, doi:10.1038/nature19800 (2016).

14 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).

15 Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290-294, doi:10.1038/nature12644 (2013).

16 Nasmyth, K. Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu Rev Genet* **35**, 673-745, doi:10.1146/annurev.genet.35.102401.091334 (2001).

17 Alipour, E. & Marko, J. F. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res* **40**, 11202-11212, doi:10.1093/nar/gks925 (2012).

18 Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e324, doi:10.1016/j.cell.2017.09.026 (2017).

19 Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol* **16**, 245-257, doi:10.1038/nrm3965 (2015).

20 Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922, doi:10.1016/j.cell.2017.05.004 (2017).

21 Vietri Rudan, M. *et al.* Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep* **10**, 1297-1309, doi:10.1016/j.celrep.2015.02.004 (2015).

22 Schmidt, D. *et al.* Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* **148**, 335-348, doi:10.1016/j.cell.2011.11.058 (2012).

23 Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557-572 e524, doi:10.1016/j.cell.2017.09.043 (2017).

24 Stadhouders, R. *et al.* Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. *Nat Genet* **50**, 238-249, doi:10.1038/s41588-017-0030-7 (2018).

25 Krijger, P. H. *et al.* Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic Cell Reprogramming. *Cell Stem Cell* **18**, 597-610, doi:10.1016/j.stem.2016.01.007 (2016).

26 Beagan, J. A. *et al.* Local Genome Topology Can Exhibit an Incompletely Rewired 3D-Folding State during Somatic Cell Reprogramming. *Cell Stem Cell* **18**, 611-624, doi:10.1016/j.stem.2016.04.004 (2016).

27 Gelb, B. D. Genetic basis of congenital heart disease. *Curr Opin Cardiol* **19**, 110-115 (2004).

28 Zaidi, S. *et al.* De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220-223, doi:10.1038/nature12141 (2013).

29 Rosa-Garrido, M. *et al.* High-Resolution Mapping of Chromatin Conformation in Cardiac Myocytes Reveals Structural Remodeling of the Epigenome in Heart Failure. *Circulation* **136**, 1613-1625, doi:10.1161/CIRCULATIONAHA.117.029430 (2017).
30 Choy, M. K. *et al.* Promoter interactome of human embryonic stem cell-derived cardiomyocytes connects GWAS regions to cardiac gene networks. *Nat Commun* **9**, 2526, doi:10.1038/s41467-018-04931-0 (2018).
31 Montefiori, L. E. *et al.* A promoter interaction map for cardiovascular disease genetics. *Elife* **7**, doi:10.7554/eLife.35788 (2018).
32 Veevers, J. *et al.* Cell-Surface Marker Signature for Enrichment of Ventricular Cardiomyocytes Derived from Human Embryonic Stem Cells. *Stem cell reports*, doi:10.1016/j.stemcr.2018.07.007 (2018).
33 Fields, P. A. *et al.* Dynamic reorganization of nuclear architecture during human cardiogenesis. *bioRxiv* (2017).
34 Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240-244, doi:10.1038/nature14450 (2015).
35 Grow, E. J. *et al.* Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221-225, doi:10.1038/nature14308 (2015).
36 Goke, J. *et al.* Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* **16**, 135-141, doi:10.1016/j.stem.2015.01.005 (2015).
37 Lu, X. *et al.* The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol* **21**, 423-425, doi:10.1038/nsmb.2799 (2014).
38 Romer, C., Singh, M., Hurst, L. D. & Izsvak, Z. How to tame an endogenous retrovirus: HERVH and the evolution of human pluripotency. *Curr Opin Virol* **25**, 49-58, doi:10.1016/j.coviro.2017.07.001 (2017).
39 Santoni, F. A., Guerra, J. & Luban, J. HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. *Retrovirology* **9**, 111, doi:10.1186/1742-4690-9-111 (2012).
40 Ohnuki, M. *et al.* Dynamic regulation of human endogenous retroviruses mediates factor-induced reprogramming and differentiation potential. *Proc Natl Acad Sci U S A* **111**, 12426-12431, doi:10.1073/pnas.1413299111 (2014).
41 Loewer, S. *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nat Genet* **42**, 1113-1117, doi:10.1038/ng.710 (2010).
42 Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134-1148, doi:10.1016/j.cell.2013.04.022 (2013).
43 Greenwald, W. W. *et al.* Integration of phased Hi-C and molecular phenotype data to study genetic and epigenetic effects on chromatin looping. *bioRxiv* (2018).
44 Lengronne, A. *et al.* Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **430**, 573-578, doi:10.1038/nature02742 (2004).
45 Busslinger, G. A. *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503-507, doi:10.1038/nature22063 (2017).
46 Jacques, P. E., Jeyakani, J. & Bourque, G. The majority of primate-specific regulatory sequences are derived from transposable elements. *PLoS Genet* **9**, e1003504, doi:10.1371/journal.pgen.1003504 (2013).
47 Barakat, T. S. *et al.* Functional Dissection of the Enhancer Repertoire in Human Embryonic Stem Cells. *Cell Stem Cell* **23**, 276-288 e278, doi:10.1016/j.stem.2018.06.014 (2018).
48 Wang, J. *et al.* Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516**, 405-409, doi:10.1038/nature13804 (2014).
49 Ng, S. Y., Johnson, R. & Stanton, L. W. Human long non-coding RNAs promote pluripotency and neuronal differentiation by association with chromatin modifiers and transcription factors. *EMBO J* **31**, 522-533, doi:10.1038/emboj.2011.459 (2012).
50 Liu, J., Li, Y., Lin, B., Sheng, Y. & Yang, L. HBL1 Is a Human Long Noncoding RNA that Modulates Cardiomyocyte Development from Pluripotent Stem Cells by Counteracting MIR1. *Dev Cell* **42**, 333-348 e335, doi:10.1016/j.devcel.2017.07.023 (2017).
51 Goodchild, N. L., Wilkinson, D. A. & Mager, D. L. Recent evolutionary expansion of a subfamily of RTVL-H human endogenous retrovirus-like elements. *Virology* **196**, 778-788, doi:10.1006/viro.1993.1535 (1993).
52 Mager, D. L. & Freeman, J. D. HERV-H endogenous retroviruses: presence in the New World branch but amplification in the Old World primate lineage. *Virology* **213**, 395-404, doi:10.1006/viro.1995.0012 (1995).
53 Kijima, T. E. & Innan, H. On the estimation of the insertion time of LTR retrotransposable elements. *Mol Biol Evol* **27**, 896-904, doi:10.1093/molbev/msp295 (2010).

879 54 McCulley, D. J. & Black, B. L. Transcription factor pathways and congenital heart disease. *Curr Top Dev*
880 *Biol* **100**, 253-277, doi:10.1016/B978-0-12-387786-4.00008-7 (2012).
881 55 Bantignies, F. *et al.* Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. *Cell*
882 **144**, 214-226, doi:10.1016/j.cell.2010.12.026 (2011).
883 56 Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem
884 cell genome. *Nat Genet* **47**, 1179-1186, doi:10.1038/ng.3393 (2015).
885 57 van der Harst, P. *et al.* 52 Genetic Loci Influencing Myocardial Mass. *J Am Coll Cardiol* **68**, 1435-1448,
886 doi:10.1016/j.jacc.2016.07.729 (2016).
887 58 Pfeiffer, M. J. *et al.* Cardiogenic programming of human pluripotent stem cells by dose-controlled
888 activation of EOMES. *Nat Commun* **9**, 440, doi:10.1038/s41467-017-02812-6 (2018).
889 59 Cohen, E. D., Tian, Y. & Morrisey, E. E. Wnt signaling: an essential regulator of cardiovascular
890 differentiation, morphogenesis and progenitor self-renewal. *Development* **135**, 789-798,
891 doi:10.1242/dev.016865 (2008).
892 60 Liao, X. *et al.* Kruppel-like factor 4 is critical for transcriptional control of cardiac mitochondrial
893 homeostasis. *J Clin Invest* **125**, 3461-3476, doi:10.1172/JCI79964 (2015).
894 61 Cordell, H. J. *et al.* Genome-wide association study of multiple congenital heart disease phenotypes
895 identifies a susceptibility locus for atrial septal defect at chromosome 4p16. *Nat Genet* **45**, 822-824,
896 doi:10.1038/ng.2637 (2013).
897 62 Boogerd, K. J. *et al.* Msx1 and Msx2 are functional interacting partners of T-box factors in the regulation
898 of Connexin43. *Cardiovasc Res* **78**, 485-493, doi:10.1093/cvr/cvn049 (2008).
899 63 Li, Q. Y. *et al.* Holt-Oram syndrome is caused by mutations in TBX5, a member of the Brachyury (T) gene
900 family. *Nat Genet* **15**, 21-29, doi:10.1038/ng0197-21 (1997).
901 64 Britten, R. J. & Davidson, E. H. Repetitive and non-repetitive DNA sequences and a speculation on the
902 origins of evolutionary novelty. *Q Rev Biol* **46**, 111-138 (1971).
903 65 King, M. C. & Wilson, A. C. Evolution at two levels in humans and chimpanzees. *Science* **188**, 107-116
904 (1975).
905 66 Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory evolution of innate immunity through co-option of
906 endogenous retroviruses. *Science* **351**, 1083-1087, doi:10.1126/science.aad5497 (2016).
907 67 Glinsky, G. V. Contribution of transposable elements and distal enhancers to evolution of human-specific
908 features of interphase chromatin architecture in embryonic stem cells. *Chromosome Res* **26**, 61-84,
909 doi:10.1007/s10577-018-9571-6 (2018).
910 68 Wang, Y. *et al.* Endogenous miRNA sponge lincRNA-RoR regulates Oct4, Nanog, and Sox2 in human
911 embryonic stem cell self-renewal. *Dev Cell* **25**, 69-80, doi:10.1016/j.devcel.2013.03.002 (2013).
912

Figure 1

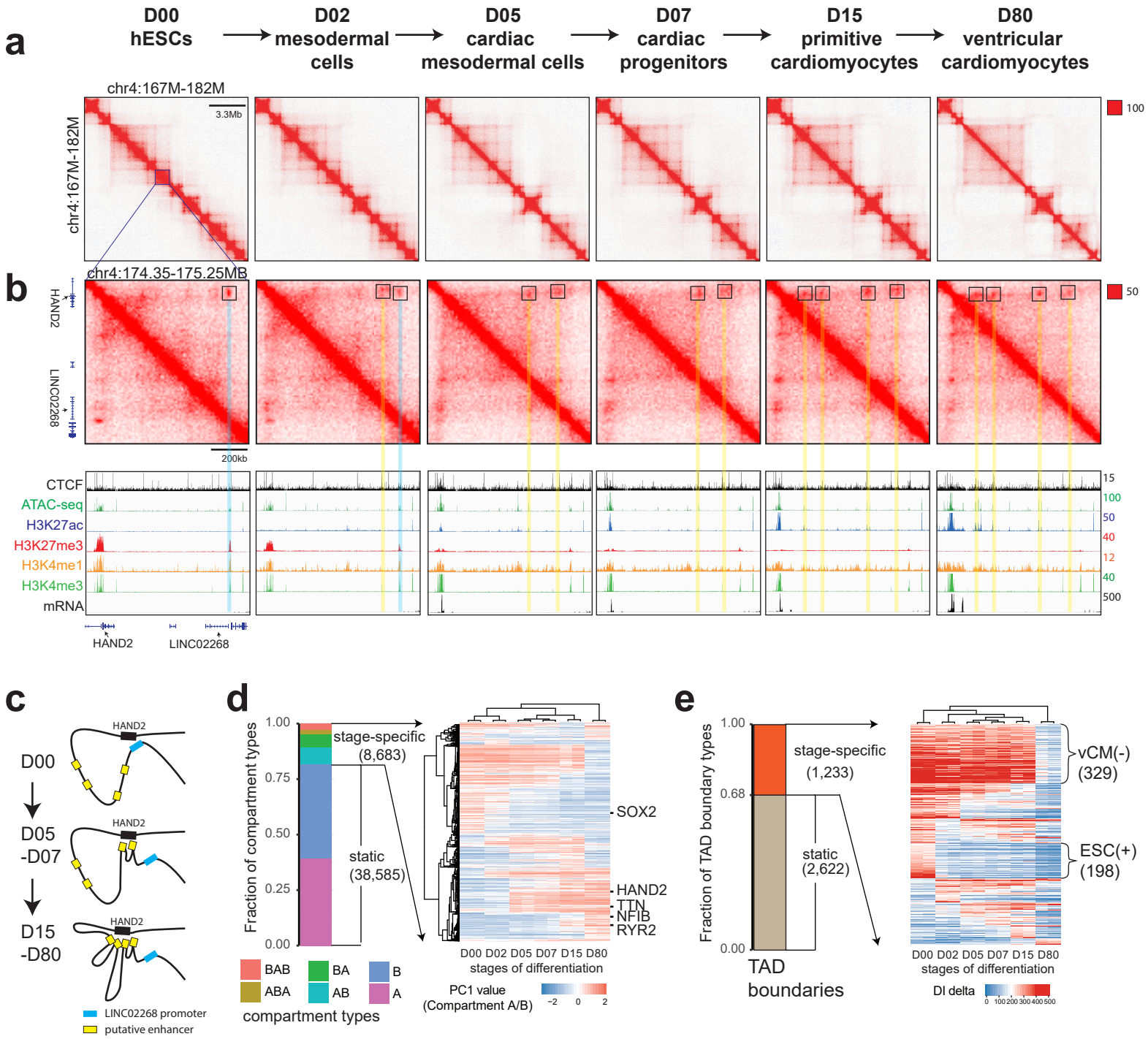

a
D00 hESCs
D02 mesodermal cells
D05 cardiac mesodermal cells
D07 cardiac progenitors
D15 primitive cardiomyocytes
D80 ventricular cardiomyocytes
chr4:167M-182M
3.3Mb
100
b
chr4:174.35-175.25MB
HAND2
LINC02268
50
200kb
CTCF
ATAC-seq
H3K27ac
H3K27me3
H3K4me1
H3K4me3
mRNA
15
100
50
40
12
40
500
c
D00
D05 -D07
D15 -D80
HAND2
LINC02268 promoter
putative enhancer
d
Fraction of compartment types
1.00
0.75
0.50
0.25
0.00
stage-specific (8,683)
static (38,585)
BAB
BA
B
ABA
AB
A
compartment types
SOX2
HAND2
TTN
NFIB
RYR2
D00 D02 D05 D07 D15 D80
stages of differentiation
PC1 value (Compartment A/B)
−2 0 2
e
Fraction of TAD boundary types
1.00
0.68
0.00
stage-specific (1,233)
static (2,622)
TAD boundaries
vCM(-) (329)
ESC(+) (198)
D00 D02 D05 D07 D15 D80
stages of differentiation
DI delta
0 100 200 300 400 500

Figure 2

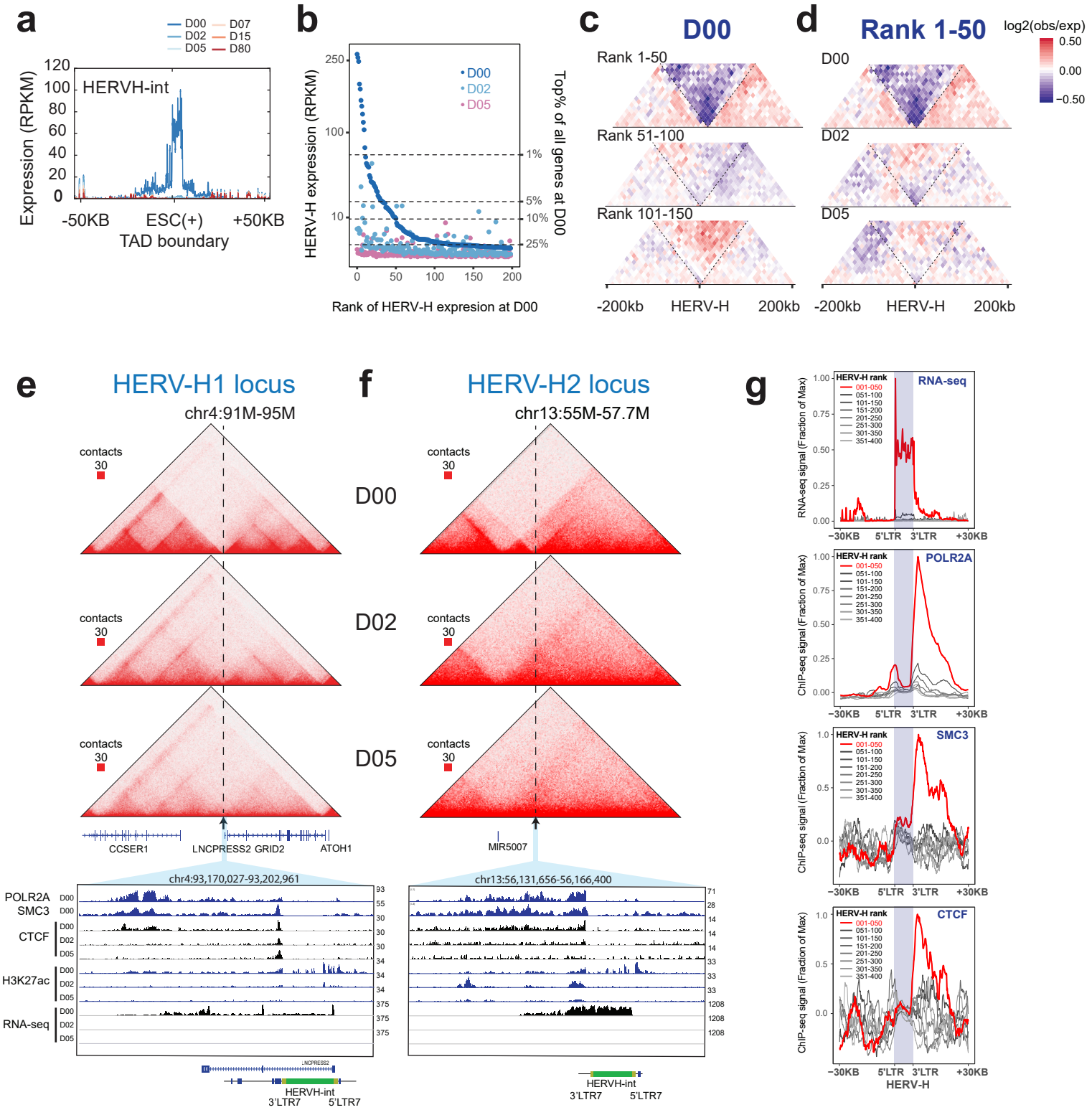

a
D00
D02
D05
D07
D15
D80
HERVH-int
Expression (RPKM)
-50KB
ESC(+)
+50KB
TAD boundary
b
D00
D02
D05
HERV-H expression (RPKM)
Top% of all genes at D00
1%
5%
10%
25%
Rank of HERV-H expresion at D00
c
D00
Rank 1-50
Rank 51-100
Rank 101-150
-200kb
HERV-H
200kb
d
Rank 1-50
log2(obs/exp)
D00
D02
D05
e
HERV-H1 locus
chr4:91M-95M
contacts
30
CCSER1
LNCPRESS2 GRID2
ATOH1
chr4:93,170,027-93,202,961
POLR2A
SMC3
CTCF
H3K27ac
RNA-seq
HERVH-int
3'LTR7
5'LTR7
f
HERV-H2 locus
chr13:55M-57.7M
MIR5007
chr13:56,131,656-56,166,400
g
RNA-seq
POLR2A
SMC3
CTCF
HERV-H rank
RNA-seq signal (Fraction of Max)
ChIP-seq signal (Fraction of Max)
-30KB
5'LTR
3'LTR
+30KB
HERV-H

Figure 3

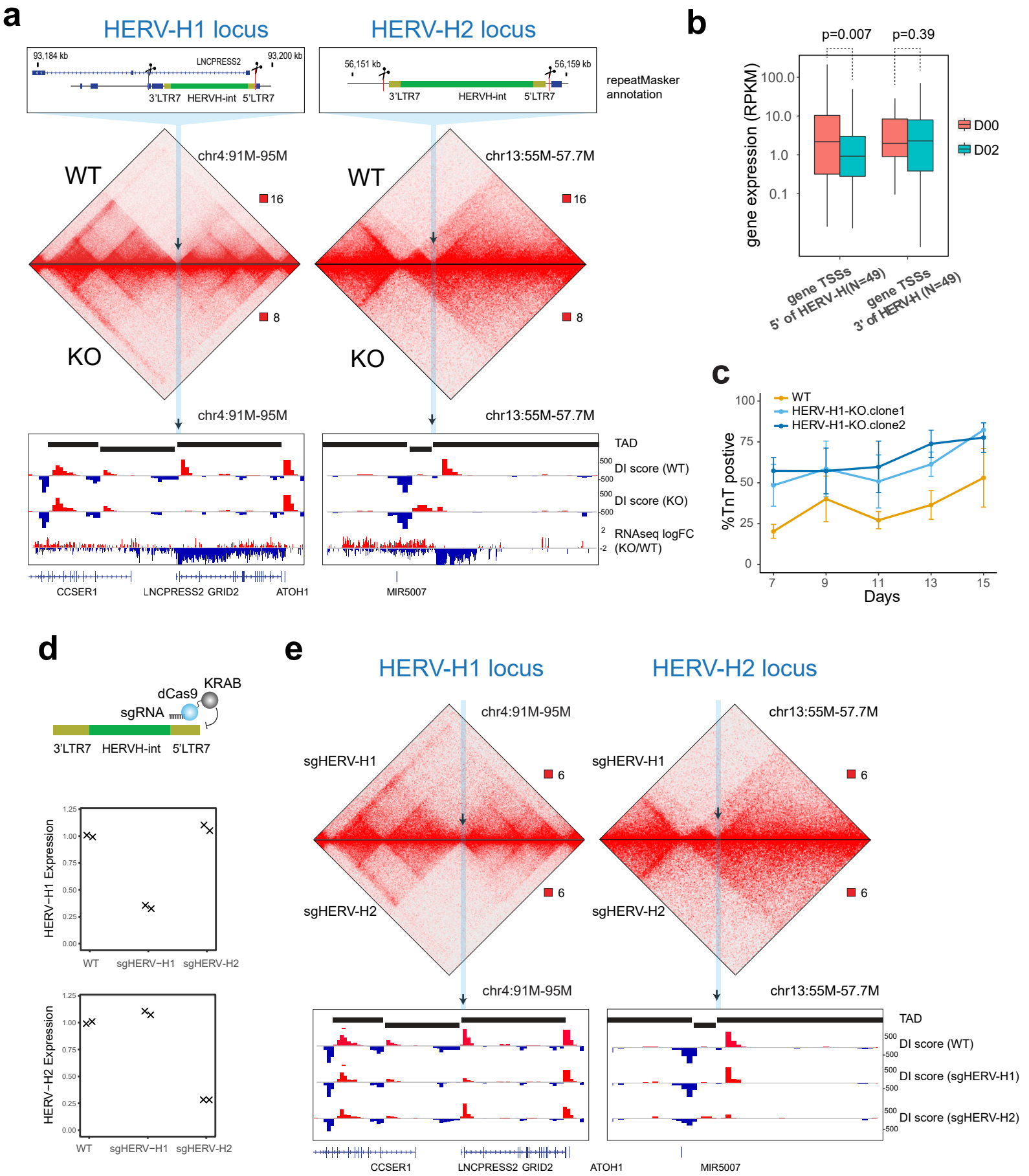

a
HERV-H1 locus
HERV-H2 locus
93,184 kb
LNCPRESS2
93,200 kb
3'LTR7 HERVH-int 5'LTR7
56,151 kb
56,159 kb
repeatMasker annotation
WT
KO
chr4:91M-95M
chr13:55M-57.7M
16
8
TAD
DI score (WT)
DI score (KO)
RNAseq logFC (KO/WT)
CCSER1
LNCPRESS2 GRID2
ATOH1
MIR5007
b
p=0.007
p=0.39
gene expression (RPKM)
100.0
10.0
1.0
0.1
D00
D02
gene TSSs 5' of HERV-H(N=49)
gene TSSs 3' of HERV-H (N=49)
c
WT
HERV-H1-KO.clone1
HERV-H1-KO.clone2
%TnT postive
Days
d
KRAB
dCas9
sgRNA
3'LTR7 HERVH-int 5'LTR7
HERV−H1 Expression
HERV−H2 Expression
WT
sgHERV−H1
sgHERV-H2
e
HERV-H1 locus
HERV-H2 locus
sgHERV-H1
sgHERV-H2
chr4:91M-95M
chr13:55M-57.7M
TAD
DI score (WT)
DI score (sgHERV-H1)
DI score (sgHERV-H2)
CCSER1
LNCPRESS2 GRID2
ATOH1
MIR5007

Figure 4

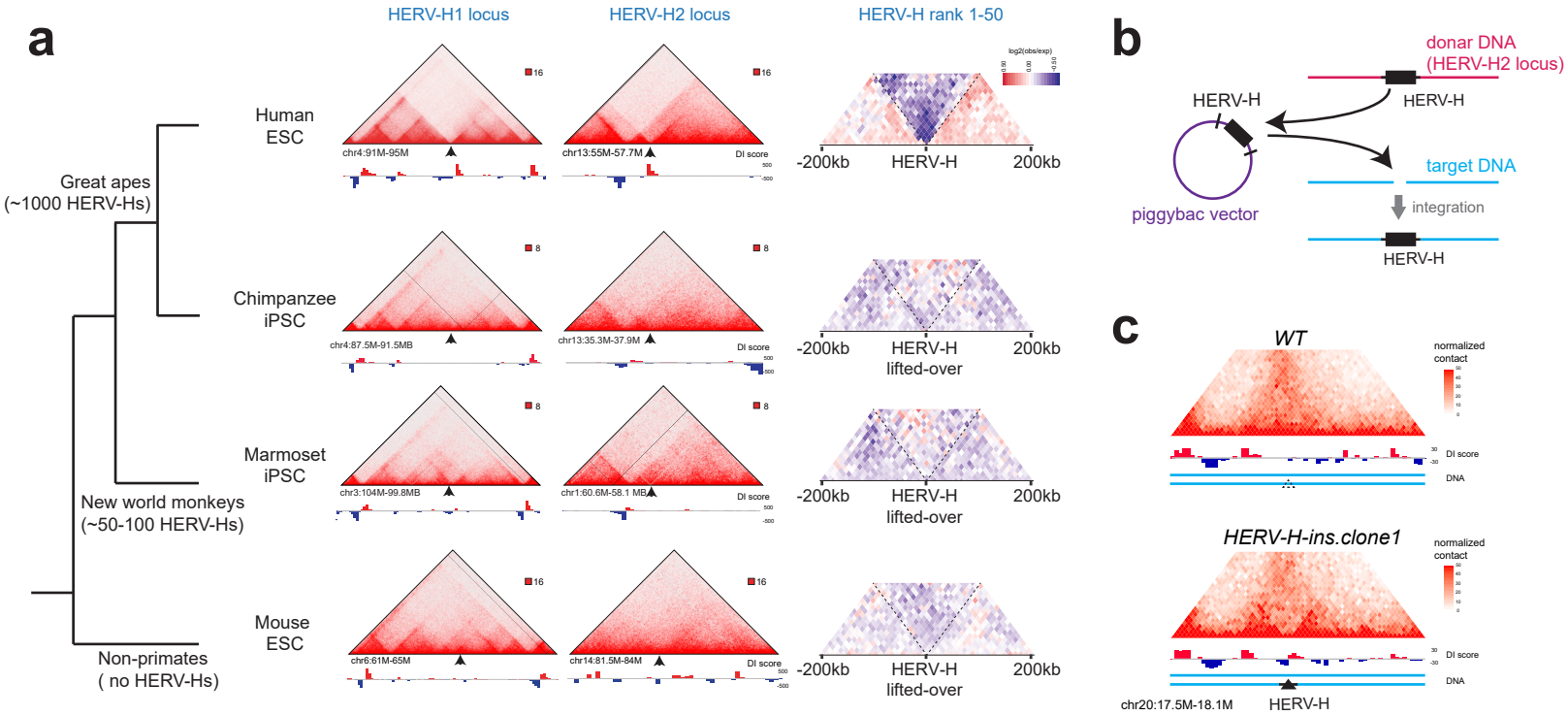

a
HERV-H1 locus
HERV-H2 locus
HERV-H rank 1-50
Great apes
(~1000 HERV-Hs)
New world monkeys
(~50-100 HERV-Hs)
Non-primates
(no HERV-Hs)
Human
ESC
Chimpanzee
iPSC
Marmoset
iPSC
Mouse
ESC
chr4:91M-95M
chr13:55M-57.7M
chr4:87.5M-91.5MB
chr13:35.3M-37.9M
chr3:104M-99.8MB
chr1:60.6M-58.1 MB
chr6:61M-65M
chr14:81.5M-84M
DI score
log2(obs/exp)
-200kb
HERV-H
200kb
lifted-over
b
donar DNA
(HERV-H2 locus)
HERV-H
target DNA
integration
piggybac vector
c
WT
HERV-H-ins.clone1
normalized
contact
DI score
DNA
chr20:17.5M-18.1M

Figure 5

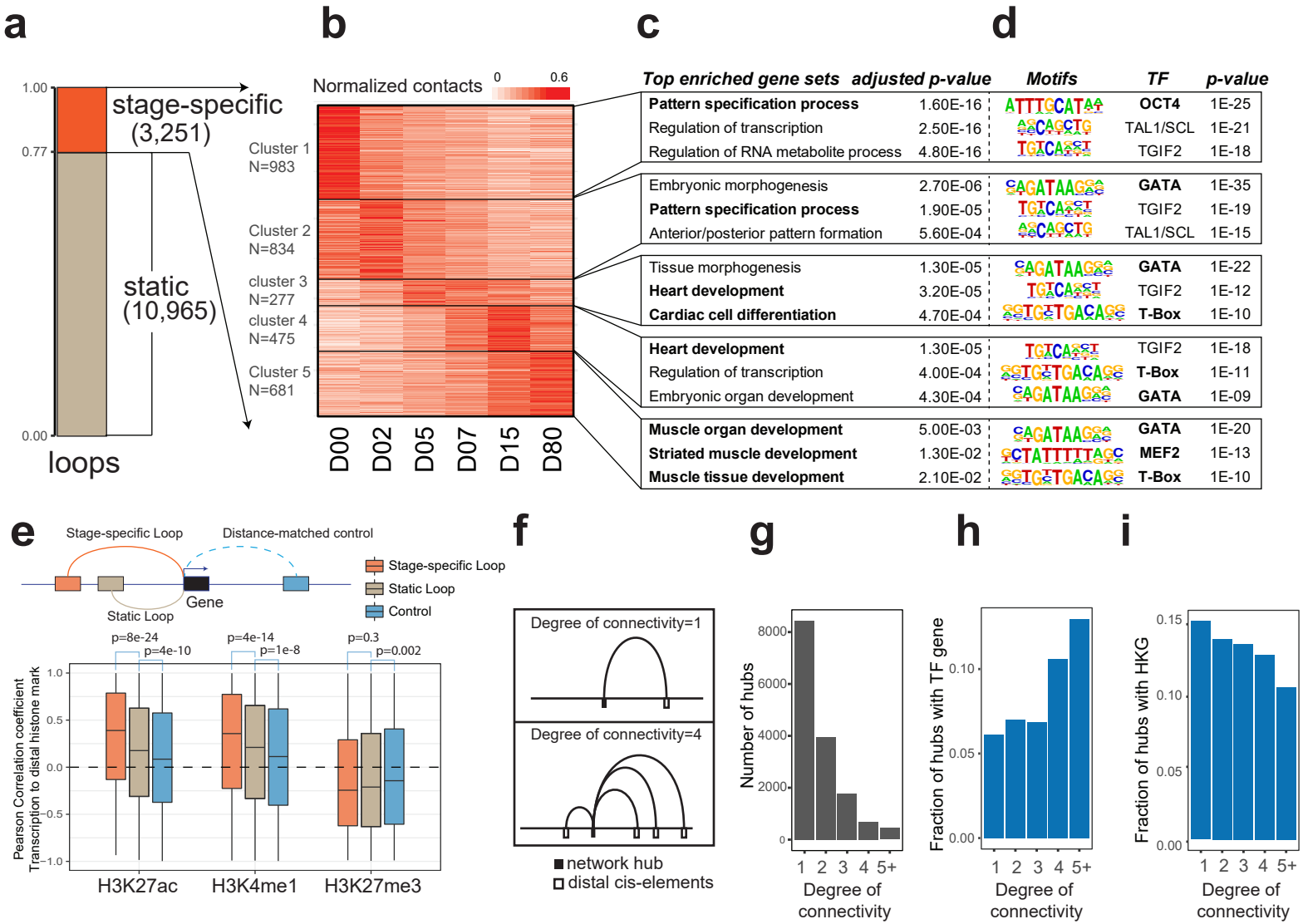
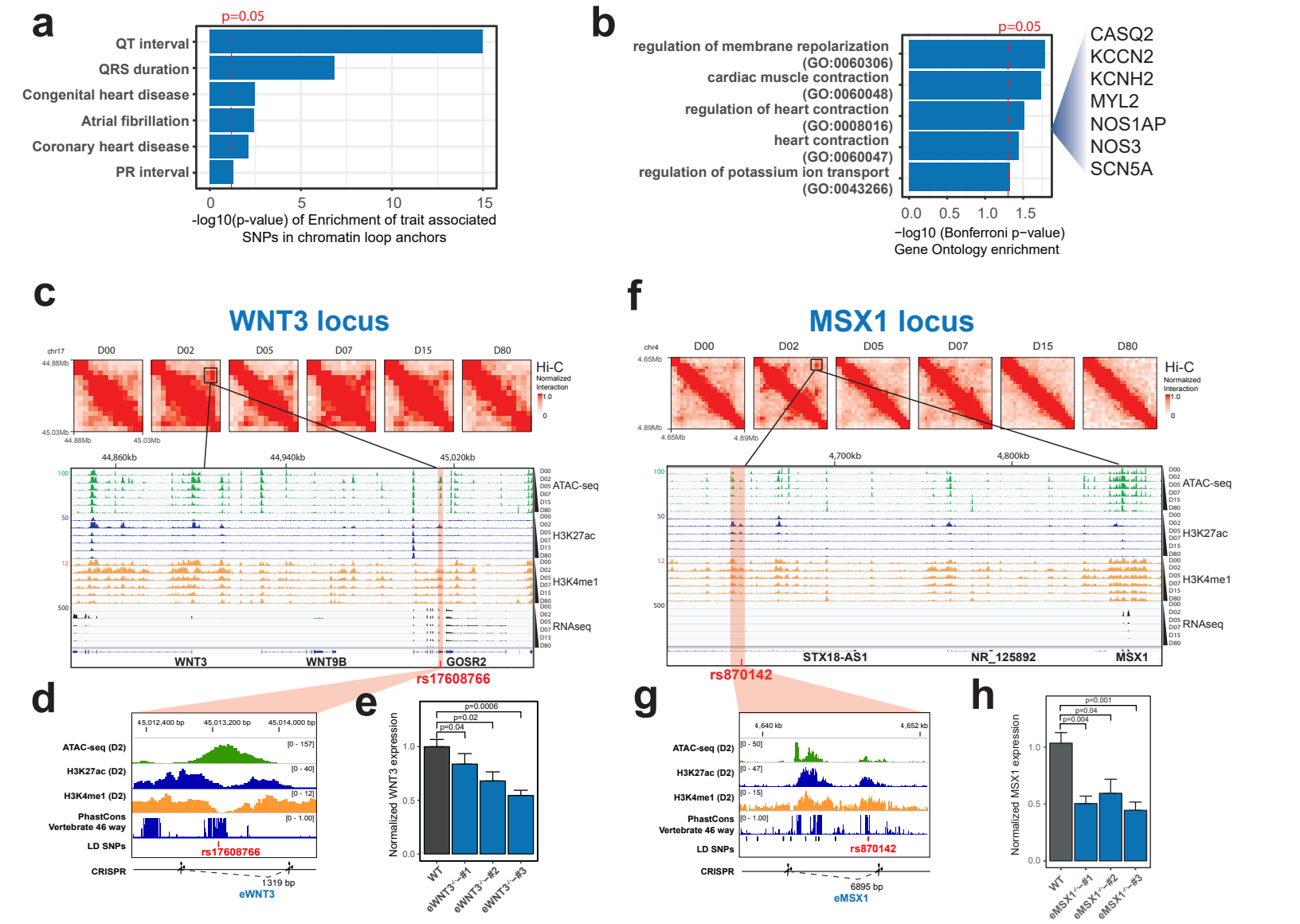
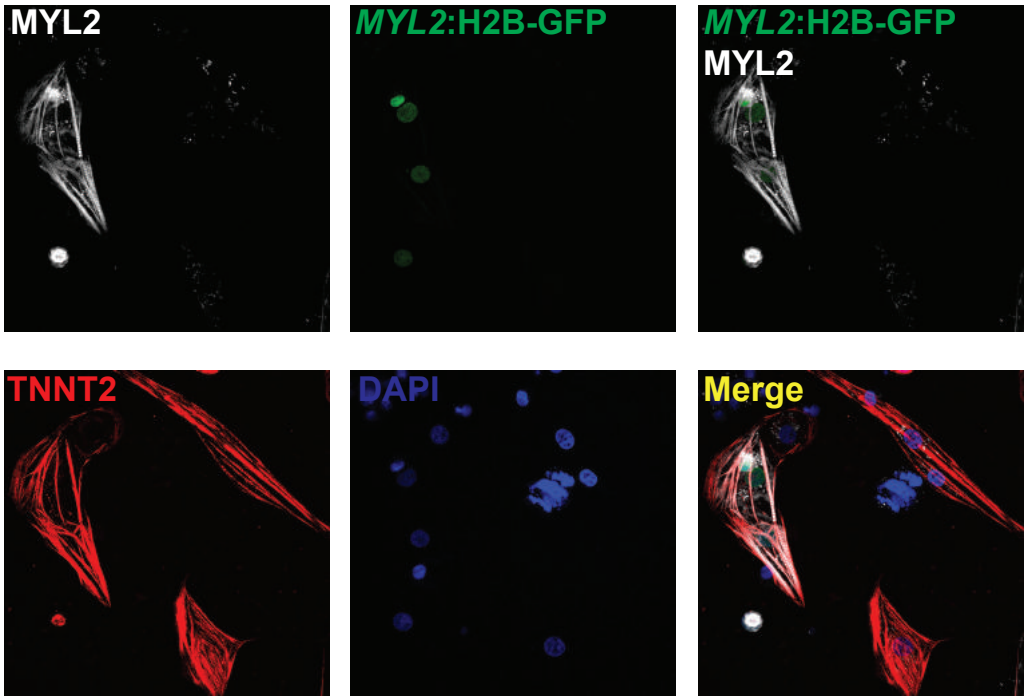

a
1.00
0.77
0.00
stage-specific
(3,251)
static
(10,965)
loops
b
Normalized contacts
0
0.6
Cluster 1
N=983
Cluster 2
N=834
cluster 3
N=277
cluster 4
N=475
Cluster 5
N=681
D00
D02
D05
D07
D15
D80
c
d
Top enriched gene sets
adjusted p-value
Motifs
TF
p-value
Pattern specification process 1.60E-16 OCT4 1E-25
Regulation of transcription 2.50E-16 TAL1/SCL 1E-21
Regulation of RNA metabolite process 4.80E-16 TGIF2 1E-18
Embryonic morphogenesis 2.70E-06 GATA 1E-35
Pattern specification process 1.90E-05 TGIF2 1E-19
Anterior/posterior pattern formation 5.60E-04 TAL1/SCL 1E-15
Tissue morphogenesis 1.30E-05 GATA 1E-22
Heart development 3.20E-05 TGIF2 1E-12
Cardiac cell differentiation 4.70E-04 T-Box 1E-10
Heart development 1.30E-05 TGIF2 1E-18
Regulation of transcription 4.00E-04 T-Box 1E-11
Embryonic organ development 4.30E-04 GATA 1E-09
Muscle organ development 5.00E-03 GATA 1E-20
Striated muscle development 1.30E-02 MEF2 1E-13
Muscle tissue development 2.10E-02 T-Box 1E-10
e
Stage-specific Loop
Distance-matched control
Gene
Static Loop
Stage-specific Loop
Static Loop
Control
p=8e-24
p=4e-10
p=4e-14
p=1e-8
p=0.3
p=0.002
Pearson Correlation coefficient
Transcription to distal histone mark
1.0
0.5
0.0
−0.5
−1.0
H3K27ac
H3K4me1
H3K27me3
f
Degree of connectivity=1
Degree of connectivity=4
network hub
distal cis-elements
g
Number of hubs
8000
6000
4000
2000
0
1 2 3 4 5+
Degree of connectivity
h
Fraction of hubs with TF gene
0.10
0.05
0.00
1 2 3 4 5+
Degree of connectivity
i
Fraction of hubs with HKG
0.15
0.10
0.05
0.00
1 2 3 4 5+
Degree of connectivity

Figure 6

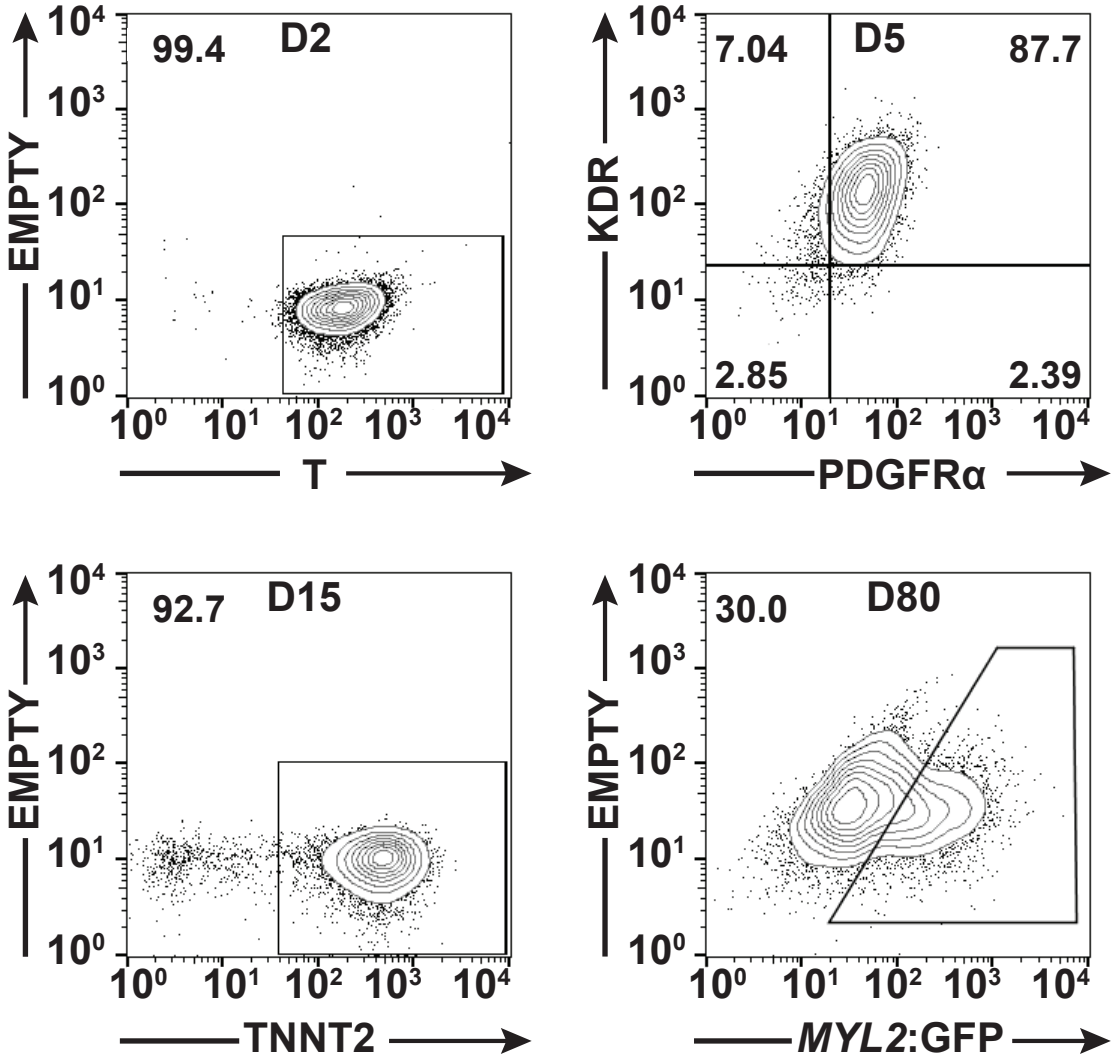

a
p=0.05
QT interval
QRS duration
Congenital heart disease
Atrial fibrillation
Coronary heart disease
PR interval
0
5
10
15
-log10(p-value) of Enrichment of trait associated SNPs in chromatin loop anchors
b
p=0.05
regulation of membrane repolarization (GO:0060306)
cardiac muscle contraction (GO:0060048)
regulation of heart contraction (GO:0008016)
heart contraction (GO:0060047)
regulation of potassium ion transport (GO:0043266)
CASQ2
KCCN2
KCNH2
MYL2
NOS1AP
NOS3
SCN5A
0.0
0.5
1.0
1.5
−log10 (Bonferroni p−value)
Gene Ontology enrichment
c
WNT3 locus
chr17
D00
D02
D05
D07
D15
D80
Hi-C
Normalized Interaction
44,860kb
44,940kb
45,020kb
ATAC-seq
H3K27ac
H3K4me1
RNAseq
WNT3
WNT9B
GOSR2
rs17608766
f
MSX1 locus
chr4
4,700kb
4,800kb
STX18-AS1
NR_125892
MSX1
rs870142
d
ATAC-seq (D2)
H3K27ac (D2)
H3K4me1 (D2)
PhastCons Vertebrate 46 way
LD SNPs
CRISPR
rs17608766
1319 bp
eWNT3
e
Normalized WNT3 expression
p=0.0006
p=0.02
p=0.04
WT
eWNT3−/−#1
eWNT3−/−#2
eWNT3−/−#3
g
rs870142
6895 bp
eMSX1
h
Normalized MSX1 expression
p=0.001
p=0.04
p=0.004
WT
eMSX1−/−#1
eMSX1−/−#2
eMSX1−/−#3

Extended Data Figure 1

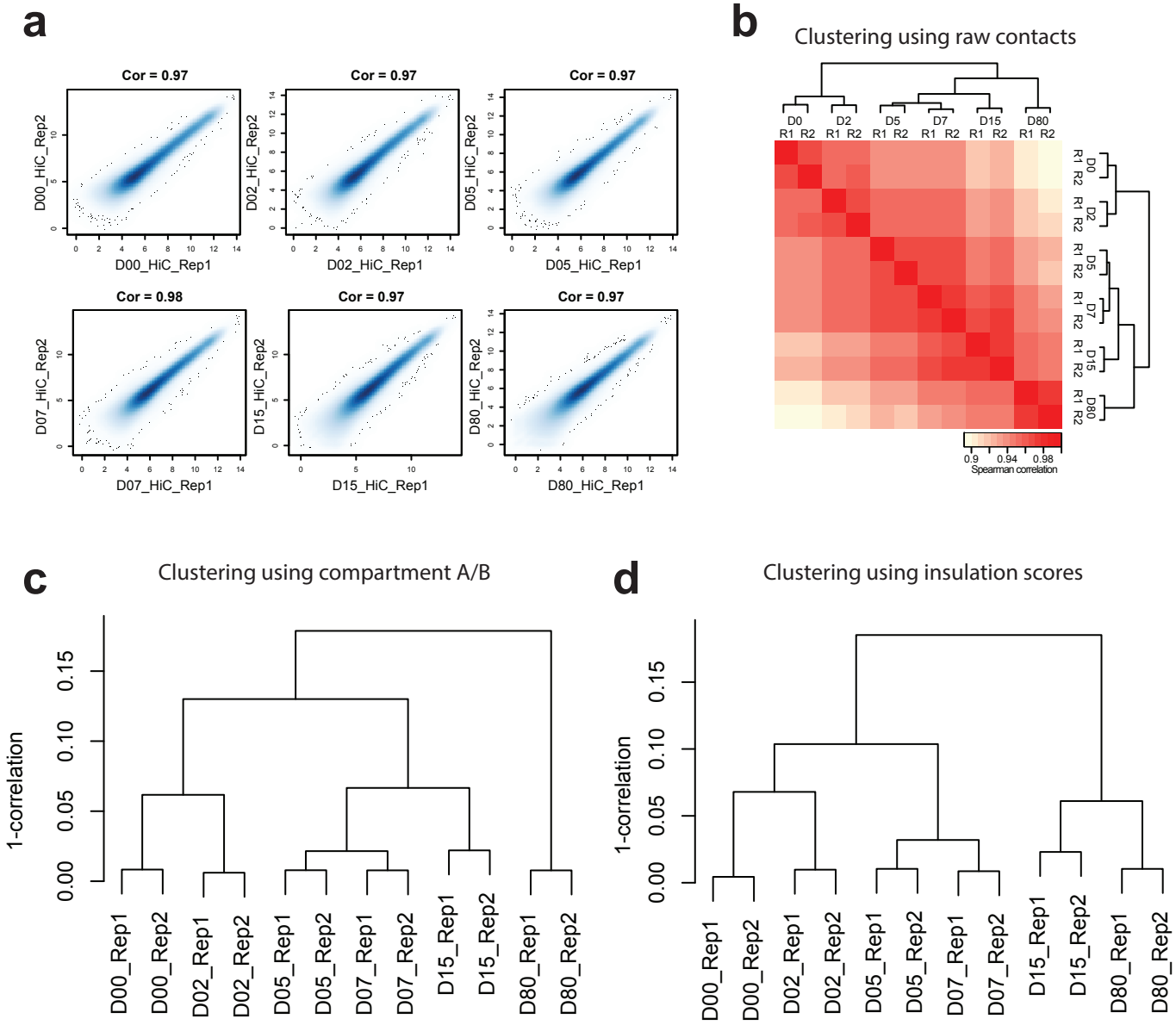
a



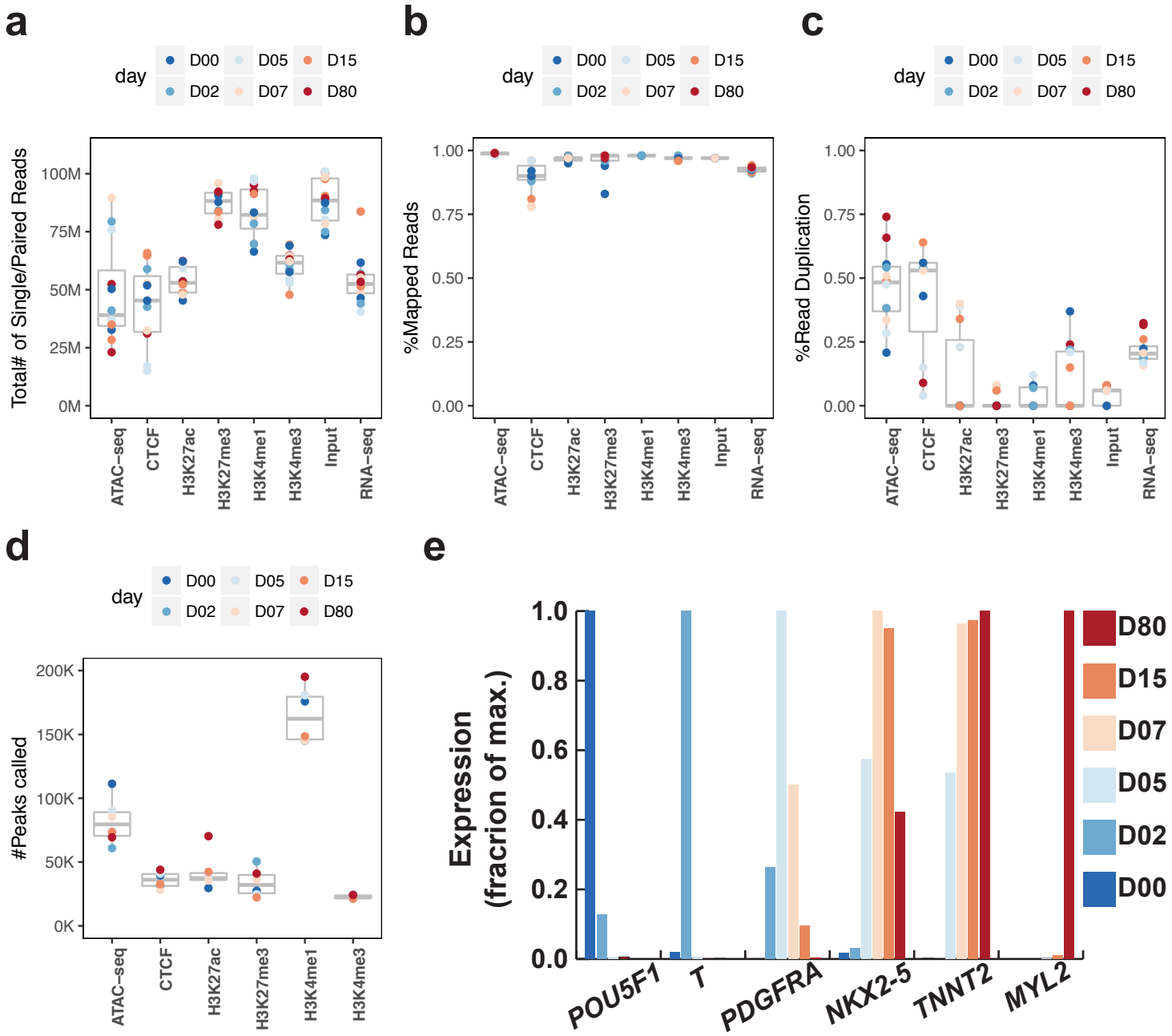
b



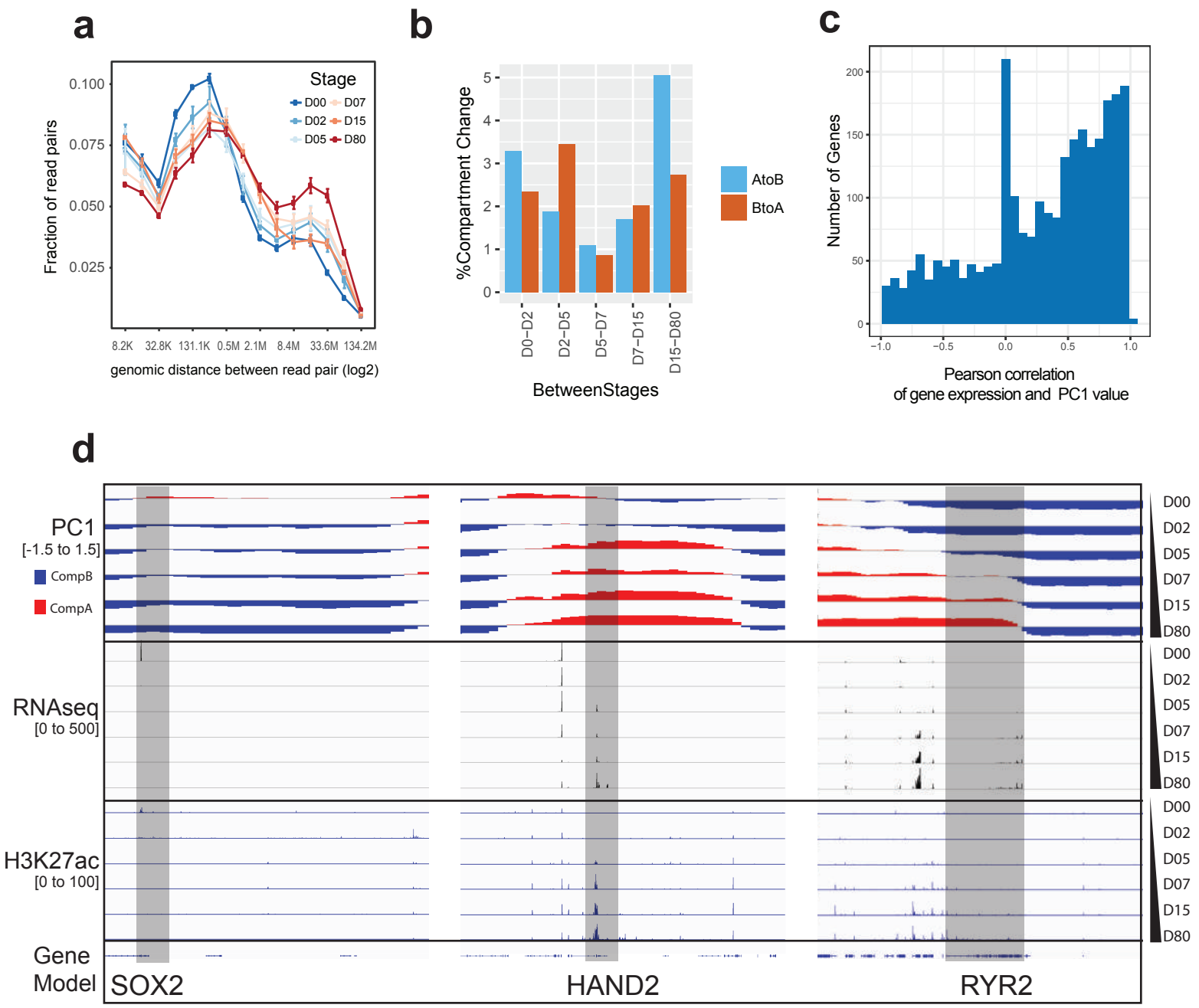
Extended Data Figure 2



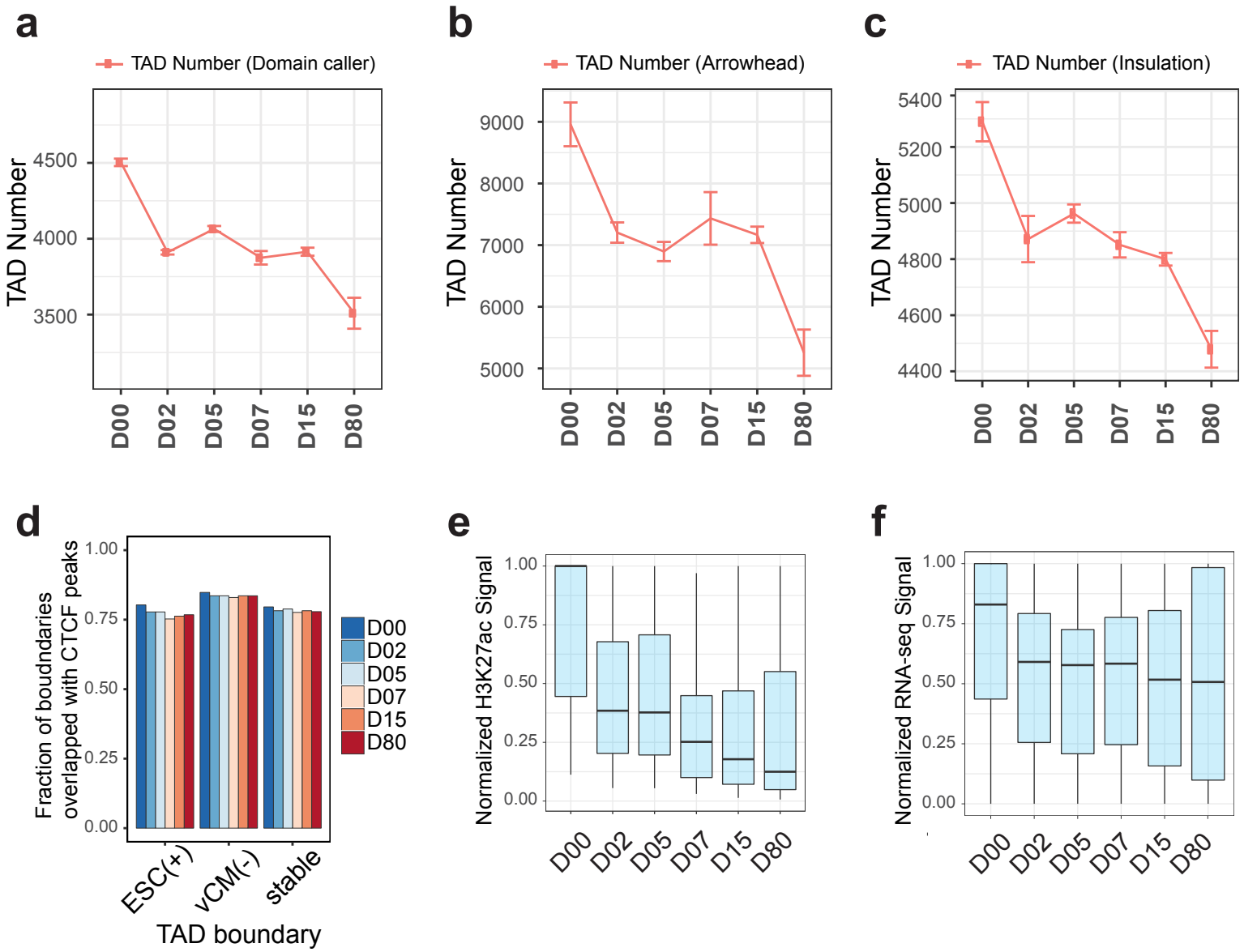
Extended Data Figure 3



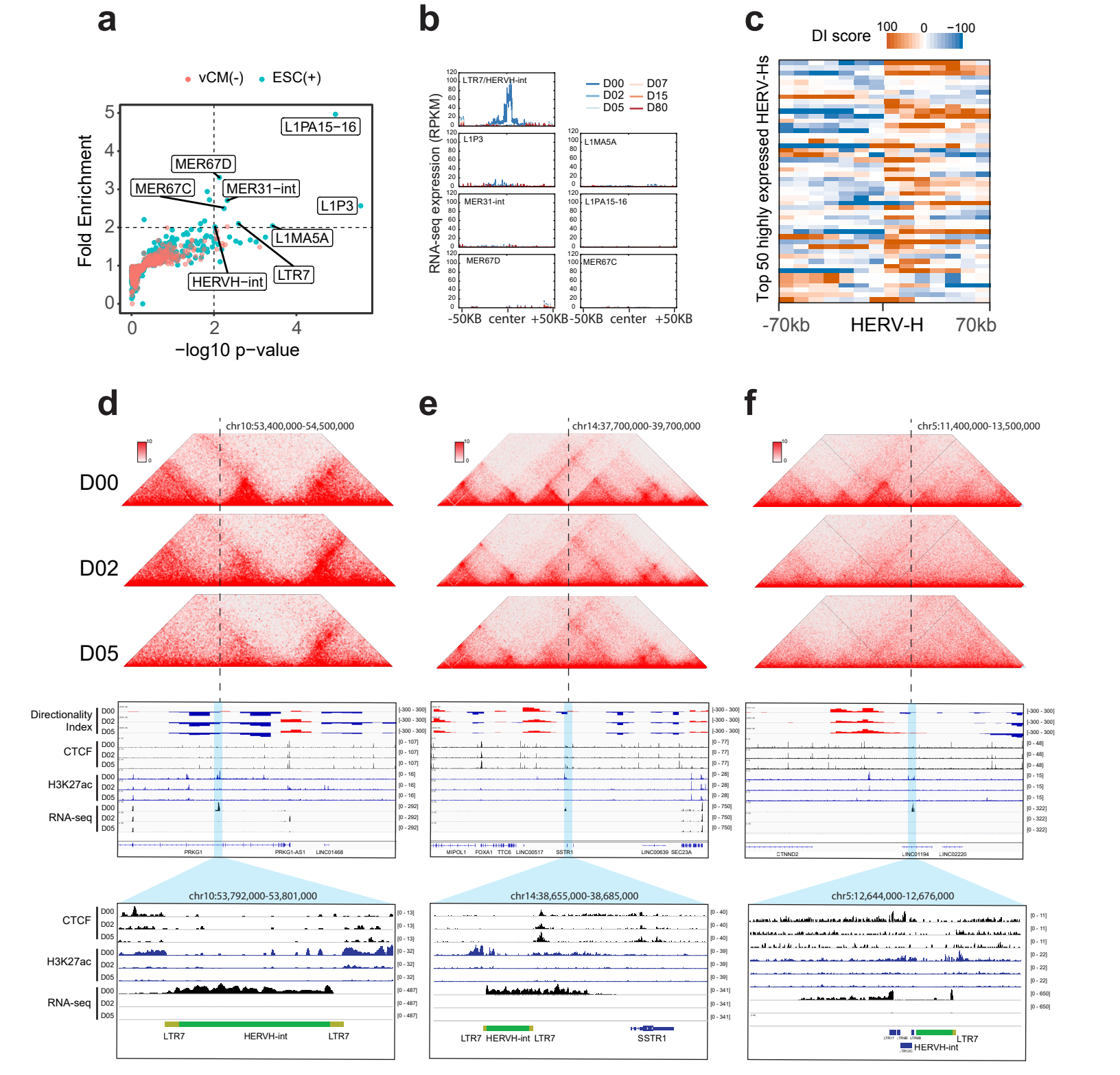
Extended Data Figure 4



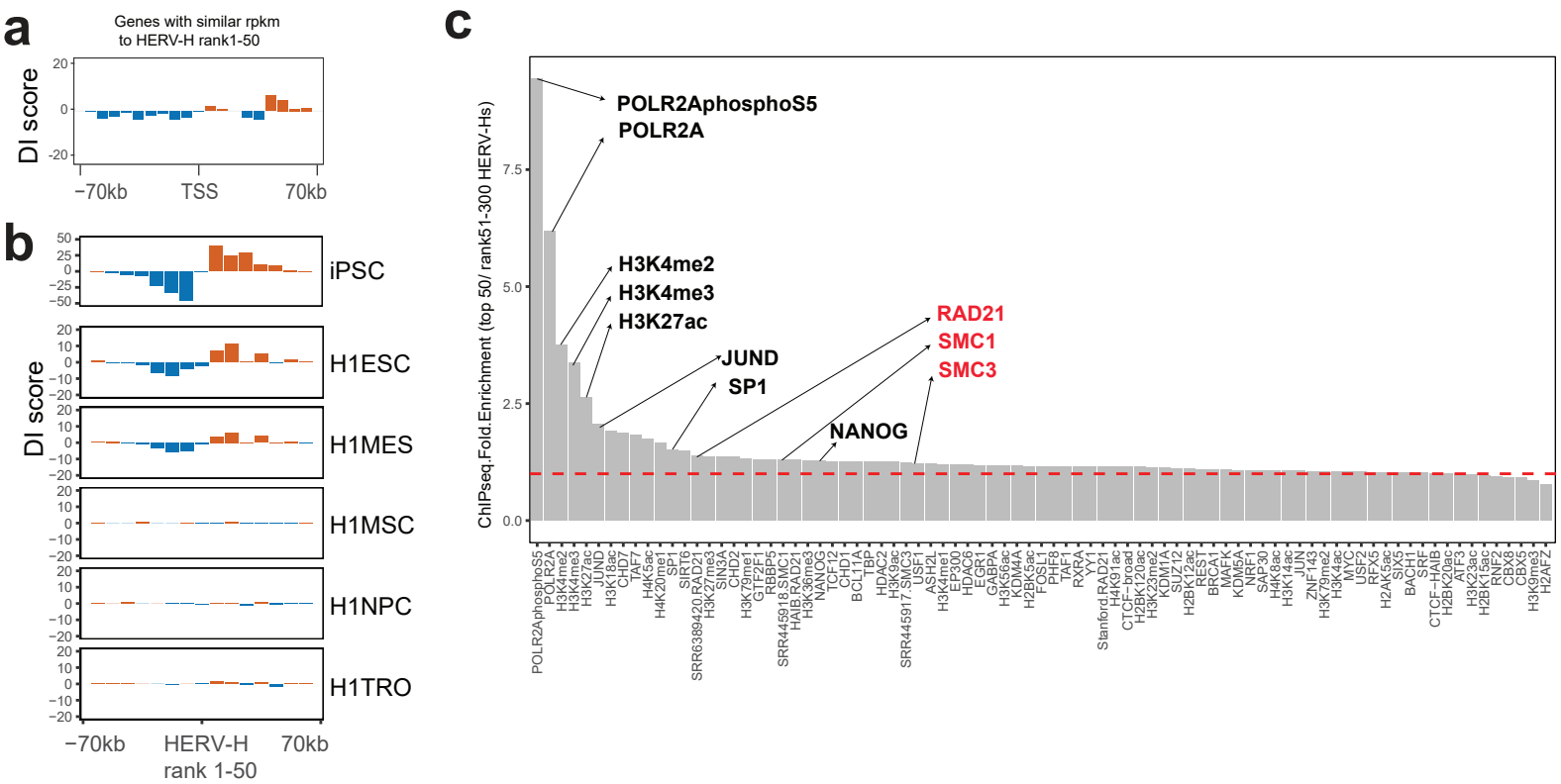
Extended Data Figure 5



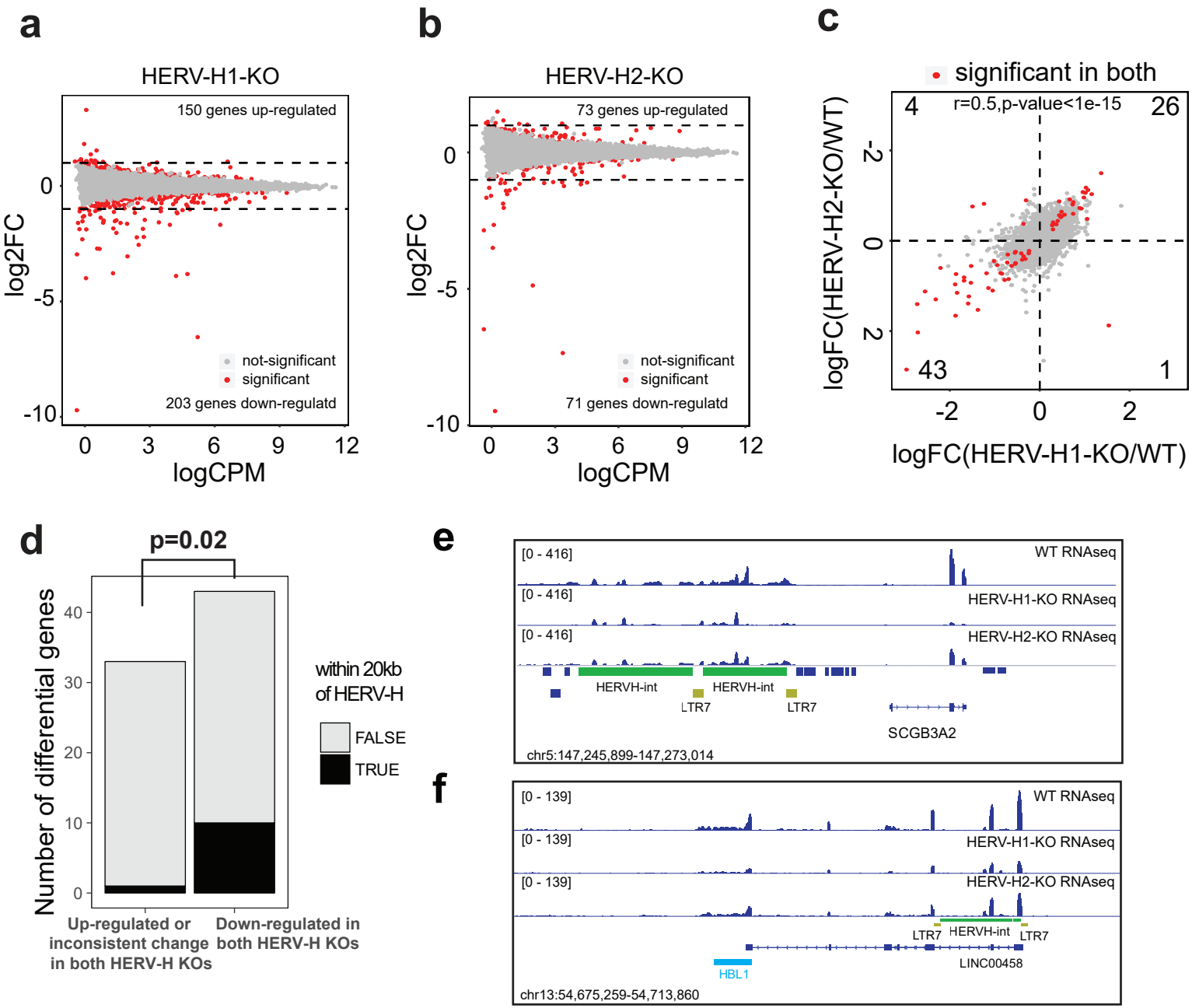
Extended Data Figure 6



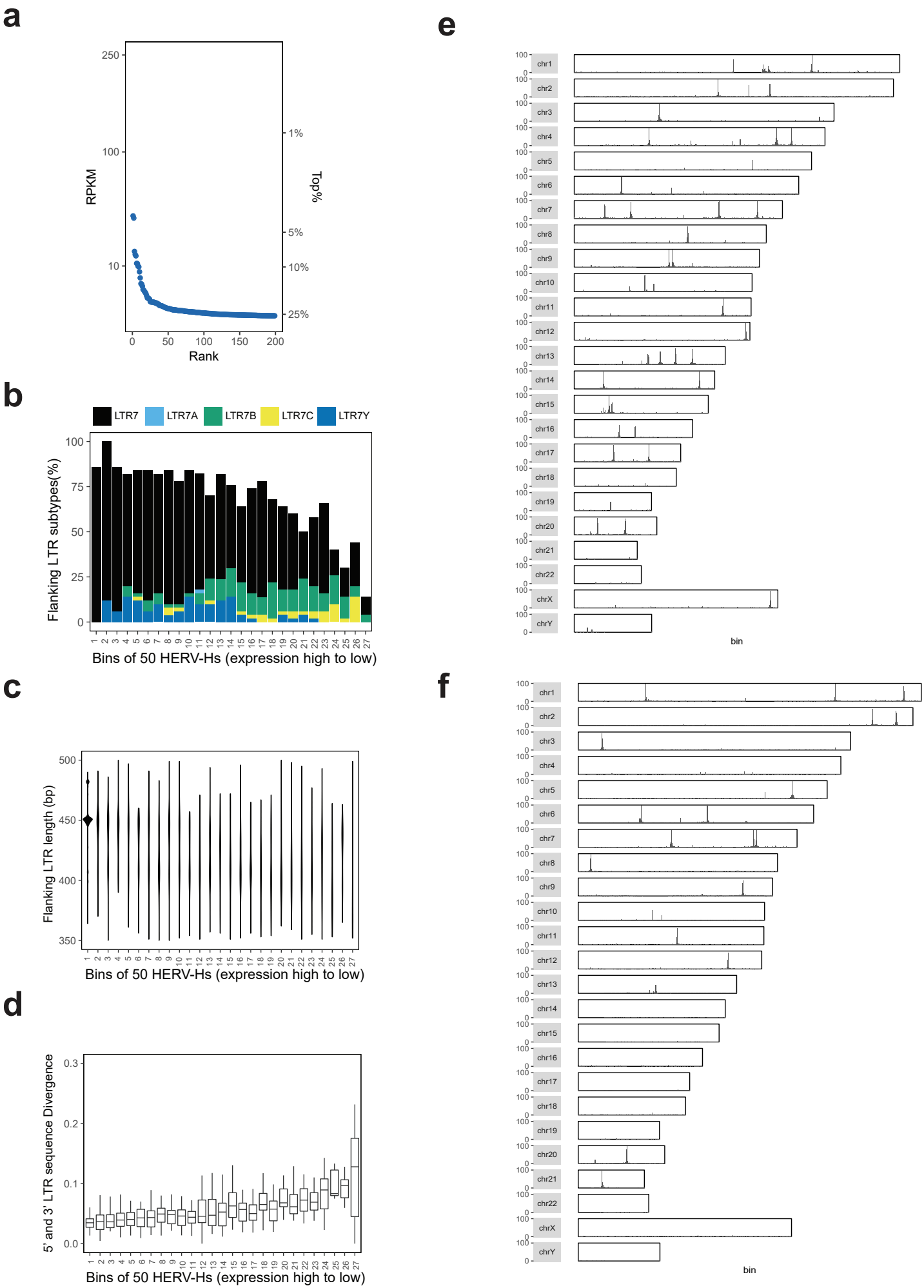
Extended Data Figure 7



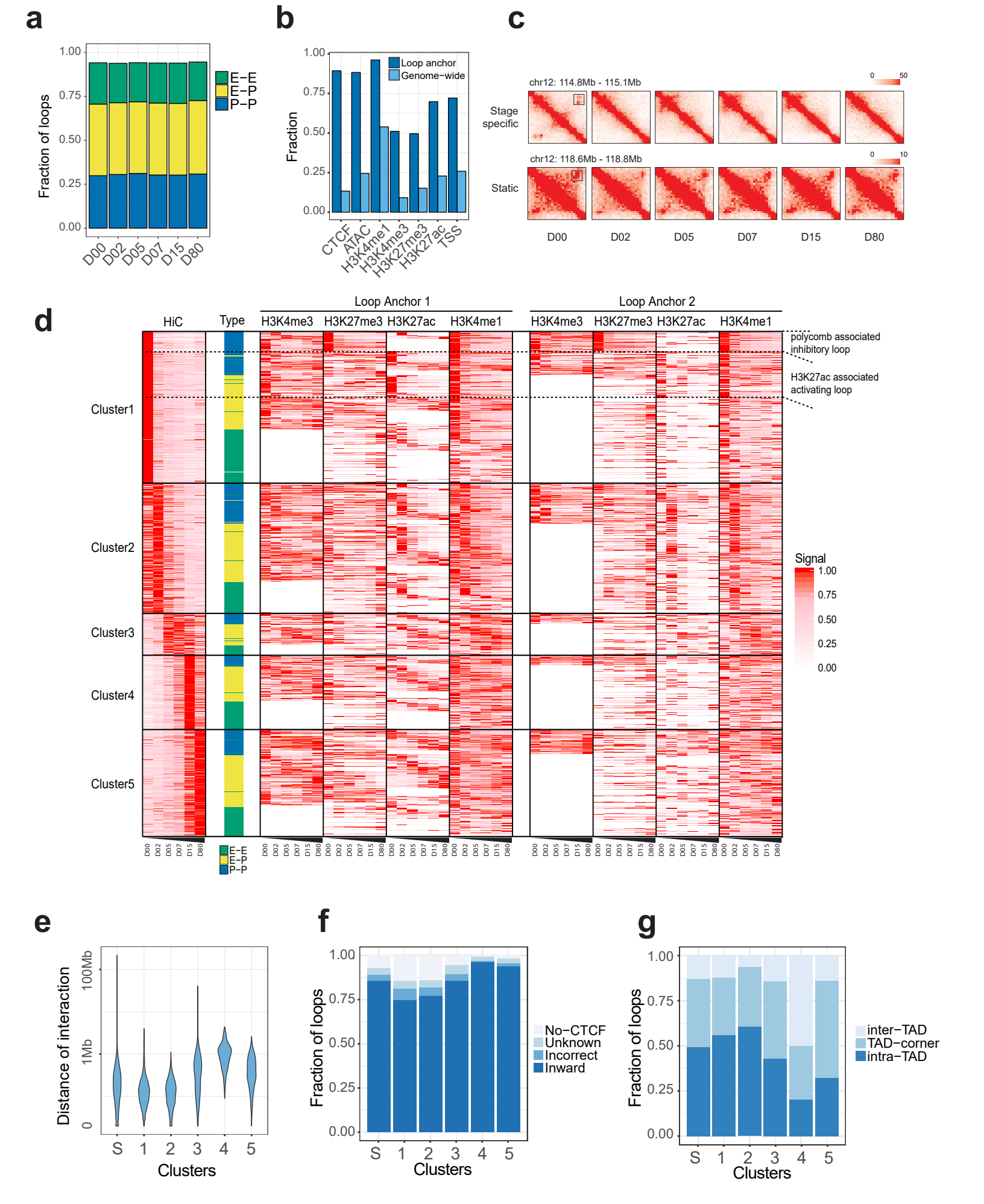
Extended Data Figure 8



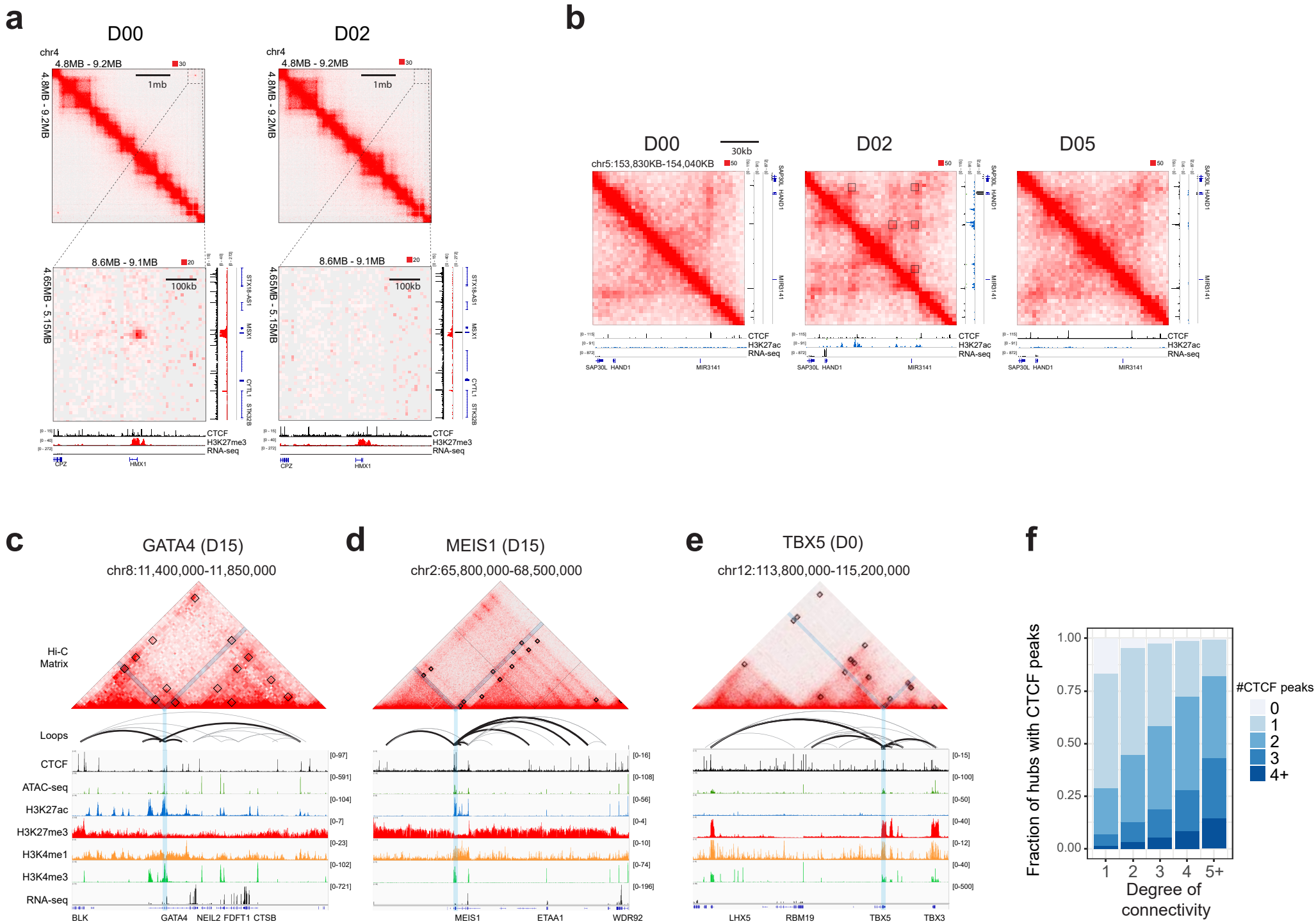
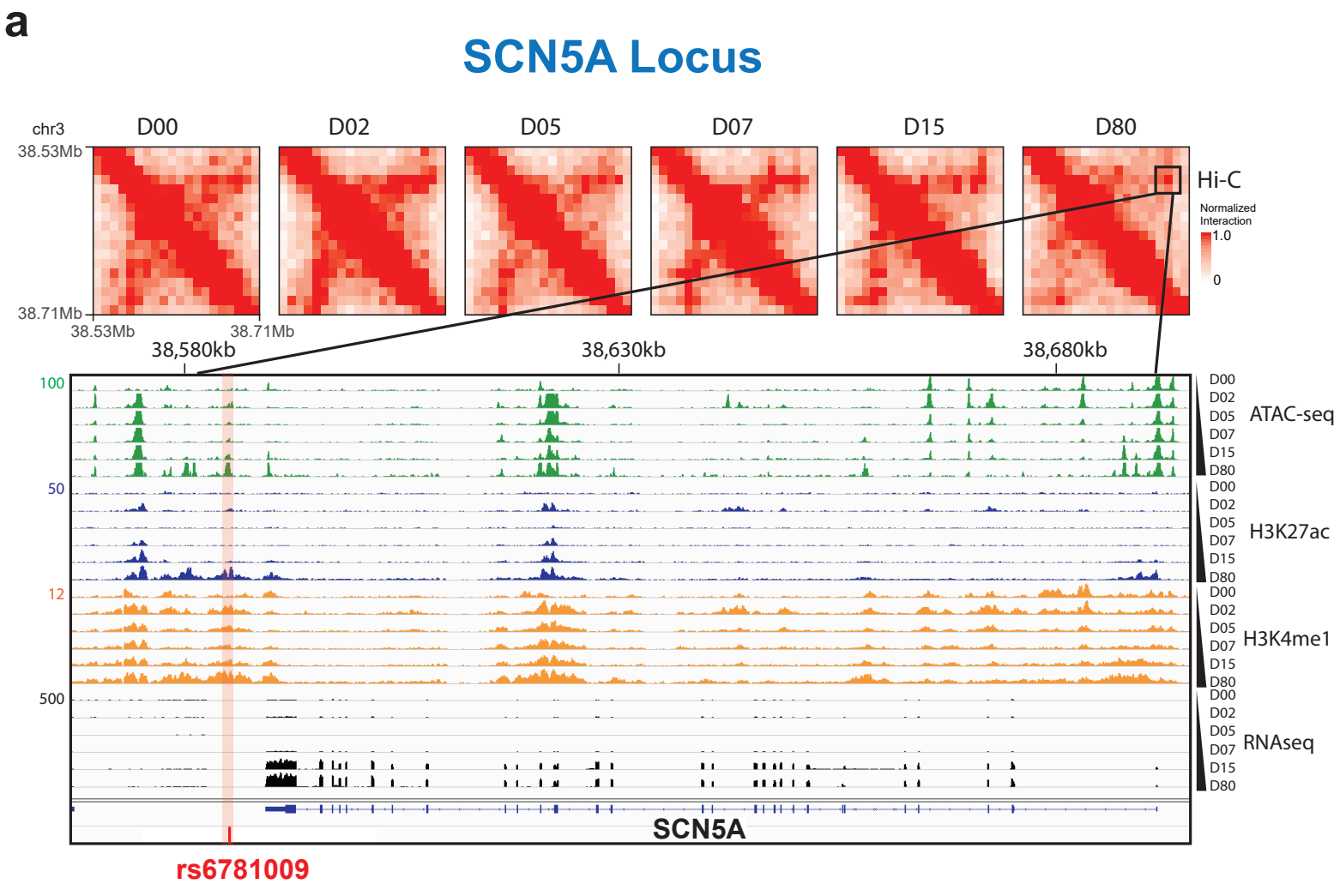
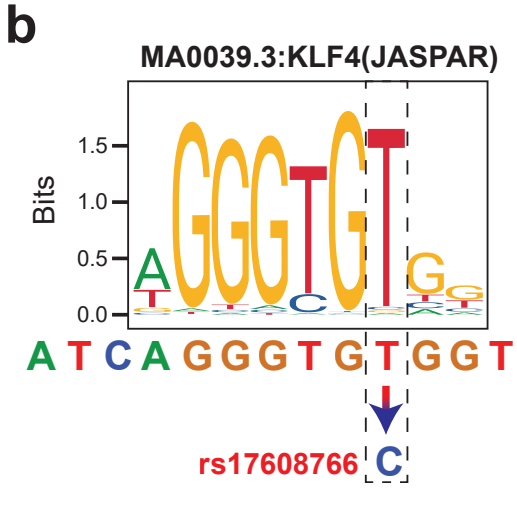
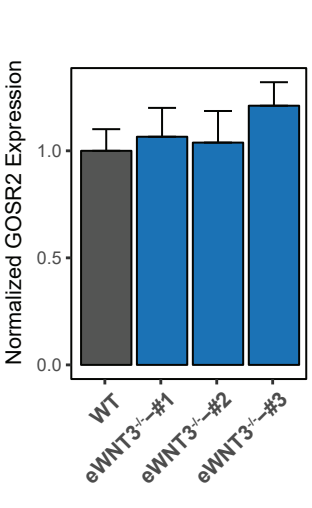
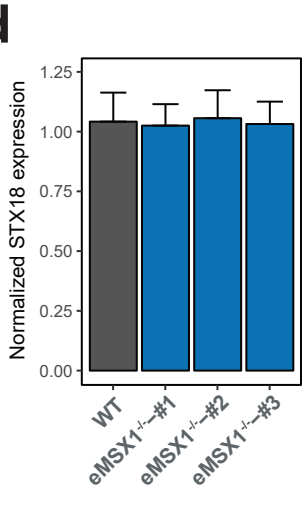
Extended Data Figure 9



Extended Data Figure 10



Extended Data Figure 11



Extended Data Figure 12