

1 **CTCF Promotes Long-range Enhancer-promoter Interactions and Lineage-specific Gene**

2 **Expression in Mammalian Cells**

3

4 Naoki Kubo¹, Haruhiko Ishii¹, Xiong Xiong², Simona Bianco³, Franz Meitinger¹, Rong Hu¹, James D.

5 Hocker¹, Mattia Conte³, David Gorkin⁴, Miao Yu¹, Bin Li¹, Jesse R. Dixon⁵, Ming Hu⁶, Mario

6 Nicodemi³, Huimin Zhao^{2,7}, and Bing Ren^{1,4,8*}

7

8 ¹Ludwig Institute for Cancer Research, La Jolla, CA, USA

9 ²Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign,

10 Urbana, IL, USA

11 ³Department of Physics, University of Naples Federico II, and INFN Complesso di Monte

12 Sant'Angelo, Naples, Italy

13 ⁴ Department of Cellular and Molecular Medicine, Center for Epigenomics, University of California

14 San Diego School of Medicine, La Jolla, CA, USA

15 ⁵Salk Institute for Biological Studies, La Jolla, CA, USA

16 ⁶Department of Quantitative Health Sciences, Lerner Research Institute, Cleveland Clinic

17 Foundation, Cleveland, OH, USA.

18 ⁷Departments of Chemistry, Biochemistry, and Bioengineering, and Carl R. Woese Institute for

19 Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA

20 ⁸Department of Cellular and Molecular Medicine, Moores Cancer Center and Institute of Genome

21 Medicine, University of California San Diego School of Medicine, La Jolla, CA, USA

22

23 *Correspondence to: biren@ucsd.edu

24

25 **[All datasets have been deposited to GEO, with accession number GSE94452, and can be**

26 **accessed [here](#)].**

27 **Abstract:**

28 **Topologically associating domains (TAD) and insulated neighborhoods (INs) have been**

29 **proposed to constrain enhancer-promoter communications to enable cell-type specific**

30 **transcription programs, but recent studies show that disruption of TADs and INs resulted in**

31 **relatively mild changes in gene expression profiles. To better understand the role of**

32 **chromatin architecture in dynamic enhancer-promoter contacts and lineage-specific gene**

33 **expression, we have utilized the auxin-inducible degron system to acutely deplete CTCF, a**

34 **key factor involved in TADs and IN formation, in mouse embryonic stem cells (mESCs) and**

35 **examined chromatin architecture and gene regulation during neural differentiation. We find**

36 **that while CTCF depletion leads to global weakening of TAD boundaries and loss of INs, only**

37 **a minor fraction of enhancer-promoter contacts are lost, affecting a small subset of genes.**

38 **The CTCF-dependent enhancer-promoter contacts tend to be long-range, spanning hundreds**

39 **of kilobases, and are established directly by CTCF binding to promoters. Disruption of CTCF**

40 **binding at the promoter reduces enhancer-promoter contacts and transcription, while**

41 **artificial tethering of CTCF to the promoter restores the enhancer-promoter contacts and**

42 **gene activation. Genome-wide analysis of CTCF binding and gene expression across multiple**

43 **mouse tissues suggests that CTCF-dependent promoter-enhancer contacts may regulate**

44 **expression of additional mouse genes, particularly those expressed in the brain. Our results**

45 **uncover both CTCF-dependent and independent enhancer-promoter contacts, and highlight a**

46 **distinct role for CTCF in promoting enhancer-promoter contacts and gene activation in**

47 **addition to its insulator function.**

48

49 **Introduction:**

50 Transcriptional regulation in mammalian cells is orchestrated by cis-regulatory elements that include
51 promoters, enhancers, insulators and other less well characterized sequences^{1,2}. Large-scale
52 projects such as ENCODE have annotated millions of candidate cis-regulatory elements in the
53 human genome and genomes of other mammalian species³⁻⁵. A majority of these candidate
54 regulatory elements are located far from transcription start sites(i.e. promoters), display tissue and
55 cell-type specific chromatin accessibility, and likely act as enhancers to regulate cell-type specific
56 gene expression. Enhancers are frequently found to be positioned close to their target gene
57 promoters in 3D space at the time of gene activation, suggesting a role for the chromatin
58 architecture in gene regulation^{6,7}. Indeed, artificially induced spatial proximity between enhancers
59 and promoters has been shown to lead to gene activation^{8,9}. Insulators, on the other hand, act to
60 block enhancer-promoter contacts to prevent ectopic gene activation¹⁰⁻¹². Clearly, in-depth
61 knowledge of the chromatin architecture in each cell type and developmental stage is necessary for
62 mechanistic understanding and functional annotation of enhancers and insulators in the genome.

63 In recent years, great strides have been made in our understanding of chromatin
64 architecture, thanks to the development of high throughput technologies to capture chromosome
65 conformation¹³⁻¹⁹. These studies have shown that interphase chromosomes reside in separate

66 nuclear space known as chromosome territories, and each chromosome is further partitioned into
67 topologically associating domains (TADs) characterized by higher levels of interactions among DNA
68 within each domain than between domains¹⁻⁵. Within TADs, genes and their regulatory elements are
69 organized into insulated neighborhoods (INs) formed by CTCF-anchored chromatin loops^{17,20}. Both
70 TADs and INs have been proposed to play important roles in gene regulation by constraining
71 enhancer-promoter contacts^{13-17,20}. Supporting this model, previous studies have shown that
72 deletion, duplication or inversion of TAD boundaries result in dysregulation of gene expression and
73 developmental disorders^{7,21-24}. Mechanistically, TADs and INs are proposed to be formed through
74 cohesin/CTCF mediated loop extrusion. In this model, the cohesin complex moves bidirectionally
75 along the chromatin fiber, and the movement is temporarily arrested by DNA-bound CTCF proteins
76 resulting extrusion of the chromatin segments between the two convergent CTCF binding sites^{25,26}.
77 Consistent with this model, acute depletion of CTCF and cohesin complex leads to global loss of
78 TADs and INs. However, the severe disruption of genome structure only affects expression of a
79 small number of genes, raising questions about the general roles of TAD and INs in gene
80 regulation^{27,28}. In addition, a recent study involving high throughput enhancer perturbation followed
81 by single cell RNA-seq analysis found that functional enhancers predominantly reside very close

82 (within 20 kb) to their target genes²⁹. These recent studies raise fundamental questions about the

83 prevalence of long-range acting enhancers and the role of TAD/INs in gene regulation.

84 Here we use two genome-wide chromatin conformation capture assays, namely in situ Hi-

85 C and PLAC-seq (also known as HiChIP)^{30,31}, to determine the chromatin architecture and

86 enhancer-promoter contacts at high-resolution during neural differentiation of mouse embryonic

87 stem cell (mESC) with continuous depletion of CTCF by auxin-inducible degron³²⁻³⁴ (Supplementary

88 Table 1). We found that most enhancer-promoter contacts, especially those at close genomic

89 distances, remained intact upon CTCF loss during cell differentiation despite the global weakening of

90 TAD boundaries. However, we also observed lost and newly formed enhancer-promoter contacts at

91 hundreds of dysregulated genes. We characterized the features of CTCF-dependent genes, and

92 found that their promoters are enriched for CTCF binding sites (CBSs) and devoid of nearby

93 enhancers. We showed that CTCF can directly establish enhancer-promoter contacts at these genes

94 since deletion of CTCF-binding site at the promoter reduces enhancer-promoter contact and gene

95 expression, while artificial tethering of CTCF to the promoter could promote enhancer-promoter

96 contacts and gene activation. Furthermore, we found over 2,300 genes that display a significant

97 correlation between CTCF occupancy at the promoter and tissue-specific gene expression patterns,

98 suggesting a role for CTCF binding in their regulation. Our findings uncovered both CTCF-

99 independent and CTCF-dependent mechanisms of enhancer-promoter communications, and
100 provided evidence for a key role for CTCF in directly promoting enhancer-promoter contacts that are
101 distinct from its function at insulator sequences.

102

103 **Results:**

104 **CTCF loss leads to weakening of TADs/INs without massive gene dysregulation during mESC
105 differentiation**

106 To investigate the functional role of CTCF in chromatin architecture and gene regulation, we utilized
107 an auxin-inducible degron system to acutely deplete CTCF protein in mESC and examined the
108 impact of CTCF loss on dynamics of gene expression and chromatin architecture during mESC
109 differentiation to neural precursor cells (NPCs) (Fig. 1a). The depletion of CTCF was verified by
110 Western blotting and chromatin occupancy of CTCF was nearly completely lost in both ESCs and
111 NPCs, along with loss of cohesin accumulation, as shown by ChIP-seq analysis (Supplementary Fig.
112 1 and Supplementary Table 2). The CTCF-depleted cells exhibited a delay in the formation of
113 neuronal axons during neural differentiation treatment with cell colonies remaining in round-shape
114 (Fig. 1b). To determine the impact of CTCF loss on chromatin architecture, we first performed Hi-C
115 analyses in mES cells before and during differentiation to NPC, in the presence and absence of

116 CTCF. TADs are characterized by strong intra-domain interactions and relatively weak inter-domain
117 interactions in Hi-C, and the strength of the TAD boundaries can be defined by the insulation score,
118 a ratio between the number of cross-border interactions and the sum of intra-domain interactions
119 within the two adjacent TADs³⁵. As shown in Supplementary Fig. 2, CTCF depletion resulted in
120 global loss of chromatin loops and contacts between convergent CTCF binding sites (genomic
121 distance > 100 kb), supporting an essential role for CTCF in the formation of these chromatin
122 organizational features (Supplementary Fig. 2a). We also observed significant weakening of TAD
123 boundaries and a dramatic loss of INs in both ESC and NPC (Supplementary Fig. 2b–e and
124 Supplementary Table 3). This result is generally consistent with previous findings indicating CTCF's
125 role in the formation of most TADs in mammalian cells²⁸ (Supplementary Fig. 2f–h).

126 We next investigated the impact of CTCF loss on gene regulation. Consistent with
127 previous reports²⁸, the vast majority of genes were expressed normally in CTCF-depleted mESC.
128 Additionally, the gene expression profiles were largely uninterrupted during cell differentiation (Fig.
129 1c). Only a small fraction of genes (382 genes, 3.0% in ESCs, 560 genes, 4.5% in NPCs) were
130 affected significantly due to CTCF loss (FDR < 0.05, fold change > 2) (Fig. 1d, e, Supplementary
131 Fig. 3, and Supplementary Table 4). Interestingly, genes that are related to neural differentiation
132 (e.g. *Pcdh* cluster genes, *Neurog*, *Neurod4*) were enriched in these CTCF-dependent genes (Fig.

133 1f), consistent with the observation that CTCF loss is accompanied by abnormal neural
134 differentiation in mESC. Thus, despite the severe disruption of TADs, INs and CTCF-mediated
135 chromatin loops, the gene regulatory programs in mESC during differentiation appear to be mostly
136 unaffected. This observation raised an important question regarding the role of chromatin domains
137 and loops in gene regulation.

138

139 **A small subset of enhancer-promoter contacts is dependent on CTCF**

140 To further delineate the relationships between the chromatin structures and gene regulation, we
141 performed quantitative analysis of enhancer-promoter contacts using PLAC-seq (also known as
142 HiChIP)^{30,31}, which interrogates chromatin contacts at select genomic regions at high resolution by
143 combining Hi-C and chromatin immunoprecipitation. We used antibodies against the histone
144 modification H3K4me3, which marks active or poised promoters, to detect chromatin contacts
145 centered on these genomic regions (Fig 2a). We obtained between 300 and 400 million paired-end
146 reads for each replicate (Supplementary Table 1). To determine the differential chromatin contacts in
147 ESCs and NPCs, we analyzed 11,900 gene promoters with similar levels of H3K4me3 ChIP-seq
148 signal using a negative binomial model for each distance-stratified 10-kb interval (Supplementary
149 Fig. 4, Methods). In total, we found 5,913 chromatin contacts between the promoters of 4,573 genes

150 and distal elements to be significantly induced during the neural differentiation (FDR < 0.05), and
151 1,594 contacts centered on 1,294 genes significantly decreased (Fig. 2b, c). Notably, over 50% of
152 these differential contacts span less than 50 kb in genomic distance (Supplementary Fig. 5a). As
153 expected, these dynamic changes of enhancer-promoter contacts were positively correlated with the
154 changes of active histone modifications such as H3K27ac and H3K4me1 (Supplementary Fig. 5b, c).

155 We confirmed previously reported dynamic enhancer-promoter contacts during mESC cell
156 differentiation (e.g. *Sox2*, *Hoxb* cluster genes, *Dnmt3b*)^{14,36} (Fig. 2b, Supplementary Fig. 5d–g).

157 Using the same approach, we determined the chromatin contacts dependent on CTCF in
158 mESCs and NPCs. The chromatin contacts between convergent CTCF-binding sites were severely
159 reduced upon CTCF loss (Supplementary Fig. 7a, b), consistent with the results from Hi-C assays.

160 However, the majority of chromatin contacts between enhancers and promoters were unchanged
161 despite the global weakening of TADs and disruption of chromatin loops and INs. Chromatin
162 contacts between 394 and 806 enhancer-promoter pairs in mESCs and NPCs, respectively,
163 decreased significantly upon CTCF loss (FDR < 0.05), while chromatin contacts between 44 and 109
164 enhancer-promoter pairs in mESCs and NPCs, respectively, increased upon CTCF loss (Fig. 2d, e,
165 Supplementary Table 5). Only 283 pairs of enhancer-promoter contacts out of 5,913 that are
166 normally induced during differentiation failed to be induced in the absence of CTCF (Fig. 2f).

167 Interestingly, genomic distances of the CTCF-dependent enhancer-promoter contacts in
168 differentiated NPCs are generally longer than that in undifferentiated ESCs (Fig. 2g). The modest
169 changes in enhancer-promoter contacts upon CTCF loss are consistent with the mild changes in
170 gene expression in these cells.

171

172 **CTCF directly promotes enhancer-promoter contacts and gene expression through binding**

173 **to gene promoters**

174 We next investigated the features of CTCF-dependent/-independent enhancer-promoter contacts
175 and gene regulation. Since ChIP-seq levels of histone modifications (H3K27ac, H3K4me1, and
176 H3K4me3) were virtually unaffected by CTCF depletion, the observed changes in enhancer-
177 promoter contacts in mESC and NPC were likely a direct consequence of CTCF loss
178 (Supplementary Fig. 4a, Supplementary Fig. 6). Consistent with this interpretation, the anchors of
179 CTCF-dependent enhancer-promoter contacts were strongly enriched for the CBSs detected by
180 ChIP-seq in the mESC and NPC (Fig. 3a). Furthermore, the degree of this enrichment increased
181 with the number of CBSs around the anchors (Fig. 3b). Many such enhancer-promoter contacts that
182 were close to CBSs at their anchor sites were identified in genes that were down-regulated upon
183 CTCF loss (Fig. 3c, d, Supplementary Fig. 7c). In these CTCF-dependent down-regulated genes, we

184 also observed many reduced enhancer-promoter contacts that had CBSs at only one side of their
185 anchor sites, preferentially at promoter side (Fig. 3c, d, Supplementary Fig. 7d, e). These findings
186 indicate that CTCF directly modulates transcription of select genes by binding to their promoters and
187 promoting long-range enhancer-promoter contacts. By contrast, the anchors of enhancer-promoter
188 contacts gained upon CTCF loss were not enriched for CBS, and they were likely a consequence of
189 loss of insulation due to the weakened TAD boundaries and INs (Supplementary Fig. 7f–h).

190 To confirm that CTCF directly mediates enhancer-promoter contacts to modulate gene
191 expression, we used CRISPR-mediated genome editing to delete a 118-bp sequence containing the
192 CTCF binding motif at the promoter of *Vcan* gene, which encodes a protein that plays an important
193 role in axonal outgrowth³⁷ and neural differentiation³⁸. *Vcan* is induced during NPC differentiation,
194 and the induction is lost upon CTCF depletion along with a long-range enhancer-promoter contact
195 (400 kb range) anchored by a CBS only on the promoter side. Polymer modelling based on the
196 strings and binders switch (SBS) model also supports such changes of chromatin contacts upon
197 CTCF depletion^{39,40} (Supplementary Fig. 8). Upon removal of the CTCF binding sequence, *Vcan*
198 expression was significantly reduced in NPC cells. This reduction in *Vcan* expression could be
199 largely restored by tethering the CTCF protein to the mutated *Vcan* promoter using a dCas9-CTCF
200 fusion and a guide RNA (gRNA) targeting a sequence adjacent to the deleted CTCF binding

201 sequence, in two different experiments using distinct gRNAs (Fig. 3e, f, Supplementary Fig. 9). The
202 rescue of the *Vcan* expression by the artificially tethered CTCF was dependent on the distal element
203 sequences (Fig. 3e, f). PLAC-seq experiments showed that the artificially tethered CTCF protein
204 could re-establish the long-range enhancer-promoter contact at *Vcan* that was lost upon deletion of
205 the CTCF binding motif near the promoter (Fig. 3g, h). Taken together, the above results
206 demonstrated that CTCF can directly promote long-range enhancer-promoter contact and gene
207 regulation by binding to the gene promoter.

208

209 **CTCF-dependent genes reside in enhancer deserts**

210 While many of the genes with CBSs at promoters were dependent on CTCF in mESCs and NPCs,
211 many others were not affected by CTCF loss despite having CTCF bindings at gene promoters
212 (Supplementary Fig. 10a). These CTCF-independent genes are generally close to the genomic
213 regions associated with the H3K27ac histone mark (≤ 50 kb, PLAC-seq peak signal p-value < 0.01)
214 (Fig. 4a, b, Supplementary Fig. 10b, c), implying that they are regulated by short-range enhancer-
215 promoter contacts formation of which are independent of CTCF (Fig. 4b, Supplementary Fig. 10d, e).
216 By contrast, CTCF-dependent genes were generally regulated by long-range enhancer-promoter
217 contacts (≥ 100 kb, Fig. 4c) especially in NPCs (Fig. 4c, Supplementary Fig. 10f). Similarly, genes

218 up-regulated upon CTCF depletion differed from those down-regulated in whether they were located
219 at enhancer desert regions or not. While the down-regulated genes tended to be located at enhancer
220 desert regions (2 enhancers or less around transcription start site (TSS) < 200 kb in Fig. 4c), the up-
221 regulated genes were close to multiple enhancers (Fig. 4c, d, and Supplementary Fig. 11 for their
222 examples).

223

224 **Promoter occupancy by CTCF correlates with expression at lineage-specific genes across**
225 **diverse mouse tissues**

226 The above findings suggest a previously under-appreciated mechanism for CTCF in gene regulation.
227 In contrast to its well-established role in forming chromatin loops, TAD boundaries and insulators, we
228 demonstrated that CTCF also directly binds to gene promoters to promote long-range enhancer-
229 promoter contacts and enable enhancer-dependent gene expression in mammalian cells. In mouse
230 ESCs and NPCs, several hundred genes are subject to regulation by this mechanism. These include
231 the proto-cadherin gene clusters that were previously reported to be regulated by CTCF binding
232 sites at the promoters and the distal enhancer⁴¹ (Fig. 2e). To further explore the extent of genes
233 subject to this CTCF-dependent mechanism, we examined public ChIP-seq datasets of CTCF
234 binding and RNA-Seq across multiple mouse tissues (9 tissue samples from ENCODE^{4,42},

235 Supplementary Table 6). Consistent with this postulated mechanism, CBSs are enriched around
236 promoters (Fig. 5a, Supplementary Fig. 12a) and ChIP-seq signals around promoters (TSS \pm 10 kb)
237 show positive correlation with gene expression in over 2,300 mouse genes in these tissues ($r > 0.6$,
238 2,332 genes), many of which could not be explained by DNA methylation levels at the promoter-
239 proximal CBSs (Fig. 5b, Supplementary Fig. 12b). Interestingly, high lineage-specificity in
240 transcription as measured by Shannon entropy⁴³ was predominantly found in the forebrain-specific
241 genes and the most enriched gene ontology (GO) term in this gene group was related to “synapse
242 assembly”. On the other hand, GO terms related to “signaling pathway” were enriched in the other
243 tissue-specific genes (Fig. 5c, Supplementary Fig. 12c–e). Many forebrain-specific and heart-specific
244 genes were down-regulated in CTCF-depleted NPCs and CTCF knockout heart tissue⁴⁴,
245 respectively (Supplementary Fig. 12f, g), supporting that many of these genes are indeed regulated
246 by CTCF binding to the gene promoters in a lineage-specific manner.

247

248 **Discussion:**

249 CTCF- and cohesin-mediated chromatin structures such as TADs and INs^{17,27,28} have been
250 postulated to play a role in constraining enhancer-promoter communications^{13–16}. However, the vast
251 majority of genes are expressed normally in the absence of CTCF or Cohesin^{27,28}, raising questions

252 about the role of chromatin architecture, especially enhancer-promoter contacts, in gene regulation.

253 Here, we provided multiple layers of evidence that CTCF not only actively forms TADs and INs, but

254 also directly promotes enhancer-promoter contacts and potentially contributes to activation of

255 thousands of lineage-specific genes. CTCF binding to the promoter of such genes is necessary and

256 sufficient for establishing their enhancer-promoter contacts. We demonstrated that artificial tethering

257 of CTCF to gene promoter could promote enhancer-promoter contacts and gene activation. Further

258 analysis of tissue-specific CTCF binding profiles and gene expression patterns across multiple

259 mouse tissues uncovered several thousand genes that might be regulated by CTCF-dependent

260 promoter-enhancer contacts.

261 Meanwhile, our study revealed that the majority of enhancer-promoter contacts are

262 independent of CTCF, which could explain the modest change of gene expression profiles upon

263 CTCF loss. Most of the enhancer-promoter contact changes during cell differentiation were

264 associated with the enhancer activities. In addition to the enhancer activity itself, it can be assumed

265 that these CTCF-independent enhancer-promoter contacts are mediated by other factors. For

266 example, Yin Yang 1 (YY1) has been shown to regulate enhancer-promoter contacts in mouse

267 embryonic stem cells^{45,46}. Another genomic interaction mediator, LIM-domain-binding protein 1

268 (LDB1), is also known to control long-range and trans interactions⁴⁷⁻⁴⁹ that regulate specific gene

269 sets such as olfactory receptor genes⁴⁹ and genes for cardiogenesis⁴⁸.

270 Our study also highlights the biological importance of the long-range chromatin contacts in

271 lineage-specific gene expression. A recent study showed that functional enhancer-promoter pairs

272 predominantly locate in very close genomic distances²⁹, despite the fact that a huge number of cis-

273 regulatory elements and interactions between promoters and distal elements have been

274 annotated^{1,2,50,51}. We show here that lineage-specific expression of many genes may be dependent

275 on long-range (>100 kb) enhancer-promoter contacts anchored by CTCF binding at the promoters. It

276 is reasonable to assume that some specific long-range enhancer-promoter contacts that potentially

277 determine lineage-specificity require the rigid structure of long-range CTCF loops^{45,46}. Consistent

278 with this model of CTCF function in promoting long-range enhancer-promoter contacts and lineage

279 specific gene expression, many previous studies that have shown that cell-type specific CTCF

280 depletion or deletion in the mouse leads to severe developmental defects⁷.

281 CTCF has been implicated in a variety of human diseases. It has been previously reported

282 that CBSs are highly mutated in several cancer types⁵²⁻⁵⁴ and somatic CTCF mutations also occur in

283 about one-quarter of endometrial carcinoma⁵⁵. Thus, further study of the mechanism for CTCF in

284 gene regulation will help to elucidate the role of CTCF and chromatin organization in tumorigenesis

285 and non-coding cancer drivers.

286

287 **Methods:**

288 **Cell culture**

289 The F1 *Mus musculus castaneus* × S129/SvJae mouse ES cells (XY; F123 cells)⁵⁶ (a gift from
290 Rudolf Jaenisch) were cultured in KnockOut Serum Replacement containing mouse ES cell media:
291 DMEM 85%, 15% KnockOut Serum Replacement (Gibco), penicillin/streptomycin (Gibco), 1× non-
292 essential amino acids (Gibco), 1× GlutaMax (Gibco), 1000 U/ml LIF (Millipore), 0.4 mM β-
293 mercaptoethanol. The cells were typically grown on 0.2% gelatin-coated plates with irradiated mouse
294 embryonic fibroblasts (MEFs) (GlobalStem). Cells were maintained by passaging using Accutase
295 (Innovative Cell Technologies) on 0.2% gelatin-coated dishes (GENTAUR) at 37°C and 5% CO₂.
296 Medium was changed daily when cells were not passaged. Cells were checked for mycoplasma
297 infection and tested negative.

298

299 **Construction of the plasmids**

300 The CRISPR/Cas9 plasmid (CTCF-mouse-3sgRNA-CRISPRexp-AID) was assembled using the
301 Multiplex CRISPR/Cas9 Assembly System kit⁵⁷ (a gift from Takashi Yamamoto, Addgene kit
302 #1000000055). Oligonucleotides for three gRNA templates were synthesized, annealed and
303 introduced into the corresponding intermediate vectors. The first gRNA matches the genome

304 sequence 23 bp upstream of the stop codon of mouse CTCF. The oligonucleotides with sequences
305 (5'-CACCGTGATCCTCAGCATGATGGAC-3') and (5'-AACGTCCATCATGCTGAGGATCAC-3')
306 were annealed. The other two gRNAs direct *in vivo* linearization of the donor vector: the first pair of
307 oligonucleotides are (5'-CACCGCTGAGGATCATCTCAGGGGC -3') and (5'-
308 AACGCCCTGAGATGATCCTCAGC -3'); the second pair is (5'-
309 CACCGATGCTGGGCCTTGCTGGC-3') and (5'-AACGCCAGCAAGGCCAGCATC-3'). The
310 three gRNA-expressing cassettes were incorporated into one single plasmid using Golden Gate
311 assembly. The donor vector (mCTCF24-AID-donor-Neo) was constructed using PCR and Gibson
312 Assembly Cloning kit (New England Biolabs). The insert cassette includes sequences that codes for
313 a 5GA linker, the auxin-induced degron (AID), a T2A peptide and the neomycin resistant marker,
314 and is flanked by 24-bp homology arms to integrate into the CTCF locus. The left and right arms
315 have sequences CCTGAGATGATCCTCAGCATGATG and GACCGGTGATGCTGGGCCTTGCT,
316 respectively. The AID coding sequence was amplified from pcDNA5-H2B-AID-EYFP³³ (a gift from
317 Don Cleveland, Addgene plasmid #47329) and the T2A-Neo^R was amplified from pAC95-pmax-
318 dCas9VP160-2A-neo⁵⁸ (a gift from Rudolf Jaenisch, Addgene plasmid #48227). The sequence for
319 the 5GA linker was included in one of the primers. The original donor backbone was a gift from Dr.
320 Ken-ichi T. Suzuki from Hiroshima University, Hiroshima, Japan.

321 The donor vector encodes the following amino acid sequence that corresponds to the 24-
322 bp left homology arm of CTCF, a 5GA linker, AID, T2A, and Neo^R:
323 PEMILSMMGAGAGAGAGAGSVELNLRETELCLGLPGGDTVAPVTGNKRGFSETVDLKLNNEPA
324 NKEGSTTHDVTTFDSKEKSACPKDPAKPPAKAQVVGWPPVRSYRKNVMVSCQKSSGGPEAAAFV
325 KVSMGDGAPYLRKIDLRLMYKSYDELSNALSNMFSSFTMGKHGGEEGMIDFMNERKLMDLVNSWDYV
326 PSYEDKDGDWMLVGDVPWPMFVDTCKRLRLMKGSDAIGLAPRAMEKCKSRAGSGEGRGSLLTCG
327 DVEENPGPRLETRMGSIAEQDGLHAGSPAAWVERLFGYDWAQQTIGCSDAAVFRLSAQGRPVLVF
328 KTDLSGALNELQDEAARLSWLATTGVPCAVALDVTEAGRDWLLLGEVPGQDLLSSH LAPAEKVSI
329 MADAMRRLHTLDPATCPFDHQAKHRIERARTRMEAGLVDQDDLDEEHQGLAPAELFARLKARMPD
330 GEDLVVTHGDACLPNIMVENGRFSGFIDCGRLGVADRYQDIALATRDIAEELGGEWADRFLVLYGIA
331 APDSQRIAFYRLLDEFF*.

332 The lentiviral vector for expressing TIR1 (Lentiv4-EFsp-Puro-2A-TIR1-9Myc) was
333 constructed using PCR and Gibson Assembly Cloning kit (New England Biolabs). The backbone
334 was modified from lentiCRISPR v2⁵⁹ (a gift from Feng Zhang, Addgene plasmid #52961) and the
335 TIR1-9myc fragment was amplified from pBabe TIR1-9myc³³ (a gift from Don Cleveland, Addgene
336 plasmid #47328). The expressing cassette includes a puromycin resistant marker followed by
337 sequences that code for P2A peptide and TIR1-9myc protein. The gene expression is driven by EFS

338 promoter in the original lentiCRISPR v2. The maps and the sequences of the plasmids are available

339 at the following URLs. CTCF-mouse-3sgRNA-CRISPRexp-AID (<https://benchling.com/s/seq->

340 1R4nJ8quYptUqerRWSDX), mCTCF24-AID-donor-Neo (<https://benchling.com/s/seq->

341 LtJu9OTscKJNCEMOk8ok), Lentiv4-EFsp-Puro-2A-TIR1-9Myc (<https://benchling.com/s/seq->

342 6wSCsW3Kr9S1igXZ8H9K).

343

344 **Transfection and establishment of CTCF-AID knock-in clones**

345 The cells were passaged once on 0.2% gelatin-coated feeder-free plates before transfection. The

346 cells were transfected using the Mouse ES Cell Nucleofector Kit (Lonza) and Amaxa Nucleofector

347 (Lonza) with 10 µg of the CRISPR plasmid and 5 µg of the donor plasmid following the

348 manufacturer's instructions. After transfection, the cells were plated on drug-resistant MEFs

349 (GlobalStem). Two days after transfection, drug selection was started by addition of 160 µg/ml G418

350 (Geneticin, Gibco) to the medium. Drug-resistant colonies were isolated and the clones with AID

351 knock-in on both alleles were found by performing PCR of the genomic DNA using primers specific

352 to sequences flanking the 3' end of the CTCF coding sequence

353 (AAATGTTAAAGTGGAGGCCTGTGAG and AAGATTGGCCGTTAACACAGC). The sequence

354 at the CTCF-AID junction on both alleles were checked by sequencing of allele-specific PCR

355 products, which were generated by using either a CTCF-129-specific
356 (CTGACTTGGGCATCACTGCTG) or a CTCF-Cast-specific
357 (GTTTGTTCTGTTGACTTAGGCATCACTGTTA) forward primer and a reverse primer in the AID
358 coding sequence (GAGGTTGGCTGGATCTTAGGACA). The expression of CTCF-AID fusion
359 protein was confirmed by observing the difference in the molecular weight compared to the control
360 cells by Western blot with anti-CTCF antibody (Millipore, 07-729).

361

362 **Lentivirus production and infection**
363 We produced the lentivirus for expressing TIR1-9myc using Lenti-X Packaging Single Shots system
364 (Clontech) and infected the CTCF-AID knock-in mESCs following the manufacturer's instructions.
365 After infection, the cells were selected by culturing with 1 µg/ml puromycin. Drug-resistant colonies
366 were isolated and expression of TIR1-9myc was confirmed by Western blot using anti-Myc antibody
367 (Santa Cruz, sc-40). Clones expressing high level of TIR1-9myc were used for the subsequent
368 experiments.

369

370 **Preparation of CTCF-depleted cells and neural progenitor cell differentiation**

371 The CTCF-AID knock-in mESCs expressing TIR1-9myc were passaged on 0.2% gelatin-coated
372 plates without MEFs. We added 1 ul 500 mM auxin (Abcam, ab146403) per 1 ml medium to deplete
373 CTCF, and changed medium with auxin every 24 hours. Cells were harvested 24, 48 or 96 hours
374 after starting auxin treatment. For NPC differentiation, the CTCF-AID knock-in mESCs were grown
375 on MEFs and passaged on 0.2% gelatin-coated plates without MEFs one day before starting
376 differentiation treatment. The cells were plated sparsely to avoid passaging to new plates during
377 neural differentiation because most of the cells failed to attach to new plates after auxin treatment.
378 On day 0, auxin was added to the CTCF-depleted cell samples, and LIF was deprived from the
379 culture medium 6 hours after adding auxin. From day 1, 5 uM retinoic acid (Sigma, R2625) was
380 added with LIF-deprived medium and auxin was also added continuously to the CTCF-depleted cell
381 samples. Cells were harvested on day 2, day 4 and day 6. To harvest auxin-washout samples, auxin
382 treatment was stopped on day 4 or day 6 and differentiation treatment was continued for another 2
383 days. Alkaline phosphatase staining was performed on each time point using the AP Staining kit II
384 (Stemgent, 00-0055).

385

386 **Antibodies**

387 Antibodies used in this study were rabbit anti-CTCF (Millipore, 07-729, for western blotting), rabbit
388 anti-Histone H3 (abcam, ab1791, for western blotting), rabbit anti-CTCF (Active Motif, 61311, for
389 microChIP-seq), rabbit anti-Rad21 (Santa Cruz, sc-98784, for microChIP-seq), rabbit anti-H3K4me1
390 (abcam, ab8895, for ChIP-seq), rabbit anti-H3K4me3 (Millipore, 04-745, for ChIP-seq), rabbit anti-
391 H3K27ac (Active Motif, 39133, for ChIP-seq), mouse anti-H3K27me3 (Active Motif, 61017, for ChIP-
392 seq), mouse anti-Myc antibody (Santa Cruz, sc-40, for western blotting), and mouse anti-Cas9 (Cell
393 Signaling, 14697, for western blotting). Goat anti-Rabbit IgG (H+L)-HRP (Bio Rad, 1706515) and
394 Goat anti-Mouse IgG (H+L)-HRP (Invitrogen, 31430) were used as secondary antibody for western
395 blotting.

396

397 **Western blotting**

398 Cells were washed with PBS and scraped in cold PBS, and pelleted to be stored at -80°C. Two
399 million cells were resuspended in 100 µL lysis buffer (20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA,
400 1mM EGTA, 1% Triton X-100, 1x complete protease inhibitor (Roche)), and sonicated for 10 minutes
401 total ON time with pulses of 15 second ON and OFF, and 40% amplitude using QSONICA 800R
402 (Qsonica). Protein concentration was measured using Pierce BCA Protein Assay Kit (Thermo
403 Fisher). Laemmli Sample Buffer (Bio-Rad) with 355 mM 2-Mercaptoethanol was mixed with 15 µg of

404 each sample and incubated for 5 minutes at 95°C. The samples were run on 4-15% Mini-
405 PROTEAN® TGX™ Precast Gels (Bio-Rad), and transferred onto nitrocellulose membranes at 100
406 V for 1 hour. The membranes were rinsed with 1x TBST and blocked with 5% dry milk at room
407 temperature for 45 minutes. After washing with TBST, the membranes were incubated with diluted
408 antibody in the blocking buffer overnight at 4°C. After overnight incubating, membranes were
409 washed 4 times 5minutes in 1x TBST at room temperature, and incubated with secondary antibody
410 in blocking buffer at room temperature for 45 minutes. After washing 4 times with TBST, the
411 substrates were detected using Pierce ECL Western Blotting Substrate (Thermo Fisher).

412

413 **Cell cycle analysis**

414 Cells were grown in 6-well plates. After dissociation with Accutase (Innovative Cell Technologies), 2-
415 5 million cells were washed with PBS and re-suspended in 300 µl ice-cold PBS. Cells were fixed for
416 a minimum of 24h at 4°C after drop-wise addition of 800 µl ice-cold ethanol. After fixation, cells were
417 pelleted and re-suspended in PBS containing 0.1% Triton X-100, 20 µg/mL Propidium iodide and 50
418 µg/ml RNase A. Cells were incubated for 30 min at 37°C before subjected to flow cytometry analysis.

419

420 **MicroChIP-seq library preparation**

421 MicroChIP-seq experiments for CTCF and Rad21 were performed as described in ENCODE
422 experiments protocols (“Ren Lab ENCODE Chromatin Immunoprecipitation Protocol for MicroChIP”
423 in <https://www.encodeproject.org/documents/>) with minor modifications. Cells were crosslinked with
424 1% formaldehyde for 10 minutes and quenched with 125mM glycine. We used 0.5 million cells for
425 microChIP. Chromatin shearing was performed using truChIP Chromatin Shearing Reagent Kit
426 (Covaris) according to the manufacturer’s instructions. Covaris M220 was used for sonication with
427 the following parameters: 10 minutes duration at 10.0% duty factor, 75.0 peak power, 200 cycles per
428 burst at 5-9°C temperature range. The chromatin was diluted with 10 mM Tris-HCl pH 7.5, 140 mM
429 NaCl, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 10% SDS, 0.1% Sodium Deoxycholate, 1x
430 complete protease inhibitor (Roche), 1 mM PMSF to adjust to 0.21% SDS concentration. We used 8
431 µL anti-rabbit IgG Dynabeads (Life Technologies) for CTCF or Rad21 antibodies and washed the
432 beads with cold RIPA buffer 1 (10 mM Tris-HCl pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.5 mM EGTA,
433 1% Triton X-100, 0.1% SDS, 0.1% Sodium Deoxycholate) for 2 times. After washing, 5 µg antibody
434 (anti-CTCF or anti-Rad21) with 95 µL RIPA buffer 1 was added to the beads and incubated on a
435 rotating platform at 4°C for 6 hours. After incubation, beads were washed once with 100 µL cold
436 RIPA buffer 1 and mixed with chromatin followed by overnight incubation on a rotating platform at
437 4°C. Beads were washed 4 times with 10 mM Tris-HCl pH 7.5, 300 mM NaCl, 1 mM EDTA, 0.5 mM

438 EGTA, 1% Triton X-100, 0.2% SDS, 0.1% Sodium Deoxycholate and washed once with 100 µL cold
439 1x TE. After removing the TE, 150 µL elution buffer 1 (20 mM Tris-HCl pH 7.5, 50 mM NaCl, 5 mM
440 EDTA, 1% SDS) was added. The input samples were processed in parallel with the ChIP samples.
441 RNase A (final conc. = 0.2 mg/mL) was added and incubated at 37°C for 1 hour with shaking at
442 1200rpm. Samples were incubated with proteinase K (final conc. = 0.13 mg/mL) at 68°C for 4 hours
443 with shaking at 1200rpm. After removal of beads, the samples were extracted with Phenol:
444 Chloroform: Isoamyl Alcohol (25:24:1) and precipitated with ethanol. To prepare Illumina sequencing
445 libraries, ThruPLEX DNA-seq 12s kit (Rubicon Genomics) was used according to the manufacturer's
446 instructions. We used 0.5-1.0 ng IP materials and 50 ng input DNA for library preparation, and 11-12
447 and 5 cycles of PCR were performed respectively. After purification by 1x AMPure Beads (Beckman
448 Coulter), library quality and quantity were estimated with TapeStation (Agilent Technologies) and
449 Qubit (Thermo Fisher Scientific) assays. Libraries were sequenced on HiSeq2500 or HiSeq4000
450 single end for 50 bp.

451
452 **ChIP-seq library preparation**
453 ChIP-seq experiments for each histone mark were performed as described in ENCODE experiment
454 protocols ("Ren Lab ENCODE Chromatin Immunoprecipitation Protocol" in

455 https://www.encodeproject.org/documents/) with minor modifications. Cells were crosslinked with 1%
456 formaldehyde for 10 minutes and quenched with 125mM glycine. We used 1.0 million cells for each
457 ChIP sample. Shearing of chromatin was performed using truChIP Chromatin Shearing Reagent Kit
458 (Covaris) according to the manufacturer's instructions. Covaris M220 was used for sonication with
459 following parameters: 10 minutes duration at 10.0% duty factor, 75.0 peak power, 200 cycles per
460 burst at 5-9°C temperature range. The concentration of fragmented DNA was diluted to 0.2 µg/µl
461 with 1x TE. For immunoprecipitation, we used 11 µL anti-rabbit or anti-mouse IgG Dynabeads (Life
462 Technologies) and wash them with cold BSA/PBS (0.5 mg / mL bovine serum albumin in 1x
463 phosphate buffered saline) for 3 times. After washing, 3 µg antibody with 147 µL cold BSA/PBS were
464 added to the beads and incubated on a rotating platform at 4°C for 2 hours. After incubation, beads
465 were washed with 150 µL cold BSA/PBS for 3 times, and mixed with 100 µL Binding Buffer (1% Triton
466 X-100, 0.1% Sodium Deoxycholate, 1x complete protease inhibitor (Roche)) plus 100 µL 0.2 µg/µl
467 chromatin followed by overnight incubation on a rotating platform at 4°C. Beads were washed 5
468 times with 50 mM Hepes pH 8.0, 1% NP-40, 1 mM EDTA, 0.70% Sodium Deoxycholate, 0.5 M LiCl,
469 1x complete protease inhibitor (Roche) and washed once with 150 µL cold 1x TE followed by
470 incubation at 65°C for 20 minutes in 150 µL ChIP elution buffer (10 mM Tris-HCl pH 8.0, 1 mM
471 EDTA, 1% SDS). The beads were removed and the samples were further incubated at 65°C

472 overnight to reverse crosslinks. The input samples were processed in parallel with the ChIP
473 samples. Samples were incubated with RNase A (final conc. = 0.2 mg/mL) at 37°C for 1 hour, and
474 Proteinase K (final conc. = 0.4 mg/mL) was added and incubated at 55°C for 1 hour. The samples
475 were extracted with phenol: chloroform: isoamyl alcohol (25:24:1) and precipitated with ethanol. We
476 used 3-5 ng of starting IP materials for preparing Illumina sequencing libraries. The End-It DNA End-
477 Repair Kit (Epicentre) was used to repair DNA fragments to blunt ends, and the samples were
478 purified by Qiagen MinElute PCR Purification kit (Qiagen, Cat#28006). A-tailing 3' end was
479 performed using Klenow Fragment (3'→5' exo-) (New England Biolabs), and then TruSeq Adapters
480 were ligated by Quick T4 DNA Ligase (New England Biolabs). Size selection using AMPure Beads
481 (Beckman Coulter) was performed to get 300-500bp DNA and PCR amplification (8-10 cycles) were
482 performed. After purification by 1x AMPure Beads (Beckman Coulter), library quality and quantity
483 were estimated with TapeStation (Agilent Technologies) and Qubit (Thermo Fisher Scientific)
484 assays. Libraries were sequenced on HiSeq4000 single end for 50 bp.

485

486 **RNA-seq library preparation**

487 Total RNA was extracted from 1–2 million cells using the AllPrep Mini kit (QIAGEN) according to the
488 manufacturer's instructions and 1 µg of total RNA was used to prepare each RNA-seq library. The

489 libraries were prepared using TruSeq Stranded mRNA Library Prep Kit (Illumina). Library quality and
490 quantity were estimated with TapeStation (Agilent Technologies) and Qubit (Thermo Fisher
491 Scientific) assays. Libraries were sequenced on HiSeq4000 using 50 bp paired-end.

492

493 **Hi-C library preparation**

494 *In situ* Hi-C experiments were performed as previously described using the MboI restriction
495 enzyme¹⁶. Cells were crosslinked with 1% formaldehyde for 10 minutes. Then 25 µL per mL of 2.5M
496 Glycine was added followed by a 5-minute incubation at room temperature and then a 15-minute
497 incubation on ice. The crosslinked pellets were washed with 1 x PBS and incubated with 200µL of
498 lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA630, 33 µL Protease Inhibitor
499 (Sigma, P8340)) on ice for 15 min, washed with 300 µL cold lysis buffer, and then incubated in 50µL
500 of 0.5% SDS for 10min at 62°C. After heating, 170 µL of 1.47% Triton X-100 was added and
501 incubated for 15min at 37°C. To digest chromatin 100U MboI and 25µL of 10X NEBuffer2 were
502 added followed by overnight incubation at 37°C with agitation at 700rpm on a thermomixer. After
503 incubation, MboI was inactivated by heating at 62°C for 20 minutes. The digested ends were filled
504 and labeled with biotin by adding 37.5µL of 0.4mM biotin-14-dATP (Life Tech), 1.5 µL of 10mM
505 dCTP, 10mM dTTP, 10mM dGTP, and 8µL of 5U/µL Klenow (New England Biolabs) and incubating

506 at 23°C for 60 minutes with shaking at 500 rpm on a thermomixer. Then the samples were mixed
507 with 1x T4 DNA ligase buffer (New England Biolabs), 0.83% Triton X-100, 0.1 mg/mL BSA, 2000U
508 T4 DNA Ligase (New England Biolabs, M0202), and incubated for at 23°C for 4 hours with shaking
509 at 300rpm on a thermomixer to ligate the ends. After the ligation reaction, samples were spun and
510 pellets were resuspended in 550uL 10 mM Tris-HCl, pH 8.0. To digest the proteins and to reverse
511 the crosslinks, 50 µL of 20mg/mL Proteinase K (New England Biolabs) and 57 µL of 10% SDS were
512 mixed with the samples, and incubated at 55°C for 30 minutes, and then 67 µL of 5M NaCl were
513 added followed by overnight incubation at 68°C. After cooling the samples, 0.8X Ampure (Beckman-
514 Coulter) purification was performed. Next, the samples were sonicated to mean fragment length of
515 400 bp using Covaris M220 with the following parameters: 70 seconds duration at 10.0% duty factor,
516 50.0 peak power, 200 cycles per burst. To collect 200-600 bp size of fragmented DNA, two rounds of
517 Ampure (Beckman-Coulter) beads purification was performed. The DNA labeled with biotin was
518 purified using 100 µL of 10 mg/mL Dynabeads My One T1 Streptavidin beads (Invitrogen). The
519 washed beads were transferred to the sample tube, incubated for 15 minutes at room temperature,
520 and the supernatant was removed. Then the beads were washed twice by 600 µL of 1x Tween
521 Wash Buffer with mixing for 2 minutes at 55°C. Then the beads were equilibrated once in 100 uL 1x
522 NEB T4 DNA ligase buffer (New England Biolabs) followed by removal of the supernatant. To repair

523 the fragmented ends and remove biotin from unligated ends, the beads were resuspended in 100 μ L
524 of the following: 88 μ L 1X NEB T4 DNA ligase buffer (New England Biolabs, B0202), 2 μ L of 25mM
525 dNTP mix, 5 μ L of 10 U/ μ L T4 PNK (New England Biolabs), 4 μ L of 3 U/ μ L NEB T4 DNA
526 Polymerase (New England Biolabs), 1 μ L of 5U/ μ L Klenow (New England Biolabs). The beads were
527 incubated for 30 minutes at room temperature. The beads were washed twice by adding 600 μ L of
528 1x Tween Wash Buffer, heating on a thermomixer for 2 minutes at 55°C with mixing. To add dA-tail,
529 the beads were resuspended in 90 μ L of 1X NEB Buffer2, 5 μ L of 10mM dATP, and 5 μ L of 5U/ μ L
530 Klenow (exo-) (New England Biolabs). The beads were incubated for 30 minutes at 37°C. The beads
531 were washed twice by adding 600 μ L of 1x Tween Wash Buffer, heating on a thermomixer for 2
532 minutes at 55°C with mixing. Following the washes, the beads were equilibrated once in 100 μ L 1x
533 NEB Quick Ligation Reaction Buffer (New England Biolabs). Then the beads were resuspended
534 again in 50 μ L 1x NEB Quick Ligation Reaction Buffer. To ligate adapters, 2 μ L of NEB DNA Quick
535 Ligase (New England Biolabs) and 3 μ L of Illumina Indexed adapter were added to the beads and
536 incubated for 15 minutes at room temperature. The beads were washed twice with 600 μ L of 1x
537 Tween Wash Buffer, heating on a thermomixer for 2 minutes at 55°C with mixing. Then the beads
538 were resuspended once in 100 μ L 10 mM Tris-HCl, pH 8.0, followed by removal of the supernatant
539 and resuspension again in 50 μ L 10 mM Tris-HCl, pH 8.0. PCR amplification (8-9 cycles) was

540 performed with 10 µL Fusion HF Buffer (New England Biolabs), 3.125 µL 10uM TruSeq Primer 1,
541 3.125 µL 10uM TruSeq Primer 2, 1 µL 10mM dNTPs, 0.5 µL Fusion HotStartII, 20.75 µL ddH₂O, 11.5
542 µL Bead-bound HiC library. Then PCR products underwent final purification using AMPure beads
543 (Beckman-Coulter). Libraries were sequenced on Illumina HiSeq 4000.

544

545 **PLAC-seq library preparation**

546 PLAC-seq experiments were performed as previously described³⁰. Cells were crosslinked with 1%
547 formaldehyde (w/v, methanol-free, ThermoFisher) for 15 minutes and quenched with 125mM
548 glycine. The crosslinked pellets (2.5–3 million cells per sample) were incubated with 300ul of lysis
549 buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% Igepal CA630, 33 µL, 1x complete protease
550 inhibitor (Roche)) on ice for 15 min, washed with 500 µL cold lysis buffer, and then incubated in 50uL
551 of 0.5% SDS for 10min at 62°C. After heating, 160 µL of 1.56% Triton X-100 was added and
552 incubated for 15min at 37°C. To digest chromatin 100U Mbol and 25uL of 10X NEBuffer2 were
553 added followed by 2 hours incubation at 37°C with agitation at 900rpm on a thermomixer. After
554 incubation, Mbol was inactivated by heating at 62°C for 20 minutes. Digestion efficiency was
555 confirmed by performing agarose gel electrophoresis of the samples. The digested ends were filled
556 and labeled with biotin by adding 37.5uL of 0.4mM biotin-14-dATP (Life Tech), 1.5 µL of 10mM

557 dCTP, 10mM dTTP, 10mM dGTP, and 8uL of 5U/uL Klenow (New England Biolabs) and incubating
558 at 37°C for 60 minutes with shaking at 900 rpm on a thermomixer. Then the samples were mixed
559 with 1x T4 DNA ligase buffer (New England Biolabs), 0.83% Triton X-100, 0.1 mg/mL BSA, 2000U
560 T4 DNA Ligase (New England Biolabs, M0202), and incubated for at room temperature for 2 hours
561 with shaking with slow rotation. The ligated cell pellets were resuspended in 125 ul of RIPA buffer
562 with protease inhibitor and incubated on ice for 10 minutes. The cell lysates were sonicated using
563 Covaris M220, and the sheared chromatins were spun at 14,000 rmp, 4°C for 15 minutes to clear the
564 cell lysate. We saved 20 ul supernatant as input, and for the rest part, 100 ul of antibody-coupled
565 beads were added to the supernatant sample, and then rotated in cold room at least 12 hours. For
566 immunoprecipitation, 300 ul of M280 sheep anti-rabbit IgG beads (ThermoFisher) was washed with
567 cold BSA/PBS (0.5 mg / mL bovine serum albumin in 1x phosphate buffered saline) for 4 times. After
568 washing, 30 ug anti-H3K4me3 (Millipore, 04-745) with 1 mL cold BSA/PBS were added to the beads
569 and incubated on a rotating platform at 4°C for at least 3 hours. After incubation, beads were
570 washed with cold BSA/PBS, and resuspended in 600 ul RIPA buffer. The beads were washed with
571 RIPA buffer (3 times), RIPA buffer + 0.16M NaCl (2 times), LiCl buffer (1 time), and TE buffer (2
572 times) at 4°C for 3 minutes at 1000 rpm. For reverse crosslinking, 163 ul extraction buffer (135 ul
573 1xTE, 15 ul 10% SDS, 12 ul 5M NaCl, 1 ul RnaseA (10mg/ml)) was added and incubated at 37°C for

574 1 hour at 1000 rpm, and 20 ug of proteinase K was added and incubated at 65°C for 2 hours at

575 1000rpm. After crosslinking, DNA was purified using Zymo DNA clean & concentrator and eluted

576 with 50 ul of 10mM Tris (pH 8.0). For biotin enrichment, 25 ul of T1 Streptavidin Beads (Invitrogen)

577 per sample were washed with 400 ul Tween wash buffer (5 mM Tris-HCl pH 8.0, 0.5 mM EDTA, 1 M

578 NaCl, 0.05% Tween-20), and resuspended in 50 ul of 2x Binding buffer (10 mM Tris-HCl pH 7.5, 1

579 mM EDTA, 2 M NaCl). The purified 50 ul DNA sample was added to the 50 ul resuspended beads

580 and incubated at room temperature for 15 minutes with rotation. The beads were washed with 500 ul

581 of Tween wash buffer twice and washed with 100 ul Low EDTA TE (supplied by Swift Biosciences

582 kit). Then beads were resuspended in 40 ul Low EDTA TE. Next, we used Swift Biosciences kit (Cat.

583 No. 21024) for library construction with modified protocol as described below. The Repair I Reaction

584 Mix was added to 40 ul sample beads and incubated at 37°C for 10 minutes at 800 rpm. The beads

585 were washed with 500 ul Tween wash buffer twice and washed with 100 ul Low EDTA TE once. The

586 Repair II Reaction Mix was added to the beads followed by incubation at 20°C for 20 minutes at 800

587 rpm. The beads were washed with 500 ul Tween wash buffer twice and washed with 100 ul Low

588 EDTA TE once. Then 25 ul of the Ligation I Reaction Mix and Reagent Y2 was added to the beads

589 followed by incubation at 25°C for 15 minutes at 800 rpm. The beads were washed with 500 ul

590 Tween wash buffer twice and washed with 100 ul Low EDTA TE once. Then 50 ul of the Ligation II

591 Reaction Mix was added to the beads followed by incubation at 40°C for 10 minutes at 800 rpm. The
592 beads were washed with 500 ul Tween wash buffer twice and washed with 100 ul Low EDTA TE
593 once, and resuspended in 21 ul 10mM Tris-HCl (pH 8.0). The amplification and purification were
594 performed according to the Swift library kit protocols. Libraries were sequenced on Illumina HiSeq
595 4000.

596
597 **ChIP-seq data analysis**

598 Each fastq file was mapped to mouse genome (mm10) with BWA⁶⁰ -aln with “-q 5 -l 32 -k 2” options.
599 PCR duplicates were removed using Picard MarkDuplicates
600 (<https://github.com/broadinstitute/picard>) and the bigWig files were created using deepTools⁶¹ with
601 following parameters: bamCompare --binSize 10 --normalizeUsing RPKM --ratio subtract (or ratio).
602 The deepTools was also used for generating heatmaps. Peaks were called with input control using
603 MACS2⁶² with regular peak calling for narrow peaks (e.g. CTCF) and broad peak calling for broad
604 peaks (e.g. H3K27me3, K3K4me1). Enhancer regions were characterized by the presence of both
605 H3K4me1 peak and H3K27ac peak. We used DEseq2⁶³ to calculate differences in peak levels
606 between samples.

607

608 **RNA-seq data analysis**

609 RNA-seq reads (paired-end, 100 bases) were aligned against the mouse mm10 genome assembly
610 using STAR⁶⁴. The mapped reads were counted using HTSeq⁶⁵ and the output files from two
611 replicates were subsequently analyzed by edgeR⁶⁶ to estimate the transcript abundance and to
612 detect the differentially expressed genes. Only genes that had H3K4me3 ChIP-seq peaks on TSS
613 were used for downstream analysis (Fig. 1d, e). Differentially expressed genes were called by FDR
614 < 0.01 and fold change > 2 thresholds. RPKM was calculated using an in-house pipeline.

615

616 **Hi-C data analysis**

617 Hi-C reads (paired-end, 50 or 100 bases) were aligned against the mm10 genome using BWA⁶⁰ -
618 mem. Reads mapped to the same fragment were removed and PCR duplicate reads were removed
619 using Picard MarkDuplicates. Raw contact matrices were constructed using in-house scripts with 10
620 or 40 kb resolution, and then normalized using HiCNorm⁶⁷. We used juicebox pre⁶⁸ to create hic file
621 with -q 30 -f options. To visualize Hi-C data, we used Juicebox⁶⁸ and 3D Genome
622 Browser (<http://www.3dgenome.org>). Topological domain boundaries were identified at 40-kb or 10-
623 kb resolution based on the directionality index (DI) score and a Hidden Markov Model as previously
624 described¹³, and they were also identified based on insulation scores using peakdet (Billauer E,

625 2012. <http://billauer.co.il/peakdet.html>). The insulation score analysis was performed as previously
626 described⁶⁹ and insulation scores on TAD boundaries were calculated by taking the average value of
627 scores that overlapped with TAD boundaries. The stripe calling was performed using a homemade
628 pipeline (shared from Feng Yue lab, Penn State University) that is based on the algorithm proposed
629 in a previous study⁷⁰. We used HiCCUPS⁶⁸ with options “-r 10000 -k KR -f 0.001 -p 2 -i 5 -d 50000”
630 to identify Hi-C peaks as chromatin loops, and then we chose CTCF associated loops among them
631 that were overlapped with convergently oriented CTCF ChIP-seq peaks in control cells. The
632 aggregate analysis of CTCF associated loops were performed using APA⁶⁸ with default parameters.

633

634 To assess global changes in TAD boundary strength between samples, we performed a comparison
635 of each samples' aggregated boundary contact profile (Supplementary Fig. 2c). First, to generate a
636 consensus set of TAD boundaries we performed a simple merge between boundaries from clone 1
637 before auxin treatment (Clone 1, 0 hr) and boundaries from clone 2 before auxin treatment (Clone 2,
638 0 hr). Two filtering steps were used to generate the final set of consensus boundaries: 1) We
639 discarded boundaries there were within 3.04 Mb of a chromosome start or end, because we would
640 not be able to extract a submatrix of the correct size for the aggregate analysis; 2) We discarded
641 boundaries > 200kb, because these often represent regions of disorganized chromatin between

642 TADs, rather than true TAD boundaries. Next, we extracted a Hi-C sub-matrix for each boundary in
643 each sample. Each sub-matrix consists of a window of 3.04 Mb centered on the midpoint of the
644 boundary in question. These boundary sub-matrices were then averaged to generate one 3.04 Mb
645 matrix representing the average boundary contact profile in a given sample. To facilitate comparison
646 between samples, average boundary contact profiles were then normalized across samples using
647 standard quantile normalization. We then made pairwise comparisons between samples by
648 subtracting the average boundary contact profile of sample 1 from the average boundary contact
649 profile of sample 2. The list of consensus TAD boundaries used here is the same as that described
650 for the aggregate boundary analysis above.

651

652 **PLAC-seq data analysis**

653 PLAC-seq reads (paired-end, 50 bases) were aligned against the mm10 genome using BWA⁶⁰ -
654 mem. Reads mapped to the same fragment were removed and PCR duplicate reads were removed
655 using Picard MarkDuplicates. Filtered reads were binned at 10 kb size to generate the contact
656 matrix. Individual bins that were overlapped with H3K4me3 peaks on TSSs were used for
657 downstream differential contact analysis. For the peak calling, we used MAPS⁷¹ with default settings
658 and FitHiChIP⁷² with coverage bias correction with default settings in 10 kb resolution.

659 For differential contact analysis, the raw contact counts in 10 kb resolution bins that have
660 the same genomic distance were used as inputs for comparison. To minimize the bias from genomic
661 distance, we stratified the inputs into every 10-kb genomic distance from 10 kb to 150 kb, and the
662 other input bins with longer distances were stratified to have uniform size of input bins that were
663 equal to that of 140–150 kb distance bins. Since each input showed negative binomial distribution,
664 we used edgeR⁶⁶ to get the initial set of differential interactions. We only used bins that have more
665 than 20 contact counts in each sample of two replicates for downstream analysis. The significances
666 of these differential interactions are either due to the difference in their H3K4me3 ChIP coverage or
667 3D contacts coverage. Therefore, the chromatin contacts overlapping with differential ChIP-seq
668 peaks ($FDR < 0.05$, $\log FC < 0.5$) were removed and only the chromatin contacts with the same level
669 of H3K4me3 ChIP-seq peaks were processed. In this differential analysis, we used all bins for inputs
670 that included non-significant interactions that were not identified by MAPS or FitHiChIP peak caller,
671 because the majority of short-range interactions were not identified as significant peaks due to their
672 high background and the changes in the short-range interactions might be also critical for gene
673 regulation. We identified a large number of differentially changed short-range interactions even
674 though many of them were not identified as significant peaks, and we observed a clear correlation
675 between these differentially changed interactions and the changes of active enhancer levels on their

676 anchor regions during neural differentiation, suggesting these interaction changes might reflect the
677 biological changes. We used significance level with change direction (-/+ log₁₀(p-value)) instead of
678 fold change to show the changes of interactions, because fold change tends to be small value
679 especially in short-range interactions even though the change is actually significant for biological
680 aspects. To visualize the PLAC-seq contacts, we used WashU Epigenome Browser⁷³.

681

682 **Active/inactive contact (AIC) ratio/value**

683 The change of chromatin contacts on enhancer-promoter (E-P) is affected by the alteration of
684 enhancer activities such as H3K27ac and H3K4me1 levels (Supplementary Fig. 5b, c), and it is also
685 well known that gene expression levels have positive correlation with these active marks around
686 their TSS. These findings suggest that information of contact counts itself should involve the
687 information of enhancer activity. Furthermore, the majority of genes have multiple E-P contacts with
688 variable changes of contact frequencies. Therefore, we designed AIC value to represent quantitative
689 activity of multiple E-P contacts and aimed to show the relationship between gene regulation change
690 and E-P contacts change without using any quantitative values of histone marks. First, we summed
691 total contact counts on active elements and promoters in each gene. As for promoter-promoter (P-P)
692 contacts, they have similar function as E-P contacts^{74,75}. However, it is still unclear that the same

693 contact frequency of P-P contacts has the same effect as that of E-P contacts. Moreover, in
694 H3K4me3 PLAC-seq datasets, P-P contacts correspond to peak-to-peak interactions that have
695 generally higher contact counts than that of peak-to-non-peak interactions. Therefore, we divided the
696 P-P contact counts by a certain integer that showed the highest correlation coefficient between gene
697 expression changes and AIC value changes before summing total active contact counts. We tested
698 integers from 0 to 10 to divide the P-P contact counts, and dividing by 3 and 8 showed the highest
699 correlation coefficient between gene expression changes and AIC value changes in ESCs and
700 NPCs, respectively. The simply summing of active contact counts is still not proper for comparison
701 between different samples because they are affected by the difference of H3K4me3 peak levels on
702 TSS in different samples. Therefore, in order to cancel the bias from the H3K4me3 peak levels in
703 different samples, we also calculated total contact counts on inactive (non-active) regions and
704 computed active/inactive contact (AIC) ratio on each gene by following formula.

705 *AIC ratio on gene A = Sum of Active Contact counts (SAC) on gene A / Sum of Inactive Contact*
706 *counts (SIC) on gene A*

707 Next, we calculated the average of SICs from the comparing two samples on each gene,
708 and multiplied them by AIC ratios to calculate AIC values. AIC values are computed as pseudo
709 contact counts to perform differential analysis by edgeR⁶⁶ after rounding them to their nearest

710 integer. The bias from different H3K4me3 levels on TSS in different samples does not occur by

711 multiplying the common average value of the SICs.

712 *AIC value on gene A = AIC ratio on gene A × Average of SICs of compared two samples on gene A*

713 We also computed the changes of AIC values using Hi-C datasets in the same way, and

714 we could observe comparable correlations with gene expression changes.

715

716 **Odds ratio calculation for CTCF-dependent E-P contacts enrichment**

717 For Fig. 3a, b all PLAC-seq contacts (10 kb resolution) on promoter and enhancer were classified

718 based on the distance from anchor sites (enhancer side or promoter side) to the nearest CTCF

719 binding sites (Fig. 3a, categorized into 4x4 bins). They are also classified based on the number of

720 CTCF motif sites around each anchor site (10 kb bin \pm 5 kb) (Fig. 3b, categorized into 5x5 bins).

721 Then, we generated 2x2 tables based on whether they are CTCF-dependent contacts or not (FDR <

722 0.05) and whether they were categorized into the bin or not. Odds ratios and p-values on each 2x2

723 tables were calculated.

724 For Fig. 3c, d all PLAC-seq contacts (10 kb resolution) on promoter and enhancer were

725 classified based on the distance from anchor sites (enhancer side or promoter side) to the nearest

726 CTCF binding sites (Fig. 3c, categorized into 4x4 bins). They are also classified based on the

727 number of CTCF motif sites around each anchor site (10 kb bin \pm 5 kb) (Fig. 3d, categorized into 5x5
728 bins). Then, we generated 2x2 tables based on whether they are on differentially down-regulated
729 genes or not (FDR < 0.05) and whether they were categorized into the bin or not. For the E-P
730 contacts on differentially down-regulated genes, chromatin contacts that were identified as
731 significant by peak calling were counted (p value < 0.01). Odds ratios and p-values on each 2x2
732 tables were calculated.

733 For Fig. 4b, all genes were classified based on the distance to the nearest interacting
734 enhancer and the number of enhancers around TSS (< 200 kb) (categorized into 3x3 bins). The
735 distance to the nearest interacting enhancer is represented by the shortest genomic distance of
736 significant PLAC-seq peaks on enhancers and promoters (p-value < 0.01). Then, we generated 2x2
737 tables based on whether they are CTCF-independent stably-regulated genes or not (FDR < 0.05)
738 and whether they were categorized into the bin or not. Odds ratios and p-values on each 2x2 tables
739 were calculated. In Fig. 4c and Supplementary Fig. 10f, the same analysis as Fig. 4b was performed
740 in CTCF-dependent down-regulated genes and CTCF-dependent up-regulated genes.

741

742 **3D modelling**

743 We used the Strings & Binders Switch (SBS) polymer model^{39,76} to dissect the 3D spatial
744 organization of the *Vcan* gene region in wild type and CTCF depleted NPC cells. In the SBS view, a
745 chromatin filament is modelled as a Self-Avoiding Walk (SAW) chain of beads, comprising different
746 specific binding sites for diffusing cognate molecular factors, called binders. Different types of
747 binding sites are visually represented by distinct colors. Beads and binders of the same color interact
748 with an attractive potential, so driving the folding of the chain. All binders also interact unspecifically
749 with all the beads of the polymer by a weaker energy affinity (see below). We estimated the optimal
750 number of distinct binding site types describing the locus and their arrangement along the polymer
751 chain by using the PRISMR algorithm, a previously described machine learning based procedure⁴⁰.
752 In brief, PRISMR takes as input a pairwise experimental contact map (e.g. Hi-C) of the studied
753 genomic region and, via a standard Simulated Annealing Monte Carlo optimization, returns the
754 minimal number of different binding site types and their arrangement along the SBS polymer chain,
755 which best reproduce the input contact map. Next, we ran Molecular Dynamics (MD) simulations of
756 the inferred SBS polymers so to produce a thermodynamics ensemble of single molecule 3D
757 conformations.
758 We focused on the genomic region chr13:89,200,000-92,000,000 (mm10) encompassing
759 the mouse *Vcan* gene, in wild type and CTCF depleted NPC cells. Applied to our Hi-C contacts data

760 of the region, at 10kb resolution, the PRISMR algorithm⁴⁰ returned in both cases polymer models
761 made of 6 different types of binding sites. In our simulations, beads and binders interact via standard
762 potentials of classical polymer physics studies⁷⁷ and the system Brownian dynamics is defined by
763 the Langevin equation. By using the LAMMPS software⁷⁸, we ran massive parallel MD simulations
764 so producing an ensemble of, at least, 10^2 independent conformations. We started our MD
765 simulations from initial SAW configurations and let the polymers evolve up to 10^8 MD time steps
766 when the equilibrium globule phase is reached. We explored a range of specific and non-specific
767 binding energies in the weak biochemical energy range, respectively from $3.0K_B T$ to $8.0K_B T$ and
768 from $0K_B T$ to $2.7K_B T$, where K_B is the Boltzmann constant and T is the system temperature. For the
769 sake of simplicity, those affinity strengths are the same for all the different types. All details about the
770 model and MD simulations are described in^{39,40}. To better highlight the locations of the *Vcan* gene
771 and its regulatory elements in the two different cases, we produced a coarse-grained version of the
772 polymers. We interpolated the coordinates of the beads with a smooth third-order polynomial curve
773 and used the POV-ray software (Persistence of Vision Raytracer Pty. Ltd) to produce the 3D images.
774 For the model derived contact maps, we computed the average contact frequencies from
775 our MD derived ensemble of 3D polymer model conformations for each cell type. We followed a
776 standard approach that considers a pair of polymer sites in contact if their physical distance is lower

777 than a threshold distance⁴⁰. To compare model contact maps with corresponding Hi-C data in each
778 cell type, we used the HiCRep stratum adjusted correlation coefficient (SCC)⁷⁹, a bias-corrected
779 correlation designed for Hi-C comparison, with a smoothing parameter h=5 and an upper bound of
780 interaction distance equal to 1.5Mb. To compute the model frequencies of multiple contacts, we
781 proceeded similarly. Specifically, fixed a viewpoint site k, we accounted for a triple contact (i,j,k)
782 between k and any pair of sites i,j along the locus if their relative physical distances were all lower
783 than the threshold distance.

784

785 **CTCF motif deletion and tethering dCas9-CTCF**

786 The CRISPR/Cas9 system was used to delete CTCF motif nearby *Vcan* promoter. The sequences of
787 the DNA targeted are listed below (the protospacer adjacent motif is underlined). The guide RNAs
788 were generated using GeneArt Precision gRNA Synthesis Kit (Invitrogen).

789 5'-TTCAGCACAAGCGGAAAATAGGG-3',

790 5'-CTGCTTGCAGTTGGGTTTTCGG-3'

791 Transfection of gRNA and Cas9 ribonucleoprotein (EnGen SpyCas9, New England Biolabs) into
792 mESCs was performed using Neon Transfection System, 10 ul tip kit (Life Technologies). The cells
793 were grown for approximately one week, and individual colonies were picked into a 96-well plate.

794 After expanding cells, genotyping by PCR and Sanger sequencing were performed to confirm the
795 motif deletion.

796 For the generation of dCas9-CTCF tethered cell lines, plasmids containing sequences for

797 dCas9 and CTCF were generated by modifying lenti-dCas-VP64-Blast (a gift from Feng Zhang,

798 Addgene #61425). The VP64 cassette was replaced by CTCF sequences to generate dCas9-CTCF

799 and neomycin resistant marker that was taken from pAC95-pmax-dCas9VP160-2A-neo (a gift from

800 Rudolf Jaenisch, Addgene 48227) was inserted. To generate gRNA plasmids to recruit dCas9, the

801 gRNA oligos were inserted into the backbone vector (pSPgRNA, a gift from Charles Gersbach,

802 Addgene Plasmid #47108). The gRNA was designed to target the top and bottom strand of *Vcan*

803 promoter-proximal region which is close to the deleted CTCF motif locus. The sequences of the DNA

804 targeted are listed below (the protospacer adjacent motif is underlined).

805 5'-CCTGCCTCCTGGACAGAGACGG-3' (for top strand)

806 5'-GTCCCTTCCGTCTGTCCAAGG-3' (for bottom strand)

807 The plasmids for dCas9-CTCF and gRNA were extracted using PureLink HiPure Plasmid Midiprep

808 Kit (Invitrogen). For the electroporation, 350 ng of dCas9-CTCF plasmids and 600 ng of gRNA

809 plasmids (1 ul) were added to 0.1–0.2 million mESCs resuspended in 10 ul Buffer R (Invitrogen),

810 and electric pulse was delivered with the setting of 1200 V, 20 ms, and 2 pulses. After culturing

811 approximately 10 days, individual colonies were picked and genotyping and western blotting were

812 performed to confirm the sequences from the transfection plasmids and their protein expression.

813 For deletion of enhancer region that is interacting with *Vcan* promoter. The sequences of

814 the DNA targeted are listed below (the protospacer adjacent motif is underlined).

815 5'- AGGAACGGCCCATTCCCGAGGGG-3',

816 5'- CAATCAATAATAAACACGCCATAGG -3'

817 Generating gRNA and transfection of gRNA and Cas9 ribonucleoprotein into mESCs were

818 performed in the same way as the deletion of CTCF motif was done. Genotyping by PCR and

819 Sanger sequencing were performed to confirm the deletion.

820

821 **Analysis of CTCF-occupied promoter (COP) genes in multiple mouse tissues**

822 To analyze the CTCF ChIP-seq signals around promoters, we calculated fold changes of sample

823 RPKM over input RPKM in each 50-bp bin and summed them in each promoter region (TSS ±10 kb)

824 when the 50-bp bins were located at the regions of optimal IDR thresholded ChIP-seq peaks. Then

825 correlation coefficient between these summed CTCF ChIP-seq signals and RNA-seq RPKM values

826 across 9 mouse tissues was computed in each gene. Heatmap was generated for genes with high

827 correlation coefficient (> 0.6). The values in the heatmap were calculated by $\log_2(\text{summed ChIP-seq}$

828 signals / average value of all tissues) for promoter-proximal CTCF signal and $\log_2(\text{RPKM} / \text{average}$

829 RPKM of all tissues) for gene expression. Lineage specificity of transcription was measured by

830 Shannon entropy⁴³. For DNA methylation levels around promoters, DNA methylation rates at CBSs

831 (motif sequences ±100 bp) were calculated and averaged in each promoter region (TSS ±10 kb).

832 PhastCons score⁸⁰ was used for conservation analysis. The highest phastCons score at each CTCF

833 motif locus was represented as the conservation score of each CBS.

834

835 **Data accessibility:**

836 All datasets generated in this study have been deposited to Gene Expression Omnibus (GEO), with

837 accession number GSE94452. Hi-C dataset analyzed in Supplementary Fig. 2f–h was provided from

838 Dr. Benoit Bruneau²⁸. All mouse tissue datasets in Fig. 5 and Supplementary Fig. 12 were

839 downloaded from the ENCODE portal⁸¹ (<https://www.encodeproject.org/>); accessions for datasets

840 are described in Supplementary Table 6.

841

842 **Acknowledgments:**

843 We thank Drs. Elphrege Nora and Benoit Bruneau for exchanging datasets and reagents. We would

844 like to give special thanks to Samantha Kuan for operating the sequencing instruments and Tristin

845 Liu and Zhen Ye for helping experiments. We would like to acknowledge the help of Drs. Victor
846 Lobanenkov and Arshad Desai for giving helpful advice and the help of Drs. Feng Yue, Xiaotao
847 Wang, Ivan Juric, and Armen Abnousi for sharing computational pipelines. We would also like to
848 give special thanks to Drs. Rongxin Fang, Yanxiao Zhang, Ramya Raviram, Anthony Schmitt, and
849 Sora Chee for sharing helpful information and protocols, as well as all the other members of the Ren
850 laboratory. This work was supported by the Ludwig Institute for Cancer Research (B.R.), NIH
851 (1U54DK107977-01) (B.R.), NIH (1U54DK107965) (H.Z.), a Ruth L. Kirschstein Institutional National
852 Research Award from the National Institute for General Medical Sciences, T32 GM008666 (J.D.H.),
853 and a Postdoc fellowship from the TOYOBO Biotechnology Foundation (N.K.).

854

855 **Author Information:**

856 The authors declare no competing financial interests. Correspondence and requests for materials
857 should be addressed to B.R. (biren@ucsd.edu)

858 **Author contributions:**

859 N.K., H.I., and B.R. conceived the project. N.K., H.I., X.X., and H.Z. engineered cell lines. N.K., R.H.,
860 J.D.H., and Z.Y. carried out library preparation. N.K. and F.M. performed cell cycle analysis. N.K.,
861 S.B., M.C., M.N., D.G., and B.L. performed data analysis. M.Y., M.H., and J.D. contributed to

862 computational analysis and experimental design. N.K. and B.R. wrote the manuscript. All authors
863 edited the manuscript.

864

865 **References:**

- 866 1. Heintzman, N.D. *et al.* Histone modifications at human enhancers reflect global cell-type-
867 specific gene expression. *Nature* **459**, 108-12 (2009).
- 868 2. Long, H.K., Prescott, S.L. & Wysocka, J. Ever-Changing Landscapes: Transcriptional
869 Enhancers in Development and Evolution. *Cell* **167**, 1170-1187 (2016).
- 870 3. Shen, Y. *et al.* A map of the cis-regulatory sequences in the mouse genome. *Nature*
871 **488**, 116-20 (2012).
- 872 4. Andersson, R. *et al.* An atlas of active enhancers across human cell types and tissues.
873 *Nature* **507**, 455-461 (2014).
- 874 5. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome.
875 *Nature* **489**, 57-74 (2012).
- 876 6. Yu, M. & Ren, B. The Three-Dimensional Organization of Mammalian Genomes. *Annu
877 Rev Cell Dev Biol* **33**, 265-289 (2017).
- 878 7. Arzate-Mejia, R.G., Recillas-Targa, F. & Corces, V.G. Developing in 3D: the role of
879 CTCF in cell differentiation. *Development* **145**(2018).
- 880 8. Deng, W. & Blobel, G.A. Manipulating nuclear architecture. *Curr Opin Genet Dev* **25**, 1-7
881 (2014).
- 882 9. Kim, J.H. *et al.* LADL: light-activated dynamic looping for endogenous gene expression
883 control. *Nat Methods* **16**, 633-639 (2019).
- 884 10. Gaszner, M. & Felsenfeld, G. Insulators: exploiting transcriptional and epigenetic
885 mechanisms. *Nat Rev Genet* **7**, 703-13 (2006).
- 886 11. Phillips-Cremins, J.E. & Corces, V.G. Chromatin insulators: linking genome organization
887 to cellular function. *Mol Cell* **50**, 461-74 (2013).
- 888 12. Yang, J. & Corces, V.G. Insulators, long-range interactions, and genome function. *Curr
889 Opin Genet Dev* **22**, 86-92 (2012).
- 890 13. Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by analysis of
891 chromatin interactions. *Nature* **485**, 376-80 (2012).

- 892 14. Phillips-Cremins, J.E. *et al.* Architectural protein subclasses shape 3D organization of
893 genomes during lineage commitment. *Cell* **153**, 1281-95 (2013).
- 894 15. Nora, E.P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation
895 centre. *Nature* **485**, 381-5 (2012).
- 896 16. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles
897 of chromatin looping. *Cell* **159**, 1665-80 (2014).
- 898 17. Dowen, J.M. *et al.* Control of cell identity genes occurs in insulated neighborhoods in
899 mammalian chromosomes. *Cell* **159**, 374-387 (2014).
- 900 18. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals
901 folding principles of the human genome. *Science* **326**, 289-93 (2009).
- 902 19. Fullwood, M.J., Wei, C.L., Liu, E.T. & Ruan, Y. Next-generation DNA sequencing of
903 paired-end tags (PET) for transcriptome and genome analyses. *Genome Res* **19**, 521-32
904 (2009).
- 905 20. Hnisz, D., Day, D.S. & Young, R.A. Insulated Neighborhoods: Structural and Functional
906 Units of Mammalian Gene Control. *Cell* **167**, 1188-1200 (2016).
- 907 21. Lupianez, D.G. *et al.* Disruptions of topological chromatin domains cause pathogenic
908 rewiring of gene-enhancer interactions. *Cell* **161**, 1012-25 (2015).
- 909 22. Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of
910 genomic duplications. *Nature* **538**, 265-269 (2016).
- 911 23. Weischenfeldt, J. *et al.* Pan-cancer analysis of somatic copy-number alterations
912 implicates IRS4 and IGF2 in enhancer hijacking. *Nat Genet* **49**, 65-74 (2017).
- 913 24. Narendra, V. *et al.* CTCF establishes discrete functional chromatin domains at the Hox
914 clusters during differentiation. *Science* **347**, 1017-21 (2015).
- 915 25. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep*
916 **15**, 2038-49 (2016).
- 917 26. Sanborn, A.L. *et al.* Chromatin extrusion explains key features of loop and domain
918 formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* **112**, E6456-
919 65 (2015).
- 920 27. Rao, S.S.P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e24
921 (2017).
- 922 28. Nora, E.P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of
923 Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e22
924 (2017).
- 925 29. Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via
926 Cellular Genetic Screens. *Cell* **176**, 1516 (2019).

- 927 30. Fang, R. *et al.* Mapping of long-range chromatin interactions by proximity ligation-
928 assisted ChIP-seq. *Cell Res* **26**, 1345-1348 (2016).
- 929 31. Mumbach, M.R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome
930 architecture. *Nat Methods* **13**, 919-922 (2016).
- 931 32. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-
932 based degron system for the rapid depletion of proteins in nonplant cells. *Nat Methods* **6**,
933 917-22 (2009).
- 934 33. Holland, A.J., Fachinetti, D., Han, J.S. & Cleveland, D.W. Inducible, reversible system
935 for the rapid and complete degradation of proteins in mammalian cells. *Proc Natl Acad
936 Sci U S A* **109**, E3350-7 (2012).
- 937 34. Nishimura, K. & Kanemaki, M.T. Rapid Depletion of Budding Yeast Proteins via the
938 Fusion of an Auxin-Inducible Degron (AID). *Curr Protoc Cell Biol* **64**, 20 9 1-16 (2014).
- 939 35. Krijger, P.H. *et al.* Cell-of-Origin-Specific 3D Genome Structure Acquired during Somatic
940 Cell Reprogramming. *Cell Stem Cell* **18**, 597-610 (2016).
- 941 36. Li, Y. *et al.* CRISPR reveals a distal super-enhancer required for Sox2 expression in
942 mouse embryonic stem cells. *PLoS One* **9**, e114485 (2014).
- 943 37. Landolt, R.M., Vaughan, L., Winterhalter, K.H. & Zimmermann, D.R. Versican is
944 selectively expressed in embryonic tissues that act as barriers to neural crest cell
945 migration and axon outgrowth. *Development* **121**, 2303-12 (1995).
- 946 38. Wu, Y. *et al.* Versican V1 isoform induces neuronal differentiation and promotes neurite
947 outgrowth. *Mol Biol Cell* **15**, 2093-104 (2004).
- 948 39. Chiariello, A.M., Annunziatella, C., Bianco, S., Esposito, A. & Nicodemi, M. Polymer
949 physics of chromosome large-scale 3D organisation. *Sci Rep* **6**, 29775 (2016).
- 950 40. Bianco, S. *et al.* Polymer physics predicts the effects of structural variants on chromatin
951 architecture. *Nat Genet* **50**, 662-667 (2018).
- 952 41. Monahan, K. *et al.* Role of CCCTC binding factor (CTCF) and cohesin in the generation
953 of single-cell diversity of protocadherin-alpha gene expression. *Proc Natl Acad Sci U S A*
954 **109**, 9125-30 (2012).
- 955 42. He, Y. *et al.* Spatiotemporal DNA Methylome Dynamics of the Developing Mammalian
956 Fetus. *bioRxiv, preprint* (2017).
- 957 43. Martinez, O. & Reyes-Valdes, M.H. Defining diversity, specialization, and gene
958 specificity in transcriptomes through information theory. *Proc Natl Acad Sci U S A* **105**,
959 9709-14 (2008).

- 960 44. Lee, D.P. *et al.* Robust CTCF-Based Chromatin Architecture Underpins Epigenetic
961 Changes in the Heart Failure Stress-Gene Response. *Circulation* **139**, 1937-1956
962 (2019).
- 963 45. Weintraub, A.S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell*
964 **171**, 1573-1588 e28 (2017).
- 965 46. Beagan, J.A. *et al.* YY1 and CTCF orchestrate a 3D chromatin looping switch during
966 early neural lineage commitment. *Genome Res* **27**, 1139-1152 (2017).
- 967 47. Deng, W. *et al.* Controlling long-range genomic interactions at a native locus by targeted
968 tethering of a looping factor. *Cell* **149**, 1233-44 (2012).
- 969 48. Caputo, L. *et al.* The Isl1/Ldb1 Complex Orchestrates Genome-wide Chromatin
970 Organization to Instruct Differentiation of Multipotent Cardiac Progenitors. *Cell Stem Cell*
971 **17**, 287-99 (2015).
- 972 49. Monahan, K., Horta, A. & Lomvardas, S. LHX2- and LDB1-mediated trans interactions
973 regulate olfactory receptor choice. *Nature* **565**, 448-453 (2019).
- 974 50. Sanyal, A., Lajoie, B.R., Jain, G. & Dekker, J. The long-range interaction landscape of
975 gene promoters. *Nature* **489**, 109-13 (2012).
- 976 51. Jung, I. *et al.* A compendium of promoter-centered long-range chromatin interactions in
977 the human genome. *Nat Genet* **51**, 1442-1449 (2019).
- 978 52. Katainen, R. *et al.* CTCF/cohesin-binding sites are frequently mutated in cancer. *Nat*
979 *Genet* **47**, 818-21 (2015).
- 980 53. Guo, Y.A. *et al.* Mutation hotspots at CTCF binding sites coupled to chromosomal
981 instability in gastrointestinal cancers. *Nat Commun* **9**, 1520 (2018).
- 982 54. Kaiser, V.B., Taylor, M.S. & Semple, C.A. Mutational Biases Drive Elevated Rates of
983 Substitution at Regulatory Sites across Cancer Types. *PLoS Genet* **12**, e1006207
984 (2016).
- 985 55. Cancer Genome Atlas Research, N. *et al.* Integrated genomic characterization of
986 endometrial carcinoma. *Nature* **497**, 67-73 (2013).
- 987 56. Gribnau, J., Hochedlinger, K., Hata, K., Li, E. & Jaenisch, R. Asynchronous replication
988 timing of imprinted loci is independent of DNA methylation, but consistent with
989 differential subnuclear localization. *Genes Dev* **17**, 759-73 (2003).
- 990 57. Sakuma, T., Nishikawa, A., Kume, S., Chayama, K. & Yamamoto, T. Multiplex genome
991 engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci Rep* **4**,
992 5400 (2014).
- 993 58. Cheng, A.W. *et al.* Multiplexed activation of endogenous genes by CRISPR-on, an RNA-
994 guided transcriptional activator system. *Cell Res* **23**, 1163-71 (2013).

- 995 59. Sanjana, N.E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for
996 CRISPR screening. *Nat Methods* **11**, 783-4 (2014).
- 997 60. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler
998 transform. *Bioinformatics* **25**, 1754-60 (2009).
- 999 61. Ramirez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data
1000 analysis. *Nucleic Acids Res* **44**, W160-5 (2016).
- 1001 62. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137
1002 (2008).
- 1003 63. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
1004 for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
- 1005 64. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21
1006 (2013).
- 1007 65. Anders, S., Pyl, P.T. & Huber, W. HTSeq--a Python framework to work with high-
1008 throughput sequencing data. *Bioinformatics* **31**, 166-9 (2015).
- 1009 66. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for
1010 differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-
1011 40 (2010).
- 1012 67. Hu, M. *et al.* HiCNorm: removing biases in Hi-C data via Poisson regression.
1013 *Bioinformatics* **28**, 3131-3 (2012).
- 1014 68. Durand, N.C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with
1015 Unlimited Zoom. *Cell Syst* **3**, 99-101 (2016).
- 1016 69. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage
1017 compensation. *Nature* **523**, 240-4 (2015).
- 1018 70. Vian, L. *et al.* The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* **175**,
1019 292-294 (2018).
- 1020 71. Juric, I. *et al.* MAPS: Model-based analysis of long-range chromatin interactions from
1021 PLAC-seq and HiChIP experiments. *PLoS Comput Biol* **15**, e1006982 (2019).
- 1022 72. Bhattacharyya, S., Chandra, V., Vijayanand, P. & Ay, F. Identification of significant
1023 chromatin contacts from HiChIP data by FitHiChIP. *Nat Commun* **10**, 4221 (2019).
- 1024 73. Zhou, X. *et al.* Exploring long-range genome interactions using the WashU Epigenome
1025 Browser. *Nat Methods* **10**, 375-6 (2013).
- 1026 74. Engreitz, J.M. *et al.* Local regulation of gene expression by lncRNA promoters,
1027 transcription and splicing. *Nature* **539**, 452-455 (2016).
- 1028 75. Diao, Y. *et al.* A tiling-deletion-based genetic screen for cis-regulatory element
1029 identification in mammalian cells. *Nat Methods* **14**, 629-635 (2017).

- 1030 76. Barbieri, M. *et al.* Complexity of chromatin folding is captured by the strings and binders
1031 switch model. *Proc Natl Acad Sci U S A* **109**, 16173-8 (2012).
- 1032 77. Kremer, K. & Grest, G.S. Dynamics of entangled linear polymer melts: A molecular -
1033 dynamics simulation. *The Journal of Chemical Physics* **92**, 5057-5086 (1990).
- 1034 78. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *Journal of*
1035 *computational physics* **117**, 1-19 (1995).
- 1036 79. Yang, T. *et al.* HiCRep: assessing the reproducibility of Hi-C data using a stratum-
1037 adjusted correlation coefficient. *Genome Res* **27**, 1939-1949 (2017).
- 1038 80. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast
1039 genomes. *Genome Res* **15**, 1034-50 (2005).
- 1040 81. Sloan, C.A. *et al.* ENCODE data at the ENCODE portal. *Nucleic Acids Res* **44**, D726-32
1041 (2016).

1042

1043 **Figure legends**

1044 **Fig. 1 | CTCF loss affects only a small fraction of genes in gene regulation during embryonic**

1045 **stem cell differentiation.**

1046 **a**, Schematic representation of experimental design and sample preparation. Auxin-inducible degron

1047 system was utilized to deplete CTCF during cell differentiation from mouse embryonic stem cells

1048 (mESCs) to neural progenitor cells (NPCs) (day 2, 4, 6). An additional 2 days of neural differentiation

1049 treatment was performed after washing out auxin in day 4 and day 6 differentiated cells. RNA-seq

1050 and *in situ* Hi-C were performed at each time point to examine the impact of CTCF loss on gene

1051 regulation and chromatin architecture.

1052

1053 **b**, Microscopic images of cell differentiation from mESCs (top) to NPCs (day 2, 4, 6) with or without

1054 auxin treatment and additional 2 days of differentiation treatment after washing out auxin from day 4,

1055 6 samples. Alkaline phosphatase staining was performed at every time point. Non-stained bright-

1056 field images of each auxin treated sample and auxin washout sample are also shown on the right.

1057

1058 **c**, Principal component analysis of gene expression profiles of control and CTCF-depleted cells at

1059 each time point of cell differentiation and 2 days after washing out auxin. Gene expression profiles in

1060 ESCs with multiple days of auxin treatment (24, 48, and 96 hours) were also analyzed. Two

1061 replicates of each sample are shown.

1062

1063 **d**, Gene expression changes between control ESCs and NPCs (day 4). Differentially up-regulated

1064 and down-regulated genes are plotted in red and blue, respectively (fold change > 2, FDR < 0.05).

1065

1066 **e**, Gene expression changes upon CTCF depletion in ESCs (left, 48 hours with or without auxin) and

1067 in differentiated cells (right, differentiation day 4 with or without auxin). Differentially up-regulated and

1068 down-regulated genes are plotted in red and blue, respectively (fold change > 2, FDR < 0.05).

1069

1070 **f**, Top two enriched GO terms of gene sets of differentially expressed genes upon CTCF loss are

1071 shown with their p values.

1072

1073 **Fig. 2 | A small number of enhancer-promoter contacts are affected upon CTCF loss despite**

1074 **the drastic weakening of TADs.**

1075 **a**, Schematic representation of study design to explore the changes of enhancer-promoter (E-P)

1076 contacts during neural differentiation in the presence and absence of CTCF. H3K4me3 PLAC-seq

1077 datasets were analyzed for changes of E-P contacts at high resolution. ChIP-seq (histone
1078 modifications) and RNA-seq (gene regulation) datasets at each time point were also used for
1079 combinational analysis.

1080

1081 **b**, Genome browser snapshots of a region around *Hoxb3–9* genes that were up-regulated during
1082 neural differentiation. The arcs show the changes of H3K4me3 PLAC-seq contacts on active
1083 elements and promoters identified in differential interaction analysis between ESCs and NPCs (see
1084 Methods). The colors of arcs represent degrees of interaction change between compared samples
1085 (blue to red, $-/+ \log_{10}(p\text{-value})$). The promoter regions of *Hoxb3–9* genes and interacting enhancer
1086 regions are shown in green and yellow shadows, respectively. CTCF, H3K4me1, H3K27ac,
1087 H3K4me3, H3K27me3 ChIP-seq and RNA-seq in ESCs and NPCs (day 4) are also shown.

1088

1089 **c**, Scatter plots showing genome-wide changes of chromatin contacts anchored on promoters and
1090 enhancers (y-axis) identified in differential interaction analysis between ESCs and NPCs. Genomic
1091 distances between their two loop anchor sites are plotted in x-axis. Significantly induced and
1092 reduced chromatin contacts are shown as red and blue dots, respectively (FDR < 0.05). The
1093 interaction changes are shown by significance value ($-/+ \log_{10}(p\text{-value})$) (see Methods).

1094

1095 **d**, Scatter plots showing genome-wide changes of chromatin contacts anchored on promoters and
1096 enhancers (*y*-axis) identified in differential interaction analysis between control and CTCF-depleted
1097 cells in ESC (left) and NPC stage (day 4) (right). Genomic distances between their two loop anchor
1098 sites are plotted in *x*-axis. Significantly induced and reduced chromatin contacts are shown as red
1099 and blue dots, respectively (FDR < 0.05).

1100

1101 **e**, Genome browser snapshots of a region around the *Pcdhg* gene cluster down-regulated upon
1102 CTCF loss in NPCs. The arcs show the changes of chromatin contacts on enhancers and promoters
1103 (E-P) and chromatin contacts on CTCF binding sites (CTCF-BS) identified in differential interaction
1104 analysis between conditions. The colors of arcs represent degrees of interaction change from control
1105 cells to CTCF-depleted cells (blue to red, $-/+ \log_{10}(p\text{-value})$). The promoter regions of *Pcdhg* gene
1106 cluster and interacting enhancer regions are shown in green and yellow shadows, respectively.
1107 CTCF, H3K4me1, H3K27ac, H3K4me3, H3K27me3 ChIP-seq, and RNA-seq in control and CTCF-
1108 depleted NPCs, and TAD boundaries in control cells are also shown.

1109

1110 **f**, On the scatter plots of changes of E-P contacts upon CTCF loss in NPCs (Fig. 2d, right panel), E-

1111 P contacts that were significantly induced during neural differentiation were plotted in red and

1112 significantly reduced contacts upon CTCF loss among them were plotted in green.

1113

1114 **g**, Histograms showing the number of significantly changed E-P contacts upon CTCF loss and their

1115 genomic distances in ESC (left) and NPC (right) stages. Significantly induced and reduced contacts

1116 are shown as red and blue bars, respectively.

1117

1118 **Fig. 3 | Enriched CBSs at promoters of CTCF-dependent genes and rescue of CTCF-**

1119 **dependent enhancer-promoter contacts and gene regulation by artificially tethered CTCF.**

1120 **a, b**, Enrichment analysis of CTCF-dependent E-P contacts categorized based on the distance from

1121 the loop anchor sites on enhancer side (vertical columns) or promoter side (horizontal columns) to

1122 the nearest CTCF binding site (CBS) (a). The same enrichment analysis categorized based on the

1123 number of CBSs around loop anchor sites (10 kb bin \pm 5 kb) on enhancer side (vertical columns) or

1124 promoter side (horizontal columns) was also performed (b). Enrichment values are shown by odds

1125 ratio (scores in boxes) and p-values (color) in ESC (left) and NPC stage (right) (see Methods).

1126

1127 **c, d**, Enrichment analysis of E-P contacts on CTCF-dependent down-regulated genes categorized
1128 based on the distance from the loop anchor sites on enhancer side (vertical columns) or promoter
1129 side (horizontal columns) to the nearest CBS (c). The same enrichment analysis categorized based
1130 on the number of CBSs around loop anchor sites (10 kb bin \pm 5 kb) on enhancer side (vertical
1131 columns) or promoter side (horizontal columns) was also performed (d). Enrichment values are
1132 shown by odds ratio (scores in boxes) and p-values (color) in ESC (left) and NPC stage (right) (see
1133 Methods).

1134

1135 **e**, Genome browser snapshots of a region around *Vcan* gene down-regulated upon CTCF loss in
1136 NPCs. The arcs show the changes of chromatin contacts on enhancers and promoters (E-P) and
1137 chromatin contacts on CTCF binding sites (CTCF-BS) identified in differential interaction analysis
1138 between conditions. The colors of arcs represent degrees of interaction change from control to
1139 CTCF-depleted cells (blue to red, $-/+ \log_{10}(p\text{-value})$). *Vcan* promoter and 400 kb downstream distal
1140 active element are shown in green and yellow shadows, respectively. CTCF, H3K4me1, H3K27ac,
1141 H3K4me3 ChIP-seq and TAD boundaries in control NPCs are also shown. Schematic representation
1142 and description of each cell line is aligned on the bottom.

1143

1144 **f**, RT-qPCR expression levels of *Vcan* gene in each cell line in NPC stage. (** p value < 0.01 and ***

1145 p value < 0.001, two-tailed t-test).

1146

1147 **g**, Changes of chromatin contacts around *Vcan* gene upon artificial tethering of CTCF at the

1148 promoter on top and bottom strand are shown. The arcs show the changes of chromatin contacts

1149 anchored on enhancers, promoters, and CTCF binding sites identified in differential interaction

1150 analysis between conditions. The colors of arcs represent degrees of interaction change from control

1151 NPCs (dCas9 alone) to dCas9-CTCF tethered NPCs (blue to red, -/+ $\log_{10}(p\text{-value})$). CTCF,

1152 H3K4me1, H3K27ac, H3K4me3, H3K27me3 ChIP-seq and TAD boundaries in NPCs are also

1153 shown.

1154

1155 **h**, Schematic representation of observed findings in the rescue experiments. The deletion of the

1156 CTCF binding motif adjacent to *Vcan* promoter disrupts its E-P contact along with loss of the CTCF-

1157 anchored loop (Supplementary Fig. 9a). Then, the artificially tethered CTCF at the promoter on top

1158 or bottom strand rescues the E-P contact and gene regulation.

1159

1160 **Fig. 4 | General features of CTCF-dependent/-independent genes.**

1161 **a**, Boxplots showing the distance from transcription start site (TSS) to the nearest H3K27ac peak in
1162 ESC (left) and NPC stage (right). Red: CTCF-dependent up-regulated genes, blue: CTCF-
1163 dependent down-regulated genes, gray: CTCF-independent stably regulated genes. (***) p
1164 value < 0.001, two-tailed t-test).

1165

1166 **b**, Enrichment analysis of CTCF-independent genes categorized based on the distance to the
1167 nearest interacting enhancer (vertical columns) and the number of enhancers around TSS (< 200 kb)
1168 (horizontal columns) in ESCs (left) and NPCs (right). Enrichment values are shown by odds ratio
1169 (scores in boxes) and p-values (color). The distance to the nearest interacting enhancer is
1170 represented by the shortest genomic distance of significant PLAC-seq peaks on enhancers and
1171 promoters (p-value < 0.01) (see Methods).

1172

1173 **c**, Enrichment analysis of CTCF-dependent down-regulated genes (left) and up-regulated genes
1174 (right) categorized based on the distance to the nearest interacting enhancer (vertical columns) and
1175 the number of enhancers around TSS (< 200 kb) (horizontal columns) in NPCs. Enrichment values
1176 are shown by odds ratio (scores in boxes) and p-values (color). The distance to the nearest
1177 interacting enhancer is represented by the shortest genomic distance of significant PLAC-seq peaks

1178 on enhancers and promoters (p -value < 0.01) (see Supplementary Fig. 10 f for the same analysis in

1179 ESCs and Methods).

1180

1181 **d**, Model for the general features of CTCF-dependent down-regulated (top), up-regulated genes

1182 (middle), and CTCF-independent stably regulated genes (bottom).

1183

1184 **Fig. 5 | CTCF bindings on gene promoters correlates with tissue-specific gene regulation.**

1185 **a**, Frequency of genomic regions and their density of CTCF motifs are plotted. The genomic regions

1186 were classified into promoters, enhancers (identified in ESCs and NPCs), gene bodies, and random

1187 regions.

1188

1189 **b**, Schematic representation of data analysis for correlation between the sum of CTCF ChIP-seq

1190 signals around TSS (< 10 kb) and gene expression levels across multiple mouse tissues (top, see

1191 Methods). Frequencies of genes are plotted based on the correlation coefficient between the CTCF

1192 ChIP-seq signals around TSS and gene expression levels across multiple mouse tissues (red line).

1193 The same plots analyzed using randomly shuffled CTCF ChIP-seq datasets are shown in gray as

1194 control (** p value < 0.001, Pearson's Chi-squared test for the comparison of fraction of genes with

1195 positive correlation coefficient ($r \geq 0.6$) or others ($r < 0.6$) between the two groups).

1196

1197 **c**, Heatmaps showing the lineage-specificity of CTCF ChIP-seq signals around promoters and their

1198 gene expression levels. Genes that had high correlation coefficient (> 0.6) in panel (b) are shown

1199 (2,332 genes). The lineage-specificity was calculated by $\log_2(\text{value} / \text{average value of all tissues})$.

1200 The violin plots also show the lineage-specificity of transcription measured by Shannon entropy in

1201 each gene group. The width is proportional to the sample size. Top enriched GO terms of each

1202 group genes are shown with fold enrichment, p-value, and their representative genes.

1203

1204 **Supplementary Fig. 1 | Depletion of CTCF and characterization of CTCF-depleted cells.**

1205 **a, b,** Western blot showing AID-tagged CTCF and wild type CTCF (a) and the expression of TIR1
1206 protein in 2 clones of mESCs. The TIR1 expression in these clones that went through multiple
1207 passages were comparable to that in control cells that have much smaller passage number.

1208

1209 **c,** Western blot showing the acute depletion of CTCF protein after 24 and 48 hours of auxin
1210 treatment.

1211

1212 **d,** Heatmaps showing CTCF ChIP-seq signals centered at all regions of CTCF peaks identified in
1213 the control cells and CTCF occupancy at the same regions in CTCF-depleted cells at each time
1214 point of ESC and NPC stages. The CTCF peaks disappeared almost completely after auxin
1215 treatment and recovered after washing out auxin.

1216

1217 **e,** Venn-diagram comparing the number of CTCF ChIP-seq peaks identified in control and CTCF-
1218 depleted ESCs at each time point.

1219

1220 **f**, Histogram showing the number of CTCF binding regions in y-axis and their CTCF ChIP-seq signal

1221 levels in x-axis. The CTCF signal levels in control cells and auxin treated cells were calculated on

1222 the CTCF peak regions identified in the control cells.

1223

1224 **g**, Heatmaps comparing the Rad21 ChIP-seq signals centered at all regions of Rad21 peaks

1225 identified in control ESCs and in CTCF-depleted ESCs at each time point (left, blue heat map). The

1226 CTCF occupancy at the same regions in the same samples are also shown (right, red heat map).

1227

1228 **h**, Growth curves of mouse ESCs with or without auxin treatment.

1229

1230 **i**, Bright-field microscopy images of mouse ESC colonies before and after multiple days of auxin

1231 treatment.

1232

1233 **j**, Cell cycle analysis by flow cytometry using propidium iodide staining in control and 24, 48, and 96

1234 hours auxin treated CTCF-depleted ESCs suggests that CTCF-depleted cells can pass through cell

1235 cycle checkpoints.

1236

1237 **Supplementary Fig. 2 | Severe disruption of chromatin architecture upon CTCF loss.**

1238 **a,** Scatter plots showing insulation scores at TAD boundaries. A higher score denotes lower
1239 insulation. Increased insulation scores in varying degrees were observed in CTCF-depleted ESCs
1240 and NPCs.

1241

1242 **b,** Aggregate boundary analysis showing the average change in boundary strength between
1243 samples. Each triangle is a contact map showing the difference in the average contact profile at TAD
1244 boundaries between two time points. The bottom column shows that there is little difference in the
1245 average boundary profile between the two control samples.

1246

1247 **c,** The number of TAD boundaries (left), stripes (middle), and insulated neighborhoods (INs) in
1248 control, CTCF-depleted, and auxin washout cells. The hatched bars indicate the number of
1249 boundaries/stripes/INs that overlapped with boundaries/stripes/INs identified in control cells.

1250

1251 **d,** Hi-C contact frequencies at each genomic distance are plotted. Overall profiles of Hi-C contact
1252 frequencies were not affected by the loss of CTCF but affected by cell differentiation.

1253

1254 **e**, APA (aggregate peak analysis) on Hi-C peak loci on convergent CTCF binding sites identified in

1255 control ESCs and NPCs (n=3185 (ESCs), 3686 (NPCs), > 100-kb looping range). APA was

1256 performed in control and CTCF-depleted cells in ESC and NPC stages and also performed after

1257 washing out auxin in NPCs. APA on Hi-C peak loci that have no CTCF binding sites (n=2874

1258 (ESCs), 2940 (NPCs), > 100-kb looping range) was also performed. The scores on the bottom

1259 represent the focal enrichment of peak pixel against pixels in its lower left.

1260

1261 **f–h**, Hi-C datasets generated in this study and Nora et al. study²⁸. Genome browser snapshots

1262 showing Hi-C contact heatmaps, TAD boundaries, directionality indices (DIs), and insulation scores

1263 analyzed in the two independent studies at the same genomic region in control and CTCF-depleted

1264 cells (f). Scatter plots showing insulation scores at all TAD boundaries (g). The number of TAD

1265 boundaries in control and CTCF-depleted cells from the two studies. The hatched bars indicate the

1266 number of boundaries that overlapped with the boundaries identified in control cells (h). The degree

1267 of weakening of TAD boundaries was comparable between the two studies.

1268

1269 **Supplementary Fig. 3 | Examples of gene expression changes during neural differentiation in**

1270 **control and CTCF-depleted cells.**

1271 **a**, Gene expression profiles of pluripotent marker genes (*Pou5f1*, *Sox2*, *Nanog*) and examples of

1272 induction failure gene upon CTCF loss that is important for nervous system development (*Neurog1*,

1273 (*Neurod4*, *Vcan*) in control and CTCF-depleted cells during differentiation from ESC to NPC and 2

1274 days after washing out auxin in NPCs.

1275

1276 **b**, Gene expression profiles of *Pcdhga* and *Hoxc* gene clusters during multiple days of auxin

1277 treatment in ESCs and during differentiation from ESC to NPC in control and CTCF-depleted cells

1278 followed by washing out auxin in NPCs

1279

1280 **Supplementary Fig. 4 | Differential chromatin interaction analysis of PLAC-seq datasets.**

1281 **a**, Schematic representation of differential interaction analysis in H3K4me3 PLAC-seq datasets. The

1282 mapped contact counts in 10-kb resolution bins that have the same genomic distance were

1283 compared separately. The input contact matrix bins were stratified into every 10-kb genomic

1284 distance from 10-kb to 150-kb and the other bins with longer distances were stratified to have a

1285 uniform number of input bins that were equal to that of 140–150 kb distance bins. Two sets of these

1286 discrete inputs of two biological replicates were compared using a negative binomial model,

1287 edgeR⁵⁸. The interactions anchored at the differential H3K4me3 ChIP-seq peaks between the

1288 conditions ($\text{FDR} < 0.05$, $|\log\text{FC}| > 0.5$) were removed and only the genes with the same level of

1289 H3K4me3 ChIP-seq peaks on promoters were processed in downstream analysis (see Methods).

1290

1291 **b**, Differential interaction analysis between ESC and NPC (top) in each distance-stratified 10-kb

1292 interval. The difference between ESC replicate 1 and replicate 2 are also shown for comparison

1293 (bottom).

1294

1295 **Supplementary Fig. 5 | Changes of enhancer-promoter contacts during neural differentiation**

1296 **were associated with the changes of chromatin modification and gene expression.**

1297 **a**, Histogram showing the number of differentially changed enhancer-promoter (E-P) contacts

1298 between control ESCs and NPCs and their genomic distances. While the majority of them (53%)

1299 were less than 50 kb distance, a large number of long-range E-P contacts were induced during

1300 differentiation. Red bars: significantly induced E-P contacts. Blue bars: significantly reduced E-P

1301 contacts.

1302

1303 **b, c**, Scatter plots showing changes of ChIP-seq peak levels of H3K27ac (c) and H3K4me1 (d) on

1304 distal element loci of induced (red) and reduced (blue) E-P contacts during neural differentiation.

1305

1306 **d, e,** Genome browser snapshots of regions around *Sox2* gene (a) and *Dnmt3b* gene (b) that were
1307 down-regulated during neural differentiation. The arcs show the changes of chromatin contacts on
1308 active elements and promoters identified in differential interaction analysis between ESCs and
1309 NPCs. The colors of arcs represent degrees of interaction change between the conditions (blue to
1310 red, $-/+ \log_{10}(p\text{-value})$). The promoter regions of these genes and interacting enhancer regions are
1311 shown in green and yellow shadows, respectively. CTCF, H3K4me1, H3K27ac, H3K4me3,
1312 H3K27me3 ChIP-seq and RNA-seq in ESCs and NPCs (day 4) are also shown.

1313

1314 **f,** Schematic representation of Active/Inactive contact (AIC) model to see the correlation between
1315 changes of multiple E-P contacts and gene expression levels. H3K27ac and H3K4me1 peaks are
1316 shown as red and yellow peaks, respectively, and regions where these two types of peaks are
1317 overlapping are defined as active elements (red colored regions). Promoter-centered chromatin
1318 contacts on these active elements are shown as red arcs (active contacts) and the other chromatin
1319 contacts are shown as blue arcs (inactive contacts). AIC ratio/value was calculated by the formulae
1320 described on the bottom (see Methods).

1321

1322 **g**, Scatter plots showing the changes of AIC values (y-axis) and gene expression levels (x-axis) in

1323 differentially expressed genes during neural differentiation with linear approximation.

1324

1325 **Supplementary Fig. 6 | CTCF loss does not alter histone modification landscape.**

1326 **a–c**, Scatter plots showing the changes of H3K27ac (a) and H3K4me1 (b) ChIP-seq signal levels

1327 upon CTCF loss on all significant peak regions in ESCs (left) and NPCs (right). The changes of

1328 H3K4me3 ChIP-seq signal levels on all peak regions on transcription start sites (TSSs) are also

1329 shown (c).

1330

1331 **d**, Boxplots showing the changes of H3K27ac (top) and H3K4me1 (bottom) ChIP-seq signal levels

1332 on distal element loci of all analyzed E-P contacts. The changes upon CTCF depletion in ESC (left)

1333 and NPC stage (middle) and the changes during neural differentiation (right) are shown. (NS not

1334 significant, *** p value < 0.001, two-tailed t-test).

1335

1336 **Supplementary Fig. 7 | Changes of chromatin contacts upon CTCF loss and features of**

1337 **CTCF-dependent enhancer-promoter contacts.**

1338 **a, b**, Scatter plots showing changes of H3K4me3 PLAC-seq contacts (y-axis) on convergently
1339 oriented CTCF binding sites and their loop ranges (x-axis). Chromatin contacts in CTCF-depleted
1340 cells were compared to the chromatin contacts in control cells in ESC (a) and NPC stage (day 4) (b).
1341 The plots were classified based on whether they are on promoters and enhancers (E-P) (left) or not
1342 (Non-E-P) (right).
1343
1344 **c**, Average enrichments of CTCF ChIP-seq signals on TSSs of CTCF-dependent up-regulated
1345 genes (red), CTCF-dependent down-regulated genes (blue), and the other CTCF-independent stably
1346 regulated genes (gray) upon CTCF loss in ESC (left) and NPC stage (right).
1347
1348 **d**, The number of CTCF-dependent differentially reduced E-P contacts anchored on convergently
1349 oriented CBSs (orange) and the CTCF-dependent E-P contacts anchored without convergently
1350 oriented CBSs (blue) in ESCs (left) and NPCs (right).
1351
1352 **e**, Scatter plots showing changes of chromatin contacts anchored on promoters of CTCF-dependent
1353 down-regulated genes and distal active elements (y-axis). Genomic distances between their two
1354 loop anchor sites are plotted in x-axis. Chromatin contacts in CTCF-depleted cells were compared to

1355 the chromatin contacts in control cells in ESC (top) and NPC stage (day 4) (bottom). The E-P
1356 contacts were classified based on following four conditions. CBSs are located at both anchor sites
1357 (10 kb bin \pm 5 kb) (red dots). CBSs are located only at promoter side of their anchor sites (green
1358 dots). CBSs are located only at distal element side of their anchor sites (blue dots). CBSs are
1359 located at neither anchor sites (gray dots). The total number of analyzed bin pairs are added.
1360
1361 **f**, Genome browser snapshots of a region around *Baiap2* gene up-regulated upon CTCF loss in
1362 NPCs. The arcs show the changes of chromatin contacts on enhancers and promoters (E-P) and
1363 chromatin contacts on CTCF binding sites (CTCF-BS). The colors of arcs represent degrees of
1364 interaction change from control cells to CTCF-depleted cells (blue to red, $-/+ \log_{10}(p\text{-value})$). The
1365 promoter regions of *Baiap2* gene and interacting enhancer regions are shown in green and yellow
1366 shadows respectively. Newly formed E-P contacts were observed in *Baiap2* gene promoter. CTCF,
1367 H3K4me1, H3K27ac, H3K4me3, H3K27me3 ChIP-seq, and RNA-seq in control and CTCF-depleted
1368 NPCs, and TAD boundaries in control cells are also shown.
1369
1370 **g**, Boxplots showing the number of CTCF binding sites that were located in between the two anchor
1371 sites of the following three types of E-P contacts. Significantly induced E-P contacts upon CTCF loss

1372 (red), significantly reduced E-P contacts upon CTCF loss (CTCF-dependent) (blue), and unchanged

1373 E-P contacts upon CTCF loss (CTCF-independent) (gray). (** p value < 0.001, two-tailed t-test).

1374

1375 **h**, The fraction of E-P contacts that were overlapped with TAD boundaries in ESC (left) and NPC

1376 stage (left). Red: significantly induced E-P contacts upon CTCF loss. Blue: significantly reduced E-P

1377 contacts upon CTCF loss (CTCF-dependent). Gray: unchanged E-P contacts upon CTCF loss

1378 (CTCF-independent).

1379

1380 **Supplementary Fig. 8 | Verification of chromatin contacts change upon CTCF loss by SBS**

1381 **polymer model.**

1382 **a**, The Strings & Binders Switch (SBS) model showing triplet interactions between *Vcan* promoter

1383 (black), distal enhancer (blue) and distal CTCFs (-) (reds), that weaken upon CTCF depletion (white

1384 arrows). Heatmaps from each viewpoint in control NPCs and CTCF-depleted NPCs are shown.

1385 CTCFs (+) (browns) and CTCFs (-) (reds) are convergently oriented.

1386

1387 **b**, Hi-C contact maps (left) of the *Vcan* locus in NPC control and auxin treated cells are well

1388 reproduced by the SBS polymer model (right) (HiCRep stratum adjusted correlation SCC = 0.76 and

1389 SCC = 0.62 respectively). Genomic positions of *Vcan* promoter (black), distal enhancer (blue) and

1390 relevant motif-oriented CTCF sites (brown and red) are shown by colored triangles.

1391

1392 **c**, SBS derived 3D structures of the control NPCs and CTCF-depleted NPCs, with relevant elements

1393 indicated by colored beads (the color code is the same as panel b). The models capture the loss of

1394 contact between *Vcan* promoter and its distal enhancer upon CTCF depletion.

1395

1396 **Supplementary Fig. 9 | Rescue experiments using dCas9-CTCF.**

1397 **a**, Genome browser snapshots of a region around *Vcan* gene. The arcs show the changes of

1398 chromatin contacts anchored on *Vcan* promoter, distal enhancer, and CTCF binding sites identified

1399 in differential interaction analysis between wild type NPCs and the promoter-proximal CTCF motif

1400 sequences deleted NPCs. The colors of arcs represent degrees of interaction change upon the

1401 deletion of CTCF motif sequences (blue to red, $-/+ \log_{10}(p\text{-value})$). The promoter region and

1402 interacting enhancer region are shown in green and yellow shadows, respectively. CTCF, H3K27ac,

1403 H3K4me1, H3K4me3, and H3K27me3 ChIP-seq, and TAD boundaries in wild type NPCs are also

1404 shown. The deletion of the single CTCF binding site leaded to disruption of its E-P contact.

1405

1406 **b**, Schematic representation of the dCas9-CTCF rescue experiments showing targeted DNA

1407 sequences for CTCF motif deletion and dCas9-CTCF recruitment.

1408

1409 **c**, Western blot showing that the cells transfected with dCas9-CTCF or dCas9 alone were

1410 successfully expressing dCas9-CTCF or dCas9 proteins.

1411

1412 **Supplementary Fig. 10 | Mechanisms of CTCF-dependent/-independent gene regulation.**

1413 **a**, Fraction of genes classified based on genomic distance from TSS to the nearest CTCF ChIP-seq

1414 peak. CTCF-dependent down-regulated genes, up-regulated genes, and CTCF-independent stably

1415 regulated genes in ESCs (top) and NPCs (bottom) are shown.

1416

1417 **b**, Average enrichments of H3K27ac ChIP-seq signals on TSSs of CTCF-dependent up-regulated

1418 genes (red), CTCF-dependent down-regulated genes (blue), and the other CTCF-independent stably

1419 regulated genes (gray) in ESC (left) and NPC stage (right). Average enrichments of H3K27me3

1420 ChIP-seq signals on TSSs and transcription end sites (TESs) are also shown (bottom).

1421

1422 **c**, Schematic representation showing two types of genes that had CTCF bindings on their TSSs (< 1
1423 kb). In Gene A, the shortest E-P contact (PLAC-seq peak signal p-value < 0.01) is shorter than 50 kb
1424 genomic distance and there are 7 enhancers or more around TSS (< 200 kb). In Gene B, the
1425 shortest E-P contact is longer than 50 kb and there are 2 enhancers or less around TSS. Boxplots
1426 showing gene expression changes upon CTCF loss in Gene A and Gene B in ESCs (left) and NPCs
1427 (right), suggesting that down-regulation upon loss of promoter-proximal CTCF is milder when genes
1428 are surrounded by proximal enhancers and regulated by short-range E-P contacts. (* p value < 0.05
1429 and *** p value < 0.001, two-tailed t-test)
1430
1431 **d, e**, Genome browser snapshot of a region around Sox2 gene (d) whose expression level was not
1432 changed significantly 24 or 48 hours after CTCF depletion in ESCs in RNA-seq and qPCR (e). The
1433 arcs show PLAC-seq contact counts in control (top) and CTCF-depleted ESCs (middle) at every 10-
1434 kb bin. The changes of chromatin contacts on enhancers and Sox2 gene promoter identified in
1435 differential interaction analysis between the two conditions are also shown (bottom). Sox2 gene
1436 promoter and interacting super enhancer are shown in green and yellow shadows, respectively.
1437 CTCF, H3K4me1, H3K27ac, H3K4me3, and H3K27me3 ChIP-seq, RNA-seq in control and CTCF-
1438 depleted ESCs are shown. Consistent with the relatively stable gene regulation upon CTCF loss, the

1439 changes of E-P contacts were also mild even though convergent CTCF binding sites are located

1440 around its anchor sites.

1441

1442 **f**, Enrichment analysis of CTCF-dependent down-regulated genes (left) and up-regulated genes

1443 (right) categorized based on the distance to the nearest interacting enhancer (vertical columns) and

1444 the number of enhancers around TSS (< 200 kb) (horizontal columns) in ESCs. Enrichment values

1445 are shown by odds ratio (scores in boxes) and p-values (color). The distance to the nearest

1446 interacting enhancer is represented by the shortest genomic distance of significant PLAC-seq peaks

1447 on enhancers and promoters (p-value < 0.01). (see Fig. 4 c for the same analysis in NPCs and

1448 Methods).

1449

1450 **Supplementary Fig. 11 | Examples of changes in enhancer-promoter contacts upon CTCF**

1451 **loss.**

1452 **a**, Genome browser snapshots of regions around *Hoxc4*, *Hoxc5*, *Hoxc8* genes and *Pax9* gene that

1453 were down-regulated upon CTCF loss in NPCs. Data display shown here is the same as

1454 Supplementary Fig. 7f.

1455

1456 **b**, Genome browser snapshots of regions around *Lmo2*, *Kif26b*, *Osr2*, *Hoxa11*, and *Hoxa13* genes

1457 that were up-regulated upon CTCF loss in ESCs or NPCs. Data display shown here is the same as

1458 Supplementary Fig. 7f.

1459

1460 **Supplementary Fig. 12 | Features of tissue-specific CTCF occupied promoter genes.**

1461 **a**, Histogram showing frequencies of genomic regions that has 2 CBSs or more in all analyzed 9

1462 tissues. They are classified based on GC content levels. Red bars show the frequencies in promoter

1463 regions (TSS ±10 kb) and blue bars show the frequencies in random regions other than promoter

1464 regions. Black line shows fold change between the two groups in each GC content level. (** p

1465 value < 0.001, two-tailed t-test).

1466

1467 **b**, Heatmap showing lineage-specificity of DNA methylation levels at CBSs (motif sequences ±100

1468 bp) in promoter regions of genes that are shown in Fig. 5c. The DNA methylation levels at multiple

1469 CBSs in the same promoter region (TSS ±10 kb) were averaged. The lineage-specificity of DNA

1470 methylation levels shown in the heatmap are calculated by $\log_2(\text{DNA methylation level} / \text{average}$

1471 methylation level of all tissues), and heatmap was sorted by correlation coefficient between CTCF

1472 ChIP-seq signal levels and the DNA methylation levels across the multiple tissues in each tissue

1473 group. Each correlation coefficient is shown in the scatter plots (right), indicating that only a partial of

1474 CTCF bindings had negative correlation with DNA methylation levels ($r < -0.5$, highlighted in blue).

1475

1476 **c**, Histogram showing frequencies of CTCF motif sequences and their PhastCons conservation

1477 scores. Red and Blue bars show CTCF motif sequences at forebrain-specific CTCF occupied gene

1478 promoters and other tissue-specific CTCF occupied gene promoters, respectively. Gray bars show

1479 CTCF motif sequences at random regions.

1480

1481 **d**, Boxplots showing gene length of lineage-specific CTCF occupied promoter genes that had high

1482 correlation coefficient (> 0.6) shown in Fig 5b, c. Forebrain-specific genes and the rest of the other

1483 lineage-specific genes are shown at right and middle, respectively. All genes whose RNA-seq RPKM

1484 value is more than 1 in at least one tissue sample were computed as control (left). (NS not

1485 significant, *** p value < 0.001 , two-tailed t-test)

1486

1487 **e**, Boxplots showing gene length of CTCF-dependent down-regulated genes, CTCF-dependent up-

1488 regulated genes, and CTCF-independent stably-regulated genes in ESCs and NPCs. (NS not

1489 significant, *** p value < 0.001 , two-tailed t-test).

1490

1491 **f**, Volcano plots showing the gene expression changes of the forebrain-specific CTCF-occupied

1492 genes between control cells and CTCF-depleted cells in ESCs (left) and NPCs (right). The larger

1493 number of forebrain-specific genes were down-regulated in NPCs upon CTCF loss.

1494

1495 **g**, Volcano plots showing gene expression changes of the heart-tissue-specific CTCF-occupied

1496 genes between control heart tissue and CTCF knockout heart tissue.

1497

1498 **Supplementary Table 1.**

1499 List of NGS sample information.

1500

1501 **Supplementary Table 2.**

1502 List of CTCF ChIP-seq peaks in control and CTCF-depleted cells.

1503

1504 **Supplementary Table 3.**

1505 List of TAD boundaries and Stripes in control and CTCF-depleted cells.

1506

1507 **Supplementary Table 4.**

1508 Gene expression changes upon CTCF loss in each time point during neural differentiation.

1509

1510 **Supplementary Table 5.**

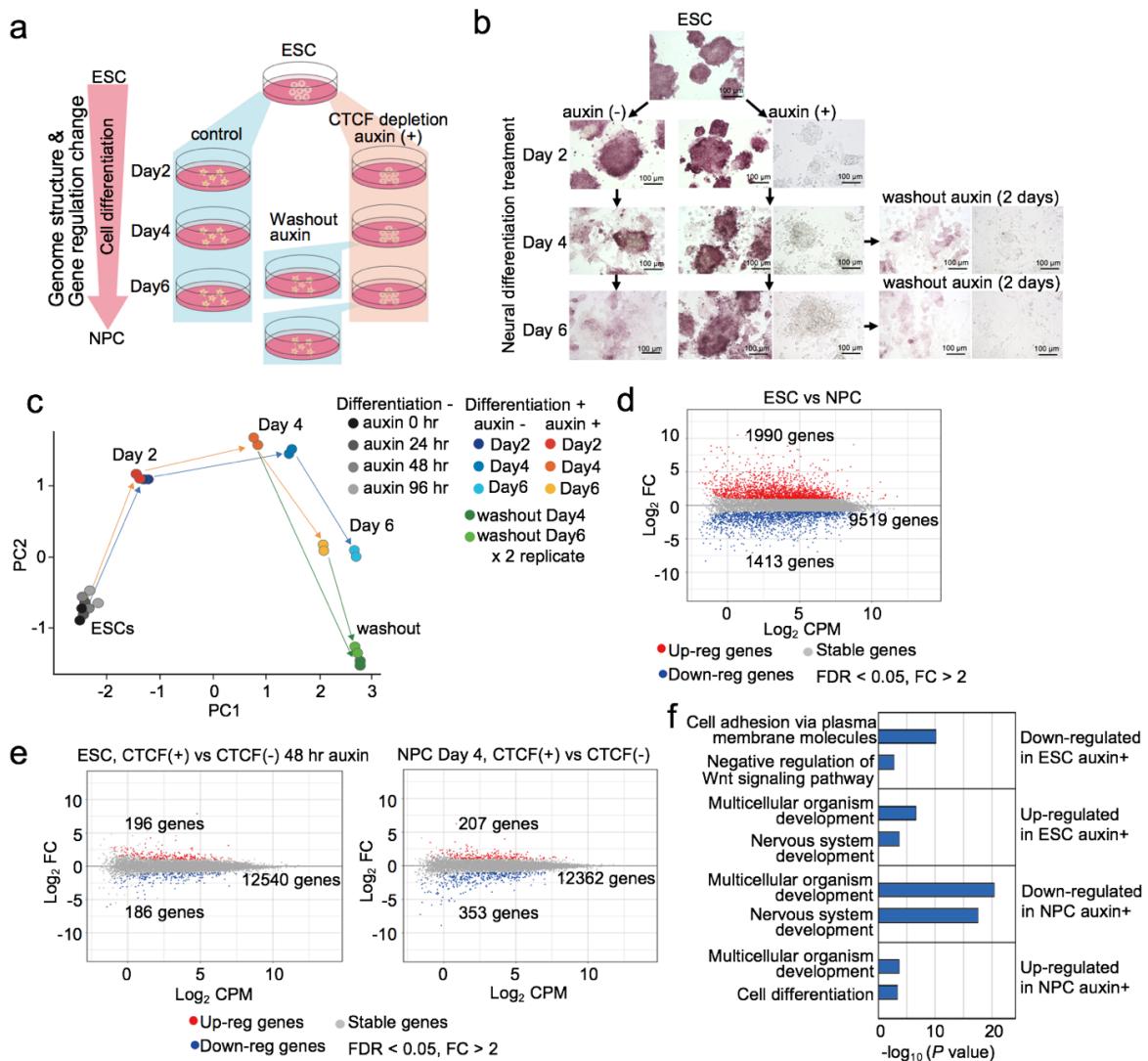
1511 List of differentially changed enhancer-promoter (promoter-promoter) interactions.

1512

1513 **Supplementary Table 6.**

1514 List of ENCODE datasets that were used for Fig. 5 and Supplementary Fig. 12.

Fig. 1



1515

Fig. 2

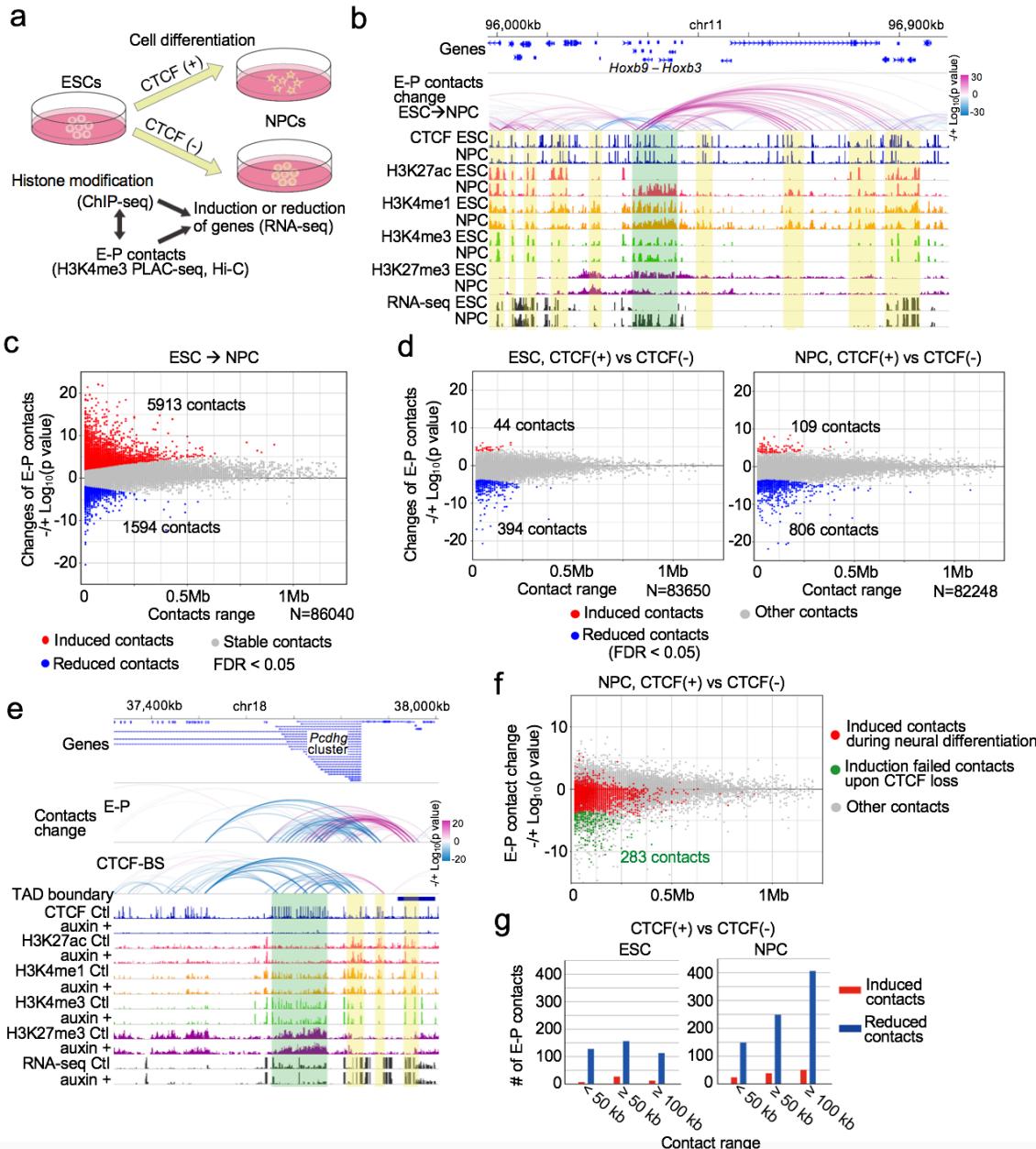


Fig. 3

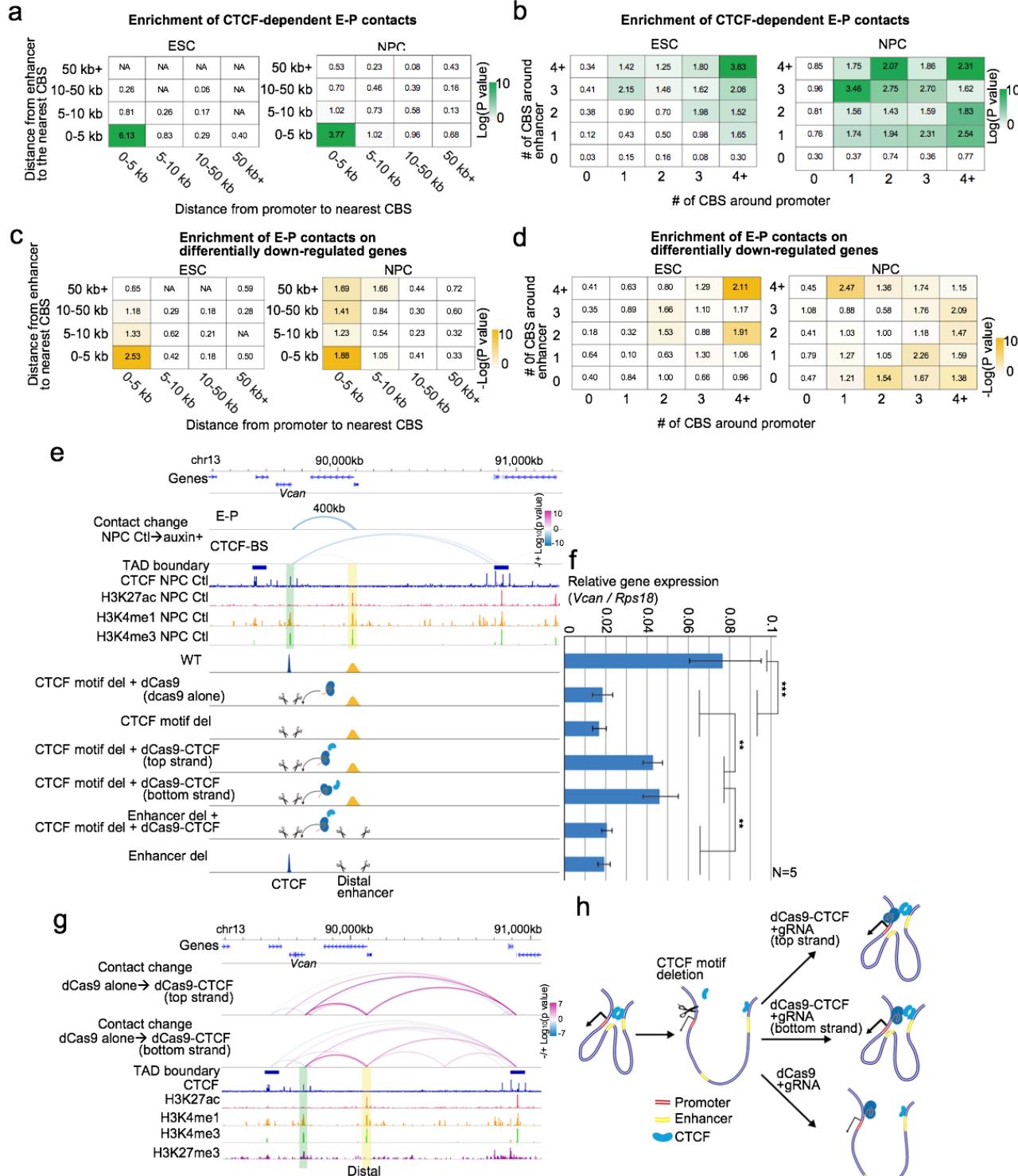
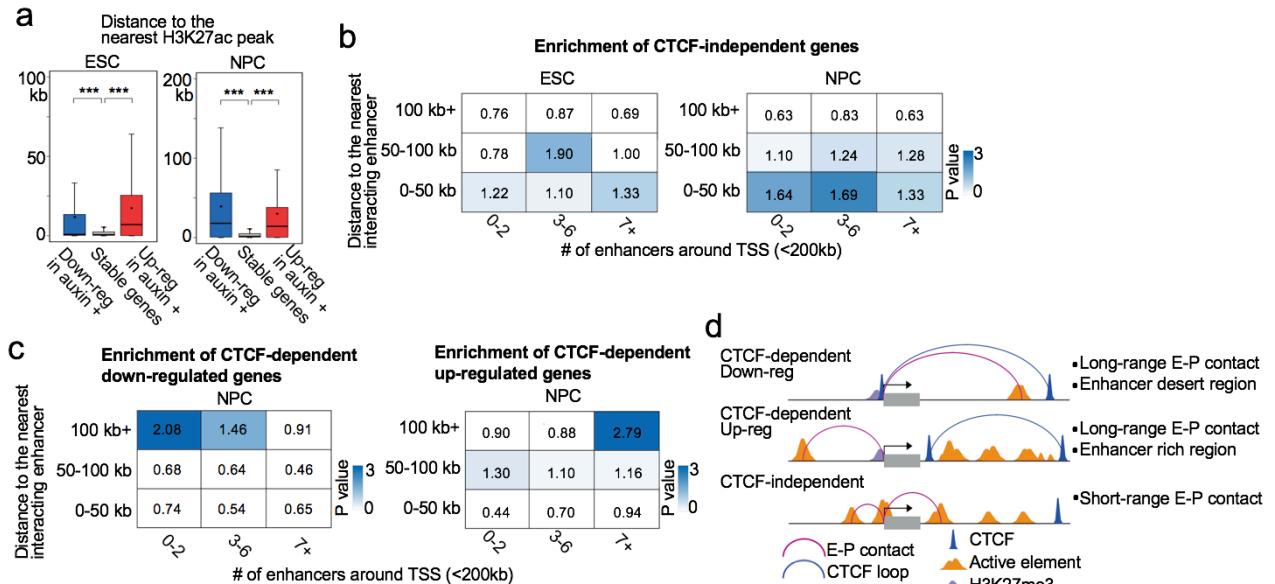
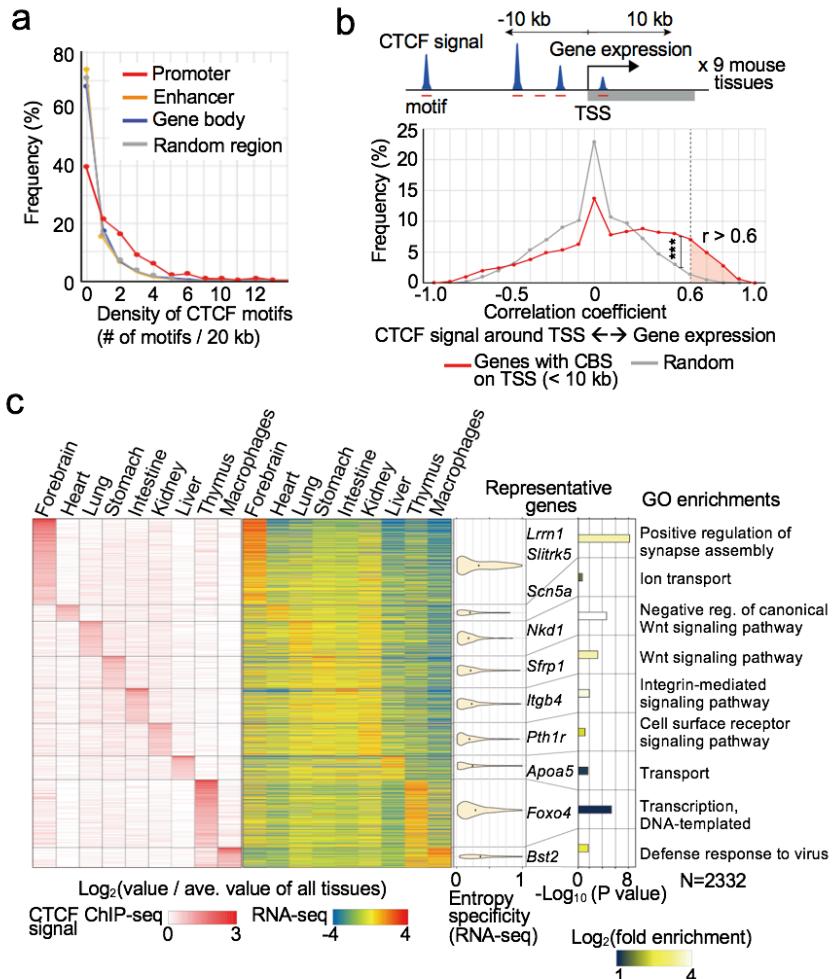


Fig. 4



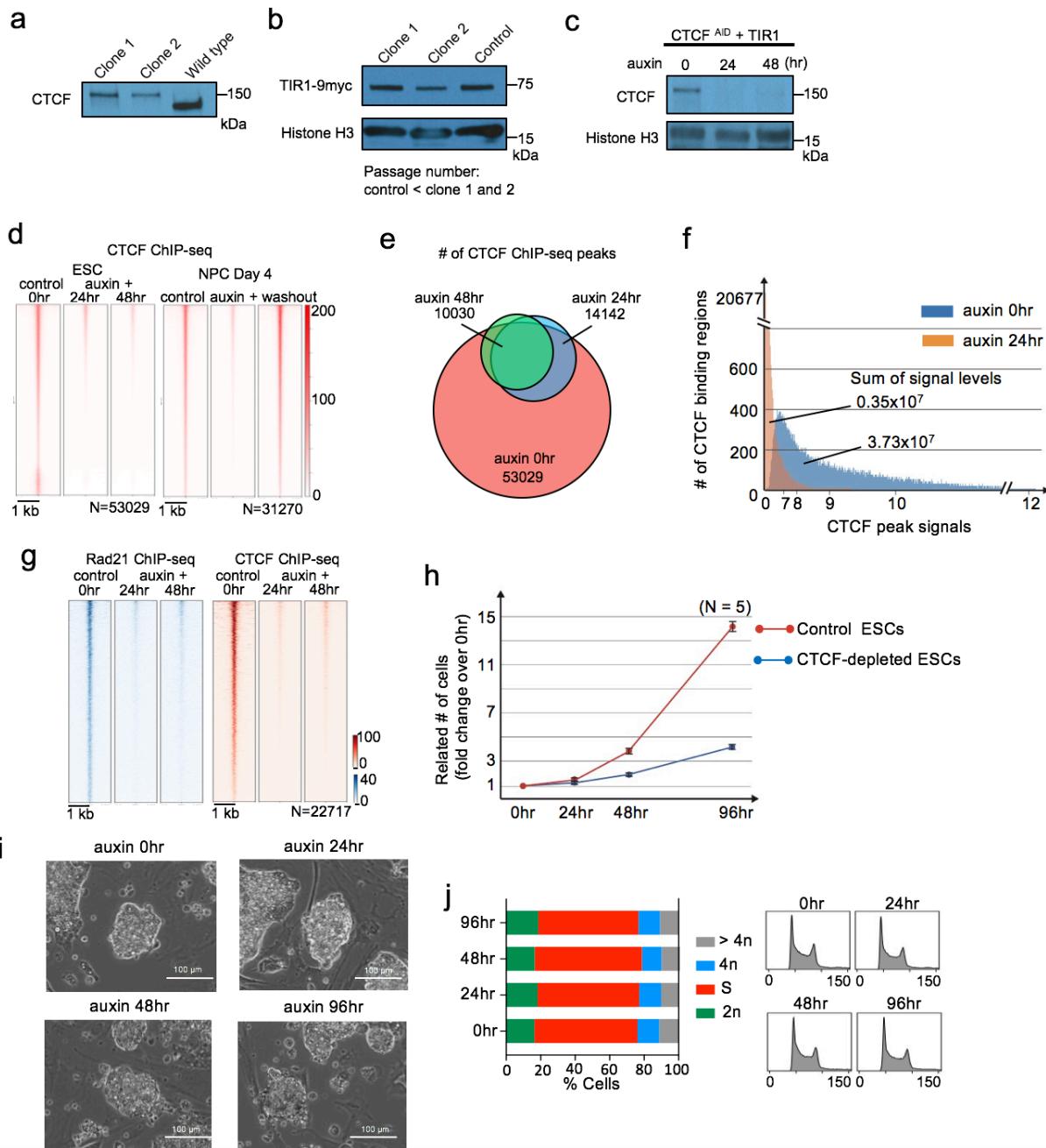
1518

Fig. 5



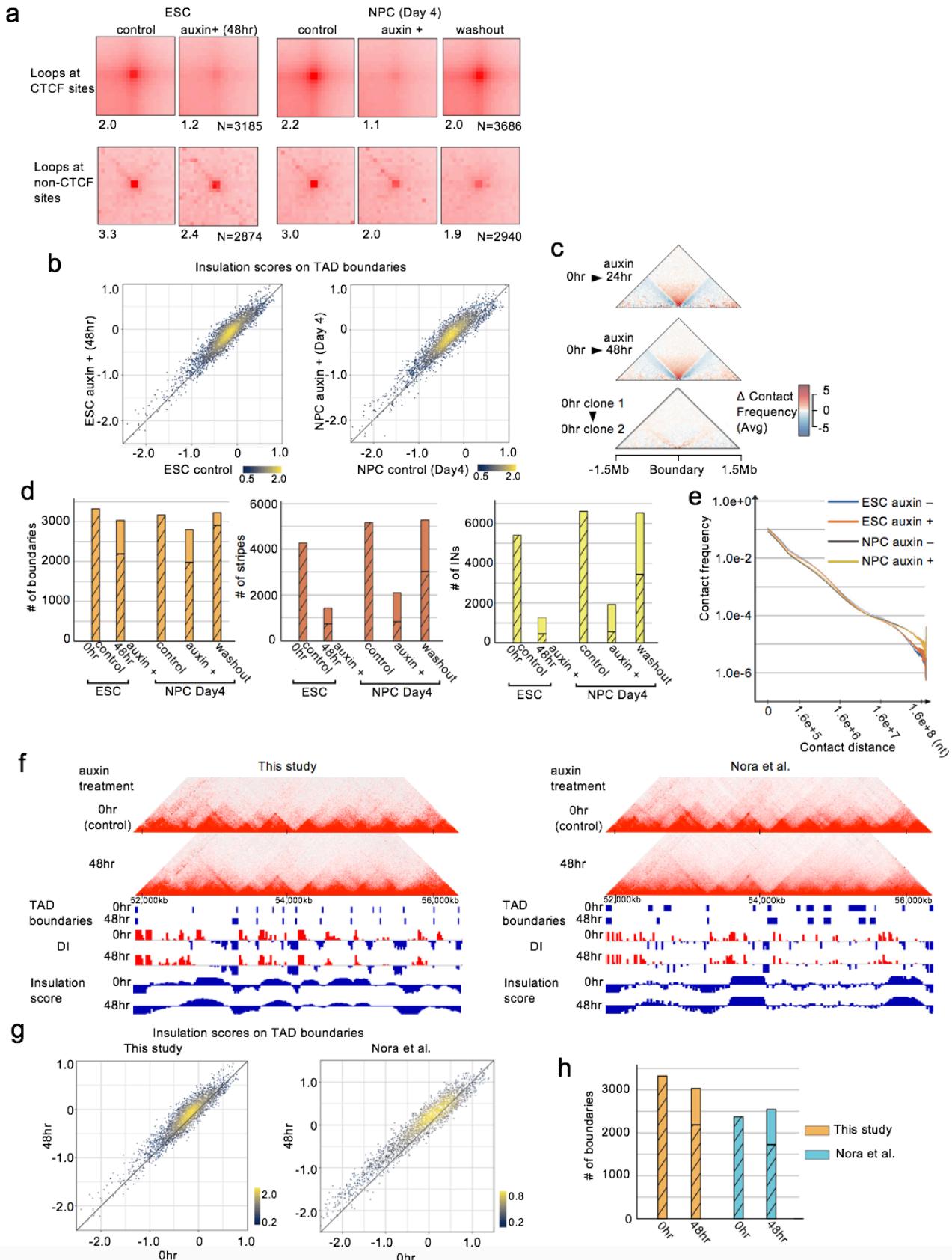
1519

Supplementary Figure 1

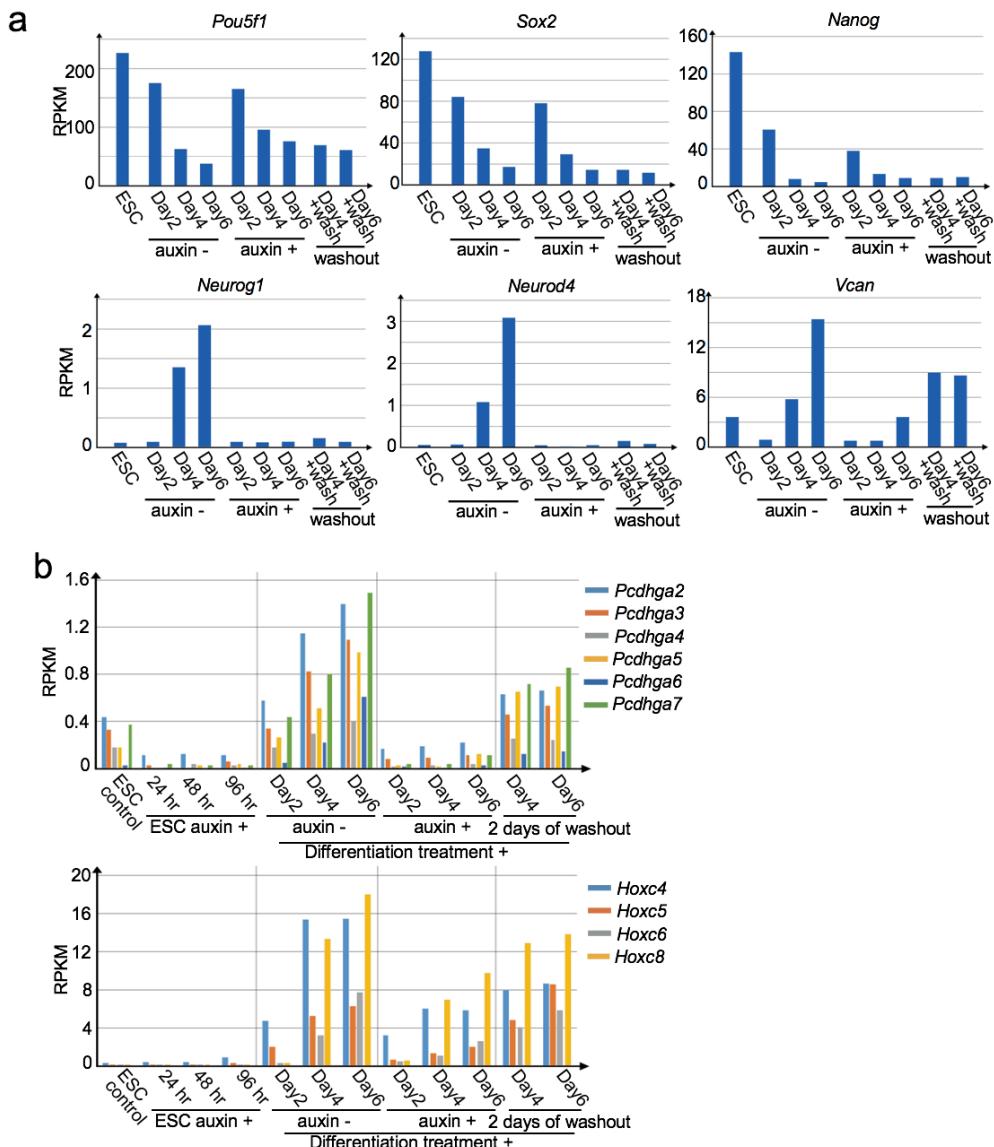


1520

Supplementary Figure 2

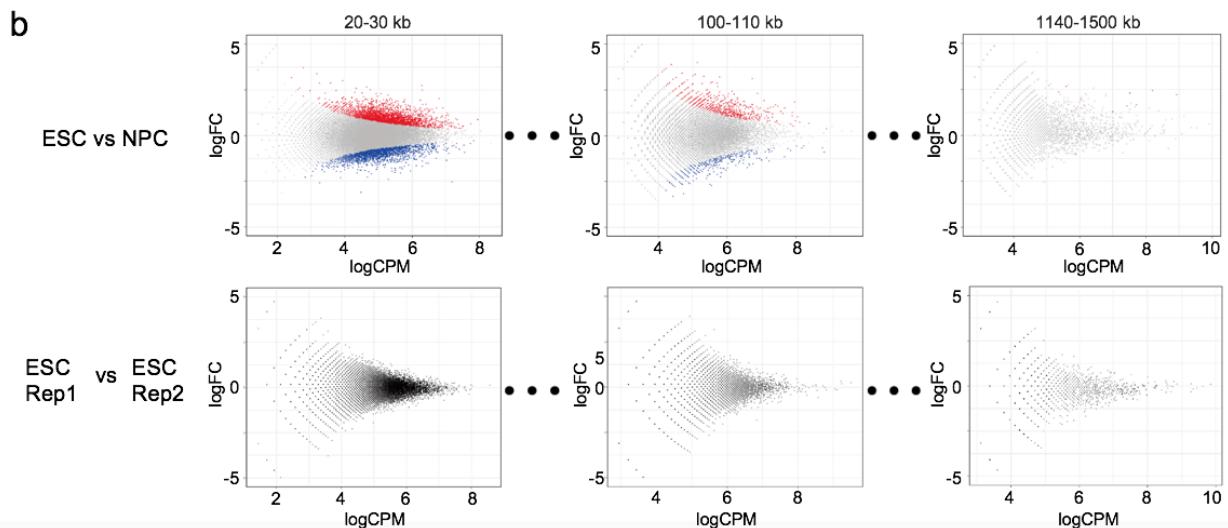
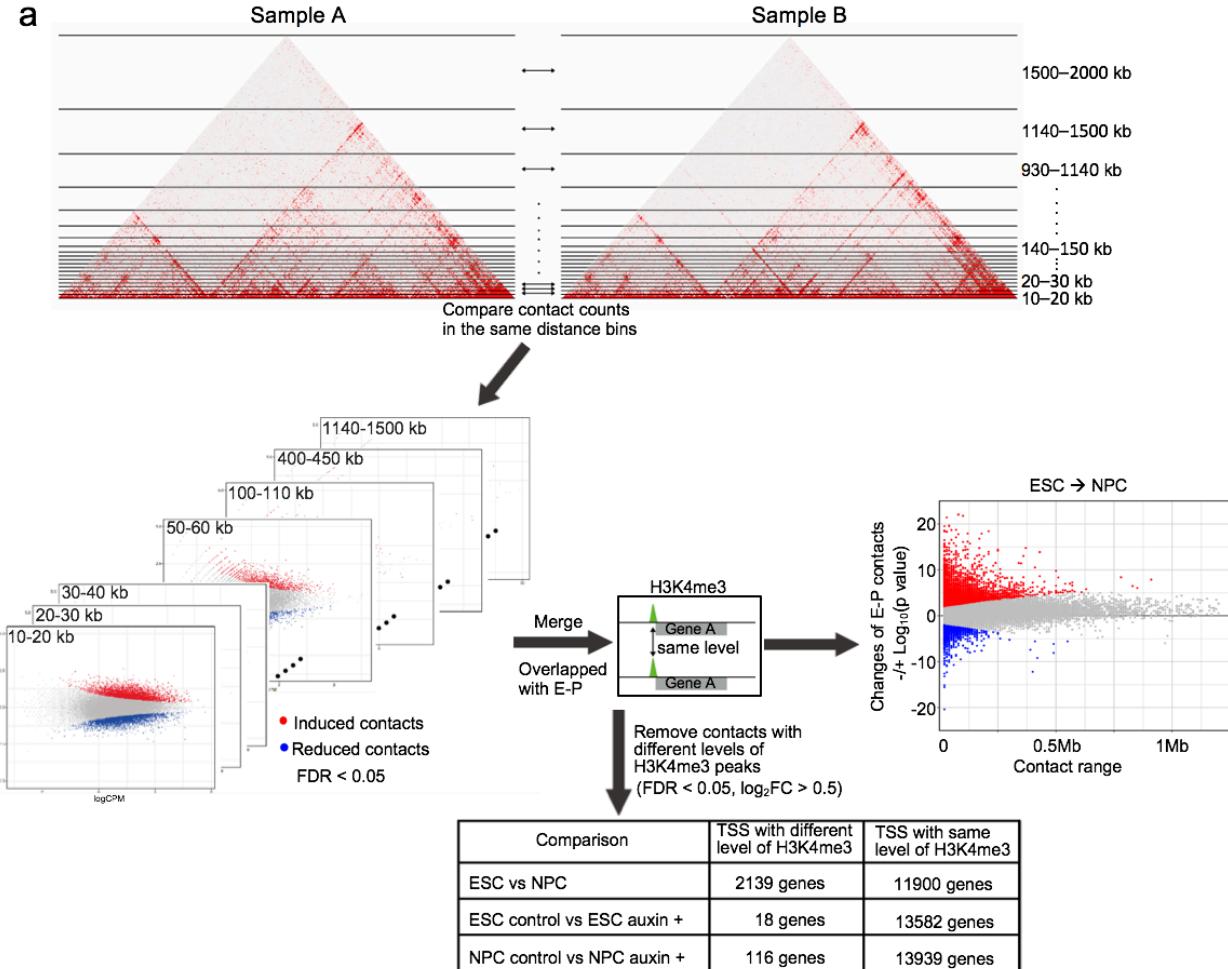


Supplementary Figure 3



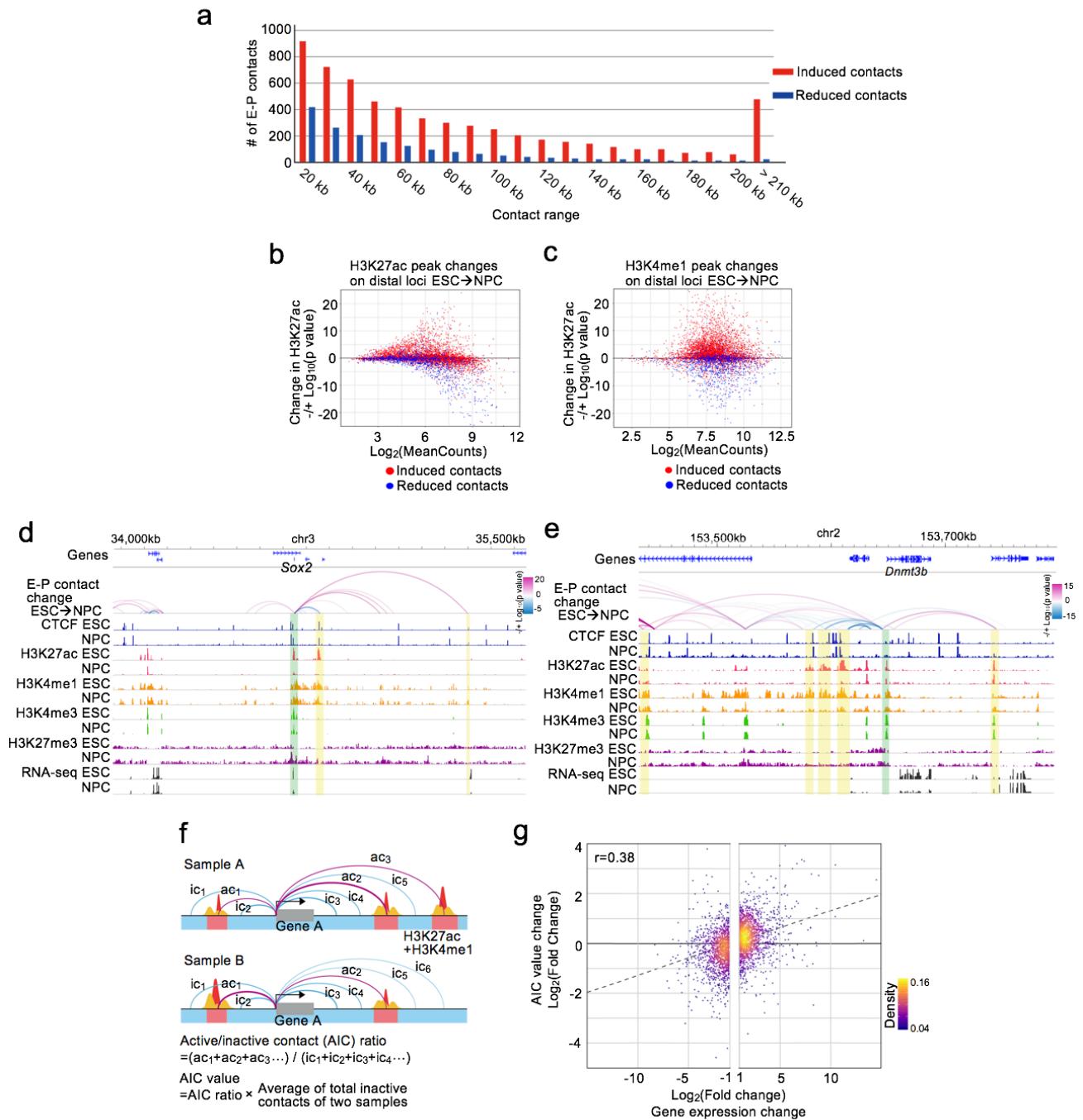
1522

Supplementary Figure 4

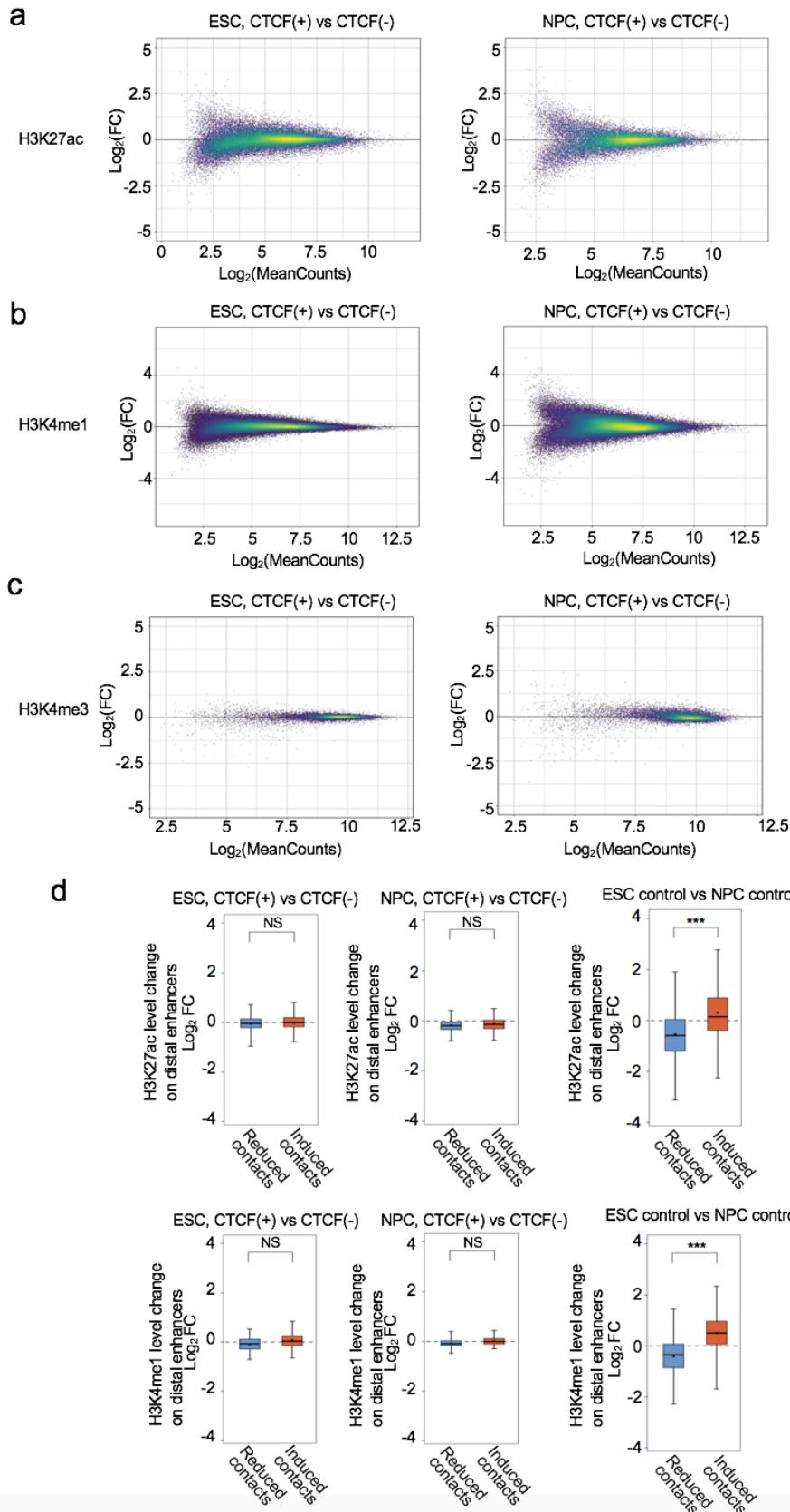


1523

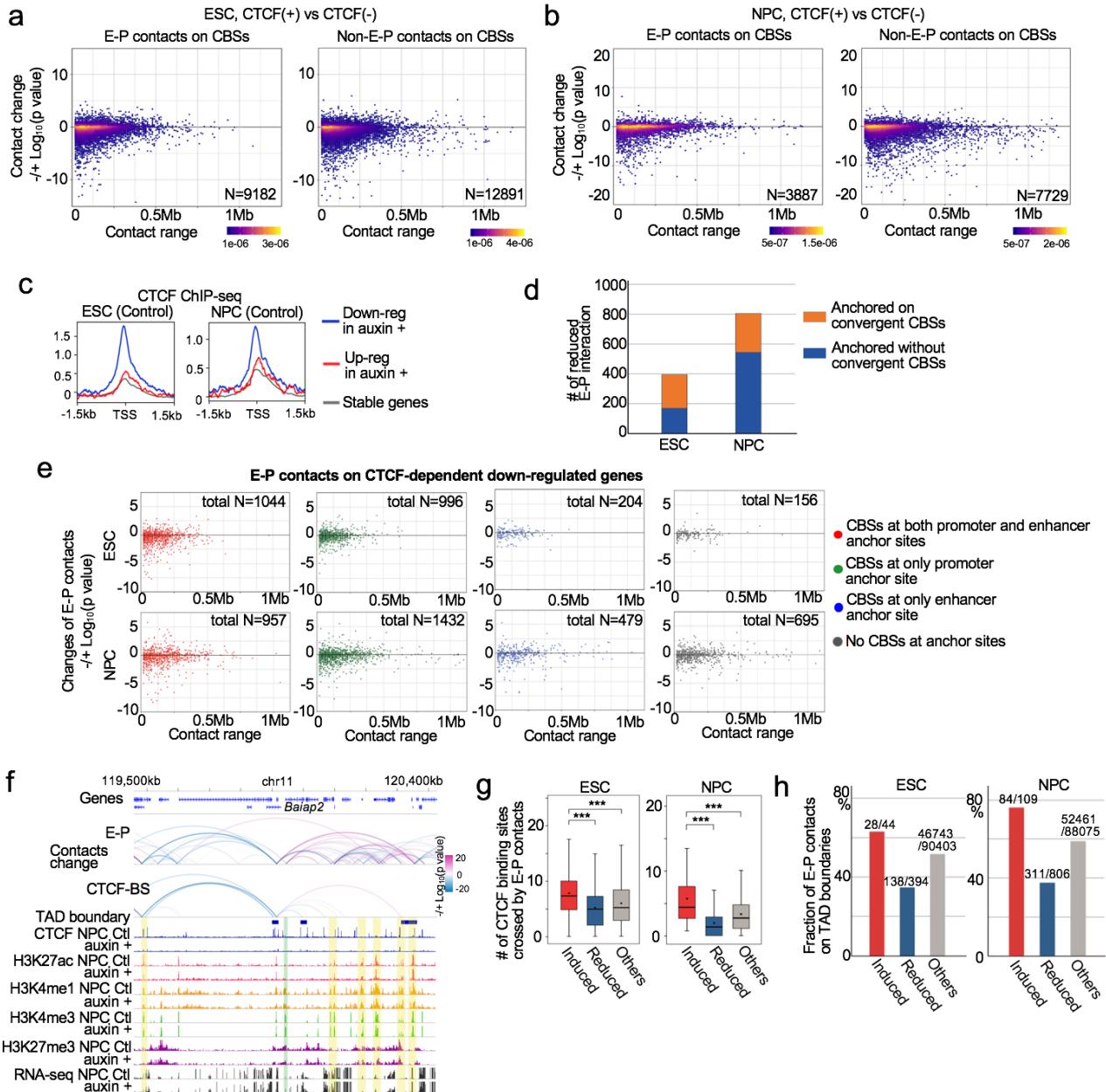
Supplementary Figure 5



Supplementary Figure 6

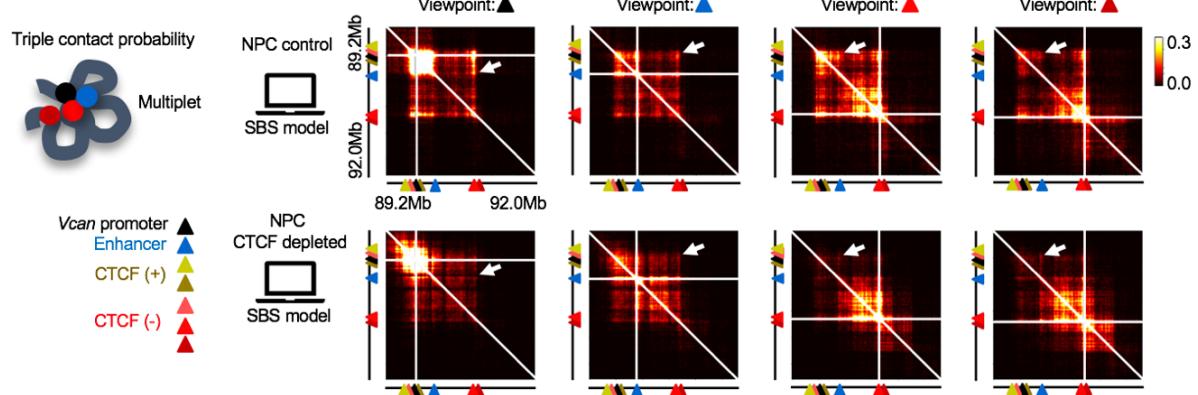


Supplementary Figure 7

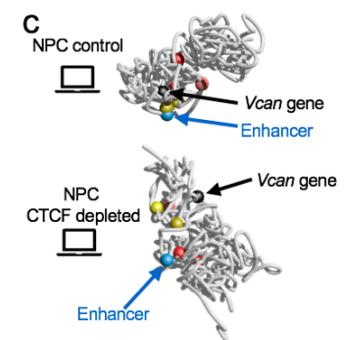
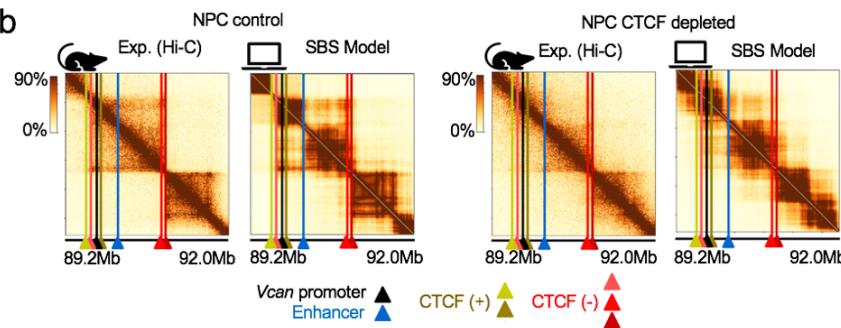


Supplementary Figure 8

a

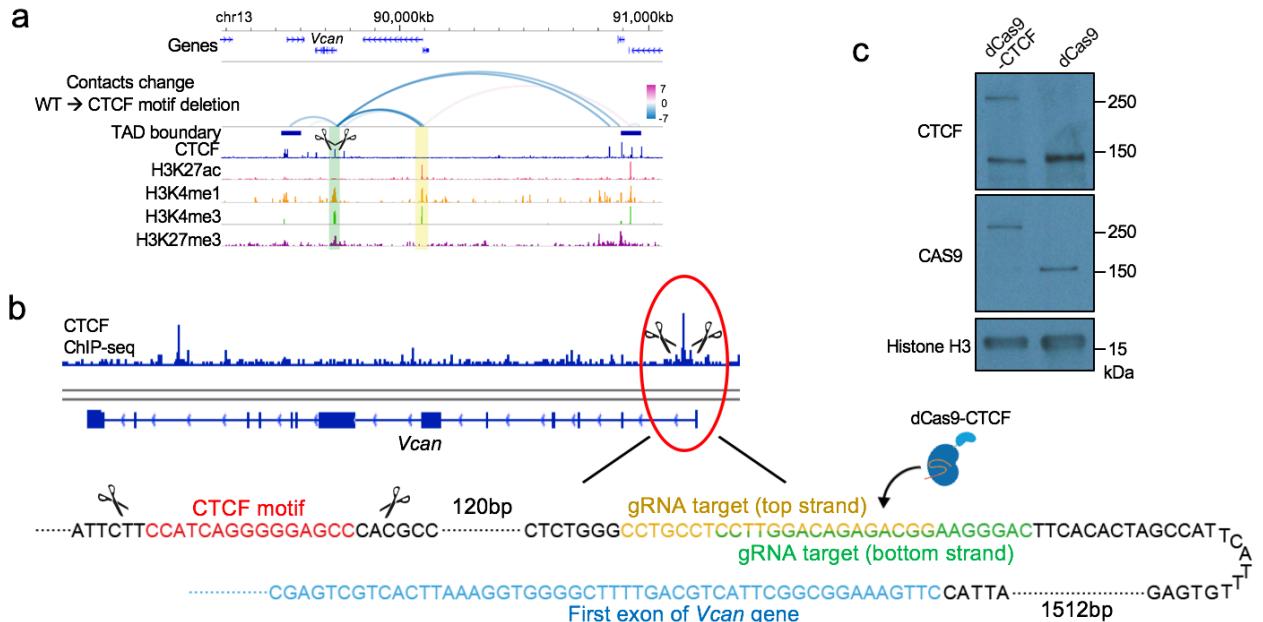


b



1527

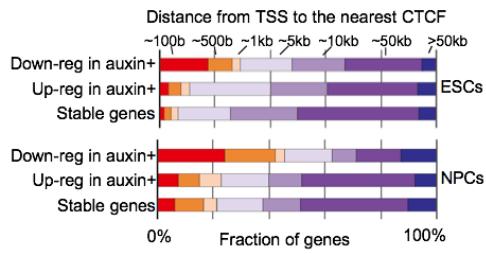
Supplementary Figure 9



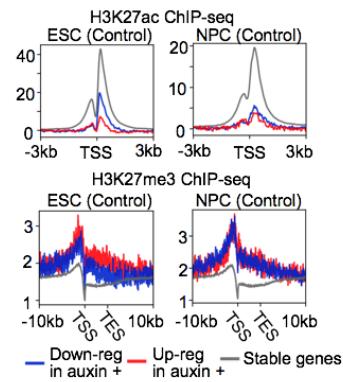
1528

Supplementary Figure 10

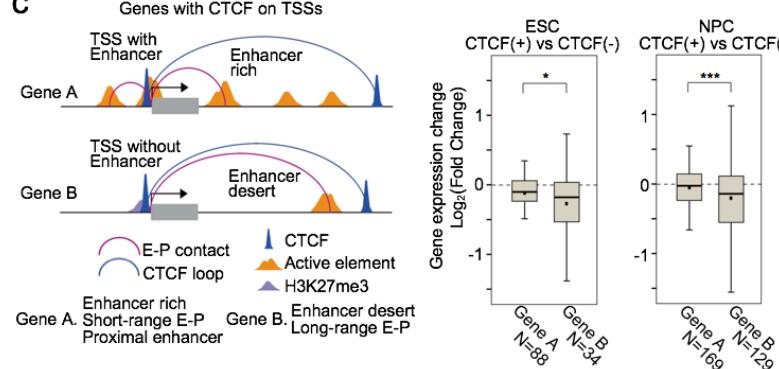
a



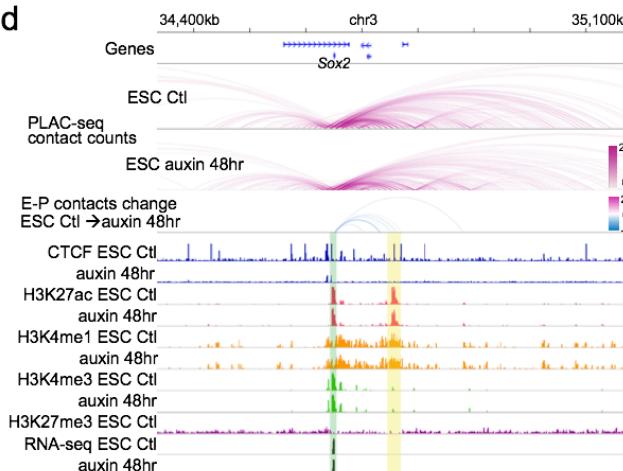
b



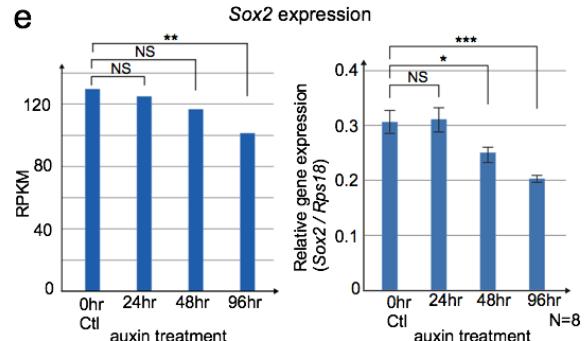
c



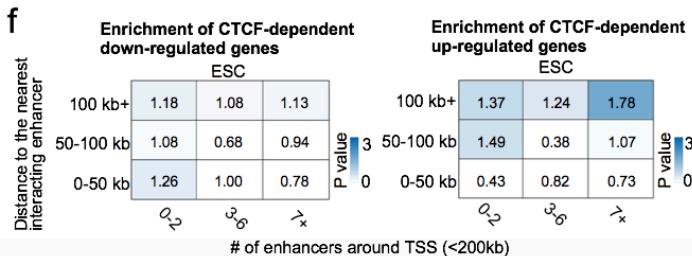
d



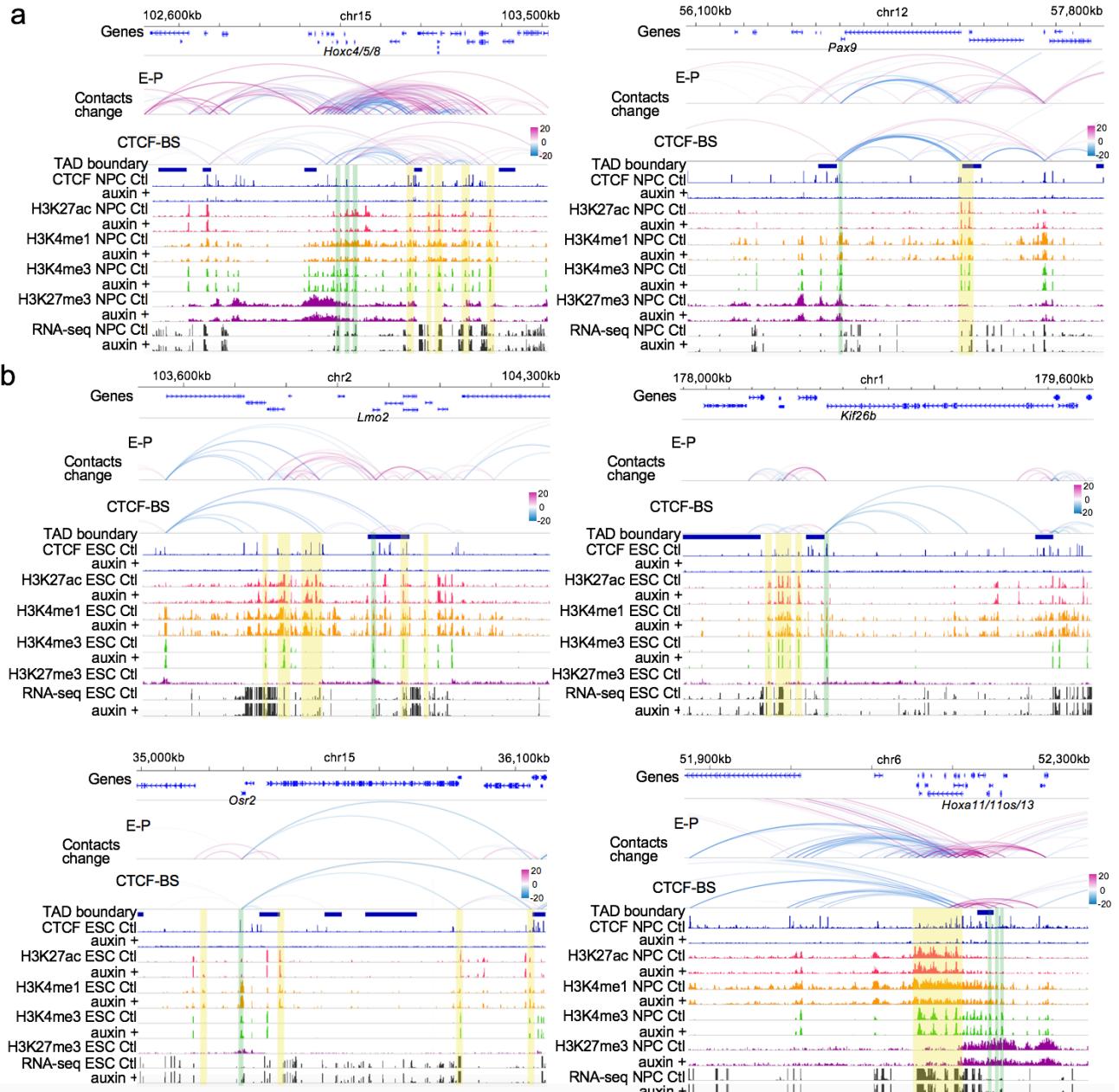
e



f

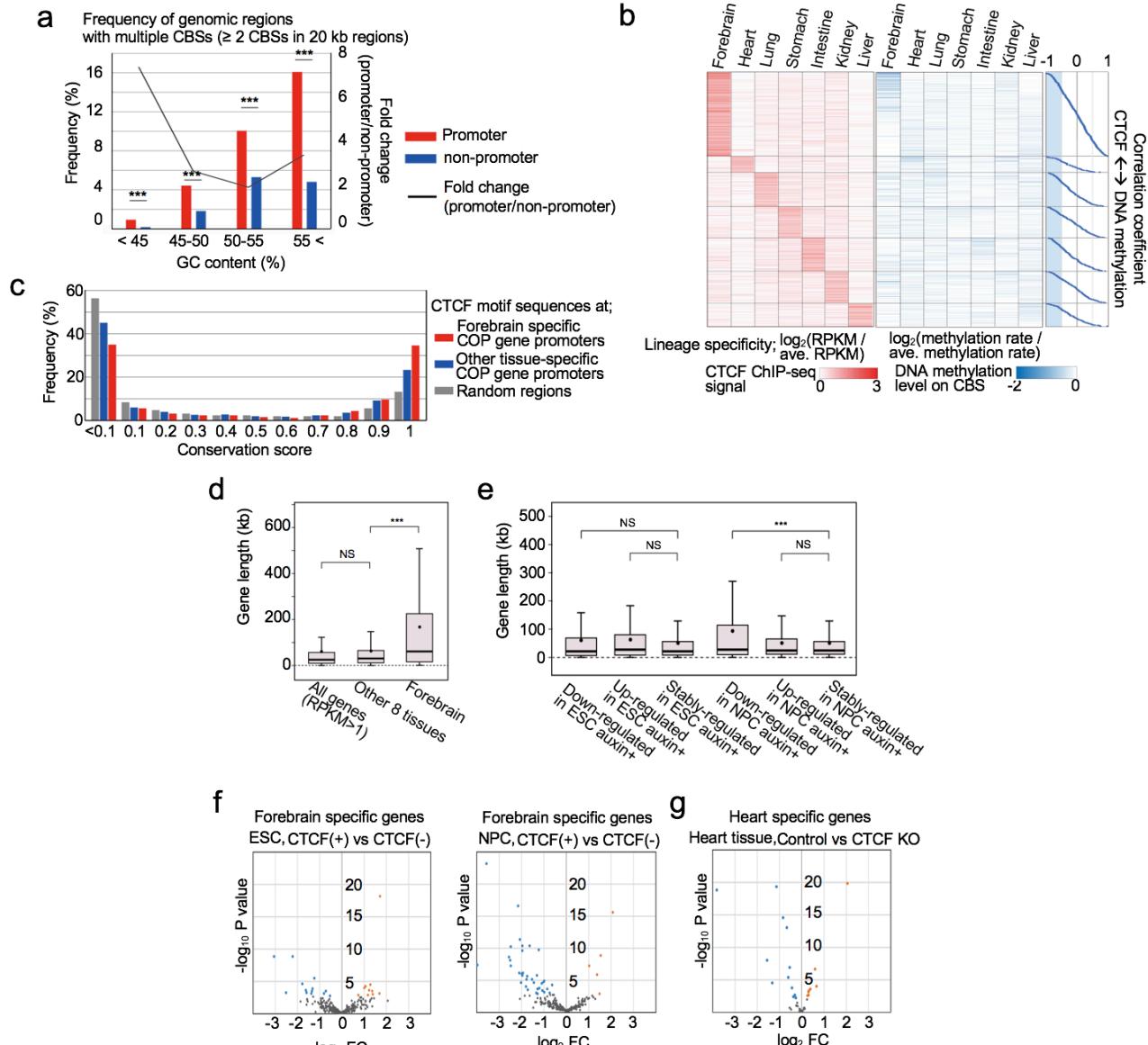


Supplementary Figure 11



1530

Supplementary Figure 12



1531