Cross-platform Hi-C meta-analysis identifies functional insulators that actively block enhancer-promoter interactions

Jian Cui[1,*], Wanying Xu[1,2,*], Xiuyuan Lang[1,2,*], Shanshan Zhang[1,2], Leina Lu[1,2], Xiaoxiao Liu[1], Yan Li[1,#], Fulai Jin[1,3,4,5,#]

[1]Department of Genetics and Genome Sciences, School of Medicine, Case Western Reserve University, Cleveland, OH 44106, USA

[2]The Biomedical Sciences Training Program (BSTP), School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106, USA

[3]Department of Computer and Data Sciences, School of Engineering, Case Western Reserve University, Cleveland, OH 44106, USA

[4]Department of Population and Quantitative Health Sciences, School of Medicine, Case Western Reserve University, Cleveland, Ohio 44106, USA

[5]Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH 44106, USA

[*] Equal contribution. [#] Co-senior authors.

Correspond with: yxl1379@case.edu (Y.L.), fxj45@case.edu (F.J.)

Summary

CTCF forms the boundaries of topologically associating domains (TADs) but their insulating function in transcription is controversial. Here we report that functional insulators (FINs) shall be determined not only by their ability to form static loops or boundaries, but also by the dynamic context if their flanking sequences form new loops upon boundary loss. We performed a meta-analysis of nine independent Hi-C and micro-C datasets upon acute CTCF- or cohesin protein-depletion and found that only at several hundred loci, newly gained enhancers-promoters (E-P) loops can be reproducibly observed upon CTCF depletion. The new E-P loops are cohesin-dependent and associated with recurrent gene activation. These findings allow us to map FINs and define their direct target genes genome-wide. FINs are mostly present in euchromatin, do not overlap with TAD boundaries, and can be perturbed by *Wapl*-depletion or CTCF-displacement. We therefore conclude that FINs, but not TAD boundaries, are *bona fide* insulators.

36 **Introduction**

37 It is generally accepted that TADs confine the communication between genes and enhancers for
38 transcriptional control[1, 2, 3, 4, 5, 6]. Most of the TADs are maintained by CTCF and cohesin, which
39 are also known as insulator proteins. Cohesin is a multimeric complex consisting of *SMC1A*,
40 *SMC3*, *RAD21* and one SA subunit (*STAG1* or *STAG2*). TAD forms when a DNA loop extrudes
41 out of the cohesin complex ring before blocked by CTCF binding sites in convergent orientation,
42 and cohesin serves as an enzymatic engine for this process. Deletion of CTCF or cohesin core
43 subunit RAD21 leads to near complete loss of TADs[7, 8]. Cohesin function is also regulated by
44 cohesin loading complex MAU2/NIPBL[9] and cohesin release factor WAPL[10, 11, 12]: TADs
45 disappear upon loss of NIPBL[13, 14], while loss of WAPL causes bigger loops because cohesin
46 can pass through CTCF blocks due to prolonged loop extrusion[14, 15].

47 Insulators are canonically defined as the DNA elements that inhibit gene expression when put
48 between enhancers and promoters[16, 17, 18]. The discovery of thousands of CTCF-demarcated
49 TADs immediately suggests a model that CTCF sites at TAD boundaries are enhancer-blocking
50 insulators, which is supported by genetic experiments focusing on selected loci[19, 20, 21, 22, 23].
51 However, this model is challenged by a genome-wide Hi-C study in mESCs (Nora *et al.*, **Fig.**
52 **1a**) which only observed mild transcription changes despite near complete loss of TADs upon
53 CTCF-depletion[7]. This conclusion is reproduced in two other genome-wide studies done with *in*
54 *situ* Hi-C and micro-C[24, 25] (**Fig. 1a**). Notably, Kubo *et al*. reported a role of CTCF in mediating E-
55 P interactions[24], and Hsieh *et al.* reported short-range cohesin-independent cross-TAD E-P
56 interactions[25]. Taken together, although all these studies questioned the insulating function of
57 TAD boundaries, they did not address the fundamental question about the whereabouts of the
58 real enhancer-blocking insulators.

59 In this study we demonstrated that “functional insulators” (FINs) can be identified by examining
60 if their flanking regions gain E-P interactions upon CTCF depletion. Challenges exist to map
61 FINs with Hi-C-like genome-wide technologies: (i) most of the public Hi-C datasets do not have
62 billion-scale read depth usually required for loop-level analysis (**Fig. 1a**); (ii) E-P interactions are
63 generally weaker than CTCF loops in Hi-C and therefore harder to detect; (iii) the gain of E-P
64 loop signal can be subtle, requiring robust quantitation of loop strength; (iv) it is difficult to rule
65 out the possibility that some loop changes are secondary effects after prolonged CTCF or
66 cohesin depletion[7, 25]. These challenges can be addressed by comparing independent datasets,
67 but existing CTCF-depletion data are from different Hi-C protocols, which perform very

68 differently in loop-level analyses with conventional pipelines[26]. We previously developed *HiCorr*
69 and *DeepLoop* pipelines to allow robust comparison of loop profiles from orthogonal Hi-C and
70 micro-C platforms, even when sequencing depth is low[27, 28]. Here we use *DeepLoop* to perform
71 a meta-analysis of orthogonal Hi-C datasets to map recurrent DNA rewiring events upon
72 depletion of CTCF and other cohesin proteins (**Fig. 1a**), thus generating a reliable genome-wide
73 map of enhancer-blocking insulators.

74

75 **Results**

76 ***DeepLoop* allows cross-platform comparison of independent CTCF- and cohesin-degron**
77 **Hi-C datasets at loop-resolution**

78 We collected nine pairs of independent Hi-C or micro-C datasets from six publications in which
79 CTCF[7, 24, 25], RAD21[25, 29, 30], or WAPL[25, 29, 31] were acutely depleted in three different mESC lines.
80 The number of contacts in these Hi-C/micro-C datasets are highly variable ranging from 43
81 million to ~1 billion. For every target protein, the choice of degron methods (AID or dTAG) and
82 the timespan of protein degradation (range from 3hr to 3 days) are also variable (**Fig. 1a**).

83 To assess how these variations complicate cross-platform comparisons. We first performed low
84 resolution compartment- and TAD-level analysis on the three CTCF-degron datasets with
85 conventional pipelines. At compartment level (500kb resolution), all datasets display rather
86 consistent compartment correlation heatmaps, and the compartmental organization is largely
87 preserved after CTCF depletion (**Supplementary Fig. 1a-c**). When plotting ICE-normalized
88 contact matrices[32] at 40kb resolution followed by DI (directionality index)-based TAD analyses,
89 all three studies consistently show similar TAD structures and a massive loss of TADs upon
90 CTCF depletion despite a trend that micro-C detected more bins with stronger directionality
91 biases (**Supplementary Fig. 1d-f**). These observations confirmed the findings from the original
92 reports and showed that conventional pipelines are overall robust across different Hi-C
93 platforms at low-resolution.

94 However, Hi-C analysis at high resolution is severely affected by data sparsity and protocol
95 biases[33, 34], thus the robustness of cross-platform comparison is highly dependent on the choice
96 of analytical pipelines. We generated 5kb resolution contact heatmaps (each pixel in a heatmap
97 represents a pair of 5kb bins) using conventional *ICE*[32] (with or without distance correction),

98 *HiCorr*[34], and *DeepLoop*[28], then used scatter plots to compare control cell heatmaps from
99 independent experiments to assess the consistency of different pipelines in analyzing
100 orthogonal Hi-C platforms (**Supplementary Fig. 2**). Here we only compare control cell
101 heatmaps because before protein depletion, the control cells from all studies are expected to be
102 at similar biological states. Protein-depletion cells from independent studies, even if they
103 depleted the same protein, can be still biologically diverse due to many variations such as time-
104 span of protein depletion, incomplete protein depletion, loss of stemness, *etc.* We found that
105 although conventional ICE pipelines show poor cross-platform consistency with R-square values
106 often close to zero, *HiCorr* and *DeepLoop* clearly improved the consistency with much higher
107 cross-platform R-square values (**Supplementary Fig. 2**, summarized in **Fig. 1b**).

108 We also use scatter plots to compare the heatmaps before and after protein depletion at pixel-
109 by-pixel level and interestingly, found that only HiCorr and DeepLoop heatmaps show
110 recognizable populations of loop pixels that retain their strength upon CTCF- or Rad21-
111 depletion (**Supplementary Fig. 2**), suggesting their robustness in revealing the unchanged loop
112 pixels that are masked by noises in the other pipelines. To further evaluate the cross-platform
113 consistency when applying different pipelines to analyze dynamic loops, we generated
114 differential heatmaps by subtracting each pair of 5kb-resolution contact heatmaps (depletion –
115 control), then computed the correlations between the differential heatmaps from all nine pairs of
116 degron experiments (**Fig. 1c**, **Supplementary Fig. 3**). Only the matrices from *HiCorr* and
117 *DeepLoop* clustered CTCF- and Rad21-degron datasets together and *Wapl*-degron datasets
118 are separated out (**Fig. 1c**, **Supplementary Fig. 3**). This is consistent with the knowledge that
119 depletion of *CTCF* or Rad21, but not Wapl, causes massive loss of TADs and CTCF loops.
120 Taking together, these results support the robustness of *DeepLoop* in cross-platform analysis
121 regardless of read depth[28].

122 **Hi-C meta-analysis reveals recurrent E-P loop-gaining events upon CTCF depletion**

123 We first focused on the CTCF-degron Hi-C datasets. In the example of **Fig. 1d** (top two rows),
124 with conventional ICE-normalization, we observed dampened TAD structure at 20kb resolution
125 as the original study reported[7], but at 5kb resolution, loop signal from ICE-normalized heatmaps
126 are weak and often invisible. From the *DeepLoop* enhanced heatmaps, we can consistently
127 observe multiple CTCF loops that appear in all control cells heatmaps but are substantially
128 weakened upon CTCF depletion. A notable change in this region is the gain of long-range loops

129 (**Fig. 1d**, second row, loop pixels in dashed circle) connecting *Notch1* gene to a distal enhancer
130 region with multiple H3K27ac peaks (**Fig. 1e**). Consistent with gain of E-P loop, *Notch1* is
131 upregulated upon CTCF-depletion (**Fig. 1e**). The gain of loop signal at this locus is also
132 observable in *Wapl*-depleted cells (**Fig. 1d**, bottom two rows) and two other CTCF-degron
133 studies (**Supp. Fig. 4**), but not in the *Rad21*-depleted cells (**Fig. 1d**, 3rd and 4th rows, **Supp. Fig.**
134 **4g**). These observations suggest that the two CTCF loop anchors at the TAD boundary region
135 (**Fig. 1d**, highlighted in pink) are blocking the *Notch1*-enhancer interaction. This example also
136 demonstrated that *DeepLoop* may detect recurrent insulating events from CTCF-degron Hi-C
137 data without the need for ultra-high sequencing depth.

138 To perform a genome-wide analysis of insulating events, we call 300K~500K loop pixels from
139 each CTCF-degron dataset and use the pixel intensity in *DeepLoop* heatmaps to measure the
140 loop strength (**Fig. 2a, Supp. Fig. 5**). Notably, *DeepLoop* outputs are enhanced from ratio
141 (obs/exp) heatmaps instead of normalized contact frequency used by ICE/KR. We previously
142 showed that this approach promises cross-platform convergence between orthogonal Hi-C and
143 micro-C data[28].

144 In all CTCF-degron studies, the scatter plots are clearly down-skewed reflecting a massive loss
145 of chromatin loops upon CTCF deletion (**Fig. 2a, Supp. Fig. 5a, d**, left panels). Using a simple
146 fold-change cutoff, we called loop pixels that lose strength ("lost"), retain strength ("retained"), or
147 gain strength ("gained") from each dataset (**Fig. 2a, Supp. Fig. 5a, d**, middle panels). We
148 further classified the loop pixels based on their epigenetic marks (**Methods**, examples of loop
149 classification in **Supp. Fig. 6**). "CTCF-only" loop pixels, defined as those with at least one
150 anchor overlapping a CTCF peak, were predominantly found among the "lost" pixels. In
151 comparison, "retained" and "gained" loop pixels are enriched with "CTCF-EP" or "EP-only" loops
152 characterized by promoter (H3K4me3) or enhancer (H3K27ac) marks at one or both anchors
153 (**Fig. 2b, Supp Fig. 5b, e**). Interestingly, in all three studies the "gained" loops are much larger
154 than the "retained" loops (**Fig. 2a, Supp. Fig. 5a, d**, right panels); we hypothesized that this is
155 because the "gained" E-P loops form through cohesin loop extrusion (more discussions below).

156 We also examine the robustness of the "lost", "retained", and "gained" pixels across
157 independent Hi-C/micro-C studies. As expected, the "lost" pixels from any study generally also
158 lose strength in the other two studies (**Fig. 2c, Supp Fig. 5c,f**, first row). However, the
159 consistency of "retained" loop pixels is study specific: (i) although the "retained" loop pixels from
160 Nora *et al.* are mostly also "retained" in Kubo *et al.*, some of them lose strength upon CTCF

161 degradation in Hsieh *et al.* (**Fig. 2c**, second row); (ii) most of the "retained" loop pixels from
162 Kubo *et al*. lose strength in the other two studies (**Supp. Fig. 5c**, second row); (iii) the "retained"
163 loops from Hsieh *et al.* remain "retained" in both Nora *et al.* and Kubo *et al.* (**Supp. Fig. 5f**,
164 second row). Since incomplete CTCF depletion causes the leftover CTCF loops[7], these
165 observations suggest that CTCF depletion in Kubo *et al.* may be less complete than the other
166 two studies. On the other hand, Hsieh *et al.* appear to have achieved better CTCF depletion,
167 perhaps due to its use of an N-terminal degron system (**Fig. 1a**), although platform-specific
168 biases, differences in assay sensitivity and sequencing depth may also explain some of the
169 discrepancies between studies.

170 Interestingly, the "gained" pixels from any one of the three studies generally also gain strength
171 in the other two studies (**Fig. 2c, Supp Fig. 5c,f,** third row). The only exception is that some
172 "gained" loop pixels from Kubo *et al*. lose strength in Hsieh *et al*. (**Supp Fig. 5c**). Again, this is
173 likely because Kubo *et al.* retains more CTCF loops due to incomplete CTCF degradation.
174 Nevertheless, most of the "gained" pixels from Nora *et al*. and Hsieh *et al*. do gain strength in
175 Kubo *et al.* (**Fig. 2c, Supp Fig. 5f**), supporting the consistency between the three studies.
176 Notably, E-P loop gaining events are largely neglected in the original Hi-C/micro-C studies
177 because the changes are often too subtle to distinguish from background with conventional
178 analytical tool (*e.g.*, **Fig. 1d**). Our meta-analyses prove that although there are only a small
179 number of newly gained E-P loops, they are reliable consequences of CTCF-depletion; affected
180 genes such as *Notch1* in **Fig. 1** are putative target genes of CTCF-mediated insulation.

181 **Hi-C Meta-analyses of *Rad21*- and *Wapl*-depletion data distinguish consensus loop**
182 **regulatory mechanisms from system variations**

183 We also performed meta-analyses of the *Rad21*- and *Wapl*-depletion Hi-C datasets. Ablation of
184 *Rad21* also causes massive loss of CTCF loops, while the "retained" and "gained" loops are
185 also enriched with E-P loops (**Supp. Fig. 7a-i**). However, there are important differences
186 between *Rad21*- and *CTCF*-depleted cells. (i) In *Rad21*-depleted cells, the "retained" and
187 "gained" loop pixels (except the "gained" loop pixels in Rhodes *et al.*) are much smaller (mostly
188 within 100kb) than those in *CTCF*-depleted cells, indicating that cohesin-independent loops are
189 short-range (**Supp. Fig. 7a, d, g**, right panels). One alternative explanation is that the greatly
190 reduced pool of cohesin only allows rare and short-lived extrusion events, resulting in smaller
191 loops. (ii) Unlike in CTCF-depleted cells, the "gained" loop pixels are NOT recurrent between

192 the three *Rad21*-degon datasets (**Supp. Fig. 7c, f, i**, last row). (iii) The "retained" and "gained"
193 loop pixels in *Rad21*-depleted cells (except the "gained" loop pixels in Rhodes *et al.*) are more
194 enriched with H3K36me3 mark, suggesting transcription associated chromatin interactions
195 (**Supp. Fig. 7b, e, h**).

196 The "gained" loops in *Rad21*-depleted cells from Rhodes *et al*. are enriched with H3K27me3
197 and tends to be larger in size (**Supp. Fig. 7a, b**, right panels). These observations are
198 consistent with the conclusion from the original study that cohesin removal enhances long-range
199 interactions between a subset of polycomb-occupied regions[30]. However, these features (the
200 larger loop size and the enrichment with H3K27me3) are specific to "gained" loops in Rhodes *et*
201 *al.* but not observed in the other two *Rad21*-degron studies (**Supp. Fig. 7d, e, g, h**, right
202 panels). We manually examined the "gained" H3K27me3 loops in Rhodes *et al.* and found that
203 in almost all cases, the long-range H3K27me3 loops are already present in the control cells
204 from Kriz *et al.* and Hsieh *et al.* before *Rad21* depletion (two examples are shown in **Supp. Fig.**
205 **8**). Therefore, at least in the control mESCs from Kriz *et al.* and Hsieh *et al.*, cohesin does not
206 prevent the formation of long-range H3K27me3 loops.

207 In contrast, ablation of cohesin unloading factor *Wapl* generates up-skewed scatterplots in all
208 three studies (**Supp. Fig. 7j, m, p**, left panels). The skewness of Hsieh *et al.* is less apparent,
209 probably because Hsieh *et al.* performed *Wapl* depletion for much shorter time (3hrs) than Liu *et*
210 *al*. (24hrs) and Kriz *et al*. (72hrs) (**Fig. 1a**). In all *Wapl*-degron studies, the "lost", "retained", and
211 "gained" loops are almost equally enriched with CTCF and E-P loops (**Supp. Fig. 7k, n, q**).
212 Consistent with the model that *Wapl*-depletion stabilizes chromatin-bound cohesin and promote
213 loop extrusion[14, 15], the "gained" loops are significantly bigger than "lost" and "retained" loops in
214 all three studies (**Supp. Fig. 7j, m, p**, right panels). The "lost", "retained", and "gained" loop
215 pixels are generally consistent between studies (**Supp. Fig. 7l, o, r**) despite some discrepancy
216 for "lost" loop pixels from Hsieh *et al.* (**Supp. Fig. 7r**), again likely because Hsieh *et al.* only
217 depleted *Wapl* for a short period of time.

218 **Newly gained E-P loops in CTCF-depleted cells are cohesin dependent.**

219 We next take the "lost", "retained", and "gained" loop pixels from *CTCF*-degron studies and
220 examine how they change in *Rad21*- and *Wapl*-degron data. Firstly, nearly all "lost" loops in
221 *CTCF*-depleted cells also disappear in *Rad21*-depleted cells (**Fig. 2d**, first row, **Supp Fig. 9a,b**,
222 first row), suggesting that most of the CTCF loops are cohesin-dependence. Secondly, a subset

223 of the “retained” loops in CTCF-depleted cells are also “retained” upon *Rad21*-depletion (**Fig.**
224 **2d**, second row, **Supp Fig. 9a, b**, second row), suggesting cohesin-independence. The other
225 “retained” loops in CTCF-depleted cells lose strength upon *Rad21*-depletion likely because they
226 are: (i) leftover CTCF loops due to incomplete CTCF-depletion, or (ii) non-CTCF loops formed
227 via cohesin-dependent mechanisms. Interestingly, most of the “retained” loops in CTCF-
228 depleted cells from Kubo *et al.* are lost upon *Rad21*-depletion (**Supp Fig. 9a**, second row), but
229 most of the “retained” loops from Hsieh *et al.* are also retained upon *Rad21*-depletion (**Supp**
230 **Fig. 9b**, second row), again suggesting that Hsieh *et al.* may have achieved more complete
231 CTCF-depletion than Kubo *et al.*

232 Finally, the “gained” loops in CTCF-degron studies, which are weak in both CTCF-degron and
233 Rad21-degron control mESCs, do not gain strength upon *Rad21*-depletion cells (**Fig. 2d, Supp**
234 **Fig. 9a, b**, third row), suggesting that the formation of these “gained” loops requires cohesin. In
235 contrast, we observed that these “gained” loops in CTCF-depleted cells showed an apparent
236 upward bias in the scatter plots of *Wapl*-degron studies (**Fig. 2d, Supp Fig. 9a, b**, third row).
237 This observation is consistent with the example in **Fig. 1d** and further supports a role of cohesin
238 in forming those “gained” loops. *Wapl*-depletion may perturb the function of insulators by
239 releasing short-range CTCF loops and allow insulated E-P interactions to skip some insulators.

240 **Define, map, and characterize functional insulators (FINs) in mESC genome**

241 Among the total 8,141 “gained” loop pixels from all three CTCF-degron Hi-C/micro-C datasets,
242 4,944 pixels are recurrent in at least two studies with a median size ~250kb (**Fig. 3a**). In each
243 dataset, the “gained” loop pixels can be merged into ~300 “loop-gaining loci” (**Methods**) and
244 there are substantial overlaps between independent studies: 174 loop-gaining loci are shared by
245 all three studies and 304 are shared by at least two studies (**Fig. 3b**). We hypothesize that
246 these recurrent E-P loop-gaining loci are caused by loss of CTCF insulation; indeed, 4,665 out
247 of the 4,944 loop-gaining pixels (94%) are insulated by at least one CTCF loop in wildtype cells
248 (**Fig. 3c**). We further curated the reproducible loop-gaining loci based on the types and numbers
249 of CREs involved (**Fig. 3d**). Nearly all the curated loci (283 of 299) gain E-P loops (2 loci gain
250 CTCF contacts and the others cannot be clearly classified). We further classified the E-P loop-
251 gaining loci into three categories (**Fig. 3d**). In *Type I*, one of the two E/P sites is insulated in a
252 CTCF loop while the other site is not (exampled in **Supp. Fig. 10a**); *Type II* involves at least two
253 CTCF loops insulating two E/P-sites (exampled in **Fig. 1e, Supp. Fig. 10b**). In contrast, the

254 gained E-P loops in *Type III* do not have insulating CTCF loops (exampled in **Supp. Fig. 10c**).
255 *Type I* and *II* represents most of the E-P loop gaining loci (263 out of 283), while only 20 loci
256 belong to *Type III*, again suggesting that most of the loop gaining loci are due to the removal of
257 CTCF insulation.

258 We therefore propose to define "functional insulators" (FINs) as the CTCF binding sites within
259 the "lost" loop anchors that when bound by CTCF, can actively block the interactions between
260 flanking *cis*-regulatory elements (CREs) (model *iii* in **Fig. 3e, Methods**). Therefore, FINs can be
261 found from "loop-gaining loci" and in fact, most "loop-gaining loci" contain at least one FIN, and
262 we therefore also call them "FIN loci" (**Methods**). The definition of FIN is based on the gain-of-
263 loop upon loss of insulator function, which is conceptually analogous to the conventional model
264 to define insulators based on the gain-of-expression upon loss of insulator function (model *i* in
265 **Fig. 3e**) but distinct from the TAD-based definition of insulators (model *ii* in **Fig. 3e**), which is a
266 static definition that does not need the knowledge about the outcomes after loss of insulators.

267 In **Fig. 3f**, we show a FIN (highlighted in pink) that is not at TAD boundary. In this example we
268 also included the browser tracks for "insulation score" (InsScore)[35] and "directionality index"
269 (DI)[1]. TAD boundaries can be defined as bins where InsScore reaches local minimum or
270 positions where DI changes its sign (**Supp. Fig. 11a**), but the FIN in this example does not fit
271 into either criterion. There are much fewer FINs than TAD boundaries (**Fig. 3g**) and only a small
272 percentage of the FINs overlap with TAD boundaries (**Fig. 3h, Supp. Fig. 11b**). These
273 observations indicate that most FINs are inside TADs and conversely, loss of TAD boundaries
274 usually does not cause loop formation across their flanking regions. Change of InsScore also
275 does not predict FINs because although we observed an overall trend of increased InsScore at
276 FINs upon CTCF depletion, TAD boundaries also show the same trend and cannot be
277 distinguished from FINs (**Fig. 3i, Supp. Fig. 11c**). Interestingly, most (96%) of the FINs are
278 within compartment A (**Fig. 3j**), indicating that active E-P blocking mainly happens within
279 euchromatin region. Regardless, it is important to note that low-resolution 3D genome features,
280 including compartment and TAD boundaries, lack the precision to reveal the dynamic changes
281 of loop profiles as we used to define FINs.

282 Cohesin degron Hi-C data further supports a cohesin-dependent mechanism of FINs. We found
283 that almost none of the curated loop-gaining events appear in the *Rad21*-degron cells (**Fig. 3d**),
284 indicating cohesin dependency. Interestingly, a significant fraction of the *Type I* and *II*, but not
285 the *Type III* "gain of E-P loop" loci, can be also observed in *Wapl*-degron datasets (**Fig. 3d**).

286 This phenomenon can be explained by previous findings that *Wapl* deletion allows loop
287 extrusion to skip CTCF blocks[14, 15]. As shown in the carton in **Fig. 3k**, because *Wapl*-depletion
288 stabilizes chromatin-bound cohesin and leads to longer loop extrusion, previously insulated
289 enhancers and promoters can interact inside the bigger CTCF loop due to "insulator skipping",
290 creating the same E-P loop as observed in CTCF-depleted cells. All these results support FINs
291 as *bona fide* insulators of cohesin-mediated loop extrusion.

292 To explore which types of enhancers or promoters would gain chromatin loops upon CTCF
293 depletion, we picked the FIN-blocked ATAC-peaks in mESCs (n = 995) and scan for enriched or
294 depleted motifs compared to all mESC ATAC-peaks (n = 59,871). Interestingly, the loop-gaining
295 ATAC-peaks are enriched with many G-rich motifs but depleted with A/T-rich motifs (**Fig. 3l**).
296 This result is reminiscent of previous reports that G-quadruplexes (G4) structure may regulate
297 chromatin looping[36, 37]. Notably, many of the putative transcription factors binding to these G-rich
298 motifs have been reported to regulate loop formation in various cell types, such as Kruppel-like
299 factors[38, 39, 40], SP1[41], MAZ and several zinc-finger proteins[42, 43, 44]. Therefore, our results
300 suggest that G-rich *cis*-regulatory elements are more likely to act as new loop anchors upon
301 CTCF-depletion.

302 **FINs are associated with consistent early transcription responses upon CTCF depletion.**

303 We found that in all CTCF-degron studies FIN proximal regions (+/- 100kb) are 2~3 fold more
304 enriched with DEGs than TAD-boundary proximal regions, and the "gained" loop anchors are 10
305 times more enriched with DEGs (**Fig. 4a**). Furthermore, early response DEGs are clearly more
306 enriched at gained loop anchors (**Fig. 4b**), suggesting that FIN-blocked enhancers and
307 promoters are the direct targets of insulators. In 16.8% of FIN loci (49 out of 299), we observed
308 the blocking of 43 super-enhancers (there are 231 super-enhancers in total) (**Fig. 4c**). As
309 expected, genes gaining super-enhancer interactions are more likely to be upregulated than
310 genes gaining interactions with regular enhancers (**Fig. 4d**).

311 Using common cutoffs (2-fold for RNA-seq and 1.5-fold for NET-seq, $p<0.05$), each of the
312 CTCF-degron studies identified a few hundred early response DEGs, and the reproducibility
313 between independent studies is limited due to complex reasons (**Fig. 4e**). Both Nora *et al.* and
314 Kubo *et al.* performed RNA-seq and the earliest timepoint was 24hr CTCF-depletion, which may
315 still include late-response DEGs caused by indirect mechanisms (*e.g.*, loss of pluripotency).
316 Among the two RNA-seq studies, Kubo *et al*. detected fewer DEGs than Nora *et al.*, but 70% of

317 Kubo *et al*.'s DEGs are also observed in Nora *et al.* (47% of the upregulated DEGs and 81% of
318 the downregulated DEGs, **Fig. 4e**). This observation is consistent with our discussion above
319 that CTCF depletion in Kubo *et al*. was less complete. Hsieh *et al*. performed both NET-seq and
320 RNA-seq after only 3hr CTCF-depletion, which should include fewer late-response genes, but
321 the short treatment time may also reduce the sensitivity of DEG detection. We use NET-seq
322 data because Hsieh *et al.* reported that NET-seq has higher sensitivity in identifying DEGs[25] as
323 it maps newly synthesized RNA, while RNA-seq measures the amount of steady state RNA in
324 the cells.

325 We decided to focus on the comparison between Hsieh *et al.* and Nora *et al.* as they represent
326 a more complete set of CTCF direct targets. We next hypothesize that DEGs near FINs are
327 more likely to be the direct targets of CTCF-mediated enhancer-promoter insulation. If this
328 hypothesis is true, DEGs near FINs should be more reproducible between independent studies.
329 Indeed, the upregulated DEGs near FINs (*i.e.*, at "gained" loop anchors upon CTCF depletion)
330 are more reproducible, but downregulated DEGs near FINs are less reproducible than controls
331 (**Fig. 4f**, red bars). In contrast, the "lost" loop anchors does not predict the reproducibility of
332 DEGs. This observation suggests that transcription activation is the primary consequence of
333 insulator loss, consistent with the convention concept that insulators block enhancer-mediated
334 gene activation.

335 In **Fig. 4g** we plotted 49 upregulated DEGs at FIN loci as putative "direct targets" of insulators
336 (from Hsieh *et al.* or Nora *et al.*), among them 34 are considered as "high-confidence" targets
337 supported by both studies. One representative example is shown in **Fig. 4h** and **Supp. Fig.**
338 **12a**, in which we observed gain of loop between two putative enhancer clusters upon CTCF
339 depletion, leading to the upregulation of *Cdiptos* gene. Similarly, we also plotted 37 reproducibly
340 down-regulated DEGs near FINs supported by both studies (**Fig. 4i**). Strikingly, 18 of the
341 recurrent down-regulated DEGs are from protocadherin-γ, a well-known insulator loci (**Fig. 4i,j**).
342 Protocadherin proteins play key roles in the formation of synaptic connections between neurons
343 and the *Pcdh-γ* genes are distinguished by their variable promoters (exon1): γ*a*'s and γ*b*'s are
344 stochastically expressed in individual neurons and the γ*c*'s are constitutively expressed[45, 46, 47,
345 48]. In mESCs, the promoters of γ*a*'s and γ*b*'s are occupied by CTCF, H3K4me3, and the
346 repressive H3K27me3 mark (**Fig. 4j**). A cluster of putative enhancers marked by H3K27ac is
347 present at the 3' of γ*c* promoters (**Fig. 4j**, highlighted in green) forming loop interactions with the
348 promoters of γ*a*'s and γ*b*'s (**Fig. 4j**, blue arches), which are restricted by a FIN with multiple

349 CTCF peaks (**Fig. 4j**, highlighted in pink). Upon CTCF depletion, the H3K27ac-interactions are
350 weakened, and the *Pcdh-*γ promoter loops extend into a distal H3K27ac-poor region marked by
351 H3K27me3 (**Fig. 4j**), consistent with the downregulation of γ*a* and γ*b* genes. To verify the
352 function of the putative FIN, we designed a pair of sgRNAs to delete the four CTCF peaks in the
353 FIN and confirm the down-regulation of all γ*a* and γ*b* genes (**Fig. 4k-l**). Taking together, these
354 results support the transcription consequences of FINs and nominate "direct targets" of FINs in
355 mESCs.

356 **Validate FINs with multiplexed CTCF displacement assay**

357 To validate the *cis*-regulatory functions of FINs, we need to verify if the gained loops are truly
358 due to the loss of insulation. Because a FIN often involves multiple CTCF binding sites that
359 contribute to the insulating function collectively, we need to simultaneously perturb multiple
360 CTCF sites in a relatively big region. To achieve this, we combine a multi-sgRNA delivering
361 system name CARGO[49] with a dCas9-based CTCF-displacement method[47, 50] to perturb
362 multiple CTCF binding sites in the same cells. In this approach, 4-10 sgRNAs exactly
363 overlapping the cognate motifs of multiple CTCF binding sites are expressed from the same
364 CARGO plasmid. When co-transfected with a dCas9 construct, the sgRNAs can displace CTCF
365 from genome at multiple target sites simultaneously (**Fig. 5a**).

366 We used this approach to validate the gain-of-loop events at two FIN loci. The first FIN is close
367 to *Cdiptos*, a recurrent upregulated DEG (**Fig. 5b**). This example belongs to a simpler type of
368 FIN in which the newly gained loop (**Fig. 5b**, red arches) is blocked by only one FIN (**Fig. 5c**).
369 We used one CARGO construct targeting 4 CTCF motifs and confirmed with CUT&RUN that we
370 successfully weakened the target CTCF peaks but did not affect other CTCF peaks in this
371 region (**Fig. 5b**, highlighted in pink). We performed 4C-seq and confirmed the gain of long-range
372 chromatin interactions consistent with Hi-C data (**Fig. 5b**, highlighted in green). We also
373 validated with RT-qPCR that *Cdiptos* is upregulated upon displacement of target CTCF binding
374 sites (**Fig. 5d**). The second example is more complex as it may involve more than two FINs
375 (**Fig. 5e,f**). In this example *Mesp1* lost its loop interaction to a nearby FIN (**Fig. 5e**, blue arches)
376 and upon CTCF depletion, gained promiscuous distal interactions to a large region that is
377 ~200kb in size (**Fig. 5e**, red arches). For this locus we used one CARGO construct to target 9
378 CTCF motifs and successfully weakened the target CTCF peaks (**Fig. 5e**, highlighted in pink).
379 Again, using 4C-seq we confirmed the gain of interactions to the large distal region containing
380 several clusters of putative enhancers (**Fig. 5e**, highlighted in green), and confirmed the

381 upregulation of *Mesp1* with RT-qPCR (**Fig. 5g**). Taking together, these experiments verified the
382 insulating functions of FINs identified from CTCF-degron Hi-C data.

383

384 **Discussion**

385 In this study we propose to define functional insulators, or FINs, as CTCF binding sites that
386 actively block the loop formation between flanking *cis*-regulatory elements (**Fig. 3e**). Mapping
387 FINs requires robust detection of newly gained E-P loops, which is challenging for Hi-C analysis
388 because E-P loops are usually weak and hard to detect. With a meta-analysis of three
389 independent CTCF-degron Hi-C datasets, we demonstrate that the subtle gain of loop signal
390 detected by *DeepLoop* are quantitatively robust across different Hi-C or micro-C datasets,
391 therefore proves the feasibility to map FINs reliably with Hi-C at intermediate read depth.

392 The concept of FIN is distinct from the common assumption that TAD boundaries are insulators.
393 Most TADs disappear upon CTCF depletion, and we typically do not observe gain of new loops
394 across TAD boundaries. In contrast, most of the FINs are anchors of CTCF loops inside TADs
395 in control cells, and we call an CTCF loop anchor "FIN" only if its flanking sequences gain new
396 loops (**Fig. 3e**). Importantly, there are only a few hundred FINs in mESCs, which readily
397 explains why CTCF-depletion only causes mild changes of transcriptome. It should be also
398 noted that the gained E-P loops do not always predict significant gene activation due to complex
399 reasons[51]. Furthermore, target gene expression can be quantitative tuned by multiple enhancers
400 in an additively fashion, and each single enhancer may only have a small effect size[52].
401 Nevertheless, we showed that FINs better explain DEGs, especially the recurrent upregulated
402 DEGs in CTCF-depleted cells. This is consistent with the classic definition of insulators as DNA
403 elements that repress enhancer-mediated gene activation.

404 It remains unclear why only a small subset of CTCF loop anchors are FINs, *i.e.*, why is gain-of-
405 loop only observed at some loci but not others? Our analysis provided some hints. Firstly, the
406 fact that almost all FINs are in compartment A (euchromatin) suggests their dependency on
407 surrounding active chromatin. Notably, promoters and enhancers may play key roles in loop
408 initiation and stalling and are known to co-localize with cohesin and its loading factors[53]. Since
409 genome compartmentalization is a highly dynamic 3D genome feature, we speculate that FINs
410 are also likely to be cell type- or cell state-specific. Secondly, meta-analysis with *Rad21*- and
411 *Wapl*-degron Hi-C data shows that the newly gained E-P loops in CTCF-depleted cells do not

form in cohesin-depleted cells and are also likely to be governed by cohesin regulators. It is therefore plausible that local cohesin activity is a key determinant. Finally, *cis*-regulatory elements at the anchors of “gained loops” are enriched with G-rich TF motifs, suggesting that some CREs are particularly capable of forming stable long-range regulatory loops but such function may undergo tight control by insulators. Taking together, all these observations suggest context-specific roles of FINs, and it will be interesting to further explore how FINs exert their specific functions in health and diseases.

419 **Methods**

420 **Experiments**

421 *Cell culture*

422 Mouse embryonic stem cells (E14Tg2a) were cultured in DMEM+Glutamax (ThermoFisher ,
423 #10566-016) with the addition of 15% Fetal Bovine Serum (ThermoFisher, # SH30071.03),
424 550μM β-mercaptoethanol (ThermoFisher, #21985-023), 1mM Sodium Pyruvate
425 (ThermoFisher, #11360-070), 1X non-essential amino acids (ThermoFisher, #11140-50), and
426 10^4 U of Leukemia inhibitory factor (Millipore, #ESG1107). The cells were maintained at a
427 density of 0.2-1.5x10^5 cells/cm^2 and were passaged every 24-48 hours using TrypLE
428 (ThermoFisher, #12563011) on dishes coated with 0.1% gelatin (Millipore, # ES-006-B) at 37°C
429 and 5% CO2. The medium was refreshed daily when cells were not being passaged. Regular
430 checks for mycoplasma infection were performed every 3-4 months, with negative results.

431 *4C-seq*

432 4C-seq was conducted following a published protocol[54]. Briefly, 5 million cells fixed with 2%
433 formaldehyde, then quenched with 125 mM glycine. The fixed cells were lysed using a cell lysis
434 buffer containing 50 mM Tris-Cl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-
435 100, and 1x protease inhibitor cocktail (Roche, Cat#11873580001) for 20–30 minutes on ice.
436 Following lysis, nuclei were collected by centrifugation at 2,500g for 5 minutes at 4°C and
437 washed once with 1x restriction enzyme buffer. The nuclei pellets were resuspended in 1x
438 restriction enzyme buffer and treated with 0.3% SDS for 1 hour at 37°C with shaking, followed
439 by an additional hour with 2.5% Triton X-100. Chromatin was then digested overnight with the
440 appropriate restriction enzyme at the recommended temperature while rotating in an air bath.
441 The specific restriction enzymes used for each locus are detailed in **Supplemental Table 6**.
442 After digestion, the enzymes were inactivated by heating at 65°C, and nuclei were ligated with
443 50 µl of T4 DNA ligase (Invitrogen, Cat# 15224–090) in a 7 ml ligation solution at 16°C
444 overnight. Reverse cross-linking was performed by treating the samples with proteinase K to
445 isolate the proximity-ligated DNA. The purified DNA was quantified and subjected to secondary
446 restriction enzyme digestion using roughly 1 unit of restriction enzyme per 1 µg of DNA at the
447 recommended temperature overnight. After enzyme inactivation, the samples were self-ligated
448 with T4 DNA ligase. The ligated DNA was recovered using sodium acetate and ethanol and
449 quantified using the Qubit dsDNA HS assay kit (Thermo Fisher, Cat# Q32851). The 4C

450 templates were amplified using designed primers to generate sequencing libraries. The primer
451 system was modified to be compatible with the Illumina Nextera system, using two sequential
452 PCRs. Locus-specific inverse PCR primers are listed in **Supplemental Table 6**. For each locus,
453 the 4C templates were amplified with locus-specific primers using 200 ng of template per
454 reaction, and products from five parallel amplifications were pooled to generate the final 4C
455 library. A 50 µl PCR product aliquot was purified with homemade Sera-Mag beads. One-fifth of
456 the purified DNA was used for a second PCR with primers N7xx and N5xx, identical to Illumina
457 Nextera sample preparation primers. The final products were purified and sequenced. Reads
458 from the first cutting site were used for data analysis.

459 *Plasmid cloning*

460 The dCas9 expression vectors used in this study were created using the Cas9 expression
461 vector backbone from the pX330 plasmid (Addgene; plasmid 42230) through the In-Fusion
462 cloning technique. The dCas9 genes were amplified from the pHAGE EF1α dCas9-KRAB
463 plasmid (Addgene; plasmid 50919) via PCR and independently cloned into the AgeI and EcoRI
464 sites of the pX330 plasmid, substituting the Cas9 ORF. Detailed information regarding the
465 primers can be found in **Supplementary Table 5**. All sgRNAs in this study were designed using
466 the CCTop-CRISPR/Cas9 target online predictor.

467 The Cas9 and *Pcdhg* loop anchor target gRNA expression vector (pDonor_Rosa26_Cas9) was
468 constructed using the backbone of the plasmid pDonor MCS Rosa26 (Addgene, plasmid
469 #37200). An EF-1α-Cas9-P2A fragment, amplified from LentiV_Cas9_Puro (Addgene, plasmid
470 #108100), and a Puromycin-SV40 poly(A) signal fragment, derived from pR26 (Addgene,
471 plasmid #127372), were inserted into the backbone. A paired *Pcdhg* loop anchor target gRNA
472 expression cassette (sequences listed in **Supplementary Table 5**) was then introduced into the
473 construct.

474 *CARGO assembly*

475 The CARGO array assembly follows a published protocol[49]. The protocol consists of two main
476 steps: constant region preparation and array assembly. In the constant region preparation step,
477 the constant region with proper sticky ends is released by digesting the pGEMT-hU6-spSL
478 (provided by Dr. Joanna Wysocka) plasmid with BsmBI. The released constant region is then
479 separated from the pGEMT backbone on a 1% agarose gel and gel purified using a commercial
480 kit (QIAGEN, Cat#28704). Oligonucleotides (50-60 bp) with variable regions were ordered

481 through IDT (**Supplementary Table 5**). The forward and reverse oligos are annealed and
482 phosphorylated and then stored at -20°C.
483 In the array assembly step, the annealed chimeric oligo is mixed with the digested constant
484 region, and a minimum of 4-hour ligation with high-concentration T4 ligase (Thermofisher
485 Scientific, Cat#EL0013) is performed. The ligation step yields ligation products, referred to as
486 minicircles. After the stoichiometry check, individual minicircles are treated with plasmid-safe
487 exonuclease to clean up the un-ligated products. Then, the plasmid-safe exonuclease treated
488 individual minicircles are pooled together and purified. The pooled circle mixes are assembled
489 directly into the digested destination vector pX332-mCherry (provided by Dr. Joanna Wysocka)
490 through an optimized golden gate assembly reaction. The product of the golden-gate reaction is
491 then subjected to plasmid-safe exonuclease (Epicentre, Cat#E3101K) treatment. After that, the
492 exonuclease-treated reaction is then transformed into competent cells (Bioline, Cat#BIO-
493 85027). For each plate, individual colonies are picked for plasmid purification using a miniprep
494 kit (QIAGEN, Cat#27204) and test digestion with Acc65I, KpnI, and XbaI. Plasmids with the
495 correct digestion pattern are then verified by Sanger sequencing

496 *ChIPmentation*

497 We employed ChIPmentation[55] to map CTCF in various samples. Briefly, cells and tissues were
498 fixed in 1% formaldehyde at room temperature for 15 minutes, followed by glycine quenching.
499 To isolate nuclei, we lysed cell with lysis buffer 1 (50 mM HEPES; pH 7.5, 140 mM NaCl, 1 mM
500 EDTA, 10% glycerol, 0.5% Nonidet-40, 0.25% Triton X-100, and a protease inhibitor cocktail) for
501 10 minutes at 4°C. The nuclei were pelleted at 1,800 g for 10 minutes at 4°C and then
502 resuspended in nuclei lysis buffer (10 mM Tris-Cl pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5 mM
503 EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine, and a protease inhibitor cocktail) for
504 sonication. For CTCF pulldown, 100-150 µg of chromatin was used. Dynabeads M-280 (Life
505 Technologies, Sheep Anti-Rabbit IgG, #11204D) were washed three times with 0.5 mg/mL of
506 BSA/PBS on ice, then incubated with the CTCF antibody (Abcam, #ab70303) for at least 2
507 hours at 4°C. The beads/antibody complexes were then washed with BSA/PBS. Pulldown was
508 performed in binding buffer (1% Triton-X 100, 0.1% Sodium Deoxycholate, and a protease
509 inhibitor cocktail in 1X TE) by mixing the beads/antibody complexes with chromatin. After
510 overnight incubation, the beads/antibody/chromatin complexes were washed with RIPA buffer
511 (50 mM HEPES pH 8.0, 1% NP-40, 0.7% Sodium Deoxycholate, 0.5 M LiCl, 1 mM EDTA, and a
512 protease inhibitor cocktail). The beads complexes were then incubated with homemade Tn5
513 transposase in tagmentation reaction buffer (10 mM Tris-Cl pH 8.0 and 5 mM MgCl2) for 10

514 minutes at 37°C. To remove free DNA, the beads were washed twice with 1X TE on ice. The
515 pulldown DNA was recovered by reversing crosslinking overnight, followed by purification with
516 PCR Clean DX beads (Aline Biosciences, #C-1003-50). ChIP-seq libraries were generated by
517 amplifying the pulldown DNA with Illumina Nextera primers. Size selection was then performed
518 with PCR Clean DX beads to select fragments ranging from 100 bp to 1000 bp.

519 *Pcdhg loop anchor deletion*

520 Mouse embryonic stem cells (E14Tg2a) were transfected with the plasmid pX330-sgR26
521 (Addgene, plasmid #127376), which expresses mammalian codon-optimized Cas9 and the
522 Rosa26 sgRNA, together with the pDonor_Rosa26_Cas9 plasmid, using Lipofectamine™ 3000
523 reagent (Invitrogen) according to the manufacturer's instructions. Two days after transfection,
524 the cells were treated with puromycin (2 μg/mL) for three days. Drug-resistant clones were then
525 dissociated into single-cell suspensions, mixed thoroughly, and replated at low density to allow
526 single-colony formation, while maintaining puromycin selection (2 μg/mL). After four days,
527 individual colonies were picked for genotyping using primers specific for the *Pcdhg* loop anchor
528 deletion (**Supplementary Table 5**). Knockout cell clones were identified and selected for
529 subsequent experiments.

530 **Data analysis**

531 *Hi-C, micro-C, and 4C-seq*

532 Hi-C and micro-C sequencing reads were mapped to mm10 reference genome separately using
533 bowtie[56] (v.1.1.2) as previously described[28]. After removing non-unique reads and PCR
534 duplication reads, the bam files are used as input into *HiCorr* in '*Bam-process-HindIII*', '*Bam-*
535 *process-DpnII*', and '*Bam-process-microC*' modes for bias correction with conventional Hi-C, *in-*
536 *situ* Hi-C, and micro-C, respectively. After bias correction, we enhance the Hi-C/micro-C data
537 with *DeepLoop*. The *DeepLoop* package provides multiple models optimized for Hi-C data with
538 different read depths[28]. (Note that we use the number of *cis*-contact within 2Mbp as the
539 measurement of read depth.) To choose the appropriate model, we usually selected the model
540 with the closest read depth. The models chosen for each dataset are listed in **Supplementary**
541 **Table 2**. Because each dataset has a pair of Hi-C experiments (control and KO conditions), if
542 the read depths of the two Hi-C experiments are significantly different, we will down-sample the
543 high-depth dataset to lower depth to reduce potential biases (**Supplementary Table 2**).

544 4C-seq data were analyzed using pipe4C[54] (v.1.1.3) to generate bam and bigwig files for data
545 visualization on UCSC genome browser.

546 *ChIP–seq and CUT&RUN*

547 ChIP-seq and CUT&RUN data were mapped to mm10 (mouse) using bowtie. After mapping
548 step, we remove PCR duplication reads and use Macs2[57] to call peak and generated bigwig to
549 for data visualization on UCSC genome browser.

550 *Motif analysis*

551 Motif enrichment analysis was performed by comparing motif occurrences at ATAC peaks
552 located on gained loop anchors against all ATAC peaks. Background motif frequencies were
553 obtained by scanning all ATAC peak sequences using the mouse HOCOMOCO v11[58] motif
554 database with FIMO (MEME suite)[59]. Motif occurrences in ATAC peaks on gained loop anchors
555 were similarly identified by FIMO scanning. Odds ratios were calculated using Fisher's exact
556 test to quantify relative motif enrichment. To determine whether motifs were significantly
557 enriched or depleted, a binomial test was applied based on the observed verse expected motif
558 frequencies, and the resulting *p*-values were adjusted to q-values for motif testing correction.

559 *RNA-seq and mNET-seq*

560 RNA-seq data are aligned to mm10 using HISAT2[60] and read counts for each gene are called
561 with featureCounts[61]. DESeq2[62] is used to identify differential expressed genes (DEGs). The
562 DEGs are identified using Wald test with the p-value <0.01 and log2FC>1. For mNET-seq
563 analysis, we followed the customized pipeline from the Hsieh *et al.*[25] RNA-seq data are aligned
564 to mm10 using HISAT2[60] and read counts for each gene are called with featureCounts[61].
565 DESeq2[62] is used to identify differential expressed genes (DEGs). The DEGs are identified
566 using Wald test with the p-value <0.01 and log2FC>1. For NET-seq analysis, we followed the
567 customized pipeline from Hsieh *et al.* with the following steps: (i) Trim the adapter using
568 TrimGalore (v.0.6.7) (https://github.com/FelixKrueger/TrimGalore) to remove adapter sequences
569 'AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC' and
570 'GATCGTCGGACTGTAGAACTCTGAAC' at each side of the reads; (ii) Map trimmed read to
571 mm10 using STAR; (iii) Identify the last nucleotide incorporated by Pol II by using mNET_snr
572 (https://github.com/tomasgomes/mNET_snr) to locate the 3′-nucleotide of the second read and
573 the strand sign of the first read. (iv) Annotate and count the reads of each gene using

574 FeatureCounts[61]; (v) Use DEseq2[62] to quantify the changes of the mNET-seq signal at the gene
575 body between control and KO cells.

576 *Pairwise loop strength comparison*

577 For each pair of CTCF/cohesin-degron study, we took top 500k loop pixels from the *DeepLoop*
578 enhanced contact matrices of control and IAA/dTAG-treated samples separately, then union
579 them together. The loop strength of each pixel in the two experiments are normalized linearly:

580 $$Y_i \sim k * X_i + \varepsilon$$

581 $$i: A\ given\ loop\ pixel$$

582 $$X_i: i\ loop\ strength\ in\ WT\ condition$$

583 $$Y_i: i\ loop\ strength\ in\ KO\ condition$$

584 However, because CTCF- and *Rad21*-depletion leads to massive loss of loop signal, much
585 fewer loop pixels are expected in the IAA/dTAG-treated cells than in control cells. To adjust for
586 this, we have to manually pick a regression slope *k* based on the shape of the scatter plots.
587 The slopes we chose are listed in **Supplementary Table 2**. We then use the minimum strength
588 of the control cell loop pixels to adjust the minimum cutoff to call loop pixels in KO samples:
589 $Y_{min} = k * \min(X_i)$. In another word, a pixel is called as loop pixel in KO sample only if its
590 strength is bigger than Y_{min}. This step leads to much fewer loop pixels in CTCF- and *Rad21*-
591 depleted cells. We typically obtain ~600k chromatin loop for each study. We used a simple fold
592 change cutoff (FC > 4) to define gained or lost loop pixels. Finally, we used the following criteria
593 to call retained loops: (i) the strength of the pixel in both WT and KO conditions must be higher
594 than their minimum cutoffs ($X_i > X_{min}$ & $Y_i > Y_{min}$); (ii) the absolute value of residual must be
595 higher than the adjusted minimum of loop pixels in KO samples ($|\varepsilon| < Y_{min}$).

596 *Classify chromatin loops based on epigenome data*

597 For the purpose of being consistent, we intersected loop anchors with the same set of mESC
598 CTCF, H3K27ac, H3K4me3, H3K27me3 and H3K36me3 ChIP-seq peaks from ENCODE[63] to
599 classify the loop pixels. We classification loop pixels in the following order:

600 (1) EP-only: both anchors have either H3K4me3 or H3K27ac mark, but they do not have
601 CTCF peaks.
602 (2) CTCF-EP: after excluding EP-only, look for pixels which only one anchor has CTCF, and
603 the other anchor has either H3K4me3 or H3K27ac marker.

604 (3) CTCF-only: after excluding EP-only and CTCF-EP, look for pixels with at least one
605 anchor bound by CTCF.
606 (4) H3K27me3: after excluding (1) - (3) above, look for pixels with at least one anchor bound
607 by H3K27me3.
608 (5) H3K36me3: after excluding (1) - (4) above, look for pixels with at least one anchor bound
609 by H3K36me3.
610 (6) The rest pixels are classified as "other".

611 *Define Loop-gaining loci, FINs, FIN loci, and FIN-blocked regular enhancers and super*
612 *enhancers (SE)*

613 Loop-gaining loci are defined from "gained" loop pixels in CTCF-depleted cells. We identified
614 thousands of "gained" loop pixels and each pixel represents a pair of 5kb-bins. However, these
615 pixels often occupy overlapping regions. We therefore merge all overlapping pixels into one
616 region and define them as loop-gaining loci. This step allows us to count the number of regions
617 in the genome containing "gained" loops. In our study, we have ~300 loop-gaining loci after
618 merging ~5,000 recurrent "gained" loop pixels.

619 We define FINs as a subset of CTCF binding sites that satisfy the following criteria: (i) they
620 located at the anchors of CTCF loops that are lost upon CTCF depletion; (ii) the FIN should be
621 inside a newly gained loop in CTCF-depleted cells; (iii) the other anchor of the CTCF loop
622 should be outside of the newly gained loop; (iv) CTCF sites satisfying (i-iii) within 20kb of each
623 other will be merged into one FIN. Note that we do not have limits on the number of CTCF sites
624 to define FINs. Therefore, although a FIN can be just one single CTCF site, most FINs have
625 multiple CTCF sites.

626 As described above, all FINs must be inside loop-gaining loci. "FIN loci" are "loop-gaining loci"
627 that containing FINs. In fact, almost all the "loop-gaining loci" are also "FIN loci". We therefore
628 often use the two terms interchangeably.

629 We call enhancers as H3K27ac peaks and use a published list of super enhancers in mESCs[64]
630 (**Supplementary Tables 11 & 12**). This allows us to determine the types of enhancers blocked
631 by FINs.

632 *Recurrent loop-gaining pixels and loop-gaining loci in CTCF depleted cells*

633 A recurrent loop-gaining pixel is called if it satisfies FC > 4 in at least two studies. We define the
634 location of each loop pixel as the entire region between the left and right anchors (including the

635 anchors themselves). To define loop-gaining loci, we merge all gained loop pixels from the
636 CTCF-degron studies using the "merge" option in *bedtools*; each loop-gaining locus usually has
637 multiple loop pixels. Loop loci with gap size < 20kb will be merge into one locus. In this study,
638 we merged all the recurrent loop-gaining pixels to obtain 299 recurrent loop events. To test the
639 consistency of loop-gaining loci between independent CTCF-degron study, we call loop-gaining
640 loci for each individual CTCF-degron study and the share loop-gaining loci from two different
641 studies are called if they have at least one base overlap.

642 *Test the recurrence of loop-gaining loci from CTCF-degron in Rad21- and Wapl- degron data*

643 Nearly all loop-gaining loci observed in Hi-C degron studies have at least one E-P loop pixels.
644 We further classify the loop-gaining loci depending on the number of CTCF loops separating the
645 two anchors of the gained E-P loops. For all loci, we also manually examined all loci to confirm
646 the classification. Several loci cannot be clearly classified into the categories shown in **Fig. 3d**
647 because they have different marks or have multiple events. To test if a loop gaining locus is
648 recurrent in one of the CTCF/cohesin-degron study in **Fig. 3d**, we examine all loop-gaining
649 pixels in the locus and if one of pixels also gain signal it will be considered recurrent.

650 *Compartment and TADs in mESCs*

651 Hi-C compartment analysis is performed in WT mESCs following previous publication[65]. After QC
652 step, we constructed 500kb resolution contact matrices for each chromosome and then
653 normalized them using the genome distance. We next generated correlation matrices for each
654 chromosome and performed the principal component analysis on the correlation matrix and then
655 assigned the genome into two compartments by using PC1. We used the density of transcription
656 starting sites (TSS) of each 500-kb bin to determine compartment A and compartment B.

657 *TAD boundary calling with Insulation score or Directionality index (DI) methods*

658 Insulation scores were computed for contact matrices using 40kb bins and 2Mb sliding window
659 using the *cooltools*[66] package. Insulating loci were identified through a local minima detection
660 procedure based on peak prominence. Highly insulating regions corresponding to strong TAD
661 boundaries were defined using the Li thresholding method implemented in the *cooltools*
662 insulation score module. In this method, TAD boundaries are a list of 40kb bins.

663 We also called TAD domains with control mESC Hi-C data using the DI method described by
664 Dixon *et al.*[1], which quantifies the degree of upstream or downstream interaction bias for each

genome bin. Directionality indices were calculated also with 40kb bins using a 2Mb sliding window. A hidden Markov model (HMM) was applied to infer hidden bias states across the genome, classifying each bin as upstream-biased, downstream-biased, or neutral. Consecutive bins sharing the same bias state were merged into domains. The start and end positions of each domain are defined as TAD boundaries. In this way, the TAD boundaries will be also a list of single bases. The TAD boundaries from the WT samples of the three datasets are listed in **Supplementary Table 9**.

To compare the TAD boundaries from two methods, we consider two TAD boundaries overlapping if their distance gap is smaller than 40kb. The same distance cutoff (40kb) is also used to decide if a FIN overlaps TAD boundary.

Ideogram plot

We generate ideogram plots with Rideogram[67] to visualized the locations of recurrent gaining-loop pixels, compartment, and TADs.

Density and enrichment of DEGs upon CTCF-depletion

For any region in the genome, the density of DEGs is defined as the average number of DEGs per Mb-region. We computed DEG density for TAD boundary +/- 100kb, FIN +/- 100kb, gained loop anchors, lost loop anchors, and retained loop anchors. We also compute the enrichment of DEGs at regions of interest as their DEG density divided by the DEG density in the whole genome.

Hi-C data visualization

Heatmap was used to visualize the raw, ratio, and *DeepLoop*-enhanced heatmaps. We use the following strategies as described before[28] to determine the color scales these heatmaps. (i) Raw or ICE-normalized heatmaps represent raw read counts; the brightest red color indicates the 98th percentile of the contact matrix. Color is proportionally scaled down to one read (white). (ii) HiCorr heatmap represents the ratio value after Hi-C bias correction; the brightest red color indicates at least twofold enrichment of a given pixel comparing to its background. Color is proportionally scaled down to onefold (no enrichment). (iii) DeepLoop heatmaps outputs ‘transformed fold change’ that represents only relative levels of signal enrichment (for example, a value of one may no longer represents no enrichment). We therefore set the brightest red

color as the lower limit of the top 300k pixels genome wide. Color is proportionally scaled down to half of that lower limit or onefold, whichever is higher. All loop arches shown in figures were generated with UCSC Genome Browser[68].

Data availability

Accession numbers for third-party data used in this study can be found in **Supplementary Table 1**. The raw data of 4C–seq and ChIP-seq generated in this study, and reanalyzed published data, can be found at accession number GSE243728. The 18 Hi-C and micro-C datasets analyzed by *DeepLoop* can be found at https://hiview.case.edu/public/DeepLoop/.

To review GEO accession GSE243728:
Go to https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE243728
Enter token uvixkcaajpajpih into the box

Code availability

The code is available is available at github (https://github.com/JinLabBioinfo/FIN_project).

References

1. Dixon JR*, et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).

2. Nora EP*, et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385 (2012).

3. Bickmore WA, van Steensel B. Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270-1284 (2013).

4. Dekker J, Mirny L. The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110-1121 (2016).

5. McCord RP, Kaplan N, Giorgetti L. Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Mol Cell* **77**, 688-708 (2020).

6. Merkenschlager M, Nora EP. CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annual review of genomics and human genetics* **17**, 17-43 (2016).

7. Nora EP*, et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922 (2017).

8. Rao SSP*, et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e324 (2017).

9. Nasmyth K, Haering CH. Cohesin: its roles and mechanisms. *Annu Rev Genet* **43**, 525-558 (2009).

10. Chan KL, Roig MB, Hu B, Beckouet F, Metson J, Nasmyth K. Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. *Cell* **150**, 961-974 (2012).

11. Huis in 't Veld PJ*, et al.* Characterization of a DNA exit gate in the human cohesin ring. *Science* **346**, 968-972 (2014).

12. Kueng S*, et al.* Wapl controls the dynamic association of cohesin with chromatin. *Cell* **127**, 955-967 (2006).

13. Schwarzer W*, et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56 (2017).

14. Haarhuis JHI*, et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693-707 e614 (2017).

15. Wutz G*, et al.* Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *The EMBO journal* **36**, 3573-3599 (2017).

16. Wallace JA, Felsenfeld G. We gather together: insulators and genome organization. *Curr Opin Genet Dev* **17**, 400-407 (2007).

17. Wendt KS*, et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**, 796-801 (2008).

18. Parelho V*, et al.* Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* **132**, 422-433 (2008).

19. Franke M*, et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265-269 (2016).

20. Weischenfeldt J*, et al.* Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat Genet* **49**, 65-74 (2017).

21. Lupianez DG*, et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025 (2015).

22. Anania C*, et al.* In vivo dissection of a clustered-CTCF domain boundary reveals developmental principles of regulatory insulation. *Nat Genet* **54**, 1026-1036 (2022).

23. Zuin J*, et al.* Nonlinear control of transcription through enhancer-promoter interactions. *Nature* **604**, 571-577 (2022).

24. Kubo N*, et al.* Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat Struct Mol Biol* **28**, 152-161 (2021).

25. Hsieh TS, Cattoglio C, Slobodyanyuk E, Hansen AS, Darzacq X, Tjian R. Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat Genet* **54**, 1919-1932 (2022).

26. Akgol Oksuz B*, et al.* Systematic evaluation of chromosome conformation capture assays. *Nature methods*, (2021).

27. Lu L*, et al.* Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of Non-coding Genome in Neural Development and Diseases. *Mol Cell* **79**, 521-534.e515 (2020).

28. Zhang S*, et al.* DeepLoop robustly maps chromatin interactions from sparse allele-resolved or single-cell Hi-C data at kilobase resolution. *Nat Genet* **54**, 1013-1025 (2022).

29. Kriz AJ, Colognori D, Sunwoo H, Nabet B, Lee JT. Balancing cohesin eviction and retention prevents aberrant chromosomal interactions, Polycomb-mediated repression, and X-inactivation. *Mol Cell* **81**, 1970-1987 e1979 (2021).

30. Rhodes JDP*, et al.* Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells. *Cell reports* **30**, 820-835 e810 (2020).

31. Liu NQ*, et al.* WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nat Genet* **53**, 100-109 (2021).

32. Imakaev M, *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature methods* **9**, 999-1003 (2012).

33. Yaffe E, Tanay A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat Genet* **43**, 1059-1065 (2011).

34. Lu L, *et al.* Robust Hi-C Maps of Enhancer-Promoter Interactions Reveal the Function of Non-coding Genome in Neural Development and Diseases. *Mol Cell* **79**, 521-534 e515 (2020).

35. Crane E, *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240-244 (2015).

36. Li L, *et al.* YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat Chem Biol* **17**, 161-168 (2021).

37. Wulfridge P, *et al.* G-quadruplexes associated with R-loops promote CTCF binding. *Mol Cell* **83**, 3064-3079 e3065 (2023).

38. Schoenfelder S, *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat Genet* **42**, 53-61 (2010).

39. Di Giammartino DC, *et al.* KLF4 is involved in the organization and regulation of pluripotency-associated three-dimensional enhancer networks. *Nat Cell Biol* **21**, 1179-1190 (2019).

40. Wei Z, *et al.* Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency. *Cell stem cell* **13**, 36-47 (2013).

41. Bryan AF, Justice M, Stutzman AV, McKay DJ, Dowen JM. Cohesin stabilization at promoters and enhancers by common transcription factors and chromatin regulators. *Epigenetics Chromatin* **18**, 33 (2025).

42. Ortabozkoyun H, *et al.* CRISPR and biochemical screens identify MAZ as a cofactor in CTCF-mediated insulation at Hox clusters. *Nat Genet* **54**, 202-212 (2022).

43. Xiao T, Li X, Felsenfeld G. The Myc-associated zinc finger protein (MAZ) works together with CTCF to control cohesin positioning and genome organization. *Proceedings of the National Academy of Sciences of the United States of America* **118**, (2021).

44. Ortabozkoyun H, *et al.* Members of an array of zinc-finger proteins specify distinct Hox chromatin boundaries. *Mol Cell* **84**, 3406-3422 e3406 (2024).

45. Hirayama T, Yagi T. Regulation of clustered protocadherin genes in individual neurons. *Seminars in cell & developmental biology* **69**, 122-130 (2017).

46. Flaherty E, Maniatis T. The role of clustered protocadherins in neurodevelopment and neuropsychiatric diseases. *Curr Opin Genet Dev* **65**, 144-150 (2020).

47. Jia Z, Li J, Ge X, Wu Y, Guo Y, Wu Q. Tandem CTCF sites function as insulators to balance spatial chromatin contacts and topological enhancer-promoter selection. *Genome Biol* **21**, 75 (2020).

48. Hirayama T, Tarusawa E, Yoshimura Y, Galjart N, Yagi T. CTCF is required for neural development and stochastic expression of clustered Pcdh genes in neurons. *Cell reports* **2**, 345-357 (2012).

49. Gu B*, et al.* Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* **359**, 1050-1055 (2018).

50. Tarjan DR, Flavahan WA, Bernstein BE. Epigenome editing strategies for the functional annotation of CTCF insulators. *Nature communications* **10**, 4258 (2019).

51. Pollex T*, et al.* Enhancer-promoter interactions become more instructive in the transition from cell-fate specification to tissue differentiation. *Nat Genet* **56**, 686-696 (2024).

52. Fulco CP*, et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat Genet* **51**, 1664-1669 (2019).

53. Kagey MH*, et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435 (2010).

54. Krijger PHL, Geeven G, Bianchi V, Hilvering CRE, de Laat W. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* **170**, 17-32 (2020).

55. Schmidl C, Rendeiro AF, Sheffield NC, Bock C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nature methods* **12**, 963-965 (2015).

56. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).

57. Zhang Y*, et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

58. Vorontsov IE*, et al.* HOCOMOCO in 2024: a rebuild of the curated collection of binding models for human and mouse transcription factors. *Nucleic acids research* **52**, D154-D163 (2024).

59. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018 (2011).

60. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nature methods* **12**, 357-360 (2015).

61. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).

62. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

63. Yue F*, et al.* A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355-364 (2014).

64. Hnisz D*, et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947 (2013).

65. Lieberman-Aiden E*, et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).

66. Abdennur N, Mirny LA. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311-316 (2020).

67. Hao Z*, et al.* RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci* **6**, e251 (2020).

68. Navarro Gonzalez J*, et al.* The UCSC Genome Browser database: 2021 update. *Nucleic acids research* **49**, D1046-D1057 (2021).

Acknowledgements

This work was supported by grants from NIH R01HG009658 and R01CA267872 (to F.J.), R01DK113185 (to Y.L.), R01DK131437, R01HG012384, and UG3NS132061 (to Y.L. and F.J.). F.J. is also supported by a subaward from NIH project (U01AG072579) through University of Miami and partially by several other NIH grants R01CA252224, R01GM148662, and P01CA272161. Y.L. is also partially supported by NIH grants R01CA264320. This work made use of the High-Performance Computing Resource in the Core Facility for Advanced Research Computing at Case Western Reserve University.

Author Contributions Statement

F.J. and Y.L conceived the project and secured funding for this project. J.C. designed and performed the FIN validation experiments, W.X. and X.L. performed the data analysis. L.L. help with the experiments and S.Z. and X.L. helped with the data analysis. F.J., W.X., J.C.,X.L., and Y.L. wrote the manuscript. F.J. and Y.L. also provided critical feedback and interpretation of results as well as helping to revise the manuscript.

Competing Interests Statement

The authors declare no competing interests

949 **Figure Legend**

950 **Figure 1. *DeepLoop* allows cross-platform comparison between independent CTCF- and**
951 **cohesin-degron Hi-C datasets at loop-resolution.**

952 **a,** A list of published Hi-C and micro-C studies with acute depletion of *Ctcf*, *Rad21*, or *Wapl* in
953 mESCs. **b,** Matrices summarizing R-square values when comparing contact heatmaps from
954 different pipelines and orthogonal Hi-C/micro-C datasets in control mESCs. **c,** Correlation
955 matrices comparing the differential matrices of all nine degron experiments. For each degron
956 experiment, the differential matrix is the difference between two Hi-C contact matrices (depletion
957 – control) at 5kb resolution. Left, Hi-C contact matrices from *ICE* are used to compute
958 differential matrices; right, Hi-C contact matrices from *DeepLoop* are used. **d,** Contact heatmaps
959 of the CTCF-, Rad21-, and Wapl-degron Hi-C data at *Notch1* locus with *ICE*, *HiCorr* and
960 *DeepLoop*. Red dashed circle: gained loops. Numbers in brackets indicate the color scale of
961 heatmaps: for *ICE* normalization and *HiCorr* heatmaps, the maximum intensity corresponds to
962 the 98th percentile of the corresponding contact matrix; for *DeepLoop* heatmaps, the maximum
963 intensity corresponds to the lower cutoff of top 500,000 pixels genome-wide. **e**, Genome
964 browser tracks showing ChIP-seq data and Hi-C loops from Nora *et al.* at *Notch1* locus. Blue
965 highlighted region showing Notch1 gene which is at the anchor of a gained loop. Pink region
966 highlights the potential insulator region that blocks the formation of *Notch1* E-P loop. RNA-seq
967 results are also shown for *Notch1* gene in WT (black) and CTCF-depleted cells (red). Bar plot:
968 quantitation of *Notch1* expression from RNA-seq. P-value: one-side Wald test provided by DE-
969 seq2.

970 **Figure 2. Identify and characterize recurrent gain of E-P loop events in CTCF-depleted**
971 **cells through cross-platform Hi-C meta-analyses.**

972 **a-c,** Meta-analysis of independent CTCF-degron Hi-C datasets. **a**, Left: scatterplot comparing
973 the strength of loop pixels between WT and CTCF-depleted mESCs (every dots represent one
974 pixel in the contact heatmap); middle: the same scatterplot as left but highlight the lost (blue),
975 retained (black), and gained (red) loop pixels; right: violin plot comparing the size of loop pixels
976 and the value below each plot indicates the median loop size for the corresponding category. **b,**
977 Pie charts showing the components of lost, retained and gained loop pixels based on epigenetic
978 marks. **c**, Scatter plots showing the consistency of lost, retained and gained loop pixels between
979 three CTCF degron studies. Plots in the left column highlight the lost, retained, and gained loop

980 pixels from Nora *et al*., the right two columns show the quantitative changes of the
981 corresponding pixels in *Kubo et al.* or *Hsieh et al*. **d,** The lost, retained, and gained loop pixels
982 from one CTCF-degron Hi-C data (Nora *et al.*) are plotted onto Hi-C scatter plots from *Rad21*-
983 or *Wapl*-degron studies to examine their quantitative changes.

984 **Figure 3. Characterization of functional insulators (FINs) in mESC genome.**

985 **a,** Heatmap showing the union of all gained loop pixels from three CTCF-degron Hi-C datasets.
986 Every row represents one CTCF-degron study, and every column represents one loop pixel.
987 Loop pixels are organized into clusters based on their reproducibility. The number of loop pixels
988 in each cluster and their median size are also shown. Color scale indicates the increase of loop
989 signal in log scale. **b**, Venn diagram showing the consistence of loop-gaining loci identified from
990 three datasets. **c,** Pie chart showing that most of the gained loops in CTCF-depleted cells have
991 at least one CTCF loop anchor in between in wildtype cells. **d,** A table summarizing the
992 classification and recurrence of loop gaining loci between independent CTCF- or cohesin-
993 degron studies. **e,** Cartoons summarizing three possible ways to define insulators. **f,** An
994 example of non-TAD boundary FIN. The FIN is highlighted in pink. The tracks of InsScore and
995 DI are also included. TAD boundaries determined from InsScore and DI are highlighted in
996 purple and indicated with black arrows, respectively. **g,** Ideogram showing the locations of loop-
997 gaining loci from three studies, TAD boundaries, and compartments on chromosome 1. FINs
998 from three different CTCF-degron studies are shown in different colors, and the color scale
999 shows the fold increase of gained loop in log scales. **h,** Pie chart showing the percentage of
1000 FINs that overlap TAD boundary. **i,** Boxplot comparing the insulation scores at FINs and TAD
1001 boundaries before and after CTCF depletion computed with HI-C data from Nora *et al*. **h,** Stack
1002 bar plot comparing the proportion of FINs and TAD boundaries in compartment A or B. **i,** A
1003 model to explain why some of the newly gained E-P loops in CTCF-depleted cells are also
1004 observed in Wapl-depleted cells. **j,** A list of TF motifs enriched or depleted in the mESC ATAC-
1005 seq peaks that will gain chromatin loops upon CTCF depletion. Enrichment or depletion are
1006 relative to all mESC ATAC-seq peaks.

1007 **Figure 4. FINs are associated with consistent early transcriptional consequences in**
1008 **independent CTCF-degron studies.**

1009 **a,** Bar plots comparing the density of DEGs upon CTCF-depletion across different genomic
1010 regions. The y-axis indicates the number of DEGs per Mb. The number of DEGs were

1011 calculated from the following datasets: 3-hour NET-seq for Hsieh *et al.*, 24-hour RNA-seq for
1012 Nora *et al*., and 24-hour RNA-seq for Kubo *et al.* **b,** Enrichment of DEGs at different types of
1013 genomic regions compared to whole genome as background. X-axis: time of IAA treatment (hrs)
1014 in each study; y-axis: fold enrichment. Each panel represents one CTCF-degron study. **c,** Pie
1015 chart showing the percentage of FINs that are blocking super enhancers; **d,** Boxplot showing
1016 the expression changes of genes gaining enhancer loop. Red: genes gaining super-enhancer
1017 loop; Blue: genes gaining regular enhancer loop. *, $p < 0.05$, Wilcoxon signed-rank test. **e,** Venn
1018 diagrams showing the overlap of all up- (left panel) and down-regulated (right panel) DEGs
1019 between three CTCF-degron studies. **f,** Bar plot comparing the reproducibility scores (measured
1020 by Jaccard Index) of up- or down-regulated DEGs between two CTCF-degron studies (Hsieh *et*
1021 *al.* and Nora *et al.*). Gray bars: all DEGs; blue bars: DEGs at the anchors of lost loops; red bars:
1022 DEGs at the anchors of gained loops. P-values are from binomial test. **g,** Heatmap showing the
1023 union of upregulated DEGs at the anchors of gained loops from *Hsieh et al* and *Nora et al*.;
1024 DEGs are clustered based on consistency. **h,** Left panel shows the browser tracks near *Cdiptos*
1025 locus and right panel shows the Hi-C heatmaps in the same region from *Hsieh et al* and *Nora et*
1026 *al.* The direction of CTCF motifs are indicated by triangles. Blue arches (left) and blue dashed
1027 circles (right): lost loop pixels; red arches and red dashed circles: gained loop. Regions
1028 highlighted in green: enhancers that gained loops; the region highlighted in pink: a CTCF loop
1029 anchor is defined as *FIN*. **i,** Similar as **g**, a heatmap showing the union of down-regulated DEGs
1030 at the anchors of gained loops from *Hsieh et al.* and *Nora et al*. Among these genes 18 genes
1031 are from the Pcdh-γ locus (shown as a separate heatmap). **j,** Left panel shows the browser
1032 tracks near Pcdh-γ locus and right panel shows the Hi-C heatmaps from two CTCF-degron
1033 studies. **k,** Design of sgRNAs for insulator deletion. **l,** RT-qPCR shows the down-regulation of
1034 Pcdh-γ genes in mESCs after insulator deletion.

1035 **Figure 5. Validate FINs with multiplexed CTCF displacement assay.**

1036 **a,** Scheme of the CARGO vector that can express multiple sgRNAs. In this system, the
1037 sequences of *n* sgRNAs are broken into left and right halves and inserted into (*n*+1) minicircles.
1038 The *i*-th minicircle contains the right half of the (*i-1*)-th sgRNA and the left half of the *i*-th sgRNA
1039 sequences divided by a *BpiI* restriction site. Final CARGO vector can be assembled by ligating
1040 the re-linearized minicircles and vector backbone; each sgRNA is driven by a U6 promoter. Co-
1041 transfect CARGO vector and a dCas9-expressing plasmid can displace multiple CTCF binding
1042 sites in the same cell. **b,** Validation of a FIN at the *Cdiptos* locus. From the top, 1st and 2nd

1043 tracks are H3K4me3 and H3K27ac ChIP-seq data showing the promoter and enhancers in
1044 mESC. 3rd and 4th tracks show CTCF ChIP-seq data before and after targeting 4 CTCF motifs in
1045 FIN (highlighted in pink) with CARGO. 5th and 6th are Hi-C arches (*DeepLoop* pixels with data
1046 from Nora *et al.*) before and after CTCF-depletion (blue: lost loops; red: gained loops). The
1047 bottom 3 tracks show 4C-seq data (bait highlighted in yellow) before and after targeting CTCF
1048 sites; regions highlighted in green gain 4C signal. The raw and ICE, HiCorr, or DeepLoop
1049 processed contact heatmaps are also included. **c,** A model cartoon showing that the gained
1050 loop at this FIN locus only involves one CTCF loop. **d,** RT-qPCR results showing that *Cdiptos* is
1051 upregulated after CTCF displacement. ** $p<0.01$; one-side t-test. **e-g,** Validation of a FIN at
1052 *Mesp1* locus similar as shown in **b-d**. Note the gained loop at this FIN locus involves at least
1053 two CTCF loops (**f**).
1054

1055 **Supplementary figure legend**

1056 **Supplementary Figure 1. Compare the compartments and TADs between orthogonal**
1057 **CTCF-degron datasets from Hi-C, in situ Hi-C, and micro-C.**

1058 **a-c,** Compartment analysis of CTCF-degron Hi-C datasets at 500kb resolution. **a,** Correlation
1059 heatmaps of the three CTCF-degron Hi-C datasets at 500kb resolution showing the
1060 compartments in control and CTCF-depleted cells. PC1 values are shown on top of each
1061 heatmap indicating compartment A (positive PC1) and B (negative PC1). **b**, Scatter plots
1062 comparing the PC1 values for each 500kb bin before and after CTCF depletion. **c,** Top row:
1063 scatter plots comparing the PC1 values of Hi-C data in control cells from three independent
1064 studies; bottom row: scatter plots comparing the PC1 values of Hi-C data in CTCF-depleted
1065 cells from three studies. **d-f,** TAD analysis of CTCF-degron Hi-C datasets at 40kb resolution. **d,**
1066 ICE normalized contact heatmaps of the three CTCF-degron Hi-C datasets at 40kb resolution.
1067 TADs are recognizable as squares along the diagonals. Note the loss of TADs in CTCF-
1068 depleted cells. **e,** Scatter plots compare the directionality index (DI) values of all 40kb bins
1069 before and after CTCF-depletion in all three studies. Note that the DI values of many bins
1070 become close to zero in CTCF-depleted cells, indicating that these bins lost directional biases,
1071 consistent with massive loss-of-TAD upon CTCF-depletion. **f,** Top row: scatter plots comparing
1072 the DI values of all 40kb bins in control cells from three independent studies; bottom row: scatter
1073 plots comparing the DI values of all 40kb bins in CTCF-depleted cells from three studies. Note
1074 that although the R-squares between Nora et al. and Kubo et al. are rather high (R-square

1075 0.718 and 0.612 for control and △CTCF cells), the DI values from micro-C data (Hsieh et al.)
1076 deviate significantly from the other two studies (R-square 0.295~0.458). In micro-C data, more
1077 bins show polarized DI values (*i.e.*, stronger directionality biases), consistent with the general
1078 belief that with ICE-based pipeline, micro-C identifies more loops than Hi-C or in situ Hi-C.

1079 **Supplementary Figure 2. Compare the loop-level analyses of orthogonal Hi-C datasets**
1080 **with ICE, HiCorr, and DeepLoop, related to Fig. 1b.**

1081 **Part I.** Cross-platform comparison of CTCF-degron datasets. **a-c,** Loop analysis of three CTCF-
1082 degron Hi-C datasets with ICE at 5kb resolution. **a,** ICE heatmaps of the three CTCF-degron Hi-
1083 C datasets at 5kb resolution. Note that there are more noises in Hi-C (Nora) and *in situ* Hi-C
1084 (Kubo) heatmaps than in micro-C (Hsieh) heatmaps. **b,** Scatter plots of ICE heatmap pixels
1085 comparing the control cells contact heatmaps from three independent studies. Note that the
1086 consistency between orthogonal Hi-C protocols is low (R-square values close to zero). **c**,
1087 Scatter plots comparing the ICE heatmaps of control and ΔCTCF cells in three different studies
1088 (5kb resolution). Every dot in the scatter plot represents one pixel in the contact matrix and the
1089 x-axis and y-axis indicate the intensity of the pixel in two experiments. Note that △CTCF cells
1090 have many pixels stronger than WT cells because the read-depth of micro-C data in △CTCF
1091 cells is higher than in control cells. **d-f,** Similar as **a-c** but the comparisons are between ICE
1092 heatmaps after correcting distance biases, which are observed/expected ratio heatmaps. Note
1093 that in **f**, the control vs. △CTCF scatter plot for Hsieh *et al.* correctly show stronger loop pixels
1094 in control cells than △CTCF cells. However, the scatter plots in **e** show that the consistency
1095 between independent studies remains poor with R-square values still close to zero, indicating
1096 that the pixel-level correlations are dominated by random noise pixels. **g-i,** the same
1097 comparisons between HiCorr-corrected heatmaps, which are also distance-corrected ratio
1098 heatmaps. Note that the scatter plots in **i** show that in each of the control vs. △CTCF
1099 comparisons, we start to observe retained pixels with similar strength before and after CTCF
1100 depletion. Furthermore, the scatter plots in **h** also show that HiCorr achieves higher R-square
1101 values across different Hi-C platforms. **j-l,** comparisons between *DeepLoop* heatmaps. Note
1102 that DeepLoop highlights loop pixels and suppresses noise pixels (**j**). The retained pixels and
1103 weakened pixels are more clearly recognizable in each of the control vs. △CTCF scatter plots
1104 (**l**). DeepLoop also shows the best cross-platform consistency at pixel level (**k**).
1105 **Part II.** Cross-platform comparison of Rad21-degron datasets. Same as Part I and all
1106 conclusions are the same.

1107 **Part III.** Cross-platform comparison of Wapl-degron datasets. Same as Part I and all
1108 conclusions are the same.

1109 **Supplementary Figure 3. Correlation matrices comparing the profiles of dynamic loops**
1110 **from all nine pairs of degron experiments, related to Fig. 1c.**

1111 For each degron experiment, we first obtain a pair of 5kb resolution Hi-C contact matrices in
1112 control and depletion cells. Next, we computed the delta matrices (depletion – control) for all
1113 nine experiments. A correlation matrix is then generated between the nine delta matrices. **a-d**
1114 are correlation matrices when contact matrices are generated with different pipelines.

1115 **Supplementary Figure 4. Recurrent gain of E-P loop at *Notch1* locus between**
1116 **independent CTCF-degron Hi-C studies, related to Fig. 1d,e.**

1117 **a,** Contact heatmaps of the CTCF-degron Hi-C data by Kubo *et al.* at *Notch1* locus with *ICE*,
1118 *HiCorr* and *DeepLoop*. Red dashed circle highlights the same locations where we observed a
1119 gained E-P loop in the contact heatmap from Nora *et al.* (shown in **Fig. 1d**). **b**, Genome browser
1120 tracks showing ChIP-seq data and Hi-C loops from Kubo *et al.* Blue highlighted region indicates
1121 *Notch1* gene. Pink region highlights the potential insulator region that blocks the formation of
1122 *Notch1* E-P loop. **c,** RNA-seq results from Kubo *et al.* are also shown as bar plot. p-value: one-
1123 side Wald test provided by DE-seq2. **d-f,** the same analysis as **a-c** at *Notch1* locus using data
1124 from Hsieh *et al.* Note that Hsieh *et al.* used mNET-seq for transcriptome profiling. **g,** Contact
1125 heatmaps of *Notch1* locus from two *Rad21*-degron datasets (Kriz *et al.* and Hsieh *et al.*) in
1126 addition to the one included in **Fig. 1d** (Rhodes *et al.*). **h,** Contact heatmaps of *Notch1* locus
1127 from two *Wapl*-degron datasets (Kriz *et al.* and Hsieh *et al.*) in addition to the one included in
1128 **Fig. 1d** (Liu *et al.*).

1129 **Supplementary Figure 5. Meta-analysis of independent CTCF-degron Hi-C datasets to**
1130 **examine the consistency and epigenetic features of dynamic loops.**

1131 **a**, Using the *in situ* Hi-C data from the CTCF-degron study by Kubo *et al.*, left: scatterplot
1132 comparing the strength of loop pixels between WT and CTCF-depleted mESCs; middle: the
1133 same scatterplot as left but highlight the lost (blue), retained (black), and gained (red) loop
1134 pixels; right: violin plot comparing the size of loop pixels. **b,** Pie charts showing the components
1135 of lost, retained and gained loop pixels based on epigenetic marks. **c**, Scatter plots in the left

1136 column highlight the lost, retained, and gained loop pixels from Kubo *et al.*, the right two
1137 columns show the quantitative changes of the corresponding pixels in Nora *et al.* or Hsieh *et al*.
1138 **d-f,** The same analyses as **a-c** are performed with CTCF-degron micro-C data by Hsieh *et al.*

1139 **Supplementary Figure 6. Examples of chromatin loop classification.**

1140 Each panel shows the genome browser tracks with loop arches. These examples include **a**, EP
1141 loop; **b**, CTCF-EP loop; **c**, CTCF loop; **d**, H3K27me3 loop; **e**, H3K36me3 loop.

1142 **Supplementary Figure 7. Meta-analysis of independent *Rad21*- and *Wapl*- degron Hi-C**
1143 **datasets to examine the consistency and epigenetic features of dynamic loops.**

1144 **Part I:** Meta-analysis of *Rad21*-degron datasets. **a,** Left and middle panels define the lost,
1145 retained, and gained loop pixels upon *Rad21*-depletion from the study by Rhodes *et al.*, and the
1146 right panel compares their loop size distribution. **b,** Pie charts showing the components of lost,
1147 retained and gained loops based on epigenetic marks. **c,** Scatter plots testing if the lost,
1148 retained and gained loop pixels from Rhodes *et al.* are quantitatively consistent in the other two
1149 *Rad21*-degron studies. **d-f,** the same analyses as **a-c** are performed with *Rad21*-degron Hi-C
1150 data by Kriz *et al.* **g-i,** the same analyses as **a-c** are performed with *Rad21*-degron micro-C data
1151 by Hsieh *et al.*
1152 **Part II:** Meta-analysis of *Wapl*-degron datasets. **j-r,** Same as Part I except that the analyses
1153 were done with Wapl-degron datasets. Note there is an overall trend that *Wapl*-depletion causes
1154 more “gained” loops.

1155 **Supplementary Figure 8. “Gained” long-range H3K27me3 loops in Rad21-depleted cells**
1156 **from Rhodes *et al.* exist before Rad21-depletion in two other studies.**

1157 Two examples are shown in **a** and **b**. In each example, Hi-C data are shown as both contact
1158 heatmaps (DeepLoop enhanced) and loop arches. Browser tracks of CTCF, H3K27ac,
1159 H3K4me3, and H3K27me3 are also included. Note that the gained H3K27me3 loop in Rhodes
1160 et al. already exist before Rad21-depletion in two other studies.

1161 **Supplementary Figure 9. Meta-analysis reveals the fate of dynamic loops from CTCF-**
1162 **degron studies upon depletion of cohesin factors.**

1163 **a,** The lost, retained, and gained loop pixels from one CTCF-degron Hi-C data (Kubo *et al.*) are
1164 plotted onto Hi-C scatter plots from *Rad21*- or *Wapl*-degron studies to examine their quantitative
1165 changes. **b,** Same analyses are done with the lost, retained, and gained loop pixels from the
1166 CTCF-degron Hi-C data by Hsieh *et al.*

1167 **Supplementary Figure 10. Examples of loop gaining loci, related to Fig. 3d.**

1168 **a,** An example of Type I EP loop gaining loci, in which one of the two E/P sites is insulated in a
1169 CTCF loop while the other site is outside the CTCF loop; **b,** An example of Type II EP-loop
1170 gaining loci, in which two E/P sites are insulated in two CTCF loops; **c,** An example of Type III
1171 EP loop gaining loci, in which neither of the two E/P sites are insulated by CTCF loops; **d,** An
1172 example of CTCF loop gaining loci, in which the newly gained loop are between incompletely
1173 degraded CTCF sites.

1174 **Supplementary Figure 11. Limited overlap between FINs and TAD boundaries, related to**
1175 **Fig. 3f,h,i.**

1176 **a,** Venn diagrams comparing the TAD boundaries called by two methods from three studies. **b,**
1177 Pie chart showing the percentage of FINs that overlap TAD boundaries called from different
1178 datasets and algorithms. **c,** Compare insulation scores at FINs and TAD boundaries in different
1179 studies.

1180 **Supplementary Figure 12. Example contact heatmaps at putative insulator target genes,**
1181 **related to Fig. 4h,j.**

1182 **a,** Raw, *ICE*, *HiCorr*, and *DeepLoop* contact heatmaps at *Cdiptos* locus in control and CTCF-
1183 depleted cells from Hsieh *et al.* and Nora *et al.* **b,** Similar as **a** but the heatmaps are at the
1184 Pcdh-γ locus.

Figure 1

a

Published Hi-C data in mESCs upon acute depleation of architectural proteins

Publication	mESC line	KO protocol	C-technique	# of contacts Ctrl	# of contacts IAA/dTAG
Nora et al.	E14TG2a	C-Ctcf AID (48hrs)	HindIII Hi-C	244M	242M
Kubo et al.	F123	C-Ctcf AID (48hrs)	in-situ Hi-C	185M	1 16M
Hsieh et al.	JM8.N4	N-Ctcf AID (3hrs)	micro-C	291M	830M
Rhodes et al.	E14TG2a	C-Rad21 AID (6hrs)	in-situ Hi-C	63M	43M
Kriz et al.	F123	C-Rad21 dTAG (8hrs)	In-situ Hi-C	136M	100M
Hsieh et al.	JM8.N4	C-Rad21 AID (3hrs)	micro-C	283M	1.00B
Liu et al.	E14TG2a	C-Wapl AID (24hrs)	in-situ Hi-C	77M	88M
Kriz et al.	F123	C-Wapl dTAG (72hrs)	in-situ Hi-C	193M	147M
Hsieh et al.	JM8.N4	N-Wapl AID (3hrs)	micro-C	298M	670M

b

Pixel-by-pixel correlation between contact heatmaps from orthogonal studies

		ICE			ICE w/ distance corr.			HiCorr			DeepLoop		
		Nora	Kubo	Hsieh	Nora	Kubo	Hsieh	Nora	Kubo	Hsieh	Nora	Kubo	Hsieh
ΔCTCF control cells	Nora	1.0	0.10	0.01	1.0	0.02	0.02	1.0	0.32	0.41	1.0	0.63	0.50
	Kubo	0.10	1.0	0.14	0.02	1.0	0.08	0.32	1.0	0.33	0.63	1.0	0.69
	Hsieh	0.01	0.14	1.0	0.02	0.08	1.0	0.41	0.33	1.0	0.50	0.69	1.0
		Rhodes	Kriz	Hsieh	Rhodes	Kriz	Hsieh	Rhodes	Kriz	Hsieh	Rhodes	Kriz	Hsieh
ΔRad21 control cells	Rhodes	1.0	0.23	0.14	1.0	0.13	0.05	1.0	0.43	0.33	1.0	0.65	0.58
	Kriz	0.23	1.0	0.25	0.13	1.0	0.05	0.43	1.0	0.37	0.65	1.0	0.63
	Hsieh	0.14	0.25	1.0	0.05	0.05	1.0	0.33	0.37	1.0	0.58	0.63	1.0
		Liu	Kriz	Hsieh	Liu	Kriz	Hsieh	Liu	Kriz	Hsieh	Liu	Kriz	Hsieh
ΔWapl control cells	Liu	1.0	0.34	0.004	1.0	0.13	0.03	1.0	0.52	0.44	1.0	0.80	0.59
	Kriz	0.34	1.0	0.01	0.13	1.0	0.03	0.52	1.0	0.51	0.80	1.0	0.62
	Hsieh	0.004	0.01	1.0	0.03	0.03	1.0	0.44	0.51	1.0	0.59	0.62	1.0

R^2 0 1

c

Correlation between differential matrices (depletion - control)

ICE

	ΔWAPL (Liu)	ΔCTCF (Nora)	ΔWAPL (Kriz)	ΔCTCF (Kubo)	ΔRAD21 (Rhodes)	ΔRAD21 (Kriz)	ΔWAPL (Hsieh)	ΔRAD21 (Hsieh)	ΔCTCF (Hsieh)
ΔWAPL (Liu)	1.0	.02	.06	.00	.04	.04	.03	.08	.12
ΔCTCF (Nora)	.02	1.0	.09	.21	.01	−.06	−.07	.04	.02
ΔWAPL (Kriz)	.06	.09	1.0	.49	.18	.25	−.29	−.35	−.44
ΔCTCF (Kubo)	.00	.21	.49	1.0	.05	−.21	−.21	−.01	−.58
ΔRAD21 (Rhodes)	.04	.01	.18	.05	1.0	.18	.14	−.15	−.18
ΔRAD21 (Kriz)	.04	−.06	.25	−.21	.18	1.0	−.22	−.56	.01
ΔWAPL (Hsieh)	.03	−.07	−.29	−.21	.14	−.22	1.0	.37	−.01
ΔRAD21 (Hsieh)	.08	.04	−.35	−.01	−.15	−.56	.37	1.0	.25
ΔCTCF (Hsieh)	.12	.02	−.44	−.58	−.18	.01	−.01	.25	1.0

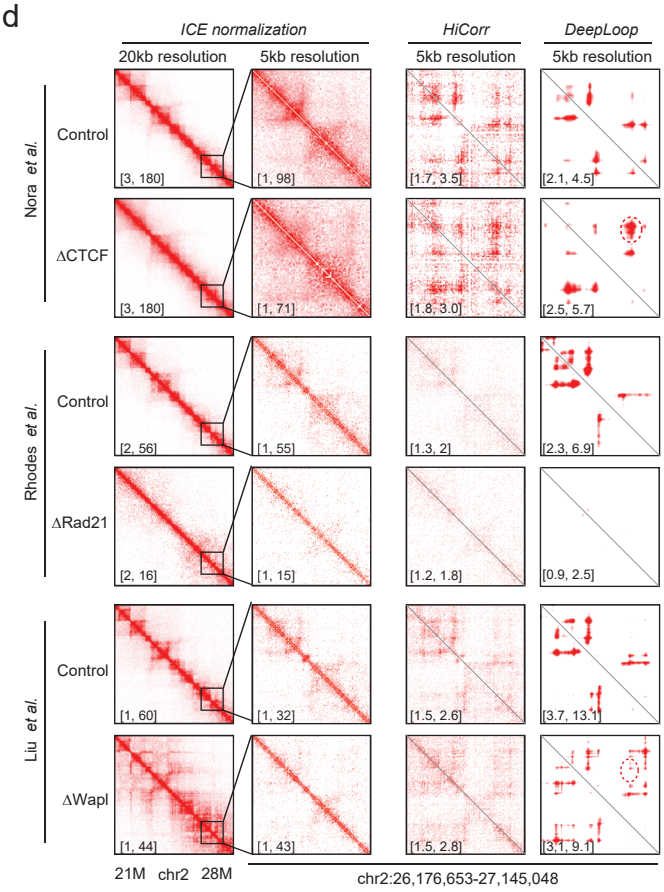
DeepLoop

	ΔCTCF (Hsieh)	ΔRAD21 (Hsieh)	ΔRAD21 (Kriz)	ΔRAD21 (Rhodes)	ΔCTCF (Kubo)	ΔCTCF (Nora)	ΔWAPL (Liu)	ΔWAPL (Kriz)	ΔWAPL (Hsieh)
ΔCTCF (Hsieh)	1.0	.79	.72	.68	.57	.62	.49	.27	−.20
ΔRAD21 (Hsieh)	.79	1.0	.74	.74	.53	.48	.49	.22	−.27
ΔRAD21 (Kriz)	.72	.74	1.0	.75	.49	.60	.44	.24	−.22
ΔRAD21 (Rhodes)	.68	.74	.75	1.0	.51	.55	.47	.19	−.33
ΔCTCF (Kubo)	.57	.53	.49	.51	1.0	.56	.51	.34	−.11
ΔCTCF (Nora)	.62	.48	.60	.55	.56	1.0	.39	.25	−.16
ΔWAPL (Liu)	.49	.49	.44	.47	.51	.39	1.0	.54	.08
ΔWAPL (Kriz)	.27	.22	.24	.19	.34	.25	.54	1.0	.34
ΔWAPL (Hsieh)	−.20	−.27	−.22	−.33	−.11	−.16	.08	.34	1.0

Spearman corr.

−1 −0.5 0 0.5 1

d



e

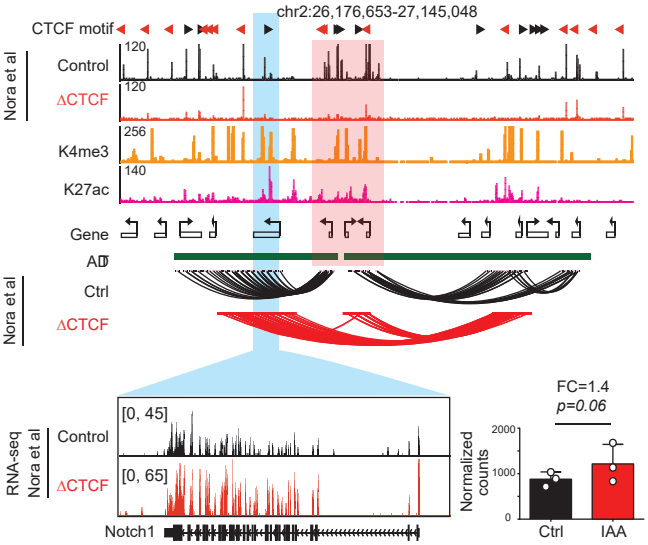


Figure 2

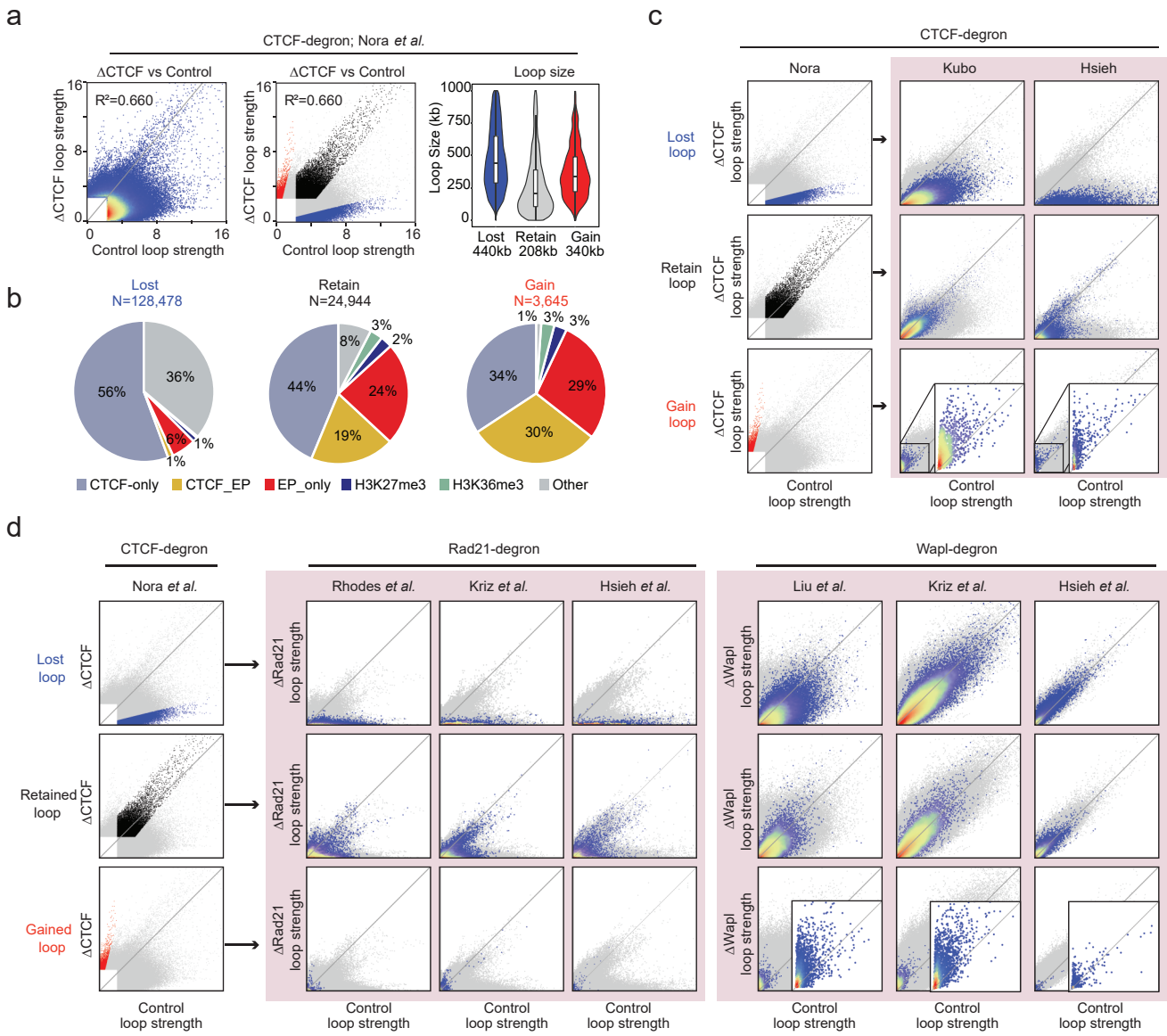

a
CTCF-degron; Nora et al.
ΔCTCF vs Control
R²=0.660
ΔCTCF loop strength
Control loop strength
Loop size
Loop Size (kb)
Lost 440kb
Retain 208kb
Gain 340kb
b
Lost N=128,478
Retain N=24,944
Gain N=3,645
56%
36%
6%
1%
1%
44%
8%
3%
2%
24%
19%
34%
1%
3%
3%
29%
30%
CTCF-only
CTCF_EP
EP_only
H3K27me3
H3K36me3
Other
c
CTCF-degron
Nora
Kubo
Hsieh
Lost loop
Retain loop
Gain loop
ΔCTCF loop strength
Control loop strength
d
CTCF-degron
Nora et al.
Lost loop
Retained loop
Gained loop
ΔCTCF
Rad21-degron
Rhodes et al.
Kriz et al.
Hsieh et al.
ΔRad21 loop strength
Wapl-degron
Liu et al.
Kriz et al.
Hsieh et al.
ΔWapl loop strength
Control loop strength

Figure 3

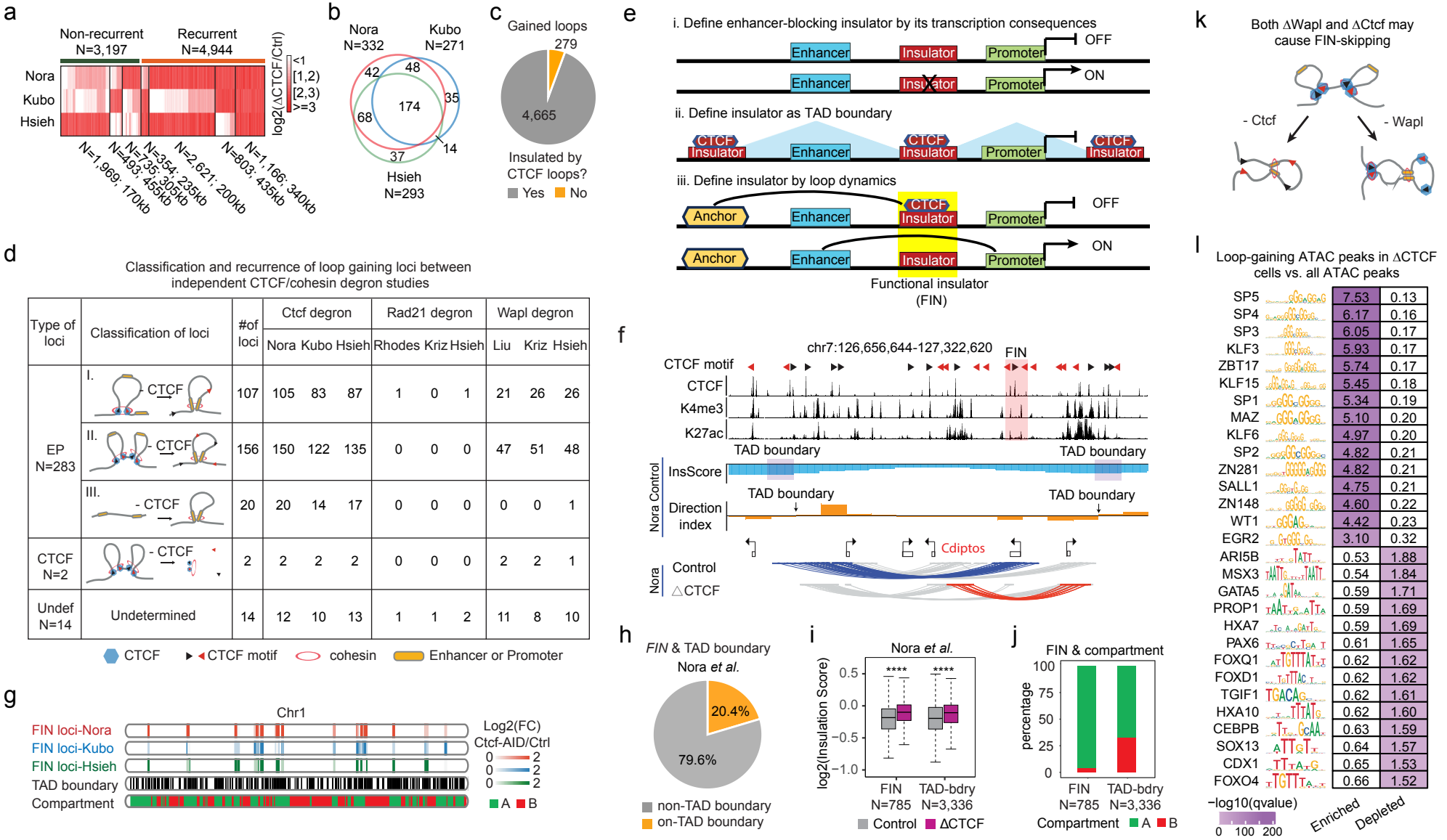

a
Non-recurrent N=3,197
Recurrent N=4,944
Nora
Kubo
Hsieh
log2(ΔCTCF/Ctrl)
<1
[1,2)
[2,3)
>=3
N=1,969; 170kb
N=493; 455kb
N=735; 305kb
N=354; 235kb
N=2,621; 200kb
N=803; 435kb
N=1,166; 340kb
b
Nora N=332
Kubo N=271
Hsieh N=293
42
48
174
35
68
37
14
c
Gained loops
279
4,665
Insulated by CTCF loops?
Yes
No
e
i. Define enhancer-blocking insulator by its transcription consequences
Enhancer
Insulator
Promoter
OFF
ON
ii. Define insulator as TAD boundary
CTCF Insulator
iii. Define insulator by loop dynamics
Anchor
Functional insulator (FIN)
k
Both ΔWapl and ΔCtcf may cause FIN-skipping
- Ctcf
- Wapl
d
Classification and recurrence of loop gaining loci between independent CTCF/cohesin degron studies
Type of loci
Classification of loci
#of loci
Ctcf degron
Rad21 degron
Wapl degron
Nora Kubo Hsieh
Rhodes Kriz Hsieh
Liu Kriz Hsieh
EP N=283
I. - CTCF 107 105 83 87 1 0 1 21 26 26
II. - CTCF 156 150 122 135 0 0 0 47 51 48
III. - CTCF 20 20 14 17 0 0 0 0 0 1
CTCF N=2
- CTCF 2 2 2 2 0 0 0 2 2 1
Undef N=14
Undetermined 14 12 10 13 1 1 2 11 8 10
CTCF
CTCF motif
cohesin
Enhancer or Promoter
f
chr7:126,656,644-127,322,620
FIN
CTCF motif
CTCF
K4me3
K27ac
TAD boundary
TAD boundary
Nora Control
InsScore
Direction index
TAD boundary
TAD boundary
Cdiptos
Nora
Control
△CTCF
g
Chr1
FIN loci-Nora
FIN loci-Kubo
FIN loci-Hsieh
TAD boundary
Compartment
Log2(FC) Ctcf-AID/Ctrl
A
B
h
FIN & TAD boundary
Nora et al.
20.4%
79.6%
non-TAD boundary
on-TAD boundary
i
Nora et al.
log2(Insulation Score)
****
****
FIN N=785
TAD-bdry N=3,336
Control
ΔCTCF
j
FIN & compartment
percentage
FIN N=785
TAD-bdry N=3,336
Compartment
A
B
l
Loop-gaining ATAC peaks in ΔCTCF cells vs. all ATAC peaks
SP5 7.53 0.13
SP4 6.17 0.16
SP3 6.05 0.17
KLF3 5.93 0.17
ZBT17 5.74 0.17
KLF15 5.45 0.18
SP1 5.34 0.19
MAZ 5.10 0.20
KLF6 4.97 0.20
SP2 4.82 0.21
ZN281 4.82 0.21
SALL1 4.75 0.21
ZN148 4.60 0.22
WT1 4.42 0.23
EGR2 3.10 0.32
ARI5B 0.53 1.88
MSX3 0.54 1.84
GATA5 0.59 1.71
PROP1 0.59 1.69
HXA7 0.59 1.69
PAX6 0.61 1.65
FOXQ1 0.62 1.62
FOXD1 0.62 1.62
TGIF1 0.62 1.61
HXA10 0.62 1.60
CEBPB 0.63 1.59
SOX13 0.64 1.57
CDX1 0.65 1.53
FOXO4 0.66 1.52
−log10(qvalue)
0 100 200
Enriched
Depleted

Figure 4

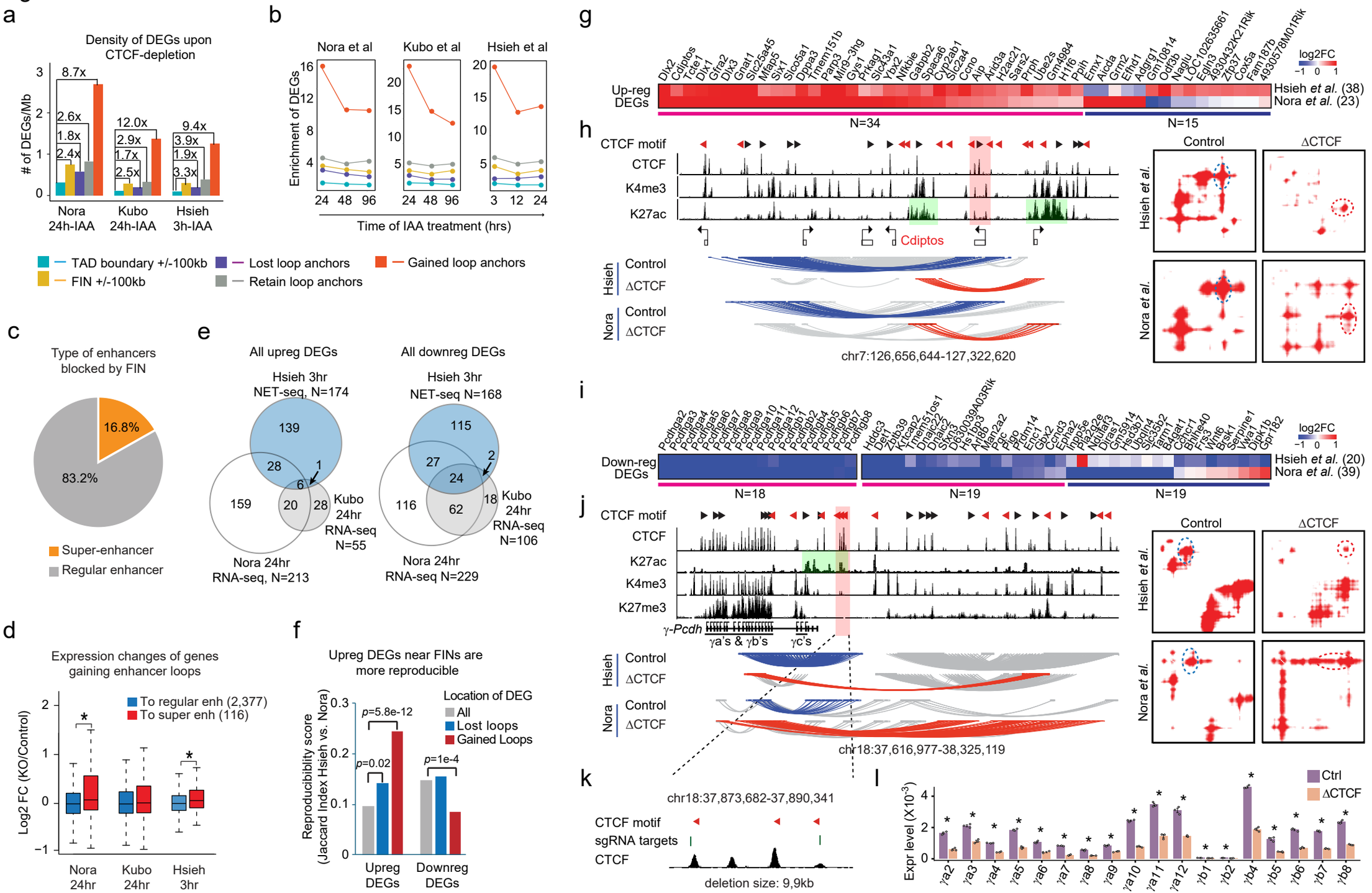
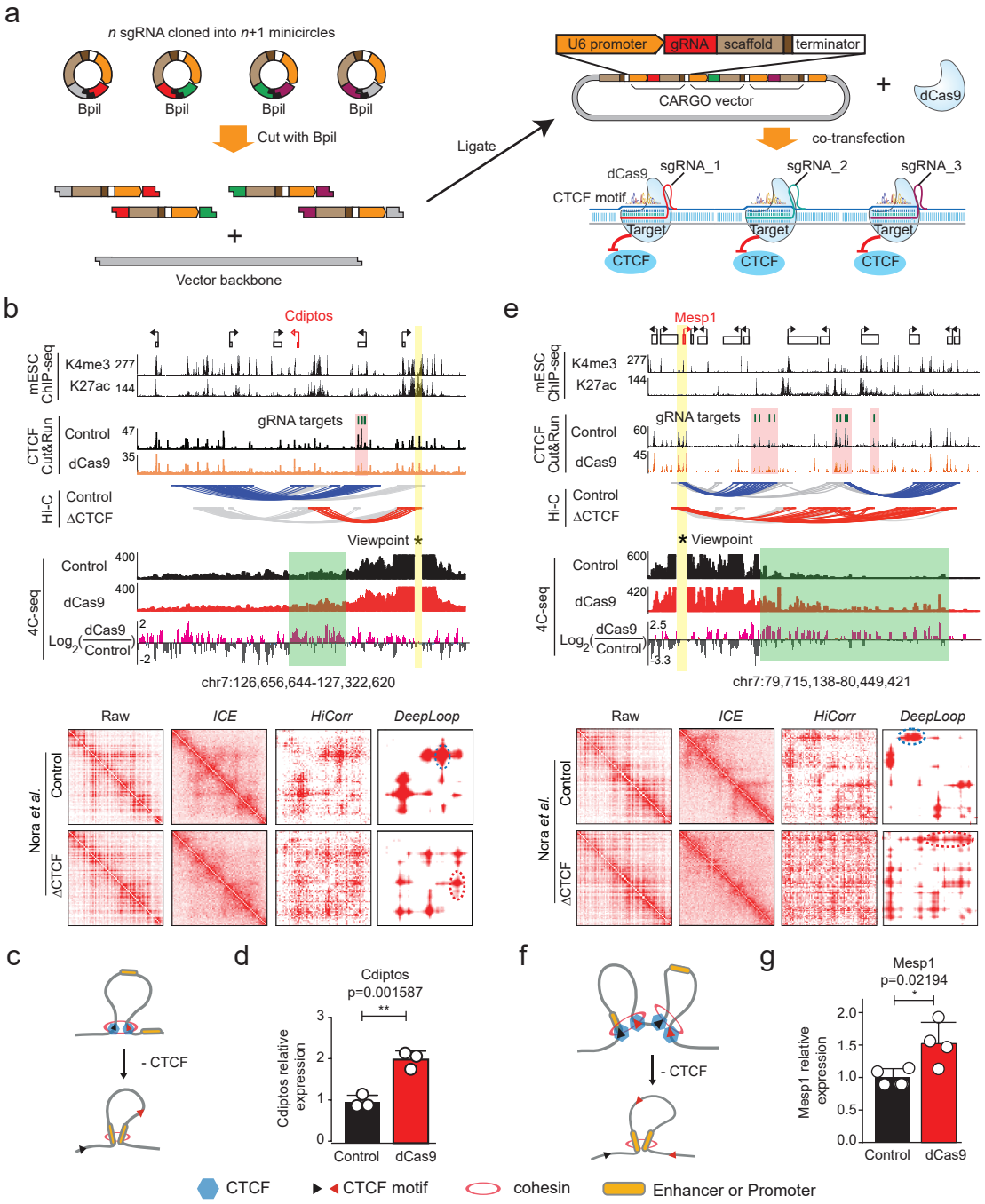

a
Density of DEGs upon CTCF-depletion
# of DEGs/Mb
8.7x
2.6x
1.8x
2.4x
12.0x
2.9x
1.7x
2.5x
9.4x
3.9x
1.9x
3.3x
Nora 24h-IAA
Kubo 24h-IAA
Hsieh 3h-IAA
b
Nora et al
Kubo et al
Hsieh et al
Enrichment of DEGs
Time of IAA treatment (hrs)
TAD boundary +/-100kb
FIN +/-100kb
Lost loop anchors
Retain loop anchors
Gained loop anchors
c
Type of enhancers blocked by FIN
16.8%
83.2%
Super-enhancer
Regular enhancer
e
All upreg DEGs
All downreg DEGs
Hsieh 3hr NET-seq, N=174
Hsieh 3hr NET-seq N=168
Kubo 24hr RNA-seq N=55
Kubo 24hr RNA-seq N=106
Nora 24hr RNA-seq, N=213
Nora 24hr RNA-seq N=229
d
Expression changes of genes gaining enhancer loops
To regular enh (2,377)
To super enh (116)
Log2 FC (KO/Control)
Nora 24hr
Kubo 24hr
Hsieh 3hr
f
Upreg DEGs near FINs are more reproducible
Location of DEG
All
Lost loops
Gained Loops
p=5.8e-12
p=0.02
p=1e-4
Reproducibility score (Jaccard Index Hsieh vs. Nora)
Upreg DEGs
Downreg DEGs
g
Up-reg DEGs
Hsieh et al. (38)
Nora et al. (23)
N=34
N=15
log2FC
h
CTCF motif
CTCF
K4me3
K27ac
Cdiptos
Control
ΔCTCF
Hsieh
Nora
chr7:126,656,644-127,322,620
i
Down-reg DEGs
Hsieh et al. (20)
Nora et al. (39)
N=18
N=19
N=19
j
CTCF motif
CTCF
K27ac
K4me3
K27me3
γ-Pcdh
γa's & γb's
γc's
chr18:37,616,977-38,325,119
k
chr18:37,873,682-37,890,341
CTCF motif
sgRNA targets
CTCF
deletion size: 9,9kb
l
Ctrl
ΔCTCF
Expr level (X10^-3)
γa2
γa3
γa4
γa5
γa6
γa7
γa8
γa9
γa10
γa11
γa12
γb1
γb2
γb4
γb5
γb6
γb7
γb8

Figure 5


a
n sgRNA cloned into n+1 minicircles
BpiI
Cut with BpiI
Ligate
Vector backbone
U6 promoter
gRNA
scaffold
terminator
CARGO vector
dCas9
co-transfection
sgRNA_1
sgRNA_2
sgRNA_3
CTCF motif
Target
CTCF
b
Cdiptos
mESC ChIP-seq
K4me3
K27ac
CTCF Cut&Run
Control
dCas9
gRNA targets
Hi-C
ΔCTCF
Viewpoint
4C-seq
Log2(dCas9/Control)
chr7:126,656,644-127,322,620
Raw
ICE
HiCorr
DeepLoop
Nora et al.
e
Mesp1
chr7:79,715,138-80,449,421
c
- CTCF
d
Cdiptos p=0.001587
Cdiptos relative expression
f
g
Mesp1 p=0.02194
Mesp1 relative expression
CTCF
CTCF motif
cohesin
Enhancer or Promoter