Map functional insulators with cross-platform Hi-C meta-analysis

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

[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, Case Western Reserve University, Cleveland, OH 44106, USA

[4]Department of Population and Quantitative Health Sciences, 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.

32 **Introduction**

33 It is generally accepted that TADs confine the communication between genes and enhancers for
34 transcriptional control(*1-6*). Most of the TADs are maintained by CTCF and cohesin, which are
35 also known as insulator proteins. Cohesin is a multimeric complex consisting of *SMC1A*, *SMC3*,
36 *RAD21* and one SA subunit (*STAG1* or *STAG2*). TAD forms when a DNA loop extrudes out of
37 the cohesin complex ring before blocked by CTCF binding sites in convergent orientation, and
38 cohesin serves as an enzymatic engine for this process(*7-15*). Deletion of CTCF or cohesin
39 core subunit RAD21 leads to near complete loss of TADs(*16, 17*). Cohesin function is also
40 regulated by cohesin loading complex MAU2/NIPBL(*18*) and cohesin release factor WAPL(*19-*
41 *21*): TADs disappear upon loss of NIPBL(*22, 23*), while loss of WAPL causes bigger loops
42 because cohesin can pass through CTCF blocks due to prolonged loop extrusion(*23, 24*).

43 Insulators are canonically defined as the DNA elements that inhibit gene expression when put
44 between enhancers and promoters(*25-27*). The discovery of thousands of CTCF-demarcated
45 TADs immediately suggests a model that CTCF sites at TAD boundaries are enhancer-blocking
46 insulators, which is supported by genetic experiments focusing on selected loci(*28-32*).
47 However, this model is challenged by a genome-wide Hi-C study in mESCs (Nora *et al.*, **Fig.**
48 **1a**) which only observed mild transcription changes despite near complete loss of TADs upon
49 CTCF-depletion(*16*). This conclusion is reproduced in two other genome-wide studies done with
50 *in situ* Hi-C and micro-C(*33, 34*) (**Fig. 1a**). Notably, Kubo *et al*. reported a role of CTCF in
51 mediating E-P interactions(*33*), and Hsieh *et al.* reported short-range cohesin-independent
52 cross-TAD E-P interactions(*34*). Taken together, although all these studies questioned the
53 insulating function of TAD boundaries, they did not address the fundamental question about the
54 whereabouts of the real enhancer-blocking insulators.

55 In this study we demonstrated that “functional insulators” (FINs) can be identified by examining
56 if their flanking regions gain E-P interactions upon CTCF depletion. Challenges exist to map
57 FINs with Hi-C-like genome-wide technologies: (i) most of the public Hi-C datasets do not have
58 billion-scale read depth usually required for loop-level analysis (**Fig. 1a**); (ii) E-P interactions are
59 generally weaker than CTCF loops in Hi-C and therefore harder to detect; (iii) the gain of E-P
60 loop signal can be subtle, requiring robust quantitation of loop strength; (iv) it is difficult to rule
61 out the possibility that some loop changes are secondary effects after prolonged CTCF or
62 cohesin depletion(*16, 34*). These challenges can be addressed by comparing independent
63 datasets, but existing CTCF-depletion data are from different Hi-C protocols, which perform very

differently in loop-level analyses with conventional pipelines(*35*). We previously developed *HiCorr* and *DeepLoop* pipelines to allow robust comparison of loop profiles from orthogonal Hi-C and micro-C platforms, even when sequencing depth is low(*36, 37*). Here we use *DeepLoop* to perform a meta-analysis of orthogonal Hi-C datasets to map recurrent DNA rewiring events upon depletion of CTCF and other cohesin proteins (**Fig. 1a**), thus generate a reliable genome-wide map of enhancer-blocking insulators.

Results

DeepLoop reveals insulated E-P loops from CTCF-degron Hi-C data

We collected nine pairs of independent Hi-C or micro-C datasets from six publications in which CTCF(*16, 33, 34*), RAD21(*34, 38, 39*), or WAPL(*34, 38, 40*) were acutely depleted in three different mESC lines. The number of contacts in these Hi-C/micro-C datasets are highly variable ranging from 43 million to ~1 billion. For every target protein, the choice of degron methods (AID or dTAG) and the timespan of protein degradation (range from 3hr to 3 days) are also variable (**Fig. 1a**). As shown in the examples in **Fig. 1b**, for Hi-C and *in situ* Hi-C studies, conventional ICE-normalization only support compartment- or TAD-level analysis at 20kb or lower resolution due to the lack of read depth. In contrast, *HiCorr*-based bias correction(*36*) followed by *DeepLoop* enhancement(*37*) reveals detailed loop profiles at 5kb resolution (**Fig. 1b**). From this example, we observed that the enhanced loop profiles in wildtype mESCs from independent studies are rather similar (**Fig. 1b**, **Supplementary Fig. 1**), supporting the robustness of *DeepLoop* in cross-platform analysis regardless of read depth(*37*). On the other hand, loss of *CTCF*, *Rad21*, and *Wapl* leads to distinct loop profiles (**Fig. 1b**).

We firstly focused on the CTCF-degron Hi-C data from Nora *et al*. In the example of **Fig. 1b** (top two rows), with conventional ICE-normalization, we observed dampened TAD structure at 20kb resolution as the original study reported(*16*), but at 5kb resolution, loop signal from ICE-normalized heatmaps are weak. From the *DeepLoop* enhanced heatmaps, we can clearly observe that multiple CTCF loops are substantially weakened upon CTCF depletion. However, the most notable change in this region is the gain of a long-range loop (**Fig. 1b**, second row, loop pixels in dashed circle) connecting *Notch1* gene to a distal enhancer region with multiple H3K27ac peaks (**Fig. 1c**). Consistent with gain of E-P loop, *Notch1* is upregulated upon CTCF-

depletion (**Fig. 1c**). The gain of loop signal at this locus is also observable in *Wapl*-depleted cells (**Fig. 1b**, bottom two rows) and two other CTCF-degron studies (**Supp. Fig. 1**), but not in the *Rad21*-depleted cells (**Fig. 1b**, 3rd and 4th rows). These observations suggest that the two CTCF loop anchors at the TAD boundary region (**Fig. 1c**, highlighted in pink) are blocking the *Notch1*-enhancer interaction. This example also demonstrated that *DeepLoop* may detect recurrent insulating events from CTCF-degron Hi-C data without the need for ultra-high sequencing depth.

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

To perform a genome-wide analysis of insulating events, we call 300K~500K loop pixels from each CTCF-degron dataset and use the pixel intensity in *DeepLoop* heatmaps to measure the loop strength (**Fig. 1d, Supp. Fig. 2**). Notably, *DeepLoop* outputs are enhanced from ratio (obs/exp) heatmaps instead of normalized contact frequency used by ICE/KR. We previously showed that this approach promises cross-platform convergence between orthogonal Hi-C and micro-C data(*37*).

In all CTCF-degron studies, the scatter plots are clearly down-skewed reflecting a massive loss of chromatin loops upon CTCF deletion (**Fig. 1d, Supp. Fig. 2a, d**, left panels). Using a simple fold-change cutoff, we called loop pixels that lose strength ("lost"), retain strength ("retained"), or gain strength ("gained") from each dataset (**Fig. 1d, Supp. Fig. 2a, d**, middle panels). We further classified the loop pixels based on their epigenetic marks and found that CTCF loops dominate the "lost" pixels, while the "retained" and "gained" pixels are enriched with enhancer or promoter marks (**Fig. 1e, Supp Fig. 2b, e**). Interestingly, in all three studies the "gained" loops are much bigger than the "retained" loops (**Fig. 1d, Supp. Fig. 2a, d**, middle panels), suggesting that these "gained" E-P loops form through loop extrusion (more discussions below).

We also examine the robustness of the "lost", "retained", and "gained" pixels across independent Hi-C/micro-C studies. As expected, the "lost" pixels from any study generally also lose strength in the other two studies (**Fig. 1f, Supp Fig. 2c,f**, first row). However, the consistency of "retained" loop pixels is study specific: (i) although the "retained" loop pixels from Nora *et al.* are mostly also "retained" in Kubo *et al.*, some of them lose strength in Hsieh *et al.* (**Fig. 1f**, second row); (ii) most of the "retained" loop pixels from Kubo *et al*. lose strength in the other two studies (**Supp. Fig. 2c**, second row); (iii) the "retained" loops from Hsieh *et al.* remain "retained" in both Nora *et al.* and Kubo *et al.* (**Supp. Fig. 2f**, second row). Since incomplete

125 CTCF depletion causes the leftover CTCF loops(*16*), these observations indicate that CTCF
126 depletion in Kubo *et al.* is less complete than the other two studies; Hsieh *et al.* achieves the
127 best CTCF depletion, perhaps because it uses N-terminal degron system (**Fig. 1a**).

128 Interestingly, the "gained" pixels from any one of the three studies generally also gain strength
129 in the other two studies (**Fig. 1f, Supp Fig. 2c,f,** third row). The only exception is that some
130 "gained" loop pixels from Kubo *et al*. lose strength in Hsieh *et al*. (**Supp Fig. 2c**). Again, this is
131 likely because Kubo *et al.* retains more CTCF loops due to incomplete CTCF degradation.
132 Nevertheless, the "gained" pixels from Nora *et al*. and Hsieh *et al*. do gain strength in Kubo *et al.*
133 (**Fig. 1f, Supp Fig. 2f**), supporting the consistency between the three studies. Notably, E-P loop
134 gaining events are largely neglected in the original Hi-C/micro-C studies because the changes
135 are often too subtle to distinguish from background with conventional analytical tool (*e.g.*, **Fig.**
136 **1b**). Our meta-analyses prove that although there are only a small number of newly gained E-P
137 loops, they are reliable consequences of CTCF-depletion; affected genes such as *Notch1* in
138 **Fig. 1** are likely to be the direct targets of insulators.

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

140 We also performed meta-analyses of the *Rad21*- and *Wapl*-depletion Hi-C datasets. Ablation of
141 *Rad21* also causes massive loss of CTCF loops, while the "retained" and "gained" loops are
142 also enriched with E-P loops (**Fig. 2a-b, Supp. Fig. 3**). However, there are important
143 differences between *Rad21*- and *CTCF*-depleted cells. (i) The "retained" and "gained" loop
144 pixels in *Rad21*-depleted cells are much smaller (mostly within 100kb) than those in *CTCF*-
145 depleted cells, indicating that cohesin-independent loops are short-range (**Fig. 2a, Supp. Fig.**
146 **3a, d**, right panels). (ii) Unlike in CTCF-depleted cells, the "gained" loop pixels are NOT
147 recurrent between the three *Rad21*-degon datasets (**Fig. 2c, Supp. Fig. 3c, f**, last row). (iii) The
148 "retained" and "gained" loop pixels in *Rad21*-depleted cells are more enriched with H3K36me3
149 mark, suggesting transcription associated chromatin interactions (**Fig. 2b, Supp. Fig. 3b, e, h**).
150 (iv) The "gained" loops from Rhodes *et al*. are enriched with H3K27me3 mark and tends to be
151 longer (**Supp. Fig. 3a, b**, right panels), which is consistent with the conclusion from the original
152 study(*39*). However, this phenomenon is specific to Rhodes *et al.* and it is not observed in the
153 other two *Rad21*-degron studies.

154 Conversely, ablation of cohesin unloading factor *Wapl* generates up-skewed scatterplots in all
155 three studies (**Fig. 2d, Supp. Fig. 4a, d**). The skewness of Hsieh *et al.* is less apparent,

156 probably because Hsieh *et al.* performed *Wapl* depletion for much shorter time (3hrs) than Liu *et*
157 *al*. (24hrs) and Kriz *et al*. (72hrs) (**Fig. 1a**). In all *Wapl*-degron studies, the "lost", "retained", and
158 "gained" loops are almost equally enriched with CTCF and E-P loops (**Fig. 2e, Supp. Fig. 4b,**
159 **e**). Consistent with the model that *Wapl*-depletion stabilizes chromatin-bound cohesin and
160 promote loop extrusion(*23, 24*), the "gained" loops are significantly bigger than "lost" and
161 "retained" loops in all three studies (**Fig. 2d, Supp. Fig. 4a, d**, right panels). The "lost",
162 "retained", and "gained" loop pixels are generally consistent between studies (**Fig. 2f, Supp.**
163 **Fig. 4c,f**) despite some discrepancy for "lost" loop pixels from Hsieh *et al.* (**Supp. Fig. 4f**), again
164 likely because Hsieh *et al.* only depleted *Wapl* for a short period of time.

165 We next take the "lost", "retained", and "gained" loop pixels from *CTCF*-degron studies and
166 examine how they change in *Rad21*- and *Wapl*-degron data. Firstly, nearly all *CTCF*-dependent
167 loops ("lost" in *CTCF*-depleted cells) are also *Rad21*-dependent (**Fig. 2g, Supp Fig. 5a,b**, first
168 row). Secondly, the *CTCF*-independent loops ("retained" in CTCF-depleted cells) include both
169 cohesin-dependent and cohesin-independent fractions, since some of them lose strength in
170 *Rad21*-degron data but some do not (**Fig. 2g, Supp Fig. 5a, b**, second row). More of the CTCF-
171 independent loops from Hsieh *et al.* are also cohesin independent (**Supp Fig. 5b**, second row),
172 again suggesting that Hsieh *et al.* achieves more complete CTCF-depletion. Finally, the
173 "gained" loops in CTCF-depleted cells do not gain strength in *Rad21*-depleted cells (**Fig. 2g,**
174 **Supp Fig. 5a, b**, third row). These results strongly indicate that the recurrently gained E-P
175 interactions upon CTCF depletion are created by loop extrusion.

176 Interestingly, the "gained" loops in CTCF-depleted cells have apparent upward biases in the
177 scatter plots from *Wapl*-degron data (**Fig. 2h, Supp Fig. 5b, d**, third row). These results are
178 consistent with the example in **Fig. 1b** and suggest that *Wapl*-depletion can perturb insulator
179 function by releasing short-range CTCF loops and allows insulated E-P interactions to skip
180 insulators (more analysis below).

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

182 Among the total 8,141 "gained" loop pixels from all three CTCF-degron Hi-C/micro-C datasets,
183 4,944 pixels are recurrent in at least two studies with a median size ~250kb (**Fig. 3a**). In each
184 dataset, the "gained" loop pixels are found in ~300 loci and there are substantial overlaps
185 between independent studies: 174 loop-gaining loci are shared by all three studies and 304 are
186 shared by at least two studies (**Fig. 3b**). We hypothesize that these recurrent E-P loop-gaining

187 loci are caused by loss of CTCF insulation; indeed, 4,665 out of the 4,944 loop-gaining pixels
188 (94%) are insulated by at least one CTCF loop in wildtype cells (**Fig. 3c**). We further curated the
189 reproducible loop-gaining loci based on the types and numbers of CREs involved (**Fig. 3d**).
190 Nearly all the curated loci (283 of 299) gain E-P loops (2 loci gain CTCF contacts and the others
191 cannot be clearly classified). We further classified the E-P loop-gaining loci into three categories
192 (**Fig. 3d**). In *Type I*, one of the two E/P sites is insulated in a CTCF loop while the other site is
193 not; *Type II* involves at least two CTCF loops insulating two E/P-sites. In contrast, the gained E-
194 P loops in *Type III* do not have insulating CTCF loops. *Type I* and *II* represents most of the E-P
195 loop gaining loci (263 out of 283), while only 20 loci belong to *Type III*, again suggesting that
196 most of the loop gaining loci are due to the removal of CTCF insulation.

197 We therefore propose to define "functional insulators" (FINs) as the CTCF loop anchors that are
198 actively blocking the interactions between flanking *cis*-regulatory elements (CREs) (model *iii* in
199 **Fig. 3e**). The definition of FIN is conceptually analogous to the conventional method to define
200 enhancer-blocking insulators by transcriptional effects (model *i* in **Fig. 3e**), but distinct from the
201 TAD-based definition of insulators (model *ii* in **Fig. 3e**). There are clearly much fewer FINs than
202 TAD boundaries (**Fig. 3f**): only 13% of the FINs overlap TAD boundaries (**Fig. 3g**). These
203 observations indicate that most FINs are inside TADs and conversely, loss of TAD boundaries
204 usually does not cause loop formation across their flanking regions. We also found that most
205 (96%) of the FIN loci are within compartment A (**Fig. 3h**), indicating that active E-P blocking
206 mainly happens within euchromatin region.

207 Cohesin degron Hi-C data further support a cohesin-dependent mechanism of FINs. We found
208 that almost none of the curated loop-gaining events appears in the *Rad21*-degron cells (**Fig.
209 3d**), indicating cohesin dependency. Interestingly, a significant fraction of the *Type I* and *II*, but
210 not the *Type III* "gain of E-P loop" loci, can be also observed in *Wapl*-degron datasets (**Fig. 3d**).
211 This phenomenon can be explained by previous findings that *Wapl* deletion allows loop
212 extrusion to skip CTCF blocks(*23, 24*). As shown in the carton in **Fig. 3i**, because *Wapl*-
213 depletion stabilizes chromatin-bound cohesin and leads to longer loop extrusion, previously
214 insulated enhancers and promoters can interact inside the bigger CTCF loop due to "insulator
215 skipping", creating the same E-P loop as observed in CTCF-depleted cells. Taken together,
216 these results support FINs as *bona fide* insulators of cohesin-mediated loop extrusion.

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

218 We found that in all CTCF-degron studies FIN proximal regions (+/- 100kb) are 2~3 fold more
219 enriched with DEGs than TAD-boundary proximal regions, and the "gained" loop anchors are 10
220 times more enriched with DEGs (**Fig. 4a**). Furthermore, early response DEGs are clearly more
221 enriched at gained loop anchors (**Fig. 4b**), suggesting that FIN-blocked enhancers and
222 promoters are the direct targets of insulators. 16.8% of FINs (49 out of 299) are blocking 43
223 super-enhancers (out of 231, or 19%) (**Fig. 4c**). As expected, genes gaining super-enhancer
224 interactions are more likely to be upregulated than genes gaining interactions with regular
225 enhancers (**Fig. 4d**).

226 Using common cutoffs (2-fold for RNA-seq and 1.5-fold for NET-seq, $p<0.05$), each of the
227 CTCF-degron studies identified a few hundred early response DEGs, and the reproducibility
228 between independent studies is limited due to complex reasons (**Fig. 4e**). Using NET-seq,
229 Hsieh *et al*. identified more DEGs than RNA-seq despite shorter time of CTCF depletion (3hr),
230 but it remains possible that some DEGs are caused by indirect mechanisms. Among the two
231 RNA-seq studies, Kubo *et al.* found fewer DEGs than Nora *et al.*, and most of the DEGs from
232 Kubo et al. are also observed in Nora *et al*. (**Fig. 4e**), consistent our conclusion above that the
233 CTCF-depletion in Kubo *et al.* is less complete. We therefore focused on the reproducible DEGs
234 between Hsieh *et al.* and Nora *et al.* as they represent a more reliable set of CTCF direct
235 targets. Interestingly, although downregulated DEGs are more reproducible than upregulated
236 DEGs between the two studies (**Fig. 4f**, grey bars), near FINs (at the anchors of gained E-P
237 loops) the upregulated DEGs are significantly more reproducible than downregulated DEGs
238 (**Fig. 4f**, red bars). This observation supports transcription activation as the primary
239 consequence of insulator loss, consistent with the convention concept that insulators block
240 enhancer-mediated gene activation.

241 In **Fig. 4g** we plotted 49 upregulated DEGs at FIN loci as direct targets of insulators (from Hsieh
242 *et al.* and Nora *et al.*), among them 34 are considered as "high-confidence" targets supported by
243 both studies. **Fig. 4h** shows one representative example, in which we observed gain of loop
244 between two enhancer clusters upon CTCF depletion, leading to the upregulation of *Cdiptos*
245 gene. Similarly, we also plotted 37 reproducibly down-regulated DEGs near FINs supported by
246 both studies (**Fig. 4i**). Strikingly, 18 of the recurrent down-regulated DEGs are from
247 protocadherin-γ, a well-known insulator loci (**Fig. 4i,j**). Protocadherin proteins play key roles in
248 the formation of synaptic connections between neurons and the *Pcdh-*γ genes are distinguished
249 by their variable promoters (exon1): γa's and γb's are stochastically expressed in individual

250 neurons and the γ*c*'s are constitutively expressed(*41-44*). In mESCs, the promoters of γ*a*'s and
251 γ*b*'s are occupied by CTCF, H3K4me3, and the repressive H3K27me3 mark (**Fig. 4j**). An
252 enhancer cluster marked by H3K27ac is present at the 3' of γ*c* promoters (**Fig. 4j**, highlighted in
253 green) forming loop interactions with the promoters of γ*a*'s and γ*b*'s, which are restricted by a
254 FIN with multiple CTCF peaks (**Fig. 4j**, blue arches). Upon CTCF depletion, the enhancer
255 interactions are weakened, and the *Pcdh*-γ promoter loops extend into a distal enhancer-poor
256 region marked by H3K27me3, which explains the downregulation of γ*a* and γ*b* genes. Taken
257 together, these results demonstrated the transcription consequences of FINs and mapped their
258 direct targets in mESCs.

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

260 To validate the *cis*-regulatory functions of FINs, we need to verify if the gained loops are truly
261 due to the loss of insulation. Because a FIN often involves multiple CTCF binding sites that
262 contribute to the insulating function collectively, we need to simultaneous perturb multiple CTCF
263 sites in a relatively big region. To achieve this, we combine a multi-sgRNA delivering system
264 name CARGO(*45*) with a dCas9-based CTCF-displacement method(*43, 46*) to perturb multiple
265 CTCF binding sites in the same cells. In this approach, 4-10 sgRNAs exactly overlapping the
266 cognate motifs of multiple CTCF binding sites are expressed from the same CARGO plasmid.
267 When co-transfected with a dCas9 construct, the sgRNAs can displace CTCF from genome at
268 multiple target sites simultaneously (**Fig. 5a-b**).

269 We used this approach to validate the gain-of-loop events at two FIN loci. The first FIN is close
270 to *Cdiptos*, a recurrent upregulated DEG (**Fig. 5c**). This example belongs to a simpler type of
271 FIN in which the newly gained loop (**Fig. 5c**, red arches) is blocked by only one FIN (**Fig. 5d**).
272 We used one CARGO construct targeting 4 CTCF motifs and confirmed with CUT&RUN that we
273 successfully weakened the target CTCF peaks but did not affect other CTCF peaks in this
274 region (**Fig. 5c**, highlighted in pink). We performed 4C-seq and confirmed the gain of long-range
275 chromatin interactions consistent with Hi-C data (**Fig. 5c**, highlighted in green). We also
276 validated with RT-qPCR that *Cdiptos* is upregulated upon displacement of target CTCF binding
277 sites (**Fig. 5e**). The second locus is more complex as the newly gained loop is blocked by more
278 than two FINs (**Fig. 5f,g**). For this locus we used one CARGO construct to target 9 CTCF motifs
279 and successfully weakened the target CTCF peaks (**Fig. 5f**, highlighted in pink). Again, we
280 confirmed the gain of loop interactions with 4C-seq (**Fig. 5f**, highlighted in green), and the

281 upregulation of *Mesp1* gene since the new loops bring distal enhancers to its proximity. Taken
282 together, these experiments verified that FINs execute their insulating functions in *cis*.

283

284 **Discussion**

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

292 The concept of FIN is distinct from the common assumption that TAD boundaries are insulators.
293 In fact, FINs are usually inside the TAD, and most TAD boundaries do not gain cross-boundary
294 loops upon CTCF depletion. Importantly, there are only a few hundred FINs in mESCs, which
295 readily explains why CTCF-depletion only cause mild changes of transcriptome. It should be
296 also noted that the gained E-P loops do not always predict significant gene activation due to
297 complex reasons(*47*). Furthermore, target gene expression can be quantitative tuned by
298 multiple enhancers in an additively fashion, and each single enhancer may only have a small
299 effect size(*48*). Nevertheless, we showed that FINs better explain DEGs, especially the
300 recurrent upregulated DEGs in CTCF-depleted cells. This is consistent with the classic definition
301 of insulators as DNA elements that repress enhancer-mediated gene activation.

302 It remains unclear why only a small subset of CTCF loop anchors are FINs, *i.e.*, why is gain-of-
303 loop only observed at some loci but not others? Our analysis provided some hints. Firstly, meta-
304 analysis with *Rad21*- and *Wapl*-degron Hi-C data shows that the newly gained E-P loops in
305 CTCF-depleted cells are cohesin-dependent and governed by cohesin regulators. It is therefore
306 plausible that local cohesin activity is a key determinant. Secondly, the fact that almost all FINs
307 are in compartment A (euchromatin) is also consistent with high cohesin activity because
308 cohesin and its loading factors are known to co-localize with promoters and enhancers(*49*).
309 Finally, since genome compartmentalization is a highly dynamic 3D genome feature, we
310 speculate that FINs are also likely to be cell type- or cell state-specific. It will be interesting to
311 further explore the context-specific roles of FINs in health and diseases.

312

313

314 **Figure Legend**

315 **Figure 1. *DeepLoop* reveals recurrent gain of E-P loops from independent CTCF-degron**
316 **Hi-C datasets.**

317 **a,** A list of published Hi-C and micro-C studies with acute depletion of *Ctcf*, *Rad21*, or *Wapl* in
318 mESCs. **b,** Contact heatmaps of the CTCF-, Rad21-, and Wapl-degron Hi-C data at *Notch1*
319 locus with *ICE*, *HiCorr* and *DeepLoop*. Red dashed circle: gained loops. **c**, Genome browser
320 tracks showing ChIP-seq data and Hi-C loops from Nora *et al.* at *Notch1* locus. Blue highlighted
321 region showing Notch1 gene which is at the anchor of a gained loop. Pink region highlights the
322 potential insulator region that blocks the formation of *Notch1* E-P loop. RNA-seq results are also
323 shown for *Notch1* gene in WT (black) and CTCF-depleted cells (red). Bar plot: quantitation of
324 *Notch1* expression from RNA-seq. P-value: one-side Wald test provided by DE-seq2. **d-f,** Meta-
325 analysis of independent CTCF-degron Hi-C datasets. **d**, Left: scatterplot comparing the strength
326 of loop pixels between WT and CTCF-depleted mESCs; middle: the same scatterplot as left but
327 highlight the lost (blue), retained (black), and gained (red) loop pixels; right: violin plot
328 comparing the size of loop pixels. **e,** Pie charts showing the components of lost, retained and
329 gained loop pixels based on epigenetic marks. **f**, Scatter plots showing the consistency of lost,
330 retained and gained loop pixels between three CTCF degron studies. Plots in the left column
331 highlight the lost, retained, and gained loop pixels from Nora *et al*., the right two columns show
332 the quantitative changes of the corresponding pixels in *Kubo et al.* or *Hsieh et al*.

333 **Figure 2. Meta-analysis of CTCF- and cohesin-degron Hi-C datasets reveals the fate and**
334 **mechanisms of dynamic loops.**

335 **a-c,** Meta-analysis of independent *Rad21* degron Hi-C datasets. **a,** Left and middle panels
336 define the lost, retained, and gained loop pixels upon *Rad21*-depletion from the study by Kriz *et*
337 *al.*, and the right panel compares their loop size distribution. **b,** Pie charts showing the
338 components of lost, retained and gained loops based on epigenetic marks. **c,** Scatter plots
339 testing if the lost, retained and gained loop pixels from Kriz et al. are consistent in the other two
340 *Rad21* degron studies. **d-e**, The same meta-analysis of three independent *Wapl*-degron Hi-C
341 datasets. **g,** The lost, retained, and gained loop pixels from one CTCF-degron Hi-C data (Nora

342 *et al.*) are plotted onto Hi-C scatter plots from *Rad21*- or *Wapl*-degron studies to examine their
343 quantitative changes.

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

345 **a,** Heatmap showing the union of all gained loop pixels from three CTCF-degron Hi-C datasets.
346 Every row represents one CTCF-degron study and every column represents one loop pixel.
347 Loop pixels are organized into clusters based on their reproducibility. The number of loop pixels
348 in each cluster and their median size are also shown. Color scale indicates the increase of loop
349 signal in log scale. **b**, Venndiagram showing the consistence of *FIN* loci identified from three
350 datasets. **c,** Pie chart showing that most of the gained loops in CTCF-depleted cells have at
351 least one CTCF loop anchor in between in wildtype cells. **d,** A table summarizing the
352 classification and recurrence of loop gaining loci between independent CTCF- or cohesin-
353 degron studies. **e,** Cartoons summarizing three possible ways to define insulators. **f,** Ideogram
354 showing the locations of FIN loci from three studies, TAD boundaries, and compartments on
355 chromosome 1. FIN from three different CTCF-degron studies are shown in different color, and
356 the color scale shows the fold increase of gained loop in log scale. **g,** Pie chart showing the
357 percentage of FINs that overlap TAD boundary. **h,** Stack bar plot comparing the proportion of
358 FINs and TAD boundaries in compartment A or B. **i,** A model to explain why some of the newly
359 gained E-P loops in CTCF-depleted cells are also observed in Wapl-depleted cells.

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

362 **a,** Bar plots comparing the density of DEGs upon CTCF-depletion at different type of genomic
363 regions. y-axis indicates the number of DEGs per Mb. **b,** Enrichment of DEGs at different types
364 of genomic regions compared to whole genome as background. X-axis: time of IAA treatment
365 (hrs) in each study; y-axis: fold enrichment. Each panel represent one CTCF-degron study. **c,**
366 Left: pie chart showing the percentage of FINs that are blocking super enhancers; Right: pie
367 chart showing the percentage of super enhancers that are blocked by *FIN*. **d,** Boxplot showing
368 the expression changes of genes gaining enhancer loop. Red: genes gaining super-enhancer
369 loop; Blue: genes gaining regular enhancer loop. *, $p < 0.05$, Wilcoxon signed-rank test. **e,** Venn
370 diagrams showing the overlap of all up- (left panel) and down-regulated (right panel) DEGs
371 between three CTCF-degron studies. **f,** Bar plot comparing the reproducibility (measured by
372 Jaccard Index) of up- or down-regulated DEGs between two CTCF-degron studies (Hsieh *et al.*

373 and Nora *et al.*). Gray bars: all DEGs; blue bars: DEGs at the anchors of lost loops; red bars:
374 DEGs at the anchors of gained loops. P-values are from binomial test. **g,** Heatmap showing the
375 union of upregulated DEGs at the anchors of gained loops from *Hsieh et al* and *Nora et al*.;
376 DEGs are clustered based on consistency. **h,** Left panel shows the browser tracks near *Cdiptos*
377 locus and right panel shows the Hi-C heatmaps in the same region from *Hsieh et al* and *Nora et*
378 *al.* The direction of CTCF motifs are indicated by triangles. Blue arches (left) and blue dashed
379 circles (right): lost loop pixels; red arches and red dashed circles: gained loop. Regions
380 highlighted in green: enhancers that gained loops; the region highlighted in pink: a CTCF loop
381 anchor is defined as *FIN*. **i,** Similar as **g**, a heatmap showing the union of down-regulated DEGs
382 at the anchors of gained loops from *Hsieh et al.* and *Nora et al*. Among these genes 18 genes
383 are from the Pcdh-γ locus (shown as a separate heatmap). **j,** Left panel shows the browser
384 tracks near Pcdh-γ locus and right panel shows the Hi-C heatmaps from two CTCF-degron
385 studies.

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

387 **a,** Scheme of the CARGO vector that can express multiple sgRNAs. In this system, the
388 sequences of *n* sgRNAs are broken into left and right halves and inserted into (*n*+1) minicircles.
389 The *i*-th minicircle contains the right half of the (*i-1*)-th sgRNA and the left half of the *i*-th sgRNA
390 sequences divided by a *BpiI* restriction site. Final CARGO vector can be assembled by ligating
391 the re-linearized minicircles and vector backbone; each sgRNA is driven by a U6 promoter. **b,**
392 Co-transfect CARGO vector and a dCas9-expressing plasmid can displace multiple CTCF
393 binding sites in the same cell. **c,** Validation of a FIN at the *Cdiptos* locus. From the top, 1st and
394 2nd tracks are H3K4me3 and H3K27ac ChIP-seq data showing the promoter and enhancers in
395 mESC. 3rd and 4th tracks show CTCF ChIP-seq data before and after targeting 4 CTCF motifs in
396 FIN (highlighted in pink) with CARGO. 5th and 6th are Hi-C arches (*DeepLoop* pixels with data
397 from Nora *et al.*) before and after CTCF-depletion (blue: lost loops; red: gained loops). The
398 bottom 3 tracks show 4C-seq data (bait highlighted in yellow) before and after targeting CTCF
399 sites; regions highlighted in green gain 4C signal. **d,** A model cartoon showing that the gained
400 loop at this FIN locus only involves one CTCF loop. **e,** RT-qPCR results showing that *Cdiptos* is
401 upregulated after CTCF displacement. ** $p<0.01$; one-side t-test. **f-h,** Validation of a FIN at
402 *Mesp1* locus similar as shown **c-e**. Note the gained loop at this FIN locus involves at least two
403 CTCF loops (**g**).
404

405 **Supplementary figure legend**

406 **Supplementary Figure 1. Recurrent gain of E-P loop at *Notch1* locus between**
407 **independent CTCF-degron Hi-C studies.**

408 **a,** Contact heatmaps of the CTCF-degron Hi-C data by Kubo *et al.* at *Notch1* locus with *ICE*,
409 *HiCorr* and *DeepLoop*. Red dashed circle highlights the same locations where we observed a
410 gained E-P loop in the contact heatmap from Nora *et al.* (shown in **Fig. 1b**). **b**, Genome browser
411 tracks showing ChIP-seq data and Hi-C loops from Kubo *et al.* Blue highlighted region indicates
412 *Notch1* gene. Pink region highlights the potential insulator region that blocks the formation of
413 *Notch1* E-P loop. **c,** RNA-seq results from Kubo *et al.* are also shown as bar plot. p-value: one-
414 side Wald test provided by DE-seq2. **d-f,** The same analysis as **a-c** at *Notch1* locus using data
415 from Hsieh *et al.* Note that Hsieh *et al.* used mNET-seq for transcriptome profiling.

416 **Supplementary Figure 2. Meta-analysis of independent CTCF-degron Hi-C datasets to**
417 **test the consistency and reveal the epigenetic features of dynamic loops.**

418 **a**, Using the *in situ* Hi-C data from the CTCF-degron study by Kubo *et al.*, left: scatterplot
419 comparing the strength of loop pixels between WT and CTCF-depleted mESCs; middle: the
420 same scatterplot as left but highlight the lost (blue), retained (black), and gained (red) loop
421 pixels; right: violin plot comparing the size of loop pixels. **b,** Pie charts showing the components
422 of lost, retained and gained loop pixels based on epigenetic marks. **c**, Scatter plots in the left
423 column highlight the lost, retained, and gained loop pixels from Kubo *et al*., the right two
424 columns show the quantitative changes of the corresponding pixels in Nora *et al.* or Hsieh *et al*.
425 **d-f,** The same analyses as **a-c** are performed with CTCF-degron micro-C data by Hsieh *et al.*

426 **Supplementary Figure 3. Meta-analysis of independent *Rad21*-degron Hi-C datasets to**
427 **test the consistency and reveal the epigenetic features of dynamic loops.**

428 **a,** Left and middle panels define the lost, retained, and gained loop pixels upon *Rad21*-depletion
429 from the study by Rhodes *et al.*, and the right panel compares their loop size distribution. **b,** Pie
430 charts showing the components of lost, retained and gained loops based on epigenetic marks.
431 **c,** Scatter plots testing if the lost, retained and gained loop pixels from Rhodes *et al.* are
432 quantitatively consistent in the other two *Rad21*-degron studies. **d-f,** The same analyses as **a-c**
433 are performed with *Rad21*-degron micro-C data by Hsieh *et al.*

434 **Supplementary Figure 4. Meta-analysis of independent *Wapl*-degron Hi-C datasets to test**
435 **the consistency and reveal the epigenetic features of dynamic loops.**

436 **a,** Left and middle panels define the lost, retained, and gained loop pixels upon *Wapl*-depletion
437 from the study by Kriz *et al.*, and the right panel compares their loop size distribution. **b,** Pie
438 charts showing the components of lost, retained and gained loops based on epigenetic marks.
439 **c,** Scatter plots testing if the lost, retained and gained loop pixels from Kriz *et al.* are
440 quantitatively consistent in the other two *Wapl*-degron studies. **d-f,** The same analyses as **a-c**
441 are performed with *Wapl*-degron micro-C data by Hsieh *et al.*

442 **Supplementary Figure 5. Meta-analysis reveals the fate of dynamic loops from CTCF-**
443 **degron studies upon depletion of cohesin factors.**

444 **a,** The lost, retained, and gained loop pixels from one CTCF-degron Hi-C data (Kubo *et al.*) are
445 plotted onto Hi-C scatter plots from *Rad21*- or *Wapl*-degron studies to examine their quantitative
446 changes. **b,** Same analyses are done with the lost, retained, and gained loop pixels from the
447 CTCF-degron Hi-C data by Hsieh *et al.*
448

449 **Method**

450 **Experiments**

451 <u>*Cell culture*</u>

452 Mouse embryonic stem cells (E14Tg2a) were cultured in DMEM+Glutamax (ThermoFisher ,
453 #10566-016) with the addition of 15% Fetal Bovine Serum (ThermoFisher, # SH30071.03),
454 550μM β-mercaptoethanol (ThermoFisher, #21985-023), 1mM Sodium Pyruvate
455 (ThermoFisher, #11360-070), 1X non-essential amino acids (ThermoFisher, #11140-50), and
456 10^4 U of Leukemia inhibitory factor (Millipore, #ESG1107). The cells were maintained at a
457 density of 0.2-1.5x10^5 cells/cm^2 and were passaged every 24-48 hours using TrypLE
458 (ThermoFisher, #12563011) on dishes coated with 0.1% gelatin (Millipore, # ES-006-B) at 37°C
459 and 5% CO2. The medium was refreshed daily when cells were not being passaged. Regular
460 checks for mycoplasma infection were performed every 3-4 months, with negative results.

461 *4C-seq*

462 4C-seq was conducted following a published protocol(*50*). Briefly, 5 million cells fixed with 2%
463 formaldehyde, then quenched with 125 mM glycine. The fixed cells were lysed using a cell lysis
464 buffer containing 50 mM Tris-Cl pH 7.5, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-
465 100, and 1x protease inhibitor cocktail (Roche, Cat#11873580001) for 20–30 minutes on ice.
466 Following lysis, nuclei were collected by centrifugation at 2,500g for 5 minutes at 4°C and
467 washed once with 1x restriction enzyme buffer. The nuclei pellets were resuspended in 1x
468 restriction enzyme buffer and treated with 0.3% SDS for 1 hour at 37°C with shaking, followed
469 by an additional hour with 2.5% Triton X-100. Chromatin was then digested overnight with the
470 appropriate restriction enzyme at the recommended temperature while rotating in an air bath.
471 The specific restriction enzymes used for each locus are detailed in **Supplemental Table 6**.
472 After digestion, the enzymes were inactivated by heating at 65°C, and nuclei were ligated with
473 50 µl of T4 DNA ligase (Invitrogen, Cat# 15224–090) in a 7 ml ligation solution at 16°C
474 overnight. Reverse cross-linking was performed by treating the samples with proteinase K to
475 isolate the proximity-ligated DNA. The purified DNA was quantified and subjected to secondary
476 restriction enzyme digestion using roughly 1 unit of restriction enzyme per 1 µg of DNA at the
477 recommended temperature overnight. After enzyme inactivation, the samples were self-ligated
478 with T4 DNA ligase. The ligated DNA was recovered using sodium acetate and ethanol and
479 quantified using the Qubit dsDNA HS assay kit (Thermo Fisher, Cat# Q32851). The 4C
480 templates were amplified using designed primers to generate sequencing libraries. The primer
481 system was modified to be compatible with the Illumina Nextera system, using two sequential
482 PCRs. Locus-specific inverse PCR primers are listed in **Supplemental Table 6**. For each locus,
483 the 4C templates were amplified with locus-specific primers using 200 ng of template per
484 reaction, and products from five parallel amplifications were pooled to generate the final 4C
485 library. A 50 µl PCR product aliquot was purified with homemade Sera-Mag beads. One-fifth of
486 the purified DNA was used for a second PCR with primers N7xx and N5xx, identical to Illumina
487 Nextera sample preparation primers. The final products were purified and sequenced. Reads
488 from the first cutting site were used for data analysis.

489 *Plasmid cloning*

490 The dCas9 expression vectors used in this study were created using the Cas9 expression
491 vector backbone from the pX330 plasmid (Addgene; plasmid 42230) through the In-Fusion
492 cloning technique. The dCas9 genes were amplified from the pHAGE EF1α dCas9-KRAB

493 plasmid (Addgene; plasmid 50919) via PCR and independently cloned into the AgeI and EcoRI
494 sites of the pX330 plasmid, substituting the Cas9 ORF. Detailed information regarding the
495 primers can be found in **Supplementary Table 5**. All sgRNAs in this study were designed using
496 the CCTop-CRISPR/Cas9 target online predictor.

497 *CARGO assembly*

498 The CARGO array assembly follows a published protocol(*45*). The protocol consists of two main
499 steps: constant region preparation and array assembly. In the constant region preparation step,
500 the constant region with proper sticky ends is released by digesting the pGEMT-hU6-spSL
501 (provided by Dr. Joanna Wysocka) plasmid with BsmBI. The released constant region is then
502 separated from the pGEMT backbone on a 1% agarose gel and gel purified using a commercial
503 kit (QIAGEN, Cat#28704). Oligonucleotides (50-60 bp) with variable regions were ordered
504 through IDT (**Supplementary Table 5**). The forward and reverse oligos are annealed and
505 phosphorylated and then stored at -20°C.
506 In the array assembly step, the annealed chimeric oligo is mixed with the digested constant
507 region, and a minimum of 4-hour ligation with high-concentration T4 ligase (Thermofisher
508 Scientific, Cat#EL0013) is performed. The ligation step yields ligation products, referred to as
509 minicircles. After the stoichiometry check, individual minicircles are treated with plasmid-safe
510 exonuclease to clean up the un-ligated products. Then, the plasmid-safe exonuclease treated
511 individual minicircles are pooled together and purified. The pooled circle mixes are assembled
512 directly into the digested destination vector pX332-mCherry (provided by Dr. Joanna Wysocka)
513 through an optimized golden gate assembly reaction. The product of the golden-gate reaction is
514 then subjected to plasmid-safe exonuclease (Epicentre, Cat#E3101K) treatment. After that, the
515 exonuclease-treated reaction is then transformed into competent cells (Bioline, Cat#BIO-
516 85027). For each plate, individual colonies are picked for plasmid purification using a miniprep
517 kit (QIAGEN, Cat#27204) and test digestion with Acc65I, KpnI, and XbaI. Plasmids with the
518 correct digestion pattern are then verified by Sanger sequencing

519 *ChIPmentation*

520 We employed ChIPmentation(*51*) to map CTCF in various samples. Briefly, cells and tissues
521 were fixed in 1% formaldehyde at room temperature for 15 minutes, followed by glycine
522 quenching. To isolate nuclei, we lysed cell with lysis buffer 1 (50 mM HEPES; pH 7.5, 140 mM
523 NaCl, 1 mM EDTA, 10% glycerol, 0.5% Nonidet-40, 0.25% Triton X-100, and a protease

524 inhibitor cocktail) for 10 minutes at 4°C. The nuclei were pelleted at 1,800 g for 10 minutes at
525 4°C and then resuspended in nuclei lysis buffer (10 mM Tris-Cl pH 8.0, 100 mM NaCl, 1 mM
526 EDTA, 0.5 mM EGTA, 0.1% Na-Deoxycholate, 0.5% N-lauroylsarcosine, and a protease
527 inhibitor cocktail) for sonication. For CTCF pulldown, 100-150 µg of chromatin was used.
528 Dynabeads M-280 (Life Technologies, Sheep Anti-Rabbit IgG, #11204D) were washed three
529 times with 0.5 mg/mL of BSA/PBS on ice, then incubated with the CTCF antibody (Abcam,
530 #ab70303) for at least 2 hours at 4°C. The beads/antibody complexes were then washed with
531 BSA/PBS. Pulldown was performed in binding buffer (1% Triton-X 100, 0.1% Sodium
532 Deoxycholate, and a protease inhibitor cocktail in 1X TE) by mixing the beads/antibody
533 complexes with chromatin. After overnight incubation, the beads/antibody/chromatin complexes
534 were washed with RIPA buffer (50 mM HEPES pH 8.0, 1% NP-40, 0.7% Sodium Deoxycholate,
535 0.5 M LiCl, 1 mM EDTA, and a protease inhibitor cocktail). The beads complexes were then
536 incubated with homemade Tn5 transposase in tagmentation reaction buffer (10 mM Tris-Cl pH
537 8.0 and 5 mM MgCl2) for 10 minutes at 37°C. To remove free DNA, the beads were washed
538 twice with 1X TE on ice. The pulldown DNA was recovered by reversing crosslinking overnight,
539 followed by purification with PCR Clean DX beads (Aline Biosciences, #C-1003-50). ChIP-seq
540 libraries were generated by amplifying the pulldown DNA with Illumina Nextera primers. Size
541 selection was then performed with PCR Clean DX beads to select fragments ranging from 100
542 bp to 1000 bp.

543 **Data analysis**

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

545 Hi-C and micro-C sequencing reads were mapped to mm10 reference genome separately using
546 bowtie(*52*) (v.1.1.2) as previously described(*37*). After removing non-unique reads and PCR
547 duplication reads, the bam files are used as input into *HiCorr* in '*Bam-process-HindIII*', '*Bam-*
548 *process-DpnII*', and '*Bam-process-microC*' modes for bias correction with conventional Hi-C, *in-*
549 *situ* Hi-C, and micro-C, respectively. After bias correction, we enhance the Hi-C/micro-C data
550 with *DeepLoop*. Note that *DeepLoop* has multiple models to enhance Hi-C data with different
551 read depth. The enhance models we chose for each dataset are listed in **Supplementary Table**
552 **2**. Because each datasets have a pair of Hi-C experiments (control and KO conditions), if the
553 read depths of the two Hi-C experiments are significantly different, we will down-sample the
554 high-depth dataset to lower depth to ensure fair comparison (**Supplementary Table 2**). 4C-seq

data were analyzed using pipe4C(*50*) (v.1.1.3) to generate bam and bigwig files for data visualization on UCSC genome browser.

ChIP–seq and CUT&RUN

ChIP-seq and CUT&RUN data were mapped to mm10 (mouse) using bowtie. After mapping step, we remove PCR duplication reads and use Macs2(*53*) to call peak and generated bigwig to for data visualization on UCSC genome browser.

RNA-seq and mNET-seq

RNA-seq data are aligned to mm10 using HISAT2(*54*) and read counts for each gene are called with featureCounts(*55*). DESeq2(*56*) is used to identify differential expressed genes (DEGs). The DEGs are identified using Wald test with the p-value <0.01 and log2FC>1. For mNET-seq analysis, we followed the customized pipeline from the Hsieh *et al.*(*34*) with the following steps: (i) Trim the adapter using TrimGalore (v.0.6.7) (https://github.com/FelixKrueger/TrimGalore) to remove adapter sequences 'AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC' and 'GATCGTCGGACTGTAGAACTCTGAAC' at each side of the reads; (ii) Map trimmed read to mm10 using STAR; (iii) Identify the last nucleotide incorporated by Pol II by using mNET_snr (https://github.com/tomasgomes/mNET_snr) to locate the 3′-nucleotide of the second read and the strand sign of the first read. (iv) Annotate and count the reads of each gene using FeatureCounts(*55*); (v) Use DEseq2(*56*) to quantify the changes of the mNET-seq signal at the gene body between control and KO cells.

Pairwise loop strength comparison

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

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

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

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

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

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

595 *Classify chromatin loops based on epigenome data*

596 For the purpose of being consistent, we intersected loop anchors with the same set of mESC
597 CTCF, H3K27ac, H3K4me3, H3K27me3 and H3K36me3 ChIP-seq peaks from ENCODE(*57*) to
598 classify the loop pixels. We classification loop pixels in the following order:

599 (1) EP-only: both anchors have either H3K4me3 or H3K27ac marker, but they do not have
600 CTCF peaks.
601 (2) CTCF-EP: after excluding EP-only, look for pixels which only one anchor has CTCF, and
602 the other anchor has either H3K4me3 or H3K27ac marker.
603 (3) CTCF-only: after excluding EP-only and CTCF-EP, look for pixels with at least one
604 anchor bound by CTCF.
605 (4) H3K27me3: after excluding (1) - (3) above, look for pixels with at least one anchor bound
606 by H3K27me3.
607 (5) H3K36me3: after excluding (1) - (4) above, look for pixels with at least one anchor bound
608 by H3K36me3.
609 (6) The rest pixels are classified as “other”.

610 *Define FINs and FIN-blocked regular enhancers and super enhancers (SE)*

611 We define FINs as a subset of CTCF binding sites that satisfy the following criteria: (i) they
612 located at the anchors of CTCF loops that are weakened upon CTCF depletion; (ii) the FIN
613 should be inside a newly gained loop in CTCF-depleted cells; (iii) the other anchor of the CTCF

614 loop should be outside of the newly gained loop. We call enhancers as H3K27ac peaks and use
615 a published list of super enhancers in mESCs(*58*) (**Supplementary Tables 11 & 12**). This
616 allows us to determine the types of enhancers blocked by FINs.

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

618 A recurrent loop-gaining pixel is called if it satisfies FC > 4 in at least two studies. We define the
619 location of each loop pixel as the entire region between the left and right anchors (including the
620 anchors themselves). To define loop-gaining loci, we merge all gained loop pixels from the
621 CTCF-degron studies using the "merge" option in *bedtools*; each loop-gaining locus usually has
622 multiple loop pixels. Loop loci with gap size < 20kb will be merge into one locus. In this study,
623 we merged all the recurrent loop-gaining pixels to obtain 299 recurrent loop events. To test the
624 consistency of loop-gaining loci between independent CTCF-degron study, we call loop-gaining
625 loci for each individual CTCF-degron study and the share loop-gaining loci from two different
626 studies are called if they have at least one base overlap.

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

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

635 *Compartment and TADs in mESCs*

636 Hi-C compartment analysis is performed in WT mESCs following previous publication(*59*). After
637 QC step, we constructed 500kb resolution contact matrices for each chromosome and then
638 normalized them using the genome distance. We next generated correlation matrices for each
639 chromosome and performed the principal component analysis on the correlation matrix and then
640 assigns the genome into two compartments by using PC1. We used the density of transcription
641 starting sites (TSS) of each 500-kb bin to determine compartment A and compartment B. For
642 TADs and their boundaries, we directly use the TADs in mESCs defined by Dixon *et al.*
643 (**Supplementary Table 9**).

644 *Ideogram plot*

645 We generate ideogram plots with Rideogram(*60*) to visualized the locations of recurrent gaining-
646 loop pixels, compartment, and TADs.

647 *Density and enrichment of DEGs upon CTCF-depletion*

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

653 *Hi-C data visualization*

654 Heatmap was used to visualize the raw, ratio, and *DeepLoop*-enhanced heatmaps. We use the
655 following strategies as described before(*37*) to determine the color scales these heatmaps. (i)
656 Raw or ICE-normalized heatmaps represent raw read counts; the brightest red color indicates
657 the 98th percentile of the contact matrix. Color is proportionally scaled down to one read (white).
658 (ii) HiCorr heatmap represents the ratio value after Hi-C bias correction; the brightest red color
659 indicates at least twofold enrichment of a given pixel comparing to its background. Color is
660 proportionally scaled down to onefold (no enrichment). (iii) DeepLoop heatmaps outputs
661 'transformed fold change' that represents only relative levels of signal enrichment (for example,
662 a value of one may no longer represents no enrichment). We therefore set the brightest red
663 color as the lower limit of the top 300k pixels genome wide. Color is proportionally scaled down
664 to half of that lower limit or onefold, whichever is higher. All loop arches shown in figures were
665 generated with UCSC Genome Browser(*61*).

666

667 **Data availability**

668 Accession numbers for third-party data used in this study can be found in **Supplementary**
669 **Table 1**. The raw data of 4C–seq and ChIP-seq generated in this study, and reanalyzed
670 published data, can be found at accession number GSE243728. The 18 Hi-C and micro-C
671 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).

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.

References

1. J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012).
2. E. P. Nora, B. R. Lajoie, E. G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N. L. van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Bluthgen, J. Dekker, E. Heard, Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385 (2012).
3. W. A. Bickmore, B. van Steensel, Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270-1284 (2013).
4. J. Dekker, L. Mirny, The 3D Genome as Moderator of Chromosomal Communication. *Cell* **164**, 1110-1121 (2016).
5. R. P. McCord, N. Kaplan, L. Giorgetti, Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Mol Cell* **77**, 688-708 (2020).
6. M. Merkenschlager, E. P. Nora, CTCF and Cohesin in Genome Folding and Transcriptional Gene Regulation. *Annual review of genomics and human genetics* **17**, 17-43 (2016).

7. Z. Tang, O. J. Luo, X. Li, M. Zheng, J. J. Zhu, P. Szalaj, P. Trzaskoma, A. Magalska, J. Wlodarczyk, B. Ruszczycki, P. Michalski, E. Piecuch, P. Wang, D. Wang, S. Z. Tian, M. Penrad-Mobayed, L. M. Sachs, X. Ruan, C. L. Wei, E. T. Liu, G. M. Wilczynski, D. Plewczynski, G. Li, Y. Ruan, CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription. *Cell* **163**, 1611-1627 (2015).
8. A. L. Sanborn, S. S. Rao, S. C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K. P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E. K. Stamenova, E. S. Lander, E. L. Aiden, Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E6456-6465 (2015).
9. J. Nuebler, G. Fudenberg, M. Imakaev, N. Abdennur, L. A. Mirny, Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proceedings of the National Academy of Sciences of the United States of America* **115**, E6697-E6706 (2018).
10. K. Nasmyth, Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Annu Rev Genet* **35**, 673-745 (2001).
11. G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, L. A. Mirny, Formation of Chromosomal Domains by Loop Extrusion. *Cell reports* **15**, 2038-2049 (2016).
12. I. F. Davidson, B. Bauer, D. Goetz, W. Tang, G. Wutz, J. M. Peters, DNA loop extrusion by human cohesin. *Science* **366**, 1338-1345 (2019).
13. Y. Kim, Z. Shi, H. Zhang, I. J. Finkelstein, H. Yu, Human cohesin compacts DNA by loop extrusion. *Science* **366**, 1345-1349 (2019).
14. M. J. Rowley, V. G. Corces, Organizational principles of 3D genome architecture. *Nat Rev Genet* **19**, 789-800 (2018).
15. L. L. P. Hanssen, M. T. Kassouf, A. M. Oudelaar, D. Biggs, C. Preece, D. J. Downes, M. Gosden, J. A. Sharpe, J. A. Sloane-Stanley, J. R. Hughes, B. Davies, D. R. Higgs, Tissue-specific CTCF-cohesin-mediated chromatin architecture delimits enhancer interactions and function in vivo. *Nat Cell Biol* **19**, 952-961 (2017).
16. E. P. Nora, A. Goloborodko, A. L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, B. G. Bruneau, Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922 (2017).
17. S. S. P. Rao, S. C. Huang, B. Glenn St Hilaire, J. M. Engreitz, E. M. Perez, K. R. Kieffer-Kwon, A. L. Sanborn, S. E. Johnstone, G. D. Bascom, I. D. Bochkov, X. Huang, M. S. Shamim, J. Shin, D. Turner, Z. Ye, A. D. Omer, J. T. Robinson, T. Schlick, B. E. Bernstein, R. Casellas, E. S. Lander, E. L. Aiden, Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e324 (2017).
18. K. Nasmyth, C. H. Haering, Cohesin: its roles and mechanisms. *Annu Rev Genet* **43**, 525-558 (2009).
19. K. L. Chan, M. B. Roig, B. Hu, F. Beckouet, J. Metson, K. Nasmyth, Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. *Cell* **150**, 961-974 (2012).
20. P. J. Huis in 't Veld, F. Herzog, R. Ladurner, I. F. Davidson, S. Piric, E. Kreidl, V. Bhaskara, R. Aebersold, J. M. Peters, Characterization of a DNA exit gate in the human cohesin ring. *Science* **346**, 968-972 (2014).
21. S. Kueng, B. Hegemann, B. H. Peters, J. J. Lipp, A. Schleiffer, K. Mechtler, J. M. Peters, Wapl controls the dynamic association of cohesin with chromatin. *Cell* **127**, 955-967 (2006).

754 22. W. Schwarzer, N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, Y. Loe-Mie,
755 N. A. Fonseca, W. Huber, H. H. C, L. Mirny, F. Spitz, Two independent modes of
756 chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56 (2017).
757 23. J. H. I. Haarhuis, R. H. van der Weide, V. A. Blomen, J. O. Yanez-Cuna, M. Amendola,
758 M. S. van Ruiten, P. H. L. Krijger, H. Teunissen, R. H. Medema, B. van Steensel, T. R.
759 Brummelkamp, E. de Wit, B. D. Rowland, The Cohesin Release Factor WAPL Restricts
760 Chromatin Loop Extension. *Cell* **169**, 693-707 e614 (2017).
761 24. G. Wutz, C. Varnai, K. Nagasaka, D. A. Cisneros, R. R. Stocsits, W. Tang, S.
762 Schoenfelder, G. Jessberger, M. Muhar, M. J. Hossain, N. Walther, B. Koch, M.
763 Kueblbeck, J. Ellenberg, J. Zuber, P. Fraser, J. M. Peters, Topologically associating
764 domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL,
765 and PDS5 proteins. *The EMBO journal* **36**, 3573-3599 (2017).
766 25. J. A. Wallace, G. Felsenfeld, We gather together: insulators and genome organization.
767 *Curr Opin Genet Dev* **17**, 400-407 (2007).
768 26. K. S. Wendt, K. Yoshida, T. Itoh, M. Bando, B. Koch, E. Schirghuber, S. Tsutsumi, G.
769 Nagae, K. Ishihara, T. Mishiro, K. Yahata, F. Imamoto, H. Aburatani, M. Nakao, N.
770 Imamoto, K. Maeshima, K. Shirahige, J. M. Peters, Cohesin mediates transcriptional
771 insulation by CCCTC-binding factor. *Nature* **451**, 796-801 (2008).
772 27. V. Parelho, S. Hadjur, M. Spivakov, M. Leleu, S. Sauer, H. C. Gregson, A. Jarmuz, C.
773 Canzonetta, Z. Webster, T. Nesterova, B. S. Cobb, K. Yokomori, N. Dillon, L. Aragon, A.
774 G. Fisher, M. Merkenschlager, Cohesins functionally associate with CTCF on
775 mammalian chromosome arms. *Cell* **132**, 422-433 (2008).
776 28. M. Franke, D. M. Ibrahim, G. Andrey, W. Schwarzer, V. Heinrich, R. Schopflin, K. Kraft,
777 R. Kempfer, I. Jerkovic, W. L. Chan, M. Spielmann, B. Timmermann, L. Wittler, I. Kurth,
778 P. Cambiaso, O. Zuffardi, G. Houge, L. Lambie, F. Brancati, A. Pombo, M. Vingron, F.
779 Spitz, S. Mundlos, Formation of new chromatin domains determines pathogenicity of
780 genomic duplications. *Nature* **538**, 265-269 (2016).
781 29. J. Weischenfeldt, T. Dubash, A. P. Drainas, B. R. Mardin, Y. Chen, A. M. Stutz, S. M.
782 Waszak, G. Bosco, A. R. Halvorsen, B. Raeder, T. Efthymiopoulos, S. Erkek, C. Siegl,
783 H. Brenner, O. T. Brustugun, S. M. Dieter, P. A. Northcott, I. Petersen, S. M. Pfister, M.
784 Schneider, S. K. Solberg, E. Thunissen, W. Weichert, T. Zichner, R. Thomas, M. Peifer,
785 A. Helland, C. R. Ball, M. Jechlinger, R. Sotillo, H. Glimm, J. O. Korbel, Pan-cancer
786 analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer
787 hijacking. *Nat Genet* **49**, 65-74 (2017).
788 30. D. G. Lupianez, K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H.
789 Kayserili, J. M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M.
790 Borschiwer, S. A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M.
791 Spielmann, A. Visel, S. Mundlos, Disruptions of topological chromatin domains cause
792 pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025 (2015).
793 31. C. Anania, R. D. Acemel, J. Jedamzick, A. Bolondi, G. Cova, N. Brieske, R. Kuhn, L.
794 Wittler, F. M. Real, D. G. Lupianez, In vivo dissection of a clustered-CTCF domain
795 boundary reveals developmental principles of regulatory insulation. *Nat Genet* **54**, 1026-
796 1036 (2022).
797 32. J. Zuin, G. Roth, Y. Zhan, J. Cramard, J. Redolfi, E. Piskadlo, P. Mach, M.
798 Kryzhanovska, G. Tihanyi, H. Kohler, M. Eder, C. Leemans, B. van Steensel, P. Meister,
799 S. Smallwood, L. Giorgetti, Nonlinear control of transcription through enhancer-promoter
800 interactions. *Nature* **604**, 571-577 (2022).
801 33. N. Kubo, H. Ishii, X. Xiong, S. Bianco, F. Meitinger, R. Hu, J. D. Hocker, M. Conte, D.
802 Gorkin, M. Yu, B. Li, J. R. Dixon, M. Hu, M. Nicodemi, H. Zhao, B. Ren, Promoter-
803 proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat Struct*
804 *Mol Biol* **28**, 152-161 (2021).

805 34. T. S. Hsieh, C. Cattoglio, E. Slobodyanyuk, A. S. Hansen, X. Darzacq, R. Tjian,
806 Enhancer-promoter interactions and transcription are largely maintained upon acute loss
807 of CTCF, cohesin, WAPL or YY1. *Nat Genet* **54**, 1919-1932 (2022).
808 35. B. Akgol Oksuz, L. Yang, S. Abraham, S. V. Venev, N. Krietenstein, K. M. Parsi, H.
809 Ozadam, M. E. Oomen, A. Nand, H. Mao, R. M. J. Genga, R. Maehr, O. J. Rando, L. A.
810 Mirny, J. H. Gibcus, J. Dekker, Systematic evaluation of chromosome conformation
811 capture assays. *Nature methods*, (2021).
812 36. L. Lu, X. Liu, W. K. Huang, P. Giusti-Rodriguez, J. Cui, S. Zhang, W. Xu, Z. Wen, S. Ma,
813 J. D. Rosen, Z. Xu, C. F. Bartels, R. Kawaguchi, M. Hu, P. C. Scacheri, Z. Rong, Y. Li,
814 P. F. Sullivan, H. Song, G. L. Ming, Y. Li, F. Jin, Robust Hi-C Maps of Enhancer-
815 Promoter Interactions Reveal the Function of Non-coding Genome in Neural
816 Development and Diseases. *Mol Cell* **79**, 521-534 e515 (2020).
817 37. S. Zhang, D. Plummer, L. Lu, J. Cui, W. Xu, M. Wang, X. Liu, N. Prabhakar, J. Shrinet,
818 D. Srinivasan, P. Fraser, Y. Li, J. Li, F. Jin, DeepLoop robustly maps chromatin
819 interactions from sparse allele-resolved or single-cell Hi-C data at kilobase resolution.
820 *Nat Genet* **54**, 1013-1025 (2022).
821 38. A. J. Kriz, D. Colognori, H. Sunwoo, B. Nabet, J. T. Lee, Balancing cohesin eviction and
822 retention prevents aberrant chromosomal interactions, Polycomb-mediated repression,
823 and X-inactivation. *Mol Cell* **81**, 1970-1987 e1979 (2021).
824 39. J. D. P. Rhodes, A. Feldmann, B. Hernandez-Rodriguez, N. Diaz, J. M. Brown, N. A.
825 Fursova, N. P. Blackledge, P. Prathapan, P. Dobrinic, M. K. Huseyin, A. Szczurek, K.
826 Kruse, K. A. Nasmyth, V. J. Buckle, J. M. Vaquerizas, R. J. Klose, Cohesin Disrupts
827 Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells. *Cell reports*
828 **30**, 820-835 e810 (2020).
829 40. N. Q. Liu, M. Maresca, T. van den Brand, L. Braccioli, M. Schijns, H. Teunissen, B. G.
830 Bruneau, E. P. Nora, E. de Wit, WAPL maintains a cohesin loading cycle to preserve
831 cell-type-specific distal gene regulation. *Nat Genet* **53**, 100-109 (2021).
832 41. T. Hirayama, T. Yagi, Regulation of clustered protocadherin genes in individual neurons.
833 *Seminars in cell & developmental biology* **69**, 122-130 (2017).
834 42. E. Flaherty, T. Maniatis, The role of clustered protocadherins in neurodevelopment and
835 neuropsychiatric diseases. *Curr Opin Genet Dev* **65**, 144-150 (2020).
836 43. Z. Jia, J. Li, X. Ge, Y. Wu, Y. Guo, Q. Wu, Tandem CTCF sites function as insulators to
837 balance spatial chromatin contacts and topological enhancer-promoter selection.
838 *Genome Biol* **21**, 75 (2020).
839 44. T. Hirayama, E. Tarusawa, Y. Yoshimura, N. Galjart, T. Yagi, CTCF is required for
840 neural development and stochastic expression of clustered Pcdh genes in neurons. *Cell*
841 *reports* **2**, 345-357 (2012).
842 45. B. Gu, T. Swigut, A. Spencley, M. R. Bauer, M. Chung, T. Meyer, J. Wysocka,
843 Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory
844 elements. *Science* **359**, 1050-1055 (2018).
845 46. D. R. Tarjan, W. A. Flavahan, B. E. Bernstein, Epigenome editing strategies for the
846 functional annotation of CTCF insulators. *Nature communications* **10**, 4258 (2019).
847 47. T. Pollex, A. Rabinowitz, M. C. Gambetta, R. Marco-Ferreres, R. R. Viales, A.
848 Jankowski, C. Schaub, E. E. M. Furlong, Enhancer-promoter interactions become more
849 instructive in the transition from cell-fate specification to tissue differentiation. *Nat Genet*
850 **56**, 686-696 (2024).
851 48. C. P. Fulco, J. Nasser, T. R. Jones, G. Munson, D. T. Bergman, V. Subramanian, S. R.
852 Grossman, R. Anyoha, B. R. Doughty, T. A. Patwardhan, T. H. Nguyen, M. Kane, E. M.
853 Perez, N. C. Durand, C. A. Lareau, E. K. Stamenova, E. L. Aiden, E. S. Lander, J. M.
854 Engreitz, Activity-by-contact model of enhancer-promoter regulation from thousands of
855 CRISPR perturbations. *Nat Genet* **51**, 1664-1669 (2019).

49. M. H. Kagey, J. J. Newman, S. Bilodeau, Y. Zhan, D. A. Orlando, N. L. van Berkum, C. C. Ebmeier, J. Goossens, P. B. Rahl, S. S. Levine, D. J. Taatjes, J. Dekker, R. A. Young, Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435 (2010).
50. P. H. L. Krijger, G. Geeven, V. Bianchi, C. R. E. Hilvering, W. de Laat, 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* **170**, 17-32 (2020).
51. C. Schmidl, A. F. Rendeiro, N. C. Sheffield, C. Bock, ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nature methods* **12**, 963-965 (2015).
52. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
53. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
54. D. Kim, B. Langmead, S. L. Salzberg, HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360 (2015).
55. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
56. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
57. F. Yue, Y. Cheng, A. Breschi, J. Vierstra, W. Wu, T. Ryba, R. Sandstrom, Z. Ma, C. Davis, B. D. Pope, Y. Shen, D. D. Pervouchine, S. Djebali, R. E. Thurman, R. Kaul, E. Rynes, A. Kirilusha, G. K. Marinov, B. A. Williams, D. Trout, H. Amrhein, K. Fisher-Aylor, I. Antoshechkin, G. DeSalvo, L. H. See, M. Fastuca, J. Drenkow, C. Zaleski, A. Dobin, P. Prieto, J. Lagarde, G. Bussotti, A. Tanzer, O. Denas, K. Li, M. A. Bender, M. Zhang, R. Byron, M. T. Groudine, D. McCleary, L. Pham, Z. Ye, S. Kuan, L. Edsall, Y. C. Wu, M. D. Rasmussen, M. S. Bansal, M. Kellis, C. A. Keller, C. S. Morrissey, T. Mishra, D. Jain, N. Dogan, R. S. Harris, P. Cayting, T. Kawli, A. P. Boyle, G. Euskirchen, A. Kundaje, S. Lin, Y. Lin, C. Jansen, V. S. Malladi, M. S. Cline, D. T. Erickson, V. M. Kirkup, K. Learned, C. A. Sloan, K. R. Rosenbloom, B. Lacerda de Sousa, K. Beal, M. Pignatelli, P. Flicek, J. Lian, T. Kahveci, D. Lee, W. J. Kent, M. Ramalho Santos, J. Herrero, C. Notredame, A. Johnson, S. Vong, K. Lee, D. Bates, F. Neri, M. Diegel, T. Canfield, P. J. Sabo, M. S. Wilken, T. A. Reh, E. Giste, A. Shafer, T. Kutyavin, E. Haugen, D. Dunn, A. P. Reynolds, S. Neph, R. Humbert, R. S. Hansen, M. De Bruijn, L. Selleri, A. Rudensky, S. Josefowicz, R. Samstein, E. E. Eichler, S. H. Orkin, D. Levasseur, T. Papayannopoulou, K. H. Chang, A. Skoultchi, S. Gosh, C. Disteche, P. Treuting, Y. Wang, M. J. Weiss, G. A. Blobel, X. Cao, S. Zhong, T. Wang, P. J. Good, R. F. Lowdon, L. B. Adams, X. Q. Zhou, M. J. Pazin, E. A. Feingold, B. Wold, J. Taylor, A. Mortazavi, S. M. Weissman, J. A. Stamatoyannopoulos, M. P. Snyder, R. Guigo, T. R. Gingeras, D. M. Gilbert, R. C. Hardison, M. A. Beer, B. Ren, E. C. Mouse, A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355-364 (2014).
58. D. Hnisz, B. J. Abraham, T. I. Lee, A. Lau, V. Saint-Andre, A. A. Sigova, H. A. Hoke, R. A. Young, Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947 (2013).
59. E. Lieberman-Aiden, N. L. van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B. R. Lajoie, P. J. Sabo, M. O. Dorschner, R. Sandstrom, B. Bernstein, M. A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L. A. Mirny, E. S. Lander, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009).

60. Z. Hao, D. Lv, Y. Ge, J. Shi, D. Weijers, G. Yu, J. Chen, RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput Sci* **6**, e251 (2020).
61. J. Navarro Gonzalez, A. S. Zweig, M. L. Speir, D. Schmelter, K. R. Rosenbloom, B. J. Raney, C. C. Powell, L. R. Nassar, N. D. Maulding, C. M. Lee, B. T. Lee, A. S. Hinrichs, A. C. Fyfe, J. D. Fernandes, M. Diekhans, H. Clawson, J. Casper, A. Benet-Pages, G. P. Barber, D. Haussler, R. M. Kuhn, M. Haeussler, W. J. Kent, The UCSC Genome Browser database: 2021 update. *Nucleic Acids Res* **49**, D1046-D1057 (2021).

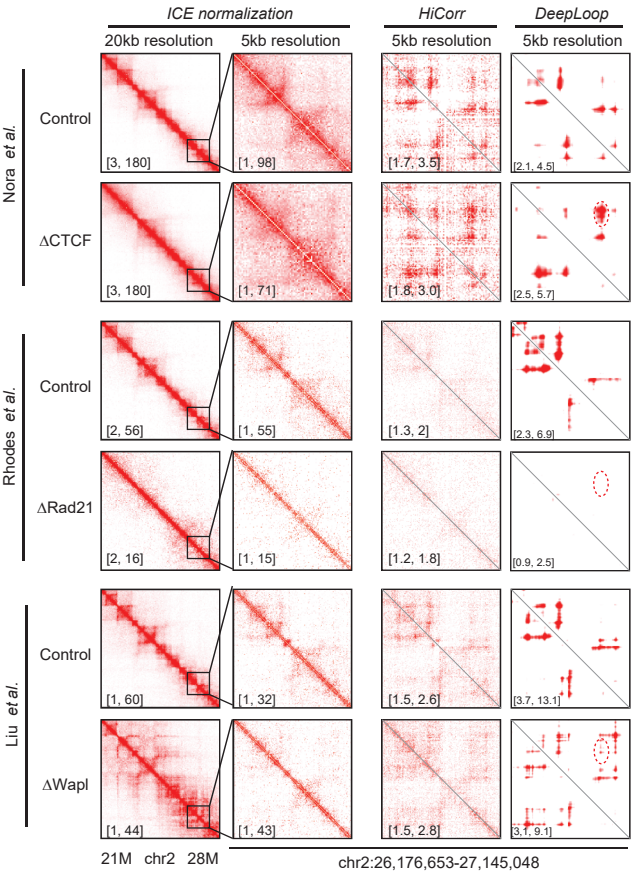
Figure 1

a

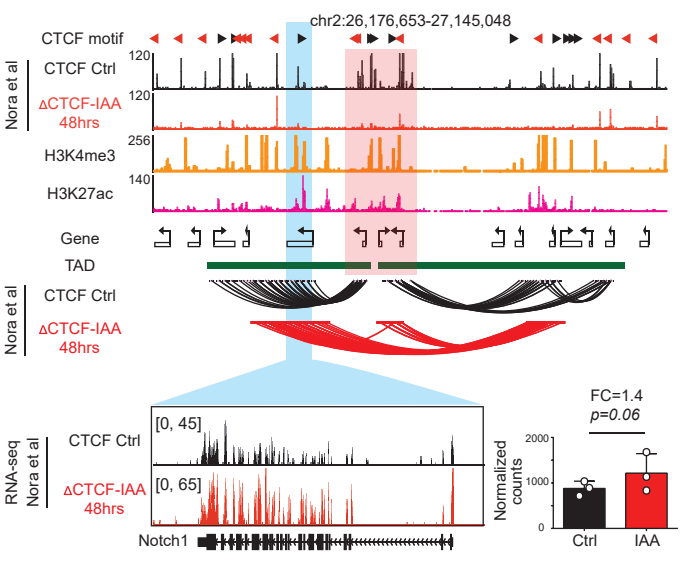
Published Hi-C data in mESCs upon acute depleation of architectural proteins

Publication	mESC line	KO protocol	C-technique	# of contacts	
				Ctrl	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	116M
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

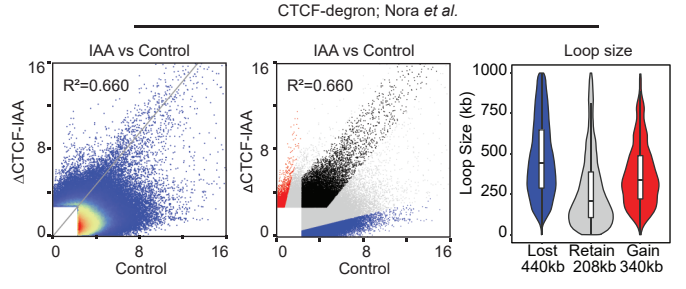
b



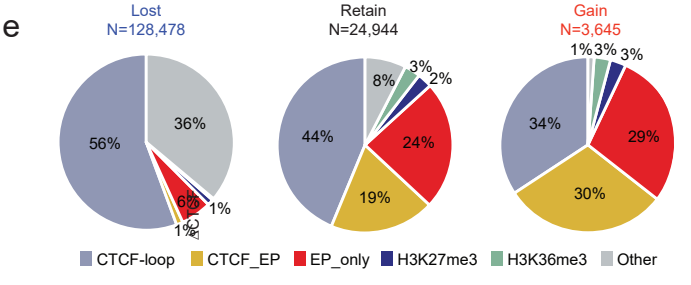
c



d



e



f

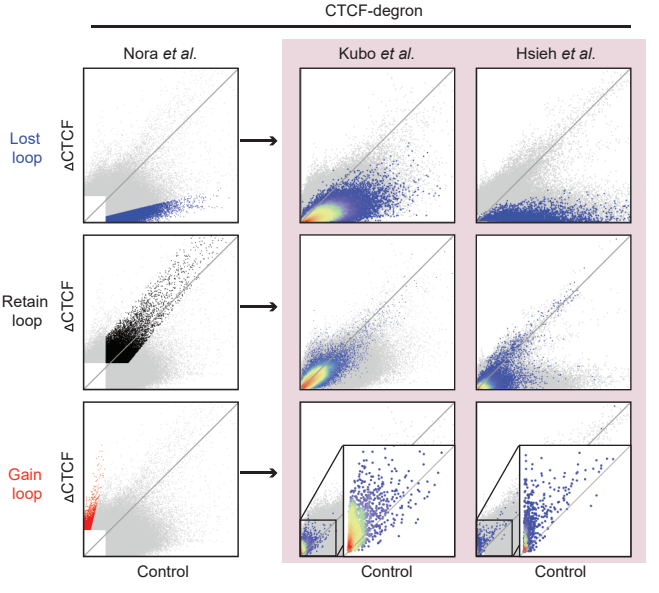


Figure 2

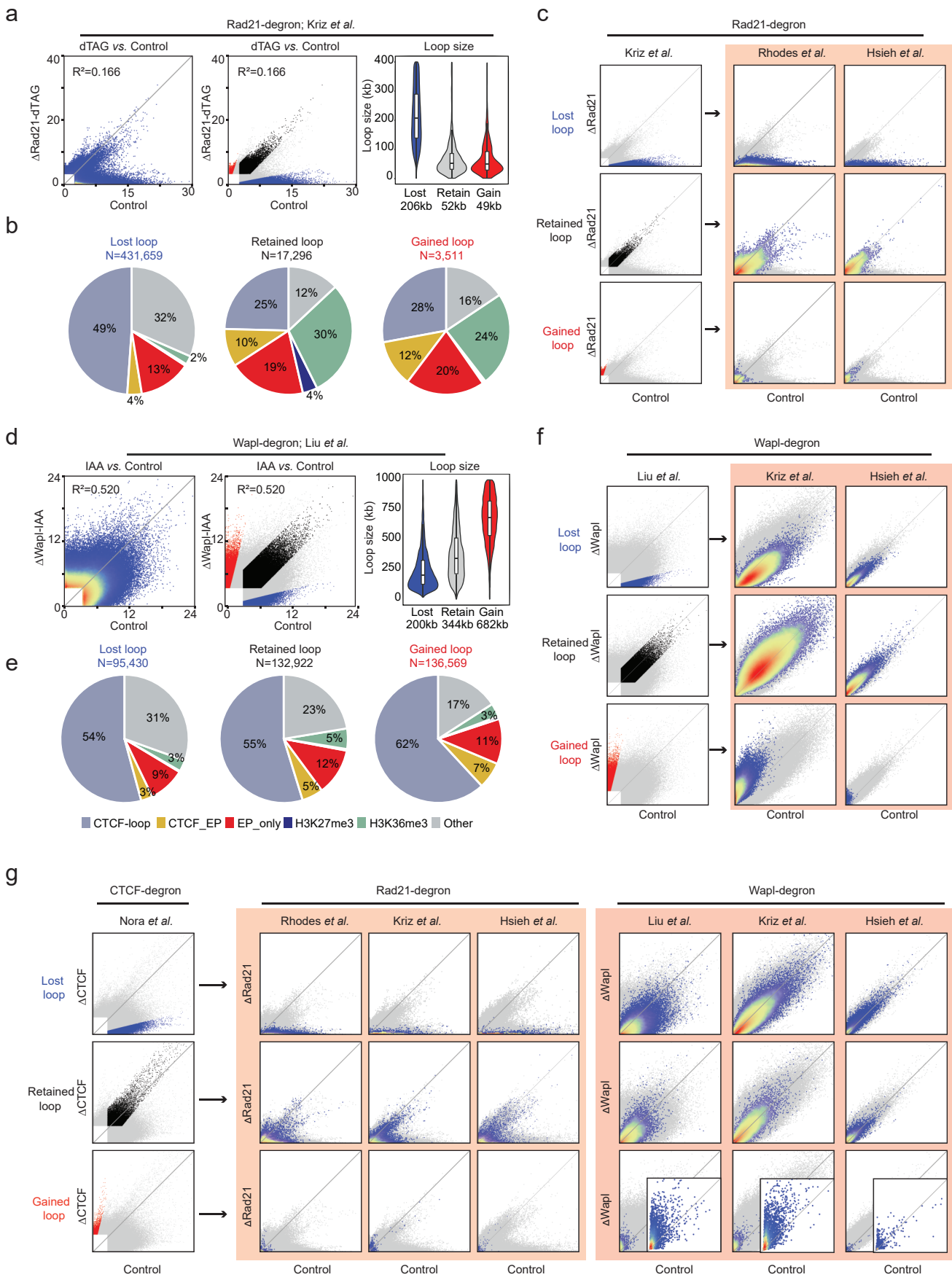
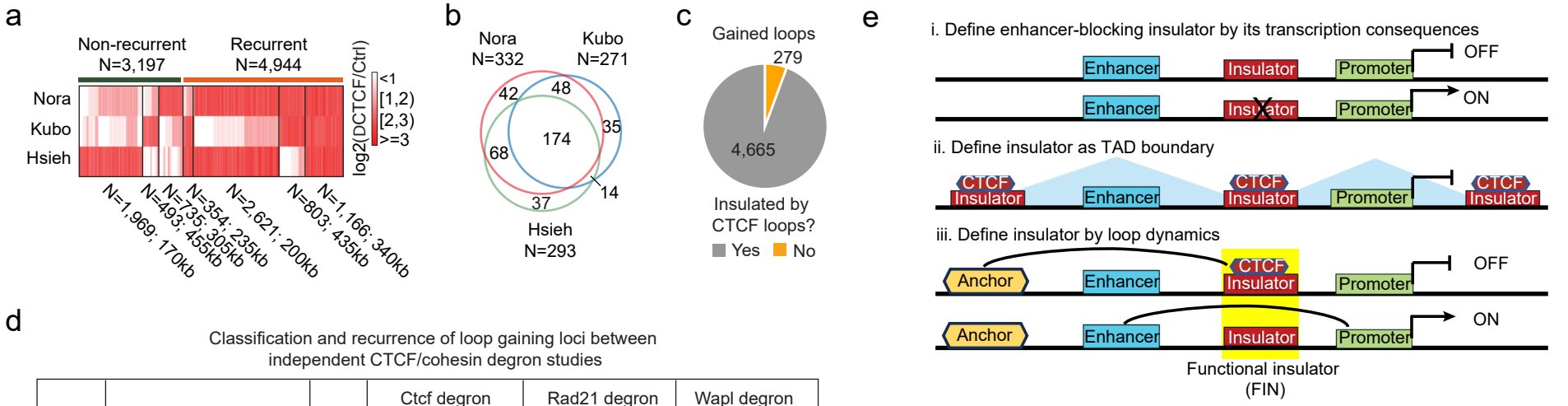

a
Rad21-degron; Kriz et al.
dTAG vs. Control
dTAG vs. Control
Loop size
R²=0.166
R²=0.166
ΔRad21-dTAG
Control
Loop size (kb)
Lost 206kb
Retain 52kb
Gain 49kb
b
Lost loop N=431,659
Retained loop N=17,296
Gained loop N=3,511
c
Rad21-degron
Kriz et al.
Rhodes et al.
Hsieh et al.
Lost loop
Retained loop
Gained loop
ΔRad21
Control
d
Wapl-degron; Liu et al.
IAA vs. Control
IAA vs. Control
Loop size
R²=0.520
R²=0.520
ΔWapl-IAA
Control
Loop size (kb)
Lost 200kb
Retain 344kb
Gain 682kb
e
Lost loop N=95,430
Retained loop N=132,922
Gained loop N=136,569
CTCF-loop
CTCF_EP
EP_only
H3K27me3
H3K36me3
Other
f
Wapl-degron
Liu et al.
Kriz et al.
Hsieh et al.
Lost loop
Retained loop
Gained loop
ΔWapl
Control
g
CTCF-degron
Nora et al.
Rad21-degron
Rhodes et al.
Kriz et al.
Hsieh et al.
Wapl-degron
Liu et al.
Kriz et al.
Hsieh et al.
Lost loop
Retained loop
Gained loop
ΔCTCF
ΔRad21
ΔWapl
Control

Figure 3



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

Figure 4

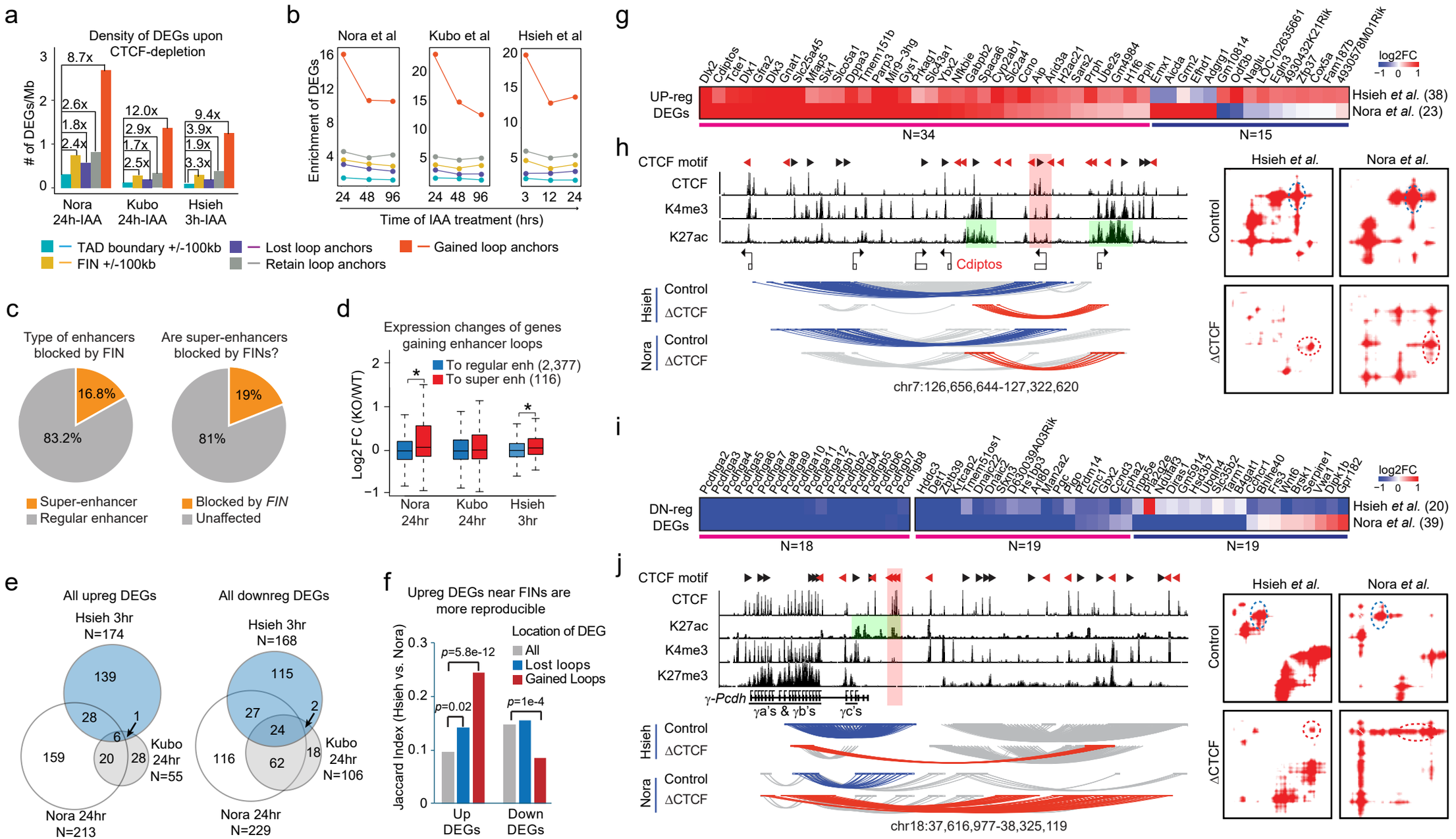
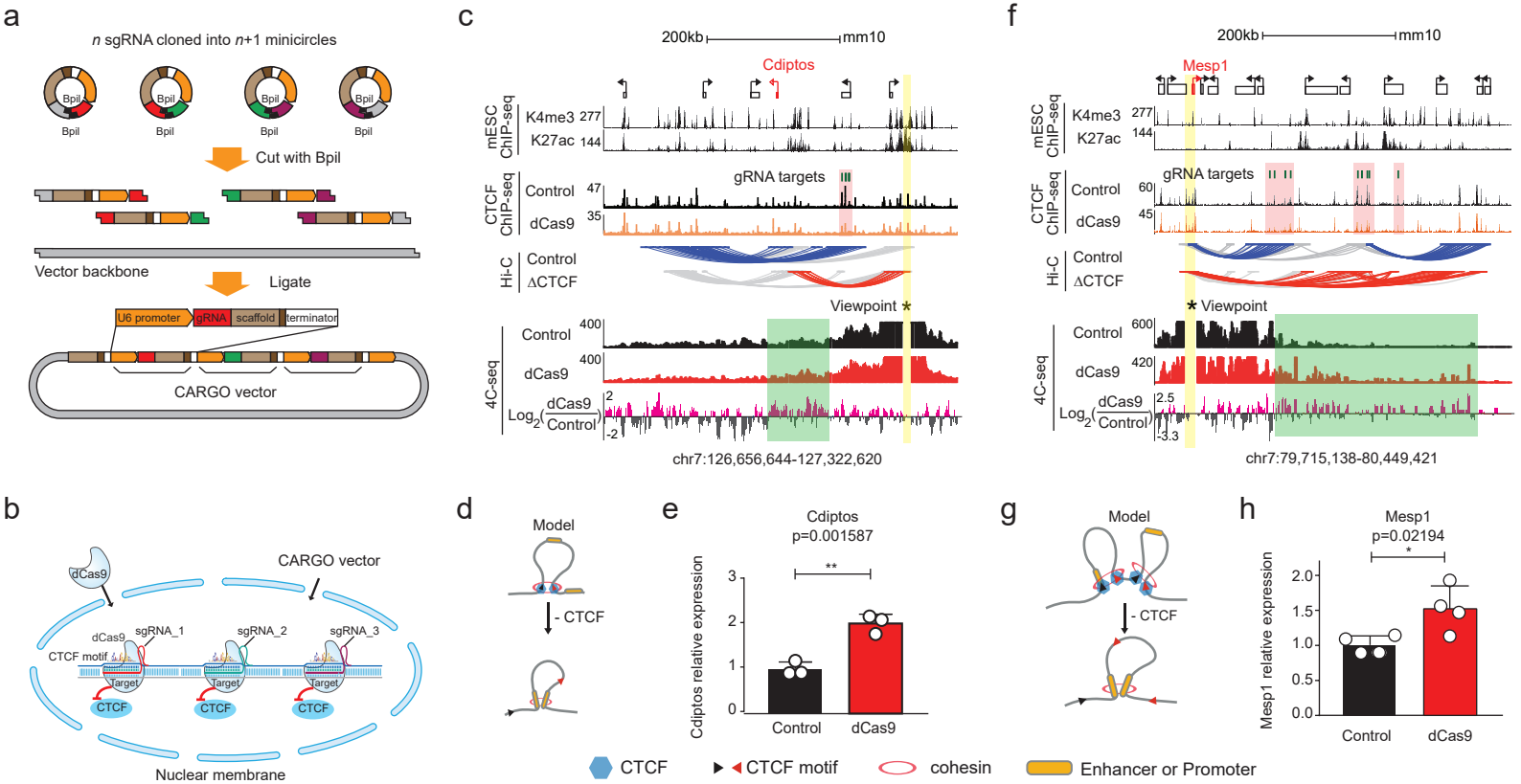
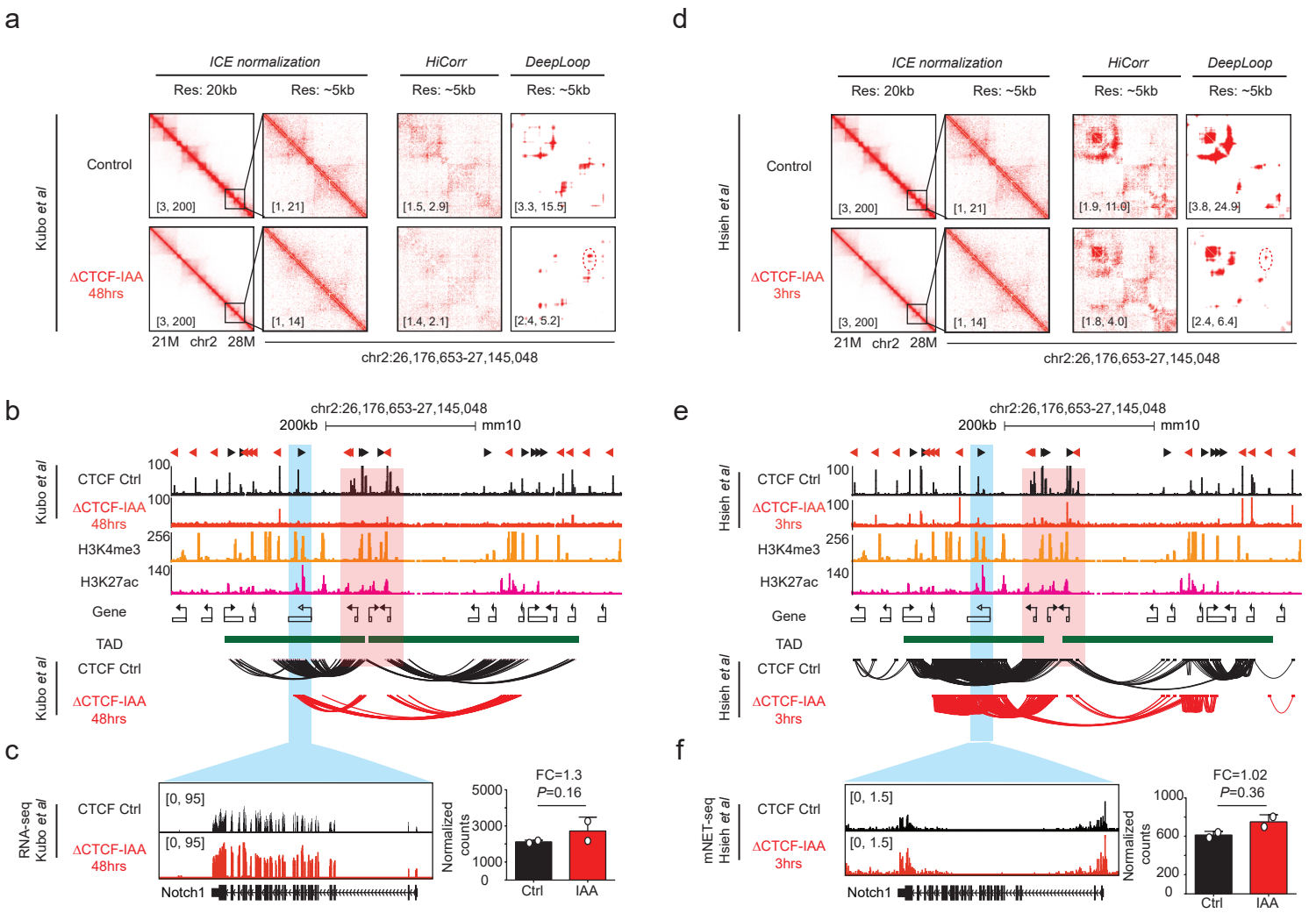

a
Density of DEGs upon CTCF-depletion
# of DEGs/Mb
8.7x
2.6x
1.8x
2.4x
12.0x
2.9x
1.7x
2.5x
9.4x
3.9x
1.9x
3.3x
Nora 24h-IAA
Kubo 24h-IAA
Hsieh 3h-IAA
b
Nora et al
Kubo et al
Hsieh et al
Enrichment of DEGs
Time of IAA treatment (hrs)
TAD boundary +/-100kb
FIN +/-100kb
Lost loop anchors
Retain loop anchors
Gained loop anchors
c
Type of enhancers blocked by FIN
16.8%
83.2%
Super-enhancer
Regular enhancer
Are super-enhancers blocked by FINs?
19%
81%
Blocked by FIN
Unaffected
d
Expression changes of genes gaining enhancer loops
To regular enh (2,377)
To super enh (116)
Log2 FC (KO/WT)
Nora 24hr
Kubo 24hr
Hsieh 3hr
e
All upreg DEGs
Hsieh 3hr N=174
139
28
1
6
159
20
28
Kubo 24hr N=55
Nora 24hr N=213
All downreg DEGs
Hsieh 3hr N=168
115
27
2
24
116
62
18
Kubo 24hr N=106
Nora 24hr N=229
f
Upreg DEGs near FINs are more reproducible
Location of DEG
All
Lost loops
Gained Loops
Jaccard Index (Hsieh vs. Nora)
p=5.8e-12
p=0.02
p=1e-4
Up DEGs
Down DEGs
g
UP-reg DEGs
log2FC
Hsieh et al. (38)
Nora et al. (23)
N=34
N=15
h
CTCF motif
CTCF
K4me3
K27ac
Cdiptos
Hsieh
Control
ΔCTCF
Nora
Control
ΔCTCF
chr7:126,656,644-127,322,620
Hsieh et al.
Nora et al.
Control
ΔCTCF
i
DN-reg DEGs
log2FC
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
Hsieh
Control
ΔCTCF
Nora
Control
ΔCTCF
chr18:37,616,977-38,325,119
Hsieh et al.
Nora et al.
Control
ΔCTCF

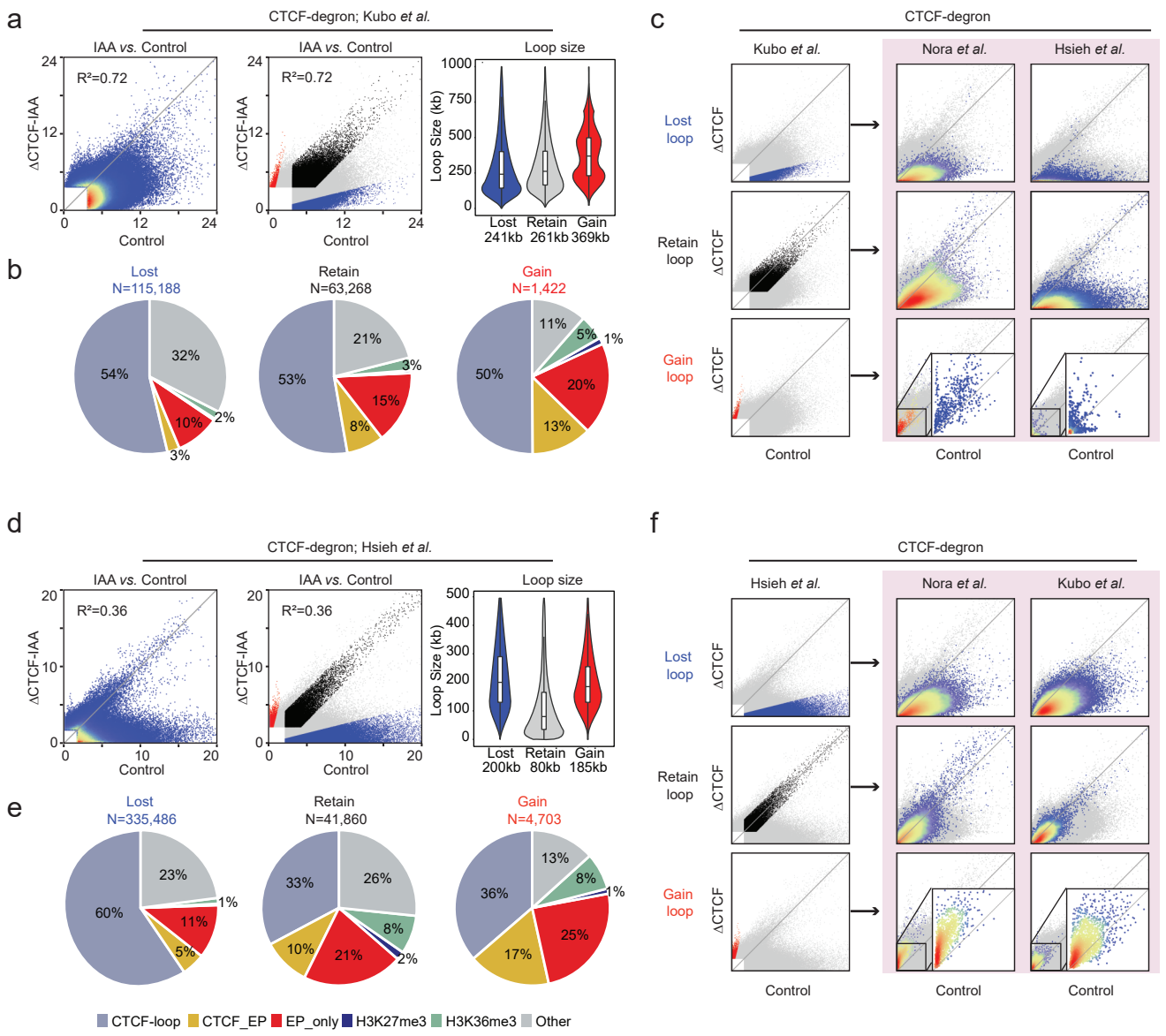
Figure 5


a
n sgRNA cloned into n+1 minicircles
BpiI
Cut with BpiI
Vector backbone
Ligate
U6 promoter
gRNA
scaffold
terminator
CARGO vector
b
dCas9
CARGO vector
sgRNA_1
sgRNA_2
sgRNA_3
CTCF motif
Target
CTCF
Nuclear membrane
c
200kb
mm10
Cdiptos
mESC ChIP-seq
K4me3
K27ac
CTCF ChIP-seq
Control
dCas9
gRNA targets
Hi-C
ΔCTCF
Viewpoint
4C-seq
Log2(dCas9/Control)
chr7:126,656,644-127,322,620
d
Model
- CTCF
e
Cdiptos
p=0.001587
Cdiptos relative expression
Control
dCas9
f
Mesp1
chr7:79,715,138-80,449,421
g
h
Mesp1
p=0.02194
Mesp1 relative expression
CTCF
CTCF motif
cohesin
Enhancer or Promoter

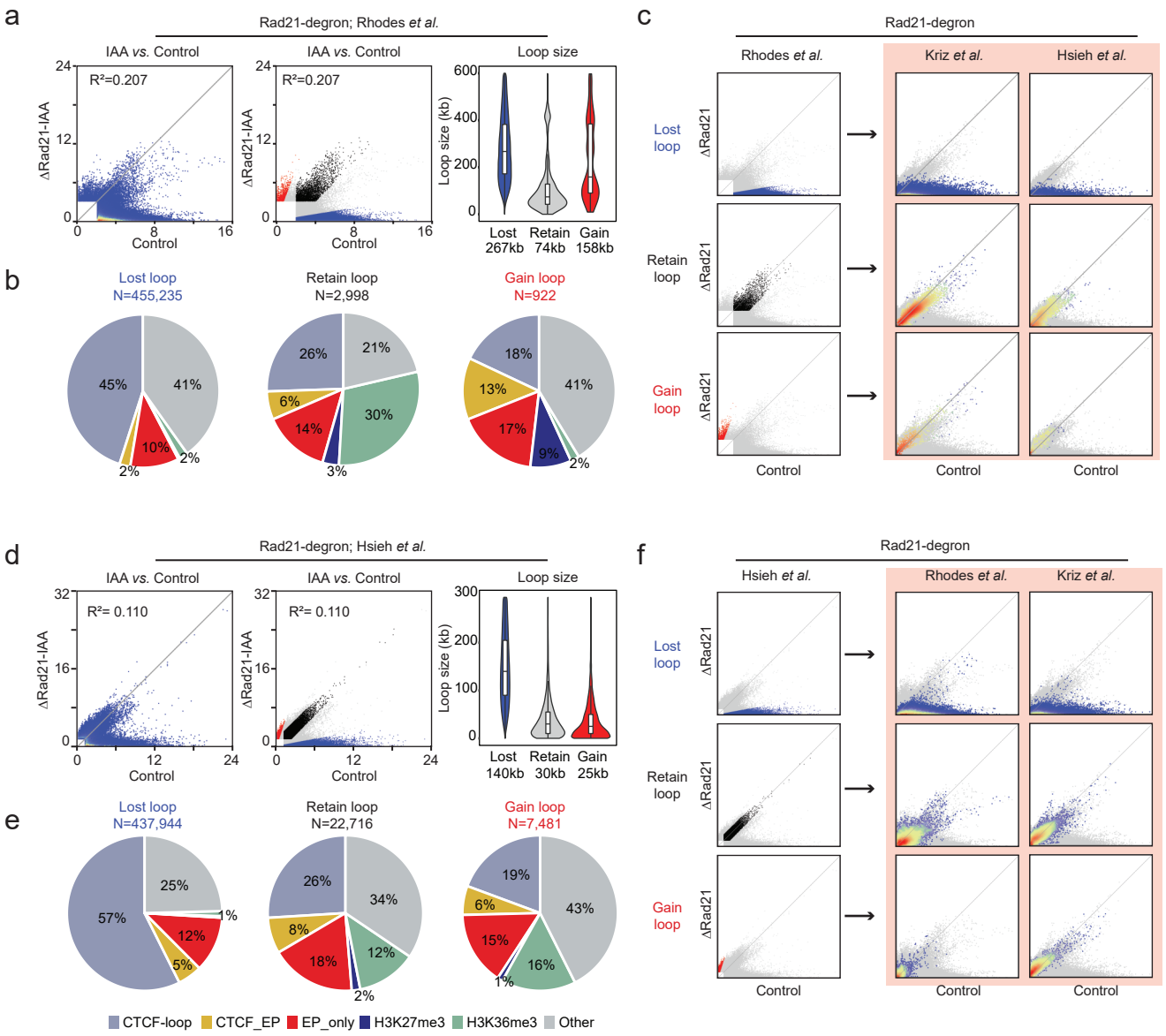
Supplementary Figure 1



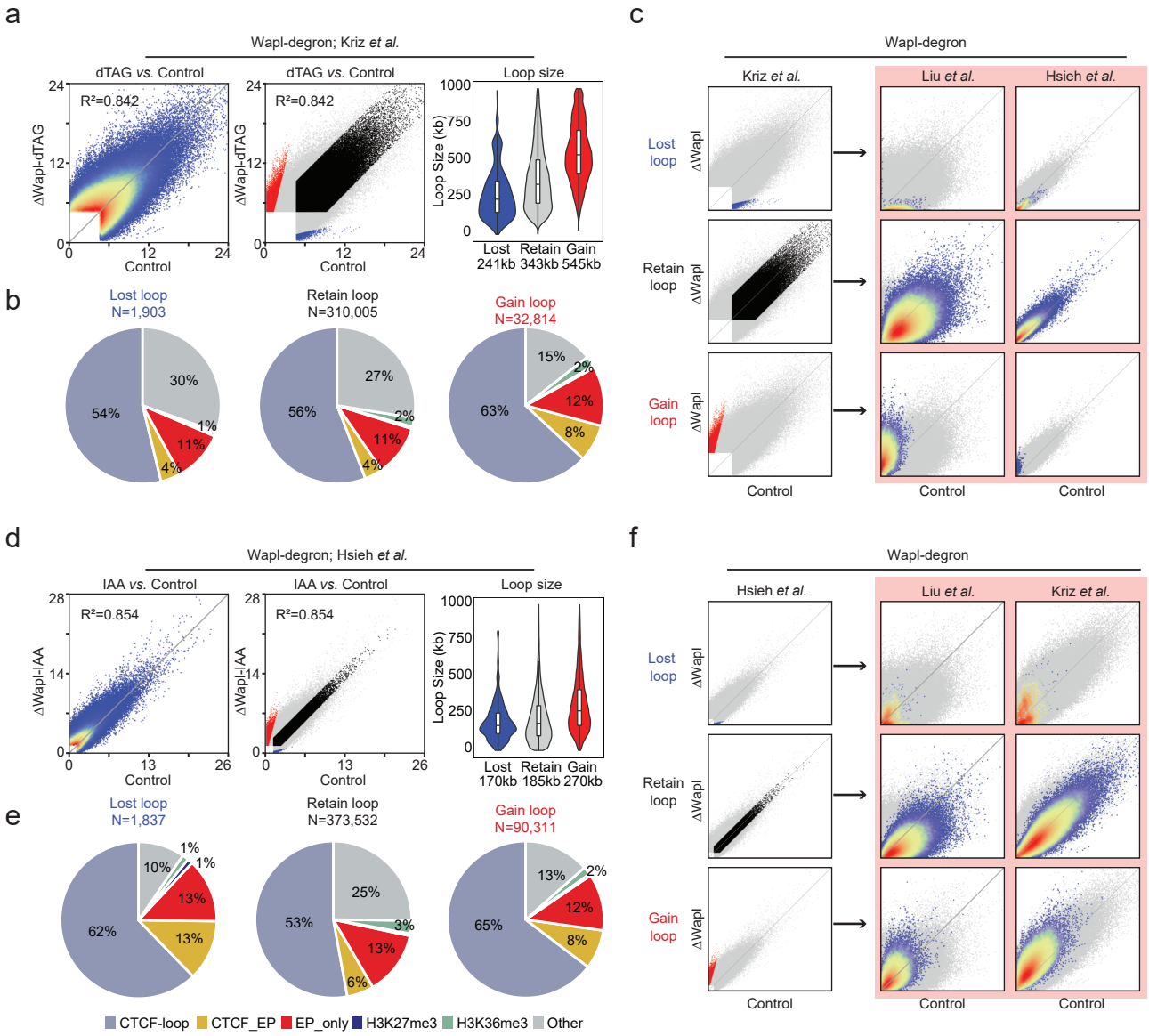
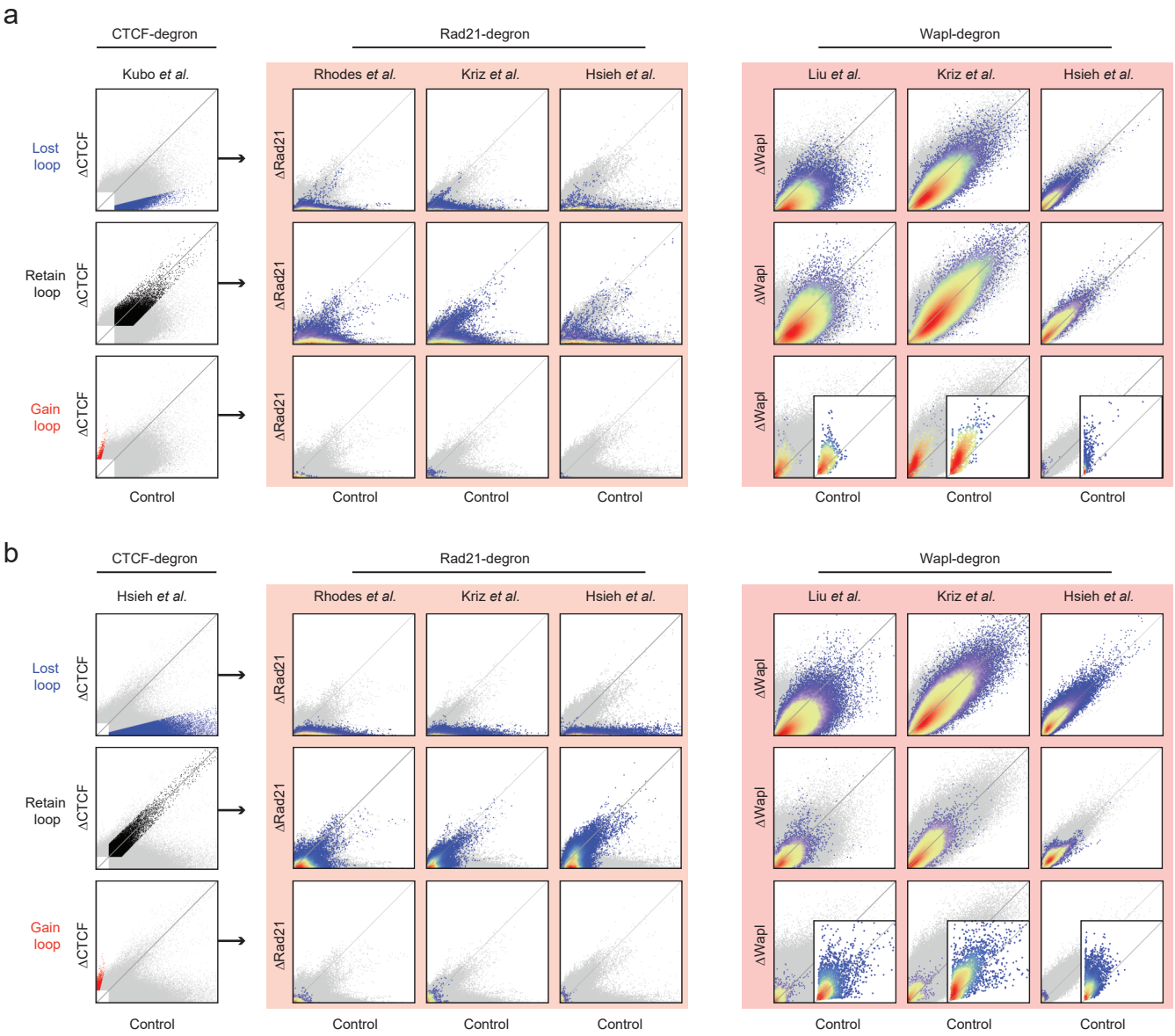
Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4



Supplementary Figure 5