DNA methylation insulates genic regions from CTCF loops near nuclear speckles

Authors: Shelby A. Roseman,[1,2,3] Allison P. Siegenfeld,[1,2,3] Ceejay Lee,[1,2] Nicholas Z. Lue,[1,2] Amanda L. Waterbury,[1,2] Brian B. Liau[1,2,*]

Affiliations:

[1]Department of Chemistry and Chemical Biology, Cambridge, MA 02138, USA

[2]Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

[3]Co-first authors

*Correspondence: liau@chemistry.harvard.edu

Running title: Methylation insulates genic CTCF loops at speckles

Abstract

CTCF occupancy is inhibited by DNA methylation, but the effects of methylation-sensitive CTCF binding on chromatin structure and transcription are poorly understood. Here, we use a selective DNMT1 inhibitor to demethylate the genome and investigate the features and functions of the resulting 'reactivated' CTCF peaks. We reveal a tendency for reactivated CTCF peaks to form chromatin loops on gene bodies and to interact with highly-looping partner peaks located in active chromatin regions near architectural stripe anchors. To dissect the potential functions of reactivated CTCF peaks, we combine DNMT1 inhibition with acute CTCF depletion and identify a subset of genes that are upregulated in a CTCF-dependent manner. These genes, as well as reactivated peaks and their highly-looping partners, are located in close proximity to nuclear speckles, prompting us to investigate whether CTCF looping and speckles are functionally linked. We find that despite the prevalence of CTCF reactivation and looping near nuclear speckles, acute depletion of speckles affects transcription with minimal disruption to CTCF binding or looping. Taken together, these findings suggest a model in which DNA methylation prevents genic CTCF occupancy and spurious interactions with regulatory elements near nuclear speckles, while CTCF looping is robust towards speckle disruption.

Introduction

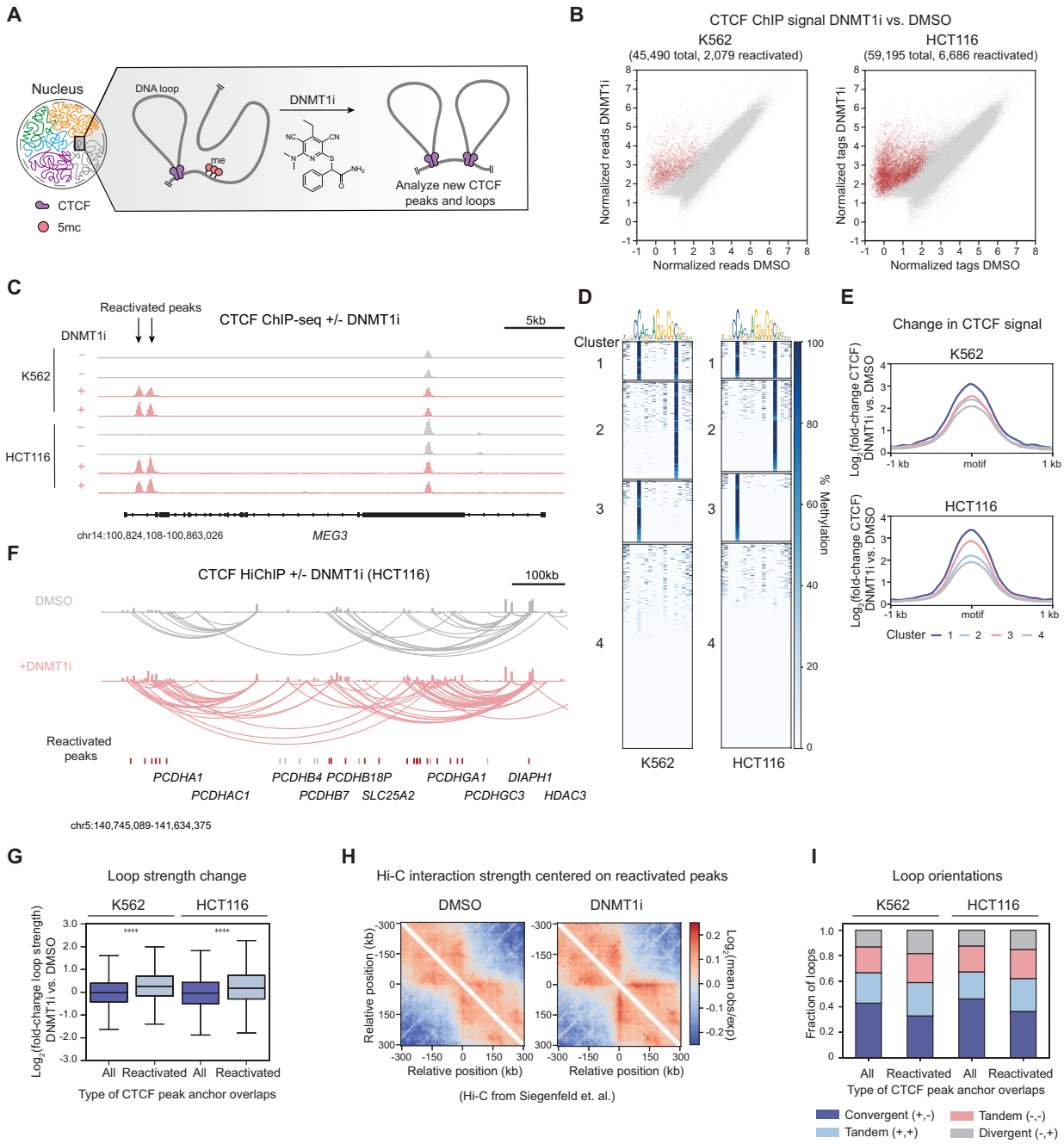
The mammalian genome is organized at multiple length scales (Rowley and Corces 2018; Rao et al. 2014; Dixon et al. 2012; Misteli 2020). This three-dimensional chromatin folding contributes to essential cellular processes such as transcription and DNA replication and is actively maintained by protein cofactors (Zheng and Xie 2019; Rao et al. 2017; Nora et al. 2017; Steensel and Belmont 2017). The architectural protein CTCF establishes the anchors of many DNA loops and topologically-associating domains (TADs) likely by blocking loop extrusion by cohesin (Wendt

et al. 2008; Li et al. 2020; Fudenberg et al. 2016; Kim et al. 2019; Davidson et al. 2019; Xiang and Corces 2021; Hansen et al. 2017; Sanborn et al. 2015; Nora et al. 2017). Consequently, perturbation of CTCF binding sites at domain boundaries can lead to aberrant cross-domain enhancer-promoter contacts, which have been linked to gene misexpression in cancer and developmental disorders (Dubois et al. 2022; Spielmann et al. 2018; Lupiáñez et al. 2015; Flavahan et al. 2016, 2019). Furthermore, CTCF binding and looping on or near genes can directly enforce enhancer-promoter loops and facilitate differential splicing and polyadenylation (Rinzema et al. 2022; Kubo et al. 2021; Shukla et al. 2011; Ruiz-Velasco et al. 2017; Nanavaty et al. 2020; Alharbi et al. 2021; Soto and Lipscombe 2020; Steensel and Furlong 2019). Although these examples demonstrate that CTCF can modulate transcription in specific contexts, acute depletion of CTCF only leads to moderate transcriptional defects, highlighting the complex and nuanced roles of CTCF in regulating gene expression (Schoenfelder and Fraser 2019; Nora et al. 2017; Rao et al. 2017; Vermunt et al. 2019).

CTCF binding is sensitive to DNA methylation, and consequently global demethylation of the genome ‘reactivates’ hundreds to thousands of previously masked CTCF peaks (Bell and Felsenfeld 2000; Hark et al. 2000; Hashimoto et al. 2017; Maurano et al. 2015; Wang et al. 2012). However, these reactivated CTCF sites comprise only a small subset of potential methylated CTCF sites that are typically occupied in other cell types, suggesting that DNA methylation serves as a specialized mechanism to regulate particular CTCF binding sites rather than global CTCF binding (Maurano et al. 2015). While methylation-sensitive CTCF peaks have been linked to gene expression and mRNA processing in some instances,(Ruiz-Velasco et al. 2017; Shukla et al. 2011; Nanavaty et al. 2020; Flavahan et al. 2019, 2016; Schuijers et al. 2018) the functions of most methylation-sensitive CTCF peaks are unknown. Moreover, the broader impacts of CTCF reactivation on chromatin looping have been largely unexplored.

53 In this study, we employ a selective inhibitor of DNMT1, the maintenance DNA methyltransferase,
54 to induce genome-wide demethylation and investigate the chromatin looping patterns and
55 functions of the resulting reactivated CTCF peaks. Through combining small molecule
56 perturbations with acute degron systems, we glean novel insights into how methylation-sensitive
57 CTCF looping interfaces with gene expression in transcriptionally active regions of the genome.
58
59 **Results**
60
61 **DNMT1 inhibition reactivates CTCF peaks and loops at previously methylated sites**
62
63 Loss of DNA methylation leads to increased CTCF occupancy at hundreds to thousands of CTCF
64 sites depending on the cell type (Maurano et al. 2015), but a global understanding of their looping
65 patterns and functions is lacking. Recently, selective, noncovalent DNMT1 inhibitors (e.g.,
66 GSK3484862) have become available, enabling dynamic control of DNA methylation with fewer
67 undesirable effects such as DNA damage and cytotoxicity (Portilho et al. 2021; Pappalardi et al.
68 2021). Consequently, we sought to employ these new DNMT1 inhibitors to probe the functions
69 and looping patterns of methylation-sensitive CTCF sites.
70
71 We treated K562 and HCT116 cells with GSK3484862 (DNMT1i (10 μM)), for three days, and
72 performed CTCF ChIP-seq and HiChIP to investigate CTCF-associated changes in chromatin
73 looping (**Fig 1A**). Under these conditions, DNMT1i treatment led to minimal growth defects in both
74 cell types (**Fig S1A**). Consistent with prior studies (Maurano et al. 2015), DNMT1i treatment
75 increased CTCF binding at thousands of sites, termed “reactivated peaks” due to their overlap
76 with CTCF peaks in other cell types (**Fig 1B,C,** 2,079 up- and 2 down-regulated in K562, 6,686
77 up- and 328 down-regulated in HCT116). As expected, reactivated CTCF sites were largely
78 methylated in wild-type cells (**Fig S1B**), and many CTCF binding motifs enriched within these

79 peak regions were methylated at two key CpG sites implicated in CTCF binding (**Fig 1D**)
80 (Spracklin et al. 2022; Wang et al. 2012). We grouped these motifs using K-means clustering of
81 methylation patterns and found clusters with methylation at zero, one, or two positions (**Fig 1D**).
82 For both cell types, CTCF emerged most strongly over the dual-methylated motifs in cluster 1
83 (**Fig 1E**). By contrast, cluster 4, which contains the lowest methylation levels, may include CTCF
84 peaks that are blocked by heterochromatin, as reported recently (Spracklin et al. 2022). Overall,
85 DNMT1i treatment led to the emergence of thousands of reactivated peaks at previously
86 predominantly methylated sites.
87
88 We next evaluated the looping patterns of reactivated CTCF peaks. CTCF HiChIP revealed that
89 a large subset of these reactivated peaks resides within at least one loop anchor and these are
90 henceforth called “looping” peaks (see Methods) (Lareau and Aryee 2018a, 2018b). An example
91 of reactivated looping CTCF peaks is shown at the *PCDHA* locus (**Fig 1F**). On average, loops
92 that are anchored by reactivated peaks increase in strength upon DNMT1 inhibition (**Fig 1G**).
93 Analysis of our published Hi-C data in K562 following 3 days of DNMT1i treatment also revealed
94 an increase in Hi-C interaction signal centered on reactivated peaks (**Fig 1H**). Additionally, loops
95 overlapping reactivated CTCF peaks are shorter than other CTCF loops, with motif cluster 1
96 reactivated peaks forming longer range loops than other reactivated peaks (**Fig S1C**). Loops
97 overlapping reactivated CTCF peaks are also more likely to be tandem or divergent loops (**Fig 1I**)
98 (Tang et al. 2015). Altogether, selective DNMT1 inhibition leads to the emergence of thousands
99 of reactivated CTCF peaks that engage in new chromatin interactions with distinctive features.
100



101

Figure 1 | DNMT1 inhibition reactivates looping CTCF peaks

A. Schematic demonstrating the experimental workflow. K562 and HCT116 cells were treated with DNMT1i for 3 days followed by CTCF ChIP-seq and CTCF HiChIP.

B. Replicate-averaged input-normalized reads within merged peaks for CTCF ChIP-seq in DNMT1i (y-axis) vs. DMSO (x-axis) for K562 cells (left) and HCT116 cells (right). Reactivated peaks are highlighted in red.

C. Example IGV CTCF ChIP-seq tracks ± DNMT1i at the MEG3 locus for K562 (top) and HCT116 (bottom). Reactivated peaks indicated with arrows.

D. Percent methylation status of CTCF motifs underlying reactivated peaks clustered by K-means.Number of CTCF sites per cluster: K562 (cluster 1: 139, cluster 2: 351, cluster 3: 221, cluster 4: 738); HCT116 (cluster 1: 405, cluster 2: 1005, cluster 3: 746, cluster 4: 2249). Bisulfite data for HCT116 from GEO GSM3317488, K562 from ENCODE ENCFF459XNY

E. CTCF replicate-averaged $\log_2$(fold-change DNMT1i/DMSO) RPKM normalized ChIP-seq signal over motif clusters from D.

F. Example IGV CTCF ChIP-seq and HiChIP tracks ± DNMT1i at the Protocadherin Alpha (PCDHA) locus (HCT116). Reactivated peaks are indicated below. Red indicates a looping reactivated peak, whereas grey indicates a nonlooping reactivated peak.

G. Boxplot showing the $\log_2$(fold-change) in loop strength DNMT1i vs. DMSO (y-axis) for loops with or without at least one anchor overlapping a reactivated peak (x-axis) for K562 (left) and HCT116 (right).

H. Heatmaps depicting $\log_2$(mean observed/expected) Hi-C interaction frequency centered on reactivated peaks for DMSO (left) and DNMT1i (right). K562 Hi-C data from Siegenfeld et al., 2022.

I. Orientations of CTCF peaks in loops for which a single CTCF motif was assigned to each anchor. + indicates a CTCF motif on the sense DNA strand, and - indicates a CTCF motif on the antisense strand.

In (**G**), the interquartile range (IQR) is depicted by the box with the median represented by the center line. Outliers are excluded. P values were calculated by a Mann-Whitney test and are annotated as follows: ns: not significant; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

Reactivated CTCF peaks form loops on genes and interact with highly-looping partners

To better understand how methylation-sensitive CTCF occupancy influences chromatin structure, we further characterized reactivated CTCF peaks and their loops. Consistent with previous findings using azacitidine (Maurano et al. 2015), loss of DNA methylation through DNMT1i treatment led to the preferential reactivation of CTCF on gene bodies, particularly in non-intron regions (exons, promoters, and TTS, **Fig 2A, S2A**). This enrichment on gene bodies was largely driven by looping reactivated CTCF peaks (**Fig 2A, Fig S2B, S2C**), highlighting an underappreciated propensity for CTCF peaks on gene bodies to engage in chromatin loops.

We next scrutinized the CTCF partners that loop to reactivated CTCF peaks. Notably, the CTCF partners of reactivated peaks tend to engage in a greater total number of different loops compared to all CTCF peaks (i.e. have more looping partners, **Fig 2B**). This trend is in part driven by the

tendency of non-intronic gene body CTCF peaks to more frequently participate in loops with highly-looping CTCF partners in comparison to intronic or intergenic CTCF peaks (**Fig S2D**). To characterize these highly-looping partners, we ranked all CTCF peaks by the number of different partners they interact with and called those with the most different partners 'highly-looping', analogously to how superenhancers or supersilencers were previously characterized (**Fig S2E, Fig 2C**) (Cai et al. 2021; Hnisz et al. 2013). HiChIP loops connecting gained peaks with highly-looping peaks increased in strength upon DNMT1i, consistent with increased interactions between these sites (**Fig S2F**).

The increased propensity for a given CTCF site to engage in many contacts may reflect architectural stripes, a chromatin structural feature identified from Hi-C data that is postulated to result from persistent loop extrusion at a single region (Vian et al. 2018; Yoon et al. 2022; Barrington et al. 2019). Stripe anchors are enriched for chromatin accessibility and histone modifications associated with active enhancers (Vian et al. 2018). Similarly, highly-looping CTCF peak regions are accessible and enriched for active chromatin marks such as H3K9ac and H3K27ac (**Fig 2D**). Furthermore, based on our previously published Hi-C data, highly-looping CTCF sites have a stronger propensity to engage in "stripe-like" chromatin interactions compared to ordinary CTCF sites and, on a global scale, do not appear to be perturbed upon DNMT1 inhibition (**Fig 2E**). In addition, stripe anchors identified from these Hi-C data for the DNMT1i treatment condition reside closer to highly-looping CTCF peaks than all peaks, consistent with the similarities between these features (**Fig 2F**). Altogether, these results show that upon DNA demethylation, reactivated CTCF sites form loops on gene bodies to highly-looping partners associated with active regulatory marks and stripe anchors.

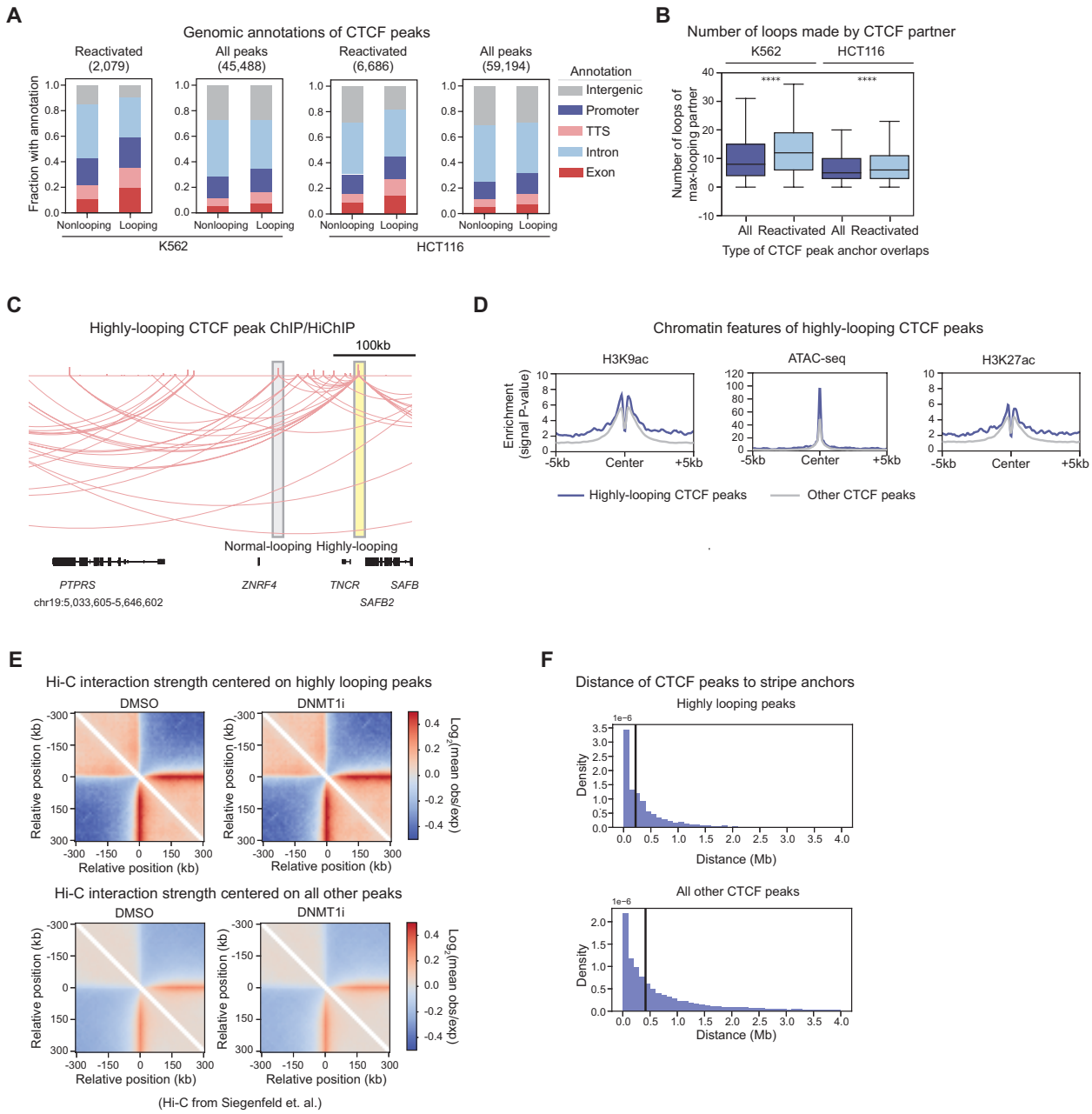


Figure 2 | Reactivated CTCF peaks loop from gene bodies and interact with stripe anchors

A. Proportion of looping and nonlooping peaks with each annotation (y-axis) for both reactivated and all CTCF peaks in K562 (left) and HCT116 (right) in DNMT1i.

B. Number of different loops (loops to different partners) the maximally-looping partner peak makes (y-axis) for all vs. reactivated CTCF peaks in DNMT1i for K562 (left) and HCT116 (right).

C. IGV representation of highly-looping (yellow) and normal-looping CTCF peaks (grey) in K562 DNMT1i-treated cells. CTCF ChIP-seq and HiChIP are shown.

D. Aggregate profile (signal P-value, y-axis) of genomic features over highly-looping peaks (blue) and all other CTCF peaks (grey). CTCF peaks were called in K562 DNMT1i treatment, and histone modifications/ATAC-seq data are from public datasets in K562 (see methods).

E. Heatmap depicting $\log_2$(mean observed/expected) Hi-C interaction frequency centered on highly-looping (top) and all other (bottom) peaks for DMSO (left) and DNMT1i (right) treatment in K562 cells. K562 Hi-C data from Siegenfeld et. al., 2022.

F. Histogram showing density (y-axis) of genomic distances (x-axis) of CTCF highly-looping peaks (left) and all other CTCF peaks (right) in DNMT1i-treated K562 cells from stripe anchors identified from DNMT1i-treated K562 Hi-C data. Median distance is shown with a solid black line. K562 LIMe-Hi-C data from Siegenfeld et. al., 2022.

In (**B**), the interquartile range (IQR) is depicted by the box with the median represented by the center line. Outliers are excluded. P values were calculated by a Mann-Whitney test and are annotated as follows: ns: not significant; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

A subset of genes depends upon CTCF for upregulation upon DNA demethylation

Prior work has shown that a majority of methylated CTCF motifs remain unoccupied even upon DNA demethylation and that the CTCF peaks that do reemerge are often present in other cell types, suggesting that cell-type specific DNA methylation may protect a “labile” set of CTCF peaks from reactivation (Maurano et al. 2015). Although methylation-dependent CTCF occupancy on gene bodies has been linked to changes in alternative splicing and polyadenylation, a global understanding of the functions of methylation-sensitive CTCF peaks remains lacking (Shukla et al. 2011; Nanavaty et al. 2020; Soto and Lipscombe 2020). Thus, we considered whether reactivated CTCF peaks in genic regions impact gene expression upon DNMT1 inhibition.

To assess correlations between reactivated CTCF binding, looping, and transcription, we performed full-length RNA-seq following 3 days of DNMT1i treatment. As expected, DNMT1 inhibition primarily caused gene upregulation, with 643 genes up- and 32 down-regulated in K562 and 3,266 genes up- and 553 down-regulated in HCT116 (Padj <0.05, |log2(fold-change)| > 1) (**Fig S3A**). On average, the genes on which reactivated CTCF peaks bind modestly increase in expression upon DNMT1i treatment relative to all genes overlapping any CTCF site (**Fig 3A, S3B**).

157

158 To more directly investigate the effects of reactivated CTCF peaks, we examined the impact of
159 DNMT1i treatment on gene expression after acute depletion of CTCF. To do so, we first generated
160 a HaloTAG-CTCF homozygous knock-in HCT116 cell line, which enabled potent degradation of
161 CTCF within 8 hours of addition of the HaloPROTAC3 ligand (**Fig 3B, S3C**). We pretreated these
162 cells with HaloPROTAC3 or vehicle for 8 hours before co-treating with DNMT1i or vehicle for 24
163 hours (32 hours total for HaloPROTAC3 treatment) and then performed RNA-seq. This shorter
164 timepoint was used to capture direct effects of CTCF degradation and DNA demethylation as well
165 as avoid toxicity caused by CTCF depletion. Consistent with prior findings (Nora et al. 2017),
166 acute CTCF degradation altered the expression of hundreds of genes with approximately equal
167 numbers up- and down-regulated (488 up-regulated, 338 down-regulated, Padj <0.05, $|\log_2$fold-
168 change$|$ > 0.5), with downregulated genes possessing higher baseline levels of CTCF at their
169 promoters (**Fig S3D**). As expected, DNMT1i treatment in this cell line for 24 hours primarily
170 resulted in gene upregulation, with 135 genes upregulated and 8 genes downregulated (Padj
171 <0.05, $|\log_2$fold-change$|$ > 0.5).

172

173 We focused our analysis on genes that were differentially expressed in DNMT1i versus DMSO
174 treatment in the presence of CTCF. Using K-means clustering, we identified four gene clusters
175 defined by unique z-score normalized expression patterns across the conditions (**Fig 3C**).
176 Clusters 1, 2, and 4 comprise genes that increase in expression upon DNMT1 inhibition, whereas
177 Cluster 3 contains the few genes that decrease in expression upon DNMT1 inhibition**.** Moreover,
178 genes in Cluster 1 are upregulated by CTCF depletion, whereas genes within Clusters 2, 3, and
179 4 are unaffected by CTCF depletion.

180

181 We focused on Clusters 2 and 4, which contain lowly-expressed genes that are upregulated upon
182 DNMT1 inhibition and are unaffected by CTCF depletion in the absence of DNMT1i (**Fig 3C, S3E**).

183 Significantly, while genes in Cluster 4 are upregulated upon DNMT1 inhibition with or without
184 CTCF present, genes in Cluster 2 depend on CTCF for DNMT1i-mediated upregulation. Cluster
185 2 genes are enriched for gene ontology annotations associated with development (**Fig S3F**). To
186 determine how CTCF binding might be related to the observed changes in gene expression, we
187 performed CTCF ChIP-seq in our HaloTag-CTCF HCT116 cells following 24 hours of DNMT1i or
188 vehicle treatment. Notably, genes in Cluster 2 preferentially gained CTCF occupancy upon
189 DNMT1 inhibition compared to genes in other clusters (**Fig 3C, S3G**). Altogether, these results
190 suggest that CTCF reactivation may contribute to DNMT1i-mediated gene activation.
191
192 To gain further insight into the features that differentiate gene Clusters 2 and 4, we next
193 considered their subnuclear localization since genic positioning within the nucleus has been linked
194 to transcriptional regulation(Lochs et al. 2019; Ilık and Aktaş 2021). Specifically, we investigated
195 the localization of genes relative to a prominent euchromatic nuclear landmark, nuclear speckles.
196 Nuclear speckles are condensate bodies of protein and RNA that have been implicated in
197 transcription and splicing (Chen et al. 2018; Chen and Belmont 2019; Zhang et al. 2020; Wang et
198 al. 2021; Spector and Lamond 2011; Alexander et al. 2021). We compared our gene clusters to
199 publicly available data from SON TSA-seq, which measures the distance of genomic regions to
200 speckles using tyramide signal amplification initiating at the core speckle protein SON.
201 Intriguingly, genes in Cluster 2 have higher SON TSA-seq levels than genes within Cluster 4 (**Fig**
202 **3D**). Genes in cluster 1, which are upregulated by CTCF depletion alone, are closest to nuclear
203 speckles. Taken together, these data support the notion that a subset of genes depend on CTCF
204 for expression upon DNMT1i treatment and are located closer to nuclear speckles than CTCF-
205 independent genes.

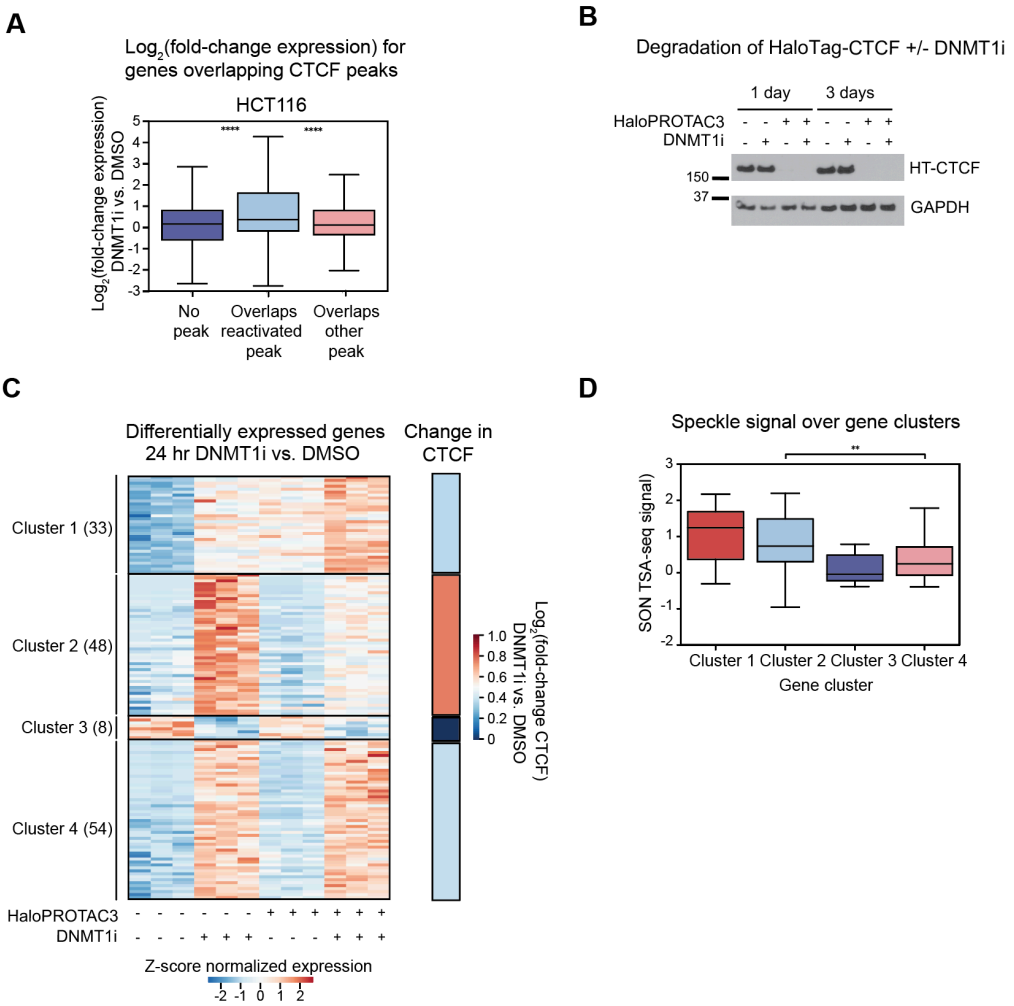


Figure 3 | Many genes depend on CTCF for upregulation upon DNMT1 inhibition

A. Boxplot showing $\log_2$(fold-change) in gene expression in DNMT1i vs. DMSO (y-axis) for genes overlapping no CTCF peak, reactivated CTCF peaks, or non-reactivated CTCF DNMT1i peaks (x-axis) for HCT116 cells.

B. Western blot showing the depletion of CTCF in HaloTag-CTCF knock-in cell line in a HaloPROTAC3-dependent and DNMT1i-independent manner (330 nM HaloPROTAC3, 10uM DNMT1i)

C. Left: Heatmap showing z-score normalized r-log transformed counts for genes that are differentially expressed in DNMT1i vs. DMSO non-HaloPROTAC3 conditions (Padj <0.05, $|\log_2\text{Foldchange}| > 0.5$) clustered by K-means. Number of genes per cluster are as follows: cluster 1: 33, cluster 2: 48, cluster 3: 6 cluster 4: 54. Right: Average log2(fold-change) in CTCF binding in DNMT1i vs. DMSO for all CTCF peaks on genes within the four gene clusters.

D. Boxplot showing normalized SON TSA-seq signal (20 kb bins, y-axis, 4D Nucleome 4DNFIBY8G6RZ) in wildtype HCT116 cells over genes within the different clusters identified in **C**.

For (**A**, **D**), the interquartile range (IQR) is depicted by the box with the median represented by the center line. Outliers are excluded. P values were calculated by a Mann-Whitney test and are annotated as follows: ns: not significant; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

206

207

208 **Reactivated peaks and their highly-looping partners are located near nuclear speckles**

209

210 Our findings suggest an interplay between CTCF reactivation and gene expression, particularly
211 for genes located near nuclear speckles. Moreover, we also had found that reactivated peaks
212 loop to highly-looping partners in active genomic regions and stripe anchors. As such, we
213 considered whether reactivated CTCF peaks and their highly-looping partners were also located
214 near nuclear speckles. Recent studies have shown that nuclear speckles are hubs of DNA-DNA
215 contacts and regions near speckles exhibit high CTCF occupancy (Spector and Lamond 2011;
216 Brown et al. 2008; Ilık and Aktaş 2021; Ilik et al. 2020; Faber et al. 2022; Quinodoz et al. 2018;
217 Kim et al. 2020). Additionally, CTCF can form stress-responsive clusters near nuclear speckles
218 in some cell types, and genomic subcompartments associated with speckles have been
219 correlated with increased CTCF binding and looping (Jabbari et al. 2019; Wang et al. 2021;
220 Lehman et al. 2021). However, a causal and direct connection between nuclear speckles and
221 CTCF binding and looping has not been explored.

222

223 To determine if reactivated peaks and highly-looping CTCF peaks are preferentially located near
224 speckles, we performed Cut&Tag for SON in K562 and HCT116 cells with and without DNMT1
225 inhibition (**Fig 4A**) (Kaya-Okur et al. 2019; Chen et al. 2018; Zhang et al. 2020; Ilik et al. 2020).
226 Consistent with localization near speckles, reactivated peaks appeared in regions of higher SON
227 signal compared to all CTCF peaks (**Fig 4B**). This trend is particularly evident for looping
228 reactivated CTCF peaks (**Fig S4A**). This enrichment of looping reactivated peaks near speckles
229 is further corroborated by publicly available SON TSA-seq data (**Fig S4B**) (Zhang et al. 2020) and
230 by an enrichment of K562 reactivated peaks in the A1 genomic subcompartment, which is known

231 to be located near speckles (**Fig S4C**) (Xiong and Ma 2019). Moreover, upon 24 hours of DNMT1i
232 treatment in our HaloTag-CTCF HCT116 knock-in cell line, reactivated peaks gain SON Cut&Tag
233 signal (**Fig 4C)**. This gain in SON signal is blocked when CTCF is degraded by HaloPROTAC3
234 treatment, suggesting that SON association of these sites upon DNMT1 inhibition is dependent
235 on CTCF. Thus, in addition to reactivated CTCF peaks occurring closer to nuclear speckles in
236 comparison to non-reactivated peaks, these regions also gain SON association in a CTCF-
237 dependent manner.
238
239 Strikingly, relative to all CTCF peaks, highly-looping CTCF peaks are enriched at nuclear speckles
240 (**Fig 4D, S4D**). More broadly, the number of different loops a CTCF peak engages in and their
241 corresponding looping strength are positively correlated with SON Cut&Tag signal (**Fig 4E, F**). In
242 addition, we observed that CTCF peaks with strong Cut&Tag SON signal preferentially engage in
243 strong shorter-range interactions in published Hi-C interaction data (**Fig 4G**). Overall, our results
244 reveal that CTCF-dependent genes, reactivated CTCF peaks, and their highly-looping partners
245 are preferentially located near nuclear speckles.
246
247
248
249
250
251
252
253
254
255
256

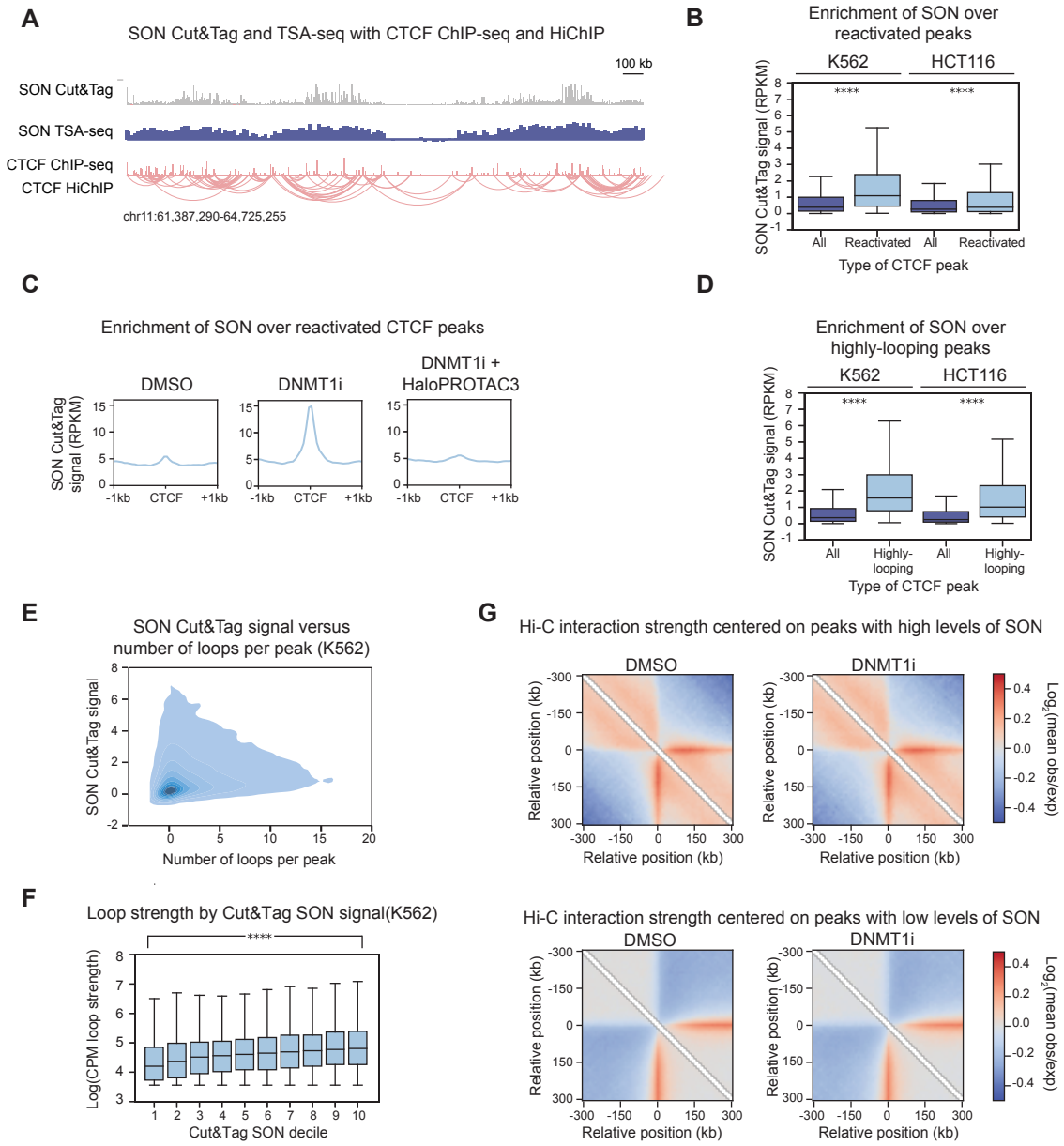


Figure 4 | Reactivated peaks and CTCF loops are located near nuclear speckles

A. Example IGV tracks depicting SON Cut&Tag signal (top), SON TSA-seq normalized counts (middle), CTCF ChIP-seq (bottom), and CTCF HiChIP for a region on chromosome 11 for K562 DMSO.

B. Boxplot showing replicate-averaged SON Cut&Tag signal (RPKM, 20 kb bins, y-axis) after DNMT1i treatment in the respective cell type over reactivated peaks vs. all other CTCF peaks called in DNMT1i treatment for K562 and HCT116 cells.

C. Aggregate profile of replicate-averaged SON Cut&Tag signal (RPKM) over 1-day reactivated CTCF peaks in HaloTag-CTCF HCT116 cells with and without CTCF degradation through HaloPROTAC3.

D. Same as **B**, but for highly-looping peaks vs. all other CTCF peaks.

E. Density plot showing the correlation between SON Cut&Tag signal (RPKM, 20 kb bins) over a given CTCF peak (y-axis) vs. the number of different loops (partners) assigned to that CTCF peak (x-axis) for K562 DNMT1i treatment.

F. Boxplot showing average $\log_2$(counts per million (CPM) loop strength) (y-axis) for CTCF HiChIP loops relative to SON Cut&Tag signal decile (x-axis). logCPM defined by Diffloop across all conditions (DMSO and DNMT1i). Loops are segregated into equally sized deciles by Cut&Tag signal in DMSO (RPKM, 20kb bins).

G. Heatmap depicting $\log_2$(mean observed/expected) Hi-C interaction frequency centered on CTCF peaks at speckles (top) and not at speckles (bottom) for DMSO (left) and DNMT1i (right) treatment. K562 Hi-C data from Siegenfeld et. al., 2022.

For boxplots, the interquartile range (IQR) is depicted by the box with the median represented by the center line. Outliers are excluded. P values were calculated by a Mann-Whitney test and are annotated as follows: ns: not significant; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

Acute disruption of nuclear speckles alters gene expression without disrupting CTCF

To explore the observed connection between speckles and CTCF looping, we tested whether nuclear speckles might play a causal role in promoting CTCF occupancy and/or looping. To accomplish this, we sought to acutely degrade speckles to evaluate whether speckles actively maintain CTCF loops while mitigating confounding factors associated with mitotic arrest caused by long-term SON depletion (Ahn et al. 2011). Consequently, we engineered a K562 knock-in cell line with the core speckle proteins, SON and SRRM2, both endogenously tagged with FKBP12^{F36V} for use with the dTAG degradation system (Ilik et al. 2020; Nabet et al. 2018). Upon addition of dTAG-13, FKBP12^{F36V}-tagged proteins are recruited to the CRBN-Cul4-Ring E3 ligase and undergo acute proteasomal degradation (**Fig 5A**). Treatment of the K562 FKBP12^{F36V} double knock-in cell line with dTAG-13 resulted in near-complete degradation of both SON and SRRM2 within 6 hours (**Fig 5B, S5A, S5B**).

We next assessed if speckle disruption directly alters gene expression by performing RNA-seq after 6 hours of dTAG-13 treatment. Upon speckle degradation, 541 genes were down-regulated and 184 genes were up-regulated (padj < 0.05, $|\log_2$fold-change$|$ > 0.25) (**Fig 5C, S5C**).

274 Consistent with a role of nuclear speckles in mediating active transcription, a large majority of
275 down-regulated genes were located within the closest decile to nuclear speckles by TSA-seq (**Fig**
276 **S5D**), and genes with the highest SON Cut&Tag signal exhibited the strongest decrease in
277 expression (**Fig 5D**). These results show that disrupting nuclear speckles has immediate effects
278 on transcription.
279
280 To determine whether speckle disruption alters CTCF binding and/or looping, we performed
281 CTCF ChIP-seq and HiChIP in the speckle dTAG cell line after treatment with dTAG-13 or DMSO
282 for 6 hours. In contrast to the immediate transcriptional effects, CTCF occupancy remained
283 unchanged upon speckle degradation, with only 2 significantly downregulated peaks (**Fig 5E,**
284 **S5E**). CTCF HiChIP also revealed that loop strength was largely unchanged across speckle
285 Cut&Tag and TSA-seq quantiles (**Fig 5F, S5F**) despite a positive correlation between speckle
286 association and baseline loop strength (**Fig 4E,F**). Even after 12 hours of dTAG treatment, there
287 was only a very subtle decrease in CTCF looping near speckles (**Fig S5F**). In addition, changes
288 in loop strength and CTCF binding were not associated with changes in gene expression (**Figure**
289 **S5G,H**). Moreover, depleting CTCF for 24 hours in our HaloTag-CTCF HCT116 knockin cell line
290 did not noticeably change speckle morphology (**Fig S5I**). Altogether, despite the strong
291 relationship between CTCF looping and speckle association, acute disruption of nuclear speckles
292 did not appreciably disrupt CTCF binding or looping, suggesting that maintenance of 3D CTCF
293 loop structure is independent of speckle association.
294
295
296
297
298

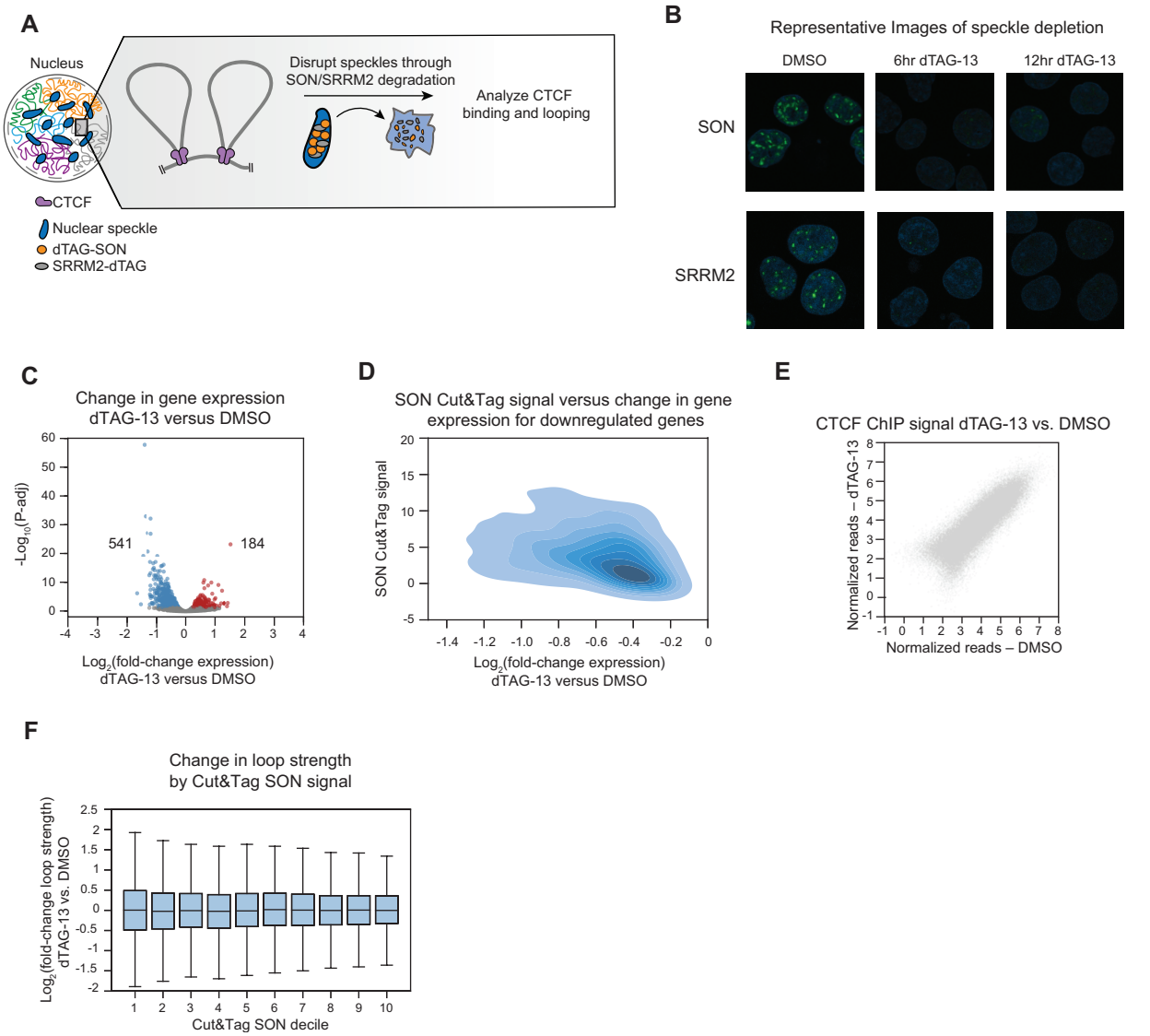


Figure 5 | Acute depletion of nuclear speckles alters gene expression without disrupting CTCF looping.

A. Schematic illustrating use of the dTAG system to acutely deplete SON and SRRM2, thus disrupting nuclear speckles.

B. Immunofluorescence for SON and SRRM2 for K562 speckle knock-in cells treated with dTAG-13 for 6 or 12 hours to deplete SON and SRRM2.

C. Volcano plot showing -$\log_{10}$(adjusted P-value) (y-axis) vs. $\log_2$(fold-change) in gene expression (x-axis) upon speckle degradation with 6 hours dTAG-13 treatment. Upregulated genes are highlighted in red, and downregulated genes are highlighted in blue (adjusted p-value <0.05, $|\log_2\text{Foldchange}| > 0.25$).

D. Density plot showing SON Cut&Tag signal in wt K562 cells treated with DMSO (RPKM 20kb bins, y-axis) vs. the change in gene expression (x-axis) for genes that decrease in expression upon dTAG-13 treatment (Spearman R, -0.43).

E. Replicate-averaged input normalized tags between CTCF ChIP-seq in dTAG-13 treatment (x-axis) vs. DMSO (y-axis) for 6 hour K562 speckle dTAG knock-in cells.

F. Boxplot showing log_2(fold-change) in loop strength for CTCF HiChIP loops in speckle knock-in K562 cells treated with dTAG-13 vs. DMSO (y-axis) relative to SON Cut&Tag signal (x-axis). Loops are segregated into equally sized deciles by Cut&Tag signal with 10 representing the decile closes to speckles (RPKM, 20kb bins).

In (**F**), the interquartile range (IQR) is depicted by the box with the median represented by the center line. Outliers are excluded. P values were calculated by a Mann-Whitney test and are annotated as follows: ns: not significant; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

Discussion

DNA methylation accounts for much of the variation in CTCF binding across cell types (Maurano et al. 2015), but the functions and features of methylation-sensitive CTCF peaks are incompletely understood. Here we leveraged a selective DNMT1 inhibitor (Pappalardi et al. 2021; Portilho et al. 2021) to gain new insights into how reactivated CTCF peaks interface with chromatin looping and transcription. This new inhibitor is a valuable tool for reactivating CTCF without the confounding effects caused by the covalent nature and toxicity of earlier hypomethylating agents.

We found that reactivated CTCF peaks preferentially form loops on gene bodies and interact with promiscuous partner peaks at stripe anchors. Architectural stripes may contribute to transcription by promoting enhancer contacts with gene bodies (Vian et al. 2018; Yoon et al. 2022; Barrington et al. 2019; Kraft et al. 2019), and may also represent a mechanism to poise transcription (Barrington et al. 2019). Thus, we propose that reactivated peaks enable transcriptional activation by interacting with regulatory elements through highly-looping CTCF partner peaks (**Fig 6**). However, DNA demethylation causes pleiotropic effects ranging from changes in gene expression to transcription factor binding (Yin et al. 2017; Portilho et al. 2021), and we cannot rule out that these other features could contribute to the effects we observe.

318 To investigate whether CTCF is causatively linked to gene expression, we coupled DNMT1
319 inhibition with CTCF degradation and unexpectedly found that a large fraction of differentially
320 expressed genes depend on CTCF for activation upon DNMT1 inhibition. These genes
321 preferentially gain CTCF binding, consistent with a role of CTCF binding in promoting gene
322 expression. More broadly, our work unmasked a subset of genes that rely on CTCF for expression
323 upon global demethylation but do not depend on CTCF under basal conditions. These results
324 help reconcile observations that CTCF is an essential protein but in cell culture its acute depletion
325 causes minimal gene expression changes (Nora et al. 2017). Consequently, CTCF-dependent
326 gene effects may often be obscured because genes are regulated by additional features (such as
327 DNA methylation). In addition, our results suggest that CTCF may play a role in positioning genes
328 close to active regulatory regions and stripe anchors to promote gene activation but may not be
329 strictly necessary for maintaining an active chromatin state.
330
331 Strikingly, CTCF-dependent genes, reactivated CTCF peaks, and their highly-looping partners
332 are located near nuclear speckles, consistent with prior studies showing that subcompartments
333 near nuclear speckles exhibit increased CTCF binding and looping (Wang et al. 2021; Jabbari et
334 al. 2019). Moreover, we show that reactivated CTCF peaks gain SON signal upon DNMT1
335 inhibition in a CTCF-dependent manner, indicating a functional tie between CTCF and speckles.
336 This is in agreement with a recent preprint reporting that CTCF helps establish basal speckle-
337 gene contacts (Yu et al. 2023). These findings add to our model by suggesting that reactivated
338 CTCF peaks could facilitate transcription of genes near speckles, perhaps by mediating the
339 engagement of genic CTCF sites with speckle-associated stripe anchors and/or splicing factors
340 (**Fig 6**). A role of CTCF in facilitating gene-body chromatin loops near speckles is synergistic with
341 proposed models suggesting that intragenic looping contributes to gene splicing (Liao and Regev
342 2020; Ruiz-Velasco et al. 2017). Thus, reactivated CTCF peaks may reorganize genes near
343 nuclear speckles as a means of influencing transcription.

344

345 Previous work knocking down SRRM2 support a causative role of speckles in modulating
346 chromatin structure but did not specifically study changes in CTCF binding or looping (Hu et al.
347 2019). Therefore, to investigate connections between speckles and CTCF, we achieved the first
348 acute degradation of nuclear speckles using the dTAG system. Although acute disruption of
349 speckles supports their direct involvement in transcription, we observed minimal impacts on CTCF
350 binding or looping. This notion is consistent with recent work showing that CTCF clustering and
351 looping is not dependent on phase separation or nuclear bodies (Lee et al. 2021; Ulianov et al.
352 2021) and that transcriptional disruption does not immediately impact chromatin looping or CTCF
353 binding (Jiang et al. 2020; Vian et al. 2018; Luan et al. 2021). Our findings also suggest that the
354 genome interactions disrupted upon SRRM2 depletion (Hu et al. 2019) are not mediated by CTCF
355 or only occur after a longer time. However, it is possible that nuclear speckles contribute to the
356 establishment, but not maintenance, of CTCF loops, which may not be evident in our system due
357 to the time points employed for speckle degradation. Moreover, other features enriched near
358 speckles, such as open chromatin, high GC content, and active gene expression, could instead
359 contribute to increased CTCF binding and looping near speckles. Our knockdown experiments
360 help clarify the directionality of the relationship between speckles and CTCF.

361

362 Taken altogether, our findings highlight a potential mechanism through which DNA methylation
363 modulates genome architecture — by protecting a subset of genic CTCF sites from engaging with
364 promiscuous CTCF peaks located near nuclear speckles. In particular, genes near nuclear
365 speckles may be susceptible to activation through CTCF-dependent mechanisms, and thus might
366 require this layer of protective regulation. Furthermore, the GC-rich composition of non-intron
367 gene bodies might increase the likelihood of CTCF motifs in comparison to other genomic regions,
368 increasing the vulnerability of these sites to ectopic CTCF binding. More broadly, our work

369 nominates new functions for the known antagonism between gene body methylation and CTCF

370 occupancy in regulating transcription through remodeling genome architecture.

371

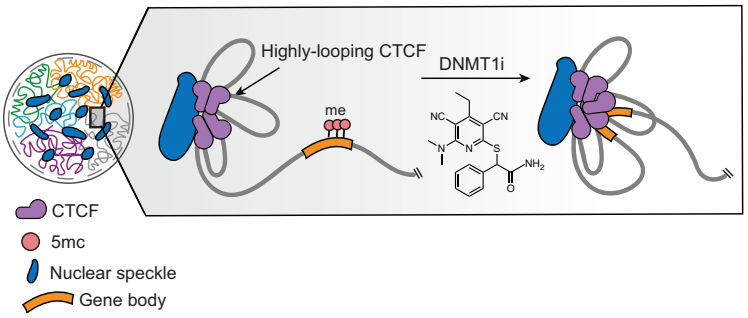


Figure 6 | Model for methylation-mediated insulation of genes from regulatory elements near nuclear speckles

372 **Methods**

373

374 **Cell culture**

375 K562 was obtained from ATCC. HEK 293 T was obtained from Life Technologies. HCT116 was
376 obtained as a gift from M. Shair. Cells were authenticated by Short Tandem Repeat profiling
377 (Genetica) and tested for mycoplasma (Sigma-Aldrich, MP0035). All cell lines were cultured in a
378 humidified 5% CO_2 incubator at 37 °C with media supplemented with 100 U/mL penicillin and
379 100 µg/mL streptomycin (Life Technologies, 15070063). K562 was cultured in RPMI-1640 (Life
380 Technologies, 11875119) with 10% fetal bovine serum (FBS) (Peak Serum, PS-FB2). HEK 293 T
381 was cultured in DMEM (Life Technologies, 1195073) with 10% FBS (Peak Serum, PS-FB2).
382 HCT116 was cultured in McCoy's 5A media (Thermo Fisher Scientific, 16600082) with 10% FBS
383 (Peak Serum, PS-FB2).

384

385 **Generating SON, SRRM2, and CTCF knock-in cell lines**

386 To generate an N-terminal Puro-P2A-dtag SON knock-in donor plasmid, SON homology arms of
387 1 kb were amplified from K562 genomic DNA and cloned into the ph1 backbone Puro-P2A-dTAG,
388 sub-cloned from pCRIS-PITChv2-Puro-dTAG, a gift from J. Bradner and B. Nabet (Addgene #
389 91793). The SON sgRNA, 5'-CACCGCTGCTCGATGTTGGTCGCCA-3' that was previously used
390 for N-terminal SON knock-in was cloned into PX459 by restriction enzyme digestion (Allen
391 Institute,https://www.allencell.org/cell-catalog.html). To generate a C-terminal dTAG-P2A-
392 blasticidin SRRM2 knock-in donor plasmid, SRRM2 homology arms of 1 kb were amplified from
393 K562 genomic DNA and cloned into the ph1 backbone dTAG-P2A-BSD, sub-cloned from pCRIS-
394 PITChv2- dTAG-BSD, a gift from J. Bradner and B. Nabet (Addgene # 91795). The sgRNA,
395 5'CACCGCCATGAGACACCGCTCCTCC -3' that was previously validated for C-terminal SRRM2
396 knock-in was cloned into PX459 by restriction enzyme digestion (Ilik et al. 2020). Notably, this
397 strategy eliminates the last four amino acids of SRRM2, which is necessary due to a sparsity of

398 guides in the region and has been used to study SRRM2 in the context of speckles (Ilik et al.
399 2020). Cells were transfected with both the donor and guide plasmids by nucleofection (Lonza,
400 program T016). Cells were selected 3 days after transfection with puromycin (2 µg/mL) (Thermo
401 Fisher Scientific, A1113803) or blasticidin (8 µg/mL) (Thermo Fisher Scientific, A1113903) for a
402 duration of 5-10 days. Single-cell clones were isolated from the bulk cell line through
403 fluorescence-activated cell sorting. The SON knock-in was generated from wildtype K562 cells,
404 and then the SRRM2 knock-in was introduced on top of the validated SON knock-in clone. To
405 generate the HT-CTCF HCT116 knock-in cell line, the sgRNA 5'-
406 CACCGTGCAGTCGAAGCCATTGTGG-3' was subcloned into PX459 and HaloTAG-CTCF was
407 subcloned into PDONR223. The guide and donor plasmids were transfected into cells using
408 Fugene HD (Promega). After seven days, single cell clones were isolated from the bulk population
409 through fluorescence-activated cell sorting using HaloTag TAMRA labeling to sort potential
410 positive clones.
411
412 **Immunoblotting**
413 Pellets were harvested for western blot analysis and washed with PBS (Corning, 21-040-CV). For
414 immunoblots of CTCF, cells were lysed on ice using RIPA buffer (Boston BioProducts)
415 supplemented with fresh HALT Protease Inhibitor (Thermo Fisher Scientific) and the lysates were
416 cleared through centrifugation. The protein concentration of the lysate was determined using the
417 BCA Protein Assay Kit (Thermo Fisher Scientific) and immunoblots for CTCF were performed
418 according to standard procedures with a 6% SureCast acrylamide page gel. For immunoblots of
419 SON and SRRM2 and corresponding loading controls, 1 million cells per condition were lysed on
420 ice for 30 minutes in 50uL RIPA buffer (Boston Bioproducts, BP-115) supplemented with 1mM
421 PMSF, 2mM EDTA, and 1X Halt Protease inhibitors (Thermo Fisher Scientific, 78429).
422 Subsequently, 50uL of wash buffer was added (50mM Tris pH 8.0, 2mM MgCl, 50mM NaCl, 1mM
423 phenylmethylsulfonyl fluoride (PMSF), 1X Halt Protease inhibitors, 1:500 Benzonase (Millipore-

424 Sigma, E1014-25KU) and the samples were rotated at room temperature for 20 minutes and then
425 clarified by full speed centrifugation at 4 °C. The protein concentration of the lysates was
426 determined using the BCA Protein Assay Kit (Thermo Fisher Scientific, 23225). Immunoblotting
427 was performed according to standard procedures with the following modifications: a 3-8% tris
428 acetate gel (Thermo Fisher Scientific, EA03752) was used and samples were run for 2 hours
429 15min at 80V at 4 °C in 1X tris-acetate running buffer (Thermo Fisher Scientific, LA0041) except
430 for the loading control which was run for 90 minutes. Transfer was conducted in 1X transfer buffer
431 (Research Product International, 501977679) containing 10% methanol and 0.05% sodium
432 dodecyl sulfate (SDS) onto an activated 0.45um PVDF membrane (Millipore sigma IPVH09120)
433 for 2 hours at 15V except for loading control which was transferred in 1X transfer buffer with 20%
434 methanol at 12V for 90 minutes. The primary antibodies used for immunoblotting are as follows:
435 CTCF (1:2000, Cell Signaling Technology D31H2); SON (1:1000, Abcam ab121759); SRRM2
436 (1:1000, ThermoFisher PA5-66827); GAPDH (1:2000, Santa Cruz Biotechnology sc-47724);
437 Vinculin (1:2000, Cell Signaling Technology 13901S).
438
439 **Immunofluorescence**
440 Immunofluoresence was performed according to standard protocol. Briefly, K562 speckle knock-
441 in cells treated with 500 nM dTAG-13 or DMSO for the indicated amount of time were cytospun
442 onto glass coverslips for 5 minutes at 800 x rcf and fixed in 4% PFA for 15 minutes. HCT116 cells
443 treated with 330 nM HaloPROTAC3 or DMSO were instead grown directly on glass coverslips
444 and fixed in 4% PFA for 15 minutes. Cells were then washed 3X in PBS, permeabilized with
445 0.25% Triton-X for 10 minutes, and subsequently washed 3X with PBS. Cells were then blocked
446 in 10% goat serum (Millipore Sigma, G9023) in PBS-T for 1 hour and incubated overnight in 5%
447 goat serum in PBS-T with primary antibody at 4°C (SON (1:300, abcam ab121759) or SRRM2
448 (1:300, ThermoFisher PA5-66827)). Following 3X PBS-T washes, cells were incubated with
449 secondary antibody for 1 hr (1:2000, Invitrgoen alexafluor488 A-11008), washed 3X in PBS-T,

450 and mounted on a coverslip with DAPI (Prolong Diamond Anti-Fade, P36961). Cells were imaged
451 on an LSM180 Confocal microscope. Image quantification was performed using CellProfiler on
452 raw Tiff images to identify nuclei and then quantify average fluorescence signal within each
453 nucleus(Carpenter et al. 2006).
454
455 **Cell growth assays**
456 K562 or HCT116 cells were plated in triplicate in a 96-well plate. Drug or vehicle (DMSO, Sigma
457 Aldrich, D8418) were dosed at the specified concentration for 3 days. After 3 days of treatment,
458 cell viability was measured using CellTiter-Glo (Promega, G7572) with the luminescence detector
459 on the SpectraMax i3x (Molecular Devices) plate reader with the SoftMax Pro (version 6.5.1)
460 software, and data were processed using Prism.
461
462 **ChIP-seq**
463 CTCF ChIP-seq was performed in duplicate. For the CTCF reactivation experiments, K562 or
464 HCT116 wildtype or HT-CTCF HCT116 knock-in cells were treated with 10 μM GSK3482364 or
465 0.1% DMSO control for the indicated time. For the speckle depletion experiments, K562 speckle
466 dTAG knock-in cells were treated with 500 nM dTAG-13 or 0.1% DMSO control for 6 or 12 hours.
467 After drug treatment, cells were resuspended and fixed at a density of 10^6 cells/mL in media
468 containing 10% FBS (Peak Serum, PS-FB2) and 1% formaldehyde (Sigma Aldrich, F8775) for
469 10 min with rotation and quenched with 0.2 M glycine and washed in PBS. 5 million cells per
470 condition were lysed on ice for 10 min in 600 uL of ChIP lysis buffer (50 mM Tris-HCl pH 7.5, 1%
471 SDS, 0.25% NaDOC) with HALT protease inhibitors (Thermo Fisher Scientific, 78429). Cells were
472 then diluted to 2mL with ChIP dilution buffer (50 mM Tris-HCL pH 7.5, 0.01% SDS, 150 mM NCl,
473 0.25% triton X-100, 1X HALT protease inhibitor) and sonicated at 4 °C with a Branson Sonicator
474 (45% amplitude, 0.7s on, 1.3s off, total time 5 min). The supernatant was clarified by centrifugation
475 at 16000 x rcf for 10 min at 4 °C . The sample was then diluted into 6 mL total with ChIP Dilution

476 Buffer and the final concentration of triton was brought to 1%. 50 µL was removed as an input
477 sample and kept at 4 °C overnight. Immunoprecipitations were carried out on the remaining
478 sample overnight at 4 °C with rotation with 5 uL of antibody (Cell Signaling Technology D31H2
479 anti-CTCF, 3418S). 50uL of Protein G Dynabeads (Thermo Fisher Scientific, 10004D) were
480 added to each sample and incubated at 4 °C with rotation for 3 hours. Supernatant was discarded
481 on a magnet and beads were then washed twice with ice-cold RIPA wash buffer (10 mM Tris-HCl
482 pH 8.1, 0.1% SDS, 150 mM NaCl, 0.1 NaDOC, 1% Triton X-100, 1 mM EDTA), once with ice-cold
483 LiCl wash buffer (10 mM Tris-HCl, pH 8.1, 250 mM LiCl, 0.5% Triton X-100, 0.5% NaDOC), and
484 once with ice cold TE buffer (10mM Tris-HCL pH 8.0, 50mM NaCl, 1mM EDTA). Chromatin was
485 eluted in 100 µL of Elution buffer (10 mM Tris-HCl pH 8.0, 0.1% SDS, 150 mM NaCl, 1 mM EDTA)
486 at 65 °C for 1 hour at 1000rpm. Input samples were brought to 100µL with Elution buffer and
487 incubated alongside ChIP samples. Supernatant was removed and 50 µg of RNAse A (Sigma
488 Aldrich, 10109142001) was added and incubated for 30 min at 37 °C. Next, 25 µg of Proteinase
489 K (Invitrogen, 25530015) was added, and the samples were incubated 63 °C for 3 hours with
490 1000rpm agitation. DNA was subsequently purified by a 2X SPRI (Omega Bio-Tek, M1378-01).
491 Sequencing libraries were then prepared from 5 ng of DNA by performing end-repair with the End-
492 it DNA End-Repair Kit (Lucigen, ER81050), followed by A tailing with NEB Klenow Fragment (3'−5'
493 exo-) (New England Biolabs, M0212), adapter ligation with NEB DNA Quick Ligase (New England
494 Biolabs, M2200), and PCR using NEB ultra mastermix (M0544S) and 10X KAPA Library Amp
495 Primer Mix (Kapa Biosystems, KK2623). Samples were sequenced using a NovaSeq SP kit
496 (Illumina) with 50 bp paired-end reads at a sequencing depth of approximately 25 million reads
497 per sample.
498
499 **HiChIP**
500 Cells were treated in duplicate. For the DNMT1i experiment, K562 or HCT116 cells were treated
501 with 10 µM GSK3482364 (ChemieTek) or 0.1% DMSO control for 3 days. For the speckle degron

502 experiment, K562 speckle dTAG knock-in cells were treated with 500 nM dTAG-13 (Sigma
503 Aldrich, SML2601) (or 0.1% DMSO control for 6 or 12 hours. CTCF HiChIP was performed on
504 each sample according to a published protocol (Mumbach et al. 2016). Specifically, cells were
505 resuspended and fixed at a density of 10^6 cell/mL in media containing 10% FBS (Peak Serum,
506 PS-FB2) and 1% formaldehyde (Sigma Aldrich, F8775) for 10 min with rotation and quenched
507 with 0.2 M glycine and washed in PBS. 10 million cells per condition were resuspended in ice-
508 cold Hi-C Lysis Buffer (10mM Tris HCl pH 8.0, 10mM NaCl, 0.2% NP-40, 1X protease inhibitor)
509 and lysed at 4 °C for 30 minutes. Tubes were centrifuged and the pellets were washed with 500
510 µL ice-cold Hi-C lysis buffer. Pellets were then resuspended in 100µL of 0.5% SDS and incubated
511 at 62 °C for 10 minutes, after which they were quenched with Triton X-100 and incubated at 37 °C
512 for 15 minutes. The samples were then digested with 375 units of DpnII (New England Biolabs,
513 R0543L) for 2 hours at 37 °C with shaking at 900 rpm and then enzyme was heat inactivated at
514 65 °C for 20 minutes. Digested ends were filled in with biotinylated dATP (Life Technologies,
515 19524-016) for 1 hour at 37 °C at 900rpm, and in situ ligation was performed with T4 DNA ligase
516 (New England Biolabs, M0202L) for 4 hours at room temperature with rotation. Nuclei were
517 pelleted by centrifugation and resuspended in Nuclear Lysis Buffer (50mM Tris-HCL pH 7.5,
518 10mM EDTA, 1% SDS, 1X protease inhibitor) and sheared in Covaris Millitubes (cat # 520135)
519 on a Covaris S220 sonicator with Fill Level 10, Duty Cycle 5, PIP 140, Cycles/Burst 200, for 4
520 minutes each. Samples were clarified by centrifugation and supernatant was split into two tubes
521 and diluted with 2X the volume of ChIP Dilution Buffer (0.01% SDS, 1.1% triton X 100, 1.2mM
522 EDTA, 16.7mM tris pH 7.5, 167mM NaCl). 30ul Pre-washed Protein G Dynabeads (Thermo Fisher
523 Scientific, 10004D) were then added to each tube and samples were rotated for 1 hour at 4 °C.
524 Supernatant was transferred to new tubes and antibody was added and allowed to bind overnight
525 at 4 °C with rotation (15uL antibody per tube, Cell Signaling Technology D31H2 anti-CTCF,
526 3418S). 30uL of pre-washed Protein G Dynabeads were added to each tube and rotated at 4 °C
527 for 2 hours. Beads were washed 3X each with Low Salt Wash Buffer (0.1% SDS, 1% triton X 100,

528 2 mM EDTA, 20 mM Tris HCL pH 7.5, 150 mM NaCl), High Salt Wash Buffer (0.1% SDS, 1%
529 triton X 100, 2 mM EDTA, 20mM Tris HCl pH 7.5, 500 mM NaCl), and LiCl Wash Buffer (10mM
530 tris pH 7.5, 250 mM LiCl, 1% NP-40, 1%Na-DOC, 1mM EDTA) at room temperature. Tubes for
531 each sample were recombined and chromatin was eluted into 100 µL of Elution Buffer (50mM
532 $NaHCO_3$, 1% SDS) per sample by incubating at room temperature for 10 minutes with rotation
533 followed by 3 minutes at 37 °C shaking. Elutions were repeated for 200 µL total volume. 10 µL of
534 Proteinase K were added per sample (Invitrogen, 25530015 10mg/mL) and the samples were
535 incubated at 55 °C for 45 minutes at 900rpm followed by 67 °C for 2 hours at 900rpm. Samples
536 were purified (Zymo #D4014) and eluted in 10 µL water and quantified by dsDNA qubit
537 quantification. Biotinylated DNA was then enriched with 5 µL pre-washed Streptavidin C1 beads
538 per sample (Life Technologies, 65001) in biotin binding buffer (5mM Tris-HCL pH 7.5, 1mM EDTA,
539 1M NaCl) for 15 minutes at room temperature with rotation. Beads were washed twice with Tween
540 Wash Buffer (5mM Tris-HCL pH 7.5, 0.5mM EDTA, 1M NaCl, 0.05% tween 100) at 55 °C for 2
541 minutes at 900rpm and were then washed with TD buffer (10mM Tris-HCL pH 7.5, 5mM MgCl2,
542 10% Dimethylformamide). Beads were resuspended in TD buffer and Tn5 (Illumina, 20034198)
543 was added at the suggested dilutions and tagmentation was allowed to occur at 55 °C for 10
544 minutes with 900 rpm shaking. Supernatant was removed and beads were washed in 50 mM
545 EDTA for 30 minutes at 50 °C and twice more for 3 minutes each. Beads were then washed twice
546 in Tween Wash Buffer at 55 °C for 2 minutes and once in 10mM Tris. DNA was amplified by PCR
547 with Phusion HF 2X MM and Nextera universal barcodes described previously and was purified
548 by two-sided SPRI purification (Omega Bio-Tek, M1378-01). Samples were sequenced using a
549 NovaSeq SP kit (Illumina) with 100bp or 50 bp paired-end reads at a sequencing depth of
550 approximately 100 million reads per sample.
551
552 **RNA-seq**

553 In triplicate separate cultures, cells were treated with the indicated compounds. For the 3-day
554 RNA-seq experiment, wildtype K562 or HCT116 cells were treated with 10 μM GSK3482364 or
555 0.1% DMSO control for 3 days. For the degron experiment, HaloTag-CTCF HCT116 knock-in
556 cells were pretreated with 330 nM HaloPROTAC3 (AOBIOUS AOB36136) or 0.1% DMSO control
557 for 8 hours, and then were treated with 10μM GSK3482364 or 0.1% DMSO in addition to +/-330
558 nM HaloPROTAC3 for 24 more hours. For the speckle knockdown experiment, K562 speckle
559 dTAG knock-in cells were treated with 500 nM dTAG-13 or 0.1% DMSO control for 6 hours. Total
560 RNA was isolated from 1 million cells per condition using the RNeasy Plus Mini Kit (Qiagen,
561 74104). mRNA was isolated using the Poly(A) RNA Selection Kit V1.5 (Lexogen, 157.96) and
562 total RNA seq libraries were prepared using the CORALL Total RNA-Seq Library Prep Kit
563 (Lexogen, 095.24). Samples were sequenced using a NovaSeq SP kit (Illumina) 50 bp paired-
564 end reads at a sequencing depth of approximately 30 million reads per sample. For the HaloTAG
565 CTCF degron experiments, samples were instead sequenced to a depth of approximately 15
566 million reads per sample with 100bp paired end reads.
567
568 **SON Cut&Tag**
569 For the wildtype cell line experiments, K562 or HCT116 wildtype cells were treated in duplicate
570 with either 10 uM GSK3482364 or 0.1% DMSO control for 3 days. For the CTCF degron
571 experiment, HaloTag-CTCF HCT116 knock-in cells were pretreated with 330 nM HaloPROTAC3
572 (AOBIOUS AOB36136) or 0.1% DMSO control for 8 hours, and then were treated with 10 μM
573 GSK3482364 or 0.1% DMSO in addition to +/- 330nM HaloPROTAC3 for 24 more hours. Cut&Tag
574 was performed according to published protocol (Kaya-Okur et al. 2019). 1 million cells were
575 harvested and washed in PBS. Cells were resuspended in ½ volume ice-cold NE1 buffer (20 mM
576 HEPES-KOH pH 7.9, 10 mM KCl, 0.5mM spermidine, 0.1% Triton-X-100, 20% glycerol with
577 Roche Complete Protease Inhibitor EDTA-free) and lysed on ice for 10 minutes. Nuclei were
578 resuspended in PBS and fixed with 0.1% formaldehyde (Sigma Aldrich, F8775) for 2 minutes at

579 room temperature and quenched with 75 mM glycine. Nuclei were centrifuged, drained, and
580 resuspended in wash buffer (20 mM HEPES-NaOH pH 7.5, 150 mM NaCl, 0.5 mM spermidine,
581 Roche Complete Protease Inhibitor EDTA-free) to ~1 million cells per mL, and 250,000 nuclei per
582 condition (as determined by hematocytometer) were used per condition moving forward.
583 Concavalin A beads (Cell Signaling Technologies, 93569S) were then activated with binding
584 buffer (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$) and 10 µL were
585 added per sample and incubated for 10 minutes at room temperature with rotation. Liquid was
586 removed on a magnet and beads were resuspended in 50uL ice-cold antibody buffer (wash buffer
587 with 2 mM EDTA and 0.1% BSA) and 1 µL of antibody (anti-SON Abcam #ab121759) was added
588 and incubated upright with gentle nutating overnight at 4 °C. Beads were resuspended in
589 secondary antibody in wash buffer (1:100 guinea pig anti-rabbit ABIN101961) and tubes were
590 nutated at room temperature for 1 hour. Liquid was removed and beads were washed three times
591 with wash buffer. pA-Tn5 (purified in-house according to published protocol (Priore et al. 2021))
592 was mixed 1:50 in Dig-300 buffer (without digitonin, 20 mM HEPES-NaOH pH 7.5, 300 mM NaCl,
593 0.5mM spermidine, Roche Complete Protease Inhibitor EDTA-free) and beads were resuspended
594 in 100 uL and nutated at room temperature for 1 hour. Beads were washed three times in Dig-
595 300 buffer. Beads were resuspended in 300 uL Tagmentation Buffer (Dig-300 buffer with 10 mM
596 $MgCl_2$) and incubated at 37 °C for one hour. 16 mM EDTA, 0.1% SDS, and 50 µg of Proteinase
597 K (Invitrogen, 25530015) was added per sample and incubated at 55 °C for one hour and then
598 DNA was purified by PCI-chloroform extraction. DNA was amplified by PCR with NEBNext HiFi
599 2x PCR Master Mix (NEB M0541S) and library was purified by 1.3X SPRI purification (Omega
600 Bio-Tek, M1378-01). Samples were sequenced using a NextSeq1000 (Illumina) with 50bp paired
601 end reads at a sequencing depth of approximately 10-20 million reads per sample.
602
603 **CTCF ChIP-seq data processing and peak calling**

604 Reads were aligned to hg38 using Burrow-Wheeler Aligner (BWA-mem, v 0.7.15)(Li 2013). Sam
605 files were converted to bam files and sorted using Samtools view and sort (v 1.5)(Li et al. 2009).
606 Identical ChIP-seq sequence reads were collapsed to avoid PCR duplicates using Picard Tools
607 MarkDuplicates(v 2.18.5) with Java (v 1.8.0) with parameters
608 VALIDATION_STRINGENCY=LENIENTASSUME_SORTED=true
609 REMOVE_DUPLICATES=true. Reads were converted to tdf files for viewing on IGV using
610 Igvtools (v 2.3.95) count with -w 25(Robinson et al. 2011). Tag directories were made from
611 deduplicated bam files with Homer (v. 4.9) makeTagDirectory with Perl (v 5.26.1) (Heinz et al.
612 2010). Peaks were called using Homer (v. 4.11) getDifferentialPeaksReplicates where -t was
613 replicate ChIP tag directories, -i was the corresponding replicate input directories, with parameters
614 -F 10 and -L 4. Reactivated peaks were called using Homer (v. 4.11)
615 getDifferentialPeaksReplicates where -t was replicate ChIP tag directories for DNMT1i treatment,
616 -b was the replicate ChIP directories for DMSO treatment, and -p was the peaks called in the
617 corresponding DNMT1i treatment condition. These peaks were further restricted to those only
618 called in DNMT1i and not DMSO by using bedtools (v 2.26.0) intersect with the -v and -wa options
619 (Quinlan and Hall 2010). For aggregate analyses, DeepTools was used to convert deduplicated
620 bam files to RPKM normalized bigwigs and to compare bigwigs to one another.
621

622 **HiChIP data preprocessing and loop calling**

623 HiChIP reads for each sample were aligned to hg38 and processed with HiC Pro (v 3.0.0) using
624 the default settings with a DpnII digested genome generated with the digest_genome script
625 (Servant et al. 2015). For most analyses, HiCPro output .allvalidpairs files for replicate samples
626 were then pooled. Loops were called using Hichipper (v 0.7.7) in call mode with default
627 parameters and the flag --input-vi, and loops for each condition were kept if they had FDR< 0.01
628 and > 4 PETs(Lareau and Aryee 2018b). For downstream foldchange and CPM analysis using
629 Diffloop, Hichipper was instead used to call loops from HiCPro outputs without pooling replicates

630 and the unfiltered loop list was used as input to Diffloop. For all Hichipper analyses, merged peak
631 files containing all peaks from DMSO and DNMT1i for a given cell line were used as the input
632 peak file.
633
634 **Hi-C analysis**
635 Published LIMe-Hi-C data for three replicates of DMSO and DNMT1 inhibitor treatment conditions
636 in K562 were employed for the analysis (Siegenfeld et al. 2022). The same treatment conditions
637 were used as employed in this study. LIMe-Hi-C samples were generated as previously
638 described(Siegenfeld et al. 2022) and sequenced to a depth of approximately 600 million reads
639 per sample. Hi-C data were processed as previously described (Siegenfeld et al. 2022) using
640 JuiceMe (v 1.0.0) to generate merged_nodups.txt.gz pairwise contact files. Contacts from all three
641 replicates were subsequently pooled together using the JuiceMe CPU "mega" script (v 1.0.0).
642 These files were then converted to the cool format at 10 kb resolution using cooler (v 0.8.11).
643 These cool files were normalized using the cooler (v 0.8.11) "balance" function and were directly
644 utilized for aggregate pileup analysis. To perform pileup analysis, CTCF peaks, identified as
645 described above, were filtered based upon the presence of a CTCF motif using the motif file
646 MA0139.1.tsv.gz downloaded from the JASPAR database and assigned to the motif with the
647 strongest signal using the bioframe (v 0.3.3) "overlap" function. In addition, CTCF sites that were
648 located within 1000 bp of a previously identified blacklisted region (ENCFF356LFX.bed.gz) were
649 removed using the bioframe "subtract" function. Pileup analysis was then performed using the
650 cooltools (v 0.5.1) "pileup" function on peaks deduplicated through the bioframe "cluster" function
651 with a "min_dist" threshold set to a resolution of 10 kb to remove duplicate peaks located within
652 the same genomic bin. Peaks for which motifs mapped to the – strand were reflected across the
653 diagonal and the pileup matrix was then plotted in python (v 3.7) using standard matplotlib (v
654 3.5.2) functions. Highly-looping and reactivated peaks were defined as specified below. Speckle

655 associated peaks were defined to be those with an average speckle SON Cut&Tag signal greater
656 than or equal to the 90th percentile genome-wide 20 kb bin average speckle signal.
657
658 The stripenn (v 1.1.65.7)(Yoon et al. 2022) “compute” function was used to identify stripes from
659 the merged DNMT1 inhibitor 10 kb cool file with the following parameters: --norm KR -m
660 0.95,0.96,0.97,0.98,0.99. The distances of highly-looping peaks and normal-looping peaks to
661 stripe anchors tabulated in the stripenn “result_filtered.tsv” output were calculated using
662 pybedtools (v 0.8.1) “closest” function and visualized in python (v 3.8) using standard plotting
663 functions.
664
665 **CTCF binding scatterplots and aggregate profile plots**
666 To make scatterplots of CTCF normalized tags, RPKM normalized bigwigs were generated for
667 ChIP samples and matched input controls from deduplicated bam files (see above) using
668 deepTools (v 3.4.3) bamCoverage. ChIP files were normalized to their matched input controls
669 using deepTools (v 3.4.3) bigwigCompare with the log2 option. A merged peak list was made by
670 combining peak files from DNMT1i and DMSO and collapsing the list with bedtools (v 2.26.0)
671 merge, and normalized ChIP-seq signal per condition was annotated onto these peaks using
672 deepTools multiBigwigSummary. To make aggregate profile plots of CTCF ChIP-seq signal over
673 genes or peak subsets, bigwig files of CTCF ChIP-seq signal were generated for each condition
674 from deduplicated bam files (see above) with deepTools (v 3.4.3) bamCoverage with flag --
675 normalizeUsing RPKM. To make aggregate plots over peak subsets, deepTools (v 3.4.3)
676 computeMatrix and plotProfile were used with the following computeMatrix parameters:
677 reference-point --referencePoint center -bs 50 -a 1000 -b 1000 --skipZeros. To make aggregate
678 plots over genes, deepTools (v 3.4.3) computeMatrix and plotprofile were used with the following
679 parameters: --scale-regions --regionBodyLength 1000-bs 50 -a 1000 -b 1000 --skipZeros.

680 Downstream processing was completed in python (v 3.7.0) (www.python.org) with pandas (v
681 1.0.3), matplotlib (v 3.2.1) and seaborn (v 0.11.2).
682
683 **Annotating CTCF peaks**
684 CTCF peaks were annotated using Homer (v 4.11) annotatePeaks using the -gtf option with an
685 Ensembl gtf file (v 104) with the standard priority order assigned (Heinz et al. 2010). Motifs were
686 assigned to CTCF peaks using gimmemotifs gimme scan function with -b and -f 0.1 and the
687 JASPAR motif pfm as input. Strand-aware motif methylation heatmaps were made using
688 Deeptools. To assess overlap of categories of CTCF peaks with a bed file of predefined
689 subcompartments from SNIPER (Xiong and Ma 2019) bedtools (v 2.26.0) map was used with the
690 – o distinct flag, and data were further processed in python (v 3.7.0) (www.python.org) and any
691 peaks overlapping multiple subcompartments were discarded. To assess overlap of categories of
692 CTCF peaks with normalized TSA-seq counts, UCSC's function bigWigToBedGraph was first
693 used to convert a bigwig file of normalized SON TSA-seq counts to a bedgraph in 20kb bins, and
694 then bedtools (v 2.26.0) map was used with – o mean flag to map TSA-seq counts onto CTCF
695 peak list and the data were further processed in python (v 3.7.0). The following published files
696 were used from 4D Nucleome for TSA-seq figures: 4DNFIVZSO9RI.bw 4DNFIBY8G6RZ.bw.
697
698 **Assigning number of loops (partners) per CTCF peak**
699 CTCF peaks were annotated with the number of different loops (partners) per peak by using
700 pgltools (v 2.2.0) intersect1D to list each loop-peak interception followed by bedtools (v 2.26.0)
701 merge on the list of peaks to sum the number of different loops in which each peak falls within an
702 anchor (Greenwald et al. 2017; Quinlan and Hall 2010). To annotate peaks with the number of
703 different loops their maximally-looping partner makes, pgltools (v 2.2.0) intersect1D was again
704 used to list each loop-peak interception. Bedtools (v 2.26.0) map was used to annotate the
705 number of different loops the peak in each partner anchor makes from the previous file using the

706 -o max flag. This resulted in one entry per loop that lists the “original” peak and the number of
707 loops its partner anchor makes. The maximally-looping partner per original peak was mapped
708 back onto the original peak list with bedtools (v 2.26.0) map using the -o max flag so that each
709 peak was listed once.
710
711 **Calling and characterizing highly-looping peaks**
712 Highly-looping peaks were called by adapting code used previously to define methylation rich
713 regions (MRRs) and superenhancers (Cai et al. 2020). Specifically, peaks were ordered by rank
714 with respect to the number of different loops they make. The tangent line was drawn where the
715 slope becomes 1 and any peaks above this cutoff were considered highly-looping for a given
716 condition. To generate aggregate profile plots of public K562 histone modification bigwig files (see
717 source data) over CTCF peaks, deepTools (v 3.4.3) computeMatrix was used with the following
718 parameters: reference-point --referencePoint center -bs 50 -a 5000 -b 5000 --skipZeros and
719 results were visualized with deepTools (v 3.4.3) plotProfile. Overlap of peaks with
720 subcompartments and TSA-seq was performed as described above for all CTCF peaks. The
721 following published files were used from the ENCODE consortium for the peak characterization
722 pileups: ENCFF239EBH.bigWig, ENCFF469JMR.bigWig, ENCFF600FDO.bigWig.
723
724 **Calculating looping changes across conditions**
725 To calculate foldchange and mean CPM loop strength between conditions, mango Hichipper
726 output files for individual replicates were input into Diffloop (v 1.16.0) and processed with the
727 loopsmake mango function (Lareau and Aryee 2018a). The full loop list was then filtered for loops
728 with FDR <0.01 and with at least 8 PETs across samples for a given cell type and loops that
729 appeared strongly (>5 PETs) in only one replicate but 0 PETS in the other were removed. The
730 quickAssoc function was then used to get loop strength foldchange and CPM values for each
731 loop. Diffloop output files obtained from the summary() function were then intersected with

genomic features (such as reactivated peak overlap, subcompartment, Cut&Tag signal) for further analyses using pgltools (v 2.2.0) intersect1D and redundant loops were merged using pgltools (v 2.2.0) merge and files were further processed in python (v 3.7.0) (www.python.org) with pandas (v1.0.3), matplotlib (v 3.2.1) seaborn (v 0.11.2) and statannot (v 0.2.3) (Greenwald et al. 2017).

Cut&Tag data processing

Paired end reads were aligned to hg38, deduplicated and converted to bam and tdf files as described above for ChIP-seq. To generate aggregate profile plots of SON Cut&Tag signal over CTCF peaks, normalized bigwig files averaging both replicates were generated from deduplicated bam files with deepTools (v 3.4.3) bamCompare with --operation mean --normalizeUsing RPKM --scaleFactorsMethod None. Then, deepTools (v 3.4.3) computeMatrix and plotProfile were used to plot SON Cut&Tag signal over categories of CTCF peaks with the following computeMatrix parameters: reference-point --referencePoint center -bs 50 -a 1000 -b 1000 –skipZeros. To compare Cut&Tag binned signal to peak and gene sets, replicate-averaged bedgraph files were generated with deepTools (v 3.4.3) bamCompare with --normalizeUsing RPKM --operation mean -of bedgraph --binSize 20000 --scaleFactorsMethod None. Binned Cut&Tag signal was mapped onto genes/peaks with bedtools map (v 2.26.0) with -o mean. To compare loop strength and loop foldchange from Diffloops summary file (described above) to binned Cut&Tag signal, the replicate-averaged Cut&Tag bedgraphs in 20 kb bins were intersected with loop files using pgltools (v 2.2.0) intersect1D with -wa -wb -allA followed by pgltools merge with -o mean (Greenwald et al. 2017). For all plots with deciles, equal deciles for each plot were determined in python (v 3.7.0) with pandas (v 1.0.3) qcut with duplicates ='drop'.(Ramírez et al. 2016) Subsequent processing for all plots was completed in python (v 3.7.0) (www.python.org) with pandas (v1.0.3), matplotlib (v 3.2.1), seaborn (v 0.11.2) and statannot (v 0.2.3).

RNA-seq processing and differential expression analysis

758 RNA-seq data were analyzed following Lexogen's recommended protocol. First, UMIs were
759 extracted with umi-tools (v 1.1.1) extract (Smith et al. 2017). Adapters were then trimmed with
760 cutadapt (v 2.7) (Martin 2011). Trimmed reads were aligned to hg38 using Star aligner (v 2.6.0)
761 with a genome generated using the Ensembl hg38 primary assembly and Ensembl GTF file v 104
762 with the following settings: --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --
763 alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.6 --
764 alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 --
765 peOverlapNbasesMin 40 --peOverlapMMp 0.8 --outSAMattributes NH HI NM MD(Dobin et al.
766 2013). Reads were then deduplicated with umi-tools dedup (v 1.1.1) with flags multimapping-
767 detection-method=NH and unpaired-reads=discard. For gene expression analysis, HTseq – count
768 (v 0.11.2) was used to count reads in genes (Anders et al. 2015). Count files were then processed
769 in DEseq2 (v 1.28) with R (v 4.0.2) using the standard workflow to get foldchange values and
770 using alpha=0.05 (Love et al. 2014; Zitovsky and Love 2019). Hierarchical clustering was
771 performed on the regularized $\log_2$-transformed counts using Euclidean distance as a metric and
772 the complete clustering method. All genes with both an adjusted $P < 0.05$, calculated using the
773 Benjamini-Hochberg correction, and $|\log_2(\text{fold-change})| > 0.5$ between wildtype DNMT1i and
774 DMSO were included in the clustered heatmap. The regularized $\log_2$-transformed counts for the
775 variable genes were grouped by *k*-means clustering ($k = 4$) and the Z-scores for the variable
776 genes, ordered according to their cluster identity, are depicted. For differential exon analysis,
777 Dexseq (v 1.34.1) was used. Files were then imported into R and foldchange values were
778 computed according to the standard Dexseq workflow. BioMart was used to map the genomic
779 ranges for each gene onto RNA-seq differential expression output files. Gene expression changes
780 were compared with Cut&Tag signal and CTCF peak overlap using bedtools (v 2.26.0) map with
781 -o mean. Gene Ontology analysis was completed on the online DAVID portal by filtering for
782 goterm_direct and goterm_BP_5 results with FDRj<0.2 (Sherman et al. 2022; Huang et al. 2009).

783 Further analysis, including volcano plots, was completed in python (v 3.7.0) (www.python.org)
784 with pandas (v 1.0.3), matplotlib (v 3.2.1) seaborn (v 0.11.2) and statannot (v 0.2.3).
785
786 **Data Access**
787
788 All sequencing data has been deposited on NCBI Gene Expression Omnibus under the super
789 series accession code GSE217760. Custom code is available at https://github.com/liaulab.
790
791 **Competing Interests**
792
793 B.B.L. holds a sponsored research project with AstraZeneca is a shareholder and member of the
794 scientific advisory board of Light Horse Therapeutics. The remaining authors declare no
795 competing interests.
796
797 **Acknowledgements**
798
799 We acknowledge members of the Liau lab for useful discussions. We acknowledge J. Nelson, C.
800 Maesner and Z. Niziolek for assistance with cell sorting. We acknowledge the FAS Division of
801 Science Research Computing Group at Harvard University for assistance with cluster computing.
802 We acknowledge S. Johnstone for support with the HiChIP protocol. We acknowledge T. Lu for
803 helpful discussions about speckle Cut&Tag. We acknowledge H.Roh for providing Protein A Tn5.
804 A.P.S and C.L. were supported by the Herchel Smith Graduate fellowships. N.Z.L. was supported
805 by a NSF Graduate Research Fellowship (grant no. DGE1745303). A.L.W was supported by a
806 Simmons award from Harvard Center for Biological Imaging. This research was supported by
807 startup funds from Harvard University.
808

Supplemental Figures

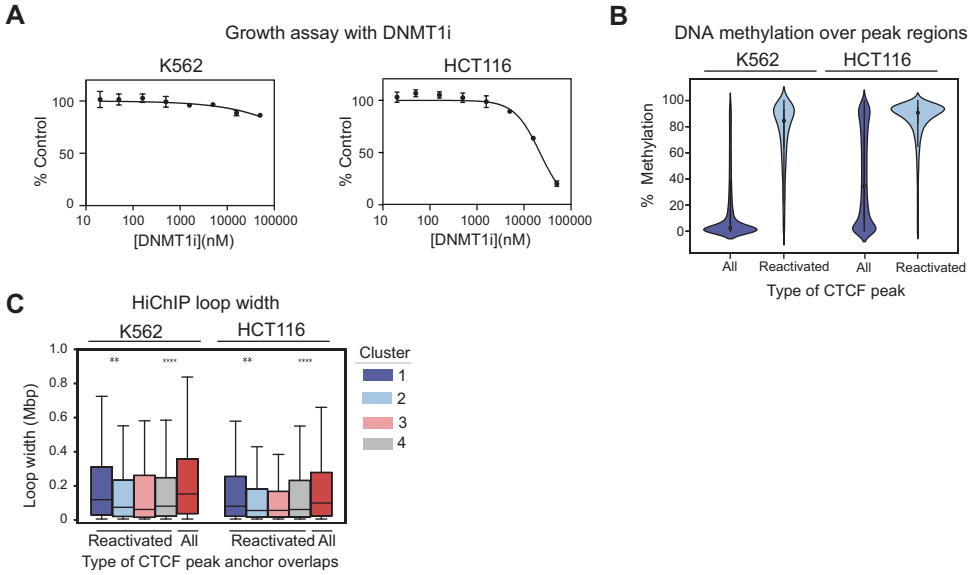


Figure S1 | DNMT1 inhibition reactivates looping CTCF peaks

A. Dose-response curve in K562 (left) and HCT116 cells (right) of percent growth (y-axis) in varying doses of DNMT1i treatment (x-axis) relative to DMSO-treated control cells after 3 days of treatment. Data represent mean ± s.e.m.

B. Percent methylation in wildtype cells on CpG sites within 300 bp CTCF peak regions for reactivated peaks and all other CTCF peaks.

C. Loop anchor-to-anchor width for loops with or without at least one anchor overlapping a reactivated peak. Loops that overlap reactivated peaks are separated into motif methylation clusters from **Fig 1D**.

For (**C**), the interquartile range (IQR) is depicted by the box with the median represented by the center line. Outliers are excluded. P values were calculated by a Mann-Whitney test and are annotated as follows: ns: not significant; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

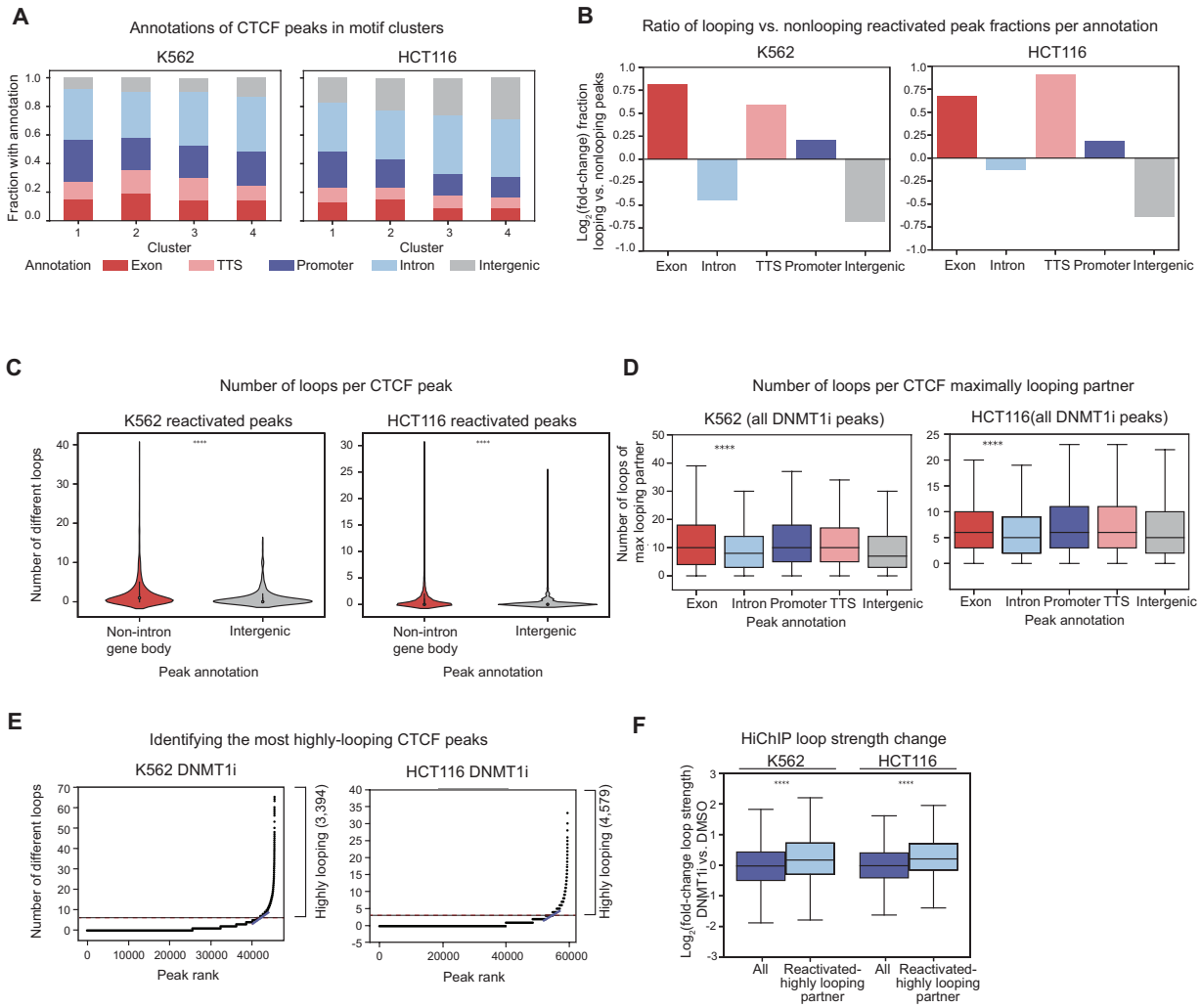


Figure S2 | Reactivated CTCF peaks loop from gene bodies and interact with stripe anchors

A. Genomic annotations of CTCF peaks in each cluster from **Fig 1D**.
B. $\log_2$(fold-change) ratio of the fractions of each annotation within looping vs. nonlooping reactivated peaks. Ratio of bars from **Fig 2A**.
C. Number of different partner peaks each CTCF peak interacts with (y-axis) for reactivated K562 (left) and reactivated HCT116 (right) peaks in DNMT1i by peak annotation (x-axis).
D. Number of different partner peaks the maximally-looping partner interacts with (y-axis) for all K562 (left) and all HCT116 (right) peaks in DNMT1i by peak annotation (x-axis).
E. Number of different partner peaks (*y*-axis) assigned to each peak by rank (*x*-axis) in DNMT1i for K562 (left) and HCT116 (right). Highly-looping peaks reside above the red dashed line.
F. Boxplot showing the $\log_2$(fold-change) in loop strength DNMT1i vs. DMSO (y-axis) for loops connecting reactivated peaks to highly-looping partners vs all other loops (x-axis) for K562 and HCT116.

For boxplots, the interquartile range (IQR) is depicted by the box with the median represented by the center line. Outliers are excluded. P values were calculated by a Mann-Whitney test and are annotated as follows: ns: not significant; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

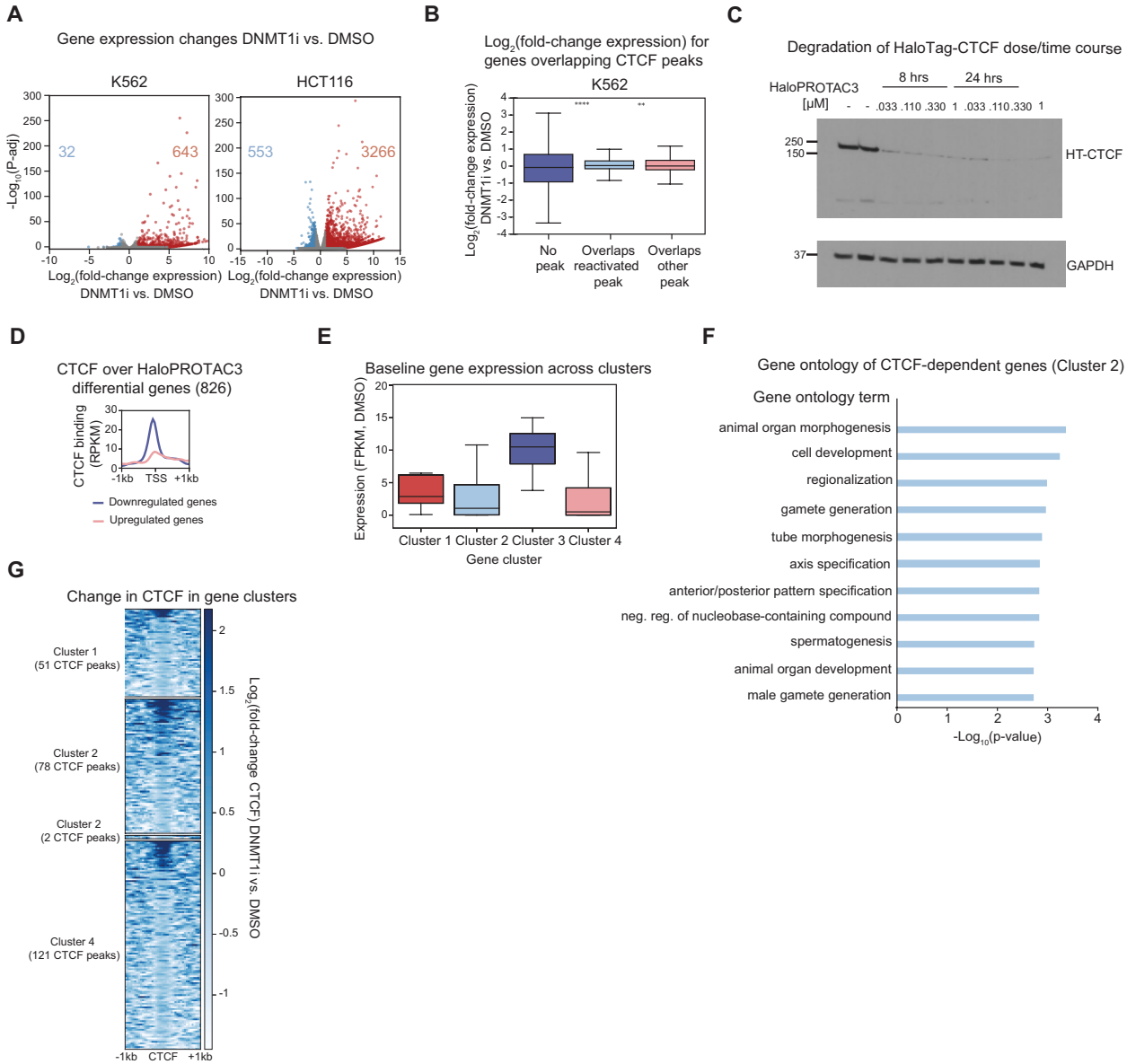


Figure S3 | Many genes depend on CTCF for upregulation upon DNMT1i

A. Volcano plot showing -$\log_{10}$(adjusted P-value) (y-axis) vs. $\log_2$(fold-change) in gene expression (x-axis) upon 3 days DNMT1i treatment for K562 (left) and HCT116 (right). Upregulated genes are highlighted in red, and downregulated genes are highlighted in blue (p-adj < 0.05, $|\log_2 \text{Foldchange}| > 1$).

B. Boxplot showing $\log_2$(fold-change) in gene expression in DNMT1i vs. DMSO (y-axis) for genes overlapping no CTCF peak, reactivated CTCF peaks, or non-reactivated CTCF DNMT1i peaks (x-axis) for K562 cells

C. Western blot for CTCF levels over a dose/time course of HaloPROTAC3 in HaloTag-CTCF knock-in HCT116 cells. GAPDH is shown as a loading control. 330nM was chosen as the final dose.

D. Aggregate profile plot of replicate-averaged RPKM normalized CTCF binding in DMSO over the transcription start sites of differential HaloPROTAC3 genes.

859 **E.** Boxplot showing expression levels of genes within clusters from **Fig 3C** in Fragments Per
860 Kilobase per Million mapped fragments (FPKM). Expression in cells treated with vehicle (-
861 HaloPROTAC3, - DNMT1i) is shown.
862 **F.** Bar plot showing the $-\log_{10}$(p-value) for the top enriched gene ontology terms determined
863 through DAVID for Cluster 2 genes from **Fig 3C**.
864 **G.** Heatmap of per-CTCF site $\log_2$(fold-change CTCF binding) after 1 day of DNMT1i treatment
865 vs. DMSO control centered over all CTCF peaks overlapping genes in the clusters from **Fig**
866 **3C**. Each gene could have multiple CTCF peaks on it.
867

For boxplots, the interquartile range (IQR) is depicted by the box with the median represented by the center line. Outliers are excluded. P values were calculated by a Mann-Whitney test and are annotated as follows: ns: not significant; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

868

869

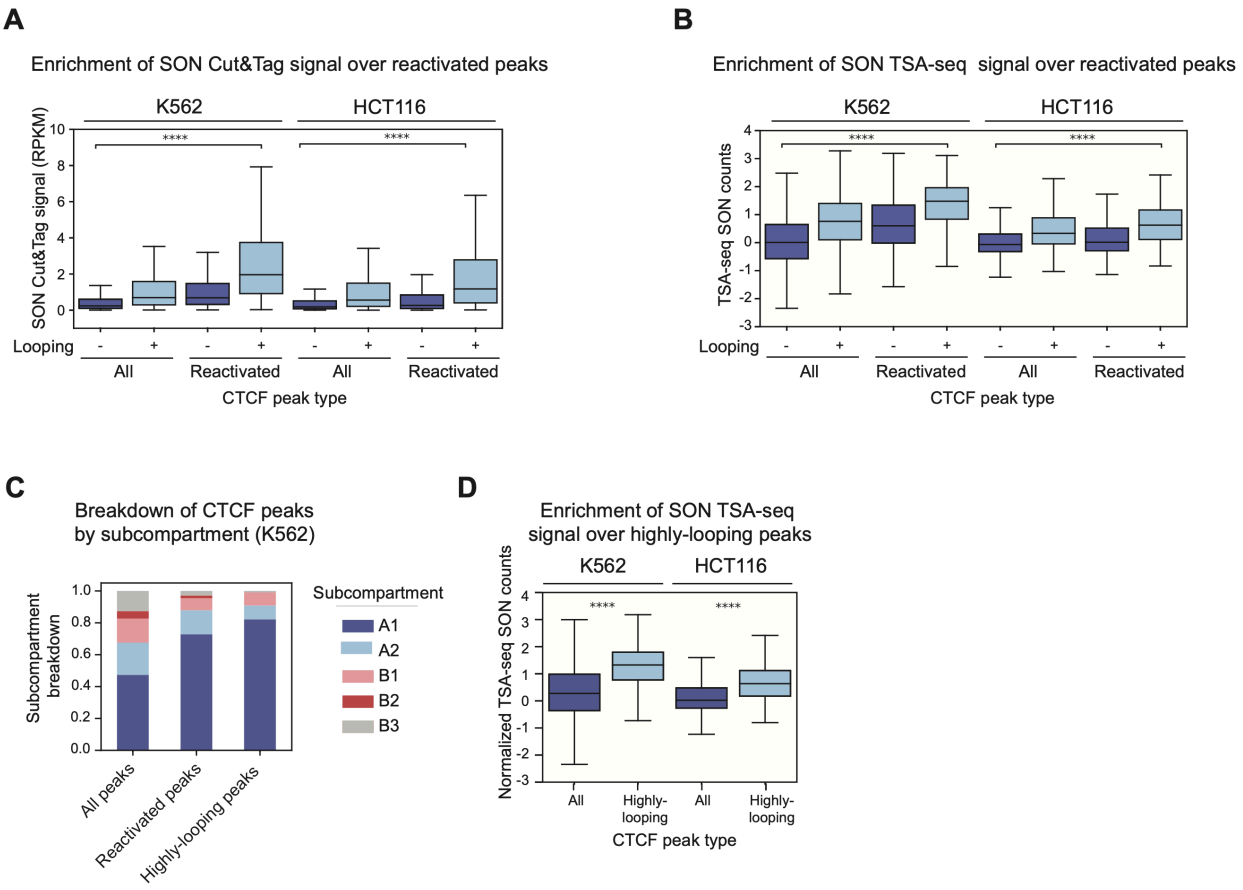


Figure S4 | Reactivated CTCF peaks loop gene bodies to stripe anchors

A. Boxplot showing replicate-averaged DNMT1i SON Cut&Tag signal (RPKM, 20 kb bins, y-axis) in the respective cell type over reactivated peaks vs. all other CTCF peaks called in DNMT1i for K562 and HCT116 cells broken down by whether the CTCF peak is in a loop anchor.
B. Boxplots showing SON TSA-seq normalized counts (y-axis) for reactivated vs. non-reactivated CTCF DNMT1i peaks for K562 and HCT116 (left) broken down by whether the CTCF peak is in a loop anchor.
C. Proportion of all, reactivated, and highly-looping CTCF K562 DNMT1i peaks in each published subcompartment (y-axis) assigned by SNIPER for wildtype K562 cells.
D. Boxplots showing SON TSA-seq normalized counts (*y*-axis) for highly-looping vs. normal-looping CTCF DNMT1i peaks for K562 and HCT116 (right). SON TSA-seq normalized counts were analyzed from public data in the untreated, wildtype respective cell type (see methods).

For boxplots, the interquartile range (IQR) is depicted by the box with the median represented by the center line. Outliers are excluded. P values were calculated by a Mann-Whitney test and are annotated as follows: ns: not significant; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

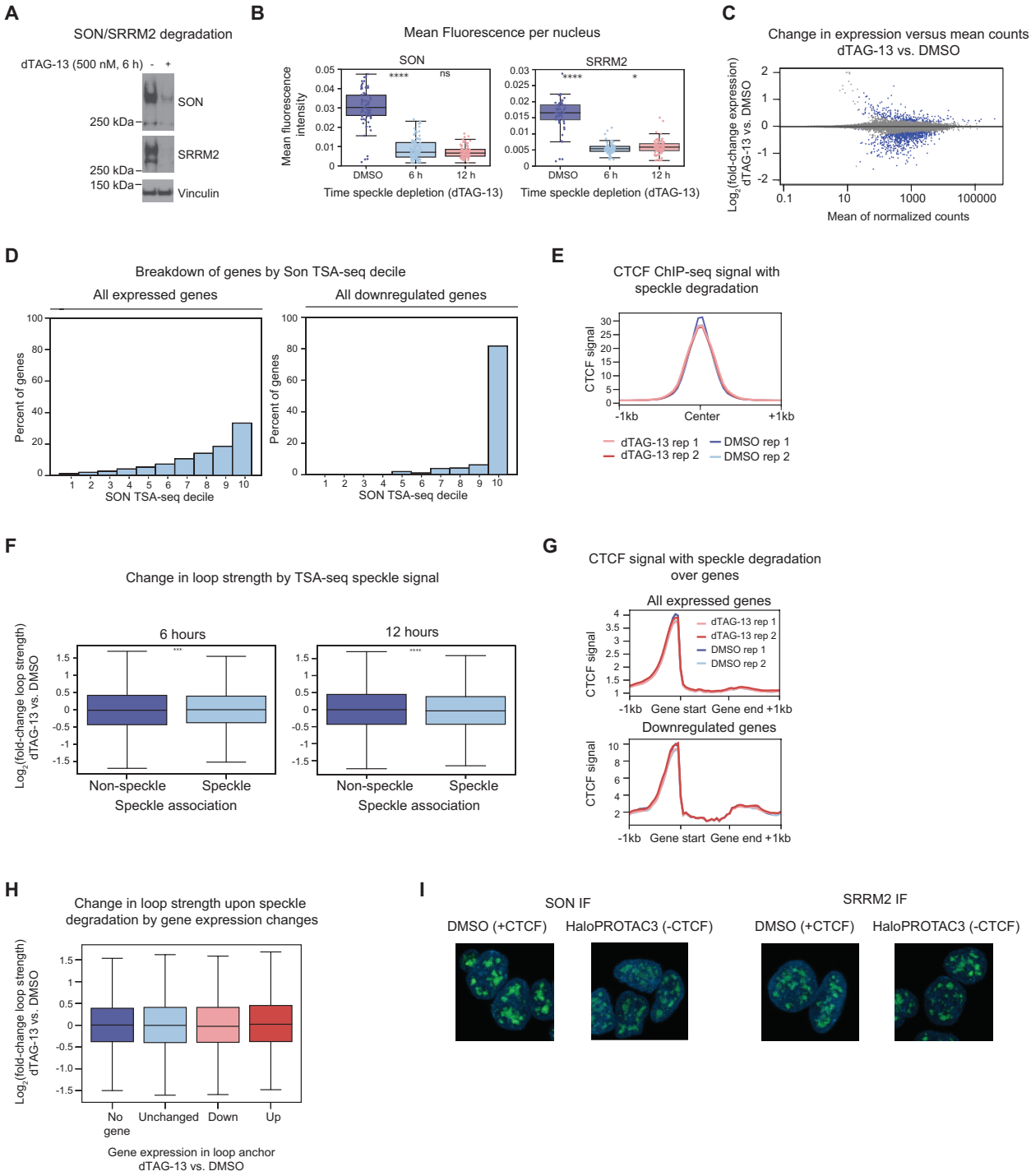


Figure S5 | Acute depletion of nuclear speckles does not disrupt CTCF binding or looping

A. Separate western blots for SON and SRRM2 in speckle dTAG knockin and wildtype (wt) K562 cells ± 6 hours of 500 nM dTAG-13 treatment. Vinculin was used as a loading control.

B. Per-nucleus mean fluorescence following immunofluorescence for SON/SRRM2 double dTAG knock-in K562 cells treated with dTAG-13 for 6 or 12 hours or DMSO for 12 hours.

899 **C.** $\log_2$(fold-change in gene expression) (y-axis) vs. baseMean gene expression for both
900 conditions (x-axis) for 6 hr dTAG-13 vs. DMSO treatment for all genes. APEGLM shrinkage
901 was applied for visualization.
902 **D.** Proportion of all nonzero (left) and all 6 hr dTAG-13 downregulated (right, padj<0.05) genes
903 (y-axis) in each SON TSA-seq decile (x-axis) from wildtype cells. 10 is the decile closest to
904 speckles.
905 **E.** Aggregate profile plot of CTCF ChIP-seq signal (RPKM) in speckle knock-in K562 cells with
906 and without 6 hr speckle degradation (y-axis) over all CTCF peaks (x-axis).
907 **F.** Boxplot showing $\log_2$(fold-change) loop strength for dTAG-13 treated vs. DMSO treated
908 speckle dTAG knock-in cells relative to TSA-seq decile after 6 or 12 hours of dTAG-13
909 treatment with matched DMSO controls. Speckle refers to the decile closest to speckles, and
910 non-speckle refers to the deciles 1-9 that are furthest from speckles.
911 **G.** Aggregate profile plot of CTCF ChIP-seq signal (RPKM) in speckle dTAG knock-in K562 cells
912 with dTAG-13 vs. DMSO treatment (y-axis) across expressed genes (top) and genes that are
913 downregulated with dTAG-13 treatment (bottom) (x-axis).
914 **H.** Boxplot showing $\log_2$(fold-change) loop strength for 6 hr dTAG-13 treated vs. DMSO treated
915 speckle dTAG knock-in cells (y-axis) for loops with anchors that do not overlap any genes,
916 overlap only genes that do not change expression, and overlap genes that increase or
917 decrease in expression in at least one anchor (x-axis).
918 **I.** Representative SON and SRRM2 immunofluorescence images (green, alexafluor488) in
919 HCT116 cells with 24 hours of 330 nM HaloPROTAC3 treatment to deplete CTCF or vehicle
920 control. Overlaid on DAPI in blue.
921

For boxplots, the interquartile range (IQR) is depicted by the box with the median represented by the center line. Outliers are excluded. P values were calculated by a Mann-Whitney test and are annotated as follows: ns: not significant; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

922
923

924 **References:**

925 Ahn E-Y, DeKelver RC, Lo M-C, Nguyen TA, Matsuura S, Boyapati A, Pandit S, Fu X-D, Zhang
926 D-E. 2011. SON Controls Cell-Cycle Progression by Coordinated Regulation of RNA
927 Splicing. *Mol Cell* 42: 185–198.

928 Alexander KA, Coté A, Nguyen SC, Zhang L, Gholamalamdari O, Agudelo-Garcia P, Lin-Shiao
929 E, Tanim KMA, Lim J, Biddle N, et al. 2021. p53 mediates target gene association with
930 nuclear speckles for amplified RNA expression. *Mol Cell* 81: 1666-1681.e6.

931 Alharbi AB, Schmitz U, Bailey CG, Rasko JEJ. 2021. CTCF as a regulator of alternative splicing:
932 new tricks for an old player. *Nucleic Acids Res* 49: gkab520-.

933 Anders S, Pyl PT, Huber W. 2015. HTSeq—a Python framework to work with high-throughput
934 sequencing data. *Bioinformatics* 31: 166–169.

935 Barrington C, Georgopoulou D, Pezic D, Varsally W, Herrero J, Hadjur S. 2019. Enhancer
936 accessibility and CTCF occupancy underlie asymmetric TAD architecture and cell type
937 specific genome topology. *Nat Commun* 10: 2908.

938 Bell AC, Felsenfeld G. 2000. Methylation of a CTCF-dependent boundary controls imprinted
939 expression of the Igf2 gene. *Nature* 405: 482–485.

940 Brown JM, Green J, Neves RP das, Wallace HAC, Smith AJH, Hughes J, Gray N, Taylor S,
941 Wood WG, Higgs DR, et al. 2008. Association between active genes occurs at nuclear
942 speckles and is modulated by chromatin environment. *J Cell Biology* 182: 1083–1097.

943 Cai Y, Zhang Y, Loh YP, Tng JQ, Lim MC, Cao Z, Raju A, Aiden EL, Li S, Manikandan L, et al.
944 2021. H3K27me3-rich genomic regions can function as silencers to repress gene expression
945 via chromatin interactions. *Nat Commun* 12: 719.

946 Cai Y, Zhang Y, Loh YP, Tng JQ, Lim MC, Cao Z, Raju A, Li S, Manikandan L, Tergaonkar V, et
947 al. 2020. H3K27me3-rich genomic regions can function as silencers to repress gene
948 expression via chromatin interactions. *Biorxiv* 684712.

949 Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH,
950 Lindquist RA, Moffat J, et al. 2006. CellProfiler: image analysis software for identifying and
951 quantifying cell phenotypes. *Genome Biol* 7: R100.

952 Chen Y, Belmont AS. 2019. Genome organization around nuclear speckles. *Curr Opin Genet*
953 *Dev* 55: 91–99.

954 Chen Y, Zhang Y, Wang Y, Zhang L, Brinkman EK, Adam SA, Goldman R, Steensel B van, Ma
955 J, Belmont AS. 2018. Mapping 3D genome organization relative to nuclear compartments
956 using TSA-Seq as a cytological ruler. *J Cell Biol* 217: jcb.201807108.

957 Davidson IF, Bauer B, Goetz D, Tang W, Wutz G, Peters J-M. 2019. DNA loop extrusion by
958 human cohesin. *Science* eaaz3418.

959 Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological
960 domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:
961 376–380.

962 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
963 TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21.

964 Dubois F, Sidiropoulos N, Weischenfeldt J, Beroukhim R. 2022. Structural variations in cancer
965 and the 3D genome. *Nat Rev Cancer* 1–14.

966 Faber GP, Nadav-Eliyahu S, Shav-Tal Y. 2022. Nuclear speckles – a driving force in gene
967 expression. *J Cell Sci* 135.

968 Flavahan WA, Drier Y, Johnstone SE, Hemming ML, Tarjan DR, Hegazi E, Shareef SJ, Javed
969 NM, Raut CP, Eschle BK, et al. 2019. Altered chromosomal topology drives oncogenic
970 programs in SDH-deficient GISTs. *Nature* 575: 229–233.

971 Flavahan WA, Drier Y, Liau BB, Gillespie SM, Venteicher AS, Stemmer-Rachamimov AO, Suvà
972 ML, Bernstein BE. 2016. Insulator dysfunction and oncogene activation in IDH mutant
973 gliomas. *Nature* 529: 110.

974 Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. 2016. Formation of
975 Chromosomal Domains by Loop Extrusion. *Cell Reports* 15: 2038–2049.

976 Greenwald WW, Li H, Smith EN, Benaglio P, Nariai N, Frazer KA. 2017. Pgltools: a genomic
977 arithmetic tool suite for manipulation of Hi-C peak and other chromatin interaction data. *Bmc*
978 *Bioinformatics* 18: 207.

979 Hansen AS, Cattoglio C, Darzacq X, Tjian R. 2017. Recent evidence that TADs and chromatin
980 loops are dynamic structures. *Nucleus* 9: 20–32.

981 Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. 2000. CTCF
982 mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 405:
983 486–489.

984 Hashimoto H, Wang D, Horton JR, Zhang X, Corces VG, Cheng X. 2017. Structural Basis for
985 the Versatile and Methylation-Dependent Binding of CTCF to DNA. *Mol Cell* 66: 711-720.e3.

986 Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass
987 CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-
988 Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* 38: 576–589.

989 Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, Young RA. 2013.
990 Super-Enhancers in the Control of Cell Identity and Disease. *Cell* 155: 934–947.

991 Hu S, Lv P, Yan Z, Wen B. 2019. Disruption of nuclear speckles reduces chromatin interactions
992 in active compartments. *Epigenet Chromatin* 12: 43.

993 Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene
994 lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44–57.

995 Ilik İA, Malszycki M, Lübke AK, Schade C, Meierhofer D, Aktaş T. 2020. SON and SRRM2 are
996 essential for nuclear speckle formation. *Elife* 9: e60579.

997 Ilık İA, Aktaş T. 2021. Nuclear speckles: dynamic hubs of gene expression regulation. *Febs J*.

998 Jabbari K, Chakraborty M, Wiehe T. 2019. DNA sequence-dependent chromatin architecture
999 and nuclear hubs formation. *Sci Rep-uk* 9: 14646.

1000 Jiang Y, Huang J, Lun K, Li B, Zheng H, Li Y, Zhou R, Duan W, Wang C, Feng Y, et al. 2020.
1001 Genome-wide analyses of chromatin interactions after the loss of Pol I, Pol II, and Pol III.
1002 *Genome Biol* 21: 158.

1003 Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, Ahmad K, Henikoff
1004 S. 2019. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat*
1005 *Commun* 10: 1930.

1006 Kim J, Venkata NC, Gonzalez GAH, Khanna N, Belmont AS. 2020. Gene expression
1007 amplification by nuclear speckle association. *J Cell Biology* 219: jcb.201904046.

1008 Kim Y, Shi Z, Zhang H, Finkelstein IJ, Yu H. 2019. Human cohesin compacts DNA by loop
1009 extrusion. *Science* 366: 1345–1349.

1010 Kraft K, Magg A, Heinrich V, Riemenschneider C, Schöpflin R, Markowski J, Ibrahim DM,
1011 Acuna-Hidalgo R, Despang A, Andrey G, et al. 2019. Serial genomic inversions induce
1012 tissue-specific architectural stripes, gene misexpression and congenital malformations. *Nat*
1013 *Cell Biol* 21: 305–310.

1014 Kubo N, Ishii H, Xiong X, Bianco S, Meitinger F, Hu R, Hocker JD, Conte M, Gorkin D, Yu M, et
1015 al. 2021. Promoter-proximal CTCF binding promotes distal enhancer-dependent gene
1016 activation. *Nat Struct Mol Biol* 1–10.

1017 Lareau CA, Aryee MJ. 2018a. diffloop: a computational framework for identifying and analyzing
1018 differential DNA loops from sequencing data. *Bioinformatics* 34: 672–674.

1019 Lareau CA, Aryee MJ. 2018b. hichipper: a preprocessing pipeline for calling DNA loops from
1020 HiChIP data. *Nat Methods* 15: 155–156.

1021 Lee R, Kang M-K, Kim Y-J, Yang B, Shim H, Kim S, Kim K, Yang CM, Min B, Jung W-J, et al.
1022 2021. CTCF-mediated chromatin looping provides a topological framework for the formation
1023 of phase-separated transcriptional condensates. *Nucleic Acids Res* 50: 207–226.

1024 Lehman BJ, Lopez-Diaz FJ, Santisakultarm TP, Fang L, Shokhirev MN, Diffenderfer KE, Manor
1025 U, Emerson BM. 2021. Dynamic regulation of CTCF stability and sub-nuclear localization in
1026 response to stress. *Plos Genet* 17: e1009277.

1027 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
1028 *Arxiv*.

1029 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R,
1030 Subgroup 1000 Genome Project Data Processing. 2009. The Sequence Alignment/Map
1031 format and SAMtools. *Bioinformatics* 25: 2078–2079.

1032 Li Y, Haarhuis JHI, Cacciatore ÁS, Oldenkamp R, Ruiten MS van, Willems L, Teunissen H, Muir
1033 KW, Wit E de, Rowland BD, et al. 2020. The structural basis for cohesin–CTCF-anchored
1034 loops. *Nature* 1–9.

1035 Liao SE, Regev O. 2020. Splicing at the phase-separated nuclear speckle interface: a model.
1036 *Nucleic Acids Res* 49: gkaa1209-.

1037 Lochs SJA, Kefalopoulou S, Kind J. 2019. Lamina Associated Domains and Gene Regulation in
1038 Development and Cancer. *Cells* 8: 271.

1039 Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for
1040 RNA-seq data with DESeq2. *Genome Biol* 15: 550.

1041 Luan J, Xiang G, Gómez-García PA, Tome JM, Zhang Z, Vermunt MW, Zhang H, Huang A,
1042 Keller CA, Giardine BM, et al. 2021. Distinct properties and functions of CTCF revealed by a
1043 rapidly inducible degron system. *Cell Reports* 34: 108783.

1044 Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz
1045 JM, Laxova R, et al. 2015. Disruptions of Topological Chromatin Domains Cause Pathogenic
1046 Rewiring of Gene-Enhancer Interactions. *Cell* 161: 1012–1025.

1047 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads |
1048 Martin | EMBnet.journal. *EMBnet.journal*.

1049 Maurano MT, Wang H, John S, Shafer A, Canfield T, Lee K, Stamatoyannopoulos JA. 2015.
1050 Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell Reports* 12:
1051 1184–1195.

1052 Misteli T. 2020. The Self-Organizing Genome: Principles of Genome Architecture and Function.
1053 *Cell* 183: 28–45.

1054 Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, Chang HY. 2016.
1055 HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods*
1056 13: 919–922.

1057 Nabet B, Roberts JM, Buckley DL, Paulk J, Dastjerdi S, Yang A, Leggett AL, Erb MA, Lawlor
1058 MA, Souza A, et al. 2018. The dTAG system for immediate and target-specific protein
1059 degradation. *Nat Chem Biol* 14: 431–441.

1060 Nanavaty V, Abrash EW, Hong C, Park S, Fink EE, Li Z, Sweet TJ, Bhasin JM, Singuri S, Lee
1061 BH, et al. 2020. DNA Methylation Regulates Alternative Polyadenylation via CTCF and the
1062 Cohesin Complex. *Mol Cell*.

1063 Nora EP, Goloborodko A, Valton A-L, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny
1064 LA, Bruneau BG. 2017. Targeted Degradation of CTCF Decouples Local Insulation of
1065 Chromosome Domains from Genomic Compartmentalization. *Cell* 169: 930-944.e22.

1066 Pappalardi MB, Keenan K, Cockerill M, Kellner WA, Stowell A, Sherk C, Wong K, Pathuri S,
1067 Briand J, Steidel M, et al. 2021. Discovery of a first-in-class reversible DNMT1-selective
1068 inhibitor with improved tolerability and efficacy in acute myeloid leukemia. *Nat Cancer* 1–16.

1069 Portilho NA, Saini D, Hossain I, Sirois J, Moraes C, Pastor WA. 2021. The DNMT1 inhibitor
1070 GSK-3484862 mediates global demethylation in murine embryonic stem cells. *Epigenet*
1071 *Chromatin* 14: 56.

1072 Priore ID, Ma S, Strecker J, Jacks T, LaFave LM, Buenrostro JD. 2021. Protocol for single-cell
1073 ATAC sequencing using combinatorial indexing in mouse lung adenocarcinoma. *Star Protoc*
1074 2: 100583.

1075 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
1076 features. *Bioinformatics* 26: 841–842.

1077 Quinodoz SA, Ollikainen N, Tabak B, Palla A, Schmidt JM, Detmar E, Lai MM, Shishkin AA,
1078 Bhat P, Takei Y, et al. 2018. Higher-Order Inter-chromosomal Hubs Shape 3D Genome
1079 Organization in the Nucleus. *Cell* 174: 744-757.e24.

1080 Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke
1081 T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis.
1082 *Nucleic Acids Res* 44: W160–W165.

1083 Rao SSP, Huang S-C, Hilaire BGS, Engreitz JM, Perez EM, Kieffer-Kwon K-R, Sanborn AL,
1084 Johnstone SE, Bascom GD, Bochkov ID, et al. 2017. Cohesin Loss Eliminates All Loop
1085 Domains. *Cell* 171: 305-320.e24.

1086 Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL,
1087 Machol I, Omer AD, Lander ES, et al. 2014. A 3D Map of the Human Genome at Kilobase
1088 Resolution Reveals Principles of Chromatin Looping. *Cell* 159: 1665–1680.

1089 Rinzema NJ, Sofiadis K, Tjalsma SJD, Verstegen MJAM, Oz Y, Valdes-Quezada C, Felder A-K,
1090 Filipovska T, Elst S van der, Ramos Z de A dos, et al. 2022. Building regulatory landscapes
1091 reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites
1092 and activate distant genes. *Nat Struct Mol Biol* 1–12.

1093 Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011.
1094 Integrative genomics viewer. *Nat Biotechnol* 29: 24–26.

1095 Rowley MJ, Corces VG. 2018. Organizational principles of 3D genome architecture. *Nat Rev*
1096 *Genet* 19: 1.

1097 Ruiz-Velasco M, Kumar M, Lai MC, Bhat P, Solis-Pinson AB, Reyes A, Kleinsorg S, Noh K-M,
1098 Gibson TJ, Zaugg JB. 2017. CTCF-Mediated Chromatin Loops between Promoter and Gene
1099 Body Regulate Alternative Splicing across Individuals. *Cell Syst* 5: 628-637.e6.

1100 Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, Bochkov ID,
1101 Chinnappan D, Cutkosky A, Li J, et al. 2015. Chromatin extrusion explains key features of
1102 loop and domain formation in wild-type and engineered genomes. *Proc National Acad Sci*
1103 112: E6456–E6465.

1104 Schoenfelder S, Fraser P. 2019. Long-range enhancer–promoter contacts in gene expression
1105 control. *Nat Rev Genet* 20: 437–455.

1106 Schuijers J, Manteiga JC, Weintraub AS, Day DS, Zamudio AV, Hnisz D, Lee TI, Young RA.
1107 2018. Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking
1108 Mechanism. *Cell Reports* 23: 349–360.

1109 Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, Barillot E.
1110 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* 16:
1111 259.

1112 Sherman BT, Hao M, Qiu J, Jiao X, Baseler MW, Lane HC, Imamichi T, Chang W. 2022.
1113 DAVID: a web server for functional enrichment analysis and functional annotation of gene
1114 lists (2021 update). *Nucleic Acids Res* 50: W216–W221.

1115 Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P,
1116 Sandberg R, Oberdoerffer S. 2011. CTCF-promoted RNA polymerase II pausing links DNA
1117 methylation to splicing. *Nature* 479: 74–79.

1118 Siegenfeld AP, Roseman SA, Roh H, Lue NZ, Wagen CC, Zhou E, Johnstone SE, Aryee MJ,
1119 Liau BB. 2022. Polycomb-lamina antagonism partitions heterochromatin at the nuclear
1120 periphery. *Nat Commun* 13: 4199.

1121 Smith T, Heger A, Sudbery I. 2017. UMI-tools: modeling sequencing errors in Unique Molecular
1122 Identifiers to improve quantification accuracy. *Genome Res* 27: 491–499.

1123 Soto EJL, Lipscombe D. 2020. Cell-specific exon methylation and CTCF binding in neurons
1124 regulates calcium ion channel splicing and function. *Elife* 9: e54879.

1125 Spector DL, Lamond AI. 2011. Nuclear Speckles. *Csh Perspect Biol* 3: a000646.

1126 Spielmann M, Lupiáñez DG, Mundlos S. 2018. Structural variation in the 3D genome. *Nat Rev*
1127 *Genet* 19: 453.

1128 Spracklin G, Abdennur N, Imakaev M, Chowdhury N, Pradhan S, Mirny LA, Dekker J. 2022.
1129 Diverse silent chromatin states modulate genome compartmentalization and loop extrusion
1130 barriers. *Nat Struct Mol Biol* 1–14.

1131 Steensel B van, Belmont AS. 2017. Lamina-Associated Domains: Links with Chromosome
1132 Architecture, Heterochromatin, and Gene Repression. *Cell* 169: 780–791.

1133 Steensel B van, Furlong EEM. 2019. The role of transcription in shaping the spatial organization
1134 of the genome. *Nat Rev Mol Cell Bio* 20: 327–337.

1135 Tang Z, Luo OJ, Li X, Zheng M, Zhu JJ, Szalaj P, Trzaskoma P, Magalska A, Wlodarczyk J,
1136 Ruszczycki B, et al. 2015. CTCF-Mediated Human 3D Genome Architecture Reveals
1137 Chromatin Topology for Transcription. *Cell* 163: 1611–1627.

1138 Ulianov SV, Velichko AK, Magnitov MD, Luzhin AV, Golov AK, Ovsyannikova N, Kireev II,
1139 Gavrikov AS, Mishin AS, Garaev AK, et al. 2021. Suppression of liquid–liquid phase
1140 separation by 1,6-hexanediol partially compromises the 3D genome organization in living
1141 cells. *Nucleic Acids Res* 49: gkab249-.

1142 Vermunt MW, Zhang D, Blobel GA. 2019. The interdependence of gene-regulatory elements
1143 and the 3D genome. *J Cell Biol* 218: 12–26.

1144 Vian L, Pękowska A, Rao SSP, Kieffer-Kwon K-R, Jung S, Baranello L, Huang S-C, Khattabi
1145 LE, Dose M, Pruett N, et al. 2018. The Energetics and Physiological Impact of Cohesin
1146 Extrusion. *Cell* 173: 1165-1178.e20.

1147 Wang H, Maurano MT, Qu H, Varley KE, Gertz J, Pauli F, Lee K, Canfield T, Weaver M,
1148 Sandstrom R, et al. 2012. Widespread plasticity in CTCF occupancy linked to DNA
1149 methylation. *Genome Res* 22: 1680–1688.

1150 Wang Y, Zhang Y, Zhang R, Schaik T van, Zhang L, Sasaki T, Peric-Hupkes D, Chen Y, Gilbert
1151 DM, Steensel B van, et al. 2021. SPIN reveals genome-wide landscape of nuclear
1152 compartmentalization. *Genome Biol* 22: 36.

1153 Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara
1154 K, Mishiro T, et al. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding
1155 factor. *Nature* 451: 796–801.

1156 Xiang J-F, Corces VG. 2021. Regulation of 3D chromatin organization by CTCF. *Curr Opin*
1157 *Genet Dev* 67: 33–40.

1158 Xiong K, Ma J. 2019. Revealing Hi-C subcompartments by imputing inter-chromosomal
1159 chromatin interactions. *Nat Commun* 10: 5069.

1160 Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave
1161 K, Zhong F, et al. 2017. Impact of cytosine methylation on DNA binding specificities of
1162 human transcription factors. *Science* 356: eaaj2239.

1163 Yoon S, Chandra A, Vahedi G. 2022. Stripenn detects architectural stripes from chromatin
1164 conformation data using computer vision. *Nat Commun* 13: 1602.

1165 Yu R, Roseman S, Siegnefeld A, Tran K, Nguyen SC, Joyce EF, Raj A, Liau B, Krantz I,
1166 Alexander K, et al. 2023. CTCF/cohesin institute the ground state of chromatin-nuclear
1167 speckle association. *BioRxiv*.

1168 Zhang L, Zhang Y, Chen Y, Gholamalamdari O, Wang Y, Ma J, Belmont AS. 2020. TSA-seq
1169 reveals a largely conserved genome organization relative to nuclear speckles with small
1170 position changes tightly correlated with gene expression changes. *Genome Res* 31:
1171 gr.266239.120.

1172 Zheng H, Xie W. 2019. The role of 3D genome organization in development and cell
1173 differentiation. *Nat Rev Mol Cell Bio* 20: 535–550.

1174 Zitovsky JP, Love MI. 2019. Fast effect size shrinkage software for beta-binomial models of
1175 allelic imbalance. *F1000research* 8: 2024.

1176