DNA methylation insulates genic regions from CTCF loops near nuclear speckles

Authors: Shelby A. Roseman,[1,2,3,6] Allison P. Siegenfeld,[1,2,4,6] Ceejay Lee,[1,2] Nicholas Z. Lue,[1,2,5] Amanda L. Waterbury,[1,2] Brian B. Liau[1,2,*]

Affiliations:

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

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

[3]Dana Farber Cancer Institute, Boston, MA 02215, USA

[4]Harvard Medical School, Boston, MA 02215, USA

[5]University of California Berkeley, Berkeley, CA 94720, USA

[6]Co-first authors

*Correspondence: liau@chemistry.harvard.edu

Abstract

The insulator protein CTCF is essential for mediating chromatin loops and regulating gene expression. While it is established that DNA methylation hinders CTCF binding, the impacts of this methylation-sensitive CTCF binding on chromatin architecture and transcription are poorly defined. Here, we used a selective DNMT1 inhibitor to investigate the characteristics and functions of 'reactivated' CTCF peaks resulting from global DNA demethylation. We found that reactivated peaks preferentially form chromatin loops on gene bodies and interact with highly-looping partner peaks located in regions of active chromatin. Notably, both reactivated CTCF peaks and their highly-looping partners are enriched near nuclear speckles, condensate bodies implicated in transcription and splicing. Utilizing targeted protein degradation, we specifically depleted CTCF and nuclear speckles to elucidate their functional interplay. By degrading CTCF upon DNMT1 inhibition, we revealed that CTCF is important for DNMT1 inhibitor-dependent interactions between chromatin and speckle proteins. Moreover, we found that CTCF promotes the activation of genes near speckles upon DNMT1 inhibition. Conversely, acute depletion of nuclear speckles revealed that they influence transcription but do not maintain CTCF binding or looping. Collectively, our study suggests a model wherein DNA methylation prevents spurious CTCF occupancy and interactions with regulatory elements near nuclear speckles, yet CTCF looping is robust towards the loss of speckles.

Introduction

The mammalian genome is organized at multiple length scales into compartments, topologically associating domains (TADs) and chromatin loops[1–4]. This three-dimensional chromatin folding contributes to essential cellular processes such as transcription and DNA replication and is

actively maintained by protein cofactors [5–8]. The architectural protein CTCF establishes the anchors of many DNA loops and TADs likely by blocking loop extrusion by cohesin[7,9–16]. 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 [17–21]. Furthermore, CTCF binding and looping on or near genes can directly enforce enhancer-promoter loops and facilitate differential splicing and polyadenylation [22–29]. 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 [6,7,30,31].

CTCF binding is sensitive to DNA methylation, and consequently global DNA demethylation 'reactivates' hundreds to thousands of previously masked CTCF peaks [32–36]. However, these reactivated CTCF sites comprise only a small subset of all potential methylated CTCF sites, suggesting that DNA methylation serves as a specialized mechanism to regulate specific CTCF binding sites [35]. While methylation-sensitive CTCF peaks have been linked to gene expression and mRNA processing in some instances,[20,21,24–26,37] the functions of most methylation-sensitive CTCF peaks are unknown. Moreover, the broader impacts of reactivated CTCF binding upon global DNA demethylation on chromatin looping have been largely unexplored.

In this study, we employed a selective inhibitor of DNMT1, the maintenance DNA methyltransferase, to induce genome-wide DNA demethylation and investigated the chromatin looping patterns of the resulting reactivated CTCF peaks. We found that reactivated CTCF peaks within genes form loops to highly-looping partner peaks located in active chromatin regions. Remarkably, reactivated CTCF peaks and their highly-looping partners are located close to nuclear speckles, condensate bodies of protein and RNA that have been implicated in

transcription and splicing [38–44]. Subnuclear localization relative to nuclear landmarks such as nuclear speckles is linked to genome structure and function, so we explored this relationship further by developing acute degron systems for both CTCF and speckles. Through these experiments, we clarify the directionality of the relationship between CTCF and speckles and glean novel insights into how methylation-sensitive CTCF contributes to chromatin structure and gene expression.

Results

DNMT1 inhibition reactivates CTCF peaks and loops on gene bodies

Loss of DNA methylation leads to increased CTCF occupancy at hundreds to thousands of CTCF sites depending on the cell type[35], but a global understanding of their looping patterns and functions is lacking. Recently, selective, noncovalent DNMT1 inhibitors (e.g., GSK3484862) have become available, enabling dynamic control of DNA methylation with fewer undesirable effects such as DNA damage and cytotoxicity [45,46]. Consequently, we employed these new DNMT1 inhibitors to probe the functions and looping patterns of methylation-sensitive CTCF sites.

We treated K562 and HCT116 cells with GSK3484862 (here after referred to as DNMT1i, 10 μM) for three days, and performed CTCF ChIP-seq and HiChIP to investigate CTCF-associated changes in chromatin looping (**Fig 1A**). Under these conditions, DNMT1i treatment led to minimal growth defects in both cell types (**Fig S1A**). Consistent with prior studies [35], DNMT1i treatment increased CTCF binding at thousands of sites, previously termed “reactivated peaks” due to their overlap with CTCF peaks in other cell types[35,47] (**Fig 1B,C,** 2,079 up- and 2 down-regulated in K562, 6,686 up- and 328 down-regulated in HCT116). As expected, reactivated CTCF sites were

largely methylated in wild-type cells (**Fig S1B**), and many CTCF binding motifs enriched within these peak regions were methylated at two key CpG sites implicated in CTCF binding (**Fig S1C**) [36,47].

We next evaluated whether these reactivated CTCF peaks form chromatin loops. CTCF HiChIP revealed that a large subset of reactivated peaks resides within at least one loop anchor and these are henceforth called "looping" peaks (see Methods) [48,49]. An example of reactivated looping CTCF peaks is shown at the *PCDHA* locus (**Fig 1D**). On average, loops that are anchored by reactivated peaks increase in strength upon DNMT1 inhibition (**Fig 1E**). We next sought to evaluate these findings using our published Hi-C dataset following DNMT1 inhibition [50]. Specifically, we analyzed Hi-C interaction strength using standard aggregate pileup analysis over all reactivated CTCF sites identified in our ChIP-seq data. Consistent with our Hi-ChIP data, this analysis revealed an increase in Hi-C interaction signal at reactivated peaks following 3 days of DNTM1 inhibition (**Fig 1F**).

We next characterized the genomic annotations of looping and nonlooping reactivated CTCF peaks. Consistent with previous findings using azacitidine [35], loss of DNA methylation through DNMT1i treatment led to the preferential reactivation of CTCF peaks on gene bodies, particularly in non-intron regions (exons, promoters, and transcription termination sites (TTS), **Fig 1G, S1D, S1E**). Notably, this enrichment on gene bodies was particularly pronounced looping reactivated CTCF peaks (**Fig 1G, Fig S1D, S1E, S1F**), highlighting an underappreciated propensity for CTCF peaks on gene bodies to engage in chromatin loops. Altogether, selective DNMT1 inhibition leads to the emergence of thousands of reactivated CTCF peaks on gene bodies that engage in new chromatin interactions.

Figure 1

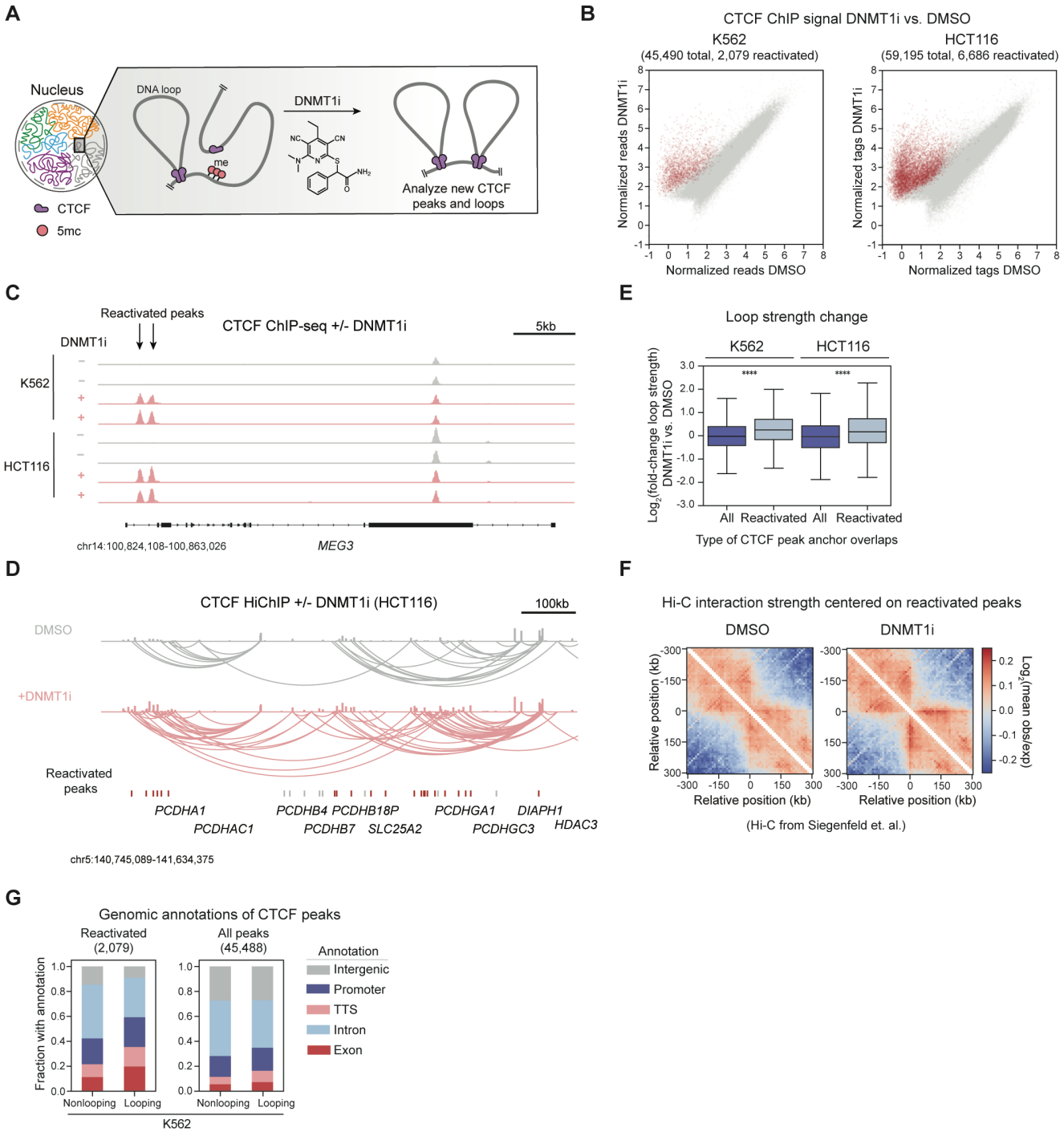
A
Nucleus
DNA loop
DNMT1i
me
Analyze new CTCF peaks and loops
CTCF
5mc
B
CTCF ChIP signal DNMT1i vs. DMSO
K562
(45,490 total, 2,079 reactivated)
HCT116
(59,195 total, 6,686 reactivated)
Normalized reads DNMT1i
Normalized reads DMSO
Normalized tags DNMT1i
Normalized tags DMSO
C
Reactivated peaks
CTCF ChIP-seq +/- DNMT1i
5kb
DNMT1i
K562
HCT116
chr14:100,824,108-100,863,026
MEG3
D
CTCF HiChIP +/- DNMT1i (HCT116)
100kb
DMSO
+DNMT1i
Reactivated peaks
PCDHA1
PCDHAC1
PCDHB4
PCDHB7
PCDHB18P
SLC25A2
PCDHGA1
PCDHGC3
DIAPH1
HDAC3
chr5:140,745,089-141,634,375
E
Loop strength change
K562
HCT116
Log2(fold-change loop strength) DNMT1i vs. DMSO
All
Reactivated
Type of CTCF peak anchor overlaps
F
Hi-C interaction strength centered on reactivated peaks
DMSO
DNMT1i
Relative position (kb)
Log2(mean obs/exp)
(Hi-C from Siegenfeld et. al.)
G
Genomic annotations of CTCF peaks
Reactivated (2,079)
All peaks (45,488)
Fraction with annotation
Nonlooping
Looping
K562
Annotation
Intergenic
Promoter
TTS
Intron
Exon

Figure 1 | DNMT1 inhibition reactivates CTCF peaks and loops on gene bodies

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 Integrated Genome Browser (IGV) CTCF ChIP-seq tracks ± DNMT1i at the MEG3 locus for K562 (top) and HCT116 (bottom). Reactivated peaks indicated with arrows.
D. 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.
E. 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).
F. 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.
G. Proportion of looping and nonlooping peaks with each annotation (y-axis) for both reactivated and all CTCF peaks in K562 in DNMT1i.

In (**E**), 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 interact with highly-looping partners near nuclear speckles

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 2A**), particularly for K562 cells. 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 S2A**). 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 partners 'highly-looping', analogously to how superenhancers or supersilencers were previously characterized (**Fig 2B, 2C, S2B**) [51,52]. HiChIP loops between reactivated peaks and highly-looping

peaks increased in strength upon DNMT1 inhibition, consistent with increased interactions between these sites (**Fig S2C**). The 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 [53–55]. Stripe anchors are enriched for chromatin accessibility and histone modifications associated with active enhancers [53]. Similarly, regions near highly-looping CTCF peaks are accessible and enriched for active chromatin marks such as H3K9ac and H3K27ac (**Fig 2D**). Furthermore, based on previously published Hi-C data [50], highly-looping CTCF sites have a stronger propensity to engage in chromatin interactions compared to ordinary CTCF sites (**Fig 2E**). In addition, stripe anchors identified from these Hi-C data also reside closer to highly-looping CTCF peaks than all peaks, consistent with the similarities between these features (**Fig S2D**). Thus, these results show that upon DNA demethylation, reactivated CTCF sites form loops to highly-looping partners associated with active regulatory marks and stripe anchors.

To understand the features that distinguish reactivated CTCF peaks and their highly-looping partners from other CTCF peaks, we next considered their subnuclear localization. To do so, we investigated their distribution across genomic subcompartments that exhibit variable transcriptional activity and association with nuclear landmarks, such as the nuclear lamina or nuclear speckles. In particular, we found a striking enrichment of reactivated CTCF peaks and highly-looping CTCF peaks in the A1 subcompartment of the genome, which is differentiated from the A2 subcompartment largely due to its localization near nuclear speckles (**Fig S2E**) [56]. Studies have shown that nuclear speckles are hubs of DNA-DNA contacts [42,44,57–61]. Additionally, CTCF can form stress-responsive clusters near nuclear speckles in some cell types, and genomic subcompartments associated with speckles have been correlated with increased CTCF binding and looping [38,41,62,63]. However, a direct connection between nuclear speckles and CTCF binding and looping has not been extensively explored.

To determine if reactivated peaks and highly-looping CTCF peaks are preferentially located near nuclear speckles, we performed Cut&Tag for SON, a core speckle protein, in K562 and HCT116 cells with and without DNMT1 inhibition (**Fig 2F**) [38,40,58,64]. Consistent with localization near speckles, reactivated peaks appeared in regions of higher SON signal compared to all CTCF peaks (**Fig 2G, S2F**). Notably, this trend is strong for looping reactivated CTCF peaks (**Fig 2G**). This enrichment of looping reactivated peaks near speckles is further corroborated by publicly available SON TSA-seq data in wildtype cells (**Fig S2G**) [40]. Thus, reactivated CTCF peaks reside closer to nuclear speckles than non-reactivated peaks.

Strikingly, relative to all CTCF peaks, highly-looping CTCF peaks are enriched at nuclear speckles (**Fig 2H, S2H**). More broadly, the number of different loops a CTCF peak engages in and their corresponding strength are positively correlated with SON Cut&Tag signal (**Fig 2I, S2I**). In addition, we observed that CTCF peaks with strong Cut&Tag SON signal preferentially engage in strong shorter-range interactions in published Hi-C interaction data (**Fig 2J**) [50]. Overall, our results reveal that reactivated CTCF peaks interact with highly-looping partners located near nuclear speckles.

Figure 2

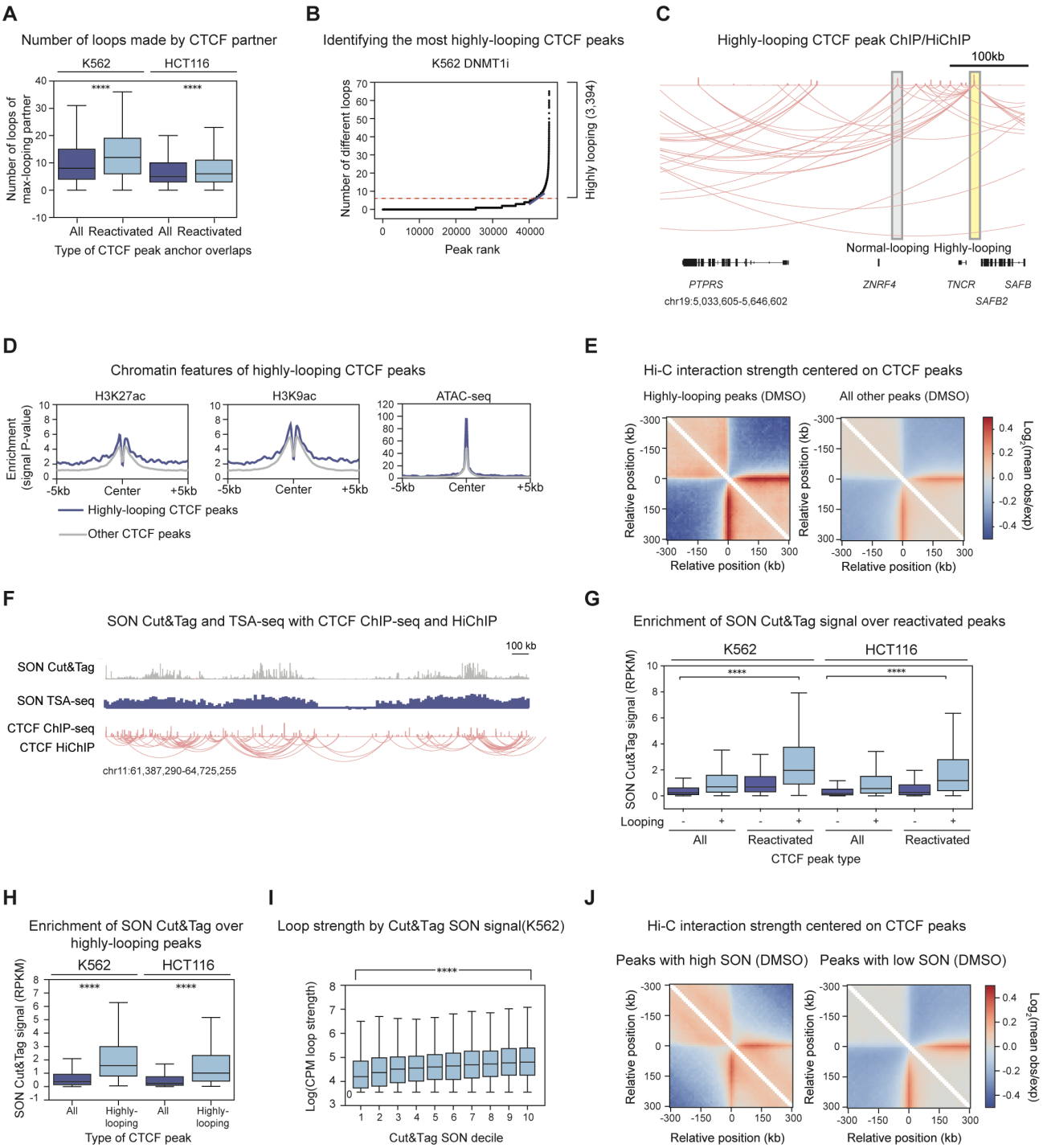
A
Number of loops made by CTCF partner
K562
HCT116
Number of loops of max-looping partner
All Reactivated All Reactivated
Type of CTCF peak anchor overlaps
B
Identifying the most highly-looping CTCF peaks
K562 DNMT1i
Number of different loops
Highly looping (3,394)
Peak rank
C
Highly-looping CTCF peak ChIP/HiChIP
100kb
Normal-looping Highly-looping
PTPRS
ZNRF4
TNCR
SAFB
SAFB2
chr19:5,033,605-5,646,602
D
Chromatin features of highly-looping CTCF peaks
H3K27ac
H3K9ac
ATAC-seq
Enrichment (signal P-value)
-5kb Center +5kb
Highly-looping CTCF peaks
Other CTCF peaks
E
Hi-C interaction strength centered on CTCF peaks
Highly-looping peaks (DMSO)
All other peaks (DMSO)
Relative position (kb)
Log2(mean obs/exp)
F
SON Cut&Tag and TSA-seq with CTCF ChIP-seq and HiChIP
100 kb
SON Cut&Tag
SON TSA-seq
CTCF ChIP-seq
CTCF HiChIP
chr11:61,387,290-64,725,255
G
Enrichment of SON Cut&Tag signal over reactivated peaks
K562
HCT116
SON Cut&Tag signal (RPKM)
Looping
All Reactivated All Reactivated
CTCF peak type
H
Enrichment of SON Cut&Tag over highly-looping peaks
K562
HCT116
SON Cut&Tag signal (RPKM)
All Highly-looping All Highly-looping
Type of CTCF peak
I
Loop strength by Cut&Tag SON signal(K562)
Log(CPM loop strength)
Cut&Tag SON decile
J
Hi-C interaction strength centered on CTCF peaks
Peaks with high SON (DMSO)
Peaks with low SON (DMSO)
Relative position (kb)
Log2(mean obs/exp)

Figure 2 | Reactivated CTCF peaks interact with highly-looping partners near nuclear speckles

- **A.** Number of different loops (loops to different partners) the maximally-looping partner peak makes (y-axis) for all vs. reactivated CTCF peaks (x-axis) in DNMT1i for K562 (left) and HCT116 (right).
- **B.** Number of different loops (number of partner peaks) (*y*-axis) assigned to each peak by rank (*x*-axis) in DNMT1i for K562. Highly-looping peaks reside above the red dashed line.
- **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 (left) and all other (right) peaks in the DMSO treatment condition in K562 cells. K562 Hi-C data from Siegenfeld et. al., 2022.
- **F.** 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.
- **G.** 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 (x-axis).
- **H.** Same as G, but for highly-looping peaks vs. all other CTCF peaks.
- **I.** 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).
- **J.** Heatmap depicting $\log_2$(mean observed/expected) Hi-C interaction frequency centered on CTCF peaks at speckles (denoted by high SON signal, left) and not at speckles (right) in DMSO. K562 Hi-C data from Siegenfeld et. al., 2022.

In (**A,G,H,I**), 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; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

DNMT1i-induced gene activation and speckle association depend on CTCF

Our data revealed that CTCF reactivation and looping are highly correlated with nuclear speckle association. As such, we next sought to study the nature of this relationship through degrading either CTCF or nuclear speckles. We first generated a HaloTAG-CTCF homozygous knock-in HCT116 cell line, which enabled the inducible recruitment of CTCF to the VHL E3 ubiquitin ligase upon addition of the small molecule HaloPROTAC3, causing subsequent proteolytic degradation within 8 hours (**Fig 3A, 3B, S3A**). We then performed SON Cut&Tag after pretreating our CTCF-HaloTag knockin cells with HaloPROTAC3 or vehicle for 8 hours before co-treating with DNMT1i or vehicle for 24 hours (32 hours total for HaloPROTAC3 treatment). This shorter timepoint was

used to capture direct effects of CTCF degradation and DNA demethylation as well as avoid toxicity caused by CTCF depletion. We found that reactivated CTCF peaks gain SON Cut&Tag signal upon DNMT1i treatment (**Fig 3C**). This gain in SON signal is blocked when CTCF is degraded by HaloPROTAC3 treatment, suggesting that speckle association at these reactivated sites is dependent on CTCF (**Fig 3C**). Thus, SON associates with reactivated CTCF peaks in a CTCF-dependent manner.

Because reactivated peaks fail to engage with nuclear speckles if CTCF is not present, we hypothesized that gene activation induced upon loss of DNA methylation may be dampened in the absence of CTCF. Prior work has suggested that cell-type specific DNA methylation may protect a “labile” set of CTCF peaks from reactivation [35] and that methylation-dependent CTCF occupancy on gene bodies has been linked to changes in gene processing, but a global understanding of the functions of methylation-sensitive CTCF peaks on transcription remains lacking [24,26,28]. Thus, we tested whether reactivated CTCF peaks promote gene activation upon DNMT1 inhibition.

To assess the role of CTCF in enabling DNMT1i-dependent gene activation, we pretreated our CTCF-HaloTag knockin cells with HaloPROTAC3 or vehicle for 8 hours and then co-treated the cells with DNMT1i or vehicle for 24 hours before performing RNA-seq (**Fig 3A, 3D**). Consistent with prior studies [7], acute CTCF degradation altered the expression of hundreds of genes with approximately equal numbers up- and down-regulated (488 up-regulated, 338 down-regulated, Padj <0.05, $|\log_2\text{fold-change}| > 0.5$) (**Fig S3B**), with downregulated genes possessing higher baseline levels of CTCF at their promoters (**Fig S3C**). As expected, DNMT1i treatment in this cell line for 24 hours primarily resulted in gene upregulation, with 135 genes upregulated and 8 genes downregulated (Padj <0.05, $|\log_2\text{fold-change}| > 0.5$) (**Fig S3B**).

We focused our analysis on genes that were differentially expressed in DNMT1i versus DMSO treatment in the presence of CTCF (**Fig S3B, left**). Using K-means clustering, we identified four gene clusters within this group defined by unique z-score normalized expression patterns across the conditions (**Fig 3D**). Clusters 1, 2, and 4 comprise genes that increase in expression upon DNMT1 inhibition, whereas Cluster 3 contains the few genes that decrease in expression upon DNMT1 inhibition**.** Moreover, genes in Cluster 1 are upregulated by CTCF depletion, whereas genes within Clusters 2, 3, and 4 are unaffected by CTCF depletion alone.

We focused on Clusters 2 and 4, which contain lowly-expressed genes that are upregulated upon DNMT1 inhibition and unaffected by CTCF depletion in the absence of DNMT1i (**Fig 3D, S3D**). Significantly, while genes in Cluster 4 are upregulated upon DNMT1 inhibition with or without CTCF present, genes in Cluster 2 depend on CTCF for DNMT1i-mediated upregulation. To determine how CTCF binding might be related to the observed changes in gene expression, we performed CTCF ChIP-seq in our HaloTag-CTCF HCT116 cells following 24 hours of DNMT1i or vehicle treatment. Notably, genes in Cluster 2 preferentially gained CTCF occupancy upon DNMT1 inhibition compared to genes in other clusters (**Fig 3D, right, S3E**). Thus, these data demonstrate that a subset of genes enriched for CTCF reactivation depend on CTCF for expression upon DNMT1i treatment.

We next assessed the correlation between CTCF-dependent genes and speckle association. Intriguingly, genes in Cluster 2 have higher SON TSA-seq levels than genes in Cluster 4 (**Fig 3E**), while genes in cluster 1, which are upregulated by CTCF depletion alone, are closest to nuclear speckles. Moreover, genes in cluster 2 gain a modest level of SON Cut&Tag signal upon DNMT1i treatment (**Fig S3F)**. Taken together, these experiments show that CTCF is necessary for both gene activation and speckle protein association upon DNMT1 inhibition.

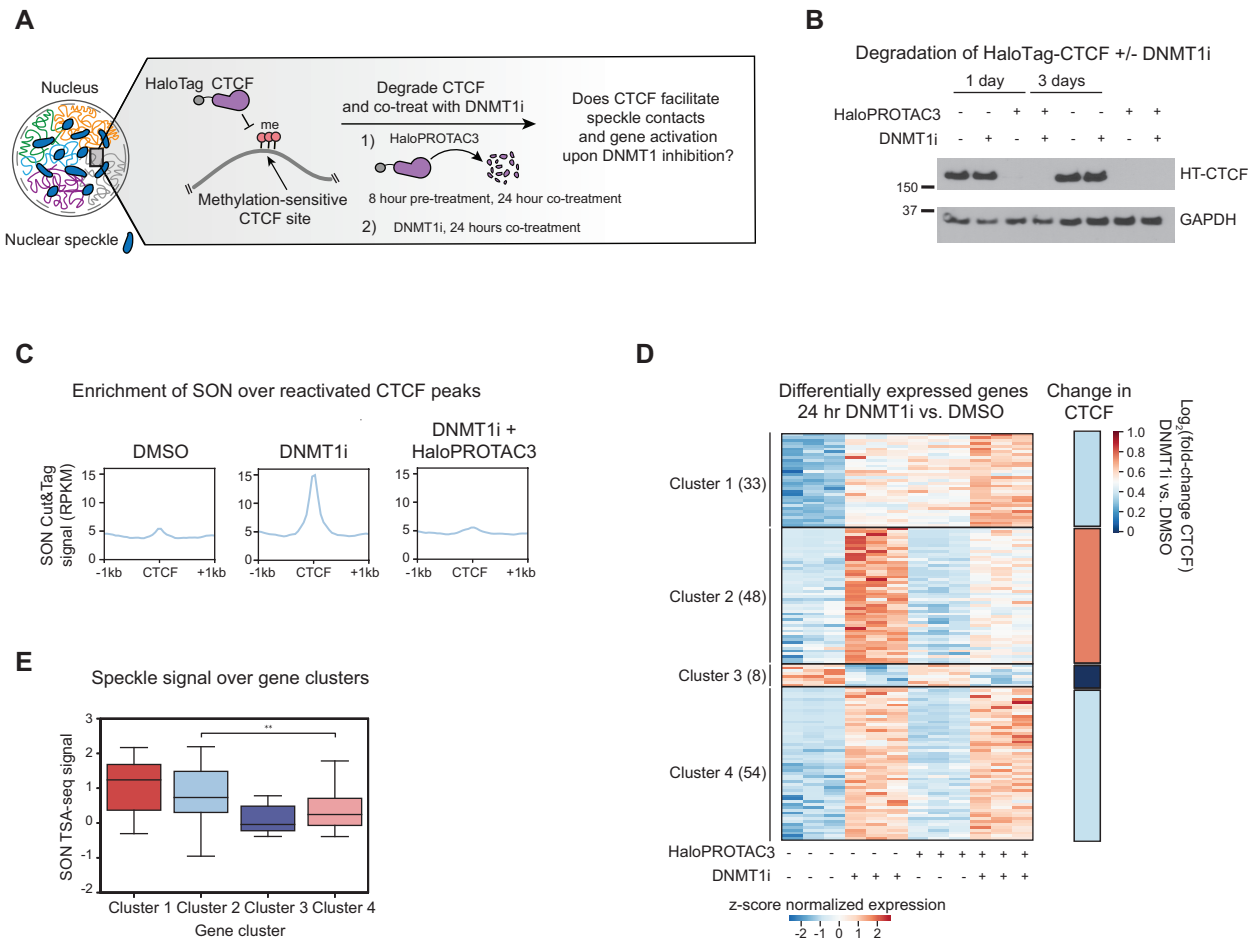

Figure 3 | DNMT1i-induced gene activation and speckle association depend on CTCF

A. Schematic showing the CTCF degradation experimental workflow. Following pre-treatment with HaloPROTAC3 for 8 hours to degrade HaloTag-CTCF, cells are co-treated with HaloPROTAC3 and DNMT1i for 24 hours. DMSO controls without CTCF depletion and/or DNMT1 inhibition were also included.

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. 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. 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, $|\log_2(\text{fold-change})| > 0.5$) 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 $\log_2$(fold-change) in CTCF binding in DNMT1i vs. DMSO for all CTCF peaks on genes within the four gene clusters.

E. 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 **D**.

For (**E**), 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 further explore the observed connection between speckles and CTCF looping, we tested whether nuclear speckles 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 [65]. 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 [58,66]. Upon addition of dTAG-13, FKBP12^{F36V}-tagged proteins are recruited to the CRBN-Cul4-Ring E3 ligase and undergo acute proteasomal degradation (**Fig 4A**). 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 4B, S4A**).

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 S4B**). Consistent with a role of nuclear speckles in mediating active transcription, a large majority of down-regulated genes were located within the closest decile to nuclear speckles by TSA-seq (**Fig S4C**), and genes with the highest SON Cut&Tag signal exhibited the strongest decrease in expression (**Fig S4D**). These results show that disrupting nuclear speckles has immediate effects on RNA abundance.

To determine whether speckle disruption alters CTCF binding and/or looping, we performed CTCF ChIP-seq and HiChIP in the speckle dTAG cell line after treatment with dTAG-13 or DMSO for 6 hours. In contrast to the immediate transcriptional effects, CTCF occupancy remained unchanged upon speckle degradation, with only 2 significantly downregulated peaks (**Fig 4C,**

S4E). CTCF HiChIP also revealed that loop strength was largely unchanged across speckle Cut&Tag and TSA-seq quantiles (**Fig 4D, S4F**) despite a positive correlation between speckle association and baseline loop strength (**Fig 2I**). Even after 12 hours of dTAG treatment, there was only a very subtle decrease in CTCF looping near speckles (**Fig S4F**). In addition, changes in gene expression were not associated with changes in loop strength or CTCF binding (**Figure S4G, H**). Conversely, depleting CTCF for 24 hours in our HaloTag-CTCF HCT116 knockin cell line did not noticeably change speckle morphology (**Fig S4I**). Altogether, acute disruption of nuclear speckles did not appreciably disrupt CTCF binding or looping, suggesting that maintenance of 3D CTCF loop structure is independent of speckle association. This complements our finding that conversely, speckle association depends upon reactivated CTCF.

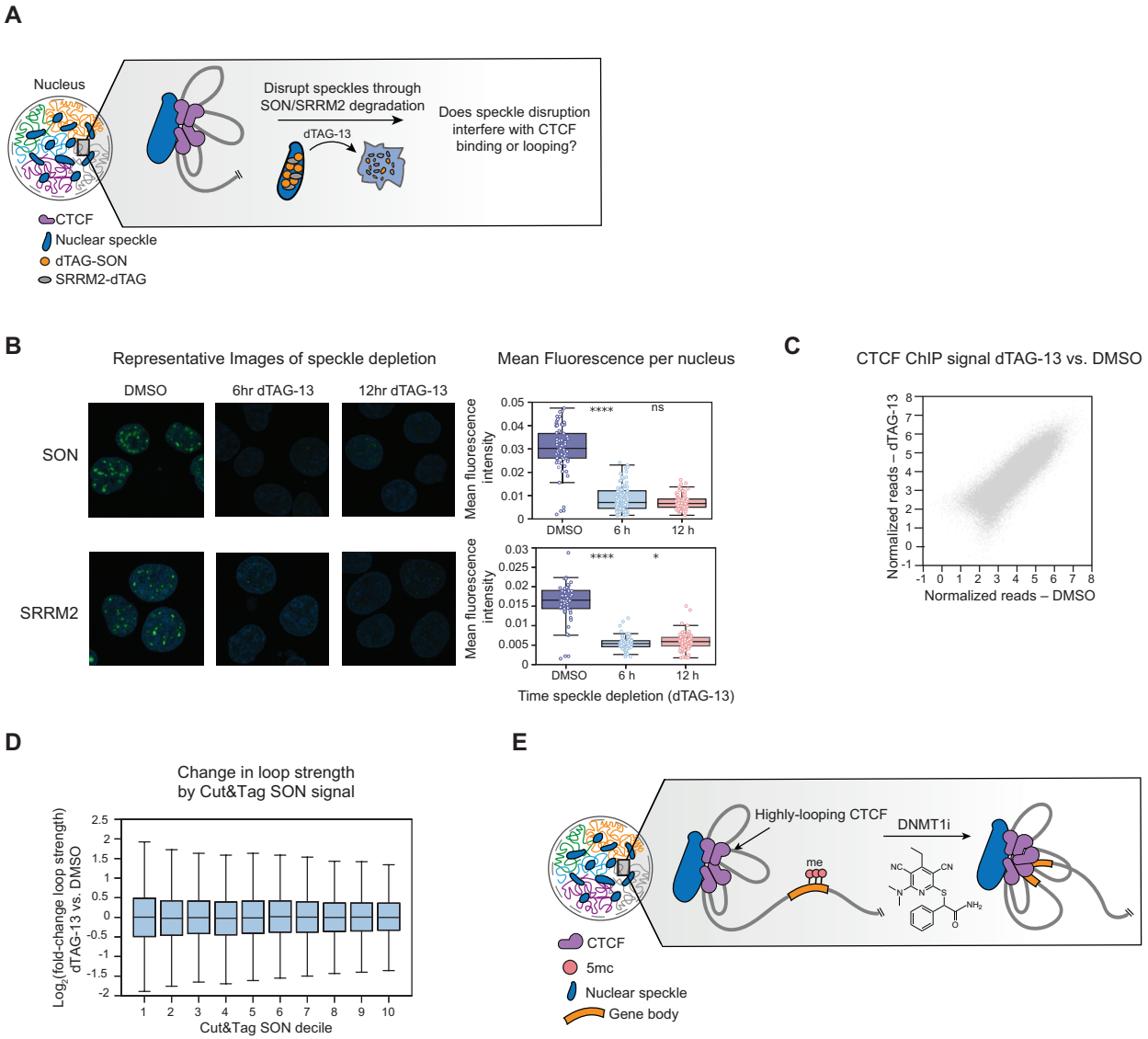

Figure 4 | Acute disruption of nuclear speckles alters gene expression without disrupting CTCF

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. Right: 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.

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

D. 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) for 6 hours relative to SON Cut&Tag signal (x-axis). Loops are segregated into equally sized deciles by Cut&Tag signal with 10 representing the decile closest to speckles (RPKM, 20kb bins).

E. Model for methylation-mediated insulation of genes from regulatory elements near nuclear speckles

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; *: $0.01 < p \leq 0.05$; **: $0.001 < p \leq 0.01$; ***: $0.0001 < p \leq 0.001$; ****: $p \leq 0.0001$.

Discussion

DNA methylation accounts for much of the variation in CTCF binding across cell types [35], but the functions and features of methylation-sensitive CTCF peaks are incompletely understood. Here we leveraged a selective DNMT1 inhibitor [45,46] 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 [53–55,67], and may also represent a mechanism to poise transcription [55]. Thus, we propose that reactivated peaks enable transcriptional activation by interacting with regulatory elements through highly-looping CTCF partner peaks (**Fig 4E)**. However, DNA demethylation causes pleiotropic effects ranging from changes in gene expression to transcription factor binding [45,68], and we cannot rule out that these other features could contribute to the effects we observe.

Strikingly, reactivated CTCF peaks and their highly-looping partners are located near nuclear speckles, consistent with prior studies showing that subcompartments near nuclear speckles exhibit increased CTCF binding and looping [41,62]. We first explored this relationship by degrading CTCF and intriguingly observed that reactivated CTCF peaks gain SON Cut&Tag signal upon DNMT1 inhibition in a CTCF-dependent manner, indicating a functional tie between reactivated CTCF and speckles. This is consistent with a recent preprint reporting that CTCF helps establish basal speckle-gene contacts[69]. Additionally, our CTCF degron experiments demonstrate that a large fraction of genes that are differentially expressed upon DNMT1 inhibition depends on CTCF

for up-regulation. These genes preferentially gain CTCF binding and are located near nuclear speckles. Consequently, CTCF peaks that are reactivated upon DNMT1 inhibition could facilitate the activation of genes near speckles, perhaps by mediating the engagement of genic CTCF sites with speckle-associated stripe anchors and/or splicing factors, although this model requires further study (**Fig 4E**). A role of CTCF in facilitating chromatin loops near speckles is synergistic with proposed models suggesting that intragenic looping contributes to gene splicing [25,70]. Thus, reactivated CTCF peaks may reorganize genes near nuclear speckles as a means of influencing transcription.

We next studied whether speckles impact CTCF binding or looping by achieving the first acute degradation of nuclear speckles. Previous work knocking down SRRM2 support a causative role of speckles in modulating chromatin structure but did not specifically study changes in CTCF binding or looping [71]. Although acute disruption of speckles supports their direct involvement in transcription, we observed minimal impacts on CTCF binding or looping. This notion is consistent with recent work showing that CTCF clustering and looping is not dependent on phase separation or nuclear bodies [72,73] and that transcriptional disruption does not immediately impact chromatin looping or CTCF binding [53,74,75]. Our findings also suggest that the genome interactions disrupted upon SRRM2 depletion [71] are not mediated by CTCF or only occur after a longer time. However, it is possible that nuclear speckles contribute to the establishment, but not maintenance, of CTCF loops, which may not be evident in our system due to the short time points employed for speckle degradation. Moreover, other features enriched near speckles, such as open chromatin, high GC content, and active gene expression, could instead contribute to increased CTCF binding and looping near speckles. Overall, our speckle degradation experiments help clarify the directionality of the relationship between speckles and CTCF.

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

Methods

Cell culture

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

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

To generate an N-terminal Puro-P2A-dtag SON knock-in donor plasmid, SON homology arms of 1 kb were amplified from K562 genomic DNA and cloned into the ph1 backbone Puro-P2A-dTAG, sub-cloned from pCRIS-PITChv2-Puro-dTAG, a gift from J. Bradner and B. Nabet (Addgene # 91793). The SON sgRNA, 5'-CACCGCTGCTCGATGTTGGTCGCCA-3' that was previously used for N-terminal SON knock-in was cloned into PX459 by restriction enzyme digestion (Allen Institute,https://www.allencell.org/cell-catalog.html). To generate a C-terminal dTAG-P2A-blasticidin SRRM2 knock-in donor plasmid, SRRM2 homology arms of 1 kb were amplified from K562 genomic DNA and cloned into the ph1 backbone dTAG-P2A-BSD, sub-cloned from pCRIS-PITChv2- dTAG-BSD, a gift from J. Bradner and B. Nabet (Addgene # 91795). The sgRNA, 5'CACCGCCATGAGACACCGCTCCTCC -3' that was previously validated for C-terminal SRRM2 knock-in\was cloned into PX459 by restriction enzyme digestion [58]. Notably, this strategy eliminates the last four amino acids of SRRM2, which is necessary due to a sparsity of guides in

the region and has been used to study SRRM2 in the context of speckles [58]. Cells were transfected with both the donor and guide plasmids by nucleofection (Lonza, program T016). Cells were selected 3 days after transfection with puromycin (2 µg/mL) (Thermo Fisher Scientific, A1113803) or blasticidin (8 µg/mL) (Thermo Fisher Scientific, A1113903) for a duration of 5-10 days. Single-cell clones were isolated from the bulk cell line through fluorescence-activated cell sorting. The SON knock-in was generated from wildtype K562 cells, and then the SRRM2 knock-in was introduced on top of the validated SON knock-in clone. To generate the HT-CTCF HCT116 knock-in cell line, the sgRNA 5'-CACCGTGCAGTCGAAGCCATTGTGG-3' was subcloned into PX459 and HaloTAG-CTCF was subcloned into PDONR223. The guide and donor plasmids were transfected into cells using Fugene HD (Promega). After seven days, single cell clones were isolated from the bulk population through fluorescence-activated cell sorting using HaloTag TAMRA labeling to sort potential positive clones.

Immunoblotting

Pellets were harvested for western blot analysis and washed with PBS (Corning, 21-040-CV). For immunoblots of CTCF, cells were lysed on ice using RIPA buffer (Boston BioProducts) supplemented with fresh HALT Protease Inhibitor (Thermo Fisher Scientific) and the lysates were cleared through centrifugation. The protein concentration of the lysate was determined using the BCA Protein Assay Kit (Thermo Fisher Scientific) and immunoblots for CTCF were performed according to standard procedures with a 6% SureCast acrylamide page gel. For immunoblots of SON and SRRM2 and corresponding loading controls, 1 million cells per condition were lysed on ice for 30 minutes in 50uL RIPA buffer (Boston Bioproducts, BP-115) supplemented with 1mM PMSF, 2mM EDTA, and 1X Halt Protease inhibitors (Thermo Fisher Scientific, 78429). Subsequently, 50uL of wash buffer was added (50mM Tris pH 8.0, 2mM MgCl, 50mM NaCl, 1mM phenylmethylsulfonyl fluoride (PMSF), 1X Halt Protease inhibitors, 1:500 Benzonase (Millipore-Sigma, E1014-25KU) and the samples were rotated at room temperature for 20 minutes and then

clarified by full speed centrifugation at 4 °C. The protein concentration of the lysates was determined using the BCA Protein Assay Kit (Thermo Fisher Scientific, 23225). Immunoblotting was performed according to standard procedures with the following modifications: a 3-8% tris acetate gel (Thermo Fisher Scientific, EA03752) was used and samples were run for 2 hours 15min at 80V at 4 °C in 1X tris-acetate running buffer (Thermo Fisher Scientific, LA0041) except for the loading control which was run for 90 minutes. Transfer was conducted in 1X transfer buffer (Research Product International, 501977679) containing 10% methanol and 0.05% sodium dodecyl sulfate (SDS) onto an activated 0.45um PVDF membrane (Millipore sigma IPVH09120) for 2 hours at 15V except for loading control which was transferred in 1X transfer buffer with 20% methanol at 12V for 90 minutes. The primary antibodies used for immunoblotting are as follows: CTCF (1:2000, Cell Signaling Technology D31H2); SON (1:1000, Abcam ab121759); SRRM2 (1:1000, ThermoFisher PA5-66827); GAPDH (1:2000, Santa Cruz Biotechnology sc-47724); Vinculin (1:2000, Cell Signaling Technology 13901S).

Immunofluorescence

Immunofluoresence was performed according to standard protocol. Briefly, K562 speckle knock-in cells treated with 500 nM dTAG-13 or DMSO for the indicated amount of time were cytospun onto glass coverslips for 5 minutes at 800 x rcf and fixed in 4% PFA for 15 minutes. HCT116 cells treated with 330 nM HaloPROTAC3 or DMSO were instead grown directly on glass coverslips and fixed in 4% PFA for 15 minutes. Cells were then washed 3X in PBS, permeabilized with 0.25% Triton-X for 10 minutes, and subsequently washed 3X with PBS. Cells were then blocked in 10% goat serum (Millipore Sigma, G9023) in PBS-T for 1 hour and incubated overnight in 5% goat serum in PBS-T with primary antibody at 4°C (SON (1:300, abcam ab121759) or SRRM2 (1:300, ThermoFisher PA5-66827)). Following 3X PBS-T washes, cells were incubated with secondary antibody for 1 hr (1:2000, Invitrgoen alexafluor488 A-11008), washed 3X in PBS-T, and mounted on a coverslip with DAPI (Prolong Diamond Anti-Fade, P36961). Cells were imaged

on an LSM180 Confocal microscope. Image quantification was performed using CellProfiler on raw Tiff images to identify nuclei and then quantify average fluorescence signal within each nucleus[76].

Cell growth assays

K562 or HCT116 cells were plated in triplicate in a 96-well plate. Drug or vehicle (DMSO, Sigma Aldrich, D8418) were dosed at the specified concentration for 3 days. After 3 days of treatment, cell viability was measured using CellTiter-Glo (Promega, G7572) with the luminescence detector on the SpectraMax i3x (Molecular Devices) plate reader with the SoftMax Pro (version 6.5.1) software, and data were processed using Prism.

ChIP-seq

CTCF ChIP-seq was performed in duplicate. For the CTCF reactivation experiments, K562 or HCT116 wildtype or HT-CTCF HCT116 knock-in cells were treated with 10 μM GSK3482364 or 0.1% DMSO control for the indicated time. For the speckle depletion experiments, K562 speckle dTAG knock-in cells were treated with 500 nM dTAG-13 or 0.1% DMSO control for 6 or 12 hours. After drug treatment, cells were resuspended and fixed at a density of 10^6 cells/mL in media containing 10% FBS (Peak Serum, PS-FB2) and 1% formaldehyde (Sigma Aldrich, F8775) for 10 min with rotation and quenched with 0.2 M glycine and washed in PBS. 5 million cells per condition were lysed on ice for 10 min in 600 uL of ChIP lysis buffer (50 mM Tris-HCl pH 7.5, 1% SDS, 0.25% NaDOC) with HALT protease inhibitors (Thermo Fisher Scientific, 78429). Cells were then diluted to 2mL with ChIP dilution buffer (50 mM Tris-HCL pH 7.5, 0.01% SDS, 150 mM NCl, 0.25% triton X-100, 1X HALT protease inhibitor) and sonicated at 4 °C with a Branson Sonicator (45% amplitude, 0.7s on, 1.3s off, total time 5 min). The supernatant was clarified by centrifugation at 16000 x rcf for 10 min at 4 °C . The sample was then diluted into 6 mL total with ChIP Dilution Buffer and the final concentration of triton was brought to 1%. 50 μL was removed as an input

sample and kept at 4 °C overnight. Immunoprecipitations were carried out on the remaining sample overnight at 4 °C with rotation with 5 uL of antibody (Cell Signaling Technology D31H2 anti-CTCF, 3418S). 50uL of Protein G Dynabeads (Thermo Fisher Scientific, 10004D) were added to each sample and incubated at 4 °C with rotation for 3 hours. Supernatant was discarded on a magnet and beads were then washed twice with ice-cold RIPA wash buffer (10 mM Tris-HCl pH 8.1, 0.1% SDS, 150 mM NaCl, 0.1 NaDOC, 1% Triton X-100, 1 mM EDTA), once with ice-cold LiCl wash buffer (10 mM Tris-HCl, pH 8.1, 250 mM LiCl, 0.5% Triton X-100, 0.5% NaDOC), and once with ice cold TE buffer (10mM Tris-HCL pH 8.0, 50mM NaCl, 1mM EDTA). Chromatin was eluted in 100 µL of Elution buffer (10 mM Tris-HCl pH 8.0, 0.1% SDS, 150 mM NaCl, 1 mM EDTA) at 65 °C for 1 hour at 1000rpm. Input samples were brought to 100µL with Elution buffer and incubated alongside ChIP samples. Supernatant was removed and 50 µg of RNAse A (Sigma Aldrich, 10109142001) was added and incubated for 30 min at 37 °C. Next, 25 µg of Proteinase K (Invitrogen, 25530015) was added, and the samples were incubated 63 °C for 3 hours with 1000rpm agitation. DNA was subsequently purified by a 2X SPRI (Omega Bio-Tek, M1378-01). Sequencing libraries were then prepared from 5 ng of DNA by performing end-repair with the End-it DNA End-Repair Kit (Lucigen, ER81050), followed by A tailing with NEB Klenow Fragment (3'−5' exo-) (New England Biolabs, M0212), adapter ligation with NEB DNA Quick Ligase (New England Biolabs, M2200), and PCR using NEB ultra mastermix (M0544S) and 10X KAPA Library Amp Primer Mix (Kapa Biosystems, KK2623). Samples were sequenced using a NovaSeq SP kit (Illumina) with 50 bp paired-end reads at a sequencing depth of approximately 25 million reads per sample.

HiChIP

Cells were treated in duplicate. For the DNMT1i experiment, K562 or HCT116 cells were treated with 10 µM GSK3482364 (ChemieTek) or 0.1% DMSO control for 3 days. For the speckle degron experiment, K562 speckle dTAG knock-in cells were treated with 500 nM dTAG-13 (Sigma

Aldrich, SML2601) (or 0.1% DMSO control for 6 or 12 hours. CTCF HiChIP was performed on each sample according to a published protocol [77]. Specifically, cells were resuspended and fixed at a density of 10^6 cell/mL in media containing 10% FBS (Peak Serum, PS-FB2) and 1% formaldehyde (Sigma Aldrich, F8775) for 10 min with rotation and quenched with 0.2 M glycine and washed in PBS. 10 million cells per condition were resuspended in ice-cold Hi-C Lysis Buffer (10mM Tris HCl pH 8.0, 10mM NaCl, 0.2% NP-40, 1X protease inhibitor) and lysed at 4 °C for 30 minutes. Tubes were centrifuged and the pellets were washed with 500 µL ice-cold Hi-C lysis buffer. Pellets were then resuspended in 100µL of 0.5% SDS and incubated at 62 °C for 10 minutes, after which they were quenched with Triton X-100 and incubated at 37 °C for 15 minutes. The samples were then digested with 375 units of DpnII (New England Biolabs, R0543L) for 2 hours at 37 °C with shaking at 900 rpm and then enzyme was heat inactivated at 65 °C for 20 minutes. Digested ends were filled in with biotinylated dATP (Life Technologies, 19524-016) for 1 hour at 37 °C at 900rpm, and in situ ligation was performed with T4 DNA ligase (New England Biolabs, M0202L) for 4 hours at room temperature with rotation. Nuclei were pelleted by centrifugation and resuspended in Nuclear Lysis Buffer (50mM Tris-HCL pH 7.5, 10mM EDTA, 1% SDS, 1X protease inhibitor) and sheared in Covaris Millitubes (cat # 520135) on a Covaris S220 sonicator with Fill Level 10, Duty Cycle 5, PIP 140, Cycles/Burst 200, for 4 minutes each. Samples were clarified by centrifugation and supernatant was split into two tubes and diluted with 2X the volume of ChIP Dilution Buffer (0.01% SDS, 1.1% triton X 100, 1.2mM EDTA, 16.7mM tris pH 7.5, 167mM NaCl). 30ul Pre-washed Protein G Dynabeads (Thermo Fisher Scientific, 10004D) were then added to each tube and samples were rotated for 1 hour at 4 °C. Supernatant was transferred to new tubes and antibody was added and allowed to bind overnight at 4 °C with rotation (15uL antibody per tube, Cell Signaling Technology D31H2 anti-CTCF, 3418S). 30uL of pre-washed Protein G Dynabeads were added to each tube and rotated at 4 °C for 2 hours. Beads were washed 3X each with Low Salt Wash Buffer (0.1% SDS, 1% triton X 100, 2 mM EDTA, 20 mM Tris HCL pH 7.5, 150 mM NaCl), High Salt Wash Buffer (0.1% SDS, 1% triton X 100, 2 mM

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

RNA-seq

In triplicate separate cultures, cells were treated with the indicated compounds. For the CTCF degron experiment, HaloTag-CTCF HCT116 knock-in cells were pretreated with 330 nM

HaloPROTAC3 (AOBIOUS AOB36136) or 0.1% DMSO control for 8 hours, and then were treated with 10μM GSK3482364 or 0.1% DMSO in addition to +/-330 nM HaloPROTAC3 for 24 more hours. For the speckle knockdown experiment, K562 speckle dTAG knock-in cells were treated with 500 nM dTAG-13 or 0.1% DMSO control for 6 hours. Total RNA was isolated from 1 million cells per condition using the RNeasy Plus Mini Kit (Qiagen, 74104). mRNA was isolated using the Poly(A) RNA Selection Kit V1.5 (Lexogen, 157.96) and total RNA seq libraries were prepared using the CORALL Total RNA-Seq Library Prep Kit (Lexogen, 095.24). Samples were sequenced using a NovaSeq SP kit (Illumina) 50 bp paired-end reads at a sequencing depth of approximately 30 million reads per sample for the speckle degron experiments. For the HaloTAG CTCF degron experiments, samples were instead sequenced to a depth of approximately 15 million reads per sample with 100bp paired end reads.

SON Cut&Tag

For the wildtype cell line experiments, K562 or HCT116 wildtype cells were treated in duplicate with either 10 uM GSK3482364 or 0.1% DMSO control for 3 days. For the CTCF degron experiment, HaloTag-CTCF HCT116 knock-in cells were pretreated with 330 nM HaloPROTAC3 (AOBIOUS AOB36136) or 0.1% DMSO control for 8 hours, and then were treated with 10 μM GSK3482364 or 0.1% DMSO in addition to +/- 330nM HaloPROTAC3 for 24 more hours. Cut&Tag was performed according to published protocol [64]. 1 million cells were harvested and washed in PBS. Cells were resuspended in ½ volume ice-cold NE1 buffer (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 0.5mM spermidine, 0.1% Triton-X-100, 20% glycerol with Roche Complete Protease Inhibitor EDTA-free) and lysed on ice for 10 minutes. Nuclei were resuspended in PBS and fixed with 0.1% formaldehyde (Sigma Aldrich, F8775) for 2 minutes at room temperature and quenched with 75 mM glycine. Nuclei were centrifuged, drained, and resuspended in wash buffer (20 mM HEPES-NaOH pH 7.5, 150 mM NaCl, 0.5 mM spermidine, Roche Complete Protease Inhibitor EDTA-free) to ~1 million cells per mL, and 250,000 nuclei per condition (as determined by

hematocytometer) were used per condition moving forward. Concavalin A beads (Cell Signaling Technologies, 93569S) were then activated with binding buffer (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$) and 10 µL were added per sample and incubated for 10 minutes at room temperature with rotation. Liquid was removed on a magnet and beads were resuspended in 50uL ice-cold antibody buffer (wash buffer with 2 mM EDTA and 0.1% BSA) and 1 µL of antibody (anti-SON Abcam #ab121759) was added and incubated upright with gentle nutating overnight at 4 °C. Beads were resuspended in secondary antibody in wash buffer (1:100 guinea pig anti-rabbit ABIN101961) and tubes were nutated at room temperature for 1 hour. Liquid was removed and beads were washed three times with wash buffer. pA-Tn5 (purified in-house according to published protocol [78]) was mixed 1:50 in Dig-300 buffer (without digitonin, 20 mM HEPES-NaOH pH 7.5, 300 mM NaCl, 0.5mM spermidine, Roche Complete Protease Inhibitor EDTA-free) and beads were resuspended in 100 uL and nutated at room temperature for 1 hour. Beads were washed three times in Dig-300 buffer. Beads were resuspended in 300 uL Tagmentation Buffer (Dig-300 buffer with 10 mM $MgCl_2$) and incubated at 37 °C for one hour. 16 mM EDTA, 0.1% SDS, and 50 µg of Proteinase K (Invitrogen, 25530015) was added per sample and incubated at 55 °C for one hour and then DNA was purified by PCI-chloroform extraction. DNA was amplified by PCR with NEBNext HiFi 2x PCR Master Mix (NEB M0541S) and library was purified by 1.3X SPRI purification (Omega Bio-Tek, M1378-01). Samples were sequenced using a NextSeq1000 (Illumina) with 50bp paired end reads at a sequencing depth of approximately 10-20 million reads per sample.

CTCF ChIP-seq data processing and peak calling

Reads were aligned to hg38 using Burrow-Wheeler Aligner (BWA-mem, v 0.7.15)[79]. Sam files were converted to bam files and sorted using Samtools view and sort (v 1.5)[80]. Identical ChIP-seq sequence reads were collapsed to avoid PCR duplicates using Picard Tools MarkDuplicates(v 2.18.5) with Java (v 1.8.0) with parameters

VALIDATION_STRINGENCY=LENIENTASSUME_SORTED=true REMOVE_DUPLICATES=true. Reads were converted to tdf files for viewing on IGV using Igvtools (v 2.3.95) count with -w 25[81]. Tag directories were made from deduplicated bam files with Homer (v. 4.9) makeTagDirectory with Perl (v 5.26.1) [82]. Peaks were called using Homer (v. 4.11) getDifferentialPeaksReplicates where -t was replicate ChIP tag directories, -i was the corresponding replicate input directories, with parameters -F 10 and -L 4. Reactivated peaks were called using Homer (v. 4.11) getDifferentialPeaksReplicates where -t was replicate ChIP tag directories for DNMT1i treatment, -b was the replicate ChIP directories for DMSO treatment, and -p was the peaks called in the corresponding DNMT1i treatment condition. These peaks were further restricted to those only called in DNMT1i and not DMSO by using bedtools (v 2.26.0) intersect with the -v and -wa options [83]. For aggregate analyses, DeepTools was used to convert deduplicated bam files to RPKM normalized bigwigs and to compare bigwigs to one another.

HiChIP data preprocessing and loop calling

HiChIP reads for each sample were aligned to hg38 and processed with HiC Pro (v 3.0.0) using the default settings with a DpnII digested genome generated with the digest_genome script [84]. For most analyses, HiCPro output .allvalidpairs files for replicate samples were then pooled. Loops were called using Hichipper (v 0.7.7) in call mode with default parameters and the flag --input-vi, and loops for each condition were kept if they had FDR< 0.01 and > 4 PETs[49]. For downstream foldchange and CPM analysis using Diffloop, Hichipper was instead used to call loops from HiCPro outputs without pooling replicates and the unfiltered loop list was used as input to Diffloop. For all Hichipper analyses, merged peak files containing all peaks from DMSO and DNMT1i for a given cell line were used as the input peak file.

Hi-C analysis

Published LIMe-Hi-C data for three replicates of DMSO and DNMT1 inhibitor treatment conditions in K562 were employed for the analysis [50]. The same treatment conditions were used as employed in this study. LIMe-Hi-C samples were generated as previously described[50] using JuiceMe (v 1.0.0) to generate merged_nodups.txt.gz pairwise contact files. Contacts from all three replicates were subsequently pooled together using the JuiceMe CPU "mega" script (v 1.0.0). These files were then converted to the cool format at 10 kb resolution using cooler (v 0.8.11). These cool files were normalized using the cooler (v 0.8.11) "balance" function and were directly utilized for aggregate pileup analysis. To perform pileup analysis, CTCF peaks, identified as described above, were filtered based upon the presence of a CTCF motif using the motif file MA0139.1.tsv.gz downloaded from the JASPAR database and assigned to the motif with the strongest signal using the bioframe (v 0.3.3) "overlap" function. In addition, CTCF sites that were located within 1000 bp of a previously identified blacklisted region (ENCFF356LFX.bed.gz) were removed using the bioframe "subtract" function. Pileup analysis was then performed using the cooltools (v 0.5.1) "pileup" function on peaks deduplicated through the bioframe "cluster" function with a "min_dist" threshold set to a resolution of 10 kb to remove duplicate peaks located within the same genomic bin[85]. Peaks for which motifs mapped to the – strand were reflected across the diagonal and the pileup matrix was then plotted in python (v 3.7) using standard matplotlib (v 3.5.2) functions. Highly-looping and reactivated peaks were defined as specified below. Speckle associated peaks were defined to be those with an average speckle SON Cut&Tag signal greater than or equal to the 90th percentile genome-wide 20 kb bin average speckle signal.

The stripenn (v 1.1.65.7)[54] "compute" function was used to identify stripes from the merged DNMT1 inhibitor 10 kb cool file with the following parameters: --norm KR -m 0.95,0.96,0.97,0.98,0.99. The distances of highly-looping peaks and normal-looping peaks to stripe anchors tabulated in the stripenn "result_filtered.tsv" output were calculated using

pybedtools (v 0.8.1) “closest” function and visualized in python (v 3.8) using standard plotting functions.

CTCF binding scatterplots and aggregate profile plots

To make scatterplots of CTCF normalized tags, RPKM normalized bigwigs were generated for ChIP samples and matched input controls from deduplicated bam files (see above) using deepTools (v 3.4.3) bamCoverage. ChIP files were normalized to their matched input controls using deepTools (v 3.4.3) bigwigCompare with the log2 option. A merged peak list was made by combining peak files from DNMT1i and DMSO and collapsing the list with bedtools (v 2.26.0) merge, and normalized ChIP-seq signal per condition was annotated onto these peaks using deepTools multiBigwigSummary. To make aggregate profile plots of CTCF ChIP-seq signal over genes or peak subsets, bigwig files of CTCF ChIP-seq signal were generated for each condition from deduplicated bam files (see above) with deepTools (v 3.4.3) bamCoverage with flag --normalizeUsing RPKM. To make aggregate plots over peak subsets, deepTools (v 3.4.3) computeMatrix and plotProfile were used with the following computeMatrix parameters: reference-point --referencePoint center -bs 50 -a 1000 -b 1000 --skipZeros. To make aggregate plots over genes, deepTools (v 3.4.3) computeMatrix and plotprofile were used with the following parameters: --scale-regions --regionBodyLength 1000-bs 50 -a 1000 -b 1000 --skipZeros. Downstream processing was completed in python (v 3.7.0) (www.python.org) with pandas (v 1.0.3), matplotlib (v 3.2.1) and seaborn (v 0.11.2).

Annotating CTCF peaks

CTCF peaks were annotated using Homer (v 4.11) annotatePeaks using the -gtf option with an Ensembl gtf file (v 104) with the standard priority order assigned [82]. Motifs were assigned to CTCF peaks using gimmemotifs gimme scan function with -b and -f 0.1 and the JASPAR motif pfm as input. Strand-aware motif methylation heatmaps were made using Deeptools. To assess overlap

of categories of CTCF peaks with a bed file of predefined subcompartments from SNIPER [56] bedtools (v 2.26.0) map was used with the – o distinct flag, and data were further processed in python (v 3.7.0) (www.python.org) and any peaks overlapping multiple subcompartments were discarded. To assess overlap of categories of CTCF peaks with normalized TSA-seq counts, UCSC's function bigWigToBedGraph was first used to convert a bigwig file of normalized SON TSA-seq counts to a bedgraph in 20kb bins, and then bedtools (v 2.26.0) map was used with – o mean flag to map TSA-seq counts onto CTCF peak list and the data were further processed in python (v 3.7.0). The following published files were used from 4D Nucleome for TSA-seq figures: 4DNFIVZSO9RI.bw 4DNFIBY8G6RZ.bw.

Assigning number of loops (partners) per CTCF peak

CTCF peaks were annotated with the number of different loops (partners) per peak by using pgltools (v 2.2.0) intersect1D to list each loop-peak interception followed by bedtools (v 2.26.0) merge on the list of peaks to sum the number of different loops in which each peak falls within an anchor [83,86]. To annotate peaks with the number of different loops their maximally-looping partner makes, pgltools (v 2.2.0) intersect1D was again used to list each loop-peak interception. Bedtools (v 2.26.0) map was used to annotate the number of different loops the peak in each partner anchor makes from the previous file using the -o max flag. This resulted in one entry per loop that lists the "original" peak and the number of loops its partner anchor makes. The maximally-looping partner per original peak was mapped back onto the original peak list with bedtools (v 2.26.0) map using the -o max flag so that each peak was listed once.

Calling and characterizing highly-looping peaks

Highly-looping peaks were called by adapting code used previously to define methylation rich regions (MRRs) and superenhancers [87]. Specifically, peaks were ordered by rank with respect to the number of different loops they make. The tangent line was drawn where the slope becomes

1 and any peaks above this cutoff were considered highly-looping for a given condition. To generate aggregate profile plots of public K562 histone modification bigwig files (see source data) over CTCF peaks, deepTools (v 3.4.3) computeMatrix was used with the following parameters: reference-point --referencePoint center -bs 50 -a 5000 -b 5000 --skipZeros and results were visualized with deepTools (v 3.4.3) plotProfile. Overlap of peaks with subcompartments and TSA-seq was performed as described above for all CTCF peaks. The following published files were used from the ENCODE consortium for the peak characterization pileups: ENCFF239EBH.bigWig, ENCFF469JMR.bigWig, ENCFF600FDO.bigWig.

Calculating looping changes across conditions

To calculate foldchange and mean CPM loop strength between conditions, mango Hichipper output files for individual replicates were input into Diffloop (v 1.16.0) and processed with the loopsmake mango function [48]. The full loop list was then filtered for loops with FDR <0.01 and with at least 8 PETs across samples for a given cell type and loops that appeared strongly (>5 PETs) in only one replicate but 0 PETS in the other were removed. The quickAssoc function was then used to get loop strength foldchange and CPM values for each 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) [86].

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 [86]. 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'.[88] 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

RNA-seq data were analyzed following Lexogen's recommended protocol. First, UMIs were extracted with umi-tools (v 1.1.1) extract [89]. Adapters were then trimmed with cutadapt (v 2.7) [90]. Trimmed reads were aligned to hg38 using Star aligner (v 2.6.0) with a genome generated using the Ensembl hg38 primary assembly and Ensembl GTF file v 104 with the following settings: --outFilterMultimapNmax 20 --alignSJoverhangMin 8 --alignSJDBoverhangMin 1 --outFilterMismatchNmax 999 --outFilterMismatchNoverLmax 0.6 --alignIntronMin 20 --alignIntronMax 1000000 --alignMatesGapMax 1000000 -- peOverlapNbasesMin 40 --peOverlapMMp 0.8 --outSAMattributes NH HI NM MD[91]. Reads were then deduplicated with umi-tools dedup (v 1.1.1) with flags multimapping-detection-method=NH and unpaired-reads=discard.

For gene expression analysis, HTseq – count (v 0.11.2) was used to count reads in genes [92]. Count files were then processed in DEseq2 (v 1.28) with R (v 4.0.2) using the standard workflow to get foldchange values and using alpha=0.05 [93,94]. Hierarchical clustering was performed on the regularized $\log_2$-transformed counts using Euclidean distance as a metric and the complete clustering method. All genes with both an adjusted $P < 0.05$, calculated using the Benjamini-Hochberg correction, and $|\log_2(\text{fold-change})| > 0.5$ between wildtype DNMT1i and DMSO were included in the clustered heatmap. The regularized $\log_2$-transformed counts for the variable genes were grouped by *k*-means clustering ($k = 4$) and the Z-scores for the variable genes, ordered according to their cluster identity, are depicted. For differential exon analysis, Dexseq (v 1.34.1) was used. Files were then imported into R and foldchange values were computed according to the standard Dexseq workflow. BioMart was used to map the genomic ranges for each gene onto RNA-seq differential expression output files. Gene expression changes were compared with Cut&Tag signal and CTCF peak overlap using bedtools (v 2.26.0) map with -o mean. Gene Ontology analysis was completed on the online DAVID portal by filtering for goterm_direct and goterm_BP_5 results with FDRj<0.2 [95,96]. Further analysis, including volcano plots, was completed in python (v 3.7.0) (www.python.org) with pandas (v 1.0.3), matplotlib (v 3.2.1) seaborn (v 0.11.2) and statannot (v 0.2.3).

Data Access

All sequencing data has been deposited on NCBI Gene Expression Omnibus under the super series accession code GSE217760. Custom code is available at https://github.com/liaulab.

Competing Interests

B.B.L. is a shareholder and member of the scientific advisory board of Light Horse Therapeutics. The remaining authors declare no competing interests.

Acknowledgements

We acknowledge members of the Liau lab for useful discussions. We acknowledge J. Nelson, C. Maesner and Z. Niziolek for assistance with cell sorting. We acknowledge the FAS Division of Science Research Computing Group at Harvard University for assistance with cluster computing. We acknowledge S. Johnstone for support with the HiChIP protocol. We acknowledge T. Lu for helpful discussions about speckle Cut&Tag. We acknowledge H.Roh for providing Protein A Tn5. A.P.S and C.L. were supported by the Herchel Smith Graduate fellowships. N.Z.L. was supported by a NSF Graduate Research Fellowship (grant no. DGE1745303). A.L.W was supported by a Simmons award from Harvard Center for Biological Imaging. This research was supported by startup funds from Harvard University.

References

1. Rowley, M.J., and Corces, V.G. (2018). Organizational principles of 3D genome architecture. Nat Rev Genet *19*, 1. https://doi.org/10.1038/s41576-018-0060-8.

2. Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021.

3. Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380. https://doi.org/10.1038/nature11082.

4. Misteli, T. (2020). The Self-Organizing Genome: Principles of Genome Architecture and Function. Cell *183*, 28–45. https://doi.org/10.1016/j.cell.2020.09.014.

5. Zheng, H., and Xie, W. (2019). The role of 3D genome organization in development and cell differentiation. Nat Rev Mol Cell Bio *20*, 535–550. https://doi.org/10.1038/s41580-019-0132-4.

6. Rao, S.S.P., Huang, S.-C., Hilaire, B.G.S., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305-320.e24. https://doi.org/10.1016/j.cell.2017.09.026.

7. Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944.e22. https://doi.org/10.1016/j.cell.2017.05.004.

8. Steensel, B. van, and Belmont, A.S. (2017). Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. Cell *169*, 780–791. https://doi.org/10.1016/j.cell.2017.04.022.

9. Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796–801. https://doi.org/10.1038/nature06634.

10. Li, Y., Haarhuis, J.H.I., Cacciatore, Á.S., Oldenkamp, R., Ruiten, M.S. van, Willems, L., Teunissen, H., Muir, K.W., Wit, E. de, Rowland, B.D., et al. (2020). The structural basis for cohesin–CTCF-anchored loops. Nature, 1–9. https://doi.org/10.1038/s41586-019-1910-z.

11. Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., and Mirny, L.A. (2016). Formation of Chromosomal Domains by Loop Extrusion. Cell Reports *15*, 2038–2049. https://doi.org/10.1016/j.celrep.2016.04.085.

12. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I.J., and Yu, H. (2019). Human cohesin compacts DNA by loop extrusion. Science *366*, 1345–1349. https://doi.org/10.1126/science.aaz4475.

13. Davidson, I.F., Bauer, B., Goetz, D., Tang, W., Wutz, G., and Peters, J.-M. (2019). DNA loop extrusion by human cohesin. Science, eaaz3418. https://doi.org/10.1126/science.aaz3418.

14. Xiang, J.-F., and Corces, V.G. (2021). Regulation of 3D chromatin organization by CTCF. Curr Opin Genet Dev *67*, 33–40. https://doi.org/10.1016/j.gde.2020.10.005.

15. Hansen, A.S., Cattoglio, C., Darzacq, X., and Tjian, R. (2017). Recent evidence that TADs and chromatin loops are dynamic structures. Nucleus *9*, 20–32. https://doi.org/10.1080/19491034.2017.1389365.

16. Sanborn, A.L., Rao, S.S.P., Huang, S.-C., Durand, N.C., Huntley, M.H., Jewett, A.I., Bochkov, I.D., Chinnappan, D., Cutkosky, A., Li, J., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc National Acad Sci *112*, E6456–E6465. https://doi.org/10.1073/pnas.1518552112.

17. Dubois, F., Sidiropoulos, N., Weischenfeldt, J., and Beroukhim, R. (2022). Structural variations in cancer and the 3D genome. Nat Rev Cancer, 1–14. https://doi.org/10.1038/s41568-022-00488-9.

18. Spielmann, M., Lupiáñez, D.G., and Mundlos, S. (2018). Structural variation in the 3D genome. Nat Rev Genet *19*, 453. https://doi.org/10.1038/s41576-018-0007-0.

19. Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. Cell *161*, 1012–1025. https://doi.org/10.1016/j.cell.2015.04.004.

20. Flavahan, W.A., Drier, Y., Liau, B.B., Gillespie, S.M., Venteicher, A.S., Stemmer-Rachamimov, A.O., Suvà, M.L., and Bernstein, B.E. (2016). Insulator dysfunction and oncogene activation in IDH mutant gliomas. Nature *529*, 110. https://doi.org/10.1038/nature16490.

21. Flavahan, W.A., Drier, Y., Johnstone, S.E., Hemming, M.L., Tarjan, D.R., Hegazi, E., Shareef, S.J., Javed, N.M., Raut, C.P., Eschle, B.K., et al. (2019). Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. Nature *575*, 229–233. https://doi.org/10.1038/s41586-019-1668-3.

22. Rinzema, N.J., Sofiadis, K., Tjalsma, S.J.D., Verstegen, M.J.A.M., Oz, Y., Valdes-Quezada, C., Felder, A.-K., Filipovska, T., Elst, S. van der, Ramos, Z. de A. dos, et al. (2022). Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. Nat Struct Mol Biol, 1–12. https://doi.org/10.1038/s41594-022-00787-7.

23. Kubo, N., Ishii, H., Xiong, X., Bianco, S., Meitinger, F., Hu, R., Hocker, J.D., Conte, M., Gorkin, D., Yu, M., et al. (2021). Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. Nat Struct Mol Biol, 1–10. https://doi.org/10.1038/s41594-020-00539-5.

24. Shukla, S., Kavak, E., Gregory, M., Imashimizu, M., Shutinoski, B., Kashlev, M., Oberdoerffer, P., Sandberg, R., and Oberdoerffer, S. (2011). CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature *479*, 74–79. https://doi.org/10.1038/nature10442.

25. Ruiz-Velasco, M., Kumar, M., Lai, M.C., Bhat, P., Solis-Pinson, A.B., Reyes, A., Kleinsorg, S., Noh, K.-M., Gibson, T.J., and Zaugg, J.B. (2017). CTCF-Mediated Chromatin Loops between Promoter and Gene Body Regulate Alternative Splicing across Individuals. Cell Syst *5*, 628-637.e6. https://doi.org/10.1016/j.cels.2017.10.018.

26. Nanavaty, V., Abrash, E.W., Hong, C., Park, S., Fink, E.E., Li, Z., Sweet, T.J., Bhasin, J.M., Singuri, S., Lee, B.H., et al. (2020). DNA Methylation Regulates Alternative Polyadenylation via CTCF and the Cohesin Complex. Mol Cell. https://doi.org/10.1016/j.molcel.2020.03.024.

27. Alharbi, A.B., Schmitz, U., Bailey, C.G., and Rasko, J.E.J. (2021). CTCF as a regulator of alternative splicing: new tricks for an old player. Nucleic Acids Res *49*, gkab520-. https://doi.org/10.1093/nar/gkab520.

28. Soto, E.J.L., and Lipscombe, D. (2020). Cell-specific exon methylation and CTCF binding in neurons regulates calcium ion channel splicing and function. Elife *9*, e54879. https://doi.org/10.7554/elife.54879.

29. Steensel, B. van, and Furlong, E.E.M. (2019). The role of transcription in shaping the spatial organization of the genome. Nat Rev Mol Cell Bio *20*, 327–337. https://doi.org/10.1038/s41580-019-0114-6.

30. Schoenfelder, S., and Fraser, P. (2019). Long-range enhancer–promoter contacts in gene expression control. Nat Rev Genet *20*, 437–455. https://doi.org/10.1038/s41576-019-0128-0.

31. Vermunt, M.W., Zhang, D., and Blobel, G.A. (2019). The interdependence of gene-regulatory elements and the 3D genome. J Cell Biol *218*, 12–26. https://doi.org/10.1083/jcb.201809040.

32. Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482–485. https://doi.org/10.1038/35013100.

33. Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature *405*, 486–489. https://doi.org/10.1038/35013106.

34. Hashimoto, H., Wang, D., Horton, J.R., Zhang, X., Corces, V.G., and Cheng, X. (2017). Structural Basis for the Versatile and Methylation-Dependent Binding of CTCF to DNA. Mol Cell *66*, 711-720.e3. https://doi.org/10.1016/j.molcel.2017.05.004.

35. Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and Stamatoyannopoulos, J.A. (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Reports *12*, 1184–1195. https://doi.org/10.1016/j.celrep.2015.07.024.

36. Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res *22*, 1680–1688. https://doi.org/10.1101/gr.136101.111.

37. Schuijers, J., Manteiga, J.C., Weintraub, A.S., Day, D.S., Zamudio, A.V., Hnisz, D., Lee, T.I., and Young, R.A. (2018). Transcriptional Dysregulation of MYC Reveals Common Enhancer-Docking Mechanism. Cell Reports *23*, 349–360. https://doi.org/10.1016/j.celrep.2018.03.056.

38. Chen, Y., Zhang, Y., Wang, Y., Zhang, L., Brinkman, E.K., Adam, S.A., Goldman, R., Steensel, B. van, Ma, J., and Belmont, A.S. (2018). Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. J Cell Biol *217*, jcb.201807108. https://doi.org/10.1083/jcb.201807108.

39. Chen, Y., and Belmont, A.S. (2019). Genome organization around nuclear speckles. Curr Opin Genet Dev *55*, 91–99. https://doi.org/10.1016/j.gde.2019.06.008.

40. Zhang, L., Zhang, Y., Chen, Y., Gholamalamdari, O., Wang, Y., Ma, J., and Belmont, A.S. (2020). TSA-seq reveals a largely conserved genome organization relative to nuclear speckles with small position changes tightly correlated with gene expression changes. Genome Res *31*, gr.266239.120. https://doi.org/10.1101/gr.266239.120.

41. Wang, Y., Zhang, Y., Zhang, R., Schaik, T. van, Zhang, L., Sasaki, T., Peric-Hupkes, D., Chen, Y., Gilbert, D.M., Steensel, B. van, et al. (2021). SPIN reveals genome-wide landscape of nuclear compartmentalization. Genome Biol *22*, 36. https://doi.org/10.1186/s13059-020-02253-3.

42. Spector, D.L., and Lamond, A.I. (2011). Nuclear Speckles. Csh Perspect Biol *3*, a000646. https://doi.org/10.1101/cshperspect.a000646.

43. Alexander, K.A., Coté, A., Nguyen, S.C., Zhang, L., Gholamalamdari, O., Agudelo-Garcia, P., Lin-Shiao, E., Tanim, K.M.A., Lim, J., Biddle, N., et al. (2021). p53 mediates target gene association with nuclear speckles for amplified RNA expression. Mol Cell *81*, 1666-1681.e6. https://doi.org/10.1016/j.molcel.2021.03.006.

44. Ilık, İ.A., and Aktaş, T. (2021). Nuclear speckles: dynamic hubs of gene expression regulation. Febs J. https://doi.org/10.1111/febs.16117.

45. Portilho, N.A., Saini, D., Hossain, I., Sirois, J., Moraes, C., and Pastor, W.A. (2021). The DNMT1 inhibitor GSK-3484862 mediates global demethylation in murine embryonic stem cells. Epigenet Chromatin *14*, 56. https://doi.org/10.1186/s13072-021-00429-0.

46. Pappalardi, M.B., Keenan, K., Cockerill, M., Kellner, W.A., Stowell, A., Sherk, C., Wong, K., Pathuri, S., Briand, J., Steidel, M., et al. (2021). Discovery of a first-in-class reversible DNMT1-selective inhibitor with improved tolerability and efficacy in acute myeloid leukemia. Nat Cancer, 1–16. https://doi.org/10.1038/s43018-021-00249-x.

47. Spracklin, G., Abdennur, N., Imakaev, M., Chowdhury, N., Pradhan, S., Mirny, L.A., and Dekker, J. (2022). Diverse silent chromatin states modulate genome compartmentalization and loop extrusion barriers. Nat Struct Mol Biol, 1–14. https://doi.org/10.1038/s41594-022-00892-7.

48. Lareau, C.A., and Aryee, M.J. (2018). diffloop: a computational framework for identifying and analyzing differential DNA loops from sequencing data. Bioinformatics *34*, 672–674. https://doi.org/10.1093/bioinformatics/btx623.

49. Lareau, C.A., and Aryee, M.J. (2018). hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. Nat Methods *15*, 155–156. https://doi.org/10.1038/nmeth.4583.

50. Siegenfeld, A.P., Roseman, S.A., Roh, H., Lue, N.Z., Wagen, C.C., Zhou, E., Johnstone, S.E., Aryee, M.J., and Liau, B.B. (2022). Polycomb-lamina antagonism partitions heterochromatin at the nuclear periphery. Nat Commun *13*, 4199. https://doi.org/10.1038/s41467-022-31857-5.

51. Cai, Y., Zhang, Y., Loh, Y.P., Tng, J.Q., Lim, M.C., Cao, Z., Raju, A., Aiden, E.L., Li, S., Manikandan, L., et al. (2021). H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. Nat Commun *12*, 719. https://doi.org/10.1038/s41467-021-20940-y.

52. Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-Enhancers in the Control of Cell Identity and Disease. Cell *155*, 934–947. https://doi.org/10.1016/j.cell.2013.09.053.

53. Vian, L., Pękowska, A., Rao, S.S.P., Kieffer-Kwon, K.-R., Jung, S., Baranello, L., Huang, S.-C., Khattabi, L.E., Dose, M., Pruett, N., et al. (2018). The Energetics and Physiological Impact of Cohesin Extrusion. Cell *173*, 1165-1178.e20. https://doi.org/10.1016/j.cell.2018.03.072.

54. Yoon, S., Chandra, A., and Vahedi, G. (2022). Stripenn detects architectural stripes from chromatin conformation data using computer vision. Nat Commun *13*, 1602. https://doi.org/10.1038/s41467-022-29258-9.

55. Barrington, C., Georgopoulou, D., Pezic, D., Varsally, W., Herrero, J., and Hadjur, S. (2019). Enhancer accessibility and CTCF occupancy underlie asymmetric TAD architecture and cell type specific genome topology. Nat Commun *10*, 2908. https://doi.org/10.1038/s41467-019-10725-9.

56. Xiong, K., and Ma, J. (2019). Revealing Hi-C subcompartments by imputing inter-chromosomal chromatin interactions. Nat Commun *10*, 5069. https://doi.org/10.1038/s41467-019-12954-4.

57. Brown, J.M., Green, J., Neves, R.P. das, Wallace, H.A.C., Smith, A.J.H., Hughes, J., Gray, N., Taylor, S., Wood, W.G., Higgs, D.R., et al. (2008). Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. J Cell Biology *182*, 1083–1097. https://doi.org/10.1083/jcb.200803174.

58. Ilik, İ.A., Malszycki, M., Lübke, A.K., Schade, C., Meierhofer, D., and Aktaş, T. (2020). SON and SRRM2 are essential for nuclear speckle formation. Elife *9*, e60579. https://doi.org/10.7554/elife.60579.

59. Faber, G.P., Nadav-Eliyahu, S., and Shav-Tal, Y. (2022). Nuclear speckles – a driving force in gene expression. J Cell Sci *135*. https://doi.org/10.1242/jcs.259594.

60. Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., et al. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell *174*, 744-757.e24. https://doi.org/10.1016/j.cell.2018.05.024.

61. Kim, J., Venkata, N.C., Gonzalez, G.A.H., Khanna, N., and Belmont, A.S. (2020). Gene expression amplification by nuclear speckle association. J Cell Biology *219*, jcb.201904046. https://doi.org/10.1083/jcb.201904046.

62. Jabbari, K., Chakraborty, M., and Wiehe, T. (2019). DNA sequence-dependent chromatin architecture and nuclear hubs formation. Sci Rep-uk *9*, 14646. https://doi.org/10.1038/s41598-019-51036-9.

63. Lehman, B.J., Lopez-Diaz, F.J., Santisakultarm, T.P., Fang, L., Shokhirev, M.N., Diffenderfer, K.E., Manor, U., and Emerson, B.M. (2021). Dynamic regulation of CTCF stability and sub-nuclear localization in response to stress. Plos Genet *17*, e1009277. https://doi.org/10.1371/journal.pgen.1009277.

64. Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G., Ahmad, K., and Henikoff, S. (2019). CUT&Tag for efficient epigenomic profiling of small samples and single cells. Nat Commun *10*, 1930. https://doi.org/10.1038/s41467-019-09982-5.

65. Ahn, E.-Y., DeKelver, R.C., Lo, M.-C., Nguyen, T.A., Matsuura, S., Boyapati, A., Pandit, S., Fu, X.-D., and Zhang, D.-E. (2011). SON Controls Cell-Cycle Progression by Coordinated Regulation of RNA Splicing. Mol Cell *42*, 185–198. https://doi.org/10.1016/j.molcel.2011.03.014.

66. Nabet, B., Roberts, J.M., Buckley, D.L., Paulk, J., Dastjerdi, S., Yang, A., Leggett, A.L., Erb, M.A., Lawlor, M.A., Souza, A., et al. (2018). The dTAG system for immediate and target-specific protein degradation. Nat Chem Biol *14*, 431–441. https://doi.org/10.1038/s41589-018-0021-8.

67. Kraft, K., Magg, A., Heinrich, V., Riemenschneider, C., Schöpflin, R., Markowski, J., Ibrahim, D.M., Acuna-Hidalgo, R., Despang, A., Andrey, G., et al. (2019). Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and congenital malformations. Nat Cell Biol *21*, 305–310. https://doi.org/10.1038/s41556-019-0273-x.

68. Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science *356*, eaaj2239. https://doi.org/10.1126/science.aaj2239.

69. Yu, R., Roseman, S., Siegnefeld, A., Tran, K., Nguyen, S.C., Joyce, E.F., Raj, A., Liau, B., Krantz, I., Alexander, K., et al. (2023). CTCF/cohesin institute the ground state of chromatin-nuclear speckle association. BioRxiv.

70. Liao, S.E., and Regev, O. (2020). Splicing at the phase-separated nuclear speckle interface: a model. Nucleic Acids Res *49*, gkaa1209-. https://doi.org/10.1093/nar/gkaa1209.

71. Hu, S., Lv, P., Yan, Z., and Wen, B. (2019). Disruption of nuclear speckles reduces chromatin interactions in active compartments. Epigenet Chromatin *12*, 43. https://doi.org/10.1186/s13072-019-0289-2.

72. Lee, R., Kang, M.-K., Kim, Y.-J., Yang, B., Shim, H., Kim, S., Kim, K., Yang, C.M., Min, B., Jung, W.-J., et al. (2021). CTCF-mediated chromatin looping provides a topological

framework for the formation of phase-separated transcriptional condensates. Nucleic Acids Res *50*, 207–226. https://doi.org/10.1093/nar/gkab1242.

73. Ulianov, S.V., Velichko, A.K., Magnitov, M.D., Luzhin, A.V., Golov, A.K., Ovsyannikova, N., Kireev, I.I., Gavrikov, A.S., Mishin, A.S., Garaev, A.K., et al. (2021). Suppression of liquid–liquid phase separation by 1,6-hexanediol partially compromises the 3D genome organization in living cells. Nucleic Acids Res *49*, gkab249-. https://doi.org/10.1093/nar/gkab249.

74. Jiang, Y., Huang, J., Lun, K., Li, B., Zheng, H., Li, Y., Zhou, R., Duan, W., Wang, C., Feng, Y., et al. (2020). Genome-wide analyses of chromatin interactions after the loss of Pol I, Pol II, and Pol III. Genome Biol *21*, 158. https://doi.org/10.1186/s13059-020-02067-3.

75. Luan, J., Xiang, G., Gómez-García, P.A., Tome, J.M., Zhang, Z., Vermunt, M.W., Zhang, H., Huang, A., Keller, C.A., Giardine, B.M., et al. (2021). Distinct properties and functions of CTCF revealed by a rapidly inducible degron system. Cell Reports *34*, 108783. https://doi.org/10.1016/j.celrep.2021.108783.

76. Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol *7*, R100. https://doi.org/10.1186/gb-2006-7-10-r100.

77. Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat Methods *13*, 919–922. https://doi.org/10.1038/nmeth.3999.

78. Priore, I.D., Ma, S., Strecker, J., Jacks, T., LaFave, L.M., and Buenrostro, J.D. (2021). Protocol for single-cell ATAC sequencing using combinatorial indexing in mouse lung adenocarcinoma. Star Protoc *2*, 100583. https://doi.org/10.1016/j.xpro.2021.100583.

79. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Arxiv. https://doi.org/10.48550/arxiv.1303.3997.

80. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Subgroup, 1000 Genome Project Data Processing (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

81. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24–26. https://doi.org/10.1038/nbt.1754.

82. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

83. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

84. Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol *16*, 259. https://doi.org/10.1186/s13059-015-0831-x.

85. Open2C, Abdennur, N., Abraham, S., Fudenberg, G., Flyamer, I.M., Galitsyna, A.A., Goloborodko, A., Imakaev, M., Oksuz, B.A., and Venev, S.V. (2022). Cooltools: enabling high-resolution Hi-C analysis in Python. bioRxiv, 2022.10.31.514564. https://doi.org/10.1101/2022.10.31.514564.

86. Greenwald, W.W., Li, H., Smith, E.N., Benaglio, P., Nariai, N., and Frazer, K.A. (2017). Pgltools: a genomic arithmetic tool suite for manipulation of Hi-C peak and other chromatin interaction data. Bmc Bioinformatics *18*, 207. https://doi.org/10.1186/s12859-017-1621-0.

87. Cai, Y., Zhang, Y., Loh, Y.P., Tng, J.Q., Lim, M.C., Cao, Z., Raju, A., Li, S., Manikandan, L., Tergaonkar, V., et al. (2020). H3K27me3-rich genomic regions can function as silencers to repress gene expression via chromatin interactions. Biorxiv, 684712. https://doi.org/10.1101/684712.

88. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160–W165. https://doi.org/10.1093/nar/gkw257.

89. Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res *27*, 491–499. https://doi.org/10.1101/gr.209601.116.

90. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads | Martin | EMBnet.journal. EMBnet.journal.

91. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

92. Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169. https://doi.org/10.1093/bioinformatics/btu638.

93. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

94. Zitovsky, J.P., and Love, M.I. (2019). Fast effect size shrinkage software for beta-binomial models of allelic imbalance. F1000research *8*, 2024. https://doi.org/10.12688/f1000research.20916.2.

95. Sherman, B.T., Hao, M., Qiu, J., Jiao, X., Baseler, M.W., Lane, H.C., Imamichi, T., and Chang, W. (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res *50*, W216–W221. https://doi.org/10.1093/nar/gkac194.

96. Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc *4*, 44–57. https://doi.org/10.1038/nprot.2008.211.

Figure S1

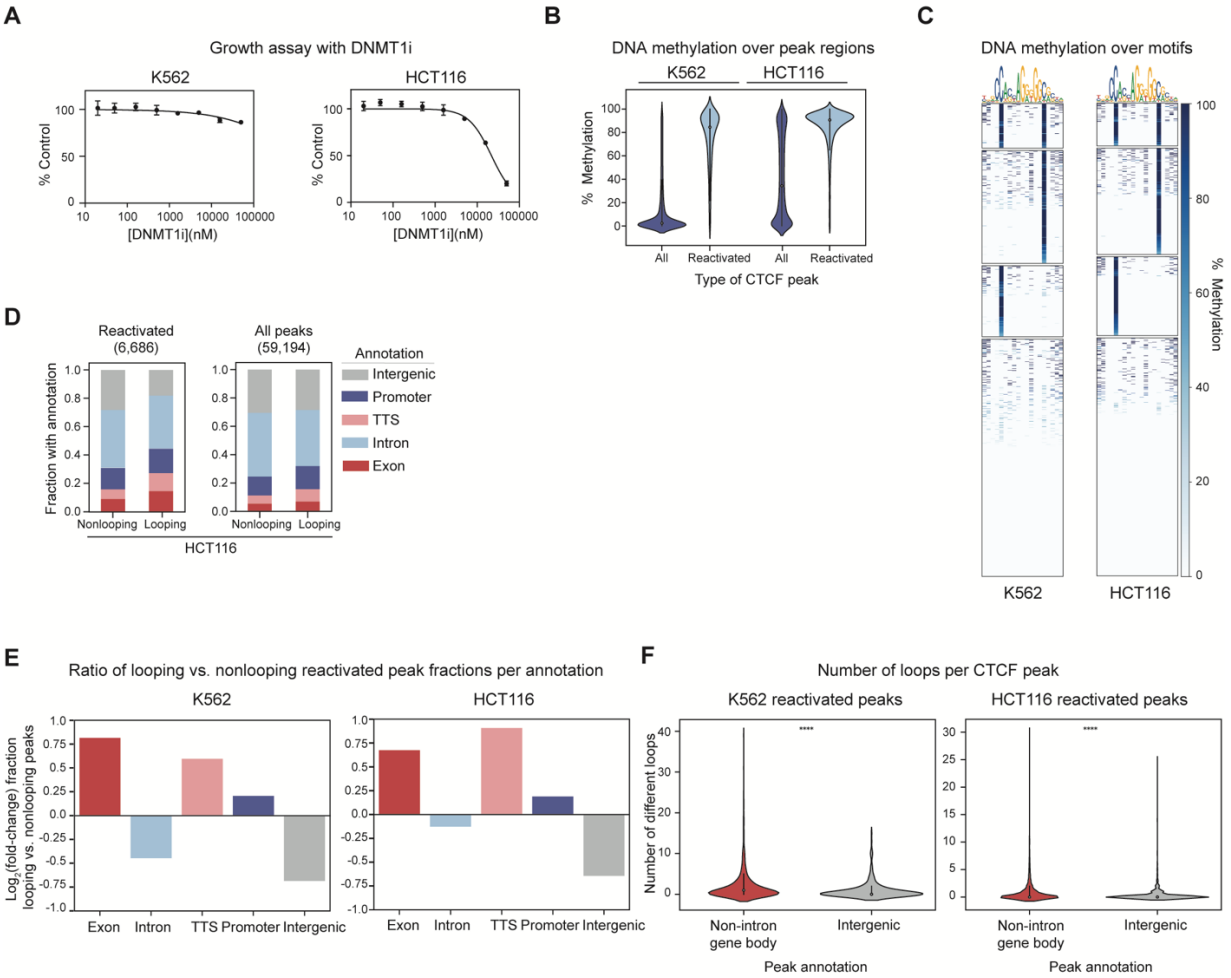

Figure S1 | DNMT1 inhibition reactivates CTCF peaks and loops on gene bodies

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 across CpG sites within 300 bp CTCF peak regions for reactivated peaks and all other CTCF peaks.
C. 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
D. Proportion of looping and nonlooping peaks with each annotation (y-axis) for both reactivated and all CTCF peaks in HCT116 in DNMT1i.
E. log2(fold-change) ratio of the fractions of each annotation within looping vs. nonlooping reactivated peaks. Ratio of bars from **Fig 1G, S1D**.
F. Number of different loops each CTCF peak makes (number of interaction partners) (y-axis) for reactivated K562 (left) and reactivated HCT116 (right) peaks in DNMT1i by peak annotation (x-axis).

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

Figure S2

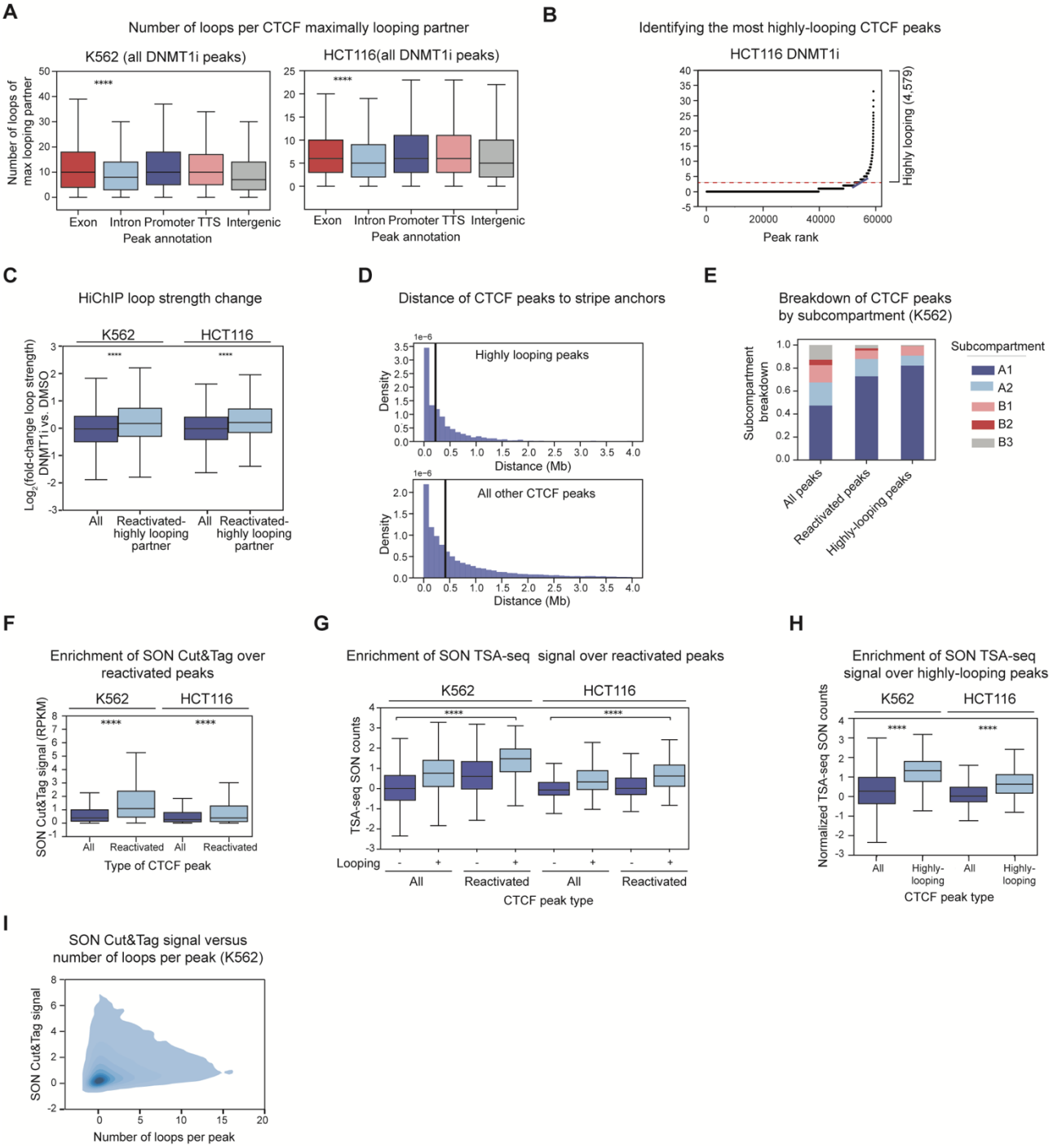
A
Number of loops per CTCF maximally looping partner
K562 (all DNMT1i peaks)
HCT116(all DNMT1i peaks)
Number of loops of max looping partner
Exon
Intron
Promoter
TTS
Intergenic
Peak annotation
B
Identifying the most highly-looping CTCF peaks
HCT116 DNMT1i
Highly looping (4,579)
Peak rank
C
HiChIP loop strength change
K562
HCT116
Log2(fold-change loop strength) DNMT1i vs. DMSO
All
Reactivated-highly looping partner
D
Distance of CTCF peaks to stripe anchors
Highly looping peaks
All other CTCF peaks
Density
Distance (Mb)
E
Breakdown of CTCF peaks by subcompartment (K562)
Subcompartment breakdown
Subcompartment
A1
A2
B1
B2
B3
All peaks
Reactivated peaks
Highly-looping peaks
F
Enrichment of SON Cut&Tag over reactivated peaks
SON Cut&Tag signal (RPKM)
Reactivated
Type of CTCF peak
G
Enrichment of SON TSA-seq signal over reactivated peaks
TSA-seq SON counts
Looping
CTCF peak type
H
Enrichment of SON TSA-seq signal over highly-looping peaks
Normalized TSA-seq SON counts
Highly-looping
I
SON Cut&Tag signal versus number of loops per peak (K562)
SON Cut&Tag signal
Number of loops per peak

Figure S2 | Reactivated CTCF peaks interact with highly-looping partners near nuclear speckles

A. 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).
B. Number of different loops (partner peaks) (*y*-axis) assigned to each peak by rank (*x*-axis) in DNMT1i for HCT116. Highly-looping peaks reside above the red dashed line.
C. 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.
D. Histogram showing density (y-axis) of genomic distances (x-axis) of CTCF highly-looping peaks (top) and all other CTCF peaks (bottom) 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.
E. Proportion of all, reactivated, and highly-looping CTCF K562 DNMT1i peaks in each published subcompartment (y-axis) assigned by SNIPER for wildtype K562 cells.
F. 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. Same as 2G, but not broken into looping categories.
G. 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.
H. 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 (4DNFIVZSO9RI.bw, 4DNFIBY8G6RZ.bw) in the untreated, wildtype respective cell type (see methods).
I. 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.

For **A**, **C**, **F**, **G**, and **H**, 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

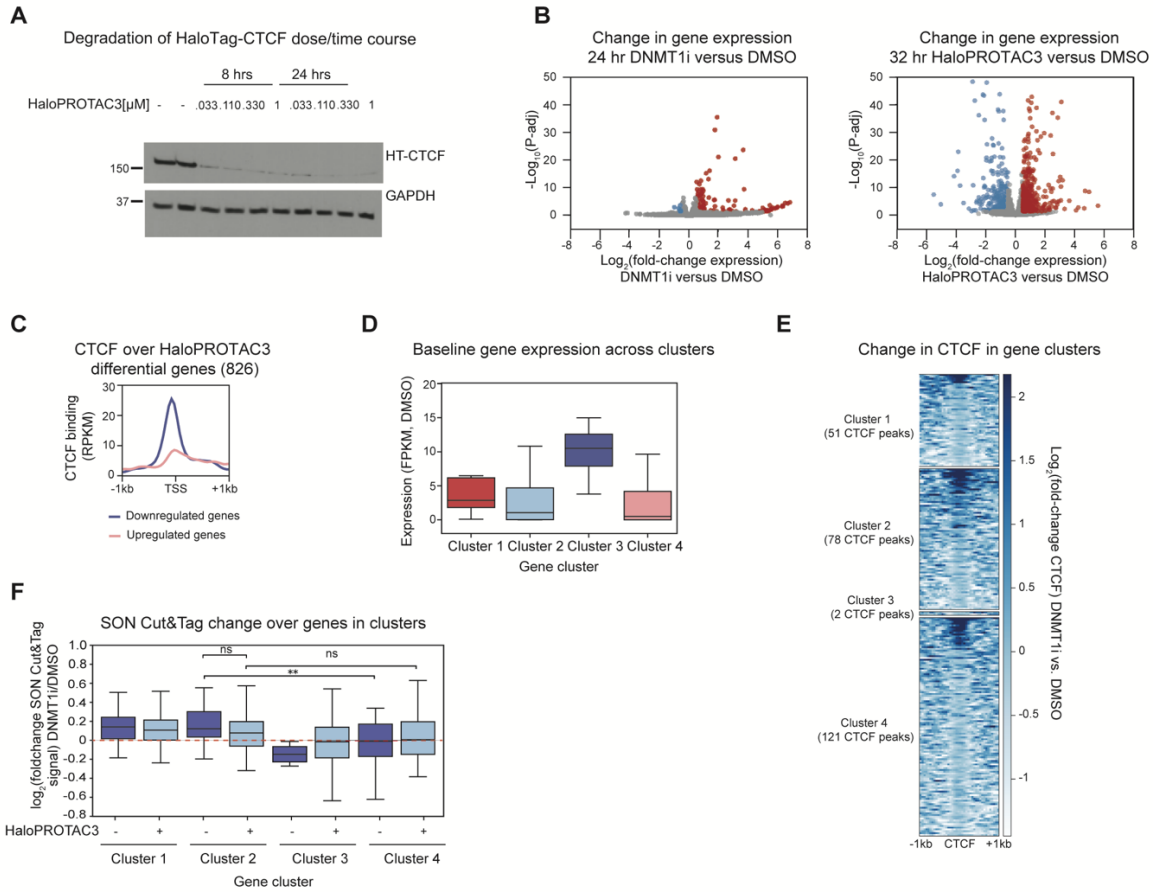

Figure S3 | DNMT1i-induced gene activation and speckle association depend on CTCF

A. 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.
B. Volcano plot showing -log10(adjusted P-value) (y-axis) vs. log2(fold-change) in gene expression (x-axis) upon 1 day DNMT1i treatment (left) or 32 hours HaloPROTAC3 treatment (right)in HaloTag-CTCF HCT116 cells. Upregulated genes are highlighted in red, and downregulated genes are highlighted in blue (p-adj < 0.05, |log2fold-change| > 0.5). <5 points per plot are omitted for visualization.
C. Aggregate profile plot of replicate-averaged RPKM normalized CTCF binding in DMSO over the transcription start sites of differential HaloPROTAC3 genes.
D. Boxplot showing expression levels of genes within clusters from **Fig 3D** in Fragments Per Kilobase per Million mapped fragments (FPKM). Expression in cells treated with vehicle (-HaloPROTAC3, - DNMT1i) is shown.
E. Heatmap of per-CTCF peak log2(fold-change CTCF binding) after 1 day of DNMT1i treatment vs. DMSO control centered over all CTCF peaks overlapping genes in the clusters from **Fig 3D**. Each gene could have multiple CTCF peaks on it.
F. Boxplot showing log2(foldchange SON Cut&Tag signal DNMT1i vs. DMSO, y axis) over gene bodies for genes in the different clusters from **Fig 3D**. Foldchange Cut&Tag signal with and without HaloPROTAC3 treatment are shown for each gene cluster. The same genes are shown for the + and – HaloPROTAC3 conditions for comparison.

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 ≤ 0.05; **: 0.001 <p ≤ 0.01; ***: 0.0001 <p ≤ 0.001; ****: p ≤ 0.0001.

Figure S4

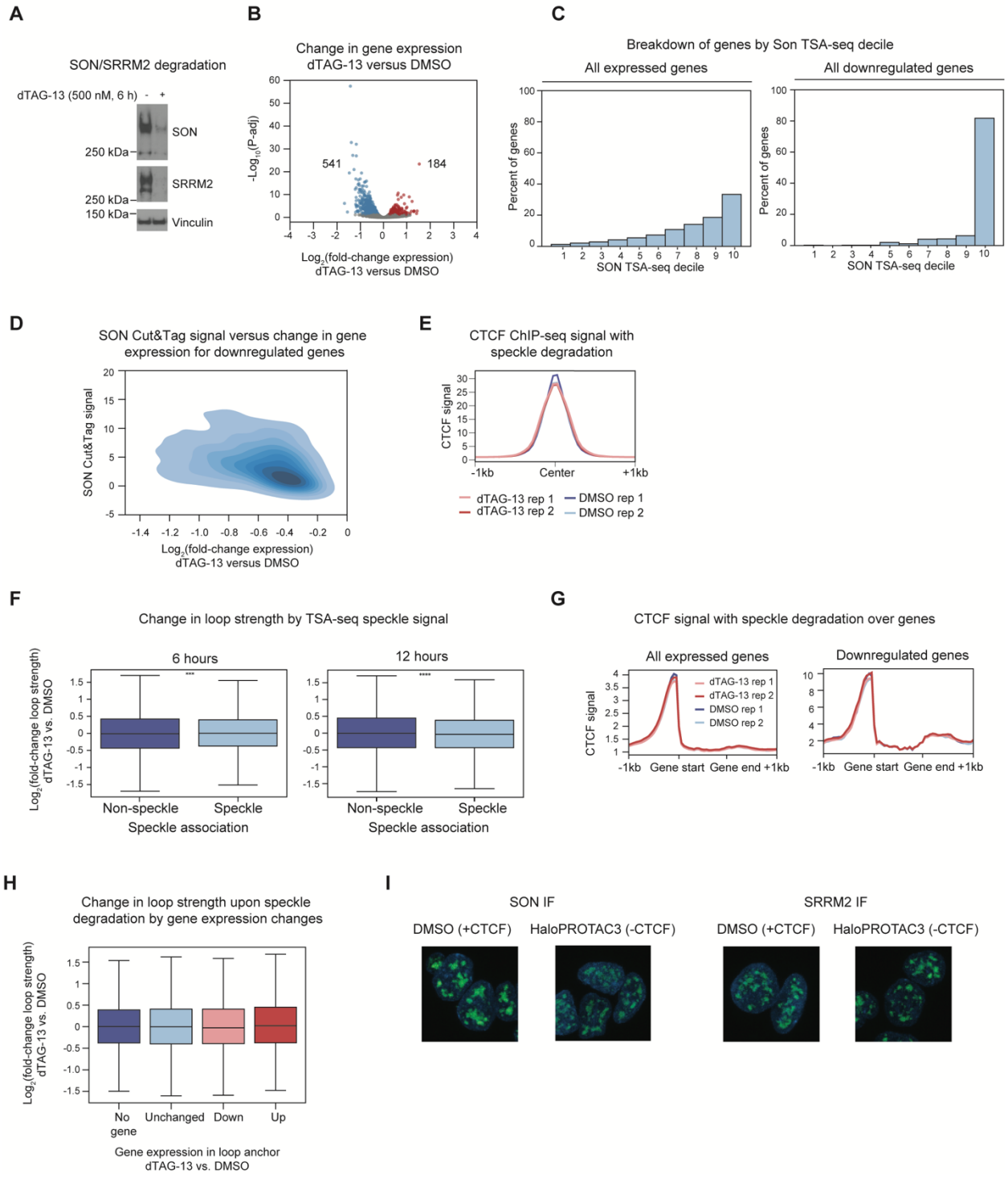
A
SON/SRRM2 degradation
dTAG-13 (500 nM, 6 h) - +
SON
250 kDa
SRRM2
250 kDa
150 kDa
Vinculin
B
Change in gene expression
dTAG-13 versus DMSO
-Log10(P-adj)
541
184
Log2(fold-change expression)
dTAG-13 versus DMSO
C
Breakdown of genes by Son TSA-seq decile
All expressed genes
All downregulated genes
Percent of genes
SON TSA-seq decile
D
SON Cut&Tag signal versus change in gene
expression for downregulated genes
SON Cut&Tag signal
Log2(fold-change expression)
dTAG-13 versus DMSO
E
CTCF ChIP-seq signal with
speckle degradation
CTCF signal
-1kb
Center
+1kb
dTAG-13 rep 1
dTAG-13 rep 2
DMSO rep 1
DMSO rep 2
F
Change in loop strength by TSA-seq speckle signal
6 hours
12 hours
Log2(fold-change loop strength)
dTAG-13 vs. DMSO
Non-speckle
Speckle
Speckle association
G
CTCF signal with speckle degradation over genes
All expressed genes
Downregulated genes
CTCF signal
dTAG-13 rep 1
dTAG-13 rep 2
DMSO rep 1
DMSO rep 2
-1kb
Gene start
Gene end
+1kb
H
Change in loop strength upon speckle
degradation by gene expression changes
Log2(fold-change loop strength)
dTAG-13 vs. DMSO
No gene
Unchanged
Down
Up
Gene expression in loop anchor
dTAG-13 vs. DMSO
I
SON IF
SRRM2 IF
DMSO (+CTCF)
HaloPROTAC3 (-CTCF)
DMSO (+CTCF)
HaloPROTAC3 (-CTCF)

Figure S4 | Acute disruption of nuclear speckles alters gene expression without disrupting CTCF

A. Separate western blots for SON and SRRM2 in speckle dTAG knock-in K562 cells ± 6 hours of 500 nM dTAG-13 treatment. Vinculin was used as a loading control.
B. 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, |$\log_2$(fold-change| > 0.25).
C. Proportion (y-axis) of all nonzero (left) and all 6 hour dTAG-13 downregulated (right, padj<0.05) genes in each SON TSA-seq decile (x-axis) from wildtype cells. 10 is the decile closest to speckles.
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. Aggregate profile plot of CTCF ChIP-seq signal (RPKM) in speckle knock-in K562 cells with and without 6-hour speckle degradation (y-axis) over all CTCF peaks (x-axis).
F. Boxplot showing $\log_2$(fold-change) loop strength for dTAG-13 treated vs. DMSO treated speckle dTAG knock-in cells relative to TSA-seq decile after 6 or 12 hours of dTAG-13 treatment with matched DMSO controls. Speckle refers to the decile closest to speckles, and non-speckle refers to the deciles 1-9 that are furthest from speckles.
G. Aggregate profile plot of CTCF ChIP-seq signal (RPKM) in speckle dTAG knock-in K562 cells with dTAG-13 vs. DMSO treatment (y-axis) across expressed genes (top) and genes that are downregulated with dTAG-13 treatment (bottom) (x-axis).
H. Boxplot showing $\log_2$(fold-change) loop strength for 6 hour dTAG-13 treated vs. DMSO treated speckle dTAG knock-in cells (y-axis) for loops with anchors that do not overlap any genes, overlap only genes that do not change expression, and overlap genes that increase or decrease in expression for at least one anchor (x-axis).
I. Representative SON and SRRM2 immunofluorescence images (green, alexafluor488) in HaloTag- CTCF HCT116 cells with 24 hours of 330 nM HaloPROTAC3 treatment to deplete CTCF or vehicle control. Overlaid on DAPI in blue.

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 ≤ 0.05; **: 0.001 <p ≤ 0.01; ***: 0.0001 <p ≤ 0.001; ****: p ≤ 0.0001.