



Published in final edited form as:

Cell Rep. 2021 February 23; 34(8): 108783. doi:10.1016/j.celrep.2021.108783.

Distinct properties and functions of CTCF revealed by a rapidly inducible degen system

Jing Luan¹, Guanjue Xiang^{2,10}, Pablo Aurelio Gómez-García^{3,10}, Jacob M. Tome^{4,11}, Zhe Zhang⁵, Marit W. Vermunt⁶, Haoyue Zhang⁶, Anran Huang⁶, Cheryl A. Keller², Belinda M. Giardine², Yu Zhang⁷, Yemin Lan⁸, John T. Lis⁴, Melike Lakadamyali⁹, Ross C. Hardison², Gerd A. Blobel^{6,12,*}

¹Medical Scientist Training Program, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

²Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA

³Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain

⁴Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY 14853, USA

⁵Department of Biomedical and Health Informatics, Children's Hospital of Pennsylvania, Philadelphia, PA, USA

⁶Division of Hematology, The Children's Hospital of Pennsylvania, Philadelphia, PA, USA

⁷Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA

⁸Penn Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

⁹Department of Physiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

¹⁰These authors contributed equally

¹¹Present address: Department of Genome Sciences, University of Washington, Seattle, WA, USA

¹²Lead contact

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*Correspondence: blobel@chop.edu.

AUTHOR CONTRIBUTIONS

J.L. and G.A.B. conceived the study and designed experiments. J.L., H.Z., and A.H. created transgenic cell lines. J.L. and C.A.K. performed ChIP-seq experiments; M.W.V. performed mapping and peak calling. G.X., R.C.H., and Y.Z. contributed to CTCF ChIP-seq normalization, clustering, and FIMO analysis. Downstream CTCF characterization was performed by J.L. with input from H.Z., M.W.V., G.X., and R.C.H. J.L. performed PRO-seq experiments with technical advice from J.M.T.; J.L. and Z.Z. analyzed PRO-seq results with input from M.W.V., J.M.T., and J.T.L. J.L. performed single-molecule imaging, with analysis performed by P.A.G.-G. and M.L. Y.L. and B.M.G. provided support for bioinformatic analyses. J.L. and G.A.B. wrote the manuscript with input from all authors.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.celrep.2021.108783>.

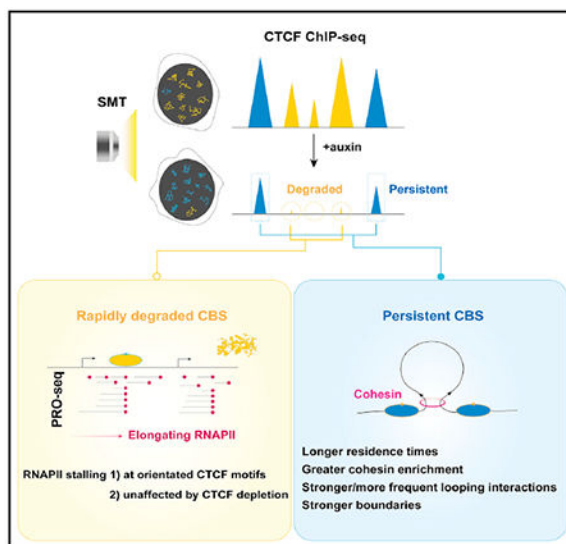
DECLARATION OF INTERESTS

The authors declare no competing interests.

SUMMARY

CCCTC-binding factor (CTCF) is a conserved zinc finger transcription factor implicated in a wide range of functions, including genome organization, transcription activation, and elongation. To explore the basis for CTCF functional diversity, we coupled an auxin-induced degron system with precision nuclear run-on. Unexpectedly, oriented CTCF motifs in gene bodies are associated with transcriptional stalling in a manner independent of bound CTCF. Moreover, CTCF at different binding sites (CBSs) displays highly variable resistance to degradation. Motif sequence does not significantly predict degradation behavior, but location at chromatin boundaries and chromatin loop anchors, as well as co-occupancy with cohesin, are associated with delayed degradation. Single-molecule tracking experiments link chromatin residence time to CTCF degradation kinetics, which has ramifications regarding architectural CTCF functions. Our study highlights the heterogeneity of CBSs, uncovers properties specific to architecturally important CBSs, and provides insights into the basic processes of genome organization and transcription regulation.

Graphical Abstract



In brief

Using an inducible CTCF degradation approach, Luan et al. show that sensitivity to degradation of CTCF on chromatin is highly variable, reflecting its diverse roles in chromatin architecture and transcription.

INTRODUCTION

CCCTC-binding factor (CTCF) is a DNA-binding protein with diverse roles in chromatin architecture and gene regulation. The contextual basis for distinct CTCF functions remains largely unknown. As an architectural protein in chromatin, one role of CTCF involves domain organization. CTCF binding sites (CBSs) are found at chromatin domain boundaries, such as topologically associating domains (TADs) and sub-TADs, where CTCF is found to function with cohesin, a ring-shaped multi-protein complex (Parelho et al., 2008;

Rubio et al., 2008; Wendt et al., 2008). Domain organization has been proposed to result from a so-called “loop-extrusion” process, during which cohesin extrudes the chromatid until it stalls at an oriented CBS, ultimately stabilizing a chromatin loop (Fudenberg et al., 2016; Sanborn et al., 2015). CTCF bordered domains can promote chromatin contacts within such domains while impeding cross-domain interactions (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012).

Despite the presence of CTCF at nearly 90% of TAD boundaries in mammalian cells (Dixon et al., 2012), the majority of CBSs reside within chromatin domains—near enhancers, promoters, or even within gene bodies (GBs)—reflecting a wide variety of activities including transcription activation, repression, elongation, and pre-mRNA splicing (Phillips and Corces, 2009). CTCF also promotes transcription activation directly at gene promoters or via fostering of contact with distal enhancers. Studies at single loci, as well as genome-wide, have implicated CTCF in both transcription initiation and RNA polymerase II (RNAPII) pause-release at the promoters, either dependent or independent of its architectural involvement (Chernukhin et al., 2007; Laitem et al., 2015; Paredes et al., 2013; Peña-Hernández et al., 2015). Within GBs, CTCF has been implicated in modulating RNAPII processivity, which, in select cases, has been linked to mRNA splicing efficiency (Kang and Lieberman, 2011; Mayer et al., 2015; Ruiz-Velasco et al., 2017; Shukla et al., 2011; Stadhouders et al., 2012).

The determinants that specify such distinct CTCF functions are still largely unknown. Various depletion strategies have been employed to interrogate CTCF requirements in genome organization and transcription (Busslinger et al., 2017; Hyle et al., 2019; Khoury et al., 2020; Kubo et al., 2021; Nora et al., 2012, 2017; Thiecke et al., 2020; Wutz et al., 2017; Zuin et al., 2014). A unifying theme among these studies has been the relatively mild effect on gene expression despite significant architectural perturbations, including the weakening of TAD boundaries. Despite the important insights gained from these studies, they typically involved depletion timescales of greater than 24 h, which might confound primary and secondary effects on both architecture and gene expression. Moreover, transcriptional changes were assessed by examining steady-state mRNA levels, which are affected by RNA processing and turnover. Finally, it has been noted that removal of CTCF from chromatin does not occur uniformly or completely (Hyle et al., 2019; Khoury et al., 2020; Kubo et al., 2021; Nora et al., 2017), and CBSs that persist tend to be located at domain boundaries. The molecular and functional basis for these observations is unknown.

Here, we combined short-timescale CTCF depletion (Nishimura et al., 2009) with genome-wide CBS profiling, PRO-seq (Kwak et al., 2013; Mahat et al., 2016b), and single-molecule tracking (SMT) (Manzo and Garcia-Parajo, 2015) to dissect CTCF function in relation to genomic contexts. We found that after initiating CTCF degradation, the chromatin persistence patterns at CBSs were highly variable, and that variation was not predicted by CTCF motif sequence alone. Persistent CBSs, often detected at chromatin boundaries and colocalized with cohesin, were associated with but not entirely predicted by strong initial signal intensities. Based on PRO-seq, acute CTCF loss resulted in only modest changes in transcriptional initiation, pause-release, and elongation. Even though in GBs, CBSs were found at sites of RNAPII stalling, unexpectedly, the CTCF motif itself appeared to sustain

stalling even upon CTCF depletion. Finally, SMT uncovered a correlation between CTCF persistence and residence times on chromatin, which has implications regarding the functional diversity of CBSs.

RESULTS

Highly variable CTCF persistence on chromatin following auxin-mediated degradation

To investigate the roles of CTCF in chromatin architecture and transcription, we fused AID and mCherry proteins at the C terminus of endogenous CTCF in an established mouse erythroblast cell line, G1E-ER4 (Weiss et al., 1997), via CRISPR/Cas9-facilitated genome editing (Zhang et al., 2019). We stably integrated a TIR-expressing construct that allows for rapid and global CTCF degradation via auxin-dependent proteasomal targeting (Figure 1A) (Morawska and Ulrich, 2013; Nishimura et al., 2009). We hereafter refer to this cell line as “CTCF-AID-mCherry+TIR.” The modified cells grew and differentiated normally, suggesting that CTCF function was not measurably perturbed (not shown), even though the steady-state levels of the fusion protein were somewhat lower compared to the endogenous CTCF in the parental cells, similar to what has been reported (Nora et al., 2017).

Within 1 h of exposure to auxin, CTCF levels became virtually undetectable by western blot of whole cell lysates (Figure 1A). Levels remained low (<1%) for any duration of treatment (Figure S1A), even when examining extracts enriched for nuclear or chromatin-associated proteins (Figure S1B). Therefore, we were surprised that chromatin immunoprecipitation sequencing (ChIP-seq) uncovered persistence of a fraction of CTCF peaks in some cases even up to 24 h, albeit at reduced intensities (Figure 1B). ChIP-qPCR at representative sites in which the signal was normalized to input chromatin confirmed these results, thus excluding potentially confounding effects of library normalization in ChIP-seq experiments. Additionally, ChIP-seq libraries at all time points were normalized to input-calibrated ChIP-qPCR of 22 sites that comprised varying signal intensities and degradation dynamics, using a linear regression model (Figure S1C; Table S1; see Method details) (Behera et al., 2019; Shao et al., 2012; Xiang et al., 2020). We further verified that the presence of persistent CBSs was not a clonal artifact by performing CTCF ChIP-qPCR in an independent clonal line and observed similar degradation patterns (Figure S1D; Table S2). Finally, persistent sites were not due to non-specific antibody recognition, as confirmed by anti-mCherry ChIP-seq, with control immunoglobulin G (IgG) ChIP-seq and chromatin input serving as negative controls (Figure S1E).

To begin to dissect the basis for non-uniform CBS degradation patterns, we first used two orthogonal methods to place the CBSs into categories based on their sensitivity to degradation. First, we employed the model-based, unsupervised clustering method Mclust (Medvedovic et al., 2004), which placed the 38,844 CBSs into six clusters (Figure 1C). The clusters differed in the rate and completeness of decay, with cluster 6 showing consistently slower rates than the rest. The initial binding intensity at 0 h also distinguished several clusters, but it was not the sole determinant of cluster placement, as clusters 5 and 6 were still partitioned separately when matched for initial binding intensities (Figure S1F; see Figure 1B for examples).

To ensure that our categorizations were robust, we also classified the CBSs by fitting their degradation kinetics to an exponential decay (ED) function (Figure 1D). The decay of most CBSs fits closely to ED ($R^2 > 0.7$; Figure S1G). For a significant number of CBSs, the initial signal intensity assumed a linear relationship with the decay coefficient $-\lambda$ (Figure 1D, the “rapidly degraded”), which directly resulted from virtually complete degradation by 4 h of auxin treatment (see mathematical proof in Method details). However, a distinct set of CBSs exhibited slower rates of decay, as shown by the points above the linear zone (Figure 1D, “cluster III”). After further stratifying the “rapidly degraded” CBSs by their initial binding enrichment, which followed a bi-modal distribution spanning a wide range, we obtained three ED-based clusters in total (Figure 1D). The CBSs in the six clusters from Mclust were largely placed into analogous clusters by the ED analysis (Figure S1H). Thus, two fundamentally different approaches placed CBSs into clusters based on their auxin-induced degradation kinetics.

In prior CTCF-AID studies in murine embryonic stem cells (mESCs), degradation-resistant CBSs were also seen but not further classified (Kubo et al., 2021; Nora et al., 2017). Since a large fraction of CBSs is tissue invariant, we were able to reprocess ChIP-seq datasets with our pipeline and found that the most persistent and enriched sites in erythroid cells (Mclust-based cluster 6) were also the most conserved and most persistent in the mESCs (Figure S1I). We also noticed that the removal of CTCF from chromatin appeared to be more complete in G1E-ER4 cells compared to mESCs, the reasons for which are unknown. A similar conclusion could be drawn when we made comparisons with a recent study that identified persistent CBSs after prolonged RNA interference (RNAi)-mediated depletion in two human cell lines (Khoury et al., 2020) (Figure S1J).

Together, these results indicate that degradation dynamics of CBSs are highly variable in a manner not solely accounted for by binding intensity, and are shared across species/cell types and different depletion modalities. This heterogeneity may reflect functional and/or location-dependent differences in CTCF chromatin binding. Additionally, our clustering strategies allow for a more refined stratification of CBSs that may be applicable to any chromatin-associated protein.

Characteristics of persistent CBSs

The amount of CTCF binding and resistance to degradation may be impacted by motif sequence, location, chromatin modifications, and contextual transcription factors (Behera et al., 2018; Ghirlando and Felsenfeld, 2016; Nakahashi et al., 2013; Plasschaert et al., 2014; Rhee and Pugh, 2011; Zuin et al., 2014).

Overall, we observed remarkable similarity between all clusters in motif sequence with minor differences in FIMO (find individual motif occurrence) scores (Figures 2A, S2A, and S2B). No significant difference was observed in the average number of high-confidence motifs under CBSs in each cluster (not shown). As upstream and downstream sequences flanking the core motif have been proposed to affect the affinity of CTCF to chromatin (Boyle et al., 2011; Kim et al., 2007; Nakahashi et al., 2013; Rhee and Pugh, 2011; Schmidt et al., 2012), they may also modulate CTCF degradation kinetics. We classified all CBS clusters based on the presence/absence of these motifs as previously described (Nakahashi et

al., 2013) and noticed a paucity of destabilizing downstream (D) motifs at persistent CBSs, suggesting that reduced chromatin binding may partially facilitate CTCF degradation (Figure S2C).

MEME-ChIP further revealed a distinguishing feature in cluster 6, which was an A/T-rich sequence located within ~200 bp of more than 60% of the CBSs (Figure S2D). We tested the role of this sequence in CTCF persistence through CRISPR/Cas9-mediated genome editing. We selected a ~100-bp-long A/T-rich sequence within the silent *Myrip* gene that has multiple strong/persistent CBSs nearby (Figures S2E and S2F). Homozygous deletion of this element led to no significant changes in either CTCF binding intensity or degradation kinetics, nor did it affect local transcription (Figures S2G and S2H). The biological significance of this element, if any, remains an open question.

We next investigated the genomic distributions and histone marks of persistent CBSs. Notably, sites with delayed degradation kinetics exhibited striking differences from others, with less representation in introns and increased presence at enhancers, in accordance with the corresponding histone modifications (Figures 2B and S2I). The differences were robust even when CBS signal intensities were matched, suggesting that they were not dependent upon initial binding enrichment.

We have previously shown that a subset of CBSs was retained on chromatin during mitosis, an interval during which most nuclear factors are evicted from chromatin (Zhang et al., 2019). We noted a striking correlation between CTCF retention on mitotic chromatin and resistance against degradation ($p < 0.0001$, Chi-squared test), even when comparing clusters 5 and 6 that have comparable CBS intensities (Figure 2C and S2J). A significant shift in the cell cycle profile was not observed and therefore does not account for this observation (Figure S2K). These results suggest that CBS degradation variability can be linked to specific genomic contexts.

Persistent CBSs are associated with chromatin loops and domain boundaries

One of CTCF's functions is to facilitate the formation of long-range chromatin loops and enforce domain boundaries. We assessed whether CBS persistence is related to any of these functions by analyzing our recently generated Hi-C data in G1E-ER4 cells (Zhang et al., 2019). We observed a significantly greater proportion of persistent CBSs at loop anchors, with persistence correlating with the number of looping interactions (Figures 2D, S2L, and S2M). We further parsed out so-called structural loops, as defined here by long-range contacts that are flanked by CBSs on both anchors, and enhancer-promoter loops with CBSs on just one or no anchor. The most persistent CBSs appeared to engage in structural loop interactions more frequently than the others (Figure S2N). The strength of structural loop interactions also correlated with the persistence of CBSs on both sides (Figures 2E and S2O).

Given that convergently oriented CBSs often function as boundaries for loop-extruding cohesin, a process essential for long-range looping interactions and domain formation, we assessed cohesin enrichment in relation to CBS degradation kinetics. Cohesin sub-unit Rad21 co-localized extensively with CBSs genome wide, with signal intensity scaling with

that of co-occupying CTCF before and after CTCF depletion (Figures 2F and S2P). Interestingly, Rad21 enrichment was greatest at persistent CBSs, even when comparing clusters 5 and 6 that had comparable CTCF binding intensities at 0 h. This suggests that cohesin binding correlates not only with CTCF enrichment, but also with persistence against degradation.

To test if increased Rad21 occupancy at persistent CBSs translates into greater domain boundary strength, we measured insulation scores (ISs) from Hi-C experiments (Mizuguchi et al., 2014; Zhang et al., 2019). Persistence of CBSs inversely correlated strongly with ISs (lower IS reflecting fewer cross-domain contacts), with cluster 6 exhibiting the strongest domain insulation (Figures 2G and S2Q). Our findings are consistent with a recent report showing that CBSs that persist following RNAi-mediated CTCF targeting are enriched at cell-type invariant loops and chromatin domain boundaries (Khoury et al., 2020).

In sum, CTCF resistance to AID-mediated degradation is associated not only with binding intensity, but also with additional features that relate to chromatin context, such as loop formation and chromatin domain boundaries. The molecular basis for CBS persistence is unknown but might be linked to chromatin binding dynamics. This possibility is further addressed below.

Effects of CTCF degradation on nascent transcription

Prior studies using mRNA-seq or micro-arrays that reported only limited transcriptional changes upon CTCF depletion were confounded by RNA stability and did not provide insights into fine-scale transcriptional dynamics. Therefore, we performed PRO-seq (Kwak et al., 2013; Mahat et al., 2016b) to provide a readout of actively transcribing RNAPII after 4 h of auxin treatment, which is significantly shorter than in similar studies and is expected to minimize secondary effects (Figure 3A). Cells lacking TIR treated with auxin served as a control. A total of 10,748 transcripts were identified, corresponding to 9,497 genes. To examine CTCF involvement in transcription initiation, pause-release, and elongation, we counted total reads over the promoter (−50 to +150 bp relative to Refseq-annotated transcription start sites [TSSs]) and GB (+200 bp relative to TSS to −500 bp relative to transcription end site [TES]) and calculated pausing index (PI) (log-transformed TSS/GB read densities) (Figure 3A).

We first investigated changes in GBs as it translates into transcriptional output. Similar to previous studies, we observed only 54 significantly upregulated (false discovery rate [FDR] < 0.05 and fold change > 2) and 68 downregulated (FDR < 0.05 and fold change < 0.5) transcripts, accounting for 1.1% of the total number of active transcripts (Figure 3B). Auxin non-specific changes were minimal (3 genes upregulated). Changes in transcription initiation and pause-release were similarly limited (Figure 3B). We further confirmed a lack of transcriptional changes at the *c-Myb* locus after CTCF depletion, where the CBS in its first intron was proposed to facilitate RNAPII elongation via distal connections with upstream enhancers (Stadhouders et al., 2012) (Figure S3A). Auxin-mediated CTCF degradation led to no significant changes in PRO-seq signal (Figure S3A). CRISPR/Cas9-mediated CBS deletion also resulted in no significant transcription perturbation (Figures S3B–S3D). Overall, we conclude that CTCF depletion does not widely perturb transcription

initiation, pause-release dynamics, or transcription elongation. It remains possible that residual CTCF binding maintains some important functions, which could only be assessed genome wide if complete degradation of CTCF were achievable or via local elimination of CBSs.

RNAPII stalls at CBSs in a CTCF-dispensable manner

The apparent lack of CTCF degradation on all transcription parameters was unexpected in light of previous literature. However, when quantifying 3' PRO-seq signals around CBSs within highly transcribed GBs, we observed significant enrichment specifically at forward-oriented CTCF motifs, suggestive of RNAPII stalling (Figures 3C and 3D) and consistent with a prior study (Mayer et al., 2015). We first asked whether the PRO-seq signal enrichment was indeed due to stalling (i.e., accumulation at the 3' end only) or to internal transcription initiation at the CBSs. In the latter case, 5' PRO-seq signals would be expected to cluster at the initiation sites on both strands (i.e., bidirectional transcription) (Core et al., 2008; Kapranov et al., 2007; Seila et al., 2008). Neither was observed in our data, arguing against this possibility (Figure S3E). This result left us with the conundrum as to why CBSs might cause RNAPII stalling while CTCF degradation had little global impact on elongation. To test whether CTCF causes local RNAPII stalling, we examined PRO-seq signals at the CBSs following CTCF degradation. Surprisingly, we found stalling at the CBSs remains fully intact despite ChIP-seq-validated loss of CTCF binding (Figures 3C, 3D, and S3F).

We further examined RNAPII stalling globally with respect to CTCF signal intensities and CBS persistence. The degree of RNAPII stalling was independent of the initial levels of CTCF binding (Figure S3F) and not predicted by CBS persistence category (not shown). Moreover, similar degrees of stalling still occurred at high-confidence CTCF motifs that were devoid of measurable CTCF binding under any condition (Figures 3E, 3F, and S3G). In all cases, however, the orientation of the CTCF motif was critical, and stalling was enriched at the G-rich section and an upstream C of the motif (Figure S3H). The amount of stalled RNAPII further correlated with the level of sequence resemblance to the canonical CTCF motif (Figure S3I). Overall, oriented CTCF motifs appear to mediate RNAPII stalling independent of CTCF binding.

CTCF binding is not impaired by elongating RNAPII

While we saw little evidence to suggest that CTCF affects RNAPII processivity, we asked whether, conversely, RNAPII might affect CTCF occupancy. Globally, there was no anti-correlation between PRO-seq signal and CTCF binding enrichment or persistence in the presence of auxin (Figures 3G and S3J). Overall, there were few significant changes in intragenic CTCF binding after transcription inhibition with 5,6-dichloro-1-D-ribofuranosyl-benzimidazole (DRB) (5 h; 11 gains/0 loss; 9,134 sites total; FDR < 0.05 & fold change > 2) and triptolide (7 h; 3 gains/0 loss; FDR < 0.05 & fold change > 2) (Figures 3H, S3K, and S3L). It thus suggests that globally, RNAPII does not displace CTCF in this cell system, even though there have been reported cases where transcription promotes or compromises CTCF occupancy (Heinz et al., 2018; Saldaña-Meyer et al., 2019).

In sum, these surprising results suggest RNAPII can stall at oriented CTCF motifs in a manner ostensibly independent of CTCF binding. Multiple possibilities might account for this observation. First, other DNA-binding proteins might compensate for CTCF function by occupying the entire CTCF motif (Kaaij et al., 2019; Loukinov et al., 2002) or sub-segments of it. These factors would have to function in an orientation-dependent manner. Second, the DNA sequence itself may present a hindrance to RNAPII processivity. Poly-G sequences have been reported to stall RNAPII via RNA:DNA hybrid formation (Belotserkovskii et al., 2010; Chen et al., 2017; Mischo et al., 2011; Skourti-Stathaki et al., 2011; Watts et al., 2019; Yonaha and Proudfoot, 1999). However, in most cases, stalling is expected to occur several base pairs downstream after the G-rich sequence exits from RNAPII, which is different from our observation where stalling occurs at G-rich positions. Additionally, GC-rich non-template DNA strands may form secondary structures to mediate RNAPII pausing (Szlachta et al., 2018), although it does not appear to explain our data based on preliminary secondary structure analysis (not shown). Third, nucleotide availability may affect RNAPII processivity. Promoter-proximal pausing has been found to preferentially occur during the incorporation of CTP (Gressel et al., 2017; Tome et al., 2018), which is the least abundant nucleotide, followed by GTP (Traut, 1994). Similarly, in GBs, it may take longer for RNAPII to sample via diffusion the less abundant nucleotides that comprise the CTCF motifs. The global lack of a stalling function by CTCF nevertheless does not preclude such a role at individual loci in select circumstances (Shukla et al., 2011). One conclusion from these results is that the occurrence of CTCF at a stalling site does not automatically allow inferences about CTCF actually functioning there. This serves as a reminder to exert caution when linking transcription factor binding to a particular phenotype and stresses the importance of using acute degradation systems in such cases.

CBS persistence is linked to residence time on chromatin

The loop-extrusion model posits that CTCF functions as a barrier to the cohesin-driven chromatid extrusion process. Our results indicate that CBSs with delayed degradation kinetics are associated with high cohesin co-occupancy, are enriched at chromatin boundaries, and are linked to long-range chromatin contacts. This, in turn, might require distinct chromatin-binding properties compared to CBSs at other locations. We hypothesized that resistance to degradation is a function of longer residence times, such that stable on-chromatin association with architectural complexes limits the accessibility of degradation machineries. We thus set out to test this hypothesis via 2D SMT at 0 h and 24 h of auxin treatment. The residence time estimated at 0 h would reflect global average binding kinetics and may exhibit a wide distribution, while that after auxin treatment would be enriched for persistent sites and may shift toward the longer spectrum.

We fused AID-HaloTag to endogenous CTCF via CRISPR/Cas9 engineering and labeled cells with HALO-ligand-Janelia Fluor 549 dye (JF549) (Grimm et al., 2015, 2016; Hansen et al., 2017) (Figure 4A). ChIP-seq confirmed that the HaloTag did not significantly alter CTCF degradation behavior even though baseline levels of the fusion protein were somewhat lower than CTCF lacking AID, as has been observed previously (Nora et al., 2017) (Figure 4B). To minimize bias associated with under-sampling after CTCF depletion, we optimized the concentrations of JF549 to sparsely label a similar density of molecules

with similar signal-to-noise ratios (SNRs) across conditions (Figure S4A). Similar to flow cytometric observations (Figure S1A), quantification of SMT signal densities per cell showed homogeneous and near-complete (>99.9%) CTCF depletion after 24 h auxin treatment (Figure S4B; Table S4).

To calculate residence times of the chromatin-bound fraction, we employed a low-excitation and slow-tracking (2 Hz with a long camera integration time of 500 ms) modality to motion-blur all freely diffusing molecules. We estimated the average residence times to be 15–20 s at baseline, which shifted toward longer times (on average, by 3–8 s) after CTCF depletion across all biological replicates and different imaging settings (Figures 4C and S4C–S4F). Of note, our long residence times are lower than previously reported (Agarwal et al., 2017; Hansen et al., 2017) and are likely underestimates due to significant photo-bleaching, thus blunting true differences between conditions. Additionally, given that SMT results are heavily dependent on experimental conditions and cell type, and are influenced by laser power and photobleaching, comparisons under the same experimental conditions (as was done here) are more meaningful than between studies.

Our results show that CTCF residence time correlates with resistance to degradation, which, as shown above, is also correlated with boundary and looping function. To be clear, we are not suggesting that CTCF is capable of stable chromatin binding for hours to still be detectable at later time points by ChIP-seq. In fact, our SMT data suggest that chromatin binding is highly dynamic even at persistent CBSs. The longer residence times may nevertheless confer protection against degradation, although the mechanism is unknown. It is tempting to speculate that chromatin-binding dynamics are functionally linked to higher order chromatin organization. Based on predictions from the loop-extrusion model (Fudenberg et al., 2016; Sanborn et al., 2015), prolonged chromatin residence may facilitate cohesin blockade and contribute to loop formation and boundary strength. This is substantiated by the increased cohesin colocalization, more frequent/stronger loop interactions, and greater insulation capacity observed at the persistent CBSs. Finally, we speculate that residence time and on/off rates both contribute to binding enrichment measured in ChIP-seq. While longer residence time may account for high ChIP-seq signal intensities at persistent CBSs, more frequent/rapid binding may account for those that are equally enriched but rapidly degraded. Overall, our SMT results suggest the chromatin-binding behavior of CTCF is diverse, the biological significance of which warrants future investigation.

In sum, the experimental and computational strategies complementing the transient depletion system of CTCF provide new insights into CTCF's variable behavior on chromatin and its diverse roles in transcription and genome organization. We envision that such studies can be broadly applied to untangle the functional diversities of other nuclear factors.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Gerd A. Blobel.

Materials availability—All unique/stable reagents generated in this study are available upon request to the lead contact.

Data and code availability—All datasets reported in this paper are available at the Gene Expression Omnibus with accession number GEO: GSE150418.

The code used for ChIP-seq normalization can be accessed through: https://github.com/guanjue/CTCF_Auxin_clustering

EXPERIMENTAL MODEL AND SUBJECT DETAILS

G1E-ER4 is an established murine erythroblast cell line (Weiss et al., 1997). G1E-ER4 cells were grown in IMDM+15% FBS, penicillin/streptomycin, Kit ligand, monothioglycerol and erythropoietin in a standard tissue culture incubator at 37C with 5% CO₂. Cells were maintained at density below 1million/ml at all times.

To increase CRISPR editing efficiency at the poly-A/T region and *c-Myb* locus, we established a Cas9-TagBFP expressing G1E-ER4 cell line. Specifically, G1E-ER4 cells were infected with EFS-Cas9-P2A-TagBFP retrovirus. To engineer EFS-Cas9-P2A-TagBFP construct, TagBFP was amplified from pKLV-U6gRNA-EF(BbsI)-PGKpuro2ABFP (RRID: Addgene_62348) and cloned into an optimized lentiviral EFS-Cas9-P2A-Puro expression vector in place of Puro via In-Fusion cloning system (Clontech: #638909). Following infection, cells were expanded for 2-3 days, with cells expressing top 1%–5% BFP signal selected by a Beckman Coulter Moflo Astrios sorter.

All sgRNA encoding oligonucleotides were inserted into a retroviral U6-sgRNA-PGK-GFP expression vector (Stonestrom et al., 2015) using a BsmBI restriction site. A Myrip-A/T_Mut (+/-) single-cell clone was generated by co-transfecting Cas9-TagBFP expressing G1E-ER4 cells with two guide RNA plasmids, followed by single-cell clone screening and genotyping by Sanger sequencing. The heterozygous clones were then edited a second time with the same guide RNA plasmids, generating both homozygous and WT clones (where the unedited allele possibly served as a repair template). After single-cell screening, a homozygous clone, Myrip-A/T_Mut (-/-), and a WT clone, Myrip-A/T_WT (+/+), were selected and confirmed by Sanger sequencing. We limited comparison to these three single-cell clones as they were derived from the same parental cell, thus minimizing clonal variation common in CRISPR editing experiments. Myb-CTCF_Mut cell lines were generated by co-transfecting Cas9-TagBFP expressing G1E-ER4 cells with two guide RNA plasmids and then sorted, screened, and sequenced as above. CTCF ChIP-qPCR was additionally performed to confirm CTCF binding disruption.

CTCF-AID-HaloTag+TIR cell line was generated by co-transfecting WT G1E-ER4 cells with 18ug pX330-GFP-sgCTCF (spCas9 with CTCF sgRNA) and 6ug of donor template amplified from a Ctfc-Halo-mAID donor template (RRID: Addgene_113103), followed by screening described above. Single-cell clones were subsequently transfected with MigR1-osTIR1-9Myc-GFP and expanded, with cells exhibiting top 1%–5% GFP signals selected via a Beckman Coulter Moflo Astrios sorter.

METHODS DETAILS

Cell culture and maintenance—G1E-ER4 cell culture has been described previously (Weiss et al., 1997). Auxin was added at 1mM.

Retroviral infection of murine cells—To generate G1E-ER4 cells stably expressing Cas9-TagBFP, cells were infected with EFS-Cas9-P2A-TagBFP retrovirus. Virus was produced in HEK293T cells grown in DMEM supplemented with 10% Fetal Bovine Serum, 2% penicillin/streptomycin, 1% L-glutamine, 100 μ M sodium pyruvate, to 90% confluence on day of transfection. Vector EFS-Cas9-P2A-TagBFP was mixed in a 4:3:2 ratio with packaging plasmids, PAX2, and VSVG envelope plasmid, respectively, in 500 μ L OPTI-MEM (ThermoFisher Scientific, Cat#31985070), and added to 500 μ L of 160 μ g/ml polyethylenimine (PEI, Polysciences Cat#23966) in OPTI-MEM for precipitation. The mixture was added to cells drop-wise and incubated at 37°C. Media was changed 6h post-transfection. Viral supernatant was harvested 24h and 48h post-transfection and pooled. For retroviral infection, 1 million G1E-ER4 cells were plated in 6-well plates with 1mL of media and 1mL of viral supernatant, supplemented with 8 μ g/ml polybrene and 10mM HEPES buffer. Spin-infections were carried out at 3,200 rpm for 1.5h at room temperature. Following infection, cells were re-suspended in fresh media and expanded for sorting.

CRISPR/Cas9-mediated genome editing—We employed CRISPR/Cas9-mediated homology-mediated repair (HDR) to fuse AID and HaloTag sequences at the C terminus of endogenous *Ctcf*. The HDR donor template was a gift from Janet Rossant (RRID: Addgene_113103) (Gu et al., 2018), with a 2.9kb region amplified by a pair of primers encompassing the AID-HaloTag sequence and 800-950 bp homology arms on both ends. The gRNA-containing plasmid was previously used to engineer CTCF-AID-mCherry+TIR (Zhang et al., 2019). Specifically, 18ug of gRNA-containing plasmid and 6ug of gel-purified linear HDR donor template were co-transfected into 3 million G1E-ER4 cells with Amaxa II electroporator (Lonza; program G-016) and Amax II Cell Line Nucleofector Kit (R) (Lonza, VCA-1001). Transfected cells were cultured in antibiotic-free medium for 72h and briefly labeled with JF5490-conjugated HALO-ligand before sorting via a Beckman Coulter Moflo Astrios sorter. Single-cell clones were then expanded for 5-7 days, followed by colony screening and genotyping by Sanger sequencing. Positive clones were subsequently expanded and transfected with a MigR1-osTIR1-9Myc-GFP construct, with those expressing top 1%–5% GFP signals selected via a Beckman Coulter Moflo Astrios sorter.

To delete A/T-rich sequences at *Mytip* locus (Figure S2D) and CBS in the first intron of *c-Myb* (Figure S3B), two gRNA-containing plasmids targeting each end of the sequences were co-transfected into Cas9-TagBFP expressing G1E-ER4 cells (see Experimental models and subject details), followed by sorting and screening as detailed above. All guide RNA sequences were picked using CRISPR design tool (<https://zlab.bio/guide-design-resources>) (Cong and Zhang, 2015).

Western blotting—Cells were washed once with PBS, pelleted, snap-froze in liquid nitrogen and stored at –80°C. Two million cells were resuspended in 100 μ L lysis buffer (20 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1mM EGTA, 1% Triton X-100, 1x complete

protease inhibitor (Roche)), and sonicated for 10 minutes total ON time with pulses of 15 s ON and OFF, and 40% amplitude with QSONICA 800R (Qsonica). Sub-cellular fractionation was accomplished using Subcellular fractionation kit (ThermoFisher; Cat#78840). The samples were run on 4%–15% Mini-PROTEAN® TGXTM Precast Gels (Bio-Rad), and transferred onto nitrocellulose membranes at 100 V for 45 min. The membranes were rinsed with 1x TBST and blocked with 5% dry milk at room temperature for 1h. After washing with TBST, membranes were incubated with anti-CTCF antibody (Millipore, 07-729) (1:1,000), anti-Rad21 antibody (Bethyl Laboratories; A302-583A) (1:1,000), anti-H3 histone antibody (Cell signaling; 9715) (1:10,000), and anti- β -Actin-Peroxidase antibody (Millipore; A3854) (1:10,000) diluted in blocking buffer overnight at 4°C. Subsequently, membranes were washed 3 times in TBST at room temperature for 10min, then incubated with secondary antibody, goat anti-Rabbit IgG (Bio-Rad) (1:10,000), in blocking buffer at room temperature for 1h. After washing 3 times in TBST, proteins of interest were detected using Pierce ECL Western Blotting Substrate (ThermoFisher; 32109).

Cell-cycle analysis by DAPI staining—Cells were pelleted and washed once in PBS. One million cells were resuspended in 500ul ice-cold PBS and then added to 4.5 mL of ice-cold methanol. Cells were kept at –20°C for 1h. Cells were then pelleted at 1200 rpm for 3 min at 4°C and resuspended in 0.5mL 0.1% Triton X-100 in PBS containing 20ng/ml DAPI. Afterward, cells were transferred on ice and used directly for flow cytometry. Cell-cycle quantification was performed using Flowjo software fitted with the Dean-Jett Fox (DJF) model.

ChIP-seq—Chromatin immunoprecipitation (ChIP) was performed as previously described (Letting et al., 2004). Antibodies include: CTCF (Millipore; 07-729), Rad21 (Bethyl Laboratories; A302-583A), mCherry (Abcam; Ab167453), RNAPII (Cell Signaling; Cat#14958), IgG from rabbit serum (Sigma; 15006). Quantitative polymerase chain reaction (qPCR) was performed using Power SYBR Green kit (Invitrogen; 4368577) with signals detected by ViiA7 System (Life Technologies). ChIP-seq libraries were prepared using Illumina’s TruSeq ChIP sample preparation kit (Illumina, Cat#IP-202-1012) according to manufacturer’s specifications, with the addition of size selection (left side at 0.9x, right side at 0.6x) using SPRIselect beads (Beckman Coulter, Cat#B23318). Library size was determined (average 351 bp, range 333-372 bp) using the Agilent Bioanalyzer 2100, followed by quantitation using real-time PCR using the KAPA Library Quant Kit for Illumina (KAPA Biosystems; Cat#KK4835). Libraries were then pooled and sequenced (1x75bp) on the Illumina NextSeq 500 platform according to manufacturer’s instructions. Bclfastq2 v 2.15.04 (default parameters) was used to convert reads to fastq.

PRO-seq—PRO-seq experiments were performed as previously reported (Mahat et al., 2016b) with modifications. For each library, 50 million cells were used together with 2 million *Drosophila* Schneider 2 (S2) cells added as spike-in to control for potential global bias associated with library scaling. After 0h and 4h auxin treatments cells were collected simultaneously and processed in pairs. Instead of collecting and storing cells in storage buffer at –80°C and resuming run-on experiments later by adding 2x nuclear run-on (NRO) buffer as described in the original protocol, we directly resuspended freshly prepared nuclei

into storage buffer/2x NRO buffer mixed in 1:1 ratio to start nuclear run-on immediately after collection. All four biotin-NTPs were supplied at equal ratio to achieve single-nucleotide resolution. To facilitate the removal of PCR duplicates, we added random hexamers to the 5' end of the VRA3 RNA adaptor during 3' adaptor ligation as the unique molecular index (UMI). Reads with the same UMI were collapsed into one. We selected fragments longer than 140bp from the PCR-amplified library. Size-selected libraries were pooled and sequenced (2x75bp) on the Illumina NextSeq 500 platform according to manufacturer's instructions to a depth of ~100 million/library.

Single-molecule tracking—CTCF-AID-HaloTag+TIR cells were labeled with JF 549 conjugated HALO-ligand for 30min, followed by 3 washes with warm PBS. Cells were resuspended in warm phenol red-free media containing 5ug/ml Hoechst 33342 (bisBenzimide H 33342 trihydrochloride, Sigma-Aldrich, ref 14533) and seeded on glass bottom wells (Thermal Fisher Scientific, 150682) that had been pre-coated with poly-D-Lysine (1:10) (Millipore, A-003-E) overnight at 4°C.

Imaging was performed using temperature-controlled Oxford Nanoimager microscope (ONI) equipped with four lasers at 405 nm, 470 nm, 561 nm and 647 nm wavelength, a high numerical aperture (NA) objective (100X, 1.4NA oil) and a high quantum efficiency (QE) scientific complementary metal-oxide semiconductor (sCMOS) detector. This camera and objective combination provide an effective pixel size of 116 nm. The microscope is also equipped with a perfect focus system (PFS) that maintains the focus with nanometer precision during the acquisition.

To avoid any misinterpretation due to the motion of cells over the acquisition times (250 s), a snapshot of the chromatin-stained nucleus (Hoechst 33342) was taken before and after the SMT acquisition. We used the 405 nm laser at very low power with 2 s of camera integration time, to avoid stressing the live cells. Only those cells with minimal change in nuclear shape and position were considered for the analysis. Those snapshots allow us to segment the trajectories and only analyze those ones that are inside the cell nuclei.

We tuned the fluorophore concentration to achieve a sparse subset of labeled molecules that are not spatially overlapping, similar number of molecules per frame and similar SNR values for the localizations. We used 5 pM for 0h condition and of 100 nM for 24h condition. For both conditions, we illuminated the samples with the 561 nm laser (Nanoimager) and used HILO illumination (Tokunaga et al., 2008) for enhancing the SNR of the localizations. We acquired 500 frames with a camera exposure time of 500 ms (camera frame rate of 2 Hz), of a 50 × 80 μm field of view. We additionally acquired 700 frames with lower laser power with results shown in Figures S4E and S4F.

It is known in the field of SMT, and we have observed it in our experiments, that residence time measurements are sensitive to the experimental conditions. In particular, photobleaching, focal plane, and non-homogeneous illumination can have a strong impact on the estimated absolute values. Thus, we performed the experiments by pairs of replicates imaged on the same day under the same conditions (see Figure 4C).

QUANTIFICATION AND STATISTICAL ANALYSIS

ChIP-seq peak calling—Bowtie 1.1.0 was used to align sequences (Langmead and Salzberg, 2012) to the mm9 reference genome. Reads with more than one mismatch or multiple alignments were excluded. Significantly enriched regions were called using MACS2 version 2.1.0 (Zhang et al., 2008) with the following parameters: $p = 10^{-5}$, $\text{extsize} = 300$ and $\text{local lambda} = 100,000$ using whole-cell extract input controls. Reads for the bigwigs were RPM normalized.

We subsequently pooled peaks from all ChIP-seq libraries together to obtain a full list of CBSs. If the distance between two peak midpoints was $\geq 250\text{bp}$, they were merged as one by bedtools (Quinlan, 2014). All merged peaks were further standardized to 500bp. Further, peaks located at regions known to be enriched with non-specific binding were removed (Xiang et al., 2019).

Adjusting variation of local background in ChIP-seq—We normalized the read counts of the IP sample to the corresponding IgG control sample so that the background signals in the IP sample and the control sample were comparable. This step was achieved by using bamCompare in deeptools with the option `scaleFactorsMethod` set to SES (Diaz et al., 2012). The SES is a commonly used method to normalize IP sample against control. It first identifies the non-peak region in the ChIP-seq data. It then calculates a scaling factor to equalize the average read counts in the nonpeak regions. The variation of the local background signal in IP sample was adjusted by taking the ratio between the normalized read counts per bp in IP sample $\lambda_{obs,i}$ and the MACS background read counts per bp in IgG control sample $\lambda_{local,i}$.

$$RC_i = \lambda_{obs,i} - \lambda_{local,i}$$

The RC_i is used as the ChIP-seq signal hereafter.

Normalizing ChIP-seq against reference ChIP-qPCR data—We recently found that ChIP-seq data could be retroactively normalized using a panel of ChIP-qPCR data to overcome global bias as a result of ChIP-seq library scaling (Behera et al., 2019). We adopted a similar strategy here with 22 genomic regions of various binding intensities and degradation dynamics selected as our reference region (Figure S1C; Table S1). Diffbind was used to call peaks from CTCF ChIP-seq data at 6 time points after auxin treatment (Stark and Brown, 2017). We first scaled ChIP-qPCR results up to levels comparable to those in ChIP-seq using the following formula:

$$qPCR_{j, scale_up} = (qPCR_{j, raw} - M_qPCR) / SD_qPCR * SD_RC + M_RC,$$

where M_qPCR and SD_qPCR are the mean and the standard deviation of qPCR signals of the reference regions, M_RC and SD_RC of corresponding ChIP-seq read counts (RC). We then applied a regression model (Shao et al., 2012; Xiang et al., 2020) to calculate the scaling factor β .

$$qPCR_{scale_up} = \beta \times RC,$$

We did not include an intercept in this model as this would artificially increase the background signals for some regions.

Factor β was then used to normalize all ChIP-seq libraries.

$$RC_{i,qPCRnorm} = \beta \times RC_i,$$

where the RC_i is the raw ChIP-seq signal in CBS i , the $RC_{i,qPCRnorm}$ is the normalized ChIP-seq signal. The performance of this normalization strategy was evaluated by the R^2 between biological replicates and showed significant improvement (data not shown).

Unsupervised clustering of CBSs based on retention dynamics—To cluster CBSs based on their signal trend after auxin treatment, we applied the Gaussian Mixture Modeling (GMM) for Model-Based Clustering (Mclust) (Medvedovic et al., 2004). We decided to use this approach because the CBS signals at different time points usually correlate with each other. Compared with other commonly used unsupervised clustering methods such as K-means and hierarchical clustering method, the multivariate Gaussian Mixture Model used in this approach can incorporate the signal correlation into the clustering step.

When directly using the signal strength to cluster the CBSs, the average signal intensity usually becomes the main factor to decide the output clusters. To address this issue, we first used DESeq2 to transform qPCR normalized read counts to the *Wald* statistics (Love et al., 2014).

$$Wald_Stat_{i,t} = \frac{\log FC_{i,t}}{\log FC_SE_{i,t}}$$

where the $\log FC_{i,t}$ is the log fold-change between the read counts at t hour and the read counts at 0 hour, the $\log FC_SE_{i,t}$ is the standard error of the log fold-change in CBS i . This transformation can first change the signal intensity of CBSs to the signal difference. It can also make the right-skewed CTCF signal become similar to the normal distribution, which better fits the assumption of the Mclust model.

To ensure reproducibility, we ran Mclust on the *Wald* statistics 30 times. The number of clusters was first determined by a Bayesian Information Criterion (BIC). In each round, the clusters were ranked based on their average signal at 0 hour. CBSs consistently grouped into the same clusters (15 rounds) were kept as peaks with robust clusters. The rest were “rescued” by assigning them to the closest reproducible clusters (Xiang et al., 2018).

Modeling degradation dynamics and clustering CBSs based on an exponential decay model—To model the degradation dynamics of CBSs, we applied a canonical exponential decay (ED) model to fit the signals of CBSs across all time points. We

chose this model because most of the CBSs follow an exponential decay pattern by calculating R^2 for the exponential model fitting (Figure S1D; more details below). From a time-series data, the ED model incorporates two parameters, an initial signal at 0 hour ($RC_{i,0}$) and a decay rate ($-\lambda$) that reflects retention dynamics. For a CBS i , we first learned the $RC_{i,0}$ from the data and then the ED model can be written as formula (1).

$$RC_{i,t} = RC_{i,0} \times e^{-\lambda \times t}, \quad (1)$$

where $RC_{i,t}$ is the qPCR normalized read counts in CBS i at time point t . After a log transformation, the $-\lambda$ can be learned by a linear model.

$$\log(RC_{i,t} + \epsilon) = -\lambda \times t + \log(RC_{i,0} + \epsilon), \quad (2)$$

where the coefficient of the linear model is the decay rate $-\lambda$, the ϵ is an added small number to avoid log transformation of zero values. The $-\lambda$ and the corresponding R^2 were calculated by a *lm* function in *R* package. We observed for the majority of the CBSs, decay rate, $-\lambda$, assumed a linear relationship with the initial binding intensity, $\log(RC_{i,0} + \epsilon)$ (Figure 1D). This indicates that these CBSs were close to being completely degraded after 4h. We show our proof in the next section below. Interestingly, many CBSs had slower rates of decay as shown by the points above the linear zone. To separate the more persistent CBSs from the rest, we used a locally estimated scatterplot smoothing (LOESS) regression model to fit the data in an iteratively matter. For each iteration, CBSs with significantly higher $-\lambda$ ($p < 1e^{-3}$) were removed from the model fitting. When the iteratively fitted LOESS model converged, we used the local average $-\lambda$ and the 3.1 ($p = 1e^{-3}$) standard deviation to identify the most persistent sites (Cluster III). We observed that the initial signal intensity of the rapidly degraded CBSs showed a bimodal distribution (Figure 1D). We thus applied Gaussian mixture model to the signal intensities, stratifying the rapidly degraded CBSs into two clusters (Clusters I and II).

Proof that the linear relationship between decay rate and initial binding intensities was due to complete degradation by 4h—Let RC_0 and RC_{t4} denote the signals in a certain CBS at 0h and 4h, and $-\lambda$ denote decay rate. When the signals after auxin treatment follows an exponential decay model, the RC_0 and $-\lambda$ will have the following relationship:

$$RC_{t4} = RC_0 \times e^{-\lambda \times t4} \quad (1)$$

$$-\lambda = \frac{\log(RC_{t4} + 1) - \log(RC_0 + 1)}{t4} \quad (2)$$

For the non-persistent CBSs, the majority of signal will be lost at the second time point $t4$ (4h). Thus, $\log(RC_{t4} + 1)$ will become close to 0.

$$\log(RC_{t4} + 1) \approx 0 \quad (3)$$

$$-\lambda = \frac{1}{t4} \times \log(RC_{t0} + 1) \quad (4)$$

Since $t4$ is a constant, there is a linear relationship between the decay rate $-\lambda$ and the initial signal intensity $\log(RC_{t0})$.

Interrogate sequence features of CBSs—We analyzed sequence features underlying CBSs of each cluster. We investigated enriched motifs within each clusters using MEME-ChIP (Machanick and Bailey, 2011). We further measured the similarity of sequences in each cluster to the canonical CTCF motif using FIMO (Grant et al., 2011). FIMO computes a p value for a given sequence based on its similarity at each position to the reference sequence. For each peak, we chose the highest $-\log_{10}$ p value as its FIMO score, which ranged from 1.8 to 10.6 in our datasets. We compared FIMO scores between clusters by two-sample t-test.

PRO-seq

Alignment: Raw data of each PRO-seq library included 98 to 169 millions of 75-base long paired end reads. FLASH2 (Mago and Salzberg, 2011) was used to merge overlapping read pairs originated from short RNA fragments. Consensus reads of merged pairs, which accounted for about 78%–85% of total reads and had 37-51 bases of average length among libraries, were saved as single end reads in a new fastq file. Unmerged read pairs were saved in their original format as two fastq files. Seqtk (<https://github.com/lh3/seqtk/blob/master/README.md>) was run to trim the 6-base UMI from the beginning of each single end read and the first read of the paired end reads.

To align trimmed reads, reference genomes, UCSC mouse mm9 and fly dm6, were downloaded from iGenomes (https://support.illumina.com/sequencing/sequencing_software/igenome.html). The two genomes were combined into a hybrid genome. Single and paired end reads were separately aligned to the hybrid genome using BWA-MEM (Li and Durbin, 2010). Only primary alignments with mapq score ≥ 20 and without INDEL were reported and saved in bam files. Duplicated reads with the same UMI and aligned to the same genomic loci were removed. They accounted for 1.4%–7.0% of aligned reads among all libraries. Overall, 43%–58% of the total reads were aligned, with 1.4 to 3.8% of them aligned to the fly genome. The percentage of fly transcripts were similar across all libraries, indicating that there was no significant global transcriptomic changes after CTCF depletion. Reads exclusively aligned to fly and mouse genomes were split and separately processed through the following steps and single and paired end reads aligned to the same genome were combined into the same bam files.

Transcription quantification: Reads aligned to reference genome were mapped to known RefSeq transcripts and genes, including both of their exons and introns, for quantitative analysis. Reads mapped to multiple genes were not counted and those mapped to sense and antisense strands of the same genes were counted separately. The dREG program was run to identify divergent transcription patterns, a feature characteristic of active TSSs and

enhancers (Danko et al., 2015). Active genes were defined as transcribing in a divergent manner at the promoter, as recognized by dREG program (Mahat et al., 2016a), and with GB read counts exceeding a pre-determined threshold (see below for threshold setting).

RefSeq TSSs within 50bp of each other were combined into unique TSS loci, and only TSS loci not overlapping with any other genes or sharing a gene with other TSSs were selected. An arbitrary window of -50 to $+150$ bp relative to Refseq-annotated TSS was used to represent transcription initiation. A window of $+200$ bp relative to TSS to -500 bp relative to annotated TES was selected to represent transcription in the gene body. We observed length-adjusted GB read counts to follow a bimodal distribution. The lower limit of the right population was used as a threshold to call active genes. Differential expression analysis was analyzed using paired DESeq2 method (Love et al., 2014). Significant upregulation was identified as $FDR < 0.05$ & fold-change > 2 ; significant downregulation as $FDR < 0.05$ & fold-change < 0.5 .

TSS-proximal pausing index calculation: Read counts of TSSs and GBs were normalized and \log_2 -transformed separately using the Rlog method. Normalized GB data were further adjusted by effective lengths as read densities. At least 6 total TSS reads and 12 gene body reads were required to be included in downstream analysis to robustly quantify pause release dynamics. Pausing index was calculated as $\log_2(\text{TSS}/\text{GB read densities})$. The paired Limma method was used to test differences in PIs between conditions. Limma was used instead of DESeq as the data were not integers. Significant upregulation was identified as $FDR < 0.05$ & fold-change > 0 ; significant downregulation as $FDR < 0.05$ & fold-change < 0 .

Intragenic PRO-seq quantification: The canonical PWM (point-weighted matrix), MA0139.1, of CTCF binding was downloaded from Jasper (<http://jaspar.genereg.net/matrix/MA0139.1/>). Motifs matching this PWM were searched in mm9 and scored by the matchPWM method (Wasserman and Sandelin, 2004). About 87% of the CTCF binding sites had at least one CTCF motif scored higher than 75 within 100 bases around their center.

A total of 6926 CTCF binding sites were located within the gene bodies of 3680 actively transcribed genes. Each of these binding sites was paired with CTCF motif with high score (mean = 82.3) in its proximity (0 to 250 bases, median = 27 bases). About half of these motifs had the same orientation as the direction of transcription. For the control group, 11626 motifs with similar matching scores, but no CTCF occupancy were selected from the same set of active genes. Intragenic pausing index at CTCF motifs was calculated as the \log_2 -ratio of reads mapped to $[+25, +75]$ bases over reads mapped to $[-50, -1]$ bases relative to the motifs. Group differences in PI were tested by paired Limma. Relative enrichment (Figure S3C) is calculated as the number of 3' reads at each base divided by the total number of 3' reads mapped to the 19bp motif. For example, if there were 100 3' reads mapped to the motif and 10 reads at the 6th base, the relative enrichment at this position is $19 \cdot (10/100) = 1.9$.

SMT

Tracking: First, the images were segmented based on the nuclei regions obtained from Hoechst 33342 fluorescence signal. The tracking step was performed individually nucleus by nucleus. For both the localization and tracking steps, we used TrackMate software (Tinevez et al., 2017). For the localization, we selected the LoG detector with sub-pixel localization, an intensity threshold of 20, which minimizes false positives, and an estimated blob diameter of 580 nm (corresponding to 5 pixels). For the tracking, we used the Simple LAP tracker algorithm with a maximum jump from frame-to-frame of 400 nm, a maximum closing gap of 4 frames and a maximum closing jump of 200 nm. The lists of trajectories were saved for each nucleus as an .xml file. The number of trajectories in each replicate ranged from 2459 to 12056.

Residence times: We calculated the residence times of chromatin-bound CTCF from the SMT trajectories. We considered that a one-frame localization is a binding event that last 500 ms. We first converted the distribution of track durations into the survival fraction of molecules defined by 1-CDF (1 – Cumulative Distribution Function) of the track lengths. Then, we fitted a two-component exponential decay function:

$$F(t) = f \cdot e^{-k_1 \cdot t} + (1 - f) \cdot e^{-k_2 \cdot t},$$

where f is the fraction belonging to each population, k_1 the short-live component associated with unspecific chromatin binding and k_2 the long-live component associated with specific chromatin binding. k_1 and k_2 are rate constants with s^{-1} units. In addition, we estimated the photobleaching rate by fitting an exponential decay function to the evolution of the number of localizations over time during the experiment (Mazza et al., 2012):

$$N(t) = A \cdot e^{-k_b \cdot t}$$

We used that value to perform a correction to the measured residence times. The corrected long-lived residence times can be obtained from the following relation:

$$k_2 = k_{corrected} + k_b.$$

where k_2 is the dissociation rate constant estimated directly from the experimental data, k_b is the photobleaching kinetics rate and $k_{corrected}$ is the dissociation rate after correction. Note that k is in s^{-1} units and the residence times are inversely proportional.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Katherine Alexander, Payel Sen, Dig B. Mahat, and Kyle A. Nilson for advice on PRO-seq; Danuta Jeziorska for advice on cell seeding for SMT; Anders Sejr Hansen for helpful SMT discussions; and Jonathan Lerner and Ricardo Linares for assistance with imaging. We are grateful to the Hardison and Blobel laboratories for

insightful discussions. This work was supported by NIH grants R01 DK054937 to G.A.B., R24 DK106766 to G.A.B. and R.C.H., and R01GM121613 to R.C.H.

REFERENCES

- Agarwal H, Reisser M, Wortmann C, and Gebhardt JCM (2017). Direct Observation of Cell-Cycle-Dependent Interactions between CTCF and Chromatin. *Biophys. J* 112, 2051–2055. [PubMed: 28487148]
- Behera V, Evans P, Face CJ, Hamagami N, Sankaranarayanan L, Keller CA, Giardine B, Tan K, Hardison RC, Shi J, and Blobel GA (2018). Exploiting genetic variation to uncover rules of transcription factor binding and chromatin accessibility. *Nat. Commun* 9, 782. [PubMed: 29472540]
- Behera V, Stonestrom AJ, Hamagami N, Hsiung CC, Keller CA, Giardine B, Sidoli S, Yuan Z-F, Bhanu NV, Werner MT, et al. (2019). Interrogating Histone Acetylation and BRD4 as Mitotic Bookmarks of Transcription. *Cell Rep.* 27,400–415.e5. [PubMed: 30970245]
- Belotserkovskii BP, Liu R, Tornaletti S, Krasilnikova MM, Mirkin SM, and Hanawalt PC (2010). Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc. Natl. Acad. Sci. USA* 107, 12816–12821. [PubMed: 20616059]
- Boyle AP, Song L, Lee B-K, London D, Keefe D, Birney E, Iyer VR, Crawford GE, and Furey TS (2011). High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Res.* 21, 456–464. [PubMed: 21106903]
- Busslinger GA, Stocsits RR, van der Lelij P, Axelsson E, Tedeschi A, Galjart N, and Peters J-M (2017). Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* 544, 503–507. [PubMed: 28424523]
- Chen L, Chen J-Y, Zhang X, Gu Y, Xiao R, Shao C, Tang P, Qian H, Luo D, Li H, et al. (2017). R-ChIP Using Inactive RNase H Reveals Dynamic Coupling of R-loops with Transcriptional Pausing at Gene Promoters. *Mol. Cell* 68, 745–757.e5. [PubMed: 29104020]
- Chernukhin I, Shamsuddin S, Kang SY, Bergström R, Kwon Y-W, Yu W, Whitehead J, Mukhopadhyay R, Docquier F, Farrar D, et al. (2007). CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol. Cell. Biol* 27, 1631–1648. [PubMed: 17210645]
- Cong L, and Zhang F (2015). Genome engineering using CRISPR-Cas9 system. *Methods Mol. Biol* 1239, 197–217. [PubMed: 25408407]
- Core LJ, Waterfall JJ, and Lis JT (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845–1848. [PubMed: 19056941]
- Danko CG, Hyland SL, Core LJ, Martins AL, Waters CT, Lee HW, Cheung VG, Kraus WL, Lis JT, and Siepel A (2015). Identification of active transcriptional regulatory elements from GRO-seq data. *Nat. Methods* 12, 433–438. [PubMed: 25799441]
- Diaz A, Nellore A, and Song JS (2012). CHANCE: comprehensive software for quality control and validation of ChIP-seq data. *Genome Biol.* 13, R98. [PubMed: 23068444]
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, and Ren B (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. [PubMed: 22495300]
- Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, and Mirny LA (2016). Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* 15, 2038–2049. [PubMed: 27210764]
- Ghirlando R, and Felsenfeld G (2016). CTCF: making the right connections. *Genes Dev.* 30, 881–891. [PubMed: 27083996]
- Grant CE, Bailey TL, and Noble WS (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. [PubMed: 21330290]
- Gressel S, Schwab B, Decker TM, Qin W, Leonhardt H, Eick D, and Cramer P (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. *eLife* 6, R106.
- Grimm JB, English BP, Chen J, Slaughter JP, Zhang Z, Revyakin A, Patel R, Macklin JJ, Normanno D, Singer RH, et al. (2015). A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat. Methods* 12, 244–250. [PubMed: 25599551]

- Grimm JB, Klein T, Kopek BG, Shtengel G, Hess HF, Sauer M, and Lavis LD (2016). Synthesis of a Far-Red Photoactivatable Silicon-Containing Rhodamine for Super-Resolution Microscopy. *Angew. Chem. Int. Ed.* 55, 1723–1727. [PubMed: 26661345]
- Gu B, Posfai E, and Rossant J (2018). Efficient generation of targeted large insertions by microinjection into two-cell-stage mouse embryos. *Nat. Biotechnol* 36, 632–637. [PubMed: 29889212]
- Hansen AS, Pustova I, Cattoglio C, Tjian R, and Darzacq X (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* 6, 2848.
- Heinz S, Texari L, Hayes MGB, Urbanowski M, Chang MW, Givarkes N, Rialdi A, White KM, Albrecht RA, Pache L, et al. (2018). Transcription Elongation Can Affect Genome 3D Structure. *Cell* 174, 1522–1536.e1522. [PubMed: 30146161]
- Hyle J, Zhang Y, Wright S, Xu B, Shao Y, Easton J, Tian L, Feng R, Xu P, and Li C (2019). Acute depletion of CTCF directly affects MYC regulation through loss of enhancer-promoter looping. *Nucleic Acids Res.* 47, 6699–6713. [PubMed: 31127282]
- Kaaij LJT, Mohn F, van der Weide RH, de Wit E, and Bühler M (2019). The ChAHP Complex Counteracts Chromatin Looping at CTCF Sites that Emerged from SINE Expansions in Mouse. *Cell* 178, 1437–1451.e14. [PubMed: 31491387]
- Kang H, and Lieberman PM (2011). Mechanism of glycyrrhizic acid inhibition of Kaposi's sarcoma-associated herpesvirus: disruption of CTCF-cohesin-mediated RNA polymerase II pausing and sister chromatid cohesion. *J. Virol* 85, 11159–11169. [PubMed: 21880767]
- Kapranov P, Cheng J, Dike S, Nix DA, Duttagupta R, Willingham AT, Stadler PF, Hertel J, Hackermüller J, Hofacker IL, et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316, 1484–1488. [PubMed: 17510325]
- Khoury A, Achinger-Kawecka J, Bert SA, Smith GC, French HJ, Luu P-L, Peters TJ, Du Q, Parry AJ, Valdes-Mora F, et al. (2020). Constitutively bound CTCF sites maintain 3D chromatin architecture and long-range epigenetically regulated domains. *Nat. Commun* 11, 54. [PubMed: 31911579]
- Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenko VV, and Ren B (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128, 1231–1245. [PubMed: 17382889]
- Kubo N, Ishii H, Xiong X, Bianco S, Meitinger F, Hu R, Hocker JD, Conte M, Gorkin D, Yu M, et al. (2021). Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat. Struct. Mol. Biol* 459, 108.
- Kwak H, Fuda NJ, Core LJ, and Lis JT (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* 339, 950–953. [PubMed: 23430654]
- Laitem C, Zaborowska J, Tellier M, Yamaguchi Y, Cao Q, Egloff S, Handa H, and Murphy S (2015). CTCF regulates NELF, DSIF and P-TEFb recruitment during transcription. *Transcription* 6, 79–90. [PubMed: 26399478]
- Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. [PubMed: 22388286]
- Letting DL, Chen Y-Y, Rakowski C, Reedy S, and Blobel GA (2004). Context-dependent regulation of GATA-1 by friend of GATA-1. *Proc. Natl. Acad. Sci. USA* 101, 476–481. [PubMed: 14695898]
- Li H, and Durbin R (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589–595. [PubMed: 20080505]
- Lorenz R, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, and Hofacker IL (2011). ViennaRNA Package 2.0. *Algorithms Mol. Biol* 6, 26. [PubMed: 22115189]
- Loukinov DI, Pugacheva E, Vatolin S, Pack SD, Moon H, Chernukhin I, Mannan P, Larsson E, Kanduri C, Vostrov AA, et al. (2002). BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc. Natl. Acad. Sci. USA* 99, 6806–6811. [PubMed: 12011441]
- Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550. [PubMed: 25516281]
- Machanic P, and Bailey TL (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 27, 1696–1697. [PubMed: 21486936]

- Mago T, and Salzberg SL (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. [PubMed: 21903629]
- Mahat DB, Salamanca HH, Duarte FM, Danko CG, and Lis JT (2016a). Mammalian Heat Shock Response and Mechanisms Underlying Its Genome-wide Transcriptional Regulation. *Mol. Cell* 62, 63–78. [PubMed: 27052732]
- Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, Waters CT, Munson K, Core LJ, and Lis JT (2016b). Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc* 11, 1455–1476. [PubMed: 27442863]
- Manzo C, and Garcia-Parajo MF (2015). A review of progress in single particle tracking: from methods to biophysical insights. *Rep. Prog. Phys* 78, 124601. [PubMed: 26511974]
- Mayer A, di Iulio J, Maleri S, Eser U, Vierstra J, Reynolds A, Sandstrom R, Stamatoyannopoulos JA, and Churchman LS (2015). Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* 161, 541–554. [PubMed: 25910208]
- Mazza D, Abernathy A, Golob N, Morisaki T, and McNally JG (2012). A benchmark for chromatin binding measurements in live cells. *Nucleic Acids Res.* 40, e119. [PubMed: 22844090]
- Medvedovic M, Yeung KY, and Bumgarner RE (2004). Bayesian mixture model based clustering of replicated microarray data. *Bioinformatics* 20, 1222–1232. [PubMed: 14871871]
- Mischo HE, Gómez-González B, Grzechnik P, Rondón AG, Wei W, Steinmetz L, Aguilera A, and Proudfoot NJ (2011). Yeast Sen1 helicase protects the genome from transcription-associated instability. *Mol. Cell* 41, 21–32. [PubMed: 21211720]
- Mizuguchi T, Fudenberg G, Mehta S, Belton J-M, Taneja N, Folco HD, FitzGerald P, Dekker J, Mirny L, Barrowman J, and Grewal SIS (2014). Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature* 516, 432–435. [PubMed: 25307058]
- Morawska M, and Ulrich HD (2013). An expanded tool kit for the auxin-inducible degron system in budding yeast. *Yeast* 30, 341–351. [PubMed: 23836714]
- Nakahashi H, Kieffer Kwon KR, Resch W, Vian L, Dose M, Stavreva D, Hakim O, Pruett N, Nelson S, Yamane A, et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. *Cell Rep.* 3, 1678–1689. [PubMed: 23707059]
- Nishimura K, Fukagawa T, Takisawa H, Kakimoto T, and Kanemaki M (2009). An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* 6, 917–922. [PubMed: 19915560]
- Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385. [PubMed: 22495304]
- Nora EP, Goloborodko A, Valton A-L, Gibcus JH, Uebersohn A, Abdennur N, Dekker J, Mirny LA, and Bruneau BG (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* 169, 930–944.e22. [PubMed: 28525758]
- Paredes SH, Melgar MF, and Sethupathy P (2013). Promoter-proximal CCCTC-factor binding is associated with an increase in the transcriptional pausing index. *Bioinformatics* 29, 1485–1487. [PubMed: 23047559]
- Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, et al. (2008). Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132, 422–433. [PubMed: 18237772]
- Peña-Hernández R, Marques M, Hilmi K, Zhao T, Saad A, Alaoui-Jamali MA, del Rincon SV, Ashworth T, Roy AL, Emerson BM, and Witcher M (2015). Genome-wide targeting of the epigenetic regulatory protein CTCF to gene promoters by the transcription factor TFII-I. *Proc. Natl. Acad. Sci. USA* 112, E677–E686. [PubMed: 25646466]
- Phillips JE, and Corces VG (2009). CTCF: master weaver of the genome. *Cell* 137, 1194–1211. [PubMed: 19563753]
- Plasschaert RN, Vigneau S, Tempera I, Gupta R, Maksimoska J, Everett L, Davuluri R, Mamorstein R, Lieberman PM, Schultz D, et al. (2014). CTCF binding site sequence differences are associated with unique regulatory and functional trends during embryonic stem cell differentiation. *Nucleic Acids Res.* 42, 774–789. [PubMed: 24121688]

- Quinlan AR (2014). BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr. Protoc. Bioinformatics* 47, 11.12.1–11.12.34.
- R Core Team (2014). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing). <http://www.R-project.org/>.
- Rhee HS, and Pugh BF (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell* 147, 1408–1419. [PubMed: 22153082]
- Rubio ED, Reiss DJ, Welch PL, Disteché CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA, and Krumm A (2008). CTCF physically links cohesin to chromatin. *Proc. Natl. Acad. Sci. USA* 105, 8309–8314. [PubMed: 18550811]
- Ruiz-Velasco M, Kumar M, Lai MC, Bhat P, Solis-Pinson AB, Reyes A, Kleinsorg S, Noh K-M, Gibson TJ, and Zaugg JB (2017). CTCF-Mediated Chromatin Loops between Promoter and Gene Body Regulate Alternative Splicing across Individuals. *Cell Syst.* 5, 628–637.e6. [PubMed: 29199022]
- Saldaña-Meyer R, Rodríguez-Hernaez J, Escobar T, Nishana M, Jácome-López K, Nora EP, Bruneau BG, Tsigirgos A, Furlan-Magaril M, Skok J, and Reinberg D (2019). RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol. Cell* 76, 412–422.e5. [PubMed: 31522988]
- Sanborn AL, Rao SSP, Huang S-C, Durand NC, Huntley MH, Jewett AI, Bochkov ID, 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. Natl. Acad. Sci. USA* 112, E6456–E6465. [PubMed: 26499245]
- Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C, Brown GD, Marshall A, Flicek P, and Odom DT (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. *Cell* 148, 335–348. [PubMed: 22244452]
- Schneider CA, Rasband WS, and Eliceiri KW (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* 9, 671–675. [PubMed: 22930834]
- Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, and Sharp PA (2008). Divergent transcription from active promoters. *Science* 322, 1849–1851. [PubMed: 19056940]
- Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, and Cavalli G (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148, 458–472. [PubMed: 22265598]
- Shao Z, Zhang Y, Yuan G-C, Orkin SH, and Waxman DJ (2012). MAnorm: a robust model for quantitative comparison of ChIP-Seq data sets. *Genome Biol.* 13, R16. [PubMed: 22424423]
- 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. [PubMed: 21964334]
- Skourti-Stathaki K, Proudfoot NJ, and Gromak N (2011). Human senataxin resolves RNA/DNA hybrids formed at transcriptional pause sites to promote Xrn2-dependent termination. *Mol. Cell* 42, 794–805. [PubMed: 21700224]
- Stadhouders R, Thongjuea S, Andrieu-Soler C, Palstra R-J, Bryne JC, van den Heuvel A, Stevens M, de Boer E, Kockx C, van der Sloot A, et al. (2012). Dynamic long-range chromatin interactions control Myb proto-oncogene transcription during erythroid development. *EMBO J.* 31, 986–999. [PubMed: 22157820]
- Stark R, and Brown GD (2017). DiffBind: differential binding analysis of ChIP-Seq peak data (Bioconductor).
- Stonestrom AJ, Hsu SC, Jahn KS, Huang P, Keller CA, Giardine BM, Kadauke S, Campbell AE, Evans P, Hardison RC, and Blobel GA (2015). Functions of BET proteins in erythroid gene expression. *Blood* 125, 2825–2834. [PubMed: 25696920]
- Szlachta K, Thys RG, Atkin ND, Pierce LCT, Bekiranov S, and Wang Y-H (2018). Alternative DNA secondary structure formation affects RNA polymerase II promoter-proximal pausing in human. *Genome Biol.* 19, 89. [PubMed: 30001206]
- Thiecke MJ, Wutz G, Muhar M, Tang W, Bevan S, Malysheva V, Stocsits R, Neumann T, Zuber J, Fraser P, et al. (2020). Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers. *Cell Rep.* 32, 107929. [PubMed: 32698000]

- Tinevez J-Y, Perry N, Schindelin J, Hoopes GM, Reynolds GD, Laplantine E, Bednarek SY, Shorte SL, and Eliceiri KW (2017). TrackMate: An open and extensible platform for single-particle tracking. *Methods* 115,80–90. [PubMed: 27713081]
- Tokunaga M, Imamoto N, and Sakata-Sogawa K (2008). Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods* 5, 159–161. [PubMed: 18176568]
- Tome JM, Tippens ND, and Lis JT (2018). Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nat. Genet* 50, 1533–1541. [PubMed: 30349116]
- Traut TW (1994). Physiological concentrations of purines and pyrimidines. *Mol. Cell. Biochem* 140, 1–22. [PubMed: 7877593]
- Wasserman WW, and Sandelin A (2004). Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet* 5, 276–287. [PubMed: 15131651]
- Watts JA, Burdick J, Daigneault J, Zhu Z, Grunseich C, Bruzel A, and Cheung VG (2019). cis Elements that Mediate RNA Polymerase II Pausing Regulate Human Gene Expression. *Am. J. Hum. Genet* 105, 677–688. [PubMed: 31495490]
- Weiss MJ, Yu C, and Orkin SH (1997). Erythroid-cell-specific properties of transcription factor GATA-1 revealed by phenotypic rescue of a gene-targeted cell line. *Mol. Cell. Biol* 17, 1642–1651. [PubMed: 9032291]
- Wendt KS, 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. [PubMed: 18235444]
- Wickham H. (2016). *ggplot2* (Springer).
- Wutz G, Varnai C, Nagasaka K, Cisneros DA, Stocsits RR, Tang W, Schoenfelder S, Jessberger G, Muhar M, Hossain MJ, et al. (2017). Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* 36, 3573–3599. [PubMed: 29217591]
- Xiang G, Giardine B, An L, Sun C, Keller CA, Heuston E, Bodine D, Hardison RC, and Zhang Y (2018). Snapshot: clustering and visualizing epigenetic history during cell differentiation. *bioRxiv*, 291880.
- Xiang G, Keller CA, Heuston E, Giardine BM, An L, Wixom AQ, Miller A, Cockburn A, Lichtenberg J, Göttgens B, et al. (2019). An integrative view of the regulatory and transcriptional landscapes in mouse hematopoiesis. *bioRxiv* 20, 731729.
- Xiang G, Keller CA, Giardine B, An L, Li Q, Zhang Y, and Hardison RC (2020). S3norm: simultaneous normalization of sequencing depth and signal-to-noise ratio in epigenomic data. *Nucleic Acids Res.* 48, e43. [PubMed: 32086521]
- Yonaha M, and Proudfoot NJ (1999). Specific transcriptional pausing activates polyadenylation in a coupled in vitro system. *Mol. Cell* 3, 593–600. [PubMed: 10360175]
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, and Liu XS (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137. [PubMed: 18798982]
- Zhang H, Emerson DJ, Gilgenast TG, Titus KR, Lan Y, Huang P, Zhang D, Wang H, Keller CA, Giardine B, et al. (2019). Chromatin structure dynamics during the mitosis-to-G1 phase transition. *Nature* 576,158–162. [PubMed: 31776509]
- Zuin J, Dixon JR, van der Reijden MIJA, Ye Z, Kolovos P, Brouwer RWW, van de Corput MPC, van de Werken HJG, Knoch TA, van IJcken WFJ, et al. (2014). Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc. Natl. Acad. Sci. USA* 111, 996–1001. [PubMed: 24335803]

Highlights

- Marked variability in CTCF persistence on chromatin upon auxin-mediated degradation
- Sites associated with persistent CTCF binding tend to be architectural
- SMT reveals prolonged chromatin residence time at persistent CTCF binding sites
- RNA polymerase II stalling occurs at oriented CTCF motifs independently of bound CTCF

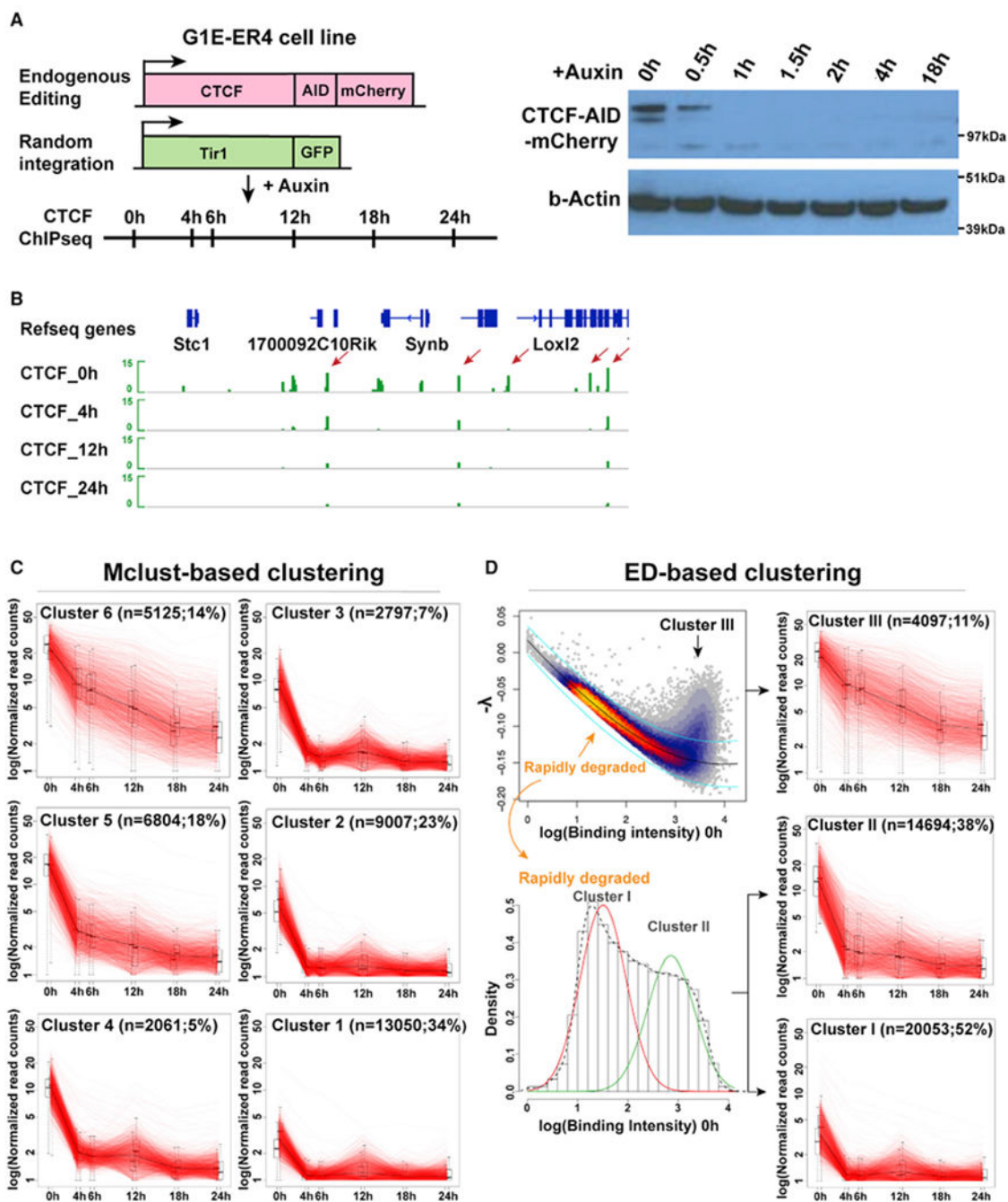


Figure 1. Highly variable CTCF persistence on chromatin following auxin-mediated degradation (A) Left: experimental design. Right: anti-CTCF western blot at indicated time points after auxin treatment. See also Figures S1A, S1B, and S4B and Table S4.

(B) A browser-track view of examples showing variability in CBS persistence. $n = 2$ biological replicates. Arrows highlight similarly enriched CBSs with different degradation dynamics.

(C) Line plot showing changes in binding enrichment of all CBSs grouped by Mclust-based clusters over time. Data were graphed as log-transformed normalized read counts under the peak.

(D) Left: illustration of ED-based clustering workflow. Right: line plot shows binding enrichment of ED-based clusters over time. Data were graphed as log-transformed normalized read counts under the peak.

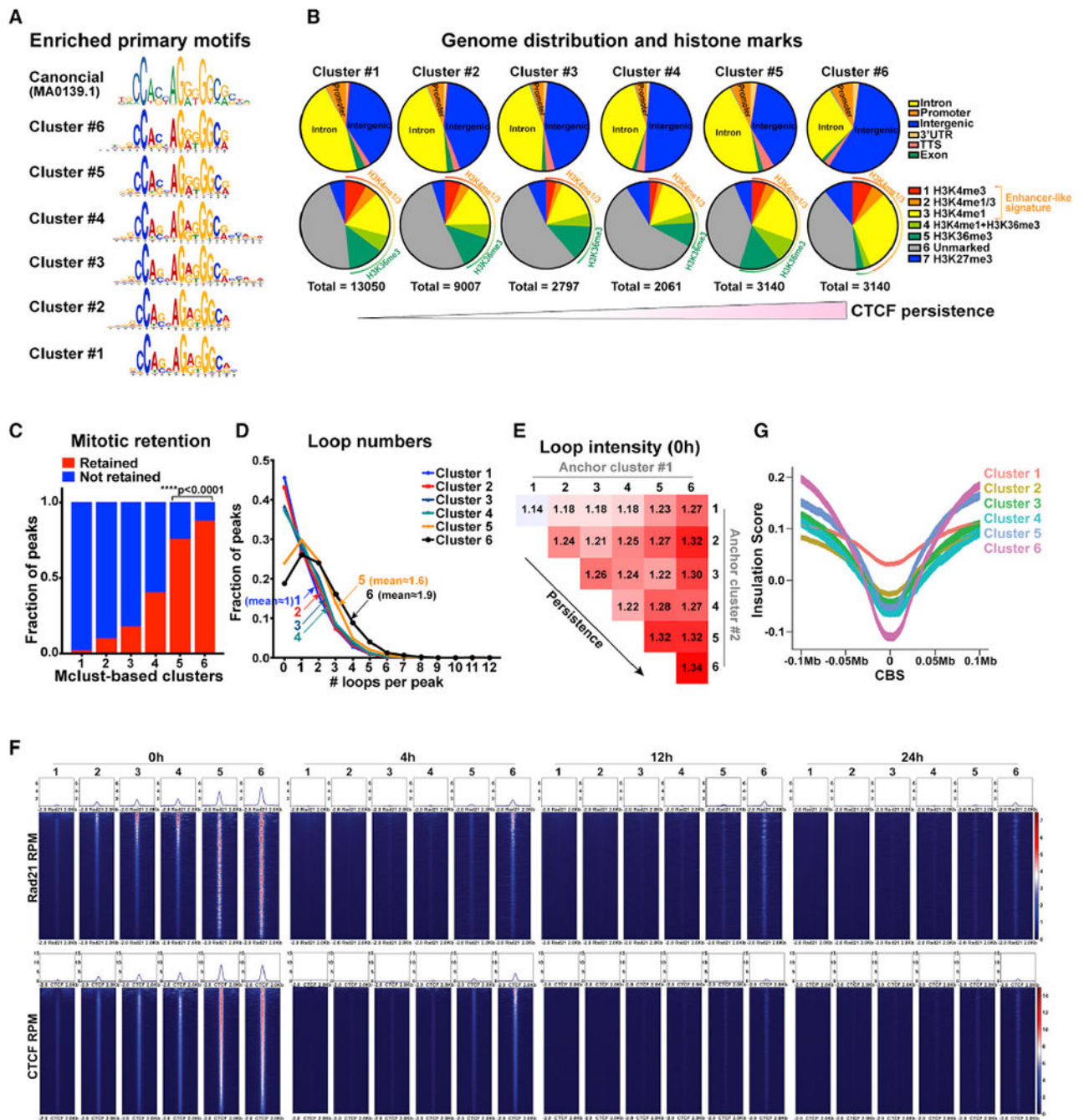


Figure 2. Characteristics of persistent CBSs

(A) Position weight matrix (PWM) of CTCF motif in Mclust-based clusters. See also Figure S2A.

(B) Genome distribution and histone marks of Mclust-based groups. See also Figure S2I.

(C) Percentage of each Mclust-based cluster retained on mitotic chromatin. p value was calculated from Chi-squared test. See also Figure S2J.

(D) Fraction of CBSs in each Mclust-based cluster engaging in indicated numbers of looping interactions. See also Figure S2M.

(E) Heatmap showing averaged loop intensity of structural loops stratified by anchor clusters (after intensity matching). See also Figure S2O.

(F) Row-linked heatmaps showing CTCF and Rad21 ChIP-seq RPMs after 0 h, 4 h, 12 h, and 24 h auxin treatment at CBSs over 4 kb genomic interval in 10-bp bins, grouped by Mclust-based clusters. See also Figure S2P.

(G) Mean (\pm SEM) IS of Mclust-based clusters centering on CBSs over 0.2-Mb genomic interval. See also Figure S2Q.

Clusters 5 and 6 were intensity matched.

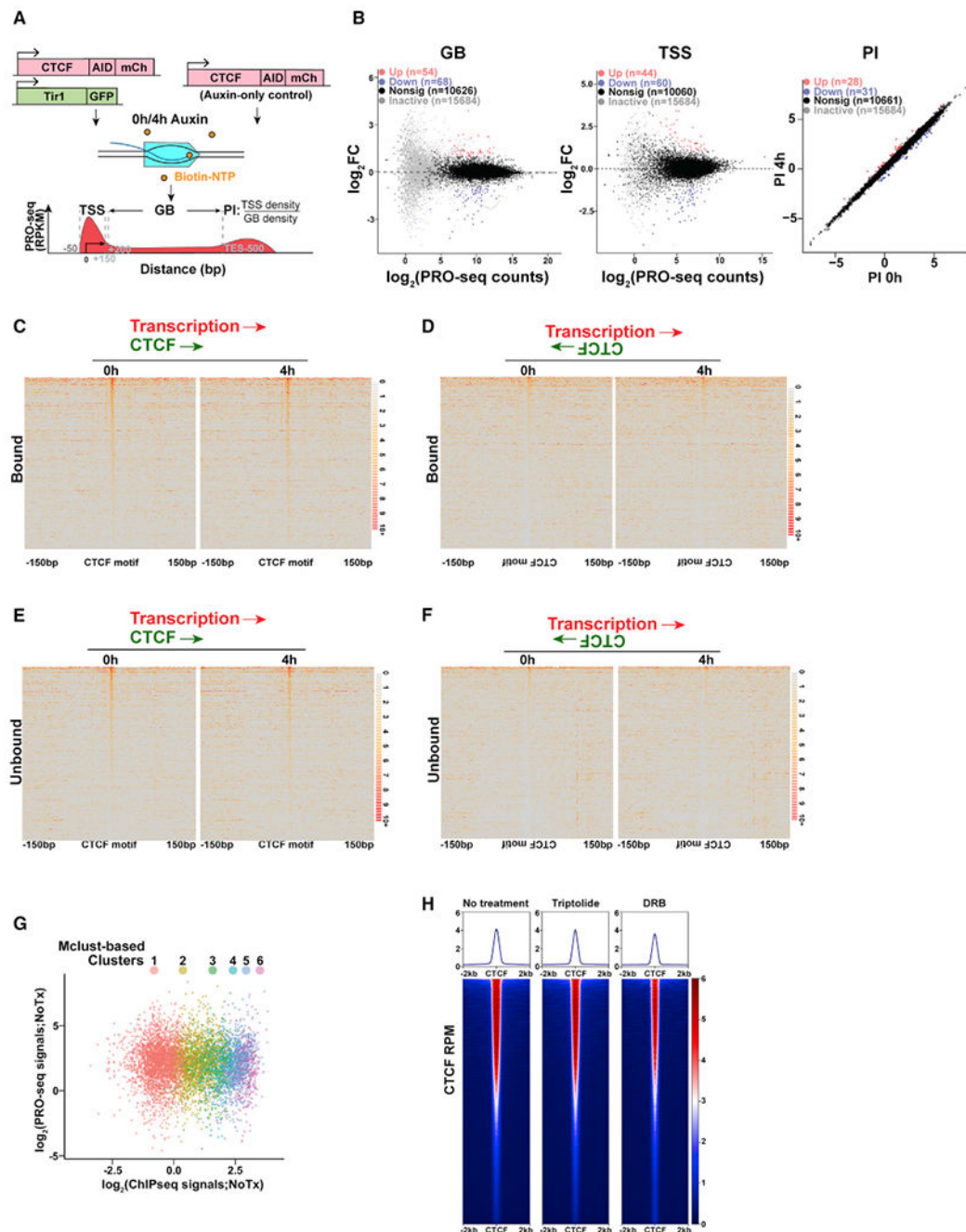


Figure 3. Effects of CTCF depletion on nascent transcription

(A) PRO-seq experimental and analytical strategy

(B) MA plots illustrating changes in GB, TSS, and PI.

(C) Row-linked heatmaps show 3' PRO-seq read counts centered on predicted CTCF motifs. Left: 0 h; right: 4 h. CBSs were oriented in the same direction as transcription. See also Figures S3E and S3F.

(D) Same analysis as (C), but at CBSs positioned in the opposite orientation as transcription. See also Figures S3E and S3F.

(E and F) Same analysis as (C) and (D), but at high-confidence CTCF motifs devoid of CTCF binding. See also Figures S3E and S3G.

(G) Scatterplot plotting CTCF binding intensity against PRO-seq signal counts over CBSs, with each Mclust-based cluster distinguished by color. NoTx, no auxin treatment. See also Figure S3J.

(H) Row-linked CTCF ChIP-seq heatmaps before and after triptolide and DRB treatment, ranked by mean intensity in the control group. All heatmaps were plotted with the same scale.

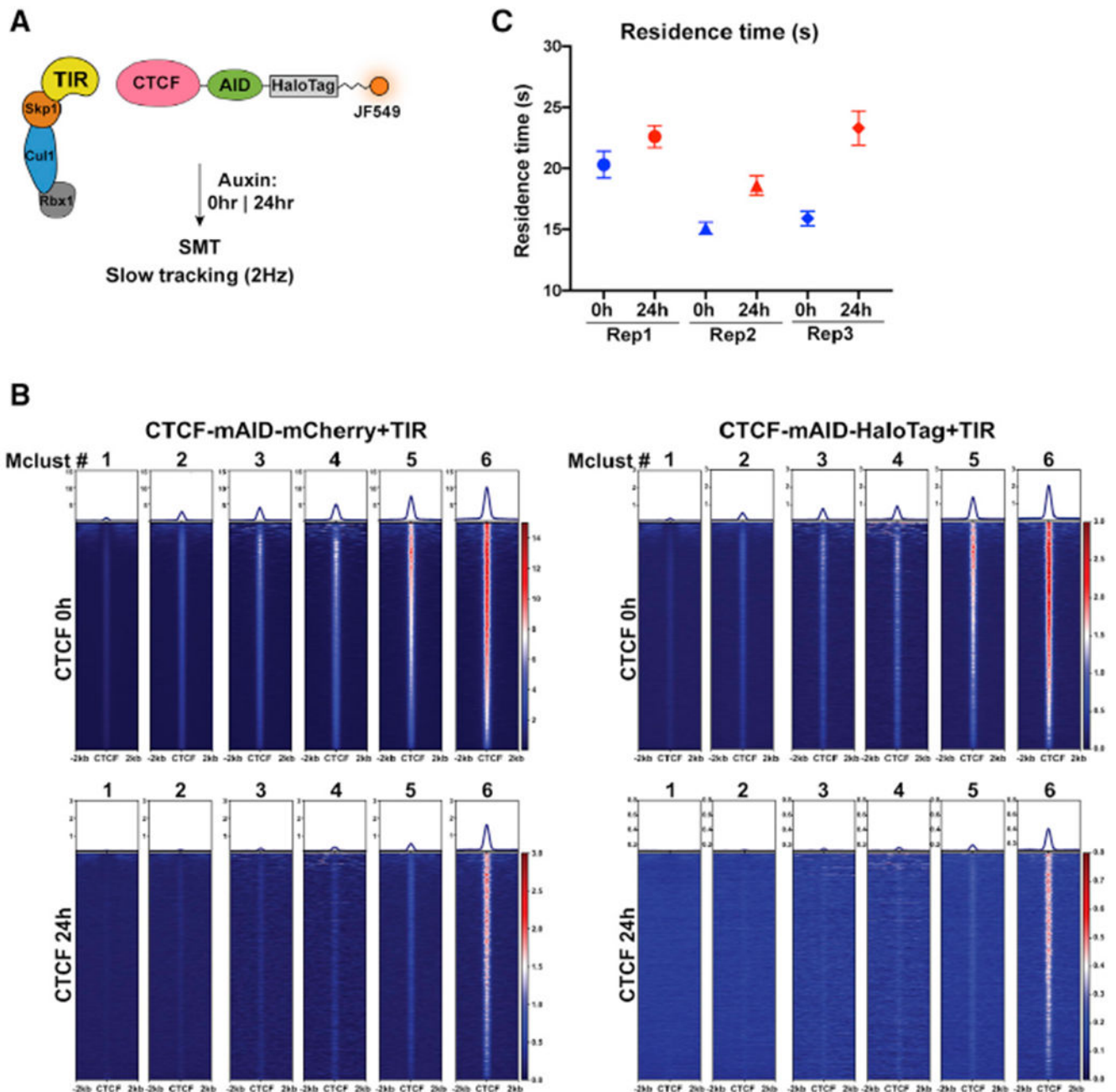


Figure 4. CBS persistence correlates with residence time on chromatin

(A) SMT experimental design.

(B) Row-linked heatmap showing CTCF ChIP-seq RPMs of all sites in CTCF-AID-mCherry+TIR and CTCF-AID-HaloTag+TIR cell lines before and after auxin treatment, grouped by Mclust-based clusters.

(C) Residence times before and after 24 h auxin treatment across three biological replicate pairs. Error bars denote 95% confidence interval. See also Figures S4C, S4D, and S4F.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-CTCF antibody, rabbit polyclonal	Millipore	07-729; RRID: AB_441965
Anti-mCherry antibody, rabbit polyclonal	Abcam	Ab167453; RRID: AB_2571870
Anti-Rad21 antibody, rabbit polyclonal	Abcam	Ab992; RRID: AB_2176601
Anti- β -Actin-Peroxidase antibody, mouse monoclonal	Millipore	A3854; RRID: AB_262011
Anti-Rpb1 NTD (D8L4Y) antibody, rabbit monoclonal	Cell Signaling	14958; RRID: AB_2687876
Anti-H3 antibody, rabbit polyclonal	Cell Signaling	9715; RRID: AB_2687876
Chemicals, peptides, and recombinant proteins		
Indole-3-acetic acid sodium salt	Sigma-Aldrich	I5148-2G
Protein A agarose beads	ThermoFisher	Cat#15918014
Protein G agarose beads	ThermoFisher	Cat#15920010
iScript Reverse Transcription Supermix	BioRad	Cat#1708841
Trizol	ThermoFisher	Cat#15596026
Power SYBR Green PCR Master Mix	ThermoFisher	Cat#4367660
Janelia Fluor® 549 HaloTag® Ligand	Promega	Cat#GA1110
Critical commercial assays		
Subcellular Protein Fractionation Kit	ThermoFisher	Cat#78840
QIAGEN PCR Purification Kit	QIAGEN	Cat#28106
QIAGEN RNeasy Kit	QIAGEN	Cat#74106
TruSeq ChIP Sample Preparation Kit	Illumina	Cat# IP 202-1012
Phusion High-Fidelity PCR Master Mix	ThermoFisher	Cat#F531S
NEBNext DNA Library Prep Master Mix for Illumina	New England BioLabs	Cat#E6040S
NEBNext Multiplex Oligos for Illumina Set1	New England BioLabs	Cat#E7335S
Cell Line Nucleofector Kit R	Lonza	Cat#VVCA-1001
In-Fusion® HD Cloning Plus	Clontech	Cat#638909
Micro Bio-Spin P-30 Gel Columns, Tris Buffer (RNase-free)	Bio-Rad	Cat#7326250
Deposited data		
CTCF ChIP-seq	Nora et al., 2017	GSE98671
CTCF ChIP-seq	Kubo et al., 2021	GSE94452
CTCF ChIP-seq	Khoury et al., 2020	GSE125641
CTCF (biotin) ChIP-seq	Nakahashi et al., 2013	GSE33819
Hi-C (late G1)	Zhang et al., 2019	GSE129997
Raw and processed sequencing data	This paper	GSE150418
Experimental models: cell lines		
G1E-ER4	Michell J. Weiss Lab	Weiss et al., 1997
CTCF-AID-mCherry+TIR	Gerd A. Blobel Lab	Zhang et al., 2019
Oligonucleotides		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ChIP-qPCR primers	This paper	Tables S1 and S2
RT-qPCR primers	This paper	Table S3
Myrip-A/T gRNA #1: TCCTGAAAATAAGACACCCC	This paper	NA
Myrip-A/T gRNA #2: CAGATATTAAGCATCCCAG	This paper	NA
Myb-CBS gRNA #1: TGACTATTGACTGCCCCCTG	This paper	NA
Myb-CBS gRNA #2: ACAAACCCCTCCCTCTCG	This paper	NA
CTCF sgRNA:GCATGATGGACCGGTGATGC	Zhang et al., 2019	N/A
Recombinant DNA		
pX330-GFP-sgCTCF (spCas9 with CTCF sgRNA)	Gerd A. Blobel Lab	Zhang et al., 2019
Ctf-Halo-mAID donor	Addgene	RRID: Addgene_113103
MigR1 guide RNA GFP	Gerd A. Blobel Lab	Stonestrom et al., 2015
EFS-Cas9-P2A-TagBFP	This paper	GenBank: MW079340
MigR1-osTIR1-9Myc-GFP	This paper	GenBank: MW079339
pKLV-U6gRNA-EF(BbsI)-PGKpuro2ABFP	Addgene	RRID: Addgene_62348
Software and algorithms		
FlowJo	FlowJo LLC	https://www.flowjo.com/
R	(R Core Team, 2014)	http://www.R-project.org/
ggplot2	Wickham, 2016	
FIMO	Grant et al., 2011	https://meme-suite.org/meme/doc/fimo.html
MEME-ChIP	Machanick and Bailey, 2011	https://meme-suite.org/tools/meme-chip
ImageJ	Schneider et al., 2012	https://imagej.nih.gov/ij/
MACS2	Zhang et al., 2008	https://github.com/macs3-project/MACS/
BedTools	Quinlan, 2014	https://bedtools.readthedocs.io/en/latest/
DESeq2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
FLASH2	Mago and Salzberg, 2011	http://www.cbcb.umd.edu/software/flash
Seqtk	https://github.com/lh3/seqtk/blob/master/README.md	https://github.com/lh3/seqtk
matchPWM	Wasserman and Sandelin, 2004	N/A
ViennaRNA	Lorenz et al., 2011	N/A