

 Open access • Posted Content • DOI:10.1101/2021.07.14.452365

## Enhancer-promoter interactions and transcription are maintained upon acute loss of CTCF, cohesin, WAPL, and YY1 — [Source link](#)

Tsung-Han S. Hsieh, Claudia Cattoglio, Elena Slobodyanyuk, Anders S. Hansen ...+2 more authors

**Institutions:** Massachusetts Institute of Technology

**Published on:** 14 Jul 2021 - bioRxiv (Cold Spring Harbor Laboratory)

**Topics:** CTCF, Cohesin, Enhancer, Transcription factor and Chromatin

Related papers:

- [Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl](#)
- [Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers](#)
- [Genome-Wide Control of RNA Polymerase II Activity by Cohesin](#)
- [The genomic landscape of cohesin-associated chromatin interactions](#)
- [The cohesin-associated protein Wapl is required for proper Polycomb-mediated gene](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/enhancer-promoter-interactions-and-transcription-are-29ex3czhm9>

# Enhancer-promoter interactions and transcription are maintained upon acute loss of CTCF, cohesin, WAPL, and YY1

Tsung-Han S. Hsieh<sup>1,2,3,4,6</sup>, Claudia Cattoglio<sup>1,2,3,4\*</sup>, Elena Slobodyanyuk<sup>1,2\*</sup>, Anders S. Hansen<sup>5</sup>, Xavier Darzacq<sup>1,2,3,6†</sup>, Robert Tjian<sup>1,2,3,4†</sup>

<sup>1</sup>Department of Molecular and Cell Biology;

<sup>2</sup>Li Ka Shing Center for Biomedical and Health Sciences;

<sup>3</sup>CIRM Center of Excellence;

<sup>4</sup>Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA 94720, USA

<sup>5</sup>Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>6</sup>Center for Computational Biology, University of California, Berkeley, Berkeley, CA 94720, USA

\*Equal contributions

†Correspondence: R.T. ([jmlim@berkeley.edu](mailto:jmlim@berkeley.edu)), X.D. ([darzacq@berkeley.edu](mailto:darzacq@berkeley.edu))

## Abstract

It remains unclear why acute depletion of CTCF and cohesin only marginally affects expression of most genes despite substantially perturbing 3D genome folding at the level of domains and structural loops. To address this conundrum, we used high-resolution Micro-C and nascent transcript profiling to find that enhancer-promoter (E-P) interactions are largely insensitive to acute (3-hour) depletion of CTCF, cohesin, and WAPL. YY1 has been proposed to be a structural regulator of E-P loops, but acute YY1 depletion also had minimal effects on E-P loops, transcription, and 3D genome folding. Strikingly, live-cell single-molecule imaging revealed that cohesin depletion reduced transcription factor binding to chromatin. Thus, although neither CTCF, cohesin, WAPL, nor YY1 are required for the short-term maintenance of most E-P interactions and gene expression, we propose that cohesin may serve as a “transcription factor binding platform” that facilitates transcription factor binding to chromatin.

## Introduction

High-throughput chromosomal conformation capture (Hi-C)-based assays have transformed our understanding of 3D genome folding<sup>1,2</sup>. Based on such studies, we can distinguish at least three levels of 3D genome folding. First, the genome is segregated into A and B compartmental domains, which largely correspond to active and inactive chromatin segments, respectively, and appear as a plaid-like pattern in Hi-C contact maps<sup>3</sup>. Second, the proteins CTCF and cohesin help fold the genome into topologically associating domains (TADs)<sup>4,5</sup> and structural chromatin loops<sup>6</sup>. Third, at a much finer scale, transcriptional elements engage in long-range chromatin interactions such as enhancer-promoter (E-P) and promoter-promoter (P-P) interactions to form local domains<sup>7-9</sup>.

Elegant experiments combining acute protein depletion of CTCF, cohesin, and cohesin regulatory proteins with Hi-C or imaging approaches have revealed the role of CTCF and cohesin in regulating the first two levels, TADs and compartments<sup>10-14</sup>. These studies have shown that while CTCF and cohesin play only a minor role in compartmentalization, their removal largely eliminates TADs and chromatin loops anchored by these proteins across the genome. CTCF and cohesin are thought to form TADs and loops through loop extrusion<sup>15,16</sup>. This model posits that cohesin extrudes bidirectional loops until it encounters convergent and occupied CTCF binding sites. When averaged across cell populations, the extruded chromatin appears to be spatially organized into a self-interacting domain (TAD or loop domain), and CTCF binding sites constitute the domain boundaries that restrict inter-domain interactions. Halting of cohesin extrusion at CTCF sites sometimes gives rise to sharp corner peaks in contact maps, known as loops or corner dots. WAPL, a cohesin unloader, releases cohesin from chromatin and WAPL depletion therefore increases cohesin residence times as well as the amount of cohesin on chromatin<sup>12,14</sup>.

However, Hi-C is ineffective for capturing the third level of 3D genome folding, the fine-scale transcriptionally important E-P and P-P interactions<sup>7,17</sup>. Indeed, a recent genetic screen found that the majority of functional E-P interactions are not identified as contacts in Hi-C data<sup>18</sup>. Our understanding of the role of CTCF and cohesin in regulating gene expression has thus mainly come from genetic experiments. Paradigmatic experiments focusing on mouse development suggested that TAD disruption through inversion or deletion of CTCF sites around developmental genes such as *Epha4*, *Kcnj*, or *Ihh* can cause severe limb malformation<sup>19</sup>. Similar studies at the locus encoding the morphogen Sonic hedgehog has led to somewhat inconsistent effects on transcription and developmental phenotypes, perhaps due to manipulation of different CTCF sites<sup>20,21</sup>. Furthermore, CTCF and cohesin appeared to be crucial to some biological processes such as neuronal maturation<sup>22</sup> and lipopolysaccharide-induced inflammatory response<sup>23</sup>, but their presence seemed dispensable in other cases such as neuronal activity-dependent transcription<sup>22</sup> and immune cell transdifferentiation<sup>23</sup>. Thus, it remained unclear if, when, where, and how CTCF/cohesin regulates E-P and P-P interactions and gene expression.

Our current lack of understanding of the role of CTCF, cohesin, and other factors in regulating E-P, P-P, and transcriptionally relevant fine-scale genome folding has largely been limited by the inability of Hi-C to directly interrogate these finer scale interactions. We recently reported that Micro-C can effectively resolve ultra-fine 3D genome folding at nucleosome resolution, including E-P and P-P interactions, thus overcoming this limitation<sup>7,24-27</sup>. Here, we used Micro-C to systematically investigate fine-scale 3D genome folding before and after acute depletion (3 hours) of CTCF, the cohesin subunit RAD21, the cohesin unloader WAPL, and the putative structural protein YY1 in mouse embryonic stem cells (mESCs). We also profiled

chromatin occupancy, total RNAs, and nascent transcripts<sup>28</sup> in the same conditions. This integrated genome-wide fine-resolution mapping approach allowed us to dissect the primary effects of depleting loop extrusion factors on gene regulatory chromatin interactions and transcription. Finally, focusing on the dynamics of YY1 uncovered an unexpected role for cohesin in facilitating transcription factor (TF) binding.

## Results

### Genome-wide identification of transcription-linked chromatin loops

We previously used Micro-C to determine fine-scale (~200 nucleotide) 3D genome structure at a resolution below TADs<sup>7</sup>. These structures correlate well with transcriptional activity and often facilitate contact between promoters and promoters (P-P) or enhancers and promoters (E-P), forming “dots” or “loops” at their intersections on Micro-C maps. Using a newly developed loop caller, Mustache<sup>29</sup>, we identified over 75,000 statistically significant dots/loops in mESCs, which is about 2.5 times greater than in our previous report<sup>7</sup>. Through analysis of local chromatin state enrichment at loop anchors, we sub-classified these loops into ~13,735 cohesin loops, ~20,369 E-P loops, ~7,433 P-P loops, and ~700 Polycomb-associated loops (**Fig. 1a-b and Extended Data Fig. 1a**). These chromatin loops span a broad range of lengths with a median size of ~160 kb for cohesin loops and ~100 kb for E-P or P-P loops (**Extended Data Fig. 1c**). Although cohesin loops exhibit the strongest interactions (~5.9-fold higher than the background), the contact intensity of E-P and P-P loops is still significantly stronger (~3-fold enrichment) than pairs between promoters and random genomic loci (**Fig. 1b**). We obtained similar results with an alternative loop calling algorithm, Chromosight<sup>30</sup> (**Extended Data Fig. 1b-c**). We note that Micro-C has much higher sensitivity for detecting E-P and P-P contacts compared to Hi-C, establishing Micro-C as a more suitable assay to study genome organization relevant to transcription regulation genome-wide in an unbiased manner (**Extended Data Fig. 1d**).

To gain a better understanding of the relationship between active transcription and chromatin loops, we profiled nascent transcription by mNET-seq<sup>28</sup> in mESCs (**Fig. 1c and Extended Data Fig. 1e**). Newly transcribed RNAs generally have a higher correlation with E-P contacts than with compartments and TADs (**Extended Data Fig. 1f**). Specifically, the strength of E-P and P-P interactions positively correlates with the level of gene expression, while cohesin loops show no such correlation (**Fig. 1d-e**).

The region around the *Klf2* gene illustrates the complexities of fine-scale 3D genome folding and how we can use Micro-C maps to identify novel E-P contacts that are mostly unresolved in Hi-C (**Fig. 1f and Extended Data Fig. 1g**). The CTCF and cohesin ChIP-seq peaks show strong contact signals between the *Ap1m1* and *Eps15l1* genes (blue arched lines and circle), which insulate the *Klf2* gene from communicating with regions outside the loop domain. However, multiple weak interactions within the downstream 150-kb region around the *Med26* gene still occur without apparent cohesin residency at their anchors (**Fig. 1f**, right panel: purple arched lines and circles), and these contacts sharply correlate with nascent transcription signals at promoters and enhancers.

To validate these observations at a genome-wide scale, we plotted ChIP-seq data for various histone marks, transcription factors, and chromatin remodelers over the major types of loop anchors (**Extended Data Fig. 1h**). Consistent with our previous characterization of the subtypes of chromatin structures below TADs<sup>7</sup> as well as a recent imaging study<sup>31</sup>, loop anchors

enriched in CTCF and cohesin generally do not colocalize with sites of active transcription. In contrast, various transcription factors, coactivators, and Pol II are associated with E-P and P-P anchors. Thus, by coupling Micro-C with nascent RNA-seq, we can more precisely delineate which chromatin loops are associated with active transcription in a given cell type.

## **E-P and P-P loops can cross TAD boundaries**

TAD boundaries formed by CTCF and cohesin are thought to regulate E-P and P-P interactions in two ways: by increasing interactions inside the TAD and by blocking interactions across TADs<sup>2</sup>. Recent single-cell Hi-C and imaging studies have argued that TADs and their boundaries are likely to vary substantially from cell to cell<sup>32–35</sup>, which may allow enhancers or promoters to escape the insulation and contact regions outside of a TAD. Thus, it remains debated whether the boundaries can prevent an enhancer from interacting with and activating a gene in another TAD. Interestingly, our genome-wide analysis uncovered that, although loop interactions largely decay across distance (**Extended Data Fig. 1i**), ~22.5% of E-P and ~33.2% of P-P loops that cross TAD boundaries retain a comparable level of contact intensity to equidistant loops within a TAD (**Fig. 1g**). Genes located at the anchors of these inter-TAD loops also show similar or even higher expression levels in nascent or total RNA analysis (**Fig. 1h**). We postulated two possibilities that could lead to this observation: the TAD boundaries that are crossed by E-P and P-P loops have lower CTCF or cohesin occupancy or have weaker insulation propensity. We first split the TAD boundaries into two groups: “crossed” or “not crossed” by loops (**Extended Data Fig. 1j**, top). Strikingly, CTCF and RAD21 occupancy at the boundaries is almost the same regardless of whether or not the boundaries are crossed by E-P, P-P, or cohesin loops (**Extended Data Fig. 1j**, bottom). The TAD boundaries crossed by either E-P or P-P loops show only slightly weaker insulation strength than the non-crossed boundaries (**Fig. 1i**). In contrast, the boundaries that insulate the cohesin loops are substantially stronger than those that allow their crossing (**Fig. 1i**). Together, these results indicate that TAD boundaries are much more effective at insulating cohesin loops than insulating E-P or P-P loops. Over 20% of enhancers can still engage with their target sites located outside of the TAD regardless of CTCF or cohesin occupancy and insulation strength at the boundaries (**Fig. 1g**), and these boundary-crossing interactions largely correlate with transcriptional output (**Fig. 1h**). We thus hypothesize that both TAD-dependent and TAD-independent mechanisms contribute to regulation of E-P interactions.

## **Acute depletion of CTCF, cohesin, and WAPL perturbs structural loops**

To test whether active loop extrusion is essential for maintaining various types of chromatin loops and how these structures impact transcription, we endogenously and homozygously tagged each of the three primary loop extrusion factors (CTCF, RAD21, or WAPL) with an auxin-inducible degron (AID) by CRISPR/Cas9-mediated genome editing in mESC lines expressing the F-box protein *OsTir1* (**Fig. 2a and Extended Data Fig. 2a-b**)<sup>36</sup>. We confirmed near-complete degradation of the AID-tagged proteins after ~3 hours of treatment with indole-3-acetic acid (IAA) by immunoblotting (**Fig. 2b and Extended Data Fig. 2c**). We note that previous studies employing acute CTCF or cohesin depletion used prolonged degradation (6 – 24 hours<sup>10,11,37</sup>), which may confound the primary molecular response with potential secondary effects<sup>38</sup>. To avoid such confounding effects, we chose to use 3 hours of depletion in this study.



We first profiled the genome-wide binding of CTCF, RAD21, SMC1A, and SMC3 using ChIP-seq in the AID-tagged lines treated with either ethanol (untreated, UT) or IAA to degrade the tagged protein. Overall, hierarchical clustering shows high quality and reproducibility between replicates (**Extended Data Fig. 2d**). CTCF and cohesin peaks were also highly reproducible between the untreated samples (Jaccard index > 0.7) (**Extended Data Fig. 2e**). We then asked how the loss of each loop extrusion factor affects the binding of each remaining factor. Consistent with previous studies<sup>10,39</sup>, both CTCF and cohesin lose their occupancy after CTCF depletion (**Fig. 2c-d and Extended Data Fig. 2f-g**). Differential peak analysis<sup>40</sup> confirmed that over 90% of CTCF peaks and 60% of cohesin peaks are significantly decreased upon loss of CTCF (adjusted p-value < 0.05) (**Fig. 2e and Extended Data Fig. 2h**). Thus, our results are in line with the widely accepted conclusion that CTCF is required to position cohesin<sup>41</sup> and perhaps protects cohesin against release by WAPL<sup>42</sup>. It has been reported that cohesin is also positioned by actively transcribing Pol II and by the transcription machinery<sup>41,43</sup> but the ChIP signal at these putative “CTCF-independent” loci is generally quite weak<sup>39</sup> (**Extended Data Fig. 2i**). Specifically, promoter-bound cohesin peaks were barely detectable in untreated cells, and though they slightly increased after CTCF depletion, we could still only detect a few thousand upregulated cohesin peaks upon CTCF degradation (**Fig. 2e and Extended Data Fig. 2i-j**)<sup>41</sup>. We thus suggest that in the absence of CTCF, the transcription machinery may halt cohesin extrusion but not as effectively as CTCF does. Whether the promoter-bound cohesin peaks detected after CTCF depletion are functional remains to be determined.

While cohesin peaks are generally lost upon CTCF degradation, CTCF binding is unaffected by altering the level of cohesin on chromosomes (**Fig. 2c-d**)<sup>11</sup>. RAD21 or WAPL depletion caused only a 10 to 20% reduction in CTCF peaks (**Fig. 2e and Extended Data Fig. 2f-g**). A recent study reported that thousands of cohesin peaks are lost near SOX2 and OCT4 binding sites after WAPL depletion<sup>44</sup>. Surprisingly, we did not observe even these minor changes as reported previously (**Fig. 2e and Extended Data Fig. 2h & 2k**), perhaps due to differences in degradation levels or the duration of depletion. Taken together, our ChIP-seq data confirmed effective degradation of loop extrusion factors within 3 hours and largely recapitulated previous observations<sup>10,11,39</sup>.

Next, we used high-resolution Micro-C to analyze the effect of CTCF, RAD21, and WAPL depletion on fine-scale 3D genome structures. Since our Micro-C data were highly reproducible across replicates (**Extended Data Fig. 3a-b**), we pooled the replicates to achieve ~1-2 billion unique reads for each sample. At the global level, our findings largely agree with previous studies<sup>10-12,14</sup>. Using contact probability  $P(s)$  analysis, we observed that CTCF depletion had minimal impact on overall interactions across the genome; RAD21 depletion reduced contact frequency in the range of 10 – 200 kb but increased interactions at 300 kb – 5 Mb; and WAPL depletion showed the opposite trend, with increased contacts at 70 – 700 kb but reduced contacts at 1 – 5 Mb (**Extended Data Fig. 3c**). In addition, loop strength analysis revealed that nearly 90% of cohesin loops were lost upon depletion of CTCF or RAD21, while most loops were retained in a similar or slightly higher strength after WAPL depletion (**Fig. 2f-g and Extended Data Fig. 3d**). Indeed, after WAPL depletion, an additional ~6,000 longer-range loops were sufficiently strengthened to meet our detection threshold (**Extended Data Fig. 3e**). These results suggest that, after CTCF depletion, cohesin-mediated DNA extrusion operates in a more unrestricted manner, resulting in the apparent loss of CTCF-anchored loops but no overall reduction of genomic interactions (**Extended Data Fig. 3c**)<sup>10,12,41</sup>. Interestingly, in comparison to a recent finding<sup>45</sup>, our data show only moderate CTCF persistence at architectural sites (e.g., loop anchors or TAD borders) upon CTCF degradation (**Extended Data Fig. 3f**). CTCF binding sites associated with

mitotic bookmarking<sup>46</sup> and frequent occupancy sites in single-molecule foot-printing (SMF) data<sup>47</sup> also did not show apparent CTCF persistence on chromatin (**Extended Data Fig 3g-h**). Given the near-complete ablation of cohesin loops after CTCF degradation (**Fig. 2g**), we suggest that the persistent residual CTCF proteins, if any, are insufficient to halt or position cohesin. Together, our Micro-C results confirmed a nearly complete disruption of cohesin loops at CTCF-anchored sites 3 hours after CTCF or RAD21 depletion and identified many emerging long-range chromatin loops in WAPL-depleted cells.

## Acute loss of CTCF, cohesin, and WAPL does not affect expression of most genes

We next asked whether acute disruption of active loop extrusion impacts the maintenance of gene expression. To capture the immediate effect of depleting loop extrusion factors on transcription, we profiled nascent transcription by mNET-seq for the control and depletion conditions along with total RNA-seq. Differential expression tests of ~30,000 genes identified that ~129 nascent transcripts changed in CTCF depletion, ~32 changed in RAD21 depletion, and only 4 changed in WAPL depletion (**Fig. 2h and Extended Data Fig. 3i-j**). Total RNA-seq found ~15 genes to be significantly deregulated in both CTCF and RAD21 depletion and ~6 genes in WAPL depletion (**Fig. 2i and Extended Data Fig. 3j**). The differentially expressed genes (DEGs) are highly consistent between CTCF and cohesin depletion and between the assays (log2 odds ratio > 6) (**Fig. 2j**). This suggests that while CTCF and cohesin are required for the transcriptional maintenance of only a small subset of genes, those genes tend to require the presence of both factors.

Furthermore, we found that the early deregulated genes upon loss of CTCF and cohesin include many cell type-specific transcription factors (e.g., *Sox21*, *Sox1*, *Myc*, *Klf4*, *Gbx2*, *Foxj1*, *Dlx3*) (**Fig. 2i**). As expected, chromatin structures around the DEGs were strongly disrupted, often featuring loss of a boundary or domain and gain of *de novo* chromatin interactions (**Fig. 2k-l**). Since perturbing the levels of these proteins in stem cells could trigger exit from the pluripotent state and/or cell death<sup>48–53</sup>, we envision that a longer time of protein degradation may induce indirect, loop-independent transcriptional responses. This finding highlights the importance of distinguishing between primary and secondary effects of perturbations in the study of loop extrusion and gene expression<sup>38</sup>, and suggests that some conclusions from studies employing 24 or 48 hours of degradation may need to be re-evaluated<sup>10,37,44</sup>. In summary, we find that, while CTCF, cohesin, and WAPL may regulate some gene expression, their acute depletion affects the transcription of only a handful of genes in mESCs, which largely encode pluripotency and differentiation factors.

## Loop extrusion factors are largely dispensable for the maintenance of enhancer-promoter and promoter-promoter interactions

The very modest transcriptional changes seen upon CTCF and cohesin degradation suggest that transcriptional E-P and P-P interactions may persist for at least 3 hours after the depletion of either CTCF, cohesin, or WAPL. To test this hypothesis, we quantified the loop strength at all 75,000 dots identified in wild type mESCs in both control and loop extrusion factor depletion conditions. About 20% of loops are significantly decreased, but the remaining 60,000 loops remain largely unaltered (**Fig. 3a**). Consistent with our previous results, the disrupted loops are CTCF- or cohesin-dependent, while the persistent loops are mostly anchored by promoters and enhancers

(**Fig. 3b and Extended Data Fig. 4a-b**). To further validate this, we specifically quantified the strength of loops that are anchored by E-P and P-P in the distance ranging from 5 kb to 2 Mb. Remarkably, acute depletion of CTCF and cohesin had a negligible impact on the E-P and P-P loops, with ~80% of E-P contacts (**Fig. 3c**) and 90% of P-P contacts (**Fig. 3d**) remaining unaltered. Interestingly, although more distal loops emerge in the WAPL depleted cells (**Extended Data Fig. 3c & 3e**), the resulting higher density of cohesin on chromatin did not increase the strength of preexisting loops (**Extended Data Fig. 4c-d**). WAPL depletion also had only a minor impact on E-P and P-P interactions (**Fig. 3c-d**).

What are these 10-20% CTCF/cohesin-sensitive E-P or P-P loops (**Fig. 3c-d**)? According to the loop extrusion model, cohesin might either directly bridge E-P interactions at TAD boundaries, or the process of extrusion might inherently increase the frequency of long-range E-P interactions<sup>54</sup>. The levels of CTCF and cohesin occupancy at the anchors of these loops are much higher than those at the unaffected loop anchors (**Extended Data Fig. 4e**), suggesting that these loops may overlap with TAD boundaries, where CTCF and cohesin proteins are highly enriched. To test this possibility, we compared the genomic features of these loops relative to TAD boundaries. The lengths of the affected E-P loops are well below the average size of a TAD, ranging from 50 to 200 kb, with CTCF-sensitive E-P loops slightly larger than cohesin-sensitive loops (**Extended Data Fig. 4f**). Surprisingly, only ~5% of the anchors are located within a 1-kb window surrounding the boundaries (**Fig. 3e-f**). Most of them are located even farther away than the unaffected loops (**Fig. 3e-f and Extended Data Fig. 4g-h**) and tend to interact with another DNA locus within the same TAD without crossing the boundary (**Fig. 3g**). We next tested whether the affected E-P or P-P interactions were associated with the DEGs in nascent RNA-seq. Indeed, E-P or P-P interactions showed a greater decrease when their associated genes were deregulated upon loss of CTCF/cohesin (**Fig. 3h**). We thus conclude that the CTCF/cohesin-sensitive E-P loops have higher CTCF and cohesin occupancy at their anchors, but these anchors are not associated with TAD boundaries. Taken together, our results suggest that E-P and P-P contacts and fine-scale gene folding largely persist and remain transcriptionally functional even after near-complete depletion of CTCF, cohesin, or WAPL.

## Probing YY1 as a candidate regulator of E-P and P-P links

Our finding that E-P and P-P interactions largely remain intact in the absence CTCF, cohesin, and WAPL suggests that other proteins are likely responsible for E-P and P-P interactions. To address this, we searched for factors specifically enriched at E-P and P-P loop anchors. BRD2, BRD4, P300, ESRRB, SP1, Mediator, YY1, pluripotency transcription factors, and chromatin remodelers are all broadly enriched at enhancer and promoter loop anchors (**Extended Data Fig. 1h**). Although Mediator has been proposed to function as a structural complex to bridge E-P interactions<sup>55</sup> and promote the folding of sub-TAD structures<sup>56</sup>, two recent studies showed that loss of Mediator does not strongly affect E-P interactions<sup>57,58</sup>.

We therefore focused on YY1, a multifunctional zinc finger-containing transcription factor (**Extended Data Fig. 5a**) that is ubiquitously expressed, highly conserved, and essential for embryonic development in mammals<sup>59</sup>. Heterozygous YY1 mutations cause Gabriele-de Vries syndrome, which is characterized by developmental delay and intellectual disability<sup>60</sup>. YY1 has been implicated in chromatin looping<sup>61</sup>, and a previous study proposed that YY1 is a master structural regulator of E-P interactions specifically<sup>62</sup>. Nevertheless, upon rapid induction (1 hour) of erythroid differentiation, YY1 triggers little or no change in H3K27ac and H3K27ac-anchored



HiChIP interactions<sup>63</sup>. These confounding results led us to investigate the role, if any, of YY1 in mediating E-P interactions using Micro-C and YY1 depletion.

## Acute YY1 depletion has little effect on global gene expression and E-P and P-P interactions

To investigate the function of YY1 in genome organization and transcriptional regulation, we fused the miniIAA7 tag<sup>64</sup> to the endogenous YY1 locus to allow for rapid protein degradation. Immunoblots and flow cytometry analysis confirmed a nearly complete degradation one hour after IAA addition (**Fig. 4a and Extended Data Fig. 5a**). We chose a 3-hour degradation time for all the following assays to avoid measuring secondary effects. ChIP-seq analysis showed a clear depletion of YY1 at its cognate sites (**Fig. 4b**), which are primarily enriched at promoters, enhancers, and bivalent loop anchors (**Fig. 4c and Extended Data Fig. 5b**), consistent with its reported role in E-P interactions. The majority of YY1 peaks were diminished, but only ~50% of peaks (n = 15075) were called significantly changed by differential peak analysis<sup>40</sup> (**Fig. 4b and Extended Data Fig. 5c**). We thus only focused on those loci that had a significant loss of YY1 occupancy. We also noticed a modest decrease in cohesin occupancy after loss of YY1 (**Fig. 4b and Extended Data Fig. 5d**), which may be associated with YY1's potential to position or halt cohesin<sup>62</sup>.

To characterize YY1's role in 3D genome organization, we acquired ~850 M unique Micro-C reads after pooling high-quality replicates from mock-treated and YY1-depleted cells (**Extended Data Fig. 5e**). We found that YY1 depletion has no strong effect on chromatin compartments, TADs, and cohesin loops (**Fig. 4d and Extended Data Fig. 5f**), and no significant change in genome-wide chromatin interaction probabilities (**Extended Data Fig. 5g**). YY1 was proposed to be a causally required structural regulator of transcription and E-P interactions<sup>62</sup>. Surprisingly, acute removal of YY1 only significantly affected ~800 loops (**Fig. 4e**), and ~325 and ~40 genes in the RNA-seq and mNET-seq profiling, respectively (**Fig. 4f-g and Extended Data Fig. 5h**). The DEGs in RNA-seq data significantly overlapped with the genes associated with differential loops or DEGs identified after 24 hrs of YY1 depletion<sup>62</sup> (**Fig. 4h and Extended Data Fig. 5h**), indicating a high correlation between E-P loops and gene expression, as well as confirming capture of an early response to YY1 degradation. More importantly, genome-wide pileup analysis for the intersections of YY1 peaks, E-P loops, and P-P loops showed only a very minor change in loop intensity after YY1 depletion (**Fig. 4i and Extended Data Fig. 5f**). This result suggests that the maintenance of most E-P loops and their regulatory functions in general do not require the presence of YY1, at least within a 3-hour depletion window. Nevertheless, a specific set of loci appears to require the presence of YY1 in order to interact with their cis-regulatory elements. For example, at the *Ifnar2* and *Ikzf2* gene loci, the weak E-P or P-P interactions were lost upon YY1 depletion and were often accompanied with either an increase or decrease in nascent transcription (**Fig. 4j and Extended Data Fig. 5i**). Taken together, although YY1 may be required for some limited set of E-P and P-P interactions, these results are not consistent with the model that YY1 is a general master structural regulator of E-P interactions in mESCs as previously proposed<sup>62</sup>.

## Single-molecule imaging reveals YY1 binding dynamics and nuclear organization

The surprisingly meager effects of YY1 on chromatin looping, E-P, and P-P interactions might result from YY1 DNA binding being very transient and/or due to only a small fraction of

YY1 proteins being bound to DNA in live cells. To better understand the dynamics and mechanisms underlying YY1 function in living cells, we tagged YY1 with HaloTag, a self-labeling protein tag that can be covalently conjugated with cell-permeable synthetic dyes suitable for single-molecule imaging in live cells (**Extended Data Fig. 6a**)<sup>65</sup>. We either knocked in a HaloTag at the N-terminus of the endogenous YY1 by CRISPR-Cas9-mediated genome editing (**Fig. 5a and Extended Data Fig. 6b**) or ectopically expressed YY1 fused with HaloTag via a minimal mammalian *L30* promoter by *PiggyBac* transposition. We isolated and validated two homozygously tagged clones, designated YN11 and YN31. We further confirmed that the Halo-tagged YY1 protein in both endogenously tagged and ectopically expressing clones was of the correct size and expressed at close to endogenous levels (**Fig. 5a and Extended Data Fig. 6c**). Live-cell confocal imaging validated that HaloTag-YY1 was predominantly localized within the nucleus and appeared to be non-homogeneously distributed throughout the nucleoplasm, with noticeable exclusion from nucleoli (**Fig. 5b-c and Extended Data Fig. 6d**). We observed bright YY1 puncta sporadically clustered within nucleoli, which is reminiscent of the nucleolar localization of TBP and its function on Pol III transcription (**Fig. 5b-c and Extended Data Fig. 6d**)<sup>66,67</sup>. We then visualized the nuclear distribution of YY1 at single-molecule resolution by using photoactivated localization microscopy (PALM) (**Fig. 5d**). YY1 displays a punctate distribution with many high-density clusters throughout the nucleoplasm. These high-density clusters are unlikely to be a fixation artifact, since such puncta were also visible in Airyscan live-cell imaging (**Fig. 5c and Extended Data Fig. 6d**). YY1 has been thought to be evicted from chromosomes during mitosis in fixed-cell imaging experiments<sup>68</sup>. However, our live-cell imaging showed continued YY1 residency on mitotic chromosomes, suggesting that YY1 may be involved in mitotic bookmarking (**Fig. 5b and Extended Data Fig. 6d**)<sup>69</sup>. Together, these results validate our homozygous HaloTag-YY1 knock-in cell lines and reveal that YY1 binds mitotic chromosomes and forms local high concentration hubs in the nucleus.

Having characterized our cell lines, we next interrogated YY1 protein dynamics and target search mechanisms. We took advantage of the stroboscopic photo-activation single-particle tracking technique (spaSPT)<sup>70,71</sup> to minimize motion-blur and tracking errors to unambiguously trace the movement of individual YY1 molecules at a frame rate of ~133 Hz (**Fig. 5e and Extended Data Fig. 6e**). Spots corresponding to HaloTag-YY1 molecules were detected and tracked to form trajectories<sup>72</sup>. We then inferred distributions of diffusion coefficients from the spaSPT data using a finite-state approximation to the Dirichlet process mixture mode (State Array) implemented in the newly developed package Spagl (**Extended Data Fig. 6e-f**)<sup>72</sup>. The diffusive mixtures can be attributed to at least two major apparent diffusion states<sup>72</sup>, including a bound population (diffusion coefficient ( $D_{bound}$ ) < 0.1  $\mu\text{m}^2\text{s}^{-1}$ ) and a broad mixture of freely diffusing molecules ( $D_{free}$  > 0.1  $\mu\text{m}^2\text{s}^{-1}$ ) (**Fig. 5f**). We found that ~31% of YY1 is in an immobile state, presumably bound to chromatin, with the remaining population exhibiting either slow diffusion (~26%;  $D_{slow}$  ~ 0.1 – 2  $\mu\text{m}^2\text{s}^{-1}$ ) or fast diffusion in the nucleoplasm (~43%;  $D_{fast}$  > 2  $\mu\text{m}^2\text{s}^{-1}$ ) (**Fig. 5f**). These measurements largely agree with kinetic modeling of displacements obtained with the Spot-On algorithm (**Extended Data Fig. 6g**)<sup>71</sup>. As an independent and orthogonal validation of these results, we biochemically probed YY1 chromatin binding affinity by subjecting freshly isolated nuclei to a series of washes with increasing salt concentrations (**Extended Data Fig. 7a**). The majority of YY1 is extracted at low salt concentrations (75 – 300 mM), but a subpopulation stayed bound on chromatin (~28%), resisting 1 M washes, consistent with the ~31% chromatin bound population estimated by spaSPT. The fraction of YY1 stably associating with chromatin is substantially lower than CTCF (~43%) and cohesin (~65%) (**Fig. 5g and Extended Data Fig. 7b**).

To further link the spatial relationship of YY1 proteins with their different diffusion rates, we reconstructed the spatial distributions of trajectories by their likelihood of diffusion coefficients at 0.02, 0.1, 1, 5, and 20  $\mu\text{m}^2\text{s}^{-1}$ . Consistent with the live-cell imaging and PALM results, the immobile fraction showed apparent cluster-like structures with  $\sim 12 - 30$  trajectories per spot (from a total of  $\sim 15,000$  trajectories per cell), while faster moving YY1 subpopulations showed weaker propensity to cluster within a constrained area (**Extended Data Fig. 7c-d**). Overlays of the immobile trajectories with the other fractions showed that the molecules within the bound regime had nearly complete overlap, but the faster-diffusing molecules were less likely to co-localize with the immobile clusters (**Fig. 5h**). The averaged (**Fig. 5h, insets**) and differential cluster signals (**Fig. 5h, bottom panel**) between the immobile and the other fractions (**Extended Data Fig. 7c-d**) further confirmed that the faster-moving molecules more frequently travel to areas in the vicinity of the immobile clusters. These results are consistent with chromatin-bound YY1 proteins forming clusters, while the diffusing YY1 molecules appear to search the 3D nuclear space outside of the clusters. We previously made similar observations for CTCF<sup>27,70,73</sup>.

The residence times of transcription factors bound at their targets often correlate with their functional outcomes<sup>74-76</sup>. To estimate the overall residence time of the bound fraction of YY1, we used fluorescence recovery after photobleaching (FRAP) to measure *in vivo* protein binding kinetics by fitting the fluorescence recovery curve to a kinetic model<sup>77,78</sup>. Using a reaction-dominant FRAP model, we estimated a residence time of  $\sim 13$  seconds for the majority of YY1 molecules (**Fig. 5i and Extended Data Fig. 7e-f**). Interestingly, we noticed that  $\sim 5-10\%$  of YY1 bound chromatin for a longer period of time with an average residence time of  $\sim 50$  seconds (**Fig. 5i and Extended Data Fig. 7e-f**). We also employed slow-SPT as an orthogonal approach to measure YY1 residence times. For slow-SPT, we imaged individual molecules at an exposure time of 100 ms, which blurs fast-moving molecules into the background and effectively captures stable binding events<sup>79</sup>. Model fitting for the survival curve of molecules<sup>79</sup> yielded a residence time of  $\sim 13$  seconds for YY1 (**Fig. 5j**). Consistent with the FRAP data showing  $\sim 5-10\%$  of long binders, slow-SPT with exposure times from 50 to 250 ms<sup>80</sup> further revealed a subpopulation of YY1 that binds chromatin for over half a minute (**Fig. 5h**).

While YY1's average residence time of  $\sim 13-40$  seconds is similar to that of many transcription factors, it is much shorter than the residence times of known structural factors such as CTCF ( $\sim 1-4$  min) and cohesin in G1 ( $\sim 20-25$  min) in mESCs<sup>70</sup>. These results may explain why YY1 depletion has a marginal effect on chromatin looping than does CTCF and cohesin depletion. CTCF/cohesin loops are generally stronger and almost completely lost upon CTCF/cohesin depletion (**Fig. 2f-g**), whereas YY1 loops tend to be weaker and less affected by YY1 depletion (**Fig. 4e&i**). Furthermore, the chromatin bound fraction of YY1 ( $\sim 30\%$ ) is considerably smaller than that for CTCF ( $\sim 50-60\%$ )<sup>70</sup>. Taken together, our imaging experiments suggest that a smaller fraction of YY1 is bound to DNA and that YY1 binding is more dynamic than CTCF, which may help explain why YY1 protein depletion has a much weaker effect on looping and 3D genome folding.

## Cohesin depletion alters YY1's chromatin binding

We recently showed that CTCF clusters are enriched in diffusive CTCF proteins near their binding sites, thereby accelerating their target search<sup>73</sup>. To test if CTCF and cohesin may similarly affect YY1's target search, we endogenously fused an auxin-inducible degron (AID) to CTCF or RAD21 in the HaloTag-YY1 parental line and confirmed  $> 90\%$  depletion after 3 hours of IAA

treatment (**Fig. 6a and Extended Data Fig. 8a**). Despite the high degradation efficiency, neither YY1's nuclear distribution nor its clustering was strongly affected after acute loss of CTCF and cohesin in either live- or fixed-cells (**Fig. 6b-c and Extended Data Fig. 8b**). This suggests that the maintenance of YY1 hubs is independent of CTCF and cohesin.

We next examined YY1 nuclear target search efficiency in the absence of CTCF and cohesin using spaSPT. Although CTCF depletion had no major effect, cohesin depletion resulted in a modest but reproducible decrease from ~33% to 22% (~31% decrease;  $p$ -value < 0.01) in the bound fraction of YY1 (**Fig. 6d and Extended Data Fig. 8c**). A lower bound fraction could either result from a shorter residence time ( $k_{\text{off}}$ ) or slower target search ( $k_{\text{on}}$ ). To distinguish between these possibilities, we analyzed the FRAP data and found YY1 residence times ( $k_{\text{off}}$ ) to be only weakly affected by CTCF and cohesin depletion (**Fig. 6e**). We therefore conclude that cohesin loss may affect YY1 target search ( $k_{\text{on}}$ ). Specifically, we estimated a ~54% decrease in  $k_{\text{on}}$  after cohesin depletion, resulting in a ~2.2-fold longer YY1 search time (UT=28 s; IAA=61 s), the time it takes YY1 on average to find and bind a cognate binding site after dissociating from DNA.

To independently test this SPT finding, we analyzed our ChIP-seq data. We found that ~3505 YY1 peaks (total peaks = ~41989) were lost after RAD21 degradation, and over 82% of these loci were associated with promoter regions (**Fig. 6f and Extended Data Fig. 8d-e**). In contrast, both CTCF and WAPL depletion had a negligible effect on YY1 occupancy (**Fig. 6f and Extended Data Fig. 8d-e**). These differences suggest that cohesin likely facilitates YY1 chromatin binding, especially at promoter regions.

Taken together, our results reveal a previously unappreciated role for cohesin in accelerating the target search of transcription factors resulting in increased YY1 chromatin binding as measured by SPT, FRAP, and ChIP-seq. These findings suggest that long-term cohesin depletion experiments must be interpreted with caution since cohesin depletion results in both direct and indirect effects including diminished general transcription factor binding to DNA (**Fig. 6g**).

## Discussion

Both the extent and mechanism by which CTCF and cohesin-mediated loop extrusion regulates transcription has remained puzzling and hotly debated<sup>10,11,22,23,31,34,37,41,45,81</sup>. Previous work has shown that CTCF or cohesin depletion abolishes TADs genome-wide and modestly affects gene expression<sup>10-14</sup>. However, since Hi-C cannot readily capture fine-scale transcriptionally relevant E-P and P-P interactions, how E-P and P-P interactions are regulated has remained unclear. Here we applied high-resolution Micro-C to overcome this limitation. Surprisingly, we found that neither CTCF, cohesin, WAPL, nor YY1 are required for the maintenance of E-P or P-P loops or transcription at least within a 3-hour window after depletion. Our findings, together with other evidence, allow us to distinguish and/or eliminate several models of E-P interactions previously assigned to these ubiquitous structural proteins (**Fig. 7**).

First, CTCF and cohesin have been proposed to either directly bridge E-P interactions<sup>55</sup> or to indirectly mediate E-P interactions by increasing contact frequency inside TADs (**Fig. 7, Model 1**)<sup>82</sup>. Our finding that acute CTCF, cohesin, and WAPL depletion minimally affects gene expression (**Fig. 2h-j**) and E-P interactions (**Fig. 3**) disfavors this model for short-term maintenance of E-P interactions, though CTCF and cohesin may still help establish E-P interactions indirectly. We propose that loop extrusion may often be a separable mechanism from



most E-P interactions and transcription, which is further supported by the following observations: 1) Over 20% of E-P or P-P loops can cross TAD boundaries and retain high contact probability and transcriptional activity (**Fig. 1**); 2) Only a very small handful of genes showed altered expression levels after CTCF, cohesin, or WAPL depletion (**Fig. 2**)<sup>10-14</sup>; 3) The majority of E-P or P-P loops persists after depletion of these structural proteins (**Fig. 3**)<sup>37,81</sup>; 4) CTCF/cohesin generally do not colocalize with transcription loci<sup>31</sup>; 5) E-P loops and transcription can be established in some cases even with no CTCF/cohesin expression<sup>22,23,83</sup>. Second, YY1 was proposed to be a master structural regulator of E-P interactions<sup>62</sup> (**Fig. 7, Model 2**). However, our Micro-C data is inconsistent with this model for the maintenance of E-P interactions, since acute YY1 depletion has little effect on E-P and P-P interactions or gene expression. Third, phase-separated transcriptional condensates have been proposed to mediate gene activation and E-P interactions<sup>84</sup>. Pol II, Mediator, BRD4, and many transcription factors containing intrinsically disordered regions (IDRs) that tend to aggregate into local-high concentration hubs in the nucleus<sup>85-87</sup>. Hub formation is thought to be a general property of TFs used to engage with regulatory elements and keep them in the spatial vicinity for gene regulation<sup>84-87</sup>. While not specifically addressed in this study, recent studies employing acute depletion/inhibition of Mediator and BRD4, which are supposed to be key players in condensate formation, also found no drastic effects on E-P interactions<sup>57,58,88</sup>. Thus, condensates also do not seem to be generally required for the maintenance of E-P interactions (**Fig. 7, Model 3**). In summary, we conclude that neither transcriptional condensates, CTCF, cohesin, WAPL, nor YY1 are generally required for the short-term maintenance of E-P interactions and the subsequent expression of most genes after acute depletion and loss of function.

Does this mean that none of these factors play any role in regulating E-P interactions and gene expression? The evidence that CTCF and cohesin can directly or indirectly regulate E-P interactions and affect gene expression in many cases is overwhelming (see examples below). To reconcile these studies with our new observations, we propose a “time-buffering” model (**Fig. 7, Model 4**). In this model, CTCF, cohesin, and architectural factors contribute to the establishment of E-P interactions, but not to their maintenance. Instead, once established, a molecular memory (e.g., histone modifications<sup>89</sup>, chromatin remodeling<sup>90-92</sup>, DNA modification<sup>93-95</sup>, lncRNAs<sup>96,97</sup>, etc.) may be sufficient to maintain E-P interactions and gene expression for several hours without the contribution of CTCF, cohesin, and other architectural factors. We propose that this time-buffering model reconciles our observations with the unambiguous genetic evidence that CTCF and cohesin regulate some E-P interactions. This evidence includes the following: 1) Insertion of CTCF sites between an enhancer and a promoter can both reduce E-P interactions and strongly reduce gene expression<sup>98-100</sup>; 2) CTCF binding site silencing<sup>101,102</sup> or genetic CTCF binding site loss<sup>21,103,104</sup> can cause aberrant E-P interactions and gene expression and drive disease; 3) Inversion or repositioning of CTCF sites can redirect E-P interactions that cause gene misexpression and diseases<sup>105,106</sup>. Two recent studies have also proposed variants of a time-buffering model based on mathematical modeling of E-P interactions and gene expression<sup>100,107</sup>. In both models, individual E-P interactions are memorized – either as long-lived promoter states<sup>100</sup> or as long-lived “promoter tags”<sup>107</sup> – such that gene expression can be temporally uncoupled from E-P interactions, yet still be causally linked. An alternative, more conservative interpretation of our data and the evidence cited above is that CTCF and cohesin only regulate a very small, unique sets of genes in specific biological processes, and their effect on a handful of loci simply cannot be generalized as a universal rule. In summary, we propose a time-buffering model for E-P interactions where CTCF, cohesin, and architectural factors contribute to establishing E-P interactions at some genes, but



where these factors are not generally required for the maintenance of E-P interactions and gene expression, except for a very small number of pluripotency genes in mESCs.

Here we also provide the first comprehensive study of YY1 dynamics and nuclear organization (**Fig. 5**). We discovered that: 1) YY1 binds mitotic chromosomes, implicating its potential function for mitotic bookmarking or transcriptional memory through the cell cycle; 2) YY1 forms nuclear hubs similar to CTCF and cohesin<sup>27,70,73</sup> that may associate with E-P or P-P interactions; 3) About 31% of YY1 proteins are bound to chromatin with an average residence time of ~13-40 seconds. Consistent with lower and less stable YY1 chromatin binding compared to CTCF and cohesin (CTCF: ~50% bound for ~1-4 min; Cohesin ~40-50% bound for ~20-25 min)<sup>70</sup>, acute YY1 depletion has little, if any, effect on 3D genome folding compared to CTCF and cohesin depletion.

Surprisingly, we found that cohesin depletion, but not CTCF depletion, significantly reduces YY1 chromatin binding and increases its target search time from 28 to 61 seconds. While more work on other transcription factors (TF) will be necessary to test the generality of this model, we propose that cohesin or its associated chromatin structures could serve as a general scaffold or platform for TFs that facilitates and increases TF binding to chromatin. Interestingly, the subunits of cohesin as well as its loading and unloading complexes are composed of multiple segments of IDRs, which have the potential to establish weak protein-protein interactions or multivalent interactions. These weak transient interactions may further engage with IDRs in TFs and thus facilitate their binding to chromatin<sup>108,109</sup>. Although more quantitative works will be necessary to unveil these mechanisms, in addition to its roles in loop extrusion, DNA repair, replication, and chromosome segregation, cohesin might also serve as a general scaffold that facilitates TF binding to chromatin and could be critical for ensuring the precise timing of gene activation and silencing during embryonic development and cell-state transitions<sup>83</sup>.

In summary, we have comprehensively investigated the role of CTCF, RAD21, WAPL, and YY1 in finer-scale chromatin structure, nascent transcription, as well as YY1 dynamics and nuclear organization. We propose a time-buffering model, where architectural proteins generally contribute to the establishment, but not the short-term maintenance, of E-P interactions and gene expression, and we also propose that cohesin plays an underappreciated role as a general scaffold that could facilitate TF binding to chromatin. The connection linking protein dynamics to chromatin structure opens a new avenue to rethink the mechanism of transcriptional regulation in the context of 3D genome organization.

## Figure legend

**Fig 1. Genome-wide identification of transcription-linked chromatin loops.** **a.** Micro-C identified over 75,190 chromatin loops, subclassified into four primary types (by Mustache loop caller<sup>29</sup>). **b.** Probability distribution of loop strength for cohesin, E-P, P-P, and random loops. The numbers of chromatin loops are shown on the left. The box plot indicates the quartiles for the loop strength score distribution. The genome-wide averaged contact signals (aggregate peak analysis (APA)) are plotted on the right. The contact map was normalized by matrix balancing and distance, with positive enrichment in red and negative signal in blue, shown as the diverging colormap with the gradient of normalized contact enrichment in log10. The ratio of contact enrichment for the center pixels is annotated within each plot. This color scheme and normalization method are used for normalized matrices throughout the manuscript unless otherwise mentioned. Asterisks denote

a  $p$ -value  $< 10^{-16}$  by the Wilcoxon test. **c.** Genome-wide averaged transcript counts for nascent transcript profiling. Genes are grouped into high, medium, and low expression level based on nascent RNA-seq data and rescaled to the same length from TSS (transcription start site) to TES (transcription end site) on the x-axis. **d.** Rank-ordered distribution of loop strength against gene expression for cohesin, E-P, and P-P loops. The distribution for each loop type was fitted and smoothed by LOESS regression. The standard error (SE) is plotted as a gray shade along with the regression line. **e.** APA are plotted by paired E-P or P-P loops and sorted by the level of nascent transcription into high, mid, and low levels. **f.** Snapshots of Micro-C maps of a ~300-kb region plotted with 800-bp resolution (left) and a ~150-kb region plotted with 200-bp resolution (zoomed-in, right). The standard heatmap shows the gradient of contact intensity for a given pair of bins. This color scheme is used for Micro-C maps throughout the manuscript. Contact maps are annotated with gene boxes and 1D chromatin tracks show the signal enrichment in the same region. Features like cohesin loops (blue arched lines and circles) and E-P/P-P loops (purple arched lines and circles) enriched at stripe intersections are highlighted. **g.** Schematic (top) showing two adjacent TADs insulated by CTCF boundaries and E-P/P-P interactions either within a TAD (intra-TAD, solid arched line) or across TADs (inter-TAD, dashed arched line). TADs called by Cooltools and Arrowhead returned similar results for the ratio of boundary-crossing E-P/P-P (see Method). APA (bottom) plotted for paired E-P/P-P that either cross (inter-TAD) or do not cross (intra-TAD) a TAD boundary. **h.** Nascent transcription ( $\pm$  strand) at the loop anchors of intra- (green) or inter-TAD (yellow) E-P/P-P loops. **i.** Heatmap and histogram profile of insulation scores at 20-kb resolutions spanning over the 1-Mb window for intra- (green) or inter-TAD (yellow) E-P/P-P loops. Colormap shows strong insulation in red and weak insulation in blue in log10.

**Fig 2. Acute depletion of loop extrusion factors affects a small set of genes.** **a.** Experimental design for the degradation of CTCF, RAD21, or WAPL. **b.** Immunoblots show the degradation levels of CTCF, RAD21, WAPL, and the loading control Actin-B after 3 hours of IAA treatment. **c.** Differential ChIP-seq signals for CTCF and RAD21 in cells depleted of CTCF, RAD21, or WAPL. The peaks called by MACS2 are plotted at the center across a  $\pm 3$ -kb region. The colormap shows an increased signal (log2) in orange and a decreased signal in purple after IAA treatment. **d.** Histogram profile of differential ChIP-seq signals for CTCF or RAD21 in cells depleted of CTCF, RAD21, or WAPL. **e.** Summary of differential ChIP-seq peak analysis. The chart shows the fraction of down-regulated, up-regulated, or unchanged peaks after IAA treatment. The total number of peaks for each protein was summed from all peaks in untreated and IAA-treated cells. **f.** Scatter plots of loop scores for cohesin loops in the untreated and IAA-treated cells. The overlaid heatmap indicates dot density, with the highest in red and the lowest in blue. Dashed lines along the diagonal mark the range that is characterized as unchanged loops. The pie chart (inset) shows the fraction of loop intensity that is increased, decreased, or unchanged after IAA treatment. The scatter plots for comparing loop intensities between two conditions are plotted in this format throughout the manuscript unless otherwise mentioned. **g.** APA are plotted by paired cohesin peaks for the untreated and IAA-treated cells. **h.** MA plots of nascent RNA-seq for CTCF or RAD21 depletion. Differentially expressed genes (DEGs) with  $q$ -value  $< 0.05$  are labeled by green (up) or pink (down). **i.** Volcano plot of total RNA-seq for CTCF or RAD21 depletion. DEGs ( $q$ -value  $< 0.01$ ) are labeled by pink with gene name. **j.** Overlap of DEGs between different depletions and assays. Colormap shows the odds ratio in log2 and the  $p$ -values are annotated on the corresponding comparisons. **k-l.** Snapshots of Micro-C maps comparing chromatin interactions in the untreated (top-right) and IAA-treated (bottom-left) cells surrounding *Enc1* or *Klf4* genes. Contact maps are

annotated with gene boxes and 1D chromatin tracks showing the ChIP-seq signal enrichment in the same region.

### **Fig 3. Enhancer and promoter proximity persists after the acute loss of loop extrusion factors.**

**a.** Scatter plots of loop scores for the called loops in the untreated and IAA-treated cells (left). The violin chart (inset) shows the distribution of loop scores for the untreated and IAA-treated conditions. APA are plotted with loops sorted by up-regulated, down-regulated, or unchanged loops (right). **b.** Enrichment of chromHMM states at loop anchors sorted by up-regulated, down-regulated, or unchanged after IAA treatment. **c-d.** Scatter plots of loop scores are plotted for paired E-P or P-P loops in the untreated and IAA-treated cells (left). APA are plotted by corresponding loop types in the untreated and IAA-treated cells (right). **e.** Length distribution of the unchanged or down-regulated E-P/P-P loops relative to TAD boundaries. **f.** Ratio of the unchanged or down-regulated E-P/P-P loop anchors located within  $\pm 1$ kb of TAD boundaries. **g.** Ratio of the unchanged or down-regulated E-P/P-P loop that can cross TAD boundaries. **h.** Cumulative distribution curves as a function of loop score for all loops or loop anchors overlapping with  $\pm 1$ kb of DEG promoters.

### **Fig 4. YY1 depletion does not immediately alter global gene expression and E-P or P-P proximity.**

**a.** Schematic for endogenous tagging for YY1 depletion and the results of immunoblots for YY1 and Actin-B. **b.** Heatmaps (left) and histogram profiles (right) of differential ChIP-seq signals for YY1, CTCF, and cohesin after YY1 depletion. **c.** Heatmaps (left) and histogram profiles (right) of differential ChIP-seq signals for YY1 around the four types of loop anchors. **d.** Overview of Micro-C contact maps at multiple resolutions in the untreated and IAA-treated cells. (left to right) Examples of Pearson's correlation matrices showing plaid-like chromosome compartments; contact matrices showing TADs along the diagonal; contacts matrices showing cohesin loops off the diagonal. **e.** APA are plotted with loops sorted by up-regulated, down-regulated, or unchanged. **f.** Volcano plot of total RNA-seq for YY1 depletion. DEGs ( $q$ -value  $< 0.01$ ) are colored by green (up) or pink (down). **g.** MA plots of nascent RNA-seq for YY1 depletion. DEGs with  $q$ -value  $< 0.05$  are colored by green (up) or pink (down). **h.** Overlap of DEGs in RNA-seq after YY1 depletion with 1) RNA-seq after 24-hour YY1 depletion, 2) genes associated with differential loops, and 3) a set of random genes. **i.** APA are plotted by corresponding loop types in the untreated and IAA-treated cells. **j.** Snapshots of Micro-C maps comparing chromatin interactions in the untreated (top) and IAA-treated (bottom) cells surrounding the *Ifnar2* or *Ikzf2* genes. Contact maps are annotated with gene boxes and genome browser tracks showing YY1 ChIP-seq signal enrichment and mNET-seq signals with the plus strand in blue and the negative strand in red.

### **Fig 5. Dynamics of YY1 protein binding.**

**a.** Schematic for endogenously tagging YY1 with HaloTag and the results of immunoblots for YY1 and TBP. **b.** Live-cell confocal imaging for HaloTag-YY1 stained with 500 nM TMR. White dashed lines label cells in interphase and blue dashed lines label mitotic cells. Scale bar: 10  $\mu$ m. **c.** Airyscan-resolved live-cell confocal imaging for YY1 ( $n=13$ ). Arrow points to sporadic loci within the nucleolus. Scale bar: 2  $\mu$ m. **d.** PALM imaging for YY1 ( $n=30$ ). Colormaps color the signal ranging from 0-100. Scale bar: 1  $\mu$ m. **e.** Overview of spaSPT with illumination patten and representative raw images for YY1 with tracking overlaid. Scale bar=1  $\mu$ m. **f.** The aggregate likelihood for diffusive YY1 molecules. The bar graph (top) shows the fractions of YY1 binned into bound, slow, and fast diffusing

subpopulations. (bottom) Estimation of YY1 diffusion coefficients by regular Brownian motion with marginalized localization errors. **g.** Immunoblots of proteins after a series of salt extractions. Cyt=cytoplasmic fraction; Ins=Insoluble fraction. **h.** Spatial reconstruction of spaSPT data for YY1's trajectory densities. YY1 trajectories are binned by diffusion coefficients as indicated. The bound trajectories are colored in red. The diffusing trajectories are colored in green and overlaid with the bound trajectories. The insets show the averaged signals for all identified clusters aggregating at the center of the plot with the same imaging overlays. Differential signals comparing each diffusing fraction to the bound fraction are plotted at the bottom. The divergent colormap shows a higher signal than the bound in red and lower than the bound in blue. **i.** FRAP analysis of YY1 bleached with a square spot. Error bars indicate the standard deviation of each acquired data point. **j.** Slow-SPT for measuring YY1's residence time (exposure time = 100 ms). Each data point indicates the unbinding rate of YY1 molecules in a single cell. The box plot shows the quartiles of data. **k.** Slow-SPT for measuring YY1's residence time with multiple exposure times.

**Fig 6. Cohesin depletion alters YY1's target search efficiency.** **a.** Schematics for endogenously tagging CTCF or cohesin with AID in the HaloTag-YY1 cell line (clone YN11) and immunoblots of CTCF, RAD21, and Actin-B. **b.** Airyscan-resolved live-cell confocal imaging for HaloTag-YY1 stained with 500 nM TMR in CTCF- or RAD21-depleted cells (n=6 for each depletion). Scale bar: 1  $\mu$ m. **c.** PALM imaging for YY1 (n=13 for each depletion). Colormaps color the signal ranging from 0-40. Scale bar: 1  $\mu$ m. **d.** The stacked bar graph shows the fractions of bound, slow, and fast diffusing YY1 in the untreated and IAA-treated cells, obtained by Spagl analysis. **e.** FRAP analysis of YY1 in the control, CTCF-depleted, or RAD21-depleted cells. **f.** Heatmaps (left) and histogram profiles (right) of differential ChIP-seq signals for YY1 after CTCF, RAD21, or WAPL depletion. **g.** Dynamic model of how cohesin or cohesin-mediated structures may regulate TF target search.

**Fig 7. Models of enhancer-promoter interactions and transcription in the context of 3D genome organization.**

## Acknowledgment

T.S.H., C.C., and E.S. conceived and designed the project. C.C. performed all biochemical and ChIP-seq experiments and analysis with E.S.'s assistance. E.S. generated all plasmids and cell lines with T.S.H.'s and Gina M. Dailey's assistance. C.C. and A.S.H. generated the parental degradation cell lines. T.S.H. performed Micro-C assays and analysis. E.S. and T.S.H. performed all imaging experiments and analyses. T.S.H. drafted the manuscript, and C.C., E.S., A.S.H., R.T. edited the manuscript. R.T. and X.D. supervised the project. Funding: This work was supported by the California Institute of Regenerative Medicine grant LA1-08013 (X.D.), the Koret UC Berkeley-Tel Aviv University Initiative grant (T.S.H. & X.D.), and by the Howard Hughes Medical Institute (003061, R.T.). T.S.H. is a postdoc fellow of the Koret UC Berkeley-Tel Aviv University Initiative. E.S. was an undergraduate fellow of SURF Rose Hills Independent at UC Berkeley. A.S.H. acknowledges support from the National Institutes of Health under Grant R00GM130896. We thank all members of the Tjian and Darzacq lab for comments on the manuscript.

## References

1. Jerkovic', I. & Cavalli, G. Understanding 3D genome organization by multidisciplinary methods. *Nat Rev Mol Cell Bio* 1–18 (2021) doi:10.1038/s41580-021-00362-w.
2. Beagan, J. A. & Phillips-Cremins, J. E. On the existence and functionality of topologically associating domains. *Nat Genet* **52**, 8–16 (2020).
3. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
4. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
5. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
6. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).
7. Hsieh, T.-H. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol Cell* **78**, 539-553.e8 (2020).
8. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet* **47**, 598–606 (2015).
9. Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res* **25**, 582–97 (2015).
10. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e22 (2017).
11. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).
12. Wutz, G. *et al.* Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *Embo J* **36**, 3573–3599 (2017).
13. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
14. Haarhuis, J. H. I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693-707.e14 (2017).



- 732 15. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports*  
733 **15**, 2038–2049 (2016).
- 734 16. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain  
735 formation in wild-type and engineered genomes. *Proc National Acad Sci* **112**, E6456–E6465  
736 (2015).
- 737 17. Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Mol*  
738 *Cell* **78**, 554–565.e7 (2020).
- 739 18. Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via Cellular  
740 Genetic Screens. *Cell* **176**, 377–390.e19 (2019).
- 741 19. Zheng, H. & Xie, W. The role of 3D genome organization in development and cell  
742 differentiation. *Nat Rev Mol Cell Bio* **20**, 535–550 (2019).
- 743 20. Robson, M. I., Ringel, A. R. & Mundlos, S. Regulatory Landscaping: How Enhancer-  
744 Promoter Communication Is Sculpted in 3D. *Mol Cell* **74**, 1110–1122 (2019).
- 745 21. Anania, C. *et al.* In vivo dissection of a clustered-CTCF domain boundary reveals  
746 developmental principles of regulatory insulation. *Biorxiv* 2021.04.14.439779 (2021)  
747 doi:10.1101/2021.04.14.439779.
- 748 22. Calderon, L. *et al.* Activity-induced gene expression and long-range enhancer-promoter  
749 contacts in cohesin-deficient neurons. *Biorxiv* 2021.02.24.432639 (2021)  
750 doi:10.1101/2021.02.24.432639.
- 751 23. Stik, G. *et al.* CTCF is dispensable for immune cell transdifferentiation but facilitates an  
752 acute inflammatory response. *Nat Genet* **52**, 655–661 (2020).
- 753 24. Hsieh, T.-H. S. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by  
754 Micro-C. *Cell* **162**, 108–119 (2015).
- 755 25. Hsieh, T.-H. S., Fudenberg, G., Goloborodko, A. & Rando, O. J. Micro-C XL: assaying  
756 chromosome conformation from the nucleosome to the entire genome. *Nat Methods* **13**, 1009–  
757 1011 (2016).
- 758 26. Costantino, L., Hsieh, T.-H. S., Lamothe, R., Darzacq, X. & Koshland, D. Cohesin residency  
759 determines chromatin loop patterns. *Elife* **9**, e59889 (2020).
- 760 27. Hansen, A. S. *et al.* Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-  
761 Binding Region in CTCF. *Mol Cell* **76**, 395–411.e13 (2019).
- 762 28. Nojima, T. *et al.* Mammalian NET-Seq Reveals Genome-wide Nascent Transcription  
763 Coupled to RNA Processing. *Cell* **161**, 526–540 (2015).

764 29. Ardakany, A. R., Gezer, H. T., Lonardi, S. & Ay, F. Mustache: multi-scale detection of  
765 chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biol*  
766 **21**, 256 (2020).

767 30. Matthey-Doret, C. *et al.* Computer vision for pattern detection in chromosome contact maps.  
768 *Nat Commun* **11**, 5795 (2020).

769 31. Gu, B. *et al.* Opposing Effects of Cohesin and Transcription on CTCF Organization Revealed  
770 by Super-resolution Imaging. *Mol Cell* **80**, 699–711.e7 (2020).

771 32. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative  
772 interactions in single cells. *Science* **362**, eaau1783 (2018).

773 33. Flyamer, I. M. *et al.* Single-nucleus Hi-C reveals unique chromatin reorganization at oocyte-  
774 to-zygote transition. *Nature* **544**, 110–114 (2017).

775 34. Luppino, J. M. *et al.* Cohesin promotes stochastic domain intermingling to ensure proper  
776 regulation of boundary-proximal genes. *Nat Genet* **52**, 840–848 (2020).

777 35. Szabo, Q. *et al.* Regulation of single-cell genome organization into TADs and chromatin  
778 nanodomains. *Nat Genet* **52**, 1151–1157 (2020).

779 36. Natsume, T., Kiyomitsu, T., Saga, Y. & Kanemaki, M. T. Rapid Protein Depletion in Human  
780 Cells by Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Reports* **15**, 210–  
781 218 (2016).

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

784 38. Jaeger, M. G. & Winter, G. E. Fast-acting chemical tools to delineate causality in  
785 transcriptional control. *Mol Cell* (2021) doi:10.1016/j.molcel.2021.02.015.

786 39. Pugacheva, E. M. *et al.* CTCF mediates chromatin looping via N-terminal domain-dependent  
787 cohesin retention. *Proc National Acad Sci* **117**, 2020–2031 (2020).

788 40. Tu, S. *et al.* MAnorm2 for quantitatively comparing groups of ChIP-seq samples. *Genome*  
789 *Res* **31**, 131–145 (2020).

790 41. Busslinger, G. A. *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF  
791 and Wapl. *Nature* **544**, 503–507 (2017).

792 42. Li, Y. *et al.* The structural basis for cohesin–CTCF-anchored loops. *Nature* **578**, 472–476  
793 (2020).

794 43. Lengronne, A. *et al.* Cohesin relocation from sites of chromosomal loading to places of  
795 convergent transcription. *Nature* **430**, 573–578 (2004).

796 44. Liu, N. Q. *et al.* WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal  
797 gene regulation. *Nat Genet* **53**, 100–109 (2021).

798 45. Luan, J. *et al.* Distinct properties and functions of CTCF revealed by a rapidly inducible  
799 degron system. *Cell Reports* **34**, 108783 (2021).

800 46. Owens, N. *et al.* CTCF confers local nucleosome resiliency after DNA replication and during  
801 mitosis. *Elife* **8**, e47898 (2019).

802 47. Sönmezer, C. *et al.* Molecular Co-occupancy Identifies Transcription Factor Binding  
803 Cooperativity In Vivo. *Mol Cell* (2020) doi:10.1016/j.molcel.2020.11.015.

804 48. Sandberg, M., Källström, M. & Muhr, J. Sox21 promotes the progression of vertebrate  
805 neurogenesis. *Nat Neurosci* **8**, 995–1001 (2005).

806 49. Bylund, M., Andersson, E., Novitch, B. G. & Muhr, J. Vertebrate neurogenesis is  
807 counteracted by Sox1–3 activity. *Nat Neurosci* **6**, 1162–1168 (2003).

808 50. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced  
809 pluripotent stem cells. *Nature* **448**, 313–317 (2007).

810 51. Jacquet, B. V. *et al.* FoxJ1-dependent gene expression is required for differentiation of radial  
811 glia into ependymal cells and a subset of astrocytes in the postnatal brain. *Development* **136**,  
812 4021–4031 (2009).

813 52. Hassan, M. Q. *et al.* Dlx3 Transcriptional Regulation of Osteoblast Differentiation: Temporal  
814 Recruitment of Msx2, Dlx3, and Dlx5 Homeodomain Proteins to Chromatin of the Osteocalcin  
815 Gene. *Mol Cell Biol* **24**, 9248–9261 (2004).

816 53. Wang, H. F. *et al.* Defining Essential Enhancers for Pluripotent Stem Cells Using a Features-  
817 Oriented CRISPR-Cas9 Screen. *Cell Reports* **33**, 108309 (2020).

818 54. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat Rev*  
819 *Genet* **19**, 789–800 (2018).

820 55. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin  
821 architecture. *Nature* **467**, 430–435 (2010).

822 56. Phillips-Cremins, J. E. *et al.* Architectural Protein Subclasses Shape 3D Organization of  
823 Genomes during Lineage Commitment. *Cell* **153**, 1281–1295 (2013).

824 57. Khattabi, L. E. *et al.* A Pliable Mediator Acts as a Functional Rather Than an Architectural  
825 Bridge between Promoters and Enhancers. *Cell* **178**, 1145–1158.e20 (2019).

826 58. Jaeger, M. G. *et al.* Selective Mediator dependence of cell-type-specifying transcription. *Nat*  
827 *Genet* 1–9 (2020) doi:10.1038/s41588-020-0635-0.

828 59. Donohoe, M. E. *et al.* Targeted Disruption of Mouse Yin Yang 1 Transcription Factor  
829 Results in Peri-Implantation Lethality. *Mol Cell Biol* **19**, 7237–7244 (1999).

830 60. Gabriele, M. *et al.* YY1 Haploinsufficiency Causes an Intellectual Disability Syndrome  
831 Featuring Transcriptional and Chromatin Dysfunction. *Am J Hum Genet* **100**, 907–925 (2017).

832 61. Beagan, J. A. *et al.* YY1 and CTCF orchestrate a 3D chromatin looping switch during early  
833 neural lineage commitment. *Genome Res* **27**, 1139–1152 (2017).

834 62. Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**,  
835 1573–1588.e28 (2017).

836 63. Perreault, A. A., Brown, J. D. & Venters, B. J. Erythropoietin Regulates Transcription and  
837 YY1 Dynamics in a Pre-established Chromatin Architecture. *Isience* **23**, 101583 (2020).

838 64. Li, S., Prasanna, X., Salo, V. T., Vattulainen, I. & Ikonen, E. An efficient auxin-inducible  
839 degron system with low basal degradation in human cells. *Nat Methods* **16**, 866–869 (2019).

840 65. Los, G. V. *et al.* HaloTag: A Novel Protein Labeling Technology for Cell Imaging and  
841 Protein Analysis. *Acs Chem Biol* **3**, 373–382 (2008).

842 66. White, R. J., Jackson, S. P. & Rigby, P. W. A role for the TATA-box-binding protein  
843 component of the transcription factor IID complex as a general RNA polymerase III transcription  
844 factor. *Proc National Acad Sci* **89**, 1949–1953 (1992).

845 67. Kieffer-Kwon, P., Martianov, I. & Davidson, I. Cell-specific Nucleolar Localization of TBP-  
846 related Factor 2. *Mol Biol Cell* **15**, 4356–4368 (2004).

847 68. Rizkallah, R. & Hurt, M. M. Regulation of the Transcription Factor YY1 in Mitosis through  
848 Phosphorylation of Its DNA-binding Domain. *Mol Biol Cell* **20**, 4766–4776 (2009).

849 69. Raccaud, M. *et al.* Mitotic chromosome binding predicts transcription factor properties in  
850 interphase. *Nat Commun* **10**, 487 (2019).

851 70. Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate  
852 chromatin loop stability with distinct dynamics. *Elife* **6**, e25776 (2017).

853 71. Hansen, A. S. *et al.* Robust model-based analysis of single-particle tracking experiments with  
854 Spot-On. *Elife* **7**, e33125 (2018).

855 72. Heckert, A., Dahal, L., Tjian, R. & Darzacq, X. Recovering mixtures of fast diffusing states  
856 from short single particle trajectories. *Biorxiv* 2021.05.03.442482 (2021)  
857 doi:10.1101/2021.05.03.442482.

858 73. Hansen, A. S., Amitai, A., Cattoglio, C., Tjian, R. & Darzacq, X. Guided nuclear exploration  
859 increases CTCF target search efficiency. *Nat Chem Biol* **16**, 257–266 (2020).

860 74. Lionnet, T. & Wu, C. Single-molecule tracking of transcription protein dynamics in living  
861 cells: seeing is believing, but what are we seeing? *Curr Opin Genet Dev* **67**, 94–102 (2021).

862 75. Loffreda, A. *et al.* Live-cell p53 single-molecule binding is modulated by C-terminal  
863 acetylation and correlates with transcriptional activity. *Nat Commun* **8**, 313 (2017).

864 76. Trojanowski, J., Frank, L., Rademacher, A., Grigaitis, P. & Rippe, K. Transcription  
865 activation is enhanced by multivalent interactions independent of phase separation. *Biorxiv*  
866 2021.01.27.428421 (2021) doi:10.1101/2021.01.27.428421.

867 77. Mueller, F., Mazza, D., Stasevich, T. J. & McNally, J. G. FRAP and kinetic modeling in the  
868 analysis of nuclear protein dynamics: what do we really know? *Curr Opin Cell Biol* **22**, 403–411  
869 (2010).

870 78. Sprague, B. L., Pego, R. L., Stavreva, D. A. & McNally, J. G. Analysis of Binding Reactions  
871 by Fluorescence Recovery after Photobleaching. *Biophys J* **86**, 3473–3495 (2004).

872 79. Chen, J. *et al.* Single-molecule dynamics of enhanceosome assembly in embryonic stem  
873 cells. *Cell* **156**, 1274–85 (2014).

874 80. Garcia, D. A. *et al.* Power-law behavior of transcription factor dynamics at the single-  
875 molecule level implies a continuum affinity model. *Nucleic Acids Res* gkab072- (2021)  
876 doi:10.1093/nar/gkab072.

877 81. Thiecke, M. J. *et al.* Cohesin-Dependent and -Independent Mechanisms Mediate  
878 Chromosomal Contacts between Promoters and Enhancers. *Cell Reports* **32**, 107929 (2020).

879 82. Symmons, O. *et al.* Functional and topological characteristics of mammalian regulatory  
880 domains. *Biotechfor* **24**, 390–400 (2014).

881 83. Rodríguez-Carballo, E. *et al.* Chromatin topology and the timing of enhancer function at the  
882 HoxD locus. *Proc National Acad Sci* **117**, 31231–31241 (2020).

883 84. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase  
884 Separation Model for Transcriptional Control. *Cell* **169**, 13–23 (2017).

885 85. Sabari, B. R. *et al.* Coactivator condensation at super-enhancers links phase separation and  
886 gene control. *Science* **361**, eaar3958 (2018).

887 86. Boija, A. *et al.* Transcription Factors Activate Genes through the Phase-Separation Capacity  
888 of Their Activation Domains. *Cell* **175**, 1842–1855.e16 (2018).

889 87. Cho, W.-K. *et al.* Mediator and RNA polymerase II clusters associate in transcription-  
890 dependent condensates. *Science* **361**, 412–415 (2018).



88. Crump, N. T. *et al.* BET inhibition disrupts transcription but retains enhancer-promoter contact. *Nat Commun* **12**, 223 (2021).

89. Bintu, L. *et al.* Dynamics of epigenetic regulation at the single-cell level. *Science* **351**, 720–724 (2016).

90. Keung, A. J., Bashor, C. J., Kiriakov, S., Collins, J. J. & Khalil, A. S. Using Targeted Chromatin Regulators to Engineer Combinatorial and Spatial Transcriptional Regulation. *Cell* **158**, 110–120 (2014).

91. Stevens, T. J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59–64 (2017).

92. Basu, S. *et al.* Live-cell 3D single-molecule tracking reveals how NuRD modulates enhancer dynamics. *Biorxiv* 2020.04.03.003178 (2020) doi:10.1101/2020.04.03.003178.

93. Shipony, Z. *et al.* Dynamic and static maintenance of epigenetic memory in pluripotent and somatic cells. *Nature* **513**, 115–119 (2014).

94. Meir, Z., Mukamel, Z., Chomsky, E., Lifshitz, A. & Tanay, A. Single-cell analysis of clonal maintenance of transcriptional and epigenetic states in cancer cells. *Nat Genet* **52**, 709–718 (2020).

95. Nuñez, J. K. *et al.* Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. *Cell* (2021) doi:10.1016/j.cell.2021.03.025.

96. Engreitz, J. M. *et al.* The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. *Science* **341**, 1237973 (2013).

97. Minajigi, A. *et al.* A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* **349**, aab2276 (2015).

98. Hou, C., Zhao, H., Tanimoto, K. & Dean, A. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc National Acad Sci* **105**, 20398–20403 (2008).

99. Huang, H. *et al.* CTCF mediates dosage- and sequence-context-dependent transcriptional insulation by forming local chromatin domains. *Nat Genet* 1–11 (2021) doi:10.1038/s41588-021-00863-6.

100. Zuin, J. *et al.* Nonlinear control of transcription through enhancer-promoter interactions. *Biorxiv* 2021.04.22.440891 (2021) doi:10.1101/2021.04.22.440891.

101. Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).

922 102. Flavahan, W. A. *et al.* Altered chromosomal topology drives oncogenic programs in SDH-  
923 deficient GISTs. *Nature* **575**, 229–233 (2019).

924 103. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic  
925 Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025 (2015).

926 104. Ushiki, A. *et al.* Deletion of CTCF sites in the SHH locus alters enhancer–promoter  
927 interactions and leads to acheiropodia. *Nat Commun* **12**, 2282 (2021).

928 105. Despang, A. *et al.* Functional dissection of the Sox9–Kcnj2 locus identifies nonessential  
929 and instructive roles of TAD architecture. *Nat Genet* **51**, 1263–1271 (2019).

930 106. Kraft, K. *et al.* Serial genomic inversions induce tissue-specific architectural stripes, gene  
931 misexpression and congenital malformations. *Nat Cell Biol* **21**, 305–310 (2019).

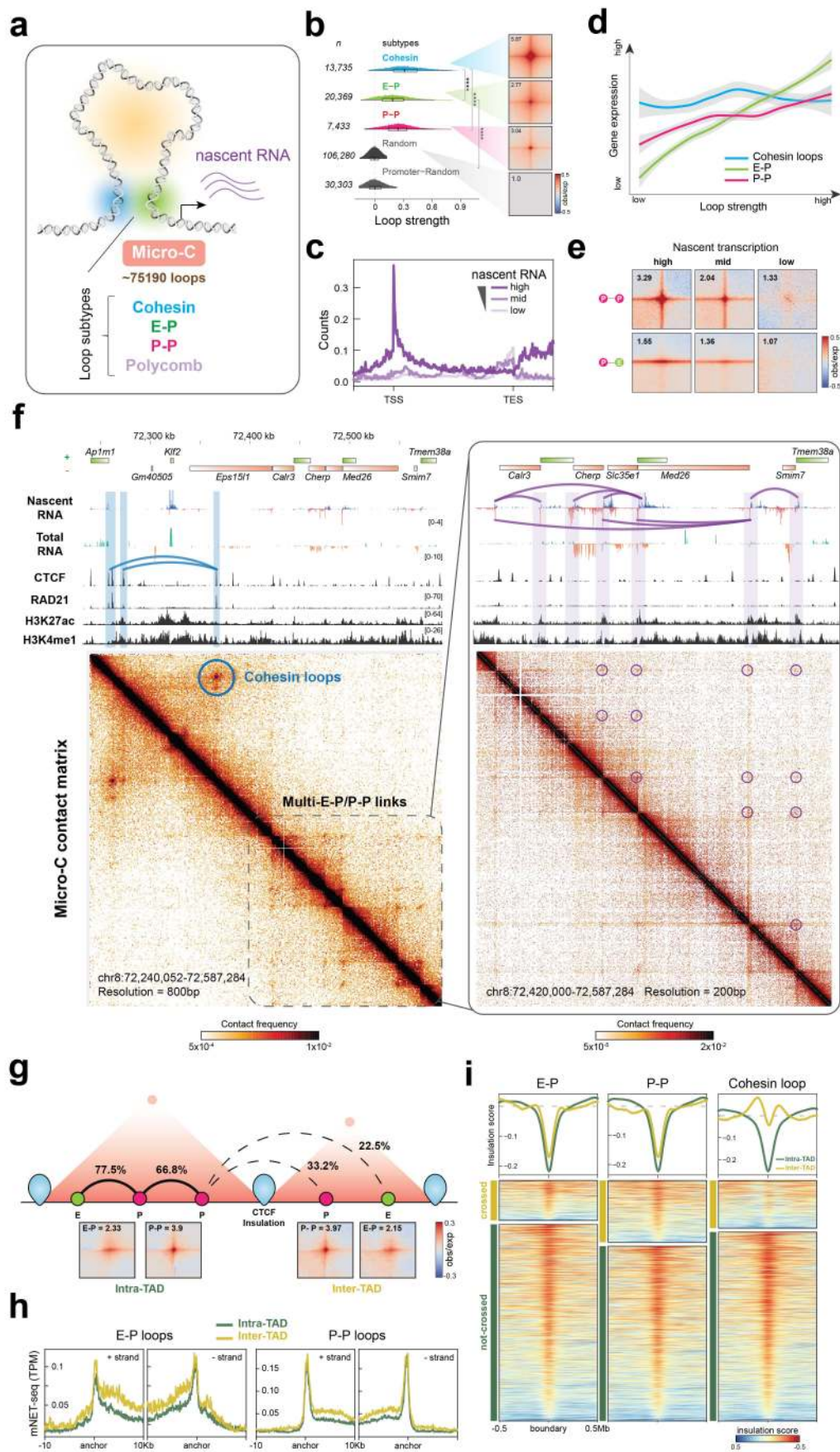
932 107. Xiao, J., Hafner, A. & Boettiger, A. N. How subtle changes in 3D structure can create large  
933 changes in transcription. *Biorxiv* 2020.10.22.351395 (2020) doi:10.1101/2020.10.22.351395.

934 108. Brodsky, S. *et al.* Intrinsically Disordered Regions Direct Transcription Factor In Vivo  
935 Binding Specificity. *Mol Cell* (2020) doi:10.1016/j.molcel.2020.05.032.

936 109. Garcia, D. A. *et al.* An intrinsically disordered region-mediated confinement state  
937 contributes to the dynamics and function of transcription factors. *Mol Cell* (2021)  
938 doi:10.1016/j.molcel.2021.01.013.

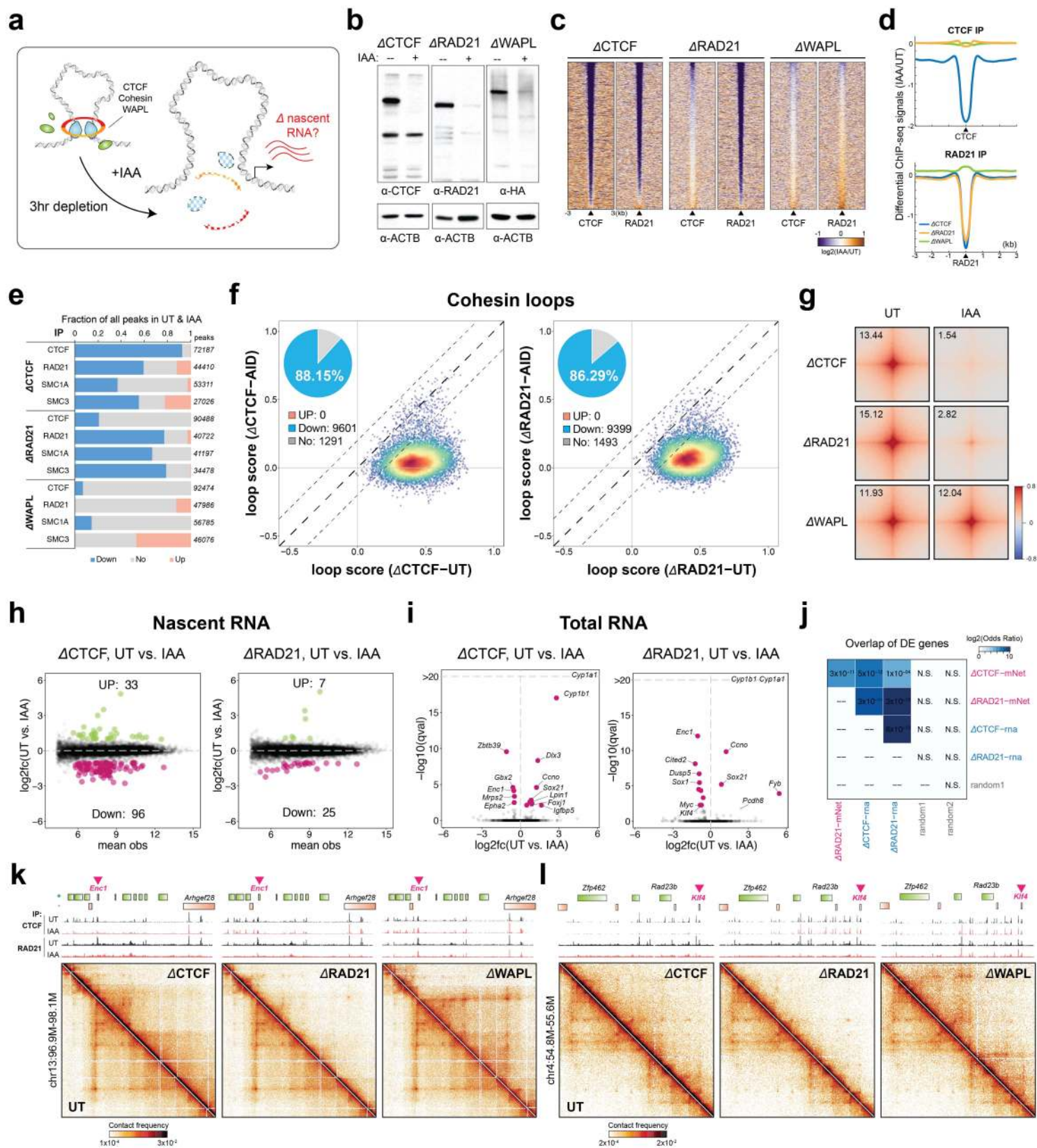
939

# Figure 1

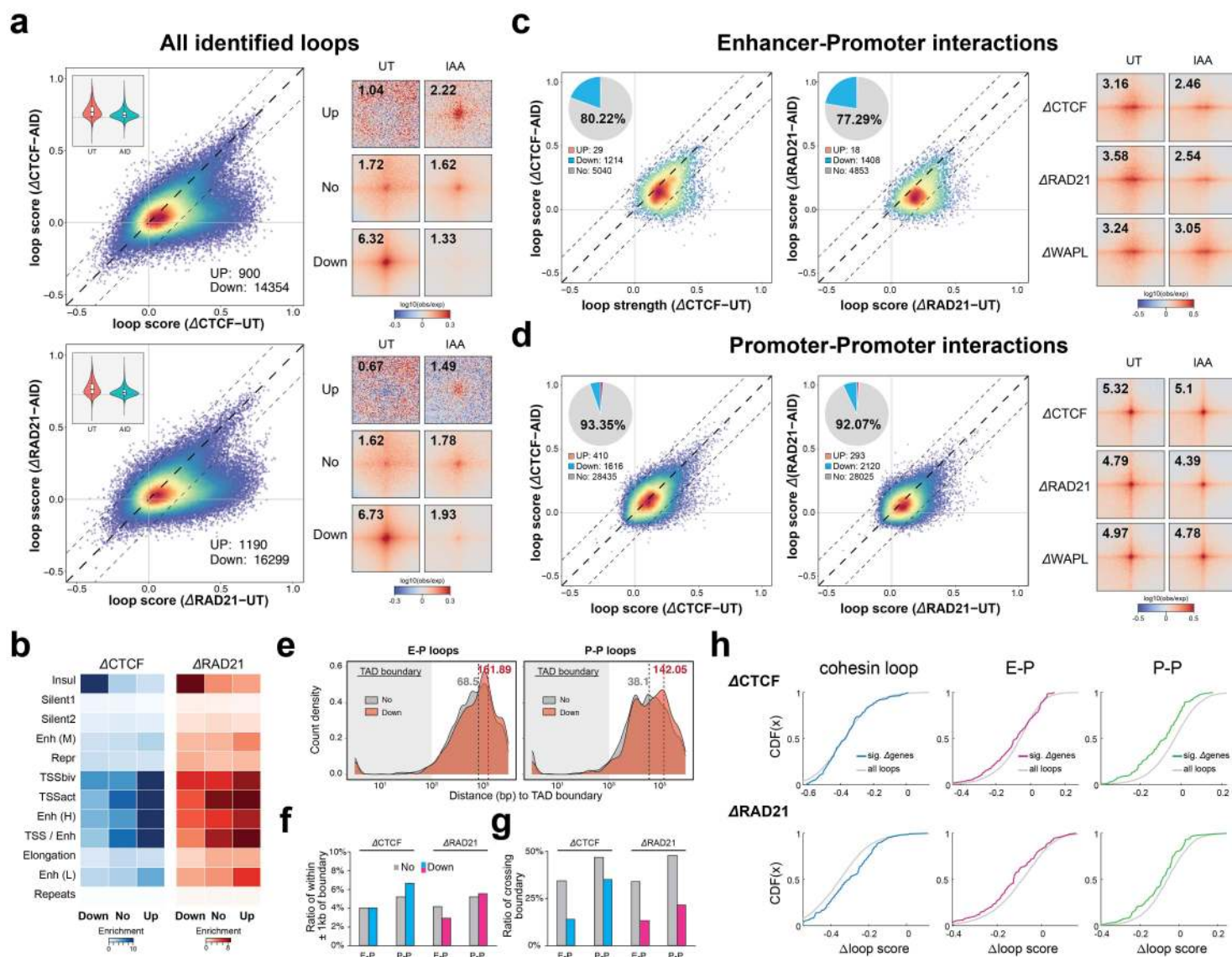




## Figure 2

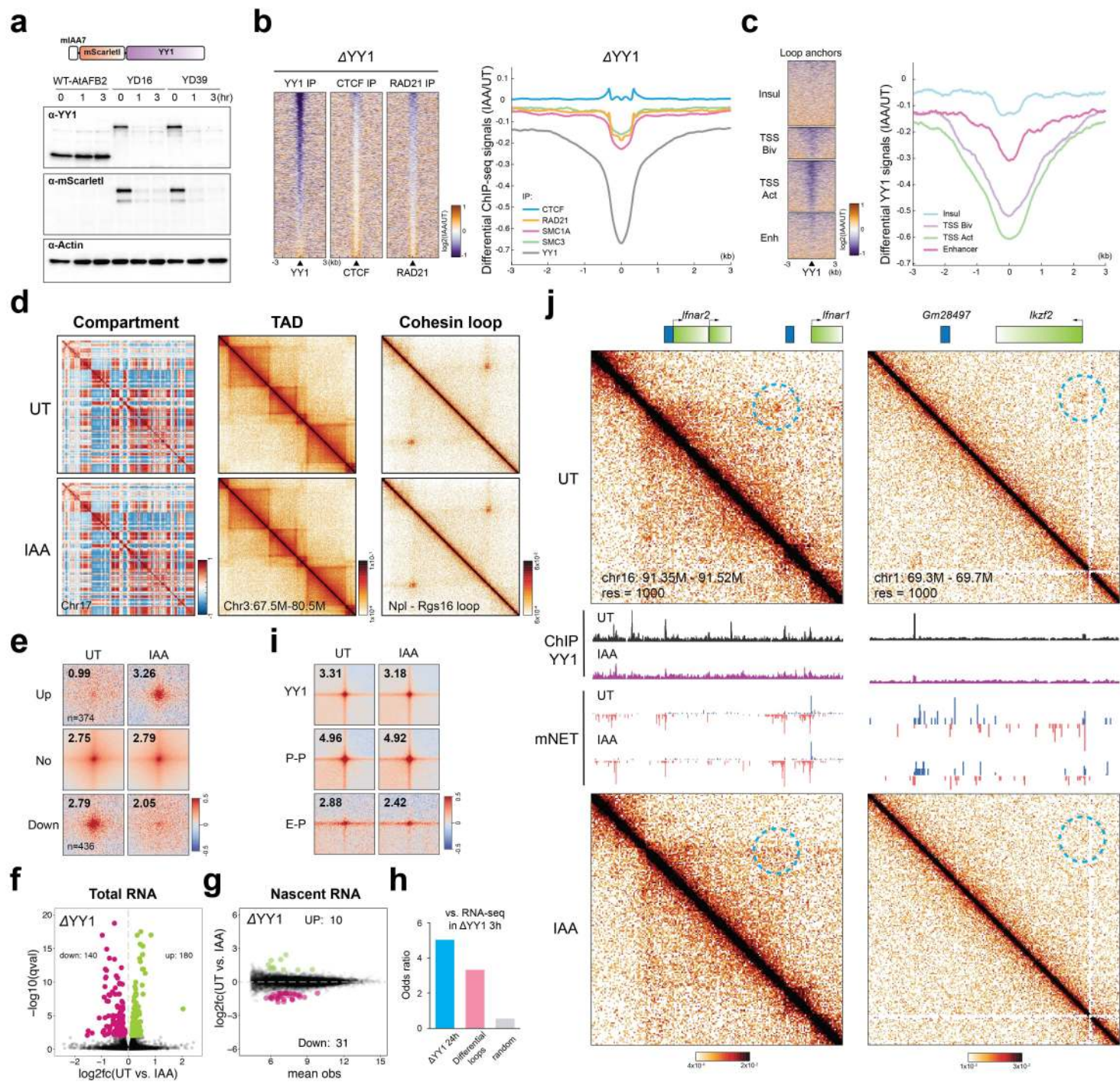


# Figure 3

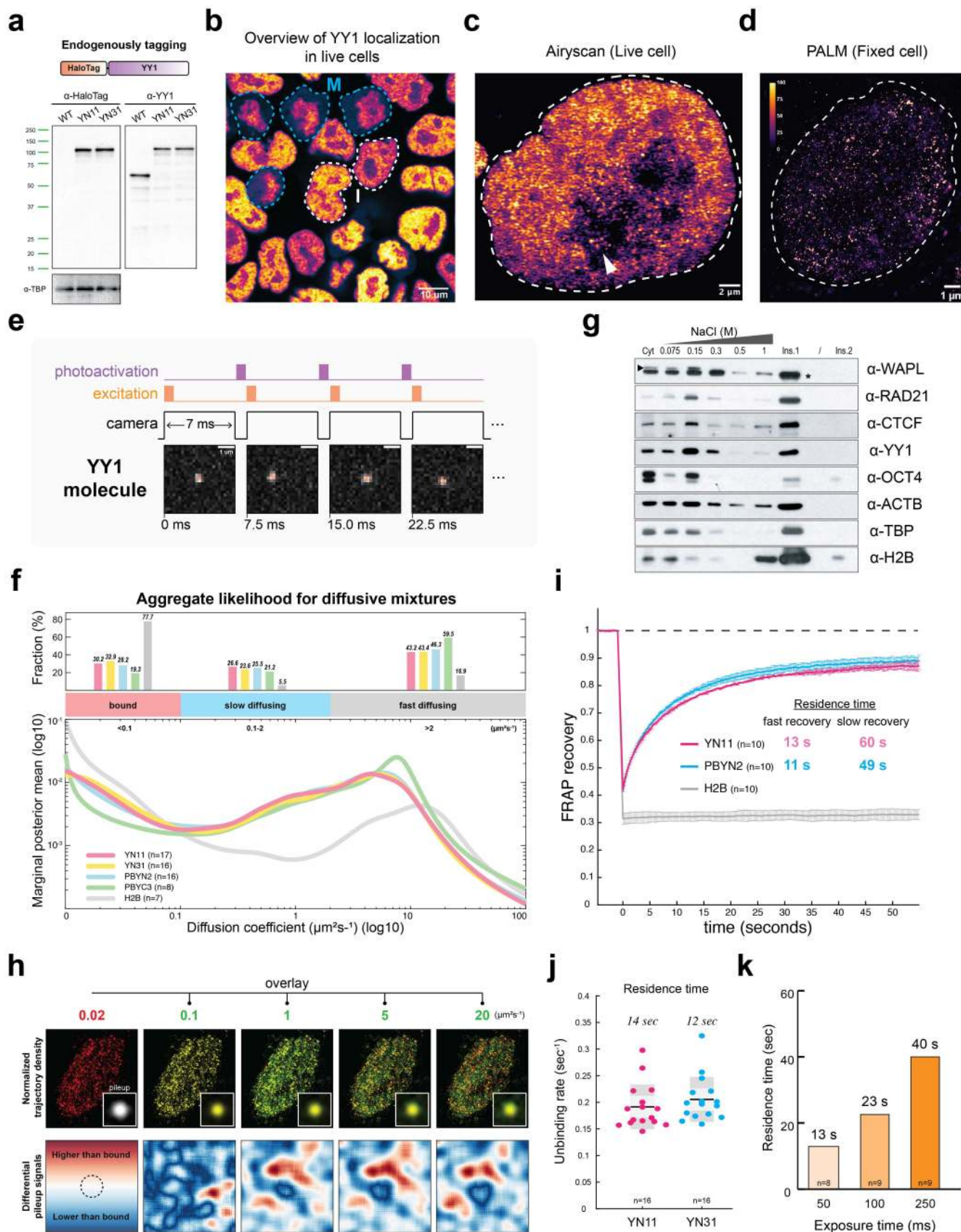




## Figure 4

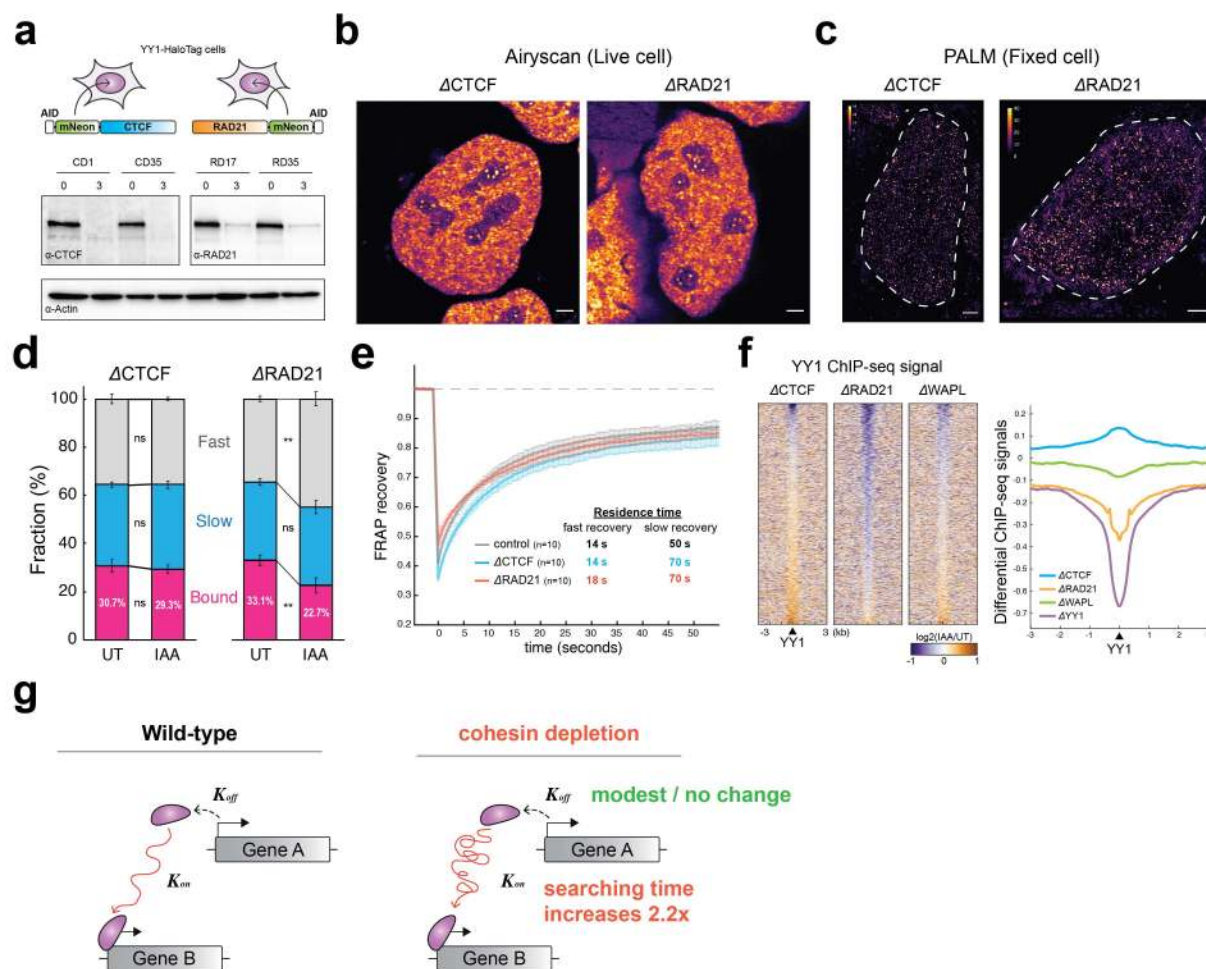


## Figure 5

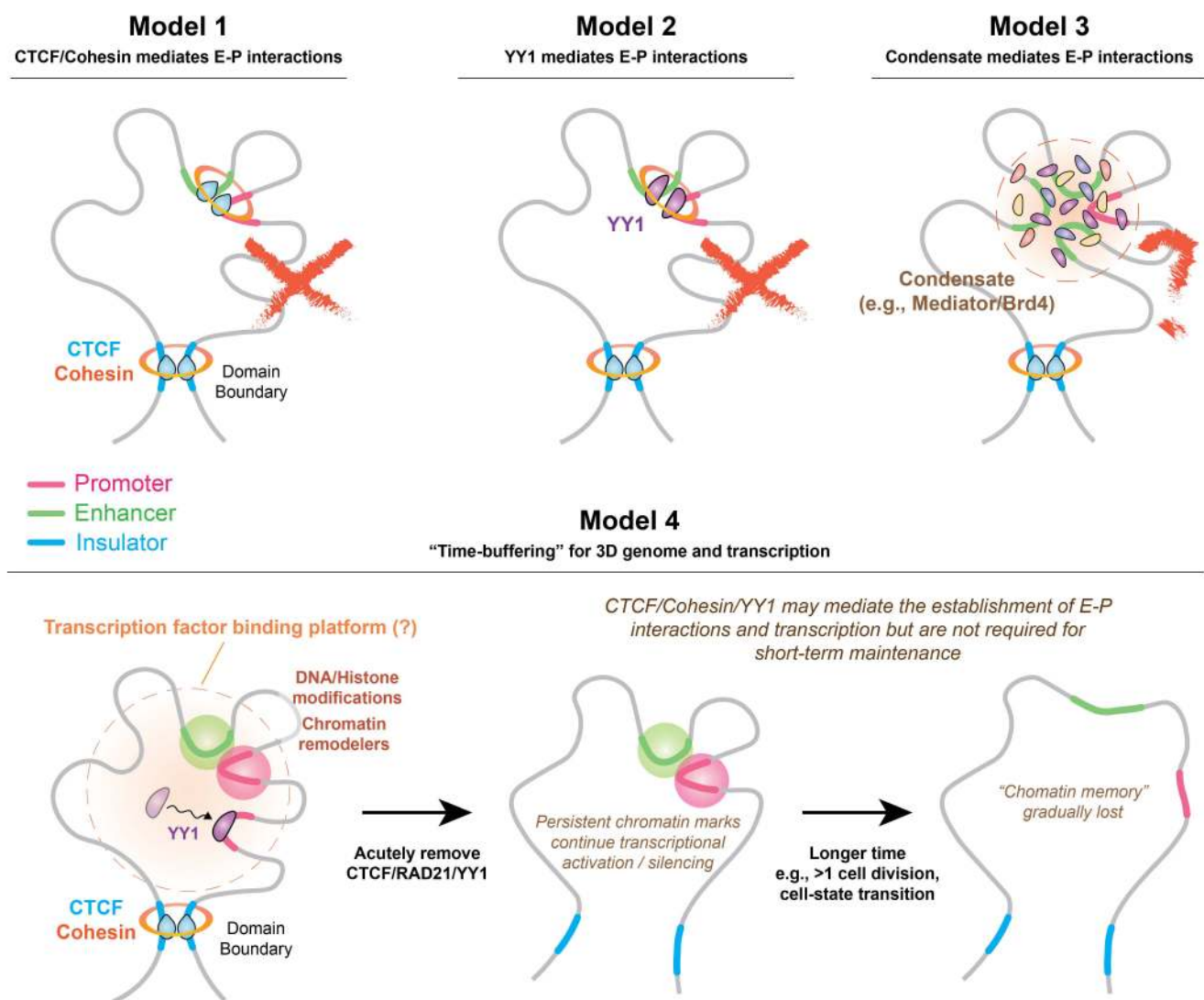




## Figure 6



## Figure 7



# Extended Data Figure Legend

## Extended Data Fig 1.

- a. Enrichment of mESC ChromHMM<sup>1</sup> states at loop anchors. The heatmap shows the log<sub>2</sub> enrichment of each state in a  $\pm 5$ -kb window around the loop anchors (loop number = 75,190 and loop anchor number = 118,733 after removing duplicates).
- b. Box plots of the loop strengths for cohesin, E-P, P-P, and random loops. Paired CTCF/cohesin, E-P, and P-P loci were quantified using Chromosight<sup>2</sup> with Micro-C data at 2-kb or 4-kb resolution, which produced comparable but not identical results as Mustache<sup>3</sup> in **Fig. 1b**. This loop list was used for the analysis in **Extended Data Fig. 1b-d**. The numbers of chromatin loops are shown on the left. The box plot indicates the quartiles for the loop strength score distribution. The genome-wide averaged contact signals (aggregate peak analysis (APA)) are plotted on the right. The contact map was normalized by matrix balancing and distance, with positive enrichment in red and negative signal in blue, shown as a diverging colormap with the gradient of normalized contact enrichment in log<sub>10</sub>. The ratio of contact enrichment for the center pixels is annotated within each plot. Asterisks denote a p-value  $< 10^{-16}$  by the Wilcoxon test.
- c. Box plots of the loop length distributions for cohesin, E-P, P-P, and random loops. The colored box represents the result using the 2-kb resolution Micro-C data and the white box using the 4-kb resolution data. The median size of loops is annotated on the right. We note that the median lengths are larger than our previous analysis with the insulation score due to the high computational expense to quantify the short-range loops with Micro-C data finer than 1-kb resolution. Asterisks denote a p-value  $< 10^{-16}$  by the Wilcoxon test.
- d. Summary of loop calling for cohesin, E-P, and P-P interactions with Micro-C (pink) or Hi-C data from Bonev et al.<sup>4</sup> (gray).
- e. Gel image representing the size distribution of the mNET-seq library on a 6% PAGE gel and the resolved bands on the Fragment Analyzer electropherogram.
- f. Heatmap of Pearson's correlation between the sequencing data (ATAC-seq, RNA-seq, Pol II ChIP-seq, and promoter pausing (pp) / gene body (gb) signal in mNET-seq) and the four levels of chromatin structures (compartment, TADs, cohesin loops, E-P and P-P loops) by Micro-C or Hi-C.
- g. Snapshots of Hi-C data in the same region as shown in **Fig. 1f**.
- h. Heatmaps of mNET-seq and ChIP-seq signal enrichments in a  $\pm 3$ -kb window around the four primary types of loop anchors.
- i. Rank-ordered distribution of loop length against loop strength for cohesin, E-P, P-P, and random loops. The distribution for each loop type was fitted and smoothed by LOESS regression. The standard error (SE) is plotted as a gray shade along with the regression line.
- j. Profiles of ChIP-seq signal for CTCF and RAD21 at TAD boundaries grouped by intra-TAD (green) / inter-TAD (gold) cohesin, E-P, or P-P loops.

## Extended Data Fig 2.

- a. Schematics for endogenously tagging CTCF, RAD21, WAPL with the mAID degron and for endogenously tagging YY1 with miniIAA7.
- b. Immunoblots of CTCF, RAD21, WAPL, YY1, and their tags (HaloTag for CTCF, V5 for RAD21, RFP (mScarletI) for YY1, and HA for WAPL) for the protein expression levels and sizes in wildtype mESCs and in clones C58, F1, YD39, and C40.



- c. Immunoblots of CTCF, RAD21 and WAPL's tag across a degradation time course from 0 (untreated) to 3 hr (IAA treatment) in clones C58, F1, and C40.
- d. Heatmap of 'all-by-all' Spearman's correlation for all ChIP-seq replicates samples (n = 96).
- e. Heatmap of Jaccard's index for the ratio of co-enriched peaks between the ChIP-seq replicates.
- f. Heatmaps of CTCF and cohesin (RAD21, SMC1A, and SMC3 subunits) ChIP-seq signal around WT-CTCF peaks called by MACS2 in the CTCF-, RAD21-, or WAPL-degron cells. The peaks called by MACS2 are plotted at the center across a  $\pm 3$ -kb region. The colormap shows the maximum signal (log2) in blue and the minimum signal in white.
- g. Heatmaps of differential ChIP-seq signals for SMC1A and SMC3 in cells depleted for CTCF, RAD21, or WAPL. The peaks called by MACS2 are plotted at the center across a  $\pm 3$ -kb region. The colormap shows an increased signal (log2) in orange and a decreased signal in purple after IAA treatment.
- h. MA plots show the differential ChIP-seq peaks between the UT and IAA-treated cells. The significantly changed peaks (adjusted  $p$ -value < 0.05) are colored in red. X-axis: mean observations of UT and IAA cells. Y-axis: log2 fold-change comparing the UT and IAA-treated cells.
- i. Snapshot of an example showing the position and level of cohesin at promoters (red shade) after CTCF degradation.
- j. MA plots showing the differential ChIP-seq peaks of RAD21 and SMC3 in the CTCF-depleted cells. The differential peaks (adjusted  $p$ -value < 0.05) are colored in red. X-axis: mean observations of UT and IAA cells. Y-axis: log2 fold-change comparing the UT and IAA-treated cells.
- k. Profiles of genome-wide averaged ChIP-seq for CTCF, RAD21, SMC1A, and input around the  $\pm 3$ -kb region of SOX2 or OCT4 peaks in WAPL-depleted cells.

### Extended Data Fig 3.

- a. Summary of Micro-C experiments for CTCF, RAD21, and WAPL depletion. Total unique reads are annotated for each replicate on the right, consisting of trans-interactions (inter-chromosome) in green, short-range cis-interactions (< 20 kb) in orange, and long-range cis-interactions (> 20 kb) in red.
- b. Summary of reproducibility tests for Micro-C. Similarity scores measured by QuASAR (top panel) or GenomeDisco (bottom panel) either between replicates (light lines) or comparing the UT and IAA-treated samples (dark lines) using 250-kb, 50-kb, 25-kb, and 10-kb resolution of Micro-C matrices.
- c. Slope distribution of  $P(s)$  curves for the UT and IAA-treated cells. Dashed lines highlight the range of genome distances affected by CTCF, RAD21, or WAPL depletion.
- d. Scatter plot of loop scores for cohesin loops in the UT and IAA-treated cells. The overlaid heatmap indicates dot density, with the highest in red and the lowest in blue. Dashed lines along the diagonal mark the range that is characterized as unchanged loops.
- e. Summary of loop numbers called by Mustache for the UT (gray) and IAA-treated (pink) cells. The additional loops (n = 5764) identified after WAPL depletion show longer length with the median at 570 kb.
- f. Profile of differential ChIP-seq signal for CTCF in a  $\pm 3$ -kb window around loop anchors (gray) or non-loop anchors (blue).
- g. APA for loops where CTCF is bound (left) or lost (right) at the anchors during mitosis<sup>5</sup>.

- h. APA for loops with either high (left) or low (right) CTCF occupancy characterized by single-molecular foot-printing (SMF) assay at their anchors<sup>6</sup>.
- i. MA plot of the mNET-seq result for WAPL depletion. DEGs with  $q$ -value  $< 0.05$  are labeled with green (up) or pink (down).
- j. Volcano plot of total RNA-seq for WAPL depletion. DEGs ( $q$ -value  $< 0.01$ ) are labeled with pink with the corresponding gene name.

#### Extended Data Fig 4.

- a. Heatmaps of differential ChIP-seq signals for CTCF, RAD21, SMC1A, and SMC3 comparing the UT and IAA-treated cells. Heatmaps were plotted across a  $\pm 3$ -kb region around four major types of loop anchors. The colormap shows an increased signal ( $\log_2$ ) in orange and a decreased signal in purple after IAA treatment.
- b. Profiles of differential ChIP-seq signals for CTCF or RAD21 comparing the UT and IAA-treated cells across the same regions as in a.
- c. Scatter plot of loop scores for the called loops in the UT and IAA-treated cells (left). The violin chart (inset) shows the distribution of loop scores for the UT and IAA-treated conditions. The pile-up contact maps are plotted with loops sorted by up-regulated, down-regulated, or unchanged loops (right).
- d. Enrichment of the chromHMM states at loop anchors sorted by up-regulated, down-regulated, or unchanged after IAA treatment.
- e. Profiles of ChIP-seq signals for CTCF (top) or RAD21 (bottom) across a  $\pm 3$ -kb region around the anchors of E-P or P-P loops that are either unchanged (gray) or reduced after CTCF (blue) or RAD21 (pink) depletion.
- f. Histogram of the length distribution for loops that are unchanged (gray) or reduced in CTCF (blue) or RAD21 (orange) depletion.
- g. Length distribution of the unchanged or down-regulated E-P/P-P loops relative to TAD boundaries.
- h. Ratio of the unchanged (gray) or down-regulated (pink) E-P/P-P loop anchors located within  $\pm 10$  kb of TAD boundaries.

#### Extended Data Fig 5.

- a. Schematic for YY1 protein domains (top) and histogram of mScarletI intensity for HaloTag-YY1 cells (clone YN11) treated with IAA for 0, 1, 2, or 3 hours.
- b. Pie chart shows the percentage of YY1 peaks enriched with four primary types of chromHMM states and silent chromatin.
- c. MA plots showing differential ChIP-seq peaks between the UT and IAA-treated cells. The significantly changed peaks (adjusted  $p$ -value  $< 0.05$ ) are colored in red. X-axis: mean observations of the UT and IAA-treated cells. Y-axis:  $\log_2$  fold-change comparing the UT and IAA-treated cells.
- d. Heatmaps of differential ChIP-seq signals for YY1, SMC1A, and SMC3 after YY1 depletion.
- e. Summary of Micro-C experiments for UT and YY1-depleted cells. Total unique reads are annotated for each replicate on the right, consisting of trans-interaction (inter-chromosome) in green, short-range cis-interaction ( $< 20$  kb) in orange, and long-range cis-interaction ( $> 20$  kb) in red.

- f. Genome-wide pileup plots for multiple levels of chromatin structures in the UT and IAA-treated cells. From left to right: 1) Saddle plot for compartmentalization strength. This shows average distance-normalized contact frequencies between 100-kb bins in cis with ascending eigenvector values (log2). Upper left and bottom right corners: contact frequency between B-B and A-A compartments. Upper right and bottom left corners: frequency of inter-compartment interactions. 2) Aggregate domain analysis (ADA) for TADs and TADs with significant YY1 ChIP-seq decrease. TADs were rescaled and aggregated at the center of the plot with matrix balancing normalization. 3) APA for cohesin, E-P, or P-P. The loops were aggregated at the center of a 20-kb window at 400-bp resolution. The genome-wide averaged loop enrichment was calculated by the fold enrichment (center pixel/expected upper-left pixels).
- g. Genome-wide contact decaying  $P(s)$  analysis (top) and the slope distributions of the  $P(s)$  curves for UT and IAA-treated cells.
- h. Overlap of DEGs between 1) mNET-seq after 3-hour YY1 depletion; 2) RNA-seq after 3-hour YY1 depletion; 3) RNA-seq after 3-hour YY1 depletion<sup>7</sup>; and 4) two sets of random genes.
- i. Snapshots of Micro-C maps comparing chromatin interactions in the UT (top) and IAA-treated (bottom) cells surrounding the *Nes* gene.

#### Extended Data Fig 6.

- a. Schematic for conjugating a fluorescent dye with the HaloTag-YY1 fusion protein, which emits fluorescence upon excitation by a specific wavelength.
- b. Schematic for endogenously fusing the N-terminus of YY1 with HaloTag.
- c. Immunoblots of wild-type (WT), HaloTag-YY1 knock-in (YN11 and YN31), and stably expressing HaloTag-YY1/YY1-HaloTag (PBYN2 and PBYC3) mESC lines for YY1, HaloTag, and FLAG proteins. TBP was used as a loading control.
- d. Confocal or Airyscan-resolved live-cell imaging for HaloTag-YY1 stained with 500 nM TMR. Arrow points to sporadic loci within the nucleolus. Images at the bottom panel are a z-projection with the mean signal.
- e. Schematic for the spaSPT experiment and the analysis pipeline with Quot and Spagl<sup>8</sup>.
- f. Heatmaps of localization errors obtained by aggregated likelihood across all trajectories (left) or posterior marginalized localization error (middle) for clone YN11, YN31, PBYN2, PBYC3 and H2B. The distribution of the likelihood of diffusion coefficients (x-axis) for single cells (each row at y-axis) is plotted on the right panel.
- g. spaSPT displacement histograms for YN11, YN31, PBYN2, PBYC3, and H2B. Raw displacement data for seven different lag times are shown with a three-state Spot-On model<sup>9</sup> fit overlaid. The inferred fractions and diffusion coefficients for each cell are shown in the table in the bottom panel.

#### Extended Data Fig 7.

- a. Schematic for biochemical fractionation experiment.
- b. Immunoblots of cytoplasmic (Cyt) and nuclear proteins dissociating from chromatin at increasing salt concentrations (75, 150, 300, 500 mM and 1M NaCl) as schematized in **a**, probed with the indicated antibodies ( $\alpha$ ). Son. Sonicated chromatin; Ins. Insoluble pellet after chromatin sonication; % of total: signal intensity of each fraction divided by the total

signal intensity across all fractions. In the top blot, the arrowhead indicates the band corresponding to WAPL; asterisks denote non-specific bands.

- c. Spatial reconstruction of spaSPT data for YY1's trajectory densities. YY1 trajectories are binned by diffusion coefficients, as indicated.
- d. Segmentation of YY1 clusters with the reconstructed images from the spaSPT data. Individual YY1 clusters are colored.
- e. Snapshots of FRAP experiments for multiple time points from "before bleach" to "50 sec after bleach".
- f. FRAP analysis of YY1 bleached with a small circular spot. Error bars indicate the standard deviation of each acquired data point.

# **Extended Data Fig 8.**

- a. Histogram of mNeonGreen intensity for cells (clones CD1 and RD35) treated with IAA for 0 or 3 hours.
- b. Airyscan-resolved live-cell imaging for HaloTag-YY1 stained with 500 nM TMR in wild type, CTCF-, or RAD21-depleted cells.
- c. Heatmaps of localization errors obtained by aggregated likelihood across all trajectories (panel one from left) or posterior marginalized localization error (panel two) for the UT or IAA-treated cells. The distribution of likelihood of diffusion coefficients (x-axis) for single cells (each y-axis row) is plotted on panel three. Estimation of YY1 diffusion coefficients by regular Brownian motion with marginalized localization errors is plotted on panel four.
- d. Heatmaps of differential ChIP-seq signals for CTCF, RAD21, SMC1a, SMC3, and YY1 in cells depleted for CTCF, RAD21, and WAPL. The peaks are centered on wildtype YY1 peaks across a  $\pm 3$ -kb region. The colormap shows an increased signal ( $\log_2$ ) in orange and a decreased signal in purple after IAA treatment.
- e. MA plot showing the differential ChIP-seq peaks between the UT and IAA-treated cells (left). The significantly changed peaks (adjusted  $p$ -value  $< 0.05$ ) are colored in red. X-axis: mean observations of UT and IAA cells. Y-axis:  $\log_2$  fold-change comparing the UT and IAA-treated cells. Pie chart shows the percentages of the down-regulated YY1 peaks enriched at promoters and enhancers (right).

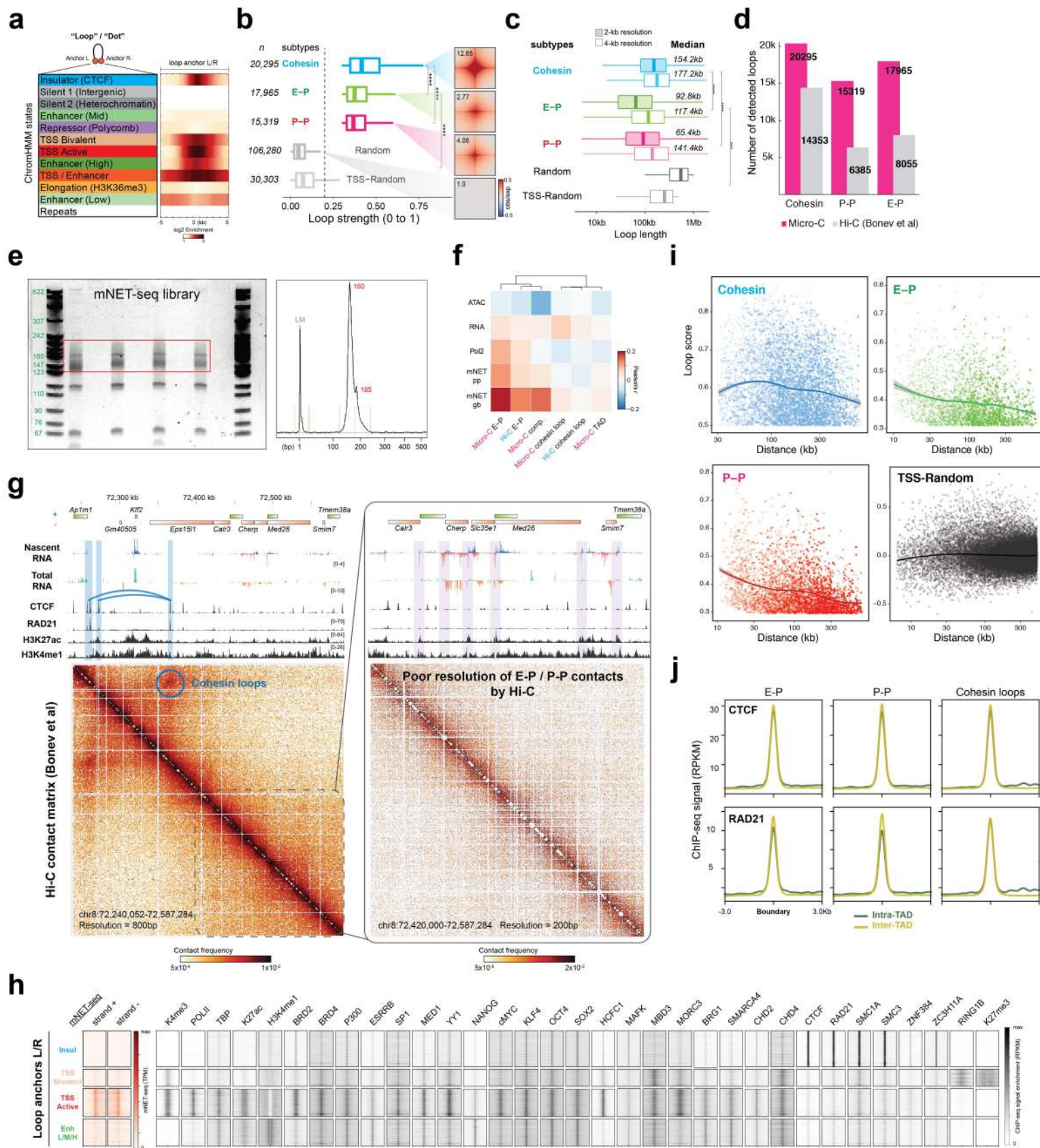
# **References for Extended Figure Legend**

1. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* **9**, 215–6 (2012).
2. Matthey-Doret, C. *et al.* Computer vision for pattern detection in chromosome contact maps. *Nat Commun* **11**, 5795 (2020).
3. Ardakany, A. R., Gezer, H. T., Lonardi, S. & Ay, F. Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biol* **21**, 256 (2020).

- 226 4. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*  
227 **171**, 557-572.e24 (2017).
- 228 5. Owens, N. *et al.* CTCF confers local nucleosome resiliency after DNA replication and during  
229 mitosis. *Elife* **8**, e47898 (2019).
- 230 6. Sönmezer, C. *et al.* Molecular Co-occupancy Identifies Transcription Factor Binding  
231 Cooperativity In Vivo. *Mol Cell* (2020) doi:10.1016/j.molcel.2020.11.015.
- 232 7. Weintraub, A. S. *et al.* YY1 Is a Structural Regulator of Enhancer-Promoter Loops. *Cell* **171**,  
233 1573-1588.e28 (2017).
- 234 8. Heckert, A., Dahal, L., Tjian, R. & Darzacq, X. Recovering mixtures of fast diffusing states  
235 from short single particle trajectories. *Biorxiv* 2021.05.03.442482 (2021)  
236 doi:10.1101/2021.05.03.442482.
- 237 9. Hansen, A. S. *et al.* Robust model-based analysis of single-particle tracking experiments with  
238 Spot-On. *Elife* **7**, e33125 (2018).

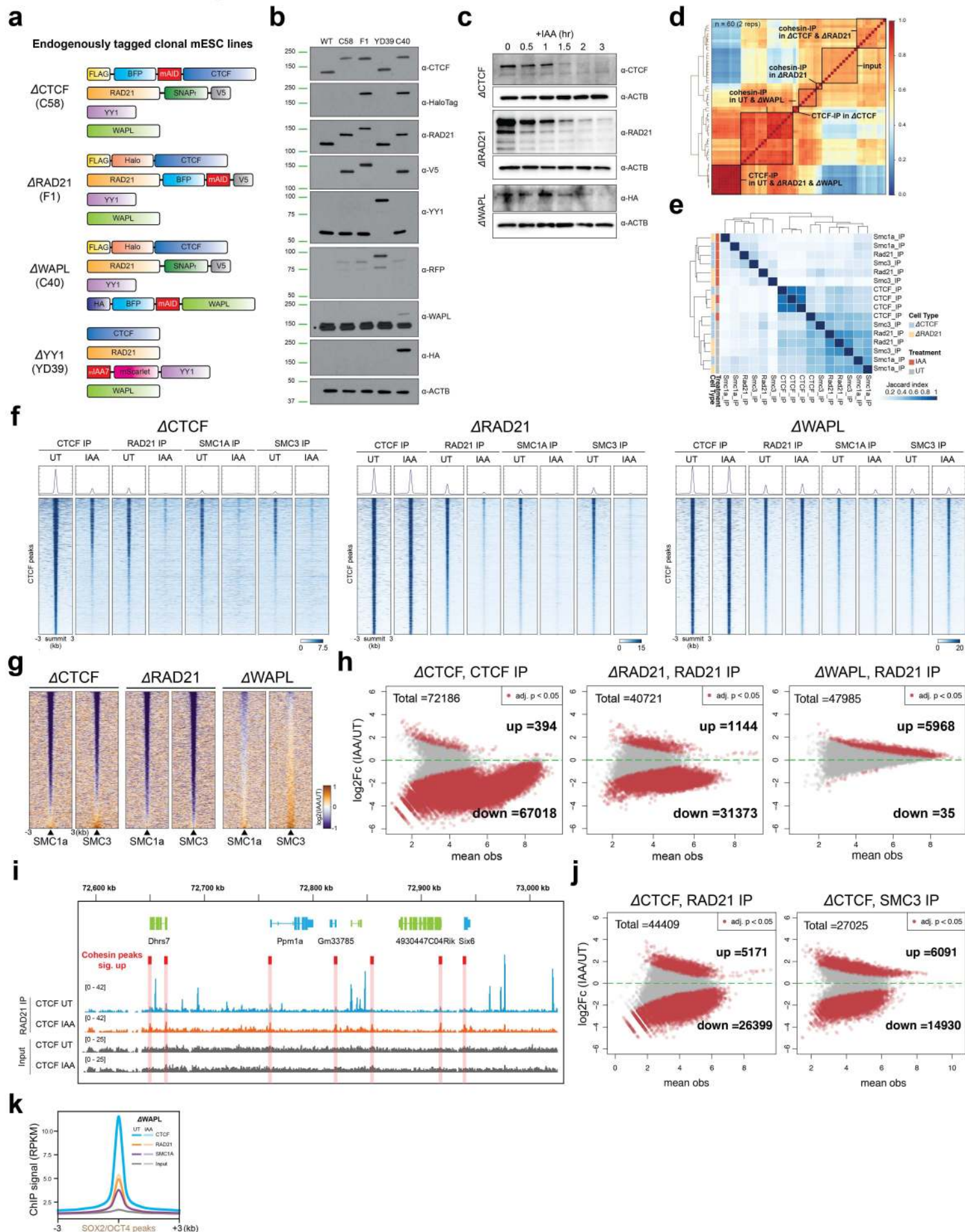


# Extended Data Figure 1

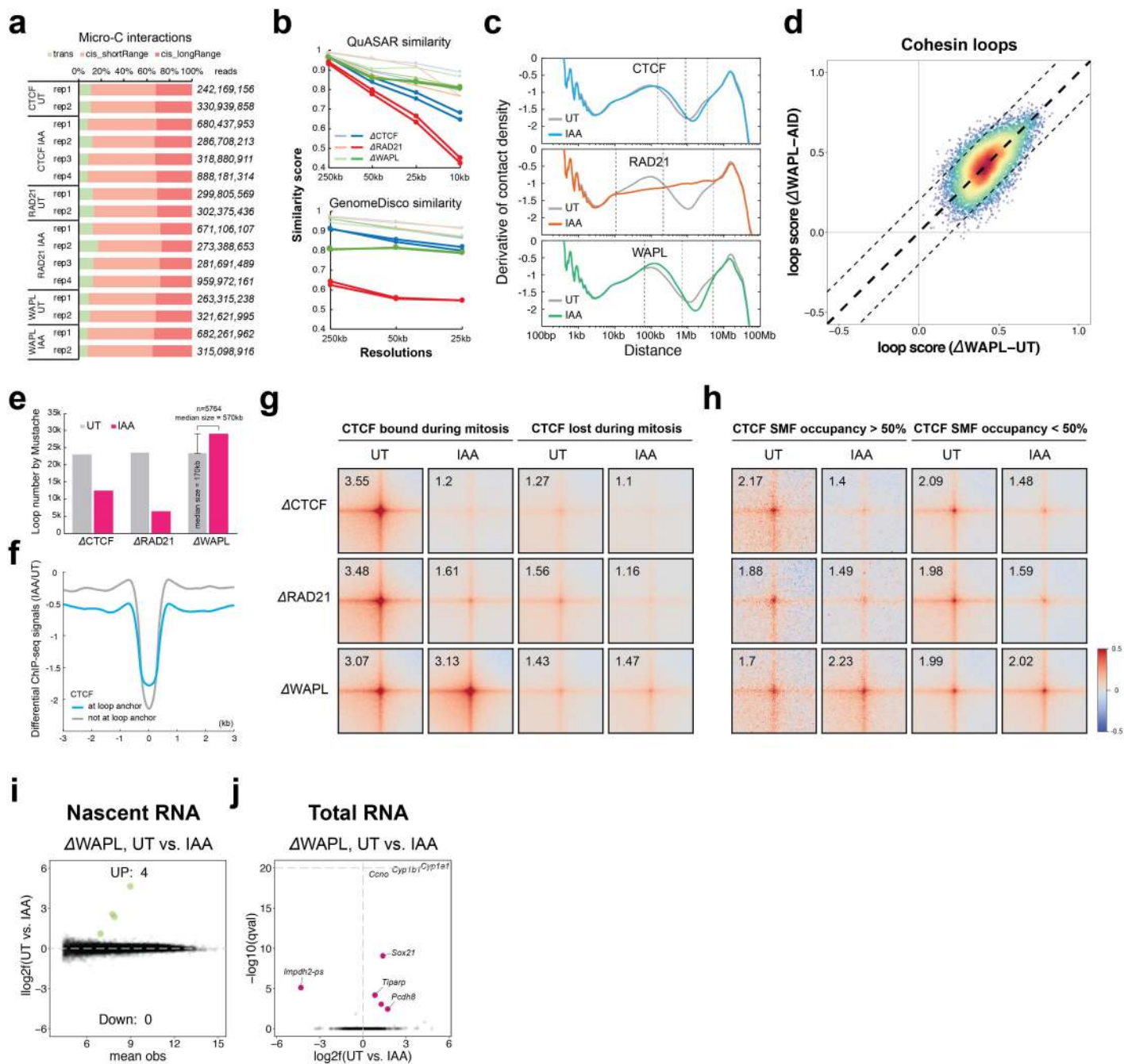




## Extended Data Figure 2

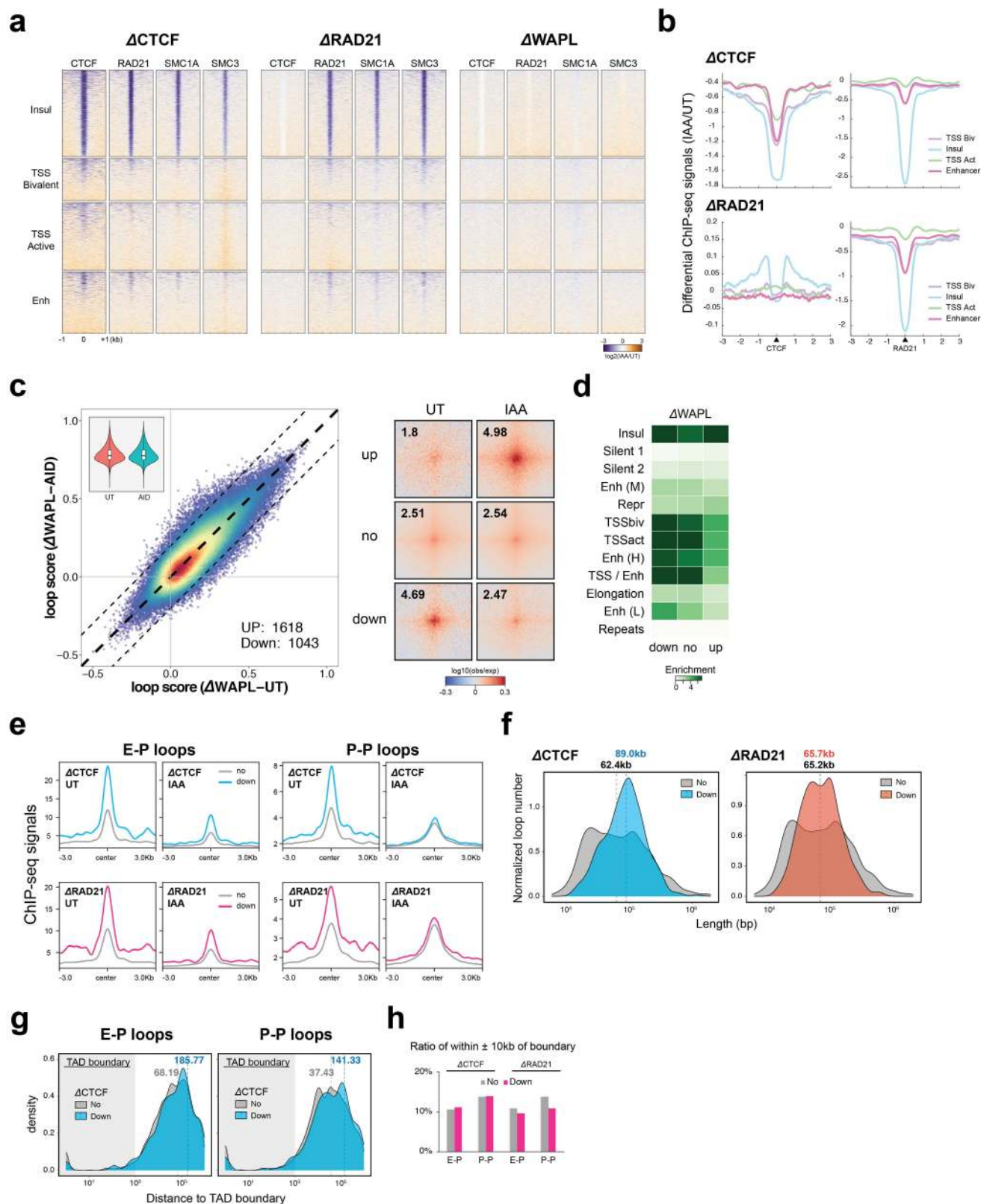


# Extended Data Figure 3

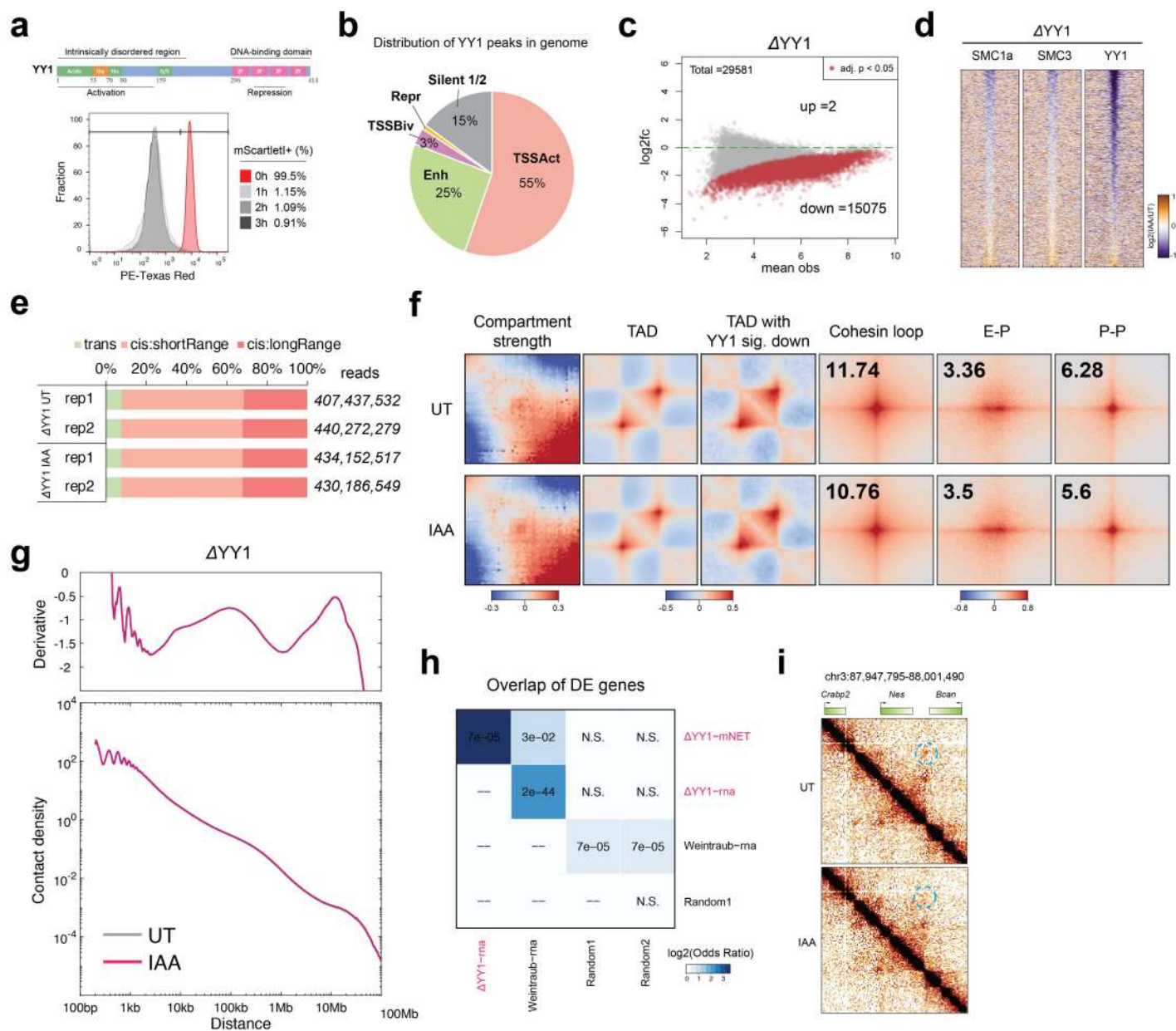




# Extended Data Figure 4

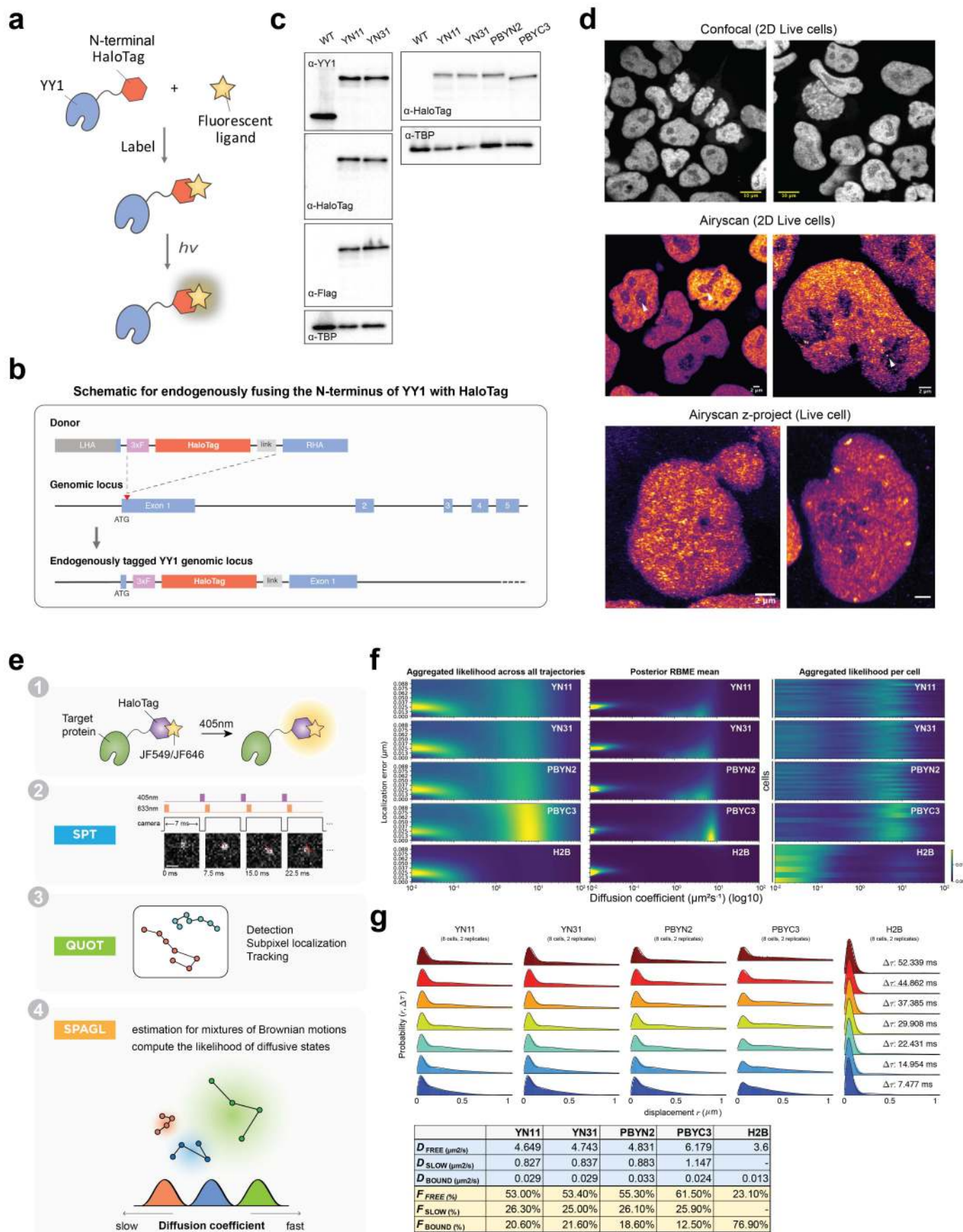


# Extended Data Figure 5



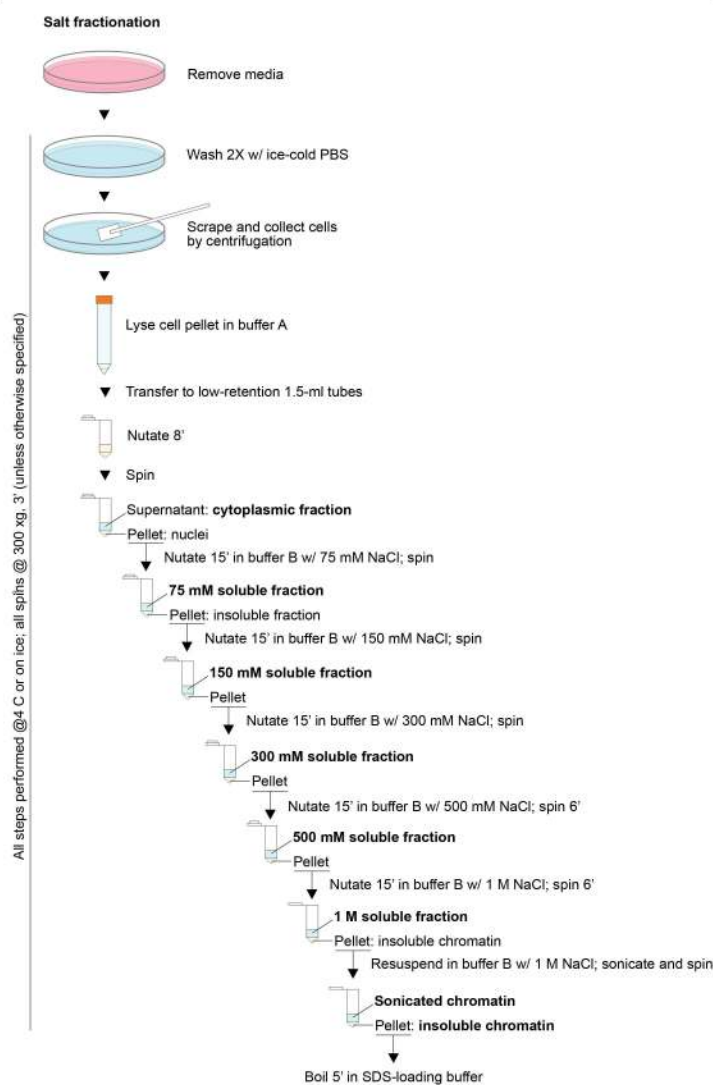


## Extended Data Figure 6

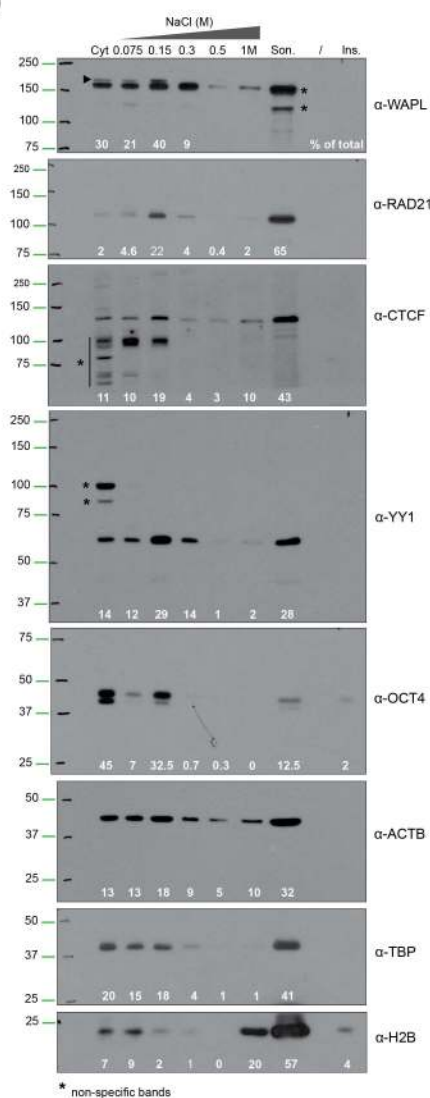


## Extended Data Figure 7

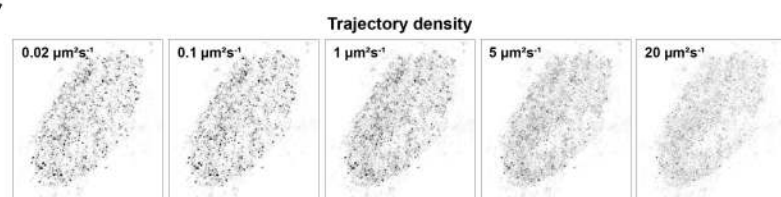
**a**



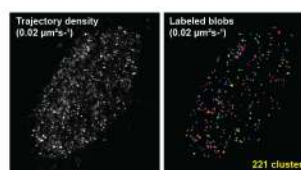
**b**



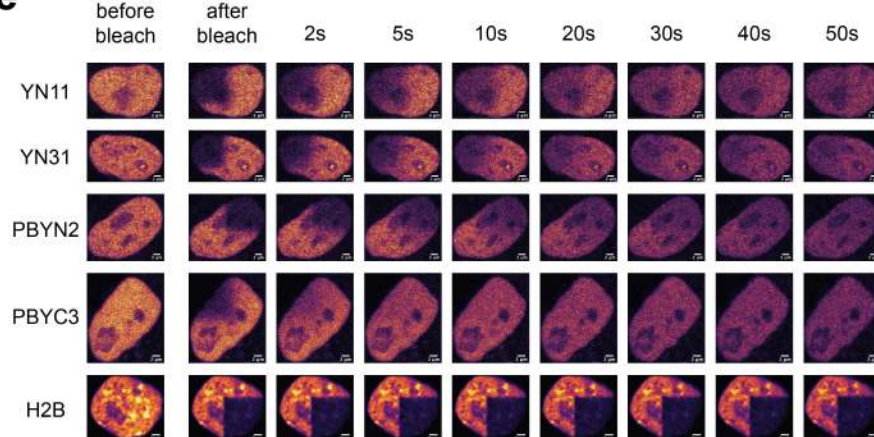
**c**



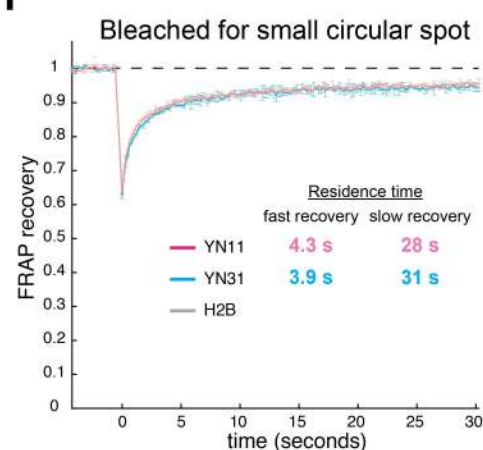
**d**



**e**

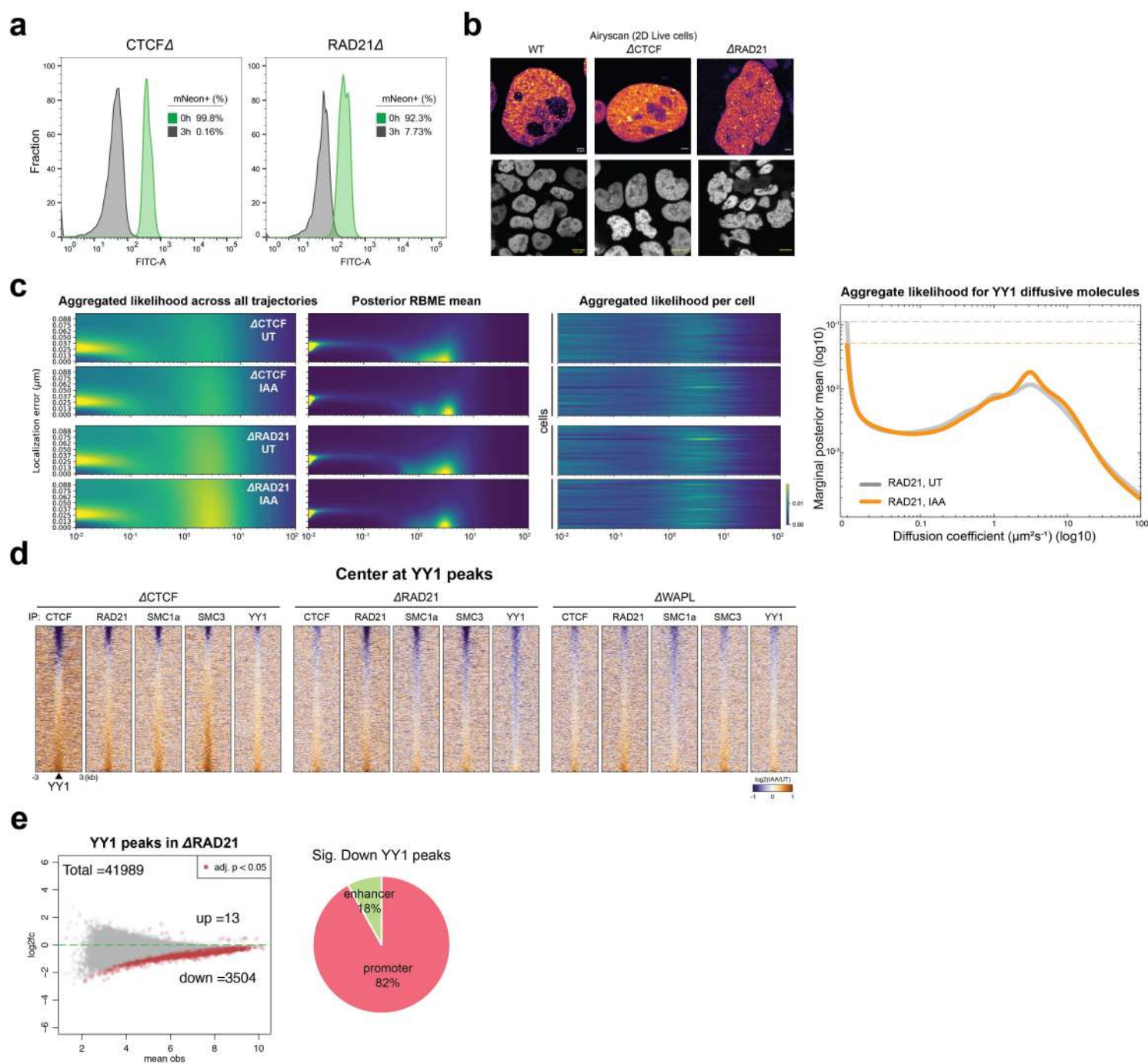


**f**





## Extended Data Figure 8



# Methods

## Cell culture, stable cell line construction and dye labeling

JM8.N4 mouse embryonic stem cells (mESCs)<sup>1</sup> (Research Resource Identifier: RRID:CVCL\_J962; obtained from the KOMP Repository at UC Davis) were used for all experiments. Cells were cultured on plates pre-coated with 0.1% gelatin (Sigma-Aldrich G9291) in knock-out DMEM (ThermoFisher #10829018) supplemented with 15% FBS (HyClone FBS SH30910.03 lot #AXJ47554), 0.1 mM MEM non-essential amino acids (ThermoFisher #11140050), 2 mM GlutaMAX (ThermoFisher #35050061), 0.1 mM 2-mercaptoethanol (Sigma-Aldrich M3148), 1% Penicillin-Streptomycin (ThermoFisher #15140122), and 1000 units of LIF (Millipore). Media was replaced daily and cells were passaged every 2 days by trypsinization. Cells were grown at 37°C and 5.5% CO<sub>2</sub> in a Sanyo copper alloy IncuSafe humidified incubator (MCO-18AIC(UV)). For imaging, the media was identical except that knock-out DMEM lacking phenol red (ThermoFisher #31053028) was used to minimize background fluorescence.

Cell lines stably expressing 3xFLAG-HaloTag-YY1 and YY1-HaloTag-3xFLAG were generated using PiggyBac transposition and drug selection. The coding sequences of YY1 and 3xFLAG-HaloTag or HaloTag-3xFLAG were cloned by Gibson Assembly into a PiggyBac vector co-expressing a puromycin resistance gene. For each construct, YY1 and HaloTag were connected by a short peptide linker sequence (GDGAGLING). 1.3 µg of PiggyBac vector was co-transfected with 0.5 µg of SuperPiggyBac transposase vector into JM8.N4 cells by nucleofection using the Lonza Mouse Embryonic Stem Cell Nucleofector Kit reagents and the Amaxa Nucleofector II device. After 24 hr, antibiotic selection was begun with 500 ng/mL of puromycin. Once cells reached ~80% confluence, 5000 cells were seeded onto a 15 cm plate and allowed to grow into colonies. After 4-6 days, single clones were isolated and seeded onto a 96-well plate. Cells were allowed to expand for 4-5 days, and successfully integrated PiggyBac constructs were validated by PCR genotyping. We selected clone PBYN2 for 3xFLAG-HaloTag-YY1 and clone PBYC3 for YY1-HaloTag-3xFLAG.

For all single-molecule experiments, cells were grown overnight on Matrigel-coated 25-mm circular no 1.5H cover glasses (High-Precision 0117650). Prior to all experiments, the cover glasses were plasma-cleaned and then stored in isopropanol until use. Cells were labeled with 50 nM PA-JFX646 (fast SPT) or 25 pM JFX549 (slow SPT) HaloTag ligand for 30 min and washed twice with PBS. After the final washes, cells were replenished with phenol-free medium prior to imaging.

For PALM experiments, cells were grown overnight on Matrigel-coated 25-mm circular no 1.5H cover glasses (High-Precision 0117650). Prior to all experiments, the cover glasses were plasma-cleaned and then stored in isopropanol until use. Cells were labeled with 500 nM PA-JFX549 HaloTag ligand for 30 min, washed twice with fresh media for 5 min, and then washed once with pH 7.4 phosphate-buffered saline (PBS). Labeled cells were fixed with 4% paraformaldehyde and 2% glutaraldehyde in PBS for 20 min at 37°C, washed once with PBS, and imaged in PBS with 0.01% (w/v) NaN<sub>3</sub>.

For FRAP experiments, cells were grown overnight on Matrigel-coated (Corning #354277; purchased from ThermoFisher #08-774-552) glass-bottom 35-mm dishes (MatTek P35G-1.5-14C). Cells were labeled with 500 nM TMR for 30 min and washed twice with PBS.

## Generation of CRISPR/Cas9-mediated knock-in cell lines

Endogenously tagged mESC lines were generated by CRISPR/Cas9-mediated genome editing as previously described<sup>2</sup> with modifications. mESCs were seeded to 6-well plates and each well was transfected with



Lipofectamine 3000 (ThermoFisher #L3000015) or nucleofection (Lonza #VPH-1001) using 1 µg of a construct encoding Cas9, Venus yellow fluorescent protein, and a sgRNA and 2 µg of a homology repair plasmid with the DNA insert of interest flanked by left and right homology arms (~500 bp each). For each insertion we designed 2-4 individual sgRNAs with the CRISPOR tool<sup>3</sup> to be transfected separately into mESCs and pooled before sorting Venus-positive cells by FACS. Sorted cells were plated at a density of ~5,000-10,000 cells per P15 plate and media replaced every other day. Single colonies were transferred to 96-well plates to be expanded and genotyped by PCR after crude DNA extraction with the DirectPCR Lysis Reagent (Viagen Biotech #302-C). Clones with a correct insert were further verified by Sanger sequencing and Western blot.

The parental cell line of CTCF, Rad21 and WAPL auxin-inducible degron (AID) mESCs was clone C59 from Hansen et al.<sup>4</sup> further modified to visualize DNA looping (Clone C36 osTir1: full details on the generation of C36 will be published elsewhere (in preparation) and are available upon request).

To endogenously tag 1) CTCF, RAD21, or WAPL with mAID<sup>5</sup> in the C36 osTir mESC line; 2) YY1 with miniIAA7<sup>6</sup> in the parental JM8.N4 with osTIR1 expression; 3) YY1 with HaloTag in the parental JM8.N4; and 4) CTCF and RAD21 with mAID in the YY1-HaloTag cell lines, we designed homology repair plasmids encoding 1) 3xFLAG-BFP-mAID-CTCF, RAD21-BFP-mAID-V5, HA-BFP-mAID-WAPL; 2) miniIAA7-mScarletI-YY1; 3) 3xFLAG-HaloTag-YY1; 4) 3xFLAG-mNeonGreen-mAID-CTCF and RAD21-mNeonGreen-mAID-V5 (plasmids, sgRNAs and primer sequences in **Supplementary Table 1**). Homozygous clones with the correct genotype were further verified by Sanger sequencing and Western blot after IAA-mediated depletion. Among the several clones generated for each AID line, we picked clone C58 for 3xFLAG-BFP-mAID-CTCF; clone F1 for RAD21-BFP-mAID-V5; clone C40 for HA-BFP-mAID-WAPL; clone YD39 for miniIAA7-mScarletI-YY1; clone YN11 for 3xFLAG-HaloTag-YY1; clone CD1 for 3xFLAG-mNeonGreen-mAID-CTCF and clone RD35 for RAD21-mNeonGreen-mAID-V5.

## Western Blotting

For Western blot analysis, cells were seeded to 6-well plates, washed with ice-cold PBS and scraped in 300 µl of low-salt lysis buffer (0.1 M NaCl, 25 mM HEPES, 1 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 0.5% NP-40 and protease inhibitors), with 125 U/mL of benzonase (Novagen, EMD Millipore), rocked at 4 °C for 1 hr and then NaCl was added to a 0.2-M final concentration. Lysates were then rocked at 4 °C for 30 min and centrifuged at maximum speed at 4 °C. Supernatants were quantified by Bradford. Alternatively, cells were lysed directly on plates with 300 µl of high-salt lysis buffer (0.5 M NaCl, 25 mM HEPES, 1 mM MgCl<sub>2</sub>, 0.2 mM EDTA, 0.5% NP-40 and protease inhibitors) and transferred to low-stick tubes with 100 µl of 4x SDS-loading buffer. Proteins were loaded onto a 9% Bis-Tris SDS-PAGE gel, transferred onto a nitrocellulose membrane (Amersham Protran 0.45 µm NC, GE Healthcare) for 1 hr at 100V, blocked in TBS-Tween with 10% milk for at least 1 hr at room temperature and blotted overnight at 4 °C with primary antibodies in TBS-T with 5% milk (antibodies listed in **Supplementary Table 1**). HRP-conjugated secondary antibodies were diluted 1:5000 in TBS-T with 5% milk and incubated with the membrane at room temperature for an hour before performing the chemiluminescence reaction (Western Lightning Plus-ECL, Enhanced Chemiluminescence Substrate, Perkin Elmer NEL105001EA). Signal was captured with either X-ray films (CL-XPosure™ Film, ThermoScientific 34091) or with a Bio-Rad ChemiDoc imaging system.

## Chromatin immunoprecipitation (ChIP)

ChIP was performed as described with a few modifications<sup>7</sup>. Cells were treated with either ethanol or 500 mM of ethanol-dissolved auxin for 3 hrs and cross-linked for 7 min at room temperature with 1% formaldehyde-containing medium; cross-linking was stopped by PBS-glycine (0.125 M final). Cells were washed twice with ice-cold PBS, scraped, centrifuged for 10 min and pellets were flash-frozen. Cell pellets were thawed and resuspended in cell lysis buffer (5 mM PIPES, pH 8.0, 85 mM KCl, and 0.5% NP-40, 1 ml/15 cm plate) with protease inhibitors and incubated for 10 min on ice. Lysates were centrifuged for 10 min at 1250 rcf and nuclear pellets were resuspended in 6 volumes of sonication buffer (50 mM Tris-HCl, pH 8.1, 10 mM EDTA, 0.1% SDS) with protease inhibitors, incubated on ice for 10 min, and sonicated to obtain DNA fragments below 2000 bp in length (Covaris S220 sonicator, 20% Duty factor, 200 cycles/burst, 150 peak incident power, 10-20 cycles 50 sec on and 30 sec off). Sonicated lysates were cleared by centrifugation and chromatin (400 µg per antibody) was diluted in RIPA buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate, 140 mM NaCl) with protease inhibitors to a final concentration of 0.8 µg/µl, precleared with Protein G sepharose (GE Healthcare) for 2 hrs at 4°C and immunoprecipitated overnight with 4 µg of specific antibodies (antibodies listed in **Supplementary Table 1**). About 4% of the precleared chromatin was saved as input. Immunoprecipitated DNA was purified with the Qiagen QIAquick PCR Purification Kit, eluted in 36 µl of 0.1X TE (1 mM Tris-HCl pH 8.0, 0.01 mM EDTA) and analyzed by qPCR together with 2% of the input chromatin prior to ChIP-seq library preparation (SYBR® Select Master Mix for CFX, ThermoFisher; ChIP-qPCR primer sequences in **Supplementary Table 1**).

ChIP-seq libraries were prepared using the NEBNext® Ultra™ II DNA Library Prep Kit for Illumina® (NEB E7645) according to manufacturer instructions with a few modifications. 20 ng of ChIP input DNA (as measured by Nanodrop) and 25 µl of the immunoprecipitated DNA were used as a starting material and the recommended reagent volumes were cut in half. The NEBNext Adaptor for Illumina was diluted 1:10 in Tris/NaCl, pH 8.0 (10 mM Tris-HCl pH 8.0, 10 mM NaCl) and the ligation step extended to 30 min. After ligation, a single purification step with 0.9X volumes of Agencourt AMPure XP PCR purification beads (Beckman Coulter A63880) was performed, eluting DNA in 22 µl of 10 mM Tris-HCl pH 8.0. 20 µl of the eluted DNA was used for the library enrichment step, performed with the KAPA HotStart PCR kit (Roche Diagnostics KK2502) in 50 µl of total reaction volume (10 µl 5X KAPA buffer, 1.5 µl 10 mM dNTPs, 0.5 µl 10 µM NEB Universal PCR primer, 0.5 µl 10 µM NEB index primer, 1 µl KAPA polymerase, 16.5 µl nuclease-free water and 20 µl sample). Samples were enriched with 9 PCR cycles (98 °C, 45 sec; [98 °C, 15 sec; 60 °C, 10 sec] x 9; 72 °C, 1 min; 4 °C, hold), purified with 0.9 volumes of AMPure XP PCR purification beads and eluted with 33 µl of 10 mM Tris-HCl pH 8.0. Library concentration, quality and fragment size were assessed by Qubit fluorometric quantification (Qubit™ dsDNA HS Assay Kit, Invitrogen™ Q32851), qPCR and Fragment analyzer™. 12 multiplexed libraries were pooled and sequenced in one lane on the Illumina HiSeq4000 sequencing platform (50-bp, single end-reads) at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

ChIP-seq raw reads from ethanol (UT)- or auxin (IAA)-treated ΔCTCF, ΔRAD21, ΔWAPL and ΔYY1 mESCs (96 libraries total, 2 biological replicates per condition) were quality-checked with FastQC and aligned onto the mouse genome (mm10 assembly) using Bowtie<sup>8</sup>, allowing for two mismatches (-n 2) and no multiple alignments (-m 1). Biological replicates were pooled and peaks were called with MACS2 (--nomodel --extsize 250)<sup>9</sup> using input DNA as a control. To create heatmaps we used deepTools (version 2.4.1)<sup>10</sup>. We first ran bamCoverage (--binSize 50 --normalizeTo1x 2150570000 - --extendReads 300 --ignoreDuplicates -of bigwig) and normalized read numbers to 1x sequencing depth, obtaining read coverage per 50-bp bins across the whole genome (bigWig files). We then

used the bigWig files to compute read numbers across 6 kb centered on either CTCF, RAD21, or YY1 peak summits as called by MACS2 in UT cells (`computeMatrix reference-point --referencePoint=TSS --upstream 3000 --downstream 3000 --missingDataAsZero --sortRegions=no`). We sorted the output matrices by decreasing UT enrichment, calculated as the total number of reads within a MACS2 called ChIP-seq peak. Finally, heatmaps were created with the `plotHeatmap` tool (`--averageTypeSummaryPlot=mean --colorMap='Blues' --sortRegions=no`). To identify the differential peaks between UT- and IAA-treated cells, we used `MANorm2`<sup>11</sup>, which employs a hierarchical strategy for normalization of ChIP-seq data and assesses within-group variability of ChIP-seq signals based on an empirical Bayes framework. We note that the total ChIP-seq peak numbers in `MANorm2` are combined from UT- and IAA-treated cells and may differ from the number of MACS2 calling.

### Biochemical fractionation

Wild type JM8.N4 mESCs were seeded to 15-cm plates, washed with ice-cold PBS, scraped in PBS and pelleted at 135 rcf, 10 min at 4 °C. Pellets were resuspended in 350 µl of cell lysis buffer A (10 mM HEPES pH 7.9, 10 mM KCl, 3 mM MgCl<sub>2</sub>, 340 mM Sucrose, 10% glycerol v/v, 1 mM DTT and freshly added 0.1% Triton X-100 v/v and protease inhibitors) and rocked for 8 min at 4 °C. Nuclei were pelleted at 3,000 xg for 3 min at 4 °C and the supernatant containing the cytoplasmic fraction was saved. Nuclei were resuspended in 350 µl of buffer B with 75 mM NaCl (9 mM EDTA, 0.2 mM EGTA, 1 mM DTT and freshly added 0.1% Triton X-100 v/v and protease inhibitors) and rocked at 4 °C for 15 min. Nuclei were pelleted again as above (supernatant saved as the 75 mM wash fraction) and washed with 350 µl of buffer B with increasing NaCl concentrations (150 mM, 300 mM, 500 mM and 1 M, see **Extended Data Figure 7** for a step-by-step procedure). After collecting the 1 M wash, the pellet was resuspended to 350 µl of 1 M buffer B and sonicated (Covaris S220 sonicator, 20% Duty factor, 200 cycles/burst, 100 peak incident power, 8 cycles 20 sec on and 40 sec off). The sonicated lysate was spun down and the insoluble pellet boiled in SDS-loading buffer. 10 µl of each fraction was added to 2 µl of 4X SDS loading buffer and subjected to Western blot as detailed above. Band intensities were quantified with the ImageJ “Analyze Gels” function<sup>12</sup>.

### Micro-C assay for mammalian cells

We briefly summarize the Micro-C experiment here. The detailed protocol and technical discussion are available in our previous study<sup>13</sup>. Mouse embryonic stem cells (JM8.N4) and the derivative genome-edited lines were cultured in the recommended conditions<sup>1</sup>. When cells grew to ~70% confluency, we resuspended them in 0.05% of trypsin, inactivated with cell culture media, and resuspended in 1% formaldehyde crosslinking media (without FBS). Cells were crosslinked for 10 min at room temperature (RT) while nutating. We then added 1 M Tris-HCl pH 7.5 (final concentration 375 mM) and incubated for 5 min at RT to quench the crosslinking reaction. Cells were spun down and washed twice with cold PBS. Cell pellets were crosslinked again with the freshly prepared 3 mM DSG crosslinking solution (in base media without FBS) for 45 min at RT while nutating. The crosslinking reaction was quenched and washed following the same steps as above. We routinely split cells into 1 million cells per vial after fixation and perform MNase titration and Micro-C with 1 million cells. Crosslinked cell pellets can be snap frozen in liquid nitrogen and stored at -80 °C for months or used immediately for the next step. We note that 1) cells directly crosslinked on the dish typically yield a similar result; 2) using freshly made formaldehyde and DSG solution is critical

to obtain high-quality Micro-C data; and 3) to avoid loss of cells, low-retention tubes and tips are strongly recommended.

Crosslinked cell pellets were resuspended in MB1 (50 mM NaCl, 10 mM Tris-HCl pH 7.5, 5 mM MgCl<sub>2</sub>, 1 mM CaCl<sub>2</sub>, 0.2% NP-40) at a concentration of 1x10<sup>6</sup> cells/100 µL and were incubated for 20 min on ice. Cells were spun down and washed once with MB1. We then added the appropriate amount of Micrococcal nuclease (MNase) and incubated the tube for 10 min at 37 °C while shaking in a thermomixer at ~850 rpm. The optimal digestion condition results in ~90% of mono-nucleosome and ~10% of di-nucleosome. The MNase reaction was inactivated by adding 4 mM EGTA and incubated for 10 min at 65 °C. Digested chromatin was spun down and washed twice with 1 mL of cold MB2 (50 mM NaCl, 10 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>). We note that and MNase titration that yields 90% monomer/10% dimers substantially reduces contamination with un-digested (un-ligated) dimers in Micro-C data.

MNase-fragmented chromatin was then subjected to the three-step end-repair protocol to generate ligatable ends filled with biotin-dNTPs. First, the pellet was resuspended in the end-repair buffer (50 mM NaCl, 10 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>, 100 µg/mL BSA, 2 mM ATP, 5 mM DTT) and the 5'-ends were phosphorylated with 25 units of T4 Polynucleotide Kinase (NEB #M0201) for 15 min at 37 °C while shaking in a thermomixer at 1000 rpm for an interval of 15 sec every 3 min. Second, to convert the mixed types of nucleosomal ends to cohesive ends, we added 25 units of DNA Polymerase I, Large (Klenow) Fragment (NEB #M0210) directly to the reaction and incubated the tube for an additional 15 min at 37 °C while shaking in a thermomixer at 1000 rpm for an interval of 15 sec every 3 min. Third, to repair the nucleosomal DNA ends to the blunt and ligatable ends, we supplemented 66 µM of dNTPs (dTTP, dGTP (NEB #N0446), biotin-dATP (Jena Bioscience #NU-835-BIO14), biotin-dCTP (Jena Bioscience #NU-809-BIOX), and 1X T4 DNA ligase reaction buffer (NEB #B0202) directly into the reaction and incubated for 45 min at 25 °C while shaking in a thermomixer at 1000 rpm for an interval of 15 sec every 3 min. The end-repair reaction was then inactivated with 30 mM EDTA for 20 min at 65°C without shaking. Next, chromatin was pelleted by centrifugation for 5 min at ~10,000xg at 4 °C and washed once with cold MB3 (50 mM Tris-HCl pH 7.5, 10 mM MgCl<sub>2</sub>).

End-repaired nucleosomes were then subjected to proximity ligation with ~5000 cohesive end units (CEU) of T4 DNA ligase (NEB #M0202) in 1X T4 DNA ligase reaction buffer (NEB #B0202) for at least 2 hrs at room temperature with slow rotation on an orbital shaker. Biotin-dNTPs at the unligated DNA termini were removed by ~500 units of Exonuclease III (NEB #M0206) in 1X NEBuffer 1 (NEB #B7001) for 15 min at 37°C while shaking at 1000 rpm for an interval of 15 sec every 3 min. Samples were then reverse crosslinked with 1X proteinase K solution (500 µg/µL Proteinase K (ThermoFisher #AM2542), 1% SDS, 0.1 µg/µL RNaseA) at 65°C overnight. DNA was extracted by the standard phenol:chloroform:isoamyl alcohol (25:24:1) and ethanol precipitation procedure. DNA was then purified again with the ZymoClean DNA Clean & Concentrator-5 Kit (Zymo #D4013). Purified DNA was separated on a 3% NuSieve GTG agarose gel (Lonza #50081). The gel band corresponding to the size of dinucleosomal DNA (~250 to 350 bp) was cut and purified with the Zymoclean Gel DNA Recovery Kit (Zymo #D4008). We note that size selection for DNA larger than 200 bp greatly reduces the ratio of unligated monomers in Micro-C data.

Micro-C sequencing libraries were generated by using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB #E7645) with some minor modifications. We first repaired the purified DNA again using the End-it DNA End-Repair Kit (Lucigen #ER0720) following the manufacturer's suggested conditions. The mix was incubated for 45 min at 25 °C and then inactivated the enzyme reaction for 10 min at 70 °C. This step is optional, but we find it increases the library yield. Biotinylated DNA was captured by Dynabeads MyOne Streptavidin C1 beads (ThermoFisher #65001) in 1X BW buffer (5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl) on a nutator for 20 min at room temperature. Beads were washed twice



with 1X TBW buffer (0.1% Tween20, 5 mM Tris-HCl pH 7.5, 0.5 mM EDTA, 1 M NaCl) for 5 min at 55 °C while shaking at 1200 rpm, rinsed once with Tris buffer (10 mM Tris-HCl pH 7.5), and then resuspended in Tris buffer.

We then performed ‘on-bead’ end-repair/A tailing and adapter ligation following the NEB protocol. After adapter ligation, beads were washed once with 1X TBW and rinsed once with Tris buffer. The Micro-C library was generated by using the KAPA HiFi HotStart ReadyMix (Roche #KK2601) or the Q5 High-Fidelity 2X Master Mix (NEB #E7645) with the manufacturer’s suggested conditions. We recommend using a minimal PCR cycle to reduce PCR duplicates, typically 8 to 12 cycles, which can generate high-quality Micro-C data. Prior to sequencing, purifying the library twice with 0.85X AMPure XP beads (Beckman #A63880) can eliminate primer dimers and adapters. We used Illumina 100 bp paired-end sequencing (PE100) to obtain ~400M reads for each replicate in this study.

## Micro-C data processing and analyses

Valid Micro-C contact read pairs were obtained from the HiC-Pro analysis pipeline<sup>14</sup>, and the detailed description and code can be found at <https://github.com/nservant/HiC-Pro>. In brief, paired fastq files were mapped to the mouse mm10 genome independently using Bowtie 2.3.0 with ‘very-sensitive-local’ mode<sup>8</sup>. Aligned reads were paired by the read names. Pairs with multiple hits, low MAPQ, singleton, dangling end, self-circle, and PCR duplicates were discarded. Paired reads with distances shorter than 100 bp (e.g., unligated mono-nucleosome) were also removed. Output files containing all valid pairs were used in downstream analyses. We recommend running a pilot sequencing run (~10M reads) and checking the following quality control statistics before moving forward to a high-coverage sequencing: (1) bowtie mapping rate; (2) reads pairing percentage; (3) ratio of sequencing artifacts; (4) ratio of cis/trans contacts; (5) unligated monomer percentage. If any of the above statistics is not optimal, one might consider checking mapping and filtering parameters or further optimizing the Micro-C experiment. The summary of Micro-C interactions in this manuscript is available in **Supplemental Table 2**.

Valid Micro-C contacts were assigned to the corresponding ‘pseudo’ nucleosome bin. The bin file was pre-generated from the mouse mm10 genome by a 100-bp window that virtually resembles the nucleosome resolution. The binned matrix can be stored in HDF5 format as a COOL file by using the COOLER package (<https://github.com/mirnylab/cooler>)<sup>15</sup> or in HIC file format by using the JUICER package (<https://github.com/aidenlab/juicer>)<sup>16</sup>. Contact matrices were then normalized by using iterative correction (IC) in COOL files<sup>17</sup> or Knight-Ruiz (KR) in HIC files<sup>18</sup>. Regions with low mappability and high noise were blocked before matrix normalization. We expect that matrix balancing normalization corrects systematic biases such as nucleosome occupancy, sequence uniqueness, GC content, or crosslinking effects<sup>17</sup>. We notice that both normalization methods produce qualitatively equal contact maps. To visualize the contact matrices, we generated a compilation of COOL files with multiple resolutions (100-bp to 12,800-bp bins) that can be browsed on the HiGlass 3D genome server (<http://higlass.io>)<sup>19</sup>. In this study, all snapshots of Micro-C or Hi-C contact maps and the 1D browser tracks (e.g., ChIP-seq) were generated by the HiGlass browser unless otherwise mentioned.

We evaluated the reproducibility and data quality for the Micro-C replicates using two published methods independently ([https://github.com/kundajelab/3DChromatin\\_ReplicateQC](https://github.com/kundajelab/3DChromatin_ReplicateQC))<sup>20</sup>. In brief, QuASAR-QC calculates the correlation of values in two distance-based transformed matrices. GenomeDISCO measures the difference in two graph diffusion-smoothed contact maps. We computed the matrix similarity scores between the biological replicates or between the untreated and IAA-treated cells for the 10-kb, 25-kb, 50-



kb, and 250-kb Micro-C data. The detailed descriptions can be found in Sauria et al. for QuASAR-QC<sup>21</sup> and Ursu et al. for GenomeDISCO<sup>22</sup>.

To analyze the genome-wide contact decaying  $P(s)$  curve, we used intra-chromosomal contact pairs to calculate the contact probability in bins with exponentially increasing widths from 100 bp to 100 Mb. Contacts shorter than 100 bp were removed from the analysis in order to minimize noise introduced by self-ligation or undigested DNA products. The orientations of ligated DNA are parsed into 'IN-IN (+/-)', 'IN-OUT' (+/+), 'OUT-IN' (-/-), and 'OUT-OUT' (-/+)' according to the readouts of Illumina sequencing<sup>23,24</sup>. 'UNI' pairs combine 'IN-OUT' and 'OUT-IN' because both orientations are theoretically interchangeable. In this study, we plotted the contact decaying curves with the 'UNI' pairs and then normalized to the total number of valid contact pairs. Slopes of contact decay curves were obtained by measuring slopes in a fixed-width window searching across the entire range of decaying curves. We then plotted the derivative slope in each window against the corresponding genomic distance.

To identify chromosome compartments, we first transformed the Observed/Expected Micro-C matrices at the 200-kb resolution to the Pearson's correlation matrices, and then obtained the eigenvector of the first principal component of the Pearson's matrix by Principal Component Analysis (PCA). The sign of the eigenvector was corrected using active histone marks (H3K27ac and H3K4me3), as positive values are the A compartment (gene-rich or active chromatin) and negative values are the B compartment (gene-poor or inactive chromatin). The detailed description can be found in Lieberman-Aiden et al.<sup>25</sup> The genome-wide compartment strength analysis shown as a saddle plot represents the rearrangement and aggregation of the genome-wide distance-normalized contact matrix with the order of increasing eigenvector values. The chromosome arm is first divided into quantiles based on the compartment score. All combinations of quantile bins are averaged and rearranged in the saddle plot. The Cooltools package (<https://github.com/mirnylab/cooltools>) has implemented the 'call-compartments' and 'compute-saddle' functions with the COOL files.

To identify chromatin domains (TADs) along the diagonal, we used insulation score analysis from the Cooltools package (<https://github.com/mirnylab/cooltools>) or arrowhead transformation analysis from the JUICER package (<https://github.com/aidenlab/juicer>)<sup>16</sup>. The detailed methods were described in Crane et al. for the insulation score analysis<sup>26</sup> or Rao et al. for the arrowhead transformation analysis<sup>27</sup>. Briefly, we analyzed the insulation profile by using a 1-Mb sliding window that scans across Micro-C contact matrices at 20-kb resolution and assigns an insulation intensity to its corresponding bin. The insulation scores were obtained and normalized as the log2 ratio of the individual score to the mean of the genome-wide averaged insulation score. Chromatin boundaries can be identified by finding the local minima along with the normalized insulation score. Boundaries overlapping with low mappability regions were removed from the downstream analysis. The arrowhead analysis defines  $A_{i,i+d} = (M^*_{i,i-d} - M^*_{i,i+d}) / (M^*_{i,i-d} + M^*_{i,i+d})$ , where  $M^*$  is the normalized contact matrix.  $A_{i,i+d}$  can be thought of as the measurement of the directionality preference of locus  $i$ , restricted to contacts at a linear distance of  $d$ .  $A_{i,i+d}$  will be strongly positive/negative if either  $i, i-d$  or  $i, i+d$  is inside the domain and the other is not, but  $A_{i,i+d}$  will be close to zero if both loci are inside or outside the domain. By assigning this query across the genome, the edges of a domain will be sharpened and TADs can be detected. For aggregate domain analysis (ADA), each domain was rescaled to a pseudo-size by  $N_{i,j} = ((C_i - D_{start}) / (D_{end} - D_{start}), (C_j - D_{start}) / (D_{end} - D_{start}))$ , where  $C_{i,j}$  is a pair of contact loci within domain  $D$  that is flanked by  $D_{start}$  and  $D_{end}$ , and  $N_{i,j}$  is a pair of the rescaled coordinates. The rescaled domains can be aggregated at the center of the plot with ICE or distance normalization. Coolpup (<https://github.com/open2c/coolpuppy>)<sup>28</sup> has implemented a handy function to perform ADA with the COOL file. The lists of TAD called by the insulation score analysis or Arrowhead are available in **Supplemental Table 3 or 4**.

To identify loops/dots, we tested two novel algorithms, Mustache (<https://github.com/ay-lab/mustache>)<sup>29</sup> and Chromosight (<https://github.com/koszulab/chromosight>)<sup>30</sup>, for the high-resolution Micro-C data. We found that both approaches outperform the HICCUPS algorithm in the JUICER package<sup>16</sup> and the ‘Call-dots’ function in the Cooltools package in sensitivity and specificity to discover focal contact enrichment. In Mustache analysis, we called loops with balanced contact matrices at resolutions of 400 bp, 600 bp, 800 bp, 1 kb, 2 kb, 4 kb, 10 kb, and 20 kb using the calling options `--pThreshold 0.1 --sparsityThreshold 0.88 --octaves 2`. We then combined all loops at different resolutions. If an interaction was detected as a loop at different resolutions, we retained the precise coordinates in finer resolutions and discarded the coarser resolution. In Chromosight analysis, we used the ‘detect’ function to call loops with balanced contact matrices at resolutions of 400 bp, 600 bp, 800 bp, 1 kb, 2 kb, 4 kb, 10 kb, and 20 kb using calling options listed in **Supplemental Table 5**. We then combined all loops at different resolutions by the same approach as described above. We applied the lists of loop anchor to many downstream analyses by using Bedtools<sup>31</sup>, R, Python, or MATLAB, including (1) comparison of loop anchors between Micro-C and Hi-C or between different chromatin states; (2) distribution of loop strength or length; (3) cross-correlation with ChIP-seq, RNA-seq, and mNET-seq data; (4) ratio of boundary crossing, etc. For analysis of paired genomic loci (e.g., paired ChIP-seq peaks, genetic features, etc.) within a distance ranging from 2 kb to 2 Mb, we used Chromosight’s ‘quantify’ function to measure the probability of loop pattern for all intersections quantitatively. The loops were filtered by the following parameters: loop score >0.35 for 10-kb resolution, loop score >0.3 for 4-kb resolution, loop score >0.2 for 2-kb resolution, and the q-value lower than 10<sup>-5</sup> for all resolutions (**Supplemental Table 5**). For aggregate peak analysis (APA) to assess genome-wide loop intensity, loops were centered and piled up on a 20-kb x 20-kb matrix with 400-bp resolution balanced data or 50-kb x 50-kb matrix with 1-kb resolution balanced data. Contacts close to the diagonal were excluded and normalized by a random shift matrix to avoid distance decay effects. The ratio of loop enrichment was calculated by dividing normalized center contacts in a searching window by the normalized corner submatrices. We used the same approach and normalization method to analyze the genome-wide target-centered loop intensity. Instead of aggregating at the intersection of loop anchors, the matrix is centered at the paired ChIP-seq peaks or genomic features. Coolpup (<https://github.com/open2c/coolpuppy>)<sup>28</sup> has implemented the APA function for the COOL file. The lists of loops called by Mustache or Chromosight are available in **Supplemental Table 6 or 7**. The lists of loop quantification for cohesin, E-P, and P-P loops are available in **Supplemental Table 8 – 10**.

## Definition of chromatin states and structure observed by Micro-C

We first used the published ChromHMM (<http://compbio.mit.edu/ChromHMM>)<sup>32,33</sup> to define the chromatin states in mESCs, which subclassifies chromatin into 12 states including: 1) CTCF/insulator, 2) active promoter (designated as “P”), 3) strong enhancer, 4) medium enhancer, 5) weak enhancer, 6) mix of promoter and enhancer, 7) bivalent promoter, 8) gene body, 9) Polycomb repressor, 10) intergenic regions, 11) heterochromatin, and 12) repeats. To simplify the analysis, we further combined the groups of strong, medium, and weak enhancers and mix of promoter and enhancer into “enhancer” (designated as “E”). In this study, we use the terms that are widely accepted in the field to describe the chromatin structures in Micro-C contact maps as well as avoid any ambiguous description that implicates their biological functions if they have not been well characterized, including: 1) Topologically-associating domain (TAD): squares along matrix diagonal enriched with self-interactions, which are defined as genomic intervals demarcated by the boundaries characterized by the insulation score analysis or the arrowhead transformation analysis; 2) Cohesin loops: focal enrichment of contacts in contact maps with the co-enrichment of CTCF/cohesin ChIP-seq peaks at loop anchors, which is thought to be formed by active loop extrusion halted by CTCF;

3) E-P/P-P loops: focal enrichment of contacts in contact maps with the co-enrichment of chromatin states for “active promoter (P)” or “enhancer (E)” at loop anchors. Although not all cohesin loops and E-P/P-P loops are formed through “looping” and some studies suggest using “dots” instead of “loops”, to simplify and be consistent with the majority of findings, we chose to use “loops” over “dots” to describe these enhanced focal contacts in this manuscript.

## RNA-seq experiments and analysis

Total RNA was extracted from  $\sim 1 \times 10^7$  mES cells ( $\sim 70\%$  confluent P10 dish) with the standard TRIzol RNA extraction protocol. The abundant rRNAs were depleted from the sample using NEBNext rRNA Depletion Kit (NEB, #E6310). The rRNA-depleted RNAs were then subjected to RNA-seq library construction using the NEBNext® Ultra II Directional RNA Library Prep Kit for Illumina® (NEB, #E7765). The final RNA-seq libraries were amplified with 7 – 8 PCR cycles.

For RNA-seq analysis, we used Kallisto<sup>34</sup> to quantify the number of transcripts and performed Sleuth<sup>35</sup> analysis for DEGs identification according to the recommended settings in the walkthrough ([https://pachterlab.github.io/sleuth\\_walkthroughs/](https://pachterlab.github.io/sleuth_walkthroughs/)). The DEGs were identified using the Wald test with the  $q$ -value  $< 0.01$ .

## Nascent RNA-seq experiment and analysis

We used the nascent RNA-seq (mNET-seq) protocol described in Nojima et al<sup>36</sup> with minor changes. In brief, the chromatin fraction was purified from  $1 \times 10^7$  mES cells by the following procedure: 1) Wash cells with cold PBS twice; 2) Resuspend cells with 4 mL cold HLB+N (10mM Tris-HCl (pH 7.5), 10 mM NaCl, 2.5 mM MgCl<sub>2</sub> and 0.5 % NP-40) and incubate for 5 min on ice; 3) Add 1 ml cold HLB+NS (10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 2.5 mM MgCl<sub>2</sub>, 0.5 % NP-40 and 10 % Sucrose) under the layer of cell lysate; 4) Spin down cells and collect the nuclear pellet; 5) Add 120  $\mu$ L cold NUN1 (20 mM Tris-HCl (pH 7.9), 75 mM NaCl, 0.5 mM EDTA and 50 % Glycerol) and transfer sample to a new tube; 6) Add 1.2 mL cold NUN2 (20 mM HEPES-KOH (pH 7.6), 300 mM NaCl, 0.2 mM EDTA, 7.5 mM MgCl<sub>2</sub>, 1 % NP-40 and 1 M Urea) and incubate for 15 min on ice while vortexing every 3 min; 7) Spin down the chromatin pellet and wash with cold PBS once. Next, chromatin and RNA were digested in 100  $\mu$ L MNase digestion solution (1x micrococcal nuclease (MNase) buffer and 40 units/ $\mu$ L MNase (NEB, #M0247)) for 5 min at 37 °C while shaking at 1,400 rpm in a thermomixer. 25 mM EGTA was then added to inactivate the reaction. Digested/solubilized chromatin was then collected by centrifugation at 16,000xg for 5 min at 4 °C. The chromatin bound-Pol II complex was purified by the following steps: 1) Dilute 100  $\mu$ L of sample with 400  $\mu$ L cold NET-2 (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 0.05 % NP-40 and 1% Empigen BB (Sigma, cat no. 30326)); 2) Add 100  $\mu$ L Pol II antibody-conjugated beads (10  $\mu$ g Pol II 8WG16 antibody conjugated with Dynabeads™ Protein G for Immunoprecipitation (ThermoFisher, #1004D)) and incubate on a tube rotator for 1 hour in the 4 °C room; 3) Wash beads with 1 mL cold NET-2 for a total of 6 washes. RNA was then phosphorylated with T4 polynucleotide kinase (T4 PNK) with the following steps: 1) Wash beads with 50  $\mu$ L cold 1X PNKT buffer (1X T4 PNK buffer and 0.1 % Tween); 2) Resuspend beads with 50  $\mu$ L PNK reaction mixture (1X T4 PNKT, 1 unit/ $\mu$ L PNK and 1 mM ATP) for 5 min at 37 °C while shaking at 1,200 rpm in a thermomixer; 3) Wash beads with 1 mL cold NET-2. RNA was isolated by the standard TRIzol RNA extraction protocol with isopropanol RNA precipitation. Purified RNA was then dissolved in 10  $\mu$ L of Urea dye (7 M Urea, 0.05 % Xylene cyanol, 0.05 % Bromophenol blue) and resolved on a 6% TBE-Urea gel at 200 V for 5 min. To size select 30-160 nt RNAs, we cut the gel between the Bromophenol

blue and the Xylene cyanol dye markers. A 0.5-mL tube was punctured with 3-4 small holes by 26G needle and inserted in a 1.5-mL tube. Gel fragments were placed in the 0.5-mL tube and broken down by centrifugation at 16,000xg for 1 min. RNA was eluted by RNA elution buffer (1 M NaOAc and 1 mM EDTA) for 1 hr at 25°C in a Thermomixer shaking at 900 rpm. Eluted RNA was purified with SpinX column (Coster, #8160) with a glass filter (Whatman, #1823-010). The eluted RNA was purified again with ethanol precipitation. RNA libraries were prepared according to the protocol of the NEBNext Small RNA Library Prep Kit (NEB, #E7330). The mNET-seq library was obtained by PCR for 12 – 14 cycles.

For mNET-seq analysis, we wrote a custom pipeline to process raw data as follows: 1) Adapter trimming: we used TrimGalore (<https://github.com/FelixKrueger/TrimGalore>) to remove sequencing adapters ‘AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC’ and ‘GATCGTCGGACTGTAGAACTCTGAAC’ at each side of the reads; 2) Mapping: Trimmed reads were mapped to the mouse mm10 reference genome with STAR RNA-seq aligner<sup>37</sup>; 3) Identifying the last nucleotide incorporated by Pol II: We used the Python script mNET\_snr ([https://github.com/tomasgomes/mNET\\_snr](https://github.com/tomasgomes/mNET_snr)) to locate the 3’ nucleotide of the second read and the strand sign of the first read. The bigWig files were generated by using Deeptools as described above. To identified DEGs in mNET-seq, we used the NRSA (Nascent RNA Sequencing Analysis)<sup>38</sup> package to statistically quantify differential changes of mNET-seq signal at the gene body between UT- and IAA-treated cells.

## Single-particle imaging experiments

All single-molecule imaging experiments were performed with a similar setting as described in our previous studies<sup>4,39</sup>. In brief, the experiments were performed with a custom-built Nikon TI microscope equipped with a 100X/NA 1.49 oil-immersion TIRF objective (Nikon apochromat CFI Apo TIRF 100X Oil), an EMCCD camera (Andor iXon Ultra897), a perfect focus system to correct for axial drift and motorized laser illumination (Ti-TIRF, Nikon), and an incubation chamber maintaining a humidified 37 °C atmosphere with 5% CO<sub>2</sub> for the sample and the objective. Excitation was achieved using the following laser lines: 561 nm (1 W, Genesis Coherent) for JF549/PA-JF549 and TMR dyes; 633 nm (1 W, Genesis Coherent) for JF646/PA-JF646 dyes; 405 nm (140 mW, OBIS, Coherent) for all photo-activation experiments. Laser intensities were controlled by an acousto-optic Tunable Filter (AA Opto-Electronic, AOTFnc-VIS-TN) and triggered with the camera TTL exposure output signal. Lasers were directed to the microscope by an optical fiber, reflected using a multi-band dichroic (405 nm/488 nm/561 nm/633 nm quad-band, Semrock) and focused in the back focal plane of the objective. Emission light was filtered using single band-pass filters placed in front of the camera (Semrock 593/40 nm for TMR and JF549/PA-JF549 and Semrock 676/37 nm for JF646/PA-JF646). The angle of incident laser was adjusted for highly inclined laminated optical sheet (HiLo) conditions<sup>40</sup>. The microscope, cameras, and hardware were controlled through the NIS-Elements software (Nikon).

For ‘fast-tracking’ stroboscopic illumination (spaSPT) at ~133 Hz, the excitation laser (633 nm for PA-JF646 or 561 nm for PA-JF549) was pulsed for 1 ms at maximum (1 W) power at the beginning of the frame interval, while the photoactivation laser (405 nm) was pulsed during the ~447 μs camera transition time. Each frame consisted of a 7-ms camera exposure time followed by a ~447-μs camera inactive time. The camera was set for frame transfer mode and vertical shift speed at 0.9 μs. With this setup, the pixel size after magnification is 160 nm and the photon-to-grayscale gain is 109. Typically, 30000 frames with this sequence were collected per nucleus, during which the 405-nm intensity was manually tuned to maintain an average molecule density of ~0.5 localizations per frame, corresponding to ~15,000 localizations per cell per movie. Maintaining a very low density of molecules is necessary to avoid tracking errors.



For ‘slow-tracking’ (slow-SPT) experiments, long exposure times (50 ms, 100 ms, and 250 ms) and low constant illumination laser intensities (0.5% - 2% of 0.5 W power) were used to measure residence time. The camera was set for normal mode and vertical shift speed at 3.3  $\mu$ s. We generally recorded each cell with 1200 frames for a 250 ms exposure time, 3000 frames for a 100 ms exposure time, or 6000 frames for a 50 ms exposure time. We included H2B-HaloTag cells for the photobleaching correction for each experiment.

For PALM experiments, continuous illumination was used for both the main excitation laser (633 nm for PA-JF646 or 561 nm for PA-JF549) and the photo-activation laser (405 nm). The intensity of the 405 nm laser was gradually increased over the course of the illumination sequence to image all molecules and avoid too many molecules being activated at any given frame. The camera was set for 25-ms exposure time, frame transfer mode, and vertical shift speed at 0.9  $\mu$ s. In total, 40,000–60,000 frames were recorded for each cell (~20–25 min), which was sufficient to image and bleach all labeled molecules.

### spaSPT analysis

For analysis of spaSPT experiments, we used the QUOT package (<https://github.com/alecheckert/quot>) to generate trajectories from raw spaSPT movies with the steps of spot detection, subpixel localization, and tracking. All localization and tracking for this manuscript were performed with the following settings: 1) Detection: generalized log likelihood ratio test with a 2D Gaussian kernel (‘11r’ with  $k = 1.0$ , pixel window size ( $w$ ) = 15, and a log ratio threshold ( $t$ ) = 26.0); 2) Subpixel localization: Levenberg-Marquardt fitting of a 2D integrated Gaussian point spread function model (‘1s\_int\_gaussian’ with pixel window size ( $w$ ) = 9, sigma = 1.0, ridge = 0.001, maximal iterations = 20 per PSF, and damping term = 0.3). 3) Tracking: we chose to use a conservative tracking algorithm with a 1.3- $\mu$ m search radius (‘conservative’ with search radius = 1.3 and maximal blinks = 0). This setting makes the algorithm search for spot reconnections unambiguously, meaning that no other reconnections are possible within the specified search radius. Jumps are discarded if other reconnection possibilities given the search radius exist.

We next used the Spagl package (<https://github.com/alecheckert/spagl>)<sup>41</sup> to estimate the likelihood of diffusion coefficients for each trajectory. The detailed discussion is available in Heckert et al<sup>41</sup>. In brief, we applied “State Array (SA)”, a grid of state parameters that span a range of diffusion coefficients (0 to 100  $\mu$ m<sup>2</sup>/s), to calculate the posterior occupations of each point in the grid. The SA method conceptually produces a similar result as the Dirichlet process mixture models (DPMM) and retains its ability to model complex diffusive mixtures, while mitigating the issue of expensive likelihood functions. Instead of allowing infinite number of states ( $K \rightarrow \infty$ ), the method fixes the number of states at a large but finite value. For each state  $j = 1, \dots, K$ , the algorithm chooses a fixed set of state parameters  $\theta_j$ . The model simplifies to:

$$\begin{aligned} \tau &\sim \text{Dirichlet}\left(\frac{\alpha}{K}, \dots, \frac{\alpha}{K}\right) \\ Z_i | \tau &\sim \text{Multinomial}(\tau, 1) \\ X_i | (Z_{ij} = 1) &\sim f_X(x | \theta_j) \end{aligned}$$

As a result, the SA method can compute more complex likelihood functions that incorporate localization error. Next, the algorithm infers the posterior  $p(Z, \tau | X)$  with the Dirichlet prior  $p(\tau)$  and corrects the systematic overestimation for the occupations of slow states due to defocalization. By marginalizing the posterior distribution on localization error, the method naturally incorporates uncertainty about the localization error of different states.

Alternatively, we analyzed the spaSPT data with the kinetic modeling framework implemented in the Spot-On package<sup>39</sup>. Briefly, the model infers the diffusion constant and relative fractions of two or three subpopulations from the distribution of displacements computed at increasing lag times ( $1\Delta\tau$ ,  $2\Delta\tau$ ,...). This is performed by fitting a semi-analytical model to the empirical histogram of displacements using non-linear least squares fitting. Defocalization is explicitly accounted for by modeling the fraction of particles that remain in focus over time as a function of their diffusion constant. We used the following setting for Spot-On analysis in the manuscript: TimePoints = 8; BinWidth = 0.010; JumpsToConsider = 4; MaxJump = 5.05; ModelFit = CDF-fitting; NumberOfStates = 3; FitIterations = 3; FitLocErrorRange = 0.010-0.075; LocError = 0.035; Dbound range = 0.0001-0.05; Dfree range = 0.5-25.

## Slow-SPT analysis

For analysis of slow-SPT experiments, we used the following tracking settings for this manuscript: 1) Detection: '11r' with  $k = 1.0$ ,  $w = 15$ ,  $t = 18$ ; 2) Subpixel localization: 'ls\_int\_gaussian' with  $w = 9$ ,  $\sigma = 1.0$ ,  $\text{ridge} = 0.001$ , maximal iteration = 20, and  $\text{damp} = 0.3$ ; 3) Tracking: 'euclidean' with search radius = 0.5, maximal blinks = 1, and maximal diffusion constant ( $\mu\text{m}^2/\text{s}$ ) = 0.08.

To extract residence times from slow-SPT data, we used long exposure times (50 ms, 100 ms, or 250 ms) to motion-blur freely diffusing molecules into the background<sup>4,42-44</sup>. We then recorded the trajectory length of each 'bound' molecule and used these to generate a survival curve (1-CDF), and performed double-exponential fitting to estimate the unbinding rates for non-specific binding ( $K_{ns}$ ) and specific binding ( $K_s$ ). We note that localization errors can cause both false-positive and false-negative detections. The  $K_{ns}$  is likely to be contaminated by localization errors (e.g., from molecules close to being out-of-focus) and experimental noise. To filter out contributions from tracking errors and slow-diffusing molecules, we applied an objective threshold as previously described to consider only particles tracked for at least  $N_{min}$  frames. To determine  $N_{min}$ , we plotted the inferred residence time as a function of  $N_{min}$  and observed convergence to a single value after  $\sim 2.5$  s. We thus used this threshold to determine the value of  $K_s$ . To correct the biases from photobleaching, cell drifting, and background fluctuating, we assume that all these factors should affect H2B-HaloTag to the same extent as those affecting YY1-HaloTag. We can use an apparent unbinding rate for H2B-HaloTag as  $K_{bias}$ , consistent with our FRAP analysis. Thus, we performed the slow-SPT experiments for YY1-HaloTag and H2B-HaloTag with the same camera and laser settings on the same day. We then obtained the residence time as:

$$\tau_s = \frac{1}{k_s - k_{bias}} = \frac{1}{k_{s, \text{true}}}$$

## Fluorescence recovery after photobleaching (FRAP) imaging analysis

FRAP was performed on an inverted Zeiss LSM 900 Axio Observer confocal microscope equipped with Airyscan 2 detector, a motorized stage, a full incubation chamber maintaining 37°C/5% CO<sub>2</sub>, a heated stage, an X-Cite 120 illumination source as well various laser lines. Images were acquired on a 40x Plan NeoFluar NA1.3 oil-immersion objective at a zoom corresponding to a 76 nm x 76 nm pixel size. The microscope was controlled using the Zeiss Zen imaging software.

In this manuscript, we recorded 60 sec of movies for YY1-HaloTag at one frame per 250 ms, corresponding to a total of 240 frames. The first 20 frames were acquired before the bleach pulse, allowing us to accurately

measure baseline fluorescence. A circular bleach spot ( $r = 6$  pixels) was chosen in a region of homogenous fluorescence at a position at least  $1 \mu\text{m}$  from nuclear or nucleolar boundaries. Alternatively, we bleached a square at one corner of nucleus, which reduces noise while introducing some uncertainty for our downstream fitting analysis. The spot was bleached using maximal laser intensity and pixel dwell time corresponding to a total bleach time of  $\sim 1$  s. We note that because the bleach duration was relatively long compared to the timescale of molecular diffusion, it is not possible to accurately estimate the bound and free fractions from our FRAP curves.

To quantify FRAP movies, we wrote a pipeline in MATLAB. Briefly, our algorithm automatically detects the bleached spot, the background spot, and the nucleus segments by Gaussian smoothing, hole-filling, and segmenting a nucleus in a FRAP movie. Cell drift is also automatically corrected by the optimal linear translation in  $x$  and  $y$ . Next, we quantify the bleach spot signal as the mean intensity of a slightly smaller region, which is more robust to lateral drift. The FRAP signal is corrected for photobleaching using the measured reduction in total nuclear fluorescence and internally normalized to its mean value during the 20 frames before bleaching. Finally, corrected FRAP curves from each single cell were averaged to generate a mean FRAP recovery. We used the mean FRAP recovery in all figures and for model-fitting.

Model selection is critical to infer the parameters from FRAP experiments. Sprague et al.<sup>45</sup> suggested that when:

$$\frac{k_{\text{ON}}^* w^2}{D_{\text{FREE}}} \ll 1 \text{ and } \frac{k_{\text{OFF}}}{k_{\text{ON}}^*} \lesssim 1$$

Then a ‘reaction dominant’ FRAP model is most appropriate. For YY1:

$$\frac{k_{\text{ON}}^* w^2}{D_{\text{FREE}}} = 0.00256 \ll 1$$

Thus, a reaction-dominant FRAP model is the most suitable choice for YY1’s FRAP modeling. Sprague et al.<sup>45</sup> demonstrated that the FRAP recovery depends only on  $k_{\text{OFF}}$  in the reaction-dominant regime. We thus fit the FRAP curves to the model and applied the slower off rate to estimate the residence time according to  $\tau_s = \frac{1}{k_{\text{off}}}$ .

### Inferring parameters related to YY1’s target search mechanism

We used the parameters inferred from our spaSPT and the residence time measurements from our FRAP or slow-SPT analysis. The detailed discussion is available in Hansen et al.<sup>4</sup>. Briefly, from the Spagl State Array analysis, we determined that the total bound fraction for YY1 is  $\sim 31\%$ . However, the total bound fraction ( $0 - 0.1 \mu\text{m}^2/\text{s}$ ) contains both YY1 molecules bound specifically to their target sites and non-specific interactions (e.g., sliding or jumping on DNA). We previously estimated the fraction that is non-specifically bound using a mutant CTCF with a His-to-Arg mutation in each of the 11 zinc-finger domains<sup>4</sup>. This CTCF mutant is virtually unable to interact specifically with its binding sites. The Spagl analysis estimated the bound fraction to be  $\sim 8.1\%$  for this mutant in mESCs. Since we did not perform the spaSPT experiments for YY1’s DNA binding domain mutants, we thus inferred the  $F_{\text{BOUND,specific}} \sim F_{\text{BOUND,total}}$  in this manuscript.

We next determined the average time for YY1 to find its cognate site after dissociating from the previous site. We will use ‘ $s$ ’ and ‘ $ns$ ’, as abbreviations for specific and non-specific binding, respectively, in the

following discussion. The pseudo-first-order rate constant for specific binding sites,  $k_{ON,s}^*$ , is related to the fraction bound by:

$$F_{BOUND,s} = \frac{k_{ON,s}^*}{k_{ON,s}^* + k_{OFF,s}^*} \Leftrightarrow k_{ON,s}^* = \frac{F_{BOUND,s} k_{OFF,s}^*}{1 - F_{BOUND,s}}$$

We determined the off-rate for a specific interaction in our residence time measurements. Thus, we can calculate  $k_{ON,s}^*$ , which is directly related to the average search time for a specific YY1-binding site:

$$\tau_{search,s} = \frac{1}{k_{ON,s}^*} = \frac{1 - F_{BOUND,s}}{F_{BOUND,s} k_{OFF,s}^*}$$

After plugging in these determined parameters of  $F_{BOUND,s}$  and  $k_{OFF,s}^*$ , we obtained total search times for YY1 of ~28.3 s in wild type mES cells, ~31.6 s in CTCF-depleted cells, ~61.4 s in RAD21-depleted cells. We inferred the residence time estimated from the slow-SPT data with 100 ms exposure time or the FRAP analysis in this manuscript.

## PALM analysis

For analysis of PALM experiments, we used the publicly available ThunderSTORM package (<https://github.com/zitmen/thunderstorm>)<sup>46</sup> with the following setting for this manuscript: 1) Image filtering: 'Wavelet filter (B-Spline)' with B-Spline order = 3 and B-Spline scale = 2.0; 2) Approximate localization: 'Local maximum' with peak intensity threshold = 1.5\*std(Wave.F1) and 8-neighbourhood connectivity; 3) Subpixel localization: 'Integrated Gaussian' with fitting radius = 3 pixels, fitting method = maximum likelihood, initial sigma = 1.6, multi-emitter analysis is disabled; 4) Image reconstruction: 'Averaged shifted histogram'. After tracking, we further filtered ambiguous emitters with the following setting: 1) Filtering: frame > 100 & intensity > 100 & sigma < 220 & uncertainty\_xy < 50; 2) Merge: Max distance = 10 & Max frame off = 1 & Max frames = 0; 3). Remove duplicates is enabled. This setting combines the blinking molecules into one and removes the multiple localizations in a frame.

## Antibodies

See **Supplementary Table 1** for a complete list of the antibodies used in this study.

## Datasets and accession numbers

The Micro-C, ChIP-seq, nascent RNA-seq and total RNA-seq data generated in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE178982. We also reanalyzed data that we previously generated in wild type mESCs (GSE130275)<sup>13</sup>. spaSPT raw data are accessible through DOI: [10.5281/zenodo.5035837](https://doi.org/10.5281/zenodo.5035837).

## Reference



- 592 1. Pettitt, S. J. *et al.* Agouti C57BL/6N embryonic stem cells for mouse genetic resources. *Nat Methods* **6**,  
593 493–495 (2009).
- 594 2. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281–2308  
595 (2013).
- 596 3. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome  
597 editing experiments and screens. *Nucleic Acids Res* **46**, gky354- (2018).
- 598 4. Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate  
599 chromatin loop stability with distinct dynamics. *Elife* **6**, e25776 (2017).
- 600 5. Natsume, T., Kiyomitsu, T., Saga, Y. & Kanemaki, M. T. Rapid Protein Depletion in Human Cells by  
601 Auxin-Inducible Degron Tagging with Short Homology Donors. *Cell Reports* **15**, 210–218 (2016).
- 602 6. Li, S., Prasanna, X., Salo, V. T., Vattulainen, I. & Ikonen, E. An efficient auxin-inducible degron  
603 system with low basal degradation in human cells. *Nat Methods* **16**, 866–869 (2019).
- 604 7. Testa, A. *et al.* Chromatin Immunoprecipitation (ChIP) on Chip Experiments Uncover a Widespread  
605 Distribution of NF-Y Binding CCAAT Sites Outside of Core Promoters\*. *J Biol Chem* **280**, 13606–13615  
606 (2005).
- 607 8. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359  
608 (2012).
- 609 9. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
- 610 10. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis.  
611 *Nucleic Acids Res* **44**, W160–W165 (2016).
- 612 11. Tu, S. *et al.* MAnorm2 for quantitatively comparing groups of ChIP-seq samples. *Genome Res* **31**,  
613 131–145 (2020).
- 614 12. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676–  
615 682 (2012).
- 616 13. Hsieh, T.-H. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin  
617 Folding. *Mol Cell* **78**, 539-553.e8 (2020).
- 618 14. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol*  
619 **16**, 259 (2015).
- 620 15. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled  
621 arrays. *Bioinformatics* (2019) doi:10.1093/bioinformatics/btz540.
- 622 16. Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C  
623 Experiments. *Cell Syst* **3**, 95–98 (2016).

624 17. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization.  
625 *Nat Methods* **9**, 999–1003 (2012).

626 18. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. *Ima J Numer Anal* **33**, 1029–1047  
627 (2012).

628 19. Kerpedjiev, P. *et al.* HiGlass: web-based visual exploration and analysis of genome interaction maps.  
629 *Genome Biol* **19**, 125 (2018).

630 20. Yardımcı, G. G. *et al.* Measuring the reproducibility and quality of Hi-C data. *Genome Biol* **20**, 57  
631 (2019).

632 21. Sauria, M. E. & Taylor, J. QuASAR: Quality Assessment of Spatial Arrangement Reproducibility in  
633 Hi-C Data. *Biorxiv* 204438 (2017) doi:10.1101/204438.

634 22. Ursu, O. *et al.* GenomeDISCO: a concordance score for chromosome conformation capture  
635 experiments using random walks on contact map graphs. *Bioinformatics* **34**, 2701–2707 (2018).

636 23. Hsieh, T.-H. S. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C.  
637 *Cell* **162**, 108–119 (2015).

638 24. Hsieh, T.-H. S., Fudenberg, G., Goloborodko, A. & Rando, O. J. Micro-C XL: assaying chromosome  
639 conformation from the nucleosome to the entire genome. *Nat Methods* **13**, 1009–1011 (2016).

640 25. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding  
641 Principles of the Human Genome. *Science* **326**, 289–293 (2009).

642 26. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage  
643 compensation. *Nature* **523**, 240–4 (2015).

644 27. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of  
645 Chromatin Looping. *Cell* **159**, 1665–1680 (2014).

646 28. Flyamer, I. M., Illingworth, R. S. & Bickmore, W. A. Coolpup.py: versatile pile-up analysis of Hi-C  
647 data. *Bioinformatics* **36**, 2980–2985 (2020).

648 29. Ardakany, A. R., Gezer, H. T., Lonardi, S. & Ay, F. Mustache: multi-scale detection of chromatin  
649 loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biol* **21**, 256 (2020).

650 30. Matthey-Doret, C. *et al.* Computer vision for pattern detection in chromosome contact maps. *Nat*  
651 *Commun* **11**, 5795 (2020).

652 31. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features.  
653 *Bioinformatics* **26**, 841–842 (2010).

654 32. Pintacuda, G. *et al.* hnRNPK Recruits PCGF3/5-PRC1 to the Xist RNA B-Repeat to Establish  
655 Polycomb-Mediated Chromosomal Silencing. *Mol Cell* **68**, 955–969.e10 (2017).

656 33. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat*  
657 *Methods* **9**, 215–6 (2012).

658 34. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq  
659 quantification. *Nat Biotechnol* **34**, 525–527 (2016).

660 35. Pimentel, H., Bray, N. L., Puente, S., Melsted, P. & Pachter, L. Differential analysis of RNA-seq  
661 incorporating quantification uncertainty. *Nat Methods* (2017) doi:10.1038/nmeth.4324.

662 36. Nojima, T., Gomes, T., Carmo-Fonseca, M. & Proudfoot, N. J. Mammalian NET-seq analysis defines  
663 nascent RNA profiles and associated RNA processing genome-wide. *Nat Protoc* **11**, 413–428 (2016).

664 37. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

665 38. Wang, J. *et al.* Nascent RNA sequencing analysis provides insights into enhancer-mediated gene  
666 regulation. *Bmc Genomics* **19**, 633 (2018).

667 39. Hansen, A. S. *et al.* Robust model-based analysis of single-particle tracking experiments with Spot-  
668 On. *Elife* **7**, e33125 (2018).

669 40. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear  
670 single-molecule imaging in cells. *Nat Methods* **5**, 159–161 (2008).

671 41. Heckert, A., Dahal, L., Tjian, R. & Darzacq, X. Recovering mixtures of fast diffusing states from  
672 short single particle trajectories. *Biorxiv* 2021.05.03.442482 (2021) doi:10.1101/2021.05.03.442482.

673 42. Sergé, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target tracing to probe  
674 spatiotemporal cartography of cell membranes. *Nat Methods* **5**, 687–694 (2008).

675 43. Chen, J. *et al.* Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell*  
676 **156**, 1274–85 (2014).

677 44. Mazza, D., Abernathy, A., Golob, N., Morisaki, T. & McNally, J. G. A benchmark for chromatin  
678 binding measurements in live cells. *Nucleic Acids Res* **40**, e119–e119 (2012).

679 45. Sprague, B. L., Pego, R. L., Stavreva, D. A. & McNally, J. G. Analysis of Binding Reactions by  
680 Fluorescence Recovery after Photobleaching. *Biophys J* **86**, 3473–3495 (2004).

681 46. Ovesný, M., Křížek, P., Borkovec, J., Švindrych, Z. & Hagen, G. M. ThunderSTORM: a  
682 comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging.  
683 *Bioinformatics* **30**, 2389–2390 (2014).