The role of cohesin loading at enhancers in the flux of loop extrusion and long-range transcriptional control

Erika C. Anderson[1], Hadi Rahmaninejad[2], Abrar Aljahani[3], Emily M. Arnold[1], Annie S. Adachi[1], Rini Shah[1], Karissa L. Hansen[1], Ivana Cavka[3], Alistair N. Boettiger[3], Geoffrey Fudenberg[2], Elphège P. Nora[1,4,5,*]

1. Cardiovascular Research Institute, University of California, San Francisco, San Francisco, CA, USA
2. Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA, USA
3. Department of Developmental Biology, Stanford University; Stanford, CA 94305, USA
4. Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA
5. Chan-Zuckerberg Biohub San Francisco, CA, USA
*correspondence: elphege.nora@ucsf.edu

Abstract

Enhancers have been proposed to act as privileged loading sites for cohesin, raising the idea that they actively fold the genome to engage distal target promoters for transcription. Supporting this idea, NIPBL/MAU2, which is required for cohesin loading, binds at enhancers in mouse embryonic stem cells. However, we find that driving cohesin recruitment near an enhancer strongly inhibits transcription from its target distal promoter, indicating that strong focal cohesin loading at enhancers is not compatible with their long-range regulatory functions. Quantitative experiments and biophysical modeling further indicate that cohesin loading at enhancers does not make major contributions to genome-wide cohesin binding and chromosome folding patterns. Instead, cohesin must load throughout the genome to extrude it, regardless of enhancer proximity, with the major determinants of cohesin traffic being extrusion barriers such as transcription and clustered CTCF sites. These findings indicate that enhancer function is ancillary to the general mechanisms of chromosome folding, informing further study of the relationship between genome architecture and transcriptional regulation.

Main Text

Introduction

Enhancers are genetic elements able to modulate the activity of target promoters over long genomic distances. Such regulatory communication is supported by chromatin loops that bring distal loci into physical proximity within the cell nucleus[1,2]. The cohesin complex plays a key role in this process by dynamically extruding chromatin loops[3,4]. Yet, how the flux of cohesin loop extrusion is controlled to support long-range enhancer-promoter communication remains unclear.

Several studies have proposed that enhancers themselves may serve as loading sites for cohesin[5–12], thereby directly initiating chromatin loops to contact distal promoters (Fig. 1A). In this view, enhancer function would extend to the reshaping of chromosome folding in ways that support the delivery of their regulatory activity to distal promoters.

Support for targeted cohesin loading at enhancers initially came from the observation that, in mammalian genomes, cohesin binds distal regulatory elements along with NIPBL/MAU2[10,13–18]. NIPBL/MAU2 was originally defined as the cohesin loader complex, as it is required to initiate (but not maintain) topological cohesin binding during sister chromatid cohesion[19,20]. In the context of loop extrusion, NIPBL/MAU2 is required for both initiation and extension of extrusion[21–24]. The presence of NIPBL/MAU2 has therefore been interpreted as an indication of cohesin loading and looping from enhancers[10,13,15]. This interpretation was reinforced by the observations that enhancers can locally increase cohesin occupancy[8,11] and create characteristic chromosome folding signatures such as Hi-C jets/fountains and stripes across organisms[10,25–30].

However, for loading at enhancers to support long-range looping with distal promoters, cohesin would need to remain stably anchored at the enhancer during extrusion; otherwise cohesin would extrude away from the enhancer before reaching the target promoter (Fig. 1A). Yet, evidence for systematic anchoring of cohesin at enhancers is lacking.

At the same time, alternative models have suggested that cohesin loading is not restricted to discrete NIPBL/MAU2 binding sites (Fig. 1A), but may instead occur in a largely uniform fashion across the genome. Indeed, quantitative simulations of chromosome folding better fit Hi-C and imaging results (including stripes) when they assume uniform rather than targeted cohesin loading[31–33]. Furthermore, many enhancers regulate promoters several hundreds of kilobases away[34–36], at distances that greatly exceed the average processivity of cohesin molecules that would load at enhancers. Such distances could be bridged by multiple cohesins that load at different loci, but not by exclusive loading at the enhancer. Finally, cohesin loading cannot simply be restricted to topologically associating domains (TADs) containing enhancers, as constitutive heterochromatin is also extruded(Spracklin et al., 2023).

Together, these considerations raise the question to what extent enhancers actively shape chromatin interactions by acting as the loading sites for extrusive cohesin, versus simply operating within the ongoing flux of cohesin extrusion. Despite the relevance to our understanding of the coupling between transcriptional regulation and chromosome folding, discriminating between these possibilities has proven challenging, as it requires interpreting steady-state measurements of cohesin binding and Hi-C within mechanistic models of extrusion dynamics.

Here, we investigate the idea that long-range transcriptional regulation in mammalian cells is facilitated by loading of extrusive cohesin at enhancers through the recruitment of NIPBL/MAU2. We find that, while transcription factors can recruit NIPBL/MAU2 at enhancers, most cohesin molecules load outside enhancers. Artificially recruiting MAU2 to increase cohesin loading and looping near enhancers fails to support distal promoter activation and is instead detrimental to transcription, indicating that highly targeted cohesin loading is incompatible with long-range regulation by enhancers. Consistent with this, genome-wide analyses and quantitative biophysical simulations show that targeted loading at enhancers contributes little, if at all, to the general flux of cohesin extrusion across chromosomes. Instead, cohesin positioning genome-wide is best explained by pervasive loading combined with blocking by transcription and clustered CTCF sites. Together, these results indicate that enhancers do not initiate cohesin loops to support their own action in long-range transcriptional control, and that instead their regulatory opportunities are largely subordinate to the global dynamics of cohesin loop extrusion.

Results

Transcription factors recruit NIPBL/MAU2 at enhancers without substantially enriching cohesin

Enhancers have initially been proposed to act as loading sites for extrusive cohesin because they can bind NIPBL/MAU2[10,15]. Given the reported technical difficulties with NIPBL ChIP-seq[31,37], we first sought to identify high-confidence NIPBL and MAU2 binding sites. We performed spike-in calibrated ChIP-seq with an HA antibody in mouse embryonic stem cells (ESCs) with either endogenous NIPBL or MAU2 tagged with both an HA epitope tag and a dTAG-inducible degron and identified ~24,000 binding sites (Extended Data Fig. 1A). These binding sites were absent in cells where NIPBL or MAU2 was degraded and in cells without an HA tag, ruling out any false positives resulting from antibody binding nonspecifically or to hyper-ChIPable open chromatin (Extended Data Fig. 1A). As expected, HA-NIPBL and HA-MAU2 binding closely matched each other (Extended Data Fig. 1A), and ChIP-seq with an antibody for endogenous NIPBL showed binding at these same sites (Extended Data Fig. 1B). Although these NIPBL and MAU2 ChIP peaks are reproducible, the ChIP signal (with multiple different antibodies) displayed much lower enrichment above background compared to ChIP for other factors (Extended Data Fig. 1F-G). This low enrichment suggests that a substantial amount of the NIPBL/MAU2 molecules bind – and possibly function – outside the peaks.

We analyzed the genomic locations of NIPBL/MAU2 binding, and found that NIPBL/MAU2 was enriched at predicted enhancers (defined as sites bound by the pluripotency transcription factors OCT4, SOX2, and NANOG[38]). In contrast, binding was much less pronounced at active transcription start sites, which show high enrichment for H3K27ac but not pluripotency transcription factors (Fig. 1B-C, Extended Data Fig. 1C-D). Therefore, NIPBL/MAU2 does not simply bind all active chromatin but tracks the binding of pluripotency transcription factors.

To further explore the specificity of NIPBL/MAU2 binding at enhancers, we tested whether transcription factor binding is sufficient to recruit NIPBL/MAU2 and load cohesin (Fig. 1D). We used human U2OS cells with a microscopically visible TetO array spanning ~4 Mb[39], a recruitment platform large enough for loaded cohesin to not simply extrude away. Tethering OCT4 and VP64 to the TetO array was sufficient to recruit endogenous NIPBL and MAU2 (Fig.

1E-F, Extended Data Fig. 2A-B) indicating that the ChIP-seq enrichment of NIPBL/MAU2 at enhancers in ESCs is indeed due to recruitment by transcription factors (directly or indirectly via transcriptional cofactors), and that NIPBL/MAU2 recruitment is not limited to pluripotency factors.

Cohesin did not, however, localize to the TetO array upon OCT4 or VP64 recruitment, despite the presence of NIPBL/MAU2. The lack of detectable cohesin upon transcription factor tethering suggests that NIPBL/MAU2 binding to active regulatory elements is not necessarily molecularly coupled to local cohesin recruitment. The low amount of NIPBL/MAU2 recruited by transcription factors is therefore insufficient to cause detectable cohesin accumulation, or NIPBL/MAU2 is not readily competent for loading cohesin when recruited by transcription factors. Recruiting larger amounts of MAU2 by direct fusion to TetR visibly enriched cohesin (Fig. 1E, Extended Data Fig. 2A-B) – demonstrating that upon sufficient loading, cohesin is readily detected by this assay.

Very strong cohesin loading sites are incompatible with regulation by enhancers

After confirming that NIPBL/MAU2 binds enhancers, we next tested the model that focal cohesin loading at enhancers increases interactions with distal promoters and benefits long-range transcriptional activation. We used the TACL system[40] to force cohesin loading at the *Car2* locus in ESCs (Fig. 2A), which depends on cohesin for transcriptional activation by distal enhancers[41]. In a cell line where one *Car2* allele encodes a GFP reporter while the other allele encodes RFP (Extended data Fig. 3A), we heterozygously inserted a TetO array either beside the promoter, between the promoter and enhancers, or beside the enhancers, and expressed TetR-MAU2 to target cohesin loading at the TetO array (Fig. 2B). In contrast to the expectation that increasing cohesin loading would bolster activation by the distal enhancers, we observed strong downregulation of *Car2* upon forcing cohesin loading from any of the three TetO locations (Fig. 2B, Extended Data Fig. 3B). Reporter expression was unaffected by TetR alone binding to the TetO array, and expression of the *Car2* reporter on the control allele without TetO insertions was unchanged, ruling out possible indirect effects of MAU2 overexpression (Extended Data Fig. 3C). Thus, despite the dependence of *Car2* expression on loop extrusion[41], strong cohesin loading at the enhancer impairs transcriptional activity.

To gain insight into the changes in chromatin folding caused by targeted cohesin loading, we used optical reconstruction of chromosome architecture (ORCA, Fig. 2C), which allowed us to specifically probe the *Car2* allele harboring the TetO array (Extended Data Fig. 4, Methods). Targeting cohesin loading either beside the enhancers or beside the promoter did not increase enhancer-promoter interactions, which remained largely unaffected (Fig. 2D and 2E). Targeting loading between the enhancers and the promoter separated them by creating a boundary that slightly reduced their interactions (Fig. 2D and 2E). Therefore, targeted loading-induced transcriptional repression, which occurs from all three configurations, cannot be simply explained by a change in enhancer-promoter interactions. Rather, high levels of extruding cohesin appear to repress transcription by disrupting enhancer or promoter activity, possibly by interfering with assembly or progression of the transcription machinery. Since strong cohesin loading at enhancers is detrimental to transcription from distal promoters, it is unlikely that enhancers generally serve as strong loading sites.

Polymer simulations show that focal cohesin loading can only moderately impact cohesin traffic

Given our observation that high targeted cohesin loading compromises enhancer function, we sought to establish how more modest loading at NIPBL/MAU2 sites would influence cohesin traffic and chromosome folding. As cohesin at NIPBL/MAU2 sites has been interpreted as a signature of targeted loading[10,42], we first used biophysical simulations[43] to explore how much cohesin binding is expected at loading sites across a range of loading rates (Fig. 3A). Using extrusion parameters selected to match measurements in mouse ESCs (Methods), we simulated a uniform background loading rate with elevated loading at a target site and found that bidirectionally extruding cohesin would not accumulate at loading sites, even with up to a 500-fold boost in on-target loading over the background loading rate (Fig. 3B). Instead, cohesin rapidly extrudes away from the loading site.

We next explored to what extent loading sites are expected to increase cohesin accumulation at nearby CTCF sites. To address this, we added to the simulations two dynamic barriers[44] flanking the loading site at varying distances (Fig. 3C). The simulations revealed that although targeted loading can, in principle, influence cohesin traffic, the impact remains limited. Even for CTCF barriers only 5 kb from the loading site, increasing cohesin accumulation two-fold would require more than a 500-fold boost in on-target loading rate (Fig. 3D). The reason focal loading has limited impact is that the amount of cohesin contributed by targeted loading sites is overshadowed by the cohesins loaded elsewhere along the chromosome. For example, with on-target loading increased 500-fold, a 250 bp loading site would not contribute more than background loading from the 125 kb of surrounding chromatin, thereby only increasing cohesin accumulation at the CTCF barrier ~two-fold. Thus, cohesin traffic to nearby barriers is expected to be dominated by chromosome-wide loading rather than by even strongly privileged loading sites.

Finally, we explored the signatures of targeted cohesin loading on Hi-C patterns. 3D polymer models indicated that high focal loading causes the emergence of fountain patterns in contact maps (Extended Data Fig. 5A-B), as observed in previous simulations [25,26]. Yet, these patterns do not increase the interactions of the loading site with its surrounding genomic environment (Extended Data Fig. 5A-B, Supplementary video), as cohesin extrudes away from the loading site. We observed that, at high loading rates, wider multi-kilobase loading sites create a boundary with stripes on contact maps (Extended Data Fig. 5B), due to multiple closely loaded cohesins colliding with each other. Although such a boundary at a large loading site can anchor cohesin, our TACL results indicate that such a configuration impairs long-range transcriptional regulation (Fig. 2).

Most cohesin resides outside CTCF sites

Having defined expectations of how targeted loading can influence cohesin positioning, we next quantified cohesin localization in ESCs to determine the extent to which NIPBL/MAU2 sites at enhancers contribute to initiating loop extrusion. In vertebrates, core cohesin subunits exhibit prominent ChIP peaks at CTCF sites [45–47], where loop extrusion is blocked. However, cohesin ChIP-seq also exhibits diffuse signal between CTCF peaks, which may reflect true translocating cohesin or non-specific background from the antibody[48]. This ambiguity complicates quantifications, as exemplified by antibodies against distinct cohesin subunits yielding markedly different fractions of reads at CTCF peaks (12% for RAD21 antibody vs. 5% for SMC1A antibody, Fig. 4A,C).

To determine how much of the detected ChIP signal is due to nonspecific antibody binding, we performed spike-in calibrated cohesin ChIP after cohesin depletion (Fig. 4B). We depleted RAD21 with an auxin-inducible degron[49–51] for 3.5 hours (Extended Data Fig. 6A-B) and performed ChIP-seq for RAD21 and SMC1A. This allowed quantifying and subtracting the antibody background from the ChIP-seq signal obtained in cells with normal cohesin levels. Using this approach, we found that only ~25% of the cohesin ChIP-seq signal is in the 0.8% of the genome bound by CTCF (Fig. 4B). This background subtraction procedure yielded strong concordance between the RAD21 and SMC1A antibodies, despite their different specificities prior to correction (Fig. 4B). We note that 25% is an upper bound for the fraction of cohesin binding in CTCF peaks, as signal remaining after auxin treatment also includes any undegraded bound cohesin (less than 4%, Extended Data Fig. 6A-B). Thus, the majority of cohesin molecules are positioned between CTCF peaks, where they are likely engaged in extrusion, transiently bound at loading sites, or stalled by collisions or other barriers, such as the transcriptional machinery.

Cohesin enrichment at NIPBL/MAU2 sites was only ~2-fold over background (Fig. 4C; Extended Data Fig. 6C-D), and most NIPBL/MAU2 peaks did not coincide with RAD21 peaks (Extended Data Fig. 1E). Since our simulations indicated that isolated loading sites are not expected to retain cohesin, we next asked whether NIPBL/MAU2 binding at enhancers nonetheless increases cohesin levels in the surrounding chromatin.

Proximity to a NIPBL/MAU2 binding site only mildly boosts cohesin levels

Guided by our simulations, we quantified RAD21 ChIP-seq enrichment at CTCF peaks, reasoning that the cohesin level, normalized to CTCF occupancy, reflects the amount of extruding cohesin blocked at each site. While CTCF and RAD21 signals were correlated, the correlation was lower than between CTCF replicates or between RAD21 replicates (Fig. 5A-B, Extended Data Fig. 6E). Accordingly, the RAD21/CTCF ratio varied 15-fold between sites (Fig. 5C), highlighting that CTCF sites with comparable occupancy block markedly different cohesin amounts across the genome (Fig. 5D). This variability is not simply explained by CTCF motif architecture, as cohesin levels were independent of the presence of upstream or downstream auxiliary motifs beyond the core CTCF motif[52] (Extended Data Fig. 6F).

Proximity to NIPBL/MAU2 sites or enhancers did not appreciably influence the RAD21/CTCF ratio: CTCF peaks within 1 kb of a NIPBL/MAU2 site only showed 1.1-fold higher cohesin enrichment compared to sites over 100 kb away from the nearest NIPBL/MAU2 binding site (Fig. 5E). We reached the same conclusion by analyzing ChIP-seq from neural progenitor cells (NPCs, Extended Data Fig. 7), indicating that NIPBL/MAU2 binding sites only have a minimal influence on the amount of cohesin that extrudes to neighboring CTCF sites.

As an orthogonal measure, we inspected whether NIPBL/MAU2 binding sites create chromatin folding signatures expected from targeted cohesin loading, such as Hi-C fountains or stripes[25–27,29,30] (Extended Data Fig. 5B). Individual strong NIPBL/MAU2 sites did not display such patterns (Extended Data Fig. 8A), although a weak fountain-like signal is present in aggregate heatmaps[25] (Extended Data Fig. 8B). This observation further supports that cohesin loading at NIPBL/MAU2 sites does not have a major role in shaping chromosome folding patterns in ESCs.

Altogether, our analyses indicate that the cohesin that loads at enhancers is not a major contributor to the flux of loop extrusion. Because other studies reported that enhancer insertions

can increase cohesin levels at nearby CTCF sites[8,11], we considered alternative mechanisms by which enhancers could increase cohesin flux. One possibility is that, instead of acting as direct loading sites, enhancers could increase cohesin recruitment over their broader genomic neighborhood, for example by shifting chromatin domains into an accessible state more permissive to loading. Consistent with this idea, simulations showed that increasing diffuse loading over a 50 kb domain by only 2-fold enriches cohesin at flanking CTCF sites as effectively as increasing targeted loading at a single site 500-fold (Fig. 3D and Extended Data Fig. 5D). Thus, our genomic analyses are consistent with a model where enhancer activation triggers modest, domain-wide modulation of cohesin loading that elevates extrusion flux, without necessarily acting as the direct initiation sites for extrusion.

Active Transcription and clustered CTCF sites impede cohesin traffic

Together, our experimental observations and simulations demonstrate that targeted loading can only explain a small portion of the 15-fold variation in cohesin enrichment measured at CTCF sites (Fig. 5). We therefore investigated to what extent this heterogeneity in cohesin traffic arises from the location of extrusion barriers, rather than cohesin loading.

Intragenic CTCF sites retained the least cohesin when most highly transcribed (Extended Data Fig. 9B), with the lowest cohesin retained at CTCF motifs oriented in the direction of transcription (Extended Data Fig. 9A). This pattern is consistent with transcribing polymerases as mobile barriers that push cohesin during elongation[9,12,31,53–55], particularly when cohesin can translocate in the direction of transcription without encountering an opposing CTCF.

Cohesin accumulation at CTCF sites was also distinctly influenced by the density and orientation of CTCF binding. CTCF sites within 1 kb of another site bound ~1.7-fold less cohesin than isolated sites (>1Mb apart), in both ESCs (Fig. 6A) and NPCs (Extended Data Fig. 7D). Analysis of tandemly oriented CTCF pairs revealed orientation and spacing-dependent effects, with the most pronounced reduction in cohesin accumulation at sites within 30 kb downstream of a co-oriented site (Fig. 6B and 6C, Extended Data Fig. 7E).

To further explore the mechanistic basis of these genomic correlations, we turned to simulations. We modeled CTCF sites as dynamic barriers where the average occupancy of a site is determined by the duration of CTCF binding (bound time) and the interval between binding events (unbound time)[44]. In simulations of a pair of tandemly oriented CTCF sites, cohesin at the downstream site was reduced up to three-fold when the barriers were immediately adjacent, and increased sharply with increasing spacing (Fig. 6D). High occupancy sites (with longer bound time) blocked most effectively. These results support that barrier interference contributes to the variation of cohesin levels at CTCF sites across the genome.

To explore the interference between CTCF barriers within complex genomic layouts, we simulated a chromosome with many randomly positioned CTCF sites with a range of CTCF occupancies. The resulting cohesin/CTCF ratios showed wide dispersion, with closely spaced CTCF sites retaining the lowest amount of cohesin, recapitulating the patterns seen in ChIP-seq data (Fig. 6E, Extended Data Fig. 10 and 11). Varying either CTCF bound time or unbound time resulted in a wide range of cohesin/CTCF ratios (Extended Data Fig. 10A-B). By contrast, regularly spaced CTCF arrays produced a narrower range of cohesin/CTCF ratios, even when using CTCF sites of variable occupancies (Fig. 6E, Extended Data Fig. 10). These simulations highlight how configurations with clustered barriers can substantially suppress cohesin traffic beyond the effect of any individual site.

Finally, we asked whether such sequence-encoded interference is captured by predictive machine learning models that are mechanism-agnostic. Using Enformer (Avsec et al., 2021), we found that introducing CTCF sequences with high cohesin binding into a recipient locus with flanking CTCF sites lowered predicted cohesin levels, mirroring experimental patterns (Extended Data Fig. 12). Removing the flanking CTCF motifs abolished this effect, illustrating that even sequence-based models implicitly capture interference between neighboring CTCF sites.

Altogether, these findings reveal that genome-wide cohesin positioning is shaped more by interference between extrusion barriers – arising from transcription and clustered CTCF sites – rather than by targeted loading, providing a simple explanation for the wide dispersion of cohesin levels at CTCF sites across the genome.

Discussion

Together, our findings argue against a model in which targeted cohesin loading at enhancers increases their functional interactions with distal promoters to boost transcription. Indeed, we observed that, at a locus that relies on loop extrusion for enhancer-promoter communication, increasing cohesin loading at the enhancer represses transcription from the target promoter 170 kb away. Recent studies have also reported that targeted cohesin loading causes transcriptional repression of neighboring promoters[40], not activation –even at physiological loading rates from endogenous enhancers[30].

Our analyses highlight that, as enhancers only constitute a small fraction of the genome, their contribution to overall cohesin on chromosomes is overshadowed by baseline loading across the rest of the genome, even with high loading rate at enhancers (Extended Data Fig. 13). Yet, our results do not rule out that, in specialized genomic or cellular contexts, focal loading at some enhancers could modulate interactions with neighboring promoters in a functionally relevant way. Indeed, small changes in enhancer-promoter contact frequency can meaningfully affect transcription – especially when the baseline contact frequency is low[56]. Focal loading at the enhancer could increase interactions with a promoter if the cohesin remained anchored at the enhancer due to an additional factor such as CTCF binding[8,10,57] or, possibly, RNA polymerase. However, most ESC enhancers do not display chromosome interaction patterns (fountains) that would be indicative of elevated interactions. Our observations indicate that long-range transcriptional regulation largely operates without relying on cohesin loading and loop initiation at enhancers, even at cohesin-dependent loci.

Several studies have demonstrated that enhancers can influence the local cohesin flux and genome folding, raising the question of how such effects can be reconciled with their minimal contribution to overall cohesin loading. Our observations are consistent with a model where enhancers modulate cohesin levels indirectly, by increasing loading across an entire domain rather than acting as discrete initiation sites for extrusion. Consistent with this view, our simulations show that modest increases in domain-wide loading produce substantially larger effects on cohesin accumulation at CTCF sites than even stronger focal loading. These simulations of a broad, but not focal, boost in cohesin loading can quantitatively account for the observation that inserting an enhancer in a silent domain leads to a three-fold increase in cohesin at a CTCF site over 100 kb away[8]. Furthermore, studies in different cell types have found that

339 inserting an enhancer can create a new self-interacting domain and increase cohesin binding at
340 sites across the domain[8,58]. Additionally, regions with enhancers or other markers of active
341 chromatin display higher domain-wide interactions[26,59], and regain domain-wide self-interactions
342 faster after transient cohesin depletion[60]. Studies in yeast demonstrated that chromatin
343 remodelers can recruit NIPBL/MAU2[61], suggesting that more accessible chromatin domains may
344 experience more pervasive cohesin loading.

345 The idea of mammalian enhancers facilitating domain-wide rather than focal cohesin
346 loading remains compatible with observed Hi-C fountains. Fountains in differentiated
347 mammalian cells[26] emerge from active regions hundreds of kb in width that are surrounded by
348 heterochromatin. Notably, such fountains are not evident at ESC enhancers in baseline
349 conditions[25] (Extended Data Fig. 8), perhaps due to their generally open chromatin[62], which
350 could result in more uniform cohesin loading. The fountains that appear in ESCs after depletion
351 of CTCF and the cohesin unloader WAPL are similarly located at multi-kilobase active domains
352 flanked by heterochromatin[29]. Altogether these considerations are consistent with active
353 enhancers inducing recruitment of cohesin across the width of the active domain, rather than
354 acting as the direct initiation sites of extrusion.

355 Our analyses illustrate how even modest increases in pervasive, domain-wide cohesin
356 loading can substantially elevate the local flux of loop extrusion. Because cohesin extrusion is
357 effectively bidirectional at the timescale of minutes[63], enhancer-promoter bridging requires
358 cohesin to load midway between the two elements[64]. While targeted loading at enhancers does
359 not support this configuration, a broad loading regime will allow a subset of cohesin molecules
360 to load at the midway position and transiently bridge enhancers to promoters (Fig. 1A).
361 Therefore, by increasing domain-wide cohesin loading, enhancers could increase chances of
362 transient but functionally productive encounters with distal promoters, without the need to
363 directly initiate cohesin loading at levels that risk interfering with transcription.

364 In this context, the functional relevance of NIPBL enrichment at regulatory elements
365 remains to be clarified. NIPBL binding at enhancers has been consistently reported across studies
366 in various cell types[15,65–67], while NIPBL binding at promoters was only reported by
367 some[15,17,37,53,65,66] but not all[31] studies. Our observation that NIPBL binding is enriched at
368 enhancers over promoters in ESCs is explained because most pluripotency transcription factor
369 binding sites are distal to promoters[68], while in other cell types, NIPBL binds with transcription
370 factors that are found at promoters[37]. NIPBL recruitment likely occurs through direct interactions
371 with cell-type specific transcription factors[69] or via intermediate co-factors present across cell
372 types, such as Mediator[15,70], BRD4[71], or the transcription complex itself[72]. Whether NIPBL has
373 cohesin-independent functions at regulatory elements and whether NIPBL/MAU2 ChIP-seq
374 peaks depend on cohesin loop extrusion requires further clarification, as studies have reported
375 conflicting results[31,40].

376 Regarding the presence of cohesin at enhancers, many studies, especially in differentiated
377 cells, reported low but detectable cohesin enrichment at enhancers[5,6,10,15–17,65,67,73–75]. Given our
378 results, extrusion blocking emerges as a more plausible explanation than targeted loading for
379 cohesin enrichment at such lineage-specific transcription factor binding sites[76]. Supporting a
380 blocking-based mechanism, deletion of transcription factor binding sites within an embryonic
381 stem cell enhancer increases cohesin traffic toward neighboring CTCF sites(Platania et al.,
382 2023), and polymer simulations better reproduce chromosome folding pattern by implementing
383 blocking rather than simply loading at enhancers alone(Zhang et al., 2023; Zhou et al., 2025).

384 Several non-exclusive mechanisms can be envisioned for how enhancers may block
385 cohesin (independently of CTCF). First, the transcription machinery itself can act as a barrier to
386 cohesin extrusion[9,12,31,55]. Second, extrusion may halt upon direct or indirect interaction between
387 cohesin and factors or structures present at regulatory elements. For example blocking at TSSs
388 could be mediated by the chromatin reader protein Phf2, which translocates with cohesin and
389 may arrest extrusion upon binding H3K4me3[77]. Finally, blocking may arise from indirect
390 physical constraints. For example, transcriptional condensates have been reported to exert forces
391 on DNA *in vitro* of up to 7pN[78], far exceeding the force sufficient to stall extrusion (~0.1-0.4 pN,
392 [79]), providing a potential explanation for the low-level and diffuse cohesin enrichment detected
393 by ChIP-seq at enhancers, without the need to invoke direct TF-cohesin interactions[18]. The
394 mechanisms of cohesin blocking at enhancers and the relevance in long-range regulation remain
395 to be investigated.

396 Our quantitative analyses of cohesin ChIP-seq reveal that most molecules (>75%) are
397 found outside CTCF sites. This result is consistent with orthogonal imaging measurements
398 indicating that TADs are rarely fully extruded even at steady-state[80–83]. By extracting
399 interpretable signatures of extrusion dynamics from snapshot cohesin ChIP-seq, our work
400 explains why cohesin occupancy is a better predictor of 3D genome folding patterns than CTCF
401 occupancy alone[84]: the impact of a CTCF site on cohesin traffic depends on the amount of the
402 extrusion flux it can intercept, which itself depends on the density, orientation, and binding
403 kinetics of nearby CTCF sites. These considerations are directly relevant to the interpretation of
404 CTCF site disruptions, with mutations of isolated CTCF sites producing larger chromosome
405 folding defects[42,85–87], and developmental loci where tuning cohesin trajectories implements
406 cellular decisions[88,89].

407 Finally, our work does not exclude functional roles for targeted cohesin loading,
408 especially in genomic processes beyond enhancer regulation. For example, targeted increases in
409 cohesin loading have been proposed to guide DNA double-strand break repair by non-
410 homologous end joining[90], a process accompanied by local transcriptional suppression[91]. We
411 speculate that DNA break-induced targeted cohesin loading may contribute to the local
412 suppression of transcription. In *S. cerevisiae,* targeted loading of condensin, another loop
413 extruder, guides recombination events underlying mating type switching[92].

414 Taken together, our work reframes the relationship between transcriptional regulatory
415 elements and cohesin biology, illuminating how genomic context shapes the local flux of loop
416 extrusion within which enhancers operate.

Methods

Plasmid Construction

Plasmids were assembled using Gibson assembly (NEB HiFi DNA Assembly Master Mix E2621) or restriction-ligation (NEB Quick Ligase M2200).

Cell culture

Parental wild-type E14Tg2a (karyotype 19, XY, 129/Ola isogenic background) *Mus musculus* embryonic stem cells and subclones were cultured in 2iSL medium [DMEM + GlutaMAX + sodium pyruvate (ThermoFisher 10569044) supplemented with 15% fetal bovine serum (Gibco SKU A5256701), 550 µM 2-mercaptoethanol (ThermoFisher 21985-023), 1X nonessential amino acids (ThermoFisher 11140-050), 10^4 U/mL of Leukemia Inhibitory Factor (Millipore ESG1107), 1 µM PD0325901 (Apex Bio/Fischer A3013-25), and 3 µM CHIR99021 (Apex Bio/Fischer A3011-100)]. Cells were maintained at a density of 0.2-1.5 x 10^5 cells/cm^2 by passaging using TrypLE (ThermoFisher 12605010) every 1-2 days on 0.1% gelatin-coated dishes (Sigma-Aldrich G1890-100G in 1XPBS) at 37°C and 5% CO_2. The medium was changed daily when cells were not passaged. Cells were checked for mycoplasma infection every 6 months and tested negative.

To establish neural progenitors (NPCs), 0.2 million WT E14 or NIPBL-FKBP ESCs were seeded in a gelatinized T25 flask in 2iSL medium. The following day, cells were rinsed twice in 1X PBS and switched to N2B27 medium [50% DMEM/F12 medium (Gibco 31330-038), 50% Neurobasal medium (Gibco 21103-049), 1X GlutaMAX (Gibco 35050061), 0.5X B27 (Gibco 17504-044), 0.5X N2 (Millipore SCM012), and 0.1 mM 2-mercaptoethanol], and media was changed daily. After 7 days, cells were detached using TrypLE, seeded in nongelatinized bacterial dishes for suspension culture at 3 million cells per 75 cm^2, and cultured in N2B27 containing 10 ng/mL EGF and FGF (Peprotech 315-09 and 100-18B). After 3 days, floating aggregates were seeded on gelatinized dishes. After 2-4 days, poorly attached aggregates were dislodged and discarded, and remaining adherent NPC cells were dissociated using Accutase (Innovative Cell Technologies AT104), passaged twice on gelatinized dishes in N2B27 + EGF/FGF and cryopreserved after expansion in 10% DMSO in N2B27 + EGF/FGF. Cells were thawed and cultured on gelatinized dishes in N2B27 + EGF/FGF and passaged using Accutase for 2-3 minutes at room temperature.

$FKBP^{12F36V}$ depletion was triggered by adding PROTAC dTAG-13 (dTAG) (Sigma-Aldrich SML2601) to a final concentration of 500 nM. AID depletion was triggered by adding indole-3-acetic acid sodium salt (IAA, auxin analog) (Tocris Bioscience 7932) to a final concentration of 500 mM.

HEK cells were grown in DMEM with 10% FBS and split when 70% confluent using TrypLE. The medium was changed every 2-3 days.

U2OS cells were grown in McCoys 5A media (Thermo Fisher 16600082) supplemented with 10% FBS at 37°C and 5% CO2. Cells were passaged every 2-3 days when 70-80% confluent using TrypLE.

ESC Genome engineering

For transfection, plasmids were prepared using the Nucleobond Midi kit (Macherey Nagel 740410.5). Transfections were done using the Neon system (Thermo Fisher) with a 100 µL tip and 1 million cells at 1400 V, 10 ms, 3 pulses. To create knock-in cells, ESCs were co-

463 transfected with 5 µg of vector expressing Cas9 and sgRNA and 20 µg of targeting construct.
464 Homology arms of targeting vectors were ~1 kb. For deletions, 2.5 µg each of two sgRNAs
465 targeting the two ends of the deletion were used.
466 After electroporation, cells were seeded in a 9 cm^2 well and left to recover for 48 hours.
467 Cells were then plated at limiting dilution and grown for 8-11 days, changing medium every
468 other day, until single colonies could be picked. Individual colonies were genotyped by PCR and
469 validated with Sanger sequencing.
470 When the targeting construct contained a resistance cassette, sgRNAs were cloned into
471 pX330, for co-expression of the sgRNA and Cas9. Cells were grown with antibiotic in the
472 medium from the time of serial dilution until the colonies were picked (Puromycin: 1 µg/mL;
473 Hygromycin: 200 µg/mL).
474 When the targeting construct did not contain a resistance cassette, we used sgRNAs
475 cloned into a Cas9-2A-puro vector (pEN243, identical to pX459 Addgene 62988). Cells were
476 treated with puromycin from ~24-72 hours after transfection to kill untransfected cells.
477 For mESC experiments using lipofectamine transfections, 360 ng total of plasmid was
478 mixed with 150 ul OptiMEM, and separately 1.8 ul lipofectamine 2000 was mixed with 150 ul
479 OptiMEM. Both mixtures were incubated in the dark for 5 minutes, then combined and
480 incubated for an additional 60 minutes in the dark at room temperature. 300 ul of the mixture
481 was then added to 0.3 million ESCs seeded in a 9 cm^2 well.
482
483 **HEK cell genome editing**
484 We created a HEK cell line in which the C-terminal of CTCF is tagged with a series of
485 epitopes (HA-FLAG-TY1-eGFP-T2A-PuroR). For genome editing, 1 million cells were
486 transfected with 20 ug of the targeting construct and 5 ug of the Cas9 sgRNA plasmid. We used
487 the Neon system with a 100 µL tip at 1100 V, 20 ms, 2 pulses. After transfection, the cells
488 recovered for four days, with 1 µg/mL puromycin added starting two days after transfection.
489 Four days after transfection, the cells were split, diluted, and dispensed into 96-well plates at a
490 concentration such that there was an average of 0.5 cells per well. After 10 days of growth, single
491 colonies were genotyped by PCR. Of 27 colonies, 10 appeared homozygous for the insertion,
492 and the remainder heterozygous for the insertion. The insertion was verified by sequencing, and
493 GFP expression was confirmed by flow cytometry.
494
495 **Calibrated ChIP-seq**
496 For all NIPBL and MAU2 ChIP experiments, cells were double crosslinked in
497 disuccinimidyl glutarate (DSG) followed by formaldehyde. First, cells were dissociated using
498 TrypLE (ESCs) or Accutase (NPCs) and resuspended in culture medium. 15 million cells were
499 pelleted at 100*g* for 5 minutes at room temperature. Supernatant was discarded and cell pellets
500 were washed once in PBS. Cells were resuspended in 10 mL PBS, and DSG (Fisher Scientific
501 AAH58208MC) dissolved in DMSO was added to a final concentration of 2 mM. Cells were
502 fixed at room temperature for 45 minutes with rotation. Cells were then pelleted at 200*g* for 5
503 minutes and washed three times with PBS (each time pelleting at 200*g* for 5 minutes). Cells were
504 then resuspended in 5 mL of 10% FBS prepared in PBS. To this, 5 mL of freshly prepared 2%
505 formaldehyde solution was added to reach a final formaldehyde concentration of 1%. 2%
506 formaldehyde was diluted from 16% methanol-free formaldehyde (Fisher Scientific PI28908) in
507 1X PBS with 10% FBS. Resuspended cells were fixed in formaldehyde for 10 minutes at room
508 temperature with rotation. To quench, 0.525 mL of 2.5 M glycine (Millipore Sigma GX0205-1)

509 solution in PBS was added (final concentration 0.125 M glycine), and cells were incubated for 5
510 minutes at room temperature followed by 15 minutes on ice. Cells were pelleted by centrifuging
511 at 200*g* for 5 minutes at 4°C. Cell pellets were resuspended in 1 mL of ice-cold 0.125M glycine
512 in PBS, transferred to 1.5 ml tubes, and centrifuged at 300*g* for 5 minutes at 4° C. Supernatant
513 was discarded and pellets were snap frozen on dry ice and stored at -70°C.
514 For CTCF and RAD21 ChIP, cells were dissociated, pelleted, and fixed in 1%
515 formaldehyde for 10 minutes with the subsequent quenching and washes as above.
516 For ChIP, fixed cells were thawed on ice, resuspended in 1 mL of ice-cold lysis buffer [10
517 mM Tris-HCl pH 7.5, 10 mM NaCl, 0.2% NP-40 (IGEPAL), 0.2% TritonX-100, 1 mM EDTA,
518 0.5 mM EGTA, supplemented with fresh 1x protease inhibitor (Thermo Scientific PI78437)] and
519 incubated on ice for 10 minutes. ~9.75 x 10^6 mouse nuclei were combined with ~0.25 x 10^6 fixed
520 HEK nuclei as spike-in. For ChIP experiments with HA antibody, HEK cells with an HA epitope
521 (cell line EA46.3) were used for the spike-in. Nuclei were pelleted at 400*g* for 5 minutes at 4°C,
522 the supernatant was discarded, and nuclear pellets were resuspended in 1 mL of shearing buffer
523 (0.5% SDS, 10 mM Tris-HCl pH 7.5, supplemented with fresh protease inhibitor). Chromatin
524 was sheared using a Covaris with settings PIP 105, duty factor 5, cycles per burst 200, total time
525 23 minutes for double-crosslinked cells or 15 minutes for formaldehyde-crosslinked cells. After
526 shearing, samples were spun at 13,000 rpm for 5 minutes at 4° C. 2.5% of the sample was
527 removed to use as input, and the remainder of the supernatant was used for ChIP. The ChIP
528 reaction was assembled as follows: 900 ul chromatin, 442 µL 5X ChIP incubation buffer (5%
529 TritonX-100, 0.75 M NaCl, 5 mM EDTA, 2.5 mM EGTA, 50 mM Tris-HCl pH7.5), 40 uL
530 Dynabeads (pre-rinsed in ChIP incubation buffer), 10 ug antibody, 22.5 uL protease inhibitor,
531 and 835.5 uL water. Tubes were incubated overnight at 4°C on a rocker.
532 The following antibodies were used for ChIP: HA (Thermo Fischer 26183), RAD21
533 (Abcam ab992), RAD21 only for experiments in RAD21-AID cell lines (abcam ab154769),
534 CTCF (Active Motif 61311), NIPBL (Bethyl A301-779A), H3K27ac (abcam ab4729).
535 Following overnight incubation, beads were collected using a magnetic rack and washed
536 sequentially with different chilled wash buffers for 5 minutes each on a rocker at 4°C. Beads
537 were washed twice with 500 µL of wash buffer 1 (0.1% SDS, 0.1% Sodium deoxycholate, 1%
538 TritonX-100, 0.15 M NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris pH 8), once with 500 µL
539 of wash buffer 2 (0.1% SDS, 0.1% Sodium deoxycholate, 1% TritonX-100, 0.5 M NaCl, 1 mM
540 EDTA, 10 mM Tris pH 8), once with 500ul of wash buffer 3 [0.25 M LiCl, 0.5% Sodium
541 deoxycholate, 0.5% NP-40 (IGEPAL), 1 mM EDTA, 0.5 mM EGTA, 10 mM Tris pH 8], and
542 twice with 500 µL of wash buffer 4 (1 mM EDTA, 0.5 mM EGTA, 10 mM Tris pH 8).
543 To elute the DNA, beads were resuspended in 200 µL of elution buffer (1% SDS, 0.1M
544 $NaHCO_3$) and incubated on a rocker for 20 minutes at room temperature. Beads were then
545 separated using a magnetic rack, and the eluate was transferred to a fresh tube while beads were
546 discarded. Input samples were thawed and elution buffer was added to bring the volume to 200
547 µL. 2 µL of 1 mg/ml RNaseA (Thermo Scientific FEREN0531) was added before incubation at
548 37°C for 30 minutes. To reverse crosslinks, 8 µL of 5 M NaCL and 5 µL of 10 mg/ml Proteinase
549 K were added, and samples were incubated overnight at 65°C. DNA was eluted using a MinElute
550 kit (Qiagen 28604).
551 Libraries were prepared using NEBNext Ultra II DNA library prep kit according to the
552 manufacturer's protocol (E7645S). The entire ChIP sample and 50 ng of the input samples were
553 used for library prep. Libraries were amplified for 16 cycles, followed by 0.9X SPRI bead
554 purification (Bulldog Bio CNGS005). The eluate was run on a 2% E-gel (Invitrogen G401002)

555 and fragments between 200-600 bp were excised and eluted using a MinElute gel extraction kit
556 (Qiagen 28604). Quality and quantity of libraries were assessed using Qubit and Bioanalyzer
557 High Sensitivity DNA kit (Agilent 5067-4626). Libraries were sequenced on a NextSeq 2000
558 using 65 bp single-end reads. Some libraries were sequenced in a second lane for additional
559 reads. Each replicate is a population of cells fixed on a separate day.
560
561 **ChIP-seq mapping**
562 Reads from fastq files were aligned to a catenated mm10 and hg38 reference genome
563 using Bowtie2. Reads with a mapq score ≥ 30 were retained and duplicates were removed, using
564 samtools version 1.21 ref[93]. ChIP bigwigs were generated with the bamCoverage function from
565 deepTools[94], with scaling factors calculated using the formula (human reads in input) / (mouse
566 reads in input) / (human reads in ChIP) x 15,000,000 ref[95]. The multiplier of 15,000,000 was
567 chosen arbitrarily to set the final scaling factors close to 1. Input bigwigs were normalized using
568 RPKM. Bigwigs were visualized using the UCSC genome browser[96].
569 To create bigwig files with multiple replicates combined, we took equal numbers of reads
570 from each replicate (samtools view with subsample option), combined them, and generated a
571 bigwig normalized by RPKM.
572
573 **Creating consensus peak lists**
574 To create lists of ChIP-seq peaks found in multiple replicates, we first used macs2
575 version 2.2.9.1[97] to call peaks on each individual ChIP replicate. For NIPBL and MAU2 peaks,
576 we used HA ChIP in untagged cells as the control. For CTCF and H3K27ac peaks, we used the
577 corresponding input sample as the control. We used bedtools multiinter to identify regions that
578 were called as peaks in the majority of replicates. To define the peak widths, we again called
579 peaks using macs2, this time including all of the replicates together. Using bedtools intersect -u,
580 we selected the peaks from the list called using combined replicates that overlap with the regions
581 called in the majority of replicates. Finally, we removed peaks in ENCODE blacklisted regions[98]
582 using bedtools intersect -v.
583 All analyses of NIPBL/MAU2 peaks use the consensus list of HA-MAU2 peaks, as the
584 HA-MAU2 ChIP was more sensitive compared to the HA-NIPBL ChIP.
585
586 **Comparing ChIP-seq binding**
587 We generated tornado plots and plots of average binding at ChIP-seq peaks using
588 computeMatrix and plotHeatmap functions from deepTools. OCT4 ChIP-seq is from[50], file name
589 GSM3992929_33_WaplC6-0h_antiOct4.bw, and H3K4me3 ChIP-seq is from ENCODE mouse
590 ES-E14 dataset ENCSR000ADL, file ENCFF001MXU, both mapped to mm10 as above. We
591 used the H3K4m3 peak list ENCFF012AZO. We generated the upset plot (Extended Data Fig.
592 1D) using Intervene[99].
593
594 **TetO array imaging**
595 We recruited transcription factors to a TetO array in U2OS 2-6-3 cells[39]. We avoided
596 using the LacO repeats that are also present in the array as we observed that LacR expression
597 alone led to cohesin recruitment, possibly because of chromosomal instability triggered by the
598 tighter LacO-LacR binding[100]. In eight-well ibidi slides (Fisher 50305795), we seeded 20,000
599 cells per well and transfected with a plasmid expressing either TetR-VP64 (pEM037), TetR-

OCT4 (pEM040), TetR (pEM041), or TetR-MAU2 (pAA014) using lipofectamine 3000 (Thermo Fisher L3000008) according to the manufacturer's instructions.

For immunofluorescence, we rinsed cells once with 250 uL of PBS and then fixed in 3% formaldehyde (Electron Microscopy Sciences 15686) diluted in PBS for 10 minutes at room temperature. Cells were rinsed three times with PBS and permeabilized with fresh 0.5% Triton X-100 in PBS for 5 minutes. After rinsing three times with PBS, cells were incubated with 1% BSA in PBS for 10 minutes. Next, cells were incubated in a 150 uL volume of primary antibody diluted 1:300 in 1% BSA in PBS for 45 minutes (antibodies: NIPBL Bethyl A301-779A, MAU2 abcam ab183033, RAD21 Millipore 05-908). Cells were washed five times with 250 uL PBS for 5 minutes each. Next, cells were incubated with secondary antibody diluted 1:1000 in 1% BSA in PBS for 45 minutes (goat anti-rabbit Invitrogen A-11012, goat anti-mouse Invitrogen A-11005) and again washed three times. Finally, cells were incubated with 1 ug/mL DAPI in PBS for 5 minutes and rinsed twice in PBS. All steps were performed at room temperature.

Images were acquired using a Zeiss spinning disk confocal microscope with 60x objective or Nikon spinning disk confocal microscope with a 60x objective with XY pixel size of 183 x 183 nm and 300 nm steps in Z. Images were analyzed in ImageJ with the JACoP plugin to calculate the Pearson correlation between red and green channels within a 12 × 12 × 8 X × Y × Z box manually placed on each GFP-positive TetO array. To calculate the intensity of signal at the array compared to background, for each cell we extracted the total signal within a box centered on the array of size 6x6 XY pixels and 4 z slices. We then chose a box outside the array with a representative level of cohesin of size 12x12 XY pixels and the same 4 z slices as the first box. To account for the different box sizes, we computed enrichment at the array as the (array intensity * 4) / (background intensity).

TACL

We heterozygously inserted a 34x TetO array at different locations in a cell line in which the two alleles of *Car2* are tagged with tdTomato and mClover reporters. To facilitate distinguishing between the two alleles by ORCA, we first created a 10 kb deletion of a neutral sequence near *Car2* enhancer 2 (chr3:15062000-15072751).

We then inserted the 34x TetO array at three locations in the *Car2* locus. To ascertain which allele the TetO array was inserted on, we used a Cas9-based strategy. For the TetO insertion 4 kb upstream of the *Car2* TSS (EA88), we performed a lipofectamine 2000 transfection of plasmids encoding Cas9 and sgRNAs to cut in the TetO repeats and downstream of the *Car2* open reading frame. For the TetO insertion 107 kb downstream of the *Car2* TSS (EA89), we used sgRNAs targeting TetO and downstream of the Car2 enhancers. To select transfected cells, we added puromycin and hygromycin from 1 day to 3 days after transfection. Five days after transfection, we performed flow cytometry and observed a population of cells with loss of either red or green expression, indicating which allele experienced large deletions and therefore harbored the TetO array. For the TetO array 172 kb downstream of the *Car2* TSS, we used a PCR strategy relying on the existing 10 kb deletion downstream of the *Car2* enhancers on the red allele.

For TACL experiments, cells were transfected with plasmids expressing either TetR-Halo-MAU2 (pAA014) or TetR-Halo (pEA117) using lipofectamine 2000 along with piggyBac transposase plasmid (PB200). Starting two days after transfection, cells were grown in 1 µg/mL puromycin to select for cells with successful integrations. Seven days after transfection, *Car2* reporter expression was measured using flow cytometry. For each location of the TetO

integration, two clones were each transfected three separate times. Flow cytometry measurements were normalized by dividing by the average reporter expression in untransfected cells of the corresponding genotype.

ORCA experiment

Cell lines and fixation

ORCA experiments were performed on clonal cell lines. For each TetO insertion position, we subcloned the pool of cells from one TetR-Halo and one TetR-Halo-MAU2 transfection. We measured Halo expression in individual clones by flow cytometry, and then for each condition selected a single clone with a similar expression level.

For fixation, we dissociated cells using TrypLE, fixed in 4% formaldehyde for 10 minutes at room temperature, washed four times in PBS (by centrifuging for five minutes at 400g), resuspended in 70% ethanol, and stored at -20°C.

Multiplexed cell labeling

Cells were labeled by adapting a previously published protocol[49,101] with minor modifications. Briefly, 3'-amine-modified oligonucleotides (Integrated DNA technologies) were conjugated to Methyltetrazine-NHS ester. The reaction was performed in sodium borate buffer (50 mM, pH 8.5) with DMSO, and the resulting methyltetrazine-oligos were purified by ethanol precipitation. The final products were resuspended in HEPES buffer.

For cell labeling, approximately 2.5-3 million fixed cells were transferred to protein LoBind tubes and 3X washed using 1% BSA in PBS and resuspended in 1X PBS. The cells were first incubated with 1 mM TCO-PEG4-TFP Ester for 5 minutes at room temperature protected from light. Subsequently, the prepared methyltetrazine-oligos were added and incubated for 30 minutes at room temperature. The reaction was quenched using Methyltetrazine-PEG4-Amine and Tris-HCl for 5 minutes. Labeled cells were washed 2X with 1% BSA in PBS and once with 1X PBS only. Cells were resuspended in 70% ethanol in 1X PBS and stored at -20 °C.

Cell plating and surface barcode normalization

Barcoded cells were pooled at equal ratios by mixing 2-4 µl of each sample in a single tube. A 10-15 µl aliquot of the pooled cell suspension was then dispensed onto the poly-D-lysine coated central region of a microscope slide and left to air dry for ~ 5 minutes before proceeding with primary probe hybridization (see below). Multiple slides were prepared in parallel (4-6), and only those with an optimal cell density forming a uniform monolayer were selected for imaging and backup acquisitions. To assess and even out the representation of all conditions prior to the main imaging experiment, barcoded cells were first imaged using readouts targeting the surface barcodes without DNA primary probe hybridization. The relative abundance of each condition was quantified from these barcode signals, and cell numbers were subsequently normalized accordingly.

Probe design

The *Car2* locus (chr3:14,753,359-15,369,435) was tiled with 72 barcodes spanning 8 kb each, with every barcode targeted by ~120 individual probes. Each probe was 100 bp in length and compromised of 4 segments: 20 bp of a reverse and forward primer sequences used during probe synthesis, 20 bp complementary to the fiducial probes, 20 bp of barcode sequence, and 40 bp of unique genomic targeting sequence. Probes were designed using previously described

692 software[102], which is accessible at https://github.com/BoettigerLab/ORCA-public. The probe
693 library was synthesized as an oligonucleotide pool by GenScript.
694
695 *DNA hybridization*
696 Hybridization with DNA primary probes was carried out following[103]. Briefly, following
697 pooling and plating, cells were fixed in 4% paraformaldehyde (PFA) for 10 minutes and washed
698 3X with 1X PBS, which was gently added to the slide by pouring along the side of the 6 cm Petri
699 dish to avoid direct flow over the cells. Cells were then premetallized in 20-40 µl of 0.5% Triton
700 X-100 in 1X PBS for 10 minutes and rinsed 3X with 1X PBS. Permeabilized cells were then
701 incubated in 20-40 µl 0.1 M HCl for 5 minutes and washed 3X with 1X PBS. To remove RNA,
702 cells were treated with 20-40 µl RNase A (10 µg ml^{-1}) for 30 minutes at 37 °C and then washed
703 3X with 2X SSC. Next, cells were equilibrated for 35 minutes in hybridization buffer (2X SSC,
704 50% formamide, 0.1% Tween-20). The volume of solution applied varied depending on the area
705 occupied by cells on each slide; solutions were added in sufficient volume to ensure complete
706 coverage of all cells.
707 Hybridization mixture was prepared by mixing 2 µl of 840 ng $µl^{-1}$ of DNA primary
708 probes and 30 µl hybridization solution (2X SSC, 50% formamide, 10% dextran sulfate, and
709 0.1% Tween-20). The mixture was then applied to the cells, covered with a coverslip treated with
710 Sigmacote, denatured for 3 minutes at 90 °C and then incubated overnight in a humidified
711 chamber at 42 °C. The following day, samples were 3X washed with 2X SSC and post-fix in a
712 mixture of 2% glutaraldehyde (GA) and 8% PFA in 1X PBS for 30 minutes. After post-fixation,
713 slides were 3X washed with 2X SSC and mounted on the microscope for sequential labeling and
714 imaging, starting with surface barcode imaging.
715
716 *ORCA Imaging setup*
717 Imaging was performed using a custom-built ORCA microscope setup with the complete
718 configuration parameters available via MicroMeta App[104]. Briefly, the system was constructed
719 around a Nikon Ti2 body with a large 32 mm aperture optics enabling the acquisition of wide-
720 field images (2,048 x 2,048-pixel area with a final pixel size of 108 nm). Illumination was
721 provided by solid-state laser arrays by Lumencor (CELESTA light engine) paired with a custom
722 polychroic mirror (Chroma, ZET408/473/545/635/750rpc-uf2) and emission filter (Chroma,
723 ZET408/473/545/635/750m). The system was equipped with a Marzhauser Wetzlar xy-stage (00-
724 24-625-000) and a Mad City Labs Inc. Z objective positioner (Nano-F200S) for z-axis control.
725 The setup was completed with a Hamamatsu ORCA-Flash4.0 V3 sCMOS camera, an IR-laser-
726 based autofocus system, and custom fluidics set up[103].
727
728 *ORCA image acquisition*
729 Primary probe hybridized samples were imaged by detecting readout sequences with dye-
730 labeled secondary oligonucleotides: AlexaFluor 750 for DNA readouts, Cy5 for surface barcodes
731 and Cy3 for fiducial markers, with the Cy3 channel used for accurate spatial registration and
732 sample drift correction. Imaging was started with surface barcodes and followed by DNA
733 tracing. Fiducial signals were detected during the first DNA imaging round together with the first
734 DNA readout and were subsequently imaged in every following cycle.
735 For each readout barcode, ~ 400 µl of probe solution (25% ethylene carbonate in 2× SSC
736 containing the readout oligonucleotide and displacement oligos complementary to the previous
737 readout) was flowed into the chamber and incubated for 30 minutes to allow hybridization and

strand displacement. All remaining wash, buffer exchange, and acquisition steps were performed as in the published protocol[103].

ORCA analysis

Raw imaging processing was performed following the image processing pipeline detailed in ref[103] with published scripts available at: https://github.com/BoettigerLab/ORCA-public. This pipeline yields 3D coordinates for every barcode within a given trace, which were subsequently transformed into pairwise distance matrices for structural analysis.

Computational correction for underperforming readouts

To ensure the integrity of population-level data, we computationally corrected all median distance and contact frequency maps for artifacts introduced by specific readout probes. For most of these readouts, the poor performance stems from flawed intrinsic properties, a recurring issue documented in prior ORCA experiments. In this study, a specific list of seven readouts (5, 10, 16, 21, 23, 48, and 65) were identified as unreliable based on visual inspection of the raw imaging data, which revealed issues such as weak signal or spurious fluorescence. The data corresponding to these specific rounds was masked and the resulting gaps in the data matrices were filled by interpolating values from neighboring, high-quality hybridization points.

Readout 38 was masked to account for the 10 kb deletion in the TetO cell lines. While the readout itself performed well, this step prevents the deletion from creating a strong predicted artifact in the differential contact and distance maps for TetO vs. non-TetO conditions.

Demultiplexing cell surface barcodes

Demultiplexing was performed using a custom MATLAB script that integrates Cellpose for nuclei segmentation[105]. For each field of view (FOV), a maximum intensity projection image was generated from the drift-corrected barcode imaging channels to facilitate robust segmentation. Cellpose was then run with its pre-trained 'nuclei' model to generate segmentation masks. For each segmented nucleus, a barcode intensity vector was created by calculating the median pixel intensity within the mask for each barcode channel. These vectors were normalized to account for variations in probe brightness. Finally, each nucleus was assigned an identity by matching its normalized barcode vector to a predefined codebook using a k-nearest neighbor approach, with an associated confidence score calculated to assess assignment ambiguity. The resulting table linked each identified spot from the primary experiment to its parent nucleus and decoded identity.

Allele disambiguation

To distinguish between control allele (non-TetO allele) and allele engineered for forced cohesin loading (TetO allele) via a 10 kb deletion, a custom Python script was used to classify chromosomal traces in each demultiplexed cell based on the presence or absence of imaging readout 38, which specifically hybridizes to the deleted region in the *Car2* control allele. To ensure high confidence in our classification, this comparison was performed exclusively between the two alleles within the same nucleus, which serves as an internal control to mitigate potential misalignment errors or technical false negatives arising from variable hybridization efficiency.

Quantification of enhancer-promoter interactions

Enhancer-promoter interactions were assessed between windows of loci centered around the *Car2* enhancers and promoter. Specifically, we used a +/- 1 bin window for each (P-window, E-window) to account for the physical span of the regulatory elements and to mitigate potential signal dropout from suboptimal probe detection efficiency during imaging.

To avoid using an arbitrary distance cutoff, a data-driven contact threshold was determined as the median of all pairwise distances between adjacent probes measured across the entire traced *Car2* domain in the control cell population, approximately 200 nm across conditions. An E-P interaction was then classified "in contact" for a given cell if the minimum distance between the E and P windows was less than or equal to this threshold.

To capture the full spectrum of interaction distances without binarization, we analyzed the distribution of minimum E-P distances. For each individual cell, we identified the single minimum distance between any locus in the E-window and any locus in the P-window. This metric represents the closest physical approach between the two regulatory regions within that cell.

Simulations

1D lattice model

We performed one-dimensional lattice simulations to model loop extrusion dynamics. Loop extrusion parameters were chosen to reflect mESC dynamics, and the various loading and barrier configurations below were used to probe different determinants of cohesin accumulation. We modeled 1D dynamics with fixed-timestep simulations on a lattice where each lattice site represented 250 base pairs of genomic sequence. All lattice sites shared a uniform unloading (or dissociation) rate for extruders; this is specified via the processivity (330 kb, equivalent to 1320 lattice sites or 660 lattice sites for each arm), the average distance extruded before dissociation. Each site could either represent a generic genomic location with baseline extruder loading probability or be designated as a boosted loading site with increased extruder loading frequency. For standard sites, cohesin loading occurred with uniform probability across the lattice. Average spacing between extruders was determined by a separation parameter, set to 185 kb (740 lattice sites), following experimental estimates[106] and previous simulations[23] in mESC.

Any extruder that dissociated from the lattice was immediately reloaded at a random un-occupied lattice site, where the likelihood of reloading at a given position was proportional to its local loading rate. Each extruder occupied two adjacent lattice sites, corresponding to its left and right legs. At each discrete time step, extruder legs not stalled by a barrier could move outward, (i.e. the left leg moves left and the right leg moves right). For example, an extruder can update from position (i, j) to (i–1, j+1). Extruder legs could not bypass one another upon collision and could not exit the first or last lattice sites, effectively confining all movement within the lattice boundaries. Positions of extruders were recorded after discarding the initial 20% of simulation steps to compute observables.

Dynamic CTCF barriers were implemented as unidirectional obstacles that selectively stalled extruder legs based on their direction of motion (as previously[44]). Dynamic barrier occupancy was modeled as a two-state process, governed by defined binding and unbinding times. Extruders encountering an unbound barrier site continued extrusion uninterrupted. An extruder leg moving in the direction opposite to the orientation of an occupied barrier became stalled upon contact. While one leg was stalled at a bound barrier, the other leg was allowed to continue moving independently.

Boosted loading site layouts

Boosted loading was reported as a fold-increase per kilobase. E.g., a loading boost of 100 on a 250 bp site is equivalent to a fold-increase of 25 per 1 kb. In simulations involving a single focal loading site, boosted loading was applied at a single 250 bp lattice site. For simulations representing larger loading regions or "active TADs" of uniformly boosted loading, multiple adjacent boosted loading sites were used. In all scenarios, boosted loading added to the total number of extruders in the simulated region. One boosted loading region or site was considered per 1,000 lattice sites. We averaged over 30,000,000 simulation snapshots for these layouts.

Quantifying cohesin at the loading site

To quantify the relative enrichment of extruders at boosted target loading sites (Fig. 3B and Extended Data Fig. 5C), we counted extruder reads within a four-site window (corresponding to 1 kb) centered on the center of each boosted target loading region. This local read count was divided by the average number of extruder reads measured in two distal control regions located 20 kb to the left and 20 kb to the right of the target region, using windows of the same size. This value was then normalized to the corresponding value obtained for the uniform loading scenario. The resulting ratio quantifies local extruder enrichment at the boosted loading site while controlling for local background read density.

Boosted loading between inward-oriented CTCF layout

To clarify the impact of targeted loading on accumulation at nearby CTCFs, we considered simplified scenarios with isolated CTCF barrier sites oriented inwards symmetrically at a defined distance from a loading site (Fig. 3D, Extended Data Fig. 5). Such arrangements were placed at the center of a 1,000 site region and tiled ten times to form a 10,000-site lattice. For these simplified scenarios, we set the CTCF bound time to be approximately one quarter of the cohesin residence time, reflecting previous estimates[44]. Unless noted, isolated barriers had an occupancy of 0.5, meaning they were CTCF-bound 50% of the time and unbound the other 50%. We calculated an enrichment ratio at CTCF sites as the accumulation of extruder legs at barrier sites in a targeted loading scenario normalized by the baseline from uniform loading. We computed in silico ChIP-seq profiles by recording the positions of extruder legs from 30,000,000 instances of 1D lattice simulations for this layout.

Tandem CTCF layout

To precisely quantify the impact of nearby CTCF sites arranged in tandem (Fig. 6D), we performed a set of simulations with otherwise isolated tandemly-oriented pairs of CTCF sites. A set of tandem sites were placed at the center of a 1,000 site region and tiled ten times to form a 10,000 site lattice. We quantified how the front CTCF can block accumulation at the back CTCF by varying the occupancy of the front CTCF while keeping the back CTCF at 0.5 occupancy. We then calculated RAD21 reads normalized by the number of times each barrier appeared at its genomic position (RAD21/CTCF). We computed in silico ChIP-seq profiles by sampling extruder leg positions from 100,000,000 1D lattice simulation instances.

Variable spacing and occupancy of CTCF barriers layouts

To develop simulations with more complex patterns of cohesin accumulation closer to those in cells, we designed simulations with randomly distributed barriers with variable bound

875 times and occupancies as follows (Fig. 6E, Extended Data Fig. 10) for a lattice with 300,000
876 sites. Barrier occupancies were independently sampled with replacement from empirical
877 measurements of CTCF binding via single-molecule footprinting[107], and the total number of
878 barriers (multiplied by their occupancy) was chosen to match experimental estimates.
879 Specifically, we sampled barriers until their combined occupancy reached 133 in 10 Mb, based
880 on estimates for the number of bound CTCF molecules in a region of this size (~217,000 total
881 CTCF molecules in a mESC nucleus, with 50% bound (based on FRAP measurements[108]), and
882 an average estimated genomic content per cell ~3 × 2.7 Gb[106]). Lattice positions and orientations
883 were randomly assigned for each sampled barrier. While per site occupancies are available
884 experimentally, per site bound and unbound times are not. Based on previous estimates[44] we thus
885 set these dynamic parameters by scaling sampled bound times such that strong sampled barriers
886 (0.65 occupancy) had a residence time (656 sec) of roughly one third the extruder lifetime, thus
887 fixing their unbound time (353 sec). Average sampled barriers (~.35 occupancy) had a
888 correspondingly lower residence time (233 sec). For varying bound time simulations, we
889 sampled site occupancies, and determined corresponding site bound times using the fixed
890 unbound time (353 sec). We then introduced variability in bound times at fixed sampled
891 occupancies by: drawing an updated bound time from a Gaussian distribution with mean equal to
892 the sampled bound time and standard deviation of ⅓ of the sampled bound time and rescaling the
893 unbound time to maintain the sampled occupancy. For the varying unbound time simulations, a
894 similar rescaling procedure was used, albeit determining site unbound times from the sampled
895 occupancy and a fixed bound time (233 sec). For constant occupancy simulations, all barriers
896 had the same occupancy, and thus the same bound (233 sec) and unbound times (353 sec). For
897 each scenario, we ran 100 independent simulations using multiprocessing, with approximately
898 153,000 recorded instances per simulation. Statistics were computed by aggregating data across
899 all runs, yielding a total of approximately 15,300,000 simulated instances.
900 To evaluate the effect of variable separation between barriers on extruder accumulation,
901 we conducted simulations using regular barrier arrangements. First, we maintained the same
902 barrier density (per 10 Mb) and positioned barriers at constant intervals of ~ 35 kb with
903 alternating orientations (right, left, right, etc.). To assess the impact of barrier spacing on the
904 dispersion of cohesin accumulation, we also increased the distance between barriers six-fold.
905 To quantify cohesin enrichment, we calculated the accumulation of extruders at barrier
906 sites. To compare simulations with experimental data, we scaled both RAD21 and CTCF read
907 counts to match the read densities measured by ChIP-seq. Specifically, we calculated the total
908 number of simulated extruder and CTCF reads per base pair within the 75 Mb simulated region,
909 and applied a normalization factor such that the simulated read density equaled the experimental
910 read density (0.016764 reads per basepair, per haploid genome). This normalization ensured that
911 simulated and experimental datasets had comparable read coverage per genomic length.
912 To further quantify the role of barrier spacing, we analyzed simulated cohesin positioning
913 at barriers similarly to experimental data. To investigate the role of nearby CTCF sites in pairs
914 oriented tandemly), we further calculated RAD21 reads normalized by how many times each
915 barrier appeared at its genomic position (RAD21/CTCF). This normalization allowed us to
916 measure RAD21 accumulation specific to each individual barrier. Specifically, we identified
917 pairs of tandemly oriented barriers, and compared the relative enrichment of extruders between
918 them by computing the ratio of extruder accumulation in the first versus the second barrier,
919 stratified by their spacing. Then we computed the cumulative distribution functions over their
920 spacing separately for right-oriented and left-oriented barriers (Extended Data Fig. 11B). Finally,

921 we assessed how extruder accumulation (for all orientations) scales with barrier proximity by
922 comparing extruders versus barriers read counts across random configurations. We binned
923 barriers by their nearest-neighbor distance and calculated the average extruder/barrier ratio in
924 each bin (Extended Data Fig. 11A).
925
926 *3D polymer simulations*
927 We performed 3D polymer simulations for the targeted loading scenarios by modeling a
928 2.5 Mb segment of chromatin as a 50 nm polymer fiber, with each monomer corresponding to
929 2.5 kb of genomic DNA or 10 sites in the corresponding 1D lattice model. To implement the 3D
930 polymer simulations, we used the Polychrom framework (https://github.com/open2c/polychrom),
931 which leverages OpenMM[109] to simulate large ensembles of chromatin conformations. To
932 simulate loop extrusion, we performed 200 3D steps between each 1D lattice update. Based on
933 previous timescale matching[44], each simulated time step in 3D represents $\delta t_{3d} = 0.015$ sec to give
934 a mean extruder residence time on chromatin of approximately 33 minutes. Loop extrusion was
935 modeled by introducing harmonic bonds between the two monomers representing the arms of
936 each extruder (positions *I* and *j*). As the loop extrusion factor progressed, these bonds were
937 dynamically updated; removing the bond between *(I, j)* and adding a new bond between *(i–1,*
938 *j+1)* to reflect loop enlargement. Under wild-type conditions, the polymer exhibited Rouse
939 dynamics.
940
941 *In silico Hi-C analysis*
942 For in silico Hi-C analysis, we generated 3,000,000 polymer conformations from our 3D
943 simulations for every parameter set we considered. Using these conformations, we constructed
944 Hi-C contact maps at a resolution corresponding to the size of one monomer (2.5 kb), applying
945 the default *polychrom* (https://github.com/open2c/polychrom) capture radius of 2.3 monomers to
946 define contact events. This threshold closely matches empirical contact distance estimates[110].
947 Simulated contact matrices were converted into .cool format using *polykit*
948 (https://github.com/open2c/polykit), facilitating analysis within the standard Hi-C data
949 processing framework[111,112].
950
951 *Code*
952 Code specifying simulations and for performing analysis are available at:
953 https://github.com/Fudenberg-Research-Group/Targeted_Cohesin_Loading
954
955 **Determining the fraction of cohesin reads at CTCF sites**
956 To measure the fraction of cohesin reads at CTCF sites, we first performed calibrated
957 ChIP-seq with RAD21 (abcam ab154769) and SMC1 (Bethyl A300-055A) antibodies in
958 untagged (E14) cells, untreated RAD21-eGFP-AID (EN272.2) cells, and EN272.2 cells in which
959 RAD21 was degraded by adding 500 mM of Indole-3-Acetic Acid sodium salt (IAA, auxin
960 analog, Tocris Bioscience 7932) to the culture media for 3.5 hours. The position of the degron
961 tag precluded use of the Abcam ab992 RAD21 antibody used in other experiments. For each of
962 two replicates, the number of cohesin reads in CTCF peaks was computed as in
963 https://github.com/Fudenberg-Research-Group/fastaFRiP using deeptools.countReadsPerBin and
964 then multiplied by the scaling factor. The background-corrected total number of reads was
965 calculated by (total mouse reads * scaling factor) – (total reads after RAD21 depletion * scaling

966 factor). Finally, the corrected fraction of reads in CTCF peaks was calculated by (cohesin reads
967 in CTCF peaks) / (background-corrected total reads).
968
969 **Western blotting**
970 We used Western blotting to quantify RAD21 and MAU2 depletion. Approximately 10
971 million ESCs were dissociated using TrypLE, resuspended in culture medium, and pelleted at
972 100g for 5 minutes at room temperature. Cell pellets were washed in PBS, pelleted again, snap
973 frozen on dry ice after aspirating the PBS, and stored at −70°C. Cell pellets were resuspended in
974 10 mM HEPES (pH 7.9), 2.5 mM MgCl2, 0.25 M sucrose, 0.1% NP40, 1 mM DTT, and 1X
975 HALT protease inhibitors (ThermoFisher 78437) and kept on ice for 10 minutes to allow
976 swelling. After centrifugation at 500g at 4°C for 5 minutes, the supernatant was discarded and
977 the pellets were resuspended on ice in 25 mM HEPES (pH 7.9), 1.5 mM $MgCl_2$, 700 mM NaCl,
978 0.5 mM DTT, 0.1 mM EDTA, 20% glycerol, 1 mM DTT, and 250 U benzonase (Sigma-Aldrich
979 E1014), and incubated on ice for 10 minutes. The nuclear lysates were centrifuged at 18,000g at
980 4°C for 10 minutes, to eliminate the cellular debris. Protein concentration from the supernatants
981 was measured using Pierce BCA Kit (ThermoFisher 23225).
982 40 ug of protein extract was combined with 1x Laemmli buffer (Biorad 1610747), heated
983 at 95°C for 10 minutes, and subjected to electrophoresis on a 4-15% polyacrylamide gel (Biorad
984 17000927). The separated proteins were transferred onto a polyvinylidene difluoride (PVDF)
985 membrane (Millipore IPFL00010) using phosphate-based transfer buffer (10 mM sodium
986 phosphate monobasic, 10 mM sodium phosphate dibasic) at 4°C, 600 mA, for 1.5 hours. After
987 the transfer, the membrane was cut so the lower molecular weight section could be blotted with
988 TBP antibody for the loading control. Membranes were blocked in 5% blocking buffer (5% skim
989 milk in TBS-T (20 mM Tris-Cl buffer-pH 7.4, 500 mM NaCl and 0.1% Tween 20)) and
990 incubated overnight at 4°C with the primary antibodies (RAD21 Millipore 05-908 1:1000,
991 RAD21 abcam ab154769 1:1000, MAU2 abcam ab183033 1:1000, or TBP Cell Signaling
992 Technology 8515S 1:2000) prepared in 5% blocking buffer. The next day, the membranes were
993 washed three times with TBS-T buffer and incubated with mouse or rabbit secondary antibodies
994 conjugated with horseradish peroxidase diluted 1:2000 (Biotium 20400, Biotium 20403) for 1 hr
995 at room temperature. The membranes were again washed three times with TBS-T buffer. The
996 blots were developed using Pierce ECL western blotting substrate (ThermoScientific 32209) and
997 detected using Chemidoc XRS+ imager (Biorad).
998 Three biological replicates of untagged (E14) and RAD21-GFP-AID cells with and
999 without auxin treatment were collected on separate days. Intensity of the bands was quantified
1000 using ImageJ. For Extended Data Fig. 6A, RAD21 intensities were normalized to TBP and then
1001 to the value in untagged cells.
1002
1003 **Flow cytometry**
1004 ESCs were dissociated with TrypLE, resuspended in culture medium, pelleted, and
1005 resuspended in 10% FBS in PBS before live cell flow cytometry on an Attune NxT instrument
1006 (Thermo Fisher). Analyses were performed using the FlowJo software. Briefly, live cells were
1007 gated using FSC-A x SSC-A and singlets were then gated using FSC-H x FSC-A before
1008 extracting median values for approximately 50,000 final gated cells. The value of
1009 autofluorescence in untagged E14 cells was subtracted.
1010 For quantification of RAD21-GFP, values were normalized by dividing by the average
1011 expression in untreated cells, and replicates were obtained by collecting cells on different days.

1012
1013 **Analysis of cohesin/CTCF ratios**
1014 We performed four replicates of calibrated ChIP-seq with CTCF (Active motif catalog
1015 number 61311) and RAD21 (Abcam ab992) antibodies in E14 ESCs. In NPCs, we performed
1016 two replicates each of CTCF and RAD21 ChIP-seq. For each protein, we combined equal
1017 numbers of reads from each of the ChIP-seq replicates. Approximately 46 million total reads
1018 were included for CTCF and RAD21 in ESC. For NPC analysis, the number of reads was
1019 normalized to match the number in ESCs. We used bedtools window to calculate the number of
1020 reads in each CTCF peak.
1021 To calculate correlations in the number of reads in CTCF sites between ChIP-seq
1022 datasets, we used multiBigwigSummary and plotCorrelation from deeptools with the Pearson
1023 correlation method.
1024 For comparing cohesin/CTCF ratios to the distance to the nearest enhancer, we used the
1025 enhancer list from[38]. For transcription start sites, we used transcript coordinates from GENCODE
1026 VM23 and included protein-coding genes with at least 1 transcript per million. Gene expression
1027 levels are from ENCODE RNA-seq in E14 mouse ESCs (identifier ENCSR000CWC)[113]. A and
1028 B compartments were called using the compartment_score function in GENOVA[114] with Hi-C
1029 data from two combined replicates of wild-type mESCs[23].
1030 For the analysis of CTCF motif variants (Extended Data Fig. 6F), we used core, forward,
1031 and reverse motifs from[52]. We used FIMO from the MEME suite[115] to identify instances of the
1032 motifs ($p < 0.0001$) in the CTCF peaks. We plotted the RAD21/CTCF ratios for peaks in the
1033 following categories: only core motifs; at least one core and at least one upstream motif but no
1034 downstream motif; at least one core and at least one downstream motif but no upstream motif; no
1035 core motifs.
1036 For analysis the of the distance over which CTCF sites can block cohesin traffic, we
1037 counted the number of forward and reverse core CTCF motifs[52] in each CTCF peak. We selected
1038 pairs of peaks that each contained only forward motifs and for which the next downstream peak
1039 was over 100 kb away, and similarly selected pairs of peaks with reverse motifs. For each pair,
1040 we calculated $\log_2$[(first peak RAD21/CTCF)/(second peak RAD21/CTCF)] and plotted the
1041 cumulative distribution of this metric based on the distance between the peaks.
1042
1043 **Micro-C analysis**
1044 We analyzed micro-C from[116], using data with two pooled replicates from untreated
1045 mESCs (GSE178982_CTCF-UT_pool.mcool) or cells with 3 hours of CTCF depletion
1046 (GSE178982_CTCF-AID_pool.mcool). Plots show data normalized by the expected interaction
1047 frequencies. We ranked MAU2 peaks based on the number of ChIP-seq reads, and generated a
1048 pileup at the top one third of sites using cooltools.pileup[112]. Control sites were selected using
1049 bedtools shuffle[117] to choose random sites within TADs that contain a top MAU2 site, with TAD
1050 definitions from[116]. Repeated permutations of random sites yielded consistent pile-up plots.
1051
1052 **Enformer**
1053 As Enformer is not trained on relevant mouse ESC datasets, we generated predictions for
1054 cohesin and CTCF binding in mouse CH12 cells. We downloaded CH12 CTCF and RAD21
1055 ChIP-seq data from the ENCODE portal (identifiers ENCSR000ERM and ENCSR000ERK),
1056 mapped the reads as above, and called peaks using macs2. We selected a CTCF peak with a low
1057 RAD21/CTCF ratio (chr15:97665011-97665743) to use as a recipient locus, and selected ten

CTCF peaks with higher CTCF/cohesin ratios to use as donor sequences. We then generated test sequences by swapping 512 bp (centered at the center of the CTCF peak) from the donor sequence into the recipient sequence. We used Enformer[118] to predict CH12 RAD21 and CTCF binding at the test sequences. We plotted the sum of the cohesin signal across the 512 bp divided by the sum of the CTCF signal across the 512 bp. We then modified the recipient sequence by replacing the CTCF motifs in the CTCF peaks flanking the central peak with neutral sequence (replacing 17 bp at each of four motifs), and repeated the swapping of the donor sequences into the modified recipient sequence.

Acknowledgements
We are grateful to members of Daniele Canzio's group for discussion and access to equipment, and to Alex Buckley, Simon Gaudin, Fernanda Vargas-Romero, Elzo de Wit, and Luca Giorgetti for critical feedback on the manuscript. We thank David Spector and Stanley Qi for the U2OS 2-6-3 cell line. Funding sources: Grant NIH R35GM142792 to EN, Grant NIH R35GM143116 to GF, the Chan-Zuckerberg Biohub San Francisco Investigator program to EN, the Sandler Program for Breakthrough Biomedical Research at UCSF, NSF EAGER grant 2313792 to EN, NIH training grant T32HL007731 (ECA), fellowships from the Helen Hay Whitney foundation (ECA), the UCSF-California Institute for Regenerative Medicine Scholars Training Program education grant EDUC4-12812 (ECA and KH), and KAUST Ibn Rushd Fellowship (AA). Part of the sequencing was performed at the UCSF CAT, supported by UCSF PBBR, RRP IMIA, and NIH 1S10OD028511-01 grants. Flow cytometry was performed in the Laboratory for Cell Analysis in the UCSF Helen Diller Family Comprehensive Cancer Center, supported by grant P30CA082103.

Author contributions
ECA performed, analyzed, or supervised all experiments except ORCA, with technical support from EMA, ASA, RS, KH, and EPN. HR performed simulations under the supervision of GF. AA designed, performed and analyzed ORCA with support from IC and under the supervision of ANB. EP designed the study with ECA and input from GF. ECA assembled figures and wrote the manuscript with assistance from EPN, GF, and input from all co-authors.

Ethics declaration
No competing interests to declare.

Data availability
Sequencing data was deposited to GEO GSE278338.

Code availability
https://github.com/hrahmanin/Target_cohesin_loading

References

1. Papantonis, A. & Oudelaar, A. M. Mechanisms of Enhancer-Mediated Gene Activation in the Context of the 3D Genome. *Annual Review of Genomics and Human Genetics* **26**, 163–188 (2025).
2. Yang, J. H. & Hansen, A. S. Enhancer selectivity in space and time: from enhancer–promoter interactions to promoter activation. *Nat Rev Mol Cell Biol* **25**, 574–591 (2024).
3. Davidson, I. F. & Peters, J.-M. Genome folding through loop extrusion by SMC complexes. *Nat Rev Mol Cell Biol* **22**, 445–464 (2021).
4. de Wit, E. & Nora, E. P. New insights into genome folding by loop extrusion from inducible degron technologies. *Nat Rev Genet* **24**, 73–85 (2023).
5. Allyn, B. M. *et al.* Promoter- and enhancer-dependent cohesin loading initiates chromosome looping to fold *Tcrb* loci for long-range recombination. *Cell Reports* **44**, 116133 (2025).
6. Hua, P. *et al.* Defining genome architecture at base-pair resolution. *Nature* **595**, 125–129 (2021).
7. Rinaldi, L. *et al.* The glucocorticoid receptor associates with the cohesin loader NIPBL to promote long-range gene regulation. *Science Advances* **8**, eabj8360 (2022).
8. Rinzema, N. J. *et al.* Building regulatory landscapes reveals that an enhancer can recruit cohesin to create contact domains, engage CTCF sites and activate distant genes. *Nat Struct Mol Biol* **29**, 563–574 (2022).
9. Valton, A.-L. *et al.* A cohesin traffic pattern genetically linked to gene regulation. *Nat Struct Mol Biol* **29**, 1239–1251 (2022).
10. Vian, L. *et al.* The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* **173**, 1165-1178.e20 (2018).
11. Vos, E. S. M. *et al.* Interplay between CTCF boundaries and a super enhancer controls cohesin extrusion trajectories and gene expression. *Molecular Cell* **81**, 3082-3095.e6 (2021).
12. Zhang, S., Übelmesser, N., Barbieri, M. & Papantonis, A. Enhancer–promoter contact formation requires RNAPII and antagonizes loop extrusion. *Nat Genet* **55**, 832–840 (2023).
13. Barrington, C. *et al.* Enhancer accessibility and CTCF occupancy underlie asymmetric TAD architecture and cell type specific genome topology. *Nat Commun* **10**, 2908 (2019).
14. Dorsett, D. & Merkenschlager, M. Cohesin at active genes: a unifying theme for cohesin and gene expression from model organisms to humans. *Current Opinion in Cell Biology* **25**, 327–333 (2013).
15. Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
16. Nitzsche, A. *et al.* RAD21 Cooperates with Pluripotency Transcription Factors in the Maintenance of Embryonic Stem Cell Identity. *PLOS ONE* **6**, e19470 (2011).
17. Schmidt, D. *et al.* A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Res.* **20**, 578–588 (2010).
18. Selivanovskiy, A. V., Molodova, M. N., Khrameeva, E. E., Ulianov, S. V. & Razin, S. V. Liquid condensates: a new barrier to loop extrusion? *Cell. Mol. Life Sci.* **82**, 80 (2025).
19. Ciosk, R. *et al.* Cohesin's Binding to Chromosomes Depends on a Separate Complex Consisting of Scc2 and Scc4 Proteins. *Molecular Cell* **5**, 243–254 (2000).
20. Murayama, Y. & Uhlmann, F. Biochemical reconstitution of topological DNA binding by the cohesin ring. *Nature* **505**, 367–371 (2014).
21. Davidson, I. F. *et al.* DNA loop extrusion by human cohesin. *Science* **366**, 1338–1345 (2019).
22. Kim, Y., Shi, Z., Zhang, H., Finkelstein, I. J. & Yu, H. Human cohesin compacts DNA by loop extrusion. *Science* **366**, 1345–1349 (2019).
23. Shah, R. *et al.* Dosage sensitivity of the loop extrusion rate confers tunability to genome folding while creating vulnerability to genetic disruption. 2025.08.14.667581 Preprint at https://doi.org/10.1101/2025.08.14.667581 (2025).
24. Wutz, G. *et al.* PDS5 proteins control genome architecture by limiting the lifetime of cohesin-NIPBL complexes. 2025.08.30.673243 Preprint at https://doi.org/10.1101/2025.08.30.673243 (2025).

1153 25. Galitsyna, A. *et al.* Extrusion fountains are hallmarks of chromosome organization emerging upon
1154 zygotic genome activation. 2023.07.15.549120 Preprint at
1155 https://doi.org/10.1101/2023.07.15.549120 (2023).
1156 26. Guo, Y. *et al.* Chromatin jets define the properties of cohesin-driven in vivo loop extrusion.
1157 *Molecular Cell* https://doi.org/10.1016/j.molcel.2022.09.003 (2022)
1158 doi:10.1016/j.molcel.2022.09.003.
1159 27. Kim, J., Wang, H. & Ercan, S. Cohesin organizes 3D DNA contacts surrounding active enhancers in
1160 C. elegans. *Genome Res* **35**, 1108–1123 (2025).
1161 28. Kraft, K. *et al.* Serial genomic inversions induce tissue-specific architectural stripes, gene
1162 misexpression and congenital malformations. *Nature Cell Biology* **21**, 305–310 (2019).
1163 29. Liu, N. Q. *et al.* Extrusion fountains are restricted by WAPL-dependent cohesin release and CTCF
1164 barriers. *Nucleic Acids Res* **53**, gkaf549 (2025).
1165 30. Lüthi, B. N. *et al.* Cohesin forms fountains at active enhancers in C. elegans. *Nat Commun*
1166 https://doi.org/10.1038/s41467-025-67302-6 (2025) doi:10.1038/s41467-025-67302-6.
1167 31. Banigan, E. J. *et al.* Transcription shapes 3D chromatin organization by interacting with loop
1168 extrusion. *Proceedings of the National Academy of Sciences* **120**, e2210480120 (2023).
1169 32. Chen, L.-F. *et al.* Structural elements promote architectural stripe formation and facilitate ultra-long-
1170 range gene regulation at a human disease locus. *Molecular Cell* **83**, 1446-1461.e6 (2023).
1171 33. Ollikainen, N., Ma, F., Braikia, F. Z. & Sen, R. Chromatin folding principles underlying the
1172 generation of antibody diversity. *Molecular Cell* https://doi.org/10.1016/j.molcel.2025.12.023
1173 (2026) doi:10.1016/j.molcel.2025.12.023.
1174 34. Bahr, C. *et al.* A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell
1175 hierarchies. *Nature* **553**, 515–520 (2018).
1176 35. Kragesteen, B. K. *et al.* Dynamic 3D chromatin architecture contributes to enhancer specificity and
1177 limb morphogenesis. *Nat Genet* **50**, 1463–1473 (2018).
1178 36. Long, H. K. *et al.* Loss of Extreme Long-Range Enhancers in Human Neural Crest Drives a
1179 Craniofacial Disorder. *Cell Stem Cell* **27**, 765-783.e14 (2020).
1180 37. Zuin, J. *et al.* A Cohesin-Independent Role for NIPBL at Promoters Provides Insights in CdLS.
1181 *PLOS Genetics* **10**, e1004153 (2014).
1182 38. Whyte, W. A. *et al.* Master Transcription Factors and Mediator Establish Super-Enhancers at Key
1183 Cell Identity Genes. *Cell* **153**, 307–319 (2013).
1184 39. Janicki, S. M. *et al.* From Silencing to Gene Expression: Real-Time Analysis in Single Cells. *Cell*
1185 **116**, 683–698 (2004).
1186 40. Han, R. *et al.* Characterization of induced cohesin loop extrusion trajectories in living cells. *Nat*
1187 *Genet* **57**, 2785–2797 (2025).
1188 41. Hansen, K. L. *et al.* Synergy between regulatory elements can render cohesin dispensable for distal
1189 enhancer function. *Science* **0**, eadt4221 (2025).
1190 42. Rekaik, H. *et al.* Sequential and directional insulation by conserved CTCF sites underlies the Hox
1191 timer in stembryos. *Nat Genet* **55**, 1164–1175 (2023).
1192 43. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports* **15**,
1193 2038–2049 (2016).
1194 44. Rahmaninejad, H., Xiao, Y., Tortora, M. M. C. & Fudenberg, G. Dynamic barriers modulate cohesin
1195 positioning and genome folding at fixed occupancy. *Genome Res.*
1196 https://doi.org/10.1101/gr.280108.124 (2025) doi:10.1101/gr.280108.124.
1197 45. Parelho, V. *et al.* Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms.
1198 *Cell* **132**, 422–433 (2008).
1199 46. Rubio, E. D. *et al.* CTCF physically links cohesin to chromatin. *Proceedings of the National*
1200 *Academy of Sciences* **105**, 8309–8314 (2008).
1201 47. Wendt, K. S. *et al.* Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature*
1202 **451**, 796–801 (2008).

48. Hu, B. *et al.* Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic Acids Res* **43**, e132–e132 (2015).
49. Hafner, A. *et al.* Loop stacking organizes genome folding from TADs to chromosomes. *Molecular Cell* **83**, 1377-1392.e6 (2023).
50. Liu, N. Q. *et al.* WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nature Genetics* **53**, 100–109 (2021).
51. Szabo, Q. *et al.* Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat Genet* **52**, 1151–1157 (2020).
52. Nakahashi, H. *et al.* A Genome-wide Map of CTCF Multivalency Redefines the CTCF Code. *Cell Reports* **3**, 1678–1689 (2013).
53. Busslinger, G. A. *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503–507 (2017).
54. Chapard, C. *et al.* RNA Pol II-based regulations of chromosome folding. *Cell Genomics* 100970 (2025) doi:10.1016/j.xgen.2025.100970.
55. Heinz, S. *et al.* Transcription Elongation Can Affect Genome 3D Structure. *Cell* **174**, 1522-1536.e22 (2018).
56. Zuin, J. *et al.* Nonlinear control of transcription through enhancer–promoter interactions. *Nature* **604**, 571–577 (2022).
57. Kubo, N. *et al.* Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat Struct Mol Biol* **28**, 152–161 (2021).
58. Georgiades, E. *et al.* Active regulatory elements recruit cohesin to establish cell specific chromatin domains. *Sci Rep* **15**, 11780 (2025).
59. Schmitt, A. D. *et al.* A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Reports* **17**, 2042–2059 (2016).
60. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e24 (2017).
61. Muñoz, S., Minamino, M., Casas-Delucchi, C. S., Patel, H. & Uhlmann, F. A Role for Chromatin Remodeling in Cohesin Loading onto Chromosomes. *Molecular Cell* **74**, 664-673.e5 (2019).
62. Efroni, S. *et al.* Global Transcription in Pluripotent Embryonic Stem Cells. *Cell Stem Cell* **2**, 437–447 (2008).
63. Barth, R. *et al.* SMC motor proteins extrude DNA asymmetrically and can switch directions. *Cell* **188**, 749-763.e21 (2025).
64. Ubertini, M. *et al.* Loop extrusion creates rare, long-lived encounters underlying enhancer-promoter communication. 2025.09.24.678119 Preprint at https://doi.org/10.1101/2025.09.24.678119 (2025).
65. Cuadrado, A. *et al.* Contribution of variant subunits and associated factors to genome-wide distribution and dynamics of cohesin. *Epigenetics & Chromatin* **15**, 37 (2022).
66. van den Berg, D. L. C. *et al.* Nipbl Interacts with Zfp609 and the Integrator Complex to Regulate Cortical Neuron Migration. *Neuron* **93**, 348–361 (2017).
67. Zhu, Y., Denholtz, M., Lu, H. & Murre, C. Calcium signaling instructs NIPBL recruitment at active enhancers and promoters via distinct mechanisms to reconstruct genome compartmentalization. *Genes Dev.* https://doi.org/10.1101/gad.343475.120 (2020) doi:10.1101/gad.343475.120.
68. Chen, X. *et al.* Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell* **133**, 1106–1117 (2008).
69. Fettweis, G. *et al.* Transcription factors form a ternary complex with NIPBL/MAU2 to localize cohesin at enhancers. *Nucleic Acids Res* **53**, gkaf415 (2025).
70. Mattingly, M. *et al.* Mediator recruits the cohesin loader Scc2 to RNA Pol II-transcribed genes and promotes sister chromatid cohesion. *Current Biology* **32**, 2884-2896.e6 (2022).
71. Linares-Saldana, R. *et al.* BRD4 orchestrates genome folding to promote neural crest differentiation. *Nat Genet* **53**, 1480–1492 (2021).
72. Tei, S. *et al.* Cohesin acts as a transcriptional gatekeeper by restraining pause–release to promote processive elongation. 2025.09.30.679672 Preprint at https://doi.org/10.1101/2025.09.30.679672 (2025).

73. Bryan, A. F., Justice, M., Stutzman, A. V., McKay, D. J. & Dowen, J. M. Cohesin stabilization at promoters and enhancers by common transcription factors and chromatin regulators. *Epigenetics & Chromatin* **18**, 33 (2025).
74. Cuadrado, A. *et al.* Specific Contributions of Cohesin-SA1 and Cohesin-SA2 to TADs and Polycomb Domains in Embryonic Stem Cells. *Cell Reports* **27**, 3500-3510.e4 (2019).
75. Grant, Z. L. *et al.* Dose-dependent sensitivity of human 3D chromatin to a heart disease-linked transcription factor. 2025.01.09.632202 Preprint at https://doi.org/10.1101/2025.01.09.632202 (2025).
76. Zhou, Y. *et al.* Lineage-determining transcription factors constrain cohesin to drive multi-enhancer oncogene regulation. *Nat Cell Biol* 1–17 (2025) doi:10.1038/s41556-025-01827-2.
77. Tang, W. *et al.* Cohesin positions the epigenetic reader Phf2 within the genome. *The EMBO Journal* **44**, 736–766 (2025).
78. Nguyen, T. *et al.* Chromatin sequesters pioneer transcription factor Sox2 from exerting force on DNA. *Nat Commun* **13**, 3988 (2022).
79. Davidson, I. F. *et al.* CTCF is a DNA-tension-dependent barrier to cohesin-mediated loop extrusion. *Nature* **616**, 822–827 (2023).
80. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).
81. Gabriele, M. *et al.* Dynamics of CTCF- and cohesin-mediated chromatin looping revealed by live-cell imaging. *Science* https://doi.org/10.1126/science.abn6583 (2022) doi:10.1126/science.abn6583.
82. Mach, P. *et al.* Cohesin and CTCF control the dynamics of chromosome folding. *Nat Genet* **54**, 1907–1918 (2022).
83. Sabaté, T. *et al.* Uniform dynamics of cohesin-mediated loop extrusion in living human cells. *Nat Genet* 1–13 (2025) doi:10.1038/s41588-025-02406-9.
84. Smaruj, P. N., Kamulegeya, F., Kelley, D. R. & Fudenberg, G. Interpreting the CTCF-mediated sequence grammar of genome folding with AkitaV2. *PLOS Computational Biology* **21**, e1012824 (2025).
85. Amândio, A. R. *et al.* Sequential in cis mutagenesis in vivo reveals various functions for CTCF sites at the mouse HoxD cluster. *Genes Dev.* **35**, 1490–1509 (2021).
86. Anania, C. *et al.* In vivo dissection of a clustered-CTCF domain boundary reveals developmental principles of regulatory insulation. *Nat Genet* **54**, 1026–1036 (2022).
87. Despang, A. *et al.* Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat Genet* **51**, 1263–1271 (2019).
88. Kiefer, L. *et al.* Tuning cohesin trajectories enables differential readout of the Pcdhα cluster across neurons. *Science* **385**, eadm9802 (2024).
89. Kiefer, L. *et al.* WAPL functions as a rheostat of Protocadherin isoform diversity that controls neural wiring. *Science* **380**, eadf8440 (2023).
90. Yang, J. H., Brandão, H. B. & Hansen, A. S. DNA double-strand break end synapsis by DNA loop extrusion. *Nat Commun* **14**, 1913 (2023).
91. Meisenberg, C. *et al.* Repression of Transcription at DNA Breaks Requires Cohesin throughout Interphase and Prevents Genome Instability. *Mol Cell* **73**, 212-223.e7 (2019).
92. Piveteau, V. *et al.* Condensin loop extrusion properties, roadblocks, and role in homology search in S. cerevisiae. 2024.09.12.612585 Preprint at https://doi.org/10.1101/2024.09.12.612585 (2026).
93. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
94. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–W165 (2016).
95. Fursova, N. A. *et al.* Synergy between Variant PRC1 Complexes Defines Polycomb-Mediated Gene Repression. *Molecular Cell* **74**, 1020-1036.e8 (2019).
96. Perez, G. *et al.* The UCSC Genome Browser database: 2025 update. *Nucleic Acids Res* **53**, D1243–D1249 (2025).

97. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biology* **9**, R137 (2008).
98. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* **9**, 9354 (2019).
99. Khan, A. & Mathelier, A. Intervene: a tool for intersection and visualization of multiple gene or genomic region sets. *BMC Bioinformatics* **18**, 287 (2017).
100. Dubarry, M., Loïodice, I., Chen, C. L., Thermes, C. & Taddei, A. Tight protein–DNA interactions favor gene silencing. *Genes Dev.* **25**, 1365–1370 (2011).
101. Gehring, J., Hwee Park, J., Chen, S., Thomson, M. & Pachter, L. Highly multiplexed single-cell RNA-seq by DNA oligonucleotide tagging of cellular proteins. *Nat Biotechnol* **38**, 35–38 (2020).
102. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
103. Mateo, L. J., Sinnott-Armstrong, N. & Boettiger, A. N. Tracing DNA paths and RNA profiles in cultured cells and tissues with ORCA. *Nat Protoc* **16**, 1647–1713 (2021).
104. Rigano, A. *et al.* Micro-Meta App: an interactive tool for collecting microscopy metadata based on community specifications. *Nat Methods* **18**, 1489–1495 (2021).
105. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* **18**, 100–106 (2021).
106. Cattoglio, C. *et al.* Determining cellular CTCF and cohesin abundances to constrain 3D genome models. *eLife* **8**, e40164 (2019).
107. Sönmezer, C. *et al.* Molecular Co-occupancy Identifies Transcription Factor Binding Cooperativity *In Vivo*. *Molecular Cell* **81**, 255-267.e6 (2021).
108. Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* **6**, e25776 (2017).
109. Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology* **13**, e1005659 (2017).
110. McCord, R. P., Kaplan, N. & Giorgetti, L. Chromosome Conformation Capture and Beyond: Toward an Integrative View of Chromosome Structure and Function. *Molecular Cell* **77**, 688–708 (2020).
111. Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311–316 (2020).
112. Open2C *et al.* Cooltools: Enabling high-resolution Hi-C analysis in Python. *PLOS Computational Biology* **20**, e1012067 (2024).
113. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
114. van der Weide, R. H. *et al.* Hi-C analyses with GENOVA: a case study with cohesin variants. *NAR Genomics and Bioinformatics* **3**, lqab040 (2021).
115. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
116. Hsieh, T.-H. S. *et al.* Enhancer-promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat Genet* **54**, 1919–1932 (2022).
117. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
118. Avsec, Ž. *et al.* Effective gene expression prediction from sequence by integrating long-range interactions. *Nat Methods* **18**, 1196–1203 (2021).
119. King, H. W. & Klose, R. J. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *eLife* **6**, e22631 (2017).

Figure 1

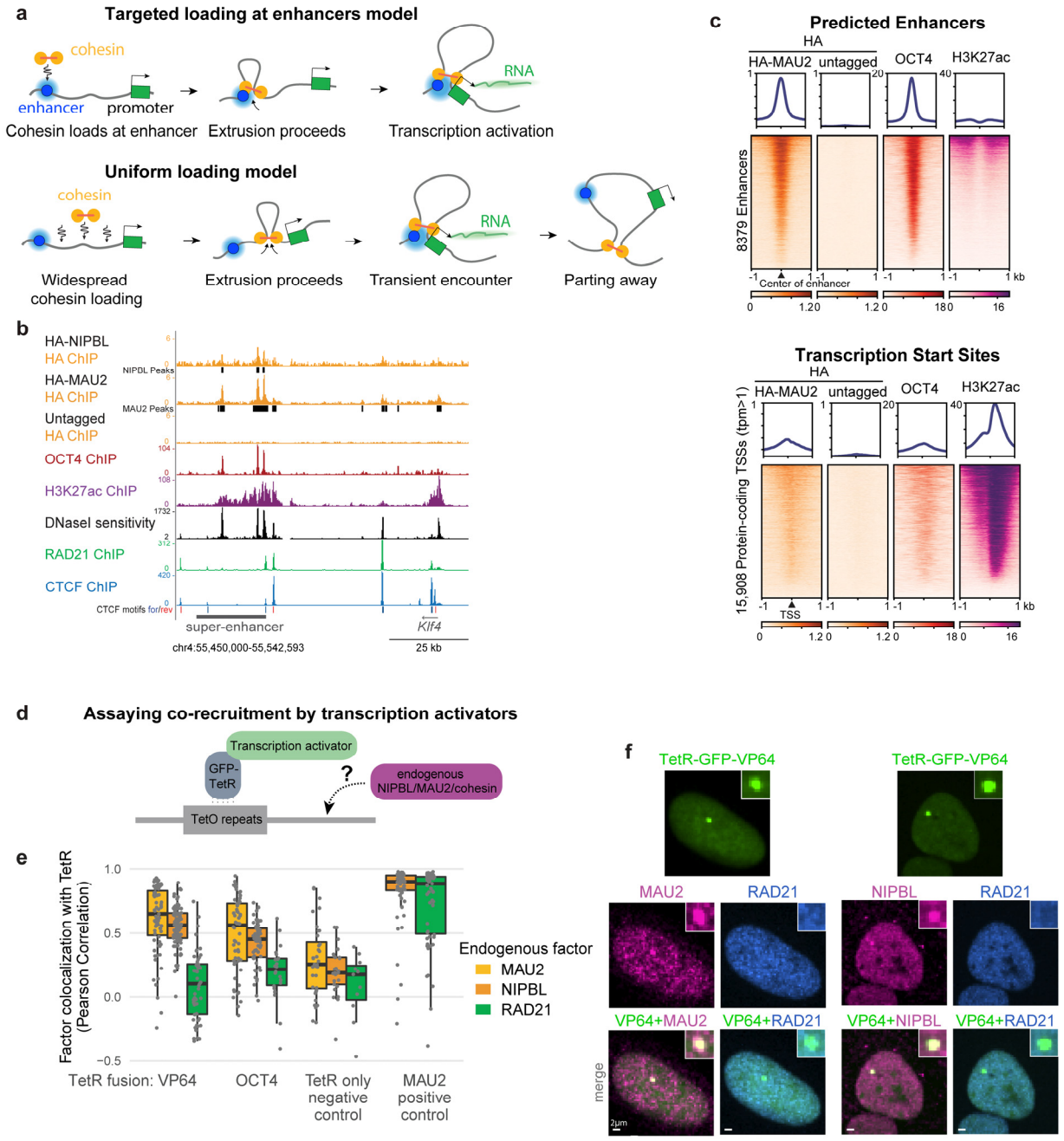


Figure 1: Enhancers bind NIPBL/MAU2 without strongly enriching cohesin

a, In the targeted loading model, cohesin is preferentially loaded at enhancers and extrudes toward target promoters to support long-range transcriptional activation. In contrast, in the uniform loading model most cohesin loads away from enhancers and bridges enhancers to their target promoters only transiently. **b**, ChIP-seq in tagged mouse embryonic stem cells (ESCs) shows high NIPBL/MAU2 binding at a distal enhancer, overlapping OCT4 binding sites, but not at a promoter. OCT4 ChIP-seq from ref[50] and DNAseI sensitivity from ref[113]. **c**, MAU2 ChIP signal is enriched at predicted pluripotency enhancers (from ref[38]) more than that at

1362 transcriptionally active TSSs. **d**, Experimental strategy to test co-recruitment of endogenous
1363 NIPBL, MAU2, and cohesin by transcriptional activators. TetR-GFP-fused transcription
1364 activators were expressed in U2OS cells with a large heterozygous TetO array[39] and endogenous
1365 NIPBL, MAU2, or cohesin were stained by immunofluorescence. **e**, Quantification of
1366 colocalization between TetR-GFP and NIPBL, MAU2, or cohesin at the TetO array. Each point
1367 denotes the Pearson correlation coefficient between the two proteins in a single cell. TetR-GFP
1368 fused to transcriptional activators (VP64 or OCT4) recruits MAU2 and NIPBL but not the
1369 cohesin subunit RAD21. TetR alone serves as a negative control, whereas TetR-MAU2 provides
1370 a positive control for cohesin recruitment. Boxes indicate interquartile range; center lines denote
1371 medians. **f**, Example cells showing that tethering VP64 recruits endogenous MAU2 and NIPBL
1372 but no detectable cohesin RAD21.
1373
1374

Figure 2

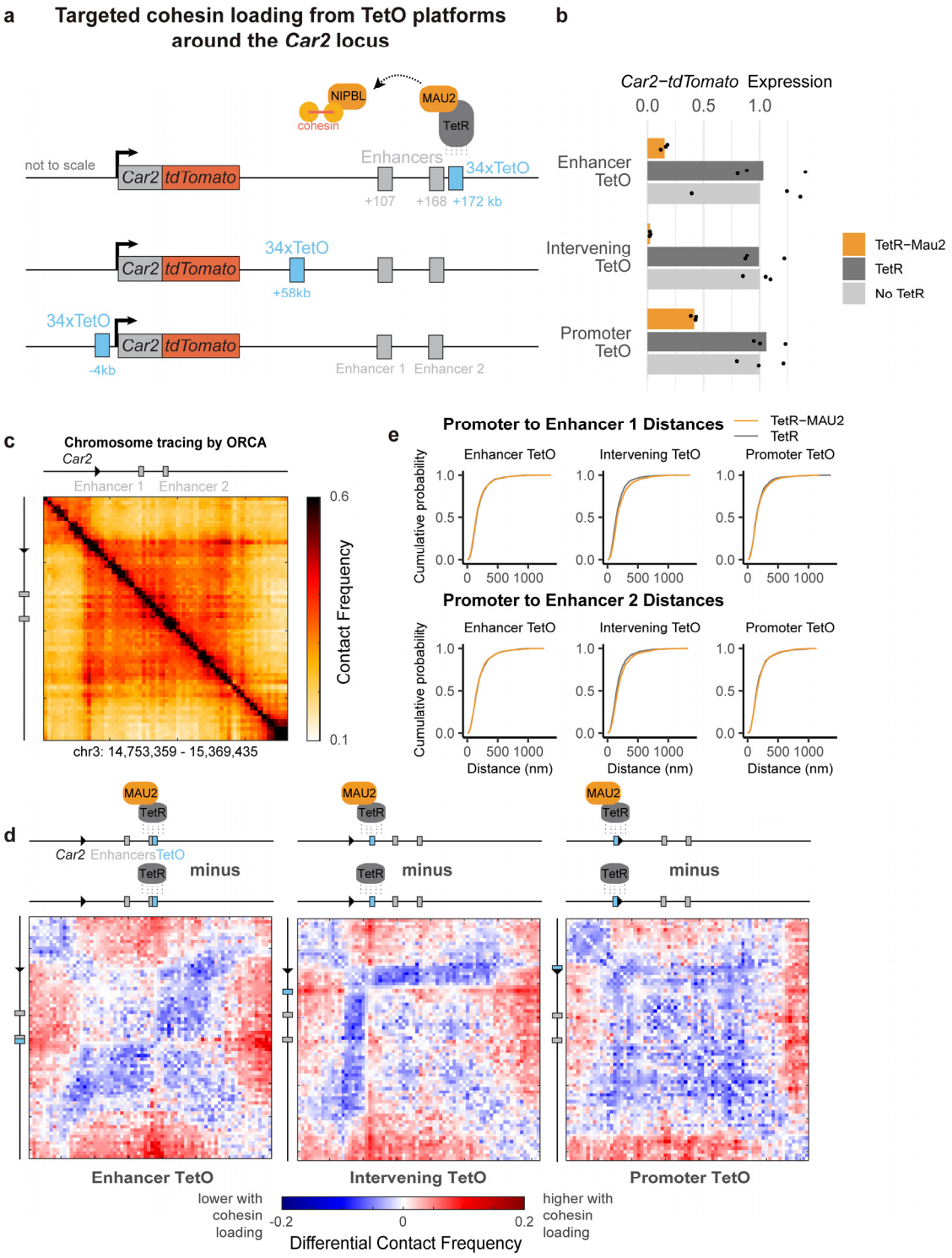


Figure 2: Targeted cohesin recruitment interferes with transcription.
a, TetR-MAU2 was stably expressed to force cohesin loading at 34x TetO arrays (2 kb) inserted at three positions across the *Car2* locus in ESCs. **b**, *Car2* expression, measured by flow cytometry, was reduced by forcing cohesin loading at all tested positions, whereas TetR binding

1380 alone had no detectable effect. Values are normalized to corresponding parental cells. Points
1381 show three biological replicates from separate transfections of the respective TetR-expressing
1382 plasmid. **c**, DNA FISH chromosome tracing (ORCA) heatmap showing contact frequencies
1383 (within ~200 nm, see methods) at the *Car2* locus (8 kb resolution). **d**, Differential contact maps
1384 comparing TetR-MAU2 cells to TetR-only controls (comprising >1300 traces per condition).
1385 Targeted cohesin loading does not increase enhancer-promoter interactions but instead creates a
1386 boundary, (except at the *Car2* promoter, which is located near an endogenous TAD boundary).
1387 The modest reduction in interactions across the TAD arises from global MAU2 overexpression
1388 rather than from targeted loading at the TetO site (see allele-specific comparisons within TetR-
1389 MAU2 cells in Extended Data Fig. 4c). **e**, Cumulative plots showing the distribution of 3D
1390 distances between the *Car2* promoter and each of its enhancers on the TetO allele with either
1391 TetR-MAU2 (forced cohesin loading) or TetR expression (control).

1392

Figure 3

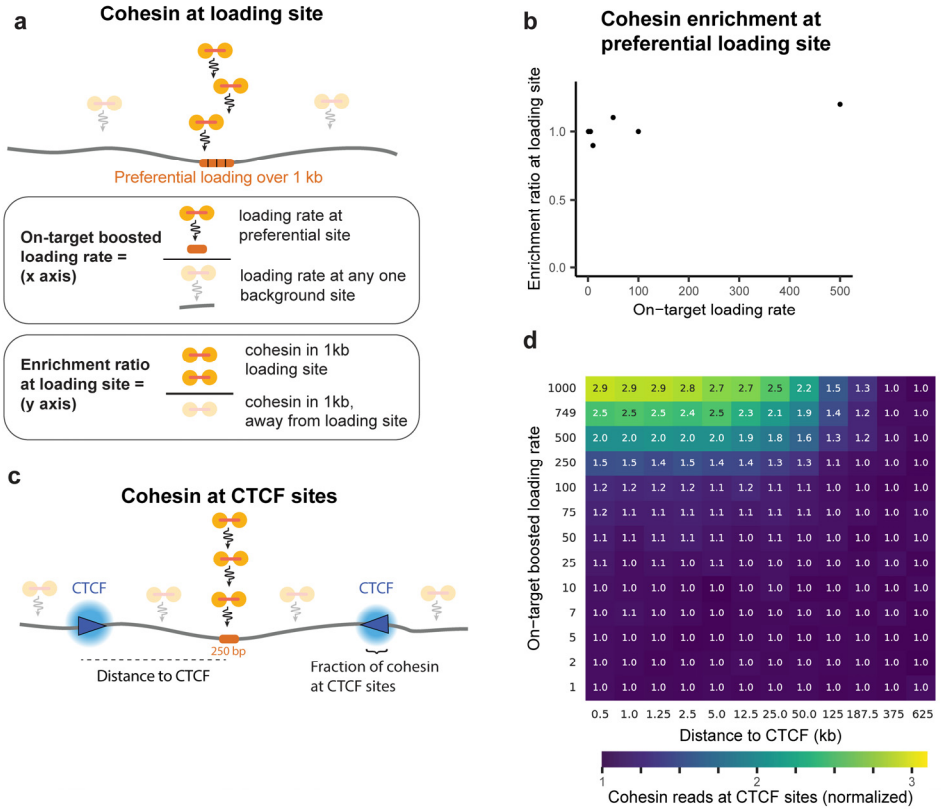


Figure 3: Simulations show minimal effect of targeted loading on cohesin traffic. **a**, and **b,** A 1D lattice simulation indicates that cohesin is not expected to accumulate at loading sites, even at high loading rates. The loading site is modeled as four adjacent monomers spanning 1 kb. The enrichment ratio is defined as the density of cohesin within a 1-kb window centered on the preferential loading site relative to background sites. **c**, Simulations exploring the effect of increasing loading rates at a single site on cohesin accumulation at neighboring CTCF sites (50% occupancy dynamic barrier, see Methods). Here, targeted loading rate at a single 250 bp monomer is increased while maintaining a constant loading rate at all other monomers. **d**, Number of cohesin reads at the CTCF site as a function of the on-target loading boost and the distance between the loading site and CTCF sites. As a reference, values are normalized to the condition with baseline loading rate (1x boost) and a CTCF site positioned 0.5 kb from the loading site.

Figure 4

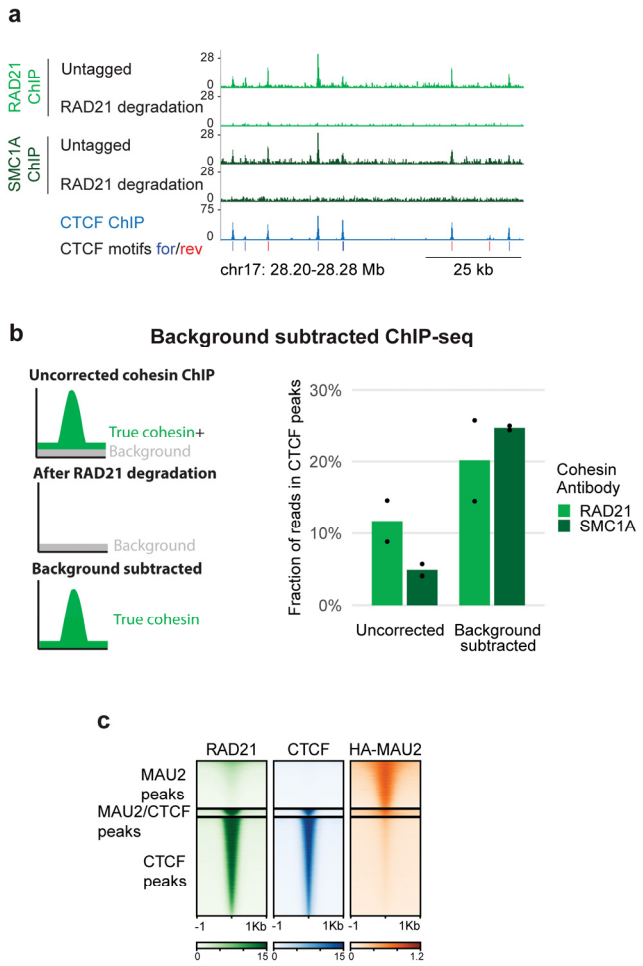


Figure 4: Cohesin binding is largely between CTCF sites

a, Spike-in calibrated ChIP-seq for cohesin subunits RAD21 and SMC1A in untagged ESCs or after cohesin degradation in RAD21-AID ESCs treated with auxin (IAA) for 3.5 hours. **b**, Strategy to calculate true cohesin binding by subtracting background antibody signal from the total reads. After background subtraction, CTCF peaks contain ~25% of cohesin ChIP-seq reads (average of two biological replicates). **c**, ChIP-seq shows little binding of the cohesin subunit RAD21 at NIPBL/MAU2 sites compared to CTCF sites.

Figure 5

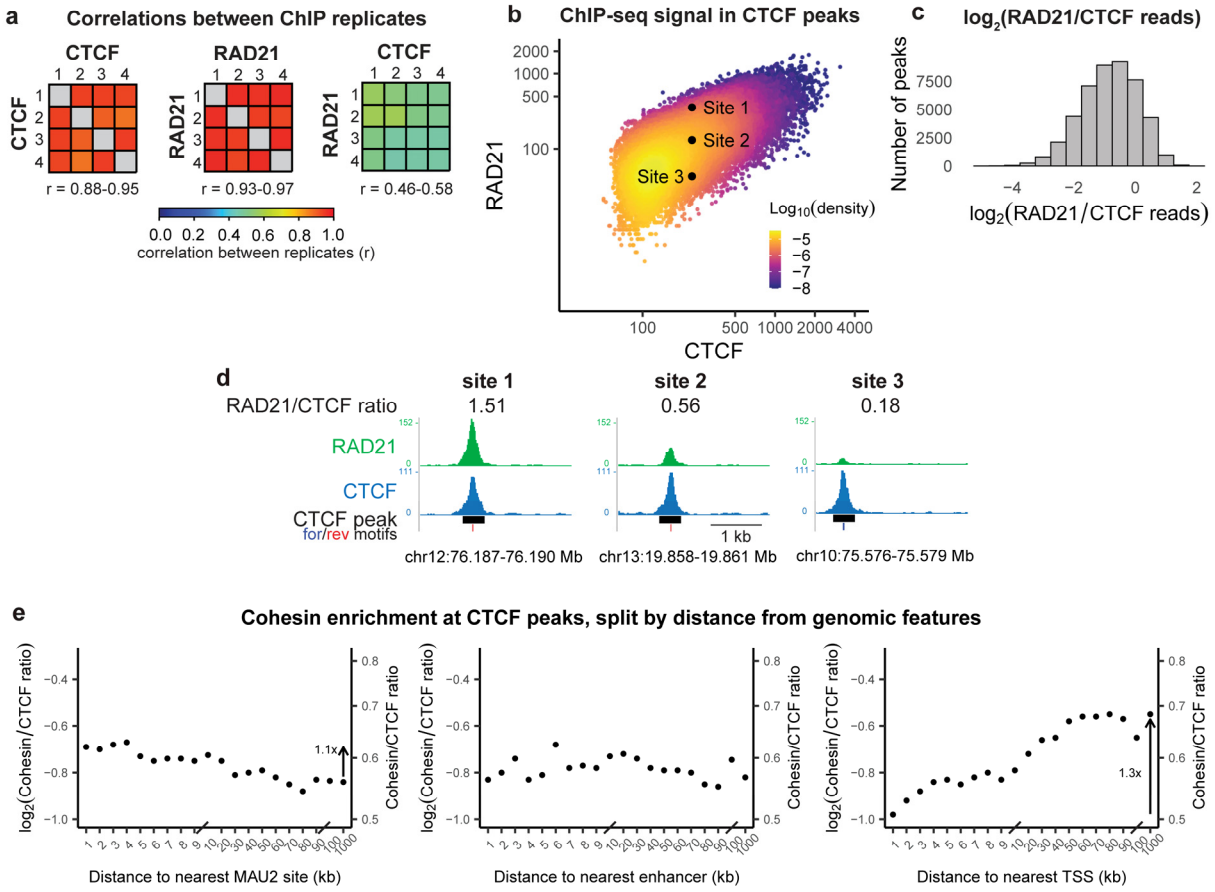


Figure 5: NIPBL/MAU2 binding sites have minimal effects on local cohesin levels. **a**, Correlation matrices (Pearson correlation coefficients) comparing ChIP–seq signal at CTCF sites across biological replicates for CTCF and RAD21, showing that replicates of the same factor correlate more strongly with each other than with the other factor. **b**, Scatter plot of RAD21 vs. CTCF reads in each CTCF peak (from four combined ChIP-seq replicates), highlighting substantial variability in cohesin occupancy at sites with similar CTCF binding. **c**, Distribution of RAD21/CTCF ratios. **d**, Example sites with the same CTCF occupancy but six-fold range in cohesin levels. **e**, Average cohesin/CTCF ratio as a function of the distance from the CTCF peak to the nearest MAU2 site, enhancer[38], or active TSS, showing minimal dependence on proximity to NIPBL/MAU2 binding and increased cohesin levels at greater distances from active TSSs.

Figure 6

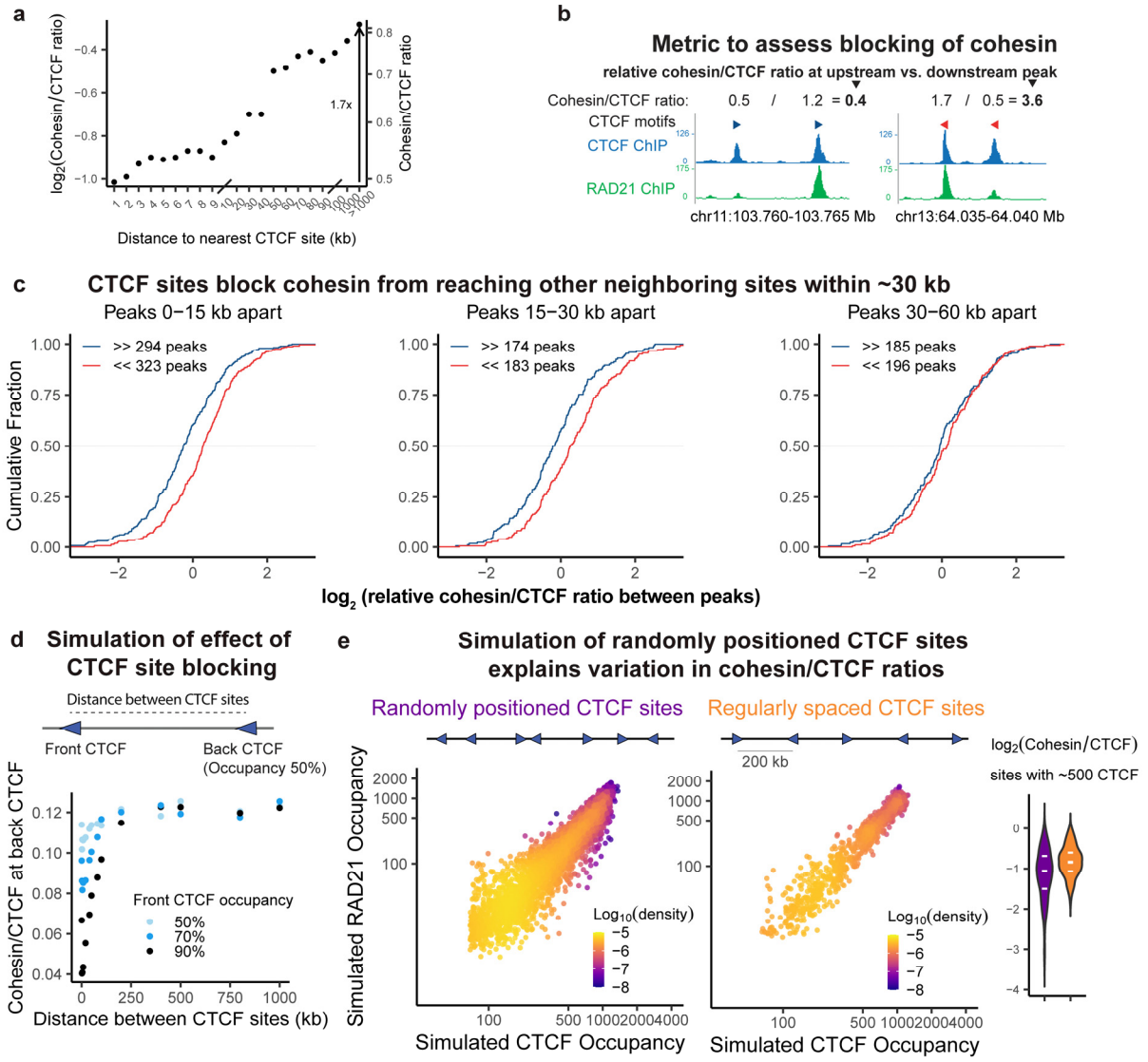
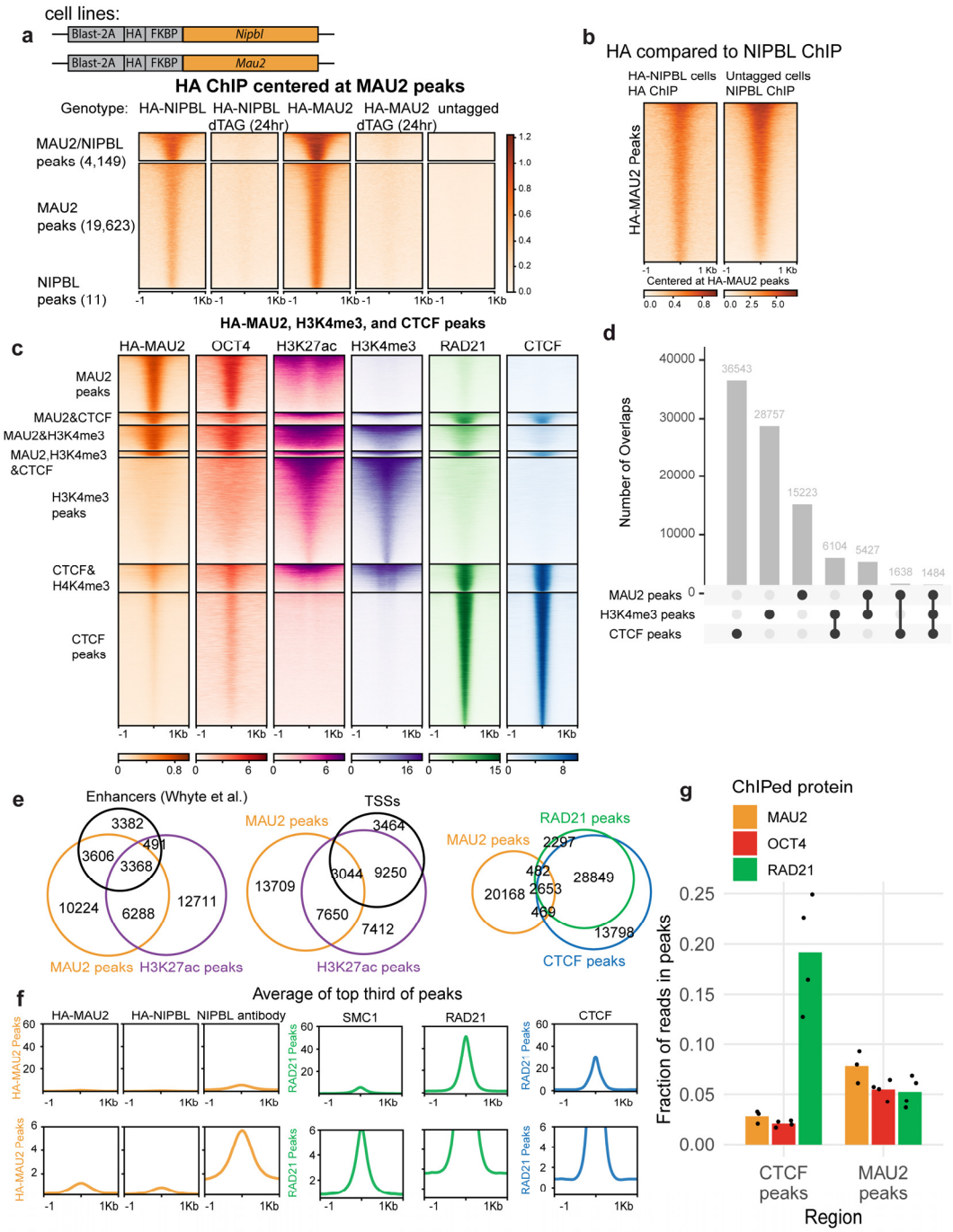


Figure 6: Clustered CTCF sites impede cohesin traffic.
a, Average cohesin/CTCF ratio at CTCF peaks as a function of distance to the nearest neighboring CTCF site, revealing reduced cohesin occupancy at closely spaced sites. **b**, We identified pairs of CTCF sites with forward motifs and where there is no other CTCF site within the downstream 100 kb. Relative cohesin/CTCF ratios compare upstream and downstream peaks based on motif orientation: a relative ratio less than 1 indicates that the rightmost peak is blocking cohesin from reaching the leftmost peak. For pairs of peaks with reverse motifs, a relative cohesin/CTCF ratio greater than 1 indicates blocking by the leftmost peak. **c**, Cumulative distributions of relative cohesin/CTCF ratios for CTCF site pairs with forward or reverse motifs, stratified by inter-site distance. Interference between neighboring CTCF sites is apparent only for sites within 30 kb of each other. **d**, In a simulation of two CTCF sites with reverse motifs, the cohesin/CTCF ratio at the rightmost site depends on the distance between sites and the occupancy of the left site. **e**, Cohesin vs. CTCF occupancy in a simulation of a 75 Mb

chromosome with randomly positioned CTCF sites shows more dispersion than a simulation with CTCF sites regularly positioned every 200 kb. The CTCF occupancy is varied by altering the bound time while keeping the CTCF unbound time constant between sites.

Extended Data Figures

Extended Data Figure 1

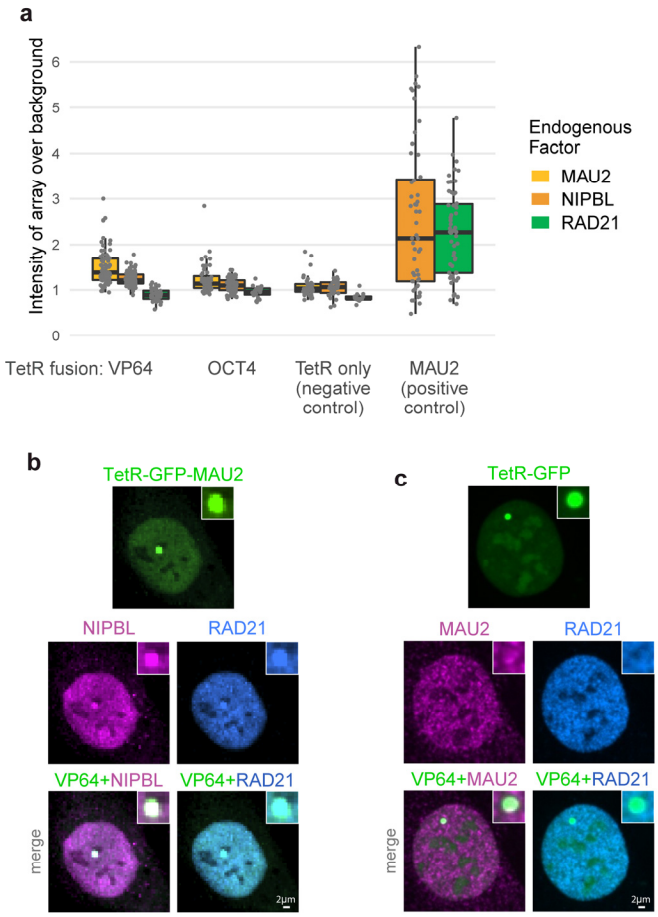


Extended Data Figure 1: Reproducible NIPBL/MAU2 binding sites overlap with enhancers in mouse ESCs.

a, HA ChIP-seq signal at MAU2 and NIPBL peaks is lost after NIPBL or MAU2 degradation (24 hr dTAG treatment). NIPBL and MAU2 co-bind the same sites, with the more sensitive HA-MAU2 ChIP-seq detecting more peaks. **b**, HA-NIPBL ChIP-seq signal closely matches ChIP-seq

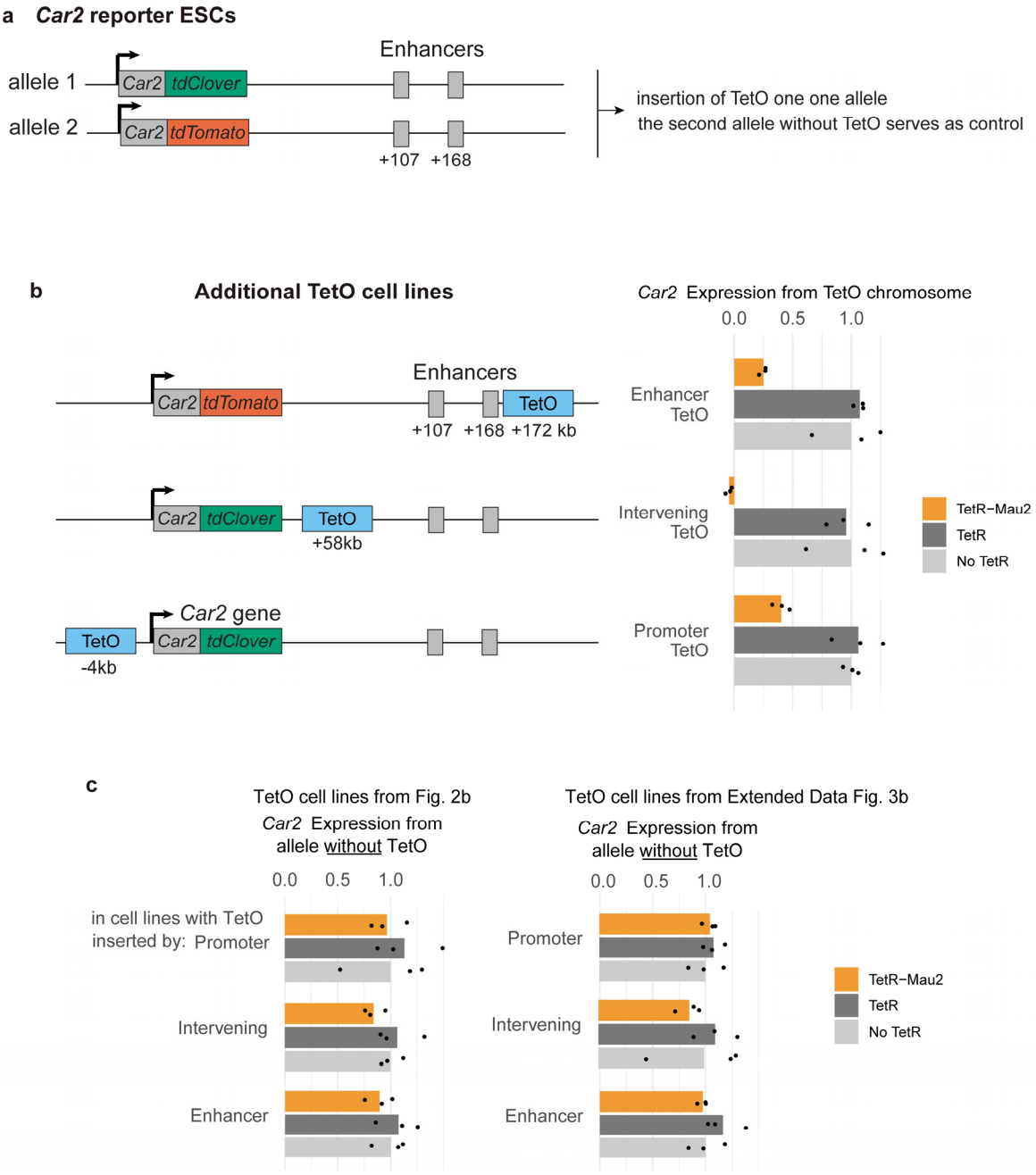
with an endogenous NIPBL antibody. **c**, Integration of ChIP-seq datasets highlighting that MAU2 binds the subset of active chromatin sites (H3K27ac) that bind OCT4 (enhancers). RAD21 (cohesin) is enriched at CTCF sites, with minimal signal at enhancers/MAU2 sites. OCT4 data is from ref[50] and H3K4me3 from ref[113]. **d**, Upset plot showing that MAU2, H3K4me3 (marking promoters), and CTCF largely bind distinct loci. **e**, Overlap of MAU2 peaks with active chromatin (H3K27ac peaks), enhancers (defined as sites bound by pluripotency transcription factors OCT4, SOX2, and NANOG from ref[38]), TSS regions (defined as 1 kb upstream of TSSs of GENCODE protein-coding genes expressed at >1 transcript per million), RAD21 peaks, and CTCF peaks. **f**, Average binding profiles of NIPBL/MAU2, cohesin subunits, and CTCF at their own peaks, showing that NIPBL/MAU2 binding sites have weak enrichment over background. Plots are shown on two different scales for comparison. **g**, Fraction of MAU2, OCT4, and CTCF ChIP-seq reads found in either CTCF or MAU2 peaks. OCT4 ChIP-seq is from ref[50,119].

Extended Data Figure 2



Extended Data Figure 2: Recruitment of cohesin by TetR-MAU2.
a, Intensity of MAU2, NIPBL, or RAD21 immunofluorescence signal at the TetO array divided by intensity of the background nuclear signal. **b**, Example cell showing that TetR-MAU2 recruits NIPBL and RAD21 to the TetO array. **c**, Example cell showing that TetR alone does not recruit MAU2 or RAD21 to the TetO array.

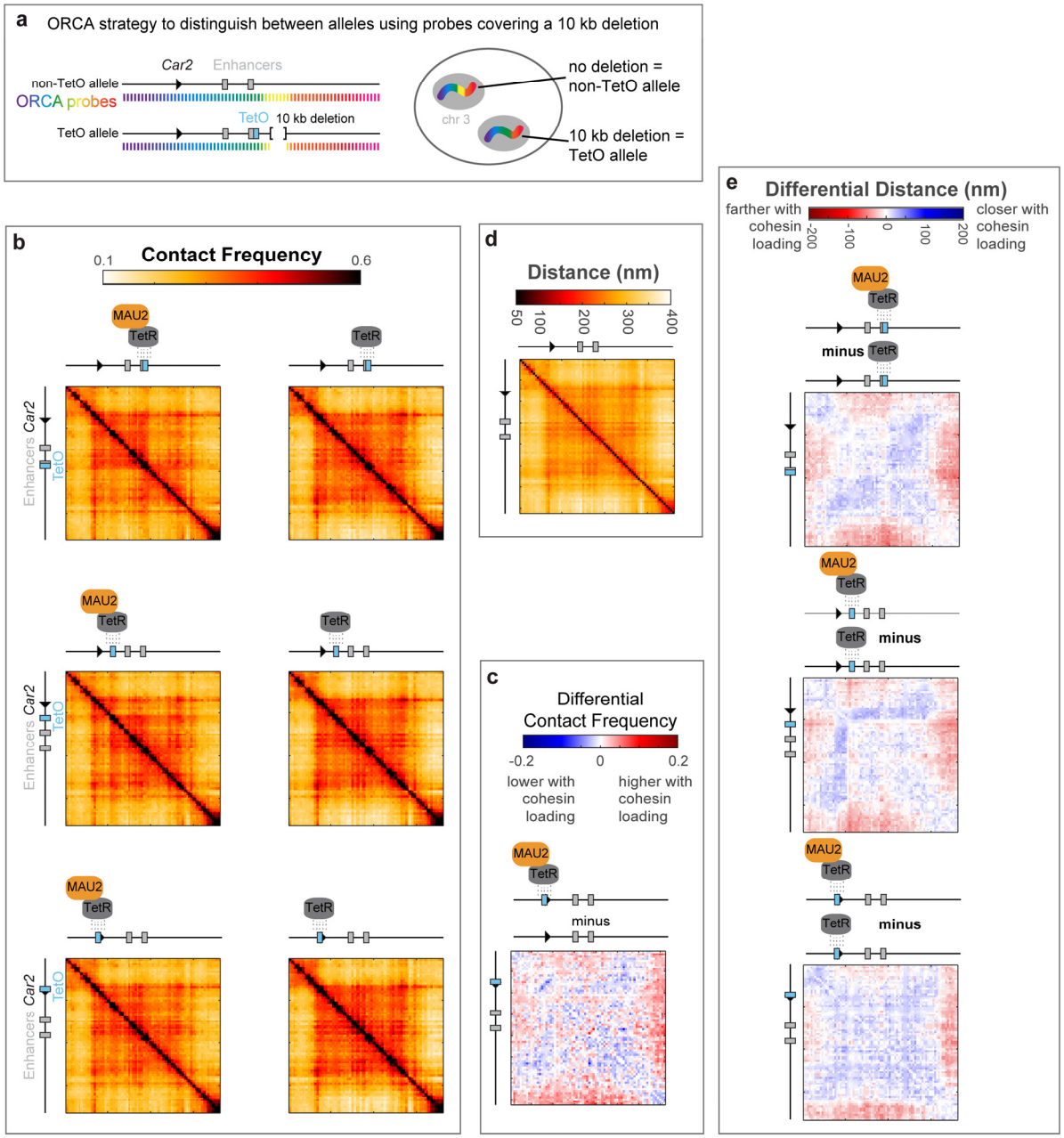
Extended Data Figure 3



Extended Data Figure 3: Additional clones for targeted cohesin recruitment. **a**, *Car2* reporter ESCs (from [41]): in each cell, one *Car2* allele harbors a tdClover reporter and the second allele harbors a tdTomato reporter. *Car2* expression in cells with heterozygous TetO insertions can therefore be directly compared between the TetO allele and control allele without the TetO. **b,** Additional cell lines with TetO insertions either on the tdClover or the tdTomato allele, replicating observations reported in Fig. 2. **c**, In the same cells that displayed reduced *Car2* expression on the TetO allele upon TetR-Mau2 expression, *Car2* expression from the

1493 control allele without the TetO insertion remains unchanged across three biological replicates
1494 from separate transfections.

Extended Data Figure 4



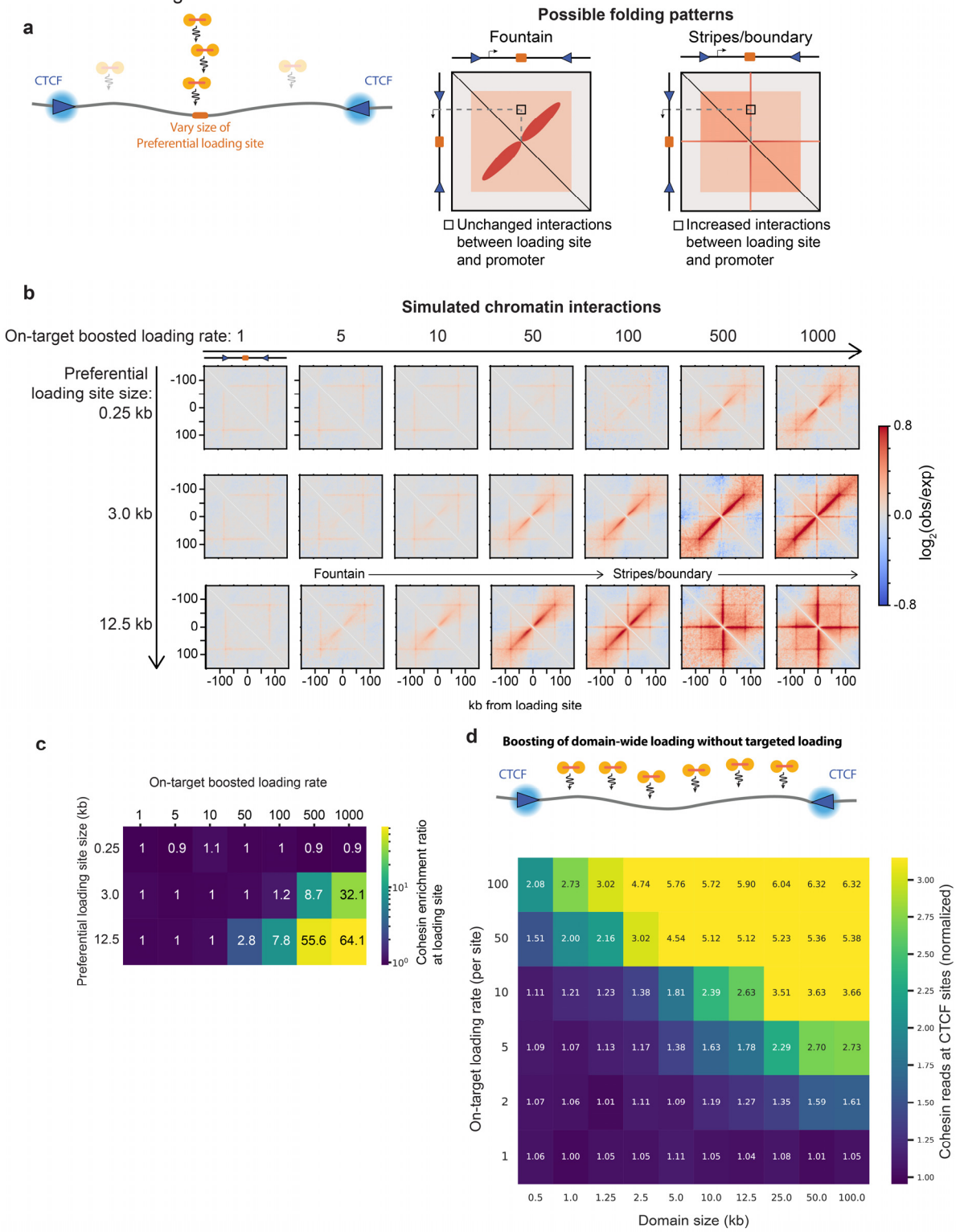
Extended Data Figure 4: ORCA measurements comparing alleles with and without high cohesin loading.

a, The TetO array used for cohesin recruitment is inserted on one allele, where we also engineered a neutral 10 kb deletion to distinguish the two alleles. This deletion enables distinguishing the two alleles, as the TetO allele will not display ORCA signal from probes covering the 10 kb deletion. **b**, ORCA contact frequencies (~200 nm, see methods) for the TetO allele in cells expressing TetR-MAU2 or TetR only (8 kb resolution). **c**, Comparing ORCA

1504 interaction frequencies between the TetO and non-TetO allele in TetR-MAU2 cells displays less
1505 change in interactions than comparing the TetO alleles between TetR-MAU2 cells and TetR
1506 control cells. This indicates that MAU2 expression slightly diminishes ORCA interaction
1507 frequencies. **d**, Heatmap showing the average distances at the *Car2* locus. **e**, Subtraction
1508 heatmaps comparing average distances in cells expressing TetR-MAU2 or TetR alone.

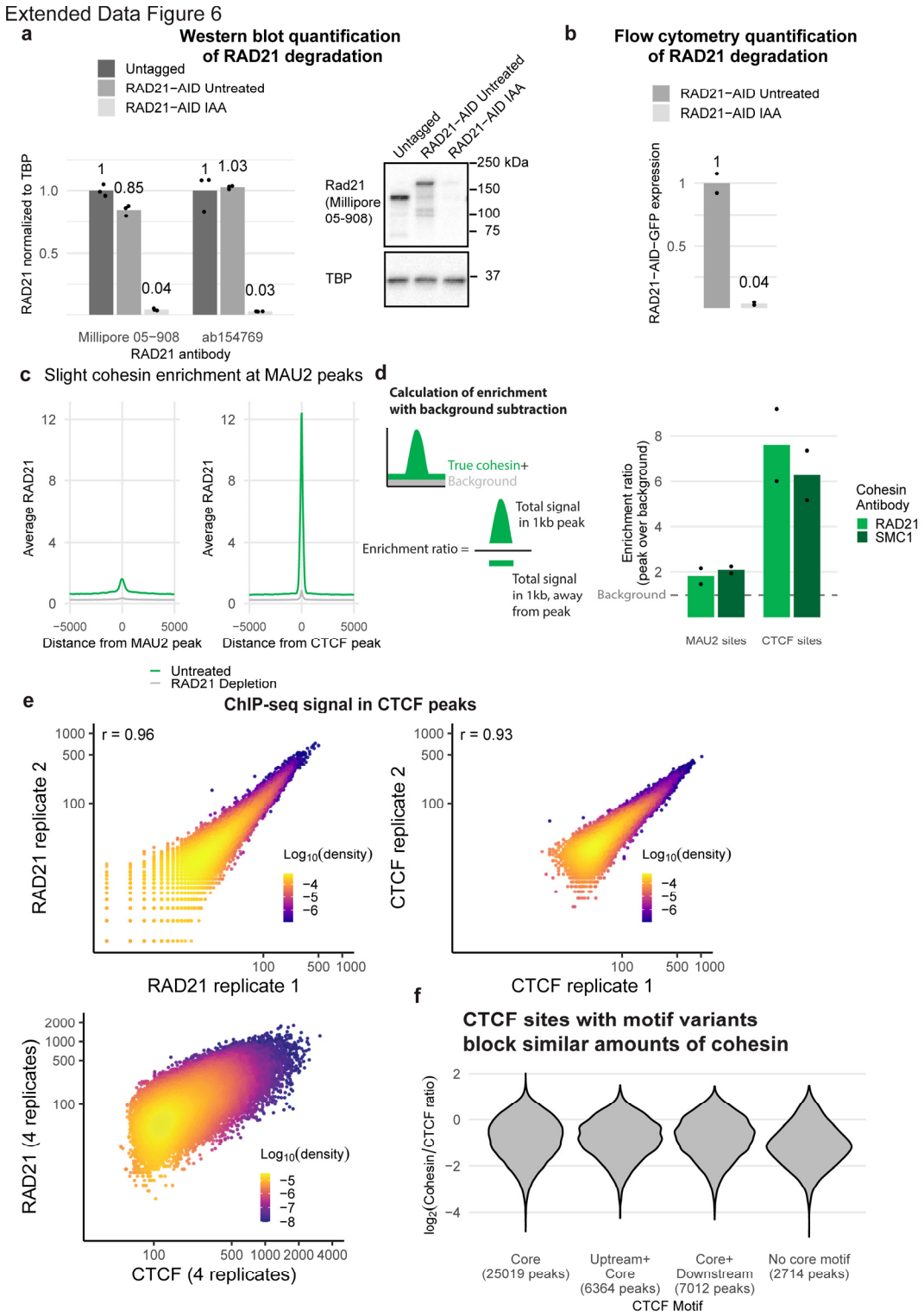
1509

Extended Data Figure 5



1510

Extended Data Figure 5: Width of cohesin loading site affects cohesin blocking and traffic. **a**, 3D polymer simulations in which the rate of cohesin loading is boosted at a loading site of varying size. **b**, Simulated Hi-C-like contact maps (2.5 kb resolution). Increasing cohesin loading creates fountain structures. For large loading sites with high loading, stripe patterns also emerge from the loading site, as colliding cohesin molecules block each other. **c**, 1D lattice simulations as in Figure 3A and 3B show that cohesin accumulates at loading sites only if loading is high across multiple adjacent monomers. The enrichment ratio at the loading site is calculated as the amount of cohesin in 1 kb at the center of the loading site divided by the amount of cohesin in 1 kb at a background site. **d**, Simulation exploring how boosting cohesin loading across domains of increasing size affects cohesin abundance at flanking CTCF sites.



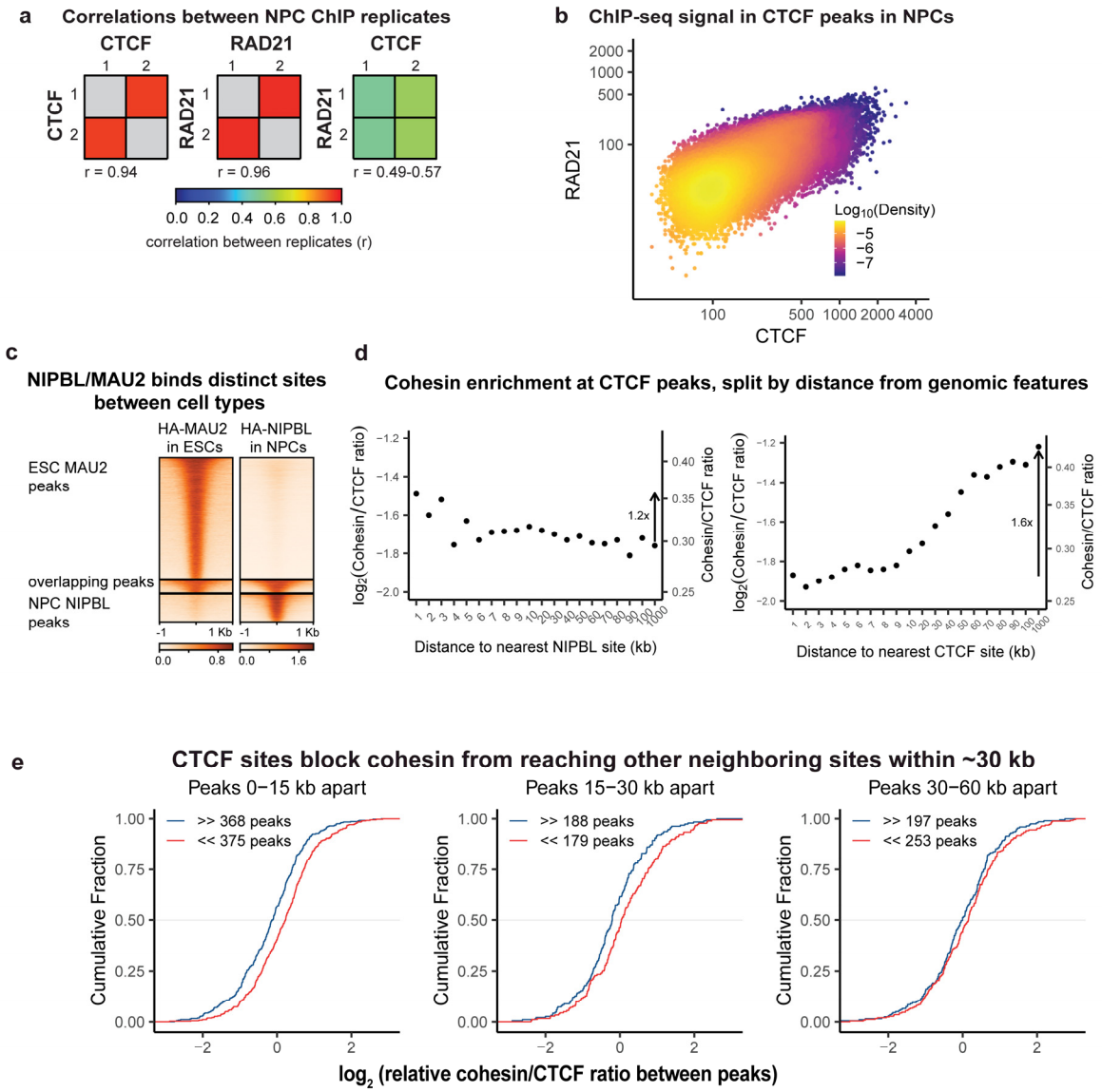
Extended Data Figure 6: Cohesin is slightly enriched at MAU2 peaks, and cohesin at CTCF peaks is weakly correlated with CTCF occupancy.
a, Western blots and **b**, flow cytometry show that in RAD21-AID-GFP cells treated with IAA for 3.5 hours, >95% of RAD21 was degraded. **c**, Average RAD21 signal at MAU2 peaks (that do not

1526 overlap CTCF peaks) or at CTCF peaks before or after 3.5 hr RAD21 depletion. **d**, Average
1527 enrichment of cohesin binding at MAU2 or CTCF peaks after background subtraction. **e**, At
1528 CTCF peaks, the number of reads is highly correlated between replicates of CTCF or RAD21
1529 ChIP-seq but shows wider dispersion between RAD21 and CTCF (four combined replicates
1530 reproduced from Fig. 5B for comparison). **f**, RAD21/CTCF ratios for CTCF sites based on the
1531 presence of upstream or downstream CTCF motifs[52]. Sites with upstream, downstream, or only
1532 core motifs show similar distributions, while peaks with no identifiable core CTCF motif block
1533 less cohesin.

1534

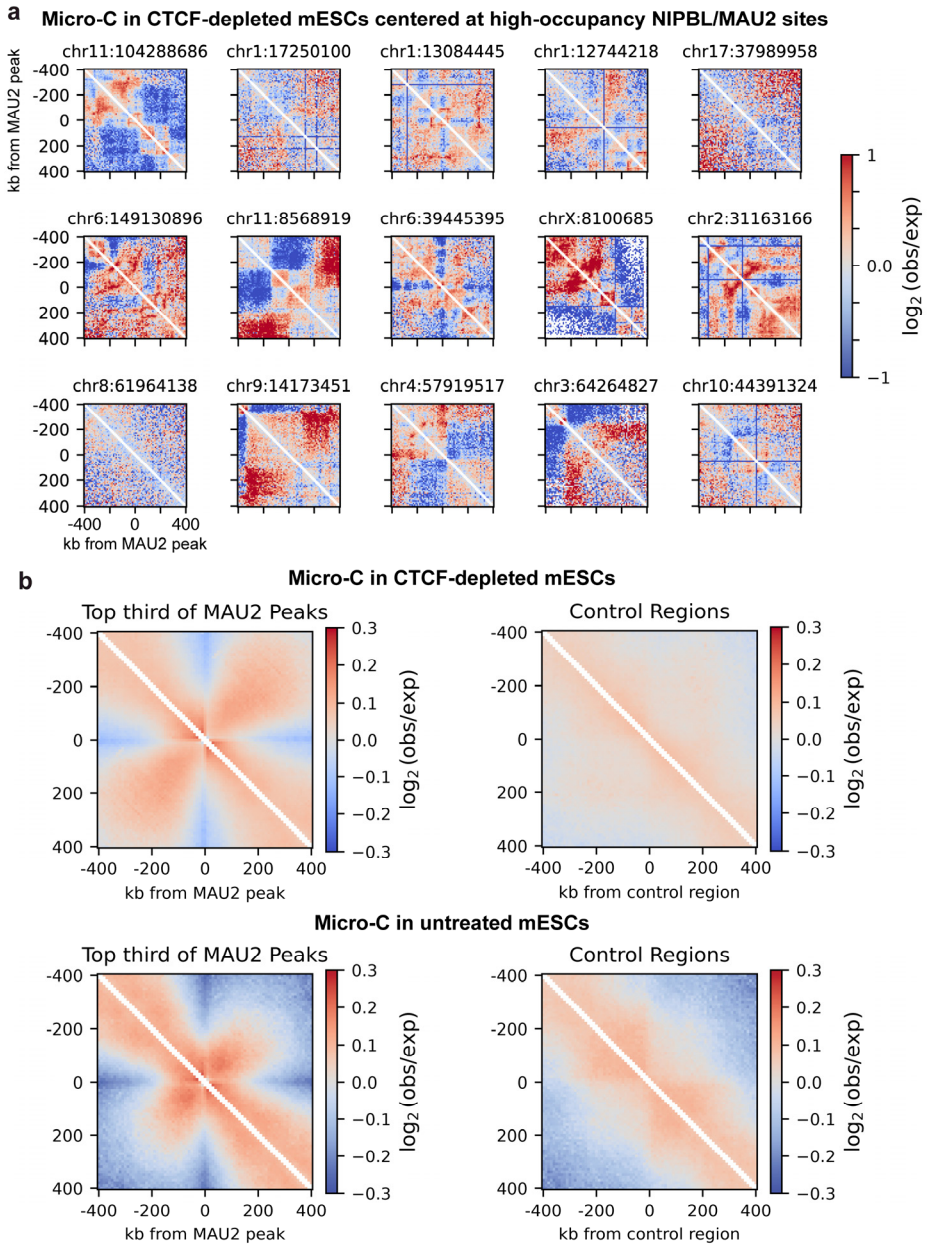
1535

Extended Data Figure 7



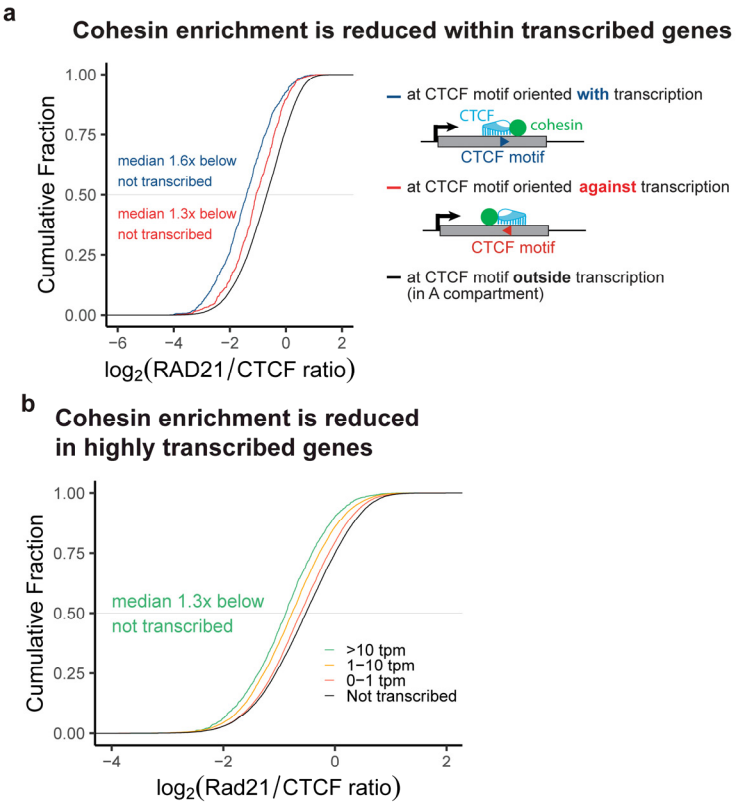
Extended Data Figure 7: CTCF sites, more than NIPBL/MAU2 sites, control cohesin traffic in Neural Progenitor Cells (NPCs) differentiated from ESCs.
a, At CTCF peaks, the number of ChIP-seq reads is more highly correlated between two CTCF replicates or two RAD21 replicates, compared to the correlation between CTCF and RAD21 (Pearson correlation coefficient). **b**, RAD21 vs. CTCF reads in each CTCF peak. Read numbers were normalized to match the number in ESCs in Figure 5B. **c**, NIPBL/MAU2 binding sites are largely not overlapping between ESCs and NPCs. **d**, Average ratio of cohesin/CTCF reads in CTCF peaks based on the distance from the peak to the nearest NIPBL or CTCF site. **e**, The distance over which CTCF sites block cohesin in NPCs was analyzed as in Figure 6C. Similarly to ESCs, pairs of CTCF sites with forward motifs have lower relative cohesin/CTCF ratios than pairs with reverse motifs only when the peaks are within 30 kb of each other.

Extended Data Figure 8



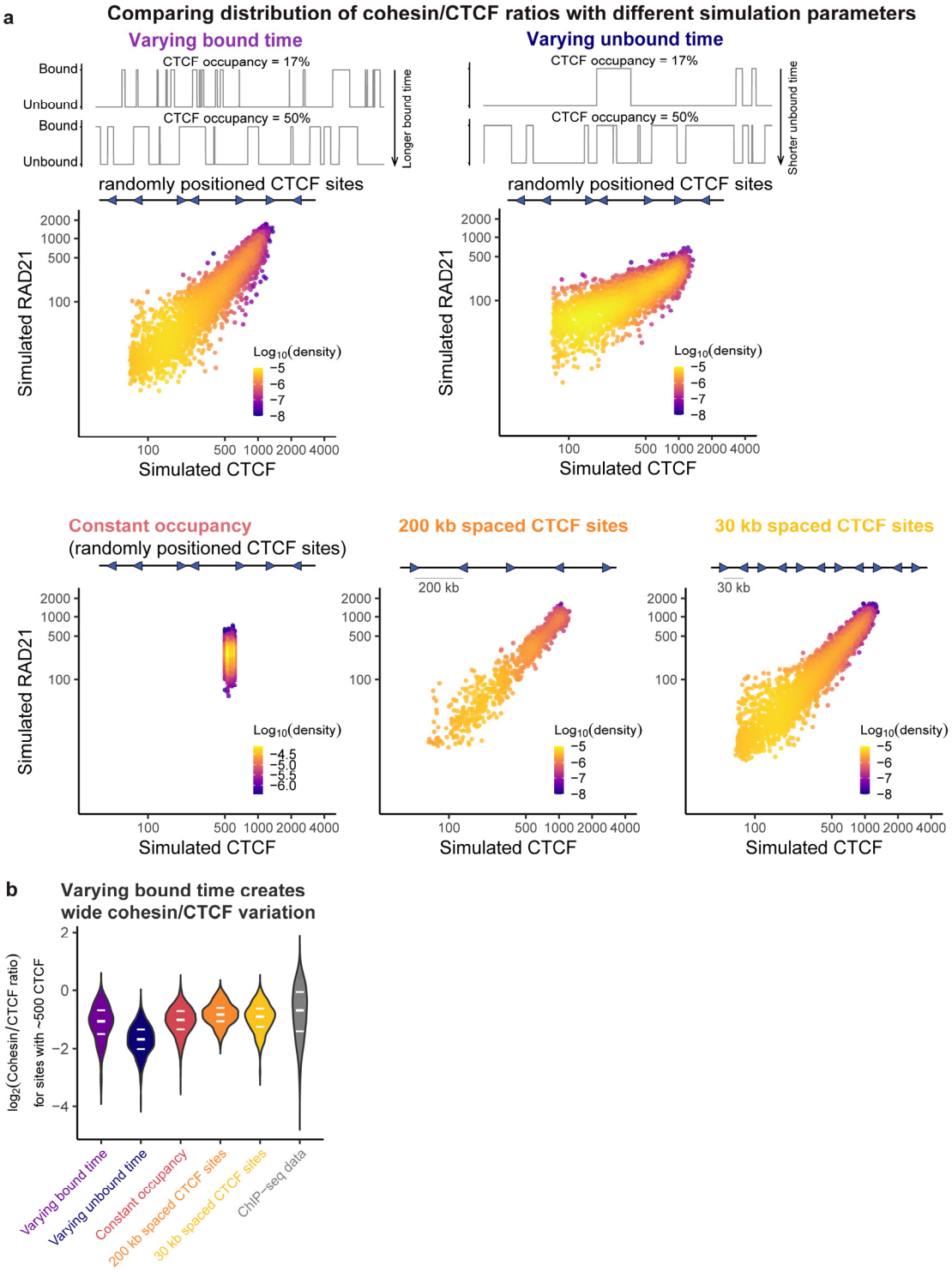
Extended Data Figure 8: NIPBL/MAU2 sites only mildly change local chromatin folding. **a**, At 15 of the highest occupancy individual MAU2 sites, micro-C in CTCF-depleted cells does not show fountain patterns. Heatmaps are plotted at 10 kb resolution. **b**, A pile-up of micro-C data in CTCF-depleted or untreated ESCs[116] at the highest-occupancy third of MAU2 sites shows a weak fountain pattern, while randomly chosen sites within the same TADs as the MAU2 sites do not.

Extended Data Figure 9



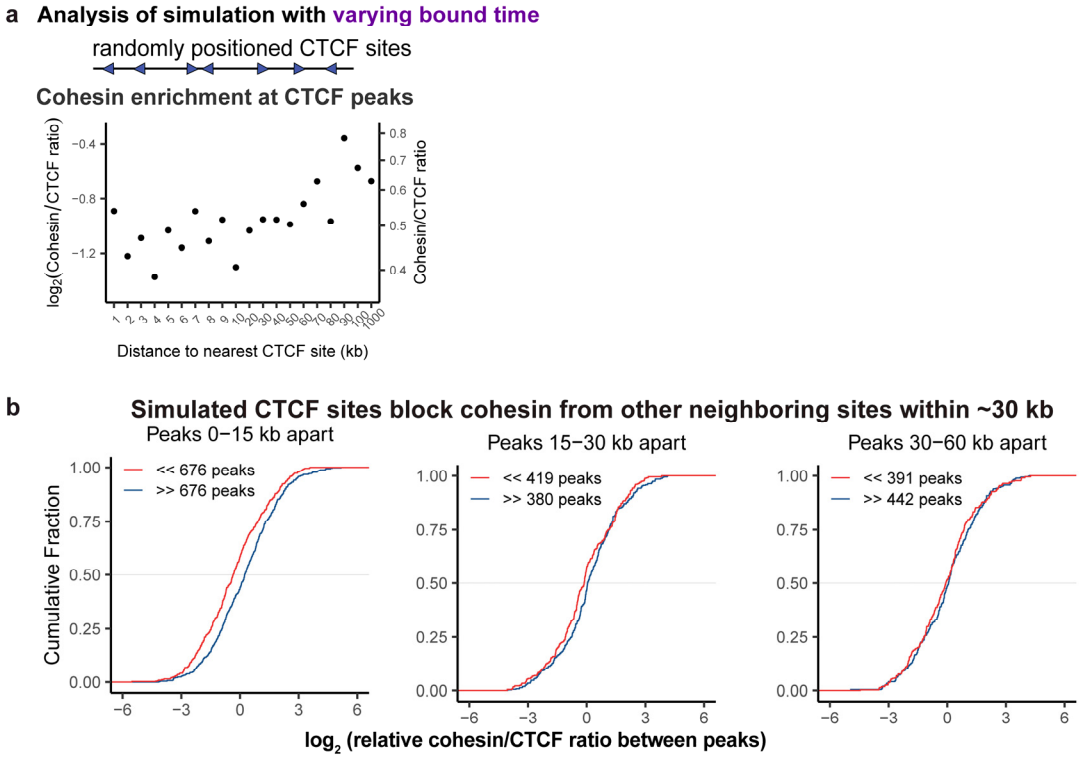
Extended Data Figure 9: Cohesin traffic is impeded in actively transcribed loci. **a**, Cohesin enrichment is reduced at CTCF sites within GENCODE protein-coding genes with RNA-seq transcription >10 tpm[113]. **b**, Cumulative plot shows lower RAD21/CTCF ratios for CTCF sites within gene bodies based on their transcription level.

Extended Data Figure 10



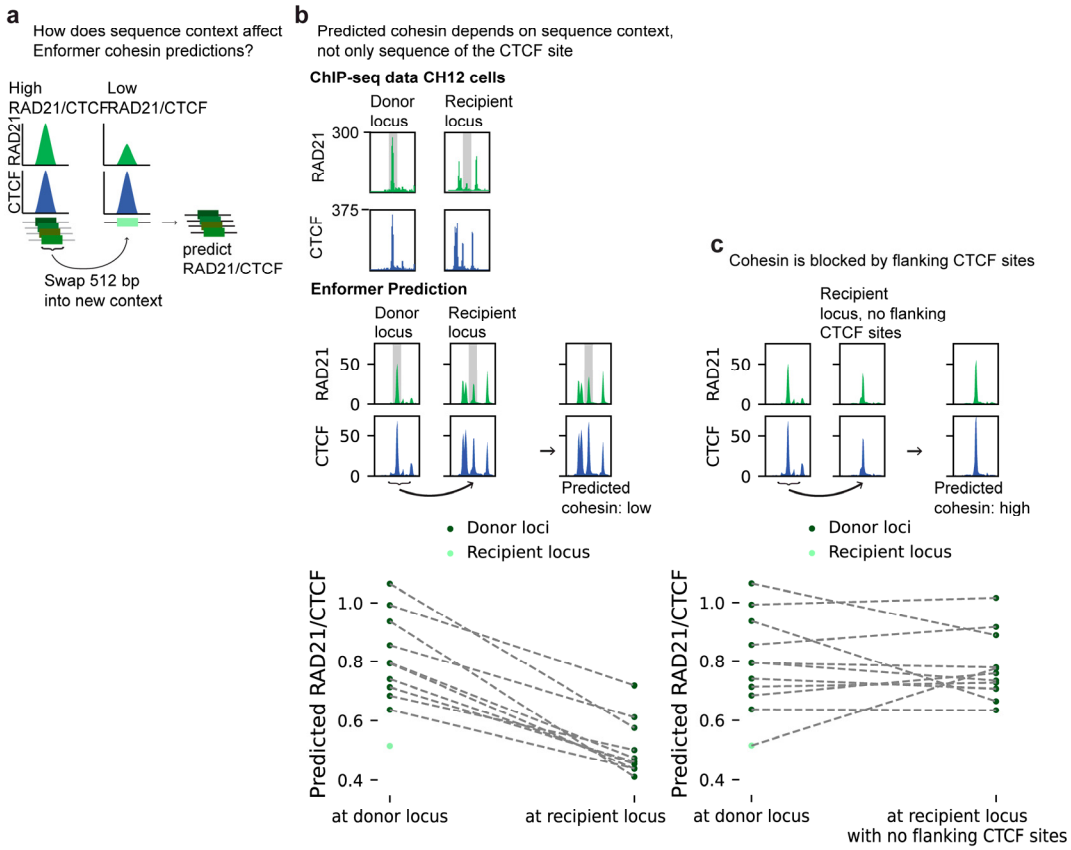
Extended Data Figure 10: Comparison of CTCF/cohesin ratios using simulations with different parameters.
a, Cohesin and CTCF occupancies at CTCF sites from simulations using different parameters for CTCF site occupancy and spacing. **b**, Distributions of CTCF/cohesin ratios for each parameter compared to ChIP-seq data.

Extended Data Figure 11



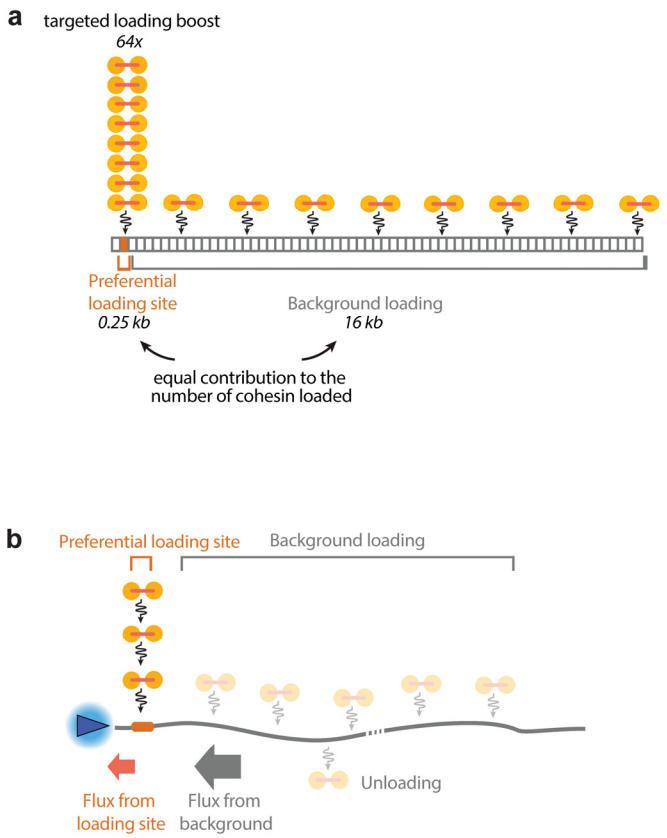
Extended Data Figure 11: In simulations, neighboring CTCF sites block cohesin traffic. **a**, Simulation quantification showing that CTCF sites near other sites have lower cohesin occupancy (compare to ChIP-seq data in Fig. 6A, Extended Data Fig. 7D). **b**, Simulation quantification showing that CTCF sites block cohesin from other CTCF sites within ~30 kb (compare to ChIP-seq data in Fig. 6C, Extended Data Fig. 7E).

Extended Data Figure 12



Extended Data Figure 12: Enformer predicts that clustered CTCF sites alter cohesin accumulation.
a, Using Enformer[118], we selected sites with high predicted CTCF and RAD21 levels in CH12 cells (a mouse B cell lymphoma line), swapped 512 bp from the CTCF binding site into a recipient sequence context where a high-occupancy CTCF site has low RAD21 binding, and predicted binding for the chimeric sequences. **b**, The chimeric sequences show low predicted RAD21, indicating that the sequence context affects the predicted cohesin binding. Cohesin and CTCF binding ChIP-seq data from CH12 cells[113] at the donor and recipient loci match the Enformer predictions. **c**, Without the CTCF motifs flanking the central site of the recipient locus, the predicted cohesin increased, indicating that Enformer predictions are influenced by CTCF motifs in the surrounding sequence.

Extended Data Figure 13



Extended Data Figure 13: Cartoons comparing the contributions of targeted versus background cohesin loading

a, As an example, with a 64-fold targeted loading boost, a 0.25 kb loading site contributes to loading as much cohesin as a 16 kb background region. **b**, Cartoon illustrating the relative contributions of targeted and background cohesin loading to the flux of cohesin reaching a CTCF site. Even when a preferential loading site exists, its contribution is easily outweighed by cohesin loaded across the surrounding chromatin at the background rate.