nature structural & molecular biology

Article

https://doi.org/10.1038/s41594-022-00890-9

A cohesin traffic pattern genetically linked to gene regulation

Received: 13 December 2021

Accepted: 1 November 2022

Published online: 8 December 2022

Check for updates

Anne-Laure Valton[1,2], Sergey V. Venev[1], Barbara Mair[3], Eraj Shafiq Khokhar[4], Amy H. Y. Tong[3], Matej Usaj[3], Katherine Chan[3], Athma A. Pai[4], Jason Moffat[3] & Job Dekker[1,2]

Cohesin-mediated loop extrusion has been shown to be blocked at specific *cis*-elements, including CTCF sites, producing patterns of loops and domain boundaries along chromosomes. Here we explore such *cis*-elements, and their role in gene regulation. We find that transcription termination sites of active genes form cohesin- and RNA polymerase II-dependent domain boundaries that do not accumulate cohesin. At these sites, cohesin is first stalled and then rapidly unloaded. Start sites of transcriptionally active genes form cohesin-bound boundaries, as shown before, but are cohesin-independent. Together with cohesin loading, possibly at enhancers, these sites create a pattern of cohesin traffic that guides enhancer-promoter interactions. Disrupting this traffic pattern, by removing CTCF, renders cells sensitive to knockout of genes involved in transcription initiation, such as the SAGA complexes, and RNA processing such DEAD/H-Box RNA helicases. Without CTCF, these factors are less efficiently recruited to active promoters.

At the scale of tens to hundreds of kilobases, the genome folds into topologically associating domains (TADs) or loop domains, as well as locus-specific chromatin loops[1–4]. TADs and loops are formed by a loop extrusion mechanism mediated by cohesin complexes[5–7]. The cohesin complex dynamically extrudes loops and accumulates most prominently at CTCF sites during interphase[8–10]. Cohesin also accumulates at active promoters and, under specific conditions, near 3′ ends of genes and sites of convergent transcription[11]. The interplay between cohesin and CTCF drives loop extrusion, leading to enrichment of interactions within TADs, depletion of interactions across TAD boundaries (insulation) and looping between CTCF sites[7,8,10,12–17].

Many open questions remain to be addressed to fully understand the process of loop extrusion. It is currently not known how and where cohesin is recruited to chromatin, how the complex actively extrudes loops, whether extrusion occurs unidirectionally or bidirectionally, or whether additional types of *cis*-element that function equivalently to CTCF-bound sites exist along chromosomes, and how they cooperate to regulate cohesin dynamics.

TADs are thought to regulate gene expression by allowing enhancer-promoter interactions within the domain, while disfavoring such interactions across their boundaries[14,18–22]. However, acute global depletion of CTCF or the cohesin subunit RAD21 leads to only a small number of gene expression changes, despite genome-wide loss of TADs and CTCF-CTCF loops (CTCF depletion[8,23]), or loss of all extrusion features (RAD21 depletion[12]). Therefore, the biological functions of loop extrusion, and cohesin blocking at specific sites, remain poorly understood.

In this Article, we analyze Hi-C data from cells depleted of CTCF, RAD21, WAPL or RNA polymerase II to describe the intricate local folding of chromosomes and the roles of different types of *cis*-element in guiding cohesin-mediated loop extrusion. From this analysis, a complex picture of cohesin trafficking along chromosomes emerges. To uncover functional roles for this intricate chromosome organization, we performed genome-wide CRISPR screens in cells with altered cohesin traffic patterns, and identified genes involved in transcription initiation and RNA processing, finding that these factors are mislocalized when the cohesin traffic pattern is disrupted.

[1]Department of Systems Biology, University of Massachusetts Chan Medical School, Worcester, MA, USA. [2]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [3]Donnelly Centre, University of Toronto, Toronto, Ontario, Canada. [4]RNA Therapeutics Institute, University of Massachusetts Chan Medical School, Worcester, MA, USA. e-mail: Job.Dekker@umassmed.edu

Results

Active transcription start sites are CTCF-independent chromatin domain boundaries

To further understand the functional roles of CTCF-CTCF chromatin loops and potentially reveal other elements that could influence or control the loop extrusion machinery, we acutely depleted CTCF using an auxin-inducible degron system from HAP1-derived human cells. This cell line, HAP1-CTCFdegron-TIR1, expresses CTCF fused to AID at both the N- and C-termini and a C-terminal green fluorescent protein (GFP) tag, as well as the TIR1 F-box protein, which mediates auxin-inducible protein degradation (Extended Data Fig. 1a). Addition of auxin to HAP1-CTCFdegron-TIR1 cells resulted in efficient depletion of CTCF (Extended Data Fig. 1b,c). We also generated the HAP1-CTCFdegron cell line without TIR1. We note that, even without addition of auxin, the CTCF protein level was reduced compared to the level observed in HAP1-CTCFdegron cells lacking TIR1 (Extended Data Fig. 1b,c). The cell-cycle profile of HAP1-CTCFdegron-TIR1 cultures was not altered after 48 h of CTCF depletion (Extended Data Fig. 1d).

We performed Hi-C on HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 cells grown in the absence or presence of auxin for 48 h. As a measure of average loop size, we analyzed the relationship of the Hi-C interaction frequency as a function of genomic distance between loci[24]. Interestingly, we found that the average loop size increased progressively when CTCF levels were reduced or entirely depleted. When CTCF is removed, CTCF-mediated blocking of loop extrusion is reduced or abolished, and longer loops are extruded (Extended Data Fig. 1e). Compartmentalization was only modestly affected (Extended Data Fig. 1f,g). Reduced CTCF levels in HAP1-CTCFdegron-TIR1 cells compared to HAP1-CTCFdegron cells, in the absence of auxin, resulted in weaker domain boundaries at CTCF sites and weaker CTCF-CTCF loops (Extended Data Fig. 1h–j). Depletion of CTCF by auxin addition resulted in near complete loss of looping interactions between CTCF sites and loss of insulation at domain boundaries. This is consistent with previous observations in CTCF degron cell lines[8,10,13].

To assess how the genomic positioning of cohesin is affected after CTCF depletion, we performed chromatin immunoprecipitation sequencing (ChIP-seq) for CTCF and the cohesin subunit RAD21 in HAP1-CTCFdegron-TIR1 cells in the presence or absence of CTCF. Only ~40% of the CTCF peaks overlapped with RAD21 peaks, showing that not all CTCF sites are associated with cohesin. Moreover, ~50% of the RAD21 peaks overlapped with CTCF peaks, suggesting that RAD21 can accumulate at locations devoid of CTCF (Fig. 1a).

We were interested in determining whether the sites that accumulate cohesin but not CTCF are able to form chromatin domain boundaries. Domain boundary formation can be quantified by the insulation metric, which measures the extent to which long-range chromatin interactions across a boundary are reduced compared to a global average[25]. To characterize the elements at which cohesin accumulates, we created a union list of all RAD21 peaks detected in either the presence or absence of CTCF and all the CTCF peaks detected in CTCF-expressing cells. We then analyzed CTCF and RAD21 accumulation and insulation for these sites in CTCF-expressing and CTCF-depleted HAP1-CTCFdegron-TIR1 cells, then ranked these sites by the level of CTCF binding in CTCF-expressing cells. We also assessed the active promoter mark H3K4me3 and coding gene locations from published datasets (Fig. 1b)[26]. We identified three major groups of elements. The first group binds both CTCF and RAD21 at high levels, with most sites included in the sets of significantly enriched CTCF and RAD21 peaks. These sites displayed strong insulation, indicating that they form domain boundaries. Sites in this group lose RAD21 binding and insulation upon CTCF depletion. The second group also bound CTCF at high levels and were often included in the set of significant CTCF peaks. However, these sites did not bind RAD21 and did not display insulation, indicating that they were not chromatin domain boundaries. The third group did not show enriched CTCF binding, but displayed relatively high levels of RAD21 binding and insulation in control cells. Most of these sites contained active promoters/transcription start sites (TSSs). Upon CTCF depletion, these sites continued to accumulate RAD21 and to display insulation. This analysis shows that active promoters/TSSs can act as domain boundaries, as was shown previously in the mouse[27,28], and these boundaries are CTCF-independent (Fig. 1b).

By aggregating Hi-C interactions at CTCF-dependent sites and at active promoters/TSSs lacking CTCF binding, and inspecting representative examples, we confirmed that insulation is lost at CTCF-dependent sites but persists at the promoter/TSS sites after CTCF depletion (Fig. 1c,d). Importantly, most of these boundaries did not overlap with compartment boundaries and therefore were bona fide cohesin-bound chromatin domain boundaries (Extended Data Fig. 2a).

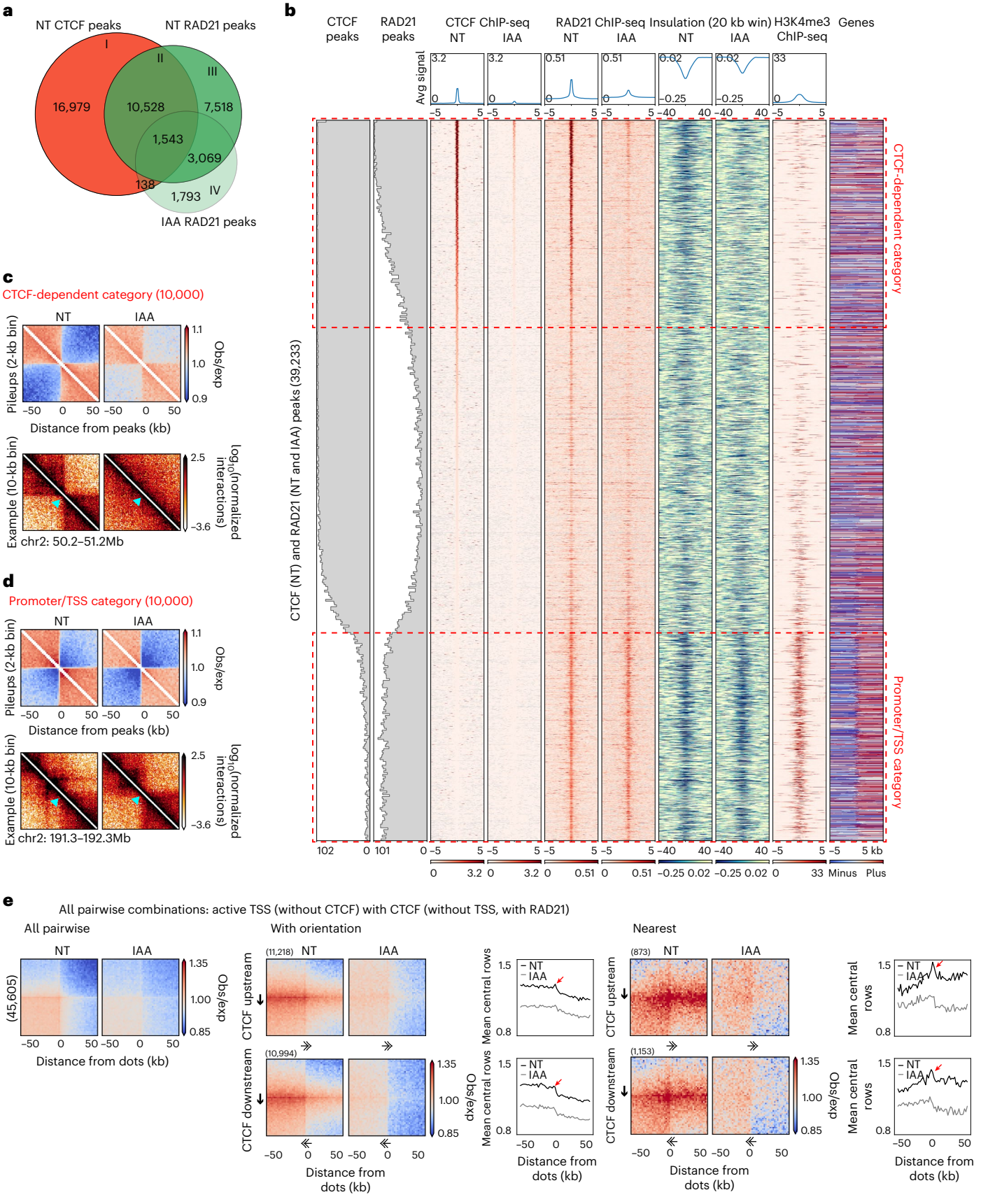
We further confirmed that these boundaries were active promoters/TSSs by analyzing CTCF and RAD21 binding and insulation in relation to H3K4me3 levels, as well as the RNA-seq signal (Extended Data Fig. 2b). We concluded that RAD21 is enriched in at least two different locations that form domain boundaries: (1) at CTCF sites, where RAD21 accumulation is dependent on CTCF, and (2) at active promoters/TSSs independent of CTCF binding[27]. Furthermore, insulation at such TSSs is maintained when CTCF is depleted, indicating it does not depend on distal CTCF sites.

We next examined long-range looping interactions between boundaries by aggregating interactions for all pairwise combinations between different types of cohesin-bound site separated by 50–500 kb. We found that active promoters/TSSs that lack CTCF binding frequently interacted with nearby CTCF sites that lack TSSs and display RAD21 binding. The interactions between active promoters/TSSs and distal CTCF sites were not due to any intervening CTCF sites or active promoters/TSSs (Fig. 1e and Extended Data Fig. 2c). As expected, all these interactions were CTCF-dependent and were most frequent when the CTCF motif was upstream of the promoter pointing towards the TSS. The orientation of the TSS itself appeared less consequential (Fig. 1e and Extended Data Fig. 2c).

Fig. 1 | Active promoters/TSSs form CTCF-independent chromatin domain boundaries. a, Venn diagram showing the overlap between CTCF peaks without auxin (non-treated, NT) and RAD21 peaks without/with auxin (NT and IAA; IAA, 3-indoleacetic acid). **b**, Stackups of the union list of all CTCF peaks called without auxin and RAD21 peaks called without/with auxin, sorted on the NT CTCF ChIP-seq signal. CTCF and RAD21 ChIP-seq signals and calculated insulation without/with auxin are plotted along with published HAP1 H3K4me3 ChIP-seq data and the genes (red, plus, forward strand; blue, minus, reverse strand; gray, no annotated transcripts)[26]. The distributions of CTCF and RAD21 peaks are plotted along the stackup. Red dashed rectangles highlight the CTCF-dependent category and the promoter/TSS category. **c**, Interaction pileup for the CTCF-dependent category (10,000 sites), without/with auxin, aggregated in a 100-kb window at 2-kb resolution and plotted along with a representative example of a boundary. The light blue arrowhead indicates the boundary location. **d**, Interaction pileup for the promoter/TSS category (10,000 sites), as described in **c**. **e**, Dot pileup aggregation plots for pairwise combinations of active TSSs (without CTCF) and CTCF peaks (without TSSs, with RAD21) separated by 50–500 kb, without/with auxin, for a 100-kb window at 2-kb resolution. Left: all pairwise interactions. Middle: with orientation. Right: CTCF (upstream or downstream)-TSS pairwise interactions, plotted with their quantification (mean of the five central bins at the CTCF site): nearest analysis. CTCF (upstream or downstream)-TSS pairwise interactions are plotted without any CTCF peaks or TSSs in between them, with their quantification (mean of the five central bins at the CTCF site). The red arrows indicate the peak of interactions between the CTCF site and TSS. The black arrows indicate the CTCF motif, and the direction of the arrow shows the motif orientation. The double-headed arrows indicate the TSS, and their direction shows the TSS orientation.

https://doi.org/10.1038/s41594-022-00890-9

In Hi-C interaction maps, lines of enriched interactions were visible from the distal CTCF sites towards the active TSSs. No such lines were detected anchored on TSSs. When we quantified the strength of this enrichment along CTCF-anchored stripes, we observed a peak in interactions centered on the TSSs (Fig. 1e). All these features disappeared when CTCF was depleted. We interpreted these results as follows: cohesin actively extrudes chromatin until it is blocked on one side by CTCF while continuing to extrude on the other side towards an active



promoter/TSS. When it reaches the active promoter/TSS, extrusion pauses and results in a local enrichment of CTCF-promoter/TSS interactions. Cohesin can subsequently occasionally extrude beyond the active promoter/TSS, leading to continuation of the CTCF-anchored stripe pattern in Hi-C beyond the TSS.

TTSs of active genes are CTCF-independent domain boundaries

By analyzing insulation profiles along genes, we found that active gene transcription termination sites (TTSs) also form domain boundaries. We calculated insulation at active TTSs that do not contain CTCF-bound sites. We detected local minima in the insulation scores, consistent with the presence of boundaries (Fig. 2a). The local insulation minima were less precisely positioned as compared to those located at active promoters/TSSs and CTCF-bound sites, and their detection required calculating insulation scores using a larger genomic window (100 kb instead of 20 kb). Insulation at TTSs was unaffected after depletion of CTCF. Strong insulating TTSs correlated with the presence of R-loops at those locations (Fig. 2a and Extended Data Fig. 3a). Active TTS domain boundaries did not overlap with compartment boundaries (Extended Data Fig. 3b).

We next plotted the average insulation profiles across distal CTCF sites and scaled active genes (Fig. 2b and Extended Data Fig. 3e). For this analysis, we only plotted data for active genes that lack CTCF binding at their promoters/TSSs and TTSs. For the HAP1-CTCF-degron cell lines, we noticed that the gene bodies display higher local interactions, with boundaries at their TSSs and TTSs, leading to the formation of gene domains. Similar observations have been reported in *Drosophila*[29]. Interestingly, this analysis revealed that depletion of CTCF not only led to insulation loss at CTCF sites, but also to reduced interactions within the active gene bodies, as reflected in a decrease in the insulation score throughout the genes.

TSSs contrary to TTSs are cohesin-independent boundaries

We did not observe RAD21 binding at active TTSs, indicating that boundary formation at these sites might not depend on cohesin (Fig. 2a). To directly determine this, we used publicly available Hi-C data obtained from RAD21-depleted HCT116 cells (HCT116-RAD21-AID)[12,30]. We observed that local insulation at active TTSs in HCT116-RAD21-AID cells was nearly lost after RAD21 depletion (Fig. 2a and Extended Data Fig. 3c). Therefore, insulation at these sites does depend on cohesin. We next analyzed the insulation profiles for these cells, as above. We found that boundary formation at active TTSs and at CTCF-bound sites both depend on cohesin (Fig. 2a,b and Extended Data Fig. 3c). However, insulation, and thus boundary formation, at TSSs was still observed even after depleting RAD21 (Fig. 2a,b and Extended Data Fig. 3c). One possible explanation is that boundary formation at TSSs is independent of cohesin. Alternately, the small amount of RAD21 remaining at TSSs after auxin treatment may be sufficient for insulation (Extended Data Fig. 3d).

Cohesin stalling and unloading at TTS boundaries

Previous studies have shown that, upon depletions of the cohesin unloader WAPL and CTCF, cohesin accumulates at the 3′ ends of active genes, implying that in wild-type cells cohesin is stalled and unloaded at TTSs[11]. To determine whether the insulation at these sites results from cohesin stalling, unloading or a combination of both, we re-analyzed Hi-C data from WAPL-depleted HAP1 cells[31]. As described previously, removing WAPL increases insulation at CTCF sites. However, we found that removing WAPL did not abolish insulation at active TTSs (Fig. 2a,b and Extended Data Fig. 3c). We conclude that insulation at active TTSs in normal cells is not simply the result of efficient cohesin unloading. Combined, our data support a model where TTSs act as sites where cohesin stalls, leading to boundary formation. Cohesin is then rapidly removed by WAPL, leading to no detectable RAD21 by ChIP-seq in normal cells.

RNA polymerase II depletion effect on chromatin boundaries

We also explored whether active transcription is necessary for boundary formation at active promoters/TSSs and TTSs. We generated a HAP1 cell line that expresses an auxin-inducible degron tagged RNA polymerase II (RNA polII) subunit, RPB1, from its endogenous promoter (HAP1-RPB1-AID). In the presence of auxin, RPB1 was efficiently depleted within 4 h. The cell-cycle profile was not altered, but cells stop growing within hours (Extended Data Fig. 4a,b). We performed Hi-C in HAP1-RPB1-AID cells without or with 4 h of auxin treatment. Hi-C interaction frequency as a function of genomic distance between loci, compartmentalization, TAD boundaries and CTCF-CTCF looping interactions only changed slightly after removal of RNA polII (Extended Data Fig. 4c–e). Disappearance of a small subset of TAD boundaries after RNA polII depletion was probably due to high levels of RPB1 accumulated at a few sites that may result in loop extrusion blocking (Extended Data Fig. 4e).

Relative to HAP1-WT cells, HAP1-RPB1-AID cells displayed weaker A and B compartments (Extended Data Fig. 4d). This may be due to lower levels of RNA polII in HAP1-RPB1-AID cells relative to the HAP1-WT cells (Extended Data Fig. 4b). Insulation profiles across CTCF sites were indistinguishable between RPB1-depleted cells and control cells. Interestingly, insulation at TSSs was unaffected by RBP1 depletion (Fig. 2b and Extended Data Fig. 3c). Insulation at active TTSs was reduced after RNA polII depletion, especially at TTSs containing R-loops (Fig. 2a,b and Extended Data Fig. 4f).

Cohesin traffic constrains promoter-enhancer interactions

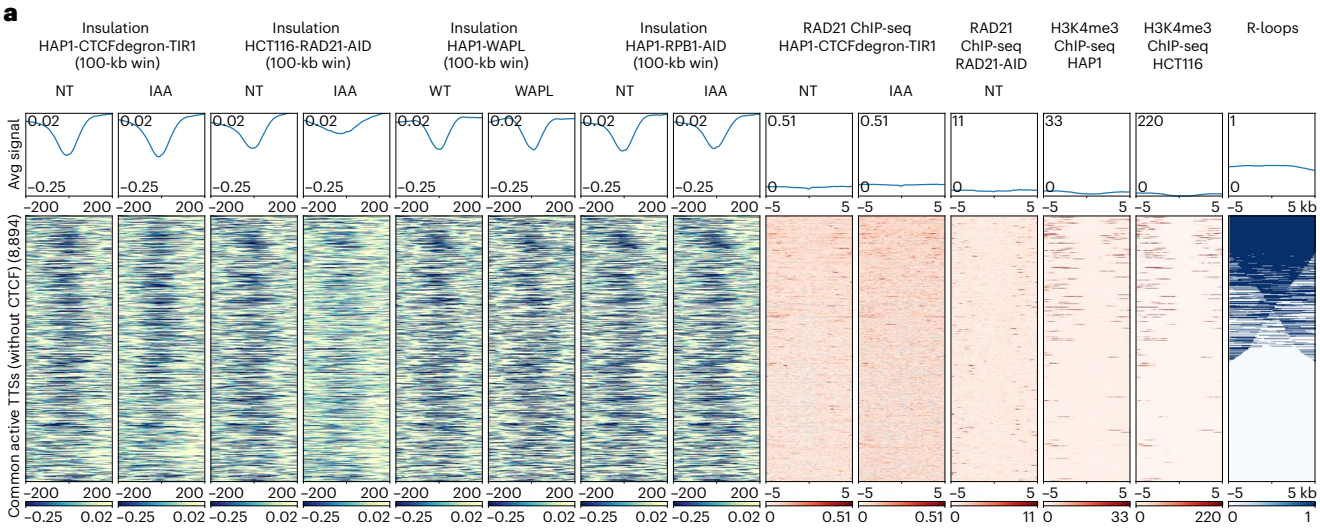
Our data provide a comprehensive view of chromatin boundaries and their relationship to cohesin movement on the chromatin creating a cohesin traffic pattern modulated by CTCF and RNA polII (Fig. 2c).

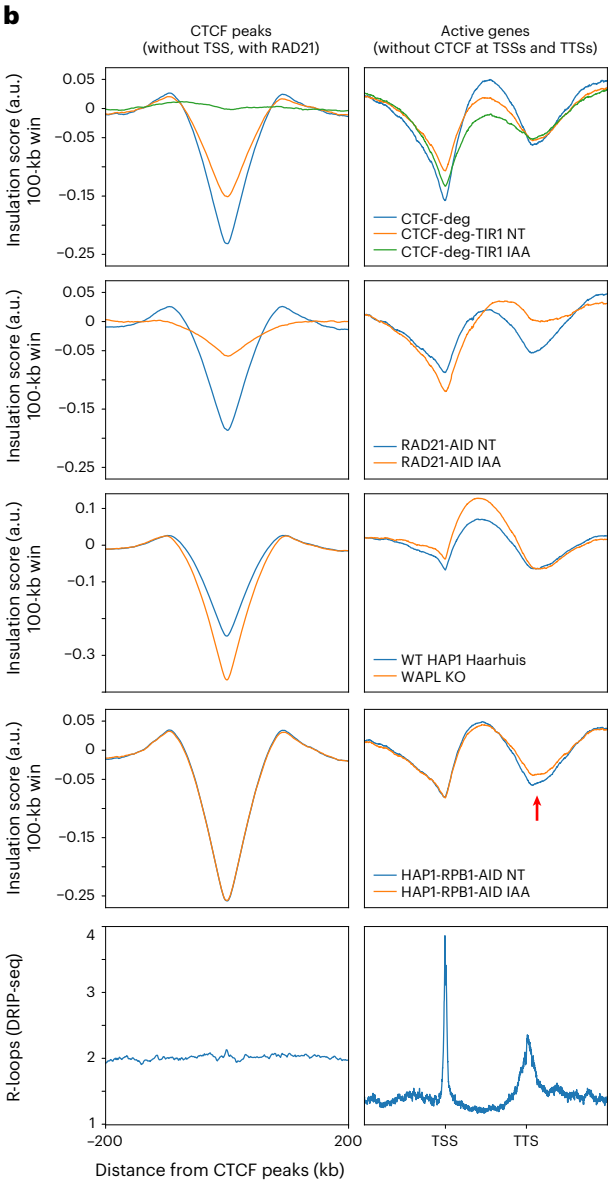
We next investigated how the altered cohesin traffic pattern after CTCF depletion affects promoter-enhancer interactions. For this analysis, we defined enhancers as sites that are DNaseI hypersensitive,

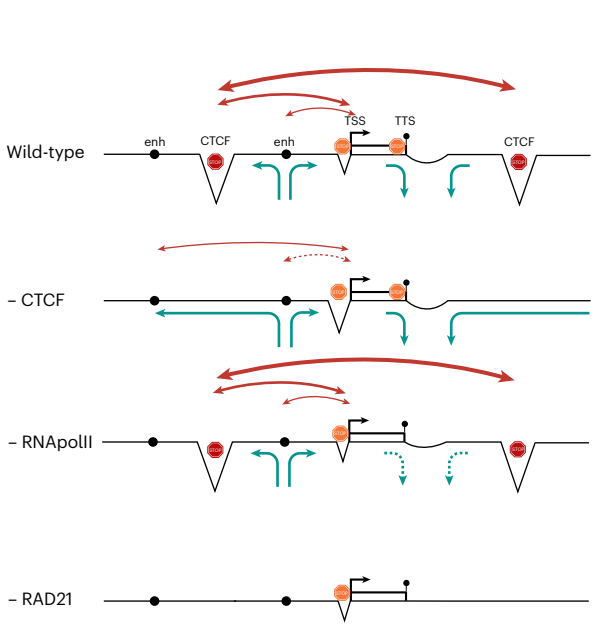
Fig. 2 | Active TTSs form chromatin boundaries, and a cohesin traffic pattern defines promoter-enhancer interactions. a, Stackups for active TTSs (without the H3K4me3 active TSS histone mark and CTCF peaks, common between HAP1 and HCT116 cell lines) sorted on the consensus list of R-loops. Insulation in HAP1-CTCF-degron-TIR1, HCT116-RAD21-AID[12] and HAP1-RPB1-AID cells without/with auxin (NT and IAA), and insulation in WT HAP1 and WAPL knockouts[31], are plotted along with RAD21 ChIP-seq signals in HAP1-CTCF-degron-TIR1 and HCT116-RAD21-AID[12], H3K4me3 in HAP1[26] and HCT116 cells, together with the consensus list of R-loop signals. Stackups are flipped according to the orientation of the genes, to have the gene body on the left of the TTSs. **b**, Average insulation profiles across CTCF peaks (without TSSs, with RAD21; left plots) and across scaled active genes without CTCF at TSSs and TTSs (right plots) at 5-kb resolution for HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 without/with auxin (first row), for the published HCT116-RAD21-AID degron without/with auxin (second row)[12], for the published WT HAP1 and WAPL knockout (third row)[31], for the HAP1-RPB1-AID without/with auxin (fourth row) and for the published R-loop DNA:RNA immunoprecipitation sequencing (DRIP-seq) in K562 (fifth row)[67]. The red arrow indicates the reduced insulation at TTSs after RNA polymerase II depletion. a.u.; arbitrary unit. **c**, Model for a cohesin traffic pattern. The cohesin traffic pattern (green arrows) in four cell lines (wild-type, depleted for CTCF (−CTCF), depleted for RNA polII (−RNA polII) and depleted for RAD21 (−RAD21)) is shown for a defined chromosomal locus. Dips in the line represent the three types of chromatin boundary we identified across the locus: at CTCF sites and at active TSSs and TTSs. In this model, cohesin is loaded at enhancers (black circles), is blocked at CTCF sites, is blocked/paused at a TSS (black arrow) and is blocked and unloaded at a TTS (black circle on a stick). This pattern of cohesin dynamics results in promoter-enhancer interactions, CTCF site-promoter interactions and CTCF-CTCF site interactions (red arrows). CTCF depletion redefines the cohesin traffic, resulting in rewired promoter-enhancer interactions. RNA polII depletion mostly affects the insulation and the cohesin trafficking at TTSs (dotted green arrows). RAD21 depletion abolishes the cohesin trafficking and only keeps the insulation at TSSs, which is mostly cohesin-independent.

https://doi.org/10.1038/s41594-022-00890-9

enriched in H3K27Ac, but not TSSs or CTCF-bound sites. We aggregated Hi-C data for all pairwise combinations of active TSSs and enhancers. We split the set of enhancer-promoter pairs into two groups: those separated by a CTCF-bound site and those without an intervening CTCF-bound site. We also analyzed enhancers located up- and downstream of the TSS separately. Finally, we examined the effects of







https://doi.org/10.1038/s41594-022-00890-9

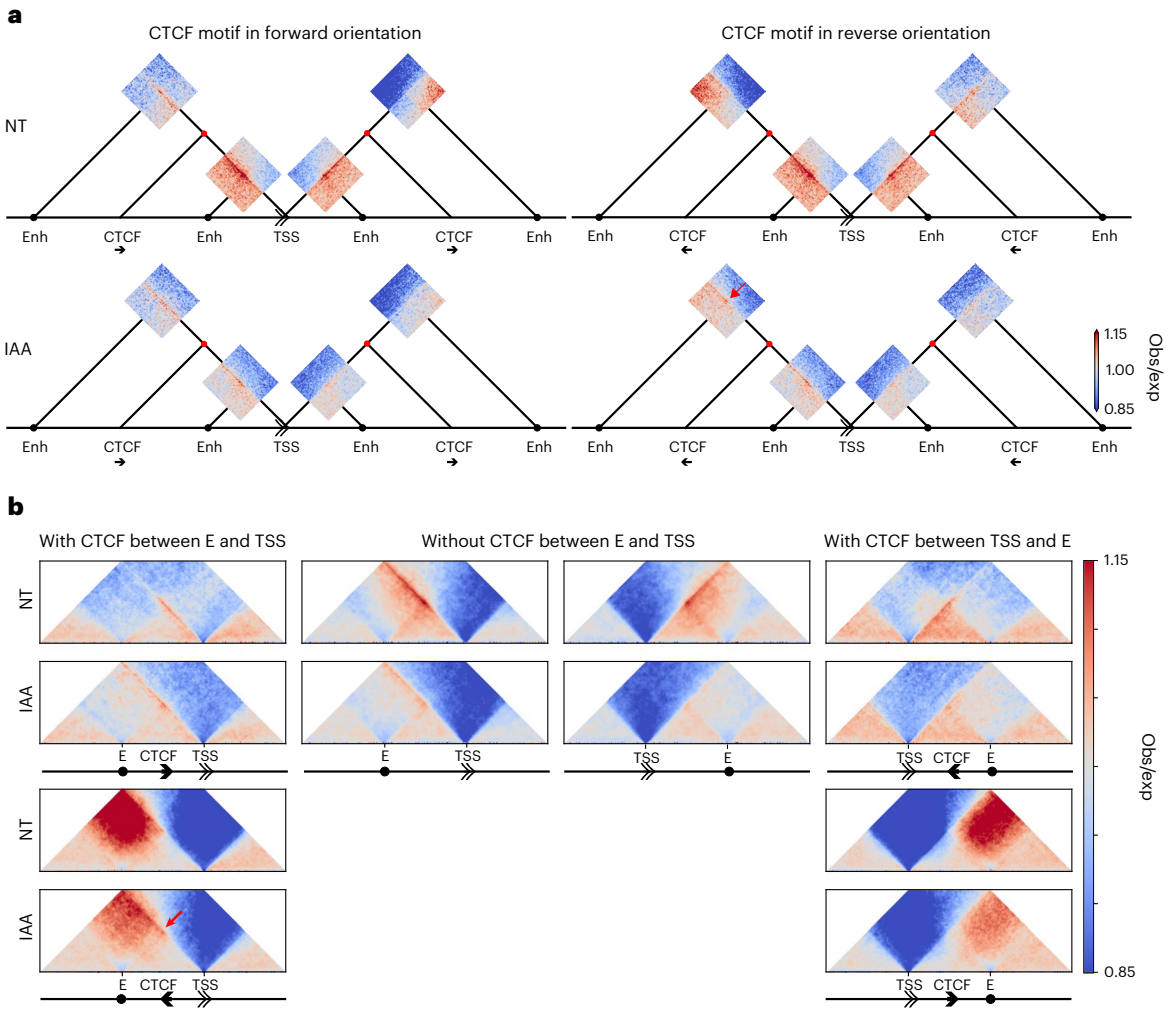


Fig. 3 | Rewiring of enhancer-promoter interactions after CTCF depletion. **a**, Dot pileups for all pairwise combinations between active TSSs (without CTCF) and enhancers (without CTCF or TSSs) separated by 50–500 kb, placed on a schematic representing the different interactions. Dot pileups are separated into upstream or downstream of the TSS and having or not having a CTCF peak with RAD21 in between the TSS and the enhancer. The promoter-enhancer interactions that had a CTCF peak with RAD21 in between them were further classified based on the orientation of the CTCF motif (four different combinations). The dots are aggregated at the center of a 100-kb window at 2-kb resolution. **b**, Aggregated and scaled Hi-C maps for enhancer (without CTCF or TSS)-TSS/promoter (without CTCF) pairs, with or without intervening CTCF peaks with RAD21 for HAP1-CTCFdegron-TIR1 cells without/with auxin (NT and IAA), at 5-kb resolution. The double-headed arrows represent the TSS and its orientation, the black arrows the CTCF peak, and the black circles the enhancer. The red dots represent interactions between CTCF and the TSS. The red arrow highlights the interaction gain after CTCF removal between enhancers and TSSs separated by CTCF sites in the reverse orientation.

the orientation of the CTCF sites located in between promoters and enhancers (Fig. 3a,b).

In cells expressing CTCF, we detected enriched interactions between promoters and enhancers only for those pairs that had no CTCF located between them (Fig. 3a,b). After CTCF depletion, enhancer-promoter interactions were rewired: interactions of promoters with upstream distal enhancers located on the other side of CTCF sites pointing towards the enhancer increased, whereas interactions between promoters and enhancers separated by CTCF sites pointing towards the TSS or not separated by any CTCF sites decreased (Fig. 3a,b). This rewiring is expected when CTCF acts as an insulator, possibly by blocking cohesin-mediated loop extrusion. The CTCF-orientation dependence suggests that these interactions are (1) mediated through cohesin-dependent loop extrusion and (2) that cohesin is extruding from a distal upstream location, for example, the enhancers, towards the TSS. Interestingly, we note that interactions with downstream enhancers are not as prominent as interactions with enhancers located upstream of the TSS, as has been observed in analyses of targeted gene sets[32].

CTCF and RNA processing proteins are genetically linked

The functions of the complex cohesin traffic pattern are not well characterized. We hypothesized that cells in which the cohesin traffic pattern is altered, for example, through CTCF depletion, would be particularly sensitive to genetic perturbations of functions that depend on these phenomena. To test this, we used genome-wide CRISPR screens based on cell proliferation. We performed these screens in HAP1-CTCFdegron-TIR1 cells expressing different levels of CTCF, then compared the results with similar screens performed in HAP1-CTCF-degron cells expressing higher levels of CTCF (Fig. 4a and Extended Data Fig. 5a). Under these conditions, cell proliferation was only slightly reduced when CTCF was depleted (Extended Data Fig. 5b). Possibly, the remaining levels of CTCF were sufficient for growth, and/or auxin resistance emerged. We sequenced the pool of guide RNAs (sgRNAs) in HAP1-CTCFdegron-TIR1 and HAP1-CTCFdegron cell populations grown in the absence or presence of auxin and identified sgRNAs that became depleted or enriched in HAP1-CTCFdegron-TIR1 cells with or without auxin relative to HAP1-CTCFdegron cells. As expected, we

https://doi.org/10.1038/s41594-022-00890-9

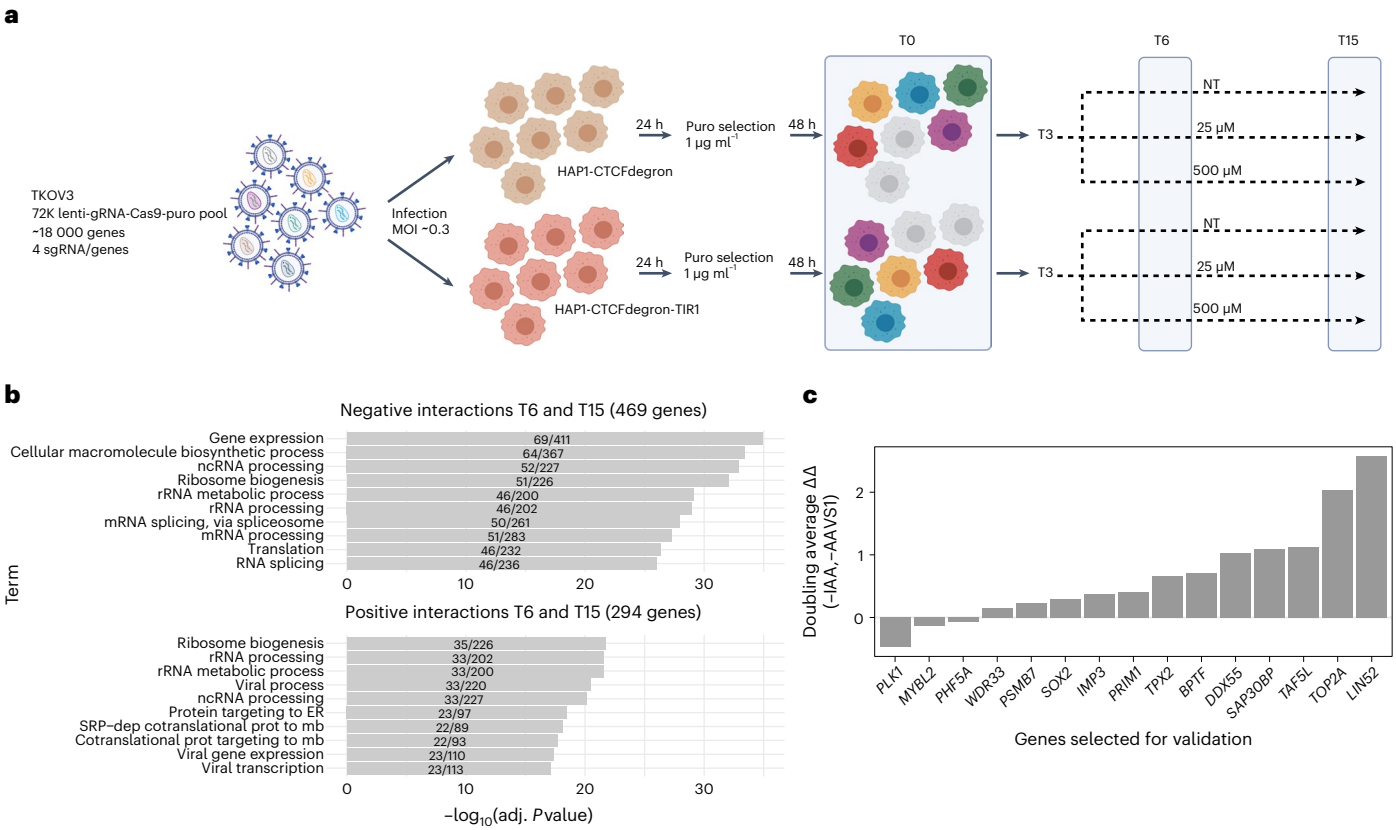


Fig. 4 | The cohesin traffic pattern is genetically linked to gene regulation factors. a, Schematic of the CRISPR screen workflow for two cell lines, HAP1-CTCFdegron (brown cells) and HAP1-CTCFdegron-TIR1 (red cells), with three auxin conditions (NT, 25 µM and 500 µM). Colored cells represent lentivirus infections and knockouts with different sgRNAs. MOI: multiplicity of infection. On the represented time points (T0, T6 and T15), cells were collected for genomic DNA extraction, PCR amplification, library preparation and next-generation sequencing. **b**, Functional enrichment analysis for negative and positive interactions identified in the screen. ER: endoplasmic reticulum. mb: membrane. SRP: signal recognition particle. **c**, Validation for a selection of genes from the gene hits. Cells with proliferation defects when CTCF and the gene hit are depleted have positive values (ΔΔ). Cells with better proliferation when CTCF and the gene hit are depleted have negative values (ΔΔ). The auxin concentration used was 25 µM. **d**, Heatmaps representing $\log_2$(fold change) ($\log_2$(FC) normalized by the HAP1-CTCFdegron (NT, 25 µM or 500 µM IAA) for DEAD/H-box family helicases, SAGA/TAF proteins, RNA polII subunits and mediator subunits with decreasing amounts of CTCF (HAP1-CTCFdegron-TIR1 without auxin (NT) and with auxin (25 µM and 500 µM IAA)). Genes indicated in red have a $\log_2$(FC) of >|0.4|.

found that sgRNAs targeting essential genes disappeared progressively over time, but most non-essential genes did not change. Our screens recovered gold-standard essential gene sets (Extended Data Fig. 5c)[33–36].

We identified a set of 469 genes whose loss reduced proliferation, as well as a set of 294 genes whose loss increased proliferation upon CTCF depletion, including genes known to be involved in CTCF-related processes, such as *SMC1A*, *Topoisomerase II*, *BPTF* and *LIN52* (Extended Data Fig. 5d,e)[37–43].

Gene ontology (GO) analysis of our screen results showed enrichment for genes involved in gene expression and RNA processing (Fig. 4b). We selected 14 hits that decreased proliferation with a broad spectrum of functions for validation, and included one hit that increased proliferation (*PLK1*). We knocked out these genes using two sgRNAs from the screens, and validated the reduced proliferation in CTCF-depleted cells for ten of them, as well as faster proliferation for *PLK1*. Four hits did not validate in this assay (Fig. 4c).

https://doi.org/10.1038/s41594-022-00890-9

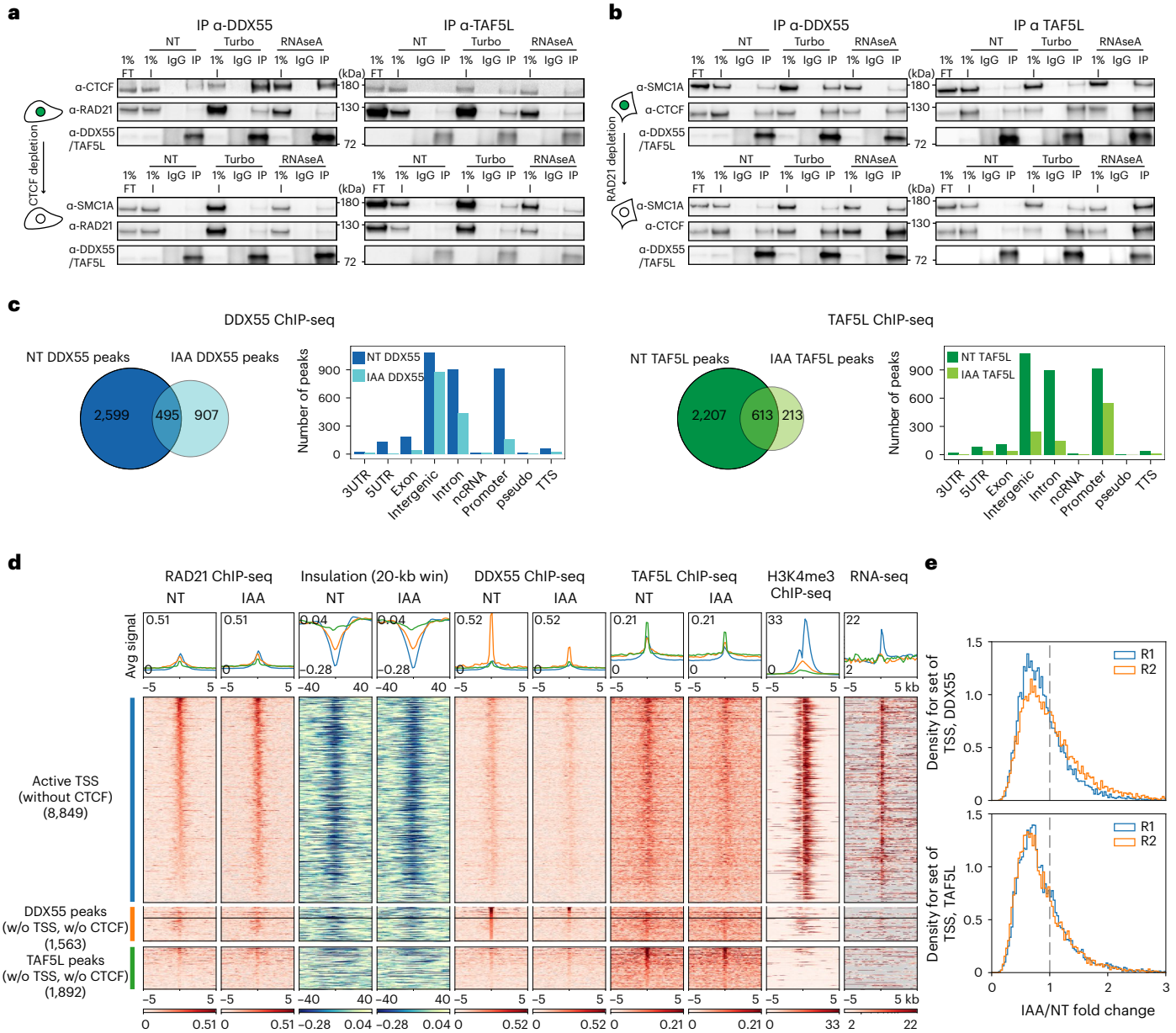


Fig. 5 | Altered cohesin traffic pattern following CTCF depletion reduces chromatin binding of DDX55 and TAF5L at active promoters/TSSs. a, Western blot co-IPs against DDX55 and TAF5L in HAP1-CTCFdegron-TIR1 without/with auxin (NT and IAA), treated with either turbonuclease (DNA− and RNA−) or RNAseA (RNA−). FT: flow through. I: input. **b**, Western blot co-IPs against DDX55 and TAF5L in RAD21-AID degron without/with auxin (NT and IAA), treated with either turbonuclease (DNA− and RNA−) or RNAseA (RNA−). **c**, Venn diagram for DDX55 (left) and TAF5L (right) ChIP-seq peaks called in HAP1-CTCFdegron-TIR1 cells without/with auxin (NT and IAA), and gene annotation bar plots for the ChIP-seq peaks called in HAP1-CTCFdegron-TIR1 without/with auxin (NT and IAA). **d**, Stackups for active TSSs without CTCF (blue) sorted on the NT RAD21 ChIP-seq signal, for DDX55 peaks (without TSSs, without CTCF; orange) sorted on NT DDX55 ChIP-seq, and for TAF5L peaks (without TSSs, without CTCF; green) sorted on NT TAF5L ChIP-seq. RAD21 ChIP-seq, calculated insulation, DDX55 ChIP-seq, TAF5L ChIP-seq and RNA-seq signals in HAP1-CTCFdegron-TIR1 cells without/with auxin (NT and IAA) are plotted along with published HAP1 H3K4me3 ChIP-seq[26]. For TSSs, genes were flipped according to their orientations to have the gene body on the right of the TSSs. Genes were not flipped for the DDX55 and TAF5L peaks. **e**, Stackup quantification for the active TSSs (without CTCF) for DDX55 and TAF5L ChIP-seq (two replicates). The distribution of the ratios (fold change) of a given signal between auxin-treated and NT conditions is shown. A fold change of <1 represents less binding of DDX55 or TAF5L at active TSSs after CTCF depletion.

Among the categories that were statistically significantly enriched, the family of DEAD/H-box helicase genes were of particular interest given that one of them (*DDX5*) had already been implicated in CTCF function[44,45]. Depletion of more than two-thirds of the studied DEAD/H-box helicases in our screens displayed proliferation effects in cells expressing lower levels of CTCF (36/50; Fig. 4d).

The screens also identified several genes involved in transcriptional regulation including subunits of the SAGA complex and TATA-box-binding protein (TBP)-associated factors (TAFs), RNA polymerase II and Mediator complexes (Fig. 4d)[46]. Mediator complexes have already been shown to be involved in cohesin-mediated interactions[47–49].

DDX55 and TAF5L physically interact with CTCF and cohesin

The results of our genome-wide screens suggest that cells with altered cohesin traffic pattern are vulnerable to defects in machineries associated with RNA processing and transcription initiation. We selected two hits for further analysis: DDX55 (a DEAD/H-box protein) and TAF5L (a subunit of the SAGA complex).

To determine whether these proteins physically associate with CTCF and/or cohesin, we performed co-immunoprecipitations (co-IP). We found that DDX55 and TAF5L both interacted with CTCF and cohesin. This interaction was not DNA- or RNA-dependent (Fig. 5a and Extended Data Fig. 6a,d,e,f). Given that CTCF and cohesin interact with each other, we next wanted to determine whether DDX55 and TAF5L require CTCF to bind to cohesin. After CTCF depletion, we performed co-IP as above and found that DDX55 and TAF5L still interacted with the cohesin complex (Fig. 5a and Extended Data Fig. 6a,d,e,f). We also performed co-IP against TAF6L, another SAGA subunit, and obtained similar results (Extended Data Fig. 6c–f). We used the HCT116-RAD21-AID cell line to determine whether the interaction of DDX55 or TAF5L with CTCF was dependent on cohesin. We found that interactions between DDX55, TAF5L and CTCF were not affected by degradation of RAD21. We conclude that DDX55 and TAF5L interact with cohesin and with CTCF independently (Fig. 5b and Extended Data Fig. 6b,d–f).

Altered cohesin traffic affects protein chromatin binding

We next wanted to assess whether chromatin binding and localization of DDX55 and TAF5L are affected by CTCF depletion. We performed DDX55 and TAF5L ChIP-seq in HAP1-CTCFdegron-TIR1 cells (Extended Data Figs. 7a–c and 8a–c). We identified 3,094 DDX55 and 2,820 TAF5L peaks, mostly at TSSs, intron and intergenic regions, many of which decreased after CTCF depletion (Fig. 5c).

Next, we determined DDX55 and TAF5L levels at active TSSs and TTSs and at CTCF sites. We detected DDX55 and TAF5L at TSSs, but very little of either protein was observed at CTCF sites and none at TTSs. Visual inspection of the ChIP-seq data suggested that, after CTCF depletion, the levels of DDX55 and TAF5L binding to TSSs and CTCF sites are reduced (Fig. 5d and Extended Data Figs. 6g, 7a–c and 8a–c). We quantified this observation by calculating the ratio of DDX55 or TAF5L levels at CTCF sites and TSSs between control cells and CTCF-depleted cells (Fig. 5e and Extended Data Fig. 6h). This ratio was found to be mostly below 1, in two independent ChIP-seq replicates, suggesting that DDX55 and TAF5L accumulation at CTCF sites and TSSs is CTCF-dependent. We noticed that the accumulation of DDX55 and TAF5L at sites that displayed DDX55 or TAF5L peaks, but did not overlap with CTCF peaks or TSSs, was also reduced after CTCF depletion (Fig. 5d).

DDX55 or TAF5L depletion modestly affects chromosome folding

We next asked whether DDX55 and TAF5L function in chromosome folding. We depleted the DDX55 and TAF5L proteins in HAP1-CTCFdegron-TIR1 cells in two ways, by using a pool of siRNAs and generating knockout clones with sgRNAs from the CRISPR screens (Extended Data Fig. 9a,b). We could only generate heterozygous knockout clones for DDX55, because DDX55 is essential[50], but succeeded in generating homozygous TAF5L knockout clones (clones). DDX55 and TAF5L depletions with siRNA or in clones did not affect the cell cycle (Extended Data Fig. 9c). We noticed that depleting DDX55 and TAF5L did not affect gene expression for most of the components of the loop extrusion machinery; however, the DDX55 and TAF5L clones showed CTCF misregulation (Extended Data Fig. 9b). We then performed Hi-C on the DDX55- and TAF5L-depleted cell lines in the presence or absence of CTCF. Depletion of DDX55 or TAF5L had only minor global effects on the Hi-C data (Fig. 6a and Extended Data Fig. 10a,b). To examine the effects on local chromatin conformation, we plotted aggregated interactions and average insulation profiles across distal CTCF sites and active and inactive genes (Fig. 6b). Insulation at CTCF sites, active TSSs or TTSs did not require DDX55 or TAF5L. When CTCF was co-depleted with DDX55 or TAF5L, insulation at CTCF sites was lost, as expected, whereas insulation at TSSs and TTSs was largely unaffected. Similar to what we observed in cells expressing normal levels of DDX55 and TAF5L, CTCF depletion decreased intragenic interactions (Fig. 2b). Therefore, the effects of CTCF depletion on intragenic interaction frequencies were independent of DDX55 and TAF5L levels. Interestingly, depletion of DDX55 and TAF5L changed the conformation of active genes. Intragenic interactions increased, similar to what we observed in WAPL-depleted cells (Fig. 2b). DDX55 or TAF5L depletions in the absence of CTCF resulted in increased intragenic interactions, similar to that observed in the presence of CTCF, suggesting that the effects of DDX55 or TAF5L depletion on intragenic interactions were independent of CTCF levels. We conclude that DDX55 and TAF5L are not required for chromatin domain boundary formation, but that CTCF, DDX55 and TAF5L independently influence the conformation of active genes.

CTCF, DDX55 and TAF5L depletion and gene expression

Finally, we assessed global gene expression after CTCF, DDX55 and TAF5L depletions by RNA-seq. Confirming previous results, CTCF depletion did not result in massive changes in gene expression (~1,300 differentially expressed genes; Extended Data Fig. 10c)[8,23]. However, the number of differentially expressed genes increased with the extent of CTCF depletion. DDX55 and TAF5L depletions modestly affected the number of differentially expressed genes but the double depletions of CTCF and DDX55 or TAF5L resulted in synergistic effects with more changes in gene expression (~600 genes; Extended Data Fig. 10c). Depleting CTCF, DDX55 or TAF5L also resulted in differential splicing of a gene set that was different from the gene set that was differentially expressed. The number of differentially spliced genes slightly increased with the double depletions (Extended Data Fig. 10d).

Discussion

Through analysis of Hi-C data obtained with cells where CTCF, RAD21, WAPL or RNA polII were rapidly depleted, we have described a complex pattern of cohesin traffic defined by different types of *cis*-element where cohesin is loaded, paused, blocked or unloaded. Cohesin may be loaded at sites distal from promoters (possibly enhancers[51]), weakly paused or blocked at active TSSs, efficiently blocked and stalled at CTCF sites, and stalled and rapidly unloaded at active TTSs (Fig. 2c). Our genome-wide genetic interaction screens in cells, with altered extrusion patterns as a result of CTCF depletion, identified genes involved in transcription initiation and RNA processing. Based on these findings, we hypothesize that the cohesin traffic pattern is functionally linked to gene control.

Three types of boundary define a cohesin traffic pattern

We describe and characterize three distinct types of domain boundary: CTCF sites and active TSSs (previously described by Bonev and colleagues[27]) and TTSs. The three elements differ in the mechanism by which they drive boundary formation.

Active TSSs display relatively strong, but highly localized insulation that is quantitatively comparable to the observed insulation at CTCF-bound sites. However, although rapid depletion of RAD21 leads to near complete loss of insulation at CTCF-bound sites, insulation at TSSs is hardly affected. Insulation at active TSSs might result from a very small amount of RAD21 binding still detected after RAD21 depletion. Alternately, insulation at active TSSs may be truly cohesin-independent: it may be driven by other loop extrusion factors or result from entirely different mechanisms, for example, specific local chromatin features that can induce chromatin domain boundary formation via yet-to-be-established processes.

At active TTSs, insulation is not affected by CTCF depletion and is quantitatively distinct from that observed at CTCF sites and TSSs: it is weaker and forms a broad zone of insulation. Intriguingly, we did not detect RAD21 at active TTSs by ChIP-seq, but insulation at TTSs is lost when RAD21 is depleted. Previous studies had shown that, upon CTCF and WAPL depletions, cohesin accumulates at 3′ ends of active genes, especially at sites of convergent transcription[11]. Moreover, insulation at active TTSs is not affected by WAPL depletion. We conclude from these observations that active TTSs are sites where, in normal cells, cohesin is

https://doi.org/10.1038/s41594-022-00890-9

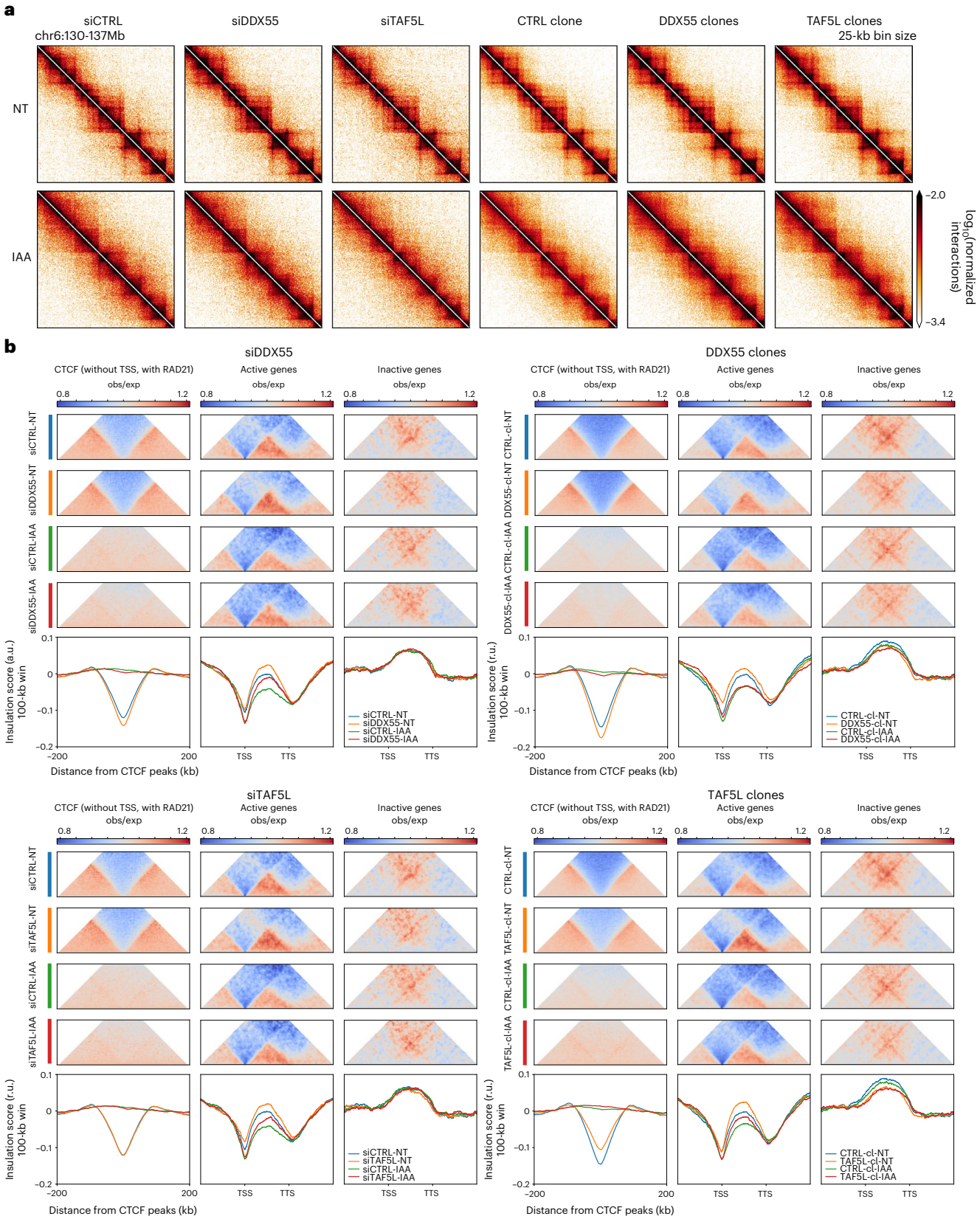


Fig. 6 | DDX55 and TAF5L depletions alter the conformation of active genes independently of CTCF. a, Hi-C contact heatmaps at 25-kb resolution for a 7-Mb region on chromosome 6 for HAP1-CTCFdegron-TIR1 cells depleted for DDX55 or TAF5L (siRNA and knockout clones) without/with CTCF (NT and IAA). **b**, Average insulation profiles across distal CTCF (without TSSs, with RAD21; left plots), across scaled active genes (without CTCF at TSSs and TTSs; middle plots) and across scaled inactive genes (without CTCF at TSSs and TTSs; right plots) for the HAP1-CTCFdegron-TIR1 without/with auxin (NT and IAA) and with DDX55 or TAF5L depleted (siRNA and clones), at 5-kb resolution. The representative scaled interaction pileups are plotted on top of the average insulation profiles. a.u.; arbitrary unit.

first blocked, leading to boundary formation and insulation, and then is unloaded by WAPL. The lack of RAD21 can be explained if unloading is fast and efficient. Insulation at active TTSs is partly dependent on RNA polII: depleting RNA polII results in weaker insulation at active TTSs. One hypothesis is that depletion of RNA polII may destabilize R-loops, which could induce local chromatin changes around TTSs, resulting in less stalling of cohesin at active TTSs, thus reducing insulation. Indeed, Busslinger and colleagues found that blocking transcription elongation using DRB in CTCF WAPL double knockout cells results in less accumulation of cohesin at TTSs[11]. An alternative hypothesis is that cohesin is pushed through the gene by RNA polII towards the TTS where it is first blocked and then rapidly unloaded. In support of this model, previous studies have shown that condensin and RNA polII can interplay and that, in yeast, cohesin could be pushed by the transcription machinery[52–54]. A role for RNA polII in cohesin positioning along active genes has also been proposed by Banigan and co-workers[55].

Enhancer-promoter interactions are directed by CTCF site orientation in a way that suggests that cohesin could be loaded at enhancers and extrude towards the promoter. This complex and dynamic cohesin traffic pattern may be important for appropriate gene regulation, for example, through recruiting and then delivering transcription-related complexes to target genes. A similar model for cohesin dynamics has been proposed by Liu and colleagues based on the analysis of cells where either WAPL or RAD21 is depleted[56].

Possible functions for the cohesin traffic pattern

Through a genetic interaction screen, we identified factors that, upon deletion, changed the growth rate of cells only when CTCF levels were low and cohesin positioning along chromosomes was altered. We identified several classes of gene involved in RNA metabolism. Among these were many DEAD/H-box-containing RNA helicases[57]. In previous studies, the DEAD/H-box helicase DDX5 and its associated RNA activator RSA were found to interact with CTCF and cohesin and to be required for insulator function, possibly by reducing cohesin localization at CTCF sites[45]. In *Drosophila*, the DDX5 ortholog Rm62 plays a role in modulating the activity of the insulator binding factor CP190[44]. We also identified a set of proteins that function in transcription initiation, including TAFs, that are part of the SAGA, TFIID and RNA polII complexes. A previous study has shown that TAF3, which is part of the core promoter recognition complex TFIID, is recruited by CTCF to promoters and mediates the looping interaction between promoter and TSS[58].

We focused on DDX55 and TAF5L for further analysis, but found that they do not appear to play a major role in chromatin folding. Therefore, their function may depend on correctly folded chromatin without playing a direct role in chromosome organization themselves. Interestingly, we found that CTCF depletion leads to reduced accumulation of DDX55 and TAF5L at both CTCF-bound and active TSSs. This observation points to an indirect role for CTCF in recruiting and positioning these factors and possibly other transcription-related complexes to distal active genes, most likely through cohesin-mediated mechanisms. DDX55 and TAF5L may be recruited to distal CTCF sites and then transported to TSSs through cohesin action. Consistent with this model, we found that DDX55 and TAF5L physically interact with both CTCF and cohesin.

Depletion of DDX55 or TAF5L, in the presence or absence of CTCF, did not result in major changes in gene expression and splicing, consistent with previous findings[8,12,23,59–64]. This may be due to redundancy with other related complexes. Alternately, acute depletion of factors that mediate enhancer-driven activation may not have a noticeable effect on transcription until many hours, or even cell cycles later, as recent analyses suggest that the transcriptional state of a TSS can be relatively long-lived[65,66].

In summary, our work delineates roles for CTCF, cohesin, WAPL and RNA polII in defining a cohesin traffic pattern constrained by different types of domain boundary at key *cis*-elements. Defects in setting up this cohesin traffic pattern correctly make cells sensitive to loss of factors involved in RNA metabolism. We propose that the complex pattern of cohesin movement along chromatin, and the roles of CTCF, WAPL and RNA polII in defining this pattern, contribute to appropriate localization of transcription and RNA-processing factors to active genes. How these phenomena control gene expression remains an open question.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-022-00890-9.

References

1. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
2. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
3. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
4. Sexton, T. et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
5. Alipour, E. & Marko, J. F. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res.* **40**, 11202–11212 (2012).
6. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
7. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, E6456–E6465 (2015).
8. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017).
9. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl Acad. Sci. USA* **115**, E6697–E6706 (2018).
10. Wutz, G. et al. Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).
11. Busslinger, G. A. et al. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, 503–507 (2017).
12. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
13. Kubo, N. et al. Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat. Struct. Mol. Biol.* **28**, 152–161 (2021).
14. Hnisz, D. et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).
15. Narendra, V. et al. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* **347**, 1017–1021 (2015).
16. Splinter, E. et al. CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus. *Genes Dev.* **20**, 2349–2354 (2006).
17. de Wit, E. et al. CTCF binding polarity determines chromatin looping. *Mol. Cell* **60**, 676–684 (2015).
18. Flavahan, W. A. et al. Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).

19. Franke, M. et al. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).
20. Lupiáñez, D. G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
21. Lupiáñez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: how alterations of chromatin domains result in disease. *Trends Genet.* **32**, 225–237 (2016).
22. Valton, A.-L. & Dekker, J. TAD disruption as oncogenic driver. *Curr. Opin. Genet. Dev.* **36**, 34–40 (2016).
23. Luan, J. et al. Distinct properties and functions of CTCF revealed by a rapidly inducible degron system. *Cell Rep.* **34**, 108783 (2021).
24. Gassler, J. et al. A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J.* **36**, 3600–3618 (2017).
25. Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
26. Campagne, A. et al. BAP1 complex promotes transcription by opposing PRC1-mediated H2A ubiquitylation. *Nat. Commun.* **10**, 348 (2019).
27. Bonev, B. et al. Multiscale 3D genome rewiring during mouse neural development. *Cell* **171**, 557–572 (2017).
28. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).
29. Rowley, M. J. et al. Condensin II counteracts cohesin and RNA polymerase II in the establishment of 3D chromatin organization. *Cell Rep.* **26**, 2890–2903 (2019).
30. Natsume, T., Kiyomitsu, T., Saga, Y. & Kanemaki, M. T. Rapid protein depletion in human cells by auxin-inducible degron tagging with short homology donors. *Cell Rep.* **15**, 210–218 (2016).
31. Haarhuis, J. H. I. et al. The cohesin release factor WAPL restricts chromatin loop extension. *Cell* **169**, 693–707 (2017).
32. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
33. Hart, T. & Moffat, J. BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics* **17**, 164 (2016).
34. Hart, T., Brown, K. R., Sircoulomb, F., Rottapel, R. & Moffat, J. Measuring error rates in genomic perturbation screens: gold standards for human functional genomics. *Mol. Syst. Biol.* **10**, 733 (2014).
35. Hart, T. et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* **163**, 1515–1526 (2015).
36. Hart, T. et al. Evaluation and design of genome-wide CRISPR/SpCas9 knockout screens. *G3 (Bethesda)* **7**, 2719–2727 (2017).
37. Uusküla-Reimand, L. et al. Topoisomerase II beta interacts with cohesin and CTCF at topological domain borders. *Genome Biol.* **17**, 182 (2016).
38. Barisic, D., Stadler, M. B., Iurlaro, M. & Schübeler, D. Mammalian ISWI and SWI/SNF selectively mediate binding of distinct transcription factors. *Nature* **569**, 136–140 (2019).
39. Wiechens, N. et al. The chromatin remodelling enzymes SNF2H and SNF2L position nucleosomes adjacent to CTCF and other transcription factors. *PLoS Genet.* **12**, e1005940 (2016).
40. Valletta, M. et al. Exploring the interaction between the SWI/SNF chromatin remodeling complex and the zinc finger factor CTCF. *Int. J. Mol. Sci.* **21**, E8950 (2020).
41. Bohla, D. et al. A functional insulator screen identifies NURF and dREAM components to be required for enhancer-blocking. *PLoS ONE* **9**, e107765 (2014).
42. Korenjak, M. et al. dREAM co-operates with insulator-binding proteins and regulates expression at divergently paired genes. *Nucleic Acids Res.* **42**, 8939–8953 (2014).
43. Li, X. et al. Chromatin boundaries require functional collaboration between the hSET1 and NURF complexes. *Blood* **118**, 1386–1394 (2011).
44. Lei, E. P. & Corces, V. G. RNA interference machinery influences the nuclear organization of a chromatin insulator. *Nat. Genet.* **38**, 936–941 (2006).
45. Yao, H. et al. Mediation of CTCF transcriptional insulation by DEAD-box RNA-binding protein p68 and steroid receptor RNA activator SRA. *Genes Dev.* **24**, 2543–2555 (2010).
46. Baker, S. P. & Grant, P. A. The SAGA continues: expanding the cellular role of a transcriptional co-activator complex. *Oncogene* **26**, 5329–5340 (2007).
47. Kagey, M. H. et al. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430–435 (2010).
48. Phillips-Cremins, J. E. et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
49. Ramasamy, S. et al. The Mediator complex regulates enhancer-promoter interactions. Preprint at https://www.biorxiv.org/content/10.1101/2022.06.15.496245v1 (2022)
50. Bartha, I., di Iulio, J., Venter, J. C. & Telenti, A. Human gene essentiality. *Nat. Rev. Genet.* **19**, 51–62 (2018).
51. 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).
52. Brandão, H. B. et al. RNA polymerases as moving barriers to condensin loop extrusion. *Proc. Natl Acad. Sci. USA* **116**, 20489–20499 (2019).
53. Lengronne, A. et al. Cohesin relocation from sites of chromosomal loading to places of convergent transcription. *Nature* **430**, 573–578 (2004).
54. Glynn, E. F. et al. Genome-wide mapping of the cohesin complex in the yeast *Saccharomyces cerevisiae*. *PLoS Biol.* **2**, E259 (2004).
55. Banigan, E. J. et al. Transcription shapes 3D chromatin organization by interacting with loop-extruding cohesin complexes. Preprint at https://www.biorxiv.org/content/10.1101/2022.01.07.475367v1 (2022).
56. Liu, N. Q. et al. WAPL maintains a cohesin loading cycle to preserve cell-type-specific distal gene regulation. *Nat. Genet.* **53**, 100–109 (2021).
57. Cargill, M., Venkataraman, R. & Lee, S. DEAD-Box RNA helicases and genome stability. *Genes* **12**, 1471 (2021).
58. Liu, Z., Scannell, D. R., Eisen, M. B. & Tjian, R. Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* **146**, 720–731 (2011).
59. Schwarzer, W. et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
60. 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).
61. Alharbi, A. B. et al. Ctcf haploinsufficiency mediates intron retention in a tissue-specific manner. *RNA Biol.* **18**, 93–103 (2021).
62. Nanavaty, V. et al. DNA methylation regulates alternative polyadenylation via CTCF and the cohesin complex. *Mol. Cell* **78**, 752–764 (2020).
63. Ruiz-Velasco, M. et al. CTCF-mediated chromatin loops between promoter and gene body regulate alternative splicing across individuals. *Cell Syst.* **5**, 628–637 (2017).

64. Shukla, S. et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**, 74–79 (2011).
65. Xiao, J. Y., Hafner, A. & Boettiger, A. N. How subtle changes in 3D structure can create large changes in transcription. *eLife* **10**, e64320 (2021).
66. Zuin, J. et al. Nonlinear control of transcription through enhancer-promoter interactions. *Nature* **604**, 571–577 (2022).
67. Sanz, L. A. et al. Prevalent, dynamic and conserved R-loop structures associate with specific epigenomic signatures in mammals. *Mol. Cell* **63**, 167–178 (2016).

Methods

Cell culture and cell lines

The human HAP1 cell line was purchased from Horizon Discovery (C859). The wild-type and mutated HAP1 cell lines (HAP1-CTCFdegron, HAP1-CTCFdegron-TIR1, HAP1-RPB1-AID, DDX55 and TAF5L knockout clones) were cultured at 37 °C with 5% CO_2 in IMDM GlutaMAX supplement (Gibco, 31980097) with 10% FBS (Gibco, 16000069) and 1% penicillin-streptomycin (Gibco, 15140122). HCT116-RAD21-AID cells were a gift from M. Kanemaki[30]. They were cultured at 37 °C with 5% CO_2 in McCoy's 5A medium GlutaMAX supplement (Gibco, 36600021) with 10% FBS (Gibco, 16000069) and 1% penicillin-streptomycin (Gibco, 15140122). The HEK293T cell line was obtained from ATCC (CRL-3216) and maintained in DMEM (Gibco, 11995065) with 10% FBS (Gibco, 16000069) and 1% penicillin-streptomycin (Gibco, 15140122). Cell lines were routinely tested for mycoplasma infection and tested negative (MycoAlertTM Mycoplasma Detection Kit, Lonza).

Antibiotic selection treatment

Blasticidin S HCl (10 mg ml^{-1}) was purchased from Thermo Fisher (A1113903), and selection was carried out with 10 µg ml^{-1} blasticidin. Puromycin dihydrochloride (10 mg ml^{-1}) was obtained from Thermo Fisher (A1113803), and selection was done with 1.5 µg ml^{-1} puromycin. Hygromycin B Gold (100 mg ml^{-1}) was purchased from Invivogen (ant-hg-1), and selection was carried out with 450 µg ml^{-1} hygromycin.

Auxin treatment

Auxin (IAA, 3-indoleacetic acid) was purchased from Millipore Sigma (45533-250MG) and dissolved in ethanol. Auxin was directly added to the cell culture plates at the indicated concentrations (25 µM for partial CTCF depletion or 500 µM for total CTCF, RPB1 and RAD21 depletions) and times (HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1, 48 h for asynchronous cells; HAP1-RPB1-AID, 4 h; HCT116-RAD21-AID, 2 h).

siRNA transfections

Pools of siRNAs were purchased from Dharmacon (siGENOME non-targeting siRNA pool #2, SMARTpool: siGENOME DDX55 siRNAl and siGENOME TAF5L siRNA). The siRNAs were resuspended in sterile ultrapure water. Transfections were carried out with lipofectamine (Lipofectamine RNAiMAX transfection reagent, Thermo Fisher Scientific, 13778075) and Opti-MEM (Thermo Fisher Scientific, 31985062) following the manufacturer's recommendations. The final concentration of siRNA used was 40 nM and the incubation time with siRNAs was 72 h. If auxin was added, the medium was removed after 24 h and replaced by auxin-containing medium for the remaining 48 h.

Plasmid construction

Each plasmid was analyzed by Sanger sequencing to confirm successful cloning.

Guide RNA cloning. sgRNAs were cloned in pSpCas9(BB)-2A-Puro (PX459) V2.0 (Feng Zhang laboratory, Addgene 62988). Briefly, the pX459 plasmid was digested with BbsI, and the sgRNA primers were phosphorylated, annealed and ligated into the BbsI linearized backbone following the Feng Zhang laboratory protocol[68].

Endogenous CTCF knockout targeting constructs. To knock out the endogenous CTCF, sgRNAs targeting the promoter and the 3′ untranslated region of the endogenous CTCF gene were cloned (~79-kb deletion).

CTCF cDNA construct, HA-AID-CTCFcDNA-AID-eGFP-blasticidin. The HA-AID-CTCFcDNA-AID-eGFP-blasticidin vector was assembled by Gibson Assembly (NEBuilder HiFi DNA Assembly Master Mix, NEB, E2621L) in pENTR221 kanamycin vector using the following templates: the CAGGS promoter (which contains the cytomegalovirus (CMV) early enhancer element, the promoter region, the first exon and the first intron of the chicken β-ACTIN gene, and the splice acceptor of the rabbit β-GLOBIN gene) was amplified from pEN396-pCAGGS-TIR1-V5-2A-PuroR (gift from E. Nora and B. Bruneau, Addgene 92142), the minimal functional AID tag (aa 71–114) was amplified with forward primer containing HA tag from pEN244-CTCF-AID[71–114]-eGFP-FRT-Blast-FRT (gift from E. Nora and B. Bruneau, Addgene 92140), the CTCF cDNA was amplified from a pCMV6-Entry vector containing CTCF cDNA (Origene, RC202416), the AID-eGFP-2A-bls was amplified from pEN244-CTCF-AID[71–114]-eGFP-FRT-Blast-FRT (gift from E. Nora and B. Bruneau, Addgene 92140), the polyA signal was amplified from pEN396-pCAGGS-TIR1-V5–2A-PuroR (gift from E. Nora and B. Bruneau, Addgene 92142). Amplifications were performed with Q5 High-Fidelity DNA polymerase (NEB, M0491L).

TIR1-hygro construct. The TIR1-hygro vector was assembled by Gibson Assembly (NEBuilder HiFi DNA Assembly Master Mix, NEB, E2621L), replacing the puromycin gene by the hygromycin gene in pEN396-pCAGGS-TIR1-V5-2A-PuroR (gift from E. Nora and B. Bruneau, Addgene 92142).

Endogenous RPB1 targeting constructs. C-terminal. To target the C-terminal part of RPB1, sgRNA targeting the last exon of *RPB1* gene was cloned.

N-terminal. To target the N-terminal part of RPB1, sgRNA targeting the first exon, around the start codon of the *RPB1* gene, was cloned.

AID C-terminal RPB1-AID-eGFP-blasticidin construct. The AID C-terminal RPB1-AID-eGFP-blasticidin vector was assembled by Gibson Assembly (NEBuilder HiFi DNA Assembly Master Mix, NEB, E2621L) in pENTR221 kanamycin vector using the following templates. The 5′ homology arm (1,680 bp) and 3′ homology arm (1,558 bp) were amplified from HAP1 genomic DNA, the minimal functional AID tag (aa 71–114)-eGFP was amplified from pEN244-CTCF-AID[71–114]-eGFP-FRT-Blast-FRT (gift from E. Nora and B. Bruneau, Addgene 92140), the T2A was amplified from pEN396-pCAGGS-TIR1-V5-2A-PuroR (gift from E. Nora and B. Bruneau, Addgene 92142), and the blasticidin resistance gene was amplified from PSF-CMV-BLAST (Sigma-Aldrich, OGS588-5UG). Amplifications were performed with the Q5 High-Fidelity DNA polymerase (NEB, M0491L).

AID N-terminal AID-RPB1 construct. The AID N-terminal AID-RPB1 vector was assembled by Gibson Assembly (NEBuilder HiFi DNA Assembly Master Mix, NEB, E2621L) in pENTR221 kanamycin vector using the following templates. The 5′ homology arm (1,079 bp) and 3′ homology arm (1,077 bp) were amplified from HAP1 genomic DNA and the minimal functional AID tag (aa 71–114)-eGFP was amplified from pEN244-CTCF-AID[71–114]-eGFP-FRT-Blast-FRT (gift from E. Nora and B. Bruneau, Addgene 92140).

AAVS1 (control locus), DDX55 and TAF5L knockout constructs. To create deletions in the AAVS1 locus, primers were designed in the AAVS1 locus. To create DDX55 knockout, the sgRNAs used in the genome-wide CRISPR screens and targeting the second exon of the *DDX55* gene were cloned. To create TAF5L knockout, the sgRNAs used in the genome-wide CRISPR screens and targeting the third exon of the *TAF5L* gene were cloned.

Genome modifications

Plasmids used for transfections were purified using the ZymoPURE II Plasmid Midiprep kit (Zymo Research, D4201). Plasmids were linearized using PvuI-HF (NEB, R3150L). Linearized plasmids were further purified with phenol chloroform extraction and ethanol precipitation. HAP1 cells were transfected using turbofectin (Origene, TF81001) following the manufacturer's recommendations.

The differences between the different construct transfections are described in the following.

HAP1-CTCFdegron-TIR1. A 1.5-µg sample of linearized HA-AID-CTCFcDNA-AID-eGFP-blasticidin vector was transfected. Twenty-four hours after transfection, blasticidin (10 µg ml^{-1})-containing medium was added, and resistant cells were selected for 48 h. A second transfection was then performed using 2 µg of four sgRNA-CRISPR-vectors (4 × 0.5 µg) on the pool of blasticidin-resistant cells to knock out the endogenous CTCF. After 24 h, puromycin (1.5 µg ml^{-1})-containing medium was added, and resistant cells were selected for 48 h. Serial dilutions were then done on 96-well plates without antibiotic selection to generate single-cell clones. To test for integration of HA-AID-CTCFcDNA-AID-eGFP-blasticidin and effective CTCF knockout, cells from individual clones were trypsinized; half was left in the 96-well plate and the other half was used for genomic DNA extraction. Clones that harbored the endogenous CTCF knockout and the integration of the HA-AID-CTCFcDNA-AID-eGFP-blasticidin construct were sequenced. Clone (referred to as HAP1-CTCFdegron in our study) with the correct sequence was used for TIR1 integration. This clone is diploid. A 2-µg sample of linearized TIR1-hygro vectors were then transfected into the HAP1-CTCFdegron clone. After 24 h, hygromycin (450 µg ml^{-1})-containing medium was added, and resistant cells were selected for 48 h. Serial dilutions were then done on 96-well plates without antibiotic selection to generate single-cell clones. Clones were then polymerase-chain-reaction (PCR) tested and sequenced for correct TIR1 integration on single clones. The diploid clone used in this study is referred to as HAP1-CTCFdegron-TIR1.

HAP1-RPB1-AID. A 1.5-µg sample of linearized RPB1-AID-eGFP-blasticidin vector and 1.5 µg of C-terminal RPB1 sgRNA were transfected into HAP1 cells. At 24 h after transfection, puromycin (1.5 µg ml^{-1})-containing medium was added, and resistant cells were selected for 48 h. Puromycin medium was then washed, and the cells were grown for 48 h without antibiotics. Blasticidin-resistant cells were selected by adding blasticidin (10 µg ml^{-1})-containing medium for seven days. The pool of blasticidin-resistant cells was then transfected with 2 µg of linearized TIR1-hygro vector. After 24 h, serial dilution of cells to select single-cell clones was performed in hygromycin (450 µg ml^{-1})-containing medium. Clones were tested by PCR and sequenced for correct AID-eGFP and TIR1 integrations on single clones. The HAP1-RPB1-AID cells are diploid.

AAVS1, DDX55 and TAF5L knockouts. sgRNAs (2 µg) targeting the AAVS1, DDX55 and TAF5L loci were transfected into HAP1 cells. Twenty-four hours after transfection, puromycin (1 µg ml^{-1})-containing medium was added and resistant cells were selected for 48 h. Serial dilutions of cells in medium without selection were then done to select single-cell clones. Clones were tested by PCR and sequenced for indels on both alleles. The AAVS1 clone (control) harbors a 23-bp deletion on both alleles. DDX55 clone 1 harbors one allele with a 3-bp deletion, deleting two amino acids (I and P) and replacing it by another one (T). The second allele has a 4-bp deletion, creating a frameshift and premature stop codon in exon 3. DDX55 clone 2 harbors one allele with a 6-bp deletion, deleting three amino acids (PLF) and replacing it by another one (L). The second allele has a 12-bp deletion deleting four amino acids (ATIP). The amount of mutated DDX55 protein is reduced in both clones. TAF5L clone 1 is homozygous with a 7-bp deletion in the third exon of the *TAF5L* gene, creating a premature stop codon. TAF5L clone 2 is homozygous with a 13-bp deletion creating a premature stop codon. These TAF5L knockout clones do not express the TAF5L protein.

Genomic DNA extraction for PCR to test clones

Cells were spun, resuspended in 30 µl of squishing buffer (SB) (10 mM Tris pH 8.0, 25 mM NaCl, 1 mM EDTA, 200 µg ml^{-1} Proteinase K), incubated for 1 h at 65 °C and 10 min at 95 °C, spun, then 1 µl of the supernatant was used for PCR.

CRISPR screen validation. Validation was performed on 16 genes, with two different sgRNAs targeting the gene of interest on HAP1-CTCFdegron-TIR1 cells.

Targeting sgRNA plasmids (2 µg) were transfected (separately for the sgRNA targeting the same gene) using turbofectin (Origene, TF81001) following the manufacturer's recommendations. After 24 h, puromycin (1.5 µg ml^{-1})-containing medium was added to select cells that integrated the plasmids. After 48 h, cells were counted and the time point considered as T0. Passaging was then performed following the scheme used in the genome-wide CRISPR screen. Three days later (T3), cells were counted and reseeded into two conditions (NT and 25 µM auxin) in duplicates in 24-well plates. The cells were counted and reseeded for the two conditions every three days until reaching T15. Cumulative growth curves were plotted with the number of counted cells. We calculated the doubling average ΔΔ by first calculating the cumulative doubling averages per gene (two sgRNAs per gene) for each time point. We then subtracted the cumulative doubling of the auxin treated from the non-treated (NT − IAA) per gene for each time point. Subsequently, we subtracted the control value (AAVS1) per gene for each time point. Finally, we calculated the mean of all time points for each experiment replicate. A positive ΔΔ value indicates a growth defect when the gene is knocked out and CTCF is depleted. A negative ΔΔ value indicates a better proliferation when the gene is knocked out and CTCF is depleted. To confirm that indels occurred, cells were collected at T15, and genomic DNA extraction was performed. PCR was then performed on the extracted gDNA from cells that passed through the transfections (mutated amplicon) and for cells that were not transfected (wild-type amplicon). PCR products were purified using the GFX PCR DNA and Gel Band Purification Kit (Cytiva, 28903470) and sent for Sanger sequencing. The Synthego tool (https://www.synthego.com/products/bioinformatics/crispr-analysis) was then used to assess the percentage of the different modified alleles in the targeted genes using the wild-type amplicons as controls[69].

Flow cytometry

Cells were dissociated with accutase (Thermo Fisher Scientific, A11105-01), resuspended in phosphate buffered saline (PBS), spun, then resuspended in 250 µl of cold PBS. To assess the cell-cycle profile (DNA content), 750 µl of 100% ethanol was slowly added to fix cells in 75% ethanol. Cells were stored at −20 °C for at least 24 h. The fixed cells were spun, resuspended in 1× PBS with propidium iodide (PI) (final concentration of 50 µg ml^{-1}) and RNAseA (0.5 mg ml^{-1}), then incubated for 30 min at room temperature (r.t.), protected from light. To assess the GFP content, cells were washed once with PBS and fixed with 4% paraformaldehyde (PFA) for 10 min. The cells were spun, then the cell pellets were resuspended in 1 ml of PBS. The cells were sorted on a FACSCALIBUR or LSRII or MACSQUANT set-up. Analysis was performed using FlowJo software (v10). The gating strategy is outlined in Supplementary Fig. 1.

Western blots

Cells were dissociated with accutase (Thermo Fisher Scientific, A11105-01), resuspended in PBS, spun, washed with PBS, spun again and kept at −20 °C. At least 1 million cells were resuspended in 100 µl of RIPA buffer (Thermo Fisher Scientific, 89900) for 30 min on ice to lyse the cells. Lysates were spun for 30 min at 4 °C, and the supernatants containing the soluble proteins were collected. Protein concentration was calculated using a Pierce BCA protein assay kit (Thermo Fisher Scientific, 23227), with 20 µg of protein loaded per lane. Samples were mixed with Pierce lane marker reducing sample buffer (Thermo Fisher Scientific, 39000) and run on a NuPAGE 3–8% Tris-acetate protein gel with NuPAGE Tris-acetate sodium-dodecyl-sulfate running buffer

(Thermo Fisher Scientific, LA0041) in an XCell SureLock Mini-Cell (Thermo Fisher Scientific, EI0001). Transfer onto a nitrocellulose membrane (0.2 µm; Bio-Rad, 1620112), was performed using the XCell SureLock Mini-Cell (Thermo Fisher Scientific, EI0001) in Pierce 10X western blot transfer buffer, methanol-free (Thermo Fisher Scientific, 35045) for 2 h at 30 V. Membranes were blocked for 2 h at r.t. with 5% milk in Tris-buffered saline with Tween (TBST) before antibody incubation overnight at 4 °C (Supplementary Table 6 presents the antibodies used). Antibodies were added in 5% milk with TBST. Membranes were washed six times for 10 min in TBST at r.t., incubated with horseradish peroxidase (HRP) secondary antibodies (Cell Signaling, 7074) 1:1,000 in 5% milk with TBST for 2 h at r.t., washed six times 10 min with TBST at room temperature, revealed with SuperSignal West Dura extended duration substrate (Thermo Fisher Scientific, 34076) and analyzed on a Bio-Rad ChemiDoc system with Image Lab 6.0.1 builder 34.

Co-immunoprecipitation

The co-IP protocol was adapted from ref. [70]. Cells were grown on 15-cm plates, washed with Dulbecco's phosphate-buffered saline (DPBS) and collected with accutase. For each co-IP step, ~30 million cells were used. Each pellet was resuspended in 1 ml of low-salt lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40 and 1X HALT protease inhibitor) and incubated on ice for 10 min. Nuclei were pelleted for 10 min at 1,500*g* at 4 °C, and resuspended in 1 ml of low-salt lysis buffer. Each set of co-IP involved three samples (NT, turbonuclease and RNAseA). The turbonuclease samples were treated with 1,200 U of turbonuclease and the RNAseA samples were treated with 0.1 mg ml^{-1} RNAseA. The samples were then incubated for 4 h at 4 °C on a rotator. After the incubation, a 50-µl sample was taken from each tube to check the efficiency of the DNA and RNA degradation. The NaCl concentration of the rest of the samples was adjusted to 200 mM, and the samples were incubated on ice for 30 min. They were then centrifuged for 10 min at maximum speed and 4 °C to extract the protein. Proteins were quantified with BCA, and 1 mg of protein was used for co-IP. A 1-mg sample of the lysate was precleared for 4 h at 4 °C with 80 µl of Protein G Dynabeads magnetic beads (10004D) washed once in co-IP buffer (0.2 M NaCl, 25 mM HEPES, 1 mM $MgCl_2$, 0.2 mM EDTA, 0.5% NP-40 and 1X HALT protease inhibitor). After pre-clearing, 1% of the input was kept to check the CTCF and RAD21 depletion and also to load onto the western blot gels. The 1-ml lysate was divided into two tubes of 500 µl and incubated overnight either with 5 µl of rabbit immunoglobulin-G (normal rabbit IgG, Cell Signaling Technology, #2729, 1 mg ml^{-1}) or 5 µl DDX55 (Bethyl 1 mg ml^{-1}, A303-027A), 15 µl TAF5L (Proteintech, 19274-1-AP, 0.333 mg ml^{-1}) or 5 µl TAF6L (ABclonal, A14369, 3.38 mg ml^{-1}). The next day, 40 µl of Protein G Dynabeads magnetic beads washed once in co-IP buffer (10004D) were added to each tube and incubated for 2 h at 4 °C. The beads were then washed five times for 5 min with 500 µl of co-IP buffer using a magnetic rack at room temperature. The flowthrough and last wash were kept for western blot gels. To elute the proteins, the beads were resuspended in 20 µl of 2× sodium dodecyl sulfate buffer, heated for 5 min at 100 °C, then the supernatants were removed after placing the tubes on the magnetic rack. The totality of the 20-µl sample was loaded onto NuPAGE Novex 3–8% Tris-acetate protein gels (1.0 mm, 12-well) and analyzed by western blot. Each co-IP was performed in two replicates (see Supplementary Table 6 for the antibodies used).

DNA and RNA extraction to check DNA and RNA degradation efficiency for co-IP

DNA was extracted using the DNA extraction kit from Qiagen (DNeasy blood and tissue kit, 69504) and resuspended in 25 µl of water. The DNA concentration was assessed with a Qubit broad range kit (dsDNA BR assay kit, Q32850, Thermo Fisher Scientific) or NanoDrop system. Of the NT samples, 100 ng was collected, as well as an equal volume from the nuclease-treated samples, for quantification by quantitative PCR (qPCR).

RNA was extracted with TRIzol following the manufacturer's recommendations. After precipitation, the RNA was resuspended in 25 µl of water. The RNA concentration was assessed with a NanoDrop system. For each reverse transcription reaction, 1 µg of the NT samples was taken, as well as an equal volume of the nuclease-treated samples. Reverse transcription was performed with a VILO IV system (SuperScript IV VILO Master Mix, 11756050, Thermo Fisher Scientific) and incubated for 20 min at 25 °C, 10 min at 50 °C and 5 min at 85 °C. cDNA was diluted by 20×, and 2 µl was used for qPCR.

Quantitative polymerase chain reaction

qPCR was directly performed on the cDNA using Fast SYBR Green Master Mix (Thermo Fisher Scientific, 4385612) and analyzed on a StepOnePlus Real-Time PCR system (Thermo Fisher Scientific) using StepOnePlus v2.3 software. See Supplementary Table 1 for the qPCR primer sequences.

Statistics and reproducibility

The co-IPs and western blots were performed at least twice for each condition. All Hi-C, RNA-seq and ChIP-seq procedures were performed in two biological replicates. CRISPR screens were performed in three technical replicates for each condition.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this Article.

Data availability

The datasets generated in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO SuperSeries accession no. GSE180691. This SuperSeries is composed of the following SubSeries: GSE180922 (Hi-C), GSE180713 (RNA-seq), GSE180690 (ChIP-seq) and GSE180657 (CRISPR screen). The following published datasets were used in this study (Supplementary Table 7): GSE72800, GSE110133, GSE70189, GSE104334, GSE104888, GSE95015; ENCODE, https://www.encodeproject.org/experiments/ENCSR131DVD/; ENCODE, https://www.encodeproject.org/experiments/ENCSR620QNS/; ENCODE, https://www.encodeproject.org/files/ENCFF176NSX/@@download/ENCFF176NSX.bigWig; ENCODE, https://www.encodeproject.org/files/ENCFF364QXM/. Data supporting the findings of this study are available from the corresponding author upon reasonable request. Source data are provided with this paper.

Code availability

Open2C scripts and notebooks used in this study are publicly available in GitHub: https://github.com/open2c and https://github.com/dekkerlab/ALV-repo.git. No other customized codes were developed for this study.

References

68. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).
69. Conant, D. et al. Inference of CRISPR Edits from Sanger Trace Data. *CRISPR J.* **5**, 123–130 (2022).
70. Hansen, A. S. et al. Distinct classes of chromatin loops revealed by deletion of an RNA-binding region in CTCF. *Mol. Cell* **76**, 395–411 (2019).
71. Fong, N. et al. Effects of transcription elongation rate and Xrn2 exonuclease activity on RNA polymerase II termination suggest widespread kinetic competition. *Mol. Cell* **60**, 256–267 (2015).

Acknowledgements

We thank members of the Dekker and the Mirny laboratories as well as members of the Open Chromosome Collective for creating a collaborative atmosphere and insightful discussions. We thank the

Flow Cytometry Core Facility for FACS sorting the cell lines, and the Deep Sequencing Core for the sequencing at the University of Massachusetts Chan Medical School. We thank C. Navarro for help with editing the manuscript. We thank M. Kanemaki (National Institute of Genetics, Mishima, Japan) for sharing the HCT116-RAD21-AID cell line. We thank E. Nora and B. Bruneau (Gladstone Institutes, San Francisco, CA, USA) for sharing plasmids. This work was supported by a grant from the National Human Genome Research Institute (NHGRI) to J.D. (HG003143) and a grant from the National Institute of General Medical Sciences (NIGMS) to A.A.P. (GM133762). J.D. is an investigator of the Howard Hughes Medical Institute. Some of the schematic figures were created with BioRender.com.

Author contributions

A.-L.V. and J.D. conceived and designed the study. A.-L.V. engineered cell lines, performed Hi-C, ChIP-seq and all the other experiments. A.-L.V. and S.V.V. analyzed Hi-C, ChIP-seq, RNA-seq and other relevant datasets. A.-L.V., B.M., A.H.Y.T. and J.M. designed the strategy for the CRISPR screens. K.C. generated the lentiviruses for the CRISPR screens. A.-L.V. and B.M. performed the CRISPR screens. A.-L.V., B.M. and M.U. analyzed the CRISPR screen data. E.S.K. and A.A.P. analyzed splicing in the RNA-seq data. A.-L.V. and J.D. wrote the manuscript with input from all the authors.

Competing interests

The authors declare no competing interests.

Additional information

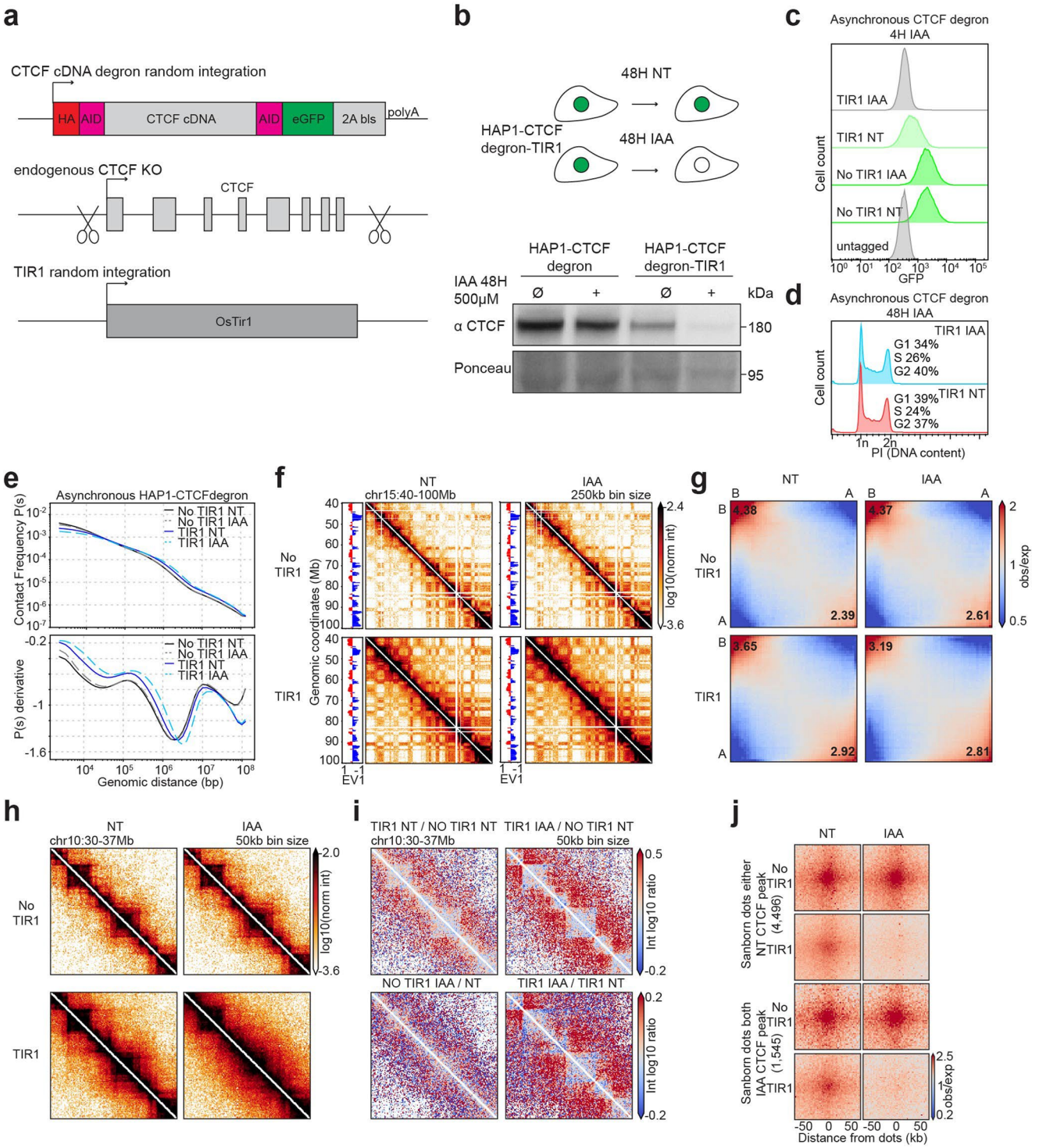
Extended data is available for this paper at https://doi.org/10.1038/s41594-022-00890-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-022-00890-9.

Correspondence and requests for materials should be addressed to Job Dekker.

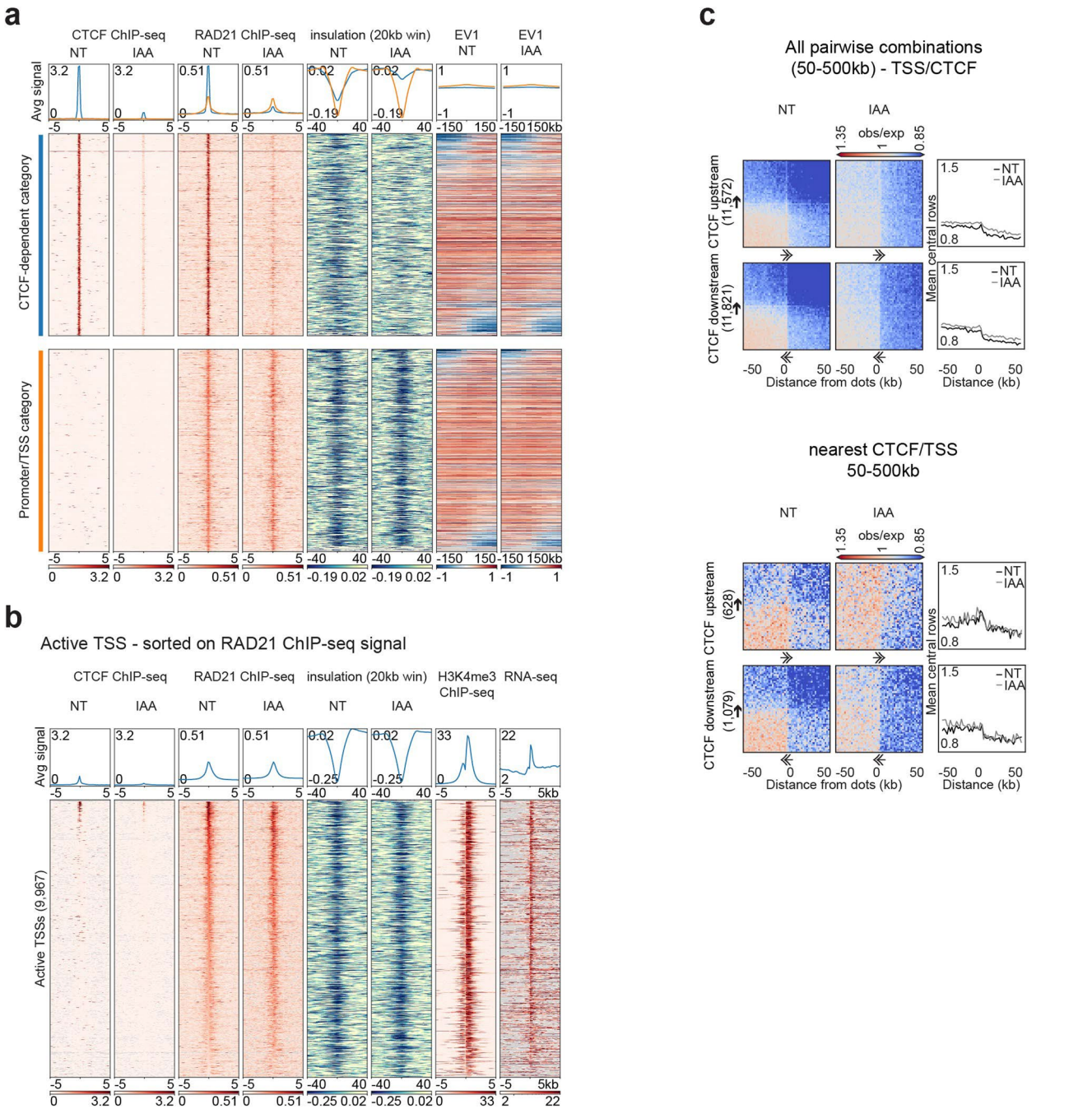
Peer review information *Nature Structural & Molecular Biology* thanks Daniel Ibrahim and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: Carolina Perdigoto, in collaboration with the *Nature Structural & Molecular Biology* team. Peer reviewer reports are available.

https://doi.org/10.1038/s41594-022-00890-9



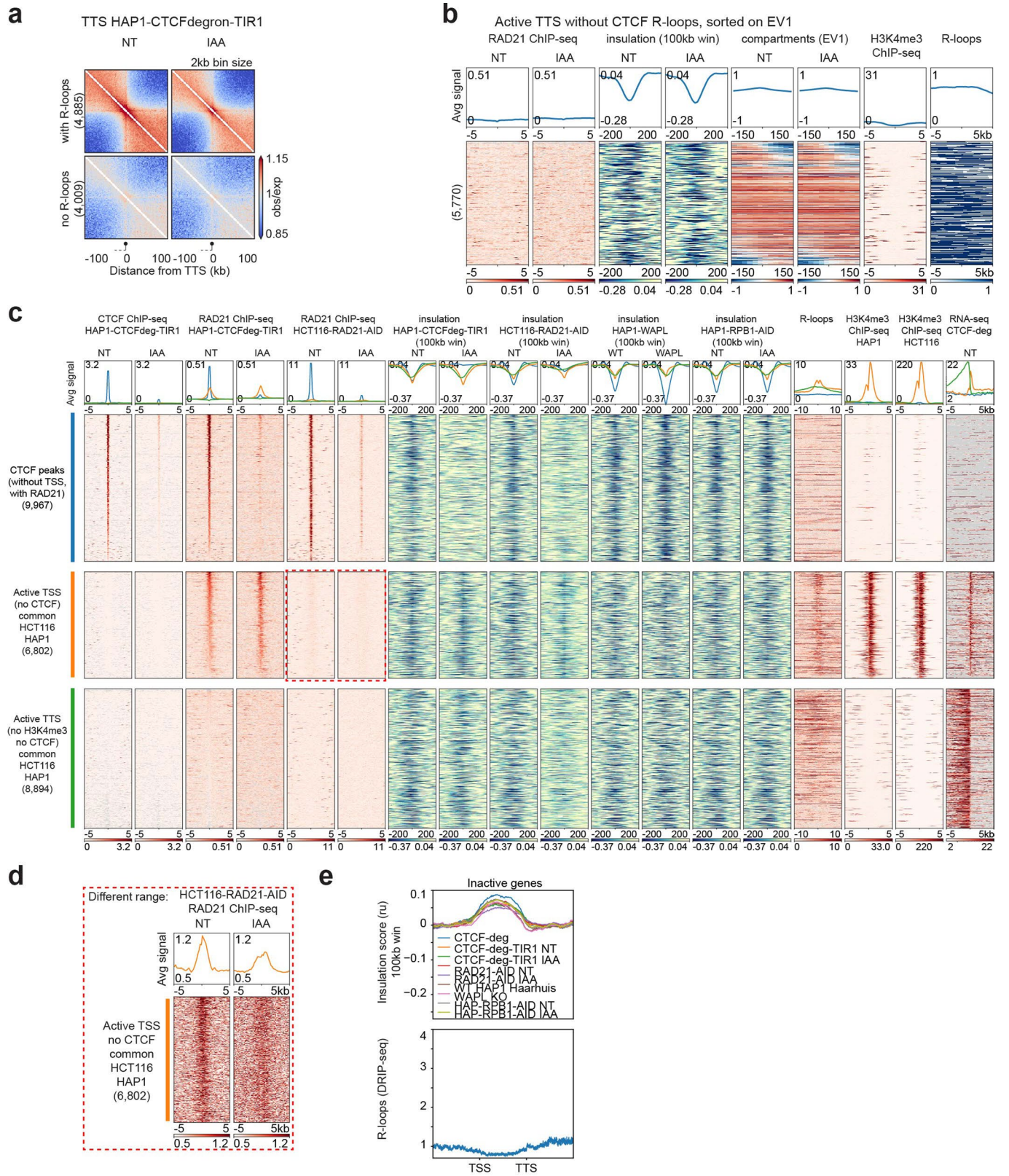
Extended Data Fig. 1 | Characterization of the HAP1-CTCFdegron cell lines. **a**, Schematic representing the strategy used to construct the HAP1-CTCFdegron-TIR1 cells. **b**, Schematic representing CTCF depletion in HAP1-CTCFdegron-TIR1 cells (top). Western blot against CTCF in HAP1-CTCFdegron cells and in HAP1-CTCFdegron-TIR1 cells without/with auxin (NT and IAA). Ponceau is shown for loading control (bottom). **c**, Flow cytometry for GFP in HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 cells without/with auxin. **d**, Flow cytometry for HAP1-CTCFdegron-TIR1 cells without/with auxin stained with Propidium Iodide (PI) to assess the DNA content for cell cycle analysis. **e**, Hi-C contact frequency as a function of genomic distance, P(s) (top) and its derivative dP/ds (bottom) for HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 cells without/with auxin. **f**, Hi-C contact heatmaps at 250kb resolution with the corresponding track of the first Eigenvector (EV1) across chromosome 15 for HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 cells without/with auxin. **g**, Genome-wide saddle plots of Hi-C data binned at 100kb resolution for HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 cells without/with auxin. The compartment strengths are indicated in the corners. **h**, Hi-C contact heatmaps at 50kb resolution for a 7Mb region on chromosome 10 for HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 cells without/with auxin. **i**, Differential interaction heatmaps for HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 cells. **j**, Dot pileups for dots characterized in HAP1 cells that have a CTCF peak in either anchor in the Non-Treated sample (4,496 dots) and that have a CTCF peak in both anchors in the auxin sample (1,545 dots) for HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 cells without/with auxin. The dots were aggregated at the center of a 100kb window at 2kb resolution.

https://doi.org/10.1038/s41594-022-00890-9



Extended Data Fig. 2 | CTCF sites and active TSSs are chromatin boundaries. **a**, Stackups for CTCF-dependent (blue) and promoter/TSS (orange) categories sorted on the change of the first Eigenvector (EV1, 25kb) from left to right flank. CTCF and RAD21 ChIP-seq, calculated insulation and EV1 in HAP1-CTCFdegron-TIR1 cells without/with auxin (NT and IAA) were plotted. **b**, Stackups for active TSSs, sorted on RAD21 ChIP-seq signal. CTCF and RAD21 ChIP-seq, calculated insulation and RNAseq in the HAP1-CTCFdegron-TIR1 cells without/with auxin were plotted along with the published HAP1 H3K4me3 ChIP-seq signals[26]. Stackups were flipped according to the orientation of the genes, to have the gene body at the right of the TSSs. **c**, Dot pileup aggregation plots for remaining CTCF motif orientations represented in Fig. 1e for a 100kb window at 2kb resolution. With orientation (top). CTCF (upstream or downstream)-TSS pairwise interactions are plotted with their quantifications (mean of the 5 central bins at the CTCF site). Nearest analysis (bottom). CTCF (upstream or dowstream)-TSS pairwise interactions are plotted without any CTCF peaks or TSSs in between them with their quantifications (mean of the 5 central bins at the CTCF site). The black arrows represent the CTCF motif and the direction of the arrow, the motif orientation. The double arrows represent the TSSs and the direction of the arrow, the TSSs orientation.

https://doi.org/10.1038/s41594-022-00890-9

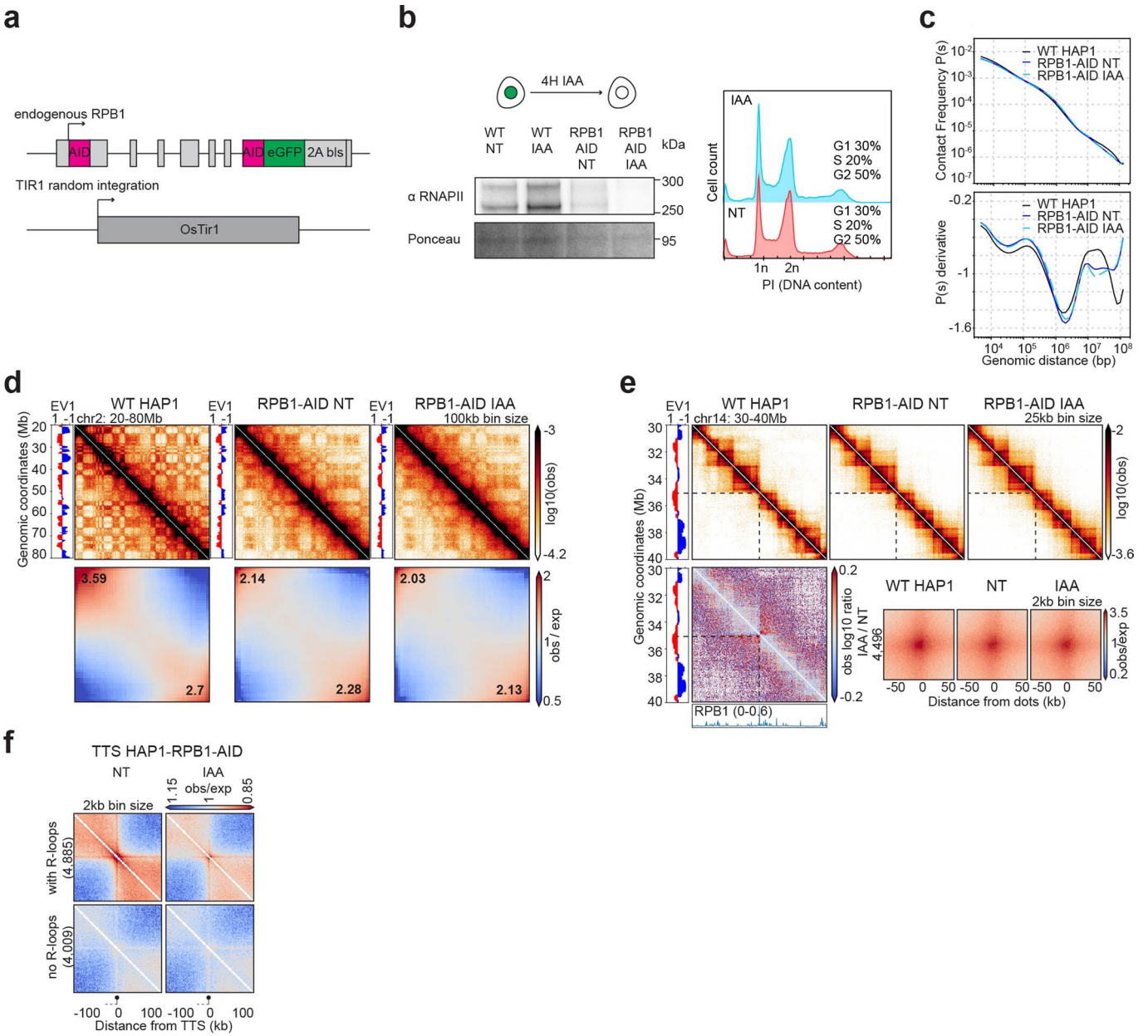


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41594-022-00890-9

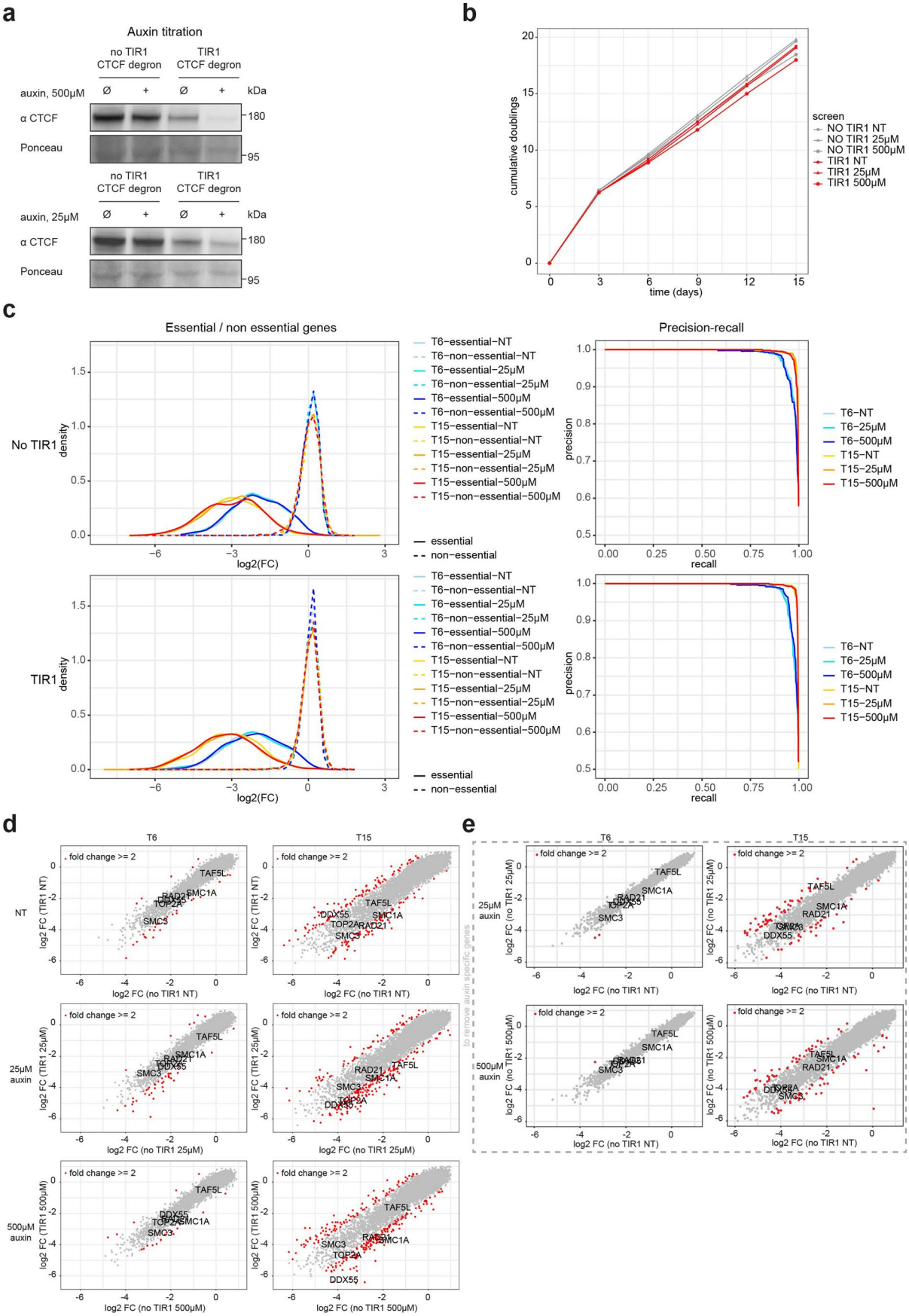
Extended Data Fig. 3 | CTCF, RAD21, WAPL and RNA polII depletion effects on the three types of chromatin boundaries. a, Oriented interaction pileups for HAP1-CTCFdegron-TIR1, without/with auxin (NT and IAA), aggregated in a 200kb window at 2kb resolution were plotted for the active TTS (without CTCF) with or without R-loops. The black circle on a stick represents the TTS and the gene body is represented by a dash line. **b**, Stackups for active TTS (without CTCF, with R-loops), sorted on the change of first Eigenvector (EV1, 25kb) signal from left to right flank. RAD21 ChIP-seq levels, calculated insulation and EV1 in HAP1-CTCFdegron-TIR1 cells without/with auxin were plotted along with the published HAP1 H3K4me3 ChIP-seq signals[26] and the consensus list of R-loops. **c**, Stackups for the three categories of insulation: the CTCF peaks (without TSSs, with RAD21) (blue), the active TSSs (orange) and TTSs (green) common between HAP1 and HCT116 cell lines, sorted on the NT RAD21 ChIP-seq signal. The CTCF and RAD21 ChIP-seq signals, the RNA-seq levels and the insulation for the described cell line and condition were plotted along with the published K562 R-loops[67] and the HAP1[26] and HCT116 H3K4me3 ChIP-seq signals. Stackups were flipped according to the orientation of the genes, to have the gene body on the right for the TSSs and on the left of the TTSs. The red dashed rectangle indicates the zoom in Extended Data Fig. 3d. **d**, Active TSSs common between HAP1 and HCT116 cell lines (without CTCF) were plotted with a different scale to show the remaining RAD21 after RAD21 depletion (red dashed rectangle). **e**, Average insulation profiles across scaled inactive genes without CTCF at TSSs and TTSs at 5kb resolution for all the Hi-C libraries plotted in Fig. 2b.

https://doi.org/10.1038/s41594-022-00890-9



Extended Data Fig. 4 | Characterization of the HAP1-RPB1-AID cell line. a, Schematic of the HAP1-RPB1-AID construct **b**, Western blot against RPB1 in WT HAP1 cells and HAP1-RPB1-AID cells showing RPB1 depletion after 4 hours of auxin treatment (IAA). Ponceau is shown for loading control. Flow cytometry for HAP1-RPB1-AID cells without/with auxin (NT and IAA) stained with Propidium Iodide (PI) to assess the DNA content for cell cycle analysis. **c**, Hi-C contact frequency as a function of genomic distance, P(s) (top) and its derivative dP/ds (bottom) for HAP1 wild-type cells and HAP1-RPB1-AID cells in absence and presence of auxin. **d**, Hi-C contact heatmaps at 100kb resolution with the corresponding track of the first Eigenvector (EV1) across a 60Mb region on chromosome 2 for HAP1 and HAP1-RPB1-AID cells without/with auxin. Genome-wide saddle plots of Hi-C data binned at 100kb resolution for HAP1 and HAP1-RPB1-AID cells without/with auxin. The compartment strengths are indicated in the corners. **e**, Hi-C contact heatmaps at 25kb resolution with the corresponding distribution of EV1 and published RPB1 ChIP-seq signal[71] for a 10Mb region on chromosome 14 for HAP1 and HAP1-RPB1-AID cells without/with auxin. The differential interaction heatmap (presence/absence of auxin) is shown on the bottom. Dot pileups for dots found in HAP1 cells that have a CTCF peak in either anchor in the CTCF degron NT sample (4,496 dots) for HAP1-RPB1-AID cells without/with auxin. The dots were aggregated at the center of a 100kb window at 2kb resolution. **f**, Oriented interaction pileups for HAP1-RPB1-AID, without/with auxin, aggregated in a 200kb window at 2kb resolution were plotted for the active TTS with or without R-loops. The black circle on a stick represents the TTS and the gene body is represented by a dash line on the left of the TTS.

https://doi.org/10.1038/s41594-022-00890-9

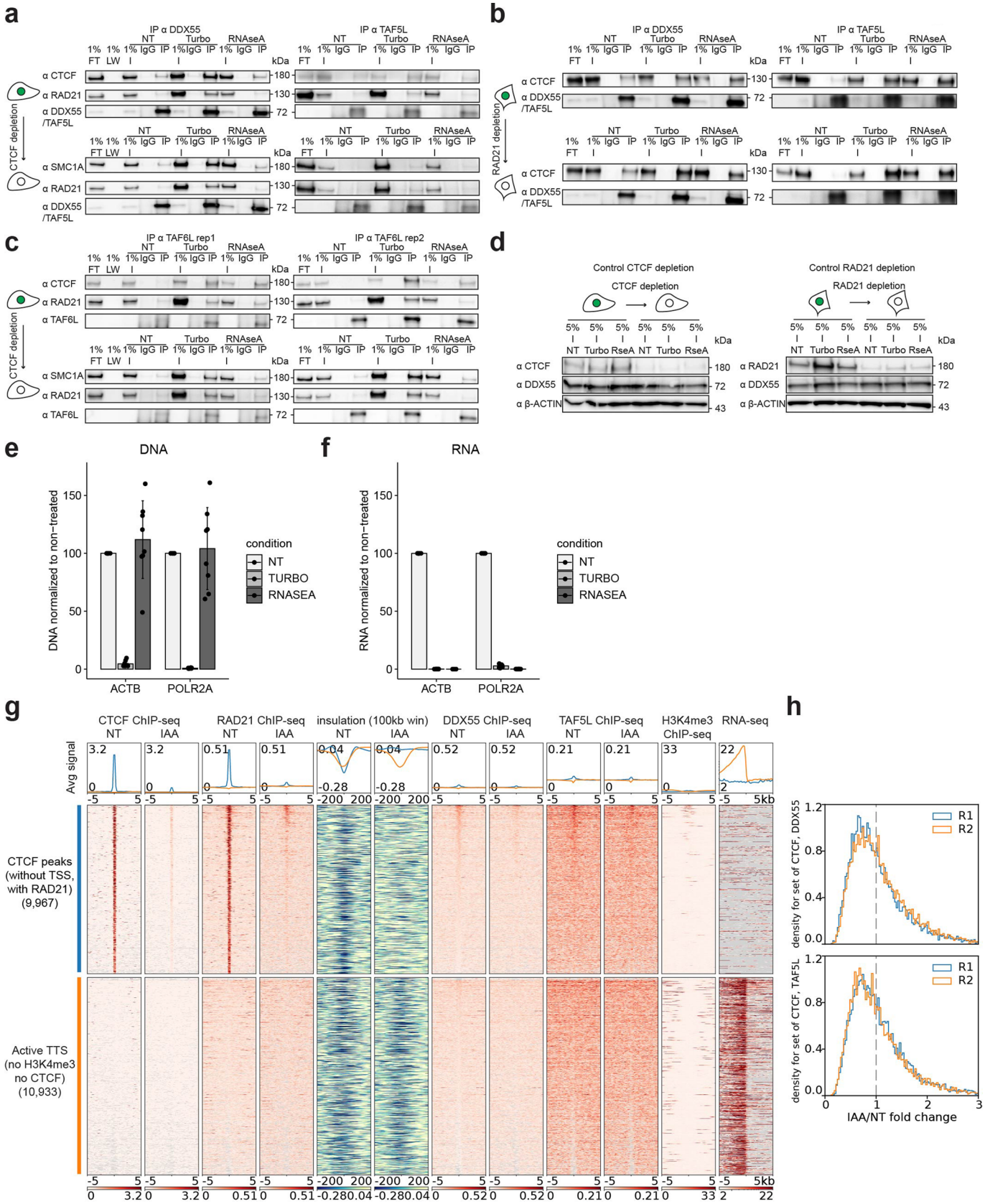


Extended Data Fig. 5 | See next page for caption.

https://doi.org/10.1038/s41594-022-00890-9

Extended Data Fig. 5 | Genome wide CRISPR screen in context of altered cohesin traffic pattern following CTCF depletion. a, Western blot against CTCF in HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 cells without/with auxin (NT and IAA) with the two different auxin concentrations (25μM and 500μM) used in the screen showing the partial CTCF depletion with 25μM IAA and the nearly total CTCF depletion with 500μM IAA. Ponceau is shown for loading control. **b**, Plot showing the cumulative number of doublings relative to the days cells were passaged during the screen for HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 for the three auxin concentrations (NT, 25μM and 500μM). **c**, Fold change distribution of essential and nonessential gene sets at indicated time points for the screens in HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 cells for the three auxin concentrations (left). Precision-recall curves based on Bayes Factors (BFs) of predefined essential and non-essential gene sets for the screens for HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 cells for the three auxin concentrations at the indicated time points (right). **d**, Scatter plots of the log2FC for HAP1-CTCFdegron-TIR1 screens against HAP1-CTCFdegron screens for T6 and T15 time points. Genes with a fold change of ≥ 2 between HAP1-CTCFdegron and HAP1-CTCFdegron-TIR1 screens are highlighted in red. Genes linked to CTCF are indicated (cohesin, *TOP2A*) along with the two studied gene hits (*DDX55* and *TAF5L*). **e**, Scatter plots of the log2FC for HAP1-CTCFdegron cells without/with auxin at T6 and T15 time points. Genes with a fold change of ≥ 2 between absence and presence of auxin are highlighted in red and considered as auxin specific genes. Genes linked to *CTCF* are indicated (cohesin genes, *TOP2A*) along with the two studied gene hits (*DDX55* and *TAF5L*).

https://doi.org/10.1038/s41594-022-00890-9

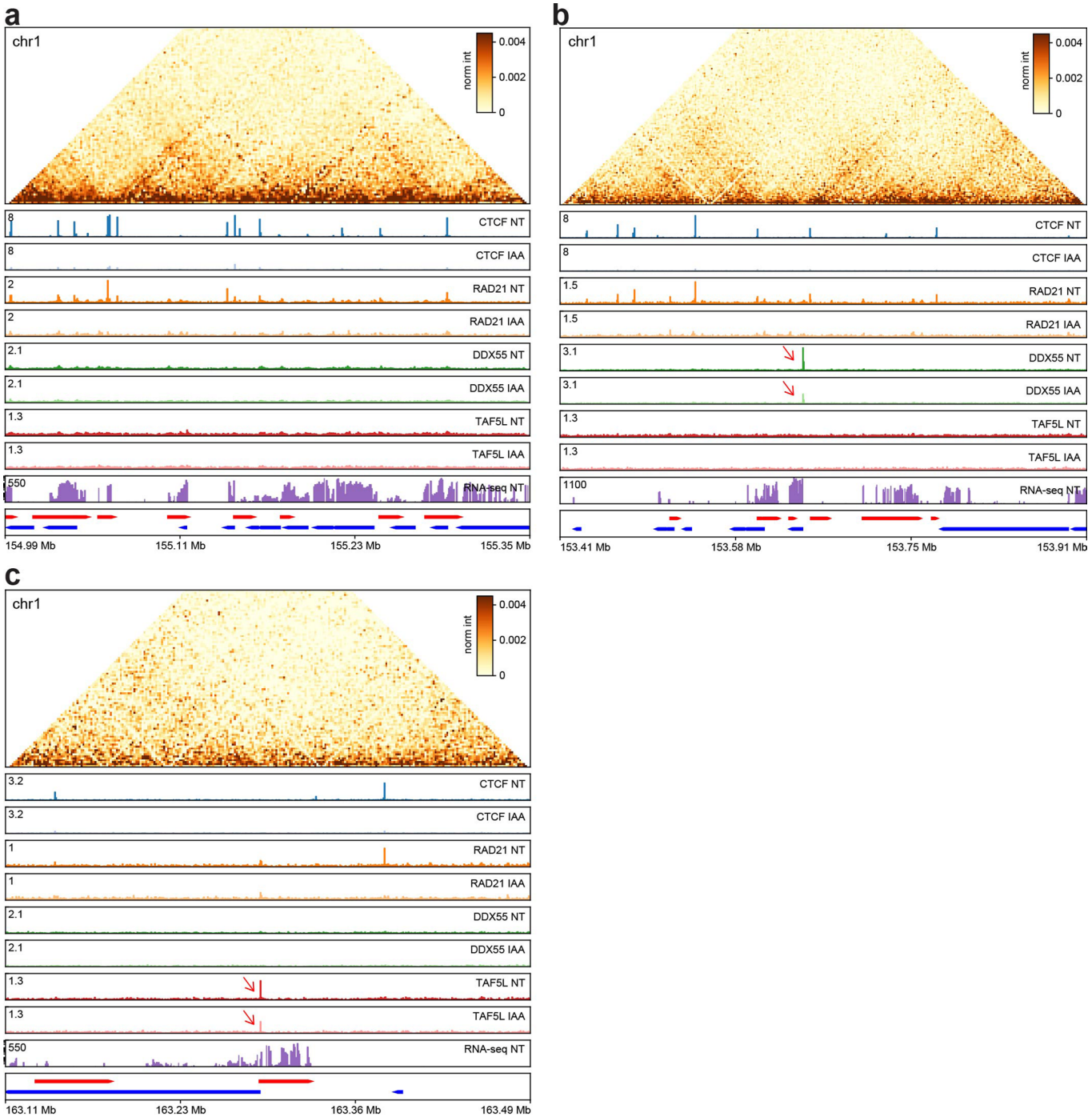


Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s41594-022-00890-9

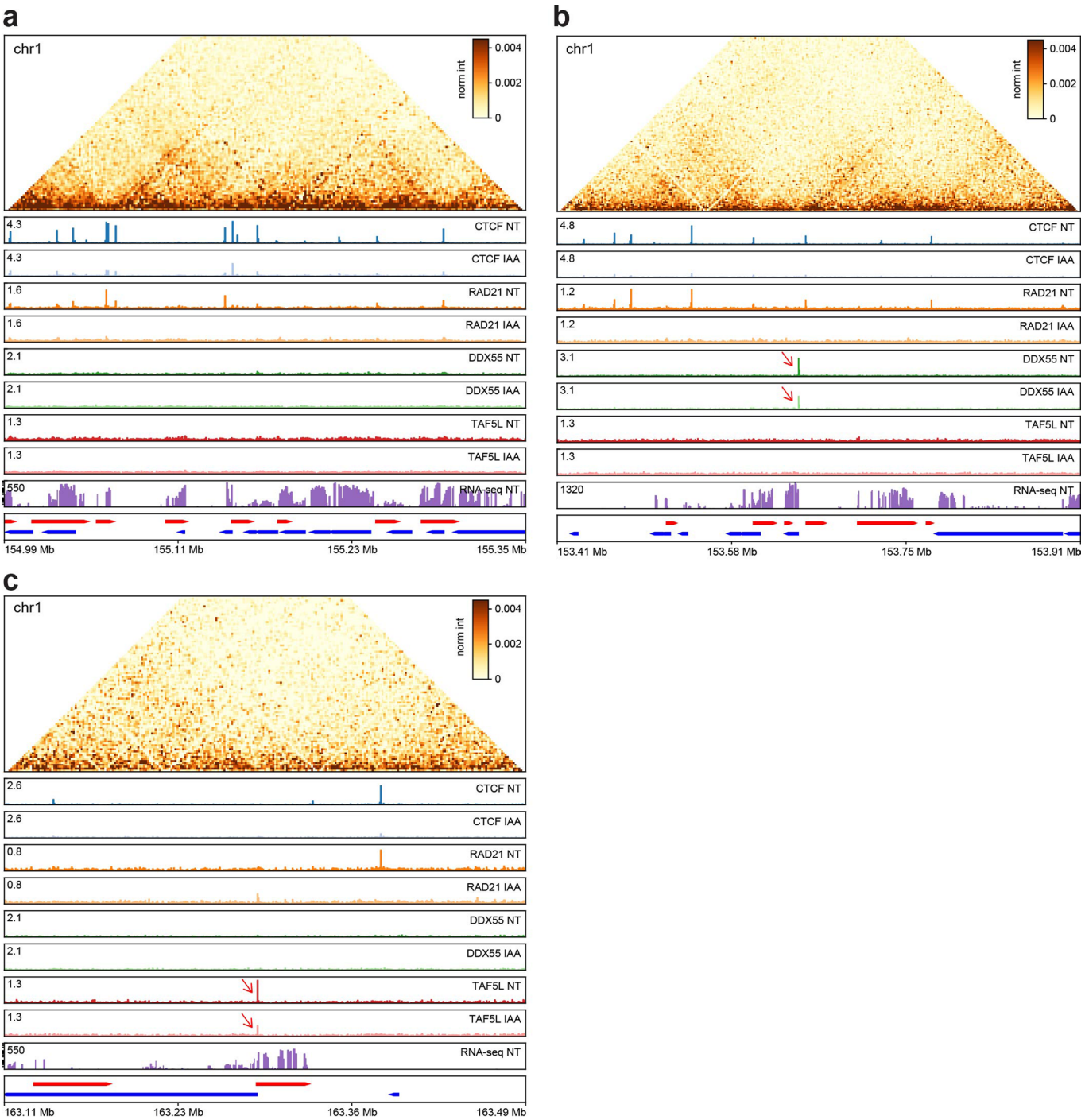
Extended Data Fig. 6 | DDX55 and TAF5L chromatin binding. a, Western blot co-IP replicates against DDX55 and TAF5L in HAP1-CTCFdegron-TIR1 without/with auxin (NT and IAA), treated with either turbonuclease (DNA - and RNA -) or RNAseA (RNA -) for CTCF, cohesin (RAD21 and SMC1A), DDX55 and TAF5L. **b**, Western blot co-IP replicates against DDX55 and TAF5L in HCT116-RAD21-AID cells without/with auxin, treated with either turbonuclease or RNAseA for CTCF, DDX55 and TAF5L. **c**, Western blot co-IP (two replicates) against TAF6L in HAP1-CTCFdegron-TIR1 cells without/with auxin, treated with either turbonuclease or RNAseA for CTCF, cohesin (RAD21 and SMC1A) and TAF6L. **d**, Representative western blots against CTCF, RAD21, DDX55 and β-ACTIN (loading control) showing the CTCF depletion efficiency in the HAP1-CTCFdegron-TIR1 co-IP (left) and the RAD21 depletion efficiency in the HCT116-RAD21-AID cells co-IP (right). **e**, Efficient DNA digestion by turbonuclease (TURBO) during DDX55, TAF5L and TAF6L co-IPs. DNA digestion was assessed by qPCR using primers specific to *ACTB* and *POLR2A* gene locations. Data are presented as mean values ± SD, n = 8 biologically independent co-IP experiments. **f**, Efficient RNA digestion by turbonuclease (TURBO) and RNaseA (RNASEA) during DDX55, TAF5L and TAF6L co-IPs. RNA digestion was assessed by qPCR using primers specific to *ACTB* and *POLR2A* genes. Data are presented as mean values ± SD, n = 8 biologically independent co-IP experiments. **g**, Stackups for CTCF (without TSSs, with RAD21), sorted on the Non-Treated (NT) DDX55 ChIP-seq signal. CTCF ChIP-seq, RAD21 ChIP-seq, calculated insulation, DDX55 ChIP-seq, TAF5L ChIP-seq and RNA seq signals in HAP1-CTCFdegron-TIR1 cells without/with auxin were plotted along with published HAP1 H3K4me3 ChIP-seq[26]. For TTSs, stackups were flipped according to the orientation of the genes, to have the gene body on the left of the TTS. **h**, Stackup quantification for CTCF (without TSSs, with RAD21) for DDX55 and TAF5L ChIP-seq for two replicates. The distribution of ratios between auxin-treated and non-treated signals is shown. A fold change < 1 represents less binding of DDX55 or TAF5L at CTCF (without TSSs, with RAD21) after CTCF depletion.

https://doi.org/10.1038/s41594-022-00890-9



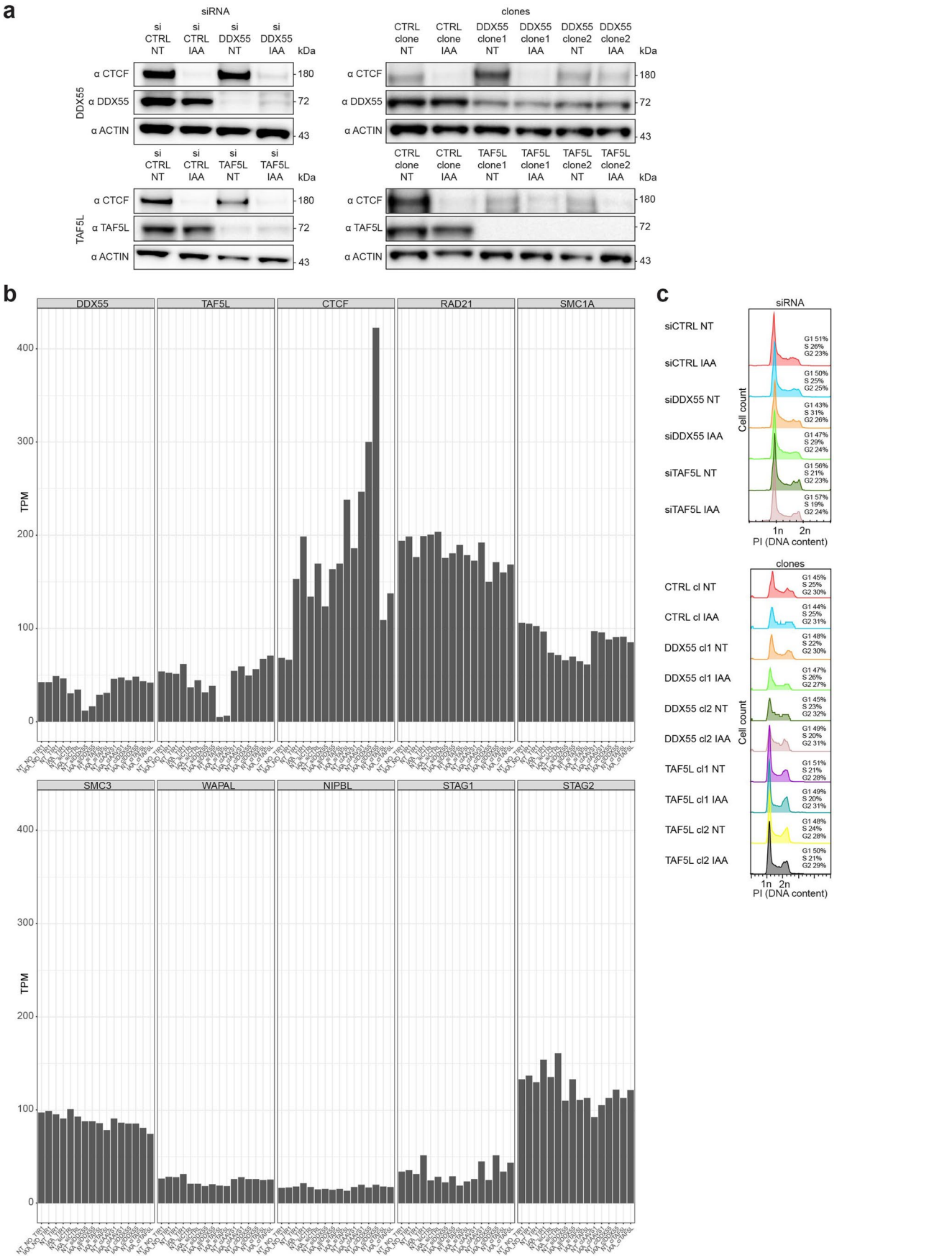
Extended Data Fig. 7 | Overview of the different genome-wide data generated in this study, ChIP-seq replicate 1. a, A genomic region on chr1 with many CTCF and RAD21 peaks. **b**, A genomic region on chr1 with a DDX55 peak at an active promoter which is reduced after CTCF depletion (red arrow). **c**, A genomic region in chr1 with a TAF5L peak at an active promoter which is reduced after CTCF depletion (red arrow). Genes on the forward strand are represented in red (plus) and genes on the reverse strand are represented in blue (minus). Hi-C contact heatmaps are binned at 2kb resolution.

https://doi.org/10.1038/s41594-022-00890-9



Extended Data Fig. 8 | Overview of the different genome-wide data generated in this study, ChIP-seq replicate 2. a, A genomic region on chr1 with many CTCF and RAD21 peaks. **b**, A genomic region on chr1 with a DDX55 peak at an active promoter which is reduced after CTCF depletion (red arrow). **c**, A genomic region in chr1 with a TAF5L peak at an active promoter which is reduced after CTCF depletion (red arrow). Genes on the forward strand are represented in red (plus) and genes on the reverse strand are represented in blue (minus). Hi-C contact heatmaps are binned at 2kb resolution.

https://doi.org/10.1038/s41594-022-00890-9

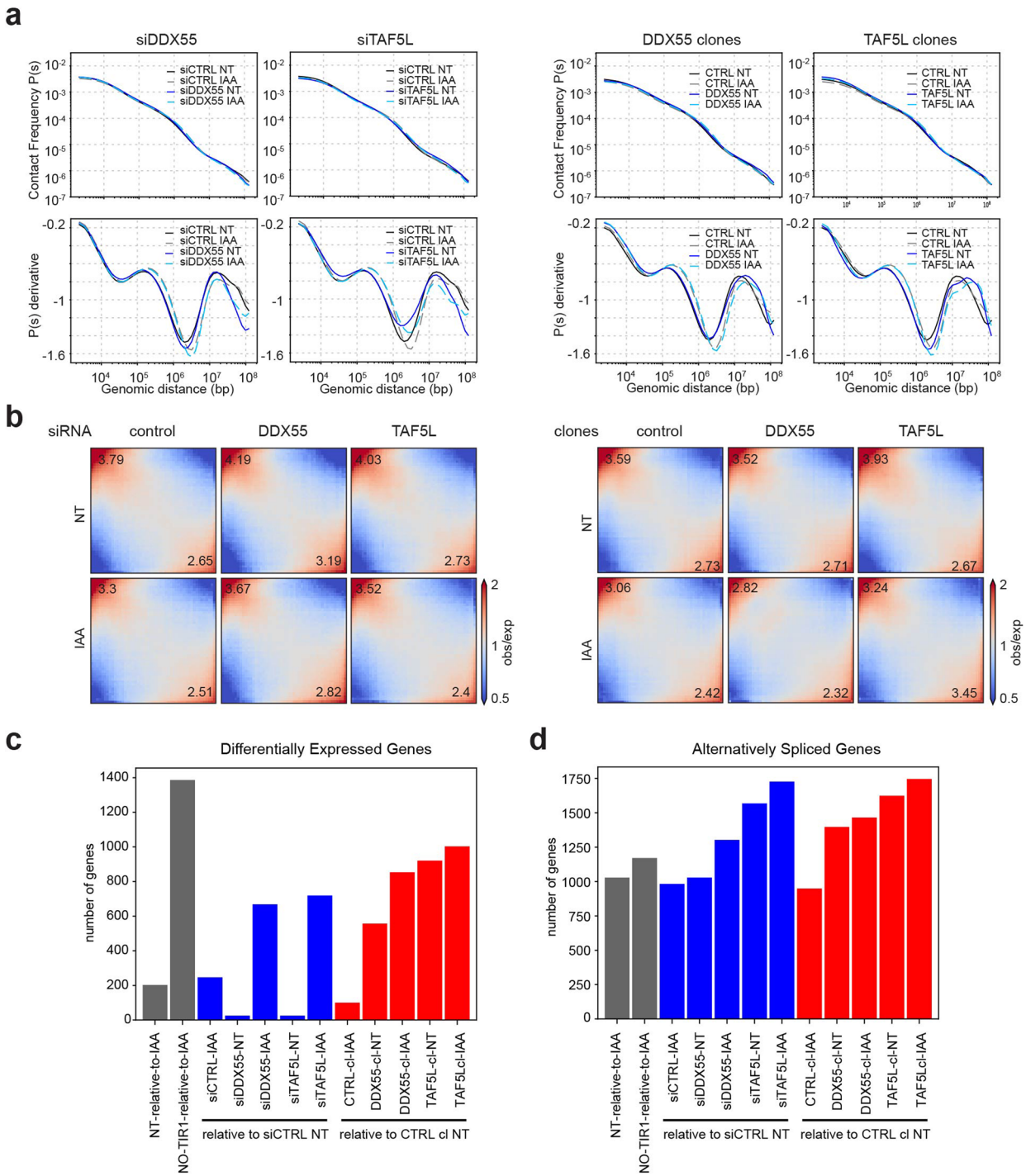


Extended Data Fig. 9 | See next page for caption.

https://doi.org/10.1038/s41594-022-00890-9

Extended Data Fig. 9 | Characterization of DDX55 and TAF5L depletions in HAP1-CTCFdegron-TIR1 cells. a, Western blots against CTCF, DDX55, TAF5L and β-ACTIN (loading control) showing DDX55 and TAF5L depletions, by siRNA and mutations in *DDX55* and *TAF5L* genes, compared to siRNA controls and mutations at the *AAVS1* non-coding sequence (CTRL clone). **b**, RNA-seq expression (TPM) for HAP1-CTCFdegron-TIR1 cells without/with auxin (NT and IAA) treated with siRNA (siCTRL, siDDX55 and siTAF5L) and CTRL, DDX55 and TAF5L clones for key genes (*DDX55*, *TAF5L*, *CTCF*, *RAD21*, *SMC1A*, *SMC3*, *WAPAL*, *NIPBL*, *STAG1* and *STAG2*). **c**, Flow cytometry for HAP1-CTCFdegron-TIR1 cells without/with auxin (NT and IAA) with siRNA (siCTRL (control), siDDX55 and siTAF5L) (top) and CTRL (control), DDX55 and TAF5L clones (bottom) stained with Propidium Iodide (PI) to assess the DNA content for cell cycle analysis.

https://doi.org/10.1038/s41594-022-00890-9



Extended Data Fig. 10 | Effect of DDX55 and TAF5L depletions on Hi-C and RNA-seq. a, Hi-C contact frequency as a function of genomic distance P(s) and its derivative dP/ds for HAP1-CTCFdegron-TIR1 cells without/with auxin (NT and IAA) treated with siRNA (siCTRL, siDDX55 and siTAF5L) (top) and CTRL, DDX55 and TAF5L clones (bottom). **b**, Genome-wide saddle plots of Hi-C data binned at 100kb resolution for HAP1-CTCFdegron-TIR1 cells without/with auxin (NT and IAA) treated with siRNA (siCTRL, siDDX55 and siTAF5L) (left) and CTRL, DDX55 and TAF5L clones (right). The compartment strengths are indicated in the corners. **c**, Number of Differentially Expressed genes in HAP1-CTCFdegron-TIR1 cells without/with auxin (NT and IAA) treated with siRNA (siCTRL, siDDX55 and siTAF5L) and CTRL, DDX55 and TAF5L clones. Gray bars indicate the CTCF depletion, blue bars indicate the siRNA depletions and red bars indicate the clones. **d**, Number of alternatively spliced genes in HAP1-CTCFdegron-TIR1 cells in absence and presence of auxin treated with siRNA (siCTRL, siDDX55 and siTAF5L) and CTRL, DDX55 and TAF5L clones. Gray bars indicate the CTCF depletion, blue bars indicate the siRNA depletions and red bars indicate the clones.

nature research

Corresponding author(s): Job Dekker

Last updated by author(s): Oct 25, 2022

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Hi-C and ChIPseq sequencing data was collected using HCS v3.4.0. CRISPR screen sequencing data was collected using HCS v2.2.68
Data analysis	Code used in this study can be found on Github: https://github.com/dekkerlab/ALV-repo.git We also used the following scripts for Hi-C analysis: Mapping and processing of Hi-C data: https://github.com/open2c/distiller-nf v0.3.3 https://github.com/open2c/pairtools v1.0.0 Storing processed Hi-C data: https://github.com/open2c/cooler v0.8.11 Analyzing Hi-C data: https://github.com/open2c/cooltools v0.5.0 https://github.com/open2c/bioframe v0.3.1 https://github.com/open2c/coolpuppy v0.9.5 pybbi v0.3.2 clodius v0.3.5 bedtools v2.29.2 HiGlass v1.11.7 matplotlib v3.5.2 ChIP-seq analysis: nf-core/chipseq v1.1.0 macs2 v2.2.7.1 meme v5.0.5

homer v4.6
intervene venn v0.6.4
bedtools v2.29.2

RNA-seq analysis:
DolphinNext: https://doi.org/10.1186/s12864-020-6714-x
STAR v2.6.1
RSEM v1.3.1
DESeq2 v3.15
DEBrowser v1.20.0
Trimmomatic v0.32
RMATS v4.1.0

CRISPR screen analysis:
bowtie v0.12.8
bagel v0.91
scikit-learn v0.23.0
scipy v1.5.2
R v4.1.0
enrichr : https://maayanlab.cloud/Enrichr/

FACS analysis:
FlowJo v10

Western blot image analysis:
Image Lab 6.0.1 builder 34

qPCR analysis:
StepOne™ Plus v2.3

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO SuperSeries accession number GSE180691. This SuperSeries is composed of the following SubSeries: GSE180922 (Hi-C), GSE180713 (RNA-seq), GSE180690 (ChIP-seq), GSE180657 (CRISPR screen). The following published datasets were used in this study (Supplementary Table 7): GSE72800, GSE110133, GSE70189, GSE104334, GSE104888, GSE95015, ENCODE: https://www.encodeproject.org/experiments/ENCSR131DVD/, ENCODE: https://www.encodeproject.org/experiments/ENCSR620QNS/, ENCODE: https://www.encodeproject.org/files/ENCFF176NSX/@@download/ENCFF176NSX.bigWig, ENCODE: https://www.encodeproject.org/files/ENCFF364QXM/. Source data are provided with this paper. All other data supporting the findings of this study are available from the corresponding author on reasonable request. All raw fastq files for CRISPR screens, RNA-seq, ChIP-seq and Hi-C data and their related Figures can be found in Supplementary Tables 2, 3, 4 and 5.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. The use of two replicates is common practice in molecular biology and genomics, given the cost and statistics are needed to be well balanced (PMID:25317452).
Data exclusions	No data were excluded from these analyses since all the data generated in this study passed quality control.
Replication	All experiments have at least two independent biological replicates. All findings described in the manuscript were confirmed in all individual replicates.
Randomization	Randomization of this study was not necessary as we did not allocate datasets into experimental groups.

Blinding

All analyses did not require blinding because results were directly linked with the data and this is neither a clinical study with large cohorts nor a genetics study with large numbers of samples.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Antibodies used in this study were:
anti-SMC1A, Bethyl Laboratories, A300-055A, lot#6, 1:2000, for Western blot
anti-RAD21, Abcam, ab992, lot#GR3310168-3, 1:1000, for Western blot and ChIP-seq
anti-CTCF, Cell signaling Technology, 2899, lot#2, 1:1000, for Western blot and ChIP-seq
anti-RPB1, Santa Cruz Biotechnology, N20 sc-899, lot#C2803, 1:500, for Western blot
anti-DDX55, Bethyl Laboratories, A303-027A, lot#A303-027A-1, 1:1000, for Western blot and ChIP-seq
anti-TAF5L, Proteintech, 19274-1-AP, 1:500, for Western blot and ChIP-seq
anti-TAF6L, ABclonal, A14369, lot#0067950201, 1:1000, for Western blot
anti-B-ACTIN, Cell signaling Technology, 8H10D10, lot#17, 1:2000, for Western blot
anti-rabbit-IgG-HRP, Cell signaling Technology, 7074, lot#31, 1:1000, for Western blot
anti-mouse-IgG-HRP, Cell signaling Technology, 7076, lot#36, 1:1000, for Western blot
anti-rabbit IgG HRP (Conformation Specific), Cell signaling Technology, 5127, lot#8, 1:1000, for Western blot
normal Rabbit IgG, Cell signaling Technology, 2729, lot#1, 1mg/mL, for co-IP

Validation

All primary antibodies were shown to react with the appropriate human proteins on the manufacturer's websites using Western blotting, co-IP and ChIP-seq. All the Western blot results showed the target protein in predicted molecular weights.
These antibodies have been validated by the supplier:
anti-SMC1A: https://www.fortislife.com/products/primary-antibodies/rabbit-anti-smc1-antibody/BETHYL-A300-055
anti-RAD21: https://www.abcam.com/rad21-antibody-ab992.html
anti-CTCF: https://www.cellsignal.com/products/primary-antibodies/ctcf-antibody/2899
anti-RPB1: https://datasheets.scbt.com/sc-899.pdf
anti-DDX55: https://www.fortislife.com/products/primary-antibodies/rabbit-anti-ddx55-antibody/BETHYL-A303-027
anti-TAF5L: https://www.ptglab.com/products/TAF5L-Antibody-19274-1-AP.htm
anti-TAF6L: https://abclonal.com/catalog-antibodies/TAF6LRabbitpAb/A14369
anti-B-ACTIN: https://www.cellsignal.com/products/primary-antibodies/b-actin-8h10d10-mouse-mab/3700
anti-rabbit-IgG-HRP: https://www.cellsignal.com/products/secondary-antibodies/anti-rabbit-igg-hrp-linked-antibody/7074
anti-mouse-IgG-HRP: https://www.cellsignal.com/products/secondary-antibodies/anti-mouse-igg-hrp-linked-antibody/7076
anti-rabbit IgG HRP: https://www.cellsignal.com/products/secondary-antibodies/mouse-anti-rabbit-igg-conformation-specific-l27a9-mab-hrp-conjugate/5127
normal rabbit IgG: https://www.cellsignal.com/products/primary-antibodies/normal-rabbit-igg/2729

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

Human HAP1 cell line was purchased from Horizon Discovery (C859).
HAP1-CTCFdegron, HAP1-CTCFdegron-TIR1, HAP1-RPB1-AID, DDX55 and TAF5L knock-out clones were generated from the Human HAP1 cell line (C859) as described in Methods.
Human HCT116-RAD21-AID cells were a gift from Masato Kanemaki (Natsume et al., Cell Rep, 2016).
HEK293T cell line was obtained from ATCC (CRL-3216).

Authentication

For Human HAP1 cell line, authentication was performed by inspection of Hi-C data obtained from HAP1 cells and detection of the appropriate translocations for this cell line.
For HAP1-CTCFdegron, HAP1-CTCFdegron-TIR1, HAP1-RPB1-AID, DDX55 and TAF5L knock-out clones, authentication was performed by inspection of Hi-C data and detecting the translocations known for the HAP1 cell line. PCR assays and sequencing were performed to detect the appropriate genetic modifications. Western blots were performed to detect the CTCF and RPB1 depletions after auxin treatment in HAP1-CTCFdegron-TIR1 and HAP1-RPB1-AID and to detect the depletion of DDX55 and TAF5L in DDX55 and TAF5L knock-out clones.

For Human HCT116-RAD21-AID cells, authentication was performed by Western blot to detect the RAD21 depletion after auxin treatment.
For HEK293T cells, authentication was performed by STR profiling at the Centre for Applied Genomics (TCAG) at the Hospital for Sick Children (SickKids) in Toronto.

Mycoplasma contamination

Cell lines were routinely tested for mycoplasma infection and tested negative (MycoAlertTM Mycoplasma Detection Kit, Lonza).

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used in this study.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

The data datasets generated in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO SuperSeries accession number GSE180691. This SuperSeries is composed of the following SubSeries: GSE180922 (Hi-C), GSE180713 (RNA-seq), GSE180690 (ChIP-seq), GSE180657 (CRISPR screen).

Files in database submission

ChIP-442-IAA-DDX55-R2-T1_S9_L001_R1_001.fastq
ChIP-442-IAA-DDX55-R2-T1_S9_L001_R2_001.fastq
ChIP-442-IAA-RAD21-R2-T1_S8_L001_R1_001.fastq
ChIP-442-IAA-RAD21-R2-T1_S8_L001_R2_001.fastq
ChIP-442-IAA-TAF5L-R2-T1_S10_L001_R1_001.fastq
ChIP-442-IAA-TAF5L-R2-T1_S10_L001_R2_001.fastq
ChIP-442-IAA-input-R2-T1_S6_L001_R1_001.fastq
ChIP-442-IAA-input-R2-T1_S6_L001_R2_001.fastq
ChIP-442-NT-DDX55-R2-T1_S4_L001_R1_001.fastq
ChIP-442-NT-DDX55-R2-T1_S4_L001_R2_001.fastq
ChIP-442-NT-RAD21-R2-T1_S3_L001_R1_001.fastq
ChIP-442-NT-RAD21-R2-T1_S3_L001_R2_001.fastq
ChIP-442-NT-TAF5L-R2-T1_S5_L001_R1_001.fastq
ChIP-442-NT-TAF5L-R2-T1_S5_L001_R2_001.fastq
ChIP-442-NT-input-R2-T1_S1_L001_R1_001.fastq
ChIP-442-NT-input-R2-T1_S1_L001_R2_001.fastq
ChIP-442-IAA-CTCF-R1-T1_S23_L005_R1_001.fastq
ChIP-442-IAA-CTCF-R1-T1_S23_L005_R2_001.fastq
ChIP-442-IAA-DDX55-R3-T1_S25_L005_R1_001.fastq
ChIP-442-IAA-DDX55-R3-T1_S25_L005_R2_001.fastq
ChIP-442-IAA-RAD21-R3-T1_S24_L005_R1_001.fastq
ChIP-442-IAA-RAD21-R3-T1_S24_L005_R2_001.fastq
ChIP-442-IAA-TAF5L-R3-T1_S26_L005_R1_001.fastq
ChIP-442-IAA-TAF5L-R3-T1_S26_L005_R2_001.fastq
ChIP-442-IAA-input-R3-T1_S22_L005_R1_001.fastq
ChIP-442-IAA-input-R3-T1_S22_L005_R2_001.fastq
ChIP-442-NT-CTCF-R1-T1_S18_L005_R1_001.fastq
ChIP-442-NT-CTCF-R1-T1_S18_L005_R2_001.fastq
ChIP-442-NT-DDX55-R3-T1_S20_L005_R1_001.fastq
ChIP-442-NT-DDX55-R3-T1_S20_L005_R2_001.fastq
ChIP-442-NT-TAF5L-R3-T1_S21_L005_R1_001.fastq
ChIP-442-NT-TAF5L-R3-T1_S21_L005_R2_001.fastq
ChIP-442-NT-input-R3-T1_S17_L005_R1_001.fastq
ChIP-442-NT-input-R3-T1_S17_L005_R2_001.fastq
ChIP-442-IAA-CTCF-R2-T1_S13_L003_R1_001.fastq
ChIP-442-IAA-CTCF-R2-T1_S13_L003_R2_001.fastq
ChIP-442-IAA-input-R4-T1_S12_L003_R1_001.fastq
ChIP-442-IAA-input-R4-T1_S12_L003_R2_001.fastq
ChIP-442-NT-CTCF-R2-T1_S10_L003_R1_001.fastq
ChIP-442-NT-CTCF-R2-T1_S10_L003_R2_001.fastq
ChIP-442-NT-RAD21-R4-T1_S11_L003_R1_001.fastq
ChIP-442-NT-RAD21-R4-T1_S11_L003_R2_001.fastq
ChIP-442-NT-input-R4-T1_S9_L003_R1_001.fastq
ChIP-442-NT-input-R4-T1_S9_L003_R2_001.fastq
NT-CTCF_R1.mLb.clN.bigWig
IAA-CTCF_R1.mLb.clN.bigWig
NT-CTCF_R2.mLb.clN.bigWig
IAA-CTCF_R2.mLb.clN.bigWig
NT-RAD21_R1.mLb.clN.bigWig
IAA-RAD21_R1.mLb.clN.bigWig
NT-RAD21_R2.mLb.clN.bigWig
IAA-RAD21_R2.mLb.clN.bigWig

NT-DDX55_R1.mLb.clN.bigWig
IAA-DDX55_R1.mLb.clN.bigWig
NT-DDX55_R2.mLb.clN.bigWig
IAA-DDX55_R2.mLb.clN.bigWig
NT-TAF5L_R1.mLb.clN.bigWig
IAA-TAF5L_R1.mLb.clN.bigWig
NT-TAF5L_R2.mLb.clN.bigWig
IAA-TAF5L_R2.mLb.clN.bigWig
NT-input_R2.mLb.clN.bigWig
IAA-input_R2.mLb.clN.bigWig
NT-input_R3.mLb.clN.bigWig
IAA-input_R3.mLb.clN.bigWig
NT-input_R4.mLb.clN.bigWig
IAA-input_R4.mLb.clN.bigWig
NT-CTCF-narrowPeaks-sort-merge-with-strength.bed
IAA-CTCF-narrowPeaks-sort-merge-with-strength.bed
NT-RAD21-narrowPeaks-sort-merge-with-strength.bed
IAA-RAD21-narrowPeaks-sort-merge-with-strength.bed
NT-DDX55-narrowPeaks-sort-merge-with-strength.bed
IAA-DDX55-narrowPeaks-sort-merge-with-strength.bed
NT-TAF5L-narrowPeaks-sort-merge-with-strength.bed
IAA-TAF5L-narrowPeaks-sort-merge-with-strength.bed

Genome browser session
(e.g. UCSC)

N/A, list of ChIP-seq peaks in presence or absence of CTCF can be uploaded to genome browser.

Methodology

Replicates

All ChIP-seq were performed on two biological replicates.

Sequencing depth

At least 55M reads were obtained fore each ChIP-seq libraries. The libraries were sequenced using 50bp paired end reads. Sequencing depth and other metrics are provided in Supplementary Table 4.

Antibodies

Antibodies used for ChIP-seq:
anti-RAD21, Abcam, ab992, lot#GR3310168-3, 4uL:5M cells
anti-CTCF, Cell signaling Technology, 2899, lot#2, 20uL:5M cells
anti-DDX55, Bethyl Laboratories, A303-027A, lot#A303-027A-1, 4uL:5M cells
anti-TAF5L, Proteintech, 19274-1-AP, 12uL:5M cells

Peak calling parameters

We used MACS2 to find the enriched ChIP-seq peaks with the following parameters: macs2 callpeak -q 0.01. Input was used as background.

Data quality

Sufficient sequencing depth was achieved according to ENCODE standards. Number of peaks obtained was comparable to published datasets.

Software

Softwares used for ChIP-seq analysis:
nf-core/chipseq v1.1.0
macs2 v2.2.7.1
meme v5.0.5
homer v4.6
intervene venn v0.6.4
bedtools v2.29.2

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Cells were dissociated with accutase (ThermoFisher Scientific, A11105-01), resuspended in PBS, spun, and resuspended in 250µL of cold PBS. To assess the cell cycle profile (DNA content), 750µL of 100% ethanol was slowly added to fix cells in 75% ethanol. Cells were stored in -20°C for at least 24 hours. Fixed cells were spun, re-suspended in 1X PBS with propidium iodide (PI) (final concentration 50µg/mL) and RNAseA (0.5mg/mL) and incubated for 30 minutes at room temperature protected from light. To assess GFP content, cells were washed once with PBS and fixed with 4% PFA for 10 minutes. Cells were spun and cell pellets

	were resuspended in 1mL of PBS. Cells were sorted on a FACSCALIBUR or LSRII or MACSQUANT. Analysis was performed using the Flowjo software.
Instrument	FACSCALIBUR or LSRII or Miltenyi MACSQuant VYB
Software	Data was analyzed using FlowJo v10 software
Cell population abundance	Cells were not sorted, only analyzed using flow cytometry.
Gating strategy	Cells were initially separated by side scatter area (y-axis) versus forward scatter area (x-axis) plot to define the population of 'live' cells. Single cells were then selected by forward scatter height (y-axis) versus forward scatter area (x-axis). Cells were represented on a histogram showing propidium iodide intercalation. Gates for G1, S and G2/M populations were finally defined. See gating strategy in Supplementary Fig. 1.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.