Cohesin acts as a transcriptional gatekeeper by restraining pause–release to promote processive elongation

Shoin Tei,[1] Toyonori Sakata,[1,2], Atsunori Yoshimura,[1] Toyoaki Natsume,[3,4,5] Masato T Kanemaki,[3,4,6] Takashi Sutani,[1] Masashige Bando,[1] and Katsuhiko Shirahige[1,2*]

[1]Laboratory of Genome Structure and Function, Institute for Quantitative Biosciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-Ku, Tokyo, 113-0032, Japan. [2]Karolinska Institutet, Department of Cell and Molecular Biology, Biomedicum, Tomtebodavägen 16, Stockholm 171 77, Sweden. [3]Department of Chromosome Science, National Institute of Genetics, Research Organization of Information and Systems (ROIS), Yata 1111, Mishima, Shizuoka 411-8540, Japan. [4]Graduate Institute for Advanced Studies, SOKENDAI, Yata 1111, Mishima, Shizuoka 411-8540, Japan. [5]Present address: Research Center for Genome & Medical Sciences, Tokyo Metropolitan Institute of Medical Science, Tokyo, Japan. [6]Department of Biological Science, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan

e-mail: kshirahi@iqb.u-tokyo.ac.jp

Abstract

Cohesin organizes 3D chromatin architecture, including promoter–enhancer loops, yet its loss causes surprisingly modest changes in steady-state gene expression. We resolve this paradox by demonstrating that cohesin performs two opposing but balancing functions through genome-wide analyses and in vitro transcription complex reconstitution. First, it promotes transcription initiation by facilitating enhancer–promoter communication that maintains active promoter chromatin. Second, it acts as a transient gate that delays pause release by briefly associating with the transcriptional machinery during the pause–

release transition. This gating occurs downstream of CDK9 and is relieved by cohesin's ATPase-dependent turnover. Kinetic modelling shows that minimal changes in steady-state gene expression result from the counterbalance between reduced initiation and enhanced pause release. In contrast, cohesin depletion impairs robust transcriptional induction in response to external stimuli. Moreover, the cohesin-mediated delay in pause release serves as a quality-control-like step by promoting elongation complex assembly and transcription processivity. These insights have important implications for cohesin-related diseases, including cancers and cohesinopathies, and broader principles of transcriptional regulation.

Introduction

Cohesin is a conserved ring-shaped SMC complex whose loading onto chromatin is mediated by the NIPBL–MAU2 heterodimer, which also stimulates cohesin's ATPase activity[1–4]. During cell division, cohesin ensures proper chromosome segregation by holding replicated sister chromatids together from S phase to anaphase. In interphase, cohesin organizes topologically associated domains (TADs) and chromatin loops, and its depletion severely disrupts these higher-order chromatin structures[5–9]. Given these architectural roles, cohesin has been proposed to regulate gene expression by modulating promoter-enhancer communication[10]. However, gene deletion or protein knockdown of cohesin affects the steady-state expression of only a small fraction of genes[5,11].

Nevertheless, observations from Cornelia de Lange syndrome (CdLS) and CHOPS syndrome suggest a transcriptional role for cohesin. CdLS is a developmental disorder caused by loss-of-function mutations in cohesin subunits and their regulatory factors[12,13]. CHOPS syndrome, in contrast, results from gain-of-function mutations in AFF4[14], the scaffold protein of the super elongation complex (SEC) that promotes RNA polymerase II (Pol II) pause release[15,16]. Notably, CdLS and CHOPS syndrome share transcriptional and phenotypic abnormalities, suggesting a common defect in transcriptional regulation. In metazoan transcription, Pol II frequently pauses shortly after initiation, stabilized by pausing factors[17,18]. Upon phosphorylation by CDK9 in SEC, Pol II is released from pausing and enters productive elongation. The

shared features of these diseases raised the possibility that cohesin also regulates Pol II pause release, emphasizing the importance of analyzing cohesin's role at distinct stages of the transcription cycle.

We investigated cohesin's transcriptional function by genome-wide profiling of transcription dynamics, histone modifications, and 3D genome architecture, together with in vitro reconstitution of the transcription machinery. Our analyses reveal that cohesin performs two opposing but balanced functions: (i) it promotes transcription initiation by facilitating enhancer–promoter communication; and (ii) it stabilizes Pol II pausing by briefly associating with the transcriptional machinery. When cohesin is depleted, reduced initiation and enhanced pause release counterbalance each other, explaining the minimal changes in steady-state gene expression. In contrast, cohesin depletion impairs robust transcriptional induction in response to external stimuli. Moreover, cohesin's transient engagement with the transcription machinery acts as a quality control step, delaying pause release to ensure proper elongation and prevent premature transcription termination. These insights have important implications for cohesin-related diseases, including cancers and cohesinopathies, and broader principles of transcriptional regulation.

Results

Cohesin is a major determinant of 3D chromatin structure

To investigate how cohesin contributes to transcriptional regulation, we acutely depleted RAD21, a cohesin subunit, in HCT-116 cells expressing an auxin-inducible degron (AID)-tagged RAD21[5,19]. Western blotting confirmed that 3-hour auxin treatment reduced RAD21 protein levels to ~6%, and ChIP-qPCR at three RAD21 ChIP-seq peaks showed that cohesin binding was reduced to ~11% (Supplementary Fig. 1a, b). FACS analysis revealed that this treatment had minimal impact on cell cycle progression (Supplementary Fig. 1c).

To examine how cohesin depletion affects higher-order genome organization, we performed Micro-C and obtained ~1.99 billion valid contacts (Supplementary Fig. 1d). As previously reported[5,20], rapid cohesin depletion altered contact distributions, decreasing short-range (<1 Mb) and increasing long-range contacts (Supplementary Fig. 1e). The number of TADs decreased from 4,537 to 2,199, and chromatin loops from 135,539 to 54,332, with the remaining structures exhibiting reduced interaction strength (Supplementary Fig. 1f-h). Upon cohesin depletion, the proportion of loops lacking cohesin peaks at their anchors increased from 40% to 62%, and their median length shortened from 78 kb to 34 kb (Supplementary Fig. 1f, i). These observations indicate that cohesin depletion preferentially disrupts cohesin-mediated loops, which are typically longer (~165 kb) than loops without cohesin (~28 kb) (Supplementary Fig. 1j). Accordingly, cohesin-anchored loops exhibited the most pronounced reduction in contact intensity after depletion (Supplementary Fig. 1k). Together, these results confirm cohesin as a key organizer of 3D genome architecture.

Cohesin depletion affects the steady-state expression of a limited number of genes

To assess how acute cohesin depletion affects gene expression, we profiled both mature and nascent transcripts using RNA-seq, TT-seq, and EU-seq. TT-seq and EU-seq detect newly synthesized RNAs through 4-thiouridine or 5-ethynyluridine (EU) metabolic labeling, respectively. Among 3,660 well-expressed, non-overlapping genes longer than 5 kb, we identified a limited number of differentially expressed genes upon cohesin depletion: 135 by TT-seq, 83 by RNA-seq, and 182 by EU-seq (Fig. 1a, Supplementary Fig. 2a, b). Of these, 28 genes were shared across all three assays (Supplementary Fig. 2c). Similarly, RNA-seq in SMC1-depleted cells revealed 85 genes with altered expression, 50 of which overlapped with those identified in RAD21-depleted cells (Supplementary Fig. 2d-g). Consistent with previous reports[5,9,21], these findings indicate that rapid cohesin depletion has only a modest effect on the transcriptome.

Cohesin promotes Pol II pausing at highly paused genes

Given this modest transcriptional impact of cohesin loss, we next examined whether cohesin depletion altered the distribution of Pol II along genes. ChIP-seq for RPB1, the largest subunit of Pol II, revealed a reduction in Pol II occupancy specifically at gene promoters, whereas occupancy along gene bodies remained largely unchanged (Supplementary Fig. 3a). To further investigate this decrease at nucleotide resolution, we performed mNET-seq, which maps transcriptionally engaged Pol II in a strand-specific manner by identifying the last incorporated nucleotide[22], using an antibody recognizing total Pol II. Consistent with the ChIP-seq results, mNET-seq signal decreased at promoters but remained unchanged across gene bodies (Fig. 1b). For more detailed analysis, we defined the +1 nucleosome as the first H3K4me3-marked nucleosome downstream of the transcription start site (TSS) and identified the pausing site as the highest mNET-seq peak between the TSS and that nucleosome (Supplementary Fig. 3b). Using this definition, promoter-focused analysis revealed a specific reduction in mNET-seq signal at pausing sites upon cohesin depletion (Fig. 1c), while TT-seq remained unchanged, suggesting reduced pausing rather than decreased transcription. To test this quantitatively, we evaluated pausing with a pausing-duration index—the mNET-seq/TT-seq ratio at pausing sites (defined over a ±25 bp window around the peak pausing site; Fig. 1d)[23,24]. This index reflects the dwell time of Pol II at pause sites in proportion to transcript output. Cohesin loss reduced this index, with the largest decrease at highly paused genes, confirming weakened pausing.

To confirm the reduction in Pol II pausing following cohesin depletion, we monitored pause release by measuring the Ser2P/RPB1 ratio at promoters, as Ser2 phosphorylation (Ser2P) marks Pol II that has exited the paused state[25]. The Ser2P/RPB1 ratio was negatively correlated with the pausing-duration index, supporting the validity of this measurement, and was increased after cohesin loss, corroborating reduced pausing (Supplementary Fig. 3c, d). As an alternative approach to quantify pausing, we estimated the half-life of paused Pol II by tracking the decay of Pol II promoter binding after triptolide treatment, which blocks new Pol II loading (Supplementary Fig. 3e)[26,27]. Cohesin depletion shortened the average half-life from 14.4 min to 10.3 min, consistent with weakened pausing (Supplementary Fig. 3f, g). Next, we evaluated Pol II pausing using RPB1 ChIP-seq by calculating a pausing index—promoter

occupancy divided by gene-body occupancy (Supplementary Fig. 3h). In line with the pausing-duration index, this index decreased after RAD21 or SMC1 depletion (Supplementary Fig. 3i, j). In contrast, the index increased upon WAPL knockdown, which elevates chromatin-bound cohesin by preventing its dissociation (Supplementary Fig. 3k-m)[28,29]. Together, these three independent approaches demonstrate that cohesin stabilizes Pol II pausing and that its loss leads to genome-wide reduction in pausing.

Given the reduced pausing upon cohesin depletion, we investigated how cohesin loss affects the pausing complex, which includes paused Pol II and its associated factors. We measured promoter occupancy relative to Pol II for NELFCD, a subunit of the NELF complex that stabilizes pausing, and for CDK9 and AFF4, components of the SEC that promote pause release. To capture the pausing state, we blocked pause release using a CDK9 inhibitor. In the absence of CDK9 inhibition, cohesin depletion decreased NELFCD but increased CDK9 and AFF4 levels at promoters (Supplementary Fig. 3n). In contrast, CDK9 inhibition restored NELFCD occupancy, whereas CDK9 and AFF4 remained high (Supplementary Fig. 3o). The rescue of NELFCD by CDK9 inhibition indicates that cohesin depletion does not impair the association between paused Pol II and NELF. Instead, the initial reduction in NELFCD likely reflects shortened pausing, as NELF dissociates when Pol II is released from pausing[30]. For SEC, the sustained increase in CDK9 and AFF4 suggests that cohesin depletion enhanced the association of SEC and Pol II.

Cohesin constrains early Pol II elongation

Given that cohesin loss shortens promoter pausing, we next examined its effect on elongating Pol II. We first quantified elongation rate using a velocity index—the TT-seq/mNET-seq ratio across gene bodies[31]. This index reflects RNA output per engaged Pol II, such that higher values indicate faster elongation (Supplementary Fig. 4a). Analysis in 1-kb bins revealed increased velocity from the TSS to +8 kb upon cohesin depletion (Supplementary Fig. 4b). To assess velocity changes across entire gene, we divided each gene into three equal regions from the TSS to the gene end (5’, Mid, and 3’). Elevated velocity near the TSS gradually returned to baseline toward the 3’ end, especially in longer genes (Supplementary Fig.

4c). To independently confirm accelerated elongation, we analyzed intronic RNA-seq coverage slopes, where co-transcriptional splicing causes a 5′ to 3′ decline in coverage[32]. In both RAD21- and SMC1-deleted cells, introns within 5 kb of the TSS showed shallower slopes, indicating faster elongation, whereas downstream introns did not (Supplementary Fig. 4d). Together, these two independent assays demonstrate that cohesin depletion accelerates Pol II progression immediately downstream of the TSS.

Cohesin stabilizes the elongation complex to suppress premature transcription termination

We next asked whether cohesin depletion alters the composition of the elongation complex. Elongation factors such as PAF1 complex, SPT6, and FACT complex facilitate processive and stable transcription by interacting with elongating Pol II[31,33]. To quantify these associations, we performed ChIP-seq for each factor and measured gene occupancy relative to Pol II. This analysis showed that cohesin loss reduced the relative level of PAF1 complex (PAF1 and CDC73) (Fig. 2a). Similar reductions were observed for SPT6 and FACT (SSRP1) (Supplementary Fig. 4e). Moreover, co-immunoprecipitation with antibodies against Pol II CTD and Ser2P further revealed that cohesin depletion weakened Pol II interaction with PAF1 components (PAF1, CDC73, or CTR9) by ~55–75% (Fig. 2b). Because PAF1 associates with Pol II only after pause release[34], these associations were undetectable during DRB treatment, which blocks Pol II release, but appeared after DRB wash-out. Together, these findings indicate that cohesin promotes the maturation of the elongation complex.

Given the incomplete assembly of the elongation complex upon cohesin depletion, we hypothesized that this would promote early transcription termination by impairing transcriptional processivity. To test this, we performed RNA-seq following co-depletion of cohesin and XRN2, a 5’-3’ exoribonuclease that degrades prematurely terminated transcripts (Supplementary Fig. 4f)[35]. As expected from its role in transcription termination, XRN2 depletion alone modestly increased signals downstream of 3′ gene ends, with a slightly greater enhancement observed in the absence of cohesin (Fig. 2c, Supplementary Fig. 4g). In contrast, regions upstream of the 3′ ends showed increased RNA-seq signal upon XRN2 knockdown only

when cohesin was depleted. This upstream accumulation was more pronounced at longer genes. To further assess transcriptomic changes, we identified genes whose expression was significantly altered in each condition. A total of 666 genes were differentially expressed only upon combined depletion, exceeding the total number of affected genes from either single condition (Fig. 2d, Supplementary Fig. 4h). At these synergistically upregulated genes, expression levels exceeded the additive effects of individual depletions (Fig. 2e). Notably, the most prominent increases occurred just upstream of the 3' ends, consistent with the genome-wide accumulation pattern seen upon co-depletion. In contrast, downregulated genes exhibited expression changes consistent with the combined effects of both treatments (Supplementary Fig. 4i). Together, these findings suggest that cohesin depletion leads to the generation of unstable transcripts within genes, likely due to the immature elongation complex, and the XRN2-mediated degradation pathway normally eliminates these transcripts.

Cohesin and its loader dynamically colocalize with RNA Polymerase II at promoters

Cohesin dynamically moves along chromosomes and is enriched at CTCF sites, enhancers, and promoters[36]. The cohesin loader, the NIPBL–MAU2 heterodimer, is required not only for loading cohesin onto DNA but also for cohesin's translocation with loop formation. Given the observed impact of cohesin depletion on Pol II pausing and elongation, we next investigated how cohesin and its loader are positioned relative to Pol II on chromatin. We performed ChIP-seq to profile RPB1, RAD21, and the loader subunits NIPBL and MAU2 under three conditions: untreated, DRB-treated, and DRB washout. DRB treatment blocks pause release, resulting in Pol II accumulation at promoters (Fig. 3a). DRB washout relieves the block, allowing Pol II to enter downstream gene bodies, thereby reducing its promoter occupancy (Supplementary Fig. 5a). RAD21, NIPBL, and MAU2 behaved similarly, albeit to a lesser extent, accumulating at promoters during DRB treatment and showing decreased promoter occupancy after washout. These changes indicate that cohesin and its loader associate with promoters in a transcription-dependent manner, similar to Pol II.

We next tested whether the cohesin loader physically associates with Pol II. Co-immunoprecipitation of FLAG-tagged endogenous NIPBL revealed its association with RPB1 and with its loader partner MAU2

(Fig. 3b)[36,37]. Triptolide treatment, which induces degradation of Pol II, abolished the NIPBL-Pol II interaction even though the NIPBL–MAU2 complex remained intact (Fig. 3c). Moreover, the CDK9 inhibitor flavopiridol greatly weakened the NIPBL–Pol II association, indicating that the interaction preferentially occurs downstream of CDK9 activation. Given their dynamic colocalization at promoters observed by ChIP–seq, it is likely that cohesin also interacts with Pol II, similarly to its loader. However, co-immunoprecipitation using antibodies against cohesin subunits did not detect a clear interaction. This likely reflects the fact that only a small fraction of total cellular cohesin is associated with Pol II. Thus, we concluded that cohesin and its loader dynamically colocalize with transcriptionally engaged Pol II at promoters, and that CDK9 activity modulates their interaction, making the association between the loader and Pol II detectable by immunoprecipitation.

Cohesin and its loader associate with the transcription complex during the pause–release transition of RNA polymerase II

Given their colocalization and physical association with Pol II, we investigated whether cohesin and its loader act as components of the transcriptional machinery. We assembled a transcriptional complex on a DNA template containing GAL4-binding sites and TATA promoter in HeLa nuclear extract (Fig. 4a). Upon activation with GAL4-VP16, we observed the recruitment of the cohesin loader, Mediator complex, general transcription factors, Pol II, and both pausing-associated (SPT5, NELFA) and release-associated (AFF4, CDK9) factors to the template (Fig. 4b). NTP addition induced Pol II Ser5 (Ser5P) and Ser2 phosphorylation (Ser2P) of Pol II, whereas DRB selectively inhibited Ser2P. These results confirm that the reconstituted system recapitulates transcription factor recruitment upon gene activation and captures Pol II phosphorylation dynamics (Supplementary Fig. 6a). Although the loader was present, cohesin itself was not detectably recruited to the template. Using recombinant NIPBL fragments, we found that the N-terminal region of NIPBL is required for loader association upon transcriptional activation (Supplementary Fig. 6b-d). Furthermore, Mediator depletion abolished the recruitment of both the loader and Pol II, suggesting that loader recruitment depends on active transcription rather than direct binding by GAL4–VP16 (Supplementary Fig. 6e).

Next, we inhibited cohesin's ATPase activity with aluminum fluoride (AlFx) to assess its role in transcription[38,39]. AlFx, in complex with ADP, mimics the transition state of ATP hydrolysis, thereby blocking cohesin's ATPase cycle. Cohesin's ability to entrap and translocate along DNA depends on its ATPase activity, which is activated by the cohesin loader[36]. AlFx treatment markedly enhanced the recruitment of cohesin, its loader, Pol II, and both pausing- and release-associated factors, making cohesin readily detectable (Supplementary Fig. 6f). This effect was most robust when AlFx was added after ADP, compared to addition before or simultaneous with ADP. Cohesin depletion and CDK9 inhibition abolished the complex (Fig. 4c, d). These findings indicate that: (i) cohesin and its loader are parts of the transcriptional complex, along with pausing- and release-associated factors; (ii) cohesin is required for the assembly of this AlFx-induced complex; and (iii) this complex is formed downstream of CDK9 activity, as shown by the enhanced formation upon AlFx addition after CDK9 activation by ADP, and its disruption upon CDK9 inhibition. Collectively, we propose that the AlFx-induced complex represents a transient intermediate state between pausing and pause release, as the NELF complex has been reported to dissociate at the time of Pol II pause release (Fig. 4e). In this model, ATP hydrolysis by cohesin triggers its disassembly as ATPase inhibition with AlFx allowed detection of this otherwise transient state. This transient intermediate may mechanistically explain the shortened Pol II pausing upon cohesin depletion (Fig. 1d). The loss of cohesin likely prevents formation of this state, thereby promoting premature pause release.

Cohesin maintains the H3K27ac mark at promoters

Because H3K27ac is a hallmark of active promoters[40], we investigated whether cohesin depletion altered this modification. ChIP-seq revealed that cohesin loss significantly reduced H3K27ac at 92% of promoters (FDR < 0.05; Fig. 5a, b). Given that chemical inhibition of H3K27ac is known to diminish promoter binding of the transcription initiation factor TFIID[41], we tested whether cohesin loss affects TFIID occupancy. Indeed, promoter occupancy of TFIID subunits TAF1 and TBP decreased at 76% and 91% of promoters, respectively (Fig. 5c, Supplementary Fig. 7a-c). As expected from this relationship, H3K27ac and TAF1 levels at promoters were strongly correlated ($\rho = 0.77$, $P < 2 \times 10^{-1}$□), as were their log□ fold-changes

upon cohesin loss ($\rho = 0.49$, $P < 2 \times 10^{-1}$; Fig. 5d). These observations suggest that cohesin depletion reduces promoter activity, as indicated by decreased H3K27ac, and consequently impairs TFIID recruitment

We hypothesized that promoter H3K27ac reflects initiation efficiency. To test this idea, we compared H3K27ac enrichment with the pausing-duration index, because genes with frequent initiation often exhibit shorter Pol II pausing durations[26,42]. As expected, these two parameters were negatively correlated ($\rho = -0.35$; P-value $< 2.2 \times 10^{-16}$; Supplementary Fig. 7d). This inverse correlation, together with the similar distribution of H3K27ac and TAF1, suggests promoter H3K27ac can serve as a proxy for initiation efficiency. Together, these results support the idea that cohesin promotes transcription initiation by maintaining promoter H3K27ac levels, thereby facilitating initiation complex recruitment.

Cohesin regulates the genomic distribution of H3K27ac via promoter–enhancer contacts

Given that cohesin depletion reduces H3K27ac at promoters, we next examined how it affects H3K27ac distribution across both promoters and enhancers, as this modification marks both regulatory elements. We annotated promoters and enhancers genome-wide using ChromHMM[43,44]. Among 40,142 combined regulatory elements, H3K27ac levels increased at 6,887 sites and decreased at 18,728 upon cohesin depletion (FDR < 0.05; Fig. 6a, Supplementary Fig. 8a). These changes were validated by ChIP–qPCR at four representative regions, two with increased and two with decreased H3K27ac (Supplementary Fig. 8b). Elements with increased H3K27ac were enriched at H3K4me1-marked enhancers, whereas those exhibiting decreases were primarily localized at H3K4me3-positive promoters, consistent with the reduction in promoter H3K27ac we observed (Supplementary Fig. 8c). BRD4, MED1, mNET-seq and TT-seq signals exhibited similar increases and decreases, while H3K4me1/3 levels themselves remained unchanged (Supplementary Fig. 8d). Next, we investigated the genomic distribution of H3K27ac-altered elements. Elements with the same H3K27ac response, particularly increased elements, were more often positioned adjacent to each other than expected by chance, assuming a uniform distribution (Fig. 6b). In

contrast, increased and decreased sites were located next to each other less often than expected. This positional bias suggests a role for 3D genome architecture in the regulation of H3K27ac by cohesin.

To test this idea, we examined chromatin contacts across the three element classes using Micro-C data. While contacts were observed between all possible pairwise combinations, elements with increased (mostly enhancers) and decreased H3K27ac (mostly promoters) formed the strongest interactions, which were markedly weakened after cohesin depletion (Fig. 6c). Analysis of contact distances revealed that interactions between increased and decreased elements spanned the longest genomic distances, both before and after cohesin loss (Fig. 6d). These observations indicate that cohesin is required to maintain long-range promoter–enhancer contacts, consistent with its established role in forming long-range chromatin loops (Supplementary Fig. 1j). We conclude that cohesin loss disrupts promoter–enhancer interactions spanning large genomic distances, resulting in increased H3K27ac at enhancers and decreased H3K27ac at promoters. Moreover, the observed increases and decreases in H3K27ac and other factors imply that enhancers serve as a source of transcriptional cofactors that are delivered to promoters through cohesin-dependent contacts. When cohesin is depleted, this supply is disrupted, causing cofactors to accumulate at enhancers and diminish at promoters (Fig. 6e).

Cohesin depletion reduces initiation but enhances pause release, counterbalancing steady-state gene expression

Thus far, our data have shown that cohesin depletion shortens Pol II pausing, accelerates early elongation, and lowers promoter H3K27ac (Supplementary Fig. 9a). However, transcriptome analysis revealed only modest changes in gene expression upon the depletion. To determine how these changes collectively yield modest transcriptional effects, we analyzed their gene-level interplay. We calculated log□ fold-changes for promoter H3K27ac, pausing-duration index, velocity index, and TT-seq expression, and then applied k-means clustering. To minimize pausing and termination effects, velocity was calculated across the gene body excluding the first and last 1□kb. Unsupervised clustering classified genes into

three classes: Cluster 1 (55%) contained genes with little or unchanged expression, while Clusters 2 (23%) and 3 (2%) included genes with overall enhancement or suppression of expression upon cohesin loss, respectively (Fig. 7a, b, Supplementary Fig. 9b, c). Consistently, significantly upregulated genes were mostly found in Cluster 2, and downregulated genes in Cluster 3. Changes in promoter H3K27ac, Pol II pausing, and velocity also showed distinct patterns across the three clusters. H3K27ac decreased moderately in Cluster 1, minimally in Cluster 2, and most strongly in Cluster 3. Pol II pausing showed the inverse pattern, being slightly reduced in Cluster 1, markedly decreased in Cluster 2, and unchanged in Cluster 3. Velocity was unchanged in Cluster 1, increased in Cluster 2, and decreased in Cluster 3. These patterns suggest that promoter H3K27ac and Pol II pausing drive gene expression in opposite directions, while transcriptional velocity amplifies the effect of pausing. In Cluster 1, comparable reductions in H3K27ac and pausing cancel each other out, and velocity changes are minimal, resulting in unchanged expression. In Cluster 2, slight H3K27ac loss, a large reduction in pausing, and increased velocity combine to upregulate gene expression. Conversely, in Cluster 3, pronounced H3K27ac decrease, little change in pausing, and a decline in velocity together suppress transcription. Promoter H3K27ac reflects transcriptional initiation efficiency, while reduced Pol II pausing indicates enhanced pause release. Thus, the gene-specific interplay between diminished initiation and accelerated pause release likely accounts for the limited impact of cohesin depletion on the steady-state transcriptome.

Simulation supports a model in which reduced initiation and enhanced pause release balance each other to maintain gene expression

To examine how reduced initiation efficiency and enhanced pause release might jointly impact transcriptional output, we simulated Pol II transitions through four states: Free, Initiation, Paused, and Elongation (Supplementary Fig. 10a). We then modulated the transition probabilities between these states to reflect changes observed upon cohesin depletion. Transition path and its probability were assigned to reflect previously reported Pol II transitions and their stationary distribution across these states (Supplementary Fig. 10b)[45]. Reducing the Free-to-Initiation probability, recapitulating lower initiation observed upon cohesin loss, decreased the Elongation fraction (Supplementary Fig. 10c). In

contrast, raising the Paused-to-Elongation probability, reflecting enhanced pause release, increased the Elongation fraction. Notably, when both initiation and release probabilities were altered together, the proportion of Pol II in Elongation returned to baseline (Fig. 7c). These simulations indicate that reduced initiation and enhanced release can counterbalance each other and maintain stable gene expression, consistent with our experimental observations. We also asked how these decreases affect the pausing index, designed to mimic the Pol II ChIP-seq pausing index. In the simulation, the pausing index was defined as the ratio of Pol II in the Initiation + Paused states to that in the Elongation state. This index responded specifically to changes in the Paused-to-Elongation probability, supporting the validity of the ChIP–seq–based pausing index (Supplementary Fig. 10d).

Cohesin loss impairs rapid gene induction upon external stimuli

Because promoter-paused Pol II is thought to serve as a preloaded complex for rapid transcriptional responses to external stimuli[46,47], we asked whether cohesin depletion affects cytokine-induced gene activation. EU-seq after TNFα treatment revealed that cohesin depletion reduced gene induction upon TNFα-treatment (Fig. 7d, e, and Supplementary Fig. 9d). Before stimulation, these genes already exhibited lower pausing indices and decreased promoter H3K27ac and TAF1 (Supplementary Fig. 9e-h), suggesting that decreased pausing and/or lowered promoter H3K27ac underlie the impaired induction. These findings suggest that cohesin is required for a robust transcriptional activation in response to external stimuli, indicating its crucial roles in stimulus-induced gene regulation.

Discussion

Here, we demonstrate that cohesin exerts two distinct functions during the transcription cycle: it simultaneously promotes transcription initiation and restrains release from promoter-proximal pausing (Supplementary Fig. 11). These opposing activities resolve the long-standing paradox that cohesin

depletion has only a limited impact on steady-state gene expression, despite severely disrupting 3D genome architecture. Upon cohesin loss, reduced initiation and enhanced pause release counterbalance each other, resulting in minimal changes in expression. However, this compensatory balance is disrupted upon external stimulation, leading to a marked reduction in transcriptional induction. Our findings indicate that cohesin regulates initiation and pausing through distinct mechanisms. To support initiation, cohesin facilitates promoter-enhancer communication, as previously proposed, thereby maintaining promoter H3K27ac. In contrast, it transiently engages the transcriptional machinery to ensure an appropriate pausing duration. Sufficient pausing is required for the proper formation of the elongation complex to prevent premature transcription termination. We therefore regard this direct engagement as a quality-control-like step that coordinates elongation factor recruitment and complete transcription.

We found that cohesin ensures proper Pol II dwell time during promoter-proximal pausing. To elucidate the underlying mechanism, our biochemical experiments identified a transcriptional complex at the pausing-to-release transition that includes cohesin, its loader, and Pol II. These findings suggest that cohesin stabilizes pausing by contributing to the formation of this intermediate complex. In the absence of cohesin, Pol II likely bypasses this step and prematurely enters elongation. We concluded that cohesin engages with the transcriptional machinery specifically between pausing and pause release, based on its CDK9 dependency and the co-occurrence of both pausing- and release-associated factors. This interpretation is further supported by two independent approaches: (i) ChIP-seq revealed that cohesin and its loader co-occupy promoters with Pol II, and that their occupancy is sensitive to transcriptional inhibition; and (ii) co-immunoprecipitation demonstrated that CDK9 activity is required for the physical association between the cohesin loader and Pol II. Together, these results suggest that CDK9 modulates the dynamic association between cohesin, its loader, and Pol II across the transcription cycle. Moreover, Inhibition of cohesin's ATPase activity made this complex detectable, suggesting that it normally disassembles upon cohesin-mediated ATP hydrolysis. Although this pausing–to–release complex has so far been detected only *in vitro*, whether an equivalent complex forms *in vivo* remains to be determined.

Proper assembly of elongation factors has been proposed to require sufficient pausing duration[34]. Accordingly, we observed incomplete formation of the elongation complex upon cohesin depletion, potentially due to insufficient pause time. Given that elongation factors facilitate processive and stable transcription by interacting with elongating Pol II[31,33], we speculate that cohesin enhances transcription processivity and prevents premature transcription termination by ensuring proper transition from pausing to productive elongation. Consistently, our RNA-seq data revealed enhanced XRN2-mediated RNA surveillance upon cohesin depletion.

Acute cohesin depletion redistributes H3K27ac by reducing it at promoters and increasing it at enhancers, and simultaneously weakens chromatin contacts between these regions. These observations suggest that cohesin mediates long-range promoter–enhancer interactions, enabling the delivery of transcriptional activation signals from enhancers to promoters (Fig. 6e). Furthermore, these interacting regions span long genomic distances, consistent with previous studies showing that cohesin facilitates the activity of distal enhancers.[48,49] A potential mechanism underlying H3K27ac redistribution involves the histone acetyltransferases p300/CBP, which localize at enhancers and deposit H3K27ac not only at enhancers but also at nearby promoters brought into proximity via promoter–enhancer interactions[41,50]. We speculate that cohesin loss reduces this spatial proximity, thereby limiting p300/CBP-mediated acetylation to enhancers. The TFIID complex, a key factor required for transcription initiation, also decreased at promoters upon cohesin loss. As inhibition of H3K27ac has been shown to diminish TFIID binding[41], the reduced H3K27ac likely contributes to the reduction in TFIID occupancy. Together, these findings suggest that cohesin maintains an active chromatin environment and supports efficient transcription initiation by sustaining regulatory communication between promoters and enhancers.

Our findings shed light on the molecular basis of CdLS, a developmental disorder caused by cohesin dysfunction. CdLS shares clinical features and a dysregulated transcriptome with CHOPS syndrome, in which SEC component AFF4 is hypermorphic[14,51]. Our results suggest that cohesin stabilizes promoter-proximal pausing and restrains pause release. Thus, aberrantly enhanced pause release may represent a

common molecular pathology of CdLS and CHOPS syndrome, as SEC triggers this transition. Consistent with this idea, genome-wide increased AFF4 and decreased NELF at promoters were observed in patient cells from both disorders[51]. This subtle yet widespread dysregulation may become particularly evident in response to external or environmental stimuli during development, as suggested by our findings. Alternatively, an immature elongation complex and destabilized productive elongation may also contribute to disease pathogenesis. Mutations in cohesin factors are also reported in cancers, sometimes co-occurring with additional mutations in other genes.[52,53] Furthermore, dysregulated gene expression caused by these mutations is proposed to contribute to tumorigenesis, although the underlying mechanism remains unclear. Our findings may offer mechanistic insight into how cohesin dysfunction contributes to transcriptional dysregulation and cancer development.

Methods

Cell culture

HCT-116-RAD21-mAID-mClover cells were described previously[19]. Cells were cultured in McCoy's 5A medium supplemented with 10% FBS and penicillin-streptomycin-l-glutamine solution (Wako) at 37°C in 5% CO2. RAD21 was depleted by treatment with 0.5 mM indole-3-acetic acid (IAA; Sigma) for 3 hours. For CDK9 inhibition, 100 μM 5,6-Dichlorobenzimidazole 1-β-D-Ribofuranoside (DRB; TCI) was added concurrently with IAA for 3 hours, or 1 μM flavopiridol (Santa Cruz) was added during the final hour of IAA treatment.

HCT116 CMV-OsTIR1 SMC1A-mAID and HCT116 Tet-OsTIR1 Stag-3FLAG-NIPBL cells were generated as previously described[19,54]. Briefly, HCT116 CMV-OsTIR1 cells were co-transfected with SMC1A CRISPR and SMC1A-mAID-Hygro donor vectors, and HCT116 Tet-OsTIR1 cells with NIPBL-N CRISPR and Hygro-P2A-Stag-3FLAG-NIPBL donor vectors. After selection with 100 μg/mL HygroGold (Invivogen), single colonies were isolated, and successful knock-in was confirmed by genomic PCR and western blotting.

siRNA-mediated knockdown

RNA interference was performed with Lipofectamine RNAiMax (Invitrogen) using siRNA duplexes at a final RNA concentration of 50 nM, according to the manufacturer's instructions. Cells were collected two days after transfection. The siRNA sequences (5′→3′) were: WAPL, UCUCCUGCUCGUUAGAAGUAAGGG; XRN2, UCGCUAUUACAUAGCUGAUCGUUUA.

Cell cycle profile analysis

Cells were fixed in 70% ethanol, washed with PBS, treated with RNase A (50 µg/mL), and stained with propidium iodide (50 µg/mL). DNA content was measured on BD Accuri C6 plus flow cytometer.

Spike-in ChIP-seq and MNase ChIP-seq

ChIP-seq was performed as described previously[55]. MNase ChIP-seq followed the same procedure, except that chromatin was fragmented by sequential micrococcal nuclease (MNase) digestion and sonication. In both assays, mouse C2C12 cells were added as a spike-in control. Briefly, HCT-116 and C2C12 cells were crosslinked with 1% formaldehyde for 10 min, quenched with 125 mM glycine for 5 min, and stored at –80°C. For standard ChIP-seq, chromatin fractionation was performed by sequential resuspension in LB1 Buffer [50 mM HEPES-KOH (pH 7.5), 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, 1 mM PMSF, 10 mM DTT, 1x cOmplete EDTA-free (Roche)], LB2 Buffer [20 mM Tris-HCl (pH 7.5), 200 mM NaCl, 1 mM EDTA, 1 mM PMSF, 1x cOmplete EDTA-free], and LB3 Buffer [20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-Deoxycholate, 0.1% SDS, 1x cOmplete EDTA-free]. Chromatin was sonicated on Sonifier 250D (BRANSON). For MNase ChIP-seq, cells were resuspended in MNase Buffer [50 mM NaCl, 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 1 mM $CaCl_2$, 0.2% NP-40, 1x cOmplete EDTA-free], digested with MNase (New England BioLabs), and sonicated. HCT-116 and C2C12 chromatin were combined at a 4:1 ratio and immunoprecipitated with antibodies conjugated to Dynabeads Protein A/G (Invitrogen) overnight at 4 °C. DNA was purified using the QIAquick PCR purification kit (QIAGEN), and libraries were prepared with

NEBNext DNA Library Prep Master Mix (New England Biolabs) and sequenced on HiSeq 2500 or NextSeq 2000 (Illumina).

The following antibodies were used: Rpb1 NTD (D8L4Y) antibody (Cell Signaling Technology, #14958), Rabbit polyclonal anti-RAD21 (Sakata et al.), Mouse monoclonal anti-MAU2 (Sakata et al.), Purified Rabbit Anti-GFP antibody (Torrey Pines Biolabs, TP401), Anti-Phospho RNA Polymerase II CTD (Ser2) (Monoclonal research institute Inc., MABI0602), anti-acetyl Histone H3 (Lys27) antibody (MBL International, MABI0309), anti-trimethyl Histone H3 (Lys4) antibody (MBL International, MABI0304), anti-monomethyl Histone H3 (Lys4) antibody (MBL International, MABI0302), Rabbit anti-MCEF Antibody (Bethyl, A302-539A), TH1L (D5G6W) Rabbit mAb (Cell Signaling Technology, #12265), Anti-Cdk9 antibody [EPR22956-37] - ChIP Grade (Abcam, #ab239364), Rabbit anti-TAF1 Antibody, Affinity Purified (Bethyl, #A303-505A), Anti-TATA binding protein TBP antibody [mAbcam51841] (Abcam, #ab300656).

Quantitative PCR

ChIP DNA was quantified using real-time PCR systems 7500 and StepOnePlus (Life Technologies) with KAPA SYBR Fast qPCR kit (KAPA Biosystems). The following primers were used (5′→3′): Rad21_1 F: GGCCTAGAAAACACGTCTACAG, R: TCCACCAAGGGACAAACAGAG; Rad21_2 F: TGCGGGTACACTTTGAGAACC, R: GTGCCACCTAGTGGTATTTCCT; Rad21_3 F: CTTTGCCCCTTTGTGGCTCA, R: ATCCTGACACTGCTGGGAATG; negative control F: ATGAGTCTGGGATTTGGCTGG, R: TCTCAGGGGAAGGAGAGAGTC.

RNA-seq

Total RNA was extracted from HCT-116 cells with TRIzol reagent (Invitrogen) and purified with NucleoSpin RNA II kit (Macherey-Nagel), following the manufacturer’s instructions. rRNA was depleted with NEBNext rRNA Depletion Kit (New England BioLabs). RNA-seq libraries were prepared using NEBNext Ultra II Directional RNA Library Prep Kit for Illumina and NEBNext Multiplex Oligos for Illumina (New England BioLabs). The libraries were sequenced on Nextseq 2000 (Illumina).

TT-seq

TT-seq was performed as described previously with minor modifications[56]. Drosophila S2 cells were used as a spike-in control. Nascent transcripts were labelled by incubating HCT-116 cells and S2 cells with 1 mM 4-thiouridine (Cayman Chemical) for 10 min and 20 min, respectively. Total RNA was extracted with TRIzol reagent (Invitrogen) and purified using Direct-zol RNA Miniprep Plus (ZYMO RESEARCH). For each library, 100 µg of HCT-116 RNA was mixed with 1 µg S2 RNA and sheared on M220 Focused-ultrasonicator (Covaris). Biotinylation of 4sU-labeled RNA was carried out with MTSEA-biotin-XX (20 mg/mL), followed by phenol–chloroform extraction and purification using the mMACS Streptavidin Kit (Miltenyi Biotec) and Direct-zol RNA Miniprep. TT-seq libraries were generated with the NEBNext Ultra II Directional RNA Library Prep Kit and Multiplex Oligos (New England Biolabs) and sequenced on a NextSeq 2000 (Illumina).

EU-seq

Drosophila S2 cells were used as a spike-in control. Nascent transcripts were labelled with 0.5 mM 5-ethynyl uridine (Thermo Fisher) for 20 min (HCT-116) or 0.2 mM for 2 hours (S2). Total RNA was extracted with TRIzol reagent (Invitrogen) and purified using NucleoSpin RNA II kit (Macherey-Nagel). rRNA depletion of HCT-116 RNA was performed with the NEBNext rRNA Depletion Kit (New England Biolabs), and S2 mRNA was purified with the Dynabeads mRNA Purification Kit (Invitrogen). For each library, 200 ng of HCT-116 RNA was mixed with 20 ng of S2 mRNA, and EU-labelled RNA was isolated with Click-iT Nascent RNA Capture Kit (Thermo Fisher). cDNA was synthesized with SuperScript VILO cDNA synthesis kit (Thermo Fisher). Second-strand cDNA synthesis and library preparation were performed with NEBNext Ultra II Directional RNA Library Prep Kit for Illumina and NEBNext Multiplex Oligos for Illumina (New England BioLabs). The library was sequenced on Nextseq 2000 (Illumina).

mNET-seq

mNET-seq was performed as described previously with minor modifications[57]. Briefly, chromatin fractionation from 1 x 10^8 HCT-116 cells were performed using HLB+N buffer [10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 2.5 mM $MgCl_2$, 0.5% NP-40, 1x cOmplete EDTA-free (Roche), 1x PhosSTOP (Roche)], HLB+NS buffer [10 mM Tris-HCl (pH 7.5), 10 mM NaCl, 2.5 mM $MgCl_2$, 0.5% NP-40, 1x cOmplete EDTA-free, 1x PhosSTOP, 10% sucrose],

NUM1 lysis buffer [20 mM Tris-HCl (pH 7.9), 75 mM NaCl, 0.5 mM EDTA (pH 8.0), 50% glycerol, 1x cOmplete EDTA-free, 1x PhosSTOP], and NUM2 lysis buffer [20 mM HEPES (pH 7.5), 300 mM NaCl, 0.2 mM EDTA (pH 8.0), 7.5 mM $MgCl_2$, 1% NP-40, 1M urea, 1x cOmplete EDTA-free, 1x PhosSTOP]. Chromatin was digested with MNase in MNase reaction buffer [1x MNase buffer, 1x BSA, 400 gel units/µL MNase (New England BioLabs)]. Pol II complexes were immunoprecipitated using Pol II CTD antibody (mAb) (Monoclonal research institute Inc.) and Dynabeads Protein G (Invitrogen). After washes, bead-bound RNA was phosphorylated with T4 PNK (NEB), extracted with TRIzol, and purified using Direct-zol RNA Miniprep (Zymo Research). Drosophila S2 mRNA was prepared as a spike-in control by extraction, poly(A) selection, fragmentation, and phosphorylation, and 2.5 ng was added to each mNET-seq sample. Libraries were constructed using the NEBNext Small RNA Library Prep Set (NEB) with size selection on AMPure XP beads (Beckman Coulter) and sequenced on a NextSeq 550 (Illumina).

Micro-C

Micro-C was performed as described previously[51]. Briefly, 5×10^6 cells were crosslinked with 1% formaldehyde for 10 min and quenched with 125 mM glycine for 5 min. After PBS washing, cells were further crosslinked with 3 mM DSG (Thermo Fisher) for 40 min, quenched with 400 mM glycine for 5 min, and stored at –80°C. Cells were resuspended in Micro-C Buffer #1 [50 mM NaCl, 10 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 1 mM CaCl2, 0.2% NP-40, 1x cOmplete EDTA-free (Roche)] and digested with MNase. DNA ends were repaired, filled in with biotinylated nucleotides, and ligated with T4 ligase (NEB). Unligated DNA ends were digested with Exonuclease III (NEB). Following proteinase K treatment and reverse cross-linking, DNA was purified with QIAquick PCR Purification Kit (QIAGEN) and size-selected with AMPure XP beads and subjected to biotin pull-down with Dynabeads MyOne Streptavidin C1

(Invitrogen). End repair, adaptor ligation, and PCR amplification were carried out on beads using the NEBNext Ultra II DNA Library Prep Kit and Multiplex Oligos (NEB) according to the manufacturer's instructions. Libraries were sequenced on a NovaSeq 6000 (Illumina).

Co-immunoprecipitation

Co-immunoprecipitation was performed using the nuclear fraction. Cells were washed in buffer A [20 mM Tris-HCl (pH 7.5), 10 mM NaCl, 1.5 mM $MgCl_2$, 10% glycerol, 10% sucrose, 1x cOmplete EDTA-free (Roche), 1x PhosSTOP (Roche), 0.2% Triton X-100] twice and stored at −80°C. The frozen pellet was lysed in lysis Benzonase buffer [20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.05% NP-40, 10% glycerol, 2.5 mM $MgCl_2$, 1x cOmplete EDTA-free, 1x PhosSTOP, Benzonase (Millipore)], and centrifuged. The pellet was further resuspended in another lysis buffer [20 mM Tris-HCl (pH 7.5), 250 mM NaCl, 0.1% NP-40, 10% glycerol, 1x cOmplete EDTA-free, 1x PhosSTOP]. Supernatants from both lysis steps were pooled and incubated with antibody-conjugated beads for 3 hours at 4°C. Beads were washed five times with buffer [20 mM Tris-HCl (pH 7.4), 200 mM NaCl, 5% glycerol, 1 mM EDTA, 0.1% NP-40], and bound proteins were eluted in SDS sample loading buffer and boiled for 5 min at 95°C.

The following antibodies were used for co-immunoprecipitation: Anti-RNA Polymerase II CTD mAb (Fujifilm, #384-09271), Anti-Phospho RNA Polymerase II CTD (Ser2) mAb (Fujifilm, #381-09281), Anti DYKDDDDK tag Antibody Magnetic Beads (Fuijifilm, #017-25151), Nipbl (Short form) Rat Monoclonal Antibody [Clone ID: KT55] (ORIGENE, #AM20797PU-N).

***In vitro* transcription assay**

Nuclear extracts from HeLa S3 cells were prepared as described by Dignam et al[58]. Cohesin was immunodepleted twice by incubating nuclear extract with protein A magnetic beads conjugated to anti-SMC1 and anti-SMC3 antibodies. The Mediator complex was immunodepleted twice using protein G magnetic beads conjugated to anti-MED4, anti-MED15, and anti-MED25 antibodies. A plasmid (p5G-sMLT) was constructed by inserting five tandem GAL4 binding sites, an 81-bp minimal promoter region

containing a TATA box, and a 245-bp *slc16a1* first intron fragment into the pUC19 multicloning site. DNA templates were amplified using a biotinylated primer 50 bp upstream of the GAL4 sites and a reverse primer 419 bp downstream of the TATA box, and conjugated to Dynabeads M-280 Streptavidin (Invitrogen) in binding buffer [10 mM Tris-HCl (pH 7.5), 1 M NaCl, 0.05% Tween-20]. After washing, templates were incubated with GAL4-VP16 in activator binding buffer [12.5 mM HEPES-KOH (pH 7.4), 62.5 mM KCl, 7.5 mM MgCl□, 12.5% glycerol, 0.025% NP-40, 100 µg/mL BSA, 0.625 mM DTT]. After washing using activator binding buffer without BSA, GAL4-VP16-bound templates were incubated with nuclear extracts in reaction buffer [12.5 mM HEPES-KOH (pH 7.4), 62.5 mM KCl, 7.5 mM MgCl□, 12.5% glycerol, 0.025% NP-40, 100 µg/mL BSA, 0.625 mM DTT, 2.5 mM NU-7441, 0.5x cOmplete EDTA-free] at 23°C. AlFx and 100 µM NTPs or ADP, prepared by mixing 1 mM AlCl□ and 5 mM NaF, were added to the reaction solution. The timing of addition is described in the corresponding figure legend. After the reaction, the beads were washed three times with activator binding buffer without BSA and eluted in SDS sample loading buffer.

Purification of recombinant proteins

NIPBL mutants N1 (1–700) and N2 (701–1000) were N-terminally FLAG-tagged and subcloned into the pFASTBac HTB vector by GenScript. NIPBL mutant C (1121–2804) was C-terminally FLAG-tagged and subcloned into the pAcHLT-B vector, also by GenScript. Wild-type MAU2 was subcloned into the pAcSG2 vector by GenScript. Recombinant baculoviruses were generated using either the flashBAC Baculovirus Expression System (Mirus) or the Bac-to-Bac Baculovirus Expression System (Gibco), according to the manufacturer's protocols. NIPBL fragments (N2 and C), as well as N1 co-expressed with MAU2, were expressed in Sf9 cells following infection with recombinant baculoviruses. Cells were lysed using a Dounce homogenizer in PB500 buffer [25□mM NaH□PO□/Na□HPO□ (pH 7.5), 500□mM NaCl, 5% glycerol] supplemented with 10□mM imidazole, 0.05% Tween-20, 0.1% NP-40, 1□mM PMSF, 3□mM 2-mercaptoethanol, and 1x cOmplete EDTA-free (Roche). Lysates were clarified by centrifugation at 35,000 × g for 1□hour and incubated with Ni-NTA Agasore (Qiagen) for 2□hours at 4□°C. The resin was washed with PB500 supplemented with 20□mM imidazole, and bound proteins were eluted with PB150 (same

composition as PB500 but with 150 mM NaCl) supplemented with 300 mM imidazole. Eluted proteins were further purified by incubation with ANTI-FLAG M2 Affinity Gel (Sigma) for 2 hours at 4 °C, washed with PB150, and eluted with PB150 containing 0.5 mg/ml 3×FLAG peptide. Final protein preparations were concentrated and buffer-exchanged into 20 mM Tris-HCl (pH 7.5), 100 mM NaCl, and 20% glycerol using Amicon Ultra Centrifugal Filters, 50kDa MWCO (Millipore).

SDS-PAGE, CBB staining, and western blotting

Protein samples were resolved by SDS–PAGE using either 4–15% Mini-PROTEAN® TGX™ Precast Protein Gels (Bio-Rad) or Perfect NT precast gels (DRC). For Coomassie Brilliant Blue (CBB) staining, gels were stained with one-step CBB stain solution (BIO CRAFT). For western blotting, proteins were transferred to Amersham Protran 0.2 NC nitrocellulose Western blotting membranes (GE Healthcare). The membranes were blocked with 3% skim milk in PBS-T and incubated with primary antibodies for several hours or overnight. After washing with PBS-T, the membranes were incubated with HRP-conjugated secondary antibodies for 1 hour. Signals were detected using SuperSignal West Pico PLUS Chemiluminescent Substrate (Thermo Scientific). When required, membranes were stripped using Stripping Solution (Wako) and reprobed.

Spike-in ChIP-seq and MNase ChIP-seq alignment and data analysis

Reads were aligned to human (hg38) and mouse (mm10) genomes with Bowtie (version 1.2.3)[59]. Normalization factors were calculated from the ratio of mouse spike-in to human reads in ChIP (IP) and input (WCE) samples, as described previously[51]. Redundant reads were removed, and read numbers were normalized to a spike-in control with parse2wig (version 3.7.2)[60]. Peak calling was performed with drompa_peakcall (version 3.7.2) in “PC_SHARP” mode[60]. Visualization was carried out with drompa+ PC_SHARP (version 1.17.2) and deepTools(version 3.5.1)[61,62].

RNA-seq alignment and data analysis

Reads were aligned to the human genome (hg38) using STAR (version 2.7.10a)[63]. Transcript abundance was quantified using RSEM (version 1.3.3)[64]. Differential expression analysis was performed with edgeR (version 3.40.2; FDR < 0.05)[65].

TT-seq alignment and data analysis

Reads were aligned to human (hg38) and Drosophila (dm6) genomes with STAR (version 2.7.1a). Mapped reads were divided into sense and anti-sense reads using samtools (version 1.15.1)[66]. Scale factors were calculated from a Drosophila gene-level count matrix generated with htseq-count (version 1.99.2) and normalized with DESeq2's estimateSizeFactors function[67,68]. A read count table was generated for genes, with TT-seq reads counted using drompa3 (version 3.7.2)[60], and differential expression analysis was performed with edgeR (version 4.2.2) with the likelihood ratio test (FDR < 0.05)[65].

EU-seq alignment and data analysis

Low-quality reads were removed with Trimmomatic (version 0.39)[69]. Reads were aligned to the Drosophila genome (dm6) using Bowtie2 (version 2.4.5) in "--very-sensitive-" mode, and the unmapped reads were aligned to the human genome (hg38)[70]. Scale factors were calculated from a Drosophila gene-level count matrix generated with htseq-count (version 1.99.2) and normalized with DESeq2's estimateSizeFactors function. Redundant reads were removed, and the remaining reads were normalized to a spike-in control with parse2wig (version 3.7.2). Differential expression analysis was performed with edgeR (version 4.2.2) with the likelihood ratio test (FDR < 0.05)[65].

mNET-seq alignment and data analysis

Reads were aligned to human (hg38) and Drosophila (dm6) genomes with STAR (version 2.7.1a). Post-processing steps, including read deduplication, RNA 3' end identification, removal of PCR internal priming,

and exclusion of reads associated with co-transcriptional splicing, were performed as described by Prudêncio et al[71]. Scale factors were calculated from a Drosophila gene-level count matrix generated with htseq-count (version 1.99.2) and normalized with DESeq2's estimateSizeFactors function. Mapped reads were divided into sense and anti-sense reads using samtools[66] (version 1.15.1).

Micro-C alignment and data analysis

Reads were aligned to the human genome (hg38) with BWA-MEM (version 0.75)[72]. Ligation-junction detection, read sorting, PCR-duplicate removal, and pair-file generation were performed using Pairtools (version 0.3.0)[73]. Matrix files were generated from the pairs files using Juicer (version 1.22.01) at 5 kb, 2 kb, and 1 kb resolutions. Loops were identified using Juicer Tools HiCCUPS and Mustache (version 1.2.0), and loops detected by both methods were combined to generate a set of Micro-C loops [74,75]. TADs were identified using Juicer Tools Arrowhead at a 25 kb resolution. Loop intensity was quantified using Chromosight[76] (version 1.6.1). For visualization, contact maps were generated using HiCExplorer (version 3.7.2), and aggregate plots were created using coolpup.py (version 1.0.0)[77,78].

Identification of +1 nucleosome and pausing site

Genomic locations of nucleosomes with H3K4me3 modifications were identified from H3K4me3 MNase ChIP-seq data with DANPOS3 (version 2.2.2)[79]. Nucleosomes located between TSS +10 bp and TSS + 300 bp with an occupancy score > 1, a P-value for occupancy < −1, a fuzziness score < 100, and a P-value for fuzziness < −2 were annotated as +1 nucleosomes.

Pausing sites were defined as the genomic position between the TSS and the +1 nucleosome dyad with the highest mNET-seq signal. mNET-seq signal intensity was quantified in 1-bp bins and smoothed using a moving average with a window size of 11. The intensity at the pausing site was required to exceed 1.2 times the median intensity of the surrounding regions.

Selection of genes for analysis

We selected 3,597 genes that met all of the following criteria: TT-seq TPM ≥ 1, gene length ≥ 5 kb, and >1 kb from neighboring genes, and the presence of a +1 nucleosome and a defined pausing site.

Calculation of the pausing-duration index, Pol II ChIP-seq pausing index, the half-time of paused Pol II, velocity index, and the slope of RNA-seq intronic reads

The pausing-duration index was calculated as the mNET-seq/TT-seq signal ratio around pausing sites for each gene. TT-seq and mNET-seq signals within ± 25 bp of pausing sites were quantified using multiBigwigSummary from deepTools.

Pol II ChIP-seq pausing index was calculated as the ratio of RPB1 occupancy at promoters to that at gene bodies. RPB1 occupancy at promoters (TSS -500 bp to 500 bp) and at gene bodies (TSS +2 kb to +3 kb) was quantified using multiBigwigSummary from deepTools.

The half-time of paused Pol II was estimated by applying a single-exponential decay model to the reduction in RPB1 occupancy at promoters over the time-course experiment following triptolide treatment. RPB1 ChIP-seq was performed 0, 5, 10, and 20 min after triptolide treatment, and RPB1 occupancy at promoters (TSS – 0.5 kb to + 0.5 kb) was quantified using multiBigwigSummary from deepTools. Genes exhibiting a continual reduction in RPB1 occupancy throughout the time-course experiment were analyzed using a single-exponential decay model fitted with nls() in R.

The velocity index was defined as the TT-seq/mNET-seq signal ratio along genes. mNET-seq and TT-seq signals were quantified using multiBigwigSummary from deepTools.

RNA-seq read intensity at introns longer than 1 kb was quantified using multiBigwigSummary from deepTools with 100-bp bins. A simple linear regression model was fitted to the quantified intensity of each intron using lm() in R. Introns with a negative slope and adjusted $R^2 \geq 0.5$ were analyzed.

K-means clustering of genes

K-means clustering (k=3) was performed in R on the log□FC of promoter H3K27ac, pausing-duration, velocity index, and TT-seq gene expression for each gene. All indices were normalized before clustering.

Identification of promoter and enhancer regions

Twelve chromatin states were defined in control and auxin-treated cells using ChromHMM (version 1.25) with ChIP-seq datasets for histone modifications (H3K27ac, H3K4me1, H3K4me3, H3K36me3, H3K79me2, H3K4me2, H2Aub, H2Bub, H3K27me3, H3K9me3), chromatin-associated factors (CTCF, BRD4, MED1), and TT-seq signal, all obtained from both conditions. Chromatin segments enriched for H3K27ac, H3K4me3, BRD4, and MED1 were annotated as active promoters, while those enriched for H3K27ac, H3K4me1, BRD4, and MED1 were classified as active enhancers.

Identification of up- or down-regulated promoters and enhancers

Active promoters and enhancers defined by ChromHMM were merged, and H3K27ac signal intensity was compared between control and auxin-treated cells using a binomial test with drompa_draw CI. FDRs were computed from p-values using p.adjust in R, and peaks with FDR < 0.05 were classified as increased or decreased based on signal changes.

Evaluation of reproducibility

Reproducibility between biological replicates was systematically assessed for TT-seq, RNA-seq, EU-seq, ChIP-seq, and mNET-seq. Reproducibility was evaluated by calculating Spearman correlations of genome-wide read coverage in 10-kb bins after excluding bins containing zero counts. For Micro-C, replicate reproducibility was quantified using QuASAR-Rep scores as implemented in 3DChromatin_ReplicateQC at 100-kb resolution[80]. All assays showed high reproducibility across conditions, with replicate correlations and QuASAR-Rep scores reported in Supplementary Tables 1and 2.

Simulation of Pol II State Transitions

We modeled transcription dynamics under cohesin depletion using a discrete-time Markov chain, in which Pol II transitions among four states: Free (1), Initiation (2), Paused (3), and Elongation (4). (Supplemental Fig. 10a). Pol II progresses sequentially through Initiation and Paused before Elongation, but can return directly to the Free from either Initiation or Paused. Dwell times in the Free, Initiation, and Paused states were variable and modeled through self-transitions, whereas Elongation had a fixed hold time (θ) before returning to Free.

We assigned the numbers 1-4 to the states Free, Initiation, Paused, and Elongation, respectively. Letting p_{ij} denote the transition probability from state *i* to state *j*,

$$\sum_j p_{ij} = 1 \text{ (for any } i)$$

Considering the constraints above, the transition probability matrix *P* is:

$$P = \begin{pmatrix} 1-p_{12} & p_{12} & 0 & 0 \\ p_{21} & 1-p_{21}-p_{23} & p_{23} & 0 \\ p_{31} & 0 & 1-p_{31}-p_{34} & p_{34} \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

This model has six independent parameters: five transition probabilities (p_{12}, p_{21}, p_{23}, p_{31}, p_{34}), and the hold time θ_4 in the Elongation state.

Next, we calibrated the parameters to reflect the actual *in vivo* behavior of Pol II reported previously[45]. One step in the Markov model was assumed to correspond to one second. The mean durations in the Initiation and Paused states were

Initiation: $\frac{1}{1-p_{22}} = \frac{1}{p_{21}+p_{23}}$, Paused: $\frac{1}{1-p_{33}} = \frac{1}{p_{31}+p_{34}}$

Reported ratios of p_{21}:p_{23}, and p_{31}:p_{34} were used to determine p_{21}, p_{23}, p_{31} and p_{34}, and θ_4 was also set to match empirical measurements. (Supplemental Fig. 10b)

The remaining parameter p_{12} was determined to reproduce the observed stationary distribution. Since the transition probability matrix has a left eigenvector with eigenvalue 1, we let this eigenvector be (α_1, α_2, α_3, α_4). The stationary distribution is then

$\pi_i = \frac{\theta_i \cdot a_i}{\sum_i \theta_i \cdot a_i}$, $\theta_i = 1$ (for i = 1, 2, 3)

p_{12} was chosen to minimize the difference between the calculated and empirical distribution. All model simulations were conducted using R.

Data availability

Micro-C, ChIP-seq, TT-seq, mNET-seq, RNA-seq, and EU-seq data were deposited in NCBI Gene Expression Omnibus (GEO) under the following accession numbers: GSE260854 (Micro-C), GSE260852 and GSE300575 (ChIP-seq), GSE300574 (TT-seq), GSE300576 (mNET-seq), GSE300577 (RNA-seq), and GSE260853 (EU-seq).

References

1. Ciosk, R. *et al.* Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins. *Mol. Cell* **5**, (2000).

2. Michaelis, C., Ciosk, R. & Nasmyth, K. Cohesins: Chromosomal proteins that prevent premature separation of sister chromatids. *Cell* **91**, (1997).

3. Guacci, V., Koshland, D. & Strunnikov, A. A direct link between sister chromatid cohesion and chromosome condensation revealed through the analysis of MCD1 in S. cerevisiae. *Cell* **91**, (1997).

4. Haering, C. H., Löwe, J., Hochwagen, A. & Nasmyth, K. Molecular architecture of SMC proteins and the yeast cohesin complex. *Mol. Cell* **9**, (2002).

5. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, (2017).

6. 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**, (2017).

7. Vian, L. *et al.* The Energetics and Physiological Impact of Cohesin Extrusion. *Cell* **173**, (2018).

8. Thiecke, M. J. *et al.* Cohesin-Dependent and -Independent Mechanisms Mediate Chromosomal Contacts between Promoters and Enhancers. *Cell Rep.* **32**, (2020).

9. Hsieh, T. H. S. *et al.* Enhancer–promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1. *Nat. Genet.* **54**, (2022).

10. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, (2018).

11. Busslinger, G. A. *et al.* Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature* **544**, (2017).

12. Tonkin, E. T., Wang, T. J., Lisgo, S., Bamshad, M. J. & Strachan, T. NIPBL, encoding a homolog of fungal Scc2-type sister chromatid cohesion proteins and fly Nipped-B, is mutated in Cornelia de Lange syndrome. *Nat. Genet.* **36**, (2004).

13. Kaur, M. *et al.* Genomic analyses in Cornelia de Lange Syndrome and related diagnoses: Novel candidate genes, genotype-phenotype correlations and common mechanisms. *Am. J. Med. Genet. A* **191**, 2113–2131 (2023).

14. Izumi, K. *et al.* Germline gain-of-function mutations in AFF4 cause a developmental syndrome functionally linking the super elongation complex and cohesin. *Nat. Genet.* **47**, (2015).

15. Lin, C. *et al.* AFF4, a Component of the ELL/P-TEFb Elongation Complex and a Shared Subunit of MLL Chimeras, Can Link Transcription Elongation to Leukemia. *Mol. Cell* **37**, (2010).

16. Liang, K. *et al.* Targeting Processive Transcription Elongation via SEC Disruption for MYC-Induced Cancer Therapy. *Cell* **175**, (2018).

17. Kwak, H. & Lis, J. T. Control of transcriptional elongation. *Annu. Rev. Genet.* **47**, 483–508 (2013).

18. Aoi, Y. & Shilatifard, A. Transcriptional elongation control in developmental gene expression, aging, and disease. *Mol. Cell* **83**, 3972–3999 (2023).

19. 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**, (2016).

20. Li, W. S. *et al.* Mature chromatin packing domains persist after RAD21 depletion in 3D. *Sci. Adv.* **11**, eadp0855 (2025).

21. Rhodes, J. D. P. *et al.* Cohesin Disrupts Polycomb-Dependent Chromosome Interactions in Embryonic Stem Cells. *Cell Rep.* **30**, (2020).

22. Nojima, T. *et al.* Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing. *Cell* **161**, (2015).

23. Caizzi, L. *et al.* Efficient RNA polymerase II pause release requires U2 snRNP function. *Mol. Cell* **81**, (2021).

24. Gressel, S. *et al.* CDK9-dependent RNA polymerase II pausing controls transcription initiation. *Elife* **6**, (2017).

25. Harlen, K. M. & Churchman, L. S. The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews Molecular Cell Biology* **18**, (2017).

26. Shao, W. & Zeitlinger, J. Paused RNA polymerase II inhibits new transcriptional initiation. *Nat. Genet.* **49**, (2017).

27. Wang, H. *et al.* H3K4me3 regulates RNA polymerase II promoter-proximal pause-release. *Nature* **615**, (2023).

28. Chan, K. L. *et al.* Cohesin's DNA exit gate is distinct from its entrance gate and is regulated by acetylation. *Cell* **150**, (2012).

29. Huis In't Veld, P. J. *et al.* Characterization of a DNA exit gate in the human cohesin ring. *Science*

(80-.). **346**, (2014).

30. Vos, S. M. *et al.* Structure of activated transcription complex Pol II–DSIF–PAF–SPT6. *Nature* **560**, (2018).

31. Žumer, K. *et al.* Two distinct mechanisms of RNA polymerase II elongation stimulation in vivo. *Mol. Cell* **81**, (2021).

32. Ameur, A. *et al.* Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat. Struct. Mol. Biol.* **18**, (2011).

33. Hou, L. *et al.* Paf1C regulates RNA polymerase II progression by modulating elongation rate. *Proc. Natl. Acad. Sci. U. S. A.* **116**, (2019).

34. Aoi, Y. *et al.* SPT6 functions in transcriptional pause/release via PAF1C recruitment. *Mol. Cell* **82**, (2022).

35. Eaton, J. D. & West, S. Termination of Transcription by RNA Polymerase II: BOOM! *Trends in Genetics* **36**, (2020).

36. Alonso-Gil, D. & Losada, A. NIPBL and cohesin: new take on a classic tale. *Trends in Cell Biology* **33**, (2023).

37. Uyehara, C. M. & Apostolou, E. 3D enhancer-promoter interactions and multi-connected hubs: Organizational principles and functional roles. *Cell Reports* **42**, (2023).

38. Çamdere, G. O., Carlborg, K. K. & Koshland, D. Intermediate step of cohesin's ATPase cycle allows cohesin to entrap DNA. *Proc. Natl. Acad. Sci. U. S. A.* **115**, (2018).

39. Higashi, T. L. *et al.* A Structure-Based Mechanism for DNA Entry into the Cohesin Ring. *Mol. Cell* **79**, (2020).

40. Kimura, H. Histone modifications for human epigenome analysis. *Journal of Human Genetics* **58**, (2013).

41. Narita, T. *et al.* Enhancers are activated by p300/CBP activity-dependent PIC assembly, RNAPII

recruitment, and pause release. *Mol. Cell* **81**, (2021).

42. Gressel, S., Schwalb, B. & Cramer, P. The pause-initiation limit restricts transcription activation in human cells. *Nat. Commun.* **10**, (2019).

43. Ernst, J. & Kellis, M. ChromHMM: Automating chromatin-state discovery and characterization. *Nature Methods* **9**, (2012).

44. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat. Protoc.* **12**, (2017).

45. Steurer, B. *et al.* Live-cell analysis of endogenous GFP-RPB1 uncovers rapid turnover of initiating and promoter-paused RNA Polymerase II. *Proc. Natl. Acad. Sci. U. S. A.* **115**, (2018).

46. Buckley, M. S., Kwak, H., Zipfel, W. R. & Lis, J. T. Kinetics of promoter Pol II on Hsp70 reveal stable pausing and key insights into its regulation. *Genes Dev.* **28**, (2014).

47. Adelman, K. *et al.* Immediate mediators of the inflammatory response are poised for gene activation through RNA polymerase II stalling. *Proc. Natl. Acad. Sci. U. S. A.* **106**, (2009).

48. Guckelberger, P. *et al.* Cohesin-mediated 3D contacts tune enhancer-promoter regulation. *bioRxiv* 2024.07.12.603288 (2024). doi:10.1101/2024.07.12.603288

49. Tjalsma, S. J. D. *et al.* Long-range enhancer-controlled genes are hypersensitive to regulatory factor perturbations. *Cell Genomics* **5**, 100778 (2025).

50. Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* **33**, (2015).

51. Sakata, T. *et al.* A common molecular mechanism underlying Cornelia de Lange and CHOPS syndromes. *Curr. Biol.* **35**, 1353-1363.e5 (2025).

52. Di Nardo, M., Pallotta, M. M. & Musio, A. The multifaceted roles of cohesin in cancer. *Journal of Experimental and Clinical Cancer Research* **41**, (2022).

53. Scott, J. S. *et al.* Emerging roles of cohesin-STAG2 in cancer. *Oncogene* **44**, 277–287 (2025).

54. Yesbolatova, A., Natsume, T., Hayashi, K. ichiro & Kanemaki, M. T. Generation of conditional auxin-inducible degron (AID) cells and tight control of degron-fused proteins using the degradation inhibitor auxinole. *Methods* **164**–**165**, (2019).

55. Sakata, T., Shirahige, K. & Sutani, T. ChIP-seq analysis of condensin complex in cultured Mammalian cells. in *Methods in Molecular Biology* **1515**, (2017).

56. Gregersen, L. H., Mitter, R. & Svejstrup, J. Q. Using TTchem-seq for profiling nascent transcription and measuring transcript elongation. *Nat. Protoc.* **15**, (2020).

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

58. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, (1983).

59. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, (2009).

60. Nakato, R., Itoh, T. & Shirahige, K. DROMPA: Easy-to-handle peak calling and visualization software for the computational analysis and validation of ChIP-seq data. *Genes to Cells* **18**, (2013).

61. Nakato, R. & Sakata, T. Methods for ChIP-seq analysis: A practical workflow and advanced applications. *Methods* **187**, (2021).

62. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, (2016).

63. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, (2013).

64. Li, B. & Dewey, C. N. RSEM Li and Dewey, 2011. *BMC Bioinformatics* **12**, (2011).

65. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, (2009).

66. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, (2009).

67. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, (2015).

68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, (2014).

69. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, (2014).

70. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, (2012).

71. Prudêncio, P., Rebelo, K., Grosso, A. R., Martinho, R. G. & Carmo-Fonseca, M. Analysis of Mammalian Native Elongating Transcript sequencing (mNET-seq) high-throughput data. *Methods* **178**, (2020).

72. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, (2010).

73. Open2C *et al.* Pairtools: from sequencing data to chromosome contacts. *bioRxiv Prepr. Serv. Biol.* (2023). doi:10.1101/2023.02.13.528389

74. Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst.* **3**, (2016).

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

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

77. Ramírez, F. *et al.* High-resolution TADs reveal DNA sequences underlying genome organization in

flies. *Nat. Commun.* **9**, (2018).

78. Flyamer, I. M., Illingworth, R. S. & Bickmore, W. A. Coolpup.py: Versatile pile-up analysis of Hi-C data. *Bioinformatics* **36**, (2020).

79. Chen, K. *et al.* DANPOS: Dynamic analysis of nucleosome position and occupancy by sequencing. *Genome Res.* **23**, (2013).

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

Acknowledgements

We would like to thank all members of Shirahige Laboratory for discussions. This work was supported by JST-CREST Grant Number JPMJCR18S5, grant-in-Aid for Transformative Research Areas (A) Grant Number 20H05940 and 20H05933, AMED-ASPIRE-A, JSPS Grant-in-Aid for Scientific Research (S) Grant Number 20H05686, and AMED BINDS grant number 22ama121020j0001.

Author Contributions

Conceptualization, S.T., T.Sakata., M.B., and K.S.; investigation, S.T., A.Y., T.Sakata., M.B.; formal analysis, T. Sutani.; resources, T.M. and M.T.K.; funding acquisition, K.S.; visualization, S.T.; writing – original draft, S.T.; writing – review & editing, S.T., T.Sakata., M.B., and K.S.; supervision, K.S.

Competing Interests statement

The authors declare no competing interests.

Figure Legends

Fig. 1 | Cohesin promotes Pol II pausing at highly paused genes. a MA plot of TT-seq showing gene expression changes upon auxin-induced cohesin depletion. HCT-116 cells expressing AID-tagged RAD21 were treated with auxin for 3 hours to degrade RAD21, and TT-seq was performed. Genes included in the analysis had TT-seq TPM ≥ 1, length ≥ 5 kb, and were ≥1 kb away from neighboring genes. Differentially expressed genes were identified using edgeR (FDR < 0.05) based on two biological replicates. **b** Heatmaps of mNET-seq signals across genes (TSS −10 kb to +100 kb) in control and auxin-treated cells. Genes are aligned at their transcription start site (TSS) and sorted by gene length. **c** Heatmaps of mNET-seq, TT-seq in control and auxin-treated cells, alongside MNase-ChIP-seq (H3K4me3) signals in control cells at promoters (TSS −250 bp to +250 bp). Genes are aligned at their TSS and sorted by the distance between TSS and pausing site. Pausing sites, defined as the highest mNET-seq peak between the TSS and +1 nucleosome, are highlighted with a dashed line. **d** Box plot of the pausing-duration index in control and auxin-treated cells (Left) and their log2 fold changes (auxin/control) (Right). The pausing-duration index is calculated as the ratio of mNET-seq to TT-seq at the pausing site (pausing site −25 bp to +25 bp). Genes were evenly divided into four groups based on their pausing-duration index in control cells. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test (Left). (**p < 0.01, ***p < 0.001). Exact p-values from left to right are: 0.0012, 3.55×10^{-27}, 5.38×10^{-43}, and 2.10×10^{-36}.

Fig. 2 | Cohesin stabilizes the elongation complex to suppress premature transcription termination. a Box plots of the relative levels of PAF1 (Left) and CDC73 (Right) to RPB1 in control and auxin-treated cells. Their relative levels were calculated at promoters (TSS -1 kb to +1 kb) and gene bodies (TSS +1 kb to 3'gene end). Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. (***p < 0.001). Exact p-values: 5.60×10^{-6} (PAF1 promoter), 4.71×10^{-121} (PAF1 gene body), 0.191 (CDC73 promoter), and 1.34×10^{-298} (CDC73 gene body). **b** Co-immunoprecipitation showing the interaction between Pol II and PAF1 complex. Cells were treated with DRB for 3 hours and harvested 20 minutes after DRB removal. Nuclear extracts were prepared, and Pol II was immunoprecipitated using antibodies against RPB1 CTD or Ser2-phosphorylated RPB1 (Ser2P). In the RPB1 IP fraction, band intensities of PAF1, CDC73, and CTR9, normalized to RPB1, were reduced to 59%, 55%, and 75% after

auxin treatment. In the Ser2P IP fraction, intensities of PAF1, CDC73, and CTR9, normalized to Ser2P, were reduced to 69%, 73%, and 72%. **c** Heatmaps of log□ fold changes (siXRN2/control and auxin + siXRN2/auxin) in RNA-seq at gene ends (3' gene end −10 kb to +5 kb). Genes are aligned at their 3'gene end and sorted by gene length. **d** Venn diagrams of the overlap among upregulated or downregulated genes identified upon auxin, siXRN2, and auxin + siXRN2 treatments. Differentially expressed genes were identified using edgeR (FDR < 0.01) based on three biological replicates. **e** Average profiles of RNA-seq log□ fold changes (auxin vs control, siXRN2 vs control, and auxin + siXRN2 vs control). Profiles are shown for genes that were upregulated or downregulated only in auxin + siXRN2-treated cells.

Fig. 3 | Cohesin and its loader dynamically colocalize with RNA Polymerase II at promoters. a Average ChIP-seq profiles of RPB1, RAD21, MAU2, and NIPBL under control, DRB treatment, and DRB washout conditions at promoters (TSS −2.5 kb to +2.5 kb). For the DRB condition, cells were treated with DRB for 3 hours. For the washout condition, cells were collected 10 minutes after DRB removal. **b** Co-immunoprecipitation showing the interaction between cohesin loader and Pol II. Cells expressing FLAG-tagged endogenous NIPBL were treated with NIPBL-targeting siRNA oligos. Nuclear extracts were prepared, and NIPBL was immunoprecipitated using either anti-FLAG or anti-NIPBL antibodies. **c** Co-immunoprecipitation showing the interaction between cohesin loader and Pol II. Cells with or without FLAG-tagged endogenous NIPBL expression were treated with flavopiridol, triptolide, or untreated. Nuclear extracts were prepared, and NIPBL was immunoprecipitated using anti-FLAG antibodies.

Fig. 4 | Cohesin and its loader associate with the transcription complex during the pause–release transition. a Schematic of the *in vitro* transcription assay. A DNA template containing five GAL4 binding sites and a TATA promoter was immobilized on magnetic beads, preloaded with GAL4-VP16, and subsequently incubated with HeLa nuclear extract. After washing, template-bound factors were eluted for analysis by Western blot. **b** *In vitro* transcription assay to assess factor recruitment upon activator addition. Template DNA, preloaded with GAL4-VP16, was incubated with nuclear extracts (Input) for the indicated times. In NTP condition, NTPs were added 40 min after the start of the incubation. In NTP+DRB condition, DRB was added at the start of incubation, and NTPs were added 40 min later. **c** *In vitro* transcription assay to assess the effect of cohesin ATPase inhibition by AlFx. Cohesin was depleted from nuclear

extract using antibodies against SMC1/SMC3. Template DNA, preloaded with GAL4-VP16, was incubated with nuclear extracts (Input). After a 20-min incubation, NTPs were added, followed 3 min later by AlFx (0, 0.5, and 1.5□mM). **d** *In vitro* transcription assay to assess the effect of CDK9 inhibition on AlFx-induced state. Cohesin was depleted from nuclear extract using antibodies against SMC1/ SMC3. Nuclear extracts were pre-incubated with CDK9 inhibitor flavopiridol (FLV). Template DNA, preloaded with GAL4-VP16, was incubated with nuclear extracts (Input). NTPs were added 20 min after the start of incubation, followed by AlFx addition 3 min after NTPs. **e** Schematic summary of factor recruitment across transcriptional stages. Presence or absence of Ser5P, Ser2P, NELF, and DSIF at each transcriptional stage was inferred from previous studies. NIPBL and MAU2 were detected upon addition of the activator, NTPs, and DRB (b), indicating that the cohesin loader is present during the pre-initiation, initiation, and pausing. The AlFx-induced state contained cohesin together with SPT5 (DSIF) and NELFA (NELF) (c), and was abolished by CDK9 inhibition (c). Based on these observations, we placed this state between pausing and release because NELF complex dissociates when Pol II is released from pausing. N.D indicates not determined.

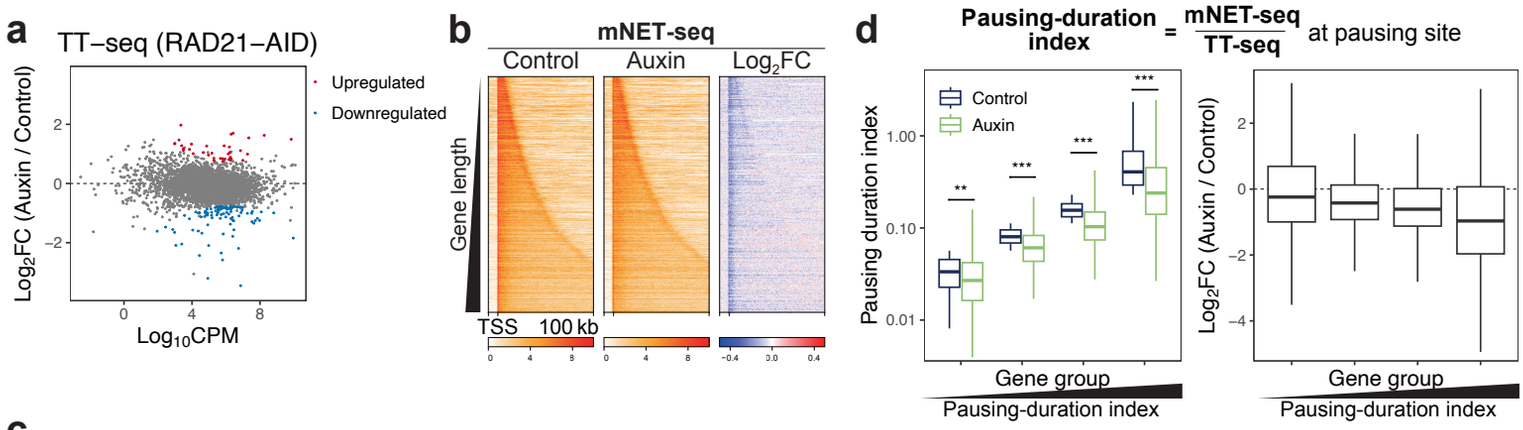
Fig. 5 | Cohesin maintains the H3K27ac mark at promoters. a Heatmaps of H3K27ac ChIP-seq signals at promoters (TSS -2.5 kb to +2.5 kb) in control and auxin-treated cells. Promoters are sorted by H3K27ac signal intensity in control cells. **b** Comparison of promoter H3K27ac (TSS -1 kb to +1 kb) between control and auxin-treated cells. Upon cohesin depletion, H3K27ac levels decreased at 92% of promoters, increased at 2%, and remained unchanged at 6% (FDR < 0.05, binomial test). **c** Heatmaps of TAF1 ChIP-seq signals at promoters (TSS -2.5 kb to +2.5 kb) in control and auxin-treated cells. Promoters are shown in the same order as in (a). **d** Comparison of promoter TAF1 and H3K27ac (TSS -1 kb to +1 kb) in control cells (Left), and their log2 fold changes (auxin/control) (Right). P-value (p) and Spearman's correlation coefficient (ρ) were calculated using Spearman's rank correlation test.

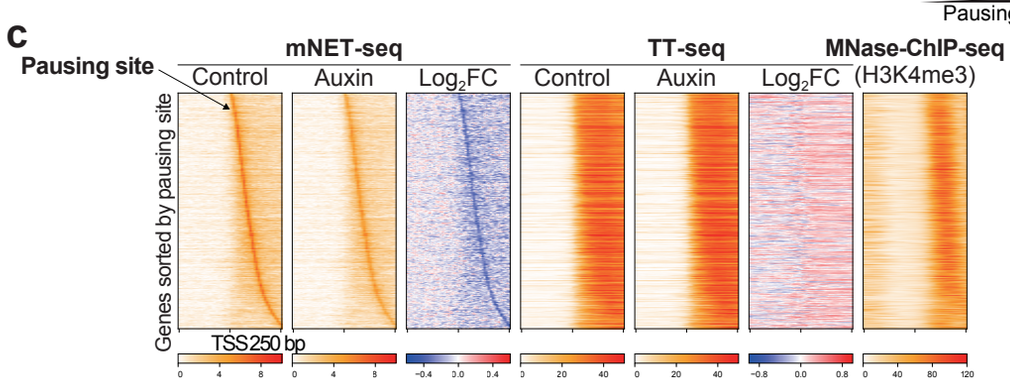
Fig. 6 | Cohesin regulates the genomic distribution of H3K27ac via promoter–enhancer contacts. a Heatmaps of H3K27ac ChIP–seq signals in control and auxin-treated cells, alongside H3K4me1 and H3K4me3 signals in control cells, at regulatory elements (promoters and enhancers). Regulatory

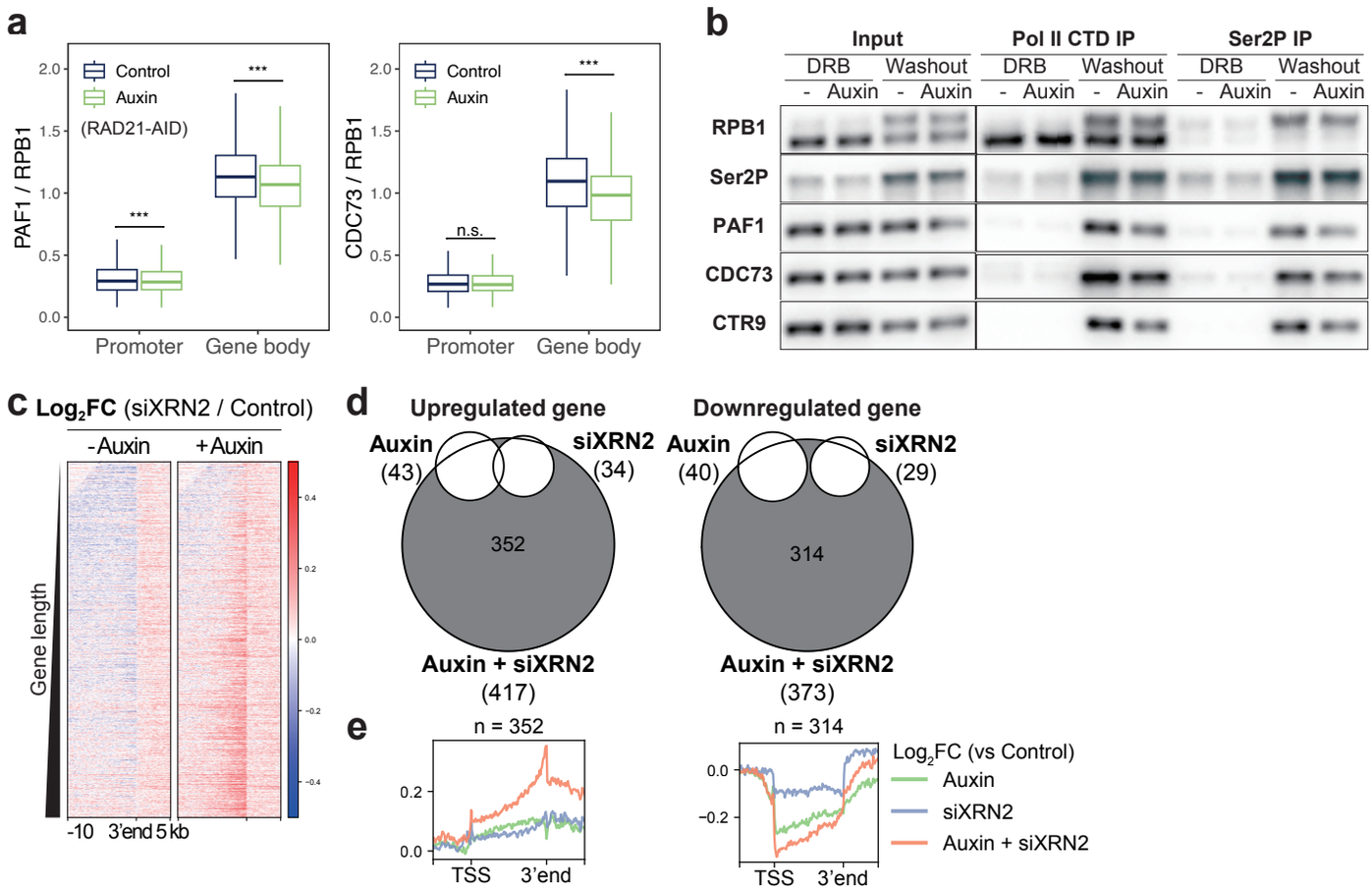
elements are sorted by fold change (auxin/control) in H3K27ac signal. **b** Heatmap of the log2 fold change (obs/exp) of transitions between consecutive regulatory elements. Regulatory elements were categorized by H3K27ac change upon cohesin depletion (increased, unchanged, or decreased). Transition probabilities were calculated from their order along the genome and compared with the expected probabilities assuming a uniform distribution. **c** Aggregate plots of Micro-C contacts between regulatory elements. Regulatory elements were categorized by H3K27ac change upon cohesin depletion (increased, unchanged, or decreased). All six possible pairwise combinations of these categories are shown. **d** Box plots of genomic distances of chromatin contacts between regulatory elements. Regulatory elements were categorized by H3K27ac change upon cohesin depletion (increased, unchanged, or decreased). Contacts were defined as paired-end Micro-C reads connecting two regulatory elements. All six possible pairwise combinations of these categories are shown. **e** Proposed model illustrating how cohesin-mediated promoter–enhancer interactions support transcription. Enhancers act as a source of transcriptional cofactors that are delivered to promoters through cohesin-dependent chromatin contacts. Upon cohesin depletion, this supply is disrupted, causing cofactors to accumulate at enhancers and decrease at promoters.

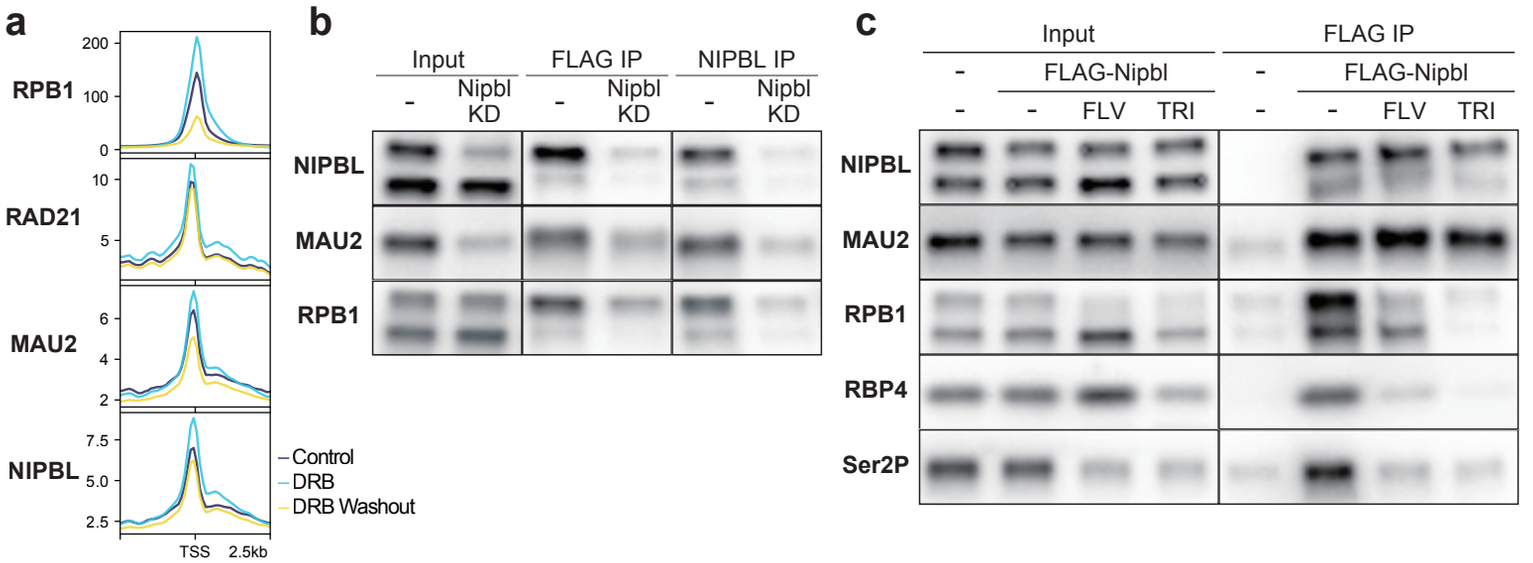
Fig. 7 | Cohesin depletion reduces initiation but enhances pause release, counterbalancing steady-state gene expression yet impairing rapid gene induction. a Heatmaps showing log□ fold changes (auxin/control) in promoter H3K27ac, pausing-duration index, and TT-seq gene expression. Genes were grouped into three clusters by K-means clustering based on log□ fold changes in these parameters and the velocity index (Supplementary Fig. 9b). The pausing-duration index was inverted (1/x) to align its directionality with the others. Upregulated or downregulated genes (Fig. 1a) are highlighted with red and blue lines, respectively. **b** Box plots showing log2 fold changes (auxin/control) in promoter H3K27ac, pausing-duration index, and TT-seq gene expression across the clusters in (a). Statistical significance was calculated using pairwise Welch's t-tests between clusters. (***$p < 0.001$). Exact p-values for promoter H3K27ac: 3.70×10^{-146} (Cluster1 vs 2), 2.67×10^{-158} (Cluster1 vs 3), 2.33×10-213 (Cluster2 vs 3); for pausing-duration index: 1.21×10^{-54} (Cluster1 vs 2), 1.99×10^{-39} (Cluster1 vs 3), 7.40×10^{-96} (Cluster2 vs 3); for TT-seq gene expression: 8.09×10^{-264} (Cluster1 vs 2), 6.08×10^{-264} (Cluster1 vs 3), 4.26×10^{-249} (Cluster2 vs 3). **c** Bar plots of simulated effects of cohesin depletion on productive

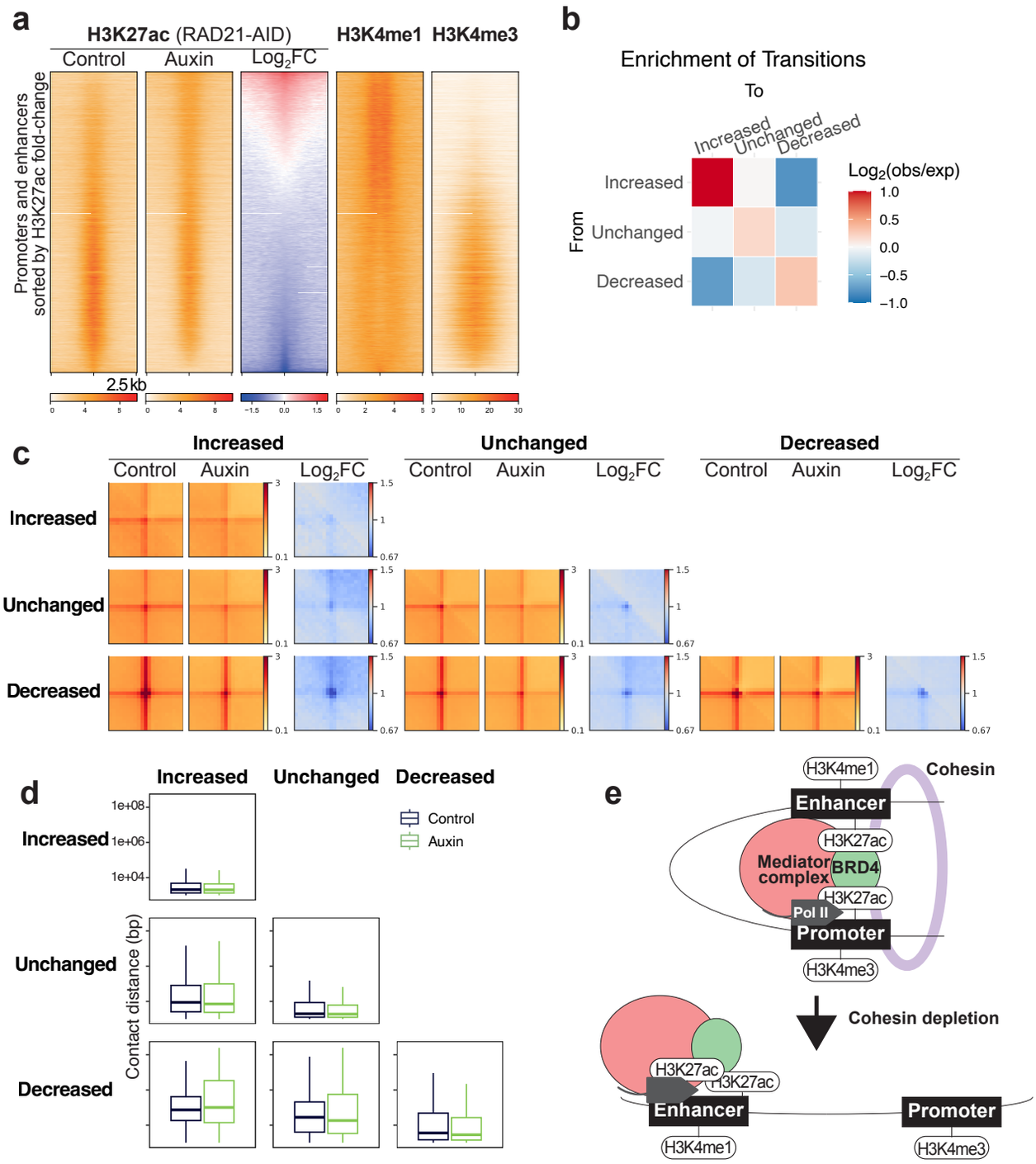
transcription. Reduction in initiation and enhancement of pause release were modeled as decreased p12 and increased p34, respectively. Productive transcription was defined as the probability of Pol II being in the Elongation state and normalized to 1 in the control. **d** Representative gene showing reduced TNFα-induced expression upon auxin treatment. EU-seq was performed at 0, 30, 60, and 90 min after TNFα treatment. **e** Line plots of gene expression fold changes (each time point/ 0 min) at TNFα-induced genes in control and auxin-treated cells. Genes upregulated upon TNFα treatment in control cells (edgeR; FDR < 0.05, fold change > 2) were classified as early, mid, or late-induced genes based on the time point of peak expression. Data represent mean ± standard error. Statistical significance was calculated using a two-sided Wilcoxon rank-sum test. (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$).

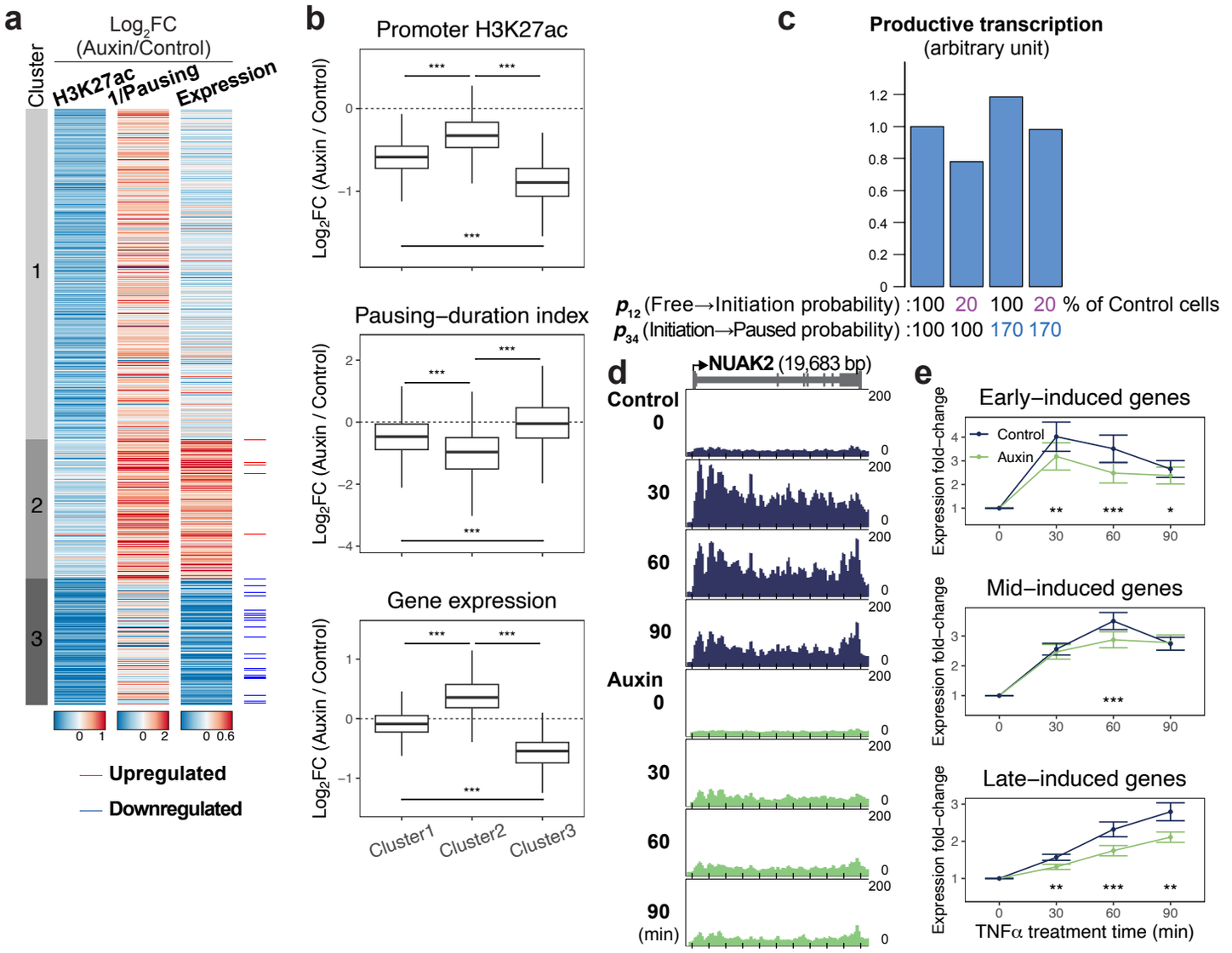

a
TT−seq (RAD21−AID)
Upregulated
Downregulated
Log2FC (Auxin / Control)
Log10CPM
b
mNET-seq
Control
Auxin
Log2FC
Gene length
TSS
100 kb
d
Pausing-duration index = mNET-seq / TT-seq at pausing site
Control
Auxin
Pausing duration index
Gene group
Pausing-duration index
Log2FC (Auxin / Control)

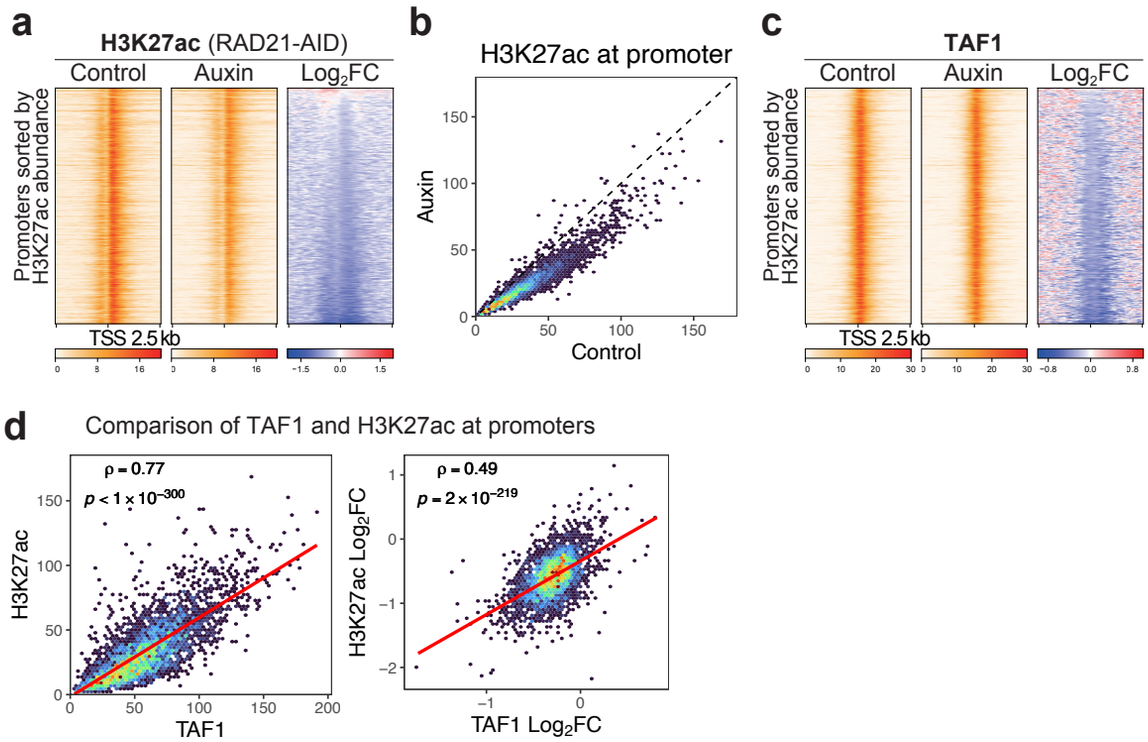

c
Pausing site
mNET-seq
Control
Auxin
Log2FC
TT-seq
Control
Auxin
Log2FC
MNase-ChIP-seq
(H3K4me3)
Genes sorted by pausing site
TSS
250 bp

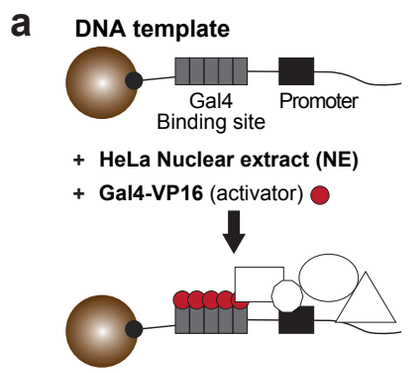

a
Control
Auxin
(RAD21-AID)
PAF1 / RPB1
CDC73 / RPB1
***
n.s.
Promoter
Gene body
b
Input
Pol II CTD IP
Ser2P IP
DRB
Washout
- Auxin
RPB1
Ser2P
PAF1
CDC73
CTR9
c
Log2FC (siXRN2 / Control)
-Auxin
+Auxin
Gene length
-10
3'end
5 kb
d
Upregulated gene
Downregulated gene
Auxin
siXRN2
(43)
(34)
352
(417)
(40)
(29)
314
(373)
Auxin + siXRN2
e
n = 352
n = 314
TSS
3'end
Log2FC (vs Control)
Auxin
siXRN2
Auxin + siXRN2

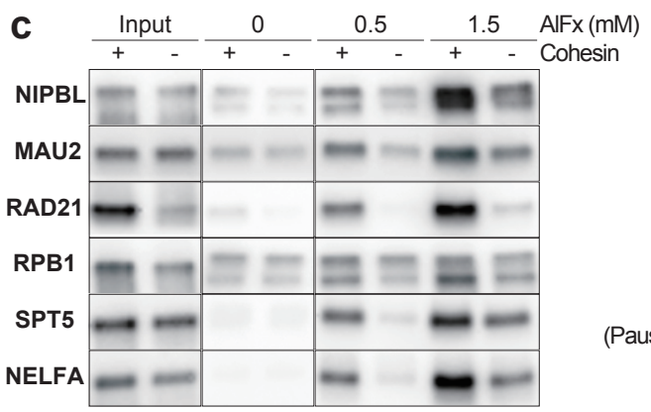

a
RPB1
RAD21
MAU2
NIPBL
200
100
0
10
5
6
4
2
7.5
5.0
2.5
TSS
2.5kb
Control
DRB
DRB Washout
b
Input
FLAG IP
NIPBL IP
-
Nipbl KD
NIPBL
MAU2
RPB1
c
Input
FLAG IP
FLAG-Nipbl
FLV
TRI
NIPBL
MAU2
RPB1
RBP4
Ser2P

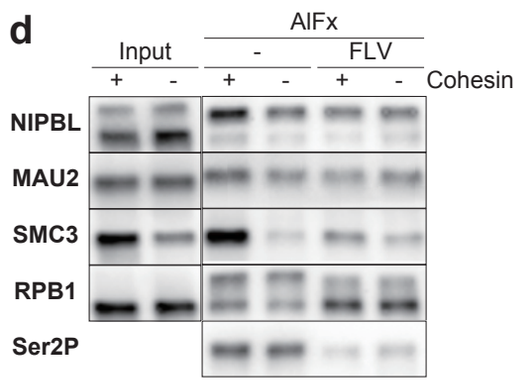

a
H3K27ac (RAD21-AID)
Control
Auxin
Log2FC
H3K4me1
H3K4me3
Promoters and enhancers sorted by H3K27ac fold-change
2.5 kb
b
Enrichment of Transitions
To
Increased
Unchanged
Decreased
From
Log2(obs/exp)
1.0
0.5
0.0
−0.5
−1.0
c
Increased
Unchanged
Decreased
d
Contact distance (bp)
1e+08
1e+06
1e+04
Control
Auxin
e
H3K4me1
Cohesin
Enhancer
H3K27ac
Mediator complex
BRD4
Pol II
Promoter
H3K4me3
Cohesin depletion

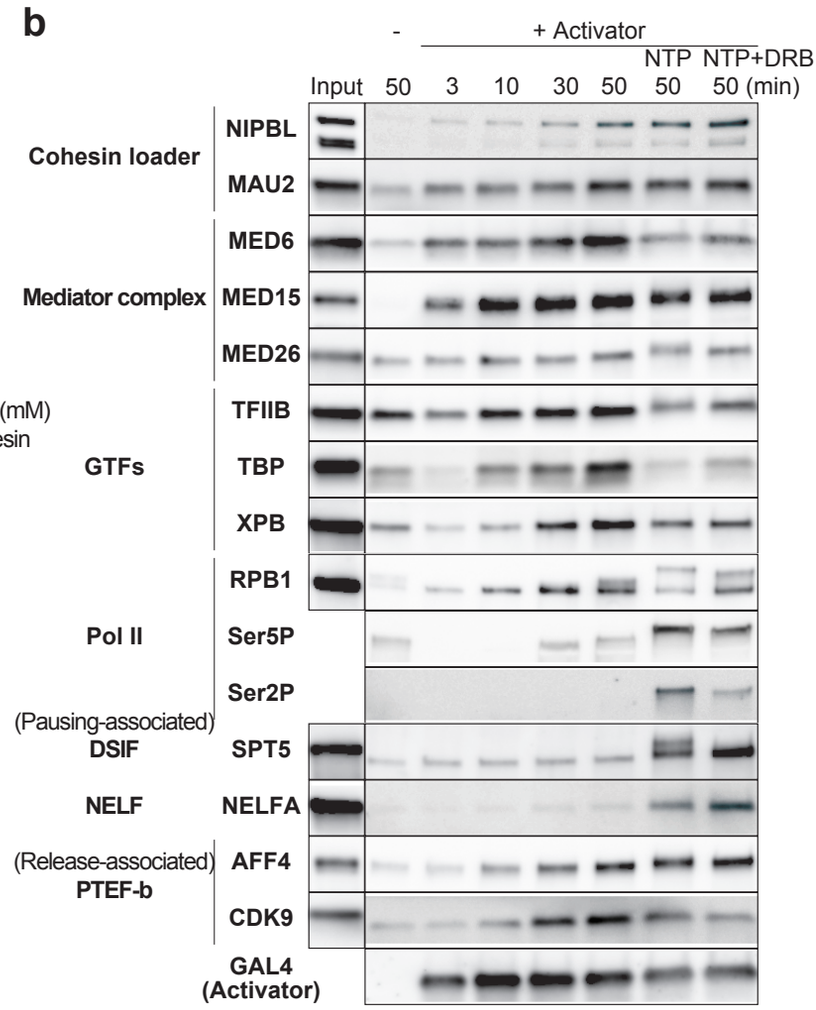

a
Log2FC (Auxin/Control)
Cluster
H3K27ac
1/Pausing
Expression
1
2
3
0 1
0 2
0 0.6
Upregulated
Downregulated
b
Promoter H3K27ac
Log2FC (Auxin / Control)
Pausing−duration index
Gene expression
Cluster1
Cluster2
Cluster3
c
Productive transcription
(arbitrary unit)
p12 (Free→Initiation probability) :100 20 100 20 % of Control cells
p34 (Initiation→Paused probability) :100 100 170 170
d
NUAK2 (19,683 bp)
Control
Auxin
0
30
60
90
(min)
e
Early−induced genes
Mid−induced genes
Late−induced genes
Control
Auxin
Expression fold−change
TNFα treatment time (min)


a
H3K27ac (RAD21-AID)
Control
Auxin
Log2FC
Promoters sorted by H3K27ac abundance
TSS 2.5 kb
b
H3K27ac at promoter
Auxin
Control
c
TAF1
Control
Auxin
Log2FC
Promoters sorted by H3K27ac abundance
TSS 2.5 kb
d
Comparison of TAF1 and H3K27ac at promoters
ρ = 0.77
p < 1 × 10−300
H3K27ac
TAF1
ρ = 0.49
p = 2 × 10−219
H3K27ac Log2FC
TAF1 Log2FC









e

FLV (iCDK9) ⊥ ADP-AlFx ⊥

pre-Initiation → Initiation → Pausing → Intermediate → Release

	pre-Initiation	Initiation	Pausing	Intermediate	Release
Ser5P	-	+	+	+	+
Ser2P	-	-	-	+	+
NELF	-	-	+	+	-
DSIF	-	-	+	+	+
Cohesin	-	-	-	+	N.D.
Cohesin loader	+	+	+	+	N.D.