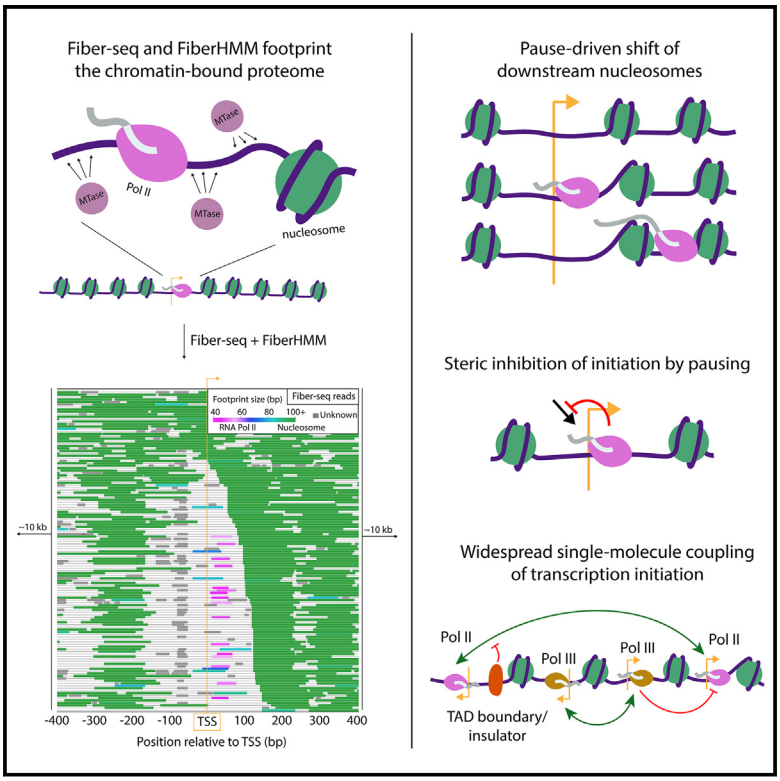
Molecular Cell

Article

# RNA polymerases reshape chromatin architecture and couple transcription on individual fibers

## Graphical abstract




## Highlights

- Fiber-seq and FiberHMM visualize the chromatin-bound proteome on 15–20 kb fibers
- Promoter-proximal pausing destabilizes and shifts downstream nucleosomes
- Transcription initiation is coupled between proximal promoter and enhancer elements
- Clustered tRNA and 5S rRNA genes exhibit coupled RNA Pol III transcription

## Authors

Thomas W. Tullius, R. Stefan Isaac, Danilo Dubocanin, Jane Ranchalis, L. Stirling Churchman, Andrew B. Stergachis

## Correspondence

churchman@genetics.med.harvard.edu (L.S.C.), absterga@uw.edu (A.B.S.)

## In brief

Tullius et al. use Fiber-seq and FiberHMM, a single-molecule, long-read genomic footprinting approach, to characterize the interplay between transcription and the broader chromatin landscape in *Drosophila*. They find that polymerase pausing destabilizes downstream nucleosomes and that transcription is coupled in a distance-dependent manner between nearby promoter and enhancer-promoter pairs.

Tullius et al., 2024, Molecular Cell *84*, 3209–3222
September 5, 2024 
https://doi.org/10.1016/j.molcel.2024.08.013

CellPress

Molecular Cell

CellPress

Article

# RNA polymerases reshape chromatin architecture and couple transcription on individual fibers

Thomas W. Tullius,[1] R. Stefan Isaac,[1] Danilo Dubocanin,[2] Jane Ranchalis,[2] L. Stirling Churchman,[1,*] and Andrew B. Stergachis[2,3,4,5,*]

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[2]Division of Medical Genetics, Department of Medicine, University of Washington, Seattle, WA, USA
[3]Department of Genome Sciences, University of Washington, Seattle, WA, USA
[4]Brotman Baty Institute for Precision Medicine, Seattle, WA, USA
[5]Lead contact
*Correspondence: churchman@genetics.med.harvard.edu (L.S.C.), absterga@uw.edu (A.B.S.)
https://doi.org/10.1016/j.molcel.2024.08.013

## SUMMARY

RNA polymerases must initiate and pause within a complex chromatin environment, surrounded by nucleosomes and other transcriptional machinery. This environment creates a spatial arrangement along individual chromatin fibers ripe for both competition and coordination, yet these relationships remain largely unknown owing to the inherent limitations of traditional structural and sequencing methodologies. To address this, we employed long-read chromatin fiber sequencing (Fiber-seq) in *Drosophila* to visualize RNA polymerase (Pol) within its native chromatin context with single-molecule precision along up to 30 kb fibers. We demonstrate that Fiber-seq enables the identification of individual Pol II, nucleosome, and transcription factor footprints, revealing Pol II pausing-driven destabilization of downstream nucleosomes. Furthermore, we demonstrate pervasive direct distance-dependent transcriptional coupling between nearby Pol II genes, Pol III genes, and transcribed enhancers, modulated by local chromatin architecture. Overall, transcription initiation reshapes surrounding nucleosome architecture and couples nearby transcriptional machinery along individual chromatin fibers.

## INTRODUCTION

Eukaryotic RNA polymerase II (Pol II) transcription initiates in a highly regulated manner along a genome that is bound by numerous DNA-associated proteins, including nucleosomes, transcription factors (TFs), chromatin remodelers, and other RNA polymerase complexes. Initiation begins with preinitiation complex (PIC) assembly, consisting of Pol II and an array of general TFs that work in unison to initiate transcription. Following PIC assembly and Pol II release, Pol II generally transcribes 25–50 bp and pauses at the promoter-proximal pause (PPP) site, a critical step in transcription that allows the integration of regulatory signals before proceeding into productive elongation.[1,2]

The chromatin environment in which transcription occurs creates a spatial arrangement conducive to both competition and coordination along individual chromatin fibers. Nucleosomes in particular serve as a physical barrier to transcription initiation, pausing, and elongation.[3–8] Genes with higher Pol II pausing rates are associated with distinct promoter-proximal nucleosome patterns.[8] However, it is unclear whether these patterns reflect paused Pol II directly modulating the positioning of assembled nucleosomes that it is competing with along an individual fiber, or if they represent a distinct promoter structure at highly paused genes. Further, transcription initiation occurs in spatial proximity to nearby transcriptionally active promoters and enhancers—a configuration that may promote direct coupling between these neighboring elements.[9–11] For example, neighboring promoters have correlated expression patterns, transcriptional activity at enhancers is associated with neighboring gene expression, and some transcriptionally active genes co-localize at specific foci within the nucleus.[12–16] It is unknown, though, if these associations on a genomic scale are merely correlative and driven by a shared permissive chromatin architecture, or if they actually represent transcriptional coupling (i.e., the coordinated synchronous occupancy of transcriptional machinery at two or more sites along an individual chromatin fiber within the nucleus).

Although pioneering cryo-electron microscopy and biophysical studies have provided a detailed single-molecule understanding of the mechanism of transcription initiation itself along individual DNA templates,[4–6,17–20] our understanding of how

this process occurs globally across multi-kilobase chromatin fibers remains limited. The global analysis of multiple genomic phenomena, such as transcription and nucleosome positioning, is limited by correlative comparisons between short-read functional genomics datasets, such as NET-seq, PRO-seq, and MNase-seq. To directly test models of how these processes interact, we need to investigate their co-occurrence across single chromatin fibers. Emerging single-molecule footprinting methods[21–25] have the potential to achieve this new genomic perspective; however, most approaches have been limited to analyzing nucleosomes due to the uncertainty of identifying smaller footprints. Short-read single-molecule footprinting of promoter regions has circumvented this issue, providing insights into local Pol II occupancy patterns during initiation,[21] but a global view of promoter patterns of Pol II, the interactions between promoter Pol II and adjacent nucleosomal arrays, and long-range interactions between promoters, enhancers, and other elements remains obscure.

In this study we leverage Fiber-seq, a high-resolution, long-read, single-molecule chromatin fiber sequencing method, to directly characterize the single-molecule competition and coordination between RNA polymerases and the surrounding chromatin proteome along individual chromatin fibers.[22] Fiber-seq uses a nonspecific $N^6$-adenine methyltransferase (m6A-MTase) to methylate accessible bases throughout the genome, which are directly read using PacBio sequencing, producing single-molecule maps of accessibility on 15–20 kb chromatin fibers.[22,26] Inaccessible patches within Fiber-seq reads represent the footprints of the chromatin-bound proteome. However, technical challenges associated with discerning the occupancy patterns of myriad chromatin components found along the same fiber have restricted previous analyses to nucleosomes. In this study we developed a hidden Markov model (HMM)-based footprint caller (FiberHMM) that allows for the identification and quantification of paused Pol II, PIC, nucleosome, TF, and Pol III transcription-associated footprints in Fiber-seq data. Using these single-molecule data in combination with targeted perturbation assays, we characterize transcription initiation dynamics and the mechanism for the inhibition of initiation by Pol II pausing. Further, we discover genome-wide, pause-driven changes to nucleosome architecture and direct transcriptional coupling between nearby Pol II genes, Pol III genes, and transcribed enhancers within spatially organized chromatin domains.

## RESULTS

### Identification of single-molecule RNA polymerase II footprints via FiberHMM

To identify single-molecule RNA polymerase-related footprints, we applied Fiber-seq to *Drosophila* S2 cells, resulting in ~160× genome-wide sequencing coverage (12 kb average read length). Nucleosomes, with ~140 bp footprints, are readily visible in Fiber-seq reads and can be identified easily due to the extremely low probability of observing on average 70 consecutive unmethylated A/T base pairs. Smaller footprints can be ambiguous, though, as they could result from sequences with a low number of adenosines, intrinsic biases of the methyltransferase, or errors in PacBio detection of methylated residues. To overcome these issues, we developed FiberHMM to detect unmethylated patches of adenines along each fiber corresponding not just to nucleosomes, as had been previously identified,[22] but also Pol II and other chromatin-bound proteins. Specifically, FiberHMM incorporates observed adenine methylation patterns as well as false-negative and false-positive methylation probabilities for each base. These probabilities are derived from control Fiber-seq data from methyltransferase-treated dechromatinized genomic DNA (gDNA) and methyltransferase-untreated genomic DNA, respectively. These false-negative and -positive methylation rates are used as fixed emission probability parameters within the HMM, while the starting and transition probabilities are trained using a subset of the Fiber-seq data. This trained model is then used to call gaps in methylation that represent protein footprints along each read at nearly single-base resolution (Figure 1A).

RNA polymerase II footprints have been previously observed using short-read MTase-based approaches.[21] As such, we sought to determine whether we could similarly identify PICs and PPPs among the FiberHMM-derived footprints at genomic sites known to harbor these complexes. When we aligned Fiber-seq footprints on a metagenomic scale around transcription start sites (TSSs), two distinct footprint populations were visible: the first directly overlapping the TSS and a second located 0–60 bp downstream of the TSS (Figure 1B). This first footprint population matches the previously observed location of PICs, whereas the latter matches the previously observed location of PPPs. The size ranges of these putative PIC and PPP footprints ranged from 60 to 80 bp and from 40 to 60 bp, respectively. Importantly, these ranges mirror the expected sizes from previous biochemical and structural studies[27–29] (Table S1).

### Validated single-molecule transcription complex occupancy

We next validated these putative PIC and PPP footprints using orthologous datasets and chemical perturbations. First, we compared their positioning to short-read sequencing-derived transcript initiation and pausing data from CAGE-seq and PRO-seq, respectively.[30–33] As expected, putative PIC footprints are selectively positioned directly over CAGE-seq peaks, and putative PPP footprints are selectively positioned directly over PRO-seq peaks (Figure 1C). Second, we evaluated the enrichment of PPP footprints within TSSs for highly paused genes. As expected, putative PPP footprints were significantly enriched at genes with a high PRO-seq-derived pause index, and gene pause index and PPP footprint enrichment showed a monotonic relationship (Figures 1D and S1C). In contrast, the enrichment of putative PIC and PPP footprints were independent of pause index (Figure 1D) and expression, respectively (Figures S1A and S1B). Finally, we performed Fiber-seq on cells treated with triptolide, an initiation-inhibiting compound that during short treatments blocks progression from the PIC to the PPP via inhibition of transcription factor IIH (TFIIH) and during longer treatments leads to degradation of RNA-binding protein 1 (RBP1).[34,35] A short, 30 min treatment with triptolide significantly reduced the levels of putative PPP footprints, while having no effect on the levels of putative PIC footprints (Figures 1E, S1D, and S1E).

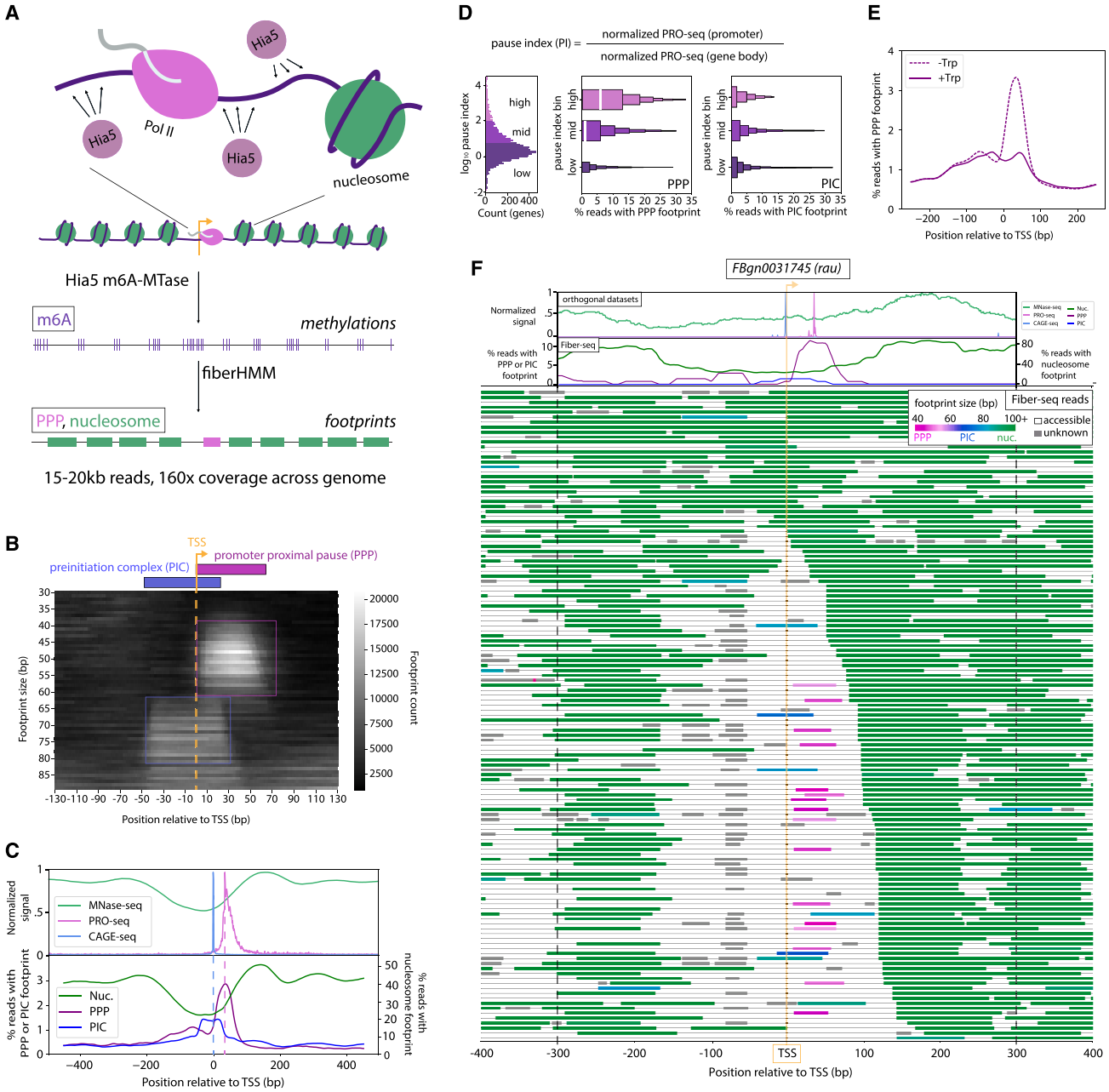



**Figure 1. Identification of Pol II footprints in Fiber-seq**

(A) Schematic of Fiber-seq. From top to bottom: *Drosophila* S2 cells were treated with Hia5 m6A methyltransferase and then sequenced using PacBio SMRT sequencing, directly detecting the methylated bases using Fibertools. Unmethylated footprints were called using FiberHMM, resulting in ~15–20 kb footprinted Fiber-seq reads.

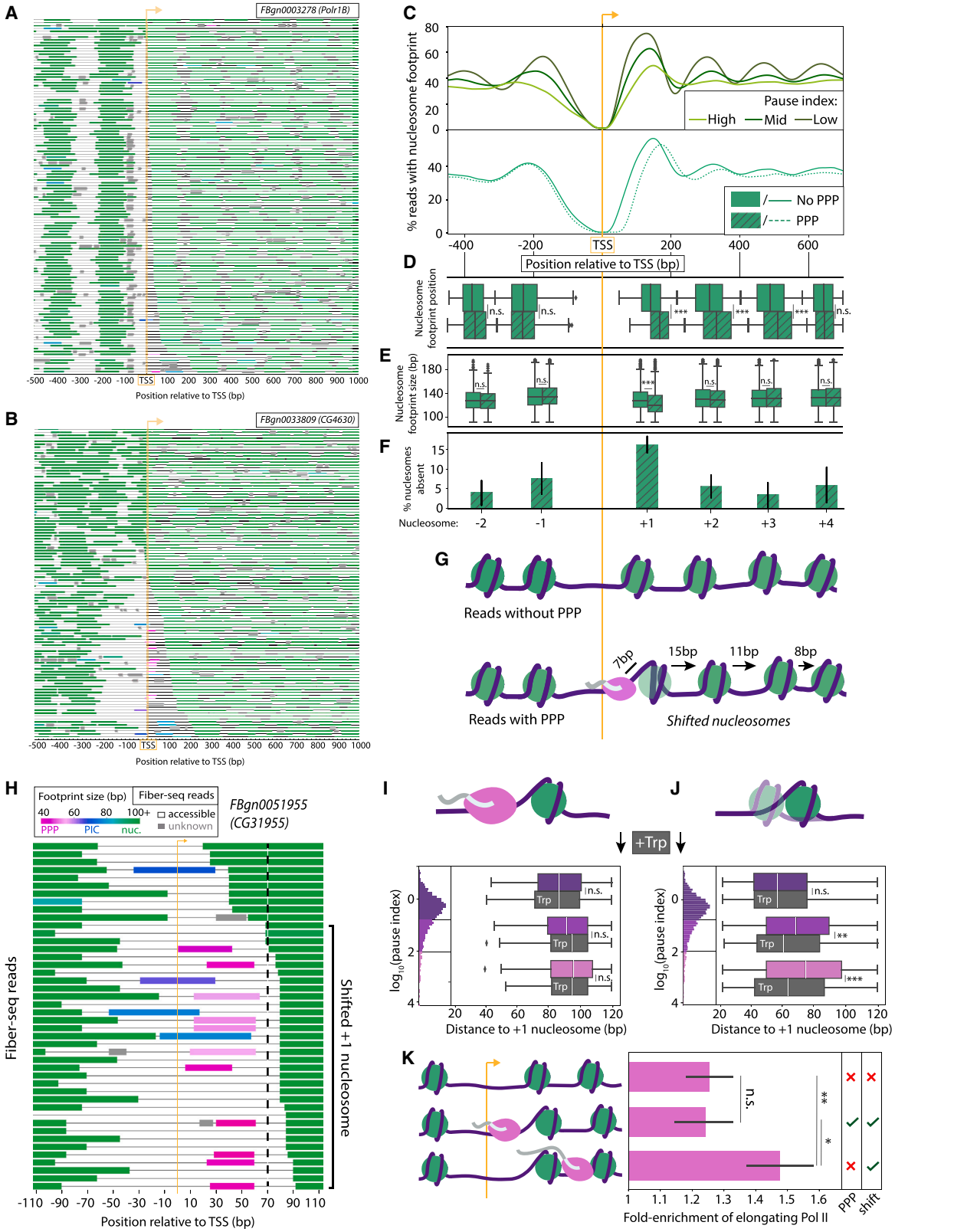
(B) Heatmap depicting the enrichment of differently sized Fiber-seq footprints around transcription start sites (TSSs) of genes with a pause index ≥10. The expected position of the PPP and PIC are schematized above, and PPP and PIC footprint populations are indicated with a box.

(C) Plot depicting (top) normalized MNase-seq, PRO-seq, and CAGE-seq signal and (bottom) the percent occupancy of nucleosome-, PPP-, and PIC-sized footprints at positions around TSSs of genes with a pause index ≥10. Dashed lines indicate CAGE-seq and PRO-seq peaks.

(D) Left: Histogram of the distribution of pause indexes of all genes, binned as follows: "high" (PI ≥ 100), "mid" (100 > PI ≥ 10), "low" (10 > PI). Letter-value plots showing the enrichment of (center) PPP and (right) PIC footprints in binned Fiber-seq reads.

(E) Plot showing the percent occupancy of PPP-sized footprints in cells untreated (dashed) or treated for 30 min with 10 μM triptolide (solid).

(F) Example Fiber-seq reads. MNase-seq, PRO-seq, and CAGE-seq signal (above) and enrichment of nucleosome-, PPP-, or PIC-sized footprints (middle). Below: Individual Fiber-seq reads with footprints colored by predicted identity (PPP, pink; PIC, blue; nucleosome, green; unknown, gray; black line, no footprint).




*(legend on next page)*

In addition to PIC and PPP footprints, we also identified nucleosome-sized footprints that are preferentially localized to sites that mirror bulk MNase-seq data, as well as smaller footprints upstream of the TSS that likely correspond to TF occupancy events (Figures 1F and S3A). Overall, each Fiber-seq read on average contains ~70 protein occupancy events defined by FiberHMM, with the majority of these being nucleosome footprints, enabling us to evaluate the single-molecule co-dependency between PIC, PPP, and nucleosome occupancy at individual loci genome wide (Figures 1F, S2O, S2P, and S4J–S4M).

## Pausing is associated with changes in nucleosome architecture

Genes with high pausing are known to be associated with altered chromatin patterns. Specifically, correlative studies using MNase-seq have reported that highly paused genes tend to have decreased promoter accessibility and shifted upstream and downstream nucleosomes.[8] However, directly disentangling whether this finding simply reflects differences in chromatin patterning at genes with a high or low pause index or actually results from Pol II pausing itself altering nucleosome positioning—an essential distinction for our understanding of eukaryotic gene regulation—remains unresolved. These models were difficult for previous studies to test, as direct connections between the presence of the PPP and nucleosome positioning on single chromatin fibers could not be measured. Instead, analyses were dependent on bulk comparisons of genes with different pause indices. A promoter's sequence, accessibility, and PPP occupancy all vary substantially between genes with a high versus low pause index, limiting the ability to translate these prior correlative findings into causal relationships. Taking advantage of the ability of Fiber-seq to capture the occupancy of individual polymerases and nucleosomes along a single chromatin fiber, we set out to test if pausing impacts nucleosome architecture.

In aggregate, we found that Fiber-seq footprints mirrored the MNase-seq-derived patterns of nucleosome enrichment seen at genes with different pausing and expression levels (Figures 2A, 2B, and S2A).[36] Similarly, we found that genes with higher pausing had more fibers with inaccessible promoters, in contrast to more highly expressed genes, which had more fibers with accessible promoters (Figures 2A, 2B, and S2B). Importantly, we observed that chromatin fibers with inaccessible versus accessible promoters had markedly distinct nucleosome architectures, which are unavoidably aggregated together when analyzing MNase-seq data (Figure S2C). To account for the impact chromatin fibers with inaccessible promoters have on nucleosome positioning, we selectively evaluated nucleosome patterns on fibers containing accessible promoters. This demonstrated that decreased promoter accessibility alone did not explain the disrupted upstream and downstream nucleosome positioning found at genes with high pause indices (Figures 2A–2C). In contrast, controlling for promoter accessibility, we found no clear difference in nucleosome architecture based on expression levels of genes, suggesting a unique association between the level of pausing at a promoter and the surrounding nucleosome landscape (Figure S2D).

Based on this analysis alone, though, we could not distinguish if these changes to nucleosome architecture reflected sequence differences of highly paused genes or if they resulted from the elevated levels of promoter-proximal pausing at those genes. To identify if there was a distinct nucleosome architecture specifically associated with pausing itself, we further divided Fiber-seq reads containing an accessible promoter based on whether they also contained a PPP footprint. We then compared the positioning of surrounding nucleosome footprints for each fiber type. The proportion of accessible fibers with a PPP footprint varies but are in the minority at all promoters (Figure 1D). To ensure that each promoter contributes the same number of fibers with and without a PPP, we subsampled an equal number

**Figure 2. Pausing drives changes to downstream nucleosomes**

(A and B) Fiber-seq reads from a low pause index gene (A) and a high pause index gene (B), with footprints colored by predicted identity (PPP, pink; PIC, blue; nucleosome, green; unknown, gray; black line, no footprint).

(C) Top: Plot showing the percent occupancy of nucleosome footprints in reads with an accessible TSS from genes binned by pause index: "high" (PI $\geq$ 100), "mid" (100 > PI $\geq$ 10), "low" (10 > PI). Bottom: Plot showing the percent occupancy of nucleosome footprints from promoter-accessible reads with (solid) or without (dashed) a PPP footprint, sampled to include an equal count of reads with or without a PPP footprint from each gene.

(D) Boxplot showing the position of nucleosomes relative to the TSS between accessible and sampled reads with (dashed) or without (solid) a PPP footprint. Nucleosomes were segmented using a Gaussian mixture model (GMM) based on a 95% posterior probability, with distributions compared with a two-sample t test (n.s. signifies $p > 0.05$, *** signifies $p$ value $<10^{-13}$).

(E) Boxplot showing the size of nucleosomes from reads with (dashed) or without (solid) a PPP footprint. Distributions were compared using a two-sample t test (n.s. signifies $p > 0.05$, *** signifies $p$ value $<10^{-14}$).

(F) Barplot showing the percent reduction of observed nucleosome footprints in reads with a PPP footprint compared to reads without a PPP footprint. Error bars represent a confidence interval derived from the bootstrapped distribution of this percentage.

(G) Cartoon summarizing the changes to nucleosome position and stability associated with the PPP.

(H) Sample of Fiber-seq reads from a representative locus, with footprints colored by predicted identity (PPP, pink; PIC, blue; nucleosome, green; unknown, gray). Reads with a shifted +1 nucleosome, defined as starting between 70 and 150 bp downstream of the TSS, are indicated with a line.

(I and J) Boxplots showing the distance to the +1 nucleosome footprint across reads with (I) or without (J) a PPP footprint. Reads are binned by pause index, indicated in the adjacent histogram: "high" (pink, PI $\geq$ 100), "mid" (purple, 100 > PI $\geq$ 10), "low" (dark purple, 10 > PI). Distances from −Trp and +Trp samples are plotted in color and in gray respectively. Significance is indicated between +Trp and −Trp pairs (two-sample t test, n.s. signifies $p > 0.05$, * signifies $p < 0.05$, ** signifies $p < 0$ 0.01, *** signifies $p < 0.0001$).

(K) Barplot showing the fold enrichment of elongating Pol II footprints within gene bodies over intergenic regions, divided into reads with no PPP and no shifted +1 nucleosome (top), a PPP and a shifted +1 nucleosome (middle), or no PPP and a shifted +1 nucleosome (bottom). Pairwise significance between bins is indicated (two-sample t test, n.s. signifies $p > 0.7$, * signifies $p$ value <0.03, ** signifies $p$ value <0.007).

of each fiber type at individual promoters. We found that PPP footprints across all genes were associated with disrupted downstream nucleosome positioning, irrespective of a gene's pause index. This effect was similar to the aggregate shift of downstream nucleosomes associated with high pause index (Figure 2C). However, no significant difference was observed for upstream nucleosomes, indicating that the shifted pattern of upstream nucleosomes at highly paused genes was independent of the PPP. In contrast, chromatin fibers containing TF footprints at GAGA factor (GAF) binding elements within the promoter had a shifted upstream nucleosome, with no significant difference in downstream nucleosome positioning, indicating that positioning of the upstream nucleosome is at least in part dependent on TF occupancy instead of PPP occupancy within the promoter (Figures S2O–S2Q).[37]

Overall, we observed that chromatin fibers containing a PPP footprint had multiple changes in the organization of the +1 to +4 nucleosomes. First, the +1 nucleosome was on average shifted downstream by ~15 bp with a footprint ~7 bp smaller, indicating that this nucleosome is often displaced and either partially unwound or a hexasome (Figures 2D, 2E, and S2F–S2I). The +2 and +3 nucleosomes also were on average significantly shifted downstream, ~11 bp and ~8 bp, respectively, but showed no size difference (Figures 2D and 2E). Second, we observed that some of these chromatin fibers were missing the +1 nucleosome, indicating that PPP footprints are associated with eviction of downstream nucleosomes (Figure 2F). In contrast, these effects were not seen upstream of the promoter or further than 1 kb downstream of the promoter and were similarly not seen in response to PIC occupancy within the promoter (Figures 2C–2F, S2E, S2J, and S2K). Taken together, these observations illustrate a nearby downstream nucleosome landscape significantly altered in association with promoter-proximal pausing at all genes (Figure 2G).

## Pausing drives the disruption of downstream nucleosomes

We next determined whether these altered downstream nucleosomes could be causally attributed to promoter-proximal pausing. Notably, at highly paused genes, we found an increase in the frequency of fibers with a shifted +1 nucleosome, even on fibers without a PPP footprint, showing that the +1 nucleosome shift can exist independently of PPP co-occupancy (Figure 2H and 2J). In contrast, at low-pause index genes, a shifted +1 nucleosome rarely occurred in the absence of a PPP footprint (Figure 2I). Together, these findings support dependence of the +1 nucleosome shift on pausing, and therefore a model whereby promoter-proximal pausing directly causes the +1 nucleosome to shift, which is stable for a brief period leading to high levels of shifted +1 nucleosomes at loci with high levels of pausing.

To test this pause-driven model directly and account for any variability in promoter architecture of high and low pausing genes, we performed Fiber-seq after reducing the formation of PPPs globally using triptolide. Triptolide treatment markedly reduced the proportion of reads containing a PPP footprint; however, the few remaining reads containing a PPP footprint still showed a downstream shift of the +1 nucleosome (Figure 2I). In contrast, far fewer chromatin fibers lacking a PPP footprint overall had a shifted +1 nucleosome, resulting in a dramatically reduced overall average shift of the +1 nucleosome across these fibers at all genes (Figure 2J). This reduction in fibers with a shifted +1 nucleosome was particularly strong at highly paused genes (Figure 2J), which also showed the largest net reduction in PPP footprints. Notably, the average + 1 nucleosome distance at highly paused genes now mirrored the distances observed on genes with low pausing in the untreated sample. Thus, the positioning of the +1 nucleosome even without an observed PPP footprint depends on a gene's overall level of pausing. Together, these data demonstrate that promoter-proximal pausing is causal of the shift in the +1 nucleosome and suggest that the shifted +1 nucleosome is stable for a brief but detectable period after the PPP has been released.

## Disrupted downstream nucleosomes are associated with active transcription

We hypothesized that if the reads with a shifted +1 nucleosome and no PPP footprint indeed represented a transient state after pausing, they would be enriched for chromatin fibers with a post-pause release Pol II in the midst of productive gene-body elongation. To test this, we first established whether Fiber-seq could identify elongating RNA Pol II footprints, which we expected to be a similar size to the PPP given their similar chromatin contacts and structure.[38,39] Across all genes and reads, we found putative elongating Pol II footprints (i.e., 40–60 bp footprints within a gene body) at a frequency of ~1.5 per 10 kb on average and markedly enriched at highly transcribed genes (Figure S1L). Furthermore, elongating Pol II footprints were depleted on gene bodies downstream of inaccessible promoters (Figure S2M) and in cells treated with triptolide (Figure S2N), suggesting that the density of elongating Pol II footprints corresponds to the level of gene body transcription.

Next, we set out to determine if reads with a shifted +1 nucleosome, but no PPP footprint, represented chromatin fibers enriched for productive gene body transcription. Compared to reads without a recently released PPP (i.e., an unshifted +1 nucleosome or a PPP footprint), reads with a recently released PPP (i.e., no PPP and a shifted +1 nucleosome) had a significantly higher density of elongating Pol II footprints (Figure 2K). In sum, these analyses conclude that Pol II pausing establishes a state of disrupted downstream nucleosomes, which upon pause release is stable and is associated with transcriptional elongation.

## Pausing sterically inhibits initiation

Having demonstrated how pausing directly alters the downstream nucleosome landscape at promoters across the genome, we next set out to address how pausing influences subsequent initiation. It is well established that pausing inhibits initiation, which is proposed to be explained, at least in part, through the steric interference between the PPP and PIC complexes.[17,40,41] This model has not been possible to test directly, though, as doing so requires the ability to simultaneously observe multiple individual paused and initiating polymerases at a single promoter. We inspected the frequency of PIC and PPP footprints within the same promoter region and found that PPPs and PICs rarely

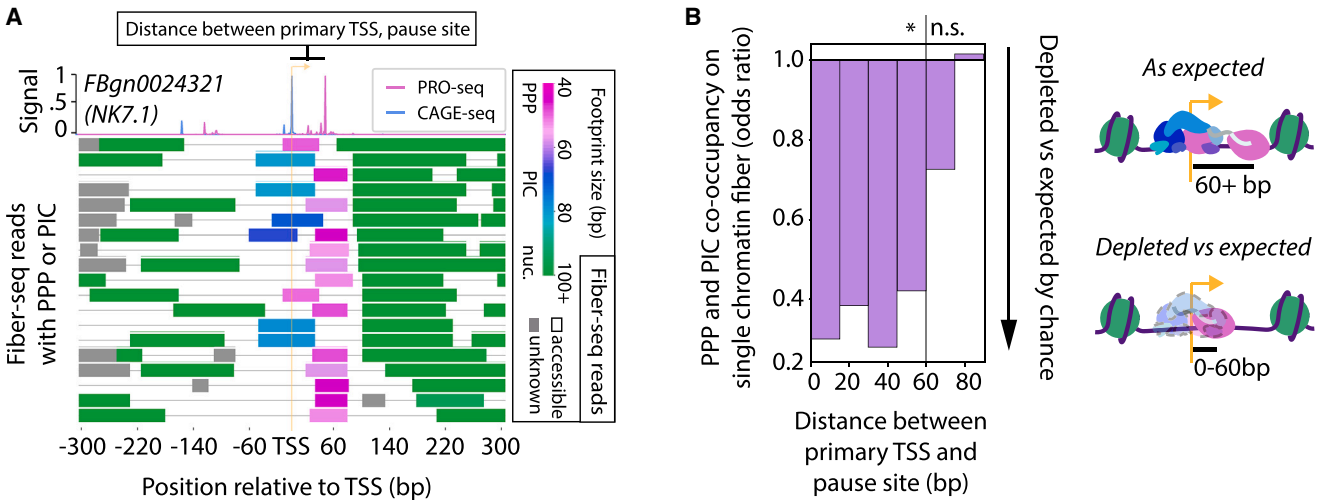



**Figure 3. Pausing sterically inhibits initiation**

(A) Sample Fiber-seq reads from a locus with simultaneous PPP and PIC footprints, with footprints colored by predicted identity (PPP, pink; PIC, blue; nucleosome, green; unknown, gray; black line, no footprint). Top: Corresponding PRO-seq (pink) and CAGE-seq (blue) signal.

(B) Bar plot, showing the Fisher's exact test odds ratio of PPP and PIC footprint co-occupancy on a single read. Genes are binned based on the distance between their primary PRO-seq and CAGE-seq peaks. Significance is indicated above (Fisher's exact test, n.s. signifies $p > 0.05$, * signifies $p$ values from $p < 0.05$ to $p < 10^{-25}$).

co-occurred on the same chromatin fiber (Figure 3A), with no evidence of a significant population of combined 100–140 bp footprints representing tightly abutted PPPs and PICs (Figure S1F). Notably, chromatin fibers containing simultaneous PPP and PIC footprints were observed as frequently as expected at loci with primary CAGE-seq and PRO-seq peaks separated by more than 60 bp, and especially more than 75 bp, a distance that would allow for unobstructed binding of both complexes (Figure 3B). In contrast, genes with primary CAGE-seq and PRO-seq peaks separated by fewer than 60 bp were significantly depleted in chromatin fibers with simultaneous PPP and PIC footprints (Figure 3B). This strong distance-dependent effect supports steric interference as the primary mechanism of pause-inhibited initiation.

## Coupled transcription initiation at neighboring genes

RNA expression studies using RNA-seq or microarray analysis have consistently shown that transcription at nearby pairs of genes is often correlated in *Drosophila* and other eukaryotes, with this effect predominantly seen at gene pairs located less than 1 kb from each other.[9,10,42] This correlation has largely been hypothesized to be a result of a shared, permissive chromatin environment between nearby genes. Additionally, imaging studies have demonstrated that Pol II localization within the nucleus is often clustered[12–16] and have captured synchronized, coupled transcription initiation at a select number of spatially co-localized genes.[43] It is unclear if these anecdotal observations of directly coupled transcription initiation represent a widespread, global phenomenon related to the correlated expression observed in bulk expression data. Given the lack of global single-molecule observations of chromatin accessibility and polymerase occupancy, it has not been possible to determine if correlated expression between nearby genes is simply explained by a shared permissive chromatin environment leading to a similar frequency of unsynchronized transcription initiation, as has been assumed, or if transcription initiation itself is coupled and synchronous.

We set out to directly test these two models using Fiber-seq. First, we tested whether spatially proximal transcribed gene promoters were preferentially accessible along the same chromatin fiber (Figure 4A). Indeed, consistent with prior work,[22] we identified that promoter accessibility was coupled along individual chromatin fibers, indicative of a permissive chromatin architecture for binding of the transcriptional machinery (Figure S4C). High promoter accessibility at a given gene did not directly translate into high levels of PPP or PIC footprints, though, with highly accessible promoters having on average three reads with Pol II footprints (Figure S4A). As such, while promoter accessibility is a prerequisite of transcription, it is insufficient to explain the gene's transcription level.

We then set out to determine if transcription initiation is also coupled between promoters of nearby genes when in a shared permissive chromatin environment. We evaluated the frequency of co-occupancy by PPP or PIC footprints at all pairs of promoters captured within a single Fiber-seq read. To isolate our analysis to transcription coupling, we only analyzed reads from chromatin fibers with both promoters accessible (Figure 4A). Given the low frequency of Pol II-related footprints at even highly accessible promoters, Pol II co-occupancy at a pair of promoters is expected to happen very infrequently in the absence of coupling (Figure S4B). In contrast, we found that pairs of nearby promoters (within 7.5 kb) were co-occupied by PPP and PIC footprints on the same fiber significantly more often than expected by chance. Furthermore, we observed a strong distance-dependent relationship to this effect, with the strongest co-occupancy seen at paired promoters located within 1 kb of

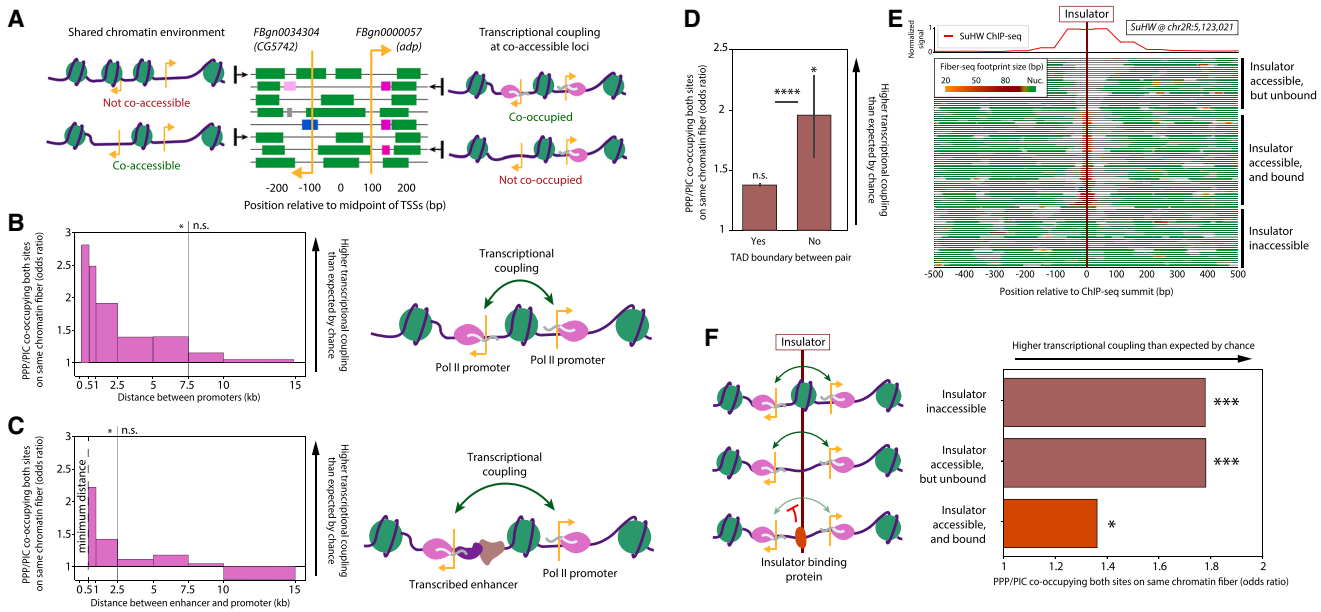



**Figure 4. Transcription initiation is coupled based on proximity**

(A) Sample Fiber-seq reads illustrating co-accessibility and PPP/PIC co-occupancy at two loci. Fiber-seq footprints are colored by predicted identity (PPP, pink; PIC, blue; nucleosome, green; unknown, gray; black line, no footprint).

(B) Barplot showing the Fisher's exact test odds ratio of PPP/PIC footprint co-occupancy on reads overlapping promoter pairs binned by distance. Significance is indicated above (Fisher's exact test, n.s. signifies $p > 0.05$, * signifies $p$ values from $p < 0.001$ to $p < 10^{-12}$).

(C) Barplot showing the Fisher's exact test odds ratio of PPP/PIC footprint co-occupancy on reads overlapping enhancer-promoter pairs binned by distance. Significance is indicated above the plot (Fisher's exact test, n.s. signifies $p > 0.05$, * signifies $p$ values ranging from $p < 0.001$ to $p < 10^{-9}$).

(D) Barplot depicting the Fisher's exact test odds ratio of PPP/PIC footprint co-occupancy on reads overlapping genes in different or the same topologically associating domain (TAD). Error bars correspond to the confidence interval calculated from 10,000× sampling iterations. Pairwise (two-sample t test, **** signifies $p$ value $<10^{-193}$) and individual (mean of sample Fisher's exact test results, n.s. signifies $p > 0.05$, * signifies $p < 0.05$) significance is indicated above.

(E) Bottom: Fiber-seq reads overlapping a SuHW binding site, with footprints colored by size (red, <90 bp footprints; green, >90 bp footprints; black line, no footprint). Top: Corresponding ChIP-seq signal.

(F) Barplot depicting the Fisher's exact test odds ratio of PPP/PIC footprint co-occupancy on reads overlapping genes separated by a ChIP-seq peak of an insulator binding protein. Reads are divided based on the illustrated state of the insulator, with significance indicated (Fisher's exact test, ** signifies $p < 10^{-5}$ and *** signifies $p < 10^{-45}$).

each other (Figure 4B). These global effects were recapitulated on a per-gene basis, which also demonstrated that transcriptional coupling can occur over longer distances, although at only select pairs of genes (Figures S4D and S4E). Together, these findings indicate that correlated expression of spatially proximal gene pairs is not exclusively a result of similar overall transcription levels mediated by a shared chromatin environment, but is driven in part by the synchronized, coupled occupancy of RNA polymerase complexes at both promoters.

## Coupled transcription initiation between enhancers and nearby genes

Many eukaryotic enhancers are transcribed by Pol II to form enhancer RNAs (eRNAs), the production of which is strongly associated with enhanced expression of nearby genes.[44–46] Given the coupling of the transcriptional machinery between nearby protein-coding gene promoters, we hypothesized that transcription initiation at transcribed enhancer-promoter pairs may be similarly coupled. Using *Drosophila* S2 cell STARR-seq and START-seq data, which map enhancer regions and nascent transcription start sites, respectively,[44,47,48] we identified ~4,000 enhancers that produce eRNAs. Overall, eRNA PPP and PIC footprints mirrored those at coding promoters, consistent with their transcription by Pol II (Figures S3B and S4F–S4I). Like paired coding promoters, we found that chromatin fibers overlapping eRNA-producing enhancers and a nearby promoter were preferentially co-occupied by initiating polymerase footprints significantly more frequently than expected based on the overall frequency of the footprints. Even more so than in the case of coding promoters, we found a strong distance-dependent relationship in this effect (Figure 4C). Together, these findings suggest that the proximity of an eRNA-producing enhancer to a promoter enables the coupling of RNA polymerase complex occupancy at both simultaneously, providing a potential mechanistic basis for how eRNAs potentiate gene expression.

## Insulator protein occupancy at TAD boundaries inhibits transcriptional coupling

While a large proportion of proximal pairs of promoters showed evidence of transcriptional coupling, we found a subset of these pairs that did not exhibit evidence of transcriptional coupling. Topologically associating domain (TAD) boundaries are associated with breaks in correlated gene expression patterns along the genome based on bulk short-read sequencing.[11,47–50] This

process has been hypothesized to result from distinct chromatin loops that form adjacent to TAD boundaries, and we sought to determine whether this correlative finding was mediated by a disruption in transcriptional coupling between promoters across a TAD boundary.

We divided pairs of genes found within 2.5 kb of each other from the previous coupling analysis based on if both were contained in the same TAD, or if they were located in separate TADs based on Hi-C data.[50] Controlling for differences in the distribution of inter-promoter distances in the same-TAD and different-TAD groups, we found that chromatin fibers overlapping pairs of genes split by TAD boundaries had significantly less evidence of transcriptional coupling (Figure 4D). However, although diminished, transcriptional coupling still occurred across TAD boundaries, indicating that the effect of TAD boundaries may be dynamic within a cell population.

TAD boundaries in *Drosophila* are generally bound by insulator proteins including CP190, CTCF, SuHW, and Mod, and insulator protein occupancy and TADs themselves are dynamic, showing substantial variability cell to cell.[50–52] We hypothesized that if the insulator occupancy defines whether a TAD is formed on a particular fiber, then those fibers should simultaneously show reduced transcriptional coupling across the associated TAD boundary. Consistent with this expectation, we observed that pairs of genes separated by bulk ChIP-seq peaks for insulator proteins showed significantly lower than expected transcriptional coupling (Figure S4B).[53–55] To determine whether transcriptional coupling is directly modulated by the occupancy of these factors, we first demonstrated that FiberHMM can readily identify single-molecule occupancy patterns of the insulator proteins CP190, CTCF, SuHW, and Mod, with these single-molecule occupancy patterns mirroring bulk measures of protein occupancy (Figures 4E and S4J–S4M). By co-evaluating the single-molecule occupancy patterns of transcriptional machinery and insulator proteins, we found that only when there is a protein directly bound at an intervening insulator element do we see a reduction in transcriptional coupling between promoter pairs (Figure 4F). Together, these results demonstrate that single-molecule transcriptional coupling is disrupted by TAD boundaries and insulator elements and that this disruption is mediated by the direct protein occupancy of intervening factors bound at these insulator elements.

## Identification of RNA polymerase III-associated footprints via FiberHMM

We next sought to determine whether this coupling was a feature of other RNA polymerases, specifically RNA polymerase III (Pol III). Pol III is primarily responsible for the transcription of tRNA and 5S rRNA genes, which are often highly expressed and spatially clustered together along the genome.[56] We first sought to determine whether FiberHMM could identify footprints at tRNA genes associated with Pol III transcription, which is mediated by TFIIIB and TFIIIC occupancy at specific sequence elements upstream of and within the tRNA gene.[57,58] Consistent with our expectations, we identified several distinct populations of footprints enriched at tRNA genes corresponding to the known binding sites of TFIIIB and TFIIIC (Figure 5A). Additionally, the size and position of these putative Pol III transcription-associated footprints was consistent across all tRNA genes (Figure 5B).

To validate whether these Pol III transcription-associated footprints were indicative of tRNA expression, we needed to test if their abundance corresponded with the expected expression of those tRNAs. However, determining the expression level of individual tRNA genes using transcript-based sequencing is difficult, as most tRNA genes in *Drosophila* have multiple copies that are identical in sequence and cannot be uniquely mapped to the genome. Furthermore, tRNA transcripts form strong secondary structures and undergo numerous modifications that impair standard reverse transcription-based transcript profiling methods.[59–62] Notably, we observed that individual copies of the same tRNA gene can have drastically different levels of these Pol III transcription-associated footprints, suggesting a high degree of variability in transcription activity between tRNA genes that produce sequence-identical transcripts that would otherwise be undistinguishable by RNA-seq (Figures 5C and S5A–S5C).

To determine whether this variability in footprint occupancy reflected the transcriptional output of each tRNA gene, we calculated a "transcription frequency" for each tRNA gene, corresponding to the fraction of reads mapping to that gene occupied by Pol III transcription-associated footprints. We found that the sum of these scores for all tRNA genes that contribute to the same codon significantly correlated with frequency of that codon within the *Drosophila* genome.[63] This correlation validates that our single-molecule measurements of RNA Pol III occupancy at individual tRNA genes accurately captures the overall demand for that tRNA in *Drosophila* (Figure S5D).

## Clustered Pol III genes exhibit coupled transcription activity

As tRNA genes are often spatially clustered along the genome, we next tested whether neighboring Pol III tRNA promoters showed transcriptional coupling similar to that observed between Pol II promoters, and if coupling explained the variance in expression we observed between sequence-identical tRNA gene copies. Overall, we observed that chromatin fibers overlapping neighboring tRNA promoters were preferentially co-bound by Pol III transcription-associated footprints significantly more frequently than expected by chance. In addition, we observed a strong distance-dependent relationship in this effect, with paired tRNA promoters located within 2.5 kb of each other showing the strongest coupling of footprint occupancy (Figure 5D). Notably, tRNA genes located within 2.5 kb of each other also have significantly higher overall single-molecule Pol III transcription-associated footprint occupancy, consistent with a model in which tRNA gene clustering enables higher overall expression via leveraging the coupled occupancy of Pol III machinery (Figure 5E).

We next wanted to determine whether the Pol III coupling observed at tRNA genes was unique to tRNA genes or a general feature of Pol III-mediated transcription initiation. Pol III is also responsible for the transcription of 5S rRNA genes, which are 120 bp genes spatially clustered together as an array of ~100 tandem copies of the same ~375 bp repeat.[64] Of note, this

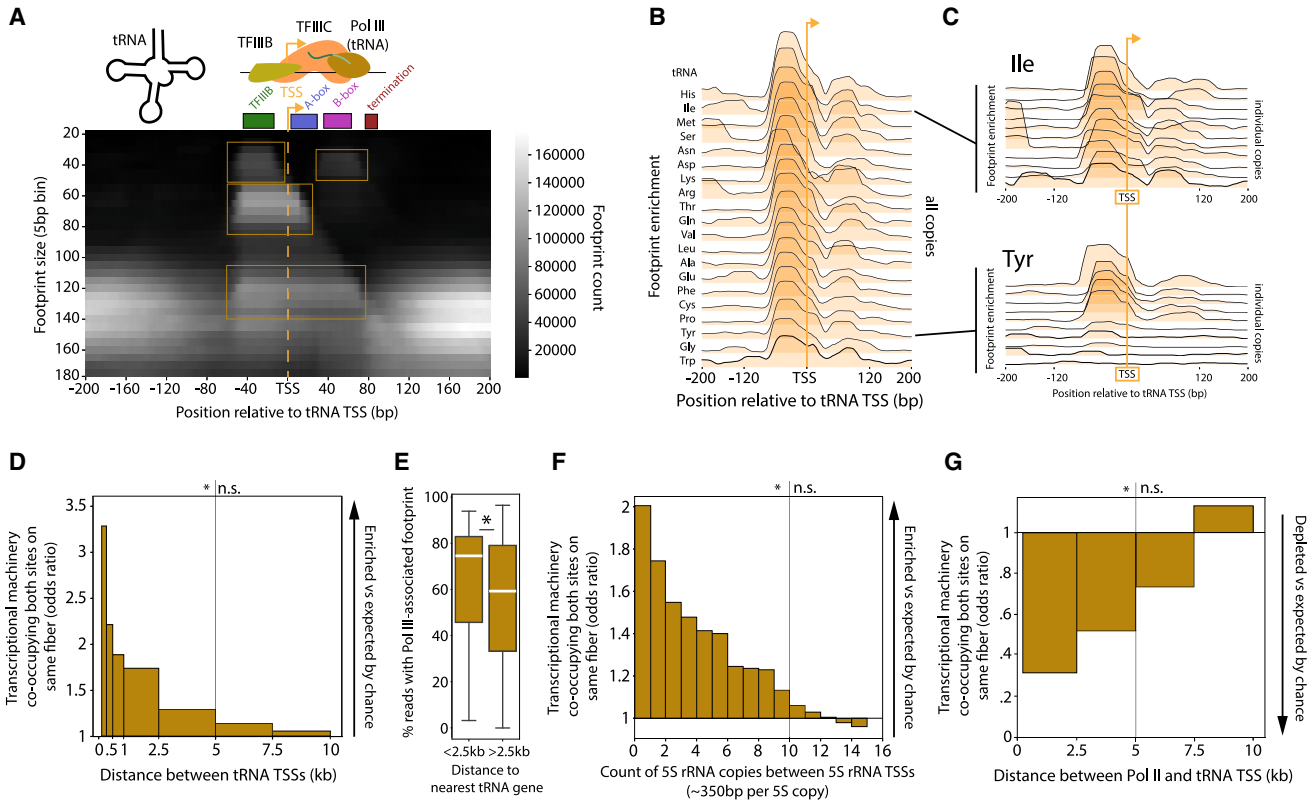



**Figure 5. Coupling of transcription activity between nearby Pol III genes**

(A) Heatmap depicting the enrichment of differently sized Fiber-seq footprints relative to tRNA TSSs. The expected position of TFIIIB, TFIIIC, terminating Pol III, and the A- and B-box are indicated above the plot. Boxes indicate enriched footprint populations.

(B) Plots showing the percent enrichment of 30–90 bp footprints relative to the TSSs of families of tRNA genes.

(C) Percent enrichment of 30–90 bp footprints relative to the TSS of individual isoleucine (top) or tyrosine (bottom) tRNA genes.

(D) Barplot depicting the Fisher's exact test odds ratio of tRNA transcription-associated footprint co-occupancy between pairs of tRNAs binned by distance, with significance indicated above (Fisher's exact test, n.s. signifies $p > 0.05$, ** signifies $p$ value $<10^{-4}$, **** signifies $p$ value $<10^{-85}$).

(E) Boxplot of the distribution of transcription frequency scores for tRNA genes separated from another tRNA gene, binned by distance. Pairwise significance is indicated (two-sample t test, * signifies $p$ value <0.05).

(F) Barplot depicting the Fisher's exact test odds ratio of 5S rRNA transcription-associated footprint co-occupancy between pairs of 5S rRNAs binned by count of intermediate 5S rRNA copies. Significance is indicated above (Fisher's exact test, n.s. signifies $p > 0.05$, ** signifies $p$ value $<10^{-4}$, **** signifies $p$ value $<10^{-251}$).

(G) Bar plot depicting the Fisher's exact test odds ratio of tRNA transcription-associated footprint co-occupancy with PPP/PIC footprints between pairs of tRNA and Pol II genes binned by distance. Significance is indicated above (Fisher's exact test, n.s. signifies $p > 0.05$, * signifies $p$ value <0.05, ** signifies $p$ value <0.01).

tandem duplication is highly repetitive, making reference-based assessments of co-occupancy using short read sequencing approaches largely impossible. At 5S rRNA genes we found a pattern of footprints similar to that at tRNA genes, corresponding to the selective occupancy at binding sites for TFIIIA, TFIIIB, and TFIIIC (Figures S5E and S5F). As individual Fiber-seq reads on average span ~30 repeats of the 5S gene, we next determined whether Pol III 5S rRNA promoters showed transcriptional coupling similar to that between Pol III tRNA promoters and between Pol II promoters. Overall, we observed that chromatin fibers overlapping nearby 5S rRNA promoters were preferentially co-bound by Pol III transcription-associated footprints significantly more frequently than expected. In addition, we observed a strong distance-dependent relationship in this effect at a range similar to that of tRNA and Pol II genes (Figure 5E). These results demonstrate that clustered, coupled occupancy by Pol III is a consistent effect across the major Pol III transcribed genes and further indicate that the condensed spatial clustering of 5S rRNA genes likely potentiates this effect.

## Transcription initiation is inversely coupled between Pol II and Pol III genes

tRNA genes are often positioned adjacent to Pol II genes, and the presence of these tRNA genes has been shown to disrupt the transcriptional output of these neighboring Pol II genes.[65,66] However, the exact mechanism driving the feedback between Pol II- and Pol III-transcribed genes occurs is unknown. Current models suggest that tRNA genes may act as boundary elements in a TFIIIC-dependent manner.[66] To determine if transcription initiation at tRNA genes influences Pol II transcription initiation at neighboring genes, we evaluated transcription initiation footprints along chromatin fibers overlapping neighboring tRNA and Pol II promoters. Overall, we observed that PPP and PIC footprints were significantly depleted along chromatin fibers

that contained Pol III transcription-associated footprints, and vice versa. This effect demonstrated a strong distance-dependent relationship, with inversely coupled Pol II and Pol III transcription initiation at pairs within 5 kb of each other (Figure 5G). Furthermore, as Pol III transcription-associated footprints at tRNA genes are significantly more prevalent than PPP and PIC footprints at Pol II genes, the dominant effect is Pol III-mediated silencing of Pol II transcription. Overall, these results demonstrate that tRNA-mediated silencing of Pol II-transcribed genes is occurring at the level of decoupled transcription initiation.

## DISCUSSION

Overall, we present Fiber-seq combined with FiberHMM, a genomic approach that allows us to delineate competition, coordination, and coupling of transcription initiation, pausing, and the surrounding nucleosome landscape along individual chromatin fibers with single-molecule and near-single-nucleotide precision. Specifically, we demonstrate a mechanism for pause-inhibited initiation as well as pause-driven changes to the positioning and stability of nucleosome arrays that compete for the same genomic real estate. In addition, we demonstrate pervasive coupling of transcription initiation at nearby pairs of promoters and enhancer-promoter pairs, and decoupling of transcription between nearby Pol II- and Pol III-transcribed genes. Together, these findings illustrate transcription initiation as not merely a passenger occupying templated chromatin environments, but as an active participant in the regulation of gene expression and establishment of chromatin architectures within these environments.

We find that Pol II pausing causes a broad destabilization of the downstream nucleosome landscape up to three nucleosomes away from the TSS. Whether this destabilization involves chromatin remodelers, local supercoiling due to Pol II transcription, or other modalities remains unclear. Upon pause release, fibers with destabilized downstream nucleosomes are actively transcribed at a higher frequency, suggesting a connection between pausing, altered downstream nucleosome architecture, and elongation activity. Further, the stability of these pause-associated changes to promoter structure likely permit a more efficient reestablishment of both initiation and pausing upon pause release, via maintaining both an accessible TSS and pause site, which would otherwise be occluded by nucleosomes.[8] Consequently, this directional, pause-specific, and expression-independent short-term disruption may serve to prime highly paused genes for rapid, regulated transcription, consistent with the patterns of expression seen at developmental and heat shock genes.[1,67]

We also demonstrate the pervasive coupling of transcription initiation between spatially proximal Pol II-transcribed promoters and enhancers. This coupling appears to be an effect shared across RNA polymerases and promoters and is most frequently seen at elements located within 7.5 kb of each other. Spatial proximity along the genome as a driver of synchronized, coupled transcription suggests a functional basis for the overall linear and spatial organization of genes, transcribed enhancers, and Pol III genes along the *Drosophila* genome. Importantly, our findings disentangle correlated transcription from true single-molecule transcriptional coupling, demonstrating a molecular basis for RNA expression studies that have consistently shown that transcription at nearby pairs of genes is often correlated in *Drosophila* and other eukaryotes.

How this synchronized transcriptional activity at nearby promoters is mediated remains unknown and an important priority for future studies. Given the strong distance dependence of coupling, there seems to be a significant level of coupling that occurs based solely on the linear arrangement of genes. However, proximity is not sufficient to explain the strong coupling that we observe with synchronous Pol II activity. Microscopy-based studies have observed similar coupling between promoter pairs or enhancer-promoter pairs upon very close contact. Pairs of promoters within the same TAD are similarly likely to come in close proximity with a distance-dependent frequency, which would explain our observations.[12–16,43,68,69] Alternatively, or in addition, supercoiling has been implicated in the coupling or inverse coupling of genes, an effect that is broadly consistent with our observations.[70] Future work will need to characterize these mechanisms on a single-fiber level, including the action of distal tethering elements or, as we observed, the occupation of insulator elements.[43,49]

Together, these findings delineate the complex relationship between polymerases and the rest of the chromatin-bound proteome along individual chromatin fibers within the cell. In addition, this work highlights footprinted Fiber-seq as a central tool for unifying prior microscopy-, structural-, and genomic-based assays of RNA polymerase—establishing a single-molecule framework of competition and coordination of polymerase complexes with the surrounding nuclear landscape along individual chromatin fibers within each cell.

### Limitations of the study

Identifying single-molecule protein occupancy footprints is inherently limited by: (1) the density of "observable" bases within the occupied region and (2) the confidence of assigning a footprint to its cognate protein. For Fiber-seq using PacBio sequencing, "observable" bases are limited to A/T base pairs, and the methyltransferase reaction may not be saturated in certain sequence contexts. The resulting ambiguity in the precise ends of each footprint likely contributes to the ~20 bp range of observed sizes for PPP and PIC footprints in our study. Our HMM-based approach addresses this ambiguity with experimentally derived emission probabilities and trained transition probabilities but is generally limited by the first-order Markov assumption—that each subsequent state only depends on the previous state. It is unclear whether more complex machine learning approaches that can take into account higher-order local and global contexts would further expand upon these successes. Of note, this problem is more challenging for Fiber-seq using simplex Oxford Nanopore sequencing, where "observable" bases are limited to only As on a single strand, or GpC-methyltransferase approaches, where "observable" bases are limited to the much more sparsely located GpC dinucleotides.

Additionally, assigning a footprint to its cognate protein is challenging, as the only available information is the footprint size, underlying genomic sequence, and prior bulk measurements of protein occupancy. Although the inherent ambiguity

of footprint identity will result in some footprints being assigned to the wrong cognate protein, overall we have demonstrated that this orthogonal information can robustly classify single-molecule footprint identity.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Hia5 expression and purification
  - Fiber-seq
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Defining CAGE-seq, START-seq and PRO-seq peaks
  - Calculating pause index, expression levels, and GAF occupancy
  - Fiber-seq processing
  - FiberHMM
  - Identifying size ranges for nucleosome footprints
  - Identifying size ranges for PPP and PIC footprints
  - Defining PPP and PIC footprints
  - Identifying size ranges and defining transcription factor footprints
  - Identifying reads with an accessible promoter or insulator
  - Validation of footprint sizes from existing structures
  - Sampling reads based on PPP and PIC footprints
  - Sampling reads based on distance
  - Segmenting downstream and upstream nucleosomes
  - +1 and −1 nucleosome size and distance on a per-gene basis
  - +1 nucleosome size and distance on a per-read basis
  - Defining elongating Pol II footprints
  - Pause-inhibited initiation quantification
  - Identifying transcribed enhancers
  - Defining tRNA Pol III transcription-associated footprints
  - Defining 5S rRNA Pol III transcription-associated footprints
  - Correlation of tRNA expression score to codon frequency
  - Global coupling analysis
  - Per-locus coupling analysis

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2024.08.013.

## ACKNOWLEDGMENTS

We thank James Warner, Ines Patop, and additional members of the Churchman lab for helpful discussions, assistance, and critical reading of the manuscript; Sean Eddy for discussions about the HMM; K. Munson; and the University of Washington PacBio Sequencing Services and University of Maryland School of Medicine Genomics Resource Center for PacBio sequencing. This work was supported by the NIH (R01-HG007173 to L.S.C., DP5-OD029630 to A.B.S.), and the Brotman Baty Institute for Precision Medicine. A.B.S. holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund and is a Pew Biomedical Scholar.

## AUTHOR CONTRIBUTIONS

Conceptualization, T.W.T., L.S.C., and A.B.S.; methodology, T.W.T. (lead), R.S.I., A.B.S., and L.S.C.; investigation, T.W.T. (lead), R.S.I., and J.R.; formal analysis, T.W.T. (lead), D.D.; writing—original draft, T.W.T.; writing—review & editing, T.W.T., R.S.I., A.B.S., and L.S.C.; funding acquisition, A.B.S. and L.S.C.; supervision, A.B.S. and L.S.C.

## DECLARATION OF INTERESTS

A.B.S. is a co-inventor on a patent relating to the Fiber-seq method (US17/995,058).

Received: February 6, 2024
Revised: July 2, 2024
Accepted: August 8, 2024
Published: August 26, 2024

## REFERENCES

1. Core, L., and Adelman, K. (2019). Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. Genes Dev. *33*, 960–982. https://doi.org/10.1101/gad.325142.119.
2. Mayer, A., Landry, H.M., and Churchman, L.S. (2017). Pause & go: from the discovery of RNA polymerase pausing to its functional implications. Curr. Opin. Cell Biol. *46*, 72–80. https://doi.org/10.1016/j.ceb.2017.03.002.
3. Petesch, S.J., and Lis, J.T. (2012). Overcoming the nucleosome barrier during transcript elongation. Trends Genet. *28*, 285–294. https://doi.org/10.1016/j.tig.2012.02.005.
4. Farnung, L., Ochmann, M., Garg, G., Vos, S.M., and Cramer, P. (2022). Structure of a backtracked hexasomal intermediate of nucleosome transcription. Mol. Cell *82*, 3126–3134.e7. https://doi.org/10.1016/j.molcel.2022.06.027.
5. Wang, H., Schilbach, S., Ninov, M., Urlaub, H., and Cramer, P. (2023). Structures of transcription preinitiation complex engaged with the +1 nucleosome. Nat. Struct. Mol. Biol. *30*, 226–232. https://doi.org/10.1038/s41594-022-00865-w.
6. Chen, Z., Gabizon, R., Brown, A.I., Lee, A., Song, A., Díaz-Celis, C., Kaplan, C.D., Koslover, E.F., Yao, T., and Bustamante, C. (2019). High-resolution and high-accuracy topographic and transcriptional maps of the nucleosome barrier. Elife *8*, e48281. https://doi.org/10.7554/eLife.48281.
7. Ramachandran, S., Ahmad, K., and Henikoff, S. (2017). Transcription and Remodeling Produce Asymmetrically Unwrapped Nucleosomal Intermediates. Mol. Cell *68*, 1038–1053.e4. https://doi.org/10.1016/j.molcel.2017.11.015.
8. Gilchrist, D.A., Dos Santos, G., Fargo, D.C., Xie, B., Gao, Y., Li, L., and Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. Cell *143*, 540–551. https://doi.org/10.1016/j.cell.2010.10.004.
9. Yang, L., and Yu, J. (2009). A comparative analysis of divergently-paired genes (DPGs) among Drosophila and vertebrate genomes. BMC Evol. Biol. *9*, 55. https://doi.org/10.1186/1471-2148-9-55.
10. Herr, D.R., and Harris, G.L. (2004). Close head-to-head juxtaposition of genes favors their coordinate regulation in Drosophila melanogaster. FEBS Lett. *572*, 147–153. https://doi.org/10.1016/j.febslet.2004.07.026.
11. Small, S., and Arnosti, D.N. (2020). Transcriptional enhancers in Drosophila. Genetics *216*, 1–26. https://doi.org/10.1534/genetics.120.301370.
12. Sutherland, H., and Bickmore, W.A. (2009). Transcription factories: gene expression in unions? Nat. Rev. Genet. *10*, 457–466. https://doi.org/10.1038/nrg2592.
13. Kimura, H., and Sato, Y. (2022). Imaging transcription elongation dynamics by new technologies unveils the organization of initiation and elongation in transcription factories. Curr. Opin. Cell Biol. *74*, 71–79. https://doi.org/10.1016/j.ceb.2022.01.002.

14. Patange, S., Ball, D.A., Karpova, T.S., and Larson, D.R. (2021). Towards a "Spot On" Understanding of Transcription in the Nucleus. J. Mol. Biol. *433*, 167016. https://doi.org/10.1016/j.jmb.2021.167016.

15. Cisse, I.I., Izeddin, I., Causse, S.Z., Boudarene, L., Senecal, A., Muresan, L., Dugast-Darzacq, C., Hajj, B., Dahan, M., and Darzacq, X. (2013). Real-time dynamics of RNA polymerase II clustering in live human cells. Science *341*, 664–667. https://doi.org/10.1126/science.1239053.

16. Cho, W.-K., Spille, J.-H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science *361*, 412–415. https://doi.org/10.1126/science.aar4199.

17. Gressel, S., Schwalb, B., and Cramer, P. (2019). The pause-initiation limit restricts transcription activation in human cells. Nat. Commun. *10*, 3603. https://doi.org/10.1038/s41467-019-11536-8.

18. Farnung, L., Vos, S.M., and Cramer, P. (2018). Structure of transcribing RNA polymerase II-nucleosome complex. Nat. Commun. *9*, 5432. https://doi.org/10.1038/s41467-018-07870-y.

19. Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M., and Bustamante, C. (2009). Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. Science *325*, 626–628. https://doi.org/10.1126/science.1172926.

20. He, Y., Fang, J., Taatjes, D.J., and Nogales, E. (2013). Structural visualization of key steps in human transcription initiation. Nature *495*, 481–486. https://doi.org/10.1038/nature11991.

21. Krebs, A.R., Imanci, D., Hoerner, L., Gaidatzis, D., Burger, L., and Schübeler, D. (2017). Genome-wide single-molecule footprinting reveals high RNA polymerase II turnover at paused promoters. Mol. Cell *67*, 411–422.e4. https://doi.org/10.1016/j.molcel.2017.06.027.

22. Stergachis, A.B., Debo, B.M., Haugen, E., Churchman, L.S., and Stamatoyannopoulos, J.A. (2020). Single-molecule regulatory architectures captured by chromatin fiber sequencing. Science *368*, 1449–1454. https://doi.org/10.1126/science.aaz1646.

23. Abdulhay, N.J., McNally, C.P., Hsieh, L.J., Kasinathan, S., Keith, A., Estes, L.S., Karimzadeh, M., Underwood, J.G., Goodarzi, H., Narlikar, G.J., and Ramani, V. (2020). Massively multiplex single-molecule oligonucleosome footprinting. Elife *9*, e59404. https://doi.org/10.7554/eLife.59404.

24. Lee, I., Razaghi, R., Gilpatrick, T., Molnar, M., Gershman, A., Sadowski, N., Sedlazeck, F.J., Hansen, K.D., Simpson, J.T., and Timp, W. (2020). Simultaneous profiling of chromatin accessibility and methylation on human cell lines with nanopore sequencing. Nat. Methods *17*, 1191–1199. https://doi.org/10.1038/s41592-020-01000-7.

25. Shipony, Z., Marinov, G.K., Swaffer, M.P., Sinnott-Armstrong, N.A., Skotheim, J.M., Kundaje, A., and Greenleaf, W.J. (2020). Long-range single-molecule mapping of chromatin accessibility in eukaryotes. Nat. Methods *17*, 319–327. https://doi.org/10.1038/s41592-019-0730-2.

26. Jha, A., Bohaczuk, S.C., Mao, Y., Ranchalis, J., Mallory, B.J., Min, A.T., Hamm, M.O., Swanson, E., Dubocanin, D., Finkbeiner, C., et al. (2023). DNA-m6A calling and integrated long-read epigenetic and genetic analysis with fibertools. Preprint at bioRxiv. https://doi.org/10.1101/2023.04.20.537673.

27. Vos, S.M., Farnung, L., Urlaub, H., and Cramer, P. (2018). Structure of paused transcription complex Pol II-DSIF-NELF. Nature *560*, 601–606. https://doi.org/10.1038/s41586-018-0442-2.

28. Rengachari, S., Schilbach, S., Aibara, S., Dienemann, C., and Cramer, P. (2021). Structure of the human Mediator-RNA polymerase II pre-initiation complex. Nature *594*, 129–133. https://doi.org/10.1038/s41586-021-03555-7.

29. Schilbach, S., Aibara, S., Dienemann, C., Grabbe, F., and Cramer, P. (2021). Structure of RNA polymerase II pre-initiation complex at 2.9 Å defines initial DNA opening. Cell *184*, 4064–4072.e28. https://doi.org/10.1016/j.cell.2021.05.012.

30. Kwak, H., Fuda, N.J., Core, L.J., and Lis, J.T. (2013). Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. Science *339*, 950–953. https://doi.org/10.1126/science.1229386.

31. Hitz, B.C., Jin-Wook, L., Jolanki, O., Kagda, M.S., Graham, K., Sud, P., Gabdank, I., Strattan, J.S., Sloan, C.A., Dreszer, T., et al. (2023). The ENCODE Uniform Analysis Pipelines. Preprint at bioRxiv. https://doi.org/10.1101/2023.04.04.535623.

32. Luo, Y., Hitz, B.C., Gabdank, I., Hilton, J.A., Kagda, M.S., Lam, B., Myers, Z., Sud, P., Jou, J., Lin, K., et al. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. Nucleic Acids Res. *48*, D882–D889. https://doi.org/10.1093/nar/gkz1062.

33. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74. https://doi.org/10.1038/nature11247.

34. Titov, D.V., Gilman, B., He, Q.-L., Bhat, S., Low, W.-K., Dang, Y., Smeaton, M., Demain, A.L., Miller, P.S., Kugel, J.F., et al. (2011). XPB, a subunit of TFIIH, is a target of the natural product triptolide. Nat. Chem. Biol. *7*, 182–188. https://doi.org/10.1038/nchembio.522.

35. Wang, Y., Lu, J.-J., He, L., and Yu, Q. (2011). Triptolide (TPL) inhibits global transcription by inducing proteasome-dependent degradation of RNA polymerase II (Pol II). PLoS One *6*, e23993. https://doi.org/10.1371/journal.pone.0023993.

36. Chereji, R.V., Bryson, T.D., and Henikoff, S. (2019). Quantitative MNase-seq accurately maps nucleosome occupancy levels. Genome Biol. *20*, 198. https://doi.org/10.1186/s13059-019-1815-z.

37. Fuda, N.J., Guertin, M.J., Sharma, S., Danko, C.G., Martins, A.L., Siepel, A., and Lis, J.T. (2015). GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. PLoS Genet. *11*, e1005108. https://doi.org/10.1371/journal.pgen.1005108.

38. Vos, S.M., Farnung, L., Boehning, M., Wigge, C., Linden, A., Urlaub, H., and Cramer, P. (2018). Structure of activated transcription complex Pol II-DSIF-PAF-SPT6. Nature *560*, 607–612. https://doi.org/10.1038/s41586-018-0440-4.

39. Vos, S.M., Farnung, L., Linden, A., Urlaub, H., and Cramer, P. (2020). Structure of complete Pol II-DSIF-PAF-SPT6 transcription complex reveals RTF1 allosteric activation. Nat. Struct. Mol. Biol. *27*, 668–677. https://doi.org/10.1038/s41594-020-0437-1.

40. Shao, W., and Zeitlinger, J. (2017). Paused RNA polymerase II inhibits new transcriptional initiation. Nat. Genet. *49*, 1045–1051. https://doi.org/10.1038/ng.3867.

41. Gressel, S., Schwalb, B., Decker, T.M., Qin, W., Leonhardt, H., Eick, D., and Cramer, P. (2017). CDK9-dependent RNA polymerase II pausing controls transcription initiation. Elife *6*, e29736. https://doi.org/10.7554/eLife.29736.

42. Lawrence, J.G. (2002). Shared strategies in gene organization among prokaryotes and eukaryotes. Cell *110*, 407–413. https://doi.org/10.1016/s0092-8674(02)00900-5.

43. Levo, M., Raimundo, J., Bing, X.Y., Sisco, Z., Batut, P.J., Ryabichko, S., Gregor, T., and Levine, M.S. (2022). Transcriptional coupling of distant regulatory genes in living embryos. Nature *605*, 754–760. https://doi.org/10.1038/s41586-022-04680-7.

44. Henriques, T., Scruggs, B.S., Inouye, M.O., Muse, G.W., Williams, L.H., Burkholder, A.B., Lavender, C.A., Fargo, D.C., and Adelman, K. (2018). Widespread transcriptional pausing and elongation control at enhancers. Genes Dev. *32*, 26–41. https://doi.org/10.1101/gad.309351.117.

45. Sartorelli, V., and Lauberth, S.M. (2020). Enhancer RNAs are an important regulatory layer of the epigenome. Nat. Struct. Mol. Biol. *27*, 521–528. https://doi.org/10.1038/s41594-020-0446-0.

46. Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I.E., Males, M., Viales, R.R., and Furlong, E.E.M. (2018). The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. Genes Dev. *32*, 42–57. https://doi.org/10.1101/gad.308619.117.

47. Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, Ł.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science *339*, 1074–1077. https://doi.org/10.1126/science.1232542.

48. Zabidi, M.A., Arnold, C.D., Schernhuber, K., Pagani, M., Rath, M., Frank, O., and Stark, A. (2015). Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. Nature *518*, 556–559. https://doi.org/10.1038/nature13994.

49. Batut, P.J., Bing, X.Y., Sisco, Z., Raimundo, J., Levo, M., and Levine, M.S. (2022). Genome organization controls transcriptional dynamics during development. Science *375*, 566–570. https://doi.org/10.1126/science.abi7178.

50. Ramírez, F., Lingg, T., Toscano, S., Lam, K.C., Georgiev, P., Chung, H.-R., Lajoie, B.R., de Wit, E., Zhan, Y., de Laat, W., et al. (2015). High-Affinity Sites Form an Interaction Network to Facilitate Spreading of the MSL Complex across the X Chromosome in Drosophila. Mol. Cell *60*, 146–162. https://doi.org/10.1016/j.molcel.2015.08.024.

51. Cattoni, D.I., Cardozo Gizzi, A.M., Georgieva, M., Di Stefano, M., Valeri, A., Chamousset, D., Houbron, C., Déjardin, S., Fiche, J.-B., González, I., et al. (2017). Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. Nat. Commun. *8*, 1753. https://doi.org/10.1038/s41467-017-01962-x.

52. Luppino, J.M., and Joyce, E.F. (2020). Single cell analysis pushes the boundaries of TAD formation and function. Curr. Opin. Genet. Dev. *61*, 25–31. https://doi.org/10.1016/j.gde.2020.03.005.

53. Yang, J., Ramos, E., and Corces, V.G. (2012). The BEAF-32 insulator coordinates genome organization and function during the evolution of Drosophila species. Genome Res. *22*, 2199–2207. https://doi.org/10.1101/gr.142125.112.

54. Ong, C.-T., Van Bortle, K., Ramos, E., and Corces, V.G. (2013). Poly(ADP-ribosyl)ation regulates insulator function and intrachromosomal interactions in Drosophila. Cell *155*, 148–159. https://doi.org/10.1016/j.cell.2013.08.052.

55. Kaushal, A., Mohana, G., Dorier, J., Özdemir, I., Omer, A., Cousin, P., Semenova, A., Taschner, M., Dergai, O., Marzetta, F., et al. (2021). CTCF loss has limited effects on global genome architecture in Drosophila despite critical regulatory functions. Nat. Commun. *12*, 1011. https://doi.org/10.1038/s41467-021-21366-2.

56. Turowski, T.W., and Tollervey, D. (2016). Transcription by RNA polymerase III: insights into mechanism and regulation. Biochem. Soc. Trans. *44*, 1367–1375. https://doi.org/10.1042/BST20160062.

57. Graczyk, D., Cieśla, M., and Boguta, M. (2018). Regulation of tRNA synthesis by the general transcription factors of RNA polymerase III - TFIIIB and TFIIIC, and by the MAF1 protein. Biochim. Biophys. Acta. Gene Regul. Mech. *1861*, 320–329. https://doi.org/10.1016/j.bbagrm.2018.01.011.

58. Male, G., von Appen, A., Glatt, S., Taylor, N.M.I., Cristovao, M., Groetsch, H., Beck, M., and Müller, C.W. (2015). Architecture of TFIIIC and its role in RNA polymerase III pre-initiation complex assembly. Nat. Commun. *6*, 7387. https://doi.org/10.1038/ncomms8387.

59. Geslain, R., and Pan, T. (2010). Functional Analysis of Human tRNA Isodecoders. J. Mol. Biol. *396*, 821–831. https://doi.org/10.1016/j.jmb.2009.12.018.

60. Marygold, S.J., Chan, P.P., and Lowe, T.M. (2022). Systematic identification of tRNA genes in Drosophila melanogaster. MicroPubl. Biol. *2022*. https://doi.org/10.17912/micropub.biology.000560.

61. Behrens, A., Rodschinka, G., and Nedialkova, D.D. (2021). High-resolution quantitative profiling of tRNA abundance and modification status in eukaryotes by mim-tRNAseq. Mol. Cell *81*, 1802–1815.e7. https://doi.org/10.1016/j.molcel.2021.01.028.

62. Lucas, M.C., Pryszcz, L.P., Medina, R., Milenkovic, I., Camacho, N., Marchand, V., Motorin, Y., Ribas de Pouplana, L., and Novoa, E.M. (2024). Quantitative analysis of tRNA abundance and modifications by nanopore RNA sequencing. Nat. Biotechnol. *42*, 72–86. https://doi.org/10.1038/s41587-023-01743-6.

63. Nakamura, Y., Gojobori, T., and Ikemura, T. (1998). Codon usage tabulated from the international DNA sequence databases. Nucleic Acids Res. *26*, 334. https://doi.org/10.1093/nar/26.1.334.

64. Procunier, J.D., and Tartof, K.D. (1975). Genetic analysis of the 5S RNA genes in Drosophila melanogaster. Genetics *81*, 515–523. https://doi.org/10.1093/genetics/81.3.515.

65. Hull, M.W., Erickson, J., Johnston, M., and Engelke, D.R. (1994). tRNA genes as transcriptional repressor elements. Mol. Cell Biol. *14*, 1266–1277. https://doi.org/10.1128/mcb.14.2.1266-1277.1994.

66. Raab, J.R., Chiu, J., Zhu, J., Katzman, S., Kurukuti, S., Wade, P.A., Haussler, D., and Kamakaka, R.T. (2012). Human tRNA genes function as chromatin insulators. EMBO J. *31*, 330–350. https://doi.org/10.1038/emboj.2011.406.

67. Gaertner, B., and Zeitlinger, J. (2014). RNA polymerase II pausing during development. Development *141*, 1179–1183. https://doi.org/10.1242/dev.088492.

68. Du, M., Stitzinger, S.H., Spille, J.-H., Cho, W.-K., Lee, C., Hijaz, M., Quintana, A., and Cissé, I.I. (2024). Direct observation of a condensate effect on super-enhancer controlled gene bursting. Cell *187*, 2595–2598. https://doi.org/10.1016/j.cell.2024.04.001.

69. Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J.B., and Gregor, T. (2018). Dynamic interplay between enhancer-promoter topology and gene activity. Nat. Genet. *50*, 1296–1303. https://doi.org/10.1038/s41588-018-0175-z.

70. Johnstone, C.P., and Galloway, K.E. (2022). Supercoiling-mediated feedback rapidly couples and tunes transcription. Cell Rep. *41*, 111492. https://doi.org/10.1016/j.celrep.2022.111492.

71. Sloan, C.A., Chan, E.T., Davidson, J.M., Malladi, V.S., Strattan, J.S., Hitz, B.C., Gabdank, I., Narayanan, A.K., Ho, M., Lee, B.T., et al. (2016). ENCODE data at the ENCODE portal. Nucleic Acids Res. *44*, 726–732. https://doi.org/10.1093/nar/gkv1160.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| Rabbit polyclonal anti-N6-methyladenosine | Active Motif | Cat# 61995; RRID: AB_2793658 |
| anti-rabbit IgG, HRP-conjugated secondary | Cell Signaling | Cat# 7074S; RRID:AB_2099233 |
| Bacterial and virus strains | | |
| T7 Express LysY/Iq | New England Biolabs | C3013I |
| Chemicals, peptides, and recombinant proteins | | |
| S-adenosyl-methionine | New England Biolabs | B9003S |
| Triptolide | Tocris | 3253 |
| Proteinase K | Millipore Sigma | 3115887001 |
| Critical commercial assays | | |
| Promega Wizard HMW DNA Extraction Kit | Promega | A2920 |
| Ni-NTA Agarose | Qiagen | 30210 |
| MWCO SnakeSkin dialysis tubing | Spectrum Laboratories | 132650 |
| 10K Amicon Ultra-15 spin concentrator | Millipore Sigma | UFC901008 |
| HiTrap Q HP | Cytiva | 17115301 |
| HiTrap SP HP | Cytiva | 17115101 |
| Deposited data | | |
| PRO-seq | Kwak et al., 2013[30] | GEO: GSE42117 |
| MNase-seq | Chereji et al., 2019[36] | GEO: GSE128689 |
| CAGE-seq | Sloan et al., 2016[71] | ENCODE: ENCSR863NJP |
| Hi-C | Ramírez et al., 2015[50] | GEO: GSE58821 |
| ChIP-seq | Ong et al., 2013[54] | GEO: GSE41354 |
| START-seq | Henriques et al., 2018[44] | GEO: GSE85191 |
| STARR-seq | Zabidi et al., 2015[48] | GEO: GSE57876 |
| Fiber-seq | This paper | GEO: GSE249742 |
| Experimental models: Cell lines | | |
| *Drosophila melanogaster* S2 | Expression systems | 94-005F |
| Recombinant DNA | | |
| pHia5ET | Stergachis et al., 2020[22] | N/A |
| Software and algorithms | | |
| FiberHMM | This study | https://doi.org/10.5281/zenodo.13259799 |
| fibertools v0.3.7 | Jha et al., 2023[26] | https://github.com/fiberseq/fibertools-rs |
| pbmm2 v1.13.1 | Pacific Biosciences | https://github.com/PacificBiosciences/pbmm2 |
| ccs v6.4.0 | Pacific Biosciences | https://github.com/PacificBiosciences/ccs |
| hmmlearn v0.2.5 | v0.2.5 | https://hmmlearn.readthedocs.io/en/latest/ |

### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Andrew Stergachis (absterga@uw.edu).

#### Materials availability

This study did not generate new unique plasmids, cell lines, or reagents.

## Data and code availability

- Fiber-seq data have been deposited at GEO and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. This paper also analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table. All data reported in this paper will be shared by the lead contact upon request.
- Code for analysis of PacBio sequencing data will be available at GitHub upon publication: https://doi.org/10.5281/zenodo.13259799.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

## EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

*Drosophila* S2 cells from Expression Systems (94-005F) were maintained in Schneider's Medium supplemented with 10% heat-inactivated Fetal Bovine Serum, 50U/mL penicillin and 50 μg/mL streptomycin (Gibco 15070063), and grown at 25°C.

## METHOD DETAILS

### Hia5 expression and purification

pHia5ET was expressed and purified as previously described (Stergachis 2020). pHia5ET was transformed into T7 Express lysY/Iq Escherichia coli cells (NEB C3013I). Overnight cultures were added to two 1 L cultures of LB medium supplemented with 50 μg/mL kanamycin and grown with shaking at 37°C to an OD600 of 0.8–1.0. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM and protein was expressed for 4 h at 20°C with shaking. Cells were pelleted at 5,000 rpm for 10 min at 4°C. The pellet was resuspended in 35 mL lysis buffer (50 mM HEPES, pH 7.5; 300 mM NaCl; 10% glycerol; 0.5% Triton X-100; 10 mM β-mercaptoethanol) supplemented with 2X Complete, EDTA-free Protease Inhibitor Cocktail (Millipore Sigma 11873580001). Cells were lysed by probe sonication (Qsonica Q125) for 10 min on ice at 50% amplitude, 30 s on/off. Lysate was clarified by centrifuging for 1 h at 40,000 x g. Ni-NTA Agarose (Qiagen 30210) was prepared by washing 5 mL slurry with 30 mL equilibration buffer (50 mM HEPES, pH 7.5; 300 mM NaCl; 20 mM Imidazole) and centrifuged at 500 x g for 3 min, repeating once. The clarified lysate and Ni-NTA agarose were combined and rotated at 4°C for 1 h. The lysate mixture was poured over a disposable gravity flow column (Bio-Rad 7321010) and washed with 20 mL Buffer 1 (50 mM HEPES, pH 7.5; 300 mM NaCl; 50 mM imidazole) and 15 mL Buffer 2 (50 mM HEPES, pH 7.5; 300 mM NaCl; 70 mM imidazole). Protein was eluted with 15 mL Elution Buffer (50 mM HEPES, pH 7.5; 300 mM NaCl; 250 mM imidazole). 6-8K MWCO SnakeSkin dialysis tubing (Spectrum Laboratories 132650) was pre-wet in dialysis buffer (50 mM Tris-HCl, pH 8; 100 mM NaCl; 1 mM DTT) and the eluate was added to the tubing and dialyzed against 2 L dialysis buffer overnight at 4°C. Dialyzed sample was concentrated using a 10K Amicon Ultra-15 spin concentrator (Millipore Sigma UFC901008) and centrifuged at 3,220 x g to concentrate to a volume of less than 1 mL. The concentrated sample was injected onto Tandem HiTrap Q HP (Cytiva 17115301) and HiTrap SP HP (Cytiva 17115101) columns equilibrated with FPLC Buffer A (50 mM Tris-HCl, pH 8; 100 mM NaCl; 1 mM DTT). Columns were washed with 5 column volumes of FPLC Buffer A. The Q column was removed and the sample eluted from the SP column using a linear gradient over 20 column volumes of 0–100% FPLC Buffer B (50 mM Tris-HCl, pH 8; 1 M NaCl; 1 mM DTT). Peak fractions were collected and concentrated to 250 μL using a 10K Amicon Ultra-15 spin concentrator. Protein was supplemented with glycerol to a final concentration of 10%, frozen, and stored at −80°C. Protein purity was assessed by SDS-PAGE, and its activity measured by an *in vitro* methylation assay.

### Fiber-seq

Fiber-seq was carried out as previously described (Stergachis et al. 2020) on *D. melanogaster* S2 cells. Cells were grown in the dark in 75 $cm^2$ flasks to a final density of 3 million cells/mL, with 8 million cells used per sample. Cells were then spun down at 250 x g for 5 min, resuspended in 1 mL of 1x DPBS, spun down again at 250 x g for 5 min, and resuspended in 150 μL Buffer A (15 mM Tris-Cl pH 8.0, 15 mM NaCl, 60mM KCl, 1 mM EDTA, 0.5 mM EGTA, 0.5 mM Spermidine). To isolate nuclei, a 0.05% final concentration of IGEPAL was added and cells were incubated for 10 min on ice and subsequently spun down at 500 x g for 5 min. The nuclei pellets were then resuspended in 60 μL Buffer A with the addition of 0.8 mM SAM and 200 U of Hia5, and incubated for 10 min at 25°C. The reaction was quenched with the addition of 1% final SDS and mixing using a wide-bore pipette tip. DNA was then purified using the Promega Wizard HMW DNA Extraction Kit (Promega A2920) and submitted for library preparation and sequencing.

Four sets of biological replicates were generated in S2 cells with no perturbations, with the latter 3 split into two technical replicates each -- a total of seven datasets overall. Two biological replicates in S2 cells treated with triptolide were also generated, with the latter biological replicate split into two technical replicates. For the samples treated with triptolide, a final concentration of 10 μM of triptolide was added to the cells 30 min prior to harvesting and an identical molarity maintained through the Fiber-seq protocol. Two control Fiber-seq datasets were generated as inputs for FiberHMM. The control Fiber-seq datasets were carried out identically, except that the no-Hia5 sample had no added Hia5 and that the dechromatinized sample was carried out on purified genomic *Drosophila* S2 genomic DNA instead. Specifically, control Fiber-seq on dechromatinized DNA was performed on genomic DNA purified from *Drosophila* S2 cells using the Promega Wizard HMW DNA Extraction Kit (Promega A2920). 2 μg of DNA was used in a final

volume of 60 μL in Buffer A with 0.8 mM S-adenosylmethionine and 200 U Hia5. Samples were incubated for 2 h at 37°C before quenching by purification with the Promega Wizard HMW DNA extraction kit.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Defining CAGE-seq, START-seq and PRO-seq peaks

To call CAGE- and PRO-seq peaks near TSSs, a sliding 20 bp window from −50 to +50 bp relative to the TSS (CAGE-seq) or 0 to 100 bp relative to the TSS (PRO-seq) around the TSS was used, assigning maxima within each window as peaks. A minimum of 10 reads was required to call a peak, with the strongest peak above that threshold being assigned as the primary PRO-seq or CAGE-seq peak. All secondary peaks were also required to have a minimum of 10-fold lower coverage than the primary peak. In the case of the primary CAGE-seq peak, the TSS used in subsequent analyses was adjusted from the reference annotation to that peak if the reference differed.

### Calculating pause index, expression levels, and GAF occupancy

Pause index and expression were both calculated using PRO-seq data. Pause index was calculated as the ratio of PRO-seq coverage in the promoter region (−100 to 300 bp relative to the TSS) to the body of the gene (300 bp after the TSS to the end of the gene), normalized by gene length. Expression was calculated based on PRO-seq coverage in the gene body (300 bp after the TSS to the end of the gene) normalized by gene length. GAF levels were calculated based on the height of the nearest GAF ChIP-seq peak.

### Fiber-seq processing

Fiber-seq datasets were assigned circular consensus sequences using the CCS tool (6.4.0) from Pacific Biosciences. They were then mapped to the dm6 genome using the pbmm2 package. Methylations were then called using the fibertools package (0.3.0) in SCNN mode and reads were converted to a BED file using a threshold of 250 for the m6A score.

### FiberHMM

Footprints were called on Fiber-seq reads using FiberHMM. FiberHMM is based on a hidden Markov model (HMM) with two hidden states-- accessible and inaccessible. To account for sequence-related biases from Hia5 or the methylation caller, at each position the model takes into account both the base and its surrounding 6 bp sequence context. The emission probabilities used in the model are the probabilities of methylation of a given base with its +/− 3 bp sequence context in an accessible or inaccessible state based on experimental-derived methylation rates from control datasets. The probability of methylation given an accessible state was based on the methylation frequency in a dataset generated from dechromatinized Drosophila S2 cell genomic DNA. The probability of methylation given an inaccessible state was based on the methylation frequency in a dataset generated from Drosophila S2 cells untreated with Hia5. Transition and starting probabilities for the HMM were trained 20 times on 1000 reads sampled in equal proportions from all untreated datasets and removed from subsequent analyses, with initial probabilities picked from the Dirichlet distribution with all parameters set to 1. The best model was chosen and then used for all subsequent footprint calling.

### Identifying size ranges for nucleosome footprints

Nucleosomes in general were defined as footprints greater than 90 bp. Most nucleosome footprints are single nucleosomes, which range from 90 to 200 bp, while greater than 200 bp footprints were infrequently seen and represented tightly spaced groups of nucleosomes. Analyses involving positioning and distance of single nucleosomes, specifically those featured in Figure 2, used a definition of 90–200 bp, otherwise nucleosomes were defined as greater than 90 bp.

### Identifying size ranges for PPP and PIC footprints

To identify the size ranges of PPP and PIC footprints, Fiber-seq reads overlapping promoters with a pause index greater than 10 were identified. These reads were then aligned around the primary TSS of each promoter, plotting the enrichment of different footprint sizes at positions around the TSS in the form of a heatmap (Figure 1C). This heatmap was then used to identify size ranges of two enriched footprint populations around the TSS corresponding to putative PPP and PIC footprints-- 40–60 bp and 60–80 bp respectively.

### Defining PPP and PIC footprints

PPP and PIC footprints were defined using two requirements. The first requirement was that the footprint must be within the expected size range, and the second was that the footprint must have a PRO-seq or CAGE-seq peak overlapping the middle 70% of the footprint and that the PRO-seq peak or CAGE-seq peak be within a range of 0–100 bp or +/− 50 bp around the TSS respectively. Due to higher false-positive rates of PPP calls at genes with a very low pause index, a slightly stricter definition of the PPP footprint was used for the analysis in Figure 2, requiring that the PRO-seq peak overlap the middle 50% of the putative PPP footprint.

### Identifying size ranges and defining transcription factor footprints

To identify the size ranges for transcription factor footprints, Fiber-seq reads overlapping ChIP-seq peak summits for the factor of interest were identified. These reads were then aligned around the summits, plotting the enrichment of different footprint sizes at positions around the summit in the form of a heatmap (Figures S2P and S4J–S4M). This heatmap was then used to identify size ranges of enriched footprint populations corresponding to putative transcription factor footprints-- generally 20–60 bp. For GAF, it was required that the footprint did not overlap a PRO-seq peak, and that the footprint was within +/−100 bp of the summit of the ChIP-seq peak. For CTCF, CP190, SuHW and Mod, the footprint was required to overlap the summit of the corresponding ChIP-seq peak. Footprints that did not fit any of these definitions, or those corresponding to PPPs, PICs, nucleosomes, or Pol III-associated footprints, were defined as "unknown" and labeled gray in plots showing Fiber-seq reads.

### Identifying reads with an accessible promoter or insulator

Reads with an accessible promoter or insulator were defined as those reads lacking a nucleosome footprint overlapping within +/−10 bp around the primary TSS of a gene, or +/− 25 bp around the insulator ChIP-seq peak summit.

### Validation of footprint sizes from existing structures

To compare footprint sizes to structures, structures from the referenced studies were accessed and downloaded from RCSB PDB (accession codes: 5C4X, 6GML, 6TED, 5OQJ, 7NVZ, 7NVS, 7NVT, and 7NVY) and visualized in PyMOL. The scaffold DNA bases obscured by the protein structure were counted manually to estimate the predicted footprint size associated with the structure.

### Sampling reads based on PPP and PIC footprints

To allow for direct comparison between features of reads with or without a PPP or PIC footprint on a global scale a sampling approach was used. At each gene all reads with an accessible promoter were identified. These reads were then split based on if they contained a PPP footprint, a PIC footprint, or neither. Reads at each gene were then sampled to match the count of reads with a PPP footprint to those with neither a PPP or PIC, or to match the count of reads with a PIC footprints to the count of those with neither a PPP or PIC. All analyses involving sampling were repeated with a different randomization seed at least 50 times to make sure that any conclusions were stable and not a result of sampling biases.

### Sampling reads based on distance

For comparison of the co-occupancy of PPP or PIC footprints at pairs of genes with or without an interrupting TAD boundary or insulator binding protein ChIP-seq peak, a similar sampling method was carried out as described above. The distance between each possible pair of genes was calculated, and the pairs included in the analysis were sampled from each group to capture an equal distribution of distances for each condition (with or without a TAD boundary or insulator binding protein ChIP-seq peak). This sampling was repeated 10,000x to capture the sampling variance and generate a confidence interval.

### Segmenting downstream and upstream nucleosomes

To identify the most likely position of nucleosomes upstream and downstream of the −1 and +1 nucleosome respectively, a Gaussian mixture model (GMM) was used. The GMM was trained on the distribution of nucleosome footprints either upstream or downstream to identify 95% confidence intervals for each nucleosome within the range. The number of states for the GMM was chosen based on a visual approximation of the number of peaks from the overall distribution.

### +1 and –1 nucleosome size and distance on a per-gene basis

To compare nucleosome footprint distance and size on a per-gene basis (Figures S2H–S2K), the size and distance of the +1 or −1 nucleosome footprints were calculated across all reads at each individual locus and a Wilcoxon Rank-Sum test was carried out, comparing the reads with a PPP footprint to those without a PPP footprint.

### +1 nucleosome size and distance on a per-read basis

To compare nucleosome footprint distance and size on a per-read basis (Figures S2F and S2G), the size and distance of the +1 nucleosome were calculated across all reads at each individual gene. These features were then bootstrap resampled across all reads containing an accessible promoter at each locus to calculate a mean and confidence interval for each feature at each locus. Each read with a PPP footprint was then tested against the corresponding mean and confidence interval to find the difference and significance of the +1 nucleosome footprint distance or size observed in that read compared to the other reads at the origin gene.

### Defining elongating Pol II footprints

Elongating Pol II footprints were defined based on matching the size range of a PPP footprint, 40–60 bp, and existing in gene bodies outside of a range of +/−300 bp around any annotated TSS or predicted eRNA TSS in the genome. The frequency of these footprints was defined as the count of the footprints per kilobase. The fold enrichment of elongating Pol II footprints was calculated based on the enrichment of identical footprints in intergenic regions outside of a range of +/−300bp from an annotated TSS or predicted eRNA TSS in the genome in the dataset from which that read originated.

### Pause-inhibited initiation quantification

To quantify the inhibition of simultaneous PPP and PIC occupancy, all genes which had at least one PPP and PIC footprint across all corresponding Fiber-seq reads were identified. These genes were then binned based on the distance between the primary CAGE-seq and PRO-seq peak and a contingency table was generated for the count of individual, simultaneous, and absent PPP and PIC footprints at these binned loci. Fisher's exact test was then used to determine the odds ratio and *p* value of having a simultaneous PPP and PIC footprint within each distance bin given the contingency table.

### Identifying transcribed enhancers

Transcribed enhancers were called by identifying TSSs within a 1kb window of enhancers from STARR-seq using START-seq data. Enhancer TSSs could not be within 500bp of any annotated TSS in the genome, including pseudogenes, noncoding RNAs, and other transcribed enhancers with higher maximum START-seq signal.

### Defining tRNA Pol III transcription-associated footprints

A similar strategy as the one employed to identify PPP and PIC footprints was used to identify Pol III transcription-associated footprints. First, a heatmap was generated to identify regions and sizes of enriched footprints aligned around all tRNA TSSs. Several unique populations of footprints were visible, overlapping known binding sites of TFIIIB and TFIIIC, with sizes ranging from 30 to 140 bp. We defined fibers with Pol III tRNA transcription-associated footprints as those with a footprint between 30 and 140 bp starting 45–55 bp upstream of the TSS, or with a <90 bp footprint at the B-box or A-box alone (a rare occurrence).

### Defining 5S rRNA Pol III transcription-associated footprints

5S rRNA genes showed a similar pattern of footprints to tRNA genes, but with slightly larger footprints overall. As such, Pol III 5S rRNA transcription-associated footprints were defined identically as those found at tRNA genes, except that the maximum size was set to 160 bp (Figure S5F).

### Correlation of tRNA expression score to codon frequency

The tRNA expression score was calculated as the fraction of reads containing a Pol III transcription-associated footprint compared to those without. All tRNA genes with coverage within the middle 95% of the overall distribution of coverage across the genome were also identified. Subsequently, all families of tRNA genes sharing a given codon sequence where all known members had an acceptable level of coverage were identified. The Pearson correlation of the overall count of copies of tRNA genes in each family with their codon frequency in the Drosophila genome was then compared to the correlation of codon frequency with the sum of expression scores for each family (Figure S5D).

### Global coupling analysis

The following describes the methodology used for the PPP/PIC coding promoter coupling analysis – however this approach holds true for all global coupling analyses, substituting the feature (PPP/PIC, promoter accessibility, Pol III transcription-associated footprints) and location (Pol II promoter, enhancer, tRNA, 5S rRNA). First, all pairs of genes with a distance between their TSSs of less than 10 kb were identified. Pair of genes were then binned by distance in such a manner as to allow for a roughly equal amount of simultaneous PPP and PIC footprints in each bin. All reads overlapping both members of each pair of genes in each bin were then identified, with the frequency of individual, simultaneous, and absent PPP/PIC counted to generate a contingency table. Only reads where both promoters were accessible were included. Fisher's exact test was then used to calculate an odds ratio and *p* value associated with simultaneous PPP or PIC occupancy in each set of reads binned by distance.

To test the effect of features like TAD boundaries or insulator occupancy on coupling, pairs of genes within 2.5 kb were identified with the feature of interest between them. In the case of TAD boundaries, reads from each population were sampled to capture an equal distribution of distances (rounded to within 100 bp) from each population 10,000x, with an empirical *p* value for the level of coupling calculated based on the average Fisher's exact test *p* value across the samplings. For insulators, as the occupancy and accessibility of the insulator was a single-molecule observation, significance was calculated using Fisher's exact test on each population.

### Per-locus coupling analysis

Coupling on a per-locus level was again quantified identically across features and locations. First, all pairs of genes with a distance between their TSSs of less than 10 kb were identified. All reads overlapping both members of each pair of genes were then identified, with only reads where both promoters were accessible included. The frequency of individual, simultaneous, and absent PPP/PIC footprints for each pair was then counted to generate a contingency table. Fisher's exact test was then used to identify an associated *p* value and odds ratio at each individual locus. Pairs of genes were then assigned as significantly coordinated if the odds ratio was greater than 1 and the *p* value was less than 0.05. Gene pairs were then binned based on distance and the fraction of pairs with a significant odds ratio was calculated for each bin.