Howard Hughes Medical Institute

Published as: *Mol Cell*. 2017 December 21; 68(6): 1038–1053.e4.

Transcription and remodeling produce asymmetrically unwrapped nucleosomal intermediates

Srinivas Ramachandran[1,2], **Kami Ahmad**[1], and **Steven Henikoff**[1,2,*]

[1]Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA

[2]Howard Hughes Medical Institute, Seattle, WA 98109, USA

Summary

Nucleosomes are disrupted during transcription and other active processes, but the structural intermediates during nucleosome disruption *in vivo* are unknown. To identify intermediates, we mapped subnucleosomal protections in *Drosophila* cells using *Micrococcal* Nuclease followed by sequencing. At the first nucleosome position downstream of the transcription start site, we identified unwrapped intermediates, including hexasomes that lack either proximal or distal contacts. Inhibiting topoisomerases or depleting histone chaperones increased unwrapping, whereas inhibiting release of paused RNAPII or reducing RNAPII elongation decreased unwrapping. Our results indicate that positive torsion generated by elongating RNAPII causes transient loss of histone-DNA contacts. Using this mapping approach, we found that nucleosomes flanking human CTCF insulation sites are similarly disrupted. We also identified diagnostic subnucleosomal particle remnants in cell-free human DNA data as a relic of transcribed genes from apoptosing cells. Thus identification of subnucleosomal fragments from nuclease protection data represents a general strategy for structural epigenomics.

Graphic abstract

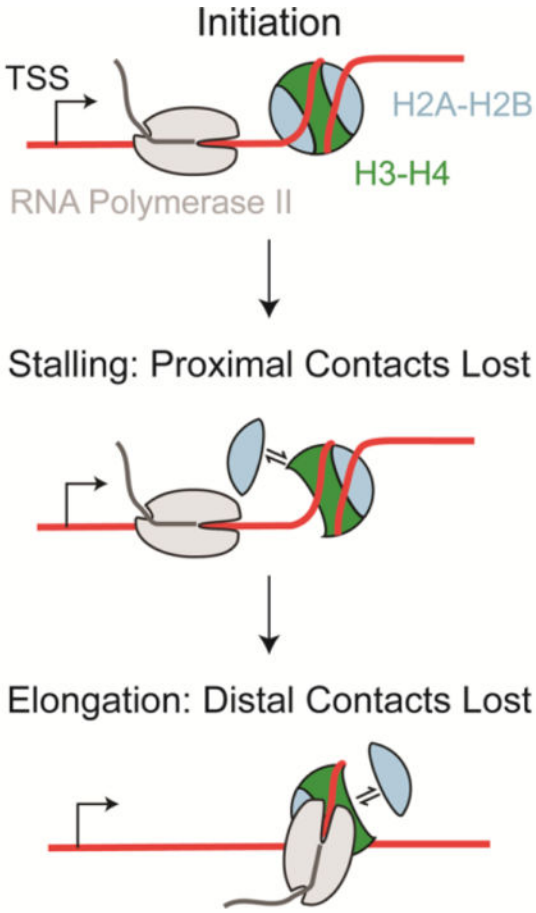
[*]Address correspondence to: Steven Henikoff, 1100 Fairview Avenue N. A1-162, Seattle WA 98109; tel: (206) 667-30 4509; fax: (206) 667-6497; steveh@fhcrc.org.

Author Contributions

S.R. conceived the study under the guidance of S.H. S.R. performed the experiments and data analysis. K.A. contributed the *Drosophila* CUT&RUN data. S.R. and S.H. wrote the manuscript.

DATA AND SOFTWARE AVAILABILITY

Datasets generated for this study have been submitted to GEO (GSE98351). All published datasets used in this paper are summarized in the Table S5.



eTOC blurb

Ramachandran et al. identify nucleosomal intermediates formed during transcription through the nucleosome *in vivo* from epigenomic datasets. They map specific histone-DNA contacts that are lost during transcription through the nucleosome that is adjacent to the promoter. They then generate transcriptional signatures from sub nucleosomal patterns in cell free DNA datasets.

Keywords

subnucleosome; hexasome; structural epigenomics; transcription elongation; torsional strain; cell free-DNA□

Introduction

Nucleosomes present formidable barriers to transcription *in vitro*. RNA polymerase II (RNAPII) stalls at the nucleosome entry site, at the H2A/H2B dimer/H3-H4 tetramer interface, and at the dyad position *in vitro* (Bondarenko et al., 2006). Productive transcription of RNAPII through a nucleosome *in vitro* requires high salt (Bondarenko et al., 2006; Hodges et al., 2009), histone chaperones (Hsieh et al., 2013) or remodelers (Gaykalova et al., 2011; Kuryan et al., 2012). Transcription through the nucleosome *in vitro*

produces hexasomes as identified by native polyacrylamide gel electrophoresis (PAGE) analysis of nucleosomes post-transcription (Kireeva et al., 2002). During multiple rounds of *E. coli* RNAP transit *in vitro*, an H2A/H2B dimer is lost and then the remaining hexasome is disrupted, resulting in nucleosome loss (Kulaeva et al., 2010).

Whereas nucleosomal intermediates formed during transcription have been identified *in vitro*, their *in vivo* counterparts have not been identified. *In vivo*, transcription elongation occurs efficiently through the chromatin template. RNAPII predominantly stalls at the entry site of the first nucleosome downstream of the TSS (+1 nucleosome) (Weber et al., 2014). Further into the gene body, RNAPII elongation speeds up (Jonkers et al., 2014). One hypothesis for the different stall location of RNAPII *in vivo* from the locations seen *in vitro,* and the high efficiency of elongation downstream of the +1 nucleosome *in vivo*, is that when RNAPII stalls at the entry of the +1 nucleosome, elongation factors, many of which include histone chaperones, facilitate nucleosome disruption and efficient elongation.

For RNAPII to transcribe through a nucleosome, all histone-DNA contacts must be transiently broken. The resulting nucleosomal intermediates are either histone octamers with unwrapped DNA or are subnucleosomal particles with less than an octamer of histones. Here we identify nucleosomal intermediates associated with transcription elongation in *Drosophila* cells to uncover the mechanism of transcriptional elongation through nucleosomes *in vivo*. We extended our approach to enhancer and insulator nucleosomes and to human cell-free DNA (cfDNA) remnants as proxies for transcriptional activity to distinguish between plasma from cancer patients and healthy individuals. The identification of subnucleosomal transcription intermediates thus represents a general strategy for epigenomic profiling.

Results

The first nucleosome of transcribed genes is enriched for subnucleosomal protections

Micrococcal nuclease preferentially degrades linker DNA and is inhibited when it encounters a protein-DNA contact. We have previously shown that the fragment lengths and fragment endpoints from MNase-seq experiments can be used to infer protein-DNA contacts in budding yeast genome-wide (Henikoff et al., 2011). We now ask if this approach can be applied to the detection and mapping of nucleosomal intermediates formed due to transcription elongation. Using MNase-seq data from *Drosophila* cells, we generated enrichment profiles of ~147 bp fragments and shorter fragments. When we plotted the enrichment relative to the first nucleosome position (+1 position) downstream of the transcription start site (TSS) ± 2000 bp, we observed that the shorter fragments were enriched at positions flanking the nucleosome depleted region (NDR) compared to positions further downstream and upstream of the NDR (Figure 1A). The +1 position is highly enriched for fragments shorter than 147 bp (Figure 1), which is also most enriched in stalled RNAPII (Weber et al., 2014). The enrichment of both subnucleosomal protections and stalled RNAPII at the +1 position suggests that the partial loss of protection from MNase cleavage might result from transcription.

In a paired-end MNase-seq experiment, the highest enrichment is for fragments that are of nucleosomal size. With lightly digested chromatin, we observed the highest enrichment at 147 and 167 bp, corresponding to the nucleosome and chromatosome, respectively (Figure 1B). We also observed shorter protections at a lower frequency. To enrich for shorter fragments that might represent nucleosomal intermediates over the +1 position, we PAGE on fragments generated from light MNase digestion of nuclei from *Drosophila* S2 cells. We excised the gel section corresponding to fragments less than ~100 bp, purified the DNA and performed paired-end Illumina sequencing. We observed robust sampling of shorter fragments from this gel-purified sample (Figure 1B), which we used to determine the distributions of the center of the short fragments relative to the +1 position. As a control, we chose nucleosomes that were as well positioned as +1 nucleosomes but which featured no transcribing RNAPII, as represented by a lack of 3'NT signal at these positions (Figure S1A-F).

Using the +1 nucleosome position as a landmark, we plotted the distributions of the fragment centers of subnucleosomal protections. We observed that fragments with lengths around 132 and 125 bp could be represented as a unimodal distribution that is centered very close to the dyad of the +1 nucleosome (Figure 1C, Figure S2A). In contrast, the distributions of the center of 112, 103, and 90 bp fragments were bimodal and could be fitted with two Gaussian curves centered on either side of the +1 nucleosome position (Figure 1E, G, Figure S2B). These bimodal distributions imply that one population represents fragments that have lost histone-DNA contacts on the edge of the +1 nucleosome proximal to the promoter and the other population represents fragments that have lost histone-DNA contacts on the edge of the +1 nucleosome distal to the promoter. At control nucleosome positions, the 132, 125, 112, 103, and 90 bp fragments are 4–5 fold depleted compared to the +1 positions (Figure 1D, F, and H, Figure S2C, D), indicating that the enrichment of shorter fragments correlates with transcription. We observed similarly enriched subnucleosomal protections at gene body nucleosomes (Figure S2E-P). To confirm that these subnucleosomal fragments at the +1 nucleosome position arise due to loss of histone-DNA contacts *in vivo* and not due to nucleosomal disassembly during the MNase reaction, we analyzed data from an experiment where chromatin was crosslinked with formaldehyde prior to treatment with MNase (Ramachandran and Henikoff 2016). In the crosslinked MNase-seq data, we observed unimodal distributions for 132 and 125 bp fragments and bimodal distributions for 90, 103, and 112 bp fragments at the +1 nucleosome position, similar to what we observed in native MNase-seq datasets, implying that these subnucleosomal protections are present *in vivo* (Figure S1H-K).

To analyze the distribution of 112, 103, and 90 bp fragments relative to the 147 bp fragments at the +1 nucleosome position independent of curve fitting, we calculated the cross-correlation of the distribution of 147 bp fragments and the subnucleosomal fragments on a gene-by-gene basis. Cross-correlation analysis determines the distribution of subnucleosomal fragments relative to nucleosomal fragments without knowledge of a "called" nucleosome position (Movie S1), thus making no assumptions about the H3-H4 dyad positions of the nucleosomes or subnucleosomes and accounting for nucleosomes with different extents of positioning. Using the *k*-means method, we could aggregate the cross-correlation profiles into two major clusters (Figure S3A-L). To determine quantitatively the

distance of the subnucleosomal particles relative to each other and to the nucleosomal particles, we calculated the cross-correlation of 90 bp fragments and 103 bp, 112 bp, and 147 bp fragments. We observed that the center of the 90 bp particles is on average 28 bp from 147 bp particles, 10 bp from 112 bp particles, and 6 bp from 103 bp particle (Figure S3M-N). The subnucleosomal and the nucleosomal fragment classes have one of their edges at the same position, whereas the other edge is closer to the nucleosomal dyad position going from 147 bp to 90 bp fragments. We infer that the positions of sub nuclosomes relative to the nucleosome arise from sequential loss of contacts from the entry site on one side of the nucleosome. Thus, the cross-correlation analysis suggests that all three subnucleosomal fragment sizes arise from loss of contacts to distal or proximal H2A-H2B dimers. In summary, we interpret the 132 and 125 bp fragments as representing nucleosomal unwrapping symmetrically on both sides of the dyad, whereas 112, 103, and 90 bp fragments represent asymmetric nucleosomal unwrapping, where DNA contacts to a single H2A/H2B dimer would be lost at increased rates at transcribed nucleosome positions.

Subnucleosomal protections reflect nucleosome dynamics

We also observed a high enrichment of subnucleosomal fragments at the −1 position (Figure 1A). In *Drosophila*, there is no upstream anti-sense transcription as observed in mammals (Nechaev et al., 2010). Thus, the subnucleosomal enrichments in the −1 position should arise independent of transcriptional elongation. When we plotted the 103 bp fragments around the −1 position, we observed a bimodal distribution representing two populations that feature loss of contacts to either the proximal or the distal H2A-H2B dimer (Figure 2A). To ask if such protection occurs at other regions of nucleosome dynamics, we analyzed DNA-binding data for the CTCF (CCCTC-binding factor) insulator protein obtained using CUT&RUN (Cleavage Under Target and Release Under Nuclease). With CUT&RUN, the chromatin of permeabilized cells or nuclei is successively bound by a factor-specific antibody and a protein-A-MNase (pA-MN) fusion protein *in situ*. The MNase is activated with calcium to produce targeted cleavages on either side of the antibody/pA-MN-bound particle, releasing it into the supernatant for paired-end DNA sequencing. Applied to CTCF in human K562 cells, CUT&RUN not only mapped CTCF binding at high-resolution, but it also specifically mapped the phased nucleosomes that are immediately adjacent to the CTCF sites (Skene and Henikoff, 2017). Thus, we could analyze fragment size distributions from the nucleosome upstream of human CTCF insulator sites genome-wide. We observed a bimodal distribution of 103 bp fragments at the first nucleosome position upstream of CTCF sites, which represents two populations featuring loss of contacts to the proximal or distal H2A-H2B dimer (Figure 2B). In summary, we observed protections corresponding to a hexamer of histones at sites of nucleosome dynamics in both *Drosophila* and human cells.

The +1 nucleosome dyad is accessible in active chromatin

We extended our analysis to 80, 68, and 58 bp particles and generated a map of representative subnucleosomal particles observed in either low-salt-soluble chromatin (Figure 2C) or total chromatin (Figure 2D). Low-salt-soluble chromatin is released when intact nuclei are washed with a buffer containing 80 mM salt after MNase treatment and is enriched for transcriptionally active chromatin (Henikoff et al., 2009; Rocha et al., 1984; Sanders, 1978). We have shown that low-salt-soluble chromatin is also enriched for

subnucleosomal protections (Teves and Henikoff, 2011; Weber et al., 2010). We observed a similar distribution of subnucleosomal fragments in both low-salt-soluble chromatin and total chromatin in fragments ranging between 132 bp and 80 bp. However, we observed strikingly different distributions for 68, and 58 bp fragments in active chromatin compared to total chromatin. In active chromatin, we observed protections on either side of the nucleosome dyad, representing two halves of the nucleosome, indicating that the nucleosome dyad is accessible (Figure 2E). In total chromatin, the major population of 68 bp particles predominantly protects the region that would correspond to the central H3-H4 tetramer in an intact nucleosome (Figure 2F). This protection is similar to that of H3-H4 tetrasomes *in vitro* (Li and Wang, 2012). Thus, bulk chromatin, which is dominated by a lack of transcription, resembles nucleosome assembly *in vitro*, whereas active chromatin is accessible at the dyad (Huang et al., 2013; Kumar and Leffak, 1986).

RNAPII elongation drives preferential loss of contacts to the distal H2A/H2B dimer

We wondered if loss of DNA contacts on the side of the nucleosome that is proximal or distal to the promoter represents the same or different modes of transcription. To compare genes with different transcription levels, we first divided genes into five clusters based on RNAPII profiles relative to the +1 nucleosome as represented by 3’NT. In Cluster 1, the active site of RNAPII peaks outside the +1 nucleosome entry site, in Cluster 2, RNAPII peaks at the entry site, in Clusters 3 and 4, RNAPII encroaches on the +1 nucleosome, and in Cluster 5, RNAPII peaks over the dyad axis (Figure 3A, B). We then plotted the distribution of centers of 112, 103, and 90 bp fragments relative to the +1 position for the five clusters (Figure 3C-E). For Clusters 1 and 2, the frequency of fragments that represent loss of contacts to the proximal dimer is higher than the frequency of fragments that feature loss of contacts to the distal dimer. For Clusters 3 and 4, we observe a gradual increase in the relative frequency of distal to proximal loss of dimer contacts, and for Cluster 5, in which RNAPII is at the dyad, the frequency of fragment centers on either side of the dyad is approximately equal. We quantified the preferential loss of contacts to the proximal dimer versus the distal dimer on a gene-by-gene basis using the asymmetry of cross-correlation on the distal side relative to the proximal side. A higher cross-correlation asymmetry means preferential loss of contacts to the distal dimer and lower cross-correlation asymmetry means preferential loss of contacts to the proximal dimer. We found that as the RNAPII peak goes from the nucleosome entry site (Cluster 1) to the dyad (Cluster 5), the asymmetry changes to more positive values (Figure 3F-H). This gradual shift implies that contacts to the distal dimer are preferentially lost as RNAPII transcribes through the nucleosome.

We then asked if the extent of transcription through the +1 nucleosome correlates with loss of contacts to the distal dimer. We identified six sets of genes based on the level of 3’NT signal at the +1 nucleosome and plotted the cross-correlation asymmetry for each set. We found the asymmetry to be more positive with higher levels of transcription (Figure 3I-K). This implies that preferential loss of contacts to the distal dimer correlates with the level of transcription through the +1 nucleosome. Previous work has shown that the +1 nucleosome presents the highest barrier to RNAPII compared to gene body nucleosomes (Weber et al., 2014). Thus, gene body nucleosomes would be associated with higher levels of transcription elongation compared to the +1 nucleosome. We plotted the cross-correlation asymmetry of

the +1, +2, +3, and +4 nucleosomes and found the +1 nucleosome to show the lowest cross-correlation asymmetry (Figure 3L). This implies that at the +1 nucleosome, which causes the highest stalling of RNAPII, there is preferential loss of contacts to the proximal dimer compared to the gene body nucleosomes, which are more subject to transcriptional elongation.

As loss of contacts to the distal dimer correlates with transcriptional elongation, we asked if the ratio of fragments representing distal or proximal loss of histone-DNA contacts changes upon perturbation of RNAPII elongation. We analyzed datasets where elongation had been inhibited by the drug DRB, which inhibits release of paused RNAPII, or by heat shock, which results in a global reduction in RNAPII elongation. We calculated the change in the ratio of loss of contacts to the distal dimer to loss of contacts to the proximal dimer for 90, 103, and 112 bp fragments. We observed a 10–50% decrease in the population representing loss of contacts to the distal dimer in the three fragment lengths upon reduction in RNAPII elongation (Figure 4A). The strength of this effect depended on the extent of subnucleosomal contacts: 112 bp fragments had the strongest effect followed by 103 bp fragments, whereas 90 bp fragments had the weakest effect. We observed similar but weaker effects at the −1 nucleosome (Figure S4). Because there is minimal transcription upstream of the promoter in *Drosophila*, we conclude that nucleosome dynamics independent of RNAPII elongation also partially contribute to the changes in loss of contacts to the distal dimer upon heat shock or treatment with DRB.

We asked if the loss of contacts to the distal H2A-H2B dimer upon heat-shock is predictive of changes in RNAPII at the +1 nucleosome. We calculated the change in cross-correlation asymmetry upon heat shock for each gene and then compared the lowest quartile, which featured genes that preferentially lost contacts to the proximal dimer, to the highest quartile, which featured genes that preferentially lost contacts to the distal dimer upon heat shock for the 112 and 103 bp fragments (Figure 4B). We observed that the lowest quartile according to change in asymmetry showed a stronger decrease in RNAPII during heat-shock compared to the highest quartile (Figure 4C). Thus, loss of contacts to the distal dimer during heat-shock predicts a weaker decrease in transcription. A subset of genes go up in expression during heat-shock. For these genes, we observed increased loss of contacts to the distal dimer (Figure 4D), which corresponds to increased transcription expected from these genes. Therefore, RNAPII elongation leads to preferential loss of contacts to distal H2A-H2B. These subnucleosomal particles could arise from disruption of the nucleosome due to RNAPII transit, from transcription-coupled nucleosome remodeling, or from nucleosome assembly taking place in the wake of RNAPII transit.

Positive torsion generated by elongating RNAPII contributes to distal dimer loss

Positive torsion can destabilize nucleosomes (Lee and Garrard, 1991; Teves and Henikoff, 2014). The formation of an ~15 bp bubble at the RNAPII active site generates positive torsion in front of elongating RNAPII. We asked if this positive torsion contributes to the preferential loss of distal H2A-H2B that is associated with RNAPII elongation. We analyzed previous datasets from our laboratory where MNase-seq was performed on nuclei after poisoning of either Topoisomerase I by Camptothecin or Topoisomerase II by ICRF-193

(Teves and Henikoff, 2014). As both topoisomerases relieve the positive torsional stress caused by RNAPII transit, topoisomerase poisoning increases positive torsion exerted on nucleosomes in actively transcribed genes (Teves and Henikoff, 2014). In both cases, when we analyzed the ratio of 90, 103, and 112 bp fragments, we observed a 20–50% increase in the loss of contacts to distal H2A-H2B compared to proximal H2A-H2B (Figure 4E).

The generation of positive torsion downstream of RNAPII is balanced by the generation of negative torsion upstream, and both torques are relieved by topoisomerases as RNAPII moves forward. Topoisomerase poisons would therefore be predicted to have a weaker effect on the nucleosome at the −1 position than at the +1 position. As subnucleosomal particles are enriched at both positions (Figure 1A), we can ask how the effects of topoisomerase poisons on the −1 nucleosome compares with their effects on the +1 nucleosome. Indeed, we observed much weaker changes in the ratio of loss of contacts to the distal dimer compared to the proximal dimer at the −1 position due to treatment with topoisomerase poisons (Figure S4).

FACT prevents dimer loss

We next used our MNase-seq fragment mapping assay to ask if histone chaperones play roles in subnucleosome production or prevention. Previous *in vitro* studies have found both nucleosome disruptive and protective activities for histone chaperones during elongation (Formosa, 2013). As we can measure the effect of RNAPII elongation on nucleosomes *in vivo*, we first asked if the extent of subnucleosome production correlates with the presence of histone chaperones FACT and Spt6, which enhance RNAPII elongation and stabilize nucleosomes during transcriptional elongation. Subunits of FACT, Spt16 and SSRP (Pob3 in *S. cerevisiae*) have been shown to bind H2A/H2B (Kemble et al., 2015), whereas Spt6 has been shown to genetically and biochemically interact with H3/H4 but not H2A/H2B (Bortvin and Winston, 1996). We used RNAi to deplete the two subunits of FACT (Spt16 and SSRP) and Spt6. On the one hand, depletion of Spt6 caused no change in the occupancy of 90 bp particles compared to control, which agrees with Spt6 mainly being a H3/H4 chaperone. On the other hand, depleting either subunit of FACT had a strong increase in 90 bp particles at the +1 nucleosome position (Figure 5A). Therefore, destabilizing nucleosomes during transcription in the absence of FACT subunits contributes to loss of H2A-H2B dimers. Interestingly, knockdown of Spt16 resulted in a 15% increase in loss of contacts to the proximal dimer, whereas depletion of SSRP resulted in a 15% increase in loss of contacts to the distal dimer (Figure 5B). Because both Spt16 and SSRP interact directly with opposite H2A-H2B dimers (Kemble et al., 2015), we infer that the increase in 90 bp particles upon knockdown of either subunit results from loss of H2A-H2B dimers to produce hexasomes.

Subnucleosomal particles contain histones

We next interrogated the histone content of the subnucleosomal particles. First we asked if the subnucleosomal fragments we observed result from extension of a histone protection by RNAPII. Assuming protection of 30 bp by RNAPII, we would expect that the center of RNAPII is positioned at a distance of ±48, ±41, ±37, and ±30 bp from the center of subnucleosomal fragments represented by 125, 112, 103, and 90 bp particles respectively if

RNAPII were extending histone protections in sub nucleosomes (schematic shown in Figure S5A). We calculated the cross-correlation between RNAPII profiles (3'NT data) and the profiles of centers of 125, 112, 103, and 90 bp fragments on a gene-by-gene basis and then plotted the aggregate cross-correlation. We observed that the cross-correlation peaks are adjacent to the distal edge of the fragments (Figure S5B-E), ruling out RNAPII to be part of the sub nucleosomes and further suggesting that RNAPII stalling at the edge of the histone contacts lost could produce these subnucleosomal species. To determine whether our cross-correlation analysis is sensitive enough to detect small populations of RNAPII contributing to subnucleosomal protections, we generated a mock dataset where 90% comes from the original 3'NT data and 10% from shifting the subnucleosome profile to where RNAPII would be sitting if it were to extend the subnucleosomal footprint. Even if only 10% of RNAPII were extending the footprint, we would see a significant cross-correlation peak at the predicted position as seen for the mock dataset, confirming the sensitivity of our cross-correlation analysis (Figure S5B-E). As the total number of subnucleosomal fragments amounts to ~40% of that of the intact +1 nucleosome, we can rule out RNAPII as part of the subnucleosomal protections of length 125, 112, 103, and 90 bp.

To ask if the subnucleosomal protections we observe contain histone H3, we first performed CUT&RUN directed against acetylated histone H3-Lysine 27 (H3K27ac), a modification that is highly enriched at enhancers and active promoters. We observed a 10-fold enrichment of H3K27ac at the +1 position of active genes and a signal close to background at inactive genes (Figure 5C). We then plotted the distribution of centers of 125 bp fragments and observed a symmetric distribution that could be fitted with a single Gaussian curve as observed in MNase-seq data (Figure 5D). In contrast, the distribution of 112, 103, and 90 bp fragments could be fitted with a mixture of two Gaussian curves (Figure 5E-G), similar to what was observed with MNase-seq data. To confirm that there are two populations featuring loss of contacts to either proximal or distal H2A-H2B dimers, we plotted the subnucleosomal distributions of the H3K27ac CUT&RUN data for the asymmetric clusters defined using the MNase-seq data (shown in Figure S3). We observed that the asymmetry in H3K27ac CUT&RUN subnucleosome protections could be predicted by MNase-seq data (Figure 5H-J), implying that the sub nuclesomal protections observed in H3K27ac CUT&RUN are identical to the subnucleosomal protections observed in MNase-seq and demonstrating that the subnucleosomal protections contain histone H3.

We further confirmed the histone H3 content of the subnucleosomal protections over the +1 nucleosome position by analyzing CATCH-IT data (Teves and Henikoff, 2011), which maps H3.3 containing particles. We observed symmetric protection for 125 bp fragments and asymmetric protections for 112, 103, and 90 bp fragments in CATCH-IT, further evidence that these particles over the +1 nucleosome position represent nucleosomal intermediates (Figure S6A-D).

To confirm the H2A/H2B content of the subnucleosomal protections, we analyzed previous H2A.Z MNase sequential ChIP-seq data from our laboratory (Weber et al., 2010). The nucleosomal intermediates associated with transcription could result from two possible structural entities. An intact histone octamer could have DNA unwrapped at either the entry or exit site of the nucleosome. Alternatively, a hexasome formed by the loss of an H2A/H2B

dimer would protect 90, 103, or 112 bp of DNA (Figure S7). In our sequential ChIP-seq experiment, cells expressed two ectopic copies of H2A.Z, one with a FLAG epitope and the other with a Biotin Ligase Recognition Peptide that gets biotinylated *in vivo*. The MNase-digested nucleosome preparation was crosslinked with formaldehyde, which prevented loss of dimers during immunoprecipitation. The first pull-down for FLAG would enrich homotypic and heterotypic nucleosomes and octamers with unwrapped DNA and hexasomes with a FLAG-tagged H2A.Z-H2B dimer. FLAG peptide was used to elute FLAG-containing particles, and in the second pull-down performed with streptavidin beads, only full nucleosomes containing both FLAG- and biotin-tagged H2A.Z/H2B dimers can be recovered (Figure 6A). Thus, if shorter fragments represent octamers with unwrapped DNA, the enrichment would not change in the second pull-down compared to the first pull-down when normalized to 147 bp particles. However, if the shorter fragments correspond to hexasomes, then they should be absent from the second ChIP (Figure 6A). We assume heterotypic particles and double-FLAG particles would be lost in both the short fragment population and in the 147 bp population during the second pull-down. Because we normalize the short fragment populations to the 147 bp population, their loss would be factored out and the loss of signal in the second ChIP gives us an upper limit estimate of the extent of hexasome production. We plotted the distributions of 112, 103, and 90 bp fragments from the first FLAG pull-down and the second streptavidin pull-down. In the first pull-down, we observed populations that represent loss of DNA contacts on both proximal and distal sides of the dyad, similar to the MNase-seq distribution (Figure 6B, C, Figure S6E). However, in the second pull-down, the populations represented by loss of DNA contacts both on the side distal to the promoter and on the side proximal to the promoter were reduced by ~50% for 90, and 103 bp fragments (Figure 6B, C). For 112 bp fragments, the population represented by loss of contacts to the proximal dimer were reduced by 54%, but the population represented by loss of contacts to the distal dimer remained the same (Figure S6E). Thus, we conclude that up to half of the fragments representing loss of proximal contacts arise from hexasomes for 90, 103, and 112 bp fragments and up to half of the fragments representing loss of distal contacts arise from hexasomes for 90 and 103 bp fragments.

Subnucleosome enrichment is a proxy for gene expression

Based on our results from *Drosophila* cells, we interpret the abundance of subnucleosomal fragments at the +1 nucleosome position to be diagnostic of transcription elongation. Recently, cell-free DNA (cfDNA) sequenced from human blood plasma revealed the presence of nucleosomal fragments (Snyder et al., 2016). Shorter nucleosome repeat lengths over gene bodies correlated with genes active in the lymphoid/myeloid cell types in healthy individuals. The blood cell origin of these DNA fragments is consistent with their generation during erythroid maturation or blood cell turnover (Underhill et al., 2016), as originally proposed by Williamson (Williamson, 1970). This active gene signature was altered in cfDNA from individuals with cancer, suggesting that other cell types were giving rise to cfDNA in disease states.

We asked if we could infer transcription status by analyzing subnucleosomal fragments from the same cfDNA datasets that were previously analyzed based on nucleosome repeat lengths. We called the +1 nucleosomes in these datasets and analyzed subnucleosomal fragments

relative to the +1 nucleosome position. We rank-ordered the genes based on average expression in lymphoid/myeloid cell types and for the top and bottom 1500 genes we plotted the length distribution of fragments that were within 50 bp of the +1 nucleosome position (Figure 7A). We observed that expressed genes were represented by a higher frequency of subnucleosomal fragments (50–100 bp) and a lower frequency of nucleosomal fragments (~150 bp) compared to non-expressed genes. There was a peak in the length distribution at 92 bp, which corresponded closely to the distribution of hexasomal fragments we observed in *Drosophila* cells. When we plotted the distribution of 92±10 bp fragments relative to the +1 nucleosome position, we observed a higher frequency of fragments from expressed genes compared to non-expressed genes as predicted from the length distributions. We also observed a bimodal distribution of fragment centers indicating loss of histone-DNA contacts with either proximal or distal H2A/H2B (Figure 7B). The population representing loss of histone-DNA contacts on the edge of the nucleosome distal to the TSS was conspicuously enriched in expressed genes compared to non-expressed genes. Thus, we are able to detect subnucleosomal protections correlating with transcription in human cfDNA, which indicates that subnucleosomal production due to transcription through the +1 nucleosome position is conserved between *Drosophila* and humans.

A motivation for sequencing cfDNA fragments from circulating blood plasma is to identify their tissue-of-origin to enable non-invasive diagnosis of diseases (Snyder et al., 2016). We asked if subnucleosomal fragment enrichment over the +1 nucleosome would provide a transcriptional signature that could distinguish healthy states from disease states. We devised a simple metric, the ratio of the number of subnucleosomal fragments to the number of nucleosomal fragments that map to ±50 bp of the +1 nucleosome of each gene and plotted this ratio to the expression level of genes from 76 cell lines and tissues. The smoothed expression level of genes from two representative datasets was plotted against subnucleosome enrichment calculated from the cfDNA dataset of a healthy donor, denoted as IH02 (Figure 7C, D). We find that the smoothed expression profile of the MOLT-4 cell line of lymphoid origin correlates well with subnucleosome enrichment of the IH02 dataset ($r^2 = 0.42$), whereas the expression profile of skin is uncorrelated ($r^2 = 0.01$) with subnucleosome enrichment of IH02. We rank-ordered the correlations between subnucleosome enrichment and gene expression for all 76 cell lines and tissues and generated receiver operating characteristic (ROC) curves. We set lymphoid/myeloid cell lines and tissues as true positives, because cfDNA in a healthy individual arises mainly from normal turnover of blood cells (Lam et al., 2017; Williamson, 1970). We observed an area under the ROC curve (AUC) of 0.91 for the healthy sample, indicating that subnucleosomal fragments mapping within the +1 nucleosome position can be used as a proxy for gene expression profile (Figure 7E). All five cfDNA samples from cancer patients had AUCs that were lower than the AUC of the healthy donor. A lower AUC means that the subnucleosomal enrichment resulted from expression of genes that are not active in the blood cells, indicating mixing of DNA from non-blood cells, which had significant aneuploidies based on metrics of chromosome balance (Snyder et al., 2016). To ensure that the AUC differentiation between healthy and disease samples is robust, we performed 100 rounds of random selection of equal numbers of reads from each sample and calculated the AUC. The AUC distribution after subsampling the cfDNA data from the healthy individual

was significantly upshifted compared to the AUC distributions of the subsampled cfDNA data from individuals with cancer (Figure 7F), indicating that subnucleosome enrichment robustly inferred contamination of blood cfDNA by cfDNA from non-blood tissues.

We also generated ROC curves using the published coefficients of correlation between fast Fourier Transformation (FFT) intensity at ~199 bp, a measure of nucleosome spacing over the gene body, and expression levels of genes (Snyder et al., 2016). Although the FFT method had a higher AUC for the healthy sample (0.96, Figure 7G), our subnucleosome enrichment detected significantly greater differences between healthy states and disease states (Figure 7H, p-value = 0.02 from paired t-test), indicating that subnucleosome enrichment at just the +1 position is better at identifying disease states compared to nucleosome repeat lengths over gene bodies. Including gene body nucleosome positions in calculating subnucleosome enrichment scores reduced the difference between healthy states and disease states (Figure 7H), indicating that the expression signature reported by subnucleosome enrichment extends only to the +1 position. This agrees with results from *Drosophila* cells, where subnucleosomes are highly enriched at the +1 position compared to the rest of the gene body. In summary, subnucleosome enrichment in cfDNA reflects transcription and can potentially indicate release of DNA from apoptotic diseased cells into a patient's blood.

Discussion

We have shown that MNase-seq can be used as a genome-scale readout of protein-DNA interactions. Our analysis has enabled the mapping of hexasomal histone-DNA intermediates at the first nucleosome position downstream of the TSS *in vivo*. Using 147 bp protections as the reference, we showed asymmetric loss of histone-DNA contacts for 112-, 103- and 90-bp fragments, which when mapped to the nucleosome structure correspond closely to protection by only a hexamer of histones. We confirmed the histone content of these intermediates using CATCH-IT, H2A.Z ChIP-seq, and H3K27ac CUT&RUN mapping data.

There are two possible hexasomal protections for a nucleosome, and because transcription is directional, we can distinguish the two protections as loss of contacts with H2A/H2B that are either proximal or distal relative to the promoter. We found that the preference for loss of contacts with the distal dimer compared to loss of contacts with the proximal dimer depends on the elongation state of RNAPII and the presence of histone chaperones. Positive torsion due to RNAPII elongation, and more elongating RNAPII compared to paused RNAPII both resulted in loss of contacts with the distal H2A/H2B dimer more than the proximal H2A/H2B dimer (Figure 6D). Depletion of either subunit of the FACT complex resulted in overall H2A/H2B dimer loss consistent with the importance of FACT in maintaining the integrity of nucleosomes during transcriptional elongation. In contrast, no effect on H2A/H2B dimer loss was observed upon depletion of Spt6, which is an H3/H4-specific chaperone. Depletion of the individual subunits of FACT affected opposite H2A/H2B dimers in that depletion of Spt16 led to increased proximal dimer loss, whereas depletion of SSRP resulted in increased distal dimer loss. Structural data for the FACT complex in budding yeast (Kemble et al., 2015) suggests that Spt16 contacts one H2A-H2B dimer and SSRP contacts the other H2A-H2B dimer of the nucleosome. Our data suggest that the choice of

H2A-H2B dimer to contact by Spt16 and SSRP is aligned to the direction of transcription, with Spt16 contacting the proximal dimer and SSRP contacting the distal dimer, because loss of Spt16 leads to increased loss of the proximal dimer and loss of SSRP leads to increased loss of the distal dimer.

The loss of H2A/H2B dimers has long been thought to be a consequence of transcription. An early analysis of the composition of histones bound to RNAPII in cultured cells showed H3/H4 to be double the molar amount of H2A/H2B, suggesting loss of dimers during transcription (Baer and Rhodes, 1983). FRAP experiments showed that H2B exchanged much faster than H3 and H4, and this fast exchange was dependent on active transcription (Kimura and Cook, 2001). But FRAP measurements themselves cannot resolve if the H2A/H2B turnover is directly due to transcription elongation. When ChIP enrichments were compared between H2B and H4 after induction of certain genes in budding yeast, the loss of H2B ChIP signal was higher than that of H4 ChIP signal after induction (Cole et al., 2014). The increased loss of H2B compared to H4 upon induction indicated the loss of H2A/H2B was greater than the loss of the whole nucleosome (which would be reported by loss of H4). Our maps of hexasomes at +1 nucleosome positions in *Drosophila* cells lead us to conclude that previous observations of H2A/H2B loss correspond to production of hexasomes, a discrete, mappable nucleosomal intermediate formed due to transcription elongation.

RNAPII elongation through a nucleosome leads to hexasome production *in vitro* while maintaining the same nucleosome position (Kireeva et al., 2002). Hexasomes can be reconstituted reliably *in vitro* (Arimura et al., 2012; Azegami et al., 2013) and furthermore, manipulation of the DNA sequence can control the distal or proximal position of the H2A/H2B dimer (Kulaeva et al., 2009; Levendosky et al., 2016). These directional hexasomes were used to ask if they presented similar barriers to elongating RNAPII *in vitro*. When H2A/H2B proximal to the promoter was missing, RNAPII arrested, but when H2A/H2B distal to the promoter was missing, RNAPII could traverse the hexasome successfully (Kulaeva et al., 2009). Our results showing that positive torsion induced loss of contacts to the distal dimer mirrors *in vitro* results (Sheinin et al., 2013), and confirms *in vivo* that distal dimer loss facilitates RNAPII elongation and that proximal dimer loss is associated with increased stalling. Consistent with increased stalling with loss of the proximal dimer, our depletion results place Spt16 proximal and SSRP distal to the promoter, and Spt16 but not the Pob3 (SSRP) subunit of FACT has a haplo-insufficiency phenotype in budding yeast (Pir et al., 2012). The chromatin remodeler Chd1 was found to slide hexasomes directionally, requiring an H2A/H2B dimer on the DNA entry side (Levendosky et al., 2016). Therefore, upon loss of the distal dimer in the wake of RNAPII, Chd1 would move the hexasome back towards the promoter, which would drive the retrograde movement of nucleosomes in the wake of RNAPII elongation that has been proposed to occur in budding yeast (Weiner et al., 2010).

The rich structural insights that can be obtained from treating nuclease-generated DNA sequencing data as a genome-wide map of protein-DNA interactions is broadly generalizable as illustrated by our uncovering of subnucleosomal protections over the +1 nucleosome of human genes in cfDNA datasets (Snyder et al., 2016). We have shown that fragments mapping within 300 bp of the TSS of genes can be used to separate disease states

from healthy states. This small footprint of the genome used in our analysis to robustly distinguish cancer from normal cfDNA is ~2 orders of magnitude less than the extent of the average transcribed region used as input for the FFT method, which reports on internucleosomal spacing. Although our subnucleosome enrichment method separates disease states somewhat better than the FFT method, the fact that they use largely non-overlapping regions of the gene means that they might be used in concert when whole genome datasets for cfDNA are available. Alternatively, our methodology of analyzing cfDNA data when coupled with targeted sequencing of promoter-proximal regions could dramatically reduce the amount of sequencing required in future applications based on cfDNA.

In cfDNA datasets, there is no experimental control over chromatin fragmentation and we can only analyze the results of genome fragmentation naturally occurring during cell turnover. We have demonstrated that this dataset revealed as much about nucleosomal intermediates formed during transcription as MNase-seq, pointing to the information on chromatin structure that is hidden in cfDNA. Subnucleosomal particles over promoter-proximal nucleosome positions are thus the relics of transcription programs that survive long after the destruction of the cell itself. By revealing the structure of nucleosome intermediates on DNA relics, the method that we introduced to analyze MNase-seq experimental data can potentially be applied to non-invasive disease diagnosis.

STAR METHODS

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Dr. Steven Henikoff (steveh@fredhutch.org).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell lines—*Drosophila melanogaster* S2 cells were grown in HyClone SFX-Insect media supplemented with 18 mM L-glutamine. *Drosophila* S2-DRSC cells (DGRC Stock #181) were grown in Shields and Sang M3 Insect Medium supplemented with 0.25% (w/v) bactopeptone, 0.1% (w/v) yeast extract and 10% FBS.

METHOD DETAILS

MNase-seq—*Drosophila melanogaster* S2 cells were grown in HyClone SFX-Insect media supplemented with 18 mM L-glutamine. Mid-log cells from a 75 mm^2 plate were harvested, centrifuged at 600g for 3 minutes and washed once with cold phosphate-buffered saline. Cells were then resuspended in cold HM2 (10 mM HEPES PH 7.4, 2 mM $MgCl_2$, 0.5 mM PMSF) and incubated for 2 minutes on ice. NP-40 was then added to a final concentration of 0.6%. Nuclei were released by intermittent pipetting for 5 minutes on ice, centrifuged at 100g, 4° C, 10 minutes and washed once with 1 ml of HM2. Nuclei were resuspended with 150 μl of HM2, $CaCl_2$ was added to a final concentration of 1 mM and incubated at 37° C for 5 minutes. Micrococcal nuclease (0.1 U, Sigma) was added and incubated for 5 minutes at 37° C. The reaction was stopped by adding EGTA to a final concentration of 1.5 mM, SDS to a final concentration of 1%, and 20 μg of proteinase K (Thermo Fisher Scientific)

per ml of the reaction. The reaction was incubated at 55° C for 15 minutes. After incubation, phenol-chloroform-isamyl alcohol (PCI) extraction was performed and 20 μg of RNase-A per ml of the aqueous phase was added and the aqueous phase was incubated at 37° C for 30 minutes. PCI extraction was performed again and the DNA in the aqueous phase was precipitated using 70% isopropanol, 30 mM sodium acetate, and glycogen, by centrifuging at 4° C, 16,000 rcf for 30 minutes. The DNA pellet was washed once with 75% ethanol and resuspended in 0.1× TE (10 mM Tris pH 8.0, 1 mM EDTA). Short fragments ~100 bp and smaller were isolated using polyacrylamide gel electrophoresis: the purified DNA was electrophoresed in 8% polyacrylamide gel made with TAE buffer. Using a 10 bp ladder as reference, the portion of the gel encompassing 50–100 bp fragments was sliced out and the DNA from the gel slice was extracted. Both the total DNA and the PAGE-extracted short DNA were sequenced on the Illumina platform.

RNAi treatment and MNase-seq—For double-stranded RNA (dsRNA) treatment, *Drosophila* S2-DRSC cells (DGRC Stock #181) were used. dsRNA amplicons for Spt16 (DRSC08714), SSRP (DRSC29959) and Spt6 (DRSC18836) were obtained from Drosophila RNAi Screening Center. Published dsRNA amplicon data were used for the GFP control (Hamada et al., 2005). dsRNA was added to a final concentration of 10 μg/1×10^6 cells in serum-free media, mixed well, and incubated for 30 minutes at room temperature, followed by supplementation with an equal volume of fresh media and fetal bovine serum to a final concentration of 10%. After 4 days of dsRNA treatment, active chromatin was profiled by treating nuclei with MNase followed by salt extraction and DNA isolation as described (Teves and Henikoff, 2012). Nuclei were released and the MNase reaction was performed as described above. The MNase reaction was stopped with 2 mM EGTA and nuclei were pelleted at 100 rcf for 10 minutes at 4° C, and the supernatant was saved as ‘S1’. The nuclei were resuspended in 80 mM extraction buffer (2 mM $MgCl_2$, 70 mM NaCl, 10 mM Tris-HCl at pH 7.4, 0.1% Triton X-100, and 2 mM EGTA) and incubated at 4° C for two hours with end-over-end mixing. After incubation, nuclei were pelleted at 100 rcf for 10 minutes at 4° C and the supernatant was clarified by centrifuging at 16000 rcf for 2 minutes at 4° C. This supernatant was combined with S1, followed by DNA extraction as described above.

CUT&RUN—1 × 10^6 S2 cells were harvested, centrifuged at 600g for 3 minutes and washed once with PBS. Cells were resuspended in NE1 buffer (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 1 mM $MgCl_2$, 20% Glycerol, 0.1% Triton X-100, 1 Roche complete protease inhibitor tablet per 10 mL just before use) and incubated in ice for 10 minutes and pelleted at 600g for 3 minutes. Cells were then resuspended in Block 1 buffer (100 mM NaCl, 20 mM HEPES-KOH pH 7.9, 5 mM EGTA, 5 mM EDTA, 0.1% Tween20, 1 Roche complete protease inhibitor tablet per 10 mL just before use, 3% bovine serum albumin (BSA)), incubated in ice for 5 minutes, and pelleted at 600g for 3 minutes. Cells were then resuspended in Block 2 buffer (100 mM NaCl, 20 mM HEPES-KOH pH 7.9, 5 mM EGTA, 5 mM EDTA, 0.1% Tween20, 1 Roche complete protease inhibitor tablet per 10 mL just before use, 0.1% BSA), incubated in ice for 5 minutes, and pelleted at 600g for 3 minutes. Cells were resuspended in Block 2 buffer, the primary antibody (antiH3K27ac, Abcam 45173) was added at 1:50 ratio, and incubated on a nutator for 1 hour at 4ºC. The cells were

then washed once with Block 2 buffer and resuspended in the Block 2 buffer. Protein A-MNase fusion (pA-MN) was added at a 1:50 ratio and incubated on a nutator for 1 hour at 4ºC. Cells were pelleted at 600g for 3 minutes and the supernatant was discarded. The cells were resuspended in Digestion buffer (100 mM NaCl, 20 mM HEPES-KOH pH 7.5, 0.1 mM EGTA, 1 Roche complete protease inhibitor tablet per 10 mL just before use), incubated for 5 minutes at room temperature, and then calcium chloride was added to a final concentration of 2 mM and the cell suspension was vortexed. The nuclease reaction was stopped after 1–5 minutes by adding EDTA and EGTA to a final concentration of 20 mM and 16mM respectively and set on ice. The cell suspension was treated with RNAse A for 30 minutes at 37ºC, then the DNA extraction buffer (300 mM NaCl, 20 mM EGTA, 10 mM EDTA, 10 mM Tris-HCl ph 8.1, 1% sodium dodecyl sulfate) and proteinase K was added and incubated at 55ºC for 1 hour. PCI extraction was performed and then the aqueous phase was mixed with half the volume of Agencourt Ampure XP beads, held 5–10 min, placed on a magnet stand, and the supernatant was retained, discarding the beads. The supernatant was mixed with ethanol at 75% and spun at 16,000g for 30 minutes to precipitate the DNA. The DNA was used for library preparation according to the published protocol (Skene and Henikoff, 2017). Briefly, libraries were constructed without size-selection, following the KAPA DNA polymerase library preparation kit protocol (https://www.kapabiosystems.com/product-applications/products/next-generation-sequencing-2/dna-library-preparation/kapa-hyper-prep-kits/). To reduce the representation of the remaining large fragments, the number of PCR cycles using the KAPA polymerase library preparation method was increased to 14 cycles.

Sequencing—Cluster generation and 25 rounds of paired-end sequencing were performed by the FHCRC Genomics Shared Resource with an Illumina Hi-Seq 2500. Processing and base-calling were performed by Illumina Eland program. Paired-end sequencing data were aligned with Novoalign (Novocraft; http://www.novocraft.com) against dmel_r5.51. All further analyses of genome-aligned paired-end data are described in the Quantification and Statistical Analysis section.

QUANTIFICATION AND STATISTICAL ANALYSIS

Analysis of fragment center distributions—The nucleosome positions were called using a custom perl script and the +1 position was defined as the first position from the TSS. The +1 positions were further filtered to include only those that were at least 50 bp from the TSS and at most 250 bp from the TSS. The perl script is available at : https://github.com/srinivasramachandran/CallNucleosomes

For each length class, the distribution of fragment centers was generated relative to the +1 nucleosome position and then smoothed by generating a running average with a window of 20 bp. This distribution was then fitted with a single and a sum of two Gaussian functions in R using the following formula:

$$Single\ Gaussian\ Curve: y = a_1 \cdot \exp\left(\frac{-1 \cdot (x - b_1)^2}{c_1^2}\right) + d$$

$$Two\ Gaussian\ Curves: y = a_1 \cdot \exp\left(\frac{-1 \cdot (x - b_1)^2}{c_1^2}\right) + a_2 \cdot \exp\left(\frac{-1 \cdot (x - b_2)^2}{c_2^2}\right) + d$$

We observed that the distribution of 90, 103 and 112 bp fragments featured significant shoulders that were well accounted for by the sum of two Gaussian curves to yield a left peak and a right peak (Figure S1G). The left peak corresponds to loss of DNA contacts with the H2A/H2B dimer distal to the TSS, and the right peak corresponds to loss of contacts between DNA and H2A/H2B dimer proximal to the TSS. Thus, the ratio of the area under the curve of the left peak to the area under the curve of the right peak represents the ratio of proximal dimer loss to distal dimer loss:

$$ratio = \frac{a_1 \cdot c_1}{a_2 \cdot c_2}$$

For each gene, the cross-correlation was calculated between the fragment center distributions of 147±5 bp fragments and subnucleosomal fragments (90±5 or 103±5 or 112±5 bp fragments) in the region ±80 bp of the +1 nucleosome center with shifts from −80 to +80 bp using R. As the peak centers were at ~±25 bp from the dyad, the cross-correlation asymmetry was calculated as follows:

$$cross correlation\ asymmetry = \sum_{i=-50}^{i=-1} r_i - \sum_{i=1}^{i=50} r_i$$

where r_i is the cross-correlation at i^{th} shift. Higher cross-correlation asymmetry values indicate a larger population on the distal side of the dyad and *vice versa*.

A set of 106 genes was obtained from Flybase under the gene ontology term “response to heat” and cross-referenced to our list of genes with RNAPII and a defined +1 nucleosome in S2 cells. This resulted in a list of 41 genes. Analyses in Figure 4F,G were performed with genes in this list that had subnucleosomal fragments.

Analysis of cell free-DNA datasets—Gene definitions were downloaded from Ensembl Biomart (http://www.ensembl.org/biomart/martview/), GRCh38.p7 assembly. Only genes that had no other genes overlapping the TSS ± 1000 bp window were considered. Fastq files of cfDNA datasets were downloaded from the SRA database and aligned to UCSC hg38 using bowtie2 with following parameters: --no-mixed --no-discordant --no-unal. After whole genome alignment, only reads aligning to the TSS ± 5000 bp of non-overlapping genes were subsetted for further analysis, the read count of which is listed in Table S4. Nucleosome positions were called using a custom perl script (https://github.com/srinivasramachandran/CallNucleosomes) and the +1 nucleosome position was defined as the first nucleosome position downstream of the TSS that is between the TSS and the TSS + 300 bp. For generating ROC curves, we first calculated the ratio of the number of 40–100 bp fragments to the number of 155–170 bp fragments over the dyad ±50 bp of +1 nucleosomes at each gene with a called +1 position (‘subnucleosome enrichment’). RNA-seq gene expression

data (Fragments per kilobase of transcript per million mapped reads (FPKM) values) for 44 human cell lines and 43 primary tissues by the Human Protein Atlas (Uhlen et al., 2015; Uhlen et al., 2010) were used. We sorted genes according to their subnucleosome enrichment and smoothed the subnucleosome enrichment and the corresponding expression levels in different cell lines and tissues with a 200 gene window. The Pearson correlation coefficient was then calculated between the smoothed subnucleosome enrichment and the smoothed gene expression values from each cell line and tissue ('subnucleosome correlation'). The celllines and tissues were sorted in ascending order based on the subnucleosome correlation. Lymphoid/myeloid cell lines and tissues were set as true positives to generate the ROC curve. To compare samples at equal numbers of reads, 10,000,000 reads were randomly sampled 100 times from each dataset (using the unix command shuf) and the correlation analysis was performed for each iteration followed by calculation of the area under the ROC curve (AUC). Then the 100 AUC values for each sample were used for generating the box plots shown in Figure 7F.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Christine Codomo for preparing Illumina sequencing libraries, Jorja Henikoff for bioinformatics, Erika Feutz for technical assistance, the Fred Hutch Genomics Shared Resource for sequencing, and Paul Talbert for comments on the manuscript.

References

Arimura Y, Tachiwana H, Oda T, Sato M, and Kurumizaka H (2012). Structural analysis of the hexasome, lacking one histone H2A/H2B dimer from the conventional nucleosome. Biochem 51, 3302–3309. [PubMed: 22448809]

Azegami N, Saikusa K, Todokoro Y, Nagadoi A, Kurumizaka H, Nishimura Y, and Akashi S (2013). Conclusive evidence of the reconstituted hexasome proven by native mass spectrometry. Biochem 52, 5155–5157. [PubMed: 23879667]

Baer BW, and Rhodes D (1983). Eukaryotic RNA polymerase II binds to nucleosome cores from transcribed genes. Nature 301, 482–488. [PubMed: 6823327]

Bondarenko VA, Steele LM, Ujvari A, Gaykalova DA, Kulaeva OI, Polikanov YS, Luse DS, and Studitsky VM (2006). Nucleosomes can form a polar barrier to transcript elongation by RNA polymerase II. Mol Cell 24, 469–479. [PubMed: 17081995]

Bortvin A, and Winston F (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. Science 272, 1473–1476. [PubMed: 8633238]

Cole HA, Ocampo J, Iben JR, Chereji RV, and Clark DJ (2014). Heavy transcription of yeast genes correlates with differential loss of histone H2B relative to H4 and queued RNA polymerases. Nucleic Acids Res 42, 12512–12522. [PubMed: 25348398]

Formosa T (2013). The role of FACT in making and breaking nucleosomes. Biochim Biophys Acta 1819, 247–255. [PubMed: 24459727]

Gaykalova DA, Nagarajavel V, Bondarenko VA, Bartholomew B, Clark DJ, and Studitsky VM (2011). A polar barrier to transcription can be circumvented by remodeler-induced nucleosome translocation. Nucleic Acids Res 39, 3520–3528. [PubMed: 21245049]

Hamada FN, Park PJ, Gordadze PR, and Kuroda MI (2005). Global regulation of X chromosomal genes by the MSL complex in Drosophila melanogaster. Genes Dev 19, 2289–2294. [PubMed: 16204180]

Henikoff JG, Belsky JA, Krassovsky K, MacAlpine DM, and Henikoff S (2011). Epigenome characterization at single base-pair resolution. Proc Natl Acad Sci U S A 108, 18318–18323. [PubMed: 22025700]

Henikoff S, Henikoff JG, Sakai A, Loeb GB, and Ahmad K (2009). Genome-wide profiling of salt fractions maps physical properties of chromatin. Genome Res 19, 460–469. [PubMed: 19088306]

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. [PubMed: 19644123]

Hsieh FK, Kulaeva OI, Patel SS, Dyer PN, Luger K, Reinberg D, and Studitsky VM (2013). Histone chaperone FACT action during transcription through chromatin by RNA polymerase II. Proc Natl Acad Sci U S A 110, 7654–7659. [PubMed: 23610384]

Huang C, Zhang Z, Xu M, Li Y, Li Z, Ma Y, Cai T, and Zhu B (2013). H3.3-h4 tetramer splitting events feature cell-type specific enhancers. PLoS genetics 9, e1003558. [PubMed: 23754967]

Jonkers I, Kwak H, and Lis JT (2014). Genome-wide dynamics of Pol II elongation and its interplay with promoter proximal pausing, chromatin, and exons. Elife 3, e02407. [PubMed: 24843027]

Kemble DJ, McCullough LL, Whitby FG, Formosa T, and Hill CP (2015). FACT Disrupts Nucleosome Structure by Binding H2A-H2B with Conserved Peptide Motifs. Mol Cell 60, 294–306. [PubMed: 26455391]

Kimura H, and Cook PR (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. J Cell Biol 153, 1341–1353. [PubMed: 11425866]

Kireeva ML, Walter W, Tchernajenko V, Bondarenko V, Kashlev M, and Studitsky VM (2002). Nucleosome remodeling induced by RNA polymerase II: loss of the H2A/H2B dimer during transcription. Mol Cell 9, 541–552. [PubMed: 11931762]

Kulaeva OI, Gaykalova DA, Pestov NA, Golovastov VV, Vassylyev DG, Artsimovitch I, and Studitsky VM (2009). Mechanism of chromatin remodeling and recovery during passage of RNA polymerase II. Nat Struct Mol Biol 16, 1272–1278. [PubMed: 19935686]

Kulaeva OI, Hsieh FK, and Studitsky VM (2010). RNA polymerase complexes cooperate to relieve the nucleosomal barrier and evict histones. Proc Natl Acad Sci U S A 107, 11325–11330. [PubMed: 20534568]

Kumar S, and Leffak M (1986). Assembly of active chromatin. Biochem 25, 2055–2060. [PubMed: 3707932]

Kuryan BG, Kim J, Tran NN, Lombardo SR, Venkatesh S, Workman JL, and Carey M (2012). Histone density is maintained during transcription mediated by the chromatin remodeler RSC and histone chaperone NAP1 in vitro. Proc Natl Acad Sci U S A 109, 1931–1936. [PubMed: 22308335]

Lam WKJ, Gai W, Sun K, Wong RSM, Chan RWY, Jiang P, Chan NPH, Hui WWI, Chan AWH, Szeto CC, et al. (2017). DNA of Erythroid Origin Is Present in Human Plasma and Informs the Types of Anemia. Clinical chemistry 63, 1614–1623. [PubMed: 28784691]

Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359. [PubMed: 22388286]

Lee MS, and Garrard WT (1991). Positive DNA supercoiling generates a chromatin conformation characteristic of highly active genes. Proc Natl Acad Sci U S A 88, 9675–9679. [PubMed: 1946386]

Levendosky RF, Sabantsev A, Deindl S, and Bowman GD (2016). The Chd1 chromatin remodeler shifts hexasomes unidirectionally. Elife 5, e21356. [PubMed: 28032848]

Li M, and Wang MD (2012). Unzipping single DNA molecules to study nucleosome structure and dynamics. Methods Enzymol 513, 29–58. [PubMed: 22929764]

Nechaev S, Fargo DC, dos Santos G, Liu L, Gao Y, and Adelman K (2010). Global analysis of short RNAs reveals widespread promoter-proximal stalling and arrest of Pol II in Drosophila. Science 327, 335–338. [PubMed: 20007866]

Pir P, Gutteridge A, Wu J, Rash B, Kell DB, Zhang N, and Oliver SG (2012). The genetic control of growth rate: a systems biology study in yeast. BMC systems biology 6, 4. [PubMed: 22244311]

Ramachandran S, and Henikoff S (2016). Transcriptional Regulators Compete with Nucleosomes Post-replication. Cell 165, 580–592. [PubMed: 27062929]

Rocha E, Davie JR, van Holde KE, and Weintraub H (1984). Differential salt fractionation of active and inactive genomic domains in chicken erythrocyte. J Biol Chem 259, 8558–8563. [PubMed: 6736042]

Sanders MM (1978). Fractionation of nucleosomes by salt elution from micrococcal nuclease-digested nuclei. J Cell Biol 79, 97–109. [PubMed: 701381]

Sheinin MY, Li M, Soltani M, Luger K, and Wang MD (2013). Torque modulates nucleosome stability and facilitates H2A/H2B dimer loss. Nature communications 4, 2579.

Skene PJ, and Henikoff S (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife 6.

Snyder MW, Kircher M, Hill AJ, Daza RM, and Shendure J (2016). Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. Cell 164, 57–68. [PubMed: 26771485]

Teves SS, and Henikoff S (2011). Heat shock reduces stalled RNA polymerase II and nucleosome turnover genome-wide. Genes Dev 25, 2387–2397. [PubMed: 22085965]

Teves SS, and Henikoff S (2012). Salt fractionation of nucleosomes for genome-wide profiling. Methods Mol Biol 833, 421–432. [PubMed: 22183608]

Teves SS, and Henikoff S (2014). Transcription-generated torsional stress destabilizes nucleosomes. Nat Struct Mol Biol 21, 88–94. [PubMed: 24317489]

Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson A, Kampf C, Sjostedt E, Asplund A, et al. (2015). Proteomics. Tissue-based map of the human proteome. Science 347, 1260419. [PubMed: 25613900]

Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S, et al. (2010). Towards a knowledge-based Human Protein Atlas. Nat Biotechnol 28, 1248–1250. [PubMed: 21139605]

Underhill HR, Kitzman JO, Hellwig S, Welker NC, Daza R, Baker DN, Gligorich KM, Rostomily RC, Bronner MP, and Shendure J (2016). Fragment Length of Circulating Tumor DNA. PLoS genetics 12, e1006162. [PubMed: 27428049]

Weber CM, Henikoff JG, and Henikoff S (2010). H2A.Z nucleosomes enriched over active genes are homotypic. Nat Struct Mol Biol 17, 1500–1507. [PubMed: 21057526]

Weber CM, Ramachandran S, and Henikoff S (2014). Nucleosomes are context-specific, H2A.Z-modulated barriers to RNA polymerase. Mol Cell 53, 819–830. [PubMed: 24606920]

Weiner A, Hughes A, Yassour M, Rando OJ, and Friedman N (2010). High-resolution nucleosome mapping reveals transcription-dependent promoter packaging. Genome Res 20, 90–100. [PubMed: 19846608]

Williamson R (1970). Properties of rapidly labelled deoxyribonucleic acid fragments isolated from the cytoplasm of primary cultures of embryonic mouse liver cells. Journal of molecular biology 51, 157–168. [PubMed: 5481278]

Highlights:

- MNase-seq can identify and precisely map nucleosomal intermediates in transcription □
- Transcription elongation causes loss of contacts to promoter-distal H2A-H2B *in vivo* □
- FACT is oriented over the +1 nucleosome and prevents H2A-H2B dimer loss *in vivo*
- Subnucleosome mapping over +1 identifies relics of transcription in cell free-DNA

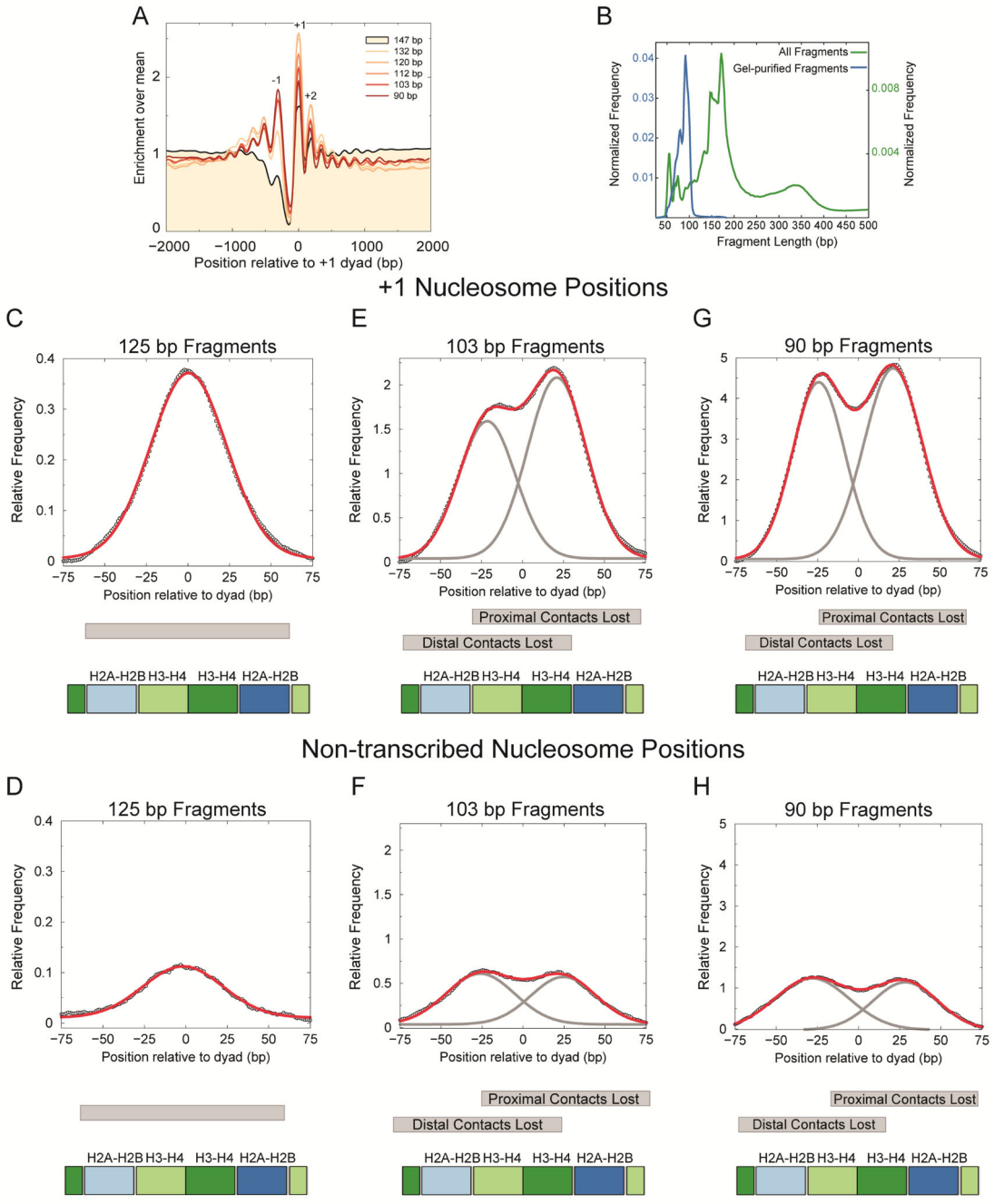


Figure 1. Asymmetric loss of histone-DNA contacts over the transcribed +1 nucleosome.

(**A**) Enrichment of different fragment lengths that are protected by MNase over the mean signal at the +1 nucleosome position ± 2000 bp plotted relative to the +1 nucleosome position. The distribution was generated from fragments that mapped to expressed genes with a defined +1 position (n=5273). Profiles are averages over a 50-bp sliding window. (**B**) Distribution of fragment lengths from a MNase-seq experiment (All Fragments) and after selection of <100 bp fragments from PAGE-purification (Gel-purified Fragments). (**C**) Distribution of 125 ± 5 bp fragment centers (open circles) plotted relative to the nucleosome

centers of expressed genes (n=5273), over which the Gaussian fit for the distribution is plotted in red. (**D**) Same as (C) for highly positioned, non-transcribed nucleosome positions selected randomly (n=5273). (**E**) Same as (C) for 103 ± 5 bp fragments. The gray curves represent the individual Gaussian functions that make up the bimodal fit plotted in red. (**F**) Same as (D) for 103 ± 5 bp fragments. (**E**) Same as (C) for 90 ± 5 bp fragments. (**G**) Same as (E) for 90 ± 5 bp fragments. Profiles are averages over a 20-bp sliding window. The gray bars below the plots represent the protection of the fragment represented by the mean position of the Gaussian distributions. The map of DNA contacts to successive histone dimers in the nucleosome structure is represented below the gray bars on the same X-axis scale as the plots. See also Figure S1, S2, and S3; Movie S1.

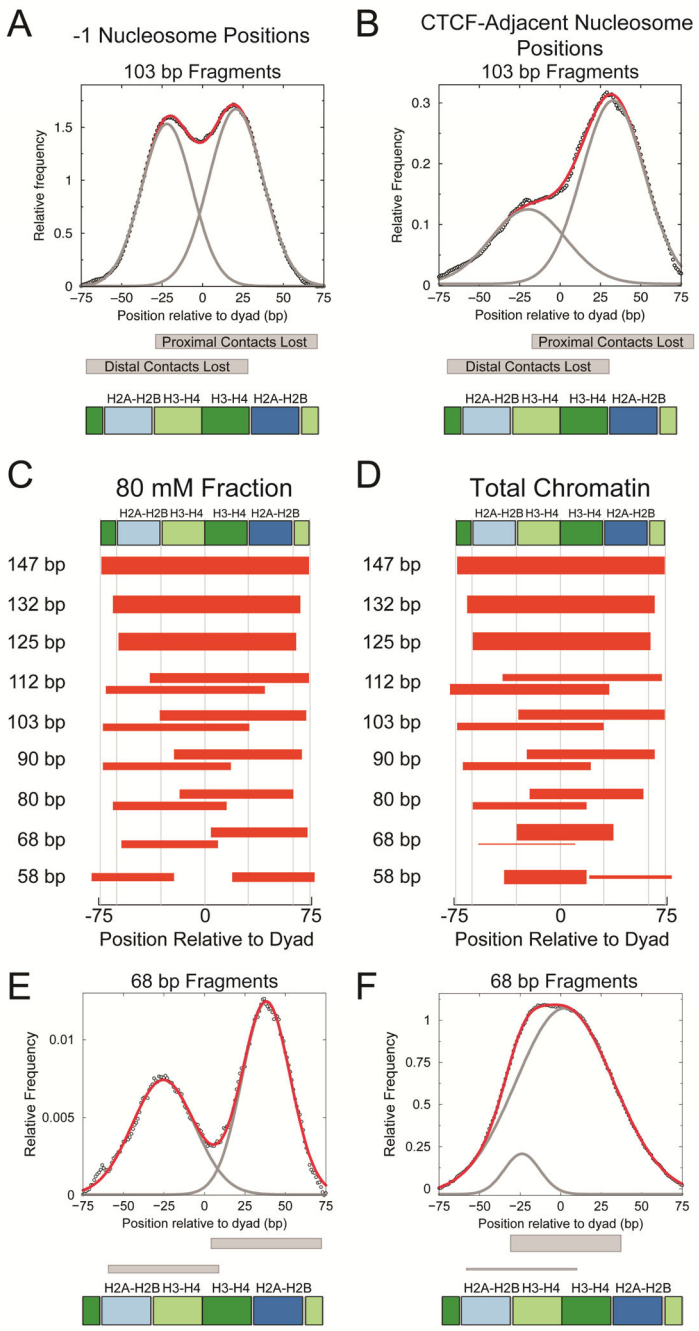


Figure 2. Subnucleosome protections in active chromatin.

A) Distribution of 103 ± 5 bp fragment centers (open circles) plotted relative to the −1 nucleosome centers of expressed Drosophila genes (n=3773), over which the Gaussian fit for the distribution is plotted in red. The gray curves represent the individual Gaussian functions that make up the bimodal fit plotted in red. **B**) Distribution of 103 ± 10 bp fragment centers (open circles) plotted relative to the −1 nucleosome centers relative to CTCF CUT&RUN sites in human K562 cells (n=15815). The distribution is obtained by subtracting the adjacent CTCF peaks. The Gaussian fit for the distribution is plotted in red. The gray curves

represent the individual Gaussian functions that make up the bimodal fit plotted in red. **C**) Histone dimers that form contacts with different regions of nucleosomal DNA in the nucleosome structure are mapped at the top. For each length class, the mean of the Gaussian distribution that fits the fragment distribution from low-salt-soluble chromatin is used to center the representative fragment. The fragment length corresponds to the mean length shown on the left. For 112, 103, 90, 80, 68, and 58 bp length classes, the distribution of fragment centers was fit as the sum of two Gaussian functions and the thickness of the representative fragments shown here are scaled to reflect the relative areas under the curve of each Gaussian function. **D**) Same as A, but for total chromatin. **E**) Distribution of 68±5 bp fragment centers (open circles) plotted relative to the nucleosome centers of expressed genes (n=5273) for low-salt-soluble chromatin, over which the Gaussian fit for the distribution is plotted in red. The gray curves represent the individual Gaussian functions that make up the bimodal fit plotted in red. **F**) Same as (C) for total chromatin.

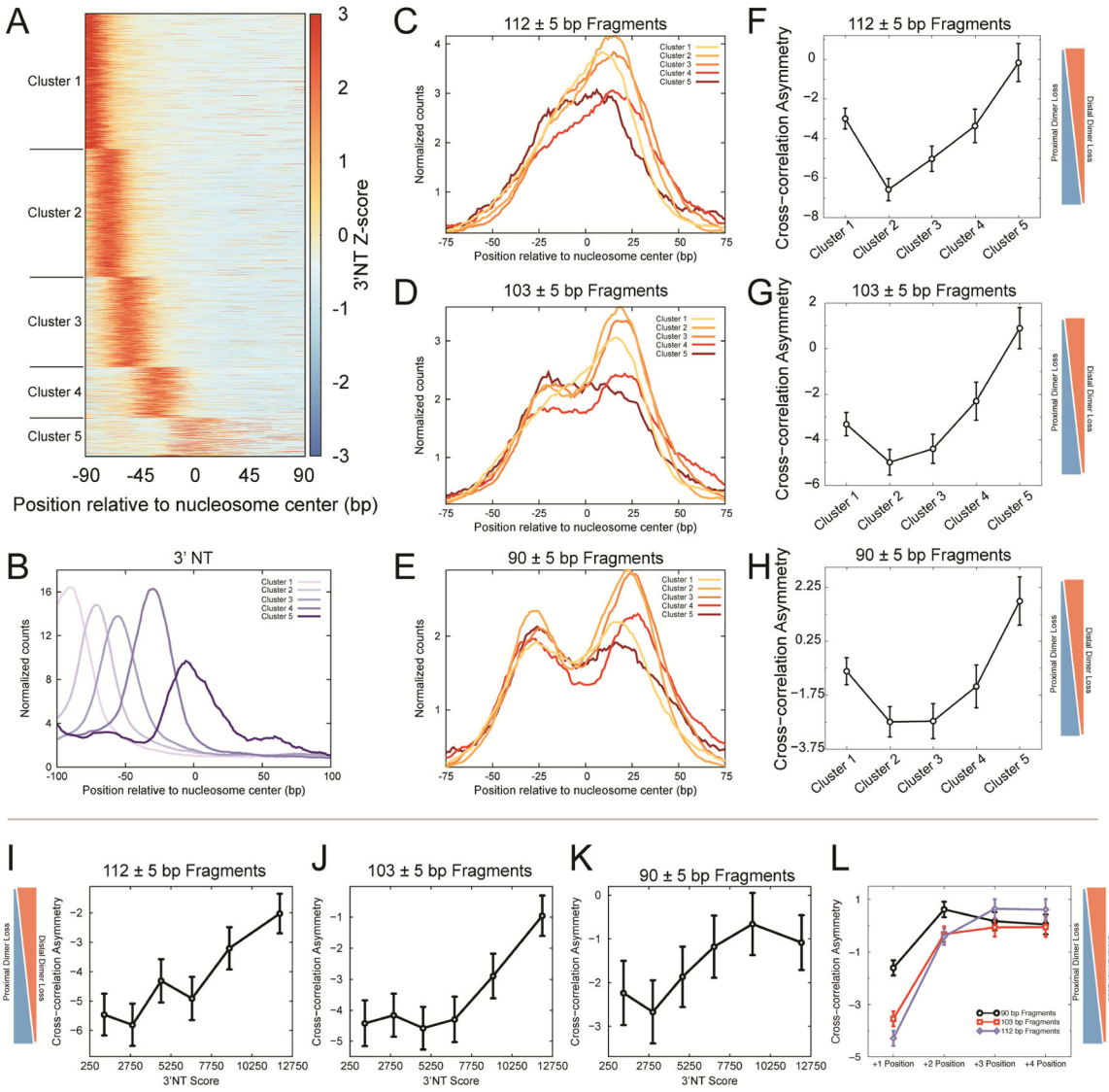


Figure 3. RNAPII drives subnucleosome production at the +1 nucleosome.

(A) Heatmap of 3'NT Z-scores relative to the +1 nucleosome position (n=5269). The Z-scores are calculated for each gene over a window of dyad±100 bp after smoothing the 3'NT profile using a 20 bp running average. The genes are divided into five clusters obtained using k-means clustering. (**B**) 3'NT normalized counts averaged over the genes in each cluster relative to the +1 nucleosome position. Profiles are averages over a 20-bp sliding window. (**C**) Distribution of 112±5 bp fragment centers for each of the five 3'NT clusters shown in (A) plotted relative to the +1 nucleosome center. Profiles are averages over a 20-bp sliding window. (**D**) Same as (C) for 103±5 bp fragment centers. (**E**) Same as (C) for 90±5 bp fragment centers. (**F**) Mean ± S.E.M. of the cross-correlation asymmetry of 112±5 bp fragments versus 147±5 fragments around the +1 nucleosome position for genes of each 3'NT cluster. (**G**) Same as (F) for 103±5 bp fragments. (**H**) Same as (F) for 90±5 bp fragments. (**I**) Genes were divided into six groups based on total 3'NT signal at the +1

nucleosome position ±90 bp. Mean ± S.E.M. of the cross-correlation asymmetry of 112±5 bp fragments versus 147±5 fragments around the +1 nucleosome position for genes in each group is plotted. (**J**) Same as (I) for 103±5 bp fragments. (**K**) Same as (I) for 90±5 bp fragments. (**L**) Mean ± S.E.M. of the cross-correlation asymmetry of subnucleosomal fragments versus 147±5 fragments around each nucleosome position for all expressed genes are plotted.

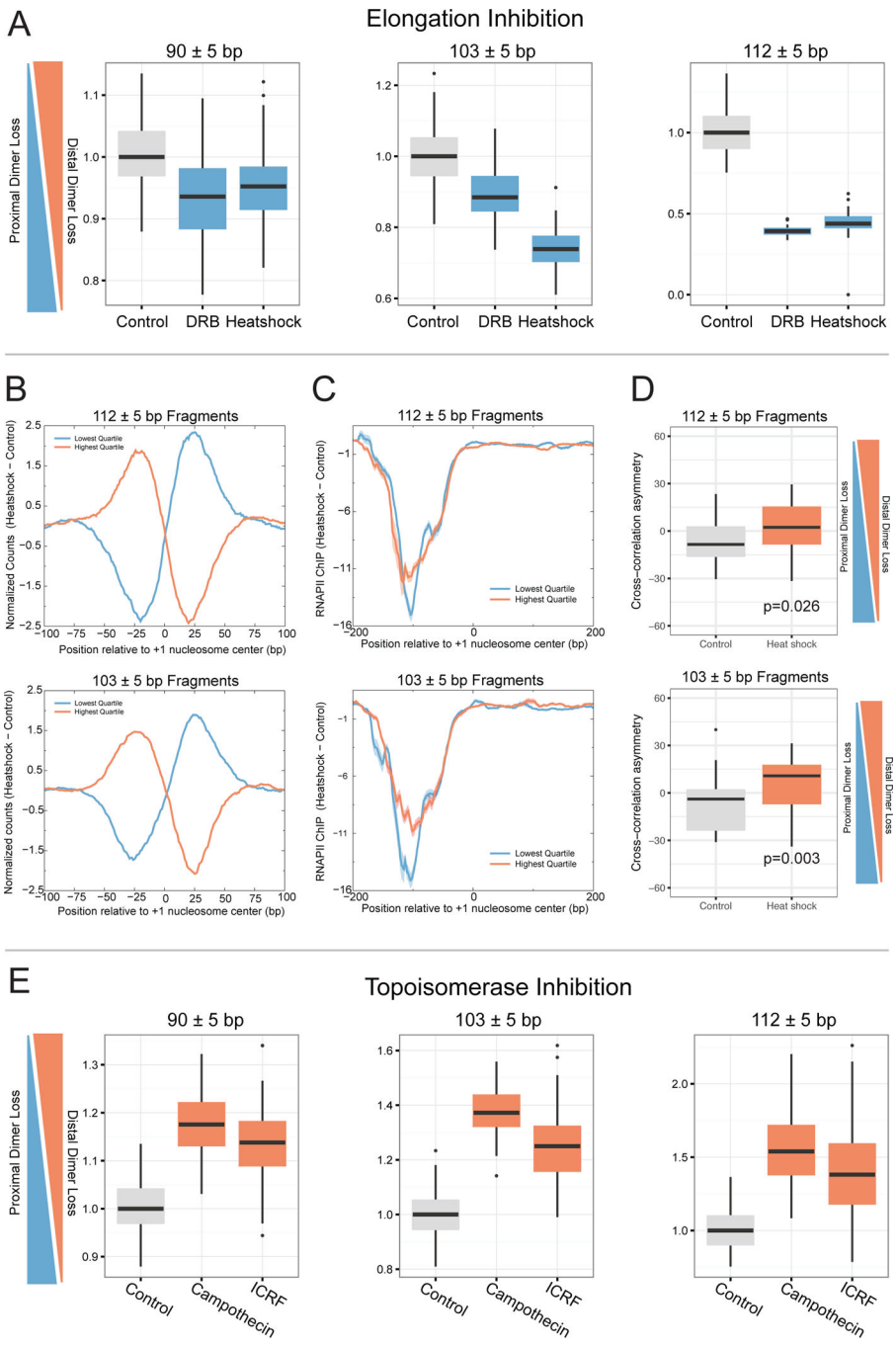


Figure 4. RNAPII elongation preferentially drives loss of contacts to the distal dimer.

(**A**) Ratio of loss of contacts to the distal dimer over the proximal dimer of the 90±5, 103±5, and 112±5 bp fragment center distributions calculated after resampling same number of fragments in each length class from the Control, DRB, and heat-shock datasets. (**B**) The 112±5 bp (top) and 103±5 bp (bottom) fragment center distribution of the control sample subtracted from the heat-shock sample, showing lowest and highest quartiles of loss of contacts to the distal dimer upon heat-shock. (**C**) Average RNAPII ChIP-seq enrichment of the control sample subtracted from the heat-shock sample plotted over the +1 nucleosome

position of genes in the highest and lowest quartiles of loss of contacts to the distal dimer upon heat shock, as defined by 112±5 bp fragments (top) and 103±5 bp fragments (bottom). (**D**) Cross-correlation asymmetry of genes responding to heat-shock in the control sample and the heat-shock sample for 112±5 bp fragments (top, n=28) and 103±5 bp fragments (bottom, n=29). P-values are from paired Wilcoxon tests. (**H**) Same as (A) for topoisomerase inhibition datasets. The p-values for comparisons of treatments vs. control are listed in Table S1. See also Figure S4.

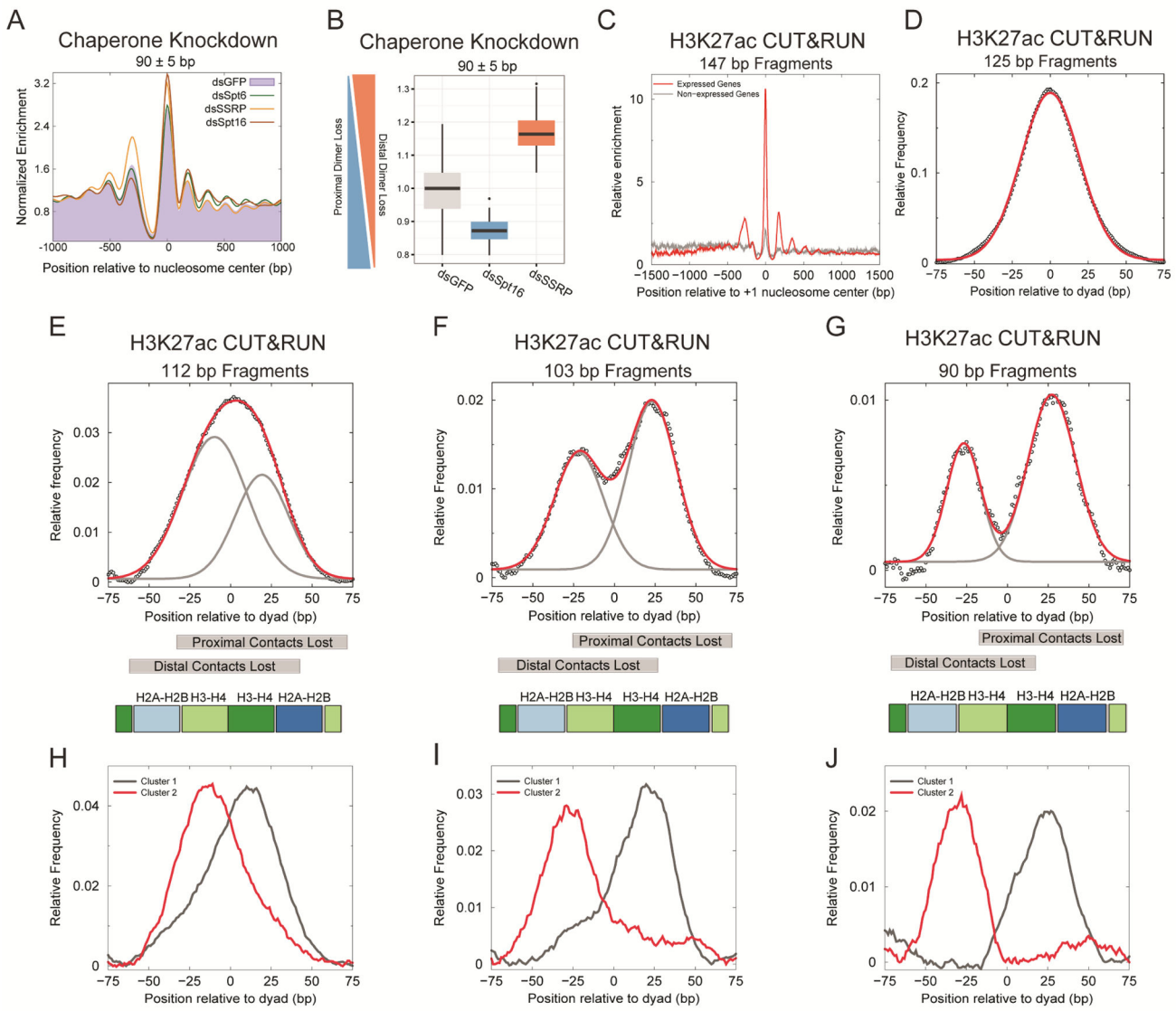
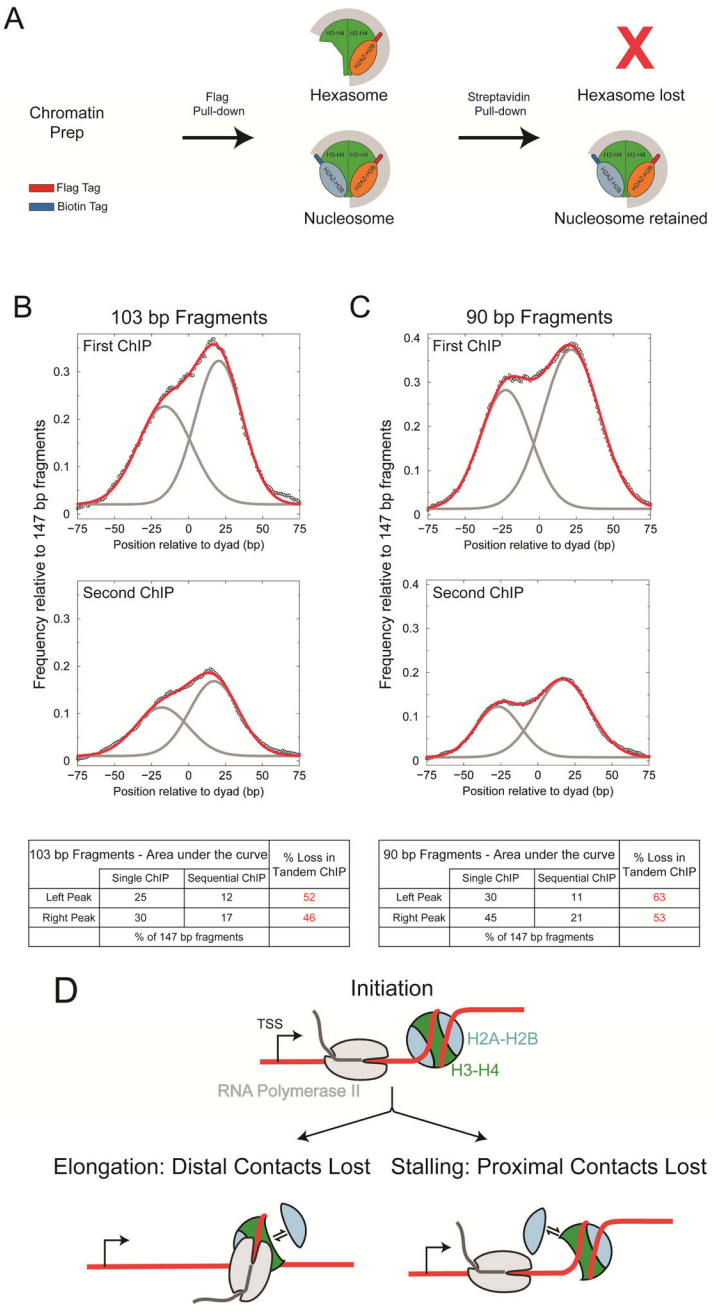


Figure 5. Chaperones protect nucleosomes during transcription.

(**A**) Enrichment of 90±5 bp fragments over the mean signal at the +1 nucleosome position ± 5000 bp plotted relative to the +1 nucleosome position. The distribution was generated for a list of expressed genes with a defined +1 position and 90 bp fragment coverage in all four datasets (n=2713). Profiles are averages over a 50-bp sliding window. (**B**) Ratio of distal over proximal dimer loss of the 90±5 bp fragment center distributions calculated after resampling same number of fragments in the length class from the Control (dsGFP), Spt16, and SSRP RNAi. The p-values for comparisons of treatments vs. control are listed in Table S2. (**C**) Enrichment of centers of 147±5 bp fragments from H3K27ac CUT&RUN experiments relative to the +1 nucleosome position. The distribution was generated from fragments that mapped to expressed genes (n=5273) and non-expressed genes (n=3091). Profiles are averages over a 20-bp sliding window. (**D**) Distribution of 125±5 bp fragment centers (open circles) plotted relative to the nucleosome centers of expressed genes

(n=5273), over which the Gaussian fit for the distribution is plotted in red. (**E**) Same as (D) for 112±5 bp fragments. The gray curves represent the individual Gaussian functions that make up the bimodal fit plotted in red. (**F**) Same as (E) for 103±5 bp fragments. (**G**) Same as (E) for 90±5 bp fragments. (**H**) Distribution of 112±5 bp fragment centers relative to the +1 nucleosome position for genes belonging to two clusters as defined in Figure S3 (**I**) Same as (F) for 103±5 bp fragments (**J**) Same as (F) for 90±5 bp fragments. See also Table S2.



103 bp Fragments - Area under the curve			% Loss in Tandem ChIP
	Single ChIP	Sequential ChIP	
Left Peak	25	12	52
Right Peak	30	17	46
	% of 147 bp fragments		

90 bp Fragments - Area under the curve			% Loss in Tandem ChIP
	Single ChIP	Sequential ChIP	
Left Peak	30	11	63
Right Peak	45	21	53
	% of 147 bp fragments		

Figure 6. Subnucleosomal particles are hexasomes and unwrapped octamers.

(**A**) Experimental schematic of the sequential ChIP protocol that can distinguish hexasomes from unwrapped octamers. (**B**) Distribution of the centers of 103±5 bp fragments from the first ChIP (top) and the second ChIP (middle); tabulation of the areas under the curve for the two Gaussian functions (left peak and right peak) normalized to the area under the curve of the Gaussian fit of centers of 147±5 bp fragments (bottom). (**C**) Same as (B) for 90±5 bp fragments. Fragment center profiles were plotted for expressed genes with a defined +1 nucleosome position (n=5273) averaged over a 20 bp sliding window. (**D**) Model for

nucleosomal intermediates formed during transcription. RNAPII elongation leads to distal dimer loss and/or loss of contacts to the distal dimer, whereas RNAPII stalling at the nucleosome entry site leads to proximal dimer loss and/or loss of contacts to the proximal dimer. See also Figures S5, S6, and S7.

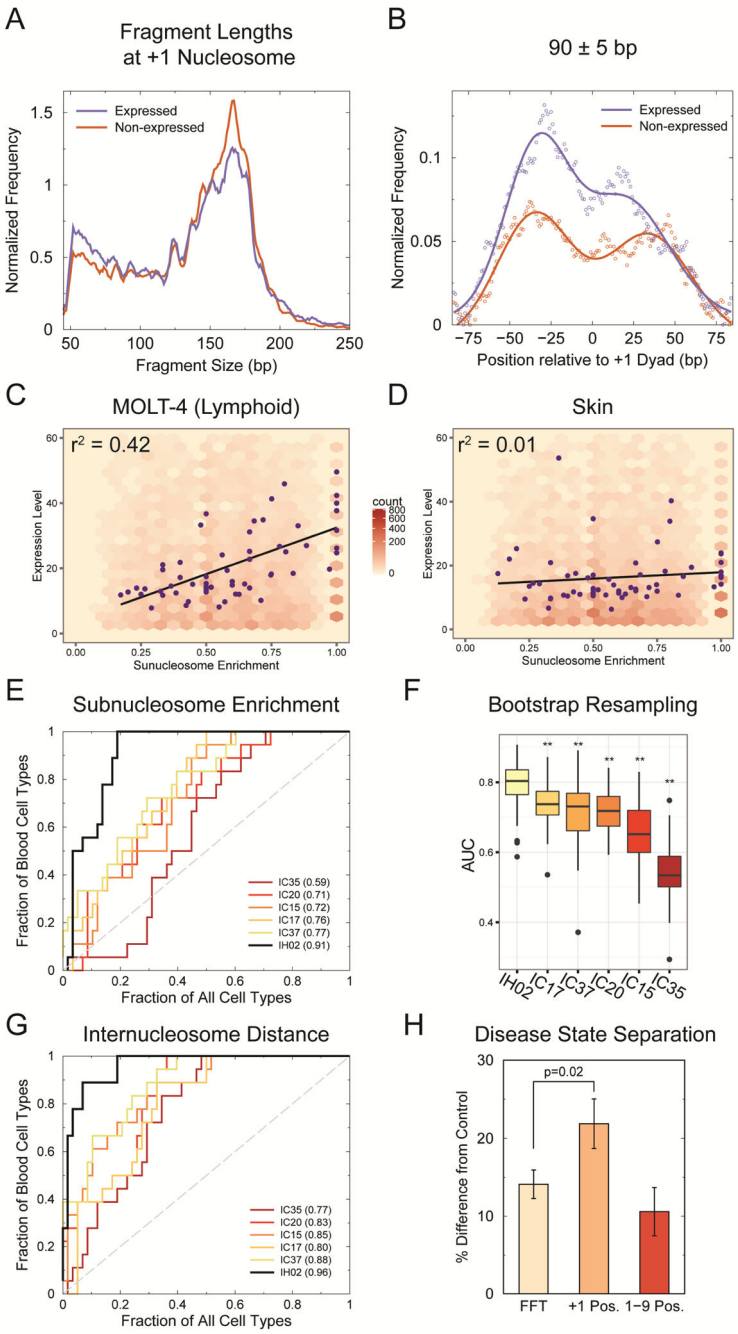


Figure 7. Cell free-DNA features subnucleosomal protections that correlate with transcription.

A) Length distribution of fragments whose centers are within the +1 dyad ± 50 bp. The top 1500 genes expressed in lymphoid/myeloid cell lines/tissues have a lower frequency of nucleosomal fragments and a higher frequency of shorter, subnucleosomal fragments compared to non-expressed genes. **B**) Distribution of centers of 92±10 bp fragments relative to the +1 dyad. A bimodal Gaussian curve fits the fragment center distribution, implying formation of hexasomes with either proximal or distal H2A/H2B dimer loss. The top 1500 genes expressed in lymphoid/myeloid cell lines/tissues have a higher frequency of 92±10 bp

fragments, especially corresponding to loss of contacts with the H2A/H2B dimer distal to the promoter, compared to non-expressed genes. **C**) Gene expression (Fragments per kilobase of transcript per million mapped reads (FPKM) values) of MOLT-4, a lymphoid-derived cell line, plotted against the subnucleosome enrichment as a 2D histogram with hexagonal binning. The blue points represent the average values of successive 200 gene bins ranked by subnucleosome enrichment. The black line is the line of best fit for the blue points. **D**) Same as (C) for expression levels of skin tissue. **E**) Receiver operator characteristic (ROC) curves are plotted for cfDNA from a healthy individual (IH02) and from individuals diagnosed with cancer (IC15, IC17, IC20, IC35, and IC37). The legend indicates the AUC for each sample. **F**) To ensure that comparisons have equal depth of sequencing, a fraction of reads was randomly chosen from all samples 100 times and the AUC calculated with the chosen reads. Box plots represent the distribution of AUC from 100 shuffles. P-values of comparison of cancer states with IH02 are listed in Table S3. **G**) Same as (E) using correlation coefficients of FFT intensity and expression levels obtained from the Supplement of (Snyder et al., 2016). **H**) Mean ± standard error of the mean plotted for the percentage difference between AUC of disease states and AUC of the healthy state. The p-value for the comparison of FFT and +1 position was calculated using paired-one-tailed student's t-test (n=5). The p-values for comparisons and read counts of different samples are listed in Table S3 and S4.