

METHOD

Open Access

# Quantitative MNase-seq accurately maps nucleosome occupancy levels



Răzvan V. Chereji<sup>1</sup>, Terri D. Bryson<sup>2</sup> and Steven Henikoff<sup>2\*</sup>

## Abstract

Micrococcal nuclease (MNase) is widely used to map nucleosomes. However, its aggressive endo-/exo-nuclease activities make MNase-seq unreliable for determining nucleosome occupancies, because cleavages within linker regions produce oligo- and mono-nucleosomes, whereas cleavages within nucleosomes destroy them. Here, we introduce a theoretical framework for predicting nucleosome occupancies and an experimental protocol with appropriate spike-in normalization that confirms our theory and provides accurate occupancy levels over an MNase digestion time course. As with human cells, we observe no overall differences in nucleosome occupancies between *Drosophila* euchromatin and heterochromatin, which implies that heterochromatic compaction does not reduce MNase accessibility of linker DNA.

**Keywords:** Chromatin, Linker DNA, Nuclease digestion kinetics

## Introduction

Nucleosomes are the basic units of DNA compaction and the fundamental constituents of chromatin [1]. They contain 147 bp of DNA, wrapped around a histone octamer comprising two copies of the four core histones: H2A, H2B, H3, and H4 [2]. DNA binding of multiple proteins is generally hindered by nucleosomes, and nucleosome positions and levels of occupancy are a common determinant of gene regulation [3–6]. Therefore, nucleosome occupancy and accessibility are important aspects of chromatin organization, with direct implications for gene regulation and all other DNA-related processes in the cell nucleus. Most yeast genes and active genes in higher eukaryotes share a stereotypical nucleosome organization of their promoters: a nucleosome-depleted region (NDR) near the transcription start site (TSS), flanked by regular arrays of nucleosomes [6–9]. This stereotypical organization has been explained by the presence of barrier complexes that bind to promoters and create an energy barrier preventing nucleosome formation [10, 11], but other determinants of nucleosome positions are also known, e.g., chromatin remodelers and DNA sequence [12].

Different cells can have different nucleosome organizational features [9, 13], and nucleosomes can occupy a given position in some cells but not others. The presence or absence of a nucleosome over a regulatory region dictates whether transcription factors may bind to the specific region or not. To obtain a quantitative understanding of gene regulation, it is extremely important to know not only the precise positions of nucleosomes, but also the percentage of cells that contain a nucleosome at a given position (nucleosome occupancy).

The most frequently used method of mapping nucleosome positions and occupancy involves digestion of chromatin with micrococcal nuclease (MNase), an endo- and exo-nuclease that preferentially digests the naked DNA between nucleosomes, releases the nucleosomes from chromatin, and enriches the nucleosome-protected DNA fragments [14]. To determine nucleosome positions and occupancy [12], the resulting undigested DNA is subjected to high throughput sequencing (MNase-seq) and mapped to the reference genome [15, 16].

Unfortunately, MNase has a sequence bias, and it cleaves DNA about 30 times faster upstream of an A or T than it does 5' of a G or C [17, 18]. Therefore, the nucleosome occupancy profiles, obtained by stacking and counting all mononucleosomal reads covering each genomic locus, are seriously affected by the level of MNase digestion [19–21]. Nucleosomes occupying the regions of the genome which are more accessible to MNase are

\* Correspondence: [steveh@fhcrc.org](mailto:steveh@fhcrc.org)

<sup>2</sup>Howard Hughes Medical Institute and Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA  
Full list of author information is available at the end of the article



released faster from chromatin, and they are enriched in the mildly digested samples, while nucleosomes located in less accessible regions are released more slowly from chromatin, and they are underrepresented in the mildly digested samples but enriched in extensively digested samples [19–21].

Therefore, a quantitative interpretation of the MNase-seq occupancy profiles is still problematic. While nucleosome occupancy should be understood as the precise fraction of cells that contain a nucleosome at a given genomic locus [12], the number of nucleosomal sequences that are obtained in MNase-seq experiments depends on the level of digestion that was used in each case. This “apparent” nucleosome occupancy is a function of the level of digestion, and thus cannot be used as a proxy for the real nucleosome occupancy, which is an intrinsic property of the chromatin organization in a population of cells and must not depend on the method of investigation that is used to observe it.

Important efforts have been made recently to distinguish between nucleosome occupancy and chromatin accessibility [21, 22]. Using a linear fit of the fragment frequency (i.e., nucleosome count) that is obtained at each location,  $N$ , as a function of the logarithm of the MNase enzyme concentration used in the digestion,  $E$ :

$$N = a \log(E) + b, \quad (1)$$

where  $a$  represents the slope, and  $b$  represents the  $y$ -intercept of the fit. Mieczkowski et al. have defined an MNase accessibility score, called *MACC* [21], which is effectively the negative of the fitted slope,  $MACC = -a = \frac{dN}{d(\log E)}$ . These authors have identified two scenarios of chromatin digestion in response to MNase titrations. If the number of nucleosomal fragments obtained from a given locus increases with the MNase concentration ( $a > 0$ ), then this locus was characterized by a negative *MACC* score and was called “inaccessible chromatin.” If the number of nucleosomal fragments decreases with the MNase concentration ( $a < 0$ ), then this locus was characterized by a positive *MACC* score and was labeled as “accessible chromatin.” As an alternative to the above classification given by Mieczkowski et al., another laboratory has separated the nucleosomes into “labile” (MNase-sensitive) and “stable,” by whether the nucleosomes were lost or not upon moderate or high MNase digestion [23].

There are other methods of mapping nucleosomes that are not affected by the MNase biases, e.g., sonication ChIP-seq and chemical cleavage mapping. Unfortunately, these methods present other inconveniences compared to MNase-seq. The resolution of sonication ChIP-seq experiments can be reduced to about 200 bp by using an extensive chromatin fragmentation procedure [24], but higher

resolution maps are hard to obtain. Moreover, sonication has been shown to have its own biases, and heterochromatic regions are typically more resistant to fragmentation [25]. Whereas the chemical cleavage method [10, 26] offers extremely high resolution for mapping nucleosomes in small genomes, it requires mutating or downregulating and replacing the histone genes [8], which in multicellular eukaryotes with large genomes is complicated because they contain many copies of histone genes that are both transcriptionally and post-transcriptionally regulated [27].

Because of the various drawbacks of other methods, we focused on understanding the results produced by MNase-seq experiments in a quantitative way. By considering the kinetics of nucleosome release from chromatin by MNase, we have developed a rigorous method for analyzing MNase digestion time course experiments to calculate the nucleosome occupancy and chromatin accessibility. Our predictions are not affected by the MNase biases for individual DNA sequences nor by the variability observed at different levels of chromatin digestion in the traditional MNase-seq experiments.

Our theoretical framework shows that using Eq. (1) to characterize the chromatin accessibility may be too simplistic, as mononucleosome counts are not predicted to follow a single slope obtained from Eq. (1) throughout the entire digestion course. In the initial stages of chromatin digestion in a population of cells, the number of nucleosomes that are released from every genomic locus will increase while the digestion proceeds, i.e., increased chromatin fragmentation results in more mononucleosomes that are released from multiple cells. However, in the later stages of digestion, after most of chromatin has been reduced to mononucleosomal fragments, continuing the digestion even further reduces the number of intact nucleosomes that are recovered from any locus, as MNase starts to over-digest and destroy the intact nucleosome core particles. Therefore, the amount of intact nucleosome fragments that are obtained in MNase-seq experiments from any genomic locus cannot be accurately modeled by a monotonic function of the level of digestion. Instead, the number of nucleosomes should be modeled by a function that initially increases from 0 to a maximum level, after which it slowly decreases until it vanishes again, at the stage of digestion when all intact nucleosomes have been destroyed by MNase.

Here, we analyze the process of chromatin digestion by MNase in a quantitative way, and we show that different properties of chromatin, such as nucleosome occupancy and accessibility of different genomic loci, can be disentangled using classical kinetics modeling of the biochemical reactions involving MNase and chromatin. To test the prediction of our theoretical model for

chromatin digestion, we have developed a quantitative MNase-seq (q-MNase-seq) method, which allows for meaningful comparisons between the number of mononucleosomes that are obtained genome-wide at different levels of digestion, unlike the traditional MNase-seq method.

## Results

### Kinetics of chromatin digestion and nucleosome release by MNase

It has been shown that a brief digestion by MNase produces globular supranucleosomal structures that may contain a large number (8–48) of nucleosomes [28]. The existence of these resistant domains suggests that the accessibility of MNase to different regions of the genome is variable, and different nucleosomes may be released from chromatin at different rates. For example, nucleosomes from unfolded arrays could be released from chromatin faster than nucleosomes forming the compact supranucleosomal structures. Moreover, the very characteristic “nucleosome ladder” pattern observed in gel electrophoresis of DNA resolved from MNase-digested chromatin suggests that different regions of the genome may be digested by MNase at different rates: while some parts of chromatin are only broken into large fragments, other parts are reduced to mononucleosome-size fragments.

To account for the fact that different regions of chromatin may be digested by MNase at different rates, we have developed a rigorous kinetic theory of nucleosome release from chromatin (Additional file 1). We model the chromatin digestion as shown in Fig. 1. Each genomic locus is characterized by a nucleosome occupancy (i.e., the fraction of cells in which this locus is occupied by a nucleosome) and a nucleosome accessibility. The

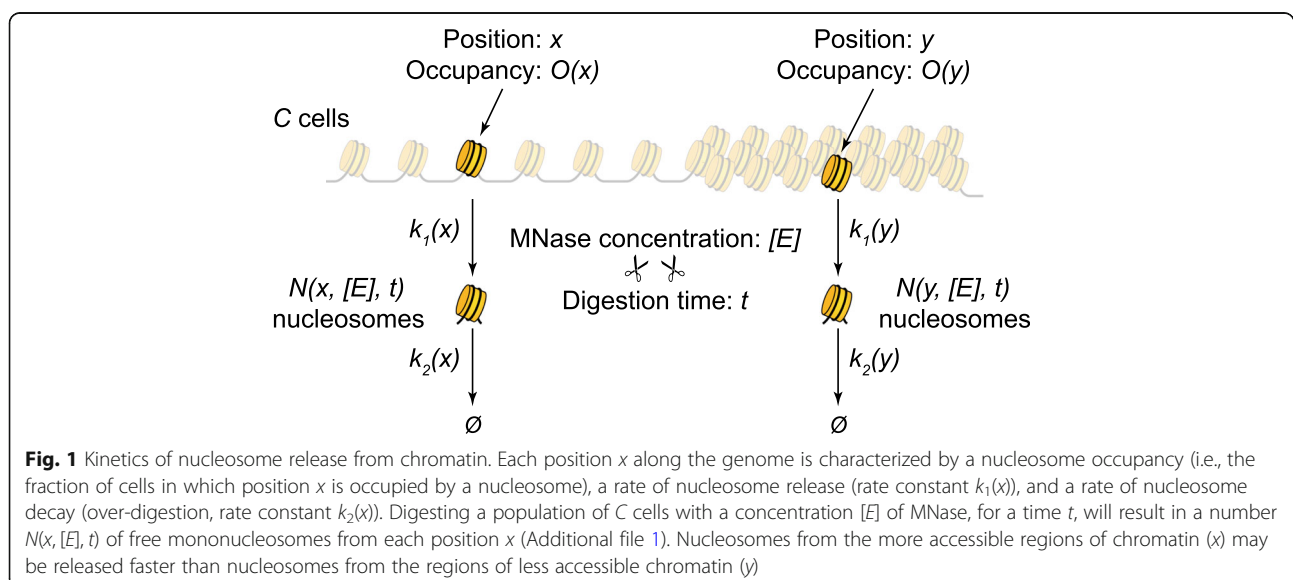
more accessible nucleosomes are the ones that are released first by MNase from chromatin.

Suppose we start with a population of  $C$  cells, and we focus on a particular genomic locus  $x$ . Let us denote the fraction of cells that contain a nucleosome (i.e., nucleosome occupancy) at this locus by  $O(x)$ . Therefore, among the  $C$  cells, only a number of  $CO(x)$  cells contain a nucleosome occupying position  $x$ , and the rest of  $C[1 - O(x)]$  cells contain a linker at position  $x$  in their genome. At the beginning of the MNase digestion reaction, all  $CO$  nucleosomes are part of intact chromatin fibers, and we denote the number of “bound” nucleosomes at time  $t = 0$  by  $B(0)$ . As digestion proceeds, after time  $t$ , we will have a number of nucleosomes,  $B(t)$ , that are still bound to chromatin and a number of nucleosomes,  $N(t)$ , that had already been released from chromatin as mononucleosomes. During digestion, nucleosomes are released from chromatin, increasing  $N(t)$  and decreasing  $B(t)$ . At the same time, MNase also attacks free mononucleosomes from the sample, slowly decaying the number of these due to over-digestion and internal cleavages followed by nucleosome disassembly. These two processes can be represented by a simple reaction chain,  $B \rightarrow N \rightarrow \emptyset$ . The corresponding concentrations of nucleosomes from the sample satisfy the following differential equations:

$$\frac{d[B]}{dt} = -k_1[B][E]$$

$$\frac{d[N]}{dt} = k_1[B][E] - k_2[N][E]$$

where  $[E]$  is the concentration of free MNase, which remains approximately constant in the experiment, and  $k_1$  and  $k_2$  represent the rate constants for the nucleosome



release and decay reactions, respectively (Additional file 1). The solutions to these equations, satisfying the initial conditions  $B(0) = CO$  and  $N(0) = 0$  are (Additional file 1):

$$\begin{aligned} [B](t) &= [C]Oe^{-k_1[E]t} \\ [N](t) &= [C]O\frac{k_1}{k_1-k_2}\left(e^{-k_2[E]t}-e^{-k_1[E]t}\right) \end{aligned} \quad (2)$$

Therefore, after time  $t$ , a fraction  $f_B(t) = \frac{[B](t)}{[B](0)}$  of the nucleosomes initially found at locus  $x$  in the population of cells are still bound to the chromatin fibers and a fraction  $f_N(t) = \frac{[N](t)}{[B](0)}$  of the nucleosomes were released from chromatin and are present as intact mononucleosomes in the sample. Additional file 1: Figure S1 shows the dependence of these nucleosome fractions on the level of digestion, denoted by  $d = [E]t$ . As seen in Additional file 1: Figure S1, the predicted fraction of mononucleosomes that are present in the sample (red line) does not vary monotonically with the degree of digestion [21] but has a more complex shape. The nucleosome counts will initially increase with the level of digestion, until they reach a maximum, for a digestion level at which the rate of nucleosome release is equal to the rate of nucleosome decay,  $k_1[B][E] = k_2[N][E]$ , which is equivalent to  $k_1e^{-k_1[E]t} = k_2e^{-k_2[E]t}$  (Additional file 1). If digestion continues further, the rate of nucleosome decay, ( $k_2[N][E]$ ), starts to be higher than the rate of release of new nucleosomes from chromatin ( $k_1[B][E]$ ), and the overall level of intact mononucleosomes starts to decrease until it completely vanishes, when MNase has destroyed all nucleosomes by over-digestion.

Different genomic loci  $x$  are characterized by different nucleosome occupancies and accessibilities, and the nucleosome counts that are obtained from each locus will depend on four parameters: the level of digestion  $d$ , nucleosome occupancy  $O(x)$ , and the two rate constants  $k_1(x)$  and  $k_2(x)$ , corresponding to the nucleosome release and decay processes, respectively. The predicted dependence of the nucleosome counts obtained from any genomic locus on the corresponding parameters  $O$ ,  $k_1$ ,  $k_2$ , and on the level of digestion  $d$ , is shown in Additional file 1: Figure S2 and S3. Additional file 1: Figure S2a shows that nucleosome occupancy  $O$  affects the overall height and area of the nucleosome count profile as a function of the digestion level, acting as a simple rescaling factor when the other parameters ( $k_1$ ,  $k_2$ ) are kept constant. Additional file 1: Figure S2b,c shows that the rate constant  $k_1$  mostly affects the initial slope of the apparent occupancy,  $N/C$ , while the decay rate constant  $k_2$  mostly affects the right tail of the apparent occupancy distribution. The fact that the three parameters  $O$ ,  $k_1$ , and  $k_2$  influence different aspects of the apparent occupancy (Additional file 1: Figure S2) guarantees the

uniqueness of the parameters obtained by fitting this distribution using Eq. (2). Additional file 1: Figure S3 reinforces this conclusion by showing that the rate constant  $k_2$  can be precisely obtained from the asymptotic behavior of the logarithm of the apparent occupancy for extensive digestions, and the estimated values of  $k_2$  are not affected by the other two parameters ( $O$ ,  $k_1$ ).

Although the solution given by Eq. (2) seems complicated, one can easily use a non-linear fit algorithm in order to obtain the parameters  $O$ ,  $k_1$ , and  $k_2$ , if the nucleosome amounts generated at multiple digestion levels are measured. For example, this non-linear fit can be obtained using the `lsqcurvefit` function in MATLAB, `nls` function in R, or `scipy.optimize.leastsq` function in Python.

Unfortunately, the traditional MNase-seq method is not appropriate for a rigorous comparison and fit of the nucleosomal counts that are obtained in different experiments. In general, when MNase-seq data are compared, these data are normalized such that the total number of reads in all experiments is set to a common value. This normalization is not valid for our purpose, as the total number of nucleosomes released from chromatin depends on the level of digestion. For this reason, we adopted a modified MNase-seq procedure, q-MNase-seq (Additional file 1: Figure S4), which allows us to do meaningful comparisons between the nucleosome counts obtained at different stages of the digestion and to fit the data in order to obtain the three parameters ( $O$ ,  $k_1$ ,  $k_2$ ) for each nucleosome, as described next.

#### Quantitative measurement of the number of nucleosomes released from chromatin at different levels of digestion

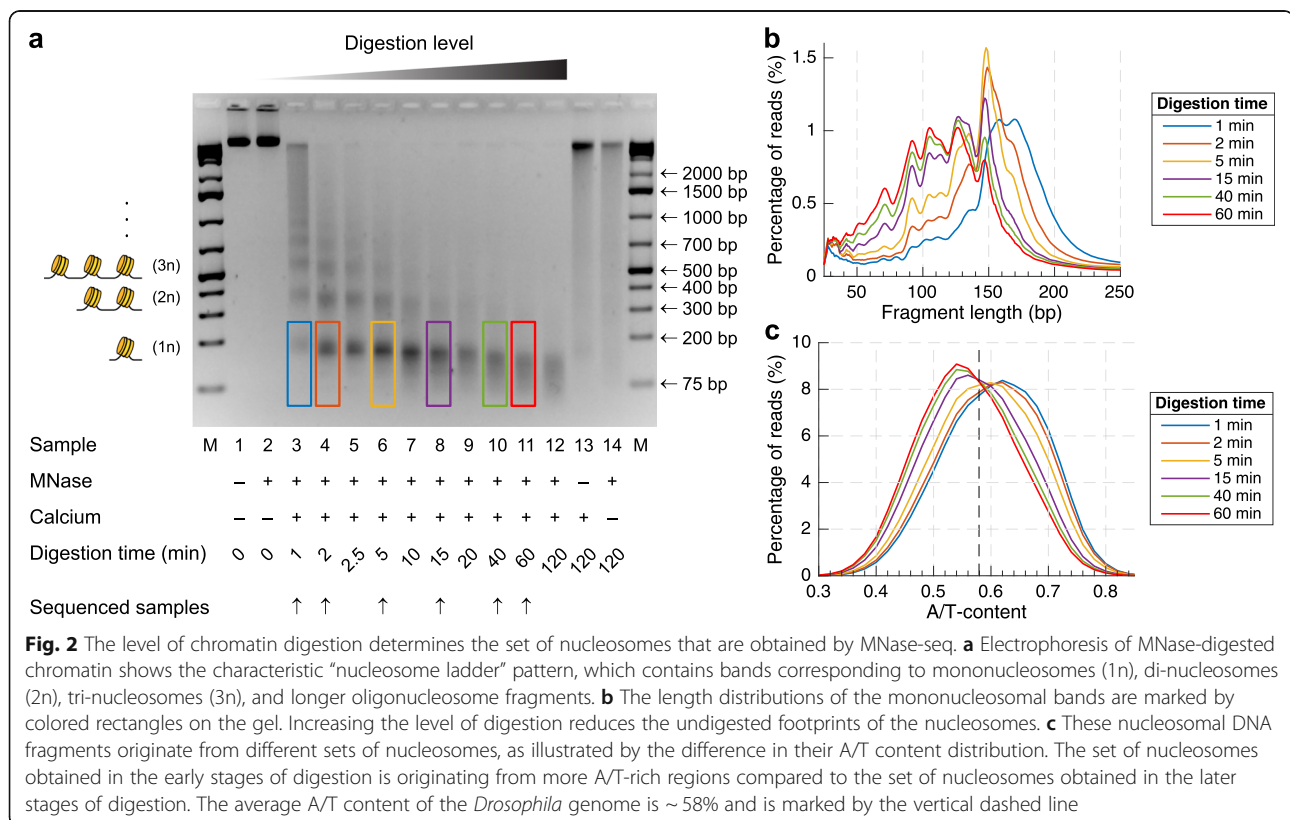
We start with similar numbers of *Drosophila* S2 cells obtained by splitting the total culture into equal volumes. These aliquots are subjected to different levels of digestion by MNase. The resulting undigested DNA is then purified. Since each sample will release a different number of nucleosomes, depending on the corresponding degree of digestion, to keep track of the relative amounts of mononucleosomes from all samples, we add equal amounts of spike-in DNA (*Saccharomyces cerevisiae* mononucleosomal DNA obtained by a traditional MNase-seq method). Then, each sample consisting of a mix of fly and yeast DNA is used to prepare the sequencing library. As each sample already contains the spike-ins, it is acceptable to use different numbers of cycles of PCR amplification, if that is necessary in order to obtain similar amounts of total DNA for sequencing. After sequencing, the DNA from each sample is aligned both to the *Drosophila* genome (version *dm6*) and to the *S. cerevisiae* genome (version *sacCer3*) to obtain the numbers of fly mononucleosomes and spike-in reads. Then, we normalize the counts of fly mononucleosomes by the amount of spike-in DNA from each sample such that all

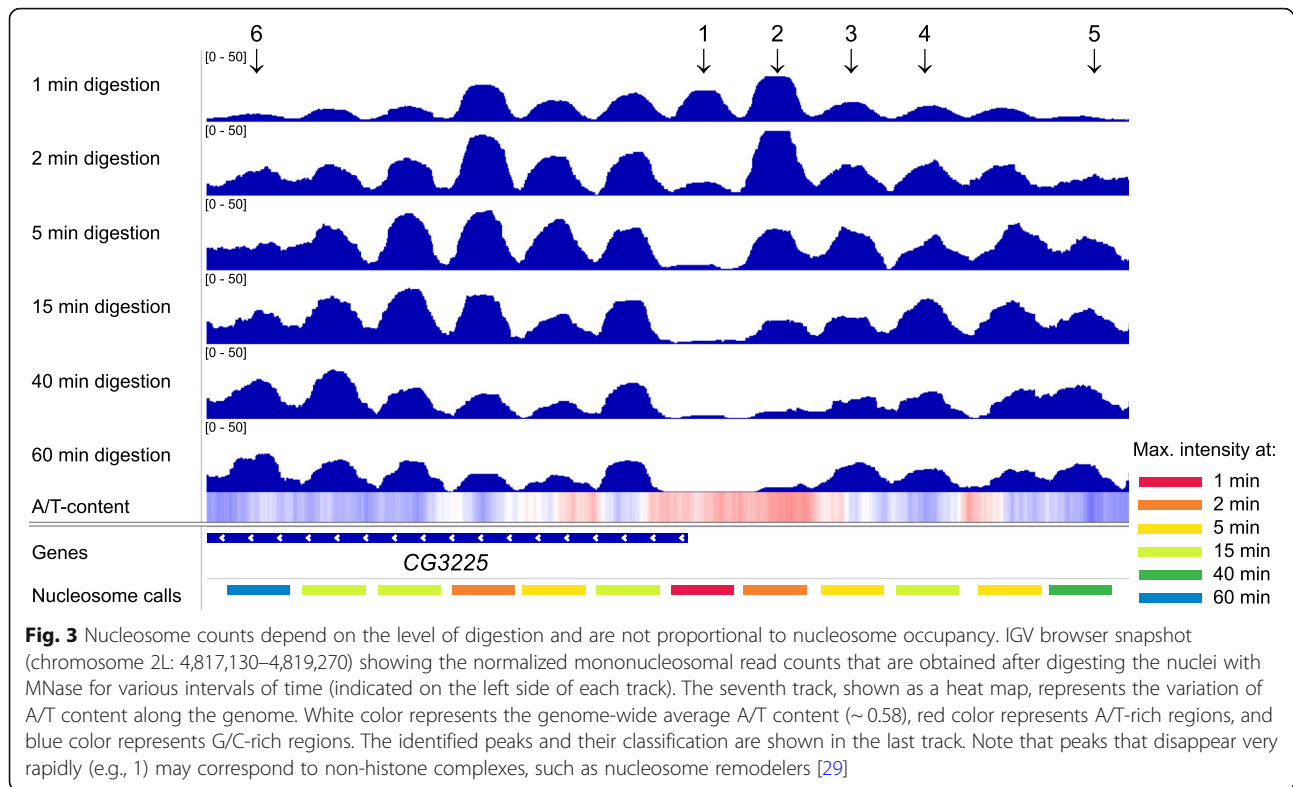
samples have the same number of spike-in fragments. Using the normalized data, we can analyze any genomic region and plot the nucleosome counts as a function of the digestion level (Additional file 1: Figure S4). These data can be fitted using Eq. (2) to obtain the nucleosome occupancy,  $O$ , and the two rate constants,  $k_1$  and  $k_2$ , for each genomic locus (Additional file 1: Figure S5).

We performed a q-MNase-seq experiment, by digesting *Drosophila* S2 cells with equal amounts of MNase for various time periods: 0 min (control), 1 min, 2 min, 2.5 min, 5 min, 10 min, 15 min, 20 min, 40 min, 60 min, and 120 min. The extent of digestion was verified by gel electrophoresis in an agarose gel (Fig. 2a). In theory, three time points are required to fit the three parameters,  $O$ ,  $k_1$ , and  $k_2$ , but because of stochastic noise, five to six time points are preferred to obtain sufficient accuracy. We selected six time points (1 min, 2 min, 5 min, 15 min, 40 min, 60 min) for which we performed end-polishing, adapter ligation, PCR amplification, and paired-end sequencing of the undigested DNA fragments without size selection. As previously observed [20], the mononucleosomal bands from the mild digestions contain longer DNA fragments (Fig. 2b) originating from the more A/T-rich regions of the genome (Fig. 2c), while the mononucleosomal bands from the more extensive digestions contain shorter DNA fragments (Fig. 2b),

and correspond to nucleosomes that occupy the more G/C-rich regions of the genome (Fig. 2c).

The nucleosome counts that were obtained in the six experiments were then normalized by the spike-in DNA amounts from each sample. An IGV browser snapshot of a representative region of the genome illustrates some of the differences among the properly normalized nucleosome counts observed after different levels of digestion (Fig. 3). The nucleosome counts vary with the level of digestion for each nucleosome. Some nucleosomes are quickly released from chromatin (Fig. 3, peaks 1 and 2), and they are enriched in the samples that were just mildly digested for 1 or 2 min. Peak 1 in Fig. 3 likely represents a non-histone barrier complex [20] that offers only a limited protection against MNase or a RSC-nucleosome complex [29] that is observed in a small number of cells during the remodeling process. Such nucleosomes are also the first ones that are over-digested by MNase and are underrepresented or completely disappear from the distributions obtained in the latest stages of the digestion (40 min, 60 min). These kinds of nucleosomes are usually denoted as MNase-sensitive nucleosomes and were shown to contain A/T-rich sequences [19, 20]. In yeast and animals, active regulatory elements are often adjacent to asymmetrically unwrapped nucleosomes [30–32]. In yeast, these nucleosomes are generally





more accessible to MNase due to the RSC complex being bound to them [29], which destabilizes the nucleosome core particle and exposes more DNA to MNase.

Using the combined nucleosome counts from the six experiments, we next identified the typical positions of nucleosomes along the *Drosophila* genome, using a custom algorithm for computing the median position for all clusters of aligned nucleosomes (described in Additional file 1). We identified about 700,000 typical nucleosome positions, covering about 73% (100 Mb) of the whole genome size. The expected fraction of the *Drosophila* genome that is covered by nucleosomes is roughly 73%, given that each nucleosome covers 147 bp out of every 200 bp (the typical nucleosome repeat length in flies). Therefore, we are confident that our nucleosome calling algorithm has detected most of the nucleosomes, including the poorly positioned ones.

In order to offer enough protection against MNase to result in a peak in MNase-seq occupancy profiles, a protein must be stably bound to DNA. The transient in vivo binding of fast transcription factors will not protect DNA during 5–60 min time course of MNase digestion. Such transient binding events may only give a significant peak in MNase-seq occupancy profiles obtained by digesting the chromatin for a very short period of time (e.g., 1 min digestion in our experiments). As we describe below, the sites that were bound by “MNase-sensitive” complexes (i.e., MNase-seq peaks that were

observed in the mild digestions and disappeared in the latter stages of digestion) accounted for fewer than 1% of the total number of identified peaks.

The bottom track in Fig. 3 shows the identified nucleosomes in the corresponding region near the gene *CG3225*. After we identified the well-positioned nucleosomes and normalized the nucleosome counts using the spike-in levels from each sample, we were able to study the evolution of the number of released nucleosomes as a function of the digestion level. Even in a small region of the genome (Fig. 3), neighboring nucleosomes are released with different kinetics, e.g., peak 1 has its maximum intensity after 1 min of digestion, peak 2 after 2 min of digestion, peak 3 after 5 min of digestion, peak 4 after 15 min of digestion, peak 5 after 40 min of digestion, and peak 6 after 60 min of digestion.

In another region of the *Drosophila* genome, the gene start and end sites are A/T-rich (Additional file 1: Figure S6), as is usually the case for many genes in diverse organisms. These loci are occupied by MNase-sensitive nucleosomes, which are underrepresented in the latest stages of the digestion. Again, A/T-rich regions that contain nucleosomes are released faster (colored in red, orange, and yellow), while G/C-rich regions that contain nucleosomes are released later in the digestion process (colored in green and blue).

Traditionally, MNase-seq experiments are performed using extensive levels of digestion (when > 80% of the

chromatin is reduced to mononucleosomes). It was previously assumed that nucleosome gaps at promoters and transcription termination regions are largely nucleosome-free [6]. We believe that this interpretation does not hold for transcription termination sites (TTSs), which are some of the most A/T-rich regions in the genome, and therefore, nucleosomes are artificially underrepresented from these loci in the extensively digested samples [10, 20]. Moreover, promoters are also A/T-rich regions, and nucleosomes from these regions may also be underrepresented in the typical MNase-seq experiments, usually corresponding to a fairly extensive level of digestion. For this reason, if an extended A/T-rich intergenic region appears to be depleted of nucleosomes in the typical MNase-seq experiments, it must be remembered that this could be just an artifact introduced by an extensive level of MNase digestion.

#### Nucleosomes occupying different DNA sequences are released from chromatin at different rates

From Fig. 3 and Additional file 1: Figure S6, one can hypothesize that nucleosomes found in A/T-rich regions are released faster from chromatin compared to nucleosomes found in G/C-rich regions. Those that are released faster could also be destroyed faster and lost from the mononucleosomal band of the extensively digested samples. We next tested this hypothesis using all the well-positioned nucleosomes that we identified. Figure 4 shows the density of sequencing reads that we obtained for each time point of the digestion, as a function of the A/T content of these nucleosomes. We see that in the initial stages of the digestion (Fig. 4, top row) the A/T-rich nucleosomes are relatively over-represented in the sample, and the mononucleosomal fragments are slightly longer than 147 bp, indicating that linker DNA adjacent to the nucleosome core particles is not yet fully trimmed. When digestion proceeds further, MNase starts to invade the nucleosome core, resulting in shorter DNA fragments (< 147 bp). The A/T-rich nucleosomes (the right-most bin) that were released first from chromatin are almost completely destroyed by MNase after 40 min of digestion (bottom two rows, 40 min and 60 min, respectively). In the latest stages of digestion (bottom row), the A/T-rich nucleosomes are underrepresented from the sample and contrary to the mild digestion (top row); now, it is the G/C-rich nucleosomes that are relatively over-represented in the sample.

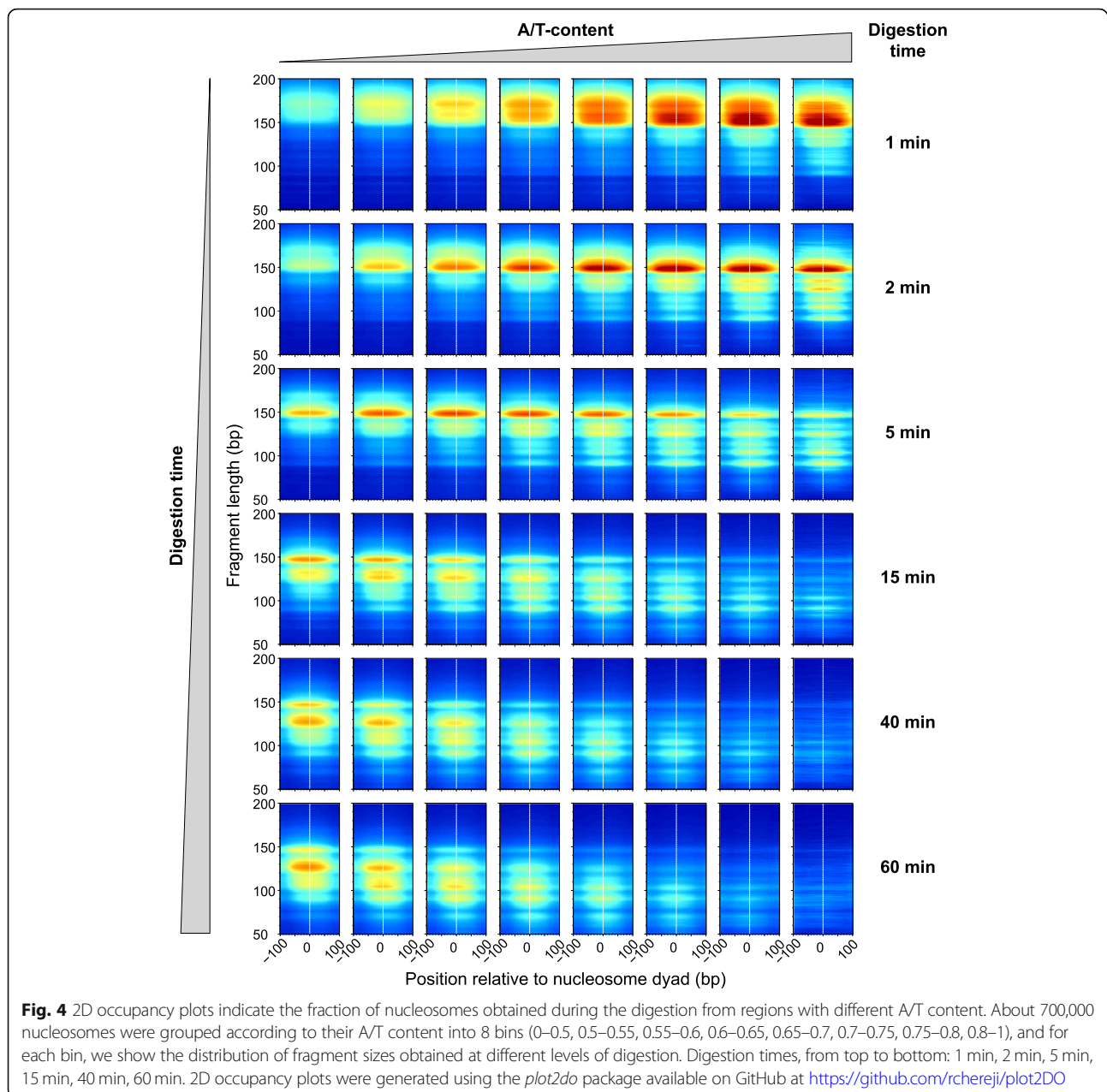
Because A/T-rich nucleosomes are destroyed first (Fig. 4) and are underrepresented in the extensively digested samples (Fig. 4, bottom row), it was previously claimed that nucleosomes have an intrinsic preference for the G/C-rich sequences [33], and models have been proposed to explain nucleosome organization based on DNA sequence alone [34, 35]. We believe that the G/C

enrichment in the nucleosomal DNA that was previously observed may be an artifact introduced by MNase, which preferentially destroyed most of the A/T-rich nucleosomes by the time the digestion reaction was stopped. Moreover, any computational model that was trained with the nucleosome occupancy profiles obtained from extensively digested chromatin, could be artificially biased to favor the G/C-rich locations on the genome and would underestimate the A/T-rich nucleosomes.

#### The experimental data confirms the prediction of our theoretical framework

Using the spike-in normalized nucleosome counts, we can test the prediction of our theoretical model. Having identified all well-positioned nucleosomes, we analyzed the digestion time course for all ~700,000 regions of the genome where we identified a nucleosome. Figure 5 shows the spike-in normalized coverage (apparent occupancy) for 6 loci, centered on 2 nucleosomes that give the maximum signal after 2 min of digestion (upper panels), 2 after 5 min of digestion (center panel), and 2 after 15 min of digestion (lower panels). From the occupancy profiles (Fig. 5, left plots from each pair), we computed the average occupancy for the center nucleosome, and we plotted these as a function of the digestion time (Fig. 5, blue circles in the right plots from each pair). These values follow the predicted dependence on the level of digestion (Eq. (2)), and using Eq. (2), we can fit the data to obtain the real nucleosome occupancy  $O$  and the two rate constants  $k_1$  and  $k_2$  characterizing the nucleosome release and decay processes.

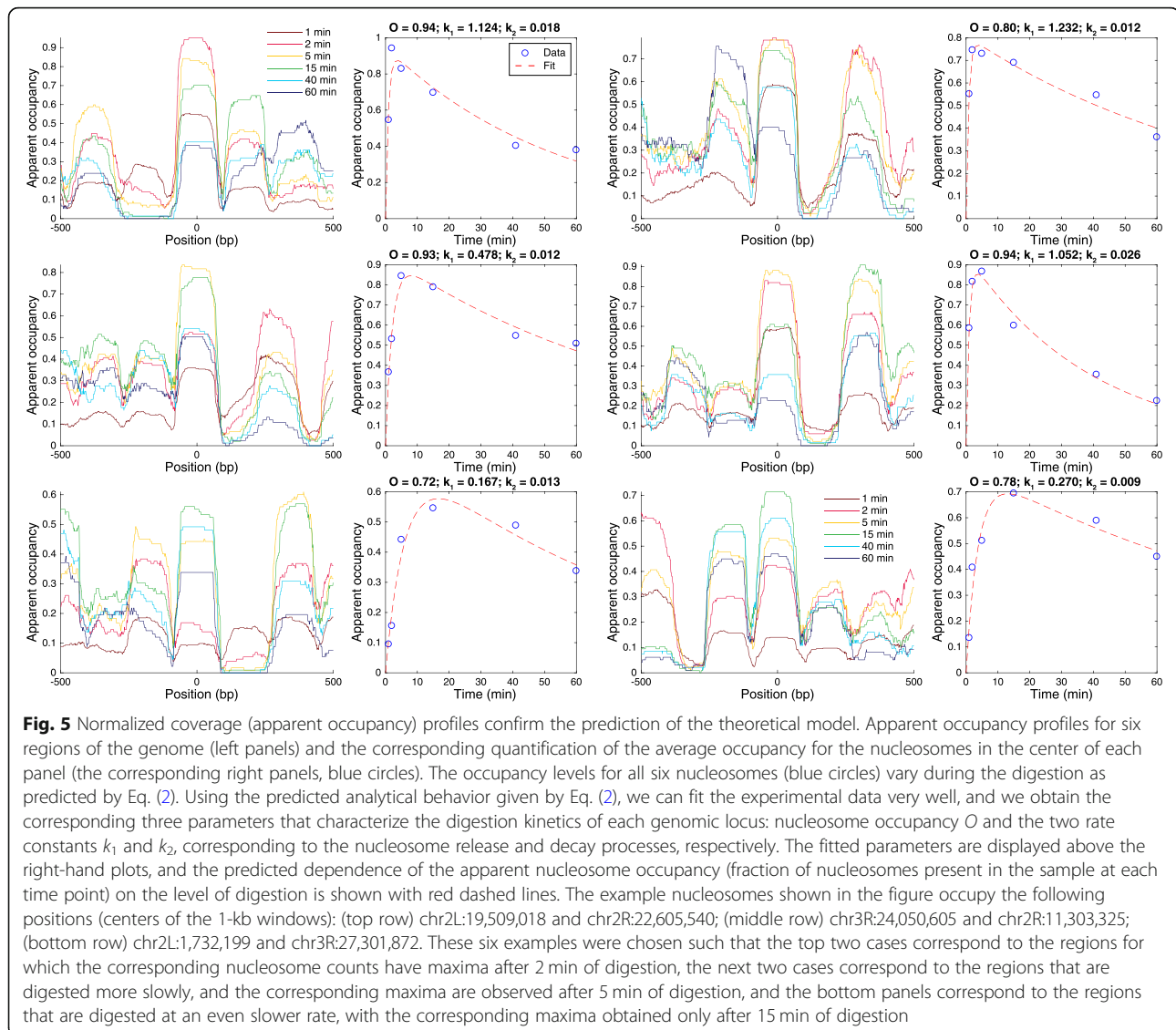
Figure 6a shows the distribution of the normalized nucleosome counts of all nucleosomes as a function of the level of digestion. As predicted by our model (Additional file 1: Figures S1 and S2), the normalized nucleosome counts are initially increasing to a maximum, and then they decrease after enough nucleosomes have been released from chromatin and when the rate of nucleosome decay is higher than the rate of nucleosome release. For more than 90% of the nucleosome loci, the maximum nucleosome count is obtained after 2 min (23.95% of nucleosomes), 5 min (34.50%), or 15 min of digestion (32.48%). A small fraction of the loci (< 1%) are characterized by a maximum nucleosome count obtained after only 1 min of digestion (the first time point that we sequenced), and another small fraction of the loci (~3.53%) give the maximum nucleosome counts after 60 min of digestion. We grouped these loci into six classes, according to the time point at which we obtained the maximum nucleosome count at the corresponding locus. Figure 6b shows the average nucleosome counts that were obtained for all six classes of nucleosomes and all six digestion levels. Apart from all other classes, the loci from class 1 show significant protection only at the initial time



point (1 min), and after that, these loci are characterized by a trough between two other nucleosomes. This class of weakly protected loci (<1% of all called MNase-seq peaks) likely represent the MNase-sensitive complexes formed by non-histone proteins [20], which still weakly hinder MNase digestion. Figure 6 suggests that neighboring nucleosomes are not released from chromatin in an independent way, as they share a common linker. If a linker is AT-rich or particularly accessible to MNase, then both of the neighboring nucleosomes are expected to be released faster from chromatin, compared to other nucleosomes, which do not have any of the two linkers preferentially cut by MNase.

Analyzing the distribution of the chromatin digestion rates obtained by fitting the nucleosome counts using Eq. (2), we observed two groups of sites (Fig. 7a). Most of the sites (>99%) are characterized by values of  $k_1$  close to 1 and represent the typical nucleosomes, which are released from chromatin at relatively slow rates. The second group of sites (<1%) is characterized by relatively high values of the chromatin digestion rate  $k_1$  (>10) and represents loci that are bound by relatively fragile complexes, which offer only a weak protection against MNase and are quickly destroyed by MNase. Figure 7b shows the distribution of apparent occupancy for these MNase-sensitive complexes (left) and the corresponding



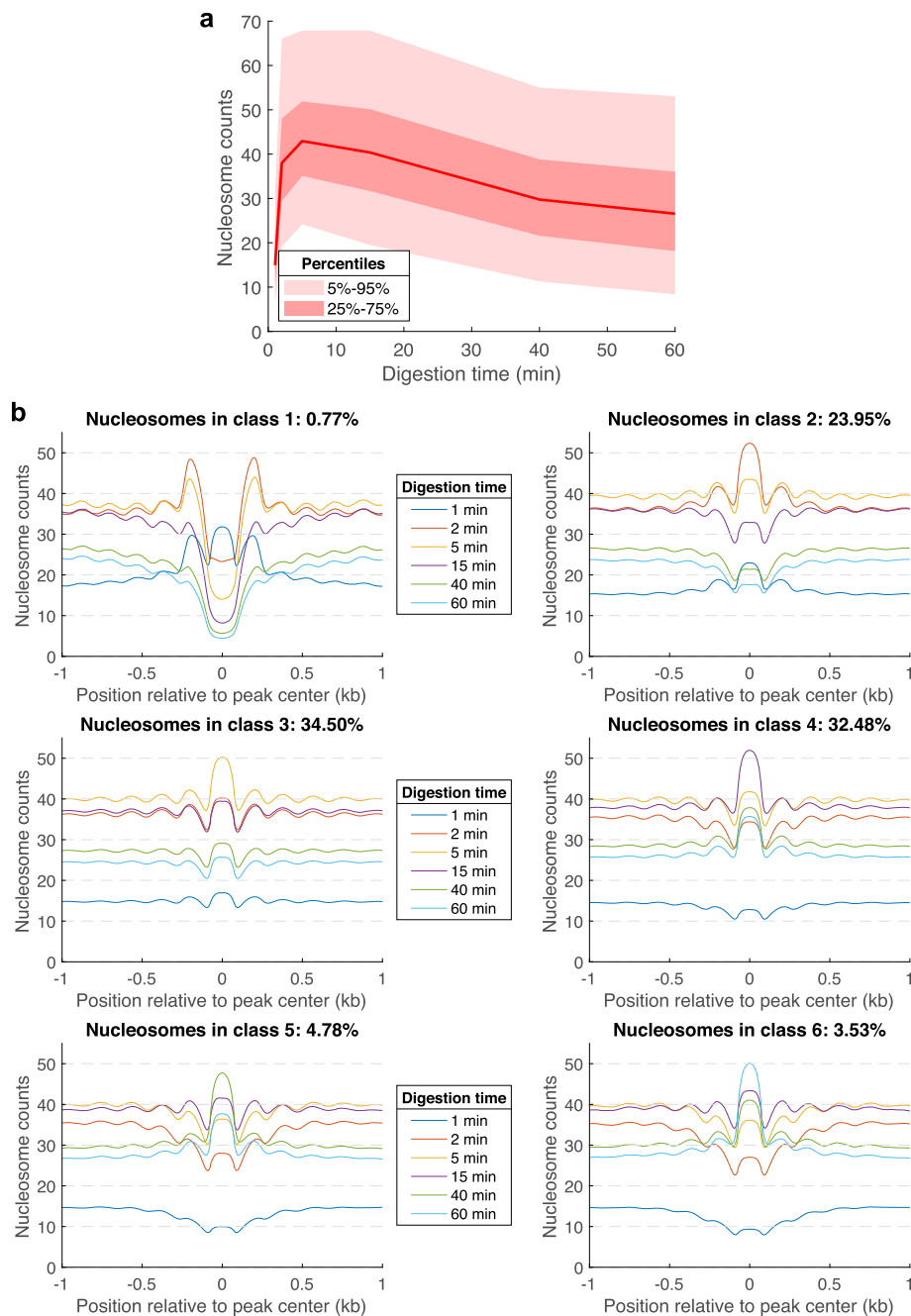


distributions that we obtained by fitting these data using Eq. (2) (right panel). Using ChIPseeker [36], we annotated the  $\sim 4500$  loci occupied by fragile complexes. We found that most of these loci are located in gene promoters ( $\sim 82\%$ ), while others are located in introns ( $\sim 11\%$ ) and intergenic regions ( $\sim 5\%$ ) (Fig. 7c). To find out whether these sites were previously annotated in the literature or not, we have downloaded the annotations of all transcription factor (TF) binding sites that are available on ChIP Atlas (<https://chip-atlas.org>). As expected, we found a strong enrichment of TF binding sites at the loci occupied by fragile complexes (Fig. 7d). Therefore, these fragile complexes are likely to represent TFs bound to promoters or enhancers, or other proteins that can bind and destabilize nucleosomes, such as SWI/SNF family chromatin remodeling complexes [29].

The bulk of chromatin is occupied by nucleosomes that are released at a relatively lower rate ( $k_1 \approx 1$ ) (Fig. 7a). Surprisingly, these nucleosomes are characterized by a wide range of occupancy levels (Fig. 8). Using the fitted values for the nucleosome occupancy, we split the  $\sim 700,000$  loci along the *Drosophila* genome according to their corresponding occupancy (Fig. 8) into nine bins. Equation (2) fitted well the apparent occupancy for all groups of nucleosomes (Fig. 8), confirming that the nucleosome counts that are obtained during MNase-seq experiments can be well modeled using our theoretical framework.

#### Different chromatin regions are characterized by similar nucleosome accessibilities

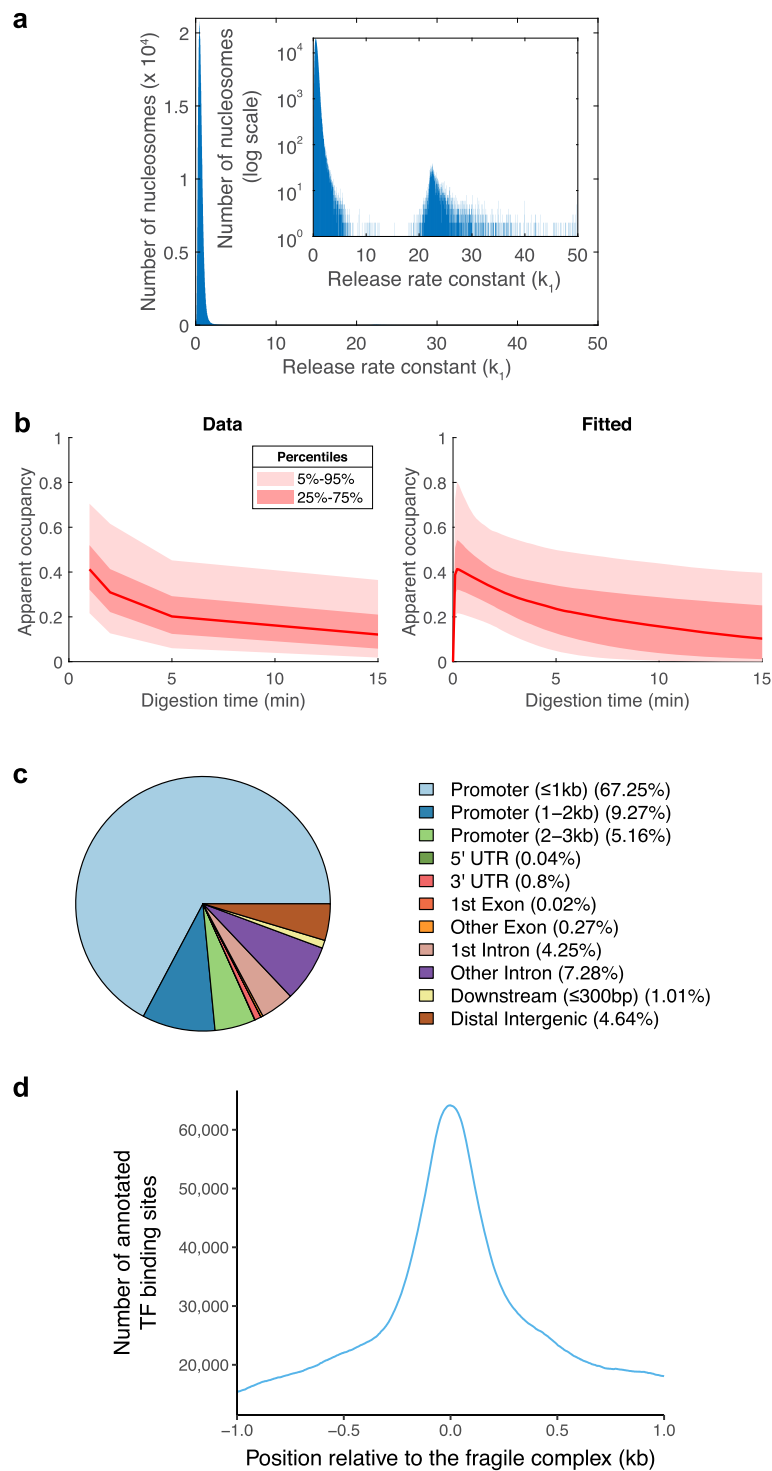
To test whether different regions of the genome have very different chromatin accessibilities, as suggested by



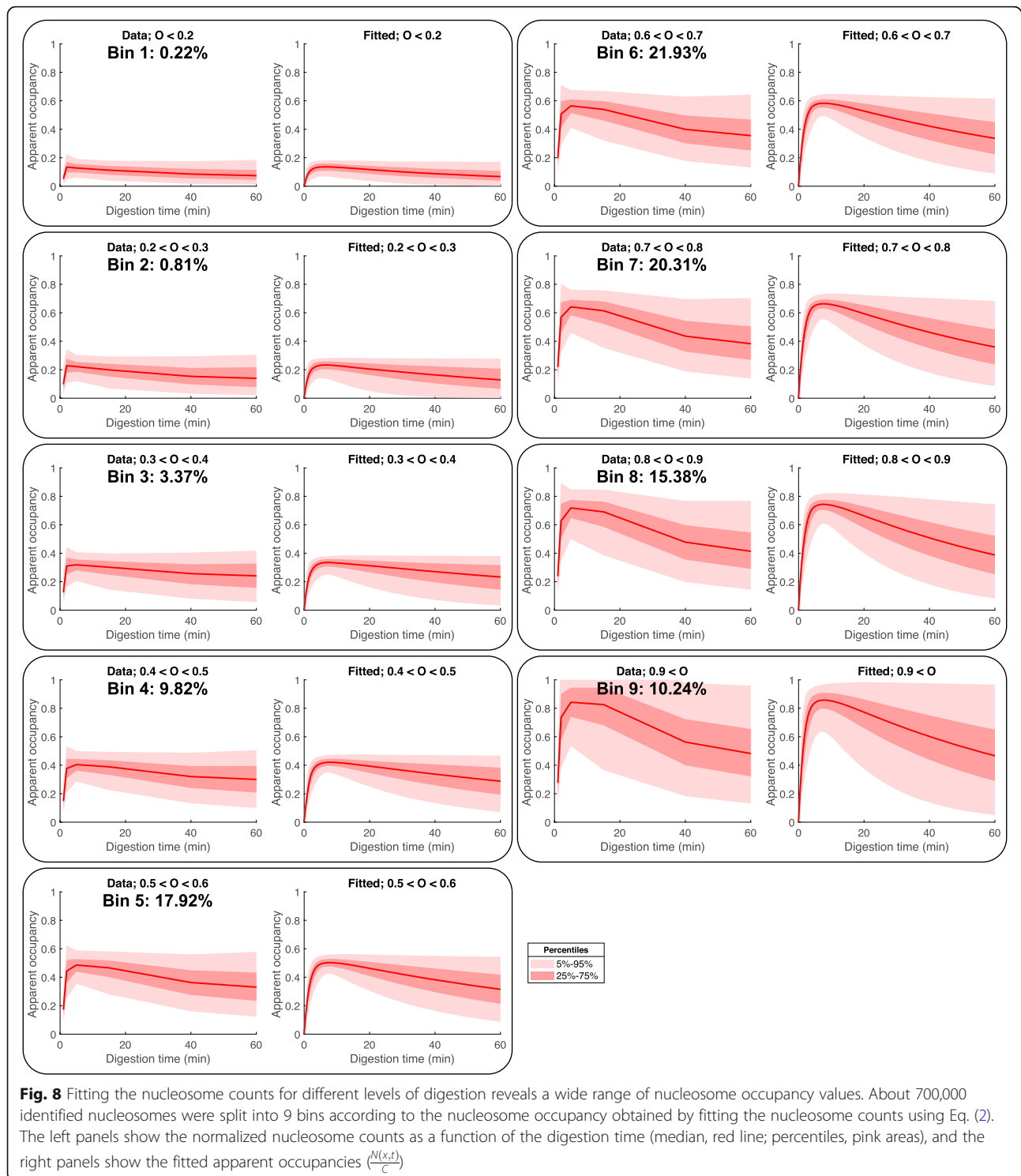
**Fig. 6** Average nucleosome counts for the six classes of nucleosomes. **a** Median nucleosome count (red line), the interquartile range (25–75%), and the 5–95% range, as a function of digestion time. **b** Average nucleosome counts for the six classes of nucleosomes that we identified according to the digestion time after which they generate the strongest signal in our experiments

DNase-seq and ATAC-seq experiments, we next analyzed the distribution of nucleosome release rates that we obtained along the genome. We first analyzed the DNaseI hypersensitive regions (DHSs) (obtained from [37]) and compared the nucleosome accessibility of the nucleosomes located less than 250 bp from DHSs and those away from DHSs. As expected, nucleosome accessibility, as measured by the nucleosome release rate

constant  $k_1$ , is higher near DHSs (Fig. 9a, left panel), but surprisingly, the accessibility of nucleosomes from DHSs is still of the same order of magnitude as the accessibility of DNaseI insensitive regions (less than twofold difference between the median levels of  $k_1$ ). The middle and right panels in Fig. 9 show the distributions of the fitted  $k_2$  and nucleosome occupancy, for the nucleosomes belonging to the indicated group.

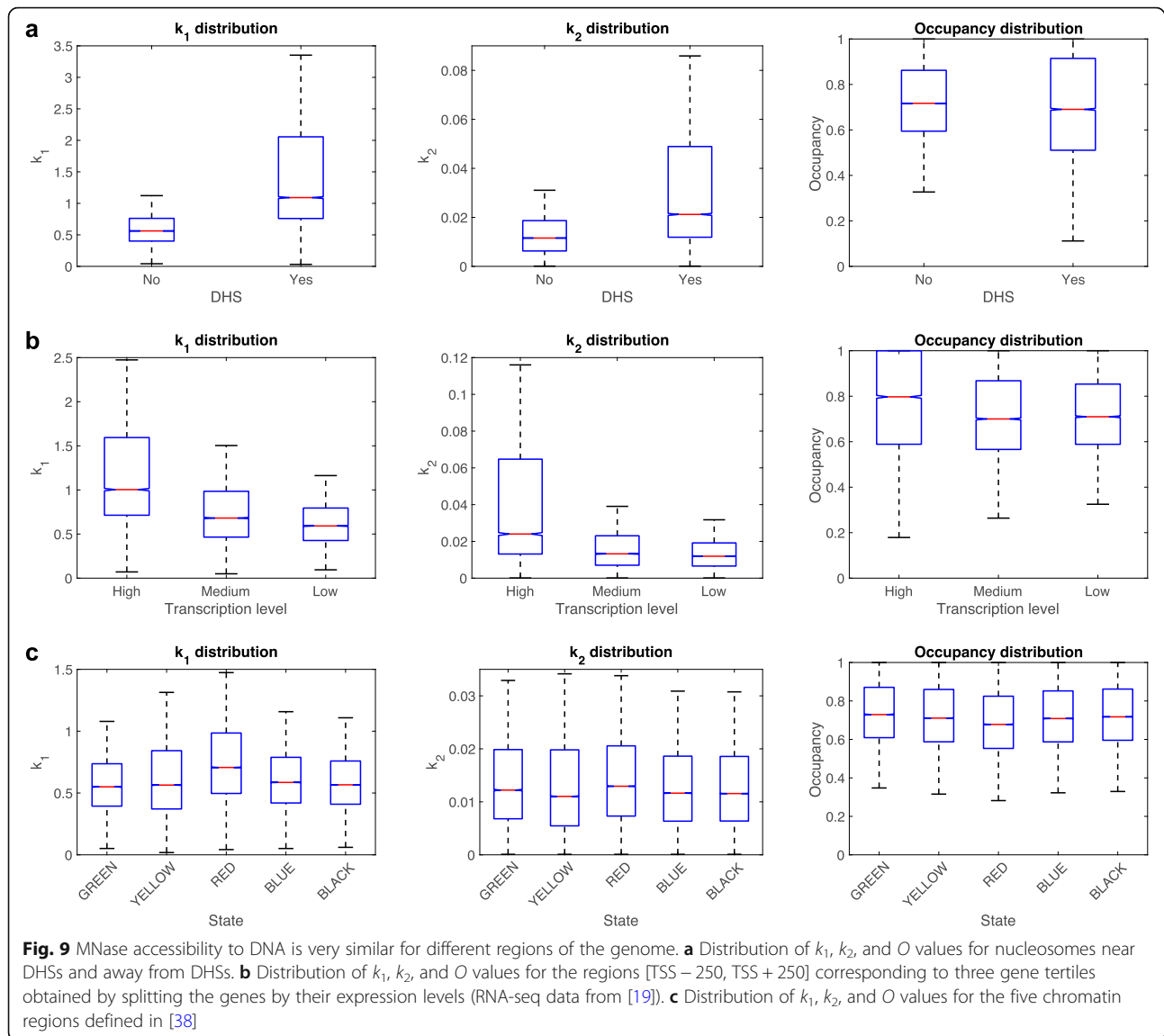


**Fig. 7** The chromatin digestion rate separates the normal nucleosomes from fragile complexes. **a** Histogram of the number of  $k_1$  values obtained by fitting the nucleosome counts using Eq. (2). A small fraction ( $< 1\%$ ) of the loci where we identified MNase-seq peaks are characterized by very high chromatin digestion rate constants ( $k_1 > 10$ ) compared to typical nucleosomes ( $k_1 \approx 1$ ). **b** The rescaled nucleosome counts (apparent occupancy,  $\frac{N(x,t)}{c}$ ) of the loci characterized by high chromatin digestion rates ( $k_1 > 10$ ) (left) and the corresponding fitted curves, given by Eq. (2) (right). **c** Annotations of the loci occupied by fragile complexes ( $k_1 > 10$ ) obtained with ChIPseeker [36]. **d** The  $\sim 60,000$  annotated binding sites of different transcription factors (annotations of TF binding sites in S2 cells were downloaded from CHIP Atlas, <https://chip-atlas.org>)



To test whether transcription affects the nucleosome accessibility, we split the genes into three tertiles according to the transcription levels obtained from RNA-seq data [19], and we compared the accessibility of the nucleosomes located near the TSSs of these genes ([TSS – 250, TSS + 250]). We observed that nucleosome accessibility

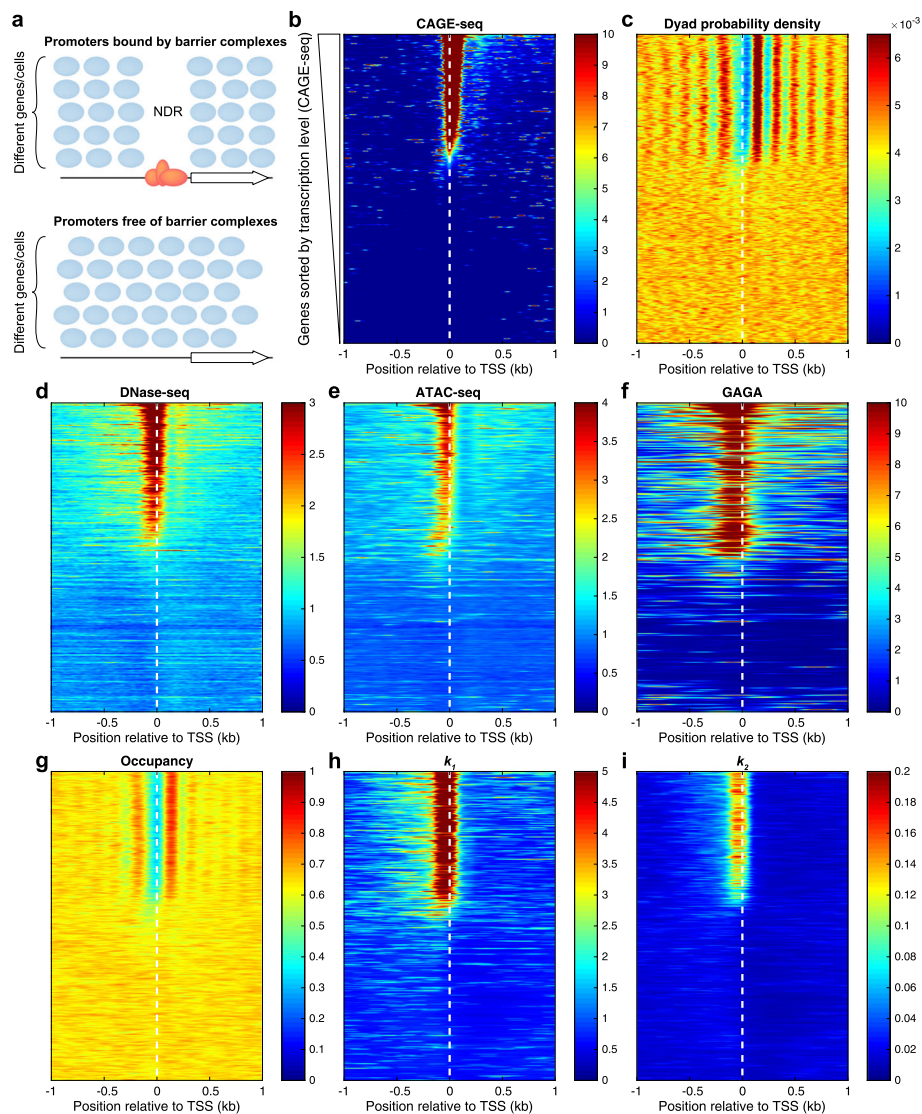
correlates with transcription, and the tertile of the most transcribed genes contained the most easily accessible nucleosomes (Fig. 9b). Again, the differences were relatively small, and the overall accessibility of the promoters corresponding to the three tertiles did not vary by orders of magnitude, but only by less than twofold.



Next, we tested whether the regions of the genome that could be away from promoters contain either highly accessible or inaccessible nucleosomes. For this, we compared the accessibility of the five previously reported chromatin states, obtained by principal component analysis and clustering of genome-wide distributions of 53 chromatin proteins, mapped using the DamID method [38]. The distribution of the  $k_1$  values for the nucleosomes corresponding to the five chromatin states are shown in Fig. 9c. Only the “RED” chromatin state contains nucleosomes that are relatively more accessible, as expected since the “RED” chromatin state contains “hubs of regulatory activity” [38] and are enriched in origins of replication. Surprisingly, the “BLUE” and “GREEN” chromatin states, which contain the known heterochromatin regions, are only slightly less accessible to MNase compared to the other three chromatin states, which are predominantly euchromatic (Fig. 9c).

### Nucleosome phasing by transcription-related barrier complexes

Previously, we have shown that nucleosome organization in *S. cerevisiae* can be predicted using a simple biophysical model of “barrier complexes” occupying promoters and steric hindrance between nucleosomes and the barrier complexes [10, 39]. Our model predicts that promoters that are occupied by such barrier complexes will be characterized by an NDR flanked by phased nucleosomes on both sides, while promoters that are not occupied by barrier complexes will not have an NDR and nucleosomes will be more disorganized (out-of-phase) (Fig. 10a). On the one hand, *S. cerevisiae* promoters are all relatively active and presumably occupied by general regulatory factors (such as Abf1, Rap1, and Reb1) and other transcription factors. Therefore, all yeast promoters are of the first type, characterized by NDRs and phased nucleosomes. On



**Fig. 10** Promoter nucleosome organization: a two-state model. **a** Our model predicts two different nucleosome organization of gene promoters. Promoters bound by barrier complexes are predicted to contain an NDR flanked by phased nucleosomes, while promoters that are free of barrier complexes are predicted to have disorganized nucleosomes and lack NDRs. **b** *Drosophila* promoters sorted by the transcription level (CAGE-seq data from [40]). **c** Nucleosome dyad organization confirms the model prediction: promoters of active genes contain NDRs that are flanked by phased nucleosomes on both sides, while promoters of inactive genes contain disorganized nucleosomes and no NDR. The promoters of active genes (top ~40% of the genes) are accessible to DNaseI (**d**) (DNase-seq data from modENCODE [41], modENCODE\_3324) and Tn5 transposase (**e**) (ATAC-seq data from [22]) and are bound by barrier complexes, such as the GAGA factor (**f**) (ChIP-seq data from [42]). **g–i** Heat maps of the nucleosome occupancy and the two fitted rates  $k_1$  and  $k_2$ , indicating **g** the probability that a site is occupied by a nucleosome, **h** chromatin accessibility to MNase, as measured by the nucleosome release rate, and **i** decay rate

the other hand, *Drosophila* has multiple tissues, and each cell type has distinct sets of active genes and silent genes, which makes S2 cells an ideal setting for testing the prediction of our nucleosome positioning model (Fig. 10a).

To test this prediction, we sorted all *Drosophila* genes by their transcription level as measured by CAGE-seq (data from [40]) and aligned the genes at their TSS (Fig. 10b). Then, we plotted the normalized nucleosome dyad distribution, using the same alignment (Fig. 10c),

which confirmed the prediction of our biophysical model. The active genes, containing barrier complexes required for gene activation at their promoters, have a distinct nucleosome organization from silent genes, as predicted by our model (Fig. 10a). While the positions of nucleosomes that are found on silent genes are determined by weak DNA sequence preferences, the positions of nucleosomes found near other strong DNA-binding proteins from promoters are determined by an interplay

of steric exclusion by the barrier complexes, statistical positioning, and the action of chromatin remodelers [12]. To confirm that the barrier complexes are present in the promoters of the active genes, we analyzed DNase-seq data [41] and ATAC-seq data [22] (Fig. 10d, e) which are usually used to detect the active functional regions and the presence of regulatory factors. Moreover, we analyzed the distribution of GAGA factor, which is one of the most abundant DNA-binding proteins in *Drosophila*. As expected, the GAGA factor was detected at the promoters of genes characterized by NDRs and phased nucleosomes, suggesting that GAGA factor is one of the components of the barrier complexes in flies.

The fitted parameters of our kinetic model of chromatin digestion by MNase (nucleosome occupancy  $O$ , nucleosome release rate constant  $k_1$ , nucleosome decay rate constant  $k_2$ ) confirmed that the most active genes have nucleosome-depleted promoters and the inactive genes have nucleosome-filled promoters (Fig. 10g). Also, the increased accessibility that we obtained for the gene promoters (Fig. 10h, i) correlate with the previous results obtained by DNase-seq and ATAC-seq (Fig. 10d, e).

Taken together, genomic data supports a statistical positioning model, in which barrier complexes occupying promoters of active genes phase nearby nucleosomes resulting in a stereotypical nucleosome organization of regular arrays of well-positioned nucleosomes flanking the nucleosome-depleted promoters.

#### **Nucleosome positions in S2 and Kc167 cells are very similar, but their occupancies are cell type-specific**

To validate the current protocol and method of analyzing nucleosome positions and occupancy, we repeated the experiments in a second biological replicate and also in a different cell line (Kc167 cells). All the results were confirmed using the new sets of experiments. Moreover, the typical nucleosome positions that we identified in the two cell types showed a good agreement. In all experiments, we identified > 677,000 nucleosomes, and about 80% of the nucleosomes have similar positions in both cell types (shifted by 20 bp or less). When nucleosome positions identified in the same cell type (different biological replicates) were compared, about 93% of the nucleosome positions found in S2 cells were shifted by 20 bp or less, and about 87% of the positions were shifted by at most 10 bp. Interestingly, when we compared the nucleosome occupancies at the common positions that were identified in the two cell types, we obtained a Pearson correlation coefficient of only 0.36 (replicate 1) and 0.30 (replicate 2). These are substantially lower than the Pearson correlation between the occupancies estimated from different replicate experiments using the same cell type (> 0.78). This suggests

that although the nucleosomes occupy about the same positions in S2 and Kc167 cells, the fractions of cells that contain a nucleosome at these positions (nucleosome occupancy) is very different in the two cell types.

#### **Discussion**

It has been traditionally assumed that in MNase-seq experiments, the nucleosome fragment counts along the genome are proportional to the fraction of cells containing a nucleosome at the corresponding locus. More recently, it was observed that the nucleosome fragment counts obtained in different experiments depend on the specific digestion conditions and on the DNA sequence of each nucleosome [19, 21], but a rigorous theory that explains the variability observed in these experiments was still missing.

The *MACC* protocol [21], which separates nucleosomes into two distinct classes—accessible (characterized by a positive *MACC* score) and inaccessible (characterized by a negative *MACC* score)—partially addresses the issue of distinguishing nucleosome occupancy from DNA accessibility. However, this simple classification was limited to mononucleosomes recovered in a digestion series and could not account for nucleosome losses that occur with over-digestion. Indeed, we observed considerable heterogeneity between loci with respect to MNase sensitivity. Whether or not nucleosome losses seriously affected their conclusions could not be ascertained in the absence of a spike-in calibration standard, which is required for quantitative comparison of samples in genome-wide profiling experiments [43, 44]. Specifically, when digestion begins, all nucleosome counts are increasing with the level of digestion (more nucleosomes are released from chromatin), while in the latest stages of digestion, all nucleosome counts are decreasing with the level of digestion (more released nucleosomes are beginning to be destroyed by MNase) [45]. In other words, all nucleosomes can be characterized by a negative *MACC* score during the initial stages of digestion and by a positive *MACC* score during the final stages of digestion.

Here, we have provided a theoretical framework that explains the observed behavior of the nucleosome counts during the entire MNase digestion (from the start, when all nucleosomes are in chromatin, to the end, when all nucleosomes have been destroyed by MNase). We have developed a quantitative way of measuring the fraction of nucleosomes that are released from chromatin (q-MNase-seq), which allowed us to test and validate the predictions of the theoretical model.

#### **Conclusions**

Using the newly developed protocol, we have found that some promoters are weakly protected by MNase-sensitive complexes, which are digested at a much higher

rate compared to normal nucleosomes (one to two orders of magnitude difference between the corresponding  $k_1$  rates). Comparing the nucleosome release rates for nucleosomes originating from different regions of the genome, we found that DHSs are also more accessible to MNase, and nucleosomes from these loci are released almost twice as fast as the nucleosomes away from DHSs. Surprisingly, when we compared the euchromatin and heterochromatin regions, we could not detect major differences between the chromatin accessibility of these regions, as previously reported in human cells [46]. The nucleosome release rates for the chromatin regions annotated as heterochromatin and euchromatin [38] are comparable to each other, indicating that MNase can access nucleosomes in these regions at similar rates. This suggests that, although heterochromatin and euchromatin appear different when observed cytologically at low resolution, at the molecular level, MNase and other proteins can access heterochromatin regions at rates similar to those of accessing euchromatin. Our findings support the conclusions that have been drawn from nucleosome resolution imaging, in which the difference between heterochromatin and euchromatin is in the density of chromatin, but not in higher-resolution features, such as the average diameter of the chromatin fiber [47]. By developing a quantitative protocol for MNase-seq, we are thus able to take advantage of the high penetrability of MNase to distinguish nucleosome position and occupancy from higher-order chromatin properties such as density and compaction.

## Methods

### Biological materials

*Drosophila melanogaster* S2 cells were obtained from the Drosophila Genome Resource Center (Stock #181) and grown in HyClone SFX-Insect media supplemented with 18 mM L-glutamine. *D. melanogaster* Kc167 cells were obtained as a gift from Lucy Cherbas and grown in HyClone SFX-Insect media supplemented with 18 mM L-glutamine and 10% fetal bovine serum.

### MNase-seq

MNase-seq was performed on unfixed cells as described [32] with the following modifications: For each time point, ~2 million cells were suspended in 10 mM HEPES pH 7.4, 0.5 mM PMSF, and 0.5% NP40 on ice in a 166- $\mu$ L volume. MNase was added at a concentration of 2.5 U per million cells. The mixture was warmed to 37 °C before the addition of 3.5  $\mu$ L 100 mM CaCl<sub>2</sub> (to 2 mM) to activate the MNase. Reactions were stopped by the addition of 170  $\mu$ L of 2XSTOP (4.35 mL TM2, 340  $\mu$ L 5 M NaCl, 200  $\mu$ L 0.5 M EDTA, 100  $\mu$ L 0.2 M EGTA, 25  $\mu$ L RNase A, and 2 pg/mL yeast mononucleosomal spike-in DNA). After phenol-chloroform-isoamyl

alcohol extraction and ethanol precipitation, the DNA pellet was dissolved in 50  $\mu$ L 1 mM Tris-HCl pH 8, 0.1 mM EDTA and used directly for 1.5% agarose gel analysis and for making sequencing libraries as described [32]. For yeast spike-in DNA, spheroplast nuclei [48] were digested with MNase down to mostly mononucleosomes and extracted with phenol-chloroform-isoamyl alcohol as described [16].

### DNA sequencing and data processing

Paired-end sequencing (PE25x25) was performed on barcoded libraries using an Illumina Hi-seq 2500. Reads were aligned to the *D. melanogaster* reference genome *dm6* and *S. cerevisiae* reference genome *sacCer3*, using bowtie2 [49] with parameter `--very-sensitive`. Digestion levels and distributions of DNA fragments were analyzed using plot2DO (<https://github.com/rchereji/plot2DO>). Nucleosome counts (defined as the number of *Drosophila* DNA fragments with the length between 100 and 200 bp) and spike-in counts (number of *S. cerevisiae* DNA fragments with the length between 100 and 200 bp) were obtained using MATLAB (Bioinformatics toolbox). Raw genome-wide nucleosome count profiles were normalized by the corresponding spike-in counts, such that the resulting number of spike-ins was 10,000 for every sample. The typical locations of all well-positioned nucleosomes were detected using a custom MATLAB script, using the map consisting of all detected nucleosomes (from all 6 levels of digestion). For every detected nucleosome (~700,000), we used Eq. (2) to fit the distribution of the nucleosome counts as a function of digestion time and obtained the three parameters: nucleosome occupancy  $O$ , nucleosome release rate constant  $k_1$ , and nucleosome decay rate constant  $k_2$ . The non-linear curve fitting for all profiles was done in MATLAB, using the function `lsqcurvefit`. To visualize specific loci, igvtools was used to create tracks (tdf files) for viewing in the IGV browser [50]. Heat maps were plotted in MATLAB using the heatmap function (<http://www.mathworks.com/matlabcentral/fileexchange/24253-customizable-heat-maps>). Additional details of the biophysical model of chromatin digestion by MNase and data analysis are available in the Additional file 1.

### Supplementary information

The online version of this article (<https://doi.org/10.1186/s13059-019-1815-z>) contains supplementary material, which is available to authorized users.

**Additional file 1:** A quantitative investigation of MNase titrations.

**Figure S1.** Dependence of the three species of nucleosomes on the digestion level. **Figure S2.** The predicted apparent nucleosome occupancy depends on the real occupancy  $O$ , and the two rates  $k_1$  and  $k_2$ . **Figure S3.** The logarithm of the apparent nucleosome occupancy has the asymptotic behavior of  $-k_2d$  for  $d \gg 1/k_2$ . **Figure S4.** Overview of a quantitative MNase-seq (q-MNase-seq) experiment. **Table S1.** The



number of paired-end sequencing reads with the length between 100 bp and 200 bp. **Figure S5.** Simulation results. **Figure S6.** IGV snapshot of chromosome 2L: 13,789,750 - 13,794,300. **Figure S7.** Correlation between sets of parameters obtained after different normalizations of the nucleosome counts. (PDF 4086 kb)

**Additional file 2:** Review history. (DOCX 30 kb)

### Acknowledgements

We thank Alan Hinnebusch, Kami Ahmad, Natalia Petrenko, and Josefina Ocampo for the discussions and helpful comments on the manuscript. This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>).

### Review history

The review history is available as Additional file 2.

### Authors' contributions

RVC and SH conceived and supervised the study. RVC developed the biophysical model and performed the data analyses. TDB performed the experiments. RVC, TDB, and SH wrote the manuscript. All authors read and approved the final manuscript.

### Funding

Intramural Research Program of the National Institute of Child Health and Human Development, National Institutes of Health (RVC); Howard Hughes Medical Institute (TDB and SH).

### Availability of data and materials

Raw and processed data: Gene Expression Omnibus [51], RNA-seq data [21], ATAC-seq data [22], CAGE-seq data [40], DNase-seq data [41], and GAGA factor ChIP-seq data [42]. Source code (Custom MATLAB, R, and Bash scripts): GitHub [52] and Zenodo [53].

### Ethics approval and consent to participate

Not applicable

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Division of Developmental Biology, Eunice Kennedy Shriver National Institute for Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892, USA. <sup>2</sup>Howard Hughes Medical Institute and Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

Received: 23 March 2019 Accepted: 4 September 2019

Published online: 13 September 2019

### References

- Kornberg RD. Chromatin structure: a repeating unit of histones and DNA. *Science*. 1974;184:868–71.
- Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*. 1997;389:251–60.
- Han M, Grunstein M. Nucleosome loss activates yeast downstream promoters in vivo. *Cell*. 1988;55:1137–45.
- Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science*. 2005;309:626–30.
- Bai L, Morozov AV. Gene regulation by nucleosome positioning. *Trends Genet*. 2010;26:476–83.
- Jiang C, Pugh BF. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet*. 2009;10:161–72.
- Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC, et al. Nucleosome organization in the *Drosophila* genome. *Nature*. 2008;453:358–62.
- Voong LN, Xi L, Sebeson AC, Xiong B, Wang JP, Wang X. Insights into nucleosome organization in mouse embryonic stem Cells through chemical mapping. *Cell*. 2016;167:1555–70 e1515.
- Lai B, Gao W, Cui K, Xie W, Tang Q, Jin W, Hu G, Ni B, Zhao K. Principles of nucleosome organization revealed by single-cell micrococcal nuclease sequencing. *Nature*. 2018;562:281–5.
- Chereji RV, Ramachandran S, Bryson TD, Henikoff S. Precise genome-wide mapping of single nucleosomes and linkers in vivo. *Genome Biol*. 2018;19:19.
- Kornberg RD, Stryer L. Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res*. 1988;16:6677–90.
- Chereji RV, Clark DJ. Major determinants of nucleosome positioning. *Biophys J*. 2018;114:2279–89.
- Small EC, Xi L, Wang JP, Widom J, Licht JD. Single-cell nucleosome mapping reveals the molecular basis of gene expression heterogeneity. *Proc Natl Acad Sci U S A*. 2014;111:E2462–71.
- Noll M. Subunit structure of chromatin. *Nature*. 1974;251:249–51.
- Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, Wang Z, Wei G, Zhao K. Dynamic regulation of nucleosome positioning in the human genome. *Cell*. 2008;132:887–98.
- Henikoff JG, Belsky JA, Krassovsky K, Macalpine DM, Henikoff S. Epigenome characterization at single base-pair resolution. *Proc Natl Acad Sci U S A*. 2011;108:18318–23.
- Dingwall C, Lomonosoff GP, Laskey RA. High sequence specificity of micrococcal nuclease. *Nucleic Acids Res*. 1981;9:2659–73.
- Horz W, Altenburger W. Sequence specific cleavage of DNA by micrococcal nuclease. *Nucleic Acids Res*. 1981;9:2643–58.
- Chereji RV, Kan TW, Grudniewska MK, Romashchenko AV, Berezikov E, Zhimulev IF, Guryev V, Morozov AV, Moshkin YM. Genome-wide profiling of nucleosome sensitivity and chromatin accessibility in *Drosophila melanogaster*. *Nucleic Acids Res*. 2016;44:1036–51.
- Chereji RV, Ocampo J, Clark DJ. MNase-sensitive complexes in yeast: nucleosomes and non-histone barriers. *Mol Cell*. 2017;65:565–77 e563.
- Mieczkowski J, Cook A, Bowman SK, Mueller B, Alver BH, Kundu S, Deaton AM, Urban JA, Larschan E, Park PJ, et al. MNase titration reveals differences between nucleosome occupancy and chromatin accessibility. *Nat Commun*. 2016;7:11485.
- Mueller B, Mieczkowski J, Kundu S, Wang P, Sadreyev R, Tolstorukov MY, Kingston RE. Widespread changes in nucleosome accessibility without changes in nucleosome occupancy during a rapid transcriptional induction. *Genes Dev*. 2017;31:451–62.
- Teif VB, Beshnova DA, Vainshtein Y, Marth C, Mallm JP, Hofer T, Rippe K. Nucleosome repositioning links DNA (de) methylation and differential CTCF binding during stem cell development. *Genome Res*. 2014;24:1285–95.
- Fan X, Lamarre-Vincent N, Wang Q, Struhl K. Extensive chromatin fragmentation improves enrichment of protein binding sites in chromatin immunoprecipitation experiments. *Nucleic Acids Res*. 2008;36:e125.
- Teytelman L, Ozaydin B, Zill O, Lefrancois P, Snyder M, Rine J, Eisen MB. Impact of chromatin structures on DNA processing for genomic analyses. *PLoS One*. 2009;4:e6700.
- Brogaard K, Xi L, Wang JP, Widom J. A map of nucleosome positions in yeast at base-pair resolution. *Nature*. 2012;486:496–501.
- Marzluff WF, Duronio RJ. Histone mRNA expression: multiple levels of cell cycle regulation and important developmental consequences. *Curr Opin Cell Biol*. 2002;14:692–9.
- Zentgraf H, Franke WW. Differences of supranucleosomal organization in different kinds of chromatin: cell type-specific globular subunits containing different numbers of nucleosomes. *J Cell Biol*. 1984;99:272–86.
- Brahma S, Henikoff S. RSC-associated subnucleosomes define MNase-sensitive promoters in yeast. *Mol Cell*. 2019;73:238–49.
- Chereji RV, Morozov AV. Ubiquitous nucleosome crowding in the yeast genome. *Proc Natl Acad Sci U S A*. 2014;111:5236–41.
- Ramachandran S, Zentner GE, Henikoff S. Asymmetric nucleosomes flank promoters in the budding yeast genome. *Genome Res*. 2015;25:381–90.
- Ramachandran S, Ahmad K, Henikoff S. Transcription and remodeling produce asymmetrically unwrapped nucleosomal intermediates. *Mol Cell*. 2017;68:1038–53 e1034.
- Tillo D, Hughes TR. G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinformatics*. 2009;10:442.
- Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, Segal E. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature*. 2009;458:362–6.

35. Locke G, Tolkunov D, Moqtaderi Z, Struhl K, Morozov AV. High-throughput sequencing reveals a simple model of nucleosome energetics. *Proc Natl Acad Sci U S A*. 2010;107:20998–1003.
36. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015;31:2382–3.
37. Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T, et al. Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature*. 2011;471:480–5.
38. Filion GJ, van Bommel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ, van Steensel B. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell*. 2010;143:212–24.
39. Chereji RV. PhD thesis. Statistical mechanics of nucleosomes. pp. ix, 148 p. New Brunswick: Rutgers, The State University of New Jersey; 2013:ix, 148 p.
40. Rennie S, Dalby M, Lloret-Llinares M, Bakoulis S, Dalager Vaagenso C, Heick Jensen T, Andersson R. Transcription start site analysis reveals widespread divergent transcription in *D. melanogaster* and core promoter-encoded enhancer activities. *Nucleic Acids Res*. 2018;46:5455–69.
41. Ho JW, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn KA, Minoda A, Tolstorukov MY, Appert A, et al. Comparative analysis of metazoan chromatin organization. *Nature*. 2014;512:449–52.
42. Zouaz A, Auradkar A, Delfini MC, Macchi M, Barthez M, Ela Akoa S, Bastianelli L, Xie G, Deng WM, Levine SS, et al. The Hox proteins Ubx and AbdA collaborate with the transcription pausing factor M1BP to regulate gene transcription. *EMBO J*. 2017;36:2887–906.
43. Hu B, Petela N, Kurze A, Chan KL, Chapard C, Nasmyth K. Biological chromodynamics: a general method for measuring protein occupancy across the genome by calibrating ChIP-seq. *Nucleic Acids Res*. 2015;43:e132.
44. Chen K, Hu Z, Xia Z, Zhao D, Li W, Tyler JK. The overlooked fact: fundamental need for spike-in control for virtually all genome-wide analyses. *Mol Cell Biol*. 2015;36:662–7.
45. Levinger LF, Carter CW Jr. Superstructural differences between chromatin in nuclei and in solution are revealed by kinetics of micrococcal nuclease digestion. *J Biol Chem*. 1979;254:9477–87.
46. Schwartz U, Nemeth A, Diermeier S, Exler JH, Hansch S, Maldonado R, Heizinger L, Merkl R, Langst G. Characterizing the nuclease accessibility of DNA in human cells to map higher order structures of chromatin. *Nucleic Acids Res*. 2019;47:1239–54.
47. Ou HD, Phan S, Deerinck TJ, Thor A, Ellisman MH, O'Shea CC. ChromEMT: visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science*. 2017;357(6349):eaag0025.
48. Furuyama S, Biggins S. Centromere identity is specified by a single centromeric nucleosome in budding yeast. *Proc Natl Acad Sci U S A*. 2007;104:14706–11.
49. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
50. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24–6.
51. Chereji RV, Bryson TD, Henikoff S. Quantitative MNase-seq accurately maps nucleosome occupancy levels. Datasets. *Gene Expression Omnibus*. 2019. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE128689>.
52. Chereji RV, Bryson TD, Henikoff S. Quantitative MNase-seq accurately maps nucleosome occupancy levels. Github. 2019; <https://github.com/rchereji/qMNase-seq>. Accessed 10 Sept 2019.
53. Chereji RV, Bryson TD, Henikoff S. Quantitative MNase-seq accurately maps nucleosome occupancy levels. Zenodo. 2019; <https://zenodo.org/record/3379053#.XWYFnc2ZPrI>. Accessed 10 Sept 2019.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.