Published online 27 November 2013
Nucleic Acids Research, 2014, Vol. 42, No. 4 2147–2158
doi:10.1093/nar/gkt1239

Quantifying the role of steric constraints in nucleosome positioning

H. Tomas Rube[1,2] and Jun S. Song[1,2,3,4,*]

[1]Institute for Human Genetics, University of California, 513 Parnassus Avenue, Box 0794, San Francisco, CA 94143-0794, USA, [2]The Eli and Edythe Broad Center of Regeneration Medicine and Stem Cell Research, University of California, 513 Parnassus Avenue, Box 0794, San Francisco, CA 94143-0794, USA, [3]Department of Epidemiology and Biostatistics, University of California, 513 Parnassus Avenue, Box 0794, San Francisco, CA 94143-0794, USA and [4]Department of Bioengineering and Therapeutic Sciences, University of California, 513 Parnassus Avenue, Box 0794, San Francisco, CA 94143-0794, USA

Received July 19, 2013; Revised October 16, 2013; Accepted November 6, 2013

ABSTRACT

Statistical positioning, the localization of nucleosomes packed against a fixed barrier, is conjectured to explain the array of well-positioned nucleosomes at the 5′ end of genes, but the extent and precise implications of statistical positioning *in vivo* are unclear. We examine this hypothesis quantitatively and generalize the idea to include moving barriers as well as nucleosomes actively packed against a barrier. Early experiments noted a similarity between the nucleosome profile aligned and averaged across genes and that predicted by statistical positioning; however, we demonstrate that aligning random nucleosomes also generates the same profile, calling the previous interpretation into question. New rigorous results reformulate statistical positioning as predictions on the variance structure of nucleosome locations in individual genes. In particular, a quantity termed the variance gradient, describing the change in variance between adjacent nucleosomes, is tested against recent high-throughput nucleosome sequencing data. Constant variance gradients provide support for generalized statistical positioning in ~50% of long genes. Genes that deviate from predictions have high nucleosome turnover and cell-to-cell gene expression variability. The observed variance gradient suggests an effective nucleosome size of 158 bp, instead of the commonly perceived 147 bp. Our analyses thus clarify the role of statistical positioning *in vivo*.

INTRODUCTION

Nucleosomes, consisting of 147 bp of DNA wrapping around histone octamers, constitute the repeating subunits of chromatin and modulate the accessibility of DNA. Nucleosome positioning in promoters and enhancers can critically regulate the activity of DNA-binding proteins and, consequently, transcription rates (1). The gene body and regulatory elements display distinct patterns of nucleosome positioning, but the forces that shape local chromatin structure are still largely unknown. Understanding the principles that guide nucleosome positioning thus remains an important problem in biology.

Recent studies highlighted the difficulty and complexity of this problem and demonstrated that no single mechanism alone can explain nucleosome positioning everywhere in the genome, suggesting that different mechanisms may play dominant roles at distinct genomic loci. For example, two salient features of chromatin observed to date are the nucleosome free regions (NFRs) upstream of transcription start sites (TSSs) and flanking arrays of well-positioned nucleosomes (2). On one hand, sequence-dependent physical properties of DNA are critical for establishing NFR; in particular, Poly(dA:dT) sequences are rigid and hinder nucleosome formation (3,4), while, to a lesser extent, certain periodic patterns with period 10 bp may facilitate wrapping around histones (5,6). On the other hand, ATP-dependent chromatin remodeling complexes can override sequence-specific biases and may be necessary for organizing nucleosome arrays, as demonstrated *in vitro* (7).

In addition to site-specific local factors, nucleosomes also interact with each other. To study the biophysical consequences of their interaction, several investigators

*To whom correspondence should be addressed. Tel: +1 217 244 2999; Fax: +1 217 244 2496; Email: songj@illinois.edu
Present address:
Jun S. Song, Departments of Bioengineering and Physics, Institute for Genomic Biology, University of Illinois, Urbana-Champaign, 1206 West Gregory Drive, Urbana, IL 61801, USA.

have modeled nucleosomes as non-overlapping one-dimensional rods, known in statistical physics as Tonks gas (8), under the influence of a one-dimensional potential. Although the predicted nucleosome occupancy generally correlates with *in vitro* data obtained from high-throughput sequencing, predicting the *in vivo* nucleosome positioning (as opposed to occupancy) remains challenging (9–12). Adding inter-nucleosomal interactions can improve the agreement between theory and observation, although a consensus on the set of interactions has yet to emerge (13–16). For example, biophysical effects, such as electrostatic attractions and histone tail interactions, may influence nucleosome positioning, but these effects are currently difficult to characterize genome wide. Every model, however, should account for the steric hindrance interaction between adjacent nucleosomes—the topic of this article—because two nucleosomes cannot occupy the same genomic sequence.

The effect of steric exclusion is especially pronounced near barriers that limit nucleosome movement. In a classic theoretical paper (17), Kornberg and Stryer studied nucleosomes without binding preference, but confined to a finite or semi-infinite genomic region by barriers, e.g. those representing DNA-binding proteins. They predicted that arrays of well-positioned nucleosomes are caused by hard barriers and steric constraints that restrict the movement of densely packed nucleosomes, a phenomenon termed statistical positioning. Along this line, one interpretation of the nucleosome organization at the 5′ end of genes is that the NFR functions as a barrier, and the downstream nucleosomes are positioned through statistical positioning. High-throughput microarray and sequencing data have recently been used to either support or refute this hypothesis (7,18–21).

In support of the hypothesis, the gene length dependence of the nucleosome occupancy is similar to that predicted by statistical positioning (20). Furthermore, the nucleosome density aligned at TSS and averaged across genes exhibits an oscillating pattern with sharp peaks close to the TSS and decreasing amplitudes farther away (18). Möbius and Gerland (21) rigorously compared this pattern to the density predicted by statistical positioning and found a striking agreement, assuming that each nucleosome occupies 147 bp. Although this analysis is compelling, we demonstrate that the observed averaged density can be an artifact of effects other than statistical positioning.

In contrast, challenging statistical positioning, Zhang *et al*. (7) found that reconstituted nucleosomes incubated with whole-cell extract can form the arrays seen *in vivo* only if ATP is added, advocating a model where nucleosomes are actively packed toward a barrier at the 5′ end of genes. Although this packing model differs from statistical positioning in that the pressure of the Tonks gas is modeled to be high only at the 5′ end of genes, it is similar in that the array of positioned nucleosomes results from the combination of a barrier, steric constraints and high nucleosome density.

The role of statistical positioning thus needs to be resolved in light of these recent studies which provide seemingly conflicting views. One approach to probing the role of steric constraints is to analyze the fuzziness, i.e. the variance in position, of individual nucleosomes. Mavrich *et al*. (19) found that the average fuzziness increases downstream of the TSS, in agreement with the heuristics of statistical positioning. This paper extends that analysis by quantitatively comparing the fuzziness profiles of individual genes to the predictions of statistical positioning. We first generalize the concept of statistical positioning to quantify how a restriction on one nucleosome influences the surrounding nucleosomes through steric hindrance. The rate at which this influence diminishes with genomic distance is captured by a quantity we term the 'variance gradient', defined to be the rate of change in fuzziness between adjacent nucleosomes. We present analytic formulae describing the variance of nucleosome locations in the general cases of both moving and fixed barriers for finite and semi-infinite nucleosome arrays. Irrespective of the nature of the restriction, the generalized model of statistical positioning predicts a constant variance gradient for long genes, with the magnitude of the gradient set by the size of the gap between regions occupied by nucleosomes. This relation between the variance gradient and gap length generalizes to the case of nucleosomes packed against a barrier by constant forces, highlighting that the notions of statistical positioning and nucleosome packing can be combined into a single theory and, thus, partially resolving the competing view points.

We then test the variance-gradient predictions against the nucleosome architecture of gene bodies experimentally observed in *Saccharomyces cerevisiae* (22). Our calculation shows that the variance gradient in individual genes is much lower than that expected for nucleosomes occupying just 147 bp. The observed variance gradient instead indicates a steric exclusion length of 158 bp. One interpretation of this result is that three-dimensional steric constraints prevent or disfavor the unbound DNA close to a nucleosome from being wrapped by the neighboring nucleosomes, thereby extending the steric footprint of the nucleosome. Because the effective length of steric hindrance determines how a restriction imposed on one nucleosome, either by chromatin remodeling or DNA sequence preference, affects other nucleosomes in an array, this finding may help explain how local chromatin structure is established *in vivo*. Our work highlights the need for more rigorous quantitative models to parallel the development in experimental data generation and warns against the common practice of averaging data to represent chromatin structure.

MATERIALS AND METHODS

Statistical positioning predicts constant variance gradients for long genes

Kornberg and Stryer (17) argued that DNA-binding proteins can act as barriers and cause the appearance of nucleosomes well positioned across a population of cells. In its simplest form, statistical positioning refers to a model in which N nucleosomes are represented as finite segments of length d and placed at random in a region of

length w, as shown in Figure 1A. The nucleosomes may be placed in any non-overlapping configuration within the prescribed region, and all allowed configurations are given equal probability.

Interestingly, the probability of finding a nucleosome centered at position x is non-uniform and oscillates with decreasing amplitude and sharpness away from the barriers (Figure 1B). If nucleosomes can be clearly resolved and labeled (+1 at the barrier), this statement can be made quantitatively sharper by considering the marginal probability $P_k(x)$ of finding the kth nucleosome centered at position x. To find $P_k(x)$, note that the problem of finding distinct configurations is equivalent to partitioning the $L = w - Nd$ base pairs uncovered by nucleosomes into $N+1$ gap regions of non-negative integer lengths $L_1, \ldots, L_{N+1}$, where d is the effective nucleosome size. The distribution of the total gap length $Y_k = \sum_{i=1}^{k} L_i$ to the left of the kth nucleosome is then

$$P(Y_k = y) = \frac{\binom{y+k-1}{k-1}\binom{L-y+N-k}{N-k}}{\binom{L+N}{N}}; \quad (1)$$

see Supplementary Methods for details. Using this distribution to calculate the mean and variance of the dyad position $X_k = (k-1/2)d + Y_k$ of the kth nucleosome gives

$$E[X_k] = k(d+\ell) - d/2$$
$$\sigma_k^2 = \mathrm{Var}[X_k] = \frac{k(N+1-k)}{N+2}\ell(\ell+1), \quad (2)$$

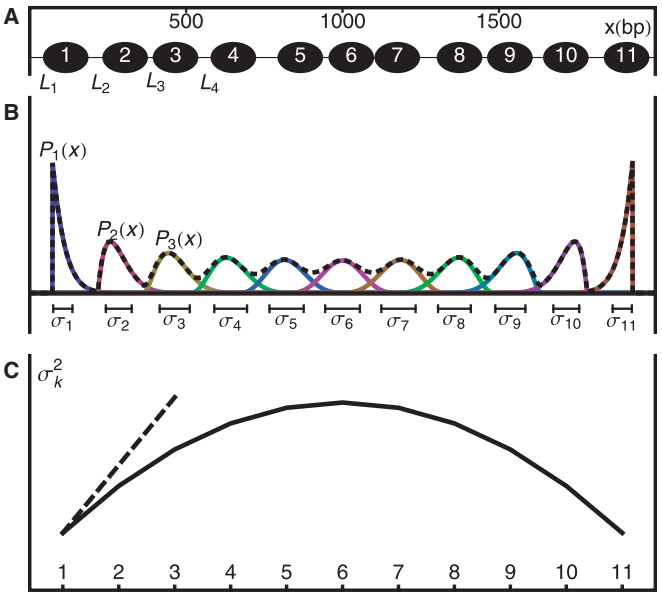


Figure 1. Hard barrier statistical positioning. **(A)** Example of a valid configuration. Nucleosomes are represented as non-overlapping intervals of width d constrained to lie between two infinite potential barriers. **(B)** The probability $P_k(x)$ of finding the kth nucleosome centered at x and the associated variance σ_k^2. Dashed black line is the mixture probability. **(C)** Solid line shows how σ_k^2 varies between two barriers. Dashed line indicates the asymptotic variance for the single barrier case, as the right barrier is moved to infinity while fixing the average nucleosome density constant.

where $\ell = L/(N+1)$ is the mean gap length (see Supplementary Methods for details). The variance expression simplifies further as the number of nucleosomes increases with the mean gap length ℓ kept fixed:

$$\lim_{N\to\infty} \sigma_k^2 = k\ell(\ell+1). \quad (3)$$

In this limit, statistical positioning predicts that the fuzziness σ_k^2 increases linearly with the nucleosome index k, with the slope depending only on the average gap length. We call this slope, $(\sigma_{k+1}^2 - \sigma_k^2) = \ell(\ell+1)$, the 'variance gradient'.

The length d is the minimal inter-nucleosome distance allowed by steric constraints, and it is related to the mean gap length ℓ and the nucleosome density ρ through $\ell = \frac{N}{N+1}(\rho^{-1} - d)$. Several hitherto-proposed models of nucleosome positioning assume this length d to be equal to the 147 bp wrapped in histones, in which case ℓ is the mean linker DNA length (9,10,11,21). However, stronger steric constraints, forcing adjacent nucleosomes to have non-zero linker DNA, would violate this assumption. In that case, the effective steric exclusion length d is > 147, and the effective gap size ℓ between nucleosomes is smaller; consequently, the observed variance gradient is lower than the case $d = 147$. Conversely, the gap length ℓ can be estimated from the observed variance gradient; and then, together with the observed inter-nucleosome distance, one can computationally infer the effective nucleosome size d, as described below.

While the above analysis assumes a fixed number N of nucleosomes, N may fluctuate in biological systems. Such variations can be modeled using the grand canonical ensemble (GCE), which allows N to vary and weighs each valid configuration by $e^{N\mu}$, where μ is the chemical potential. Note that this assumption does not affect the single barrier result, as the GCE is equivalent to the canonical ensemble in the thermodynamic limit, with $\mu = \ln[(\ell+1)^{d-1}/\ell^d]$; see Supplementary Methods for details. For a pair of barriers, multiple N values can give contributions of similar magnitude, resulting in the appearance of dephased nucleosomes, as shown in Supplementary Figure S1A and B (20). Such mixing occurs for only certain values of w, and the occurrence is attenuated with increasing μ. For example, for $\mu = 2$, one nucleosome count dominates ($\max_N P(N) > 0.9$) for 64% of interval length w in the range $w \in [200, 1000]$; for $\mu = 8$, however, this fraction is 93% of the range, and the nucleosome count does not fluctuate much (Supplementary Figure S1C). Supplementary Figure S1D and E shows that Equation (2) is still accurate in a broad range of interval length w for which one nucleosome count dominates the GCE. We shall subsequently demonstrate that the chemical potential in *S. cerevisiae* is actually very high in long genes, further justifying our simpler canonical ensemble approach.

Actively packed nucleosomes

A key assumption of statistical positioning is bidirectionality; non-overlapping nucleosomes move freely in both directions. However, reconstituted nucleosomes

incubated with whole-cell extract require ATP to form ordered arrays; and, within the arrays, the nucleosomes preferentially accumulate at the 5′ end of genes, suggesting that nucleosomes are unidirectionally packed against a barrier at the 5′ end (7). Furthermore, maps of nucleosome movement reveal that ancestral nucleosomes accumulate at the 5′ end of genes, and quantitative modeling suggests transcription-related retrograde nucleosome movement may be the cause of the directionality (23).

To model the effect of actively packed nucleosomes, consider a generalization of the single flexible barrier model, where the nucleosomes are pushed against the barrier by position-independent forces f_k. In this model, each valid configuration has a relative weight $e^{-\sum_k X_k f_k}$, where X_k is the distance between the kth nucleosome and the barrier. The forces modulate the nucleosome spacing and variance gradient, the effects that can be described in terms of the pressure p_k on the kth nucleosome exerted by the $k' > k$ nucleosomes and the packing force f_k. It can be seen that p_k satisfies the recursion relation

$$p_k = p_{k+1} + f_k \tag{4}$$

and that

$$e^{-p_k} = \frac{E[L_k]}{E[L_k]+1}, \tag{5}$$

from which it follows that

$$\sigma_k^2 = \sigma_{k-1}^2 + E[L_k](E[L_k]+1); \tag{6}$$

see Supplementary Methods for details. Equation (4) indicates that the pressure change is the greatest near regions where the packing force is the greatest and that in the absence of any packing force, the pressure is constant across nucleosomes. In turn, Equation (5) suggests that higher pressure on the kth nucleosome leads to a smaller gap relative to the $(k-1)$ th nucleosome and that in the absence of any packing force, the expected gap distance between adjacent nucleosomes is constant. Finally, combined with these observations, Equation (6) implies that the variance gradient of nucleosome positioning is small near regions with high packing force and increases as the packing force decreases, eventually becoming constant as the packing force vanishes. Small forces acting on all nucleosomes along a gene will give a slowly varying pressure and a nucleosome distribution that is similar to that of statistical positioning near the 5′ end. Large forces applied only on the 5′ most nucleosomes will, on the other hand, give rapidly increasing variance gradient and nucleosome spacing away from the 5′ end. Thus, while statistical positioning and active nucleosome packing differ in the source of the nucleosome pressure, they share the common feature that the variance gradient is a reflection of the gap between the regions sterically occupied by nucleosomes.

Statistical positioning with moving barriers

In its original form, statistical positioning only considers fixed barriers. However, a number of factors can affect the nucleosome distribution at gene boundaries and make barrier nucleosomes neither free nor fixed: (i) ATP-dependent remodeling may move the first nucleosome in a gene, (ii) Poly(dA:dT) sequences common in the NFR increase the free energy of nucleosome formation and (iii) the nucleosome promoter architecture may vary in an ensemble of cells. To model such 'soft' barriers, we consider $N+2$ nucleosomes, where the first and last nucleosomes act as moving barriers. We assume that the barrier nucleosomes, indexed as the 0th and $(N+1)$ th, are positioned according to a joint probability distribution $P(X_0, X_{N+1})$ and that the in-between nucleosomes are distributed through statistical positioning conditioned on X_0 and X_{N+1}. The variance of the kth nucleosome, $k = 0, \ldots, N+1$, is then:

$$\sigma_k^2 = \frac{k(N+1-k)}{N+2}\left(E[\ell](E[\ell]+1) - \frac{\sigma_L^2}{N+1}\right) + \frac{(N+1-k)\sigma_0^2 + k\sigma_{N+1}^2}{N+1}, \tag{7}$$

where σ_L^2 is the variance of the total gap length (see Supplementary Methods for details). Note that σ_L^2 simplifies to $\sigma_0^2+\sigma_{N+1}^2$ when the barrier nucleosomes move independently. The first term in the above expression has the same parabolic k dependence as in the fixed barrier case, with ℓ replaced by its expected value and offset by a small contribution. The second term is a linear interpolation of the fuzziness of the barrier nucleosomes.

The variance structure downstream of a single soft barrier is found in the large N limit of Equation (7), with mean gap length and σ_L^2 kept fixed:

$$\lim_{N\gg 1} \sigma_k^2 = \sigma_0^2 + k\, E[\ell](E[\ell]+1). \tag{8}$$

A partial restriction on the first nucleosome thus increases the overall fuzziness of the downstream nucleosomes, but leaves the variance gradient unchanged. Furthermore, statistical positioning, in essence, is the statement that a restricted nucleosome also restricts its neighbors and, consequently, that the positions of tightly packed nucleosomes are correlated. Equation (8) thus describes how the organizing effect of statistical positioning degrades with genomic distance. In fact, the constancy of the variance gradient in Equation (8) generalizes to models with more complex nearest-neighbor interactions (see Supplementary Methods for a simple derivation).

Quantifying variance gradient using paired-end sequenced data

Nucleosome position and fuzziness were calculated from paired-end MNase-seq data in *S. cerevisiae* (GEO accession number GSM756482) (22). Nucleosomes were identified by scanning the strand-specific read density profiles for pairs of offset peaks on opposite strands. Peaks were located using the statistic

$$T(x,s) = \frac{2\,n(x,s)}{n(x-50,s)+n(x+50,s)},$$

where $s = +$ and $s = -$ correspond to the positive and negative strands, respectively, and $n(x,s)$ is the number

of s-strand reads starting (for $s = +$) or ending (for $s = -$) in a window of width 50 bp centered at x. A greedy search was performed by (i) finding the first peak at $x_1, s_1 = \arg\max_{x,s} T(x,s)$, (ii) finding an offset peak on the opposite strand at $x_2 = \arg\max_x T(x, -s_1)$, where $x_2 \in [x_1{+}50, x_1{+}200]$ or $x_2 \in [x_1 - 200, x_1 - 50]$ depending on s_1 and (iii) iteratively repeating the first two steps to find new pairs of peaks under the constraint that new peaks are at least 100 bp away from previous peaks and have $T \geq 1.25$. Next we identified the region of support (ROS) around each peak to be the interval between two flanking troughs at $\arg\min_x T(x,s)$ where $|x - x_{1,2}| \in [50, 120]$. We associated to each nucleosome all paired-end reads starting and ending in the corresponding paired ROS.

MNase digestion is an inherently stochastic process that cuts unprotected DNA fragments randomly, albeit with some sequence bias. Sequencing provides direct information about the locations of flanking cleavage sites, denoted as forward F and reverse R random variables; but, importantly, the dyad location X itself is not observable. The locations F and R are related to X via offset random variables Δ_F and Δ_R as $F = X - \Delta_F$ and $R = X{+}\Delta_R$. Paired-end sequencing provides an empirical joint distribution of F and R, from which one can compute the covariance $\mathrm{cov}(F,R)$. If we assume that the distance between a cleavage site and the dyad of a nucleosome is independent of the genomic location of the nucleosome, then Δ_F and Δ_R are uncorrelated with X. If we further assume that MNase digestion on one side of the nucleosome does not influence the digestion on the other side, then Δ_F are Δ_R are also uncorrelated. Under these assumptions, we obtain $\sigma_X^2 = \mathrm{cov}(F,R)$. The fuzziness of each nucleosome was estimated using this formula and the set of paired-end reads spanning between the corresponding ROSs. We used the MNase-seq data without size selection (22), but restricted our analysis to reads of length 120–200 bp to minimize the contribution from DNA-binding proteins other than histones. Nucleosomes with less than a quarter of the median number of reads were discarded.

For accurate estimates of nucleosome fuzziness, it is important that the sequenced reads are assigned to correct nucleosomes. Assigning reads to the closest peak is reasonable for well-separated nucleosomes, but it will lead to an underestimation of σ_k^2 for fuzzy nucleosomes, as illustrated in Supplementary Figure S2A. [This effect may explain why the fuzziness approaches a fixed value far from the TSS in (19).] We thus estimate the range $[0, \sigma_{\max}^2]$ of fuzziness where each read can be accurately assigned to a single nucleosome, and denote nucleosomes with greater fuzziness as delocalized. To estimate $\sigma_{\max}$, note that a uniform random variable defined on $[0, w]$ has variance $w^2/12$. Taking w to be the mean inter-nucleosome distance 167 bp, a completely delocalized nucleosome uniformly distributed in the interval will have variance $\sim 2324\,\mathrm{bp}^2$. A conservative estimate (see Supplementary Figure S2B) is to set $\sigma_{\max}^2$ to be roughly half this number, i.e. $\sigma_{\max}^2 \sim 1200\mathrm{bp}^2$, which was sufficiently large to consider up to eight nucleosomes away from barriers.

Nucleosomes were assigned to genes using TSS and transcription termination site (TTS) annotations based on RNA-seq (24), allowing dyads to be at most 50 bp outside the annotated gene bodies. Only genes classified as 'verified' in the Saccharomyces Genome Database (SGD, R61) were used. To prevent ambiguous reads from complicating our analysis, genes overlapping 'dubious' genes or regions classified as long terminal repeats, LTR-retrotransposon, repeat regions, transposable element gene or rRNA as defined by SGD were removed; genes overlapping simple repeats longer than 500 bp as defined in the UCSC Genome Browser (sacCer2) also were removed. The variance gradient a was calculated at the 5′ and 3′ ends of each gene separately by performing least-squares fitting of the truncated linear function $\min(ak{+}b, \sigma_{\max}^2)$ to the observed values of the variance σ_k^2 of the first and last eight nucleosomes in the gene body. The R^2 values are the fraction of variance explained by these fits.

Nomenclature and notation

By +1 nucleosome, we mean the first nucleosome after TSS. By −1 nucleosome, we mean the last nucleosome just before the TTS. For positive k, we count the kth nucleosome downstream from +1; for negative k, we count the kth nucleosome upstream from −1. We use $\hat{N}$ to denote the number of nucleosomes observed from sequencing data.

RESULTS

Averaged read density is not suitable for validating statistical positioning

A commonly used method for analyzing nucleosome mapping data is to align and average the density profiles across many genes at either the TSS (18,19) or the +1 nucleosome (21), yielding an oscillating pattern with decreasing amplitudes, as shown in Figure 2A. Ostensibly, this alignment agrees well with the nucleosome distribution predicted by the single barrier statistical positioning with exclusion length $d = 147$ bp and mean gap length $\ell = 30$ bp, corresponding to a variance gradient of $930\,\mathrm{bp}^2/\mathrm{nucl}$ and in line with the previous study (21).

However, we shall subsequently show that this estimate of variance gradient is an order of magnitude larger than that actually found in individual genes, indicating that this averaging approach is not suitable for studying statistical positioning. Moreover, averaging the nucleosome profile across different genes blurs the distinction between gene-to-gene variations in nucleosome location with respect to TSS (or +1 nucleosome) and nucleosome fuzziness within a single gene, making it difficult to interpret the peaks in the resulting profile. That is, the increasing width of the peaks in Figure 2A can result from either (i) increasingly fuzzy nucleosomes centered at the same relative positions across genes or (ii) sharply positioned nucleosomes that differ in relative positions across genes. Whereas the former is a signature of statistical positioning, the latter could result from a multitude of reasons, such as differences among genes in the remodeling of nucleosome arrays relative to TSS.

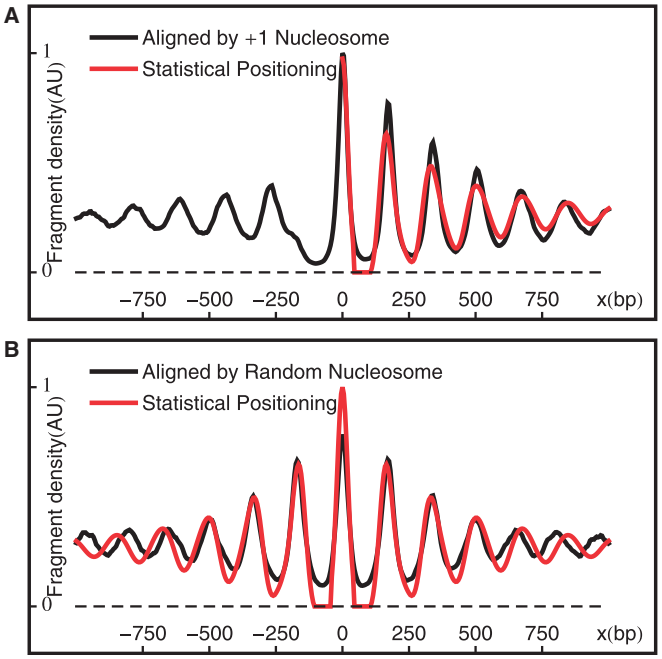


Figure 2. Aligned and averaged density of paired-end read centers. **(A)** Black curve shows the alignment across genes at the +1 nucleosomes. Red curve shows the single-barrier statistical positioning prediction with $\ell = 30$ bp and $d = 147$bp, smoothed with a Gaussian filter. **(B)** Black curve shows the average density aligned at randomly chosen downstream ($k > 1$) nucleosomes. Red curve shows the same statistical positioning model as in (A).

Even if all nucleosomes are equally well positioned, across-gene variations in nucleosome spacing can indeed induce decaying oscillation in averaged nucleosome occupancy, with the variance gradient equal to the variance of spacing across genes. Because the observed variance in spacing across genes is 770 bp^2, comparable to the variance gradient of 930 bp^2 in Figure 2A, most of the decay in oscillations in Figure 2A can be accounted for by the variations in spacing. To further test this fact, we aligned and averaged the nucleosome profile at random nucleosomes sampled from gene bodies, thereby averaging over trends in nucleosome fuzziness away from the randomly chosen nucleosomes. Figure 2B shows that the resulting profile is almost identical to that found by aligning at the +1 nucleosomes and artificially fits the predicted probability distribution of statistical positioning. This analysis shows that the variation in inter-nucleosome distance among genes by itself, without increasing fuzziness in nucleosome positioning away from barriers, is sufficient to generate the damped oscillation pattern that has been previously attributed to statistical positioning.

Low variance gradient is observed at the 5′ end of long genes

To avoid the aforementioned problems associated with averaged read densities, we measured the fuzziness of each nucleosome and analyzed whether its variation along individual genes agrees with Equation (8) (Figure 3A; 'Materials and Methods' section). For long genes with 16 or more nucleosomes, the median fuzziness σ_k^2 stays constant between the +1 and +2 nucleosomes and increases linearly between the +2 and +8 nucleosomes with a variance gradient of 88 bp^2/nucl (Figure 3B). This rate is close to Mavrich *et al.*'s earlier calculation showing that σ_k increases from 20 to 29 bp between the +1 and +5 nucleosomes, corresponding to a variance gradient of 110 bp^2/nucl (19). The observed linear increase in σ_k^2 fits Equation (8) well for $\ell \sim 9$ bp, indicating that single-barrier statistical positioning may explain the array of nucleosomes downstream of the 5′ end of long genes. Because the median distance between well-positioned nucleosomes is 167 bp, this interpretation implies an effective steric exclusion length $d = 158$ bp. Interestingly, nucleosomes can be packed this tightly only with the large chemical potential $\mu = 14$ (Supplementary Figure S4).

To check that the observed low variance gradient is not an artifact of averaging, we calculated the variance gradients in individual genes using linear regression truncated at $\sigma^2_{\max}$, the upper boundary of the range of sensitivity within which we can confidently assign each read to a single nucleosome ('Materials and Methods' section). The distribution of variance gradients across genes with 16 or more nucleosomes is shifted distinctly toward positive values compared to the null distribution of permuted nucleosomes (Supplementary Figure S5; Kolmogorov–Smirnov test $p < 10^{-35}$). For each gene, we also calculated the goodness-of-fit R^2 value of the regression and compared the observed distributions of the variance gradient and R^2 with those obtained by permuting the nucleosomes; 46% of genes have significantly high R^2 (at 5% significance after 100 permutations). Although the median R^2 is high (0.48, compared to 0.08 after permutation), many genes have a relatively poor fit, low variance gradient and poorly positioned +1 nucleosome (Supplementary Figure S5). Because statistical positioning can give arrays of positioned nucleosomes only if the +1 nucleosome is well positioned, we restricted our attention to genes where $\sigma_1^2 < 600$bp^2 (roughly 50% of long genes). This set of genes has a roughly linear fuzziness profile (median $R^2 = 0.68$) and a median variance gradient of 93 bp^2/nucl, consistent with statistical positioning with $\ell \sim 9$ bp (Figure 3C); these values were insensitive to the σ_1 cutoff. In sharp contrast, fitting a statistical positioning model to the averaged read density gave $\ell = 30$ bp and thus a variance gradient of 930 bp^2/nucl, almost one order of magnitude larger than the observed value. This discrepancy further demonstrates the difficulty of studying trends in nucleosome fuzziness using an averaged read density profile.

Nucleosomes at the 3′ end of genes are overall fuzzier than those at the 5′ end, with 49% of the −1 nucleosomes classified as delocalized ('Materials and Methods' section), compared to 18% at the 5′ end. Nevertheless, in genes where the last nucleosome is well positioned, i.e. $\sigma^2_{-1} < 600$bp^2, the quality of fit is high (median $R^2 = 0.74$), and the variance gradient is typically greater than that at the 5′ end (median gradient = 127 bp^2/nucl), as is the average inter-nucleosome distance, suggesting that statistical positioning may play a role also at the 3′ end of this subset of genes, albeit at a lower nucleosome density. It should be noted, however, that the median

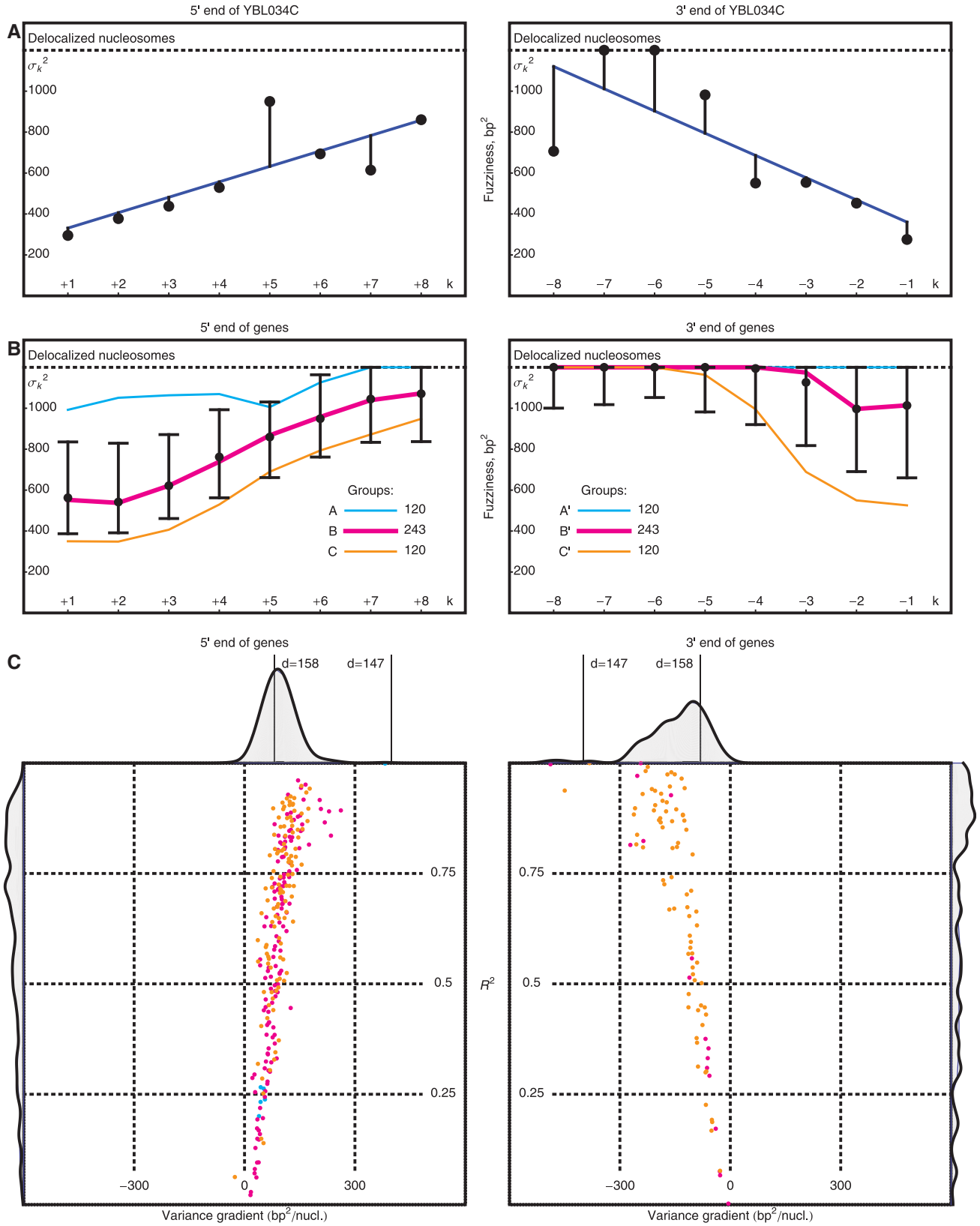


Figure 3. Variation of nucleosome fuzziness along long genes. **(A)** Nucleosome fuzziness (black dots) for the first and last eight nucleosomes in the gene YBL034C. Nucleosomes that are poorly resolved ($\sigma_k^2 > \sigma_{max}^2 = 1200\text{bp}^2$; 'Materials and Methods' section) are designated as delocalized. Variance gradient a for each gene is found by fitting the truncated linear function $\min(ak+b, \sigma_{max}^2)$ to the observed values of σ_k^2 using least squares (blue line). **(B)** Distribution of fuzziness (σ_k^2) for 483 genes with 16 or more nucleosomes. Genes are grouped based on the quartiles of the projection of eight-dimensional fuzziness profiles onto the first principal component at the 5′ and 3′ ends, separately (Supplementary Figure S3).

(continued)

distance between the −1 nucleosome and the TSS of the nearest downstream gene is only 446 bp for genes with $\sigma^2_{-1} < 600\text{bp}^2$, compared to 1200 bp for all genes (see Supplementary Figure S6). This observation suggests that it may actually be the next TSS, and not the TTS, that acts as a barrier at the 3′ end, in line with (21).

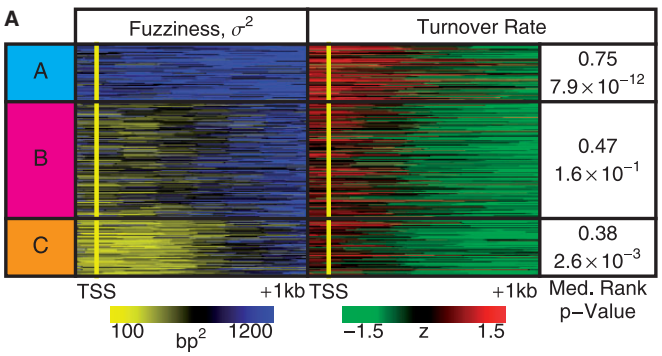
Delocalized nucleosomes are associated with high nucleosome turnover and gene expression variability

To characterize the variation in nucleosome organization across genes, we performed principal component analysis of the fuzziness profiles at the 5′ and 3′ ends separately. The fuzziness profiles projected along the first principal component were then used to group genes: in Figure 3B, Groups A and C contain the genes with fuzziness in the upper and lower quartiles, respectively, while Group B contains the remaining genes (Supplementary Figure S3). Group A is characterized by chaotic nucleosome organization with no clear trend in fuzziness (median $R^2 = 0.13$, median variance gradient = 14 bp^2/nucl) and has 39% of nucleosomes classified as delocalized. Groups B and C have arrays of well-positioned nucleosomes, differing in the fuzziness of the +1 nucleosome, but having similar variance gradients (median of 72 bp^2/nucl and 96 bp^2/nucl, respectively).

Statistical positioning requires the nucleosomes to be in thermal equilibrium. Earlier studies have found two contrasting promoter architectures: promoters with a distinct nucleosome depletion next to the TSS are associated with well-positioned nucleosomes, low histone turnover rate and low transcriptional plasticity, whereas promoters lacking such depletion have a more chaotic chromatin architecture and transcription signature (25). Similarly, to investigate how these activities are associated with statistical positioning, we ranked all genes from 0.0 to 1.0 according to the turnover rate (26) of nucleosomes in the interval [−100 bp, +1 kb] relative to the TSS and calculated the median rank of the genes in each group. Consistent with the fact that Group A exhibits significant deviation from statistical positioning, Figure 4A shows that this group is enriched for high histone turnover rate (median rank = 0.75, $p = 8 \times 10^{-12}$, see Supplementary Methods), suggesting that persistent perturbations at TSS may disrupt nucleosome arrays. Statistical positioning also assumes homogeneous conditions across a cell population, and the strong deviation of Group A genes from the predicted nucleosome fuzziness profile indicated that those genes may be subjected to high cell-to-cell transcriptional variability. Comparing our groups of genes to published measures of variability (27), we found that Group A is significantly enriched for genes with high intrinsic stochastic noise (median rank = 0.68, $p = 9 \times 10^{-4}$) and responsiveness (median rank = 0.67, $p = 1 \times 10^{-5}$), as defined in (27) (Figure 4B). Conversely, Group C, having the most well-positioned nucleosomes, was depleted of genes with high stochastic noise (median rank = 0.3, $p = 5 \times 10^{-4}$). TATA-containing genes are known to be associated with stress response (28) and were highly enriched in Group A relative to Groups B and C (25% in A versus 7% in B and 3% in C, $p = 1 \times 10^{-8}$), in agreement with earlier observations for aligned reads (29).

Fuzziness profile is highly asymmetric between the 5′ and 3′ ends of short genes

Statistical positioning predicts the variance gradient to be lower for a pair of barriers compared to a single



B

	Stochastic Noise	Responsiveness	Stress Response	Trans Variability	
A	0.68	0.67	0.63	0.67	Median Rank
	9e−4	1e−5	1e−3	2e−4	p–Value
B	0.47	0.42	0.45	0.44	
	2e−1	3e−4	4e−2	2e−2	
C	0.3	0.49	0.45	0.52	
	5e−4	7e−1	2e−1	6e−1	

Figure 4. Functions associated with nucleosome profiles. **(A)** Nucleosome fuzziness and histone turnover rate (26) in the [−100 bp, +1 kb] region relative to TSS for the gene groups defined in Figure 3. Genes were ranked between 0 and 1 by their turnover rate in the [−100 bp, +1 kb] region; median rank of each group and its *P*-value were calculated ('Materials and Methods' section). The Group A genes that deviate from statistical positioning have a significantly high median turnover rate. **(B)** Enrichment analysis for gene variability [intrinsic stochastic noise, responsiveness, stress response and trans-variability from (27)]. The Group A genes again show a significantly high level of variability.

Figure 3. Continued
The groups A and A′ correspond to the first quartile, B and B′ to the two central quartiles and C and C′ to the last quartile. Colored lines show the median σ^2_k of the genes in the respective groups. Black vertical bars show the first and third quartiles of fuzziness at each nucleosome. **(C)** Scatter plot of variance gradient [slope in (A)] and R^2 for fitting statistical positioning at the 5′ (left) and 3′ (right) ends of genes, colored by the groups in (B). Only genes with well-positioned +1 or −1 nucleosomes (σ^2_{+1} or $\sigma^2_{-1} < 600\text{bp}^2$) are used in the respective plots. Horizontal and vertical plots show the marginal distributions of variance gradient and R^2, respectively. Vertical lines in the top marginal plots show the theoretical variance gradient predictions from single-barrier statistical positioning for different values of the steric exclusion length *d*, using an average inter-nucleosome distance of 167 bp.

barrier (Figure 1C). It also predicts that as the barrier separation length w increases, the gap length ℓ and fuzziness should both increase until an additional nucleosome is added, whereupon both quantities drop and the cycle restarts (20).

To test these predictions, we use the observed distance $\hat{w}$ between the +1 and −1 nucleosomes as a proxy for w and the number $\hat{N}$ of identified nucleosome peaks as an estimate of N. For $\hat{N} \lesssim 9$, the $\hat{w}$ axis is partitioned into alternating stable regions, dominated by a single $\hat{N}$ value, and transitional regions, where $\hat{N}$ may vary (Figure 5A). In the language of statistical physics, for each $\hat{w}$, we consider a GCE of nucleosomes, where the total number of nucleosomes can fluctuate. The stable regions, centered at $\hat{w}_{\hat{N}}$, are spaced 170 bp apart. Figure 5A shows that the median fuzziness mostly increases with increasing $\hat{w}$ within each stable region, as predicted by statistical positioning. For $\hat{N} = 3, \ldots, 6$, however, the fuzziness increases slightly with decreasing $\hat{w}$ for $\hat{w} \lesssim \hat{w}_{\hat{N}}$, suggesting that a nucleosome may get evicted as $\hat{w}$ decreases and thus that the GCE may be important for these short genes. Taken together, these results show that the confining effect of double barriers can be detected in the nucleosome organization of short genes, in line with (20). To make a quantitative assessment of whether this effect agrees with statistical positioning, we next examine if the observed dependence of fuzziness on $\hat{N}$ and k is consistent with the prediction in Equation (7). Because Equation (7) assumes fixed N, we focus on genes with $\hat{w}$ in the stable regions, i.e. $\hat{w} \approx \hat{w}_N$. For these genes, the median fuzziness profile is highly asymmetric between the 5′ and 3′ ends and has the following three main features: (i) a linear increase from the 5′ end across most of the gene body with variance gradient 76 bp^2/nucl, (ii) a rapid decrease down to 1000 bp^2 at the 3′ end for sufficiently long genes and (iii) constant values between the +1 and +2 nucleosomes (see Figure 5B and Supplementary Figure S7A and B). It needs to be checked whether these features are consistent with statistical positioning.

Statistical positioning predicts that nucleosomes positioned by a pair of sharp barriers should have a symmetric fuzziness profile. Restricting the +1 nucleosome ($\sigma_{+1} < \sigma_{\max}$) suppresses the linearly increasing part, while restricting the − 1 nucleosome ($\sigma_{-1} < \sigma_{\max}$) makes the 3′ drop in fuzziness steeper (see Figure 5C and Supplementary Figure S7C), consistent with statistical positioning. However, the fuzziness continues to increase linearly from the 5′ end into the gene body for a much greater distance than predicted by statistical positioning. The rapid decrease in median fuzziness at the 3′ end is present mostly for long genes ($\hat{N} \gtrsim 10$) where the linear increase exceeds 1000 bp^2; in shorter genes, the characteristic profile is truncated in the linear stage ($\hat{N} \lesssim 7$) or is followed by a plateau at 1000 bp^2. This asymmetric pattern and the extensive linearity of increasing variance at the 5′ end thus deviate from the parabolic variance structure predicted by double-barrier statistical positioning.

More precisely, Figure 5D shows the best fit of Equation (7) to the median fuzziness of short ($\hat{N} \leq 9$)

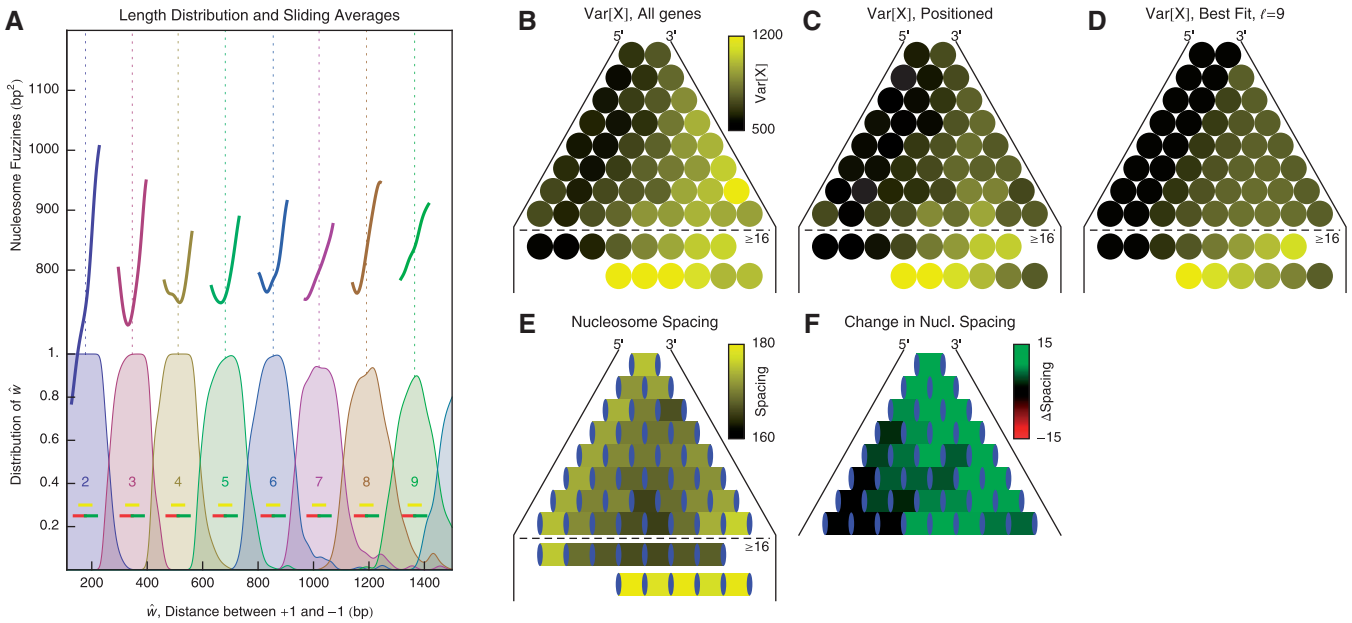


Figure 5. Nucleosome distribution in short genes, organized by the total nucleosome count. **(A)** Top: Sliding average (Gaussian weighted, $\sigma = 15$bp) of fuzziness as a function of $\hat{w}$, the distance between the +1 and −1 nucleosomes and the number of identified nucleosomes $\hat{N}$. Bottom: Distribution of $\hat{w}$ for different values of $\hat{N}$. Dashed vertical lines indicate $\hat{w}_{\hat{N}}$. Horizontal yellow bars indicate genes used in **(B–E)** and red and green bars indicate genes used in **(F)**. (B) Median nucleosome fuzziness of the kth nucleosome (disks, left to right) out of N total (rows), calculated from genes with $|\hat{w} - \hat{w}_N| \leq 20$bp. Last two lines indicate the values for the first eight and last six nucleosomes in long genes with $N \geq 16$. (C) Same as (B), restricted to genes with well-positioned first and last nucleosomes ($\sigma_{+1}, \sigma_{-1} < \sigma_{\max}$). (D) Best fit of the model in Equation (7) to the values in (C), giving parameters $\sigma_{+2}^2 = 560\text{bp}^2$, $\sigma_{-1}^2 = 760\text{bp}^2$ and $\ell = 9$bp. (E) Mean nucleosome spacing (colored boxes) for the set of genes used in (B). Blue ovals represent nucleosomes. (F) Difference in mean nucleosome spacing (colored boxes) between genes slightly longer ($\hat{w}_N < \hat{w} \leq w_N + 40$bp) than w_N relative to genes slightly shorter ($\hat{w}_N - 40\text{bp} \leq \hat{w} < \hat{w}_N$).

genes in Figure 5C with well-positioned +1 and −1 nucleosomes, assuming that $\sigma_{+1}^2 = \sigma_{+2}^2$ and that the term containing σ_L^2 is negligible; the latter assumption is natural for long genes, where the term is inversely suppressed, and it can also apply to short genes if the movement of the first nucleosome causes a back reaction on the last nucleosome. While the agreement with data is decent ($R^2 = 0.34$) and the fitted gap length $\ell = 9$ consistent with the long gene results, this model severely under-estimates the fuzziness profile in the central regions of medium-long genes ($10 \leq \hat{N} \leq 15$) (cf. Supplementary Figure S7C versus Supplementary Figure S7E). Including longer genes in the fit gives $\ell = 12$ and thus alleviates the fuzziness suppression in the middle of gene body, but it still fails to recapitulate the steady linear increase in fuzziness extending along a much greater distance than predicted by statistical positioning. The linear increase observed for a wide range of $\hat{N}$ instead suggests that the nucleosomes are packed against the 5′ barrier and that the effect of a second barrier, located either at the TTS or further downstream, is limited because of low 3′ nucleosome pressure.

Contrary to the even nucleosome spacing predicted by statistical positioning, the nucleosome spacing has been observed to be suppressed at the 5′ end of genes (7). In genes with length $\hat{w}$ in the sable region, we find that nucleosomes are relatively evenly distributed in short genes except for a slight decrease in spacing in the middle of gene body (Figure 5E). For longer genes ($\hat{N} \gtrsim 10$), however, the spacing is contracted away from the 5′ end and then expanded toward the 3′ end (Supplementary Figure S7D), significantly deviating from the prediction of statistical positioning. If this spacing asymmetry results from packing of nucleosomes toward the 5′ end, the resulting pressure variations could also explain the observed differences in variance gradient at the 5′ and 3′ ends of genes.

Finally, the +1 and +2 nucleosomes behave differently from other nucleosomes; they have similar fuzziness, and the spacing between them is larger than that found for downstream nucleosomes (see Supplementary Figure S7D). Comparing the nucleosome spacing between two groups of genes, those with length $\hat{w}$ either slightly above or below the characteristic length $\hat{w}_{\hat{N}}$ (Figure 5F), we find that the spacing between the +1 and +2 nucleosomes increases only slightly and that most additional space is distributed further downstream. This finding suggests that the +1 and +2 nucleosomes are relatively fixed and that those two nucleosomes are not likely to be positioned by statistical positioning.

DISCUSSION

Although the initial theoretical (17) and experimental (30) studies of nucleosome positioning in the vicinity of barriers are more than two decades old, recent experimental progress, based on both high throughput sequencing and alternative experimental techniques (31), has sparked a renewed interest in the role of steric constraints in nucleosome positioning.

This article rigorously examines the theory of statistical positioning and presents a strikingly simple, yet robust, prediction: given that the position of one nucleosome is restricted, the variance in the position of surrounding nucleosomes increases at a constant rate away from the restricted nucleosome. The variance gradient—the rate of this increase—is entirely determined by the average size of the gap between the regions sterically occupied by nucleosomes. While the original formulation of statistical positioning models the restriction as a sharp repulsive barrier, our generalized result is also valid for a partial restriction, may it be due to sequence preference or chromatin remodeling, as long as the adjacent nucleosomes can move freely.

To test this generalized model, we calculated the empirical value of the variance gradient and R^2 in individual genes in *S. cerevisiae*. About 50% of long genes exhibited constant variance gradient at the 5′ end, with a median value of ∼100 bp^2/nucl. In contrast, statistical positioning based on each nucleosome occupying 147 bp at the observed density of one nucleosome per 167 bp would yield a variance gradient of 420 bp^2/nucl. One interpretation of the observed low variance gradient is that nucleosomes effectively occupy around 158 bp and impose steric constraints longer than the 147 bp that are directly bound by histones. This extension of steric exclusion can easily arise from simple geometric considerations of how nucleosomal DNA wraps 1.6 times around a histone octamer, such that two nucleosomes may collide well before they bind adjacent base pairs. Furthermore, just like beads on a necklace can make it semi-rigid, small inter-nucleosomal separations curtail the allowed three-dimensional chromatin configurations, inducing a free energy cost for small separations and widening the region effectively occupied by a nucleosome. Interestingly, most models in the literature describe nucleosomes as non-overlapping finite intervals distributed along the genome according to a free energy potential (9–11). While the methods used for determining the potential differ, the region occupied by the nucleosome is commonly set to 147 bp, potentially underestimating the effect of steric constraints.

Nucleosome breathing, or transient unwrapping of nucleosomal DNA, also affects the effective size of nucleosomes. Cross-species variations in the aligned read density can be explained by letting the average length of unwrapped DNA depend on the species-specific nucleosome density, as previously demonstrated (16). It was found that the free energy potential associated with unwrapping spans 167 bp, in line with our conclusion that the region of steric constraints extends beyond 147 bp. However, it is difficult to infer the dynamics governing nucleosome positioning from the aligned and averaged sequencing read density; if the inter-nucleosomal distance varies across genes while the individual nucleosomes are well positioned, the averaged profile will appear fuzzy, yielding a fictitious profile mimicking the effects of statistical positioning, as demonstrated by aligning random nucleosomes. The fact that the variance gradient observed in individual genes is one order of magnitude lower than that seen in the averaged read density illustrates how the averaging procedure can lead to incorrect

conclusions and demonstrates the importance of testing models of nucleosome positioning in individual genes.

While the steady variance gradient in the 5′ end is consistent with statistical positioning, the pattern of variance gradient in short genes, overall fuzziness and nucleosome spacing in the 3′ end suggest that the full gene body is not described by statistical positioning in its original form. An alternative interpretation of these findings is that ATP-dependent remodeling push nucleosomes toward the 5′ end and cause arrays of well-positioned nucleosomes (7). While differing from statistical positioning in its original form, even in this scenario, the arrays of well-positioned nucleosomes result from high nucleosome density, steric constraints and barriers. One reflection of this similarity is that the relation among steric constraints, nucleosome spacing and the variance gradient for packed nucleosomes in the case of a single barrier, as described by Equations (4–6), is the same as in statistical positioning. Therefore, our interpretation that the low and constant variance gradient reflects strong steric constraints is unchanged even when packing forces are present. The relatively even variance gradient and nucleosome spacing from the +2 through +8 nucleosomes near the 5′ end of long genes indicate that the 'nucleosome pressure' does not decrease abruptly in this range. The 5′/3′ asymmetry in nucleosome spacing is negligible for short genes, but becomes more pronounced with increasing gene length. This observation suggests a small packing force acting on all nucleosomes; the 5′/3′ asymmetry induced by such a force grows with the number of nucleosomes, creating wider nucleosome spacing and higher variance gradient in the 3′ end of long genes.

While arrays of well-positioned nucleosomes are ubiquitous at the 5′ end of genes, the 3′ end has a much higher fraction of delocalized nucleosomes. In short genes, where the effect of a second barrier at the TTS is predicted to be most pronounced and suppress the overall variance gradient to a parabola, the observed median fuzziness actually increases steadily toward the 3′ end. This disagreement indicates that most short genes do not have a sharp barrier at the TTS. In long genes, for which the ordering influence of the distant 5′ barrier is presumably smaller at the 3′ end, the median fuzziness drops toward the 3′ end, suggesting either a partial restriction at the TTS or a barrier further downstream. The latter hypothesis is supported by the correlation between the fuzziness of the last nucleosome and the distance to the closest downstream TSS (Supplementary Figure S6), in line with (21).

The +1 and +2 nucleosomes have very similar fuzziness, and perturbations in gene length leaves the spacing between them fixed, suggesting packing of the +2 nucleosome. However, the large spacing between these two nucleosomes seems to contradict this interpretation. Further studies would be needed to understand how these special nucleosomes are positioned.

Finally, although the simplified model of freely moving nucleosomes may be too crude to account for the full nucleosome landscape, steric restriction will be an important part of any description of nucleosome positioning; fixing the location of one nucleosome will certainly affect the possible configurations of its neighbors. In finding the balance between statistical positioning and other factors, such as histone binding preferences or higher order chromatin structure, this article shows that the effect of the former is characterized by the variance gradient $\ell(\ell+1)$, where ℓ is the mean gap distance between sterically constrained regions. Measuring the variance gradient will thus give important clues about nucleosome dynamics, as demonstrated in *S. cerevisiae*.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Abhinav Nellore, Aaron Diaz and Courtney Onodera for helpful discussion.

FUNDING

NSF CAREER Award [1144866]; NIH [R01CA163336]. Funding for open access charge: NSF CAREER Award [1144866].

Conflict of interest statement. None declared.

REFERENCES

1. Radman-Livaja,M. and Rando,O.J. (2010) Nucleosome positioning: how is it established, and why does it matter? *Dev. Biol.*, **339**, 258–266.
2. Rando,O.J. and Chang,H.Y. (2009) Genome-wide views of chromatin structure. *Annu. Rev. Biochem.*, **78**, 245–271.
3. Iyer,V. and Struhl,K. (1995) Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J.*, **14**, 2570–2579.
4. Anderson,J.D. and Widom,J. (2001) Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Mol. Cell. Biol.*, **21**, 3830–3839.
5. Satchwell,S.C., Drew,H.R. and Travers,A.A. (1986) Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.*, **191**, 659–675.
6. Ioshikhes,I., Bolshoy,A., Derenshteyn,K., Borodovsky,M. and Trifonov,E.N. (1996) Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. *J. Mol. Biol.*, **262**, 129–139.
7. Zhang,Z., Wippo,C.J., Wal,M., Ward,E., Korber,P. and Pugh,B.F. (2011) A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science*, **332**, 977–980.
8. Tonks,L. (1936) The complete equation of state of one, two and three-dimensional gases of hard elastic spheres. *Phys. Rev.*, **50**, 955.
9. Segal,E., Fondufe-Mittendorf,Y., Chen,L., Thastrom,A., Field,Y., Moore,I.K., Wang,J.P. and Widom,J. (2006) A genomic code for nucleosome positioning. *Nature*, **442**, 772–778.
10. Field,Y., Kaplan,N., Fondufe-Mittendorf,Y., Moore,I.K., Sharon,E., Lubling,Y., Widom,J. and Segal,E. (2008) Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.*, **4**, e1000216.
11. Locke,G., Tolkunov,D., Moqtaderi,Z., Struhl,K. and Morozov,A.V. (2010) High-throughput sequencing reveals a simple model of nucleosome energetics. *Proc. Natl Acad. Sci. U.S.A.*, **107**, 20998–21003.
12. Zhang,Y., Moqtaderi,Z., Rattner,B.P., Euskirchen,G., Snyder,M., Kadonaga,J.T., Liu,X.S. and Struhl,K. (2010) Evidence against a

genomic code for nucleosome positioning Reply to [ldquo] Nucleosome sequence preferences influence in vivo nucleosome organization [rdquo]. *Nat. Struct. Mol.Biol.*, **17**, 920–923.
13. Lubliner,S. and Segal,E. (2009) Modeling interactions between adjacent nucleosomes improves genome-wide predictions of nucleosome occupancy. *Bioinformatics*, **25**, i348–i355.
14. Chereji,R.V., Tolkunov,D., Locke,G. and Morozov,A.V. (2011) Statistical mechanics of nucleosome ordering by chromatin-structure-induced two-body interactions. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, **83(5 Pt 1)**, 050903.
15. Chereji,R.V. and Morozov,A.V. (2011) Statistical mechanics of nucleosomes constrained by higher-order chromatin structure. *J. Stat. Phys.*, **144**, 379–404.
16. Möbius,W., Osberg,B., Tsankov,A.M., Rando,O.J. and Gerland,U. (2013) Toward a unified physical model of nucleosome patterns flanking transcription start sites. *Proc. Natl Acad. Sci. U.S.A.*, **110**, 5719–5724.
17. Kornberg,R.D. and Stryer,L. (1988) Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.*, **16**, 6677–6690.
18. Yuan,G.C., Liu,Y.J., Dion,M.F., Slack,M.D., Wu,L.F., Altschuler,S.J. and Rando,O.J. (2005) Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science*, **309**, 626–630.
19. Mavrich,T.N., Ioshikhes,I.P., Venters,B.J., Jiang,C., Tomsho,L.P., Qi,J., Schuster,S.C., Albert,I. and Pugh,B.F. (2008) A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.*, **18**, 1073–1083.
20. Chevereau,G., Palmeira,L., Thermes,C., Arneodo,A. and Vaillant,C. (2009) Thermodynamics of intragenic nucleosome ordering. *Phys. Rev. Lett.*, **103**, 188103.
21. Möbius,W. and Gerland,U. (2010) Quantitative test of the barrier nucleosome model for statistical positioning of nucleosomes up- and downstream of transcription start sites. *PLoS Comput. Biol.*, **6**, e1000891.
22. Krassovsky,K., Henikoff,J.G. and Henikoff,S. (2012) Tripartite organization of centromeric chromatin in budding yeast. *Proc. Natl Acad. Sci. U.S.A.*, **109**, 243–248.
23. Radman-Livaja,M., Verzijlbergen,K.F., Weiner,A., van Welsem,T., Friedman,N., Rando,O.J. and van Leeuwen,F. (2011) Patterns and mechanisms of ancestral histone protein inheritance in budding yeast. *PLoS Biol.*, **9**, e1001075.
24. Nagalakshmi,U., Wang,Z., Waern,K., Shou,C., Raha,D., Gerstein,M. and Snyder,M. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, **320**, 1344–1349.
25. Tirosh,I. and Barkai,N. (2008) Two strategies for gene regulation by promoter nucleosomes. *Genome Res.*, **18**, 1084–1091.
26. Dion,M.F., Kaplan,T., Kim,M., Buratowski,S., Friedman,N. and Rando,O.J. (2007) Dynamics of replication-independent histone turnover in budding yeast. *Science*, **315**, 1405–1408.
27. Choi,J.K. and Kim,Y.J. (2009) Intrinsic variability of gene expression encoded in nucleosome positioning sequences. *Nat. Genet.*, **41**, 498–503.
28. Basehoar,A.D., Zanton,S.J. and Pugh,B.F. (2004) Identification and distinct regulation of yeast TATA box-containing genes. *Cell*, **116**, 699–709.
29. Ioshikhes,I.P., Albert,I., Zanton,S.J. and Pugh,B.F. (2006) Nucleosome positions predicted through comparative genomics. *Nat. Genet.*, **38**, 1210–1215.
30. Fedor,M.J., Lue,N.F. and Kornberg,R.D. (1988) Statistical positioning of nucleosomes by specific protein-binding to an upstream activating sequence in yeast. *J. Mol. Biol.*, **204**, 109–127.
31. Milani,P., Chevereau,G., Vaillant,C., Audit,B., Haftek-Terreau,Z., Marilley,M., Bouvet,P., Argoul,F. and Arneodo,A. (2009) Nucleosome positioning by genomic excluding-energy barriers. *Proc. Natl Acad. Sci. U.S.A.*, **106**, 22257–22262.