



A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome

Travis N. Mavrich, Ilya P. Ioshikhes, Bryan J. Venters, et al.

Genome Res. 2008 18: 1073-1083 originally published online June 12, 2008

Access the most recent version at doi:[10.1101/gr.078261.108](https://doi.org/10.1101/gr.078261.108)

References This article cites 54 articles, 15 of which can be accessed free at:
<http://genome.cshlp.org/content/18/7/1073.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © 2008, Cold Spring Harbor Laboratory Press

A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome

Travis N. Mavrich,^{1,2} Ilya P. Ioshikhes,³ Bryan J. Venters,^{1,2} Cizhong Jiang,^{1,2} Lynn P. Tomsho,² Ji Qi,² Stephan C. Schuster,² Istvan Albert,² and B. Franklin Pugh^{1,2,4}

¹Center for Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA; ²Center for Comparative Genomics and Bioinformatics, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, Pennsylvania 16802, USA; ³Department of Biomedical Informatics and Molecular & Cellular Biochemistry, Davis Heart and Lung Research Institute, The Ohio State University, Columbus, Ohio 43210, USA

Most nucleosomes are well-organized at the 5' ends of *S. cerevisiae* genes where “–1” and “+1” nucleosomes bracket a nucleosome-free promoter region (NFR). How nucleosomal organization is specified by the genome is less clear. Here we establish and inter-relate rules governing genomic nucleosome organization by sequencing DNA from more than one million immunopurified *S. cerevisiae* nucleosomes (displayed at <http://atlas.bx.psu.edu/>). Evidence is presented that the organization of nucleosomes throughout genes is largely a consequence of statistical packing principles. The genomic sequence specifies the location of the –1 and +1 nucleosomes. The +1 nucleosome forms a barrier against which nucleosomes are packed, resulting in uniform positioning, which decays at farther distances from the barrier. We present evidence for a novel 3' NFR that is present at >95% of all genes. 3' NFRs may be important for transcription termination and anti-sense initiation. We present a high-resolution genome-wide map of TFIIB locations that implicates 3' NFRs in gene looping.

[Supplemental material is available online at www.genome.org.]

Where nucleosomes are located in a genome, their functional significance, and what rules determine their location, have been long-standing issues in chromatin biology (Simpson 1991; Lu et al. 1994; Kiyama and Trifonov 2002; Rando and Ahmad 2007). In the 1980s, Kornberg and colleagues used computational modeling to propose a statistical basis for nucleosome positioning, in which fixed barriers on chromosomes generate adjacent ordered arrays of nucleosomes in an otherwise disordered arrangement (Kornberg 1981; Kornberg and Stryer 1988). Potential fixed barriers include (1) underlying nucleosome positioning sequences in which patterns of certain dinucleotides (e.g., AA/TT) impart DNA bendability and facilitate DNA wrapping around the histone core (Simpson and Stafford 1983; Drew and Travers 1985; Satchwell et al. 1986; Ioshikhes et al. 1996; Widom 2001), (2) adjacent rigid DNA sequences such as poly(dA:dT) tracts that exclude nucleosomes (Iyer and Struhl 1995; Suter et al. 2000; Anderson and Widom 2001; Bao et al. 2006; Lee et al. 2007), or (3) gene regulatory proteins bound to specific DNA sequences which would act as an intrinsic barrier or position a nucleosome barrier (Fedor et al. 1988; Roth et al. 1990; Pazin et al. 1997).

The simple presence of a barrier, on statistical grounds, restricts the number of positions a nucleosome can occupy at locations adjacent to the barrier. The next adjacent nucleosome is then restricted by the statistical limitations imposed upon the preceding nucleosome, and so on. There is a natural order decay of positioning within the nucleosomal array that depends upon

nucleosome density and distance from the barrier (Kornberg and Stryer 1988). This statistical model for nucleosome positioning could, in principle, establish the entire nucleosome organization for a gene with a single well-placed barrier. Alternatively or in addition to a barrier, nucleosome positioning throughout a gene might be imparted by positioning elements associated with each nucleosome. It has been suggested that as much as 50% of the nucleosome positions in *Saccharomyces* can be accounted for solely by sequence preference (Segal et al. 2006). However, it remains to be determined the extent to which sequence-based principles vs. packaging principles shape the organization of nucleosomes in the genome. Here we address the inter-relationship of these potential positioning mechanisms and their relationship to chromatin organization across the *S. cerevisiae* genome.

Results

A genome-wide map of nucleosome locations in *Saccharomyces*

A number of high-resolution genome-wide maps of nucleosome locations in *Saccharomyces* have recently been experimentally determined (Raisner et al. 2005; Yuan et al. 2005; Albert et al. 2007; Lee et al. 2007; Whitehouse et al. 2007; Shivaswamy et al. 2008). All of these maps agree with regard to the basic organization of nucleosomes, but differ in their focus and emphasis. Here we concentrate on the organizing principles of nucleosomes across the *Saccharomyces* genome.

Nucleosomes were crosslinked to DNA in vivo with formaldehyde, trimmed to ~147 bp core particles with micrococcal

⁴Corresponding author.

E-mail bfp2@psu.edu; fax (814) 863-7024.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.078261.108>.

nuclease (MNase), immunopurified with antibodies against tagged histones H3 and H4, and 1,206,057 individual nucleosomal DNAs were sequenced ("ChiP-seq," see Methods). 53,026 consensus nucleosome locations were identified by at least three sequencing reads of >100 bp each. The long sequencing reads afforded by the GS20 genome sequencer (454 Life Sciences) allow nucleosome identification in regions of low complexity DNA sequence (e.g., telomeric regions, transposons, tRNA genes, paralogous genes, etc.), which are problematic for other approaches that employ hybridization (microarrays) or identify short (~25 bp) sequence tags.

A median of 14 sequence reads per nucleosome was attained (Supplemental Fig. S1; Supplemental Table S1), resulting in >93% sequencing coverage. Yeast therefore have ~57,000 nucleosomes, which agrees with recent reports (Lee et al. 2007; Shivaswamy et al. 2008). Nucleosome locations also agreed well with other data sets (Supplemental Fig. S2) (Raisner et al. 2005; Albert et al. 2007; Lee et al. 2007; Shivaswamy et al. 2008).

The distribution of reads and nucleosome locations at genomic sites is presented in Figure 1A and Supplemental Figure S3. Data and nucleosome distributions for any locus in the genome can be visualized at <http://atlas.bx.psu.edu>. The data suggest that nucleosomes do not exist as distinct populations in which roughly half are well-positioned and half are delocalized (Yuan et al. 2005; Segal et al. 2006; Lee et al. 2007), although for purposes of analysis it is often useful to categorize them as such. Instead, nucleosome positions exist as a continuum from low to high delocalization (fuzziness).

Patterns of nucleosome organization across the genome

The arrangement of nucleosomes along chromosomal DNA has several physical properties that include location, occupancy level, fuzziness/phasing, spacing, and width (Fig. 1B; see Table 1 for operational and descriptive definitions). These properties take on significance in the context of chromosomal features (e.g., beginning and end of genes). Figure 1C shows a composite distribution of nucleosomes relative to the 4799 transcriptional start sites (TSS) reported elsewhere (David et al. 2006). A number of

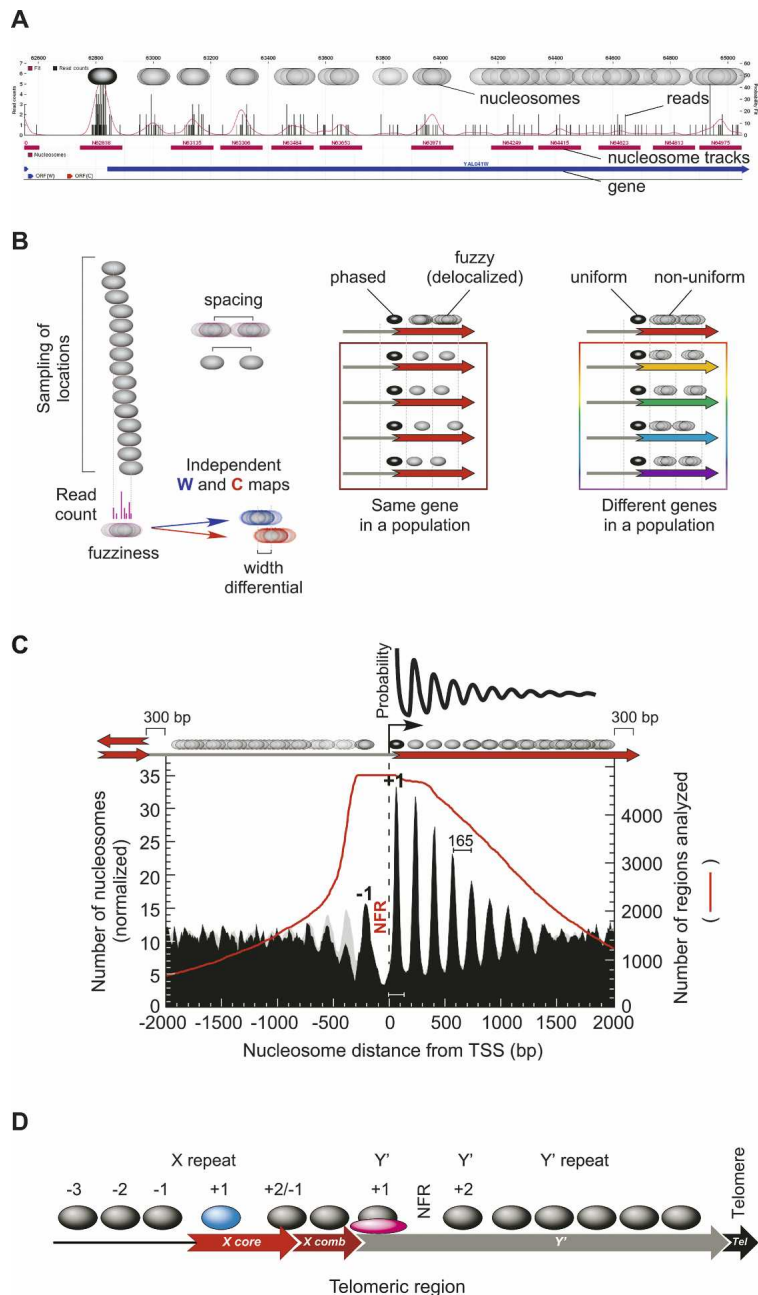


Figure 1. Nucleosome organization around the 5' end of genes. (A) Browser shot showing the distribution of sequencing reads along 2.5 kb of an arbitrary gene (5' end of YAL041W shown in blue). The read count at each coordinate is shown as a bar graph. The distribution was smoothed using a correction factor for MNase bias. Peaks correspond to assigned nucleosome locations based upon a user-defined threshold (purple tracks). Nucleosomes that are found in broad peaks or plateaus are assigned a specific location, although their actual position is delocalized (illustrated by overlapping ovals). Additional browser shots are shown in Supplemental Figure S3 and for any queried locus at <http://atlas.bx.psu.edu/>. (B) Illustration of the physical properties associated with nucleosome positions, which are defined in Table 1. (C) Distribution of nucleosome locations relative to transcriptional start sites (TSS) (David et al. 2006). Nucleosome distances were binned and the count divided by 100, then normalized to the number of regions (red line) present in each bin, and plotted as a smoothed distribution (black-filled plot). In an effort to represent a "pure" pattern, regions <300 bp from an adjacent TSS or TTS (transcript termination site) but also >300 bp from the reference TSS, were removed from the analysis. The same plot was generated without these filters (gray-filled plot, in background). An illustration of the statistical distribution of nucleosomes reported by Kornberg and Stryer (1988) is shown at the top of the panel. (D) Schematic illustration of nucleosome organization at telomeric regions. The arrangement of repeat elements is not representative of all telomeres, but represents a common arrangement. Positions of nucleosomes are approximate and reflect general themes, such as nucleosome-free zones, and noteworthy nucleosomes (colored blue, or bound by another protein in pink).

Table 1. Physical properties of a nucleosome located in a genome

General description		Technical description
Occupancy	Amount of nucleosome binding.	Number of sequencing reads used to assign a nucleosome location
Fuzziness/phasing	A measure of how delocalized or spread out a nucleosome position is	The standard deviation of all read coordinates that contribute to a nucleosome location
Width	A measure of whether the edges of a nucleosome are more or less protected from MNase compared to a standard 147 bp core particle	The number of base pairs between the consensus left and right borders of a nucleosome (measured separately on the W and C strand)
Spacing	Distance between nucleosomes	Number of base pairs between the midpoints of two adjacent nucleosomes
Feature distance	For example, how far a nucleosome is from a transcription start or end site	Number of base pairs between a nucleosome midpoint and a specified chromosomal coordinate

patterns were recognizable that have been described in previous studies (Lee et al. 2004; Sekinger et al. 2005; Yuan et al. 2005; Ioshikhes et al. 2006; Segal et al. 2006; Albert et al. 2007; Shivswamy et al. 2008), and so will not be discussed here. These include (1) a depletion of nucleosomes immediately upstream of the -1 nucleosome at places where adjacent upstream genes are absent (i.e., intergenic regions), (2) a ~ 140 bp nucleosome-free promoter region (NFR), which was found at $\sim 95\%$ of all 5761 genes, (3) a TSS that is ~ 13 bp inside the upstream border of the $+1$ nucleosome, (4) uniform 165 bp spacing of nucleosomes (18 bp linker) near the 5' end of genes that extends to as many as nine nucleosomes, and (5) a loss of uniformity (peak-trough diminution) toward the 3' end. Such properties represent the predominant genome-wide pattern.

The uniformity of positioning with respect to the TSS is much greater here than reported in any of the other studies. Several factors might contribute to this, including immunopurification of crosslinked nucleosomes, MNase bias correction, and methods of peak calling. Since positioning is determined with respect to experimentally measured TSSs, which have inherent experimental error, the actual uniformity of positioning could be stronger than that shown in Figure 1C.

We examined a variety of subclasses of genes as well as chromosomal elements to see whether certain types of genes or elements possessed alternative organizational arrangements (Supplemental Fig. S4). RNA polymerase II subclasses included TATA-containing, TATA-less, lowly transcribed, highly transcribed, cytosolic ribosomal protein encoding, mitochondrial ribosomal protein encoding, sporulation-specific, filamentation-specific, histone H1-enriched. In most cases, the canonical nucleosome organization pattern was observed. However, there were some differences. TATA-containing genes tended to have less of a canonical pattern, which may be related to their greater dependency on chromatin remodeling complexes than most other genes (Basehoar et al. 2004). Genes encoding cytosolic ribosomal proteins were depleted of nucleosomes, and were missing a -1 nucleosome (Supplemental Fig. S4E), probably owing to their extremely high transcription rate. Genes containing histone H1 lacked a consensus organizational pattern around the -1 nucleosome (Supplemental Fig. S4I), suggesting that H1 might be contributing to gene-specific chromatin organization in the upstream regions of promoters. snRNA and snoRNA genes lacked a consensus organization pattern (Supplemental Fig. S4J), while tRNA genes were largely devoid of nucleosomes, although a consensus organization was evident for ~ 1 kb upstream (Supplemental Fig. S4K). Replication origins (ARS) were devoid of nucleosomes, but were bracketed by positioned nucleosomes (Supplemental Fig. S4L).

Yeast telomeric regions are composed of four classes of repetitive elements (as listed in SGD, see Fig. 1D): telomeric (~ 76 bp), Y' (~ 6400 bp), X combination (~ 278 bp), and X core (~ 461 bp). Because of their repetitive nature, telomeric regions have been largely excluded from genome-wide mapping studies using tiled microarrays or short sequence tags, and so little is known about telomeric nucleosome organization. We found telomeric repeats to be relatively depleted of nucleosomes, although a positioned nucleosome was centered ~ 150 bp upstream of the beginning of these repeats (Supplemental Fig. S4M). Y' elements contained well-positioned nucleosomes in which the $+1$ and $+2$ nucleosomes (defined relative to the start of the element) were located at $+80$ and $+440$ bp from the start (Fig. 1D; Supplemental Fig. S4N). As a result there is a ~ 200 bp NFR. Interestingly, the Y' $+1$ nucleosome protects an extra 40 bp of DNA from MNase digestion, suggesting that it is part of a larger stable complex. Canonically spaced and positioned nucleosomes were found upstream of X core elements (designated as -3 , -2 , and -1 positions in Fig. 1D, see also Supplemental Fig. S4O,P). X core elements started at the downstream border of the -1 nucleosome. Inside the X core repeat, the $+1$ nucleosome is extremely well-positioned and normally sized (~ 150 bp), but spaced ~ 250 bp downstream from the -1 nucleosome. The $+2$ nucleosome, which is equivalent to the -1 nucleosome of the X combination repeats, resides another 330 bp further downstream, after which subsequent nucleosomes are normally spaced. Thus, the $+1$ nucleosome of the telomeric X core repeats appears to be unusually isolated and stable (blue oval in Fig. 1D).

Nucleosome organization at the 5' end of genes conforms to the principles of statistical positioning

Figure 1C illustrates two potential examples of statistical positioning emanating from promoters: (1) dense nucleosome packing over genes which drives uniform spacing, but decays further from the barrier (peak-trough diminution between the $+1$ nucleosome and $+2000$ bp), and (2) lower nucleosome density over intergenic regions in which uniform spacing is diminished upstream of the -1 nucleosome. The statistical positioning model further predicts that nucleosome fuzziness on individual genes will increase when moving away from a barrier position (Yuan et al. 2005), such as that presented by the $+1$ nucleosome. Indeed nucleosome fuzziness, as measured by the standard deviation of sequencing read locations that define each nucleosome position (Fig. 1B; Table 1), increased from the TSS to the 3' end of the gene (Fig. 2A, general trend of the trace from the TSS to $+2000$ bp). The decay in phasing, as measured by the change in fuzziness over distance, was three times greater in the upstream

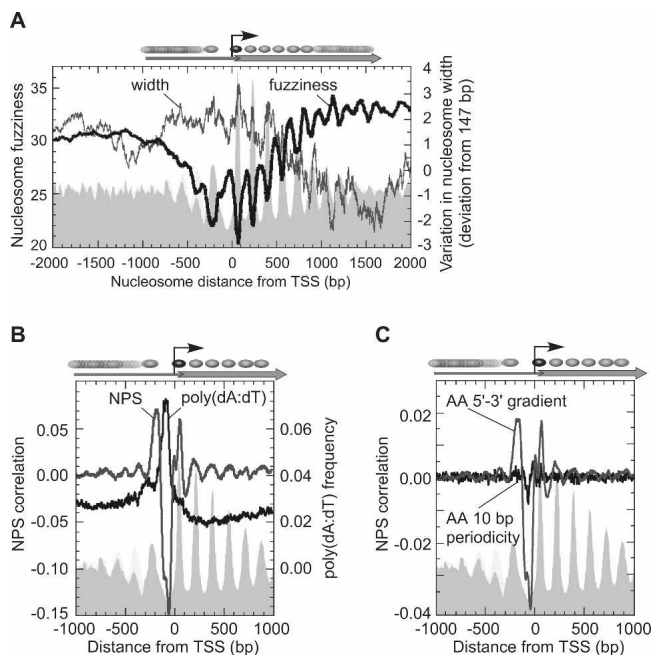


Figure 2. Statistical positioning emanating from barrier nucleosomes. (A) Nucleosome fuzziness and width relative to the TSS. Fuzziness (standard deviation of read locations for those nucleosomes defined by three or more reads) is plotted, against a gray backdrop of Figure 1C. Data are plotted as a moving average of 500 nucleosomes. Deviations from the canonical 147 bp nucleosomal length (in bp) are displayed as a 2000-nucleosome moving average, and represent the distance between the nucleosomal calls made separately on the W and C strand. (B) Averaged NPS correlation score and poly(dA:dT) density for genes aligned by the TSS. The NPS correlation employed an updated AA/TT nucleosomal distribution pattern (Supplemental Fig. S6C). Plot of AAAA/TTTT per bp frequency (11 bp moving average) is shown. Longer poly(dA:dT) tracts gave similar results (data not shown). The gray backdrop reflects nucleosome positions from Figure 1C. NPS scores and fuzziness are not correlated (Supplemental Fig. S7), indicating that the diminishing NPS correlation at distances distal to the $-1/+1$ nucleosome is unlikely to cause the increased fuzziness shown in panel A. (C) NPS correlation profiles broken out by AA dinucleotide enrichment near 5' ends vs. 10 bp periodic spacing of AA dinucleotides.

direction (between the -200 and -800 bp) than in the downstream direction (between the $+100$ and $+700$ bp), which is in accord with statistical positioning at the lower nucleosome density occurring immediately upstream of the -1 nucleosome.

Another expectation of the statistical positioning model is that the lower nucleosome density seen over highly transcribed genes, as reported elsewhere (Bernstein et al. 2004; Lee et al. 2004; Sekinger et al. 2005), would be associated with increased nucleosome fuzziness. Ribosomal protein genes are among the most highly transcribed genes in the genome (Li et al. 1999). These genes displayed a twofold reduction in nucleosome content, and a steep increase in nucleosome fuzziness from the 5' end to the 3' end (Supplemental Fig. S4E). This sharp increase in fuzziness further supports the notion that nucleosomes downstream from the $+1$ position are positioned by the principles of statistical positioning. It is surprising that the $+1$ nucleosome at the ribosomal protein genes remains uniformly positioned and highly phased despite the high level of transcription (Supplemental Fig. S4E). A similar observation was reported elsewhere (Shivaswamy et al. 2008). The stable positioning of the $+1$ nucleosome suggests that the high state of transcription does not

directly cause nucleosome fuzziness, and that the return of a nucleosome to the $+1$ position after passage of RNA polymerase is likely to be an immediate and controlled process.

An intriguing pattern evident in Figure 2A is the undulating periodic fuzziness directed from the TSS toward the 3' end. The peaks coincided with the valleys of nucleosome locations, indicating that at genes where nucleosomes are not at the consensus location, they are significantly delocalized. This delocalization of nucleosomes in consensus linker regions is not artifactual fuzziness associated with randomly distributed contaminating DNA (Supplemental Fig. S5). Thus, the positional relationship of the $+1$ nucleosome to the TSS may be of fundamental importance to genes, to the point where "displaced" nucleosomes are positionally unstable.

Our nucleosome map reports the consensus length of DNA in the nucleosome core particle that is protected from MNase digestion (defined here as nucleosome "width"; Table 1). As shown in Figure 2A and Supplemental Figure S4, nucleosomal width tended to decrease toward the 3' end of genes, suggesting that the DNA entering and exiting the nucleosome may be less tightly associated and thus more MNase susceptible than nucleosomes near the 5' end of genes. "Looser" entry/exits points might be linked in some way to the underlying sequence, altered histone composition and/or modifications, or nucleosome fuzziness that changes from the 5' to the 3' ends of genes (Lieb and Clarke 2005; Yuan et al. 2005), and this might help RNA polymerase II traverse nucleosomes (Li et al. 2005).

DNA sequences that organize nucleosomes are concentrated at promoter regions

Repeating 10 bp patterns of AA/TT dinucleotides create DNA curvature that favors nucleosome formation and rotational positioning of the DNA (Drew and Travers 1985; Satchwell et al. 1986; Shrader and Crothers 1989; Widom 1996; Cohan et al. 2005). Previous experimentally derived frequency distributions of AA/TT dinucleotides across nucleosomal DNA show the expected 10 bp pattern (Ioshikhes et al. 1996; Satchwell et al. 1986). However, within the same pattern there is also an enrichment of AA dinucleotides toward the 5' end of nucleosomal DNA (see Supplemental Fig. S6A), which has not been investigated. Correlations to the experimental AA/TT pattern are strongest at the -1 and $+1$ nucleosome positions (Ioshikhes et al. 2006), suggesting that such patterns contribute to nucleosome positioning in the *Saccharomyces* genome. Whether 10 bp periodicities of AA/TT dinucleotides or 5' enrichment of AA provides the predominant contribution to positioning is determined below.

In contrast to AA/TT patterns, contiguous poly(dA:dT) runs are rigid and exclude nucleosomes in core promoter regions (Iyer and Struhl 1995; Suter et al. 2000; Anderson and Widom 2001; Sekinger et al. 2005; Bao et al. 2006). Therefore, AA/TT NPS and poly(dA:dT) patterns are inversely related. We measured poly(dA:dT) density and AA/TT NPS correlations of genes aligned by TSS. We employed an updated AA/TT NPS pattern that took advantage of our larger collection of well-positioned nucleosomes (Albert et al. 2007). As expected from prior studies (Yuan et al. 2005; Ioshikhes et al. 2006), there was an enrichment of poly(dA:dT) and a negative NPS correlation in the core promoter region (Fig. 2B, from -150 to 0). Thus, the underlying DNA sequence of core promoters might be designed to exclude nucleosomes.

In contrast to the core promoter region, positive NPS corre-

lations aligned with the location of the -1 and $+1$ nucleosomes, as expected. Regions further downstream had AA/TT NPS correlations of significantly lower magnitude and lacked substantial poly(dA:dT) enrichment between nucleosomes, indicating that their direct underlying or adjacent contributions to the highly positioned $+2$, $+3$, and $+4$ nucleosomes (see Fig. 1C) are modest. The combination of a poly(dA:dT)-rich NFR and an NPS-directed $+1$ nucleosome may help create the barrier necessary for statistical positioning of downstream nucleosomes.

When analyzing the iteratively improved AA/TT NPS pattern (Supplemental Fig. S6C), we noticed that the 10 bp periodicity of AA dinucleotides had diminished, and that the slope of the AA frequency distribution had become more pronounced. To explore whether AA 10-bp periodicities or 5' AA enrichment was the predominant parameter in the pattern, we either smoothed out the AA periodicities (using an 11 bp moving average) or eliminated the 5' AA enrichment by symmetrizing the pattern (Supplemental Fig. S6D). As shown in Figure 2C, enrichment of AA dinucleotides toward the 5' end of the nucleosome (and TT enrichment toward the 3' end) was a better descriptor of positioning than 10 bp periodicities of AA/TT. Therefore, we surmise that AA enrichment toward the 5' ends (and TT enrichment toward the 3' ends) provides a larger contribution to positioning than 10 bp periodicities. The 10 bp AA/TT periodicity, which is rather evident in well-positioned nucleosomes (Satchwell et al. 1986; Ioshikhes et al. 2006; Albert et al. 2007), might instead contribute to rotational positioning rather than translational positioning of the -1 and $+1$ nucleosomes.

Nucleosomal dinucleotide patterns are evident for the -1 and $+1$ nucleosomes, but largely absent from genic nucleosomes

Many dinucleotides are distributed nonrandomly across yeast nucleosomal DNA (Albert et al. 2007). We determined the extent to which each of the 16 possible dinucleotides might contribute to nucleosome positioning at -1 , $+1$, and all other genic positions. Figure 3A and Supplemental Figure S8 reveal a nonrandom distribution of certain dinucleotides across genomic regions that have been aligned at the midpoints of 38,200 nucleosomes. Plots have been separated into the -1 , $+1$, and genic nucleosomal groups. We first focus on the nucleosomal region located between -73 and $+73$ bp from the nucleosome dyad. The reciprocal 5'-3' gradient of AA vs. TT dinucleotides was evident for the -1 and $+1$ nucleosomes, and was largely absent from genic nucleosomes as evident by the relatively flat line between -73 and $+73$ bp (Supplemental Fig. S8B).

From these distributions, we calculated that $\sim 8\%$ of all nucleosomal dinucleotides (i.e., those between -73 and $+73$), by virtue of being either present or absent at specific locations, contribute to positioning of the -1 and $+1$ nucleosomes, whereas only $\sim 2\%$ contribute to positioning of the remaining genic nucleosomes (Supplemental Table S4). This translates to ~ 12 and ~ 3.6 dinucleotides per nucleosome for the $-1/+1$ and genic nucleosomes, respectively. Of these, AA/TT made the largest contributions (approximately three and one per nucleosome, respectively). These findings are in agreement with other studies demonstrating a limited role for positioning sequences genome-wide (Lowary and Widom 1997; Peckham et al. 2007; Shivaswamy et al. 2008), and compare favorably with estimates derived from a smaller data set (Bolshoy et al. 1996). The current analysis does not preclude the possibility that the contribution of dinucleo-

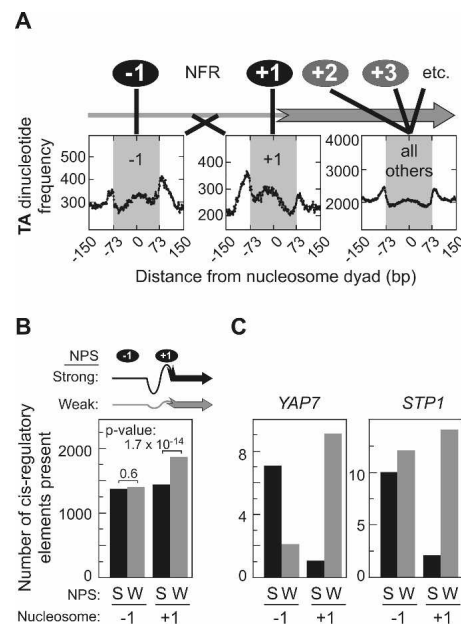


Figure 3. Dinucleotides and *cis*-regulatory elements linked to nucleosome positioning. (A) “TA” dinucleotide distribution across three classes of nucleosomal DNA. From left to right: plots are for -1 , $+1$, and all other genic nucleosomes. To maintain directionality relative to the TSS, only the transcribed strand was compiled. This and the remaining dinucleotide plots are presented in Supplemental Fig. S8. The Y-axis is scaled such that the ratio of the upper and lower range are the same in all plots, allowing them to be compared directly. (B) Enrichment of *cis*-regulatory elements at $+1$ nucleosomes having low NPS scores. See Methods for definition of strong (S) and weak (W) NPS scores. (C) Examples of *cis*-regulatory elements that are enriched at $+1$ nucleosomes having weak NPS scores.

tides associated with the -1 and $+1$ nucleosome is due to sequence bias associated with other properties of the genome that are co-incident with nucleosomes, although we are unaware of any such properties.

AT and TA dinucleotides are particularly enriched at $-1/+1$ nucleosome borders

Many dinucleotides in linker regions (analyzed from ± 73 to ± 90 relative to the nucleosome dyad) were also distributed nonrandomly (Fig. 3A; Supplemental Fig. S8). In particular, AT and TA dinucleotides peaked at nucleosome borders that are adjacent to core promoter regions and to a lesser extent in genic regions. A similar pattern occurred at the linker/nucleosome borders in *C. elegans* (Johnson et al. 2006a). For reasons explained in the Supplemental Figure S8 legend, this is unlikely to be a consequence of MNase preference for cleavage of A/T dinucleotides. Thus, AT and TA dinucleotides mark linker/nucleosome borders to a greater degree than found by chance, but not all or even most nucleosomes are marked in such a way. It may be relevant that TA dinucleotides impart more flexibility into DNA, and AT imparts more rigidity, than other dinucleotides (Packer et al. 2000; Fujii et al. 2007). Such dinucleotides, when present at selective nucleosome borders, might influence nucleosome positioning.

Cis-regulatory elements are enriched at $+1$ nucleosomes that lack strong positioning sequences

DNA sequence-directed nucleosome positioning does not exclude contributions from sequence-specific DNA binding pro-

teins that might position nucleosomes (Fedor et al. 1988; Roth et al. 1990). To explore potential contributions from proteins, we looked for enrichment of *cis*-regulatory element motifs (MacIsaac et al. 2006) in regions having weak intrinsic nucleosome positioning (Ioshikhes et al. 2006). We counted ~1900 conserved *cis*-regulatory elements that reside in the vicinity of the +1 nucleosome for genes that had weak NPS correlations in the same region, and compared it to those with strong NPS correlations. As shown in Figure 3B, genes with weak NPS correlations had 30% more conserved *cis*-regulatory elements than strong NPS genes ($P < 10^{-13}$). For the same collection of genes no difference in *cis*-regulatory elements was detected at the -1 nucleosome ($P = 0.6$). Thus, positioning of the +1 nucleosome may be endowed in part by sequence-specific DNA binding proteins in addition to intrinsic positioning sequences and adjacent exclusion sequences.

The distribution of individual classes of *cis*-regulatory elements is presented in Supplemental Figure S9. Several individual classes of elements stood out. Examples include binding sites for Yap7, which belongs to a class of bZIP stress response regulators (Fernandes et al. 1997), and the zinc-finger transcription factor Stp1 (Fig. 3C) (Wang et al. 1992).

Nucleosome organization at the 3' end of genes conforms to the principles of statistical positioning

To further test the concept of statistical positioning, we examined nucleosome positions near the ends of genes, where nucleosomes are generally fuzzy. If statistical positioning were not occurring, we would expect that in regions of predominantly fuzzy nucleosomes, the infrequent well-positioned nucleosome would represent an isolated event. As such, neighboring nucleosomes would not be influenced by the positioned nucleosome and would therefore be fuzzy.

To avoid potential influence of positioning occurring at the 5' end, short genes (<1.3 kb) were excluded from the analysis. The remaining 2027 genes were partitioned into three groups having either highly phased, moderately phased, or very fuzzy terminal nucleosomes, as assessed by their percent rank of fuzziness (standard deviation of read locations). Nucleosome spacing upstream of the terminal nucleosome (i.e., internal to the genes) was then assessed relative to the terminal nucleosome. As shown by the filled plots in Figure 4A, uniformly spaced nucleosomes (well-defined peaks at -180, -340, and -510, with well-defined intervening troughs) were found adjacent to highly phased terminal nucleosomes (filled plot in the lower panel). Fuzzy terminal nucleosomes lacked such uniformity (filled plot in the upper panel). This provides further evidence for statistical positioning as a general rule governing nucleosome organization.

We next looked at nucleosome fuzziness as a function of distance from the terminal nucleosome (Fig. 4B). Relatively low fuzziness persisted over a greater genomic distance from highly phased terminal nucleosomes than the persistence emanating from very fuzzy terminal nucleosomes. Thus, the infrequent occurrence of a positioned nucleosome far from promoter regions creates a higher occurrence of adjacently well-positioned nucleosomes, which is predicted by the statistical positioning model.

Nucleosome positioning signals at the end of genes are antagonized by cleavage and polyadenylation signals

To assess whether the phased terminal and adjacent upstream nucleosomes were positioned by the underlying DNA sequence,

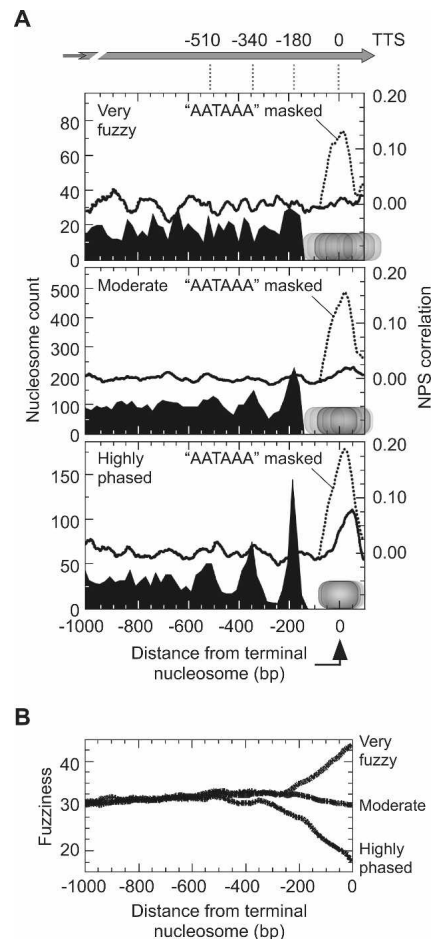


Figure 4. Statistical positioning emanating from positioned terminal nucleosomes. (A) Distribution of nucleosomes upstream of the terminal nucleosome. The terminal nucleosome was defined as the closest upstream nucleosome to the transcript termination site (TTS) (David et al. 2006) for transcripts of at least 1.3 kb. Terminal nucleosomes that were within 500 bp of the TTS and defined by >6 reads (to achieve a statistically robust fuzziness value) were divided into three groups based upon percent rank of fuzziness (<15%, highly phased; 15%–85%, moderate; >85%, very fuzzy). Upstream nucleosomes were binned based upon distance from the terminal nucleosome. Bin counts (left axis) were not normalized to the number of genes analyzed. NPS correlation plots for the same set of genes aligned by the terminal nucleosome are shown as solid traces (right axis). Dashed traces have AATAAA (and related sequences) masked, as described in the Methods. (B) Sliding window analysis of average nucleosome fuzziness as a function of distance from the terminal nucleosome. The low fuzziness around the penultimate nucleosome (from 150 to 250 bp upstream of the terminal nucleosome) was significant at 1 SD but not at 2 SD.

we generated AA/TT NPS correlation profiles of genic regions that were aligned by their terminal nucleosome. Highly phased terminal nucleosomes possessed stronger NPS correlations (solid trace in lower panel of Fig. 4A) than the very fuzzy terminal nucleosomes (solid trace in upper panel of Fig. 4A). NPS correlations were of similar magnitude to those found at the -1 and +1 nucleosomes (Fig. 2B), indicating that AA/TT positioning sequences account for well-positioned terminal nucleosomes. However, AA/TT positioning sequences were relatively flat in adjacent upstream genic regions, indicating that they contribute little toward positioning of adjacent phased nucleosomes, which further implicates statistical positioning.

The NPS correlation peak was not exactly co-incident with the measured location of the well-positioned terminal nucleosome (solid trace in lower panel of Fig. 4A). We investigated whether this discrepancy was due to the presence of cleavage and polyadenylation signals (AATAAA and related sequences) that populate this region. Due to their similarity to poly(dA:dT) tracts, they might be antagonistic to nucleosome positioning (Peckham et al. 2007). When AATAAA and related sequences were masked (see Methods), the NPS correlation peak not only shifted closer to the expected position (dotted traces), but also became stronger and more evident even at fuzzier nucleosomes. Regions further upstream were unaffected. This finding suggests that nucleosome positioning is partially “hardwired” at the end of genes, but is antagonized by cleavage and polyadenylation signals. Antagonism between DNA sequences in positioning the terminal nucleosomes is in line with the notion that the terminal nucleosome may be highly regulated at certain genes, and connected to transcription termination (Alen et al. 2002).

Most *Saccharomyces* genes contain a novel NFR at the end of their genes

While much attention has focused on nucleosome organization at the beginning of genes, little is known about their organization at the end of genes. Interestingly, nucleosome occupancy dropped precipitously at the 3' end of genes, essentially where the cleavage and polyadenylation site is found (Fig. 5A). A similar observation was recently reported elsewhere (Shivaswamy et al. 2008). The median length of these 3' terminal NFRs was 173 bp (distance between flanking nucleosome borders, with the downstream nucleosome typically being a part of the next downstream gene or promoter) (Supplemental Fig. S10). 5232 of 5507 genes (95% of all genes) had an identifiable 3' terminal NFR of >50 bp. Both the 5' and 3' NFR increased in size with highly transcribed genes (Supplemental Fig. S11), indicating that they are dynamic and linked to transcription. The 3' NFR might be a site for disassembling the terminating RNA polymerase, much as the 5' NFR is the site for assembly. An additional possibility is that the 3' NFR is where anti-sense pre-initiation complexes assemble. Consistent with the latter possibility, we identified 71 anti-sense transcripts (Perocchi et al. 2007) that initiated within 300 bp of a 3' NFR midpoint (Fig. 5B). Thus, for some genes the 3' NFR might represent the promoter region for anti-sense transcription.

Selective enrichment of TFIIB at 3' NFRs supports a gene looping model

Recent studies suggest that a terminating RNA polymerase II might recycle directly back to its promoter via DNA looping that involves general transcription factor TFIIB (Singh and Hampsey 2007). Looped genes contain TFIIB at their 5' and 3' ends. However, other initiation factors, such as TBP, reside only at the promoter. To explore whether the 3' NFRs identified here might accommodate such a model, we conducted high-resolution genome-wide location analysis (ChIP-chip) on TFIIB and TBP, using microarray probes spaced every 5 bp throughout the *S. cerevisiae* genome. We searched for genes that met three criteria: (1) The 3' NFR resided in an intergenic region that either lacked a promoter or was far from a promoter, so as to avoid complications associated with TFIIB bound to nearby promoters; (2) TFIIB was present at the promoter, which is a requirement of the looping model (Singh and Hampsey 2007); and (3) TFIIB was present at the 3'

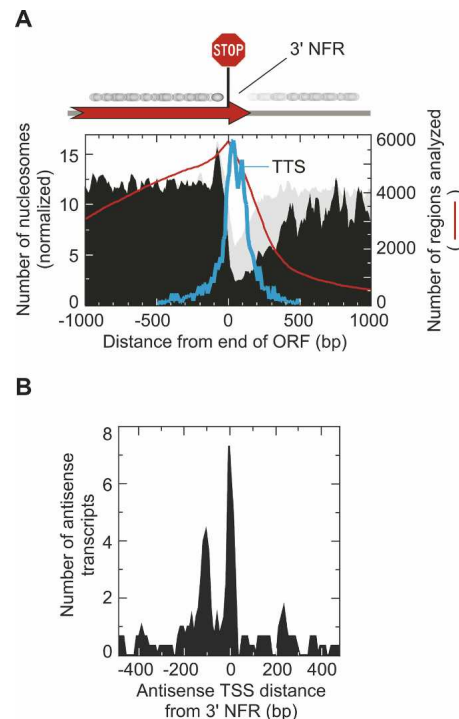


Figure 5. Distribution of nucleosomes and anti-sense TSS around 3' end of genes. (A) Distribution of nucleosome locations relative to the 3' end of open reading frames (ORFs). Nucleosome distances relative to the ORF stop codon were binned in 10 bp intervals, normalized to the number of regions represented in each bin (red line), then smoothed using a three-bin moving average. Regions internal to adjacent genes or within 360 bp of another ORF start site (whichever is closer) were either removed (black fill) or not removed (gray fill) from the analysis. The distribution of polyA sites, where transcripts terminate (TTS), is shown as a blue trace. (B) Distribution of anti-sense TSS distances from 3' NFRs. Anti-sense TSS coordinates were from (Perocchi et al. 2007). Distances were binned (10 bp bin) and smoothed (three-bin moving average).

NFR, but TBP was absent. The absence of TBP selects against 3' NFRs that might be sites of anti-sense initiation.

Remarkably, 120 genes met the looping prediction criteria ($P < 10^{-8}$; Fig. 6A; Supplemental Table S6). At these genes, TFIIB was found just downstream from the cleavage and polyadenylation site (Fig. 6B), which is where RNA polymerase II is likely to terminate transcription. These data meet one prediction of the gene looping model. If reflective of gene looping, the results suggest that the coordination of initiation and termination via gene looping between the 5' and 3' NFRs may be common in yeast, although not necessarily universal. The physiological importance of looping remains to be determined. We found nothing extraordinary about the transcription frequency or gene ontologies of these genes (data not shown). Taken together, our results suggest that the 3' NFR may be as “busy” a place as the 5' NFR during the transcription cycle.

Discussion

We propose that the genomic organization of nucleosomes in *Saccharomyces* largely originates from the 5' end of genes, where a combination of poly (dA:dT) exclusion sequences in the core promoter, AA positioning sequences near the 5' end of nucleosomal DNA (TT near the 3' end), AT and TA dinucleotides near the nucleosome borders, and transcription factor binding sites

define the location of the $-1/+1$ nucleosomes and the nucleosome-free promoter region (Fig. 7). The $+1$ nucleosome establishes a barrier that largely relies on the principles of statistical positioning to position densely packed downstream nucleosomes for as far as ~ 1 kb, with little contribution from the underlying DNA sequence. Most yeast genes are ~ 1 kb or less. Beyond 1 kb, genic nucleosomes lack positional relationships with the TSS and tend to be less phased than the $+1$ nucleosome. The DNA entering and exiting the downstream nucleosomes tends to be more susceptible to MNase digestion, suggesting that the DNA is more loosely associated with these nucleosomes. Conceivably, positioning sequences near the 5' end of genes, or histone modifications/variants that are enriched at the 5' end of genes, could contribute to a "stronger" wrap of the DNA helix on the histone surface.

Relocation of the $+1$ nucleosome by chromatin remodeling complexes would therefore also relocate an array of downstream nucleosomes in the same direction, creating a mechanism for long distance control through a single nucleosome. However, this would need to be tested directly, perhaps by using positioning sequences to reposition a nucleosome and examining its impact on adjacent nucleosomes. Recent studies have shown that the ISW2 complex repositions nucleosome arrays in the yeast genome (Whitehouse et al. 2007). If the principles of statistical positioning are in play, ISW2 needs to act on only one nucleosome in each array in order to reposition the entire array. Nevertheless nucleosome mobility in general may be facilitated by global chromatin remodeling complexes such as RSC (Parnell et al. 2008). A second observation that appears to be a manifestation of statistical positioning is the repositioning of downstream

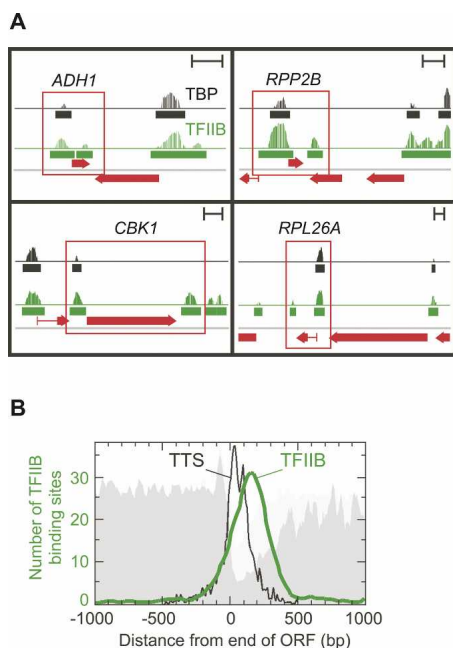


Figure 6. Distribution of TFIIIB around 3' end of genes. (A) High-resolution mapping of TFIIIB at genomic loci. The genome-wide location of TFIIIB (Sua7) and TBP were determined by CHIP-chip using high density Affymetrix tiling arrays (5 bp average probe spacing, 3.2 million probes). Results for four potentially looped genes are shown (boxed in red). Vertical bars reflect binding strength (TFIIIB in green; TBP in black), and horizontal bars represent binding locations. Scales in upper right corner of each box represent 500 bp. (B) Composite distribution of TFIIIB at the end of genes. Plots of nucleosomes (gray backdrop), TTS (black trace), and TFIIIB (green trace) are shown.

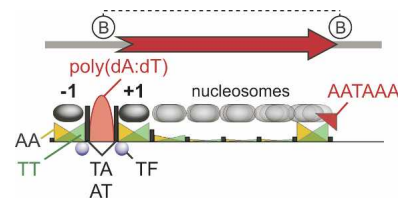


Figure 7. Model depicting various contributors to nucleosome organization in the *S. cerevisiae* genome. Yellow and green mirrored triangles represent increased AA and TT dinucleotide enrichment toward the 5' and 3' ends of nucleosomal DNA, respectively. Their 10-bp periodical placement would rotationally phase the DNA on the nucleosomal surface, although rotational phasing may contribute modestly to translational positioning (Tanaka et al. 1992; Lee et al. 2007). The red half-ellipse represents the distribution of poly(dA:dT) tracts that exclude nucleosomes from the promoter. Black vertical bars represent AT and TA dinucleotide enrichment at nucleosome borders, which might contribute to translational positioning. The contributions of all A/T dinucleotides are diminished beyond the $+1$ nucleosome, in which statistical phasing takes over and decays toward the 3' end of the gene (represented as fuzzier nucleosome ovals in gray). AA/TT positioning at the 3' end of genes is depicted as being antagonized by AATAAA-related cleavage and polyadenylation signals. Transcription factors (TF) are shown to contribute to -1 and $+1$ positioning. TFIIIB-linked looping (β) is shown.

nucleosomes when a paused RNA polymerase II contacts and repositions the $+1$ nucleosome in *Drosophila* (Mavrich et al. 2008).

While the entire length of nucleosomal DNA contributes to its overall affinity for the histone core, by our calculations only a few dinucleotides contribute to positioning. Of these, AA/TT contribute the most, with ~ 3 AA or TT dinucleotides helping position the $+1$ nucleosome. This contribution seems rather small to impart intrinsic curvature over the entire length of nucleosomal DNA. Thus, intrinsic DNA curvature, if present, might be limited to ~ 50 bp regions internal to where the DNA enters and exits the nucleosome. This would allow more stable packaging of the -1 and $+1$ nucleosome.

Toward the 3' ends of genes, nucleosomes progressively acquire more random locations relative to the TSS, and are less phased. Nucleosome occupancy drops precipitously at the 3' end of genes, forming a 3' NFR. The 3' NFR may participate in transcription termination, anti-sense initiation, and recycling of RNA polymerase to the promoter by DNA looping. Thus, as much as the -1 and $+1$ nucleosomes may be important regulators of transcription initiation, the terminal nucleosomes that flank the 3' NFR may be important regulators of transcription termination (Alen et al. 2002). AATAAA and related elements play a clear role in transcription termination by promoting cleavage and polyadenylation of the nascent mRNA transcript. These elements might also facilitate termination by antagonizing nucleosome formation over the termination region. Whether and how the terminal nucleosome of a gene is linked to termination remains to be determined.

Methods

Nucleosome mapping in vivo

Three independent nucleosome preparations were made from BY4741 strains containing C-terminal TAP tags on histone H3 or H4. Maps produced from H3 were indistinguishable from H4 maps, and so the data were combined. Details of formaldehyde crosslinking, MNase digestion, TAP immunopurification, gel purification of mono-nucleosome sized DNA, DNA sequencing using the Roche GS20 (454 Life Sciences), mapping to genomic

coordinates from www.yeastgenome.org, correction for MNase bias, and making calls on nucleosome locations are described elsewhere (Albert et al. 2007), with the following modifications: 0.05% sodium dodecyl sulfate was included in the ChIP. Soluble chromatin was prepared from 3 L of cells, precleared with Sepharose 4B for 15 min at 4°C, then immunoprecipitated with 600 μ L of bed volume IgG Sepharose Fast Flow. Nucleosome core particles were eluted with 3 kU TEV Protease at 25°C for 2 h. Raw sequencing reads can be accessed through NCBI Trace Archives TI SRA001003, Sequencing Center “CCGB.” Bulk downloads or specific queries of nucleosome positions can be accessed from <http://atlas.bx.psu.edu> or Supplemental Table S1.

Nucleosome distributions

For every “feature” start or end coordinate, defined by the *Saccharomyces* Genome Database (as of May 13, 2007), a search region was defined using the following criteria. First, minimum and maximum search distances were set at ± 300 bp and 3000 bp, respectively, relative to the feature coordinate. Second, between the minimum and maximum, the actual range was set to be 300 bp from the next feature. Examples of “the next feature” include start or end coordinates of ARS elements, retrotransposons, telomeres, snoRNAs, tRNAs, and either ORFs or mRNA transcripts (Supplemental Table S2) (David et al. 2006). A 300 bp border from the next feature was chosen so as to minimize inclusion of potential $-1/\text{NFR}/+1$ nucleosomal regions, which tend to be within 300 bp of an ORF start site or TSS. Third, the nucleosome count along the chromosome coordinate (bin size = 20 bp) in the allowed region for each feature was determined and summed across all bins having the same distance from the feature. Fourth, the nucleosome count was normalized to the number of regions in each bin (displayed in Figs. 1C, 5A, and Supplemental Fig. S4), then multiplied by 100. Line plots of three bin moving averages were then drawn representing the normalized bin counts at varying distances from the feature coordinate. Distances were measured to the center of each bin.

NPS correlation plots

NPS correlation profiles were calculated using a derivative of an earlier AA/TT pattern (Supplemental Fig. S6A) (Ioshikhes et al. 1996). The updated pattern was derived by first acquiring 2285 of the most highly positioned H2A.Z-containing nucleosomes identified elsewhere (Supplemental Table S3) (Albert et al. 2007) and ~200 other well-defined nucleosomes (Ioshikhes et al. 1996). The sequence of both DNA strands was filtered to retain a subset that positively correlated with the original AA/TT patterns (Ioshikhes et al. 1996, 2006). From that subset a new empirical distribution pattern of AA and TT dinucleotides was determined (Supplemental Fig. S6B). This new pattern was used to refilter the subset for positively correlating sequences, and a newer AA/TT pattern derived. After six iterations, new patterns originating from the two nucleosome data sets became virtually indistinguishable. The new pattern (Supplemental Fig. S6C) was then used to calculate the NPS correlation profiles as described elsewhere (Ioshikhes et al. 2006). Distances represent the distance from the feature to the center of the NPS correlation pattern. When masking sequences related to AATAAA (at the 3' end of genes), trinucleotides AAT, TAA, and AAA located downstream from the terminal nucleosome were replaced by NNT, TNN, and NNA, respectively, and the NPS correlation was calculated.

Dinucleotide counts

The sum total count of each dinucleotide present at the indicated distance from the nucleosome midpoint was determined (Supple-

mental Table S4). Subgroups were identified as nucleosomes located from -300 to -100 (-1 position), and -100 to $+150$ ($+1$ position). All other genic nucleosomes include those between the $+1$ nucleosome and the ORF end site. Binned counts were smoothed using a three-bin moving average. The fractional contribution of each dinucleotide to positioning relative to the overall dinucleotide count was determined by first calculating the absolute difference between the individual dinucleotide counts (e.g., AA_{-73} , located at the 5' nucleosome border of the transcribed strand) at a given position along the 146 bp nucleosomal DNA and the average individual dinucleotide count (e.g., AA_{ave}) in the 147-bp window (i.e., deviations from the average). These position-specific differences were summed across the 147 bp nucleosomal length of DNA, and divided by the overall dinucleotide count in the 147-bp window (NN_{total}). Thus, the fractional contribution of AA = $(|AA_{-73} - AA_{ave}| + \dots + |AA_{+73} - AA_{ave}|) / NN_{total}$. Summation over all 16 dinucleotides provided the fractional contribution of all dinucleotides to positioning. This value multiplied by 147 provided the number of dinucleotides per nucleosome that contribute to positioning.

Cis-regulatory element enrichment at +1 nucleosomes with low NPS scores

Gene lists and corresponding NPS correlation profiles were taken from Figure 2 of Ioshikhes et al. (2006), in which raw profiles for individual genes had been smoothed (50 bp moving average) then binned in 10 bp intervals (Supplemental Table S5). The standard deviation of the resulting smoothed and binned NPS correlation values was determined for each gene over the interval -100 to $+100$. The genes were then rank ordered by standard deviation of the aggregate NPS correlation score. Genes in the lowest and highest 20th percentile (defined as “Weak” and “Strong,” respectively) were selected for further analysis. For these genes, we counted the number of conserved cis-regulatory elements defined in (Maclsaac et al. 2006) as “c3” (regardless of binding status) across the interval of -350 to -150 (defined as the location of the -1 nucleosome) and from -150 to $+150$ (defined as the location of the $+1$ nucleosome), relative to that ATG translational start site.

Definition of 3' NFR

A 3' NFR is defined here as the “closest” linker DNA border to the 3' end of the ORF. For purposes of “closeness” we masked those linkers that were <50 bp in length. If no linker >50 bp long was within 500 bp of the ORF end point, then the closest linker that was <50 bp was used as the 3' NFR. This strategy therefore preferentially looks for NFRs that are >50 bp, but if it does not find one, it will allow one that is <50 bp in length.

Genome-wide survey for gene looping using TFIIB and TBP ChIP-chip

TFIIB (Sua7-TAP) and TBP ChIP material was prepared by standard methods (Zanton and Pugh 2006), and hybridized to Affymetrix *S. cerevisiae* Tiling 1.0R Arrays. Signal analysis, interval analysis, and peak calling were performed using Model-based Analysis of Tiling-arrays software (Johnson et al. 2006b). The bandwidth and max gap parameters had a value of 75 bp, and peaks were called using a significance threshold of $P < 10^{-2}$. To minimize false positives due to confounding contributions from nearby promoters, genes were filtered using the following criteria: (1) only the 5082 genes with experimentally determined transcription termination sites (TTS) were used (David et al. 2006), (2) genes whose TTS was within 300 bp of a downstream ORF start site were removed, and (3) genes containing a TBP peak

within ± 150 bp of the TTS were removed. 2679 TFIIIB peaks were mapped to the promoter and/or TTS for the 4151 genes that met the above criteria. Promoter regions were arbitrarily defined as the 300 bp window immediately upstream of the ORF start site, and the TTS window was 150 bp upstream of, and downstream from the TTS. TFIIIB occupied both the promoter region and the TTS of 120 genes (Supplemental Table S6). The CHITEST function in Excel was used to determine the probability that TFIIIB would fall within the promoter region and the TTS of the same gene by chance. The distance from the nearest ORF end site was calculated for the 224 Sua7 peaks that were found within 150 bp of the TTS. The composite distribution profile in Figure 6B was generated by binning these distance values in 50 bp intervals. Microarray data are deposited in ArrayExpress under accession no. E-MEXP-1584.

Acknowledgments

This work was supported by a grant from NIH (HG004160). We thank S. Zanton for help in generating figures. We thank J. Reese, D. Gilmour, S. Tan, and T. Koerber for many helpful comments.

References

- Albert, I., Mavrich, T.N., Tomsho, L.P., Qi, J., Zanton, S.J., Schuster, S.C., and Pugh, B.F. 2007. Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* **446**: 572–576.
- Alen, C., Kent, N.A., Jones, H.S., O'Sullivan, J., Aranda, A., and Proudfoot, N.J. 2002. A role for chromatin remodeling in transcriptional termination by RNA polymerase II. *Mol. Cell* **10**: 1441–1452.
- 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.
- Bao, Y., White, C.L., and Luger, K. 2006. Nucleosome core particles containing a poly(dA-dT) sequence element exhibit a locally distorted DNA structure. *J. Mol. Biol.* **361**: 617–624.
- 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.
- Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. 2004. Global nucleosome occupancy in yeast. *Genome Biol.* **5**: R62. doi: 10.1186/gb-2004-5-9-r62.
- Bolshoy, A., Ioshikhes, I., and Trifonov, E.N. 1996. Applicability of the multiple alignment algorithm for detection of weak patterns: Periodically distributed DNA pattern as a study case. *Comput. Appl. Biosci.* **12**: 383–389.
- Cohanim, A.B., Kashi, Y., and Trifonov, E.N. 2005. Yeast nucleosome DNA pattern: Deconvolution from genome sequences of *S. cerevisiae*. *J. Biomol. Struct. Dyn.* **22**: 687–694.
- David, L., Huber, W., Granovskaia, M., Toedling, J., Palm, C.J., Bofkin, L., Jones, T., Davis, R.W., and Steinmetz, L.M. 2006. A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci.* **103**: 5320–5325.
- Drew, H.R. and Travers, A.A. 1985. DNA bending and its relation to nucleosome positioning. *J. Mol. Biol.* **186**: 773–790.
- 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.
- Fernandes, L., Rodrigues-Pousada, C., and Struhl, K. 1997. Yap, a novel family of eight bZIP proteins in *Saccharomyces cerevisiae* with distinct biological functions. *Mol. Cell. Biol.* **17**: 6982–6993.
- Fujii, S., Kono, H., Takenaka, S., Go, N., and Sarai, A. 2007. Sequence-dependent DNA deformability studied using molecular dynamics simulations. *Nucleic Acids Res.* **35**: 6063–6074.
- 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.
- Ioshikhes, I.P., Albert, I., Zanton, S.J., and Pugh, B.F. 2006. Nucleosome positions predicted through comparative genomics. *Nat. Genet.* **38**: 1210–1215.
- 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.
- Johnson, S.M., Tan, F.J., McCullough, H.L., Riordan, D.P., and Fire, A.Z. 2006a. Flexibility and constraint in the nucleosome core landscape of *Caenorhabditis elegans* chromatin. *Genome Res.* **16**: 1505–1516.
- Johnson, W.E., Li, W., Meyer, C.A., Gottardo, R., Carroll, J.S., Brown, M., and Liu, X.S. 2006b. Model-based analysis of tiling-arrays for ChIP-chip. *Proc. Natl. Acad. Sci.* **103**: 12457–12462.
- Kiyama, R. and Trifonov, E.N. 2002. What positions nucleosomes?—A model. *FEBS Lett.* **523**: 7–11.
- Kornberg, R. 1981. The location of nucleosomes in chromatin: Specific or statistical. *Nature* **292**: 579–580.
- Kornberg, R.D. and Stryer, L. 1988. Statistical distributions of nucleosomes: Nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* **16**: 6677–6690.
- Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. 2004. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* **36**: 900–905.
- Lee, W., Tillo, D., Bray, N., Morse, R.H., Davis, R.W., Hughes, T.R., and Nislow, C. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.* **39**: 1235–1244.
- Li, B., Nierras, C.R., and Warner, J.R. 1999. Transcriptional elements involved in the repression of ribosomal protein synthesis. *Mol. Cell. Biol.* **19**: 5393–5404.
- Li, G., Levitus, M., Bustamante, C., and Widom, J. 2005. Rapid spontaneous accessibility of nucleosomal DNA. *Nat. Struct. Mol. Biol.* **12**: 46–53.
- Lieb, J.D. and Clarke, N.D. 2005. Control of transcription through intragenic patterns of nucleosome composition. *Cell* **123**: 1187–1190.
- Lowary, P.T. and Widom, J. 1997. Nucleosome packaging and nucleosome positioning of genomic DNA. *Proc. Natl. Acad. Sci.* **94**: 1183–1188.
- Lu, Q., Wallrath, L.L., and Elgin, S.C. 1994. Nucleosome positioning and gene regulation. *J. Cell. Biochem.* **55**: 83–92.
- MacIsaac, K.D., Wang, T., Gordon, D.B., Gifford, D.K., Stormo, G.D., and Fraenkel, E. 2006. An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* **7**: 113. doi: 10.1186/1471-2105-7-113.
- Mavrich, T.N., Jiang, C., Ioshikhes, I.P., Li, X., Venter, B.J., Zanton, S.J., Tomsho, L.P., Qi, J., Glaser, R., Schuster, S.C., et al. 2008. Nucleosome organization in the *Drosophila* genome. *Nature* **453**: 358–362.
- Packer, M.J., Dauncey, M.P., and Hunter, C.A. 2000. Sequence-dependent DNA structure: Dinucleotide conformational maps. *J. Mol. Biol.* **295**: 71–83.
- Parnell, T.J., Huff, J.T., and Cairns, B.R. 2008. RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. *EMBO J.* **27**: 100–110.
- Pazin, M.J., Bhargava, P., Geiduschek, E.P., and Kadonaga, J.T. 1997. Nucleosome mobility and the maintenance of nucleosome positioning. *Science* **276**: 809–812.
- Peckham, H.E., Thurman, R.E., Fu, Y., Stamatoyanopoulos, J.A., Noble, W.S., Struhl, K., and Weng, Z. 2007. Nucleosome positioning signals in genomic DNA. *Genome Res.* **17**: 1170–1177.
- Perocchi, F., Xu, Z., Clauder-Munster, S., and Steinmetz, L.M. 2007. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res.* **35**: e128. doi: 10.1093/nar/gkm683.
- Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J., and Madhani, H.D. 2005. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* **123**: 233–248.
- Rando, O.J. and Ahmad, K. 2007. Rules and regulation in the primary structure of chromatin. *Curr. Opin. Cell Biol.* **19**: 250–256.
- Roth, S.Y., Dean, A., and Simpson, R.T. 1990. Yeast alpha 2 repressor positions nucleosomes in TRP1/ARS1 chromatin. *Mol. Cell. Biol.* **10**: 2247–2260.
- Satchwell, S.C., Drew, H.R., and Travers, A.A. 1986. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191**: 659–675.
- 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.
- Sekinger, E.A., Moqtaderi, Z., and Struhl, K. 2005. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol. Cell* **18**: 735–748.
- Shivaswamy, S., Bhinge, A., Zhao, Y., Jones, S., Hirst, M., and Iyer, V.R. 2008. Dynamic remodeling of individual nucleosomes across a eukaryotic genome in response to transcriptional perturbation. *PLoS*

- Biol.* **6**: e65. doi: 10.1371/journal.pbio.0060065.
- Shrader, T.E. and Crothers, D.M. 1989. Artificial nucleosome positioning sequences. *Proc. Natl. Acad. Sci.* **86**: 7418–7422.
- Simpson, R.T. 1991. Nucleosome positioning: Occurrence, mechanisms, and functional consequences. *Prog. Nucleic Acid Res. Mol. Biol.* **40**: 143–184.
- Simpson, R.T. and Stafford, D.W. 1983. Structural features of a phased nucleosome core particle. *Proc. Natl. Acad. Sci.* **80**: 51–55.
- Singh, B.N. and Hampsey, M. 2007. A transcription-independent role for TFIIB in gene looping. *Mol. Cell* **27**: 806–816.
- Suter, B., Schnappauf, G., and Thoma, F. 2000. Poly(dA·dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic Acids Res.* **28**: 4083–4089.
- Tanaka, S., Zatchej, M., and Thoma, F. 1992. Artificial nucleosome positioning sequences tested in yeast minichromosomes: A strong rotational setting is not sufficient to position nucleosomes in vivo. *EMBO J.* **11**: 1187–1193.
- Wang, S.S., Stanford, D.R., Silvers, C.D., and Hopper, A.K. 1992. STP1, a gene involved in pre-tRNA processing, encodes a nuclear protein containing zinc finger motifs. *Mol. Cell. Biol.* **12**: 2633–2643.
- Whitehouse, I., Rando, O.J., Delrow, J., and Tsukiyama, T. 2007. Chromatin remodelling at promoters suppresses antisense transcription. *Nature* **450**: 1031–1035.
- Widom, J. 1996. Short-range order in two eukaryotic genomes: Relation to chromosome structure. *J. Mol. Biol.* **259**: 579–588.
- Widom, J. 2001. Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.* **34**: 269–324.
- 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.
- Zanton, S.J. and Pugh, B.F. 2006. Full and partial genome-wide assembly and disassembly of the yeast transcription machinery in response to heat shock. *Genes & Dev.* **20**: 2250–2265.

Received March 9, 2008; accepted in revised form April 15, 2008.