



## Positioned and G/C-capped poly(dA:dT) tracts associate with the centers of nucleosome-free regions in yeast promoters

Randy Wu and Hao Li

*Genome Res.* 2010 20: 473-484 originally published online February 4, 2010

Access the most recent version at doi:[10.1101/gr.103226.109](https://doi.org/10.1101/gr.103226.109)

---

**Supplemental  
Material**

<http://genome.cshlp.org/content/suppl/2010/02/05/gr.103226.109.DC1.html>

**References**

This article cites 37 articles, 9 of which can be accessed free at:

<http://genome.cshlp.org/content/20/4/473.full.html#ref-list-1>

**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:  
<http://genome.cshlp.org/subscriptions>

---

# Positioned and G/C-capped poly(dA:dT) tracts associate with the centers of nucleosome-free regions in yeast promoters

Randy Wu and Hao Li<sup>1</sup>

Department of Biochemistry and Biophysics, University of California San Francisco, San Francisco, California 94143-2542, USA

Eukaryotic transcriptional regulation is mediated by the organization of nucleosomes in promoter regions. Most *Saccharomyces cerevisiae* promoters have a highly stereotyped chromatin organization, where nucleosome-free regions (NFR) are flanked by well-ordered nucleosomes. We have found that yeast promoters fall into two classes differing in NFR sharpness, and that this distinction follows a known transcriptional dichotomy in yeast genes. A class of yeast promoters having well-defined NFRs are characterized by positioned patterns of poly(dA:dT) tracts with several novel features. First, poly(dA:dT) tracts are localized in a strand-dependent manner, with poly(dA) tracts lying proximal to transcriptional start sites and poly(dT) tracts lying distal, and collectively define a symmetry axis that is coincident with NFR centers. Second, poly(dA:dT) tracts are preferentially “capped” by G:C residues on the terminus proximal to the symmetry axis. Both signature features co-vary with fine positional variations between NFRs, establishing a closely knit relationship between poly(dA:dT) tracts, their capping patterns, and the central coordinates of NFRs. We found that these features are unique to promoters with well-defined NFRs, and that these promoters display significant difference between *in vitro* and *in vivo* nucleosome occupancy patterns. These observations are consistent with a model in which localized and G:C-capped poly(dA:dT) tracts initiate or facilitate the formation of NFRs at their center, possibly with chromatin remodeling and transcriptional machines involved.

[Supplemental material is available online at <http://www.genome.org>.]

Eukaryotic DNA is packaged as chromatin: highly organized arrays of nucleosomes which profoundly affect the functions of the underlying sequence (Luger et al. 1997; Kornberg and Lorch 1999; Li et al. 2007). The recent emergence of genome-wide nucleosome maps in a variety of organisms have made it possible to study, at a global level, how nucleosome positioning emanates from sequence (Yuan et al. 2005; Albert et al. 2007; Lee et al. 2007; Kaplan et al. 2008; Mavrich et al. 2008a,b; Shivaswamy et al. 2008). Factors contributing to *in vivo* nucleosome positioning fall into two categories: intrinsic positioning effects arising from differential DNA–histone affinities imparted by sequence (Widom 2001) and extrinsic positioning effects not directly dependent on the quality of DNA–histone interactions. Intrinsic nucleosome positioning effects include 10-base pair (bp) periodic dinucleotides that contribute to rotational phasing (Ioshikhes et al. 2006; Segal et al. 2006), and poly(dA:dT) tracts that are thought to destabilize nucleosomes because of their intrinsic structure (Iyer and Struhl 1995; Suter et al. 2000). Extrinsic nucleosome positioning effects include passive exclusion (e.g., nucleosome occlusion by a bound transcriptional regulator) and active remodeling (e.g., nucleosome sliding/removal by ATP-dependent chromatin remodeling enzymes or transcriptional machines). Depending on genomic context, these intrinsic and extrinsic effects must be carefully balanced in order to achieve the nucleosome configurations necessary for biological function.

Nucleosomes at yeast promoters have been extensively studied as paradigms of intrinsic and extrinsic positioning. At most yeast promoters, nucleosomes are highly ordered and are largely defined around a nucleosome-free region (NFR). The NFR appears to be

a conserved mode of promoter nucleosome organization in most eukaryotes. NFRs are also closely related to the well-characterized DNase I hypersensitive sites in eukaryotes, where regulatory activity is manifested as local chromatin changes (Elgin 1981; Weintraub et al. 1981). In *Saccharomyces cerevisiae*, NFRs appear in up to 95% of promoters (Mavrich et al. 2008a) and typically span 140 bp (Lee et al. 2007; Shivaswamy et al. 2008). Furthermore, it has been shown that the conserved histone variant H2A.Z is heavily incorporated into the +1 and –1 nucleosomes flanking each NFR (Raisner et al. 2005). NFRs are organizing centers for the assembly of transcriptional machinery (Venters and Pugh 2009) and are thought to determine the locations of transcriptional start sites (TSSs) (Jiang and Pugh 2009).

Mavrich et al. (2008a) have proposed a “barrier nucleosome” model of promoter chromatin organization, where strongly positioned +1 and –1 nucleosomes guide the positioning of flanking nucleosomes through statistical packing. This model accounts for the well-defined positions of NFRs, as well as the long-range order of nucleosomes, but requires robust mechanisms to correctly position the +1 and –1 “barrier nucleosomes” at each promoter. The question remains as to how the +1 and –1 nucleosomes are positioned, and whether this positioning occurs through intrinsic or extrinsic mechanisms or a combination of both. Mavrich and colleagues made the interesting observation that barrier nucleosome positions correlate with certain sequence features, such as nucleosome positioning sequence (NPS) and AA/TT dinucleotide gradient. However, the influences of NFR sequences and extrinsic nucleosome positioning factors on key promoter nucleosome positions have not been explored in detail.

Extrinsic nucleosome positioning effects can be deconvoluted from intrinsic ones by comparative analysis of *in vivo* and *in vitro* nucleosome data. Such a study was recently performed (Kaplan et al. 2008) that found good overall correlation between *in*

## <sup>1</sup>Corresponding author.

E-mail [haoli@genome.ucsf.edu](mailto:haoli@genome.ucsf.edu); fax (415) 514-2617.

Article published online before print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.103226.109>.

vivo and in vitro (assembled on purified genome DNA) nucleosome positions, suggesting that intrinsic DNA sequence preference is an important determinant of nucleosome positioning. We re-examined this data at finer granularity in promoters and found important differences between in vitro and in vivo nucleosome positions (Fig. 1). In many promoters, NFRs with sharply defined boundaries in vivo become very fuzzy in vitro, and the long-range ordering of nucleosomes flanking the NFRs is lost (Mavrich et al. 2008a). These observations suggest that active remodeling/transcriptional processes may be involved in shaping NFRs in the promoter regions.

While it is not fully understood how NFRs are specified by DNA sequence signals, poly(dA:dT) tracts—homopolymeric stretches of deoxyadenosine—are highly abundant in NFRs and have been strongly implicated as possible NFR positioning sequences (Behe 1995; Yuan et al. 2005; Lee et al. 2007; Segal and Widom 2009). Poly(dA:dT) tracts exhibit structural properties (Nelson et al. 1987; Woods et al. 2004), which could make them resistant to nucleosome incorporation and increase their accessibility to transcription factors (Iyer and Struhl 1995; Suter et al. 2000; Anderson and Widom 2001; Bao et al. 2006; Lee et al. 2007). In addition, computational approaches to nucleosome prediction have found that poly(dA:dT)

tracts as short as length 3 have significant discriminative power in distinguishing nucleosomal versus non-nucleosomal sequence (Peckham et al. 2007; Yuan and Liu 2008). These studies suggest a role of poly(dA:dT) in the establishment of NFRs.

In this study, we explore how poly(dA:dT) tracts contribute intrinsically and extrinsically to the formation of yeast NFRs by systematically analyzing the positions of these tracts relative to NFRs at a high resolution. We begin by comparing in vivo to in vitro nucleosome occupancy in promoters genome-wide (Kaplan et al. 2008). We found that promoters fall into two classes differing in the sharpness of NFRs in vivo. One class (class I) is enriched for TATA-less and constitutively expressed genes and shows stronger differences between in vivo and in vitro nucleosome occupancy patterns, compared to the latter class (class II). We then describe several novel characteristics of poly(dA:dT) tracts and show that these characteristics are correlated with NFR patterns unique to class I promoters. First, we show that poly(dA) and poly(dT) tracts have distinct localized spatial distributions in class I promoters, and that these distributions are symmetric across a central axis in the NFR. Second, we demonstrate that poly(dA:dT) tracts in NFRs exhibit an oriented, terminal specific “capping” by G:C. We demonstrate that the positions of G:C-capped poly(dA:dT) tracts

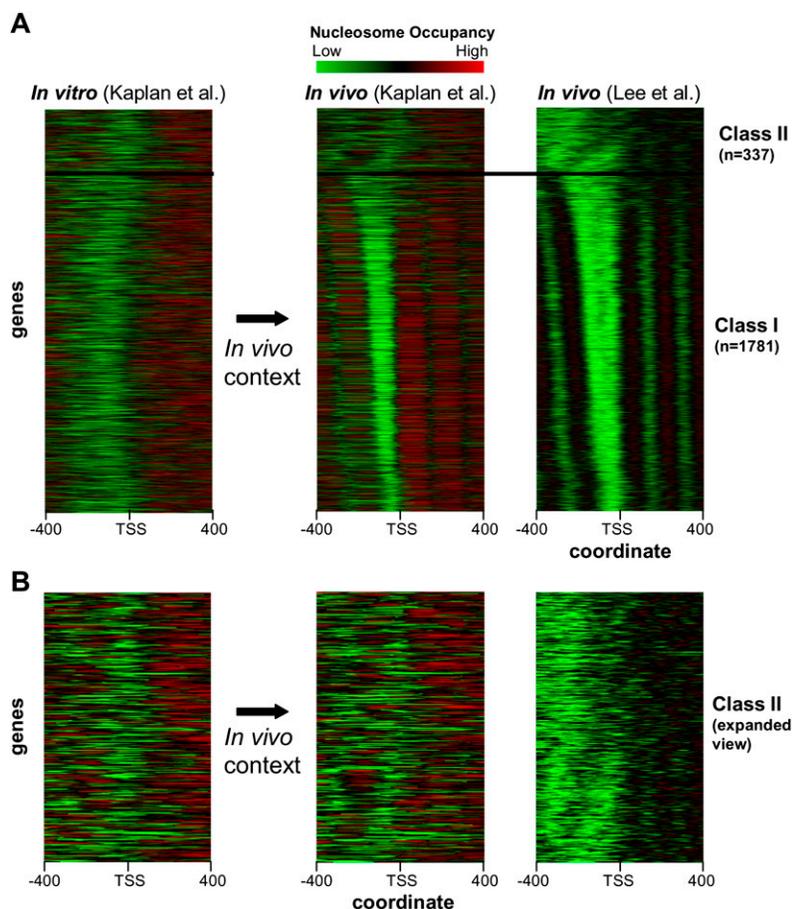
are strongly correlated with NFR center positions. The localized and symmetric placement of poly(dA:dT) tracts in promoters, their orientation-specific base-capping characteristics, and their correlations with NFRs of class I promoters suggest that poly(dA:dT) tracts are important in directing NFRs in these promoters. Models of poly(dA:dT)-directed NFR formation and mechanistic implications are discussed.

## Results

### Extrinsic positioning effects differ between two promoter classes

We re-examined in vivo and in vitro genome-wide *S. cerevisiae* nucleosome positioning data (Lee et al. 2007; Kaplan et al. 2008), using in vivo data from Lee et al. to order 2118 yeast promoters in a visually coherent manner with a self-organizing map (SOM) (Holdaway and White 1990). We also generated promoter nucleosome maps for the Kaplan et al. (2008) (in vivo and in vitro) data using the same promoter ordering (Fig. 1A). Promoters were aligned according to their TSSs, which were previously determined experimentally (David et al. 2006).

After the promoters were aligned and ordered, we were able to demarcate them into two classes on the basis of clear visual distinctions between their in vivo chromatin patterns. Promoters in class I are defined by the presence of a single NFR with sharply defined 5' and 3' boundaries, where the 3' boundary is situated at the TSS. (Although slight variations in



**Figure 1.** Comparison of in vitro and in vivo nucleosome maps at yeast promoters. (A) Nucleosome maps from tiling array data (Lee et al. 2007) for 2118 *S. cerevisiae* promoters were clustered using a self-organizing map (SOM) and visually partitioned into two groups (class I and class II). In vivo and in vitro nucleosome maps (Kaplan et al. 2008) were ordered according to the SOM. Maps were aligned according to transcriptional start sites determined using tiling arrays (David et al. 2006). (B) Close-up of the class II promoter class.

3' NFR boundaries are visible, these may be attributable to experimental noise in nucleosome position mapping/TSS determination.) The positions of 5' NFR boundaries vary substantially, giving rise to NFRs ranging between 100 and 200 bp in width. This effect is easily visible in the SOM, where class I promoters are ordered by decreasing NFR width. In contrast, class II promoters lack well-defined NFR boundaries, giving the group a fuzzy visual appearance. Nucleosome density throughout class II promoters appears uneven and haphazard, and their 3' NFR boundaries do not cleanly line up with their TSSs. The visual distinctions between these two promoter classes were reproducible between the Lee et al. (2007) and Kaplan et al. (2008) data sets, suggesting that these results are robust with respect to different experimental settings.

Next, we compared *in vivo* nucleosome data in class I and class II promoters with the Kaplan et al. (2008) *in vitro* nucleosome data. In class I promoters, the sharply defined NFRs seen *in vivo* are replaced with diffuse nucleosome-depletion throughout the entire promoter *in vitro*. Furthermore, the defined 5' NFR boundaries in class I promoters *in vivo* are lost, as shown in the SOM by the loss of the progressively decreasing NFR sizes (Fig. 1). In contrast, the appearance of nucleosomes in class II promoters *in vitro* does not significantly differ from those *in vivo*. In fact, a close examination of class II promoters shows that many fine details in nucleosome positioning *in vitro* are kept intact *in vivo* (see the expanded view of class II promoters in Fig. 1B).

These observations collectively suggest that the two promoter groups, defined initially on the basis of visual differences in promoter chromatin, reflect two modes of *in vivo* nucleosome organization. To quantify this dichotomy, we computed the Pearson correlations between *in vivo* and *in vitro* nucleosome occupancy data (Kaplan et al. 2008) across each promoter. Smaller correlation indicates a greater divergence between *in vivo* and *in vitro* data and indicates the presence of stronger extrinsic nucleosome positioning effects. The mean *in vivo*-*in vitro* correlation for class II promoters was 0.64, in comparison to 0.58 for class I. Therefore, class I promoters have a stronger dependence on extrinsic nucleosome positioning factors *in vivo* than class II promoters.

### Distinctions between the two promoter classes reflect a known transcriptional dichotomy in yeast

It has been recognized that transcriptional processes at a *S. cerevisiae* promoter can be understood to occur in one of two distinct modes. Such transcriptional dichotomy was first highlighted in TATA-box-containing versus TATA-less genes, where it was shown that TATA-containing genes tend to be more stress-responsive, have greater ranges of expression, and are more regulated by nucleosomes, chromatin regulators, and TBP regulators than TATA-less genes (Basehoar et al. 2004). A recent work has shown that this dichotomy is also related to promoter nucleosome patterns, where promoters with strong nucleosome depletion proximal to the TSS correspond to the TATA-less gene class (Tirosh and Barkai 2008). We found that our dichotomy between class I and class II promoters is fundamentally similar to both of these previously established dichotomies (Supplemental Fig. 1). Class I promoters are more likely to be TATA-less with highly defined patterns of TSS-proximal nucleosome depletion than class II promoters. Therefore, it is likely that a promoter's dependence on intrinsic and extrinsic nucleosome positioning factors is an additional dichotomy, which can be drawn between two fundamentally distinct modes of transcription.

### Poly(dA:dT) tracts in class I promoters occupy defined, symmetric positions in NFRs

To see how poly(dA:dT) tracts contribute to intrinsic vs. extrinsic nucleosome-positioning effects in NFRs, we first computed poly(dA:dT) tract frequencies in class I and class II promoters as functions of distance from the TSS. We considered the two poly(dA:dT) orientations separately, labeling them as either poly(dA) or poly(dT) in accord with the direction defined by downstream transcription. To facilitate comparison between tracts of differing lengths, we normalized each tract frequency as a percent enrichment relative to its corresponding background frequency in all yeast intergenic sequences.

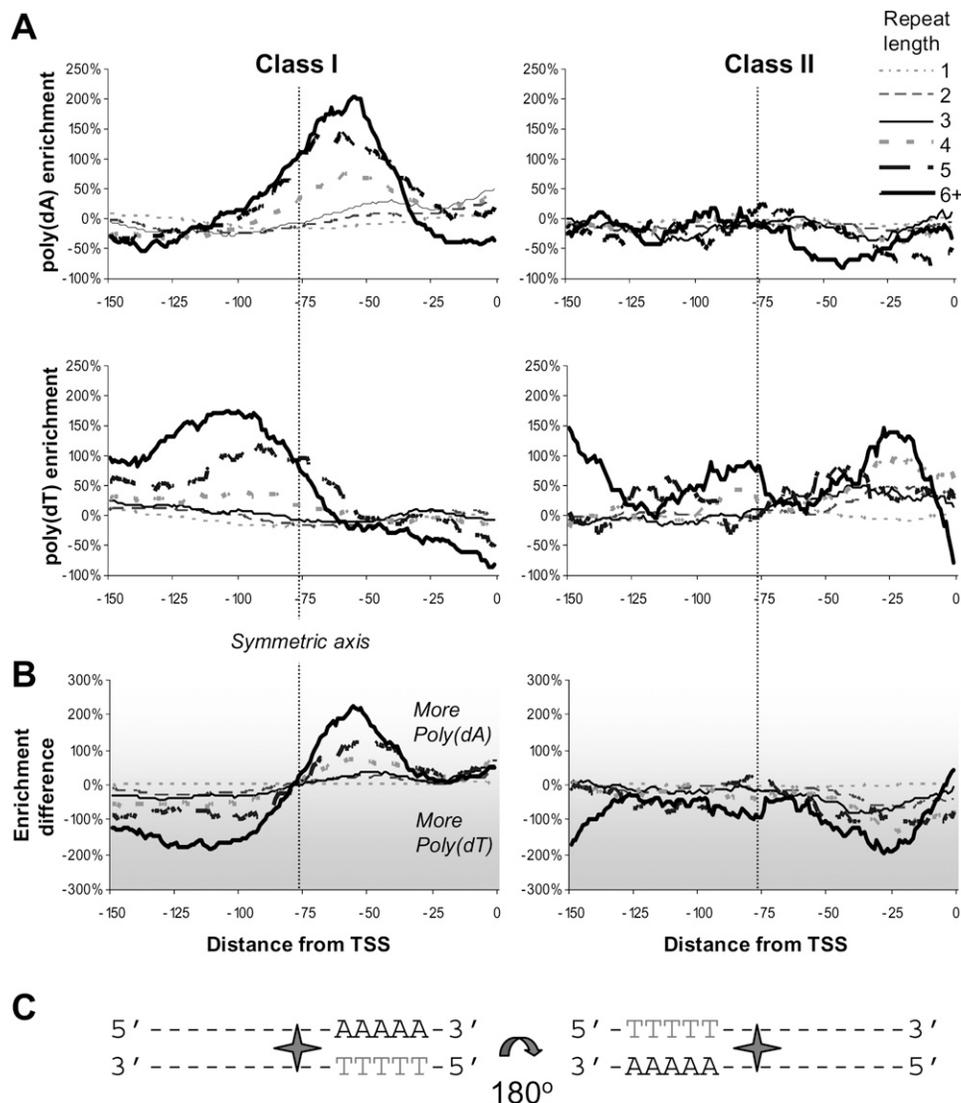
The emerging poly(dA:dT) localization patterns revealed a strong contrast between class I and class II promoters (Fig. 2A). Poly(dA:dT) tracts are not only significantly more abundant in class I promoters, but also exhibit a striking localized symmetry specific to these promoters. Poly(dA:dT) tracts in class I promoters are symmetric about an axis located ~75 bp upstream from the TSS, at the center of the average NFR. Poly(dA:dT) tracts are arranged about this symmetric axis in an orientation-dependent manner: poly(dA) sequences tend to be downstream of this axis (peak near -60), while poly(dT) sequences tend to be upstream of the axis (peak near -90). The magnitudes of these peaks increase with longer tract lengths. This localized, symmetric arrangement of poly(dA:dT) tracts is not present in class II promoters, where poly(dA:dT) tracts are less abundant.

It is useful to summarize this localized, symmetric poly(dA:dT) tract arrangement by taking the difference between the poly(dA) and poly(dT) enrichments curves (Fig. 2B). In these enrichment difference curves, the horizontal intercept indicates the promoter coordinate where poly(dA:dT) tracts in both orientations are equal in abundance. The enrichment difference curves of all tract lengths share an approximate 180° rotational symmetry (i.e.,  $C_2$  symmetry) about the symmetric axis. This result demonstrates that there is a 180° rotational symmetry in the arrangement of poly(dA:dT) tracts in the underlying sequence: After rotation by 180°, a poly(dT) tract on one side of the symmetric axis becomes a poly(dA) tract on the other side (illustrated in Fig. 2C). Such a symmetry implies that, on average, poly(dT) tracts "point to" NFR centers downstream, and such a relationship holds true on either strand, independent of the direction of the transcription.

While NFRs are also known to be present in the 3' untranslated regions of genes, we did not observe localized, symmetric poly(dA:dT) tract enrichment there (Supplemental Fig. 2). This suggests that the presently observed relationship between poly(dA:dT) tracts and NFRs may be specific to those NFRs in 5' promoter regions.

### Positions of poly(dA:dT) tracts co-vary with *in vivo* NFR positions in class I promoters

The observation that the poly(dA:dT) symmetric axis (-75) coincides with the average NFR central coordinate led us to examine how poly(dA:dT) tract positions correspond with *in vivo* NFR positions at a more fine-grained level. We segmented the SOM-ordered class I promoters into six equal subgroups (I-VI), where the NFR of each subgroup narrows over the previous subgroup (Fig. 3A) by ~17 bp, and plotted poly(dA:dT) enrichment difference curves for each subgroup (Fig. 3B; for raw enrichment values, see Supplemental Fig. 3). Finally, the locations of relevant NFR features in each subgroup are summarized in Figure 3C, which schematically represents six elements: the 5' and 3' nucleosome boundaries, the



**Figure 2.** Poly(dA:dT) tract enrichment patterns in yeast promoters. (A) Poly(dA:dT) tract frequencies in class I and class II promoters are represented as relative percent enrichments over background tract frequencies. Tracts in the two orientations are considered separately and referenced as either poly(dA) or poly(dT) according to the downstream direction of transcription. Data are smoothed over 21-bp windows. Tracts of lengths 6 and greater were considered collectively for statistical accuracy. Coordinates are relative to transcription start sites (TSS). (B) Enrichment differences between poly(dA) and poly(dT) tracts as a function of promoter position; values above the  $x$ -axis indicate greater poly(dA) enrichment. The symmetric axis near  $-75$  is indicated with a dashed line. (C) Illustrating the  $180^\circ$  rotational (i.e.,  $C_2$ ) symmetry of poly(dA:dT) tracts with respect to the symmetric axis.

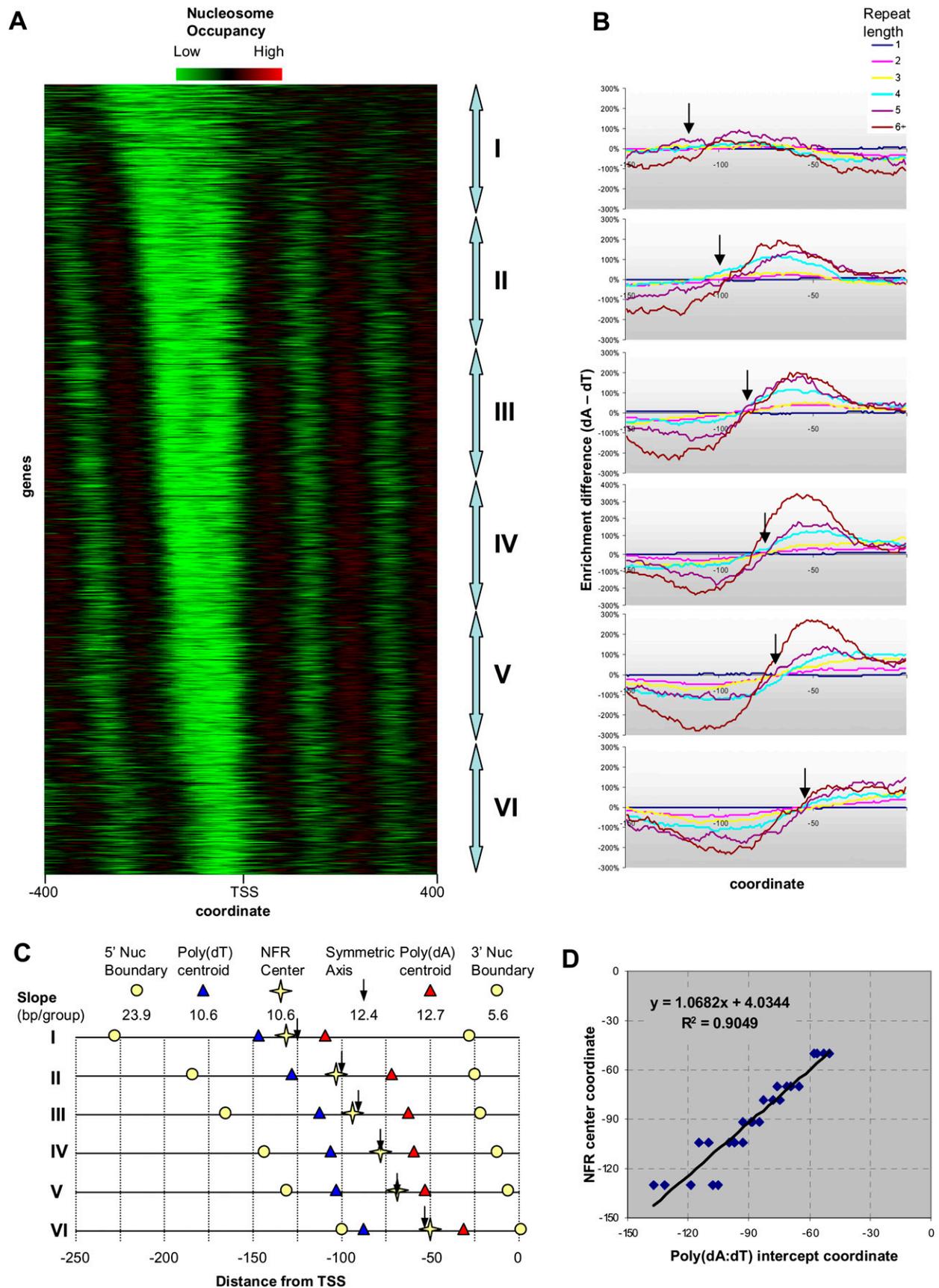
NFR center, the symmetric axis, and the centroids of the poly(dA) and poly(dT) enrichment curves (which represent the expected tract positions). We also assigned a best-fit slope to each of the six promoter elements using linear regression, where the slope of each element represents the average number of base pairs the element shifts toward the TSS over the  $\sim 297$  promoters of the subgroup.

Several trends among NFR elements were notable. First, a plot of NFR center coordinates against poly(dA:dT) intercept coordinates (Fig. 3D) has a slope of approximately one ( $r^2 = 0.90$ ). Second, poly(dA) and poly(dT) tract centroids tended to remain at relatively fixed distances ( $\sim 20$ – $25$  bp) from the symmetric axis. This is shown by the slopes of the symmetric axis (12.4 bp/group), NFR center (14.7 bp/group), and tract centroids (10.6 and 12.7 bp/group), which have comparable values. These results stand in contrast with the slopes of the NFR boundaries: The 3' boundary moves very little (5.6 bp/group), where the 5' boundary moves dramatically (23.9 bp/group). Thus,

poly(dA:dT) tracts do not seem to act as boundary elements because they do not abut NFR boundaries. Instead, their influence on nucleosome positioning seems more closely tied to the NFR central coordinate. Finally, as nucleosome occupancy patterns do not shift across subgroups in vitro (see Supplemental Fig. 4), these influences are likely to be extrinsic.

#### Poly(dA:dT) tracts in promoters are G:C-capped in a terminal-dependent manner

Given the low information content and ubiquitous nature of poly(dA:dT) tracts in *S. cerevisiae* intergenic sequences, we investigated the possibility that certain poly(dA:dT) tracts may contain additional sequence signals in flanking regions, which enhance the tracts' specificities and distinguish them from the background. To investigate this hypothesis, we examined base pair



**Figure 3.** Poly(dA:dT) tracts track fine variations in NFR positions. (A) The class I promoters, which show a progressive narrowing of the NFR, are divided into six equal subgroups, I–VI. (B) Poly(dA:dT) tract enrichment differences for each subgroup. Arrows denote locations of symmetric axes for individual subgroups. (C) Schematic summary of the locations of six promoter elements in each promoter subgroup. Slopes are derived by linear regression and represent the average number of base pairs an element shifts per subgroup of promoters. (D) Plot of NFR center coordinates versus poly(dA:dT) intercept coordinates. Multiple points for each subgroup represent intercepts for different tract length.

compositions at the 5' and 3' flanking regions of all promoter poly(dA:dT) tracts. Intriguingly, we found that poly(dA:dT) tracts in yeast intergenic sequences preferentially incorporate G:C nucleotides at terminal positions, a phenomenon we refer to as "G:C capping." G:C capping is strand-specific in the sense that G residues preferentially cap the poly(dA) strand and, accordingly, C residues preferentially cap the complementary poly(dT) strand. Capping can occur in either a 5' sense or a 3' sense relative to the poly(dA) strand. By convention, we designate sequences of the form  $GA_n$  and  $T_nC$  (where  $n$  is an integer) as "5' G:C-capped" tracts and sequences of the form  $A_nG$  and  $CT_n$  as "3' G:C-capped" tracts (see Fig. 4A). Furthermore, we define the G:C-capping rate as the fraction of tracts in a given population that are G:C-capped.

We computed poly(dA:dT) G:C-capping rates in two kinds of intergenic regions: promoter sequences (TSS up to -150) and nonpromoter sequences (downstream of stop codons and nonoverlapping with promoter sequences). As a background measure, the expected G:C-capping rate is 25.7% based on the overall distribution of single nucleotides in yeast intergenic regions. In both kinds of intergenic regions, we found that G:C-capping rates can significantly exceed the expected capping rate. G:C-capping rates are strongly dependent on poly(dA:dT) tract lengths (Fig. 4B): As

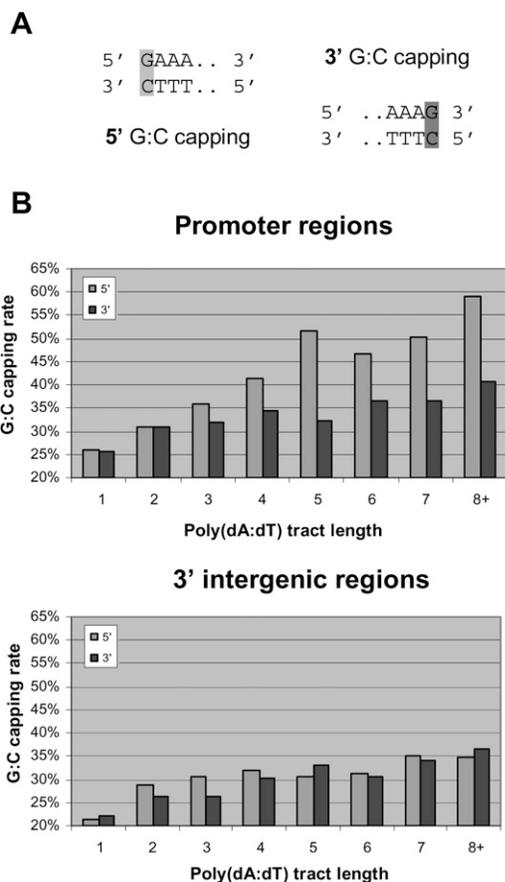
tracts get longer, they are capped more frequently in both promoter and nonpromoter sequences. However, G:C-capping trends differ between promoter sequences and nonpromoter intergenic sequences in two important ways. First, G:C-capping rates are significantly higher in promoters than in nonpromoters at all tract lengths. Second, 5' and 3' G:C-capping rates rapidly diverge in promoter sequences for tracts longer than 2, where 5' capping rates begin to rise dramatically faster than 3' capping rates. In longer poly(dA:dT) tracts (length 5 and above), 5' G:C-capping rates increase to 50% and above, which are far higher than the corresponding 3' G:C-capping rates. The largest discrepancies between 5' and 3' capping rates reach nearly 20% (for tracts length 8 and longer). By contrast, this terminus dependence is not seen in nonpromoter sequences, where the lower 5' and 3' G:C-capping rates do not diverge as tract lengths increase. Thus, 5' G:C-capping, if not necessarily G:C capping in general, appears to be a promoter-specific phenomenon. Based on these observations, we believe that 5' G:C capping could be a means of adding promoter-specific information to particular poly(dA:dT) tracts, allowing those tracts important for NFR determination to be distinguished from other poly(dA:dT) tracts in the background.

### 5', but not 3', G:C capping of poly(dA:dT) tracts is strongly localized to NFRs

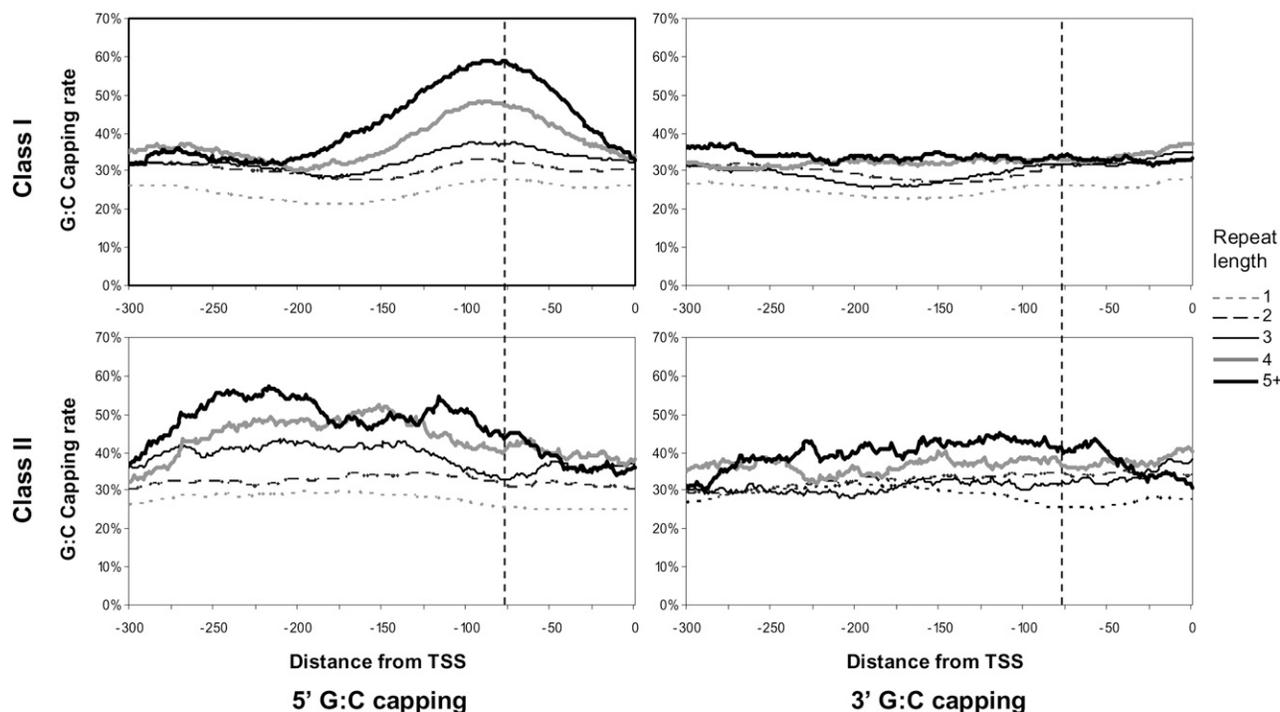
To further test the concept that 5' G:C capping can accentuate poly(dA:dT) tracts in a way that relates to *in vivo* NFR determination, we looked at G:C-capping rates as functions of distance from the TSS, examining 5' and 3' G:C capping in both class I and class II promoters (Fig. 5). As expected, 3' G:C-capping rates lack significant dependence on distance from the TSS; in both promoter classes, 3' capping rates are uniform up to 300 bp upstream from the TSS, with only modest dependence on tract length. In contrast, 5' G:C-capping rates are strongly distance dependent. In class I promoters, 5' capping rates rise conspicuously at a single peak centered at -75, where the longest tracts approach capping rates of nearly 60%. At distal promoter regions (past -200), however, 5' capping rates fall back down to background levels equal to 3' capping rates. Thus, the asymmetry between G:C-capping rates at 5' and 3' tract termini is a highly localized phenomenon that is coincident with the symmetric axis: Tracts that are closer to the symmetric axis are more likely to be 5' G:C-capped. Class II promoters also show somewhat higher levels of 5' G:C capping than background. However, rates are not strongly localized and appear to be higher in distal promoter regions. These observations partially reflect the noisier capping data due to the smaller sample size (also visible in 3' capping), but may also be indicative of some functional role of G:C capping in class II promoters as well.

### 5' G:C-capping localization superimposed on poly(dA:dT) tract localization: An overall view

We have shown that there are two separate trends in poly(dA:dT) tract positioning in class I promoters. First, poly(dA:dT) tracts, depending on orientation, are placed on opposite sides of the symmetric axis. Second, tracts that are closer to the symmetric axis are more likely to be 5' G:C-capped than those that are farther from the symmetric axis. The overall result of these two effects is given in Figure 6A, where the frequencies of tracts in both orientations are each broken down by the 5' capping base. As Figure 6 illustrates, the overall effect of 5' G:C capping is to subdivide each tract population into two subpopulations: a group of 5' G:C-capped



**Figure 4.** G:C capping of poly(dA:dT) tracts. (A) Illustrating the concept of poly(dA:dT) G:C capping. The poly(dA) strand preferentially terminates with G residues at both ends, while the poly(dT) strand preferentially terminates with C. The capping terminal is designated relative to the poly(dA) strand:  $GA_n$  and  $T_nC$  tracts, 5' capping;  $A_nG$  and  $CT_n$ , 3' capping. (B) G:C-capping rates at 5' and 3' termini over different tract lengths in promoter intergenic regions (from TSS to -150) and 3' intergenic regions (nonoverlapping with promoter sequences).



**Figure 5.** 5' and 3' G:C-capping rates, as functions from the TSS in class I and class II promoters. Patterns denote different poly(dA:dT) tract lengths; tracts 5 and longer were pooled for statistical accuracy. Vertical dashed line indicates the position of the symmetric axis for the average class I promoter.

tracts proximal to the symmetric axis (i.e., closer to the center of the NFR) and a separate group of uncapped tracts distal from the symmetric axis (i.e., close to the boundaries of the NFR). Figure 6A also illustrates that most, but not all, of the localized poly(dA:dT) tract enrichments seen in class I promoters (Fig. 2) are due to 5' G:C-capped tracts.

The spatial relationships between poly(dA:dT) tracts, 5' G:C capping, and nucleosome positions are summarized in Figure 6B, which gives a composite view of these elements in the average *in vivo* class I promoter. The average class I promoter with a NFR 150 bp wide can be thought to be split into upstream and downstream halves by the imaginary symmetric axis at  $-75$ . Flanking the symmetric axis are poly(dA:dT) tracts, where poly(dA)-oriented tracts predominately exist on the downstream half of the NFR, and poly(dT)-oriented tracts predominately exist on the upstream half. Within each half, the oriented tracts are then further divided according to the presence or absence of 5' G:C capping. Capped tracts tend to be proximal to the symmetry axis (15 bp away on average), whereas uncapped tracts are typically distal to the symmetric axis (between 25 and 35 bp away).

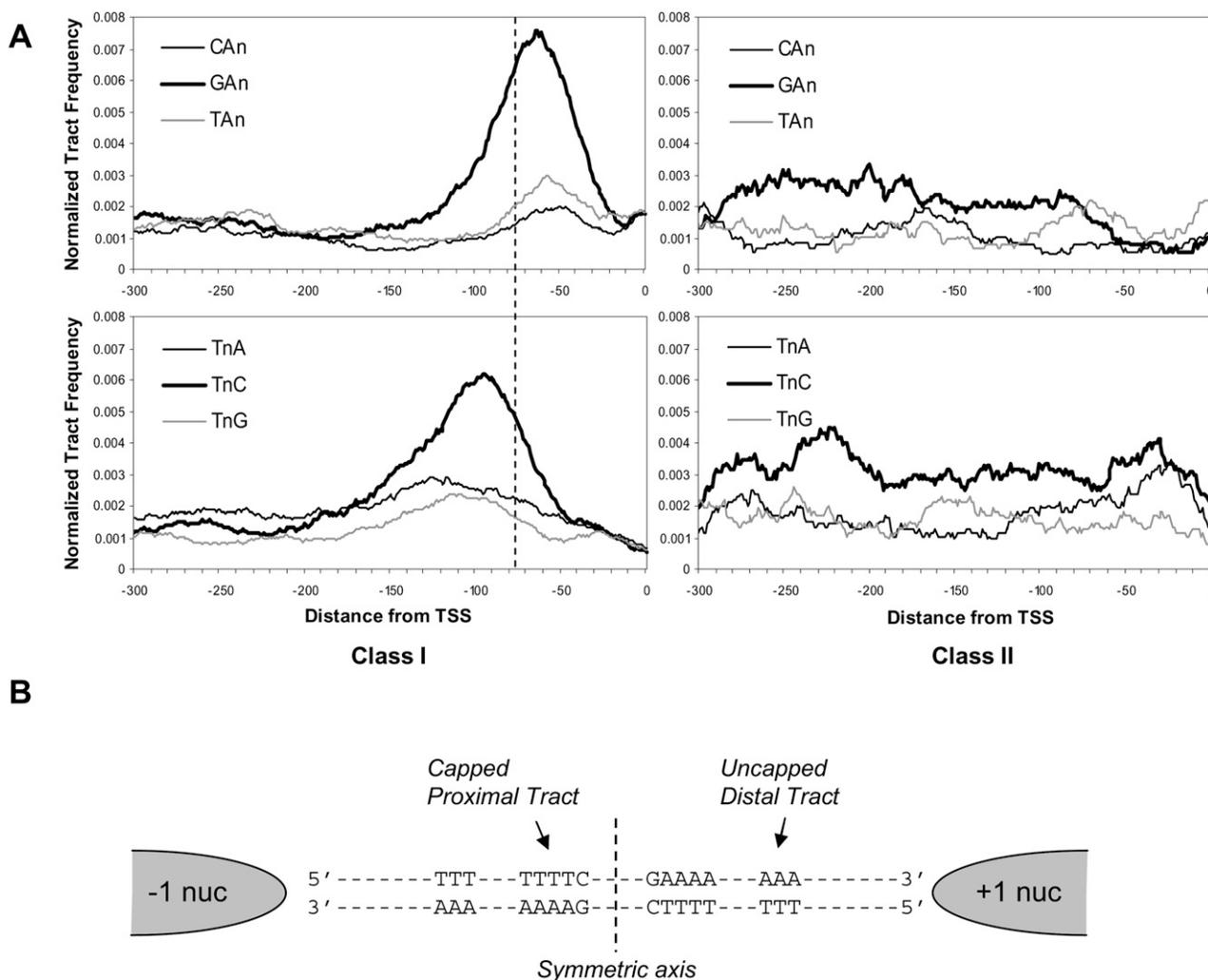
### Poly(dA) and poly(dT) tracts assort independently

Because poly(dA:dT) tract symmetries have been derived by averaging over large number of promoters, it does not follow that promoters are individually symmetric. To address this question, we asked whether the number of promoters containing both poly(dA) and poly(dT) sequences exceed the expected number given independent tract assortment. For this purpose we analyze the 297 well-aligned promoter sequences of subgroup IV of the class I promoters: Because these promoters have highly uniform NFR positions, we can estimate the number of functional poly(dA:dT) tracts in each promoter by counting them in fixed windows

centered 15 bp from the symmetric axis (see data in Table 1). For every tract length, the fraction of promoters in which poly(dA) and poly(dT) both occur in the same promoter is comparable to (and often lower than) the fraction that is expected from independent assortment. Therefore, a given poly(dA:dT) tract on one side of the symmetry axis is not necessarily accompanied by an inverted tract on the opposite side of the symmetry axis. This counting analysis thus disfavors any mechanistic model where poly(dA:dT) tracts are constrained to act as inverted pairs in individual promoters.

### Poly(dA:dT) tract patterns show independence from transcription factor binding sites

Because poly(dA:dT) tracts appear to be closely associated with extrinsic nucleosome positioning effects, we also assessed whether the observed poly(dA:dT) enrichment patterns could be dependent on the presence of binding sites for known transcription factors (TF). Any TF that is functionally associated with the *in vivo* nucleosome positioning role of poly(dA:dT) tracts is likely to be found preferentially in class I promoters over class II promoters. Thus, we used this criteria to screen TFs by using an annotated set of bound and functionally conserved transcription factor binding sites (TFBS) (Fig. 7A; MacIsaac et al. 2006). Binding sites for the majority of TFs are vastly overrepresented in class II promoters; for example, Skn7 favors occupancy in class II promoters by more than 10:1. Apparently, the majority of TFBS are concentrated in a relatively few number of class II promoters, which is consistent with the notion that these genes tend to be more highly regulated. Only four of 45 TFs examined (Reb1, Hsf1, Abf1, and Rpn4) were overrepresented in the class I promoters, and of these only Reb1 and Abf1 have bound and functionally conserved sites in a significant number of promoters genome-wide (226 and 209, respectively).



**Figure 6.** 5' G:C capping splits poly(dA:dT) tracts into subpopulations proximal and distal to the symmetric axis. (A) Frequencies of poly(dA:dT) tracts ( $n = 4$ ) with different 5' capping bases were computed in class I and class II promoters. Frequencies represented on the vertical axis are numbers of motifs per promoter per base smoothed across a 21-bp window. Bold black curves represent 5' G:C capping. (B) Schematic representation of the data from (A) showing a hypothetical class I promoter with 5' G:C-capped and -uncapped poly(dA:dT) tracts arranged across the symmetric axis.

Do the extrinsic nucleosome-positioning effects exerted by poly(dA:dT) tracts depend on the presence of certain TFs? If this were the case, poly(dA:dT) tracts enrichments would be expected to dramatically increase in promoters containing binding sites for such TF (e.g., Reb1 or Abf1). Furthermore, poly(dA:dT) enrichment sig-

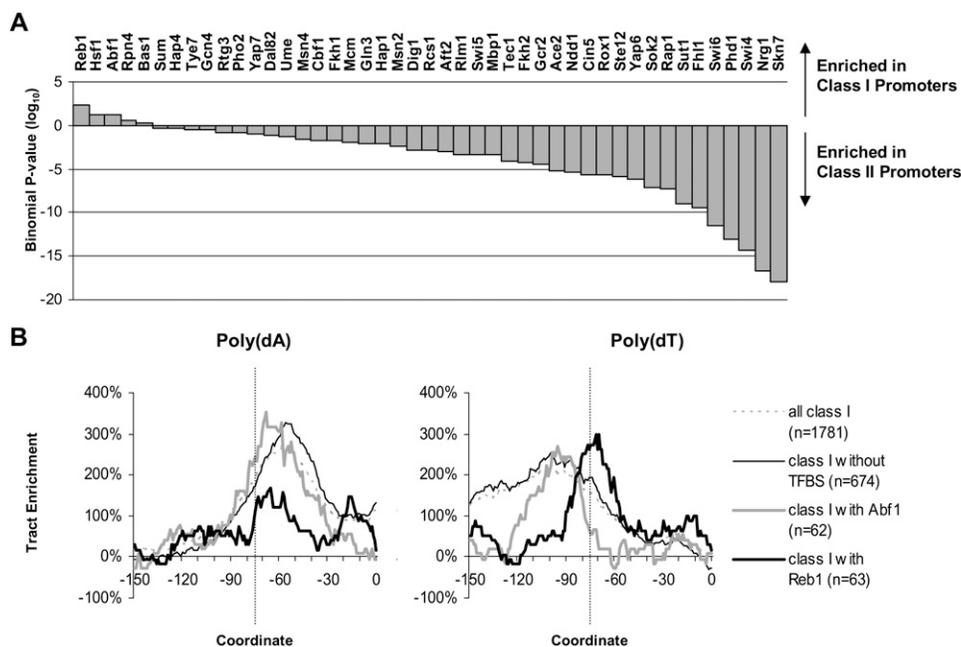
nals would be reduced to background levels in promoters not containing any TFBS. However, this does not seem to be the case: Tract enrichments in promoters containing Reb1 or Abf1 TFBS are similar in magnitude to background (Fig. 7B). Similarly, tract enrichments in TFBS-depleted promoters (in which any promoter containing

a TFBS with a binding  $P$ -value less than 0.005, and not necessarily conserved, is removed) do not decrease. Thus, it appears that the functions of poly(dA:dT) tracts do not directly depend on any of the 118 TFs studied in the MacIssac survey.

Curiously, each of the Abf1- and Reb1-specific tract enrichment profiles have deviations from the overall class I promoter profile. Both Abf1 and Reb1 have sharper poly(dT) enrichments than background. The Reb1 poly(dT) peak is shifted by  $\sim 30$  bp to the opposite side of the symmetric axis, and Reb1 also lacks a prominent poly(dA)

**Table 1.** Poly(dA:dT) coverage (fraction of promoters containing tract) and copy number (average number of tracts per promoter) in class I subgroup IV core promoters ( $n = 297$ )

Tract length cutoff	Poly(dA) coverage	Poly(dT) coverage	dA or dT coverage	dA + dT coverage (expected)	dA avg. copies	dT avg. copies
3	96%	96%	100%	92%(92%)	3.02	3
4	80%	80%	96%	65%(64%)	1.71	1.44
5	59%	57%	83%	32%(33%)	0.89	0.83
6	34%	31%	58%	7%(11%)	0.42	0.37
7	23%	21%	41%	3%(5%)	0.27	0.23
8	15%	13%	27%	1%(2%)	0.16	0.14
9	10%	8%	18%	0%(1%)	0.11	0.084
10	9%	6%	14%	0%(1%)	0.088	0.064



**Figure 7.** Poly(dA:dT) enrichments occur independently of transcription factor binding sites. (A) Ranking of transcription factors by overrepresentation of bound and functionally conserved sites in class I promoters based on annotated sites from MacIsaac et al. (2006).  $P$ -values were computed by modeling the distribution of TFBS between class I and class II promoters using a binomial distribution. The vertical axis gives the logarithm of the cumulative binomial probability of having the observed number of TFBS in a given class. Overrepresentation in class I promoters is shown by positive values and in class II promoters by negative values. (B) Poly(dA:dT) tract enrichments (length  $\geq 4$ ) for class I promoters and subsets thereof: Abf1-containing promoters, Reb1-containing promoters, and TFBS-depleted promoters. TFBS-depleted promoters were selected by excluding promoters containing moderately bound ( $P < 0.005$ ) binding sites (no conservation requirement) for any of 118 TFs.

peak. Thus, even while poly(dA:dT) tracts do not seem to generally depend on any TFs, TF-specific tract patterns may still exist among some factors.

## Discussion

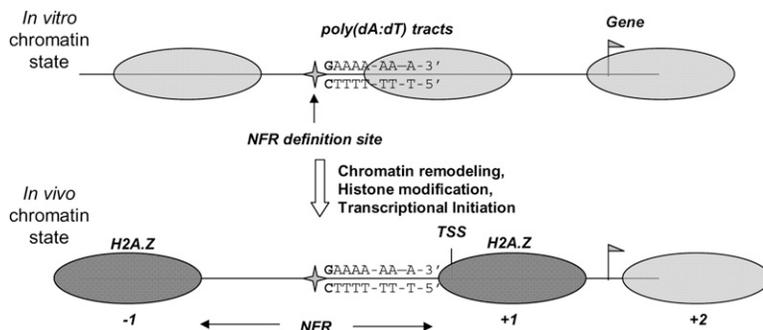
We have used multiple sources of high-resolution nucleosome occupancy data to synthesize a detailed portrait of how poly(dA:dT) tracts relate to NFR positions in yeast promoters. Our analysis is based on our classification of yeast promoters into two distinct classes differing in the sharpness of their NFR boundaries when aligned by transcriptional start sites. By comparing *in vivo* to *in vitro* nucleosome occupancy data, we found that these two classes of promoters also differ in their degree of dependence on extrinsic nucleosome positioning effects. Class I promoters are characterized by sharply defined *in vivo* NFR boundaries that are absent *in vitro*, and thus appear to rely strongly on extrinsic factors. In contrast, class II promoters have similar *in vivo* and *in vitro* nucleosome patterns, both of which lack sharply defined NFR boundaries, suggesting that class II promoters do not strongly depend on extrinsic means of nucleosome positioning.

We have shown that, on a sequence level, these two promoter classes dramatically differ in their distribution of poly(dA:dT) tracts. Poly(dA:dT) tracts are not only more abundant in class I promoters, but also exhibit highly localized, symmetric distributions that are not present in class II promoters and 3' intergenic regions. Importantly, the symmetry axes defined by poly(dA:dT) tracts closely track *in vivo* NFR positions. Interestingly, quantitative analyses of the positions of symmetry axes relative to NFR boundaries and centers suggest that poly(dA:dT) tract positions are most strongly correlated with NFR centers rather than boundaries.

Our analysis also revealed a novel phenomenon, poly(dA:dT) tract G:C capping, especially in class I promoters. G:C capping is terminal-specific, with G capping at the 5' end of poly(dA) tracts and C capping at the 3' end of poly(dT) tracts. Importantly, G:C capping is highly localized at the symmetric axis in these promoters, showing that poly(dA:dT) tracts closer to the symmetric axis are more likely to be capped. Because these capping rates were computed after normalizing for tract abundance, their localization at symmetry axes independently accentuates the organizational significance of the NFR centers. In this respect, the phenomenon of G:C capping adds further nuance to the already intricate fine structure of poly(dA:dT) tracts in class I promoters.

The organization of poly(dA:dT) tracts may prove insightful in deciphering how these sequences can give rise to the extrinsic nucleosome positioning effects seen in class I promoters *in vivo*. Based on our analysis that poly(dA) and poly(dT) tracts assort independently (that is, they need not both occur on the same promoter) and appear in multiple copies in each promoter, we propose a hypothetical NFR definition model consistent with our data. In this model, a group of poly(dA:dT) tracts specify an NFR definition site in the promoter (Fig. 8) that becomes the NFR center after downstream extrinsic nucleosome positioning events. The set of local poly(dA:dT) can be thought of as components of a hypothetical "arrow" in the promoter that "points" to the NFR center, where G:C-capped tract is the "head" of the arrow positioned close to the center. Our data in Figure 3 suggest that this "central" NFR definition model is more likely than a model where poly(dA:dT) tracts directly act as nucleosome boundary elements.

Although it is beyond the scope of this paper to pursue the mechanistic origins of the present phenomena, we briefly mention a few potential classes of regulation that might be involved.



**Figure 8.** A model of how poly(dA:dT) tracts mediate extrinsic nucleosome-positioning effects. Illustrating a hypothetical role of poly(dA:dT) tracts in facilitating a the formation of an *in vivo* NFR during nucleosome assembly or remodeling. 5' G:C-capped poly(dA:dT) tracts, in conjunction with downstream uncapped tracts, may guide the activity of a chromatin-remodeling complex or transcription factor. The 5' capping residue may signal the initiation or termination of remodeling activity. This process may be coupled to the deposition of histone H2A.Z deposition into boundary nucleosomes. The NFR, once established, serves as the basis for subsequent transcriptional initiation events.

Regarding transcription factors, our survey of TF distribution across class I and class II promoters showed that only a small fraction of TFs preferentially occupy the former. Of these, the so-called “general” transcription factors Reb1 and Abf1 have recently been discussed in the literature as frequent occupants of NFRs, in which they are highly localized (Lee et al. 2007). We have shown, however, that the enrichment of positioned poly(dA:dT) tracts are not generally dependent on any particular TF, or even all TFs collectively. Thus, it may be that poly(dA:dT) tracts and Reb1/Abf1 perform overlapping functions in extrinsic nucleosome organization. However, notable deviations in poly(dA:dT) tract patterns in Reb1- and Abf1-containing promoters suggest that these TFs may be functionally dependent on poly(dA:dT) tracts in some way.

Another way for poly(dA:dT) tracts to direct extrinsic nucleosome positioning is by facilitating interactions between promoter sequences and ATP-dependent chromatin remodeling enzymes, which may remove nucleosomes and/or slide them away from the NFR center. Yeast contains numerous candidate chromatin remodeling complexes including Swi2/Snf2, Ino80, Isw1, Isw2, and RSC (Narlikar et al. 2002). For example, RSC is known to mediate nucleosome sliding at Pol II promoters (Cairns et al. 1996; Parnell et al. 2008). Interestingly, its ATPase subunit, Sth1, can track along one strand of duplex DNA with 3' to 5' polarity (Saha et al. 2005). It is possible that G:C-capped poly(dA:dT) tracts may mark an initiation point for the remodeling activity of ATP-dependent chromatin remodelers such as RSC.

It is known that most of the  $-1$  and  $+1$  nucleosomes flanking yeast promoter NFRs contain the conserved histone variant H2A.Z in place of H2A. Although it is currently unclear whether H2A.Z deposition precedes or follows NFR formation, the presence of H2A.Z at the specifically positioned nucleosomes that define the NFR is consistent with the idea that correct formation of NFRs is inextricably tied to regulated, extrinsic processes. In particular, it has been shown that the introduction of an ectopic poly(dA:dT) 7-mer sequence in conjunction with a modified Reb1 binding sequence into a non-NFR locus is sufficient for local NFR induction and H2A.Z deposition. These data, which corroborate our notion poly(dA:dT) tracts may extrinsically position nucleosomes into NFR structures, also suggest that H2A.Z incorporation might be tied to the mechanism of such remodeling.

We emphasize that an extrinsic role for poly(dA:dT) tracts in nucleosome positioning is not inconsistent with the established

intrinsic nucleosome positioning properties of these tracts. As illustrated in Figure 1, intrinsic nucleosome positioning effects generally deplete promoters of nucleosomes *in vitro*, while extrinsic effects refine these structures into stereotyped NFRs at select promoters *in vivo*. Poly(dA:dT) tracts may exert a similar balance of intrinsic and extrinsic effects by generally repelling nucleosomes, while mediating interaction between promoter DNA and transcription factors or chromatin remodeling machinery in select contexts. Furthermore, the hypothesis that poly(dA:dT) tracts play an extrinsic role in the establishment or maintenance of NFRs is consistent with the observation by Mavrich et al. (2008a) that NPS and dinucleotide gradient underlie boundary nucleosomes.

We observed that when the  $+1$  and  $-1$  nucleosomes are grouped separately, a dinucleotide gradient is apparent only on the side of the nucleosome facing NFR, i.e., 3' for the  $-1$  nucleosome and 5' for the  $+1$  nucleosome (see Supplemental Fig. 5a). In addition, for nucleosomes flanking wider NFRs, the gradient shifts away from the nucleosome and toward the center of NFR (Supplemental Fig. 5b). These observations suggest that the gradient originates from the NFR regions and is more likely to repel nucleosomes rather than attract them. It is plausible that a combination of repulsion by the NFR (with some extrinsic effects) and attraction by the nucleosome positioning sequence act together to position the boundary nucleosomes.

Finally, our division of promoters into two classes is largely consistent with previously reported transcriptional dichotomies, such as that between TATA-containing and TATA-less genes. TATA box-containing genes (corresponding to class II promoters), tend to be stress-induced, expressed at very high or low levels, under more evolutionary selective pressure, and more regulated by nucleosomes and chromatin regulators (Basehoar et al. 2004) compared to TATA-less genes (corresponding to class I promoters). The presence of poly(dA:dT) mediated extrinsic NFR formation in some genes, but not others, likely reflects the differential regulatory requirements of these genes. TATA-less promoters may be more likely to require an extrinsically mediated open NFR to maintain high levels of TFIID and basal transcriptional levels commensurate with housekeeping functions. As shown in Figure 1, these extrinsic effects are also likely to be important in defining the transcriptional start site in these genes. These observations also raise the possibility that NFR formation in class I promoters may be tied to the transcriptional activity at these promoters, with poly(dA:dT) tracts marking a site for the assembly of transcriptional machineries. In contrast, class II promoters may require more diverse and individualized nucleosome structures because of their greater regulatory requirements, and do not utilize poly(dA:dT) mediated extrinsic NFR formation. Thus, our findings reinforce the notion that the yeast transcriptome is fundamentally bipolar.

Despite unresolved mechanistic details, we have provided insight into a class of promoter-specific sequences that correspond to the positions of key nucleosomes. By postulating that poly(dA:dT) motifs form the basis of dynamic, extrinsic mechanisms of NFR formation, we have provided a fresh perspective on this ubiquitous class of sequences. Finally, NFRs are known to be present in the promoters of multicellular eukaryotes. It will be interesting to see whether

similar poly(dA:dT) sequence patterns are seen in NFR-containing promoters of higher organisms. Even if the sequence patterns themselves have diverged, common mechanistic principles controlling extrinsic NFR formation may be shared among other eukaryotes.

## Methods

### Promoter classification

*S. cerevisiae* nucleosome positioning data mapped at 4-bp resolution (Lee et al. 2007) was aligned relative to 4799 mapped transcriptional start sites (David et al. 2006) from -400 to +400. For all single direction promoters we considered the distance between the TSS and the corresponding start codon, filtering out promoters where the TSS is downstream or >500 bp upstream of ATG. Furthermore, we excluded short promoters of less than 150 bp, yielding a total of 1842 filtered single direction promoters. This analysis was repeated in divergently transcribed intergenic regions where the TSS to TSS distance was greater than 1 kb. In these cases the opposing termini were extracted separately to give 276 additional promoters. All promoters were oriented relative to the direction of transcription. Self-organizing Map analysis was performed using Cluster with default parameters (Xdim = 49 and Ydim = 1) and visualized with Treeview; both are available from the Eisen Lab (<http://rana.lbl.gov/EisenSoftware.htm>). Additional *in vivo* and *in vitro* nucleosome maps (Kaplan et al. 2008) were presented by ordering the data in the same promoter order as the SOM.

### Poly(dA:dT) enrichment analysis

We computed background frequencies of all poly(dA:dT) sequences within all yeast intergenic sequences (3.5 Mb) as motif count per total sequence length. Motif counts were tabulated only according to the longest observed tracts; e.g., GAAAAT registers as AAAA, but not AAA or AA. For the background set, poly(dA) and poly(dT) were pooled due to lack of reference direction. Position-dependent frequencies in promoter sequences were computed in 21-bp sliding windows centered at the reference position. Any poly(dA:dT) tract whose 5' end was located in this window was tabulated to occur in this window. For each window, frequencies were computed as total motif number over all sequences per window size. For both background and location-specific analyses tracts of  $l \geq 6$  were counted as a single class. Enrichment scores were computed as

$$Enrichment = \frac{freq_{obs} - freq_{background}}{freq_{background}}$$

Enrichment difference curves were derived by subtracting poly(dT) enrichments from poly(dA) enrichments.

### Promoter element location analysis

#### Nucleosomal boundaries

For each subgroup, we computed the average nucleosome occupancy profile using aligned data (Lee et al. 2007). The maxima of peaks corresponding to +1 and -1 nucleosomes were taken to be their average central coordinates within each subgroup. Boundary coordinates were inferred by adjusting central coordinates by 73 bp. The NFR center was taken as the average of the 5' and 3' NFR boundary coordinates.

#### Symmetric axes

For each subgroup, we found the horizontal intercepts of the enrichment difference curves of each tract length class. The coor-

dinate of the symmetric axis for each subgroup was taken as the median value among the intercepts of the different length classes.

#### Poly(dA:dT) centroids

For each subgroup, we first computed an extended motif enrichment profile from -300 to +50, which was then smoothed using a Gaussian ( $\sigma = 5$  bp). We then selected a promoter region that was inclusive of the major enrichment peaks for length classes 4, 5, and 6+. This region is defined according to the largest region spanned by the  $x$ -intercepts of the smoothed enrichment curves of these three length classes. The centroids for each length class were calculated within this region as follows:

$$Centroid_k = \frac{\sum_i i E_{ik}}{\sum_i E_{ik}}$$

where  $i$  is the promoter position relative to the TSS,  $k = 4, 5,$  or  $6+$  is an index over motif length classes, and  $E$  is the enrichment. To filter out background signals, only locations with enrichment values above 0% were tabulated. The final centroid position for the subgroup was taken as the median value among the three length classes.

#### Tract counting analysis

Nonoverlapping tracts were counted in fixed windows of the 297 aligned promoters of class I subgroup IV. Windows were 80-bp wide and centered at -60 for poly(dA) and at -90 for poly(dT). The expected number of promoters with both poly(dA) and poly(dT) coverage given independent assortment of poly(dA) and poly(dT) is given by:

$$E = \frac{n_A n_T}{n_{total}}$$

where  $n_A$  is the number of promoters containing poly(dA) tracts,  $n_T$  is the number of promoters containing poly(dT) tracts, and  $n_{total}$  is the total number of promoters.

#### G:C-capping analysis

Background capping rates were found by pooling all instances of poly(dA) and poly(dT) in all yeast intergenic sequences and computing 5' and 3' terminal base compositions separately for each length class. To compute the expected background capping rate  $r$ , we renormalize background single base frequencies  $f_\sigma$  given that the capping base  $X$  is different than the tract base  $Y$ :

$$r(X, Y) = \frac{f_X}{\sum_{\sigma \neq Y} f_\sigma}$$

thus,

$$r(G, A) = \frac{f_G}{f_C + f_G + f_T} = \frac{0.1738}{0.1779 + 0.1738 + 0.3245} = 0.257 \approx r(C, T).$$

In promoter sequences, we define capping rates at upstream and downstream termini  $c_L^{up}$  and  $c_L^{down}$  by base  $X$  for  $Y$  tracts of length  $L$  at coordinate  $i$  as

$$c_L^{up}(X, Y, i) = p(\sigma_{i-1} = X | \sigma_{i-1} \neq Y, \sigma_{i+L} \neq Y, \sigma_k = Y, i \leq k < i+L) \text{ and}$$

$$c_L^{down}(X, Y, i) = p(\sigma_{i+L} = X | \sigma_{i-1} \neq Y, \sigma_{i+L} \neq Y, \sigma_k = Y, i \leq k < i+L),$$

that is, the probability that the upstream or downstream capping base is  $X$  given contiguous bases  $Y$  at positions  $i$  through  $i+L-1$

flanked by non-Y bases at  $i - 1$  and  $i + L$ . Capping rates shown were smoothed over 41-bp windows centered at  $i$ .

### Transcription factor binding site analysis

We used TFBS annotation from MacIssac et al. (2006) with chromatin immunoprecipitation with microarray hybridization (ChIP-chip)  $P < 0.001$  and conservation in two additional yeast species ([http://fraenkel.mit.edu/improved\\_map/](http://fraenkel.mit.edu/improved_map/)). For each TF we counted the number of promoters containing bound and conserved binding sites ("orfs\_by\_factor\_p0.001\_cons2.txt") in class I and class II promoter classes. Only TF with sites in more than 30 promoters genome-wide (45 TFs total meet this criterion) were considered. The statistical significance of TF overrepresentation was calculated using a binomial distribution with  $N$  total TFBS in class I and II promoters and probably  $P$  of occurring in a class II promoter. For TFs overrepresented in class II promoters, the  $P$ -value is  $\log_{10}[P(X \leq n)]$ , where  $P(X \leq n)$  is the cumulative binomial probability of observing at least  $n$  of  $N$  total sites in class II promoters. For TFs overrepresented in class I promoters, the  $P$ -value is  $-\log_{10}[P(X \geq n)]$ . For the TFBS exclusion analysis we excluded from the class I promoters all promoters with a binding  $P$ -value of 0.005 or less, without any constraint on conservation ("orfs\_by\_factor\_p0.005\_cons0.txt").

### Acknowledgments

We thank William Lee of Stanford for access to the yeast whole-genome nucleosome tiling array data. We also thank Paul Hartley and Hiten Madhani of UCSF and Jeffrey Chuang of Boston College for discussion and feedback.

### References

Albert I, Mavrich TN, Tomsho LP, Qi J, Zanton SJ, Schuster SC, Pugh BF. 2007. Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. *Nature* **446**: 572–576.

Anderson JD, 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 CL, 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 AD, Zanton SJ, Pugh BF. 2004. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* **116**: 699–709.

Behe MJ. 1995. An overabundance of long oligopurine tracts occurs in the genome of simple and complex eukaryotes. *Nucleic Acids Res* **23**: 689–695.

Cairns BR, Lorch Y, Li Y, Zhang M, Lacomis L, Erdjument-Bromage H, Tempst P, Du J, Laurent B, Kornberg RD. 1996. RSC, an essential, abundant chromatin-remodeling complex. *Cell* **87**: 1249–1260.

David L, Huber W, Granovskaia M, Toedling J, Palm CJ, Bofkin L, Jones T, Davis RW, Steinmetz LM. 2006. A high-resolution map of transcription in the yeast genome. *Proc Natl Acad Sci* **103**: 5320–5325.

Elgin SC. 1981. DNAase I-hypersensitive sites of chromatin. *Cell* **27**: 413–415.

Holdaway RM, White MW. 1990. Computational neural networks: Enhancing supervised learning algorithms via self-organization. *Int J Biomed Comput* **25**: 151–167.

Ioshikhes IP, Albert I, Zanton SJ, Pugh BF. 2006. Nucleosome positions predicted through comparative genomics. *Nat Genet* **38**: 1210–1215.

Iyer V, Struhl K. 1995. Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J* **14**: 2570–2579.

Jiang C, Pugh BF. 2009. Nucleosome positioning and gene regulation: Advances through genomics. *Nat Rev Genet* **10**: 161–172.

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, Leproust EM, Hughes TR, Lieb JD, Widom J, et al. 2008. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**: 362–366.

Kornberg RD, Lorch Y. 1999. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**: 285–294.

Lee W, Tillo D, Bray N, Morse RH, Davis RW, Hughes TR, Nislow C. 2007. A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* **39**: 1235–1244.

Li B, Carey M, Workman JL. 2007. The role of chromatin during transcription. *Cell* **128**: 707–719.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**: 251–260.

MacIsaac KD, Wang T, Gordon DB, Gifford DK, Stormo GD, 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 TN, Ioshikhes IP, Venters BJ, Jiang C, Tomsho LP, Qi J, Schuster SC, Albert I, Pugh BF. 2008a. A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res* **18**: 1073–1083.

Mavrich TN, Jiang C, Ioshikhes IP, Li X, Venters BJ, Zanton SJ, Tomsho LP, Qi J, Glaser RL, Schuster SC, et al. 2008b. Nucleosome organization in the *Drosophila* genome. *Nature* **453**: 358–362.

Narlikar GJ, Fan HY, Kingston RE. 2002. Cooperation between complexes that regulate chromatin structure and transcription. *Cell* **108**: 475–487.

Nelson HC, Finch JT, Luisi BF, Klug A. 1987. The structure of an oligo(dA)-oligo(dT) tract and its biological implications. *Nature* **330**: 221–226.

Parnell TJ, Huff JT, Cairns BR. 2008. RSC regulates nucleosome positioning at Pol II genes and density at Pol III genes. *EMBO J* **27**: 100–110.

Peckham HE, Thurman RE, Fu Y, Stamatoyannopoulos JA, Noble WS, Struhl K, Weng Z. 2007. Nucleosome positioning signals in genomic DNA. *Genome Res* **17**: 1170–1177.

Raisner RM, Hartley PD, Meneghini MD, Bao MZ, Liu CL, Schreiber SL, Rando OJ, Madhani HD. 2005. Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. *Cell* **123**: 233–248.

Saha A, Wittmeyer J, Cairns BR. 2005. Chromatin remodeling through directional DNA translocation from an internal nucleosomal site. *Nat Struct Mol Biol* **12**: 747–755.

Segal E, Widom J. 2009. Poly(dA:dT) tracts: Major determinants of nucleosome organization. *Curr Opin Struct Biol* **19**: 65–71.

Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang JP, Widom J. 2006. A genomic code for nucleosome positioning. *Nature* **442**: 772–778.

Shivaswamy S, Bhinge A, Zhao Y, Jones S, Hirst M, Iyer VR. 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.

Suter B, Schnappauf G, 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.

Tirosh I, Barkai N. 2008. Two strategies for gene regulation by promoter nucleosomes. *Genome Res* **18**: 1084–1091.

Venters BJ, Pugh BF. 2009. A canonical promoter organization of the transcription machinery and its regulators in the *Saccharomyces* genome. *Genome Res* **19**: 360–371.

Weintraub H, Larsen A, Groudine M. 1981.  $\alpha$ -Globin-gene switching during the development of chicken embryos: Expression and chromosome structure. *Cell* **24**: 333–344.

Widom J. 2001. Role of DNA sequence in nucleosome stability and dynamics. *Q Rev Biophys* **34**: 269–324.

Woods KK, Maehigashi T, Howerton SB, Sines CC, Tannenbaum S, Williams LD. 2004. High-resolution structure of an extended A-tract: [d(CGCAAATTTGCG)]<sub>2</sub>. *J Am Chem Soc* **126**: 15330–15331.

Yuan GC, Liu JS. 2008. Genomic sequence is highly predictive of local nucleosome depletion. *PLoS Comput Biol* **4**: e13. doi: 10.1371/journal.pcbi.0040013.

Yuan GC, Liu YJ, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. 2005. Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**: 626–630.

Received November 14, 2009; accepted in revised form February 2, 2010.