Nucleosome rotational positioning is influenced by *cis*- and *trans*-acting factors

Gregory M. Wright and Feng Cui*

Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, 85 Lomb Memorial Drive, Rochester, NY 14623, USA

*To whom correspondence should be addressed:

Tel: +1 585 475 4115; Fax: +1 585 475 2398; Email: fxcsbi@rit.edu

Running title: Factors determining nucleosome rotational positioning

Keywords: Nucleosome, rotational positioning, chromatin remodelers, DNA patterns

Supplemental data for this article can be accessed on the publisher's website.

Abstract

The rotational positioning of nucleosomes is associated with certain periodic DNA patterns, with the well-known preference of WW dinucleotides (where W is A or T) and SS dinucleotides (where S is G or C) to occur at the sites where nucleosomeal DNA bends into the minor and major grooves, respectively. However, *cis*- and *trans*-acting factors governing this nucleosome positioning sequence (NPS) pattern remain unclear. To identify these factors, we performed a systematic analysis of WW/SS patterns within 147-bp nucleosome core particle fragments in different eukaryotes under multiple cellular conditions. We find that >50% of nucleosomes in the genomes do not have this NPS pattern and >20% of nucleosomes take the inverse, anti-NPS pattern. The fraction of anti-NPS nucleosomes is increased in cells lacking one or more chromatin remodelers and in genes with high RNA polymerase II (Pol II) transcription frequencies. Remarkably, anti-NPS nucleosomes exhibit distinctive distributions across the yeast and human genomes: the nucleosomes are distributed uniformly in yeast, while in humans they are more enriched at promoters and coding regions than at repetitive DNA elements. Such differences can be explained by intrinsic DNA sequence patterns. Thus, our results suggest that the rotational positioning of nucleosomes can be influenced, in a synergic manner, by *cis* and *trans* determinants including DNA sequence, ATP-dependent nucleosome remodeling enzymes and RNA Pol II.

Introduction

One of the most intriguing questions in eukaryotic cell biology is to understand what determines the precise positions of nucleosomes and how specific positioning of nucleosomes affects transcription factor (TF) binding and gene expression in the context of chromatin. The basic structural unit of chromatin is the nucleosome core particle (NCP) that consists of a histone octamer, around which 147 bp of DNA is wrapped in ~1.7 turns (Luger et al. 1997). Genome-wide nucleosome maps for budding yeast reveal the presence of ~200-bp nucleosome-depleted regions (NDR) at the promoters of most genes (Sekinger et al. 2005; Yuan et al. 2005). The NDR is typically flanked by the -1 and +1 nucleosomes that are located at highly preferred positions, and together with downstream nucleosomes, they form a regular array extending from the 5' to 3' end of the coding region of genes (Mavrich et al. 2008a; Shivaswamy et al. 2008). This stereotypical feature of chromatin structure is well conserved in different species (Albert et al. 2007; Mavrich et al. 2008b; Schones et al. 2008). Historically, nucleosome positioning is usually characterized by two parameters: rotational positioning, defined by the side of the DNA helix that faces the histones, and translational positioning, defined by the nucleosome midpoint (or dyad) with regard to the DNA sequence (Lu et al. 1994). Several determinants of nucleosome positioning have been reviewed in literature (Struhl and Segal 2013; Radman-Livaja and Rando 2009).

One of the most well-known *cis* factors for the translational positioning of nucleosomes is homopolymeric sequences poly(dA:dT) that are characterized with stiff structures and are strongly inhibitory for nucleosome formation (Nelson et al. 1987; Suter et al. 2000; Segal and Widom 2009). This poly(dA:dT)-dependent nucleosome exclusion has been observed in many genome-wide nucleosome mapping studies *in vivo* (Field et al. 2008; Raveh-Sadka et al. 2012) and *in vitro* (Kaplan et al. 2011; Zhang et al. 2011). Another important sequence feature is G+C content, which has emerged as an important factor for predicting nucleosome occupancy *in vivo* (Tillo and Hughes 2009; Tillo et al. 2010). Structurally, GC-rich sequence elements are able to adopt energetically favorable configurations at most

locations around histone octamers (Chua et al. 2012), which may account for the strong positioning and high binding affinity of the '601' nucleosome (Lowary and Widom 1998).

Trans-acting factors that govern the translational positioning of nucleosomes include chromatin remodelers and RNA polymerase II (Pol II) (Segal and Struhl 2013). In particular, snf2-related chromatin remodelers such as Isw1 and Chd1 are required for the establishment of regularly spaced nucleosome array around the transcription start site (TSS) (Gkikopoulos et al. 2011). This genomic positioning pattern is weaker in the *S. cerevisiae* single mutants lacking Isw1 or Chd1 and dramatically worse in double mutants (Ocampo et al. 2016). On the other hand, this *in vivo* pattern can be partially reconstituted *in vitro* if a yeast crude extract and ATP are added to purified histones and DNA (Zhang et al. 2011). In other words, chromatin remodeling activities can reconstitute the positioning pattern around TSS (i.e., -1 nucleosome, NDR, +1 and +2 nucleosomes), but not for downstream nucleosomes (i.e., +3 nucleosome and beyond). These observations indicate that chromatin remodelers are necessary but not sufficient to establish the *in vivo* pattern of nucleosome positioning. Recently, it has been shown that that RNA Pol II transcription is involved in establishing the *in vivo* pattern. In particular, basal transcription factors and preinitiation complexes help to position the +1 nucleosome (Zhang et al. 2011; Hughes et al. 2012; Rhee and Pugh 2012), whereas Pol II elongation facilitates the formation of nucleosome arrays (Vaillant et al. 2010; Hughes et al. 2012). Overall, *cis* and *trans* factors act synergistically to determine nucleosome translational positioning, with DNA sequence helping to establish NDR and *trans*-acting factors such as chromatin remodelers and RNA Pol II directing the nucleosome array around TSS.

With regard to the rotational positioning of nucleosomes, it is well established that DNA sequence is a critical determinant (Trifonov 2011). Various nucleosome positioning sequence (NPS) patterns associated with DNA bending propensities have been proposed (Mengeritsky and Trifonov 1983; Zhurkin 1983; Drew and Travers 1985; Calladine and Drew 1986; Uberbacher et al. 1988; Ioshikhes et al. 1992; Baldi et al. 1996; Ioshikhes et al. 1996; Lowary and Widom 1998; Kogan et al. 2006; Ioshikhes et al. 2011), including the WW/SS pattern (where W is A or T and S is G or C), which was first described by Travers

and colleagues (Satchwell et al. 1996). In this pattern, WW dinucleotides tend to occur at the sites where nucleosomal DNA bends into the minor grooves, while SS dinucleotides are often positioned at the sites where nucleosomal DNA bends into in the major grooves. While this pattern has moderate success in predicting the translational positioning of nucleosomes (Segal et al. 2006), it has been successfully used to predict the rotational positioning of nucleosomes (Cui et al. 2013). Structurally, highly conserved "sprocket" arginine residues (Muthurajan et al. 2003; Sullivan and Landsman 2003) insert into the minor grooves of DNA (Davey et al. 2002). Short poly(dA:dT) stretches help to narrow the DNA minor grooves, potentially enhancing electrostatic interactions between DNA phosphate backbone and sprocket arginine residues (Rohs et al. 2009; Wang et al. 2010; West et al. 2010). These arginine-DNA contacts that occur every ~10 base pairs (bp) within nucleosome core particles (NCP) may play an important role in the rotational positioning of nucleosomes (Luger et al. 1997; Rohs et al. 2009; Wang et al. 2010; West et al. 2010). Note that the rotational settings of nucleosomes are critical for the binding of TFs such as the glucocorticoid receptor (GR) (Li and Wrange 1995) and the tumor suppressor p53 (Sahu et al. 2010; Cui and Zhurkin 2014). However, what factor(s) influence the rotational positioning of nucleosomes remains elusive.

Here, we propose that the rotational positioning of nucleosomes have both *cis* and *trans* determinants. We hypothesize that *trans*-acting factors such as chromatin remodelers and RNA Pol II can influence the rotational settings of nucleosomes, while *cis*-acting factors such as DNA sequence play a decisive role. We provide evidence in support of this hypothesis by analyzing genome-wide nucleosome locations mapped by the paired-end sequencing technique. We focused on the 147-bp NCP fragments because the minor- and major-groove bending sites on the fragments can be precisely defined (see Methods). Systematic evaluation of the occurrence of WW and SS dinucleotides in these sites helps to differentiate the canonical WW/SS-based NPS patterns (Satchwell et al. 1986) from other patterns. Nucleosomes with non-NPS patterns are predominant across eukaryotic genomes. Moreover, anti-NPS nucleosomes are often found in cells that lack of one or more chromatin remodelers and have increased Pol II transcription

frequencies. These nucleosomes appear to be distributed differently in yeast and human genomes, which may be explained by intrinsic DNA sequence patterns in the two genomes. Our results suggest that both *cis* and *trans* determinants cooperate to control the rotational positioning of nucleosomes.

Results

Non-NPS nucleosomes are predominantly in eukaryotes

We set out to define 12 minor-groove bending sites (minor-GBS) and 12 major-groove bending sites (major-GBS) along the 147-bp DNA fragment co-crystalized with a histone octamer (PDB ID: 1KX5) (Supplemental Table S1, Supplemental Fig. S1). Depending on the abundance of WW and SS dinucleotides occurring in minor- and major-GBS, we were able to divide nucleosomal DNA into 4 pattern types, Types 1 – 4 (Table 1). The Type 1 pattern represents the conventional NPS pattern with WW preferentially occurring in minor-GBS and SS in major-GBS (Fig. 1A and Supplemental Fig. S2A). The Type 4 pattern is the inverse of Type 1 with WW and SS dinucleotides occurring in major- and minor-GBS respectively (Fig. 1D and Supplemental Fig. S2D). For simplicity, nucleosomes with the Type 1 pattern are denoted as NPS nucleosomes, whereas those with the Type 4 pattern are denoted as anti-NPS nucleosomes. Nucleosomes with a non-Type-1 pattern are collectively called non-NPS nucleosomes.

We quantified the 4 types of nucleosomes in several eukaryotes including yeast, nematode, fruit fly, mouse and human (Supplemental Tables S2 and S3), and had several interesting findings. First, the fractions of the 4 pattern type are conserved between two biological replicates for each species, which highlights the small variability and high quality of selected datasets (Fig. 2A). Second, for all the species, NPS nucleosomes account for <50% of total nucleosomes, indicating that non-NPS nucleosomes are predominant across eukaryotic genomes (Fig. 2A). Third, the fractions of Type 2 and Type 3 nucleosomes are relatively unchanged across species, which prompts us to develop a measure, ΔNPS, to capture the difference in the fractions of Type 1 and Type 4 nucleosomes (that is, ΔNPS (%) = Type 1 – Type 4). The ΔNPS values can be viewed as a representation of relative abundance of NPS *versus* anti-NPS nucleosomes in a given genomic region. Fourth, ΔNPS values vary substantially between ~5% (fruit fly and mouse) and ~20% (yeast, nematode and human) (Fig. 2B). The low ΔNPS values (~5%) observed in

fruit fly and mouse could be related to the embryonic nature of S2 cells (fruit fly) and embryonic stem cells (mouse), because these cells are characterized by "hyperdynamic" chromatin in which histone proteins exchange rapidly on and off the genome (Meshorer et al. 2006). Therefore we are focused on two highly divergent eukaryotic species, yeast *Saccharomyces cerevisiae* and human.

In summary, our results have shown that NPS nucleosomes are not the most frequent type of nucleosomes in eukaryotes. By contrast, non-NPS nucleosomes are prevalent and predominant in eukaryotic genomes. Remarkably, both yeast and human have a substantial amount of anti-NPS nucleosomes (~20%). These nucleosomes possess DNA sequence patterns inverse to the canonical NSP patterns (that is, WW in major-GBS and SS in minor-GBS). Theoretically, the anti-NPS nucleosomes have unfavorable electrostatic interactions between sprocket arginine residues and DNA (see Introduction). We next sought to address what influences the abundance of anti-NPS nucleosomes and how they are distributed in the yeast and human genomes.

Chromatin remodelers influence the abundance of anti-NPS nucleosomes in yeast

We first examined if yeast mutants of ATP-dependent chromatin remodelers favor anti-NPS nucleosomes genome-wide. We calculated the 4 sequence types (Supplemental Fig. S3) and the average ΔNPS values (Fig. 3) of all nucleosomes from yeast strains containing eight chromatin remodeler mutants, which include four single mutants (*isw1Δ*, *isw2Δ*, *chd1Δ*, and *rsc8Δ*), three double mutants (*isw1Δ isw2Δ*, *isw1Δ chd1Δ*, and *isw2Δ chd1Δ*), and one triple mutant (*isw1Δ isw2Δ chd1Δ*) (Ocampo et al. 2016). All mutants except *isw2Δ* have significantly different ΔNPS values from the wildtype strain (Student's t-test $p < 10^{-5}$) (Fig. 3), indicating that deletion of chromatin remodeler(s) often changes the fraction of anti-NPS nucleosomes.

Detailed analysis of ΔNPS values reveals several intriguing observations. First, all mutants have a small deviation of ΔNPS values (<5%) from the wildtype, suggesting that ATP-dependent chromatin remodeling complexes are unable to dramatically change the rotational settings of nucleosomes in yeast.

Second, the mutants can be divided into four groups based on their ΔNPS values (denoted by ovals in Fig. 3). For instance, the single mutant *isw2Δ* is grouped with the wildtype, indicating that *isw2Δ* has a minimal impact on the rotational positioning of nucleosomes. The single mutant *isw1Δ* has a higher ΔNPS value than the wildtype strain, whereas the mutant *chd1Δ* has a lower ΔNPS value, as if the two chromatin remodelers have an opposite effect on the nucleosome rotational settings. The double mutant *isw1Δ chd1Δ* have the lowest ΔNPS value, suggesting that it has the biggest impact on the rotational settings of nucleosomes. Note that the aforementioned patterns have an interesting association with recent findings on nucleosome spacing set by these mutants (Ocampo et al. 2016). That is, ISW2 plays a minor role in setting nucleosome spacing, whereas CHD1 and ISW1 compete to set the spacing. Moreover, the *isw1Δ chd1Δ* double mutant shows heavily compromised spacing of nucleosomes. Our results clearly show that chromatin remodelers have an impact on both translational and rotational positioning of nucleosomes.

Uniform distributions of anti-NPS nucleosomes in yeast

To study how anti-NPS nucleosomes are distributed in the yeast genome, we examined the ΔNPS values around TSS. We first calculated and profiled the overall normalized nucleosome occupancy around TSS, together with the profiles separated into quartiles by transcriptional frequencies based on RNA-seq data (Fig. 4A-B and Supplemental Fig. S4). The nucleosome profiles from the wildtype strain reveal a well-established pattern (Yuan et al. 2005; Mavrich et al. 2008a): the presence of a ~200-bp nucleosome-depleted region (NDR), flanked by phased nucleosomes, which form a highly regular array extending into gene bodies (Fig. 4A). By contrast, the profiles from the *isw1Δ chd1Δ* double mutant have shown the destruction of phased nucleosomes (Fig. 4B), consistent with earlier studies (Ocampo et al. 2016)

Next we profiled the average ΔNPS values for nucleosomes -1 to +5 (Fig. 4C-D) and had the following findings. First, the ΔNPS values vary slightly from the genome average for both wildtype (Fig. 4C) and mutant (Fig. 4D) strains. Although the phased nucleosomes (especially for nucleosomes +2 to +5) are

disrupted in *isw1Δ chd1Δ* double mutant (Fig. 4B), the levels of ΔNPS values have no substantial changes from their genome averages. This trend holds for other mutants (Supplemental Fig. S5), which further supports our finding that chromatin remodelers are not major players in determining nucleosome rotational settings. Second, no noticeable changes in the ΔNPS values are seen between highly and lowly transcribed genes, indicating that RNA Pol II transcription has little impact on nucleosome rotational settings (Fig. 4C-D and Supplemental Fig. S5). Third, nucleosomes -1 and +1 often have a higher ΔNPS value than the genome average, whereas nucleosomes +2 to +5 have a ΔNPS value close to or below the genome average (see solid and dashed black lines in Fig. 4C-D and Supplemental Fig. S5). This tendency is more pronounced when the normalized ΔNPS values are summed for nucleosomes ±1 and nucleosomes 2-5 (Fig. 5). It shows that nucleosomes in promoters (i.e., nucleosomes ±1) tend to have higher ΔNPS values than nucleosomes in gene bodies (i.e., nucleosomes 2-5). If this pattern is caused by chromatin remodelers, one would expect to see a uniform, asymmetric binding pattern of chromatin remodelers around TSS. However, the ChIP-seq data of these remodelers reveal that three remodelers (Chd1, Isw2 and Rsc8) have a higher occupancy in promoters than in gene bodies, whereas one remodeler, Isw1, has a higher occupancy in gene bodies than in promoters (Supplemental Fig. S6), suggesting that the observed feature of ΔNPS values is unlikely due to chromatin remodelers themselves. Rather, our data shown below suggest that it might be associated with the high occurrence of AT-containing fragments in the promoter regions.

Enrichment of anti-NPS nucleosomes around TSS in humans

We further examined the distribution of anti-NPS nucleosomes around human TSS. Nucleosome occupancy profiles separated into quartiles by transcriptional frequencies based on RNA-seq data (Supplemental Fig. S7) reveal a NDR flanked by weakly phased nucleosomes (Fig. 6A), as illustrated in previous studies (Valouev et al. 2011). Visual inspection of ΔNPS values reveals a dramatic decrease (>10%) for all nucleosomes in the array compared to the genome average (see solid black line in Fig. 6B). This pattern is in marked contrast to that of yeast (compare Fig. 4C and 6B), indicating that anti-NPS

nucleosomes are enriched around TSS in humans. Interestingly, lowly transcribed genes have higher ΔNPS values than highly transcribed genes except for nucleosome +5 (compare blue line and red line in Fig. 6B), suggesting that nucleosomes are more likely to adopt the canonical NPS pattern in lowly transcribed genes than in highly transcribed genes. This result indicates that RNA Pol II transcription plays a role in the rotational settings of nucleosomes in humans.

Overall, our data have shown a clear difference in the distribution of anti-NPS nucleosomes in the yeast and human genomes. That is, anti-NPS nucleosomes are distributed nearly uniform across the yeast genome, whereas in humans, these nucleosomes are more enriched around genes (i.e., in promoters and gene bodies). Moreover, RNA Pol II transcription frequencies have an impact on rotational setting of nucleosomes in human (Fig. 6B), and to a lesser extent, in yeast (Fig. 4C).

Differential DNA sequence patterns around yeast and human genes

To check if the observed difference in the anti-NPS nucleosome distribution between the yeast and human genome is due to underlying DNA sequence, we calculated the distance auto-correlation (DAC) functions (Cui et al. 2012) of WW (or SS) dinucleotides in the TSS-surrounding regions (i.e., between -500 bp and +1000 bp relative to TSS at position 0). Inspection of the DAC functions in yeast reveals a very clear fine structure, with peaks at 11, 21, 31, 42, 52 etc (Fig. 7A and 7B). This pattern suggests that WW (or SS) are separated from each other by a multiple of ~10 bp (denoted as “in phase”). This in-phase pattern for WW and SS dinucleotides is found across the yeast genome (Supplemental Fig. S8). We also calculated the distance cross-correlation (DCC) functions between WW and SS dinucleotides, which show peaks at 3, 16, 26, 36, 47, 56 etc (Supplemental Fig. S9A). This pattern suggests that WW and SS are separated by ~10 × n + 5 bp (where n = 0, 1, 2, …), which is simply denoted as “out of phase”. Our results indicate that yeast DNA is organized in such a way that WW (or SS) are in phase, whereas WW and SS are out of phase. This pattern facilitates the formation of nucleosomes with an optimal rotational setting. Such specific organization of WW and SS fragments across the yeast genome may explain the small deviation

of the ΔNPS values of nucleosomes around TSS from the genome average (Fig. 4C-D and Supplemental Fig. S5).

To understand why there is a small increase of the ΔNPS values in promoters compared to those in gene bodies (Figs. 4C-D and 5), we compared the DAC function of WW dinucleotides in two segments: between -500 bp to 0 bp (TSS) and between 0 bp (TSS) and +1000 bp (Supplemental Fig. S10). We found that although both segments exhibit ~10-bp periodicity of WW dinucleotides, the [-500, 0] segment has higher function values at the peaks than the [0, +1000] segment. This result indicates that the [-500, 0] segment has more in-phase WW dinucleotides than the [0, +1000] segment, suggesting that nucleosomes are more likely to have the NPS pattern compared to the anti-NPS pattern. This may account for the slightly higher ΔNPS values in nucleosomes ±1 compared to nucleosomes +2 to +5 (Fig. 4C-D and 5).

We further examined the DAC functions of WW (or SS) in the TSS-surrounding regions in humans. We found that both DAC functions do not show any recognizable peaks around positions 10, 20, etc (Fig. 7C-D), consistent with earlier studies (Tolstorukov et al. 2009; Reynolds et al. 2009). Similar results were obtained for the DCC function between WW and SS (Supplemental Fig. S9B). This lack of sequence periodicities suggests that nucleosomes can be formed with any rotational settings, which may explain why nucleosomes have low ΔNPS values around TSS (Fig. 6B).

Taken together, we have observed a drastic difference in DNA sequence patterns around TSS in the yeast and human genome. The yeast DNA exhibits an in-phase pattern of WW (or SS) dinucleotides and an out-of-phase pattern between WW and SS fragments. By contrast, no periodic WW or SS pattern is observed in TSS-surrounding regions in humans. This difference clearly explains the patterns of ΔNPS values around TSS in yeast and humans (Fig. 4C and 6B).

Human NPS nucleosomes are often found in repetitive DNA elements

Our data have shown that the average ΔNPS value of all human nucleosomes is ~17% (Fig. 2B), while the average ΔNPS value of nucleosomes -1 to +5 is only ~7% (Fig. 6B). This result suggests that a large

amount of NPS nucleosomes are located far from genic regions. To address where these NPS nucleosomes are preferentially located, we examined repetitive DNA elements that account for ~50% of the human genome (Treangen and Salzberg 2012). We compared repeat families in size (Fig. 8A) with the number of nucleosomes residing in these families (Fig. 8B) and found that nucleosomes are depleted in simple repeats and long terminal repeats (LTR), consistent with previous studies (Samans et al. 2014). We also observed that nucleosomes are enriched in elements of "other repeats" such as MuDR, PiggyBac, TcMar-Mariner, hAT-Charlie, but not in short interspersed nuclear elements (SINE) and long interspersed nuclear elements (LINE) (Fig. 8A-B).

We calculated the fractions of 4 nucleosome types and the average ΔNPS value of nucleosomes residing in each family of repetitive elements (Fig. 8C-D). We found that all families except simple repeats have the average ΔNPS value close to the genome average. LINE-bearing nucleosomes have the highest average ΔNPS value (Fig. 8D), indicating that these nucleosomes are more likely to have the canonical NPS pattern. This finding is consistent with our previous observation that NPS nucleosomes are often found in lowly transcribed genes (Fig. 6B), because transposable elements including SINE and LINE are usually associated with silent chromatin (Slotkin and Martienssen 2007).

Examination of the DAC function of WW dinucleotides in repeats shows that LINE and, to a lesser extent, SINE elements have periodic patterns with peaks around 10 bp (Supplemental Fig. S11C-D). By contrast, the simple repeats have a ~5-bp periodicity (Supplemental Fig. S11A), whereas the LTR and "other repeat" elements have relatively weak periodic patterns (Supplemental Fig. S11B, E). All these results indicate that NPS nucleosomes are more often found in repetitive DNA (especially in LINE elements) than in genic regions, which can be explained by intrinsic DNA sequence patterns in the two groups of genomic regions (compare Supplemental Fig. 11 with Fig. 7C-D).

Discussions

A model for nucleosome positioning determinants in eukaryotic genomes

We have conducted a comprehensive analysis of WW/SS-based sequence patterns from various eukaryotes including yeast and humans. Our work demonstrates that nucleosomes with a non-canonical rotational positioning pattern (i.e., a non-Type-1 pattern) are predominant in the genomes (Fig. 2). Moreover, anti-NPS nucleosomes (i.e., Type 4) are distributed differently around TSS in the yeast and human genomes (Fig. 4 and 6), and such differences are likely associated with intrinsic DNA sequence patterns (Fig. 7). In light of these findings, we propose a model for nucleosome positioning in eukaryotic genomes (Fig. 9).

Both the translational and the rotational positioning nucleosomes can be influenced by *cis*- and *trans*-acting factors such as DNA sequence, chromatin remodeling complexes, and RNA Pol II. However, the relative importance of these factors differs between the two types of nucleosome positioning. For the translational positioning of nucleosomes, *trans* determinants such as chromatin remodeling complexes and RNA Pol II transcription play a major role, as suggested before (Segal and Struhl 2013), whereas for the rotational positioning of nucleosomes, *cis* determinants such as DNA sequence is more important (Fig. 9A).

The WW/SS-based DNA sequence patterns are distributed differently across the yeast and human genome (Fig. 9B). In yeast, WW (or SS) occur with a ~10-bp periodicity across the genome, which favor the formation of nucleosomes with an optimal rotational setting (Fig. 9B). The relative abundance between NPS and anti-NPS nucleosomes varies little across the genome. By contrast, the periodic WW (or SS) patterns are missing in human promoters and genic regions. As a result, nucleosomes can take any rotational setting, which leads to an increase of anti-NPS nucleosomes. Because anti-NPS nucleosomes have unfavorable histone-DNA interactions, they are prone to slide. NPS nucleosomes are often found in repetitive DNA elements (such as LINE and SINE) where periodic WW patterns are present (Fig. 9B).

Association between translational positioning and rotational settings of nucleosomes

It is well established that nucleosome positions is often described by two parameters: translational positioning and rotational settings (Lu et al. 1994). However, how the two aspects of nucleosome positioning are related has never been examined on a genomic scale.

For the first time, our study made the direct comparison between the rotational settings of nucleosomes and their preferred positions around TSS. Specifically, we calculated the ΔNPS values of nucleosomes (e.g., nucleosome -1, +1 to +5) in the array. We found that when nucleosomes have a good translational pattern, like the one around yeast TSS (Fig. 4A), the average ΔNPS value is high (~20%) (Fig. 4C), indicating that nucleosomes prefer to have an optimal rotational setting. When nucleosomes have a poor translational pattern, like the one around human TSS (Fig. 6A), the average ΔNPS value is low (~7%) (Fig. 6D), indicating that an increased number of nucleosomes has the anti-NPS pattern.

This correlation between the translational and the rotational positioning of nucleosomes can be easily explained by our model, in which DNA sequence pattern influences both of them (Fig. 9A). That is, if DNA lacks periodic patterns of WW (and/or SS) dinucleotides, favorable interactions between histones (i.e., sprocket arginine residues) and DNA are unlikely to form, which facilitates the sliding of nucleosomes and therefore contributes to a weak *in vivo* translational nucleosome positioning pattern.

Translational and rotational positioning of nucleosomes in human repetitive elements

Repetitive elements account for ~50% of the human genome (Treangen and Salzberg 2012). Earlier studies have shown that Alu elements that consist of ~10% of the human genome harbor two nucleosomes (Englander et al. 1993; Englander and Howard 1995; Tanaka et al. 2010). Recent studies showed that nucleosomes are depleted in certain types of elements (e.g., simple repeats and LTR), but not in other types (e.g., SINE and LINE) in human sperms (Samans et al. 2014). In our analysis, the human nucleosome data, which are from lymphoblastoid cell lines (Gaffney et al. 2012), are in general agreement with these findings in sperms. It would be interesting to check if nucleosomes in other cell

types such as embryonic stem cells also follow this occupancy pattern in repeat families. It has been known that various reprogramming factors such as Oct4, Sox2, Klf4 preferentially bind to repetitive DNA elements (Kunarso et al. 2010), and they are able to bind nucleosomal DNA (Soufi et al. 2015).

We and other groups have shown that the rotational positioning of nucleosomes is critical for TF binding including p53 (Sahu et al. 2009; Cui and Zhurkin 2014). We have found that an Alu-residing nucleosome, if taking the optimal rotational setting, help to expose the putative binding sites of p53 (Cui et al. 2010). The present study further illustrates that nucleosomes are likely to take the optimal rotational setting in repeats (Fig. 8D), which facilitates the binding of repeat elements by different TFs including p53. Further analyses are needed to address the biological roles of anti-NPS nucleosomes to see if, in these cases, TF binding sites are buried inside the nucleosome and inhibitory for protein binding, as we suggested for p53 (Cui and Zhurkin 2014). Understanding the translational and rotational positioning of nucleosomes in human repeats will shed new lights on mechanisms regulating gene transcription in repeats in normal and diseased cellular context.

Materials and Methods

Minor- and major-groove bending sites in a nucleosomal DNA fragment

Minor- and major-groove bending sites (GBS) in a 147-bp nucleosomal DNA fragment were defined previously (Cui and Zhurkin 2010). Briefly, the fragments whose minor grooves face the histone octamer are called minor-GBS, located at the superhelical locations (SHL) ±0.5, ±1.5, …, ±6.5, while the fragments whose minor grooves face away from the histone are called major-GBS, located at SHL ±1, ±2, …, ±6. Each site is 3 or 4 base pairs in length. In total, there are 14 minor-GBS and 12 major-GBS along a 147-bp NCP fragment. Note that the minor-GBS at SHL ±0.5 (shown in grey in Supplemental Figure S1) are not included for analysis because out-of-phase WW/SS peaks are observed at this site (Satchwell et al. 1986). As a result, there are 12 minor-GBS covering 48 base pairs, and 12 major-GBS covering 44 base pairs (Supplemental Table S1). To account for this size difference between minor- and major-GBS, we assign a coefficient 1.0909 (=48/44) to major-GBS (Table 1).

Classification of nucleosomes with different NPS patterns

Nucleosomal DNA sequences are divided into 4 types based on relative abundance of WW and SS dinucleotides in minor- and major-GBS (Table 1). Type 1 nucleosomes have the canonical WW/SS pattern (i.e., WW are abundant in minor-GBS, whereas SS are abundant in major-GBS). Type 4 nucleosomes have the anti-NPS pattern (i.e., WW are abundant in major-GBS, whereas SS are abundant in minor-GBS). Type 2 and 3 nucleosomes have ‘mixed’ patterns, in which both WW and SS are abundant in minor-GBS (Type 2) or in major-GBS (Type 3). The difference between Type 1 and Type 4 nucleosomes in percentage (%) is denoted as ΔNPS. That is, ΔNPS = Type 1 (%) – Type 4 (%).

High-resolution nucleosome datasets in yeast and higher eukaryotes

High-resolution nucleosomal DNA datasets from various eukaryotes were used in this study (Supplemental Table S2), all of which were mapped by the paired-end sequencing technique. Because this

technique generates short reads on both ends of a nucleosomal DNA fragment, the length of the fragment can be precisely determined. This length information is critically important because only the fragments with 147 bp in length were analyzed in this study. This is because that on these fragments, minor- and major-GBS can be determined in an unambiguous manner (Cui and Zhurkin 2010, Supplemental Fig. 1).

Paired-end nucleosome sequence data from yeast include wildtype cells (YDC111 (Cole et al. 2011), YPE458 (Ocampo et al. 2016), and MCY3647 (Ganguli et al. 2015) and mutant (MT) cells (Ganguli et al. 2015; Ocampo et al. 2016) including single mutants (*isw1Δ*, *isw2Δ*, *chd1Δ*, *rsc8Δ*), double mutants (*isw1Δ isw2Δ, isw1Δ chd1Δ, isw2Δ chd1Δ*), and triple mutants (*isw1Δ isw2Δ chd1Δ*). The paired reads were aligned to the *Saccharomyces cerevisiae* genome using ELAND (Illumina). Only the reads uniquely aligned to the genome with no mismatch were selected. Note that all yeast strains (wildtype or mutants) were grown to mid-log phase in synthetic complete (SC) medium. All of the wildtype strains (YDC111, YPE458 and MCY3647) should be equivalent for most purposes, since they grew in the same SC medium.

Paired-end MNase-seq data were also taken from *C. elegans* (Steiner and Henikoff 2014), *Drosophila* S2 cells (Chereji et al. 2016), mouse E14 ES cells (West et al. 2014), and human lymphoblastoid cell lines (Gaffney et al. 2012). BAM files were either downloaded from NCBI GEO database or obtained by mapping raw reads to the corresponding genomes using the default setting of Bowtie (Langmead et al. 2009)

Chromatin remodeler ChIP-seq datasets

Raw ChIP-seq data for yeast chromatin remodelers including Rsc (Ramachandran et al. 2015), Isw1, Isw2 and Chd1 (Zentner et al. 2012) were mapped to the genome using the default setting of Bowtie (Langmead et al. 2009). The $\log_2$(IP/input) ChIP-seq signal for each base pair ±1 kb from the validated TSS of yeast genes, separated into quartiles by transcriptional frequencies, was averaged to made the aggregate plots.

Genome-wide gene expression (RNA-seq) data in yeast and humans

RNA-seq data of human granulocytes were used in this study, and the RPKM (Reads Per Kilobase of transcript per Million) values were taken from the literature (Valouev et al. 2010). Transcriptional landscapes of the yeast genome have been studied by RNA-seq experiments in yeast cells growing under 18 different environment conditions, including the SC medium. Because nucleosomes were extracted from cells growing in the SC medium, the RNA-seq data for the SC condition were taken for analysis (Waern and Snyder 2013). Raw reads were mapped to the yeast genome by the default setting of Bowtie (Langmead et al. 2009). The RPKM values of the yeast data were calculated following the same method as the human data (Valouev et al. 2010).

Repeat elements in the human genome

Human repetitive region positions in human genome assembly hg19 were downloaded from the UCSC Genome Browser. The repeat elements were identified using RepeatMasker (v3.2.7) and Repbase Update (9.11). The major types of repeat elements were selected for analysis, including SINE (Alu, MIR), LINE (CR1, L1, L2 and RTE), LTR (ERV1, ERVK, ERVL and Gypsy), Simple Repeat ($(TG)_n$, $(TCG)_n$, $(CACTC)_n$, $(GAGTG)_n$, and $(TATATG)_n$). The remaining repeat elements were included into the "Other" category including MuDR, PiggyBac, TcMar-Mariner, hAT-Charlie. Only the elements with >150 bp in length (i.e., approximately the size of one nucleosome) were selected for analysis.

Since human nucleosomes were mapped to the genome assembly hg18 (Gaffney et al. 2012), the UCSC LiftOver utility was used along with the hg18-hg19 chain file to convert the hg19 repeat elements to their corresponding positions in human genome assembly hg18 (Supplemental Table S4). Nucleosomes overlapped with these repeat families were summarized in Supplemental Table S5.

Distance auto- and cross-correlation function

The distance auto-correlation (DAC) and distance cross-correlation (DCC) functions have been discussed previously (Zhurkin 1981; Cui et al. 2012). In the present study, DAC function was used to calculate the correlation between WW or SS dinucleotides. That is, if a WW dinucleotide (AA, TT, AT or TA) is separated from another WW dinucleotide by a distance *d*, one occurrence is counted for that distance in the DAC function. For the DCC function, we counted how many times a WW dinucleotide is separated from a SS dinucleotide (GG, CC, GC or CG) by a distance *d*, where *d* is between 1 and 150 bp.

Disclosure of potential conflict of interest

No potential conflicts of interest were disclosed.

Funding

The research was supported by a NIH grant R15GM116102 (to F.C.)

References

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

Baldi P, Brunak S, Chauvin Y, Krogh A. 1996. Naturally occurring nucleosome positioning signals in human exons and introns. *J Mol Biol* **263**: 503-510.

Calladine CR, Drew HR. 1986. Principles of sequence-dependent flexure of DNA. *J Mol Biol* **192**: 907-918.

Chereji RV, Kan TW, Grudniewska MK, Romashchenko AV, Berezikov E, Zhimulev IF, Guryev V, Morozov AV, Moshkin YM. 2016. Genome-wide profiling of nucleosome sensitivity and chromatin accessibility in Drosophila melanogaster. *Nucleic Acids Res* **44**: 1036-1051.

Chua EYD, Vasudevan D, Davey GE, Wu B, Davey CA. 2012. The mechanics behind DNA sequence-dependent properties of the nucleosome. *Nucleic Acids Res* **40**: 6338-6352.

Cole HA, Howard BH, Clark DJ. 2011. The centromeric nucleosome of budding yeast is perfectly positioned and covers the entire centromere. *Proc Natl Acad Sci USA* **108**: 12687-12692.

Cui F, Chen L, LoVerso PR, Zhurkin VB. 2014. Prediction of nucleosome rotational positioning in yeast and human genomes based on sequence-dependent DNA anisotropy. *BMC Bioinformatics* **15**: 313.

Cui F, Sirotin MV, Zhurkin VB. 2011. Impact of Alu repeats on the evolution of human p53 binding sites. *Biol Direct* **6**:2

Cui F, Zhurkin VB. 2010. Structure-based analysis of DNA sequence patterns guiding nucleosome positioning in vitro. *J Biomol Struct Dyn* **27**: 821-841.

Cui F, Zhurkin VB. 2014. Rotational positioning of nucleosomes facilitates selective binding of p53 to response elements associated with cell cycle arrest. *Nucleic Acids Res* **42**: 836-847.

Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ. 2002. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9Å resolution. *J Mol Biol* **391**: 1097-1113.

Drew HR, Travers AA. 1985. DNA bending and its relation to nucleosome positioning. *J Mol Biol* **186**: 773-790.

Englander EW, Howard BH 1995. Nucleosome positioning by human Alu elements in chromatin. *J Biol Chem* **270**: 10091-10096.

Englander EW, Wolffe AP, Howard BH. 1993. Nucleosome interactions with a human Alu element. *J Biol Chem* **268**: 19565-19573.

Field Y, Kaplan N, Fondufe-Mittendorf Y, Moore IK, Sharon E, Lubling Y, Widom J, Segal E. 2008. Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput Biol* **4**: e1000216.

Gaffney DJ, McVicker G, Pai AA, Fondufe-Mittendorf YN, Lewellen N, Michelini K, Widom J, Gilad Y, Pritchard JK. 2012. Controls of nucleosome positioning in the human genome. *PLoS Genet* **8**: e1003036.

Ganguli D, Chereji RV, Iben JR, Cole HA, Clark DJ. 2014. RSC-dependent constructive and destructive interference between opposing arrays of phased nucleosomes in yeast. *Genome Res* **24**: 1637-1649.

Gkikopoulos T, Schofield P, Singh V, Pinskaya M, Mellor J, Smolle M, Workman JL, Barton GJ, Owen-Hughes T. 2011. A role for Snf2-related nucleosome-spacing enzymes in genome-wide nucleosome organization. *Science* **333**, 1758-1760.

Hughes A, Jin Y, Rando OJ, Struhl K. 2012. A functional evolutionary approach to identify determinants of nucleosome positioning: a unifying model for establishing the genome-wide pattern. *Mol. Cell* **48**: 5-15.

Ioshikhes I, Bolshoy A, Derenshteyn K, Borodovsky M, Trifonov EN. 1996. Nucleosome DNA sequence pattern revealed by multiple alignment of experimentally mapped sequences. *J Mol Biol* **262**: 129-139.

Ioshikhes I, Bolshoy A, Trifonov EN. 1992. Preferred positions of AA and TT dinucleotides in aligned nucleosome DNA sequences. *J Biomol Struct Dyn* **9**: 1111-1117.

Ioshikhes I, Hosid S, Pugh BF. 2011. Variety of genomic DNA patterns for nucleosome positioning. *Genome Res* **21**: 1863-1871.

Jiang C, Pugh BF. 2009. A compiled and systematic reference map of nucleosome positions across the Saccharomyces cerevisiae genome. *Genome Biol* **10**: R109.

Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Tillo D, Field Y, LeProust EM, Hughes TR, Lieb JD, Widom J, Segal E. 2009. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**: 362-366.

Kogan SB, Kato M, Kiyama R, Trifonov EN. 2006. Sequence structure of human nucleosome DNA. *J Biomol Struct Dyn* **24**: 43-48.

Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G. 2010. Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat Genet* **42**: 631-634.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Li Q, Wrange O. 1995. Accessibility of a glucocorticoid response element in a nucleosome depends on its rotational positioning. *Mol Cell Biol* **15**: 4375-4384.

Lowary PT, Widom J. 1998. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* **276**: 19-42.

Lu Q, Wallrath LL, Elgin SC. 1994. Nucleosome positioning and gene regulation. *J Cell Biochem* **55**: 83-92.

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

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, Gilmour DS, Albert I, Pugh BF. 2008b. Nucleosome organization in the Drosophila genome. *Nature* **453**: 358-362.

Mengeritsky G, Trifonov EN. 1983. Nucleotide sequence-directed mapping of the nucleosomes. *Nucleic Acids Res* **11**: 3833-3851.

Muthurajan UM, Bao Y, Forsberg LJ, Edayathumangalam RK, Suto RK, Chakravarthy S. Dyer PN, Luger K. 2003. Structure and dynamics of nucleosomal DNA. *Biopolymers* **68**: 547-556.

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.

Ocampo J, Chereji RV, Eriksson PR, Clark DJ. 2016. The ISW1 and CHD1 ATP-dependent chromatin remodelers compete to set nucleosome spacing in vivo. *Nucleic Acids Res* **44**: 4625-4635.

Radman-Livaja M, Rando OJ. 2009. Nucleosome positioning: How is it established, and why does it matter? *Dev Biol* **339**: 258-266.

Ramachandran S, Zentner G, Henikoff S. 2015. Asymmetric nucleosomes flank promoters in the budding yeast genome. *Genome Res* **25**: 381-390.

Raveh-Sadka T, Levo M, Shabi U, Shany B, Keren L, Lotan-Pompan M, Zeevi D, Sharon E, Weinberger A, Segal E. 2012. Manipulating nucleosome disfavoring sequences allows fine-tune regulation of gene expression in yeast. *Nat Genet* **44**: 743-750.

Reynolds SM, Bilmes JA, Noble WS. 2009. On the Relationship between DNA Periodicity and Local Chromatin Structure. In: Batzoglou S. (eds) Research in Computational Molecular Biology. RECOMB 2009. *Lecture Notes in Computer Science*, vol. 5541. Springer, Berlin, Heidelberg

Rhee HS, Pugh BF. 2012. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**: 295-301.

Rohs R, West SM, Sosinsky A, Liu P, Mann RS, Honig B. 2009. The role of DNA shape in protein-DNA recognition. *Nature* **461**: 1248-1253.

Sahu G, Wang D, Chen CB, Zhurkin VB, Harrington RE, Appella E, Hager GL, Nagaich AK. 2009. p53 binding to nucleosomal DNA depends on the rotational positioning of DNA response element. *J Biol Chem* **285**:1321-1332.

Samans B, Yang Y, Krebs S, Sarode GV, Blum H, Reichenbach M, Wolf E, Steger K, Sansranjavin T, Schagdarsurengin U. 2014. Uniformity of nucleosome preservation pattern in mammalian sperm and its connection to repetitive DNA elements. *Dev Cell* **30**: 23-35.

Satchwell SC, Drew HR, Travers AA. 1986. Sequence periodicities in chicken nucleosome core DNA. *J Mol Biol* **191**: 659-675.

Schones DE, Cui K, Cuddapah S, Roh T, Barski A, Wang Z, Wei G, Zhao K. 2008 Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**: 887-898.

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

Sekinger EA, Moqtaderi Z, Struhl K. 2005. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. *Mol Cell* **18**: 736-748.

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.

Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**: 272-285.

Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. 2015. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**: 555-568.

Steiner FA, Henikoff S. 2014. Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. *Elife* **3**: e02025.

Struhl K, Segal E. 2013. Determinants of nucleosome positioning. *Nat Struct Mol Biol* **20**: 267-273.

Sullivan SA and Landsman D. 2003. Characterization of sequence variability in nucleosome core histone folds. *Proteins* **52**: 454-465.

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.

Tanaka Y, Yamashita R, Suzuki Y, Nakai K. 2010. Effects of Alu elements on global nucleosome positioning in the human genome. *BMC Genomics* **11**: 309.

Tillo D and Hughes TR. 2009. G+C content dominates intrinsic nucleosome occupancy. *BMC Bioinformatics* **10**: 442.

Tillo D, Kaplan N, Moore IK, Fondufe-Mittendorf Y, Gossett AJ, Field Y, Lieb JD, Widom J, Segal E, Hughes TR. 2010. High nucleosome occupancy is encoded at human regulatory sequence. *PLoS One* **5**: e9129.

Tolstorukov MY, Kharchenko PV, Goldman JA, Kingston RE, Park PJ. 2009. Comparative analysis of H2A.Z nucleosome organization in the human and yeast genomes. *Genome Res* **19**: 967-977.

Treangen TJ, Salzberg SL. 2012. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* **13**: 36-46.

Trifonov EN. 2011. Thirty years of multiple sequence codes. *Genomics Proteomics Bioinformatics* **9**: 1-6.

Uberbacher EC, Harp JM, Bunick GJ. 1988. DNA sequence patterns in pre*cis*ely positioned nucleosome. *J Biomol Struct Dyn* **6**: 105-120.

Vaillant C, Palmeira L, Chevereau G, Audit B, d'Aubenton-Carafa Y, Thermes C, Arneodo A. 2010. A novel strategy of transcription regulation by intragenic nucleosome ordering. *Genome Res* **20**: 59-67.

Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. 2011. Determinants of nucleosome organization in primary human cells. *Nature* **474**: 516-520.

Waern K, Snyder M. 2013. Extensive transcript diversity and novel upstream open reading frame regulation in yeast. *G3 Genes Genom Genet* **3**: 343-352.

Wang D, Ulyanov NB, Zhurkin VB. 2010. Sequence-dependent Kink-and-Slide deformations of nucleosomal DNA facilitated by histone arginines bound in the minor groove. *J Biomol Struct Dyn* **27**: 843-859.

West JA, Cook A, Alver BH, Stadtfeld M, Deaton AM, Hochedlinger K, Park PJ, Tolstorukov MY, Kingston RE. 2014. Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nat Commun* **5**: 4719.

West SM, Rohs R, Mann RS, Honig B. 2010. Electrostatic interactions between arginines and the minor groove in the nucleosomes. J Biomol Struct Dyn 27: 861-866.

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

Zentner GE, Tsukiyama T, Henikoff S. 2012. ISW1 and CHD chromatin remodelers bind promoters but act in gene bodies. *PLoS Genet* **9**: e1003317.

Zhang Y, Mogtaderi Z, Rattner BP, Euskirchen G, Snyder M, Kadonaga JT, Liu XS, Struhl K. 2009. Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions in vivo. *Nat Struct Mol Biol* **16**: 847-852.

Zhang Z, Wippo CJ, Wal M, Ward E, Korber P, Pugh BF. 2011 A packing mechanism for nucleosome organization reconstituted across a eukaryotic genome. *Science* **20**: 977-980.

Zhurkin VB. 1981. Periodicity in DNA primary structure is defined by secondary structure of the coded protein. *Nucleic Acids Res* **9**: 1963-1971.

Zhurkin VB. 1983. Specific alignment of nucleosomes on DNA correlates with periodic distribution of purine-pyrimidine and pyrimidine-purine dimers.

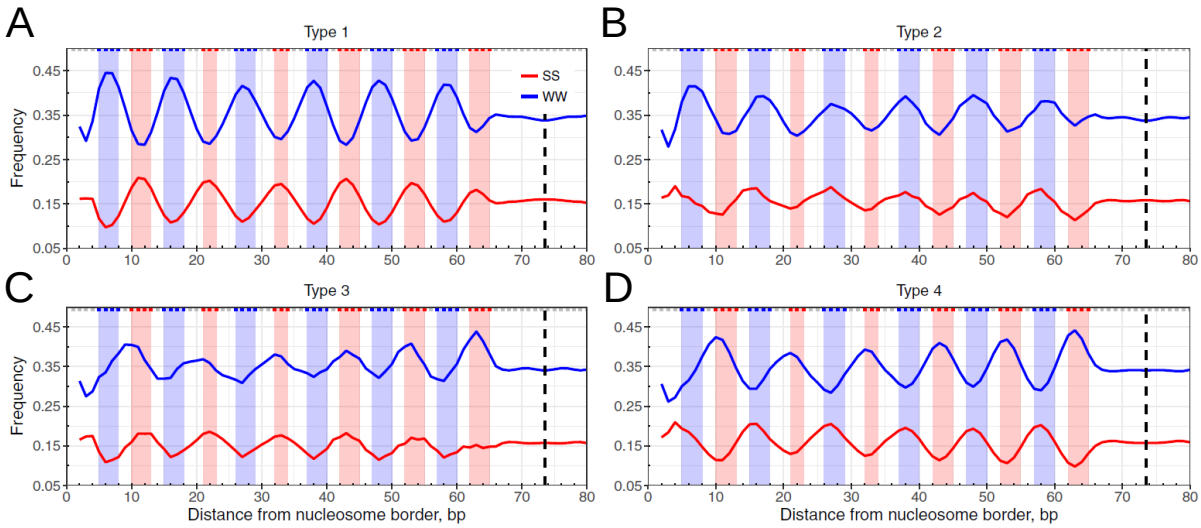

Figure 1. Four sequence patterns of nucleosomal DNA in yeast. Smoothed frequencies of the combined number of AA, TT, AT, and TA (denoted as WW, shown in blue) and GG, CC, GC and CG (denoted as SS, shown in red) at each nucleotide position were 'symmetrized' with respect to the dyad (dashed lines). Three base-pair running averages of the WW and SS frequencies were calculated for Type 1 (A), Type 2 (B), Type 3 (C) and Type 4 (D) nucleosomes. The blue and red shade areas represent the locations of minor- and major-GBS, respectively (see Methods).

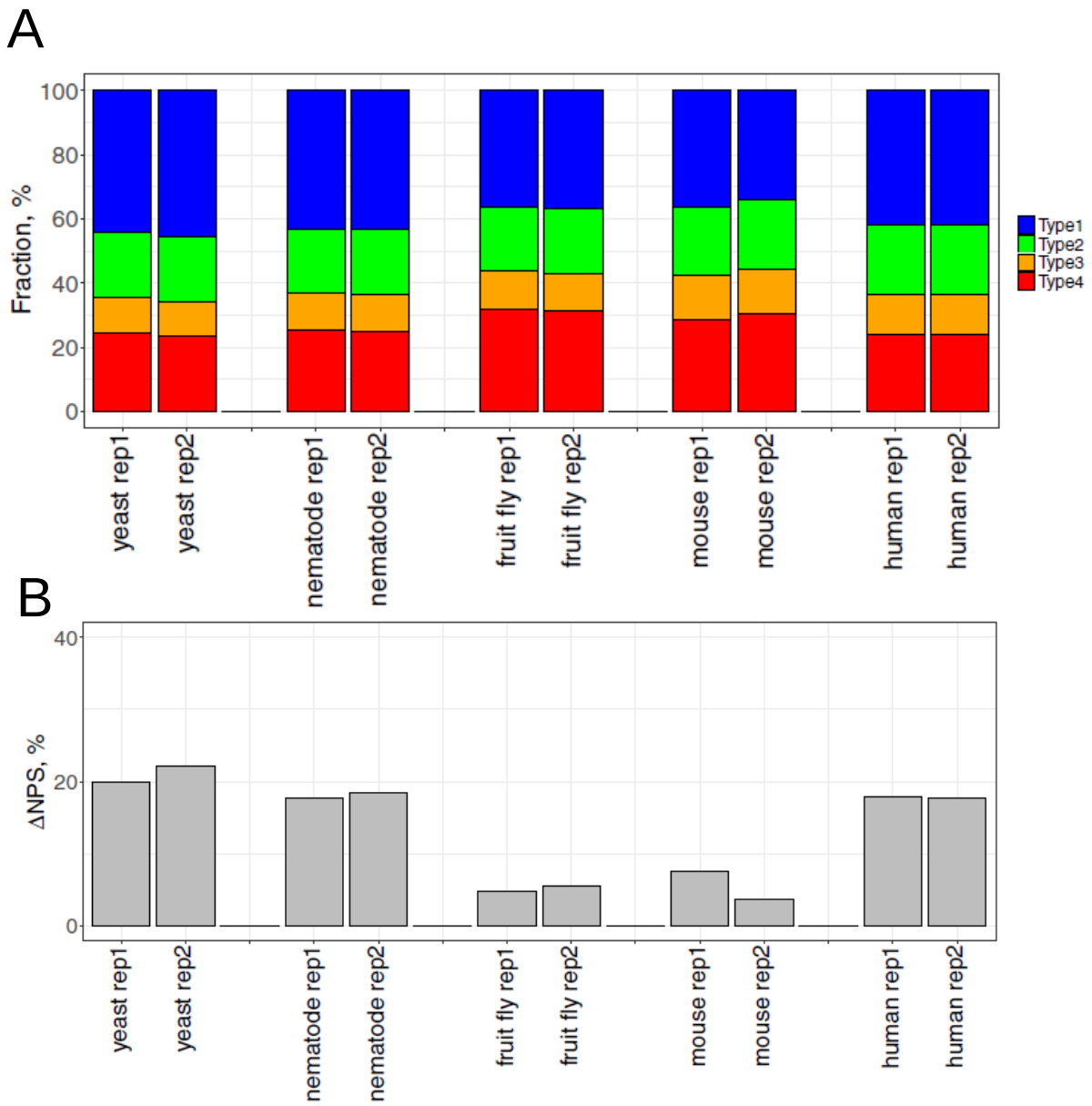

Figure 2. Comparison of four nucleosomal DNA patterns in eukaryotes. (A) Fractions of four sequence patterns in nucleosomal DNA from yeast, nematode, fruit fly, mouse and human. For each species, two biological replicates were used for analysis. (B) ΔNPS values of the replicates.

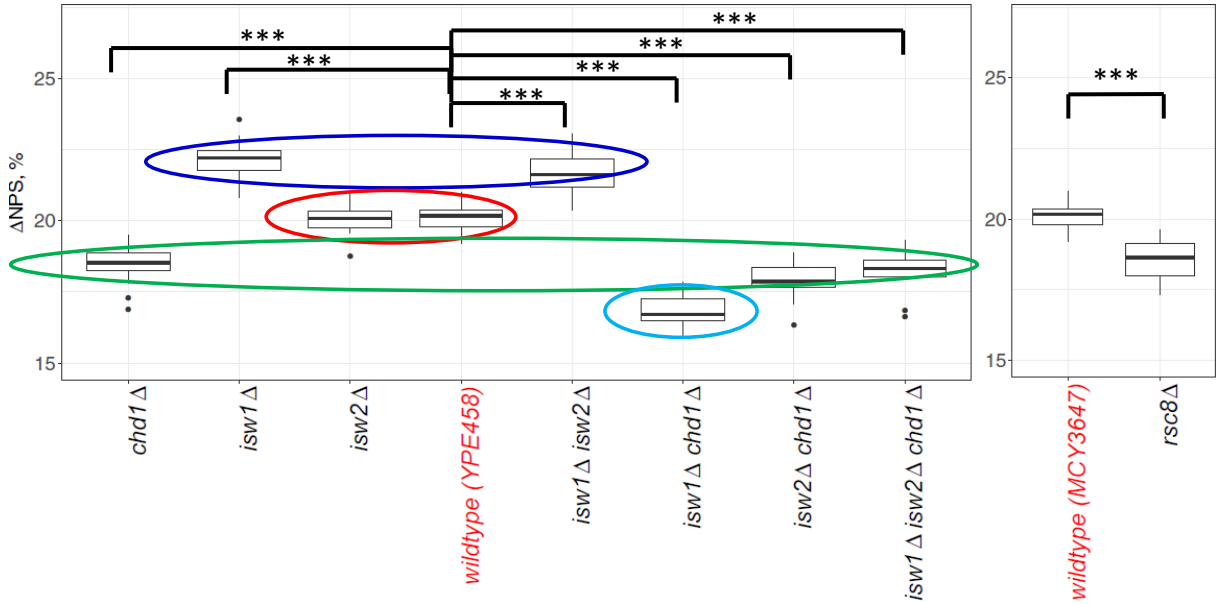

Figure 3. Average ΔNPS values of nucleosomes in wildtype and mutant strains. ΔNPS values were calculated by chromosomes and shown in box-whisker plots. Yeast strains (wildtype and mutants) were clustered into groups (denoted by ovals) based on their ΔNPS values. Triple stars represent $p < 10^{-5}$ (Student's t-test).

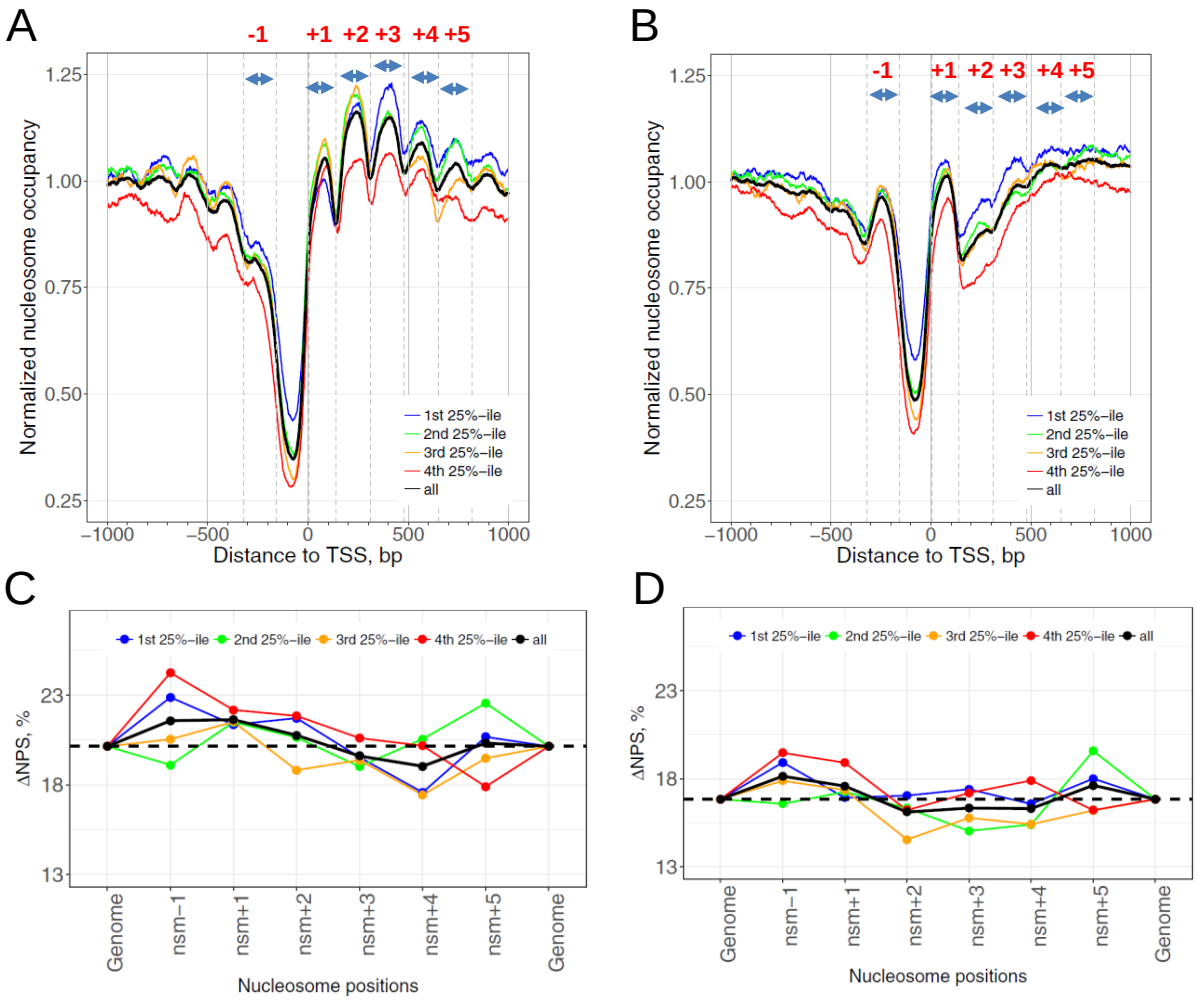

Figure 4. Occupancy profiles and average ΔNPS values of nucleosomes around TSS in yeast. Aggregate plots of nucleosome occupancy are shown for the wildtype (A) and mutant strains (*isw1Δ chd1Δ* double mutant) (B). Nucleosome occupancy signals ±1 kb of verified TSSs are separated into quartiles by transcriptional frequency based on RNA-seq data (Waern and Snyder 2013). Nucleosomes -1 to +5 are demarcated by vertical dashed lines and arrows (A, B). (C, D) Average ΔNPS values for the nucleosomes -1 to +5 of the wildtype (C) and the mutant (D) strain. The genome average of the ΔNPS values is denoted by horizontal dashed lines (C, D). Nucleosome zones that demarcate nucleosomes -1 to +5 are obtained following the methods used in previous studies (Jiang and Pugh 2009). These definable zones relative to the TSS (position 0) to which a nucleosome midpoint may be assigned are: -1 (from -320 to -160), +1 (from +5 to +140), +2 (from +141 to +310), +3 (from +311 to +480), +4 (from +481 to +650), and +5 (from +651 to +820). For the sake of comparison, the nucleosome zones are the same for both the wildtype and mutant strains.

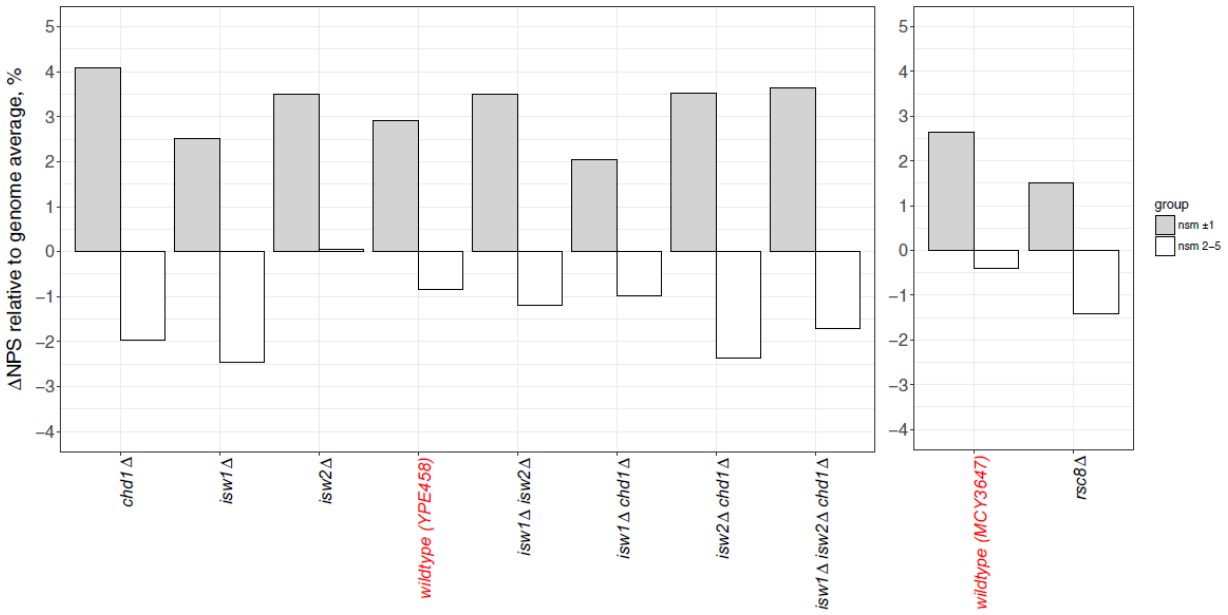

Figure 5. Normalized ΔNPS values of nucleosomes in promoters and in gene bodies. The ΔNPS values were normalized by subtracting the genome-average ΔNPS value. Nucleosomes ±1 are considered to be in promoters (shown in gray) and nucleosomes +2 to +5 are considered to be in gene bodies (shown in white). The normalized ΔNPS values in each group were summed up.

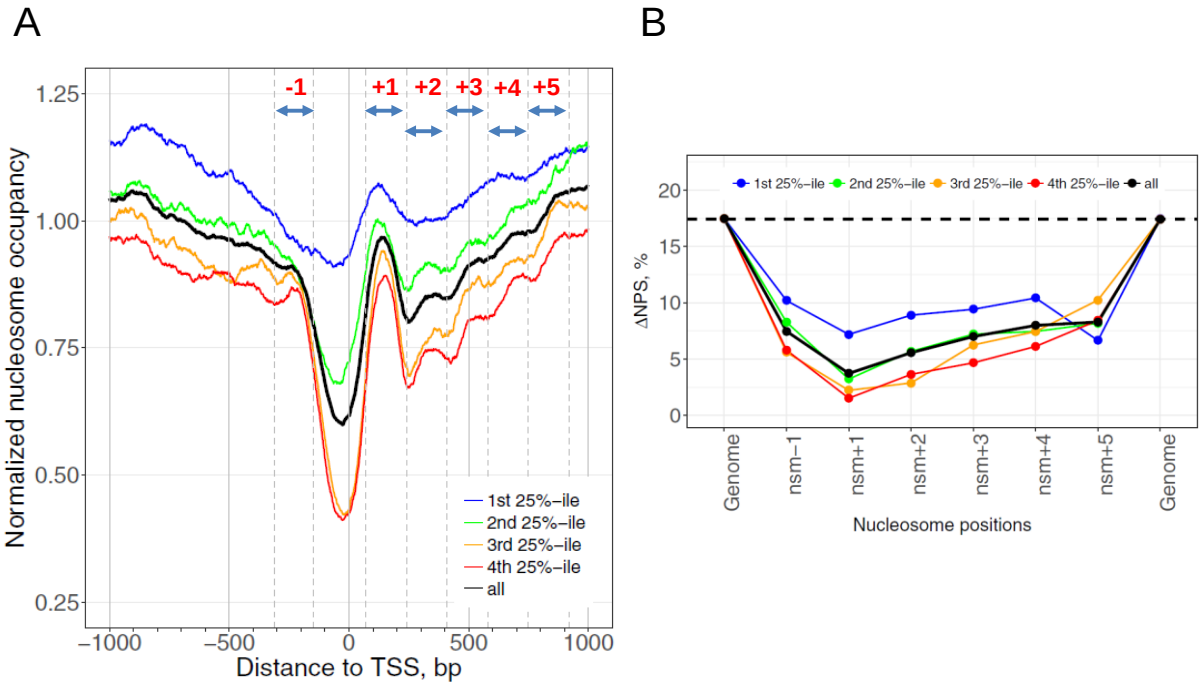

Figure 6. Occupancy profiles and the average ΔNPS values of nucleosomes around TSS in human. (A) Aggregate plot of nucleosome occupancy signals ±1 kb of verified TSS, separated into quartiles by transcriptional frequency based on RNA-seq data (Valouev et al. 2010). Nucleosomes -1 to +5 are demarcated by vertical dashed lines and arrows. (B) The average ΔNPS values of the nucleosomes around TSS. The genome average ΔNPS value is denoted by a horizontal dashed line. Definable nucleosome zones relative to the TSS (position 0) to which a nucleosome midpoint may be assigned are: -1 (from -310 to -150), +1 (from +70 to +240), +2 (from +241 to +410), +3 (from +411 to +580), +4 (from +581 to +750), and +5 (from +751 to +920).

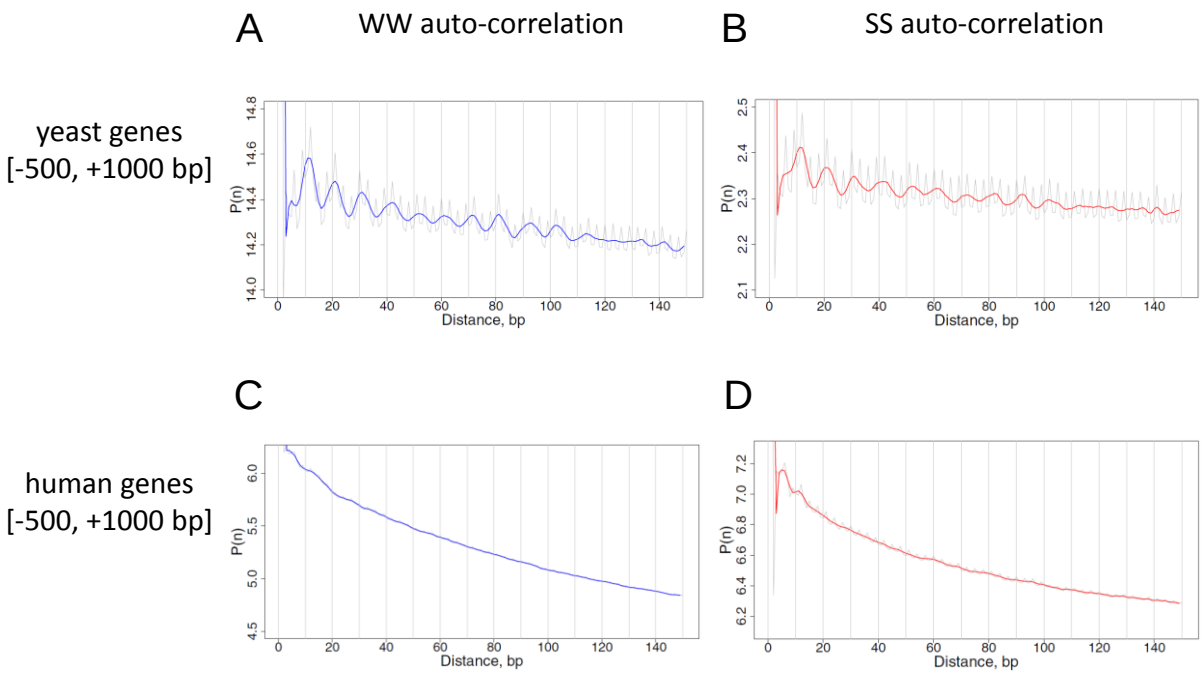

Figure 7. Distance auto-correlation function for WW (blue) and SS (red) dinucleotides in yeast (A-B) and human (C-D) genes. Genomic fragments [-500 bp, +1000 bp] relative to verified TSSs (position 0) were used for analysis. Both raw function values (in gray) and 3-bp running average values (in colors) were profiled.

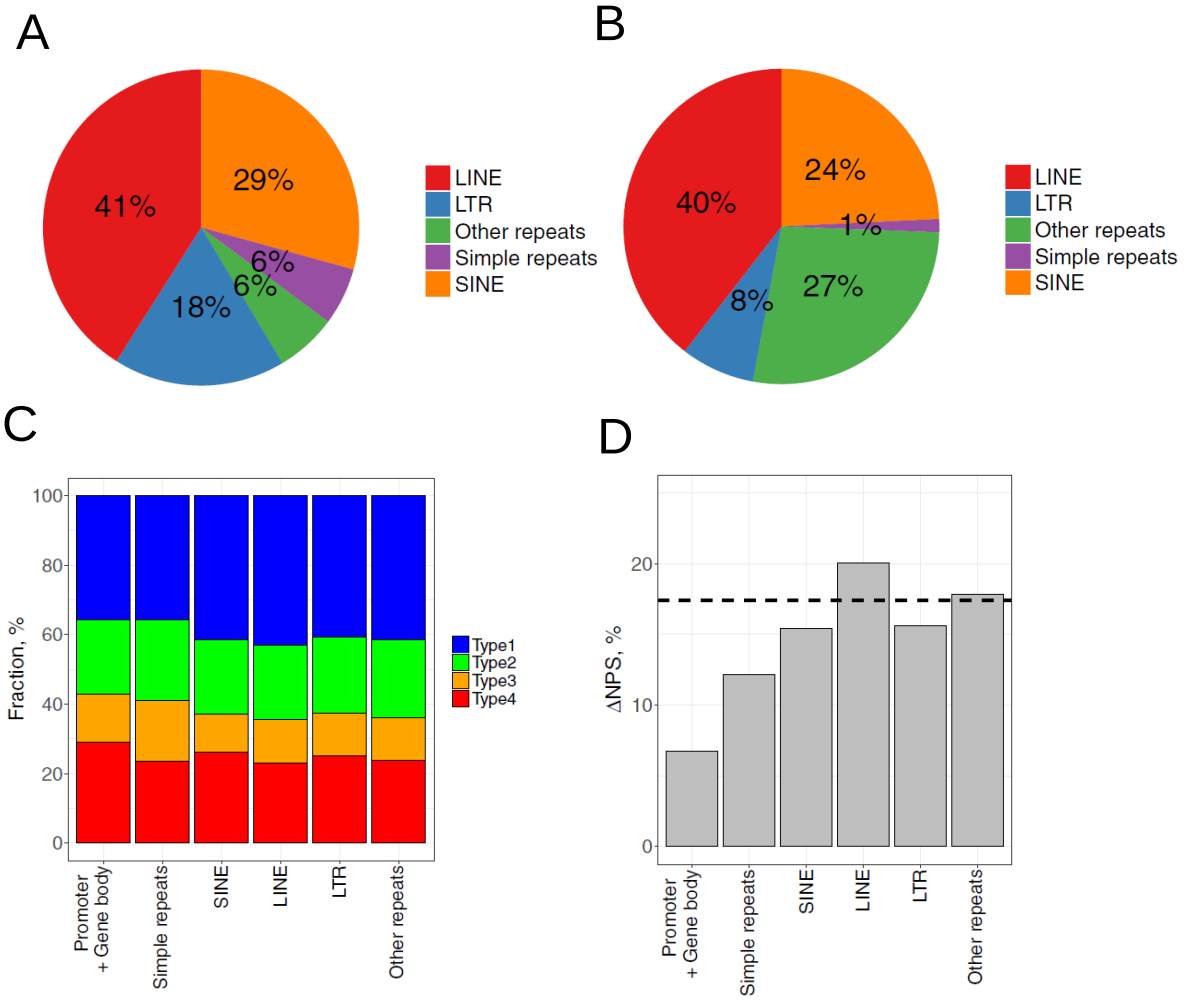

Figure 8. Types of nucleosomal DNA sequence patterns in human transposable elements (TEs). (A) Fractions of human TE families. The fractions of TE families in the human genome were taken from literature (Treangen and Salzberg 2012). (B) Fractions of 147-bp human nucleosomes residing in TEs grouped by their families. (C) Fractions of 4 types of nucleosomes in TEs grouped by families. (D) Average ΔNPS values for nucleosomes in genic and repetitive DNA regions. The genome average ΔNPS value is indicated by horizontal dashed lines.

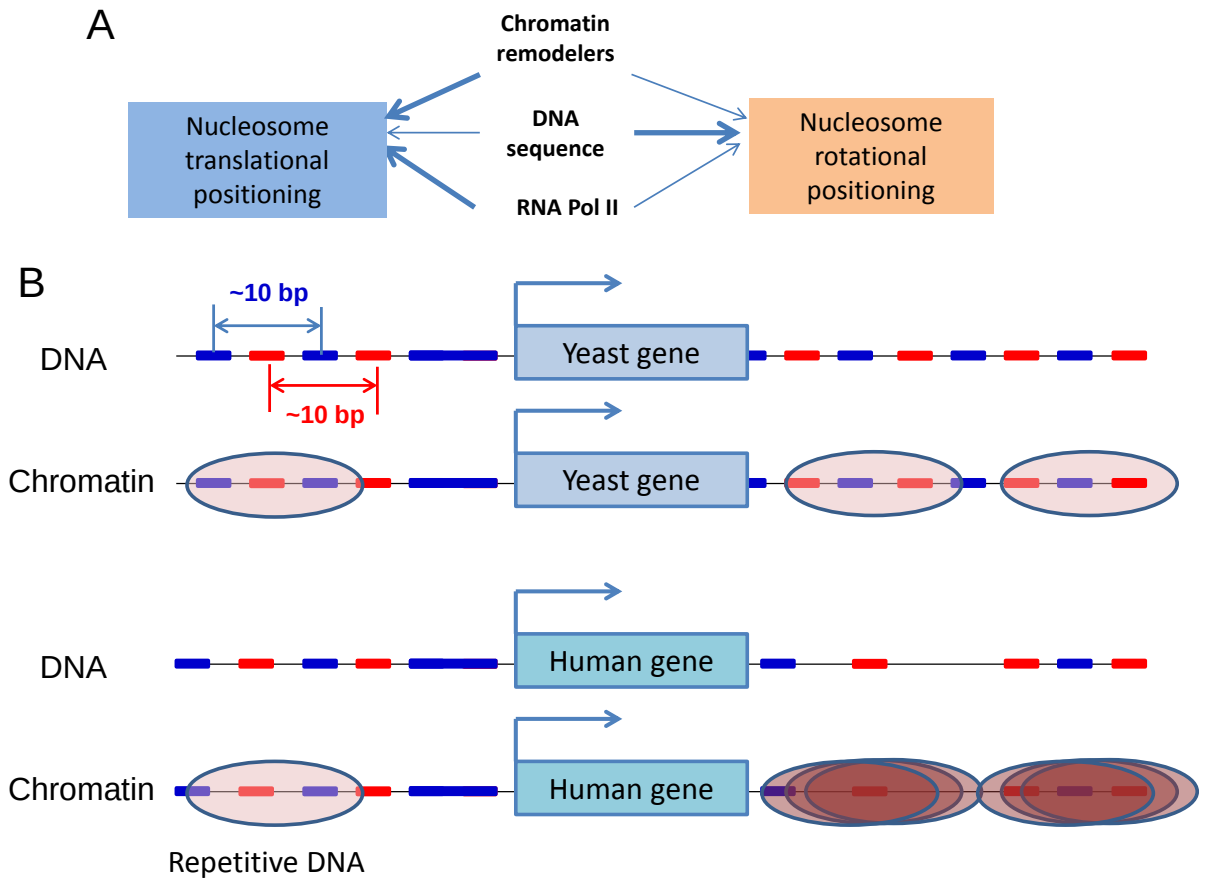

Figure 9. Models for *cis* and *trans* nucleosome positioning determinants. (A) *Cis* and *trans* factors associated with nucleosome rotational settings and translational positioning. The *cis*-acting factors such as DNA sequence pattern play a major role (indicated by thick arrow), whereas trans-acting factors such as chromatin remodeling complexes and RNA Pol II transcription play minor roles (indicated by thin arrows) in establishing nucleosome rotational setting. By contrast, the *trans* factors play a major role in determining translational positioning of nucleosomes. (B) Differential rotational positioning of nucleosomes around yeast and human genes. In yeast DNA, both WW (blue) and SS (red) dinucleotides are regularly positioned with a ~10-bp periodicity around genes and across the genome. WW dinucleotides are preferentially out of phase with SS dinucleotides (i.e., separated by 10n + 5 bp, where n ≥ 1). This unique DNA pattern facilitates the formation of nucleosomes with optimal rotational settings (represented by well-separated ovals). In human, however, a ~10-bp periodicity of WW or SS dinucleotides is not present around TSS, suggesting that nucleosomes can take any rotational setting (represented by overlapping ovals). Nucleosomes with optimal rotational settings are often observed in repetitive DNA regions including SINEs and LINEs, which are usually associated with silent chromatin (Slotkin and Martienssen 2007).

Table 1. Classification of nucleosomal DNA sequence patterns

Sequence Type	WW	SS
Type 1	minor WW >= Major WW * coef[†]	minor SS <= Major SS * coef
Type 2	minor WW >= Major WW * coef	minor SS > Major SS * coef
Type 3	minor WW < Major WW * coef	minor SS <= Major SS * coef
Type 4	minor WW < Major WW * coef	minor SS > Major SS * coef

[†]coef: the coefficient is a constant of 1.0909 (=48/44). It refers to the ratio between 48 nucleotide positions in minor-groove bending sites and 44 nucleotide positions in major-groove bending sites (see Supplemental Table S1).