

# SCIENTIFIC REPORTS



OPEN

## The implication of DNA bending energy for nucleosome positioning and sliding

Guoqing Liu<sup>1,2</sup>, Yongqiang Xing<sup>1,2</sup>, Hongyu Zhao<sup>1,2</sup>, Lu Cai<sup>1,2</sup> & Jianying Wang<sup>1</sup>

Nucleosome not only directly affects cellular processes, such as DNA replication, recombination, and transcription, but also serves as a fundamentally important target of epigenetic modifications. Our previous study indicated that the bending property of DNA is important in nucleosome formation, particularly in predicting the dyad positions of nucleosomes on a DNA segment. Here, we investigated the role of bending energy in nucleosome positioning and sliding in depth to decipher sequence-directed mechanism. The results show that bending energy is a good physical index to predict the free energy in the process of nucleosome reconstitution *in vitro*. Our data also imply that there are at least 20% of the nucleosomes in budding yeast do not adopt canonical positioning, in which underlying sequences wrapped around histones are structurally symmetric. We also revealed distinct patterns of bending energy profile for distinctly organized chromatin structures, such as well-positioned nucleosomes, fuzzy nucleosomes, and linker regions and discussed nucleosome sliding in terms of bending energy. We proposed that the stability of a nucleosome is positively correlated with the strength of the bending anisotropy of DNA segment, and both accessibility and directionality of nucleosome sliding is likely to be modulated by diverse patterns of DNA bending energy profile.

By folding and wrapping with the aid of proteins, eukaryotic genomes are packaged into high-order chromatin structure. Genome packaging is of great interest not because it condenses long DNA molecules to accommodate to the confined space of nucleus, but because chromatin architecture, which is the product of the packaging, regulates or gets involved in many fundamentally important molecular processes, such as DNA replication, transcription, and recombination. Genome is packaged under the control of packaging code in the cell, which indirectly ensures that the packaged chromatin has to be accessible for DNA-involved processes at some regions. Obviously, the current way of the genome packaging as well as the packaging code is the product of the natural selection. The mechanism of the packaging has been an intensively studied issue in the past thirty years<sup>1–5</sup>, but still remains elusive in many aspects, even at the level of nucleosome positioning. Nucleosome, the repeating unit of chromatin, is the first-step product of the packaging and act as a fine-scale regulator of various molecular processes. For example, nucleosome depletion at upstream region of a gene is known to facilitate transcription by increasing chromatin accessibility for transcription factor binding<sup>2</sup>. Numerous factors have been shown to affect nucleosome positioning. Intrinsic DNA properties<sup>6–10</sup> known to affect nucleosome positioning include 10-bp periodical occurrence of dinucleotides (AA/TT/TA/AT or CC/GG/CG/GC), nucleosome positioning motifs, poly(A) tract, etc. Some non-DNA factors<sup>11–21</sup> can also affect nucleosome positions. For example, the chromatin remodeling complexes can displace a nucleosome in a ATP-dependent manner<sup>18–21</sup>, while chemical modifications to histones in a nucleosome may recruit other molecules to alter chromatin structure indirectly<sup>17</sup>. The binding competency of genomic regions with histones and other proteins, such as transcription factors and RNA polymerases, can also affect nucleosome formation<sup>13,14</sup>. Besides, nucleosome organization is subjected to both the concentration of histones in a cell and the kinetics of nucleosome formation<sup>4,22,23</sup>. What we focus on in this study is physical aspects of the sequence preference. Most of the positioning signals encoded in DNA sequence can be explained from the perspective of physics. For example, the 10-bp periodicity and positioning motifs in nucleosomal DNA facilitate the wrapping of the sequence around a histone octamer by increasing sequence curvature<sup>5,24</sup>. Among the existing models<sup>25–46</sup> designed to predict nucleosome positions, physical models<sup>32–46</sup> are preferable in terms of

<sup>1</sup>School of Life Science and Technology, Inner Mongolia University of Science and Technology, Baotou, 014010, China. <sup>2</sup>Inner Mongolia Key Laboratory for Functional Genome Bioinformatics, Inner Mongolia University of Science and Technology, Baotou, 014010, China. Correspondence and requests for materials should be addressed to G.L. (email: [gqliu1010@163.com](mailto:gqliu1010@163.com))

providing a much deeper and intuitive insight into the nucleosome positioning mechanism. For example, a physical model in which both DNA elastic energy and histone-DNA interaction term used to penalize the deviation of nucleosomal DNA from ideal superhelix were considered successfully predicted *in vitro* nucleosome positions and free energies<sup>40</sup>. In another biophysical model, in addition to sequence-dependent harmonic energy, interactions between histones and nucleosomal DNA at their binding sites were considered<sup>41</sup>. Using the model, the authors predicted positions of several nucleosomes with a high accuracy and investigated the nucleosome sliding supposed to take place through twist defect diffusion. In our previous study, we presented a deformation energy model and successfully applied it to the prediction of nucleosome dyad positions and occupancy<sup>46</sup>. Particularly, bending energy performed very well in the dyad position prediction<sup>46</sup>. Although a number of energetics models were proposed to predict nucleosome-forming ability of DNA sequences<sup>32–46</sup>, to our best knowledge, none of them attempted to discuss the link between DNA deformation energy profile and chromatin remodeling. In this study, we characterized distinctly-organized chromatin structure in terms of DNA bending energy, and explored bending energy-related properties in nucleosome sliding. We also predicted nucleosome formation free energy using our bending energy model. Our results highlight the roles of bending energy in rotational positioning of nucleosomes, nucleosome reconstitution *in vitro* and nucleosome sliding.

## Materials and Methods

**Materials.** A well-established nucleosome positioning sequence, 601 sequence, was taken from van der Heijden *et al.*<sup>31</sup>. A 601 sequence-based DNA segment used in a nucleosome sliding experiment was taken from Blosser *et al.*<sup>47</sup>. A unique nucleosome map with base-pair resolution for the yeast genome (sacCer2 version) was taken from Brogaard *et al.*<sup>48</sup>. From the map, we selected top 5000 strong nucleosomes (ratio > 3.604), 5000 medium nucleosomes (1.551 < ratio < 1.793) and bottom 5000 weak nucleosomes (ratio < 0.553) according to NCP/noise ratio. The 5000 medium nucleosomes have ratio values around the mean of all ratio values. NCP here represents nucleosome center positioning score, which measures the nucleosome positioning signal strength and the noise represents the level of background non-specific cleavage of DNA backbone<sup>48</sup>. The larger the ratio is, the higher the strength of placing a nucleosome center at the position. Genomic coordinates of well-positioned nucleosomes, fuzzy nucleosomes and linker regions (S288C) were taken from Lee *et al.*<sup>49</sup>. The genomes of budding yeast (sacCer2 and S288C) were downloaded respectively from UCSC (<http://genome.ucsc.edu/>) and Saccharomyces Genome Database (<https://www.yeastgenome.org/>). The sequences used in nucleosome re-constitution experiments *in vitro* and corresponding free energy data were taken from references<sup>50–52</sup>.

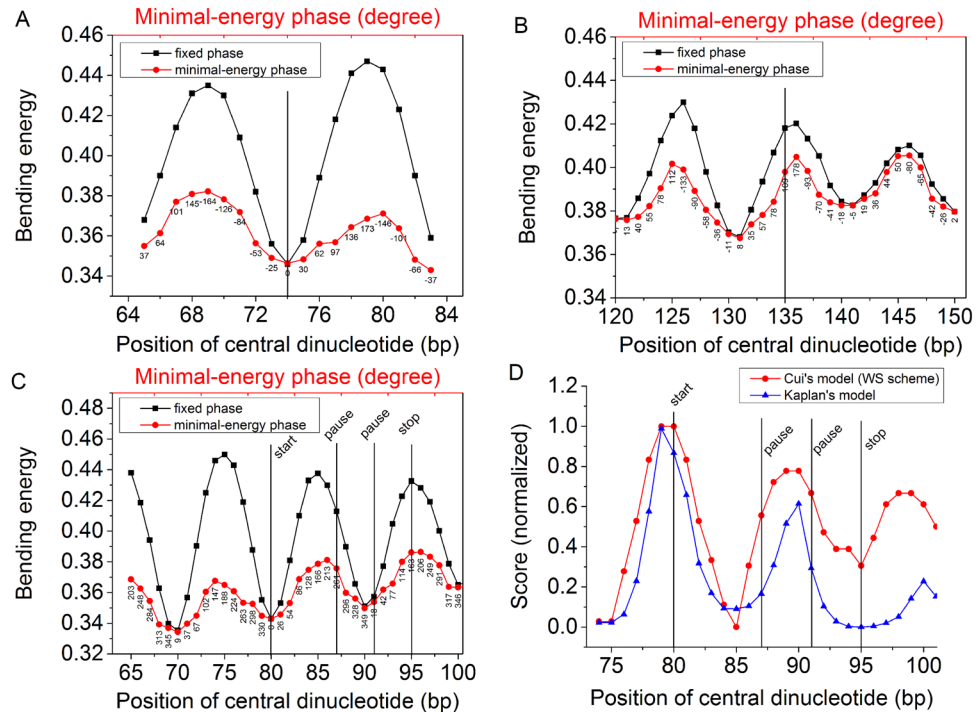
Transcription start sites (TSS) for protein-coding genes in budding yeast were obtained from high-resolution transcription map<sup>53</sup>. We selected 4,197 poly(A) RNA hybridization-based transcription segments, which overlap >50% of protein-coding regions of experimentally verified genes located on the same strand. The end sites of the transcription segments, which were located at the 5' side of the genes, were defined as TSS. Note that in order to analyze the nucleosome map<sup>48</sup> (sacCer2-based) at the aforementioned TSS (sacCer1-based), the genomic coordinates for the 4,197 transcription segments were converted from sacCer1 to sacCer2 version using LiftOver at UCSC (<http://genome.ucsc.edu/>). Numerous studies<sup>4,12,48</sup> shows that most NFR in yeast are roughly located at the position –175 to –25 relative to TSS. Accordingly, the nucleosome-free region (NFR) at 5' end of protein-coding genes in this study were defined as the region located from –175 to –25, relative to the TSS. The –1 nucleosomes and +1 nucleosomes surrounding the NFR were defined as those whose central genomic positions are located, respectively, from –325 to –175 and –25 to 125. We also analyzed precisely identified –1/+1 nucleosomes, whose genomic coordinates (sacCer3-based) and corresponding version of the genome (sacCer3 version) were derived, respectively, from Table S1 of the literature<sup>54</sup> and UCSC.

**Bending energy calculation.** A nucleosome consists of a histone octamer and a DNA segment of 147 bp that is sharply bent and tightly wrapped ~1.7 times around the histone octamer in a left-handed superhelix. Two 9-bp ends of nucleosomal DNA have little contribution to its curvature<sup>24,46</sup> and hence a 129-bp window is used in bending energy calculation. The bending energy is formulated as

$$E_b(i) = \sum_{i=1}^{128} \left[ \frac{1}{2} k_\rho(i) [\rho(i) - \rho_0(i)]^2 + \frac{1}{2} k_\tau(i) [\tau(i) - \tau_0(i)]^2 \right] \quad (1)$$

where  $\rho(i)$  and  $\tau(i)$  are, respectively, the predicted roll and tilt angles at dinucleotide step  $i$  in a 129-bp DNA segment assumed to be subject to a constraint of curvature of  $579^\circ$ , which is the same as that for the central 129-bp part of the ideal superhelix that best fits the core DNA in the nucleosome core particle<sup>24</sup>;  $\rho_0(i)$  and  $\tau_0(i)$  are equilibrium values of roll and tilt respectively for the dinucleotide at step  $i$ ;  $k_\rho(i)$  and  $k_\tau(i)$  are the dinucleotide-dependent force constants. Here, roll and tilt are two of the six degrees of freedom used to describe the geometry of DNA double helix according to Cambridge Convention<sup>55</sup>. The sequence-dependent force constants and equilibrium values of roll, tilt and twist used in this study were all taken from Liu *et al.*<sup>46</sup>, which were estimated by using the structures of protein-DNA complexes.

As previously done<sup>46</sup>,  $\rho(i)$  and  $\tau(i)$  in Eq. 1 were estimated by using the structure constraint (Eq. 2) derived from a nucleosome crystal structure and the relation (Eq. 3) between deformations in roll and tilt and bending force supposed to be uniformly distributed along the DNA sequence analyzed. Cumulative helical twist at step  $i$ ,  $\Omega_i$ , was calculated by adding up the equilibrium twists of the dinucleotide steps counted from the central base-pair of the sequence.



**Figure 1.** Bending energy profiles for sequences. The units of bending energies reported throughout the results are  $k_b T/\text{bps}$ . **(A)** Bending energies for 601 sequence were calculated using “fixed model” and “minimal-energy model”. The phases of the central dinucleotide of 129-bp window corresponding to minimal energies in “minimal-energy” model was indicated under the plot symbols. **(B)** Bending energies for a 5S oocyte sequence. **(C)** Bending energies for a 601 sequence-based DNA segment, which was used in a nucleosome sliding experiment *in vitro*<sup>47</sup>. The “start”, “pause” and “stop” in the figure represent the positions at which the nucleosome was initially positioned, transiently paused, and finally reached, respectively. **(D)** For the 601 sequence-based DNA segment, score of a nucleosome center being at the positions along the DNA sequence were calculated using Cui’s model<sup>55</sup> and Kaplan’s model<sup>4</sup>. The original scores were normalized to the range of 0–1 for the convenience of display.

$$579^\circ = \sum_{i=1}^{128} [\rho(i) \cos \Omega_i + \tau(i) \sin \Omega_i] \quad (2)$$

$$\begin{cases} \rho(i) - \rho_0(i) = F_b \cos \Omega_i / k_\rho(i) \\ \tau(i) - \tau_0(i) = F_b \sin \Omega_i / k_\tau(i) \end{cases} \quad (3)$$

Bending energy of any sequence segment, 129 bp in size, is computed by combining the aforementioned three equations. The unit of force constants used in our model is  $k_b T/\text{degree}^2$ , and hence the unit of bending energy calculated with Eq. 1 is  $k_b T$ , where  $k_b$  is Boltzmann constant and  $T$  is effective temperature. After dividing the bending energy by 128, the number of base-pair steps of the sequence segment, we obtain average bending energy per base-pair step, whose unit is  $k_b T/\text{bps}$  where bps denotes base-pair step.

Bending energy is inversely correlated with nucleosome-forming ability of a DNA segment. In other words, the lower the bending energy of a sequence is, the easier the sequence is to form a nucleosome. The equation for bending energy calculation where the central dinucleotide of analyzed DNA segment was assigned a fixed phase of zero was designed to estimate the bending energy of the DNA segment which bends toward the major groove at its central dinucleotide (see Liu *et al.*<sup>46</sup> for details). To assess the maximal bendability that a DNA segment can have, a minor modification was made: we first calculated bending energies corresponding to every bending direction toward which the central dinucleotide bends, by altering the phase of the central dinucleotide from  $-180$  degrees to  $180$  degrees, and then picked out the minimal energy to assess the maximal bendability. In a word, the present study is mainly an application of our previous model and only one methodological extension of the model is the measurement of minimal bending energy for a DNA sequence by selecting the smallest one from computed bending energies that correspond to different directions in which the DNA bends. In this study, the previous bending energy model is denoted as “fixed phase” model, and the modified one as “minimal energy phase” model. Throughout the manuscript, “fixed phase” model was used unless stated. To be specific, “minimal energy phase” model was used only in Fig. 1. The bending energy profile difference between the two models was shown taking “601” sequence as an example (Fig. 1A). Although minor difference exists, both models indicated a local energy minimum at its central position, at which the “601” sequence bends toward its major groove.

In our model, twist deformation and DNA stretching that play some roles in nucleosome positioning are not considered. The twist term, however, is coupled with other degrees of freedom used to describe DNA structure, particularly with the roll as illustrated by crystal structure of NCP147. There is a strong correlation between roll and twist<sup>32</sup>. So, the inclusion of roll, which is a major contributor to DNA bending, in our model automatically takes at least a significant part of the twist effect into account. Furthermore, in the calculation of bending energy, we used dinucleotide-dependent cumulative twists that indicate proper phasing of dinucleotides relative to the dyad position of a nucleosome, and the use of the dinucleotide-dependent twists instead of a constant twist for all dinucleotide steps represents the inclusion of twist effect to some extent in our model. In addition, what we focus on in the study is the effect of DNA bending on rotational positioning and sliding of nucleosomes, in which the stretching and shearing of DNA sequence are not as important as bending.

**Distribution of 10-bp periodicity, fluctuation amplitude and minimal value of bending energies along genomic sequences.** We first obtained the series of bending energies for each of the genomic regions for well-positioned nucleosomes, fuzzy nucleosomes and linkers using “fixed phase model”, and then carried out following analyses.

Fast Fourier Transform algorithm was employed to detect periodical signals in each series of bending energies<sup>46</sup> using a sliding window of 50. For  $N$ -sized numerical data, its Fourier transform  $F(k)$  is given by

$$F(k) = \sum_{m=0}^{N-1} E(m)e^{-2\pi mik/N}, \quad k = 0, 1, \dots, N - 1 \quad (4)$$

where  $N$  is a positive even number. Periodicity is denoted as  $f = N/k$ . Practically, it is enough to analyze the Fourier transform with  $k$  in the range  $[0, N/2]$ , as the amplitude of the Fourier transform,  $|F(k)|$ , is symmetric with respect to  $k = N/2$ . To capture an amplitude corresponding to  $\sim 10$ -bp periodicity, we selected the maximal value of the amplitudes in the periodicity range between 9.5 and 11.5 to represent the intensity of  $\sim 10$ -bp periodicity.

The distribution of fluctuation amplitudes of bending energies along the genomic sequence was obtained by consecutively computing the absolute of the difference between each pair of maximum and minimum within a sliding window of 12 along the series of bending energies. The window size, 12, was selected to capture the local maximum and minimum within each oscillating period ( $\sim 10$  bp) of the calculated bending energies, which corresponds to the bending anisotropy of DNA. The distribution of energy minima was obtained similarly using a sliding window of 12 along the series of bending energies by selecting minimum within the window. Average distribution profiles of aforementioned three indexes for each kind of genomic sequences, such as well-positioned nucleosomes, fuzzy nucleosomes and linkers, were obtained respectively by averaging over the same kind of genomic sequences.

**Definition of canonical positioning/alternative positioning.** Based on the bending energies derived from the fixed-phase model, we defined two nucleosome positioning modes: canonical positioning and alternative positioning. A nucleosome is considered to adopt canonical positioning mode if its nucleosomal DNA sequence has a local energy minimum for its central subsequence of 129 bp, and if no local energy minimum is detected at the central position, it is called alternative positioning. A maximal uncertainty of 2 bp is allowed for the “central position”. Besides, in order to exclude biologically meaningless local energy minima which are the result of random fluctuation of bending energies, the “local energy minimum” here is defined as the lowest and central value among five consecutive bending energies, which monotonically decrease from its both sides.

**Spearman rank correlation.** Spearman rank correlation analysis was performed using SPSS 11.5 to test the correlation between two  $n$ -sized variables, and correlation coefficients ( $R$ ) and  $P$  values were reported. Whether the observed Spearman rank correlation coefficient is significantly different from zero was tested using  $t = R \sqrt{\frac{n-2}{1-R^2}}$ , which is distributed approximately as Student’s  $t$  distribution with  $n - 2$  degrees of freedom under the null hypothesis stating that there is no relationship between the two variables.

## Results and Discussion

**Prediction of nucleosome dyad positions.** In our previous study<sup>46</sup>, we successfully predicted dyad positions for 19 out of 20 nucleosomes that were re-constituted *in vitro*. To test if the model established on minimal energy phase frame could give any implication on the positioning mechanism of one wrongly predicted nucleosome whose dyad position is at 135 bp on 5S oocyte sequence, we presented the bending energy profile for the nucleosome. Unfortunately, no expected local energy minimum was detected at the position 135 (Fig. 1B).

**Bending energy profile in nucleosome sliding.** What we are interested in this study is “can we infer some information from bending energy profile of sequence if nucleosome sliding is influenced by DNA sequence?”. Motivated by this question, we analyzed bending energy profile for a 601 sequence-based DNA segment, which was used in a nucleosome sliding experiment *in vitro*<sup>47</sup>. A nucleosome was assembled on the sequence, and then remodeled under the ATP-dependent action of a remodeler protein, ACF. The nucleosome was initially positioned at a position denoted as “start” in Fig. 1C, and then moved and transiently paused twice before it reached its final state (denoted as “stop” in Fig. 1C) on the sequence. We expect possible bending energy minima at the positions including start, pause and stop positions, which indicate relatively stable states of binding between histone octamer and underlying sequence. We really observed local energy minima at the initially positioned position and one intermediately paused position as expected. However, the bending energies at the other intermediately paused position and finally reached position do not tend to adopt low bending energy, but instead show relatively high local energies. In addition, we tried two other state of the art models<sup>6,56</sup> for predicting

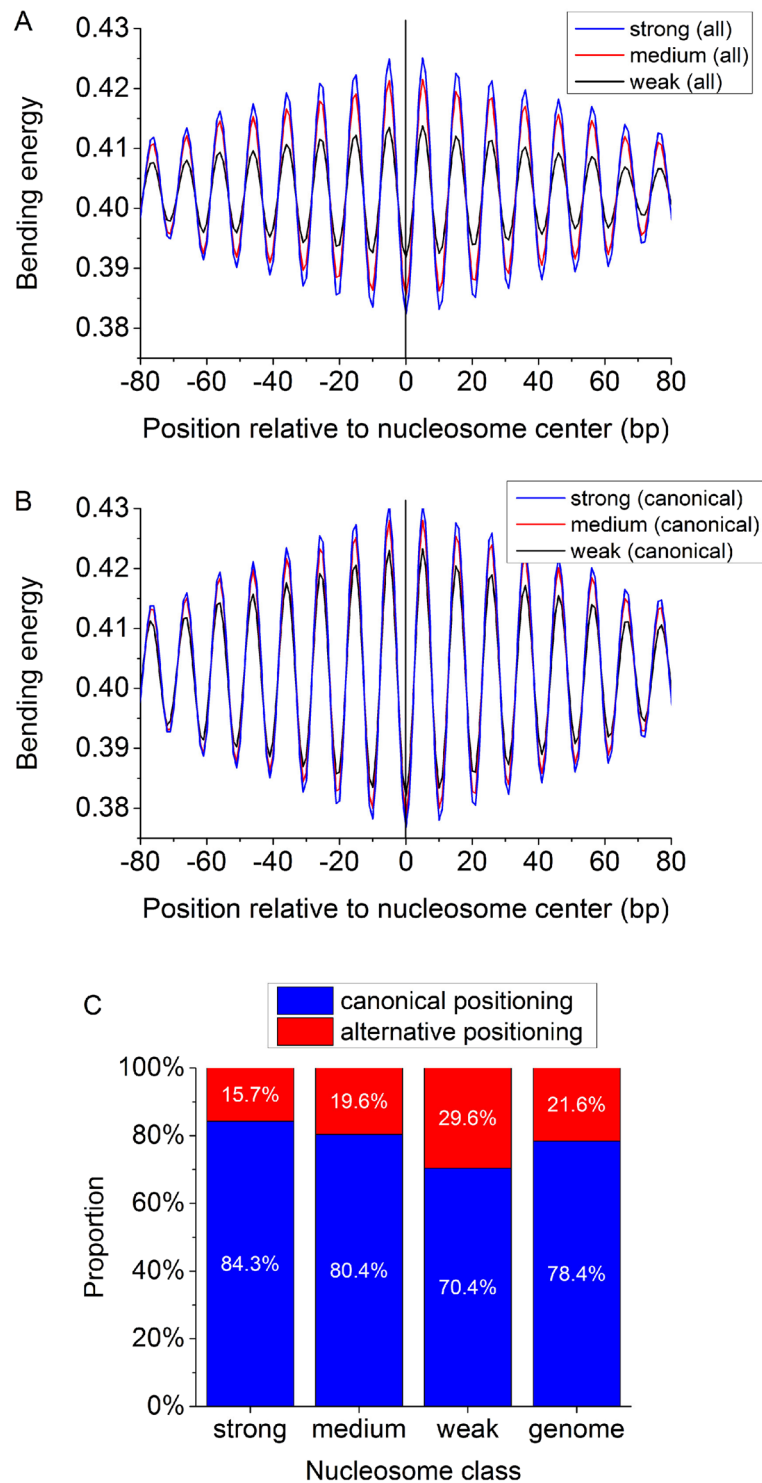


nucleosome center positions and found that the models also failed to predict a peak of positioning score at the “stop” position (Fig. 1D). These results implicate that nucleosomes, particularly *in vivo*, do not necessarily have a representative feature of minimal energy. One possible explanation for this is the energy barriers associated with the DNA-histone interaction, which might occur in the process of nucleosome sliding and block the access of the nucleosome to a minimal energy state. Moreover, for *in vivo* system, some other factors such as the dynamics of nucleosomes may also account for the observed results. Nucleosome maps obtained in different studies show a certain extent of discrepancy, which can be caused by several kinds of noise, such as different experimental conditions, sequencing data processing, cell cycle phase, gene transcription rate, nucleosome dynamics and sample variations produced by differences in the growth media<sup>57</sup>. Flores *et al.*<sup>57</sup> proposed that nucleosomes in most cases move along the one-dimensional DNA fiber and are mainly positioned at specific places in response to strong nucleosome depletion signals, such as intrinsic properties of DNA, the competition of histones with DNA-binding proteins and chromatin-remodelers. This means nucleosome is a highly dynamic structure in the cell and the noise mentioned above is likely to impede at least a minor proportion of nucleosomes from adopting minimal energy states. For example, if a transcription factor bound to genomic DNA impedes the sliding of a nucleosome before it reaches the state having local energy minimum, the nucleosome may position for a long or short time at a place where there is no local energy minimum. We therefore do not believe that nucleosomes in such a dynamic *in vivo* system are all positioned at genomic sites having local minima of bending energy.

Given that not all nucleosomes can be characterized by a local bending energy minimum, it is interesting to see what proportion of a genome is occupied by such non-canonical nucleosomes without local energy minimum. It is an important question because the extent of the prevalence of this non-canonical positioning pattern can not only help us understand the chromatin structure organization and structure-mediated processes, but also directly affect the accuracy of various sequence-dependent models designed to predict nucleosome positions. It is generally accepted that strong nucleosomal DNA sequences have strong periodical signals<sup>6,11</sup> and the periodicity encoded in DNA sequences for nucleosome positioning has a close relationship with its bending energy. Dividing 67,543 nucleosomes from the base-pair resolution unique map for yeast<sup>48</sup> into three categories according to nucleosome center positioning score/noise ratio, we found that as the positioning signal in sequences becomes stronger, the amplitude of the bending energy profile becomes greater (Fig. 2A). There are three possible reasons for the reduced amplitude for the weak nucleosomes. Firstly, the amplitudes of bending energy profiles for individual sequences tend to decrease as the positioning signals in sequences become weaker. Secondly, the bending energy of weak nucleosomal sequences does not show as regular appearance of 10–11 bp periodical oscillation as strong nucleosomes. Thirdly, weak nucleosomes do not have consistently low bending energy at their aligned sites (nucleosome centers), which can result in the averaged weak amplitude profile for weak nucleosomes even if they have equivalent extend of 10-bp periodicity as strong nucleosomes. We tested if the three categories of nucleosomes differ in the proportion of local minimum of bending energy at the aligned sites. Here, a nucleosome positioning mode is called “canonical positioning” if corresponding nucleosomal sequence has a local energy minimum at its center, otherwise called “alternative positioning”. As shown in Fig. 2C, as nucleosome center positioning signal weakens, the proportion of alternative positioning nucleosomes increases from 15.7% to 29.6%, indicating the appearance of local energy minimum at nucleosome centers is not an exclusive unique mode of nucleosome positioning. In yeast, at genome-wide level, about 21% nucleosomes adopt such kind of alternative positioning. Accordingly, we propose that people should be cautious in using a framework which can only predict canonical positioning of nucleosomes and explaining the corresponding results. Moreover, we carried out a similar analysis excluding the alternatively positioned nucleosomes, and reproduced a similar pattern (Fig. 2B) as in Fig. 2A, suggesting that: (1) the difference in bending energy between three categories of nucleosomes cannot be ascribed solely to the alternatively positioned nucleosomes; (2) bending energy is an indicator of nucleosome positioning ability of DNA sequence even when the effect of alternative positioning was removed.

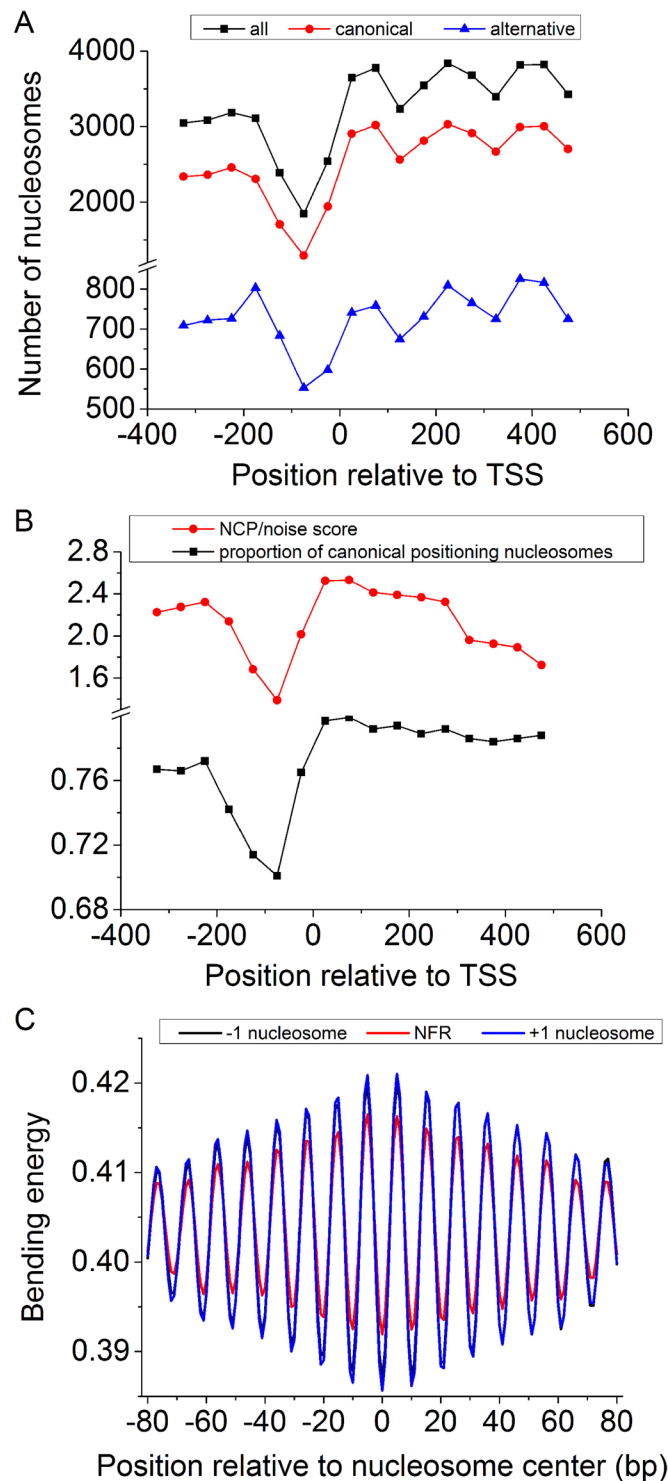
In order to see to what extent +1 nucleosomes positioned at the 5' end of genes are dictated by the bending energy, we analyzed the distribution of nucleosomes around transcription start sites and corresponding DNA bending energies. Our results show that both canonical and alternative nucleosomes are depleted at the upstream of the TSS (from position -125 to -25 in Fig. 3A), and a much higher proportion of +1 and -1 nucleosomes tend to be canonically positioned as compared to the nucleosomes found in the nucleosome-depleted region (Fig. 3B). Furthermore, as compared with the nucleosome free region (NFR), +1 and -1 nucleosomes can be characterized with a lower level of bending energy minima and higher bending anisotropy indicated by the higher amplitude in the bending energy oscillation (Fig. 3C). We also show that precisely identified +1/-1 nucleosomes have slightly enhanced 10-bp oscillation in bending energy than the +1/-1 nucleosomes defined in this study (supplementary Fig. S1).

We are also interested in characterizing linker regions and well-positioned and fuzzy nucleosomes in terms of bending energy. Because 10-bp periodical oscillation of bending energy reflects directly the bending anisotropy of a DNA sequence, which has a close relationship with the rotational positioning of the nucleosome, we analyzed the ~10 periodicity in the bending energies calculated along the sequences by using Fourier transform. As expected, the bending energy of well-positioned nucleosomes exhibits much stronger ~10-bp periodicity than fuzzy nucleosomes and linker sequences (Fig. 4A). The two sides encompassing the linker regions have much elevated ~10-bp periodicity in bending energy (Fig. 4A), which can be ascribed to nucleosome positioning sequences flanking the linker sequences. Consistently, the amplitude of bending energy for well-positioned nucleosomes is the greatest and that for linkers is the smallest (Fig. 4B). It is worth noting that strong periodicity of bending energy does not necessarily mean a low bending energy. As depicted in Fig. 4C, we found an interesting result that the averaged profile of the local bending energy minima is the lowest for fuzzy nucleosomes. According to the aforementioned results, we presented bending energy profiles for diverse nucleosome positioning patterns in Fig. 5B, which differs largely from previously proposed one (Fig. 5A) in the context of fuzzy nucleosomes.

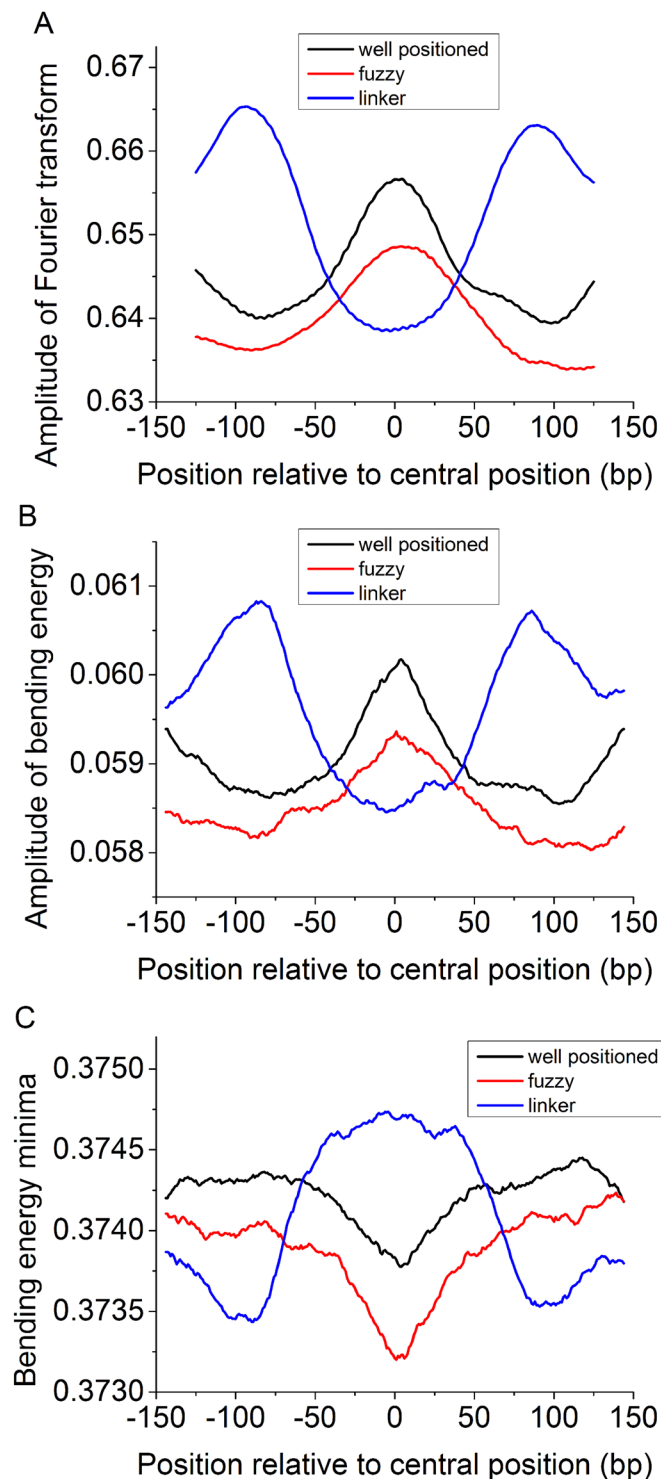


**Figure 2.** (A) Average bending energy profile for nucleosome positioning sequences in the nucleosome map of Brogaard *et al.*<sup>48</sup>. The nucleosome positioning sequences were divided into three categories according to the intensity of nucleosome positioning signal (see “Materials” for details). (B) The same as (A), but alternative positioning nucleosomes were not included in the analysis. (C) The proportion of alternative positioning nucleosomes increases as the nucleosome center positioning signal weakens from strong, medium to weak. “canonical positioning” represents a nucleosome positioning mode in which the nucleosomal sequence has a local bending energy minimum at its center. In contrast, “alternative positioning” does not have a local bending energy minimum at the center of the sequence.

Is it possible to give a further insight into chromatin remodeling from the perspective of sequence bending energy? As previously found, it is difficult for a nucleosome positioned at a genomic region with low bending energy to slide towards regions with high bending energy, particularly when the nucleosome is rotationally locked



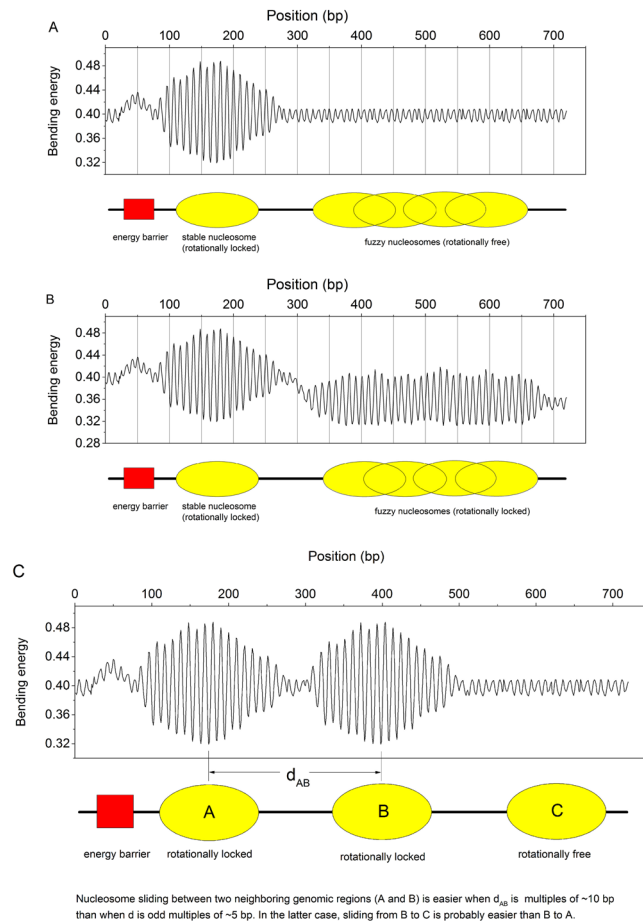
**Figure 3.** Distribution of nucleosomes around transcription start sites (TSS). The unique nucleosome map<sup>48</sup> and a total of 4,197 TSS for protein-coding genes<sup>53</sup> were analyzed here (see Materials for details). **(A)** Both canonical and alternative nucleosomes show similar distribution patterns around TSS. They are depleted at the upstream of the TSS (from position  $-125$  to  $-25$ ), and enriched in gene body regions. A window of 150 bp with a moving step of 50 bp was used in nucleosome counting. If the center of a nucleosome is located in the predefined window, it is considered to be one hit for that window. **(B)** The proportion of canonical nucleosomes in the predefined window varies around the TSS. The proportion is low for the nucleosome-depleted region and high for surrounding regions where  $+1$  and  $-1$  nucleosomes with high nucleosome center positioning scores (NCP/noise) are located. **(C)** DNA bending energy difference between the nucleosome free regions (NFR, from position  $-175$  to  $-25$ ),  $-1$  nucleosomes (from  $-325$  to  $-175$ ), and  $+1$  nucleosomes (from  $-25$  to  $125$ ).



**Figure 4.** (A) Distribution of  $\sim 10$ -bp periodicity amplitudes of the Fourier transform of bending energies for distinctly organized nucleosome regions. (B) Distribution of fluctuation amplitudes of bending energies for distinctly organized nucleosome regions. (C) Distribution of bending energy minimum for distinctly organized nucleosome regions. See “Materials and Methods” for methodology adopted here.

(Fig. 5C). Besides, given that the rotational setting of the underlying sequence with respect to histone surface can hardly change in the course of nucleosome sliding<sup>18</sup>, we propose here that there is a simple bending energy-based rule directing nucleosome sliding (Fig. 5C): a nucleosome is easy to slide between two neighboring genomic positions if they have the same rotational setting, otherwise, nucleosome sliding is difficult. If a nucleosome slides between two positions with opposite rotational settings, the underlying DNA sequence has to rotate 180 degrees relative to the double-helix axis to accommodate its rotational setting to histone surface, which can

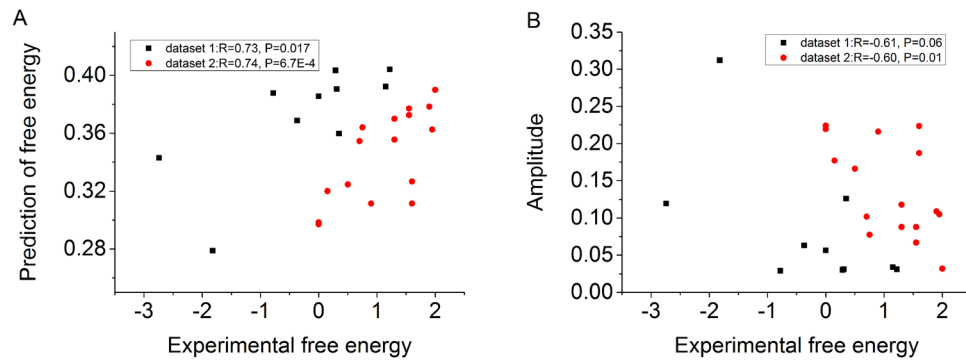




**Figure 5.** Cartoons of bending energy profiles for diverse nucleosome positioning patterns. The cartoons were drawn only for illustration of nucleosome positioning patterns for different genomic regions and bending energy-dependent model for nucleosome sliding, and therefore the bending energies in the Figure do not represent stringent quantitative predictions of the deformation energy model. (A) Pattern of deformation energy profiles (adapted from the reference<sup>18</sup>) for different genomic regions, such as energy barrier regions, well-positioned nucleosomes and fuzzy nucleosomes. (B) Pattern of bending energy profiles proposed in this study. (C) A bending energy-based model for nucleosome sliding. Nucleosome sliding between two neighboring positioning sites (A,B) is favored when they have the same rotational setting (their distance is multiples of  $\sim 10$  bp). In contrast, if they have opposite rotational setting (their distance is odd multiples of  $\sim 5$  bp) sliding between them is disfavored, and in this case, sliding from (B) to a rotationally free site (C) is probably easier than sliding to a rotationally opposite site (A). In Fig. 5C, the yellow ovals represent three genomic sites, which are likely to accommodate nucleosomes. According to the bending energy profile and our discussion, we know that well-positioned nucleosomes are favored at site (A and B), while site (C) neither represents a well-positioning region, nor a fuzzy nucleosome region. Although the site (C) is less favorable for positioning a nucleosome than site (A and B), it is also represented by an oval to illustrate our nucleosome sliding model.

apparently introduce a large energy barrier for nucleosome sliding. In other words, apart from remodeler activity, bending energy profile of DNA sequence may also play a important role in the accessibility and directionality of nucleosome sliding. Note that the model of sequence-dependent nucleosome sliding proposed above is for ATP-dependent sliding of nucleosomes, in which rotational setting of DNA on the histone core remains unaltered as described in loop/bugle propagation model<sup>58</sup>. Among several models proposed for nucleosome sliding<sup>58</sup>, twist defect diffusion model and loop/bugle propagation model are of great interest. Both of the models are supported by some experimental evidence, but a major difference between the models is: the diffusion model for nucleosome sliding along DNA in response to thermal fluctuations predicts an alteration of the rotational phasing of the DNA on the histone core, while the loop propagation model for ATP-dependent sliding of nucleosomes allows DNA to maintain the rotational phasing<sup>58</sup>.

**Nucleosome free energy prediction.** In our previous study, our model performs well in predicting nucleosome dyad positions and occupancy<sup>46</sup>. The model also successfully discriminated nucleosome-enriched regions from nucleosome-depleted regions. However, it is unclear how the model performs in nucleosome free energy prediction. The free energy refers to the change of free energy of the experimental system used to re-constitute



**Figure 6.** Bending energy is a good indicator of free energy in nucleosome re-constitution experiments *in vitro*. Dataset 1 is a sequence set from reference<sup>50</sup>, and Dataset 2 is a sequence collection from references<sup>51,52</sup>. (A) The mean of local minima of bending energies has a significant rank correlation with nucleosome free energies. Spearman correlations coefficients and corresponding *P*-values were reported in the figure (see Method section for the significance test). (B) The amplitude of bending energy profile also has a significant rank correlation with nucleosome free energies.

nucleosomes *in vitro*. Generally, DNA sequences with low bending energy tend to form nucleosomes and have low free energy in the process of nucleosome assembly *in vitro*. Given the majority of the nucleosomes tend to adopt canonical positioning, the rotational positioning of which can be more clearly captured by our fixed-phase model than minimal-phase model, we developed a free energy prediction model based on the fixed-phase model. A local minimum of calculated bending energy profile represent the ability of the region centering at the minimum to form a nucleosome. In our collected free-energy data, the precise positions of the re-constituted nucleosomes are unknown, raising a question that which bending energy data is suitable for predicting free energy. Because the bending energy minima are all possible positioning sites in nucleosome assembly experiments, we decided to rank nucleosome free energy by the mean of local minima of bending energies. This approach successfully predicted the ranks of the nucleosome free energies (Fig. 6A). We also show that the amplitude of the bending energy profile is significantly anti-correlated with free energy (Fig. 6B), indicating that bending anisotropy of DNA segment can significantly affect nucleosome formation.

## Conclusion

To conclude, we investigated the DNA-encoded signals in chromatin remodeling, characterized distinctly-organized chromatin structure and predicted nucleosome free energies using a DNA bending energy model. Our results show that well-positioned nucleosomes, fuzzy nucleosomes and linker regions have distinct bending energy profiles; approximately, 21% of the nucleosomes in yeast adopted an alternative positioning mode that is different from canonical positioning; DNA bending energy alone can serve as a good indicator of nucleosome forming ability. Our data suggest that the stability of a nucleosome is positively correlated with the strength of the bending anisotropy of DNA segment, and both accessibility and directionality of nucleosome sliding are likely to be modulated by diverse patterns of DNA bending energy profile.

## References

- Kornberg, R. D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285–294 (1999).
- Lee, C. K., Shibata, Y., Rao, B., Strahl, B. D. & Lieb, J. D. Evidence for nucleosome depletion at active regulatory regions genome-wide. *Nat. Genet.* **36**, 900–905 (2004).
- Yuan, G. C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).
- Kaplan, N. *et al.* The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* **458**, 362–366 (2009).
- Trifonov, E. N. Cracking the chromatin code: precise rule of nucleosome positioning. *Phys Life Rev.* **8**(1), 39–50 (2011).
- Segal, E. *et al.* A genomic code for nucleosome positioning. *Nature* **442**, 772–778 (2006).
- Rapoport, A. E., Frenkel, Z. M. & Trifonov, E. N. Nucleosome positioning pattern derived from oligonucleotide compositions of genomic sequences. *J. Biomol. Struct. Dyn.* **28**, 567–574 (2011).
- Satchwell, S. C., Drew, H. R. & Travers, A. A. Sequence periodicities in chicken nucleosome core DNA. *J. Mol. Biol.* **191**, 659–675 (1986).
- Segal, E. & Widom, J. Poly(dA:dT) tracts: major determinants of nucleosome organization. *Curr. Opin. Struct. Biol.* **19**, 65–71 (2009).
- Bao, Y., White, C. L. & Luger, K. Nucleosome core particles containing a poly(dA:dT) sequence element exhibit a locally distorted DNA structure. *J. Mol. Biol.* **361**, 617–624 (2006).
- Struhl, K. & Segal, E. Determinants of nucleosome positioning. *Nat Struct Mol Biol.* **20**(3), 267–273 (2013).
- Mavrich, T. N. *et al.* A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* **18**, 1073–1083 (2008).
- Zhang, Y. *et al.* Intrinsic histone-DNA interactions are not the major determinant of nucleosome positions *in vivo*. *Nat. Struct. Mol. Biol.* **16**, 847–852 (2009).
- Schones, D. E. *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898 (2008).
- Vaillant, C. *et al.* A novel strategy of transcription regulation by intra-genic nucleosome ordering. *Genome Res.* **20**, 59–67 (2010).
- Zhang, Y., Shin, H., Song, J. S., Lei, Y. & Liu, X. S. Identifying positioned nucleosomes with epigenetic marks in human from ChIP-Seq. *BMC Genomics* **9**, 537 (2008).
- Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693–705 (2007).

18. Partensky, P. D. & Narlikar, G. J. Chromatin remodelers act globally, sequence positions nucleosomes locally. *J. Mol. Biol.* **391**, 12–25 (2009).
19. Narlikar, G. J., Sundaramoorthy, R. & Owen-Hughes, T. Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* **154**, 490–503 (2013).
20. Racki, L. R. *et al.* The chromatin remodeler ACF acts as a dimeric motor to space nucleosomes. *Nature* **462**, 1016–1021 (2009).
21. Yen, K., Vinayachandran, V., Batta, K., Koerber, R. T. & Pugh, B. F. Genome-wide nucleosome specificity and directionality of chromatin remodelers. *Cell* **149**, 1461–1473 (2012).
22. Dong, F. & van Holde, K. E. Nucleosome positioning is determined by the (H3-H4)<sub>2</sub> tetramer. *Proc. Natl. Acad. Sci. USA* **88**, 10596–10600 (1991).
23. Böhm, V. *et al.* Nucleosome accessibility governed by the dimer/tetramer interface. *Nucleic Acids Res.* **39**, 3093–3102 (2011).
24. Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423**, 145–150 (2003).
25. Peckham, H. E. *et al.* Nucleosome positioning signals in genomic DNA. *Genome Res.* **17**, 1170–1177 (2007).
26. Xing, Y., Zhao, X. & Cai, L. Prediction of nucleosome occupancy in *Saccharomyces cerevisiae* using position-correlation scoring function. *Genomics* **98**, 359–366 (2011).
27. Liu, H. *et al.* A comparative evaluation on prediction methods of nucleosome positioning. *Brief. Bioinform.* **15**, 1014–1027 (2014).
28. Guo, S. H., Deng, E. Z., Xu, L. Q., Ding, H. & Lin, H. iNuc-PseKNC: a sequence-based predictor for predicting nucleosome positioning in genomes with pseudo k-tuple nucleotide composition. *Bioinformatics* **30**, 1522–1529 (2014).
29. Chen, W., Lin, H., Feng, P. M. & Ding, C. iNuc-PhysChem: A sequence-based predictor for identifying nucleosomes via physicochemical properties. *PLoS ONE* **7**, e47843 (2012).
30. Gabdank, I., Barash, D. & Trifonov, E. N. Single-base resolution nucleosome mapping on DNA sequences. *J. Biomol. Struct. Dyn.* **28**, 107–122 (2010).
31. van der Heijden, T., van Vugt, J. J., Logie, C. & van Noort, J. Sequence-based prediction of single nucleosome positioning and genome-wide nucleosome occupancy. *Proc. Natl. Acad. Sci. USA* **109**, E2514–E2522 (2012).
32. Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M. & Zhurkin, V. B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. USA* **95**, 11163–11168 (1998).
33. Anselmi, C., Bocchinfuso, G., De Santis, P., Savino, M. & Scipioni, A. Dual role of DNA intrinsic curvature and flexibility in determining nucleosome stability. *J. Mol. Biol.* **286**, 1293–1301 (1999).
34. Milani, P. *et al.* Nucleosome positioning by genomic excluding-energy barriers. *Proc. Natl. Acad. Sci. USA* **106**, 22257–22262 (2009).
35. Scipioni, A., Morosetti, S. & De Santis, P. A statistical thermodynamic approach for predicting the sequence-dependent nucleosome positioning along genomes. *Biopolymers* **91**, 1143–1153 (2009).
36. Scipioni, A. & De Santis, P. Predicting nucleosome positioning in genomes: physical and bioinformatic approaches. *Biophysical Chemistry* **155**, 53–64 (2011).
37. Sereda, Y. V. & Bishop, T. C. Evaluation of elastic rod models with long range interactions for predicting nucleosome stability. *J. Biomol. Struct. Dyn.* **27**, 867–887 (2010).
38. Miele, V., Vaillant, C., d'Aubenton-Carafa, Y., Thermes, C. & Grange, T. DNA physical properties determine nucleosome occupancy from yeast to fly. *Nucleic Acids Res.* **36**, 3746–3756 (2008).
39. Tolstorukov, M. Y., Colasanti, A. V., McCandlish, D., Olson, W. K. & Zhurkin, V. B. A novel 'Roll-and-Slide' mechanism of DNA folding in chromatin. Implications for nucleosome positioning. *J. Mol. Biol.* **371**, 725–738 (2007).
40. Morozov, A. V. *et al.* Using DNA mechanics to predict *in vitro* nucleosome positions and formation energies. *Nucleic Acids Res.* **37**, 4707–4722 (2009).
41. Fathizadeh, A., Berdy Besya, A., Reza Eftehadi, M. & Schiessel, H. Rigid-body molecular dynamics of DNA inside a nucleosome. *Eur. Phys. J. E* **36**, 21 (2013).
42. Vaillant, C., Audit, B. & Arneodo, A. Experiments confirm the influence of genome long-range correlations on nucleosome positioning. *Phys. Rev. Lett.* **99**, 218103 (2007).
43. Wang, J. Y., Wang, J. & Liu, G. Calculation of nucleosomal DNA deformation energy: its implication for nucleosome positioning. *Chromosome Res.* **20**, 889–902 (2012).
44. Liu, G., Feng, F., Zhao, X. & Cai, L. Nucleosome organization around pseudogenes in the human genome. *BioMed Res. Int.* 821596 (2015).
45. Liu, G., Cui, X., Li, H. & Cai, L. Evolutionary direction of processed pseudogenes. *Sci. China Life Sci.* **59**, 839–849 (2016).
46. Liu, G. *et al.* A deformation energy-based model for predicting nucleosome dyads and occupancy. *Sci. Rep.* **6**, 24133 (2016).
47. Blosser, T. R., Yang, J. G., Stone, M. D., Narlikar, G. J. & Zhuang, X. Dynamics of nucleosome remodelling by individual ACF complexes. *Nature* **462**, 1022–1027 (2009).
48. Brogaard, K., Xi, L., Wang, J. P. & Widom, J. A map of nucleosome positions in yeast at base-pair resolution. *Nature* **486**, 496–501 (2012).
49. Lee, W. *et al.* A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.* **39**, 1235–1244 (2007).
50. Thåström, A. *et al.* Sequence motifs and free energies of selected natural and non-natural nucleosome positioning DNA sequences. *J. Mol. Biol.* **288**, 213–229 (1999).
51. Shrader, T. E. & Crothers, D. M. Artificial nucleosome positioning sequences. *Proc. Natl. Acad. Sci. USA* **86**, 7418–7422 (1989).
52. Shrader, T. E. & Crothers, D. M. Effects of DNA sequence and histone-histone interactions on nucleosome placement. *J. Mol. Biol.* **216**, 69–84 (1990).
53. David, L. *et al.* A high-resolution map of transcription in the yeast genome. *Proc. Natl. Acad. Sci. USA* **103**, 5320–5325 (2006).
54. Chereji, R. V., Ramachandran, S., Bryson, T. D. & Henikoff, S. Precise genome-wide mapping of single nucleosomes and linkers *in vivo*. *Genome Biol.* **19**, 19 (2018).
55. Dickerson, R. E. Definitions and nomenclature of nucleic acid structure parameters. *J. Biomol. Struct. Dyn.* **6**, 627–634 (1989).
56. Cui, F., Chen, L., LoVerso, P. R. & Zhurkin, V. B. Prediction of nucleosome rotational positioning in yeast and human genomes based on sequence-dependent DNA anisotropy. *BMC Bioinformatics* **15**, 313 (2014).
57. Flores, O., Deniz, Ö., Soler-López, M. & Orozco, M. Fuzziness and noise in nucleosomal architecture. *Nucleic Acids Research* **42**, 4934–4946 (2014).
58. Bowman, G. D. Mechanisms of ATP-dependent nucleosome sliding. *Curr. Opin. Struct. Biol.* **20**, 73–81 (2010).

## Acknowledgements

We would like to thank Xiujuan Zhao and Xiangjun Cui for their helpful discussions. This work was supported by grants from the National Natural Science Foundation (31660322, 61102162, 61662055, 31760247) and Science Foundation for Excellent Youth Scholars of Inner Mongolia University of Science and Technology (2016YQL06).

## Author Contributions

G.L. designed the study, did computational analysis and wrote the manuscript. Y.X. and H.Z. participated in the data preparation and analysis, L.C. and J.W. participated in the data analysis and discussion.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-27247-x>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018