Front. Electr. Electron. Eng. China 2008, 3(4): 388–393
DOI 10.1007/s11460-008-0085-4

RESEARCH ARTICLE

Zhiming DAI, Xianhua DAI, Jihua FENG, Qian XIANG, Yangyang DENG, Jiang WANG

Insights into DNA signals for nucleosome positioning

Abstract The nucleosome is the fundamental unit of eukaryotic genomes. Its positioning in the promoter region plays a central role in regulating gene transcription. Experimental evidence suggests that the genomic DNA sequence is one important determinant of nucleosome positioning. Several approaches have been developed to predict nucleosome positions based on DNA sequence features, but the results indicate that there is room for improvement. This paper presents a new computational approach to predict genome-wide nucleosome locations in promoter regions. Importantly, the proposed approach outperforms existing approaches in yeast. Further analysis demonstrates that DNA signals for nucleosome positioning vary with species and composition of histones. Analysis of individual genes reveals that the role of the underlying DNA sequence in nucleosome positioning varies with genes.

Keywords nucleosome, DNA sequence preferences, nucleosome positioning signals

1 Introduction

Eukaryotic genomes are packaged into chromatin, the major structural element of which is nucleosome [1]. Each nucleosome consists of a core particle and linker DNA naked or associated with histone H1 [2]. Nucleosomes and additional proteins can interact to form a highly folded chromatin structure, which is the substrate for the essential biological processes of DNA replication, recombination, repair, transcription and chromosome segregation, and cell division [3]. The structure of the nucleosome core particle, composed of a 147 bp stretch of DNA tightly wrapped around a histone octamer, has been solved by X-ray crystallography at atomic resolution [4]. Genome-scale nucleosome positioning in yeast (*Saccharomyces cerevisiae*) and human cells have been identified by advanced experimental approaches [5–8].

Nucleosomes in promoter regions are considered to limit accessibility to regulatory factors and recruit other proteins for histone modifications. Nucleosome positioning along genomic DNA sequence thus plays an important role in both positive and negative gene regulation [9]. There are three main ways in which cells overcome the nucleosomal barrier to regulate gene expression. One way is through chromatin remodeling, using the energy of ATP hydrolysis to change the nucleosome positions on DNA [10]. Another way involves chemically modifying the tails of histones, such as by acetylation, methylation, sumoylation, phosphorylation, ubiquitination, and adenosine-diphosphate ribosylation [11]. The complex combinations of modifications may form a 'histone code', leading to epigenetic and heritable changes in chromatin domains and recruitment of structural proteins and enzymes to the chromatin [12]. The third consists of incorporation of histone variants (e.g. H2A.Z and H3.3) into nucleosomes, often resulting in changes in nucleosome stability and patterns of gene expression [13].

Experimental evidence indicates that certain DNA sequences have strong ability to bend and twist [14]. Consequently, DNA sequences differ greatly in their ability to wrap around histones. The binding affinities can vary by several magnitudes [15]. Furthermore, most identified positioned-nucleosomes are conserved across several human cell lines [7], suggesting that nucleosome positioning may be partially encoded in genomic DNA sequence. The genomic code may be represented by short DNA motifs with repetitive appearances at ~10 bp intervals [16–18]. Several approaches have been proposed to predict nucleosome positions based on DNA sequence features [18–21]. In particular, yeast genome-wide nucleosome positions have been predicted and their links with specific chromosome functions have also been reported [18,19]. However, it is clear from previous work that only a subset of experimentally determined nucleosome positions are accurately predicted, thus there is room for

Received June 12, 2008; accepted July 18, 2008

Zhiming DAI, Xianhua DAI (✉), Jihua FENG, Qian XIANG, Yangyang DENG, Jiang WANG
Department of Electronic Engineering, Sun Yat-sen University, Guangzhou 510275, China
E-mail: issdxh@mail.sysu.edu.cn

improvement. In addition, it is not clear whether the underlying DNA sequence plays similar roles in nucleosome positioning of different genes.

In this study, first, a strong periodic pattern in a set of nucleosomal DNA sequences was found to develop a new computational approach, which successfully predicted the yeast genome-wide, experimentally-determined nucleosome positions across promoter regions. Second, DNA signals for nucleosome positioning among species and roles of the underlying DNA sequence in nucleosome positioning of different genes were analyzed.

2 DNA sequence pattern for nucleosome packaging

Many DNA sequence features, including TGGA [22], VWG [23], CTG [24] and the AA/TT/TA dinucleotides [18], are proposed as DNA signals for nucleosome packaging. Specifically, ~10 bp periodic AA/TT/TA dinucleotides recur remarkably along the nucleosomal DNA sequences. The feature is widely used to predict nucleosome positions [19,20]. However, previous studies show that existing predictive approaches are far from perfect [19,20]. One possible explanation is that the underlying DNA sequence itself is just among the dominant determinants of nucleosome positioning, and another one is that the dinucleotide pattern may not completely reflect the DNA signals for nucleosome packaging. To enhance prediction accuracy, it is necessary to find an additional nucleosomal DNA sequence pattern.

We used a collection of mononucleosome DNA sequences on the yeast genome, measured by an accurate experimental method [18], to find a remarkable sequence pattern (Fig. 1(a)). It includes ~10-bp periodic alternating (A,T)-rich and (A,T)-poor sequences at the DNA helical repeat (hereinafter referred to as (A,T)-alternating pattern). In other words, the fractions of A/T nucleotides along nucleosomal DNA sequences exhibit recurring valley-peak patterns. These DNA sequence preferences to yeast nucleosomes may correspond to preferences to a minor groove for the (A,T)-rich sequences bending via negative base-pair roll and a major groove for the (G,C)-rich sequences bending via positive base-pair roll. To test the significance of the (A,T)-alternating pattern, we randomly sampled 129-bp sequences from yeast genome. The mean number of valley-peak patterns from 10000 random experiments fell to ~4 compared with 10 observed in yeast nucleosomal DNA, which indicates that the (A,T)-alternating pattern is a unique characteristic of nucleosomal DNA sequences. Moreover, another strong (A,T)-alternating pattern was derived from a collection of nucleosomal DNA sequences from chicken (Fig. 1(e)).

3 Predicting genome-wide nucleosome positions

As shown in Figs. 1(a) and 1(e), the significantly discriminative (A,T)-alternating pattern can be employed to predict nucleosome positions. Although a similar pattern has been found to facilitate DNA wrapping around histones [16,17], this feature has not yet been applied to the prediction of nucleosome positions. Using the (A,T)-alternating pattern, we developed a new computational approach to predict yeast genome-wide nucleosome positions in promoter regions spanning between −1200 and +200 (relative to the +1 ATG translational start codon). The predictions showed significant correspondence with the experimentally determined nucleosome locations measured in a genome-scale study [5] (Fig. 2). We also made a similar analysis of the nucleosome positioning data predicted by two popular computational approaches: one was based on the AA/TT dinucleotide pattern and the other used a thermodynamic nucleosome-DNA interaction model [18,19]. Overall, ~51% of our predicted nucleosomes were within 35 bp of those determined experimentally [5] compared with ~45% and ~39% in the two previous studies. This result estimates that the (A,T)-alternating pattern reflects the DNA signals for nucleosome packaging more effectively in yeast promoter regions.

To examine whether our approach is also applicable to other chromosome regions, we predicted the genome-wide nucleosome organization in yeast. The identification of high-resolution nucleosome positions throughout chromosome III in yeast provided an opportunity to assess our method [5], of which the accuracy of predicting yeast genome-scale nucleosome locations showed a slight decline, yet it was similar to those in the two other studies [18,29]. ~49% of our predicted nucleosomes were within 35 bp of those determined experimentally [5] compared with ~46% and ~48% in the two studies respectively. The decline is attributed to the potential unknown characteristic of DNA signals for nucleosome packaging in chromosome III and the prevalence of (A,T)-alternating pattern for nucleosome packaging in promoter regions.

Peckham et al. presented a new computational method for the prediction of nucleosome locations [30]. They predicted the nucleosome formation potential of any given DNA sequence by a support vector machine (SVM) trained on a subset of nucleosomal DNA sequences. The SVM used all possible k-mers ($k = 1$ to 6) as features for prediction. Since we both used the same yeast nucleosome positioning data throughout chromosome III for assessment, the two approaches were compared and their criteria for performance evaluation were used. Overall, 45.4% and 55.8% of our predicted nucleosomes were within 30 and 40 bp of those genome-scale experimentally determined nucleosome positions [5], corresponding to 41.3% and 49.8% in their report. In addition, a high-resolution map

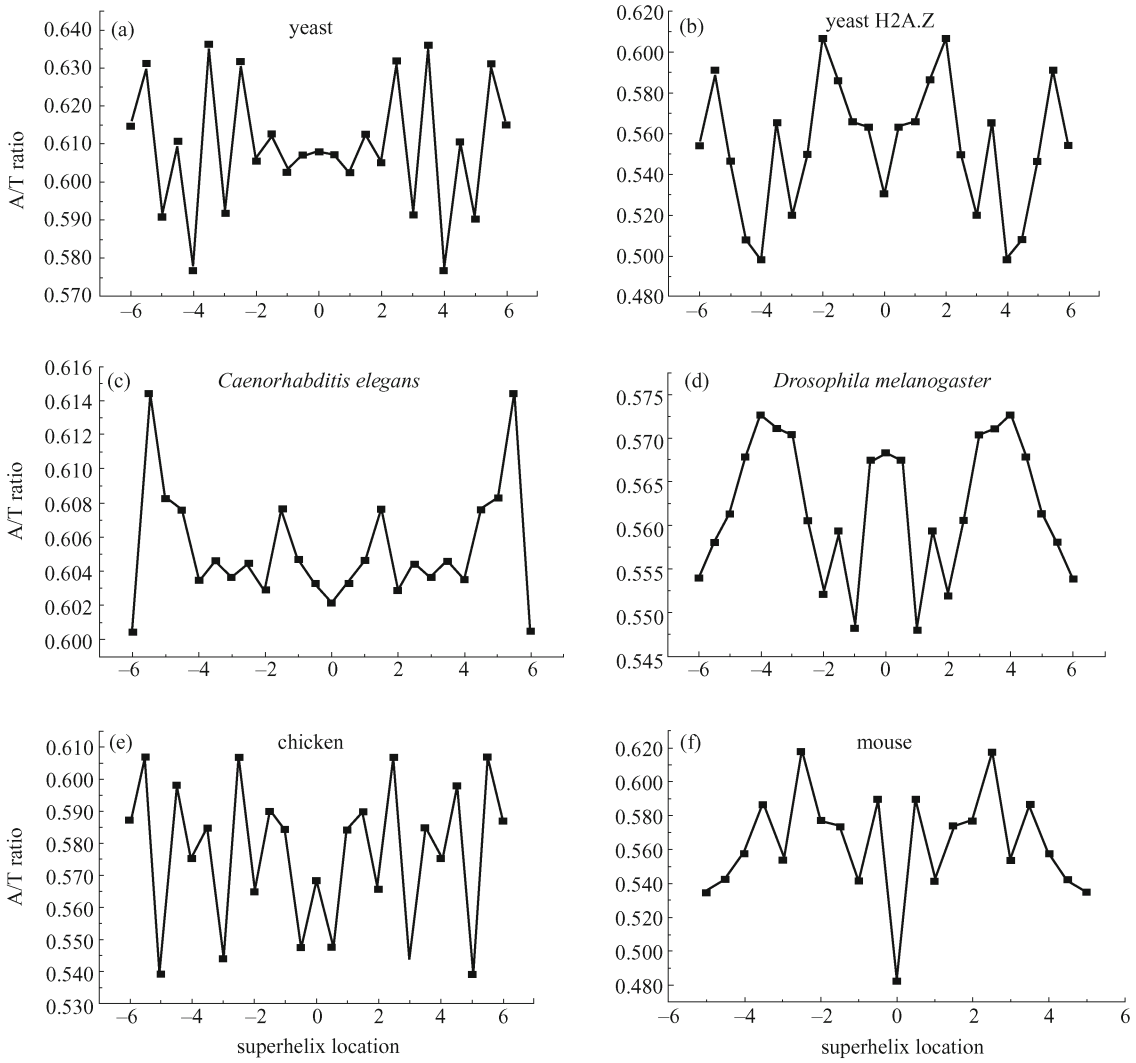

Fig. 1 Landscape of (A,T)-alternating patterns. (a)–(f) Fraction of A/T nucleotides at each half of superhelix location (SHL, each half of SHL represents a minor groove or a major groove, see Ref. [25]) of centre-aligned 129-bp-long yeast (a, Ref. [18]), yeast H2A.Z (b, the topmost 199 H2A.Z nucleosomes with the highest score from Ref. [6]), *Caenorhabditis elegans* (c, Ref. [26]), *Drosophila melanogaster* (d, Ref. [27]), chicken (e, Ref. [17]) and mouse (f, 109-bp long, Ref. [28]) nucleosome-bound sequences, where both the original form and the reverse complement form of each sequence are added to the centre alignment, showing strong recurring valley-peak patterns along yeast and chicken sequences. For these two species, valley-peak pattern appears at 11 and 10 out of 12 SHLs respectively, compared with 4 expected by chance

of genome-wide nucleosome occupancy in yeast has been completed [8]. ~51% of our predicted nucleosomes were within 35 bp of those determined experimentally across promoter regions. These results indicate that the (A,T)-alternating pattern may be a prevalent DNA signal for yeast nucleosome packaging.

4 DNA signals for nucleosome positioning vary with species and composition of histones

Do DNA signals for nucleosome packaging correlate with the composition of histones? Recently, histone variant H2A.Z nucleosomes, constituting about 10% of all cellular H2A molecules in S-phase [31], have been measured across the yeast genome [6]. Yeast H2A.Z nucleosomal sequences did not exhibit a strong (A,T)-alternating pattern (Fig. 1(b)). Both the proposed approach and other approaches [18] showed modest decline in prediction: ~44% of our predicted nucleosomes and ~42% of theirs were respectively within 35 bp of experimentally determined H2A.Z nucleosome positions. This result indicates that the genomic code for H2A.Z nucleosome positioning is different.

To study whether the DNA signals for nucleosome packaging vary with species, nucleosome locations in human promoters were predicted. We then compared the predictions with a collection of high-resolution positioned-nucleosomes identified in approximately 4000 human promoters [7]. The (A,T)-alternating

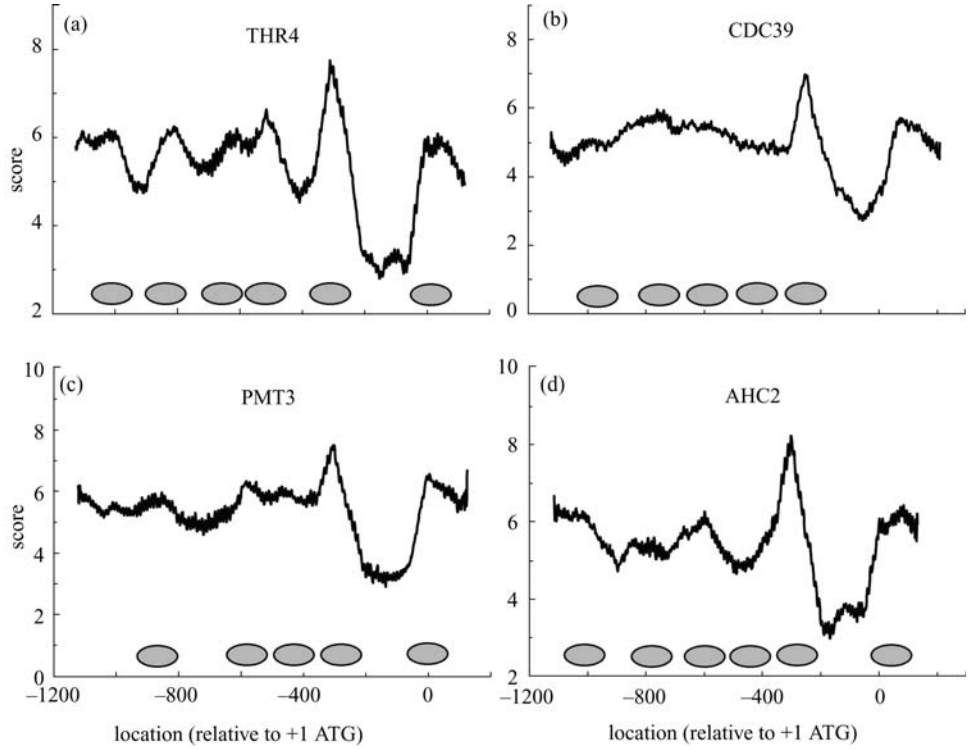

Fig. 2 Scoring profiles and experimental positioned-nucleosomes at selected promoter regions. (a)-(d) Black traces correspond to scores calculated by our approach. For each point, its score is a number proportional with the probability of being the nucleosome midpoint. The scoring profiles are smoothed by a 51-bp sliding window with 1-bp steps. Gray ovals represent nucleosomes identified in Ref. [5]. For most local maxima in scoring profiles, they correspond to experimentally determined nucleosomes

pattern became weak for nucleosome packaging in human promoter regions: ~38% of the predicted nucleosomes were within 35 bp of experimentally determined nucleosome positions. Additionally, nucleosomal DNA sequences from *Caenorhabditis elegans* exhibited a moderate (A,T)-alternating pattern (Fig. 1(c)), while those from *Drosophila melanogaster* and mouse displayed weak (A,T)-alternating patterns (Figs. 1(d) and 1(f)). Together with the remarkable (A,T)-alternating patterns of chicken and yeast, these results suggest that DNA signals for nucleosome packaging may vary with species.

The results demonstrate that DNA signals for nucleosome positioning vary with species and composition of histones. There may not be a uniform model for nucleosome-DNA interaction. However, the (A,T)-alternating pattern is a strong DNA signal for nucleosome positioning in some species from the clear periodic patterns of chicken and yeast nucleosome-bound sequences. A previous study showed that the AA/TT/TA dinucleotide pattern might be the dominant DNA signal for nucleosome packaging in yeast [18], but our approach showed superior results to the approach based on the dinucleotide pattern [19]. One possible interpretation is that the (A,T)-alternating pattern also contains periodic short DNA motifs (e.g. AA/TT/TA dinucleotide pattern) that help nucleosome packaging, suggesting that the (A,T)-alternating pattern captures most DNA signals for nucleosome formation in yeast.

5 Role of underlying DNA sequence in nucleosome positioning varies with genes

The results above demonstrate that intrinsic genomic organization can explain only a subset of nucleosome positions, which is consistent with other studies thus arriving at the conclusion that the underlying DNA sequence is not the sole determinant of nucleosome positioning [8,30]. We calculated the prediction accuracy for every gene promoter. Since the prediction is based on DNA sequence features, the prediction accuracy of every gene reflects the dependence of its nucleosome positioning on the underlying DNA sequence. The higher the accuracy is, the more the dependence is. Importantly, the distribution of accuracies is dispersive (Fig. 3), suggesting that the role of the intrinsic DNA sequence in nucleosome positioning varies with genes.

6 Conclusions

A novel computational approach was developed to predict nucleosome positions in yeast promoter regions. The reliable nucleosome predictions by the proposed method lay a good foundation for consequent analysis. The results show that DNA signals for nucleosome packaging correlate with species and composition of histones. It is suggested that there may not be a unified DNA sequence

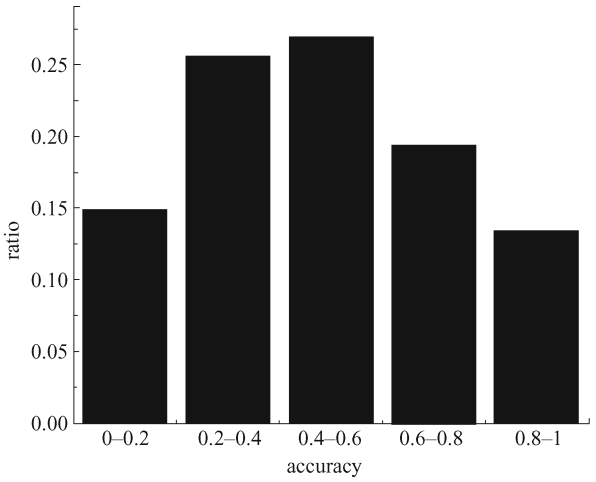

Fig. 3 Distribution of nucleosomal predictive accuracies of genes. Correct predictions are whose centers locate at 35 bp intervals from experimentally-determined nucleosome midpoints. The accuracy of one gene gives the fraction of experimentally determined nucleosomes that are correctly predicted in the promoter region

pattern for nucleosome positioning. Thus, it is necessary to develop specific models for predicting nucleosomes in some species. It is possible to extend our methodology to address this issue. It is also revealed that the underlying DNA sequence plays different roles in positioning nucleosomes of different genes. It will be worthy further study whether this difference is linked to different properties of the corresponding genes.

Appendix A Materials and methods

A.1 Data preparation

Yeast genome sequences were downloaded from the *Saccharomyces* Genome Database (http://www.yeastgenome.org). Human promoter sequences were downloaded from the UCSC Genome Browser (http://www.genome.ucsc.edu). Nucleosome positioning data in Fig. 1 and those for assessing the proposed method were downloaded from web supplements of concerned journals and papers.

A.2 Proposed approach for predicting nucleosome positions

The central 129 of 147 bp contribute significantly to the accommodation of the DNA super-helical path in the nucleosome than both 9 bp terminal segments [25]. We started by scanning along the DNA sequence with 1 bp steps with a 129 bp sliding window. In this way, all possible 129 bp sub-sequences in the DNA sequence were obtained. Each sub-sequence was assumed as central nucleosomal sequence and the fraction of A/T nucleotides at each half of the supposed SHL was calculated. We then assigned the number of SHLs displaying valley-peak pattern in the fraction of A/T nucleotides to this sub-sequence as its original score. In addition, we employed some complements to the scoring scheme, including penalizing the score of -2 for the NFR region (from -200 to -50) and favoring both 50 bp flanks of NFR region by adding 2 to the score. We chose these parameter values to enhance prediction accuracy. The score of each sub-sequence was assigned to its central location. The scoring profile was smoothed by a 51 bp sliding window with 1 bp steps. The scoring data were sorted in decreasing order. We determined the location with the highest score as the first nucleosome midpoint, and then iterated over the scoring data to determine the remaining nucleosome positions as long as the new nucleosome did not overlap with any previously determined nucleosomes. This process proceeded until no more nucleosomes could be laid (the algorithm implemented in Matlab is available upon request).

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 60474075).

References

1. Kornberg R D, Lorch Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. Cell, 1999, 98(3): 285–294
2. Khorasanizadeh S. The nucleosome: from genomic organization to genomic regulation. Cell, 2004, 116(2): 259–272
3. Luger K, Hansen J C. Nucleosome and chromatin fiber dynamics. Current Opinion in Structural Biology, 2005, 15(2): 188–196
4. Richmond T J, Davey C A. The structure of DNA in the nucleosome core. Nature, 2003, 423(6936): 145–150
5. Yuan G C, Liu Y J, Dion M F, et al. Genome-scale identification of nucleosome positions in *S. cerevisiae*. Science, 2005, 309(5734): 626–630
6. Albert I, Mavrich T N, Tomsho L P, et al. Translational and rotational settings of H2A.Z nucleosomes across the *Saccharomyces cerevisiae* genome. Nature, 2007, 446(7135): 572–576
7. Ozsolak F, Song J S, Liu X S, et al. High-throughput mapping of the chromatin structure of human promoters. Nature Biotechnology, 2007, 25(2): 244–248
8. Lee W, Tillo D, Bray N, et al. A high-resolution atlas of nucleosome occupancy in yeast. Nature Genetics, 2007, 39(10): 1235–1244
9. Wyrick J J, Holstege F C P, Jennings E G, et al. Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. Nature, 1999, 402(6760): 418–421
10. Flaus A, Owen-Hughes T. Mechanisms for ATP-dependent chromatin remodelling. Current Opinion in Genetics and Development, 2001, 11(2): 148–154
11. Strahl B D, Allis C D. The language of covalent histone modifications. Nature, 2000, 403(6765): 41–45
12. Jenuwein T, Allis C D. Translating the histone code. Science, 2001, 293(5532): 1074–1080
13. Henikoff S, Ahmad K. Assembly of variant histones into chromatin. Annual Review of Cell and Developmental Biology, 2005, 21: 133–153

14. Widom J. Role of DNA sequence in nucleosome stability and dynamics. Quarterly Reviews of Biophysics, 2001, 34(3): 269–324
15. Davey C, Pennings S, Meersseman G, et al. Periodicity of strong nucleosome positioning sites around the chicken adult globin gene may encode regularly spaced chromatin. Proceedings of the National Academy of Sciences of the United States of America, 1995, 92(24): 11210–11214
16. Roychoudhury M, Sitlani A, Lapham J, et al. Global structure and mechanical properties of a 10-bp nucleosome positioning motif. Proceedings of the National Academy of Sciences of the United States of America, 2000, 97(25): 13608–13613
17. Satchwell S C, Drew H R, Travers A A. Sequence periodicities in chicken nucleosome core DNA. Journal of Molecular Biology, 1986, 191(4): 659–675
18. Segal E, Fondufe-Mittendorf Y, Chen L, et al. A genomic code for nucleosome positioning. Nature, 2006, 442(7104): 772–778
19. Ioshikhes I P, Albert I, Zanton S J, et al. Nucleosome positions predicted through comparative genomics. Nature Genetics, 2006, 38(10): 1210–1215
20. Wang J Z, Widom J. Improved alignment of nucleosome DNA sequences using a mixture model. Nucleic Acids Research, 2005, 33(21): 6743–6755
21. Levitsky V G. RECON: a program for prediction of nucleosome formation potential. Nucleic Acids Research, 2004, 32: W346–349
22. Cao H, Widlund H R, Simonsson T, et al. TGGA repeats impair nucleosome formation. Journal of Molecular Biology, 1998, 281(2): 253–260
23. Baldi P, Brunak S, Chauvin Y, et al. Naturally occurring nucleosome positioning signals in human exons and introns. Journal of Molecular Biology, l996, 263(4): 503–510
24. Godde J S, Wolffe A P. Nucleosome assembly on CTG triplet repeats. Journal of Biological Chemistry, 1996, 271(25): 15222–15229
25. Luger K, Mäder A W, Richmond R K, et al. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature, 1997, 389(6648): 251–260
26. Johnson S M, Tan F J, McCullough H L, et al. Flexibility and constraint in the nucleosome core landscape of *Caenorhabditis elegans* chromatin. Genome Research, 2006, 16(12): 1505–1516
27. Mavrich T N, Jiang C, Ioshikhes I P, et al. Nucleosome organization in the *Drosophila* genome. Nature, 2008, 453(7193): 358–362
28. Widlund H R, Cao H, Simonsson S, et al. Identification and characterization of genomic nucleosome-positioning sequences. Journal of Molecular Biology, 1997, 267(4): 807–817
29. Yuan G C, Liu J S. Genomic sequence is highly predictive of local nucleosome depletion. PLoS Computational Biology, 2008, 4(1): e13. doi:10.1371/journal.pcbi.0040013
30. Peckham H E, Thurman R E, Fu Y, et al. Nucleosome positioning signals in genomic DNA. Genome Research, 2007, 17(8): 1170–1177
31. Redon C, Pilch D, Rogakou E, et al. Histone H2A variants H2AX and H2AZ. Current Opinion in Genetics and Development, 2002, 12(2): 162–169