Nucleosome Positioning of Intronless Genes in the Human Genome

Xiangfei Cheng, Yue Hou, Yumin Nie, Yiru Zhang, Huan Huang, Hongde Liu, and Xiao Sun

Abstract—Nucleosomes, the basic units of chromatin, are involved in transcription regulation and DNA replication. Intronless genes, which constitute 3 percent of the human genome, differ from intron-containing genes in evolution and function. Our analysis reveals that nucleosome positioning shows a distinct pattern in intronless and intron-containing genes. The nucleosome occupancy upstream of transcription start sites of intronless genes is lower than that of intron-containing genes. In contrast, high occupancy and well positioned nucleosomes are observed along the gene body of intronless genes, which is perfectly consistent with the barrier nucleosome model. Intronless genes have a significantly lower expression level than intron-containing genes and most of them are not expressed in CD4+ T cell lines and GM12878 cell lines, which results from their tissue specificity. However, the highly expressed genes are at the same expression level between the two types of genes. The highly expressed intronless genes require a higher density of RNA Pol II in an elongating state to compensate for the lack of introns. Additionally, 5' and 3' nucleosome depleted regions of highly expressed intronless genes are deeper than those of highly expressed intron-containing genes.

Index Terms—Epigenetics, intron-containing genes, intronless genes, nucleosome occupancy

1 Introduction

Nucleosomes are the basic packaging units of eukaryotic chromatin, which consist of approximately 147 base pairs (bp) of DNA wrapped around an octamer composed of eight histones [1]. There are many factors that contribute to nucleosome positioning, such as intrinsic DNA sequence preference [2], [3], transcription factors (TFs) [4], [5], RNA Pol II [6], chromatin remodelers [7], [8], [9], [10], and DNA methylation [11], [12]. Adding nucleosome disfavoring sequences to a yeast promoter results in a higher frequency of promoter transitions between inactive and active states [13]. Nucleosome organization plays a crucial role in gene regulation. The crystal structure of nucleosomes indicates that nucleosomes can spontaneously unwrap and rewrap their terminal DNA segments, allowing TFs to access their binding sites on nucleosomal sequences [14]. The closed or open state of chromatin may affect gene expression and transcriptional noise [15]. The advancement of high throughput technology allows us to generate genome-wide maps of nucleosome positioning [16], [17], [18]. Several general characteristics have been revealed about the nucleosome organization. First, a nucleosome depleted region (NDR) upstream of the transcription start site (TSS), followed by a regularly spaced nucleosome array, is involved in gene expression [16], [18], [19]. Second, there is a NDR also detected in the 3' untranslated region of many genes [20]. Additionally special nucleosome positioning patterns were discovered around some transcription termination sites (TTSs), which can be correlated with distinct polymerase occupancy and expression levels [21], [22], [23], [24]. Third, nucleosomes are well positioned at both ends of protein-encoding exons regardless of transcription level, while nucleosomes are depleted in the vicinity of the flanking introns [25], [26].

Eukaryotic genes often have an interrupted gene structure with multiple exons separated by introns. The exon-intron structure makes alternative splicing possible increasing the diversity of protein products [27]. Additionally, as a result of the functional coupling between splicing and other events in gene expression, the intron-containing structure influences other expression processes including mRNA export and transcription rate [28], [29]. The transcription efficiency of intron-containing transgenes is 10-100 times higher than that of their intronless counterpart [30]. In the human genome, approximately 3 percent of genes are intronless genes [27]. Human Intronless Genes Database includes 687 human intronless genes. Among them, only 323 genes can be functionally analyzed [31]. Intronless genes often encode signal transducing proteins, receptors, and regulatory molecules in growth and development [31], [32]. The fact that intronless genes have no introns naturally makes them special when compared with intron-containing genes. Intronless genes are typically expressed at a lower level and in a narrower breadth than intron-containing genes [33]. Intronless genes, without spliceosome recruiting conserved transcription/export complex, need a portion of the coding region called cytoplasmic accumulation region (CAR) to help mRNAs to export from the nucleus and keep stable [34]. A 10-nt consensus sequence found in the CAR can promote the stable accumulation of β-globin mRNA in the cytoplasm [35]. Intronless genes emerged after a burst of retrotransposition in primates approximately 38-50 million years ago. Intronless genes,

• *The authors are with the State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing 210096, China. E-mail: {chengxiangfeiangel, houyue_seu, zhangyiru_seu, ahuanghuan}@163.com, nieyum@126.com, {liuhongde, xsun}@seu.edu.cn.*

Manuscript received 10 Dec. 2014; revised 2 May 2015; accepted 5 Aug. 2015. Date of publication 25 Sept. 2015; date of current version 6 Aug. 2018.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TCBB.2015.2476811

1545-5963 © 2015 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

which are inferred to originate from the retroposition of cellular mRNA by retrotransposable elements, happen to reposition around some preexisting elements that promote their transcription [27].

Despite many researchers have focused on the mystery of intronless genes, the nucleosome organization of intronless genes has not been investigated. In this paper, we use publicly available data to examine how nucleosomes are positioned around intronless genes. Distinct nucleosome positioning patterns are found in intronless and intron-containing genes, which might result from intrinsic DNA sequence preference, DNA methylation and barrier nucleosome model. In general, intronless genes are unexpressed or lowly expressed, but when they are highly expressed, they need more RNA pol II that must be active to compensate for the shortage of introns.

TABLE 1
Nucleosome Occupancy Differences between Intronless Genes and Intron-Containing Genes[A]

Cell Lines	Region	P-value
Resting CD4+ T	promoter	6.4308×10^{-5}
Resting CD4+ T	gene body	1.3899×10^{-19}
Resting CD4+ T	downstream TTS	7.6802×10^{-10}
Activated CD4+ T	promoter	6.8019×10^{-7}
Activated CD4+ T	gene body	1.7129×10^{-5}
Activated CD4+ T	downstream TTS	1.5497×10^{-5}
GM12878	promoter	3.3976×10^{-6}
GM12878	gene body	8.1203×10^{-51}
GM12878	downstream TTS	2.0929×10^{-37}

[A] *We used the Kolmogorov-Smirnov test to check the difference in nucleosome occupancy of various regions between intronless and intron-containing genes.*

2 Results and Discussion

2.1 Nucleosome Occupancy Around TSSs and TTSs of Intronless Genes

We calculated the average nucleosome occupancy around TSSs and TTSs of both intronless and intron-containing genes in CD4+ T cells and GM12878 cells (Fig. 1). The differences in the nucleosome occupancy of variant regions between intronless and intron-containing genes were calculated (Table 1). The nucleosome occupancy of intronless and intron-containing genes is quite different in promoters and gene bodies. The P values also proved the same conclusion. The nucleosome occupancy pattern of intron-containing genes is consistent with that showed in previous studies [10], [36], [37]. The first nucleosome upstream of the TSS (−1 nucleosome) is well positioned in front of the obvious NDR of intron-containing genes. The first nucleosome downstream of the TSS (+1 nucleosome) is also well positioned and the downstream nucleosome occupancy displays oscillations with amplitude that decays with

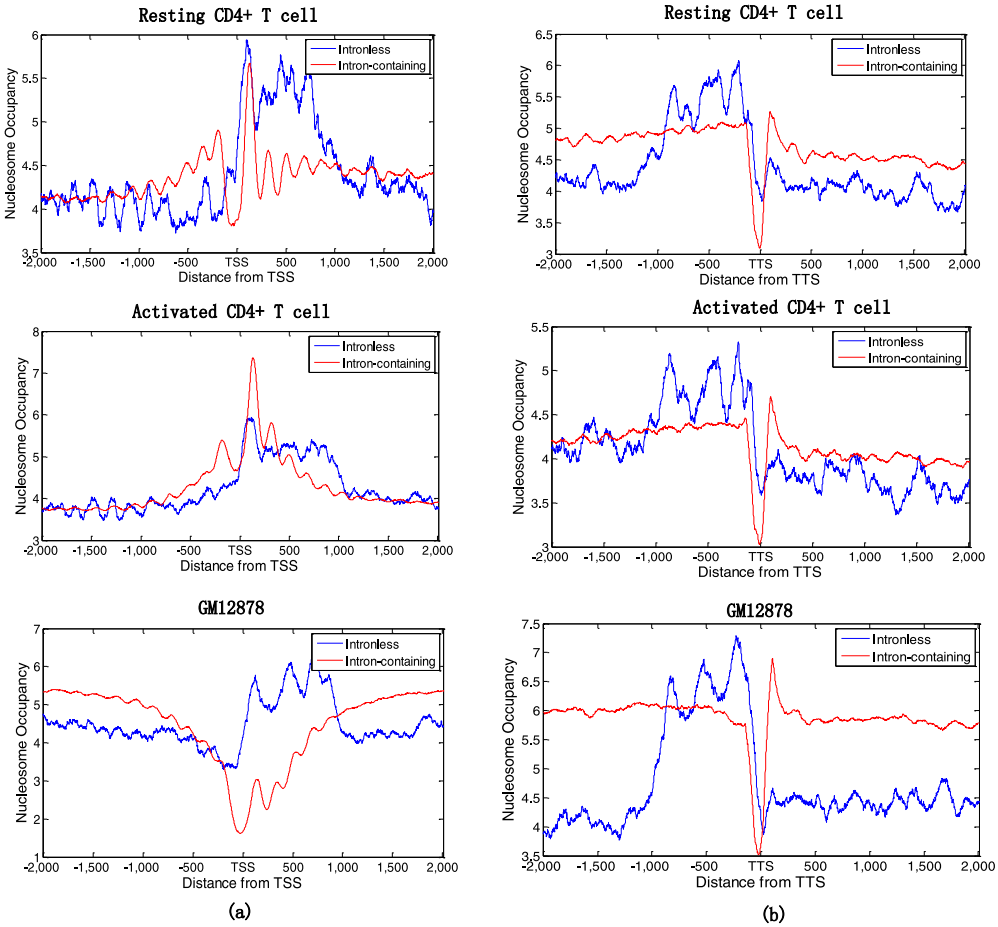


Fig. 1. The nucleosome occupancy around transcription start sites and transcription termination sites of intronless genes (blue line) and intron-containing genes (red line). (a) The nucleosome occupancy around TSSs. (b) The nucleosome occupancy around TTSs.

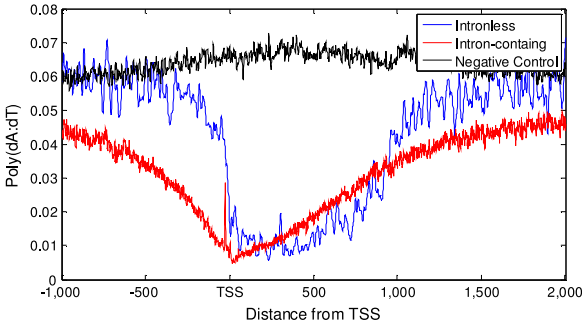


Fig. 2. The poly (dA:dT) distribution of intronless genes (blue) and intron-containing genes (red). $P\_\text{promoter} = 6.4565 \times 10^{-74}$ (*P* value of poly (dA:dT) tracts in the promoter between intronless genes and intron-containing genes). *P*_body = 0.0436 (*P* value in the gene body between intronless genes and intron-containing genes). *P* values were calculated using the Kolmogorov-Smirnov test.

increasing distance from the TSS (Fig. 1a). There is also an obvious NDR immediately around the TTSs of intron-containing genes (Fig. 1b). A well-positioned nucleosome downstream of the NDR of TTS is associated with the identification of TTSs. The pattern for intronless genes is distinct from that for intron-containing genes.

First, nucleosome occupancy in the promoters of intronless genes is significantly lower than that of intron-containing genes (Fig. 1, Table 1). Because of the nature of nucleosome depletion in CpG islands around the TSS [38], [39], we inferred that the CpG islands in promoters might be responsible for the low nucleosome occupancy in the promoter of intronless genes. However, only a small portion of intronless genes have CpG islands (250 of 939 [see *Data and methods*]). Moreover, when just considering the CpG island-free genes, intronless genes have lower nucleosome occupancy in the promoter regions than intron-containing genes (**Fig. S1,** which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TCBB.2015.2476811). This indicates that there must be factors other than CpG islands in promoters that cause the low nucleosome occupancy in the promoters of intronless genes. Poly (dA:dT) tracts and their flanking DNA are relatively depleted of nucleosomes *in vivo* because of their rigid properties [40]. In the promoters, the poly (dA:dT) tracts in intronless genes are significantly higher than those in intron-containing ones (Fig. 2), while the differences in poly (dA:dT) tracts in the gene body are not so obvious. Compared with generic sequence DNA, poly (dA:dT) tracts have a shorter helical repeat, a narrow minor groove, a distinct spine of hydration within the minor grove, and maximal overlap of the bases separately within each strand [40]. These unusual properties allow poly (dA:dT) tracts to intrinsically resist the structure deformations that are required for nucleosome formation. We concluded that the poly (dA:dT) tracts in promoters might be an important sequence preference that influences the differences in nucleosome occupancy in the promoters of intronless and intron-containing genes. The differences in nucleosome occupancy of gene bodies between intronless and intron-containing genes are influenced by other factors.

The +1 nucleosome and −1 nucleosome of transcribed genes are acetylated and methylated [16], [41]; these nucleosomes are recognized by TFIID [42], [43], [44]. In TATA box-containing genes, TFIID then delivers TBP to promoters [45]. TBP binds TFIIB and places it immediately downstream of the TSSs, and TFIIB positions Pol II at the promoter [46]. Some evidences suggest that TFIIB controls TSSs selection. Therefore, the well-positioned nucleosome around an NDR contributes to the recognition and selection of a TSS. The fuzzy promoter nucleosomes of intronless genes indicate that the recognition of intronless TSS may be inconvenient. When intronless genes are highly expressed, the −1 nucleosome and +1 nucleosome become well positioned (see *Highly expressed intronless genes are associated with elongating Pol II*), which can help the transcriptional initial complex to recognize the TSSs.

Second, intronless genes have shallower NDRs around the TTSs and lower nucleosome occupancy downstream of the NDRs than intron-containing genes (Table 1). Because most intronless genes are not expressed in the CD4+ T cell lines, the nucleosome occupancy around TTSs of intronless genes (Fig. 1b) is almost the same as that of unexpressed genes (more details will be discussed in *Highly expressed intronless genes are associated with elongating Pol II)*. The nucleosome levels immediately downstream of TTSs of intronless genes are lower than those of intron-containing genes because of the sequence preference (see *Highly expressed intronless genes are associated with elongating Pol II*). And we suspect that the low nucleosome occupancy downstream of TTSs leads to inefficient polyadenylation of intronless transcripts [47]. Besides, some intronless genes (45 intronless genes) have protein heterodimerization activity (see **Table S2,** available in the online supplemental material), including the genes encoding histones. Transcripts of such histone genes contain a unique 3′ end instead of being polyadenylated [48], and nucleosomes immediately downstream of TTSs do not act as the pause site, which is the role they play in intron-containing genes that allows RNA Pol II more time to process 3′ end−event, including the addition of a poly(A) tail. For these reasons, nucleosome occupancy downstream of TTSs of intronless genes is lower than that of intron-containing genes. The distinct nucleosome positioning pattern around TTSs indicates that intronless genes might use a different way to identify the TTSs and to promote the formation of mRNA 3′ end.

Third, the nucleosome occupancy in the region of 0-1,000 bp downstream of TSSs was higher than that in the flanking region of gene body in intronless genes. Additionally, a higher level of nucleosome occupancy about 1,000 bp upstream of TTSs was observed as well. Given the shorter length of intronless genes, we speculate that the higher nucleosome occupancy level downstream TSSs of intronless genes is related to their gene body, which will be discussed in detail in the following section.

2.2 Nucleosomes Occupancy Is High in the Gene Body of Intronless Genes

As discussed above, the higher nucleosome occupancy downstream of TSSs of the intronless gene was suspected to be related to their gene body. To test our hypothesis, we examined the length of intronless genes. Most intronless genes are ~1,000 bp (Fig. 3), while the size of intron-containing genes

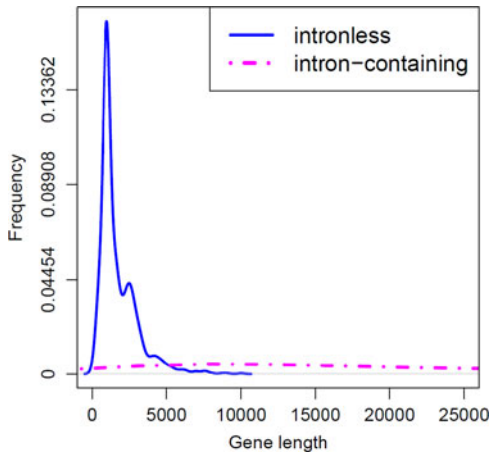


Fig. 3. The length of intronless genes and intron-containing genes. $P = 1.5678 \times 10^{-58}$. The *P* value was calculated using the Student's t-test.

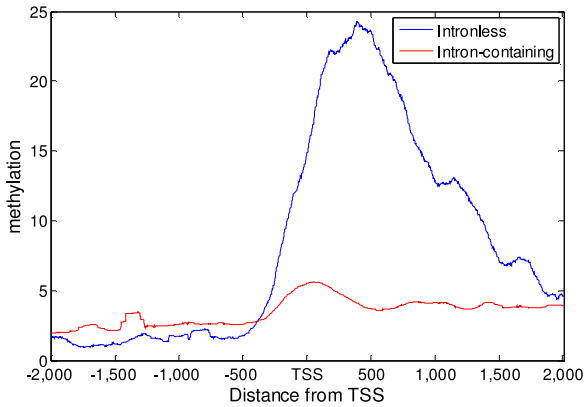


Fig. 5. The methylation around the TSSs of intronless genes (blue) and intron-containing genes (red) in resting CD4+ T cells. We also calculated the differences of the methylation in gene body of intronless genes and intron-containing genes using the Kolmogorov-Smirnov test. $P\_\text{body} = 6.0579 \times 10^{-16}$.

is evenly distributed. The length distribution of intronless genes and intron-containing genes is quite different (Fig. 3, $P < 0.01$).

Nucleosome positioning is partly determined by intrinsic DNA sequence preferences. Previous studies showed a genomic code for nucleosome positioning [2], [3]. The feature of 10-bp periodicities of specific dinucleotides along the nucleosome was observed in many genomes [8], [49]. AT-rich DNA sequences are characteristic of NDRs because of their rigid and bent property, making them less compatible with the extreme bending that is required for nucleosome formation [50], [51]. GC content is a major determinant of nucleosome positioning because of its influences on many DNA structure preferences [51]. We analyzed *in vitro* reconstructed nucleosome positioning patterns and the predicted nucleosome positioning patterns based only on DNA sequence. The *in vitro* reconstruction and predicted nucleosome occupancy of intronless genes and intron-containing genes are significantly different in the gene body, just like the nucleosome occupancy *in vivo* [50] (Fig. 4, $P < 0.01$). The *in vitro* reconstruction and predicted nucleosome occupancy in the region of 0-1,000 bp downstream of TSSs of intronless genes was higher than that of other regions, which was consistent with the *in vivo* nucleosome occupancy of intronless genes (Fig. 1a). In addition, most intronless genes are not expressed in human CD4+ T and GM12878 cells; in contrast, only 16 percent of intron-containing genes are not expressed and there are almost no TFs and RNA Pol II binding in the unexpressed genes (see *Highly expressed intronless genes are associated with elongating Pol II*). Therefore TFs and RNA Pol II should not be the main reasons that lead to the average nucleosome positioning of whole intronless genes. In conclusion, intron-containing genes are obviously more affected by other factors than intronless ones; because their nucleosome occupancy along the gene body *in vitro* or predicted by DNA sequence is inconsistent with the nucleosome occupancy *in vivo*.

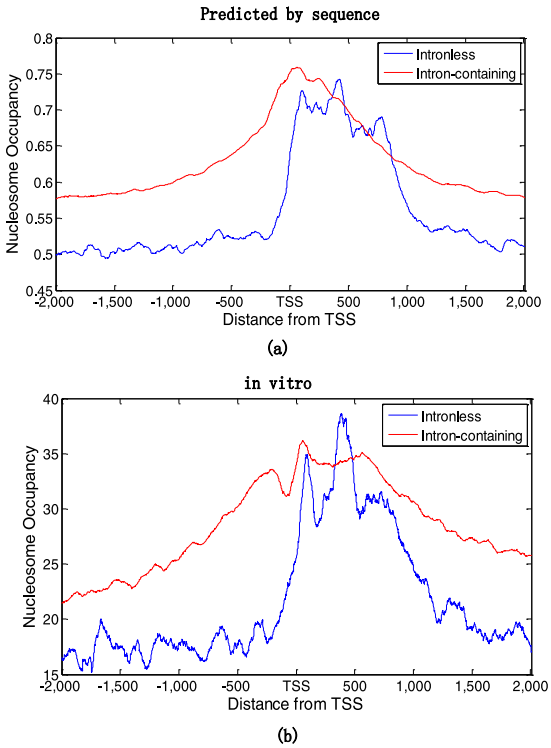


Fig. 4. Intrinsic nucleosome positioning pattern. (a) The nucleosome positioning pattern of intronless genes (blue) and intron-containing genes (red) predicted by DNA sequence. $P\_\text{body} = 3.5668 \times 10^{-4}$ (*P*-value in the gene body of intronless genes and intron-containing genes). (b) The *in vitro* nucleosome pattern of intronless genes (blue) and intron-containing genes (red). $P\_\text{body} = 1.9270 \times 10^{-6}$. *P* values were calculated using the Kolmogorov-Smirnov test.

Methylation of DNA at CpG dinucleotides represents an epigenetic remark that is involved in the regulation of gene expression in vertebrate cells. It is because DNA methylation can alter the properties of the DNA sequence. Nucleosome positioning is also affected by 5-methycytosine. DNA methylation has been reported to promote the compaction and stabilization of nucleosomes by causing the overwrapping of DNA around the histone octamer [12], [52]. As seen in Fig. 5, the gene bodies of intronless genes are hypermethylated compared to the methylation of intron-containing genes ($P\_\text{body} = 6.0579 \times 10^{-16}$). In addition, our result agrees with that of a genome wide analysis of the relationship between nucleosome positioning and DNA methylation that showed that exons were enriched in nucleosomes that are

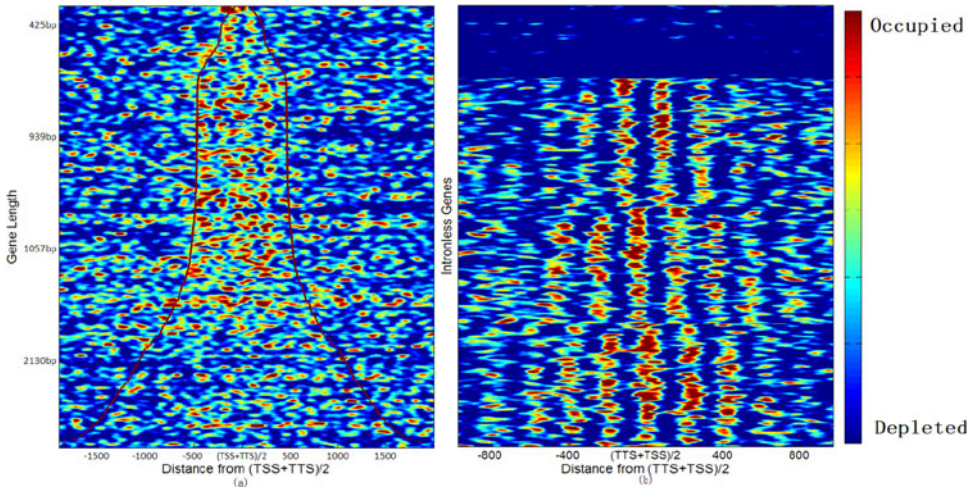


Fig. 6. The heatmap of nucleosome occupancy in intronless genes in resting CD4+ T cell lines. (a) Nucleosome occupancy in the gene body of intronless genes; the gene length increases from top to bottom, and the red line on the plot marks the sites of TSSs and TTSs. (b) K-means cluster for the nucleosome occupancy in the gene body of intronless genes ($k = 4$).

preferentially positioned in methylated sequences rather than unmethylated sequences [11]. The hypermethylation of the gene body of intronless genes is an important factor for the high nucleosome occupancy of intronless gene bodies.

DNA sequence preference and DNA methylation can partly interpret the *in vivo* nucleosome positioning [2], [7], [53]. However evidences suggest that the organization of nucleosome in specific regulatory regions is largely a consequence of statistical packing principles. The barrier nucleosome model indicates that the positioning of nucleosomes depends on barriers that can be formed by well positioned nucleosomes, TFBSs, AT-enriched sequence, RNA Pol II, and etcetera [8], [10], [36], [54]. We observed that the +1 nucleosome (Fig. 1a) and the nucleosomes immediately upstream of TTSs (Fig. 1b) were well positioned; these results were consistent with the barrier nucleosome model. When intronless genes are short, these two nucleosomes act as barriers that form a "box" in which nucleosomes must be posited (Fig. 3b). However, when intronless genes become longer, the nucleosomes' roles as barriers fade, resulting in a fuzzily positioned nucleosome. To illustrate this view, we plotted the heatmap of the nucleosome occupancy in intronless genes and intron-containing genes in resting CD4+ T cell lines (Fig. 6).

The nucleosome occupancy along the gene body is higher than that of the flanking region of the genes, and nucleosomes prefer a periodicity in the gene body (Fig. 6a). This phenomenon perfectly proves the nucleosome barrier model. As the gene length increases, the nucleosomes along the gene body become fuzzy and non-periodic, suggesting that the exclusion power from the barriers forces the nucleosomes to be periodic. However, when the gene length exceeds the influential region of the barriers, the impact from barriers weakens. When the gene length is longer than 2,000 bp, the confinement induced by barriers will not be strong enough to restrict the nucleosomes on gene bodies, which leads to fuzzier nucleosome occupancy. To understand nucleosome barriers more clearly, we clustered the nucleosome signals along the gene body using the K-means ($k = 4$) method (Fig. 6b and Fig. S2, available in the online supplemental material, we also checked the heatmap when $k = 3$ and $k = 5$. The figure was showed in Fig. S2, available in the online supplemental material). The nucleosome peaks were found using iNPS, a program that determines accurate nucleosome positioning from sequence data [55]. In Fig. 6, all intronless genes were aligned at the middle of the gene bodies. In Fig. S2, available in the online supplemental material, each gene was aligned at the +1 nucleosome and the last nucleosome of the gene body respectively. The 5′ and 3′ nucleosomes were positioned using the methods proposed by Vaillant and Palmeira [56]. After clustering the signal, the periodicity of nucleosomes in the gene body is observed more clearly (Fig. 6b). The nucleosomes along the gene body are forced to be periodic by the barriers at the two ends of the genes. When nucleosomes are closer to the barrier, they prefer higher and more periodic. This can be seen in Fig. S2, available in the online supplemental material. +1 nucleosome and the first nucleosome upstream the TTS act as the barriers. The nucleosomes that are close to these two nucleosomes have a significant periodicity. The amplitudes of +1 nucleosome and the last nucleosome in gene body are higher than that of other nucleosomes, and the nucleosome occupancy signal decays as the distance to the barriers increase.

In conclusion, the sequence preference, DNA methylation, and barrier nucleosome model all influence the nucleosome occupancy along the gene body of intronless genes. These factors contribute to the unique nucleosome occupancy of intronless genes.

2.3 Highly Expressed Intronless Genes Are Associated with Elongating Pol II

Introns facilitate transcription in *S.cerevisiae* [57] and *Drosophila* [58]; this is also true in the human genome. To examine the nucleosome dynamical repositioning with gene expression, genes were divided into three groups (see Data and methods) based on gene expression values (Reads Per-Kilobase of exon model per million mapped reads, RPKM): highly expressed genes (intronless gene: 116, intron-containing gene 7,235), lowly expressed genes (intronless gene: 264, intron-containing gene: 7,591) and unexpressed genes (intronless genes: 555, intron-containing gene: 2,954). Most intronless genes are unexpressed. Of the expressed

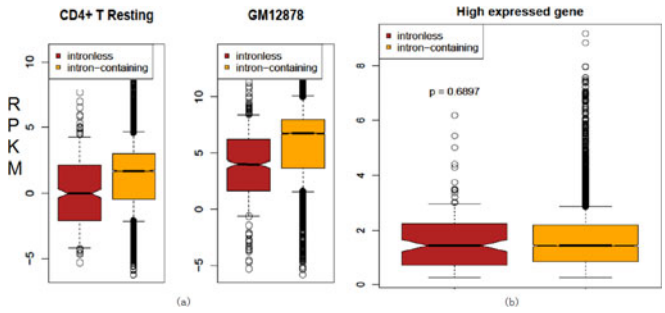


Fig. 7. The expression of intronless genes (red) and intron-containing genes(orange). (a) All expressed genes of resting CD4+ T cells (left) and GM12878 cells (right). Both P values are 2.2×10^{-16} (rank-sum test). (b) Highly expressed genes of Resting CD4+ T cell. P value = 0.6897 (rank-sum test).

intronless genes, the expression levels are obviously lower than those of intron-containing genes (Fig. 7a). The most plausible explanation for this phenomenon is that intronless gene expression is tissue-specific. Therefore most intronless genes are expressed at a low level in CD4+ T and GM12878 cells. To confirm this view, intronless genes were analyzed in the Kyoto Encyclopedia of Genes and Genomes (KEGG) by DAVID [59], [60], an online bioinformatic tool to annotate and integrate gene analysis. Totally, 628 intronless genes were found in the KEGG pathways by DAVID, which included 342 genes that are related to olfactory transduction (see **Table S1,** available in the online supplemental material). Other genes mostly take part in taste transduction (24 intronless genes), systemic lupus erythematosus (31 intronless genes), neuroactive ligand-receptor interaction (34 intronless genes). Gene Ontology (GO) analysis revealed that 359 intronless genes have olfactory receptor activity (see **Table S2,** available in the online supplemental material) [61].

Even though the intronless genes have tissue-specificity and express at a low level in CD4+ T and GM12878 cell lines, the highly expressed intronless genes and intron-containing genes seem to have similar average expression levels (Fig. 7b). Then, we analyzed the nucleosome organization of genes with different expression levels and calculated the P value of nucleosome occupancy according to different expression levels between intronless genes and intron-containing genes (Fig. 8, Table S3, available in the online supplemental material). Immediately around the TSSs of both intron-containing and intronless genes with a high expression level, we observed noticeable NDRs, accommodating TFs binding to TFBSs. The amplitude of the $+1$ nucleosome and -1 nucleosome is the highest among these groups of genes in both of intronless genes and intron-containing genes. This indicates that the $+1$ nucleosome and -1 nucleosome might play an important role in gene expression. The depths of NDRs and the $+1$ nucleosome amplitudes around TSSs of lowly expressed genes are decayed, although they still can be observed. In unexpressed gene, the NDR disappeared. There was only one high nucleosome occupancy downstream TSSs of intron-containing genes, and the nucleosome occupancy of unexpressed intronless genes was similar to that of all intronless genes. This indicates that the nucleosome positioning patterns are dynamically adjusted to allow the binding of transcription factors and RNA Pol II.

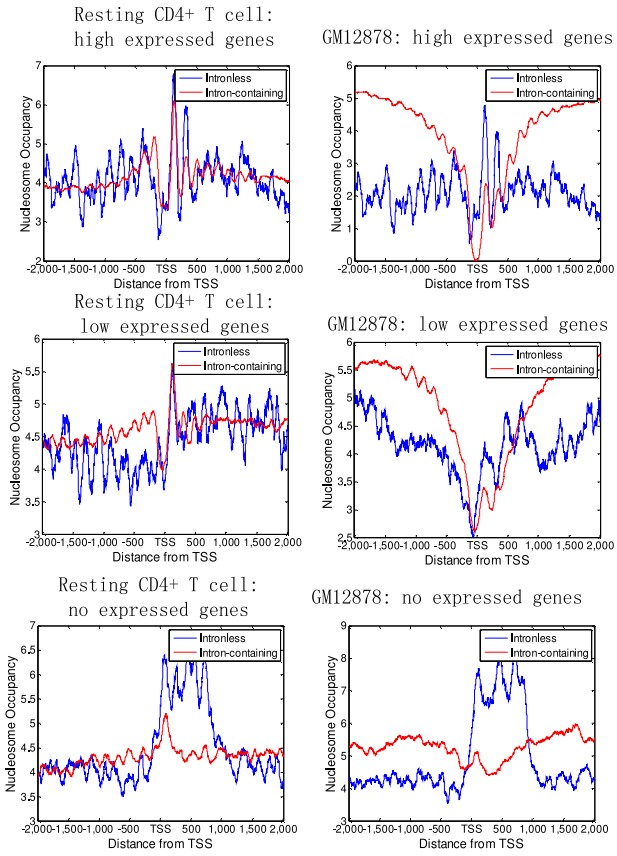


Fig. 8. The nucleosome positioning patterns around TSSs in different expressed levels. (a) The nucleosome occupancy of genes in resting CD4+ T cell. (b) The nucleosome occupancy of genes in GM12878 cells.

In addition, from Fig. 8a, -1 nucleosome of highly expressed intronless genes disappear in CD4+ T cell, wider and deeper NDRs and higher $+1$ nucleosome are observed in highly expressed intronless genes in both CD4+ T and GM12878 cell lines. For intron-containing genes, the process of splicing promotes gene expression. Because intronless genes have no introns, they need some other ways to compensate for this shortage.

Nucleosome phasing and positioning surrounding the TSSs may be related to Pol II binding. Therefore, we examined the RNA Pol II tags in highly expressed promoters, lowly expressed promoters, and unexpressed promoters (Fig. 9). The RNA Pol II density decreases when the expression level is low in both intronless and intron-containing genes. It is intriguing to note that the RNA Pol II density of intronless genes is higher than that of intron-containing genes in expressed genes. These results suggest that more RNA Pol II is required in transcription of intronless genes.

Previous studies suggested that a fraction of promoters are associated with poised or stalled RNA Pol II [62], and that the activity of RNA Pol II influences gene expression profiles [16]. Therefore, we adopted the scenario described in data and methods to determine RNA Pol II activity (see *Data and Methods*). Over 80 percent highly expressed intronless genes are associated with elongating Pol II, while only 44 percent of highly expressed intron-containing genes are associated with elongating Pol II. These data suggest that intronless genes

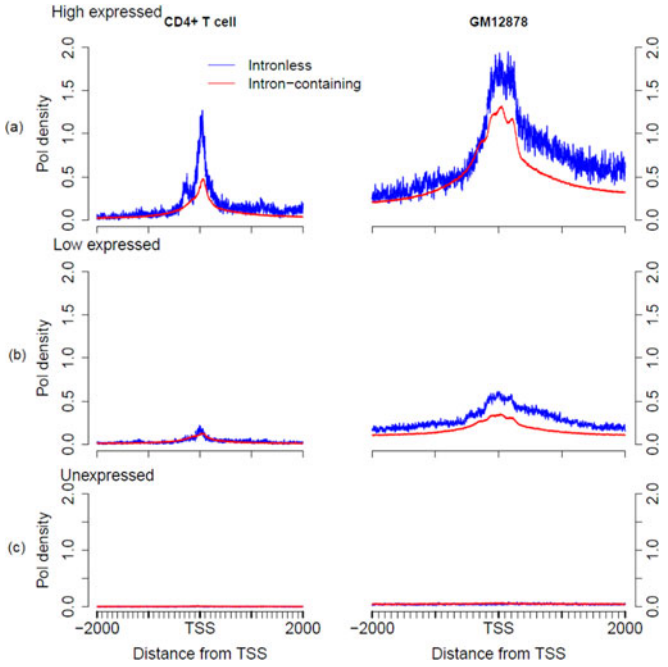


Fig. 9. The RNA Pol II density around TSS of intronless genes (blue) and intron-containing genes (red). Left Panel: Resting CD4+ T cell; Right Panel: GM12878 cells. (a) Highly expressed genes. (b) Lowly expressed genes. (c) Unexpressed genes.

require not only a high density of RNA Pol II, but also Pol II in an elongating state. Introns can enhance almost every step of gene expression, from transcription to translation [28], [63], [64]. The higher density and elongating RNA Pol II compensate for the shortage in the introns and splicing for intronless genes.

The binding of RNA Pol II and TFs often reposition nucleosome around TSSs. The nucleosome positioning patterns of unexpressed genes are similar to those of the predicted and *in vitro* nucleosome positioning. The DNA sequence preference may play an important role in nucleosome positioning. If TSSs are masked by nucleosomes, unnecessary or incorrect transcription will be reduced [65]. Once one gene is prepared to express, some pioneer TFs may bind to the nucleosomes around TSSs and open the local chromatin [66], [67] to enable the access of TFs and/or RNA Pol II to DNA. TFs and RNA Pol II will recruit other needed proteins to form the transcriptional complex and chromatin modifiers to initial transcription and dynamically repositioning the nucleosomes. The binding of RNA Pol II is often associated with strong positioning of the +1 nucleosome and wider NDR, while the absence of RNA Pol II is associated with fuzzy positioning [9], [16]. As a result, wide and deep NDRs appear in highly expressed genes, a well-positioned +1 nucleosome and subsequent nucleosome array also appear. The higher density RNA Pol II might be a reason for the higher occupancy of +1 nucleosome and wider NDR of highly expressed intronless genes.

To further study the relationship between nucleosomes around TTSs of intronless genes and gene expression levels, we plotted the nucleosome occupancy around the TTSs of genes with different expression levels (see Fig. 10). It seems that the 3′ NDR upstream of TTSs of intronless genes is strongly correlated with the expression level. Higher intronless gene expression tended to be associated with deeper 3′

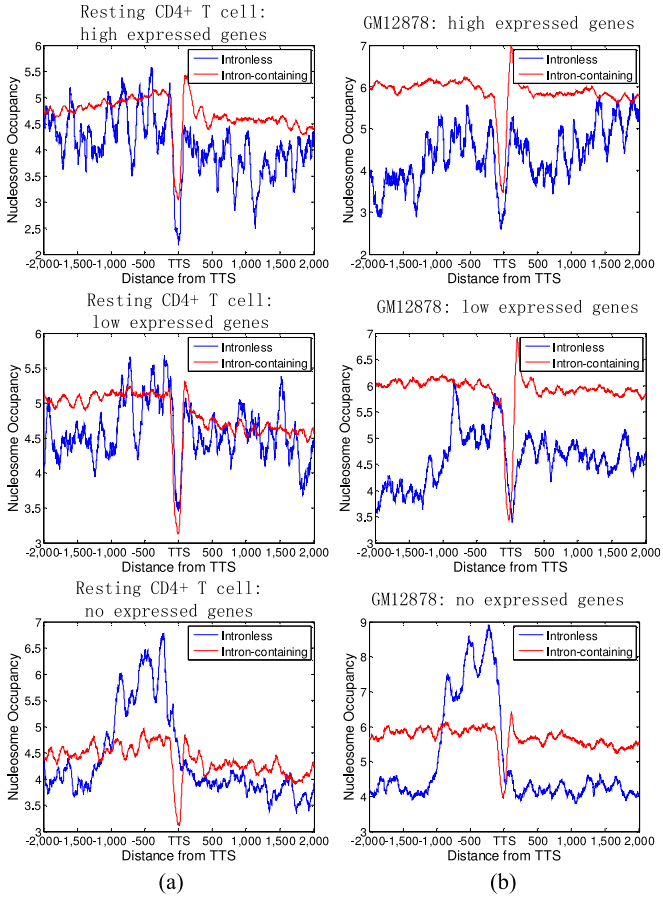


Fig. 10. The nucleosome occupancy around TTSs of intronless genes (blue) and intron-containing genes (red) at different expression levels in Resting CD4+ T cells and GM12878 cells. (a) Resting CD4+ T cells, and (b) GM12878 cells. We calculated *P* values of nucleosome occupancy downstream of TTSs of intronless genes and intron-containing genes (Table S4, available in the online supplemental material).

NDRs. In contrast, this property was not observed in intron-containing genes. The nucleosomes upstream TTSs of intron-containing genes have almost the same pattern even though they are expressed in different expression levels, which indicated that nucleosome occupancy around TTSs of intron-containing genes is more affected by the sequence preference. This is supported by the GC content around TTSs (Fig. S3, available in the online supplemental material). Some researchers suggest that the 3′ NDR may be a site for RNA polymerase disassembly [54]. Because most intronless genes are unexpressed in the CD4+ cell lines and GM12878 cell lines, the nucleosome occupancy around the TTSs of intronless genes (Fig. 1b) is the same as that of unexpressed genes (Fig. 10).

The 3′ NDRs of highly expressed intronless genes are stronger than those of intron-containing genes (Fig. 10). In fact, we suspect the conclusion should be adapted for all intronless and intron-containing genes that are expressed at the same levels. Even though the lowly expressed intron-containing genes have deeper 3′ NDRs, their expression levels are also significantly higher than lowly expressed intronless genes (see **Fig. S4,** available in the online supplemental material) which may lead to deeper 3′ NDRs.

In addition, the nucleosome occupancy immediately downstream of the TTSs of intronless genes is lower than that

of intron-containing genes because of the sequence preference (see **Fig. S3**, available in the online supplemental material). The GC content downstream of the TTSs of intronless genes is obviously lower than that of intron-containing genes. Additionally some transcripts of intronless genes, especially histone genes, are not polyadenylated at the natural site [47], [48]. Therefore the nucleosomes immediately downstream TTS do not act as the pause site as it does in intron-containing genes for the addition of the 3′ poly(A) tail. Nucleosomes positioned downstream of the TTSs might act as RNA Pol II pause sites and allow the RNA Pol II more time to process 3′ end events [20], [68]. Nucleosome occupancy near TTSs regulates the transcription termination of protein-coding genes. While the nucleosome occupancy for intronless genes displays a different pattern from that of intron-containing genes, it will be interesting to examine whether it implies a different transcription termination mechanism for intronless genes.

3 Conclusion

Intronless genes are a special group of genes in the eukaryotic organism. Therefore we investigated the nucleosome occupancy patterns of intronless genes. The nucleosome positioning pattern of intronless genes is distinct from that of intron-containing genes. The nucleosome levels in promoters of intronless genes are lower than those of intron-containing genes (Fig. 1a), resulting in genes that are unexpressed or expressed at a low level. The abundance of poly (dA:dT) in intronless gene promoters might be the decisive factor for the lower nucleosome occupancy. In addition, the nucleosome positioning pattern around TTSs of intronless genes is also different from that of intron-containing genes, which may indicate a different way to identify TTSs for intronless genes. When the genes are highly expressed in the cells, intronless genes always have deeper 3′ NDRs upstream the TTSs.

Intronless genes in the 900-bp region downstream of TSSs not only have a higher nucleosome occupancy than other regions but also have well positioned nucleosomes when these genes are shorter than 2,000 bp (Fig. 1a), which is correlated with intronless gene bodies (Fig. 6). Our results demonstrate that the intrinsic DNA sequence preferences partly determine the nucleosome positioning pattern (Fig. 4) and that the NDRs near TSSs and TTSs act as barriers that result in periodic nucleosome positioning along intronless gene bodies. In addition, the hypermethylated intronless gene bodies (Fig. 5) promote the compaction and stabilization of nucleosomes and contribute to the high nucleosome occupancy of intronless gene bodies.

In general, intronless genes are expressed at a lower level (Fig. 7) and in a narrower range than intron-containing genes [33], because most intronless genes are tissue-specific. However the expression values of highly expressed genes of intronless genes and intron-containing genes were almost the same (the low expressed genes are not (see **Fig. S3,** available in the online supplemental material)). Highly expressed intronless genes have deeper NDRs and higher +1 nucleosome occupancy than intron-containing genes, which may accommodate more TFs and RNA Pol II (Fig. 9). Additionally, most highly expressed intronless genes are associated with elongating RNA Pol II. In this way, intronless genes may compensate for the lack of introns.

When a gene is unexpressed *in vivo*, there is almost no Pol II binding at the promoter. Hence, DNA sequence preference may determine the nucleosome positioning and phasing of unexpressed genes. Once a gene is prepared for expression, some pioneer factor will bind to its target site in nucleosomal DNA and recruit ATP-dependent chromatin remodelers to displace nucleosomes and allow other TFs and RNA Pol II to access their sites. Then, a NDR will be formed around the TSS, and the nucleosome array will be positioned flanking the NDR with the assistant of Isw2.

These results expand our knowledge of intronless genes and nucleosome organization *in vivo*.

4 Data and Methods

4.1 Identification of Intronless Genes

The genes were downloaded from RefSeq Genes of hg18 datasets [69]. We chose all protein-coding genes as our objects and excluded non-coding genes. We then selected genes with only one exon in the RefSeq Gene as the candidates of intronless genes. Pseudogene and some protein coding genes whose TSSs were contained within other genes were abandoned by manual work. Finally we obtained 939 intronless genes from the hg18 human genome. Genes with more than one exon were regarded as intron-containing genes. If a gene had more than one TSS, we selected the TSS that was the closet to the 5′ end. Finally 17,776 intron-containing genes were obtained from hg18. The TTSs sites were also obtained from RefSeq Genes.

4.2 Nucleosome Level

The nucleosome signal profiles in CD4+ resting and activated cells were obtained from the literature [16], which were generated by Solexa (accession no. GSE10437). The nucleosome data of GM12878 cells were generated by Snyder lab and were downloaded from the UCSC ftp server. The short reads was extended to 150 bp from 5′ to 3′, and we selected 75 bp in the middle to calculate the nucleosome occupancy aligned with TSSs and TTSs. The nucleosome positioning patterns that were predicted by sequences were predicted using the software developed by Segal lab [50], and the DNA sequences in the promoters were extracted from whole genome sequence which was obtained from UCSC.

Because the nucleosome data of GM12878 cells were mapped with hg19 while other data were mapped with hg18, we used the liftOver tool, which was downloaded from UCSC, to transform the genome coordinates of GM12878 cells data from hg19 to hg18.

4.3 P Values between Intronless Genes and Intron-Containing Genes

We calculated the *P* values in variant regions of intronless genes and intron-containing genes. The average value of nucleosomes, which referred to the reads number in the region divided by the length of the region, indicated the nucleosome occupancy value in the region. All of the average values in the intronless group and intron-containing group were calculated. Finally, we compared these two groups and obtained *P* values. The promoters included the region 0-1,000 bp upstream of the TSSs. The gene bodies included the region 0-1,000 bp downstream of the TSSs.

Downstream of the TTSs included the region 0-1,000 bp downstream of the TTSs. The methylation and predicted nucleosome occupancy was also calculated in the same way.

4.4 CpG Islands

The distributions of CpG islands in the human genome were obtained from the UCSC human genome database. A gene was regarded as a CpG island containing genes when the TSS was in a CpG island.

4.5 Poly(dA:dT)

Promoters in the region of –1,000 ~2,000 bp were scanned for continuous motifs of poly(dA:dT) tracks with a size of 6 bp.

4.6 RNA Pol II

The RNA Pol II data in the CD4+ T resting cells were obtained from Schones et al. [16] (accession no. GSE10437). The RNA Pol II data of GM12878 generated by Snyder lab were downloaded from the UCSC ftp server. A sliding window of 5 bp was used around TSSs, and all reads originated in these windows were calculated. We used a scheme described in the reference [16] to identify the activity of RNA Pol II. We calculated the Pol II levels as the sum of all Pol II signals in a 1 kilobase (kb) region surrounding the TSS and calculated the average Pol II level (APL) for 1 Kb window through the gene body. A stalling index (SI) was then defined as the ratio of the promoter Pol II level over the average gene body level. Genes with stalled Pol II were defined as those with $SI > 10$ and $APL < 5$. Genes with elongating Pol II were defined as those with $1 < SI < 10$ and $APL > 5$. Genes with no promoter Pol II were defined as those with $APL = 0$.

4.7 DNA Methylation

The DNA methylation data in the resting CD4+ T cell lines were obtained from Choi et al. [70]. Sliding windows of 5 bp were applied in the regions of TSSs and all reads that originated in these windows were calculated. Total counts were divided by the number of genes in each set.

4.8 Analysis of Gene Expression Levels Using mRNA-seq Data

The RNA-Seq data from CD4+ T cells were obtained from I. Chepelev's work [71] (accession no. GSE16190), and the RNA-Seq data of GM12878 were generated by Gingeras-CSHL [71], [72], which were downloaded from UCSC ftp server. The expression level for each gene was quantified in reads per kilobase of exon model per million mapped sequence reads (RPKM) [73]. All genes in the cells were categorized to three groups: no expressed genes, lowly expressed genes and highly expressed genes according to their expression levels. No expressed genes referred to the genes whose RPKM was 0. Rest genes were averaged to two groups by RMKM, 50 percent of the genes with lower RPKM were defined as lowly expressed genes and 50 percent of the genes with higher RPKM were defined as highly expressed genes.

Acknowledgments

This work was supported by the National Basic Research Program of China (2012CB316501) and the National Nature Science Foundation of China (61472078 and 61073141). Xiao Sun is the corresponding author of this work. Xiangfei Cheng and Yue Hou contributed equally to the work.

References

[1] R. D. Kornberg and Y. Lorch, "Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome," *Cell*, vol. 98, no. 3, pp. 285–294, Aug. 6, 1999.

[2] E. Segal, Y. Fondufe-Mittendorf, L. Chen, A. Thastrom, Y. Field, I. K. Moore, J. P. Wang, and J. Widom, "A genomic code for nucleosome positioning," *Nature*, vol. 442, no. 7104, pp. 772–778, Aug. 17, 2006.

[3] H. Liu, X. Duan, S. Yu, and X. Sun, "Analysis of nucleosome positioning determined by DNA helix curvature in the human genome," *BMC Genomics*, vol. 12, p. 72, 2011.

[4] Y. Nie, X. Cheng, J. Chen, and X. Sun, "Nucleosome organization in the vicinity of transcription factor binding sites in the human genome," *BMC Genomics*, vol. 15, p. 493, 2014.

[5] Y. Nie, H. Liu, and X. Sun, "The patterns of histone modifications in the vicinity of transcription factor binding sites in human lymphoblastoid cell lines," *PLoS One*, vol. 8, no. 3, p. e60002, 2013.

[6] M. A. Schwabish and K. Struhl, "Evidence for eviction and rapid deposition of histones upon transcriptional elongation by RNA polymerase II," *Mol. Cell Biol.*, vol. 24, no. 23, pp. 10111–10117, Dec 2004.

[7] E. Segal and J. Widom, "What controls nucleosome positions?" *Trends Genet*, vol. 25, no. 8, pp. 335–343, Aug. 2009.

[8] D. J. Gaffney, G. McVicker, A. A. Pai, Y. N. Fondufe-Mittendorf, N. Lewellen, K. Michelini, J. Widom, Y. Gilad, and J. K. Pritchard, "Controls of nucleosome positioning in the human genome," *PLoS Genet.*, vol. 8, no. 11, p. e1003036, 2012.

[9] G. Arya, A. Maitra, and S. A. Grigoryev, "A structural perspective on the where, how, why, and what of nucleosome positioning," *J. Biomol. Struct. Dyn.*, vol. 27, no. 6, pp. 803–820, Jun. 2010.

[10] A. Valouev, S. M. Johnson, S. D. Boyd, C. L. Smith, A. Z. Fire, and A. Sidow, "Determinants of nucleosome organization in primary human cells," *Nature*, vol. 474, no. 7352, pp. 516–520, Jun. 23, 2011.

[11] C. K. Collings, P. J. Waddell, and J. N. Anderson, "Effects of DNA methylation on nucleosome stability," *Nucleic Acids Res.*, vol. 41, no. 5, pp. 2918–2931, Mar. 1, 2013.

[12] J. S. Choy, S. Wei, J. Y. Lee, S. Tan, S. Chu, and T. H. Lee, "DNA methylation increases nucleosome compaction and rigidity," *J. Am. Chem. Soc.*, vol. 132, no. 6, pp. 1782–1783, Feb. 17, 2010.

[13] M. Dadiani, D. van Dijk, B. Segal, Y. Field, G. Ben-Artzi, T. Raveh-Sadka, M. Levo, I. Kaplow, A. Weinberger, and E. Segal, "Two DNA-encoded strategies for increasing expression with opposing effects on promoter dynamics and transcriptional noise," *Genome Res.*, vol. 23, no. 6, pp. 966–976, Jun. 2013.

[14] K. Luger, M. L. Dechassa, and D. J. Tremethick, "New insights into nucleosome and chromatin structure: An ordered state or a disordered affair?" *Nat. Rev. Mol. Cell. Biol.*, vol. 13, no. 7, pp.436–447, Jul. 2012.

[15] J. B. Zaugg and N. M. Luscombe, "A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast," *Genome Res.*, vol. 22, no. 1, pp. 84–94, Jan. 2012.

[16] D. E. Schones, K. Cui, S. Cuddapah, T. Y. Roh, A. Barski, Z. Wang, G. Wei, and K. Zhao, "Dynamic regulation of nucleosome positioning in the human genome," *Cell*, vol. 132, no. 5, pp. 887–898, Mar. 7, 2008.

[17] K. Brogaard, L. Xi, J. P. Wang, and J. Widom, "A map of nucleosome positions in yeast at base-pair resolution," *Nature*, vol. 486, no. 7404, pp. 496–501, Jun. 28, 2012.

[18] G. C. Yuan, Y. J. Liu, M. F. Dion, M. D. Slack, L. F. Wu, S. J. Altschuler, and O. J. Rando, "Genome-scale identification of nucleosome positions in S. cerevisiae," *Science*, vol. 309, no. 5734, pp. 626–630, Jul. 22, 2005.

[19] T. K. Kelly, Y. Liu, F. D. Lay, G. Liang, B. P. Berman, and P. A. Jones, "Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules," *Genome Res.*, vol. 22, no. 12, pp. 2497–2506, Dec. 2012.

[20] H. Huang, H. Liu, and X. Sun, "Nucleosome distribution near the 3′ ends of genes in the human genome," *Biosci. Biotechnol. Biochem.*, vol. 77, no. 10, pp. 2051–2055, 2013.

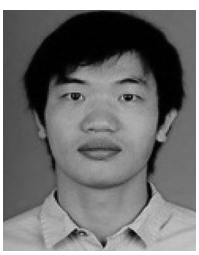
[21] H. E. Mischo and N. J. Proudfoot, "Disengaging polymerase: Terminating RNA polymerase II transcription in budding yeast," *Biochim. Biophys. Acta*, vol. 1829, no. 1, pp. 174–185, Jan. 2013.

[22] A. R. Grosso, S. F. de Almeida, J. Braga, and M. Carmo-Fonseca, "Dynamic transitions in RNA polymerase II density profiles during transcription termination," *Genome Res.*, vol. 22, no. 8, pp. 1447–1456, Aug. 2012.

[23] S. C. Murray, A. Serra Barros, D. A. Brown, P. Dudek, J. Ayling, and J. Mellor, "A pre-initiation complex at the 3'-end of genes drives antisense transcription independent of divergent sense transcription," *Nucleic Acids Res.*, vol. 40, no. 6, pp. 2432–2444, Mar. 2012.
[24] A. Weiner, A. Hughes, M. Yassour, O. J. Rando, and N. Friedman, "High-resolution nucleosome mapping reveals transcription-dependent promoter packaging," *Genome Res.*, vol. 20, no. 1, pp. 90–100, Jan. 2010.
[25] R. Andersson, S. Enroth, A. Rada-Iglesias, C. Wadelius, and J. Komorowski, "Nucleosomes are well positioned in exons and carry characteristic histone modifications," *Genome Res.*, vol. 19, no. 10, pp. 1732–1741, Oct. 2009.
[26] H. Huang, S. Yu, H. Liu, and X. Sun, "Nucleosome organization in sequences of alternative events in human genome," *Biosystems*, vol. 109, no. 2, pp. 214–219, Aug. 2012.
[27] E. A. Grzybowska, "Human intronless genes: functional groups, associated diseases, evolution, and mRNA processing in absence of splicing," *Biochem. Biophys. Res. Commun.*, vol. 424, no. 1, pp. 1–6, Jul. 20, 2012.
[28] H. Le Hir, A. Nott, and M. J. Moore, "How introns influence and enhance eukaryotic gene expression," *Trends Biochem. Sci.*, vol. 28, no. 4, pp. 215–220, Apr. 2003.
[29] D. K. Niu and Y. F. Yang, "Why eukaryotic cells use introns to enhance gene expression: Splicing reduces transcription-associated mutagenesis by inhibiting topoisomerase I cutting activity," *Biol. Direct*, vol. 6, p. 24, 2011.
[30] R. L. Brinster, J. M. Allen, R. R. Behringer, R. E. Gelinas, and R. D. Palmiter, "Introns increase transcriptional efficiency in transgenic mice," in *Proc. Nat. Acad. Sci. USA*, vol. 85, no. 3, pp. 836–840, Feb. 1988.
[31] A. Louhichi, A. Fourati, and A. Rebai, "IGD: A resource for intronless genes in the human genome," *Gene*, vol. 488, nos. 1/2, pp. 35–40, Nov. 15, 2011.
[32] M. K. Sakharkar and K. R. Sakharkar, "A tale of intronless genes in eukaryotic genomes," in *Proc. Int. Conf. Biomed. Pharmaceutical Eng.*, 2006, pp. 461–466.
[33] S. A. Shabalina, A. Y. Ogurtsov, A. N. Spiridonov, P. S. Novichkov, N. A. Spiridonov, and E. V. Koonin, "Distinct patterns of expression and evolution of intronless and intron-containing mammalian genes," *Mol. Biol. Evol.*, vol. 27, no. 8, pp. 1745–1749, Aug. 2010.
[34] H. Lei, A. P. Dias, and R. Reed, "Export and stability of naturally intronless mRNAs require specific coding region sequences and the TREX mRNA export complex," in *Proc. Nat. Acad. Sci. USA*, vol. 108, no. 44, pp. 17985–17990, Nov. 1, 2011.
[35] H. Lei, B. Zhai, S. Yin, S. Gygi, and R. Reed, "Evidence that a consensus element found in naturally intronless mRNAs promotes mRNA export," *Nucleic Acids Res.*, vol. 41, no. 4, pp. 2517–2525, Feb. 1, 2013.
[36] W. Mobius and U. Gerland, "Quantitative test of the barrier nucleosome model for statistical positioning of nucleosomes up- and downstream of transcription start sites," *PLoS Comput. Biol.*, vol. 6, no. 8, p. e1000891, 2010.
[37] V. R. Iyer, "Nucleosome positioning: bringing order to the eukaryotic genome," *Trends Cell Biol.*, vol. 22, no. 5, pp. 250–256, May 2012.
[38] J. K. Choi, "Contrasting chromatin organization of CpG islands and exons in the human genome," *Genome Biol.*, vol. 11, no. 7, p. R70, 2010.
[39] A. M. Deaton and A. Bird, "CpG islands and the regulation of transcription," *Genes Develop.*, vol. 25, no. 10, pp. 1010–1022, May 15, 2011.
[40] E. Segal and J. Widom, "Poly(dA:dT) tracts: Major determinants of nucleosome organization," *Curr. Opin. Struct. Biol.*, vol. 19, no. 1, pp. 65–71, Feb. 2009.
[41] A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao, "High-resolution profiling of histone methylations in the human genome," *Cell*, vol. 129, no. 4, pp. 823–837, May 18, 2007.
[42] R. H. Jacobson, A. G. Ladurner, D. S. King, and R. Tjian, "Structure and function of a human TAFII250 double bromodomain module," *Science*, vol. 288, no. 5470, pp. 1422–1425, May 26, 2000.
[43] O. Matangkasombut, R. M. Buratowski, N. W. Swilling, and S. Buratowski, "Bromodomain factor 1 corresponds to a missing piece of yeast TFIID," *Genes Develop.*, vol. 14, no. 8, pp. 951–962, Apr. 15, 2000.
[44] M. Vermeulen, K. W. Mulder, S. Denissov, W. W. Pijnappel, F. M. van Schaik, R. A. Varier, M. P. Baltissen, H. G. Stunnenberg, M. Mann, and H. T. Timmers, "Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4," *Cell*, vol. 131, no. 1, pp. 58–69, Oct. 5, 2007.
[45] R. D. Kornberg, "Mediator and the mechanism of transcriptional activation," *Trends Biochem. Sci.*, vol. 30, no. 5, pp. 235–239, May 2005.
[46] D. A. Bushnell, P. Cramer, and R. D. Kornberg, "Structural basis of transcription: Alpha-amanitin-RNA polymerase II cocrystal at 2.8 A resolution," in *Proc. Nat. Acad. Sci. USA*, vol. 99, no. 3, pp. 1218–1222, Feb. 5, 2002.
[47] Y. Huang, K. M. Wimler, and G. G. Carmichael, "Intronless mRNA transport elements may affect multiple steps of pre-mRNA processing," *EMBO J.*, vol. 18, no. 6, pp. 1642–1652, Mar. 15, 1999.
[48] W. F. Marzluff, E. J. Wagner, and R. J. Duronio, "Metabolism and regulation of canonical histone mRNAs: Life without a poly(A) tail," *Nat. Rev. Genet.*, vol. 9, no. 11, pp. 843–854, Nov. 2008.
[49] N. I. Bieberstein, F. Carrillo Oesterreich, K. Straube, and K. M. Neugebauer, "First exon length controls active chromatin signatures and transcription," *Cell Rep*, vol. 2, no. 1, pp. 62–68, Jul. 26, 2012.
[50] N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, D. Tillo, Y. Field, E. M. LeProust, T. R. Hughes, J. D. Lieb, J. Widom, and E. Segal, "The DNA-encoded nucleosome organization of a eukaryotic genome," *Nature*, vol. 458, no. 7236, pp. 362–366, Mar. 19, 2009.
[51] D. Tillo and T. R. Hughes, "G+C content dominates intrinsic nucleosome occupancy," *BMC Bioinf.*, vol. 10, p. 442, 2009.
[52] J. Y. Lee and T. H. Lee, "Effects of DNA methylation on the structure of nucleosomes," *J. Am. Chem. Soc.*, vol. 134, no. 1, pp. 173–175, Jan. 11, 2012.
[53] K. Struhl and E. Segal, "Determinants of nucleosome positioning," *Nat. Struct. Mol. Biol.*, vol. 20, no. 3, pp. 267–73, Mar. 2013.
[54] T. N. Mavrich, I. P. Ioshikhes, B. J. Venters, C. Jiang, L. P. Tomsho, J. Qi, S. C. Schuster, I. Albert, and B. F. Pugh, "A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome," *Genome Res.*, vol. 18, no. 7, pp. 1073–1083, Jul. 2008.
[55] W. Chen, Y. Liu, S. Zhu, C. D. Green, G. Wei, and J. D. Han, "Improved nucleosome-positioning algorithm iNPS for accurate nucleosome positioning from sequencing data," *Nat. Commun.*, vol. 5, p. 4909, 2014.
[56] C. Vaillant, L. Palmeira, G. Chevereau, B. Audit, Y. d'Aubenton-Carafa, C. Thermes, and A. Arneodo, "A novel strategy of transcription regulation by intragenic nucleosome ordering," *Genome Res.*, vol. 20, no. 1, pp. 59–67, Jan. 2010.
[57] K. Juneau, M. Miranda, M. E. Hillenmeyer, C. Nislow, and R. W. Davis, "Introns regulate RNA and protein abundance in yeast," *Genetics*, vol. 174, no. 1, pp. 511–518, Sep. 2006.
[58] R. W. McKenzie and M. D. Brennan, "The two small introns of the Drosophila affinidisjuncta Adh gene are required for normal transcription," *Nucleic Acids Res.*, vol. 24, no. 18, pp. 3635–3642, Sep. 15, 1996.
[59] M. Kanehisa and S. Goto, "KEGG: Kyoto encyclopedia of genes and genomes," *Nucleic Acids Res.*, vol. 28, no. 1, pp. 27–30, Jan. 1, 2000.
[60] G. Dennis Jr., B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, and R. A. Lempicki, "DAVID: Database for Annotation, Visualization, and Integrated Discovery," *Genome Biol.*, vol. 4, no. 5, p. P3, 2003.
[61] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium," *Nat. Genet.*, vol. 25, no. 1, pp. 25–29, May 2000.
[62] J. Zeitlinger, A. Stark, M. Kellis, J. W. Hong, S. Nechaev, K. Adelman, M. Levine, and R. A. Young, "RNA polymerase stalling at developmental control genes in the Drosophila melanogaster embryo," *Nat. Genet.*, vol. 39, no. 12, pp. 1512–1516, Dec. 2007.
[63] H. F. Wang, L. Feng, and D. K. Niu, "Relationship between mRNA stability and intron presence," *Biochem. Biophys. Res. Commun.*, vol. 354, no. 1, pp. 203–208, Mar. 2, 2007.

[64] A. B. Rose, T. Elfersi, G. Parra, and I. Korf, "Promoter-proximal introns in Arabidopsis thaliana are enriched in dispersed signals that elevate gene expression," *Plant Cell*, vol. 20, no. 3, pp. 543–551, Mar, 2008.
[65] D. Tillo, N. Kaplan, I. K. Moore, Y. Fondufe-Mittendorf, A. J. Gossett, Y. Field, J. D. Lieb, J. Widom, E. Segal, and T. R. Hughes, "High nucleosome occupancy is encoded at human regulatory sequences," *PLoS One*, vol. 5, no. 2, p. e9129, 2010.
[66] H. H. He, C. A. Meyer, H. Shin, S. T. Bailey, G. Wei, Q. Wang, Y. Zhang, K. Xu, M. Ni, M. Lupien, P. Mieczkowski, J. D. Lieb, K. Zhao, M. Brown, and X. S. Liu, "Nucleosome dynamics define transcriptional enhancers," *Nat. Genet.*, vol. 42, no. 4, pp. 343–347, Apr. 2010.
[67] K. S. Zaret and J. S. Carroll, "Pioneer transcription factors: Establishing competence for gene expression," *Genes Dev*, vol. 25, no. 21, pp. 2227–2241, Nov. 1, 2011.
[68] H. Huang, J. Chen, H. Liu, and X. Sun, "The nucleosome regulates the usage of polyadenylation sites in the human genome," *BMC Genomics*, vol. 14, p. 912, 2013.
[69] K. D. Pruitt, T. Tatusova, and D. R. Maglott, "NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins," *Nucleic Acids Res.*, vol. 35, no. Database issue, pp. D61–5, Jan. 2007.
[70] J. K. Choi, J. B. Bae, J. Lyu, T. Y. Kim, and Y. J. Kim, "Nucleosome deposition and DNA methylation at coding region boundaries," *Genome Biol*, vol. 10, no. 9, p. R89, 2009.
[71] I. Chepelev, G. Wei, Q. Tang, and K. Zhao, "Detection of single nucleotide variations in expressed exons of the human genome using RNA-Seq," *Nucleic Acids Res.*, vol. 37, no. 16, p. e106, Sep. 2009.
[72] L. Jiang, F. Schlesinger, C. A. Davis, Y. Zhang, R. Li, M. Salit, T. R. Gingeras, and B. Oliver, "Synthetic spike-in standards for RNA-seq experiments," *Genome Res*, vol. 21, no. 9, pp. 1543–1551, Sep. 2011.
[73] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold, "Mapping and quantifying mammalian transcriptomes by RNA-Seq," *Nat. Methods*, vol. 5, no. 7, pp. 621–628, Jul. 2008.

Xiangfei Cheng received the master's degree from the School of Biological Science and Medical Engineering, Southeast University. Her main research interests include epigenetics and computational biology.



Yue Hou received the master's degree from the School of Biological Science and Medical Engineering from Southeast University. He is currently working toward the PhD degree in Southeast University. His main research interests include epigenetics and computational biology.

Yumin Nie received the PhD degree from the School of Biological Science and Medical Engineering from Southeast University. Her main research interests focus on DNA methylation. She published four papers in journals and conferences.

Yiru Zhang received the master's degree from the School of Biological Science and Medical Engineering, Southeast University.

Huan Huang received the PhD degree from the School of Biological Science and Medical Engineering, Southeast University. Her research interests mainly focus on epigenetics. She published four papers in journals and conferences.

Hongde Liu received the PhD degree in biomedical engineering in 2006 from Northwest Normal University. He is currently an associate professor of biological science and medical engineering in Southeast University. His research focuses on nucleosome and gene regulation. He published more than 30 papers in journals and conferences in these areas.

Xiao Sun received the BS degree in computer science in 1984 and the PhD degree in biomedical engineering in 1993 from Southeast University. Since 2001, he has been a professor in the School of Biological Science and Medical Engineering, Southeast University. His major research interests include bioinformatics and biomedical signal processing. He has been in charge of and completed many projects which were funded by the Natural Science Foundation of China, National High-tech R&D Program (863 Program), or National key Basic Research Program of China (973 program). He is the author or corresponding author of more than 80 papers. He also published two books *Fundamentals of Bioinformatics* and *R & BioConductor and Their Applications in Genome Analysis.*

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.