Biochemical and Biophysical Research Communications 391 (2010) 884–889

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

A novel mechanism of epigenetic regulation: Nucleosome-space occupancy

Peng Cui [a,1], Lingfang Zhang [a,b,1], Qiang Lin [a,b,1], Feng Ding [a], Chengqi Xin [a,b], Xiangdong Fang [a], Songnian Hu [a,*], Jun Yu [a,*]

[a] The CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, 100029 Beijing, China
[b] Graduate School of the Chinese Academy of Sciences, 100029 Beijing, China

ARTICLE INFO

Article history:
Received 19 November 2009
Available online 3 December 2009

Keywords:
Nucleosome
Histone modification
Epigenetic regulation

ABSTRACT

Nucleosome positioning around the gene space (or transcriptional unit) plays a crucial role for gene regulation but we do not know if the spatial organization—nucleosome-space occupancy or nucleosome density in a defined sequence unit length—contributes to the regulation complexity of mammalian gene expression. Using our own rmRNA-Seq (ribosomal RNA-minus RNA sequencing) and publically available ChIP-Seq (H3) data from mouse stem cells, we discovered a non-random distribution of nucleosomes along chromosomes, and further genome-wide studies on histone modifications, DNA methylation, transcriptional activity, gene density, and base compositional dynamics, demonstrated that nucleosome-space occupancy of genomic regions—clustered genes and their intergenic spaces—show distinctive features, where a high occupancy coincides with active transcription, intensive histone modifications, poor DNA methylation, and higher GC contents as compared to the nucleosome-poor regions. We therefore proposed that nucleosome-space occupancy as a novel mechanism of epigenetic gene regulation, creating a vital environment for transcriptional activation.

Introduction

Nucleosome, as a basic building unit of eukaryotic chromatins, not only plays a structural role, but also participates in regulating transcription through its deposition over nucleosome-space (DNA sequence where nucleosomes are bound to) and the chemical modification of its components—histones [1–3]. Genome-wide investigations of nucleosome positioning demonstrated that nucleosomes are regularly arranged around protein-coding genes from the transcriptional start site (TSS) to transcriptional termination site (TTS), and free in the regions flanking the genes [4]. The nucleosome predominantly resides the upstream of TSS and regulates the accessibility of regulatory elements (such as promoters) through histone modifications (such as H3K4me3 and H3K36me3 for active genes, and H3K27me3, H3K9me2, and H3K9me3 for inactive genes) [5,6]. In the interior of genes, nucleosomes strongly prefer to occupy exon starts, suggesting a potential role in splicing [4,7]. These noticeable organizational patterns provide clues into mechanistic principles of nucleosome-related gene regulations.

Recent studies have been focused on nucleosome positioning around genes [4,8,9]; however, knowledge on how nucleosomes are organized locally across genes beyond individual transcriptional units and to what extents nucleosome-space are occupied among actively transcribed genes remains to be elucidated. We used our own rmRNA-Seq data in combination with other publicly available ChIP-Seq (H3) data to conduct a survey for conditional occupancy of the nucleosome-space at whole-genome level, and identified nucleosome-rich regions among mouse ESCs. Combining data from other genome-wide studies for histone modifications, DNA methylation, transcriptional activity, gene density, and base compositional dynamics, we discovered that nucleosome-rich regions possess distinctive features, including active transcription, intensive histone modifications, poor DNA methylation, higher GC content, and higher gene density as compared to nucleosome-poor regions. These analyses suggested that nucleosome-space occupancy in general may also play a pivotal role in regulating gene expression.

Materials and methods

Data sources. We generated rmRNA-Seq (ribosomal RNA-minus) data from S129 mouse stem cell, using SOLiD (Life Technologies), and other data are from public databases. Nucleosome (pan-H3), histone modification, transcriptome, and DNA methylation data for mouse stem cell [10,11] are from public database NCBI. Other expression sequence tags were mapped to the mouse genome using the SOLiD analysis pipeline [12] and annotated based on Refseq [13]. We used a Poisson background model to evaluate

* Corresponding authors. Address: Beijing Institute of Genomics, Chinese Academy of Sciences, No. 7 Beitucheng West Road, Chaoyang, Beijing, China. Fax: +86 10 82995362 (S. Hu), +86 10 82995357 (J. Yu).
E-mail addresses: husn@big.ac.cn (S. Hu), junyu@big.ac.cn (J. Yu).
[1] These authors contributed equally to this work.

0006-291X/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2009.11.157

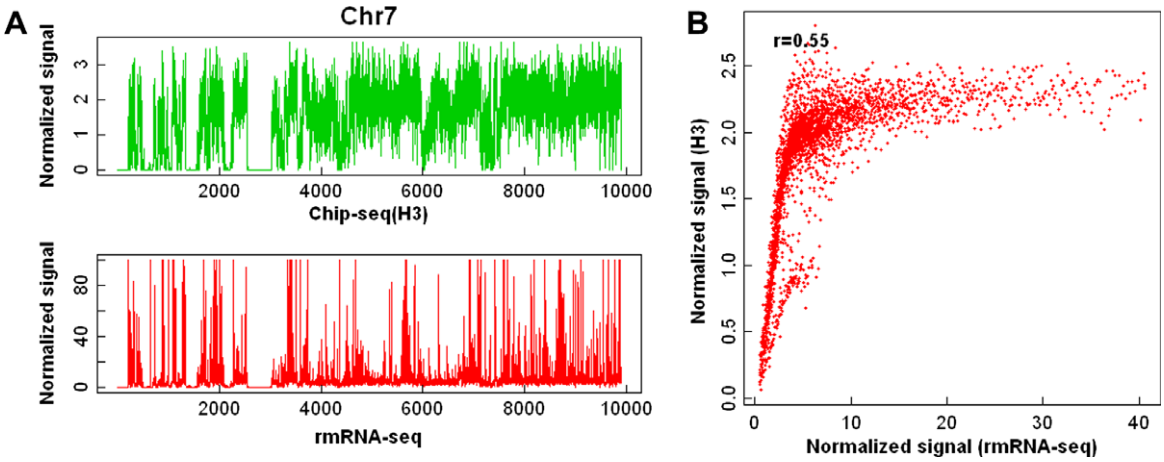

Fig. 1. Profiles of nucleosome and transcription along chromosome 7 and their correlation analysis (A). The whole chromosome is equally divided into 10,000 windows (resolution = 1 e−5); in each window, tag density of ChIP-Seq (pan-H3) and rmRNA-Seq, separately indicating the level of nucleosome and transcription, are determined and plotted along the whole chromosome. (B) Firstly, all windows were grouped to 5000 window sets according to expression level. There is a good correlation ($P < 0.01$) when each chromosome is equally divided into 10000 windows (resolution = 1 e−4). The *x* axis indicates the transcriptional level, while the *y* axis indicates nucleosome level.

whether a gene is active or inactive [14], and obtained 6047 inactive and 12,994 active genes. To assess the expression level of a gene, we counted its mapped sequence tags and normalized over its transcript length. We classified the Refseq-defined genes into the HCP, LCP, and ICP genes according to a published method [15].

Correlation analyses. We divided chromosome into numbers of windows based on the defined resolution (resolution = 1/numbers of windows), and counted number of tags in each window as "Tag Density" or the signal of nucleosomes, transcriptional activity, RNAPII, histone modifications or DNA methylation. Then, we

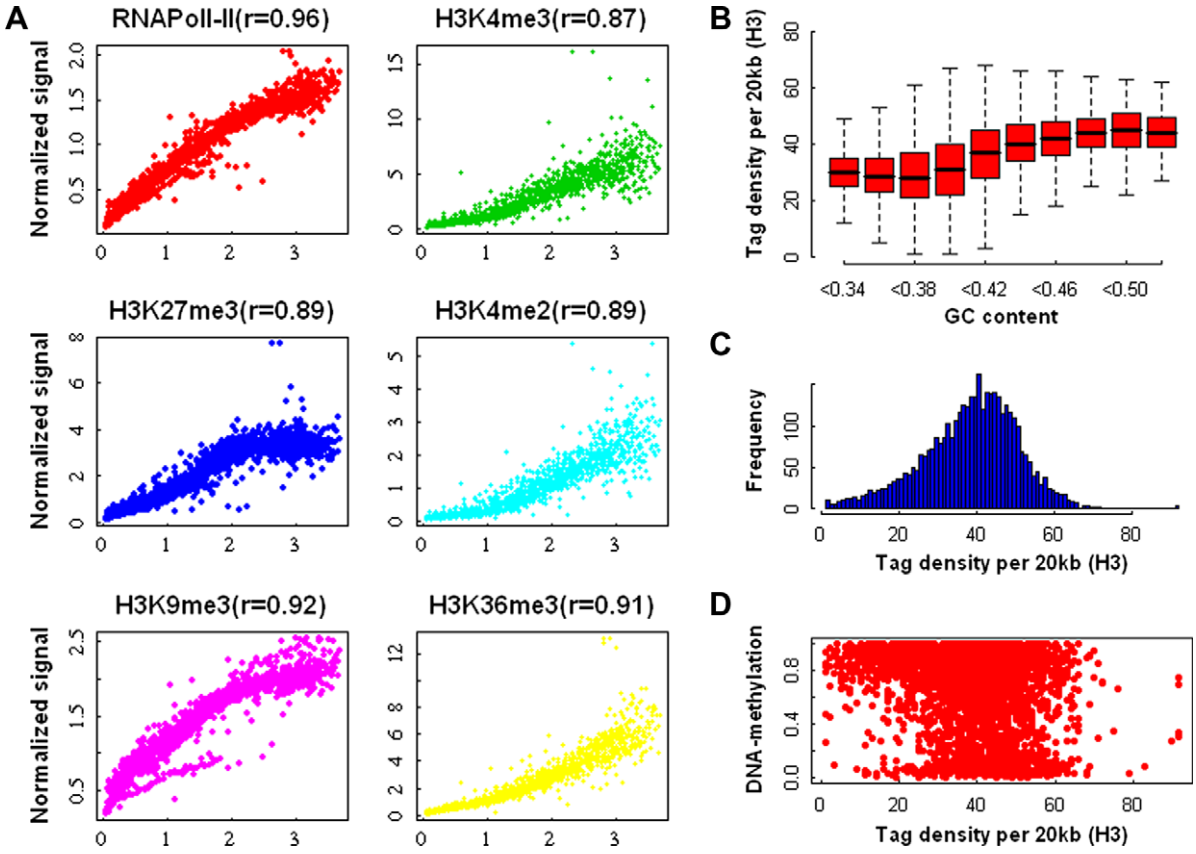

Fig. 2. Correlation between nucleosome-space occupancy and RNAPII binding signals, histone modifications, DNA methylation, GC content. (A) Windows were grouped to 5000 gene window sets according to nucleosome enrichment. The *x* axis indicates the nucleosome level, while the *y* axis indicates histone modification level. (B) The nucleosome density was box-plotted according to GC content. Firstly, each chromosome was equally divided into 10,000 window; the nucleosome density and GC content were calculated in each window. The GC content is divided into several intervals indicated on the *x* axis, and the corresponding nucleosome density was box-plotted. (C) We investigated the correlation between the nucleosome density and DNA methylation level. Firstly, each chromosome is equally divided into 10,000 windows, and the density of nucleosome was assessed in each window. Then, we investigated the distribution of nucleosome density, and found most genomic windows have a moderate nucleosome density. (D) The mean methylation level of each window is calculated and plotted against nucleosome density. The nucleosome-poor region show highly methylated.

performed Pearson correlation test to evaluate the correlation between nucleosome signals and transcriptional activity, RNAPII binding, histone modifications and DNA methylation, and used hierarchical clustering strategy for the analysis of genomic, epigenetic, and transcriptional features. All statistical analyses and drawing were performed using R tools [16].

Nucleosome positioning around TSS, exon–intron junction, and TTS. We categorized genes into nucleosome-rich and nucleosome-poor clusters based on their expression levels, aligned transcript-specific positions (TSS, internal exon–intron junction by eliminating first and last exons, and TTS) in a ±1-kb window, and counted the tags in a 10-bp window.

Results

Actively transcribed genes have higher occupancy of their nucleosome-spaces

To investigate the nucleosome distribution along chromosomes, we acquired rmRNA-Seq data and also used the publicly available ChIP-Seq data (H3) from the mouse ESCs [10]. We analyzed nucleosome signals as the number of tags in a 20-kb window at a resolution of 0.0001 along chromosomes (Figs. 1A and S1) and clearly defined nucleosome-space occupancy as nucleosome-rich or nucleosome-poor regions as if nucleosomes preferentially bind to certain genomic regions, leading to a non-random distribution. In the interior space of the nucleosome-rich regions, there is an approximately Gaussian (normal) distribution of nucleosome-space occupancy across their genic and intergenic sequences (an example is shown in Fig. S2). At a gene locus (*lin28*), the nucleosome phasing can be clearly observed, and it extends to the next neighboring gene (*aim* and *cd52*) across an intergenic sequence. Therefore, nucleosome enrichment seems occurring at relatively larger genomic segments, and is not limited to individual genes. This observation suggests that nucleosome-space occupancy is beyond the organization of individual genes. We further defined transcriptional signals along each chromosome at the same resolution into "actively transcribed chromosomal blocks" or ACB to show that they are associated with so-called nucleosome-rich genomic regions (Figs. 1A, and S1). Our correlation analysis demonstrated that their transcriptional activities (at a resolution of 0.0001 in a 20-kb window) are associated with a high density (the number of nucleosomes over a defined space unit other than the nucleosome coverage over a defined binding site) of nucleosome-space occupancy (Fig. 1B).

Nucleosome-space occupancy is correlated with epigenetic mechanisms including histone modifications and DNA methylation

The active ACB (Fig. S1) not only has a strong correlation with RNAPII binding signals (Fig. 2A) but also with higher GC content (Fig. 2B), higher gene density, and open chromatin structure that has been well described in recent studies [17,18]. We used the publicly available histone modification data (H3K27me3, H3K4me3, H3K9me3, H3K36me3, and H3K4me2) from mouse ESCs to examine the nucleosome binding, based on the number of tags at a resolution of 0.0001 (Fig. 2A). In general, H3K27me3 and H3K9me3 are associated with transcriptional suppression, whereas H3K4me3, H3K36me3, and H3K4me2 often indicate transcriptional activation [5,19]. Obviously, nucleosome-space occupancy provides more opportunities for histone modifications to regulate gene expression [20]. In addition, a positive correlation of H3K4 to nucleosome-space occupancy shed lights on why nucleosome enrichment may facilitate RNAPII binding, forming ACB. Since H3K4 is often associated with active promoters and RNAPII

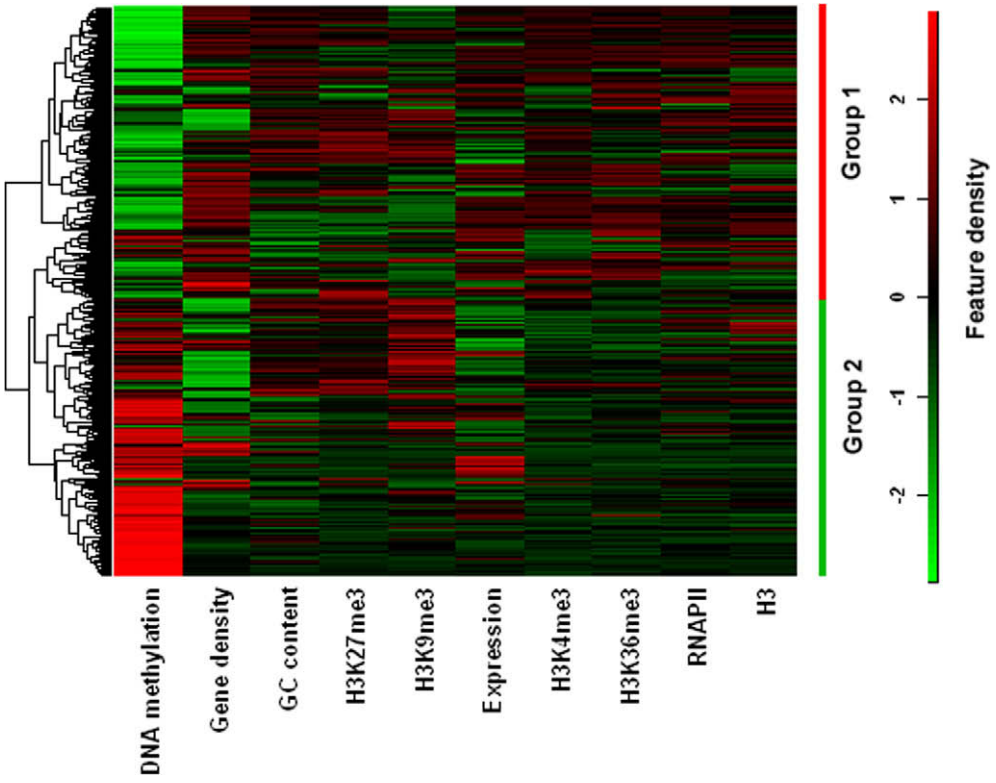

Fig. 3. Clustering of genomic, epigenetic, and transcriptional features. The signals of nucleosome occupancy, DNA methylation, RNAPII binding, histone modification (H3K27me3, H3K9me3, H3K4me3, and H3K36me3) and sequence tags from rmRNA-Seq are aligned in a 100-kb window. The gene count in each window indicates gene density. Criteria about a gene-centric grouping scheme (Group1 and Group2) are discussed in the main text.

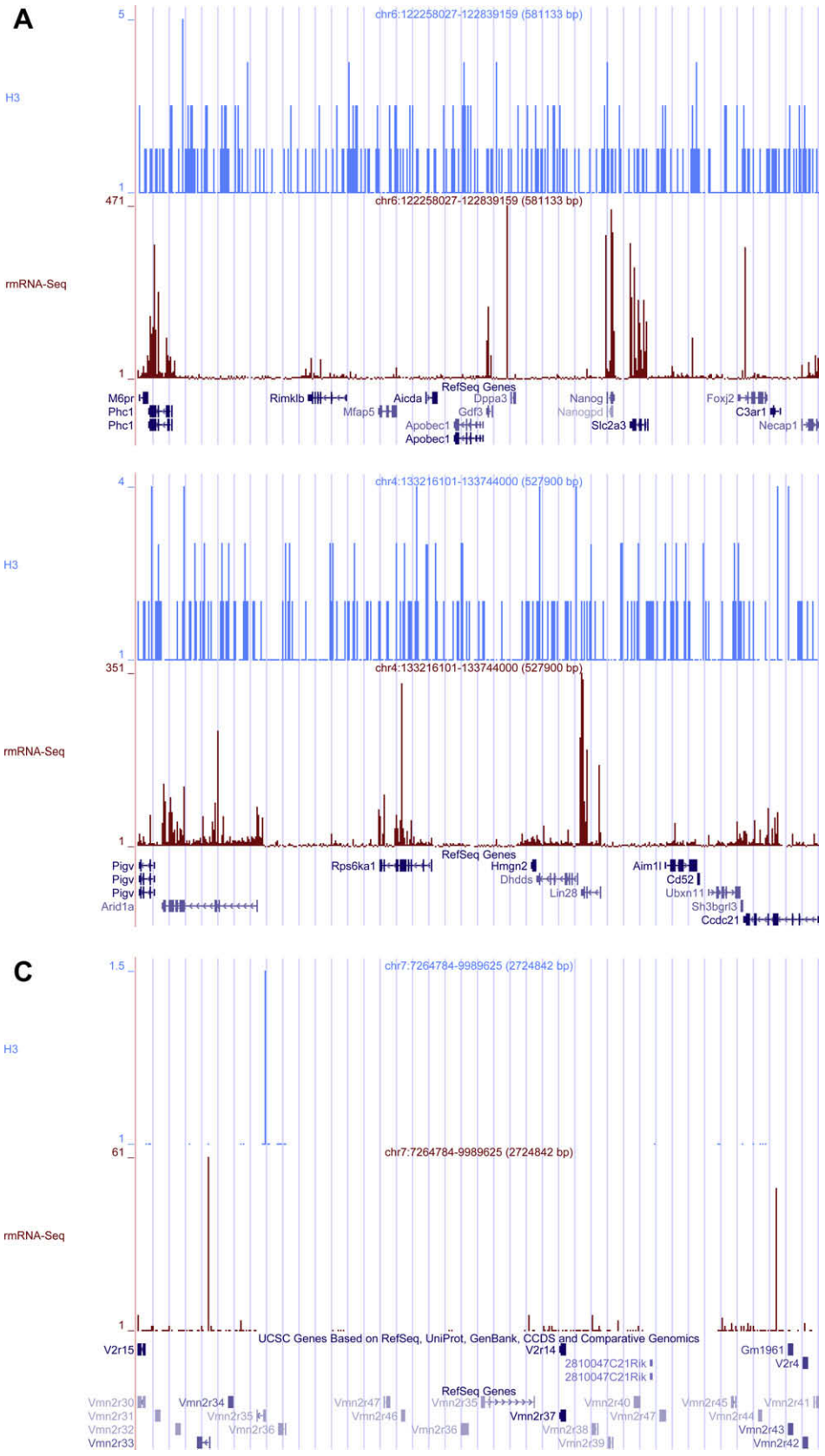

Fig. 4. Elevated and reduced nucleosome signals in selected HCP and LCP gene loci. The two actively transcribed genomic regions around two HCP genes, *nanog* (A) and *lin28* (B), exhibiting "nucleosomal islands" phenomenon in contrast to weaker nucleosome signals of the inactive genomic regions around a LCP gene *vmn2r*. The *x* axis shows the genomic positions of relevant genes and the *y* axis indicates normalized nucleosome and transcription signals.

binding [21], we considered that intensive RNAPII binding may be a consequence of H3K4 enrichment within nucleosome-rich regions. We also showed that nucleosome-space occupancy correlates well with the level of DNA methylation (Fig. 2C). As recent

reports have suggested that CpG methylation and nucleotide composition (such as GC content) both affects nucleosome formation and positioning [22,23], we believe that DNA methylation and nucleotide composition dynamics may directly or indirectly contribute to the regulation of nucleosome remodeling.

We have further performed a hierarchical clustering of the relevant genomic, epigenetic, and transcriptional features, and were able to divide two broad categories of nucleosome-space occupancy (Fig. 3). Group 1 represents gene-rich sequences that have highly-expressed genes, harbor more nucleosomes, show intensive histone modifications, especially H3K4me3 and H3K36me3 (marking transcriptional activation), possess strong RNAPII binding, exhibit lower level of DNA methylation, and are high GC-content sequences. Group 2 genomic sequences however show the opposite trends for all these features as compared to Group1.

The nucleosome-space occupancy is highly relevant in regulating gene expression in ESCs

We identified 12,994 genes that are actively transcribed in the mouse ESCs acquired so far, of which 80% are high-CpG-density promoter (HCP) genes, and 6% are low-CpG-density promoter (LCP) genes. The HCP genes are associated with two categories: ubiquitous 'housekeeping' genes and highly regulated 'key developmental' genes [15]; both are highly expressed in mouse ESCs. The LCP genes are generally associated with tissue-specific genes showing lower transcription activity. This pattern of gene expression is consistent with the previous report [11]. We also plotted the nucleosome signals at TSS (transcription start sites), the exon–intron junctions, and TTS (transcription termination sites) of the genes. For the nucleosome-rich genes or ACB, we observed that the nucleosome signals are strong across the exon–intron junctions, and distinctively lower at the promoter or transcription starts (Fig. S3). However, these nucleosome positioning patterns are lost among the nucleosome-poor genes.

To provide insights into the role of nucleosome-space occupancy in regulating gene expression, we compared nucleosome distribution between the active and inactive genes and found that most active genes (90%) form clusters (>5 genes) along chromosomes, where nucleosomes are constantly enriched but most inactive genes, though also clustered, have poorly occupied nucleosome space (Figs. S4 and S5). This is also evident from the inspection of selected HCP and LCP gene regions (Fig. 4). For instance, the two active genomic regions around *nanog* and *lin28* genes, transcriptional factors for keeping pluripotency of ESCs [24,25], showed the "nucleosomal island" effect that is equivalent to ACB as we defined here. In contrast, the nucleosome signals are dramatically reduced in inactive genomic regions around the V2R gene families specifically expressed in the mammalian vomeronasal organ [26]. These analyses pointed to a widespread elevated nucleosome signals in the active HCP genes.

Discussion

Epigenetic control plays very important role in regulating gene expression. Over the past decade, mounting evidence has pointed to at least four levels of epigenetic gene expression regulation: chromosome architecture, nucleosome positioning, histone modification, and DNA methylation [2]. We are now adding the fifth one—nucleosome-space occupancy measured by nucleosome distribution over chromosomal segments where actively transcribed genes are clustered forming ACB. Using data from mouse ESCs, we identified positive correlations of nucleosome-space occupancy to transcriptional activity, histone modifications, and RNAII binding and we therefore hypothesize that, in a defined gene cluster, the degree of its nucleosome-space occupancy alters transcriptional activity and it works in concert with histone modifications, chromosome-architecture control elements, and other transcriptional regulating factors for regulating gene expression.

The genome-wide maps of nucleosome positioning suggest a regularly arrangement of nucleosomes around protein-coding genes (from TSS to TTS) and a poor presence in the flanking sequences [3]. In the present study, we propose nucleosome-space occupancy among gene clusters and their intergenic sequences as an important parameter in studying mechanisms of epigenetics-based gene regulation. The distinctive patterns of nucleosome-space occupancy between the active and inactive genomic regions support the idea that intensive nucleosome-space occupancy creates a widely accessible environment facilitating transcriptional activation and it is an additional mechanism for nucleosome organization over genes and their intergenic sequences aside from regulating transcriptional initiation and transcript splicing. Moreover, based on the observation that nucleosomes are irregularly positioned among nucleosome-poor genes, we predict that nucleosome-space occupancy plays a necessary role for nucleosome positioning and transcriptional gene regulation, perhaps through some organizing or neighboring effects.

Accession numbers

NCBI SRA ID: SRX012691 (rmRNA-Seq data).

Acknowledgments

The study is supported by grants from the National Basic Research Program (973 Program; 2006CB910401, 2006CB910403, and 2006CB910404), the Ministry of Science and Technology of the People's Republic of China.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2009.11.157.

References

[1] Q. Lu, L.L. Wallrath, S.C. Elgin, Nucleosome positioning and gene regulation, J. Cell Biochem. 55 (1994) 83–92.
[2] B.E. Bernstein, A. Meissner, E.S. Lander, The mammalian epigenome, Cell 128 (2007) 669–681.
[3] D.E. Schones, K. Cui, S. Cuddapah, T.Y. Roh, A. Barski, Z. Wang, G. Wei, K. Zhao, Dynamic regulation of nucleosome positioning in the human genome, Cell 132 (2008) 887–898.
[4] R. Andersson, S. Enroth, A. Rada-Iglesias, C. Wadelius, J. Komorowski, Nucleosomes are well positioned in exons and carry characteristic histone modifications, Genome Res. 19 (2009) 1732–1741.
[5] A. Barski, S. Cuddapah, K. Cui, T.Y. Roh, D.E. Schones, Z. Wang, G. Wei, I. Chepelev, K. Zhao, High-resolution profiling of histone methylations in the human genome, Cell 129 (2007) 823–837.
[6] A.J. Bannister, R. Schneider, F.A. Myers, A.W. Thorne, C. Crane-Robinson, T. Kouzarides, Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes, J. Biol. Chem. 280 (2005) 17732–17736.
[7] S.M. Berget, Exon recognition in vertebrate splicing, J. Biol. Chem. 270 (1995) 2411–2414.
[8] C. Jiang, B.F. Pugh, Nucleosome positioning and gene regulation: advances through genomics, Nat. Rev. Genet. 10 (2009) 161–172.
[9] M. Shimizu, Functional structure of chromatin in gene expression and regulation—nucleosome positioning, Tanpakushitsu Kakusan Koso 38 (1993) 971–981.
[10] T.S. Mikkelsen, M. Ku, D.B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T.K. Kim, R.P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E.S. Lander, B.E. Bernstein, Genome-wide maps of chromatin state in pluripotent and lineage-committed cells, Nature 448 (2007) 553–560.
[11] A. Meissner, T.S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B.E. Bernstein, C. Nusbaum, D.B. Jaffe, A. Gnirke, R. Jaenisch, E.S. Lander, Genome-scale DNA methylation maps of pluripotent and differentiated cells, Nature 454 (2008) 766–770.

[12] N. Cloonan, A.R. Forrest, G. Kolle, B.B. Gardiner, G.J. Faulkner, M.K. Brown, D.F. Taylor, A.L. Steptoe, S. Wani, G. Bethel, A.J. Robertson, A.C. Perkins, S.J. Bruce, C.C. Lee, S.S. Ranade, H.E. Peckham, J.M. Manning, K.J. McKernan, S.M. Grimmond, Stem cell transcriptome profiling via massive-scale mRNA sequencing, Nat. Methods 5 (2008) 613–619.
[13] K.D. Pruitt, T. Tatusova, D.R. Maglott, NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic Acids Res. 35 (2007) D61–D65.
[14] L.J. Core, J.J. Waterfall, J.T. Lis, Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters, Science 322 (2008) 1845–1848.
[15] S. Saxonov, P. Berg, D.L. Brutlag, A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters, Proc. Natl. Acad. Sci. USA 103 (2006) 1412–1417.
[16] R.B. Dessau, C.B. Pipper, "R"—project for statistical computing, Ugeskr. Laeger 170 (2008) 328–330.
[17] N. Gilbert, S. Boyle, H. Fiegler, K. Woodfine, N.P. Carter, W.A. Bickmore, Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers, Cell 118 (2004) 555–566.
[18] J. Fantes, B. Redeker, M. Breen, S. Boyle, J. Brown, J. Fletcher, S. Jones, W. Bickmore, Y. Fukushima, M. Mannens, et al., Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype, Hum. Mol. Genet. 4 (1995) 415–422.
[19] C.R. Vakoc, M.M. Sachdeva, H. Wang, G.A. Blobel, Profile of histone lysine methylation across transcribed mammalian chromatin, Mol. Cell. Biol. 26 (2006) 9185–9195.
[20] T. Kouzarides, Chromatin modifications and their function, Cell 128 (2007) 693–705.
[21] N.D. Heintzman, R.K. Stuart, G. Hon, Y. Fu, C.W. Ching, R.D. Hawkins, L.O. Barrera, S. Van Calcar, C. Qu, K.A. Ching, W. Wang, Z. Weng, R.D. Green, G.E. Crawford, B. Ren, Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome, Nat. Genet. 39 (2007) 311–318.
[22] S. Pennings, J. Allan, C.S. Davey, DNA methylation, nucleosome formation and positioning, Brief. Funct. Genomic Proteomic 3 (2005) 351–361.
[23] H.R. Chung, M. Vingron, Sequence-dependent nucleosome positioning, J. Mol. Biol. 386 (2009) 1411–1422.
[24] G. Pan, D. Pei, The stem cell pluripotency factor NANOG activates transcription with two unusually potent subdomains at its C terminus, J. Biol. Chem. 280 (2005) 1401–1407.
[25] K. Zheng, X. Wu, K.H. Kaestner, P.J. Wang, The pluripotency factor LIN28 marks undifferentiated spermatogonia in mouse, BMC Dev. Biol. 9 (2009) 38.
[26] N. Thorne, H. Amrein, Vomeronasal organ: pheromone recognition with a twist, Curr. Biol. 13 (2003) R220–R222.