Genomics 101 (2013) 113–119

Contents lists available at SciVerse ScienceDirect

Genomics

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

Quantitative analysis reveals increased histone modifications and a broad nucleosome-free region bound by histone acetylases in highly expressed genes in human CD4$^+$ T cells

Hongde Liu [a,*,1], Kun Luo [b,1], Hao Wen [c], Xin Ma [a], Jianming Xie [a], Xiao Sun [a]

[a] State Key Laboratory of Bioelectronics, Southeast University, Nanjing 210096, China
[b] Department of Neurosurgery, Xinjiang Evidence-based Medicine Research Institute, The First Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, China
[c] State Key Laboratory Incubation Base of Xinjiang Major Diseases Research, First Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, China

ARTICLE INFO

Article history:
Received 9 July 2012
Accepted 15 November 2012
Available online 27 November 2012

Keywords:
Histone
Nucleosomes
Gene expression
Single nucleotide polymorphism

ABSTRACT

Genome-wide mapping of nucleosomes and histone modifications revealed meaningful patterns. Despite advances in resolving the associations between chromatin and transcription, quantitative chromatin dynamics have not been well defined. We quantitatively determined differences in histone modifications, nucleosome positions, DNA methylation, and transcription factor binding in highly expressed and repressed genes in human CD4$^+$ T cells. We showed that the first (−1) nucleosome upstream of the transcription start site (TSS) is shifted to the 5′ direction, thus forming a broad nucleosome-free region (NFR) near the TSS in highly expressed genes in CD4$^+$ T cells. Moreover, the transcription factor YY1 and histone acetyltransferases bind the NFR with high affinity. Most of histone acetylations drastically increase in transcription activation (>5 folds). We also suggested that single nucleotide polymorphisms (SNPs) occur at a much lower frequency in highly expressed genes than in repressed genes. Our analysis quantitatively revealed details of chromatin dynamics.

1. Introduction

Chromatin structure is deeply affected by nucleosome positioning and various epigenetic modifications and is closely correlated with transcription. Chromatin structure dynamically changes in gene regulatory regions. Genome-wide mapping of nucleosomes and histone modifications revealed some key patterns in a variety of cell types [1]. In the vicinity of the transcription start site (TSS), there is a nucleosome-free (depleted) region (NFR or NDR) flanked by two well-positioned nucleosomes that often contain the histone variant H2A.Z [2]. In human CD4$^+$ T cells, the first (+1) nucleosome downstream of the TSS exhibits a differential pattern in active and silent genes. H2A.Z-containing and modified nucleosomes are preferentially lost from the first (−1) nucleosome position upstream of TSS [3]. In yeast, the size of the NFR varies, and expressed genes tend to have larger NFRs than unexpressed ones [4]. Histone acetylation (ac), H3 lysine 4 dimethylation (H3K4me2), and H3K4me3 are associated with active promoters [1,5–9]. Repressed genes are marked with H3K9me2/3, H3K27me3, and DNA methylation [6–9]. Bivalent modifications carrying tri-methylations of both H3K4 and H3K27 were revealed at development-related promoters in embryonic stem cells [10]. Transcribed regions enrich H3K36me3 [11]. A combinatorial signature of H3K4me3, which is enriched at promoters, and H3K36me3, which covers the transcribed region, was used to identify novel non-coding RNAs [12]. The models involving H3K4me3 and H3K79me1 are the most predictive of the expression levels in low CpG content promoters, whereas high CpG content promoters require H3K27ac and H4K20me1 [13]. A study in both cancer cells and stem cells reported that H3K4me1, H3K4me3, and H3K27ac at promoters are generally cell-type invariant, whereas those modifications at enhancers are cell type-specific [14]. The combinatorial patterns of epigenetic modifications can define the distinct chromatin states [15]. Chromatin structure was also suggested in genetic variation [16].

Despite the quickly advancing revelations of the associations between chromatin and transcription, a quantitative comparison of epigenetic modifications has not been performed in expressed genes and repressed genes. We are interested in the extent to which the change in the density of a specific chromatin mark causes productive transcription.

In this paper, we quantitatively determined the differences in histone modifications, nucleosome positions, DNA methylation, and transcription factor binding in highly expressed genes and repressed genes. Instead of choosing genes by sorting expression levels, we used housekeeping genes (HKGs) and tissue-specific genes (TSGs) to represent the highly expressed genes and the repressed genes, respectively. HKGs are compact in gene structure and are ubiquitously expressed in all tissues [17,18]. Moreover, their high expression is not due to the redundancy of gene copies [18]. Most TSGs are silent or

* Corresponding author.
E-mail address: liuhongde@seu.edu.cn (H. Liu).
[1] The authors contributed equally to the work.

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ygeno.2012.11.007

expressed at low levels, and only a small fraction of them are expressed in a specific tissue. Thus, the average expression level of TSGs is low [17]. We also examined genetic variations in the two types of genes and their associations with epigenetic factors. Our analysis provided a quantitative comparison and revealed some interesting details about chromatin dynamics.

2. Materials and methods

2.1. Datasets

Lists of 1522 HKGs and 975 TSGs were retrieved from the literature [17]. Only the genes that could be uniquely mapped to probe identifiers of the Affymetrix Human Genome U133A Array were used, resulting in a list of 1352 HKGs and 858 TSGs. The genomic coordinates of TSSs of the genes were retrieved through "Tables" function from Genome Browser (http://genome.ucsc.edu) of the University of California, Santa Cruz (UCSC) (NCBI36/hg18). Expression data in $CD4^+$ T cells was downloaded from Gene Expression Omnibus (GEO) of NCBI (accession no. GSE10437) [3]. It contains raw expression values of two replicates for each state of cell (resting and activated). In using, the expression values were averaged over two replicates. In order to compare epigenetic patterns between HKGs and $CD4^+$ T cell-specific highly expressed genes, firstly, top 1000 highly expressed genes and top 1000 lowly expressed genes were selected by sorting gene expression in resting and activated $CD4^+$ T cells, respectively, resulting in 4 sets of genes lists. Then, HKGs and TSGs are excluded from the 4 sets of genes lists (Table 1).

The coordinates of ChIP-Seq tags were taken from genome-wide studies of the distributions of 20 histone methylations, one H2A.Z histone variant, Pol II [6], 18 histone acetylations in CD4+ T cells [9], and nucleosomes in both resting and activated $CD4^+$ T cells [3]. Data of the DNA methylation levels (the mean methylation percentages) at the promoter (−1 kbp to +0.6 kbp relative to TSS) and gene body in naïve $CD4^+$ T cells were retrieved from GEO of NCBI (accession no. GSM871287) [19]. Datasets recording the binding locations of 5 histone acetyltransferases (HATs) and 4 histone deacetylases (HDACs) in $CD4^+$ T cells were retrieved from the literature (accession no. GSE15735) [20]. Data for the binding sites of the transcription factor YY1 were retrieved from GEO (accession no. GSM630810) [21].

Datasets of single nucleotide polymorphism (SNP), insertion (In), deletion (Del), and insertion–deletion (In-del) were downloaded through "Tables" function of Genome Browser of UCSC (NCBI36/hg18, SNP130). Only the verified data was used.

2.2. Chromatin mark profiles of histone modifications, nucleosome positioning, and binding of transcription factors (TFs) around TSSs

For each type of chromatin mark (histone modifications, nucleosomes, Pol II, CTCF), the number of sequencing tags was counted at every genomic location in a 4 k bp region surrounding the TSS. The process was carried out on both positive and negative strands, thus resulting in two profiles of marks. The two profiles were then aligned by oppositely moving with an appropriate shift so that the Euler distance between the corresponding parts of the two profiles was minimized (the shift value for each type of chromatin mark was shown in Table s1). The two profiles were then summed at every corresponding location to generate a final profile. For quantitative comparisons, the profiles for HKGs and TSGs were averaged by dividing by the number of HKGs and TSGs, respectively. The coordinates of the nucleosome tags were extended to 146 bp in the 3′ direction prior to profiling, and the shift was limited at 73 bp. For TF YY1, 5 HATs (P300, CBP, PCAF, Tip60, and MOF) and 4 HDACs (HDAC1/2/3/6), their binding profiles were generated in 4 k bp regions around TSSs using the coordinates of ChIP-Seq tags [20,21].

We called these profiles of histone modifications, nucleosome positions, and TF binding sites "chromatin mark profiles." Similarly, chromatin mark profiles were also computed for the $CD4^+$ T cell-specific highly expressed genes (Table 1).

2.3. Density profiles of SNPs, Dels, and In-dels

Density profiles of SNPs, Dels, and In-dels were calculated by counting the total number of each variation that occurred at every genomic location in a 4 k bp region around TSSs.

2.4. Comparisons of chromatin mark profiles in highly and lowly expressed genes (HKGs and TSGs)

The profiles for nucleosomes, CTCF, Pol II, and 38 histone modifications were clustered using a *k*-means algorithm. Parameter *k* in the algorithm was estimated by choosing representative profiles by comparing profile shapes in HKGs. Firstly, profiles of H3K4me3, H3K79me3, H3K27me1, H3R2me2, H3K9me1, nucleosome and Pol II were chosen, so initial *k* is 7. Then after the first clustering, the cluster number was manually adjusted so as to make it more reasonable. At last, profiles for nucleosome and Pol II were set as a single cluster, respectively. The clustering profiles were summed to gain an average profile. Pearson correlation coefficients (PCC) were used to estimate the correlations between the profiles. The Wilcoxon rank sum test was used to evaluate the difference between any two profiles.

Fold changes of the profiles of histone modifications in HKGs and TSGs were examined at the −1 nucleosome, in the NFR, at the +1 and +2 nucleosomes, and in a range of −1 kbp to +0.6 kbp regions around the TSS. For the −1, +1, and +2 nucleosomes, the density of histone marks was counted in a 147 bp region surrounding nucleosome dyad. Fold changes of the histone modifications were represented by ratios of counts in HKGs to counts in TSGs.

3. Results and discussion

3.1. Highly expressed genes were associated with a broader nucleosome-free region in the vicinity of the TSS

Averagely, the expression level of HKGs is high and that of TSGs is low. In both resting and activated $CD4^+$ T cells, as expected, the expression levels of HKGs are higher than that of TSGs ($p = 4.0\times10^{-232}$, *t*-test) (Fig. 1), indicating that the two groups of genes indeed represent highly and lowly expressed genes and can be used in further analysis. HKGs have a higher GC content than TSGs (*t*-test, $p = 1.5\times10^{-12}$) (Fig. s1). The profiles of nucleosome marks in HKGs show a similar shape regardless of the resting or activated cell states (Fig. 2A). For TSGs, the profiles also match in resting and activated cells, especially

Table 1
Amount of HKGs and TSGs that overlap with the $CD4^+$ T cell-specific highly and lowly expressed genes. In profiling chromatin marks, both HKGs and TSGs are excluded from the highly and lowly expressed genes lists. Fisher's exact test suggests the proportions of both HKGs and TSGs are significantly different in the highly and lowly expressed genes.

	Top 1000 highly expressed gene (HKGs, TSGs)	Top 1000 lowly expressed gene (HKGs, TSGs)	*p*-value of Fisher's exact test	
			HKGs in the highly and lowly expressed genes	TSGs in the highly and lowly expressed genes
In resting $CD4^+$ T cells	1000 (181, 4)	1000 (21, 34)	1.0×10^{-35}	1.0×10^{-6}
In activated $CD4^+$ T cells	1000 (189, 1)	1000 (23, 28)	1.0×10^{-36}	1.0×10^{-7}

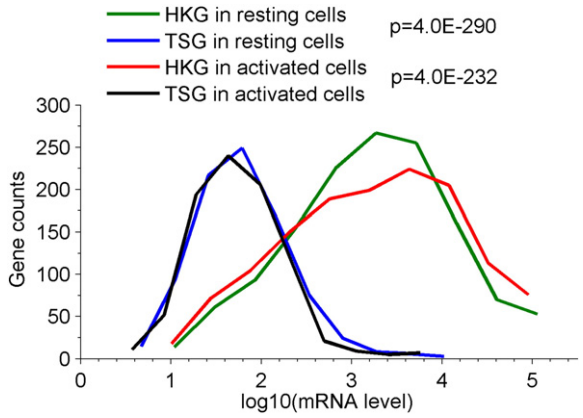

Fig. 1. Gene expression levels of HKGs and TSGs in both resting and activated CD4$^+$ T cells. The distributions of the logarithm of mRNA levels in both resting and activated CD4$^+$ T cells are shown. The difference between these distributions was tested by a two-sample *t*-test. The mean mRNA levels of HKGs and TSGs are 1698.2 and 63.1 in resting CD4+ T cells.

in a 0.4 kbp region surrounding the TSS. These results suggest that most nucleosomes have stable positions even when the cell states change.

We manually measured nucleosomes positions by searching peaks in the nucleosome profiles in Fig. 2A. The +1 nucleosome is well positioned at +135±4 bp downstream of TSSs in both HKGs and TSGs. In Zhao et al.'s study [3], the +1 nucleosome exhibited differential positioning in active and silent genes. However, in our result, only the tag density (the normalized tag count) increased ~20%, and the position did not significantly change (variation <8/147=6%). The divergence was probably caused by the different methods that were used to choose contrasting genes. In a "barrier" model [22,23], the +1 nucleosome serves as a "barrier" and is involved in forming uniform positioning downstream of the TSS. In our study, the spacing of neighbor nucleosomes varied from 180 to 200 bp. Because the +1 nucleosome was consistently aligned relative to the TSS, we speculated that the +1 nucleosome associated with the site of transcription initiation.

Significantly, the first (−1) nucleosome upstream of TSS showed differential positioning. In TSGs, it positions at −110 bp, while in HKGs, it shifts to −200 bp upstream of the TSS (Fig. 2A). Considering the +1 nucleosome well positions at ~+135 bp downstream of TSSs in both HKGs and TSGs, we inferred that the NFR in HKG was 90 bp (200−110) broader than that in TSGs.

We then asked if the broader NFR in HKGs is also in the CD4$^+$ T cell-specific highly expressed genes. Using the highly expressed genes selected by ranking expression data (see Part 2.1, Table 1), we tested the speculation. Results indicate that the highly expressed genes also show a broader NFR near TSSs than the lowly expressed genes (Fig. 2B). In resting T cells, dyad of the +1 nucleosome is at ~+135 bp in both types of genes. Of the −1 nucleosome, the dyad is at −100 bp in the lowly expressed genes, while it is at −175 bp in the highly expressed genes, suggesting a broad NFR (Fig. 2B). In the activated cells, the result is similar. This is consistent with the finding that a broad NFR at promoter regions correlates with high expression of genes in yeast [4]. Also, it was found during differentiation of hematopoietic stem cells, a longer nucleosome linker region was generated by shifting the flanking nucleosomes away at enhancer GATA1 sites [23].

The tags density of the −1 nucleosome shows a comparable level between HKGs and TSGs in both resting cells ($p=0.31$, Wilcoxon rank sum test) and activated cells ($p=0.39$) (Fig. 2A). But it is different between the T cell-specific highly and lowly expressed genes

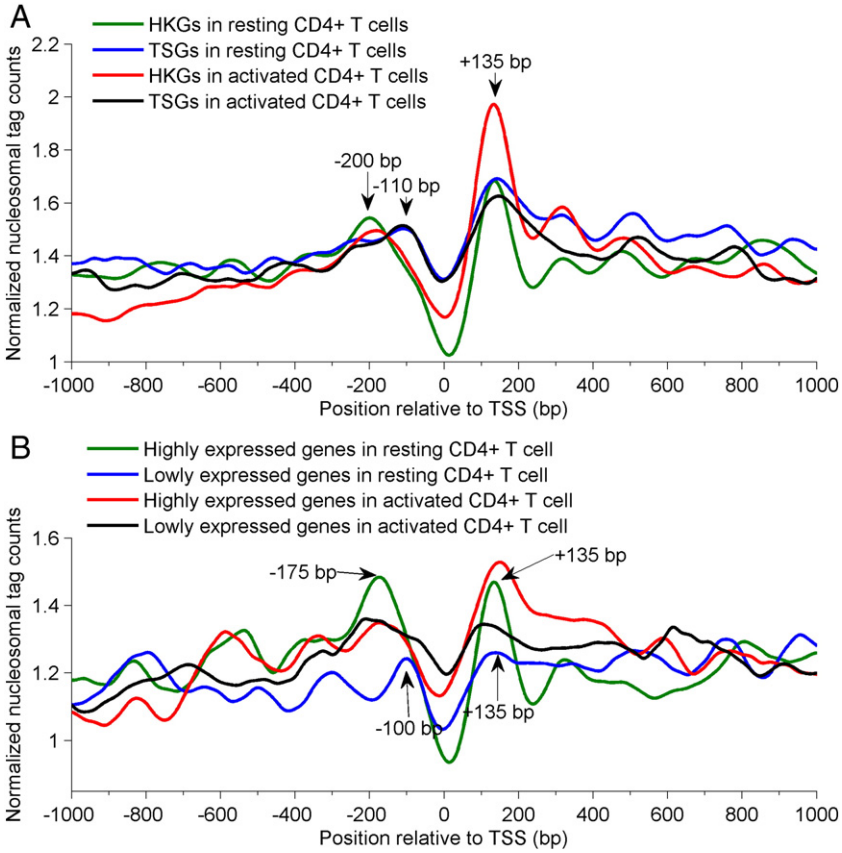

Fig. 2. Profiles of nucleosomes near TSSs for genes with different expression levels in both resting and activated CD4$^+$ T cells; subplot A is for HKGs and TSGs; subplot B is for the CD4$^+$ T cell-specific highly and lowly expressed genes (see Table 1). Arrows indicate the dyad positions of nucleosomes.

(Fig. 2B), suggesting this characteristic observed in HKG and TSG cannot be generalized to the $CD4^+$ T cell-specific highly and lowly expressed genes.

Zhao et al. suggested that the deposition of histone variant H2A.Z or modification by H3K4me3 may facilitate nucleosome eviction or repositioning in the −1 nucleosome region [3]. We calculated the fractions of H2A.Z-containing nucleosome by dividing the tag density of H2A.Z by that of the nucleosome at the dyad of the −1 nucleosome (Fig. s2). Result indicates 68% and 61% of nucleosomes contains H2A.Z variant in HKGs and the $CD4^+$ T cell-specific highly expressed genes, respectively, whereas only 9% of nucleosomes in TSGs contain H2A.Z (Fig. s2). Also, the level of H3K4me3 at the −1 nucleosome shows a 7-fold increase in HKGs compared to TSGs. These results are consistent with Zhao et al.'s findings [3].

Taken together, both HKGs and the $CD4^+$ T cell-specific highly expressed genes show a broad NFR near TSS. The broad NFR is essential for transcription. In the process of activating a repressed gene, the −1 nucleosome probably undergoes a shift to upstream (5'), thus resulting a broad NFR.

3.2. Patterns of chromatin marks

We clustered 42 profiles of chromatin marks using *k*-means algorithm. The profiles for HKGs exhibit obvious patterns (Fig. 3). The profile of each type of chromatin mark is shown in Fig. s2. Histone acetylation, variant H2A.Z, and H3K4me3 have similar profiles and cluster together (red line in Fig. 3). These marks are involved in transcription activation [5–7,9]. The average profile of the marks drops near the TSS, well corresponding to the NFR, and peaks at the −1 and +1 nucleosomes, and then drastically decreases to the background level in a 1.5 k bp region flanking the TSS. The similarity of these profiles was also evidenced by the high correlation coefficients between the profiles (average coefficient >0.788) (Fig. s3). This also suggests that the active chromatin marks tend to co-occur to form a combinatorial pattern [24,25].

The profile of Pol II peaks at +40 bp, which is close to the +1 nucleosome. The profiles of H3K79me1/2/3 and H3K36me3 show low levels upstream of the TSS, and the levels sharply increase downstream of the +3 nucleosome. H3K36me3 is a hallmark of transcribed regions and is used to identify the TSS [11,12]. Here, we suggested that methylation of H3K79 also can be an indicator of gene coding regions. Enrichment of H4K20me1and H2BK5me1 is observed downstream of the +3 nucleosome. H3K27me3 and H3K9me3 are associated with heterochromatin [5–7,9]. Repressive histone modifications cluster into two subgroups (blue line and yellow line in Fig. 3) that include H3K4me1/2, H3K27me1/2, H3K9me1, and H3R2me2. These modifications are depleted in a broad range from 0.5 kbp upstream to 0.8 kbp downstream of the TSS.

Moreover, we profiled the histone modifications tags near TSSs for the resting $CD4^+$ T cell-specific highly expressed genes (see middle column in Fig. s2) and computed the correlation coefficients of the profiles between HKGs and the highly expressed genes (Fig. s4).

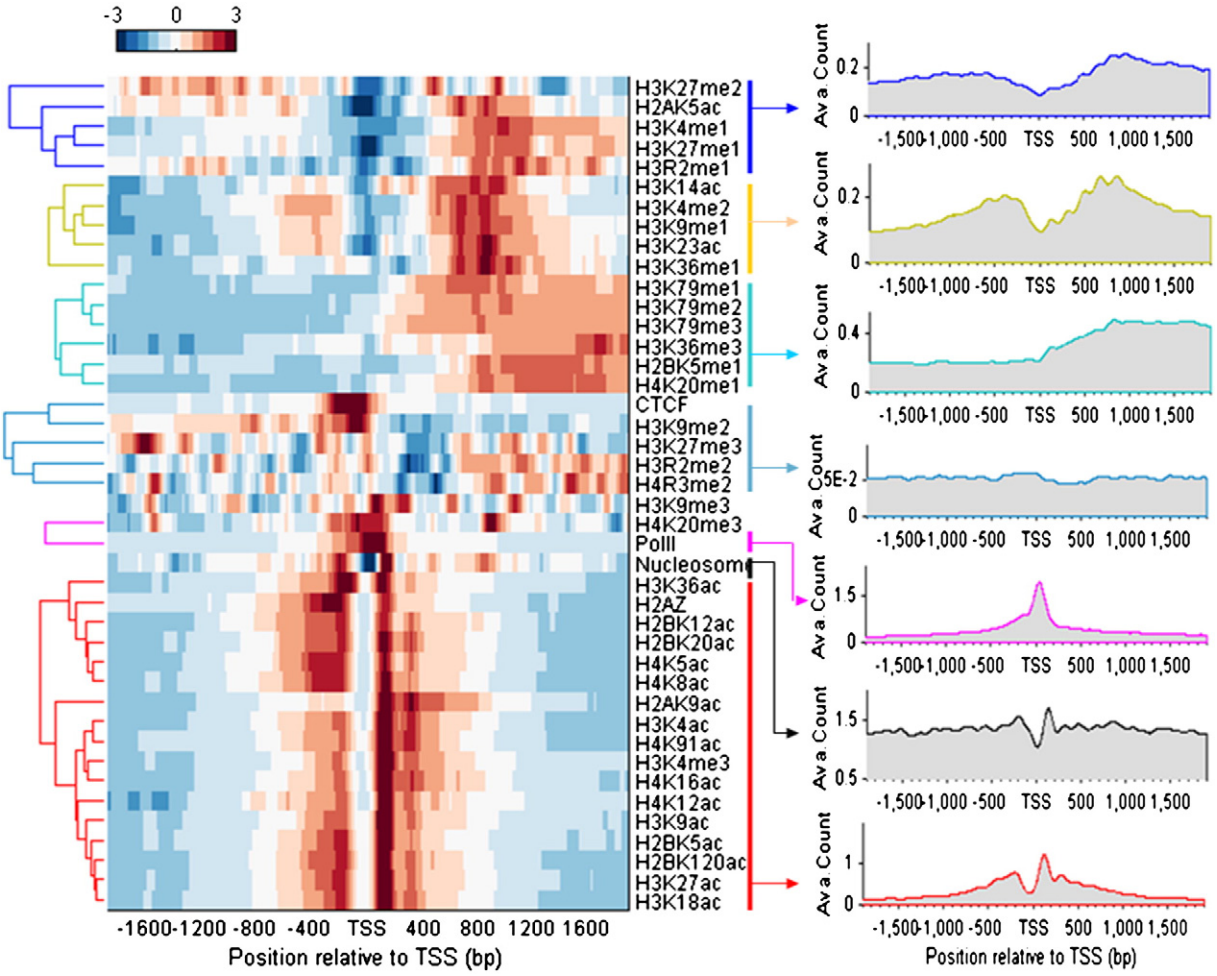

Fig. 3. Clustering of 42 chromatin mark profiles for HKGs with *k*-means algorithm (left). The average profiles of subgroups are shown (right). The colored vertical lines and arrows in the middle indicate each of the subgroups, and profiles for both nucleosome and Pol II are shown separately.

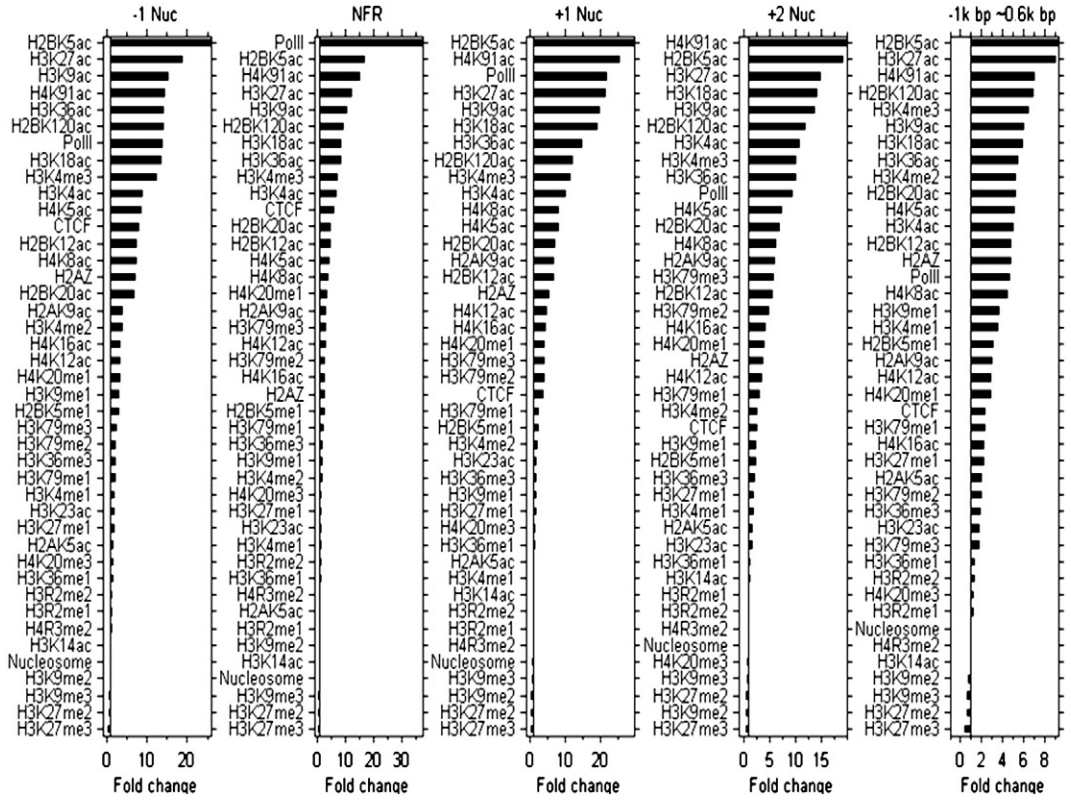

Fig. 4. Fold changes of HKG compared to TSG in levels of chromatin marks at the −1, +1, and +2 nucleosomes, in the NFR, and in a region spanning −1 kbp to +0.6 kbp of the TSS. For HKGs, the centre positions of −1 nucleosome, the NFR, +1 and +2 nucleosomes are at −200, 12, 135 and 315 bp, respectively; for TSGs, they are at −110, −2, 135 and 315 bp, respectively.

There are 31 types of chromatin marks of which the correlation coefficients are more than 0.80 (Fig. s4), indicating that the highly expressed genes and HKGs share the similar epigenetic patterns.

In TSGs, partly due to the low density of tags, a clear clustering map was not observed for profiles of chromatin marks (Fig. s5). The marks do not show an obvious enriched distribution. The repressive modifications accumulate near the TSS. Interestingly, Pol II in TSGs shows a peak at −580 bp upstream of the TSS.

3.3. HKGs enrich most of the histone modification marks

We quantitatively determined the fold changes in enrichment of the chromatin marks at the −1, +1, and +2 nucleosomes, in the NFR, and in the range from −1 k bp to +0.6 kbp relative to the TSS in HKGs and TSGs (Fig. 4). Compared with changes in histone methylation, histone acetylation is drastically changed between HKGs and TSGs, which is consistent with the literature [6]. For some types of acetylation, such as H3K27, H2AK5, and H4K91, acetylation levels are 10 folds greater in HKGs than in TSGs. It is exceptional to note that H3K14ac remains nearly constant in the region between the −1 and +1 nucleosomes.

The H3K4me2/3, H3K79me1/2/3, H3K27ac, H4K20me1, and H2BK5ac modifications were suggested to be the most predictive of gene expression [11,13,26]. In our quantitative analysis, the modifications except acetylation of H3K27 and H2BK5 showed a small fold change (<5). The modifications with a low fold change are more sensitive in predicting expression level than those with large changes. Taking H3K79me and H3K27ac as examples, every twofold increase in H3K79me at the +1 nucleosome means a 27 folds increase in gene expression (Figs. 1 and 4), whereas H3K27ac needs to increase 20 folds to reach the same increase in gene expression. From this view, a model involving modifications of H3K4me2/3, H3K79me1/2/3, and H4K20me1 will have a better capacity of predicting gene expression level than a model using H3K27ac and H2BK5ac. In order to verify the hypothesis, the expression levels of 1027 genes of human chromosome 6 were linearly regressed with H3K79me1 and H3K27ac tags density at the +1 nucleosome, respectively (Fig. s6). Result indicated the regression with H3K79me1 showed a better performance ($\log_{10}(\text{mRNA level}) = 2.5346 + 11.0188 \times \text{H3K79me1}_{\text{tag\_density}}$, $p = 5.14 \times 10^{-9}$ (F-test)). H2BK5ac levels changed by more than 25 folds; this modification was suggested in highly expressed promoters [9,20].

It was reported that pericentromeres enrich H3R2me2, H4K20me3, and H4R3me2 [27]. In our analysis, the levels of arginine methylation (H3R2 and H4R3) remained nearly constant, possibly suggesting that H3R2me2 and H4R3me2 do not have roles in transcription. Repressive H3K9me2/3 and H3K27me2/3 also exhibited small changes of less than 2 folds. Considering the similar epigenetic patterns between the resting cell-specific highly expressed genes and HKGs (Figs. s2 and s4), we conclude that both HKGs and the highly expressed genes enrich histone acetylations and some types of histone methylations.

3.4. DNA methylation highly accumulates at promoters of TSGs

We also compared the levels of DNA methylation (the mean methylation percentages) in HKGs and TSGs. At promoters of HKGs, less than 27.6% of DNA was methylated, while at promoters of TSGs, over 51.9% of DNA was methylated (Fig. 5), showing a significant accumulation ($p = 1.0 \times 10^{-9}$, t-test) in TSGs. In contrast, the levels of DNA methylation of the gene body in HKGs and TSGs are 45% and 43%, respectively, which are not significantly different ($p = 0.31$, t-test) (Fig. 5). This result indicates that DNA methylation accumulates at lowly expressed promoters.

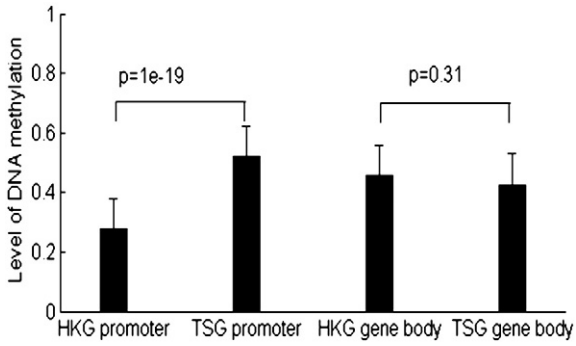

Fig. 5. DNA methylation levels (the mean methylation percentages) in promoters and gene bodies of HKGs and TSGs, the significance of difference is tested by two-sample *t*-test, the bar on the pillar indicates standard error of mean.

3.5. Both histone acetyltransferase and deacetylase reside in the NFR near the TSS

As shown in Fig. 2, both the CD4$^+$ T cell-specific highly expressed gene and HKG are characterized by a broad NFR in the vicinity of TSS. We were interested in learning which TFs bind in the NFR. First, we examined binding data for the TF YY1, a ubiquitous transcription factor [21]. Cuddapah et al. identified that the binding sites of YY1 colocalized with a high mobility group protein (HMGN1) [21]. We found that the binding sites of YY1 are significantly concentrated in the broad NFR of both HKGs and the highly expressed genes (Fig. 6A). Then, considering that HATs associate with histone acetylations and accounts for the high acetylation level at promoters of HKGs (Fig. 4), we examined the profile of HAT binding sites. Wang et al. revealed that HATs and HDACs are both targeted to transcribed regions of active genes by phosphorylated RNA Pol II [20]. We further found that profiles for the binding sites of HATs (encoded by the genes EP300, CREBBP, KAT2B, KAT5 and KAT8) (Figs. 6B–F) and HDACs (HDAC1,HDAC2 and HDAC3) (Figs. 6G–I) were significantly high in the region corresponding to the NFR in both HKGs and the CD4$^+$ T cell-specific highly expressed genes. In contrast, their binding sites in TSGs were very limited and scattered across a large region (Fig. 6). Because of this, we speculated that YY1 plays a role in recruiting the HATs. The increased width of the NFR in highly expressed genes was related to the binding of both YY1 and HATs. However, whether the broad NFR is a consequence or a cause of binding of YY1 and HATs remains unknown.

3.6. Highly expressed genes associate with low SNP density

Transcription is coupled with DNA repair [16,28], and chromatin structure plays roles in transcription. Highly expressed genes have a greater opportunity to repair DNA damage. Therefore, we hypothesized that the frequency of genetic variation in a gene was correlated with the gene's expression and thus was affected by chromatin structure. In order to validate this hypothesis, we determined the frequency of SNPs in a 4 kbp region surrounding the TSS. As expected, SNPs occur at a much lower frequency in highly expressed genes than in repressed genes ($p = 6.1 \times 10^{-37}$, Wilcoxon rank sum test) (Figs. 7A and s7), suggesting that the highly transcribed genes can efficiently repair DNA damage.

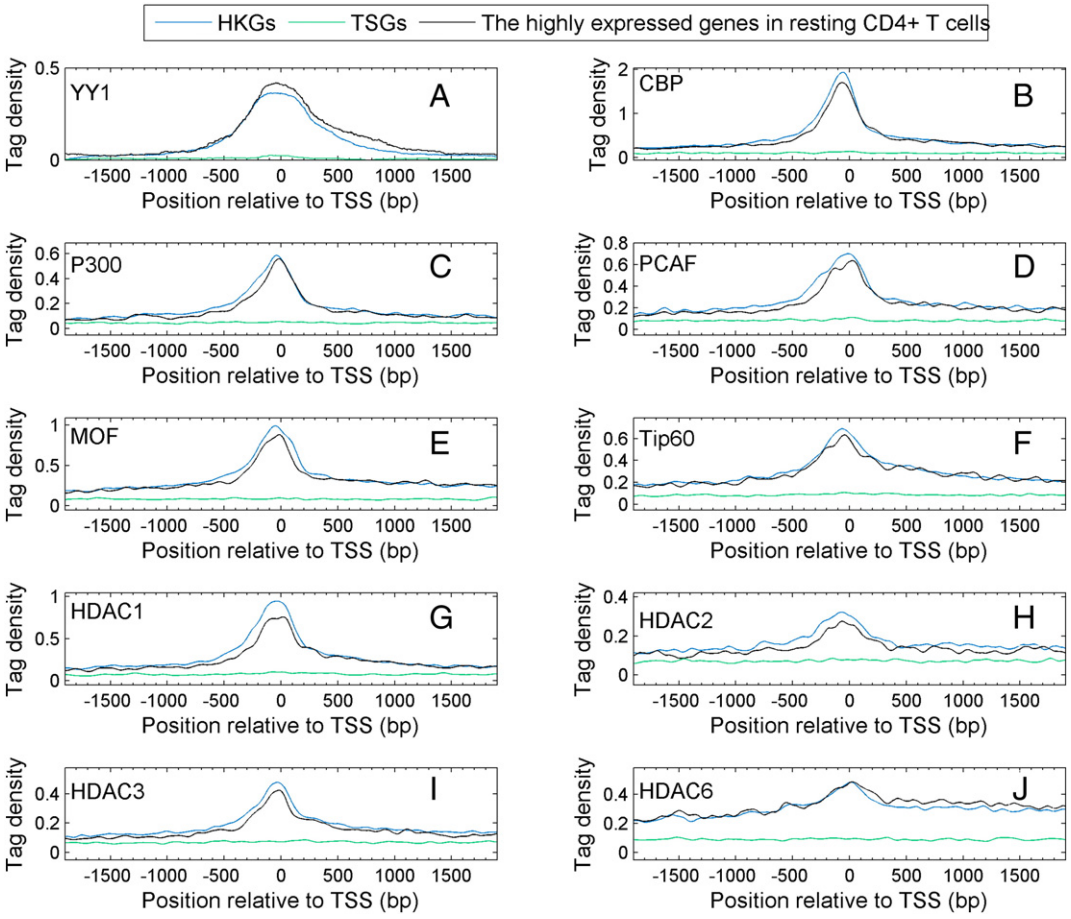

Fig. 6. Binding profiles of transcription factor YY1 and histone acetyltransferases at promoters of HKGs and TSGs. A, YY1; B, CBP, C, P300; D, PCAF; E, MOF; F, Tip60; G, HDAC1; H, HDAC2; I, HDAC3; J, HDAC6; black lines indicate the profiles for the resting CD4$^+$ T cell-specific highly expressed genes.

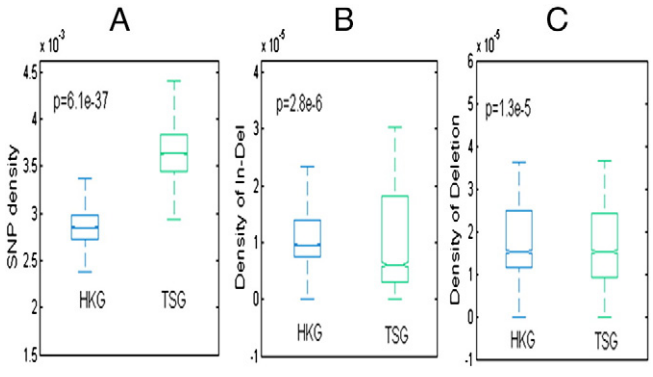

Fig. 7. Frequency of genetic variations in HKGs and TSGs; shown are boxplots of genetic variation frequency in HKGs and TSGs, A, SNP; B, In-Del; C, Del; *p*-value is from Wilcoxon rank sum test.

Other variations including Dels and In-dels also exhibited a similar tendency (Figs. 7B and C and s7). Then, we examined the correlation between histone modifications and the frequency of variation (Table s2). The frequency of SNPs negatively correlates with histone acetylation, H3K4me3, and H3K79me1, suggesting that the genomic region with the histone modifications is not prone to accumulating nucleotide variations.

4. Conclusion

We quantitatively determined the differences of chromatin marks between highly expressed genes and repressed genes. We revealed that the NFR near TSSs of highly expressed genes was broader than the NFR near TSSs of repressed genes. The $CD4^+$ T cell-specific highly expressed genes and HKGs share the similar epigenetic patterns. Interestingly, binding sites for the TF YY1 and HATs are enriched in the NFR, suggesting a close relationship between the binding and the repositioning of the first nucleosome upstream of the TSS. Most of histone acetylations associate with a productive transcription. In addition, our analysis indicates that a model involving H3K4me2/3, H3K79me1/2/3, and H4K20me1 will be more predictive of gene expression than models involving H3K27ac and H2BK5ac. SNPs occur at a much lower frequency in HKGs than in TSGs. Our analysis quantitatively revealed details of chromatin dynamics.

Acknowledgments

This work was supported by Natural Science foundation of Xinjiang Uygur Autonomous Region (No.: 2012211A076), the Natural Science Foundation of China (No.: 31100914), and the National Basic Research Program of China (No.: 2012CB316501).

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2012.11.007.

References

[1] W.V. Zhou, A. Goren, B.E. Bernstein, Charting histone modifications and the functional organization of mammalian genomes, Nat. Rev. Genet. 12 (2011) 7–18.
[2] P.D. Hartley, H.D. Madhani, Mechanisms that specify promoter nucleosome location and identity, Cell 137 (3) (2009) 445–458.
[3] D.E. Schones, K.R. Cui, S. Cuddapah, T.Y. Roh, A. Barski, Z.B. Wang, et al., Dynamic regulation of nucleosome positioning in the human genome, Cell 132 (2008) 887–898.
[4] B.J. Zaugg, M.N. Luscombe, A genomic model of condition-specific nucleosome behavior explains transcriptional activity in yeast, Genome Res. 22 (2012) 84–94.
[5] J. Ernst, P. Kheradpour, T.S. Mikkelsen, N. Shoresh, L.D. Ward, C.B. Epstein, et al., Mapping and analysis of chromatin state dynamics in nine human cell types, Nature 473 (2011) 43–49.
[6] A. Barski, S. Cuddapah, K.R. Cui, T.Y. Roh, E.D. Schones, Z.B. Wang, et al., High-resolution profiling of histone methylations in the human genome, Cell 129 (2007) 823–837.
[7] T.Y. Roh, S. Cuddapah, K.R. Cui, K.J. Zhao, The genomic landscape of histone modifications in human T cells, Proc. Natl. Acad. Sci. U. S. A. 13 (43) (2006) 15782–15787.
[8] A. Meissner, S.T. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, et al., Genome-scale DNA methylation maps of pluripotent and differentiated cells, Nature 454 (2008) 766–770.
[9] Z.B. Wang, C.Z. Zang, A.J. Rosenfeld, E.D. Schones, A. Barski, S. Cuddapah, et al., Combinatorial patterns of histone acetylations and methylations in the human genome, Nat. Genet. 40 (7) (2008) 897–903.
[10] X.D. Zhao, X. Han, J.L. Chew, J. Liu, K.P. Chiu, A. Choo, et al., Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells, Cell Stem Cell 1 (2007) 286–298.
[11] S. Schwartz, E. Meshorer, G. Ast, Chromatin organization marks exon–intron structure, Nat. Struct. Mol. Biol. 16 (2009) 990–995.
[12] M. Guttman, I. Amit, M. Garber, C. French, F.M. Lin, D. Feldser, et al., Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals, Nature 458 (2009) 223–227.
[13] R. Karlić, H.R. Chung, J. Lasserre, K. Vlahovíček, M. Vingron, Histone modification levels are predictive for gene expression, Proc. Natl. Acad. Sci. USA 107 (7) (2010) 2926–2931.
[14] D.N. Heintzman, C.G. Hon, D.R. Hawkins, P. Kheradpour, A. Stark, F.L. Harp, et al., Histone modifications at human enhancers reflect global cell-type-specific gene expression, Nature 459 (2009) 108–112.
[15] J. Ernst, M. Kellis, Discovery and characterization of chromatin states for systematic annotation of the human genome, Nat. Biotechnol. 28 (2010) 817–825.
[16] Y.T. Michael, V. Natalia, M.S. Robert, J.P. Peter, Impact of chromatin structure on sequence variability in the human genome, Nat. Struct. Mol. Biol. 18 (4) (2011) 510–515.
[17] X.W. She, A.C. Rohl, C.J. Castle, V.A. Kulkarni, M.J. Johnson, R.H. Chen, Definition, conservation and epigenetics of housekeeping and tissue-enriched genes, BMC Genomics 10 (2009) 269.
[18] E. Eisenberg, Y.E. Levanon, Human housekeeping genes are compact, Trends Genet. 19 (7) (2003) 362–365.
[19] J. Han, S.G. Park, J.B. Bae, J. Choi, J.M. Lyu, S.H. Park, et al., The characteristics of genome-wide DNA methylation in naïve CD4+ T cells of patients with psoriasis or atopic dermatitis, Biochem. Biophys. Res. Commun. 422 (1) (2012) 157–163.
[20] Z.B. Wang, C.Z. Zang, K.R. Cui, E.D. Schones, A. Barski, W.Q. Peng, et al., Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes, Cell 138 (5) (2009) 1019–1031.
[21] S. Cuddapah, D.E. Schones, K.R. Cui, T.Y. Roh, A. Barski, G. Wei, et al., Genomic profiling of HMGN1 reveals an association with chromatin at regulatory regions, Mol. Cell. Biol. 31 (4) (2011) 700–709.
[22] T.N. Mavrich, I.P. Ioshikhes, B.J. Venters, C.Z. Jiang, L.P. Tomsho, J. Qi, A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome, Genome Res. 18 (2008) 1073–1083.
[23] A. Valouev, S.M. Johnson, S.D. Boyd, C.L. Smith, A.Z. Fire, A. Sidow, Determinants of nucleosome organization in primary human cells, Nature 474 (2011) 516–520.
[24] D. Ucar, Q.Y. Hu, K. Tan, Combinatorial chromatin modification patterns in the human genome revealed by subspace clustering, Nucleic Acids Res. 39 (10) (2011) 4063–4075.
[25] A. Lennartsson, K. Ekwall, Histone modification patterns and epigenetic codes, Biochim. Biophys. Acta 1790 (2009) 863–868.
[26] C. Cheng, K.K. Yan, K.Y. Yip, J. Rozowsky, R. Alexander, C. Shou, et al., A statistical framework for modeling gene expression using chromatin features and application to modENCODE datasets, Genome Biol. 12 (2) (2011) R15.
[27] J.A. Rosenfeld, Z.B. Wang, D.E. Schones, K.J. Zhao, R. DeSalle, et al., Determination of enriched histone modifications in non-genic portions of the human genome, BMC Genomics 10 (2009) 143.
[28] P.C. Hanawalt, G. Spivak, Transcription-coupled DNA repair: two decades of progress and surprises, Nat. Rev. Mol. Cell Biol. 9 (2008) 958–970.