Resource

Cell

High-Resolution Profiling of Histone Methylations in the Human Genome

Artem Barski,[1,3] Suresh Cuddapah,[1,3] Kairong Cui,[1,3] Tae-Young Roh,[1,3] Dustin E. Schones,[1,3] Zhibin Wang,[1,3] Gang Wei,[1,3] Iouri Chepelev,[2] and Keji Zhao[1,*]

[1] Laboratory of Molecular Immunology, National Heart, Lung, and Blood Institute, NIH, Bethesda, MD 20892, USA

[2] Department of Human Genetics, Gonda Neuroscience and Genetics Research Center, University of California, Los Angeles, Los Angeles, CA 90095, USA

[3] These authors contributed equally to this work and are listed alphabetically.

*Correspondence: zhaok@nhlbi.nih.gov

DOI 10.1016/j.cell.2007.05.009

SUMMARY

Histone modifications are implicated in influencing gene expression. We have generated high-resolution maps for the genome-wide distribution of 20 histone lysine and arginine methylations as well as histone variant H2A.Z, RNA polymerase II, and the insulator binding protein CTCF across the human genome using the Solexa 1G sequencing technology. Typical patterns of histone methylations exhibited at promoters, insulators, enhancers, and transcribed regions are identified. The monomethylations of H3K27, H3K9, H4K20, H3K79, and H2BK5 are all linked to gene activation, whereas trimethylations of H3K27, H3K9, and H3K79 are linked to repression. H2A.Z associates with functional regulatory elements, and CTCF marks boundaries of histone methylation domains. Chromosome banding patterns are correlated with unique patterns of histone modifications. Chromosome breakpoints detected in T cell cancers frequently reside in chromatin regions associated with H3K4 methylations. Our data provide new insights into the function of histone methylation and chromatin organization in genome function.

INTRODUCTION

Eukaryotic DNA is packaged into a chromatin structure consisting of repeating nucleosomes formed by wrapping 146 base pairs of DNA around an octamer of four core histones (H2A, H2B, H3, and H4). The histones, particularly their N-terminal tails, are subject to a large number of posttranslational modifications (Kouzarides, 2007). Histone modifications are implicated in influencing gene expression and genome function by establishing global chromatin environments and orchestrating DNA-based biological processes. Among the various modifications, histone methylations at lysine and arginine residues are relatively stable and are therefore considered potential marks for carrying the epigenetic information that is stable through cell divisions. Indeed, enzymes that catalyze the methylation reaction have been implicated in playing critical roles in development and pathological processes.

Remarkable progress has been made during the past few years in the characterization of histone modifications on a genome-wide scale. The main driving force has been the development and improvement of the "ChIP-on-chip" technique by combining chromatin immunoprecipitation (ChIP) and DNA-microarray analysis (chip). With almost complete coverage of the yeast genome on DNA microarrays, its histone modification patterns have been extensively studied. The general picture emerging from these studies is that promoter regions of active genes have reduced nucleosome occupancy and elevated histone acetylation (Bernstein et al., 2002, 2004; Lee et al., 2004; Liu et al., 2005; Pokholok et al., 2005; Sekinger et al., 2005; Yuan et al., 2005). High levels of H3K4me1, H3K4me2, and H3K4me3 are detected surrounding transcription start sites (TSSs), whereas H3K36me3 peaks near the 3′ end of genes.

Significant progress has also been made in characterizing global levels of histone modifications in mammals. Several large-scale studies have revealed interesting insights into the complex relationship between gene expression and histone modifications. Generally, high levels of histone acetylation and H3K4 methylation are detected in promoter regions of active genes (Bernstein et al., 2005; Kim et al., 2005; Roh et al., 2005, 2006), whereas elevated levels of H3K27 methylation correlates with gene repression (Boyer et al., 2006; Lee et al., 2006; Roh et al., 2006). In addition to the promoter regions, these modifications are also detected in intergenic regions as both sharply localized peaks and wide-spread domains. The H3 acetylation and H3K4me1 signals outside of promoter regions have been correlated with functional enhancers in various cell types (Heintzman et al., 2007; Roh et al., 2005; Roh et al., 2007). The apparently opposite modifications, H3K4me3 and H3K27me3, colocalize in regions termed

"bivalent domains" in embryonic stem cells, which have been suggested to function in the differentiation of these cells (Bernstein et al., 2006). Similar domains also exist in differentiated T cells (Roh et al., 2006). Although H3K9 methylation has been implicated in heterochromatin formation and gene silencing, a large-scale analysis suggested that H3K9me3 is enriched in many active promoters (Squazzo et al., 2006). However, extensive location analysis of only a few histone methylations has been completed at a limited genome coverage and resolution in human cells (Bernstein et al., 2007). Comprehensive and high-resolution colocalization analysis of all histone methylations for the entire human genome is required to understand the functional correlation of various histone lysine and arginine methylations in processes such as transcription and DNA repair.

ChIP-on-chip has a number of limitations, including the large sets of arrays needed to cover the mammalian genome and the potential bias introduced by amplification (Bernstein et al., 2007). We have previously successfully combined ChIP with serial analysis of gene expression (SAGE) in a method termed genome-wide mapping technique (GMAT) (Roh et al., 2004), to map several histone H3 modifications in human T cells, including H3 K9/K14 diacetylation (Roh et al., 2005), H3K4 trimethylation, and H3K27 trimethylation (Roh et al., 2006). However, the method is limited by a relatively low resolution of 500–1000 bp and the considerable cost of sequencing in order to cover the entire genome. These limitations made a high-resolution map of histone modifications in the human genome unapproachable. To overcome these problems, we have employed the Solexa 1G Genome Analyzer, which performs massive parallel signature sequencing on unique molecular arrays, to directly sequence ChIP DNA from mononucleosomes generated by micrococcal nuclease (MNase) digestion of native chromatin. We demonstrate that this is a comprehensive, quantitative, and cost-effective method to analyze histone modification and protein target sites in large genomes such as the human genome. Our data indicate novel functions of histone methylation and chromatin organization in gene expression. The data are made available at http://dir.nhlbi.nih.gov/papers/lmi/epigenomes/hgtcell.html.

RESULTS

Direct Sequencing Analysis of ChIP DNA Samples Using Solexa 1G Genome Analyzer (ChIP-Seq)

To resolve histone modification signals to individual nucleosomes, we used mononucleosome templates generated by MNase digestion of native chromatin for ChIP. All the ChIP samples were confirmed using real-time quantitative PCR analysis of known target sites before further analysis using the Solexa 1G Genome Analyzer. The sequencing procedure requires a one-step adaptor ligation and limited PCR amplification (17 cycles) of ChIP DNA molecules, followed by cluster generation and sequencing-by-synthesis (Figure 1A). The image files generated by the analyzer are processed to produce DNA sequence data using the Solexa Analysis Pipeline (see Experimental Procedures). One sequencing run of this procedure can generate more than 20 million sequence tags of up to 36 bp each. A large majority of these short tags contain sufficient sequence information to be mapped unambiguously to the human genome. To distinguish between this procedure and our previous procedure, GMAT, or other similar procedures that incorporate the SAGE protocol, we have termed this direct sequencing procedure "ChIP-Seq." Since the ChIP-Seq method is analogous to direct counting of the molecules in the ChIP DNA samples, it requires minimal normalization. The number of tags detected for a particular nucleosome is directly proportional to the modification level of that nucleosome.

To determine the reproducibility and reliability of ChIP-Seq, we compared two independent experiments carried out using the H3K4me3 antibody by scatter analysis. The high degree of correlation shown in Figure 1B indicates that ChIP-Seq is indeed a reliable method for analyzing histone modifications in the human genome. To further determine whether ChIP-Seq can reproduce our previous GMAT results (Roh et al., 2006), we compared the H3K4me3 distribution patterns at a locus on chromosome 19 obtained from both methods. The data obtained using ChIP-Seq are largely consistent with the GMAT data (Figure 1C). However, the resolution with ChIP-Seq was dramatically higher. Furthermore, ChIP-Seq was more sensitive and generated less false-negative regions. For example, an H3K4me3 modification peak was found localized to the enhancer region downstream of the *IL-13* gene using ChIP-Seq, while no signal was detected using GMAT (data not shown). The *IL-2* gene, which is rapidly induced by T cell receptor signaling, exhibited a strong peak of H3K4me3 signals about 1.5 kb upstream of its TSS with the ChIP-Seq analysis, whereas only a single copy tag was detected in the same region using GMAT (Figure 1D). The improved sensitivity and specificity can be attributed to using native mononucleosomes for ChIP and obtaining 10-fold more tags.

Correlation of Histone Methylations at Promoter Region with Gene Expression

Gene promoters near transcription start sites contain critical regulatory elements necessary for transcription. Previous large-scale and genome-wide analyses have revealed that several histone modifications including H3K4 methylation and H3K9 acetylation are enriched in promoter regions, whereas H3K36me3 is elevated in the transcribed regions of active genes. To evaluate the contribution of other histone methylations and histone variants in human cells, we have now determined the genome-wide distribution at single-nucleosome resolution of 20 histone lysine and arginine methylations; one histone variant (H2A.Z); RNA polymerase II (Pol II); and the insulator binding protein, CTCF. To correlate these modifications with gene transcription, 12,726 human genes, whose expression levels in human resting $CD4^+$ T cells are known

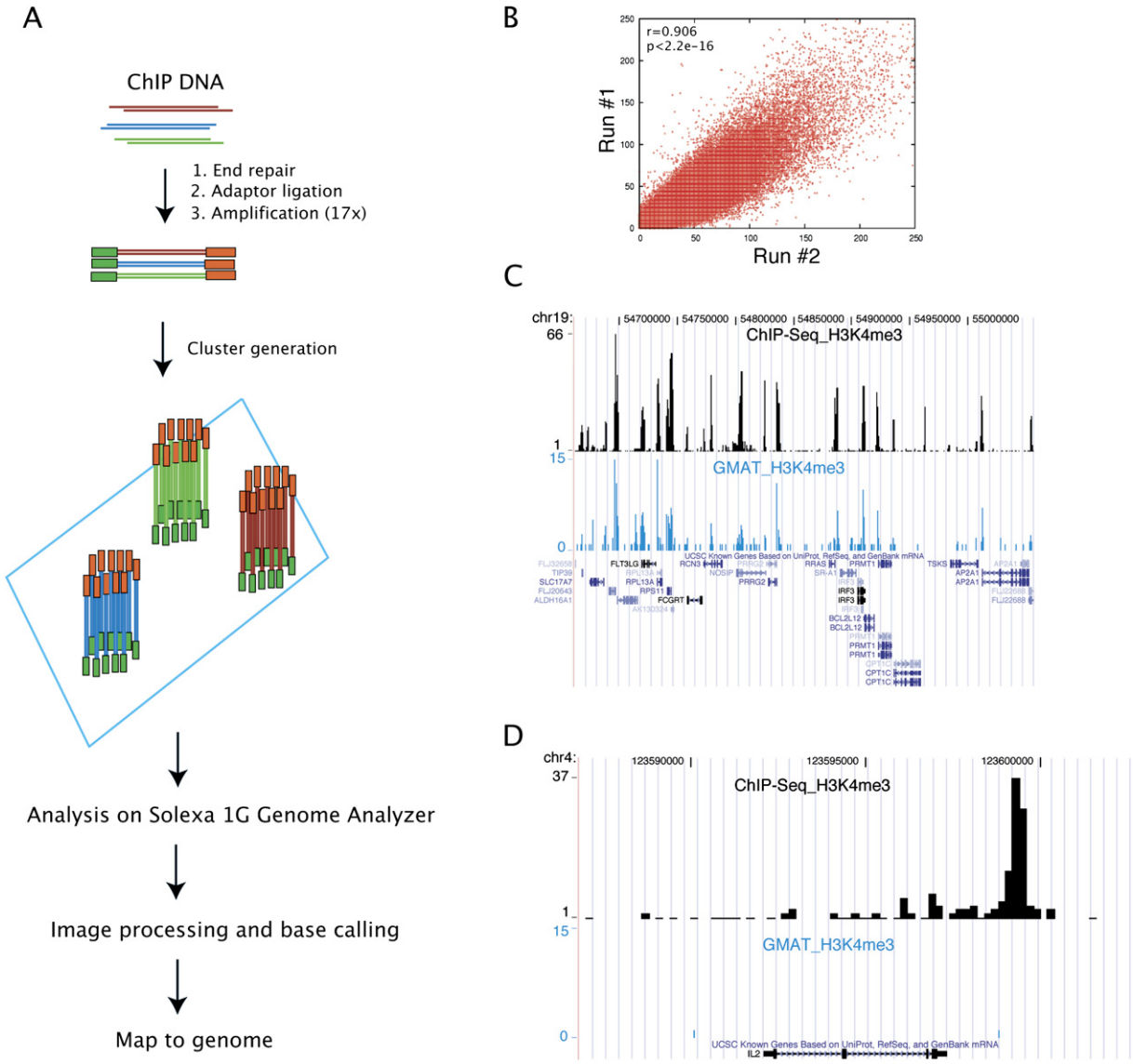


Figure 1. Analysis of Histone Modifications and Protein Location by ChIP and Direct Sequencing Using Solexa 1G Genome Analyzer (ChIP-Seq)

(A) The flow chart of ChIP-Seq.

(B) Scatter analysis of two independent analyses of H3K4me3 distribution. Pearon's product-moment correlation between the two samples: r = 0.906 (p-value < 2.2e-16).

(C) Comparison of H3K4me3 modification patterns in a 350 kb region on chromosome 19 using ChIP-Seq (upper panel) and using GMAT (lower panel). The data are displayed as custom tracks on the UCSC genome browser (Kent et al., 2002).

(D) Comparison of GMAT and ChIP-Seq data at the *IL-2* locus.

(Su et al., 2004), were separated into 12 groups of 1,000 genes according to their expression levels. The genes in each group were aligned relative to the TSS, and the detected tag numbers in 5 bp windows were calculated near the TSS for each group (see Experimental Procedures).

Pol II Binding

We identified 35,961 Pol II islands (see Experimental Procedures). As expected, Pol II binding was positively correlated with gene expression levels (Figure 2A). Interestingly, 37% of the silent promoters contained Pol II islands, suggesting that they may be poised for activation or retain a memory of past transcription. About 91% of all the Pol II islands were correlated with H3K4me3 islands (Figure S1).

H3K4 Methylations

All the three states of H3K4 methylation were elevated surrounding the TSSs of known genes (Figures 2B, 2C,

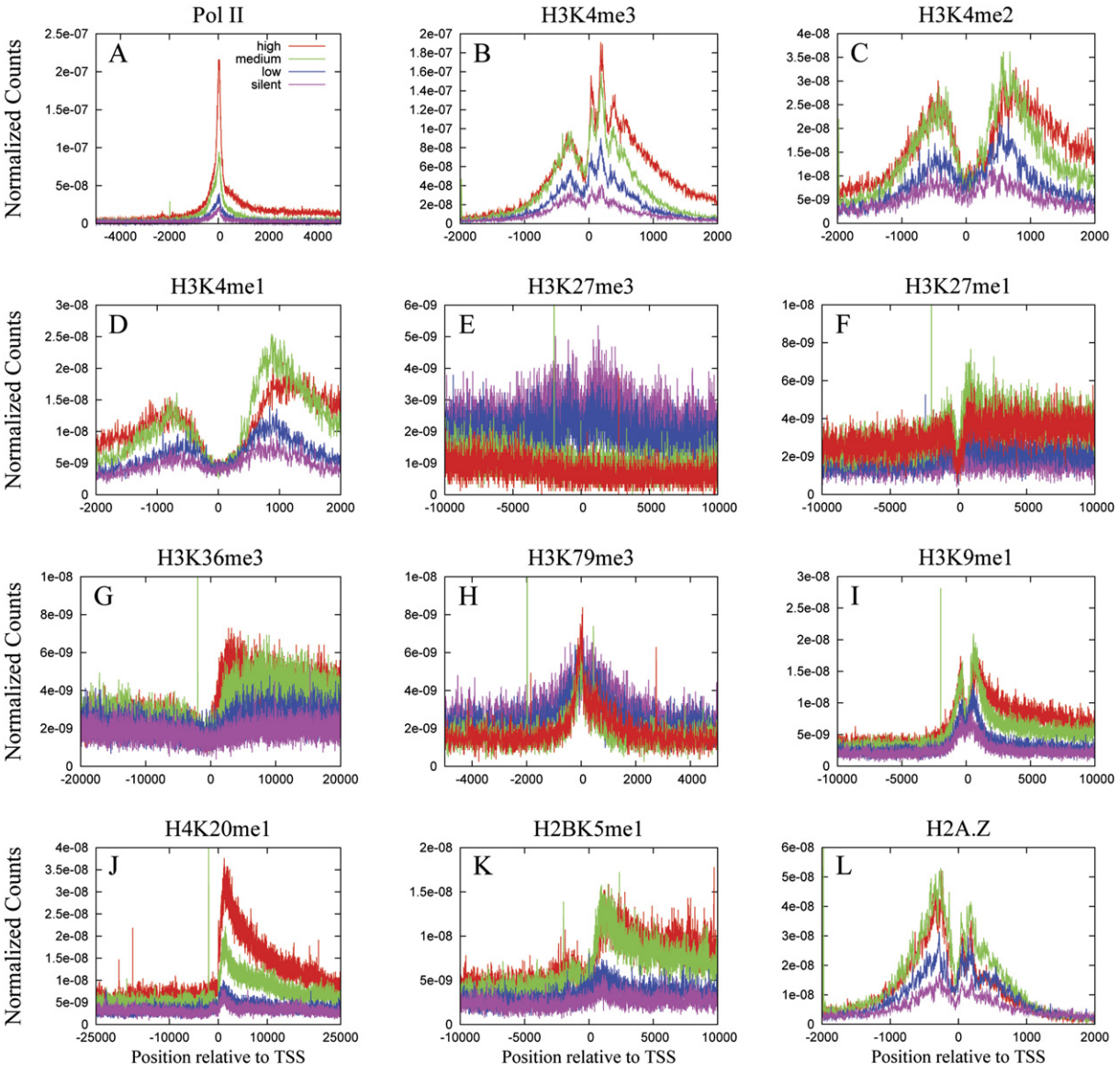


Figure 2. Histone Methylation near Transcription Start Sites

(A)–(L) Profiles of the histone methylation indicated above each panel across the TSS for highly active, two stages of intermediately active and silent genes are shown. Twelve thousand human genes were separated into twelve groups of one thousand genes according to their expression levels (see Experimental Procedures).

and 2D). H3K4me3 positively correlated with gene expression. A significant dip in the signal was observed between −200 to +50 for H3K4me3, which correlated with the nucleosome loss in active genes. A series of peaks of H4K4me3 signals at +50, +210, and +360 were detected, suggesting similar nucleosome positioning relative to TSS in active genes. Similar to Pol II binding, H3K4me3 islands were detected in 59% of silent promoters (Figure S1).

While the levels of H3K4me1 and H3K4me2 positively correlated with transcriptional levels, the signals downstream of TSSs were higher in the intermediately active group of genes than in the highly active genes. This might be caused by the high levels of H3K4me3 in the highly active group, since the three methylation states compete for the single lysine. Two major peaks were detected for each modification: −900 and +1000 for H3K4me1, −500 and +700 for H3K4me2, and −300 and +100 for H3K4me3. The signals are progressively more localized to the vicinity of TSSs as the modification moves from mono- to di- to trimethylation, which is consistent with the results from human HeLa cells (Heintzman et al., 2007).

H3K27 Methylations

Previous studies have suggested that methylation of H3K27 correlated with gene repression (Boyer et al., 2006; Lee et al., 2006; Roh et al., 2006). Indeed, H3K27me3 levels were higher at silent promoters than at active promoters (Figure 2E). The H3K27me3 signals were modestly elevated at silent promoter and reduced at active promoters and genic regions, whereas not much change was observed in intergenic regions. Unexpectedly, H3K27me1 signals were higher at active promoters than silent promoters, particularly downstream of the TSS (Figure 2F). H3K27me2 had a similar distribution with H3K27me3, though less biased toward silent genes (Figure S2C).

H3K9 Methylations

Methylation of H3K9 has been implicated in heterochromatin formation and gene silencing (Bannister et al., 2001). The TSS alignment indicates that the signals of both H3K9me2 and H3K9me3 were higher in silent genes than active genes in a region of 10 kb surrounding the TSS (Figures S2A and S2B). Surprisingly, higher H3K9me1 levels were detected in more active promoters surrounding the TSS (Figure 2I), suggesting that this modification may be associated with transcriptional activation.

H3K36 Methylations

H3K36me3 signals were sharply elevated after TSSs in active genes (Figure 2G), which is in agreement with a previous observation that H3K36me3 associates with actively transcribed regions (Bannister et al., 2005). However, H3K36me1 only showed a slight preference toward active promoters (Figure S2D).

H3K79 Methylations

H3K79me3 was more localized at active promoters than silent promoters, although the silent promoters appear to have higher H3K79me3 signals than active promoters except in a narrow region surrounding TSS (Figures 2H). H3K79me1 and H3K79me2 did not show significant preference toward either active or silent promoters (Figures S2E and S2F).

H3R2 Methylations

Methylation of H3R2 is catalyzed by CARM1, which coactivates nuclear hormone receptor-mediated transcription (An et al., 2004), suggesting that the H3R2 methylation may be involved in transcriptional activation. However, the TSS alignment did not show an enrichment of H3R2me1 or H3R2me2(as) in the promoters of active genes over the silent genes (Figures S2G and S2H).

H4K20 Methylations

Previous studies suggest association of H4K20 methylation with repressive chromatin. Recently, it was found that H4K20 mono- and trimethylation have a distinct distribution through the whole genome. H4K20me3 is associated with heterochromatin, while H4K20me1 is enriched in promoter or coding regions of active genes (Schotta et al., 2004; Talasz et al., 2005; Vakoc et al., 2006). H4K20me1 was also found colocalized with H3K9me1 (Sims et al., 2006). Our data indicate that H3K20me3 did not show an apparent association with either active or silent promoters (Figure S2I). However, H4K20me1 strongly correlated with gene activation in the regions downstream of the TSS (Figure 2J), consistent with it being an activation mark (Vakoc et al., 2006).

H4R3 Methylation

H4R3 can be methylated by the protein arginine methyltransferases PRMT1 and PRMT5. It has been suggested to be involved in nuclear receptor-mediated transcriptional activation (Lee et al., 2002). H4R3 methylation stimulates subsequent acetylation of H3 and H4, resulting in gene activation (Huang et al., 2005). The antibody we used recognizes both H2A and H4 (symmetric dimethyl R3). Our analysis indicates that these modifications did not show preference for either active or silent promoters (Figure S2J).

H2BK5 Methylation

The methylation of H2BK5 has not been reported previously. Our data show that H2BK5me1 associated with active promoters downstream of the TSS (Figure 2K), suggesting that it is a novel activation mark.

Histone Variant H2A.Z

H2A.Z, a histone H2A variant, is highly conserved through evolution. It was exclusively localized in the transcriptionally active macronucleus of the protozoan *T. thermophila* (Allis et al., 1980). In yeast, it is bound to regions near telomeres and large numbers of both active and silent yeast promoters and has been suggested to act as an antisilencing factor (Meneghini et al., 2003), reviewed by (Guillemette and Gaudreau, 2006). In higher eukaryotes, it is detected at insulators in the chicken β-globin locus (Bruce et al., 2005) and enriched in pericentric heterochromatin during the initial stages of mouse development (Rangasamy et al., 2003). Interestingly, our investigation indicates that H2A.Z was highly enriched at promoter regions both upstream and downstream of TSS, which is similar with its distribution in yeast (Raisner et al., 2005). However, in contrast to yeast, its binding levels correlated with gene activity in humans (Figure 2L) (also see Figure 4A).

Comparison between Gene Activity and Histone Methylations in Transcribed Regions

Our analyses indicate that levels of several modifications at promoters did not correlate with gene activity. Previous studies suggest that some modifications such as H3K36me3 are enriched only in actively transcribed regions in yeast (Bannister et al., 2005). To examine this in human T cells, we separated 12,726 human genes into different activity groups and plotted against their average modification levels in their transcribed regions (Figure 3; see Experimental Procedures).

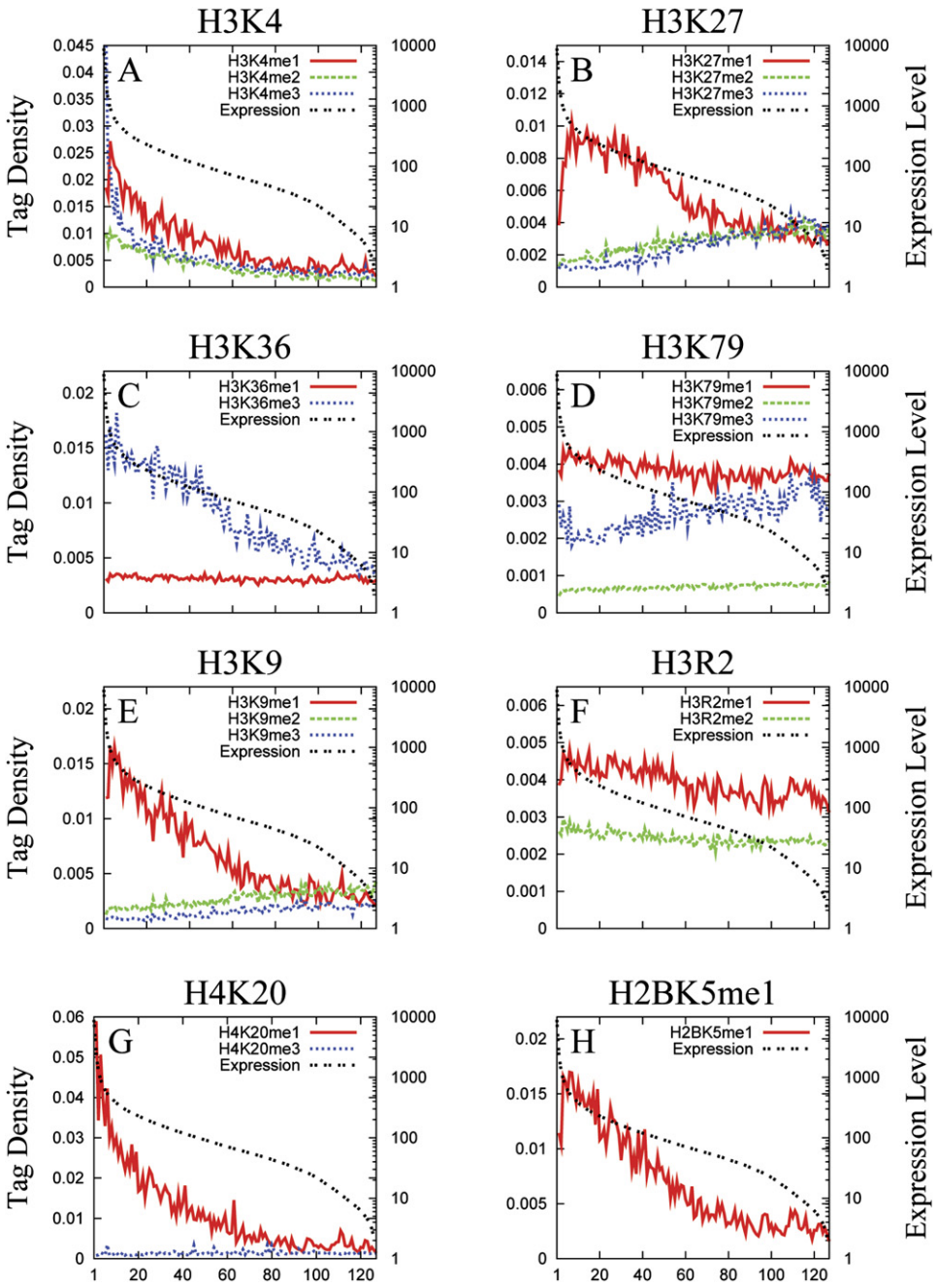


Figure 3. Correlation of Histone Methylations in Transcribed Regions with Expression Levels

(A)–(H) Genes were separated into groups of 100 genes based on their expression levels from high to low (left to right on the x axis). The right y axis indicates the expression level of each group. The left y axis indicates the histone methylation level that was derived by summing up all the detected tags within the entire transcribed gene-body regions and normalized to the total base pair number of the 100 genes. The histone methylation is indicated above each panel from (A) to (H).

Modifications Associated with Gene Silencing

We observed from this analysis that the levels of H3K27me2 and H3K27me3 correlated with gene silencing (Figure 3B). Also, a modest correlation between H3K9me3 and H3K9me2 levels and gene silence was observed (Figure 3E). However, we did detect highly localized H3K9me3 peaks in some active genes such as *STAT1* and *STAT4* (Figure 4A), which is consistent with previous observation (Squazzo et al., 2006). Unexpectedly, H3K79me3, which is associated with actively transcribed genes in yeast, modestly correlates with gene silencing in human T cells except for a small group of the most active genes that

were associated with higher levels of H3K79me3 (Figure 3D). In contrast to its association with the promoter regions, H2A.Z association within gene-body regions modestly correlates with gene silencing (Figure S3B).

Modifications Associated with Gene Activation

As expected, H3K4me1, H3K4me2, H3K4me3, and H3K36me3 signals correlated with gene activation (Figures 3A and 3C). Consistent with the TSS alignment analysis, H3K27me1, H3K9me1, H4K20me1, and H2BK5me1 modifications all positively correlated with the levels of gene expression (Figures 3B, 3E, 3G, and 3H). Both H3R2me1 and H3R2me2(as) modestly correlated with gene activation (Figure 3F).

Histone Methylation Patterns of Active and Silent Genes

Figure 4A shows a typical example of histone methylations of the genomic locus that contains the active *STAT1* and *STAT4* genes and silent *MYO1B* gene. The *STAT1* and *STAT4* genes were marked with high levels of H3K4me1, H3K4me2, H3K4me3, and Pol II at their promoter regions and high levels of H3K4me1, H3K4me2, H3K27me1, H3K9me1, H3K36me3, H2BK5me1, and H4K20me1 within their transcribed regions. Association of these activation marks was further exemplified by the *RUNX1* gene locus on Chromosome 21 (Figure S4A). In striking contrast, the *MYO1B* gene was associated with high levels of H3K27me2, H3K27me3, H3K9me2, and H3K79me3. Interestingly, H2A.Z was detected at high levels as localized peaks in the promoter regions of *STAT1* and *STAT4* and the intergenic region between *STAT4* and *MYO1B*, whereas more uniformly distributed but lower levels of H2A.Z were detected throughout the entire *MOY1B* gene. To extend these observations to more genes, we generated composite profiles of histone methylations for the 1000 most active and 1000 least active genes, respectively, spanning their gene bodies and extending 5 kb upstream and 5 kb downstream (Figures 4B and S4B). The tag numbers detected in every 5% of the gene-body region and every 1 kb outside of the gene-body region were summed to obtain methylation levels. These numbers were then normalized by the total number of base pairs in each region. It is notable that H2BK5me1, H3K27me1, H3K9me1, and H4K20me1 signals peaked near the 5′ end, whereas H3K36me3 signals peaked near the 3′ end of active genes.

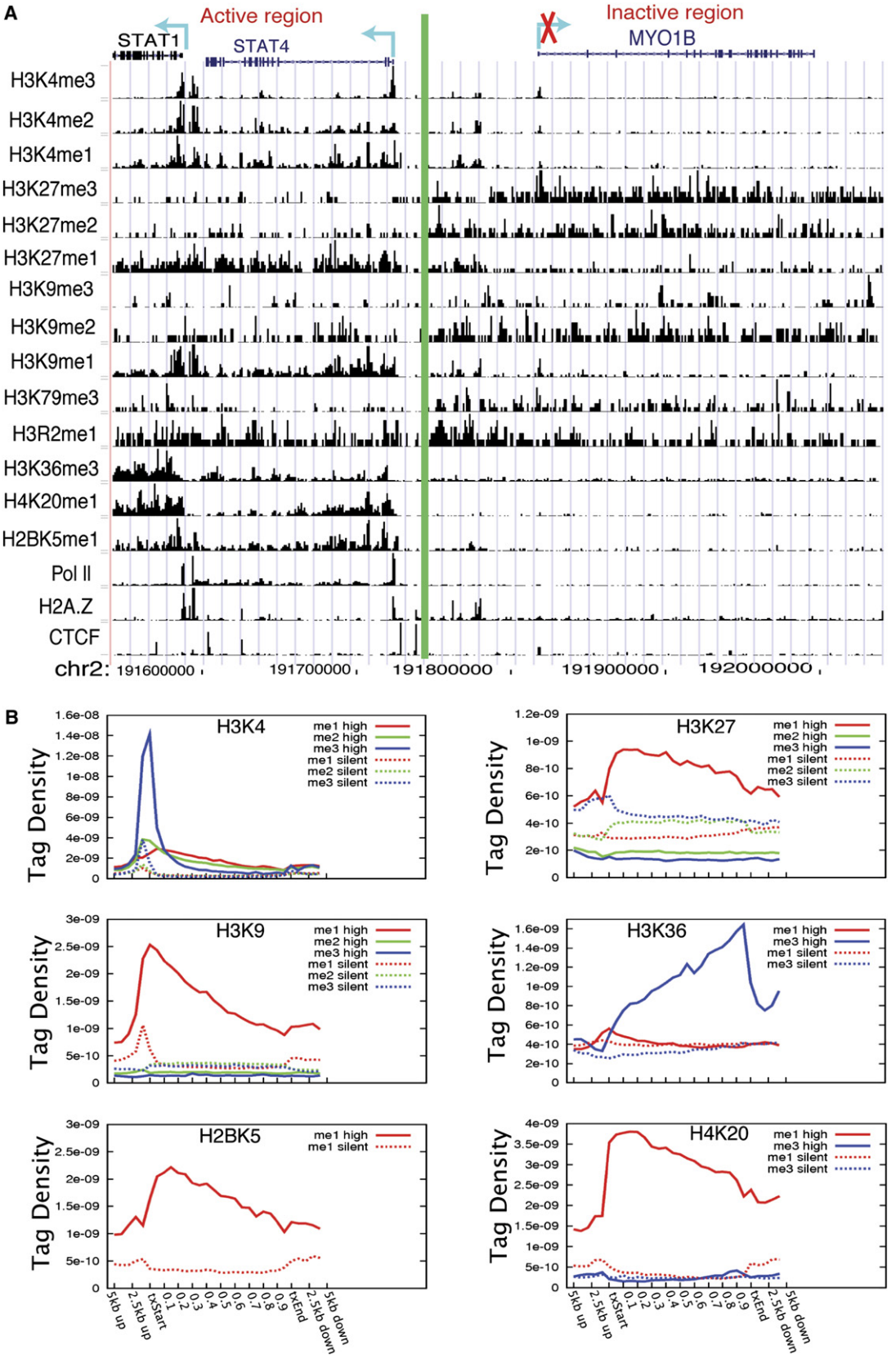
Histone Methylations at Insulators

Insulators are DNA elements that separate different chromatin domains. CTCF is a multi-zinc finger protein that binds the insulators in vertebrates (Gaszner and Felsenfeld, 2006; Kim et al., 2007). To investigate whether histone methylation-defined chromatin domains are demarcated by CTCF-bound insulators, we have determined the genome-wide distribution of CTCF in human $CD4^+$ T cells using ChIP-Seq. We identified 20,262 CTCF target sites, out of which 8,308 are located in intergenic regions, 6,305 in transcribed regions, and 5,649 within 2 kb of TSSs. Many of the CTCF-binding sites were indeed detected between active and silent chromatin domains. For example, the actively transcribed *PPP5C* gene is bracketed by repressive chromatin domains, marked with high levels of H3K27me3 signals. The CTCF-binding sites (highlighted in green boxes) clearly demarcate the boundaries of these two different chromatin domains (Figure 5A). Similarly CTCF binding marks distinct boundaries between the highly expressed *PAK4* gene region and the surrounding high H3K27me3 regions (Figure S5A). Another example is the binding of CTCF to the first intron of the *GNAS* gene, which appears to block the spreading of the upstream H3K27me3 signals and keeps the downstream alternative TSS active (Figure S5B). These observations are consistent with the idea that CTCF-bound insulators may functionally separate active from silent chromatin domains.

To identify the general histone methylation patterns at insulators, we aligned all the identified CTCF islands excluding those near Pol II islands and compared with all histone methylations. As shown in Figure 5B, all the three states of H3K4 methylation were highly enriched. Interestingly, H2A.Z was also highly enriched at insulators. It is interesting to note that another histone variant, H3.3, was reported at boundaries of *cis*-regulatory element domains in *Drosophila* (Mito et al., 2007). Unexpectedly, we found that H3K9me1 but not H3K9me2 and H3K9me3 were enriched at insulators. The percentage of the CTCF islands associated with various histone methylations is shown in the inset of Figure 5B.

Histone Methylations at Enhancers

We have previously used genome-wide histone H3 acetylation patterns to predict functional enhancers in human T cells with high confidence (Roh et al., 2005, 2007). A recent study has suggested high levels of H3K4me1 combined with low levels of H3K4me3 as a signature for predicting enhancers in HeLa cells (Heintzman et al., 2007). To better describe the histone modification patterns at enhancer elements, we first examined several known functional enhancers for their association with modified histones including H3K4me3, H3K4me1, and histone variant H2A.Z in T cells. Three distal enhancers, termed CNS1, CNS2, and CNS22, have been found to regulate the expression of interferon-γ in T cells (Hatton et al., 2006; Shnyreva et al., 2004). Our data indicate that each of these three regions was associated with H3K4me1, H3K4me3, and H2A.Z (Figure 5C). The acetylation island 1 enhancer downstream of the *IL-13* gene, which was previously identified using histone H3 acetylation patterns (Roh et al., 2005), also associated with H3K4me1, H3K4me3, and H2A.Z (Figure S6). In addition, all the known enhancers at the well-characterized *CD4* and *IL-2* receptor loci were associated with the three marks examined here (data not shown). DNase hypersensitivity has been successfully used as a marker of functional enhancer elements by many studies. To identify histone methylations associated with enhancers in an unbiased


A
Active region
Inactive region
STAT1
STAT4
MYO1B
H3K4me3
H3K4me2
H3K4me1
H3K27me3
H3K27me2
H3K27me1
H3K9me3
H3K9me2
H3K9me1
H3K79me3
H3R2me1
H3K36me3
H4K20me1
H2BK5me1
Pol II
H2A.Z
CTCF
chr2: 191600000 191700000 191800000 191900000 192000000
B
H3K4
H3K27
H3K9
H3K36
H2BK5
H4K20
Tag Density
me1 high
me2 high
me3 high
me1 silent
me2 silent
me3 silent
5kb up
2.5kb up
txStart
txEnd
2.5kb down
5kb down

way, we compared histone methylation patterns with a list of 14,184 DNase hypersensitive (HS) sites in human $CD4^+$ T cells (Crawford et al., 2006). Because insulators and promoters also show hypersensitivity to DNase digestion, we removed the HS sites that overlapped with CTCF and Pol II-binding sites and obtained 3,507 sites for further analysis. All three methylation states of H3K4 were elevated in these HS sites, with H3K4me2 and H3K4me3 more localized and H3K4me1 spanning a broader region. In addition, H2A.Z and H3K9me1 were highly enriched in these HS sites (Figure 5D). The percentage of DNase HS sites associated with each of these histone methylations is shown in the inset of Figure 5D.

Based on the analyses above, we summarize the epigenetic characteristics at insulators, enhancers, promoters, and transcribed regions of active and inactive genes (Figures 5E and 5F). Active genes are characterized by high levels of H3K4me1, H3K4me2, H3K4me3, H3K9me1, and H2A.Z surrounding TSSs and elevated levels of H2BK5me1, H3K36me3, H3K27me1, and H4K20me1 downstream of TSS and throughout the entire transcribed regions.

In contrast, inactive genes are characterized by low or negligible levels of H3K4 methylation at promoter regions, high levels of H3K27me3 and H3K79me3 in promoter and gene-body regions; low or negligible levels of H3K36me3, H3K27me1, K3K9me1, and H4K20me1 in gene-body regions; and uniformly distributed and low levels of H2A.Z.

Bivalent Domains with Both H3K4me3 and H3K27me3 in Differentiated Cells

H3K4me3 and H3K27me3 are a pair of "Yin-Yang" modifications with regard to gene activity. The bivalent domains, where H3K4me3 and H3K27me3 signals exist next to each other, are suggested to play regulatory roles for the differentiation of embryonic stem cells (Bernstein et al., 2006). We previously identified more than 3000 promoters that were associated with both H3K4me3 and H3K27me3 signals in differentiated T cells (Roh et al., 2006). However, whether differentiated cells have bivalent domains or not remains an unanswered question. Our current investigation confirmed that the expression of the promoters that were associated with both H3K4me3 and H3K27me3 modifications were significantly lower than the others in the genome (Figure S7A). While the H3K4me3 signals were highly localized to promoter and other regions, the H3K27me3 signals tended to be distributed over a broader region, in agreement with previous observations (Boyer et al., 2006; Lee et al., 2006; Roh et al., 2006). Figure S7B exemplifies the colocalization of the H3K4me3 and H3K27me3 signals in the intergenic region between the *RAB6IP2* and *FBXL14* genes. A blow-up map revealed that these two modifications actually occupied essentially nonoverlapping subregions. We identified 5736 such regions with overlapping H3K4me3 and H3K27me3 in resting $CD4^+$ T cells. These data suggest that there may be unexpected plasticity in the epigenetic memory of differentiated cells.

Chromosome Banding Patterns Correlate with Histone Methylation Patterns

Human chromosomes demonstrate a series of light (R) and dark (Q) bands when subjected to limited trypsin digestion and Giemsa staining. The R and Q bands are thought to represent separate compartments of the human genome, which are correlated with chromosome condensation, replication timing, and gene composition and activity (Craig and Bickmore, 1993). However, the molecular basis of their differences is not clear. We compared the banding pattern and histone methylation patterns and found that most band boundaries correlated with the transition of histone H3 lysine 4, 9, and 27 methylation patterns (Figure 6A). Q bands usually correlated with the depletion of H3K4 methylation and continuously widespread H3K9me3 signals, though the levels of H3K9me3 may not be necessarily higher than the peaks often found in active chromatin domains. In contrast, R bands usually correlated with high levels of H3K4 methylation and lower or more punctuate H3K9me3 signals. Because our analysis of histone modification were performed with resting T cells, which are in G0/G1 phase, and chromosome banding is assayed on mitotic cells, the strong correlation between histone modification and chromosome banding patterns indicates that at least some of these modifications are stable through mitosis. Gene desert regions are often located in Q bands and therefore associated with histone methylations typical of Q band patterns (Figure S8).

Although H3K27me3, H3K9me3, and H4K20me3 are all considered marks for transcriptional repression, our data showed that they can occupy both similar and distinct chromatin domains. For example, high levels of H3K27me3 could be found in both Q and R bands. However, it often demonstrated a complementary pattern with the H3K9me3 signals (Figures 6A and S8). H4K20me3 is colocalized with H3K9me3 in large repeats-related chromatin domains such as those near centromeric regions or *ZNF* repeats (Figures 6B and S9). Interestingly, most sharply localized H3K9me3 signals, which were often detected in active chromatin domains, were not colocalized with the H4K20me3 signals (Figures 6B and S9).

Chromosome Break Points Are Localized in Active Chromatin Domains

Chromosome translocation is a major cause of certain subtypes of lymphomas and leukemia. Specific chromosomal rearrangements frequently occur at some hot spots. It has been speculated that these specific break

Figure 4. Histone Methylation Patterns of Active and Inactive Genes

(A) A typical example of histone methylation patterns at active and inactive genes shown as custom tracks on the UCSC genome browser.

(B) Profiles of histone methylation patterns of 1000 highly active or silent genes across the gene bodies (see Experimental Procedures). Density plots extend 5 kb 5′ and 3′ of the gene bodies.

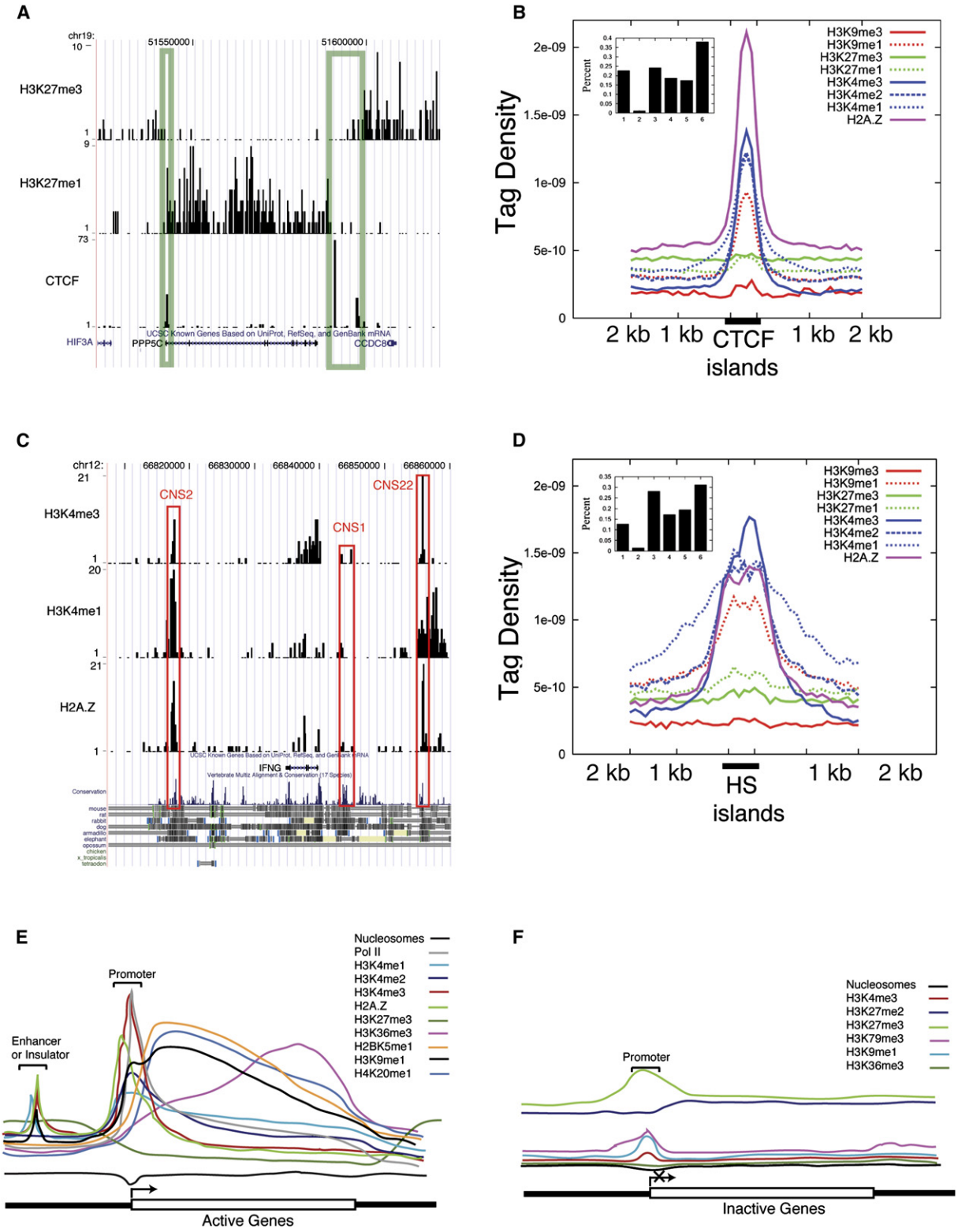


Figure 5. Epigenetic Modifications at Insulators and Enhancers

(A) CTCF binding marks the boundaries of active and inactive domains. H3K4me3, H3K27me1, H3K27me3, and CTCF binding were shown at the *PPP5C* locus.

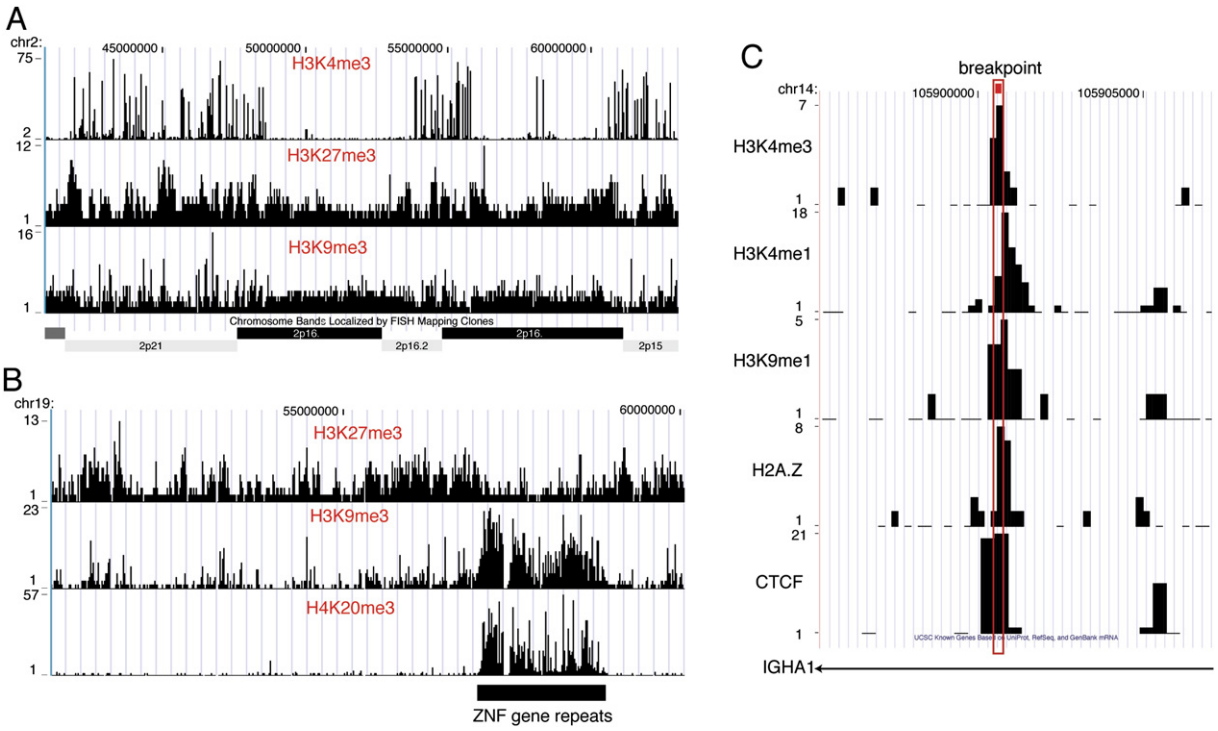


Figure 6. Correlation of Histone Methylation Patterns with Chromosomal Features

(A) Correlation of chromosome banding with histone methylations. Distribution of H3K4me3, H3K27me3, and H3K9me3 are shown as custom tracks on the UCSC genome browser.

(B) *ZNF* gene repeats are marked by high levels of H3K9me3 and H4K20me3.

(C) Chromosome breakpoints are located in regions with histone modifications associated with open chromatin. The breakpoint identified within the *IGHA1* gene associated with T cell cancers is highlighted in the red box.

points may be associated with open chromatin structure. Indeed, some breakpoints were found to colocalize with DNase hypersensitive sites (Zhang and Rowley, 2006). Thus, we decided to compare the locations of chromosome breakpoints listed in the Gross Rearrangement Breakpoint Database (Abeysinghe et al., 2004) with our histone methylation maps. Chromosome translocation within the IgA heavy chain gene (*IGHA1*) is linked to a number of leukemia and lymphomas. Interestingly, we found that the breakpoint associated with T cell lymphoma at the *IGHA1* locus is located within peaks of activating histone modifications, such as H3K4me1, H3K4me2, H3K4me3, and H3K9me1, whereas the breakpoints at this locus in non-T cell cancers are not associated with these activating modifications (Figure 6C). It is also associated with H2A.Z and CTCF (Figure 6C). Such peaks can mark promoters as well as putative enhancers and insulators. Among the 84 somatic breakpoints found in T cell cancers, 52 (62%) are colocalized with the H3K4me3 signals. In contrast, only 26% of the 180 breakpoints in non-T cells cancers are associated with H3K4me3. This result suggests that open chromatin structure maintained by the *MLL* family proteins and long-range chromatin organization by CTCF may contribute to chromosome breaks and may explain why certain translocations are found in the cancers of specific cell types.

(B) Histone modification patterns at CTCF-bound insulators. All the CTCF islands excluding those overlapping Pol II islands were aligned, and the tag densities of histone methylations were measured across the islands. The methylation patterns are also shown for 2 kb of DNA on each side of the island. (Inset shows the percentage of CTCF islands associated with different modifications. The numbers on the x axis correspond as follows: 1 = H2A.Z, 2 = H3K9me3, 3 = H3K9me1, 4 = H3K4me3, 5 = H3K4me2, 6 = H3K4me1).

(C) Association of H3K4me3, H3K4me1, and H2A.Z with known enhancers of the *IFN-γ* gene. The three known enhancers of *IFN-γ* were highlighted in red box and labeled as CNS1, CNS2, and CNS22. The data are displayed as custom tracks on the UCSC genome browser.

(D) Histone methylations patterns at enhancers represented by DNase HS sites (Inset shows the percentage of DNase islands associated with different modifications. The numbers on the x axis correspond as follows: 1 = H2A.Z, 2 = H3K9me3, 3 = H3K9me1, 4 = H3K4me3, 5 = H3K4me2, 6 = H3K4me1).

(E and F) Cartoons showing characteristic histone methylation patterns of active and silent genes. Promoter and enhancer or insulator regions are indicated.

Prediction of Novel Transcription Units Using the Histone Methylation Patterns

The accuracy of gene prediction is crucial to the understanding of the human genome. Gene annotation using multiple criteria including sequence analysis and experimental methods has resulted in the Ensembl prediction of 22,184 genes. However, the human UniGene database, Build 181, contains 52,924 UniGene clusters (Zhang et al., 2006). A recent study suggests the existence of a large number of uncharacterized intergenic transcripts (Cheng et al., 2005). However, it is not known whether they are products of leakage transcription or if they originate from novel TSSs of uncharacterized transcription units. Furthermore, most known genes have alternative TSSs, and it is not clear which one is used in a particular tissue. We investigated the possibility of confirming a known TSS or predicting novel TSSs using histone methylation patterns. As shown in Figure S5B, the upstream TSS of *GNAS* is associated with high levels of H3K27me3, suggesting that transcription from this TSS is inhibited. Instead, the downstream TSS is associated with high levels of H3K4me3 and H2A.Z and characterized by high levels of H2BK5me1, H3K36me3, H3K9me1, H3K27me1, and H4K20me1 in the transcribed region, which suggests that it is the active site of transcriptional initiation. Indeed, we observed high levels of Pol II binding of the downstream TSS (Figure S5B). In order to predict novel TSSs, we examined a few loci with characteristic patterns of histone methylations of an active TSS, which either colocalized with Unigene starts or do not colocalize with any known features and tested the existence of transcripts using RT-PCR. The *DHX37* gene was associated with high levels of H3K36me3 and H2BK5me1 (Figure 7A). However, only very low levels of Pol II binding and H3K4me3 were detected at its TSS. Moreover, the H3K36me3 level was higher near this TSS than the 3′ end of the gene, which is opposite from the expected pattern. In fact, there were strong signals about 10 kb downstream of the *DHX37* gene, which does not colocalize with any predicted gene units or transcripts. RT-PCR analysis shown in Figure 7C (lane 1) confirmed the existence of transcripts in the region highlighted in red. These data collectively suggest transcription is initiated from this TSS in the opposite direction of the *DHX37* gene. Another example is the novel TSS downstream of the *PRRG2* gene (Figure 7B), where transcripts were also detected (Figure 7C, lane 2). Figure S10 shows the strong signals of H3K4me3 and Pol II colocalized with a UniGene start, suggesting that transcription may be initiated from this site instead of a downstream known TSS of the *C12orf51* gene. Indeed, transcripts were detected downstream of this site but upstream of the known TSS (Figure 7C, lane 3).

DISCUSSION

Work using various model organisms has linked H3K4 methylation to promoter regions and H3K36 methylation to transcribed regions of active genes. Even though H4K20 methylation has been associated with heterochromatin formation and gene silencing, two recent reports suggested that H4K20me1 may be involved in transcriptional activation by being enriched in active promoters (Talasz et al., 2005) or in transcribed regions (Vakoc et al., 2006). Previous evidence indicates that H3K27me1 is enriched in heterochromatin (Peters et al., 2003) or distributed broadly in euchromatic region but selectively depleted in active promoter regions (Vakoc et al., 2006). Our data indicate that H3K27me1 and H4K20me1 are associated with actively transcribed regions. Moreover, H3K9me1 and H2BK5me1 are also associated with transcribed regions. H3K79me3 does not show a correlation with either active or silent genes in yeast (Pokholok et al., 2005), while H3K79me2 was linked to active transcription in *Drosophila* (Schubeler et al., 2004) and in humans (Okada et al., 2005). However, our data suggest that H3K79me1 is modestly associated with activation while H3K79me3 is associated with repression in human cells. H3K79me2 did not show any preference toward either active or silent genes. Therefore, actively transcribed regions are marked with high levels of H3K36me3, H3K27me1, H3K9me1, H4K20me1, and H2BK5me1. H3K36me3 exhibits a more 3′ distribution, whereas H3K9me1 and H2BK5me1 exhibit higher levels near the 5′ end, and H3K27me1 distribute more evenly throughout the transcribed regions. Evidence from yeast suggests that SET2 brings in the H3K36me3 modification, and it may be recruited by direct interaction with the phosphorylated C-terminal domain (Strahl et al., 2002). Though it is not clear which enzymes are responsible for the monomethylation of these histone lysines in T cells, our data suggest that their activities are recruited to actively transcribed regions, and one attractive hypothesis is that these enzymes may travel with the elongating RNA polymerase.

An active subject of research in the postgenomic era is to identify functional regulatory elements for human gene expression. Comparative genomics has been used to predict a set of highly conserved noncoding sequences as enhancers (Loots et al., 2000; Miller et al., 2004; Pennacchio and Rubin, 2001). However, this strategy has several limitations, including degeneracy of functional sequences and the fact that it cannot identify species-specific enhancers. Furthermore, it does not provide the information about the function of a potential enhancer in a particular tissue. Histone acetylation has been used to successfully predict a large set of conserved and nonconserved functional enhancers in human and mouse T cells (Roh et al., 2005, 2007). Several regions associated with H3K4me2 signals identified in a large-scale analysis were also suggested to be enhancers (Bernstein et al., 2005). A recent study on the 30 Mb human ENCODE regions suggests that enhancers are marked by monomethylation, but not trimethylation of H3K4 (Heintzman et al., 2007). To obtain a comprehensive methylation profile for enhancers, we analyzed the DNase HS sites in human $CD4^+$ T cells for their methylation patterns by removing all potential

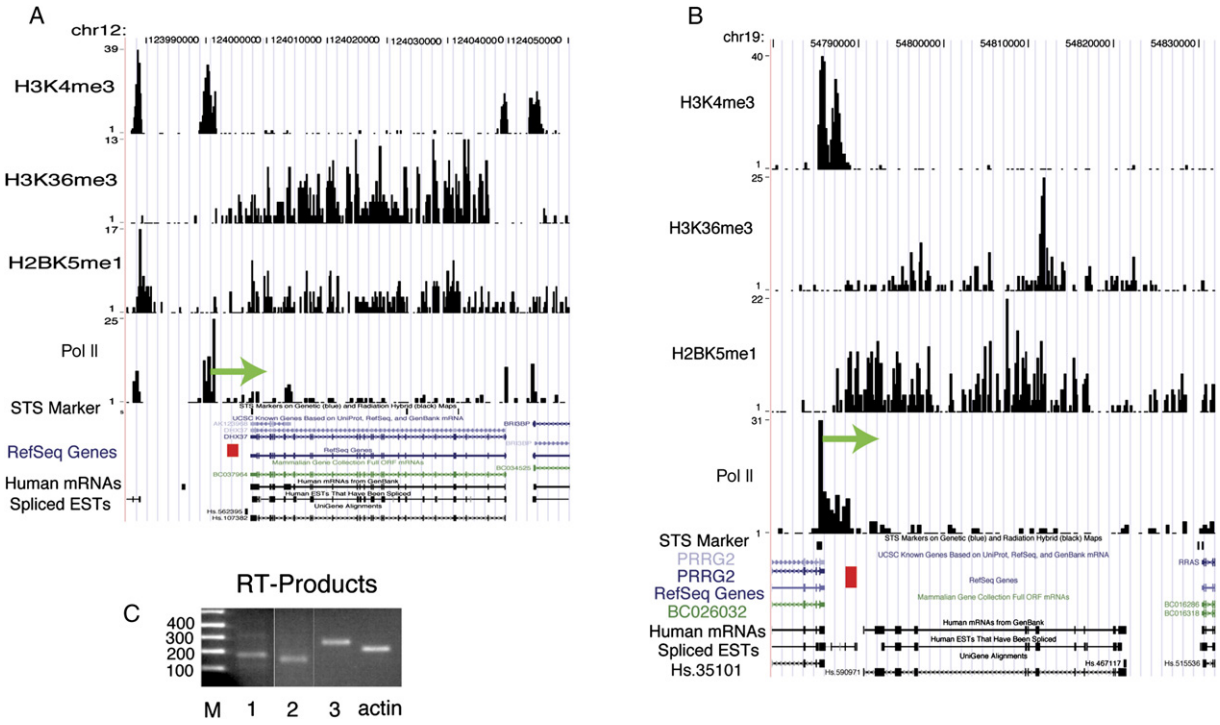


Figure 7. Prediction of Novel Transcription Units Using Epigenetic Profiles

(A) A potential novel TSS downstream of the *DHX37* gene. The H3K4me3, H3K36me3, H2BK5me1, and Pol II signals are shown. The potential direction of transcription is indicated by the arrow near the Pol II signals. The region tested for existence of transcripts is marked as a red box. The data are shown as custom tracks on the UCSC genome browser.

(B) A potential novel TSS downstream of the *PRRG2* gene.

(C) RT-PCR analysis of the suspected transcription units. Potential transcripts were analyzed by RT-PCR using primers indicated in panels (A) and (B) as well as in Figure S8. β-actin was used as a control. The DNA size makers are indicated as M.

promoter regions and insulator elements. In contrast to the previous report, our data show that all the three states of H3K4 methylation are highly enriched in these potential enhancers marked by DNase HS sites. Examination of the known enhancers of the *IFN-γ* and other genes confirmed this conclusion. In addition, high levels of H3K9me1 and H2A.Z are all enriched in the HS sites. These observations provide more support for the idea that the critical regulatory elements, such as enhancers, insulators, and promoters, are active sites undergoing various chromatin modifications (Mito et al., 2007).

What is the difference between an active enhancer and an active promoter? Though both the promoters and enhancers are associated with H3K4 methylation, H2A.Z, and H3K9me1, active promoters are characterized by high levels of H3K27me1, H3K36me3, H3K9me1, H4K20me1, and H2BK5me1 downstream of TSSs (see summary in Figure 5E). This particular pattern for active promoters can be used to identify novel transcription units or confirm transcription start sites for suspected genes, as we demonstrate in Figure 7.

In this study, we demonstrated that direct sequencing of ChIP DNA using the Solexa 1G Genome Analyzer is an efficient method for mapping genome-wide distributions of histone modifications and chromatin protein target sites. The technique is particularly well suited for analyzing protein locations and chromatin modification in large genomes such as the human genome. This cost-effective method produces digital-quality data and should find broad applications in our efforts to understand the contribution of the human epigenomes in gene expression and epigenetic inheritance. Indeed, we have already used it here to provide the most comprehensive genome-wide data for more than 20 epigenetic marks in human T cells.

EXPERIMENTAL PROCEDURES

ChIP and Template Preparation for Sequencing Analysis

$CD4^+$ T cells were purified from human blood using human $CD4^+$ T cell isolation kit II kits (Miltenyi). The cells were digested with MNase to generate mainly mononucleosomes with minor fraction of dinucleosomes for histone modification mapping. For mapping enzyme target sites, the cells were crosslinked with formaldehyde treatment and chromatin fragmented to 200 to 300 bp by sonication. Chromatin from 2×10^7 cells was used for each ChIP experiment, which yielded approximately 200 ng of DNA. The ChIP DNA ends were repaired using PNK and Klenow enzyme, followed by treatment with Taq polymerase to generate a protruding 3′ A base used for adaptor ligation. Following

ligation of a pair of Solexa adaptors to the repaired ends, the ChIP DNA was amplified using the adaptor primers for 17 cycles and the fragments around 220 bp (mononucleosome + adaptors) isolated from agarose gel. The purified DNA was used directly for cluster generation and sequencing analysis using the Solexa 1G Genome Analyzer following manufacturer protocols. All the antibodies used and the total tag number obtained for each sample are listed in Table S1.

Data Analysis

Solexa Pipeline Analysis

Sequence tags were obtained and mapped to the human genome using the Solexa Analysis Pipeline. The output of the Solexa Analysis Pipeline was converted to browser extensible data (BED) files for viewing the data in the UCSC genome browser. Data for each of the histone modifications is presented in BED files detailing the genomic coordinates of each tag as well as in BED files detailing summary windows displaying the number of tags in 200 bp windows. Summary files for CTCF and RNA Polymerase II were created using 400 bp windows.

Histone Methylation near Transcription Start Sites

Expression information was obtained for $CD4^+$ T cells from the GNF SymAtlas gene expression atlas (Su et al., 2004). GNF expression probes were mapped to UCSC Known Genes to obtain expression information in $CD4^+$ T cells for 12,726 known genes. These genes were broken up into 12 sets of 1000 genes by ranking the genes by expression levels and assigning genes to sets of 1000. The four sets shown in Figures 2 and S2 correspond to highly expressed and two stages of intermediately expressed and silent genes. The genes in each set were aligned relative to the TSS. For each modification and enzyme, the tag density (number of tags per base pair) was calculated in 5 bp windows relative to the TSS.

Correlation of Histone Methylations in Transcribed Regions with Expression Levels

The 12,726 UCSC known genes for which expression information was obtained were broken up into 127 groups of 100 genes based on expression. The average tag density of each modification within the gene bodies for genes within each set was calculated.

Tag Islands

In an attempt to eliminate background noise signals from our analysis, we defined "islands" of tags for each of the modifications as well as RNA Polymerase II and CTCF signals. An island was defined for each sample by first finding all summary windows that have more than one tag and then grouping consecutive windows above this threshold and allowing a distance of one window of empty signal. Lastly, all islands that did not contain at least eight tags were eliminated.

Bivalent Promoters

All overlapping islands of H3K4me3 and H3K27me3 were found and defined as bivalent domains (Bernstein et al., 2006; Roh et al., 2006). Genes with bivalent domains within 1 kb of their TSS were classified as having bivalent promoters.

Supplemental Data

Supplemental Data include one table, ten figures, and Supplemental References and can be found with this article online at http://www.cell.com/cgi/content/full/129/4/823/DC1/.

ACKNOWLEDGMENTS

We thank Dr. Warren Leonard for critical reading of the manuscript. This work was supported by the Intramural Research Program of the NIH, National Heart, Lung, and Blood Institute.

Received: April 20, 2007
Revised: May 3, 2007
Accepted: May 3, 2007
Published: May 17, 2007

REFERENCES

Abeysinghe, S.S., Stenson, P.D., Krawczak, M., and Cooper, D.N. (2004). Gross Rearrangement Breakpoint Database (GRaBD). Hum. Mutat. *23*, 219–221.

Allis, C.D., Glover, C.V., Bowen, J.K., and Gorovsky, M.A. (1980). Histone variants specific to the transcriptionally active, amitotically dividing macronucleus of the unicellular eucaryote, Tetrahymena thermophila. Cell *20*, 609–617.

An, W., Kim, J., and Roeder, R.G. (2004). Ordered cooperative functions of PRMT1, p300 and CARM1 in transcriptional activation by p53. Cell *117*, 735–748.

Bannister, A.J., Schneider, R., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2005). Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes. J. Biol. Chem. *280*, 17732–17736.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

Bernstein, B.E., Humphrey, E.L., Erlich, R.L., Schneider, R., Bouman, P., Liu, J.S., Kouzarides, T., and Schreiber, S.L. (2002). Methylation of histone H3 Lys 4 in coding regions of active genes. Proc. Natl. Acad. Sci. USA *99*, 8695–8700.

Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. Cell *120*, 169–181.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bernstein, B.E., Meissner, A., and Lander, E.S. (2007). The mammalian epigenome. Cell *128*, 669–681.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

Bruce, K., Myers, F.A., Mantouvalou, E., Lefevre, P., Greaves, I., Bonifer, C., Tremethick, D.J., Thorne, A.W., and Crane-Robinson, C. (2005). The replacement histone H2A.Z in a hyperacetylated form is a feature of active genes in the chicken. Nucleic Acids Res. *33*, 5633–5639.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., et al. (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science *308*, 1149–1154.

Craig, J.M., and Bickmore, W.A. (1993). Chromosome bands – flavours to savour. Bioessays *15*, 349–354.

Crawford, G.E., Holt, I.E., Whittle, J., Webb, B.D., Tai, D., Davis, S., Margulies, E.H., Chen, Y., Bernat, J.A., Ginsburg, D., et al. (2006). Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). Genome Res. *16*, 123–131.

Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. Nat. Rev. Genet. *7*, 703–713.

Guillemette, B., and Gaudreau, L. (2006). Reuniting the contrasting functions of H2A.Z. Biochem Cell Biol *84*, 528–535.

Hatton, R.D., Harrington, L.E., Luther, R.J., Wakefield, T., Janowski, K.M., Oliver, J.R., Lallone, R.L., Murphy, K.M., and Weaver, C.T. (2006). A distal conserved sequence element controls Ifng gene expression by T cells and NK cells. Immunity *25*, 717–729.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007).

Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Huang, S., Litt, M., and Felsenfeld, G. (2005). Methylation of histone H4 by arginine methyltransferase PRMT1 is essential in vivo for many subsequent histone modifications. Genes Dev. *19*, 1885–1893.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The Human Genome Browser at UCSC. Genome Res. *12*, 996–1006.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005). A high-resolution map of active promoters in the human genome. Nature *436*, 876–880.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231–1245.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Lee, Y.H., Koh, S.S., Zhang, X., Cheng, X., and Stallcup, M.R. (2002). Synergy among nuclear receptor coactivators: selective requirement for protein methyltransferase and acetyltransferase activities. Mol. Cell. Biol. *22*, 3621–3632.

Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D., and Lieb, J.D. (2004). Evidence for nucleosome depletion at active regulatory regions genome-wide. Nat. Genet. *36*, 900–905.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301–313.

Liu, C.L., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S.L., Friedman, N., and Rando, O.J. (2005). Single-nucleosome mapping of histone modifications in S. cerevisiae. PLoS Biol. *3*, e328.

Loots, G.G., Locksley, R.M., Blankespoor, C.M., Wang, Z.E., Miller, W., Rubin, E.M., and Frazer, K.A. (2000). Identification of a coordinate regulator of interleukins *4*, 13 and 5 by cross-species sequence comparisons. Science *288*, 136–140.

Meneghini, M.D., Wu, M., and Madhani, H.D. (2003). Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. Cell *112*, 725–736.

Miller, W., Makova, K.D., Nekrutenko, A., and Hardison, R.C. (2004). Comparative genomics. Annu. Rev. Genomics Hum. Genet. *5*, 15–56.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. Science *315*, 1408–1411.

Okada, Y., Feng, Q., Lin, Y., Jiang, Q., Li, Y., Coffield, V.M., Su, L., Xu, G., and Zhang, Y. (2005). hDOT1L links histone methylation to leukemogenesis. Cell *121*, 167–178.

Pennacchio, L.A., and Rubin, E.M. (2001). Genomic strategies to identify mammalian regulatory sequences. Nat. Rev. Genet. *2*, 100–109.

Peters, A.H., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., et al. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol. Cell *12*, 1577–1589.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genome-wide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517–527.

Raisner, R.M., Hartley, P.D., Meneghini, M.D., Bao, M.Z., Liu, C.L., Schreiber, S.L., Rando, O.J., and Madhani, H.D. (2005). Histone variant H2A.Z marks the 5′ ends of both active and inactive genes in euchromatin. Cell *123*, 233–248.

Rangasamy, D., Berven, L., Ridgway, P., and Tremethick, D.J. (2003). Pericentric heterochromatin becomes enriched with H2A.Z during early mammalian development. EMBO J. *22*, 1599–1607.

Roh, T.Y., Cuddapah, S., and Zhao, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev. *19*, 542–552.

Roh, T.Y., Ngau, W.C., Cui, K., Landsman, D., and Zhao, K. (2004). High-resolution genome-wide mapping of histone modifications. Nat. Biotechnol. *22*, 1013–1016.

Roh, T.Y., Cuddapah, S., Cui, K., and Zhao, K. (2006). The genomic landscape of histone modifications in human T cells. Proc. Natl. Acad. Sci. USA *103*, 15782–15787.

Roh, T.Y., Wei, G., Farrell, C.M., and Zhao, K. (2007). Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. Genome Res. *17*, 74–81.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004). A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. Genes Dev. *18*, 1251–1262.

Schubeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., Delrow, J., et al. (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev. *18*, 1263–1271.

Sekinger, E.A., Moqtaderi, Z., and Struhl, K. (2005). Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. Mol. Cell *18*, 735–748.

Shnyreva, M., Weaver, W.M., Blanchette, M., Taylor, S.L., Tompa, M., Fitzpatrick, D.R., and Wilson, C.B. (2004). Evolutionarily conserved sequence elements that positively regulate IFN-gamma expression in T cells. Proc. Natl. Acad. Sci. USA *101*, 12622–12627.

Sims, J.K., Houston, S.I., Magazinnik, T., and Rice, J.C. (2006). A trans-tail histone code defined by monomethylated H4 Lys-20 and H3 Lys-9 demarcates distinct regions of silent chromatin. J. Biol. Chem. *281*, 12760–12766.

Squazzo, S.L., O'Geen, H., Komashko, V.M., Krig, S.R., Jin, V.X., Jang, S.W., Margueron, R., Reinberg, D., Green, R., and Farnham, P.J. (2006). Suz12 binds to silenced regions of the genome in a cell-type-specific manner. Genome Res. *16*, 890–900.

Strahl, B.D., Grant, P.A., Briggs, S.D., Sun, Z.W., Bone, J.R., Caldwell, J.A., Mollah, S., Cook, R.G., Shabanowitz, J., Hunt, D.F., and Allis, C.D. (2002). Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol. Cell. Biol. *22*, 1298–1306.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. USA *101*, 6062–6067.

Talasz, H., Lindner, H.H., Sarg, B., and Helliger, W. (2005). Histone H4-lysine 20 monomethylation is increased in promoter and coding - regions of active genes and correlates with hyperacetylation. J. Biol. Chem. *280*, 38814–38822.

Vakoc, C.R., Sachdeva, M.M., Wang, H., and Blobel, G.A. (2006). Profile of histone lysine methylation across transcribed mammalian chromatin. Mol. Cell. Biol. *26*, 9185–9195.

Yuan, G.C., Liu, Y.J., Dion, M.F., Slack, M.D., Wu, L.F., Altschuler, S.J., and Rando, O.J. (2005). Genome-scale identification of nucleosome positions in S. cerevisiae. Science *309*, 626–630.

Zhang, Y., and Rowley, J.D. (2006). Chromatin structural elements and chromosomal translocations in leukemia. DNA Repair (Amst.) *5*, 1282–1297.

Zhang, J., Zhang, L., and Coombes, K.R. (2006). Gene sequence signatures revealed by mining the UniGene affiliation network. Bioinformatics *22*, 385–391.