

The genomic landscape of histone modifications in human T cells

Tae-Young Roh, Suresh Cuddapah, Kairong Cui, and Keji Zhao*

Laboratory of Molecular Immunology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD 20892

Communicated by Carl Wu, National Institutes of Health, Bethesda, MD, September 5, 2006 (received for review June 23, 2006)

To understand the molecular basis that supports the dynamic gene expression programs unique to T cells, we investigated the genomic landscape of activating histone modifications, including histone H3 K9/K14 diacetylation (H3K9acK14ac), H3 K4 trimethylation (H3K4me3), and the repressive histone modification H3 K27 trimethylation (H3K27me3) in primary human T cells. We show that H3K9acK14ac and H3K4me3 are associated with active genes required for T cell function and development, whereas H3K27me3 is associated with silent genes that are involved in development in other cell types. Unexpectedly, we find that 3,330 gene promoters are associated with all of these histone modifications. The gene expression levels are correlated with both the absolute and relative levels of the activating H3K4me3 and the repressive H3K27me3 modifications. Our data reveal that rapidly inducible genes are associated with the H3 acetylation and H3K4me3 modifications, suggesting they assume a chromatin structure poised for activation. In addition, we identified a subpopulation of chromatin regions that are associated with high levels of H3K4me3 and H3K27me3 but low levels of H3K9acK14ac. Therefore, these regions have a distinctive chromatin modification pattern and thus may represent a distinct class of chromatin domains.

chromatin landscape | epigenome | gene expression | genome-wide mapping

Numerous covalent modifications on histones have been discovered, most of which occur at the N-terminal tails of histones (1). Depending on the nature and position of the modification, they are linked to different gene activities. For example, acetylation of histone H3 on lysines 9 and 14 and methylation on lysines 4, 36, and 79 (activating marks) are linked to active transcription, whereas methylation on lysines 9 and 27 (repressive marks) is linked to transcriptional repression (1–3). Histone acetylation and deacetylation are highly dynamic processes that are controlled by the counteracting action of histone acetyltransferases (HATs) and histone deacetylases (HDACs) (4). Histone methylation catalyzed by histone methyltransferases (HMTs) is believed to be a more stable epigenetic mark and can be associated with either gene activation or repression, depending on the nature and position of the modification (5). Activating marks such as histone acetylation and H3 K4 trimethylation (H3K4me3) are colocalized in active chromatin (6–9), whereas repressive marks, including H3 K9 trimethylation (H3K9me3) and H3 K27 trimethylation (H3K27me3), are found in inactive chromatin (10–12). Different hypotheses, including histone code and signaling platform models, have been proposed for mechanisms used in regulation of gene expression by histone modifications (13–16).

Human resting T cells circulating in the bloodstream have a unique gene expression program that distinguishes them from other cell types (17). Activation of T cells upon encountering specific antigens rapidly induces or represses thousands of genes. Accompanying the gene expression changes during T cell activation and differentiation, extensive changes in histone modifications have been detected in several lymphocyte-specific genes (18–24). To understand the relationship between histone modifications and gene expression on a whole-genome scale, we have

investigated the landscape of genome-wide histone modifications, including both the activating marks H3K9acK14ac and H3K4me3 and the repressive mark H3K27me3, by using an unbiased genome-wide mapping technique (GMAT) that combines the chromatin immunoprecipitation (ChIP) and serial analysis of gene expression (SAGE) protocols (25, 26). Our data indicate that constitutively expressed or rapidly inducible genes, which are required for T cell development and function, are associated with high levels of H3 acetylation and H3K4me3 signals, whereas permanently silenced genes, which are involved in the development of other cell types, are associated with high levels of H3K27me3 signals. Unexpectedly, we found that not only are the H3K4me3 and H3K27me3 modifications colocalized at thousands of promoters in differentiated human T cells, but also both their absolute and relative levels correlate with the expression levels of these genes. Our data suggest that these histone modifications together may define the chromatin landscape important for gene expression in T cells.

Results

Both Activating and Repressive Epigenetic Marks Are Enriched in the Promoter Region. GMAT analysis involves sequencing ChIP DNA samples by using the long-serial analysis of gene expression protocol (SAGE), which detects 21-bp sequence tags from each DNA fragment pulled-down by specific antibodies. The detection frequency of each tag represents the level of a particular histone modification at the genomic locus containing the sequence tag. Our previous analysis with GMAT revealed that histone acetylation islands are correlated with transcription/chromatin regulatory elements, including promoters, enhancers, and locus control regions (LCRs) (25). We now have determined the distribution of other important histone modifications, including H3K4me3 and H3K27me3, in the entire genome of human primary T cells. A comprehensive epigenomic map was constructed based on these data (see Fig. 7, which is published as supporting information on the PNAS web site; a searchable and interactive high-resolution map can be found at <http://dir.nhlbi.nih.gov/papers/lmi/epigenomes>).

Activating marks, including H3K9acK14ac and H3K4me3, are concentrated in promoter regions in the human genome (7, 9, 25). To compare the distribution between these activating marks to those of repressive marks, we aligned 21,355 annotated genes relative to their transcription start sites (TSSs) and plotted the levels of modification across a 10-kb region. Interestingly, all marks examined were enriched in a 2-kb promoter region surrounding the TSSs (Fig. 1 *A* and *B*), indicating that the

Author contributions: T.-Y.R. and S.C. contributed equally to this work. T.-Y.R., S.C., and K.Z. designed research; T.-Y.R., S.C., and K.Z. performed research; K.C. contributed new reagents/analytic tools; T.-Y.R., S.C., K.C., and K.Z. analyzed data; and T.-Y.R., S.C., and K.Z. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Abbreviations: GMAT, genome-wide mapping technique; ChIP, chromatin immunoprecipitation; TSS, transcription start site; TCR, T cell receptor.

*To whom correspondence should be addressed at: Laboratory of Molecular Immunology, National Heart, Lung, and Blood Institute, National Institutes of Health, Building 10, Room 7N311, 9000 Rockville Pike, Bethesda, MD 20892. E-mail: zhaok@nhlbi.nih.gov.

suggesting that they need to be repressed for normal T cell development. We note that many critical transcription factors involved in development also are associated with H3K27me3 in human ES cells (28). Interestingly, only about one-third of K27-methylated regions in the human ES cells overlapped with highly K27-methylated regions in T cells, which mainly contained the genes involved in transcription regulation (data not shown). Many genes encoding cell membrane proteins involved in receptor activity were associated with high levels of H3K27me3 signals in T cells but not in ES cells (data not shown). Our results therefore reveal that the genes involved in T cell development and function are associated with the H3K4me3 modification, whereas the genes involved in development of other cell types are associated with the H3K27me3 modification.

Rapidly Inducible Genes Are Associated with H3 Acetylation and H3K4me3 Modifications. Human genes can be classified into expressed and nonexpressed genes; however, some of the expressed genes are constitutively expressed, whereas others need to be turned off during development or in response to environmental stimuli. Analogously, some of the nonexpressed genes are constitutively silent in certain cell lineages because their ectopic expression might be very toxic, whereas other nonexpressed genes may need to be rapidly turned on during development or in response to stimuli. The complex and diverse patterns of gene expression may be associated with diverse patterns of histone modifications. For example, the constitutively expressed house-keeping β -actin gene exists in a highly open chromatin structure characterized by high levels of H3 acetylation and H3K4me3 throughout the locus (data not shown). The *NFATC1*, *NFATC2*, and *NFATC3* genes, which are required for T cell activation and differentiation (29, 30), are transcribed only at minimal levels in resting T cells but are rapidly induced by TCR signaling. Interestingly, they were associated with high levels of H3K4me3 signals, especially at their promoter regions (Fig. 3A–C). It also is interesting to note that we detected clusters of H3K4me3 signals near the 3' ends and in the transcribed regions of these *NFAT* genes, which may denote regulatory elements for these genes. In contrast, *NFATC4*, which is not expressed in T cells, is associated with only low levels of H3K4me3 but high levels of H3K27me3 signals (Fig. 3D and data not shown). The high levels of the H3K4me3 signal of the *NFATC1*, *NFATC2*, and *NFATC3* genes suggest that their minimal levels of transcription in resting T cells may result from a “lack of activation.” When an activator becomes available, as in the case of TCR signaling, their transcription can be rapidly activated. In contrast, the silent *NFATC4* and *NeuroD* genes were associated with high levels of H3K27me3 signals (Fig. 3D and data not shown) to ensure their silent states in T cells. Together, these results are consistent with the notion that the chromatin modified with H3K4me3 and H3 acetylation is competent for transcription, and the chromatin marked with H3K27me3 is incompetent for transcription.

Genome-Wide Colocalization Analysis of the H3K9acK14ac, H3K4me3, and H3K27me3 Modifications. To examine the location relationship between these different histone modifications in an unbiased way for the entire human genome in T cells, we analyzed whole-genome scattergrams of pairwise marks by comparing their average modification levels in 3-kb sequence windows. The results show that $\approx 90\%$ of the H3K9acK14ac and H3K4me3 signals were colocalized in the human genome (Fig. 4A). Circle I highlighted 14,528 such genomic regions. Interestingly, a distinct population of 357 chromatin regions was associated with high levels of H3K4me3 but with no or low levels of H3K9acK14ac (Fig. 4A, highlighted in circle II). Because these regions are far from any identified genes, it is possible that H3K4me3 has a function in addition to its participation in transcriptional activation of promoters. The scatter analysis

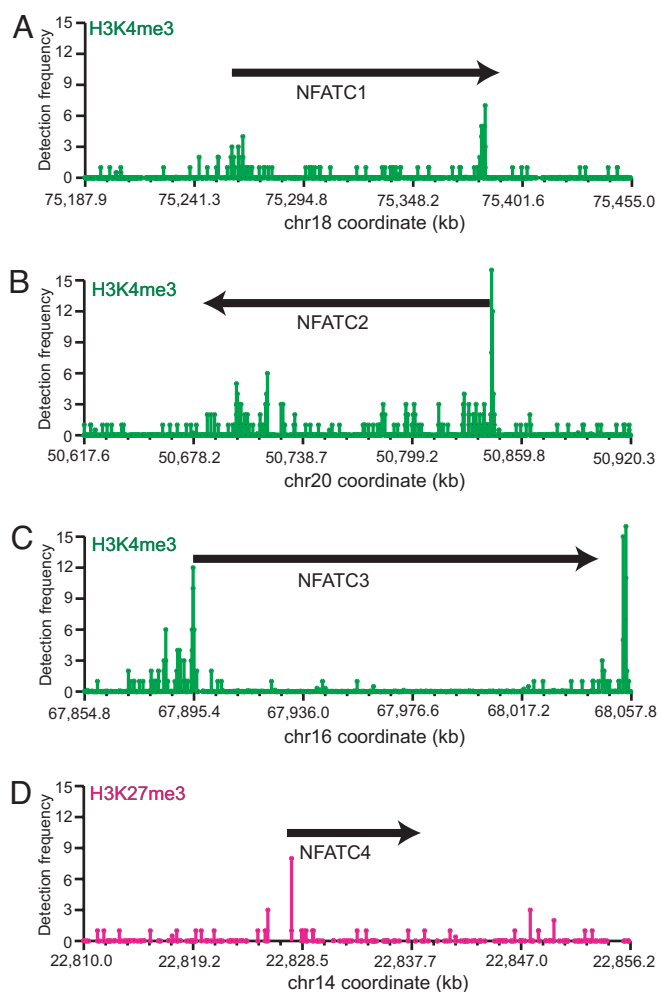


Fig. 3. The H3K4me3 modification for the TCR-inducible *NFATC1* (A), *NFATC2* (B), and *NFATC3* (C) genes and H3K27me3 modification for the silent and noninducible *NFATC4* gene (D) are shown. The positions of these genes are indicated. x axis, chromosome coordinate of the locus. y axis, level of the modification as indicated by the detection frequency of the GMAT tags.

between H3K9acK14ac and H3K27me3 indicates that the highly K27-methylated regions did not overlap with the highly acetylated chromatin regions (Fig. 4B, circle III). Circle IV Fig. 4B *Inset* highlights 8,636 genomic regions with low to moderate levels (3 to 50 detected tag counts) of both H3K9acK14ac and H3K27me3. As we showed previously (25), most tags (even single-copy tags) are true positive signals. Moreover, the data in Fig. 6A and B also show that this level of modification is well correlated with gene expression. Therefore, the signals in this region are not just noise but true signals. Our data indicate that the H3K9acK14ac and H3K27me3 modifications colocalize in a significant population of chromatin regions. The comparison between H3K4me3 and H3K27me3 revealed extensive colocalization of these two marks in the regions highlighted as V and VI (Fig. 4C). Most of the regions in V are identical with those in region II (Fig. 4A). Therefore, circle VI represents chromatin regions with low to moderate levels of H3K9acK14ac, H3K4me3, and H3K27me3 modifications, many of which are located in gene promoter regions, whereas circle V marks chromatin regions with high levels of H3K4me3 and H3K27me3 but low levels of H3 acetylation signals, which are located far from any known genes. It was recently observed that the H3K4me3 and H3K27me3 modifications colocalize at several genes critical for

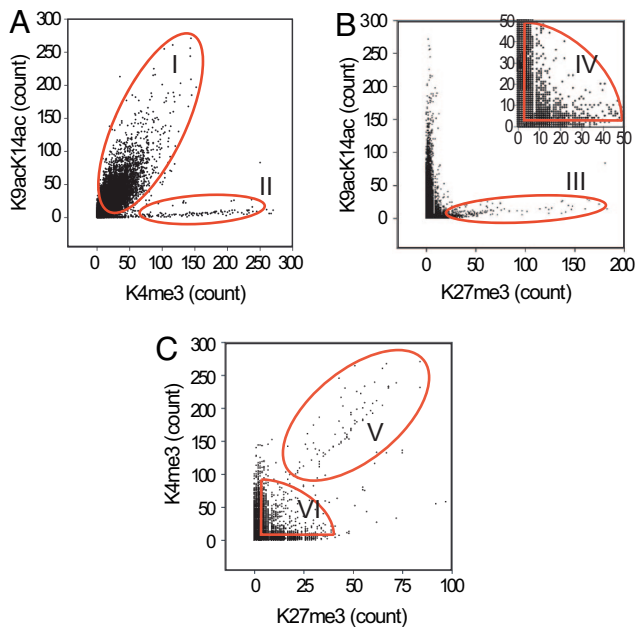


Fig. 4. Genome-wide colocalization analysis of the H3K9acK14ac, H3K4me3, and H3K27me3 modifications. Pairwise scatter analyses of the histone modifications indicated were analyzed by using GMAT. The number of tags (tag count) detected in 3-kb windows was calculated across the entire human genome. The scatter analyses of the tag counts were performed by comparing two histone modifications in A–C, as indicated. *B* Inset shows the windows with low to moderate tag numbers between 3 and 50.

development in human ES cells, suggesting that the H3K4me3 signal may prime the chromatin structure for the genes to be activated during development (31). However, our discovery of the extensive colocalization of these two marks suggests that these modifications together may define chromatin landscape in a wide range of genomic regions.

Sequential ChIP Assays Confirm the Coexistence of Activating and Repressive Marks at the Same Promoters. The detection of these apparently opposing modifications at the same genomic loci raises the possibility that their coexistence may be caused by cellular heterogeneity. The CD3⁺ T cells we used for the genome-wide analysis consisted of $\approx 33\%$ CD8⁺ T cells and 67% CD4⁺ T cells. To confirm that the colocalization of the H3K4me3 and H3K27me3 modifications was not caused by cellular heterogeneity, we analyzed several promoters, which were associated with moderate levels of H3K9acK14ac, H3K4me3, and H3K27me3 modifications in the GMAT libraries (Circles I, IV, and VI in Fig. 4), by using ChIP assays with purified CD4⁺ T cells that are major components of the CD3⁺ T cells. Indeed, our data confirmed that most tested promoters were associated with both H3K4me3 and H3K27me3 modifications (Fig. 5 *A* and *B*). To determine whether these two modifications are located on the same allele and are not independent modifications of the two alleles in CD4⁺ T cells, we further analyzed these promoters by using sequential ChIP assays. As a positive control, we performed the first ChIP assay by using an anti-H3K4me3 antibody and the second ChIP by using an anti-H3K9acK14ac antibody. This analysis confirmed the colocalization of H3K4me3 and H3K9acK14ac at all of the tested promoters (data not shown). When an anti-H3K27me3 antibody was used for the second ChIP, we detected the colocalization of H3K4me3 and H3K27me3 at all tested promoters except CSF1 (Fig. 5C). Similar results were obtained when the

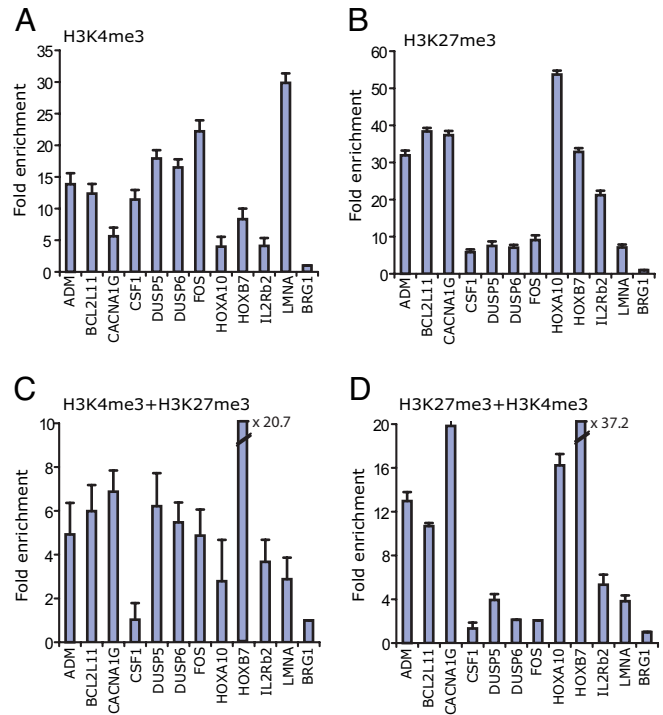


Fig. 5. Colocalization of H3K4me3 and H3K27me3 signals is confirmed by sequential ChIP assays. (*A* and *B*) ChIP assays were performed by using anti-H3K4me3 (*A*) and anti-H3K27me3 (*B*) antibodies. The fold enrichment of different promoters relative to a BRG1 upstream region was determined by using real-time quantitative PCR. (*C* and *D*) Sequential ChIP experiments. (*C*) H3K4me3 + H3K27me3, sequential ChIP using first the H3K4me3 antibody followed by the H3K27me3 antibody. (*D*) H3K27me3 + H3K4me3, sequential ChIP using first the H3K27me3 antibody followed by the H3K4me3 antibody. The fold enrichment was determined as described above. The experiments were performed in triplicate.

first ChIP was performed by using an anti-H3K27me3 antibody and the second ChIP by using an anti-H3K4me3 antibody (Fig. 5D). These data indicate that these two modifications occur on the same allele. However, our data do not address whether these modifications actually occur on the same nucleosome or on neighboring nucleosomes or whether they are attributable to the cross-linking of distant chromatin fragments in the ChIP assays. At some promoters, it is possible that the detection of these two modifications may be caused by the H3K4me3 modification occurring on one allele and the H3K27me3 modification occurring on the other allele, for example, in a situation of imprinting or monoallelic expression.

Gene Expression Levels Correlate with both the Absolute and Relative Levels of the H3K27me3 and H3K4me3 Modifications. To clarify whether the colocalization of the H3K4me3 and H3K27me3 signals correlates with gene expression, we compared the histone modification patterns with gene expression levels of 16,566 genes, which can be identified as annotated genes from a T cell gene expression database (17). The H3K4me3 signal was detected in 9,910 promoters (60% of all promoters) (Fig. 6A). H3K27me3 was detected in 5,252 promoters (32% of all promoters) (Fig. 6A). Surprisingly, 3,330 promoters were associated with both the activating (H3K4me3) and the repressive (H3K27me3) mark. Because H3K4me3 and H3K9acK14ac were highly correlated at $>95\%$ of the promoters, we did not distinguish between the two in the following analysis.

The Venn diagram in Fig. 6A shows three groups of genes based on their H3K4me3 and H3K27me3 modification patterns.

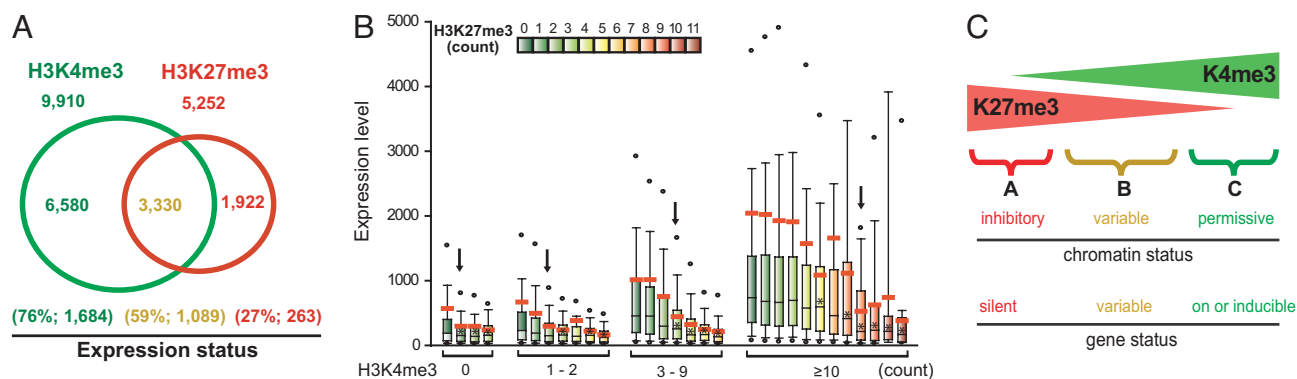


Fig. 6. Gene expression levels correlate with both the absolute and the relative levels of the H3K27me3 and H3K4me3 modifications. (A) The Venn diagram depicts the promoters associated with three tags or more of H3K4me3 (green circle) and H3K27me3 (red circle). The percentage of genes expressed and the average expression levels detected by DNA microarrays are indicated in parentheses below the circles. (B) Box-and-whisker plot showing the range of gene expression levels correlated with the histone modifications. The genes are grouped first according to H3K4me3 levels (by the total GMAT tag count in a 3-kb promoter region) and then subgrouped according to H3K27me3 levels (by the total GMAT tag count in a 3-kb promoter region). The y axis represents the arbitrary gene expression levels detected by DNA microarray analysis. The number of H3K27me3 tags corresponding to an $\approx 50\%$ decrease in expression levels is indicated by the arrow above the column. The boundary of the box closest to 0 indicates the 25th percentile; the line within the box indicates the median; the boundary of the box farthest from 0 indicates the 75th percentile; and whiskers above and below the box indicate the 5th and 95th percentiles, respectively. The outlying points are shown as filled circles. The red thick lines indicate the mean values of each group. * indicates that the decrease in gene expression levels compared with no H3K27me3 tags is significant ($P < 0.01$). (C) A Model. An inhibitory or permissive chromatin structure is generated by coexistence of the H3K27me3 and H3K4me3 modifications.

The expression levels of each group, which was measured by the average expression levels detected by DNA microarray analysis (17) and the percentage of the genes expressed in the group (250 as a cutoff value for expression), is indicated below the groups. The analysis revealed that $\approx 76\%$ of the K4-methylated genes were expressed at an average level of 1,684 (arbitrary units), whereas only 27% of the K27-methylated promoters were expressed at a level of only 263. The genes with both the H3K4me3 and H3K27me3 modifications were expressed at intermediate levels (59%; 1,089). To see whether there is a quantitative correlation between gene expression and these two modifications, we used a box-and-whisker plot to show the distribution of gene expression levels relative to the levels of the H3K4me3 and H3K27me3 modifications. The 16,566 genes first were grouped according to the sum of the H3K4me3 tag counts detected in 4-kb promoter regions and then subgrouped according to the sum of the H3K27me3 tag counts in the same regions by using colors from green to red (Fig. 6B). The results indicated that genes with higher levels of the H3K4me3 signal tend to be expressed at higher levels. The presence of H3K27me3 signals tended to correlate with decreased levels of expression in each group. The higher the number of H3K27me3 tags, the lower the expression levels. Interestingly, it seems that the H3K27me3 levels corresponding to a 50% decrease in expression levels depended on the H3K4me3 level. For example, when there were no H3K4me3 signals detected, the presence of one H3K27me3 tag correlated with a 50% decrease in expression ($P < 0.01$). When there were two tags of H3K4me3, two H3K27me3 tags correlated with a 50% decrease in expression ($P < 0.01$). When there were >10 H3K4me3 tags, the number of H3K27me3 tags that correlated with a 50% decrease in gene expression increased to 8 ($P < 0.01$). These data suggest that H3K27me3 acts as a repressive mark, and H3K4me3 correlates with gene activation. Moreover, not only the absolute level of these histone modifications but also the ratios of H3K4me3 to H3K27me3 signals appear to be critical factors associated with chromatin activity.

Discussion

In this study, we have examined and compared the distribution of several critical histone modifications in the entire genome of human T cells. Previous analyses by using ChIP-on-chip assays

with tiling arrays covering either chromosomes 21 and 22 or 9,328 1-kb promoters indicate that H3 acetylation and H3K4me3 signals are highly correlated in the promoter regions (7, 9). Our entire genome analysis confirmed the high correlation between the acetylation and H3K4me3 signals in the promoter region. Furthermore, our data indicate that a subpopulation of chromatin regions, which are located far away from any known promoters, are associated with high levels of H3K4me3 and H3K27me3 but low levels of H3K9acK14ac. The histone modifications in these regions differ from both heterochromatin and euchromatin as they are usually described in the literature. A similar case reported in the literature is that centromeric chromatin in *Drosophila* is associated with H3K4me2 and H3K9me2,3 but depleted of H3K4me3 and H3K9ac (32), indicating that different chromosomal locations may be characterized with different combinations of histone modifications. Therefore, the regions marked by H3K4me3 and H3K27me3 but not H3K9ac may represent chromosomal locations with possibly novel functions.

Correlation of histone modifications with gene expression patterns indicates that H3K4me3 generally is associated with gene expression, whereas H3K27me3 generally is correlated with gene silencing. We find in this study that thousands of promoters are associated with both the H3K4me3 and H3K27me3 modifications in differentiated human T cells, indicating that coexistence of these apparently opposing marks is a widespread phenomenon. Moreover, our data indicate that the levels of gene expression are correlated with both the absolute and relative levels of the H3K4me3 and H3K27me3 signals.

Is histone modification a step of chromatin modification before gene activation or is it simply a consequence of transcriptional activation or repression? Transcriptional activation results in recruitment of histone modification enzymes and, therefore, increased levels of modifications in yeast (33). Indeed, elevated levels of the histone acetylation and H3K4me3 modifications are detected at many promoters and other regulatory regions upon gene activation by TCR signaling (ref. 25 and data not shown). However, we did not detect significant increases in the histone modification levels of many induced genes upon TCR signaling (data not shown), suggesting that histone modifications occur at an earlier step than transcriptional activation for these

