Patterns of Histone H3 Lysine 27 Monomethylation and Erythroid Cell Type-specific Gene Expression*[S]

Laurie A. Steiner[‡1], **Vincent P. Schulz**[‡1], **Yelena Maksimova**[‡], **Clara Wong**[‡], and **Patrick G. Gallagher**[‡§2]

From the Departments of ‡*Pediatrics and* §*Genetics, Yale University School of Medicine, New Haven, Connecticut 06520*

Post-translational histone modifications, acting alone or in a context-dependent manner, influence numerous cellular processes via their regulation of gene expression. Monomethylation of histone H3 lysine 27 (K27me1) is a poorly understood histone modification. Some reports describe depletion of K27Me1 at promoters and transcription start sites (TSS), implying its depletion at TSS is necessary for active transcription, while others have associated enrichment of H3K27me1 at TSS with increased levels of mRNA expression. Tissue- and gene-specific patterns of H3K27me1 enrichment and their correlation with gene expression were determined via chromatin immunoprecipitation on chip microarray (ChIP-chip) and human mRNA expression array analyses. Results from erythroid cells were compared with those in neural and muscle cells. H3K27me1 enrichment varied depending on levels of cell-type specific gene expression, with highest enrichment over transcriptionally active genes. Over individual genes, the highest levels of H3K27me1 enrichment were found over the gene bodies of highly expressed genes. In contrast to H3K4me3, which was highly enriched at the TSS of actively transcribing genes, H3K27me1 was selectively depleted at the TSS of actively transcribed genes. There was markedly decreased to no H3K27me1 enrichment in genes with low expression. At some locations, H3K27 monomethylation was also found to be associated with chromatin signatures of gene enhancers.

The dynamic state of chromatin is influenced by many factors including nucleosomal configuration (1). Nucleosomes are composed of repetitive histone octamer units wrapped by DNA. Histones are evolutionarily conserved proteins with a characteristic histone fold and flexible NH$_2$-terminal tails. Post-translational modifications of histones (2) acting alone or in a context-dependent manner, influence various cellular processes such as transcription, replication, DNA repair, and cell cycle progression by regulating the intrinsic chromatin environment or modifying the binding of non-histone, chromatin-modifying protein complexes (3, 4). Although some histone modifications correlate with specific biologic activities, it is clear that the complexity of histone modifications, and their interactions are just beginning to be revealed (4–7). Understanding the effect these modifications have on cellular processes will provide important insights into both normal and disease-associated processes (8).

Methylation of position-specific lysine residues in NH$_2$-terminal histone tails is a critical post translational modification that can be associated with either active or repressed chromatin (9). For example, modifications such as trimethylation of histone H3 lysine 4 and acetylation of histone H3 lysine 9 are generally associated with euchromatin and gene activation (2) while trimethylation of histone H3 lysine 9 is generally associated with hetero-chromatin and gene repression.

In addition to the unmodified form, histone H3 K27 exists in three methylation states, monomethylated, dimethylated, and trimethylated. Dimethylated H3K27 is the most abundant form (10) while trimethylated H3K27 has been the best characterized. Trimethylated histone H3 lysine 27 (H3K27me3)[3] is associated with regions of heterochromatin and plays a critical role in gene silencing (11) recruiting and/or stabilizing the polycomb repressor complex 1 on chromatin, thereby mediating transcriptional repression (12–15). It plays a critical role in epigenetic programs in development, X chromosome inactivation, and imprinting (11, 16, 17).

Monomethylation of histone H3 K27 is a poorly characterized post-translational histone modification for which variable associations of gene expression and patterns of gene enrichment have been observed (18). Immunofluorescence studies localized H3K27me1 primarily to areas of pericentric heterochromatin where it was thought to be a marker of gene repression (9) yet it was also found in regions of transcriptionally permissive euchromatin. Follow up studies of H3K27me1 enrichment and gene expression have been conflicting. Selective depletion of K27Me1 at the promoters and transcription start sites (TSS) of several genes has been observed, including the active beta-globin and GATA2 genes in G1E cells, implying that depletion of H3K27me1 at the TSS is necessary for active transcription (19) while others associate increased enrichment for H3K27me1 at promoter regions with increased levels of mRNA expression (20). Studies during cellular differentiation positively correlated H3K27me1 enrichment across gene bodies with expression of less active genes but negatively correlated

[3] The abbreviations used are: H3K27me3, trimethylated histone H3 lysine 27; H3K27me1, monomethylated histone H3 lysine 27; H3K4me3, trimethylated histone H3 lysine 4; TSS, transcriptional start site; ChIP-chip, chromatin immunoprecipitation on microarray chip analysis.

HeK27me1 gene body enrichment across more active genes (21). H3K27me1 enrichment over inactive genes has been studied in less detail.

This report characterizes patterns of H3K27me1 enrichment in erythroid and non-erythroid cells using chromatin immunoprecipitation on chip microarray correlated with mRNA transcriptome analyses. These experiments revealed that the degree of H3K27me1 enrichment fluctuated depending on levels of cell-type gene expression. The location of H3K27me1 enrichment varied depending on levels of cell-type specific mRNA expression, with the highest levels of H3K27me1 enrichment found over the bodies of highly expressed genes. H3K27me1 was selectively depleted at the TSS of actively transcribed genes, which was in sharp distinction to patterns of histone H4 lysine 4 trimethylation (H3K4me3), which were enriched at TSS. There was markedly decreased to no H3K27me1 enrichment in genes with low expression. At some locations, H3K27 monomethylation was also found associated with chromatin signatures of gene enhancers.

EXPERIMENTAL PROCEDURES

Cell Culture and Selection—K562 (erythroid) and RD cells (muscle, rhabdomyosarcoma) were maintained in RPMI 1640 medium with 10% fetal calf serum. SH-SY5Y cells (neural, neuroblastoma) were maintained in a 1:1 mixture of DMEM and Ham's F12 media with 10% fetal calf serum (22). Human CD34-selected stem and progenitor cells were obtained from the Yale Center from Excellence in Molecular Hematology. Cells were cultured in StemSpan SF expansion media (StemSpan, 09650) with estradiol (100 ng/ml), dexamethasone (10 ng/ml), human transferrin (200 ng/ml), insulin (10 ng/ml), Flt3 ligand (100 ng/ml), stem cell factor (100 ng/ml), IL3 (50 ng/ml), IL6 (20 ng/ml), insulin like growth factor-1 (50 ng/ml), and erythropoietin (3units/ml) for 9–14 days. FACS analysis was used to analyze cellular expression of CD71 (transferrin receptor) and CD235a (glycophorin A) as described (22).

RNA Isolation and Amplification—RNA was prepared using the Qiagen RNeasy kit (Qiagen, Valencia, CA) according to the manufacturer's instructions. For quality control, RNA purity and integrity were verified by denaturing gel electrophoresis, OD 260/280 ratio, and analysis on an Agilent 2100 Bioanalyzer Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). Double-stranded cDNA and biotin-labeled cRNA were synthesized and purified according to the recommended Illumina protocol using a TotalPrep RNA Amplification kit (Applied Biosystems). Triplicate samples of 400 ng of total RNA per cell type were reverse transcribed to cDNA using a T7 oligo(dT) primer. Second-strand cDNA was synthesized, *in vitro* transcribed, and labeled via incorporation of biotin-16-UTP. Integrity of purified cRNA was assessed on an Agilent 2100 Bioanalyzer prior to hybridization.

Primary Microarray Data Acquisition and Analyses—Labeled cRNA samples were hybridized to Illumina HumanWG-6 v2.0 Expression Bead Chip genome-wide arrays using standardized Illumina reagents and protocols according to the manufacturer's instructions. After washing and staining, BeadChips were scanned on the Illumina Iscan. Scanned files were loaded into BeadStudio software for analysis.

Gene Expression Microarray Quality Control and Data Analyses—Quality control, probe mapping, and transformation of array data were performed using the Bioconductor lumi version 1.14 R version 2.11.0 package specifically designed for analysis of Illumina Bead Arrays. Determination of the sample mean expression, number of expressed genes, distance to sample mean, sample standard deviation, and plots of signal density, pair-wise correlation and sample clustering were generated to identify any possible outlier samples. Data were subjected to the VST variance-stabilizing transformation, then quantile normalized using procedures in the lumi package. Genes with a detection *p* value of 0.01 (the default lumi cutoff) in two or more of the three replicates were called present, and were otherwise called absent. For analysis of high and low expressed genes, median values of the three sample replicate values were used for each probe. The highest 25% and lowest 25% expression value genes that were also present on the NimbleGen ChIP-chip array were identified using a custom R script.

Transcript Validation—Quantitative real-time quantitative PCR was performed to confirm expression levels of RNA transcripts. RNA prepared from K562 cells, SY5Y, and RD cells was treated with amplification grade DNase I and reverse transcribed with an oligo(dT) primer using the SuperScript First-Strand Synthesis System (Invitrogen). Primer pairs for 15 representative genes were created using Primer 3 software, designed to amplify ~150-bp fragments, each spanning an intron. Reverse transcription products were amplified by real-time PCR using an iCycler (Bio-Rad) with the primers in supplemental Table S1. PCR specificity was verified by assessing amplification product melting curves. Real-time PCR data were normalized to an ornithine decarboxylase antizyme 1 (OAZ1) mRNA control. The fold changes in specific mRNA levels were calculated using the ΔCT method, with results presented as mean ± S.E. of the fold changes. Results were normalized to the highest expressed gene in each group. Triplicate analyses were performed for each target (22, 23).

Chromatin Immunoprecipitation (ChIP)—ChIP assays were performed as previously described. Antibodies utilized for immunoprecipitation included histone 3 monomethyl lysine 27 (H3K27me1, Upstate 07–448), histone 3 tri-methyl lysine 4 (H3K4me3, ABCAM, ab8580), and nonspecific rabbit IgG (Santa Cruz, sc-2091). Antibody-bound DNA-protein complexes were collected using protein A- or G-agarose beads, washed, eluted from the beads, and cross-linking of DNA-protein adducts reversed by incubation at 65 °C for 4 h. DNA was cleaned with the QIAquick PCR purification kit (Qiagen) according to manufacturer's instructions and amplified with the GenomePlex Whole Genome Amplification kit (Sigma) according to manufacturer's instructions. Amplified DNA was cleaned using the QIAquick PCR purification kit (Qiagen) before amplification, labeling, and hybridization to arrays or before qPCR analyses.

ChIP-chip Analyses—A custom high-density genomic tiling array containing probes for the genomic regions of 117 targeted genes (supplemental Table S2) was designed with NimbleGen systems software. Probes, typically ~50 bp in length, were tiled with 10–100 bp spacing, typically ~65 bp. 10–100 kb of flank-
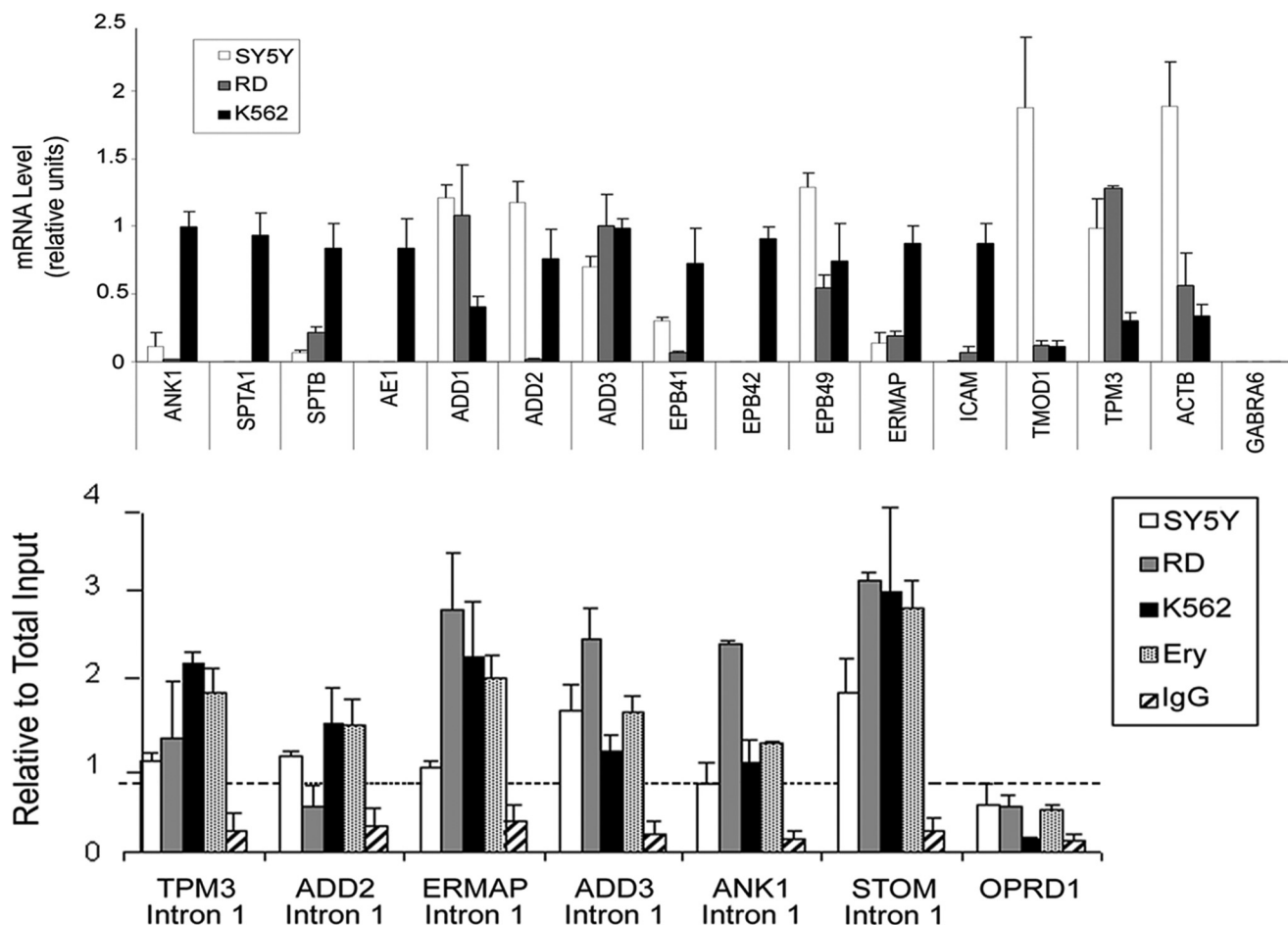
FIGURE 1. **Validation of mRNA expression arrays and ChIP-chip of major erythrocyte membrane protein genes.** *A*, reverse transcription-PCR was used to validate mRNA expression levels of major erythrocyte membrane protein genes in SHSY5Y (neural), RD (muscle), and K562 (erythroid) cells. A region from the cerebellum-specific GABRA6 gene was included as a negative control. K562 cell data are from Steiner L. *et al.* 22. ANK1 expression was not detected in RD cell mRNA, as corresponding primers were located 5′ of the muscle-specific cDNA isoform. *B*, quantitative ChIP-PCR validation of H3K27me1 enrichment at select gene loci in SY5Y, RD, K562, and human primary erythroid (Ery) cell chromatin. A region from the opioid receptor delta 1 (*OPRD1*) gene was included as a negative control. *White bars* represent H3K27me1-precipitated SY5Y chromatin. *Gray bars* represent H3K27me1-precipitated RD chromatin. *Black bars* represent H3K27me1-precipitated K562 cell chromatin. *Stippled bars* represent H3K27me1-precipitated human primary erythroid cell chromatin. *Striped bars* denote representative IgG-precipitated chromatin from all three cell types. Gene Abbreviations: *ANK1*: ankyrin-1; *SPTA1*: alpha spectrin; *SPTB*: beta spectrin; *AE1*: band 3; *ADD1*: alpha-adducin; *ADD2*: beta-adducin; *ADD3*: gamma-adducin; *EPB41*: protein 4.1R; *EPB42*: protein 4.2; *EPB49*: dematin; *ERMAP*: erythroblast membrane-associated protein; *ICAM*: intercellular adhesion molecule 4; *TMOD1*: tropomodulin-1; *TPM3*: tropomyosin-3; *GABRA6*: gamma-aminobutyric acid A receptor, alpha 6; *OPRD1*: opioid receptor, delta 1.

ing DNA was included on the chip for each locus. Because of the flanking DNA, a total of 234 genes are fully contained on the array. Each probe was duplicated on the array. Labeling and hybridization of DNA samples for ChIP-chip analysis was performed as described (22).

Data obtained from ChIP-chip experiments were analyzed using the Tamalpais Peak calling algorithm to determine areas of DNA-protein interaction (24). Control and immunoprecipitation paired data files were processed with the R Smudgekit version 2.4 software to remove chip hybridization artifacts. Ratio gff files of control and experimental data were generated and the three replicate data files were processed together using the Tamalpais web server to generate candidate binding regions. Peaks identified at all 4 levels of stringency were subjected to additional analyses because even though levels L1, L2, and L3 are more stringent and identify regions of binding with higher accuracy, L4 peaks also often yield valid binding sites (24).

Wig format signal profile files were generated using an R script as follows: The log2 transformed ChIP and input signal data were quantile normalized using the preprocessCore package (25). The median input signal for all replicates was subtracted from the median Chip signal for all replicates to generate log ratio wig files. Aggregate data for various genomic regions for the wig files were generated using the Galaxy analysis tool (26). Analysis of the genomic distribution of binding sites identified by ChIP-chip was performed using the Cis-regulatory Element Annotation System (CEAS)(27, 28) using the wig files described above.

Validation of ChIP-chip Results—Primers were designed for representative peaks where H3K27me1 binding was identified by the Tamalpais peak calling algorithm (supplemental Table S3). Immunoprecipitated DNA was analyzed by quantitative real-time PCR (iCycler, Bio-Rad) using the appropriate primers as previously described. SYBR green fluorescence in 25 μl of PCR reactions was determined and the amount of product was
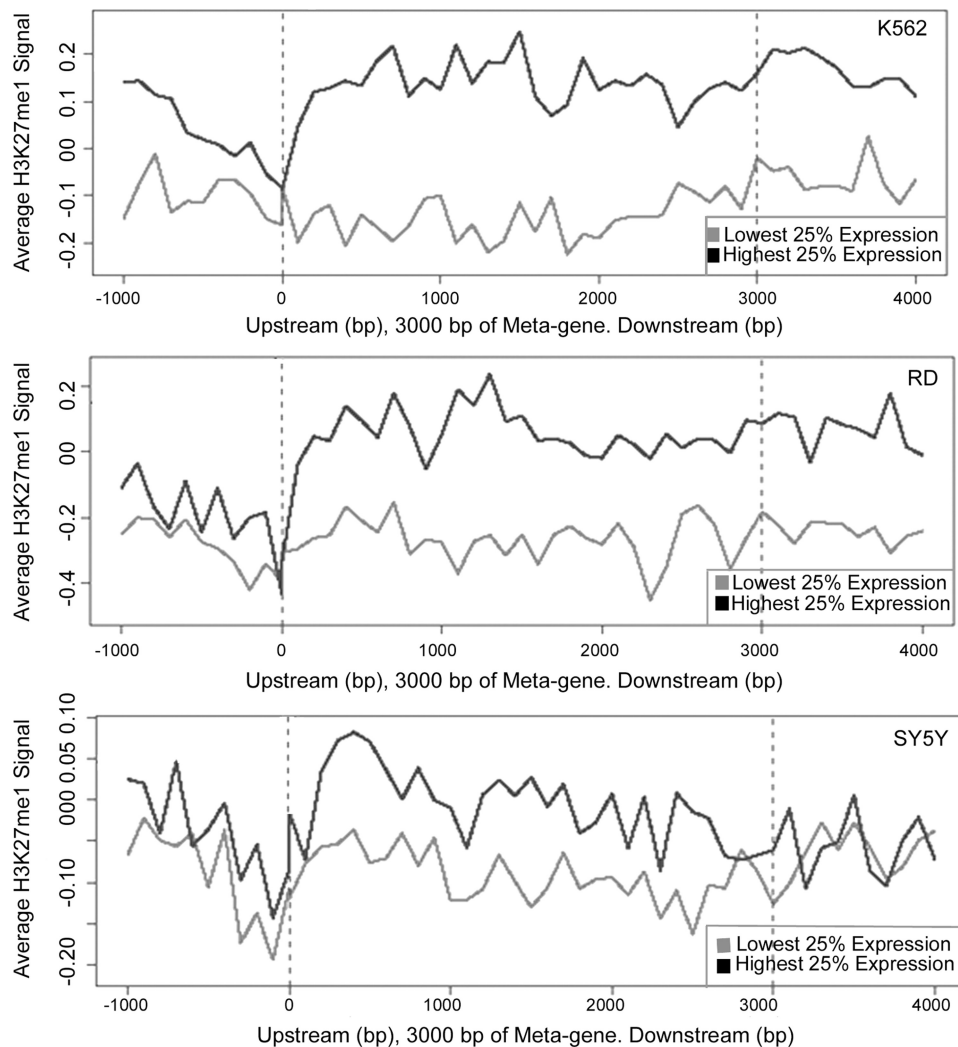
FIGURE 2. **Metagene profiling of H3K27me1 enrichment *versus* mRNA expression.** Levels of H3K27me1 enrichment were plotted on metagene profiles for the highest 25% (*black*) and lowest 25% (*gray*) expressed genes in erythroid (K562), muscle (RD), and neural (SY5Y) cells.

determined relative to a standard curve generated from a titration of input chromatin. Amplification of a single amplification product was confirmed by dissociation curve analysis. Enrichment of binding sites in target DNA over input was determined using ΔCT analysis. Results presented as mean $\pm$ S.E. of the fold enrichment with triplicate analyses performed for each binding site. Student's *t* test was used to compare fold enrichment attained with each specific antibody to fold enrichment attained using nonspecific IgG. *p* values > 0.05 were considered to be significant.

Data Access—The raw data files generated by the array and ChIP-chip analyses have been submitted to Gene Expression Omnibus (GEO) with the accession number of GSE32919. The mRNA microarray experiments comply with MIAME (Minimum Information About a Microarray Experiment) standards (29).

RESULTS

Transcriptome Analyses and ChIP-chip—Levels of mRNA expression were assessed by transcript profiling via hybridization of mRNA from three different cell types, K562 (erythroid),

SY5Y (neural), and RD (muscle) cells, to Illumina HumanWG-6 v2.0 microarrays. Levels of expression were assigned absent or present calls using the Illumina detection *p* values that are based on negative control hybridization probes. Genes were also ranked by expression level, and compared by quartiles. Quantitative real-time PCR was performed to validate expression levels of representative mRNA transcripts assigned by the expression arrays (Fig. 1*A*). Results from 54 of 57 probes from the microarrays were validated by real-time PCR in the 3 cell types (supplemental Table S4). Only one of the Illumina ADD2 probes failed. Quantitative RT-PCR detected ADD2 mRNA in K562 cell and SY5Y mRNA, paralleling previously reported expression limited to nervous system and hematopoietic cells.

To characterize H3K27me1 enrichment in erythroid and nonerythroid cells, ChIP-chip was performed. ChIP was done using chromatin from K562, SY5Y, and RD cells using an antibody specific to H3K27me1. The resulting DNA was applied to a custom designed NimbleGen array which contained the loci of 117 genes spanning 16.0 MB of the human genome and representing all the autosomal chromosomes except 13 and 18. Levels of H3K27me1 enrichment were assessed using the CEAS

A

SY5Y

RD

K562

ADD3

B

SY5Y

RD

K562

ADD2

C

SY5Y

RD

K562

ANK1   ANK1
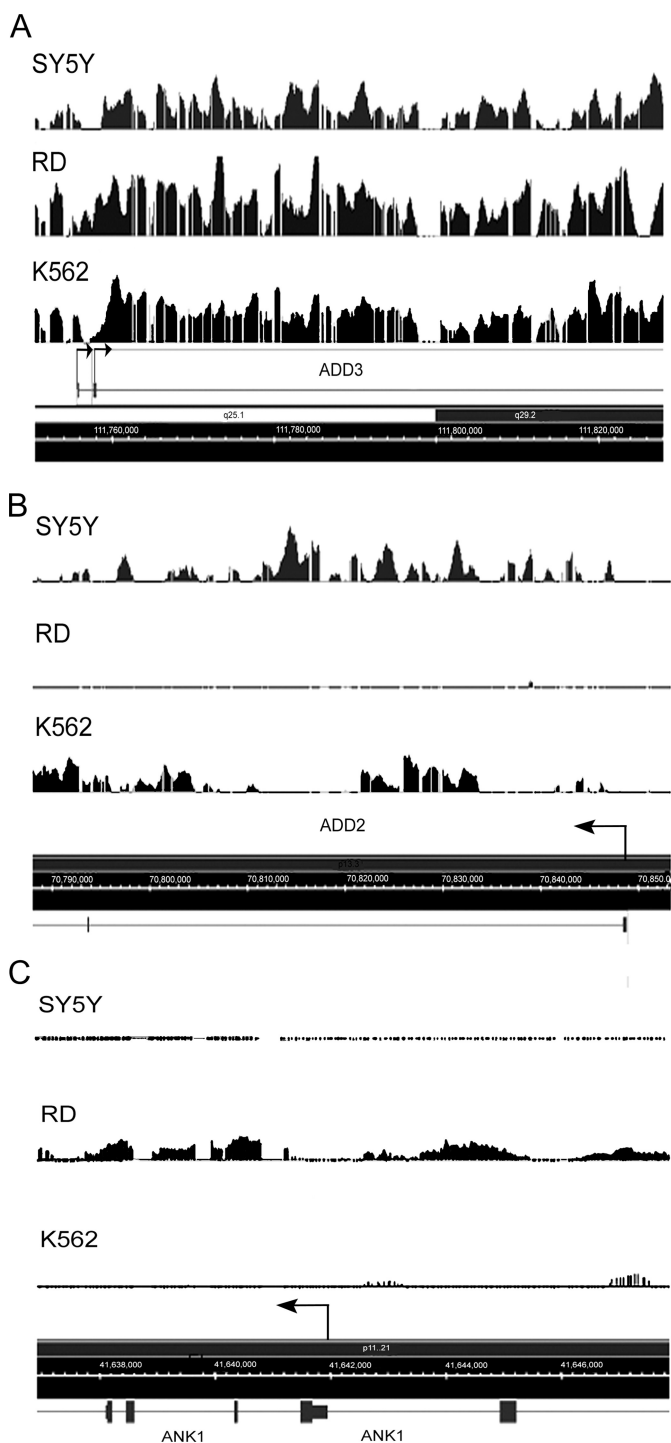
FIGURE 3. **Representative integrated genome browser (*IGB*) views of H3K27me1 enrichment.** IGB view of H3K27me1 enrichment at the *ADD3* (*A*), *ADD2* (*B*), and *ANK1* (*C*) gene loci in neural (SY5Y), muscle (*RD*), and erythroid (*K562*) cell chromatin. Arrows indicate transcriptional start sites. There is *ADD3* mRNA expression in all three cell types, paralleled by broad regions of H3K27me1 enrichment in chromatin from all three cell types. H3K27me1 enrichment correlated with cell-type specific mRNA expression in the *ADD2* gene, expressed only in K562 and SY5Y cell mRNA, but not RD cell mRNA, paralleling H3K27me1 enrichment at the *ADD2* locus in K562 and SY5Y but not RD cell chromatin. Similarly, H3K27me1 enrichment was found only in the region of a muscle-specific transcript in RD cells at the very 3′-end of the *ANK1* gene locus.

algorithm. ChIP Q-PCR validated H3K27me1 enrichment in six representative gene loci assigned by the ChIP-chip arrays (Fig. 1*B*).

As noted above, all 6 regions of H3K27me1 enrichment in K562 chromatin that were examined were validated by use of ChIP-qPCR (Fig. 1*B*). K562 cells have been utilized as models of erythroid cells in numerous studies of gene structure and function. However, variation in copy number produced by karyotypic abnormalities acquired over time may alter the results of ChIP-qPCR, necessitating study in primary human erythroid cells to confirm biologic relevance. The six regions of H3K27me1 enrichment identified for membrane protein genes in K562 cells were examined in chromatin from R3/R4 stage primary, cultured human erythroid cells by use of ChIP-qPCR. All six regions of H3K27me1 enrichment identified in K562 chromatin were also enriched in chromatin of primary human erythroid cells (Fig. 1*B*).

H3K27me1 Enrichment Correlates with Cell Type-specific Gene Expression—To correlate H3K27me1 enrichment with patterns of gene expression, the level of enrichment for H3K27me1 for each locus was plotted against the mRNA expression level over a meta-gene profile in all three cell types. In these meta-gene analyses, average signals from continuous ChIP enrichment across every element (gene) are plotted and compared against background percentages normalized to genes of the same length (*i.e.* 3 kb). H3K27me1 demonstrated a consistent pattern of enrichment in highly expressed genes in all three cells types (Fig. 2).

The correlation of H3K27me1 binding with cell-type specific expression was validated across several gene loci by ChIP-qPCR. The membrane protein gene adducin-3 (*ADD3*) had significant mRNA expression in all three cell types, paralleled by H3K27me1 enrichment in all three cell types (Fig. 3*A*). The adducin-2 (*ADD2*) gene demonstrated significant expression in K562 and SY5Y but not RD cell mRNA, a pattern that correlated with K27me1 enrichment in K562 and SY5Y, but not RD cell chromatin (Fig. 3*B*). The human ankyrin-1 gene locus (*ANK1*) is spread over 220 kb and the erythroid transcript is encoded by 42 exons (30). At the 3′-end of the ankyrin-1 locus, a short muscle-specific mRNA isoform transcribed off a muscle specific exon followed by the last 4 ankyrin-1 exons driven by a muscle-specific promoter is found (31). K27me1 enrichment is present across this region in muscle, but not neural or erythroid cell chromatin (Fig. 3*C*).

Increased Levels of H3K27me1 Expression Over the Gene Body Correlate with Increased Gene Expression—To specifically assess the correlation between the *location* of H3K27me1 enrichment in/around gene loci and gene expression, H3K27me1 enrichment in three regions of each gene: the 5′ flanking region (designated −1000 bp to −250 bp), the core promoter/transcription start site (TSS, designated −250 to +250), and over the body of the gene (+250 to end of gene), was correlated with mRNA expression. Over the gene body, increasing levels of H3K27me1 correlated with increased levels of gene expression in K562 cells, with the level of H3K27me1 enrichment over the body of the top quartile of expressed genes significantly higher than the level of H3K27me1 enrichment over the bottom quartile of expressed genes (p value = 1.756e-10)(Fig. 4*A*). Similar results were observed in SY5Y and RD cells (not shown). Similarly, H3K27me1 enrichment over genes bodies was positively correlated to gene expression when mRNA
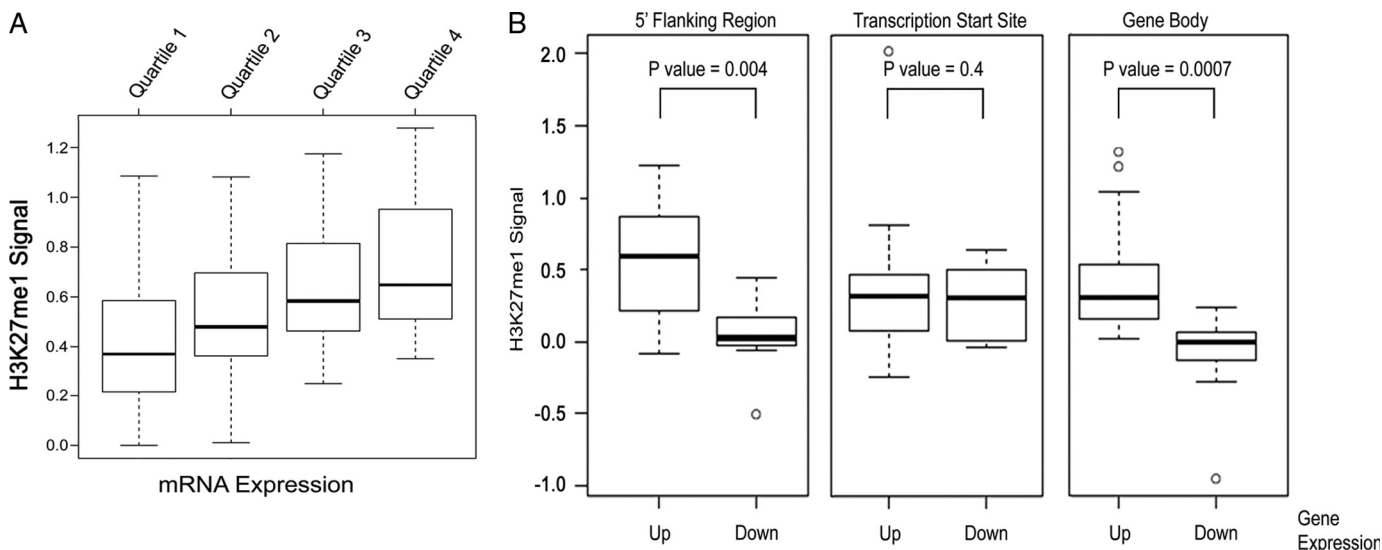
FIGURE 4. **Box plots depicting the relationship between levels of H3K27me1 enrichment over gene bodies and levels of mRNA expression.** *A*, levels of H3K27me1 enrichment over gene bodies in K562 cell chromatin were plotted against mRNA expression. Gene expression, divided into quartiles, was ranked from low (quartile 1) to high (quartile 4) from *left* to *right*. Increasing levels of H3K27me1 enrichment correlated with levels of gene expression divided by quartiles, with the level of H3K27me1 enrichment over the body of the top quartile of expressed genes being significantly higher than the level of H3K27me1 enrichment over the bottom quartile of expressed genes. *B*, levels of H3K27me1 enrichment in genes differentially expressed between K562 and RD cells. Genes with a 2-fold difference in expression between cell types (at an adjusted *p* value <0.05) were identified, and levels of H3K27me1 enrichment compared.

expression was assigned as present or absent (*p* value = 2.816e-06, not shown).

Levels of H3K27me1 enrichment were also assessed in genes differentially expressed between the three different cell types. Genes with a 2-fold difference (at adjusted *p* value<0.05) in expression between cell types were identified and levels of H3K27me1 enrichment compared. When comparing expression of genes up-regulated *versus* down regulated in K562 *versus* RD cells (Fig. 4*B*), H3K27me1 enrichment was highest in the 5′ flanking region (*p* value = 0.004) and over the body of the gene (*p* value <0.0007). This increase in H3K27me1 enrichment was not seen at the TSS (*p* value = 0.4, not significant). Similar correlation of gene expression with levels and sites of H3K27me1 enrichment were obtained when genes differentially expressed in K562 cells were compared with genes in SY5Y cells (not shown).

H3K27me1 Is Selectively Depleted at the Transcription Start Site of Actively Transcribed Genes—There have been variable, limited reports of the role of H3K27me1 at the TSS, with some implying that higher levels of H3K27me1 at the TSS are associated with actively transcribed genes, and other reports speculating that depletion of H3K27me1 at the promoter was necessary for active transcription to occur (19, 20). In our dataset, actively transcribed genes demonstrated selective depletion of H3K27me1 expression at the TSS in all three cell types (Fig. 5*A*). This is in sharp contrast to H3K4Me3, which accumulates at the transcription start site of active genes (Fig. 5*B*). An example of TSS depletion is shown at the 5′-end of the beta-spectrin gene in K562 cell chromatin (Fig. 6).

Genes with Little to No H3K27me1 Enrichment—Although H3K27me1 was originally found to be localized primarily to regions of heterochromatin, it has not been studied in detail in transcriptionally silent genes. In K562 cells, 94 genes were

below the threshold for expression and 140 were above the threshold of expression (absent/present calls). The majority of nonexpressed genes had no H3K27me1 enrichment and only 12.7% of nonexpressed genes had any H3K27me1 enrichment at all, and then at very low levels (levels between 0.2 and 0.5 log ChIP/input).

Nuclear staining suggests that H3K27me1 is distributed widely throughout the genome (9) and it has been suggested that monomethylation of H3K27 is the default state of chromatin, with TSS depletion necessary for active transcription (19). Interestingly, our study identified 12 genes with housekeeping-like function, *e.g.* beta-actin, which lacked any H3K27me1 enrichment across their entire gene loci in all 3 cell types, yet demonstrated significant mRNA expression in all three cell types. These findings imply that H3K27me1 enrichment is not always necessary for transcription.

Enhancer Signatures and H3K27me1 Enrichment—Analysis of the chromatin signature of transcriptional enhancers has been a topic of significant effort in recent reports. One report described HeK27me1, H3K4me1, and H3K9me1 enrichment at enhancers of differentiation genes prior to their activation, suggesting that these monomethylations participate in maintaining the activation state necessary for cellular differentiation (21). To determine if H3K27me1 enrichment is part of a chromatin signature for tissue-specific enhancers, we compared its enrichment to a previously reported enhancer signature in K562 cell chromatin (32). In the region contained on our chip, there were 399 K562 cell enhancer signature peaks and 275 H3K27me1 Tamalpais L1-L4 peaks with 82 overlapping peaks. Compared with a randomized version of the 399 K562 cell enhancer signature peaks, randomized within the same erythroid chip regions and with the same lengths of regions, there were only 17 overlaps (*p* value< 2.2e-16, 95% confidence intervals: 0.1969188, 0.2891735),

A   H3K27me1
      Transcriptional Start Site

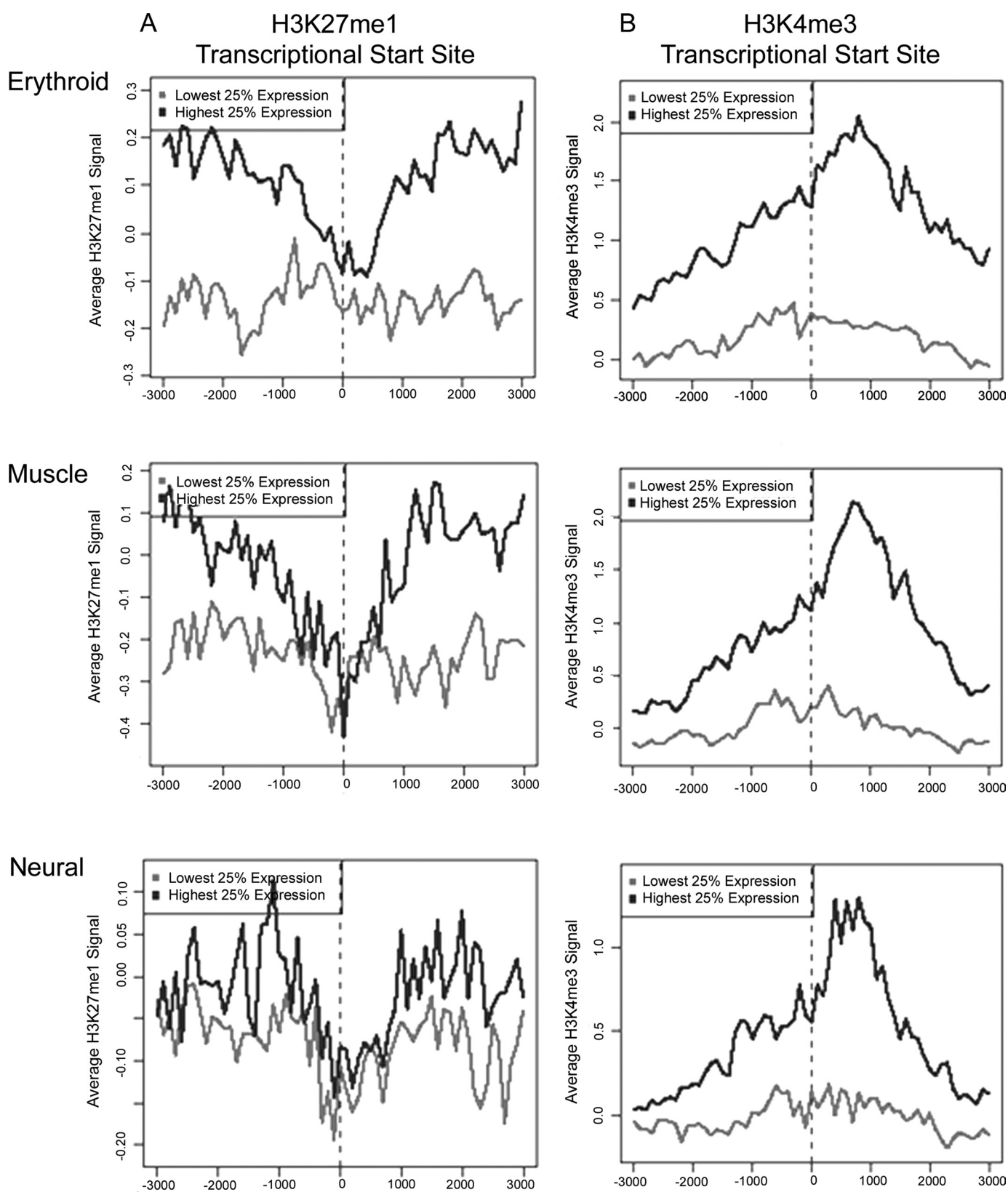B   H3K4me3
      Transcriptional Start Site

FIGURE 5. **Profiling of H3K27me1 and H3K4me3 enrichment *versus* mRNA expression at promoter region/transcriptional start sites.** *A*, levels of H3K27me1 enrichment at promoter region/transcriptional start sites were plotted for the highest 25% (*black*) and lowest 25% (*gray*) expressed genes in erythroid (*K562*), muscle (*RD*), and neural (*SY5Y*) cells. *B*, levels of H3K4me3 enrichment at promoter region/transcriptional start sites were plotted for the highest 25% (*black*) and lowest 25% (*gray*) expressed genes in erythroid (*K562*), muscle (*RD*), and neural (*SY5Y*) cells.
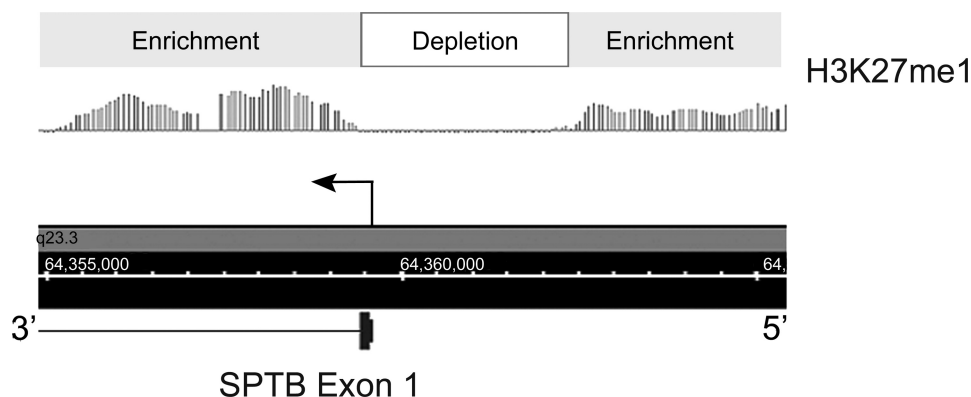
FIGURE 6. **Representative integrated genome browser (*IGB*) view of H3K27me1 enrichment at a promoter region/transcriptional start site in K562 cells.** Representative IGB view of H3K27me1 enrichment at the beta-spectrin gene (*SPTB*) promoter region/transcriptional start site. The erythroid exon 1 is noted below. The *arrow* indicates the transcriptional start site of the erythroid mRNA. Regions of H3K27me1 enrichment (*shaded*) and depletion (*white*) are noted above the gene.

indicating that a statistically significant number of K562 cell enhancer signatures contain H3K27me1. Average H3K27me1 signals from the enhancer and a shuffled enhancer control were extracted and compared in box plot format, confirming the differences in signal at enhancers (*p* value from multiple randomizations <0.001).

DISCUSSION

Patterns of H3K27 monomethylation vary across the genomes of various organisms. In *Caenorhabditis elegans*, H3K27me1 is significantly enriched at highly expressed X-linked genes in the early embryo, where it may play a role in germ line repression of the X chromosome (33). In *Arabidopsis*, H3K27me1 is enriched at chromocenters (34), where it has been proposed to play a role in maintaining heterochromatin condensation and gene silencing (18). In mammalian cells, H3K27me1 has been localized by immunofluorescence studies both to areas of pericentric heterochromain, where it was thought to be a marker of gene repression (9) and to regions of euchromatin. Our data, obtained in three different mammalian cell types, indicate that H3K27 monomethylation exhibits a wide distribution throughout euchromatin, with the location and degree of H3K27me1 enrichment correlating directly with levels of gene expression, with higher levels of enrichment at the loci of highly expressed genes.

Our data also directly address the reported inconsistencies in the location of H3K27me1 enrichment across gene loci and its relationship to levels of gene expression. Vakoc *et al.* (19) described H3K27 monomethylation at euchromatic regions of the beta-major globin, the polyadenylate-binding protein-1, and the GATA-2 gene loci, with enrichment throughout the bodies of these genes during active transcription. They also observed that H3K27me1 was selectively removed in the vicinity of the transcription start sites of active genes. Profiling the methylation status of histones in human primary CD4+ T lymphocytes, Barski *et al.* (20) also observed H3K27me1 enrichment in actively transcribing genes. In contrast, their studies suggested that H3K27me1 enrichment was higher at the core promoters and TSS of active genes compared with transcriptionally inactive genes, particularly immediately downstream of the transcription start site.

The data in this report reveal that in mammalian cells, the location and degree of H3K27me1 enrichment varies with levels of cell-type specific gene expression. The highest levels of H3K27me1 enrichment were found over the gene bodies of highly expressed genes, followed by enrichment in the 5′ flanking region. There was markedly decreased to no H3K27me1 enrichment in genes with low expression.

In contrast, in highly expressed, actively transcribing genes, H3K27me1 was selectively depleted at the TSS. This is sharp contrast to H3K4me3, which was highly enriched at the TSS of actively transcribing genes. Thus our data agree with that of Vakoc *et al.* (19) who also observed selective removal of H3K27me1 in the vicinity of the TSS of actively transcribing genes, and support the suggestion that levels of H3K27 monomethylation at TSS undergo dynamic changes on activation or repression of transcription. Together, our data indicate that H3K27me1 is a regionally dynamic histone modification with levels inversely correlating with levels of gene expression.

Not surprisingly, the mechanisms that control deposition of the H3K27 monomethylation mark also vary across species. In *Arabidopsis*, this process is regulated by ATXR5 and ATXR6, yet these proteins and their homologues are lacking in metazoans (18, 34). In *Drosophila*, HeK27 monomethylation is dependent on E(z) (35–37). In mammals, homologues of E(z), EZH1, and EZH2, regulate H3K27 monomethylation (38). The polycomb repressive complex 2 (PRC2) is also required for mammalian H3K27 monomethylation (39). Our data demonstrate that monomethylation of H3K27me1 and trimethylation of H3K4 are mirror images of each other. These observations are of note because in *Drosophila*, the enzymes that regulate H3K27 monomethylation and H3K4 trimethylation, Polycomb, and Trithorax, respectively, are well known antagonists (40).

These studies drive the important question of what is/are the functional role(s) of H3K27me1 in mammalian cells. Our data indicate that removal of H3K27me1 at TSS is required for active transcription. Beyond this, additional studies of chromatin architecture and gene expression are needed to fully understand the role of H3K27me1. Data continue to accumulate indicating that individual methyl marks on histones, considered alone, may have limited biological significance. For instance,

H3K9me3, like H3K27me1, originally considered a mark of constitutive heterochromatin, was later found to sometimes mark transcriptionally active genes (11). Thus, although any single histone mark, such as H3K27me1, might not *always* correlate with transcriptional state, in combination with other histone modifications and regulatory protein binding, it may contribute to the overall transcriptional regulatory program in a cell (4, 11). Recent work demonstrates how specific properties of the polyvalent chromatin fiber and its associated effector proteins are able to amplify small differences in histone methyl lysine recognition, influencing the dynamic state of associated chromatin architecture (12). Ultimately, understanding the complexity of histone modifications in cells of different types and their functional contribution to regulation of gene expression will contribute to our knowledge or normal and perturbed hematopoiesis.

REFERENCES

1. Bai, L., and Morozov, A. V. (2010) *Trends Genet* **26,** 476–483
2. Zhou, V. W., Goren, A., and Bernstein, B. E. (2010) *Nat. Rev. Genet* **12,** 7–18
3. Campos, E. I., and Reinberg, D. (2009) *Annu. Rev. Genet* **43,** 559–599
4. Suganuma, T., and Workman, J. L. (2008) *Cell* **135,** 604–607
5. Berger, S. L. (2007) *Nature* **447,** 407–412
6. Berger, S. L., Kouzarides, T., Shiekhattar, R., and Shilatifard, A. (2009) *Genes Dev.* **23,** 781–783
7. Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Peng, W., Zhang, M. Q., and Zhao, K. (2008) *Nat. Genet.* **40,** 897–903
8. Bhaumik, S. R., Smith, E., and Shilatifard, A. (2007) *Nat. Struct. Mol. Biol.* **14,** 1008–1016
9. 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., Martens, J. H., and Jenuwein, T. (2003) *Mol. Cell* **12,** 1577–1589
10. Papp, B., and Müller, J. (2006) *Genes Dev.* **20,** 2041–2054
11. Trojer, P., and Reinberg, D. (2006) *Cell* **125,** 213–217
12. Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007) *Cell* **128,** 735–745
13. Schwartz, Y. B., and Pirrotta, V. (2007) *Nat. Rev. Genet* **8,** 9–22
14. Ringrose, L., Ehret, H., and Paro, R. (2004) *Mol. Cell* **16,** 641–653
15. 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., Koseki, H., Fuchikami, T., Abe, K., Murray, H. L., Zucker, J. P., Yuan, B., Bell, G. W., Herbolsheimer, E., Hannett, N. M., Sun, K., Odom, D. T., Otte, A. P., Volkert, T. L., Bartel, D. P., Melton, D. A., Gifford, D. K., Jaenisch, R., and Young, R. A. (2006) *Cell* **125,** 301–313
16. Plath, K., Fang, J., Mlynarczyk-Evans, S. K., Cao, R., Worringer, K. A., Wang, H., de la Cruz, C. C., Otte, A. P., Panning, B., and Zhang, Y. (2003) *Science* **300,** 131–135
17. Umlauf, D., Goto, Y., Cao, R., Cerqueira, F., Wagschal, A., Zhang, Y., and Feil, R. (2004) *Nat. Genet.* **36,** 1296–1300
18. Jacob, Y., and Michaels, S. D. (2009) *Epigenetics* **4,** 366–369
19. Vakoc, C. R., Sachdeva, M. M., Wang, H., and Blobel, G. A. (2006) *Mol. Cell Biol.* **26,** 9185–9195
20. Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007) *Cell* **129,** 823–837
21. Cui, K., Zang, C., Roh, T. Y., Schones, D. E., Childs, R. W., Peng, W., and Zhao, K. (2009) *Cell Stem Cell* **4,** 80–93
22. Steiner, L. A., Maksimova, Y., Schulz, V., Wong, C., Raha, D., Mahajan, M. C., Weissman, S. M., and Gallagher, P. G. (2009) *Mol. Cell Biol.* **29,** 5399–5412
23. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002) *Genome Biol.* **3,** RESEARCH0034
24. Bieda, M., Xu, X., Singer, M. A., Green, R., and Farnham, P. J. (2006) *Genome Res.* **16,** 595–605
25. Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003) *Bioinformatics* **19,** 185–193
26. Goecks, J., Nekrutenko, A., and Taylor, J. *Genome Biol.* **11,** R86
27. Ji, X., Li, W., Song, J., Wei, L., and Liu, X. S. (2006) *Nucleic Acids Res.* **34,** (Web Server issue), W551–W554
28. Shin, H., Liu, T., Manrai, A. K., and Liu, X. S. (2009) *Bioinformatics* **25,** 2605–2606
29. Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M. (2001) *Nat. Genet.* **29,** 365–371
30. Gallagher, P. G., Tse, W. T., Scarpa, A. L., Lux, S. E., and Forget, B. G. (1997) *J. Biol. Chem.* **272,** 19220–19228
31. Gallagher, P. G., and Forget, B. G. (1998) *J. Biol. Chem.* **273,** 1339–1348
32. Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., Lobanenkov, V. V., Stewart, R., Thomson, J. A., Crawford, G. E., Kellis, M., and Ren, B. (2009) *Nature* **459,** 108–112
33. Liu, T., Rechtsteiner, A., Egelhofer, T. A., Vielle, A., Latorre, I., Cheung, M. S., Ercan, S., Ikegami, K., Jensen, M., Kolasinska-Zwierz, P., Rosenbaum, H., Shin, H., Taing, S., Takasaki, T., Iniguez, A. L., Desai, A., Dernburg, A. F., Kimura, H., Lieb, J. D., Ahringer, J., Strome, S., and Liu, X. S. (2011) *Genome Res.* **21,** 227–236
34. Jacob, Y., Feng, S., LeBlanc, C. A., Bernatavichute, Y. V., Stroud, H., Cokus, S., Johnson, L. M., Pellegrini, M., Jacobsen, S. E., and Michaels, S. D. (2009) *Nat. Struct. Mol. Biol.* **16,** 763–768
35. Ebert, A., Schotta, G., Lein, S., Kubicek, S., Krauss, V., Jenuwein, T., and Reuter, G. (2004) *Genes Dev.* **18,** 2973–2983
36. Müller, J., Hart, C. M., Francis, N. J., Vargas, M. L., Sengupta, A., Wild, B., Miller, E. L., O'Connor, M. B., Kingston, R. E., and Simon, J. A. (2002) *Cell* **111,** 197–208
37. Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002) *Cell* **111,** 185–196
38. Shen, X., Liu, Y., Hsu, Y. J., Fujiwara, Y., Kim, J., Mao, X., Yuan, G. C., and Orkin, S. H. (2008) *Mol. Cell* **32,** 491–502
39. Montgomery, N. D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S. J., Otte, A. P., and Magnuson, T. (2005) *Curr. Biol.* **15,** 942–947
40. Ruthenburg, A. J., Allis, C. D., and Wysocka, J. (2007) *Mol. Cell* **25,** 15–30