View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Elsevier - Publisher Connector

Cell Stem Cell
Article

# Whole-Genome Analysis of Histone H3 Lysine 4 and Lysine 27 Methylation in Human Embryonic Stem Cells

Guangjin Pan,[1] Shulan Tian,[2] Jeff Nie,[2] Chuhu Yang,[2] Victor Ruotti,[2] Hairong Wei,[2] Gudrun A. Jonsdottir,[2] Ron Stewart,[2] and James A. Thomson[1,2,3,4,*]
[1]Genome Center of Wisconsin, University of Wisconsin-Madison, 425 Henry Mall, Madison, WI 53706-1580, USA
[2]WiCell Research Institute, P.O. Box 7365, Madison, WI 53707-7365, USA
[3]Wisconsin National Primate Research Center, University of Wisconsin-Madison, 1220 Capitol Court, Madison, WI 53715-1299, USA
[4]Department of Anatomy, University of Wisconsin School of Medicine and Public Health, 470 North Charter Street, Madison, WI 53706-1509, USA
*Correspondence: thomson@primate.wisc.edu
DOI 10.1016/j.stem.2007.08.003

## SUMMARY

**We mapped Polycomb-associated H3K27 trimethylation (H3K27me3) and Trithorax-associated H3K4 trimethylation (H3K4me3) across the whole genome in human embryonic stem (ES) cells. The vast majority of H3K27me3 colocalized on genes modified with H3K4me3. These comodified genes displayed low expression levels and were enriched in developmental function. Another significant set of genes lacked both modifications and was also expressed at low levels in ES cells but was enriched for gene function in physiological responses rather than development. Comodified genes could change expression levels rapidly during differentiation, but so could a substantial number of genes in other modification categories. *SOX2*, *POU5F1*, and *NANOG*, pluripotency-associated genes, shifted from modification by H3K4me3 alone to colocalization of both modifications as they were repressed during differentiation. Our results demonstrate that H3K27me3 modifications change during early differentiation, both relieving existing repressive domains and imparting new ones, and that colocalization with H3K4me3 is not restricted to pluripotent cells.**

## INTRODUCTION

During development from a single cell to a complex organism, mammals establish a progressively increasing number of distinct cell types with progressively decreasing developmental potential. In the adult, the identities of distinct cell types are generally stable, implying that the underlying transcriptional networks that help impart those identities must be stable, in spite of the ability of some genes to respond rapidly to physiological and environmental stimuli. However, the cloning of Dolly demonstrated that, under some conditions, those transcriptional networks could be altered to impart a dramatic change in identity. It is a central goal of regenerative medicine to understand how cellular identity is maintained and how that identity might be changed to mediate the repair of damaged tissue.

Human embryonic stem (ES) cells offer an important model for examining transitions of cellular identity in human material (Thomson et al., 1998). Unlike somatic cells, ES cells have the potential to differentiate into any cell type of the body, a potential that remains poorly understood. It has been suggested that this potential is related to an unusual chromatin structure or to specific histone modifications (Azuara et al., 2006; Lee et al., 2004; Martens et al., 2005; Meshorer and Misteli, 2006). Posttranslational modifications of core histones, such as methylation and acetylation, are correlated with the conformational state of chromatin and the transcriptional status of genes (Cosgrove and Wolberger, 2005; Fischle et al., 2003; Jenuwein and Allis, 2001; Margueron et al., 2005; Turner, 2000). Compared to differentiated cells, ES cells have highly dispersed, euchromatic nuclei, elevated amounts of histone modifications associated with open chromatin, such as acetylated histone H3 and H4, and reduced amounts of histone modifications associated with heterochromatin, such as histone H3 lysine9 trimethylation (H3K9me3) (Meshorer et al., 2006). Polycomb repressive complexes (PRCs), which catalyze H3K27me3 and were initially discovered as regulators of the Homeotic (*HOX*) genes in *Drosophilia* (Jurgens, 1985; Lewis, 1978), repress a wide range of targets in mammals (Kirmizis et al., 2004; Schwartz and Pirrotta, 2007; Squazzo et al., 2006). Genome-wide mapping of components of PRC2 reveals a dramatic enrichment for key developmental regulators in both ES cells and in diploid human fibroblasts, and these target genes are expressed at significantly lower average levels than nontarget genes (Boyer et al., 2006; Bracken et al., 2006; Lee et al., 2006). In ES cells, repression of these developmental regulators is proposed to be key in maintaining the undifferentiated, pluripotent state,

as their expression might otherwise cause differentiation. In ES cells, H3K27me3 is shown to be coextensive with binding of PRC2 components around the transcription start site (TSS) (Lee et al., 2006). In contrast to the repressive H3K27me3 modification, H3K4me3, catalyzed by Trithorax protein complexes, is associated with active transcription (Byrd and Shearn, 2003; Nagy et al., 2002). Genome-wide mapping of the MLL1 methyltransferase, a mammalian homolog to Trithorax, demonstrates a strong association with the promoters of active genes and confirms colocalization with H3K4me3 (Guenther et al., 2005).

The recent mapping of both H3K4me3- and H3K27me3-modified regions across highly conserved noncoding elements (HCNEs) in mouse ES cells revealed an unexpectedly high frequency of colocalization of these "activating" and "repressive" modifications (Bernstein et al., 2006). These "bivalent" modifications were proposed to represent genes specially poised to initiate transcription, and this poised state was proposed to be central to the developmental potential of ES cells. To gain further insight into the whole-genome distribution of H3K4me3 and H3K27me3, we mapped these modifications across the entire genome in human ES cells. Our results confirm that Polycomb-mediated H3K27me3 is associated with widespread suppression of developmental genes, and we extend previous findings in showing that colocalization of H3K27me3 with H3K4me3 at promoters in human ES cells is the rule rather than the exception. We also found that genes initially marked with both H3K4 and H3K27 modifications could indeed change rapidly with differentiation; however, many genes that lack this particular combination of marks were capable of rapid change as well. In addition, transcription factors involved in pluripotency—*OCT4* (*POU5F1*), *NANOG*, *SOX2*—that were initially marked only by H3K4me3 became marked with H3K27me3 as well with the onset of transcriptional suppression of these genes during differentiation. Our finding that pluripotency-associated transcription factors become modified by both H3K4me3 and H3K27me3 as they are downregulated during differentiation demonstrates that not all potential Polycomb targets relevant to developmental regulation are suppressed in ES cells and that the recruitment of the Polycomb complexes changes rapidly during early differentiation, both to relieve existing and create new repressive domains. We also find a substantial class of genes initially marked by neither H3K4me3 nor H3K27me3, which is poorly expressed but highly enriched for genes that can respond to physiological stimuli, suggesting fundamental differences in the transcriptional regulatory mechanisms that control genes that can respond to transient signals and those that specify or maintain cell identity.

## RESULTS

### Whole-Genome Mapping of H3K4me3 and H3K27me3 in Human ES Cells

By using specific antibodies, we immunoprecipitated H3K4me3- or H3K27me3-associated genomic DNA (chromatin immunoprecipitation, ChIP) from H1 human ES cells maintained in defined medium (TeSR medium), (Figure 1A and see Figure S1A in the Supplemental Data available with this article online) (Ludwig et al., 2006a; Ludwig et al., 2006b). Antibody-enriched DNA was then sonicated (resulting in fragments between 250 and 500 bp in size), amplified, labeled, and mixed with equal amounts of differentially labeled input DNA and hybridized to microarrays containing roughly 14.5 million 50-mer oligonucleotides, representing all the nonrepeat DNA across the human genome at a 100 bp resolution (Figures 1A and 1B) (Kim et al., 2005). We defined the H3K4me3 or H3K27me3 peaks using a simple algorithm that requires a stretch of at least three of four neighboring probes to have signals significantly above the background. A p value for each peak was calculated using the Wilcoxon matched-pairs signed rank test (see the Experimental Procedures and the Supplemental Experimental Procedures for more information). A threshold of $p \leq 0.078$ was chosen for enriched H3K4me3 or H3K27me3 modification regions, which represents a false discovery rate (FDR) of ~8% based on the stepwise p value method proposed by Benjamini and Hochberg (1995). A total of 14,106 H3K27me3 peaks were identified, with a size range of 250–42,630 bp. Of these peaks, 75% are larger than 628 bp with a median of 1050 bp. The combined length of all H3K27me3 peaks (22,700,677 bp) represents 0.75% of the genome (Figure 2A). For H3K4me3, 23,115 peaks were detected with a size range of 250–29,060 bp. Of these peaks, 75% are larger than 850 bp with a median of 1604 bp. The total length of the H3K4me3 peaks (43,223,975 bp) represents 1.4% of the genome (Figure 2A).

In order to match these peaks to genes, we first examined the H3K4me3 and H3K27me3 peak distribution near the TSS (±5 kb) of all 23,302 well-annotated RefSeq genes. We found that H3K4me3 peaks were enriched in the region closest to the TSS sites (±2 kb) (Figure 2B), consistent with previous results (Liu et al., 2005). H3K27me3 peaks were also enriched in a band centered around the TSS, but with a greater width and with central depressed signal intensity immediately over the TSS site, suggesting at least partial exclusion of this modification at the TSS (Figure 2B). This partial exclusion could be related to H3K4me3 occupancy or the presence of a preinitiation complex.

Using a region extending 2.5 kb in each direction from the 5′ end of genes, 57% of H3K27 and 73% of H3K4 trimethylation peaks corresponded to known or predicted 5′ ends. When the peaks were randomized across the genome, only 5.8% of H3K27 and 5.7% of H3K4 trimethylation peaks corresponded to known or predicted gene 5′ ends (±2.5 kb). H3K4 and/or H3K27 trimethylation marked 67% of all the UCSC known genes (Figure 2C). Notably, a majority (65%) of the UCSC known genes were marked with H3K4me3 (alone or in combination with H3K27me3), and only a small number of the UCSC genes (2%) were marked with H3K27me3 alone. We found a substantial portion of the UCSC genes (16%) that is associated with both modifications in human ES cells and a third (33%) that is not associated with either modification (Tables S1–S4).

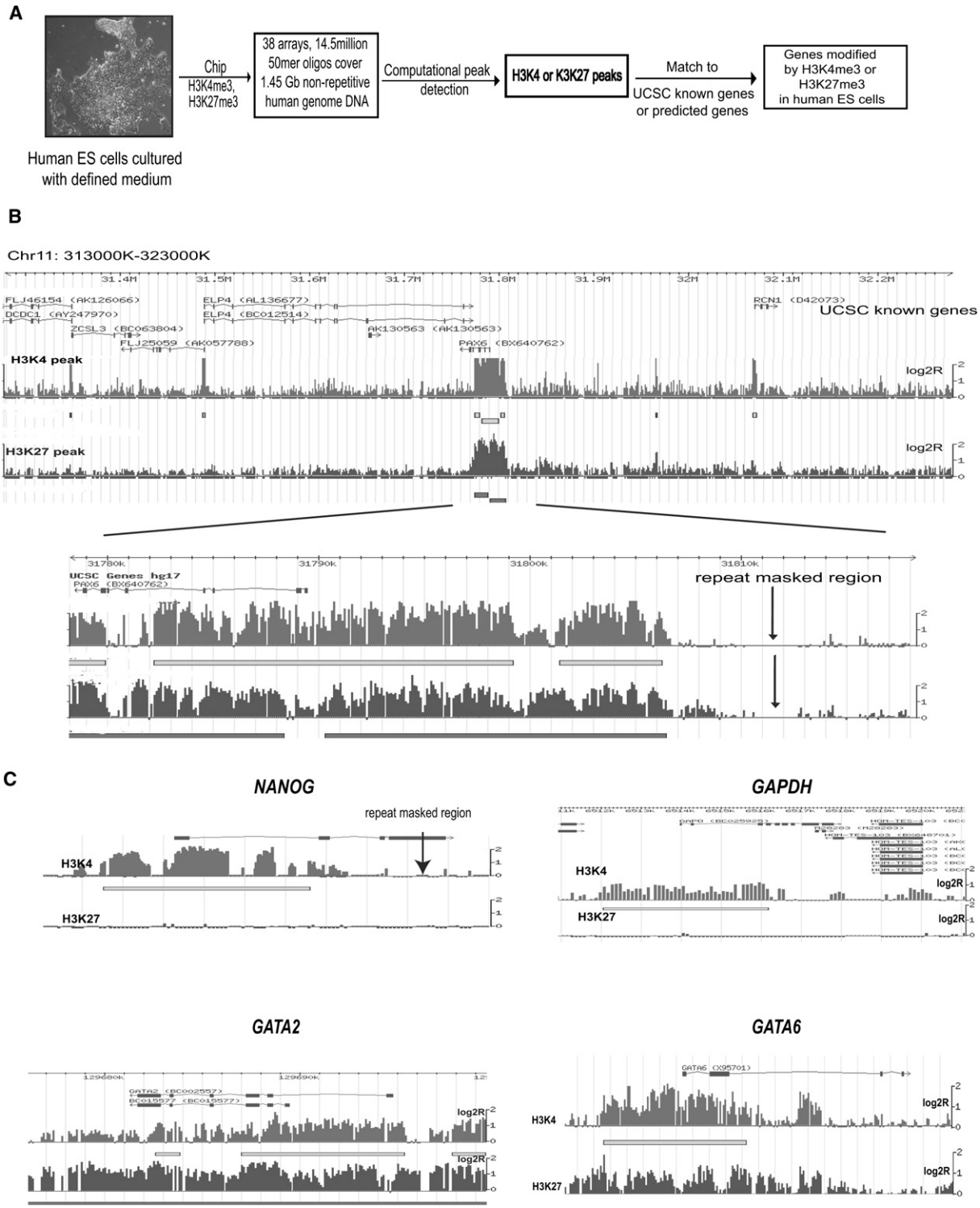



**Figure 1. Whole-Genome Histone Modification Mapping in Human ES Cells**

(A) Strategy of mapping of H3K4 and H3K27 across the whole genome in human ES cells.

(B) Representative view of results of the whole-genome histone ChIP-chip experiments. The logarithmic ratio between H3K4 (gray)- or H3K27 (black)-enriched DNA and input DNA are shown in the middle. UCSC known gene annotations are indicated on the top panel. Bottom panel is a close-up view of gene region with both K4 and K27 enrichments. Arrows point to a representative repeat masked genomic region.

(C) Representative views of histone modifications of pluripotency-associated gene *NANOG*, housekeeping gene *GAPDH*, and early lineage genes *GATA2* and *GATA6*.

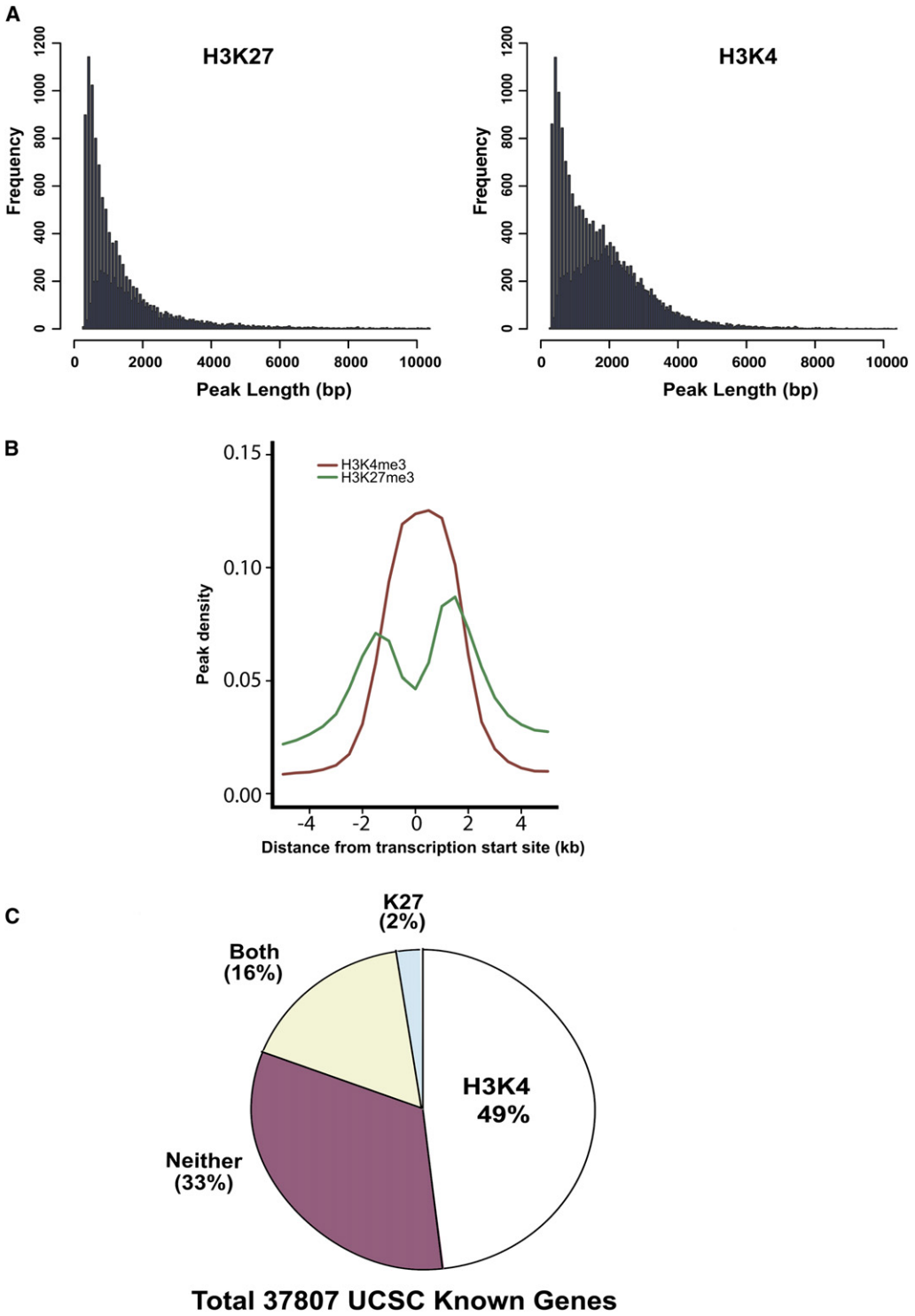



**Figure 2. Whole-Genome Distributions of H3K4me3 and H3K27me3 in Human ES Cells**

(A) Total H3K4- or H3K27-modified regions in human ES cells.

(B) Distributions of H3K4 or H3K27 peaks relative to the TSS of RefSeq annotations. Almost all of the peaks mapped within the 10 kb window around the TSSs fall within ±2 kb of the TSS (99.7% of H3K4me3 peaks, 97.9% of H3K27me3 peaks).

(C) Histone peak patterns of the total UCSC known genes.

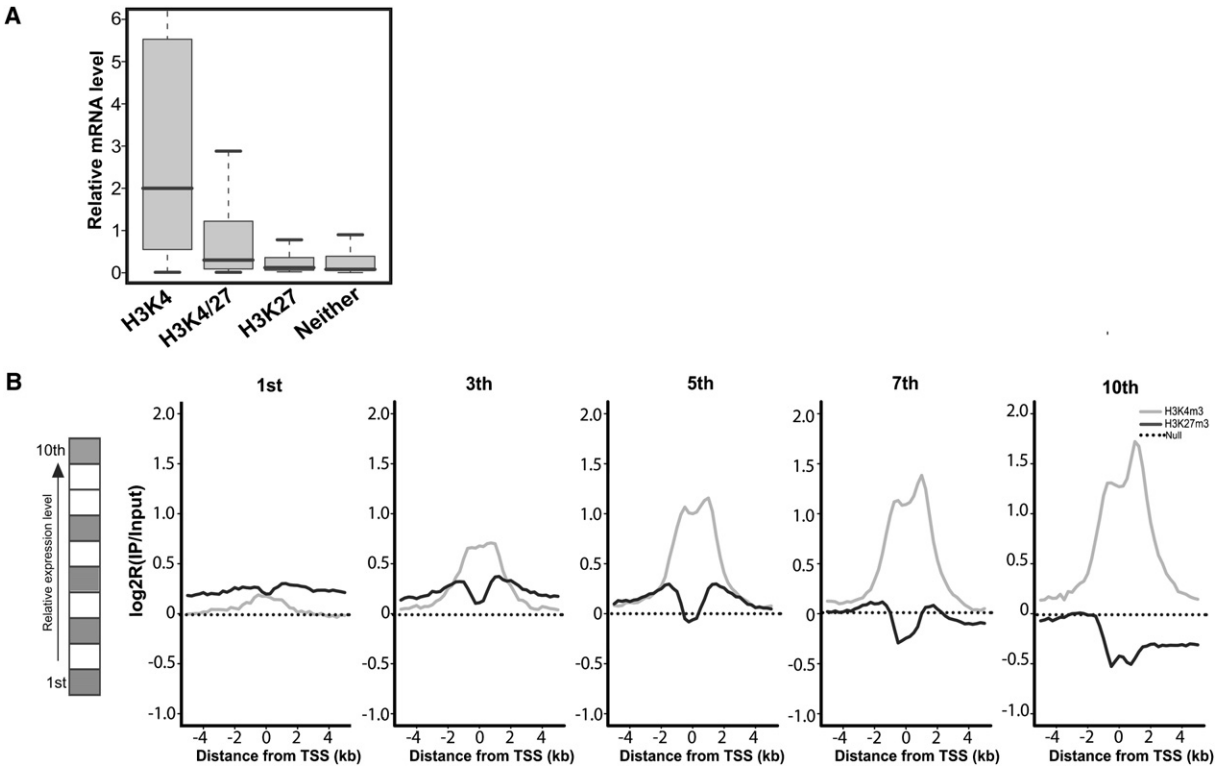



**Figure 3. Correlation of H3K4me3 or H3K27me3 with Expression Levels in Human ES Cells**

(A) Expression levels of genes associated with different histone patterns. Box plot showing 25th, 50th, and 75th percentiles of relative expression levels for genes associated with four different patterns. The error bars represent 1.5 times the interquartile range above and below the mean. The bold line is the mean. The gray box represents the quartiles directly below and above the mean.

(B) The average signal intensity of H3K4me3 and H3K27me3 ±5 kb around the TSS of all genes from the UCSC Known Gene database is displayed for genes broken down by decile of expression (1st being lowest 10% and 10th being highest 10%). Note the dip in H3K27me3 average signal around the TSS as expression increases.

We also compared the genomic positions of H3K4me3 and H3K27me3 peak regions with enhancers and CpG islands (see the Supplemental Experimental Procedures for details). H3K4/H3K27 peak regions are considered to be overlapping with enhancers or CpG islands if they overlap or if the distance between them is less than 50 bp. Of 23,115 H3K4me3 peak regions, 1022 overlap with enhancers and 13,252 overlap with CpG islands. Of these 1022 H3K4me3 regions that overlap with enhancers, 795 also overlap with CpG islands. Of 14,106 H3K27me3 peak regions, 302 overlap with enhancers and 1900 overlap with CpG islands. Of these 302 H3K27me3 regions that overlap with enhancers, 116 also overlap with CpG islands. For H3K4me3 peaks that are at least 2 kb away from the nearest TSS and are associated with enhancers, 93.4% overlap with CpG islands. For H3K27me3 peaks that are at least 2 kb away from the nearest TSS and are associated with enhancers, 70.6% overlap with CpG islands.

## Correlation of H3K4 and K27 Trimethylations with Gene Activity

Previous analysis in other model systems and a smaller-scale study using human cells suggested that H3K4me3 associates with active promoters, while H3K27me3 associates with silenced promoters (Santos-Rosa et al., 2002; Schneider et al., 2003; Schubeler et al., 2004). To examine the relationships of promoter status to the H3K4me3 and H3K27me3 patterns in human ES cells, we compared expression levels of genes associated with different H3K4me3 and/or H3K27me3 modification patterns. Based on the gene expression microarray data of human ES cells, we found that genes associated with just H3K4me3 had the highest level of expression, while those with H3K27me3 alone had the lowest (Figure 3A), consistent with the previous findings. Genes with both histone modifications had relatively low levels of expression that were generally intermediate between expression levels of genes with the single modifications. Interestingly, a substantial portion of genes that lacked both H3K4me3 and H3K27me3 had a very low average expression level that was similar to those genes marked by H3K27me3 alone. Comparisons of the four categories of histone-marked genes (H3K4me3 only, H3K27me3 only, colocalized, and neither mark) with publicly available SAGE data for human ES cells from the Cancer Genome Anatomy Project (CGAP) gave very similar results to the microarray-based comparisons (data not shown).

In addition to analyzing the expression level of genes divided into four categories, we also plotted the average signal intensities for H3K4me3 and H3K27me3 around the TSS for all known genes broken down into five distinct expression categories by decile (1st, 3rd, 5th, 7th, and 10th decile where 1st is lowest expressors and 10th is highest expressors) (Figure 3B). As shown in Figure 3B, the lowest expressors tend to be H3K27me3 dominant or lack both modifications around the TSS (Figure 3B, leftmost panel). As the expression level increases, the average H3K4me3 signal increases, with a concomitant dip in H3K27me3 average signal at the TSS. This H3K27me3 dip actually extends below background for the highly expressed genes (Figure 3B, rightmost panel) and extends well into the gene, suggesting there is a low level of H3K27me3 modifications throughout the genome that is actively removed at the TSS and within genes as genes are expressed.

## Colocalization of H3K27me3 with H3K4me3 on Same Promoters in Human ES Cells

Previous studies on HCNEs in mouse ES cells (2.5% of whole genome) revealed a fair number of genomic regions associated with both H3K4me3 and H3K27me3 (Bernstein et al., 2006). Through whole-genome analysis in human ES cells, we found that most H3K27me3 peaks (90%) localized on promoters that were already marked with H3K4me3, suggesting that colocalization of H3K27me3 with H3K4me3 on the same promoters is a rule in ES cells rather than an exception. A gene is considered to be colocalized if it contains both an H3K27me3 peak and an H3K4me3 peak within ±2.5 kb of its TSS. The vast majority (5659 out of 6060, 93%) of these colocalized genes have an H3K27me3 peak that overlaps at least 1 bp with the H3K4me3 peak. It is possible that the percentage of overlapping regions would be higher if repetitive regions that are masked out of the arrays could be probed. To further examine this colocalization, we plotted the signal intensities around the TSS for both of the modifications for genes in each of the four H3K4me3/H3K27me3 categories to examine whether the threshold used for our peak-finding algorithm produced false negatives that resulted in genes being assigned to incorrect categories (Figure 4A). In the case of the H3K27me3-only category (Figure 4A, second panel), there was an average enrichment for H3K4me3 centered around the TSS that was higher than background, suggesting that at least some genes in this category were misassigned, as some appear to have modest H3K4me3 enrichment that falls below the threshold of our peak finder. In contrast, the gene category lacking both H3K4me3 and H3K27me3 peaks demonstrated no signal enrichment around the TSS for either modification (Figure 4A, right panel), further increasing our confidence that the "neither" category is real. Genes closely associated with pluripotency, such as *NANOG*, *POU5F1* (*OCT4*), and *SOX2*, and housekeeping genes, such as *GAPDH*, were marked solely by H3K4me3 (Figure 1C and Figure S2). Many known developmental control genes, such as *SOX1*, *PAX6* (neural ectoderm); *T*, *HAND1* (mesoderm); *GATA4*, *GATA6* (endoderm); and *GATA2*, *GATA3*, *CDX2* (trophectoderm) were associated with both modifications (Figure 1C and Figure S2). Through ChIP-combined QPCR, we confirmed the H3K4me3-only patterns on the promoters of pluripotent genes *NANOG*, *POU5F1*, and *SOX2* and housekeeping gene *GAPDH* and colocalization of both markers on the promoters of developmental genes (Figure 4B). To exclude the possibility that this colocalization might be due to mixed-cell populations, we performed sequential ChIP first to H3K27me3 then to H3K4me3 (Figure 4C) on the same sample. Compared with genes solely modified by H3K4me3, each developmental gene again showed significant enrichments of H3K27me3 modifications on their promoters (Figure 4C).

## Distinct Biological Functions of Genes that Are Associated with Different Combinations of H3K4 and H3K27 Trimethylation

To gain insight into biological functions of genes associated with different combinations of histone modifications, we performed gene ontology (GO) analysis of the genes in each of the four categories. We found that significant enrichments for genes associated with both modifications were related to developmental functions (Figure 5). Many known critical transcription factors were identified from this gene set. In all, we identified 587 factors with combined H3K4 and H3K27 trimethylation that are involved in transcriptional regulation, including key regulators of early development such as HOX factors, GATA factors, forkhead factors, and homeodomain factors (Table S5). Many of these factors were previously identified to be targets of the Polycomb complex both in human and mouse ES cells (Boyer et al., 2006; Lee et al., 2006), indicating that, in ES cells, the Polycomb complex-based modifications correlate with genes modified by H3K4me3.

In contrast to the H3K4/H3K27me3-overlapping genes, genes that carried neither H3K4 nor H3K27 trimethylation were not enriched in developmental function but in functions related to signal transduction and physiological response to stimuli (Figure 5). These genes were expressed at low levels in ES cells. For example, rhodopsin-like G protein receptor (GPCR) superfamily members, olfactory receptors, and immune-response factors were highly enriched in this category (Table S6). The genes with only the H3K4 modification were strongly associated with basic physiologic processes common to all cells, probably reflecting the fact that a large number of expressed housekeeping genes fall into this category. Also, in addition to transcription factors related to ES pluripotency, genes in signaling pathways active in human ES cells, such as in the FGF-MAP kinase pathway, were generally marked by H3K4me3 alone (Figure S3). Thus, H3K4me3 marks a set of genes that maintain fundamental properties of human ES cells. GO terms for the genes marked solely with H3K27me3 had enrichment with p values of much lower significance (data not shown), possibly reflecting the small size of this category but also decreasing our confidence that this represents a real category.

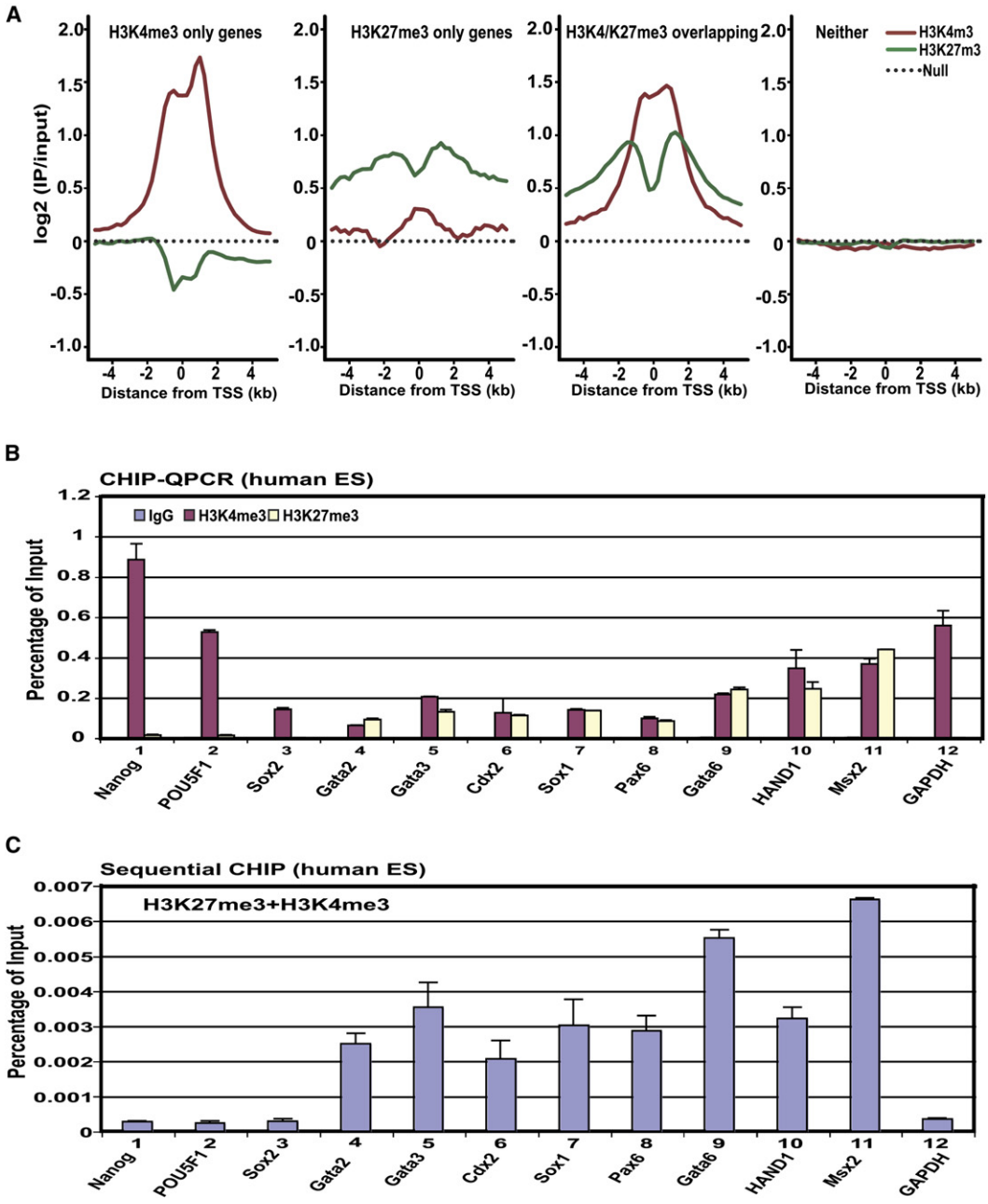



**Figure 4. Colocalization of H3K27me3 and H3K4me3 in Human ES Cells**

(A) The average signal intensity of H3K4me3 and H3K27me3 ±5 kb around the TSS for genes in each of the four different H3K4/H3K27 categories. The dotted black line represents the normalized average signal over the whole chip.

(B) ChIP-QPCR data confirmed the histone modifications on the promoters identified by whole-genome analysis.

(C) Sequential ChIP-combined QPCR data showed histone modification enrichments on the H3K4me3/H3K27me3 promoters, not the H3K4me3-only promoters. The error bars indicate the SD (standard deviation) on three triplicate PCR results.

## Genes Can Change Expression Levels Rapidly during Differentiation Regardless of Their H3K4/H3K27 Trimethylation Status

Previous studies in mouse ES cells proposed that H3K4me3/H3K27me3 overlapping on the same promoter (bivalent) kept the target gene in a poised state for later activation upon ES cell differentiation (Bernstein et al., 2006).

To examine the relationship between specific histone modifications and developmental timing of gene expression, we monitored gene expression changes of H1 human ES cells during retinoic acid (RA)-induced differentiation and correlated those changes with initial H3K4/H3K27 trimethylation status. By 5 days of RA treatment, H1 ES cells changed morphology and downregulated *OCT4*

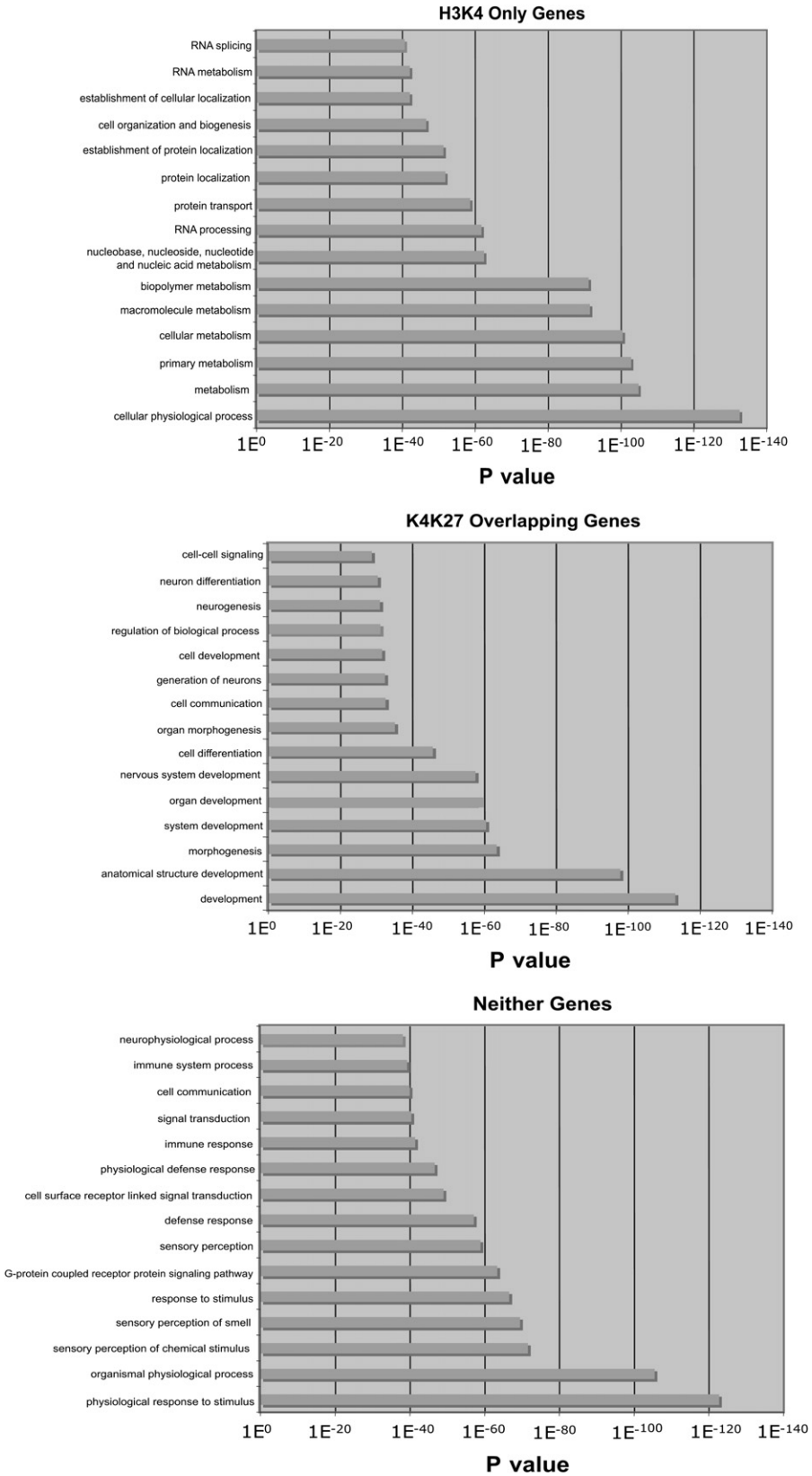



**Figure 5. GO Analysis of Genes Associated with Different Histone Patterns**

The y axis shows the GO term, and the x axis shows the p value for the significance of enrichment for the top 15 GO terms.

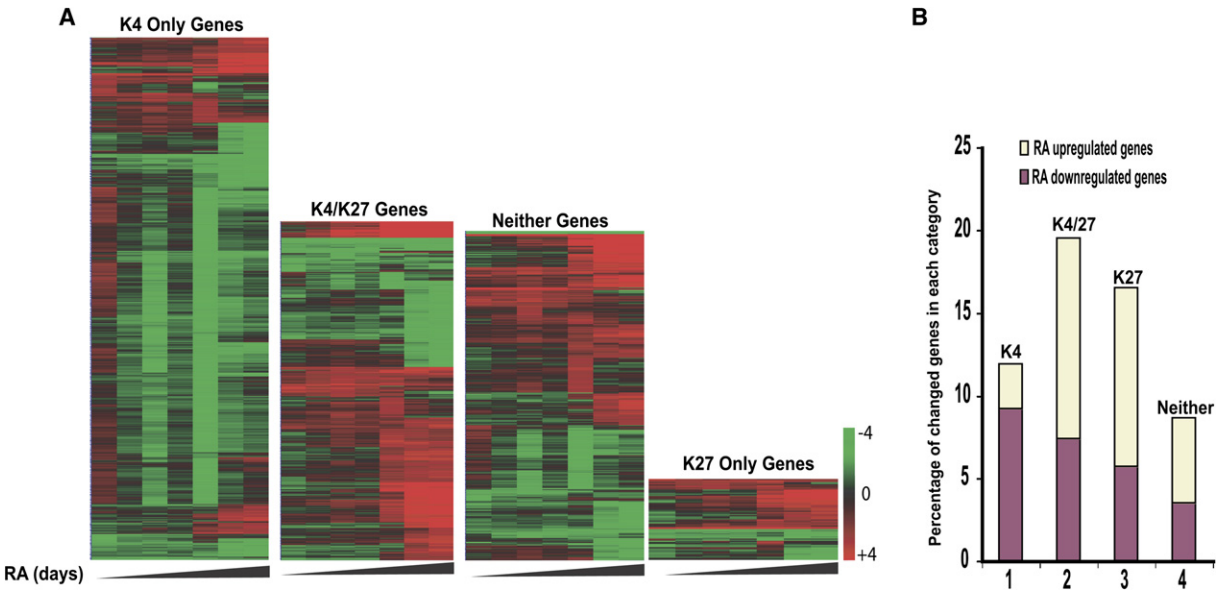



**Figure 6. Regulation Profiles of Genes Associated with Different Histone Patterns during ES Cell Differentiation**

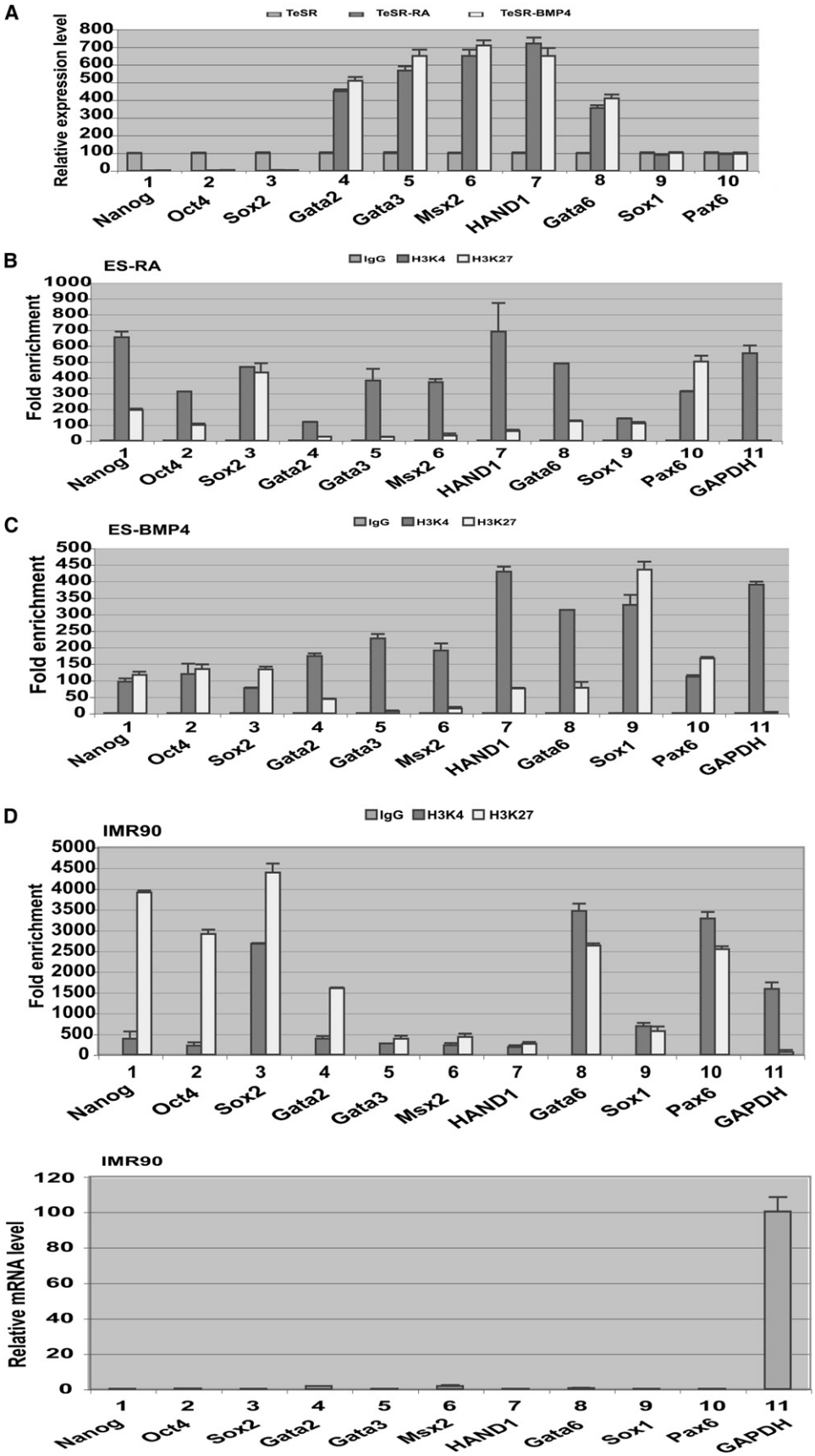
(A) Cluster analysis of gene expression profiles of changed genes in each category during RA-induced differentiation.

(B) Percentage of genes in each category that changed expression at least 3-fold at one or more time points. Genes that were upregulated or downregulated at least 3-fold in one or more tested time points were separated and shown in different colors.

(Figure S1). During a 5 day time course of RA differentiation, 11.9% of H3K4-only genes, 19.5% of H3K4/H3K27-overlapping genes, 16.5% of H3K27-only genes, and 8.7% of genes marked with neither modification showed at least a 3-fold change for one or more time points examined (Figure 6 and Table S7). For the genes that altered expression, genes modified solely by H3K4me3 tended to be downregulated, perhaps not surprisingly, as this gene set contained many pluripotency-associated genes that were expressed at high levels initially. However, a significant number of H3K4me3-only genes were nonetheless upregulated. In contrast, most genes that changed expression and initially were associated with combined H3K4/H3K27 trimethylation or H3K27me3 alone were upregulated during differentiation, again perhaps not surprisingly, as these genes were expressed at low levels initially and related to development function. Interestingly, a fair portion of genes that lack both modifications can also rapidly change their expression upon differentiation, albeit the percentage of changed genes was lower than that of H3K4/H3K27me3-overlapping genes (Figure 6B). Thus, although the genes marked with combined H3K4/H3K27 trimethylation were relatively enriched for genes that upregulated during differentiation, a significant number of genes in each of the four initial histone modification categories could change expression levels rapidly during differentiation. Our data suggest that regulation profiles of gene expression during ES cell differentiation are related to their biological function (i.e., developmental genes turn on during development) but that the relative transcriptional responsiveness of genes is not strongly related to initial H3K4/H3K27 trimethylation status.

## Changes of H3K4/H3K27 Trimethylation Patterns upon Differentiation—Combined H3K4/H3K27 Trimethylation Is a General Indicator of Repression

To examine how the patterns of H3K4/H3K27 trimethylation changed upon differentiation, we examined several transcription factors that were upregulated (*GATA2*, *GATA3*, *MSX2*, *HAND1*, and *GATA6*), downregulated (*NANOG*, *OCT4*, and *SOX2*), or remained not expressed or expressed at low levels (*SOX1* and *PAX6*) during RA- or BMP4-induced differentiation (Figure 7A and Figure S1). ChIP combined with quantitative PCR confirmed that, in ES cells, housekeeping gene *GAPDH* and ES-cell-enriched genes *NANOG*, *OCT4*, and *SOX2* exhibited H3K4me3 alone, while lineage-specific transcription factors were marked with both H3K4 and H3K27 trimethylation (Figure 4B). However, in both RA- and BMP4-differentiated cells when the ES-cell-specific genes were downregulated, they became marked by both H3K4 and H3K27 trimethylation (Figures 7B and 7C). The Oct4-negative status of the RA- and BMP4-differentiated cells rules out that this is due to a mixed population of cells after differentiation (Figure S1). In contrast, lineage-specific transcription factors that were upregulated in the differentiated cells became marked by H3K4me3 alone (Figures 7B and 7C). The housekeeping gene *GAPDH* maintained the H3K4me3 pattern in both undifferentiated and differentiated cells. The neural lineage-specific genes, *SOX1* and *PAX6*, that were not induced kept the combined H3K4/H3K27 trimethylation pattern in differentiated cells. Thus, bivalent H3K4/H3K27 trimethylation is not unique to ES cells, but ES-cell-specific genes and developmental genes suppressed during differentiation can also maintain or


A
TeSR
TeSR-RA
TeSR-BMP4
Relative expression level
800
700
600
500
400
300
200
100
0
1 2 3 4 5 6 7 8 9 10
Nanog Oct4 Sox2 Gata2 Gata3 Msx2 HAND1 Gata6 Sox1 Pax6
B
IgG H3K4 H3K27
ES-RA
Fold enrichment
1000
900
800
700
600
500
400
300
200
100
0
1 2 3 4 5 6 7 8 9 10 11
Nanog Oct4 Sox2 Gata2 Gata3 Msx2 HAND1 Gata6 Sox1 Pax6 GAPDH
C
IgG H3K4 H3K27
ES-BMP4
Fold enrichment
500
450
400
350
300
250
200
150
100
50
0
1 2 3 4 5 6 7 8 9 10 11
Nanog Oct4 Sox2 Gata2 Gata3 Msx2 HAND1 Gata6 Sox1 Pax6 GAPDH
D
IgG H3K4 H3K27
IMR90
Fold enrichment
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
0
1 2 3 4 5 6 7 8 9 10 11
Nanog Oct4 Sox2 Gata2 Gata3 Msx2 HAND1 Gata6 Sox1 Pax6 GAPDH
IMR90
Relative mRNA level
120
100
80
60
40
20
0
1 2 3 4 5 6 7 8 9 10 11
Nanog Oct4 Sox2 Gata2 Gata3 Msx2 HAND1 Gata6 Sox1 Pax6 GAPDH

acquire this dual mark during differentiation. The colocalization of H3K4/H3K27 trimethylation is not likely to be caused by a heterogeneous population of differentiated cells, as FACS analysis shows a clear shift of OCT4 signal upon differentiation (Figure S1).

To rule out the possibility that the change of the H3K4/H3K27 trimethylation patterns in differentiated cells was merely an artifact arising from in vitro ES cell differentiation, we also examined the same set of transcription factors in IMR90 cells, a diploid human lung fibroblast cell line. In the IMR90 cells, none of the selected ES-cell-specific or lineage-specific genes were expressed at significant levels (Figure 7D). We found that ES-cell-specific genes still exhibited both H3K4 and H3K27 trimethylation in the IMR90 cells, but the enrichments for H3K27me3 were often increased compared to that for H3K4me3. Lineage-specific transcription factors that were not expressed in IMR90 cells also exhibited detectable levels of both H3K4 and H3K27 enrichments. Housekeeping gene *GAPDH* continued to exhibit H3K4me3 alone.

## DISCUSSION

Previous work documented the unexpected colocalization of H3K4me3 and H3K27me3 in mouse ES cells in a highly conserved region consisting of about 2.5% of the genome, and it was suggested that this bivalent state was essential for genes in pluripotent cells to remain "poised" for activation (Bernstein et al., 2006). We extended those previous studies by analyzing these modifications across the whole genome in human ES cells and found that there are indeed a substantial number of genes marked by both H3K4 and H3K27 trimethylation and that these genes are globally enriched for genes involved in development. In fact, of the genes that were marked by H3K27me3, 90% were also marked by H3K4 trimethylation, and these colocalized modifications centered on the TSS. Although we failed to detect colocalized H4K4me3 peaks for 10% of the H3K27me3-associated genes, it is worth noting that even genes expressed at the lowest levels had on average an enrichment for H3K4me3 signal around the TSS (Figure 3B), suggesting that at least some of these genes are also comodified but were missed because of global optimization of false-positive and false-negative rates during peak finding or other limitations of the technology. Thus, in the vast majority of cases in which H3K27me3 is present, it colocalizes with genes marked with H3K4me3 in human ES cells.

Transcription factors directly involved in pluripotency, such as *POU5F1*, *NANOG*, and *SOX2*, were initially marked by H3K4me3 in ES cells but became modified by both H3K4me3 and H3K27me3 as they were downregulated during differentiation. These results indicate a role for Polycomb group proteins in suppressing pluripotency-associated genes during differentiation and indicate that the colocalization of H3K4 and H3K27 trimethylation is not only restricted to ES cells to keep developmental genes poised for activation, as these genes are not destined to be reactivated during development. Others have previously reported the localization of H3K27me3 at the promoters of *POU5F1* and *SOX2* in human T cells (Azuara et al., 2006), but this modification was not detected in mouse neuronal precursor cells (Boyer et al., 2006), suggesting tissue specificity in either the establishment or the maintenance of this modification.

In RA- and BMP4-differentiated ES cells, lineage-specific genes that were not induced (*SOX1* and *PAX6*) continue to be marked by both H3K4 and H3K27 trimethylation. Similarly, both modifications were generally detected for lineage-specific genes that were not expressed in a human fibroblast cell line, although the ratios of enrichment sometimes changed favoring the H3K27 modification (Figure 7). A recent report also suggests that the colocalization of the H3K4 and H3K27 trimethylation in human T cells is not unusual (Roh et al., 2006). Roh et al. used a genome-wide mapping technique (GMAT) that combines ChIP with serial analysis of gene expression protocols and identified 5252 promoters marked with H3K27me3 in primary human T cells, and 3330 of them colocalized with H3K4me3. Clearly, how widespread these dual modifications are in differentiated cells and how the ratio and distribution of the two modifications change with lineage restriction merits additional study, but these results nonetheless indicate that this pattern of modification is not highly specific to pluripotent cells but reflects more general regulatory mechanisms. Consistent with our findings, very recently Mikkelsen and colleagues mapped H3K4me3 and H3K27me3 modifications genome-wide in mouse ES cells, neural progenitor cells (NPC), and mouse embryonic fibroblast cells (MEF) and found that a small portion (4%) of genes marked with H3K4me3 only in ES cells acquired H3K27me3 in NPC cells or MEF cells as these genes were suppressed (Mikkelsen et al., 2007), indicating a conserved role for the PcG complex in suppressing genes during mouse and human development. In addition, they also showed in their work that mouse promoters containing high CpG island content (HCP genes), intermediate CpG island content (ICP genes), and low CpG island content (LCP genes) differ in their correlation with H3K4me3 and H3K27me modification. We compared the H3K4me3

**Figure 7. Control of Lineage-Specific Genes by Histone Patterns during ES Cell Differentiation**

(A) Relative expression levels of indicated pluripotent genes and lineage genes in undifferentiated ES cells (TeSR) and differentiated cells triggered by RA or BMP4. Expression levels of all tested genes in undifferentiated cells were set to 100 for comparison.

(B and C) H3K4 or H3K27 trimethylations of pluripotent genes and lineage genes in undifferentiated and differentiated cells triggered by RA or BMP4. The histone states of the indicated genes were determined by ChIP and real-time PCR in H1 ES cells maintained in TeSR medium or treated with RA (1 μM) or BMP4 (50 ng/ml) for 5 days.

(D) Histone states and expression levels of indicated pluripotent genes and lineage genes in terminally differentiated cells, IMR90. The error bars indicate the SD (standard deviation) on three triplicate PCR results.

andH3K27me3 status of the human orthologs of mouse HCP, ICP, and LCP genes. In general, our results in human are consistent with those from mouse, specifically (1) a high percentage (93% for human, 99% for mouse) of HCP genes contain H3K4me3 marks, (2) a low percentage (28% for human, 7% for mouse) of LCP genes contain H3K4me3 marks, (3) a near equal percentage of HCP genes contain colocalized marks (24% for human, 22% for mouse), and (4) H3K27me3 marks alone are near absent from both HCP and LCP genes (see Table S12 and the Supplemental Experimental Procedures). These results suggest that, like the mouse, human HCP and LCP genes may have distinct modes of regulation.

We also identified a substantial class of genes in human ES cells (33% of the UCSC known gene set) that lacked both H3K4 and H3K27 trimethylation. The exceptionally low expression profile (Figure 3A), the complete absence of H3K4me3 and H3K27me3 signal enrichment (Figure 3B), and the distinctive functional classification (Figure 5) all suggest that this represents a real group. Unlike genes marked by both H3K4 and H3K27 trimethylation, GO analysis of this group of genes failed to reveal a significant enrichment for genes involved in development but enriched for categories of genes involved in the response to transient external stimuli (Figure 5). For example, rhodopsin-like G protein receptors that represent a widespread protein family that transduce extracellular signals were highly enriched in this category, as were cytokines involved in the immune response (Figure S4). At least for human T cells, genes in this category tend to acquire H3K4me3 when they become expressed (Roh et al., 2006). Since this group lacks H3K27me3 in human ES cells, their relative inactivity appears to be Polycomb independent, yet how they are repressed remains unknown. However, this class of genes suggests fundamental differences in gene regulation during development and in response to transient stimuli.

Our results offer additional evidence that H3K27me3 catalyzed by Polycomb complexes mediates transcriptional repression of key developmental control genes and indicate that this repression is accomplished generally in the presence of H3K4me3 in human ES cells. Our results suggest that, although colocalization of H3K4me3 and H3K27me3 does enrich for genes that can upregulate during early differentiation, a substantial number of genes that lack this specific combination of modifications in human ES cells can respond quickly also. Our finding that key transcription factors involved in pluripotency acquire H3K27me3 and maintain H3K4me3 after downregulation during differentiation demonstrates that only a subset of potential Polycomb complex target genes is actively repressed in human ES cells and that recruitment of Polycomb complexes can quickly change, either to relieve existing or to create new repressive domains. Given that the pluripotency-associated transcription factors are generally not reactivated during development, it is hard to argue that these particular genes are physiologically poised for activation, although genes maintaining this dual mark might indeed be more amenable to reprogramming. Identifying the transcriptional partners that mediate these early dynamic switches in Polycomb specificity should offer further important insight into how the pluripotent state is maintained and exited or how it can be achieved through reprogramming of differentiated cells.

## EXPERIMENTAL PROCEDURES

### Cells and Cell Culture

#### *Antibodies, Chromatin Immunoprecipitation, and Whole-Genome Microarrays*

Trimethyl-histone H3K4 or K27 antibodies used for the ChIP experiments were purchased from Upstate. ChIP-chip analysis was performed as described (see reference, Turner, 2000) with slight modification. Briefly, $1 \times 10^9$ H1 cells were collected and chemically crosslinked by 1% formaldehyde. Cells were rinsed twice with PBS ($-Ca^{2+}-Mg^{2+}$) and lysed using lysis buffer (0.1% SDS, 0.5% Triton X-100, 20 mM Tris-HCl [pH 8.1], 150 mM NaCl, protease inhibitor, Roche). Lysed cells were sonicated using a Misonix Sonicator 3000 at power 4 for 20 15 s pulses with 30 s intervals between each pulse in ice water. Sonicated fragments ranged in size from 250 to 500 bp. After sonication, samples were centrifuged at 13,000 × g for 10 min. The supernatant was preabsorbed by 50 μl protein A/G beads (Santa Cruz) and incubated with 10 μg antibodies (antiH3K4me3, Upstate, 07-473, antiH3K27me3, Upstate 05-851) overnight at 4°C. The immunocomplex was collected by 100 μl protein A/G beads (Santa Cruz). The beads were then washed four times by lysis buffer, two times by LiCl buffer (0.25 M LiCl, 1% NP-40, 1% deoxycholate, 1 mM EDTA, 10 mM Tris-HCl [pH 8.1]), and three times by TE buffer. The bound immunocomplex was eluted by adding 300 μl fresh elution buffer (1% SDS, 0.1 M $NaHCO_3$). Twenty microliters of 5M NaCl was then added to the eluted product and incubated at 65°C overnight to reverse the crosslinking. Immunoprecipitated genomic DNA was then purified using a QIAGEN Purification Kit, blunted, ligated to linker, and amplified by PCR (ligation-medicated PCR). Additional PCR reactions were performed to generate a total 160 μg of ChIP DNA for the final hybridization.

Of immunoprecipitated or total genomic LM-PCR DNA, 4 μg was labeled with Cy3 or Cy5 and combined to hybridize to NimbleGen human whole-genome arrays that were manufactured by NimbleGen Systems (www.nimblegen.com) as described (Kim et al., 2005). Array design, data extraction, and specification can be found in the Supplemental Experimental Procedures (Kim et al., 2005).

For the analysis of histone states in differentiated cells, ChIP combined with quantitative PCR was used. The amplified genomic regions for the tested genes were chosen based on the ChIP-chip data. Oligonucleotide sequences for the PCR are provided in the Supplemental Data. For the sequential ChIP, crosslinked chromatin from ES cells was first immunoprecipitated by anti-H3K27me3 as regular ChIP except that the chromatin was eluted by 30 mM DTT, 500 mM NaCl, and 0.1% SDS. The eluted chromatin was then subject to a second round of immunoprecipitation with anti-H3K4me3 and followed by standard ChIP procedure.

#### *Identification of H3K4- and H3K27-Associated Regions and Mapping of Genes*

Detailed descriptions of H3K4 or K27 peak finding are provided in the Supplemental Data. Briefly, after scanning and image extraction, IP and Input signal values for each of the 38 arrays were normalized according to Qspline (Workman et al., 2002). Each pair of normalized probe signals between IP and Input was then converted into log ratios based on the following function: LogR(i) = Log (IP[i]/Input[i]). To identify the H3K4- or K27-associated genomic regions, we developed a simple statistical approach to define putative H3K4- and H3K27-modified regions. For each array, the initial window was set as regions with a minimum of four continuous probes (representing 350 bp) with logR >1. If two adjacent peaks were separated by a probe-containing gap of <500 bp, they were merged into one peak. Furthermore, if two peaks

were separated by a repeat-masked region (for which there are no probes) of less than <1000 bp, these were also merged into one region. For gene mapping, we choose the gene region extending ±2.5 kb from the TSS to measure their modification state.

### Gene Expression Analysis

For gene expression analysis, we isolated the total RNA from H1 ES cells or RA-treated cells using RNeasy Mini Kit (QIAGEN, Valencia, CA) according to the manufacturer's recommendations. PolyA RNA was then isolated using Oligotex mRNA Mini Kit (QIAGEN). The mRNA were then reverse transcribed, labeled, mixed with differently labeled sonicated genomic DNA, and hybridized to a single standard NimbleGen HG17 gene expression array that covers transcripts from ~36,000 human loci (NimbleGen Systems). Detailed descriptions of array design and data normalization are provided in the Supplemental Experimental Procedures. The relative expression level was calculated as described in the Supplemental Experimental Procedures. We set the expression level of genes in undifferentiated cells as 1 and calculated the fold change of individual genes in each treatment.

### Analysis of the Expression of Pluripotent or Lineage Genes in Undifferentiated or Differentiated Cells

To test the transcript levels of genes in ES cells or factor-treated cells, total RNA was isolated from $10^6$ H1 ES cells maintained in TeSR or treated by BMP4 (50 ng/ml) or RA (1 μM) for 5 days using QIAGEN RNeasy Mini Kit in combination with on-column DNase treatment according to the manufacturer's instructions. Superscript II reverse transcriptase (Invitrogen) and Oligo-dT were used to synthesis the first strand of cDNA. Real-time PCR was performed on the 7300 ABI detection system using SYBR green PCR master mix (ABI). The primers used for PCR can be found in the Table S9.

### Supplemental Data

Supplemental Data include Supplemental Experimental Procedures, 12 tables, and 5 figures and can be found with this article online at http://www.cellstemcell.com/content/full/1/3/299/DC1/.

## ACKNOWLEDGMENTS

We thank Palmer Yu and Garrett Lee for technical support and Deborah Faupel for critical reading of the manuscript. We thank Karen Jenny Heidarsdottir and Bryndis Krogh Gisladottir for producing the microarray data. We thank Bing Ren for sharing H3K4me3 HeLa validation data prior to publication and for comments on the manuscript. This work was supported by funding from private donations, the W.M. Keck Foundation, and the WiCell Research Institute. Please address G.P. with requests for materials. J.A.T. is a cofounder and shareholder of Cellular Dynamics International and Stem Cell Products, Inc.

Received: June 15, 2007
Revised: July 13, 2007
Accepted: August 10, 2007
Published: September 12, 2007

## REFERENCES

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., et al. (2006). Chromatin signatures of pluripotent cell lines. Nat. Cell Biol. *8*, 532–538.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. Roy. Statist. Soc. Ser. B. Methodological *57*, 289–300.

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.

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.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. *20*, 1123–1136.

Byrd, K.N., and Shearn, A. (2003). ASH1, a Drosophila trithorax group protein, is required for methylation of lysine 4 residues on histone H3. Proc. Natl. Acad. Sci. USA *100*, 11535–11540.

Cosgrove, M.S., and Wolberger, C. (2005). How does the histone code work? Biochem. Cell Biol. *83*, 468–476.

Fischle, W., Wang, Y., and Allis, C.D. (2003). Histone and chromatin cross-talk. Curr. Opin. Cell Biol. *15*, 172–183.

Guenther, M.G., Jenner, R.G., Chevalier, B., Nakamura, T., Croce, C.M., Canaani, E., and Young, R.A. (2005). Global and Hox-specific roles for the MLL1 methyltransferase. Proc. Natl. Acad. Sci. USA *102*, 8603–8608.

Jenuwein, T., and Allis, C.D. (2001). Translating the histone code. Science *293*, 1074–1080.

Jurgens, G. (1985). A group of genes controlling the spatial expression of the bithorax complex in Drosophila. Nature *316*, 153–155.

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.

Kirmizis, A., Bartley, S.M., Kuzmichev, A., Margueron, R., Reinberg, D., Green, R., and Farnham, P.J. (2004). Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. Genes Dev. *18*, 1592–1605.

Lee, J.H., Hart, S.R., and Skalnik, D.G. (2004). Histone deacetylase activity is required for embryonic stem cell differentiation. Genesis *38*, 32–38.

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.

Lewis, E.B. (1978). A gene complex controlling segmentation in Drosophila. Nature *276*, 565–570.

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. 10.1371/journal.pbio.0030328.

Ludwig, T.E., Bergendahl, V., Levenstein, M.E., Yu, J., Probasco, M.D., and Thomson, J.A. (2006a). Feeder-independent culture of human embryonic stem cells. Nat. Methods *3*, 637–646.

Ludwig, T.E., Levenstein, M.E., Jones, J.M., Berggren, W.T., Mitchen, E.R., Frane, J.L., Crandall, L.J., Daigh, C.A., Conard, K.R., Piekarczyk, M.S., et al. (2006b). Derivation of human embryonic stem cells in defined conditions. Nat. Biotechnol. *24*, 185–187.

Margueron, R., Trojer, P., and Reinberg, D. (2005). The key to development: interpreting the histone code? Curr. Opin. Genet. Dev. *15*, 163–176.

Martens, J.H., O'Sullivan, R.J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J. *24*, 800–812.

Meshorer, E., and Misteli, T. (2006). Chromatin in pluripotent embryonic stem cells and differentiation. Nat. Rev. Mol. Cell Biol. *7*, 540–546.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. Dev. Cell *10*, 105–116.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Nagy, P.L., Griesenbeck, J., Kornberg, R.D., and Cleary, M.L. (2002). A trithorax-group complex purified from Saccharomyces cerevisiae is required for methylation of histone H3. Proc. Natl. Acad. Sci. USA *99*, 90–94.

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.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. Nature *419*, 407–411.

Schneider, R., Bannister, A.J., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2003). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat. Cell Biol. *6*, 73–77.

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.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat. Rev. Genet. *8*, 9–22.

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.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *482*, 1145–1147.

Turner, B.M. (2000). Histone acetylation and an epigenetic code. Bioessays *22*, 836–845.

Workman, C., Jensen, L.J., Jarmer, H., Berka, R., Gautier, L., Nielser, H.B., Saxild, H.H., Nielsen, C., Brunak, S., and Knudsen, S. (2002). A new non-linear normalization method for reducing variability in DNA microarray experiments. Genome Biol. *3*, research0048. Published online August 30, 2002. 10.1186/gb-2002-3-9-research0048.

## Accession Numbers

The raw microarray data are available in the public microarray Gene Expression Omnibus database under accession numbers GSE8439 and GSE8463.