Cell, Vol. 120, 169–181, January 28, 2005, Copyright ©2005 by Elsevier Inc. DOI 10.1016/j.cell.2005.01.001

Genomic Maps and Comparative Analysis of Histone Modifications in Human and Mouse

Bradley E. Bernstein,[1,2,3,6,*] Michael Kamal,[2,6] Kerstin Lindblad-Toh,[2] Stefan Bekiranov,[4] Dione K. Bailey,[4] Dana J. Huebert,[1,2] Scott McMahon,[1,2] Elinor K. Karlsson,[2] Edward J. Kulbokas III,[2] Thomas R. Gingeras,[4] Stuart L. Schreiber,[1,2] and Eric S. Lander[2,5,*]

[1]Howard Hughes Medical Institute at the Department of Chemistry and Chemical Biology Harvard University 12 Oxford Street Cambridge, Massachusetts 02138

[2]Broad Institute of Harvard and MIT Cambridge, Massachusetts 02139

[3]Department of Pathology Brigham and Women's Hospital and Harvard Medical School Boston, Massachusetts 02115

[4]Affymetrix 3380 Central Expressway Santa Clara, California 95051

[5]Whitehead Institute for Biomedical Research Cambridge, Massachusetts 02139

Summary

We mapped histone H3 lysine 4 di- and trimethylation and lysine 9/14 acetylation across the nonrepetitive portions of human chromosomes 21 and 22 and compared patterns of lysine 4 dimethylation for several orthologous human and mouse loci. Both chromosomes show punctate sites enriched for modified histones. Sites showing trimethylation correlate with transcription starts, while those showing mainly dimethylation occur elsewhere in the vicinity of active genes. Punctate methylation patterns are also evident at the cytokine and IL-4 receptor loci. The Hox clusters present a strikingly different picture, with broad lysine 4-methylated regions that overlay multiple active genes. We suggest these regions represent active chromatin domains required for the maintenance of Hox gene expression. Methylation patterns at orthologous loci are strongly conserved between human and mouse even though many methylated sites do not show sequence conservation notably higher than background. This suggests that the DNA elements that direct the methylation represent only a small fraction of the region or lie at some distance from the site.

Introduction

Eukaryotic genomes are organized into chromatin, a heterogeneous higher-order structure with central roles in transcription, replication, and repair. The fundamental unit of chromatin is the nucleosome, which consists of 146 base pairs (bp) of DNA wrapped around an octamer of histone proteins. Histones are subject to a variety of posttranslation modifications including acetylation, methylation, and phosphorylation. These modifications regulate chromatin structure and function in part by recruiting additional proteins (Jenuwein and Allis, 2001; Schreiber and Bernstein, 2002; Turner, 2002). Since histone modifications can be maintained through cell division, they may encode heritable "epigenetic" information. The importance of these and other epigenetic mechanisms in cancer and other diseases is becoming increasingly recognized (Feinberg and Tycko, 2004; Wolffe and Matzke, 1999).

Genomic studies of chromatin in model systems have provided general insight regarding the functions of specific histone modifications. For example, chromatin immunoprecipitation (ChIP) has been combined with DNA microarrays to map the locations of modified histones genome-wide in yeast and flies (Bernstein et al., 2002; Schubeler et al., 2004). These studies showed that histone H3 lysine 4 (Lys4) methylation is associated with the transcribed portions of active genes. Quantitative PCR-based approaches have been used to map histone modification patterns at the β-globin locus in mouse and chicken (Bulger et al., 2003; Litt et al., 2001; Schneider et al., 2004). Although of more modest scales, these studies identified associations between domains, genes and regulatory elements, and modification patterns not found in yeast or flies. Recently, a genomic study of chromatin fiber structure in human cells showed that regions of high gene density tend to adopt an open chromatin structure not necessarily dependent on the activation state of the genes (Gilbert et al., 2004).

Because these findings suggest that chromatin in higher eukaryotes has distinct properties from that in model systems, we carried out a large-scale study of histone modification patterns in human and mouse cells. A combination of ChIP, linear DNA amplification, and tiling oligonucleotide arrays was used to interrogate a series of genomic regions at high resolution. Here, we present maps of H3 Lys4 di- and trimethylation, and H3 Lys9/14 acetylation for human chromosomes 21 and 22, generated for the human hepatoma cell line HepG2. In addition, we present maps of Lys4 dimethylation for six genomic loci that are orthologous in human and mouse, including the cytokine cluster, the IL-4 receptor (IL4R) region, and all four Hox clusters. These were generated for primary human and mouse fibroblasts, both of lung origin. Together, the maps detail more than 39 million base pairs (MB) of genome and thus represent coverage three orders of magnitude greater than in prior studies.

This large-scale study revealed a number of remarkable features of mammalian chromatin structure. Lys4 methylation occurs in several distinct genomic distributions. Punctate sites that show trimethylation strongly correlate with transcription starts of active genes and are also enriched for H3 acetylation and RNA polymerase II phosphorylated at serine 5 (Ser5). Punctate sites that show mainly dimethylation reside elsewhere in the vicinity of active genes. In marked contrast, broad meth-

*Correspondence: lander@broad.mit.edu (E.S.L.); bbernst@fas.harvard.edu (B.E.B.)

[6]These authors contributed equally to this work.

ylated regions overlaying multiple genes are evident at several Hox clusters but not in any of the other regions examined. These methylated Hox regions, which are lineage dependent, may represent active chromatin domains that function in the maintenance of Hox gene expression. Finally, comparisons of methylation patterns at orthologous loci in human and mouse indicate that methylation patterns are strongly conserved and thus likely to be functional. Nonetheless, we find that the overall degree of sequence conservation across the methylated sites is often no higher than the background rate. This result contrasts with the high degree of sequence conservation seen at other functional elements and suggests that comparative genomic analysis may not suffice to define the complete extent of such elements.

Results

Genomic Maps of Histone Modifications

Genomic maps of H3 Lys4 di- and trimethylation and H3 Lys9/14 acetylation were generated for human chromosomes 21 and 22 by ChIP and array hybridization. Crosslinked chromatin from HepG2 cells was immunoprecipitated with antibodies against histone H3 dimethylated at Lys4, trimethylated at Lys4, or acetylated at Lys9 and Lys14. Thus isolated DNA and nonenriched control DNA (whole cell extract) were amplified by in vitro transcription and hybridized separately to oligonucleotide arrays that tile nonrepetitive portions of chromosomes 21 and 22 at 35 bp intervals. Genomic sites enriched for a given modification were identified by sliding a 400 bp window across the interrogated regions and testing whether hybridization signals of oligonucleotide probes within the window were significantly greater for the ChIP sample. Overlapping windows meeting a given significance threshold were merged to form a single extended site. This approach is well suited for chromatin analysis, as it identifies genomic sites associated with modified histones in an unbiased manner and also defines their extents. We used a significance threshold of $p < 10^{-4}$ because this cutoff has a low false-positive rate (Figure 1; see below). We identified 430 dimethyl Lys4 sites, 504 trimethyl Lys4 sites, and 694 acetylated sites with median sizes of 605, 659, and 703 bp, respectively (Table 1). The modified sites cover approximately 1% of the nonrepetitive portions of these chromosomes (Figures 2A and 2B).

There is a significant correspondence between the dimethyl and trimethyl Lys4 sites, and 308 locations are enriched for both methylation states (Table 1). The correspondence between methylation and acetylation is also high, with more than 90% of Lys4-methylated sites also showing Lys9/14 acetylation. In contrast, when we used a general, modification-insensitive H3 antibody to generate an additional set of maps for chromosomes 21 and 22, we found just 11 significantly enriched sites. We note that the general and modification state-specific antibodies used in this study also recognize the histone variant H3.3, which differs from H3 by only four amino acids. It is likely that genomic sites enriched for Lys4 methylation and Lys9/14 acetylation are also enriched for H3.3, as a high proportion of this variant exhibits these modifications in vivo (McKittrick et al., 2004).

In addition to these chromosome-wide analyses, we mapped dimethyl Lys4 for the cytokine cluster, the IL4R region, and all four Hox loci in primary human fibroblasts. We identified a total of 116 dimethyl Lys4 sites at the 10^{-4} threshold that cover approximately 10% of these loci (Figures 2C and 2D). These regions were selected not only because they play important roles in immunol-

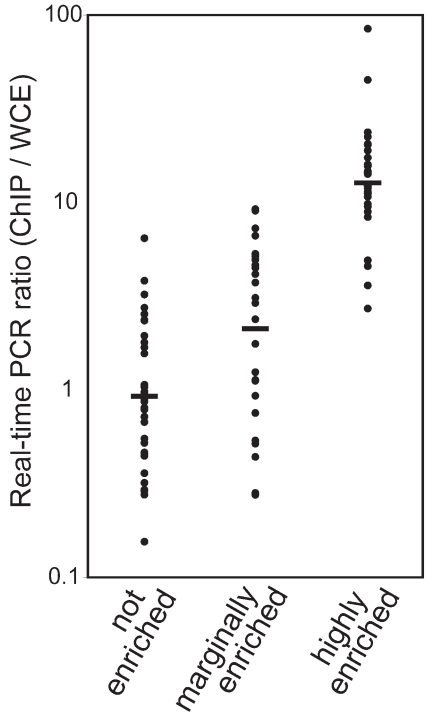

Figure 1. Enrichment of Selected Genomic Regions in a Dimethyl Lys4 ChIP

Real-time PCR enrichment ratios for selected genomic regions found to be not enriched ($p > 10^{-2}$), marginally enriched ($10^{-4} < p < 10^{-2}$), or highly enriched ($p < 10^{-4}$) in the array-based study are plotted in three aligned columns. Median values are indicated by bars.

Table 1. Genomic Sites on Chromosomes 21 and 22 Enriched for Modified Histones

Type	Description	Number of Sites	Median Size	Percent Coinciding with Gene Start
Me2	All dimethyl Lys4 sites (10^{-4} threshold)	430	605	46%
Me3	All trimethyl Lys4 sites (10^{-4} threshold)	504	659	66%
Me2+3	Trimethyl near dimethyl (10^{-4} threshold)	308	714	68%
Me3>>2	Trimethyl (10^{-4}) not near marginal dimethyl (10^{-2})	49	536	39%
Me2>>3	Dimethyl (10^{-4}) not near marginal trimethyl (10^{-2})	56	538	14%
Acetyl	All acetyl Lys9/14 sites (10^{-4} threshold)	694	703	58%

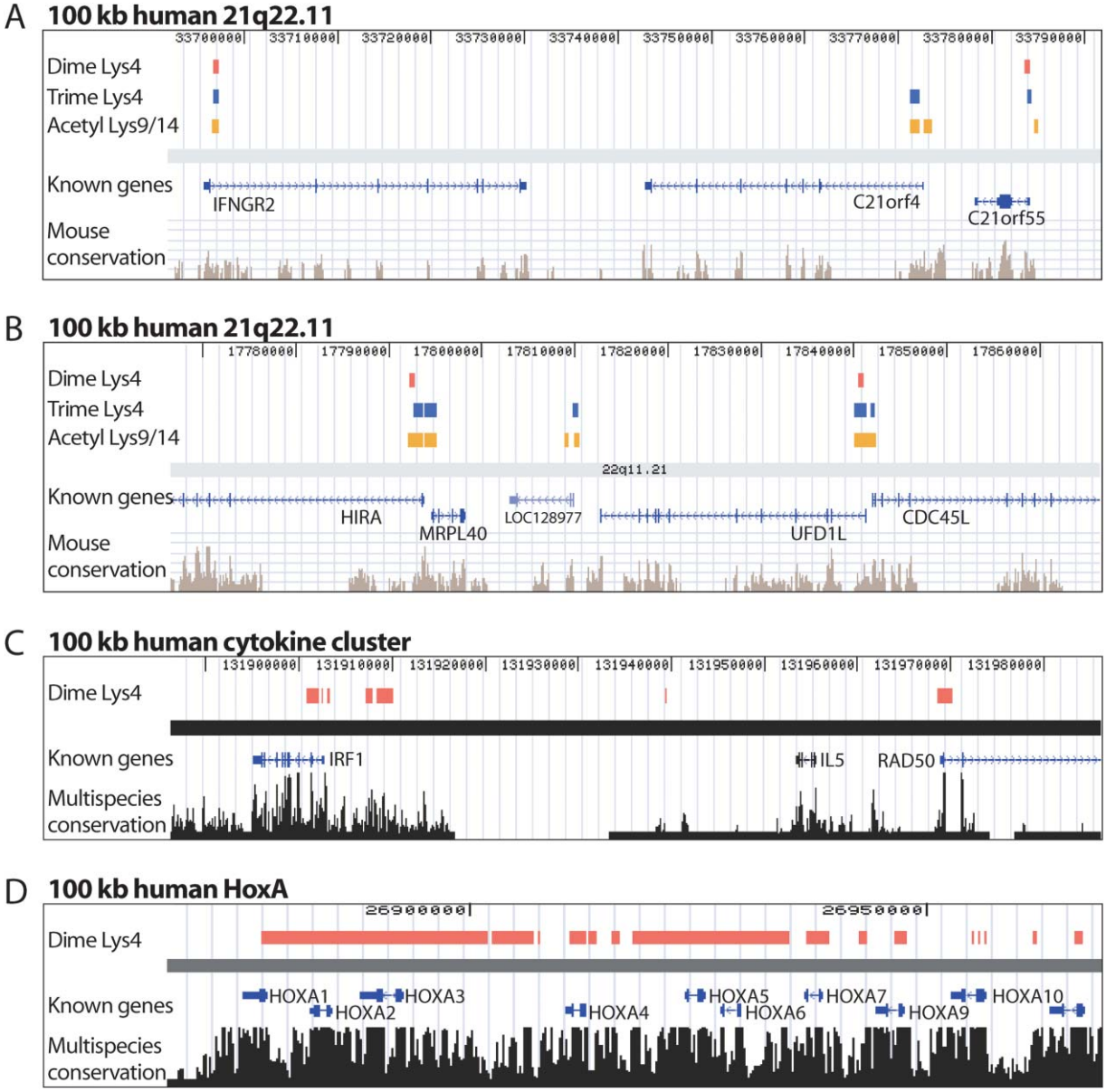

Figure 2. Representative Views of Equivalently Sized Genomic Regions

(A and B) 100 kb regions of human chromosomes 21 and 22. Dimethyl Lys4, trimethyl Lys4, and acetyl Lys9/14 sites meeting the high significance threshold (10^{-4}) are shown along with a mouse-human sequence conservation track (Kent et al., 2002; Waterston et al., 2002). (C and D) 100 kb regions of the human cytokine and HoxA clusters showing dimethyl Lys4 at the high significance threshold and a multispecies sequence conservation track (Blanchette et al., 2004).

ogy and development but also because they are clearly orthologous in human and mouse (i.e., have conserved gene orders). Species comparisons of histone methylation patterns in these regions should therefore provide general insight into the conservation properties of epigenetic marks. Thus, we also mapped dimethyl Lys4 for these loci in primary mouse fibroblasts, using analogous tiling arrays developed for mouse. The mouse maps show 97 dimethyl Lys4 sites at the 10^{-4} threshold that cover approximately 10% of these loci. All data are publicly available at http://www.broad.mit.edu/cell/chromatin_study.

Real-Time PCR Validation

We used a standard approach that incorporates ChIP and real-time PCR to assess the reliability of the 10^{-4} threshold and the accuracy of the array-based mapping technique in general. For this evaluation, we focused specifically on the maps of the human cytokine, IL4R, and Hox loci. We arbitrarily selected 30 highly enriched ($p < 10^{-4}$), 26 marginally enriched ($10^{-4} < p < 10^{-2}$), and 32 nonenriched ($p > 10^{-2}$) sites from the dimethyl Lys4 maps of these regions. Primer pairs that amplify sequences within each site were designed. Chromatin from primary human fibroblasts was immunoprecipi-

tated with antibody against dimethyl Lys4 using a standard ChIP protocol. Relative enrichment was quantified for each site in real-time PCR reactions using either 0.5 ng ChIP DNA or 0.5 ng nonenriched control DNA as template (Figure 1). Of sites identified as highly enriched by array, 93% were confirmed by real-time PCR to have enrichment ratios of at least 4-fold. Of sites identified as nonenriched, just 3% had ratios above 4-fold. Of sites identified as marginally enriched, 42% had ratios above 4-fold. Hence, the array-based approach reliably identifies highly enriched sites of methylation and rules out nonenriched sites. Furthermore, the intermediate distribution of marginally enriched sites suggests the maps contain quantitative information. Although the accuracy of the technique may partially depend on the genomic region examined, these results should be applicable also to the chromosome-wide data that were generated using similar methods. We focused further analysis on the highly enriched sites ($p < 10^{-4}$).

Trimethyl Lys4 Sites Correlate with 5′ Ends of Active Genes on Human Chromosomes 21 and 22

A systematic analysis of the locations of Lys4-methylated sites on chromosomes 21 and 22 revealed a striking correlation with the 5′ ends of annotated genes. Specifically, we examined regions residing within 1 kb of the transcription starts of currently known genes, as listed at the UCSC Human Genome Browser Gateway (Benson et al., 2004). We found that 66% of the trimethyl Lys4 sites and 46% of the dimethyl Lys4 sites overlap these regions (compared to 7% expected at random). The majority of methylated gene starts exhibit both methylation states. These data are consistent with recent studies that documented Lys4 di- and trimethylation at the 5′ ends of a limited number of genes in chicken and human cells (Liang et al., 2004; Schneider et al., 2004). Although these studies did not detect any difference in the distributions of di- and trimethylation, our analysis of a much larger number of genes suggests that trimethylation correlates better with gene starts. To investigate this further, we categorized methylated sites according to their degree of di- and trimethylation. We found that sites showing both methylation states and those showing mainly trimethylation strongly correlate with gene starts (68% and 39%, respectively). In contrast, sites showing mainly dimethylation do not correlate with gene starts much better than expected at random (14%). Hence, although di- and trimethyl Lys4 both correlate with gene starts, the trimethyl mark is more specific and appears to be the primary predictor of a coincident start site.

Still, a significant fraction of trimethyl Lys4 sites on chromosomes 21 and 22 do not coincide with the start of a known gene. To investigate the possibility that these additional sites coincide with genes not yet defined in the known gene set, we examined two other sets of gene predictions, made by Ensembl (Hubbard et al., 2002) and Acembly (http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly). We found that 56% of trimethylated sites that do not reside near the start of a known gene do reside within 1 kb of a transcription start of one of the gene predictions (compared to 22% expected at random). We also studied the proximity of methylated sites to potential transcriptional units, as identified by hybridizing polyA mRNA from HepG2 cells to the tiling arrays (Kampa et al., 2004; Kapranov et al., 2002). We found that 38% of trimethyl sites that do not reside near the start of a known gene do reside within 1 kb of one of these potential transcriptional units (compared to 22% expected at random). Together, these data suggest that transcription starts underlie most if not all trimethyl Lys4 sites on human chromosomes 21 and 22.

We next sought to correlate methylation status with transcriptional activity. We estimated transcription levels of known genes along chromosomes 21 and 22 using the mRNA maps generated for HepG2 cells. We found that genes with trimethylation at their 5′ end have significantly higher transcriptional rates than those without. In yeast, Lys4 methylation appears to be targeted to the 5' ends of active genes via an interaction between the Ser5-phosphorylated C-terminal domain (CTD) of RNA polymerase II and a methylase-containing complex (Krogan et al., 2003; Ng et al., 2003). We thus used ChIP and real-time PCR to test for enrichment of this RNA polymerase II isoform at the 5′ ends of human genes. We specifically examined three genes that show strong Lys4 methylation in HepG2 cells (*ETS2*, *CSTB*, and *GAPDH*). We found that the 5′ end of each gene is significantly enriched (6- to 8-fold) for Ser5-phosphorylated CTD. These data are consistent with the possibility that Lys4 methylation is targeted to the 5′ ends of human genes via a conserved mechanism involving this RNA polymerase II isoform.

Dimethyl Lys4 Sites in the Vicinity of Active Genes

The observation that sites showing mainly dimethylation do not correlate with 5′ ends of genes raised the possibility that these reflect other determinants. We therefore examined the set of preferentially dimethylated sites (Me2>>3 sites), which we defined as those showing dimethylation at the high significance threshold ($p < 10^{-4}$) but no trimethylation even at the marginal threshold ($p < 10^{-2}$). Of the 56 Me2>>3 sites on chromosomes 21 and 22, only eight are located within 1 kb of the start of a known gene. Nonetheless, nearly all of the remaining sites reside in the vicinity of known genes that are highly expressed in HepG2 cells (Figure 3).

We selected six Me2>>3 sites for further analysis. Five reside nearby or within introns of active genes, while the sixth is near several annotated ESTs. Using real-time PCR, we confirmed that each site is markedly enriched by a dimethyl Lys4 ChIP but far less enriched by a trimethyl Lys4 ChIP. We also used ChIP to test for enrichment of Ser5-phosphorylated CTD. In contrast to the trimethyl sites examined above, these Me2>>3 sites are not enriched for this RNA polymerase II isoform, suggesting they do not represent transcription starts. Although we cannot rule out the possibility that Me2>>3 sites reflect weak transcriptional activity, we note that these sites do not correlate with 5′ ends of known genes on chromosomes 21 and 22, many of which appear weakly active in HepG2.

We next examined the Me2>>3 sites in three additional cell types, including primary human fibroblasts, a melanoma cell line (A375), and an ovarian carcinoma line (NIH:OVCAR3). We found the methylation status of

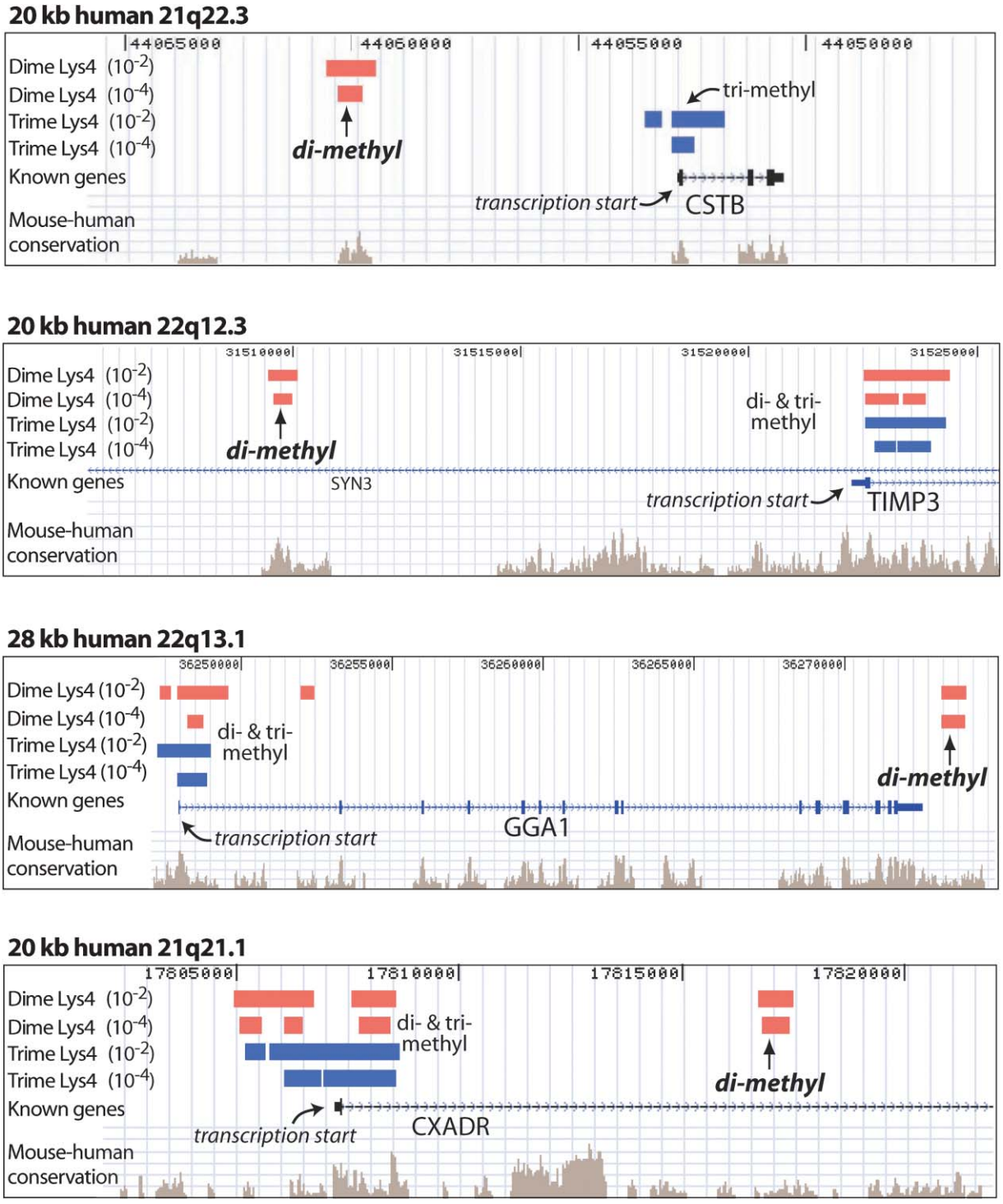

Figure 3. Genomic Views Showing Preferentially Dimethylated Sites on Chromosomes 21 and 22

Tracks for dimethyl Lys4 (red bars) and trimethyl Lys4 (blue bars) at the designated enrichment threshold of 10^{-4} and the more permissive 10^{-2} threshold are shown. Also shown is a mouse-human sequence conservation track (Kent et al., 2002; Waterston et al., 2002). Highlighted are preferentially dimethylated sites (Me2>>3) upstream of *CSTB* and *TIMP3*, downstream of *GGA1*, and within an intron of *CXADR*. The *CSTB* site is also 7 kb upstream of *NOP52*, the *TIMP3* site is in an intron of *SYN3*, and the *GGA1* site is 7 kb upstream of *SH3BP1*.

these sites to be highly dependent on cell type. Of the six sites examined, four are dimethylated in the fibroblasts, three are dimethylated in the melanoma cells, and two are dimethylated in the carcinoma cells. Together, these global and focused analyses suggest that preferentially dimethylated sites represent novel markers of cell state that occur within active chromosomal regions.

Our findings also raise the possibility that some

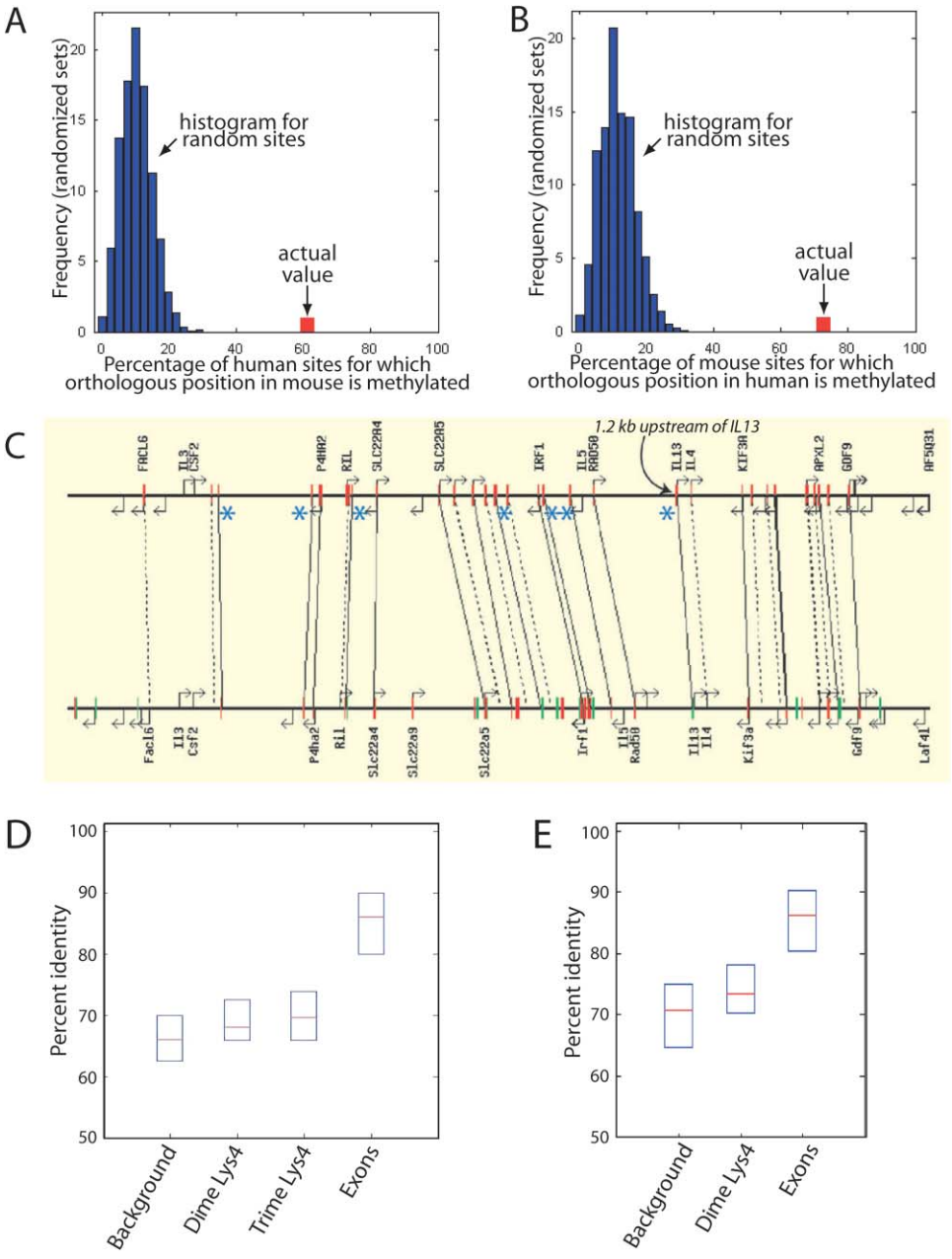

Figure 4. Conservation Properties of Lys4-Methylated Sites

(A and B) The extent to which the locations of Lys4-methylated sites are conserved in the orthologous human and mouse cytokine and IL4R regions is shown (red). The plots depict the percent of methylated sites in one organism for which the orthologous position in the other organism is methylated at the 10^{-2} threshold. As controls, histograms show the percent of sites in randomly generated sets for which the orthologous position in the other species is methylated at the 10^{-2} threshold (blue).

(C) Landscape plot showing correspondence between dimethyl Lys4 sites in the human cytokine cluster and sites in mouse at the 10^{-4} (red) and 10^{-2} (green) thresholds. Methylated sites that are not near gene starts are indicated by blue asterisks.

(D) Box plot showing 25th, 50th, and 75th percentile median percent sequence identities (mouse-human) for background regions, dimethyl and trimethyl Lys4 sites, and exons on chromosomes 21 and 22.

(E) Analogous plot for the cytokine and IL4R regions showing sequence identities for background regions, dimethyl Lys4 sites, and exons.

Me2>>3 sites might promote gene activity. We initially examined a site upstream of the *CSTB* gene (Figure 3). This site coincides with a highly conserved sequence element (in contrast to many Me2>>3 sites, see below). We cloned a 750 bp fragment containing this sequence element to test its regulatory potential. When placed upstream of a luciferase reporter, this DNA fragment leads to a 1.7-fold (±0.2) increase in gene expression. This result does not address the function of the coincident methylation, which would require stable incorpora-

tion or other chromatin-based assay. However, it suggests that some preferentially dimethylated sites may mark functional elements in the DNA.

H3 Acetylation Is Also Enriched at 5′ Ends of Genes and Strongly Correlates with Lys4 Methylation

We next examined the maps of H3 Lys9/14 acetylation on chromosomes 21 and 22 (Figure 2). Like the maps of H3 Lys4 methylation, these contain punctate sites around 1 kb in size that frequently coincide with transcription starts. The acetylated sites correlate well with di- and trimethylated sites but are particularly linked to the latter. Of the 504 trimethyl Lys4 sites, 495 coincide with acetylated sites. Like trimethylated sites, acetylated sites strongly correlate with 5′ gene ends, with 58% residing within 1 kb of the start of a known gene. This is consistent with a chromatin-scanning experiment that localized H3 acetylation to the 5′ ends of a limited number of human genes (Liang et al., 2004) and with the results of a genome-wide study in yeast (Roh et al., 2004). This systematic colocalization of methyl and acetyl marks could reflect recruitment of complexes containing Lys4 methylases and acetylases, such as TAC1 in flies (Smith et al., 2004) and/or crosstalk between adjacent modifications (Fischle et al., 2003). Active genes in flies are also systematically enriched for these marks and for Lys79 methylation (Schubeler et al., 2004). It will be important to investigate the extent to which the many other modifications on H3 and the other histones also correlate with these marks.

Relatively Uniform Histone Density across Chromosomes 21 and 22

A final set of maps for chromosomes 21 and 22 was generated by immunoprecipitating chromatin from HepG2 cells with a general histone H3 antibody. These were initially determined as controls to ensure that the previous maps reflect the modifications, rather than histone density or accessibility. The general H3 maps show only 11 significantly enriched sites across these chromosomes and therefore support the specificity of the modification maps.

However, prior studies have shown that promoters of active genes in yeast are depleted in a general H3 ChIP, as they tend to have low nucleosome occupancy (Bernstein et al., 2004; Lee et al., 2004). We therefore sought to determine whether promoters of active human genes are also depleted in a general H3 ChIP. We identified depleted sites by testing whether hybridization signals of oligonucleotide probes within a sliding window were significantly lower for the H3 ChIP sample than for the control. Although 67 sites on chromosomes 21 and 22 were depleted in the H3 ChIP at the 10^{-4} threshold, these do not correspond to gene promoters. We note that mammalian promoters are less constrained in size than those in yeast. Subtle nucleosome depletion may be distributed across these larger regions but remain undetected in our assay. Still, these data suggest that the marked nucleosome depletion evident in active promoters in yeast is not characteristic of mammalian chromatin.

Lys4-Methylated Sites Show Stronger Conservation of Location than Underlying Sequence

To gain further insight into the functional properties of genomic sites enriched for modified histones, we examined their conservation properties. For this analysis, we focused on the dimethyl Lys4 maps of the human and mouse cytokine and IL4R loci generated from primary lung fibroblasts of each species. We reasoned that *functional* sites of Lys4 methylation in the human genome should be conserved in mouse; i.e., the orthologous position in the mouse genome should also be enriched for Lys4 methylation. Like the chromosome-wide maps, maps of these loci show punctate methylated sites (49 in human, median size 626 bp; 41 in mouse, median size 675 bp), many of which reside near transcription starts. The locations of the sites are strongly conserved in human and mouse, with 55% of human sites showing methylation at the orthologous location in mouse and 68% of mouse sites showing methylation at the orthologous location in human (Figures 4A–4C). The proportion is slightly higher (61% and 73%) if a more permissive threshold ($p < 10^{-2}$) is allowed for defining methylated sites in the second species. Lys4-methylated sites near gene starts are conserved at roughly the same rate as sites far from gene starts. The correlations are highly significant: methylated sites in one species correspond to methylated sites in the other at a rate that is roughly 7-fold higher than expected at random. The correspondence strongly suggests that Lys4-methylated sites reflect conserved functions.

We next studied whether the genomic sequences at the methylated sites in human show an unusually high degree of evolutionary conservation in mouse. For each site, we determined whether an orthologous sequence could be identified in mouse, and, if so, we calculated the percentage of sequence identity. Orthologous sequence in mouse could be identified for 85% of the sites on human chromosomes 21 and 22 and for 96% of the sites in the human cytokine and IL4R regions. The distribution of sequence identity is shown in Figure 4. A large proportion of the sites (73% on chromosomes 21 and 22 and 74% in the two other regions) show no higher degree of conservation than seen in random background sequence (defined as within one standard deviation of the mean for such sequence). This is the case both for methylated sites that are near gene starts and for sites far from starts. Notably, *functional* conservation of methylation at orthologous human and mouse sites does not correlate with higher levels of overall *sequence* conservation. When comparing the cytokine and IL4R regions, human methylated sites that show methylation in mouse do not have significantly higher sequence identity than those that do not show methylation in mouse.

The relatively low degree of sequence conservation at many of the methylated sites has important implications. It contrasts with the fact that other functional elements show a high degree of sequence conservation across their entire lengths (which may be long, in the case of exons, or short, in the case of transcription factor binding sites). The fact that methylated sites show a low overall degree of sequence conservation suggests that

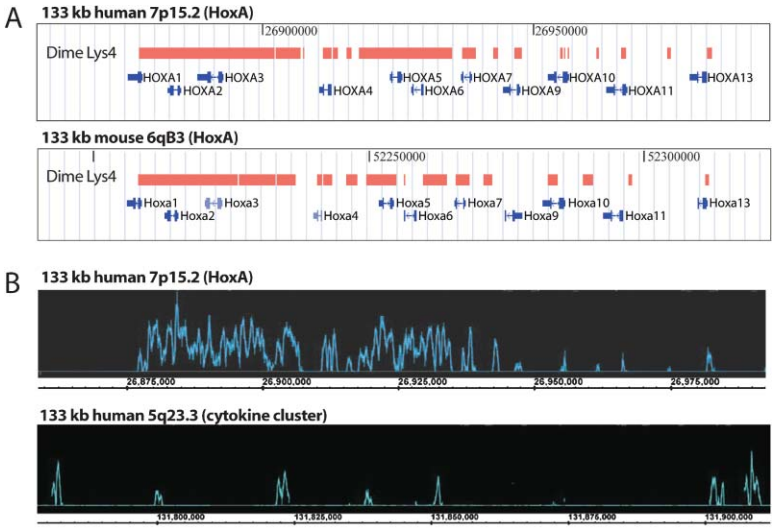

Figure 5. Broad Lys4-Methylated Regions in Hox

(A) Genomic views of human and mouse HoxA. Regions enriched for dimethyl Lys4 at the high-significance threshold are indicated (red bar).

(B) Dimethyl Lys4 signal intensity is shown for the same portion of human HoxA and for an equivalently sized portion of the cytokine cluster.

the critical functional elements in the DNA either constitute a small portion of the overall methylated region or lie at some distance from the region itself. Moreover, it indicates that crossspecies genome sequence analysis will likely not suffice to identify the complete extent of these methylated sites. Rather, comparative analysis of the methylated regions themselves in different species may help confirm these regions and define their functions.

Broad Methylated Regions Overlay Substantial Portions of the Human and Mouse Hox Clusters

Analysis of the remaining orthologous regions—the Hox clusters—revealed methylation patterns that are completely unique among the 39 MB of genome examined. Whereas Lys4-methylated sites in the other regions (chromosomes 21 and 22, cytokine cluster, IL4R region) are punctate, several of the Hox clusters contain huge methylated regions that encompass multiple genes. The contrast is readily seen by examining the N50 size of methylated sites (the size x such that 50% of the bases in methylated sites reside in sites of size $\geq$ x). Whereas the methylated sites in the human cytokine and IL4R regions have an N50 size of 1.3 kb, the sites in the human Hox loci have a remarkable N50 size of 17 kb. Detailed analysis of the map of human HoxA (generated for primary fibroblasts from human lung) reveals a nearly continuous Lys4-methylated region 60 kb in length that stretches from *HOXA1* to *HOXA7* (Figure 5). More than 85% of this region is enriched for dimethyl Lys4 at the 10^{-4} threshold.

Like the punctate methylated sites in the cytokine and IL4R loci, the broad Lys4-methylated regions in Hox appear to be evolutionarily conserved. When we examined the map of mouse HoxA (generated for primary fibroblasts from mouse lung), we found it contains a large, nearly continuous methylated region of comparable location and extent to that in human (Figure 5). Large methylated regions are also evident in the human and mouse HoxB clusters and in the mouse HoxC cluster. In each case, the methylation overlays the 3′ portions of the clusters that contain the lower numbered Hox genes.

The Hox loci are notable in that they exhibit a conserved gene order that is colinear with the time and space of expression during development (Kmita and Duboule, 2003). Chromatin is thought to play a prominent role in the maintenance of Hox gene expression (Francis and Kingston, 2001). Indeed, the Lys4-methylated regions we observe in human HoxA and HoxB appear to overlay active genes, as indicated by a previous transcriptional study that found 3′ HoxA and HoxB genes to be active in this fibroblast lineage (Chang et al., 2002). That study found that other fibroblast lineages have distinct patterns of Hox expression. For example, fibroblasts from gum, toe, and foreskin do not express 3′ HoxA and HoxB genes. We hypothesized that the Lys4 methylation overlaying these genes in the lung-derived cells would not be present in these other lineages. To test this, we examined a series of genic and intergenic locations within the 3′ portions of HoxA and HoxB by ChIP and real-time PCR. Whereas these sites show intense Lys4 methylation in the fibroblasts from lung (20- to 60-fold enrichment), they are not methylated in fibroblasts from gum, toe, or foreskin (Figure 6). We also examined an intergenic site located within the 5′ portion of HoxA. This site is strongly methylated in fibroblasts from toe and foreskin, which both express 5′ HoxA genes, but not in the lung or gum cells, which do not express these genes (Chang et al., 2002). Hence, broad Lys4-methylated regions in Hox are lineage dependent and appear to correlate with expression of underlying genes.

Since Lys4 methylation in Hox overlays genic and intergenic regions, we considered the possibility that robust transcription also occurs outside of the genes. We examined the transcriptional status of the entire HoxA and HoxB loci by hybridizing RNA from the lung fibroblasts to the tiling oligonucleotide arrays. Transcription was detected at 33% of intergenic bases within methylated portions of the human HoxA and HoxB regions, whereas the proportion in nonmethylated areas was only $\sim$10%. Similar enrichments of intergenic transcription were evident in methylated portions of the mouse HoxA and HoxB loci. In contrast, intergenic tran-

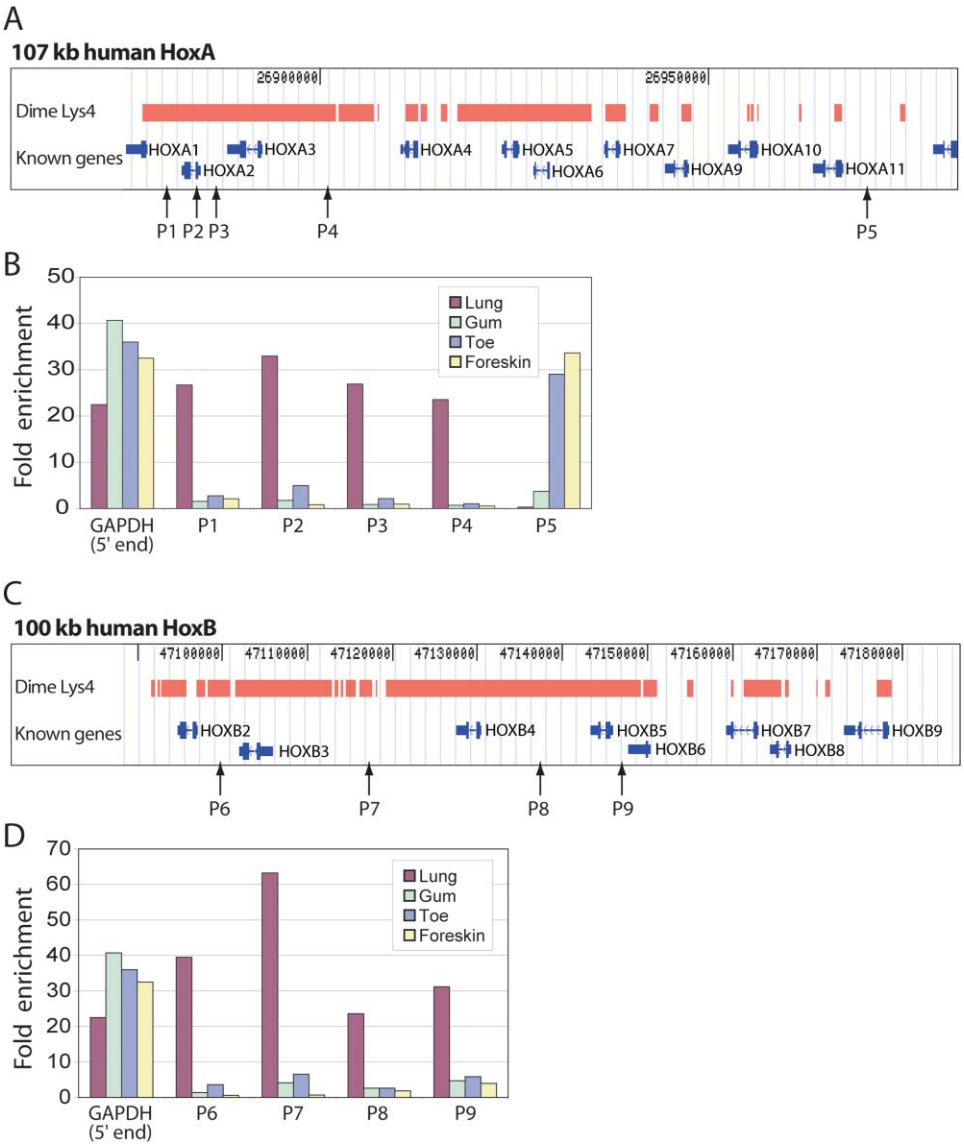

Figure 6. Hox Methylation in Different Fibroblast Lineages

(A) Genomic view of human HoxA. Regions enriched for dimethyl Lys4 in fibroblasts from lung are shown (red bar). Genomic sites examined by ChIP and real-time PCR in four fibroblast lineages are indicated (arrows).

(B) Real-time PCR data showing enrichment of indicated HoxA sites in dimethyl Lys4 ChIP assays carried out in the specified fibroblast lineages (data for GAPDH shown as a positive control).

(C and D) Genomic view and real-time PCR data for human HoxB.

scription levels in the cytokine and IL4R loci were comparable to those in unmethylated portions of Hox. These findings suggest that uniquely high levels of intergenic transcription occur in the Lys4-methylated portions of Hox. Indeed, transcription in regulatory portions of the *Drosophila* bithorax complex has been observed previously (Bae et al., 2002; Rank et al., 2002). We speculate that intensely methylated, highly transcribed regions in Hox represent active chromatin domains required for the maintenance of Hox gene expression.

Discussion

We have determined high-resolution maps of histone H3 Lys4 di- and trimethylation and H3 Lys9/14 acetylation for two human chromosomes, and maps of Lys4 dimethylation for several orthologous human and mouse loci, including the cytokine cluster, the IL4R region, and all four Hox clusters. These maps, generated by combining ChIP, linear DNA amplification, and tiling oligonucleotide arrays, comprise a total of 39 MB of genome. The result is an unprecedented long-range view, three orders of magnitude greater in scale than prior studies, that reveals a series of insights into the structure and function of vertebrate chromatin.

Lys4 Methylation at Transcription Starts and Elsewhere in the Vicinity of Active Genes

Prior studies have documented H3 Lys4 methylation at the 5′ ends of a few dozen genes in higher eukaryotes

(Liang et al., 2004; Schneider et al., 2004). Here, we greatly expand this association by showing that the 5′ ends of more than 300 human genes, approximately 30% of known genes on chromosomes 21 and 22, are highly enriched for Lys4-methylated histones. The trimethylated state appears to be the more specific predictor of a coincident gene start, and genes that show trimethylation at their 5′ ends are significantly more active than genes that do not.

Lys4-methylated histones are also associated with actively transcribed regions in yeast (Bernstein et al., 2002; Ng et al., 2003; Santos-Rosa et al., 2002). In that model organism, an interaction between the Ser5-phosphorylated RNA polymerase II CTD and a methylase-containing complex is thought to target methylation to the 5' ends of active genes (Krogan et al., 2003; Ng et al., 2003). An analogous mechanism may underlie the establishment of Lys4 methylation in mammals, as we find this RNA polymerase II isoform is enriched at methylated starts of active genes in human cells. Although its functions remain poorly understood, Lys4 methylation may facilitate elongation and mRNA processing (Hampsey and Reinberg, 2003), possibly via recruitment of nucleosome remodeling complexes (Santos-Rosa et al., 2003).

In contrast to genomic sites showing Lys4 trimethylation with or without dimethylation, sites showing mainly dimethylation tend not to coincide with gene starts. Nonetheless, the vast majority reside nearby or within introns of highly expressed genes. These sites do not show enrichment of Ser5-phosphorylated CTD or other evidence of transcription. Their methylation status is highly dependent on cell type, suggesting that preferentially dimethylated sites may represent novel markers of cell state that correlate with active chromosomal regions. Some dimethyl Lys4 sites coincide with highly conserved sequences and may mark functional elements in the DNA. However, many do not show conservation of underlying sequence much above random. These may reflect lateral propagation or a higher-order chromatin structure that makes associated histones accessible to modifying enzymes or places them in close three-dimensional proximity to the 5′ ends of nearby genes.

The generality of these findings, derived through comprehensive analysis of the nonrepetitive portions of chromosomes 21 and 22, is an important question. Although the interrogated regions constitute approximately 2% of the human genome, they are not known to contain large gene clusters and are not representative of heterochromatin. Analysis of a series of gene clusters suggests that the punctate distribution of methylated sites observed throughout chromosomes 21 and 22 is extensible to some (cytokine, IL4R) but clearly not all (Hox) loci. Finally, it should be noted that the chromosome-wide maps were generated for the HepG2 cell line. We found similar methylation patterns for the cytokine and IL4R loci in primary fibroblasts and confirmed several methylated sites on chromosomes 21 and 22 in these cells. However, some aspects of chromatin in the neoplastic cells may not be relevant to normal cells.

Methylated Sites Are Functionally Conserved Even When Underlying Sequence Is Not Highly Conserved

We systematically examined the conservation properties of methylated sites. In the regions that were compared between the human and mouse genomes, we observed a strong correlation between the presence of methylation in one genome and the presence of methylation at the orthologous site in the other genome. The correlation was observed for sites near or far from gene starts. This functional conservation supports the biological importance of these sites.

Despite this observation, the methylated sites across chromosomes 21 and 22 as well as the cytokine and IL4R regions often show sequence conservation that is not significantly higher than the background rate for orthologous sites. These results suggest that the critical functional elements in the DNA responsible for the methylated regions either represent only a small fraction of the region or else lie at some distance from the site; they may act by lateral propagation or by dictating higher-ordered chromatin structure. In any case, this suggests that comparative genomic analysis alone will not suffice to identify the precise locations and boundaries of the methylated regions. Moreover, crossspecies comparisons of methylated regions may help to validate the regions and understand their functions.

Distinctive Chromatin Structures in Human and Mouse Hox

The final set of orthologous loci that we examined were the Hox clusters. These loci are remarkable in that the Hox genes are arranged in a conserved order that is colinear with time and space of expression (Kmita and Duboule, 2003). Chromatin structure in Hox is of great interest and is thought to play a critical role in maintaining gene expression patterns (Francis and Kingston, 2001).

We observed completely unique methylation patterns in Hox, characterized by broad Lys4-methylated regions encompassing multiple genes (Figure 5). The locations and extents of these regions are largely conserved in the analogous human and mouse cells examined (primary fibroblasts from lung). In the human cells, the methylation occurs throughout the 3′ portions of HoxA and HoxB and thus overlays genes known to be active in this cell type (Chang et al., 2002). These regions do not show methylation in other fibroblast lineages (gum, toe, and foreskin) that do not express 3′ HoxA and HoxB genes. Chambeyron and Bickmore (2004) recently described a visible decondensation of chromatin at the HoxB locus during embryonic stem cell differentiation that was accompanied by increased acetylation and Lys4 methylation of the *HoxB1* and *HoxB9* genes. We suggest that the extensive methylated regions we observe in primary fibroblasts reflect active chromatin domains established during differentiation in a lineage-dependent manner to maintain Hox expression.

How these chromatin domains influence Hox expression remains unclear. Lys4 methylation is generally associated with active regions and may maintain chromatin accessibility by recruiting chromatin remodeling

enzymes (Santos-Rosa et al., 2003), promoting acetylation of adjacent lysines and/or preventing the establishment of repressive modifications (Fischle et al., 2003). The methylated domains also show uniquely high levels of intergenic transcription. Elongation promotes chromatin opening (Sims et al., 2004), and nongenic transcripts appear to enhance the accessibility of the β-globin and antigen receptor loci (Bolland et al., 2004; Gribnau et al., 2000). Importantly, Lys4-methylated domains in Hox could reflect regions where the trithorax group protein MLL functions. MLL catalyzes Lys4 methylation and is required for the expression of individual Hox genes (Milne et al., 2002; Nakamura et al., 2002; Yu et al., 1995). The broad distribution of Lys4 methylation discovered here raises the possibility that MLL functions globally in Hox clusters to activate and/or maintain the expression of multiple adjacent genes. Further studies are needed to define what roles each of these features play in determining Hox chromatin structure and how they may contribute to the uniquely colinear regulation of the Hox genes.

Conclusion

With the ability to perform large-scale analysis of chromatin structure, it should become possible to gain systematic insight into the overall structure and function of chromatin. The present study of more than 37 MB of human genome, as well as 2 MB of orthologous mouse genome, reveals diverse findings, including punctate methylated sites at transcription starts and elsewhere in the vicinity of active genes, large methylated domains likely to play seminal roles in Hox regulation, and functional conservation of many sites across species. Further large-scale studies across the human and mouse genomes will be needed to evaluate the generality of the patterns observed here and identify other general features in mammalian chromatin.

Experimental Procedures

Chromatin Immunoprecipitation

Primary human (CRL-1485) and mouse (CCL-206) fibroblasts from lung, human hepatocellular carcinoma cells (HepG2), human melanoma cells (A-375), and human ovarian carcinoma cells (NIH:OVCAR3) were obtained from the American Type Culture Collection. Primary human fibroblasts from gum (AG9319) and toe (AG9309) were obtained from Coriell Cell Repositories, and human foreskin fibroblasts (hTERT-BJ1) were obtained from Promega. Chromatin was immunoprecipitated using a variation of the protocol described at http://www.upstate.com. Approximately 10^8 cells grown to 80% confluence in DMEM with 10% FBS and 1% sodium pyruvate were fixed with 1% formaldehyde, washed with ice-cold PBS, harvested by scraping, pelleted, and resuspended in SDS lysis buffer (see Supplemental Data at http://www.cell.com/cgi/content/120/2/169/DC1/). Samples were sonicated with a Branson 250 Sonifier for 8 × 25 s cycles at 70% duty, output 3.5, centrifuged at 13,000 × g for 10 min, and diluted 10-fold in ChIP dilution buffer. After removing a control aliquot (whole-cell extract), sample was incubated at 4°C overnight with antibodies against dimethyl H3 Lys4 (Upstate), trimethyl H3 Lys4 (Abcam), acetyl H3 Lys9/14 (Upstate), the H3 C terminus (Abcam), or Ser5-phosphorylated RNA polymerase II (Covance). Complexes were precipitated with protein A-Sepharose or anti-mouse IgM-agarose beads. Beads were washed sequentially with low-salt immune complex wash, high-salt immune complex wash, LiCl immune complex wash, and TE (Supplemental Data). Immunoprecipitated chromatin was eluted in elution buffer, incubated at 65°C for 8 hr, and treated with proteinase K. DNA was purified by extracting twice with phenol and once with chloroform and precipitating in ethanol. The DNA was treated with RNase and purified with a MinElute Kit (Qiagen).

DNA Amplification and Array Hybridization

ChIP (enriched) and control (whole-cell extract) DNA samples were amplified by in vitro transcription as described (Liu et al., 2003). Amplified RNA was purified using an RNeasy Mini Kit (Qiagen), converted into single-stranded cDNA with random primers, fragmented with DNase I, and end labeled with biotin as described (Kapranov et al., 2002). ChIP and control samples (~2 mg) were applied to separate oligonucleotide arrays. Arrays were hybridized 16–18 hr at 45°C, washed using the antibody amplification protocol described in the Affymetrix Expression Analysis Technical Manual, and scanned using an Affymetrix Gene Array scanner. For each mapping experiment, two independent biological replicates were hybridized in triplicate.

Analysis of ChIP Tiling Array Data

Raw array data were quantile normalized within enriched/control replicate groups (Bolstad et al., 2003) and scaled to a median feature intensity of 525. (PM, MM) intensity pairs were mapped to the genome using exact 25-mer matching to hs.NCBIv33 (chromosomes 21 and 22), hs.NCBIv34 (human loci), or mm3 (mouse loci). For each genomic position to which a probe pair mapped, a data set was generated consisting of all (PM, MM) pairs mapping within a window of ±200 bp. A Wilcoxon Rank Sum test (Hollander and Wolfe, 1999) was applied to the transformation log2(max[PM-MM,1]) for data from the six treatment (ChIP samples) and six control DNA (whole-cell extract) arrays within the local data set, testing the null hypothesis that treatment and control data come from the same probability distribution against the alternative that the probability distribution of the treatment is shifted such that it is greater than that of the control. The Wilcoxon test was applied in a sliding window across genomic coordinates as described (Cawley et al., 2004). Genomic positions belonging to enriched replicated regions were defined by applying a p value cutoff of 10^{-4}, and resultant positions separated by <200 bp were merged to form a predicted dimethyl Lys4, trimethyl Lys4, or acetyl Lys9/14 enriched region. An identical method was used to identify regions enriched by the general H3 antibody. Regions depleted in the general H3 ChIP were identified by reversing treatment and control data. Genomic alignments, sequence conservation analysis, and randomization trials are described in the Supplemental Data on the *Cell* web site.

Collection and Analysis of RNA Tiling Array Data

Cytosolic RNA was isolated from HepG2, CRL-1485, and CCL-206 cells grown under the conditions described as described (Kapranov et al., 2002). Biological replicates were hybridized in triplicate to arrays tiling human chromosomes 21 and 22 (HepG2), the human orthologous loci (CRL-1485), or the mouse orthologous loci (CCL-206). Intensity data were quantile normalized (Bolstad et al., 2003), and the expression level was estimated for every mapped probe position by calculating the median of all pairwise average values of PM − MM (Hollander and Wolfe, 1999) of probes that mapped to positions within a sliding window of ±50 bp. The human and mouse data for the orthologous loci were normalized to a median intensity of 550.

Real-Time PCR

PCR primer pairs were designed to amplify 150–200 bp fragments from selected genomic regions (Supplemental Data). Real-time PCR reactions were carried out using Quantitect SYBR green PCR mix (Qiagen) in an MJ Research Opticon Instrument according to manufacturer's instructions. Fold enrichment was determined in real-time PCR reactions using either 0.5 ng ChIP DNA or 0.5 ng unenriched DNA as template. Each enrichment ratio was determined from two independent ChIP assays evaluated in duplicate by real-time PCR.

Luciferase Reporter Assays

The 750 bp genomic region that coincides with the Me2>>3 site upstream of *CSTB* and a control region that includes 750 bp DNA adjacent to this site were cloned upstream of the promoters in the pGL3 promoter luciferase construct (Supplemental Data). Cotransfections of reporter plasmids and a Renilla vector were carried out in HepG2 cells using Lipofectamine 2000 (Invitrogen). Luciferase and Renilla signals were quantified after 48 hr using the Dual-Luciferase Assay Reporter Kit (Promega). Relative reporter expression was quantified by dividing luciferase by Renilla signal and averaging across eight replicates. The control region did not significantly alter luciferase expression.

Acknowledgments

We are grateful for insightful comments and suggestions from Duncan Odom, Rick Young, Kevin Struhl, Philipp Kapranov, Jay Bradner, Mike Zody, Pablo Tamayo, Pratima Nemani, and members of the Bauer Center for Genomic Research. B.E.B. is supported by a K08 Development Award from the NCI. S.L.S. is an investigator in the Howard Hughes Medical Institute. This work was supported in part by funds from the NIGMS (R01 GM038627); the NCI (N01-C0-12400); and Affymetrix, Inc.

Received: October 7, 2004
Revised: December 10, 2004
Accepted: January 4, 2005
Published: January 27, 2005

References

Bae, E., Calhoun, V.C., Levine, M., Lewis, E.B., and Drewell, R.A. (2002). Characterization of the intergenic RNA profile at abdominal-A and Abdominal-B in the *Drosophila* bithorax complex. Proc. Natl. Acad. Sci. USA *99*, 16847–16852.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Wheeler, D.L. (2004). GenBank: update. Nucleic Acids Res. *32*, D23–26.

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

Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O., and Schreiber, S.L. (2004). Global nucleosome occupancy in yeast. Genome Biol. *5*, R62. Published online August 20, 2004. 10.1186/gb-2004-5-9-r62

Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., et al. (2004). Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. *14*, 708–715.

Bolland, D.J., Wood, A.L., Johnston, C.M., Bunting, S.F., Morgan, G., Chakalova, L., Fraser, P.J., and Corcoran, A.E. (2004). Antisense intergenic transcription in V(D)J recombination. Nat. Immunol. *5*, 630–637.

Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185–193.

Bulger, M., Schubeler, D., Bender, M.A., Hamilton, J., Farrell, C.M., Hardison, R.C., and Groudine, M. (2003). A complex chromatin landscape revealed by patterns of nuclease sensitivity and histone modification within the mouse beta-globin locus. Mol. Cell. Biol. *23*, 5234–5244.

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., et al. (2004). Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499–509.

Chambeyron, S., and Bickmore, W.A. (2004). Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev. *18*, 1119–1130.

Chang, H.Y., Chi, J.T., Dudoit, S., Bondre, C., van de Rijn, M., Botstein, D., and Brown, P.O. (2002). Diversity, topographic differentiation, and positional memory in human fibroblasts. Proc. Natl. Acad. Sci. USA *99*, 12877–12882.

Feinberg, A.P., and Tycko, B. (2004). The history of cancer epigenetics. Nat. Rev. Cancer *4*, 143–153.

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

Francis, N.J., and Kingston, R.E. (2001). Mechanisms of transcriptional memory. Nat. Rev. Mol. Cell Biol. *2*, 409–421.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. (2004). Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. Cell *118*, 555–566.

Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R., and Fraser, P. (2000). Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. Mol. Cell *5*, 377–386.

Hampsey, M., and Reinberg, D. (2003). Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation. Cell *113*, 429–432.

Hollander, M., and Wolfe, D.A. (1999). Nonparametric Statistical Methods, Second Edition (New York: John Wiley and Sons, Inc.).

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., Cox, T., Cuff, J., Curwen, V., Down, T., et al. (2002). The Ensembl genome database project. Nucleic Acids Res. *30*, 38–41.

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

Kampa, D., Cheng, J., Kapranov, P., Yamanaka, M., Brubaker, S., Cawley, S., Drenkow, J., Piccolboni, A., Bekiranov, S., Helt, G., et al. (2004). Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. Genome Res. *14*, 331–342.

Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P., and Gingeras, T.R. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. Science *296*, 916–919.

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

Kmita, M., and Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. Science *301*, 331–333.

Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M., et al. (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol. Cell *11*, 721–729.

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

Liang, G., Lin, J.C., Wei, V., Yoo, C., Cheng, J.C., Nguyen, C.T., Weisenberger, D.J., Egger, G., Takai, D., Gonzales, F.A., and Jones, P.A. (2004). Distinct localization of histone H3 acetylation and H3–K4 methylation to the transcription start sites in the human genome. Proc. Natl. Acad. Sci. USA *101*, 7357–7362.

Litt, M.D., Simpson, M., Gaszner, M., Allis, C.D., and Felsenfeld, G. (2001). Correlation between histone lysine methylation and developmental changes at the chicken beta-globin locus. Science *293*, 2453–2455.

Liu, C.L., Schreiber, S.L., and Bernstein, B.E. (2003). Development and validation of a T7 based linear amplification for genomic DNA. BMC Genomics *4*, 19.

McKittrick, E., Gafken, P.R., Ahmad, K., and Henikoff, S. (2004). Histone H3.3 is enriched in covalent modifications associated with active chromatin. Proc. Natl. Acad. Sci. USA *101*, 1525–1530.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D., and Hess, J.L. (2002). MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol. Cell *10*, 1107–1117.

Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C.M., and Canaani, E.

(2002). ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol. Cell *10*, 1119–1128.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol. Cell *11*, 709–719.

Rank, G., Prestel, M., and Paro, R. (2002). Transcription through intergenic chromosomal memory elements of the *Drosophila* bithorax complex correlates with an epigenetic switch. Mol. Cell. Biol. *22*, 8026–8034.

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

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.

Santos-Rosa, H., Schneider, R., Bernstein, B.E., Karabetsou, N., Morillon, A., Weise, C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2003). Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. Mol. Cell *12*, 1325–1332.

Schneider, R., Bannister, A.J., Myers, F.A., Thorne, A.W., Crane-Robinson, C., and Kouzarides, T. (2004). Histone H3 lysine 4 methylation patterns in higher eukaryotic genes. Nat. Cell Biol. *6*, 73–77.

Schreiber, S.L., and Bernstein, B.E. (2002). Signaling network model of chromatin. Cell *111*, 771–778.

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.

Sims, R.J., 3rd, Belotserkovskaya, R., and Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. Genes Dev. *18*, 2437–2468.

Smith, S.T., Petruk, S., Sedkov, Y., Cho, E., Tillib, S., Canaani, E., and Mazo, A. (2004). Modulation of heat shock gene expression by the TAC1 chromatin-modifying complex. Nat. Cell Biol. *6*, 162–167.

Turner, B.M. (2002). Cellular memory and the histone code. Cell *111*, 285–291.

Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J.F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. Nature *420*, 520–562.

Wolffe, A.P., and Matzke, M.A. (1999). Epigenetics: regulation through repression. Science *286*, 481–486.

Yu, B.D., Hess, J.L., Horning, S.E., Brown, G.A., and Korsmeyer, S.J. (1995). Altered Hox expression and segmental identity in Mll-mutant mice. Nature *378*, 505–508.