Cell

A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells

Bradley E. Bernstein,[1,2,3,*] Tarjei S. Mikkelsen,[3,4] Xiaohui Xie,[3] Michael Kamal,[3] Dana J. Huebert,[1] James Cuff,[3] Ben Fry,[3] Alex Meissner,[5] Marius Wernig,[5] Kathrin Plath,[5] Rudolf Jaenisch,[5] Alexandre Wagschal,[6] Robert Feil,[6] Stuart L. Schreiber,[3,7] and Eric S. Lander[3,5]

[1] Molecular Pathology Unit and Center for Cancer Research, Massachusetts General Hospital, Charlestown, MA 02129, USA
[2] Department of Pathology, Harvard Medical School, Boston, MA 02115, USA
[3] Broad Institute of Harvard and MIT, Cambridge, MA 02139, USA
[4] Division of Health Sciences and Technology, MIT, Cambridge, MA 02139, USA
[5] Whitehead Institute for Biomedical Research, MIT, Cambridge, MA 02139, USA
[6] Institute of Molecular Genetics, CNRS UMR-5535 and University of Montpellier-II, Montpellier, France
[7] Howard Hughes Medical Institute at the Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA
*Contact: bbernstein@partners.org
DOI 10.1016/j.cell.2006.02.041

SUMMARY

The most highly conserved noncoding elements (HCNEs) in mammalian genomes cluster within regions enriched for genes encoding developmentally important transcription factors (TFs). This suggests that HCNE-rich regions may contain key regulatory controls involved in development. We explored this by examining histone methylation in mouse embryonic stem (ES) cells across 56 large HCNE-rich loci. We identified a specific modification pattern, termed "bivalent domains," consisting of large regions of H3 lysine 27 methylation harboring smaller regions of H3 lysine 4 methylation. Bivalent domains tend to coincide with TF genes expressed at low levels. We propose that bivalent domains silence developmental genes in ES cells while keeping them poised for activation. We also found striking correspondences between genome sequence and histone methylation in ES cells, which become notably weaker in differentiated cells. These results highlight the importance of DNA sequence in defining the initial epigenetic landscape and suggest a novel chromatin-based mechanism for maintaining pluripotency.

INTRODUCTION

Epigenetic regulation of gene expression is mediated in part by posttranslational modifications of histone proteins, which in turn modulate chromatin structure (Jenuwein and Allis, 2001; Margueron et al., 2005). The core histones H2A, H2B, H3, and H4 are subject to dozens of different modifications, including acetylation, methylation, and phosphorylation. Histone H3 lysine 4 (Lys4) and lysine 27 (Lys27) methylation are of particular interest as these modifications are catalyzed, respectively, by trithorax- and Polycomb-group proteins, which mediate mitotic inheritance of lineage-specific gene expression programs and have key developmental functions (Ringrose and Paro, 2004). Lys4 methylation positively regulates transcription by recruiting nucleosome remodeling enzymes and histone acetylases (Santos-Rosa et al., 2003; Pray-Grant et al., 2005; Sims et al., 2005; Wysocka et al., 2005), while Lys27 methylation negatively regulates transcription by promoting a compact chromatin structure (Francis et al., 2004; Ringrose et al., 2004).

Various observations suggest that chromatin undergoes important alterations during mammalian development (Delaval and Feil, 2004; Margueron et al., 2005; Sado and Ferguson-Smith, 2005). Embryonic stem (ES) cell differentiation is accompanied by changes in chromatin accessibility at several key developmental genes, including a large-scale opening of the HoxB locus (Chambeyron and Bickmore, 2004; Perry et al., 2004). Furthermore, Polycomb-group proteins play an essential role in maintaining the pluripotent state of ES cells and show markedly reduced expression upon differentiation (O'Carroll et al., 2001; Silva et al., 2003; Valk-Lingbeek et al., 2004). However, little is known about the overall structure of ES cell chromatin, how it is established, or how it contributes to the maintenance of pluripotency (Szutorisz and Dillon, 2005).

Large-scale studies of mammalian chromatin have recently become possible with the combination of

chromatin immunoprecipitation (ChIP) and DNA microarrays. Initial studies in primary fibroblasts revealed thousands of genomic sites associated with Lys4 methylation (Bernstein et al., 2005; Kim et al., 2005). The vast majority show a "punctate" pattern, typically occurring at sites of ~1-2 kb near promoters of active genes. Lys27 methylation is implicated in X chromosome inactivation and imprinting (Plath et al., 2003; Umlauf et al., 2004). However, little is known about the overall genomic distribution of this repressive mark. Gene-specific and limited microarray studies have reported that Lys27 methylation tends to occur at punctate sites near promoters of repressed genes (Cao and Zhang, 2004; Kimura et al., 2004; Kirmizis et al., 2004; Koyanagi et al., 2005). Based on such studies, the distributions of Lys4 and Lys27 methylation have been thought to be nonoverlapping.

A notable exception to the punctate pattern of histone modifications is evident at the Hox gene clusters: These loci contain large, cell type-specific Lys4 methylated regions, up to 60 kb in length, that overlay multiple Hox genes (Bernstein et al., 2005; Guenther et al., 2005). These regions likely reflect accessible chromatin domains established during embryonic development to maintain Hox gene expression programs (Chambeyron and Bickmore, 2004). However, the extent to which large domains of chromatin modifications represent a general feature of mammalian genomes remains unclear.

Recent studies have revealed that the most highly conserved noncoding elements (HCNEs) in mammalian genomes cluster within ~200 HCNE-rich genomic loci, which include all four Hox clusters (Nobrega et al., 2003; Bejerano et al., 2004; Lindblad-Toh et al., 2005; Woolfe et al., 2005). These regions tend to be gene-poor but are highly enriched for genes encoding transcription factors (TFs) implicated in embryonic development.

These findings suggest that the HCNE-rich regions or the TF genes within them may contain key epigenetic regulatory controls involved in development. We explored this by mapping histone methylation patterns in mouse ES cells across 61 large regions (~2.5% of the genome). The results reveal a novel chromatin modification pattern that we term "bivalent domains," consisting of large regions of Lys27 methylation harboring smaller regions of Lys4 methylation. In ES cells, bivalent domains frequently overlay developmental TF genes expressed at very low levels. Bivalent domains tend to resolve during ES cell differentiation and, in differentiated cells, developmental genes are typically marked by broad regions selectively enriched for either Lys27 or Lys4 methylation. We suggest that bivalent domains silence developmental genes in ES cells while keeping them poised for activation. Finally, we analyzed the relationship between histone methylation and the underlying DNA sequence in both ES and differentiated cells. This analysis suggests that DNA sequence largely defines the initial epigenetic state in ES cells, which is subsequently altered upon differentiation, presumably in response to lineage-specific gene expression programs and environmental cues.

RESULTS

Bivalent Domains in ES Cells Contain Repressive and Activating Histone Modifications

Histone H3 Lys4 and Lys27 methylation patterns in ES cells were examined across a subset of HCNE-rich loci using a combination of ChIP and tiling oligonucleotide arrays. The arrays tile 61 large genomic regions, totaling 60.3 Mb, at a density of approximately one probe per 30 bases (Table S1). The regions consist of the four Hox clusters (1.3 Mb encoding 43 genes), 52 additional HCNE-rich regions (55 Mb encoding 169 genes), and five "control" regions that do not show high HCNE density (4 Mb encoding 95 genes). We isolated genomic DNA associated with either trimethylated Lys4 or trimethylated Lys27 by immunoprecipitating cross-linked chromatin, and we then hybridized these DNA fractions to the tiling arrays (see Experimental Procedures). We identified regions of Lys4 or Lys27 methylation by comparing these hybridization results to those obtained for total genomic DNA. Experiments were performed in duplicate and analyzed using previously validated criteria (Bernstein et al., 2005). The resulting maps of ES cell chromatin (Figures 1 and S1) show a number of important features, many of which were unexpected.

We found a total of 343 sites of Lys4 methylation, ranging in size from 1 kb to 14 kb with a median size of 3.4 kb (Table S2). Of these, 63% correspond to transcription start sites (TSSs) of known genes. Conversely, 80% of the TSSs are covered by Lys4 sites. Because Lys4 sites and TSSs each cover only ~2% of the genomic regions, this concordance is highly significant. We and others have previously noted a global concordance between Lys4-methylated sites and TSSs in differentiated mammalian cells (Bernstein et al., 2005; Kim et al., 2005).

We also found 192 Lys27-methylated sites across these regions. These tend to affect much larger genomic regions than the Lys4 sites. The median Lys27 site is smaller in the control regions (5 kb) but twice as large in the HCNE-rich regions (10 kb) and still larger in the Hox regions (18 kb). Overall, 75% of Lys27 sites are larger than 5 kb. We will refer to these large regions as "Lys27 domains." There are 123 in the HCNE regions, 14 in the Hox regions, and 7 in the control regions (Table S2).

Comparison of the two datasets revealed many instances of a previously undescribed pattern of chromatin modifications: Three-quarters of the Lys27 domains contain Lys4 sites within them. These regions thus harbor both a "repressive" and an "activating" chromatin modification; we therefore termed them "bivalent domains." There are 95 in the HCNE regions, 9 in the Hox regions, and 5 in the control regions (Table S2).

Bivalent Domains Overlay Developmentally Important TF Genes in HCNE-Rich Regions

Roughly three-quarters of the bivalent domains in the HCNE regions (69/95) overlap TSSs of known genes, with the Lys4 sites typically positioned directly at the

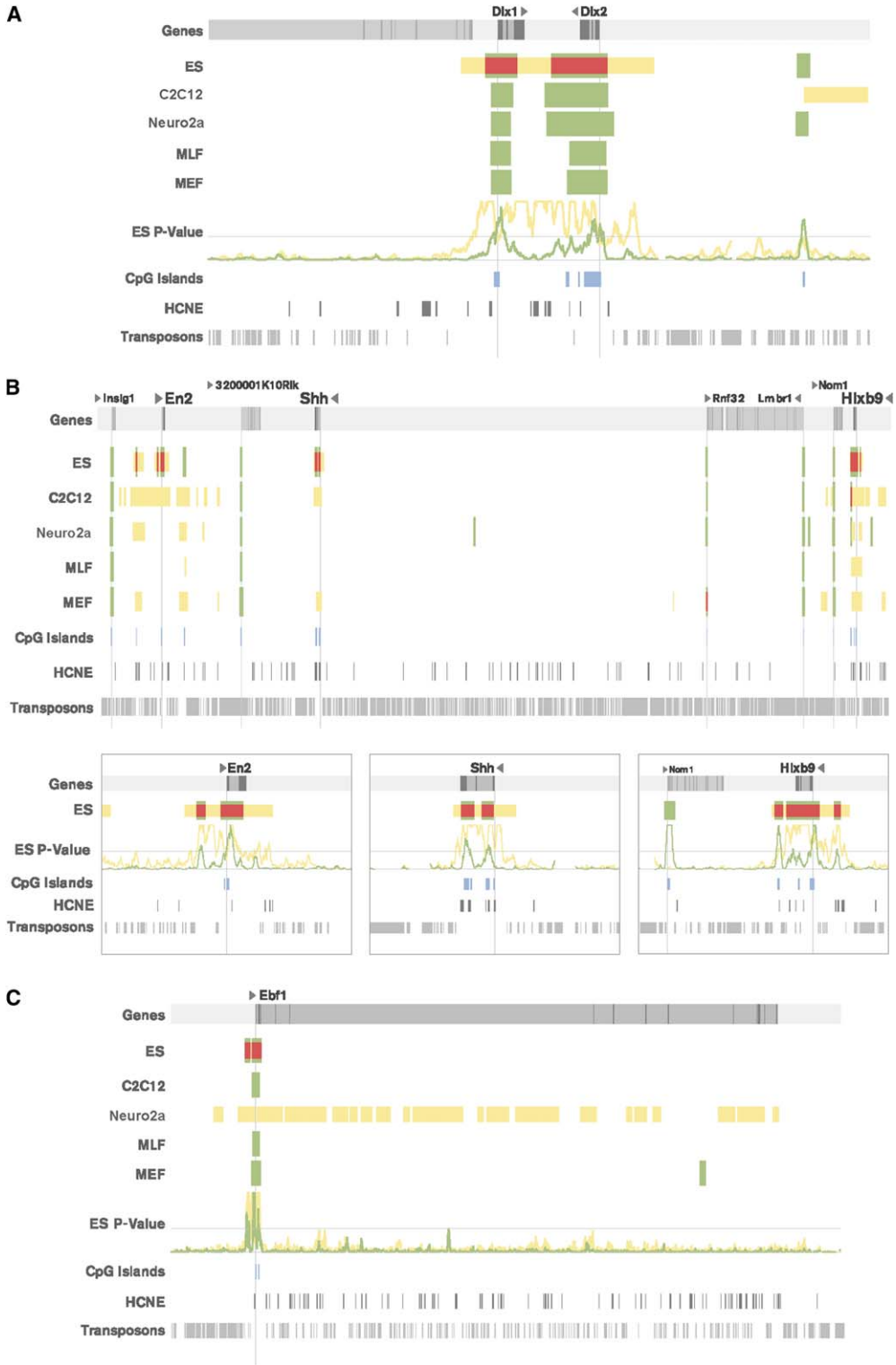

Figure 1. Representative Views of Histone Methylation Patterns across HCNE-Rich Regions in ES and Differentiated Cells

(A) Dlx1-Dlx2 gene cluster (Region 47, 112 kb). For each cell type, tracks show regions associated with Lys27 methylation (yellow), Lys4 methylation (green), or both modifications (red). For ES cells only, the raw p-value signals for Lys27 (yellow) and Lys4 methylation (green) are also shown. Genes (TSSs indicated by long vertical lines; exons indicated as dark), CpG islands, HCNEs (Lindblad-Toh et al., 2005), and transposable elements are also shown.

(B) En2, Shh, Hlxb9 (Region 48, 1.5 MB). Expanded views show 75 kb around each gene.

(C) Ebf1 (Region 31, 497 kb). Note expansive Lys27 methylated region in the Neuro2a cells.

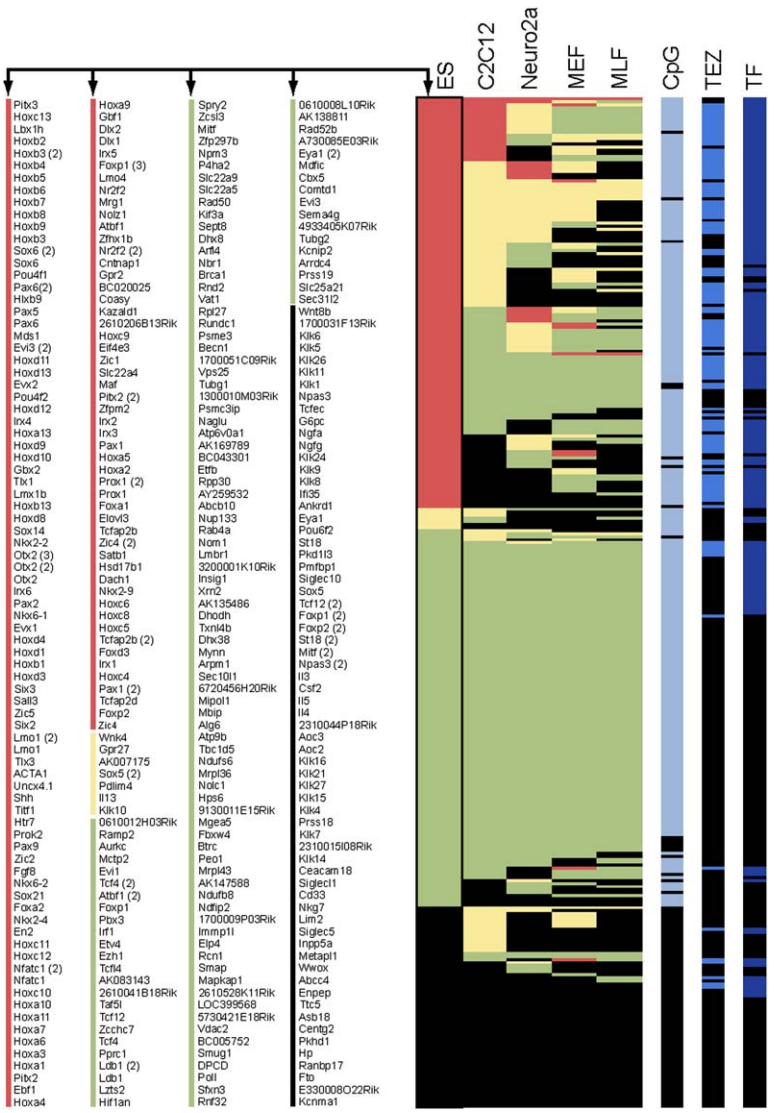

Figure 2. Histone Methylation Status of Transcription Start Sites

Methylation status is shown for the 332 known TSSs in the 61 examined regions in ES cells, C2C12 myoblasts, Neuro2a neuroblastoma cells, mouse embryonic fibroblasts (MEF), and mouse lung fibroblasts (MLF). Red indicates the presence of a bivalent domain; yellow indicates Lys27 methylation only; green indicates Lys4 methylation only; and black indicates no detected methylation. Blue rows in the three rightmost columns indicate TSSs that correspond to TF genes, contain CpG islands, or coincide with transposon-exclusion zones (TEZs). This figure and the corresponding Table S3 show the strong correlation among bivalent domains, TF genes, CpG islands, and TEZs.

TSS. Of these, a full 93% (64/69) occur at genes that encode TFs, including Sox, Fox, Pax, Irx, and Pou gene family members, even though TF genes make up only half of the genes in the regions examined (Figure 2). The 26 bivalent domains that do not occur at known TSSs are also of interest: Four occur at the 3′-ends of developmental genes (Npas3, Meis2, Pax2, and Wnt8b), and ten occur in locations that show strong evidence of encoding transcripts (including the presence of mRNA transcripts, CpG islands, and high levels of sequence conservation). Among the nonTF genes associated with bivalent domains are genes implicated in neural development, such as Fgf8 and Prok1. In the Hox regions, the observed bivalent domains are especially large and overlap multiple TSSs of known genes, all of which are TFs. The five bivalent domains in the control regions are quite short; they all overlap gene starts, although these genes do not encode TFs.

The chromatin analysis thus reveals that ES cells contain many bivalent domains. In HCNE-rich regions, these domains are particularly large and highly enriched at developmentally important genes that establish cell identity. The bivalent nature of this novel epigenetic pattern raises the possibility that the associated genes are poised in a bipotential state, which may be resolved differently in different cell lineages. This hypothesis predicts that differentiated cells would contain few, if any, bivalent domains.

In Differentiated Cells, TF Genes Are Marked by Either Repressive or Activating Modifications

We next examined Lys4 and Lys27 methylation patterns across these same regions in a collection of differentiated cell types, including mouse embryonic fibroblasts (MEFs), mouse primary lung fibroblasts (MLFs), C2C12 myoblasts, and Neuro2a neuroblastoma cells. We identified multiple

Lys4 and Lys27 methylated sites in each cell type, many of which are large. However, in marked contrast to the ES cell data, we found few bivalent domains in the differentiated cells (6 in MEFs, 1 in MLFs, 13 in myoblasts, and 12 in the neuroblastoma cells).

Thus, the majority of TSSs that show bivalent domains in ES cells do not show bivalent domains in the differentiated cells. The vast majority of these (93/97) instead show either a Lys27- or a Lys4-methylated site in at least one of the differentiated cell types (Figure 2 and Table S3). These "monovalent" sites tend to be large, with median sizes of 19.4 kb and 7.4 kb for Lys27 and Lys4, respectively (compared to 6.7 kb and 3.4 kb over all sites in the differentiated cells). Thus, bivalent domains appear largely specific to ES cells and, in differentiated cells, developmental genes are instead frequently organized within expansive regions showing either repressive or activating modifications.

Bivalent Modification Patterns in ES Cells Confirmed by Alternate Techniques

Given the novel nature of the bivalent domains, we sought to confirm our results using completely different reagents and protocols (see Experimental Procedures). Specifically, we used an independent source of ES cells with a different genotype; we refer to the first source as ES1 and the second as ES2. We also used an alternative ChIP procedure carried out on micrococcal nuclease-digested nucleosomes that had not been subjected to cross-linking and performed the immunoprecipitation with antisera from different sources. This alternative ChIP technique controls for nucleosome occupancy and is not subject to potential artifacts of cross-linking and sonication (O'Neill and Turner, 2003). The ES2 data also show a large number of bivalent domains, and these correspond closely to those seen in the ES1 data. Importantly, 94 of the 95 bivalent domains in the ES2 data correspond to bivalent domains in the ES1 cells.

We next sought to test whether the observed bivalent domain structure truly reflects the simultaneous presence of both Lys4 and Lys27 methylation on the same physical chromosomes. It is formally possible that the bivalent domains could instead reflect the presence of either two subpopulations with distinct character or one population alternating between two states. To rule out this possibility, we carried out a sequential ChIP in which ES cell chromatin was immunoprecipitated first with Lys27 tri-methyl antibody and second with Lys4 tri-methyl antibody. This sequential purification is designed to retain only chromatin that concomitantly carries both kinds of modifications. Using real-time PCR, we tested three TSSs associated with bivalent domains (Irx2, Dlx1, and Hlxb9). Each was significantly enriched relative to the controls (genes enriched for only Lys27 or only Lys4) (see Figure 3 and Experimental Procedures). For example, Irx2 is enriched ~10-fold in the primary (Lys27) ChIP and further enriched >30-fold (relative to control) in the secondary (Lys4) ChIP. This shows that a large proportion of Irx2 chromatin that contains Lys27 methylation also contains Lys4 methylation—at least 30-fold more than the control. (Of course, the technique cannot prove that 100% of all Irx2 species in ES cells carry both modifications because of inherent limitations due to background). We also tested the Irx2 TSS by repeating the sequential ChIP with the order of the immunoprecipitations reversed and again found significant enrichment (see Experimental Procedures). Together, the experiments above suggest that the bivalent domains accurately represent the epigenetic state at many TF genes in ES cells.

Bivalent Domains Are Associated with Low Levels of Gene Expression

To gain insight into the functional significance of bivalent domains, we examined gene expression patterns across the three cell types with at least ten bivalent domains (ES cells, C2C12, and Neuro2a) (Mogass et al., 2004; Tomczak et al., 2004; Perez-Iratxeta et al., 2005). Within each cell type, we found that genes marked by Lys4 methylation tend to be expressed at significantly higher levels than those associated with Lys27 methylation (Figure 4). A good example is the Ebf1 gene, which encodes a TF implicated in multiple differentiation pathways: In MEFs, MLFs, and myoblasts, it is expressed at relatively high levels (Schraets et al., 2003; Koli et al., 2004) and is associated with relatively large Lys4 sites (>5 kb), while in neuroblastoma cells it is expressed at an essentially undetectable level and is associated with an expansive Lys27 domain.

We next examined the expression levels of genes marked by bivalent domains. These show low levels of expression, with the overall distribution being similar to that for genes marked by Lys27 methylation alone (Figure 4). Thus, the presence of Lys4 methylation at a TSS is typically associated with high gene activity when it occurs in the absence of Lys27 methylation but with low gene activity when it occurs together with Lys27 methylation (that is, the repressive effect of Lys27 appears to be epistatic to the activating effect of Lys4 methylation in a bivalent domain). These results raise the possibility that bivalent domains function to silence developmental genes in ES cells while keeping them poised for induction upon initiation of specific developmental pathways.

Resolution of Bivalent Domains during ES Cell Differentiation

Our analysis of ES cells and four differentiated cell types suggests that bivalent domains are characteristic of pluripotent cells, that they silence developmental genes while keeping them poised, and that they tend to resolve upon ES cell differentiation into Lys4 or Lys27 methylation, in accordance with associated changes in gene expression. We sought to study whether the resolution of bivalent domains can be observed soon after ES cell differentiation by examining a differentiated cell type obtained directly from ES cells. Specifically, we differentiated ES cells along a neural pathway in serum-free culture and generated a homogenous population of multipotent neural precursor cells maintained in FGF2- and EGF-containing media, as

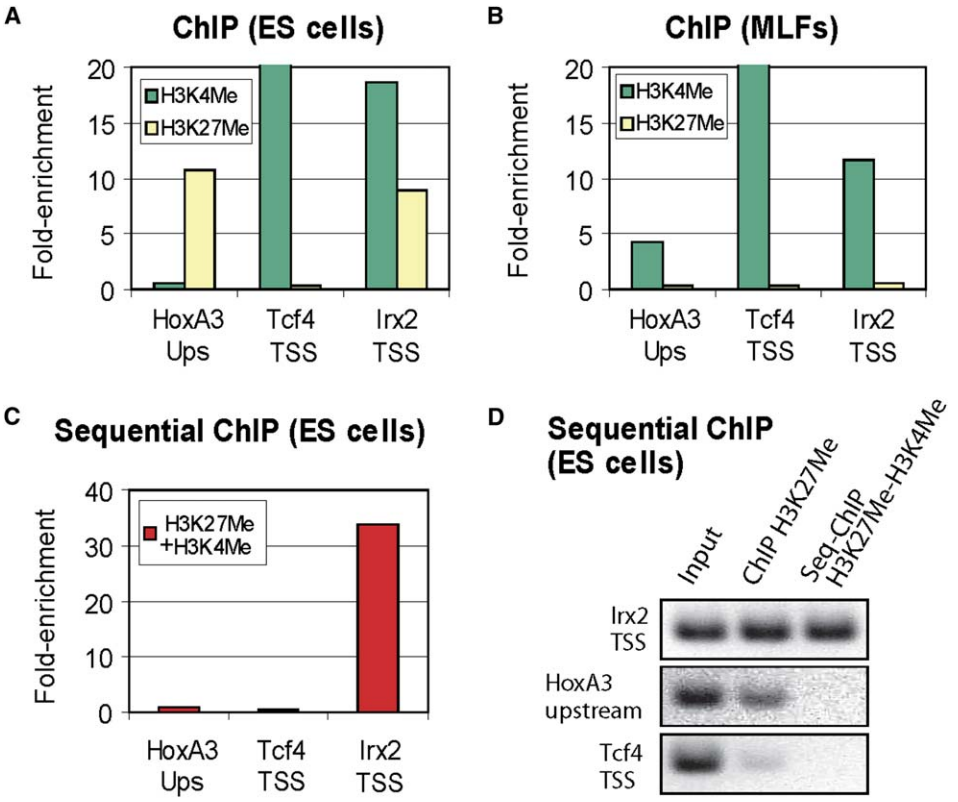

Figure 3. Characterization of the Irx2 Bivalent Domain

ChIP and sequential ChIP were used to examine the methylation status of the Irx2 TSS (bivalent), the Tcf4 TSS (Lys4 only), and a site upstream of HoxA3 (Lys27 only).

(A) Real-time PCR ratios reflect the enrichment of indicated sites when ES cells are subjected to ChIP with trimethyl Lys4 antibody or trimethyl Lys27 antibody.

(B) Corresponding data for mouse lung fibroblasts.

(C) Real-time PCR ratios reflect the relative enrichment of indicated sites after sequential immunoprecipitations with trimethyl Lys27 antibody and then trimethyl Lys4 antibody (see Experimental Procedures).

(D) PCR products amplified from control (input) DNA, Lys27 ChIP DNA, and Lys27-Lys4 sequential ChIP DNA are shown; the same samples were used as real-time PCR template in (C). The sequential ChIP results indicate that, in ES cells, the Irx2 TSS is associated with chromatin marked by both Lys27 and Lys4 methylation.

described previously (Conti et al., 2005). We focused on seven genes associated with bivalent domains in ES cells (Figure 5). These include three genes that are markedly induced during differentiation (Nkx2.2, Sox21, and Zfpm2), one that is weakly induced (Dlx1), and three that are not induced (Pax5, Lbx1h, and Evx1). Using ChIP and real-time PCR, we first confirmed that the TSS of each gene is indeed associated with both Lys4 and Lys27 methylation in the original ES cells, and we then examined the methylation status of these genes in the neural precursor cells. For the three genes whose expression is markedly induced, the TSS becomes specifically associated with Lys4 methylation in these differentiated cells. For the three genes that are not induced, the TSS becomes specifically associated with Lys27 methylation. Interestingly, the TSS of the weakly induced gene, Dlx1, remains associated with both methylation marks in the neural precursor cells, although the Lys4 signal is significantly stronger. These data support a model in which bivalent domains are largely specific to ES cells and tend to resolve upon ES cell differentiation according to pathway-specific gene expression programs.

Epigenetic Modifications in ES Cells Strongly Correlate with Underlying DNA Sequence

Because the Lys4- and Lys27-methylated sites in ES cells seem to represent important initial conditions for development, we searched for DNA sequence features that might underlie or predict the establishment of these epigenetic marks across the genome.

We found a strong positive correlation between the presence of Lys4 methylation in ES cells and the density of CpG dinucleotides in the underlying DNA sequence (median 8% versus 2% expected) (Figure S2). Strikingly, 95% of TSSs with Lys4 sites have CpG islands, and 91% of TSSs with CpG islands also have Lys4 sites

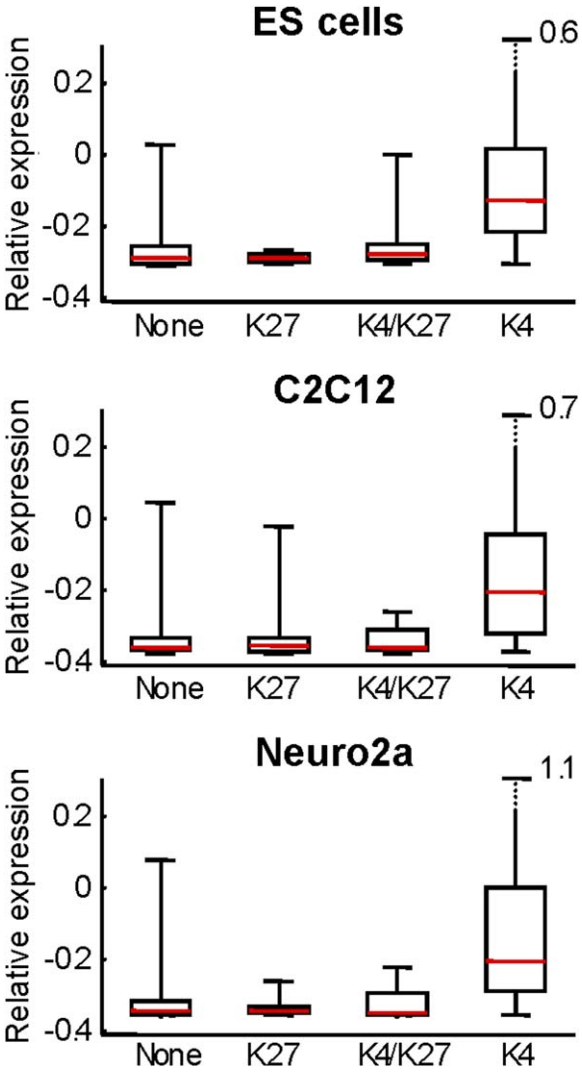

Figure 4. Gene Expression as a Function of Histone Methylation Status

Box plot showing 25th, 50th, and 75th percentile expression levels in ES cells, myoblasts, and neuroblastoma cells for genes associated with no histone methylation, Lys27 methylation, bivalent domains, or Lys4 methylation. Whiskers show 2.5th and 97.5th percentiles. Expression data (*y* axis) were determined from published expression profiles (Mogass et al., 2004; Tomczak et al., 2004; Perez-Iratxeta et al., 2005), uniformly normalized to a mean of 0 and a standard deviation of 1 for all probes on each array.

(r_{phi} = 0.73) (see Table S4 and Experimental Procedures). Moreover, the lengths of the two features are significantly correlated where they overlap (r = 0.50). By contrast, the correlation is weaker in the differentiated cells; this is primarily due to loss of Lys4 methylation at 20%–35% of CpG islands (r_{phi} = 0.40 for MLFs) (Table S4). We note that a recent genome-wide study (Roh et al., 2005) of histone H3 acetylation in T cells observed a correlation with CpG islands at a similar level to that seen in the differentiated cells examined here.

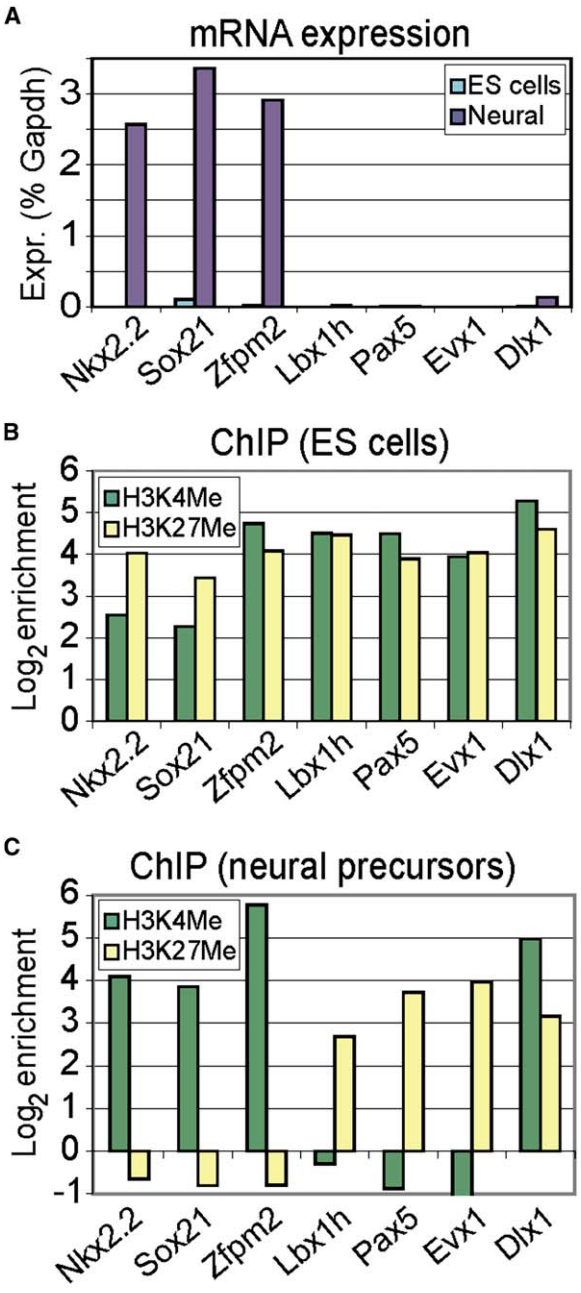

Figure 5. Resolution of Bivalent Domains during ES Cell Differentiation

ES cells were differentiated along a neural pathway in serum-free culture, and a homogenous population of multipotent neural precursor cells were maintained in FGF2- and EGF-containing media, as described in Conti et al., 2005. Several loci showing bivalent domains in ES cells were examined in the differentiated cells.

(A) Expression levels (relative to Gapdh) were determined by RT-PCR for the indicated genes in ES cells and in neural precursors. The methlyation states of the indicated genes were determined by ChIP and real-time PCR in ES cells (B) and in neural precursors (C). The data suggest that bivalent domains tend to resolve during ES cell differentiation in accordance with associated changes in gene expression.

We also found that Lys27-methylated regions in ES cells show a strikingly low density of transposon-derived sequence (median 6% versus 22% expected) (Figures S2 and S3). The most extreme example is found at the Hox clusters, which are known to have the lowest density of transposon-derived sequence in the mouse and human genomes (Lander et al., 2001; Waterston et al., 2002) and which have the largest Lys27 domains (up to 141 kb) in our sample. Most of the Lys27 domains contain long stretches (>10 kb) with little or no identifiable transposon-derived sequence. We defined such regions as "transposon exclusion zones" (TEZs) (see Experimental Procedures). Within the loci examined here, 89% of TSSs with a TEZ have a Lys27 domain in ES cells, and 73% of TSSs with a Lys27 domain have a TEZ (r_{phi} = 0.69) (Table S4). The lengths of these two features are significantly correlated where they overlap (r = 0.78). Interestingly, we note that the small number of Lys27 domains found only in differentiated cells does not appear to overlap particularly transposon-poor sequence (Figure 1C and Table S4).

We tested if the TEZs represent conserved genomic features by examining the orthologous sequence in the human and dog genome. The frequency of lineage-specific repeats provides an independent test of whether transposons are tolerated in these regions. The TEZs show a clear deficit of lineage-specific repeats in both human (1.3% versus 15.2% expected) and dog (1.0% versus 9.1% expected), confirming that this property is strongly conserved across mammals.

We then searched for TEZs across the entire mouse genome. We identified 710 TEZs, of which 328 overlap TSSs of known genes (Table S5). Strikingly, the vast majority of these genes encode developmental and tissue-specific TFs (189), proteins involved in axon guidance and neuronal function (65), and other cell signaling-related proteins such as growth factors (25), including Fgf8, Fgf10, Fgf14, and the imprinted gene Igf2. Notably, they include ~70% of the developmental regulators previously identified within 204 HCNE-rich loci (Lindblad-Toh et al., 2005). We predict that most of these genes will harbor Lys27 domains or bivalent domains in ES cells.

Colocalization of Bivalent Domains with Oct4 and Nanog

Finally, we examined the relationship of bivalent domains to the reported binding sites of certain pluripotent TFs. A recent genomic analysis in human ES cells found that Oct4, Nanog, and Sox2 are frequently associated with developmentally important genes (Boyer et al., 2005). We mapped the Oct4, Nanog, and Sox2 binding sites reported in that study to orthologous positions in the mouse genome and examined their overlap with bivalent domains. About 50% of bivalent domains coincide with binding sites of at least one of the pluripotent TFs, a highly significant correspondence ($p < 10^{-9}$). The correlation is primarily due to Oct4 and Nanog and actually becomes more significant when Sox2 is removed from the analysis. Interestingly, although many of the genes targeted by these pluripotent factors are actively expressed in ES cells, those that are also associated with a bivalent domain tend to be silenced ($p < 5 \times 10^{-3}$). This suggests that the bivalent domains may override any activation potential these TFs might have but also raises the possibility that the pluripotent TFs may help keep these genes in a poised state. Notably, a full 50% of bivalent domains are not associated with any of the three pluripotent TFs. It will be interesting to see if these coincide with binding sites of other important TFs.

DISCUSSION

Our results shed light on chromatin structure in ES cells and raise intriguing hypotheses about its establishment and function during development. The bivalent domains reported here have many notable features: They combine both "repressive" and "activating" modifications; they are highly enriched in ES cells relative to differentiated cells; and they are associated with genes encoding TFs with roles in embryonic development and lineage specification. In differentiated cells, these TF genes instead tend to be associated with large regions carrying either an activating or a repressive methylation mark. We propose that bivalent domains silence developmental genes in ES cells while preserving their potential to become activated upon initiation of specific differentiation programs. Bivalent domains may be related to a phenomenon observed at the bithorax complex in early fly development, where silenced Polycomb response elements are nonetheless associated with trithorax-group proteins and low-level transcription. Remarkably, both of these activities appear to be required for subsequent gene activation during development (Orlando et al., 1998; Schmitt et al., 2005). By analogy, Lys4 methylation within bivalent domains and associated trithorax activities may keep silenced developmental genes poised in ES cells. Our analyses of differentiated cells suggest that bivalent domains largely resolve during differentiation into large regions of either Lys27 or Lys4 methylation. These modified regions may provide a robust epigenetic memory to maintain lineage-specific expression or repression of these critical genes. Their large size would ensure that each daughter chromosome would likely inherit a substantial proportion of the modified histones, which could then promote similar modification of new histones in the immediate vicinity (Henikoff et al., 2004; van Steensel, 2005).

A fundamental issue that remains is to understand the mechanism by which the initial conditions are established in ES cells. The analysis here suggests that some of the answer can be read directly from the genome sequence. The strong association of Lys4 methylation with CpG islands may well be directly causal, inasmuch as the trithorax complexes that methylate Lys4 are reported to associate with CpG-rich DNA (Ayton et al., 2004; Lee and Skalnik, 2005). The strong association of Lys27 methylation with transposon-exclusion zones

may instead reflect strong evolutionary pressure against the presence of transposon-derived sequence in these regions. Repetitive sequences are subject to repressive epigenetic modifications (Arnaud et al., 2000; Lippman et al., 2004; Martens et al., 2005), which might interfere with the function of the bivalent domains and thus be eliminated by selection. It has been reported previously that imprinted loci, while significantly depleted for short interspersed transposable elements (SINEs), are permissive to L1 long interspersed transposable elements (LINEs) (Greally, 2002). In contrast, we find that both classes of transposons tend to be excluded from regions associated with Lys27 methylation or bivalent domains in ES cells. The direct signal for Lys27 methylation remains unclear (although we cannot exclude the possibility that the deficit of transposon-related chromatin modifications in some fashion promotes the association of the Polycomb complex that methylates Lys27). The correlations between the histone modifications and the genomic features are notably weaker in differentiated cells (Bernstein et al., 2005). We suggest that while the embryonic state may be largely defined by DNA sequence, it is subsequently altered in response to lineage-specific transcriptional programs and environmental cues and epigenetically maintained.

Our study was motivated by the suspicion that HCNE-rich regions might be particularly fruitful targets for studying chromatin structure in ES cells; this has indeed been borne out. However, the results here do not explain the functional role of the HCNEs themselves. Although HCNEs are markedly enriched at many of the Lys27 and Lys4 sites in both ES and differentiated cells, they tend overall to be distributed across much larger regions. One possibility is that some of the HCNEs dictate chromosome conformation or nuclear localization in a manner that facilitates robust gene regulation and/or epigenetic switching (Chambeyron and Bickmore, 2004; Kosak and Groudine, 2004).

Further studies will be needed to define bivalent domains and related features. It will be important to examine the entire genome in ES cells as well as to follow their fate during development and differentiation. In particular, it will be interesting to determine whether the bivalent domains that persist following ES cell differentiation correspond to genes that remain poised for later induction. In addition, it will be valuable to characterize the bivalent domains with respect to other epigenetic modifications and the binding sites of additional TFs. We note that preliminary studies of H3 Lys9 methylation show no evidence of association with bivalent domains.

A deeper understanding of bivalent domains may shed light on mechanisms that underlie the maintenance of pluripotency in ES cells and lineage fidelity in differentiated cells. Moreover, a comprehensive inventory of the presence or absence of bivalent domains over key developmental genes may provide valuable markers of cell identity and differentiation potential, both in normal and pathologic states.

EXPERIMENTAL PROCEDURES

Cell Culture

The first source of ES cells (ES1) were V6.5 murine ES cells (genotype 129SvJae × C57BL/6; male; passages 10–15). They were cultivated in 5% CO_2 at 37° on irradiated MEFs in DMEM containing 15% FCS, leukemia-inhibiting factor, penicillin/streptomycin, L-glutamine, and nonessential amino acids (Rideout et al., 2000). At least two to three passages under feeder-free conditions on 0.2% gelatin were used to exclude feeder contamination. The second source of ES cells (ES2, used for micrococcal nuclease-ChIP) was SF1-1 murine ES cells (genotype C57BL/6 × M. spretus F1; male; passages 11–16) grown in the absence of feeder cells on gelatinized plates as described previously (Umlauf et al., 2004). Primary mouse lung fibroblasts (ATCC #CCL-206), mouse embryonic fibroblasts (10.5 p.c.) immortalized with polyoma virus, C2C12 myoblasts (ATCC #CRL-1772), and Neuro2a neuroblastoma cells (ATCC #CCL-131) were grown in DMEM with 10% fetal bovine serum and penicillin/streptomycin at 37°, 5% CO_2. ES1 cells were differentiated into pan-neural precursor cells through embryoid body formation for 4 days and selection in ITSFn media for 5–7 days (Okabe et al., 1996), and they were maintained in FGF2 and EGF2 (both from R&D Systems) containing chemically defined media as described (Conti et al., 2005). These cells uniformly express nestin and Sox2 and upon growth factor withdrawal differentiate into neurons, astrocytes, and oligodendrocytes (Brustle et al., 1999; Conti et al., 2005).

Chromatin Immunoprecipitation

ChIP experiments for all cells except ES2 were carried out as described in Bernstein et al., 2005 and at http://www.upstate.com. Briefly, ~5 × 10^7 cells were trypsinized, fixed with 1% formaldehyde, resuspended in Lysis Buffer, and fragmented with a Branson 250 Sonifier to a size range of 200 to 1,000 bases. Solubilized chromatin was diluted 10-fold in ChIP dilution buffer and, after removal of a control aliquot, incubated at 4°C overnight with antibody against trimethyl Lys4 (Abcam #8580) or trimethyl Lys27 (Upstate #07-449). Immune complexes were precipitated with Protein A-sepharose, washed sequentially with Low Salt Immune Complex Wash, LiCl Immune Complex Wash, and TE, and then eluted in Elution Buffer. After cross-link reversal and Proteinase K treatment, ChIP and control DNA samples were extracted with phenol-chloroform, precipitated under ethanol, treated with RNase and Calf Intestinal Alkaline Phosphatase, and purified with a MinElute Kit (Qiagen).

Micrococcal Nuclease-ChIP

Chromatin fragments of one to six nucleosomes were prepared from unfixed chromatin from ES2 cells by micrococcal nuclease digestion and immunoprecipitated using antibody against trimethyl Lys27 (Plath et al., 2003) or dimethyl Lys4 (Upstate #07-030) as described (Umlauf et al., 2004). Immunoprecipitated DNA fractions and a control DNA sample enriched using unrelated antisera (against chicken antibodies) were extracted and purified as described above.

Sequential ChIP

Cross-linked chromatin from ES cells was immunoprecipitated with antibody against trimethyl Lys27 as described above (see "Chromatin Immunoprecipitation") except that chromatin was eluted in a solution of 30 mM DTT, 500 mM NaCl, and 0.1% SDS at 37°. Eluted chromatin was diluted 50-fold, subjected to a second immunoprecipitation with antibody against trimethyl Lys4, and then eluted with standard Elution Buffer. The isolated DNA was extracted and purified as above. In addition, a "reverse" sequential ChIP was carried out in which chromatin was immunoprecipitated first with antibody against trimethyl Lys4 and then with antibody against trimethyl Lys27.

Real-Time PCR

PCR primers for evaluating ChIP assays were designed to amplify 150–200 base pair fragments from the indicated genomic regions. Real-time PCR was carried out using Quantitect SYBR green PCR mix (Qiagen) in an MJ Research Opticon Instrument. For ChIP experiments, either 0.5 ng ChIP DNA or 0.5 ng control DNA was used as template, and fold-enrichments were determined by the $2^{-\Delta CT}$ method described in the Applied Biosystems User Bulletin. For sequential ChIP experiments, 2 μl sequential ChIP DNA or 2 μl of a 1:100 dilution of input DNA was used as template, and relative fold-enrichments were determined by the $2^{-\Delta\Delta CT}$ method, using HoxA3 Ups as the normalizer. Ratios were determined from two independent ChIP or sequential ChIP assays, each evaluated in duplicate by real-time PCR. Primers corresponding to the Irx2, Dlx1, or Hlxb9 TSSs were used to test for enrichment of Lys9 methylation (see Supplemental Data). RT-PCR was used to measure gene expression in ES cells and neural precursor cells. Briefly, RNA was isolated using an RNeasy mini kit (Qiagen), reverse transcribed, and quantified using SYBR green PCR master mix on a 7000 ABI detection system. Primer sequences are available in Supplemental Data.

Region Selection and Array Design

The identification of 204 HCNE-rich loci on the basis of sequence comparisons across the human, mouse, and dog genomes was reported previously (Lindblad-Toh et al., 2005). For the current study, we selected 56 HCNE-rich loci, including all four Hox clusters, as well as 5 control loci that do not show unusual HCNE density (ACTA locus, chromosome 19 gene desert, CD33r locus, BRCA1 locus, cytokine cluster). Each region was mapped to the mouse genome using mm5 coordinates (Table S1). Custom tiling arrays for these regions were obtained from Affymetrix Inc. (Santa Clara, CA). They contain approximately 1.3 million probe pairs, each consisting of perfect match (PM) and single base mismatch (MM) 25-mer oligonucleotides, designed to interrogate the unique sequence in these regions at approximately 30 base intervals (Kapranov et al., 2002).

DNA Amplification and Array Hybridization

ChIP and control DNA samples were amplified by in vitro transcription, converted into double-stranded cDNA with random primers, fragmented with DNase I, and end-labeled with biotin as described (Kapranov et al., 2002; Liu et al., 2003; Cawley et al., 2004; Bernstein et al., 2005). ChIP and control samples (5–10 μg) were hybridized to separate oligonucleotide arrays. Arrays were hybridized 16–18 hr at 45°C, washed, stained, and scanned using an Affymetrix GeneChip Scanner 3000 7G as described in the Affymetrix Expression Analysis Technical Manual.

Analysis of ChIP Tiling Array Data

Raw array data were quantile-normalized, scaled, and analyzed as described (Cawley et al., 2004; Bernstein et al., 2005). Enrichment was quantified using a Wilcoxon Rank Sum test applied to the transformation log2(max[PM-MM,1]) for data from ChIP and control arrays within a window of ±500 base pairs, testing the null hypothesis that ChIP and control data come from the same probability distribution. Genomic positions belonging to enriched regions were defined by applying a high P-value cutoff of 10^{-4}. These regions were extended locally by merging adjacent windows with P-values of at least 10^{-2}, and resultant positions separated by <2 kb were merged to form a predicted Lys4 or Lys27 methylated site. Bivalent domains were defined as Lys27 methylated regions > 5 kb that overlap Lys4 sites > 1 kb. Histone methylation data are available as interactive tracks at: http://www.broad.mit.edu/cell/hcne_chromatin.

Genomic Analysis

We collated a list of known TSSs based on RefSeq and Genbank mRNAs aligned to the examined regions in mouse (mm5) and the orthologous regions in human (hg17; alignments obtained from the UCSC genome browser). The methylation status of each TSS was based on the presence of significantly enriched Lys4 or Lys27 sites or bivalent domains within 2 kb upstream or downstream. The expected density of CpG and transposable elements at methylated sites was determined from random intervals of the same size, anchored in nonrepetitive sequence. CpG islands were defined as in Takai and Jones, 2002. TEZs (transposon exclusion zones) were defined as regions satisfying one of two criteria: (1) regions of at least 10 kb without any transposable elements; (2) regions of at least 15 kb with no more than 250 bases annotated as transposable elements. Identified regions were then merged together as one TEZ if they were within 1kb. For the genome-wide search, only criterion (1) was used. Supplemental Data are available at: http://www.broad.mit.edu/cell/hcne_chromatin.

Supplemental Data

Supplemental Data include three figures and five tables and can be found with this article online at http://www.cell.com/cgi/content/full/125/2/315/DC1/.

ACKNOWLEDGMENTS

We thank Aisling O'Donovan, Jen Couget, Phil Kapranov, Rob Schneider, Andi Gnirke, Christina Hughes, Leslie Gaffney, and Michelle Clamp for insightful comments and suggestions. We gratefully acknowledge Kathryn Coser and Toshi Shioda at the Massachusetts General Hospital Cancer Center microarray core facility for assistance with hybridizations and Claire Reardon at the Bauer Center for Genomics Research for assistance with real-time PCR. B.E.B. is supported by a K08 Development Award from the National Cancer Institute. A.M. was supported by a Boehringer Ingelheim Fonds (BIF) PhD fellowship. M.W. is supported by a longterm fellowship provided by the Human Frontier Science Program Organization (HFSP). R.F. acknowledges grant funding from the European Science Foundation (EuroSTELLS) and from ARC (France). S.L.S. is an investigator in the Howard Hughes Medical Institute. This work was supported in part by grants from the National Cancer Institute (CA84198 to R.J.), the National Institute of Child Health and Human Development (HD045022 to R.J.), the National Institute for General Medical Sciences (GM38627 to S.L.S.), the National Human Genome Research Institute (to E.S.L.), and by funds from the Broad Institute of MIT and Harvard.

Received: November 4, 2005
Revised: January 18, 2006
Accepted: February 23, 2006
Published: April 20, 2006

REFERENCES

Arnaud, P., Goubely, C., Pelissier, T., and Deragon, J.M. (2000). SINE retroposons can be used in vivo as nucleation centers for de novo methylation. Mol. Cell. Biol. *20*, 3434–3441.

Ayton, P.M., Chen, E.H., and Cleary, M.L. (2004). Binding to nonmethylated CpG DNA is essential for target recognition, transactivation, and myeloid transformation by an MLL oncoprotein. Mol. Cell. Biol. *24*, 10470–10478.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. (2004). Ultraconserved elements in the human genome. Science *304*, 1321–1325.

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

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al.

(2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947–956.

Brustle, O., Jones, K.N., Learish, R.D., Karram, K., Choudhary, K., Wiestler, O.D., Duncan, I.D., and McKay, R.D. (1999). Embryonic stem cell-derived glial precursors: a source of myelinating transplants. Science *285*, 754–756.

Cao, R., and Zhang, Y. (2004). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol. Cell *15*, 57–67.

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.

Conti, L., Pollard, S.M., Gorba, T., Reitano, E., Toselli, M., Biella, G., Sun, Y., Sanzone, S., Ying, Q.L., Cattaneo, E., and Smith, A. (2005). Niche-independent symmetrical self-renewal of a mammalian tissue stem cell. PLoS Biol. *3*, e283.

Delaval, K., and Feil, R. (2004). Epigenetic regulation of mammalian genomic imprinting. Curr. Opin. Genet. Dev. *14*, 188–195.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a polycomb group protein complex. Science *306*, 1574–1577.

Greally, J.M. (2002). Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. Proc. Natl. Acad. Sci. USA *99*, 327–332.

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.

Henikoff, S., Furuyama, T., and Ahmad, K. (2004). Histone variants, nucleosome assembly and epigenetic inheritance. Trends Genet. *20*, 320–326.

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

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.

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.

Kimura, H., Tada, M., Nakatsuji, N., and Tada, T. (2004). Histone code modifications on pluripotential nuclei of reprogrammed somatic cells. Mol. Cell. Biol. *24*, 5710–5720.

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.

Koli, K., Wempe, F., Sterner-Kock, A., Kantola, A., Komor, M., Hofmann, W.K., von Melchner, H., and Keski-Oja, J. (2004). Disruption of LTBP-4 function reduces TGF-beta activation and enhances BMP-4 signaling in the lung. J. Cell Biol. *167*, 123–133.

Kosak, S.T., and Groudine, M. (2004). Gene order and dynamic domains. Science *306*, 644–647.

Koyanagi, M., Baguet, A., Martens, J., Margueron, R., Jenuwein, T., and Bix, M. (2005). EZH2 and histone 3 trimethyl lysine 27 associated with Il4 and Il13 gene silencing in T(H)1 cells. J. Biol. Chem. *280*, 31470–31477.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Lee, J.H., and Skalnik, D.G. (2005). CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J. Biol. Chem. *280*, 41725–41731.

Lindblad-Toh, K., Wade, C.M., Mikkelsen, T.S., Karlsson, E.K., Jaffe, D.B., Kamal, M., Clamp, M., Chang, J.L., Kulbokas, E.J., 3rd, Zody, M.C., et al. (2005). Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature *438*, 803–819.

Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., et al. (2004). Role of transposable elements in heterochromatin and epigenetic control. Nature *430*, 471–476.

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.

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.

Mogass, M., York, T.P., Li, L., Rujirabanjerd, S., and Shiang, R. (2004). Genomewide analysis of gene expression associated with Tcof1 in mouse neuroblastoma. Biochem. Biophys. Res. Commun. *325*, 124–132.

Nobrega, M.A., Ovcharenko, I., Afzal, V., and Rubin, E.M. (2003). Scanning human gene deserts for long-range enhancers. Science *302*, 413.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S.C., Surani, M.A., and Jenuwein, T. (2001). The polycomb-group gene Ezh2 is required for early mouse development. Mol. Cell. Biol. *21*, 4330–4336.

Okabe, S., Forsberg-Nilsson, K., Spiro, A.C., Segal, M., and McKay, R.D. (1996). Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. Mech. Dev. *59*, 89–102.

O'Neill, L.P., and Turner, B.M. (2003). Immunoprecipitation of native chromatin: NChIP. Methods *31*, 76–82.

Orlando, V., Jane, E.P., Chinwalla, V., Harte, P.J., and Paro, R. (1998). Binding of trithorax and Polycomb proteins to the bithorax complex: dynamic changes during early Drosophila embryogenesis. EMBO J. *17*, 5141–5150.

Perez-Iratxeta, C., Palidwor, G., Porter, C.J., Sanche, N.A., Huska, M.R., Suomela, B.P., Muro, E.M., Krzyzanowski, P.M., Hughes, E., Campbell, P.A., et al. (2005). Study of stem cell function using microarray experiments. FEBS Lett. *579*, 1795–1801.

Perry, P., Sauer, S., Billon, N., Richardson, W.D., Spivakov, M., Warnes, G., Livesey, F.J., Merkenschlager, M., Fisher, A.G., and Azuara, V. (2004). A dynamic switch in the replication timing of key regulator genes in embryonic stem cells upon neural induction. Cell Cycle *3*, 1645–1650.

Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. Science *300*, 131–135.

Pray-Grant, M.G., Daniel, J.A., Schieltz, D., Yates, J.R., 3rd, and Grant, P.A. (2005). Chd1 chromodomain links histone H3 methylation with SAGA- and SLIK-dependent acetylation. Nature *433*, 434–438.

Rideout, W.M., 3rd, Wakayama, T., Wutz, A., Eggan, K., Jackson-Grusby, L., Dausman, J., Yanagimachi, R., and Jaenisch, R. (2000). Generation of mice from wild-type and targeted ES cells by nuclear cloning. Nat. Genet. *24*, 109–110.

Ringrose, L., Ehret, H., and Paro, R. (2004). Distinct contributions of histone H3 lysine 9 and 27 methylation to locus-specific stability of polycomb complexes. Mol. Cell *16*, 641–653.

Ringrose, L., and Paro, R. (2004). Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu. Rev. Genet. *38*, 413–443.

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

Sado, T., and Ferguson-Smith, A.C. (2005). Imprinted X inactivation and reprogramming in the preimplantation mouse embryo. Hum. Mol. Genet. *14*, R59–R64.

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.

Schmitt, S., Prestel, M., and Paro, R. (2005). Intergenic transcription through a polycomb group response element counteracts silencing. Genes Dev. *19*, 697–708.

Schraets, D., Lehmann, T., Dingermann, T., and Marschalek, R. (2003). MLL-mediated transcriptional gene regulation investigated by gene expression profiling. Oncogene *22*, 3655–3668.

Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. Dev. Cell *4*, 481–495.

Sims, R.J., 3rd, Chen, C.F., Santos-Rosa, H., Kouzarides, T., Patel, S.S., and Reinberg, D. (2005). Human but not yeast CHD1 binds directly and selectively to histone H3 methylated at lysine 4 via its tandem chromodomains. J. Biol. Chem. *280*, 41789–41792.

Szutorisz, H., and Dillon, N. (2005). The epigenetic basis for embryonic stem cell pluripotency. Bioessays *27*, 1286–1293.

Takai, D., and Jones, P.A. (2002). Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc. Natl. Acad. Sci. USA *99*, 3740–3745.

Tomczak, K.K., Marinescu, V.D., Ramoni, M.F., Sanoudou, D., Montanaro, F., Han, M., Kunkel, L.M., Kohane, I.S., and Beggs, A.H. (2004). Expression profiling and identification of novel genes involved in myogenic differentiation. FASEB J. *18*, 403–405.

Umlauf, D., Goto, Y., Cao, R., Cerqueira, F., Wagschal, A., Zhang, Y., and Feil, R. (2004). Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. Nat. Genet. *36*, 1296–1300.

Valk-Lingbeek, M.E., Bruggeman, S.W., and van Lohuizen, M. (2004). Stem cells and cancer; the polycomb connection. Cell *118*, 409–418.

van Steensel, B. (2005). Mapping of genetic and epigenetic regulatory networks using microarrays. Nat. Genet. Suppl. *37*, S18–S24.

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.

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., et al. (2005). Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. *3*, e7.

Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H., and Allis, C.D. (2005). WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. Cell *121*, 859–872.