Leading Edge
Review

Cell

The Mammalian Epigenome

Bradley E. Bernstein,[1,2,3,*] Alexander Meissner,[4] and Eric S. Lander[3,4]
[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] Whitehead Institute for Biomedical Research, MIT, Cambridge, MA 02139, USA
*Correspondence: bbernstein@partners.org
DOI 10.1016/j.cell.2007.01.033

Chemical modifications to DNA and histone proteins form a complex regulatory network that modulates chromatin structure and genome function. The epigenome refers to the complete description of these potentially heritable changes across the genome. The composition of the epigenome within a given cell is a function of genetic determinants, lineage, and environment. With the sequencing of the human genome completed, investigators now seek a comprehensive view of the epigenetic changes that determine how genetic information is made manifest across an incredibly varied background of developmental stages, tissue types, and disease states. Here we review current research efforts, with an emphasis on large-scale studies, emerging technologies, and challenges ahead.

Introduction

The sequencing of the human genome is now essentially complete (Lander et al., 2001; McPherson et al., 2001). Yet, the primary sequence is only a foundation for understanding how the genetic program is read. Superimposed upon the DNA sequence is a layer of heritable ''epigenetic'' information that we have only just begun to read and appreciate. This epigenetic information is stored as chemical modifications to cytosine bases and to the histone proteins that package the genome. By regulating chromatin structure and DNA accessibility, these chemical changes influence how the genome is made manifest across a diverse array of developmental stages, tissue types, and disease states (Bird, 2002; Goll and Bestor, 2005; Margueron et al., 2005).

The past few years have seen remarkable progress in our ability to characterize epigenetic modifications at a global scale, and some enlightening patterns have begun to emerge. In this review, we discuss large-scale studies of cytosine methylation and histone modifications in mammalian cells. These investigations have revealed diverse epigenetic controls ranging from hypermethylated DNA at promoters of silenced tumor suppressor genes to broad domains of modified histones at developmental loci (see also the Review by P.A. Jones and S.B. Baylin, page 683 of this issue; and the Review by M.A. Surani et al., page 747 of this issue). This review emphasizes both fundamental biological insights as well as the technologies that have enabled these studies. We conclude with a look forward at the technological and organizational challenges that must be addressed to achieve a ''whole-genome'' understanding of the epigenetic mechanisms that regulate normal physiology and human disease.

Epigenetic Modifications in Mammalian Genomes

Epigenetic modifications fall into two main categories: DNA methylation and histone modifications. In vertebrates, DNA methylation occurs almost exclusively in the context of CpG dinucleotides, and most CpGs in the genome are methylated (Bird, 2002; Goll and Bestor, 2005). Non-CpG methylation (CNG and CNN) has an established functional role in plants (Chan et al., 2005) and might also act in mammals. It has been observed at a low frequency in the early mouse embryo (Haines et al., 2001) and embryonic stem (ES) cells, but is significantly decreased in somatic tissues (Ramsahoye et al., 2000). A recent study implicated CpA methylation as means for allelic exclusion in sensory neurons (Lomvardas et al., 2006). If non-CpG methylation plays a functional role in vertebrate genomes, it is important to note that only some of the approaches described here can detect this modification.

The core histones that make up the nucleosome are subject to more than 100 different posttranslational modifications, including acetylation, methylation, phosphorylation, and ubiquitination (see also the Review by T. Kouzarides, page 693 of this issue). These occur primarily at specific positions within the amino-terminal histone tails. Although the vast majority of these modifications remain poorly understood, recent years have seen considerable progress in the understanding of lysine acetylation and methylation. Whereas lysine acetylation almost always correlates with chromatin accessibility and transcriptional activity, lysine methylation can have different effects depending on which residue is modified (Figure 1). Methylation of histone H3 lysine 4 (H3K4) and H3 lysine 36 is associated with transcribed chromatin. In contrast, methylation of H3 lysine 9 (H3K9), H3 lysine 27 (H3K27), and H4 lysine 20 (H4K20)

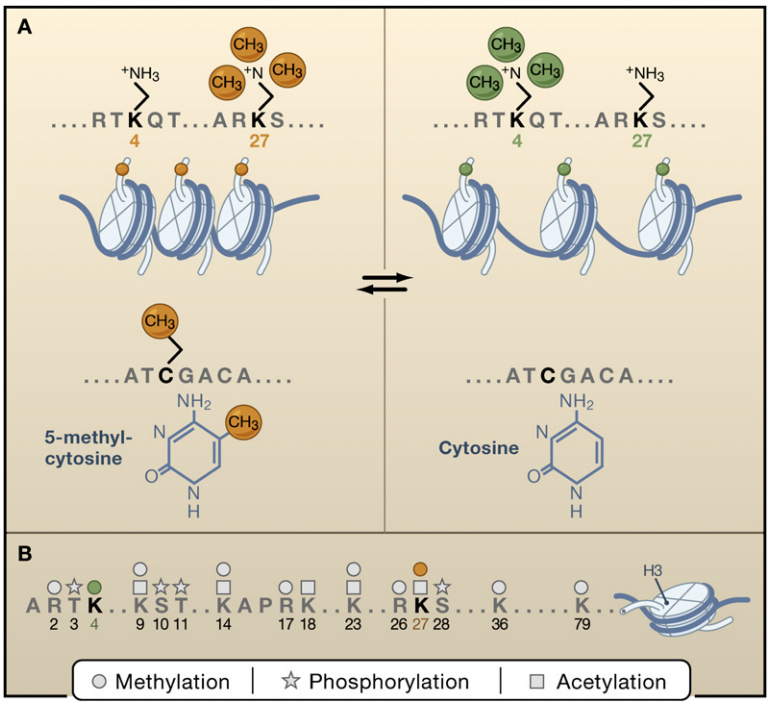

Figure 1. Cytosine and Histone Methylation

Cytosine methylation is the only known covalent modification of DNA in mammals. In contrast, histones are subject to hundreds of modifications, including acetylation, methylation, phosphorylation, and ubiquitination. (A) illustrates the structures and effects of cytosine methylation (repressive/red) and two histone marks: H3K27 methylation (repressive/red) and H3K4 methylation (activating/green). (B) illustrates the diversity of histone H3 modifications.

generally correlate with repression. Distinct histone modifications can influence each other and may also interact with DNA methylation, in part through the activities of protein complexes that bind modified histones or methylated cytosines (Figure 2) (see also the Review by B. Li et al., page 707 of this issue).

The effect of lysine acetylation reflects, in part, neutralization of the charge interaction between the DNA backbone and the histone tails. However, histone modifications also function to recruit other enzymes to specific regions of the genome. Acetylated lysines are recognized by bromodomains within nucleosome remodeling complexes. An interaction between methylated H3K4 and the Chd1 chromodomain appears to recruit activating complexes to chromatin (Pray-Grant et al., 2005; Sims et al., 2005). In contrast, methylated H3K9 and H3K27 are bound by HP1 and Polycomb, respectively, which mediate chromatin compaction (Margueron et al., 2005). A given lysine can have up to three methyl groups, and this "methyl state" can influence chromodomain binding. Polycomb preferentially interacts with trimethylated H3K27, while HP1 shows preference for both di- and trimethylated H3K9 (Fischle et al., 2003).

Epigenetic Inheritance

Chemical modifications to histone proteins and cytosine bases provide heritable epigenetic information that is not encoded in the nucleotide sequence. Cytosine methylation patterns are clearly propagated through cell division. Their preservation involves the "maintenance" methyltransferase Dnmt1, which has specificity for hemi-methylated CpG dinucleotides; the enzyme thus can methylate CpGs in a newly synthesized DNA strand based on the presence of methylation in the CpG dinucleotide in the complementary template strand (Bird, 2002; Goll and Bestor, 2005).

A subset of histone modifications also appears to show epigenetic inheritance. In yeast (which lack DNA methylation), interactions between hypoacetylated histones and SIR proteins (*S. cerevisiae*) or between H3K9 methylated histones and the Swi6 chromodomain (*S. pombe*) maintain the heterochromatic state through cell division (Grewal and Moazed, 2003). The latter situation appears to involve a positive feedback mechanism in which (1) H3K9 methylation recruits Swi6 via its chromodomain; (2) Swi6 in turn recruits the H3K9 methyltransferase Clr4; and (3) Clr4 in turn modifies H3K9 on other histones in the vicinity. Because histones segregate randomly during cell division, each daughter chromosome should inherit some modified histones (provided the modified region is sufficiently large); the modification state could then "spread" locally to newly deposited histones. Indeed, multiple protein complexes in chromatin have complementary binding and modifying activities and may thus contribute to the epigenetic maintenance of histone modification patterns (see also the Review by A. Groth et al., page 721 of this issue).

In addition, compelling evidence supports the heritability of specific histone modifications in multicellular organisms. In particular, H3K27 and H3K4 methylation are catalyzed by Polycomb-group (PcG) and trithorax-group (trxG) protein complexes, which mediate mitotic inheritance of lineage-specific gene expression patterns (Ringrose and Paro, 2004; see also the Review by B. Schuettengruber et al., page 735 of this issue). PcG

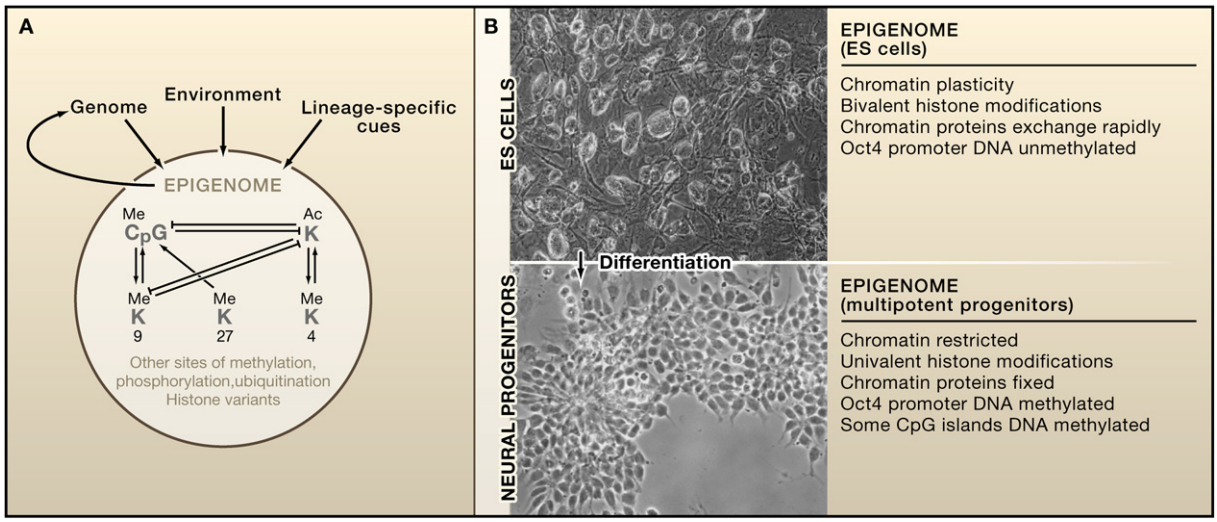

Figure 2. The Epigenome Is a Complete Description of "Heritable" Modifications to DNA and Histone Proteins as They Occur across the Genome

(A) The epigenome's makeup within a given cell is a function of genetic determinants, lineage-specific cues, and environment. The different chemical changes interact to form a complex regulatory network that modulates chromatin structure and genome function (Margueron et al., 2005).

(B) Several lines of evidence suggest that the epigenome of pluripotent embryonic stem cells is uniquely plastic (Meshorer et al., 2006; Vire et al., 2006; and references in text).

proteins dissociate from chromosomes during mitosis, raising the question of how they maintain this information. A physical interaction between PcG complexes and methylated histones retained within the chromatin could direct them back to their target sites after cell division. Still, the details of this intriguing model remain elusive and, furthermore, the extent to which other modifications are heritable remains enigmatic. Models of inheritance are further obscured by replication-independent histone deposition and by the potentially significant role of histone variants (Henikoff et al., 2004). Nevertheless, the term epigenome is used loosely to refer to cytosine methylation and the full repertoire of histone modifications, with the expectation that only a subset of the latter modifications will have epigenetic inheritance.

A Dynamic Landscape of Cytosine Methylation

DNA Methylation in Development and Disease

Mammalian DNA methylation has been implicated in a diverse range of cellular functions and pathologies, including tissue-specific gene expression, cell differentiation, genomic imprinting, X chromosome inactivation, regulation of chromatin structure, carcinogenesis, and aging (Bird, 2002). It is essential for normal development (Li et al., 1992; Okano et al., 1999) and remains indispensable for the survival of differentiated cells (Jackson-Grusby et al., 2001). Mechanistically, a methylated cytosine base can function to promote or preclude recruitment of regulatory proteins. In the former case, the methyl mark can be read through a family of methyl-CpG binding proteins thought to mediate transcriptional repression through interactions with histone deacetylases (reviewed in Bird, 2002). Alternatively, the methyl mark can exclude DNA binding proteins from their target sites, as has been shown for CTCF binding at the H19 locus (Hark et al., 2000).

DNA methylation patterns are dynamic in development and disease. In early mammalian development the paternal genome is actively demethylated shortly after protamine-histone exchange in the male pronucleus. The maternal genome subsequently becomes demethylated, presumably through a passive DNA replication mechanism (Reik et al., 2001; Santos et al., 2002). Genomewide methylation levels increase rapidly in the blastocyst, establishing a differential pattern between the cells of the ICM and those of the trophectoderm, and ultimately resulting in the formation of methylation patterns found in the adult. In addition to the coordinated changes during normal development, the DNA methylome undergoes characteristic changes in pathologies such as cancer. These include genomewide loss of methylation and aberrant local gain of methylation. In particular, tumor suppressor gene promoters are targets of hypermethylation, which typically results in their silencing (Jones and Baylin, 2002). An even more central and early role for epigenetic changes in tumor development has recently been proposed. This model suggests that cancer may evolve from a population of nonneoplastic, polyclonal, epigenetically disrupted stem/progenitor cells, potentially with additional genetic lesions (Feinberg et al., 2006; see also the Review by P.A. Jones and S.B. Baylin, page 683 of this issue).

CpG Islands

CpGs tend to cluster in regions, termed CpG islands, that are characterized by high (G+C) and CpG content (Bird, 2002). CpG islands cover about 0.7% of the human

genome (depending on the precise definition), but contain 7% of the CpG dinucleotides (Fazzari and Greally, 2004; Lander et al., 2001). The enrichment of CpG dinucleotides in these regions suggests that they are unmethylated, at least in the germ line, and thus evade the high divergence rate for methylated CpGs. This divergence is due to the fact that the mismatch repair system can accurately recognize and correct the deamination product of cytosine bases (uracil), but not the deamination product of methylcytosine (thymine). About 60% of human gene promoters are associated with CpG islands. Although it has been suggested that most CpG islands are always unmethylated, a subset have been shown to be subject to tissue-specific methylation during development (Bird, 2002; Strichman-Almashanu et al., 2002). A computational analysis of CpG occurrences and restriction site distributions in mammalian genomes suggested further that a substantial proportion of CpG islands become methylated in differentiated tissues (Fazzari and Greally, 2004). This view is supported by recent experimental studies of the DNA methylome (see below). In addition, a significant fraction of CpG dinucleotides reside within repetitive elements, but these are heavily methylated in somatic tissues.

Various authors have proposed specific definitions of CpG islands based on sequence features, although it is important to remember that these computational criteria are not a perfect predictor of the methylation status. The original criteria defined CpG islands as regions of at least 200 bases with a (G+C)-content of at least 50% and a ratio of observed CpG frequency to expected CpG frequency of at least 0.6 (Gardiner-Garden and Frommer, 1987). A more refined definition, the Takai-Jones criteria, provides a better association with 5′ regions of genes and excludes most Alu repeats (Takai and Jones, 2002). Computational definitions of CpG islands are somewhat arbitrary and thus exclude many 5′ regions with only a limited number CpGs. Many promoters that lack strictly defined CpG islands have nonetheless been shown to have tissue-specific methylation patterns that strongly correlate with transcriptional activity. For example, the methylation status of the Oct-4 and Nanog promoters correlates well with expression, though neither contains an annotated CpG island (Blelloch et al., 2006; Hattori et al., 2004). Conversely, tissue-specific demethylation is associated with transcriptional activation of IL2 and Sry (Bruniquel and Schwartz, 2003; Nishino et al., 2004). The functional relevance of such regions awaits further study. In the end, the best definition of CpG islands will be based not on computational prediction, but on direct experimental evaluation of methylation status.

Studying the DNA Methylome

There are a number of techniques for studying cytosine methylation at specific loci, several of which have been adapted for large-scale analyses. Beck and colleagues have undertaken an ambitious collaborative study of cytosine methylation within CpG islands and non-CpG islands in normal and diseased tissues. The group is using "bisulfite sequencing" as a gold standard approach. DNA is treated with sodium bisulfite to convert unmethylated cytosines to uracils and then subjected to conventional DNA sequencing; unmethylated cytosines will be read as thymine, while methylated cytosines will be read as cytosine. Initial studies focused on the human MHC locus (Rakyan et al., 2004), but a recent scale-up analyzed the methylation status of about 40,000 CpGs on chromosomes 6, 20, and 22 in various tissues (Eckhardt et al., 2006). In both studies the majority of analyzed loci showed a bimodal methylation distribution profile (either hypermethylated or hypomethylated). Only a small fraction (9.2%) of the 511 CpG islands was found to be methylated. In contrast, almost 50% of non-CpG islands containing 5′ UTRs were hypermethylated. The group extrapolated from a subset of amplicons that about 70% may have conserved methylation profiles between mouse and human.

Though highly accurate, this gold standard sequencing approach is not readily scalable—at least, not with the current generation of DNA sequencing technology. Hence, several groups have used a variety of other approaches to generate large-scale cytosine methylation datasets in recent years. These typically involve fractionation of methylated and unmethylated portions of the genome by methyl-sensitive restriction or antibodies, followed by microarray- or sequencing-based analysis (see the Emerging Technologies subsection).

The overall aims of these studies were to compare either different cell types (Bibikova et al., 2006; Ching et al., 2005; Khulan et al., 2006; Rollins et al., 2006; Strichman-Almashanu et al., 2002; Yamada et al., 2004) or normal and tumor samples (Hu et al., 2005; Keshet et al., 2006; Weber et al., 2005; Weisenberger et al., 2006). Bibikova et al. identified characteristic epigenetic profiles for ES cells and differentiated cells. Moreover, they found that the average methylation level of the analyzed CpGs (selected from 5′ regions of 371 genes) was about 35%. Similarly, Yamada et al. showed that a significant fraction of CpG islands on chromosome 21 is methylated in a variety of tissues, including 31/149 in peripheral blood cells. The occurrence of CpG island methylation in these studies is higher than reported in the study by Eckhardt et al. (2006). Therefore, more detailed and genomewide analyses are required to determine the full extent of CpG island methylation. Several groups observed distinct epigenetic signatures associated with specific tumors. For instance, Keshet et al. could detect de novo methylation at 135 promoters (of which 127 contain CpG islands) when comparing a colon cancer cell line with normal colon. These investigators found that tumor-specific methylated genes fall into distinct functional classes and tend to cluster along chromosomes. Hu et al. found that DNA methylation patterns in breast carcinoma varied markedly with tumor stage and type. Weisenberger et al. used a multiplexed PCR-based approach to distinguish a subset of colorectal tumors with high frequencies of CpG island methylation. Though each of the studies only touched upon the vast landscape of cytosine methylation, they are nonetheless

highly informative and support and justify more comprehensive and coordinated epigenome studies.

Global Insights into Histone Biology

Methods for Large-Scale Analysis of Histone Modifications

The past several years have also brought considerable progress in the development of large-scale tools for analyzing histone modifications. These tools rely heavily on a procedure called chromatin immunoprecipitation (chromatin IP or ChIP) in which chromatin is immunoprecipitated with antibody against a transcription factor, a chromatin-associated protein, or a modified histone. PCR can then be used to query for the presence or absence (or relative enrichment) of a predefined sequence in the chromatin IP DNA. Alternatively, a panel of primers can be used to interrogate a given locus.

Far more extensive analyses can be achieved using microarrays (ChIP-on-chip) or by sequencing the chromatin IP DNA. Tiling oligonucleotide arrays that cover the entire nonrepetitive portions of the human and mouse genomes are now available from several sources. These are not restricted to currently annotated genes but may detect epigenetic changes associated with uncharacterized transcriptional units or regulatory elements. Moreover, they generate continuous data along chromosomes and can thus define the extents and boundaries of genomic regions with modified histones. The ChIP-on-chip assay can in principle be applied to any histone modification for which an effective antibody is available, though in practice data quality is highly dependent on which modification is being analyzed and the efficiency of the antibody pulldown. Sequencing technologies are advancing rapidly and hold great promise for epigenome study. However, a very large number of sequencing reads are required for sufficient coverage of a mammalian genome (see Emerging Technologies subsection).

Landscape of Activating Histone Modifications

Systematic studies of chromatin modifications have revealed a complex landscape including punctate sites of modified histones at transcription start sites, distal regulatory elements and conserved sequences, and broad domains at gene clusters and developmental loci. Initial small-scale studies of the murine β-globin locus revealed acetylated histones associated with globin gene promoters, the locus control region, and extended subdomains in a tissue-specific and developmentally regulated fashion (Bulger et al., 2003; Forsberg et al., 2000). Roh and colleagues used a sequencing method to map histone H3 acetylation in human T cells (Roh et al., 2005). They identified nearly 50,000 acetylated sites in the human genome that correlate with active transcription start sites and CpG islands and tend to cluster within gene-rich loci. Nearly half of the acetylated sites were intergenic, frequently colocalizing with known T cell regulatory elements, DNase hypersensitive sites, and other sequences showing strong evolutionary conservation between human and mouse. The authors also compared acetylation patterns in resting and activated T cells and found roughly 4000 sites unique to the activated cells.

In parallel studies, ChIP-on-chip analysis was used to map H3 acetylation and H3K4 methylation in cultured human and mouse cells. High-resolution tiling oligonucleotide arrays were used to interrogate either all active promoters (Kim et al., 2005) or the nonrepetitive portions of human chromosomes 21 and 22 plus several orthologous human and mouse loci (Bernstein et al., 2005). Many of the findings were consistent with the sequencing analysis, with modified histones mapping to transcription start sites and putative regulatory elements in a cell-type-specific manner. Patterns of H3 acetylation and H3K4 trimethylation were nearly identical. Global studies in yeast and flies have also demonstrated colocalization of various activating histone modifications (Liu et al., 2005; Pokholok et al., 2005; Schubeler et al., 2004). These findings suggest that multiple active histone modifications combine redundantly to ensure robust chromatin regulation (Schreiber and Bernstein, 2002). However, the extent to which other histone modifications may contribute to greater functional complexity in chromatin [also referred to as the "histone code" (Jenuwein and Allis, 2001)] remains unclear, in part due to a lack of data on their global distributions.

Notably, a significant proportion of sites enriched for activating modifications do not coincide with conserved genomic sequence. Nevertheless, two lines of evidence suggest these sites may also be functional. First, a comparison of H3K4 methylation at orthologous loci in analogous human and mouse cells (primary fibroblasts from lung) revealed striking conservation of methylation patterns, even in regions where the underlying sequence is only modestly conserved (Bernstein et al., 2005). Second, a sequence element underlying an acetylated site in T cells that did not show an unusual degree of sequence conservation was nonetheless found to function effectively as an enhancer in a reporter assay (Roh et al., 2005). These findings illustrate the potential of epigenomic analysis to identify novel regulatory elements that may not be readily discernable through comparative genomics.

Genomewide Targets of PcG Complexes

PcG proteins play essential roles in development and in the epigenetic maintenance of lineage-specific gene repression (Ringrose and Paro, 2004; see also the Review by B. Schuettengruber et al., page 735 of this issue). They are required for ES cell pluripotency and are markedly downregulated upon differentiation (Valk-Lingbeek et al., 2004). Polycomb repressive complex 2 (PRC2) catalyzes H3K27 methylation, while PRC1 binds methylated H3K27 and mediates chromatin compaction (Margueron et al., 2005; Ringrose and Paro, 2004). Several recent studies applied ChIP-on-chip analysis to identify regulatory targets of PcG complexes. Young and colleagues used genomewide tiling arrays to map PRC2 binding in human ES cells (Lee et al., 2006). They identified more than 1000 gene targets, most of which were also enriched for H3K27 trimethylation. These include a large number of

genes encoding developmental regulators such as homeobox transcription factors and key signaling proteins. This work was complemented by a parallel study from Jaenisch and colleagues that used promoter arrays to map PRC2 and PRC1 binding in murine ES cells and also identified a large number of developmentally important gene targets (Boyer et al., 2006). The targets are largely transcriptionally silent in ES cells, but many were activated in ES cells lacking a critical PRC2 component.

Farnham and colleagues used ChIP-on-chip analysis to examine PRC2 localization in embryonal carcinoma cells and several tumor lines (Squazzo et al., 2006). Identified targets included transcriptionally silenced genes encoding developmentally important transcription factors, as well as glycoproteins and immunoglobulin receptors. These investigators also found a strong association between PRC2 binding and H3K27 trimethylation, but importantly, ruled out concomitant trimethylation of H3K9 (at least in the embryonal carcinoma cells). An unexpected global correlation was observed between H3K9 trimethylation and RNA polymerase II occupancy. Although H3K9 methylation has a clear role in heterochromatin formation, this result is consistent with a prior study that observed H3K9 trimethylation and HP1γ binding within a number of actively transcribed regions (Vakoc et al., 2005).

Helin and colleagues examined PRC2 and PRC1 binding in mouse embryonic fibroblasts (Bracken et al., 2006). These researchers found that both complexes colocalize along with H3K27 trimethylation to around 1000 genes, many with developmental functions. They showed that a common subset of the genes is de-repressed by siRNA-mediated knockdown of components of either complex, suggestive of a functional link between PRC2-mediated histone methylation and PRC1-mediated chromatin compaction in gene silencing.

Chromatin Domains and Cellular Memory

A common theme of the PcG complex studies is that binding and associated H3K27 methylation often involves expansive genomic regions. This was nicely illustrated by the genomewide tiling array analysis of PRC2, which showed that binding at developmental regulator genes extends 2 to 35 kb from the promoters, while binding at other genes occurs in a more punctate fashion (Lee et al., 2006). These ''repressive'' domains at developmental regulator genes are comparable in size to ''activating'' domains of H3K4 methylation previously identified in the Hox clusters (Bernstein et al., 2005). Notably, these activating domains are also occupied by the trxG protein MLL1 (Guenther et al., 2005). Subsequent studies have shown that developmental regulator genes in differentiated cells are frequently associated with broad domains enriched for either trimethylated H3K27 or trimethylated H3K4 (Bernstein et al., 2006). The domains are highly cell-type-specific, with H3K27 domains marking genes repressed in a given lineage, and H3K4 domains marking active ones. Similarly expansive regions of chromatin modification also affect other highly regulated loci in mammalian genomes (Morshead et al., 2003; Szutorisz et al., 2005).

Chromatin domains could theoretically provide a robust epigenetic memory to maintain expression or repression of critical lineage-specifying genes. While punctate modification sites of just a few adjacent histones could easily be lost during mitosis when histones segregate randomly to the daughter strands, large domains with significant numbers of modified histones would likely be inherited by both daughter strands and could promote similar modification of newly deposited histones (Henikoff et al., 2004). The epigenome studies presented above support a central role for chromatin domains with PcG or trxG proteins in the epigenetic control of developmental regulator genes. Notably, this paradigm appears well conserved in the fruit fly *D. melanogaster*, which has been the subject of many seminal observations on PcG function in metazoan development. In particular, recent studies have revealed that PcG complexes bind across similarly expansive regions that also encode transcription factors (Negre et al., 2006; Schwartz et al., 2006; Tolhuis et al., 2006).

Epigenetic Mechanisms of Genome Plasticity

The studies described above thus reveal a central role for domains of PcG complexes in the repression of developmental regulator genes in both differentiated cells and undifferentiated ES cells. However, several lines of evidence suggest that the domains in ES cells have an unusual structure and plasticity that may contribute to pluripotency. For example, ChIP-on-chip analysis revealed that large H3K27 trimethylated regions in murine ES cells frequently overlap smaller H3K4 methylated sites (Bernstein et al., 2006). Sequential chromatin IP assays confirmed that these opposing modifications coexist at the same locus on the same chromosome. These regions, termed ''bivalent domains,'' overlay developmental regulator genes that are largely silent in ES cells. Fisher and colleagues have also reported colocalization of these opposing histone modifications specifically in pluripotent cells (Azuara et al., 2006). Notably, these researchers found that the bivalent regions adopt an open structure, as judged by their early replication status, which is atypical of PcG-associated chromatin.

Hence, although they are associated with gene repression, H3K27 methylated domains in ES cells retain H3K4 methylation and other characteristics of active chromatin. There is an interesting analogy here to early fly development. Prior studies have shown that the bithorax locus is co-occupied by PcG and trxG proteins, with the latter being essential for subsequent gene induction (Orlando et al., 1998; Ringrose and Paro, 2004). By analogy, trxG proteins and H3K4 methylation within bivalent domains may keep developmental regulator genes poised for induction in ES cells.

A key issue remains whether bivalent chromatin is unique to ES cells (Figure 2). Limited studies in multipotent neural and hematopoietic cells did not reveal evidence of bivalent domains (Azuara et al., 2006; Bernstein et al., 2006). However, a recent ChIP-sequencing study by Zhao and colleagues suggested that some promoters

Table 1. Genome to Epigenome

Genetic Feature	Epigenetic Feature	Correlation of Genetic and Epigenetic Feature		Notation
		ES Cells	Differentiated Cells	
CpG islands	DNA methylation-free	Strong	Moderate	Genomic principles suggest CpG islands methylation-free, yet some are methylated in differentiated cells
CpG islands	H3K4 methylation	Strong	Moderate	In differentiated cells, some CpG islands lose H3K4 methylation
Transposon-free regions	H3K27 methylation	Strong	Moderate	In differentiated cells, some transposon-free regions lose H3K27 methylation, while other regions gain this mark
Conserved noncoding elements	H3K27 methylation	Variable	Variable	Domains of PcG proteins and H3K27 methylation coincide with clusters of highly conserved noncoding elements
Repetitive elements	DNA methylation	Strong	Strong	Repetitive elements silenced by DNA methylation
Repetitive elements	H3K9, H3K27, H4K20 methylation	Variable	Variable	Dependent on repeat class and developmental stage

Correlations between genetic and epigenetic features are shown for pluripotent ES cells and differentiated cells. The correlations tend to be more significant in ES cells, suggesting a potentially important role for DNA sequence in defining the newly reprogrammed epigenome (see text for references).

in primary human T cells may exhibit both H3K27 and H3K4 methylation (Roh et al., 2006). In particular, the HOXB7 promoter showed robust consecutive enrichments in a sequential chromatin IP. The authors draw an analogy to the proposed role of bivalent chromatin in ES cells and suggest that a similar mechanism may prime the dynamic gene expression changes that occur in T cells upon antigen recognition (Roh et al., 2006). It remains unclear whether the "bivalent promoters" defined by Roh and colleagues are equivalent to the structures observed in ES cells or represent a different kind of structure. It should also be noted that the T cell study used a relatively permissive threshold for declaring H3K27 and H3K4 methylated promoters. Further studies and improved analysis tools are needed to define the roles of bivalent histone modifications and other aspects of chromatin plasticity in pluripotency, multipotency, and cell fate decisions during development.

The Relationship between Genome and Epigenome

The Newly Reprogrammed Epigenetic State

Germ cell development and early embryogenesis both involve genomewide epigenetic reprogramming that is intimately tied to changes in the developmental potency (Mager and Bartolomei, 2005; Reik et al., 2001). Current technologies are unable to study the epigenome during reprogramming as they require too many cells. However, pluripotent ES cells, derived from the inner cell mass where remethylation begins in early development, are representative of the newly reprogrammed state.

The role of DNA sequence in defining the ES cell epigenome is likely to be significant. Accordingly, striking associations have been identified between specific genomic features and histone methylation patterns in ES cells. H3K4 methylation coincides to a remarkable extent with CpG islands (Bernstein et al., 2006). This association may reflect a causal relationship inasmuch as trxG complexes that catalyze H3K4 methylation contain domains that bind unmethylated CpG dinucleotides (Birke et al., 2002; Lee and Skalnik, 2005). The correlation is much weaker in differentiated cells, largely due to loss of H3K4 methylation at a subset of CpG islands (Table 1).

H3K27 methylation in ES cells correlates with a distinct genomic feature that has only recently been appreciated: large regions that are strongly depleted, or free, of transposable elements (Bernstein et al., 2006; Simons et al., 2006). Most transcription start sites with H3K27 methylated domains coincide with regions of more than 10 kb with little or no identifiable transposon-derived sequence. There could be a paucity of transposons because transposition is incompatible with the chromatin structures or because transposon insertion near developmental genes is subject to strong negative selection. Alternatively, transposons could interfere with the spreading of H3K27 methylation, once this modification is initiated through other mechanisms. Notably, the *gypsy* transposable element in *D. melanogaster* insulates against Polycomb-mediated repression and can halt the spreading of H3K27 methylation (Kahn et al., 2006; Mallin et al., 1998; Sigrist and Pirrotta, 1997).

Highly conserved noncoding sequence elements in mammalian genomes may also play a role in defining the epigenome. These elements are prevalent within regions that are PRC2 bound and H3K27 methylated in ES cells

(Bernstein et al., 2006; Lee et al., 2006). A subset could function in PRC2 recruitment in a way analogous to the Polycomb response elements in flies (Ringrose and Paro, 2004). However, highly conserved elements in mammals are distributed across vast genomic regions much larger than the chromatin domains, and thus are likely to have additional functions (Lindblad-Toh et al., 2005; Woolfe et al., 2005). Potential examples include a role in higher-order genome organization or other long-range epigenetic silencing phenomena. Notably, Clark and colleagues recently identified a 4 MB band of human chromosome 2q.14.2, replete with conserved elements, that is subject to widespread silencing by DNA methylation in colorectal cancers (Frigola et al., 2006).

Repetitive Elements

Close to 50% of the human genome consists of transposable elements and other repetitive DNA. The functions of these sequences remain poorly understood, and they may play significant roles in global epigenetic control. Transposon-derived DNA sequence is typically highly methylated in somatic tissues (Bird, 2002).

Repetitive sequences in the genome are also associated with characteristic histone modifications. Tandem satellite repeats are found within centric and pericentric heterochromatin, marked by H3K9 and H4K20 trimethylation. A detailed study by Jenuwein and colleagues revealed that LINEs, SINEs, and other interspersed repeats have variable degrees of H3K9, H3K27, and H4K20 methylation (Martens et al., 2005). Even within a given repeat class, modification status appeared to depend on the cellular differentiation state. Most array studies ignore repetitive sequence, for the technical reason that they cross-hybridize to one another. However, a tiling array analysis of *A. thaliana* chromatin suggested that H3K9 methylation may spread from tandem repeats and epigenetically silence nearby genes (Lippman et al., 2004). Another array study found that the H3K9 methyl-interacting protein HP1 associates with repeat-dense regions of the *D. melanogaster* genome (de Wit et al., 2005).

Predicting Cytosine Methylation from DNA Sequences

CpG islands are generally assumed to be unmethylated in the germ line (except imprinted loci). However, several groups have attempted to predict cytosine methylation patterns in differentiated cells based on DNA sequence. Bock and colleagues sought to discriminate CpG islands that are prone to methylation in human lymphocytes from those that remain unmethylated (Bock et al., 2006). They found that CpG islands that remain unmethylated are particularly GC- and CpG-rich. Conversely, islands that are prone to methylation in the lymphocytes are enriched for segmental duplications, tandem repeats, and sequences with multiple self-alignments. Das and colleagues attempted to identify general sequence predictors of DNA methylation in brain tissue (Das et al., 2006). As expected, they found that unmethylated sequences are enriched for CpG islands, and depleted of Alu elements. These investigators also identified short sequence motifs that help discriminate between methylated and unmethylated DNA fractions from brain tissue, though their functional significance remains unclear.

Emerging Technologies in Epigenome Research

Cytosine Methylation

An increasing number of techniques to detect and compare DNA methylation on a larger scale have been reported in the past years. Most are derived from previously established techniques that have been combined with either microarrays or high-throughput sequencing technology. For example, methylated or unmethylated DNA fractions can be enriched by digesting DNA samples with methylation-sensitive (or methylation-dependent) restriction enzymes (MSREs), and then analyzed by array-hybridization or sequencing (Lippman et al., 2004; Rollins et al., 2006; Strichman-Almashanu et al., 2002; Yamada et al., 2004).

A limitation of MSRE-based methods is that, while they discriminate for or against methylation at the recognition site of the particular enzyme used, they cannot directly reveal the methylation status of cytosines or CpG dinucleotides outside the restriction site. An alternative approach for distinguishing methylated and unmethylated fractions involves immunoprecipitation with a methylcytosine antibody (Keshet et al., 2006; Weber et al., 2005). Ecker and colleagues recently combined this method with a whole-genome *A. thaliana* tiling array in the first complete high-resolution analysis of a DNA methylome (Zhang et al., 2006). This technique is theoretically less biased in that it can detect any methylated site. However, it depends on the specificity of the antibody, and resolution is limited by the size of the immunoprecipitated DNA fragments and by the coverage of the microarrays.

The most precise way to map methylcytosines is to use bisulfite sequencing, as described above, which provides information at the nucleotide level. Although highly informative in the largest such study to date (Eckhardt et al., 2006), the scaling of this method is challenging because it involves so much locus-specific amplification. An alternative approach is simply to perform ''shotgun sequencing'' on a sample of bisulfite-treated DNA. The approach may be applied to the entire genome or a reproducible subset of the genome—for example, the DNA contained within restriction fragments of a given size range (termed reduced representation sequencing). In a small pilot study, a number of key facts about bisulfite sequencing were established: (1) by using an ES cell line deficient for all active DNA methyltransferases, it could be demonstrated that near-complete bisulfite conversion (>99.9%) can be achieved; (2) amplification bias appears minimal; and (3) the technique readily scales with increasing sequencing capacity, and is automatable as no locus-specific primers are needed for PCR or sequencing (Meissner et al., 2005).

The relative value of specific techniques will depend on the application—for example, whether the aim is to profile many samples at distinct loci or to determine the exact

genomewide distribution of all methylcytosines. The latter will likely involve analyzing many epigenomes at nucleotide resolution, potentially through genomewide bisulfite sequencing. Although such an undertaking represents a significant technical challenge, recent advances in bisulfite conversion, library generation, and high-throughput sequencing suggest its ultimate feasibility (Margulies et al., 2005; Meissner et al., 2005; Rakyan et al., 2004; Shendure et al., 2005). The apparently bimodal status of cytosine methylation (Bird, 2002; Rakyan et al., 2004) suggests that even moderate sequencing coverage may be sufficient for most of the genome. By contrast, deeper analysis of cytosine methylation patterns across multiple tissue types will likely be achieved through complementary approaches that rely on microarrays or other multiplexed technologies. It is through the integration of these distinct but complementary approaches that the landscape and dynamics of the DNA methylome will most clearly be elucidated.

Histone Modifications

Chromatin IP remains the primary method used by researchers to interrogate the modification status of histones associated with a specific gene, regulatory element, or genomic position. Standard chromatin IP protocols involve treating cells with formaldehyde to cross-link DNA and nearby proteins, sonicating the chromatin to generate fragments just a few nucleosomes in size, and then immunoprecipitating with antibody against a particular histone modification.

However, the methodology has a number of limitations. Efficacy depends on the precise epitope and the quality of the antibody; better reagents are needed for many specific epigenetic modifications. Fixation and fragmentation may introduce biases, and certain modifications may be "masked" by chromatin proteins that bind modified histones. These issues can be partially overcome through the use of alternate techniques, such as N-ChIP, biotin-tag affinity purification, or DamID (Mito et al., 2005; O'Neill and Turner, 2003; van Steensel et al., 2001). The N-ChIP technique is of particular value for analyzing histone modifications because it fragments chromatin by micrococcal nuclease digestion rather than sonication and does not require cross-linking (O'Neill and Turner, 2003). Biotin-tag affinity purification has been used to map histone variants fused to a biotin ligase recognition peptide (Mito et al., 2005). The technique achieves high specificity by relying on streptavidin pull-down rather than immunoprecipitation. DamID is an alternative method for mapping chromatin-associated proteins or transcription factors in which the protein of interest is fused to a DNA adenine methylase, and its DNA targets identified by restriction with adenine methylation-sensitive enzymes (van Steensel et al., 2001).

Chromatin IP is scalable for global analysis with microarrays (ChIP-on-chip) or, to an increasing extent, sequencing technologies. Oligonucleotide tiling arrays are of particular value, as they offer high-resolution, comprehensive coverage. Limitations of ChIP-on-chip analysis include insensitivity to repetitive elements, susceptibility to amplification bias, and the fact that many epitopes currently lack suitable antibody reagents. The technique also requires millions of cells. Although Turner and colleagues recently reported a chromatin IP protocol applicable to significantly fewer cells (O'Neill et al., 2006), their approach results in modest enrichments and thus may not scale effectively.

Sequencing-based methods have also been used to map histone modifications (Roh et al., 2005), although their use has been limited by the large numbers of reads required. The approach involves sequencing chromatin IP DNA and aligning the reads to the genome. The probability that a given genomic region was enriched in the chromatin IP is determined from the number of sequencing reads that fall within the region. The total number of reads needed to map a given modification genomewide can be modeled on a Poisson distribution as a function of both chromatin IP enrichment and the desired resolution and accuracy. Consider a modification such as H3K4 trimethyl that enriches target regions efficiently in a chromatin IP assay (e.g., 30-fold). Mapping such a modification to the human genome at a resolution of 500 base pairs and a sensitivity and specificity of 90% would theoretically require roughly one million reads. However, mapping a modification with antibodies that provide less efficient enrichment (e.g., 10-fold) would require roughly three million reads. In the latter case, only a small minority of individual reads would be expected to align to true target regions. Still, more advanced sequencing technologies with significantly greater read capacities have shown promise for transcription factor mapping (Ng et al., 2006) and have good potential for epigenome analysis.

Future Perspectives and Challenges

The epigenome studies reviewed here—although informative, diverse, and impressive in scope—only begin to describe the immensely complex epigenetic regulatory network controlling genome function in development and disease. A more complete understanding of the roles played by cytosine methylation and the different histone modifications in normal development and disease will require further study and, in particular, improved methodologies for cell and molecular biology, genomics, and computation.

One of the most important issues will be reduction of the number of cells required for histone modification studies. This would enable epigenetic profiling of limited populations of stem cells and other primary tissues, as well as profiling of microdissected tumor samples. Innovative cell and molecular biology techniques, such as more efficient and faithful amplification procedures, should help in this regard. In addition, better antibody reagents, including monoclonal antibodies, are needed to increase sensitivity, as well as to enable interrogation of additional modifications and to improve consistency among laboratories.

A second challenge will be to develop better tools for highly parallel DNA analyses. Oligonucleotide tiling arrays

with increasingly high probe densities will continue to improve coverage, resolution, and cost-effectiveness of ChIP-on-chip studies. These platforms are also enabling high-resolution studies of cytosine methylation through restriction- and antibody-based assays. Massively parallel sequencing technology is also advancing at a rapid pace and becoming increasingly attractive for epigenome analysis. Moreover, various multiplexed technologies being developed for expression analysis, copy number measurements, and mutation detections are also likely to play important roles (Fan et al., 2006).

Finally, computational tools are urgently needed to analyze and integrate the diverse epigenomic data being generated. Existing tools have largely been adapted from platforms designed for other problems and are suboptimal. It will be important to develop new approaches that consider the unique attributes of epigenomic data and can integrate high-resolution data on cytosine methylation, histone modifications, chromatin domains, and boundaries across different tissues and developmental stages. An effective platform will need to incorporate epigenomic data in the context of other genomic information such as RNA expression, DNA copy number, mutations, and so on.

The studies reviewed here have broken important new ground, but leave unanswered many central questions regarding the global distributions, dynamics, and regulatory functions of these diverse modifications. These issues are currently being addressed by many individual laboratories. However, their efforts could benefit from a larger framework to encourage, coordinate, standardize, and integrate the various studies and the diverse data that they are generating. We note recent discussions about an ''epigenome project'' that would produce draft analyses of cytosine methylation, key histone modifications and variants, and chromatin-associated proteins in carefully chosen cell states (Jones and Martienssen, 2005; Qiu, 2006). Such information could offer critical insights into the relationships between genotype, phenotype, and environment, and serve as a catalyst for future studies of the epigenetic mechanisms that regulate normal physiology and human disease. Though complex in execution and interpretation, a concerted effort toward understanding the epigenome would ultimately be rewarded with a far richer understanding of how the genetic code is made manifest across an incredibly varied background of developmental stages, tissue types, and disease states.

ACKNOWLEDGMENTS

We thank Rudolf Jaenisch, Tarjei Mikkelsen, John Greally, Peter Jones, Konrad Hochedlinger, and Aviva Presser for helpful comments and critical readings of the manuscript. We also acknowledge our colleagues at Massachusetts General Hospital, the Broad Institute, and the Whitehead Institute for stimulating discussions and critical contributions to some of the work reviewed here. B.E.B. is supported by funds from the National Cancer Institute, the Burroughs Wellcome Fund, the Harvard Stem Cell Institute, and Massachusetts General Hospital. A.M. was supported by a Boehringer Ingelheim Fonds (BIF) PhD fellowship. E.S.L. is supported by funds from the National Human Genome Research Institute and the Broad Institute of MIT and Harvard.

REFERENCES

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

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.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell *125*, 315–326.

Bibikova, M., Chudin, E., Wu, B., Zhou, L., Garcia, E.W., Liu, Y., Shin, S., Plaia, T.W., Auerbach, J.M., Arking, D.E., et al. (2006). Human embryonic stem cells have a unique epigenetic signature. Genome Res. *16*, 1075–1083.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Birke, M., Schreiner, S., Garcia-Cuellar, M.P., Mahr, K., Titgemeyer, F., and Slany, R.K. (2002). The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic Acids Res. *30*, 958–965.

Blelloch, R., Wang, Z., Meissner, A., Pollard, S., Smith, A., and Jaenisch, R. (2006). Reprogramming efficiency following somatic cell nuclear transfer is influenced by the differentiation and methylation state of the donor nucleus. Stem Cells *24*, 2007–2013. Published online May 18, 2006. 10.1634/stemcells.2006-0050.

Bock, C., Paulsen, M., Tierling, S., Mikeska, T., Lengauer, T., and Walter, J. (2006). CpG island methylation in human lymphocytes is highly correlated with DNA sequence, repeats, and predicted DNA structure. PLoS Genet *2*, e26. 10.1371/journal.pgen.0020026.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353. Published online April 19, 2006. 10.1038/nature04733.

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

Bruniquel, D., and Schwartz, R.H. (2003). Selective, stable demethylation of the interleukin-2 gene enhances transcription by an active process. Nat. Immunol. *4*, 235–240.

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.

Chan, S.W., Henderson, I.R., and Jacobsen, S.E. (2005). Gardening the genome: DNA methylation in Arabidopsis thaliana. Nat. Rev. Genet. *6*, 351–360.

Ching, T.T., Maunakea, A.K., Jun, P., Hong, C., Zardo, G., Pinkel, D., Albertson, D.G., Fridlyand, J., Mao, J.H., Shchors, K., et al. (2005). Epigenome analyses using BAC microarrays identify evolutionary conservation of tissue-specific methylation of SHANK3. Nat. Genet. *37*, 645–651.

Das, R., Dimitrova, N., Xuan, Z., Rollins, R.A., Haghighi, F., Edwards, J.R., Ju, J., Bestor, T.H., and Zhang, M.Q. (2006). Computational

prediction of methylation status in human genomic sequences. Proc. Natl. Acad. Sci. USA *103*, 10713–10716.

de Wit, E., Greil, F., and van Steensel, B. (2005). Genomewide HP1 binding in Drosophila: developmental plasticity and genomic targeting signals. Genome Res. *15*, 1265–1273.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V.K., Attwood, J., Burger, M., Burton, J., Cox, T.V., Davies, R., Down, T.A., et al. (2006). DNA methylation profiling of human chromosomes *6*, 20 and 22. Nat. Genet. *38*, 1378–1385.

Fan, J.B., Chee, M.S., and Gunderson, K.L. (2006). Highly parallel genomic assays. Nat. Rev. Genet. *7*, 632–644.

Fazzari, M.J., and Greally, J.M. (2004). Epigenomics: beyond CpG islands. Nat. Rev. Genet. *5*, 446–455.

Feinberg, A.P., Ohlsson, R., and Henikoff, S. (2006). The epigenetic progenitor origin of human cancer. Nat. Rev. Genet. *7*, 21–33.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev. *17*, 1870–1881.

Forsberg, E.C., Downs, K.M., Christensen, H.M., Im, H., Nuzzi, P.A., and Bresnick, E.H. (2000). Developmentally dynamic histone acetylation pattern of a tissue-specific chromatin domain. Proc. Natl. Acad. Sci. USA *97*, 14494–14499.

Frigola, J., Song, J., Stirzaker, C., Hinshelwood, R.A., Peinado, M.A., and Clark, S.J. (2006). Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. Nat. Genet. *38*, 540–549.

Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. J. Mol. Biol. *196*, 261–282.

Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. Annu. Rev. Biochem. *74*, 481–514.

Grewal, S.I., and Moazed, D. (2003). Heterochromatin and epigenetic control of gene expression. Science *301*, 798–802.

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.

Haines, T.R., Rodenhiser, D.I., and Ainsworth, P.J. (2001). Allele-specific non-CpG methylation of the Nf1 gene during early mouse development. Dev. Biol. *240*, 585–598.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature *405*, 486–489.

Hattori, N., Nishino, K., Ko, Y.G., Ohgane, J., Tanaka, S., and Shiota, K. (2004). Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells. J. Biol. Chem. *279*, 17063–17069.

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

Hu, M., Yao, J., Cai, L., Bachman, K.E., van den Brule, F., Velculescu, V., and Polyak, K. (2005). Distinct epigenetic changes in the stromal cells of breast cancers. Nat. Genet. *37*, 899–905.

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E., and Jaenisch, R. (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat. Genet. *27*, 31–39.

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

Jones, P.A., and Baylin, S.B. (2002). The fundamental role of epigenetic events in cancer. Nat. Rev. Genet. *3*, 415–428.

Jones, P.A., and Martienssen, R. (2005). A blueprint for a Human Epigenome Project: the AACR Human Epigenome Workshop. Cancer Res. *65*, 11241–11246.

Kahn, T.G., Schwartz, Y.B., Dellino, G.I., and Pirrotta, V. (2006). Polycomb complexes and the propagation of the methylation mark at the Drosophila ubx gene. J. Biol. Chem. *281*, 29064–29075.

Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R.A., Niveleau, A., Cedar, H., and Simon, I. (2006). Evidence for an instructive mechanism of de novo methylation in cancer cells. Nat. Genet. *38*, 149–153.

Khulan, B., Thompson, R.F., Ye, K., Fazzari, M.J., Suzuki, M., Stasiek, E., Figueroa, M.E., Glass, J.L., Chen, Q., Montagna, C., et al. (2006). Comparative isoschizomer profiling of cytosine methylation: the HELP assay. Genome Res. *16*, 1046–1055.

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.

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.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., et al. (2006). Control of developmental regulators by polycomb in human embryonic stem cells. Cell *125*, 301–313.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

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., Kaplan, T., Kim, M., Buratowski, S., Schreiber, S.L., Friedman, N., and Rando, O.J. (2005). Single-nucleosome mapping of histone modifications in S. cerevisiae. PLoS Biol. *3*, e328. 10.1371/journal.pbio.0030328.

Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J., and Axel, R. (2006). Interchromosomal interactions and olfactory receptor choice. Cell *126*, 403–413.

Mager, J., and Bartolomei, M.S. (2005). Strategies for dissecting epigenetic mechanisms in the mouse. Nat. Genet. *37*, 1194–1200.

Mallin, D.R., Myung, J.S., Patton, J.S., and Geyer, P.K. (1998). Polycomb group repression is blocked by the Drosophila suppressor of Hairy-wing [su(Hw)] insulator. Genetics *148*, 331–339.

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

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376–380. Published online July 31, 2005. 10.1038/nature03959.

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.

McPherson, J.D., Marra, M., Hillier, L., Waterston, R.H., Chinwalla, A., Wallis, J., Sekhon, M., Wylie, K., Mardis, E.R., Wilson, R.K., et al. (2001). A physical map of the human genome. Nature *409*, 934–941.

Meissner, A., Gnirke, A., Bell, G.W., Ramsahoye, B., Lander, E.S., and Jaenisch, R. (2005). Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. Nucleic Acids Res. *33*, 5868–5877.

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

Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Genome-scale profiling of histone H3.3 replacement patterns. Nat. Genet. *37*, 1090–1097.

Morshead, K.B., Ciccone, D.N., Taverna, S.D., Allis, C.D., and Oettinger, M.A. (2003). Antigen receptor loci poised for V(D)J rearrangement are broadly associated with BRG1 and flanked by peaks of histone H3 dimethylated at lysine 4. Proc. Natl. Acad. Sci. USA *100*, 11577–11582.

Negre, N., Hennetin, J., Sun, L.V., Lavrov, S., Bellis, M., White, K.P., and Cavalli, G. (2006). Chromosomal distribution of PcG proteins during Drosophila development. PLoS Biol. *4*, e170. 10.1371/journal.pbio.0040170.

Ng, P., Tan, J.J., Ooi, H.S., Lee, Y.L., Chiu, K.P., Fullwood, M.J., Srinivasan, K.G., Perbost, C., Du, L., Sung, W.K., et al. (2006). Multiplex sequencing of paired-end ditags (MS-PET): a strategy for the ultra-high-throughput analysis of transcriptomes and genomes. Nucleic Acids Res. *34*, e84.

Nishino, K., Hattori, N., Tanaka, S., and Shiota, K. (2004). DNA methylation-mediated control of Sry gene expression in mouse gonadal development. J. Biol. Chem. *279*, 22306–22313.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

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

O'Neill, L.P., VerMilyea, M.D., and Turner, B.M. (2006). Epigenetic characterization of the early embryo with a chromatin immunoprecipitation protocol applicable to small cell populations. Nat. Genet. *38*, 835–841.

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.

Pokholok, D.K., Harbison, C.T., Levine, S., Cole, M., Hannett, N.M., Lee, T.I., Bell, G.W., Walker, K., Rolfe, P.A., Herbolsheimer, E., et al. (2005). Genomewide map of nucleosome acetylation and methylation in yeast. Cell *122*, 517–527.

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.

Qiu, J. (2006). Epigenetics: unfinished symphony. Nature *441*, 143–145.

Rakyan, V.K., Hildmann, T., Novik, K.L., Lewin, J., Tost, J., Cox, A.V., Andrews, T.D., Howe, K.L., Otto, T., Olek, A., et al. (2004). DNA methylation profiling of the human major histocompatibility complex: a pilot study for the human epigenome project. PLoS Biol. *2*, e405. 10.1371/journal.pbio.0020405.

Ramsahoye, B.H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A.P., and Jaenisch, R. (2000). Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. Proc. Natl. Acad. Sci. USA *97*, 5237–5242.

Reik, W., Dean, W., and Walter, J. (2001). Epigenetic reprogramming in mammalian development. Science *293*, 1089–1093.

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 genomewide mapping. Genes Dev. *19*, 542–552.

Roh, T.Y., Cuddapah, S., Cui, K., and Zhao, K. (2006). The genomic landscape of histone modifications in human T cells. Proc. Natl. Acad. Sci. USA *103*, 12782–15787. Published online October 16, 2006. 10.1073/pnas.0607617103.

Rollins, R.A., Haghighi, F., Edwards, J.R., Das, R., Zhang, M.Q., Ju, J., and Bestor, T.H. (2006). Large-scale structure of genomic methylation patterns. Genome Res. *16*, 157–163.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev. Biol. *241*, 172–182.

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 genomewide chromatin analysis of a higher eukaryote. Genes Dev. *18*, 1263–1271.

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genomewide analysis of Polycomb targets in Drosophila melanogaster. Nat. Genet. *38*, 700–705.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., and Church, G.M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. Science *309*, 1728–1732.

Sigrist, C.J., and Pirrotta, V. (1997). Chromatin insulator elements block the silencing of a target gene by the Drosophila polycomb response element (PRE) but allow trans interactions between PREs on different chromosomes. Genetics *147*, 209–221.

Simons, C., Pheasant, M., Makunin, I.V., and Mattick, J.S. (2006). Transposon-free regions in mammalian genomes. Genome Res. *16*, 164–172.

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.

Squazzo, S.L., O'Geen, H., Komashko, V.M., Krig, S.R., Jin, V.X., Jang, S.W., Margueron, R., Reinberg, D., Green, R., and Farnham, P.J. (2006). Suz12 binds to silenced regions of the genome in a cell-type-specific manner. Genome Res. *16*, 890–900.

Strichman-Almashanu, L.Z., Lee, R.S., Onyango, P.O., Perlman, E., Flam, F., Frieman, M.B., and Feinberg, A.P. (2002). A genomewide screen for normally methylated human CpG islands that can identify novel imprinted genes. Genome Res. *12*, 543–554.

Szutorisz, H., Canzonetta, C., Georgiou, A., Chow, C.M., Tora, L., and Dillon, N. (2005). Formation of an active tissue-specific chromatin domain initiated by epigenetic marking at the embryonic stem cell stage. Mol. Cell. Biol. *25*, 1804–1820.

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.

Tolhuis, B., Muijrers, I., de Wit, E., Teunissen, H., Talhout, W., van Steensel, B., and van Lohuizen, M. (2006). Genomewide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. Nat. Genet. *38*, 694–699.

Vakoc, C.R., Mandat, S.A., Olenchock, B.A., and Blobel, G.A. (2005). Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol. Cell *19*, 381–391.

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., Delrow, J., and Henikoff, S. (2001). Chromatin profiling using targeted DNA adenine methyltransferase. Nat. Genet. *27*, 304–308.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.M., et al. (2006). The Ploycomb group protein EZH2 directly controls DNA methylation. Nature *439*, 871–874.

Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat. Genet. *37*, 853–862.

Weisenberger, D.J., Siegmund, K.D., Campan, M., Young, J., Long, T.I., Faasse, M.A., Kang, G.H., Widschwendter, M., Weener, D., Buchanan, D., et al. (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nat. Genet. *38*, 787–793.

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. 10.1371/journal.pbio.0030007.

Yamada, Y., Watanabe, H., Miura, F., Soejima, H., Uchiyama, M., Iwasaka, T., Mukai, T., Sakaki, Y., and Ito, T. (2004). A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. Genome Res. *14*, 247–266.

Zhang, X., Yazaki, J., Sundaresan, A., Cokus, S., Chan, S.W., Chen, H., Henderson, I.R., Shinn, P., Pellegrini, M., Jacobsen, S.E., and Ecker, J.R. (2006). Genomewide high-resolution mapping and functional analysis of dna methylation in arabidopsis. Cell *126*, 1189–1201.