Published in final edited form as:

Trends Genet. 2011 June ; 27(6): 242–250. doi:10.1016/j.tig.2011.03.001.

Emerging patterns of epigenomic variation

Aleksandar Milosavljevic

Molecular and Human Genetics Department, Baylor College of Medicine, Houston, Texas 77030

Abstract

Fuelled by new sequencing technologies, epigenome mapping projects are revealing epigenomic variation at all levels of biological complexity, from species to cells. Comparisons of methylation profiles among species reveal evolutionary conservation of gene body methylation patterns, pointing to the fundamental role of epigenomes in gene regulation. At the human population level, epigenomic changes provide footprints of the effects of genomic variants within the vast non-protein coding fraction of the genome while comparisons of the epigenomes of parents and their offspring point to quantitative epigenomic parent-of-origin effects confounding classical Mendelian genetics. At the organismal level, comparisons of epigenomes from diverse cell types provide insights into cellular differentiation. Finally, comparisons of epigenomes from monozygotic twins help dissect genetic and environmental influences on human phenotypes and longitudinal comparisons reveal aging-associated epigenomic drift. The development of new bioinformatic frameworks for comparative epigenome analysis is putting epigenome maps within reach of researchers across a wide spectrum of biological disciplines.

Spotting epigenetic memory within layers of epigenomic information

Epigenomics is a new arrival into an old lineage that includes the classical concept of epigenesis and the evolving concept of epigenetics. Within mainstream molecular biology, epigenetics currently has several specific definitions, all focusing on mechanisms other than changes in DNA sequence that perpetuate altered cellular activity states [1] [2] [3] [4]. These mechanisms are essential for differentiated cells, which acquire distinct “programs”, not encoded in DNA, to guide their cell-type specific homeostatic behavior within a multicellular organism. In humans and other species epigenetic memory integrates genetic, physiological and environmental influences through the lifetime of an organism and is particularly sensitive to those influences during early development.

Epigenetic memory may reside in various cellular compartments and, most conspicuously, within the nucleus. The genomic DNA sequence in a living cell is tagged by methylation of specific cytosines, and by modifications of DNA-attached protein and RNA fractions of the chromatin. Some layers of this information, such as cytosine methylation tags within more than twenty eight million cytosine-phosphate-guanine (CpG) dinucleotides in the human genome, are replicated concurrently with the DNA. Other layers may reflect altered epigenetic states that may be propagated over long periods of time and during cell division by other mechanisms. An epigenome of a specific cell population is a comprehensive genome-wide map of these chromatin tags [5] [6].

© 2011 Elsevier Ltd. All rights reserved.

Corresponding Author: Milosavljevic, A. (amilosav@bcm.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Methylation of CpG dinucleotides in genomic DNA is the most studied epigenomic mark [7]. In humans, methylation, hydroxymethylation, and possibly other modifications of the fifth carbon of cytosine occur within CpG dinucleotides. CpG methylation patterns may be reprogrammed during cellular differentiation but generally tend to be stably replicated during mitosis. A genome-wide map of cytosine methylation within a cell type comprises a methylome [8], a key component of the epigenome.

Additional layers of epigenomic information are contained within the protein fraction of chromatin. Stretches of about 147bp of DNA sequence wrap around nucleosomes to form nucleosome core particles, which in turn may form higher levels of periodic and aperiodic chromatin structure filling the nucleus [9]. Each nucleosome contains eight histone proteins (typically two H2A, H2B, H3 and H4) with globular cores and unstructured amino-terminal tails. Numerous chemical modifications of specific amino acids in histone tails such as trimethylation of the fourth lysine on histone H3 (denoted H3K4me3) and trimethylation of the twenty seventh lysine (H3K27me3) occur in specific combinations along the genome and participate in gene expression regulation and all major processes within the nucleus [10].

In addition to the DNA and protein fractions, chromatin also includes an RNA fraction which plays an important role in epigenome regulation [11] [12]. These and other layers of epigenomic information (Box 1) are reprogrammed by rearrangement and chemical modifications during cellular differentiation [13] [14]. There is an extensive cross-talk between DNA methylation changes, histone modifications, non-coding RNAs, chromatin accessibility and the three-dimensional conformation of the chromatin [15] [16].

Box 1

Five key epigenomic components

Each of the following five key epigenomic components consists of a layer of information about state of chromatin other than the information about the genomic nucleotide sequence.

Genomic DNA methylation

Methylation of cytosines in genomic DNA [7]. In differentiated human cell types cytosines that occur in a majority of close to thirty million 5'-CpG-3' sites may be methylated. Approximately half of human genes contain CpG islands, regions highly enriched for CpG dinucleotides which tend not to be methylated across tissues. Mammalian tissues can be distinguished by their methylation patterns [25, 26].

DNA-bound protein and its modifications

Includes both binding of master regulators and nucleosomes to DNA as well as their modifications. The modifications include methylation, acetylation, and other modifications of specific amino acids in nucleosomal histones [10].

Non-coding RNA

Both small [11] and large non-coding RNA [12, 100] species are involved in a diversity of mechanisms of epigenome regulation including chromatin silencing, imprinting control and X chromosome inactivation in mammals, and regulation of developmental regulators by the Polycomb-Trithorax system [101].

Chromatin accessibility

Degree of accessibility of the genomic DNA fraction of chromatin. The genomic DNA that is tightly bound by the protein component of the chromatin is less accessible to

transcription machinery and other regulatory proteins [102]. Chromatin accessibility is regulated during cellular differentiation. It associates with genomic variants and is heritable [35].

Spatial organization of the chromatin

Proximity of chromosomal loci is experimentally determined by methods such as 3C-seq (Table 1). Spatial organization itself is dynamic, may affect gene expression, correlates with genomic methylation [15] and is determined at least in part by the action of master regulators such as estrogen receptor [65].

The rapidly decreasing cost of DNA sequencing is enabling wider adoption of technologies for sequencing-based mapping of epigenomes. The most widely used technologies using sequencing in the readout step (such technologies are usually suffixed by “-seq”) include whole-genome sequencing of bisulfite-treated genomic DNA (MethylC-seq) for mapping DNA methylomes, chromatin immuno-precipitation and sequencing (ChIP-seq) for mapping histone marks and DNA binding sites of regulatory protein complexes, RNA-sequencing (RNA-seq) for mapping the RNA fraction of the epigenome and others (listed in Table 1). The recently initiated NIH Roadmap Epigenomics Project [6] is applying such assays on a large scale to map the landscape of epigenomic variation (Figure 1) and to construct the Human Epigenome Atlas (<http://www.epigenomeatlas.org>). European and other worldwide epigenome mapping projects are also being initiated [17].

Here, I review some recent epigenomic advances with a primary focus on human and mouse epigenomic variation not directly related to disease, emphasizing studies of methylomes and to a lesser degree of histone marks. I discuss fresh insights gained by epigenome mapping and comparison, particularly those overturning long-held assumptions. I start with the comparison among species, which reveals evolutionarily conserved gene body methylation patterns, pointing to the fundamental role of epigenomes in gene regulation. I then move to the human population level, examining epigenomic footprints of genomic variation outside of traditionally better understood protein coding regions. Studies of parent-offspring trios reveal intricate tissue-specific parent-of-origin epigenomic effects confounding classical Mendelian genetics. At the level of individual organisms, analysis of epigenomes from diverse cell types helps map the pathways of cellular differentiation, providing insights into mammalian development. Influences of the environment and aging on human phenotypes are being dissected by longitudinal epigenome comparison and by comparing of epigenomes of monozygotic twins.

Epigenome mapping is fuelled by rapid assay technology advances but those require matching advances in bioinformatic technologies for downstream analysis of high volumes of data. I review bioinformatic resources and methodological frameworks for epigenomics, including methods for “deciphering” epigenomic “codes” and open issues in comparative epigenome analysis. I conclude by proposing a computationally-inspired perspective on the epigenome as the working memory of the cell.

Evolutionarily conserved epigenomic patterns among species

It is widely recognized that DNA methylation, histone modifications and other layers of epigenomic information appear to be generally present across eukaryotes, but much of our understanding of epigenomic variation was until recently extrapolated to many organisms from studying just a handful of model organisms [18]. Historically, DNA methylation has been associated with transcriptional silencing of transposons [19], but recent systematic mapping of DNA methylation across a number of animals, plants and fungi using whole-genome bisulfite sequencing [20] [21] provided strong evidence that invertebrates generally

do not follow this rule. The studies instead confirm the ancestrality and ubiquity of another pattern – methylation of the gene bodies of active constitutively transcribed housekeeping genes [5]. Exons of such genes tend to be more methylated than introns and methylation tends to be excluded from transcription start and termination sites pointing to the role of methylation in regulating gene expression both at level of the whole gene and at the level of individual exons. The functional significance of gene body methylation has been most dramatically demonstrated in honeybees where knockdown of the DNA methyltransferase Dnmt3 diminishes gene body methylation and mimics the effect of royal jelly by causing larvae to develop into queens [22].

The pattern of preferential exon methylation in gene bodies is paralleled by preferential nucleosome positioning over exons and by enrichment of the histone mark H3K36me3 over transcribed exons. These patterns indicate an intimate connection between epigenomic changes and regulation of transcription at the level of whole genes and individual exons [23]. A key mechanistic link between histone marks and exon splicing was provided in a breakthrough study [24] which identified adaptor molecules that recognize exonic histone mark signature involving H3K36me3 and affect tissue-specific pre-mRNA splicing. These results reinforce the hypothesis that for some genes the epigenome harbors a “splicing code” [23] that determines tissue-specific splicing outcomes.

Tissue-specific DNA methylation patterns are sufficiently conserved between human and mouse to completely discriminate tissue types regardless of species origin [25]. Remarkably, comparative analyses revealed conservation of DNA methylation patterns between the two species even at loci where genomic DNA sequences diverged [26]. These findings are consistent with observed conservation between human and mouse of tissue-specific spatial chromatin organization into coordinately replicated megabase-sized “replication domains” [27].

Epigenomes of non-human primates are yet to be mapped at high resolution. A comparison of DNA methylation patterns of 36 genes in brain, liver and lymphocytes between chimpanzee and human [28] revealed that divergence in methylation patterns is most pronounced in the brain, hinting that the rapid evolution of the human brain may have left significant epigenomic footprints.

In summary, evolutionary patterns of epigenome variation among species are revealing evolutionarily conserved roles of epigenomes in regulation of gene transcription, splicing, tissue specificity, development, and pointing to rapid epigenomic changes that may have accompanied human evolution.

Epigenomic footprints of genomic variation among individuals of a species

Our ability to interpret functional effects of genomic sequence variants in the human genome is still largely limited to those occurring in exons and splice sites and thus affecting peptide sequences. But protein coding regions comprise less than two percent of the human genome. The effects of a fraction of sequence variants throughout the remaining fraction of the genome may now also be in principle discerned by observing associated epigenomic changes. Known examples of such effects include variants that cause disease via changes of chromatin structure in *cis*, such as CGG triplet repeat expansions in the Fragile X syndrome and repeat contractions in Facioscapulohumeral dystrophy [29]. These disease variants create allelic imbalances between epigenomes on homologous chromosomes. Such allelic imbalances are sequence-associated and therefore distinct from other allelic imbalances such as in imprinting, where epigenomic state and gene expression depend on parent of origin, from randomized patterns of X inactivation in females, and from randomized patterns of monoallelic expression of genes involved in olfaction, immunity, and other processes [30].

The results of a first high-resolution genome-wide survey of allele-specific DNA methylation in humans using the methylation-sensitive single nucleotide polymorphism (SNP) analysis based method [31] were recently published [32]. A large degree of quantitative allelic methylation imbalances was observed. The study showed that a large fraction of the imbalances was associated with sequence variation in *cis*, indicating that many sequence variants that reside on either maternal or paternal chromosomes may affect methylation levels on the same chromosome. Similar patterns were found in a study of peripheral blood mononuclear cells using whole-genome bisulfite sequencing [33]. Both studies demonstrate not only allelic DNA methylation but also allele-specific expression of affected genes. A detailed plan has been proposed [34] to utilize the information about methylation imbalances in order to identify the SNPs of functional significance within critical regions detected by genome-wide association studies (GWAS). This is important because a large proportion of critical regions identified by GWAS fall outside of protein-coding regions. In such regions, epigenomic changes may provide the only lead toward variants that have functional consequences and may therefore be causing disease.

Genomic variants also have widespread effects on chromatin accessibility and transcription factor binding in lymphoblastoid cell lines [35]. It is not clear at this point how changes of various layers of the epigenome, including histone marks, relate to each other genome-wide. However, locus-specific studies indicate coordinate allele-specific changes in DNA methylation and histone marks [36] [37].

A recent study [38] provides indications that structural genomic variants also affect the epigenome. Indeed, the effects of copy number variants on gene expression are not limited to the genes within copy-altered loci, as commonly assumed. In fact, many affected genes reside far away from the structural alteration, leading the authors to hypothesize that structural variant effects on gene expression may be mediated by changes in chromatin structure spanning hundreds of kilobasepairs. Future epigenome mapping and comparison studies should shed more light on this exciting hypothesis.

Imprinting and complex parent-of-origin effects

Genomic imprinting is a mechanism of transcriptional regulation of mammalian genes through which expression of a set of genes is restricted to one parental allele [39]. Imprinted loci typically contain multiple genes regulated by parent-of-origin specific DNA methylation of an imprinting center [40], typically a CpG island. While the allelic methylation patterns of the imprinting center tend to be shared by all somatic cells, mono-allelic (maternal or paternal) expression of some imprinted genes may be observed only in specific tissues such as brain.

Classical on/off imprinting may only be a tip of an iceberg of complex and quantitative parent-of-origin effects, as revealed recently in a study of allelic expression in mouse using the mRNA-seq method [41] [42]. The study reports a list of 1300 genes with parent-of-origin expression patterns, a thirteen-fold increase over the current number of about one hundred known imprinted genes in mouse. The deep mRNA sequencing revealed many subtle quantitative allelic imbalances, indicating that typical parent-of-origin effects are not binary. Alleles of maternal origin were preferentially expressed in embryonic brains, whereas paternal alleles were preferentially expressed in adults. Three times more genes with a sex-specific imprinting status were imprinted in female than in male offspring. The sex-specific influence was observed in the hypothalamus of female offspring, indicating a parent-specific influence on hypothalamic function in daughters but not in sons. The parent-of-origin effects revealed by these studies point to complex parental influences on behavior

and physiology of their offspring, with obvious implications for studies of human phenotypes and diseases, particularly those affecting brain function and development.

Unexpectedly, parent-of-origin preference was detected even for different transcript isoforms of the same gene. This implicates the epigenome in splicing regulation, consistent with independent mechanistic studies pointing to an epigenomic “splicing code” [23] discussed above. Mapping of CpG methylation and additional epigenomic marks will be required in order to fully understand the extent and nature of these complex parent-of-origin effects in mammals.

Epigenomic footprints of cellular differentiation

Cellular differentiation leaves highly informative footprints on the epigenome at all stages of development. Mammalian epigenome experiences two major waves of genomic DNA demethylation and chromatin remodeling during embryogenesis, one upon fertilization [43] setting the stage for cellular differentiation during development and the other in primordial germ cells (PGCs) [43], setting the stage for imprinting and return to totipotency. A pioneering methylome comparison study [44] profiled DNA methylation throughout the genomes in embryos deficient in Activation-Induced Cytidine Deaminase (AID) and wild-type mice by whole-genome bisulfite sequencing. The comparison of epigenomes revealed significantly reduced genome-wide erasure of DNA methylation in PGCs affected by AID deficiency, confirming the role of AID in demethylation. A related study [45] has argued that deamination may be a triggering event and provided evidence that base-excision repair pathway plays a key role in genomic demethylation of PGCs as well as in genomic demethylation of paternal pronuclei upon oocyte fertilization.

Comprehensive mapping of methylomes of embryonic stem cells and differentiated fibroblasts [46] [47] was a milestone in epigenomic studies of development. High degree of genome-wide methylation in non-CpG cytosines in embryonic stem cells but not in fibroblasts was detected at basepair level of resolution and for the first time characterized genome-wide [46]. Reduced methylation levels in fibroblasts were associated with lower transcriptional activity of the affected genes. Regions proximal to genes involved in pluripotency were differentially methylated, confirming the role of DNA methylation in cell fate determination.

Histone modifications also undergo changes during differentiation. The “bivalent” or “poised” state consisting of the H3K27me/H3K4me combination is particularly abundant in pluripotent cells [48]. This state is inherited by daughter cells during mitosis and makes genes responsive to certain intrinsic or environmental stimuli during cellular differentiation. There is also an extensive cross-talk between histone modifications and DNA methylation changes during cellular differentiation [16]. About half of all human genes, among them many key regulators of development regulated by the Polycomb/Trithorax system [49] contain in their promoter regions a CpG island [50]. These CpG islands are mostly unmethylated across different mammalian tissue types but may become methylated during cellular differentiation, in aging tissues, and in cancer [51] [52].

Recent comparison of methylomes and histone marks in the H1 human embryonic stem cell (hESC) line and differentiated fibroblast cell line IMR90 [53] indicates that the most dynamic DNA methylation changes occur at enhancers and other distant regulatory regions, involving complex cell-type specific interactions with H3K4me1 and H3K4me2 marks. The differentiation of hESC into fibroblasts also involves extensive spreading of repressive chromatin marks. The repressive H3K27me3 mark covers only 119 Mbp of the genome of the H1 human embryonic stem cell line but spreads over 394Mbp in differentiated fibroblast cell line IMR90. The repressive H3K9me3 mark similarly spreads from 148Mbp to

510Mbp. The gain of the repressive marks is associated with corresponding decrease in DNA methylation, challenging the common assumption that gene repression is associated with DNA methylation.

The expansion of repressive H3K9me3 and H3K27me3 marks in IMR90 and CD4+ T cells selectively affects developmental genes. Distinct genes were repressed by expanded H3K27me3 domains in the two cell types, hinting at the role of chromatin-mediated repression in cell fate determination. This is consistent with a comparison of H3K27me3 marks in pancreatic beta cells and acinar cells [54] revealing an extensive H3K27me3 program shared by the two endodermal cell types. In sharp contrast to the shared endodermal epigenomic program, the gene expression signature of beta cells resembles that of ectoderm-derived neural tissues. The study further showed that the neural expression program was co-opted during late pancreatic cell differentiation through activation of a small number of transcriptional regulators by selective removal of inactivating H3K27me3 marks.

A pioneering study of twelve human tissues [25] identified tissue-specific differentially methylated genomic regions (T-DMRs) in 17% of 873 genes examined on chromosomes 6, 20 and 22. DNA methylation in about one third of 5' untranslated regions was inversely proportional to gene expression. A bimodal distribution of DNA methylation levels with transcription start sites (TSS) was observed showing a hypomethylated core region of about 1,000bp symmetrically surrounding the sites. Several epigenome mapping studies reported T-DMRs at low-CpG promoters [55] [56] [57], CpG island shores [58] and at promoter-distal sites that exhibit enhancer activity [25] [59].

Cellular differentiation involves interaction between epigenomic changes and the action of master regulators of development including key transcription factors regulating pluripotency and micro RNAs [60] [61]. The list of prominent master regulators also includes the enhancer blocker and domain barrier associated protein CTCF [62], enhancer associated pioneer factor FOXA1 [63], master regulator of erythroid development GATA1 [64], estrogen receptor [65] and others. Binding of these regulators to typically many thousands of genomic loci is revealed by ChIP-seq assays. Regulator binding is typically accompanied by alterations in multiple histone marks, changes in chromatin accessibility, and spatial reorganization of chromosomes [65].

In summary, accumulating epigenomes from diverse cell types are providing raw material for detailed maps of cellular differentiation. More insights into cellular differentiation are expected to emerge from comparative analyses of epigenomes of differentiated cell types (Box 2 lists some issues in comparative analysis) that are being produced by current epigenome mapping projects (Table 2).

Box 2

Key issues in comparative epigenome analysis

Any specific comparative analysis of epigenomes may require that some key experimental and computational issues be addressed. In the following we briefly review those issues.

Tissue heterogeneity

Tissue samples may contain multiple distinct cell types. Consequently, differences detected by comparison of two samples may be due to different proportions of epigenomically distinct cell types within each sample. If the epigenomic signatures of constituent cell types are known in advance, multiple linear regression methods may help estimate relative abundance of different cell types within the tissue.

Assay heterogeneity

Various methylation assays differ in resolution and cover different but overlapping genomic regions [8, 37] [103] (Table 1). When comparing methylomes obtained by two different assays, comparison may be performed at resolution (window size) that equals the lower of the two. To avoid detecting spurious differences, comparison may be limited to loci equally ascertained by both assays.

Epigenome representation

For the purpose of comparison, two epigenomes may be converted into lists of numbers, each number corresponding to an average value of an epigenomic marks over a window of fixed size or over a feature such as an enhancer, a promoter or a CpG island. The choice of features may depend on the goal of comparison.

Allele-specific epigenomic states

Each cell contains two epigenomes, one on the paternal and the other on the maternal chromosome. Assays with sequencing readouts provide means of distinguishing the two signals at heterozygous loci. Pash 3.0 program [104] maps both regular and bisulfite-treated reads while detecting basepair and indel variants and can be used for calling allele-specific methylation.

Sensitive pairwise comparison

Epigenome comparison may be a search for a small or even a single important difference between two epigenomes. For example, a single difference between methylomes of monozygotic twins may be responsible for a phenotype distinguishing them. Sensitive well-characterized pipelines for such comparisons are yet to be developed. The best current solution is to perform multiple comparisons using different tools to improve the chance of detecting causative differences [73].

Multi-epigenome comparison

Appropriate algorithms must be chosen for comparative analyses involving more than 2 epigenomes. Random forest algorithms have been used to cluster collections of epigenomes by tissue type and age [78]. The use of cladistic tree reconstruction methods has been suggested for the purpose of reconstructing the bifurcating tree of cellular differentiation [93].

Epigenomic changes due to environmental influences and aging

A combination of genetic, physiological and environmental influences through the lifetime of an organism may have an effect on the epigenomic “program” acquired by a cell during differentiation. The epigenomic program may be particularly sensitive to environmental influences during early stages of development, but significant environmentally-induced and stochastic changes may accumulate over a lifetime of a mature organism, increasing risk of aging-associated diseases.

The epigenome has now been established as a key mediator of environmental influences on phenotypes in plants and animals, including mammals. Increased licking, grooming and nursing by rat mothers improved the stress response of their pups via DNA methylation and histone modifications in the promoter of a glucocorticoid receptor gene expressed in the hippocampus [66]. Both mouse [67] and human studies [68] show that maternal nutrition during gestation may leave epigenomic footprints causing obesity in offspring. Transmission of environmental influences by the epigenome sometimes carries across multiple generations [69].

A powerful approach to identify environmental influences on the epigenome and aging-associated epigenomic drift in humans is to compare epigenomes of monozygotic (MZ) twins. A pioneering study [70] applied restriction landmark genome scanning (RLGS), an early method for CpG methylation mapping using methylation-sensitive restriction enzymes, to obtain epigenome maps in a large cohort of MZ twins. Although the twins were epigenomically indistinguishable during their early years of life in terms of CpG methylation and histone acetylation, epigenomes significantly diverged with age across various tissue types. The role of the environment is indicated by the fact that the epigenomes of twins that were older, had different lifestyles, and spent less time together diverged more.

Large discordance between MZ twins in heritable psychiatric disorders, including schizophrenia and bipolar disorder provided an early rationale for epigenomic comparisons [71] [72]. The first comprehensive characterization of epigenomic differences between MZ twins discordant for a disease came from MZ twins discordant for multiple sclerosis (MS) [73], where the relevant cell types are more readily accessible. Comparison of CD4+ T-cells (cells involved in the MS-causing auto-immune reaction) involved complete genomic DNA sequences, DNA methylomes determined by the RRBS method, and transcriptomes sequenced by RNA-seq. Observed differences failed to explain the discordance between the twins. The search for causative changes revealed open methodological problems that need to be addressed at both the experimental and algorithmic levels (Box 2).

Evidence for aging-associated epigenomic changes has been accumulating [74] [70, 75] [76] [77] [78]. Several recent studies on humans [79] [80] and mouse [81] identified CpG islands associated with “bivalent” histone marks and key developmental regulators (genes typically silenced in stem cells by the Polycomb/Trithorax system and not by CpG island methylation) as being preferentially methylated with aging. The methylation of CpG islands silences developmental regulators, thus conferring stem-cell like properties to both aging and cancer cells. The patterns of CpG island methylation are therefore consistent with aging-associated cancer risk. Mapping of methylation patterns in the CpG-poor fraction of the genome [82] revealed even more pervasive aging-associated tissue-dependent patterns of hypomethylation affecting expression of metabolism-related genes, pointing to yet another connection between epigenomic changes and aging-associated pathophysiology.

In summary the pervasive aging-associated methylation changes discovered by the first comprehensive epigenome comparisons between MZ twins discordant for a human disease may help explain aging-associated disease risk. With the lowering of experimental and bioinformatic barriers, many more such studies are soon likely to provide insights into epigenomic connections between aging, environment, and human health.

Bioinformatic frameworks for epigenome analysis

As evident from the results reviewed so far, epigenome mapping and analysis is picking up pace. A number of large collaborative projects [17], including the NIH Roadmap Epigenomics Initiative [6] are contributing toward a map of epigenomic variation (Table 2). The International Human Epigenome Consortium (<http://ihc-epigenomes.org>) has been organized with the goal to coordinate and provide standards for international epigenome mapping efforts. Tools for epigenome analysis and visualization are becoming available and have recently been reviewed [83] [84]. In the following we briefly review three conceptual frameworks for epigenome analysis.

The significance of epigenomic variation is commonly evaluated based on its immediate effects on gene expression. But, the causative link may also go in the other direction – the epigenomic state may be a consequence of active transcription. Even in cases where

epigenomic state is causative, the change of the epigenome may be separated in time from its effect on gene expression. Also, the final effects may depend not on a single change but on a combination of epigenomic changes. Moreover, the effects of an epigenomic change, say during development, may be contingent on future intrinsic or environmental inputs. Based on these and additional theoretical considerations, criteria for biologically meaningful functional “codes” have been proposed [85]. One recently discovered combinatorial chromatin state that meets the proposed criteria is the “bivalent” or “poised” state consisting of the H3K27me/H3K4me combination [48]. This state is inherited by daughter cells during mitosis and makes genes responsive to certain intrinsic or environmental stimuli during cellular differentiation. Newly discovered biologically meaningful recurrent combinations of histone marks that may belong to an epigenomic “code” include signatures of active and inactive core and proximal promoters [56]; active and inactive enhancers [86] [87] [88]; gene-body signatures of active transcription [89]; gene regulation by alternative promoters [56] and alternative splicing [24] as well as gene silencing within larger heterochromatic domains conserved across cell types [90]. A “code-book” of fifty one epigenomic states, each state defined by a combination of epigenomic marks, has been inferred using an unsupervised learning approach based on the Hidden Markov Model approach [91]. Similar codes can also be inferred using a Bayes Network approach [92]. Using such “codes”, epigenomes can be “parsed” into segments corresponding to individually defined and biologically meaningful states.

The second framework is comparative analysis of epigenomes. This framework will gain importance with the increasing density of the sampling of the epigenomic variation space by the growing number of epigenomic projects. A key requirement for “meta-comparison” of epigenomes produced by different projects will be adherence to common data and metadata standards and the development of an informatic infrastructure for comparative epigenome analysis [93]. A number of issues of relevance for the comparative analysis of epigenomes are listed in Box 2.

The third framework that is yet to be tested on epigenomic data is systems modeling. One class of promising models, already used in tissue engineering, are cellular automata [94, 95]. Equipped with programmable epigenomes that are responsive to extra-cellular stimuli—which are currently lacking in these models—cellular automata may help bridge the chasm between the microcosm of the cell and the macrocosm of tissues and higher levels of organization of multi-cellular organisms.

Concluding remarks: Epigenome as the working memory of the cell

In summary, the emerging patterns of epigenomic variation are revealing epigenomic “codes” and providing insights through comparative analysis across a broad range of fields. With predictable technological advances over the coming years, patterns of epigenomic variation will be mapped at an ever accelerating rate. But will epigenomics live up to current high expectations?

Reasons for optimism are suggested by the central role of the epigenome in a cell-as-computer analogy. By combining the attributes of stability and programmability, the epigenome corresponds to computer’s working memory, which typically resides in computer RAM and includes the complete record of all active processes. In this key role, epigenome bridges the gap between the “read-only memory” or “CPU firmware” of genomic DNA and the transient or “memoryless” intra- and inter-cellular signaling processes. Cell’s “working memory” includes both the cell-type specific “program” that guides cell’s behavior and a record of key biological processes and environmental influences. It is therefore reasonable to

expect that reading and interpreting epigenomes will help close knowledge gaps across a broad spectrum of biology.

Acknowledgments

The author thanks Dr. R. Alan Harris for thoughtful comments and corrections and acknowledges support from the NIH Roadmap Epigenomics Roadmap Initiative grant U01 DA025956.

Glossary

Allelic imbalance	The epigenomic state of maternal and paternal chromosomes within a nucleus of a diploid organism may differ. An epigenomic difference in a specific locus is referred to as allelic imbalance. An allelic imbalance is experimentally detectable if it spans a heterozygous locus (such as a single nucleotide polymorphism) and if the readout technology (such as sequencing) provides allelic information. This allelic information allows separation of an epigenomic signal over a heterozygous locus into signals residing on maternal and paternal chromosomes. An allelic imbalance may be binary (on/off) or quantitative, where the signals from both the maternal and paternal chromosomes are greater than zero but of different magnitude
Cellular automata	An early model of computing popularized by Conway's Game of Life [96] and useful for modeling inter-cellular interactions and extra-cellular signaling during tissue growth and remodeling [94] [95]
CpG island	No standard definition of CpG islands has been agreed on. The definition usually takes the following form: "an X-bp stretch of DNA with a C+G content of Y% and an observed CpG/expected CpG in excess of Z". An early definition [97] suggests X=200, Y=50 and Z=0.6. A revision to X=500, Y=55, Z=0.65 has been proposed more recently [98]. CpG islands occur in the promoter regions of close to half of human genes and tend not to be methylated across tissues
Epigenome comparison	Refers to either (1) a pairwise comparison of two epigenomes to determine loci with specific differences and similarities in any of the epigenomic components (Box 1); (2) comparison of an individual epigenome against a database such as the Human Epigenome Atlas to determine the most similar epigenomes; (3) or any multi-way comparison such as clustering of epigenomes to determine their mutual relationship
Human Epigenome Atlas	A compendium of reference epigenomes produced by the NIH Epigenomics Roadmap Initiative (http://www.roadmapepigenomics.org , see Table 2 for more details) suitable for integrative analysis, visualization, and epigenome comparison. A schematic representation of Atlas contents is in Figure 1. The Atlas is accessible at (http://www.epigenomeatlas.org)
Polycomb/Trithorax system	The system consists of Polycomb group (PcG) and trithorax (trxG) group proteins, critical regulators of numerous developmental genes. The trxG has been associated with gene activation and the H3K4me3 mark and PcG with gene repression and the H3K27me3 mark [49]. The Polycomb/Trithorax system is conserved across the animal kingdom. The patterns of histone marks maintained by the Polycomb/Trithorax act as a kind of

cellular GPS system specifying the anterior/posterior and proximal/distal coordinates of the cell within the body [99]

**Structural
variant**

Denotes a genomic variant that is larger than about ten thousand basepairs. Includes duplications, deletions, inversions, interchromosomal translocations and other interchromosomal rearrangements

References

1. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet.* 2003; 33(Suppl):245–54. [PubMed: 12610534]
2. Bird A. Perceptions of epigenetics. *Nature.* 2007; 447(7143):396–8. [PubMed: 17522671]
3. Berger SL, et al. An operational definition of epigenetics. *Genes Dev.* 2009; 23(7):781–3. [PubMed: 19339683]
4. Bonasio R, Tu S, Reinberg D. Molecular signals of epigenetic states. *Science.* 2010; 330(6004):612–6. [PubMed: 21030644]
5. Suzuki MM, Bird A. DNA methylation landscapes: provocative insights from epigenomics. *Nat Rev Genet.* 2008; 9(6):465–76. [PubMed: 18463664]
6. Bernstein BE, et al. The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol.* 2010; 28(10):1045–8. [PubMed: 20944595]
7. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev.* 2002; 16(1):6–21. [PubMed: 11782440]
8. Beck S. Taking the measure of the methylome. *Nat Biotechnol.* 2010; 28(10):1026–8. [PubMed: 20944589]
9. Woodcock CL. Chromatin architecture. *Curr Opin Struct Biol.* 2006; 16(2):213–20. [PubMed: 16540311]
10. Kouzarides T. Chromatin modifications and their function. *Cell.* 2007; 128(4):693–705. [PubMed: 17320507]
11. Bourc'his D, Voinnet O. A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science.* 2010; 330(6004):617–22. [PubMed: 21030645]
12. Guttman M, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature.* 2009; 458(7235):223–7. [PubMed: 19182780]
13. Bernstein BE, Meissner A, Lander ES. The mammalian epigenome. *Cell.* 2007; 128(4):669–81. [PubMed: 17320505]
14. Feng S, Jacobsen SE, Reik W. Epigenetic reprogramming in plant and animal development. *Science.* 2010; 330(6004):622–7. [PubMed: 21030646]
15. Espada J, Esteller M. DNA methylation and the functional organization of the nuclear compartment. *Semin Cell Dev Biol.* 2010; 21(2):238–46. [PubMed: 19892028]
16. Ikegami K, et al. Interplay between DNA methylation, histone modification and chromatin remodeling in stem cells and during development. *Int J Dev Biol.* 2009; 53(2–3):203–14. [PubMed: 19412882]
17. Satterlee JS, Schubeler D, Ng HH. Tackling the epigenome: challenges and opportunities for collaboration. *Nat Biotechnol.* 2010; 28(10):1039–44. [PubMed: 20944594]
18. Allis, CD., et al. *Epigenetics.* Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press; 2007.
19. Zemach A, Zilberman D. Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr Biol.* 2010; 20(17):R780–5. [PubMed: 20833323]
20. Zemach A, et al. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science.* 2010; 328(5980):916–9. [PubMed: 20395474]
21. Feng S, et al. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A.* 2010; 107(19):8689–94. [PubMed: 20395551]

22. Kucharski R, et al. Nutritional control of reproductive status in honeybees via DNA methylation. *Science*. 2008; 319(5871):1827–30. [PubMed: 18339900]
23. Fox-Walsh K, Fu XD. Chromatin: the final frontier in splicing regulation? *Dev Cell*. 2010; 18(3):336–8. [PubMed: 20230741]
24. Luco RF, et al. Regulation of alternative splicing by histone modifications. *Science*. 2010; 327(5968):996–1000. [PubMed: 20133523]
25. Eckhardt F, et al. DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat Genet*. 2006; 38(12):1378–85. [PubMed: 17072317]
26. Bernstein BE, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*. 2005; 120(2):169–81. [PubMed: 15680324]
27. Ryba T, et al. Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res*. 2010; 20(6):761–70. [PubMed: 20430782]
28. Enard W, Paabo S. Comparative primate genomics. *Annu Rev Genomics Hum Genet*. 2004; 5:351–78. [PubMed: 15485353]
29. Zoghbi, HY.; Beaudet, AL. Epigenetics and Human Disease. In: Allis, CD., et al., editors. *Epigenetics*. Cold Spring Harbor laboratory Press; Cold Spring Harbor, New York: 2007. p. 435–456.
30. Gimelbrant A, et al. Widespread monoallelic expression on human autosomes. *Science*. 2007; 318(5853):1136–40. [PubMed: 18006746]
31. Kerkel K, et al. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet*. 2008; 40(7):904–8. [PubMed: 18568024]
32. Schalkwyk LC, et al. Allelic skewing of DNA methylation is widespread across the genome. *Am J Hum Genet*. 2010; 86(2):196–212. [PubMed: 20159110]
33. Li Y, et al. The DNA methylome of human peripheral blood mononuclear cells. *PLoS Biol*. 2010; 8(11):e1000533. [PubMed: 21085693]
34. Tycko B. Mapping allele-specific DNA methylation: a new tool for maximizing information from GWAS. *Am J Hum Genet*. 2010; 86(2):109–12. [PubMed: 20159108]
35. McDaniell R, et al. Heritable individual-specific and allele-specific chromatin signatures in humans. *Science*. 2010; 328(5975):235–9. [PubMed: 20299549]
36. Dindot SV, et al. Epigenetic profiling at mouse imprinted gene clusters reveals novel epigenetic and genetic features at differentially methylated regions. *Genome Res*. 2009; 19(8):1374–83. [PubMed: 19542493]
37. Harris RA, et al. Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications. *Nat Biotechnol*. 2010; 28(10):1097–105. [PubMed: 20852635]
38. Cahan P, et al. The impact of copy number variation on local gene expression in mouse hematopoietic stem and progenitor cells. *Nat Genet*. 2009; 41(4):430–7. [PubMed: 19270704]
39. Wan LB, Bartolomei MS. Regulation of imprinting in clusters: noncoding RNAs versus insulators. *Adv Genet*. 2008; 61:207–23. [PubMed: 18282507]
40. Lewis A, Reik W. How imprinting centres work. *Cytogenet Genome Res*. 2006; 113(1–4):81–9. [PubMed: 16575166]
41. Gregg C, et al. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science*. 2010; 329(5992):643–8. [PubMed: 20616232]
42. Gregg C, et al. Sex-specific parent-of-origin allelic expression in the mouse brain. *Science*. 2010; 329(5992):682–5. [PubMed: 20616234]
43. Abdalla H, Yoshizawa Y, Hochi S. Active demethylation of paternal genome in mammalian zygotes. *J Reprod Dev*. 2009; 55(4):356–60. [PubMed: 19721335]
44. Popp C, et al. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature*. 2010; 463(7284):1101–5. [PubMed: 20098412]
45. Hajkova P, et al. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science*. 2010; 329(5987):78–82. [PubMed: 20595612]

46. Lister R, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature*. 2009; 462(7271):315–22. [PubMed: 19829295]
47. Laurent L, et al. Dynamic changes in the human methylome during differentiation. *Genome Res*. 2010; 20(3):320–31. [PubMed: 20133333]
48. Bernstein BE, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006; 125(2):315–26. [PubMed: 16630819]
49. Schuettengruber B, et al. Genome regulation by polycomb and trithorax proteins. *Cell*. 2007; 128(4):735–45. [PubMed: 17320510]
50. Illingworth RS, Bird AP. CpG islands—‘a rough guide’. *FEBS Lett*. 2009; 583(11):1713–20. [PubMed: 19376112]
51. Jones PA, Baylin SB. The epigenomics of cancer. *Cell*. 2007; 128(4):683–92. [PubMed: 17320506]
52. Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol*. 2010; 28(10):1057–68. [PubMed: 20944598]
53. Hawkins RD, et al. Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell*. 2010; 6(5):479–91. [PubMed: 20452322]
54. van Arensbergen J, et al. Derepression of Polycomb targets during pancreatic organogenesis allows insulin-producing beta-cells to adopt a neural gene activity program. *Genome Res*. 2010; 20(6):722–32. [PubMed: 20395405]
55. Weber M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet*. 2007; 39(4):457–66. [PubMed: 17334365]
56. Yagi S, et al. DNA methylation profile of tissue-dependent and differentially methylated regions (T-DMRs) in mouse promoter regions demonstrating tissue-specific gene expression. *Genome Res*. 2008; 18(12):1969–78. [PubMed: 18971312]
57. Brunner AL, et al. Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver. *Genome Res*. 2009; 19(6):1044–56. [PubMed: 19273619]
58. Ji H, et al. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature*. 2010; 467(7313):338–42. [PubMed: 20720541]
59. Schmidl C, et al. Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res*. 2009; 19(7):1165–74. [PubMed: 19494038]
60. Marson A, et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell*. 2008; 134(3):521–33. [PubMed: 18692474]
61. Mohammad HP, Baylin SB. Linking cell signaling and the epigenetic machinery. *Nat Biotechnol*. 2010; 28(10):1033–8. [PubMed: 20944593]
62. Cuddapah S, et al. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res*. 2009; 19(1):24–32. [PubMed: 19056695]
63. Eeckhoutte J, et al. Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers. *Genome Res*. 2009; 19(3):372–80. [PubMed: 19129543]
64. Cheng Y, et al. Erythroid GATA1 function revealed by genome-wide analysis of transcription factor occupancy, histone modifications, and mRNA expression. *Genome Res*. 2009; 19(12):2172–84. [PubMed: 19887574]
65. Hsu PY, et al. Estrogen-mediated epigenetic repression of large chromosomal regions through DNA looping. *Genome Res*. 2010; 20(6):733–44. [PubMed: 20442245]
66. Weaver IC, et al. Epigenetic programming by maternal behavior. *Nat Neurosci*. 2004; 7(8):847–54. [PubMed: 15220929]
67. Jirtle RL, Skinner MK. Environmental epigenomics and disease susceptibility. *Nat Rev Genet*. 2007; 8(4):253–62. [PubMed: 17363974]
68. Tobi EW, et al. DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Hum Mol Genet*. 2009; 18(21):4046–53. [PubMed: 19656776]
69. Daxinger L, Whitelaw E. Transgenerational epigenetic inheritance: more questions than answers. *Genome Res*. 2010; 20(12):1623–8. [PubMed: 21041414]

70. Fraga MF, et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A*. 2005; 102(30):10604–9. [PubMed: 16009939]
71. Petronis A. The origin of schizophrenia: genetic thesis, epigenetic antithesis, and resolving synthesis. *Biol Psychiatry*. 2004; 55(10):965–70. [PubMed: 15121478]
72. Petronis A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature*. 2010; 465(7299):721–7. [PubMed: 20535201]
73. Baranzini SE, et al. Genome, epigenome and RNA sequences of monozygotic twins discordant for multiple sclerosis. *Nature*. 2010; 464(7293):1351–6. [PubMed: 20428171]
74. Issa JP, et al. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat Genet*. 1994; 7(4):536–40. [PubMed: 7951326]
75. Kwabi-Addo B, et al. Age-related DNA methylation changes in normal human prostate tissues. *Clin Cancer Res*. 2007; 13(13):3796–802. [PubMed: 17606710]
76. Bjornsson HT, et al. Intra-individual change over time in DNA methylation with familial clustering. *JAMA*. 2008; 299(24):2877–83. [PubMed: 18577732]
77. Boks MP, et al. The relationship of DNA methylation with age, gender and genotype in twins and healthy controls. *PLoS One*. 2009; 4(8):e6767. [PubMed: 19774229]
78. Christensen BC, et al. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet*. 2009; 5(8):e1000602. [PubMed: 19680444]
79. Rakyan VK, et al. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res*. 2010; 20(4):434–9. [PubMed: 20219945]
80. Teschendorff AE, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res*. 2010; 20(4):440–6. [PubMed: 20219944]
81. Maegawa S, et al. Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res*. 2010; 20(3):332–40. [PubMed: 20107151]
82. Thompson RF, et al. Tissue-specific dysregulation of DNA methylation in aging. *Aging Cell*. 2010; 9(4):506–18. [PubMed: 20497131]
83. Bock C, Lengauer T. Computational epigenetics. *Bioinformatics*. 2008; 24(1):1–10. [PubMed: 18024971]
84. Lim SJ, Tan TW, Tong JC. Computational Epigenetics: the new scientific paradigm. *Bioinformation*. 2010; 4(7):331–7. [PubMed: 20978607]
85. Turner BM. Defining an epigenetic code. *Nat Cell Biol*. 2007; 9(1):2–6. [PubMed: 17199124]
86. Heintzman ND, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*. 2007; 39(3):311–8. [PubMed: 17277777]
87. Koch CM, et al. The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res*. 2007; 17(6):691–707. [PubMed: 17567990]
88. Hoffman BG, et al. Locus co-occupancy, nucleosome positioning, and H3K4me1 regulate the functionality of FOXA2-, HNF4A-, and PDX1-bound loci in islets and liver. *Genome Res*. 2010; 20(8):1037–51. [PubMed: 20551221]
89. Mikkelsen TS, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*. 2007; 448(7153):553–60. [PubMed: 17603471]
90. Thurman RE, et al. Identification of higher-order functional domains in the human ENCODE regions. *Genome Res*. 2007; 17(6):917–27. [PubMed: 17568007]
91. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol*. 2010; 28(8):817–25. [PubMed: 20657582]
92. van Steensel B, et al. Bayesian network analysis of targeting interactions in chromatin. *Genome Res*. 2010; 20(2):190–200. [PubMed: 20007327]
93. Milosavljevic A. Putting epigenome comparison into practice. *Nat Biotechnol*. 2010; 28(10):1053–6. [PubMed: 20944597]
94. Cowin SC. Tissue growth and remodeling. *Annu Rev Biomed Eng*. 2004; 6:77–107. [PubMed: 15255763]
95. Peirce SM, Van Gieson EJ, Skalak TC. Multicellular simulation predicts microvascular patterning and in silico tissue assembly. *FASEB J*. 2004; 18(6):731–3. [PubMed: 14766791]

96. Mitchell, M. Complexity : a guided tour. Vol. xvi. Oxford England; New York: Oxford University Press; 2009. p. 349
97. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. *J Mol Biol.* 1987; 196(2): 261–82. [PubMed: 3656447]
98. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci U S A.* 2002; 99(6):3740–5. [PubMed: 11891299]
99. Chang HY. Anatomic demarcation of cells: genes to patterns. *Science.* 2009; 326(5957):1206–7. [PubMed: 19965461]
100. Chen LL, Carmichael GG. Decoding the function of nuclear long non-coding RNAs. *Curr Opin Cell Biol.* 2010; 22(3):357–64. [PubMed: 20356723]
101. Hekimoglu B, Ringrose L. Non-coding RNAs in polycomb/trithorax regulation. *RNA Biol.* 2009; 6(2):129–37. [PubMed: 19270511]
102. John S, et al. Interaction of the glucocorticoid receptor with the chromatin landscape. *Mol Cell.* 2008; 29(5):611–24. [PubMed: 18342607]
103. Bock C, et al. Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol.* 2010; 28(10):1106–14. [PubMed: 20852634]
104. Coarfa C, et al. Pash 3.0: A versatile software package for read mapping and integrative analysis of genomic and epigenomic variation using massively parallel DNA sequencing. *BMC Bioinformatics.* 2010; 11:572. [PubMed: 21092284]
105. Polikepahad S, et al. Proinflammatory role for let-7 microRNAs in experimental asthma. *J Biol Chem.* 2010; 285(39):30139–49. [PubMed: 20630862]
106. Robertson G, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat Methods.* 2007; 4(8):651–7. [PubMed: 17558387]
107. Hesselberth JR, et al. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. *Nat Methods.* 2009; 6(4):283–9. [PubMed: 19305407]
108. Miele A, et al. Mapping chromatin interactions by chromosome conformation capture. *Curr Protoc Mol Biol.* 2006; Chapter 21(Unit 21):11. [PubMed: 18265379]
109. Rakyan VK, et al. An integrated resource for genome-wide identification and analysis of human tissue-specific differentially methylated regions (tDMRs). *Genome Res.* 2008; 18(9):1518–29. [PubMed: 18577705]
110. Flicek P, et al. Ensembl 2011. *Nucleic Acids Res.* 2011; 39(Database issue):D800–6. [PubMed: 21045057]
111. Fingerman IM, et al. NCBI Epigenomics: a new public resource for exploring epigenomic data sets. *Nucleic Acids Res.* 2011; 39(Database issue):D908–12. [PubMed: 21075792]
112. Raney BJ, et al. ENCODE whole-genome data in the UCSC genome browser (2011 update). *Nucleic Acids Res.* 2011; 39(Database issue):D871–5. [PubMed: 21037257]

Figure 1.

A schematic illustration of the space of epigenomic variation. The displayed DNA methylation and histone marks (from [53]) were remapped on the hg19 reference genome assembly and are available for browsing and download from the Human Epigenome Atlas Release 2 site (accessible at <http://www.epigenomeatlas.org> and other sites). Key developmental regulators *HOXA1- HOXA13* occur in left-to right order in the *HOXA* cluster. The cluster is in a “poised” state as indicated by the combination of active (H3K4me3) and inactive (H3K27me3) marks. Increase in active and decrease in inactive marks over *HOXA1- HOXA9* genes occurs in fibroblasts. Parallel changes in methylation can also be observed.

Table 1

Epigenomic data layers and assays with sequencing readouts

Data layer	Assay	Assay description	References
Genomic DNA methylation	MeDIP-seq	Methylated DNA immuno-precipitation and sequencing. Methylated cytosines in sheared DNA are immuno- precipitated with a methylcytosine antibody. The immuno-precipitated DNA is sequenced to reveal extent of methylation genome-wide at about 200bp resolution level.	[37]
	MBD-seq	Similar to MeDIP-seq except that methylated cytosines are recognized by the methyl-binding domain (MBD) of biotin tagged MBD2 protein prior to fractionation by affinity binding to streptavidin beads. The sensitivity of MeDIP-seq and MBD-seq are similar.	[37]
	MRE-seq	DNA fragments resulting from digestions of genomic DNA by methylation-sensitive restriction enzymes (MREs) HpaII (C ⁺ CGG), HinfI (G ⁺ CGC) and AclI (C ⁺ CGC) are end-sequenced. The end-sequences from each digest are mapped onto a reference genome to identify unmethylated CpGs.	[37]
	MethylC-seq	Whole-genome sequencing of bisulfite- treated genomic DNA to identify methylation status of cytosines. The bisulfite treatment converts unmethylated cytosines into uracil, which is ultimately converted into thymine. Methylated cytosines are protected from bisulfite conversion. Bisulfite-treated DNA is sequenced and the reads are mapped onto unique genomic positions in a reference sequence. Ratio of cytosines and thymines in resulting sequences indicates degree of methylation.	[46]
	RRBS-seq	“Reduced representation” bisulfite sequencing is a variation on the Methyl-C-seq protocol with an additional complexity reduction step. Specifically, prior to MethylC-seq, genomic DNA is first subjected to digestion using a CpG methylation insensitive enzyme (typically MspI) and size-selection. This method provides information about methylation levels in the CG-rich fraction of the genome.	[37, 103]
Chromatin- associated RNA	RNA-seq	Sequencing of RNA species including both small RNA and long RNAs with various roles in chromatin regulation.	[105]
DNA-bound protein and its modifications	ChIP-seq	Immuno-precipitation of cross-linked and sheared chromatin. DNA fragments cross-linked to immuno- precipitated protein are sequenced to reveal localization of specific proteins genome-wide. Applications include mapping of transcription factor binding sites and of modified histones positions.	[106]
Chromatin accessibility	DNase- sensitivity- seq	Involves sequencing of short DNA fragments produced by DNase I cleavage of intact genomic DNA. Sequenced fragments are mapped onto reference genomic sequences. High density of mapped fragments at a specific position indicates lack of protection of genomic DNA from cleavage by bound protein.	[107]
Spatial organization of chromatin	3C-seq	Chromosome conformation capture. Chromatin is cross-linked and sheared. Inter- and intra-chromosomal proximity of DNA strands is detected by their ligation and sequencing of ligation products to detect intra- and inter- chromosomal chromatin interactions. More recent variations on 3C include 4C and 5C methods.	[108]

Table 2

Sources of epigenomic information

Epigenomic Data Resource	Description	References
Ensembl http://www.ensembl.org	Ensembl uptakes the epigenomic data tracks from ENCODE and the NIH Roadmap Epigenomics Project as well as smaller hypothesis-driven projects. Hundreds of tracks are available, primarily from ChIP-seq and DNase-seq assays for human and mouse cell types. Ensembl also contains a resource for identification of human tissue- specific differentially methylated regions [109].	[110]
NCBI epigenomics portal http://www.ncbi.nlm.nih.gov/epigenomics	Curated, annotated and organized epigenetics- specific data selected from general-purpose archives, such as the Gene Expression Omnibus, and Sequence Read Archives, including the data from the NIH Roadmap Epigenomics Project. Includes 337 samples and over 1100 data tracks from five well-studied species.	[111]
NIH Roadmap Epigenomics Project http://www.roadmapepigenomics.org Human Epigenome Atlas http://www.epigenomeatlas.org	The largest current epigenomics project with an exclusive focus on producing a comprehensive catalogue of human epigenomic maps, including methylomes, histone marks, chromatin accessibility and RNA-seq data for hundreds of human cell types. The epigenome maps produced by the NIH Roadmap and other projects are being integrated into the Human Epigenome Atlas.	[6]
UCSC genome browse http://www.genome.ucsc.edu	A collection of high- throughput, genome-wide data generated on a panel of human cell lines by the ENCODE project including histone marks, chromatin accessibility, DNA methylation, RNA expression, RNA binding and other cell-state indicators.	[112]