Epigenome Mapping in Normal and Disease States

Arika K. Maunakea, Iouri Chepelev, Keji Zhao

Abstract: Epigenomes are comprised, in part, of all genome-wide chromatin modifications, including DNA methylation and histone modifications. Unlike the genome, epigenomes are dynamic during development and differentiation to establish and maintain cell type–specific gene expression states that underlie cellular identity and function. Chromatin modifications are particularly labile, providing a mechanism for organisms to respond and adapt to environmental cues. Results from studies in animal models clearly demonstrate that epigenomic variability leads to phenotypic variability, including susceptibility to disease that is not recognized at the DNA sequence level. Thus, capturing epigenomic information is invaluable for comprehensively understanding development, differentiation, and disease. Herein, we provide a brief overview of epigenetic processes, how they are relevant to human health, and review studies using technologies that enable epigenome mapping. We conclude by describing feasible applications of epigenome mapping, focusing on epigenome-wide association studies (eGWAS), which have the potential to revolutionize current studies of human diseases and will likely promote the discovery of novel diagnostic, preventative, and treatment strategies. (Circ Res. 2010;107:327-339.)

Key Words: epigenetics ■ genetics ■ gene expression ■ gene regulation

In contrast to the genome, which remains largely unchanged in most cells of an organism, epigenomes are the product of a gradual commitment of cell lineages to more constrained patterns of gene expression throughout development that is, in part, shaped by the environment. An epigenome can be defined as the combination of all genome-wide chromatin modifications in any given cell type that directs its unique gene expression pattern. Epigenomes are labile during development, are responsive to extrinsic factors, and even differ among individuals with identical genetic composition.

The principal chromatin modifications include DNA methylation and posttranslational modifications of histones that package the DNA in nucleosomal units. Although both DNA methylation and histone modifications appear to be meiotically and/or mitotically heritable, only the former epigenetic process is backed with strong mechanistic support for heritability. As we discuss in this review, accumulating evidence from mammalian studies indicate that variability of epigenetic modifications of chromatin during development and in response to distinct environmental factors directly contribute to adult phenotypic variability and disease susceptibility that could not previously be accounted for by DNA sequence alone. Thus, characterizing genome-wide chromatin modification patterns (ie, epigenome mapping) may aid in the discovery of disease-causing genes in humans for which nongenetic factors are clearly involved and confound DNA sequence–based association studies of disease. Epigenome mapping across different cell types and developmental periods from normal individuals, in a manner analogous to the efforts of sequencing the human genome, is an essential prerequisite.

Technological advances allowed characterization of the first human epigenome in CD4+ T cells, which describes 38
histone modifications, including both histone methylation and acetylation.\textsuperscript{8–10} Similar technologies have now enabled truly genome-wide analyses of DNA methylation.\textsuperscript{11,12} Taking advantage of these and other technologies, the recently launched International Human Epigenome Consortium aims to map 1000 reference epigenomes within a decade.\textsuperscript{13} Herein, we provide a brief overview of epigenetic processes and how they are relevant to human health and review initial studies using epigenome mapping. In particular, we discuss the evidence supporting that epigenetic processes offer a mechanistic link between genetic determinants and environment during development, and we address how results of epigenome mapping might be applied to measure previously unobservable potential interindividual variability that could account for differences to disease susceptibility.

**Epigenetic Processes**

In general, epigenetic processes including DNA methylation and histone modifications are thought to modulate the accessibility of cis-DNA elements to trans-acting factors via regulating chromatin structure.\textsuperscript{14} In plants, RNA interference contributes to epigenetic regulation; however, whether there is a conserved mammalian equivalent remains unclear\textsuperscript{15} and will not be further discussed here. Similarly, there is considerable debate of whether chromatin remodeling\textsuperscript{1} should be considered an epigenetic mechanism because it is unknown whether this process can transmit memory of cell fate from one generation to the next.

Epigenetic regulation of chromatin structure consequently influences gene expression. This type of control is exemplified by the phenomenon of genomic imprinting, the monoallelic expression of genes in a parent-of-origin–dependent manner. Critical cis-elements that regulate expression of imprinted genes called imprinting control regions exhibit parental-specific chromatin modifications including DNA methylation and histone modifications that govern their activity.\textsuperscript{16} Similarly, as a form of dosage compensation between female and male eutherians, X-chromosome inactivation (XCI), mediated primarily by Xist and Tsix genes, is accompanied by chromosome-specific histone modifications associated with heterochromatin and gene silencing at the 2-cell stage of early embryonic development.\textsuperscript{17} Once established, XCI is maintained by DNA methylation–mediated gene silencing in cells of the embryo proper. Because genomic imprinting and XCI have been reviewed extensively elsewhere,\textsuperscript{16,17} we focus on the global distribution and functional importance of DNA methylation and histone modifications as they relate to cellular phenotype.

**DNA Methylation**

Perhaps the best understood epigenetic mechanism is DNA methylation, which in mammals occurs almost exclusively within 5′-cytosine–guanine-3′ dinucleotides (CpGs), although CpNpG methylation has also been detected.\textsuperscript{18} In general, DNA methylation is typically associated with gene silencing by affecting the binding of methylation-sensitive DNA binding proteins and/or by interacting with various modifications of histone proteins that alter DNA accessibility to promoters.\textsuperscript{19} Once established by the de novo methyltransferases DNA methyltransferase (DNMT)\textsuperscript{3a and DNMT3b},\textsuperscript{20} DNA methylation is maintained through mitosis primarily by the DNMT1 enzyme, which associates with PCNA and the replication foci and has a significant preference for action on hemi-methylated DNA following DNA replication.\textsuperscript{21} This mechanism allows for perpetuation of the DNA methylation state in newly formed cells, a well-established mode of epigenetic inheritance.

Several studies suggest multiple functional roles for DNA methylation, including silencing transposable elements,\textsuperscript{22} mediating developmental gene regulation,\textsuperscript{23} and reducing transcriptional noise.\textsuperscript{24} Indeed, DNA methylation in mammals is essential for embryonic development,\textsuperscript{25} differentiation,\textsuperscript{26} and cell cycle control.\textsuperscript{27} Additionally, DNA methylation plays critical roles in maintaining transcriptional silencing of genes on the inactive X-chromosome and imprinted genes.\textsuperscript{28} Human diseases have been associated with abnormal DNA methylation patterns, including cancer,\textsuperscript{29} ICF (immunodeficiency, centromeric instability, facial anomalies) syndrome, ATRX (alpha-thalassemia, mental retardation) syndrome, and fragile X syndrome.\textsuperscript{30} Even mutations in a gene that encodes a protein that “reads” DNA methylation signals and thought to act as a transcriptional silencer, MeCP2, are associated with the autism spectrum disorder, Rett syndrome.\textsuperscript{31} In plants, DNA methylation deficiency results in spurious transcription initiation from cryptic sites, demonstrating a role for methylation in reducing transcriptional noise.\textsuperscript{24} A central theme of these findings is that DNA methylation functions to maintain a repressed chromatin state, thereby stably silencing promoter activity.\textsuperscript{32} It is important to note, however, that DNA methylation is not always associated with gene silencing. Some studies have demonstrated that DNA methylation can augment expression of an imprinted gene by blocking the binding of repressor proteins to silenced elements within the gene.\textsuperscript{33} Furthermore, additional novel roles of DNA methylation have been suggested by recent epigenome mapping studies (discussed below).

**Histone Modifications**

In addition to the covalent modification of DNA, histone proteins that package nDNA are known to be subjected to a
wide array of posttranslational modifications on specific residues along their NH$_2$-terminal “tails” that project outside of the nucleosome core. Such modifications include acetylation, methylation, phosphorylation, ubiquitination, sumoylation, and potentially others that remain to be discovered. Since these modifications contribute to chromatin conformation, the potential information stored within different combinatorial patterns of these modifications led to the hypothesis of a “histone code,” which proposes that specific combinations of these modifications dictate locus-specific transcriptional competence. Although the specific functional consequences of each of these modifications are insufficiently understood, many of the enzymes responsible for adding and removing them have been identified. In addition, alterations of the genes encoding histone-modifying enzymes and perturbations to the modification patterns they produce are associated with disease, underscoring the importance of histone modifications in normal development.

For example, the Polycomb-group protein EZH2, a histone H3 lysine 27 (H3K27) methyltransferase, was overexpressed in prostate cancer, whereas global loss of H4K16 acetylation and H4K20 trimethylation (H4K20me3) was observed in lymphoma and colon cancer. One of the best understood histone modification is the reversible acetylation of lysine residues on histones H3 and H4. Acetylation is accomplished by specific histone acetyltransferases that neutralizes the net positive charge of the histone tail and facilitates access of transcription factors to the underlying DNA. Whereas histone acetylation is clearly associated with transcriptional activation, the reverse reaction catalyzed by histone deacetylases (HDACs) increases the transcription potential of the underlying DNA, and is associated with a transcriptionally repressed state. Distinct from acetylation, methylation of histones can be associated with transcriptional activation or repression depending on the specific residue targeted and degree of methylation. For example, H3K4me3 is frequently enriched at gene promoters, many of which are active. Localization of H3K4me3 at inactive promoters, however, facilitates the transient binding of histone acetyltransferases and histone deacetylases to maintain the promoters in a repressed but poised state for future activation. How histone methylation accomplishes gene repression is not fully understood; however, it is known that methylation of certain lysine residues, such as H3K9, acts as a docking site for heterochromatin protein (HP1), which in turn recruits histone methyltransferases. Other modifications are not restricted to promoters, including the “activating” modification H3K36 which is distributed over the gene bodies of actively expressed genes. Although histone phosphorylation, ubiquitylation, and sumoylation of specific serine, threonine, or lysine residues are involved in transcriptional regulation, modifications on other residues are associated with mitosis, DNA repair, and apoptosis. The precise functions of many of these modifications remain under investigation.

New evidence suggests the possibility that histone modifications could be mitotically heritable, although this is debated. In yeast for example, the heterochromatic states mediated by the interactions of hypoacetylated histones and silent-information-regulator proteins or H3K9 methylation and the Swi6 chromodomain are maintained through cell division. In flies, H3K27 and H3K4 methylation, catalyzed by Polycomb-group and trithorax-group protein complexes respectively, mediate mitotic inheritance of lineage-specific gene expression patterns. This is accomplished, in part, by EED, a constituent of the PRC2 (Polycomb repressive complex 2), which binds specifically to histone tails, activates the methyltransferase activity of PRC2, and facilitates propagation of the H3K27me3 mark on mitosis. The extent to which other histone modifications are heritable remains unclear. However, histone modifications along with DNA methylation undergo considerable changes during development when the epigenome is particularly vulnerable to environmental influences, as discussed later.

Epigenome Mapping

Recent technological advances in DNA sequencing, in part, have enabled epigenome mapping. Data derived from these technologies provide unprecedented insight into the distribution, interplay, and potential novel functions of chromatin modifications and associated proteins. Whereas other reviews explain these technologies in greater detail, we focus on the currently available methods that use next-generation sequencing to map epigenomes and discuss unexpected findings.

Detecting DNA Methylation Genome-Wide

Several strategies that allow detection of DNA methylation include digestion of DNA by methylation-sensitive or –insensitive restriction endonucleases, the chemical modification of DNA by sodium bisulfite, immunoprecipitation of 5-methylcytosine to separate unmethylated and methylated fractions of the genome, and enrichment of methylated DNA using DNA binding proteins. These strategies have been coupled to high-throughput technologies. Although the genomic coverage and resolution varies among the different strategies, each offers unique advantages. Thus, these approaches are complementary and would potentially generate mutually reinforcing and comprehensive genome-wide DNA methylation data.

Of these strategies, one that provides the best resolution of cytosine methylation incorporates sodium bisulfite treatment of DNA. Sodium bisulfite chemically converts all unmethylated cytosines to uracil by hydrolytic deamination of the 5,6-dihydrocytosine-6-sulfonate product at the C4 position of the pyrimidine ring. However, a methyl group at the C5 position inhibits this reaction, protecting methylated cytosines from conversion. After PCR amplification using primers corresponding to the bisulfite-converted sequences of interest, the products are traditionally cloned and sequenced to evaluate CpG methylation throughout the amplified region. Several applications exploiting this “gold-standard” bisulfite-conversion method have been adapted to evaluate methylation globally using next-generation sequencing technology including a shotgun-sequencing approach, a reduced representation shotgun-sequencing approach, and a targeted approach using bisulfite padlock probes. Bisulfite treatment of genomic DNA combined with ultrahigh-throughput sequencing using Illumina technology allows sensitive mea-
measure of cytosine methylation on a genome-wide scale within specific sequence contexts at single nucleotide resolution.\textsuperscript{66} This technology has been successfully applied to the \textit{Arabidopsis thaliana} genome,\textsuperscript{66,69} to the mouse genome,\textsuperscript{66} and, more recently, to the human genome,\textsuperscript{11,12} impressively accounting for \textgreater\textasciitilde90\% of all cytosines genome-wide. However, a major prohibitive disadvantage of this approach is the high cost because a very large amount of sequencing is required to yield comprehensive and accurate methylation data. To overcome this limitation, a semimarrow approach termed “reduced representation bisulfite sequencing” (RRBS) has been devised.\textsuperscript{67} In RRBS, genomic DNA is first digested with specific restriction enzymes whose recognition sequences are disproportionately represented in GC-rich DNA, such as \textit{BgII} or \textit{MspI}, fragments of a certain size range are then selected and subjected to adaptor ligation, treated with sodium bisulfite, PCR-amplified, and sequenced. Although this approach provides useful nucleotide-level quantitation of cytosine methylation, the selection of fragments in a given size range explicitly excludes entire genome coverage and introduces a bias toward CpG-rich regions of the genome that typically include unmethylated CpG islands in normal cells. However, RRBS has proven useful for cancer studies that require the rapid, high-throughput analysis of aberrant DNA methylation patterns in multiple samples.\textsuperscript{70} An alternative approach that eliminates the potential bias of methods targeting CpG islands yet offers nucleotide-level resolution involves the use of bisulfite padlock probes (BSPPs) to capture selected locations of the genome for methylation profiling.\textsuperscript{68} However, the genome-wide coverage of DNA methylation is limited to the number of bisulfite padlock probes used per reaction, thus precluding a complete analysis of DNA methylation throughout the genome.

Other strategies avoid the use of sodium bisulfite altogether and instead enrich for methylated DNA. One approach takes advantage of an antibody targeted to single methylated cytosines to directly capture methylated sequences within genomic DNA. Termed methylated DNA immunoprecipitation (MeDIP), this approach allows for relative quantitation of DNA methylation genome-wide when coupled to comparative genomic hybridization microarrays as was applied to human\textsuperscript{81} and plant\textsuperscript{71} samples. More recently, MeDIP has been adapted to high-throughput sequencing using Illumina technology (MeDIP-Seq),\textsuperscript{72} providing a more complete genome-wide coverage of DNA methylation than microarrays can achieve. Another way to enrich for methylated DNA is using a methyl-binding domain fused to human IgG\textsuperscript{3} or methyl-binding domain proteins bound to a sepharose matrix\textsuperscript{62} before hybridizing onto an array. Irrespective of the platforms used to measure output DNA methylation signals, limitations of these affinity approaches include the bias that methylated CpG-rich sequences give higher enrichments than equally methylated yet CpG-poor sequences, the inability to measure DNA methylation of individual repetitive elements, and the lower resolution of methylation detection relative to bisulfite-sequencing. To overcome these limitations, a novel approach called Methyl-MAPS (methylation mapping analysis by paired-end sequencing) has recently been developed that provides single nucleotide resolution of DNA methylation, covers \textgreater80\% of CpG sites within mammalian genomes, and enables quantitation of methylation at repetitive elements in addition to single-copy loci by combining enzymatic fractionation and deep sequencing.\textsuperscript{74}

These epigenome mapping techniques afforded an unprecedented opportunity to observe the distribution of DNA methylation over the majority of the genome leading to some surprising discoveries. The majority of gene promoters were depleted of DNA methylation, whereas there was markedly increased cytosine methylation within gene bodies. The large scope of DNA methylation within gene bodies, as observed for a third of all genes in the Arabidopsis genome\textsuperscript{66,69} and for a substantial number of genes in humans,\textsuperscript{11,12,68,75} suggests a potentially novel role for DNA methylation that seemingly contradicts the dominant model for DNA methylation–mediated gene silencing. Specifically, higher levels of DNA methylation were detected within gene bodies of actively expressed genes compared to silent genes. Suppression of spurious initiation of transcription within highly active genes,\textsuperscript{76} modulation of transcriptional elongation,\textsuperscript{77} regulation of pre-mRNA splicing,\textsuperscript{12} and tissue-specific alternative promoter usage\textsuperscript{77a} are potential novel roles of gene body methylation. That substantial gene body methylation is an evolutionarily conserved feature of eukaryotic genomes\textsuperscript{78–80} further supports a biological role and warrants continued investigation.

Another unexpected common finding of these genome-wide DNA methylation studies was that, in contrast to the long-held view that cytosines within CpG islands remain free of DNA methylation, a subset of CpG island-containing promoters were in fact methylated in normal cells.\textsuperscript{81–84} Consistent with previous gene-centric studies in cancer, this normal promoter methylation was, in general, negatively correlated with the expression state of the underlying gene. In studying normal embryonic stem cell differentiation, gain or loss of promoter methylation was observed concomitant with loss or gain of expression, respectively, of the associated gene,\textsuperscript{11,12} many of which have previously defined roles in embryonic stem cell function. Interestingly, it was also recognized that unlike somatic cells, embryonic stem cells harbored substantial levels of non-CpG cytosine methylation.\textsuperscript{11,12} Genome-wide, non-CpG methylation was diminished in human embryonic stem cell–differentiated fibroblasts\textsuperscript{12} and significantly gained in induced pluripotent embryonic-like stem cells generated from primary human fibroblasts,\textsuperscript{11} which together indicates a novel role of non-CpG methylation in the origin and maintenance of the pluripotent stem cell state.

Finally, a landmark study using a more limited genome-wide DNA methylation fingerprinting technique called AIMS (amplification of intermethylated sites), observed remarkable differences in the DNA methylation profiles of monozygotic twins, particularly those that were older, had different lifestyles, and had spent less of their lives together.\textsuperscript{3} These results implicate a probabilistic accumulation of epigenetic variability likely attributable to environmental differences experienced by these genetically identical individuals during their lifetime and may be involved in the etiology of monozygotic twin discordance for common diseases and traits.\textsuperscript{85}
Further support of this notion comes from a recent DNA methylation profiling study that uncovered epigenetic differences that might contribute to the discordance for the autoimmune disease systemic lupus erythematosus (SLE) among monozygotic twins.86

Additional epigenome mapping of DNA methylation profiles across a larger number of cell types and individuals, coupled with global gene expression data, should reveal a more comprehensive understanding of the prevalence of normal promoter methylation, non-CpG methylation, and interindivual variability to elucidate the epigenetic determinants of normal differentiation and disease susceptibility. Notably, the studies of DNA methylation described above focus exclusively on 5-methylcytosine. However, 5-hydroxymethylcytosine has recently been detected in mammalian DNA,87,88 yet its in vivo function remains unclear, in part, because current technologies, including the gold-standard bisulfite-sequencing approach, fail to distinguish between the 2 types of DNA modifications.89 MeDIP approaches using recently introduced anti-5-hydroxymethylcytosine antibodies may overcome this critical limitation.

Detecting Histone Modifications Genome-Wide
Most of the existing methods for studying histone modifications on a genomic scale combine the use of chromatin immunoprecipitation (ChIP) with high-throughput technologies including DNA microarrays (ChIP-chip) or massively parallel sequencing (ChIP-Seq). ChIP technique relies on the isolation of individual chromatin fragments using an antibody specific to a particular feature of the chromatin fragments, including DNA-binding proteins, histone modifications, and nucleosomes. Because these techniques have been reviewed elsewhere,56,90,91 we will only mention that applications of these techniques across multiple species including yeast,92–94 fly,95 mouse,96,97 and human44,98–100 have revealed that distinct genomic regions such as enhancers, promoters, and gene bodies have distinct histone modification patterns.8,9,101–104

Thus, mapping these modifications globally may provide a more precise functional annotation of genomes. In addition, detecting lesser-known histone modifications including phosphorylation and sumoylation using chemical approaches such as peptide synthesis, mutagenesis, and in vitro nucleosomal arrays could complement in vivo ChIP results in elucidating their function(s).105

Our epigenome mapping data of 38 histone modifications including methylation and acetylation using ChIP-Seq in normal primary human CD4+ T cells have revealed remarkably consistent categorical patterns across gene regions depending on the transcriptional activity of the underlying gene.8,9 For example, histone modifications across actively expressed genes can be separated into at least 4 categories based on their general patterns and may correspond to their distinct functions in transcription (Figure 1). “Active” histone modification marks highly enriched within gene promoters, for instance, may be involved in transcription initiation ( eg, H3K4me3), whereas those that are intragenic may be involved in elongation, termination, or pre-mRNA splicing ( eg, H3K36me3).8,9,46,106,107 Similarly, other histone modifications across silent genes adopt distinct patterns possibly impairing or preventing transcription initiation when enriched at promoters ( eg, H3K27me3) or disrupting elongation when enriched throughout gene bodies. “Silent” histone marks such as H3K27me3 seem to function dominantly over “active” histone marks such as H3K4me3, as exemplified by the colocalization of both marks (ie, bivalent domains) over inactive promoters.8,44,97,103 These results suggest that histone modifications delineate functional features of genes, includ-
These studies suggest that enhancers may be involved in transcriptional changes in response to differentiation, histone modifications associated with gene silencing (eg, H3K27me3) and active gene expression (eg, H3K4me3, H3K36me3, H3K9me1, and H4K20me1) are gained and lost, respectively, in downregulated genes (eg, CD34) (A), whereas the opposite is true for upregulated genes even when the gene promoter region harbors activating histone marks (eg, H3K4me1) and silencing marks such as H3K27me1 (eg, CD36) (B). Differentiation of CD34+ to CD36- cells is depicted from top to bottom (blue arrow) for both italicized letter A and italicized letter B. Other loci experienced similar changes of these and other histone marks including genes in the HOX cluster, transcription factors Gata1 and Gata3, and other lineage-specific genes.115
tionally, our recent epigenomic data on various T helper cell subsets demonstrated that signature cytokine genes adopted a histone modification pattern consistent with their expression status within their respective lineage on differentiation, whereas promoters of key transcription factor genes were marked by both H3K4me3 and H3K27me3 even when they were not expressed in the cell. These bivalently modified transcription factor genes can be induced under appropriate growth conditions and underlie the plasticity of T cells. Altogether, these studies support a recurring theme for histone modifications in restricting multipotency and defining the developmental potential of progenitor cells. These modifications also interact with DNA methylation to reinforce lineage commitment as was observed in neuronal progenitors.

Epigenome, Environment, and Disease

During and following the programmed changes to the mammalian epigenome in development, epigenetic processes respond to extrinsic factors including environment, diet, and even behavior. This plasticity is thought to allow the organism to respond and adapt quickly to external stimuli, yet also confer the organism, and even in some cases its offspring, with the ability to “memorize” contacts with such stimuli into adulthood. Recent studies in mammalian systems provide explicit examples of this metastability that contrasts with the neo-Darwinian concept of inheritance. In particular, evidence that fetal environment influences epigenetic processes that subsequently alter susceptibility to chronic disorders are mounting. These results have important implications for human health and provide further impetus to map human epigenomes.

Perhaps the most compelling initial data demonstrating that epigenetic modifications are sensitive to environmental stimuli comes from studies of so-called metastable epialleles identified in mice. As the name suggests, metastable epialleles are alleles that are variably expressed because of epigenetic modifications that are established in early development, giving rise to heritable phenotypic mosaicism between cells and between individuals in the absence of genetic heterogeneity. So far, such alleles include murine A\(^\text{\textsuperscript{v}}\), Axin\(^\text{\textsuperscript{pa}}\), and Cabp\(^{\text{IAP}}\), all of which are associated with contraoriented IAP (intercisternal A particle) insertions whose activities are regulated by DNA methylation and influence gene expression. DNA methylation levels at these IAP elements vary stochastically and depend on maternal nutrition and environmental exposures during early development, contributing directly to variable expressivity and phenotypic variation. For instance, agouti dams fed a diet high in folate and betaine, which are metabolized into methyl donors required for the catalytic processes of DNA and histone methylation, shifts the coat color distribution of offspring from yellow toward brown and decreases their risk for obesity by increasing DNA methylation near the viable yellow agouti (A\(^\text{V}\)) locus. Maternal dietary supplementation with genistein found in soy produced similar results. Surprisingly, opposite results were observed in offspring of dams fed a diet comprising an endocrine active compound bisphenol A, linking the impact of environmental exposure to xenobiotics to potentially deleterious consequences on the epigenome and health of mammals. Indeed, fetal exposure to xenobiotics alters epigenetic states in the offspring that produces deleterious affects on the health of the offspring and on the health of their descendants through transgenerational transmission of the epigenetic (ie, DNA methylation) perturbation in the germline. Intriguingly, bisphenol A-induced hypomethylation in the fetal epigenome could be abolished by maternal dietary supplementation with methyl donors, demonstrating that changes in diet can offset the potentially deleterious effects of environmental toxicants on the developing fetus. How this works at the molecular level is still not quite understood, but other examples of the plasticity of the fetal epigenome abound and are not necessarily restricted to metastable epialleles.

Animal models mimicking human pathology by alterations in maternal nutrition during pregnancy and weaning have consistently demonstrated significant responses on epigenetic modifications to chromatin and subsequent effects of offspring phenotype. Of particular interest are models recapitulating hypertension, which is a major risk factor for cardiovascular and cerebrovascular disease and to which genetic and environmental factors clearly contribute. In rodents, subjecting mothers to relative undernutrition during pregnancy may predispose the developing offspring to cardiovascular disease later in life. For example, utero-placental insufficiency reduced the activity of DNMT1 in the kidneys of offspring, causing hypomethylation of the p53 promoter and increased p53 expression, presumably contributing to elevated p53-mediated apoptosis thereby reducing glomeruli number and causing hypertension. In primates, a maternal high-fat diet contributes to increased histone H3 acetylation and decreased histone deacetylase activity consistent with increased expression of genes relevant to impaired lipid metabolism in the fetus. These and other similar studies unequivocally show that maternal nutrition environment can significantly impact the fetal epigenome contributing directly to the health of the offspring.

Although maternal nutrition contributes to epigenetic modifications of the developing embryo in utero, maternal behavior can alter epigenetic patterns of gene expression in young neonates that, once established, also persists into adulthood. So far, examples of this phenomenon are limited to studies of rodents. Rat offspring of mothers with low pup licking and grooming (LG) behavior showed high levels of DNA methylation of the hippocampal glucocorticoid receptor (GR) gene promoter, concomitant with low expression levels of the associated gene, and exhibited poor response to stress later in life, whereas the opposite was observed in offspring whose mothers displayed more frequent LG behavior. Additionally, cross-fostering experiments, where pups of low LG mothers were swapped with pups of high LG mothers and vice versa, convincingly revealed that the offspring GR methylation pattern was mediated by maternal behavior, rather than genetics, in the early postnatal environment during a critical window of postnatal development. This is, the DNA methylation pattern of the GR promoter was exclusively labile during the first week of birth, but once established remained unchanged into adulthood. Whether maternal phenotype contributes to epigenetic and phenotypic variabil-
ity of offspring in humans, however, has yet to be directly determined.

Taken together, results from these studies lead to the conclusion that epigenetic modifications caused by the environment of the organism at particularly vulnerable or epigenetically labile periods of development are involved in the etiology of adult disease, which may be easily prevented with dietary supplementation at these critical developmental windows. Importantly, in specific cases, the induced metastable epigenetic alteration is transgenerationally heritable. A priori knowledge as to which epigenetically labile loci are associated with disease outcome in adulthood is essential for preventative or therapeutic intervention. This can be achieved by epigenome mapping studies in development.

Relevance to Human Disease
The above results from animal models are relevant to humans and are entirely compatible with the developmental origins of disease hypothesis postulating that nutrition and other environmental factors during human prenatal and early postnatal development influence “developmental plasticity” and alter susceptibility to disease including adult cardiovascular disease, type 2 diabetes, and obesity. The role of epigenetic variability in contributing to disease susceptibility in humans has largely been unexplored. Yet results of epidemiological studies implicate epigenetic mechanisms in the etiology of disease.

Analogous to rodent studies of maternal behavior and offspring’s response to stress, the severity of symptoms arising from post-traumatic stress disorder that influences maternal cortisol production during pregnancy can predict the levels of cortisol excretion in babies, while women who experience severe stress during pregnancy give birth to offspring that experience altered activity of the HPA (hypothalamus–pituitary–adrenal) axis, which is regulated by the GR gene, in childhood. Importantly, low cortisol levels are significantly associated with maternal post-traumatic stress disorder implicating the involvement of epigenetic mechanisms in mediating this response. Also, consistent with the described animal studies supporting a major role of the intrauterine environment contributing to the health of progeny via epigenetic alterations, lower birth weight in black populations, in particular, predicts higher blood pressure, elevated cortisol reactivity, and early signs of diabetes as children and related cardiovascular conditions as adults. Additionally, studies of multigenerational trends of birth outcomes suggest a potentially nongenetic mode of inheritance because maternal fetal growth rate predicts that of their offspring. Together, these and other epidemiological data provide preliminary support that environmental factors impacting gene expression variability related to phenotype. Thus, integrating DNA sequence, transcriptome, and copy number variation information with epigenome mapping data will provide a powerful approach, called epigenome-wide association studies (eGWAS), toward comprehensively understanding disease. Indeed, accounting for nongenetic factors improves the statistical power of studies measuring gene expression variability related to phenotype. Furthermore, transcriptional variability can be inherited, presumably by epigenetics. To complement proposed theoretical frameworks and ongoing eGWAS efforts for integrating genetic and epigenetic information on a genome-wide scale, we briefly highlight 3 feasible additional applications of eGWAS given current technology.

Genome-Wide Association Studies Using Epigenetic Modifications Can Increase Statistical Power to Detect Disease-Causing Genomic Loci
In conventional genome-wide association studies (GWAS), a large number of single nucleotide polymorphisms (SNPs), approximately one million, are genotyped in thousands of cases and controls, and are statistically tested one-by-one for an association with the disease status. Because of a large
number of statistical tests performed, the resulting probability values must be adjusted for multiple testing. Typically, only a few SNPs are found to be significantly associated with a disease, and the cumulative effect size of these SNPs constitutes only a very small fraction of the total expected effect size attributable to genetic factors. The missing fraction of the genetic factors that contribute to the disease risk is believed to come from epistatic interactions of multiple genetic loci. In principle, one can attempt an epistatic association analysis with different combinations of SNPs. However, this becomes exponentially infeasible both computationally and statistically. As described in the previous sections, extrinsic signals are transmitted through intracellular transduction networks, which in turn are fed into gene regulatory cascades that impinge on the chromatin state via various epigenetic modifications (Figure 3). Thus, the epigenomic variations caused by epistatic interactions of multiple genetic loci may be identified through epigenome mapping, which could pinpoint the affected regions and thus increase the statistical power of capturing the complex disease-causing epistatic interactions associated with genetic and environmental factors.

**Epigenetic Modifications Could Serve As Proxies for Unknown Environmental Factors Associated With Disease**
Because there can be multiple unknown environmental factors that predispose to disease, it can be difficult to profile all possible factors for association with the disease status. The environment influences epigenetic states and thus these states are inherently quantifiable variables that can be used as proxies for diverse unknown disease predisposing environmental factors. Mapping these observable states would enable conventional DNA sequence–based studies to take into account the many unknown environmental variables that otherwise confound such studies.

**Epigenetic Modifications Are Observable Marks of Functional Regulatory Elements in the Genome**
It is widely believed that the majority of disease susceptibility loci for common diseases are in noncoding functional regulatory elements of the genome. Indeed, population studies have focused on identifying cis-acting regulatory variation that might associate with gene expression variability and disease susceptibility. Therefore, the task of finding such regulatory elements is paramount. Computational methods for regulatory element identification based on DNA sequence alone are plagued by high false-positive rates. One of the reasons for the failure of DNA sequence–based methods for detecting regulatory elements is that regulatory information is nonlocally encoded in linear DNA sequence. An epigenome integrates nonlocal information into local states. Indeed, it was observed that there are some regions in the genome, including known enhancer elements, where histone modification patterns are conserved between humans and mice, whereas the underlying local DNA sequence is not. Complex interactions of regulatory elements in the genome with the local chromatin environment and the gene regulatory network create distinct chromatin states in the vicinity of these elements. As we have discussed, histone modification patterns have been used for predictions of functional elements; thus, epigenome mapping studies will further aid in the identification of functional regulatory sequences throughout the genome which may provide insight into disease.

**Future Perspective**
There is now precedence of interindividual epigenetic variability, which may influence disease susceptibility. In fact, 3 classes of epigenetic variation and examples of each have been described. As we have discussed, tools for detecting global epigenetic diversity are already in place, with many of them continually being improved on. We expect that
Sources of Funding

This work was supported by the Division of Intramural Research Program of the NIH, National Heart, Lung, and Blood Institute.

Disclosures

None.

References


Epigenome Mapping in Normal and Disease States
Alika K. Maunakea, Iouri Chepelev and Keji Zhao

Circ Res. 2010;107:327-339
doi: 10.1161/CIRCRESAHA.110.222463

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/107/3/327

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org/subscriptions/