DOCUMENT SUMMARY This 2012 review paper provides a detailed overview of the basic concepts of **epigenetics** and how environmental signals influence gene expression. It describes the complex molecular machinery that controls whether genes are turned on or off, including **DNA methylation** by **DNMTs**, the role of **methyl-CpG-binding proteins (MBPs)** like **MeCP2**, and the vast array of **histone modifications** that form the "histone code." The paper explains how these elements work together with nucleosomal remodeling complexes to create either condensed, silenced **heterochromatin** or accessible, active **euchromatin**, ultimately shaping an organism's phenotype in response to its environment.

FILENAME mazzio_2012_research_report_epigenetics_gene_expression.md

METADATA Category: RESEARCH Type: report Relevance: Reference Update Frequency: Static Tags: #epigenetics #gene-expression #dna-methylation #histone-modification #chromatin #environmental-signals #gene-silencing #mecp2 #hpa-axis #molecular-biology Related Docs: This document provides foundational concepts for Murgatroyd et al., 2009 and Hull et al., 2017. Supersedes: None specified

FORMATTED CONTENT

Basic concepts of epigenetics: Impact of environmental signals on gene expression

Elizabeth A. Mazzio and Karam F.A. Soliman*

Key words: DNA methyltransferases, **methyl-CpG-binding proteins**, H2A.Z, high mobility group proteins, SWI-SNF, nucleosomal remodeling complexes, **heterochromatin** proteins, **epigenetics**

Abstract

Through epigenetic modifications, specific long-term phenotypic consequences can arise from environmental influence on slowly evolving genomic DNA. Heritable epigenetic information regulates nucleosomal arrangement around DNA and determines patterns of gene silencing or active transcription. One of the greatest challenges in the study of **epigenetics** as it relates to disease is the enormous diversity of proteins, **histone modifications** and **DNA methylation** patterns associated with each unique maladaptive phenotype. This is further complicated by a limitless combination of environmental cues that could alter the epigenome of specific cell types, tissues, organs and systems. In addition, complexities arise from the interpretation of studies describing analogous but not identical processes in flies, plants, worms, yeast, ciliated protozoans, tumor cells and mammals. This review integrates fundamental basic concepts of **epigenetics** with specific focus on how the epigenetic machinery interacts and operates in continuity to silence or activate gene expression. Topics covered include the connection between **DNA methylation**, **methyl-CpG-binding proteins**, transcriptional repression complexes, histone residues, **histone modifications** that mediate gene repression or relaxation, histone core variant stability, H1 histone linker flexibility, FACT complex, nucleosomal remodeling complexes, HP1 and nuclear lamins.

Introduction

The human genome is composed of billions of sequence arrangements containing a bioinformatics code that controls how genes are expressed. This code is further dependent upon heritable non-static epigenetic arrangement of histone scaffolding that surrounds the DNA and comprises the "epigenome." While the evolution of DNA occurs at a slow pace, expedient heritable changes to the epigenome allow dynamic and flexible modification to suit rapid environmental adaptation. While the epigenome has more influence on the temporal phenotype, the collective effects of change to the genome and the epigenome contribute to observable physical or biochemical characteristics of an organism.

Throughout the life cycle, dynamic epigenetic control over the phenotype is influenced by a time component responsible for maturation and senescence from conception to adulthood. Environmental epigenetic factors affecting long-term phenotypic change are largely initiated during in utero/perinatal periods, when introduction to the external world is being established. It is believed that since epigenetic patterns are inherited through mitosis, the earlier the stage of development, the more critical the environmental impact on the resulting phenotype.

> During fetal development, environmental cues can induce the modification of a pliable epigenome, which can result in long-term changes in gene expression that occur in a self-sustaining manner in the absence of the original stimulus.

Adverse gestational conditions that arise from inadequate healthcare, poor nutrition, socioeconomic disadvantage and racial disparities are often associated with long-lasting phenotypic consequences in adults, yielding greater risk of diabetes and heart disease, as well as low birth weight and congenital defects in progeny. It is now becoming evident that these effects are inextricably linked to altered epigenetic patterns.

Once established in the offspring, epigenetic marks can become trans-generational, continuing transmittance to future descendants including the very trait of maternal nurturing in females. The longevity of transgenerational epigenomic inheritance pattern is further influenced by the severity and repetition of a similar environmental stimulus among individuals of the same lineage. If the stimuli are discontinued, phenotypic traits could dissipate after the first or second generation.

The purpose of this review is to simplify the enormous complexity of epigenetic biochemistry that links nuclear DNA to the environment. On one hand, the concept of **epigenetics** is relatively simple in that it describes a means by which genes are either turned on or off by a heritable epigenome. On the other hand, the environmental and biological controls that mediate these events are extraordinary in number.

DNA Methylation

More than 30 million base pairs of the human genome are CG dinucleotides (CpGs). CpG sites are target platforms for methylation, which consist of the covalent attachment of a methyl group to the 5' position of the cytosine (C). CpG methylation patterns correlate with transcriptional silencing patterns observed in 60-90% of the human genes. Methylated CpGs reinforce silencing by a number of processes including direct ability to block transcription initiation complexes from binding to DNA promoter regions, inhibition of RNA polymerase (RNAP) and recruitment of transcriptional repressor complexes.

Methylation of CpG sites is performed by **DNA methyltransferases DNMT3A** and **DNMT3B**, which carry out de novo methylation, and **DNMT1**, which ensures heritable epigenome replication through cell division. A closer look at **DNMT1** shows that this enzyme not only methylates DNA, but also docks directly to **methyl-CpG binding proteins (MBPs)**, such as **MeCP2**, MBD2 and MBD3. **MBPs** can equally dock to constrictive histone enzymes, such as histone deacetylases HDAC1 and HDAC2, human H3K9 methyltransferase (Suv39 h1) and **heterochromatin** protein 1 (HP1), all synergistic components of gene silencing.

Once silencing expression patterns are established by **DNA methylation**, epigenetic inheritance patterns are further solidified by small non-coding RNAs (miRNAs) that ensure regional silencing through degradation of unwanted mRNAs, stunting mRNA maturation or blocking promoter areas associated with mRNA to be silenced.

Methyl-CpG-binding proteins

Once CpG methylation patterns are established, silencing is continually reinforced by a series of **methyl-CpG-binding proteins** that directly dock to CpGs through their methyl-CpG binding domains (MBD). In humans, MBD proteins include MBD1, MBD2, MBD3, MBD4 and **MeCP2**. The basic function of MBD proteins is to secure a primary attachment to methylated CpG and a secondary attachment to the surrounding histone scaffolding in order to enable constriction by further docking of **DNMTs**, histone methyltransferases, HDACs and ATPase **chromatin** remodeling complexes. These components work together to compress **chromatin** into **heterochromatin** at transcription start sites throughout the genome.

- **MBD1**: MBD1 forms a primary attachment to methylated CpG sites, a secondary attachment to histone constriction enzymes such as SETDB1/Suv39 h1, which methylates H3K9, a tertiary attachment to **DNMT1**, and a quaternary attachment to a TRD protein.
- **MBD2-MBD4**: MBD2 forms a primary attachment to methylated CpG and a secondary attachment to the Mi2-NuRD complex. The Mi2-NuRD complex houses the histone constriction enzymes HDAC1 and HDAC2. MBD4 is unique in that it functions as a DNA repair enzyme that maintains methylated CpG motifs.
- **MeCP2**: **MeCP2** is a prominent silencing mark that is heavily embedded in **heterochromatin**. Similar to MBD1, **MeCP2** also contains a TRD unit that prevents transcription by RNAP II and docks to the paired amphipathic helix protein Sin3A. Yeast Sin3 (Sin3A and Sin3B) serve as co-repressors that bring histone deacetylase activity in very close proximity to genes targeted for silencing.
- **Other human MBD proteins**: Additional human proteins that contain MBD domains have been identified, suggesting a role in collaborative silencing. These include: (1) BAZ2A/TIP5 and BAZ2B; (2) KMT1F/CLLD8, KMTIE/SETDB1 and (3) KIAA1461/MBD5 and KIAA1887/MBD6.

Mutations or defects in any of the above mentioned processes, including **DNMTs** or **MBPs**, can lead to developmental disorders such as Rett syndrome (**MeCP2**), Angelman syndrome, defects in brain development, autism (MBD1), or immunodeficiency-centromeric instability and facial anomalies syndrome (DNMT3b).

Histone Modifications

Once CpG sites are methylated and associate with MPD-transcription repression complexes, the next order of macromolecular control is given by the distribution and stability of the histone units. Each individual histone octamer is comprised of two copies of H2A/H2B dimer cores and H3/H4 tetramers, which wrap around 146 base pairs of DNA. Repeating histone units make up the composition of **nucleosomes** and **nucleosomes** make up higher order **chromatin**. Histone octamer components contain a structured domain and an unstructured N-terminal tail of varying length that protrudes outward from the nucleosome, being readily subject to modifications known as "histone marks."

Histone marks are established by covalent interactions that alter the electrostatic charge and, therefore, histone shape and DNA-histone affinity. The compilation of histone marks make up what is termed "**the histone code**." The number of possible combinations comprising **the histone code** is extraordinarily high. The most extensively studied modifications to H3/H4 tetramers include (1) hyperacetylation, mediated by histone acetylases (HATs) such as GNAT/PCAF and (2) deacetylation by HDACs. A second set of highly investigated modifications to H3/H4 tails involve (1) the transfer of methyl groups by histone methyltransferases (HMTs) and (2) removal of methyl groups by histone demethylases (HKDMs). The methylation of histones can either provoke gene repression (as in the case of H3K9me3) or gene expression (as in the case of H3K4me3).

Figure 1. H3 histone modifications. Slice numbers represent residue positions. Green denotes histone marks typically associated with gene activation; red represents histone marks generally associated with transcriptional repression. Modifications are denoted at specific histone residues as Iso, isomerated; Ac, acetylated; Me, methylated (variable methylation can include me1, me2, me3); P, phosphorylated; me-, de-methylated and AC- de-acetylated. Residues are K, lysine; R, Arginine; S, Serine; T, Threonine; P, Proline.

Figure 2. H4 histone modifications. Slice numbers represent residue positions. Green denotes histone marks typically associated with gene activation; red represents histone marks generally associated with transcriptional repression. Modifications are denoted at specific histone residues as Bio, biotinylated; AC, acetylated; Me, methylated (variable methylation can include me1, me2, me3); Ac-, de-acetylated and SUMO, sumoylated.

Nucleosomal Positioning

The nucleosomal structure is predicted by the characteristics of the collective assembly of individual histone octamers that circumscribe 146 base pairs of DNA, which are further adjoined by a piece of 15 base pairs of DNA woven through the connective linker histone H1. Constricted **nucleosomes** are packaged tightly around DNA, serving as a multi-composite mechanical barricade by which transcription initiation complexes are denied access to DNA. In the opposite manner, expansive, nucleosome-free, non-methylated CpG islands are associated with active gene expression.

The three most recognized events associated with how **nucleosomes** twist and slide involve (1) the stability of the histone cores H2A/H2B; (2) the integrity of the H1 linker and (3) the supporting tortional movement enabled by the ATPase-driven **chromatin** remodeling machinery.

Histone core stability

Nucleosomal ejection is initiated by a destabilized H2A histone core that occurs through an exchange process with an unstable dimer. The unstable dimer is collapsible, highly acetylated and enables **chromatin** expansion and nucleosome ejection at areas of unmethylated DNA targeted for active expression. Unstable histone variants such as H2A.Z most often coincide with transcription start sites and are highly concentrated in **euchromatin**. In contrast, core variant exchange with an H2A of greater stability is often associated with gene silencing.

H1-the histone linker

Disruption of the H1 linker is evident along transcription start sites that are aligned in proximity to unstable histone core variants such as H2A.Z109 and hyper-acetylated H3 and H4 tails. In mammals, these modifications occur alongside high mobility group (HMG) proteins, which serve to loosen tension of the H1 linker, enabling nucleosomal expansion.

Nucleosomal remodeling complexes

- **SWI-SNF**: ATP-dependent **chromatin** remodeling complexes were first discovered in yeast and named SWI (after yeast mating-type switching)-SNF (after sucrose non fermenting) nucleosome remodeling complexes. SWI/SNF BRG1 and hBRM ATPases are typically associated with gene expression and reportedly present in the vicinity of unmethylated CpG islands and heavily acetylated histone tails.
- **ACF**: Other remodeling complexes are involved in gene silencing, such as human ATP-dependent **chromatin** assembly and remodeling factor, hACF, or its yeast counterpart, ISW2, which arrange **nucleosomes** by a sliding motion toward formation of organized **heterochromatin**.
- **CHD/Mi2-NuRD**: Repressive remodeling complexes of the SNF2-like DNA helicase/ATPase category also include NuRD, which contains Mi-2 subunits and histone deacetylases HDAC1 and HDAC2.

Heterochromatin Proteins

Heterochromatin proteins HP1a and HPIβ are the major tethering elements that bring together most, if not all, silencing elements by adjoining Suv39 h1, H3K9me3 by a HP1 N-terminal chromodomain, **MeCP2**, methylated CpGs, MBD1/SETDB1/CAF-1 and human Swi/Snf ATPases, via their chromoshadow PXVXL motif. HP1 proteins are involved in nucleosomal constriction and counteract expansive **chromatin** remodeling complexes such as Brg1 and Brm.

Macromolecular Chromatin

Euchromatin

At the macromolecular level, transcriptionally active **euchromatin** is housed at the inner nucleoplasm/center of chromosome territories, while silenced **heterochromatin** is arranged at the outer circular perimeter of the nuclear membrane. **Euchromatin** contains a combination of unstable core variants, HATs, G9a, HMG proteins and a low occupancy of linker H1.

Heterochromatin and lamins

Silenced **heterochromatin** is comprised of two categories. Constitutive **heterochromatin** is considered to be static and irreversible through the lifetime of the organism, whereas facultative **heterochromatin** can change patterns through epigenetic alteration. At the nuclear periphery, silenced **heterochromatin** contains a combination of MBD proteins, repression complexes, H3K9me2, H3K9me3, HP1a, Suv39 h1 and methylated DNA. The nuclear lamina is required to lock the peripheral position of **heterochromatin** within the nuclear envelope to maintain silencing.

Summary

> **Epigenetics** may be conceptualized as control over how genomic DNA is expressed.

This control is initiated by **DNMTs** that methylate CpGs, which are then tagged by MBD proteins attached to potent repression complexes. Repression complexes, in turn, control modifications to histones H3 and H4 tails, which perpetuate constriction and make stable modifications to the histone cores H2A and H2B to prevent histone unit ejection/nucleosomal displacement. These collective events are associated with stabilized tension of the histone H1 linker. HP1a and HP1B proteins tether silencing elements from methylated CpGs to ATPase remodeling machinery-together, in order to tightly crowd methylated DNA close to the **nucleosomes**, thereby blocking transcription elements. The positioning of silenced **heterochromatin** along the nuclear envelope is carried on by lamins; centrally positioned **euchromatin** remaining transcriptionally active. Silencing and transcription patterns are altered by environmental cues and the transmittance of epigenetic information is heritable through mitosis and meiosis. While this describes a very basic model, the complexities of epigenetic control in regulating the phenotype seem infinite.