Genome-Wide Epigenetics

Brian C. Capell[1,2] and Shelley L. Berger[1]

Journal of Investigative Dermatology (2013) **133**, e9. doi:10.1038/jid.2013.173

WHAT IS EPIGENETICS?

The term "epigenetics" was coined by Conrad Waddington to describe "the branch of biology which studies the causal interactions between genes and their products, which bring the phenotype into being" (Goldberg *et al.*, 2007). Very broadly, the word has come to refer to the study of the regulation of genes, their expression, and how that translates into particular phenotypes, independent of any change to the underlying DNA sequence. More simply stated, epigenetics is the study of functionally relevant changes in gene expression (with subsequent changes in cellular phenotype) that result from mechanisms other than from changes in the underlying DNA nucleotide sequence.

Despite the fact that there is no change in the nucleotide sequence, epigenetic modifications may be heritable and can be passed down to subsequent generations through cell replication and division of alternative chromatin states. This "turning on or off" of genes explains why, despite having the same underlying DNA sequence, a keratinocyte looks and behaves so differently than a hepatocyte and why the epigenetic state is carried over to maintain cell- and tissue-type specification. Although a given cell's (or individual's) genome remains relatively stable over time, the epigenome can and does vary depending on a number of factors, including environmental conditions. These processes allow for many "good" functions, including normal organism development; however, aberrant epigenetic mechanisms are implicated in different disease processes, including malignancies.

This article provides a brief overview of the field of epigenetics and offers a glimpse into some of the major techniques used to study it, with a particular focus on chromatin immunoprecipitation followed by sequencing (ChIP-seq), the current standard method for studying proteins and other epigenetic factors that bind to DNA.

At the heart of epigenetic control is the organization of DNA into chromatin. This begins with 147 base pairs of DNA wrapped around eight histone proteins, which include the core histones H2A, H2B, H3, and H4 (Figure 1). Each of these histone octamers is referred to as a nucleosome. The nucleosomes are packaged tightly into even more compact fibers known as chromatin. Through this complex structure, epigenetic regulation occurs primarily through four mechanisms. First, DNA can undergo direct chemical modification by cytosine methylation, which is a general marker of gene silencing. Second, posttranslational modifications of the core histones can occur, primarily through methylation, acetylation, ubiquitylation, and phosphorylation, making up the primary chromatin structure (Figure 1). Acting in concert with these two aspects of the epigenetic machinery, noncoding RNAs contribute to the regulation of these processes (Greer and Shi, 2012). Finally, the chromatin is then packaged, via long-range interactions, into a higher-order structure within the cell nucleus.

All of these organizational steps serve to modulate DNA accessibility and thus control gene expression. More open regions of chromatin, or euchromatin, are poised for activation by the transcriptional machinery, whereas more

ADVANTAGES OF CHIP-SEQ

- ChIP-seq provides a powerful tool with which to assess with excellent resolution the binding of any transcription factor, histone modification, or other DNA-binding protein of interest across the entire genome.
- It allows investigators to assess how these modifications or transcription factors affect different phenotypes or disease states.

LIMITATIONS

- ChIP-seq reports only relative, not absolute, values of the bound proteins.
- Investigators must know the factor of interest in advance.
- ChIP-seq can be limited by the availability and quality of the antibody to the protein or modification of interest.

[1]*Epigenetics Program, Department of Cell and Developmental Biology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA and* [2]*Department of Dermatology, University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania, USA*

Correspondence: Brian C. Capell, 2 Maloney Building, 3600 Spruce Street, Philadelphia, Pennsylvania 19104, USA. E-mail: brian.capell@uphs.upenn.edu

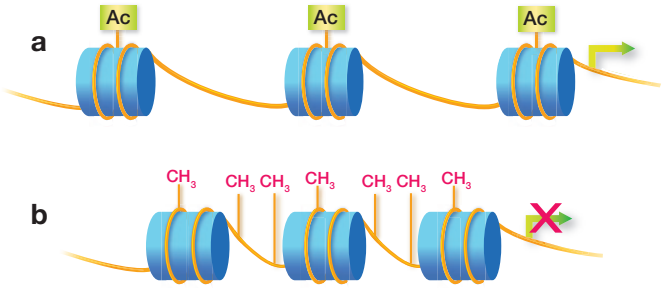

Figure 1. Major epigenetic modifications. (**a**) Histone modification. A total of 147 base pairs of DNA (depicted as yellow lines) wrap around eight histone proteins (depicted as blue cylinders) to form nucleosomes. When nucleosomes are packaged together, they form chromatin. Posttranslational modifications (e.g., addition of acetyl, methyl, or ubiquitin groups) of histones alter local chromatin conformation. Variability in chromatin compaction affects the accessibility of genes. Loosely condensed regions (euchromatin) are more actively expressed and tightly condensed regions (heterochromatin) are repressed. The image depicts histone acetylation, associated with loosening of local chromatin and more active gene expression. (**b**) DNA methylation. DNA methylation occurs through the addition of a methyl group to the C5 position of cytosine to form 5-methylcytosine, typically at cytosine phosphate guanine dinucleotides. In promoter regions, DNA methylation silences genes by interfering with transcription factor binding. DNA methylation is typically associated with tightly condensed chromatin. Reprinted with permission from Cheng and Cho, 2012.

closed regions generally exist as more stably repressed regions known as constitutive heterochromatin, or regions that are repressed but poised for activation, known as facultative heterochromatin. Certain histone modifications tend to be associated with particular transcriptional states, such as trimethylation of the histone H3 at lysine number 4 (H3K4me3) with transcriptionally active euchromatin, trimethylation of histone H3 at lysine 9 (H3K9me3) with transcriptionally repressed constitutive heterochromatic domains, and trimethylation of histone H3 at lysine 27 (H3K27me3) with facultative heterochromatin. Acetylated histones tend to mark active genes. These associations are sometimes referred to as the histone code or—more accurately, because of the complexity of the histone modifications—as a histone language (Berger, 2007).

Each step of this organization may provide sophisticated layers of regulation with potentially profound implications ranging from the maintenance of cell fate during cellular differentiation in development to the turning on or off of tumor suppressors or oncogenes in cancer. For example, hypermethylation of the *FAS* gene decreases its expression and results in a reduced ability to undergo apoptosis in Sézary syndrome (Jones *et al.*, 2010). That epigenetic factors may be modifiable or reversible makes their study particularly promising from a disease perspective. The histone deacetylase inhibitors vorinostat and romidepsin used in the treatment of cutaneous T-cell lymphoma are two primary examples, and numerous other compounds are currently under development (Dummer *et al.*, 2012).

ChIP BASICS

Several techniques are used in genome-wide epigenetic studies. ChIP, in which an antibody to a specific histone modification or other DNA-binding protein of interest is used, lies at the heart of many of these technologies. Initially, ChIP was employed to analyze chromatin structure at discrete genomic loci. It has been used more recently in conjunction with microarrays to analyze gene expression (ChIP-chip). However, with the advent of next-generation sequencing technology and its continuously declining costs, ChIP-seq has become the standard method of analyzing genome-wide maps of DNA-binding proteins and chromatin modification enrichment.

Offering much greater resolution and depth of coverage, ChIP-seq has enabled tremendous progress beyond ChIP-chip. This comprehensive mapping strategy has allowed investigators to ascribe associations between particular histone modifications and either active or repressed transcription, as well as assign them to particular locations across the genome, including enhancers, promoters, gene bodies, and insulators. Numerous variations on the standard ChIP

Table 1. ChIP-seq variations

Technique	Basic overview
DNase-seq	Uses DNase1 endonuclease to digest the DNA in areas depleted of nucleosomes, thus identifying areas genome wide of open chromatin where regulatory factors typically bind, although it does not identify the specific bound factors.
ChIP-exo	Uses λ-phage endonuclease digestion of the DNA rather than sonication to greatly improve the resolution of precisely where the factor of interest binds and to remove contaminating DNA.
FAIRE-seq	Formaldehyde-assisted identification of regulatory elements is similar to DNase-seq in that the formaldehyde identifies nucleosome-depleted regulatory regions by extracting the DNA that is not cross-linked to nucleosomes.
ChIP-MNase	Micrococcal nuclease is used to digest the DNA and determine where nucleosomes are present.
Chromatin-capture technologies (3C, 5C, HiC)	These techniques help map higher-order chromatin structure. 3C, or chromosome conformation capture, is a method to map local chromosome interactions based on the increased frequency of molecular interactions between chromosome fragments in close three-dimensional proximity in the nucleus. 5C extends the amount of the genome that can be assayed and HiC allows the entire genome to be assayed, although at a limited resolution.
ChIA-PET	(Chromatin interaction analysis with paired-end tag sequencing.) This technique is similar to the chromatin capture technologies above, but it employs a chromatin immunoprecipitation step to show how higher-order chromatin structure affects transcription.