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

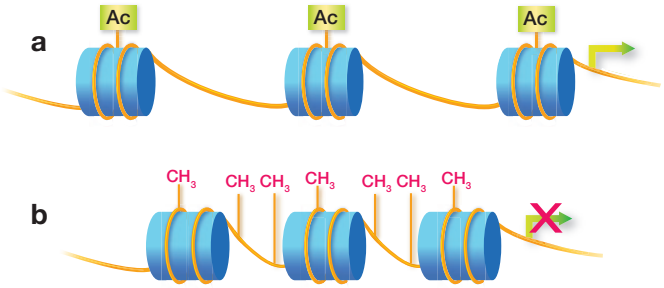

Figure 1. Major epigenetic modifications. (**a**) Histone modification. A total of 147 base pairs of DNA (depicted as yellow lines) wrap around eight histone proteins (depicted as blue cylinders) to form nucleosomes. When nucleosomes are packaged together, they form chromatin. Posttranslational modifications (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.

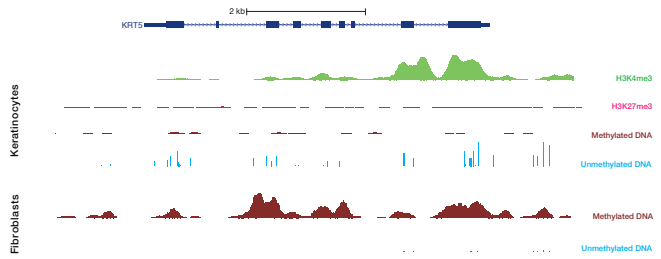

Figure 2. Basic epigenome browser view. This view highlights an ~9-Kb segment of the *KRT5* gene, with the dark blue rectangles and lines depicting the gene structure. Profiled cells are primary cultured neonatal foreskin keratinocytes (KCs, top) and fibroblasts (bottom). These KCs express *KRT5*, whereas fibroblasts do not. Starting from the top, there are high levels of H3K4me3 histone modification signal (associated with active transcription) at the promoter in KCs (depicted in green). In pink, H3K27me3 signal (associated with gene repression) shows low levels in KCs. Regions of methylated DNA are depicted in brown, with minimal signal in KCs and high levels in fibroblasts. Unmethylated DNA is depicted with light blue vertical bars and shows a high number of sequencing peaks for KCs compared with a low number for fibroblasts. Reprinted with permission from Cheng and Cho, 2012.

protocol have created complementary assays that have enhanced our ability to interrogate chromatin modifications and gain different information (Table 1).

The initial step in ChIP-seq involves cross-linking DNA and protein in a population of cells with formaldehyde. The chromatin is then sonicated into small fragments of roughly 200–600 base pairs. An antibody to the protein or histone modification of interest is used to immunoprecipitate the DNA–protein complex. Input DNA (the part of the sample that is removed prior to the immunoprecipitation step) is the most common control. The cross-links are then reversed to allow the DNA to be assayed by creating a DNA library for next-generation sequencing. During library preparation, the DNA is selected for its size by gel electrophoresis, typically in the range of 200–300 base pairs. The DNA is then amplified by polymerase chain reaction prior to sequencing (Furey, 2012; Park, 2009).

Most ChIP-seq experiments have been performed with the Illumina Genome Analyzer, and they typically produce 8–15 million sequence reads, which are then mapped back to the human genome to look for peaks of signal enrichment of statistical significance relative to the control. Establishing these peaks is highly dependent on threshold settings. Sharp peaks often define transcription factor binding sites, whereas broad peaks often represent large domains (Figure 2). In general, the more common a histone modification is across the genome, the greater the number of sequence reads required to accurately map it.

LIMITATIONS

ChIP reports only relative, and not absolute, enrichment of particular modifications across the genome; thus, it is not always possible to infer biological relevance when signals are more subtle. The technique is inherently biased in that one must know the particular protein or histone modification one seeks prior to starting the procedure. Similarly, the specificity and sensitivity (i.e., affinity) of the polyclonal antibodies used is one of the most important determinants of the success of the experiment. Antibodies can vary not only from different companies, but also from batch to batch (Furey, 2012; Park, 2009). Development of monoclonal antibodies for histone modifications is helping to ameliorate these antibody variation issues.

Both underlying genetic and environmental factors can influence the underlying chromatin state. Therefore, it is important to keep in mind that ChIP captures a particular modification at a particular moment in time, and its results are limited by the fact that chromatin can have stable differences both across varying cell populations and within a population of a particular cell type that may be of variable heterogeneity (Zhou *et al.*, 2011). Thus, to make final conclusions, the results of multiple ChIP experiments must be aligned and compared. Other quality-control measures include testing independent antibodies, given the possibility of nonspecific binding and cross-reactivity of similar histone modifications, and RNA interference against the enzymes responsible for adding the modifying group or, alternatively, mass spectrometry.

QUESTIONS

Answers are available as supplementary material online and at http://www.scilogs.com/jid/.

1. Heterochromatin refers to:

A. Actively transcribed regions of DNA to which transcription factors actively bind.

B. The core structure around which 147 base pairs of DNA bind.

C. Constitutively closed and transcriptionally repressed areas of genome organization.

D. Areas of the genome marked by the histone modification H3K4me3.

2. The major role of ChIP-seq experiments lies in:

A. The ability to map DNA methylation patterns across the genome.

B. The ability to map the bound histone modifications and other proteins bound to chromatin across the genome.

C. The ability to analyze gene expression changes with changes in chromatin modifications.

D. The ability to associate single-nucleotide polymorphisms with various disease states.

THE FUTURE

ChIP and its variations (discussed in Table 1) are providing investigators with powerful tools to study the epigenome and its role in normal and diseased states. Next-generation sequencing technologies have allowed comprehensive genome-wide perspectives through both ChIP-seq and its variations (Table 1), as well as other techniques such as whole-genome bisulfite sequencing for the study of DNA methylation. These techniques are likely to become more common and streamlined as sequencing technology continues to improve—for example, as sensitive and fast techniques such as nanopore sequencing become available. Many important questions remain, from how epigenetic marks are inherited to dissecting precisely how various epigenetic modifications and global changes in chromatin organization are regulated by various signals, including environmental exposures and developmental cues. As these mysteries are unraveled, they should have important implications for many fundamental processes inherent to dermatology, including processes as diverse as hair follicle stem-cell differentiation, skin cancer, and skin aging (Botchkarev *et al*., 2012; Cheng and Cho, 2012).

CONFLICT OF INTEREST

The authors state no conflict of interest.

ACKNOWLEDGMENTS

This work was supported by a NIAMS T32 grant to the University of Pennsylvania Department of Dermatology and NIA P01 grant P01AG031862 to SLB.

SUPPLEMENTARY MATERIAL

Answers and a PowerPoint slide presentation appropriate for journal club or other teaching exercises are available at http://dx.doi.org/10.1038/jid.2013.173.

REFERENCES

Berger SL (2007) The complex language of chromatin regulation during transcription. *Nature* 447:407–12

Botchkarev VA, Gdula MR, Mardaryev AN *et al.* (2012) Epigenetic regulation of gene expression in keratinocytes. *J Invest Dermatol* 132:2505–21

Cheng JB, Cho RJ (2012) Genetics and epigenetics of the skin meet deep sequence. *J Invest Dermatol* 132:923–32

Dummer R, Goldinger SM, Cozzio A *et al.* (2012) Cutaneous lymphomas: molecular pathways leading to new drugs. *J Invest Dermatol* 132: 517–25

Furey TS (2012) ChIP-seq and beyond: new and improved methodologies to detect and characterize protein–DNA interactions. *Nat Rev Genet* 13:840–52

Goldberg AD, Allis CD, Bernstein E (2007) Epigenetics: a landscape takes shape. *Cell* 128:635–8

Greer EL, Shi Y (2012) Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet* 13:343–57

Jones CL, Wain EM, Chu CC *et al.* (2010). Downregulation of Fas gene expression in Sezary syndrome is associated with promoter hypermethylation. *J Invest Dermatol* 130:1116–25

Park PJ (2009) ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 10:669–80

Zhou VW, Goren A, Bernstein BE (2011) Charting histone modifications and the functional organization of mammalian genomes. *Nat Rev Genet* 12:7–18