CHAPTER 3

Overview and Concepts

C. David Allis,[1] Thomas Jenuwein,[2] and Danny Reinberg[3]

[1]The Rockefeller University, New York, New York; [2]Research Institute of Molecular Pathology, Vienna, Austria; [3]UMDNJ-Robert Wood Johnson Medical School, Piscataway, New Jersey

CONTENTS

GENERAL SUMMARY

The DNA sequencing of the human genome and the genomes of many model organisms has generated considerable excitement within the biomedical community and the general public over the past several years. These genetic "blueprints" that exhibit the well-accepted rules of Mendelian inheritance are now readily available for close inspection, opening the door to improved understanding of human biology and disease. This knowledge is also generating renewed hope for novel therapeutic strategies and treatments. Many fundamental questions nonetheless remain. For example, how does normal development proceed, given that every cell has the same genetic information, yet follows a different developmental pathway, realized with exact temporal and spatial precision? How does a cell decide when to divide and differentiate, or when to retain an unchanged cellular identity, responding and expressing according to its normal developmental program? Mistakes made in the above processes can lead to the generation of disease states such as cancer. Are these mistakes encoded in faulty genetic blueprints that we inherited from one or both of our parents, or are there other layers of regulatory information that are not being properly read and decoded?

In humans, the genetic information (DNA) is organized into 23 chromosome pairs consisting of approximately 25,000 genes. These chromosomes can be compared to libraries with different sets of books that together instruct the development of a complete human being. The DNA sequence of our genome is composed of about 3×10^9 bases, abbreviated by the four letters (or bases) A, C, G, and T within its sequence, giving rise to well-defined words (genes), sentences, chapters, and books. However, what dictates when the different books are read, and in what order, remains far from clear. Meeting this extraordinary challenge is likely to reveal insights into how cellular events are coordinated during normal and abnormal development.

When summed across all chromosomes, the DNA molecule in higher eukaryotes is about 2 meters long and therefore needs to be maximally condensed about 10,000-fold to fit into a cell's nucleus, the compartment of a cell that stores our genetic material. The wrapping of DNA around "spools" of proteins, so-called histone proteins, provides an elegant solution to this packaging problem, giving rise to a repeating protein:DNA polymer known as chromatin. However, in packaging DNA to better fit into a confined space, a problem develops, much as when one packs too many books onto library shelves: It becomes harder to find and read the book of choice, and thus, an indexing system is needed. Chromatin, as a genome-organizing platform, provides this indexing. Chromatin is not uniform in structure; it comes in different packaging designs from a highly condensed chromatin fiber (known as heterochromatin) to a less compacted type where genes are typically expressed (known as euchromatin). Variation can enter into the basic chromatin polymer through the introduction of unusual histone proteins (known as histone variants), altered chromatin structures (known as chromatin remodeling), and the addition of chemical flags to the histone proteins themselves (known as covalent modifications). Moreover, addition of a methyl group directly to a cytosine (C) base in the DNA template (known as DNA methylation) can provide docking sites for proteins to alter the chromatin state or affect the covalent modification of resident histones. Recent evidence suggests that noncoding RNAs can "guide" specialized regions of the genome into more compacted chromatin states. Thus, chromatin should be viewed as a dynamic polymer that can index the genome and potentiate signals from the environment, ultimately determining which genes are expressed and which are not.

Together, these regulatory options provide chromatin with an organizing principle for genomes known as "epigenetics," the subject of this book. In some cases, epigenetic indexing patterns appear to be inherited through cell divisions, providing cellular "memory" that may extend the heritable information potential of the genetic (DNA) code. Epigenetics can thus be narrowly defined as changes in gene transcription through modulation of chromatin, which is not brought about by changes in the DNA sequence.

In this overview, we explain the basic concepts of chromatin and epigenetics, and we discuss how epigenetic control may give us the clues to solve some long-standing mysteries, such as cellular identity, tumorigenesis, stem cell plasticity, regeneration, and aging. As readers comb through the chapters that follow, we encourage them to note the wide range of biological phenomena uncovered in a diverse range of experimental models that seem to have an epigenetic (non-DNA) basis. Understanding how epigenetics operates in mechanistic terms will likely have important and far-reaching implications for human biology and human disease in this "post-genomic" era.

1 Genetics Versus Epigenetics

Determining the structural details of the DNA double helix stands as one of the landmark discoveries in all of biology. DNA is the prime macromolecule that stores genetic information (Avery et al. 1944), and it propagates this stored information to the next generation through the germ line. From this and other findings, the "central dogma" of modern biology emerged. This dogma encapsulates the processes involved in maintaining and translating the genetic template required for life. The essential stages are (1) the self-propagation of DNA by semiconservative replication; (2) transcription in a unidirectional 5′ to 3′ direction, templated by the genetic code (DNA), generation of an intermediary messenger RNA (mRNA); (3) translation of mRNA to produce polypeptides consisting of linear amino to carboxyl strings of amino acids that are colinear with the 5′ to 3′ order of DNA. In simple terms: *DNA* ↔ *RNA* → *protein*. The central dogma accommodates feedback from RNA to DNA by the process of reverse transcription, followed by integration into existing DNA (as demonstrated by retroviruses and retrotransposons). However, this dogma disavows feedback from protein to DNA, although a new twist to the genetic dogma is that rare proteins, known as prions, can be inherited in the absence of a DNA or RNA template. Thus, these specialized self-aggregating proteins have properties that resemble some properties of DNA itself, including a mechanism for replication and information storage (Cohen and Prusiner 1998; Shorter and Lindquist 2005). Additionally, emerging evidence suggests that a remarkably large fraction of our genome is transcribed into "noncoding" RNAs. The function of these noncoding RNAs (i.e., non-protein-encoding except tRNAs, rRNAs, snoRNAs) is under active investigation and is only beginning to become clear in a limited number of cases.

The origin of epigenetics stems from long-standing studies of seemingly anomalous (i.e., non-Mendelian) and disparate patterns of inheritance in many organisms (see Chapters 1 and 2 for a historical overview). Classic Mendelian inheritance of phenotypic traits (e.g., pea color, number of digits, or hemoglobin insufficiency) results from allelic differences caused by mutations of the DNA sequence. Collectively, mutations underlie the definition of phenotypic traits, which contributes to the determination of species boundaries. These boundaries are then shaped by the pressures of natural selection, as explained by Darwin's theory of evolution. Such concepts place mutations at the heart of classic genetics. In contrast, non-Mendelian inheritance (e.g., variation of embryonic growth, mosaic skin coloring, random X inactivation, plant paramutation) (Fig. 1) can manifest, to take one example, from the expression of only one (of two) alleles within the same nuclear environment. Importantly, in these circumstances, the DNA sequence is not altered. This is distinct from another commonly referred to non-Mendelian inheritance pattern that arises from the maternal inheritance of mitochondria (Birky 2001).

The challenge for epigenetic research is captured by the selective regulation of one allele within a nucleus. What distinguishes two identical alleles, and how is this distinction mechanistically established and maintained through successive cell generations? What underlies differences observed in monozygotic ("identical") twins that make them not totally identical? Epigenetics is sometimes cited as one explanation for the differences in outward traits, by translating the influence of the environment, diet, and potentially other external sources to the expression of the genome (Klar 2004; see Chapters 23 and 24). Determining what components are affected at a molecular level, and how alterations in these components affect human biology and human disease, is a major challenge for future studies.

Another key question in the field is, How important is the contribution of epigenetic information for normal development? How do normal pathways become dysfunctional, leading to abnormal development and neoplastic transformation (i.e., cancer)? As mentioned above, "identical" twins share the same DNA sequence, and as such, their phenotypic identity is often used to underscore the defining power of genetics. However, even twins such as these can exhibit outward phenotypic differences, likely imparted by epigenetic modifications that occur over the lifetime of the individuals (Fraga et al. 2005). Thus, the extent to which epigenetics is important in defining cell fate, identity, and phenotype remains to be fully understood. In the case of tissue regeneration and aging, it remains unclear whether these processes are dictated by alterations in the genetic program of cells or by epigenetic modifications. The intensity of research on a global scale testifies to the recognition that the field of epigenetics is a critical new frontier in this post-genomic era.

In the words of others, "We are more than the sum of our genes" (Klar 1998), or "You can inherit something beyond the DNA sequence. That's where the real excitement in genetics is now" (Watson 2003). The overriding motivation for deciding to edit this book was the general belief that we and all the contributors to this volume could transmit this excitement to future generations of students, scientists, and physicians, most of whom were taught genetic, but not epigenetic, principles governing inheritance and chromosome segregation.

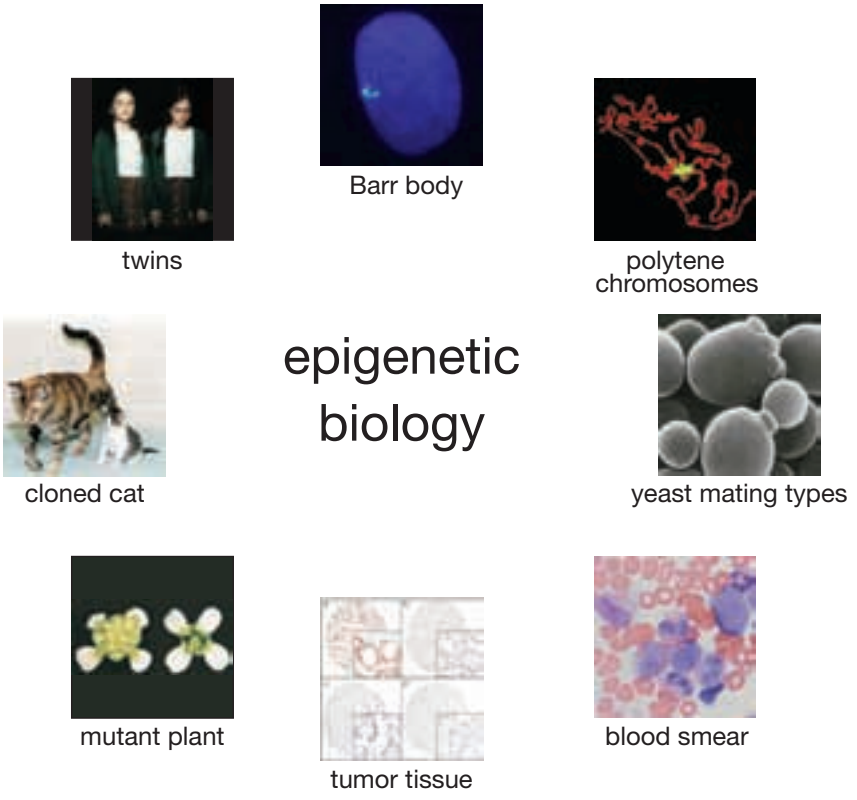

Figure 1. Biological Examples of Epigenetic Phenotypes

Epigenetic phenotypes in a range of organisms and cell types, all attributable to *non-genetic* differences. *Twins:* Slight variations partially attributable to epigenetics (©Randy Harris, New York). *Barr body:* The epigenetically silenced X chromosome in female mammalian cells, visible cytologically as condensed heterochromatin. *Polytene chromosomes:* Giant chromosomes in *Drosophila* salivary glands, ideally suited for correlating genes with epigenetic marks (reprinted from Schotta et al. 2003 [©Springer]). *Yeast mating type:* Sex is determined by the active MAT locus, while copies of both mating-type genes are epigenetically silenced (©Alan Wheals, University of Bath). *Blood smear:* Heterogeneous cells of the same genotype, but epigenetically determined to serve different functions (courtesy Prof. Christian Sillaber). *Tumor tissue:* Metastatic cells (*left*) showing elevated levels of epigenetic marks in the tissue section (reprinted, with permission, from Seligson et al. 2005 [©Macmillan]). *Mutant plant: Arabidopsis* flower epiphenotypes, genetically identical, with epigenetically caused mutations (reprinted, with permission, from Jackson et al. 2002 [©Macmillan]). *Cloned cat:* Genetically identical, but with varying coat-color phenotype (reprinted, with permission, from Shin et al. 2002 [©Macmillan]).

2 Model Systems for the Study of Epigenetics

The study of epigenetics necessarily requires good experimental models, and as often is the case, these models seem at first sight far removed from studies using human (or mammalian) cells. Collectively, however, results from many systems have yielded a wealth of knowledge. The historical overviews (Chapters 1 and 2) make reference to several important landmark discoveries that have emerged from early cytology, the growth of genetics, the birth of molecular biology, and relatively new advances in chromatin-mediated gene regulation. Different model organisms (Fig. 2) have been pivotal in addressing and solving the various questions raised by epigenetic research. Indeed, seemingly disparate epigenetic discoveries made in various model organisms have served to unite the research community. The purpose of this section is to highlight some of these major findings, which are discussed in more detail in the following chapters of this book. As readers note these discoveries, they should focus on the fundamental principles that investigations using these model systems have exposed; their collective contributions point more often to common concepts than to diverging details.

Unicellular and "lower" eukaryotic organisms—*Saccharomyces cerevisiae, Schizosaccharomyces pombe,* and *Neurospora crassa*—permit powerful genetic analyses, in part facilitated by a short life cycle. Mating-type (MAT) switching that occurs in *S. cerevisiae* (Chapter 3) and *S. pombe* (Chapter 6) has provided remarkably instructive examples, demonstrating the importance of chromatin-mediated gene control. In the budding yeast *S. cerevisiae*, the unique silent information regulator (SIR) proteins were shown to engage specific modified histones. This was preceded by elegant experiments using genetics to document the active participation of histone proteins in gene regulation (Clark-Adams et al. 1988; Kayne et al. 1988). In the fission yeast *S. pombe*, the patterns of histone modification operating as activating and repressing signals are remarkably similar to those in metazoan organisms. This has opened the door for powerful genetic screens being employed to look for gene products that suppress or enhance the silencing of genes. Most recently, a wealth of mechanistic insights linking the RNA interference (RNAi) machinery to the induction of histone modifications acting to repress gene expression was discovered in fission yeast (Hall et al. 2002; Volpe et al. 2002). Shortly afterward, the RNAi machinery was also implicated in transcriptional gene silencing in the plant *Arabidopsis thaliana*, underscoring the potential importance of this regulation in a wide range of organisms (see Section 10).

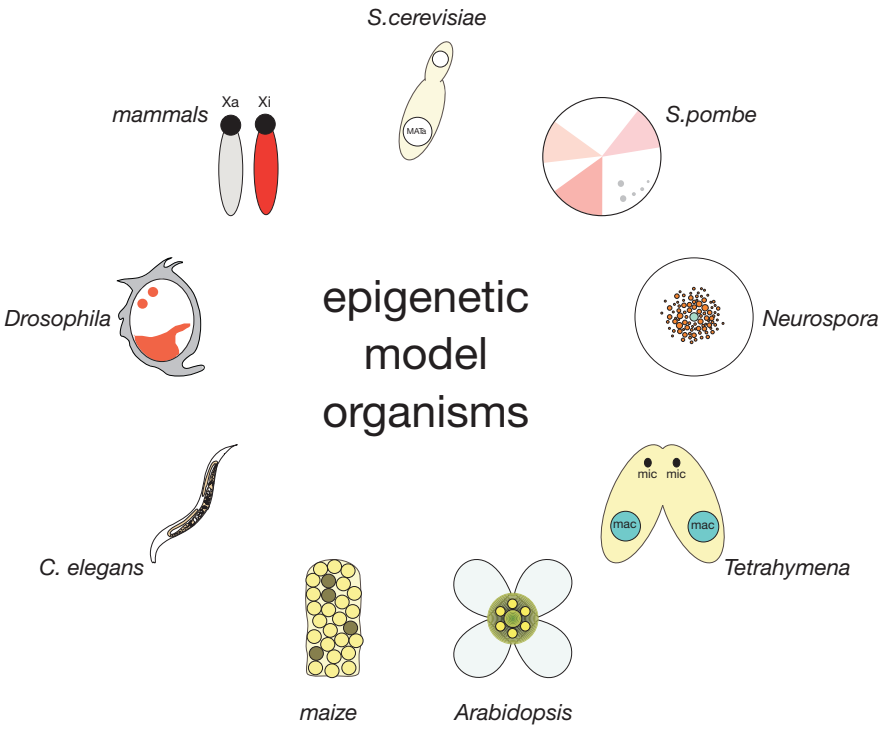

Figure 2. Model Organisms Used in Epigenetic Research

Schematic representation of model organisms used in epigenetic research. *S. cerevisiae*: Mating-type switching to study epigenetic chromatin control. *S. pombe:* Variegated gene silencing manifests as colony sectoring. *Neurospora crassa:* Epigenetic genome defense systems include repeat-induced point mutation, quelling, and meiotic silencing of unpaired DNA, revealing an interplay between RNAi pathways, DNA and histone methylation. *Tetrahymena:* Chromatin in somatic and germ-line nuclei are distinguished by epigenetically regulated mechanisms. *Arabidopsis:* Model for repression by DNA, histone, and RNA-guided silencing mechanisms. *Maize:* Model for imprinting, paramutation, and transposon-induced gene silencing. *C. elegans:* Epigenetic regulation in the germ line. *Drosophila:* Position-effect variegation (PEV) manifest by clonal patches of expression and silencing of the white gene in the eye. *Mammals:* X-chromosome inactivation.

Other "off-beat" organisms have also made disproportionate contributions toward unraveling epigenetic pathways that at first seemed peculiar. The fungal species, *N. crassa*, revealed the unusual non-Mendelian phenomenon of repeat-induced point mutation (RIP) as a model for studying epigenetic control (Chapter 6). Later, this organism was used to demonstrate the first functional connection between histone modifications and DNA methylation (Tamaru and Selker 2001), a finding later extended to "higher" organisms (Jackson et al. 2002). Ciliated protozoa, such as *Tetrahymena* and *Paramecium*, commonly used in biology laboratories as convenient microscopy specimens, facilitated important epigenetic discoveries because of their unique nuclear dimorphism. Each cell carries two nuclei: a somatic macronucleus that is transcriptionally active, and a germ-line micronucleus that is transcriptionally inactive. Using macronuclei as an enriched starting source of "active" chromatin, the biochemical purification of the first nuclear histone-modifying enzyme—a histone acetyltransferase or HAT—was made (Brownell et al. 1996). Ciliates are also well known for their peculiar phenomenon of programmed DNA elimination during their sexual life cycle, triggered by small noncoding RNAs and histone modifications (Chapter 7).

In multicellular organisms, genome size and organismal complexity generally increase from invertebrate (*Caenorhabditis elegans*, *Drosophila melanogaster*) or plant (*A. thaliana*) species to "higher," and to some, "more relevant," vertebrate organisms (mammals). Plants, however, have been pivotal to the field of epigenetics, providing a particularly rich source of epigenetic discoveries (Chapter 9) ranging from transposable elements and paramutation (McClintock 1951) to the first description of noncoding RNAs involved in transcriptional silencing (Ratcliff et al. 1997). Crucial links between DNA methylation, histone modification, and components of the RNAi machinery came through plant studies. The discovery of plant epialleles, with comic names such as SUPERMAN and KRYPTONITE (e.g., Jackson et al. 2002), and several vernalizing genes (Bastow et al. 2004; Sung and Amasino 2004) have further provided the research field with insights into understanding the developmental role of epigenetics and cellular memory. Plant meristem cells have also offered the opportunity to study crucial questions such as somatic regeneration and stem cell plasticity (see Chapters 9 and 11).

For understanding animal development, *Drosophila* has been an early and continuous genetic powerhouse. Based on the pioneering work of Muller (1930), many developmental mutations were generated, including the homeotic transformations and position-effect variegation (PEV) mutants explained below (also see Chapter 5). The homeotic transformation mutants led to the idea that there could be regulatory mechanisms for establishing and

maintaining cellular identity/memory which was later shown to be regulated by the Polycomb and trithorax systems (see Chapters 11 and 12). For PEV, gene activity is dictated by the surrounding chromatin structure and not by primary DNA sequence. This system has been a particularly informative source for dissecting factors involved in epigenetic control (Chapter 5). Over 100 suppressors of variegation [*Su(var)*] genes are believed to encode components of heterochromatin. Without the foundation established by these landmark studies, the discovery of the first histone lysine methyltransferases (HKMTs) (Rea et al. 2000) and the resultant advances in histone lysine methylation would not have been possible. As is often the case in biology, comparable screens have been carried out in fission yeast and in plants, identifying silencing mutants with functional conservation with the *Drosophila Su(var)* genes.

The use of reverse genetics via RNAi libraries in the nematode worm *C. elegans* has contributed to our understanding of epigenetic regulation in metazoan development. There, comprehensive cell-fate tracking studies, detailing all the developmental pathways of each cell, have highlighted the fact that Polycomb and trithorax systems probably arose with the emergence of multicellularity (see Sections 12 and 13). In particular, these mechanisms of epigenetic control are essential for gene regulation in the germ line (see Chapter 15).

The role of epigenetics in mammalian development has mostly been elucidated in the mouse, although a number of studies have been translated to diverse human cell lines and primary cell cultures. The advent of gene "knock-out" and "knock-in" technologies has been instrumental for the functional dissection of key epigenetic regulators. For instance, the Dnmt1 DNA methyltransferase mutant mouse provided functional insight for the role of DNA methylation in mammals (Li et al. 1992). It is embryonic-lethal and shows impaired imprinting (see Chapter 18). Disruption of DNA methylation has also been shown to cause genomic instability and reanimation of transposon activity, particularly in germ cells (Walsh et al. 1998; Bourc'his and Bestor 2004). There are approximately 100 characterized chromatin-regulating factors (i.e., histone and DNA-modifying enzymes, components of nucleosome remodeling complexes and of the RNAi machinery) that have been disrupted in the mouse. The mutant phenotypes affect cell proliferation, lineage commitment, stem cell plasticity, genomic stability, DNA repair, and chromosome segregation processes, in both somatic and germ cell lineages. Not surprisingly, most of these mutants are also involved in disease development and cancer. Thus, many of the key advances in epigenetic control took advantage of unique biological features exhibited by many, if not all, of the above-mentioned model organisms. Without these biological processes and the functional analyses (genetic and biochemical) that delved into them, many of the recent advances in epigenetic control would have remained elusive.

3 Defining Epigenetics

The above discussion begs the question, What is the common thread that allows diverse eukaryotic organisms to be connected with respect to fundamental epigenetic principles? Different epigenetic phenomena are linked largely by the fact that DNA is not "naked" in all organisms that maintain a true nucleus (eukaryotes). Instead, the DNA exists as an intimate complex with specialized proteins, which together comprise chromatin. In its simplest form, chromatin—i.e., DNA spooled around nucleosomal units consisting of small histone proteins (Kornberg 1974)—was initially regarded as a passive packaging molecule to wrap and organize the DNA. Distinctive forms of chromatin arise, however, through an array of covalent and non-covalent mechanisms that are being uncovered at a rapid pace (see Section 6). This includes a plethora of posttranslational histone modifications, energy-dependent chromatin-remodeling steps that mobilize or alter nucleosome structures, the dynamic shuffling of new histones (variants) in and out of nucleosomes, and the targeting role of small noncoding RNAs. DNA itself can also be modified covalently in many higher eukaryotes, by methylation at the cytosine residue, usually but not always, of CpG dinucleotides. Together, these mechanisms provide a set of interrelated pathways that all create variation in the chromatin polymer (Fig. 3).

Many, but not all, of these modifications and chromatin changes are reversible and, therefore, are unlikely to be propagated through the germ line. Transitory marks are attractive because they impose changes to the chromatin template in response to intrinsic and external stimuli (Jaenisch and Bird 2003), and in so doing, regulate the access and/or processivity of the transcriptional machinery, needed to "read" the underlying DNA template (Sims et al. 2004; Chapter 10). Some histone modifications (like lysine methylation), methylated DNA regions, and altered nucleosome structures can, however, be stable through several cell divisions. This establishes "epigenetic states" or means of achieving cellular memory, which remain poorly appreciated or understood. From this perspective, chromatin "signatures" can be viewed as a highly organized system of information storage that can index distinct regions

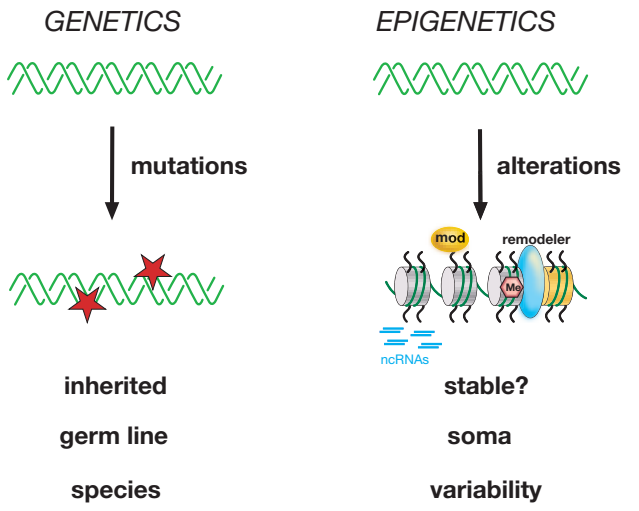

Figure 3. Genetics Versus Epigenetics

GENETICS: Mutations (*red stars*) of the DNA template (*green helix*) are heritable somatically and through the germ line. *EPIGENETICS:* Variations in chromatin structure modulate the use of the genome by (1) histone modifications (mod), (2) chromatin remodeling (*remodeler*), (3) histone variant composition (*yellow nucleosome*), (4) DNA methylation (*Me*), and (5) noncoding RNAs. Marks on the chromatin template may be heritable through cell division and collectively contribute to determining cellular phenotype.

of the genome and accommodate a response to environmental signals that dictate gene expression programs.

The significance of having a chromatin template that can potentiate the genetic information is that it provides multidimensional layers to the readout of DNA. This is perhaps a necessity, given the vast size and complexity of the eukaryotic genome, particularly for multicellular organisms (see Section 11 for further details). In such organisms, a fertilized egg progresses through development, starting with a single genome that becomes epigenetically programmed to generate a multitude of distinct "epigenomes" in more than 200 different types of cells (Fig. 4). This programmed variation has been proposed to constitute an "epigenetic code" that significantly extends the information potential of the genetic code (Strahl and Allis 2000; Turner 2000; Jenuwein and Allis 2001). Although this is an attractive hypothesis, we stress that more work is needed to test this and related provocative theories. Other alternative viewpoints are being advanced which argue that clear combinatorial "codes," like the triplet genetic code, are not likely in histones or are far from established (Schreiber and Bernstein 2002; Henikoff 2005). Despite these uncertainties, we favor the general view that a combination of covalent and non-covalent mechanisms will act to create chromatin states that can be templated through cell division and development by mechanisms that are just beginning to be defined. Exactly how these altered chromatin states are faithfully propagated during DNA replication and mitosis remains one of the fundamental challenges of future studies.

The phenotypic alterations that occur from cell to cell during the course of development in a multicellular organism were described by Waddington as the "epigenetic landscape" (Waddington 1957). Yet the spectrum of cells, from stem cells to fully differentiated cells, all share identical DNA sequences but differ remarkably in the profile of genes that they actually express. With this knowledge, epigenetics later came to be defined as the "Nuclear inheritance which is not based on differences in DNA sequence" (Holliday 1994).

Since the discovery of the DNA double helix and the early explanations of epigenetics, our understanding of epigenetic control and its underlying mechanisms has greatly increased, causing some to describe it in more lofty terms as a "field" rather than just "phenomena" (see Wolffe and Matzke 1999; Roloff and Nuber 2005; Chapter 1). In the past decade, considerable progress has been gained regarding the many enzyme families that actively modify chromatin (see below). Thus, in today's modern terms, epigenetics can be molecularly (mechanistically) defined as "The sum of the alterations to the chromatin template that collectively establish and propagate different patterns of gene expression (transcription) and silencing from the same genome."

4 The Chromatin Template

The nucleosome is the fundamental repeating unit of chromatin (Kornberg 1974). On the one hand, the basic chromatin unit consists of a protein octamer containing two molecules of each canonical (or core) histone (H2A, H2B, H3, and H4), around which is wrapped 147 bp of DNA. Detailed intermolecular interactions between the core histones and the DNA were determined from landmark studies leading to an atomic (2.8 Å) resolution X-ray picture of the nucleosome assembled from recombinant parts (Fig. 5) (Luger et al. 1997). Higher-resolution images of mononucleosomes, as well as emerging higher-order structures (tetranucleosomes) (Schalch et al. 2005), continue to capture our attention, promising to better explain the physiologically relevant substrate upon which most, if not all, of the chromatin remodeling and transcriptional machinery operates.

The core histone proteins that make up the nucleosome are small and highly basic. They are composed of a

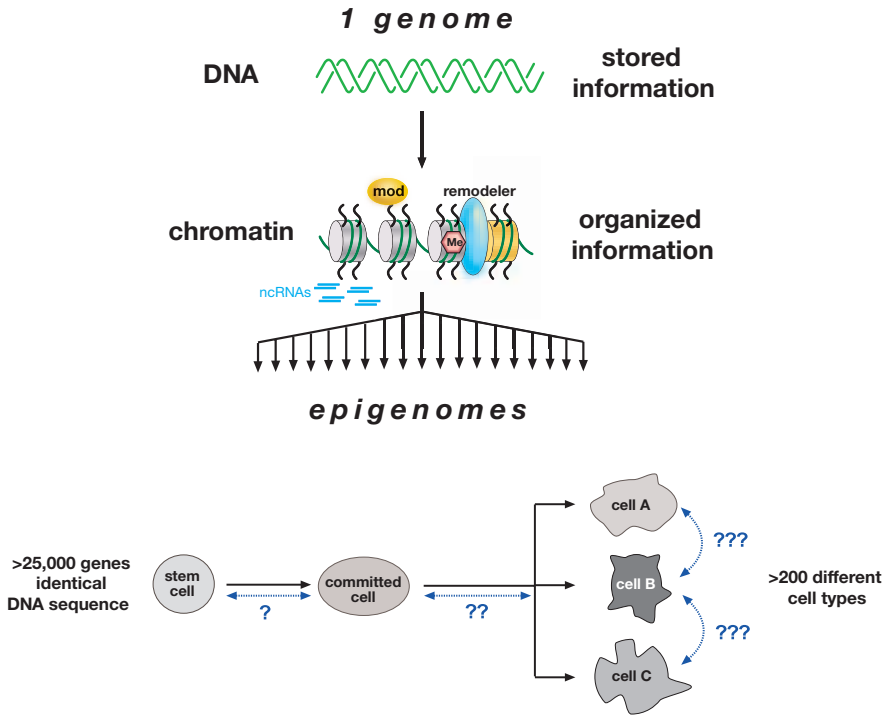

Figure 4. DNA Versus Chromatin

The genome: Invariant DNA sequence (*green double helix*) of an individual. The epigenome: The overall chromatin composition, which indexes the entire genome in any given cell. It varies according to cell type, and response to internal and external signals it receives. (*Lower panel*) Epigenome diversification occurs during development in multicellular organisms as differentiation proceeds from a single stem cell (the fertilized embryo) to more committed cells. Reversal of differentiation or transdifferentiation (*blue lines*) requires the reprogramming of the cell's epigenome.

globular domain and flexible (relatively unstructured) "histone tails," which protrude from the surface of the nucleosome (Fig. 5). Based on amino acid sequence, histone proteins are highly conserved from yeast to humans. Such a high degree of conservation lends support to the general view that these proteins, even the unstructured tail domains, are likely to serve critical functions. The tails, particularly of histones H3 and H4, in fact hold important clues to nucleosomal variability (and hence chromatin), as many of the residues are subject to extensive posttranslational modifications (see back end paper for standard nomenclature used in this textbook and Appendix 2 for a listing of known histone modifications).

Acetylation and methylation of core histones, notably H3 and H4, were among the first covalent modifications to be described, and were long proposed to correlate with positive and negative changes in transcriptional activity. Since the pioneering studies of Allfrey and coworkers (Allfrey et al. 1964), many types of covalent histone modifications have been identified and characterized; these include histone phosphorylation, ubiquitination, sumoylation, ADP-ribosylation, biotinylation, proline isomerization, and likely others that await description (Vaquero et al. 2003). These modifications occur at specific sites and residues, some of which are illustrated in Figure 6 and listed in Appendix 2. Specific enzymes and enzymatic complexes, some of which are highlighted in the following overview and individual chapters, catalyze these covalent markings. Because these lists will continue to grow in years to come, our intent was to mention only individual marks and enzymes that can illustrate what we feel are important general concepts and principles.

In certain chromatin regions, nucleosomes may contain histone variant proteins in place of a core (canonical) histone. Ongoing research is showing that this compositional difference contributes to marking regions of the

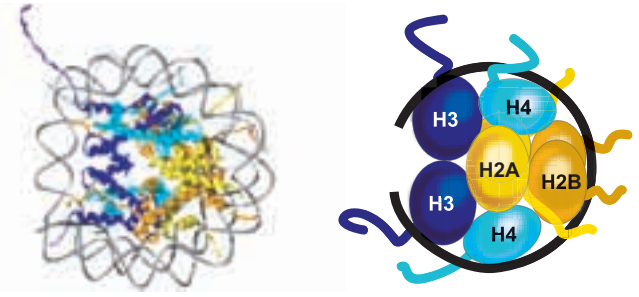

Figure 5. Nucleosome Structure

(*Left*) A 2.8 Å model of a nucleosome. (*Right*) A schematic representation of histone organization within the octamer core around which the DNA (*black line*) is wrapped. Nucleosome formation occurs first through the deposition of an H3/H4 tetramer on the DNA, followed by two sets of H2A/H2B dimers. Unstructured amino-terminal histone tails extrude from the nucleosome core, which consists of structured globular domains of the eight histone proteins.

chromosomes for specialized functions. Variant proteins for core histones H2A and H3 are currently known, but none exists for histones H2B and H4. We suspect that histone variants, although often minor in terms of amount and accordingly more difficult to study, are bountiful in the information they contain and essential to contributing to epigenetic regulation (for more detail, see Section 8 and Chapter 13).

5 Higher-Order Chromatin Organization

Chromatin, the DNA-nucleosome polymer, is a dynamic molecule existing in many configurations. Historically, chromatin has been classified as either euchromatic or heterochromatic, stemming from the nuclear staining patterns of dyes used by cytologists to visualize DNA. Euchromatin is decondensed chromatin, although it may be transcriptionally active or inactive. Heterochromatin can broadly be defined as highly compacted and silenced chromatin. It may exist as permanently silent chromatin (constitutive heterochromatin), where genes will rarely be expressed in any cell type of the organism, or repressed (facultative heterochromatin) in some cells during a specific cell cycle or developmental stage. Thus, there is a spectrum of chromatin states and a long-standing literature suggesting that chromatin is a highly dynamic macromolecular structure, prone to remodeling and restructuring as it receives physiologically relevant input from upstream signaling pathways. Only recently, however, has excellent progress been made unraveling molecular mechanisms that govern these remodeling steps.

The textbook, 11-nm "beads on a string" template represents an active and largely "unfolded" interphase configuration wherein DNA is periodically wrapped around repeating units of nucleosomes (Fig. 7). The chromatin fiber, however, is not always made up of regularly spaced nucleosomal arrays. Nucleosomes may be irregularly packed and fold into higher-order structures that are only beginning to be observed at atomic resolution (Khorasanizadeh 2004). Differential and higher-order chromatin conformations occur in diverse regions of the genome during cell-fate specification or in distinct stages of the cell cycle (interphase versus mitotic chromatin).

The arrangement of nucleosomes on the 11-nm template can be altered by *cis*-effects and *trans*-effects of covalently modified histone tails (Fig. 8). *cis*-Effects are brought about by changes in the physical properties of modified histone tails, such as a modulation in the electrostatic charge or tail structure that, in turn, alters internucleosomal contacts. A well-known example, histone acetylation, has long been suspected to neutralize positive charges of highly basic histone tails, generating a localized expansion of the chromatin fiber, thereby enabling better access of

Figure 6. Sites of Histone Tail Modifications

The amino-terminal tails of histones account for a quarter of the nucleosome mass. They host the vast majority of known covalent modification sites as illustrated. Modifications do also occur in the globular domain (*boxed*), some of which are indicated. In general, active marks include acetylation (*turquoise Ac flag*), arginine methylation (*yellow Me hexagon*), and some lysine methylation such as H3K4 and H3K36 (*green Me hexagon*). H3K79 in the globular domain has anti-silencing function. Repressive marks include H3K9, H3K27, and H4K20 (*red Me hexagon*), Green = active mark, red = repressive mark.

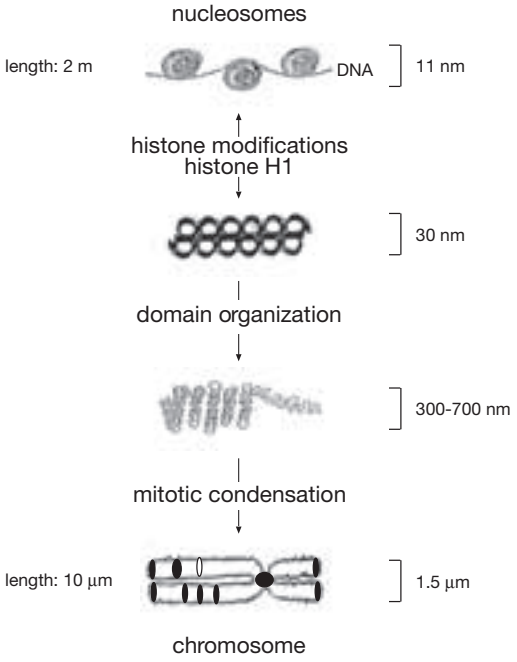

Figure 7. Higher-Order Structuring of Chromatin

The 11-nm fiber represents DNA wrapped around nucleosomes. The 30-nm fiber is further compacted into an as-yet-unconfirmed structure (illustrated as solenoid conformation here), involving linker histone H1. The 300–700-nm fiber represents dynamic higher-order looping that occurs in both interphase and metaphase chromatin. The 1.5-μm condensed chromosome represents the most compacted form of chromatin that occurs only during nuclear division (mitosis or meiosis). It is not yet clear how mitotic chromosome-banding patterns (i.e., G- or R-banding) correlate with particular chromatin structures.

transcription machinery to the DNA double helix. Phosphorylation, through the addition of net negative charge, can generate "charge patches" (Dou and Gorovsky 2000) that are believed to alter nucleosome packaging or to expose histone amino termini by altering the higher-order folded state of the chromatin polymer (Wei et al. 1999; Nowak and Corces 2004). In much the same way, linker histones (H1) are believed to promote the packaging of higher-order fibers by shielding the negative charge of linker DNA between adjacent nucleosomes (Thomas 1999; Khochbin 2001; Harvey and Downs 2004; Kimmins and Sassone-Corsi 2005). The addition of bulky adducts, such as ubiquitin and ADP-ribose, may also induce different arrangements of the histone tails and open up nucleosome arrays. The extent to which histone tails can induce chromatin compaction through modification-dependent and -independent mechanisms is not clear.

Histone modifications may also elicit what we refer to as *trans*-effects by the recruitment of modification-binding partners to the chromatin. This can be viewed as "reading" a particular covalent histone mark in a context-dependent fashion. Certain binding partners have a particular affinity and hence are known to "dock" onto specific histone tails and often do so by serving as the chromatin "Velcro" for one polypeptide within a much larger enzymatic complex that needs to engage the chromatin polymer. For instance, the bromodomain—a motif that recognizes acetylated histone residues—is often, but not always, part of a histone acetyltransferase (HAT) enzyme that exists to acetylate target histones (see Fig. 10 in Section 7) as part of a larger chromatin-remodeling complex (Dhalluin et al. 1999; Jacobson et al. 2000). Similarly, methylated lysine residues embedded in histone tails can be read by chromodomains (Bannister et al. 2001; Lachner et al. 2001; Nakayama et al. 2001) or similar domains (e.g., MBT, tudor) (Maurer-Stroh et al. 2003; Kim et al. 2006) to facilitate downstream chromatin-modulating events. In some cases, for instance, the association of chromodomain proteins precipitates the spreading of heterochromatin by the histone methyltransferase (HKMT)-catalyzed methylation of adjacent histones which can then be read by chromodomain proteins (Chapter 5).

Histone modifications of both the tail regions and the globular core region (Cosgrove et al. 2004) can also target ATP-dependent remodeling complexes to the 11-nm fiber required for the transition from poised euchromatin to a transcriptionally active state. This mobilization of nucleosomes may occur by octamer sliding, alteration of nucleosome structure by DNA looping (for more detail, see Chapter 12) or replacement of specific core histones with histone variants (Chapter 13). ATP-dependent chromatin remodelers (such as SWI/SNF, an historically important example) hydrolyze energy to bring about significant changes in histone:DNA contacts, resulting in looping, twisting, and sliding of nucleosomes. These non-covalent mechanisms have been shown to be critically important for gene regulatory events (Narlikar et al. 2002) as much as those involving covalent histone modifications (see Chapter 10). The finding that specific ATP-dependent remodelers can shuffle histone variants into and out of chromatin provides a means to link *cis*, *trans*, and remodeling mechanisms. Understanding, in turn, how these interconnected mechanisms act in a concerted fashion to vary epigenetic states in chromatin is far from complete.

More compact and repressive higher-order chromatin structures (30-nm) can also be achieved through the recruitment of linker histone H1 and/or modification-dependent or "architectural" chromatin-associated factors

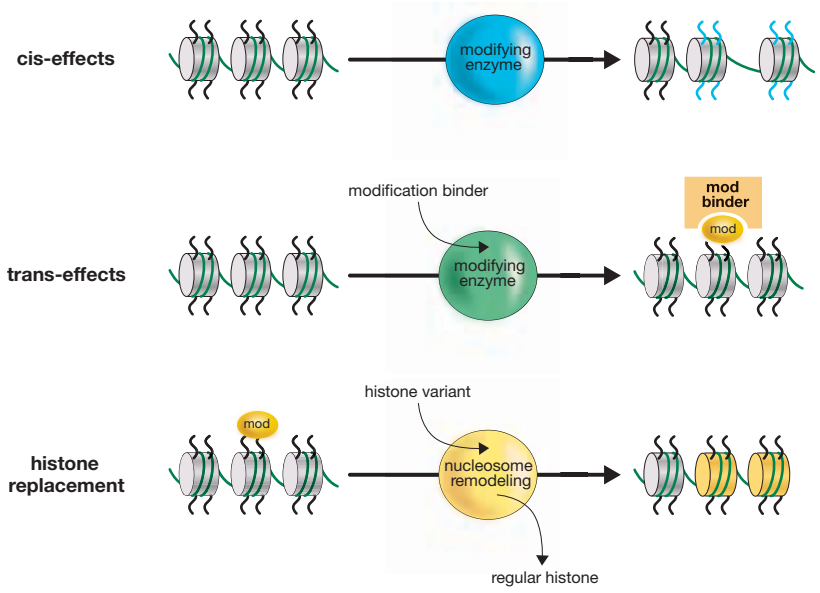

Figure 8. Transitions in the Chromatin Template (*cis/trans*)

cis-effects: A covalent modification of a histone tail residue results in an altered structure or charge that manifests as a change in chromatin organization. *trans-effects:* The enzymatic modification of a histone tail residue (e.g., H3K9 methylation) results in an affinity for chromatin-associated protein (mod binder, e.g., HP1). The association of a mod binder (or associated protein complexes) causes downstream alterations in chromatin structure. *Histone replacement:* A covalent histone modification (or other stimulus) can signal the replacement of a core histone with a histone variant through a nucleosome-remodeling exchanger complex.

such as heterochromatin protein 1 (HP1) or Polycomb (PC). Although it is commonly held that compaction of nucleosomal chromatin (11-nm) into a 30-nm transcriptionally incompetent conformation is accomplished by the incorporation of linker histone H1 during interphase, the functional and structural dissection of this histone has, until recently, been difficult (Fan et al. 2005). One likely problem underlying these studies is the fact that histone H1 occurs as different isoforms (~8 in mammals), making it difficult to do detailed genetic analyses. Thus, there is redundancy between some H1 isoforms whereas others may hold tissue-specific functions (Kimmins and Sassone-Corsi 2005). Interestingly, H1 itself can be covalently modified (phosphorylated, methylated, poly(ADP) ribosylated, etc.), raising the possibility that *cis* and *trans* mechanisms currently being dissected on core histones may well extend to this important class of linker histone, and also to non-histone proteins (Sterner and Berger 2000).

Considerable debate has taken place over the details of the way in which the 30-nm chromatin fiber is organized. In general, either "solenoid" (one-start helix) models, wherein the nucleosomes are gradually coiled around a central axis (6–8 nucleosomes/turn), or more open "zigzag" models, which adopt higher-order self-assemblies (two-start helix), have been described. New evidence, including that collected from X-ray structure using a model system containing four nucleosomes, suggests a fiber arrangement more consistent with a two-start, zigzag arrangement of linker DNA connecting two stacks of nucleosome particles (Khorasanizadeh 2004; Schalch et al. 2005). Despite this progress, we note that linker histone is not present in the current structures, and even if it were present, the 30-nm chromatin fiber compacts the DNA only approximately 50-fold. Thus, considerably more levels of higher-order chromatin organization exist that have yet to be resolved outside of light- and electron-microscopic examination, whether leading to interphase or mitotic chromatin states. Despite structural uncertainties, recent results in living cells have now established the existence of multiple levels of chromatin folding above the 30-nm fiber within interphase chromosomes. A noteworthy advance was the development of new approaches to label specific DNA sequences in live cells, making it possible to study the dynamics of chromatin opening and closing in vivo in real time. Interestingly, these results reveal a dynamic interplay of positive and negative chromatin-remodeling factors in setting higher-order chromatin structures for states more or less compatible with gene expression (Fisher and Merkenschlager 2002; Felsenfeld and Groudine 2003; Misteli 2004).

Organization into larger looped chromatin domains (300–700 nm) occurs, perhaps through anchoring the chromatin fiber to the nuclear periphery or other nuclear scaffolds via chromatin-associated proteins such as nuclear lamins. The extent to which these associations give rise to meaningful functional "chromosome territories" remains unclear, but numerous reports are showing that this concept deserves serious attention. For instance, clustering of multiple active chromatin sites to RNA polymerase II (RNA pol II) transcription factors has been observed, and similar concepts seem to apply to the clustering around replicating DNA and DNA polymerase. In

contrast, clustering of "silent" heterochromatin (particularly pericentromeric foci) and genes localized in *trans* has also been documented (see Chapters 4 and 21). How these associations are controlled and the extent to which nuclear localization of chromatin domains affects genome regulation are not yet clear. There is, nonetheless, an increasing body of evidence showing correlations of an active or silent chromatin configuration with a particular nuclear territory (Cremer and Cremer 2001; Gilbert et al. 2004; Janicki et al. 2004; Chakalova et al. 2005).

The most condensed DNA structure is observed during the metaphase stage of mitosis or meiosis. This permits the faithful segregation of exact copies of our genome (one or two copies of each chromosome, depending on the division at hand), via chromosomes, to each daughter cell. This condensation involves a dramatic restructuring of the DNA from a 2-m molecule when fully extended, into discrete chromosomes measuring on average 1.5 μm in diameter (Fig. 7). This is no less than a 10,000-fold compaction and is achieved by the hyperphosphorylation of linker (H1) and core histone H3, and the ATP-dependent action of the condensin and cohesin complexes, and topoisomerase II. Exactly how non-histone complexes engage mitotic chromatin (or M-phase chromatin modifications), and what rules dictate their association and release from chromatin in a cell-cycle-regulated fashion, remain to be determined (Bernard et al. 2001; Watanabe et al. 2001). Here, the well-known mitotic phosphorylation of histone H3 (i.e., serines 10 and 28) and members of the H1 family may provide important clues, but genetic and biochemical experiments have yet to yield full insights into what the function of these mitotic marks is. Interestingly, a formal theory has been proposed that specific methylation marks, when paired with more dynamic and reversible phosphorylation marks, may act as a "binary switch" in histone proteins, governing the binding and release of downstream effectors that engage the chromatin template (Fischle et al. 2003a). Using HP1 binding to methylated histone H3 on lysine 9 (H3K9me) and mitotic serine 10 phosphorylation (H3S10ph) as a paradigm, evidence in support of a mitotic "methyl/phos switch" has recently been provided (Daujat et al. 2005; Fischle et al. 2005; Hirota et al. 2005).

Specialized chromosomal domains, such as telomeres and centromeres, serve distinct functions dedicated to proper chromosome dynamics. Telomeres act as chromosomal ends, providing protection and unique solutions to how the very ends of DNA molecules are replicated. Centromeres provide an attachment anchor for spindle microtubules during nuclear division. Both of these specialized domains have a fundamental role in the events that lead to faithful chromosome segregation. Interestingly, both telomeric and centromeric heterochromatin is distinguishable from euchromatin, and even other heterochromatic regions (see below), by the presence of unique chromatin structures that are largely repressive for gene activity and recombination. Moving expressed genes from their normal positions in euchromatin to new positions at or near centromeric and telomeric heterochromatin (see Chapters 4–6) can silence these genes, giving rise to powerful screens described earlier that sought to identify suppressors or enhancers of position-effect variegation (PEV) or telomere-position effects (TPE; Gottschling et al. 1990; Aparicio et al. 1991). Centromeres and telomeres have molecular signatures that include, for example, hypoacetylated histones. Interestingly, centromeres are also "marked" by the presence of the histone variant CENP-A, which plays an active role in chromosome segregation (Chapter 14). Thus, the proper assembly and maintenance of distinct centromeric and pericentromeric heterochromatin is critical for the completion of mitosis or meiosis, and hence, cellular viability. In addition to the well-studied centromeric and pericentromeric forms of constitutive heterochromatin, progress is being made into mechanisms of epigenetic control for centromeric (and telomeric) "identity." Clever experiments have shown that "neocentromeres" can function in place of normal centromeres, demonstrating that DNA sequences do not dictate the identity of centromeres (Chapters 13 and 14). Instead, epigenetic hallmarks, including centromere-specific modification patterns and histone variants, mark this specialized chromosomal domain. Considerable progress is being made into how other coding, noncoding, and repetitive regions of chromatin contribute to these epigenetic signatures. How any of these mechanisms relate, if at all, to chromosomal banding patterns is not known, but remains an intriguing possibility. Achieving an understanding of the epigenetic regulation of these portions of unique chromosomal regions is needed, highlighted by the fact that numerous human cancers are characterized by genomic instability, which is a hallmark of certain disease progression and neoplasia.

6 The Distinction between Euchromatin and Heterochromatin

This overview has been divided into discussions of euchromatin and heterochromatin, although we acknowledge that multiple forms of both classes of chromatins exist. Euchromatin, or "active" chromatin, consists

largely of coding sequences, which only account for a small fraction (less than 4%) of the genome in mammals. What molecular signals then mark coding sequences with the potential for productive transcription, and how does chromatin structure contribute to the process? An extensive literature has suggested that euchromatin exists in an "open" (decompacted), more nuclease-sensitive configuration, making it "poised" for gene expression, although not necessarily transcriptionally active. Some of the genes are ubiquitously expressed (housekeeping genes); others are developmentally regulated or stress-induced in response to environmental cues. The cooperation of selected *cis*-acting DNA sequences (promoters, enhancers, and locus control regions), bound by combinations of *trans*-acting factors, triggers gene transcription in concert with RNA polymerase and associated factors (Sims et al. 2004). Together these factors have been highly selected during evolution to orchestrate an elaborate series of biochemical reactions that must occur in the appropriate spatial and temporal setting. Does chromatin provide an "indexing system" which better ensures that the above machinery can access its target sequences in the appropriate cell type?

At the DNA level, the AT-rich vicinity of promoters is often devoid of nucleosomes and may exist in a rigid noncanonical B-form DNA configuration, promoting transcription factor (TF) occupancy (Mito et al. 2005; Sekinger et al. 2005). However, TF occupancy is not enough to ensure transcription. The recruitment of nucleosome-remodeling machines, through the induction of activating histone modifications (e.g., acetylation and H3K4 methylation), facilitates the engagement of the transcription machinery by pathways that are currently being defined (Fig. 9 and Chapter 10). Exchange of displaced histones with histone variants after the transcription machinery has unraveled and transcribed the chromatin fiber ensures integrity of the chromatin template (Ahmad and Henikoff 2002). Achieving fully mature mRNAs, however, also requires posttranscriptional processes involving splicing, polyadenylation, and nuclear export. Thus, the collective term "euchromatin" likely represents a complex chromatin state(s) that encompasses a dynamic and elaborate mixture of dedicated machines that interact together and closely with the chromatin fiber to bring about the transcription of functional RNAs. Learning the "rules" as to how, in the most general sense, the "activating machinery" interacts with the transcription apparatus as well as the chromatin template is an exciting area of current research, although due to its dynamic nature, it may not strictly classify as epigenetics, but more as transcription and chromatin dynamics studies.

What then defines "heterochromatin?" Although it is historically less well studied than euchromatin, new insights suggest that heterochromatin plays a critically important role in the organization and proper functioning of genomes from yeast to humans (although *S. cerevisiae* has a distinct form of heterochromatin). Underscoring its potential importance is the fact that 96% of the mammalian genome consists of noncoding and repetitive sequences. New mechanistic insights, underlying the formation of heterochromatin, have revealed unexpected findings. For example, non-sequence-specific transcription, which produces double-stranded RNA (dsRNA), is subject to silencing by an RNA interference (RNAi)-like mechanism (see Section 10 below). The production of such dsRNAs acts as an "alarm signal" reflecting the fact that the underlying DNA sequence cannot generate a functional product, or has been invaded by RNA transposons or viruses. The dsRNA is then processed by Dicer and targeted to chromatin by complexes dedicated to initiating a cascade of events leading to the formation of heterochromatin. Using a variety of model systems, remarkable progress has been made dissecting what appears to be a highly conserved pathway leading to a heterochromatin "locked-down" state. Although the exact order and details may vary, this general pathway involves histone tail deacetylation, methylation of specific lysine residues (e.g., H3K9), recruitment of heterochromatin-associated proteins (e.g., HP1), and establishment of DNA methylation (Fig. 9). It is likely that sequestering of selective genomic regions to repressive nuclear domains or territories may enhance heterochromatin formation. Interestingly, increasing evidence suggests that heterochromatin may be the "default state," at least in higher organisms, and that the presence of a strong promoter or enhancer, producing a productive transcript, can override heterochromatin. Even in lower eukaryotes, the general concepts of heterochromatin assembly seem to apply. Hallmark features include hypoacetylated histone tails, followed by the binding of acetylation-sensitive heterochromatin proteins (e.g., SIR proteins; for details, see Chapter 4). Depending on the fungal species (e.g., budding vs. fission yeast), varying amounts of histone methylation and HP1-like proteins exist. Even though these genomes are more set to a general default state of being poised for transcription, some heterochromatin-like genomic regions are present (mating loci, telomeres, centromeres, etc.) that are able to suppress gene transcription and genetic recombination when test genes are placed in these new neighborhoods.

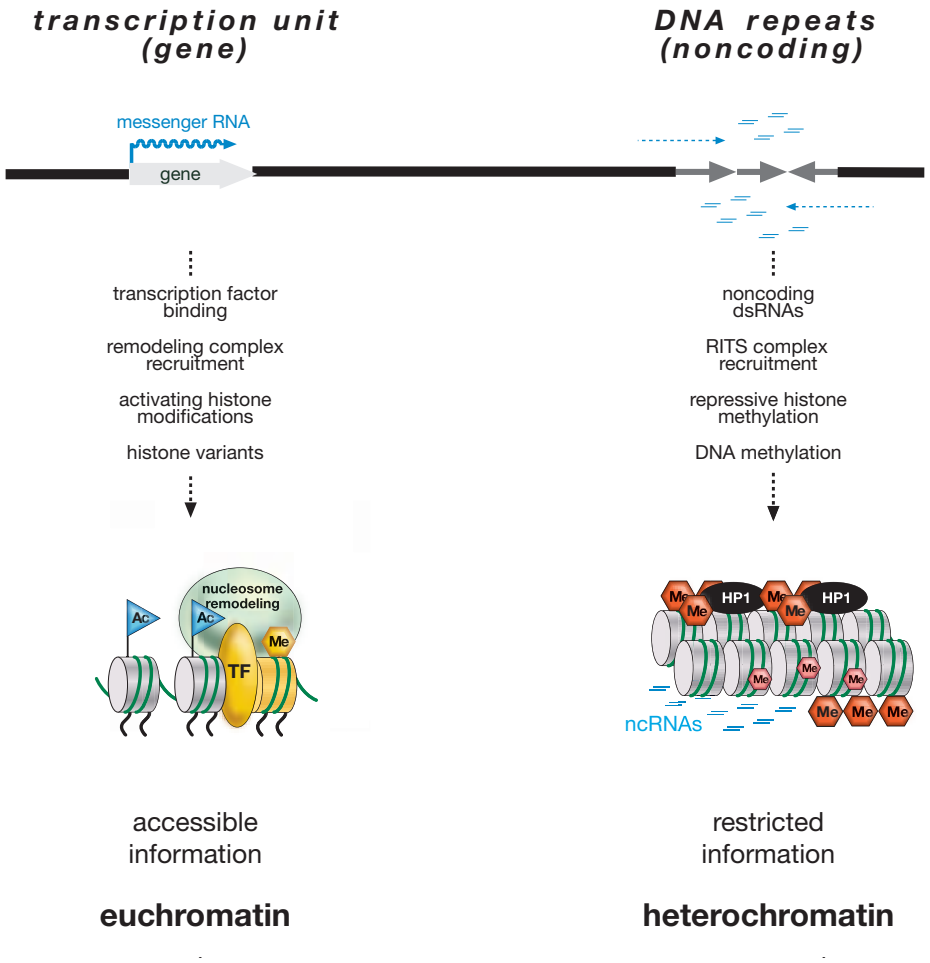

Figure 9. Distinction between Euchromatic and Heterochromatic Domains

Summary of common differences between euchromatin and constitutive heterochromatin. This includes differences in the type of transcripts produced, recruitment of DNA-binding proteins (i.e., transcription factor [TF]), chromatin-associated proteins and complexes, covalent histone modifications, and histone variant composition.

What useful functions might heterochromatin serve? The definition of centromeres, a region of constitutive heterochromatin, correlates well with a heritable epigenetic state and is thought to be evolutionarily driven by the largest clustering of repeats and repetitive elements on a chromosome. This partitioning ensures large and relatively stable heterochromatic domains marked by repressive "epigenetic signatures," facilitating chromosome segregation during mitosis and meiosis (Chapter 14). Here, it is noteworthy that centromeric repeats and the corresponding epigenetic marks that associate with them have been duplicated and moved onto other chromosome arms to create "silencing domains" in organisms such as fission yeast. Constitutive heterochromatin at telomeres (the protective ends of chromosomes) similarly ensures stability of the genome by serving as chromosomal "caps." Last, heterochromatin formation is known to be a defense mechanism against invading DNA. Collectively, these findings underscore a general view that heterochromatin serves important genome maintenance functions which may rival even that of euchromatin itself.

In summary, the broad functional distinction between euchromatin and heterochromatin can thus far be attributed to three known characteristics of chromatin. First is the nature of the DNA sequence—e.g., whether it contains AT-rich "rigid" DNA around promoters, repetitive sequences and/or repressor-binding sequences that signal factor association. Second, the quality of the RNA produced during transcription determines whether it is fully processed into an mRNA that can be translated, or whether the RNA is degraded or earmarked for use by the RNAi machinery to target heterochromatinization. Third, spatial organization within the nucleus can play a significant sequestering role for the maintenance of localized chromatin configurations.

7 Histone Modifications and the Histone Code

We have explored how histone modifications may change the chromatin template by *cis*-effects that alter internucleosomal contacts and spacing, or the *trans*-effects caused by histone and non-histone protein associations

with the template. What is the contribution and biological output of histone modifications? Patterns of chromatin structure that correlate with histone tail modifications have emerged from studies using bulk histones, suggesting that epigenetic marks may provide "ON" (i.e., active) or "OFF" (inactive) signatures. This has come through a long history of mostly correlative studies showing that certain histone modifications, notably histone acetylation, are associated with active chromatin domains or regions that are generally permissive for transcription. In contrast, other marks, such as certain phosphorylated histone residues, have long been associated with condensed chromatin that, in general, fails to support transcriptional activity. The histone modifications shown in Appendix 2 summarize the sites of modification that are known at this time. Here, we stress that these reflect modifications and sites that may well not be exhibited by every organism.

How are histone modifications established or removed in the first place? A wealth of work in the chromatin field has suggested that histone tail modifications are established ("written") or removed ("erased") by the catalytic action of chromatin-associated enzymatic systems. However, the identity of these enzymes eluded researchers for years. Over the last decade, a remarkably large number of chromatin-modifying enzymes have been identified from many sources, most of which are compiled in Appendix 2. This has been achieved through numerous biochemical and genetic studies. The enzymes often reside in large multi-subunit complexes that can catalyze the incorporation or removal of covalent modifications from both histone and non-histone targets. Moreover, many of these enzymes catalyze their reactions with remarkable specificity to target residue and cellular context (i.e., dependent on external or intrinsic signals). For clarity, and by way of example, we discuss briefly the four major enzymatic systems that catalyze histone modifications, together with their counterpart enzymatic systems that reverse the modifications (Fig. 10) (Vaquero et al. 2003; Holbert and Marmorstein 2005). Together, these antagonistic activities govern the steady-state balance of each modification in question.

Histone acetylases (HATs) acetylate specific lysine residues in histone substrates (Roth et al. 2001) and are reversed by the action of histone deacetylases (HDACs) (Grozinger and Schreiber 2002). The histone kinase family of enzymes phosphorylate specific serine or threonine residues, and the phosphatases (PPTases) remove phosphorylation marks. Particularly well known are the mitotic kinases, such as cyclin-dependent kinase or aurora kinase, which catalyze the phosphorylation of core (H3) and linker (H1) histones. Less clear in each case are the opposing PPTases that act to reverse these phosphorylations as cells exit mitosis.

Two general classes of methylating enzymes have been described: the PRMTs (protein arginine methyltransferases) whose substrate is arginine (Lee et al. 2005), and the HKMTs (histone lysine methyltransferases) that act on lysine residues (Lachner et al. 2003). Arginine methylation is indirectly reversed by the action of deiminases, which convert methyl-arginine (or arginine) to a citrulline residue (Bannister and Kouzarides 2005). Methylated lysine residues appear to be chemically more stable. Lysine methylation has been shown to be present in mono-, di-, or tri-methylated states. Several tri-methylated residues in the H3 and H4 animo termini appear to have the potential to be stably propagated during cell divisions (Lachner et al. 2004), as well as the H4K20me1 mark in *Drosophila* imaginal discs (Reinberg et al. 2004). Recently, a lysine-specific "demethylase" (LSD1) was described as an amine oxidase that is able to remove H3K4 methylation (Shi et al. 2004). The enzyme acts by FAD-dependent oxidative destabilization of the aminomethyl bond, resulting in the formation of unmodified lysine and formaldehyde. LSD1 was shown to be selective for the activating H3K4 methylation mark and can only destabilize mono- and di-, but not tri-methylation. This demethylase is part of a large repressive protein complex that also contains HDACs and other enzymes. Other evidence suggests that LSD1 can associate in a complex together with the androgen receptor at target loci and demethylate the H3K9me2 repressive histone mark to contribute to transcriptional activation (Metzger et al. 2005). A different class of histone demethylases has been characterized to work via a more potent mechanism—radical attack—known as hydroxylases or dioxygenases (Tsukada et al. 2006). One of these only destabilizes H3K36me2 (an active mark), but not in the tri-methyl state. This novel jumonji histone demethylase (JHDM1) contains the conserved jumonji domain, of which there are around 30 known in the mammalian genome, suggesting that some of these enzymes may also be able to attack other residues as well as a tri-methyl state (Fodor et al 2006; Whetstine et al 2006).

Considerable progress has been made in dissecting the enzyme systems that govern the steady-state balance of these modifications, and we suspect that much more progress will be made in this exciting area. It remains a challenge to understand how these enzyme complexes are regulated and how their physiologically relevant substrates

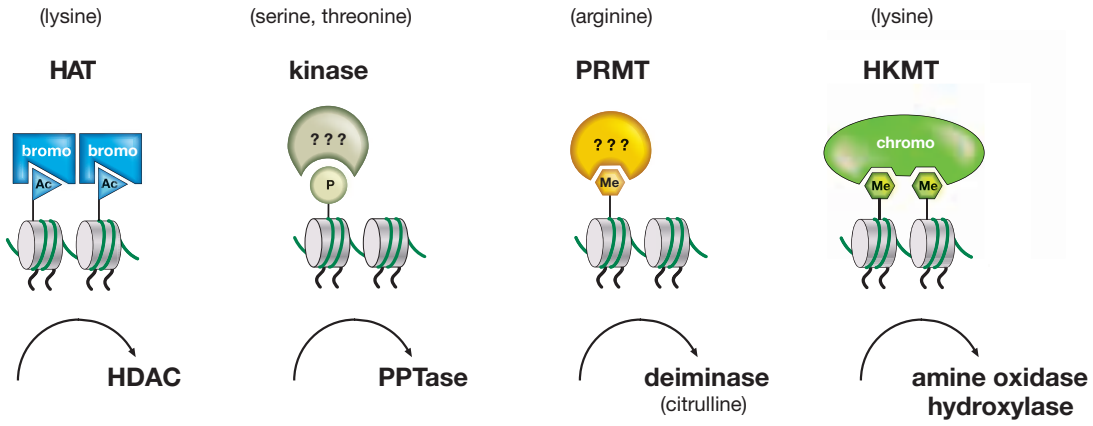

Figure 10. Histone-modifying Enzymes

Covalent histone modifications are transduced by histone-modifying enzymes ("writers") and removed by antagonizing activities. They are classified into families according to the type of enzymatic action (e.g., acetylation or phosphorylation). Protein domains with specific affinity for a histone tail modification are termed "readers." (HAT) Histone acetyltransferase; (PRMT) protein arginine methyltransferase; (HKMT) histone lysine methyltransferase; (HDAC) histone deacetylase; (PPTase) protein phosphatases; (Ac) acetylation; (P) phosphorylation; (Me) methylation.

and sites are targeted. In addition, it remains unclear how covalent mechanisms affect epigenetic phenomena.

Histone modifications do not occur in isolation, but rather in a combinatorial manner as proposed for modification cassettes (i.e., covalent modifications in adjacent residues of a particular histone tail, e.g., H3K9me and H3S10ph or H4S1ph, H4R3me, and H4K4ac) and *trans* histone pathways (covalent modifications between different histone tails or nucleosomes; see Fig. 11). Intriguingly, almost all of the known histone modifications correlate with activating or repressive function, dependent on which amino acid residue(s) in the histone amino termini is modified. Both synergistic and antagonistic pathways have been described (Zhang and Reinberg 2001; Berger 2002; Fischle et al. 2003b) that can progressively induce combinations of active marks, while simultaneously counteracting repressive modifications. It is, however, not known how many distinct combinations of modifications across the various amino-terminal histone positions exist for any given nucleosome, because most of the studies have been carried out on bulk histone preparations. In addition to the amino termini, modifications in the globular histone fold domains have recently been shown to affect chromatin structure and assembly (Cosgrove et al. 2004), thereby influencing gene expression and DNA damage repair (van Attikum and Gasser 2005; Vidanes et al. 2005). It is also worth noting that several of the histone-modifying enzymes also target non-histone substrates (Sterner and Berger 2000; Chuikov et al. 2004). Figure 11 illustrates two examples of established hierarchies of histone modifications that seem to index transcription of active chromatin or, in contrast, pattern heterochromatic domains.

These studies provoke the question of whether there is a "histone code" or even an "epigenetic code." Although this theoretical concept has been highly stimulating, and has been shown to be correct in some of its predictions, the issue as to whether a code actually exists has remained largely open. As a comparison, the genetic code has proven extremely useful, because of its predictability and near universality. It uses for the most part a four-base "alphabet" in the DNA (i.e., nucleotides), forming what is generally an invariant and nearly universal language. In contrast, current evidence suggests that histone-modification patterns are likely to vary considerably from one organism to the next, especially between lower and higher eukaryotes, such as yeast and humans. Thus, even if a histone code exists, it is not likely to be universal. This situation is made considerably more complicated when one considers the dynamic nature of histone modifications, varying in space and in time. Furthermore, the chromatin template engages a staggering array of remodeling factors (Vignali et al. 2000; Narlikar et al. 2002; Langst and Becker 2004; Smith and Peterson 2005). However, chromatin immunoprecipitation assays (ChIP), when examined on genome-wide levels (ChIP on chip), have begun to decipher nonrandom and somewhat predictable patterns in several genomes (e.g., *S. pombe*, *A. thaliana*, mammalian cells), such as strong correlations of H3K4me3 with activated promoter regions (Strahl et al. 1999; Santos-Rosa et al. 2002; Bernstein et al. 2005) and of H3K9 (Hall et al. 2002; Lippman et al. 2004; Martens et

al. 2005) and H3K27 (Litt et al. 2001; Ringrose et al. 2004) methylation with silenced heterochromatin. Perhaps the limitation of the histone code is that one modification does not invariantly translate to one biological output. However, modifications combinatorially or cumulatively do appear to define and contribute to biological functions (Henikoff 2005).

8 Chromatin-remodeling Complexes and Histone Variants

Another major mechanism by which transitions in the chromatin template are induced is by signaling the recruitment of chromatin "remodeling" complexes that use energy (ATP-hydrolysis) to change chromatin and nucleosome composition in a non-covalent manner. Nucleosomes, particularly when bound by repressive chromatin-associated factors, often impose an intrinsic inhibition to the transcription machinery. Hence, only some sequence-specific transcription factors and regulators (although not the basal transcription machinery) are able to gain access to their binding site(s). This accessibility problem is solved, in part, by protein complexes that mobilize nucleosomes and/or alter nucleosomal structure. Chromatin-remodeling activities often work in concert with activating chromatin-modifying enzymes and can generally be categorized into two families: the SNF2H or ISWI, and the Brahma or SWI/SNF family. The SNF2H/ISWI family mobilizes nucleosomes along the DNA (Tsukiyama et al. 1995; Varga-Weisz et al. 1997), whereas Brahma/SWI/SNF transiently alter the structure of the nucleosome, thereby exposing DNA:histone contacts in ways that are currently being unraveled (see Chapter 12).

Additionally, some of the ATP-hydrolyzing activities resemble "exchanger complexes" that are themselves dedicated to the replacement of conventional core histones with specialized histone "variant" proteins. This ATP-costing shuffle may actually be a means by which existing modified histone tails are replaced with a clean slate of variant histones (Schwartz and Ahmad 2005). Alternatively, recruitment of chromatin-remodeling complexes, such as SAGA (Spt-Ada-Gcn5-acetyltransferase) can also be enhanced by preexisting histone modifications to ensure transcriptional competence of targeted promoters (Grant et al. 1997; Hassan et al. 2002).

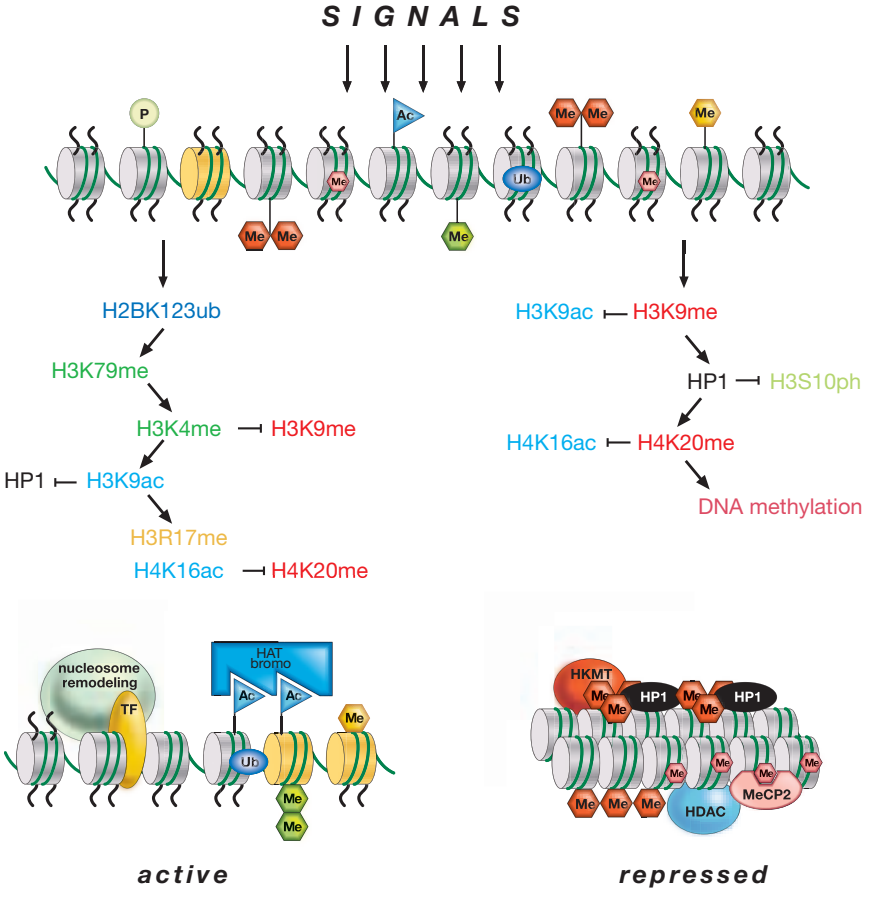

Figure 11. Coordinated Modification of Chromatin

The transition of a naïve chromatin template to active euchromatin (*left*) or the establishment of repressive heterochromatin (*right*), involving a series of coordinated chromatin modifications. In the case of transcriptional activation, this is accompanied by the action of nucleosome-remodeling complexes and the replacement of core histones with histone variants (*yellow*, namely H3.3).

In addition to transcriptional initiation and establishing the primary contact with a promoter region, the passage of RNA pol II (or of RNA pol I) during transcriptional elongation is further obstructed by the presence of nucleosomes. Mechanisms are therefore required to ensure the completion of nascent transcripts (particularly of long genes). In particular, a series of histone modifications and docking effectors act in concert with chromatin-remodeling complexes such as SAGA and FACT (for facilitate chromatin transcription) (Orphanides et al. 1998) to allow RNA pol II passage through nucleosomal arrays. These concerted activities will, for example, induce increased nucleosomal mobility, displace H2A/H2B dimers, and promote the exchange of core histones with histone variants. As such, they provide an excellent example of the close interplay between histone modifications, chromatin remodeling, and histone variant exchange to facilitate transcriptional initiation and elongation (Sims et al. 2004). Other remodeling complexes have also been characterized, such as Mi-2 (Zhang et al. 1998; Wade et al. 1999) and INO-80 (Shen et al. 2000), which are involved in stabilizing repressed rather than active chromatin.

Compositional differences of the chromatin fiber that occur through the presence of histone variants contribute to the indexing of chromosome regions for specialized functions. Each histone variant represents a substitute for a particular core histone (Fig. 12), although histone variants are often a minor proportion of the bulk histone content, and thus more difficult to study than regular histones. An increasing body of literature (for review, see Henikoff and Ahmad 2005; Sarma and Reinberg 2005) documents that histone variants have their own pattern of susceptibility to modifications, likely specified by the small number of amino acid changes that distinguish them from their family members. On the other hand, some histone variants have distinct amino- and carboxy-terminal domains with unique chromatin-regulating activity and different affinities to binding factors. By way of example, transcriptionally active genes have general histone H3 exchanged by the H3.3 variant, in a transcription-coupled mechanism that does not require DNA replication (Ahmad and Henikoff 2002). The replacement of core histone H2A with the H2A.Z variant correlates with transcriptional activity and can index the 5′ end of nucleosome-free promoters. However, H2A.Z has also been associated with repressed chromatin. CENP-A, the centromere-specific H3 variant, is essential for centromeric function and hence chromosome segregation. H2A.X, together with other histone marks, is associated with

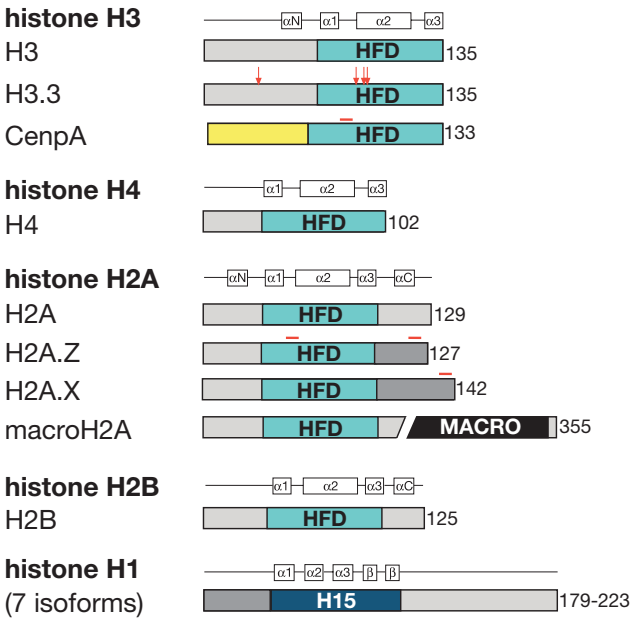

Figure 12. Histone Variants

Protein domain structure for the core histones (H3, H4, H2A, H2B), linker histone H1, and variants of histones H3 and H2A. The histone fold domain (HFD) where histone dimerization occurs, and regions of the protein that differ in histone variants (shown in *red*) are indicated.

sensing DNA damage and appears to index a DNA lesion for recruitment of DNA repair complexes. MacroH2A is a histone variant that specifically associates with the inactive X chromosome (Xi) in mammals (for more details on histone variants, see Chapter 13).

Importantly, and in contrast to the commonly held textbook notion that histones are synthesized and deposited only during S phase, synthesis and substitution of many of these histone variants occurs independently of DNA replication. Hence, the replacement of core histones by histone variants is not restricted to cell cycle stages (i.e., S phase), but can take immediate effect in response to ongoing mechanisms (e.g., transcriptional activity or kinetochore tension during cell division) or stress signals (e.g., DNA damage or nutrient starvation). Elegant biochemical studies have documented chromatin remodeling or exchanger complexes that are specific for replacement of distinct histone variants, such as H3.3, H2A.Z, or H2A.X (Cairns 2005; Henikoff and Ahmad 2005; Sarma and Reinberg 2005). For instance, replacement of H3 with the H3.3 variant occurs via the action of the HIRA (histone regulator A) exchanger complex (Tagami et al. 2004), and H2A is replaced by H2A.Z through the activity of the SWR1 (Swi2/Snf2-related ATPase 1) exchanger complex