Overview and Concepts

C. David Allis,\textsuperscript{1} Thomas Jenuwein,\textsuperscript{2} and Danny Reinberg\textsuperscript{3}

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

CONTENTS

1. Genetics Versus Epigenetics, 25
2. Model Systems for the Study of Epigenetics, 26
3. Defining Epigenetics, 28
4. The Chromatin Template, 29
5. Higher-Order Chromatin Organization, 31
6. The Distinction between Euchromatin and Heterochromatin, 34
7. Histone Modifications and the Histone Code, 36
8. Chromatin-remodeling Complexes and Histone Variants, 39
9. DNA Methylation, 41
10. RNAi and RNA-directed Gene Silencing, 42
11. From Unicellular to Multicellular Systems, 44
12. Polycomb and Trithorax, 45
13. X Inactivation and Facultative Heterochromatin, 47
14. Reprogramming of Cell Fates, 49
15. Cancer, 50
16. What Does Epigenetic Control Actually Do? , 52
17. Big Questions in Epigenetic Research, 55
References, 56
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 \(3 \times 10^9\) 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 discounts 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.
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).
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 (Ratcliffe 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 \( \text{Su(var)} \) 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 \textit{Drosophila} \text{Su(var)} genes.

The use of reverse genetics via RNAi libraries in the nematode worm \textit{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 reactivation 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 post-translational 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...
GENETICS
mutations

EPIGENETICS
alterations

inherited
germ line
species

stable?
soma
variability

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 Wolfe 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
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 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.
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](image)

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.
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-bind-
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 chromatin 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.
transcription unit
(gene)

DNA repeats
(noncoding)

messenger RNA

transcription factor
binding
remodeling complex
recruitment
activating histone
modifications
histone variants

noncoding
dsRNAs
RITS complex
recruitment
repressive histone
methylation
DNA methylation

accessible
information

restricted
information

accessible
information

restricted
information

euchromatin

heterochromatin

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 modify-
cations, and histone variant composition.

What useful functions might heterochromatin serve?
The definition of centromeres, a region of constitutive het-
rochromatin, 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 "epi-
genetic signatures," facilitating chromosome segregation
during mitosis and meiosis (Chapter 14). Here, it is note-
worthy that centromeric repeats and the corresponding
epigenetic marks that associate with them have been dupli-
cated 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, het-
rochromatin formation is known to be a defense mecha-
nism 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 attrib-
uted to three known characteristics of chromatin. First is
the nature of the DNA sequence—e.g., whether it con-
tains AT-rich "rigid" DNA around promoters, repetitive
sequences and/or repressor-binding sequences that signal
factor association. Second, the quality of the RNA pro-
duced 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 signifi-
cant 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 internu-
cleosomal 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 deiminas, 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 amino 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 amino-methyl 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 dioxygenases or hydroxylases (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; Whetstone 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
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; Chuklov 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
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 coordinatened 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 Ml-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 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.
DNA methylation is the oldest epigenetic mechanism known to correlate with gene repression (Razin and Riggs 1980). It is present to varying degrees in all eukaryotes except yeast. This modification consists of the addition of a methyl group at cytosine residues of the DNA template. It occurs at CpG dinucleotides in mammals, whereas other symmetric, asymmetric, and non-CpG methylation patterns are known in *N. crassa* and plants. The distribution of methylated DNA along the genome shows enrichment at noncoding regions (e.g., centromeric heterochromatin) and interspersed repetitive elements (transposons) but not in the CpG islands of active genes (Bird 1986). In fact, the increasing levels of DNA methylation correlate with a relative increase in noncoding and repetitive DNA content in the genomes of higher eukaryotes (see Fig. 15 in Section 11). Experimental evidence indicates that this is because DNA methylation serves mainly as a host defense mechanism to silence much of the genome of foreign origin (i.e., replicated transposable elements, viral sequences, and other repeated sequences).

DNA methyltransferases (DNMTs) are the “effectors” of DNA methylation, and catalyze either de novo (i.e., at novel sites) or maintenance methylation of hemimethylated DNA following DNA replication (see Chapter 18). Loss of the ability to maintain DNA methylation can result in several diseases, such as ICF (Immunodeficiency, Centromeric instability, and Facial abnormalities) (see Chapters 18 and 23). Deregulation in the levels of DNA methylation is also a contributing factor to cancer progression (Chapter 24).

What are the signals that direct DNMTs to methylate certain regions of DNA? Currently it is known that highly repetitive tandem repeat sequences of the genome (e.g., pericentromeric heterochromatin) rely on the repressive H3K9 methylation marks to direct DNA methylation de novo, as evidenced in *N. crassa* and plants (see Chapters 6 and 9). Interspersed repeats can also signal de novo DNA methylation, described in the context of RIP in *N. crassa* (Tamaru and Selker 2001) and retrotransposon silencing in the male germ line of mammals. A protein responsible for the latter has been identified—Dnmt3L—and may function by scanning the genome to identify high levels of homology–heterology junctions that signal the requirement for DNA methylation (Bourc'his and Bestor 2004). In plants, RNAs provide the signal for de novo DNA methylation, through a unique mechanism termed RNA-dependent DNA methylation (RdDM; see Chapter 9). There is evidence that chromatin remodelers of the SWI/SNF family are in some way necessary for the global patterns of DNA methylation, as demonstrated in plants (Jeddeloh et al. 1999) for the DDM1 protein and mammals by the Lsh1 homolog (Yan et al. 2003). Last, the HKMT-PcG protein Ezh2 may also be involved in directing DNA methylation at certain promoters in mammals (Vire et al. 2005).

Once established, the way in which DNA methylation may function to silence chromatin is not entirely clear, although evidence points to trans-regulation. Binders of methylated cytosines, called methyl-CpG-binding domain proteins (MBD), can be considered the DNA methylation equivalent to binders (or readers) of modified histone motifs (Fig. 13). For example, the methylcytosine-binding protein (MeCP2) binds methylated CpGs and recruits HDACs to mediate repressive histone marks (see Chapter 18). DNA methylation is also known to disturb the recognition sites of transcriptional regulators (e.g., CTCF) that are involved in genomic imprinting (see Chapter 19).

The existence of methylated DNA at imprinted loci that silence either the maternal or paternal allele in plants and placental mammals suggests that in the course of evolution they uniquely harnessed this epigenetic mechanism to stabilize gene repression. Interestingly, in marsupials, there is a lack of DNA methylation at imprinted loci, indicating that its involvement in mammalian imprinting is a relatively recent evolutionary event (see Chapters 17 and 19). Conversely, in dipteran insects such as *Drosophila*, DNA methylation has largely been lost as a functional epigenetic mechanism (Lyko 2001).

Highly repetitive regions of the mammalian genome that are typically methylated become increasingly mutagenic when unmethylated, to the extent of causing global genomic instability (Chen et al. 1998). Chromosomal abnormalities ensue, which are a major cause of many diseases and cancer progression (see Section 15). This underlines the crucial role that DNA methylation plays in genome integrity. Conversely, individual methylated cyto-
sine bases have a high propensity to spontaneously mutate. Thus, over time, C-T transitions occur through a deamination reaction (Fig. 13), but this characteristic is also thought to be beneficial for protecting the host genome because it permanently deactivates parasitic DNA sequences such as transposons. In a different context, this same chemical reaction is actively catalyzed by the activation-induced deaminase, or AID. Expression of the enzyme in B and T cells causes "somatic hypermutation" at the immunoglobulin (Ig) locus. This is an important mechanism for expanding the repertoire of antigen receptors and hence strengthening the immunity of mammals (Petersen-Mahrt 2005; Chapter 21). AID expression observed in early mammalian development has led to the suggestion that it may provide an alternative route to demethylating DNA, although this would happen at the risk of increased point mutation rates.

DNA methylation and histone methylation are prominent mechanisms for epigenetic regulation of the genome. Noncoding RNAs, as described in the next section, are important primary triggers for inducing silent chromatin. It is also known that RNA molecules can be heavily methylated at the sugar or nucleoside backbone. Moreover, methylation at the 3' end of small noncoding RNAs has been shown to stabilize these molecules (Yu et al. 2005). Intriguingly, Dnmt2 was recently identified as a tRNA methyltransferase (Goll et al. 2006). It is therefore plausible that, similar to DNMTs and HKMTs, RNA methyltransferases may exist as "writers" of epigenetic information, although there is no direct evidence for this. However, RNA methylation appears to be "sensed" by certain Toll-like receptors (transmembrane receptors that recognize common pathogen molecular motifs) to mediate innate immunity (Ishii and Akira 2005), corroborating such a hypothesis. This raises the interesting possibility that RNA methylation may yet prove to be a third form of methylation-based epigenetic modulation.

10 RNAi and RNA-directed Gene Silencing

The knowledge that constitutive heterochromatin at centromeres and telomeres plays an instructive role in genome integrity has contributed to a paradigm shift in the way that repetitive noncoding "junk" DNA is viewed. Is it possible that these repetitive sequences serve a nonwasteful purpose that is only beginning to be elucidated? Is it even possible that such DNA sequences are not completely "silent?"

This possibility has stemmed from a fundamental series of discoveries that linked RNAi to the formation of silent chromatin (heterochromatin). RNAi is a host defense mechanism that breaks down dsRNA species into small RNA molecules (known as short interfering RNA or siRNA). This process ultimately leads to RNA degradation or the use of the small RNAs to inhibit translation, known as posttranscriptional silencing (PTGS). The more recently discovered transcriptional gene silencing (TGS) mechanism, leading to heterochromatin formation, was discovered through the convergence of independent lines of investigation into chromatin and the RNAi machinery. On the one hand, much was known about repressive DNA methylation (in fungi, plants, and mammals), chromatin modifications (e.g., H3K9me3), and chromatin-associated factors (HP1) that are characteristic of heterochromatin domains. On the other hand, researchers were making headway in identifying factors of the RNAi machinery (e.g., Dicer, Argonaute, RNA-dependent RNA polymerase or RdRP). The most convincing progress that tied together these two seemingly divergent fields came from elegant studies in S. pombe, where mutations of any component of the RNAi machinery resulted in defects in chromosome segregation (Hall et al. 2002; Reinhart and Bartel 2002; Volpe et al. 2002). This was brought about by the inability to stabilize centromeric heterochromatin and underscored the likely widespread role of RNAi-mediated mechanisms in producing silent heterochromatin domains. It also highlighted the importance of heterochromatin beyond transcriptionally silencing.
genes, to a role in maintaining genome integrity and hence viability, as shown by the requirement of centromeric heterochromatin for the process of chromosome segregation. Emerging evidence also suggests that siRNAs are required in defining other specialized regions of functional heterochromatin, such as telomeres.

Transcription from both DNA strands of *S. pombe* pericentromeric repeats and the detection of processed siRNA derivatives provided strong evidence that the dsRNA derivative was the critical substrate to target the RITS complex to the centromeres for silencing (Fig. 14) (Verdel et al. 2004). Furthermore, *clr4* mutants (the *S. pombe* ortholog of mammalian *Swi39h HKMT*) failed to process dsRNA into siRNAs, strengthening the case for the interplay between the RNAi machinery and heterochromatin assembly (Motamedi et al. 2004). Exactly how siRNAs, generated by the RNAi machinery (i.e., Dicer, Argonaute, RdRP), initiate heterochromatin assembly or guide it to appropriate genomic loci is still unknown. A model has emerged in which a complex interaction between the RNAi machinery complex RITS and centromeric repeats leads to a self-reinforcing cycle of heterochromatin formation involving *Clr4*, HDACs, *Swi6* (ortholog to mammalian *HP1*), and cohesin, probably via Ago-directed annealing of RNA:RNA hybrids to the nascent transcript (Fig. 14) (see Chapters 6 and 8).

In *Tetrahymena*, a similar RNA-mediated targeting mechanism has been recognized to direct the unique case of DNA elimination that occurs in the somatic nucleus. In this case, transcription occurs from both strands of the internal eliminated segment (IES) sequences in the "silent," germ-line (micronuclear) genome at the appropriate stage of the sexual pathway (Chalker and Yao 2001; Mochizuki et al. 2002). Along the same lines as the RNAi-dependent TGS model, a scan RNA (scnRNA) model was proposed to explain how DNA sequences in the parental micronucleus can epigenetically control genomic alterations in the new micronucleus, involving small RNAs (for more detail, see Chapter 7). These exciting results provide the first demonstration of an RNAi-like process directly altering a somatic genome. This raises the intriguing possibility that intergenic RNAs produced at the V-DJ locus (Bolland et al. 2004) may potentially direct DNA sequence elimination during V-DJ recombination of the immunoglobulin heavy chain (IGH) locus in B cells and the T-cell receptor (TCR) loci in T cells.

In plants, there are a number of orthologs for many of the RNAi components, resulting in a variety of RNA silencing pathways that can act with greater specificity for particular DNA sequences, although there is some redundancy between factors. Studies of RNAi-mediated TGS in plants have revealed a novel class of RNA polymerases—RNA polymerase IV (or RNA pol IV)—that may transcribe DNA solely at heterochromatic regions (Herr et al. 2005; Pontier et al. 2005). Also unique to plants is the demonstration that RNAi pathways directly affect DNA methylation (Chan et al. 2004) (explained in detail in Chapter 9).

RNAi-like chromatin effects have also been uncovered in *Drosophila* and mammals. For instance, RNase A treatment of permeabilized mammalian cells rapidly removes heterochromatic H3K9me3 marks, suggesting that an RNA moiety may be a structural component of pericentromeric heterochromatin (Maison et al. 2002). Ablation of siRNA processing factors in vertebrates impairs H3K9 methylation and HP1 binding at pericentromeric heterochromatin (Fukagawa et al. 2004). Intriguingly, embryonic stem (ES) cells still proliferate, but fail to differentiate, in dicer-null mutants (Kanellopoulou et al. 2005), suggesting a currently not-understood connection between the RNAi machinery, noncoding RNAs, and mammalian development. In *Drosophila*, silencing of the tandem arrays of the *miniwhite* gene, subject to PEV also appears to be dependent on the RNAi machinery (Pal-Bhadra et al. 2004).
Collectively, these studies indicate a crucial, and probably primary, role for noncoding RNAs in triggering epigenetic transitions and heritably maintaining specific chromatin states of the chromatin template. In fact, these noncoding RNAs have provided the answer for how diverse repetitive sequences in different organisms achieve heterochromatinization through an RNA-targeted mechanism. In an effort to identify more targets of RNAi, the sequencing of small RNAs has revealed that they are largely transcribed from endogenous transposons and other repetitive sequences in plants, Drosophila, and mammals, among other organisms (Almeida and Allshire 2005; Bernstein and Allis 2005). Together, these results indicate that RNAi has evolved, in part, to maintain genomic stability by silencing mobile DNA elements and viruses, and is a conserved mechanism across most eukaryotic species. It now appears, however, not only that RNA silencing represses invading sequences, but also that this basic mechanism has been harnessed by the cell for the heterochromatinization of centromeres, thereby ensuring correct chromosome segregation and genome integrity.

Together, the above examples indicate a striking variation to the central dogma of gene control that is beginning to emerge as follows: DNA → noncoding RNA → chromatin → gene function. The idea that noncoding RNAs would actively participate in RNAi-like mechanisms which also target locus-specific domains for chromatin remodeling and gene silencing was never anticipated.

11 From Unicellular to Multicellular Systems

The 5,000–6,000 genes contained in the genomes of budding and fission yeasts are sufficient to regulate basic metabolic and cell division processes. There is, however, no requirement for cell differentiation, because these unicellular organisms are essentially clonal and, as such, repetitive “immortal” entities. In contrast, mammals code for ~25,000 genes required in ~200 different cell types. Understanding how multicellular complexity is generated and coordinated from the same genetic template is a key question in epigenetic research.

A comparison of the genome sizes between yeasts, flies, plants, and mammals indicates that genome size significantly expands with the complexity of the respective organism. There is a more than 300-fold difference between the genome sizes of yeast and mammals, but only a modest 4–5-fold increase in overall gene number (Fig. 15). However, the ratio of coding to noncoding and repetitive sequences is indicative of the complexity of the genome: The largely “open” genomes of unicellular fungi have relatively little noncoding DNA compared with the highly heterochromatic genomes of multicellular organisms. In particular, mammals have accumulated considerable repetitive elements and noncoding regions, which account for the majority of its DNA sequence (52% noncoding and 44% repetitive DNA). Only 4% of the mammalian genome thus encodes for protein function (including intronic sequences). This massive expansion of repetitive and noncoding sequences in multicellular organisms is most likely due to the incorporation of invasive elements, such as DNA transposons, retroposons, and other repetitive elements. Although these represent a burden for coordinated gene expression programs, they also allow genome evolution and plasticity, and a certain degree of stochastic gene regulation. The expansion of repetitive elements has even infiltrated the transcriptional units of the mammalian genome. This results in transcription units that are frequently much larger (30–200 kb), commonly containing multiple promoters and DNA repeats within untranslated introns. In contrast, plants, with similarly large genomes, generally possess smaller transcription units with smaller introns, because they have evolved defense mechanisms to ensure that transposon insertion within transcription units is not tolerated.

There are important organismal differences that manifest in the types of epigenetic pathways utilized despite high degrees of functional conservation for many mechanisms across species. Differences are, in part, believed to be related to genome size. The vast expansion of the genome with noncoding and repetitive DNA in higher eukaryotes requires more extensive epigenetic silencing mechanisms. This correlates with the fact that mammals and plants employ a full range of repressive histone lysine methylation, DNA methylation, and RNAi silencing mechanisms. Another challenge accompanying multicellularity is how to coordinate and maintain multiple cell types (cellular identity). This is a delicate balance, involving the Polycomb (PcG) and Trithorax (trxG) groups of protein complexes for genome regulation. The PcG proteins, in particular, correlate with the emergence of multicellularity (see Section 12).

Cells within multicellular organisms can be functionally divided into two major compartments: germ cells (totipotent and required for transmission of genetic information to the next generation) and somatic cells (the differentiated “powerhouse” of an organism). There are important questions of how the germ-cell compartment...
maintains totipotency of its epigenome and what mechanisms are involved in erasing, establishing, and maintaining cell fate (cell memory). Because one germ cell can give rise to another germ cell, it essentially has an infinite proliferative potential, as do unicellular “immortal” organisms. However, to fulfill this role, germ cells are for the most part “resting” and unresponsive to external stimuli, so that integrity of their epigenome can be protected. Indeed, mammalian oocytes can be retained in a resting state for more than 40 years. Similarly, adult stem cells (multipotent) are largely a dormant cell population, proliferating (and self-renewing) only when activated by mitogenic stimuli to enter a restricted number of cell divisions. Thus, the makeup of the epigenome is challenged by many intrinsic (e.g., transcription, DNA replication, chromosome segregation) and external (e.g., cytokines, hormones, DNA damage, or general stress responses) signals, particularly if somatic differentiation has forced cells to leave the protective germ-cell and stem-cell environment.

12 Polycomb and Trithorax

Among some of the main effectors that can transduce signals to the chromatin template and participate in maintaining cellular identity (i.e., provide cellular memory) are members of the PcG and trxG groups of genes (Ringrose and Paro 2004). These genes were discovered in Drosophila by virtue of their role in the developmental regulation of the Hox gene cluster and homeotic gene regulation. PcG and trxG have since been shown to be key regulators for cell proliferation and cellular identity in multicellular eukaryotes. In addition, these groups of genes are involved in several signaling cascades that respond to mitogens and morphogens; regulate stem cell identity and proliferation, vernalization in plants, homeotic transformations and transdetermination, lineage commitment during B- and T-cell differentiation, and many other aspects of metazoan development (see Chapters 11 and 12). We now briefly address what is known about how the PcG and trxG families of genes convert developmental cues into an "epigenetic memory" through chromatin structure.

The PcG and trxG groups of proteins function for the most part antagonistically: The PcG family of proteins establish a silenced chromatin state and the trxG family of proteins in general propagate gene activity. The molecular identification of the Pc gene known to stabilize patterns of gene repression over several cell generations provided the first evidence for a molecular mechanism for cellular or epigenetic memory. As well, PC provided an example of a chromodomain-containing protein with a high degree of similarity to the chromodomain of the heterochromatin-associated protein HP1 (Paro and Hog-
ness 1991). As mentioned above, chromodomains are well documented to be specific histone methyl-lysine binding modules (illustrated in Fig. 10).

Approximately 20 PcG genes and at least 15 distinct trxG genes have been identified in Drosophila. Functional analyses have shown that these groups of genes constitute a spectrum of diverse proteins yet are highly conserved between eukaryotes. PcG genes encode products that include DNA-binding proteins (e.g., YY1), histone-modifying enzymes (e.g., Ezh2), and other repressive chromatin-associated factors that contain a chromodomain with affinity for H3K27me3 (e.g., PC). trxG genes encode transcription factors (e.g., GAGA or Zeste), ATP-dependent chromatin-remodeling enzymes (e.g., Brahma), and HKMTs such as Ash1 and Trx (or its mammalian homologs MLL, Set1, and the MLL family). In most instances, the trxG and PcG families of proteins function as components of diverse complexes to establish stable chromatin structures that facilitate the expression or silencing of developmentally regulated genes (see Chapters 11 and 12).

Despite recent advances, the mechanism by which PcG- or trxG-containing complexes are targeted to developmentally regulated chromatin regions is not well understood. In Drosophila, heritable gene repression requires the recruitment of PcG protein complexes to DNA elements called polycomb response elements (PREs). Equivalent sequences in mammals have remained elusive. It is unclear how PcG protein complexes cause long-range silencing in a PRE-dependent manner, because PREs are usually located kilobases from the transcription start site of target genes. It can be postulated that repulsion or recruitment of PcG complexes may be discerned by changes in transcriptional activity, or differences in productive versus nonproductive mRNA processing (Pirrotta 1998; Dellino et al. 2004; Schmitt et al. 2005). Current models support PcG binding through interaction with DNA-binding proteins and the affinity of the chromodomain within the PC protein for H3K27me3-modified histones (Cao et al. 2002). However, PcG complexes can also associate in vitro with nucleosomes that lack histone tails (Francis et al. 2004), and furthermore, PRE elements have reduced nucleosome density (Schwartz et al. 2005). The most logical explanation for some of these disparate observations is that PcG binding in vivo would initially require interaction with DNA-bound factors that is then stabilized by association with nucleosomes and modified H3K27me3 in the adjacent chromatin region. Clearly, more research is needed to link existing evidence of how PcG complexes are targeted to regions of chromatin and how they mediate repression. This is likely to be organism-dependent, because there is great heterogeneity in the PcG complexes (see Chapter 11).

Trithorax group proteins maintain in general an active state of gene expression at target genes and overcome (or prevent) PcG-mediated silencing. This transition is even less well understood, but recent evidence suggests that an RNA-based mechanism could provide the trigger for the recruitment of Ash1 to target promoters (Sanchez-Elsner et al. 2006). A number of transient and stable changes in chromatin structure are thought to ensue, perhaps facilitated by intergenic transcription that can establish an open chromatin domain and mediate active histone replacement. Documented chromatin changes include the incorporation of "active" histone-lysine methylation marks by trxG HKMTs such as Trx and Ash1, and the reading of these marks (e.g., the WDR5 recognition of H3K4me; Wysocka et al. 2005). The action of trxG ATP-dependent chromatin-remodeling factors such as Brahma is also required, although how these mechanisms interrelate has yet to be fully determined (for more detail, see Chapter 12).

Many PcG and trxG proteins cooperate to maintain a tightly controlled level of repressed heterochromatin versus active euchromatin in a normal cell. In mammalian somatic interphase nuclei, the nuclear morphology reveals that constitutive domains of pericentromeric heterochromatin are grouped into 15–20 foci (see Fig. 16). Deregulation of cell fate and proliferation control, which leads to developmental abnormalities and cancer, frequently displays abnormal nuclear morphologies. For example, the nuclear organization in PML-leukemia (related to mixed lymphocyte leukemia [MLL]) cells shows an absence of pericentromeric foci (Di Croce 2005). In contrast, senescent (nonproliferating) cells display a nuclear morphology with large ectopic heterochromatin clusters (Narita et al. 2003; Scaffidi et al. 2005). Thus, nuclear morphology appears to be a good marker for distinguishing between normal and aberrant cell states, indicating that nuclear architecture may yet play a regulatory role in maintaining specialized domains of chromatin.

The study of histone modification levels is another indicator of cell normality or abnormality. Many of these changes are attributed to the deregulation of PcG (e.g., Ezh2) or trxG (e.g., MLL) HKMTs, contributing to the progression and even metastatic potential of a tumor (see Section 15). Indeed, the increase in overall levels of either of the above-mentioned proteins is associated with increased risk of prostate cancer, breast cancer, multiple myeloma, or leukemia (Lund and van Lohuisen 2004;
In other cases of neoplastic transformation, there is a manifest decrease in repressive histone marks and increase in overall acetylation states (Seligson et al. 2005) causing elevated levels of gene transcription and genomic instability. Clearly, changes in the global control of chromatin, possibly through perturbation of histone-modifying enzymes, affects the functionality of the genome and disrupts the proper gene expression profile of a normal cell.

In the case of cellular senescence, an increase in repressive histone marks is also an indicator of cellular dysfunction. This, concomitant with reduced definition of histone acetylation, can reinforce and even increase the levels of silent chromatin, blocking cellular plasticity and driving cells into an antiproliferative state (Scaffidi et al. 2005). This is largely an age-related effect, although the disease state, progeria, can prematurely advance aging. Conversely, when repressive pericentromeric methyl marks are decreased in mutants lacking the transducing enzyme (Suv39h), cells display increased rates of immortalization, no longer senesce, and show greater rates of genomic instability (Braig et al. 2005). These examples illustrate that chromatin deregulation, demonstrated by the levels of characteristic histone marks, often transduced by PcG and trxG enzymes, and nuclear morphology, is proving to be an important indicator of disease progression.

13 X Inactivation and Facultative Heterochromatin

PcG-mediated gene silencing and X-chromosome inactivation are prime examples for developmentally regulated transitions between active and inactive chromatin states (see Fig. 17), often referred to as facultative heterochromatin. This is in contrast to constitutive heterochromatin (at, e.g., pericentromeric domains), which may by default be induced at noncoding and highly repetitive regions. Facultative heterochromatin occurs at coding regions of the genome, where gene silencing is dependent on, and sometimes reversible by, developmental decisions specifying distinct cell fates.

One of the best-studied examples for facultative heterochromatin formation is the inactivation of one of the two X chromosomes in female mammals to equalize the dosage of X-linked gene expression with males that possess only one X (and a heteromorphic Y) chromosome (Chapter 17). Here, chromosome-wide gene silencing of

![Figure 16. Cellular Identity by PcG and trxG Proteins](image)

Two cell compartments are established during embryogenesis, distinguished by their differentiation potency: They are germ cells (totipotent) and somatic cells (including stem cells) with restricted differentiation potentials. The plasticity of a germ or stem cell's genome expression potential is reflected in reduced levels of repressive histone marks which are no longer visible at pericentromeric foci. Normal proliferating cells typically have a nuclear morphology showing 15-20 heterochromatic foci. Polycomb- and Trithorax-containing complexes operate in specifying the epigenetic and, hence, cellular identity of different lineages. They also function in response to external "stress" stimuli, promoting cellular proliferation and appropriate gene expression. Loss of genome plasticity and proliferation potential occurs in senescent (aging) cells, reflected by abnormally large heterochromatic foci and an overall increased level of repressive histone marks. Highly proliferating tumor cells, however, exhibit changes in the balance of repressive and activating histone marks through the deregulation of PcG and trxG histone-modifying enzymes. This is accompanied by perturbed nuclear morphology.
the inactive X chromosome (Xi) induces a high degree of Xi compaction that is visible as the Barr body, localized in the nuclear periphery of female mammalian cells. How the two alleles of the X chromosomes are counted and how one particular X chromosome is chosen for inactivation are challenging questions in today’s epigenetic research.

X inactivation involves a large (~17 kb) noncoding RNA, Xist, which appears to act as the primary trigger for chromatin remodeling at the Xi. Although there is the potential to form dsRNA between Xist and the antisense transcript Tsix (expressed only before the onset of X inactivation), no compelling evidence exists for RNA-dependent mechanisms being involved in the initiation of X inactivation. The X-inactivation center (XIC) and likely DNA “entry” or “docking” sites (postulated to be specialized repetitive DNA elements that are enriched on the X chromosomes) play a role for Xist RNA to associate and function as a scaffolding molecule, decorating the Xi in cis. Xist promotes the recruitment and action of both PRC1 (polycomb repressive complex) and PRC2 complexes, involved in establishing a stable inactive X chromosome. PRC2 components include, for example, the HKMT chromatin-modifying enzyme, EZH2, which catalyzes H3K27me3. PRC1 complex binding may be promoted by both H3K27me3 and histone-modification-independent means, whereas other components of the complex, such as the Ring1 proteins, ubiquitin H2A. Such is the heterogeneity of PcG complexes that different components can act independently of other complex components. The chromatin modifications, PcG complex binding, the subsequent incorporation of the histone variant macroH2A along the Xi, and extensive DNA methylation all contribute to generating a facultative heterochromatin structure along the entire Xi chromosome. Once a stable heterochromatic structure is established, Xist RNA is no longer required for its maintenance (Avner and Heard 2001; Heard 2005). A similar form of monoallelic silencing is genomic imprinting, which also uses a noncoding or antisense RNA to silence one allelic copy in a parent-of-origin-specific manner (Chapter 19). It is currently not clear whether and how Dicer-mutant mouse ES cells would affect the processes of X inactivation or genomic imprinting.

The general paradigm of dosage compensation, a classical epigenetically controlled mechanism, has also been addressed in other model organisms, notably C. elegans (Meyer et al. 2004; Chapter 15) and Drosophila (Gilfillan et al. 2004; Chapter 16). It is not yet clear whether dosage compensation occurs in birds, despite the fact that they are heterogametic organisms. In Drosophila, dosage compensation between the sexes occurs not by X inactivation in the female, but by a twofold up-regulation from the single X chromosome in the male. Intriguingly, two non-

---

**Figure 17. RNA Directed Induction of Repressed Chromatin States**

Different forms of silent chromatin have different primary signals, but many are likely to be RNA transcript-related (from aberrant transcripts, to Xist RNA, to dsRNAs), depending on the nature of the underlying DNA sequence. This triggers the establishment of a collection of chromatin changes, including a combination of histone modifications (H3K9, H3K27, and H4K20 methylation), the binding of repressive proteins or complexes (e.g., PC or HP1) to the chromatin, DNA methylation, and the presence of histone variants (e.g., macroH2A on the inactive X chromosome). Facultative or constitutive heterochromatin shows visible clustering in the nucleus. Euchromatic repression cannot be determined by nuclear morphology patterns.
coding RNAs, roX1 and roX2, are known to be essential components, and their expression is male-specific. Although similar mechanistic details probably exist between flies and mammals, it is clear that activating chromatin remodeling and histone modifications, notably MOF-dependent H4K16 acetylation on the male X chromosome, plays a key role in Drosophila dosage compensation. Exactly how histone-modifying activities, such as the MOF histone acetyltransferase, are targeted to the male X chromosome remains a challenge for future studies. Furthermore, ATP-dependent chromatin-remodeling activities, such as nucleosome-remodeling factor (NURF), are thought to antagonize the activities of the dosage compensation complex (DCC).

Together, this section and Sections 10 and 11 have described mechanisms for RNA-directed chromatin modifications, as they occur for constitutive heterochromatin, the Xi chromosome, and, possibly, also PcG-mediated gene silencing. On the basis of the intriguing parallels, one might postulate that an RNA moiety(s) or unpaired DNA would provide an attractive primary trigger for stabilizing PcG complexes at PREs or compromised promoter function, where they may "sense" the quality of transcriptional processing. Aberrant or stalled elongation and/or splicing errors could spur the interaction between PRE-bound PcG and a promoter, resulting in transcriptional shutdown. Thus, initiation of PcG silencing would be induced by the transition from productive to nonproductive transcription. The extent to which trxG complexes may utilize RNA quality control and/or processing of primary RNA transcripts as part of maintaining transcriptional "ON" states is beginning to be unraveled (Sanchez-Elsner et al. 2006).

14 Reprogramming of Cell Fates

The question of how cell fate can be altered or reversed has long intrigued scientists. The germ cell and early embryonic cells distinguish themselves from other cell compartments as the "ultimate" stem cell by their innate totipotency. Although cell-fate specification in mammals allows for around 200 different cell types, there are, in principle, two major differentiation transitions: from a stem (pluripotent) cell to a fully differentiated cell, and between a resting (quiescent or G₀) and a proliferating cell. These represent the extreme endpoints among many intermediates, consistent with a multitude of different makeups of the epigenome in mammalian development. During embryogenesis, a dynamic increase of epigenetic modifications is detected in the transition from the fertilized oocyte to the blastocyst stage, and then at implantation, gastrulation, organ development, and fetal growth. Most of these modifications or imprints may be erased via transfer of a differentiated cell nucleus to the cytoplasm of an enucleated oocyte. However, some marks may persist, thereby restricting normal development of cloned embryos, and a few could even be inherited as germ-line modifications (g-mod) (see Fig. 18), which, in mammals, are likely to include DNA methylation.

Liver regeneration and muscle cell repair are exceptions of mammalian tissues that can regenerate in response to damage or injury, although most other tissues are unable to be reprogrammed. In other organisms, such as plants and Axolotl, certain somatic cells can actually reprogram their epigenome and reenter the cell cycle to regenerate lost or damaged tissue (Tanaka 2003). In general, however, reprogramming of somatic cells is not possible unless they are engineered to recapitulate early development upon nuclear transfer (NT) into an enucleated oocyte. This was first demonstrated in cloned frogs (Xenopus), and more recently by the generation of Dolly, the first cloned mammal (Campbell et al. 1996; see Chapter 22).

Three major obstacles to efficient somatic reprogramming in mammals have been identified. First, certain somatic epigenetic marks (e.g., repressive H3K9me3) are stably transmitted through somatic cell divisions and resist reprogramming in the oocyte. Second, a somatic cell nucleus is unable to recapitulate the asymmetry of reprogramming that occurs in the fertilized embryo as a consequence of the differential epigenetic marks inherited by the male and female haploid genomes (see Mayer et al. 2000; van der Heijden et al. 2005; Chapter 20). Third, transmission of imprinted loci that are particularly important in fetal and placental development is not faithfully maintained upon NT (Morgan et al. 2005). Most cloned embryos abort, suggesting that perturbed epigenetic imprints represent a major bottleneck for normal development and could be the cause for the poor efficiencies of assisted reproductive technologies (ART) and the reduced vigor of cloned animals.

The use of embryonic stem cells versus somatic cells shows greatly enhanced reprogramming potential. The demonstration that quiescent cells (a frequent characteristic of stem cells) have a reduction in global H3K9me3 and H4K20me3 states could be a factor indicating enhanced plasticity of the epigenome (Baxter et al. 2004). This is also consistent with the fact that "immortal" unicellular organisms (e.g., yeast) with a largely open and active genome lack several repressive epigenetic mechanisms.
During the lifetime of an individual, epigenetic modifications (mod) are acquired in different cell lineages (left). Nuclear transfer (NT) of a somatic cell reverses the process of terminal differentiation, eradicating the majority of epigenetic marks (mod); however, some modification that would also be present in the germ line (g-mod) cannot be removed. During neoplastic transformation (from a normal to tumor cell), caused by a series of genetic mutations (red stars), epigenetic lesions accumulate. The epigenetic lesions (mod), but not the mutations, can be erased through reprogramming upon NT. This approach evaluates the interplay between genetic and epigenetic contributions to tumorigenesis. (Figure adapted from R. Jaenisch.)

Another feature of normal epigenetic reprogramming in mammals, postfertilization, is its distinct asymmetry. This can first be attributed to different programs of epigenetic specification in the male and female germ cells (Chapters 19 and 20). The sperm genome is largely made up of protamines, although there is a residual but significant level of CENP-A (an H3 histone variant) and other putative epigenetic imprints (Kimmins and Sassone-Corsi 2005), whereas the oocyte is made up of regular nucleosome-containing chromatin. Once fertilized, the sperm and oocyte haploid genomes have another cycle of reprogramming involving DNA demethylation and exchange of histone variants. The modifications can either enhance or balance epigenetic differences of the two parental genomes before nuclear fusion, in the first cell cycle. During differentiation of embryonic (i.e., inner cell mass [ICM]) and extraembryonic (i.e., trophectoderm [TE] and placenta) tissues, different DNA-methylation and histone-modification profiles are established between lineages (Morgan et al. 2005). Somatic cloning cannot faithfully recapitulate these patterns of reprogramming, showing rapid but less extensive demethylation in the first cell cycle, and perturbed DNA methylation and histone lysine methylation between ICM and TE cells.

A closely related concern in somatic cell reprogramming is the fate of imprinted gene loci. For normal embryonic development to proceed, correct allelic expression at imprinted loci is required (Chapter 19).

This was demonstrated by the seminal experiments that generated uniparental embryos (Barton et al. 1984; McGrath and Solter 1984; Surani et al. 1984). Androgenetic embryos (both genomes are of male origin) exhibited retarded embryonic development but hyperproliferation of extraembryonic tissues (e.g., placenta). In gyno- or parthenogenetic embryos (both genomes are of female origin), the placenta is underdeveloped. A parent-specific imprint must therefore be established in the germ cell following erasure of preexisting marks (Chapter 20). It is believed that this occurs for approximately 100 or more imprinted genes, largely involved in systems of resource provision for embryonic and placental development (e.g., Igf2 growth factor). Intriguingly, there is evidence that imprinting may be perturbed during in vitro culture of embryos produced by ART or nuclear transfer (Maher 2005).

15 Cancer

There is a delicate balance between self-renewal and differentiation. Neoplastic transformation (also similarly referred to as tumorigenesis) is regarded as the process whereby cells undergo a change involving uncontrolled cell proliferation, a loss of checkpoint control tolerating the accumulation of chromosomal aberrations and genomic aneuploidies, and mis-regulated differentiation (Lengauer et al. 1998). It is commonly thought to be caused by at least one genetic lesion, such as a point mutation, a deletion, or a translocation, disrupting either a tumor suppressor gene or an oncogene (Hanahan and Weinberg 2000). Tumor suppressor genes become silenced in tumor cells. Oncogenes are activated through dominant mutations or overexpression of a normal gene (proto-oncogene). Importantly, an accumulation of aberrant epigenetic modifications is also associated with tumor cells (see Chapter 24). The epigenetic changes involve altered DNA methylation patterns, histone modifications, and chromatin structure (see Fig. 19). Thus, neoplastic transformation is a complex multistep process involving the random activation of oncogenes and/or the silencing of tumor suppressor genes, through genetic or epigenetic events, and is referred to as the "Knudson two-hit" theory (Feinberg 2004; Feinberg and Tyko 2004). To illustrate, silencing of the retinoblastoma (Rb) gene, a tumor suppressor, causes loss of checkpoint control, which not only provides a proliferative advantage, but also promotes a "second hit" by affecting downstream functions related to chromatin structure which maintain genome integrity (Gonzalo and Blasco 2005). Inappro-
prial activation of an oncogenic product such as the myc
gene can have a similar effect (Knoepfler et al. 2006).

One question raised by current research is, To what
extent do aberrant epigenetic changes contribute to the
incidence and overall behavior of a tumor? This was
addressed by NT experiments using a melanoma cell
nucleus as the donor (Hochdorfer et al. 2004). Any
genetic lesions of the donor cell remain; however, NT
erases the epigenetic makeup. The tumor incidence of
cloned mouse fetuses was then studied, indicating that the
spectrum of tumors that arose de novo varied greatly, con-
sistent with different contributions of epigenetic modific-
ations in different tissues that trigger neoplastic progressi-
on.

DNA hypomethylation (as opposed to hypermethyla-
tion) can occur at discrete loci or over widespread chro-
mosomal regions. DNA hypomethylation was, in fact, the
first type of epigenetic transition to be associated with
cancer (Feinberg and Vogelstein 1983). This has turned
g out to be a widespread phenotype of cancer cells. At the
individual gene level, DNA hypomethylation can be neo-
plastic due to the activation of proto-oncogenes, the
derpression of genes that cause aberrant cell function,
or the biallelic expression of imprinted genes (also
 termed loss of imprinting or LOI) (see Chapters 23 and
24). On a more global genomic scale, broad DNA
hypomethylation, particularly at regions of constitutive
heterochromatin, predisposes cells to chromosomal
translocations and aneuploidies that contribute to cancer
progression. This effect is recapitulated in Dnmt1
mutants (Chen et al. 1998). The genomic instability that
ensues when there is DNA hypomethylation is due likely
to the mutagenic effect of transposon reactivation. With
attention turning to the essential role that repressive his-
tone modifications play in maintaining heterochromatin
at centromeres and telomeres, evidence has emerged that if
these marks are lost, genome instability also results, con-
tributing to cancer progression (Gonzalo and Blasco 2005).

Conversely, DNA hypermethylation is concentrated at
the promoter regions of CpG islands in many cancers.
Silencing of tumor suppressor genes through such aber-
rant DNA hypermethylation is particularly critical in can-
cer progression. Recent studies have revealed that there is
considerable cross talk between chromatin modifications
and DNA methylation, demonstrating that more than
one epigenetic mechanism can be involved in the silenc-
ing of a tumor suppressor gene. As an illustration, it is
known that the tumor suppressor genes, p16 and
hMLH1, are silenced by both DNA methylation and
repressive histone lysine methylation in cancer (McGar-
vey et al. 2006).

The deregulation of chromatin modifiers is implicated
in many forms of cancer. Certain histone-modifying
enzymes become oncogenic, such as the PcG protein
EZH2 and the trxG protein MLL, and exert their effect
through perturbing a cell’s epigenetic identity, which con-
sequently either transcriptionally silences or activates
inappropriate genes (Schneider et al. 2002; Valk-Lingbeek
et al. 2004). It is clear that the epigenetic identity is crucial
to cellular function. In fact, the pattern of global acetyl
and methyl histone marks is proving to be a hallmark for the
progression of certain cancers, as demonstrated by a study
in prostate tumor progression (Seligson et al. 2005).

![Diagram](image)

**Figure 19. Epigenetic Modifications in Cancer**

(a) Aberrant epigenetic marks at cancer-
causing loci typically involve the derepres-
sion of oncogenes or silencing of tumor
suppressor genes. Epigenetic marks known
to alter a normal cell include DNA methyl-
ation, repressive histone methylation, and
histone deacetylation. (b) The use of epi-
genetic therapeutic agents for the treat-
ment of cancer has consequences on the
chromatin template, illustrated for a tumor
suppressor locus. Exposure to Dnmt
inhibitors results in a loss of DNA methyl-
ation, and exposure to HDAC inhibitors
results in the acquisition of histone acetyl
marks and subsequent downstream modi-
fications, including active histone methyl
marks and the incorporation of histone
variants. The cumulative chromatin
changes lead to gene re-expression.
The development of drug targets inhibiting the function of the chromatin-modifying effector enzymes has opened up a new horizon for cancer therapeutics (see Fig. 19). The use of DNMT and HDAC inhibitors is in the most advanced stages of clinical trials in this new generation of cancer therapeutics. Zebularine and SAHA are, respectively, two such inhibitors. They are particularly beneficial for cancer cells that have repressed tumor suppressor genes (J.C. Cheng et al. 2004; Garcia-Manero and Issa 2005; Marks and Jiang 2005), because treatment leads to transcriptional stimulation. A major proportion of repressive histone lysine methylation is lost during treatment, most probably due to transcription-coupled histone exchange and nucleosome replacement; however, these inhibitors do not significantly alter H3K9me3 at target promoter regions (McGarvey et al. 2006). It remains to be resolved whether repressive marks that persist could induce subsequent re-silencing of tumor suppressor genes when treatment is paused, thereby counteracting the benefit of "epigenetic therapy." It is possible that a dual epigenetic therapy strategy, using DNMT and HDAC inhibitors, may promise a better prognosis in clinical trials.

Identification of inhibitors to other classes of histone modifiers, namely HKMTs and PRMTs, is currently in the development phase. There are approximately 50 SET domain HKMTs alone in the mammalian genome. Most of the well-characterized enzymes, such as SUV39H, EZH2, MLL, and RIZ, have already been implicated in tumor development (Schneider et al. 2002). Thus, high-throughput screens (HTS) are being employed in efforts to identify small-molecule inhibitors that could be used in exploratory research and, eventually, cancer therapy. All the classes of histone-modifying enzymes are suited for such an approach, as their specific substrate-binding sites (i.e., to histone peptides), in contrast to generic cofactor (e.g., acetyl-CoA and SAM) binding sites, would allow more selective drug development. HTS have been successful for HDACs (Su et al. 2000), PRMTs (D. Cheng et al. 2004), and HKMTs (Greiner et al. 2005).

For the transfer of knowledge to occur from basic to applied research, both hypothesis-driven and empirical approaches are required to ultimately define the efficacy and usefulness of any histone-modifying enzyme inhibitor. For instance, selective HKMT inhibitors against MLL or EZH2 may be valuable therapeutic agents for leukemia or prostate cancer. Alternatively, the use of a SUV39H HKMT inhibitor, which would seem counter-intuitive because of the necessity of this enzyme in maintaining constitutive heterochromatin and genome stability, may still preferentially sensitize tumor cells. In addition, analysis of the HDAC inhibitor SAHA has revealed that it may operate through additional pathways that are distinct from transcriptional reactivation (Marks and Jiang 2005). For example, HDAC inhibitors can also sensitize chromatin lesions, inhibiting efficient DNA repair and permitting genomic instabilities that can trigger apoptosis in tumor cells. These observations will have to be monitored when assessing the efficacy of dual combination therapies. Judging from the results to date, however, it is conceivable that combination therapy using HDAC and HKMT inhibitors may be more selective in killing pro-neoplastic cells by driving them into information overflow and chromatin catastrophe. It is hoped that continued research will identify the viable candidates for efficient epigenetic cancer therapy.

16 What Does Epigenetic Control Actually Do?

Approximately 10% of the protein pool encoded by the mammalian genome plays a role in transcription or chromatin regulation (Swiss-Prot database). Given that the mammalian genome consists of $3 \times 10^6$ bp, it must accommodate ~$1 \times 10^7$ nucleosomes. This gives rise to an overwhelming array of possible regulatory messages, including DNA-binding interactions, histone modifications, histone variants, nucleosome remodeling, DNA methylation, and noncoding RNAs. Yet, the process of transcriptional regulation alone is quite intricate, often requiring the assembly of large multiprotein complexes (>100 proteins) to ensure initiation, elongation, and correct processing of messenger RNA from a single selected promoter. If DNA sequence-specific regulation is so elaborate, one would expect the lower-affinity associations along the dynamic DNA–histone polymer to be even more so. On the basis of these considerations, rarely will there be one modification that correlates with one epigenetic state. More likely, and as experimental evidence suggests, it is the combination or cumulative effect of several (probably many) signals over an extended chromatin region that stabilizes and propagates epigenetic states (Fischle et al. 2003b; Lachner et al. 2003; Henikoff 2005).

For the most part, transcription factor binding is transient and lost in successive cell divisions. For persistent gene expression patterns, transcription factors are required at each subsequent cell division. As such, epigenetic control can potentiate a primary signal (e.g., promoter stimulation, gene silencing, centromere definition) to successive (but not indefinite) cell generations by the
heritable transmission of information through the chromatin template (Fig. 20). Interestingly, in *S. pombe*, Swi6-dependent epigenetic variegation can be suppressed for many cell divisions during both mitosis and meiosis (Grewal and Klar 1996) by histone modifications (most probably H3K9me2). Analogous studies were performed in *Drosophila* using a pulse of an activating transcription factor to transmit cellular memory for *Hox* gene expression during the female germ line (Cavalli and Paro 1999). In both of these examples, epigenetic memory is mediated by chromatin alterations that comprise distinct histone modifications and, most likely, also the incorporation of histone variants.

If histone modifications function together, an imprint may be left on the chromatin template that will help to mark nucleosomes, particularly if a signal is reestablished after DNA replication (Fig. 20). For even more stable inheritance, collaboration between histone modifications, histone variant incorporation, and chromatin remodeling will convert an extended chromatin region into persistent structural alterations that can then be propagated over many cell divisions. Although explained for the inheritance of transcriptional “ON” states, a similar synergy between repressive epigenetic mechanisms will more stably lock silenced chromatin regions, which is further reinforced by additional DNA methylation.

The DNA double helix can be viewed then as a self-organizing polymer which, through its ordering into chromatin, can respond to epigenetic control and amplify a primary signal into a more long-term “memory.” In addition, many histone modifications probably evolved in response to intrinsic and external stimuli. In keeping with this, chromatin-modifying enzymes require cofactors, such as ATP (kinases), acetyl-CoA (HATs), and SAM (HKMTs), whose levels are dictated by environmental changes (e.g., diet). Thus, the altered conditions can be translated into a more dynamic or stable DNA–histone polymer. An excellent example is the NAD-dependent HDAC, Sir2, which acts as “sensor” for nutrients and life span/aged cells (Guarente and Picard 2005; Rine 2005). Understanding how these environmental cues are cast into biologically relevant epigenetic signatures, and how they are read, translated, and inherited, lies at the heart of current epigenetic research. It is, however, important to stress that epigenetic control requires an intricate balance between many factors and that functional interaction is not always faithfully reestablished after each cell division. This is a functional contrast with genetics, which involves alteration of the DNA sequence, which is always stably propagated through mitosis and meiosis, if the mutation occurs in the germ line.

An important question arising from the above considerations is how the information contained in the chromatin is maintained from mother to daughter cells. If a cell loses its identity, through disease, misregulation, or reprogramming, is this identity loss accompanied by changes in chromatin structure? Bulk synthesis of most core histones is highly regulated during the cell cycle. Transcription of the core histone genes generally occurs during the S phase, the stage when DNA is replicated (replication coupled). This “coordination” assures that as the amount of DNA is doubled in the cell, there are sufficient core histones to be deposited onto the newly replicated DNA, and thus, the packaging of the DNA occurs simultaneously with DNA replication. As presented above, various regions of chromatin may have distinct differences in histone modifications that program the region to be either transcribed or not. How do domains of the newly synthesized daughter chromatin retain this crucial information for appropriate gene expression? How is the program faithfully templated from one cell generation to the next, or through meiosis and germ-cell formation (sperm and egg)? These central questions await future investigation.

Although initial studies indicated a semiconservative process, wherein a new H3/H4 tetramer is deposited, followed by the incorporation of two new H2A/H2B dimers, recent data have challenged this hypothesis. In this recent model, the “new” H3 and H4 polypeptides, which may already carry several posttranslational modifications, are incorporated as newly synthesized H3/H4 histone dimers together with the “old” H3/H4 dimers segregating between the mother and daughter DNA. If this is the case, then the modified, parental H3/H4 dimers would now also be present with the newly synthesized dimers on the same DNA. Their co-presence may then dictate that appropriate modifications are placed on the newly added dimers (Tagami et al. 2004). This model is attractive and might help explain the inheritance of histone modifications, and thus, the propagation of epigenetic information through DNA replication and cell division. However, more evidence is needed to support the validity of this or other intriguing models to explain the transmission of chromatin marks through cell division.

In closing this chapter, we ask, Does epigenetic control differ in a fundamental way from basic genetic principles? Although we may wish to view Waddington’s epigenetic landscape as being demarcated patches of activating versus repressive histone modifications along
Figure 20. Epigenetic Potentiation of a Primary Signal (Memory/Inheritance)

Classic genetics predicts that gene expression is dependent on the availability and binding of the appropriate panel of transcription factors (TF). Removal of such factors (i.e., a primary signal) results in the loss of gene expression, and thus constitutes a transient activating signal (top). Chromatin structure contributes to gene expression, where some conformations are repressive and others active. The activation of a locus may therefore occur through a primary signal and result in the downstream change in chromatin structure, involving active covalent histone marks (mod) and the replacement of core histones with variants (e.g., H3.3). Through cell division, this chromatin structure may only be reestablished in the presence of an activating signal (denoted "recurring signal"). Epigenetic memory results in the maintenance of a chromatin state through cell division, even in the absence of the primary activating signal. Such a memory system is not absolute, but involves multiple levels of epigenetic regulation for remodeling chromatin structure. The dynamic nature of chromatin means that although a chromatin state may be mitotically stable, it is nonetheless prone to change, hence affecting the longevity of epigenetic memory.

The continuum of the chromatin polymer, this notion could easily be overinterpreted. It is only in recent years that we have learned about the major enzymatic systems through which histone modifications might be propagated. This has shaped our current thinking about the stability, and hence the inheritance, of certain histone marks. In addition, it is underscored by the recent discoveries showing that mutations in chromatin-modifying activities, such as nucleosome remodelers (Cho et al. 2004; Mohrnan and Verrijzer 2005), DNMTs (Robertson 2005), HDACs or HMTs (Schneider et al. 2002), as they are frequently found in abnormal development and neoplasia, are telling examples of the ultimate power of genetic control. As such, tumor incidence in these mutant mice is generally regarded as a genetic disease. In contrast, alterations in nucleosome structure, DNA methylation, and histone modification profiles—that are not caused by a mutated gene—would classify as "true" epigenetic aberrations. Excellent examples of these more plastic systems are stochastic decisions in early embryonic development, reprogramming by nuclear transfer, transcriptional memory, genomic imprinting, mosaic X inactivation, centromere identity, and tumor progression. Genetics and epigenetics are thus closely related phenomena, and inherent to both is their propagation through cell division, which, for genetic control, also comprises the germ line, if mutations occur in germ cells. In the case of other—often too easily categorized—epigenetic modifications, we do not know whether they only reflect a minor and transient response to changes in the external environment or significantly contribute to phenotypic differences that can then be maintained over many, but not indefinite, somatic cell divisions, and occasionally affect the germ line. Even with our greatly improved knowledge of epigenetic mechanisms today, there is little, or no, novel support for Lamarckism.
17 Big Questions in Epigenetic Research

This book discusses the fundamental concepts and general principles that explain how epigenetic phenomena occur, as puzzling as they may seem. Our ultimate goal is to expose the reader to the current understanding of mechanisms that guide and shape these concepts, drawing upon the rich biology from which they emerge. In just a few years, epigenetic research has prompted exciting and remarkable insights and breakthrough discoveries, yet many long-standing questions remain unanswered (see Fig. 21). Although it is tempting to draw broad-brush conclusions and to propound general rules from this progress, we caution against this tendency, suspecting that there will be many exceptions that break the rules. For example, it is clear that striking organismal differences occur. Notably, from unicellular to multicellular organisms, the extent and type of histone modifications, histone variants, DNA methylation, and use of the RNAi machinery does vary.

There are, however, plenty of reasons for renewed energy in research programs designed to gain molecular insights into epigenetic phenomena. Elegant biochemical and genetic studies have already successfully dissected many of the functional aspects of these pathways, in an unprecedented manner. It could therefore be predicted that careful analysis of epigenetic transitions in different cell types (e.g., stem versus differentiated; resting versus proliferating) will uncover hallmarks of pluripotency (Bernstein et al. 2006; Boyer et al. 2006; Lee et al. 2006). This will most likely be valuable in diagnosing which chromatin alterations are significant during normal differentiation as compared with disease states and tumorigenesis. For example, using large-scale mapping approaches with normal, tumor, or ES cells—"epigenetic landscaping" along entire chromosomes (Brach et al. 2006b; Squazzo et al. 2006; Epigenomics AG, ENCODE, GEN-AU, EPIGENOME NoE)—it is anticipated that the knowledge generated could be harnessed for novel therapeutic intervention approaches and work toward promoting a worldwide consortium to map the entire human epigenome (Jones and Martienssen 2005). It is conceivable that differences in the relative abundance between distinct histone modifications, such as the apparent underrepresentation of repressive histone lysine tri-methylation in S. pombe and A. thaliana, may reflect the greater proliferative and regenerative potential in these organisms as compared to the more restricted developmental programs of metazoan systems. In addition, the functional links between the RNAi machinery, histone lysine methylation, and DNA methylation will continue to provide exciting surprises into the complex mechanisms required for cell-fate determination during development. Similarly, an enhanced understanding of the dynamics and specificity of nucleosome-remodeling machines will contribute to this end. We predict that more "exotic" enzymatic activities will be uncovered, catalyzing epigenetic transitions through modifications of histone and non-histone substrates. It would appear that chromatin alterations, as induced by the above mechanisms, act largely as a response filter to the environment. Thus, it is hoped that this knowledge can ultimately be applied to enhanced therapeutic strategies for resetting

---

Figure 21. Big Questions in Epigenetic Research

The many experimental systems used in epigenetic research have unveiled numerous pathways and novel insights into the mechanisms of epigenetic control. Many questions, as shown in the figure, still remain and require further elucidation or substantiation in new and existing model systems and methods.
some of an individual's epigenetic response that contribute to aging, disease, and cancer. This includes tissue regeneration, therapeutic cloning (using ES cells and their derivatives), and adult stem cell therapy strategies. It is believed such strategies will extend cellular life span, modulate stress responses to external stimuli, reverse disease progression, and improve assisted reproductive technologies. We predict that understanding the "chromatin basis" of pluripotency and totipotency will lie at the heart of understanding stem cell biology and its potential for therapeutic intervention.

Many fundamental epigenetic questions remain. For example, What distinguishes one chromatin strand from the other allele when both contain the same DNA sequence in the same nuclear environment? What defines the mechanisms conferring inheritance and propagation of epigenetic information? What is the molecular nature of cellular memory? Are there epigenetic imprints in the germ line that serve to keep this genome in a totipotent state? If so, how are these marks erased during development? Alternatively, or in addition, are new imprints added during development that serve to "lock in" differentiated states? We look forward to the next generation of studies (and students) bold enough to tackle these questions with the heart and passion of previous generations of genetic and epigenetic researchers.

In summary, the genetic principles described by Mendel likely govern the vast majority of our development and our outward phenotypes. However, exceptions to the rule can sometimes reveal new principles and new mechanisms leading to inheritance that have been underestimated, and in some cases, poorly understood previously. This book hopes to expose its readers to the newly appreciated basis of phenotypic variation—one that lies outside of DNA alteration. It is our hope that the systems and concepts described in this book will provide a useful foundation for future generations of students and researchers alike who become intrigued by the curiosities of epigenetic phenomena.

References


Brownell J.E., Zhou J., Ranalli T., Kobayashi R., Edmondson D.G.,


