Current Trends
in Science

PLATINUM JUBILEE SPECIAL

Divide and (epigenetic) rule: Chromatin domains as functional and structural units of genomes

RAKESH K MISHRA[1] and SANJEEV GALANDE[2]

[1]*Centre for Cellular and Molecular Biology, Uppal Road, Hyderabad, India.*
e-mail: mishra@ccmb.res.in
[2]*National Centre for Cell Science, Ganeshkhind, Pune, India.*
e-mail: sanjeev@nccs.res.in

Mapping and sequencing of genomes from a large number of evolutionarily diverse species in the past decade revealed that sequence *per se* is not sufficient to understand genome function, the higher-order organization of the genome and its various modifications are also important. In eukaryotic nuclei, genome is packaged by association with a number of basic proteins to form chromatin with nucleosomes as its repeating structural units. However, chromatin is highly heterogeneous at both micro and macro levels due to differential chemical modifications of DNA and histones, which can mark various functional states of chromatin. Distinct functional states ranging from 'highly active' to 'completely silenced' can be associated with specific nucleosome rearrangements, histone variants, histone post-translational modifications, and interactions of non-histone regulators. This projects a very dynamic scenario in which the environmental and cell-type specific signals can inflate the finite coding capacity of the genome into an epigenome with virtually infinite possibilities of combinations and regulations. Thus, chromatin structure has emerged as a key player in the transmission of heritable gene expression patterns. In this paper, we discuss emerging concepts of how DNA sequence can dictate chromatin organization at the domain level. Specific emphasis is put on the regulatory elements such as boundary elements, that mark the limits of chromatin domains and divide the genome into functional domains. The implications of these in development, differentiation, and disease are also discussed.

1. The Genome era

The Human Genome Project produced a reference sequence of the euchromatic human genome, which was the first major genome sequenced (International Human Genome Sequencing Consortium 2001, 2004; Venter *et al* 2001; McPherson *et al* 2001). Since then, a large number of genomes from evolutionarily diverse organisms have been sequenced. Interestingly, the estimated number of human genes has been repeatedly revised down from initial predictions of 100,000 or more as genome sequence quality and gene finding methods have improved. According to the current predictions, the haploid human genome contains about 20,000–25,000 protein-coding genes, far less than had originally been expected before its sequencing (International Human Genome Sequencing Consortium 2004). In fact, it is estimated that only about 1.5% of the human genome actually codes for proteins, while the rest consists of RNA genes, regulatory sequences, introns and (controversially) 'junk' DNA (International Human Genome Sequencing Consortium 2001; 2004). Another striking observation from the genome sequencing projects revealed, that the total number of genes in human genome was not more than twice that of many other simpler organisms, such as the roundworm and the fruit fly. Also, the genome sizes of several

Keywords. Epigenetics; genome organization; chromatin architecture; gene regulation.

lower eukaryotes are larger than that of humans (Gregory T R 2006 Animal Genome Size Database http://www.genomesize.com.). Thus, genome complexity does not necessarily reflect organismal complexity. However, human cells make extensive use of alternative splicing to produce several different proteins from a single gene, and the human proteome is believed to be much larger and more complex than those of the simpler organisms. Functional mapping of the genome has revealed that human genes are distributed unevenly across the chromosomes. Each chromosome contains various gene-rich and gene-poor regions, which seem to be correlated with chromosome bands and GC-content. The significance of these non-random patterns of gene density is yet to be studied, and it is believed that they may dictate the chromatin domain structures and higher-order assembly. In addition to protein coding genes, the human genome contains thousands of RNA genes, including tRNA, ribosomal RNA, microRNA, and other non-coding RNA genes (International Human Genome Sequencing Consortium 2001, 2004; Venter *et al* 2001).

Eukaryotic genomes are not as tidy as the genomes of prokaryotes. In higher eukaryotes, a large proportion of the genome does not code for any proteins (figure 1A). Composition of human genome based on different class of DNA sequences is as follows: coding $< 2\%$, retrotransposons 45%, minisatellites 7%, microsatellite/simple sequence repeats (SSRs) 3%. The remaining 42% is non-repetitive, heterogeneous DNA of unknown function. These non-coding DNA sequences have been referred to as 'junk', 'selfish' or 'parasitic' DNA as no functional significance of this part of the genome was obvious. 'Junk' DNA is a provisional label introduced 37 years ago by Susumu Ohno (Ohno S 1972), for the portions of the DNA sequence of a chromosome or a genome that has not been assigned any function. However, as we learn more about genomic organization, chromatin structure, nuclear architecture, maintenance of genomic information and gene regulation, it is revealed that a large proportion of the non-coding DNA is functional at various aspects of these processes and the so called 'junk' DNA is not so redundant after all.

Firstly the composition of eukaryotic genome would suggest that excess of DNA is often, not subject to a strong negative selection. Considering that a big proportion of these non-coding sequences is transcribed, it will be novice to think that in eukaryotic cell, where genes are regulated with extremely fine precision, accumulation of 'junk' DNA and even its transcription will not be subject to negative selection. On the contrary, accumulation and maintenance of these sequences is suggestive of advantages offered by these elements to the cell. For example, retroposition helps to maintain the complexity and fluidity of eukaryotic genomes by generating genes, pseudogenes, transposable elements, and novel combinations of DNA sequences. Resulting wealth of genetic variations serve as raw material for positive and negative selection and neutral drift.

As we examine the genomic organization of different organisms, it turns out that complexity of highly evolved organism is not reflected by the number of genes that they are made of. For example, worm (*C. elegans*) has more genes than flies (*Drosophila melanogaster*), although flies are relatively more evolved creatures and display far more complex body structures and behavior. Human genome consists of about 25000 genes. This is $\sim$1.5 fold the number of genes found in flies, although the human genome itself is $\sim$20 times bigger in size compared to that of flies (Gregory T R 2006 Animal Genome Size Database http://www.genomesize.com). This suggests that a higher number of non-coding DNA and fewer genes were incorporated in the genome of evolving organisms (figure 1B). It is also becoming increasingly clear that regulation of gene expression in higher eukaryotes is more complex, and that this complexity is achieved through epigenetic mechanisms, enabled by the additional non-coding part of the genome.

An inspection of the repetitive DNA also suggests that organization and composition of this part of non-coding DNA is not random, and hence has selection pressure suggesting a role for these sequences in the normal functioning of the cell. For example, certain kinds of SSRs are highly under represented, while others are in high abundance. Similarly, distribution of the association of the repetitive DNA with other non-coding DNA in the context of the genes on the chromosome has led to the belief that at least some of these sequences may serve as landmarks for higher-order regulatory mechanisms, regulating large number of genes during development. In cancer cells, a large number of genes are dysregulated and the transformation of a normal cell into cancer cell is often associated with deletion of large amount of the repetitive sequences (Thibodeau *et al* 1993). Systematic investigation of the vast quantity of sequence data in the human genome is currently a major thrust area of scientific inquiry, as its function remains unknown. The ENCODE project is the first such effort to understand the function of regulatory elements at a genome-wide level (The ENCODE Project Consortium 2007).

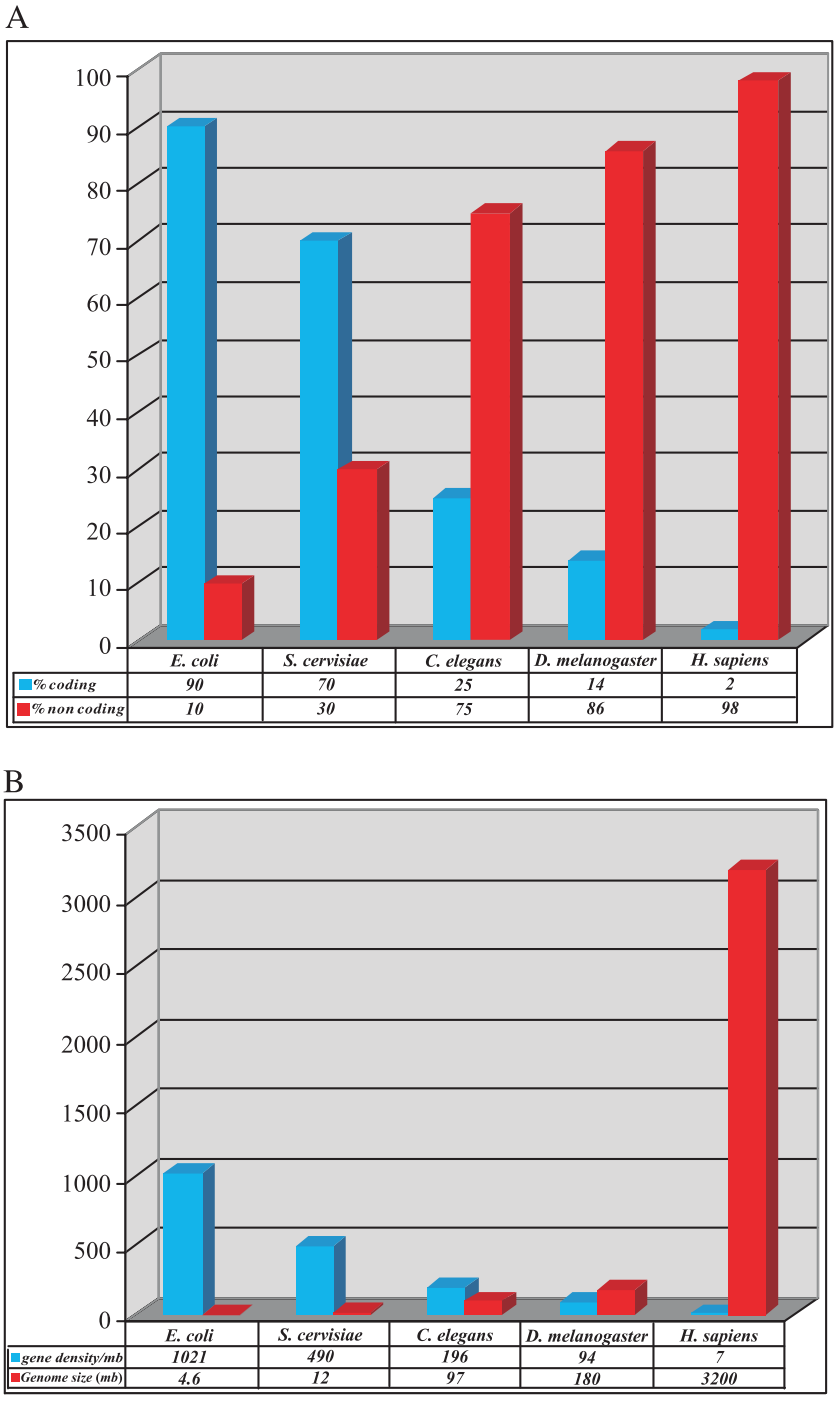

Figure 1. Genome size and organismal complexity are not directly correlated. Genome size is increased in complex organisms but this increase is not due to more number of genes or coding DNA. **(A)**. Graph depicting percentage of total genomic DNA in various model organisms plotted as coding vs non-coding. It can easily be seen that from simple to complex organisms, proportion of non-coding part of genome has increased, while that of coding part has decreased consistently. **(B)**. Graph depicting genome size of various model organisms and gene density of corresponding genomes. Number of genes also does not rise in proportion to the genome size. However, in general the gene density (number of genes/Mb of genome) falls dramatically from simple to complex organisms.

2. Higher-order chromatin organization of genome

In eukaryotes, the genomic DNA is packaged at a basal level by winding of the DNA around the histone octamer into nucleosomes in a 10 nm beads-on-a-string fiber, and at a second level by folding of the nucleoprotein-filament into 30 nm fibers. Each nucleosome contains a core histone octamer composed of histones H2A, H2B, H3,

and H4 (Luger *et al* 1997). This basic building block of chromatin is then packaged in a series of higher-order structures that give rise to functional entities. Further compaction of interphase chromatin is effected by formation of loops which operates independent of the basal level packaging. The looping of genomic DNA is a natural process, and is observed from simple prokaryotes to very complex organisms like humans. The nuclear substructure which provides a skeletal framework for periodic attachment of the 30 nm chromatin fiber to form loop structures has been termed as the nuclear matrix (reviewed by Galande 2007). The bases of loops that are attached to the nuclear matrix contain specialized DNA sequences called Matrix Attachment Regions (MARs) or Scaffold Attachment Regions (SARs). The formation of loop domains involving nuclear matrix and associated MARs and their dynamic reorganization has recently shown to be important for gene transcription (Cai *et al* 2006; Kumar *et al* 2007; Galande *et al* 2007). Most importantly, recent genome-wide studies have indicated that this kind of higher-order assembly reflects functional compartmentalization of the genome (The ENCODE Project Consortium 2007). The ENCODE project revealed that different chromatin features, particularly histone modifications correlate with chromatin structure, both over short and long distances.

3. Active versus inactive chromatin domains

Regulation of gene expression is undoubtedly one of the most intensely studied phenomena in modern biology. Post human genome sequence, it was quickly appreciated that primary DNA sequence is necessary, but not sufficient to understand how the genetic programs unfold. Thus, issues related with the structure and organization of the genome came into prominence. A layer of heritable 'epigenetic' information is superimposed on the primary DNA sequence that we have begun to understand and appreciate in recent years (Bernstein *et al* 2007). Understanding how eukaryotic genome is packaged into higher-order chromatin structures and what the functional consequences of this organization are, has begun to emerge recently (Spector 2003; Goetze *et al* 2007; Galande *et al* 2007). The epigenetic information is added as chemical modifications to cytosine bases in DNA, and to multiple amino acid residues in histone proteins that package the genome. These chemical modifications, impart their effect by regulating chromatin structure and DNA accessibility, and thereby orchestrate a diverse array of developmental stages, tissue types, and disease states (Bird 2007; Goll and Bestor 2005; Margueron *et al* 2005).

Recent years have witnessed tremendous progress in our ability to characterize the epigenetic modifications in both DNA and proteins at global scales. Such studies have also yielded underlying patterns in epigenetic mechanisms of genome regulation. The core histones are subjected to more than 100 different post-translational modifications (PTMs) including acetylation, phosphorylation, methylation, ubiquitination, sumoylation and poly(ADP)-ribosylation (Kouzarides 2007). Most of these modifications occur within the amino-terminal tails of histones that protrude out of the nucleosome (Luger *et al* 1997). Different histone modifications have diverse effects on transcription. Acetylation and methylation of lysines within the amino-terminal histone tails are the most widely studied PTMs, that reveal distinct distributions demarcating euchromatin and heterochromatin (Jenuwein and Allis 2001; The ENCODE Project Consortium 2007; The AACR Human Epigenome Task Force and EUNE 2008). Acetylation of lysines typically correlates with chromatin accessibility and transcriptional activity, whereas methylation of lysines can lead to contrasting effects depending upon the position of the lysine. Methylation of histone H3 lysine 9 (H3K9), H3 lysine 27 (H3K27), and histone H4 lysine 20 (H4K20) is usually associated with repression. In contrast, methylation of H3 lysine 4 (H3K4) and H3 lysine 36 (H3K36) is associated with transcriptionally active chromatin (Bernstein *et al* and Li *et al* 2007) (figure 2).

During interphase, chromatin can be classified into two functional and structural states, although the structural distinction is yet to be understood. Euchromatin or active chromatin, represents portion of the genome where the DNA is accessible due to relatively open nucleosome conformation. Genomic regions within the euchromatin compartment are more flexible, and contain genes in transcriptionally poised state, which indicates that they are not necessarily 'on', but can be activated if required (The ENCODE Project Consortium 2007). In contrast, genomic regions within the heterochromatin compartment are very rigid due to tight packaging. The high state of condensation renders chromatin inaccessible to transcription factors or other chromatin associated proteins (The ENCODE Project Consortium 2007; Talbert and Henikoff 2006; Huang *et al* 2004). Such regions are characteristically enriched in repetitive sequences and silenced genes associated with morphogenesis or differentiation (Reik 2007; Feinberg and Tycko 2004). Imprinted genes and X-chromosome inactivation are examples of the above category.

	Attribute	Euchromatin	Heterochromatin
Chromatin features	A. Structure	1. Less condensed 2. Open 3. Accesible	1. Highly condensed 2. Closed 3. Inaccesible
	B. DNA sequence	Gene-rich	Non-coding DNA Repetitive elements
	C. Activity	1. Expressed 2. Active	1. Repressed 2. Silent
Epigenetic markers	A. DNA methylation	Hypomethylation	Hypermethylation
	B. Histone acetylation	Hyperacetylation of histones H3, H4	Hypoacetylation of histones H3, H4
	C. Histone methylation	H3K4me2, H3K4me3, H3K9me1	H3K27me2, H3K27me3, H3K9me2, H3K9me3

Figure 2. The ABCs of epigenetics—Signature epigenetic modifications that demarcate euchromatin and heterochromatin. The plethora of chemical modifications in DNA, and proteins demarcate chromatin into two states depending upon its transcriptional activity — active or euchromatin and inactive or heterochromatin. Methylation can occur at one, two or three sites in the same lysine residue and are denoted as me1 (mono), me2 (di) and me3 (tri) respectively. Mono, di, and tri methylations are brought about by different methyltransferases and have diverse consequences (Shilatifard 2006). Figure adopted from *Pathways*^{TM} 2008 SABiosciences Corporation **8** 1–5.

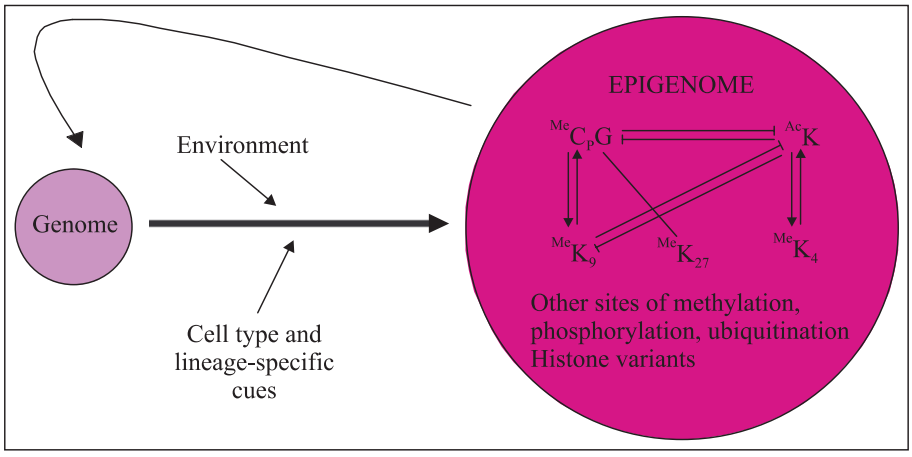

Figure 3. From genome to epigenome. Relatively finite coding capacity of the genome is converted into a virtually unlimited number of possibilities for coding various functions by using the combinatorial epigenetic marks. The making of the epigenome within any given cell type is a function of not only its genetic make-up, but to a large extent also on the environmental and lineage-specific cues. In such scenario, the plethora of chemical modifications in DNA, and proteome interact to form a complex regulatory network, that modulates chromatin structure and genome function (Margueron *et al* 2005; Bernstein *et al* 2007; Wang *et al* 2008). The dynamic interplay between histone and DNA modifications is also depicted as agonist and antagonist effects by arrows and lines, respectively. The genome is heavily marked with myriad of epigenetic modifications giving rise to epigenome, which in turn affects the genome such that the effects are heritable. (For details, see text.)

Complex modes of regulation arise from the combinatorial modifications of different histone residues and therefore give rise to the larger 'epi'genome (Wang *et al* 2008). This is possible because various histone modifications influence each other and may also be dependent on DNA methylation (Cedar and Bergman 2009) (figure 3). This is orchestrated by activities of protein complexes, that specifically bind modified histones or methylated cytosines (Li *et al* 2007).

4. The ENCyclopedia Of DNA Elements (ENCODE) project

To understand the human genome, and by extension of the biological processes it mediates and the ways in which its defects can give rise to disease, we need a more transparent and comprehensive view of the information it encodes. The Encyclopedia of DNA Elements (ENCODE) project was started in 2003 with an aim to provide a

more biologically informative representation of the human genome, by using high-throughput methods to identify and catalog the functional elements encoded. A consortium of scientists from 35 groups representing ten countries systematically analyzed 1% (∼30 Mb) of the human genome, by combination of microarray and sequencing approaches with computational analyses. This pilot project to systematically analyze regulatory regions in 1% of the genome revealed new insights into how the information coded in the DNA blueprint, is turned into functioning systems in the living cell (The ENCODE Project Consortium 2007). This preliminary study revealed that genome is much more than a mere vehicle for genes, and also provided insights into the elaborate molecular mechanisms that operate prior to gene expression (Greally 2007). The most striking findings of this project were — **(1)** The genome is extensively transcribed, substantially more regions of the genome are transcribed into RNA, than had previously been recognized. **(2)** Regulatory functions for DNA surrounding promoters were also found without any bias towards upstream regions. Interestingly, the epigenetic marks at the promoters differ from the non-promoter regulatory sequences indicating their distinct functions. **(3)** Chromatin accessibility and histone modification patterns are highly predictive of both the presence and activity of transcription start sites. **(4)** Distal DNaseI hypersensitive sites have characteristic histone modification patterns that reliably distinguish them from promoters. Interestingly, some of these distal sites show marks consistent with insulator function. **(5)** The data further corroborated the chromatin domain hypothesis, that the genome is subdivided into functional entities by virtue of larger scale organization similar to that observed during DNA replication (White *et al* 2004). The precise overlay of histone modifications with DNA replication data, revealed that DNA replication timing is correlated with chromatin structure. Indeed, the coordinated ENCODE approach successfully correlated DNA replication with higher-order patterns of the organization of transcriptional regulators in the same cell-types. Thus, early-replicating regions are marked with histone modifications associated with gene activation and late-replicating regions are enriched in repressive modifications (The ENCODE Project Consortium 2007).

5. Transcending from genome to epigenome

Embryonic developmental program that starts with a single cell, the zygote, leads to differentiation into various cell types that work coordinately to form a living organism. For example, identical genome in human gives rise to more than 200 types of terminally differentiated cells. Each of these cell types is the consequence of specific genome packaging leading to cell type specific chromatin structures and gene expression profile. One single genome, therefore, gives rise to a number of epigenomes during embryonic development that sum up into a huge epigenome (figure 3). Various external cues such as environmental factors and cell-type or lineage specific signals, also contribute in a similar manner to the epigenome (Bernstein *et al* 2007). Various features of the human genome that transcend its primary DNA sequence, such as chromatin packaging, histone modifications, and DNA methylation, are important in regulating gene expression, genome replication and other cellular processes are collectively referred to as epigenetic marks (Misteli 2007; Bernstein *et al* 2007). For years, the link between DNA methylation and histone modification was not clear although both of them were known to be involved in establishing patterns of gene expression during development. Only recently it has become apparent that DNA methylation and histone modification pathways can function in an interdependent manner (Cedar and Bergman 2009). This 'communication' can occur in both directions: histone methylation can influence DNA methylation patterns and DNA methylation may provide template for certain histone modifications after DNA replication (figure 3). At molecular level, such crosstalk has recently been shown to be manifested via direct interactions between histone and DNA methyltransferases (reviewed in Cedar and Bergman 2009). Failures in these 'epigenetic' features and crosstalks are thought to be linked to cancer and other abnormalities, and some may be heritable across generations. In this review, we have highlighted the epigenetic mechanisms that govern formation of chromatin domains, with specific emphasis on histone modifications.

6. Gene environment interaction and epigenetics

All organisms have to interact with their environment. The response of the genome depends on two important factors: **(1)** it cannot be programmed like embryonic development, as change in the environment is an external factor not in control of the organism and **(2)** it has to be rapid. These features make gene and environment interaction very special and much of this response has to be mediated by epigenetic mechanisms. This constitutes an exciting area of research.

Apart from their role in gene expression, chromatin structure, and epigenetic modifications, epigenetic mechanisms have also been implicated in distinct aspects of DNA metabolism, such as DNA replication, repair and recombination. Chadwick and Wade (2007), use the example of Rett syndrome, a neurological disease caused by loss of function of a methyl-DNA-binding protein, to discuss the concept of 'chromatin diseases'. The role of chromatin condensation, histone variants and modifications in the repair of DNA breaks has been appreciated recently (Bao and Shen 2007). A particularly exciting role for chromatin structure and modification has emerged from the study of mechanisms involved in the establishment and maintenance of pluripotency in embryonic stem cells and lineage specification. Embryonic stem cells appear to have a characteristic set of chromatin properties, including opposing activating and repressive histone modifications on lineage-specific genes, that seem to be crucial for their timed expression (Giadrossi *et al* 2007).

7. Chromatin elements that execute epigenetic regulation

Early during development, a number of mechanisms converge to set up the expression pattern of genes in cell-type specific manner initially by establishing the expression profile of key regulatory genes, like homeotic genes, are *hox* genes. Once the pattern is initiated, the factors that actually set up this pattern disappear, but the pattern set by them is maintained. This maintenance is carried out by cellular memory mechanisms, that included *cis* elements like cellular memory elements or Polycomb response elements, and the *trans* acting factors such as the Polycomb group (PcG) of genes that maintains the repressive state and the Trithorax (Trx) group of genes that maintains the active state of chromatin. This complex interplay of variety of *cis* elements and proteins machinery that enforces regulatory chromatin features operates in defined regions of the genome — the functional domains. It is well established that eukaryotic regulatory elements, including enhancers and silencers can function over a long distance. In their genomic context, however, they are restricted to their functional domains. The chromatin elements that define the limits of these domains are called 'chromatin domains boundaries'. The concept of 'chromatin domains' – the topologically independent structural unit – is the hallmark of higher-order chromatin organization. The concept that this structural unit also coincides with the functional unit of genomes offers a useful framework in dissecting the structure-function relationship. The spatial organization of eukaryotic genomes is tightly linked with their transcriptional regulation (Spector 2003; Goetze *et al* 2007). In higher eukaryotes such as mammals, transcription-related chromatin positioning is regulated largely at the level of chromatin sub-domains and also at the level of individual genes. Compartmentalization of multiple neighboring genes into a shared chromatin environment facilitates coordinated regulation of their expression (Spector 2003; Goetze *et al* 2007). Recent years have witnessed an explosive growth in the development of genome-scale tools for monitoring protein occupancy and histone modifications, enabling identification of a variety of chromatin domains based on the genomic binding patterns of various regulatory proteins and histone modifications. Distinct functional states of chromatin ranging from 'highly active' to 'completely silenced' can be associated with specific nucleosome rearrangements, histone variants, histone post-translational modifications, and interactions of non-histone regulators (reviewed by Bernstein *et al* 2007).

The expression of genes stably integrated into cultured cells or whole organisms is subject to chromosomal position effects. At some sites of insertion, only negligible levels of transgene expression are observed, while at other sites high levels of expression are obtained. Several factors probably contribute to these chromosomal position effects. Firstly, nearby regulatory elements (enhancers and silencers) may act on the transgene promoter, either potentiating or interfering with promoter activity. Secondly, these nearby regulatory elements may inappropriately interact with the regulatory elements that control expression of the transgene. Thirdly, if the transgene is inserted at a region in the chromosome, that is in an inactive chromatin configuration, this chromatin configuration may invade the transgene, eventually shutting off its promoter. This spreading of an inactive chromatin configuration may be analogous to that observed, when a chromosomal inversion places euchromatic genes near heterochromatin (position effect variegation) (Henikoff 1992; Reuter and Spierer 1992). Generally, the formation of large inactive chromatin domains may be a normal mechanism of gene regulation in higher eukaryotes. For example, the segment specific regulation of developmental loci, such as the bithorax complex (BX-C) in *Drosophila* or the *Hox* gene complexes in mammals appear to involve establishment and maintenance of inactive chromatin domains (Mihaly *et al* 1998).

What defines the boundaries of the chromatin domains? Two special chromatin structure elements, SCS and SCS' flanking the 87A7 heat shock locus of *Drosophila* and characterized by a set

of DNaseI hypersensitive sites, were proposed to function as boundaries of this chromatin domain (Udvardy *et al* 1985). Subsequently, SCS and SCS' have been demonstrated to provide position independent expression of a gene that is flanked by such elements (Kellum and Schedl 1991) and also block the action of an enhancer from acting across on a promoter (Kellum and Schedl 1992). These observations suggest that SCS and SCS' function as boundaries of the chromatin domains *in vivo* even at ectopic locations. Subsequently several boundary elements have been identified from various organisms.

While no consensus sequence has been found among different boundaries, two general conclusions can be made from the number of studies: **(1)** boundaries work across the species and **(2)** small sequence motifs cluster together to create a boundary. It is likely that while having some characteristics in common, individual chromatin domain boundaries may be associated with other regulatory elements. Many of the BX-C boundaries, for example, have associated *Polycomb*-Response Elements (PREs). Similarly, boundaries near the mammalian Locus Control Regions (LCRs) are associated with various regulatory elements (Iqbal and Mishra 2007). Different kinds of boundaries with subtle variations in their regulatory properties thus appear to be a common feature of eukaryotic chromatin organization.

8. How do chromatin domain boundaries function?

We can make speculative models, but in order to really answer this question we must know the components of the boundaries. Only a few boundary elements have been studied in some detail and very few interacting proteins have been identified so far. Among the boundaries identified till now, there is no significant sequence similarity although several of them have at least one common functional characteristic—the enhancer blocking ability in transgene assays. Clearly, we need to know more about the organization of these elements and interacting proteins in order to understand the mechanism of boundary function. Genetic and molecular information currently available, does not allow us to discount various speculations about the organization and function of chromatin domain boundaries (Gerasimova and Corces 2001; Iqbal and Mishra 2007; Mishra and Karch 1999).

Recent studies indicate that boundaries may function by arranging chromosomal domains in nuclear compartments in a way that regions requiring similar regulatory environment are together. This may involve clustering of boundary elements. This compartmentalization or clustering can take place in association with nuclear matrix through SAR/MAR like elements. Indeed, human MARs have been shown to insulate transgene expression from chromosomal position effects in *Drosophila* (Namciu *et al* 1998).

Since boundaries appear to be landmark regulatory and structural feature in the cell-type specific epigenetic state, several other regulatory elements can be analyzed in the context of boundaries. One such element is PRE. PREs play an important role in development as they function to maintain a predetermined state of activity at the level of chromatin structure by a mechanism that is proposed to involve epigenetic cellular memory. In transgenic assays, PREs show position-dependent and pairing-dependent properties, suggesting that these elements have very high tendency to interact with similar genomic elements. It is also speculated that such elements may bring together chromatin regions and play an important role in nuclear compartmentalization. The molecular basis of these interactions remains to be understood. Association of BX-C boundaries with PRE and ability of the PREs to cluster together may suggest that clustering/pairing of chromatin structure may initiate from one boundary and end at another one in *cis*. Recently, a Pc-G mutation, pleiohomeotic (pho) has been shown to be a sequence specific DNA binding protein, whose binding site is conserved among PREs (Brown *et al* 1998; Mihaly *et al* 1998b). This protein is homologous to a multifunctional mammalian protein YY-1. As YY-1 is known to be a matrix associated protein, it raises the possibility that PREs and the associated boundaries may function by sequestering DNA along nuclear matrix.

9. Proof of principle for epigenetic regulation of gene expression: Position effect variegation

Classical example of epigenetic effect is the position effect variegation of *white* gene in *Drosophila*. In a chromosomal inversion, the *white* gene was brought in the proximity of centromeric heterochromatin on the X-chromosome (figure 4). This created new context for the *white* gene, which normally is located far away from the centromere. In this new context, during development heterochromatin spreads beyond the *white* gene and inactivates it, while in some cells it does not reach *white* gene and maintains its active status. This functional state is clonally inherited later during the development and rest of the life in the fly. In the adult eye where *white* gene is responsible

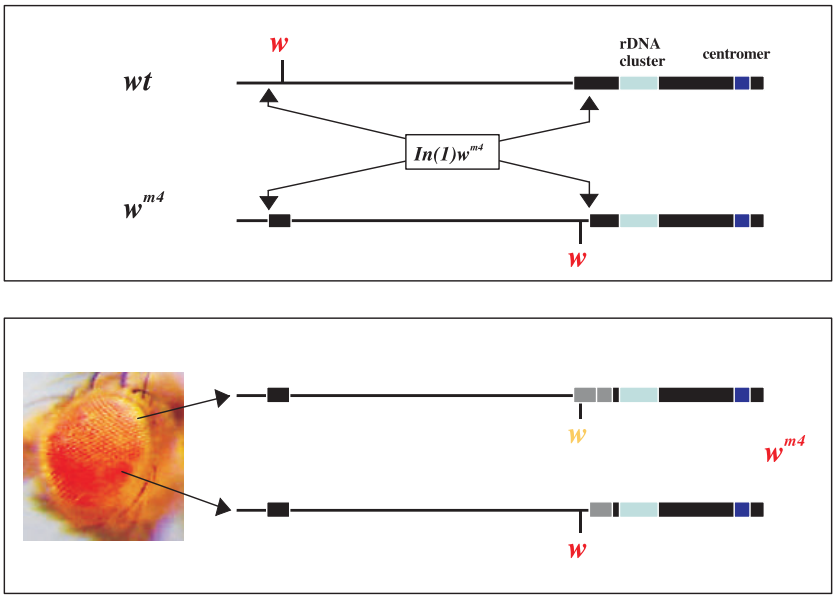

Figure 4. Position effect variegation (PEV). Example of PEV as seen in $In(1)w^{m4}$ chromosomal inversion on the X-chromosome of *Drosophila melanogaster*. Upper box shows the map of wt and $In(1)w^{m4}$ inversion breakpoints. The *white* gene, w, in the wild type chromosome is located near the telomere. Due to the inversion in the $In(1)w^{m4}$ chromosome, w is brought near the centromeric heterochromatin (black box). In the mutant flies part of the eye w gene is repressed as the heterochromatin spreads beyond it (shown as gray box) leading to lighter pigmentation while in other regions w gene is fully active (lower box), leading to the variegated phenotype. While all the cells have identical genotype, early during development the expression state of w gene is set differentially and is maintained epigenetically during later stages.

for red pigmentation, this early epigenetic event and its clonal inheritance is reflected as variegated eye color. While all the cells in the eye carry identical genome, they express different phenotype in terms of pigmentation. This epigenetic phenomenon has been studied extensively and has resulted in remarkable advancement in our understanding of epigenetic processes, genome organization, heterochromatin-euchromatin interaction, etc. Genetic dissection of PEV has opened ways to study chromatin structure and function (Henikoff 1992; Reuter and Spierer 1992).

10. Principles of epigenetic regulation

10.1 *Developmental aspects of epigenetic regulation*

Studies over the past few years have provided important insights into the epigenetic regulation of development. Studies using embryonic stem (ES) cells have revealed how the dynamic interplay of histone modifications orchestrates their maintenance and differentiation. A bivalent epigenetic mark of transcriptional activity in form of two contrasting modifications namely H3K4me3 and H3K27me3 was found to co-exist on multiple essential developmental genes expressed at low levels in mouse ES cells (Bernstein *et al* 2006) and on certain development-associated genes in human induced pluripotent ES cells (Takahashi *et al* 2007). Such epigenetic marks are important components of the molecular circuitry of pluripotency and are implicated in regulation of ES cell differentiation by maintaining a 'poised' low level expression of key genes (Jaenisch and Young 2008). Similarly, two histone modifications H3K9/14ac and H3K4me3 that are associated with constitutively expressed or inducible genes are implicated during T-cell development and function (Roh *et al* 2005). Conversely, H3K27me3 is enriched in permanently silenced genes that are involved in development of other cell-types. Interestingly, in differentiated T-cells the bivalent mark occurred simultaneously on many promoters like in ES-cells, and the genes regulated by such promoters were expressed at much lower levels (Roh *et al* 2005) suggesting that the bivalent mark of histone modifications regulates genes in undifferentiated ES cells as well as differentiated cell-types. These genes presumably can be switched on or off as required by the cell and may therefore carry these contrasting epigenetic marks.

11. Regulation of gene clusters

The interactive and interdependent nature of epigenetic modification, chromatin structure, and

genome organization has emerged from several studies. Most useful insights have been obtained from studies on regulation of gene clusters. *Hox* genes are among the best studied in this context, and have lead to the significant understanding of the epigenetic regulation during development. *Hox* genes exist as clusters wherein 5′ end of the complex has the gene that determines the posterior of most part of the body axis, while the 3′ end gene specifies most part of the anterior. This colinearity of organization, expression and function of *hox* genes was initially discovered in *Drosophila*, and has later been found to be conserved in all bilaterians. Early during development in flies gap genes, pair rule genes and the segments polarity genes (together referred to as segmentation genes) set the expression pattern of *hox* genes. Once this pattern is initiated, the segmentation gene products disappear and the maintenance of the *hox* genes is carried out by the two groups of genes—the Polycomb group (PcG) and trithorax group (trxG). PcG genes maintain the repressed state while trxG genes maintain the active state of genes by maintaining the appropriate epigenetic state of the chromatin (Brock and Fisher 2005). PcG group genes interpret the repressed state and put the H3K27me3 mark that is recognized by the chromodomain of PcG, which binds to such regions of genome to maintain the repressed state. The active state is interpreted by the trxG proteins that add the H3K9Ac mark which is recognized by the bromodomains of activator protein such as Brm. Interestingly, this epigenetic mechanism of maintaining the expression 'state of cellular memory' is conserved during evolution (Vasanthi and Mishra 2008). For example, each PcG and trxG member of *Drosophila* is present as one or more homologs in humans (Gould 1997; Mishra *et al* 2007; Ringrose and Paro 2004). Since PcG or trxG genes are direct regulators of homeotic genes, mutations in PcG/trxG genes cause misexpression of *hox* genes that leads to homeotic phenotype in flies as well as in mice. Differentially expressed state of closely spaced *hox* genes is maintained in the unit of chromatin domains that are separated by domain boundaries (Iqbal and Mishra 2007; Mihaly *et al* 1998). Deletion mutations in such boundaries lead to mix-up of the adjacent regulatory domains causing misexpression of homeotic genes that leads to homeotic phenotypes.

Gene activation is often accompanied by large-scale changes in chromatin conformation (Spector 2003; Goetze *et al* 2007). Moreover, large domains (50 to 200 kb) of generalized DNaseI hypersensitivity have been detected around developmentally regulated gene clusters (Dillon 2006), suggesting that the genome is organized into 'open' and 'closed' chromatin territories that represent higher-order functional domains. Technological advances in recent years have provided unprecedented insights into the role of chromatin organization and interactions of various structural-functional components towards gene regulation. Accumulating evidence in the past few years has revealed that activation or repression of genes involves change in formation of chromatin loops through long-range interactions between regulatory elements (Carter *et al* 2002; Tolhuis *et al* 2002; Spilianakis *et al* 2004; Cai *et al* 2006; Kumar *et al* 2007). The MARs are one of such elements that act at the bases of loops, where specific MAR binding proteins bind and tether these elements to the nuclear matrix, providing a higher-order compaction of chromatin. These loop configurations are altered during some cellular processes in a similar fashion as observed in the case of enhancer looping (Tolhuis *et al* 2002). The dynamic reorganization of higher-order chromatin has been documented using several gene clusters as model genomic *loci*. For example, during T_H2 cell activation, several DNA segments in the T_H2 cytokine cluster get freshly associated with the nuclear matrix, which is required for the coordinated expression of T_H2 locus genes (Cai *et al* 2006). Recent studies have unraveled the role of SATB1 in organization of chromatin 'loopscape' and its dynamic nature in response to physiological stimuli. The mammalian major histocompatibility complex (MHC) locus is a supercluster comprising several clusters of structurally unrelated genes, and has been the preferred model for investigating chromatin-based mechanisms that regulate gene clusters and expression of genes within them. SATB1, the global chromatin organizer and transcription factor, organizes the MHC class-I locus into distinct chromatin loops by tethering MARs to nuclear matrix at fixed distances. The MHC class I locus is reorganized during IFNγ treatment or RNAi mediated knockdown of SATB1 or PML isoforms. Silencing of SATB1 mimics the effects of IFNγ treatment on chromatin loop architecture of the MHC class-I locus and altered expression of genes within the locus (Kumar *et al* 2007 and Galande *et al* 2007). Thus, SATB1 has emerged as a key factor integrating higher-order chromatin architecture with gene regulation. At genome-wide level, SATB1 seems to play a role in organization of the transcriptionally 'poised' chromatin in a cell-type specific manner (Galande *et al* 2007). Thus, SATB1 and other multifaceted chromatin organizers may play a critical role in linking higher-order organization of genome with its function.

12. Evolutionary conservation of epigenetic mechanisms

The key players of epigenetic regulatory mechanisms are highly conserved across the species. Boundary elements from one species function equally well in another suggesting that key components are present in both. Similarly, the PcG and trxG members are well conserved from plants to flies and mammals (Mishra *et al* 2007; Ringrose and Paro 2004). Among best studied cases in context conservation of epigenetic mechanisms are the *Hox* complexes. The regulatory elements like PREs and boundaries not only function between fly and mouse, the proteins involved in their regulation are also conserved and have similar role. These observations lead to the idea that many of the epigenetic regulatory mechanisms and machineries have been well conserved during evolution as it plays a key role in development (Dean *et al* 2001; Levenson and Sweatt 2006; Vasanthi and Mishra 2008). Integration of the ENCODE data, in particular with respect to mammalian evolution based on inter- and intra-species sequence comparisons, has yielded new mechanistic and evolutionary insights concerning the functional landscape of the human genome, corroborating the above notion. However, it should be noted that the biggest surprise of the pilot phase of the ENCODE project was that many functional elements are seemingly unconstrained across mammalian evolution (The ENCODE Project Consortium 2007). In evolutionary terms, this pool of genomic elements may serve as a 'warehouse' for natural selection, potentially acting as the source of lineage-specific elements and functionally conserved but nonorthologous elements between species.

13. Epigenetics and disease

13.1 *Aging*

Developmental program leads to differentiation of totipotent cells to differentiated and functionally specialized cells. In the heart of this differentiation is the cell-type specific genomic packaging and the epigenome that expresses the selected set of genes and keeps the rest repressed. After the completion of the development that establishes the cell-type specific expression state, maintenance of this state is also critical for the appropriate functioning of the cell, during the life of an organism. Weakening of this epigenetic memory may lead to improper functioning of a large number of genes—a situation similar to aging, where a large degree of misregulation of genes takes place. Since PcG and trxG genes maintain the expression state, it is reasonable to ask if this class of proteins contribute to aging. During the lifetime of a cell or organism, damage by several processes keeps accumulating, that at some point starts to weaken the repair or maintenance system. This can happen, for example, when the efficiency of PcG/trxG system is overpowered by the degree of damage due to extrinsic agents or intrinsic processes. Changes in the epigenetic state of genome packaging can also lead to a similar situation altering the activity/expression state of PcG/trxG proteins and influence the process of aging (Mishra and Mishra 2009).

14. Chromatin and cancer

Journey of a genome from gamete to zygote, and then to germ line or terminal differentiation is accompanied by alterations in a parallel fashion in the organization of chromatin in these cells. It is becoming clear that some key steps of developmental program are implemented at the level of chromatin structure, which is then transmitted to daughter cells through variety of epigenetic 'memory' mechanisms acting at the level of chromatin organization as discussed above. During transformation of cells this implementation somehow fails, and the control of cell division is lost, leading to transformation. Several lines of recent observations suggest that different kinds of malignancies are associated with alteration in chromatin organization, few of which are discussed here.

The retinoblastoma (Rb) is a tumor suppressor protein that represses gene expression by modulating the architecture of chromatin. Rb recruits E2F protein to the histone deacetylase HDAC1 to form a complex, that prevents expression from E2F bound promoters (Brehm *et al* 1998; Magnaghi-Jaulin *et al* 1998). This repression is released when Rb is phosphorylated by CDKs. Viral oncogenes have been shown to bind to Rb and hence release it from the HDAC1-Rb-E2F complex, which allows transcription to occur from the E2F bound promoters, thus explaining why almost all cancer cells have to work their way past Rb protein.

At DNA level also, various kinds of alterations are observed in tumors. Parent-of-origin specific cytosine methylation patterns (imprinting) of DNA are relaxed in some tumors. Certain cancers are associated with the loss of imprinting of the parentally imprinted genes (Ogawa *et al* 1993; Rainier *et al* 1993). Repetitive or satellite DNA forms a large portion of the chromatin in humans. Loss of satellite DNA sequences have been reported in colon cancer (Thibodeau *et al* 1993). Rearrangements or deletions involving a break point at the chromosomal band 11q23 are

associated with a variety of hematopoietic malignancies. Cloning of the gene, involved in 11q23 chromosomal translocations in acute leukemias, shows that the gene in question is a homolog of *trithorax (trx)* gene of *Drosophila* (Djabali *et al* 1992). The *trx* – a component involved in the maintenance of active chromatin – is an important gene involved in the process of development. Topoisomerase II is an important component of chromatin scaffold and has been suggested to bind near the chromatin domain boundaries. Interestingly, in patients treated with topoisomerase II inhibitors, frequent occurrence of 11q23 aberrations in secondary leukemias, has been reported (Pui *et al* 1991).

Number of epigenetic modifications have been reported to be altered during tumorigenesis, however, further studies will be required to establish their cause and effect relationship. Hypermethylation of CpG islands and genome-wide hypomethylation are common epigenetic features of cancer cells. Loss of acetylation at lysine 16 (H4K16ac) and trimethylation at lysine 20 (H4K20me3) was identified as a hallmark of human cancer cells and was associated with DNA hypomethylation at repetitive sequences (Fraga *et al* 2005). Epigenetic silencing in cancer cells is mediated by at least two distinct histone modifications, polycomb-based histone H3 lysine 27 trimethylation (H3K27me3) and lysine 9 dimethylation (H3K9me2). Using chromatin immunoprecipitation microarrays (ChIP-on-chip) in prostate cancer cells compared to normal prostate (Kondo *et al* 2008), found that up to 5% of promoters were enriched with H3K27me3. Among these promoters, 16% possessed CpG islands and 84% were non-CpG islands, and these genes were silenced specifically in prostate cancer. These studies implicated polycomb (EZH2)-mediated H3K27me3 as a mechanism of tissue and/or cell-type specific tumor-suppressor gene silencing in cancer that is potentially independent of DNA methylation status of the promoter (Kondo *et al* 2008). Several members of the Polycomb group (PcG) of genes have been implicated in cancer (Valk-Lingbeek *et al* 2004). Over-expression of BMI-1 and EZH2 has been linked to breast and prostate cancers (Glinsky *et al* 2003). BMI-1 is also reported to be over-expressed in several other cancers such as non-small-cell lung cancer, colorectal cancer, nasopharyngeal carcinoma, and oral cancer (Kang *et al* 2007; Song *et al* 2006). Large body of evidence has indicated that proteins from the PcG are epigenetic chromatin modifiers involved in cancer development and also in the maintenance of embryonic and adult stem cells underlining common regulatory mechanisms in cancer cells and stem cells (reviewed by Valk-Lingbeek *et al* 2004).

Thus, the combinatorial and interdependent nature of manifestation of the epigenetic code provides vast number of 'readouts' to translate the information encoded in the genome. Further studies and alliance projects such as ENCODE (The ENCODE Project Consortium 2007) and Alliance for the Human Epigenome and Disease (AHEAD) (The AACR Human Epigenome Task Force and EUNE 2008) of truly global nature will unravel the cause and effect relationship between the epigenetic modifications and various developmental disorders and diseases. Certain epigenetic marks can also be used as therapeutic targets (Mulero-Navarro and Esteller 2008). Technological breakthroughs during the past few years have enabled considerable advances in identification of global patterns of epigenetic patterns. The major challenge that lies ahead is to decode the epigenome and to understand its regulatory principles, which await development of new models, analysis tools, and systems approach.

References

Bao Y and Shen X 2007 Chromatin remodeling in DNA double-strand break repair; *Curr. Opin. Genet. Dev.* **17** 126–131.

Bernstein B E, Meissner A and Lander E S 2007 The mammalian epigenome; *Cell* **128** 669–681.

Bernstein B E, Mikkelsen T S, Xie X, Kamal M, Huebert D J, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber S L and Lander E S 2006 A bivalent chromatin structure marks key developmental genes in embryonic stem cells; *Cell* **125** 315–326.

Bird A 2007 Perceptions of epigenetics; *Nature* **447** 396–398.

Brehm A, Miska E A, McCance D J, Reid J L, Bannister A J and Kouzarides T 1998 Retinoblastoma protein recruits histone deacetylase to repress transcription; *Nature* **391** 597–601.

Brock H W and Fisher C L 2005 Maintenance of gene expression patterns; *Dev. Dyn.* **232** 633–655.

Brown J L, Mucci D, Whiteley M, Dirksen M L and Kassis J A 1998 The *Drosophila* polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1; *Mol. Cell* **1** 1057–1064.

Cai S, Lee C C and Kohwi-Shigematsu T 2006 SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes; *Nat. Genet.* **38** 1278–1288.

Carter D, Chakalova L, Osborne C S, Dai Y F and Fraser P 2002 Long-range chromatin regulatory interactions *in vivo*; *Nat. Genet.* **32** 623–626.

Cedar H and Bergman Y 2009 Linking DNA methylation and histone modification: Patterns and paradigms; *Nature Rev. Genet.* **10** 295–304.

Chadwick L H and Wade P A 2007 MeCP2 in Rett syndrome: Transcriptional repressor or chromatin architectural protein? *Curr. Opin. Genet. Dev.* **17** 121–125.

Dean W, Santos F, Stojkovic M, Zakhartchenko V, Walter J, Wolf E and Reik W 2001 Conservation of methylation reprogramming in mammalian development:

Aberrant reprogramming in cloned embryos; *Proc. Natl. Acad. Sci. USA* **98** 13734–13738.

Dillon N 2006 Gene regulation and large-scale chromatin organization in the nucleus; *Chromosome Res.* **14** 117–126.

Djabali M, Selleri L, Parry P, Bower M, Young B and Evans G A 1992 A trithorax-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias; *Nat. Genet.* **2** 113–118.

Feinberg A P and Tycko B 2004 The history of cancer epigenetics; *Nat. Rev. Cancer* **4** 143–153.

Fraga M F, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K, Iyer N G, Pérez-Rosado A, Calvo E, Lopez J A, Cano A, Calasanz M J, Colomer D, Piris M A, Ahn N, Imhof A, Caldas C, Jenuwein T and Esteller M 2005 Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer; *Nat. Genet.* **37** 391–400.

Galande S, Purbey P K, Notani D and Kumar P P 2007 The third dimension of gene regulation: Organization of dynamic chromatin loopscape by SATB1; *Curr. Opin. Genet. Dev.* **17** 408–417.

Gerasimova T I and Corces V G 2001 Chromatin insulators and boundaries: Effects on transcription and nuclear organization; *Ann. Rev. Genet.* **35** 193–208.

Giadrossi S, Dvorkina M and Fisher A G 2007 Chromatin organization and differentiation in embryonic stem cell models; *Curr. Opin. Genet. Dev.* **17** 132–138.

Glinsky G V, Ivanova Y A and Glinskii A B 2003 Common malignancy-associated regions of transcriptional activation (MARTA) in human prostate, breast, ovarian, and colon cancers are targets for DNA amplification; *Cancer Lett.* **201** 67–77.

Goetze S, Mateos-Langerak J and van Driel R 2007 Three-dimensional genome organization in interphase and its relation to genome function; *Semin. Cell Dev. Biol.* **18** 707–714.

Goll M G and Bestor T H 2005 Eukaryotic cytosine methyltransferases; *Ann. Rev. Biochem.* **74** 481–514.

Gould A 1997 Functions of mammalian polycomb group and trithorax group related genes; *Curr. Opin. Genet. Dev.* **7** 488–494.

Greally J M 2007 Encyclopedia of humble DNA; *Nature* **447** 782–783.

Henikoff S 1992 Position effect and related phenomena; *Curr. Opin. Genet. Dev.* **2** 907–912.

Huang J, Fan T, Yan Q, Zhu H, Fox S, Issaq H J, Best L, Gangi L, Munroe D and Muegge K 2004 Lsh, an epigenetic guardian of repetitive elements; *Nucleic Acids Res.* **32** 5019–5028.

International Human Genome Sequencing Consortium 2001 Initial sequencing and analysis of the human genome; *Nature* **409** 860–921.

International Human Genome Sequencing Consortium 2004 Finishing the euchromatic sequence of the human genome; *Nature* **431** 931–945.

Iqbal H and Mishra R K 2007 Chromatin domain boundaries: Defining the functional domains in genome; *Proc. Indian Natl. Sci. Acad.* **73** 239–253.

Jaenisch R and Young R 2008 Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming; *Cell* **132** 567–582.

Jenuwein T and Allis C D 2001 Translating the histone code; *Science* **293** 1074–1080.

Kang M K, Kim R H, Kim S J, Yip F K, Shin K H, Dimri G P, Christensen R, Han T and Park N H 2007 Elevated Bmi-1 expression is associated with dysplastic cell transformation during oral carcinogenesis and is required for cancer cell replication and survival; *Br. J. Cancer* **96** 126–133.

Kellum R and Schedl P 1991 A position-effect assay for boundaries of higher order chromosomal domains; *Cell* **64** 941–950.

Kellum R and Schedl P 1992 A group of SCS elements function as domain boundaries in an enhancer-blocking assay; *Mol. Cell Biol.* **12** 2424–2431.

Kondo Y, Shen L, Cheng A S, Ahmed S, Boumber Y, Charo C, Yamochi T, Urano T, Furukawa K, Kwabi-Addo B, Gold D L, Sekido Y, Huang T H and Issa J P 2008 Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation; *Nat. Genet.* **40** 741–750.

Krivtsov A V, Feng Z, Lemieux M E, Faber J, Vempati S, Sinha A U, Xia X, Jesneck J, Bracken A P, Silverman L B, Kutok J L, Kung A L and Armstrong S A 2008 H3K79 methylation profiles define murine and human MLL-AF4 leukemias; *Cancer Cell* **14** 355–368.

Kumar P P, Bischof O, Purbey P K, Notani D, Urlaub H, Dejean A and Galande S 2007 Functional interaction between PML and SATB1 regulates chromatin-loop architecture and transcription of the MHC class I locus; *Nat. Cell Biol.* **9** 45–56.

Levenson J M and Sweatt J D 2006 Epigenetic mechanisms: A common theme in vertebrate and invertebrate memory formation; *Cell. Mol. Life Sci.* **63** 1009–1016.

Li B, Carey M and Workman J L 2007 The role of chromatin during transcription; *Cell* **128** 707–719.

Luger K, Mäder A W, Richmond R K, Sargent D F and Richmond T J 1997 Crystal structure of the nucleosome core particle at 2.8 Å resolution; *Nature* **389** 231–233.

Magnaghi-Jaulin L, Groisman R, Naguibneva I, Robin P, Lorain S, Le Villain J P, Troalen F, Trouche D and Harel-Bellan A 1998 Retinoblastoma protein represses transcription by recruiting a histone deacetylase; *Nature* **391** 601–605.

Margueron R, Trojer P and Reinberg D 2005 The key to development: Interpreting the histone code? *Curr. Opin. Genet. Dev.* **15** 163–176.

McPherson J D, Marra M, Hillier L, Waterston R H, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis E R and Wilson R K (International Human Genome Mapping Consortium) 2001 A physical map of the human genome; *Nature* **409** 934–941.

Mihaly J, Hogga I, Barges S, Galloni M, Mishra R K, Hagstrom K, Muller M, Schedl P, Sipos L and Gausz J 1998 Chromatin domain boundaries in the Bithorax complex; *Cell Mol. Life Sci.* **54** 60–70.

Mihaly J, Mishra R K and Karch F 1998b A conserved sequence motif in polycomb-response elements; *Mol. Cell.* **1** 1065–1066.

Mishra K and Mishra R K 2009 In: (ed.) Tollefsbol T O *Epigenetics of Aging*. Springer, pp. 135–150.

Mishra R K and Karch F 1999 Boundaries that demarcate structural and functional domains of chromatin; *J. Biosci.* **24** 377–399.

Mishra R K, Yamagishi T, Vasanthi D, Ohtsuka C and Kondo T 2007 Involvement of polycomb-group genes in establishing HoxD temporal colinearity; *Genesis* **45** 570–576.

Misteli T 2007 Beyond the sequence: Cellular organization of genome function; *Cell* **128** 787–800.

Mulero-Navarro S and Esteller M 2008 Epigenetic biomarkers for human cancer: The time is now; *Crit. Rev. Oncol. Hematol.* **68** 1–11.

Namciu S J, Blochlinger K B and Fournier R E 1998 Human matrix attachment regions insulate transgene expression from chromosomal position effects in *Drosophila melanogaster; Mol. Cell. Biol.* **18** 2382–2391.

Ogawa O, Eccles M R, Szeto J, McNoe L A, Yun K, Maw M A, Smith P J and Reeve A E 1993 Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour; *Nature* **362** 749–751.

Ohno S 1972 So much 'junk' DNA in our genome; In *Evolution of genetic systems* (ed.) Smith H H, pp. 366–370.

Pui C H, Ribeiro R C, Hancock M L, Rivera G K, Evans W E, Raimondi S C, Head D R, Behm F G, Mahmoud M H and Sandlund J T 1991 Acute myeloid leukemia in children treated with epipodophyllotoxins for acute lymphoblastic leukemia; *N. Engl. J. Med.* **325** 1682–1687.

Rainier S, Johnson L A, Dobry C J, Ping A J, Grundy P E and Feinberg A P 1993 Relaxation of imprinted genes in human cancer; *Nature* **362** 747–749.

Reik W 2007 Stability and flexibility of epigenetic gene regulation in mammalian development; *Nature* **447** 425–432.

Reuter G and Spierer P 1992 Position effect variegation and chromatin proteins; *Bioessays* **14** 605–612.

Ringrose L and Paro R 2004 Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins; *Ann. Rev. Genet.* **38** 413–443.

Roh T Y, Cuddapah S and Zhao K 2005 Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping; *Genes Dev.* **19** 542–552.

Shilatifard A 2006 Chromatin modifications by methylation and ubiquitination: Implications in the regulation of gene expression; *Ann. Rev. Biochem.* **75** 243–269.

Song L B, Zeng M S, Liao W T, Zhang L, Mo H Y, Liu W L, Shao J Y, Wu Q L, Li M Z, Xia Y F, Fu L W, Huang W L, Dimri G P, Band V and Zeng Y X 2006 Bmi-1 is a novel molecular marker of nasopharyngeal carcinoma progression and immortalizes primary human nasopharyngeal epithelial cells; *Cancer Res.* **66** 6225–6232.

Spector D L 2003 The dynamics of chromosome organization and gene regulation; *Ann. Rev. Biochem.* **72** 573–608.

Spilianakis C G and Flavell R A 2004 Long-range intrachromosomal interactions in the T helper type 2 cytokine locus; *Nat. Immunol.* **5** 1017–1027.

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K and Yamanaka S 2007 Induction of pluripotent stem cells from adult human fibroblasts by defined factors; *Cell* **131** 861–872.

Talbert P B and Henikoff S 2006 Spreading of silent chromatin: Inaction at a distance; *Nat. Rev. Genet.* **7** 793–803.

The American Association for Cancer Research Human Epigenome Task Force & European Union, Network of Excellence, Scientific Advisory Board 2008 Moving AHEAD with an international human epigenome project; *Nature* **454** 711–715.

The ENCODE Project Consortium 2007 Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project; *Nature* **447** 799–816.

Thibodeau S N, Bren G and Schaid D 1993 Microsatellite instability in cancer of the proximal colon; *Science* **260** 816–819.

Tolhuis B, Palstra R J, Splinter E, Grosveld F and de Laat W 2002 Looping and interaction between hypersensitive sites in the active beta-globin locus; *Mol. Cell* **10** 1453–1465.

Udvardy A, Maine E and Schedl P 1985 The 87A7 chromomere. Identification of novel chromatin structures flanking the heat shock locus that may define the boundaries of higher order domains; *J. Mol. Biol.* **185** 341–358.

Valk-Lingbeek M E, Bruggeman S W and van Lohuizen M 2004 Stem cells and cancer; the polycomb connection; *Cell* **118** 409–418.

Vasanthi D and Mishra R K 2008 Epigenetic regulation of genes during development: A conserved theme from flies to mammals; *J. Genet. Genomics* **35** 413–429.

Venter J C, Adams M D and Myers E W 2001 The sequence of the human genome; *Science* **291** 1304–1351.

Wang Z, Zang C, Rosenfeld J A, Schones D E, Barski A, Cuddapah S, Cui K, Roh T Y, Peng W, Zhang M Q and Zhao K 2008 Combinatorial patterns of histone acetylations and methylations in the human genome; *Nat. Genet.* **40** 897–903.

White E J, Emanuelsson O, Scalzo D, Royce T, Kosak S, Oakeley E J, Weissman S, Gerstein M, Groudine M, Snyder M and Schübeler D 2004 DNA replication-timing analysis of human chromosome 22 at high resolution and different developmental states; *Proc. Natl. Acad. Sci. USA* **101** 17771–17776.