Spatial epigenetics: linking nuclear structure and function in higher eukaryotes

Dean A. Jackson

University of Manchester, Faculty of Life Sciences, 131 Princess St, Manchester M1 7DN, U.K.

Abstract

Eukaryotic cells are defined by the genetic information that is stored in their DNA. To function, this genetic information must be decoded. In doing this, the information encoded in DNA is copied first into RNA, during RNA transcription. Primary RNA transcripts are generated within transcription factories, where they are also processed into mature mRNAs, which then pass to the cytoplasm. In the cytoplasm these mRNAs can finally be translated into protein in order to express the genetic information as a functional product. With only rare exceptions, the cells of an individual multicellular eukaryote contain identical genetic information. However, as different genes must be expressed in different cell types to define the structure and function of individual tissues, it is clear that mechanisms must have evolved to regulate gene expression. In higher eukaryotes, mechanisms that regulate the interaction of DNA with the sites where nuclear functions are performed provide one such layer of regulation. In this chapter, I evaluate how a detailed understanding of nuclear structure and chromatin dynamics are beginning to reveal how spatial mechanisms link chromatin structure and function. As these mechanisms operate to modulate the genetic information in DNA, the regulation of chromatin function by nuclear architecture defines the concept of ‘spatial epigenetics’.

1email dean.jackson@manchester.ac.uk
Introduction

All organisms are defined by unique information that is encoded in their DNA. However, in order to express individual features, the genetic information within DNA must be decoded. During this process, the information embedded in DNA is first transferred to RNA and then later to protein. The proteins that are generated from individual genetic units – called genes – then provide specific functions that in turn define the structure and function of cells and the physiological capabilities of the organism. The overall process by which information in DNA is transferred to functional proteins is called gene expression.

In simplistic terms, the genes of multicellular organisms – know technically as metazoa – can be divided into two broad classes: (i) genes that are expressed in all cells and provide fundamental function, such as enzymes of essential metabolic pathways, are referred to as housekeeping genes; and (ii) other genes are specialized to perform a specific function in particular cell types. The activity of these is defined by cell differentiation, and expression of these genes is defined by the lineage of the cells and their developmental history.

In humans, approximately 250 different cell types can be described based on their developmental lineage and the constellation of cell-type-specific genes that they express. The different cell types are defined by the functions that they perform, as a consequence of their expression profiles. Gene expression within the haematopoietic cell lineage provides an excellent example of this. As cells develop by specialization and eventual terminal differentiation of more primitive precursor cells – these are the haematopoietic stem cells – particular cell-type-specific genes that support discrete functions are expressed [1]. Well-known examples of these cell-type-specific genes in this developmental lineage are globin gene expression in erythroblasts, which eventually mature into fully differentiated red blood cells, and immunoglobulin gene expression in lymphocytes, which express antibodies.

With only rare exceptions, differentiated cells retain all of their genetic material during differentiation, so that profiles of tissue-specific gene expression are defined at the level of transcriptional control. Cells that do change their genetic content do so because of the function they perform. For example, mature red blood cells have to pass through very narrow capillaries to fulfil their roles in tissue oxygenation and this is only possible because the precursor cells are enucleated during differentiation. This, however, is a rare example of a cell type that alters its genetic content during differentiation. The vast majority of differentiated cells retain a full genetic complement. This fact can be demonstrated using somatic nuclear transfer techniques, where the differentiated nuclei are treated in a way that allows them to de-differentiate and undergo a full developmental programme after being introduced into a suitable enucleated egg. This approach of somatic cell nuclear transfer for animal cloning was made famous by the cloning of Dolly the sheep in 1997 [2]. Even though
we know the majority of the DNA sequence of many organisms (some large DNA repeat blocks are extremely difficult to sequence and are missing from the published genomes), we still have only limited understanding of how this information can be used to predict levels and cell-type-specific patterns of gene expression. Even so, key mechanisms that control gene expression are inherent to the genetic code itself [3–5]. Most notably, gene expression is activated when transcription factors interact with their cognate recognition motifs in gene promoters and other activating sequences – such as enhancers – to form a complex that recruits the transcription machinery to a gene. This process of sequence-dependent association of specific proteins with defined sequence elements within DNA can be studied in molecular detail, and is conceptually simple to understand. However, inside the cell, this feature of control is itself regulated by chromatin structure and folding inside the nucleus. This means that even when specific DNA motifs and appropriate binding factors are present within a cell, it is not inevitable that their interaction will be permitted and specific functions performed.

In fact, it is now well-documented that the mechanisms of DNA packaging in chromatin and factors that control how the chromatin fibre is then organized within nuclei can exert profound influences on patterns and levels of gene expression. These so called ‘epigenetic’ factors arise from a number of sources. The most obvious, the post-translational modification of histones, has been reviewed extensively over the years [6–8] and will be noted only in passing in the present chapter. But in addition to this, it has been known for many years that different chromosomal locations – and perhaps different nuclear positions – have quite different functional capabilities. Many observations imply that chromosome structure and nuclear architecture have co-evolved to ensure that genes are expressed from their natural loci to provide the requisite levels of expression in appropriate circumstances. This chapter presents a review of our present understanding of chromosome structure and nuclear architecture and attempts to set out some general principles that will allow us to explore how nuclear structure and function can be linked.

Basic features of nuclear organization

The nucleus

A view of the mammalian cell nucleus is shown in Figure 1. This image was generated using classical electron microscopy techniques, which have been developed over the past 50 years to visualize cellular compartments at high resolution; with the fixation and staining techniques used in Figure 1, the resolution is at the nanometer (nm) range. The section shown in Figure 1 gives a classical view of nuclear structure as it is seen in a mammalian cell that is capable of continuous proliferation; this example is from a culture of human HeLa cells, a cancer cell line that was established from a cervical carcinoma. In this section, the structure of the two major cellular compartments – the nucleus and cytoplasm – is very clear. The nucleus is located at the centre of the cell and
Figure 1. Electron micrograph showing a typical section through a human HeLa cell

Proliferating human HeLa cells were processed using routine electron microscopy techniques. Samples were stained using uranyl acetate and lead citrate and a thin section visualized using transmission electron microscopy. The central nucleus (Nu) is enclosed by a double lipid bilayer that forms the nuclear membrane (NM). The most prominent nuclear feature is a large central nucleolus (No). This highly structured compartment contains classical fibrillar centres with their surrounding dense fibrillar components and a more diffuse granular component. The structure of these compartments is defined by the functional organization of the nucleolus during rRNA synthesis and metabolism. Dense clumps of heterochromatin form along the inner face of the nuclear membrane, around nucleoli and at a small number of sites that are scattered throughout the nucleus. Active genes within more open euchromatin are dispersed throughout the nuclear interior and mixed with the interchromatin domain. These two compartments have an amorphous granular appearance with little obvious structure. The nucleus is surrounded by the cytoplasm (C), where many membrane-bound organelles are found. The nuclear diameter in this image is 10 µm.

is bounded by a double lipid membrane, making it the largest cellular organelle (by classical definition organelles have a surrounding membrane). This very special membrane is punctuated by a few thousand nuclear pores, which regulate access between the two compartments. The pores allow free transit of molecules with sizes of $\lesssim 50$ kDa, and nuclear import and export of larger molecules is regulated by highly sophisticated transport mechanisms. The cytoplasm, notably, contains numerous other organelles, such as mitochondria, Golgi apparatus and endosomes. The nucleus, in contrast, has no internal membrane-bound structures and the major nuclear structures are defined by the association properties of their component parts.

Within nuclei, the sites of rRNA (ribosomal RNA) synthesis and processing within nucleoli provide the most obvious compartment [9,10]. In this
image a single huge nucleolus is seen and this can also be seen to contain the three characteristic nucleolar compartments: fibrillar centres, dense fibrillar component and granular component. The structure of these sub-compartments within nucleoli is defined by the association of the active rRNA genes with the transcription machinery, in the fibrillar centres, and the dense structure of the associated nascent transcripts, which form the dense fibrillar component. The nascent transcripts are processed and packaged into mature sub-ribosomal particles (the 40S and 60S subunits in human cells) within the granular component. Nucleoli thus provide an excellent paradigm for understanding nuclear architecture, in that, as with other nuclear compartments, their structure is a product of the functions that nucleoli perform. Nucleoli represent approx. 5% of the nuclear volume in proliferating mammalian cells. The remainder of the volume is occupied by chromatin, proteins that drive chromatin function and the products of these functions, such as nascent and mature mRNA molecules. DNA is packaged as chromatin that is folded into highly condensed (heterochromatin) or more open and dynamic (euchromatin) forms. The dense chromatin contains very few active genes and is usually assumed to represent an inactive nuclear compartment; as seen here this is often clustered around the inner nuclear membrane and along the surface of nucleoli, which essentially represent the surfaces on the chromatin-containing compartment. The open chromatin occupies the nuclear interior. This chromatin has a dispersed punctate appearance and is interspersed with granular material of slightly different texture, which represents local sites where different nuclear functions are performed.

Revealing the active sites of nuclear function
Because of the complexity of components within nuclei, it is not possible to recognize individual active sites unless specific components of the nucleus are first labelled in some way. Specific nuclear compartments [11] can be labelled using antibodies. To do this, specific antibodies must be raised to unique regions of the protein under study, this prevents cross-reaction with other proteins in the cell. Cells are fixed on to some suitable surface, usually a glass slide, and then the antibody is incubated with the sample so that it is able to bind to the specific target proteins. After washing away any unbound antibody, the location of the bound antibody is revealed by incubating with a fluorescent secondary antibody (for example a fluorescein-conjugated donkey anti-mouse antibody) and visualized using a fluorescent microscope. An example of how this approach can be used to study nuclear compartments is shown in Figure 2. In this case, a major nuclear compartment, which is referred to as the nuclear speckles [12] because of its appearance in such images, is revealed by staining with an antibody which binds to the Sm (human autoimmune serum) epitope that is found in many RNA-binding proteins. The same samples can be co-stained with another antibody that reveals sites of transcription, which were labelled by treating the cell with
BrUTP before fixation. The Br-containing transcription sites [13] were also revealed by indirect labelling using a different fluorescent reporter. Note here that transcription does not occur within nuclear speckles, but the speckles provide protein machinery that is required for RNA processing and assembly of mature mRNA–protein complexes that migrate to the cytoplasm. Hence, the relative distribution of sites of RNA synthesis and sites that are rich in the machinery used during mRNA maturation contribute to the overall efficiency of mRNA synthesis.

**The sites of gene expression**

Classical electron microscopy techniques (such as in Figure 1 with immunostaining for electron microscopy) have shown that gene expression generally occurs at the borders of condensed chromatin in association with perichromatin fibrils. The ability to label the nascent RNA (Figure 3) using BrUTP *in vitro* or BrU *in vivo* has shown that a proliferative mammalian cell has approx. 1000–2000 active sites of RNA polymerase II activity for each haploid chromosome set [14–16]. As cells have at least five times this number of active RNA polymerase II holoenzyme complexes [14], each active centre must represent a nuclear compartment where the transcripts from groups of genes are generated and processed together. This spatial co-ordination of the different steps required to produce mature mRNAs at specific nuclear sites forms the basis of the concept of transcription factories [17,18]. Within these structures, transcripts are polymerized, processed and assembled into the required mRNA–protein complex before being released from the site to engage the downstream export pathway. For a typical transcript, events occurring at the transcription site take approx. 15 min to complete, although
for many of the longer genes the formation of mature mRNAs may take many hours.

In mammalian cells, three distinct classes of RNA transcript are generated by different polymerases. Interestingly, these three classes of polymerase are seen to occupy discrete transcription factories [15]. As we have seen, mRNA transcripts are generated throughout the nucleoplasm as discrete spots of activity within RNA polymerase II-containing transcription factories (Figure 3). RNA polymerase III factories, with similar structure, are engaged in the synthesis of many very small transcripts, such as tRNA (transfer RNA) molecules. Finally, RNA polymerase I factories within nucleoli are required to generate the rRNA molecules that form the basis of ribosome structure.

As noted above, the structure of nucleoli is defined by the synthesis of rRNA and processing of the primary transcripts to form assembled sub-ribosomal particles, which can then pass to the cytoplasm. Labelling studies using BrUTP incorporation and both light and electron microscopy for analysis demonstrate that the dense fibrillar components contain the nascent RNP (ribonucleoprotein) and active genes (visualized by DNA in situ hybridization) [19]. However, this apparently structured view of nucleolar organization belies the dynamic properties that are required during ribosome biosynthesis [20].

In attempting to understand the structure of nuclei, an evolving concept of nuclear compartments involves the principle of molecular self assembly. Nuclear compartments such as transcription centres are not membrane-bound. Instead, they appear to persist because of the ‘self-assembly’ characteristics of their components [21,22] and architectural constraints imposed by nuclear and chromosome structure (see below). Nuclear speckles provide an excellent

Figure 3. Transcription sites in human cells
HeLa cells were permeabilized and the nascent transcripts labelled as described in Figure 2. To provide optimal resolution, immunostaining was performed on cryosections of ~90 nm and Br-RNA immunolabelled with Cy3 (indocarbocyanine; shown green) and DNA counterstained with TOTO-3 (shown red). Fluorescence images were collected using confocal microscopy and two colour merges generated. The sections shown are from three samples that were treated independently with α-amanitin during transcription, to differentially inhibit RNA polymerases I, II and III. In the control sample, without inhibitor (0), staining from all three classes of RNA polymerase is seen. With inhibitor added at 2 µg/ml (2) synthesis by RNA polymerase II is inhibited completely and nucleoplasmic sites reflect synthesis by RNA polymerase III. Finally, with inhibitor at 250 µg/ml (250) both RNA polymerase II and III activities are eliminated and the only synthesis that remains is the resistant RNA polymerase I activity within nucleoli. Image kindly provided by Ana Pombo; see [15] for details.
example of this architectural theme. Proteins of this compartment occupy the interchromatin spaces where they may accumulate to form large, clustered aggregates or contribute to a more diffuse pool. Once formed, the aggregates are stable (i.e. they occupy a particular area of nuclear space over time), even though the components within them are extremely dynamic. However, the structures are also plastic and can change dramatically under conditions where RNA synthesis is inhibited and the need for splicing is lost [12]. The expression status of a gene can also be shown to correlate with its nuclear location. Studies on the lymphoid lineage of mammalian cells have demonstrated that the inactivation of gene expression correlates with relocation of a gene to heterochromatic nuclear sites; although the reverse process occurs during gene activation [23,24]. During this process, the sequence-specific transcription factor Ikaros becomes associated both with the silenced gene and local centromeric heterochromatin. Ikaros is able to bind both a target promoter and sites within the $\alpha$ satellite repeats, so providing a means of driving appropriate genes into inactive heterochromatic sites. During B lymphocyte development the IgH (Ig is immunoglobulin) and IgK loci are located at the nuclear periphery in haematopoietic progenitors and pro-T-cells, and in the nuclear interior in pro-B nuclei [25]. The inactive loci are associated with the nuclear lamina and must move to active sites within the nuclear interior before recombination and transcription of the IgH and IgK loci can occur.

**Chromatin folding and chromosome territories**

Gene expression demands the highly orchestrated interaction between the active centres that contain the synthetic machinery and the chromatin template. Hence, in order to understand how the interactions can be regulated, it is essential to understand how chromatin is folded and packaged within mammalian nuclei [24,26–28].

Mammalian genomes are so huge that they would extend for some 2 m if their DNA was stretched out. However, inside a cell, this length of DNA is folded to occupy a nucleus that typically will be only 10 μm (i.e. $10^{-6}$ m) in diameter. This huge condensation is achieved using a hierarchical series of DNA packaging features: (i) DNA first associates with nucleosomes (composed of structural histone octamers) that contain ~160 bp of DNA that wraps as approximately two DNA turns around the nucleosome core; (ii) the primary nucleosomal fibre folds into more condensed higher-order structures; (iii) the higher-order structures might be constrained locally as a series of chromatin loops [this occurs via multiple chromatin interactions and also interactions with structural components of the nucleoskeleton (see below)]; and (iv) chromatin fibres and loops are also constrained locally into higher-order structures called DNA foci that typically contain ~1 Mbp of DNA [26]. These higher-order structures appear to be functional units during DNA synthesis and are thought to be stabilized by chromatin-based interactions that rely on the post-translational modification of histones.
The structure of DNA foci within mammalian cells can be revealed by in situ labelling techniques that define the nuclear distribution of DNA from a specific chromosomal region or locus [26,27]. The data presented in [29] provide an excellent example of this level of DNA folding. This study was designed to explore the distribution of active euchromatin and more inactive heterochromatin, within the mammalian cell nucleus. The regions chosen for analysis contained adjacent gene-rich (called islands) and gene-poor (called deserts) domains that contained approx. 10000 kbp of DNA. To visualize the distribution of these regions within nuclei, the respective regions were labelled homogeneously using probes that were hybridized to the DNA within fixed nuclei. Before hybridization, the nuclei were treated to make their DNA single-stranded so that binding can occur, and after hybridization the samples were washed and further processed (similar to Figure 2) to reveal the sites of hybridization. Three key features are seen in such experiments: first, the different chromosomal domains are largely self-contained and interact with, but do not really mix with, adjacent regions that maintain different chromatin features; secondly, within both gene-rich and gene-poor domains the chromatin is folded into DNA-rich foci that typically occupy ~1 Mbp of DNA; and thirdly, the foci within the gene-rich domains occupy a larger nuclear volume (by ~3-fold) than their gene-poor counterparts, showing that the gene-rich chromatin has a more open and accessible configuration.

Such gene-rich and gene-poor domains are also known to correlate with cytologically defined sub-chromosomal domains that can be visualized as chromosomal bands. These structures define the chromosome-specific organization that is used to generate a species-specific chromosome profile or ‘karyotype’. Classically, karyotypes are generated by staining chromosomes with the dye Giemsa after first spreading them on to glass slides. The stain binds more strongly to the more condensed chromatin within the heterochromatin domains, which are thus revealed as Giemsa dark or G-bands; such bands are very variable in size but most fall in the range 5–20 Mbp of DNA. These bands are interspersed with Giemsa light or reverse bands. So called R-bands are gene-rich and contain most of the expressed genes. Importantly, individual chromosomes within nuclei do not mix freely with other chromosomes, but for the most part remain folded in a way that means that they occupy self-contained territories [26]. Territories do not occupy specific nuclear positions, although gene activity can influence interphase chromosome location: chromosomes with a high density of active genes tend to be located towards the nuclear centre, whereas those with a low density tend to lie closer to the nuclear periphery [27]. In addition, the chromatin within individual territories is often polarized, so that the gene-rich and gene-poor regions tend to cluster within discrete regions of individual territories [29,30].

Although chromosome territories generally maintain a specific steady-state structure within nuclei, this does not mean that they should be considered as static structures. In fact, the molecular interactions that define chromosome
shape are not well-established and it appears that chromosomes are able to assume a wide range of different shapes during the cell cycle. However, the extent to which changes in the position and shape of individual chromosome territories can influence their function is presently unclear [31]. Interestingly, in living cells DNA foci are often seen to be dynamic over a range 0–2.5 μm, so that in some cells chromatin can be seen to escape from individual territories to form loops that might spread for many microns into the surrounding nucleoplasm [32]. This looping behaviour occurs in response to gene expression and there is some evidence that the organization of foci can be altered or unfolded in response to changes in their functional status [33]. Although it is extremely difficult to estimate the amount of unfolded DNA that is not constrained within structural foci, this feature of organization does have important consequences as it allows the formation of nuclear compartments where DNA from adjacent chromosome territories is able to interact [34]. In fact, many recent studies have shown that it is not unusual for genes on different chromosomes to interact at common nuclear sites. When such interactions are seen, the common sites appear to be transcription factories, where, as noted above, many (typically ~5–10) active genes are transcribed together. Detailed analysis of such interactions shows that transcription within mammalian cells might often involve co-regulation of genes that can be classified as gene-expression networks based on their interaction with specific transcription factors during their expression [24,35]. Many fascinating questions concerning the structure and function of chromatin within mammalian nuclei remain to be resolved. However, there is no doubt that using new techniques which allow us to analyse specific DNA–DNA interactions with great efficiency, the secrets of chromosome architecture will continue to be revealed [36].

The nuclear matrix and the interchromatin domain
DNA organization is clearly hierarchical in higher eukaryotes and the different levels of folding will inevitably contribute to the efficiency of chromatin function. Indeed, many aspects of chromatin function are so complex when considered in molecular detail, that the compartmentalization, and hence isolation, of different functions is also likely to be required in order to preserve genome integrity. As we have seen, both the DNA and major centres of nuclear function – such as sites of RNA transcription (above) and also DNA replication [37] – can be visualized as self-contained structures. Of course, as the template for RNA and DNA synthesis is contained within the DNA foci it is inevitable that the DNA-containing and synthetic compartments must interact. However, to fully understand how nuclear organization defines the efficiency of chromatin function it is essential to understand how these interactions are controlled. The best way to approach this is to imagine the chromatin and interchromatin compartments as being independent nuclear domains, with the essential nuclear activities being performed within the boundary regions where the components of these two compartments are able...
to interact [24,31]. We have seen already how the structures of DNA foci are largely maintained by the association of specific classes of chromatin, no doubt based on the histone epi-states that are defined by their post-translational modifications. Protein structures in the interchromatin domain are largely dependent on the biophysical properties of their components and in particular their propensity to co-associate or self-assemble [21].

However, a number of compelling observations suggest that the principles of self-assembly are not sufficient to explain all features of nuclear organization, and many studies support the possibility that fundamental features of chromatin function are reliant on a global organization with properties similar to those performed by the cytoskeleton in the cytoplasm. This idea that a structural framework contributes to the regulation of nuclear functions is especially compelling if we consider that connections from the nucleus to the cytoplasm and through the cytoplasm to the extracellular matrix within tissues is able to provide a contiguous structural network that might facilitate wide-scale regulation of nuclear function [38]. In principle, this type of structural network would be able to transmit information over a long range and between different cells based on their tissue architecture.

The structural framework within higher eukaryotes has been referred to by various names, which are based on the different extractions used during their preparation; nuclear extraction is actually essential, as without removing chromatin it is not possible to see any underlying structure. Historically, the earliest extractions were performed using hypertonic extraction (with 2 M NaCl) and nuclease digestion to reveal a nuclear matrix [39]. Later studies developed detergent-based extractions, to reveal a nuclear scaffold [40], and physiological procedures to reveal a nucleoskeleton (Figure 4) [41,42]. Because of the treatments used, these three preparations contain both common and unique components. For simplicity, the following analysis focuses on the general concepts of structural organization and provides key examples to exemplify the major organizational principles.

Arguments developed above are predicated on the existence of a structural network that is able to connect throughout cells and tissues. However, although the cytoskeleton and extracellular matrix are understood in detail, the existence of a pervasive nucleoskeleton remains a matter of some debate. The nuclear lamin proteins are the best candidates to provide the structural basis of nuclear architecture [43,44]. These proteins are class V intermediate filaments that form stable ~10 nm fibres. Four major lamin proteins, called lamin A, lamin C (A and C are expressed by differential splicing from the lamin A gene, LMNA), lamin B1 (expressed from LMNB1) and lamin B2 (expressed from LMNB2) are expressed in mammalian cells. Broadly speaking, the B-type lamins are expressed in more primitive cells and the A-type lamins are expressed as cells become more specialized and differentiate. These proteins interact to form a stable network along the inner face on the nuclear membrane as the nuclear lamina and also form structures within the nucleoplasm, although detailed
molecular analysis of the networks remains elusive. Within both the nuclear lamina and internal nucleoskeleton the interaction of the lamin proteins with a wide variety of interacting protein complexes is responsible for defining the links between nuclear structure and function. The roles performed by the lamin proteins are complex and have been reviewed extensively. For the sake of simplicity, it is appropriate here to note that key functions include maintaining the structural architecture of the nuclear compartment, the spatial organization of chromatin, chromatin domains and chromosome territories, and contributing to the structure and function of centres where major chromatin functions are performed. In addition to these intermediate filament proteins, nuclei also contain nuclear-specific isoforms of actin and myosin, which are thought to regulate global nuclear architecture and chromatin dynamics [45,46].

**Proteins of the nuclear matrix**

Many proteins have been classified as components of the nuclear matrix [47,48] and nucleoskeleton [49]. This literature is too complex to discuss in detail here. However, it is worth considering one example that exemplifies how the

---

**Figure 4. The nucleoskeleton**

The nucleoskeleton of HeLa cells can be revealed if permeabilized cells are treated with nucleases to fragment and remove chromatin. In this ~250 nm thick resinless electron micrograph, the spherical, central nucleus (n) is demarcated by a dense nuclear lamina (nl). Notably, nuclear architecture is preserved even though most chromatin – approximately half of the nuclear mass – is removed. The residual nucleoskeleton is seen as a diffuse network of coated filaments that pervades the nuclear volume and maintains the spatial organization of major nuclear compartments such as nucleoli (no), replication and transcription factories, and nuclear speckles; note that the distribution of these nuclear domains is preserved even though almost all chromatin has been removed. To preserve the fragile cellular remnants during electron microscopy processing, cells were embedded in agarose micro-beads; the residual agarose filaments can be seen surrounding the cell. Image kindly provided by Pavel Hozak; see [42] for details.
nuclear matrix might present sites that are essential for correct regulation of gene function. In leading this field of research, Gary Stein and his colleagues have studied the behaviour of the runt-related transcription factors (RUNX/CBFA/AML), which play essential roles in cellular differentiation and fetal development [50]. These studies have characterized a domain within the RUNX transcriptional activators that targets the protein to discrete subnuclear foci. Removal of the targeting domain results in lethal haematopoietic and skeletal phenotypes, and implies that for this class of protein the correct nuclear location is critical for function [51]. RUNX proteins also interact with Smads, a family of signalling proteins that regulate various developmental and biological processes in response to the TGFβ (transforming growth factor β)/bone morphogenic protein family of growth factors. Interestingly, the RUNX proteins are required to target Smads to nuclear sites where transcription is performed. This implies that gene expression involves the in situ integration of critical signals through the assembly of regulatory complexes at transcriptionally active subnuclear sites [52]. This work and other studies [53,54] suggest that activating factors engage appropriate nuclear sites that are competent to perform gene expression and that the process of transcription can only occur once a gene has been recruited to the active centre.

**Gene expression and chromatin domains**

Different genes are expressed to quite different extents and many are expressed in specific cells and at precise times during development. This begs the question: how are different levels of gene expression maintained? For a specific gene, chromatin status and the availability of activating transcription factors will combine to establish an engaged transcription complex that drives RNA synthesis. The role of chromatin structure is clearly implicit to the success of this process. It has been known for many years that active and inactive genes have quite distinct chromatin states and that inert chromatin is able to spread, and so down-regulate, previously active genes [55]. Two general mechanisms are fundamental to protecting the functional status of active genes. The first principle concerns the maintenance of a fluid and active chromatin state [5–8]. This is achieved through dynamic histone modifications; histone hyperacetylation in particular correlates with an active chromatin status. The second requires that genes are organized into chromatin domains that might be regarded as the functional units of gene expression [24,31]. Chromatin status and domain structure are likely to be intimately linked, one dependent on the other. To understand how, it is necessary to appreciate the mechanisms by which chromatin domains can be established. Gene domains, at least within the confines of their natural chromosomal locus, ensure that genes are expressed at the desired levels in appropriate tissues and at the required times during development. In many cases, domain structure and chromosome organization will have evolved together to generate the desired levels of gene expression. Within domains, the maintenance of distinct chromatin states
that correlate with inert or active chromatin is fundamental to the expression

process.

Superimposed on the wide-scale chromatin domains are a series of interactions that drive the formation of chromatin loops. The formation of chromatin loops is complex, but in simplistic terms the generation of a loop will occur when two binding sites within a chromosomal locus – for example a gene domain – interact within a single protein complex, such as a transcription factory. This is exemplified by the formation of loops during β-globin gene expression, when the promoters of the active genes interact within a protein complex that also contains sequences from the upstream regulatory sequence called the LCR (locus control region) [56]. During expression, such protein complexes with multiple DNA–protein interactions have been termed the active chromatin hub [56,57]; although they clearly have many features in common with transcription factories discussed above. In addition to regulatory sequences, such as promoters, enhancers and LCRs, various structural DNA motifs also appear to contribute to the architecture of chromatin domains. These can be classified together as nuclear matrix and scaffold-attachment regions [58], which will often serve to isolate and/or insulate individual chromatin domains from the influence of neighbouring sequences [59]. Over recent years, the insulator protein CTCF (CCCTC-binding factor) has emerged as a good candidate to define boundary elements that punctuate the genome to form higher-order chromatin domains [60]. Intriguingly, sites of CTCF binding have also been shown to be sites of cohesin accumulation, suggesting that they might assume special structural properties that contribute to the architecture of chromatin loops [61]. In addition, hotspots of CTCF binding have been shown to establish unique features in the local chromatin environment, which might contribute to the formation of higher-order chromatin conformations.

Conclusions

The success of higher eukaryotes is defined by the co-operation of cells that are adapted to perform a wide variety of specialized roles. The functional abilities of such differentiated cells are defined by the specific combination of genes that they express. Hence, the regulation of gene expression during differentiation is fundamental to the success of higher forms of life. The process by which stably differentiated cells arise during development and then regulate expression of the appropriate genes is undoubtedly complex. The developmental history of each cell plays a significant role in this process, as the developmental lineage of each cell type gradually restricts their capacity for differential gene expression. Eventually, as cells approach their terminal state of differentiation, patterns of gene expression become fixed. At this point, active genes are defined by transcription factors and other regulatory proteins that are expressed. As differentiation proceeds, active genes assume a specific chromatin status that supports their transcription, whereas genes that are not expressed are condensed into inactive heterochromatin. These different
chromatin compartments have distinct properties that define the functional capabilities of the genes within them.

As almost all differentiated cells have the same gene content this regulation must operate epigenetically. In considering this epigenetic regulation in the present chapter, I have focussed on the mechanisms that are influenced by the structure of nuclei to control how the different nuclear compartments are able to interact. For simplicity, nuclei can be partitioned into two fundamental compartments: the chromatin-rich and interchromatin domains. In the chromatin compartment, DNA is folded into a network of chromatin foci, the sub-units of chromosome structure, which are linked in three-dimensional nuclear space in a way that must reflect the linear continuity of chromosomal DNA. This folding leads to the formation of discrete chromosome territories. During interphase, individual chromosomes clearly occupy a larger volume than their mitotic counterparts. However, as chromosomes swell during nuclear re-assembly following mitosis, the chromatin foci are maintained and the major changes in volume reflect their separation and formation of the interchromatin domain. This interchromatin compartment supports all the major chromatin functions, and it is within this compartment that a structural nuclear network is formed. Perhaps importantly, the separation of the chromatin and interchromatin compartments ensures that the specific interstitial zone, where these two major compartments interact, will be the most active synthetic nuclear domain. Of course, because of their distinct spatial architecture these components must be dynamic in order to interact. Interestingly, while traditionally chromatin is perceived as being a huge complex and proteins relatively small, it appears that the organization of higher-order chromatin-containing structures has evolved to accommodate a local dynamic behaviour. In clear contrast, the formation of protein super-complexes, often in association with the nucleoskeleton or nuclear matrix, will restrict the local mobility of active sites. These spatially stable nuclear domains must be targeted by dynamic chromatin during the initiation phase of gene expression. This simple arrangement ensures that the euchromatin, which inevitably maintains a more dynamic structure because of expression-related histone modifications, will be far more likely to engage the active synthetic sites than heterochromatin, which is stored in aggregated foci that have very limited mobility.

Although the rudiments of this general spatial organization have been recognized for many years, the biological consequences are only now beginning to emerge. Notably, this organization facilitates the formation of gene networks that can form in response to changes in cell signalling and so allow cells to adapt the patterns of gene expression that are induced in response to environmental cues [34, 62]. In some cases, the gene networks that interact within common active sites appears to modulate the amount of expression from individual genes, perhaps by locating them within a nuclear domain that maintains an appropriate concentration of essential transcription factors. Interestingly, in many differentiated cells the tissue-specific factors are seen to
accumulate at spatially stable and discrete nuclear sites. These sites frequently form in association with the nuclear matrix and provide a target or platform where the members of relevant gene networks are able to interact within individual transcription factories. This view of global nuclear structure provides an integrated model in which the spatial architecture of nuclear compartments represents a fundamental epigenetic mechanism to control the optimal levels of gene expression in differentiated cells.

Summary

- Individual cell types arise as a result of cell differentiation along cell lineages that reflect the developmental history of precursor stem cells.
- The genetic information within DNA must be folded inside nuclei in order to facilitate the expression of genes that define individual cell types.
- Gene expression in higher eukaryotes is performed in huge protein complexes – called transcription factories – where many genes are transcribed and processed together.
- Appropriate regulation of gene expression demands that DNA is folded into chromatin domains that also represent fundamental units of chromosome structure.
- Active sites of gene expression are spatially constrained in the interchromatin domain of mammalian nuclei through interactions with the nucleoskeleton.
- As chromatin domains and transcription factories are spatially discrete structures, chromatin must be dynamic in order to engage the active sites where gene expression is performed.

References


© The Authors Journal compilation © 2010 Biochemical Society


© The Authors Journal compilation © 2010 Biochemical Society

© The Authors Journal compilation © 2010 Biochemical Society