EMBO Member's Review

Chromatin: constructing the big picture

Bas van Steensel*

Division of Gene Regulation, Netherlands Cancer Institute, Amsterdam, The Netherlands

Chromatin is the ensemble of genomic DNA and a large number of proteins. Various genome-wide mapping techniques have begun to reveal that, despite the tremendous complexity, chromatin organization is governed by simple principles. This review discusses the principles that drive the spatial architecture of chromatin, as well as genome-wide-binding patterns of chromatin proteins.

The EMBO Journal (2011) **30**, 1885–1895. doi:10.1038/emboj.2011.135; Published online 28 April 2011

Subject Categories: chromatin & transcription

Keywords: chromatin; genomes; transcription

Introduction

Chromatin is probably the most complex molecular ensemble in the cell. It consists of genomic DNA together with all directly or indirectly associated protein and RNA molecules. This includes histones, DNA-binding factors (DBFs), the basal transcription machinery and its nascent transcripts, replication and repair machineries that copy and maintain DNA, and many other molecules that interact with any of these components. All of these components work in concert, and cannot be fully understood unless they are studied in their complete context.

The past decade has seen the development of new powerful genomics-based techniques that have substantially scaled up our ability to perform measurements on chromatin. These and other new methods have enabled us to tackle the complexity of chromatin more efficiently, and we are now beginning to see the first glimpses of the 'big picture' of chromatin. Basic global principles, some of which may be different than previously thought, are emerging. Here, these principles (summarized in Box 1) will be discussed. We will focus on two general aspects of chromatin: structure and composition. First, we will consider the spatial folding of the chromatin fibre inside the nucleus, at scales from ~1 kb to several megabases. Among others, we will reconsider the widely assumed dichotomy of 'compacted' heterochromatin versus 'open' euchromatin, an oversimplified model that may prove to be largely incorrect. Second, we will look into the proteome of chromatin, and compare the major types of chro-

matin as defined by local protein composition. Characteristics and biological functions of these chromatin types will be discussed. Finally, a model will be presented on how chromatin may target transcription factors to specific genes without a major role for chromatin (in)accessibility.

Chromatin structure: polymer physics and specific interactions

At its heart, chromatin consists of nucleosomes, which consist of ~147 bp of DNA wrapped around histone octamers. About 20–50 bp of linker DNA connects the nucleosomes. This 'beads-on-a-string' configuration can be regarded as a highly flexible polymer (Figure 1A), and to a large extent the three-dimensional folding of chromatin is driven by the principles of polymer physics (Langowski and Heermann, 2007). All available evidence indicates that a chromosome does not adopt one single reproducible configuration; instead, it can have many different ones. Nevertheless, polymer physics imposes some restrictions to the possible architecture. Computer simulations have shown that some features of chromosomes, such as aggregation of heterochromatic chromocenters or the segregation of chromosomes into distinct territories, may be explained by basic polymer behaviour with only a few additional assumptions (Cook and Marenduzzo, 2009; de Noijer *et al.*, 2009; Bohn and Heermann, 2010). But other features of chromosome folding cannot be accounted for by polymer physics alone; for example, over megabase distances, interphase chromosomes are more compacted than is predicted by computer simulations of simple polymers (Mateos-Langerak *et al.*, 2009). Indeed, it is now clear that specific biochemical mechanisms exist that contribute to the folding of chromatin. In essence, these mechanisms come in three types: local

Box 1 Principles of chromatin organization

I Three-dimensional architecture

- Architecture is driven by a combination of polymer biophysics and biochemical interactions.
- Most DNA is in a beads-on-a-string configuration with varying degrees of poorly understood local compaction.
- There are many long-range contacts between genomic loci, but most of these contacts are transient.
- Tethering to landmark structures contributes to the overall folding of chromosomes.

II Chromatin composition

- Chromatin harbours thousands of proteins and dozens of histone marks.
- Distinct combinations of proteins and histone marks form a limited number of chromatin types.
- Some chromatin types mediate gene repression, others are conducive to transcription.
- Each chromatin type assists the binding of specific DBFs to their cognate DNA motifs.

*Corresponding author. Division of Gene Regulation, Netherlands Cancer Institute, Plesmanlaan 121, Amsterdam 1066 CX, The Netherlands. Tel.: +31 20 512 2040; Fax: +31 20 512 1989; E-mail: b.v.steensel@nki.nl

Received: 10 February 2011; accepted: 6 April 2011; published online: 28 April 2011

compaction, long-range interactions and anchoring to nuclear scaffolds.

Local chromatin compaction

A popular notion is that the nucleosomal fibre is locally coiled to form a solenoid or other repetitive structure with a diameter of ~ 30 nm. Theoretically, there are several ways in which nucleosomes could be stacked to form such a fibre, and evidence for some configurations has been obtained *in vitro* using pure polynucleosomes (reviewed in Fussner *et al*, 2011). However, advanced electron microscopy methods have failed to detect significant amounts of 30 nm fibres in unfixed nuclei or mitotic chromosomes (Maeshima *et al*, 2010; Fussner *et al*, 2011). Such fibres may thus be rare, or be restricted to very specific cell types.

There may, however, be other kinds of local compaction that are more common: electron micrographs of most nuclei show clear patches of densely stained chromatin, and bio-physical fractionation of chromatin fragments also points to differential compaction (Gilbert *et al*, 2004). This local compaction may not be in the form of a regular higher-order fibre, but may instead consist of rather irregular patches of aggregated nucleosomes (Figure 1B). A variety of specialized proteins may promote such compaction, such as the linker histone H1 and Polycomb Group (PcG) protein complexes (Francis *et al*, 2004; Hizume *et al*, 2005).

Long-range interactions

A second biochemical mechanism that is important for chromosome architecture involves physical contacts between

two sequence elements that are distant on the linear chromosome (Figure 1C). Such contacts can be detected using a technique named chromatin conformation capture, of which there are now many derivatives (reviewed in Simonis *et al*, 2007; Fullwood and Ruan, 2009; van Steensel and Dekker, 2010). Application of this technology has revealed that there are many physical interactions between enhancers and promoters (Hakim *et al*, 2010). There are also frequent interactions within chromosomes over longer distances (up to ~ 10 Mb) and even between loci on different chromosomes, although the latter tend to be less abundant. Surveys in mammalian cells suggest that there may be many thousands of long-range interactions (Simonis *et al*, 2006; Fullwood *et al*, 2009; Lieberman-Aiden *et al*, 2009). Interestingly, these interactions show a global domain pattern along the chromosomes, where regions of inactive and active chromatin each preferentially interact among themselves. This suggests that chromatin is segregated into two distinct compartments (Simonis *et al*, 2006; Lieberman-Aiden *et al*, 2009). A variety of proteins is thought to act as bridging factors that mediate long-range interactions, including the insulator protein CTCF, several transcription regulators, cohesin, polycomb protein complexes and others (Sexton *et al*, 2009; Hakim *et al*, 2010).

It is important to note that most of these long-range contacts are in part stochastic: even though they are specific, most of the reported interactions between two loci occur only in a small fraction of cells at any point in time (Simonis *et al*, 2006; Lieberman-Aiden *et al*, 2009; Bantignies *et al*, 2011). It is, therefore, difficult to imagine how such

Figure 1 Principles of 3D organization of chromatin. (A, B) Variation of local compaction, illustrated by 3D computer simulations of nucleosome fibres. The high flexibility of the fibre allows it to adopt many different configurations, which may be extended (A) or compacted (B). Both fibres in these cartoons consist of 60 nucleosomes and only differ in the degree of bending of the linker DNA. (C) Long-range contacts. Linearly distant loci can physically interact in nuclear space, often held together by a specific protein complex. (D) Anchoring of specific genomic segments to fixed nuclear landmarks such as the NL (depicted by the grey lattice). Red spheres in (C) and (D) depict hypothetical anchoring proteins or protein complexes. 3D rendering movies of (A–C) are available as Supplementary data.

individual pairwise contacts contribute substantially to the regulation of gene activity, unless short contacts are somehow memorized as a long-lasting epigenetic imprint (Bantignies *et al*, 2003).

Attachments to nuclear landmarks

A third mechanism that contributes to the folding of chromosomes involves interactions of the genome with relatively fixed nuclear 'landmarks' that may act as anchoring sites (Figure 1D) (reviewed in Deniaud and Bickmore, 2009; Towbin *et al*, 2009; Kind and van Steensel, 2010; van Steensel and Dekker, 2010). One such landmark is the nuclear lamina (NL), which consists of lamin polymers. The NL coats most of the inner nuclear membrane, providing a large surface area for potential contacts with the genome. Indeed, molecular mapping studies have shown that the genomes of *Drosophila*, mouse and humans interact extensively with the NL, via so-called lamina-associated domains (LADs) (Pickersgill *et al*, 2006; Guelen *et al*, 2008; Peric-Hupkes *et al*, 2010; van Bemmelen *et al*, 2010). In mammals, there are ~1300 LADs, which have a median size of about 0.5 Mb. Combined, mammalian LADs cover about 35–40% of the genome. LADs generally have sharp borders that are marked by specific sequence elements that include binding sites for CTCF, CpG islands and promoters. Most genes in LADs are transcriptionally silent. This indicates that the NL contributes to gene repression, which is supported by de-repression of NL-associated genes in flies lacking one of the lamins (Shevelyov *et al*, 2009) and by downregulation of certain genes that are artificially tethered to the NL (Finlan *et al*, 2008; Kumaran and Spector, 2008; Reddy *et al*, 2008; Dialynas *et al*, 2010). In accordance with a role for the NL in gene regulation, hundreds of genes were observed to move towards or from the NL during differentiation (Peric-Hupkes *et al*, 2010). It is still unclear how these dynamic interactions of the genome with the NL are mediated and regulated.

Nuclear pore complexes (NPCs) also interact with specific genomic loci (Liang and Hetzer, 2010). These loci are different from LADs, which is consistent with the observation that NPCs are primarily located in gaps in the NL and thus represent a distinct micro-environment (Schermelleh *et al*, 2008). Genome-wide mapping efforts in yeast, flies and mammals have identified many genes that are bound by NPC proteins (Casolari *et al*, 2004; Capelson *et al*, 2010; Kalverda *et al*, 2010; Vaquerizas *et al*, 2010), but it has become clear that most of these proteins freely diffuse through the nucleoplasm and often interact with their target genes in the nuclear interior rather than at the NPCs (Capelson *et al*, 2010; Kalverda *et al*, 2010). Genomic maps of NPC protein interactions should, therefore, be interpreted with caution, as they only partially reflect the spatial positioning of the genome relative to the nuclear envelope. Future studies need to address the functions and mechanisms of genome–NPC interactions.

Another nuclear structure that may provide a relatively fixed anchoring platform for the genome is the nucleolus. Recent mapping studies have identified many nucleolus-associated domains (NADs), which tend to be large domains and harbour specific gene sets (Nemeth *et al*, 2010; van Koningsbruggen *et al*, 2010). Some of the NADs may overlap with LADs, suggesting that there are

genomic regions that have alternating contacts with both the NL and nucleoli.

Chromatin accessibility

Often, spatial configuration of DNA is confused with accessibility of DNA, but these are distinct—although sometimes related—aspects of chromatin structure. Chromatin accessibility is best defined as the availability of DNA sequences for molecular interactions, typically by DBFs. Nucleosomes are major determinants of *local* DNA accessibility: a DNA sequence tightly wrapped around a nucleosome is less easily bound by a DBF than the same sequence in a nucleosome-free stretch of DNA; indeed this model is supported by ample experimental data (Hayes and Wolffe, 1992; Lieb and Clarke, 2005; Rando and Ahmad, 2007). Nucleosome remodelling factors are protein complexes that can alter the positioning of nucleosomes along the DNA (Maier *et al*, 2008; Clapier and Cairns, 2009).

It is widely believed that there are mechanisms that control accessibility beyond the single nucleosome, over extended chromosomal regions. Theoretically, access could be blocked by specialized proteins that form a dense coating around a series of nucleosomes. Alternatively, a stretch of nucleosomes may be tightly compacted, thereby leaving little space for other proteins to access the DNA. These mechanisms are easy to imagine, but is the existence of such 'inaccessible heterochromatin' really supported by experimental data?

Accessibility of chromatin is often measured by exposing detergent-permeabilized nuclei to endonucleases such as DNase I, after which the cutting sites are quantitatively mapped (Weintraub and Groudine, 1976). The idea is that the cutting frequency is proportional to the accessibility of the DNA. However, permeabilization causes the dissociation of many proteins—due non-physiological buffer conditions and a simple dilution effect—and will thus create a substantially distorted picture. This does not mean that DNase I hypersensitivity mapping is useless; it has proven to be a powerful method to identify regulatory elements in the genome (Hesselberth *et al*, 2009; Song and Crawford, 2010), but one should just not forget the caveats when interpreting the results in terms of *in vivo* chromatin accessibility.

Mapping of *in vivo* methylation by DNA adenine methyltransferase (Dam) (Gottschling, 1992; Singh and Klar, 1992; Kladde and Simpson, 1994; Wines *et al*, 1996) does not have the drawbacks of endonucleases. Dam can be expressed *in vivo* (i.e. without permeabilization) and it marks any accessible GATC motif by adding a small adenine-methylation tag that does not occur endogenously in most eukaryotes. The adenine-methylation pattern can be mapped afterwards by a variety of methods. A recent genome-wide survey of Dam methylation in three *Caenorhabditis elegans* tissues (Sha *et al*, 2010) established that most of the genome is accessible, with only about two-fold variation in Dam methylation signals. Much of this variation is explained by the local positioning of nucleosomes (consistent with earlier Dam assays in yeast; Kladde and Simpson, 1994), and not by effects over extended genomic regions. A carefully controlled study in budding yeast also found less than two-fold differences in accessibility to both Dam between euchromatin and

repressive heterochromatin marked by SIR proteins (Chen and Widom, 2005).

Other DNA-interacting proteins were also found to have normal access to repressive chromatin. Exogenously expressed DBFs bind to their recognition motifs with almost the same affinity in active and repressive chromatin contexts (Chen and Widom, 2005; Filion *et al*, 2010), and also several endogenous factors were found to bind to promoters in repressive chromatin (Chen and Widom, 2005; Gao and Gross, 2008).

These results are in overall agreement with studies of chromatin accessibility at the microscopy level, done by comparing rates of diffusion of fluorescently labelled macromolecules inside morphologically defined heterochromatin and euchromatin. Such studies failed to detect substantial differences in accessibility for molecules with radial sizes up to ~ 10 nm, which corresponds to protein complexes of roughly 1 MDa (Verschure *et al*, 2003; Gorisch *et al*, 2005; Pack *et al*, 2006). Only molecules of larger sizes appear to be specifically excluded from heterochromatic regions (Gorisch *et al*, 2005). Even mitotic chromosomes, which are highly compacted, were found to be readily accessible to transcription factor binding (Chen *et al*, 2005).

In summary, while steric hindrance of DBFs may occur locally at individual positioned nucleosomes, it seems unlikely that regulated accessibility through changes in chromatin compaction constitutes a major mechanism to control the binding of proteins to DNA. Before we consider an alternative model that explains how the binding of DBFs may be regulated, we need to take a close look at the protein composition of chromatin.

Chromatin proteome

A human nucleus has about 3×10^7 nucleosomes, but there are many other chromatin proteins. For the 'big picture', it is worth to consider how many protein molecules are part of the chromatin in a typical human nucleus. As an upper limit, we will take the total number of protein molecules estimated to be present in a nucleus. A human cell is thought to contain nearly 10^{10} protein molecules (Lodish *et al*, 2000), so if we assume that 10% of these make up the cell nucleus, there are roughly 10^9 protein molecules per nucleus, or about 1 per 6bp of DNA and about 30 per nucleosome (Figure 2). Of course, not all nuclear proteins are part of chromatin, but the majority probably is—even the RNA processing machinery is associated with chromatin (Perales and Bentley, 2009; Luco *et al*, 2011)—so this estimate cannot be too far off. The abundance of individual proteins varies over several orders of magnitude. Most transcription factors are present as 10^3 – 10^5 molecules per nucleus, while some linker histones and high-mobility group proteins (which are both thought to modulate chromatin structure) are almost as abundant as nucleosomes (Kuehl *et al*, 1984; Paull *et al*, 1996).

It is also useful to consider the *diversity* of proteins that are part of chromatin. A recent proteomics study conservatively identified >500 proteins associated with purified human metaphase chromosomes (Ohta *et al*, 2010). Interphase chromatin is likely to be much more complex, because many proteins are known to dissociate from chromatin during mitosis. Systematic surveys in yeast indicate that about 1/3 of all cellular proteins are located in the nucleus (Huh

Figure 2 How many protein molecules are present in a stretch of chromatin? Computer-simulated model of nucleosomal fibre (60 nucleosomes) without (A), with 5 (B), or with 25 (C) proteins per nucleosome on average. Proteins (grey spheres) were simulated to attach randomly to DNA, histones or to other proteins, the latter yielding simulated protein 'complexes' of various sizes. The model in (C) is most in agreement with estimates of protein molecule numbers inside the nucleus. A 3D rendering movie is available as Supplementary data.

et al, 2003). If one extrapolates this to human cells, then roughly 8000 different proteins are nuclear and thus candidates to be part of human chromatin. Of these, ~ 1400 are catalogued as sequence-specific DBFs (Vaquerizas *et al*, 2009). Taken together, it is reasonable to assume that chromatin in a typical human cell consists of several thousands of different proteins.

All of these proteins exert their functions by forming complexes with DNA, RNA and other proteins. Biochemically purified chromatin protein complex typically consist of 5–50 proteins. However, these only include stably interacting proteins; current estimates suggest that the entire human interactome may contain 130 000 interactions (Venkatesan *et al*, 2009). So in chromatin, there may be several tens of thousands of distinct pairwise protein–protein interactions, of which we currently know only a tiny fraction. If we also consider the many non-coding RNA molecules that are being discovered as part of chromatin (Hung and Chang, 2010; Koziol and Rinn, 2010), it becomes clear that chromatin is an incredibly complex macromolecule.

Principal chromatin types

The myriad of chromatin proteins and possible interactions among them raises the theoretical possibility that every bit of the genome is bound by a distinct, unique combination of proteins. This idea of a 'combinatorial code' was articulated about a decade ago for histone modifications, which by themselves could potentially form many more combinations than the total number of nucleosomes in a nucleus (Strahl and Allis, 2000).

A systematic study of the combinatorial complexity of chromatin proteins was recently conducted in cultured *Drosophila* cells (Filion *et al.*, 2010). Using the DamID technology (van Steensel *et al.*, 2001), a broad set of 53 proteins, representing a cross-section of the known functional classes of chromatin components, was mapped genome wide. Several histone marks were also mapped. Integrative analysis of the maps of the 53 proteins revealed that much of the binding profiles is explained by five principal chromatin types, which are each made up of unique combinations of proteins. While some proteins mark only one of these chromatin types, many proteins are shared by two to four types. Because the Greek word *chroma* means 'colour', the five principal chromatin types are indicated by the colours YELLOW, RED, BLUE, BLACK and GREEN (Figure 3). As outlined below, these five types have distinct and sometimes remarkable characteristics.

YELLOW and RED (high-occupancy target) chromatin

Most transcriptionally active genes are marked by YELLOW or RED chromatin. YELLOW chromatin fits our general concept of active genes: it harbours components of basal transcription machinery, some DBFs, and enzymes that control histone acetylation. Perhaps surprisingly, this includes the histone deacetylases (HDACs) RPD3 and SIR2—these were originally thought to be involved in repression of gene activity, but HDACs have now also been found to be abundantly present on active genes in human cells (Wang *et al.*, 2009).

RED chromatin is also primarily associated with active genes. Yet, RED chromatin has several unique features that distinguish it from YELLOW chromatin. First, it marks primarily genes that are tissue specific, in contrast to YELLOW chromatin, which has a strong preference for ubiquitously expressed housekeeping genes. Second, while RED and YELLOW genes cover a similar range of expression levels, genes in RED chromatin mostly lack the histone mark H3K36me3, as well as MRG15, a chromodomain protein that binds H3K36me3 *in vitro*. This absence of H3K36me3

is surprising, because it was previously thought to be a general mark of transcription elongation and have a role in the repression of antisense cryptic promoters in the transcription units (Li *et al.*, 2007). Why this mark is lacking from active genes in RED chromatin is unclear.

The third and most puzzling property of RED chromatin is its enormous diversity of proteins. On average, regions marked by RED chromatin are bound by about 60% of all proteins that were tested. Recent mapping of ~50 additional chromatin proteins again detected a similar fraction in RED chromatin (BvS lab, unpublished results). Extrapolation of these results implies that there must be hundreds of different proteins that associate with each RED chromatin region. Many of these proteins seem to be functionally unrelated: for example, the DNA-binding factors GAGA factor, ecdysone receptor, MNT and Jun-related antigen all co-localize in red chromatin, yet there is no evidence for functional relationships between these factors. Hotspots where seemingly unrelated regulatory factors co-localize were observed first in fly cells (Moorman *et al.*, 2006), and more recently in *Drosophila* embryos, mouse ES cells and *C. elegans* (Chen *et al.*, 2008; Li *et al.*, 2008; MacArthur *et al.*, 2009; Gerstein *et al.*, 2010; modENCODE Consortium, 2010). Sometimes they are referred to as 'high-occupancy target' (HOT) regions. They have been observed by two independent mapping techniques (ChIP and DamID), ruling out a technical artefact. Interestingly, re-analysis of previously published ChIP-chip data (Harbison *et al.*, 2004) indicates that also the budding yeast genome harbours hotspots that are massively bound by a variety of factors—this feature was originally missed due to the use of a normalization algorithm (Harbison *et al.*, 2004) that masks sites of high occupancy (H Bussemaker, personal communication). Thus, equivalents of RED chromatin appear to exist in many species.

RED/HOT regions have reduced nucleosome density, higher nucleosome turnover and enrichment of histone variant H3.3, which is linked to nucleosome displacement (Moorman *et al.*, 2006; Filion *et al.*, 2010; modENCODE Consortium, 2010). One might, therefore, postulate that RED regions are

Figure 3 Cartoon model of the five principal chromatin types as identified in *Drosophila* cells. Some proteins are shared between two or more chromatin types. BLACK chromatin has a preference to be in contact with the NL (grey lattice, bottom left). RED chromatin has the highest overall protein occupancy.

highly accessible and thus available for binding by many proteins. However, as discussed above, differences in accessibility along the genome appear of minor influence to DBFs. Moreover, this logic is self-contradictory: it is difficult to imagine a DNA sequence being highly accessible (i.e. not bound by anything) and at the same time occupied by dozens of proteins. Indeed, it was empirically determined that the DNA-binding domain of yeast Gal4 does not have preferential access to its binding motif in RED chromatin of fly cells (Filion *et al.*, 2010). We, therefore, need to invoke a different, more specific mechanism. For example, the formation of RED chromatin may be started by a few 'pioneer' DBFs that bind to specific sequence motifs, which in turn recruit other proteins, which in turn recruit yet other proteins. A multitude of protein-protein interactions may thus drive the formation of a large, complex aggregate of proteins. It was found that the DNA-binding domain of BICOID is largely dispensable for the targeting to hotspots (Moorman *et al.*, 2006), which illustrates the prominent role of protein-protein interactions in the formation of this baffling type of chromatin.

What does RED/HOT chromatin do? Most likely, it is involved in the control of its target genes; perhaps RED regions represent a specialized type of enhancers. An interesting possibility is that RED regions may spatially cluster in so-called transcription factories, which are nuclear foci where RNA polymerase II, various regulatory proteins and multiple transcribed genes come together (Sutherland and Bickmore, 2009; Cook, 2010). In addition, RED chromatin is the first to replicate during S-phase, and enriched for origin recognition complex proteins (Filion *et al.*, 2010; modENCODE Consortium, 2010, 21177974), raising the possibility that RED chromatin regions serve as early-firing replication origins. Given the complexity of RED chromatin, it will be a substantial challenge to unravel its functions and molecular architecture.

BLUE: polycomb chromatin

BLUE chromatin is characterized by the presence of PcG proteins and the histone mark H3K27me3 (Filion *et al.*, 2010). PcG proteins primarily repress transcription and exhibit a strong preference to bind to genes involved in the regulation of developmental processes. Detailed reviews of the molecular biology of PcG proteins can be found elsewhere (Muller and Verrijzer, 2009; Morey and Helin, 2010; Sawarkar and Paro, 2010).

BLACK chromatin

A fascinating new type of chromatin is BLACK chromatin (Filion *et al.*, 2010). In *Drosophila* Kc167 cells, this type of chromatin covers nearly half of the non-repetitive genome. Essentially, all genes embedded in BLACK chromatin are transcriptionally silent; in total, BLACK chromatin contains nearly two-thirds of all silent genes, and thus is the most prevalent repressive chromatin type. A survey of gene expression data in a variety of tissues has indicated that genes that are located in BLACK chromatin (in Kc167 cells) are generally expressed in only a few tissues. Thus, BLACK chromatin appears to be dedicated to the repression of a large set of tissue-specific genes. Importantly, a transposable element carrying a reporter gene is more likely to be silenced when integrated in BLACK regions than in any of the four other major chromatin types (Filion *et al.*, 2010). This indicates that BLACK chromatin is not just a passive marker of

transcriptionally inactive chromatin, but rather makes an active contribution to repression of transcription.

How this repression is achieved is still unclear, but hints come from the handful of proteins that are now known to mark BLACK chromatin. One of these is the linker histone H1, which has previously been linked to repression of transcription (Woodcock *et al.*, 2006). Another is Lamin, the main component of the NL, which implies that BLACK chromatin is preferentially located at the nuclear periphery (Figure 3). Other BLACK marker proteins include the AT-hook protein D1, which has been implicated in higher-order packaging of chromatin (Smith and Weiler, 2010) and SuUR, a regulator of replication of heterochromatic regions in polytene chromosomes (Zhimulev *et al.*, 2003).

So far, no histone marks have been identified in BLACK chromatin, which may explain why this chromatin type has remained unnoticed until recently. It is, however, quite possible that BLACK-specific histone modifications remain to be discovered. For example, EFFETE, a highly conserved homologue of yeast Ubc4/5 that exhibits histone ubiquitination activity *in vitro*, binds along BLACK chromatin, suggesting that one of the histones may carry a specific ubiquitination mark. It is probable that additional components of BLACK chromatin remain to be discovered. Clearly, there is still much to be learned about BLACK chromatin.

GREEN chromatin: HP1 and partners are conducive to transcription

GREEN chromatin is specifically marked by HP1 and SU(VAR)3-9, together with several HP1-associated proteins (HP2 through 6) and the histone modifications H3K9me2 and H3K9me3 (Greil *et al.*, 2007; Filion *et al.*, 2010; Riddle *et al.*, 2010). It is traditionally referred to as 'heterochromatin', but with current knowledge it may be better to abandon this term (see below). GREEN chromatin occupies large domains in pericentric regions and on the 'dot' chromosome 4, both of which are rich in repetitive sequences. In addition, it is found on hundreds of genes scattered along the chromosome arms.

HP1 and SU(VAR)3-9 are well known for their ability to repress reporter genes that are integrated into HP1-rich regions (Girton and Johansen, 2008). Paradoxically, several genome-wide studies in *Drosophila* have now established that most genes that are naturally bound by HP1 are transcriptionally active (Greil *et al.*, 2003; de Wit *et al.*, 2007; Johansson *et al.*, 2007; Piacentini *et al.*, 2009; Filion *et al.*, 2010; Riddle *et al.*, 2010). This is agreement with earlier observations of individual active genes in pericentric heterochromatin (Wakimoto and Hearn, 1990; Clegg *et al.*, 1998). In cultured Kc167 cells, we have found that HP1 knockdown slightly reduces the expression of HP1-bound genes (G Hogan and BvS, unpublished data), and in flies, pericentric genes show lowered activity in HP1 mutant flies (Lu *et al.*, 2000; Schulze *et al.*, 2005). Thus, at least in *Drosophila*, HP1 is primarily a modest activator of its natural target genes. With this new insight, the term 'heterochromatin' has become inappropriate, because it is traditionally used for repressive chromatin; the term GREEN chromatin provides a neutral alternative.

HP1 paradox

How can it be that HP1 (together with its partner proteins) represses certain reporter genes, while it is conducive to transcription of the vast majority of its natural target genes?

The first important part of an explanation is the realization that all examples of heterochromatin-repressed genes in *Drosophila* were historically *selected for a repressive phenotype*. For example, out of hundreds of flies that were treated with ionizing radiation to induce chromosome rearrangements, the rare ones with a variegating expression pattern were hand picked (for good reasons: the variegating phenotype is fascinating!). In another strategy, reporter genes were randomly inserted into the genome using transposable elements as vehicles, and again flies with a variegating phenotype were selected. By this strategy, about 65 different genes were discovered that variegate when inserted near centromeres, telomeres or on the 'dot' chromosome 4 (Tweedie *et al*, 2009). This may be a substantial number, but we should not forget that thousands of other genes have *not* been reported to variegate. Thus, could it be that only certain genes are repressed when placed in a GREEN chromatin environment? Detailed investigation of one classic position effect variegation (PEV) model suggests that this may indeed be the case (Vogel *et al*, 2009). The w^{m4} mutants, originally discovered due to variegation of the *white* eye colour gene, are caused by a translocation that joins *white* and flanking genes to pericentric heterochromatin. Molecular mapping revealed that HP1 spreads across the junction into the formerly euchromatic region and covers about 20 genes including *white*. Remarkably, only the expression of *white* is reduced, while the other genes remain essentially unaffected.

Thus, at least in *Drosophila*, the dogma that HP1 and partner proteins inhibit transcription was based on eye-catching, yet non-representative reporter gene data. At present, we can only speculate how these reporter genes are repressed by HP1, while natural target genes are not. Most of the genes that exhibit PEV when exposed to HP1 are tissue-specific genes, which tend to be located in RED chromatin. It is striking that GREEN chromatin has a very different protein composition than RED chromatin, even though both are transcriptionally active. Moreover, GREEN and RED chromatin rarely are direct neighbours (U Braunschweig, GJ Filion, JG van Bemmelen, BvS, unpublished data). These observations suggest that RED and GREEN chromatin may be incompatible. It is thus conceivable that insertion of a gene that normally carries RED chromatin into a GREEN chromatin environment may lead to destabilization of the RED chromatin, and as a consequence mis-regulation of the gene. Future experiments should reveal whether this 'incompatible chromatin types' model provides a good explanation of PEV.

At present, it is still poorly understood why a specific set of several hundreds of active genes scattered throughout the fly genome are bound by HP1 and partners. HP1 target genes tend to be long and exon rich, raising the possibility that HP1 is needed to stimulate transcription elongation or regulation of RNA processing. Indeed, in *Drosophila*, HP1 interacts with the FACT elongation complex (Kwon *et al*, 2010). Furthermore, HP1 and partner proteins may have other functions, such as regulation of replication timing (Schwaiger *et al*, 2010).

Other chromatin classifications: *Drosophila* and other species

The five principal chromatin types that were identified in *Drosophila* cells represent the *major* combinations of proteins

found among the set of 53 tested proteins. It may be possible to further refine this classification by a more fine-grained algorithm. However, any sub-classification of these five states will require objective statistical or biological criteria in order to avoid arbitrariness. Importantly, we have found recently that extending the set of binding maps with 50 additional (again broadly selected) proteins does not uncover additional chromatin types (JG van Bemmelen, GJ Filion, A Rosado, W Talhout, BvS, unpublished data), indicating that the five types represent indeed the principal types.

Several other teams recently reported the classification of chromatin states based on large genome-wide data sets. In these instances, the data sets were mostly restricted to histone modifications. A set of 18 histone marks in *Drosophila* cells yielded 9 or 30 combinatorial states depending on the algorithm used (Kharchenko *et al*, 2010). Because these data were obtained in a different cell line (S2 cells), comparison to the five colours in Kc167 cells should be interpreted with caution. Nevertheless, four of the nine states in S2 cells resemble BLACK, GREEN (2 ×) and BLUE chromatin in Kc cells. The remaining five states may represent sub-states of RED and YELLOW chromatin. Because the large majority of histone marks is associated with active genes, there may have been a 'magnifying glass' effect on the classification of active chromatin, explaining why this study (Kharchenko *et al*, 2010) identified a larger number of active chromatin states than the study based on a broad selection of 53 proteins (Filion *et al*, 2010).

Interestingly, a recent study reported in this issue of *EMBO Journal* suggests that the plant *Arabidopsis thaliana* may also have a limited number of principal chromatin types. A survey of 11 histone marks together with DNA methylation identified four major combinations that together cover most of the genome (Roudier *et al*, 2011). Two of these states roughly resemble the GREEN and BLUE chromatin types found in *Drosophila*. A third state, which generally lacks any of the tested histone marks, may be the equivalent of BLACK chromatin. Finally, only one major type of transcriptionally active chromatin was found in *Arabidopsis*, unlike the two distinct states in *Drosophila*.

As many as 51 combinatorial chromatin states were identified in human lymphocytes, based on maps of 38 histone marks (Ernst and Kellis, 2010). This may be an over-classification, because several of the combinations appeared highly similar. The authors grouped the 51 combinations into 5 major classes, which roughly correspond to active promoters, active transcription units, intergenic regions near active genes, transcriptionally inactive regions and repetitive sequences. In *C. elegans*, a more conservative analysis of 33 genome-wide maps (mostly of histone marks) identified three major combinations (Gerstein *et al*, 2010).

It is not always straightforward to compare the identified chromatin types across species, because some proteins and histone marks have adopted new functions and interactions during evolution. For example, in *Drosophila*, lamin B is primarily associated with BLACK chromatin, whereas H3K9me2 is located in GREEN chromatin. In contrast, in mouse and human cells H3K9me2 and lamin B1 binding patterns overlap substantially (Guelen *et al*, 2008; Wen *et al*, 2009; Peric-Hupkes *et al*, 2010). An intriguing possibility is that BLACK and GREEN chromatin have merged in mammals; it is also conceivable that only a few proteins/marks

Figure 4 Chromatin types may act as specific guides for DBFs. Cartoon model illustrating how each chromatin type selectively supports DNA binding by DBFs through specific cooperative interactions. Chromatin types are shown as large coloured blocks surrounding the DNA double helix. Two different DBFs are depicted in orange and purple, with their corresponding recognition motifs highlighted in the same colours underneath the DNA. While most motifs occur in all chromatin types and are in principle accessible for binding, each chromatin type enhances motif binding of specific DBFs, by virtue of cooperative protein–protein interactions (indicated as matching DBF shapes and niches in the chromatin types). Thus, chromatin types help to direct each DBF to specific subsets of its recognition motif.

have swapped between chromatin types. Along the same lines, the HP1 orthologue in *A. thaliana* has evolved to bind to H3K27me3 (Zhang *et al*, 2007), pointing to task switching of components between GREEN and BLUE chromatin as defined in *Drosophila*. These examples illustrate that it will be necessary to map large and unbiased sets of chromatin components systematically in every species (and in every cell type) in order to define and compare the major chromatin types.

DBF guidance by chromatin types

DBFs are key players in gene regulation: they are the proteins that read the instructions encoded in regulatory DNA. It is, therefore, important to understand how DBFs interact with DNA. DBFs generally recognize a 4–8 bp sequence motif, often with substantial tolerance for variations in the motif. A typical 6 bp motif occurs by chance every ~4 kb, so the human genome harbours nearly 1 million motif occurrences that are potentially bound by a DBF. Yet *in vivo* the vast majority of these motifs is not occupied by the DBF. Thus, additional specificity mechanisms must exist that direct DBFs to only a subset of its motif occurrences. One such mechanism is the occlusion of individual binding motifs by positioned nucleosomes (Hayes and Wolffe, 1992; Lieb and Clarke, 2005; Rando and Ahmad, 2007). As argued above, compaction of larger chromatin segments does not appear to be an effective mechanism to prevent access of DBFs. Instead, the different chromatin types may act as DBF-specific guides through positive targeting mechanisms (Figure 4).

A survey of DBF–motif interactions has revealed that each DBF has a preference to bind to its motif in certain chromatin types only. For example, MNT binds to its motif preferentially in YELLOW and RED chromatin, while SU(HW) binds to its motif primarily in RED, BLUE and BLACK chromatin, even though the cognate motifs for both DBFs are present in all five chromatin types (Filion *et al*, 2010). Thus, each DBF has a preferred chromatin context in which it binds to its motif. These preferences are most likely due to the presence of chromatin-associated ‘helper’ proteins that assist the

DBFs. A helper protein may physically interact with the DBF and stabilize the binding of the DBFs to its motifs. Because each DBF may have different helper proteins, each chromatin type augments the interactions of a specific subset of DBFs with their motifs. Together, the chromatin types thus constitute a selection system that guides each DBF to its binding motifs in only some regions of the genome.

Big picture of chromatin: what’s next?

Chromatin is highly complex, yet the simple global principles that have emerged provide a strong foundation for future research. Among these are several principles of spatial folding and chromatin composition (Box 1). Yet many questions wait to be answered. For example: what are the molecular mechanisms that drive the assembly of the principal chromatin types? How do these types differ between cell types and species? How do they affect gene expression and other DNA-associated functions such as replication and repair? Do they contribute to epigenetic memory (Kaufman and Rando, 2010; Margueron and Reinberg, 2010), and if so, by what mechanism? How important are long-range chromatin contacts for gene regulation? How are interactions of the genome with nuclear landmarks controlled? The rapid pace of technology development suggests that we will soon be able to get answers to these important questions, and perhaps we will uncover additional basic principles.

Supplementary data

Supplementary data are available at *The EMBO Journal* Online (<http://www.embojournal.org>).

Acknowledgements

I thank members of my laboratory for helpful comments. This study was supported by an NWO-ALW VICI grant.

Conflict of interest

The author declares that he has no conflict of interest.

References

- Bantignies F, Grimaud C, Lavrov S, Gabut M, Cavalli G (2003) Inheritance of polycomb-dependent chromosomal interactions in *Drosophila*. *Genes Dev* 17: 2406–2420
- Bantignies F, Roure V, Comet I, Leblanc B, Schuettengruber B, Bonnet J, Tixier V, Mas A, Cavalli G (2011) Polycomb-dependent

- regulatory contacts between distant Hox Loci in *Drosophila*. *Cell* 144: 214–226
- Bohn M, Heermann DW (2010) Repulsive forces between looping chromosomes induce entropy-driven segregation. *PLoS One* 6: e14428

- Capelson M, Liang Y, Schulte R, Mair W, Wagner U, Hetzer MW (2010) Chromatin-bound nuclear pore components regulate gene expression in higher eukaryotes. *Cell* **140**: 372–383
- Casolari JM, Brown CR, Komili S, West J, Hieronymus H, Silver PA (2004) Genome-wide localization of the nuclear transport machinery couples transcriptional status and nuclear organization. *Cell* **117**: 427–439
- Chen D, Dundr M, Wang C, Leung A, Lamond A, Misteli T, Huang S (2005) Condensed mitotic chromatin is accessible to transcription factors and chromatin structural proteins. *J Cell Biol* **168**: 41–54
- Chen L, Widom J (2005) Mechanism of transcriptional silencing in yeast. *Cell* **120**: 37–48
- Chen X, Xu H, Yuan P, Fang F, Huss M, Vega VB, Wong E, Orlov YL, Zhang W, Jiang W, Loh YH, Yeo HC, Yeo ZX, Narang V, Govindarajan KR, Leong B, Shahab A, Ruan Y, Bourque G, Sung WK *et al* (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**: 1106–1117
- Clapier CR, Cairns BR (2009) The biology of chromatin remodeling complexes. *Annu Rev Biochem* **78**: 273–304
- Clegg NJ, Honda BM, Whitehead IP, Grigliatti TA, Wakimoto B, Brock HW, Lloyd VK, Sinclair DA (1998) Suppressors of position-effect variegation in *Drosophila melanogaster* affect expression of the heterochromatic gene light in the absence of a chromosome rearrangement. *Genome* **41**: 495–503
- Cook PR (2010) A model for all genomes: the role of transcription factories. *J Mol Biol* **395**: 1–10
- Cook PR, Marenduzzo D (2009) Entropic organization of interphase chromosomes. *J Cell Biol* **186**: 825–834
- de Noijer S, Wellink J, Mulder B, Bisseling T (2009) Non-specific interactions are sufficient to explain the position of heterochromatic chromocenters and nucleoli in interphase nuclei. *Nucleic Acids Res* **37**: 3558–3568
- de Wit E, Greil F, van Steensel B (2007) High-resolution mapping reveals links of HP1 with active and inactive chromatin components. *PLoS Genet* **3**: e38
- Deniaud E, Bickmore WA (2009) Transcription and the nuclear periphery: edge of darkness? *Curr Opin Genet Dev* **19**: 187–191
- Dialynas G, Speese S, Budnik V, Geyer PK, Wallrath LL (2010) The role of *Drosophila* Lamin C in muscle function and gene expression. *Development* **137**: 3067–3077
- Ernst J, Kellis M (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28**: 817–825
- Filion GJ, van Bommel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ, van Steensel B (2010) Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**: 212–224
- Finlan LE, Sproul D, Thomson I, Boyle S, Kerr E, Perry P, Ylstra B, Chubb JR, Bickmore WA (2008) Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet* **4**: e1000039
- Francis NJ, Kingston RE, Woodcock CL (2004) Chromatin compaction by a polycomb group protein complex. *Science* **306**: 1574–1577
- Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EG, Huang PY, Welboren WJ, Han Y, Ooi HS, Ariyaratne PN, Vega VB, Luo Y, Tan PY, Choy PY *et al* (2009) An oestrogen-receptor- α -bound human chromatin interactome. *Nature* **462**: 58–64
- Fullwood MJ, Ruan Y (2009) ChIP-based methods for the identification of long-range chromatin interactions. *J Cell Biochem* **107**: 30–39
- Fussner E, Ching RW, Bazett-Jones DP (2011) Living without 30 nm chromatin fibers. *Trends Biochem Sci* **36**: 1–6
- Gao L, Gross DS (2008) Sir2 silences gene transcription by targeting the transition between RNA polymerase II initiation and elongation. *Mol Cell Biol* **28**: 3979–3994
- Gerstein MB, Lu ZJ, Van Nostrand EL, Cheng C, Arshinoff BI, Liu T, Yip KY, Robilotto R, Rechtsteiner A, Ikegami K, Alves P, Chateigner A, Perry M, Morris M, Auerbach RK, Feng X, Leng J, Vielle A, Niu W, Rhrissorakrai K *et al* (2010) Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**: 1775–1787
- Gilbert N, Boyle S, Fiegler H, Woodfine K, Carter NP, Bickmore WA (2004) Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. *Cell* **118**: 555–566
- Girton JR, Johansen KM (2008) Chromatin structure and the regulation of gene expression: the lessons of PEV in *Drosophila*. *Adv Genet* **61**: 1–43
- Gorisch SM, Wachsmuth M, Toth KF, Lichter P, Rippe K (2005) Histone acetylation increases chromatin accessibility. *J Cell Sci* **118**: 5825–5834
- Gottschling DE (1992) Telomere-proximal DNA in *Saccharomyces cerevisiae* is refractory to methyltransferase activity *in vivo*. *Proc Natl Acad Sci USA* **89**: 4062–4065
- Greil F, de Wit E, Bussemaker HJ, van Steensel B (2007) HP1 controls genomic targeting of four novel heterochromatin proteins in *Drosophila*. *EMBO J* **26**: 741–751
- Greil F, van der Kraan I, Delrow J, Smothers JF, de Wit E, Bussemaker HJ, van Driel R, Henikoff S, van Steensel B (2003) Distinct HP1 and Su(var)3-9 complexes bind to sets of developmentally coexpressed genes depending on chromosomal location. *Genes Dev* **17**: 2825–2838
- Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, van Steensel B (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**: 948–951
- Hakim O, Sung MH, Hager GL (2010) 3D shortcuts to gene regulation. *Curr Opin Cell Biol* **22**: 305–313
- Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004) Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**: 99–104
- Hayes JJ, Wolffe AP (1992) The interaction of transcription factors with nucleosomal DNA. *Bioessays* **14**: 597–603
- Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS, Fields S, Stamatoyannopoulos JA (2009) Global mapping of protein-DNA interactions *in vivo* by digital genomic footprinting. *Nat Methods* **6**: 283–289
- Hizume K, Yoshimura SH, Takeyasu K (2005) Linker histone H1 *per se* can induce three-dimensional folding of chromatin fiber. *Biochemistry* **44**: 12978–12989
- Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK (2003) Global analysis of protein localization in budding yeast. *Nature* **425**: 686–691
- Hung T, Chang HY (2010) Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol* **7**: 582–585
- Johansson AM, Stenberg P, Pettersson F, Larsson J (2007) POF and HP1 bind expressed exons, suggesting a balancing mechanism for gene regulation. *PLoS Genet* **3**: e209
- Kalverda B, Pickersgill H, Shloma VV, Fornerod M (2010) Nucleoporins directly stimulate expression of developmental and cell-cycle genes inside the nucleoplasm. *Cell* **140**: 360–371
- Kaufman PD, Rando OJ (2010) Chromatin as a potential carrier of heritable information. *Curr Opin Cell Biol* **22**: 284–290
- Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T, Linder-Basso D, Plachetka A, Shanower G, Tolstorukov MY, Luquette LJ, Xi R, Jung YL, Park RW, Bishop EP, Canfield TP *et al* (2010) Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**: 480–485
- Kind J, van Steensel B (2010) Genome-nuclear lamina interactions and gene regulation. *Curr Opin Cell Biol* **22**: 320–325
- Kladde MP, Simpson RT (1994) Positioned nucleosomes inhibit Dam methylation *in vivo*. *Proc Natl Acad Sci USA* **91**: 1361–1365
- Kozioł MJ, Rinn JL (2010) RNA traffic control of chromatin complexes. *Curr Opin Genet Dev* **20**: 142–148
- Kuehl L, Salmond B, Tran L (1984) Concentrations of high-mobility-group proteins in the nucleus and cytoplasm of several rat tissues. *J Cell Biol* **99**: 648–654
- Kumaran RI, Spector DL (2008) A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *J Cell Biol* **180**: 51–65
- Kwon SH, Florens L, Swanson SK, Washburn MP, Abmayr SM, Workman JL (2010) Heterochromatin protein 1 (HP1) connects the FACT histone chaperone complex to the phosphorylated CTD of RNA polymerase II. *Genes Dev* **24**: 2133–2145

- Langowski J, Heermann DW (2007) Computational modeling of the chromatin fiber. *Semin Cell Dev Biol* **18**: 659–667
- Li B, Gogol M, Carey M, Pattenden SG, Seidel C, Workman JL (2007) Infrequently transcribed long genes depend on the Set2/Rpd3S pathway for accurate transcription. *Genes Dev* **21**: 1422–1430
- Li XY, MacArthur S, Bourgon R, Nix D, Pollard AD, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, Chu HC, Ogawa N, Inwood W, Sementchenko V, Beaton A, Weiszmann R, Celniker SE, Knowles DW, Gingeras T, Speed TP *et al* (2008) Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. *PLoS Biol* **6**: e27
- Liang Y, Hetzer MW (2010) Functional interactions between nucleoporins and chromatin. *Curr Opin Cell Biol* **23**: 65–70
- Lieb JD, Clarke ND (2005) Control of transcription through intragenic patterns of nucleosome composition. *Cell* **123**: 1187–1190
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289–293
- Lodish H, Berk A, Zipursky SL, Matsudaira P, Baltimore D, Darnell J (2000) *Molecular Cell Biology*, 4th edn New York: W.H. Freeman
- Lu BY, Emtage PC, Duyf BJ, Hilliker AJ, Eissenberg JC (2000) Heterochromatin protein 1 is required for the normal expression of two heterochromatin genes in Drosophila. *Genetics* **155**: 699–708
- Luco RF, Allo M, Schor IE, Kornblihtt AR, Misteli T (2011) Epigenetics in alternative pre-mRNA splicing. *Cell* **144**: 16–26
- MacArthur S, Li XY, Li J, Brown JB, Chu HC, Zeng L, Grondana BP, Hechmer A, Simirenko L, Keranen SV, Knowles DW, Stapleton M, Bickel P, Biggin MD, Eisen MB (2009) Developmental roles of 21 Drosophila transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol* **10**: R80
- Maeshima K, Hihara S, Eltsov M (2010) Chromatin structure: does the 30-nm fibre exist *in vivo*? *Curr Opin Cell Biol* **22**: 291–297
- Maier VK, Chioda M, Becker PB (2008) ATP-dependent chromatin remodeling. *Biol Chem* **389**: 345–352
- Margueron R, Reinberg D (2010) Chromatin structure and the inheritance of epigenetic information. *Nat Rev Genet* **11**: 285–296
- Mateos-Langerak J, Bohn M, de Leeuw W, Giromus O, Manders EM, Verschure PJ, Indemans MH, Gierman HJ, Heermann DW, van Driel R, Goetze S (2009) Spatially confined folding of chromatin in the interphase nucleus. *Proc Natl Acad Sci USA* **106**: 3812–3817
- modENCODE Consortium (2010) Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* **330**: 1787–1797
- Moorman C, Sun LV, Wang J, de Wit E, Talhout W, Ward LD, Greil F, Lu XJ, White KP, Bussemaker HJ, van Steensel B (2006) Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. *Proc Natl Acad Sci USA* **103**: 12027–12032
- Morey L, Helin K (2010) Polycomb group protein-mediated repression of transcription. *Trends Biochem Sci* **35**: 323–332
- Muller J, Verrijzer P (2009) Biochemical mechanisms of gene regulation by polycomb group protein complexes. *Curr Opin Genet Dev* **19**: 150–158
- Nemeth A, Conesa A, Santoyo-Lopez J, Medina I, Montaner D, Peterfia B, Solovei I, Cremer T, Dopazo J, Langst G (2010) Initial genomics of the human nucleolus. *PLoS Genet* **6**: e1000889
- Ohta S, Bukowski-Wills JC, Sanchez-Pulido L, Alves Fde L, Wood L, Chen ZA, Platani M, Fischer L, Hudson DF, Ponting CP, Fukagawa T, Earnshaw WC, Rappsilber J (2010) The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. *Cell* **142**: 810–821
- Pack C, Saito K, Tamura M, Kinjo M (2006) Microenvironment and effect of energy depletion in the nucleus analyzed by mobility of multiple oligomeric EGFPs. *Biophys J* **91**: 3921–3936
- Paull TT, Carey M, Johnson RC (1996) Yeast HMG proteins NHP6A/B potentiate promoter-specific transcriptional activation *in vivo* and assembly of preinitiation complexes *in vitro*. *Genes Dev* **10**: 2769–2781
- Perales R, Bentley D (2009) ‘Cotranscriptionality’: the transcription elongation complex as a nexus for nuclear transactions. *Mol Cell* **36**: 178–191
- Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SW, Solovei I, Brugman W, Graf S, Flicek P, Kerkhoven RM, van Lohuizen M, Reinders M, Wessels L, van Steensel B (2010) Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* **38**: 603–613
- Piacentini L, Fanti L, Negri R, Del Vescovo V, Fatica A, Altieri F, Pimpinelli S (2009) Heterochromatin protein 1 (HP1a) positively regulates euchromatic gene expression through RNA transcript association and interaction with hnRNPs in Drosophila. *PLoS Genet* **5**: e1000670
- Pickersgill H, Kalverda B, de Wit E, Talhout W, Fornerod M, van Steensel B (2006) Characterization of the Drosophila melanogaster genome at the nuclear lamina. *Nat Genet* **38**: 1005–1014
- Rando OJ, Ahmad K (2007) Rules and regulation in the primary structure of chromatin. *Curr Opin Cell Biol* **19**: 250–256
- Reddy KL, Zullo JM, Bertolino A, Singh H (2008) Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* **452**: 243–247
- Riddle NC, Minoda A, Kharchenko PV, Alekseyenko AA, Schwartz YB, Tolstoukhov MY, Gorchakov AA, Jaffe JD, Kennedy C, Linder-Basso D, Peach SE, Shanower G, Zheng H, Kuroda MI, Pirrotta V, Park PJ, Elgin SC, Karpen GH (2010) Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. *Genome Res* **21**: 147–163
- Roudier F, Roudier F, Ahmed I, Bérard C, Sarazin A, Mary-Huard T, Cortijo S, Bouyer D, Caillieux E, Duvernois-Berthet E, Al-Shikhley L, Giraut L, Després B, Drevensek S, Barneche F, Derozier S, Brunaud V, Aubourg S, Schnitzger A, Bowler C *et al* (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J* (advance online publication, 12 April 2011; doi:10.1038/emboj.2011.103)
- Sawarkar R, Paro R (2010) Interpretation of developmental signaling at chromatin: the Polycomb perspective. *Dev Cell* **19**: 651–661
- Schermlle L, Carlton PM, Haase S, Shao L, Winoto L, Kner P, Burke B, Cardoso MC, Agard DA, Gustafsson MG, Leonhardt H, Sedat JW (2008) Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* **320**: 1332–1336
- Schulze SR, Sinclair DA, Fitzpatrick KA, Honda BM (2005) A genetic and molecular characterization of two proximal heterochromatic genes on chromosome 3 of Drosophila melanogaster. *Genetics* **169**: 2165–2177
- Schwaiger M, Kohler H, Oakeley EJ, Stadler MB, Schubeler D (2010) Heterochromatin protein 1 (HP1) modulates replication timing of the Drosophila genome. *Genome Res* **20**: 771–780
- Sexton T, Bantignies F, Cavalli G (2009) Genomic interactions: chromatin loops and gene meeting points in transcriptional regulation. *Semin Cell Dev Biol* **20**: 849–855
- Sha K, Gu SG, Pantalena-Filho LC, Goh A, Fleenor J, Blanchard D, Krishna C, Fire A (2010) Distributed probing of chromatin structure *in vivo* reveals pervasive chromatin accessibility for expressed and non-expressed genes during tissue differentiation in *C. elegans*. *BMC Genomics* **11**: 465
- Shevelyov YY, Lavrov SA, Mikhaylova LM, Nurminsky ID, Kulathinal RJ, Egorova KS, Rozovsky YM, Nurminsky DI (2009) The B-type lamin is required for somatic repression of testis-specific gene clusters. *Proc Natl Acad Sci USA* **106**: 3282–3287
- Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**: 1348–1354
- Simonis M, Kooren J, de Laat W (2007) An evaluation of 3C-based methods to capture DNA interactions. *Nat Methods* **4**: 895–901
- Singh J, Klar AJ (1992) Active genes in budding yeast display enhanced *in vivo* accessibility to foreign DNA methylases: a novel *in vivo* probe for chromatin structure of yeast. *Genes Dev* **6**: 186–196
- Smith MB, Weiler KS (2010) Drosophila D1 overexpression induces ectopic pairing of polytene chromosomes and is deleterious to development. *Chromosoma* **119**: 287–309
- Song L, Crawford GE (2010) DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harb Protoc* **2010**, pdb.prot5384
- Strahl BD, Allis CD (2000) The language of covalent histone modifications. *Nature* **403**: 41–45
- Sutherland H, Bickmore WA (2009) Transcription factories: gene expression in unions? *Nat Rev Genet* **10**: 457–466
- Towbin BD, Meister P, Gasser SM (2009) The nuclear envelope—a scaffold for silencing? *Curr Opin Genet Dev* **19**: 180–186

- Tweedie S, Ashburner M, Falls K, Leyland P, McQuilton P, Marygold S, Millburn G, Osumi-Sutherland D, Schroeder A, Seal R, Zhang H (2009) FlyBase: enhancing *Drosophila* gene ontology annotations. *Nucleic Acids Res* **37**: D555–D559
- van Bemmel JG, Pagie L, Braunschweig U, Brugman W, Meuleman W, Kerkhoven RM, van Steensel B (2010) The insulator protein SU(HW) fine-tunes nuclear lamina interactions of the *Drosophila* genome. *PLoS One* **5**: e15013
- van Koningsbruggen S, Gierlinski M, Schofield P, Martin D, Barton GJ, Ariyurek Y, den Dunnen JT, Lamond AI (2010) High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol Biol Cell* **21**: 3735–3748
- van Steensel B, Dekker J (2010) Genomics tools for unraveling chromosome architecture. *Nat Biotechnol* **28**: 1089–1095
- van Steensel B, Delrow J, Henikoff S (2001) Chromatin profiling using targeted DNA adenine methyltransferase. *Nat Genet* **27**: 304–308
- Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM (2009) A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**: 252–263
- Vaquerizas JM, Suyama R, Kind J, Miura K, Luscombe NM, Akhtar A (2010) Nuclear pore proteins nup153 and megator define transcriptionally active regions in the *Drosophila* genome. *PLoS Genet* **6**: e1000846
- Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI, Yildirim MA, Simonis N, Heinzmann K, Gebreab F, Sahalie JM, Cevik S, Simon C, de Smet AS, Dann E, Smolyar A *et al* (2009) An empirical framework for binary interactome mapping. *Nat Methods* **6**: 83–90
- Verschure PJ, van der Kraan I, Manders EM, Hoogstraten D, Houtsmuller AB, van Driel R (2003) Condensed chromatin domains in the mammalian nucleus are accessible to large macromolecules. *EMBO Rep* **4**: 861–866
- Vogel MJ, Pagie L, Talhout W, Nieuwland M, Kerkhoven RM, van Steensel B (2009) High-resolution mapping of heterochromatin redistribution in a *Drosophila* position-effect variegation model. *Epigenetics Chromatin* **2**: 1
- Wakimoto BT, Hearn MG (1990) The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of *Drosophila melanogaster*. *Genetics* **125**: 141–154
- Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W, Zhao K (2009) Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138**: 1019–1031
- Weintraub H, Groudine M (1976) Chromosomal subunits in active genes have an altered conformation. *Science* **193**: 848–856
- Wen B, Wu H, Shinkai Y, Irizarry RA, Feinberg AP (2009) Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat Genet* **41**: 246–250
- Wines DR, Talbert PB, Clark DV, Henikoff S (1996) Introduction of a DNA methyltransferase into *Drosophila* to probe chromatin structure *in vivo*. *Chromosoma* **104**: 332–340
- Woodcock CL, Skoultschi AI, Fan Y (2006) Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res* **14**: 17–25
- Zhang X, Germann S, Blus BJ, Khorasanizadeh S, Gaudin V, Jacobsen SE (2007) The Arabidopsis LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat Struct Mol Biol* **14**: 869–871
- Zhimulev IF, Belyaeva ES, Makunin IV, Pirrotta V, Volkova EI, Alekseyenko AA, Andreyeva EN, Makarevich GF, Boldyreva LV, Nanayev RA, Demakova OV (2003) Influence of the SuUR gene on intercalary heterochromatin in *Drosophila melanogaster* polytene chromosomes. *Chromosoma* **111**: 377–398