Genome architecture: from linear organisation of chromatin to the 3D assembly in the nucleus

Joana Sequeira-Mendes¹ · Crisanto Gutierrez¹

Received: 26 February 2015 / Revised: 9 August 2015 / Accepted: 12 August 2015
© Springer-Verlag Berlin Heidelberg 2015

Abstract The genetic information is stored in the eukaryotic nucleus in the form of chromatin. This is a macromolecular entity that includes genomic DNA and histone proteins that form nucleosomes, plus a large variety of chromatin-associated non-histone proteins. Chromatin is structurally and functionally organised at various levels. One reveals the linear topography of DNA, histones and their post-translational modifications and non-histone proteins along each chromosome. This level provides regulatory information about the association of genomic elements with particular signatures that have been used to define chromatin states. Importantly, these chromatin states correlate with structural and functional genomic features. Another regulatory layer is established at the level of the 3D organisation of chromatin within the nucleus, which has been revealed clearly as non-random. Instead, a variety of intra- and inter-chromosomal genomic domains with specific epigenetic and functional properties has been identified. In this review, we discuss how the recent advances in genomic approaches have contributed to our understanding of these two levels of genome architecture. We have emphasised our analysis with the aim of integrating information available for yeast, *Arabidopsis*, *Drosophila*, and mammalian cells. We consider that this comparative study helps define common and unique features in each system, providing a basis to better understand the complexity of genome organisation.

Introduction

Chromatin is a highly dynamic structure constituted by genomic DNA and a variety of proteins, among which are histones and a myriad of DNA-binding factors. The basic unit of chromatin is the nucleosome, comprised of 147 bp of DNA wrapped around a histone octamer of the four core histones: H2A, H2B, H3 and H4 (Kouzarides 2007). The linker histone H1 asymmetrically binds to the DNA at the entering/exiting duplex fibre adjacent to the nucleosome core, and it is thought to play a key role on chromatin condensation and formation of higher-order structures (Happel and Doenecke 2009). Histones are globular proteins with protruding unstructured N-terminal tails that are strikingly prone to post-translational modifications (PTMs) in a multitude of residues (for review, see (Bannister and Kouzarides 2011; Feng and Jacobsen 2011; Gurard-Levin and Almouzni 2014)). Post-translational modifications of histones can occur in any histone, but they are largely concentrated in H3. The most frequent modifications are acetylations, methylations, and phosphorylations though others such as ubiquitylation, citrullination or nitrosylation can also produce functional consequences (Bannister and Kouzarides 2011; Christophorou et al. 2014). Additionally, the core histones may be substituted by histone variants, introducing a surplus range of flavours in the epigenome that contributes to the regulation of genome function (Gurard-Levin and Almouzni 2014; Maze et al. 2014). The functional relevance of PTMs depends on a strict balance between enzymes responsible for modifying histone residues (writers and erasers) and proteins that recognise and bind modified histones (readers) and transmit to a downstream signalling cascade. For example, in the case of gene expression control, some marks are typically associated with gene repression while others with gene activation. Altogether, the covalent alterations of histone aminoacids, along with modifications

✉ Joana Sequeira-Mendes
jsequeiramendes@cbm.csic.es

✉ Crisanto Gutierrez
cgutierrez@cbm.csic.es

¹ Centro de Biología Molecular Severo Ochoa, CSIC-UAM, Nicolás Cabrera 1, Cantoblanco, 28049 Madrid, Spain

in the DNA molecule and the action of small regulatory RNAs or chromatin remodelling complexes, can define distinct chromatin properties. This high-order complex is directly involved in crucial cellular processes such as transcription, replication, endoreplication or chromosome segregation, and it orchestrates diverse patterns of cell differentiation and plasticity, in a fine tune regulation of the genome activity.

Chromatin organisation and topography

Over the last years, the generation of genome-wide datasets for an increasing number of epigenetic marks, small RNA molecules, and chromatin-associated proteins immeasurably contributed to the current knowledge of the molecular mechanisms controlling genome function. Preliminary analysis of these data showed that certain histone modifications were often found in combination with other marks, while others were mutually exclusive. The generation of new computational algorithms allowed a step ahead in the meta-analysis of the available large-scale data. Such analyses led to the decoding of combinatorial patterns of histone modifications and DNA-binding factors that, in turn, reflect chromatin organisation and function. Some examples of such algorithms include ChromaSig, an unsupervised learning method of unbiased clustering of chromatin signatures (Hon et al. 2008); ChromHMM, a procedure based on a multivariate hidden Markov model (HMM) (Ernst and Kellis 2012); or more recently, and expanding the HMM methodology, a Bayesian network that models observed combinations of chromatin modifications (Biesinger et al. 2013). The latter defines not only the chromatin states but also the state transition propensities between different cell types.

In yeast, mapping and evaluation by Principal Component Analysis (PCA) of histone modifications at single nucleosome resolution showed a correlation between the different patterns and their locations throughout the gene bodies or promoter regions. Moreover, the observed combinations were foretelling of the transcriptional activity of the corresponding gene (Liu et al. 2005).

A comprehensive profiling describing the differences in the chromatin states and transcriptional competence of pluripotent and lineage-committed mouse cells revealed that differentiation is accompanied by chromatin changes in bivalent promoter regions (chromatin territories that harbour marks present in active—H3K4me3—and repressed—H3K27me3—chromatin). The observed changes reflect modifications in the transcriptional status of the corresponding genes (Mikkelsen et al. 2007). This study focused mainly on the regulatory regions of different cell populations and alleles and on the prediction of new promoter and enhancer regions. Since then, many other surveys with an unbiased genome-wide scope have followed. Indeed, the whole mouse ESC genome was later on

categorised into three domains (active, non-active, and null), based on multiple histone modification maps (Larson and Yuan 2010). On the other hand, and taking advantage of the massive amount of large-scale histone modification data generated, the human genome has been thoroughly analysed and sorted into a variable number of chromatin signatures in diverse cell types that correlate with functional elements (Ernst and Kellis 2010; Ernst and Kellis 2013; Ernst et al. 2011; Hon et al. 2008; McDaniell et al. 2010). In fact, the Roadmap Epigenomics Consortium has just published a comprehensive and unifying analysis of more than a hundred epigenomes (including 111 generated by the programme and 16 epigenomes previously generated by ENCODE). In this endeavour, histone modifications, DNA accessibility, DNA methylation, and expression profiles from distinct cell types and tissues were integrated. The high-resolution maps obtained shed light into the key role of the epigenetic layer of information in gene regulation, development, and disease (Roadmap Epigenomics Consortium et al. 2015).

In *Drosophila*, a comprehensive analysis of chromatin organisation was first performed by integrating the genome-wide binding maps of 53 chromatin-associated proteins and the profiles of four key histone marks as global validators of the defined states (Filion et al. 2010). Besides the non-overlapping repressive states associated with polycomb group (PcG) proteins and HP1 heterochromatin, the authors described two classes of transcriptionally active euchromatin that mainly differ in the classes of genes they harbour. As a striking feature, almost 50 % of the genome is part of a gene-poor, silent chromatin state that is depleted in most of the proteins analysed and could play a role in active transcriptional silencing. Interestingly enough, a subset of just five proteins can restate the classification obtained with the systematic mapping. This strongly suggests that the backbone of chromatin organisation is driven by a small number of master regulators and that co-localising protein factors may be interaction partners or downstream elements with no direct role in the local architecture. A succeeding analysis of the chromatin landscape in flies, based on 18 histone modification patterns, and integrated with additional data, defined nine prevalent combinations of marks (Kharchenko et al. 2010). In this study, the active chromatin was further divided in distinct classes that correlate the observed chromatin signatures with gene length, gene features, and milieu. Moreover, the authors outlined distinct combinations of marks within PcG-associated chromatin.

The analysis of the distribution of several chromatin-linked proteins, histone variants, and histone modifications in two developmental stages of *Caenorhabditis elegans* led to the identification of five main clusters that correlated with gene expression, gene repression, and the X chromosome (Liu et al. 2011).

The wide range of information collected on epigenetic marks in *Arabidopsis thaliana*, together with the availability of a large collection of gain-of-function and loss-of-function

transgenic lines for functional studies, has made this organism one of the preferred model systems in the epigenetic field. *Arabidopsis* chromosome 4 was categorised into four main chromatin states, based on the combinatorial preferences of DNA methylation and several histone marks (Roudier et al. 2011). The described domains, representing active, repressed, silent, and intergenic regions, were shown to be predominantly small and scattered along the chromosome, reflecting the compact nature of the *Arabidopsis* genome. A subsequent analysis comprehensively integrated the information of several histone marks, histone variants, DNA methylation, and primary DNA sequence features (Sequeira-Mendes et al. 2014). In this study, the epigenome of *Arabidopsis* was sorted into nine prevalent combinations of marks, outside the limits of classical euchromatin and heterochromatin early described by the botanist Emil Heitz in 1928 (Heitz 1928). Each chromatin state was associated with distinct levels of DNaseI chromatin accessibility (Shu et al. 2012) and with described functional genomic elements: promoters, transcription start sites (TSS), gene bodies of active and repressed genes, PcG-repressed domains, intergenic regions, and two classes of heterochromatin that mainly differ in their nucleotide content (Sequeira-Mendes et al. 2014). Analysis of the linear relationships between chromatin states demonstrated the preference for a small number of motifs that describe active, PcG-regulated, and heterochromatin regions in the genome. Interestingly, an intergenic state enriched in PcG marks seems to play an important role as a hub (i.e. the state that connects all other groups of states) between active chromatin and heterochromatin. Although some evidence is starting to be reported (see “Polycomb group-regulated chromatin” section), detailed studies will be required to deepen the knowledge on the role of H3K27me3 in the chromatin topology.

The defined number of chromatin signatures can change for every given genome depending on the resolution of the survey, the mathematical model, and the set and number of histone marks or protein-binding factors examined. Albeit these differences, the highlight should reside on the integration of chromatin signature and genomic function.

The analyses of the prevalent combinations of epigenetic modifications in the various model organisms are unravelling patterns of nucleosome flavours that reflect distinct genomic elements. We will now focus on the main characteristics of the fundamental elements in the eukaryotic genome. Summary and integration of the key chromatin signatures defining genomic elements per model system are depicted in Fig. 1.

Promoters and TSS

Promoters are functional elements typically located immediately upstream (5'-) of the TSS of a gene, in the same DNA strand. Encompassing binding sites for the RNA polymerase and for transcription factors, promoters contain the elements

responsible for gene expression control and their size and structure vary widely. The particular combinations of histone modifications found at promoters reveal the regulatory status of these elements and associated genes.

In yeast, the proximal domain of most RNA PolII promoters comprises two hypo-acetylated nucleosomes in H2BK16, H4K8, and H4K16. These nucleosomes, located immediately upstream of the coding region, are also depleted in H3K4me1/2 and do not correlate with transcription levels (Liu et al. 2005). Notably, the surrounding nucleosomes tend to be highly acetylated and the observed transitions in the histone modification patterns are quite sharp around the TSS. When nucleosomes around the TSS are analysed separately by transcriptional level, main differences arise and two distinct chromatin states are defined: nucleosomes surrounding TSS of poorly expressed genes are depleted in H3K4me3 and hypo-acetylated at H3K9/K14/K18, H4K5/K12, and H2AK7, and on the other hand, these same residues are hyper-acetylated in highly expressed genes and the corresponding nucleosomes are enriched in H3K4me3 (Liu et al. 2005).

Several chromatin states define human promoter regions (Ernst and Kellis 2010; Ernst and Kellis 2013; Ernst et al. 2011). In common, they are defined by enrichments in H3K4me3 and they are globally demarcated by regions of open chromatin, sustained by the observed hypersensitivity to DNaseI digestion. Human promoters are rich in CpG islands, transcription factor-binding sites, and conserved elements (Ernst and Kellis 2010). Nonetheless, they differ in particular combinations of histone modifications, such as H3K79me2/3, H4K20me1, H3K4me1/2, and H3K9me1, or various acetylations that mirror the transcriptional levels of the associated genes. Thus, they can comprehensively be clustered into active (enriched in H3K9ac, H3K27ac, H3K4me2/3), weak (enriched in H3K4me2/3 with moderate enrichments in H3K9ac and H3K4me1), and poised (enriched in H3K4me2 and H3K27me3 with modest enrichments in H3K4me1/3) states (Ernst et al. 2011). Functionally, the antagonistic co-existence of the H3K27me3 and the H3K4me2/3 modifications defines a poised promoter while the H3K4me2/3 together with H3K9ac and H3K27ac signature indicates promoter activation.

In flies, the first study to systematically categorise chromatin sorted actively transcribed chromatin into two main states that encompass promoters (Filion et al. 2010). These domains, vastly gene-dense and with similar expression levels, were globally characterised by enrichments in H3K4me2 and H3K79me3, and depleted in the H3K27me3 and H3K9me2 marks, although they display striking differences in the associated proteins and DNA replication timing. A subsequent study at higher resolution identified a chromatin state typical of active promoter and TSS-proximal regions defined by high levels of H3K4me2/3 and H3K9ac (Kharchenko et al. 2010), as observed in human promoters. It is noteworthy that the list

Fig. 1 A key to identify functional genomic elements in different model systems based on chromatin states. The epigenetic signatures of the associated functional genomic elements are represented as a heatmap. The information depicted in this figure is based on the available chromatin profiling data for the different model organisms (Brown and

Bachtrög 2014; Ernst and Kellis 2010; Ernst et al. 2011; Filion et al. 2010; Gu and Fire 2010; Ho et al. 2014; Kharchenko et al. 2010; Liu et al. 2011; Roudier et al. 2011; Sequeira-Mendes et al. 2014). 1) *A. thaliana* and *C. elegans* lack CTCF, NA no available data in the published chromatin states datasets

of genes associated with this active promoter signature is most conserved between sexes and across fly species (Brown and Bachtrög 2014).

The combination of marks present in a promoter is very conserved across kingdoms. In *Arabidopsis*, active euchromatin is characterised by high levels of open chromatin marks such as H3K4me2/3, H3K36me3, H3K9me3, H2Bub, and H3K56ac (Roudier et al. 2011). This state was eventually refined to sort out chromatin signatures representative of

proximal and distal promoter regions and TSS (Fig. 2) (Sequeira-Mendes et al. 2014). Active proximal promoter regions and TSS vicinities are particularly sensitive to DNaseI digestion, and thus highly accessible (Shu et al. 2012). These regions are mainly defined by the combination of H3K4me2 and H3K4me3, H3 acetylation, H3K36me3, and H2Bub, and by a relatively low nucleosome occupancy, enriched in the variants H3.3 and H2A.Z (Sequeira-Mendes et al. 2014; Shu et al. 2014). Similarly to what is observed in mammalian

Fig. 2 Functional genomic elements are characterised by specific combinations of epigenetic marks. A schematic representation of the distinct domains in a higher eukaryote genome is presented, using *Arabidopsis thaliana* as model. In each chromatin state, histone modifications, histone variants and DNA features that are enriched over the genome average appear in *vermillion*, whereas the ones that are depleted over the genome average are represented in *blue*. The transcriptional unit in active euchromatin is divided (from *right to left*)

into distal promoter, proximal promoter, TSS/5'-end of gene, transcription elongation, gene body and introns and gene body/3'-end of gene/TTS. The repressed euchromatin is divided into genic units and intergenic regions. In heterochromatin, TE density exemplifies two different chromatin types: the single TE denotes lower TE enrichment in the AT-rich heterochromatin, whereas several TEs represent the GC-rich heterochromatin

poised promoters, a set of active modifications but also comprising the repressive PcG mark H3K27me3 defines two bivalent promoter states. These states encompass most promoter regions in the *Arabidopsis* genome and mostly differ in the levels of marks typical of active transcription and in their relative localisation, defining two signatures for the most proximal and distal parts of the promoter elements.

The pattern of marks associated to promoters of highly expressed genes in *C. elegans* also defines a chromatin state that contains H3K4me2 and H3K4me3, H3K27ac, H4K8ac, H4K16ac, H4tetra-ac and HTZ-1/H2A.Z (Liu et al. 2011). Based on a comparative analysis of metazoan chromatin organisation (human, fly, and worm), the promoter state was comprehensively defined by H3K4me3, H3K4me1, H3K27ac, and H3K79me2 (Ho et al. 2014). Moreover, metazoans share architectural similarities, exhibiting a well-positioned +1 nucleosome at the TSS of expressed genes. Although many features in the chromatin organisation of metazoan promoters are shared, there are also some differences. Namely, the H3K4me3 enrichment profiles around the TSS in worms and human display a bimodal distribution, while in flies, they show a single peak downstream the TSS. However, these differences remain unexplained, and bidirectional transcription, chromatin accessibility or nucleotide content were discarded as possible factors (Ho et al. 2014).

Enhancers

Enhancers are among the most scrutinised *cis*-regulatory elements. These sequences can be located several kilobases away from the TSS, establishing contacts with promoter sequences through chromatin loops that potentiate transcription of the target genes. Their activity is crucial in the maintenance of specific patterns of gene expression, so critical in the processes of differentiation and development (for review, see (Calo and Wysocka 2013; Spitz and Furlong 2012)).

In mammalian genomes, enhancers are characterised by high levels of H3K4me1, the histone variant H2A.Z, various acetylated residues (i.e. H3K27ac), and the presence of CTCF. Compared to active promoters, active enhancers display a higher ratio of H3K4me1 to H3K4me3. Although enhancers represent a small fraction of the genome, their chromatin signature is distinctive enough to sort well-defined chromatin states in the human genome (Ernst and Kellis 2010). In common, they have an increased DNaseI hypersensitivity and high levels of H3K4me1. Additionally, they are enriched in transcription factor-binding motifs and regulatory sequences. Human enhancers can be globally clustered according to the transcriptional level of the proximal genes (Ernst et al. 2011). Strong enhancer state is defined by high levels of H3K4me1 and H3K4me2 (with or without H3K4me3), H3K9ac, H3K27ac, and the variant H2A.Z, and it is occupied by

RNA PolIII. On the other hand, weak or poised enhancers only contain enrichments in H3K4me1 and H3K4me2 and no detectable RNA PolIII is found at those sequences. These states are frequently interchangeable depending on the cell type analysed, in agreement with the enhancer function in establishing spatial and temporal specific expression patterns. The chromatin signature of enhancers described above is largely conserved across metazoans (Ho et al. 2014), reflecting the functional conservation of these regulatory elements.

Prevalent combinations of chromatin marks that coincide both with enhancer activity in functional assays and with the binding of an identified enhancer binding protein—the histone acetyltransferase p300, from the coactivator family CBP/p300 (Visel et al. 2009)—were also found in *Drosophila* (Kharchenko et al. 2010). Largely located inside intronic regions, these small-size domains show a high nucleosome turnover, lack H3K4me3, and are enriched in H3K4me1, H3K18ac and H3K27ac, as occurs with mammalian enhancers. The fraction of genes in the same chromatin state is conserved between male and female flies and across *Drosophila* species (Brown and Bachtrog 2014).

A chromatin signature that is depleted in some of the classical marks associated with active chromatin but enriched in H2A.Z and H3K4me1 also exists in *Arabidopsis* (Sequeira-Mendes et al. 2014). It is represented mainly around transcription termination sites (TTS) of genes and some intergenic regions (Fig. 2). However, very little is known about enhancer elements in *Arabidopsis*, with only few examples deriving from enhancer trap assays (Ott and Chua 1990; Sundaresan et al. 1995). Although the described chromatin fingerprint could represent potential enhancer regions, functional assays are still required to validate this hypothesis.

Active genes

Open chromatin regions are usually associated with high expression levels of the residing genes. In yeast, a gradient of chromatin signatures can distinguish the nucleosomes in the 5'-end, the middle, and the 3'-end of transcribed genes (Liu et al. 2005). The nucleosomes in the most upstream regions of the gene show enrichments in H3K4me3 and several acetylations in histones H3 and H4, including the previously described H3K9/14ac. The trimethylation of H3K4 turns to dimethylation and finally monomethylation, in a 5' to 3' gradient.

In human cells, chromatin is also sorted into various patterns of histone modifications that correlate with their location over the bodies of active genes as well as with the expression levels (Ernst and Kellis 2010; Ernst et al. 2011). Transcription-associated regions are defined by different combinations of the H3K79me, H3K27me1, H2BK5me1, H4K20me1, and elongation mark H3K36me3. These classify 5' proximal, 5' distal, spliced exon-enriched, and 3'-end states. TTS cluster in

a domain which is characterised by the presence of PolIII and high frequencies of H3K36me3 and H4K20me1, and by the absence of H3K4 methylation. Intriguingly, the chromatin state that tends to be enriched in TTS in *Arabidopsis* is mostly enriched in H3K4me1 and H2A.Z, maybe reflecting the loss of histone code sharpness by the compact nature of this model organism's genome. As a striking curiosity, human zinc-finger genes seem to contain a specific chromatin signature, grounded on the association of H3K9me3 and H4K20me3 with these genes, the presence of H3K36me3, and lower levels of other marks (Ernst and Kellis 2010). Indeed, a strong association with the repressive marks H3K9me3 and H4K20me3 had been reported for ZNF gene repeats (Barski et al. 2007). It would be interesting to address whether this is a universal feature of this gene family, as many of the members have orthologues in animals, plants, and fungi.

Active euchromatin in *Drosophila* largely presents high frequencies of H3K4me2 and H3K79me3, and depletion of the H3K9me2 and H3K27me3 marks (Filion et al. 2010). The specific signatures of actively transcribed chromatin, as observed in other systems, correlate with the genomic milieu, gene size, and regulatory functions. Exons of transcribed genes define a state particularly enriched in the elongation mark H3K36me3, whereas intronic regions cluster in a state with high levels of H3K4me1, H3K18ac, and H3K27ac, the typical pattern of mammalian enhancers (Kharchenko et al. 2010). An additional state distinguished by enrichment in H3K36me1 forms patches with other transcription-associated signatures within the gene bodies of long expressed genes. When focusing on genes that present sex-influenced or sex-specific expression in *Drosophila miranda*, no enrichment in marks associated with transcriptional activity is observed in either sex, independently of the gene expression levels (Brown and Bachtrog 2014). This suggests that gender-regulated genes may have a stricter epigenetic control that avoids leakiness and assures the transcriptional activity is fine-tuned in the specific gender. Moreover, one of the outputs of the comparative analysis of chromatin landscape between sexes and species of *Drosophila* is the definition of a chromatin signature typical of dosage compensation in the male X chromosome, characterised by high levels of H4K16ac only. As expected, genes with similar expression levels across species show high enrichments in active histone marks whereas non-transcribed genes have a silent chromatin environment, with no enrichment in any particular modification.

C. elegans also shares common features with other model systems in the organisation of chromatin patterns across transcribed genes, as the observed enrichments in the methylation states of H3K36 and H3K79 (Liu et al. 2011).

In a recent analysis of metazoan chromatin organisation, it was reported that H3K9me3 is present at the gene bodies of both transcribed and silent genes in heterochromatin domains in human cells (Ho et al. 2014). Similar observations had been

previously published in *Drosophila*, where the most typical heterochromatin marks are H3K9me2 and H3K9me3 (Riddle et al. 2011). Recent literature suggests a relationship between H3K9me3 and splicing: high levels of this histone modification are found in the alternative exons of genes, with concomitant accumulation of a member of the Heterochromatin Protein 1 (HP1) family at those regions—HP1 γ (Saint-Andre et al. 2011). HP1 recruits hnRNPs in *Drosophila* (Piacentini et al. 2009), and HP1 β was also reported to interact with ASF/SF2 in human cells (Loomis et al. 2009). H3K9me3 could thus be a regulator of splicing, promoting the recruitment of splicing factors through the interaction with particular HP1 members. HP1 could also facilitate the inclusion of the alternative exons through a decrease in the elongation rate of RNA PolII at the H3K9me3-marked regions (Saint-Andre et al. 2011).

Still, the presence of H3K9me3 in actively transcribed gene bodies is not a general rule for all metazoans. In *C. elegans*, transcribed genes in heterochromatic regions display lower H3K9me3 enrichments over gene bodies relative to silent genes. A similar situation concerns H4K20me1, which is enriched on both transcribed and silent gene bodies in human cells, but only present in expressed genes in flies and worms. Clearly, these epigenetic marks characteristic of repressed regions are not directly responsible of silencing mechanisms. Indeed, in *Drosophila*, many genes covered by HP1 and H3K9me3 are still transcriptionally active. One possible hypothesis is that the silencing of H3K9me3-containing genes is caused rather by DNA methylation. This does not seem to be the case in *Arabidopsis*: a subset of long genes residing in euchromatic arms contains high levels of H3K9me2, a mark typically present in heterochromatin; although the DNA encoding these genes is densely methylated and the chromatin is less accessible, they are still actively transcribed (Shu et al. 2012). It would be interesting to address whether these genes are enriched in splicing variants and compare their RNA PolII elongation rates and their nucleosome organisation vs. H3K9me2-depleted gene bodies.

Euchromatin in plants is mainly characterised by the combination of marks typical of active chromatin such as H3K4me2/3, H3K36me3, H3K9ac/me3, H3K56ac, 5mC, and H2Bub, and by the presence of the histone variant H3.3 (Roudier et al. 2011; Sequeira-Mendes et al. 2014; Stroud et al. 2012; Wang et al. 2014; Wollmann et al. 2012). Nevertheless, specific patterns are associated with particular regions within genes, forming motifs that globally describe transcription units (Sequeira-Mendes et al. 2014). Regarding the gene bodies, they are typically characterised by chromatin highly enriched in H3K4me2/3, H3K36me3, H2Bub, and in the histone variant H2A.Z around the 5'UTR of genes, followed by a highly similar signature but already containing enrichments in H3K4me1 and near background levels of H2A.Z (Fig. 2). This observation is in agreement with the described

gradients of histone methylation observed in gene bodies (Henikoff and Shilatifard 2011). A combination where H3K4me1, H2Bub, and H3K36me3 are the prominent marks describes coding sequences and intronic regions of *Arabidopsis* genes. This landscape is particularly present in longer than average transcription units. Medium-sized genes tend to contain around their TTS chromatin domains depleted of most activating marks with the exception of H3K4me1 and the variant H2A.Z. Since this chromatin state is also found in intergenic domains, it remains to be answered whether these regions are enriched in enhancer sequences or whether it is merely a reflection of the compact type of genome.

Repressed genes

The transcriptional status of a gene reflects in the combinations of marks observed in the residing nucleosomes. Unlike the methylation of H3K4 and the global hyperacetylation seen in actively transcribed domains, the promoters and 5' coding regions of repressed genes in yeast are massively hypoacetylated (Liu et al. 2005). This hypoacetylation, a global feature of transcriptionally repressed chromatin in all model organisms analysed, is also detected at classical heterochromatic domains as the subtelomeric regions and the silent mating type loci.

In *Drosophila*, large domains of relatively gene-poor chromatin constitute a transcriptionally silent state characterised by the absence of enrichment in any of the analysed histone modifications (Brown and Bachtrög 2014; Filion et al. 2010; Kharchenko et al. 2010) and by the association with H1 and nuclear lamina proteins (Filion et al. 2010). This chromatin is actively capable of silencing transgenes and is intrinsically different from the previously described heterochromatin, as it is depleted in PcG proteins or Heterochromatin Protein 1 (HP1). The switch to active chromatin state in the genes under the regulation of this silent chromatin state tends to be tissue-specific (Brown and Bachtrög 2014; Filion et al. 2010). It is worth noting that the human genome also contains silenced chromatin states that are globally devoid of histone modifications and show moderate enrichments in the nuclear lamina. However, the major association with nuclear lamina is observed for heterochromatin and repetitive states in human cells (see “Heterochromatin” section).

In worms, silent genes tend to be associated with chromatin clusters enriched in either H3K9me3 or H3K27me3 (Liu et al. 2011). Notably, there is a chromatin signature specific of the X chromosome, enriched in H3K27me1 and H4K20me1, that coincides with high levels of X-linked gene repression proteins, suggesting a role of these marks in dosage compensation mechanisms in this organism.

In plants, a cluster of chromatin associated with weakly transcribed genes resembles the background chromatin in flies by the absence of prevalent modifications (Roudier et al.

2011; Wang et al. 2014). However, repressed genes also fall in other chromatin signatures typical of heterochromatin or PcG-regulated (Fig. 2; see “Polycomb group-regulated chromatin” and “Heterochromatin” sections).

Polycomb group-regulated chromatin

In multicellular organisms, targeted gene repression through PcG group proteins sorts out specific chromatin clusters highly enriched in H3K27me₃, either in the presence or absence of typical marks of active chromatin, such as H3K4me_{2/3}. These domains are usually correlated with cell type-specific silencing of developmentally regulated genes.

In *Drosophila*, around 15 % of the chromatin is bound by PcG proteins, and is highly enriched in H3K27me₃ (Filion et al. 2010; Kharchenko et al. 2010). Within the domains of this cluster, subsets of combinatorial patterns around TSS can be observed, from fully repressed to paused or expressed genes (Kharchenko et al. 2010). Interestingly, *D. miranda* female flies have a higher portion of euchromatin covered by PcG marks than males (Brown and Bachtrög 2014). Whether this is true for other *Drosophila* species and/or it is related to repressed genes in the X and neo-X chromosomes is currently unknown.

In human cells, the genes falling in the PcG-repressed cluster are in general developmentally regulated and the chromatin in these domains is hypo-acetylated (Ernst and Kellis 2010; Ernst et al. 2011). Notably, the classical PcG state is depleted in embryonic stem cells whereas the bivalent states are more abundant, in agreement with the pluripotent potential of these cells.

In *Arabidopsis*, H3K27me₂ and H3K27me₃ cover most repressed or weakly expressed genes (Roudier et al. 2011). This prevalent combination establishes a chromatin state that covers around 20 % of the genome. H3K27me₃ domains in plants are much smaller than their counterparts in animals, many coinciding with single transcription units. This typical PcG-repressed euchromatin is characterised by high frequencies of H3K27me₃ and depletion of activating and heterochromatin marks, in an H3.1-rich nucleosome context that covers genes and intergenic regions (Fig. 2). Additionally, other H3K27me₃-containing states with distinct chromatin patterns exist (Sequeira-Mendes et al. 2014). A set of domains mainly localising upstream of the TSS and at the proximal regions of promoters is highly enriched in both H3K27me₃ repressive and H3K4me₃ activating marks, representing a bivalent chromatin state that harbours both modifications in the same chromatin fibre. An additional cluster characterised by H3K27me₃ and with moderate enrichments in activating histone marks is typically found at intergenic regions or at the distal parts of the promoters in plants (Fig. 2). Different PcG-containing states characterised by the presence or absence of activating marks were also described for human, *Drosophila*, and *C. elegans* in

a comprehensive study of metazoan chromatin organisation (Ho et al. 2014). Distinctively, *C. elegans* chromatin comprises states where heterochromatin (H3K9me₃) and PcG (H3K27me₃) marks strongly associate, in what appears to be a particular configuration of silent chromatin domains.

Intergenic domains

Many studies on chromatin organisation focus on the combinations of marks that predominate in transcriptional units. Still, some information can be gathered on the most frequent chromatin patterns at intergenic regions.

The so-called active intergenic states in human cells were associated globally with higher levels of H3K4me₁, the variant H2A.Z, several acetylation marks and/or the presence of CTCF, representing candidate enhancer or insulator regions ((Ernst and Kellis 2010); see “Enhancers” section). However, there is no prevalent combination of marks that globally describes intergenic regions, and they are probably being sorted into more than one state.

Drosophila, *Arabidopsis*, and *C. elegans* all present compact genomes with average short intergenic regions. Many of these intergenic regions were globally catalogued as background/null chromatin, characterised by the absence of histone modifications (Brown and Bachtrög 2014; Filion et al. 2010; Roudier et al. 2011; Wang et al. 2014). Still, higher resolution studies identified signatures that sort these short intergenic regions into active, PcG-regulated or classical heterochromatin states (Fig. 2) (Sequeira-Mendes et al. 2014), likely because the chromatin fingerprints of intergenic regions correlate with the milieu of neighbouring genomic elements.

Heterochromatin

Although most of the *Saccharomyces cerevisiae* genome is described as decondensed chromatin, the most compact domains that represent heterochromatin correspond to poorly transcribed regions characterised by large territories of hypo-acetylated histones (Liu et al. 2005). These repressed regions are the subtelomeric sequences, the silent mating type loci, and the rDNA repeats.

Heterochromatic regions correspond to gene-poor areas and repetitive elements in the human genome, where one of the most frequent modifications is H3K9me₃ (Ernst and Kellis 2010; Ernst et al. 2011). These territories, with a higher than average A+T content, account for more than 70 % of the genome and tend to be excluded from the nuclear core. Such domains tend to be much larger than active regions of the genome (e.g. promoter domains have an average size of 500 bp, whereas heterochromatin domains are ~10 kb long). In human cells, these large inactive territories show a strong association with the nuclear lamina in the nuclear periphery, coincident with their silent transcriptional status. This

observation is similar in the silenced (or null) chromatin in *Drosophila* (see “[Repressed genes](#)” section; (Filion et al. 2010)). The inactive domains in human cells can be separated into different states based on their chromatin signatures, namely heterochromatin and repetitive states.

In flies, the chromatin state globally defined as heterochromatin was sorted by the dense coverage by HP1 and associated proteins, as well as by a specific enrichment in H3K9me2, a histone mark typically present in condensed chromatin in *Drosophila* (Filion et al. 2010). This modification is generated by the histone methylase SU(VAR)3-9, a protein that is highly enriched in this state. SU(VAR)3-9 is known to interact with HDAC1, responsible for the overall deacetylation found at heterochromatin (Czermin et al. 2001). Notably, classic heterochromatin in *Drosophila* is not physically associated with the nuclear lamina. As observed in the human genome, this state is represented by broad domains and is localised mainly around centromeres. *C. elegans* also displays repetitive regions that associate with the nuclear membrane and are covered by H3K9 methylation (Liu et al. 2011). However, pericentric heterochromatin is absent in *C. elegans* and silent regions are spread along the chromosome arms, with centromeres disseminated along them (Ho et al. 2014). Notably, these regions are specifically enriched in both H3K9me3 and H3K27me3, a so-far exclusive feature of *C. elegans* heterochromatin organisation (see “[Polycomb group-regulated chromatin](#)” section). We can only speculate whether this unique silencing mechanism is a reflection of the topographical differences in the heterochromatin organisation in worms. Subsequent studies in flies distinguish between the pericentromeric/chromosome 4 heterochromatin, which is highly enriched in H3K9me2/3, and the heterochromatin-like regions. The latter are located within euchromatin and contain moderate levels of H3K9me2/3 marks (Kharchenko et al. 2010). Without apparent selective pressure, heterochromatin displays little conservation between *Drosophila* species and gender (Brown and Bachtrog 2014). Indeed, different sets of genes reside in heterochromatin in the two sexes and female flies show almost doubled content of constitutive heterochromatin relative to males. Females tend to acquire heterochromatin structure in euchromatic regions (as suggested by the higher silencing rates in positional effect variegation assays), and it is hypothesised that the absence of the Y chromosome acting as an attractor of heterochromatin-related proteins is the reason for this gender-dependent feature.

Plants heterochromatin defines a specific chromatin signature comprised by the depletion in active marks and by high levels of DNA methylation, H3K9me2, H3K27me1/2, and H4K20me1 (Fig. 2) (Bernatavichute et al. 2008; Fischer et al. 2006; Johnson et al. 2002; Roudier et al. 2011; Sequeira-Mendes et al. 2014; Wang et al. 2014; Zhou et al. 2010). This chromatin state forms large domains around centromeres, as observed in other multicellular organisms, and

also small domains scattered in euchromatin where it also silences transposable elements and other repetitive sequences. Heterochromatin can be further separated into two distinct types that are distinguished by their average G+C content (Fig. 2) (Sequeira-Mendes et al. 2014). GC-rich heterochromatin constitutes the least DNaseI-accessible portion of the genome, with large domains that concentrate at pericentromeric regions, either intergenic or enriched in transposable elements. On the other hand, AT-rich heterochromatin, less inaccessible than the former, contains moderate enrichments in repressive marks and smaller domain sizes. A similar classification of heterochromatin, independent of GC content and based on the enrichment levels of repressive marks, was recently described elsewhere (Wang et al. 2014). Notably, these states form linear motifs that lead us to speculate whether this AT-rich inactive state could facilitate the access of protein factors to heterochromatin.

From linear to 3D organisation of chromatin in the nucleus

Nuclear structure and organisation have been a subject of interest ever since the early observations of this organelle in red blood cells of salmon, by Antonie van Leeuwenhoek in 1682: “... I could not make out what parts these oval particles consisted, for it seemed to me that some of them enclosed in a small space a little round body or globule...”. But it was not until 1831 that the botanist Robert Brown described and named this round structure the cell nucleus, in a paper read to the Linnean Society and acknowledging also Franz Bauer’s drawings of this organelle. The relevance of this sub-cellular structure was highly increased with the finding that it contains the genetic material of eukaryotic cells. Packing the large DNA fibres and associated proteins within the eukaryotic nucleus implies a remarkable topological challenge. But one step ahead of the primary fitting task, nuclear architecture reveals a close relation between spatial organisation and genomic function. Basic chromosome topology and the discovery of differentially condensed regions (heterochromatin and euchromatin) in post-mitotic chromosomes were first described in mosses by Emil Heitz (Heitz 1928). Advances in cytogenetic techniques allowed the visual inspection of particular chromosome portions and their architecture, through fluorescence in situ hybridisation (FISH) procedures. Still, with microscopy techniques revealing to be necessary but insufficient to shed light into the rules governing nuclear topology, new genomic approaches were developed and are under continuous update and optimisation. We will succinctly go through these novel methods for the systematic inspection of nuclear organisation (for a thorough review and illustrations of the techniques, refer to de Wit and de Laat (2012) and Dekker et al. (2013)). We will follow with a global view of the interaction data and its

functional significance in different model systems. For a discussion on the basic differences, complementation, and controversy between microscopy-based and high-throughput genomic methodologies, the reader is referred to some excellent papers (Belmont 2014; Pombo and Dillon 2015; Williamson et al. 2014).

The first high-throughput methodology that allowed the global analysis of chromatin interactions and spatial organisation of chromosomes was chromosome conformation capture (3C) (Dekker et al. 2002). This technique (and all the derived methodologies) relies on the formaldehyde fixation of chromatin within intact nuclei or cells, which cross-links all the regions of chromosomes that were contacting via DNA or protein factors. The chromatin is then isolated and digested by an appropriate restriction enzyme, and the cross-linked fragments are subjected to intramolecular ligation, resulting in an “interaction library”. The frequency of contact between two given loci (one vs. one analysis) can then be determined by several PCR-based techniques using specific primers, measuring the abundance of the ligation product (Dekker et al. 2002; Hagège et al. 2007; Louwers et al. 2009). The 3C methodology sets the base for all the three-dimensional genomic techniques currently used. The circular chromosome conformation capture (4C) adds a second digestion and circularisation step (Zhao et al. 2006). Primers within the *locus* of interest, also called *viewpoint*, and close to the restriction target sequences will amplify by inverse PCR the pool of contacting sites to be analysed downstream. 4C is also an acronym for 3C-on chip, a technology where the library of interactions is subjected to microarray hybridisation (Simonis et al. 2006) or more recently to next-generation sequencing (NGS) (Splinter et al. 2011). This allows a blind screening of the physical interactors of each viewpoint, in a one vs. all survey. Another high-throughput 3C-based methodology is 3C-Carbon Copy (5C) (Dostie et al. 2006). This approach allows the simultaneous identification of the interaction profiles between various regions (many vs. many). Following the described generation of the 3C ligation library, a mix of primers containing universal tails at their 5'-ends, and each partially overlapping a different restriction site in the locus of interest, is hybridised and anneals with the DNA templates. If two regions were interacting in the chromatin, the primer pair will be located contiguously and can thus be ligated. These ligated primers are then amplified and analysed by microarrays or NGS. Although with lower resolution than other 3C-based methodologies, 5C presents the improvement of gathering the information on the frequencies of interaction for several pairs of sites. The final output is a network of interactions that allows the reconstruction of the three-dimensional conformation of larger genomic regions.

The first method that allowed the analysis of all the genomic interactions within the nucleus was Hi-C (all vs. all) (Lieberman-Aiden et al. 2009). During the generation of the

3C library, an end-repair reaction in the presence of biotin-labelled nucleotides precedes the blunt ligation of interacting segments. The purified DNA is then sheared and the biotin-containing ligation junctions purified with streptavidin for further pair-end NGS analysis. Paired reads assembling in two separate restriction fragments represent an interaction between these sites. The resulting matrix of interactions gives rise to a global representation of all the inter- and intra-chromosomal contacts in the genome and to the average three-dimensional conformation of the nucleus of the analysed system. The first Hi-C analysis had a resolution of ~1 Mb due to sequencing depth limitations, but subsequent studies have achieved greater resolution, varying from a few hundred base pairs in *Drosophila* to a few kilobases in mammals. Variations of this methodology with a 4C-based protocol were also described (Duan et al. 2010). Very recently, an *in situ* Hi-C library that included the step of proximal DNA ligation in intact nuclei was sequenced in great depth, allowing a remarkable increase in what the authors call “map resolution” of the Hi-C maps, of ~5 kb and down to ~1 kb (Rao et al. 2014). More targeted conformation capture techniques as ChIA-loop (Horike et al. 2005) and ChIA-PET (Fullwood et al. 2009) include a chromatin immunoprecipitation step that restricts the analysis of interactions exclusively to those pairs of sites bound by a protein of interest (in a many vs. many approach). Although presenting obvious limitations, these ChIP-dependent methods also show advantages with respect to 3C-based technologies, as they can detect specific chromatin loops in pools of cells independently on the average frequency of contacts in the whole population. The reader is referred to two comprehensive reviews that focus exclusively in chromosome conformation capture techniques and show the workflows of 3C-derived methods (de Wit and de Laat 2012; Dekker et al. 2013).

Focusing now on the latest chromosome conformation capture data, we will briefly discuss the state of the art in the field of nuclear three-dimensional organisation from the standpoint of epigenetic context.

The 3C technique was first reported in a pioneer study of the topological organisation of the budding yeast chromosome III (Dekker et al. 2002). A subsequent analysis of all the intra- and inter-chromosomal interactions in *S. cerevisiae* at ~1 kb resolution led to the proposal of an average 3D model of the whole genome (Duan et al. 2010). The snapshot of this highly compacted genome shows that chromosomes are gathered through the centromeres in one nuclear pole, with highly flexible chromosome arms. Intra-chromosomal interactions prevail over inter-chromosomal ones, although this tendency decreases as one moves towards the chromosome ends. The nucleolus at the opposite pole of the nucleus is specifically occupied by a section of chromosome XII containing the rDNA repeats. These observations are coherent with prior data based on high-resolution probabilistic gene maps built

through the automated analysis of microscopy images (Berger et al. 2008). Besides the centromeres, DNA replication origins that fire specifically in early S phase, DNA fragile sites, and tRNA genes are involved in high-frequency inter-chromosomal contacts. It should be noted that no structured territories were observed in yeast, although this is not the case for other model systems.

In *Drosophila*, while physical interactions are largely limited to the same chromosome arm, PcG-regulated genes specifically cluster in the same nuclear space. This organisation, directed by global chromosome architecture, cooperates in the preservation of the epigenetic silencing (Bantignies et al. 2011; Cheutin and Cavalli 2014; Tolhuis et al. 2011). A high-resolution chromosomal contact map confirmed the PcG domains and also known cytological observations, such as the clustering of centromeres and the heterochromatin-enriched chromosome 4 or the interactions between telomeres (Sexton et al. 2012). Additionally, it unravelled the existence of delimited territories that correlate with the underlying epigenetic states (Filion et al. 2010; Sexton et al. 2012). Actively transcribed genes tend to establish interactions with other active elements in the same arm. Unlike repressive PcG domains, which are formed mostly by long-range intra-chromosomal interactions between target genes in euchromatic chromosome arms, centromeric HP1 heterochromatin folds into chromosomal territories where inter-chromosomal contacts prevail. Additionally, a third class of repressed chromatin (null chromatin; see “Repressed genes” section) with no specific epigenetic signature forms independent territories in the nuclear periphery. Together, the repressive domains hierarchically cluster. In a chromosome arm, domains are organised into active and inactive higher-order territories with different folding properties. The active domains establish inter-chromosomal contacts with other active domains whereas the repressed territories tend to be limited to their chromosomal domains. This is consistent with previous studies in human cells describing compartmentalisation of interactions in the nucleus, with physical segregation of active and heterochromatin domains into the so-called A and B territories, respectively (Lieberman-Aiden et al. 2009). This segregation was also observed in mouse nuclei (Simonis et al. 2006; Zhang et al. 2012).

The topologically associating domains (TADs) show a comparable modular organisation in metazoans and are frequently delimited by insulator elements, such as CP190 and Chromator in *Drosophila* or CTCF in mammals (Dixon et al. 2012; Nora et al. 2012; Sexton et al. 2012). TADs range from ~100 kb in flies to ~1 Mb in mammals, a feature that probably reflects the differences in genes and genome sizes. ChIA-PET sequencing of CTCF- or cohesin-interacting regions in mammals reiterated the structural role of these proteins in the establishment or maintenance of specific chromatin domains and in the regulation of chromosome architecture in

development (DeMare et al. 2013; Downen et al. 2014; Handoko et al. 2011). CTCF is known to be involved in the formation of chromatin loops and mutation of CTCF-binding sites or conditional deletion of the protein destabilises long-range interactions, demonstrating a direct role in chromosome folding (Splinter et al. 2006). Moreover, Hi-C analysis of cells depleted of CTCF or cohesin showed a differential role of these proteins in the chromatin architecture (Zuin et al. 2014). While depletion of CTCF interferes differentially in the balance of both intra- and interdomain interactions, cohesin appears to play a role only in the maintenance of local loops. Microscopical analysis revealed a global compaction of chromatin upon knock down of CTCF and the cohesin subunit Rad21, contrary to what would be expected (Tark-Dame et al. 2014). It is possible that the decrease in short-range contacts increases the probability of long-range loop formation (as observed by (Zuin et al. 2014)), and this in turn promotes chromatin compaction.

In mammals, domain boundaries are also enriched in housekeeping genes, tRNAs, and short interspersed element (SINE) retrotransposons, suggesting a role of these elements in nuclear architecture (Dixon et al. 2012). Moreover, these borders sign a limit for chromatin spreading as they distinctly segregate typical heterochromatin marks. Active gene-rich domains are clearly defined in the nucleus and mostly organised into enhancer-promoter loops that regulate gene expression (Li et al. 2012; Sanyal et al. 2012; Smemo et al. 2014; Tolhuis et al. 2011). In plants, high-order chromatin structures that regulate the transcriptional status of a gene have also been described (Crevillén et al. 2012). Repressed chromatin, which tends to be gene-poor and late-replicating, clusters into domains that associate in the nuclear periphery, as observed in *Drosophila* with the null (or silent) chromatin (see “Repressed genes” section). Accordingly, the interaction map between the human genome and nuclear lamina components unveiled the lamina-associated domains (LADs). These discrete domains, described as repressed chromatin regions, form structural units bordered by the insulator CTCF (Guelen et al. 2008). These observations relied on a 3C-independent method (DamID). Early- and late-replicating domains are also clearly spatially separated in the nucleus, and specific switches in the replication timing programme during differentiation are intimately associated with changes in the spatial organisation of chromatin (Pope and Gilbert 2013; Ryba et al. 2010).

Increasing the resolution for the generation of an improved 3D map of the human genome showed that the nuclear organisation goes beyond active and repressed chromatin territories, with sub-compartments with a median length of 300 kb that clearly correlate with distinct epigenetic signatures (Rao et al. 2014). Notably, this study reveals that mammals also exhibit chromosome domains that are PcG-related, a characteristic that was until now unremarkable perhaps due to resolution limitations.

It is worth noting that all 3C-based studies have reconstructed chromosome conformation based on the average contacts from a population of nuclei. Single-cell Hi-C showed that chromosomes globally maintain domain organisation in the cell population, but cell-to-cell variability exists in the chromosome structure (Nagano et al. 2013). Single-cell Hi-C provides a *bona fide* individual interaction map that could help reconcile non-compatible views obtained with FISH vs. 3C-based techniques.

In *Arabidopsis*, overall rules of chromosome interactions follow the observations described in other higher eukaryotes. Euchromatin and heterochromatin are found physically separated in the nucleus, and the frequency of intra-chromosomal interactions decreases towards the chromosome ends, as inter-chromosomal contacts increase in these distal segments (Grob et al. 2013). Pericentromeric regions cluster within the nucleus and telomeres show high interaction frequencies (except the ones from the short arms of chromosomes 2 and 4 that contain the rDNA repeats, which interact at lower frequencies). Hi-C analyses in *Arabidopsis* clearly support these previous observations, point out distinctive features in plants but also further sustain general rules governing nuclear architecture in multicellular organisms (Feng et al. 2014; Grob et al. 2014; Wang et al. 2014). The *Arabidopsis* chromatin is organised in compartmentalised territories, although in a smaller scale that does not allow the designation of TADs as defined in animal nuclear organisation. The nonexistence of these topological domains in *Arabidopsis* could be due to the absence of the insulator protein CTCF in the plant kingdom, which is present at the boundaries of TADs in animals. Additionally, it could be a matter of linear organisation of the chromatin, as the average gene unit size and the compact nature of its genome most certainly impose topographical constraints that rule chromosome folding. Nonetheless, although the local chromatin packing differs from animal patterns, insulator-like, TAD-boundary-like, and TAD-interior-like regions were recently described in *Arabidopsis*, each enriched in specific epigenetic states (Wang et al. 2014). As in human cells and in *Drosophila*, boundary- and insulator-like regions are enriched in active genes.

Apart from the physical separation between heterochromatin in pericentromeric regions and euchromatin in chromosome arms, two main territories can be described within the latter: loose structural domains that represent active territories with low interaction frequencies within themselves but that reach out for distal contacts both in *cis* and in *trans* and compacted structural domains highly interactive among each other but isolated from the rest of the genome (Feng et al. 2014; Grob et al. 2014). These territories correlate with epigenetic signatures typical of active and repressed euchromatin, respectively. Moreover, the contact maps disclosed intra- and inter-chromosomal long-range interactions between small heterochromatin islands scattered throughout the euchromatic

arms (Feng et al. 2014; Grob et al. 2014; Wang et al. 2014). These islands were differently characterised either by enrichment in transposable elements (TEs), smRNAs, and H3K27me1 (Grob et al. 2014) or by high H3K9me2 and TE content (Feng et al. 2014). In one of these studies, the interacting structure representing highly specific and exclusive contacts between heterochromatic regions within chromosome arms was identified as the KNOT (Grob et al. 2014). The KNOT is thought to play a role in TE defence, in a mechanism presumably shared with *Drosophila* where piRNA clusters involved in TE regulation also interact at the three-dimensional level (Grob et al. 2014; Sexton et al. 2012). It should be noted though that not all patches of heterochromatin embedded in chromosome arms show these strong interaction profiles (Feng et al. 2014). Globally, heterochromatin state seems to prevail in the long-range interactions that outline the overall folding of chromosomes, whereas PcG-regulated regions seem to be involved in short-range interactions (Wang et al. 2014). Additionally, strong interaction frequencies were observed within a small subset of regions heavily marked by H3K27me3 (Feng et al. 2014; Wang et al. 2014). Whether these H3K27me3-containing interactive regions are related to the PcG bodies observed in animals and could represent 3D domains that facilitate gene silencing and regulate genome function during proliferation and differentiation is still an unanswered question (Cheutin and Cavalli 2014; Wang et al. 2014). In fact, the precise influence of the epigenetic states in the three-dimensional folding of chromatin fibres in the nucleus is now starting to be unravelled thanks to the availability of mutants in protein factors involved in epigenetic regulation, where Hi-C experiments can be performed. Consistent with a possible dependence of the chromatin environment in chromosome packing, mutants lacking H3K27me3 lose the interaction domains constituted by PcG target genes observed in wild-type nuclei (Feng et al. 2014). Whether this is due to a direct role of the PcG silencing mark or to the subsequent loss of proteins that might mediate these interactions it is still unclear. Likewise, mutants of proteins involved in the condensation of heterochromatin (*atmorc6*) or global DNA methylation (*met1* and *ddm1*) show decreased interaction frequencies within pericentromeric regions and increased contacts between these regions and euchromatic arms, due to the decondensation of heterochromatin and consequent derepression (Moissiard et al. 2012). Finally, Hi-C experiments performed in mutants involved in nuclei size control (*crwn1* and *crwn4*) suggest that the epigenetic landscape plays a more important role than the nuclear morphology in the maintenance of domain structure (Grob et al. 2014).

The increasing amount of structural and molecular information, and recently the genomic data on high-order organisation of the genome, clearly demonstrates the functional implications. The significance of nuclear territories clearly goes beyond the organisation of the genome into smaller subunits

because they have functional consequences in terms of accessibility to cellular factors, gene expression, and functional coordination. The 3D organisation of the nucleus apparently optimises its compartmentalisation from a structural and functional point of view. Moreover, it is conceivable that future research will identify crucial features that are regulated by or regulate specific cell differentiation states or the signalling of developmental cues. If so, the possibility of identifying disease conditions or response to the environment that ultimately depend on changes in genome architecture may not be too far in time.

Acknowledgments We deeply thank all C.G. laboratory members for the insightful discussions and B. Desvoves for the help with Fig. 2. Research at C.G. lab is supported by Ministerio de Economía y Competitividad (grant BFU2012–34821) and an institutional grant of Fundación Ramón Areces to the Centro de Biología Molecular Severo Ochoa.

Compliance with ethical standards This is a review article that does not provide original data and is not a research involving human and/or animals. Informed consent is not applicable.

Conflict of interest The authors declare that they have no competing interests.

References

- Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. *Cell Res* 21:381–395. doi:10.1038/cr.2011.22
- Bantignies F et al (2011) Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell* 144:214–226. doi:10.1016/j.cell.2010.12.026
- Barski A et al (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129:823–837. doi:10.1016/j.cell.2007.05.009
- Belmont AS (2014) Large-scale chromatin organization: the good, the surprising, and the still perplexing. *Curr Opin Cell Biol* 26:69–78. doi:10.1016/j.ccb.2013.10.002
- Berger AB et al (2008) High-resolution statistical mapping reveals gene territories in live yeast. *Nat Methods* 5:1031–1037. doi:10.1038/nmeth.1266
- Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS One* 3:e3156. doi:10.1371/journal.pone.0003156
- Biesinger J, Wang Y, Xie X (2013) Discovering and mapping chromatin states using a tree hidden Markov model. *BMC Bioinformatics* 14(Suppl 5):S4. doi:10.1186/1471-2105-14-S5-S4
- Brown EJ, Bachtrog D (2014) The chromatin landscape of *Drosophila*: comparisons between species, sexes, and chromosomes. *Genome Res* 24:1125–1137. doi:10.1101/gr.172155.114
- Calo E, Wysocka J (2013) Modification of enhancer chromatin: what, how, and why? *Mol Cell* 49:825–837. doi:10.1016/j.molcel.2013.01.038
- Cheutin T, Cavalli G (2014) Polycomb silencing: from linear chromatin domains to 3D chromosome folding. *Curr Opin Genetics Dev* 25:30–37. doi:10.1016/j.gde.2013.11.016
- Christophorou MA et al (2014) Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* 507:104–108. doi:10.1038/nature12942
- Crevillén P, Sonmez C, Wu Z, Dean C (2012) A gene loop containing the floral repressor FLC is disrupted in the early phase of vernalization. *EMBO J* 32:140–148. doi:10.1038/emboj.2012.324
- Czermin B, Schotta G, Hülsmann BB, Brehm A, Becker PB, Reuter G, Imhof A (2001) Physical and functional association of SU(VAR)3-9 and HDAC1 in *Drosophila*. *EMBO Rep* 2:915–919. doi:10.1093/embo-reports/kve210
- de Wit E, de Laat W (2012) A decade of 3C technologies: insights into nuclear organization. *Genes Dev* 26:11–24. doi:10.1101/gad.179804.111
- Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. *Science* 295:1306–1311. doi:10.1126/science.1067799
- Dekker J, Marti-Renom MA, Mirny LA (2013) Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nature Rev Genet* 14:390–403. doi:10.1038/nrg3454
- DeMare LE, Leng J, Cotney J, Reilly SK, Yin J, Sarro R, Noonan JP (2013) The genomic landscape of cohesin-associated chromatin interactions. *Genome Res* 23:1224–1234. doi:10.1101/gr.156570.113
- Dixon JR et al (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–380. doi:10.1038/nature11082
- Dostie J et al (2006) Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* 16:1299–1309. doi:10.1101/gr.5571506
- Dowen JM et al (2014) Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159:374–387. doi:10.1016/j.cell.2014.09.030
- Duan Z et al (2010) A three-dimensional model of the yeast genome. *Nature* 465:363–367. doi:10.1038/nature08973
- Ernst J, Kellis M (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* 28:817–825. doi:10.1038/nbt.1662
- Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 9:215–216. doi:10.1038/nmeth.1906
- Ernst J, Kellis M (2013) Interplay between chromatin state, regulator binding, and regulatory motifs in six human cell types. *Genome Res* 23:1142–1154. doi:10.1101/gr.144840.112
- Ernst J et al (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473:43–49. doi:10.1038/nature09906
- Feng S, Jacobsen SE (2011) Epigenetic modifications in plants: an evolutionary perspective. *Curr Opin Plant Biol* 14:179–186. doi:10.1016/j.pbi.2010.12.002
- Feng S, Cokus SJ, Schubert V, Zhai J, Pellegrini M, Jacobsen SE (2014) Genome-wide Hi-C analyses in wild-type and mutants reveal high-resolution chromatin interactions in *Arabidopsis*. *Mol Cell* 55:694–707. doi:10.1016/j.molcel.2014.07.008
- Filion GJ et al (2010) Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* 143:212–224. doi:10.1016/j.cell.2010.09.009
- Fischer A, Hofmann I, Naumann K, Reuter G (2006) Heterochromatin proteins and the control of heterochromatic gene silencing in *Arabidopsis*. *J Plant Physiol* 163:358–368. doi:10.1016/j.jplph.2005.10.015
- Fullwood MJ et al (2009) An oestrogen-receptor- α -bound human chromatin interactome. *Nature* 462:58–64. doi:10.1038/nature08497
- Grob S, Schmid MW, Luedtke NW, Wicker T, Grossniklaus U (2013) Characterization of chromosomal architecture in *Arabidopsis* by chromosome conformation capture. *Genome Biol* 14:R129. doi:10.1186/gb-2013-14-11-r129

- Grob S, Schmid MW, Grossniklaus U (2014) Hi-C analysis in Arabidopsis identifies the KNOT, a structure with similarities to the flamenco locus of Drosophila. *Mol Cell* 55:678–693. doi:10.1016/j.molcel.2014.07.009
- Gu SG, Fire A (2010) Partitioning the *C. elegans* genome by nucleosome modification, occupancy, and positioning. *Chromosoma* 119:73–87. doi:10.1007/s00412-009-0235-3
- Guelen L et al (2008) Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* 453:948–951. doi:10.1038/nature06947
- Gurard-Levin ZA, Almouzni G (2014) Histone modifications and a choice of variant: a language that helps the genome express itself. *F1000Prime Rep* 6:76 doi:10.12703/P6-76
- Hagège H et al (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* 2:1722–1733. doi:10.1038/nprot.2007.243
- Handoko L et al (2011) CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* 43:630–638. doi:10.1038/ng.857
- Happel N, Doecke D (2009) Histone H1 and its isoforms: contribution to chromatin structure and function. *Gene* 431:1–12
- Heitz E (1928) Das Heterochromatin der Moose. *Jahrb Wiss Botanik* 69:762–818
- Henikoff S, Shilatifard A (2011) Histone modification: cause or cog? *Trends Genet* 27:389–396. doi:10.1016/j.tig.2011.06.006
- Ho JWK et al (2014) Comparative analysis of metazoan chromatin organization. *Nature* 512:449–452. doi:10.1038/nature13415
- Hon G, Ren B, Wang W (2008) ChromaSig: a probabilistic approach to finding common chromatin signatures in the human genome. *PLoS Computat Biol* 4:e1000201. doi:10.1371/journal.pcbi.1000201
- Horiike S-i, Cai S, Miyano M, Cheng J-F, Kohwi-Shigematsu T (2005) Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nature Genet* 37:31–40. doi:10.1038/ng1491
- Johnson L, Cao X, Jacobsen S (2002) Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr Biol* 12:1360–1367
- Kharchenko PV et al (2010) Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* 471:480–485. doi:10.1038/nature09725
- Kouzarides T (2007) Chromatin modifications and their function. *Cell* 128:693–705. doi:10.1016/j.cell.2007.02.005
- Larson JL, Yuan G-C (2010) Epigenetic domains found in mouse embryonic stem cells via a hidden Markov model. *BMC Bioinformatics* 11:557. doi:10.1186/1471-2105-11-557
- Li G et al (2012) Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148:84–98. doi:10.1016/j.cell.2011.12.014
- Lieberman-Aiden E et al (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289–293. doi:10.1126/science.1181369
- Liu CL, Kaplan T, Kim M, Buratowski S, Schreiber SL (2005) PLOS Biology: single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol* 3:e328. doi:10.1371/journal.pbio.0030328
- Liu T et al (2011) Broad chromosomal domains of histone modification patterns in *C. elegans*. *Genome Res* 21:227–236
- Loomis RJ et al (2009) Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation. *Mol Cell* 33:450–461. doi:10.1016/j.molcel.2009.02.003
- Louwers M, Splinter E, van Driel R, de Laat W, Stam M (2009) Studying physical chromatin interactions in plants using chromosome conformation capture (3C). *Nat Protoc* 4:1216–1229. doi:10.1038/nprot.2009.113
- Maze I, Noh K-M, Soshnev AA, Allis CD (2014) Every amino acid matters: essential contributions of histone variants to mammalian development and disease. *Nature Rev Genet* 15:259–271. doi:10.1038/nrg3673
- McDaniell R et al (2010) Heritable individual-specific and allele-specific chromatin signatures in humans. *Science* 328:235–239. doi:10.1126/science.1184655
- Mikkelsen TS et al (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448:553–560. doi:10.1038/nature06008
- Moissiard G et al (2012) MORC family ATPases required for heterochromatin condensation and gene silencing. *Science* 336:1448–1451. doi:10.1126/science.1221472
- Nagano T et al (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502:59–64. doi:10.1038/nature12593
- Nora EP et al (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485:381–385. doi:10.1038/nature11049
- Ott RW, Chua NH (1990) Enhancer sequences from Arabidopsis thaliana obtained by library transformation of *Nicotiana tabacum*. *Mol Gen Genet* 223:169–179
- 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. doi:10.1371/journal.pgen.1000670
- Pombo A, Dillon N (2015) Three-dimensional genome architecture: players and mechanisms. *Nat Rev Mol Cell Biol* 16:245–257. doi:10.1038/nrm3965
- Pope BD, Gilbert DM (2013) The replication domain model: regulating replicon firing in the context of large-scale chromosome architecture. *J Mol Biol* 425:4690–4695. doi:10.1016/j.jmb.2013.04.014
- Rao SSP et al (2014) A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159:1665–1680. doi:10.1016/j.cell.2014.11.021
- Riddle NC et al (2011) Plasticity in patterns of histone modifications and chromosomal proteins in *Drosophila* heterochromatin. *Genome Res* 21:147–163. doi:10.1101/gr.110098.110
- RoadmapEpigenomicsConsortium et al (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518:317–330. doi:10.1038/nature14248
- Roudier F et al (2011) Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J* 30:1928–1938. doi:10.1038/emboj.2011.103
- Ryba T et al (2010) Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res* 20:761–770. doi:10.1101/gr.099655.109
- Saint-Andre V, Batsche E, Rachez C, Muchardt C (2011) Histone H3 lysine 9 trimethylation and HP1 gamma favor inclusion of alternative exons. *Nat Struct Mol Biol* 18:337–344. doi:10.1038/nsmb.1995
- Sanyal A, Lajoie BR, Jain G, Dekker J (2012) The long-range interaction landscape of gene promoters. *Nature* 489:109–113. doi:10.1038/nature11279
- Sequeira-Mendes J et al (2014) The functional topography of the Arabidopsis genome is organized in a reduced number of linear motifs of chromatin states. *Plant Cell* 26:2351–2366. doi:10.1105/tpc.114.124578
- Sexton T et al (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148:458–472. doi:10.1016/j.cell.2012.01.010
- Shu H, Wildhaber T, Siretskiy A, Gruissem W, Hennig L (2012) Distinct modes of DNA accessibility in plant chromatin. *Nat Commun* 3:1281. doi:10.1038/ncomms2259
- Shu H et al (2014) Arabidopsis replacement histone variant H3.3 occupies promoters of regulated genes. *Genome Biol* 15:R62. doi:10.1186/gb-2014-15-4-r62

- Simonis M et al (2006) Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genet* 38:1348–1354. doi:[10.1038/ng1896](https://doi.org/10.1038/ng1896)
- Smemo S et al (2014) Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* 507:371–375. doi:[10.1038/nature13138](https://doi.org/10.1038/nature13138)
- Spitz F, Furlong EEM (2012) Transcription factors: from enhancer binding to developmental control. *Nature Rev Genet* 13:613–626. doi:[10.1038/nrg3207](https://doi.org/10.1038/nrg3207)
- Splinter E et al (2006) CTCF mediates long-range chromatin looping and local histone modification in the beta-globin locus. *Genes Dev* 20:2349–2354. doi:[10.1101/gad.399506](https://doi.org/10.1101/gad.399506)
- Splinter E et al (2011) The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. *Genes Dev* 25:1371–1383. doi:[10.1101/gad.633311](https://doi.org/10.1101/gad.633311)
- Stroud H, Otero S, Desvoyes B, Ramirez-Parra E, Jacobsen SE, Gutierrez C (2012) Genome-wide analysis of histone H3.1 and H3.3 variants in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A* 109:5370–5375. doi:[10.1073/pnas.1203145109](https://doi.org/10.1073/pnas.1203145109)
- Sundaresan V et al (1995) Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev* 9:1797–1810
- Tark-Dame M, Jerabek H, Manders EM, Heermann DW, van Driel R (2014) Depletion of the chromatin looping proteins CTCF and cohesin causes chromatin compaction: insight into chromatin folding by polymer modelling. *PLoS Comput Biol* 10:e1003877. doi:[10.1371/journal.pcbi.1003877](https://doi.org/10.1371/journal.pcbi.1003877)
- Tolhuis B et al (2011) Interactions among Polycomb domains are guided by chromosome architecture. *PLoS Genet* 7:e1001343. doi:[10.1371/journal.pgen.1001343](https://doi.org/10.1371/journal.pgen.1001343)
- Visel A et al (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457:854–858. doi:[10.1038/nature07730](https://doi.org/10.1038/nature07730)
- Wang C et al (2014) Genome-wide analysis of local chromatin packing in *Arabidopsis thaliana*. *Genome Res*. doi:[10.1101/gr.170332.113](https://doi.org/10.1101/gr.170332.113)
- Williamson I et al (2014) Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev* 28:2778–2791. doi:[10.1101/gad.251694.114](https://doi.org/10.1101/gad.251694.114)
- Wollmann H, Holec S, Alden K, Clarke ND, Jacques P-É, Berger F (2012) Dynamic deposition of histone variant H3.3 accompanies developmental remodeling of the *Arabidopsis* transcriptome. *PLoS Genet* 8:e1002658. doi:[10.1371/journal.pgen.1002658](https://doi.org/10.1371/journal.pgen.1002658)
- Zhang Y et al (2012) Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* 148:908–921. doi:[10.1016/j.cell.2012.02.002](https://doi.org/10.1016/j.cell.2012.02.002)
- Zhao Z et al (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genet* 38:1341–1347. doi:[10.1038/ng1891](https://doi.org/10.1038/ng1891)
- Zhou J, Wang X, He K, Charron J-BF, Elling AA, Deng XW (2010) Genome-wide profiling of histone H3 lysine 9 acetylation and dimethylation in *Arabidopsis* reveals correlation between multiple histone marks and gene expression. *Plant Mol Biol* 72:585–595. doi:[10.1007/s11103-009-9594-7](https://doi.org/10.1007/s11103-009-9594-7)
- Zuin J et al (2014) Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proc Natl Acad Sci U S A* 111:996–1001. doi:[10.1073/pnas.1317788111](https://doi.org/10.1073/pnas.1317788111)