Packaging the fly genome: domains and dynamics

Rob White

Advance Access publication date 3 September 2012

Abstract

Two independent genomic approaches have recently converged to provide insight into the domain organization of the Drosophila genome. Genome-wide mapping of chromosomal proteins and histone modifications has generated detailed maps of the Drosophila chromatin landscape and has led to the identification of a number of different chromatin states and their distribution in domains across the genome. A remarkably similar domain organization is derived from whole genome mapping of chromatin interactions that reveals the segmentation of the genome into structural domains. This review focuses on our current understanding of this domain architecture which provides a foundation for our understanding of the link between chromatin organization and the dynamic activity of the genome.

Keywords: Drosophila; chromatin; genome organization; insulator; Polycomb; polytene chromosome

Within the nucleus, DNA is packaged by histones into nucleosomes and, in association with a large number of non-histone proteins, is dynamically assembled into a variety of different chromatin states. The nature of these chromatin states, their composition, how they are organized along the genome and how they relate to the functions of the genome are all key current questions. Over recent years a detailed genome-wide map of the Drosophila chromatin landscape has been emerging and is beginning to reveal insights into the functional packaging of the fly genome.

Knowledge of the chromatin landscape is derived from a variety of approaches to map features onto the genome. Chromatin immunoprecipitation coupled with either genomic microarray analysis (ChIP-array) or sequencing (ChIP-Seq) has been the predominant approach to map the sites where chromatin-associated proteins are bound and to identify the genomic locations of specific modified histones which play key roles in forming a variety of chromatin states [1–3]. The DamID approach has also been used to map the binding sites of a large number of chromatin-associated proteins [4]. Other techniques have been developed to probe chromatin structure rather than composition and these include genome-wide DNase1 sensitivity assays [5], FAIRE [6] and profiling chromatin solubility [7]. Some of the major datasets are detailed in Table 1. Despite the success of mapping a large number (>100) of DNA-binding proteins and other chromatin-associated factors, it should be borne in mind that this is only a small proportion of the total; for example, for direct DNA-binding proteins alone there are >700 in the Drosophila genome [8]. However, integrating available mapping data with the wealth of functional genetic and genomic information available in Drosophila has provided a powerful approach towards interpreting the functional organization of the genome.

THE STATES OF CHROMATIN

Different chromatin states have long been recognized with the classical distinction between euchromatin and heterochromatin. Recent mapping studies have
revealed further subdivisions and, taking DamID mapping data for 53 chromatin components in the Drosophila Kc cell line, Filion et al. [4] demonstrated that the genome is broadly segmented into five principal chromatin types. Each chromatin type has a combinatorial signature of bound proteins, with few individual proteins exclusively associated with a single type. The different chromatin types are named according to colours: green, blue and black are generally associated with gene repression, whereas red and yellow correspond to regions containing active genes. Green chromatin contains classic heterochromatin marker proteins such as HP1 and Su(var)3-9. It is predominantly found in known heterochromatic areas such as pericentric regions and is associated with the heterochromatin histone modification, H3K9me2. Blue chromatin is associated with the Polycomb silencing machinery and is enriched in the histone mark H3K27me3 imparted by Polycomb repressive complex 2 (PRC2). Black repressive chromatin is surprisingly widespread, occupying 48% of the non-repetitive genome, and is associated with Histone H1 and the chromatin proteins D1, Ial and SuUR. The relatively few genes it contains show little or no expression. Of the active chromatin states, yellow contains predominantly house-keeping genes and is associated with the H3K36me3 mark linked to elongating RNA polymerase, whereas red chromatin contains active genes with more complex expression, is enriched in specific gene-regulatory components and transcription factors and surprisingly has little H3K36me3. These chromatin states segment the genome into an array of domains with a median domain length of around 7 kb but with some domains extending over 700 kb. A similar set of epigenetically marked domains was identified by the Drosophila ModENCODE consortium who profiled 18 histone modifications in two cell lines, S2 and BG3 [3]. The combinatorial patterns segregated into 9 chromatin states and integration with mapping data on 16 chromosomal proteins and gene expression data revealed specific domain signatures. These nine states are clearly related to the five chromatin colours of Filion et al., with the principal difference that the red and yellow states are further divided into a number of sub-states; for example, a state distinguished by high H4K16ac together with H3K36me3 that is markedly enriched on the X chromosome and associated with dosage compensation in the male.

### Structural Domains Based on Proximity Mapping

The identified chromatin states provide a map of the domain organization of the genome but how does this relate to the structural organization of chromatin? A key technique to probe large-scale chromatin organization is chromatin conformation capture (3C), which is based on the enhanced ligation of chromatin fragments that are in close proximity in cross-linked chromatin [9]. Originally developed to monitor interactions between two specific sites, the protocol has been developed in numerous variants (4C, 5C, etc.) that enable progressively more extensive interaction analysis. A recent variant, Hi-C, combining the 3C approach with high-throughput sequencing enables global mapping of interactions across the entire genome [10]. A modified Hi-C approach has been applied to Drosophila embryo chromatin to derive a high-resolution genome-wide contact map [11]. This map segments the 130 Mb non-repetitive Drosophila genome into >1000 physical interaction domains, each domain representing a region of significant interaction flanked by domain borders. The domain borders are defined by the relative lack of interactions across them.

By correlating these physical domains with mapping profiles for chromosomal proteins and histone modifications, Sexton et al. [11] define four classes of domain; null, active, Polycomb-Group (Pc-G) and...
HP1/Centromere. In general, these domain classes map well onto the Filion et al. chromatin states (with ‘active’ being a combination of the red and yellow states) with clear examples of boundaries of chromatin marks aligning precisely with borders of the physical domains. This is exciting as it brings together the physical and epigenetic maps and provides a fundamental insight into chromatin domains as the functional building blocks of genome organization (Figure 1). Interestingly, the active domains show intra-domain interaction properties that differ from the repressive domain classes (Null, Pc-G and HP1/Centromere) supporting the idea that different epigenetic states are associated with specific types of chromatin folding.

**DOMAIN BOUNDARIES**

Sexton et al. investigated which epigenetic marks are associated with domain borders and found clear enrichment of the insulator components CP190, BEAF-32 and CTCF and the chromodomain protein Chromator (also known as Chriz). Insulators have been widely proposed as key elements that may organize the domain architecture of the genome [12, 13]. Insulator elements are functionally defined by their ability to block enhancer-promoter interaction and by their capacity to prevent the spreading of repressive chromatin states. Genomic mapping studies have shown in both Drosophila and vertebrates that insulators flank regulatory domains (e.g. [14, 15]) and that insulator component binding sites are enriched at boundaries between epigenetically marked chromatin states [16–18].

It is still unclear how insulators function. There is considerable evidence associating insulators with the formation of chromatin loops (reviewed in [13, 19]. How does this fit with the observations that insulators flank physical and epigenetically marked domains? A simple model would be that each domain is a loop with insulators forming the tie that holds the ends of the loop, and this fits with classical models of chromatin structure. The topological constraint formed by the insulator tie could be sufficient to enable loops to behave as isolated domains capable of independently adopting different chromatin states. The interaction map of Sexton et al. sheds some light on this as they observe relatively uniform contact intensities across domains and no evidence for heightened interaction between borders. This does not easily fit with the insulator-tied loop model where a predominant interaction could be expected linking the beginning and end of domains, although it is possible that this is a technical issue and that greater resolution will be required to detect such specific localized interactions. In

![Physical interaction domains from proximity mapping](#)

![Chromatin States from chromosomal protein and histone modification mapping](#)

**Figure 1:** The domain organization of the genome. In the domain model, the different chromatin states are represented by domains of different chromatin folding, separated by boundaries (open dots).
addition, although insulators are enriched at domain boundaries, Sexton et al. found that most insulator protein binding sites are actually intra-domain suggesting that the insulators that flank domains may be a special subclass. Indeed, insulator proteins appear to serve a wide variety of functions [13] so there may be many roles for loop formation by intra-domain insulators (Figure 2). Alternatively, it is possible that intra-domain insulator protein binding sites represent latent domain boundaries poised in an inactive state but capable of activation in particular cell types with the potential of generating cell-specific domain architectures.

POLYTENE CHROMOSOMES: MODELS OF INTERPHASE CHROMOSOMAL ORGANIZATION?

The association of the Chromator protein with the boundaries of interaction domains forges a link between diploid cell interphase chromatin organization and the structure of polytene chromosomes. The polytene chromosomes in fly salivary gland nuclei have long provided a dramatic map of chromatin organization across the genome. They consist of over a thousand chromatin strands lying side by side, and they reveal a conspicuous readout of chromatin organization in their banding pattern that has been used for decades to map Drosophila genes. The Chromator protein binds to interbands on polytene chromosomes (i.e. the regions lying between the chromosome bands) and Chromator mutations disrupt polytene chromosome structure [20, 21]. Chromator forms a complex with the zinc-finger protein Z4, and together they recruit the JIL-1 kinase, leading to H3S10 phosphorylation in interbands [22].

The polytene chromosome map consisting of approximately 3000 bands is generally only aligned at low resolution with the genome sequence, making detailed comparison of genomic features and polytene structure difficult. However, recently a focused analysis of an approximately 400-kb region of chromosome X has been performed [23]. The available genome-wide mapping for a set of proteins, including Chromator, that had previously been associated with interbands in polytene chromosomes enabled the prediction of the genomic locations of the boundaries of two major bands, 10A1-2 and 10B1-2. Hybridization of probes from these regions to polytene chromosomes demonstrated that they indeed flanked these bands and that they co-localized with Chromator binding. This makes a clear link not only between the epigenetic mapping in diploid cells and the distribution of proteins such as Chromator in the polytene chromosomes but also between the epigenetic map and the physical organization of chromatin reflected in the banding pattern of the polytene chromosomes. Comparison of the polytene banding with the five-colour map of chromatin state reveals that these two major polytene bands are associated with long domains consisting primarily of black chromatin. Other smaller bands with less compact chromatin lying between the two major bands generally lie in a yellow chromatin domain, and these bands too appear to be flanked by Chromator binding.

Overall, it appears that the visible structure in the banding pattern in polytene chromosomes and the domain architecture identified by proximity mapping in embryos both relate to the chromatin state landscape identified in diploid cells. This indicates similarity in the organization and suggests that the polytene banding may be generally revealing the underlying common domain organization of the genome (Figure 1).

CHROMATIN DYNAMICS

The overall similarity in domain organization between different cell types points to a relatively
constant chromatin architecture but the most intriguing questions concern the dynamics of chromatin architecture as cells change their gene expression profiles; for example, as cells differentiate during development. The developmental and physiological chromatin dynamics are only beginning to be investigated.

There are many questions. With respect to the domain organization, does the genome have a fixed set of domain boundaries providing a constant architectural framework within which chromatin state in any domain can vary? How do chromatin states change: what are the pathways of state change, what are the rules for which states can be converted into which other states and what are the mechanisms involved?

A start has been made by examining the constancy of some of the features in the epigenetic landscape. For example, how constant are the binding profiles of insulator components that are proposed as key determinants of domain boundaries. Comparison of insulator binding profiles between Mbn2 cells, a tumourous hematopoietic cell line, and Kc cells, derived from embryo neuronal tissue, revealed that while most binding sites are common between these two cell types, there is a significant fraction of differential sites [24]. For example, 18% of the CTCF sites in Kc cells were found to be unique. Similar levels of cell type-specific CTCF sites were found comparing Kc cells with S2 cells, a haemocyte-like mesodermal cell line [18]. However, these site comparisons are likely to over-represent differential sites due to thresholding problems. Using a non-thresholding comparative approach, Negre et al. found a smaller but still significant number (6%) of differential CTCF sites [18]. In summary, there appears to be a small number of sites with variable CTCF occupancy. Insulator function, however, does not simply depend on binding of one of the DNA-binding insulator components such as CTCF. There are additional components that are required for the construction of a functional insulator complex, such as the BTB-domain protein CP190 [25]. Some CTCF sites that were common to the two cell lines nevertheless showed differential CP190 binding indicating that insulator complex formation can be regulated at several levels potentially enlarging the scope for dynamic insulator function. Recently, Wood et al. [26] examined the changes in insulator component binding associated with cell differentiation induced by the hormone 20-hydroxyecdysone and found evidence for both induction of binding by DNA-binding insulator components [CTCF, Su(Hw) and BEAF] and also for recruitment of CP190 to sites already bound by one of the DNA-binding insulator components. Overall, it seems clear that insulator complexes can be flexible. However, as insulators appear to serve a variety of functions [13] and as only a minority of insulator component binding sites are located at domain boundaries, the implications of this flexibility for the stability of the domain organization of the genome are not yet clear.

LONG-RANGE INTERACTIONS

Proximity mapping approaches not only detected local interactions that defined the domain organization but also identified long-range interactions both between sites on the same chromosome and also between chromosomes. These interactions can be classified according to different types of large-scale chromatin organization within the nucleus. One clear level of contact enhancement reflects linear proximity, and chromosome arms behave as relatively isolated interaction domains. This fits well with the chromosome territories observed in mammalian nuclei [27]. A second type of organization reflects the separation of inactive 'heterochromatin' and active 'euchromatin'. This is seen in the enhanced interaction between centromeres embedded in pericentric heterochromatin, the association of centromeres with the largely heterochromatic chromosome 4 and in the clustering of telomeres. In addition, within each chromosome arm, domains bearing active chromatin marks cluster separately from domains bearing inactive chromatin marks. This separation of inactive and active domains is also a strong feature of the human interaction map [10]. A related third type of organization is the long-range interaction between domains carrying a specific epigenetic mark, and this has been most studied in the case of interactions between Pc target genes.

Interaction between Pc targets was initially proposed based on genetic studies, the occurrence of aggregates of Pc (Pc bodies) within nuclei and fluorescent in situ hybridization studies showing co-localization of Pc target genes (reviewed in [28]). Proximity mapping approaches have revealed an extensive set of interactions between Pc targets that define a 3D Pc interactome that may play an
important role in the organization of chromatin within the nucleus and in the stabilization of gene repression [29, 30]. Using a 4C protocol with an anchor biotinylated primer extension step that allows the identification of genome-wide interactions with a specified 4C bait primer site, Bantignies et al. [29] investigated interactions with a well-studied Pc target, the Fab-7 Pc Response Element (PRE) in the Bithorax Complex (BX-C) on chromosome 3R. The Fab-7 PRE regulates the silencing of the Abd-B Hox gene and the experiment used larval tissues where Abd-B and the other genes of the BX-C are switched off. The strongest interactions were found within the 300 kb domain encompassing the three silenced Hox genes of the BX-C. Outside this domain, long-range interactions were found along the length of the 28 Mb 3R chromosome arm and showed a clear correlation with the locations of Pc target genes. The top eight long-range interaction sites all map to previously identified Pc-bound regions, and in global analysis, the interaction sites were found to be highly enriched for Pc-binding and for the H3K27me3 histone modification associated with the activity of PRC2. In a similar 4C experiment, again using larval tissues where the Abd-B gene is repressed, Tolhuis et al. [30] identified long-range interactions using the Abd-B promoter region as bait. They demonstrated 17 regions dispersed over chromosome 3R that significantly interact with Abd-B, and they showed that these regions were highly enriched for Pc targets. Similar interactions between Pc targets were found using other Pc targets (e.g. the ANT-C Hox complex) as the 4C bait. The selectivity for interaction with Pc targets did not simply reflect clustering of inactive loci as using repressed non-Pc target genes as baits revealed long-range interactions that were not enriched in Pc targets. Thus, the basis for the long-range contacts appears to be interaction between regions carrying specific epigenetic marks. However, the interactions between Pc targets do not represent co-clustering of Pc targets sites throughout the whole genome. Tolhuis et al. emphasize that there is a strong topological constraint on these interactions as they are predominantly restricted to contacts between loci on the same chromosome arm. In a revealing experiment, they studied long-range interactions with the ANT-C Pc target in a fly strain that is homozygous for a chromosome inversion that moves the ANT-C locus from the right arm of chromosome 3 to the left arm. In the inversion strain, ANT-C now formed novel interactions with Pc targets on 3L and generally lost the interactions with 3R Pc targets that had been seen in the wild type. So chromatin interactions within the nucleus are not simply determined by affinity between specific epigenetic complexes, but rather topology plays a predominant role in determining the interaction frequency.

Li et al. [31] have analysed the basis for the specificity of interactions between Pc targets. Using transgenes carrying boundary element fragments from the BX-C, they investigated interactions both between transgenes, and between transgenes and endogenous elements. In this system, both intrachromosomal and inter-chromosomal 3C interactions were observed. Two boundary elements were studied, Fab-7 and Mcp, each containing both a Pc-binding PRE and an insulator element. Dissecting the sequence requirements for these long-range interactions revealed, surprisingly, that the interactions were dependent on the insulator elements and not on the PREs. How insulator-dependent interactions can form a specific set of contacts between Pc target genes is not clear, but the diversity of insulator components found in insulator complexes could provide a basis for selective interactions.

An interesting question is how important are these long-range interactions for the control of gene expression. The inversion experiment described above gave Tolhuis et al. an opportunity to test this as the inversion clearly had a strong effect on specific interactions; for example, the interaction between the BX-C and ANT-C. Gene expression profiling by microarray did not detect significant changes and a phenotypic test based on the activity of Pc target gene, Scr in the ANT-C, in determining the number of sex combs on the first thoracic legs of male flies also detected no effect. In contrast, Bantignies et al. [29] detected effects of deleting the Fab-7 PRE in the BX-C on the expression of genes in the ANT-C and a small effect in a phenotypic assay based on Scr function. Further work is required to clarify the effects of interfering with specific long-range contacts, and a different experimental strategy is required for testing the role of Pc target interaction in general as, potentially, specific contacts may not be as important as generalized contact between Pc-bound regions for stabilizing Pc-mediated gene silencing.
HOTSPOTS AND CHROMATIN ACCESSIBILITY

Chromosomal protein mapping studies have revealed that some regions of the genome stand out as hotspots of protein binding [1, 2, 4, 32, 33]. These regions are a feature of red chromatin, and they are also characterized by decreased nucleosome density, increased nucleosome turnover and increased histone variant H3.3, which is associated with nucleosome displacement. These hotspots and the association with ‘open’ chromatin raise the important issue of how the packaging of the genome into a variety of chromatin states affects the accessibility of DNA and the targeting of chromosomal proteins such as DNA-binding transcription factors. This issue was highlighted in the studies mapping the genome-wide binding profiles of 21 transcription factors in the early Drosophila embryo [1, 33]. Despite their diversity in terms of DNA-binding domains, binding specificities and involvement in different developmental pathways, these transcription factors showed highly overlapping binding profiles. Using a high-throughput assay (DNase-Seq) measuring sensitivity of chromatin to DNase1 digestion, genome wide maps of DNA accessibility were generated for five stages of early embryogenesis [5]. Comparing the DNA accessibility and transcription factor profiles reveals a striking 87% overlap [34]. This fits with the idea that, for many transcription factors, chromatin inaccessibility would effectively limit the available target DNA. In stage 5 (blastoderm), embryos only ~12% of the genome is classified as DNase1 accessible; thus, this effect could have a major impact on the target genes available to be bound and regulated by a specific transcription factor. In this way, chromatin accessibility would be a major contributor to the design of developmental gene regulatory networks. A key role for chromatin accessibility is supported by the observation that in vivo binding profiles for transcription factors are better correlated to chromatin accessibility than to the occurrence of their specific recognition sequences in genomic DNA [34]. As DNA accessibility is dynamic during development, some DNA-binding proteins should be able to bind ‘inaccessible’ DNA, and such pioneer factors would be responsible for initiating the formation of novel accessible regions. For such factors, a correlation between recognition site occurrence and chromatin accessibility should be observed, and this has identified a few potential pioneers, Hairy, Runt and Snail, among the early embryo transcription factors.

The heat shock system has been exploited to investigate the chromatin states that permit transcription factor binding [35]. Upon heat shock in S2 cells, heat shock factor (HSF) binds to ~15% of its binding motifs in the genome. This inducible system enables characterization of the chromatin state prior to transcription factor binding, and this reveals a strong correlation between the prior presence of active chromatin marks (e.g. H3K4me3, H3K27ac and H3K9ac) and HSF binding. By switching a HSF binding motif in an ec dysone-inducible gene from an inactive to an active chromatin state by ec dysone treatment now enables HSF to bind to the motif, emphasizing the role of the chromatin landscape in determining the availability of transcription factor target sites.

The relationship between chromatin accessibility and the chromatin state domains is complicated. The DNase1-Seq analysis has concentrated on the peaks of accessibility (DNase1 hypersensitive sites), which may mark individual elements, such as enhancers, rather than on wider regions of DNase1 accessibility that may be associated with chromatin state domains. Nevertheless, domains rich in DNase1 hypersensitive sites are related to red or active chromatin and plasticity in DNase1 hypersensitive sites between cell lines is associated with transitions between chromatin states [3]. A link between chromatin state and the domain-wide binding of transcription factors is seen in the binding of the homeodomain proteins Ubx and Hth to the Pc-regulated domains of the BX-C [36]. Using chromatin from specific larval segments, where different BX-C domains are active, allowed investigation of binding in domains in both active and repressed states. Widespread binding was observed in active domains, but homeodomain protein binding appeared to be excluded from repressed domains suggesting that, at least for Pc domains, epigenetic state affects domain-wide availability of the DNA for transcription factor binding.

PROSPECTS

The coming together of chromosomal protein and histone modification genomic profiles with proximity mapping has produced a clear view of the functional packaging of the genome into physical domains that may adopt a variety of chromatin states. This identification of domains as a fundamental unit of genome architecture is an important step forwards in our understanding but also leads to many
further questions. It is important to determine what functions actually operate at the domain level. Domains of different chromatin states are distinguished by combinations of associated proteins and histone modifications. Some of these features can be co-extensive with the domains but other features are more local; for example, specific marks associated with enhancers, promoters, or elongating polymerase. The latter features occur within domains and can contribute to the domain signature allocating it to a particular chromatin state, but they may be distinct from processes that act at the domain level, e.g. to open up an inactive domain or to mediate the transition from one chromatin state to another. It will be interesting to see if there is a discrete set of functions of a more architectural nature that operate on domains to permit or restrict the activities of the functional elements that reside within domains. There are a variety of candidates for domain-level functions, Pc-repression, Lamin-binding, presence of linker histone H1 but for several proteins, for example CTCF, it has been difficult to disentangle their potential architectural roles from more local gene regulatory functions. In general, although some of the molecules involved have been identified, there remains much to discover about domain functions, domain interactions and the organization of domains within the nucleus.

A key area for the future is the investigation of domain dynamics. Current studies comparing different cell lines or in vivo developmental stages give an indication of the plasticity of domain states but are of limited use for the investigation of mechanisms of domain state transitions and their role in developmental cell fate decisions. Recent developments of genome-wide mapping on selected cell populations provides a basis for the analysis of in vivo domain dynamics [37]. Continuing technical advances together with the weight of existing studies on the Drosophila genome make the fly a major model for the analysis of the many remaining questions concerning how genome packaging relates to genome function.

Key points

- Analysis of genome-wide maps of chromosomal protein binding and histone modification reveals the basic packaging of the genome into a small number of chromatin states.
- Whole genome mapping of chromatin interactions reveals the organization of the genome into structural domains.

FUNDING

This work was supported by the Wellcome Trust [project grant: 089834/Z/09/Z].

References


