Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Genetics & Development

Chromosomal domains: epigenetic contexts and functional implications of genomic compartmentalization

Amos Tanay[1,2] and Giacomo Cavalli[3]

We review recent developments in mapping chromosomal contacts and compare emerging insights on chromosomal contact domain organization in *Drosophila* and mammalian cells. Potential scenarios leading to the observation of Hi-C domains and their association with the epigenomic context of the chromosomal elements involved are discussed. We argue that even though the mechanisms and precise physical structure underlying chromosomal domain demarcation are yet to be fully resolved, the implications to genome regulation and genome evolution are profound. Specifically, we hypothesize that domains are facilitating genomic compartmentalization that support the implementation of complex, modular, and tissue specific transcriptional program in metazoa.

Addresses
[1] Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel
[2] Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel
[3] Institut de Génétique Humaine, UPR 1142, CNRS, 141, rue de la Cardonille, 34396 Montpellier, France

Corresponding authors: Tanay, Amos (amos.tanay@weizmann.ac.il) and Cavalli, Giacomo (giacomo.cavalli@igh.cnrs.fr)

Current Opinion in Genetics & Development 2013, **23**:197–203

This review comes from a themed issue on **Genome architecture and expression**

Edited by **Genevieve Almouzni** and **Frederick Alt**

For a complete overview see the Issue and the Editorial

Available online 14th February 2013

0959-437X

http://dx.doi.org/10.1016/j.gde.2012.12.009

Introduction

Genomes employ remarkably diverse architectures to store information in DNA sequences and direct all forms of biological function across the tree of life. Information is stored concisely and directly at most bacterial species genomes, where genome evolution favors concise organization and functional specialization. As organisms' complexity increase, and in particular in multi-cellular eukaryotes, genomes are expanding mildly in terms of new genes, but scale up by two to three orders of magnitudes in size from millions to billions of bases. Genetic information is then embedded into long and complex DNA sequences in a redundant and indirect fashion. Although the implications of such sparse encoding are widely believed to be profound, it was so far difficult to describe them precisely. Mechanisms that are capable or processing and possibly taking advantage of fragmented and patchy genomic encodings (e.g. RNA splicing) promote the notion that genome sequences are heterogeneous in their information content, ranging from perfectly optimized elements similar those making up bacterial genomes to 'junk'-like sequences spanning millions of bases with seemingly no direct function. In contrast, numerous recent studies are utilizing high throughput sequencing to generate rich maps of genomic and epigenomic activity, suggesting that much of the genome is under selection [1,2] and involved in gene regulation. Ultimately, understanding genome function, and describing how and why metazoan genomes are so large, complex and redundant, must be achieved through physical characterization of genome and chromosome structure. In this short review we survey recent technological and analytical advances leading to new insight into the structure of complex chromosomes. By mapping chromosomal contacts, we propose, geneticists and epigeneticists are finding vital clues that may lead to integrative, physical and mechanistic models of genome function.

Genomic techniques are revolutionizing the study of chromosomal architectures

Historically, the study of chromosomal architectures relied on structural and biochemical studies of nucleosomes and their modifications at the local level (reviewed in [3]) and on fluorescence-based microcopy (reviewed in [4]) for studying longer range contacts and global chromosomal organization. The development of chromosome conformation capture [5] by Dekker and others and the combination of 3C with powerful genomics approaches [6••,7••,8••,9–11] facilitated the quantification of chromatin contacts at unprecedented scale and breadth. 3C is performed through fragmentation of the genome (using, e.g. sequence specific restriction enzymes) followed by re-ligation of DNA fragments that were crosslinked together, owing to physical proximity at the time of nuclei fixation. Through the genomic approach, proximity-ligation events for millions of loci can be assayed simultaneously across millions of cells, comparing to only few pairwise contacts that can be approached via traditional FISH. Moreover, the granularities at which 3C experiments are performed depend on the genome fragmentation and can therefore theoretically approach the kilobase scale [8••] or even better, comparing favorably to diffraction limited traditional microcopy or even refined imaging techniques [12].

3C is providing biased probabilistic indications of proximity. The extensive genomic coverage and high-resolution restriction site grid provide 3C-based techniques with a remarkable potential to revolutionize chromosome research. Despite this potential, physical interpretation of 3C data, and modeling of chromosomal architectures based on it remains challenging. Any 3C experiment (regardless of the downstream genomic processing performed) involves quantification of re-ligation frequency between pairs of genomic fragments. Globally, these frequencies are known to be correlated with physical proximity (e.g. as demonstrated by many FISH experiments) [8••,9,13]. At a more quantitative level however, it is clear that physical proximity is not the only factor affecting 3C contact frequencies. For example, some natural genomic parameters, including the size of the restriction fragments and nucleotide composition, correlate strongly with 3C-ligation frequencies and can be shown to contribute probabilistically to a variation in contact intensities spanning more than an order of magnitude (in Hi-C [14] or 4C-seq [15•] experiments). It is currently not well understood to what extent other factors, including those linked with epigenomic features like nucleosome composition, replication timing, and binding by trans-factors, can contribute to enhanced crosslinking, fragmentation, or successful recovery of 3C-aggregates. Such uncharacterized biases will need to be further resolved and clarified in future studies. Even more fundamentally, the statistical nature of 3C, which is averaging chromosomal conformation over millions of nuclei, requires particular attention by analysts and modelers. Current methods cannot distinguish between strong contacts occurring at low frequencies and weak contacts occurring consistently within the nuclei population – since both scenarios can generate a similar number of contacts on average. Likewise, equally strong contacts in terms of molecular affinity ('on rates') might potentially last more or less time ('off rates') if the overall or the local chromatin mobility is different. Once again, variations in chromatin dynamics may thus result in variations in 3C signal strength. Modeling of 3C-contacts must take these aspects into account, considering the variation in the structure of individual nuclei as documented by years of microcopy studies. In summary, current data support the idea that 3C contact frequencies are robust indications for chromosomal proximity, but also suggest that more work is needed to bolster our ability to interpret 3C data in the context of quantitative models for chromosome architecture.

3C maps consistently reveal chromosomal domain structure. Scaling up 3C experiments using large 5C libraries [16–18,19••] or combining 3C into open-ended protocols generated comprehensive 3C contact maps encompassing many megabases of chromosomal territories in yeast, *Drosophila*, Mouse and Human cells [6••,7••,8••,9]. The analysis of such maps first reconfirmed known physical properties of chromosomes, and then proposed significant genome wide generalization and higher resolution refinements of these properties. The maps confirmed a strong presence of chromosomal territories, clearly distinguishing contacts between elements in the same chromosome and contacts crossing chromosomal boundaries. Chromosomes were then shown to divide according to activity patterns, with chromosomal elements harboring actively transcribed genes tending to contact other such elements more often than regions lacking active genes [8••,20]. Going beyond these coarse grained models of chromosome structure, higher resolution analysis revealed novel modular structures that package genomic regions into domains with strong internal connectivity and limited external interactions. The resulting physical or topological domains (Figure 1) create an attractive framework for modeling chromosome structure, simplifying (at least theoretically) the problem into understanding how domains contact each other to form together higher order structures. In *Drosophila*, about 1000 domains sizing around 100KB each were described. In human and mouse, 2000–3000 domains were described, measuring around 1MB on average, suggesting a modular chromatin organization similar to *Drosophila*, but with modules of larger size. Interestingly, mammalian genes are also about one order of magnitude larger than their fly counterparts. Whether the conserved ratio between domain and gene sizes is circumstantial or more deeply linked to how domains are established remains unknown. Importantly however, no domain structure was described in yeast [21], where a compact and gene-packed genome is divided into chromosomes that are typically in the size of one *Drosophila* domain.

The epigenomics of 3C domains. The consistent evidence for 3C contact domains in *Drosophila* and mammals led to many questions regarding the physical structure underlying such domains, and the implication of such structures on genome function. 3C domains were found to correlate strongly with *linear* epigenetic marks, including histone modification enrichments, active gene density, lamina interaction, replication time, nucleotide and repetitive element composition [8••]. The combination of these marks, that were previously studied statistically to extract epigenomic domains and classify them [10,11,22••], was shown to distinguish many of the identified 3C domains, allowing their broad classification into groups. In flies, such classification included domains showing active transcription (active domains), repressive domains localized toward the nuclear periphery and lacking specific epigenetic enrichment (null domains), domains bound by Polycomb complexes (Polycomb domains) and domains enriched with heterochromatic marks (Hp1 domains). In mammals, clear identification of active and null domains is evident, while Polycomb and Hp1 domains, if exists, are likely to be smaller than 1MB in scale, making their detection using current maps difficult. The correlations