Review

CellPress

A matter of scale: how emerging technologies are redefining our view of chromosome architecture

Diego I. Cattoni, Alessandro Valeri, Antoine Le Gall, and Marcelo Nollmann

Centre de Biochimie Structurale, CNRS UMR5048, INSERM U1054, Université de Montpellier, 29 rue de Navacelles, 34090 Montpellier, France

The 3D folding of the genome and its relation to fundamental processes such as gene regulation, replication, and segregation remains one of the most puzzling and exciting questions in genetics. In this review, we describe how the use of new technologies is starting to revolutionize the field of chromosome organization, and to shed light on the mechanisms of transcription, replication, and repair. In particular, we concentrate on recent studies using genome-wide methods, single-molecule technologies, and super-resolution microscopy (SRM). We summarize some of the main concerns when employing these techniques, and discuss potential new and exciting perspectives that illuminate the connection between 3D genomic organization and gene regulation.

The importance of context

In addition to the information encoded in the genetic sequence, chromosomes encompass other channels of information, including epigenetic modifications, regulatory sequences, and spatial organization. These distinct information channels, directly or indirectly, define how the genetic code is interpreted. The spatial organization of genomes influences, and is affected by, many fundamental cellular processes, such as gene expression, DNA repair, DNA replication, and chromosome segregation. Recent technological advances have considerably accelerated our understanding of how chromosomes are organized in the cell, revealing several novel scales of DNA organization, and are starting to unveil the molecular mechanisms involved in this complex organization. In this review, we describe some of the most recent state-of-the-art technologies and discuss how these have been applied to unravel new principles of chromosome organization. In particular, we concentrate on genome-wide, single-molecule, and SRM methods and address their advantages, limitations, and future challenges.

General chromosome organization

The DNA of a eukaryotic cell must be compacted by four orders of magnitude to fit into a micrometer-size nucleus. This is mainly accomplished by the hierarchical folding of DNA. At the lowest level of organization, naked double-stranded DNA is wrapped around an octamer of core histone proteins, which are the primary architectural elements of the chromatin fiber, to form a nucleosome [1]. In eukaryotes, chromosome folding changes dramatically during the cell cycle, with interphase and metaphase chromosomes displaying probably the most well-defined and distinct architectures. During interphase, chromosomes occupy discrete, albeit loose, nuclear regions termed chromosome territories [2,3]. Within territories, chromosome segments adopt complex organizations and topographies depending on their degree of compaction and can be dynamically repositioned to regulate gene expression [4]. Before cell division, a remarkable structural transformation takes place and chromosomes reorganize spatially and converge to a highly condensed mitotic chromosome state adopting a characteristic X-shaped morphology before cytokinesis takes place [5].

Glossary

3C: chromosome conformation capture. Method enabling the determination of interaction frequencies between pairs of genomic loci.
CRISPR-Cas: clustered, regularly interspaced, short palindromic repeat technology. A new approach for generating RNA-guided nucleases that allow for editing, regulating and targeting genomes (more widely defined as RNA-guided engineered nucleases RGENs).
FCS: fluorescence correlation spectroscopy. FCS is a microscopy-based methodology that by correlation analysis of fluorescence intensity fluctuations (in time) of diluted solutions of fluorophores permits the measurement of diffusion coefficients and average concentrations.
FCCS: fluorescence cross-correlation spectroscopy is an extension of the FCS technology that measures the correlation between nonspectrally overlapping fluorophores providing a highly sensitive measurement of molecular interactions.
Genomic 3C: 3C methods allowing for the identification of chromatin interactions across the entire genome (including Hi-C and 3C-seq).
Oligopaint: A PCR-based method for generating highly efficient single-stranded DNA FISH probes from complex DNA libraries. This flexible method enables the visualization of large genomic regions, ranging in size from tens of kilobases to megabases.
PALM: photo-activated localization microscopy and **STORM:** stochastic optical reconstruction microscopy are wide-field fluorescence microscopy imaging methods that rely on the localization of single molecules to achieve spatial resolutions below 20–30 nm.
RICS: raster image correlation spectroscopy measures correlations in space to measure molecular dynamics and concentrations from fluorescence confocal images.
SIM: structured illumination microscopy. Super-resolution imaging technology enabling multicolor, 3D imaging at resolutions of ~100 nm.

Corresponding author: Nollmann, M. (marcelo.nollmann@cbs.cnrs.fr).
Keywords: chromatin structure; chromosome conformation capture; super-resolution microscopy; single molecule microscopy; gene expression.

0168-9525/
 http://dx.doi.org/10.1016/j.tig.2015.05.011

In contrast to eukaryotes, bacteria do not go through a defined interphase period and most DNA processes (e.g., replication, transcription, and repair) occur simultaneously making studies of chromosome folding and regulation more challenging. Bacteria organize their genetic material following a hierarchical organization similar to that observed for eukaryotes: at the lowest level of organization chromosomes are condensed mainly by negative supercoiling and by histone-like, nucleoid-associated proteins (NAPs), while at higher levels they are segmented into micro- (10–20 kbp in sizes, [6]) and macrodomains (hundreds of kbp in size, [7]) that define topological domains and genetic insulation of different chromosomal regions.

Genome-wide methods to study 3D folding of DNA

Chromosome conformation capture (3C; see Glossary) was developed in the early 2000s as a biochemical strategy to analyze the physical contacts between different regions of a chromosome and between the different chromosomes in cell populations [8]. Briefly, 3C methods seize a snapshot of the chromosome conformation at a given moment by crosslinking (by means of chemically fixative agents) DNA segments in close spatial proximity followed by a re-ligation step and PCR amplification or DNA sequencing (Figure 1A). The results obtained are then depicted as frequency contact maps (Figure 1B) representing the probability of two DNA segments being close in space. More recently, the coupling of 3C to high-throughput sequencing [9–11] permitted the detection of genome-wide chromatin interactions and the study of local chromatin folding at scales ranging from tens to hundreds of kilobase pairs with resolutions that have recently reached 2 kbp [12]. Currently, the two most commonly used strategies for genome-wide studies, Hi-C and 3C-seq [10,11], differ in that Hi-C includes a step to introduce biotinylated nucleotides at ligation junctions, enabling their specific purification, whereas in 3C-seq, the digested DNA is re-ligated and directly sequenced without biotin incorporation (Figure 1A). The application of these genomic 3C methods (Hi-C, and 3C-seq and derivatives) led to the discovery of a new level of organization for interphase chromosomes that is characterized by chromosomal regions displaying high frequencies of self-interactions termed physical domains, topologically associating domains (TADs), or topological domains [11,13,14]. In certain cases, TADs correlate with active and repressive chromatin states [11,13–15]. TADs can vary greatly in size, ranging from megabases to hundreds of kilobases depending on the organism (1 Mbp and ~100 kbp in human/mouse and *Drosophila*, respectively [11,14,16]).

Importantly, genomic 3C methods have also been instrumental in the study of the principles and mechanisms regulating gene expression. In human and mouse cells, over 10^5 and 10^6 loops or long-range interactions (LRIs) linking transcription start sites and distal elements were observed [17,18]. By ligating DNA in intact nuclei (*in situ* Hi-C), Rao *et al*. [19] generated much denser Hi-C maps reaching resolutions of ~1 kbp and found a much lower number of interactions for the whole human genome (~10^4 loops). Although there are significant differences in the total number of LRIs, probably associated with the limitations of each method and the lack of standardized data analysis strategies, all of these studies reinforce the previous hypothesis proposing that the main mechanism by which regulatory elements communicate with their cognate target genes is by chromatin looping [20]. Over the past decade, several other genome-wide technologies have been developed that complement well the different 3C methodologies. ChIP-microarray (ChIP-Chip [21]) and ChIP-sequencing (ChIP-PET and ChIP-Seq [22], Figure 1A) allow for the identification of specific protein binding sites, whereas gene expression measurements allow for the differentiation of active or repressed transcription chromatin regions. The combined use of genomic 3C, Chip, and gene expression datasets indicates that loops frequently link promoters and enhancers, correlate with gene activation and are enriched with insulator elements, particularly the insulator binding factor CTCF (CCCTC factor).

Chromosomes are dynamic structures and repositioning of a locus relative to nuclear compartments and other genomic loci can regulate its function. A recent study further unveiled the structure of interphase chromosomes and showed that during replication, timing transition regions (i.e., regions of the chromosome separating early and late replication) share a near one-to-one correlation with TAD boundaries and that early replication is associated with diminished interactions with the nuclear lamina [23]. By combining 5C and Hi-C, Naumova *et al*. [24] shed further light on the two distinct chromosome scaffolds observed before and after replication (i.e., the interphase and mitotic stages of the cell cycle). Interestingly, when evaluating different human cell lines, it was observed that chromosomes differed largely in their compartmental and TADs organizations during interphase, while during mitotic phase all cell types converged to virtually identical homogeneous interaction maps for all chromosomes. These results suggest that during the cell cycle chromosomes alternate between cell type-specific and locus-specific interphase organizations, and puts forward the existence of a universal cell-type and locus invariant mitotic conformation.

Long-range gene regulation may involve epigenetic components including proteins of the Polycomb group. These proteins are organized into nuclear foci (termed bodies) that contain silenced Polycomb-targeted chromatin (i.e., repressed genes). Recently, Sexton *et al*. [11], by employing 3C-seq, revealed the organization of the *Drosophila* embryonic nucleus and found that Polycomb proteins establish long-range contacts between silenced TADs (i.e., chromatin regions marked with the repressive histone H3K27me3, Figure 1B) whereas insulator proteins [e.g., centrosomal protein (CP)190, *Drosophila* homolog of CTCF (dCTCF), and boundary element-associated factor (BEAF)-32; see correlation methods section below] are likely to play a relevant role in separating TADs and eventually in organizing the internal architecture of TADs [11,15,25].

For prokaryotic organisms, studies of chromosome architecture by conformation capture are just starting to emerge. Initial models of global chromosome folding and organization in bacteria have been proposed, but only for a limited number of species [26–28]. Importantly, a recent study revealed the existence of eukaryotic-like TADs in