Understanding 3D genome organization by multidisciplinary methods

Ivana Jerković, Giacomo Cavalli

Understanding 3D genome organization by multidisciplinary methods

Ivana Jerković and Giacomo Cavalli

Institute of Human Genetics, CNRS, University of Montpellier, Montpellier, France
giacomo.cavalli@igh.cnrs.fr

Abstract

Understanding how chromatin is folded in the nucleus is fundamental to understanding its function. Although 3D genome organization has been historically difficult to study owing to lack of relevant methodologies, major technological breakthroughs in genome-wide mapping of chromatin contacts and advances in imaging technologies in the 21st century considerably improved our understanding of chromosome conformation and nuclear architecture. In this Review, we discuss methods of 3D genome organization analysis, including **sequencing-based techniques**, such as Hi-C and its derivatives, micro-C, DamID and others; **microscopy-based techniques**, such as super-resolution imaging coupled with fluorescent *in situ* hybridization (FISH), multiplex FISH, *in situ* genome sequencing and live microscopy methods; and **computational and modeling approaches**. We describe the most commonly used techniques and their contribution to our current knowledge of nuclear architecture and, finally, we provide a perspective on up-and-coming methods that open possibilities for future major discoveries.

Introduction

Euchromatin, heterochromatin and the hypothesis of individual-chromosome territories in the nucleus have been observed and suggested already at the end of the 19th and the beginning of the 20th century using light microscopy and chromatin dyes [1–3]. However, by the middle of 20th century, the chromosome territories hypothesis was largely abandoned as conventional electron microscopy failed to confirm it. It was not until the 1980s and the development of a novel imaging technique named fluorescence in-situ hybridization (FISH), that the chromosome territories hypothesis was finally validated, thereby instigating the study of nuclear architecture as we know it today [4–6]. The FISH method demonstrated the existence of chromosome territories and chromosome intermingling at territory edges, and also indicated that chromosomal regions rich in active genes largely reside at the nuclear interior, whereas chromosomal regions rich in inactive genes largely reside at the nuclear periphery [7–16] .

About two decades later, completion of sequencing of the human genome and the subsequent genome-wide characterization of genetic variations and epigenetic transcription regulation propelled the study of nuclear architecture into a new era. One breakthrough was the development of new techniques based on the principles of ligation of linearly-distal genomic regions that come into 3D spatial proximity in order to map genome organization and, at the same time, assay the functionality of this 3D organization [17,18]. However, despite the tremendous advancements made in this relative short time, the full complexity of the biophysical principles underlying the strong

3D compaction of the long linear genomic DNA into a micrometer-sized nucleus remain to be elucidated. Furthermore, we do not know the rules by which the structures imposed by general biophysical laws can be modified at specific genes in order to drive changes in gene expression programs that underlie cell fate and plasticity. These major challenges stimulate methodological improvements and invention of new experimental techniques and modeling approaches. On the other hand, fast and numerous technical developments generate confusion in the field of genome organization research, since it is unclear what information can each method provide, and the scarce comparison of concurrent methods hampers progress. Given the technological challenges and limitations of each of technology, the multiscale nature of genome organization and the multifaceted regulation of genome function, only the deployment of the full discourse of experimental and theoretical genome analysis approaches will allow us to reach a complete appreciation of genome function and the capacity to harness it in order to deliver not only fundamental knowledge but also valuable biomedical applications.

In this Review, we discuss techniques for high-throughput chromatin contacts analysis and highly multiplexed, super-resolution and live-imaging methods. The applicability of these techniques is intimately linked with computational tools, including machine learning and mathematical modeling based on first principles or driven by quantitative data. We discuss relevant technological progress, provide a view of the current state of 3D genome organization research, and discuss promising future developments.

Studying the multilayered 3D genome

The idea of the nucleus as a highly organized organelle existed for over a century [1–3]. However, so far, we grasp only a part of the principles that govern nuclear organization, and the emergence of new evidence is tightly connected with the development of new methods.

A major breakthrough in chromatin biology was the establishment of chromosome conformation capture (3C) — a nuclear ligation assay in conjunction with PCR, which marked the beginning of the era of high-throughput next-generation sequencing-based techniques for the investigation of chromosome conformation [17,18]. Indeed, a series of 3C derivatives (from here on defined as C-based techniques) were developed to assay contact frequency between multiple genomic loci, including circular 3C (4C) [19], which measures interaction frequencies of one locus with many loci ('one-vs-many'); many-vs-many assays (3C carbon copy (5C) [20], Capture-C [21–25], Capture-Hi-C [26], etc.) and genome-wide, all-vs-all assays [27,28] (Hi-C) (Box 1) [29]. Progressively, these techniques were tweaked to allow for enrichment of specific contacts driven by proteins of interest (many-vs-all),including chromatin immunoprecipitation (ChIP-loop [30]), chromatin interaction analysis with paired-end tag (ChIA-PET [31]), HiChIP [32] and proximity ligation assisted ChIP-seq (PLAC-seq [33]); or of contacts focused on selected genomic locations (Capture-C [21–25], Capture-Hi-C [26])[29]

Simultaneously with the development of the C-based techniques, ligation-independent techniques were invented to assay not only chromosome conformation in general, but also the nuclear position of chromatin contacts (tyramide signal amplification (TSA), DNA adenine methyltransferase identification (DamID), split-pool recognition of interactions by tag extension (SPRITE)) and multi-way contacts (SPRITE and genome architecture mapping (GAM)), which are not assayed effectively using ligation-based techniques [34–42]. Finally, the recent advancement of super-resolution microscopy and imaging techniques allowed us to investigate chromatin conformation of single cells at extremely high resolution and at a higher throughput than ever before [12,14,15,43–48]. In addition to improvements in spatial resolution, live-imaging in combination with genome-engineering using CRISPR–Cas9 systems facilitated and improved the study of chromatin-contact dynamics [49–51].

Owing to these methodological and technological advancements, it is not surprising that the past decade has provided major revelations in 3D genome organization and function. Most notably is the finding that chromosomes in interphase predominantly fold into two compartments, A and B, which respectively consist of predominantly gene-active and gene-inactive regions[27] (Figure 1). Furthermore, parts of compartments, from the same or different chromosomes, can come together and create hubs, which are connected by multiple chromatin interactions, thereby sharing a common function (for example, gene repression) and coalescing around different nuclear bodies such as nuclear speckles [35,37,52]. On a scale below the compartments, chromatin interactions were found to be enriched within domains of 100 kb to 1 Mb in length termed topologically associating domains (TADs); these partially insulated domains are subdivided into smaller chromatin nanodomains (CNDs) [43,53–57] (Figure 1). Both of these layers of organization — compartments and TADs/CNDs — were confirmed to be genomic features present across cell lines and species, but the principles that govern their folding are just beginning to be elucidated [58–66]. Chromatin-loop extrusion has been shown to be one mechanism responsible for folding at Mb scale. In interphase it is mediated by cohesin complex which can be blocked by CCCTC-binding factor (CTCF) bound to sequence motifs in convergent orientation, thereby demarcating TAD boundaries[58,59,72–75,61,62,64,67–71] (Figure 1). Importantly, these features are not only of structural nature, but are functional as well, as compartments are quite homogenously comprised of gene-active or gene-inactive regions and TADs can facilitate the formation of enhancer–promoter contacts within their borders (Box 2; Figure 1). Although it is not entirely clear what the relationship is between TAD boundaries, insulation and disease, structural variations perturbing TAD boundaries and changes in CTCF binding and insulation can alter gene expression and lead to developmental defects and disease [76–84] (Box 2). These exemplary findings clearly demonstrate the importance of method development and choice in studying 3D genome organization. Below, we discuss established and more recently-developed methods in detail.

2 Sequencing-based techniques

The most common sequencing-based approach to assay chromatin architecture involves crosslinking of spatially proximal chromatin fragments followed by their isolation and sequencing, which is then used as a proxy to estimate

contact frequency. Some sequencing-based techniques allow the investigation of chromatin conformation genome-wide (non-enrichment methods), whereas others depend on isolation and identification of a subset of contacts (enrichment methods). Another important distinction is between methods based on ligation of formaldehyde cross-linked chromatin fragments (C-based) and methods that do not involve ligation of cross-linked fragments (non-C-based).

2.1 Non-enrichment methods

Non-enrichment methods assay chromatin conformation at once across entire genomes. The first techniques used ligation to retain together spatially proximal fragments prior to their isolation and sequencing. Subsequently, ligation-free techniques have been developed as well. Both types of approaches capture the majority of 3D features, but they do not deliver identical information due to their inherent technical differences.

C-based methods: Hi-C and Micro-C

Hi-C is frequently used to identify 3D chromatin contacts genome-wide [27,85] (Figure 2). The original Hi-C protocol included dilution during proximity-ligation, a step introduced originally in 3C and 4C in order to favor ligation of intramolecular chromatin contacts to reduce artifacts. However, this dilution was not very effective, as it was shown that around 60% of contacts originate from inter-chromosomal interactions[85–87]. This problem was circumvented by the omission of the SDS treatment just before ligation, which allowed permeabilization of the nuclear membrane and thus chromatin digestion and ligation *in situ* [85,86]. This modification is justified by the fact that ligation occurs in fixed nuclei, where molecular diffusion is virtually absent, and by the fact that the cross-linking reaction occurs between reactive moieties that are located in the nanometer range. Therefore, *in* situ ligation enabled a more efficient capture of true contacts, thereby delivering higher resolution for the same sequencing depth. After this methodological improvement, the omission of SDS treatment was adopted in virtually all subsequent C-based techniques. *In situ* chromatin processing allowed further development of single-cell Hi-C, which was the first sequencing-based single-cell chromatin analysis technique to be successfully established [87].

Although Hi-C is suitable for the detection of compartments and TADs, its resolution is intrinsically linked with the use of restriction enzymes and to sequencing depth. Therefore, even if Hi-C libraries are sequenced by the billions of reads, the unbiased detection of local interactions, like enhancer–promoter contacts in the sub-TAD range, largely depends on the distribution of restriction sites, which in turn depends on the underlying sequence composition and thus is not uniform across the genome. This issue was first addressed concomitantly in two techniques, DNase Hi-C and Micro-C, both based on chromatin fragmentation without using restriction enzymes[88–91]. Micro-C introduced double cross-linking and replaced the restriction enzymes used in Hi-C with micrococcal nuclease digestion[91–94] (Figure 2). This produces a fairly uniform fragmentation down to the nucleosome level, which increases local resolution. In addition, Micro-C (and theoretically DNase Hi-C) also retains information on nucleosome positioning, which can be jointly analyzed with chromatin contact information from a single data set. However, although this technique is a promising improvement for the study of local chromatin topology, according to a recent preprint

article, it is less efficient in capturing long-distance and inter-chromosomal contacts compared with Hi-C [95]. Therefore, careful framing of the research questions is needed to accurately select between different non-enrichment methods.

Non-C-based methods: SPRITE and GAM

Although C-based methods have been extremely successful; they have intrinsic limitations and potential sources of bias. First, like many other chromatin analysis methods, they rely on mild formaldehyde cross-linking, which is powerful but potentially limited in capturing interactions of proteins with short residence time on chromatin or containing a low fraction of amino acids that can be crosslinked[96,97]. Second, these methods require ligation of genomic fragments prior to sequencing, a procedure that is only partially efficient. Third, they depend on short paired-end sequencing, which provides information only on bipartite interactions, whereas multipartite *in vivo* chromatin interactions escape the analysis.

Several methods that do not rely on ligation of chromatin fragments allow the detection of dual or multiple interactions. In SPRITE, crosslinked nuclei are isolated and fragmented, then individual crosslinked pieces of chromatin are uniquely barcoded using multiple cycles of a split-and-pool strategy; after high-throughput sequencing, reads carrying the same combination of barcodes represent genomic sites that are a part of the same crosslinked cluster (Figure 3). Since the method does not select for specific sequence sizes, it yields bipartite as well as multipartite contacts. SPRITE has been further adapted to facilitate the capture of DNA–DNA, RNA–DNA and RNA–RNA interactions, thereby allowing to determine whether RNAs of interest are associated with a subset of genome interactions and what the relationship of the RNA is to nuclear landmarks [37]. Finally, a single cell version of SPRITE allows the study of multi-way contacts in individual cells [36]. Future work applying this method to a broad range of cell types and analyzing two-way and multi-way contacts to a deeper level is required in order to fully exploit its advantages and appreciate its limitations [52]. GAM is an orthogonal method that can also provide frequencies of multivalent interactions[35]. In GAM, fixed cells are embedded in sucrose, frozen and cryo-sectioned, and the DNA is extracted and sequenced from each section[41,98] (Figure 3). Loci that are closer to each other in the nuclear space are co-sequenced more frequently than distant loci. Since sections are taken from multiple nuclei sliced at random orientations, the co-segregation of all possible pairs of loci among a large collection of nuclear section profiles is used to generate a matrix of inferred locus proximities. GAM matrices produce maps similar to Hi-C maps but require fewer cells — few hundred nuclei produce maps that approximate those obtained from large populations of cells in Hi-C. Like SPRITE, GAM can identify multiple interactions, thereby enabling the direct study of multivalent enhancer-promoter interactions and of higher-order chromatin structures.

2.2 Enrichment methods

The above-described techniques detect chromatin contacts present in the nucleus irrespective of genomic location, nuclear topography or the underlying protein binding. Yet, to fine-scale map chromatin folding and understand some of its functional aspects, it is necessary to detect specific contacts using enrichment approaches, thereby amplifying

the contacts signal in a specific genomic region of interest (Capture-C, capture Hi-C (cHi-C)) or for a specific protein of interest (ChIA-PET, ChIP-loop, HiChIP, PLAC-seq, DamID, DamC, tyramide signal amplification (TSA-seq))[21,22,38,39,42,99–101,23–26,30,32–34] (Figure 2).

C-based methods: HiChIP, Capture-C and cHi-C

The first C-based enrichment-dependent techniques to be developed combined proximity-ligation with ChIP, for example the low-throughput ChIP-loop and later, with higher-throughput, ChIA-PET, which was subsequently improved to allow for more efficient mapping and detection of single-nucleotide polymorphism[30,31,102]. However, proximity-ligation in these techniques was performed in non-optimal conditions, with ChIP and sonication done preceding ligation, possibly affecting the accuracy of captured interactomes. These shortcomings were addressed in the next generation of protein-enrichment C-based techniques, HiChIP and PLAC-seq, in which the C-technique is performed first and *in situ*, thereby creating an optimal environment for proximity-ligation, followed by protein enrichment[32,33,102]. HiChIP and PLAC-seq allow the identification of a subset of interactions forming in association with specific protein binding, but can only be used as a proxy, not a definitive proof, that a protein of interest mediates the captured chromatin contacts. In addition to protein-mediated enrichment, chromatin contacts can also be enriched for a specific genomic location(s) using techniques such as Capture-C and cHi-C[21–26] (Figure 2). In these techniques, a Hi-C library is first generated and then hybridized to specifically designed baits (RNA or DNA) corresponding to either one large region (several Mbs) of interest in the case of cHi-C, or to multiple specific sites in the genome (for example, a collection of specific gene promoters) in the case of Capture-C. This allows mapping of contacts in fine detail, which would normally require 20–50 fold more sequencing without enrichment [26,76,103]. In the future, capture approaches could be combined with other techniques to adjust them for specific needs.

Non-C-based methods: TSA and DamID

Ligation-independent techniques such as TSA-Seq and DamID can also enrich for contacts associated with specific proteins and map the nuclear topology[34,38,39,42,100]. TSA-Seq relies on tyramide signal amplification [34,42].Cells are first crosslinked, stained with a primary antibody against a protein of interest and then with a horseradish peroxidase (HRP)-conjugated secondary antibody[34,42]. HRP catalyzes the formation of tyramide–biotin free radicals, which diffuse and covalently link to nearby proteins, DNA and RNA. The biotin moiety can be used to stain nuclei as well as purify and sequence the associated DNA. Since the amount of tyramide–biotin signal decreases with increasing distance from the antibody localization source, TSA-Seq read frequencies can be transformed into 3D distances from nuclear landmarks of interest upon appropriate calibration. A second and orthogonal, cross-linking-independent technique is DamID, which involves tethering *Escherichia coli* DNA adenine methyltransferase (Dam) to a chromatin protein; the Dam moiety methylates adenines at GATC consensus DNA sites surrounding the chromatin protein of interest[38]. Application of DamID to proteins with distinct nuclear compartmentalization allows identifying genomic domains associated to nuclear landmarks, such as the lamina associated domains (LADs)[40]. More recent applications of DamID also enable identifying LADs in single cells and to simultaneously quantify protein–DNA contacts and RNA expression in the same cell[100,104]. Similar to DamID, DamC has been established as a cross-linking and ligation free

technique that can replace 4C; in DamC, a fusion protein of Dam and reverse tetracycline receptor (rTetR) is recruited to Tet operator sequences (TetOs) ectopically inserted at a genomic site of interest[101]. Methylated DNA is then detected by high-throughput sequencing, and scoring of the methylated Dam target sites around the TetOs allows quantifying chromatin contacts. DamC may be of great interest for low cell number or tissue-specific applications.

3 Super-resolution microscopy methods

In DNA FISH, DNA probes are hybridized to cognate genomic regions of interest and visualized by fluorescence microscopy, which allows measuring localization, shape and inter-probe distances [4–6]. However, the study of chromosome conformation and of individual chromatin contacts under the microscope has been limited by the low number of loci that can be probed simultaneously owing to the low number of available independent fluorescence channels, and by the limited spatial resolution of traditional light microscopes. These limitations have been removed thanks to major technological advancements in light microscopy applications.

Light emitted by any point source is diffracted such that the point will appear in an image as a so-called airy diffraction pattern, the size of which is proportional to the wavelength. In practice, this property of light, together with aberrations of optical systems and light scattering, limits the resolution (the minimal distance at which two signals can be distinguished), following a formula derived by the physicist Ernst Abbe, to approximately 250 nm in the *x* and *y* axes (lateral resolution) and 600 nm in the *z* axis (axial resolution), even when using the best confocal microscopes and image processing software. For decades, the resolution of light microscopy was believed to be intrinsically limited by diffraction, but imaging technologies have progressed at a remarkable speed, allowing the detection of increasing number of nuclear components at a spatial resolution surpassing the Abbe limit. These methods, collectively called super-resolution microscopy, increase spatial resolution mainly in three different ways, and here we discuss their applications in 3D genome organization research (Figure 4a). Detailed descriptions of these methods can be found in more specialized microscopy reviews [105,106].

Structured illumination microscopy (SIM) is an approach to super-resolution microscopy that increases resolution by a factor of two in each axis by exploiting a non-uniform illumination pattern: the sample is serially exposed to light from different angles and different axial phases [107–109] (Figure 4a). This illumination pattern interferes with the sample in a manner that can be conveniently analyzed in the Fourier mathematical space to improve resolution. Despite its complexity, the advent of commercially available SIM microscopes and software and the fact that the technology is compatible with standard fluorophores, labeling procedures and multi-color imaging has stimulated the widespread application of SIM. SIM has allowed to assess unprecedented details of chromatin and nuclear organization, such as the relations between chromatin and the nuclear periphery or the physical structure of TADs in single cells [43,53,110]. In particular, the analysis of mammalian TADs using SIM has revealed that TADs are subdivided into the smaller CNDs: as CND boundaries vary from cell to cell, ensemble Hi-C experiments blur their identification, thereby illustrating the power of single-cell, super-resolution imaging to illuminate 3D genome organization features that are inaccessible to cell–population-based technologies [57,111].

A second family of super-resolution microscopy methods is called single molecule localization microscopy (SMLM); it includes stochastic optical reconstruction microscopy (STORM), photo-activated localization microscopy (PALM) and fluorescence photo-activated localization microscopy (FPALM) [112–114]. All three methods use fluorophores that can be converted from a fluorescent (or activated) state to a dark (or inactivated) state (and vice versa) and rely on the stochastic excitation and detection of spatially separated single fluorophores. The spatial separation is achieved by making sure that only a small fraction of the total population of fluorophores in the sample can emit light, so that individual emitters do not overlap in a single imaging frame [115]. Sequential imaging using cycles of activation and inactivation of the fluorophores, followed by the precise localization of the center of emission of the individual fluorophores and on the superposition of all imaging cycles, generates super-resolution images that can reach a lateral resolution of 20 nm in biological samples ((3D images can be obtained using various methods)[115] (Figure 4a). The application of these methods has provided crucial insights into the fundamental folding of chromatin in the nucleus. Nucleosomes were shown to transiently interact to form clutches of various sizes interspersed with nucleosome-depleted regions [116]. At a higher scale of organization, nucleosomes were found to form CNDs, i.e., aggregates of a diameter of ~160 nm, within which individual nucleosomes display highly correlated motion in live cells [117]. This organization suggests that CNDs may arise from a coordinated behaviour that might reflect multiple, dynamic nucleosome interactions, consistent with a recent analysis of fixed chromatin by SIM [57,111]. It will be interesting to study whether CNDs might regulate genome functions such as gene expression (in particular, the frequency of Enhancer-Promoter contacts) or DNA replication.

The third approach to super-resolution microscopy is stimulated emission depletion (STED). This technique uses a configuration similar to confocal microscopy, but with an additional laser called the depletion beam, which illuminates the sample in a donut shape that has zero intensity at the center of the excitation laser [118,119]. This illumination provokes the depletion of emissions in the periphery and only allows emission in the center, thereby generating a sub-diffractive point spread function. In most applications, STED reaches a lateral resolution of 30–50 nm and high axial resolution can also be obtained (Figure 4a). The drawbacks of this method are its strong laser intensity and the requirement for specific fluorophores. A modified version called reversible saturable optical fluorescence transitions (RESOLFT) allows using lower beam intensities [120]. Importantly, SMLM and STED do not have a theoretical resolution limit and a combination of the two methods has achieved axial and lateral resolutions below 3 nm in cells [121]. Therefore, imaging applications are quickly enabling to investigate chromatin and nuclear architecture at the macromolecular scale.

A remaining limitation of these super-resolution microscopy methods is throughput, as they are limited to using two or three colors, which restricts the number of loci that can be analyzed simultaneously. Recently, however, oligonucleotide-based FISH protocols called oligopaints were coupled with microfluidics to allow multiple cycles of hybridization[122] (Figure 4b, 4c). This conjunction of methods enabled sequential probe hybridization and assaying multiple different loci with high precision in the 3D nuclear space [12,15,44–48,123]. These techniques allowed obtaining distance distribution maps among all imaged points that recapitulate the maps obtained from Hi-C experiments and

provide 3D trajectories of chromatin molecules at high resolution and in thousands of cells, something which is unattainable even in single-cell Hi-C (Figure 4c, 4d). These techniques include Multiplex FISH imaging, which helps establish high resolution tracing of chromatin folding of megabase-size genomic domains by labeling several tens of locations in the genome simultaneously [44,46,48,123] optical reconstruction of chromatin architecture (ORCA) [47]; Hi-M [45], which is a multiplexed, sequential imaging approach; and oligopaint fluorescence *in situ* sequencing (OligoFISSEQ) [14] (Figure 4b, 4c, 4d). Importantly, OligoFISSEQ is a method combining barcoded Oligopaint to *in situ* sequencing technology, which is used to read out the barcode carried by the oligos (Figure 4d). This makes OligoFISSEQ a powerful high-multiplexing technology [14]. *In situ* sequencing can also be coupled to Tn5 transposase-mediated random insertion of DNA-sequencing adapters into hundreds of positions of fixed genomic DNA. *In situ* amplification can then be used to insert unique molecular identifiers that are sequenced *in situ* prior to *ex situ* sequencing in order to identify the genomic region of Tn5 insertion. This allows obtaining the 3D location of hundreds of loci per cell, a powerful alternative to hybridization-based imaging methods [16].

Currently, it is possible to simultaneously visualize thousands of DNA loci, hundreds of different RNA molecules and several proteins or histone modifications, thereby enabling high-throughput structure–function analyses in thousands of single cells and truly inaugurating the field of spatial imaging-based 3D genomics [12,15]. These light microscopy methods are accompanied by developments in electron microscopy, with electron-microscopy tomography allowing the study of chromatin at nanometer resolution [124]. Each of these methods has advantages and limitations. For instance, methods with very high spatial resolution are typically not optimal for the description of architectures of large domains owing to a slow acquisition process and to the intrinsic noise in the images that are obtained. However, advanced OligoSTORM imaging provides powerful information on 3D genome organization that is complementary to molecular techniques such as Hi-C, thereby enabling the investigation of genome architecture and function to a degree that was unthinkable a decade ago [14].

4 Computational analysis and modelling

Although the methods discussed above provided important insights into 3D genome organization and function, they are still limited in their ability to describe how the chromatin fiber folds in the 3D space of the nucleus and they cannot predict structural changes that would result from perturbations such as mutations in genes or in gene-regulatory components. Evaluating the impact of architecture on genome function remains even more inaccessible at present. All these limitations have stimulated computational analyses and the development of mathematical modelling which, in conjunction with experiments, might help achieve a quantitative and predictive understanding of chromosome architecture and function.

4.1 Analysis of Hi-C data

The advent of Hi-C and related technologies has raised strong interest in the development of matching computational analysis tools, owing to the inherent complexity of Hi-C data. The achievable spatial resolution of Hi-C is affected by sequencing depth, library complexity and the DNA-cutting frequency of the enzyme used for chromatin fragmentation. Since the number of possible chromatin-fragment interactions is extremely high (> 10^{14} in the human genome, when using 4-base cutter restriction enzymes) and the sequencing depth of a typical experiment is limited, Hi-C matrices are sparse, that is, many entries in the matrices are 0, if they are not represented at the appropriate resolution. This makes it impossible to distinguish between genuine absence of contacts and absence of contacts owing to low sequencing depth. Furthermore, the different sizes of restriction fragments across the genome, differences in mappability between regions with high or low density of repetitive elements, and the decay of interaction frequencies with increase in genomic distance make Hi-C matrices typically very heterogeneous in terms of contacts at different genomic locations or across different distances. All these factors limit the resolution with which one can call contact regions or domain boundaries, and generate difficulties in defining the precise locations of compartments, TADs and chromatin loops. The first Hi-C study[27], which used a 6-base cutter and achieved low sequencing depth, produced reliable matrices at the resolution of 100kb and identified the compartments. To this end, Hi-C matrices were normalized by genomic distance, converted to correlation matrices and subjected to principal component analysis, which distinguished the active (A_ and inactive (B) compartment types. Later algorithms additionally applied clustering steps like Gaussian hidden Markov modelling to Hi-C maps based on much deeper sequencing, leading to further specification of epigenetic compartment signatures and more detailed stratification of the A compartment into two sub-compartments and B compartment into three sub-compartments [85,125–129].

The computational identification of TADs required high resolution maps, which were published three years after the first low-resolution HiC experiment [53–55,130]. Although nowadays TAD calling is done routinely, there is no clear consensus method but rather numerous TAD callers that are based on different principles. Initial computational approaches such as insulation score and directionality index determined TAD boundaries by defining a one dimensional linear score of a bin-fractionated genome, where the bin of the local minima (for the insulation score) or the one between local minima and maxima (for the directionality index) would determine the boundary position [55,131,132]. These approaches, however, could not inform on TAD hierarchy and missed identifying nested TADs. Subsequently, other computational approaches were developed to address this issue either by further developing the linear score approach (Matryoshka [133]), by clustering contact map data (ICFinder [134], TADpole [135]) or by using graph theory-based algorithms that identify nested TADs as contact subnetworks connecting to form larger TADs (3DNetMod [136]) . Furthermore, numerous other computational tools have been developed that combine these approaches with different efficiencies at different scales and resolutions [137–140].

The third major feature of Hi-C data are chromatin contacts and loops, which, like TADs, became detectable as the data became more resolutive. Specific chromatin contacts are defined as statistically significant increases in contact

detection in comparison with a general background model. This is the basis of Fit-Hi-C, a computational tool that assigns a statistical confidence to a contact by using random polymer modeling while accounting for known Hi-C biases such as genomic distance; an adapted version, HiC-DC, additionally accounts for sparsity and over-dispersion and yields a more conservative statistical significance estimate [141,142]. However, a locally enriched contact — a chromatin loop — might elude algorithms that use only a general background model to estimate statistical significance of enrichment. HiCCUPS, one of the first loop-dedicated algorithms, identifies a chromatin loop as the most enriched bin in comparison to its immediate neighborhood while using a high-resolution 5kb Hi-C data as an input [85]. This algorithm helped drive the discovery of a specific subtype of loops, CTCF loops, and contributed to the development of the loop-extrusion model, thereby demonstrating the importance of specialized algorithm development in order to grasp the full biological significance of the experimental data. In-depth reviews and websites have compiled and compared available compartment, TAD and loop callers [139,143–147]. The increasing robustness of primary computational tools to identify 3D genome features has led to a blooming of applications aimed to identify DNA sequences of biological significance linked with genomic regions in spatial proximity as mapped by Hi-C techniques. These applications have enabled associating genes with putative enhancers based on contact frequency, epigenomic and DNA features; identifying novel regulatory elements from genome-wide association studies; and assigning a potential role for short tandem repeats in genome organization [148–155].

Although these computational tools can provide precious information on genome structure and function, the cost of obtaining high-resolution Hi-C maps can become prohibitive, particularly when many experimental conditions need to be compared. In order to tackle this problem, a machine learning approach based on deep convolutional neural networks has been used to impute higher-coverage Hi-C maps from low-coverage data in order to increase the resolution with which loops or TAD borders can be defined [156]. This function is important in order to identify genomic features that might be involved in the regulation of these structures even if the data is not sufficiently resolutive.

4.2 3D chromatin modelling

The computational tools discussed above provide new information about regulatory elements and their function, but they do not inform on 3D architecture. Three main types of modelling strategies have been used to infer 3D genome folding, although some methods can blur this classification as they have characteristics belonging to more than one type [137,157–160]. The first modelling strategy is the bottom-up approach of polymer modelling, which attempts to infer and understand chromosome architecture from first principles, typically modelling chromosomes as self-avoiding polymers moving in a confined space that represents the nucleus (Figure 5a). Applications aim to identify components regulating 3D folding, and thus they try to reduce the number of variable parameters describing the polymer behavior while maximizing the fit between experimental data, typically Hi-C contact matrices, and analogous matrices that are derived from measuring contacts in snapshots taken at given times after starting polymer motion simulations. This modelling strategy has been applied to rationalize the decay in contact frequencies

between chromatin regions as a function of the linear distance separating them [161,162]. More recently, these models have offered an explanation for the formation of domains of active and inactive chromatin, the generation of TADs through loop extrusion and the contribution of epigenetic features such as chromatin types to the formation of TADs and compartments [58,67,128,163–166]. The combined role of loop extrusion and of active and inactive compartments in the determination of global chromosome organization has also been studied[68]. Furthermore, in addition to intra-chromosomal contact frequencies, this strategy allowed the investigation of inter-chromosomal contacts[167]. A current limitation of these models is that they can typically reproduce and predict some, but not all the features of 3D chromosome folding and in particular, they usually do not perform equally well at different scales (loops, TADs, compartments, chromosome territories) [137,138]. This is partly due to the considerable computational time required for the iterative simulation processes that are involved in generating the models, and interesting ongoing developments involve accelerating computation[168].

An alternative, physics-based bottom-up modelling strategy does not aim at minimizing the number of parameters to describe polymer behavior. Instead, it models chromosomal regions as polymers, in which each monomer represents a genomic region of fixed size and can interact with any other monomer with a specific energy. Each of the interaction energies can be adjusted until the configuration ensemble of the polymer produces a contact matrix that resembles the Hi-C data matrix. This approach allows searching for the monomers that have the most crucial role in driving the specific 3D configuration defining the genomic region of interest [169–171]. Furthermore, it can also be used to study the 3D path of the chromatin fiber in the resulting polymer models and to compare it with 3D data such as those provided by imaging methods in order to relate contacts to 3D architectural features of the region of interest [43].

The second type of modelling strategy is the top-down approach of restraint modelling, starting from data, which are often derived from Hi-C maps and sometimes integrated by maps of chromatin –nuclear lamina interactions, in order to infer the 3D architecture of genomic loci, entire chromosomes or the whole nucleus (Figure 5b). In some applications, the contact maps are used to set restraints that the models must satisfy in setting the 3D folding paths of chromatin fibers. The modelling result can be a consensus genome structure, or an ensemble of structures, which reflects the structural variability among cells or during time [28,172–179]. This type of modelling can also deliver information about chromatin folding dynamics, provided Hi-C data from time-course experiments are available. This is possible by interpolating the restraints through each of the time points [180]. Another interesting data-driven modelling uses a population deconvolution approach, in which Hi-C data are used to generate a large population of structures which, together, reproduce the experimental interaction patterns. This approach has been extended to incorporate chromatin–lamina interaction and imaging data [46,181,182].

The recent progress in computation speed led to the deployment of another cohort of 3D chromosome folding prediction tools, which are based on machine learning methods that use epigenomic and chromosome

conformation information as input and display the predicted 3D architecture as output [171,183–188]. Such computational tools rely on input data obtained from several different cell lines to train their models and identify the minimum necessary signature to accurately predict an enhancer–promoter pair, promoter–promoter pair and CTCF loops as well as for contact quantification [183,184,187,188]. Recently, two tools, Akita and DeepC, used convolutional neural networks to predict 3D folding solely on the basis of DNA sequence [185,186]. These kinds of computational tools could become very important to enable making predictions from experimental samples, for which the full epigenome datasets are not available because of sample-quantity limitations, for example in the case of patient samples.

A third type of modelling strategy, which combines both top-down data driven models and bottom-up physical models has been described [189]. This model uses parameters derived entirely from a Hi-C experiment as input, but also factors in a polymer's energy function like in bottom-up approaches. However, in this case the energy function is designed strictly from biological factors that have been demonstrated to have a role in 3D genome organization, which ensures biological relevance while simultaneously allowing for mechanical investigation. With monomers of 1 Mb in size, this model successfully reconstituted the radial positioning of entire chromosomes and uncovered previously unknown contributions of distinct biological processes (separation of A and B compartment, centromere clustering, inter-chromosomal contacts). However, the large monomer size limited resolution and prevented the capture of more local features such as sub-TAD or loop structures, but the currently available computational power could allow decreasing the monomer size and testing whether mechanistic insights into more local structures can be correctly modelled.

Improved algorithms and the ever-increasing computation power will soon allow modelling the dynamics of whole-genome folding at high spatial and temporal resolution, making computational methods crucial complementary tools to the experimental methods.

5 Emerging genome structure technologies

Many outstanding questions remain in the research of nuclear architecture, and it is therefore not surprising that new sequencing-based methods, microscopy-based methods and computational methods are continually being developed.

One outstanding question in the field is how to address single-cell variability while not compromising high-throughput. Sequencing-based techniques address this through the adaptation of C-based techniques and C-independent techniques for single cells [36,87,190]. The first of these single-cell adaptations was single-cell Hi-C (scHi-C), which revealed high inter-cell contact and TAD variability and indicated that TADs are highly stochastic domains [87]. This finding, however, put into question whether this major Hi-C feature, TADs, represents actual physical structures or reflects statistical average rather than a physical reality. Later, microscopy studies settled this controversy by showing that TADs do correspond with physical domains, but also that they have highly variable structures, clearly

emphasizing the need to focus on techniques that provide information on large numbers of single cells [43,44,57]. Indeed, an increasing number of chromatin analysis techniques are being developed into single-cell applications to address this issue and study chromatin conformation stochasticity and inter-cell variability [36,190–193].

Another major goal pertaining to C-based methods, is to overcome the resolution limit and potential cross-linking bias. Although alternative fragmentation techniques (Micro-C, DNase-Hi-C) successfully deal with the resolution problem, every technique that relies on crosslinking is inherently biasing fragmentation towards open-chromatin regions [88,91]. A recently published method, Cap-C, approaches this problem by exchanging the standard formaldehyde crosslinking with dendrimer crosslinking. By using three differently sized molecules, Cap-C allows homogeneous crosslinking of open and closed chromatin, thereby achieving more uniform fragmentation and higher resolution [88–91,194]. Since this is a straightforward change to current C-based protocols, it has the potential to be widely implemented.

The question of how to obtain spatial coordinates of the chromatin in the nucleus in a high-throughput manner and how to integrate chromosome conformation data with the spatial position has been difficult to answer using sequencing-based techniques. Recently, two new methods to study nuclear topology and higher-order organization based on ligation-free methodology were published, genomic loci positioning by sequencing (GPSeq) [195] , and chromatin interaction analysis by droplet based genomic analysis (ChiA-Drop) [192], which provides information on multivalent interactions (similar to SPRITE). In ChiA-Drop, cross-linked and fragmented chromatin is loaded onto a microfluidics device so that individual cross-linked molecules are partitioned into droplets that contain unique barcoding reagents (Figure 3). After pooling, high-throughput sequencing and identification of reads carrying the same barcodes, putative 3D interactions are identified. In *Drosophila melanogaster*, ChiA-Drop was performed using less than 10,000 cells and thus it could be suitable for analyzing rare cell types. Furthermore, the possibility to enrich for interactions that depend on specific proteins allows to infer the relative position of the interacting regions relative to nuclear bodies or landmarks [192]. GPSeq primarily focuses on the study of radial chromosomal positions in the nucleus by performing restriction enzyme digestion in a time course *in situ*, which allows the capture of the most nuclear-periphery-adjacent chromatin following short-term digestion, whereas the longer digestion times capture progressively more interior parts of the genome. However, to infer distances to the periphery correctly it is necessary to perform a YFISH (in which a Y-looking FISH adapter is ligated on the restriction enzyme overhang while the other side of the adapter interacts with FISH probes) coupled with super-resolution imaging in a time course. Using this approach, it is possible to investigate not only the radial position of chromosomes but also the radial positions of DNA replication, double-stranded DNA breaks (DSBs) and mutations [195]. Recently, a computational method called SPIN (spatial position inference of the nuclear genome) has been developed to predict genome-wide spatial positioning in the nucleus. The method integrates spatial multi-omics data including TSA-seq, DamID and Hi-C in a computational framework based on Hidden Markov random field to localize clusters of chromatin contacts relative to nuclear bodies such as nuclear speckles or the lamina [196]. This complementary tool of experimental and

computational multi-omics methods might provide the essential missing components in the nuclear organization research toolbox.

Finally, the last outstanding questions in the field we shall discuss are how to achieve higher spatial resolution in microscopy, and how to assay chromatin dynamics of individual genomic loci. In order to achieve spatial resolution, one has to be able to image beyond the diffraction limit. In addition to super resolution microscopy, several recent publications reported an alternative approach called expansion microscopy (ExM), in which the sample is embedded in a polyelectrolyte gel that expands 4–5 times when immersed in water [197–199] . ExM offers imaging of structures that are beyond the diffraction limit using conventional microscopy, and according to a recent preprint article, when combined with super-resolution microscopy, ExM achieved resolution of 5 nm [198]. Furthermore, ExM can be extremely powerful for spatially precise positioning of RNA species *in situ* and has been recently used in combination with FISSEQ technology to perform RNA *in situ* sequencing in an unbiased manner [200]. This technology conglomeration offers great promise for the future, as it combines different principles to achieve sub-diffractive resolution and multiplexing. One could easily imagine extending its applications to multiplex DNA FISH or *in situ* sequencing methods in order to analyze the traces of chromosomes, chromosome domains or individual loci at high resolution. However, it is important to keep in mind that sample expansion can alter the ultrastructure of the chromatin and it will be necessary to ensure that the structure remains preserved under standard conditions.

There are two major bottlenecks to studying chromatin contact dynamics of individual loci. First, in order to visualize such contacts, cells usually had to be subjected to heavy genome engineering to insert either lac-O or tet-O arrays [50,201–203]. Second, the signal must be sufficiently strong in order to visualize individual loci. Multiple different methods have been developed to reach this goal[49,50,204–207]. Chimeric array of gRNA-oligo (CARGO) and CRISPR–Cas-mediated Live FISH are examples of two independently developed live-imaging techniques that addressed these limitations by using the CRISPR–Cas9 system. CARGO uses multiple guide RNAs (gRNAs) targeting nuclease-dead Cas9 (dCas9)-EGFP to certain genomic locus to achieve efficient fluorescence-signal amplification and circumvent the need for strenuous genome engineering [49] (Figure 6a). This method enabled the detection of *cis*-regulatory element mobility during cell differentiation in relation to their expression, and offers great promise as it is relatively simple [49]. Similarly, CRISPR–Cas-mediated Live FISH utilizes dCas9 to target a region of interest, but here the gRNAs are fluorescently labeled, thereby amplifying the signal more than fourfold [206] (Figure 6b). Furthermore, the use of catalytically active Cas9 together with dCas9 allowed the simultaneous visualization of DSBs and fluorescent DSB-repair proteins, practically creating a live Immuno-FISH [206]. Finally, Live FISH was further expanded by coupling the dCas9–gRNAs with the dCas13–gRNA system, thereby granting visualization of both DNA and nascent RNA transcripts in live cells [206,207]. Together, these and similar dCas9-based techniques might become valuable for studying chromatin and transcription dynamics in live cells, and open venues for application ranging from basic science to diagnostics [50,208].

6 Future directions

Genome architecture as a field of research has come a long way in a remarkably short period of time thanks to multi-disciplinarity that was driven by technological advancements. In this Review, we discussed major discoveries in chromatin conformation and nuclear topology through a technical prism. However, there is still so much we do not understand and that is not accessible to us owing to methodological shortcomings.

The next decade will likely see a continued expansion of imaging-driven techniques with a strong emphasis on multiplexing and on live microscopy, especially in conjunction with sub-diffractive resolution. As already now we see the implementation of Live FISH, further development and specialization of these methods could possibly help to study enhancer–promoter dynamics in respect to transcriptional output [206,207]. Furthermore, live-microscopy-based techniques could be well suited to study the kinetics of transcription factor binding to chromatin, a subject that is poorly understood. In addition to live microscopy, the throughput of the super-resolution FISH-based methods will likely further increase and help study inter-cell variability. However, these predictions are based on the premise that the hardware necessary for these techniques will become more available and affordable.

Regarding sequencing-based methods, there is already a strong tendency to use different methods on the same sample, for example nucleosome occupancy and Hi-C (Micro-C, DNase-Hi-C) or bisulfite sequencing and Hi-C, in order to obtain different types of information from the same sample, but also to expand the amount of information that can be extracted from limited material such as patient or rare-cell samples [91–93,191]. This trend will likely continue, and new multifaceted approaches will emerge allowing the collection of complex data. Furthermore, existing techniques are constantly being adapted to extend their applicability. For instance, two recent Hi-C-based techniques and a SIM imaging-based technique have been described that allow inspection of sister chromatid topology at the cell population level [73,74,209].These techniques mark a significant milestone because, unlike other sequencing-based techniques, they allow the study of chromatin conformation during S phase of the cell cycle. It is possible that the implementation of these techniques will fuel discoveries related to S phase and DNA replication that were unattainable with the previously available techniques.

An important open question is the causality between topological insulation and transcription, which currently is difficult to properly address. Ideally, a single-locus proteomics approach would be appropriate to investigate the underlying proteome of a TAD border or a local chromatin insulator region in order to identify candidate insulator factors and analyze whether they play a causal role to determine insulation. Existing single-locus proteomics techniques to study topological insulation and transcription are available, but they are incredibly laborious and complicated to implement [210–212]. A major breakthrough in proteomics or in wet lab protocols to decrease input material is required for single-locus proteomics techniques to become widely applicable. However, it is an exciting lane of research that will certainly help explain the functional aspect of chromatin conformation.

Finally, we argue that not only methods driven by a technological boom, but also different and unconventional points of view should coalesce to invent new approaches and fuel milestone discoveries in genome architecture research. The importance of the interdisciplinary approaches described above will become even more prominent with future technological developments. Indeed, in order to re-orientate our field of research in the interdisciplinary direction, large consortia are being organized with the purpose of connecting different expertise and points of view. We believe that such developments should be highly encouraged and adopted even in individual laboratories, as they might promote individual projects and, in turn, in the field itself.

References:

1. Flemming, W. *Zellsubstanz, Kern und Zelltheilung*. *F.C.W. Vogel, Leipzig* (1882).
2. Heitz, E. Das Heterochromatin der Moose. *Jahrbücher für wissenschaftliche Bot.* (1928).
3. Boveri, T. *Die Blastomerenkerne von Ascaris megalocephala und die Theorie der Chromosomenindividualität*. *Engelmann* (1909).
4. Cremer, T. *et al.* Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. *Hum. Genet.* **62**, 201–209 (1982).
5. Manuelidis, L. Individual interphase chromosome domains revealed by in situ hybridization. *Hum. Genet.* **71**, 288–293 (1985).
6. Schardin, M., Cremer, T., Hager, H. D. & Lang, M. Specific staining of human chromosomes in Chinese hamster x man hybrid cell lines demonstrates interphase chromosome territories. *Hum. Genet.* **71**, 281–287 (1985).
7. Cremer, M. *et al.* Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. in *The Nucleus. Methods in Molecular Biology* 205–239 (2012). doi:10.1007/978-1-59745-406-3_15
8. Branco, M. R. & Pombo, A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* **4**, e138 (2006).
9. Chambeyron, S. & Bickmore, W. A. Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev.* **18**, 1119–1130 (2004).
10. Ferrai, C. *et al.* Poised transcription factories prime silent uPA gene prior to activation. *PLoS Biol.* **8**, e1000270 (2010).
11. Rosin, L. F., Nguyen, S. C. & Joyce, E. F. Condensin II drives large-scale folding and spatial partitioning of interphase chromosomes in Drosophila nuclei. *PLOS Genet.* **14**, e1007393 (2018).
12. Su, J. H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182**, 1641–1659 (2020).
13. Fritz, A. J., Sehgal, N., Pliss, A., Xu, J. & Berezney, R. Chromosome territories and the global regulation of the genome. *Genes, Chromosom. Cancer* **58**, 407–426 (2019).
14. Nguyen, H. Q. *et al.* 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat. Methods* **17**, 822–832 (2020).

First mention of OligoFISSEQ (DNA in situ sequencing); showcasing its multiplexing possibilities by imaging 249 loci simultaneously.

15. Takei, Y. *et al.* Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344–350 (2021).
16. Payne, A. C. *et al.* In situ genome sequencing resolves DNA sequence and structure in intact biological samples. *Science* **371**, eaay3446 (2020).
17. Cullen, K. E., Kladde, M. P. & Seyfred, M. A. Interaction between transcription regulatory regions of prolactin chromatin. *Science* **261**, 203–206 (1993).
18. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
19. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**, 1348–1354 (2006).

20. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299–1309 (2006).
21. Schoenfelder, S. *et al.* The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements. *Genome Res.* **25**, 582–97 (2015).
22. Schoenfelder, S. *et al.* Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat Genet* **47**, 1179–1186 (2015).
23. Mifsud, B. *et al.* Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat Genet* **47**, 598–606 (2015).
24. Hughes, J. R. *et al.* Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* **46**, 205–212 (2014).
25. Davies, J. O. J. *et al.* Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat. Methods* **13**, 74–80 (2015).
26. Jäger, R. *et al.* Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat. Commun.* **6**, 6178 (2015).
27. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Revelas Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
28. Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nature* **465**, 363–367 (2010).
29. Sati, S. & Cavalli, G. Chromosome conformation capture technologies and their impact in understanding genome function. *Chromosoma* **126**, 33–44 (2017).
30. Horike, S. I., Cai, S., Miyano, M., Cheng, J. F. & Kohwi-Shigematsu, T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nat. Genet.* **37**, 31–40 (2005).
31. Fullwood, M. J., Wei, C. L., Liu, E. T. & Ruan, Y. Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses. *Genome Res.* **19**, 521–532 (2009).
32. Mumbach, M. R. *et al.* HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
33. Fang, R. *et al.* Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res.* **26**, 1345–1348 (2016).
34. Chen, Y. *et al.* Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J. Cell Biol.* **217**, 4025–4048 (2018).

Introduction of TSA-seq, a cytological ruler for nuclear speckles and lamina. It is the first genomics method that enabled transforming sequencing reads into physical distances in the nucleus.

35. Quinodoz, S. A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744–757 (2018).

Introduction of SPRITE and discovery of inter-chromosomal hubs around the nucleolus and nuclear specles.

36. Arrastia, M. V *et al.* A single-cell method to map higher-order 3D genome organization in thousands of individual cells reveals structural heterogeneity in mouse ES cells. *Preprint at bioRxiv* doi:10.1101/2020.08.11.242081 (2020).
37. Quinodoz, S. A. *et al.* RNA promotes the formation of spatial compartments in the nucleus. *Preprint at bioRxiv* doi:10.1101/2020.08.25.267435 (2020).
38. Van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nat. Biotechnol.* **14**, 424–428 (2000).

Introduction of DamIG, which is used for spatial positioning of HP1 on chromatin *in vitro* and *in vivo* in *Drosophila melanogaster*.

39. Vogel, M. J., Peric-Hupkes, D. & van Steensel, B. Detection of in vivo protein - DNA interactions using DamID in mammalian cells. *Nat. Protoc.* **2**, 1467–1478 (2007).
40. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
41. Beagrie, R. A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519–524 (2017).

Introduction of the GAM method and discovery of a richness of multi-way contacts in the nucleus.

42. Zhang, L. *et al.* TSA-seq reveals a largely conserved genome organization relative to nuclear speckles with small position changes tightly correlated with gene expression changes. *Genome Res.* **31**, 251–264 (2021).
43. Szabo, Q. *et al.* TADs are 3D structural units of higher-order chromosome organization in *Drosophila*. *Sci. Adv.*

4, eaar8082 (2018).
44. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).
This study uses high-throroughput oligopaint technology to study chromatin dynamics and showes that TADs are highly stochastic in single cells, but predictable at the population level.
45. Cardozo Gizzi, A. M. *et al.* Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Mol. Cell* **74**, 212-222.e5 (2019).
Highly multiplexed sequential Oligo-FISH (RNA and DNA) based on the conjunction of microscopy and microfluidics in *Drosophila melanogaster* embryos.
46. Nir, G. *et al.* Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet.* **14**, e1007872 (2018).
47. Mateo, L. J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49–54 (2019).
48. Sawh, A. N. *et al.* Lamina-Dependent Stretching and Unconventional Chromosome Compartments in Early C. elegans Embryos. *Mol. Cell* **78**, 96–111 (2020).
49. Gu, B. *et al.* Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* **359**, 1050–1055 (2018).
50. Shaban, H. A. & Seeber, A. Monitoring the spatio-temporal organization and dynamics of the genome. *Nucleic Acids Res.* **48**, 3423–3434 (2020).
51. Barth, R., Bystricky, K. & Shaban, H. A. Coupling chromatin structure and dynamics by live super-resolution imaging. *Sci. Adv.* **6**, eaaz2196 (2020).
52. Vangala, P. *et al.* High-Resolution Mapping of Multiway Enhancer-Promoter Interactions Regulating Pathogen Detection. *Mol. Cell* **80**, 359–373 (2020).
53. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
54. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* **148**, 458–472 (2012).
55. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
56. Kalhor, R. *et al.* Developmental barcoding of whole mouse via homing CRISPR. *Science* **361**, eaat9804 (2018).
57. Szabo, Q. *et al.* Regulation of single-cell genome organization into TADs and chromatin nanodomains. *Nat. Genet.* **52**, 1151–1157 (2020).
First description of chromatin nanodomains in mammals using Oligopaints in conjunction with SIM.
58. Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Rep.* **15**, 2038–49 (2016).
59. Wutz, G. *et al.* Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* **36**, 3573–3599 (2017).
60. Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305–320 (2017).
61. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51–56 (2017).
62. Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900–910 (2015).
63. Haarhuis, J. H. I. *et al.* The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension. *Cell* **169**, 693–707 (2017).
64. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930–944 (2017).
65. Bonev, B. *et al.* Multiscale 3D Genome Rewiring during Mouse Article Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell* **171**, 557–572 (2017).
66. Ulianov, S. V. *et al.* Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. *Genome Res.* **26**, 70–84 (2016).
67. Sanborn, A. L. *et al.* Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci U S A* **112**, 6456–6465 (2015).
68. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 6697–6706 (2018).

69. Rowley, M. J. *et al.* Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Mol. Cell* **67**, 837–852 (2017).
70. Benedetti, F., Racko, D., Dorier, J., Burnier, Y. & Stasiak, A. Transcription-induced supercoiling explains formation of self-interacting chromatin domains in S. Pombe. *Nucleic Acids Res.* **45**, 9850–9859 (2017).
71. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).
72. Jerković, I., Szabo, Q., Bantignies, F. & Cavalli, G. Higher-Order Chromosomal Structures Mediate Genome Function. *J. Mol. Biol.* **432**, 676–681 (2020).
73. Oomen, M. E., Hedger, A. K., Watts, J. K. & Dekker, J. Detecting chromatin interactions between and along sister chromatids with SisterC. *Nat. Methods* **17**, 1002–1009 (2020).
74. Mitter, M. *et al.* Conformation of sister chromatids in the replicated human genome. *Nature* **586**, 139–144 (2020).
75. AlHaj Abed, J. *et al.* Highly structured homolog pairing reflects functional organization of the Drosophila genome. *Nat. Commun.* **10**, 1–14 (2019).
76. Franke, M. *et al.* Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature* **538**, 265–269 (2016).
77. Lupiáñez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet.* **32**, 225–237 (2016).
78. Despang, A. *et al.* Functional dissection of the Sox9–Kcnj2 locus identifies nonessential and instructive roles of TAD architecture. *Nat. Genet.* **51**, 1263–1271 (2019).
79. Andrey, G. & Mundlos, S. The three-dimensional genome: regulating gene expression during pluripotency and development. *Development* **144**, 3646–3658 (2017).
80. Weischenfeldt, J. *et al.* Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. *Nat. Genet.* **49**, 65–74 (2016).
81. Liu, X. S. *et al.* Rescue of Fragile X Syndrome Neurons by DNA Methylation Editing of the FMR1 Gene. *Cell* **172**, 979–992 (2018).
82. Bruneau, B. G. & Nora, E. P. Chromatin Domains Go on Repeat in Disease. *Cell* **175**, 38–40 (2018).
83. Akdemir, K. C. *et al.* Somatic mutation distributions in cancer genomes vary with three-dimensional chromatin structure. *Nat. Genet.* **52**, 1178–1188 (2020).
84. Valton, A. L. & Dekker, J. TAD disruption as oncogenic driver. *Curr. Opin. Genet. Dev.* **36**, 34–40 (2016).
85. Rao, S. S. P. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).

This paper showcased the first use of the *in situ* Hi-C protocol, and introduced HICCUPS, a contacts caller dedicated to chromatin loops.

86. Nagano, T. *et al.* Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biol.* **16**, 175 (2015).
87. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59–64 (2013).
88. Ma, W. *et al.* Fine-scale chromatin interaction maps reveal the cis-regulatory landscape of human lincRNA genes. *Nat. Methods* **12**, 71–78 (2014).
89. Ramani, V. *et al.* Mapping 3D genome architecture through in situ DNase Hi-C. *Nat. Protoc.* **11**, 2104–2121 (2016).
90. Deng, X. *et al.* Bipartite structure of the inactive mouse X chromosome. *Genome Biol.* **16**, 152 (2015).
91. Hsieh, T. H. S. *et al.* Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C. *Cell* **162**, 108–119 (2015).
92. Hsieh, T.-H. S. *et al.* Resolving the 3D Landscape of Transcription-Linked Mammalian Chromatin Folding. *Mol. Cell* **78**, 539–553 (2020).

High resolution Micro-C performed in mESC with concomitant analysis of nucleosome positioning.

93. Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Mol. Cell* **78**, 554–565 (2020).

High resolution Micro-C performed in human ESC and fibroblasts with concomitant analysis of nucleosome positioning.

94. Goel, V. Y. & Hansen, A. S. The macro and micro of chromosome conformation capture. *WIREs Dev. Biol.*

e395 (2020). doi:10.1002/wdev.395
95. Akgol Oksuz, B. *et al.* Systematic evaluation of chromosome conformation capture assays. *Preprint at bioRxiv* doi:10.1101/2020.12.26.424448 (2020).
96. Baranello, L., Kouzine, F., Sanford, S. & Levens, D. ChIP bias as a function of cross-linking time. *Chromosom. Res.* **24**, 175–181 (2016).
97. Gavrilov, A., Razin, S. V. & Cavalli, G. In vivo formaldehyde cross-linking: It is time for black box analysis. *Brief. Funct. Genomics* **14**, 163–165 (2015).
98. Kempfer, R. & Pombo, A. Methods for mapping 3D chromosome architecture. *Nat. Rev. Genet.* **21**, 207–226 (2020).
99. Fullwood, M. J. & Ruan, Y. ChIP-based methods for the identification of long-range chromatin interactions. *J. Cell. Biochem.* **107**, 30–39 (2009).
100. Kind, J. *et al.* Genome-wide Maps of Nuclear Lamina Interactions in Single Human Cells. *Cell* **163**, 134–147 (2015).
101. Redolfi, J. *et al.* DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. *Nat. Struct. Mol. Biol.* **26**, 471–480 (2019).
102. Li, X. *et al.* Long-read ChIA-PET for base-pair-resolution mapping of haplotype-specific chromatin interactions. *Nat. Protoc.* **12**, 899–915 (2017).
103. Andrey, G. *et al.* Characterization of hundreds of regulatory landscapes in developing limbs reveals two regimes of chromatin folding. *Genome Res.* **27**, 223–233 (2017).
104. Rooijers, K. *et al.* Simultaneous quantification of protein–DNA contacts and transcriptomes in single cells. *Nat. Biotechnol.* **37**, 766–772 (2019).
105. Sahl, S. J., Hell, S. W. & Jakobs, S. Fluorescence nanoscopy in cell biology. *Nature Reviews Molecular Cell Biology* **18**, 685–701 (2017).
106. Liu, Z., Lavis, L. D. & Betzig, E. Imaging Live-Cell Dynamics and Structure at the Single-Molecule Level. *Molecular Cell* **58**, 644–659 (2015).
107. Heintzmann, R. & Cremer, C. G. Laterally modulated excitation microscopy: improvement of resolution by using a diffraction grating. in *Optical Biopsies and Microscopic Techniques III* **3568**, (SPIE, 1999).
108. Gustafsson, M. G. L. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. *J. Microsc.* **198**, 82–87 (2000).
109. Gustafsson, M. G. L. *et al.* Three-dimensional resolution doubling in wide-field fluorescence microscopy by structured illumination. *Biophys. J.* **94**, 4957–4970 (2008).
110. Schermelleh, L. *et al.* Subdiffraction multicolor imaging of the nuclear periphery with 3D structured illumination microscopy. *Science* **320**, 1332–1336 (2008).
111. Miron, E. *et al.* Chromatin arranges in chains of mesoscale domains with nanoscale functional topography independent of cohesin. *Sci. Adv.* **6**, eaba8811 (2020).
112. Rust, M. J., Bates, M. & Zhuang, X. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nat. Methods* **3**, 793–795 (2006).
113. Betzig, E. *et al.* Imaging intracellular fluorescent proteins at nanometer resolution. *Science* **313**, 1642–1645 (2006).
114. Hess, S. T., Girirajan, T. P. K. & Mason, M. D. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.* **91**, 4258–4272 (2006).
115. Huang, B., Babcock, H. & Zhuang, X. Breaking the diffraction barrier: Super-resolution imaging of cells. *Cell* **143**, 1047–1058 (2010).
116. Ricci, M. A., Manzo, C., Lakadamyali, M. & Cosma, M. P. Chromatin Fibers Are Formed by Heterogeneous Groups of Nucleosomes In Vivo. *Cell* **160**, 1145–1158 (2015).
117. Nozaki, T. *et al.* Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging. *Mol. Cell* **67**, 282–293 (2017).
118. Hell, S. W. & Wichmann, J. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.* **19**, 780–782 (1994).
119. Klar, T. A., Jakobs, S., Dyba, M., Egner, A. & Hell, S. W. Fluorescence microscopy with diffraction resolution barrier broken by stimulated emission. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 8206–8210 (2000).
120. Hofmann, M., Eggeling, C., Jakobs, S. & Hell, S. W. Breaking the diffraction barrier in fluorescence microscopy at low light intensities by using reversibly photoswitchable proteins. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17565–

17569 (2005).
121. Gwosch, C. K. *et al.* MINFLUX nanoscopy delivers multicolor nanometer 3D-resolution in (living) cells. *Nat. Methods* **17**, 217–224 (2020).
122. Beliveau, B. J. *et al.* Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 21301–21306 (2012).
123. Wang, S. *et al.* Spatial organization of chromatin domains and compartments in single chromosomes. *Science* **353**, 598–602 (2016).
124. Ou, H. D. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, eaag0025 (2017).
125. Yaffe, E. & Tanay, A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat. Genet.* **43**, 1059–1065 (2011).
126. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
127. Ryba, T. *et al.* Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. *Genome Res.* **20**, 761–770 (2010).
128. Di Pierro, M., Cheng, R. R., Aiden, E. L., Wolynes, P. G. & Onuchic, J. N. De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 12126–12131 (2017).
129. Lin, D., Bonora, G., Yardimci, G. G. & Noble, W. S. Computational methods for analyzing and modeling genome structure and organization. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **11**, e1435 (2019).
130. Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* **30**, 90–98 (2012).
131. Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240–244 (2015).
132. Shin, H. *et al.* TopDom: An efficient and deterministic method for identifying topological domains in genomes. *Nucleic Acids Res.* **44**, e70 (2015).
One of the most reliable and popular TAD callers.
133. Malik, L. & Patro, R. Rich chromatin structure prediction from Hi-C data. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **16**, 1448–1458 (2019).
134. Haddad, N., Vaillant, C. & Jost, D. IC-Finder: Inferring robustly the hierarchical organization of chromatin folding. *Nucleic Acids Res.* **45**, 81 (2017).
135. Soler-Vila, P., Cuscó, P., Farabella, I., Di Stefano, M. & Marti-Renom, M. A. Hierarchical chromatin organization detected by TADpole. *Nucleic Acids Res.* **48**, e39 (2020).
One of the most recently developed TAD callers; this paper is extremely valuable for its benchmarking quality and for multiple TAD caller comparisons.
136. Norton, H. K. *et al.* Detecting hierarchical genome folding with network modularity. *Nat. Methods* **15**, 119–122 (2018).
137. Moller, J. & de Pablo, J. J. Bottom-Up Meets Top-Down: The Crossroads of Multiscale Chromatin Modeling. *Biophysical Journal* **118**, 2057–2065 (2020).
138. Bendandi, A., Dante, S., Zia, S. R., Diaspro, A. & Rocchia, W. Chromatin Compaction Multiscale Modeling: A Complex Synergy Between Theory, Simulation, and Experiment. *Frontiers in Molecular Biosciences* **7**, 15 (2020).
139. Zufferey, M., Tavernari, D., Oricchio, E. & Ciriello, G. Comparison of computational methods for the identification of topologically associating domains. *Genome Biol.* **19**, 217 (2018).
140. Forcato, M. *et al.* Comparison of computational methods for Hi-C data analysis. *Nat. Methods* **14**, 679–685 (2017).
141. Ay, F., Bailey, T. L. & Noble, W. S. Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res.* **24**, 999–1011 (2014).
142. Carty, M. *et al.* An integrated model for detecting significant chromatin interactions from high-resolution Hi-C data. *Nat. Commun.* **8**, 1–10 (2017).
143. Pal, K., Forcato, M. & Ferrari, F. Hi-C analysis: from data generation to integration. *Biophysical Reviews* **11**, 67–78 (2019).
144. Dozmorov, M., Sirusb; & Benfeitas, R. Hi-C data analysis tools and papers. Available at:

https://github.com/mdozmorov/HiC_tools/blob/master/README.md.
145. Roayaei Ardakany, A., Gezer, H. T., Lonardi, S. & Ay, F. Mustache: Multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation. *Genome Biol.* **21**, 256 (2020).
146. Jordan Rowley, M. *et al.* Analysis of Hi-C data using SIP effectively identifies loops in organisms from C. Elegans to mammals. *Genome Res.* **31**, (2020).
147. Ay, F. & Noble, W. S. Analysis methods for studying the 3D architecture of the genome. *Genome Biology* **16**, 183 (2015).
148. Nikumbh, S. & Pfeifer, N. Genetic sequence-based prediction of long-range chromatin interactions suggests a potential role of short tandem repeat sequences in genome organization. *BMC Bioinformatics* **18**, 218 (2017).
149. Rhie, S. K. *et al.* Using 3D epigenomic maps of primary olfactory neuronal cells from living individuals to understand gene regulation. *Sci. Adv.* **4**, eaav8550 (2018).
150. Hafez, D. *et al.* McEnhancer: Predicting gene expression via semi-supervised assignment of enhancers to target genes. *Genome Biol.* **18**, 199 (2017).
151. Di Iulio, J. *et al.* The human noncoding genome defined by genetic diversity. *Nat. Genet.* **50**, 333–337 (2018).
152. Javierre, B. M. *et al.* Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell* **167**, 1369–1384 (2016).
153. Mitchelmore, J., Grinberg, N. F., Wallace, C. & Spivakov, M. Functional effects of variation in transcription factor binding highlight long-range gene regulation by epromoters. *Nucleic Acids Res.* **48**, 2866–2879 (2020).
154. Malysheva, V., Mendoza-Parra, M. A., Blum, M., Spivakov, M. & Gronemeyer, H. Gene regulatory network reconstruction incorporating 3D chromosomal architecture reveals key transcription factors and DNA elements driving neural lineage commitment. *Preprint at bioRxiv* doi:10.1101/303842 (2019).
155. Madrid-Mencía, M., Raineri, E., Cao, T. B. N. & Pancaldi, V. Using GARDEN-NET and ChAseR to explore human haematopoietic 3D chromatin interaction networks. *Nucleic Acids Res.* **48**, 4066–4080 (2020).
156. Zhang, Y. *et al.* Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. *Nat. Commun.* **9**, 1–9 (2018).
157. Oluwadare, O., Highsmith, M. & Cheng, J. An Overview of Methods for Reconstructing 3-D Chromosome and Genome Structures from Hi-C Data. *Biological Procedures Online* **21**, 1–20 (2019).
158. Marti-Renom, M. A. *et al.* Challenges and guidelines toward 4D nucleome data and model standards. *Nature Genetics* **50**, 1352–1358 (2018).
159. MacPherson, Q., Beltran, B. & Spakowitz, A. J. Bottom–up modeling of chromatin segregation due to epigenetic modifications. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 12739–12744 (2018).
160. Bianco, S., Chiariello, A. M., Annunziatella, C., Esposito, A. & Nicodemi, M. Predicting chromatin architecture from models of polymer physics. *Chromosome Research* **25**, 25–34 (2017).
161. Mirny, L. A. The fractal globule as a model of chromatin architecture in the cell. *Chromosom. Res.* **19**, 37–51 (2011).
162. Barbieri, M. *et al.* A model of the large-scale organization of chromatin. *Biochemical Society Transactions* **41**, 508–512 (2013).
163. Jost, D., Carrivain, P., Cavalli, G. & Vaillant, C. Modeling epigenome folding: Formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.* **42**, 9553–9561 (2014).
164. Jost, D. & Vaillant, C. Epigenomics in 3D: Importance of long-range spreading and specific interactions in epigenomic maintenance. *Nucleic Acids Res.* **46**, 2252–2264 (2018).
165. Barbieri, M. *et al.* Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* **24**, 515–524 (2017).
166. Falk, M. *et al.* Heterochromatin drives compartmentalization of inverted and conventional nuclei. *Nature* **570**, 395–399 (2019).
167. Oliveira Junior, A. B., Contessoto, V. G., Mello, M. F. & Onuchic, J. N. A Scalable Computational Approach for Simulating Complexes of Multiple Chromosomes. *J. Mol. Biol.* (2020). doi:10.1016/j.jmb.2020.10.034
168. Ghosh, S. K. & Jost, D. How epigenome drives chromatin folding and dynamics, insights from efficient coarse-grained models of chromosomes. *PLoS Comput. Biol.* **14**, e1006159 (2018).
169. Giorgetti, L. *et al.* Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* **157**, 950–963 (2014).
170. Bianco, S. *et al.* Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat.*

Genet. **50**, 662–667 (2018).
171. Contessoto, V. G. *et al.* The Nucleome Data Bank: web-based resources to simulate and analyze the three-dimensional genome. *Nucleic Acids Res.* **49**, 172–182 (2020).
172. Rousseau, M., Fraser, J., Ferraiuolo, M. A., Dostie, J. & Blanchette, M. Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. *BMC Bioinformatics* **12**, 414 (2011).
173. Baú, D. *et al.* The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.* **18**, 107–115 (2011).
174. Trussart, M. *et al.* Defined chromosome structure in the genome-reduced bacterium Mycoplasma pneumoniae. *Nat. Commun.* **8**, 14665 (2017).
175. Zhu, G. *et al.* Reconstructing spatial organizations of chromosomes through manifold learning. *Nucleic Acids Res.* **46**, e50 (2018).
176. Paulsen, J. *et al.* Long-range interactions between topologically associating domains shape the four-dimensional genome during differentiation. *Nat. Genet.* **51**, 835–843 (2019).
177. Lesne, A., Riposo, J., Roger, P., Cournac, A. & Mozziconacci, J. 3D genome reconstruction from chromosomal contacts. *Nat. Methods* **11**, 1141–1143 (2014).
178. Trieu, T., Oluwadare, O. & Cheng, J. Hierarchical Reconstruction of High-Resolution 3D Models of Large Chromosomes. *Sci. Rep.* **9**, 1–12 (2019).
179. Tan, L., Xing, D., Chang, C.-H. H., Li, H. & Xie, X. S. Three-dimensional genome structures of single diploid human cells. *Science* **361**, 924–928 (2018).
180. Di Stefano, M. *et al.* Transcriptional activation during cell reprogramming correlates with the formation of 3D open chromatin hubs. *Nat. Commun.* **11**, 1–12 (2020).
181. Li, Q. *et al.* The three-dimensional genome organization of Drosophila melanogaster through data integration. *Genome Biol.* **18**, 145 (2017).
182. Hua, N. *et al.* Producing genome structure populations with the dynamic and automated PGS software. *Nat. Protoc.* **13**, 915–926 (2018).
183. Whalen, S., Truty, R. M. & Pollard, K. S. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat. Genet.* **48**, 488–496 (2016).
184. Li, W., Wong, W. H. & Jiang, R. DeepTACT: Predicting 3D chromatin contacts via bootstrapping deep learning. *Nucleic Acids Res.* **47**, e60 (2019).
185. Fudenberg, G., Kelley, D. R. & Pollard, K. S. Predicting 3D genome folding from DNA sequence with Akita. *Nat. Methods* **17**, 1111–1117 (2020).
First computational modelling paper describing modelling of chromatin conformation solely from the DNA
186. Schwessinger, R. *et al.* DeepC: predicting 3D genome folding using megabase-scale transfer learning. *Nat. Methods* **17**, 1118–1124 (2020).
One of the first computational tools to model chromatin conformation solely from the DNA sequence. However, unlike Akita, it requires training on the epigenetics data.
187. Zhang, S., Chasman, D., Knaack, S. & Roy, S. In silico prediction of high-resolution Hi-C interaction matrices. *Nat. Commun.* **10**, 1–18 (2019).
188. Belokopytova, P. S., Nuriddinov, M. A., Mozheiko, E. A., Fishman, D. & Fishman, V. Quantitative prediction of enhancer–promoter interactions. *Genome Res.* **30**, 72–84 (2020).
189. Qi, Y. *et al.* Data-driven Polymer Model for Mechanistic Exploration of Diploid Genome Organization. *Biophys. J.* **119**, 1905–1916 (2020).
190. de Luca, K. L. & Kind, J. Single-cell damid to capture contacts between dna and the nuclear lamina in individual mammalian cells. in *Methods in Molecular Biology* **2157**, 159–172 (Humana Press Inc., 2021).
191. Li, G. *et al.* Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat. Methods* **16**, 991–993 (2019).
192. Zheng, M. *et al.* Multiplex chromatin interactions with single-molecule precision. *Nature* **566**, 558–562 (2019).
193. Ramani, V. *et al.* Massively multiplex single-cell Hi-C. *Nat. Methods* **14**, 263–266 (2017).
194. You, Q. *et al.* Direct DNA crosslinking with CAP-C uncovers transcription-dependent chromatin organization at high resolution. *Nat. Biotechnol.* **39**, 225–235 (2020).
195. Girelli, G. *et al.* GPSeq reveals the radial organization of chromatin in the cell nucleus. *Nat. Biotechnol.* **38**,

1184–1193 (2020).
196. Wang, Y. *et al.* SPIN reveals genome-wide landscape of nuclear compartmentalization. *Genome Biol.* **22**, 36 (2021).
First tool that combines nuclear architecture (positioning) data from TSA-seq and DamID in order to compute spatial compartmentalization of chromatin in respect to the lamina or nuclear bodies such as nuclear speckles.
197. Kubalová, I. *et al.* Prospects and limitations of expansion microscopy in chromatin ultrastructure determination. *Chromosom. Res.* **28**, 355–368 (2020).
198. Shi, X. *et al.* Label-retention expansion microscopy. *Preprint at bioRxiv* doi:10.1101/687954 (2021).
199. Wassie, A. T., Zhao, Y. & Boyden, E. S. Expansion microscopy: principles and uses in biological research. *Nature Methods* **16**, 33–41 (2019).
200. Alon, S. *et al.* Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. *Science* **371**, eaax2656 (2021).
201. Robinett, C. C. *et al.* In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J. Cell Biol.* **135**, 1685–1700 (1996).
202. Belmont, A. S. & Straight, A. F. In vivo visualization of chromosomes using lac operator-repressor binding. *Trends Cell Biol.* **8**, 121–124 (1998).
203. Lucas, J. S., Zhang, Y., Dudko, O. K. & Murre, C. 3D trajectories adopted by coding and regulatory DNA elements: First-passage times for genomic interactions. *Cell* **158**, 339–352 (2014).
204. Chen, H. *et al.* Dynamic interplay between enhancer–promoter topology and gene activity. *Nat. Genet.* **50**, 1296–1303 (2018).
205. Maass, P. G. *et al.* Spatiotemporal allele organization by allele-specific CRISPR live-cell imaging (SNP-CLING). *Nat. Struct. Mol. Biol.* **25**, 176–184 (2018).
206. Wang, H. *et al.* CRISPR-mediated live imaging of genome editing and transcription. *Science* **365**, 1301–1305 (2019).
207. Geng, Y. & Pertsinidis, A. CAS-LiveFISH: Simple and versatile imaging of genomic loci in live mammalian cells and early pre-implantation embryos. *Preprint at bioRxiv* doi:10.1101/2020.08.25.265306 (2020).
208. Neguembor, M. V. *et al.* (Po)STAC (Polycistronic SunTAg modified CRISPR) enables live-cell and fixed-cell super-resolution imaging of multiple genes. *Nucleic Acids Res.* **46**, e30 (2018).
209. Chu, L. *et al.* The 3D Topography of Mitotic Chromosomes. *Mol. Cell* **79**, 902-916.e6 (2020).
210. Liu, X. *et al.* In Situ Capture of Chromatin Interactions by Biotinylated dCas9. *Cell* **170**, 1028–1043 (2017).
211. Tsui, C. *et al.* DCas9-targeted locus-specific protein isolation method identifies histone gene regulators. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 2734–2741 (2018).
212. Myers, S. A. *et al.* Discovery of proteins associated with a predefined genomic locus via dCas9-APEX-mediated proximity labeling. *Nat. Methods* **15**, 437–439 (2018).
213. de Wit, E. *et al.* CTCF Binding Polarity Determines Chromatin Looping. *Mol. Cell* **60**, 676–684 (2015).
214. Gómez-Marín, C. *et al.* Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proc Natl Acad Sci U S A* **112**, 7542–7547 (2015).
215. Bantignies, F. *et al.* Polycomb-dependent regulatory contacts between distant hox loci in drosophila. *Cell* **144**, 214–226 (2011).
216. Monahan, K., Horta, A. & Lomvardas, S. LHX2- and LDB1-mediated trans interactions regulate olfactory receptor choice. *Nature* **565**, 448–453 (2019).
217. Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
218. Bhattacharyya, S., Chandra, V., Vijayanand, P. & Ay, F. Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nat. Commun.* **10**, 1–14 (2019).
219. Takei, Y. *et al.* Global architecture of the nucleus in single cells by DNA seqFISH+ and multiplexed immunofluorescence. *Preprint at bioRxiv* doi:10.1101/2020.11.29.403055 (2020).
220. Tolhuis, B., Palstra, R. J., Splinter, E., Grosveld, F. & de Laat, W. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol Cell* **10**, 1453–1465 (2002).
221. Palstra, R. J. *et al.* The β-globin nuclear compartment in development and erythroid differentiation. *Nat. Genet.* **35**, 190–194 (2003).
222. Splinter, E. *et al.* CTCF mediates long-range chromatin looping and local histone modification in the β-globin

locus. *Genes Dev.* **20**, 2349–2354 (2006).
223. Vakoc, C. R. *et al.* Proximity among distant regulatory elements at the β-globin locus requires GATA-1 and FOG-1. *Mol. Cell* **17**, 453–462 (2005).
224. Splinter, E. & De Laat, W. The complex transcription regulatory landscape of our genome: Control in three dimensions. *EMBO J.* **30**, 4345–4355 (2011).
225. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet.* **38**, 1341–1347 (2006).
226. Andrey, G. *et al.* A switch between topological domains underlies HoxD genes collinearity in mouse limbs. *Science* **340**, 1234167 (2013).
227. Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96–100 (2014).
228. Apostolou, E. *et al.* Genome-wide chromatin interactions of the nanog locus in pluripotency, differentiation, and reprogramming. *Cell Stem Cell* **12**, 699–712 (2013).
229. Van De Werken, H. J. G. *et al.* 4C technology: Protocols and data analysis. in *Methods in Enzymology* **513**, 89–112 (Academic Press Inc., 2012).
230. Bonev, B. & Cavalli, G. Organization and function of the 3D genome. *Nat. Rev. Genet.* **17**, 661–678 (2016).
231. Y, G.-H. *et al.* Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nat. Genet.* **51**, 1272–1282 (2019).
232. Lupiáñez, D. G. *et al.* Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012–1025 (2015).
233. Ogiyama, Y., Schuettengruber, B., Papadopoulos, G. L., Chang, J.-M. & Cavalli, G. Polycomb-Dependent Chromatin Looping Contributes to Gene Silencing during Drosophila Development. *Mol. Cell* **71**, 73–88 (2018).
234. Flavahan, W. A. *et al.* Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature* **575**, 229–233 (2019).
235. Tarjan, D. R., Flavahan, W. A. & Bernstein, B. E. Epigenome editing strategies for the functional annotation of CTCF insulators. *Nat. Commun.* **10**, 4258 (2019).
236. Kim, J. H. *et al.* LADL: light-activated dynamic looping for endogenous gene expression control. *Nat. Methods* **16**, 633–639 (2019).
237. Wang, H. *et al.* CRISPR-Mediated Programmable 3D Genome Positioning and Nuclear Organization. *Cell* **175**, 1405–1417 (2018).

Acknowledgements

We thank Marco Di-Stefano for critical reading of the computational approaches section. We thank Bernd Schuttengrueber and Frederic Bantignies for input on figures. We thank Quentin Szabo for help with and input on Figure 4. I.J. was supported by EMBO Long-Term Fellowship (ALTF 559-2018) and the Laboratory of Excellence EpiGenMed. Research in the G.C. laboratory is supported by grants from the European Research Council (Advanced Grant 3DEpi, under grant agreement No 788972), the European Union's Horizon 2020 research and innovation programme (MuG, under grant agreement No 676556 and ChromDesign, under the Martie Sklodowska-Curie grant agreement No 813327), the Agence Nationale de la Recherche (ANR-15-CE12-0006 EpiDevoMath), the Fondation pour la Recherche Médicale

(DEI20151234396), the MSDAVENIR foundation (Project GENE-IGH), the INSERM and the French National Cancer Institute (INCa).

Author contributions

The authors contributed equally to all aspects of the article.

Competing interests

The authors declare no competing interests.

Peer review information

Nature Reviews Molecular Cell Biology thanks Zhijun Duan; Bing Ren; Juan Vaquerizas, who co-reviewed with Elizabeth Ing-Simmons; and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

Publisher's note

Figure legends:

Figure 1 Multiple levels of genome organization and the methods to study them.

DNA folds at multiple scales (indicated in the left) to build chromosomes. DNA winding around histones forms nucleosomes, which are organized into clutches, each containing ~1–2 kb of DNA[116]. Nucleosome clutches form chromatin nanodomains (CNDs) of ~100 kb in size, where most enhancer–promoter (E–P) contacts take place[43,57,117]. At the scale of ~1 Mb, CNDs and CTCF–cohesin-dependent chromatin loops form topologically associating domains (TADs) [43,57,58,67,116,213,214]. On the higher scale of up to 100s of Mbs, chromatin segregates into gene active and inactive compartments (A and B, respectively) and into compartment-specific contact hubs (not shown). At the highest topological level, the nucleus is organised into chromosome territories[27,35]. Different techniques can be used to study different genome organization levels, and some techniques can be used to study different organization levels. Although Hi-C, genome architecture mapping (GAM) and other sequencing-based techniques can be used to detect chromosome territories, 3D fluorescent *in situ* hybridization (FISH) is most useful to study this level of organization, as it provides direct spatial information [7,27,41]. Compartments and hubs are usually studied with Hi-C, GAM, split-pool recognition of interactions by tag extension (SPRITE) and several multiplexed super-resolution

FISH techniques[14,15,123,27,35,41,44–48]. At the more functional, 1 Mb scale, most informative are capture Hi-C (cHi-C), Capture-C, micrococcal nuclease chromosome conformation assay (Micro-C) and super-resolution FISH approaches [21–26,91]. On the other hand, protein-driven enrichment techniques such as Hi-C chromatin immunoprecipitation (HiChIP), chromatin interaction analysis by paired-end tag sequencing (ChIA-PET), proximity ligation assisted chromatin immunoprecipitation (PLAC-seq) and chromatin-interaction analysis via droplet-based and barcode-linked sequencing (ChIA-Drop) can be used to study different levels of folding depending on whether the protein is associated with relatively local folding (e.g., MED12 in E–P contacts), more long-range contacts (e.g., Polycomb proteins) or inter-chromosomal hubs (e.g., LHX2, LDB1) [31–33,102,192,215–218]. CNDs have been discovered only recently. Owing to their stochastic nature and to inter-cell variability, only super-resolution microscopy FISH has so far been able to detect CNDs[43,57,117]. DNA adenine methyltransferase identification (DamID), Genomic loci positioning by sequencing (GPSeq), Tyramide signal amplification (TSA).

Figure 2 Main C-based methods for interrogation of 3D genome organization.

For the application of every chromosome conformation capture (C)-based method, chromatin must first be cross-linked, either with one cross-linker (in most methods) or with two cross-linkers in the case of micrococcal nuclease C (Micro-C) and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET)[85,91,102]. Hi-C derivatives then use digestion by restriction enzymes (RE) to fragment the chromatin, fill-in fragment ends with biotin and perform ligation, all *in situ*. The sample is then sonicated, de-crosslinked and enriched for informative fragments through biotin pulldown; the resulting chromatin-fragments library is subjected to amplification and sequencing [85]. Hi-C chromatin immunoprecipitation (HiChIP) differs from this standard protocol by introducing an immunoprecipitation (enrichment) step just after ligation, whereas capture Hi-C (cHi-C) includes an enrichment step involving hybridization to RNA baits that represent a genomic region of interest, followed by pulldown that is performed on the final Hi-C library [26,32]. The resulting libraries will therefore be enriched either for all genomic contacts (in the case of Hi-C), for chromatin contacts at genomic regions where the protein of interest binds (HiChIP), or for contacts at a specific region of interest (cHi-C). Micro-C follows an almost identical procedure as Hi-C, with distinctions in the steps of fixation and of digestion, in which the restriction enzymes used in the Hi-C protocol are replaced with micrococcal nuclease (MNase)-mediated chromatin fragmentation[91]. This digestion modification results in a more uniform fragmentation of the genome, thereby allowing Micro-C to achieve higher resolution of local contacts. ChIA-PET provides similar information to HiChIP, but in ChIA-PET immunoprecipitation is performed immediately after fixation and sonication; and while the sample is

still on the pull-down beads, a linker with biotin is added and ligation is performed on-beads in order to reduce the amount of random ligation products[102]. Concomitant with adapter addition, the sample is de-crosslinked and fragmented using a transposase (Tn5). Finally, the sample is enriched for informative fragments through biotin pulldown and undergoes library amplification and sequencing. DSG, disuccinimidyl glutarate; EGS, ethylene glycol bis(succinimidyl succinate).

Figure 3 Main ligation-independent methods for interrogation of 3D genome organization.

Ligation-independent methods have been developed in order to study multi-way contacts that are inaccessible to ligation-based methods. The most common ligation-independent techniques are split-pool recognition of interactions by tag extension (SPRITE), chromatin-interaction analysis via droplet-based and barcode-linked sequencing (ChIA-Drop) and genome architecture mapping (GAM)[35,41,192]. In all three, nuclei are fixed. Then, SPRITE and ChiA-Drop proceed with sonication and chromatin digestion by DNase I. From this point on, the two techniques take advantage of different approaches to retaining information on multi-way contacts. SPRITE utilizes a split-and-pool strategy, in which every sample is split, barcoded and pooled together five times. This results in unique barcoding of all fragments that crosslinked together and thus to the identification of DNA sequences that were involved in the same multi-way contacts [35]. ChIA-Drop uses microfluidics to produce a droplet carrying a unique barcode, adapters and material for DNA amplification reactions used to label a single chromatin-interaction knot, thereby allowing the identification of all DNA sequences that have been crosslinked together. GAM utilizes a completely different strategy to assay genome architecture and is suitable for investigating multi-way contacts, higher-order chromatin structures as well as more local contacts. In GAM, the sample is fixed, embedded in sucrose and cryo-sectioned to obtain thin slices, from which individual nuclear slices are laser micro-dissected. Genomic DNA is then extracted from a single-nucleus slice followed by whole-genome amplification and sequencing. The data obtained from the different sections of a nucleus are pulled together and interactions are identified as DNA sequences that co-segregate more often than others.

Figure 4 Microscopy and FISH-based methods for 3D genome investigation.

a) Super-resolution microscopy techniques overcome the diffraction limit using different approaches. In structured illumination microscopy (SIM), the sample is exposed to a series of non-uniform illumination from different angles and axial phases [107–109]. The resulting light pattern is analysed by Fourier transformation to achieve a final mathematical reconstruction of the image, which improves lateral and axial resolution by two-fold. Single molecule localization microscopy (SMLM) uses low excitation energy

that causes a stochastic excitation of photoswitchable fluorophores and allows for the precise localization of the centre of emission [112–115]. Sequential images of the sample are taken, in which fluorophores turn either bright or dark and the final image is created by a superposition of all imaging cycles. In most practical applications, this method yields a lateral resolution of up to 20 nm. Stimulated emission depletion (STED) uses stimulated emission depletion through the combination of two lasers: an excitation laser illuminates the sample in the middle and a doughnut-shaped depletion beam depletes the surrounding signal [118,119]. In practice, STED reaches lateral resolution of about 50 nm and axial resolution of about 80 nm to 600 nm. **b)** Oligopaints are fluorescently labelled synthetic DNA oligonucleotides that can be combined with fluorescence *in situ* hybridization (FISH) to label, visualize and measure the distances between genomic regions [122]. **c)** Oligopaints can be further modified to allow sequential imaging by incorporating sample bleaching or automatized microfluidics with microscopy. Following every round of imaging, the oligonucleotides are washed out before proceeding with a new round of hybridization and imaging[44–48,123]. Distant genomic regions are imaged simultaneously using different detection fluorophores while their neighbouring regions are imaged in the next round, thereby allowing chromatin tracing. **d)** In oligo fluorescence *in situ* sequencing (OligoFISSEQ), oligonucleotides contain barcodes that can be read through hybridization with a set of specific primers, to which fluorescently labelled dinucleotides are ligated. The fluorescent signal is then imaged and cleaved off. This process is repeated until the barcode is read in full (inferred from the specific combination of fluorophores) [14,15,219] gDNA, genomic DNA; PSF, point spread function; SPDM, spectral position determination microscopy.

Figure 5. Computational modelling of the 3D genome.

Two main computational strategies have been devised to investigate chromosome folding and the mechanisms driving it: a bottom-up strategy (polymer modelling) and a top-down strategy (restraint-based modelling) [137,157,158]. **a)** Polymer modelling mimics the physical behaviour of the chromatin fibre that is represented as series of monomers (beads) on a string and in which attraction or repulsion forces can be attributed based on first principles. The behaviour of the polymer is influenced by non-specific restraints, e.g. imposing that the beads must be self-avoiding, by specific restraints, such as cohesin-dependent loop extrusion that is blocked at convergent CTCF sites, and by specific interactions, such as attraction between beads sharing the same epigenetic modifications. Such strategy can produce virtual contact maps that can be compared with Hi-C interaction matrices. This process is reiterated until the input parameters can reconstitute a simulation that optimally recapitulates experimental maps. Polymer modelling allows to infer the mechanisms and estimate the forces that are necessary or sufficient to

achieve the chromatin conformation closest to the experimental data[159]. Beads of different colours indicate regions carrying different epigenetic modifications **b)** Restraint-based modelling uses experimental data such as Hi-C contact maps or fluorescent in situ hybridization (FISH)-measured nuclear distances to infer all the spatial restraints necessary to determine the structure of the genomic region of interest. The resulting model is an optimally reconstructed 3D folded chromatin fibre that gives information on spatial positioning of chromatin regions.

Figure 6. Live microscopy using CRISPR–dCas9 to study the 4D genome.

Live-cell microscopy is one of the few approaches that informs on the dynamics of chromatin contacts. Multiple different methods allow the study of chromatin in 4D (changes of 3D chromatin structure over time) [50,204–207]. Here, we illustrate two such methods based on nuclease-dead Cas9 (dCas9) recruitment to the chromatin: chimeric array of gRNA oligonucleotides (CARGO) and Live FISH. **a)** CARGO utilizes a single plasmid encoding multiple guide RNA (gRNAs) targeting the same genomic region of interest, for example an enhancer, in order to amplify the fluorescence provided by dCas9-EGFP [49]. The region of interest can be imaged over time to study its spatial dynamics. **b)** Live FISH utilizes two fluorescently labelled gRNAs (one red (Cy3) and one green (A488)) targeting different regions of interest [206]. The two regions can be labelled and imaged simultaneously, which allows tracking the dynamics of the regions over time. D, distance.

Box 1. Overview of proximity-ligation-based methods for mapping chromatin interactions

The development of chromosome conformation capture (3C), which detects pair-wise interactions between select loci ('one-vs-one') through nuclear-proximity ligation in combination with semi-quantitative PCR, marks the onset of the eponymous C-based techniques in the early 2000s[17,18,29]. Using 3C, the locus control region of the β-globin locus was shown for the first time to form chromatin loops with and thus to activate its promoter, and to form an active chromatin hub that dynamically follows transcription during differentiation and is stabilized by transcription factors [220–223]. However, 3C is low-throughput and cannot successfully detect long-distance contacts. Circular 3C (4C) overcame these limitations by using primers in order to detect genome-wide contacts formed by a single 'viewpoint' to (one-vs-all) [19,224,225]. Later, 4C was combined with next generation sequencing , and was used to describe the dynamics of chromatin contacts during development [226–229]. However, the most influential technique in 3D genome organization research is Hi-C, in which the DNA interactome of the entire genome is assayed (all-vs-all) [27]. Hi-C led to the identification of genomic compartments and topologically associating

domains and to the development of the loop-extrusion model [27,53,130,54,55,58,59,61,64,67,85]. Finally, several techniques were developed, which combined Hi-C with chromatin immunoprecipitation, thereby allowing the interrogation of chromatin contact frequencies based on the presence of a specific protein [31–33]. It is important to note that in all C-based techniques that include protein enrichment, the mapped contacts are probabilistic rather than deterministic features and it is impossible to predict how and whether these contacts will translate into function.

Together, proximity-ligation based techniques fueled most of the discoveries in 3D genome organization research during the past fifteen years. The ease of application of these techniques is anti-correlated with the richness of data obtained from them, which is likely the reason why so many different C-based technique adaptations exist. With the recent development of microscopy-related techniques, the proximity ligation-based data is even more valuable as it will offer an imaging-complementary information that is invaluable for achieving a better understanding of genome folding.

Box 2 Manipulation of genome architecture

A fundamental question that has been very difficult to address in the past is whether genome architecture changes that are detected in different cellular conditions are a cause or a consequence of changes in gene expression [65,230]. Use of Hi-C in *Drosophila melanogaster* lines that carry a set of known genomic alterations has allowed to address this question. Intriguingly, genome topology changes stemming from chromosomal inversions or other mutations can be buffered to a large extent, such that gene expression remains robust in most, although not all genes[231]. Whereas this analysis used pre-existing mutant lines, CRISPR–Cas technology has also been used extensively to manipulate genome architecture. Inducing specific mutations at critical genome architecture regulatory regions has shown that genome organization into topologically associating domains with specific boundaries contributes to correct cell-type and tissue-specific gene regulation[213,232]. In *D. melanogaster*, the deletion of specific chromatin-loop anchor sequences or the insertion of boundary elements that prevent loop formation, showed that Polycomb-dependent genomic loops can contribute to gene silencing during development [233]. The same approach showed that CTCF-binding sites are required for correct insulation of gene expression and that their deletion activates an oncogenic gene-expression program [234].

CRISPR–Cas was used not only to edit the genome, but also the epigenome, as in the case of using a fusion of nuclease-dead Cas9 (dCas9) with DNA (cytosine-5)-methyltransferase 3A, which targets DNA methylation to specific regions and displaces CTCF binding[235]. Another powerful experimental tool combines CRISPR–Cas with optogenetics to induce chromatin looping upon stimulation with blue light and

study its functional consequence[236]. Finally, the CRISPR-genome organization (CRISPR-GO) system enables inducible and reversible repositioning of dCas9-targeted genomic regions in the nucleus; CRISPR-GO was shown to reposition a locus of interest to the nuclear lamina, thereby perturbing its function, and to Cajal bodies **[G]** and PML bodies [237]. This system could be used to target loci to other nuclear compartments in order to study the functional consequences of their relocation. Combined with the many experimental tools that enable measuring the effects of perturbation of genome structure and function, these techniques will be crucial for differentiating between cause and consequence of 3D genome organization and gene regulation.

GLOSSARY

First principles

Basic building blocks of knowledge that cannot be deduced from any other preposition used for mathematical modeling of polymer behaviors.

Airy diffraction pattern

A diffused circle surrounded by rings of decreasing intensity generated when a laser passes through a circular opening.

Cajal bodies

Nuclear bodies of 0.3 to 1μm in size, containing RNAs and proteins and involved in RNA-related metabolic processes.

Dendrimer crosslinking

A procedure in which formaldehyde crosslinking can be followed or replaced by crosslinking with dendrimers, i.e. highly ordered, branched polymeric molecules of different sizes.

Diffraction limit

The points where two Airy patterns are too close to be distinguishable.

Lamina associated domains

Chromosome domains associated to the nuclear lamina in the 3D nuclear space.

Loop extrusion model

Model suggesting that motors such as cohesin or condensin form a ring around chromatin and use the energy of ATP to slide through it while extruding the intervening region.

Multi-way contacts

Chromatin contacts involving more than two chromatin fragments.

Nuclear speckles

Nuclear foci enriched in pre-mRNA splicing factors located in the nucleoplasm of eukaryotic cells.

Point spread function (PSF)

The response of an imaging system to a point object. If the object is below the microscope resolution it will appear larger than it really is..

PML bodies

Nuclear bodies of 0.1 to 1μm in size, containing many components, including the promyelocytic leukemia protein (PML) and frequently associated to Cajal bodies.

Polycomb

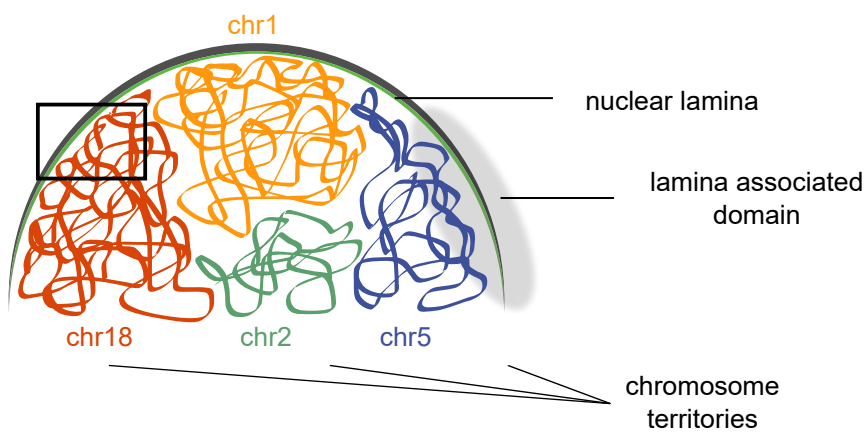
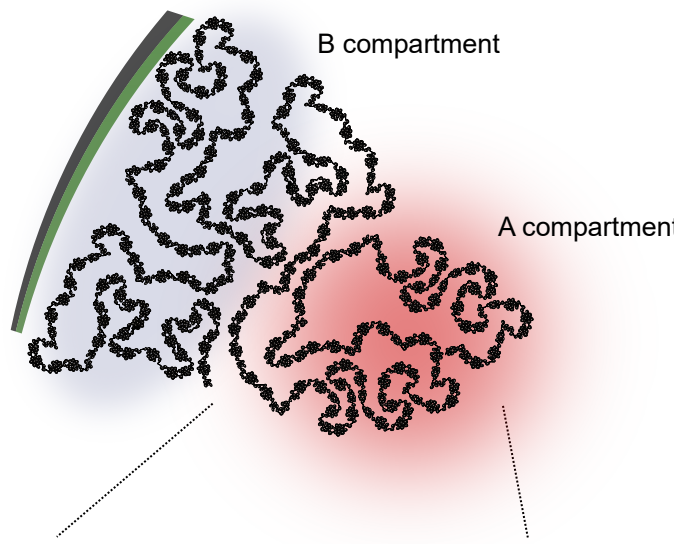
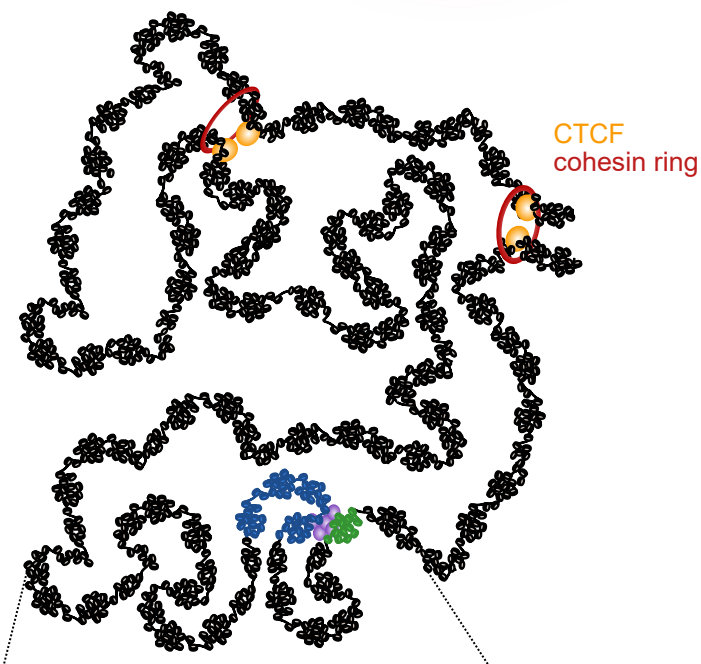
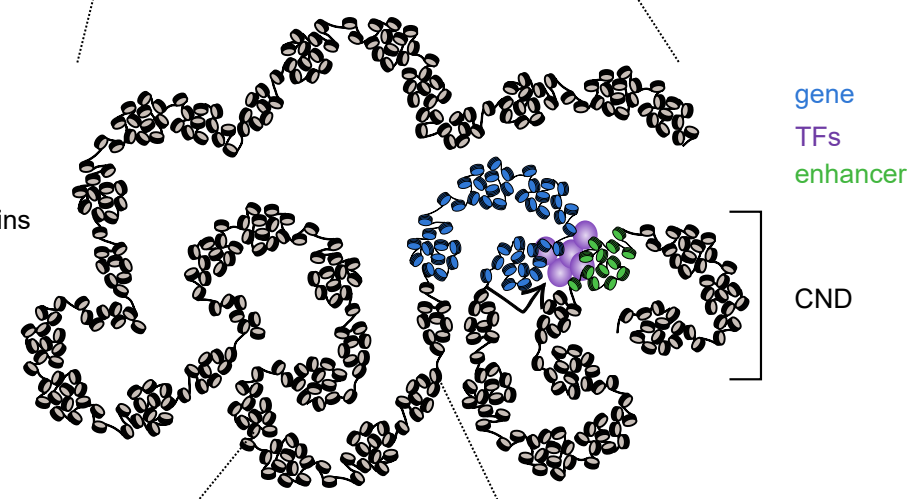
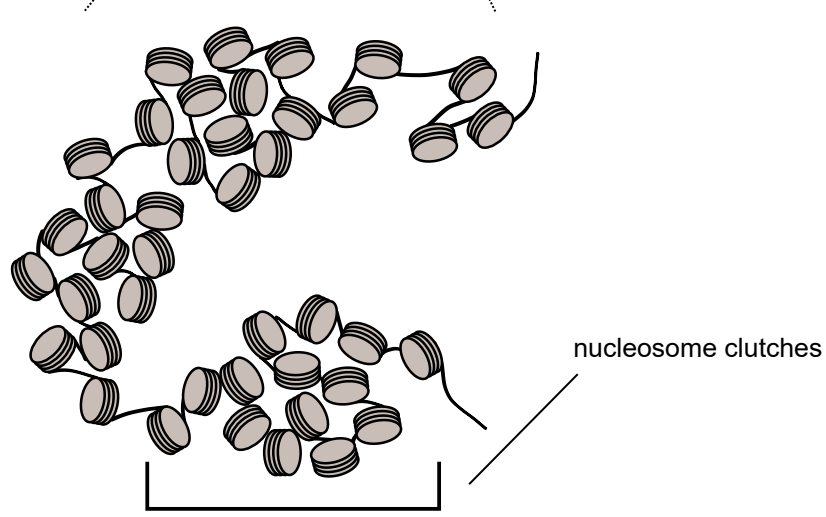
An evolutionarily conserved group of proteins involved in the regulation of a large group of target genes.

Sub-diffractive point spread function

A point spread function of smaller size than that generated by diffraction-limited systems.

Tyramide signal amplification

A method enabling sensitive detection of low-abundance molecules in fluorescent immunocytochemistry applications.

			METHODS		
			NGS		MICROSCOPY
			enrichment	non-enrichment	
entire chromosomes	chromosome territories			GAM GP-seq Hi-C	3D FISH
up to 100s Mbs	compartments hubs		DamID TSA-seq	GAM Hi-C Micro-C GP-seq SPRITE	oligopaints multiplexed super-resolution FISH electron microscopy
100 kb - few Mb	topologically associating domains (TADs) loops		Capture-C cHi-C 4C	GAM SPRITE Hi-C Micro-C	oligopaints multiplexed super-resolution FISH
10 - 100 kb	chromatin nanodomains (CNDs) functional loops (E-P contacts ...)		Capture-C cHi-C 4C	GAM Micro-C	oligopaints multiplexed super-resolution FISH
<10 kb	nucleosome clutches			Micro-C	oligopaints multiplexed super-resolution FISH electron microscopy

Nucleus

Micro-C

ChIA-PET

Formaldehyde fixing

Formaldehyde & DSG fixing

Formaldehyde & EGS fixing

RE digestion

MNase digestion

Sonication & Immunoprecipitation

biotin

biotin

biotin linker

Biotin fill-in & ligation

Biotin fill-in & ligation

Biotinilated linker addition & on-bead ligation

in-situ

in-situ

HiChIP

Hi-C

cHi-C

biotin

Sonication & Immunoprecipitation

Sonication & de-crosslinking

Sonication & de-crosslinking

De-crosslinking

Decrosslinking & biotin pull-down

Biotin pull-down

Biotin pull-down

Tn5 tagmentation & adapter addition

HiChIP library amplification

HiC library amplification

HiC library amplification

Biotin pull-down

RNA bait pulldown

cHiC library amplification

ChIA-PET library amlification

Sequencing

Sequencing

Sequencing

Sequencing

Sequencing

Genome wide

Genome wide

Region of interest

Genome wide

Genome wide

Cross-linked nuclei

Sonication & DNaseI digestion

SPRITE

Add barcodes round 1

Add barcodes round 2

Split and pool to add barcodes n=5

Sequential barcoding

Sequencing

Genome wide + multi-way & higher order contacts & hubs

ChIA-Drop

Microfluidics based individual molecule barcoding

Sequencing

Genome wide + multi-way & higher order contacts

GAM

Cross-linked cells

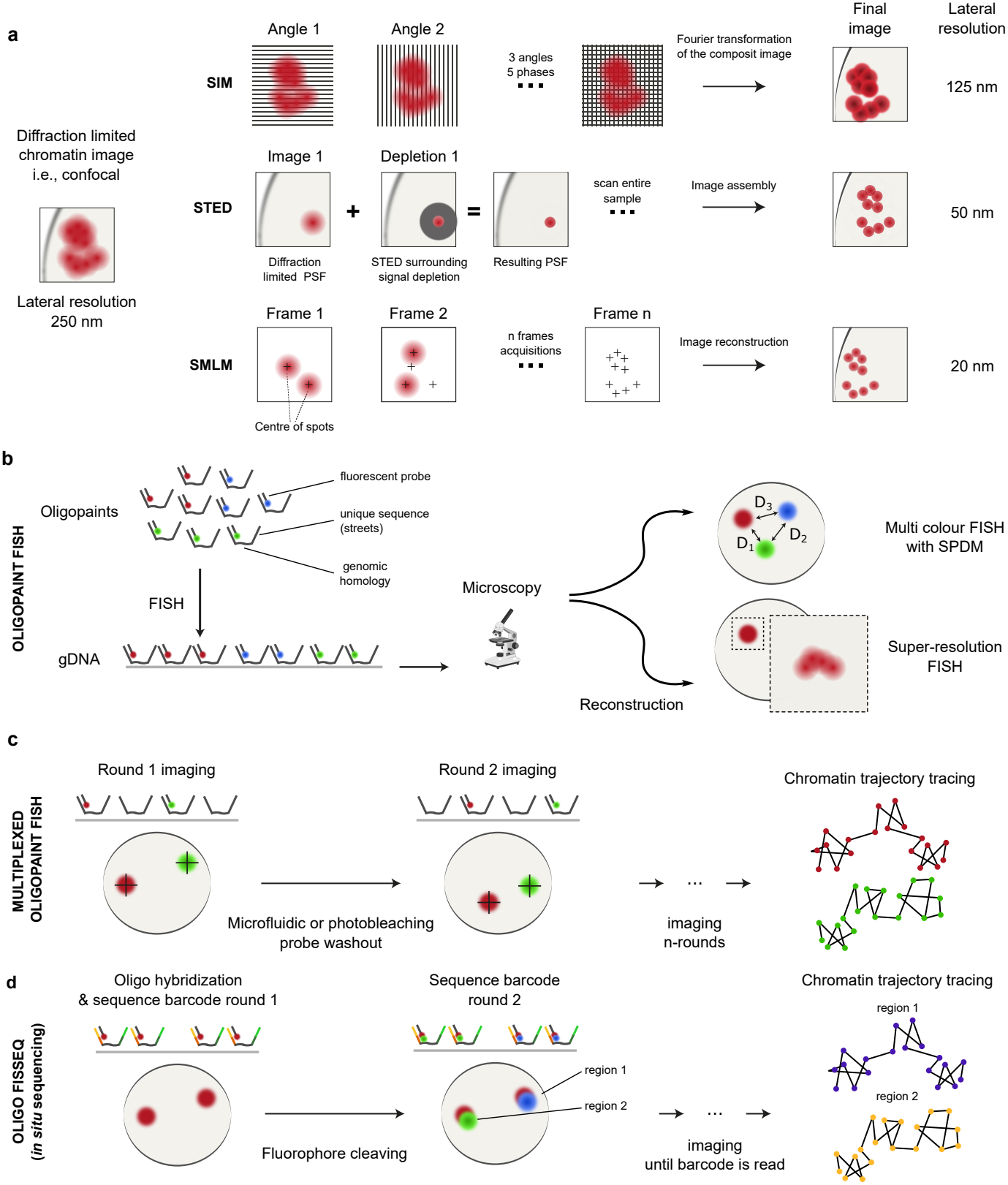
Cryo sectioning

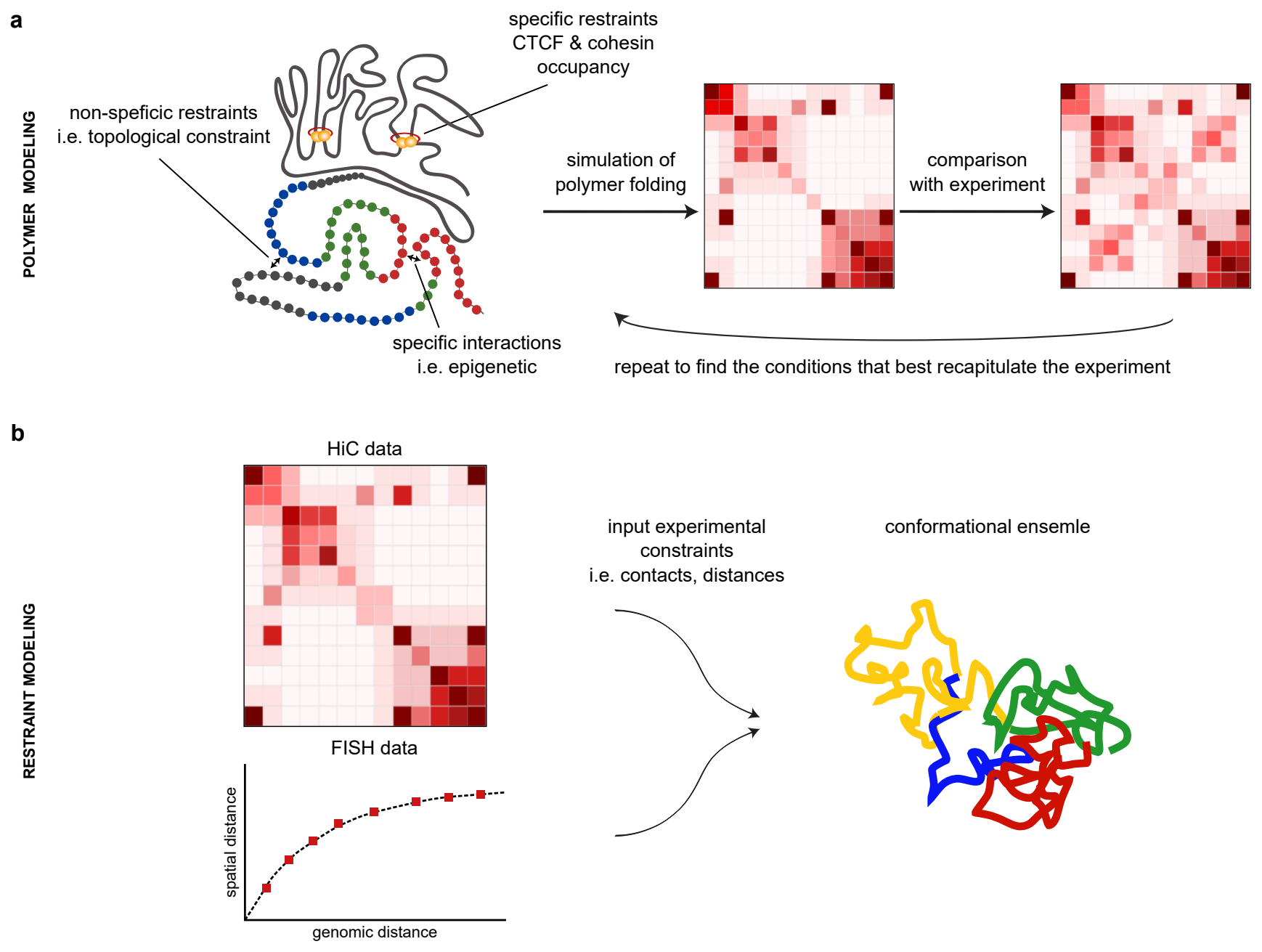
Individual nuclei laser microdissection

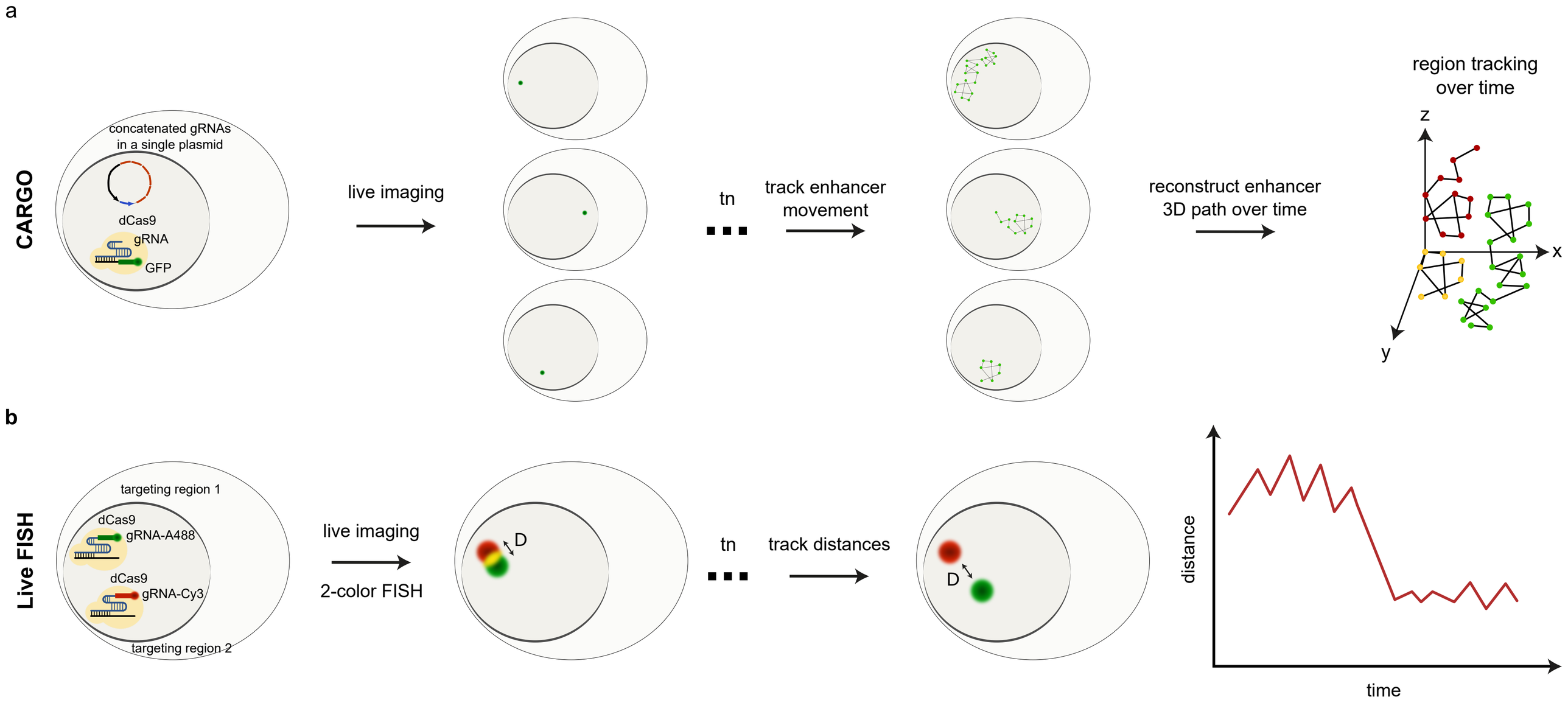
Individual nuclei whole genome amplification & brcoding

Sequencing

Genome wide + multi-way & higher order contacts


a
Angle 1
Angle 2
Final image
Lateral resolution
SIM
3 angles
5 phases
Fourier transformation of the composit image
125 nm
Diffraction limited chromatin image i.e., confocal
Image 1
Depletion 1
STED
scan entire sample
Image assembly
50 nm
Diffraction limited PSF
STED surrounding signal depletion
Resulting PSF
Lateral resolution 250 nm
Frame 1
Frame 2
Frame n
SMLM
n frames acquisitions
Image reconstruction
20 nm
Centre of spots
b
OLIGOPAINT FISH
Oligopaints
fluorescent probe
unique sequence (streets)
genomic homology
FISH
gDNA
Microscopy
D3
D1
D2
Multi colour FISH with SPDM
Super-resolution FISH
Reconstruction
c
MULTIPLEXED OLIGOPAINT FISH
Round 1 imaging
Round 2 imaging
Chromatin trajectory tracing
Microfluidic or photobleaching probe washout
imaging n-rounds
d
OLIGO FISSEQ (in situ sequencing)
Oligo hybridization & sequence barcode round 1
Sequence barcode round 2
Chromatin trajectory tracing
region 1
region 2
Fluorophore cleaving
imaging until barcode is read


a
POLYMER MODELING
specific restraints
CTCF & cohesin
occupancy
non-speficic restraints
i.e. topological constraint
specific interactions
i.e. epigenetic
simulation of
polymer folding
comparison
with experiment
repeat to find the conditions that best recapitulate the experiment
b
RESTRAINT MODELING
HiC data
FISH data
spatial distance
genomic distance
input experimental
constraints
i.e. contacts, distances
conformational ensemle


a
CARGO
concatenated gRNAs
in a single plasmid
dCas9
gRNA
GFP
live imaging
tn
track enhancer
movement
reconstruct enhancer
3D path over time
region tracking
over time
z
x
y
b
Live FISH
targeting region 1
dCas9
gRNA-A488
dCas9
gRNA-Cy3
targeting region 2
live imaging
2-color FISH
D
tn
track distances
D
distance
time