Advanced Review

3D modeling of chromatin structure: is there a way to integrate and reconcile single cell and population experimental data?

François Le Dily,[1,2*] François Serra[1,2,3] and Marc A. Marti-Renom[1,2,3,4*]

The genome is organized in a hierarchical fashion within the nucleus in interphase. This nonrandom folding of the chromatin fiber is thought to play important roles in the processing of the genetic information. Therefore, a better knowledge of the mechanisms underlying the three-dimensional structure of the genome appears essential to fully understand the nuclear processes including transcription and replication. Fluorescent *in situ* hybridization (FISH) and molecular biology methods deriving from the Chromosome Conformation Capture technique are the methods of choice to study genome 3D organization at different levels. Although these single cell and population methods allowed to highlight similar chromatin structures, they also show frequent discrepancies which might be better understood by improving the capacity to generate actual 3D models of organization based on the different types of data available. This review aims at giving an overview of the principles, advantages, and limits of microscopy and molecular biology methods of analysis of genome structure and at discussing the different approaches of modeling of chromatin classically used and the improvements that are necessary to reach a better understanding on the links between chromatin structure and its spatial organization. © 2017 John Wiley & Sons, Ltd

How to cite this article:
WIREs Comput Mol Sci 2017, e1308. doi: 10.1002/wcms.1308

INTRODUCTION

Processing the genomic information in the nucleus of eukaryotic cells requires regulatory proteins to access DNA assembled into chromatin. Although the chromatin fiber is highly packed within the nuclear space, its density is heterogeneous. For example, looser chromatin regions tend to occupy the center of the nucleus, whereas compact chromatin is enriched close to the nuclear periphery and to the nucleolus. Moreover, it is now evident that the organization of the genome in the nucleus is generally hierarchical and nonrandom: chromosomes are organized as territories occupying preferential positions depending on their gene-density, with gene-rich small chromosomes frequently located at the center of the nucleus and larger gene-poor chromosomal domains tending to occupy the nuclear periphery.[1,2] Within chromosomes, the chromatin fiber is further segmented in mega-base-sized domains where loci interact more frequently between them inside the domain than with the neighboring regions. These so called topologically associated domains (TADs) are distributed in different chromatin compartments depending on their epigenetic state.[3–5] The organization in TADs is thought

*Correspondence to: francois.ledily@crg.eu; martirenom@cnag.crg.eu

[1]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona, Spain

[2]Universitat Pompeu Fabra (UPF), Barcelona, Spain

[3]Structural Genomic Group, CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Baldiri Reixac 4, Barcelona, Spain

[4]Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluis Companys 23, Barcelona, Spain

Conflict of interest: The authors have declared no conflicts of interest for this article.

to favor the establishment of regulatory loops between enhancers and promoters, thus being involved in the regulation of transcription and in the specificity of cell transcriptomes.[6] Although globally organized, this three-dimensional (3D) organization has yet strong dynamic and stochastic components blurring the general patterns and their correlation with nuclear processes.[7] Nonetheless, characterizing how the genome is folded in the nucleus is essential to determine how it can constrain or regulate the nuclear processing of DNA including transcription as well as replication and repair.[8,9]

Two main approaches allow gaining insight on chromatin conformation and organization of chromosomal domains: direct visualization of those structures in the nucleus of cells by microscopy-based methodologies, in particular Fluorescent *in situ* hybridization or FISH[10] and molecular biology methods, principally derived from the Chromosome Conformation Capture or 3C.[11] Due to the 3D nature of genome organization, it appears that establishing actual 3D models greatly helps in understanding their functional implication and in predicting the consequences of changes in the conformation. Increasing attempts are made to generate such models from the analysis of local region of chromatin,[12] chromosomal domains or full chromosomes,[13] as well as complete genome in the confined nuclear space.[14] These 3D modeling approaches are of several types based on different methodologies and assumptions. However they generally use as input data coming from 3C-derived methods thus bringing limitations due to the population nature of these experiments. Indeed, and in contrast to FISH that allows single cell analysis, 3C-derived results generally consist of a superimposed picture of structures present in a population of cells. And this methodological difference can lead to discrepancies between observations based on FISH and observations based on 3C. Although recovering the natural heterogeneity of genome structure from 3C-based dataset is a challenge, we speculate that accurate modeling could resolve these discrepancies between FISH and 3C. Future implementations of 3D modeling should account for the population aspect of a given experiment and on the structural variability inherent to the chromatin.

In this review, after reminding the principles, advantages, and limitations of FISH and 3C-based approaches, we will discuss the differences between these two approaches. We will further give an overview of the principles of the main 3D modeling methods used to reconstruct genome conformation and we will focus on how, by integrating data from various origins, integrative modeling shall permit to obtain a clearer view of the relation between structure and function of the genome.

SINGLE CELL AND POPULATION ANALYSIS OF THE 3D GENOME ARCHITECTURE

Microscopy-Based Single Cell Methods

FISH allows the visualization of specific DNA loci within the nucleus of fixed cells by the use of complementary probes labeled with modified nucleotides.[15] This technique has been widely used to study the organization of the genome at very different scales covering from chromosomes, to specific pairs of loci (Figure 1). FISH notably permitted to validate the existence of chromosome territories.[16] Moreover, closer analysis showed that chromatin domains within those territories could further exhibit differences in condensation or in relative localization, which may mainly depend on their gene content.[17,18] Additionally, FISH permitted, for example, to elucidate the reorganization of gene positioning during the process of differentiation[19] and to demonstrate the propensity of related genes to be transcribed in

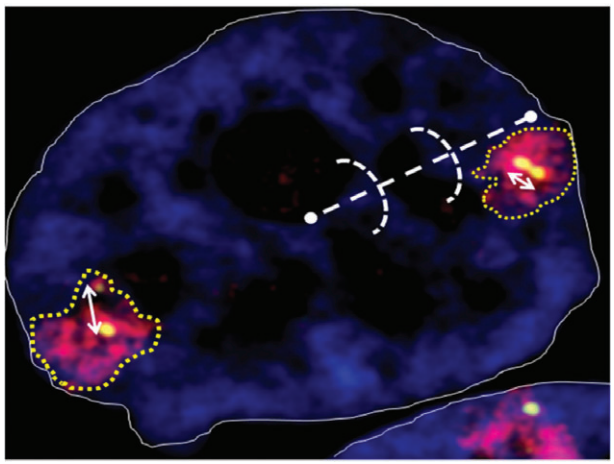

FIGURE 1 | Use of FISH to study the spatial organization of the genome in single cell. Fluorescent *in situ* hybridization (FISH) can be performed on fixed cells to analyze the different parameters of the spatial organization of chromosomes or genomic domains. For example, whole chromosome painting can be used to study the position of chromosome territories relative to nuclear structures as well as their radial positioning from nuclear center to periphery (dashed lines—in this example, human chromosome 2 was labeled in red). Multiple specific genomic regions can be labeled in parallel to measure the distances that separate them from each other (double arrows—in this example, two different loci of human chromosome 2 were labeled in green).

© 2017 John Wiley & Sons, Ltd

transcription factories.[20] The applicability of FISH is, however, limited by the sensitivity of fluorescently labeled probes or antibodies and by the microscope resolution. This limits the size of the loci to be probed as well as the physical distances that can be resolved. The emerging of high-resolution microscopy methods as well as the development of new labeling strategies are now overcoming these limitations.[21–24] For example, DNA-paint, which uses oligonucleotide-based 'Oligopaint' probes, has been recently used to analyze the folding of chromatin domains with different epigenetic states at high resolution.[22] These single cell methods were mainly limited to the analysis of genomic loci in fixed cells, and were not showing the dynamic nature of the chromatin fiber. This limitation has been overcome with the development of genome editing methods and biochemical tools allowing to track specific fluorescently labeled DNA region in living cells.[25]

As most single cell approaches, microscopy images are strongly affected by the cell-to-cell variability of the genome. Indeed, despite common rules of organization, the intrinsic dynamic properties of the chromatin fiber as well as the stochastic nature inherent to each of its level of organization are clearly highlighted by the observed variations in most of FISH-based studies. Therefore, establishment of models of organization based on FISH data requires the analysis of large number of cells. This need of higher throughput has invariably resulted in the development of methods to facilitate the automation of both the acquisition of images and their analysis.[26] However, even allowing the concomitant study of several thousands of cells, such methods remain limited in terms of throughput and resolution in comparison with molecular biology techniques.

Molecular-Based Population Methods

In parallel to the microscopy-based methods described above, the knowledge on genome organization was highly benefited from molecular biology techniques. In particular from the derivatives of the original Chromosome Conformation Capture (3C) assay was developed 15 years ago by Dekker et al.[11] These 3C methods are based on the cross-link of chromatin with formaldehyde, which bridges together DNA loci found in close spatial localization within the cell nucleus (Figure 2(a)). The frequency of cross-link between two DNA fragments is proportional to the frequency of contact within a population of cells, and is thus assumed to reflect an average physical distance in the nucleus. After chromatin cross-linking, DNA is further digested with restriction enzymes recognizing target sequences of 4 or 6 bases. For example, digestion of the human genome by these so called 4 or 6 bp-cutters restriction enzymes results in DNA restriction fragments of around 0.4 or 4 kb average length, respectively. The generated DNA ends that have been cross-linked together are further ligated and the ligation events are analyzed (Figure 2(b)). The original 3C method allowed the analysis of contacts between two given loci using primers designed to amplify regions close to the restriction sites of interest. It therefore limited the analysis to few thousands of bases around a given locus. Further implementations allowed analyzing the structure of larger domains or, more recently, assessing the interactions in a genomic context by coupling the experiment with high-throughput sequencing (Figure 2(b)). The most used 3C derivative methods are (Figure 2(b)): (1) 4C, which allows to interrogate the interactions of a locus with the rest of the genome by the use of inverted PCR primers[27]; (2) 5C which uses multiplexed primers covering from hundreds of kilobases to several mega-bases, and therefore allows to study interactions within a limited region of a genome[28]; and (3) Hi-C where the digested DNA fragments are biotin labeled at their restriction ends, allowing their selection after ligation and permitting the use of the selected material to high-throughput sequencing.[4] In this last case, the ligation step is performed, either in diluted conditions after partial cell lysis as in the original method[4] or directly in intact nuclei—also called *in situ* or *in nuclei* Hi-C -.[29] Although each 3C-derived method has specific advantages depending on the type of analysis required, Hi-C is nowadays becoming the method of choice. The resolution of the Hi-C, which aims at an unbiased detection of all interactions between any pair of restriction fragments in the genome, depends on the sequencing depth. Indeed, identifying all interactions at the fragment-based resolution using Hi-C would require an immense sequencing effort. Hi-C results are thus classically represented as matrices of pairwise contacts resulting from the binning of observed ligation events at a given resolution (Figure 3). Recently, high coverage Hi-C maps have been generated at 1-kb resolution[29] but more frequently, they range from 5 to 100 kb.[3,30]

These molecular methods confirmed several levels of organization already observed by FISH as the organization of chromosomes in territories occupying preferential and distinct positions in the nucleus (Figure 3). Also *de novo* identified the preferential segregation of chromatin domains[4] in compartments (Figure 3) and the existence of a functional segmentation of chromosomes in self-interacting Mb-sized domains called TADs (Figure 3).[3,5,30,31] In addition,

© 2017 John Wiley & Sons, Ltd

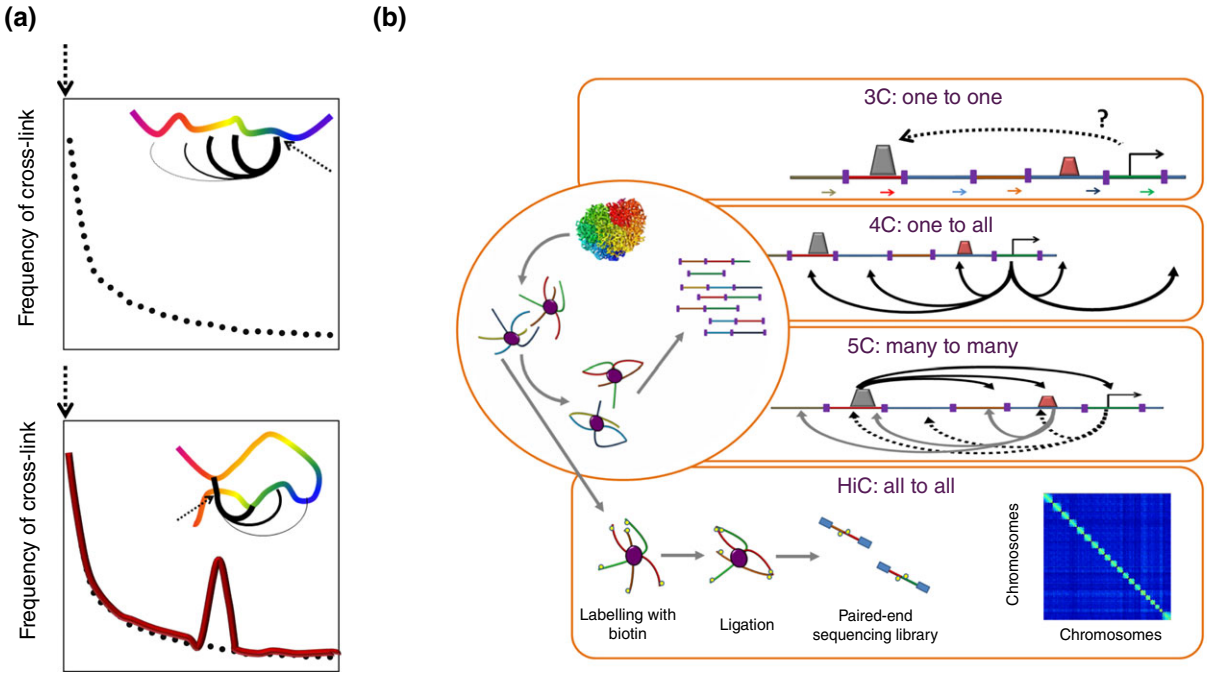

FIGURE 2 | Principle of Chromosome Conformation Capture and derivative methods. (a) 3C-based methods are based on the cross-link of chromatin with formaldehyde. The probability of cross-link between two regions is dependent on the spatial distance that separates them. In the case of a randomly organized molecule (top panel), the frequency of cross-linking will decrease as a function of the genomic distance that separate two loci whereas this relationship is broken in the case two region are establishing preferential contacts (bottom panel). (b) 3C-based methods use the frequency of ligation of cross-linked DNA after digestion by restriction enzyme (RE) to recover this information in a quantitative output. If the original 3C methods uses specific primers to detect one-to-one those potential interactions, further derivative allows to analyze the interactions at higher levels, either one-to-many/all (4C), many-to-many (5C), or all-to-all (Hi-C) depending on the detection method used.

high-resolution analysis demonstrated the existence of specific loops between regulatory sites, enhancer and promoters.[29,30] In summary, 3C-derived methods are highly resolutive methods to study the organization of the genome at different scales. Although the resulting frequencies of contacts between loci is assumed to reflect their spatial distance in the nucleus, it is important to keep in mind that this final static picture of contacts actually represents the average of events in the cell population. Attempts have been done to question the heterogeneity and commonality of these contacts by establishing single cell Hi-C methods.[32] Such analysis confirmed observations made at the population levels but also highlighted the stochastic and dynamic nature of these contacts.[32]

Agreements and Discrepancies between FISH and 3C-Based Results

As mentioned above, Hi-C experiments and FISH are both able to demonstrate the organization of chromosomes into territories as well as their preferential localization within the nuclear space. Additionally, Hi-C and 5C experiments demonstrated that chromosomes are segmented in TADs where the interactions are more frequent between loci within the domain rather than with the neighboring domains.[3,5,30,31] The use of FISH probes for labeling either loci located within a single TAD or within contiguous TADs confirmed their spatial isolation.[5] Another result extrapolated from 3C data, and confirmed *in situ*, is the co-localization of regulatory elements with their target promoters located several kilo-bases away. FISH has also confirmed juxtapositions of loci located on different chromosomes.[19] However, co-localization of loci was found to occur in a much lower proportion of cells in FISH[19] than it could be speculated from 3C-based results, highlighting the importance of taking precautions in the interpretation of 3C data obtained in a population of cells.

Several examples of more discrepancies between 3C-based data and *in situ* visualizations have been observed. In a recent systematic comparison of data obtained by 5C and FISH for several regions of the genome in different cell types, it was observed that, in many cases, the two approaches agreed.[33] However, the work also highlighted discrepancies between the interpretations of the data

© 2017 John Wiley & Sons, Ltd

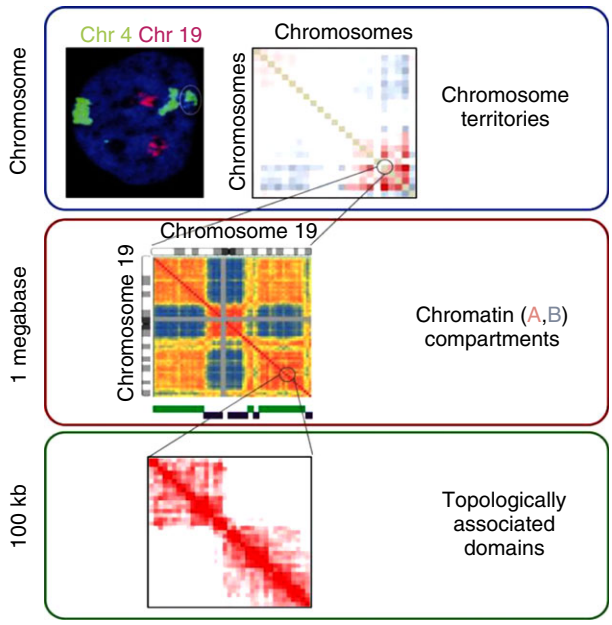

FIGURE 3 | Hi-C allows to study genome conformation at different scales and demonstrates a hierarchical organization. Binning the results of Hi-C experiment at different scale (e.g., whole chromosome, 1 Megabase, 100 kb windows or lower) permits to identify various levels of organization of the genome: chromosomes are organized as territories which occupy preferential relative positions. Chromatin is segregated into two compartments depending on its active or inactive state and chromosomes are segmented in topologically associating domains (TADs).

obtained by each of the two methods. For example, in some cases, regions appearing as highly interacting in 3C-derived data corresponded to highly unfolded regions in the FISH analysis.[33] One could hypothesize that these apparent discrepancies are linked to the technical differences in the processing of the samples, as for example, the thermic treatment required for the hybridization in FISH, even if such process has been shown to mainly preserve the nuclear architecture.[34] Or, on the side of 3C-derived methods, fixation process itself could bring artifacts at the time of the interpretation of data. Indeed, although formaldehyde is the most common cross-linking agent used in the study of chromatin, its exact chemical activity in cells remains poorly understood. This agent is known to generate DNA-protein as well as protein–protein cross-links. However, some studies revealed different efficiencies of cross-link depending on the binding dynamics of the proteins involved.[35] It is therefore possible that subnuclear environments with enrichments of specific proteins may have different cross-link propensity. In addition to the uncertain relationship between interaction frequency and spatial distance, the dynamic and stochastic nature of chromatin is also reflected in the single, agglomerative, conformation yielded by the 3C-based approaches. For example, a given interaction frequency between two loci could emerge from various possibilities within the cell population (Figure 4(a)). Additionally, although some attempts are made to analyze multiple pairs of concomitant contacts in Hi-C data,[36] the analysis is generally performed in a pair-wise manner, which does not allow to decipher whether an interaction between three or more genomic regions occur at once in a same cell or by pairs in different subpopulations of cells (Figure 4(b)).

In summary, due to its population nature and the potential biases in the cross-link procedure, the interpretation of 3C results remains challenging. Although the differences observed are probably less relevant at low resolution, it is likely that the discrepancies will increase with the resolution in both imaging and molecular approaches.

3D MODELING OF 3C-BASED DATA

Three-dimensional (3D) spatial representation of the data should help in extracting information not intuitively accessible only from 3C-based experiments. Increasing efforts are made to generate 3D objects that classify, filter, and recapitulate the information contained in the single 2D interaction matrices from 3C-based experiments conducted over millions of cells. Ideally, it should allow the generation of models that are not only a comprehensive and accurate representation of the genome structure but also permit to establish new hypothesis on the rules that govern the establishment and maintenance of such structures. Such ideal models would, in addition, help to understand the differences between single cell observations and population analysis.

Two main global approaches of genome structure modeling have emerged in the last few years. The first one, bottom-up, aim at using the physical and chemical properties of the chromatin fiber to build 3D models. Such approach is particularly useful to test specific biological hypothesis, as for example, the supercoiling of TADs,[37] the loop extrusion model,[29,38,39] the hierarchical organization of TADs,[40] or the spatial co-localization of co-regulated genes.[41] The second, which is the main subject of this review, correspond to top-down modeling approaches. In this case, the aim is to translate the experimental data into 3D Euclidean distances or spatial contact-maps and to use them to build the models. The projection of an interaction matrix into a list of 3D coordinates is a simple mathematical conversion called multidimensional scaling. Such

© 2017 John Wiley & Sons, Ltd

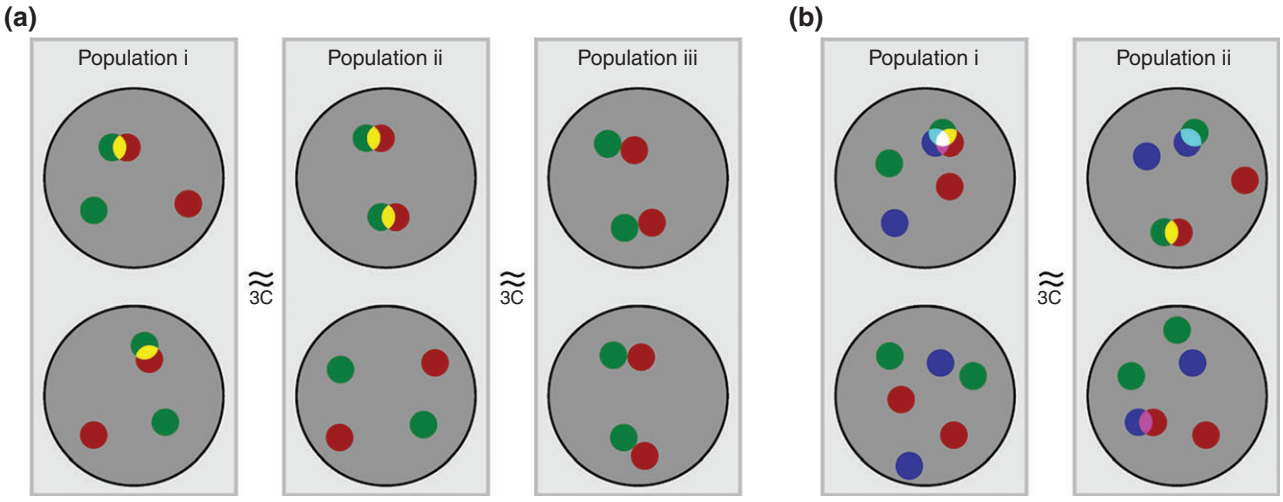

FIGURE 4 | Interpretation of 3C-based results as compared to single cell FISH. Classical 3C-based data do not allow to determine whether co-localization of genomic element occur within a single cell or if they represent the sum of distinct organization co-existing in the population. (a) i, ii, and iii correspond to different models of organization of two loci (green and red) which would give relatively similar results in 3C-derived datasets: in i, the red and green loci are co-localized in every cell of the population whereas the two other loci are located further away from each other. In ii, two subpopulations of cells are depicted, one presenting the co-localization of the two loci (green and red), the other without any colocalization. In iii, the green and red loci do not colocalize but their movement are restrained in a limited area which favor their random collision and increase the probability of cross-link. (b) Classically 3C-derived data give pair-wise information which do not permit to decipher whether three loci (blue, red, and green) colocalize in a same cell at the same time (population i) or whether two of them are interacting in different cells or at different time (population ii).

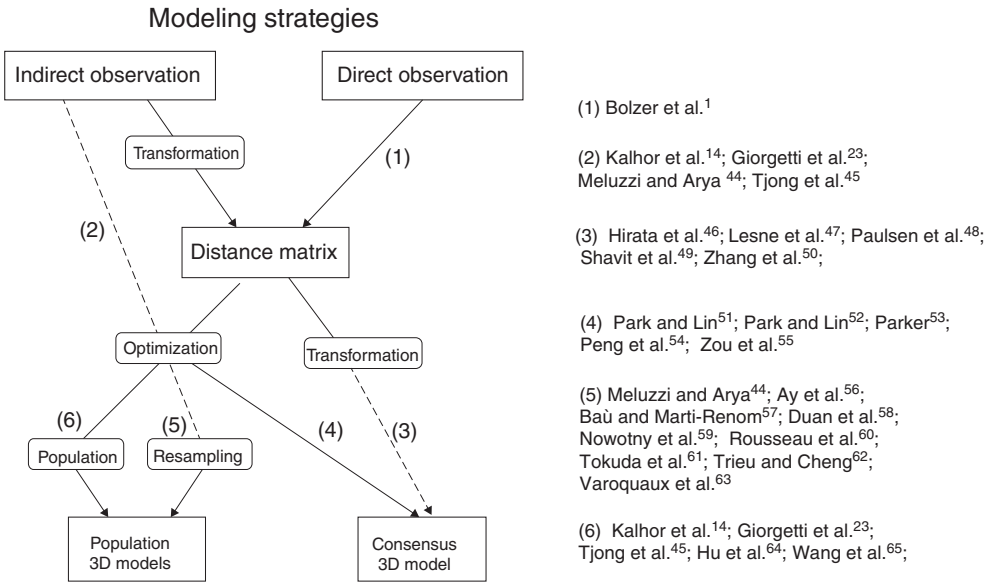

FIGURE 5 | Modeling strategies. Six main different modeling strategies are currently available for modeling genomes and genomic domains. Those methods diverse in their representation, scoring and sampling of the data. [1],[1] [2],[14,23,44,45] [3],[46–50] [4],[51–55] [5],[44,56–63] and [6].[14,23,45,64,65] Refer to the text for detailed explanation of each strategy.

methods are already largely used in the field of protein structure determination.[42] However, this approach is very sensitive to interaction matrices with large structural variability or noise. In this case, a single 3D model cannot explain all the interactions.[43] Such problem is precisely one of the limitations of classic 3C-derived experiments where the matrices generated represent a sum over millions of cells and where the number of captured interactions per cell is limited.

© 2017 John Wiley & Sons, Ltd

Modeling Strategies

Top-down modeling strategies consist in either integrating the experimental measures to generate a global solution, that is, one consensus model (Figure 5 [3] and [4]) or in considering that only a heterogeneous population of 3D models account for the variability of the experimental measurements (Figure 5 [5] and [6]).[45,57] Most of the modeling approaches require a distance- or contact-matrix to define the restraints that apply to the 3D models (Figure 5). Since in the case of 3C-derived experiments, the experimental data consist of indirect observations, (i.e., frequency of ligation), the first step usually consists of transforming the observed frequencies into distances. The transformation of interaction data into distances/contacts appears to have reached a global consensus; with the most generally used formula being:

$$D_{ij} \propto \left(\frac{1}{F_{ij}}\right)^{\alpha}$$

where F_{ij} is the frequency of interactions observed between the particles i and j, D_{ij} is the distance between these particles, and α is a correcting factor which reflects the correlation between interaction frequency and 3D distance. Based on polymer models, the α value is usually set to 1.0 in accordance with crumpled or fractal globule models.[4,66,67] This parameter α could, however, be fixed or optimized depending on the type of data (see Ref 68 for a recent review on the different strategies of optimization). The alternative of transforming interaction matrices into a continuous distance metric is to directly use the observed interactions of the Hi-C experiment. In this case the 3D models are built and optimized by trying to match the observed

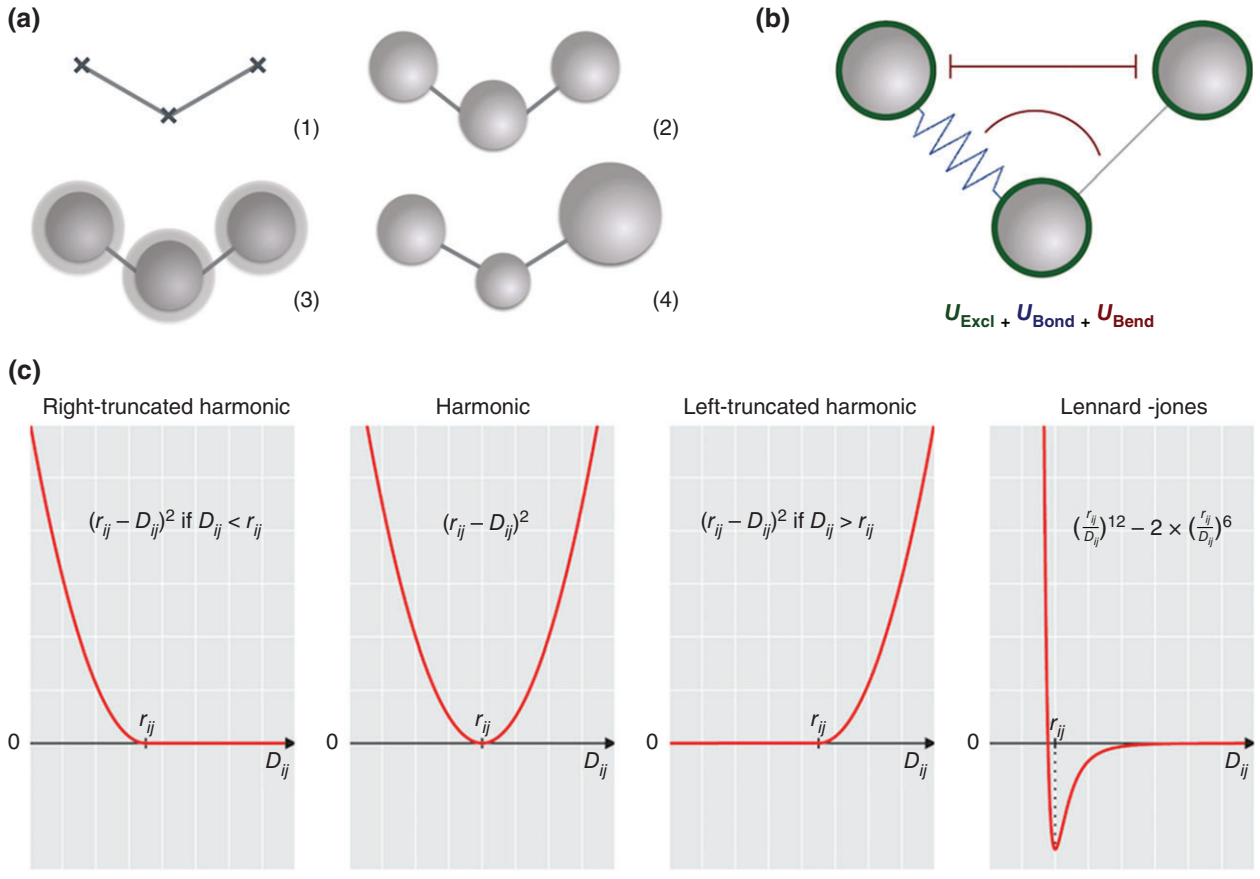

FIGURE 6 | Chromatin model representation. (a) Different ways of representing modeled chromatin loci: [1] as dots, with no physical consistency,[46–52,55,59,60,63,64] [2] as spheres,[14,16,23,44,45,54,56–58,61,62,65] [3] as spheres with complex physical properties, for example, a soft external surface and a hard core,[14,23] and [4] as sphere of different sizes (either representing different content in nucleotides, or different levels of compaction).[14,45,57] (b) Schematic representation of the application of the three main physical restraints, the excluded volume (U_{Excl}), the bond restraint (U_{Bond}) here represented either as a spring that could be modeled by a Lennard–Jones potential or as a harmonic restraint, the bending (U_{Bend}) that could in this example be modeled explicitly imposing an angle between three particles or by limiting the distance between two nonconsequtive particles (i, i + 2) using a right-truncated harmonic. (c) Different kind of potentials to be applied between particles i and j. D_{ij} being the Euclidian distance between them, r_{ij} the optimal distance inferred from, for example, the experimental data.

© 2017 John Wiley & Sons, Ltd

interactions between chromatin loci with the contacts in the modeled chromatin (Figure 5 [2]).

Once converted to distances, two alternative strategies are possible. The first consists in using multidimensional scaling to directly project the distance matrix into 3D coordinates[47,50](Figure 5 [3]). The second consists in optimizing the 3D conformation of chromatin models according to a set of restraints defined from the distance matrix[57] or directly from interaction frequencies (Figure 5 [5] and [6]).[14,44,45]

Independent of the model strategy, an initial step shared by all is to define a representation of these virtual segments of chromatin. Each bin of the distance or interaction matrix represents a segment of chromatin fiber, and each of these segment is generally represented either as a dot (Figure 6(a) [1]) or as a sphere (Figure 6(a) [2]). In the case where segments are represented as tangible objects, other layers of complexity can be added as, for example, a soft layer allowing a partial overlap between segments or a heterogeneous representation of segments into particles of different sizes (Figure 6(a) [3] and [4]). Another aspect of the representation is the type of restraints to be applied between particles (Figure 6 (b)). The physical restraints (U_{PHYS}) that are generally applied can be divided into three types:

$$U_{\mathrm{PHYS}} = U_{\mathrm{Excl}} + U_{\mathrm{Bond}} + U_{\mathrm{Bend}}$$

where U_{Excl} are the restraints corresponding to the excluded volume forbidding two particles to occupy the same space, U_{Bond} are the restraints related to the bond between neighboring particles forcing contiguous particle to stay close, and U_{Bend} defines the rigidity of the modeled chromatin fiber (see Figure 6 (b) and (c), and Ref 68 for detailed explanation of the application of restraints). Once defined, different optimization algorithms can be used to apply these restraints on 3D representations of the chromatin and fit their shape with the input interaction matrices (see Ref 68 for details). For the optimization, the application of the restraints can be done on a single model (Figure 5 [3] and [4]) or on population of models (Figure 5 [5] and [6]). In the case of population-based modeling, several 3D models are optimized together. At the end of the optimization, the sum of all their conformations should reproduce the input distance- or interaction-matrix. In a sense, during the optimization, the models are competing for the input restraints, and this allows that each of the models explains only a portion of the interactions of the Hi-C experiment. With this modeling strategy, it is important to consider that an increase in the number of models involves an increase in the number of free parameters and in the computation time. An alternative to the optimization of several models together is the resampling strategy (Figure 5 [5]). In this case, the number of free parameters is the same as for a single model but the output is a population of models. Methodologically, it simply consists of applying the consensus modeling approach repeatedly with different initial states. Resampling approaches are based on the assumption that each local minima of the optimization process corresponds to a subpopulation of cells with similar chromatin folding.

Data Quality Inference

In all cases, data quality is an important parameter to take into account when choosing for the suitable modeling strategy.[43,47,50,52] In particular, a recent work has shown the possibility to assess the accuracy of resulting 3D models by using statistical properties of the input interaction matrices.[43] Indeed, a combination of descriptive statistics extracted from the input interaction-matrix can help to understand the quality of the experimental data and its suitability for 3D modeling. For example, low structural variability in the folding of the chromatin is usually accompanied with low values of skewness, kurtosis, and high number of significant Eigenvalues calculated from the input matrices.[43] The results of this work also indicated that structural variability was the principal source of error in 3D modeling.

Resolving Structural Variability

As mentioned above, the structural variability inherent to the population-based 3C-derived methods appear to be an important source of inaccuracy of 3D models.[43] Therefore, modeling strategies that aim at identifying population of models within the datasets are, in theory, more adapted than strategies developing single consensus models. In this case, 3D modeling should allow deconvoluting subpopulation of cells with distinct global chromatin folding. With this idea, many population-based modeling strategies have been implemented in the last years (Figure 5 [5]). These techniques were shown to be successful at identifying the subpopulation of models within single 3C-derived experiment. However, a particular caution must be taken with the choice of the number of models to optimize as the number of free parameters increases with each of them. Even though the latest modeling approaches are designed specifically to deal with these very low reads per cell ratio,[48,52] the main problem of population-based modeling when

© 2017 John Wiley & Sons, Ltd

deconvoluting 3C-based matrices is that the original data consist of the sum very sparse single cell matrices with a very low ratio of observations per cell[32] and precautions should still be taken at the time of the interpretation.

Validation of 3D Models

Despite the efforts put in developing realistic modeling strategies, it remains essential to assess the accuracy of the models generated with data obtained independently. Typically, observations from *in situ* by FISH are used to validate inferred 3D models. This methodology can, for example, involve a correlation between the distances of two or more FISH probes and the distances measured on the corresponding positions in a set of 3D models.[5,12,14,23,54,56,60,69–71] However, as previously discussed, and beyond the differences in resolution, the main issue with this validation strategy is the low throughput of the FISH experiment. A large number of experimental replicates are usually needed to confirm the predicted folding of the chromatin. It is expected that with the emergence of new *in situ* methods with increased throughput, larger-scale assessment of 3D modeling accuracy will be available.

Integration of Distinct Types of Data in the Models

Restraint-based modeling primarily uses 3C-based data. Integrating other kind of data prior to modeling would help in increasing the accuracy of the models generated and in optimizing the modeling parameters, as for example, the physical scale or degree of compaction of the fiber. One could envisage that distinct kind of data can be used as, for example, direct microscopy observations (e.g., FISH), general features of polymer physics models (e.g., rigidity of the fiber), or biological knowledge (e.g., nucleus size or known chromosome tethering). Indeed, several studies of polymer physic-based modeling already integrated known biological properties,[72] TADs,[40,73] and epigenetic state.[74] An example of integration of FISH data for chromatin modeling is the work from Bolzer et al. where chromosome territories where observed by FISH, and used as confinement areas for the polymer models representing chromatin fiber.[1] It is clear that there is still work to be done to translate distinct kind of data into sets of restraints that would be integrated into a single modeling protocol. However, this translation should be the hardest part as restraints-based modeling is already able to integrate data from different sources to define and apply restraints to 3D models.[75] Beyond the methodological difficulties, the main problem with the integration of distinct types of data in the modeling is the decision of which data should be used. For example, and as mentioned previously, FISH data are usually used to validate the models, and thus also cannot be used to build the models. Therefore, when additional kinds of data are available, a decision has to be made: use the data to validate the model or use the data to build the model.

FUTURE DIRECTIONS

Recent advances in imaging techniques have increased our capacity to describe the mechanisms of dynamic changes of chromatin.[76] These advances have been complemented with *in silico* models of the chromatin or naked DNA fiber, which have been used to mechanistically explain the physics behind conformational changes of the genome.[77,78] Altogether, both imaging and computational modeling could be used to assess the source of structural variability in 3C-derived experiments, which is likely arising from the polymorphic nature of the DNA flexibility and dynamics. This chromatin heterogeneity, already clear between heterochromatin and euchromatin, is a factor that has not yet been integrated in any of the here-discussed modeling strategies, although it has been shown to explain an important part of the observed variability in 3C-based interaction matrices.[61] In addition, it may be one the principal sources of discrepancies between direct and indirect observations.[35,79] From the perspective of 3C-based detection of interactions, more open and thus more dynamic region could result in increased contacts.[33] This observation might explain some of the discrepancies between 3C-derived experiment and FISH observations.

From the modeling perspective, this polymorphic nature of the chromatin could be addressed simply by varying the scale parameter (number of nucleotides per nanometer), or the α parameter (see above) and optimizing these modeling parameters independently for different chromatin regions. Different chromatin regions could thus be defined based on prior knowledge obtained from their epigenetic state. Indeed, chromatin structural features are related to the chromatin epigenetic state.[22,54,80] Being able to predict chromatin structure from epigenetic state is one of the many areas toward which structural biology, and epigenetics are heading. Although tools to do the conversion are already

© 2017 John Wiley & Sons, Ltd

available,[81] their application is still limited to interacting pairs of promoters or, and, enhancers within a single TAD. Such predictive tools are still thus in their early stages, but this kind of analysis could help in validating or even *de novo* reconstructing 3D models of the chromatin.

ACKNOWLEDGMENTS

All the authors contributed equally to this work. F.L.D. thanks Miguel Beato and the members of the Chromatin and Gene expression Group (CRG, Barcelona). This review has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement 609989. The content of this review reflects only the author's views and the Union is not liable for any use that may be made of the information contained therein. We acknowledge funding from the Spanish Ministry of Economy and Competitiveness (BFU2013-47736-P) and the European Union's Horizon 2020 research and innovation programme under grant agreement No 676556 and support of the Spanish Ministry of Economy and Competitiveness, 'Centro de Excelencia Severo Ochoa 2013-2017', SEV-2012-0208 and of the CERCA Programme/Generalitat de Catalunya to the CRG.

REFERENCES

1. Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, Muller S, Eils R, Cremer C, Speicher MR, et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* 2005, 3:e157.
2. Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* 1999, 145:1119–1131.
3. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 2012, 485:376–380.
4. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 2009, 326:289–293.
5. Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 2012a, 485:381–385.
6. Lupianez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 2015, 161:1012–1025.
7. Thomson I, Gilchrist S, Bickmore WA, Chubb JR. The radial positioning of chromatin is not inherited through mitosis but is established de novo in early G1. *Curr Biol* 2004, 14:166–172.
8. Cavalli G, Misteli T. Functional implications of genome topology. *Nat Struct Mol Biol* 2013, 20:290–299.
9. Pope BD, Ryba T, Dileep V, Yue F, Wu W, Denas O, Vera DL, Wang Y, Hansen RS, Canfield TK, et al. Topologically associating domains are stable units of replication-timing regulation. *Nature* 2014, 515:402–405.
10. Solovei I, Cavallo A, Schermelleh L, Jaunin F, Scasselati C, Cmarko D, Cremer C, Fakan S, Cremer T. Spatial preservation of nuclear chromatin architecture during three-dimensional fluorescence in situ hybridization (3D-FISH). *Exp Cell Res* 2002, 276:10–23.
11. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation. *Science* 2002, 295:1306–1311.
12. Baù D, Sanyal A, Lajoie BR, Capriotti E, Byron M, Lawrence JB, Dekker J, Marti-Renom MA. The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. *Nat Struct Mol Biol* 2011a, 18:107–114.
13. Belton JM, Lajoie BR, Audibert S, Cantaloube S, Lassadi I, Goiffon I, Bau D, Marti-Renom MA, Bystricky K, Dekker J. The conformation of yeast chromosome iii is mating type dependent and controlled by the recombination enhancer. *Cell Rep* 2015, 13:1855–1867.
14. Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat Biotechnol* 2012, 30:90–98.
15. Markaki Y, Smeets D, Fiedler S, Schmid VJ, Schermelleh L, Cremer T, Cremer M. The potential of 3D-FISH and super-resolution structured illumination

 © 2017 John Wiley & Sons, Ltd

microscopy for studies of 3D nuclear architecture: 3D structured illumination microscopy of defined chromosomal structures visualized by 3D (immuno)-FISH opens new perspectives for studies of nuclear architecture. *Bioessays* 2012, 34:412–426.

16. Cremer M, Grasser F, Lanctôt C, Müller S, Neusser M, Zinner R, Solovei I, Cremer T. Multicolor 3D fluorescence in situ hybridization for imaging interphase chromosomes. *Methods Mol Biol* 2008, 463:205–239.
17. Goetze S, Mateos-Langerak J, Gierman HJ, de Leeuw W, Giromus O, Indemans MH, Koster J, Ondrej V, Versteeg R, van Driel R. The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Mol Cell Biol* 2007, 27:4475–4487.
18. Mahy NL, Perry PE, Gilchrist S, Baldock RA, Bickmore WA. Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *J Cell Biol* 2002, 157:579–589.
19. Osborne CS, Chakalova L, Mitchell JA, Horton A, Wood AL, Bolland DJ, Corcoran AE, Fraser P. Myc dynamically and preferentially relocates to a transcription factory occupied by Igh. *PLoS Biol* 2007, 5:e192.
20. Osborne CS, Chakalova L, Brown KE, Carter D, Horton A, Debrand E, Goyenechea B, Mitchell JA, Lopes S, Reik W, et al. Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat Genet* 2004, 36:1065–1071.
21. Beliveau BJ, Apostolopoulos N, Wu CT. Visualizing genomes with Oligopaint FISH probes. *Curr Protoc Mol Biol* 2014, 105:14.23.
22. Boettiger AN, Bintu B, Moffitt JR, Wang S, Beliveau BJ, Fudenberg G, Imakaev M, Mirny LA, Wu CT, Zhuang X. Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 2016, 529:418–422.
23. Giorgetti L, Galupa R, Nora EP, Piolot T, Lam F, Dekker J, Tiana G, Heard E. Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* 2014, 157:950–963.
24. Hajjoul H, Mathon J, Ranchon H, Goiffon I, Mozziconacci J, Albert B, Carrivain P, Victor JM, Gadal O, Bystricky K, et al. High-throughput chromatin motion tracking in living yeast reveals the flexibility of the fiber throughout the genome. *Genome Res* 2013, 23:1829–1838.
25. Bystricky K. Chromosome dynamics and folding in eukaryotes: insights from live cell microscopy. *FEBS Lett* 2015, 589:3014–3022.
26. Shachar S, Pegoraro G, Misteli T. HIPMap: a high-throughput imaging method for mapping spatial gene positions. *Cold Spring Harb Symp Quant Biol* 2015, 80:73–81.
27. Zhao Z, Tavoosidana G, Sjölinder M, Göndör A, Mariano P, Wang S, Kanduri C, Lezcano M, Sandhu KS, Singh U, et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat Genet* 2006, 38:1341–1347.
28. Dostie J, Ta R, Ra A, Selzer RR, Lee WL, Ta H, Rubio ED, Krumm A, Lamb J, Nusbaum C, et al. Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* 2006, 16:1299–1309.
29. Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 2014, 159:1665–1680.
30. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 2013, 503:290–294.
31. Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* 2012, 148:458–472.
32. Nagano T, Lubling Y, Stevens TJ, Schoenfelder S, Yaffe E, Dean W, Laue ED, Tanay A, Fraser P. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 2013, 502:59–64.
33. Williamson I, Berlivet S, Eskeland R, Boyle S, Illingworth RS, Paquette D, Dostie J, Bickmore WA. Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization. *Genes Dev* 2014, 28:2778–2791.
34. Cremer M, Muller S, Kohler D, Brero A, Solovei I. Cell preparation and multicolor FISH in 3D preserved cultured mammalian cells. *CSH Protoc* 2007, 2007:pdb prot4723.
35. Gavrilov A, Razin SV, Cavalli G. In vivo formaldehyde cross-linking: it is time for black box analysis. *Brief Funct Genomics* 2015, 14:163–165.
36. Darrow EM, Huntley MH, Dudchenko O, Stamenova EK, Durand NC, Sun Z, Huang S-C, Sanborn AL, Machol I, Shamim M, et al. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc Natl Acad Sci U S A* 2016, 113:E4504–E4512.
37. Benedetti F, Dorier J, Burnier Y, Stasiak A. Models that include supercoiling of topological domains reproduce several known features of interphase chromosomes. *Nucleic Acids Res* 2014, 42:2848–2855.
38. Fudenberg G, Imakaev M, Lu C, Goloborodko A, Abdennur N, Mirny LA. Formation of chromosomal domains by loop extrusion. *Cell Rep* 2016, 15:2038–2049.

© 2017 John Wiley & Sons, Ltd

39. Fudenberg G, Mirny L. Higher-order chromatin structure: bridging physics and biology. *Curr Opin Genet Dev* 2012, 22:115–124.
40. Fraser J, Ferrai C, Chiariello AM, Schueler M, Rito T, Laudanno G, Barbieri M, Moore BL, Kraemer DC, Aitken S, et al. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Mol Syst Biol* 2015, 11:1–14.
41. Di Stefano M, Rosa A, Belcastro V, di Bernardo D, Micheletti C. Colocalization of coregulated genes: a steered molecular dynamics study of human chromosome 19. *PLoS Comput Biol* 2013, 9:e1003019.
42. Havel TF, Kuntz ID, Crippen GM. The theory and practice of distance geometry. *Bull Math Biol* 1983, 45:665–720.
43. Trussart M, Serra F, Baù D, Junier I, Serrano L, Marti-Renom MA. Assessing the limits of restraint-based 3D modeling of genomes and genomic domains. *Nucleic Acids Res* 2015, 43:3465–3477.
44. Meluzzi D, Arya G. Recovering ensembles of chromatin conformations from contact probabilities. *Nucleic Acids Res* 2013, 41:63–75.
45. Tjong H, Li W, Kalhor R, Dai C, Hao S, Gong K, Zhou Y, Li H, Zhou XJ, Le Gros MA, et al. Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *Proc Natl Acad Sci U S A* 2016, 113:E1663–E1672.
46. Hirata Y, Oda A, Ohta K, Aihara K. Three-dimensional reconstruction of single-cell chromosome structure using recurrence plots. *Sci Rep* 2016, 6:34982.
47. Lesne A, Riposo J, Roger P, Cournac A, Mozziconacci J. 3D genome reconstruction from chromosomal contacts. *Nat Methods* 2014, 4:10–13.
48. Paulsen J, Gramstad O, Collas P. Manifold based optimization for single-cell 3D genome reconstruction. *PLoS Comput Biol* 2015, 11:1–19.
49. Shavit Y, Hamey FK, Lio P. FisHiCal: an R package for iterative FISH-based calibration of Hi-C data. *Bioinformatics* 2014, 30:3120–3122.
50. Zhang Z, Li G, Toh K-C, Sung W-K. 3D chromosome modeling with semi-definite programming and Hi-C data. *J Comput Biol* 2013, 20:831–846.
51. Park J, Lin S. Statistical inference on three-dimensional structure of genome by truncated poisson architecture model. In: Choudhary P, Nagaraja C, Ng H, eds. *Ordered Data Analysis, Modeling and Health Research Methods*. Springer Proceedings in Mathematics & Statistics, vol. 149. Cham: Springer; 2015, 245–261.
52. Park J, Lin S. Impact of data resolution on three-dimensional structure inference methods. *BMC Bioinformatics* 2016, 17:70.
53. Parker GA. Sperm competition and its evolutionary consequences in the insects. *Biol Rev* 1970, 45:525–567.
54. Peng C, Fu L-Y, Dong P-F, Deng Z-L, Li J-X, Wang X-T, Zhang H-Y. The sequencing bias relaxed characteristics of Hi-C derived data and implications for chromatin 3D modeling. *Nucleic Acids Res* 2013, 41:e183.
55. Zou C, Zhang Y, Ouyang Z. HSA: integrating multitrack Hi-C data for genome-scale reconstruction of 3D chromatin structure. *Genome Biol* 2016, 17:40.
56. Ay F, Bunnik EM, Varoquaux N, Bol SM, Prudhomme J, Vert JP, Noble WS, Le Roch KG. Three-dimensional modeling of the *P. falciparum* genome during the erythrocytic cycle reveals a strong connection between genome architecture and gene expression. *Genome Res* 2014, 24:974–988.
57. Baù D, Marti-Renom M. Genome structure determination via 3C-based data integration by the integrative modeling platform. *Methods* 2012, 58:300–306.
58. Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, Shendure J, Fields S, Blau CA, Noble WS. A three-dimensional model of the yeast genome. *Nature* 2010, 465:363–367.
59. Nowotny J, Ahmed S, Xu L, Oluwadare O, Chen H, Hensley N, Trieu T, Cao R, Cheng J. Iterative reconstruction of three-dimensional models of human chromosomes from chromosomal contact data. *BMC Bioinformatics* 2015, 16:338.
60. Rousseau M, Fraser J, Ma F, Dostie J, Blanchette M. Three-dimensional modeling of chromatin structure from interaction frequency data using Markov chain Monte Carlo sampling. *BMC Bioinformatics* 2011, 12:414.
61. Tokuda N, Terada TP, Sasai M. Dynamical modeling of three-dimensional genome organization in interphase budding yeast. *Biophys J* 2012, 102:296–304.
62. Trieu T, Cheng J. MOGEN : a tool for reconstructing 3D models of genomes from chromosomal conformation capturing data. *Bioinformatics* 2015, 32:1–7.
63. Varoquaux N, Ay F, Noble WS, Vert J-P. A statistical approach for inferring the 3D structure of the genome. *Bioinformatics* 2014, 30:i26–33.
64. Hu M, Deng K, Qin Z, Dixon J, Selvaraj S, Fang J, Ren B, Liu JS. Bayesian inference of spatial organizations of chromosomes. *PLoS Comput Biol* 2013a, 9: e1002893.
65. Wang S, Xu J, Zeng J. Inferential modeling of 3D chromatin structure. *Nucleic Acids Res* 2015, 43:1–12.
66. Grosberg AY, Nechaev SK, Shakhnovich EI. The role of topological constraints in the kinetics of collapse of macromolecules. *J Phys* 1988, 49:2095–2100.
67. Mirny L. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res* 2011, 19:37–51.

© 2017 John Wiley & Sons, Ltd

68. Serra F, Di Stefano M, Spill YG, Cuartero Y, Goodstadt M, Baù D, Marti-Renom MA. Restraint-based three-dimensional modeling of genomes and genomic domains. *FEBS Lett* 2015, 589:2987–2995.

69. Hu M, Deng K, Qin Z, Liu JS. Understanding spatial organizations of chromosomes via statistical analysis of Hi-C data. *Quant Biol* 2013b, 1:156–174.

70. Le Dily F, Baù D, Pohl A, Vicent GP, Serra F, Soronellas D, Castellano G, Wright RHG, Ballare C, Filion G, et al. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev* 2014, 28:2151–2162.

71. Umbarger M, Toro E, Wright M, Porreca GJ, Baù D, Hong S-H, Fero MJ, Zhu LJ, Marti-Renom M, McAdams HH, et al. The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Mol Cell* 2011, 44:252–264.

72. Tjong H, Gong K, Chen L, Alber F. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res* 2012, 22:1295–1305.

73. Barbieri M, Chotalia M, Fraser J, Lavitas LM, Dostie J, Pombo A, Nicodemi M. Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci U S A* 2012, 109:16173–16178.

74. Jost D, Carrivain P, Cavalli G, Vaillant C. Modeling epigenome folding: formation and dynamics of topologically-associated chromatin domains. *Nucleic Acids Res* 2014, 42:9553–9561.

75. Russel D, Lasker K, Webb B, Velázquez-Muriel J, Tjioe E, Schneidman-Duhovny D, Peterson B, Sali A. Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol* 2012, 10:e1001244.

76. Dion V, Gasser SM. Chromatin movement in the maintenance of genome stability. *Cell* 2013, 152:1355–1364.

77. Amitai A, Toulouze M, Dubrana K, Holcman D. Analysis of single locus trajectories for extracting in vivo chromatin tethering interactions. *PLoS Comput Biol* 2015, 11:e1004433.

78. Shinkai S, Nozaki T, Maeshima K, Togashi Y. Dynamic nucleosome movement provides structural information of topological chromatin domains in living human cells. *PLoS Comput Biol* 2016, 12:e1005136.

79. Sexton T, Cavalli G. The role of chromosome domains in shaping the functional genome. *Cell* 2015, 160:1049–1059.

80. Serra F, Baù D, Filion G, Marti-Renom MA. Structural features of the fly chromatin colors revealed by automatic three-dimensional modeling. *bioRxiv* 2016:1–29.

81. Zhu Y, Chen Z, Zhang K, Wang M, Medovoy D, Whitaker JW, Ding B, Li N, Zheng L, Wang W. Constructing 3D interaction maps from 1D epigenomes. *Nat Commun* 2016, 7:10812.

© 2017 John Wiley & Sons, Ltd