

# Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data

Job Dekker<sup>1</sup>, Marc A. Marti-Renom<sup>2,3</sup> and Leonid A. Mirny<sup>4</sup>

**Abstract** | How DNA is organized in three dimensions inside the cell nucleus and how this affects the ways in which cells access, read and interpret genetic information are among the longest standing questions in cell biology. Using newly developed molecular, genomic and computational approaches based on the chromosome conformation capture technology (such as 3C, 4C, 5C and Hi-C), the spatial organization of genomes is being explored at unprecedented resolution. Interpreting the increasingly large chromatin interaction data sets is now posing novel challenges. Here we describe several types of statistical and computational approaches that have recently been developed to analyse chromatin interaction data.

## Restraint-based modelling

A computational method to model the three-dimensional structure of an object represented by points and restraints between them.

<sup>1</sup>Program in Systems Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 368 Plantation Street, Worcester, Massachusetts 01605–2324, USA.

<sup>2</sup>Genome Biology Group, Centre Nacional d'Anàlisi Genòmic (CNAG), Baldiri Reixac 4, 08028 Barcelona, Spain.

<sup>3</sup>Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), Dr. Aiguader 88, 08003 Barcelona, Spain.

<sup>4</sup>Institute for Medical Engineering and Science, and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. Correspondence to J.D.

e-mail: [job.dekker@umassmed.edu](mailto:job.dekker@umassmed.edu)

doi:10.1038/nrg3454

Published online 9 May 2013

Chromosomes are some of the most complex molecular entities in the cell: the molecular composition of the chromatin fibre is highly diverse along its length, and the fibre is intricately folded in three dimensions. Tremendous efforts are being devoted to mapping the local structure of chromatin by analysing the complement of DNA-associated proteins and their modifications along chromosomes. Such studies allow the identification of genomic locations of genes and regulatory elements that are active in a given cell type, and they have started to uncover comprehensive sets of functional elements of the human genome and the genomes of several model organisms (for example, REFS 1–3). Only over the past decade has a series of molecular and genomic approaches been developed that can be used to study three-dimensional (3D) chromosome folding at increasing resolution and throughput; these methods are all based on chromosome conformation capture (3C). These methods allow the determination of the frequency with which any pair of loci in the genome is in close enough physical proximity (probably in the range of 10–100 nm) to become crosslinked<sup>4,5</sup> (BOX 1).

These 3C-based methods are starting to generate vast amounts of genome-wide interaction data. Here we briefly describe the main experimental approaches and then describe in more depth recently developed analytical, computational and modelling approaches for analysis of comprehensive chromatin interaction data sets. We discuss three emerging approaches to analyse 3C-based data sets. The first approach simply aims to

identify pairs or sets of loci that interact more frequently than would otherwise be expected, which points to chromatin looping or specific co-location events. Analysis of groups of preferentially interacting loci has been used to identify higher-order chromosomal domains. The other two approaches — restraint-based modelling and approaches that model chromatin as a polymer — use all of the interaction data, including baseline and non-specific interactions, to build ensembles of spatial models of chromosomes. 3D models can then be used to identify higher-order structural features and DNA elements that are involved in organizing chromosomes and to estimate chromatin dynamics within one cell as well as cell-to-cell variability in folding. We discuss how the application of these approaches is starting to uncover principles that determine the spatial organization of chromosomes, to reveal novel layers of chromatin structure and to relate these structures to gene expression and regulation.

## Studying chromosome organization

**Insights from imaging.** When chromosomes are observed in living cells, they can appear highly variable between cells<sup>6,7</sup>, and this could be interpreted as reflecting a general lack of organization. However, detailed studies using various improved imaging techniques have revealed several organizational principles of chromosomes at the scale of the whole nucleus<sup>7</sup>. First, in interphase cells of many organisms, chromosomes do not readily mix but instead occupy their own separate territories<sup>8</sup>. Second, where chromosome territories

Box 1 | 3C-based methods

**a 3C: converting chromatin interactions into ligation products**



**b Ligation product detection methods**

3C	4C	5C	ChIA-PET	Hi-C
One-by-one All-by-all	One-by-all	Many-by-many	Many-by-many	All-by-all
			<ul style="list-style-type: none"> <li>• DNA shearing</li> <li>• Immunoprecipitation</li> </ul>	<ul style="list-style-type: none"> <li>• Biotin labelling of ends</li> <li>• DNA shearing</li> </ul>
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing	Sequencing	Sequencing

In chromosome conformation capture (3C)-based methods (see panel **a** of the figure), cells are crosslinked with formaldehyde to link chromatin segments covalently that are in close spatial proximity. Next, chromatin is fragmented by restriction digestion or sonication. Crosslinked fragments are then ligated to form unique hybrid DNA molecules. Finally, the DNA is purified and analysed. The different 3C-based methods differ only in the way that hybrid DNA molecules, each corresponding to an interaction event of a pair of loci, are detected and quantified (see panel **b** of the figure). In classical 3C experiments, single ligation products are detected by PCR one at the time using locus-specific primers. Given that 3C can be laborious, most 3C analyses typically cover only tens to several hundreds of kilobases. 4C (also known as ‘circular 3C’ or ‘3C-on-chip’) uses inverse PCR to generate genome-wide interaction profiles for single loci<sup>48,105,106</sup>. 5C combines 3C with hybrid capture approaches to identify up to millions of interactions in parallel between two large sets of loci: for example, between a set of promoters and a set of distal regulatory elements<sup>46,107,108</sup>. 4C approaches are genome-wide but are anchored on a single locus. 5C analyses typically involve two sets of hundreds to thousands of restriction fragments to interrogate up to millions of long-range interactions that can cover up to tens of megabases and that can be contiguous or scattered among loci of interest throughout the genome. The Hi-C method was the first unbiased and genome-wide adaptation of 3C and includes a unique step in which, after restriction digestion, the staggered DNA ends are filled in with biotinylated nucleotides (as shown by the asterisks)<sup>64</sup>. This facilitates selective purification of ligation junctions that are then directly sequenced. Hi-C provides a true all-by-all genome-wide interaction map, but the resolution of this map depends on the depth of sequencing. When several hundred million read pairs are obtained, as is currently routine, chromatin interactions in the mouse or human genome can be detected at 100 kb resolution.

Other 3C variants have recently been described that differ in molecular details but that all generate comprehensive and genome-wide interaction maps<sup>28,47,57,75</sup>. Interestingly, technology development has now gone full circle back to 3C: the classical 3C method is no longer used only for analysing interactions one at the time by PCR but is now also used for genome-wide interaction mapping as the resulting complete 3C DNA ligation mixture can be directly sequenced on modern deep-sequencing platforms<sup>57</sup>. Finally, various approaches combine 3C with chromatin immunoprecipitation to enrich for chromatin interactions between loci bound by specific proteins of interest<sup>109,110</sup>. For instance, the chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) method allows for genome-wide analysis of long-range interactions between sites bound by a protein of interest. Because ChIA-PET data represent a selected subset of interactions that occur in the genome, the three analysis approaches described in this article cannot directly be applied to this data type. LMA, ligation-mediated amplification.

**Chromosome territories**  
Each territory is the domain of a nucleus occupied by a chromosome.

**Polycomb bodies**  
Discrete nuclear foci containing Polycomb proteins and their silenced target genes. Polycomb bodies have been observed in *Drosophila melanogaster* and human cells by imaging and *in situ* hybridization.

touch, they can form areas in which intermingling occurs, providing opportunities for potentially functional interactions between loci located on different chromosomes<sup>9</sup>. Third, transcription does not occur diffusely throughout the nucleus but happens at subnuclear sites enriched in RNA polymerase II and other components of the transcription and RNA-processing machinery<sup>10–12</sup>. This implies that actively transcribed genes tend to co-localize, possibly in specific groups

related to their transcriptional regulators<sup>13</sup>. Finally, transcriptionally inactive segments of the genome also tend to associate with each other and often can be found localized at the nuclear periphery<sup>14</sup>, around nucleoli<sup>15,16</sup> or, in *Drosophila melanogaster*, at subnuclear structures such as Polycomb bodies<sup>17–19</sup>. These observations point to a spatially and functionally compartmentalized nucleus, in which subnuclear positioning of loci is correlated with gene expression.

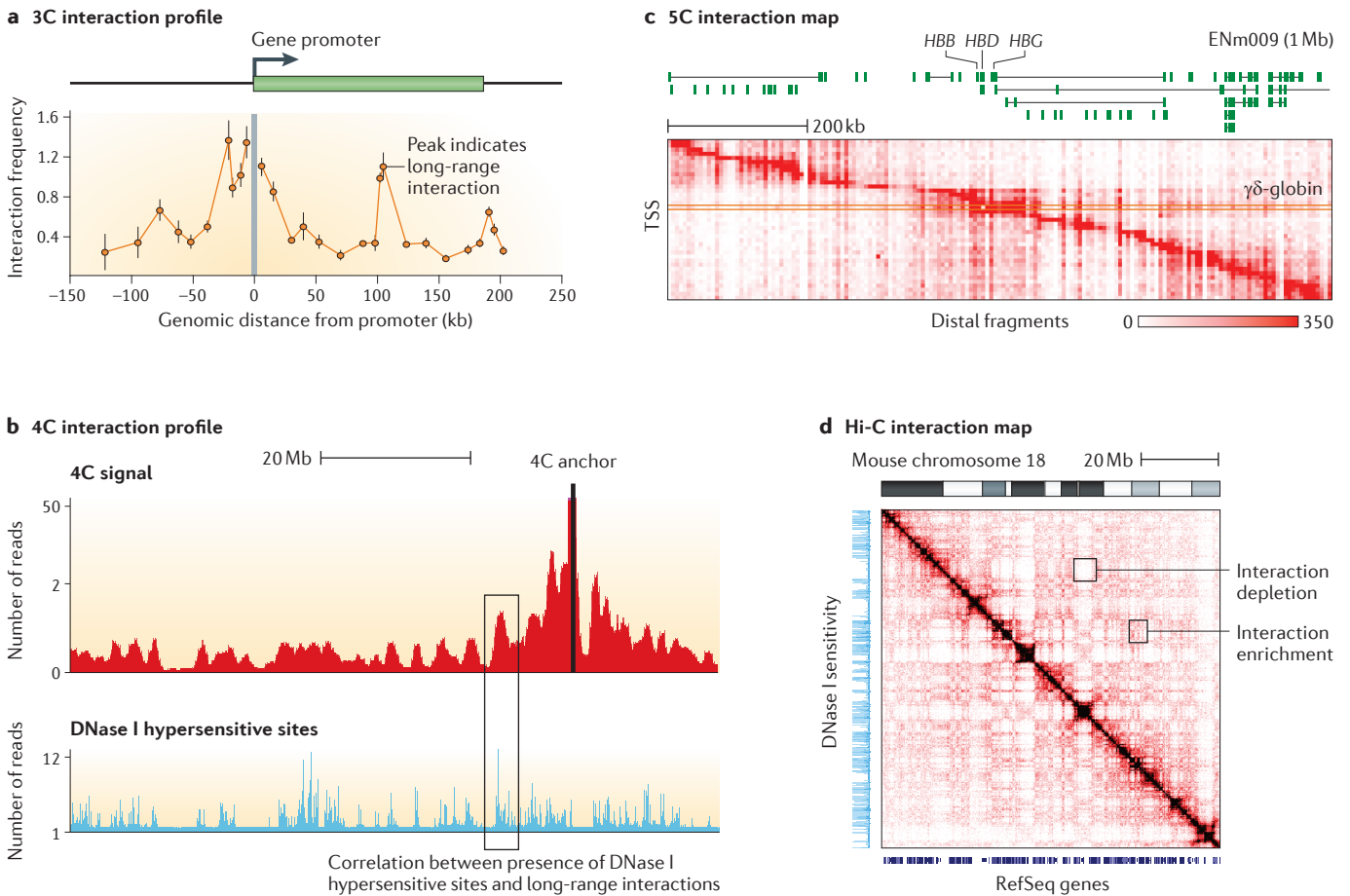
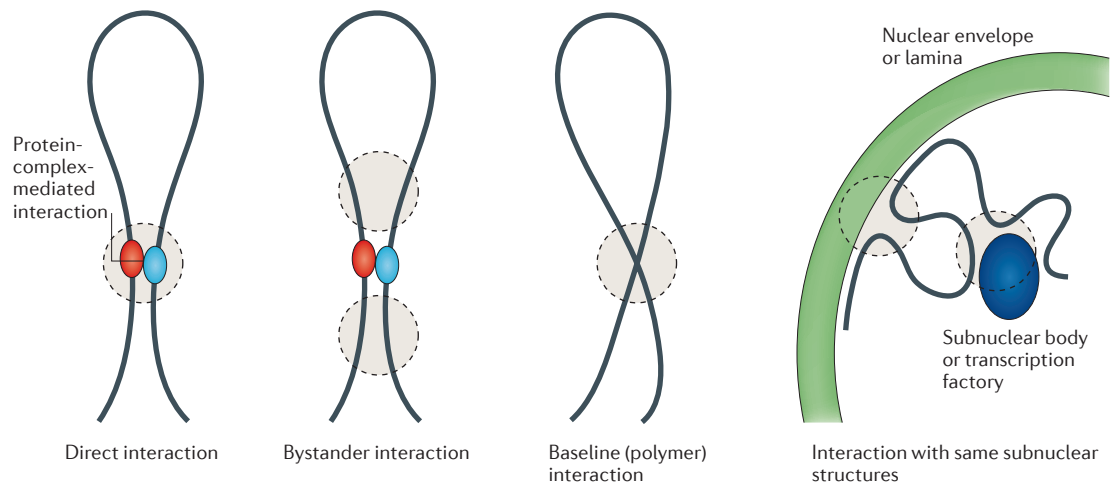


Figure 1 | **Examples of 3C, 4C, 5C and Hi-C data sets.** **a** | Chromosome conformation capture (3C) data for the *CFTR* gene in Caco-2 cells (which are a human colon adenocarcinoma cell line)<sup>34</sup>. **b** | 4C data from the mouse genome and DNase I hypersensitivity data from the same region, simulated from data from REF. 112. **c** | An example of a 5C interaction map for the ENCODE ENm009 region in K562 cells (which are a human erythroleukaemia cell line) based on data from REF. 46. Each row represents an interaction profile of a transcription start site (TSS) across the 1 Mb region on human chromosome 11 that contains the  $\beta$ -globin locus. **d** | Hi-C data from mouse chromosome 18 from REF. 111. 3C and 4C data are linear profiles along chromosomes and can be directly compared to other genomic tracks such as DNase I sensitivity. 5C and Hi-C data are often represented as two-dimensional heat maps. Other genomic features, such as positions of genes or the location of DNase I hypersensitive sites, can be displayed along the axes for visual analysis of chromosome structural features. DNase I data are taken from the Mouse ENCODE Consortium, from the laboratory of J. Stamatoyannopoulos<sup>113</sup>. Part **a** is modified, with permission, from REF. 34 © (2010) Oxford Univ. Press. Part **d** is modified, with permission, from REF. 112 © (2012) Macmillan Publishers Ltd. All rights reserved.

**3C-based technologies.** Imaging approaches do not readily allow a comprehensive analysis of the 3D folding of complete genomes or determination of the organization of entire chromosomes within their territories at kilobase resolution. To overcome these limitations, approaches based on 3C have been developed that allow the mapping of chromosome folding at sufficient resolution to observe individual genes and regulatory elements and that can operate at a genome-wide scale<sup>4,5</sup>. The rationale of 3C-based approaches is that when a sufficient number of pairwise interaction frequencies are determined for a genomic region, chromosome or whole genome, its 3D arrangement can be inferred. 3C-based methods have been extensively reviewed and discussed elsewhere<sup>5,20–22</sup> and are summarized in BOX 1.

3C and 4C generate single interaction profiles for individual loci. For instance, 3C typically yields a long-range interaction profile of a selected gene promoter or other genomic element of interest versus surrounding chromatin (FIG. 1a), whereas 4C generates a genome-wide interaction profile for a single locus (FIG. 1b). These data sets can be represented as single tracks that can be plotted along the genome and compared to other genomic features such as DNase I hypersensitive sites (which are hallmarks of gene regulatory elements<sup>23</sup>) or genes. 5C and Hi-C methods are not anchored on a single locus of interest but instead generate matrices of interaction frequencies that can be represented as two-dimensional heat maps with genomic positions along the two axes (FIG. 1c,d).



**Figure 2 | Processes leading to close spatial proximity of loci.** Chromosome conformation capture (3C)-based technologies capture loci that are in close spatial proximity. Various biologically and structurally distinct examples are shown in which loci are in close spatial proximity. Analysis and interpretation of 3C data sets need to take these different scenarios into consideration.

### Interpreting chromatin interaction data

Before analysing chromatin interaction data, it is important to consider carefully what 3C-based assays capture (FIG. 2). These methods report on the relative frequency in the cell population by which two loci are in close spatial proximity, but they do not distinguish functional from non-functional associations nor do they reveal the mechanisms that led to their co-localization. Close spatial proximity can be the result of direct and specific contacts between two loci, mediated by protein complexes that bind them or can be the result of indirect co-localization of pairs of loci to the same subnuclear structure, such as the nuclear lamina, nucleolus or transcription factory. In addition, co-localization in a given cell can be a nonspecific result of the packing and folding of the chromatin fibre, as determined by other (nearby) specific long-range interactions or other constraints, or can be due to random (nonspecific) collisions in the crowded nucleus. Further, one of the defining features of chromosomes is that they are very long and flexible chromatin fibres. This feature — the polymer nature of chromosomes — also determines to a significant extent the frequency with which pairs of loci interact even in the absence of any specific higher-order structures<sup>24,25</sup>.

Finally, the precise 3D path of a chromatin fibre is highly variable even between otherwise identical cells and is locally dynamic (up to a megabase or so) within cells<sup>26,27</sup>. This explains why comprehensive chromatin interaction data sets typically show that a locus has some non-zero probability to interact with almost any other locus in the genome, although this probability of course widely varies, reflecting the overall nonrandom conformation of the genome<sup>24,25,28,29</sup>. Each instance of a chromatin interaction, or ligation product, that is detected represents an interaction involving a pair of loci in a single cell in the population. Thus, 3C interaction frequency data represent the fraction of cells in which pairs of loci are in close spatial proximity at the time the

cells are fixed, and the data can be understood only when genome folding displays enormous cell-to-cell heterogeneity (see REFS 28,29 and below). These considerations highlight the complex nature of comprehensive chromatin interaction data sets: the data represent the sum of interactions across a large cell population, and in each cell chromosome conformation is determined by many different constraints that act on the chromatin fibre.

Currently, the challenge of analysing chromosome conformation is shifting from developing experimental approaches for generating increasingly comprehensive and quantitative data sets to building analytical tools to interpret the interaction data. The first approach we consider is used to identify point-by-point looping interactions: for example, between promoters and gene regulatory elements.

### Linking regulatory elements to target genes

**Identifying looping interactions.** In genomes of Metazoa, each gene is surrounded by large numbers of elements<sup>1–3</sup>, and a major question is what principles determine which elements regulate any given gene at a given time. From detailed analyses of single genes over the past decade, and more comprehensive genome-wide studies reported more recently, the main mechanism by which regulatory elements communicate with their cognate target genes is through chromatin looping, which brings elements that are widely spaced in the linear genome into close spatial proximity<sup>30,31</sup>.

In many single-locus studies, classical 3C is used to quantify interaction frequencies between an element of interest — for example, a promoter — and flanking chromatin extending up to hundreds of kilobases (see example in FIG. 1a). Analysis of such ‘anchored’ interaction profiles can then point to distal loci that interact with the anchor locus more frequently than expected, suggesting a looping interaction (for example, see REFS 4,32–34). In general, it has been found that interaction frequencies

#### Nuclear lamina

A scaffold of lamin proteins predominantly found in the nuclear periphery associated with the inner surface of the nuclear membrane.

#### Transcription factory

A nuclear compartments in which active transcription takes place; it has a high concentration of RNA polymerase II.

#### Constraints

Forces (or scoring functions) that restrict the movement of objects (or points) that they apply to. Often used synonymously with ‘restraint’.

exponentially decay with increasing genomic distance. In many studies, looping interactions are inferred when a local peak is observed on top of the overall decaying baseline of interactions<sup>35</sup>. Most single-locus 3C analyses are qualitative in nature, and simple visual inspection of interaction profiles is used to identify peaks in interaction frequencies. Comparison of interaction profiles obtained in different cells or under different conditions can then provide further support, including statistical quantitative support, of the looping interaction when the long-range contact is condition- or cell-type-specific. FIGURE 1a shows a typical example of such looping interaction analysis of the *CFTR* locus<sup>34</sup>.

**Examples of looping at specific loci.** One of the best-studied examples is the long-range interaction between the locus control region (LCR) and the set of distal  $\beta$ -globin genes located 40–80 kb away. 3C studies in mouse and human detected prominent interactions between these elements in globin-expressing cells, and these interactions were significantly less frequent in cells that do not express these genes: for example, cells in the brain<sup>32,36</sup>. These interactions are mediated by specific transcription factors, including KLF1 and GATA1, that bind the LCR and the gene promoters<sup>37,38</sup>. Further, the looping interaction directly stimulates transcription by facilitating recruitment and phosphorylation of RNA polymerase II<sup>39</sup>. Looping has been found in many other cases in a range of species: for instance, the  $\alpha$ -globin genes<sup>40</sup>, the *CFTR* gene<sup>33,34</sup>, the interleukin gene cluster<sup>41</sup>, the *MYC* gene<sup>42,43</sup>, the major histocompatibility complex (MHC) II genes<sup>44</sup> and the yeast silent mating type loci *HML* and *HMR*<sup>45</sup>. Thus, chromatin looping constitutes a common mechanism by which gene regulatory elements control genes over large genomic distances.

### Comprehensive analysis of looping

**Analysing looping with 5C.** 5C has allowed more comprehensive analysis of chromatin looping for large numbers of genes by measuring many anchored interaction profiles in parallel (FIG. 1c). For example, in a recent study, interaction profiles for over 600 gene promoters were mapped in three human cell lines and at the resolution of single restriction fragments (~4 kb in size)<sup>46</sup>. The baseline of interaction frequencies could be estimated from the entire data set by assuming that the large majority of interrogated interactions were not specific looping interactions. This led to an estimate for the baseline interaction frequency for each genomic distance (FIG. 3a). Looping interactions were then identified by detection of signals that are significantly higher than this baseline, at a chosen *P* value and false discovery rate. This approach provides a more statistically rigorous analysis of identifying significant peaks on top of this baseline compared with classical 3C single-gene studies. A similar analysis was used for identification of sets of significant interactions in the yeast genome<sup>47</sup>. These approaches can identify pairs of loci that interact more frequently than expected, but they are limited by the models and assumptions that are used to define the expected interaction frequencies. Another limitation is that interaction frequencies are obtained in

arbitrary units, and thus the real interaction frequency in the examined cell population (that is, the percentage of cells in which the loci interact) remains unknown and can be quite low, as shown by fluorescence *in situ* hybridization (FISH; for example, see REFS 45, 48, 49). This makes it difficult to assess the functional role of these interactions in any given cell (see REF. 50 for more consideration of this issue).

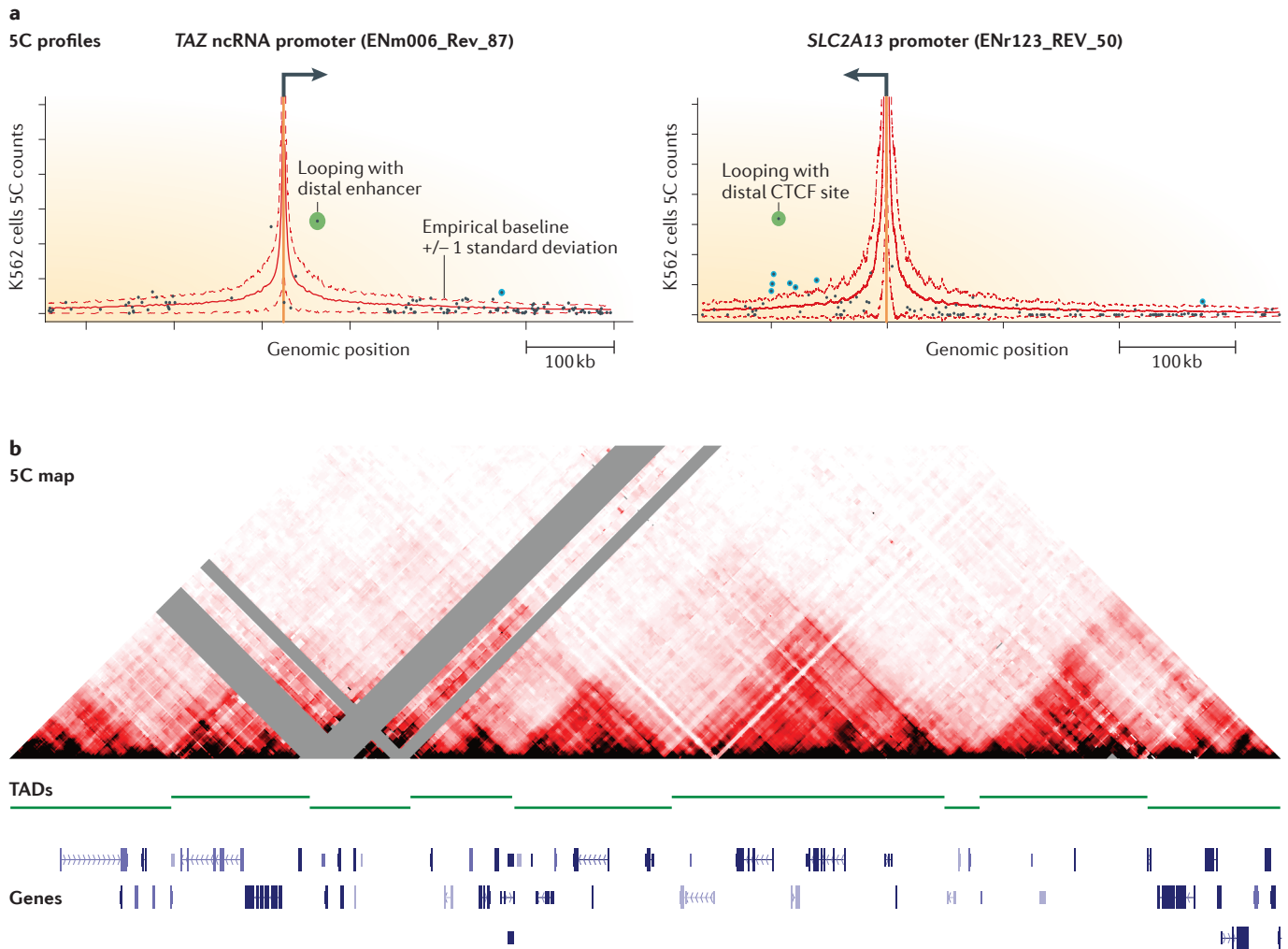
**Insights into looping landscapes.** Despite its limitations, comprehensive looping analyses are now starting to reveal common principles of long-range interactions involved in gene expression. A study of the human genome<sup>46</sup> identified thousands of significant long-range looping interactions between gene promoters and distal loci, reinforcing the notion that many, if not all, gene promoters engage with distal elements through looping. Analysis of this large set of looping interactions identified important general concepts of long-range gene regulation and also countered some long-held ideas. First, many of the looping events are cell-type-specific interactions between active gene promoters and distal elements resembling active enhancers; this is consistent with a role of these chromosome structures in gene activation. Second, one abundant class of long-range interactions involves promoters looping to sites bound by the insulator protein CTCF. The role of this class of looping interactions in gene regulation is not fully understood, but a general architectural role seems likely<sup>31,51,52</sup>. Third, regulatory elements are often assumed to regulate the nearest gene, even though previous genetic studies have provided examples in which this is not the case<sup>53</sup>. However, looping interactions often skip one or more genes, suggesting that the linear arrangement of genes and elements is a fairly poor predictor of their functional and structural interactions. Finally, relationships between genes and regulatory elements are far from exclusive: genes can interact with multiple distal elements, and elements can interact with multiple genes. Computational predictions based on correlations between gene activity and activity of distal elements across panels of cell lines also led to the prediction that genes are regulated by multiple distal elements<sup>54–56</sup>.

In addition, it was found that the average pattern of looping interactions around promoters is asymmetric: promoters interact with distal elements that can be located upstream or downstream of the transcription start site (TSS), but the most frequent looping interactions are observed with elements located ~120 kb upstream of the TSS. Why the looping landscapes of promoters display this asymmetry is not clear, but it may point to some form of directionality in the mechanism by which transcriptional looping interactions are formed.

From these studies, a picture emerges of chromosomes as highly complex 3D networks driven by long-range interactions. This view raises many new questions related to the processes that determine the specificity of gene–element interactions, the proteins that mediate them and how these looping interactions contribute to gene regulation.

**Locus control region (LCR).** A *cis*-acting element that organizes a gene cluster into an active chromatin domain and enhances transcription in a tissue-specific manner.

**CTCF**  
A highly conserved zinc finger protein that influences chromatin organization and architecture and is implicated in diverse regulatory functions, including transcriptional activation, repression and insulation.



**Figure 3 | Chromatin looping interactions and topologically associating domains.** **a** | Examples of long-range interaction profiles in the human genome, as determined by 5C. The orange vertical bar indicates the position of the gene promoters, the solid red line indicates the empirically estimated level of baseline interactions, and the dashed red lines indicate baseline plus or minus 1 standard deviation. The presence of a looping interaction is inferred when a pair of loci interact statistically more frequently than would be expected on the basis of the baseline frequency. The green data points represent significant looping interactions. Data are taken from REF. 46. **b** | A dense 5C interaction map of a 4.5 Mb region on the mouse X chromosome containing the X-chromosome inactivation centre. In red is the interaction frequency between pairs of loci, grey represents missing data due to low mappability. The interaction map is cut in half at the diagonal to facilitate alignment with genomic features. Visual inspection reveals the presence of triangles, which correspond to regions (topologically associating domains (TADs)) in which loci frequently interact with each other. Loci located in different TADs do not interact frequently. TAD boundaries have been determined by computationally determining the asymmetry between up- and downstream interactions around them<sup>59</sup>. ncRNA, non-coding RNA. Data are taken from REF. 58.

**Topologically associating domains**

Methods including 5C and Hi-C, which map all interactions in a genomic region of interest or in complete genomes in an unbiased fashion, can be analysed in various ways to identify structural features of chromosomes. One prominent feature of metazoan genomes is the formation of various types of chromosomal domains<sup>50</sup> (BOX 2). Studies using these approaches for *D. melanogaster*, mouse and human chromosomes have recently discovered that chromosomes are composed of discrete topologically associating domains (TADs), which can be hundreds of kilobases in size<sup>57–60</sup> (FIG. 3b).

Visual inspection of a high-resolution 5C interaction map of a 4.5 Mb region encompassing the mouse X-chromosome inactivation centre revealed a series of large structural domains<sup>58</sup>. Loci located within these TADs tend to interact frequently with each other, but they interact much less frequently with loci located outside their domain. This feature enabled researchers to identify TADs throughout the human and mouse genomes by analysing lower-resolution, but genome-wide, Hi-C interaction maps in combination with a hidden Markov model approach<sup>59</sup>. This analysis showed that TADs are universal building blocks of chromosomes<sup>59</sup>; the human

**X-chromosome inactivation centre**

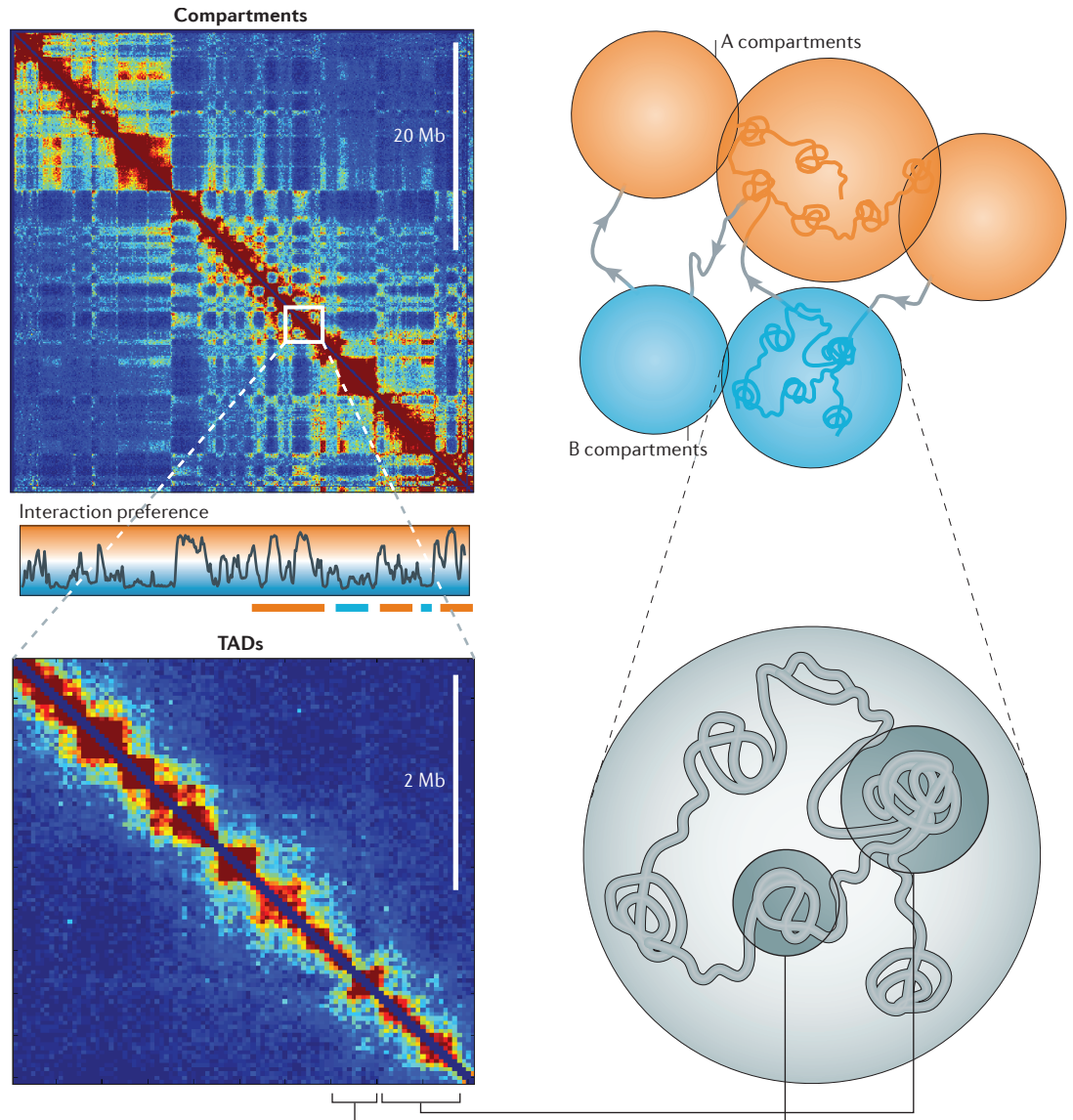
A genetically defined locus of several megabases on the X chromosome of mammals that is required to initiate transcriptional repression along a single X chromosome in female cells.

Box 2 | Genome compartments

Inter- and intrachromosomal interaction maps for mammalian genomes<sup>28,64,111</sup> have revealed a pattern of interactions that can be approximated by two compartments — A and B — that alternate along chromosomes and have a characteristic size of ~5 Mb each (as shown by the compartment graph below top heat map in the figure). A compartments (shown in orange) preferentially interact with other A compartments throughout the genome. Similarly, B compartments (shown in blue) associate with other B compartments. Compartment signal can be quantified by eigenvector expansion of the interaction map<sup>64,111,112</sup>. The A or B compartment signal is not simply biphasic (representing just two states) but is continuous<sup>112</sup> and correlates with indicators of transcriptional activity, such as DNA accessibility, gene density, replication timing, GC content and several histone marks. These indicators suggest that A compartments are largely euchromatic, transcriptionally active regions.

Topologically associating domains (TADs) are distinct from the larger A and B compartments. First, analysis of embryonic stem cells, brain tissue and fibroblasts suggests that most, but not all, TADs are tissue-invariant<sup>58,59</sup>, whereas A and B compartments are tissue-specific domains of active and inactive chromatin that are correlated with cell-type-specific gene expression patterns<sup>64</sup>. Second, A and B compartments are large (often several megabases) and form an alternating pattern of active and inactive domains along chromosomes. By contrast, TADs are smaller (median size around 400–500 kb; see zoomed in section of heat map in the figure) and can be active or inactive, and adjacent TADs are not necessarily of opposite chromatin status. Thus, it seems that TADs are hard-wired features of chromosomes, and groups of adjacent TADs can organize in A and B compartments (see REF. 50 for a more extensive discussion).

Shown in the figure are data for human chromosome 14 for IMR90 cells (data taken from REF. 59). In the top panel, Hi-C data were binned at 200 kb resolution, corrected using iterative correction and eigenvector decomposition (ICE), and the compartment graph was computed as described in REF. 112. The lower panel shows a blow up of a 4 Mb fragment of chromosome 14 (specifically, 74.4 Mb to 78.4 Mb) binned at 40 kb.



and mouse genomes are each composed of over 2,000 TADs, covering over 90% of the genome.

TADs are defined by genetically encoded boundary elements. This was directly demonstrated by deletion of a boundary between two TADs in the X-chromosome inactivation centre<sup>58</sup>, which led to partial fusion of the two flanking TADs. The two TADs did not fully merge, suggesting that a new boundary was activated. Further, genome-wide analysis of boundary regions indicated that they are enriched in CTCF-bound loci, although CTCF also frequently binds sites within TADs. This suggests that at least some CTCF-bound elements may indeed act as boundary elements, as has long been hypothesized<sup>51,61</sup>. However, CTCF-bound sites are certainly not the only genomic elements enriched near TAD boundaries<sup>59,60</sup>, and the mechanisms that establish TAD boundaries are still undefined.

The existence of TADs also suggests constraints on which looping interactions between genes and distal regulatory elements can occur. It is tempting to speculate that looping interactions would be limited to elements located within the same TAD. Indeed, an initial analysis in the mouse genome suggests that enhancer–promoter interactions are particularly frequent within TADs<sup>56</sup>. If correct, this would point to a major role for TADs in regulation of gene expression by limiting genes to only a certain set of distal regulatory elements. Consistent with this idea, analysis of the TADs in the X-chromosome inactivation centre showed that genes within the same TAD tend to be coordinately expressed during cell differentiation<sup>58</sup>, possibly because they share the same set of gene regulatory elements. The presence of TADs could provide a chromatin structural explanation for the long-standing observation that groups of neighbouring genes are often correlated in expression across cell types<sup>62,63</sup>.

### Building three-dimensional models of chromatin

Several analytical approaches are being developed that use comprehensive interaction data sets — not only those interactions that occur significantly more frequently than expected — to generate ensembles of 3D conformations of loci, chromosomes or whole genomes. These 3D representations can lead to the identification of higher-order features of chromosome conformation, such as formation of globular domains and chromosome territories, and may help to identify the sequence elements and processes that are involved in folding.

3D modelling approaches can be divided into roughly two types of methods. In the first approach — discussed in this section — a chromatin interaction data set is used to derive a population-averaged 3D conformation. In the second approach (discussed below), chromatin interaction data are analysed in statistical terms of polymer ensembles.

#### *Restraint-based three-dimensional model building.*

Comprehensive interaction maps reflect the population-averaged co-location frequencies of loci, which tend to be inversely related to average spatial distance (for example, see REFS 45,58,59,64). Interaction frequencies, or average spatial distances inferred from them,

can therefore be used as restraints to build 3D models that place loci in relative 3D space in a way that is most consistent with their interaction probabilities<sup>65</sup>. In this context, restraints refer to forces in the modelling that are applied to pairs of loci that will position them according to their average spatial distance, as inferred from their interaction frequency. Such approaches aim at finding 3D models of chromatin by treating them as a computational optimization problem. Therefore, optimal 3D models of genomic domains or genomes can be generated by minimizing a scoring function that is proportional to the violation of the imposed spatial restraints.

Generally, such 3D modelling follows an iterative process that cycles over four stages: information gathering (experiments), model representation and scoring, model optimization and model analysis (FIG. 4A). After experimental chromatin interaction or distance data have been obtained (usually by light microscopy or 3C-based methods), a genomic domain is then represented as a string of particles and spatial restraints between them<sup>66</sup>. Such representation needs to be adequate to the resolution of the input experimental data so that the use of the available information makes an exhaustive search of the 3D conformational space computationally feasible. For instance, the depth of DNA sequencing and size of the genomic region will determine the maximal resolution at which models can be built; the region is divided into the smallest particles that each still have sufficient long-range interaction data. For 5C data sets, each restriction fragment can be used as a particle, whereas for genome-wide data sets larger bins are often used: for example, 1 Mb for the human genome or 10–30 kb for the smaller genome of yeast. Next, it is necessary to determine a scoring function that will affect the spatial restraints between the particles. To this end, the experimental observations about the genomic domain or genome need to be translated into measurable relationships between the particles. The functional forms of restraints may be diverse to accommodate the integration of diverse sets of experimental observations: for example, real average distances between some of the loci as determined by light microscopy. After the system has been represented at the appropriate scale, and the relationships between the particles have been formulated on the basis of the observations, the final structure of the modelled object is obtained by minimizing the scoring function: that is, by simultaneously reducing the violations of all imposed restraints. The resulting algorithmically optimal models can be refined and further analysed using additional experimental observations that were not used during model building.

*Restraint-based modelling of genomic regions.* A pioneer implementation of restraint-based 3D modelling of a genomic domain was the spatial analysis of the human immunoglobulin H (*IGH*) locus using distance measurements obtained by light microscopic imaging of a set of 12 positions across the locus<sup>67</sup>. The resulting images were integrated with computational simulations to propose that the *IGH* locus is organized into compartments containing clusters of loops separated by linkers.

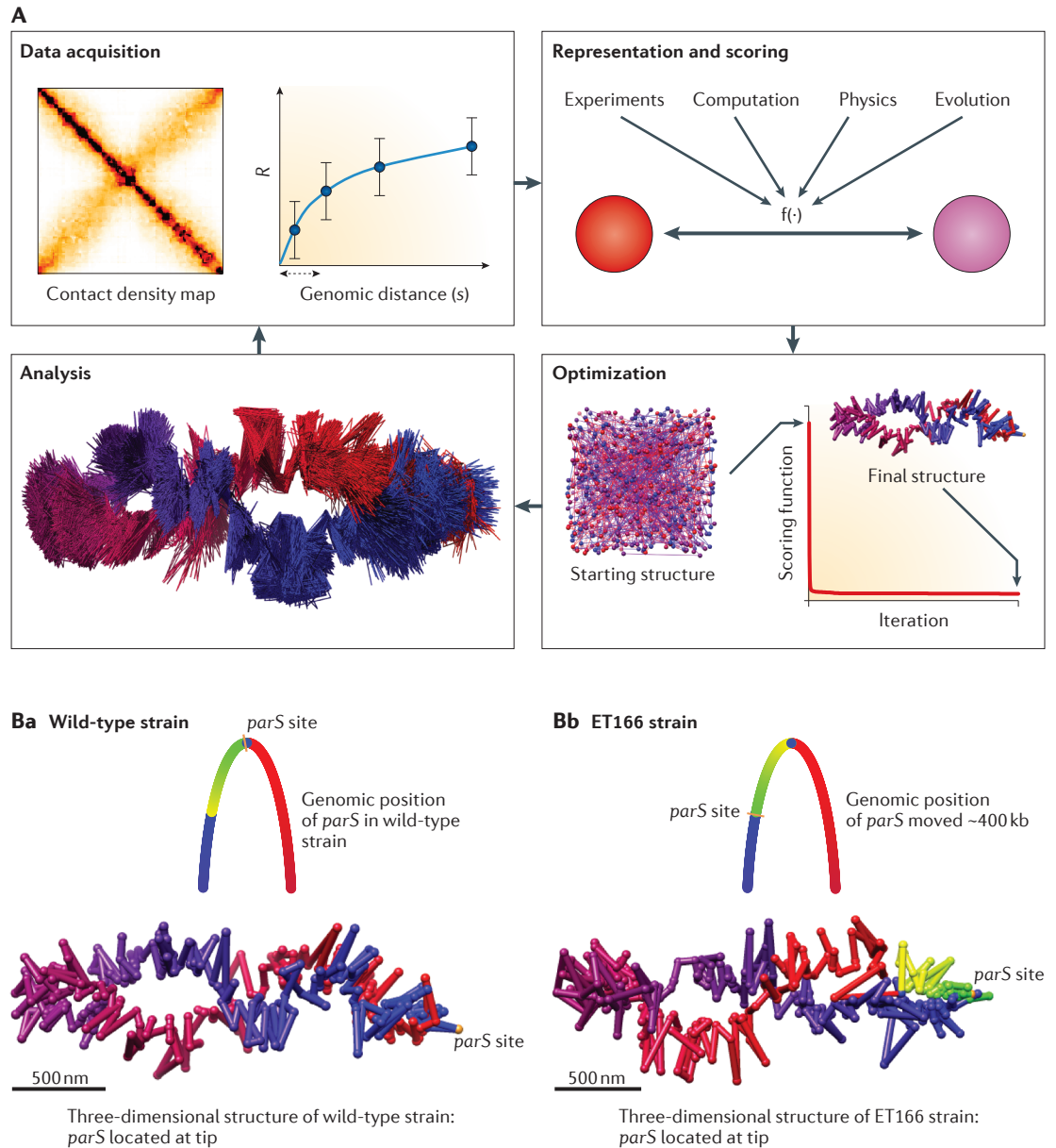
#### Boundary elements

DNA elements that lie between two gene-controlling elements, such as a promoter and an enhancer, or between two large chromosomal domains, preventing their communication or interaction. The function of boundary elements is usually mediated by the binding of specific factors.

#### Restraints

Forces (or scoring functions) that maintain the objects (or points) to which they apply at their position of equilibrium.





**Figure 4 | Three-dimensional modelling of genomes and genomic domains. A** | Iterative and integrative process for model building. The iterative process consists of data acquisition, model representation and scoring, model optimization and model analysis. **Ba** | A three-dimensional (3D) model of the wild-type *Caulobacter crescentus* genome, highlighting the position of the  $parS$  site located at the tip of the elliptical 3D structure of the genome. **Bb** | A 3D model of the ET166 strain of *C. crescentus* in which the  $parS$  site has been moved ~400 kb of its original locus (indicated in the schematic diagram of the genome). In the 3D structure of genome of the ET166 strain, the  $parS$  site is found at the tip of the structure again, which required a genome-wide rotation. The 3D models of *C. crescentus* are described in REF. 72. The models in this figure are reproduced, with permission, from REF. 72 © (2011) Elsevier.

Another study used a conceptually similar approach, but with 5C data, for analysis of the 3D organization of the human homeobox A (*HOXA*) gene cluster<sup>68</sup>. The models indicated that the chromatin conformation of the *HOXA* cluster changes during cell differentiation<sup>68</sup>. Also, 5C interaction maps for the human  $\alpha$ -globin region were used to build 3D models with the Integrative Modeling Platform<sup>69</sup>. The models demonstrated that long-range interactions among sets of widely spaced active functional

elements are sufficient to drive folding of local chromatin domains into compact globular states<sup>70,71</sup>. It is tempting to suggest that these globular conformations are related to TADs. The models also confirmed that the  $\alpha$ -globin genes were in close spatial proximity to their cognate long-range acting enhancers, as has been discovered from analysis of pairs of loci that interact more frequently than expected (as described above<sup>46</sup>). Importantly, the formation of globular domains could not readily be inferred

from analysis of only significant pairwise looping interactions and thus highlights how 3D model building helps to gain insights into higher-order chromosome structures beyond the formation of chromatin loops.

**Restraint-based modelling of genomes.** With the availability of high-resolution interaction maps for entire genomes, the first genome-wide 3D models were built on the basis of the same principles of data integration used previously to study genomic domains. The 3D structure of the *Caulobacter crescentus* genome was determined by combining genome-wide 5C chromatin interaction data, live-cell imaging and computational modelling<sup>72</sup>. The resulting models demonstrated that the bacterial genome is ellipsoidal with periodically arranged arms. The ellipsoidal structure predicted that specific *cis*-regulatory elements must be located at the tips of the arms, and further analyses showed that *parS* sequence elements have a role in chromosome folding<sup>72–74</sup> (FIG. 4B). This work provided one of the first examples in which structural analysis directly led to the identification of DNA elements involved in chromosome folding and suggests that structure–function studies, which are more typically done for proteins, may be feasible for whole chromosomes.

3D models have also been generated for several eukaryotic genomes, including the fission and budding yeast genomes<sup>29,47,75,76</sup> and, at a much lower resolution, the human genome<sup>28</sup>. The first budding yeast 3D genome model was a coarse-grained static snapshot of the genome, but it recapitulated the known features of its organization into a *Rabl* configuration and identified additional features, such as clustering of origins of early replication and tRNA genes<sup>47</sup>. A 3D model for the fission yeast genome was built using a genome-wide chromatin interaction data set<sup>75</sup> and showed a global genome organization that is similar to budding yeast, with prominent centromere clustering. Interestingly, the model revealed statistically significant interactions among highly expressed functionally related genes that may be reminiscent of the formation of transcription foci in the nuclei of mammalian cells<sup>29</sup>. These models all confirmed previously described characteristics of the yeast nucleus as observed in microscopic studies<sup>77,78</sup>, but importantly they demonstrated that a small set of spatial constraints is sufficient to yield a highly organized genome architecture<sup>29</sup>. A model of the human genome at low resolution based on a genome-wide chromatin interaction data set<sup>28</sup> and statistical analysis showed that nonspecific interchromosomal interactions are consistent with known architectural features.

Structural models of chromatin provide the opportunity to place linear annotations of the genome, such as positions of genes and gene regulatory elements, into a 3D context. Therefore, further developments in 3D model building will help to define the various levels of chromosome organization (including looping events, globules or TADs and higher-order compartments) to pinpoint sequence elements that determine these structures and to place widely spaced genomic loci in a spatial context that can reveal potentially functional long-range relationships.

## Polymer approaches

Although bottom-up restraint-based approaches have proved to be informative for building models of fairly stable chromosomal domains, top-down polymer approaches provide insight into statistical organizational features of folding states of chromosomes, their cell-to-cell variability and their dynamics within one cell. The application of polymer physics to chromosome research has a long history. Early works addressed such questions as the organization of interphase chromosomes, mechanisms of mitotic condensation, roles of topological constraints and DNA supercoiling<sup>79–86</sup>. Other studies have used simulations of polymer rings to suggest that chromosomal territories can be formed owing to topological constraints that prevent mixing<sup>87</sup> of individual chromosomes<sup>88–90</sup>. Polymer simulations are also being used to investigate how location of chromosomes can be influenced by properties of the chromatin fibre, its local folding and specific interactions between chromosomes<sup>88,91,92</sup>.

**The equilibrium globule state.** Several studies have sought to find a polymer model of interphase chromosomes that is consistent with FISH data on spatial distances between loci as a function of their genomic distances<sup>79,81,83</sup>. These studies considered equilibrium states of a homopolymer, such as a self-avoiding chain in a good solvent (known as a swollen coil), a non-interacting chain (known as an ideal chain) and a polymer in a poor solvent or that is externally confined (known as an equilibrium globule). The main feature of FISH data is a steady increase in the spatial distance with genomic separation of up to 10 Mb, followed by a more gradual increase or a plateau for genomic separation above 10 Mb. Earlier studies suggested that these features are consistent with the equilibrium globule<sup>79</sup>. Recent studies invoked much more complex models of polymer condensation<sup>93–95</sup>, which essentially leads to the equilibrium globule state. These models qualitatively reproduce the ‘rise-and-plateau’ pattern but otherwise fit FISH data rather poorly<sup>94</sup>. Available FISH data, however, are limited to a small number of loci and suffer from large cell-to-cell variability, making it hard to differentiate between these polymer models and alternatives discussed below.

**Interpretation of interaction data using polymer physics.** With the emergence of 3C methods, approaches of polymer physics are being used to rationalize measured probabilities of spatial interactions<sup>4,64,96</sup>. Measured contact frequencies are used to determine and to characterize the ensemble of chromatin conformations. The first question to be asked is whether conformations of a chromosomal locus are all similar to each other, like conformations of a single protein folded into native structure, or as diverse as conformations of a random polymer coil. Hi-C data show a lack of specific contacts among loci >1 Mb apart, whereas specific interactions are detected at smaller scales (for example, TADs and loops between genes and regulatory elements generally involve loci separated by <1 Mb) (FIG. 5a). The absence of reproducible contacts at larger length scales makes

### Rabl configuration

A pattern of nuclear organization in which centromeres of all chromosomes are spatially clustered and their arms run in parallel. This organization has been proposed to be a passive consequence of chromosome segregation but can also be actively maintained by mechanisms that cluster centromeres.

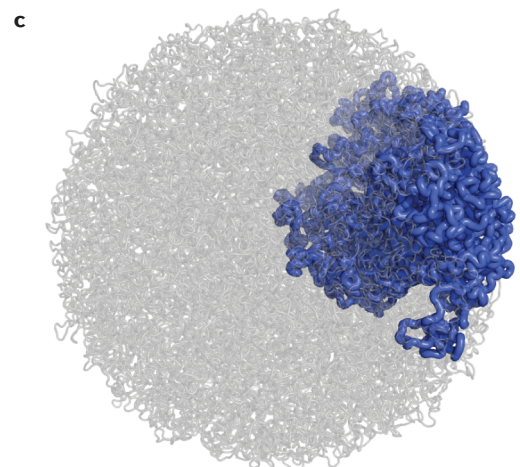
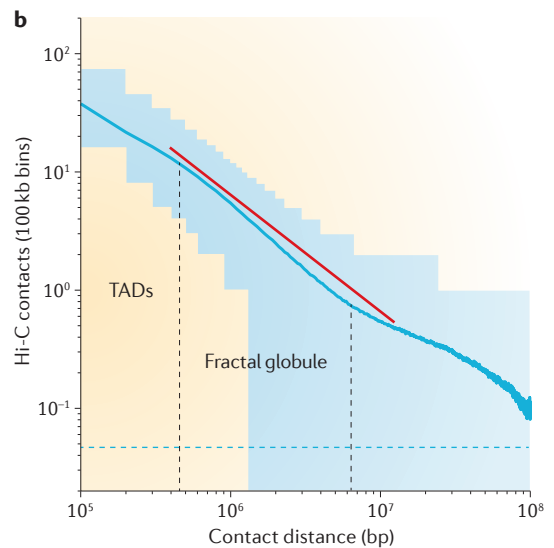
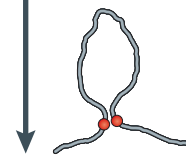
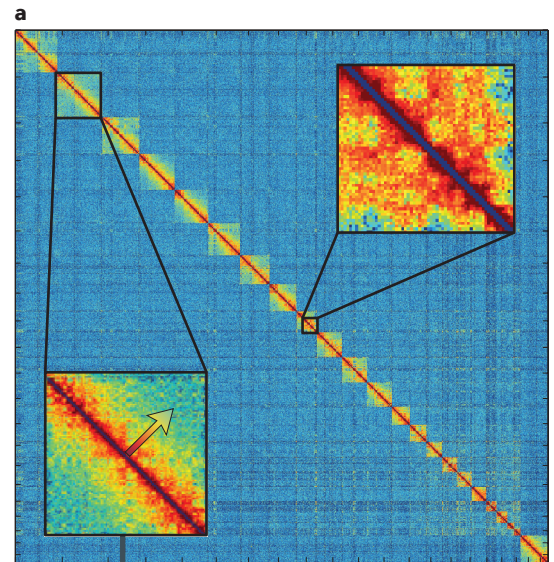
higher-order chromosome conformations very different from conformations of a single folded protein, suggesting that chromatin at large scales (>1 Mb) can be better characterized as a statistical ensemble of diverse conformations, probably reflecting differences between individual cells that collectively possess some specific statistical, spatial or topological properties.

**Contact probability and genomic distance: the fractal globule.** Interactions within single chromosomal arms exhibit a striking 100-fold decrease of the contact probability  $P$  with genomic distance  $s$ , making it the most prominent feature of intrachromosomal interactions. Hi-C data for non-synchronized human cells<sup>64</sup> show three regimes that each exhibit a decline in the contact probability (FIG. 5a,b): first, a shallow decline for  $s < 0.7$  Mb, corresponding to TADs<sup>58</sup>; second, a steeper decline of the contact probability for  $0.7 \text{ Mb} < s < 10 \text{ Mb}$ , corresponding to some globular organization of chromatin; and, third, a shallow decline at distance  $s > 10 \text{ Mb}$ , but at these distances the interaction frequencies are very low, so the statistics are not robust. Importantly, the intermediate regime  $0.7 \text{ Mb} < s < 10 \text{ Mb}$  is characterized by a power-law scaling,  $P(s) \sim s^{-1}$ , of the contact probability.

The power-law scaling of the contact probability is not surprising, as contact probability in many polymer systems follows power-law dependencies, and the exact value of the power is indicative of the polymer state (that is, a specific ensemble of configurations). Polymer simulations can be used to build various conformational ensembles. In these simulations, chromatin is represented by a 10 nm fibre with one monomer corresponding to 2–5 nucleosomes<sup>64</sup>. A 10 Mb region is modelled by thousands of monomers that have excluded volume, are connected into a polymer chain and are subjected to external forces and constraints. The folding and dynamics of the fibre is simulated by Monte Carlo or Brownian dynamics; these are standard simulation techniques in which each monomer experiences forces acting on it, including random Brownian fluctuations, and moves in response to these forces. An ensemble of obtained conformations is used to calculate a map of contact frequency, and its features — for example,  $P(s)$  — are compared with those of experimental Hi-C maps.

Simulations and theory demonstrated that the scaling observed in Hi-C for 0.7–10 Mb range is inconsistent with the equilibrium states (that is, conformational ensembles) of a homopolymer, such as the ideal chain, the swollen coil and the equilibrium globule. A non-equilibrium state called the fractal globule, which was conjectured in 1988 (REF. 97), was simulated and found to recapitulate contact probability<sup>64,98</sup>. These simulations studied condensation of a chromatin fibre of 4–16 Mb, which was represented by a polymer chain of  $N=4,000$ – $32,000$  monomers. Such long chains are essential to capture statistical properties of polymers.

The fractal globule, which emerges as a result of polymer condensation during which topological constraints prevent knotting and slow down equilibration of the polymer, has a number of important properties. First, dense and uniform packing of chromatin at the scale of



**Fractal globule**

A dense, non-equilibrium polymer state, which emerges as a result of a polymer condensation. In this state, the polymer is unknotted and each region of the chain is locally compact, allowing easy opening and closing of chromosomal regions.

◀ **Figure 5 | Large-scale features of genome folding.** **a** | Whole-genome map of relative contact probabilities obtained by Hi-C (normalized by iterative correction and eigenvector decomposition (ICE))<sup>112</sup>. Insets show two of the most prominent features: intrachromosomal decline of the contact probability; and a compartment pattern of interactions observed inter- and intrachromosomally. **b** | Contact probability  $P(s)$  as a function of genomic separation  $s$ . The mean contact probability for each separation is shown by the blue line, with the distribution shown by 75% quantiles in light blue. The red line shows  $P(s) \sim s^{-1}$  scaling. Two characteristic regimes corresponding to topologically associating domains (TADs; <0.7 Mb) and the fractal globule (between 0.7 and 7 Mb) are labelled. **c** | Polymer model of the fractal globule of 10 Mb of a genome (one monomer representing two nucleosomes), with a 1 Mb region shown in blue, illustrating its compactness within the globule. Part **a** of the figure is modified, with permission, from REF. 112 © (2012) Macmillan Publishers Ltd. All rights reserved.

<10 Mb is consistent with observed chromatin globules of about 1  $\mu\text{m}$  in diameter<sup>99</sup> (assuming a realistic 5–10% chromatin volume density). Second, the unknotted conformation of the fractal globules (which is not a feature of equilibrium globules) allows easy opening and closing or translocation of chromosomal regions over large distances in the nucleus<sup>100,101</sup>. Third, dense packing of segments of the fractal globule implies that continuous regions of the genome (in the size range 1–10 Mb) are compactly folded (FIG. 5c) rather than being spread. This property distinguishes the fractal globule from the equilibrium globule, where individual segments are akin to random walks: that is, they are extended and intermixed. Although available FISH data do not allow to discriminate between these models<sup>102</sup>, staining of continuous genomic regions<sup>103</sup> shows a great deal of compactness and little intermixing, strongly supporting the fractal globule. One of the limitations of the original fractal globule model is that the fractal globule is formed during condensation, rather than decondensation of the chromatin from mitotic chromosomes. However, it has been demonstrated that a similar organization could emerge when several initially condensed chromosomes were allowed to decondense into the nuclear volume<sup>90</sup>. It has been suggested that topological interactions between chromosomes prevent their mixing and equilibration during biologically relevant timescales<sup>90</sup>. The fractal globule state can also emerge as an equilibrium state of a polymer ring in a melt of other such rings<sup>89</sup>, in which rings model stable chromatin loops. What unites all of these models is a central role of topological constraints in ‘crumpling’ interphase chromatin.

Another study that aimed to explain the scaling of Hi-C contact probability used an equilibrium homopolymer model and suggested that the fractal globule

emerges in equilibrium, at the transition between the open and condensed states<sup>95</sup>. This result, however, contradicts well-known facts in polymer physics<sup>104</sup> and is probably a result of a poor statistics owing to very short chains ( $N=512$ ) used in simulations.

Connections of the fractal globule conformation to much smaller TADs and much larger chromosomal territories and compartments are yet to be established. It also remains to be seen whether the fractal globule is susceptible to slow ‘melting’ over long times or because of topoisomerase II enzyme activity, or whether specific biological mechanisms are responsible for its maintenance.

### Future perspective

In the coming years, we can expect a wealth of chromatin interaction data to become available. With expected further increases in sequencing capacity and reduction in cost, chromatin interaction maps will become available for even the largest genomes at increasing resolution. Analysing these data sets will become the major challenge, requiring new developments in bioinformatics, computational biology and biophysics. The approaches described here are only a starting point, and we envision a rapid expansion in efforts to explore the 3D folding of chromosomes and the effects on the biology of genomes. Further improvements in both experimental and computational data analysis approaches will facilitate addressing several important questions that the field of genome regulation is currently grappling with. For instance, most 3C-based studies do not directly allow measurement of the dynamics and cell-to-cell variation in chromosome folding, and thus it is currently largely unknown how stable looping interactions and chromatin domains are within individual cells or how stochastic they are between cells. Further, the relative contributions of genomic sequence and transcriptional activity in establishing the compartmentalized architecture of chromosomes are yet to be determined. The roles of lamina association, direct or mediated colocalization of transcribed regions and other molecular mechanisms shaping activity-associated organization of the nucleus need to be established. Finally, we still know little about how chromosome structure changes during development and on perturbation (for example, as cells respond to signals), and how chromosomes fold and unfold during the cell cycle. With the rapid technological developments in this field, we may get some answers to these questions in the years ahead.

1. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
2. Nègre, N. *et al.* A cis-regulatory map of the *Drosophila* genome. *Nature* **471**, 527–531 (2011).
3. Gerstein, M. B. *et al.* Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE project. *Science* **330**, 1775–1787 (2010).
4. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306–1311 (2002).
5. van Steensel, B. & Dekker, J. Genomics tools for unraveling chromosome architecture. *Nature Biotech.* **28**, 1089–1095 (2010).
6. Müller, I., Boyle, S., Singer, R. H., Bickmore, W. A. & Chubb, J. R. Stable morphology, but dynamic internal reorganisation, of interphase human chromosomes in living cells. *PLoS ONE* **5**, e11560 (2010).
7. Boyle, S., Rodesch, M. J., Halvensleben, H. A., Jeddelloh, J. A. & Bickmore, W. A. Fluorescence *in situ* hybridization with high-complexity repeat-free oligonucleotide probes generated by massively parallel synthesis. *Chromosome Res.* **19**, 901–909 (2011).
8. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Rev. Genet.* **2**, 292–301 (2001).
9. Branco, M. R. & Pombo, A. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS Biol.* **4**, e138 (2006).
10. Iborra, F. J., Pombo, A., Jackson, D. A. & Cook, P. R. Active RNA polymerases are localized within discrete transcription ‘factories’ in human nuclei. *J. Cell Sci.* **109**, 1427–1436 (1996).
11. Fraser, P. & Bickmore, W. Nuclear organization of the genome and the potential for gene regulation. *Nature* **447**, 413–417 (2007).
12. Brown, J. M. *et al.* Association between active genes occurs at nuclear speckles and is modulated by chromatin environment. *J. Cell Biol.* **182**, 1083–1097 (2008).

13. Schoenfelder, S. *et al.* Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genet.* **42**, 53–61 (2010).
14. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948–951 (2008).
15. Németh, A. *et al.* Initial genomics of the human nucleolus. *PLoS Genet.* **6**, e1000889 (2010).
16. van Koningsbruggen, S. *et al.* High-resolution whole-genome sequencing reveals that specific chromatin domains from most human chromosomes associate with nucleoli. *Mol. Biol. Cell.* **21**, 3735–3748 (2010).
17. Tolhuis, B. *et al.* Interactions among Polycomb domains are guided by chromosome architecture. *PLoS Genet.* **7**, e1001543 (2011).
18. Bantignies, F. *et al.* Polycomb-dependent regulatory contacts between distant *Hox* loci in *Drosophila*. *Cell* **144**, 214–226 (2011).
19. Pirrotta, V. & Li, H. B. A view of nuclear Polycomb bodies. *Curr. Opin. Genet. Dev.* **22**, 101–109 (2012).
20. de Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26**, 11–24 (2012).
21. Hakim, O. & Misteli, T. SnapShot: chromosome conformation capture. *Cell* **148**, 1068–1068.e2 (2012).
22. Ethier, S. D., Miura, H. & Dostie, J. Discovering genome regulation with 3C and 3C-related technologies. *Biochim. Biophys. Acta.* **1819**, 401–410 (2012).
23. Felsenfeld, G. & Groudine, M. Controlling the double helix. *Nature* **421**, 448–453 (2003).
24. Rippe, K. Making contacts on a nucleic acid polymer. *Trends Biochem. Sci.* **26**, 733–740 (2001).
25. Fudenberg, G. & Mirny, L. A. Higher-order chromatin structure: bridging physics and biology. *Curr. Opin. Genet. Dev.* **22**, 115–124 (2012).
26. Chubb, J. R., Boyle, S., Perry, P. & Bickmore, W. A. Chromatin motion is constrained by association with nuclear compartments in human cells. *Curr. Biol.* **12**, 439–445 (2002).
27. Marshall, W. F. *et al.* Interphase chromosomes undergo constrained diffusional motion in living cells. *Curr. Biol.* **7**, 930–939 (1997).
28. Kalhor, R., Tjong, H., Jayatilaka, N., Alber, F. & Chen, L. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature Biotech.* **30**, 90–98 (2011).
- These authors apply simulations to analyse genome-wide chromatin interaction data to generate spatial models of nuclear organization that also capture the cell-to-cell variability in chromosome organization.**
29. Tjong, H., Gong, K., Chen, L. & Alber, F. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res.* **22**, 1295–1305 (2012).
30. Miele, A. & Dekker, J. Long-range chromosomal interactions and gene regulation. *Mol. Biosyst.* **4**, 1046–1057 (2008).
31. Krivega, I. & Dean, A. Enhancer and promoter interactions—long distance calls. *Curr. Opin. Genet. Dev.* **22**, 79–85 (2012).
32. 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).
33. Ott, C. J. *et al.* Intronic enhancers coordinate epithelial-specific looping of the active *CFTR* locus. *Proc. Natl Acad. Sci. USA* **106**, 19934–19939 (2009).
34. Gheldof, N. *et al.* Cell-type-specific long-range looping interactions identify distant regulatory elements of the *CFTR* gene. *Nucleic Acids Res.* **38**, 4235–4336 (2010).
35. Dekker, J. The 3C's of chromosome conformation capture: controls, controls, controls. *Nature Methods* **3**, 17–21 (2006).
36. Palstra, R. J. *et al.* The  $\beta$ -globin nuclear compartment in development and erythroid differentiation. *Nature Genet.* **35**, 190–194 (2003).
37. Drissen, R. *et al.* The active spatial organization of the  $\beta$ -globin locus requires the transcription factor EKLF. *Genes Dev.* **18**, 2485–2490 (2004).
38. Vakoc, C. R. *et al.* Proximity among distant regulatory elements at the  $\beta$ -globin locus requires GATA-1 and FOG-1. *Mol. Cell* **17**, 453–462 (2005).
39. Deng, W. *et al.* Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233–1244 (2012).
- In this work, the authors show that physical tethering of an enhancer to its target promoter can activate the gene, providing one of the first direct mechanistic insights into the role of chromatin looping in gene control.**
40. Vernimmen, D., De Gobbi, M., Sloane-Stanley, J. A., Wood, W. G. & Higgs, D. R. Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J.* **26**, 2041–2051 (2007).
41. Spilianakis, C. G. & Flavell, R. A. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nature Immunol.* **5**, 1017–1027 (2004).
42. Ahmadiyeh, N. *et al.* 8q24 prostate, breast, and colon cancer risk loci show tissue-specific long-range interaction with MYC. *Proc. Natl Acad. Sci. USA* **107**, 9742–9746 (2010).
43. Wright, J. B., Brown, S. J. & Cole, M. D. Upregulation of c-MYC in *cis* through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells. *Mol. Cell Biol.* **30**, 1411–1420 (2010).
44. Majumder, P., Gomez, J. A., Chadwick, B. P. & Boss, J. M. The insulator factor CTCF controls MHC class II gene expression and is required for the formation of long-distance chromatin interactions. *J. Exp. Med.* **205**, 785–798 (2008).
45. Miele, A., Bystrycky, K. & Dekker, J. Yeast silent mating type loci form heterochromatic clusters through silencer protein-dependent long-range interactions. *PLoS Genet.* **5**, e1000478 (2009).
46. Sanyal, A., Lajoie, B. R., Jain, G. & Dekker, J. The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).
- In this paper, thousands of long-range interactions across 30 Mb in the human genome are discovered. This paper describes some of the statistical approaches that can be used to identify significant locus–locus interactions in comprehensive chromatin interaction data sets.**
47. Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nature* **465**, 363–367 (2010).
48. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nature Genet.* **38**, 1348–1354 (2006).
49. Hakim, O. *et al.* Diverse gene reprogramming events occur in the same spatial clusters of distal regulatory elements. *Genome Res.* **21**, 697–706 (2011).
50. Gibcus, J. H. & Dekker, J. The hierarchy of the 3D genome. *Mol. Cell* **49**, 773–782 (2013).
51. Phillips, J. E. & Corces, V. G. CTCF: master weaver of the genome. *Cell* **137**, 1194–1211 (2009).
52. Handoko, L. *et al.* CTCF-mediated functional chromatin interactome in pluripotent cells. *Nature Genet.* **43**, 630–638 (2011).
53. Kleinjan, D. A. & van Heyningen, V. Long-range control of gene expression: emerging mechanisms and disruption in disease. *Am. J. Hum. Genet.* **76**, 8–32 (2005).
54. Gerstein, M. B. *et al.* Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**, 91–100 (2012).
55. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
56. Shen, Y. *et al.* A map of the *cis*-regulatory sequences in the mouse genome. *Nature* **488**, 116–120 (2012).
57. Sexton, T. *et al.* Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
58. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
- This paper describes the discovery of TADs using 5C and shows that TAD boundaries are independent of chromatin modification but are defined by genetic *cis*-elements.**
59. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
- This paper describes the discovery of TADs and discusses a computational strategy to identify TAD boundaries using Hi-C data sets.**
60. Hou, C., Li, L., Qin, Z. S. & Corces, V. G. Gene density, transcription, and insulators contribute to the partition of the *Drosophila* genome into physical domains. *Mol. Cell* **48**, 471–484 (2012).
61. Gaszner, M. & Felsenfeld, G. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nature Rev. Genet.* **7**, 703–713 (2006).
62. Caron, H. *et al.* The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science* **291**, 1289–1292 (2001).
63. Spellman, P. T. & Rubin, G. M. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol.* **1**, 5 (2002).
64. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
- This work describes development of the Hi-C method and how polymer simulations can be used to analyse chromatin interaction data. This work also described the fractal globule state of chromatin at the 1–10 Mb scale.**
65. Marti-Renom, M. A. & Mirny, L. A. Bridging the resolution gap in structural modeling of 3D genome organization. *PLoS Comput. Biol.* **7**, e1002125 (2011).
66. Baù, D. & Marti-Renom, M. A. Structure determination of genomic domains by satisfaction of spatial restraints. *Chromosome Res.* **19**, 25–35 (2011).
67. Jhunjhunwala, S. *et al.* The 3D structure of the immunoglobulin heavy-chain locus: implications for long-range genomic interactions. *Cell* **133**, 265–279 (2008).
- This worked combined FISH data and polymer modelling to obtain spatial models for the immunoglobulin heavy-chain locus.**
68. Fraser, J. *et al.* Chromatin conformation signatures of cellular differentiation. *Genome Biol.* **10**, R37 (2009).
69. Russel, D. *et al.* Putting the pieces together: integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* **10**, e1001244 (2012).
70. Sanyal, A., Baù, D., Marti-Renom, M. A. & Dekker, J. Chromatin globules: a common motif of higher order chromosome structure? *Curr. Opin. Cell Biol.* **23**, 325–331 (2011).
71. Baù, D. *et al.* The three-dimensional folding of the  $\alpha$ -globin gene domain reveals formation of chromatin globules. *Nature Struct. Mol. Biol.* **18**, 107–114 (2011).
- These authors describe a restraint-based modelling approach to use chromatin interaction data to derive spatial models of chromatin domains.**
72. Umbarger, M. A. *et al.* The three-dimensional architecture of a bacterial genome and its alteration by genetic perturbation. *Mol. Cell* **44**, 252–264 (2011).
73. Ebersbach, G., Briegel, A., Jensen, G. J. & Jacobs-Wagner, C. A self-associating protein critical for chromosome attachment, division, and polar organization in *Caulobacter*. *Cell* **134**, 956–968 (2008).
74. Bowman, G. R. *et al.* A polymeric protein anchors the chromosomal origin/ParB complex at a bacterial cell pole. *Cell* **134**, 945–955 (2008).
75. Tanizawa, H. *et al.* Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic Acids Res.* **38**, 8164–8177 (2010).
76. Hu, M. *et al.* Bayesian inference of spatial organizations of chromosomes. *PLoS Comput. Biol.* **9**, e1002893 (2013).
77. Jin, Q. W., Fuchs, J. & Loidl, J. Centromere clustering is a major determinant of yeast interphase nuclear organization. *J. Cell Sci.* **113**, 1903–1912 (2000).
78. Taddei, A. & Gasser, S. M. Structure and function in the budding yeast nucleus. *Genetics* **192**, 107–129 (2012).
79. van den Engh, G., Sachs, R. & Trask, B. J. Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. *Science* **257**, 1410 (1992).
80. McManus, J. *et al.* Unusual chromosome structure of fission yeast DNA in mouse cells. *J. Cell Sci.* **107**, 469–486 (1994).
81. Hahnfeldt, P. *et al.* Polymer models for interphase chromosomes. *Proc. Natl Acad. Sci. USA* **90**, 7854–7858 (1993).
82. Marko, J. F. & Siggia, E. D. Polymer models of meiotic and mitotic chromosomes. *Mol. Biol. Cell* **8**, 2217–2231 (1997).
83. Sachs, R. K. *et al.* A random-walk/giant-loop model for interphase chromosomes. *Proc. Natl Acad. Sci. USA* **92**, 2710–2714 (1995).
84. Sikorav, J. L. & Jannink, G. Kinetics of chromosome condensation in the presence of topoisomerases: a phantom chain model. *Biophys. J.* **66**, 827 (1994).

85. Grosberg, A., Rabin, Y., Havlin, S. & Neer, A. Crumpled globule model of the three-dimensional structure of DNA. *Europhys. Lett.* **23**, 373 (1993).
86. Vologodskii, A. V., Levene, S. D., Klenin, K. V., Frank-Kamenetskii, M. & Cozzarelli, N. R. Conformational and thermodynamic properties of supercoiled DNA. *J. Mol. Biol.* **227**, 1224–1243 (1992).
87. Bohn, M. & Heermann, D. W. Repulsive forces between looping chromosomes induce entropy-driven segregation. *PLoS ONE* **6**, e14428 (2011).
88. Dorier, J. & Stasiak, A. The role of transcription factories-mediated interchromosomal contacts in the organization of nuclear architecture. *Nucleic Acids Res.* **38**, 7410–7421 (2010).
89. Vettorel, T., Grosberg, A. Y. & Kremer, K. Statistics of polymer rings in the melt: a numerical simulation study. *Phys. Biol.* **6**, 025013 (2009).
90. Rosa, A. & Everaers, R. Structure and dynamics of interphase chromosomes. *PLoS Comput. Biol.* **4**, e1000153 (2008).
91. Cook, P. R. & Marenduzzo, D. Entropic organization of interphase chromosomes. *J. Cell Biol.* **186**, 825–834 (2009).
92. Jerabek, H. & Heermann, D. W. Expression-dependent folding of interphase chromatin. *PLoS ONE* **7**, e37525 (2012).
93. Bohn, M. & Heermann, D. W. Diffusion-driven looping provides a consistent framework for chromatin organization. *PLoS ONE* **5**, e12218 (2010).
94. Mateos-Langerak, J. *et al.* Spatially confined folding of chromatin in the interphase nucleus. *Proc. Natl Acad. Sci. USA* **106**, 3812–3817 (2009).
95. Barbieri, M. *et al.* Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl Acad. Sci.* **109**, 16173–16178 (2012).
96. Rosa, A., Becker, N. B. & Everaers, R. Looping probabilities in model. Interphase chromosomes. *Biophys. J.* **98**, 2410–2419 (2010).
97. Grosberg, A. Y., Nechaev, S. K. & Shakhnovich, E. I. The role of topological constraints in the kinetics of collapse of macromolecules. *J. Physique* **49**, 2095–2100 (1988).
98. Mirny, L. A. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res.* **19**, 37–51 (2011).
99. Rapkin, L. M., Anchel, D. R. P., Li, R. & Bazett-Jones, D. P. A view of the chromatin landscape. *Micron* **43**, 150–158 (2012).
100. Belmont, A. S. *et al.* Insights into interphase large-scale chromatin structure from analysis of engineered chromosome regions. *Cold Spring Harbor Symp. Quant. Biol.* **75**, 453–460 (2011).
101. Towbin, B. D. *et al.* Step-wise methylation of histone h3k9 positions heterochromatin at the nuclear periphery. *Cell* **150**, 934–947 (2012).
102. Emanuel, M., Radja, N. H., Henriksson, A. & Schiessel, H. The physics behind the larger scale organization of DNA in eukaryotes. *Phys. Biol.* **6**, 025008 (2009).
103. Shopland, L. S. *et al.* Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence. *J. Cell Biol.* **174**, 27–38 (2006).
104. Rubinstein, M. & Colby, R. H. *Polymer Physics* (Oxford Univ. Press, 2003).
105. Würtele, H. & Chartrand, P. Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended chromosome conformation capture methodology. *Chromosome Res.* **14**, 477–495 (2006).
106. Zhao, Z. *et al.* Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genet.* **38**, 1341–1347 (2006).
107. 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).
108. Lajoie, B. R., van Berkum, N. L., Sanyal, A. & Dekker, J. My5C: web tools for chromosome conformation capture studies. *Nature Methods* **6**, 690–691 (2009).
109. Horike, S., Cai, S., Miyano, M., Cheng, J. F. & Kohwi-Shigematsu, T. Loss of silent-chromatin looping and impaired imprinting of DLX5 in Rett syndrome. *Nature Genet.* **37**, 31–40 (2005).
110. Fullwood, M. J. *et al.* An oestrogen-receptor- $\alpha$ -bound human chromatin interactome. *Nature* **462**, 58–64 (2009).
111. Zhang, Y. *et al.* Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**, 908–921 (2012).
112. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods* **9**, 999–1003 (2012).
113. Mouse ENCODE Consortium. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biol.* **13**, 418 (2012).

#### Acknowledgements

We are grateful to all the members of our groups for many discussions about three-dimensional genomics. Supported by grants from the US National Institutes of Health (NIH), US National Human Genome Research Institute (HG003143 and HG003143-06S1) and a W.M. Keck Foundation distinguished young scholar in medical research grant to J.D.; financial support from the Spanish Ministerio de Ciencia e Innovación (BFU2010-19310/BMC), the Human Frontiers Science Program (RGP0044/2011) and the BLUEPRINT project (EU FP7 grant agreement 282510) to M.A.M.-R.; and a grant from the NIH National Cancer Institute (Physical Sciences–Oncology Center at Massachusetts Institutes of Technology Grant U54CA143874) to L.A.M.

#### Competing interests statement

The authors declare no competing financial interests.

#### FURTHER INFORMATION

Job Dekker's homepage: <http://3dg.umassmed.edu/welcome/welcome.php>

Marc A. Marti-Renom's homepage: <http://marciuslab.org>

Leonid A. Mirny's homepage: <http://mirnylab.mit.edu>

ALL LINKS ARE ACTIVE IN THE ONLINE PDF