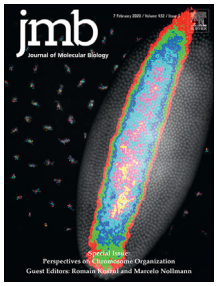

# Higher-order Chromosome Structures Investigated by Polymer Physics in Cellular Morphogenesis and Differentiation

**Andrea Esposito, Andrea M. Chiariello, Mattia Conte, Luca Fiorillo, Francesco Musella, Renato Sciarretta and Simona Bianco**

***Dipartimento di Fisica,*** *Università di Napoli Federico II, and INFN Napoli, Complesso Universitario di Monte Sant'Angelo, 80126, Naples, Italy*

***Correspondence to Andrea Esposito and Simona Bianco:*** andresposito@na.infn.it, biancos@na.infn.it
https://doi.org/10.1016/j.jmb.2019.12.017
***Edited by Marcelo Nollmann***

## Abstract

Experimental advances in Molecular Biology demonstrated that chromatin architecture and gene regulation are deeply related. Hi-C data, for instance, returned a scenario where chromosomes form a complex pattern of interactions, including TADs, metaTADs, and compartments, correlated with genomic and epigenomic features. Here, we discuss the emerging hierarchical organization of chromatin and show how it remains partially conserved during mouse neuronal differentiation with changes highly related to modifications in gene expression. In this scenario, models of polymer physics, such as the Strings & Binders (SBS) model, can be a crucial instrument to understand the molecular mechanisms underlying the formation of such a higher order 3D structure. In particular, we focus on the case study of the murine *Pitx1* genomic region. At this locus, two alternative spatial conformations take place in the hindlimb and forelimb tissues, corresponding to two different transcriptional states of *Pitx1*. We finally show how the structural variants can affect the locus 3D organization leading to ectopic gene expression and limb malformations.



## Introduction

The recent development of genome-wide technologies to map chromosomal contacts such as Hi-C, GAM, and SPRITE [1–3], has opened the investigation of the role of 3D genome organization in orchestrating the transcriptional activities of the cell. Typically, gene regulation relies on specific DNA segments, such as enhancers, that control the activation of genes by coming into physical proximity with their promoters, despite being located up to several hundreds of kilobases away [4]. This mechanism of regulation is reflected in complex patterns of contacts, as measured by the above techniques, and gives rise to the formation of peculiar structures such as *A/B compartments* at large-scale, Topologically Associating Domains (TADs) loops and stripes at the mega-base (Mb) scale and micro-TADs at smaller scales [1,5–8]. The functional role of those structures is extensively investigated, and it is known that the perturbation of TADs or loops structure can lead to gene misexpression and disease [9–12]. Notably, analyses of Hi-C datasets beyond the TADs organization level revealed that TADs interact with each other forming nonrandom higher order contacts among regions with similar genetic and epigenetic marks. Here we review the hierarchical structure of such a higher order organization, formed of domains-within-domains, named meta-TADs, which extends up to the range of entire chromosomes [13].

A common goal in this field is the development of a comprehensive biophysical model to understand the basic mechanisms shaping the chromosome 3D structure at the different length scales. To this aim, several approaches based on concepts from polymer physics have been proposed in order to provide a quantitative framework to explain the complexity of the experimental pattern of contacts [14–23]. Here, in particular, we review the main features of the Strings and Binders Switch (SBS) model [24], where chromatin is schematized as a self-avoiding polymer

chain and its spatial conformations arise through the attachment of dispersed molecular factors to binding sites placed along the polymer. By using the SBS model, experimental contact matrices can be reproduced with high accuracy [10,22]. As a case study where 3D structures beyond TADs are relevant, we review the organization of the genomic region around the *Pitx1* gene during morphogenesis and differentiation in mouse limbs [9]. At this locus, the gene is regulated by a *pan*-limb enhancer (Pen), which displays activity in both forelimb and hindlimb tissues, but *Pitx1* is expressed in hindlimb only. Interestingly, it has been found that the restriction of the enhancer activity to the hindlimb is associated with tissue-specific differences in their 3D chromatin structure. Finally, a structural variant located on a small region of the forelimb tissue can convert the forelimb-specific three-dimensional inactive structure to the hindlimb active one, causing gene misexpression and limb malformation. Our approach based on polymer physics allows us not only to dissect chromatin tissue-specific arrangements, but also to predict the effects of genomic mutations, i.e. deletions, inversions or duplications, on chromatin architecture.

## The Hierarchical Architecture of The Genome

Early analyses of Hi-C data revealed the presence in the genome of 10 Mb sized regions, belonging to two different classes named A and B *compartments* [1], the former linked to more open, actively transcribed chromatin, and the latter to more closed, repressed chromatin. Since they tend to interact with the homologous type of domain (A with A and B with B), they form a "plaid" pattern well visible in the contact maps, suggesting a physical separation among the A and B domains recently observed with super-resolution microscopy [25]. Subsequent analyses led to the identification of smaller domains named TADs [5,6], whose existence has later been supported by several independent methods [2,3,26]. The TADs were defined as 0.5−1 Mb long chromatin regions exhibiting a high level of Hi-C self-interaction and insulation from each other by boundary elements. Notably, boundaries are enriched for housekeeping genes, transfer RNAs, and the CTCF binding protein [5]. Representing a specific, largely tissue-invariant scale of genome architecture, TADs act as constraints of the genomic space sampled by each locus, enabling the control of genes by enhancers inside the TAD, but limiting their interaction with spurious enhancers located in different TADs [27,28]. However, a deeper look at Hi-C contact matrices suggests that contact patterns also exist inside and across TADs and extend up to the scale of entire chromosomes. By starting from this observation, the hypothesis of larger structures behind TADs has recently been proposed [13], and their functional role further investigated by subsequent studies [29−31].

To quantitatively define such larger structures, a clustering procedure of Hi-C contact domains has been implemented through the mouse neuronal differentiation [13]. Precisely, for each chromosome, the contacts between all the possible pairs of neighboring TADs were computed and the most strongly interacting pair thought to be the best candidate to form a higher-order structure, was merged into a larger domain, named meta-TAD (Fig. 1a). This procedure is repeated, iteratively clustering together pairs of TADs and meta-TADs until finding a meta-TAD of the same size of the whole chromosome. A tree of domains-within-domains (Fig. 1b) is then obtained, indicating the way in which the different TADs colocalize together to form higher order structures up to the chromosomal scale.

The meta-TADs structure has been validated statistically by using different measures [13]. For instance, it has been shown that Hi-C interactions between meta-TADs are significantly above a background reference, measured in randomized Hi-C matrices. In addition, a comparison between the meta-TADs tree and A/B compartment structure showed that TADs within a meta-TAD frequently belong to the same compartment (see e.g. Fig. 1a), as expected from the observation that regions belonging to the same type of compartment tend to interact with each other [1]. These results suggest that the hierarchical organization in domains-within-domains is a general feature of the genome, and the concept of meta-TAD correctly captures this type of architecture.

Notably, the meta-TADs hierarchy was shown to aggregate together chromatin domains with similar biological properties [13]. To this aim, different genomic, epigenomic, and expression features were measured, and they were found to correlate over longer genomic distances within meta-TADs trees than along the linear DNA sequence. These features are also found to be much more enriched across meta-TAD boundaries compared to TAD boundaries (Fig. 1c), suggesting an important functional role of the meta-TAD organization.

Similarly to the TAD structure, the meta-TAD hierarchical organization has been confirmed during different points of mouse neuronal differentiation, and its variability across cell type has been investigated. The results, based on the measure of the cophenetic correlation between meta-TAD trees, showed that meta-TADs are only relatively conserved (Fig. 1d), with a value of the correlation, which stood around 82%, still comparable to the variation of TAD boundaries, which are 70% conserved.

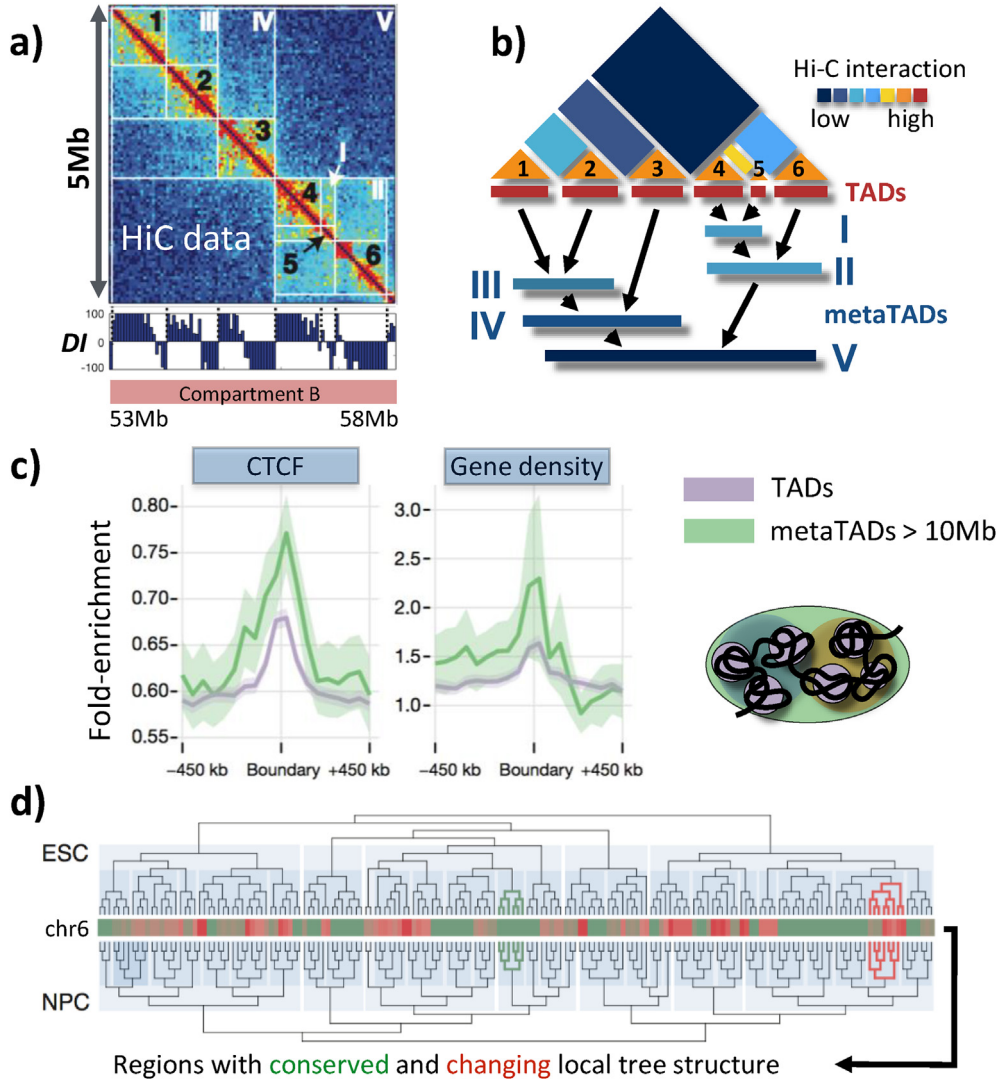


**Fig. 1. The hierarchical folding of the genome. (a)** Hi-C contact matrix of the chr2:53000000–58000000 region in mESC [13] with TADs (Arabic numbers) and meta-TADs (Roman numbers) highlighted. The middle panel shows the directionality index (DI) [5] used to identify the different TADs. The bottom panel shows that the whole illustrated region belongs to a B compartment. **(b)** meta-TADs have been identified by iteratively clustering the most strongly interacting pair of neighboring TADs, forming a tree of domains-within-domains (schematically shown in the panel). (c) The enrichment of the chromatin features at the TAD (violet) and meta-TAD (green) boundaries shows that the meta-TAD boundaries are even more correlated with genomic and epigenomic features than the TAD boundaries. (d) Comparison between the meta-TAD trees of chr6 in mESC (above) and neuronal precursors cells (below) to show that the meta-TAD structure is only partially conserved (green regions) during differentiation (the cophenetic correlation is around 80%). The central heatmap highlights the regions which are changing their meta-TAD structure (colored in red) and the regions conserving it (colored in green). As an example, the tree structure of two of these regions is also highlighted. Both change and conservation correlate with expression, but TADs in conserved meta-TADs have expression changes coherent in sign. (Adapted from Ref. [13]).

Finally, the relationship between meta-TADs rewiring and gene expression has been examined [13], finding that changes in gene expression correlate with tree changes as well as with tree conservation. The latter case was explained by observing that if a pair of TADs belongs to a conserved meta-TAD, the expression change tends to be in the same direction that is altered or maintained on both TADs.

In view of the presented results, the concept of meta-TAD captures the hierarchical properties of chromosomes folding. This tree-like structure of domains-within-domains correlates with genetics, epigenetics, and transcription data, hence

highlighting its functional role. Moreover, the meta-TAD organization is detected in different cell types and in both humans and mice [13], suggesting that this topology can be a general organizational principle of genomes.

## The SBS Model of Chromatin

The definition of TADs and meta-TADs, while representing interesting features of the way chromosomes fold, is directly based on the Hi-C contact matrices. In this heuristic approach, the experimental data represent the starting point to identify contact patterns attributable to the three-dimensional genomic structures (like TADs and meta-TADs). Other approaches, based on first principles, have been proposed in the last few years [14], with the potential advantage to suggest the molecular mechanisms beyond the architecture. In this section, we briefly review some key features of the Strings and Binders (SBS) model, extensively tested in previous works [22,24,32,33], and show its application to a real genomic region.

In the SBS framework, a chromatin filament is represented as a self-avoiding string of consecutive beads, along which binding sites for the attachment of specific molecules (binders) are distributed (Fig. 2a). Being subject to Brownian motion, these binders, biologically related to real DNA binding molecules, such as Transcription Factors, can bridge two or more polymer sites, forming loops. In this way, the interaction among beads and binders drives the folding of the chromosomes. The statistical properties of such a system are investigated by molecular dynamics simulations, i.e. by solving the

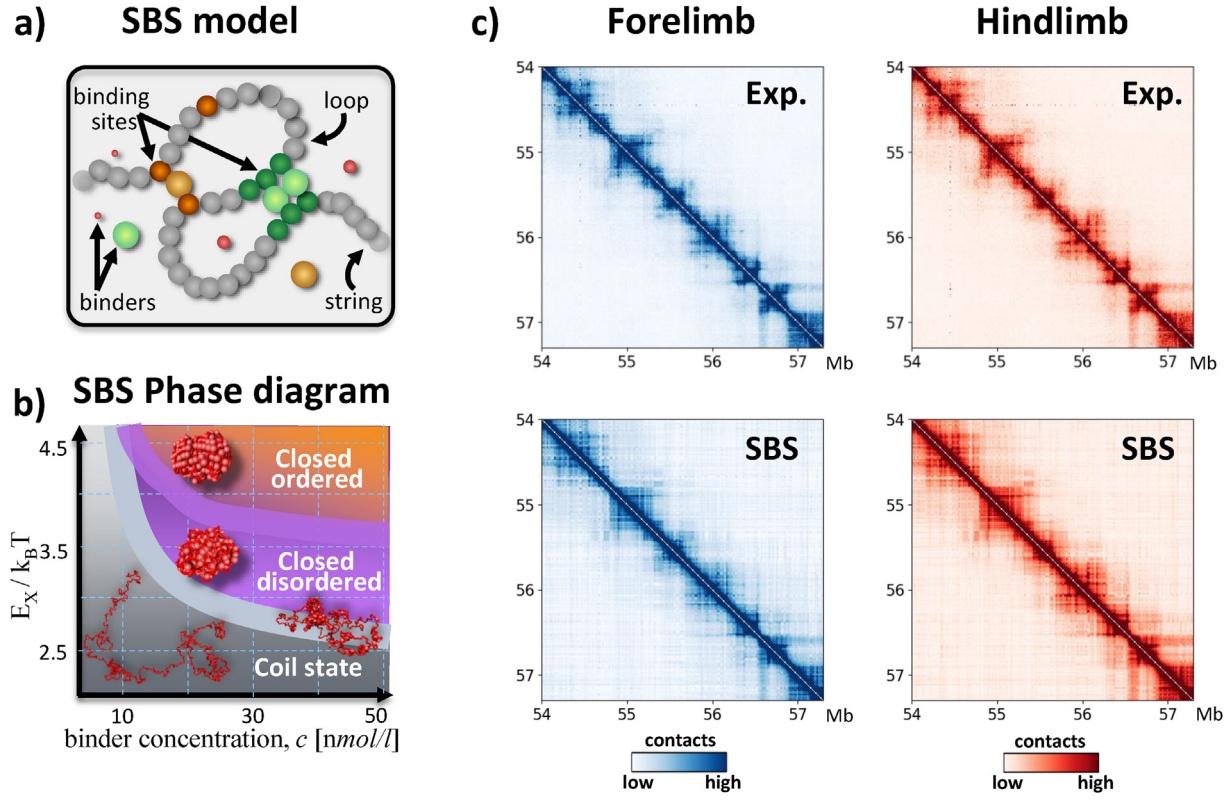


**Fig. 2. The SBS model explains the folding of the *Pitx1* gene region. (a)** Schematic representation of the SBS polymer model: a chromatin filament is represented by a self-avoiding chain of beads having binding sites (green and orange sites in the cartoon) for the attachment of dispersed molecules. The interactions between beads and binders bridge the different regions of the polymer chain shaping its architecture. (b) The phase diagram of an SBS polymer with a simplified distribution of the colored sites (only one type of interaction) shows three thermodynamic phases as a function of binder concentration $c$ and interaction energy $E_X$: a coil phase, where open polymer conformations arise; a closed-disordered phase, where the polymer folds in a more compact structure and disordered binders; a closed-ordered phase, where the polymer is compact but the binders spatial distribution is ordered (Adapted from Ref. [22]). (c) An example of the application of the SBS model on the chr13:54000000–57300000 mouse genomic region around the *Pitx1* gene. Top: experimental CHi-C maps [9] in the forelimb and hindlimb tissues. Bottom: the SBS model-derived contact maps show high similarity to the experimental ones, with a Pearson correlation coefficient equal to 0.98 in both cases. The bin size of the shown contact maps is 10 kb.

Langevin equations of motion with specific potentials, whose details are fully described in Ref. [34]. A simplified version of the model, with only one type of binding site, provides a phase diagram where different classes emerge as a function of the concentration of binders, $c$, and their interaction energy with the polymer chain, $E_X$. More precisely, three main thermodynamic phases are found (Fig. 2b): a *coiled* state, where the polymer falls in an open self-avoiding walk, at small $c$ or $E_X$, and a closed *globule* state, where the polymer chain folds in a more compact structure and binders are organized in either a disordered or ordered assembly, according to the value of concentrations and energies of interaction. The different thermodynamic phases correspond to the different conformational classes of Polymer Physics, and the transition from one phase to another can be obtained by crossing the phase boundary, without fine-tuning parameters.

Such a model, albeit simple, can easily explain the formation of chromatin structures as TADs, metaTADs, and compartments with an appropriate distribution of the interacting sites along the polymer chain. For instance, an SBS model with two different types of binding sites arranged in two contiguous blocks (e.g. red-green) will form two spatially separated domains (two TADs) due to the independent interaction of the respective block with its cognate binders [24], a concept known in polymer physics as microphase separation [36]. Interestingly, recent experiments supported the role of microphase separation in chromatin organization [35–38]. By considering a slightly more complex distribution of the binding sites, higher-order structures as metaTADs can also be recapitulated [22,39]. For example, by adding a third type of beads (e.g. blue), interspersed in the two contiguous blocks, the metaTADs structure spontaneously arises from the higher order interaction between the two homologous blocks (mediated by the blue binders). Compartments can be easily reproduced by the model as well, by considering more adjacent blocks of two alternating types of binding sites (e.g. red-green-red-green). In fact, the long-range interactions between blocks of the same type will lead to the formation of structures resembling compartments, with the typical plaid pattern arising from the model average contact map. Finally, by increasing the number of types of interactions and further complicating the binding sites distribution, an even more complex pattern of contacts can be described, up to real genomic regions.

## A Case Study: Folding of the *Pitx1* Locus in Mouse Limbs

To describe the folding of real genomic loci, the PRISMR method, based on Monte Carlo Simulated Annealing optimization procedure, has been developed [10]. Briefly, starting from an experimental contact matrix, the algorithm searches in the huge space of all the possible polymer models for the one that best describes the input matrix. The output consists of an SBS polymer with a minimal number of different types of interactions and the precise distribution of the binding sites along the polymer chain. All the details of the PRISMR minimization procedure can be found in Ref. [10].

As an example of the above-discussed method, here we review its application on a genomic region of the mouse chromosome 13 containing the *Pitx1* gene. Since *Pitx1* is found to be expressed in hindlimbs, but not expressed in forelimbs, its regulation is crucial to ensure a correct identity and differentiation of mouse limbs [40]. The *Pitx1* regulatory landscape has been shown to extend over 400 kb and to form several specific chromatin loops termed as regulatory anchors (*RA*) 1–5, with *RA2* representing the gene promoter and RA5/Pen its enhancer [9]. In order to investigate the spatial conformation of this locus, Capture Hi-C (CHi-C) interaction maps encompassing a 3 Mb region around Pitx1 have been produced in the mouse forelimb and hindlimb tissues. The corresponding 3D structures are then obtained, from the data at 10 kb resolution, deriving with PRISMR the best SBS polymer models of both loci. To take into account the effects of cell population heterogeneity, we considered the coil/globule mixture, which maximizes the matrices similarity, obtaining an 80–20% mixture for both forelimb and hindlimb. In Fig. 2c, we show the experimental contact matrices (top panel) and the matrices inferred from the model (bottom panel) of the entire region considered chr13:54000000–57300000 (mm9). The contact pattern is well recapitulated by the SBS model in both cases, as also shown by the high values of the Pearson correlation coefficient r between the experimental and model matrices ($r = 0.98$ in both forelimb and hindlimb).

## Ectopic Gene-Enhancer Interaction Causes Limb Malformation

The mechanisms of regulation at the *Pitx1* gene region has been further investigated by looking at 3D polymer models of a restricted 1 Mb wide region where the differences between the forelimb and hindlimb interaction maps were marked. Precisely, in the forelimb tissue interactions are found between *Pitx1* and the repressed gene Neurog1, but not with *Pitx1* and Pen (Fig. 3a left), while in the hindlimbs it is the opposite, with *Pitx1* showing contacts with its enhancer Pen, but not with Neurog1 (Fig. 3b left). The derived ensembles of polymer structures in the two tissues allowed to visualize the conformational

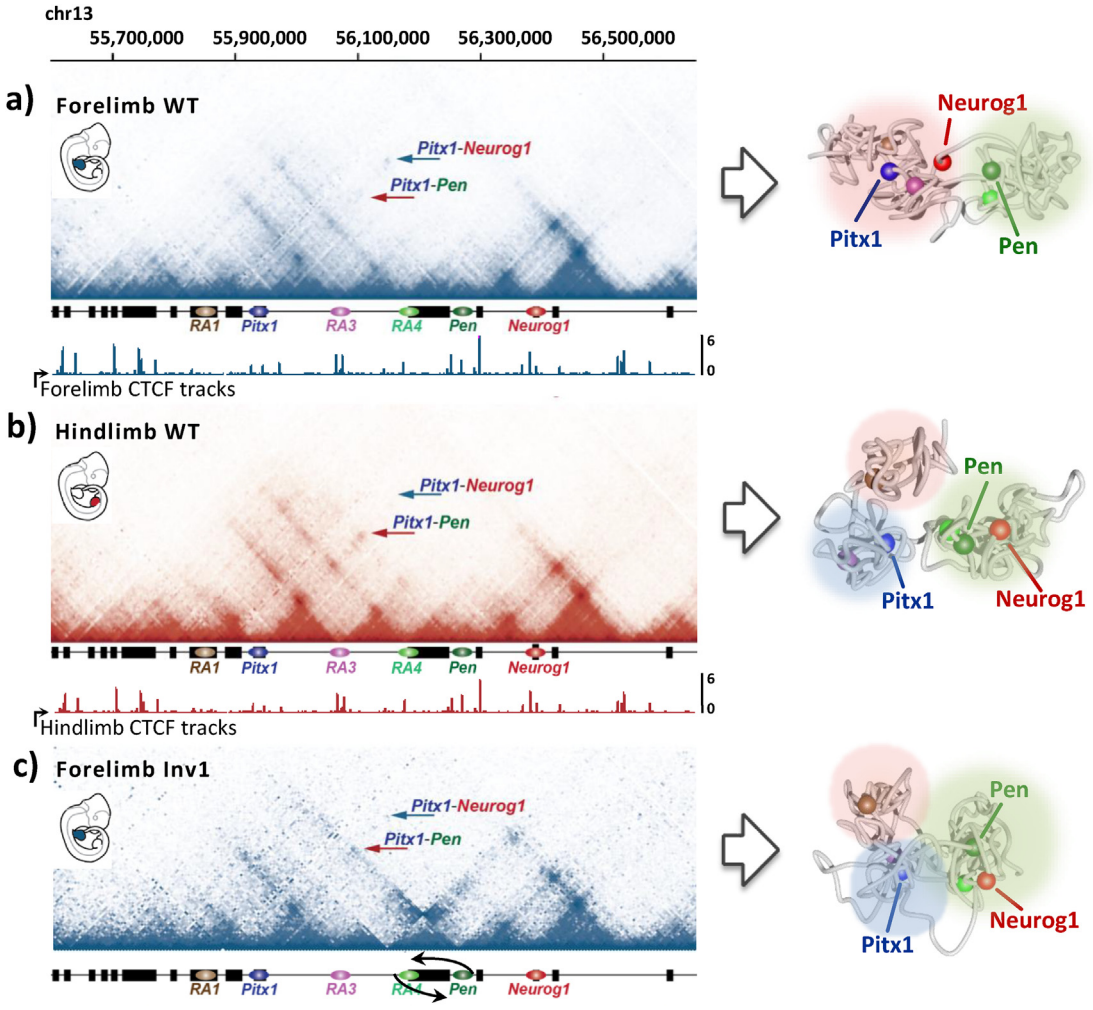


**Fig. 3. Tissue-specific 3D conformations modulate gene-enhancer interaction (a)** Left: CHi-C maps of the restricted forelimb region encompassing the *Pitx1* gene with some relevant interactions highlighted. The bar plot under the matrix is the distribution of the transcription factor CTCF along the genomic sequence in the forelimb. Right: representative 3D structure of the forelimb genomic region showing a two-hub conformation. Note that the repressed Neurog1 gene prevents the contact between *Pitx1* and Pen, which are separated from each other. (b) Left: CHi-C maps of the restricted hindlimb region with some relevant interactions highlighted. The bar plot under the matrix is the distribution of the transcription factor CTCF along the genomic sequence in the hindlimb. Right: representative 3D structure of the hindlimb genomic region showing a three-hub conformation. Note the proximity between *Pitx1* and Pen, which can correctly activate its target gene. (c) Left: CHi-C maps of the Inv1 forelimb case with some relevant interactions highlighted. The inverted region, indicated in the bar below the map, involves RA4 and Pen and moves Pen nearer to *Pitx1*. Right: representative 3D structure of the Inv1 forelimb genomic region showing that the locus fold in a hindlimb-like three-hub conformation, with increased proximity between Pen and *Pitx1*, which is ectopically activated with a consequent limb malformation (a partial arm to leg transformation). The bin size of the shown contact matrices is 10 kb (Adapted from Ref. [9]).

changes in the 3D space and to perform quantitative measures, such as physical distances among the regions of interest, helping to provide a clearer biological interpretation. As shown in Fig. 3a (right), in the forelimbs, the locus segregates into two chromatin hubs, containing (1) *Pitx1*, RA3, and Neurog1 (blue, pink, and red spheres, respectively) and (2) Pen and RA4 (dark green and light green spheres, respectively). This spatial conformation is such that Pen and *Pitx1* are separated from each other, and the repressed gene Neurog1 is close to Pitx1, so preventing its activation. Conversely, the hindlimb-specific configuration is partitioned in three major hubs (Fig. 3b, right), one containing RA1 only, another one containing *Pitx1* and RA3, and the last one RA4, Pen, and Neurog1. Here, the physical proximity between *Pitx1* and its enhancer Pen ensures a correct regulation of the gene. Therefore,

chromatin spatial configuration restricts the activity of Pen to the hindlimb tissue by separating the enhancer from its promoter in the forelimbs.

The relationship between architecture and function is even more highlighted by a structural variant acting on this locus. Indeed, a 113-kb long inversion (Inv1) of a fragment containing Pen and RA4 alters the inactive forelimb conformation, which becomes nearly identical to the wild type hindlimb. Indeed, the interaction map of Inv1 forelimbs shows clear hindlimb-specific features, i.e. the ectopic contact between *Pitx1* and Pen and the absence of interaction between *Pitx1* and Neurog1 (Fig. 3c, left). The 3D reconstruction of the locus bearing the inversion shows a hindlimb-like architecture with the formation of three chromatin hubs and a closer proximity between Pen and *Pitx1* compared to the wild type forelimb (Fig. 3c, right). Because of this increased proximity, the gene displayed a 44-fold increase in expression, and limb malformation was detected on Inv1 mice [9].

To conclude, the application of the SBS model to the *Pitx1* gene region allowed us to translate the pairwise information in the ensemble of 3D conformations of the region under study, obtaining significant insights into the spatial organization of the locus, as the hindlimb three hubs and forelimb two hubs organization, not directly accessible from the experimental contact data. Additionally, the physical distances among any region of the locus can be computed from the ensemble of polymers, confirming, in this case, the greater proximity between Pen and *Pitx1* in hindlimb and between Neurog1 and *Pitx1* in forelimb with respect to the other tissue [9].

## Different Models of Chromatin Architecture

As broadly discussed above, the SBS model can describe many features of chromosome architecture across genomic scales. However, it cannot explain, for example, the so-called *CTCF* convergence-bias (see below) without extra, ad hoc hypothesis. In the following, we give a brief overview of the alternative polymer physics models and discuss how their integration could provide a more comprehensive description of the chromatin organization. Some of these models are essentially based on the homotypic interaction among particles, as in the SBS scenario. An example is a block-copolymer model [17], where the self-avoiding chain representing a chromatin filament includes blocks of beads of the same type (or color) and the interactions, only allowed among equal colored beads, occur directly between chromatin loci, without the requirement of mediators. With an appropriate coloring, such a system can be used to describe the folding of real chromatin regions. For instance, using the epigenetic data to derive the location of the different blocks, it has been shown in Ref. [17] that the contact matrices of the *Drosophila melanogaster* genome are in good agreement with the matrices predicted from the block-copolymer model.

A different way to look into the molecular mechanisms of chromatin folding is provided by the loop extrusion (*LE*) model [18,21]. Here a loop extruding factor, such as the mammalian ring-structured *cohesin* complex, anchors in a point of the chain and starts to extrude a chromatin loop via an ATP driven process. The extrusion persists until some roadblocks are reached, enabling the regions between the barriers to contact themselves and giving rise to the formation of chromatin contact domains. In mammals, the zinc finger protein CTCF has been identified as the roadblock factor which halts the extrusion process in an orientation-dependent manner, that is, the CTCF sites at the loop borders need to be inward-oriented (convergent) to allow the blocking action. Since in the LE model a significant energy consumption is required to let the ATP motor work, an alternative scenario has been proposed by the slip-link (*SL*) model [41,42]. In the SL model, instead of active extrusion, the cohesin complex simply slides over the chromatin fiber in a diffusive fashion, with no ATP usage. Remarkably, LE-based models were shown to successfully explain the formation of chromatin structures such as TADs, loops, and stripes [18,21]. Recent observations that chromatin organization at the TAD level is largely erased upon cohesin depletion [43,44], also point toward a role of LE processes in shaping those structures. Nevertheless, recent super-resolution microscopy studies [26] showed that TADs disruption as a consequence of cohesin removal only holds in population-averaged contact matrices while not preventing domain formation at the single-cell level, albeit with a notably more random distribution along the genome. This finding suggests that the cohesin/CTCF extrusion mechanism plays a fundamental role in determining the preferential TAD boundaries locations rather than being the basic mechanism of TAD formation. Thus, it is likely that other mechanisms, for example, homotypic interactions, act in parallel to the LE to drive chromatin organization at the TAD level. Moreover, differently from TADs, experiments have shown that the larger scale organization of chromatin in A/B compartments is largely preserved upon cohesin depletion, thus excluding LE as a mechanism for their formation while supporting the homotypic interaction scenario.

While the SBS and the block-copolymer models envisage a scenario in which the 3D chromatin structures are conformations at thermodynamic equilibrium, in the LE, as well as in the SL model, the final conformations are far from equilibrium.

Since both frameworks can explain critical properties of chromatin folding which the other approach cannot explain (e.g. A/B compartments for the equilibrium models or the CTCF convergence bias for the LE ones) combined models have been proposed, in which multiple homotypic interactions and chromatin extrusion coexist [10,43,45]. However, it is still not clear if loop extrusion and homotypic interaction are competitive or cooperative mechanisms, and the study of their precise interplay will be crucial to understand the principles of genome architecture and regulation.

## Discussion

We reviewed and discussed the recent advancements based on the analysis of Hi-C data showing that chromosomes fold in a hierarchy of domains-within-domains named meta-TADs, extending from the scale of the TAD up to the size of entire chromosomes. By correlating with genomic and epigenomic features, meta-TAD tree structure slightly reorganizes during differentiation, and these changes correlate with the changes in expression. The advantages of such a hierarchical organization may lie in facilitating the general compaction of chromatin while preserving, at the same time, the specificity and the efficiency of mechanisms needed to access specific genomic region to activate or silence target genes.

Polymer physics-based models are becoming an increasingly important tool to study the molecular mechanisms regulating the genome spatial organization [17,18,24,41]. Here, we reviewed some of the key aspects of the SBS model, a polymer model where folding is driven by a specific arrangement of binding domains and molecular factors by phase separation [10,22,32,46]. Other models provide different scenarios. For instance, in the loop-extrusion model [18], the cohesin complex progressively extrudes the chromatin filament until reaching convergently oriented CTCF boundaries that block its extruding activity. Although these models have been shown to describe multiple observations [21,43,47], a unified model explaining all the features of the 3D genome organization is still lacking.

Finally, we reviewed how polymer physics within the SBS scenario captures well the essentials of chromatin folding at the *Pitx1* genomic region, crucially involved in developing limb buds. Here, the spatial reconstruction of the forelimb and hindlimb locus architecture allows the interpretation of the different state of activity of *Pitx1*, which are associated with its different spatial positioning with respect to the *pan*-limb enhancer Pen and the repressed Neurog1 gene. Precisely, while in the hindlimbs, *Pitx1* is close to Pen, an opposite condition turns out in the forelimbs, where it is physically disconnected from the enhancer [9]. A structural variant repositioning Pen genomically closer to *Pitx1* in the forelimb can convert the locus architecture in the hindlimb-like structure, leading to pathogenic *Pitx1* activation. This suggests that tissue-specific spatial conformation can modulate enhancer-promoter contacts, and hence, having an active regulatory role.

Novel technologies are emerging that can investigate chromatin pairwise and multiway contacts at the single-cell level, such as multicontact 4C [48], TriC [49] and super-resolution microscopy [26] at an increasingly higher resolution. These experimental advancements, in combination with polymer physics and computational approaches, will hopefully lead to the development of comprehensive models of chromatin folding and shed light on the role of architecture on cellular functions and diseases.

## Founding Information

We acknowledge grants from the National Institutes of Health (NIH) (1U54DK107977-01), the EU H2020 Marie Curie ITN (813282), the Einstein BIH Fellowship Award (EVF-BIH-2016-282).

## Acknowledgments

We acknowledge computer resources from INFN, CINECA, Scope/Recas and CRESCO/ENEAGRID [50]. We thank Prof. Mario Nicodemi for the critical discussion and useful comments about the manuscript.

*Received 16 July 2019;*
*Received in revised form 25 October 2019;*
*Accepted 11 December 2019*
Available online 18 December 2019

***Keywords:***
Hierarchical folding;
Genome architecture;
Principled approach;
Structural variants;
*Pitx1*

## References

[1] E. Lieberman-Aiden, N.L. Van Berkum, L. Williams, M. Imakaev, T. Ragoczy, A. Telling, I. Amit, B.R. Lajoie, P.J. Sabo, M.O. Dorschner, R. Sandstrom, B. Bernstein, M.A. Bender, M. Groudine, A. Gnirke, J. Stamatoyannopoulos, L.A. Mirny, E.S. Lander, J. Dekker, Comprehensive mapping of long-range interactions reveals folding principles of the human

genome, Science 326 (2009) 289–293, https://doi.org/10.1126/science.1181369, 80-.

[2] R.A. Beagrie, A. Scialdone, M. Schueler, D.C.A. Kraemer, M. Chotalia, S.Q. Xie, M. Barbieri, I. De Santiago, L.M. Lavitas, M.R. Branco, J. Fraser, J. Dostie, L. Game, N. Dillon, P.A.W. Edwards, M. Nicodemi, A. Pombo, Complex multi-enhancer contacts captured by genome architecture mapping, Nature 543 (2017) 519–524, https://doi.org/10.1038/nature21411.

[3] S.A. Quinodoz, N. Ollikainen, B. Tabak, A. Palla, J.M. Schmidt, E. Detmar, M.M. Lai, A.A. Shishkin, P. Bhat, Y. Takei, V. Trinh, E. Aznauryan, P. Russell, C. Cheng, M. Jovanovic, A. Chow, L. Cai, P. McDonel, M. Garber, M. Guttman, Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus, Cell 174 (2018) 744–757, https://doi.org/10.1016/j.cell.2018.05.024, e24.

[4] J. Dekker, L. Mirny, The 3D genome as moderator of chromosomal communication, Cell 164 (2016) 1110–1121, https://doi.org/10.1016/j.cell.2016.02.007.

[5] J.R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J.S. Liu, B. Ren, Topological domains in mammalian genomes identified by analysis of chromatin interactions, Nature 485 (2012) 376–380, https://doi.org/10.1038/nature11082.

[6] E.P. Nora, B.R. Lajoie, E.G. Schulz, L. Giorgetti, I. Okamoto, N. Servant, T. Piolot, N.L. Van Berkum, J. Meisig, J. Sedat, J. Gribnau, E. Barillot, N. Blüthgen, J. Dekker, E. Heard, Spatial partitioning of the regulatory landscape of the X-inactivation centre, Nature 485 (2012) 381–385, https://doi.org/10.1038/nature11049.

[7] S.S.P.P. Rao, M.H. Huntley, N.C. Durand, E.K. Stamenova, I.D. Bochkov, J.T. Robinson, A.L. Sanborn, I. Machol, A.D. Omer, E.S. Lander, E.L. Aiden, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping, Cell 159 (2014) 1665–1680, https://doi.org/10.1016/j.cell.2014.11.021.

[8] T.-H.S. Hsieh, E. Slobodyanyuk, A.S. Hansen, C. Cattoglio, O.J. Rando, R. Tjian, X. Darzacq, Resolving the 3D landscape of transcription-linked mammalian chromatin folding, BioRxiv (2019), https://doi.org/10.1101/638775.

[9] B.K. Kragesteen, M. Spielmann, C. Paliou, V. Heinrich, R. Schöpflin, A. Esposito, C. Annunziatella, S. Bianco, A.M. Chiariello, I. Jerković, I. Harabula, P. Guckelberger, M. Pechstein, L. Wittler, W.L. Chan, M. Franke, D.G. Lupiáñez, K. Kraft, B. Timmermann, M. Vingron, A. Visel, M. Nicodemi, S. Mundlos, G. Andrey, Dynamic 3D chromatin architecture contributes to enhancer specificity and limb morphogenesis, Nat. Genet. 50 (2018) 1463–1473, https://doi.org/10.1038/s41588-018-0221-x.

[10] S. Bianco, D.G. Lupiáñez, A.M. Chiariello, C. Annunziatella, K. Kraft, R. Schöpflin, L. Wittler, G. Andrey, M. Vingron, A. Pombo, S. Mundlos, M. Nicodemi, Polymer physics predicts the effects of structural variants on chromatin architecture, Nat. Genet. 50 (2018) 662–667, https://doi.org/10.1038/s41588-018-0098-8.

[11] D.G. Lupiáñez, K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J.M. Opitz, R. Laxova, F. Santos-Simarro, B. Gilbert-Dussardier, L. Wittler, M. Borschiwer, S.A. Haas, M. Osterwalder, M. Franke, B. Timmermann, J. Hecht, M. Spielmann, A. Visel, S. Mundlos, Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions, Cell 161 (2015) 1012–1025, https://doi.org/10.1016/j.cell.2015.04.004.

[12] D. Hnisz, A.S. Weintrau, D.S. Day, A.L. Valton, R.O. Bak, C.H. Li, J. Goldmann, B.R. Lajoie, Z.P. Fan, A.A. Sigova, J. Reddy, D. Borges-Rivera, T.I. Lee, R. Jaenisch, M.H. Porteus, J. Dekker, R.A. Young, Activation of proto-oncogenes by disruption of chromosome neighborhoods, Science 351 (2016) 1454–1458, https://doi.org/10.1126/science.aad9024, 80-.

[13] J. Fraser, C. Ferrai, A.M. Chiariello, M. Schueler, T. Rito, G. Laudanno, M. Barbieri, B.L. Moore, D.C. Kraemer, S. Aitken, S.Q. Xie, K.J. Morris, M. Itoh, H. Kawaji, I. Jaeger, Y. Hayashizaki, P. Carninci, A.R. Forrest, C.A. Semple, J. Dostie, A. Pombo, M. Nicodemi, Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation, Mol. Syst. Biol. 11 (2015), https://doi.org/10.15252/msb.20156492, 852–852.

[14] M. Nicodemi, A. Pombo, Models of chromosome structure, Curr. Opin. Cell Biol. 28 (2014) 90–95, https://doi.org/10.1016/j.ceb.2014.04.004.

[15] M. Bohn, D.W. Heermann, Diffusion-driven looping provides a consistent framework for chromatin organization, PLoS One 5 (2010), https://doi.org/10.1371/journal.pone.0012218.

[16] C.A. Brackley, S. Taylor, A. Papantonis, P.R. Cook, D. Marenduzzo, Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organization, Proc. Natl. Acad. Sci. 110 (2013) E3605–E3611, https://doi.org/10.1073/pnas.1302950110.

[17] D. Jost, P. Carrivain, G. Cavalli, C. Vaillant, Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains, Nucleic Acids Res. 42 (2014) 9553–9561, https://doi.org/10.1093/nar/gku698.

[18] A.L. Sanborn, S.S.P. Rao, S.-C. Huang, N.C. Durand, M.H. Huntley, A.I. Jewett, I.D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, K.P. Geeting, A. Gnirke, A. Melnikov, D. McKenna, E.K. Stamenova, E.S. Lander, E.L. Aiden, Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes, Proc. Natl. Acad. Sci. 112 (2015) E6456–E6465, https://doi.org/10.1073/pnas.1518552112.

[19] C.A. Brackley, J. Johnson, S. Kelly, P.R. Cook, D. Marenduzzo, Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains, Nucleic Acids Res. 44 (2016) 3503–3512, https://doi.org/10.1093/nar/gkw135.

[20] C.A. Brackley, J.M. Brown, D. Waithe, C. Babbs, J. Davies, J.R. Hughes, V.J. Buckle, D. Marenduzzo, Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models, Genome Biol. 17 (2016) 59, https://doi.org/10.1186/s13059-016-0909-0.

[21] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, L.A. Mirny, formation of chromosomal domains by loop extrusion, Cell Rep. 15 (2016) 2038–2049, https://doi.org/10.1016/j.celrep.2016.04.085.

[22] A.M. Chiariello, C. Annunziatella, S. Bianco, A. Esposito, M. Nicodemi, Polymer physics of chromosome large-scale 3D organisation, Sci. Rep. 6 (2016), https://doi.org/10.1038/srep29775.

[23] M. Di Stefano, J. Paulsen, T.G. Lien, E. Hovig, C. Micheletti, Hi-C-constrained physical models of human chromosomes

recover functionally-related properties of genome organization, Sci. Rep. 6 (2016), https://doi.org/10.1038/srep35985.
[24] M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, M. Nicodemi, Complexity of chromatin folding is captured by the strings and binders switch model, Proc. Natl. Acad. Sci. 109 (2012) 16173–16178, https://doi.org/10.1073/pnas.1204799109.
[25] G. Nir, I. Farabella, C. Pérez Estrada, C.G. Ebeling, B.J. Beliveau, H.M. Sasaki, S.H. Lee, S.C. Nguyen, R.B. McCole, S. Chattoraj, J. Erceg, J. AlHaj Abed, N.M.C. Martins, H.Q. Nguyen, M.A. Hannan, S. Russell, N.C. Durand, S.S.P. Rao, J.Y. Kishi, P. Soler-Vila, M. Di Pierro, J.N. Onuchic, S.P. Callahan, J.M. Schreiner, J.A. Stuckey, P. Yin, E.L. Aiden, M.A. Marti-Renom, C.T. Wu, Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling, PLoS Genet. (2018), https://doi.org/10.1371/journal.pgen.1007872.
[26] B. Bintu, L.J. Mateo, J.-H. Su, N.A. Sinnott-Armstrong, M. Parker, S. Kinrot, K. Yamaya, A.N. Boettiger, X. Zhuang, Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells, Science 80– (2018) 362, https://doi.org/10.1126/science.aau1783, eaau1783.
[27] O. Symmons, V.V. Uslu, T. Tsujimura, S. Ruf, S. Nassari, W. Schwarzer, L. Ettwiller, F. Spitz, Functional and topological characteristics of mammalian regulatory domains, Genome Res. (2014), https://doi.org/10.1101/gr.163519.113.
[28] W. Akhtar, J. De Jong, A.V. Pindyurin, L. Pagie, W. Meuleman, J. De Ridder, A. Berns, L.F.A. Wessels, M. Van Lohuizen, B. Van Steensel, Chromatin position effects assayed by thousands of reporters integrated in parallel, Cell (2013), https://doi.org/10.1016/j.cell.2013.07.018.
[29] Y. Zhan, L. Mariani, I. Barozzi, E.G. Schulz, N. Blüthgen, M. Stadler, G. Tiana, L. Giorgetti, Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes, Genome Res. (2017), https://doi.org/10.1101/gr.212803.116.
[30] A.M. Chiariello, S. Bianco, C. Annunziatella, A. Esposito, M. Nicodemi, The scaling features of the 3D organization of chromosomes are highlighted by a transformation a la Kadanoff of Hi-C data, Epl 120 (2017), https://doi.org/10.1209/0295-5075/120/40004.
[31] O. Shukron, V. Piras, D. Noordermeer, D. Holcman, Statistics of chromatin organization during cell differentiation revealed by heterogeneous cross-linked polymers, Nat. Commun. 10 (2019), https://doi.org/10.1038/s41467-019-10402-x.
[32] C. Annunziatella, A.M. Chiariello, S. Bianco, M. Nicodemi, Polymer models of the hierarchical folding of the Hox-B chromosomal locus, Phys. Rev. E. 94 (2016), https://doi.org/10.1103/PhysRevE.94.042402.
[33] S. Bianco, A.M.A.M. Chiariello, C. Annunziatella, A. Esposito, M. Nicodemi, Predicting chromatin architecture from models of polymer physics, Chromosome Res. 25 (2017) 25–34, https://doi.org/10.1007/s10577-016-9545-5.
[34] C. Annunziatella, A.M. Chiariello, A. Esposito, S. Bianco, L. Fiorillo, M. Nicodemi, Molecular dynamics simulations of the strings and binders switch model of chromatin, Methods 142 (2018) 81–88, https://doi.org/10.1016/j.ymeth.2018.02.024.
[35] D. Hnisz, K. Shrinivas, R.A. Young, A.K. Chakraborty, P.A. Sharp, A phase separation model for transcriptional control, Cell 169 (2017) 13–23, https://doi.org/10.1016/j.cell.2017.02.007.
[36] A.G. Larson, D. Elnatan, M.M. Keenen, M.J. Trnka, J.B. Johnston, A.L. Burlingame, D.A. Agard, S. Redding, G.J. Narlikar, Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin, Nature 547 (2017) 236–240, https://doi.org/10.1038/nature22822.
[37] A.R. Strom, A.V. Emelyanov, M. Mir, D.V. Fyodorov, X. Darzacq, G.H. Karpen, Phase separation drives heterochromatin domain formation, Nature 547 (2017) 241–245, https://doi.org/10.1038/nature22989.
[38] B.A. Gibson, L.K. Doolittle, M.W.G. Schneider, L.E. Jensen, N. Gamarra, L. Henry, D.W. Gerlich, S. Redding, M.K. Rosen, Organization of chromatin by intrinsic and regulated phase separation, Cell 179 (2019) 470–484, https://doi.org/10.1016/j.cell.2019.08.037, e21.
[39] M. Barbieri, S.Q. Xie, E. Torlai Triglia, A.M. Chiariello, S. Bianco, I. De Santiago, M.R. Branco, D. Rueda, M. Nicodemi, A. Pombo, Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells, Nat. Struct. Mol. Biol. 24 (2017) 515–524, https://doi.org/10.1038/nsmb.3402.
[40] A. DeLaurier, R. Schweitzer, M. Logan, Pitx1 determines the morphology of muscle, tendon, and bones of the hindlimb, Dev. Biol. 299 (2006) 22–34, https://doi.org/10.1016/j.ydbio.2006.06.055.
[41] C.A. Brackley, J. Johnson, D. Michieletto, A.N. Morozov, M. Nicodemi, P.R. Cook, D. Marenduzzo, Nonequilibrium chromosome looping via molecular slip links, Phys. Rev. Lett. 119 (2017), https://doi.org/10.1103/PhysRevLett.119.138101.
[42] C.A. Brackley, J. Johnson, D. Michieletto, A.N. Morozov, M. Nicodemi, P.R. Cook, D. Marenduzzo, Extrusion without a motor: a new take on the loop extrusion model of genome organization, Nucleus (2018), https://doi.org/10.1080/19491034.2017.1421825.
[43] W. Schwarzer, N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, Y. Loe-Mie, N.A. Fonseca, W. Huber, C. H Haering, L. Mirny, F. Spitz, Two independent modes of chromatin organization revealed by cohesin removal, Nature 551 (2017) 51–56, https://doi.org/10.1038/nature24281.
[44] S.S.P. Rao, S.C. Huang, B. Glenn St Hilaire, J.M. Engreitz, E.M. Perez, K.R. Kieffer-Kwon, A.L. Sanborn, S.E. Johnstone, G.D. Bascom, I.D. Bochkov, X. Huang, M.S. Shamim, J. Shin, D. Turner, Z. Ye, A.D. Omer, J.T. Robinson, T. Schlick, B.E. Bernstein, R. Casellas, E.S. Lander, E.L. Aiden, Cohesin loss eliminates all loop domains, Cell 171 (2017) 305–320, https://doi.org/10.1016/j.cell.2017.09.026, e24.
[45] J. Nuebler, G. Fudenberg, M. Imakaev, N. Abdennur, L.A. Mirny, Chromatin organization by an interplay of loop extrusion and compartmental segregation, Proc. Natl. Acad. Sci. U. S. A (2018), https://doi.org/10.1073/pnas.1717730115.
[46] A.M. Chiariello, A. Esposito, C. Annunziatella, S. Bianco, L. Fiorillo, A. Prisco, M. Nicodemi, A polymer physics investigation of the architecture of the murine orthologue of the 7q11.23 human locus, Front. Neurosci. (2017), https://doi.org/10.3389/fnins.2017.00559.
[47] K. Kraft, A. Magg, V. Heinrich, C. Riemenschneider, R. Schöpflin, J. Markowski, D.M. Ibrahim, R. Acuna-Hidalgo, A. Despang, G. Andrey, L. Wittler, B. Timmermann, M. Vingron, S. Mundlos, Serial genomic inversions induce tissue-specific architectural stripes, gene misexpression and

congenital malformations, Nat. Cell Biol. 21 (2019) 305–310, https://doi.org/10.1038/s41556-019-0273-x.

[48] A. Allahyar, C. Vermeulen, B.A.M. Bouwman, P.H.L. Krijger, M.J.A.M. Verstegen, G. Geeven, M. van Kranenburg, M. Pieterse, R. Straver, J.H.I. Haarhuis, K. Jalink, H. Teunissen, I.J. Renkens, W.P. Kloosterman, B.D. Rowland, E. de Wit, J. de Ridder, W. de Laat, Enhancer hubs and loop collisions identified from single-allele topologies, Nat. Genet. (2018), https://doi.org/10.1038/s41588-018-0161-5.

[49] A.M. Oudelaar, J.O.J. Davies, L.L.P. Hanssen, J.M. Telenius, R. Schwessinger, Y. Liu, J.M. Brown, D.J. Downes, A.M. Chiariello, S. Bianco, M. Nicodemi, V.J. Buckle, J. Dekker, D.R. Higgs, J.R. Hughes, Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains, Nat. Genet. (2018), https://doi.org/10.1038/s41588-018-0253-2.

[50] G. Ponti, F. Palombi, D. Abate, F. Ambrosino, G. Aprea, T. Bastianelli, F. Beone, R. Bertini, G. Bracco, M. Caporicci, B. Calosso, M. Chinnici, A. Colavincenzo, A. Cucurullo, P. Dangelo, M. De Rosa, P. De Michele, A. Funel, G. Furini, D. Giammattei, S. Giusepponi, R. Guadagni, G. Guarnieri, A. Italiano, S. Magagnino, A. Mariano, G. Mencuccini, C. Mercuri, S. Migliori, P. Ornelli, S. Pecoraro, A. Perozziello, S. Pierattini, S. Podda, F. Poggi, A. Quintiliani, A. Rocchi, C. Scio, F. Simoni, A. Vita, The role of medium size facilities in the HPC ecosystem: the case of the new CRESCO4 cluster integrated in the ENEAGRID infrastructure, in: Proc. 2014 Int. Conf. High Perform. Comput. Simulation, HPCS, 2014, p. 2014, https://doi.org/10.1109/HPCSim.2014.6903807.