



# A Polymer Physics Investigation of the Architecture of the Murine Orthologue of the 7q11.23 Human Locus

Andrea M. Chiariello<sup>1,2†</sup>, Andrea Esposito<sup>1,2,3†</sup>, Carlo Annunziatella<sup>1,2†</sup>, Simona Bianco<sup>1,2†</sup>, Luca Fiorillo<sup>1,2</sup>, Antonella Prisco<sup>4</sup> and Mario Nicodemi<sup>1,2,5\*</sup>

<sup>1</sup> Dipartimento di Fisica, Università di Napoli Federico II, Naples, Italy, <sup>2</sup> Istituto Nazionale Di Fisica Nucleare Napoli (INFN), Complesso Universitario di Monte Sant'Angelo, Naples, Italy, <sup>3</sup> Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC), Berlin, Germany, <sup>4</sup> Institute of Genetics and Biophysics, Consiglio Nazionale Delle Ricerche (CNR), Naples, Italy, <sup>5</sup> Berlin Institute of Health, Berlin, Germany

## OPEN ACCESS

### Edited by:

Umberto di Porzio,  
Consiglio Nazionale Delle Ricerche  
(CNR), Italy

### Reviewed by:

Argyris Papantonis,  
University of Cologne, Germany  
Alba Di Pardo,  
Centre for Neurogenetics and Rare  
Diseases, Italy

### \*Correspondence:

Mario Nicodemi  
[mario.nicodemi@na.infn.it](mailto:mario.nicodemi@na.infn.it)

<sup>†</sup>These authors have contributed  
equally to this work.

### Specialty section:

This article was submitted to  
Neurodegeneration,  
a section of the journal  
Frontiers in Neuroscience

Received: 27 July 2017

Accepted: 25 September 2017

Published: 10 October 2017

### Citation:

Chiariello AM, Esposito A,  
Annunziatella C, Bianco S, Fiorillo L,  
Prisco A and Nicodemi M (2017) A  
Polymer Physics Investigation of the  
Architecture of the Murine Orthologue  
of the 7q11.23 Human Locus.  
*Front. Neurosci.* 11:559.  
doi: 10.3389/fnins.2017.00559

In the last decade, the developments of novel technologies, such as Hi-C or GAM methods, allowed to discover that chromosomes in the nucleus of mammalian cells have a complex spatial organization, encompassing the functional contacts between genes and regulators. In this work, we review recent progresses in chromosome modeling based on polymer physics to understand chromatin structure and folding mechanisms. As an example, we derive in mouse embryonic stem cells the full 3D structure of the *Bmp7* locus, a genomic region that plays a key role in osteoblastic differentiation. Next, as an application to Neuroscience, we present the first 3D model for the mouse orthologue of the Williams–Beuren syndrome 7q11.23 human locus. Deletions and duplications of the 7q11.23 region generate neurodevelopmental disorders with multi-system involvement and variable expressivity, and with autism. Understanding the impact of such mutations on the rewiring of the interactions of genes and regulators could be a new key to make sense of their related diseases, with potential applications in biomedicine.

**Keywords:** polymer physics, chromatin, Neuroscience, congenital disease, 7q11.23 locus

## INTRODUCTION

Novel and different technologies, such as Hi-C (Lieberman-Aiden et al., 2009) and GAM (Beagrie et al., 2017), are revealing that, in higher organisms, chromatin is folded in the nucleus of cells with complex 3D spatial organization (Lanctôt et al., 2007; Misteli, 2007; Bickmore and van Steensel, 2013; Tanay and Cavalli, 2013; Dekker and Mirny, 2016). Chromosomes are organized into arrays of megabase-sized topologically associating domains (TADs), characterized by strong local interactions (Dixon et al., 2012; Nora et al., 2012; Beagrie et al., 2017); TADs, in turn, can interact between each other generating higher order structure, called meta-TADs, extending across genomic scales (Fraser et al., 2015); and patterns are seen also within TADs (Sexton et al., 2012; Phillips-Cremins et al., 2013). The 3D structure of chromatin has crucial functional roles since, for instance, gene activity can be controlled by chromosomal interactions through the formation of long-range contacts between gene and their regulators. Nevertheless, the mechanisms regulating chromosomes architecture are only partially understood.

In order to elucidate the genome-wide contact data and to identify the mechanism underlying chromatin 3D organization, quantitative models, based on principles of polymer physics, have been

developed in the last few years (reviewed, e.g., in Nicodemi and Pombo, 2014). Some of them focus on the molecular mechanisms driving chromatin folding, such as the interactions with binding molecules (Nicodemi and Prisco, 2009; Barbieri et al., 2012; Brackley et al., 2013; Jost et al., 2014; Chiariello et al., 2016); some others consider dynamic processes, involving the transient formation of loops (Bohn and Heermann, 2010) or mechanisms where the polymer is extruded dynamically by specific extruding molecular factors (Sanborn et al., 2015; Fudenberg et al., 2016). In others entire chromosomes are modeled, based topological and kinetic constraints (Rosa and Everaers, 2008; Di Stefano et al., 2016). Here we focus on the String&Binders Switch (SBS) model, which appears to be able to explain in a quantitative way Hi-C, GAM and FISH data within a single framework (Nicodemi and Prisco, 2009; Barbieri et al., 2012; Fraser et al., 2015; Annunziatella et al., 2016; Chiariello et al., 2016; Beagrie et al., 2017). In the SBS model specific interactions between DNA-binding molecules determine the formation of non-random chromatin loops (Barbieri et al., 2012; Annunziatella et al., 2016; Chiariello et al., 2016). Within the SBS model, contact probabilities from many independent Hi-C datasets (Annunziatella et al., 2016; Chiariello et al., 2016) can be explained with a high degree of accuracy. Here we review the basic underlying concepts of polymer physics and, in particular, we report our results about the 3D structure of *Bmp7* locus (Chiariello et al., 2016) in mouse embryonic stem cells (ESC-46C). As an application of potential interest to neurogenetics, we reconstruct for the first time the three dimensional structure of the *7q11.23* locus, where structural variants are associated with severe disorders such as autism and Williams-Beuren syndrome (e.g., Sanders et al., 2011 and ref.s therein).

## POLYMER MODELS OF CHROMATIN ORGANIZATION

In the SBS model a chromatin filament is represented as a Self-Avoiding Walk (SAW) polymer made of consecutive beads which can interact with diffusing particles (binders, **Figure 1A**). The interaction between binders and bead chain spontaneously gives rise to the typical loops that are present in the genome. In the simplest version, only one type of binders are present, and can interact with all the beads of the chain, homopolymer model (Nicodemi and Prisco, 2009; Barbieri et al., 2012, 2013; Annunziatella et al., 2016; Chiariello et al., 2016). The binding-molecules have a concentration  $c$  and a binding affinity  $E_X$ . By tuning the control parameters  $c$  and  $E_X$  the system folds in its different stable conformational states, as dictated by polymer physics. It is possible to show that at least three stable states exist (Annunziatella et al., 2016; Chiariello et al., 2016; Bianco et al., 2017) the coil state, observed for low values of affinities  $E_X$  and concentrations  $c$ , where the chain behaves as a SAW polymer because of binder-beads loops are unstable; the globular state, where the interaction is strong enough to fold the polymer in a compact conformation. In this phase the binders can form a disordered structure or, for higher  $c$  and, especially,  $E_X$  values, an ordered structure (crystalline-like) although they have not any

direct interaction with each other (Chiariello et al., 2016; Bianco et al., 2017). Different polymer conformations in the same class are thermodynamically indistinguishable. In **Figure 1B** three exemplificative structures are shown, one for each stable class.

## CHROMATIN AS A MIXTURE OF DIFFERENTLY FOLDED REGIONS

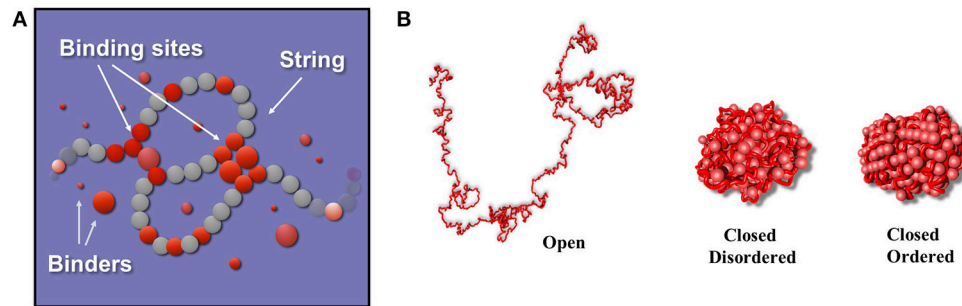
The thermodynamic classes just described represent the stable conformations of the polymer. It is well established from microscopy and Hi-C experiments that chromosomes are typically organized in eu- and heterochromatin, that is genome regions where chromatin is in an more open and in a more compact state respectively. This consideration points toward a simple, yet approximated picture of chromatin where single chromosomes are mixtures of differently folded regions, belonging to the stable classes (*pure states*) predicted by polymer physics (Annunziatella et al., 2016; Chiariello et al., 2016), as schematically depicted in **Figure 2A**.

To test this *pure state mixture* model of chromatin we compare the experimental average pairwise contact probability  $P(s)$  of two generic loci separated by a genomic distance  $s$  with the theoretical  $P(s)$ . In this scenario, the average  $P(s)$  results simply from a linear combination of the contact probabilities in the three pure state regimes. Thus, the theoretical  $P(s)$  depends only on the fractions of these states in the combination, which are best fit parameters (Barbieri et al., 2012; Chiariello et al., 2016). In **Figure 2B** we show the results of this approach in the analysis of another recent published Hi-C dataset in different chromosomes, from fibroblast (IMR90) cells (Rao et al., 2014). Superimposed in black are the corresponding mixture model fits, which highly match the experimental behavior (chi-squared test across all chromosomes ranging from  $6 \times 10^{-3}$  to  $2.5 \times 10^{-1}$ ). This analysis shows that this simple model quantitatively explains the average genome-wide or single chromosome contact data for a wide range of genomic lengths, from 0.5 Mb up to chromosomal scale, as found previously for other analyzed datasets (Chiariello et al., 2016).

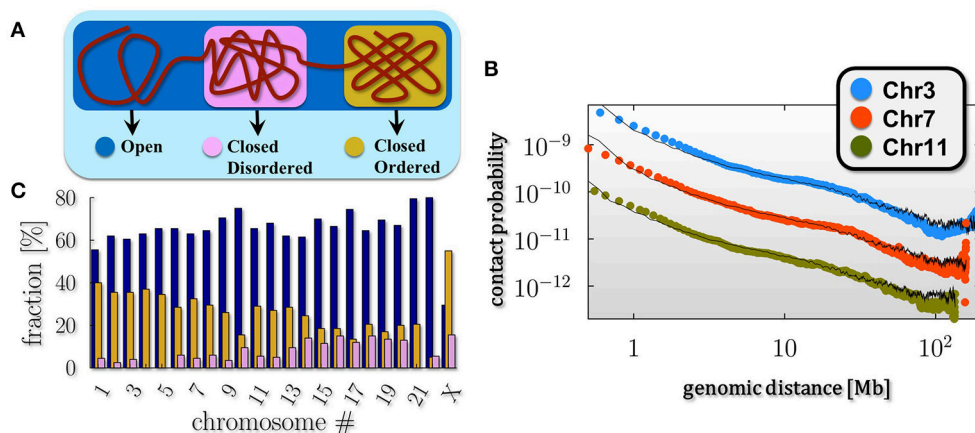
The mixture composition that best describes the experimental data for the chosen cell line is estimated (**Figure 2C**; errors below 10% of signal, not shown in the plot). It results to be strongly dependent on the considered chromosome: for instance, chromosome X is composed by mostly closed regions (above the 60%) with both the closed states (ordered and disordered) present (in a ratio about 3/2) while chromosomes richer in genes, such as chromosome 17 or 19, are up to 70% open in IMR90 cells. Briefly, the open-close combination reflects the distribution of differently folded domains along the chromosomes, across their thermodynamics states as also recently confirmed by the completely different experimental technique GAM (Beagrie et al., 2017).

## TOPOLOGICAL DOMAINS AND HIGHER-ORDER STRUCTURES

Beyond the average trend of contact frequencies with genomic separation, the SBS model can explain other specific features



**FIGURE 1** | The Strings and Binders model of chromatin. **(A)** The StringandBinders (SBS) model aims to quantify the scenario where chromatin folding is established by diffusing molecular factors such as Transcription Factors. The figure shows a toy version of the model including only one type of binders (red spheres):  $c$  is the binder concentration and  $E_X$  their binding affinity. **(B)** Such a simple toy model has three main stable conformational classes (images adapted from Chiariello et al., 2016): a *coil* state where the polymer conformation is an open Self-Avoiding Walk (SAW) chain randomly folded (polymer on the left); a disordered *globule* state where the polymer is closed in a compact, disordered conformation (center) and an ordered *globule* state where the polymer and the binders are in an ordered, crystalline-like, conformation (right).

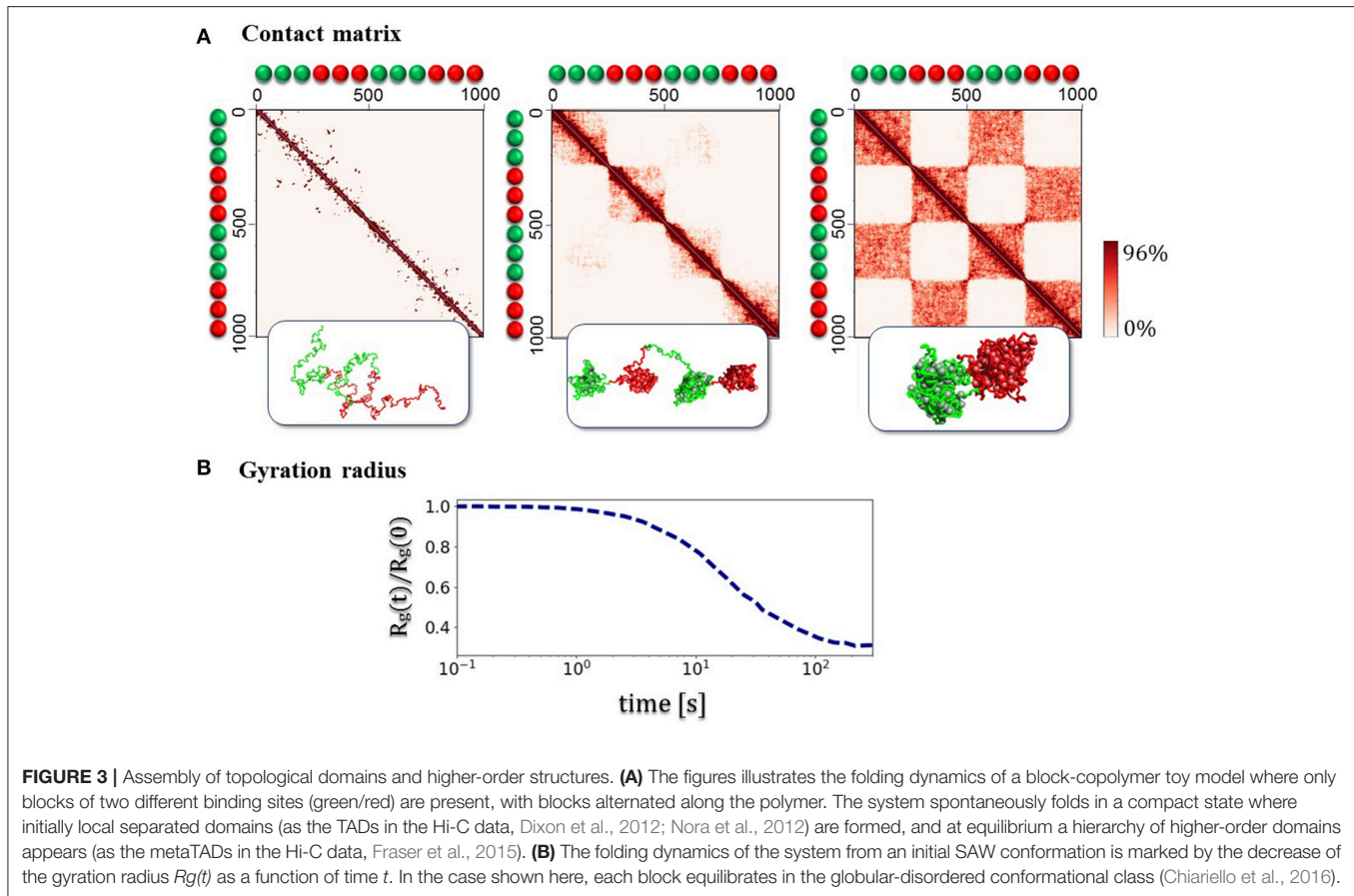


**FIGURE 2** | Chromosomes are a mixture of differently folded regions. **(A)** The polymer mixture model of chromatin poses that a chromosome is a mixture of differently folded region, each belonging to the stable folding classes described by polymer physics (pure states), as discussed in **Figure 1**. In this scenario the average pairwise contact frequency is determined only by the relative abundances of the pure states in the mixture. **(B)** Within such a model, the decay of the average pairwise contact data as a function of genomic separation of single chromosomes from in-situ Hi-C data in fibroblast (IMR90) cells (data from Rao et al., 2014) can be fitted over three orders of magnitudes in genomic separation. **(C)** The barplot shows the relative composition of the pure states across the different chromosomes for the IMR90 cell line: chromosome richer in genes, such as chr17 or chr19, result to be composed, for the most part, by open regions (up to 70%), while chromosome X, as expected, has a more compact conformation compared to the other chromosome, with percentage of closed state above the 60%.

of chromatin folding, such as the formation of Topological Domains (TADs, Dixon et al., 2012; Nora et al., 2012) and their higher-order hierarchical structure recently discovered (Fraser et al., 2015). To illustrate the basic concepts involved without delving into polymer physics details (see Nicodemi and Prisco, 2009; Barbieri et al., 2012, 2013; Chiariello et al., 2015, 2016; Annunziatella et al., 2016), we consider first a toy polymer of chromatin with two different types of beads and binders (block-copolymer model, **Figure 3**). The beads along the polymer are alternated in two pairs of blocks, which can interact only with the corresponding type of binders. Each polymer block can fold in different ways, according to thermodynamics phase it belongs to. The folding dynamics from the SAW initial configuration to the equilibrium state reveals some

relevant aspects of chromatin architecture. First, consecutive globular structures, visually similar to the TADs, experimentally observed in Hi-C and GAM data, appear. Then, higher order structures appear, generating the typical hierarchical structure from interactions between the different blocks (**Figure 3**). In the toy model discussed here, the equilibrium contact matrix presents a chessboard-like pattern (**Figure 3A**), which reflects the hierarchical organization of higher-order structures, resulting from intra- and inter-domain interactions, even at lower energies. The example shown in **Figure 3** refers to the case where an initially open polymer folds into the closed disordered state, as discussed before.

The simplified description of folding of the above toy model schematically illustrates the basic polymer mechanisms that



could underlie the formation of complex patterns seen in the experimental contact data. All details from polymer physics, and the underlying theory, can be found in the cited original papers (e.g., Nicodemi and Prisco, 2009; Barbieri et al., 2012; Annunziatella et al., 2016; Chiariello et al., 2016).

### 3D GENOME RECONSTRUCTION: THE *BMP7* LOCUS IN ESC-46C MURINE CELLS

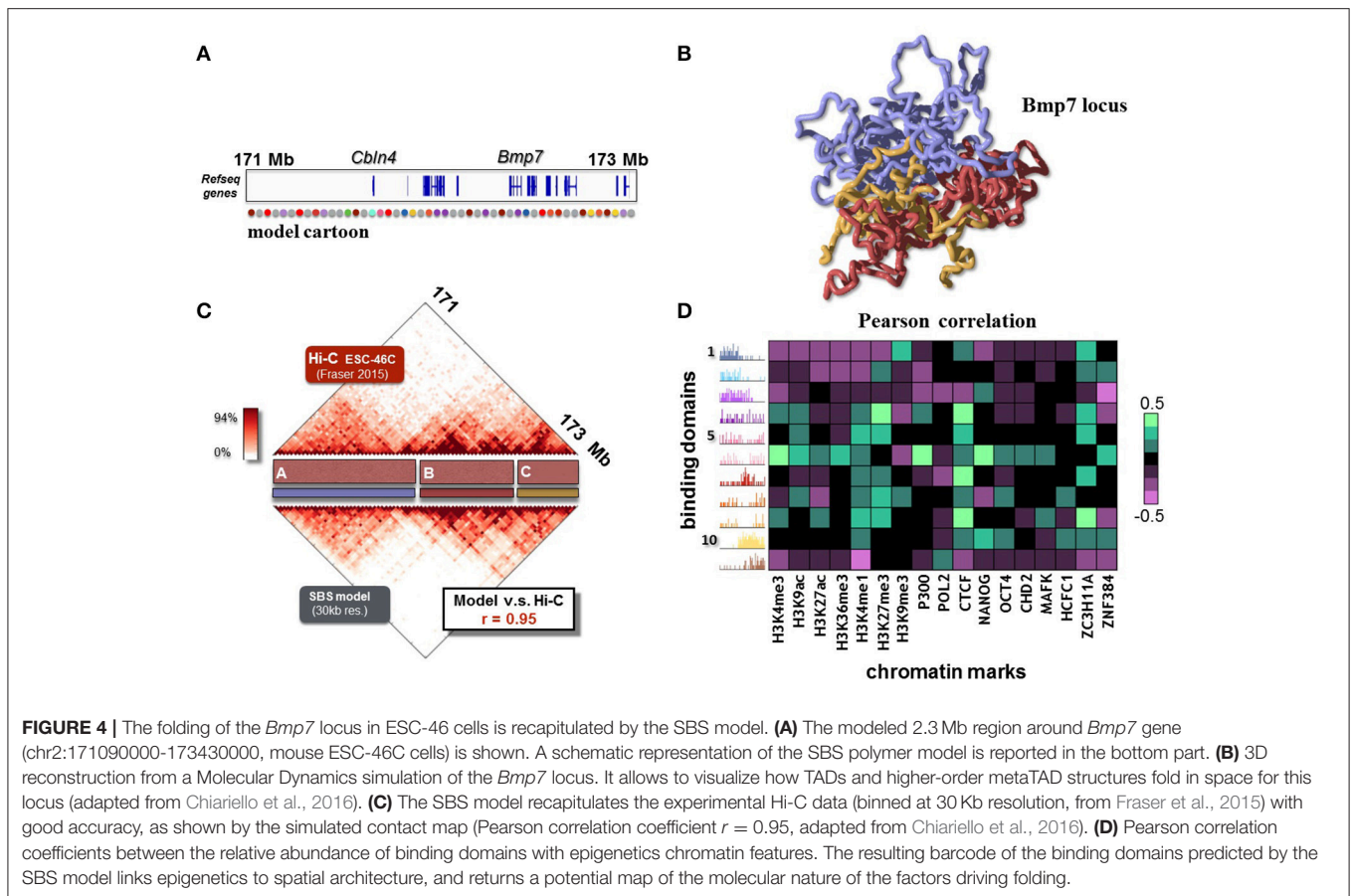
After reviewing the general mechanisms of polymer physics, within the framework of the SBS model, that can explain chromatin folding, next we focus on the description of experimental contact data of specific loci (Annunziatella et al., 2016; Chiariello et al., 2016). Polymer modeling can, in fact, explain real data to a good degree of accuracy and return the full 3D structure of the loci of interest. As example, we briefly review the results about the *Bmp7* locus, a 2.3 Mb region around the *Bmp7* gene (chr2:171090000-173430000, **Figure 4A**), that is important for osteoblastic differentiation (Bandyopadhyay et al., 2006). The SBS model can reproduce the Hi-C data from ESC-46C murine cells (at 30 Kb resolution) with high accuracy (Pearson correlation coefficient  $r = 0.95$ , **Figure 4B**).

In order to explain the finer details of the locus experimental Hi-C matrix, the model is of course more complex than the

two color toy polymer considered in the previous section. The *Bmp7* locus model requires 11 different types of binding sites (represented as colors in **Figure 4C**), and their corresponding binding molecules. The positions of binding sites on the polymer are estimated from the Hi-C data through a statistical inference algorithm and Monte-Carlo optimization (Annunziatella et al., 2016; Chiariello et al., 2016; Bianco et al., 2017). Interestingly, the inferred binding domains overlap roughly with the TADs visible in the Hi-C matrix. Nevertheless, they also compenentrate and partially overlap with each other, giving rise to the visible higher-order structures (Annunziatella et al., 2016; Chiariello et al., 2016). The 3D reconstruction of the model (**Figure 4B**) gives the 3D representation of the patterns contained in the Hi-C data. For example, the three principal domains of interactions labeled as A, B and C in **Figure 4C**, are clearly visible in the polymer snapshot (colored in medium purple, red and gold respectively). In SBS model these interactions naturally emerge because of the presence of molecules (binders) that mediate longer-range, higher-order interactions.

Finally, to give a sense of the molecular nature of the binding sites, we cross epigenomic databases of chromatin features with the relative abundance of the binding sites (Bianco et al., 2017). The heatmap of **Figure 4D** represents the Pearson correlation coefficients between the relative abundance of binding domains and chromatin features from ENCODE along the *Bmp7* region





for ESC-46C cells. Each binding domain is characterized by specific combinations of epigenomic features: some domains, in fact, correlate with active marks, while others are characterized by more repressive marks. Interestingly many domains correlate strongly with CTCF, a known chromatin organizer, whilst others are not linked to it. That suggests, for instance, that while CTCF has an important role in the folding of chromatin (Tark-Dame et al., 2014), additional remodeling factors also play a role, beyond CTCF, as shown in recent studies (Barbieri et al., 2017).

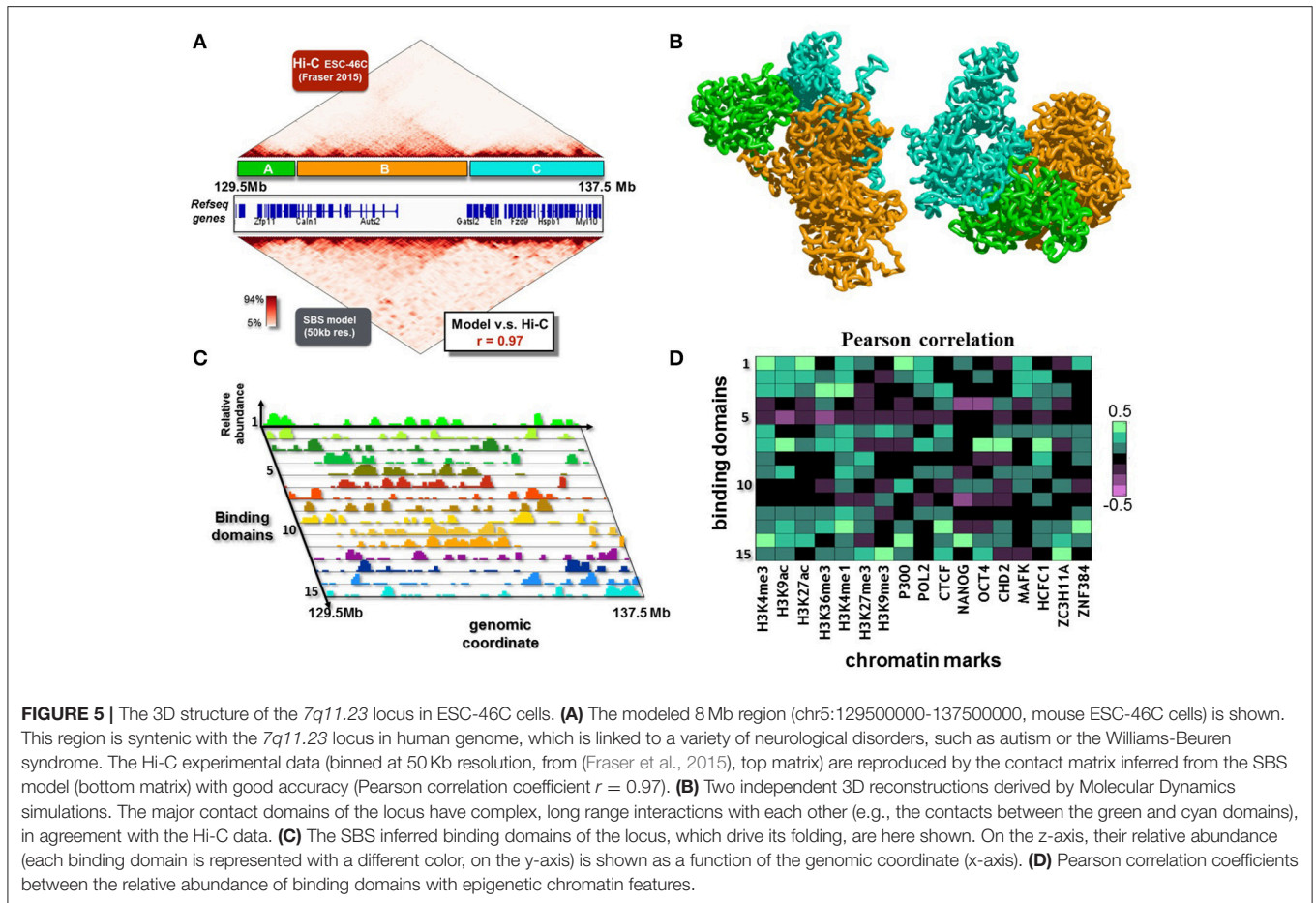
## APPLICATION TO NEUROGENETICS: THE 7Q11.23 LOCUS

As a final application of the model, we present here the first, albeit initial exploration of the 3D structure of the *7q11.23* locus, which is involved in severe neuropsychiatric disorders. Structural variants at the *7q11.23* locus can cause a variety of neurological, behavioral and other problems. For example, the *7q11.23* duplication syndrome is associated with speech problems and behavioral issues such as increased anxiety levels or autism (see Berg et al., 2007; Merla et al., 2010; Ramocki et al., 2010; Ebert et al., 2014). The Williams-Beuren syndrome (WSB) is a complex developmental disorder associated to the deletion of 1.5–1.8 Mb in the *7q11.23* locus, encompassing a couple of

dozens genes (see Nature Research Highlights, 2011; Sanders et al., 2011; Chailangkarn et al., 2016; and ref.s therein).

The attribution of the various features of WBS to specific genes is a complex, on-going effort, relying, among other strategies, on the phenotype-genotype analysis of patients with atypical deletions, and on genetically modified mouse models. Besides the role of the genes in the deleted/duplicated region or epigenetic mechanisms, a factor that may be implicated in determining the genotype-phenotype relationship is the effect of deletions and duplication on the 3D architecture of the genomic region. For such a reason, we report a first investigation of the architecture of the locus.

Here, we consider a 8 Mb region (chr5:129500000-137500000, **Figure 5A**) from mouse ESC-46C cell line, syntenic with the *7q11.23* locus in human genome. The dataset used is from Fraser et al. (2015), binned at 50 kb. The inferred contact matrix is highly correlated with the experimental matrix (**Figure 5A**), with a Pearson correlation coefficient  $r = 0.97$ . The polymer model of the locus involves 15 different types of binding sites, whose position and abundance along the DNA sequence is represented by the different color histograms in **Figure 5C**. The associated inferred conformations of the polymer model help explaining the 3D features of this locus. In **Figure 5B**, two possible configurations, obtained from independent simulations, are showed. For sake of clarity, we color in green, orange and



cyan respectively the three major domains visible in the matrix (labeled again as A, B and C, **Figure 5A**), so we can easily compare the contact pattern with the spatial reconstruction. At a first visual analysis, we can recognize the A, B and C domains as distinguishable and individual blocks, in agreement with the experimental data. Interestingly, a more deeper inspection reveal a non-random contact between domain A and domain C, which is again in agreement with the long range interaction contained in the experimental matrix, even though it is an higher-order detail. Importantly, a collection of much smaller domains (the strong red triangles in the contact matrix) close to the diagonal is evident, and a complex pattern of higher order interactions among them is present, so to give the typical hierarchical internal substructure to the major domains. The model also captures such lower level organization. For sake of clarity, we do not color all these domains in the polymer representation of **Figure 5B**. As the model does capture not only general aspects of the locus organization, but also its finer features, it can be used to derive relevant biological implications of the 3D structure.

To this aim, we analyse epigenomic databases of chromatin features and cross them with the relative abundance of the binding sites (Bianco et al., 2017), as made in the previous section. In **Figure 5D** are reported the results, which reveal a complex,

not trivial, pattern of correlations. Interestingly, several binding domains exhibit high correlations with many of the considered features, reflecting the biological complexity of the locus, highly enriched in genes (**Figure 5A**). Conversely, few of the binding domains (e.g., type 5 and 11, **Figures 5C,D**) do not correlate with the considered epigenomic features, and result to be associated with the central, gene poor, region of the locus. In summary, our polymer model of the murine genomic region syntenic with the *7q11.23* locus in human provides a first reconstruction of the ensemble of 3D conformation of the locus. In particular, a complex network of higher-order interactions of the locus emerges from our analysis, whose rewiring could be important to understand the effects of disease associated structural variants.

## DISCUSSION

In conclusion, we reviewed recent developments in polymer physics models to understand the 3D structure of genomic loci and the connection to human diseases. In particular, we discussed a polymer model of chromatin, the SBS model, where folding is determined by specific interaction with DNA-binding sites. Genome-wide and loci specific chromatin contact data can be explained over orders of magnitude in genomic

separation, and also other chromatin features as TADs (Dixon et al., 2012; Nora et al., 2012) and spontaneous hierarchical domains formation (Fraser et al., 2015). As mentioned in the Introduction, other polymer models, have been proposed. These models also reproduce successfully experimental data for different genomic regions. For instance, in the model proposed in Brackley et al. (2013), the  $\alpha$  and  $\beta$  globin gene regions in mouse erythroblasts are explained with a high degree of accuracy (Brackley et al., 2016), while the loop extrusion model reproduces the contact data of loci where CTCF factor is known to play an important role (Sanborn et al., 2015; Fudenberg et al., 2016). Another important example is the *Xist* locus in mouse embryonic stem cells (Scialdone et al., 2011; Giorgiotti et al., 2014; Chiariello et al., 2016). Non mammalian genomes have also been modeled, as *Drosophila* (Jost et al., 2014) and budding yeast (Kozioł et al., 2005; Cheng et al., 2015).

The SBS is, of course, a simplified polymer model of chromatin and many other factors are present in real cells, as confinement, crowding and entanglement effects. Additionally, off-equilibrium phenomena are likely to be present as well-known from the studies of complex fluids (see Caglioti et al., 1998; Nicodemi, 1998; Nicodemi and Coniglio, 1998; Coniglio and Nicodemi, 2000; Nicodemi and Jensen, 2001; Tarzia et al., 2004; Nicodemi and Prisco, 2007; Grebenkov et al., 2008, and references therein). The SBS has been also used to describe symmetry-break mechanisms in the *Xist* region during X-Chromosome Inactivation (Nicodemi et al., 2008), and to model the chromosome recognition and pairing in mitosis and meiosis (Ong and Corces, 2014). The SBS model has been employed to explain folding of a set of important loci, such as the *Hoxb* locus (Annunziatella et al., 2016; Barbieri et al., 2017) and the *Sox9* locus (Chiariello et al., 2016). Here, we illustrated the case of the *Bmp7* locus in ESC-46C murine cells (Chiariello et al., 2016).

As a novel application of potential interest to neurogenetics, we presented the reconstruction of the *7q11.23* locus architecture, where structural variants (deletions and duplications) are known to be linked to neurocognitive disorders such as autism spectrum disorders or the Williams–Beuren syndrome (WBS). For example, a clear genotype–phenotype correlation has been determined in WBS only for some genes (e.g., the elastin gene, which give rise to the vascular and connective tissue abnormalities). The molecular substrates carrying the other clinical aspects of *7q11.23* copy number variations (CNVs), including the neurocognitive phenotypes, are still not fully understood. Recent studies suggest that other factors, as regulatory sequences or epigenetic mechanisms, could have an important role in producing the variable expressivity of *7q11.23* CNV phenotypes, besides the role of the genes in the deleted/duplicated interval. Understanding the folding mechanism that regulates the 3D conformation of this region can also help a better comprehension of the link between the structural variants and the rewiring of

the contacts between the locus genes and their regulators. To that aim, we employed our polymer physics model. We showed that a complex pattern of interaction in this large genomic region is present. Investigating the spatial structure of this locus, and analogously of loci whose variants are involved in diseases, can help to better understand the contact landscape between their regulatory elements. In this way, the effect of structural variants on the spatial organization can be quantitatively studied, and the comprehension of the mechanisms causing the disease can be potentially widely improved.

Importantly, in our model it is not required additional *a priori* information about the molecular factors shaping the genome (for instance, CTCF sites position and other epigenetics features). Furthermore, it can be used also for modeling non mammalian genomes, whereas contact data are available, as in yeast (Hsieh et al., 2015). The putative responsible factors are purely derived by experimental contact data and their nature guessed by correlating their position with epigenomics datasets. That can be important to novel applications in biomedicine to diagnose *in-silico* diseases associated to improper chromatin folding, as cancer (Valton and Dekker, 2016) and congenital disorders (Ong and Corces, 2014; Lupiáñez et al., 2015).

## MATERIALS AND METHODS

We investigated the SBS polymer model by Brownian Molecular Dynamics simulations, implemented using LAMMPS (Plimpton, 1995). The complete description of methods and details about the model are provided in the cited references.

The data files of chromatin features in the *Bmp7* and *7q11.23* regions in mouse ESC used here are available online from ENCODE (The ENCODE Project Consortium, 2012), with the accession numbers: ENCFF434DJA, ENCFF916EVG, ENCFF945LRL, ENCFF001XWR, ENCFF001YIX, ENCFF001YAD, ENCFF251JZH, ENCFF817CZF, ENCFF001XWO, ENCFF001YIY, ENCFF001YIV, ENCFF830QDG, ENCFF001YAE, ENCFF796LDS, ENCFF883ROD, ENCFF530LPO, ENCFF854IVF.

## AUTHOR CONTRIBUTIONS

MN designed the study; CA, AE, SB, AC, and MN developed the project; AC, CA, SB, AE, and LF run the computer simulations and performed the analyses; AC, CA, AE, SB, AP, and MN wrote the manuscript.

## ACKNOWLEDGMENTS

Work supported by the NIH ID 1U54DK107977-01, CINECA ISCRA ID HP10CYFPS5 and HP10CRTY8P, and Einstein BIH Fellowship grants to MN, computer resources from INFN, CINECA, and *Scope* at the University of Naples.



## REFERENCES

- Annunziatella, C., Chiariello, A. M., Bianco, S., and Nicodemi, M. (2016). Polymer models of the hierarchical folding of the Hox-B chromosomal locus. *Phys. Rev. E* 94:042402. doi: 10.1103/PhysRevE.94.042402
- Bandyopadhyay, A., Tsuji, K., Cox, K., Harfe, B. D., Rosen, V., and Tabin, C. J. (2006). Genetic analysis of the roles of BMP2, BMP4, and BMP7 in limb patterning and skeletogenesis. *PLoS Genet.* 2:e216. doi: 10.1371/journal.pgen.0020216
- Barbieri, M., Chotalia, M., Fraser, J., Lavitas, L. M., Dostie, J., Pombo, A., et al. (2012). Complexity of chromatin folding is captured by the strings and binders switch model. *Proc. Natl. Acad. Sci. U.S.A.* 109, 16173–16178. doi: 10.1073/pnas.1204799109
- Barbieri, M., Scialdone, A., Piccolo, A., Chiariello, A. M., di Lanno, C., Prisco, A., et al. (2013). Polymer models of chromatin organization. *Front. Genet.* 4:113. doi: 10.3389/fgene.2013.00113
- Barbieri, M., Xie, S. Q., Torlai Triglia, E., Chiariello, A. M., Bianco, S., de Santiago, I., et al. (2017). Active and poised promoter states drive folding of the extended HoxB locus in mouse embryonic stem cells. *Nat. Struct. Mol. Biol.* 24, 515–524. doi: 10.1038/nsmb.3402
- Beagrie, R. A., Scialdone, A., Schueler, M., Kraemer, D. C. A., Chotalia, M., Xie, S. Q., et al. (2017). Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* 543, 519–524. doi: 10.1038/nature21411
- Berg, J. S., Brunetti-Pierri, N., Peters, S. U., Kang, S. H., Fong, C. T., Salamone, J., et al. (2007). Speech delay and autism spectrum behaviors are frequently associated with duplication of the *Tq11.23* Williams-Beuren syndrome region. *Genet. Med.* 9, 427–441. doi: 10.1097/GIM.0b013e3180986192
- Bianco, S., Chiariello, A. M., and Annunziatella, C., Esposito, A., and Nicodemi, M. (2017). Predicting chromatin architecture from models of polymer physics. *Chromosome Res.* 25, 25–34. doi: 10.1007/s10577-016-9545-5
- Bickmore, W. A., and van Steensel, B. (2013). Genome architecture: domain organization of interphase chromosomes. *Cell* 152, 1270–1284. doi: 10.1016/j.cell.2013.02.001
- Bohn, M., and Heermann, D. W. (2010). Diffusion-driven looping provides a consistent framework for chromatin organisation. *PLoS ONE* 5:e12218. doi: 10.1371/journal.pone.0012218
- Brackley, C. A., Brown, J. M., Waithe, D., Babbs, C., Davies, J., Hughes, J. R., et al. (2016). Predicting the three-dimensional folding of cis-regulatory regions in mammalian genomes using bioinformatic data and polymer models. *Genome Biol.* 17:59. doi: 10.1186/s13059-016-0909-0
- Brackley, C. A., Taylor, S., Papantonis, A., Cook, P. R., and Marenduzzo, D. (2013). Nonspecific bridging-induced attraction drives clustering of DNA-binding proteins and genome organisation. *Proc. Natl. Acad. Sci. U.S.A.* 110, E3605–E3611. doi: 10.1073/pnas.1302950110
- Caglioti, A., Coniglio, A., Herrmann, H. J., Loreto, V., and Nicodemi, M. (1998). Segregation of granular mixtures in presence of compaction. *Europhys. Lett.* 43, 591–597.
- Chailangkarn, T., Trujillo, C. A., Freitas, B. C., Hrvovj-Mihic, B., Herai, R. H., Yu, D. X., et al. (2016). A human neurodevelopmental model for Williams syndrome. *Nature* 536, 338–343. doi: 10.1038/nature19067
- Cheng, T. M. K., Heeger, S., Chaleil, R. A. G., Matthews, N., Stewart, A., Wright, J., et al. (2015). A simple biophysical model emulates budding yeast chromosome condensation. *Elife* 4:e05565. doi: 10.7554/eLife.05565
- Chiariello, A. M., Annunziatella, C., Bianco, S., Esposito, A., and Nicodemi, M. (2016). Polymer physics of chromosome large-scale 3d organisation. *Sci. Rep.* 6:29775. doi: 10.1038/srep29775
- Chiariello, A. M., Bianco, S., Piccolo, A., Annunziatella, C., Barbieri, M., Pombo, A., et al. (2015). Polymer models of the organization of chromosomes in the nucleus of cells. *Mod. Phys. Lett. B* 29:1530003. doi: 10.1142/S0217984915300033
- Coniglio, A., and Nicodemi, M. (2000). The jamming transition of granular media. *J. Phys.* 12, 6601–6610. doi: 10.1088/0953-8984/12/29/331
- Dekker, J., and Mirny, L. (2016). 3D Genome as Moderator of Chromosomal Communication. *Cell* 164, 1110–1121. doi: 10.1016/j.cell.2016.02.007
- Di Stefano, M., Paulsen, J., Lien, T. G., Hovig, E., and Micheletti, C. (2016). Hi-C-constrained physical models of human chromosomes recover functionally-related properties of genome organization. *Sci. Rep.* 6:35985. doi: 10.1038/srep35985
- Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., et al. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380. doi: 10.1038/nature11082
- Ebert, G., Steininger, A., Weißmann, R., Boldt, V., Lind-Thomsen, A., Grune, J., et al. (2014). Distribution of segmental duplications in the context of higher order chromatin organisation of human chromosome 7. *BMC Genomics* 15:537. doi: 10.1186/1471-2164-15-537
- Fraser, J., Ferrai, C., Chiariello, A. M., Schueler, M., Rito, T., Laudanno, G., et al. (2015). Hierarchical folding and reorganisation of chromosomes are linked to transcriptional changes during cellular differentiation. *Mol. Syst. Biol.* 11:852. doi: 10.15252/msb.20156492
- Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., Mirny, L. A., et al. (2016). Formation of chromosomal domains by loop extrusion. *Cell Rep.* 15, 1–12. doi: 10.1016/j.celrep.2016.04.085
- Giorgietti, L., Galupa, R., Nora, E. P., Piolot, T., Lam, F., Dekker, J., et al. (2014). Predictive Polymer modeling reveals coupled fluctuations in chromosome conformation and transcription. *Cell* 157, 950–963. doi: 10.1016/j.cell.2014.03.025
- Grebakov, D. S., Ciamarra, M. P., Nicodemi, M., and Coniglio, A. (2008). Force correlations and arch formation in granular assemblies. *Phys. Rev. Lett.* 100:078001. doi: 10.1103/PhysRevLett.100.078001
- Hsieh, T. H., Weiner, A., Lajoie, B., Dekker, J., Friedman, N., and Rando, O. J. (2015). Mapping nucleosome resolution chromosome folding in yeast by Micro-C. *Cell* 162, 108–119. doi: 10.1016/j.cell.2015.05.048
- Jost, D., Carrivain, P., Cavalli, G., and Vaillant, C. (2014). Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.* 42, 9553–9561. doi: 10.1093/nar/gku698
- Kozioł, S., Zagulski, M., Bilinski, T., and Bartosz, G. (2005). Antioxidants protect the yeast *Saccharomyces cerevisiae* against hypertonic stress. *Free Radic. Res.* 39, 365–371. doi: 10.1080/10715760500045855
- Lancôt, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.* 8, 104–115. doi: 10.1038/nrg2041
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragozcy, T., Telling, A., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293. doi: 10.1126/science.1181369
- Lupiani, D. G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161, 1012–1025. doi: 10.1016/j.cell.2015.04.004
- Merla, G., Brunetti-Pierri, N., Micale, L., and Fusco, C. (2010). Copy number variants at Williams-Beuren syndrome *Tq11.23* region. *Hum. Genet.* 128, 3–26. doi: 10.1007/s00439-010-0827-2
- Mistel, T. (2007). Beyond the sequence: cellular organization of genome function. *Cell* 128, 787–800. doi: 10.1016/j.cell.2007.01.028
- Nature Research Highlights (2011). Neurogenetics: extended hunt for autism genes. *Nature* 474, 254–255. doi: 10.1038/474254d
- Nicodemi, M. (1998). Force correlations and arch formation in granular assemblies. *Phys. Rev. Lett.* 80, 1340–1343. doi: 10.1103/PhysRevLett.80.1340
- Nicodemi, M., and Coniglio, A. (1998). Macroscopic glassy relaxations and microscopic motions in a frustrated lattice gas. *Phys. Rev. E* 57:R39.
- Nicodemi, M., and Jensen, H. J. (2001). Creep of superconducting vortices in the limit of vanishing temperature: a fingerprint of off-equilibrium dynamics. *Phys. Rev. Lett.* 86:4378. doi: 10.1103/PhysRevLett.86.4378
- Nicodemi, M., and Pombo, A. (2014). Models of chromosome structure. *Curr. Opin. Cell Biol.* 28C, 90–95. doi: 10.1016/j.celb.2014.04.004
- Nicodemi, M., and Prisco, A. (2007). Symmetry-breaking model for X-chromosome inactivation. *Phys. Rev. Lett.* 98:108104. doi: 10.1103/PhysRevLett.98.108104
- Nicodemi, M., and Prisco, A. (2009). Thermodynamic pathways to genome spatial organization in the cell nucleus. *Biophys. J.* 96, 2168–2177. doi: 10.1016/j.bpj.2008.12.3919
- Nicodemi, M., Panning, B., and Prisco, A. (2008). A thermodynamic switch for chromosome colocalization. *Genetics* 179, 717–721. doi: 10.1534/genetics.107.083154



- Nora, E. P., Lajoie, B. R., Schulz, E. G., Giorgetti, L., Okamoto, I., Servant, N., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385. doi: 10.1038/nature11049
- Ong, C. T., and Corces, V. G. (2014). CTCF: an architectural protein bridging genome topology and function. *Nat. Rev. Genet.* 15, 234–246. doi: 10.1038/nrg3663
- Phillips-Cremins, J. E., Sauria, M. E., Sanyal, A., Gerasimova, T. I., Lajoie, B. R., Bell, J. S., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281–1295. doi: 10.1016/j.cell.2013.04.053
- Plimpton, S. (1995). Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* 117, 1–19.
- Ramocki, M. B., Bartnik, M., Szafranski, P., Kolodziejska, K. E., Xia, Z., Bravo, J., et al. (2010). Recurrent distal *7q11.23* deletion including HIP1 and YWHAG identified in patients with intellectual disabilities, epilepsy, and neurobehavioral problems. *Am. J. Hum. Genet.* 87, 857–865. doi: 10.1016/j.ajhg.2010.10.019
- Rao, S. S., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., et al. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680. doi: 10.1016/j.cell.2014.11.021
- Rosa, A., and Everaers, R. (2008). Structure and dynamics of interphase chromosomes. *PLoS Comput. Biol.* 4:e1000153. doi: 10.1371/journal.pcbi.1000153
- Sanborn, A. L., Rao, S. S. P., Huang, S.-C., Duranda, N. C., Jewetta, A. I., Bochkova, I. D., et al. (2015). Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U.S.A.* 112, E6456–E6465. doi: 10.1073/pnas.1518552112
- Sanders, S. J., Ercan-Sencicek, A. G., Hus, V., Luo, R., Murtha, M. T., Moreno-De-Luca, D., et al. (2011). Multiple recurrent de novo CNVs, including duplications of the *7q11.23* Williams syndrome region, are strongly associated with autism. *Neuron* 70, 863–885. doi: 10.1016/j.neuron.2011.05.002
- Scialdone, A., Cataudella, I., Barbier, M., Prisco, A., and Nicodemi, M. (2011). Conformation regulation of the X chromosome inactivation center: a model. *PLoS Comput. Biol.* 7:e1002229. doi: 10.1371/journal.pcbi.1002229
- Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., et al. (2012). Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148, 458–472. doi: 10.1016/j.cell.2012.01.010
- Tanay, A., and Cavalli, G. (2013). Chromosomal domains: epigenetic contexts and functional implications of genomic compartmentalization. *Curr. Opin. Genet. Dev.* 23, 197–203. doi: 10.1016/j.gde.2012.12.009
- Tark-Dame, M., Jerabek, H., Manders, E. M., van der Wateren, I. M., Heermann, D. W., and van Driel, R. (2014). Depletion of the chromatin looping proteins CTCF and cohesin causes chromatin compaction: insight into chromatin folding by polymer modeling. *PLoS Comput. Biol.* 12:e1004716. doi: 10.1371/journal.pcbi.1003877
- Tarzia, M., de Candia, A., Fierro, A., Nicodemi, M., and Coniglio, A. (2004). Glass transition in granular media. *Europhys. Lett.* 66:531. doi: 10.1209/epl/i2004-10015-y
- The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. doi: 10.1038/nature11247
- Valton, A. L., and Dekker, J. (2016). TAD disruption as oncogenic driver. *Curr. Opin. Genet. Dev.* 36, 34–40. doi: 10.1016/j.gde.2016.03.008

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Chiariello, Esposito, Annunziatella, Bianco, Fiorillo, Prisco and Nicodemi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.