Bifractal nature of chromosome contact maps

Simone Pigolotti
*Biological Complexity Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan**

Mogens H. Jensen
The Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen Ø, Denmark

Yinxiu Zhan
Friedrich Miescher Institute for Biomedical Research, Maulbeerstrasse 66, 4058 Basel, Switzerland

Guido Tiana
Department of Physics and Center for Complexity and Biosystems, Università degli Studi di Milano and INFN, via Celoria 16, 20133 Milano, Italy

(Dated: October 13, 2020)

Modern biological techniques such as Hi–C permit to measure probabilities that different chromosomal regions are close in space. These probabilities can be visualised as matrices called contact maps. In this paper, we introduce a multifractal analysis of chromosomal contact maps. Our analysis reveals that Hi–C maps are bifractal, i.e. complex geometrical objects characterized by two distinct fractal dimensions. To rationalize this observation, we introduce a model that describes chromosomes as a hierarchical set of nested domains and we solve it exactly. The predicted multifractal spectrum is in excellent quantitative agreement with experimental data. Moreover, we show that our theory yields to a more robust estimation of the scaling exponent of the contact probability than existing methods. By applying this method to experimental data, we detect subtle conformational changes among chromosomes during differentiation of human stem cells.

I. INTRODUCTION

During cellular interphase, mammalian chromosomes assume a globular structure in the nucleus [1–5]. Their conformational properties can be studied *in vivo* with a set of techniques called chromosome conformation capture (3C) [6], most notably their genome-wide version called Hi–C [7]. Results of Hi–C experiments can be represented as matrices whose elements are proportional to the probability that two chromosomal regions are in contact in space [8]. Hi–C measurements have paved the way for a mechanistic understanding of chromosome folding [9–12]. In particular, they have revealed that mammalian chromosomes are characterized by a hierarchy of nested, tightly connected structures [13, 14]. At the scale of tens of kilobases one identifies 'contact domains' [15, 16]. Structures at the hundreds kilobases scale are usually called 'topological associating domains' (TADs) [17–19]. TADs have been extensively studied due to their essential role in gene regulation [12, 20–24], although they do not seem to be privileged over other levels in the hierarchy from a structural point of view [14]. At scales of few to tens of megabases, Hi–C experiments have identified 'compartments' [7]. Compartments are checkboard–like domains that are thought to be driven by mutually exclusive association between active and inactive chromatin [7, 25]. At the scale of the whole nucleus, microscopy and Hi–C experiments showed that chromosomes occupy distinct 'chromosomal territories' [7, 26].

A simpler approach to characterize the behavior of Hi–C matrices at different scales is to study the average contact probability $p(\ell)$ of pairs of chromosomal regions i and j with respect to their genomic distance $\ell = |i - j|$. Such decay appears to follow a power law

$$p(\ell) \sim \ell^{-\beta}. \tag{1}$$

The contact probability exponent β is often estimated to be slightly smaller than one [7, 15, 27, 28]. Such low values are incompatible with simple equilibrium homopolymeric models [29]. In contrast, non-equilibrium models such as the crumpled globule [7, 29] are able to account for such low exponents. Other proposed mechanisms include mediation of polymer interaction by other molecules [30], active loop-extrusion [31], and finite–size effects in heteropolymers [32]. In any case, the apparent power-law range of the contact probability is usually limited to one decade or less, see e.g. [7, 15, 27]. Therefore, estimates of the exponent β are rather sensitive to the choice of the fitting range.

A single physical mechanism is unlikely to explain the structure of chromatin at all scales. In fact, selective removal of proteins known to be involved in protein architecture can affect some levels of the hierarchy and not others [33–36]. Because of their importance in the control of gene expression, we focus our attention on the scales associated with TADs, i.e. below the megabase scale. It has been suggested that this hierarchy of structure can be analyzed by comparing statistical properties of Hi–C

* simone.pigolotti@oist.jp

matrices at different resolutions. [37].

In this paper, we robustly characterize scale-invariant properties of chromosomes at the scale of TADs using the theory of multifractals. This theory has been developed to characterize heterogeneous systems characterized by scale invariance [38–40]. This analysis reveals that chromosome contact maps are *bifractal*, i.e. geometric objects characterized by two distinct fractal dimensions. Bifractal behavior has been previously reported in studies of surface roughness [41], distribution of matter in the universe [42] and turbulence [43]. We show that multifractal theory also provides a robust estimation of the scaling of the contact probability at the scale of TADs.

The manuscript is organized as follows. In Section II we introduce the multifractal analysis using as example a Hi–C map of chromosome 1 of mouse embryonic stem cells. In Section III we introduce the hierarchical domain model, compute its multifractal spectrum and the scaling of its contact probability. We also show that the model predicts very accurately the multifractal spectrum of Hi–C maps. In Section IV, we apply our findings to a broader range of experimental datasets. We show how our theory can be used as a computational method to discern differences among Hi–C maps in different experiments. We show in particular that our method is able to characterize differentiation of human stem cells. Section V is devoted to conclusions and perspectives.

II. MULTIFRACTAL ANALYSIS

We introduce our idea using a Hi–C map of chromosome 1 in mouse embryonic stem cells, see Fig. 1a and Appendix A. The contact probability seems to decay as a power law, at least for relatively short genomic distances, see Fig. 1b. However, the local logarithmic slope of the contact probability does not present the clear plateau characteristic of a power law, see Fig. 1c.

To characterize scaling properties of chromosomes in a more robust way, we study the Hi–C map as a multifractal. A multifractal is a system described in terms of a density, that in our case is the density of counts in the Hi–C map. To study structures at different scales, we construct two-dimensional histograms of the Hi–C map with bins of different linear resolution ϵ, see Fig. 1d. Geometrical structures of linear size smaller than ϵ are not resolved in these maps. The smallest possible value of ϵ is the resolution of the original Hi–C map, in our case $\epsilon = 4 \cdot 10^4$ bp. We define the probability $p_{ij}(\epsilon)$ in bin at coordinates i, j at resolution ϵ. We always work with normalized maps, so that $\sum_{ij} p_{ij}(\epsilon) = 1$ for all choices of ϵ.

We assume the density to be scale invariant, at least for ϵ small enough:

$$p_{ij}(\epsilon) \sim \epsilon^{\alpha}, \tag{2}$$

where $\sim$ denotes the leading behavior and α is the scaling exponent associated with the density. Since the map is not homogeneous, different bins can be in principle characterized by different values of α. The number of bins $N(\alpha)$ associated with a given value of α must also scale as a power of ϵ

$$N(\alpha) \sim \rho(\alpha)\epsilon^{-D(\alpha)}. \tag{3}$$

The exponent $D(\alpha)$ characterizes the scaling of the number of bins of linear size ϵ necessary to cover the set with density exponent α and therefore can be interpreted as the fractal dimension associated with this set. The quantity $\rho(\alpha)$ is a prefactor independent of ϵ.

Computing $D(\alpha)$ directly is often unpractical. A convenient related quantity is the partition function $Z(q, \epsilon)$, defined by [39, 40]

$$Z(q, \epsilon) = \sum_{ij} [p_{ij}(\epsilon)]^q . \tag{4}$$

The name "partition function" originates from an analogy with statistical physics, where the exponent q plays the role of an inverse temperature. Indeed, in an non-homogeneous system, for $q \to 0$ (high temperature), all bins give similar contributions to the sum in Eq. (4), whereas for large q (low temperature) the sum is dominated by relatively few terms characterized by largest values of the measure p_{ij}. For a scale-invariant system, one also expects a power-law scaling of the partition function with the resolution,

$$Z(q, \epsilon) \sim \epsilon^{K(q)}, \tag{5}$$

at least for small ϵ. This happens to be the case for our Hi–C map, see Fig. 1e. In this case, for ϵ slightly larger than its minimum value, the local logarithmic slope of the partition function is essentially flat for a broad range of scales encompassing the typical sizes of TADs, see Fig. 1f. This signals that the power–law behaviour in Eq. (5) is much more robust than that of $p(\ell)$.

In the theory of multifractals, the function $K(q)$ defined in Eq. (5) is called the *multifractal spectrum*. The multifractal spectrum is related with the fractal dimensions $D(\alpha)$ by a Legendre transform. To show that, we collect in Eq. (4) all terms with the same value of α:

$$Z(q, \epsilon) \sim \sum_{\alpha} \rho(\alpha)\epsilon^{q\alpha - D(\alpha)}. \tag{6}$$

A saddle-point evaluation of Eq. (6) reveals that

$$K(q) = \min_{\alpha}[q\alpha - D(\alpha)], \tag{7}$$

as anticipated. As a consequence, a linear multifractal spectrum indicates that the system is homogeneous, i.e. all of its parts are characterized by the same fractal dimension $D = \alpha$ and hence by the same scaling behavior $N \sim \epsilon^{-D}$ and $p_{ij} \sim \epsilon^{D}$. Conversely, a non-linear multifractal spectrum signals a diversity of scaling exponents and associated dimensions. The multifractal spectrum of chromosome 1 presents two different linear regimes,

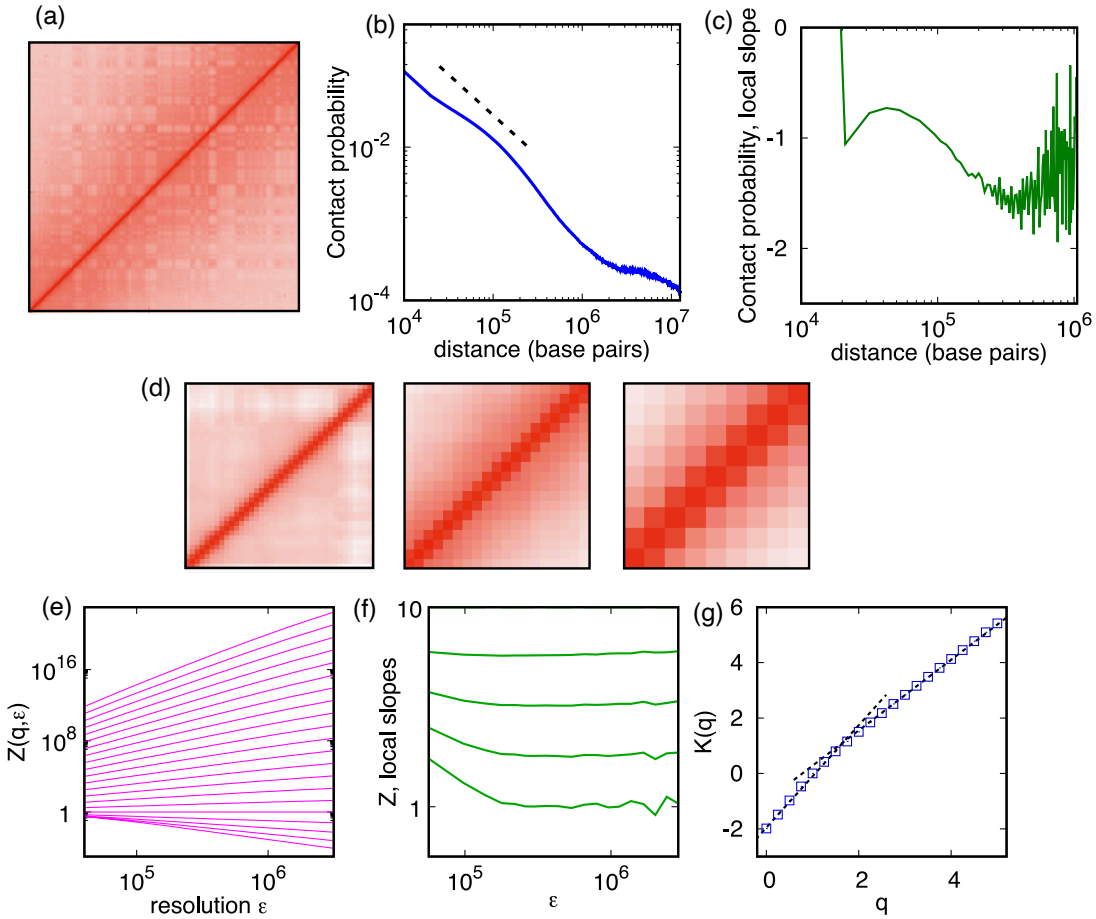

FIG. 1. **Contact probability scaling versus multifractal analysis of a Hi–C map.** (a) Hi–C map of chromosome 1 of mouse embryonic stem cells (mESC) [44]. Darker shades of red correspond to higher values of the contact probability p_{ij} (b) Scaling of the average contact probability $p(\ell)$ as a function of the distance $\ell = |i - j|$. The continuous line is a best-fit of a power law, Eq. (1), in the range of distances $[10^5, 10^6]$, yielding $\beta \approx 1.06$. (c) Local logarithmic slope $\mathrm{d}\ln P(\ell)/\mathrm{d}(\ln \ell)$ of the contact probability. (d) Coarse-graining of the Hi–C map at increasing values of the resolution $\epsilon = 10^6, 10^7, 2\ 10^7$ bp. (e) Scaling of $Z(q, \epsilon)$ as a function of the resolution ϵ, see Eqs. (4) and (5). Different lines correspond to different values of q, increasing from top to bottom from $q = 0$ to $q = 5$ at intervals of $\Delta q = 0.25$. (f) Local logarithmic slopes of the first momenta of $Z(q)$ ($q = 0, 0.25, 0.5, 0.75$). (g) Corresponding multifractal spectrum $K(q)$. Here and throughout the paper, spectra are obtained by a fit in log-log scale of $Z(q, \epsilon)$ versus ϵ in the range $\epsilon \in [10^5, 10^7]$ unless specified otherwise. Two straight dashed lines are shown to guide the eye. Python code to compute the multifractal spectrum is available at [45].

see Fig. 1g and can therefore be associated with two distinct fractal dimensions. A system with such properties is named a bifractal [41–43]. Since the exponent q is analogous to an inverse temperature, the sharp change of slope in the spectrum of a bifractal system is akin to a phase transition in equilibrium statistical physics [46].

III. HIERARCHICAL DOMAIN MODEL

To rationalize these observations, we introduce a hierarchical domain model of Hi–C maps. We define the model by an iterative transformation of a measure on the unit square $[0, 1] \times [0, 1]$. At the first iteration, the measure is given by a 2×2 matrix with diagonal element a and off–diagonal elements b, with $a > b > 0$. At each following iteration n we generate a $2^n \times 2^n$ matrix. The off-diagonal element of the matrix at the previous iteration becomes a 2×2 block of identical values in the matrix at the n-th iteration. The diagonal blocks are further multiplied by the original matrix. The procedure is illustrated in Fig. 2a and Supplementary Fig. S1. We impose $b = 1/2 - a$, so that the measure remains normalized at each iteration. This means that, effectively, the model is defined in terms of a single free parameter a. Because of the normalization and the requirement that the measure should be concentrated along the diagonal, such parameter falls in the range $1/4 \leq a \leq 1/2$.

Physically, the parameter a represents the weight of domains compared to the rest of the Hi–C map. Matrices with larger a have a measure more concentrated

along the diagonal, whereas matrices with smaller a are characterized by a more uniform measure, see Fig. 2a. In the limiting case $a = 1/4$ the matrix is uniform at all iterations, whereas for $a = 1/2$ the matrix is characterized by a uniform measure on the diagonal and all the off-diagonal elements are zero.

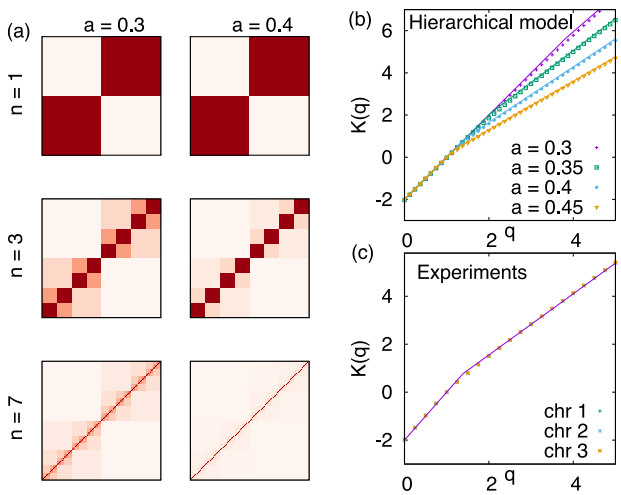

FIG. 2. **Hierarchical domain model.** (a) Construction of the hierarchical domain model for two different values of the parameter a. (b) Colored symbols are the multifractal spectrum calculated by the hierarchical domain mode at different values of a, the solid curve is the prediction of Eqs. (11) and (12). (c) Multifractal spectra of the first three chromosomes of mESC (colored symbols). The solid curve is a fit of Eqs. (11) and (12), resulting in $a \approx 0.413$. Here and in the following, all fits are performed using the least square method, unless specified otherwise.

We now analytically compute the multifractal spectrum of the hierarchical domain model. The partition function can be expressed as

$$Z(q,n) = \sum_{k=0}^{n} \exp[\xi(k)], \tag{8}$$

where we defined the exponent

$$\xi(k) = [n + \max(n-k-1,0)] \ln 2 + kq \ln a + \\ + (1-\delta_{kn}) q \ln b - \max(n-k-1,0) q \ln 4. \tag{9}$$

A saddle–point estimation of the partition function gives

$$\frac{d\xi}{dk} = -\ln 2 + q \ln 4a, \tag{10}$$

which is positive for $q > q_c = (\ln 2)/(\ln 4a)$. This means that, for $q \geq q_c$, the leading term is either $\xi(n)$ or $\xi(n-1)$. Since $\xi(n) - \xi(n-1) = q \ln(a/b)$, the maximum of the exponent is attained at $k = n$. Thus, for large n, the partition function scales as $Z(q,n) \sim \epsilon^{K(q)}$ with the length scale $\epsilon = 2^{-n}$ and the multifractal spectrum

$$K(q) = -q \frac{\ln a}{\ln 2} - 1. \tag{11}$$

For $a = 1/2$, the matrix $p_{ij}(\epsilon)$ is diagonal at each iteration. In this case, Eq. (11) predicts a linear spectrum with slope $D = 1$, consistent with the fact that the geometric set is equivalent to a one dimensional line. For $a = 1/4$ the distribution is uniform on the square, and Eq. (11) correctly returns $D = 2$. Between these two limiting cases, Eq. (11) predicts a fractal distribution with a dimension $D = -\ln a/\ln 2$ between 1 and 2.

We now focus on the "high temperature phase" where $-\ln 2 + q \ln 4a < 0$. In this case, the term dominating the scaling is $k = 0$, so that

$$K(q) = 2(q-1). \tag{12}$$

Since in the high temperature phase the scaling is determined by the terms far from the diagonal, the spectrum is that of a regular two-dimensional set.

Summarizing, the predicted multifractal spectrum is characterized by two linear regimes: one with slope $-\log(a)/\log(2)$ for $q > q_c = (\ln 2)/(\ln 4a)$, Eq. (11), and one with slope 2 for $q < q_c$, Eq. (12). Such predictions are in excellent agreement with numerical simulations, as shown in Fig. 2b, with very small discrepancies for high values of q and low values of a arising from finite size effects. These results confirm the validity of our saddle point approximation.

Strikingly, our theory predicted extremely well also the multifractal spectra of real chromosomes, with a fitted value of $a \approx 0.425$ and very little variability among the first three chromosomes of mouse embryonic stem cells (mean squared displacement MSD ≈ 0.0015 in all three cases), see Fig. 2c. To test whether this result is a unique signature of a hierarchical mechanism, we numerically computed the spectra of two null models. In the first null model, the contact probabilities are expressed as

$$p_{ij} = |i-j|^{-\beta} + b_{ij}, \tag{13}$$

where b_{ij} is equal to 1 if i and j belong to the same block of linear size M and zero otherwise. In the second null model, the contact probabilities are given by

$$p_{ij} = |i-j|^{-\beta} + \xi_{ij}. \tag{14}$$

The terms ξ_{ij} are independent, identically distributed random variables with average 0 and standard deviation σ. If the resulting value of p_{ij} is negative, it is rounded up to zero. In both models, the contact probabilities are normalized at the end of the procedure.

The solution of Eqs. (11) and (12) provides a poor fit to the first null model (MSD in the range $0.01 \sim 0.02$ depending on parameters), see Fig. 3a and 3b. The second null model provides a better fit, comparable with that of chromosomes for some values of the parameters, see Fig. 3. However, for realistic values of parameters ($\sigma \approx 3\ 10^{-5}$ and and $\beta \lesssim 1$) the quality of the fit is appreciably worse than that of maps from a real chromosome, see Fig. 3a and 3c).

Our theory accurately describes also Hi–C maps of Drosophila chromosomes (average MSD ≈ 0.001 for the

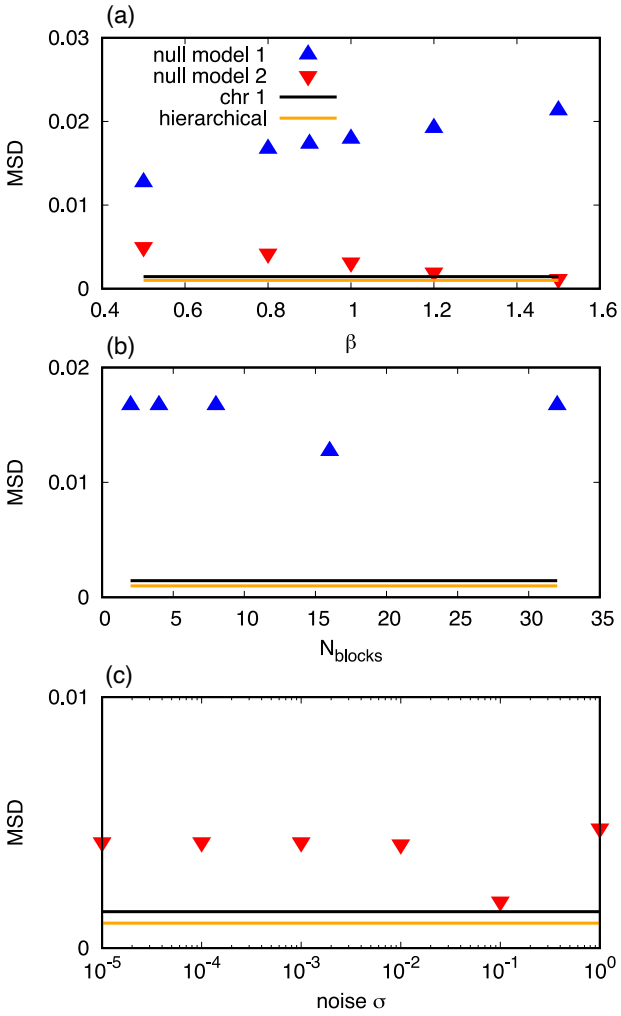

FIG. 3. **Performance of null models.** Points and curves show the mean square deviation (MSD) between the multifractal spectrum computed from different sets of data and the one predicted by Eqs. (11) and (12) upon fitting the parameter a. Blue upward and red downward triangles denote fits of the multifractal spectrum computed from the null models expressed by Eqs. (13) and (14), respectively. In (a), fits are shown as a function of β for fixed $N_{\rm blocks} = N/M = 15$ for the first null model, and fixed $\sigma = 3\ 10^{-5}$ for the second null model. In (b), fits of the first null model are shown as a function of $N_{\rm blocks}$ for fixed $\beta = 0.5$. In (c), fits of the second null model are shown as a function of σ for fixed $\beta = 0.8$. In all panels, the dark black line indicates the MSD of a fit of experimental data from mouse chromosome 1. The light orange line indicates the MSD associated with a numerical fit of the hierarchical model itself, with $a = 0.45$ and $n = 10$ iterations.

first four chromosomes, see Supplementary Fig. S2) and of human chromosomes (average MSD ≈ 0.005 for all chromosomes except chromosome Y, see Supplementary Fig. S3). These observations support that the bifractal spectrum predicted by the hierarchical domain model is compatible with that observed in a broad class of higher organisms.

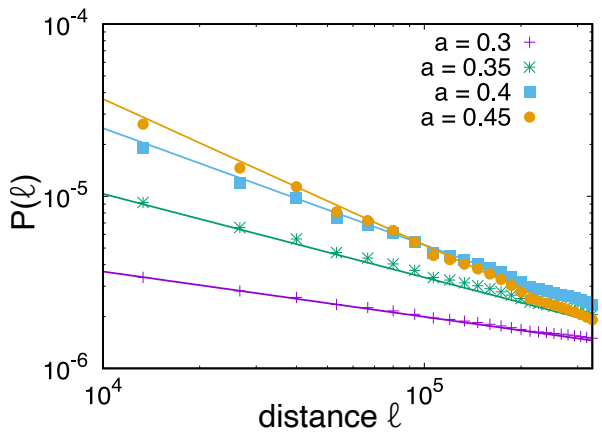

FIG. 4. **Contact probability in the hierarchical domain model.** Decay of contact probability $P(\ell; n)$ as a function of the genomic distance ℓ in the hierarchical model for $n = 10$ and different values of the parameter a, shown in the figure legend. Continuous lines are the exponent predictions of Eq. (15).

In the hierarchical domain model, the contact probability $P(\ell; n)$ decays as a power law of the genomic distance (see Eq. (1)), with an exponent

$$\beta = \frac{\ln(4a)}{\ln(2)}. \tag{15}$$

Derivation of Eq. (15) is presented in Appendix B. For $1/2 \leq a \leq 1$, Eq. (15) predicts contact probability exponents in the range $\beta \in [0, 1]$. This prediction is supported by numerical simulations, see Fig. 4. This means that the hierarchical structure of contact matrices automatically leads to exponents $\beta \leq 1$, as observed in chromosomes.

IV. COMPARISON WITH EXPERIMENTAL DATA

We test more extensively our method using datasets from different Hi–C experiments. We fit the multifractal spectrum of each dataset and obtain the corresponding exponent β via Eq. (15). Python scripts to perform this analysis on any Hi–C dataset are freely available [45]. We first test the robustness of the multifractal analysis across two Hi–C experiments in mouse embryonic stem cells (mESC) from two different labs [33, 44], both following the Hi–C protocol *in solution*, see Fig. 5a. Both the multifractal and the power law methods predict that the variability of the exponent β across chromosomes is significantly larger than the variability of the exponent across replicates. To determine which method performs best, we implement a bootstrap test with the null hypothesis that the values of β of different chromosomes are paired at random. The multifractal method permits to exclude

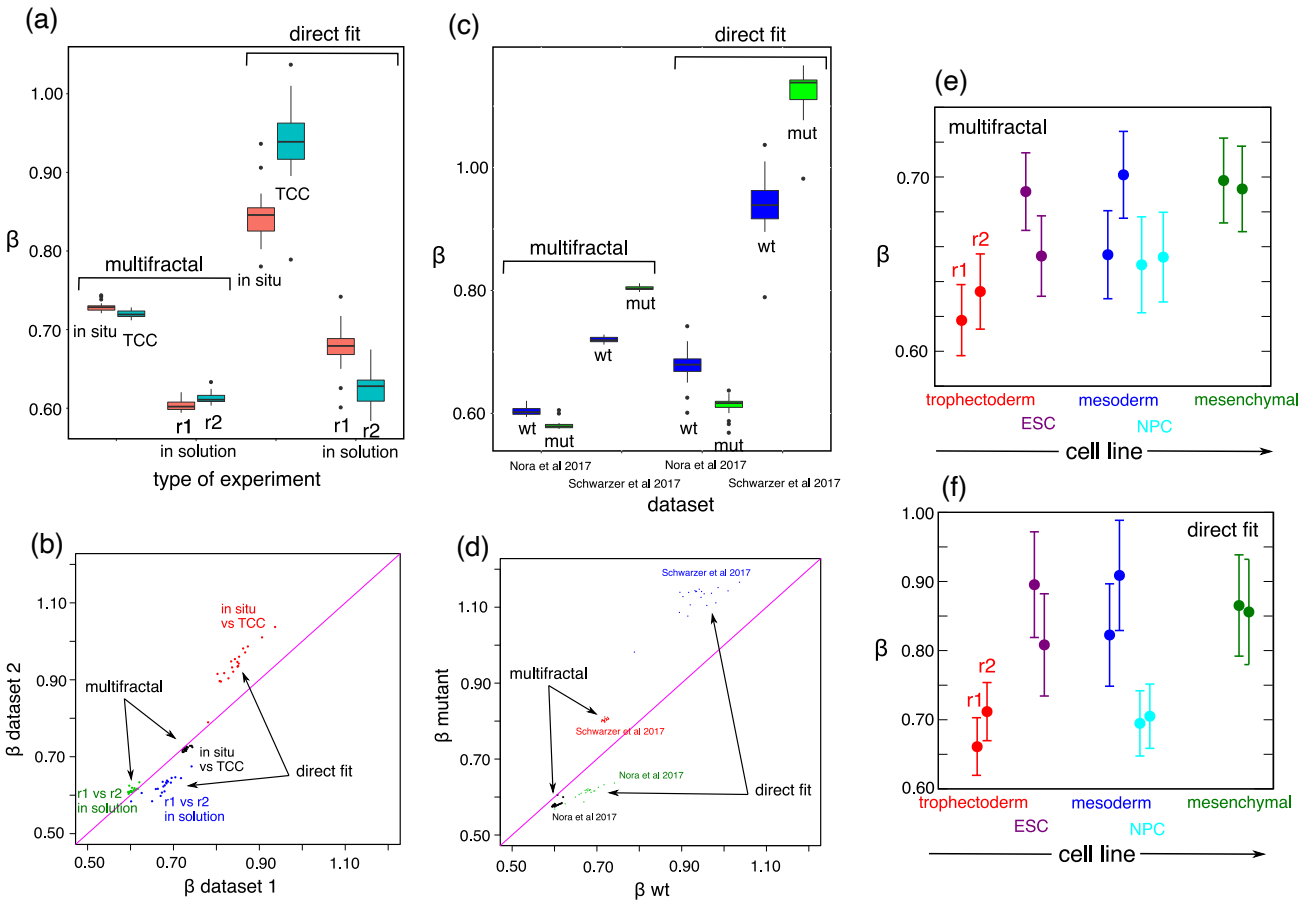

FIG. 5. **Application of multifractal analysis to experimental Hi–C maps.** (a) Exponents β calculated from the multifractal spectrum using Eq. (15) and from a direct fit of mouse embryonic stem cells (as in Fig. 1b) measured with different Hi–C methods. In particular, we analyze results of Hi–C experiments of in situ [8], TCC [47] and two independent replicates of in solution [33, 44]. Black bars indicate medians over chromosomes, colored boxes mark the first and third quartiles and black dots are outliers. (b) Scatter plot of the data displayed in panel a. (c) Value of β obtained for the wild–type (wt) system and for a mutant (mut) in which either CTCF [33] or Nipbl [47] is depleted in experiments. (c) (d) Scatter plot of the data displayed in panel c. (e) Exponents β obtained with the multifractal method for cell lines at different stages of human stem cell development [48]. (f) Exponents obtained by direct fit of the same data.

random pairing of chromosomes (p–value $5.2{\cdot}10^{-4}$) much better than the direct fit (p–value 0.11), see Supplementary Table S1. Moreover, the mean squared difference of β between the two replicates is smaller in the multi–fractal case compared to direct fit (0.009 vs 0.04). The χ^2 associated with the direct fit is affected by a strong systematic error, although remaining quite correlated. This effect is much milder in the multifractal approach, see Fig. 5b. The multifractal and direct fit methods are similarly robust with respect to varying the resolution of Hi–C maps, see Supplementary Fig. S4.

We apply the two methods to attempt to distinguish among experiments on the same cell lines, but following different experimental protocols. These different protocols mainly differ in the ligation step of digested DNA fragments. In the original "in solution" protocol, the ligation is performed in a diluted solution [49]. In other protocols, the ligation step is either carried out in intact nuclei ("in situ" protocol [49]) or on solid substrates ("TCC" protocol [50]). The two latter protocols are able to produce Hi-C maps with better signal-to-noise ratio compared to the in solution protocol [49, 50]. Both the multifractal and the direct fit methods show that the values of β obtained from in situ and TCC experiments are markedly different from those obtained in solution, see Fig. 5a). The multifractal method estimates compatible scaling exponents for in situ and TCC protocols. The p–value associated with the null hypothesis that chromosomes are paired at random is 0.021, see Supplementary Table S2 and Fig. 5b. A direct fit does not permit to draw this conclusion (p–value 0.14). This result suggests that the in situ and TCC protocols result in statistically

compatible Hi–C maps due to their high signal-to-noise ratio.

We compare the two methods in detecting differences between wild–type and mutant cell lines, in which either the Nipbl gene [47] or the CTCF gene [33] is knocked down (see Fig. 5c). Knock–down of these genes has been shown to disrupt chromosome folding. In particular, Nipbl knock–down leads to loss of TAD structures and global changes in scaling properties [47]. We quantify average differences in exponents between the wild–type and the mutant in terms of χ^2 of pairs of chromosomes, weighted by their mean squared errors, see Supplementary Fig. S5. Multifractal analysis detects more marked differences ($\chi^2 = 214$) compared to the direct fit ($\chi^2 = 4.7$), although both methods highlight a statistically significant difference between the two sets (p–values $1.3 \cdot 10^{-5}$ for the direct fit and $< 10^{-6}$ for the multifractal). Knock–down of CTCF also causes a loss of TAD structure, but without a clear effect on genome–wide scaling [33]. Nonetheless, the values of β of different chromosomes obtained by multifractal analysis reveal a large ($\chi^2 = 86$) and significant (p–value $< 10^{-6}$) difference between the scaling properties of wild–type and CTCF–deficient cells. Also in this case, the direct fit detects less marked differences ($\chi^2 = 25$), see Fig. 5c and Fig. 5d).

We apply the multifractal analysis to elucidate how chromosome structure changes during cellular differentiation. To this aim, we analyze Hi–C data obtained from different human cell lines at different stages of early development [48]. In temporal order, one differentiation branch includes the cell lines: trophectoderm, embyonic stem cells (ESC), mesoderm, and mesenchymal. Another differentiation branch is in order trophectoderm, embyonic stem cells (ESC), neuronal precursor cells (NPC) cells. The value of β obtained by multifractal analysis (see Fig. 5e) tends to increase upon differentiation. We first tested for significance applying Kendall's tau test to all chromosome pairs in different cell lines, excluding Mesoderm-NPC pairs since they are at a similar differentiation stages. We obtained a p–value $< 10^{-15}$ for the multifractal test versus 10^{-3} for the power law fit, see Fig. 5f. Testing for ordering of cell lines in each differentiation branch separately results in p–values $< 10^{-18}$ (multifractal) versus $2 \cdot 10^{-6}$ (power law fit) for the first branch, and 10^{-4} (multifractal) versus 0.5 (power law fit) for the second branch. In this latter case, the direct power law fit is not sensitive enough to detect a significant ordering. The increase of β with the developmental stage points to the idea that more differentiated cell types require more insulated domains to achieve more specialised function.

V. CONCLUSIONS

Multifractal analysis of Hi–C data is a powerful statistical tool to characterize scaling properties of chromosomes. We have shown that contact maps of chromosomes are bifractal, i.e. they are characterized by two distinct fractal dimensions. To understand this observation, we proposed a hierarchical domain model, whose analytical solution is in striking quantitative agreement with observed multifractal spectra. Our theory implicates a power-law scaling of the contact probability with exponents β lower than one, in agreement with experimental observations. The multifractal method is sensitive enough to discard a null model with power-law decay of the contact probability, but domains on a single length scale only. We found that another null model, in which the contact probability is a sum of a power law plus noise, is able to produce a similar spectrum, but only for unrealistically large values of either the exponent β or the noise intensity. The predicted form of the multifractal spectrum provides a more stringent prediction than the contact probability exponent alone. The analysis proposed here can be used as a stringent benchmark to select among different polymer models that provide similar values of β [29–32].

Our results indicate that scaling properties of chromosomes are a direct consequence of the hierarchical structure of chromosome domains, at least at the level of TADS [14, 16]. Recent work has suggested that such domain structure is generated by a "hierarchical folding" mechanism, mediated by different proteins such as cohesin and CTCF [10]. The precise mechanism driving the folding of chromosomes has been subject of debate [11]. It will be interesting to test whether the activity of these factors can produce self-similar structures compatible with our observations. The shape of the multifractal spectrum is controlled by a single parameter, that also controls the contact probability exponent β via Eq. (15). The determination of β from the fit of the multifractal spectrum is a much more robust way of characterizing the Hi–C map than the direct fit of β. Multifractal analysis is also more sensitive in highlighting subtle structural differences as shown in the case of Nipbl and CTCF knock–down. The analysis of the multifractal spectrum is simple and robust enough to become a routinely tool to analyze contact maps, both from polymer models and experiments, and capture subtle differences among them.

Appendix A: Hi–C data analysis

We reanalyzed published datasets (Table 1) using HiC-Pro v.2.11.1 [51] to maintain a consistent data processing pipeline. Briefly, we mapped reads to the corresponding genome (mm10 for mouse, hg19 for human and dm6 for fly) retrieving chimeric reads and keeping only unique mappable reads. We divided genomes into bins of fixed sizes (40kb unless specified otherwise) and made an histograms of reads. We applied Ice normalization [52] to the binned matrices. We then applied library size normalization to allow comparison across samples.

Cell type	Organism	GEO	Condition	Protocol	Study
ES cells	Mus Musculus	GSE128015	Wild Type	In situ	[8]
ES cells	Mus Musculus	GSE93431	Wild Type	TCC	[47]
ES cells	Mus Musculus	GSE72697	Wild Type	In solution	[44]
ES cells	Mus Musculus	GSE98671	Wild Type	In solution	[33]
ES cells	Mus Musculus	GSE98671	CTCF KD	In solution	[33]
ES cells	Mus Musculus	GSE93431	Nipbl KD	TCC	[47]
Trophectoderm	Homo sapiens	GSE52457	Wild Type	In solution	[48]
ES cells	Homo sapiens	GSE52457	Wild Type	In solution	[48]
NP cells	Homo sapiens	GSE52457	Wild Type	In solution	[48]
Mesoderm	Homo sapiens	GSE52457	Wild Type	In solution	[48]
Mesenchymal	Homo sapiens	GSE52457	Wild Type	In solution	[48]

TABLE I. Hi–C datasets analysed in this study. Raw data are downloaded from the GEO database at the relative accession number.

Appendix B: Scaling of contact probability in the hierarchical domain model

We derive the scaling of the contact probability in the hierarchical domain model. The contact probability $P(\ell;n)$ at the n-th iteration can be expressed by summing the probabilities of blocks at a genomic distance ℓ from the diagonal:

$$P(\ell;n) = \sum_{ij} p_{ij}(n)\delta_{|i-j|,\ell} \tag{B1}$$

with the distance ℓ. To find an explicit expression for $P(\ell;n)$, we start from the expression of the partition function in Eqs. (8) and (9). The only difference between $P(\ell;n)$ and $Z(1,n)$ is the Kronecker delta in (B1), that selects a subset of terms at a given genomic distance. It can be seen from the structure of the matrix that the number of terms with a given power of a first increases linearly with ℓ up to a maximum, then decreases linearly to zero. This means that $P(\ell;n)$ can be expressed as

$$P(\ell;n) = \sum_{k=0}^{n-1} 2^n g(\ell\ 2^{k-n}) \frac{a^k b^{(1-\delta_{kn})}}{4^{\max(n-k-1,0)}}, \tag{B2}$$

where $g(x)$ is defined to be x if $0 \le x \le 1/2$ and $1-x$ if $1/2 < x \le 1$. We now approximate the sum with an integral:

$$\begin{aligned} P(\ell;n) \approx & \int_{-\infty}^{n-\frac{\ln 2\ell}{\ln 2}} \mathrm{d}k\ \ell \exp\left[k\ln 2 + k\ln a + \right. \\ & \left. + (1-\delta_{kn})\ln b - \max(n-k-1,0)\ln 4\right] \\ & + \int_{n-\frac{\ln 2\ell}{\ln 2}}^{n-\frac{\ln \ell}{\ln 2}} \mathrm{d}k\ \exp\left[n\ln 2 + \ln\left(1-\ell\ e^{(k-n)\ln 2}\right)\right. \\ & \left. + k\ln a + (1-\delta_{kn})\ln b - \max(n-k-1,0)\ln 4\right] \end{aligned} \tag{B3}$$

For large n, the scaling is determined by the maximum of the argument of the exponential. This maximum is attained at a value k^* given by

$$\ell\ 2^{(k^*-n)} = \frac{\ln(4a)}{\ln(4a)+\ln 2}. \tag{B4}$$

Therefore the decay exponent of the contact probability with the genomic distance is $\beta = \ln(4a)/\ln(2)$, as reported in Section III.

ACKNOWLEDGMENTS

We thank Luca Giorgetti for help with the data.

[1] L. Giorgetti, R. Galupa, E. P. Nora, T. Piolot, F. Lam, J. Dekker, G. Tiana, and E. Heard, Cell **157**, 950 (2014).
[2] J. Dekker, Epigenetics & chromatin **7**, 25 (2014).
[3] G. Tiana, A. Amitai, T. Pollex, T. Piolot, D. Holcman, E. Heard, and L. Giorgetti, Biophysical journal **110**, 1234 (2016).
[4] J. Dekker and L. Mirny, Cell **164**, 1110 (2016).
[5] R. P. McCord, N. Kaplan, and L. Giorgetti, Molecular Cell **77** (2020).
[6] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, science **295**, 1306 (2002).
[7] 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, *et al.*, science **326**, 289 (2009).
[8] J. Redolfi, Y. Zhan, C. Valdes-Quezada, M. Kryzhanovska, I. Guerreiro, V. Iesmantavicius, T. Pollex, R. S. Grand, E. Mulugeta, J. Kind, *et al.*, Nature structural & molecular biology **26**, 471 (2019).
[9] A. Tanay and G. Cavalli, Current opinion in genetics & development **23**, 197 (2013).
[10] B. Bonev and G. Cavalli, Nature Reviews Genetics **17**,

661 (2016).
[11] M. J. Rowley and V. G. Corces, Nature Reviews Genetics **19**, 789 (2018).
[12] Q. Szabo, F. Bantignies, and G. Cavalli, Science advances **5**, eaaw1668 (2019).
[13] J. Fraser, C. Ferrai, A. M. Chiariello, M. Schueler, T. Rito, G. Laudanno, M. Barbieri, B. L. Moore, D. C. Kraemer, S. Aitken, *et al.*, Molecular systems biology **11** (2015).
[14] Y. Zhan, L. Mariani, I. Barozzi, E. G. Schulz, N. Blüthgen, M. Stadler, G. Tiana, and L. Giorgetti, Genome research **27**, 479 (2017).
[15] A. L. Sanborn, S. S. Rao, S.-C. Huang, N. C. Durand, M. H. Huntley, A. I. Jewett, I. D. Bochkov, D. Chinnappan, A. Cutkosky, J. Li, *et al.*, Proceedings of the National Academy of Sciences **112**, E6456 (2015).
[16] S. S. 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, *et al.*, Cell **159**, 1665 (2014).
[17] J. R. Dixon, S. Selvaraj, F. Yue, A. Kim, Y. Li, Y. Shen, M. Hu, J. S. Liu, and B. Ren, Nature **485**, 376 (2012).
[18] 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, *et al.*, Nature **485**, 381 (2012).
[19] T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc, M. Hoichman, H. Parrinello, A. Tanay, and G. Cavalli, Cell **148**, 458 (2012).
[20] D. G. Lupiáñez, K. Kraft, V. Heinrich, P. Krawitz, F. Brancati, E. Klopocki, D. Horn, H. Kayserili, J. M. Opitz, R. Laxova, *et al.*, Cell **161**, 1012 (2015).
[21] O. Symmons, L. Pan, S. Remeseiro, T. Aktas, F. Klein, W. Huber, and F. Spitz, Developmental cell **39**, 529 (2016).
[22] W. A. Flavahan, Y. Drier, B. B. Liau, S. M. Gillespie, A. S. Venteicher, A. O. Stemmer-Rachamimov, M. L. Suvà, and B. E. Bernstein, Nature **529**, 110 (2016).
[23] D. Hnisz, A. S. Weintraub, D. S. Day, A.-L. Valton, R. O. Bak, C. H. Li, J. Goldmann, B. R. Lajoie, Z. P. Fan, A. A. Sigova, *et al.*, Science **351**, 1454 (2016).
[24] L. L. Hanssen, M. T. Kassouf, A. M. Oudelaar, D. Biggs, C. Preece, D. J. Downes, M. Gosden, J. A. Sharpe, J. A. Sloane-Stanley, J. R. Hughes, *et al.*, Nature cell biology **19**, 952 (2017).
[25] M. Falk, Y. Feodorova, N. Naumova, M. Imakaev, B. R. Lajoie, H. Leonhardt, B. Joffe, J. Dekker, G. Fudenberg, I. Solovei, *et al.*, Nature **570**, 395 (2019).
[26] T. Cremer and M. Cremer, Cold Spring Harb. Perspect. Biol. **2**, a003889 (2010).
[27] B. Bonev, N. M. Cohen, Q. Szabo, L. Fritsch, G. L. Papadopoulos, Y. Lubling, X. Xu, X. Lv, J.-P. Hugnot, A. Tanay, *et al.*, Cell **171**, 557 (2017).
[28] Y. Zhan, L. Giorgetti, and G. Tiana, Chromosome Research **25**, 5 (2017).
[29] L. A. Mirny, Chromosome research **19**, 37 (2011).
[30] M. Barbieri, M. Chotalia, J. Fraser, L.-M. Lavitas, J. Dostie, A. Pombo, and M. Nicodemi, Proceedings of the National Academy of Sciences **109**, 16173 (2012).
[31] G. Fudenberg, M. Imakaev, C. Lu, A. Goloborodko, N. Abdennur, and L. A. Mirny, Cell reports **15**, 2038 (2016).
[32] Y. Zhan, L. Giorgetti, and G. Tiana, Physical Review E **94**, 032402 (2016).
[33] E. P. Nora, A. Goloborodko, A.-L. Valton, J. H. Gibcus, A. Uebersohn, N. Abdennur, J. Dekker, L. A. Mirny, and B. G. Bruneau, Cell **169**, 930 (2017).
[34] W. Schwarzer, N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, Y. Loe-Mie, N. A. Fonseca, W. Huber, C. Haering, L. Mirny, and F. Spitz, Nature **152**, 1270 (2017).
[35] J. H. I. Haarhuis, R. H. van der Weide, V. A. Blomen, J. O. Yáñez-Cuna, M. Amendola, M. S. van Ruiten, P. H. L. Krijger, H. Teunissen, R. H. Medema, B. van Steensel, T. R. Brummelkamp, E. de Wit, and B. D. Rowland, Cell **169**, 693 (2017).
[36] 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, and E. L. Aiden, Cell **171**, 305 (2017).
[37] A. M. Chiariello, S. Bianco, C. Annunziatella, A. Esposito, and M. Nicodemi, EPL (Europhysics Letters) **120**, 40004 (2018).
[38] G. Parisi and U. Frish, Proc. Intern. School of Physics' Enrico Fermi', 1983, Varenna, Italy (1985).
[39] R. Benzi, G. Paladin, G. Parisi, and A. Vulpiani, Journal of Physics A: Mathematical and General **17**, 3521 (1984).
[40] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, Physical Review A **33**, 1141 (1986).
[41] B. Bhushan and A. Majumdar, Wear **153**, 53 (1992).
[42] R. Balian and R. Schaeffer, Astronomy and Astrophysics **226**, 373 (1989).
[43] K. P. Iyer, K. R. Sreenivasan, and P. Yeung, Physical Review X **9**, 041006 (2019).
[44] L. Giorgetti, B. R. Lajoie, A. C. Carter, M. Attia, Y. Zhan, J. Xu, C. J. Chen, N. Kaplan, H. Y. Chang, E. Heard, *et al.*, Nature **535**, 575 (2016).
[45] `https://github.com/spigolotti/multifractal_hic`, accessed: 2020-09-07.
[46] T. Bohr and M. H. Jensen, Physical Review A **36**, 4904 (1987).
[47] W. Schwarzer, N. Abdennur, A. Goloborodko, A. Pekowska, G. Fudenberg, Y. Loe-Mie, N. A. Fonseca, W. Huber, C. H. Haering, L. Mirny, *et al.*, Nature **551**, 51 (2017).
[48] J. R. Dixon, I. Jung, S. Selvaraj, Y. Shen, J. E. Antosiewicz-Bourget, A. Y. Lee, Z. Ye, A. Kim, N. Rajagopal, W. Xie, *et al.*, Nature **518**, 331 (2015).
[49] T. Nagano, C. Várnai, S. Schoenfelder, B.-M. Javierre, S. W. Wingett, and P. Fraser, Genome biology **16**, 175 (2015).
[50] R. Kalhor, H. Tjong, N. Jayathilaka, F. Alber, and L. Chen, Nature biotechnology **30**, 90 (2012).
[51] N. Servant, N. Varoquaux, B. R. Lajoie, E. Viara, C.-J. Chen, J.-P. Vert, E. Heard, J. Dekker, and E. Barillot, Genome biology **16**, 259 (2015).
[52] M. Imakaev, G. Fudenberg, R. P. McCord, N. Naumova, A. Goloborodko, B. R. Lajoie, J. Dekker, and L. A. Mirny, Nature methods **9**, 999 (2012).