Supplementary Materials for

A Transferable Model For Chromosome Architecture

Michele Di Pierro[a,1,2], Bin Zhang[a,1], Erez Lieberman Aiden[a,b], Peter G. Wolynes[a,c,2], José N. Onuchic[a,c]

Affiliations:

[a]Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005

[b] The Center for Genome Architecture, Baylor College of Medicine, Houston, Texas 77030

[c] Departments of Physics & Astronomy, Chemistry, and Biosciences, Rice University, Houston, Texas 77005

[1] These authors contributed equally.

[2] Correspondence to: Michele.DiPierro@rice.edu, pwolynes@rice.edu

This file includes:

Supplementary Text

In this section we show how the Maximum Entropy Principle can be used to derive an effective potential energy function for the chromosome. To do so we start from the potential associated with a generic polymer and we introduce information specific to the chromosome as it emerges from experimental evidence present in the literature.

The Maximum Entropy Effective Energy

Let us start our derivation from the potential describing a homopolymer $U_{HP}(\vec{r})$; this potential includes soft-core interactions, bonds and angles energies. For the homo-polymer potential and for a generic phase space observable function $\varphi(\vec{r})$, a simulation in the canonical ensemble yields the following value:

$$\varphi_{HP} = \left\langle \varphi(\vec{r}) \right\rangle_{U_{HP}} = \frac{\int \varphi(\vec{r}) e^{-\beta U_{HP}(\vec{r})} d\vec{r}}{\int e^{-\beta U_{HP}(\vec{r})} d\vec{r}} = \int \varphi(\vec{r}) \pi^{HP}(\vec{r}) d\vec{r}$$

where $\vec{r}$ is the vector characterizing the positions in Cartesian space of all the loci in the chromosome, $\pi^{HP}(\vec{r})$ is the probability density for the homo-polymer model and $\beta = 1/K_B T$.

Following the Maximum Entropy approach introduced by Jaynes (1), we consider the probability density $\pi^{\mathrm{ME}}(\vec{r})$ that reproduces the experimental values of a certain set of n observables $\varphi_i(\vec{r})$. Such a probability density is defined by the following constraint equations:

$$c_0 = \int \pi^{\mathrm{ME}}(\vec{r}) d\vec{r} - 1$$

$$c_1 = \int U^{HP}(\vec{r}) \pi^{\mathrm{ME}}(\vec{r}) d\vec{r} - \frac{3}{2} N k_b T$$

$$c_i^{DATA} = \int \varphi_i(\vec{r}) \pi^{\mathrm{ME}}(\vec{r}) d\vec{r} - \varphi_i^{\exp} \qquad i = 1,...,n$$

Each of the constraint equations must be equal to zero; the first equation ensures that $\pi^{\mathrm{ME}}(\vec{r})$ is normalized, the second fixes the average potential energy to be equal to the thermal energy $\frac{3}{2} N k_b T$ and the last n equations ensure that the expectation values and the experimental values coincide for each one of the considered observables.

To determine the probability density $\pi^{ME}(\vec{r})$ we maximize the information entropy

$$S = -\int \pi^{ME}(\vec{r}) \ln\left(\pi^{ME}(\vec{r})\right) d\vec{r}$$

subject to the constraints. This approach is based on the fact that a constrained maximization of the entropy is equivalent to minimizing the amount of additional information built into the distribution other than the one contained in the constraints themselves.

Using Lagrange multipliers we obtain the following extremum condition

$$\frac{\partial S}{\partial \pi^{ME}} - \lambda_0 \frac{\partial c_0}{\partial \pi^{ME}} - \lambda_1 \frac{\partial c_1}{\partial \pi^{ME}} - \sum_{i=1}^{n} \lambda_i^D \frac{\partial c_i^{DATA}}{\partial \pi^{ME}} = 0$$

which leads to the probability distribution

$$\pi^{ME}(\vec{r}) = \frac{e^{-\lambda_1 U_{HP}(\vec{r}) - \sum_{i=1}^{n} \lambda_i^D \varphi_i(\vec{r})}}{\int d\vec{r}\; e^{-\lambda_1 U_{HP}(\vec{r}) - \sum_{i=1}^{n} \lambda_i^D \varphi_i(\vec{r})}}$$

Recognizing that λ_1 coincides with β, we can think at the maximum entropy probability distribution $\pi^{ME}(\vec{r})$ as the distribution sampled from the maximum entropy potential energy

$$U_{ME}(\vec{r}) = U_{HP}(\vec{r}) + \frac{1}{\beta} \sum_{i=1}^{n} \lambda_i^D \varphi_i(\vec{r})$$

Derivation of the MiChroM Effective Energy

For the interpretation of Hi-C experiments and the subsequent definition of the constraints the probability of crosslinking is the key phase space observable. Little is known about the functional relationship between crosslinking probability and geometric distance. Such relationship depends on multiple factors related to the experimental set-up and the reagents and everything is further complicated by the coarse grained nature of our model. However, it is safe to assume that the probability of a crosslinking event to happen between two loci i and j is a decreasing function of their geometric distance r_{ij}.

We postulate that the probability of crosslinking can be approximated by the scalar function

$$f(r_{ij}) = \frac{1}{2}\left(1 + \tanh\left[\mu\left(r_c - r_{ij}\right)\right]\right)$$

This functional form is a switch function; at short distance the probability of crosslinking is close to unity (sure event) while at long distance the probability of a crosslinking event is zero. Such a shape for the probability is consistent with the fact that crosslinking is only possible in case of physical proximity.

In a chromosome conformation capture experiment (Hi-C in the case of the experimental data used in this manuscript) the number of crosslinking event is counted over a set of many cells. The sampling of the crosslinking probability is therefore averaged over a set of different chromosome conformations to produce the measured probabilities:

$$p_{ij} = \left\langle f\left(r_{ij}\right)\right\rangle = \frac{\int f\left(r_{ij}\right) e^{-\beta U(\vec{r})}\, d\vec{r}}{\int e^{-\beta U(\vec{r})}\, d\vec{r}}$$

where we have used the canonical ensemble and where $\vec{r}$ is the vector characterizing the positions in Cartesian space of all the loci in the chromosome, $U\left(\vec{r}\right)$ is the potential energy of the system and $\beta = 1/K_B T$.

We now chose a set of constraints that will lead us to the MiChroM information theoretic energy function. We make three physical assumptions about of the process of chromatin folding. First, chromatin can be classified into a few different chromatin types; each chromatin type possesses specific interaction patterns with the other types and is characterized by specific biochemical properties. In our model we will assume that when two segments of chromatin come into contact, the effective free energy change due to this contact depends exclusively on the chromatin type identity of each segment. Second, loop formation happens between specific loci and it is related to the activity of the protein CTCF; every time the pair of loci constituting the two ends of a loop comes into contact, there is a specific gain in the effective free energy. The third assumption is relative to the ideal chromosome model: every time a pair of loci comes into contact there is a gain/loss of $\gamma\left(d\right)$ effective free energy that depends only on the genomic distance d.

These three physical assumptions are related to three phase-space observables define using the probability of cross-linking. We then constrain the expectation value of these observables to their experimentally determined values extracted from the contact probabilities measured through chromosome conformation capture. In this way we introduce three classes of constraints that recapitulate the assumptions made about of the process of chromatin folding.

For a given configuration $\vec{r}$ of the chromosome, the average number of crosslinking events happening between two chromatin types k and l is:

$$T_{kl}\left(\vec{r}\right) = \sum_{\substack{i\in\{\text{Loci of Type } k\} \\ j\in\{\text{Loci of Type } l\}}} f\left(r_{ij}\right)$$

The expectation value of T_{kl} is a proxy for the energy of the contact of type kl. The number of crosslinking events of type kl is proportional to the total number of contacts of such kind; the average number of contacts is in turn proportional to the energy of that kind of contact. For this reason constraining T_{kl} to its experimental value is similar to constraining the effective contact energy to the correct value.

Similarly, the total number of crosslinking events happening between loci that are known to form a loop is:

$$L(\vec{r}) = \sum_{(i,j)\in \{\text{Loops Sites}\}} f(r_{ij})$$

The third class of observables we wish to constrain is the number of contacts that happen at a certain fixed genomic distance[1] d:

$$G_d(\vec{r}) = \sum_i f(r_{i,\,i+d})$$

The set of constraints generated by all the observables above is the following:

$$\begin{aligned}
c_0 &= \int \pi^{\text{MiChroM}}(\vec{r})\,d\vec{r} - 1 \\
c_1 &= \int U_{HP}(\vec{r})\,\pi^{\text{MiChroM}}(\vec{r})\,d\vec{r} - \frac{3}{2}Nk_bT \\
c_T^{kl} &= \int T_{kl}(\vec{r})\,\pi^{\text{MiChroM}}(\vec{r})\,d\vec{r} - T_{kl}^{\text{exp}} \qquad \{\forall k,l \in \text{ Set of Types: } l \geq k\} \\
c_L &= \int L(\vec{r})\,\pi^{\text{MiChroM}}(\vec{r})\,d\vec{r} - L^{\text{exp}} \\
c_G^d &= \int G_d(\vec{r})\,\pi^{\text{MiChroM}}(\vec{r})\,d\vec{r} - G_d^{\text{exp}} \qquad 3 \leq d \leq d_{cutoff}
\end{aligned}$$

These constraints define the MiChroM probability distribution and the information theoretic energy function is accordingly:

$$U_{MiChroM}(\vec{r}) = U_{HP}(\vec{r}) + \sum_{\substack{k\geq l \\ k,l\,\in \text{ Types}}} \alpha_{kl} \sum_{\substack{i\in\{\text{Loci of Type } k\} \\ j\in\{\text{Loci of Type } l\}}} f(r_{ij}) + \chi \sum_{(i,j)\in \{\text{Loops Sites}\}} f(r_{ij}) + \sum_{d=3}^{d_{cutoff}} \gamma_d \sum_i f(r_{i,\,i+d})$$

The MiChroM potential energy is therefore composed of 4 terms; homo-polymer potential, type-to-type interactions, loops interactions and ideal chromosome term.

Crosslinking can only act at short range, consequently all the terms contained in the effective energy only act at short range. This very appealing feature of the MiChroM energy function is consistent with the idea that differential binding of proteins on chromatin is the driving mechanism behind chromosomes organization.

It is also worth to notice that the first three terms in the energy function only depend on Cartesian distance while the last term is the only one depending on both Cartesian and genomic distance.

The Lagrange multipliers α's, χ and γ's remain to be determined and the procedure to do so is discussed in the next section.

Optimization Procedure

[1] If we label loci of size Δd in a chromosome by using an increasing integer index i then the distance between a locus i and a locus j is represented by $d = j - i$ in units Δd.

For the generic observable $\varphi_i(\vec{r})$ we wish to determine the parameters λ_i contained in the potential energy

$$U_{ME}(\vec{r},\vec{\lambda}) = U_{HP}(\vec{r}) + \sum_{i=1}^{n} \lambda_i \varphi_i(\vec{r})$$

so that the expectation values of the observables coincide with its experimental values:

$$\langle \varphi_i \rangle_{\vec{\lambda}} = \frac{1}{Z(\vec{\lambda})} \int e^{-\beta U_{ME}(\vec{r},\vec{\lambda})} \varphi_i(\vec{r}) d\vec{r} = \varphi_i^{\exp} .$$

where we have used the partition function

$$Z(\vec{\lambda}) = \int e^{-\beta U_{ME}(\vec{r},\vec{\lambda})} d\vec{r} .$$

Following a standard approach (2, 3) we define the convex objective function θ

$$\theta(\vec{\lambda}) = \ln Z(\vec{\lambda}) + \beta \sum_{i=1}^{n} \lambda_i \varphi_i^{\exp} .$$

The partial derivatives of the target function are

$$g_i(\vec{\lambda}) = \frac{\partial \theta}{\partial \lambda_i} = -\beta \langle \varphi_i \rangle_{\vec{\lambda}} + \beta \varphi_i^{\exp}$$

If the constraint equations $\langle \varphi_i \rangle_{\vec{\lambda}^*} = \varphi_i^{\exp}$ are all satisfied for a certain parameter vector $\vec{\lambda}^*$ then the target function has a stationary point for that value. However, we must note that there is no guarantee for the existence of such parameter vector. The Hessian matrix

$$B_{ij}(\vec{\lambda}) = \frac{\partial^2 \theta}{\partial \lambda_i \partial \lambda_j} = \beta^2 \left(\langle \varphi_i \varphi_j \rangle_{\vec{\lambda}} - \langle \varphi_i \rangle_{\vec{\lambda}} \langle \varphi_j \rangle_{\vec{\lambda}} \right)$$

is positive semi-definite by construction; if a stationary point exists then it is a global minimum.. To find the optimal set of parameters $\vec{\lambda}^*$ we numerically minimize the target function θ using Newton's method; i.e. we use iterative scheme

$$\lambda^{l+1} = \lambda^l - \gamma B^{-1}(\lambda^l) g(\lambda^l)$$

where l is the iteration number and $\gamma \in (0,1)$ is dampening parameter included to increase stability. In practice, for the optimization of MiChroM γ was set to several values always comprised between 0.2 and 0.66.

Methods

Chromosome Model and Computer Simulations

The MiChroM energy function is the following:

$$U_{MiChroM}(\vec{r}) = U_{HP}(\vec{r}) + \sum_{\substack{k \geq l \\ k,l \in \text{Types}}} \alpha_{kl} \sum_{\substack{i \in \{\text{Loci of Type } k\} \\ j \in \{\text{Loci of Type } l\}}} f(r_{ij}) + \chi \cdot \sum_{(i,j) \in \{\text{Loops Sites}\}} f(r_{ij}) + \sum_{d=3}^{500} \gamma(d) \sum_{i} f(r_{i,\,i+d})$$

Homopolymer Model

The homo-polymer potential models a generic bead-spring polymer in which each bead represents a genomic segment of 50 Kb in sequence.
The potential energy $U_{HP}(\vec{r})$ describes a spatially confined self-avoiding polymer (no non bonding attractive term is included here) and serves as a support for the features added by using the maximum entropy principle.
The potential $U_{HP}(\vec{r})$ consists of the following five terms, U_{FENE}, U_{Angle}, U_{hc}, U_{sc} and U_{c} (4).

$$U_{HP}(\vec{r}) = \sum_{i \in \{\text{Loci}\}} U_{FENE}(r_{i,i+1}) + \sum_{i \in \{\text{Loci}\}} U_{hc}(r_{i,i+1}) + \sum_{i \in \{Angles\}} U_{Angle}(\theta_i) + \sum_{\substack{i,j \in \{\text{Loci}\} \\ j > i+2}} U_{sc}(r_{i,j}) + \sum_{i \in \{\text{Loci}\}} U_{c}(\vec{r}_i)$$

U_{FENE} (Finite Extensible Nonlinear Elastic potential) is the bonding potential applied between two consecutive monomers, and it has the following form:

$$U_{FENE}(r_{i,j}) = \begin{cases} -\frac{1}{2} k_b R_0^2 \ln\left[1 - \left(\frac{r_{i,j}}{R_0}\right)\right] & \text{for } r_{i,j} \leq R_0 \\ 0 & \text{for } r_{i,j} > R_0 \end{cases}$$

Additionally the hard-core repulsive potential

$$U_{hc}\left(r_{i,j}\right)=\begin{cases} 4\varepsilon\left[\left(\dfrac{\sigma}{r_{i,j}}\right)^{12}-\left(\dfrac{\sigma}{r_{i,j}}\right)^{6}+\dfrac{1}{4}\right] & \text{for } r_{i,j}\le\sigma 2^{\frac{1}{6}} \\ 0 & \text{for } r_{i,j}>\sigma 2^{\frac{1}{6}} \end{cases}$$

is added between bonded monomers to avoid overlap.
A three-body potential is applied to all connected three consecutive monomers in the following form

$$U_{Angle}\left(\theta_i\right)=k_a\left[1-\cos\left(\theta_i-\theta_0\right)\right]$$

where θ_i is the angle defined by the two vectors $\vec{r}_{i,i+1}$ and $\vec{r}_{i,i-1}$.

Finally, all the non-bonded pairs experience a soft-core repulsive interaction via the potential (5)

$$U_{sc}\left(r_{i,j}\right)=\begin{cases} \dfrac{1}{2}E_{cut}\left[1+\tanh\left(\dfrac{2U_{LJ}\left(r_{i,j}\right)}{E_{cut}}-1\right)\right] & r_{i,j}<r_0 \\ U_{LJ}\left(r_{i,j}\right) & r_0\le r_{i,j}\le\sigma 2^{\frac{1}{6}} \\ 0 & r_{i,j}>\sigma 2^{\frac{1}{6}} \end{cases}$$

The above expression corresponds to the Lennard-Jones potential $U_{LJ}\left(r_{i,j}\right)=4\varepsilon\left[\left(\dfrac{\sigma}{r_{i,j}}\right)^{12}-\left(\dfrac{\sigma}{r_{i,j}}\right)^{6}+\dfrac{1}{4}\right]$ capped off at a finite distance, thus allowing for chain crossing at finite energetic cost. r_0 is chosen as the distance at which $U_{LJ}\left(r_{i,j}\right)=\dfrac{1}{2}E_{cut}$.

The confinement potential U_c describes the interaction between the chromosome and a spherical wall, whose size is chosen to enforce a volume fraction of 0.1; which is consistent with the experimentally determined density of chromatin (6). The spherical wall is included to mimic the interaction with the nucleus envelope that chromosomes experience inside the cell. Each monomer i of the chromosome interacts with its nearest point on the wall $\vec{r}_{np}$ through the potential $U_{hc}\left(r_{i,np}\right)$.

Probability of Crosslinking

The function representing the probability of crosslinking

$$f\left(r_{ij}\right)=\frac{1}{2}\left(1+\tanh\left[\mu\left(r_c-r_{ij}\right)\right]\right)$$

is tuned accordingly to the model and experimentally determined contact probabilities. Specifically, the function is tuned to return 1 when two beads are in contact (two beads are in contact when the distance between their centers is equal to 1 in reduced units):

$$f\left(1\right)=1$$

As a second constraint, the function f is tuned to return the correct experimental probability for the next nearest neighbors. The maximum distance between next nearest neighbors is 2 and the function f is monotonically decreasing with distance; therefore the minimum of the experimental probabilities for next nearest neighbors must correspond to the configuration in which they are at maximum distance:

$$f\left(2\right)=\min\left\{P_{i,i+2}^{\exp}\right\}$$

The resulting parameters adjusted for the contact maps of GM12878 B-lymphoblastoid cells in dataset GSE63525 (7) are $\mu=3.22$ and $r_c=1.78$.

MiChroM Parameter Set

We consider 5 chromatin types A1, A2, B1, B2, B3 plus a non-specific type NA. Chromatin type B4 was detected in minimal amounts in chromosome 19 by Rao *et al* (7) but given the small amount of data available for the optimization we decided to treat B4 as B3 (See Figure S48).

The parameters α 's governing the type-to-type interactions are:

	A1	A2	B1	B2	B3	NA
A1	-0.268028	-0.274604	-0.262513	-0.258880	-0.266760	-0.225646
A2	-0.274604	-0.299261	-0.286952	-0.281154	-0.301320	-0.245080
B1	-0.262513	-0.286952	-0.342020	-0.321726	-0.336630	-0.209919
B2	-0.258880	-0.281154	-0.321726	-0.330443	-0.329350	-0.282536
B3	-0.266760	-0.301320	-0.336630	-0.329350	-0.341230	-0.349490
NA	-0.225646	-0.245080	-0.209919	-0.282536	-0.349490	-0.255994

The parameter χ governing the loop interactions is equal to -1.612990.

Ideal Chromosome Term

All parameters in MiChroM are found by using Newton's method; as explained below, for the Ideal Chromosome interaction term we did however use a two-step procedure to reduce the number of parameters.

During the calibration phase the local compaction term $\sum_{d=3}^{d_{cutoff}} \gamma_d \sum_i f\left(r_{i,\,i+d}\right)$ is added for the first 200 neighbors ($d_{cutoff} = 200$); the optimization described above converges to the set of parameters $\left\{\gamma_d, d = 3,...,200\right\}$. After optimization we further reduce the number of parameters in MiChroM by fitting the set $\left\{\gamma_d, d = 3,...,200\right\}$ with the function:

$$\gamma(d) = \frac{\gamma_1}{\log(d)} + \frac{\gamma_2}{d} + \frac{\gamma_3}{d^2}$$

The best fit was obtained with the parameters $\gamma_1 = -0.030$, $\gamma_2 = -0.351$, $\gamma_3 = -3.727$.

In the sampling phase the local compaction term is added for the first 500 neighbors and is calculated via the function $\gamma(d)$.

As shown in figure S50, the data set $\left\{\gamma_d, d = 3,...,200\right\}$ is quite noisy and it can be approximated with similar accuracy by several different functional forms that all return very similar values for the energies $\gamma(d)$ (See also figure S51).

It is worth stressing that the function $\gamma(d)$ represents the effective energy gain happening when a contact randomly forms between two loci at a genomic distance d. Since this energy is modulated by the switch function $f\left(r_{i,\,i+d}\right)$ there is no long distance attraction or repulsion due to this (or any other) term of the MiChroM energy function.

Simulation Details

The initial configurations for our simulations were prepared as described below. First, we condense the polymer from an extended configuration initialized as a straight line. To condense the polymer, we perform $2\text{x}10^4$ step MD simulation under the potential energy function

$$U_{Eq}(\vec{r}) = \sum_{i \in \{\text{Loci}\}} U_{FENE}\left(r_{i,i+1}\right) + \sum_{i \in \{\text{Loci}\}} U_{hc}\left(r_{i,i+1}\right) + \sum_{i \in \{Angles\}} U_{Angle}\left(\theta_i\right) + \sum_{\substack{i,j \in \{\text{Loci}\} \\ j>i+2}} U_{sc}\left(r_{i,j}\right) + \frac{1}{2} K_{Eq}\left(R_g - R_g^0\right)^2$$

which is the homopolymer potential with the presence of a harmonic bias on the radius of gyration R_g. We choose $K_{Eq} = 200\varepsilon/\sigma^2$ and $R_g^0 = 1$. The spherical wall mimicking the nucleus confinement is not present in this phase. Then, starting from the condensed polymer configuration obtained in the first step, we perform 20 million steps equilibration with the potential energy function $U_{HP}(\vec{r})$, which also includes the confinement potential. In each chromosome, the radius of the confinement potential was set to reproduce the experimentally determined density of chromatin (6) corresponding to a volume ratio of 0.1.
The optimization of the parameters set in order to reproduce the experimental map of chromosome 10 required approximately 20 iterations in which sampling was collected using MD simulation and the parameter set was updated following the procedure already described. Subsequently, the parameter set was used to sample the chromosome conformations for the remaining autosomal chromosomes.
All chromosome simulations were performed using the molecular dynamics package LAMMPS (8). Reduced units were used during the simulation, with

$$k_a = 2\varepsilon \qquad k_b = \frac{30\varepsilon}{\sigma^2} \qquad E_{cut} = 4\varepsilon \qquad \varepsilon = K_B T$$
$$R_0 = 1.5\sigma \qquad \sigma = 1 \qquad \theta_0 = \pi$$

Simulations were maintained at a constant temperature $T = 1.0$ via Langevin dynamics with a damping coefficient of 10.0τ, where τ is the time unit. A time step $\Delta t = 0.01\tau$ was used for all the simulations. All MD simulations (in the optimization runs as well as in the sampling runs) were run until convergence was achieved as tested by verifying that different replicas reported similar results. Given the different sizes of different chromosomes, this consisted in a total simulation length comprised between $1\text{x}10^8$ and $4\text{x}10^8$ steps for each chromosome. A short equilibration (10^6 time steps) at high temperature (T=10) was performed in each simulation before starting sampling chromosome conformations.

Symmetry Score

To measure the symmetry of a matrix M we first calculate its symmetric part, S, and its antisymmetric part, A. We then compute $\frac{||S||}{||S||+||A||}$, which returns a symmetry score of 1 for a perfectly symmetric matrix and a score of 0 for a perfectly anti-symmetric matrix.

The symmetry score can be computed using different matrix norms returning different results for partially symmetric matrices. We used both the 2-norm and the Frobenius norm.

Experimental data

Contact Probabilities, Loop Locations, and Chromatin Type Sequences

We used contact maps, chromatin types sequences and loops locations obtained from human GM12878 B-lymphoblastoid cells and reported by Rao *et al.* (7). The Gene Expression Omnibus (GEO) accession number for the data sets is GSE63525 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63525).

To obtain the experimental contact probabilities we started from the raw counts of detected ligations at 50 kb resolution and we balanced the matrix by enforcing it to be doubly stochastic (each of its rows and columns must sum to a constant); this was done using the Knight and Ruiz (KR) vectors reported by the authors. To extract a contact probability matrix $P^{\exp}$ from the stochastic matrix C we then divided each row i by its maximum entry, typically $C_{i,i+1}$. In this way we assign probability 1 (sure event) to the contact that was observed most often and consistently with our beads-on-a-string model, we obtain that neighboring beads are in contact.

Gene Expression

Gene expression data as measured by RNA-Seq (9) were downloaded from ENCODE (Encyclopedia of DNA Elements) website (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCaltechRnaSeq/).

The expression level for each bead in our model, which represents a genomic segment of 50 Kb in length, was determined as the arithmetic mean of expression over the entire segment. For the convenience of further analysis, we partitioned the genomic segments into ON and OFF genes using a threshold of 20 for the gene expression level.

Structural Characterization of the Chromosome

Clustering Algorithm

We developed the following algorithm to detect spatial clusters formed by chromatin types or active genes.
First, given the set of Cartesian coordinates for all genomic loci, we compute the Voronoi diagram that tessellates the three dimensional space occupied by a chromosome (10). This diagram provides each locus a corresponding Voronoi cell, the region enclosed by which consists of points closer to that locus than to any other loci. When computing the Voronoi diagram, a boundary enclosing a finite space with all the genomic loci inside needs to be specified. If the boundary is simply chosen as a cubic box or a sphere, artifacts can be introduced due to the irregular packing of the chromosome and the Voronoi cells for loci at the edges of the packing would be very large and extend off to the edges of the container. To remedy this issue and avoid creating very large Voronoi cells, we applied upper bounds for the volume of each Voronoi cell as a dodecahedron.
Second, we identify nearest neighbors of a given locus as the set of loci whose Voronoi cells share an interface with the cell of that locus. Because the two cells share an interface, there are no other loci in between them.

Third, by connecting nearest neighbors with edges, we can build a graph for a given chromatin type by including only loci of that type as the network nodes. From this graph, we use a depth-first search algorithm (11) to identify the largest connected component, which is a subgraph in which any two nodes are connected to each other by paths. Therefore, in this subgraph, one can go from a genomic locus to any other locus of the same chromatin type, by following nearest neighbors of the same type. We therefore use the size of this subgraph as a measure for the degree of clustering of a given chromatin type.

Radial Density Profile

The radial density profile is calculated as

$$\rho(r) = \frac{\langle N(r) \rangle}{4\pi r^2 \Delta r}$$

where r is the distance from a genomic locus to the centroid, $N(r)$ is the number of genomic loci found in the spherical shell r, $r + \Delta r$, and $\Delta r = R_0 / 200$ is the size of the numerical grid. R_0 is the size of the confinement to ensure a density of 0.1, and is chromosome dependent. $\langle\ \rangle$ indicates averaging over the ensemble of chromosome conformations. When calculating the density profile for a given chromatin type, only loci of that type is included in counting $N(r)$. To compare among different chromatin types, and loci with different gene expressions, we normalize the radial profiles by the number of loci of a give type N_{Type}.

Knot Invariants

We used the following knot invariants to characterize the topology of simulated chromosome conformations.

Minimum rope length. The idea behind this invariant is simple: as the complexity of a knot increases, a longer rope of fixed diameter will be needed to form that knot. Therefore, it is expected that there exists a minimal length required to form a given knot type from rope of a given diameter. This correlation between knot complexity and rope length motivates the proposition that the minimal length/diameter ratio is a knot invariant. The minimal rope length is the smallest contour length a polymer can have that would preserve the original topology while keeping its diameter fixed. This invariant has been successfully applied to other DNA topological problems. In a previous manuscript, we used this invariant to study the topology of interphase chromosomes for stem cells and mature cells (12). Following the same protocol, we calculate this knot invariant using the Shrink-On-No-Overlaps algorithm (13).

Alexander polynomial. This is a knot invariant that assigns a polynomial with integer coefficients to each knot type. Alexander polynomial is superior when compared to the minimum rope length in distinguishing non-trivial knot types. To calculate the alexander polynomial, we use algorithm explained in ref (14, 15).

To determine either one of the two knot invariants listed above, we first need to close the linear chromosome. We connect the two ends of each chromosome conformation with a

procedure used in studying protein knots (16), which we outline as follows. First, starting at one end of the polymer, we gradually add new beads along the vector pointing from the center of mass of the chromosome to the chromosome end. The extension procedure terminates when the distance from the last added bead to the chromosome center of mass exceeds the radius of the spherical confinement R by 4σ. These additional beads are placed with a spacing of σ. The same procedure is then repeated for the other end of the chromosome. Finally, the two newly extended ends are connected along the arc with radius R+4σ.

Reproducibility Of The Probability Peaks Generated By Looping Interactions In Simulated Maps

In order to assess the efficacy of the looping interactions in the MiChroM Hamiltonian in creating chromatin loops we study how many of the input loops sites do indeed manifest a local peak in the contact probability.
One location of the contact map represents a local peak if its contact probability is significantly increased with respect to its neighbors.
In order to measure contact enrichment we first calculate the normalized contact map $M_{i,j} = \frac{P^{Sim}_{i,j}}{S(j-i)}$, where P^{Sim} is the simulated contact probability matrix and $S(d)$ is the average contact probability (in simulations) as a function of the genomic distance d .
A reasonable estimator for the average contact probability of the neighborhood around position (i,j) is

$$E_{i,\mathrm{j}} = \frac{M_{i+1,j+1} + M_{i-1,j-1} + M_{i-1,j+1} + M_{i+1,j-1}}{4};$$

if $M_{i,j}$ is significantly larger than $E_{i,\mathrm{j}}$ then the position (i,j) is likely to be a local peak (7).
For the simulation of the 22 chromosomes we used as input the positions of 7827 loops; in simulated maps, 6772 of those sites (approximately 87%) show a contacts enrichment larger than 30%. In comparison, only 2% of the non looping positions within a distance of 2 Mb from each other show a similar contact enrichment.
The vast majority of the loops that are not reproduced by MiChroM are very short loops; in 89% of the missing loops the distance between the two anchors of the loop is less than 100 kb, indicating that the loop in our model consists of one or zero beads between the two anchors.
These results indicate that a peak in contact probability is found in simulated maps in almost all the situations in which the model has sufficient resolution to describe a meaningful loop and confirm that the looping interactions in MiChroM can effectively reproduce loop features of Hi-C maps genome-wide. (See also figure S49)

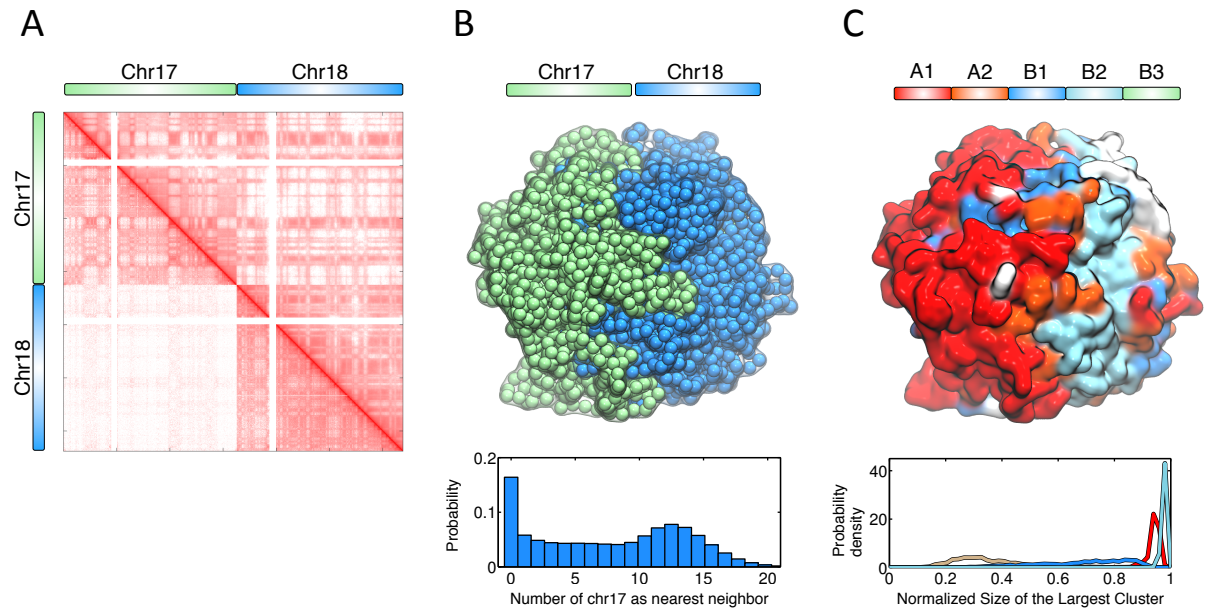

Fig. S1. Structural characterization of the simulated conformations for chromosome 17 and 18. A) Comparison between contact probabilities obtained from simulation (upper triangle) and experiment (lower triangle).
The intra-chromosomal contact maps are essentially the same as those simulated in isolation. Spatial confinement introduces artifacts in the frequency of inter-chromosomal contacts; as a result, the inter-chromosomal contact map from simulation shows increased probabilities with respect to Hi-C. Despite the biased intensity, the inter-chromosomal map reproduces the correct pattern of interactions. Overall Pearson's correlation between the two datasets is 0.96.
B) Chromosomes form territories. Chromosomes 17 and 18, colored in green and blue respectively, phase separate from each other. The bottom panel plots the probabilities for each locus, belonging to either chromosome 17 or 18, to find *N* genomic loci from chromosome 17 as nearest neighbors. The bimodal shape is consistent with the phase separation shown in the top.

C) Phase separation of chromatin types persists in the presence of multiple chromosomes. (*Top*) Surface plot for chromosome 17 and 18 combined. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated conformations for the two chromosomes.

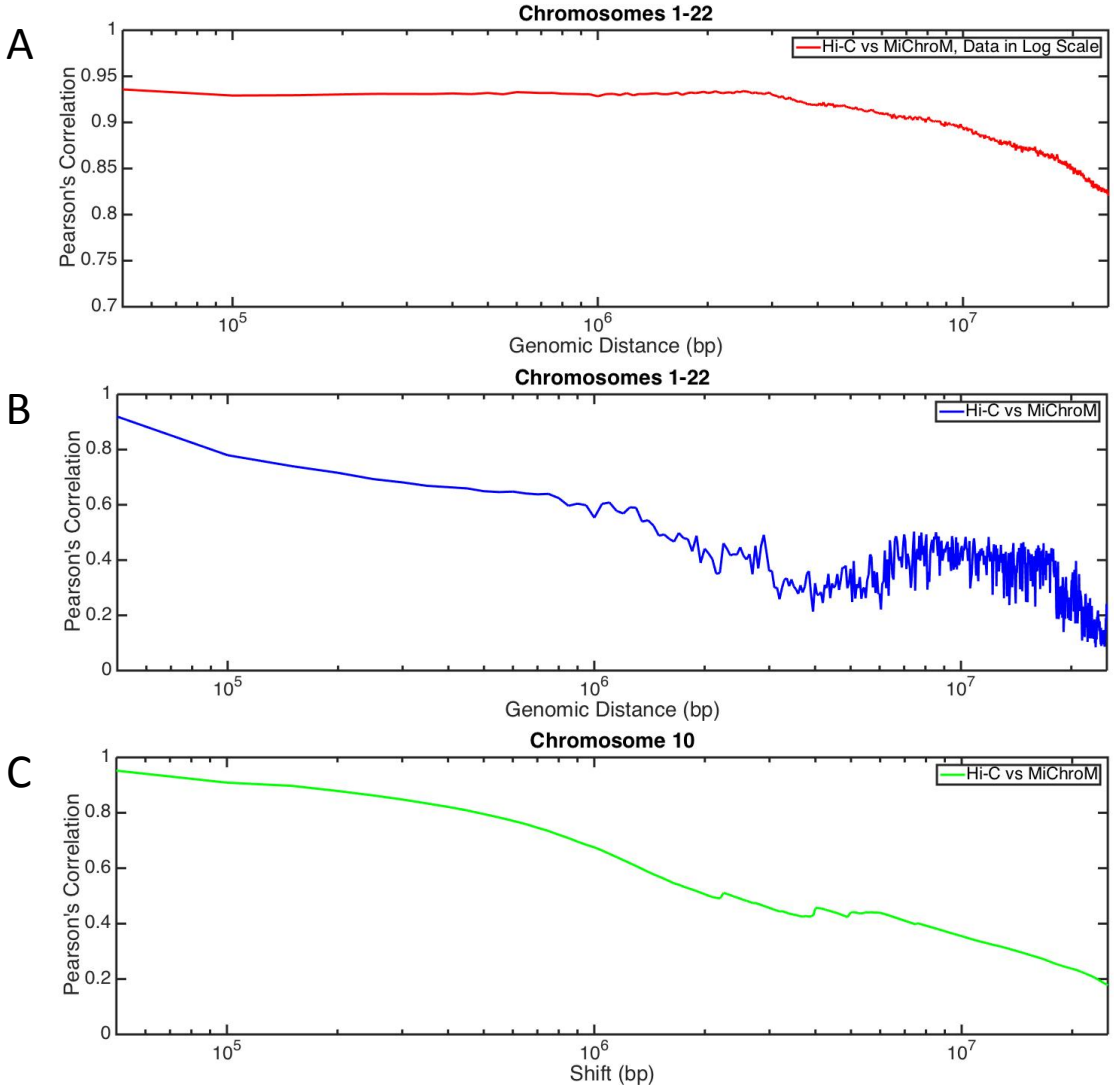

Fig. S2. The Pearson's correlation between experimental (7) and simulated data sets is shown as a function of genomic distance. The resolution is the maximum allowed by our current model, i.e. 50 kb. A) For all autosomal chromosomes we show the correlation between the experimental contact probabilities and those generated *de novo* by MiChroM. Data are represented in log scale. B) For all autosomal chromosomes we show the correlation between the experimental contact probabilities and those generated *de novo* by MiChroM. C) For chromosome 10 we show the correlation between the set of experimental probabilities $\left\{P_{i,j}^{\exp} : i \in [1,2712], j \in [i+d,2712]\right\}$ and the set of computational contact probabilities $\left\{P_{i,j}^{MiChroM} : i \in [1,2712], j \in [i+d,2712]\right\}$ as a function of the shift d. All panels show that

MiChroM generates contact maps that are highly correlated with those measured in Hi-C experiments and support the plausibility of the physical mechanism underlying MiChroM.

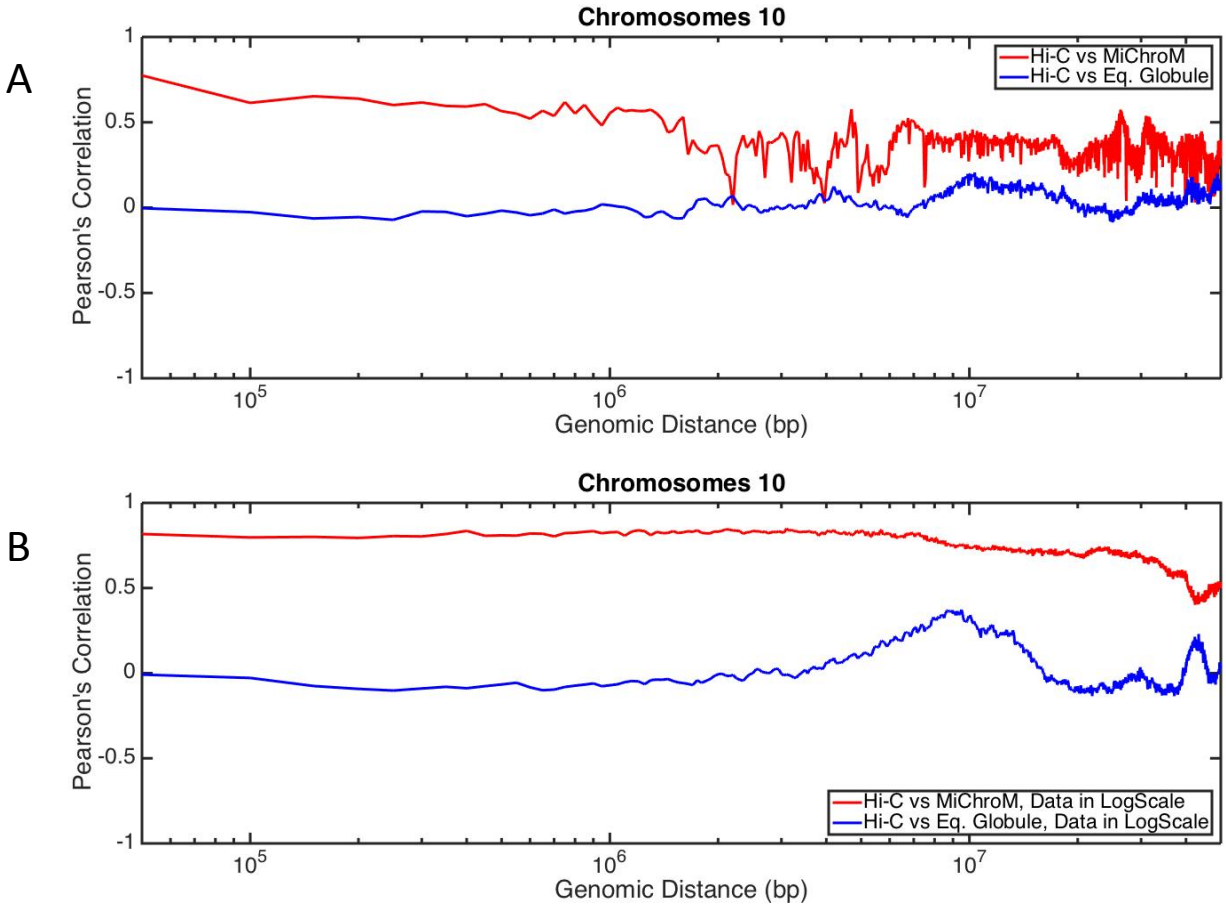

Fig. S3. For chromosome 10 we show a comparison between the equilibrium globule and the simulated chromosome. Resolution is 50 kb. A) Pearson's correlation between experimental (7) and simulated contact probabilities as a function of the genomic distance B) Pearson's correlation between experimental and simulated contact probabilities as a function of the genomic distance when the two data sets are represented in log scale. The comparison between the equilibrium ensemble generated by MiChroM and the one generated by the homopolymer model shows that our information theoretical approach was very successful in improving the quality of the model.

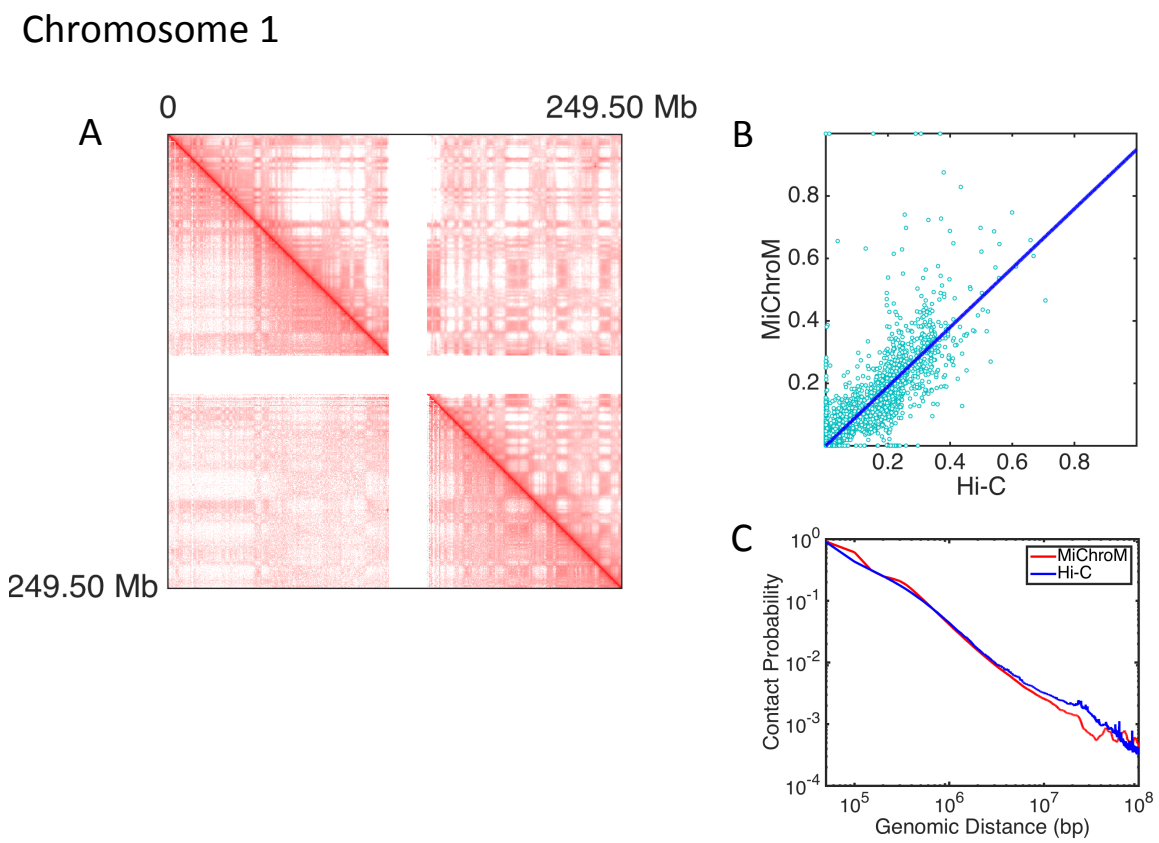

Fig. S4. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 1 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.955 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 2% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

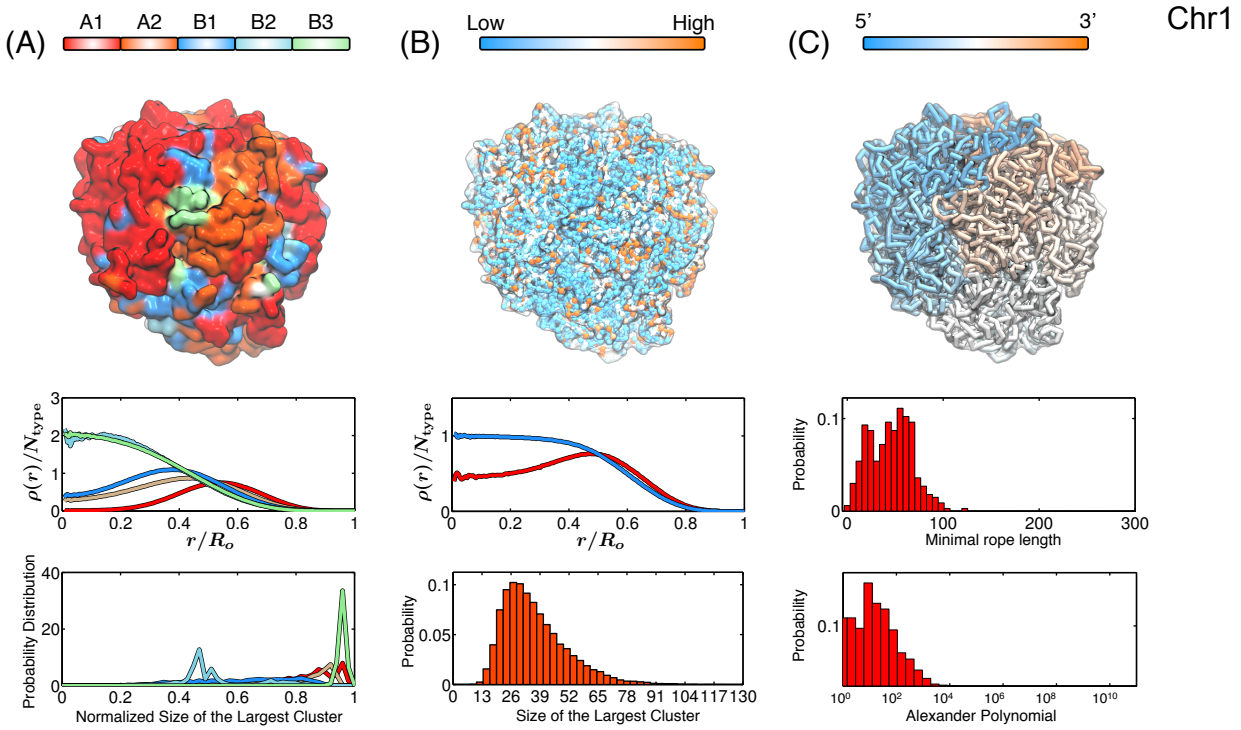

Fig. S5

Structural characterization of the simulated conformations for chromosome 1.
A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

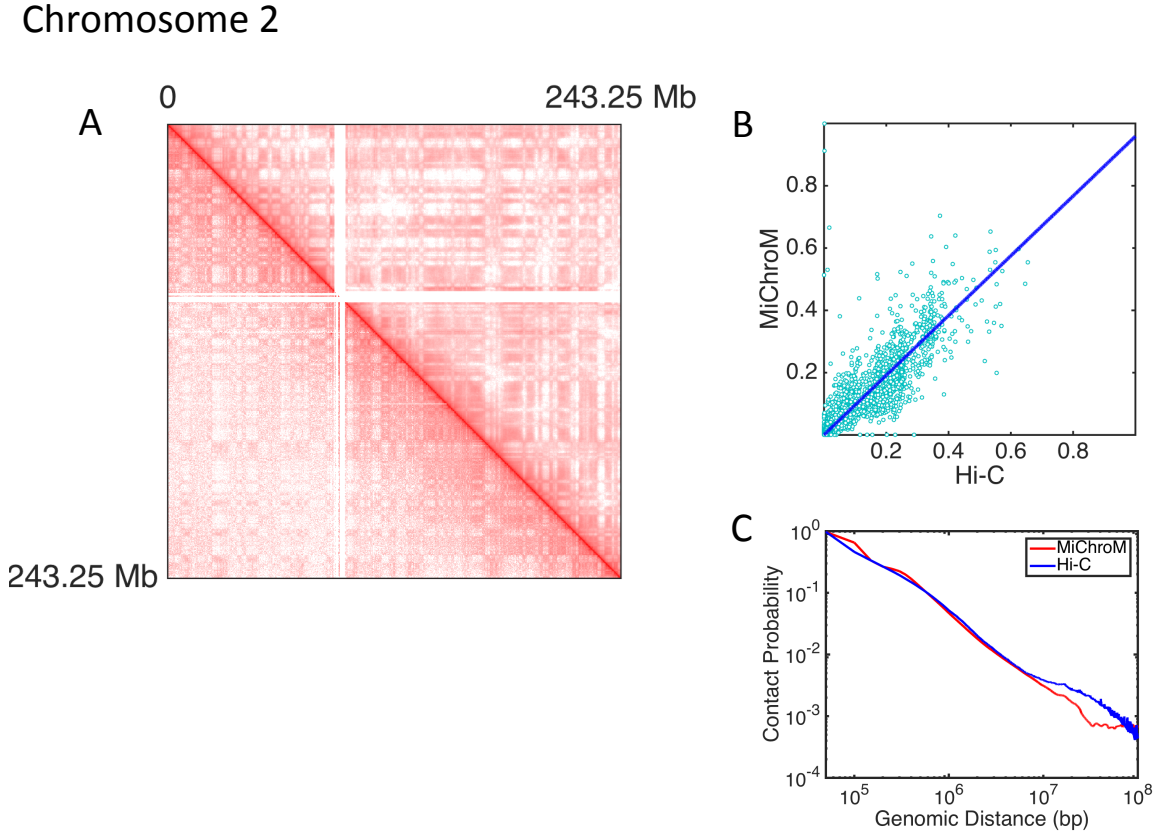

Fig. S6. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 2 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.965 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 2% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

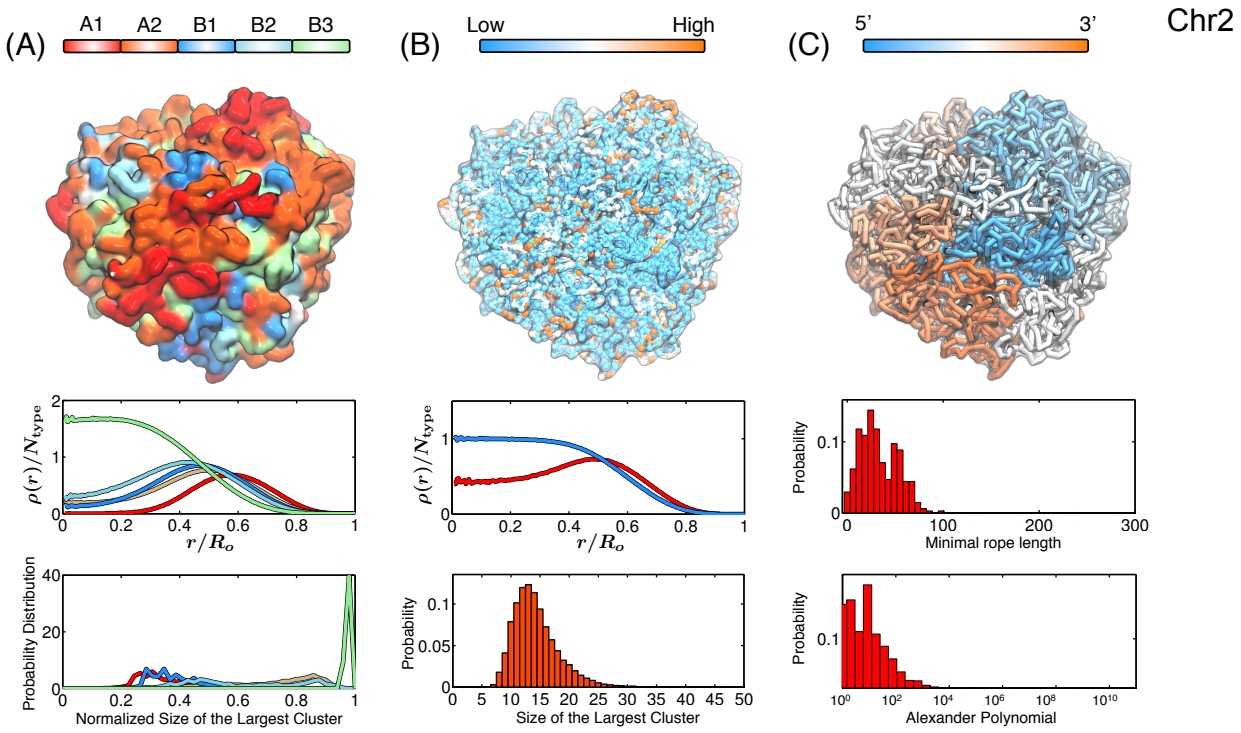

Fig. S7. Structural characterization of the simulated conformations for chromosome 2.
A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

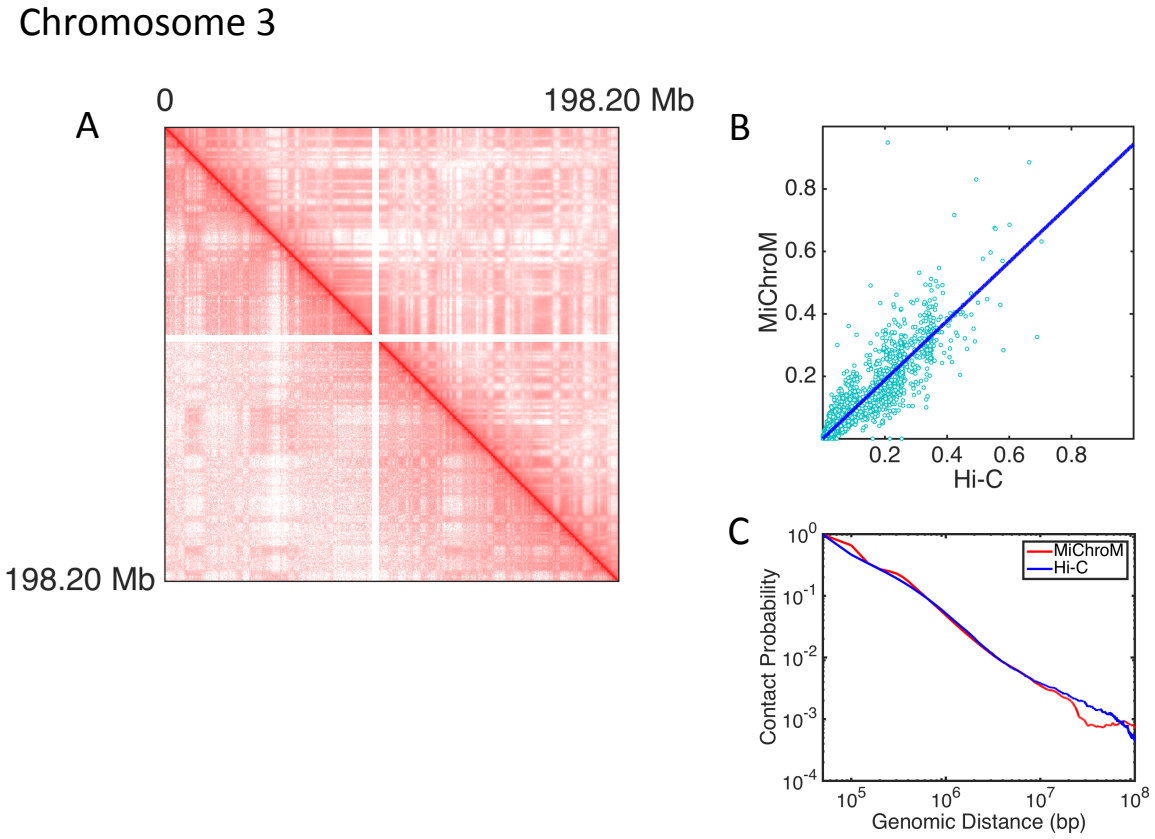

Fig. S8. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 3 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.976 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 2% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

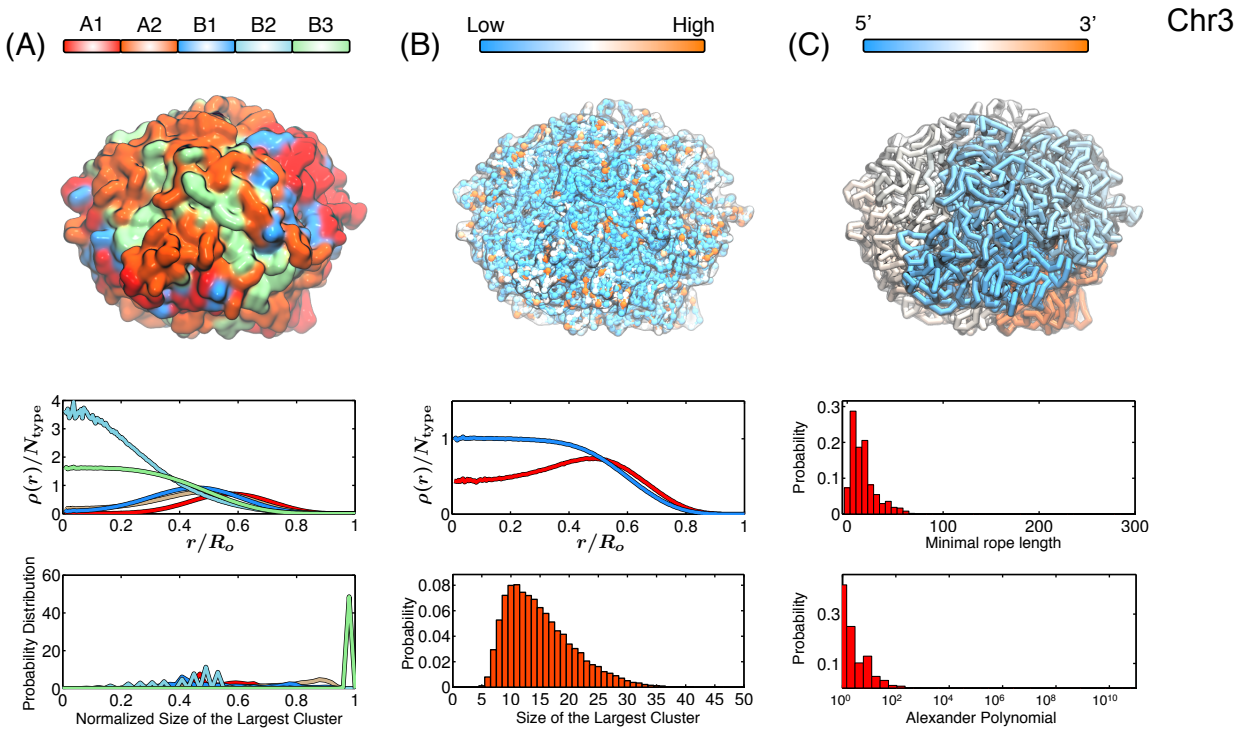

Fig. S9. Structural characterization of the simulated conformations for chromosome 3.
A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

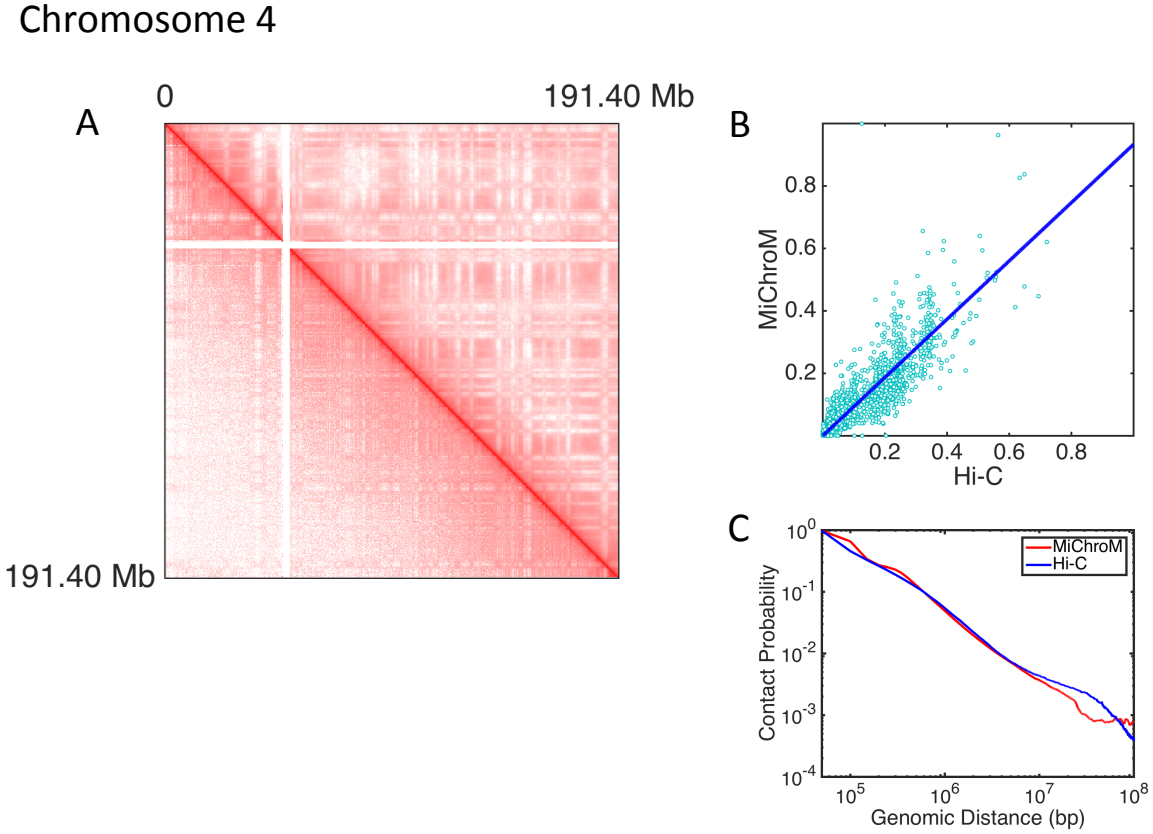

Fig. S10. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 4 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.977 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 2% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

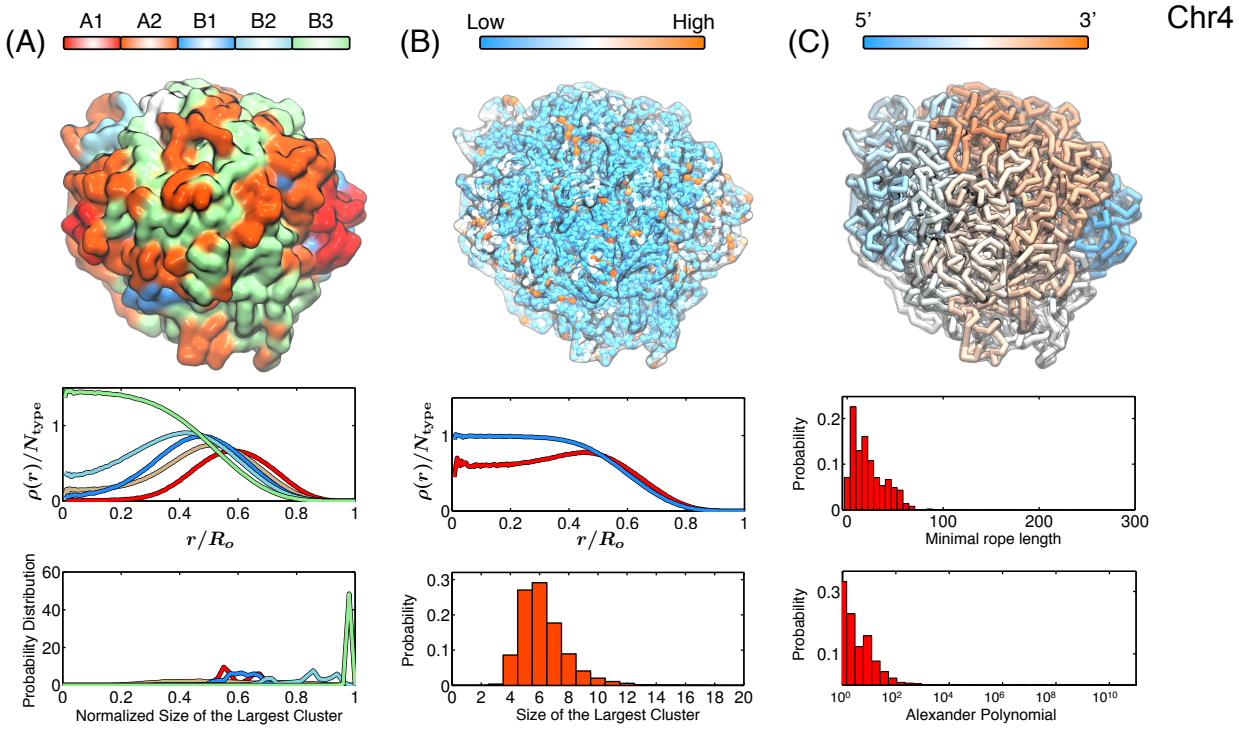

Fig. S11. Structural characterization of the simulated conformations for chromosome 4. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

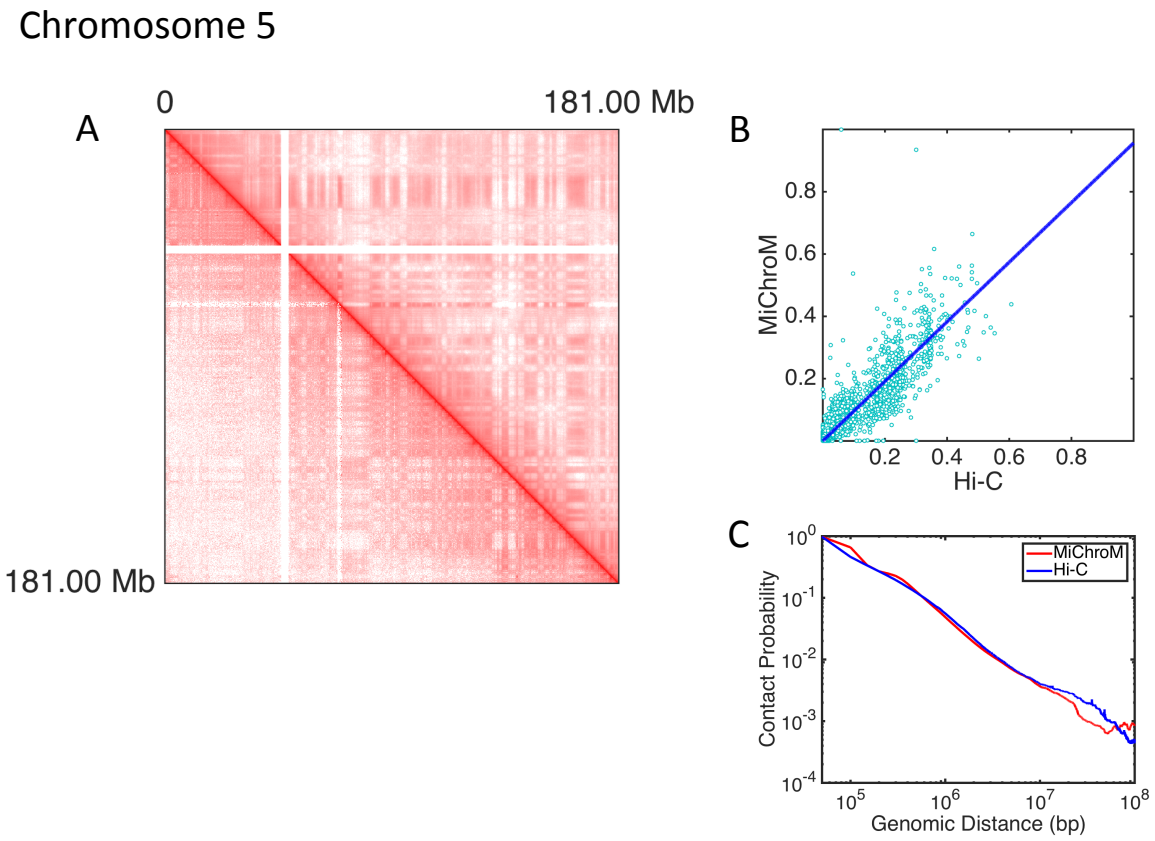

Fig. S12. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 5 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.969 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 2% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

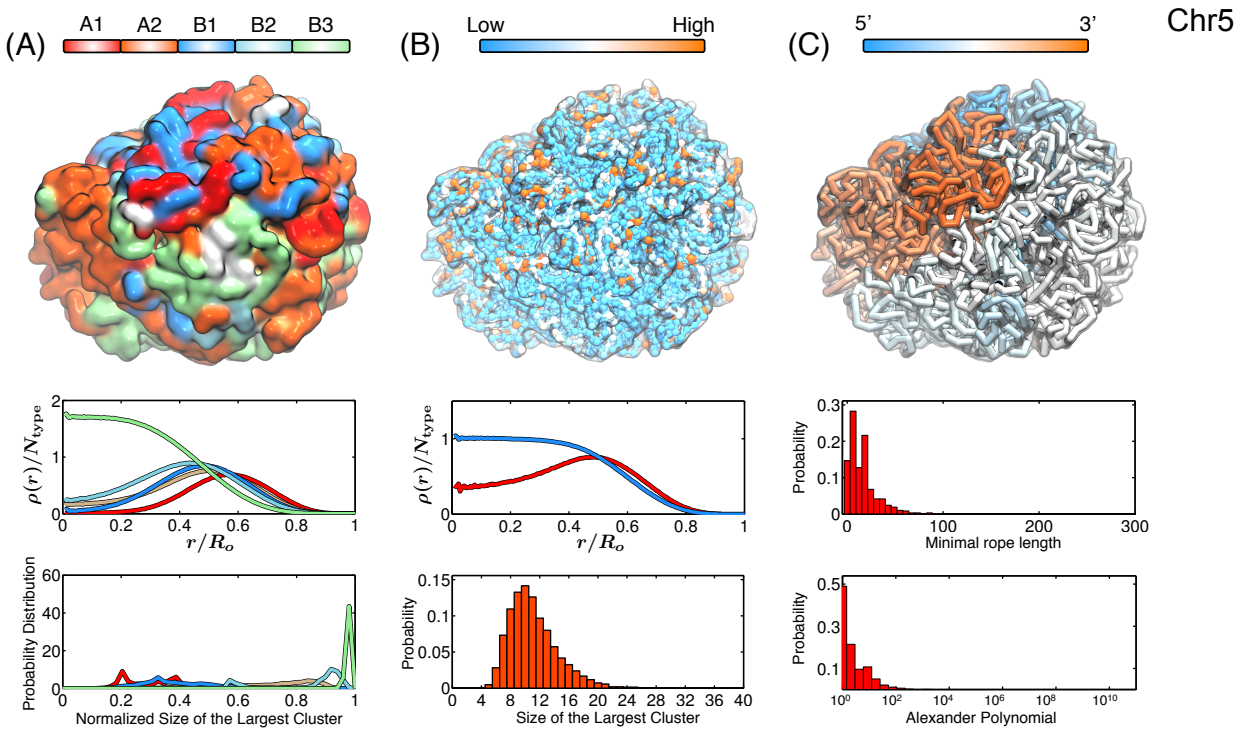

Fig. S13. Structural characterization of the simulated conformations for chromosome 5. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

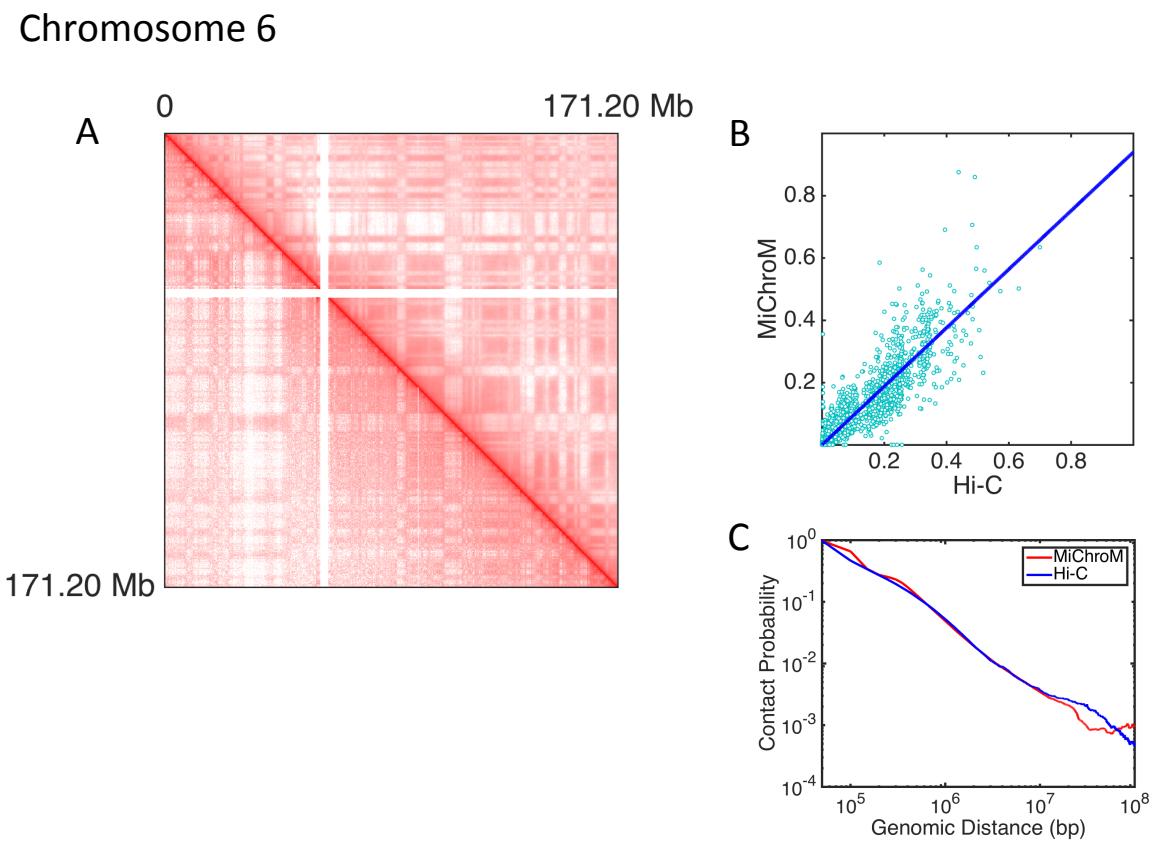

Fig. S14. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 6 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.971 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 2% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

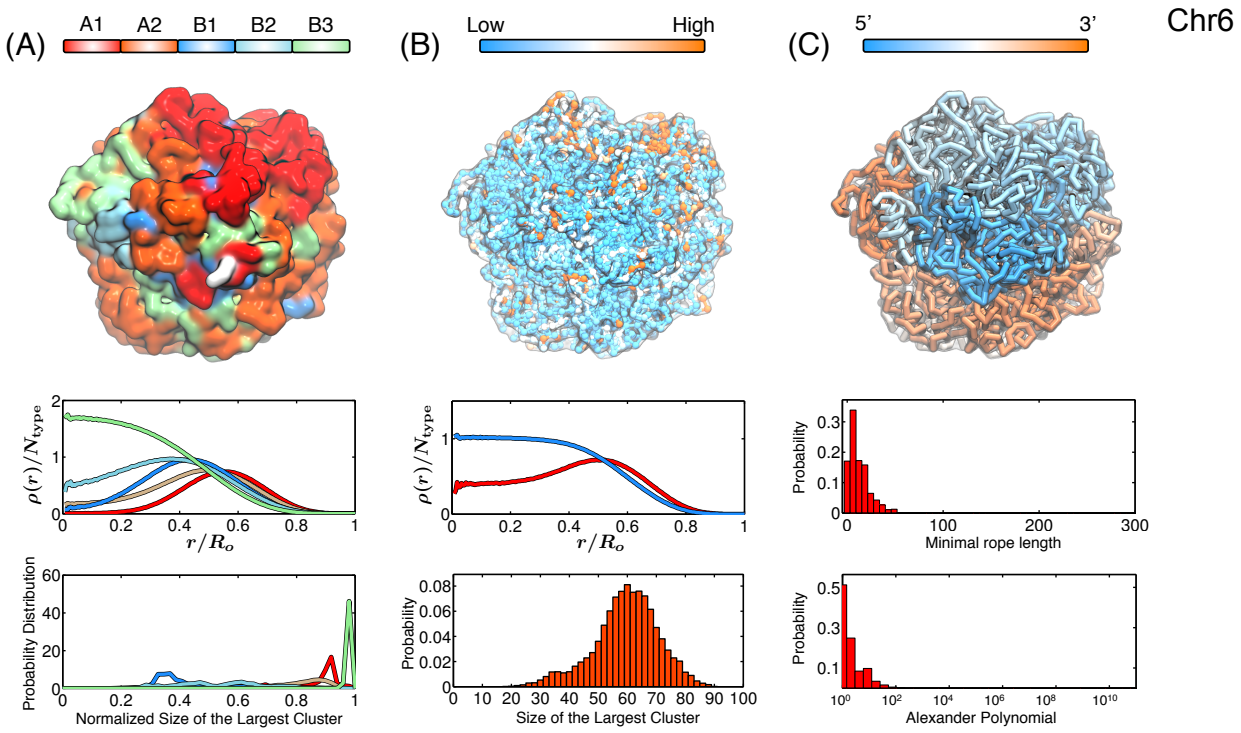

Fig. S15. Structural characterization of the simulated conformations for chromosome 6. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

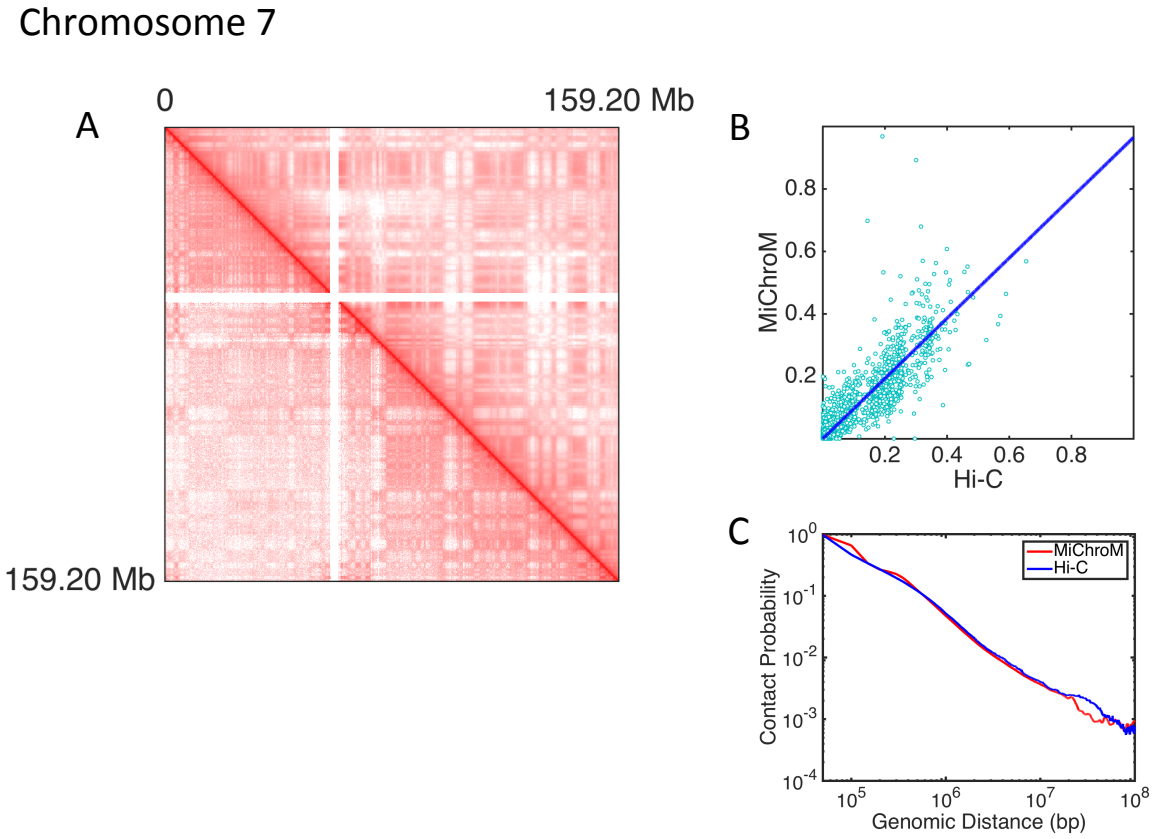

Fig. S16. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 7 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.966 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 2% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

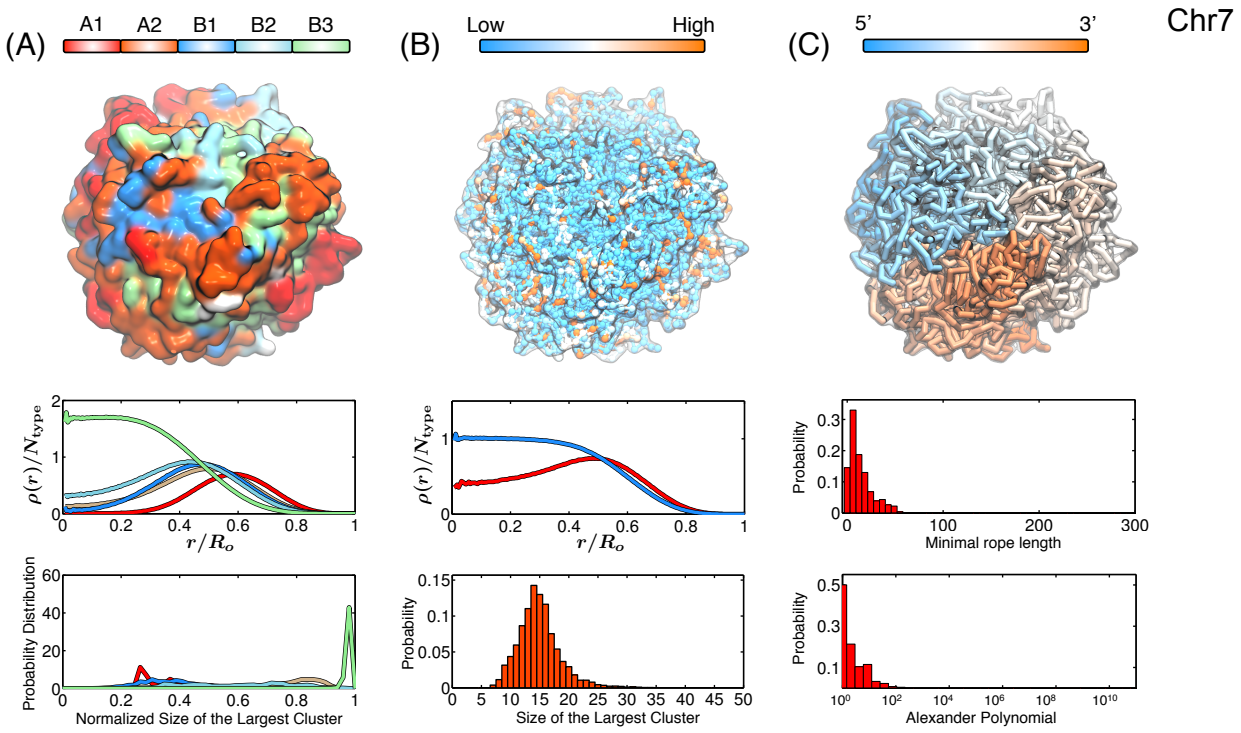

Fig. S17. Structural characterization of the simulated conformations for chromosome 7.
A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

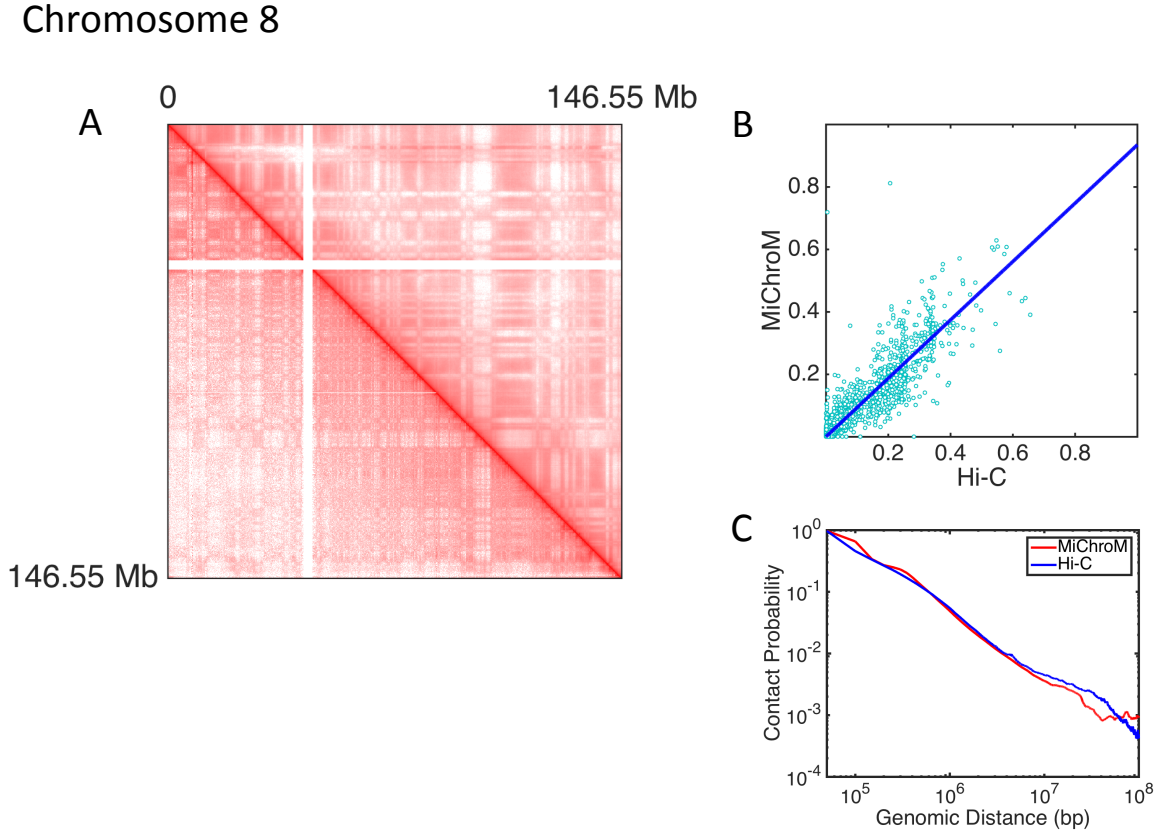

Fig. S18. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 8 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.968 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 2% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

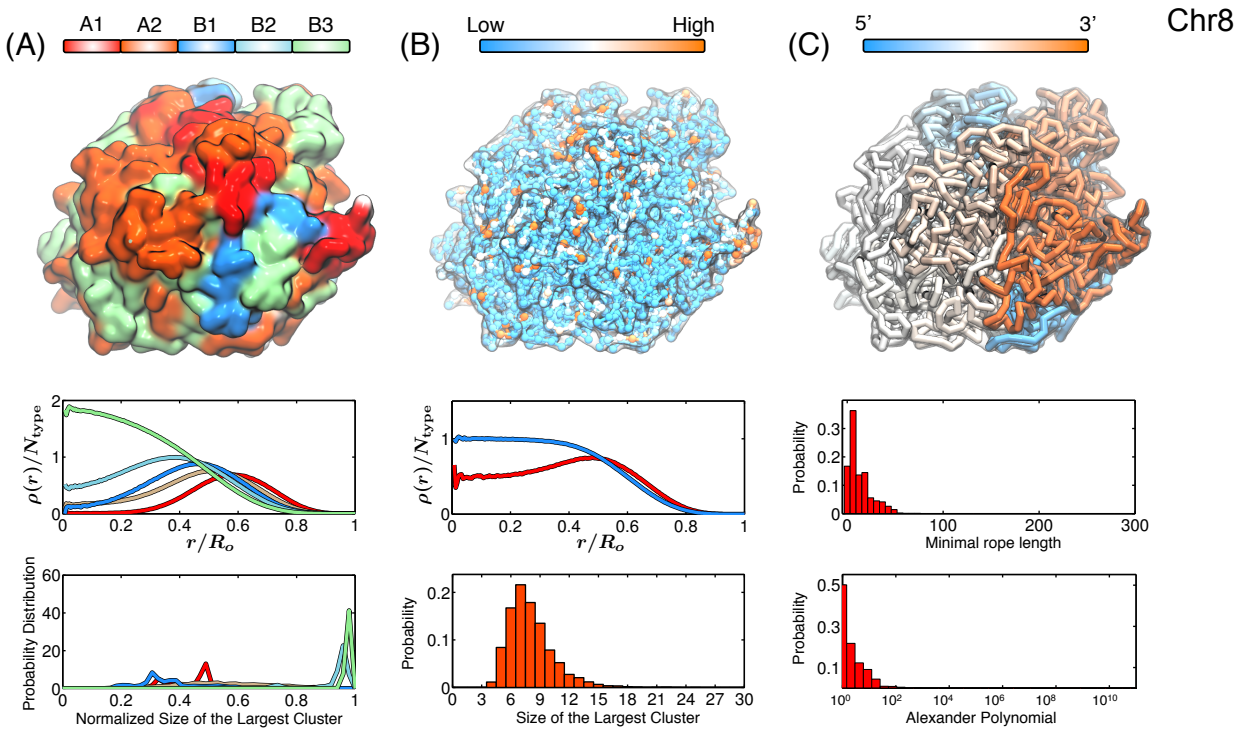

Fig. S19. Structural characterization of the simulated conformations for chromosome 8. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

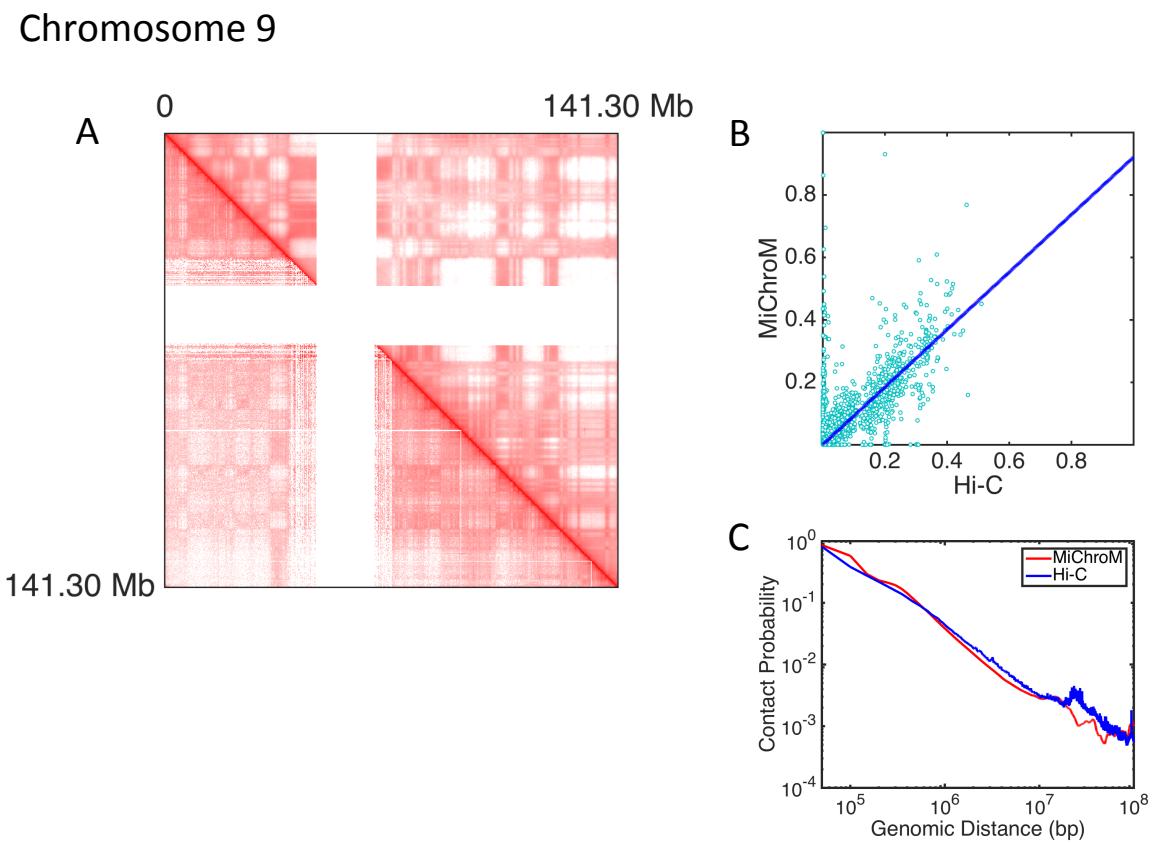

Fig. S20. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 9 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.888 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 2% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

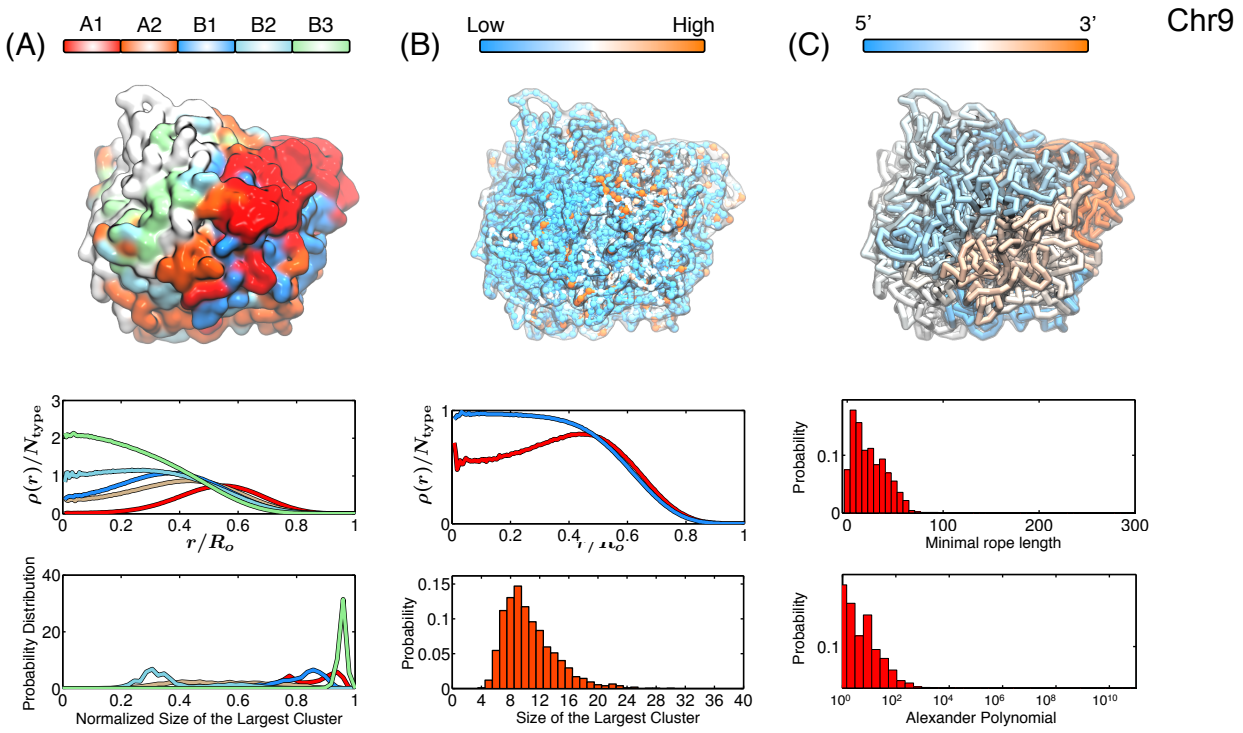

Fig. S21. Structural characterization of the simulated conformations for chromosome 9.
A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

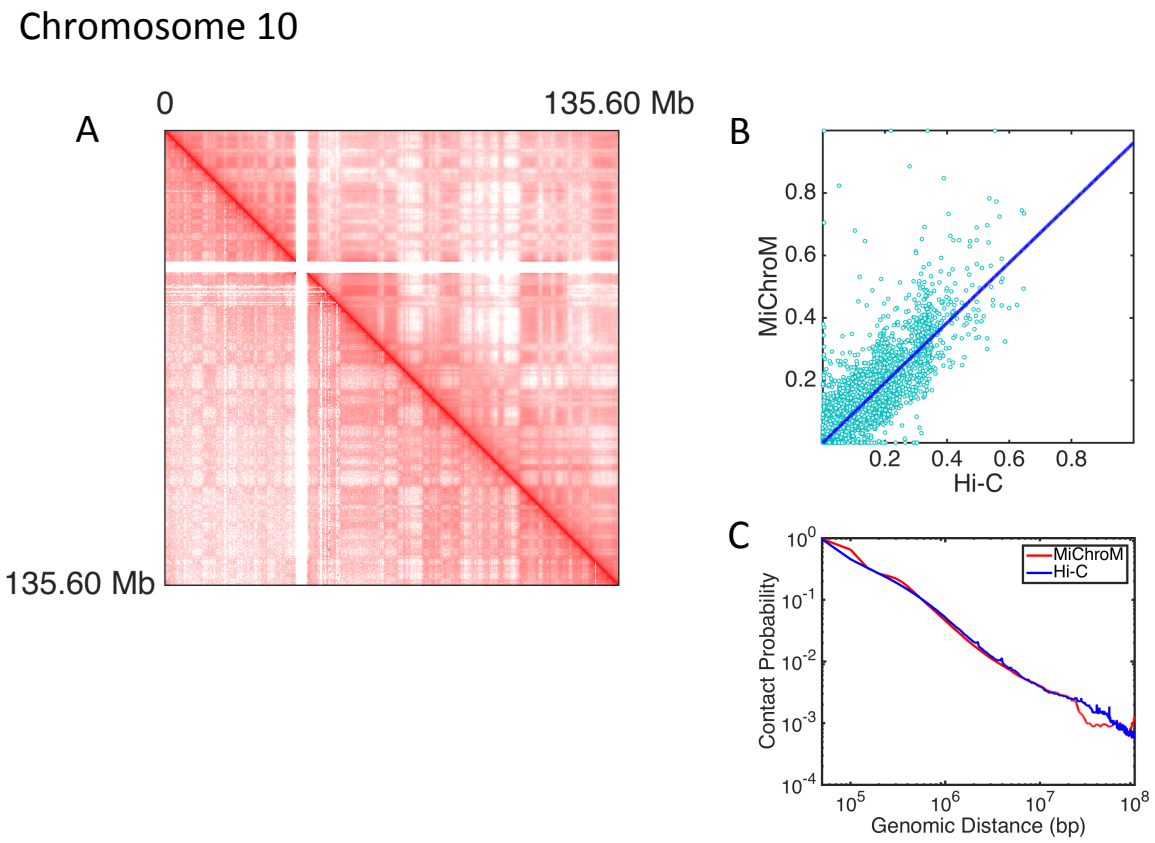

Fig. S22. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 10 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.957 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 10% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

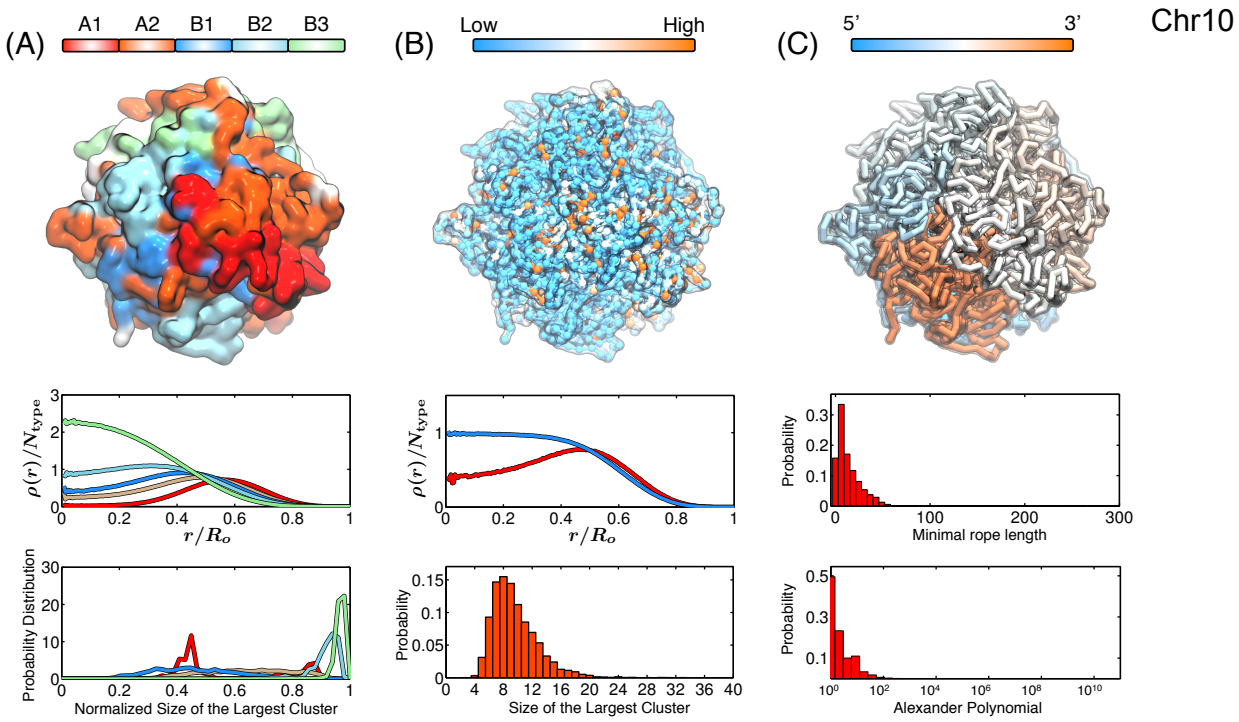

Fig. S23. Structural characterization of the simulated conformations for chromosome 10. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

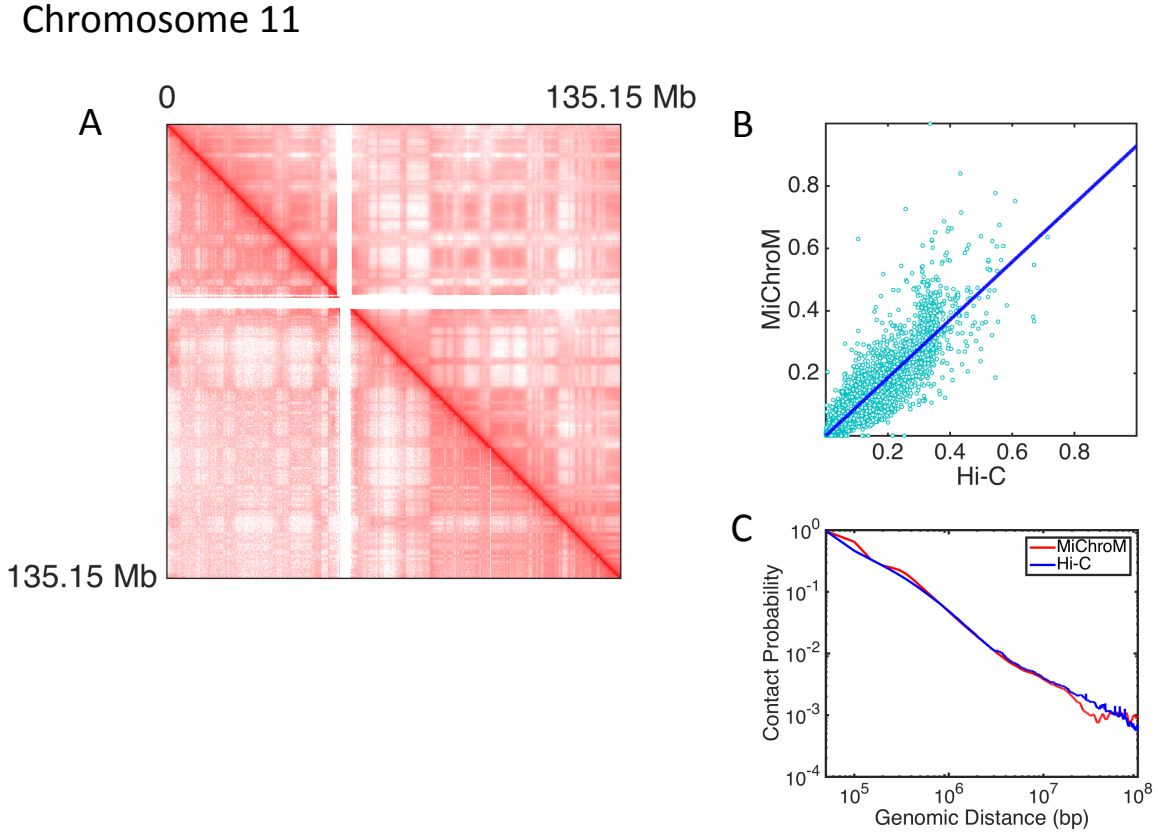

Fig. S24. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 11 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.970 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 10% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

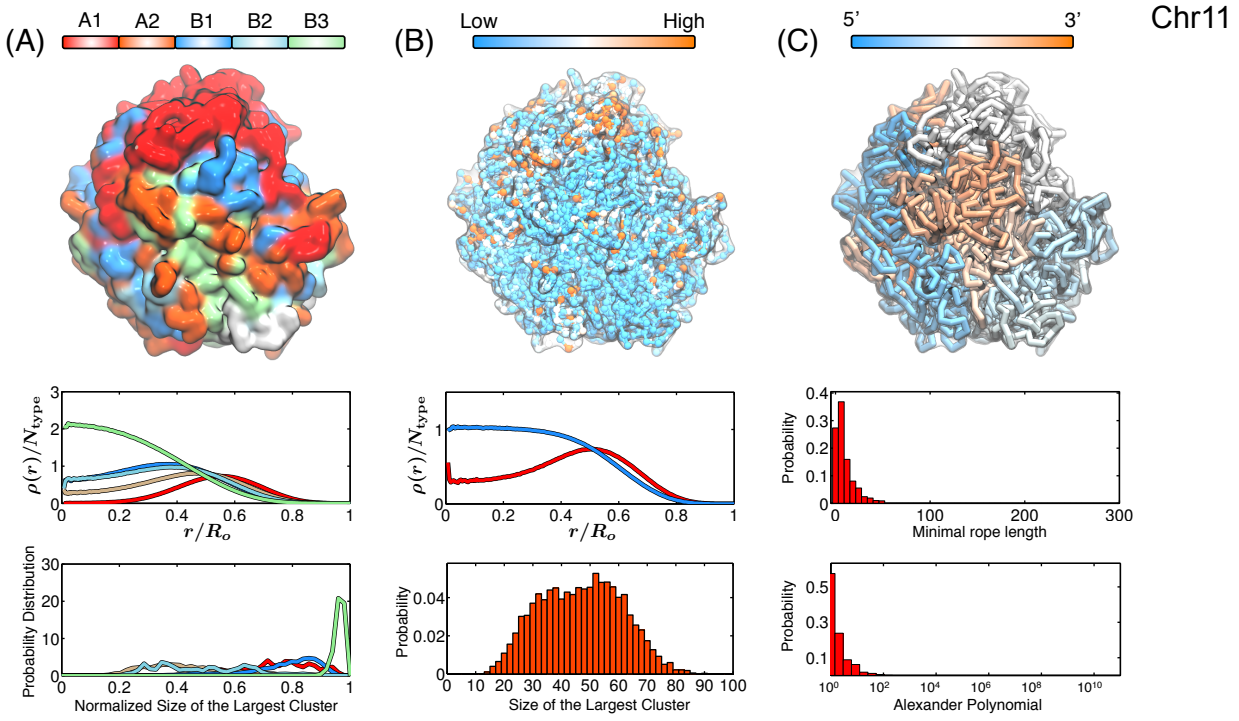

Fig. S25. Structural characterization of the simulated conformations for chromosome 11. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

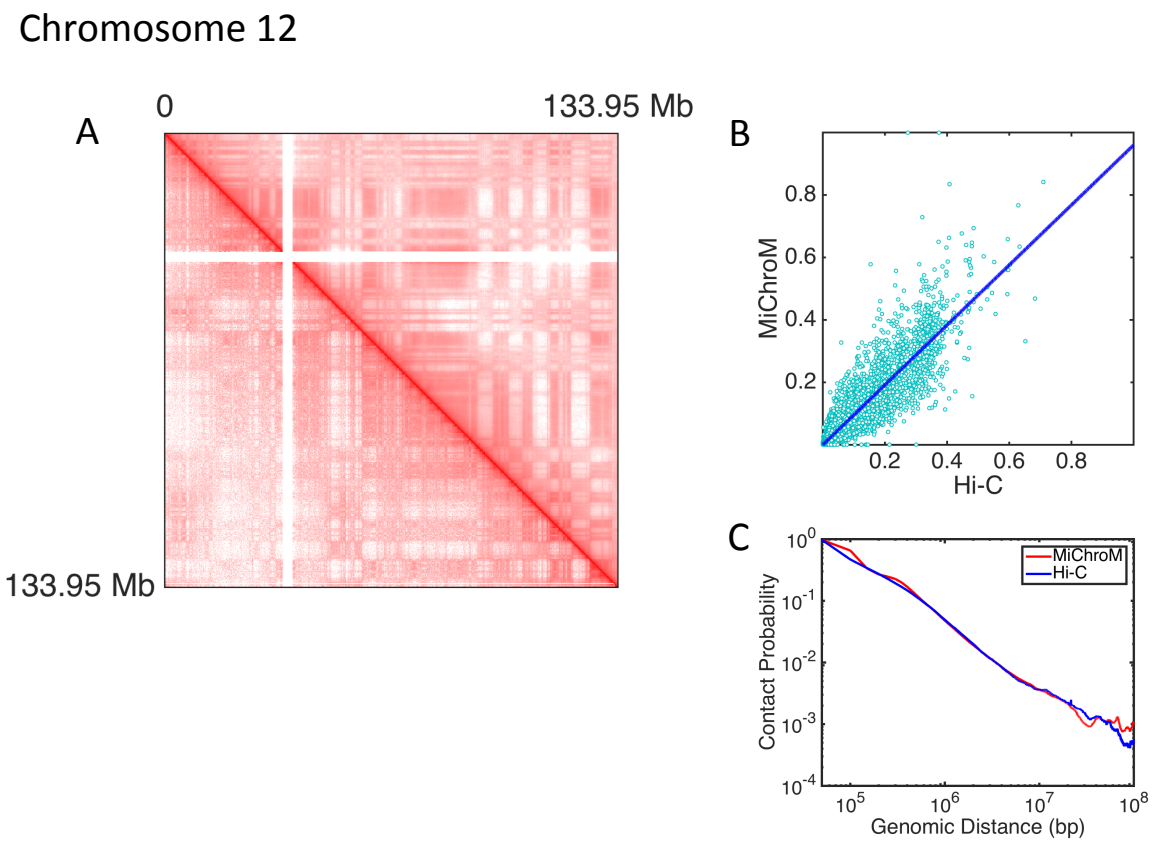

Fig. S26. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 12 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.974 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 10% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

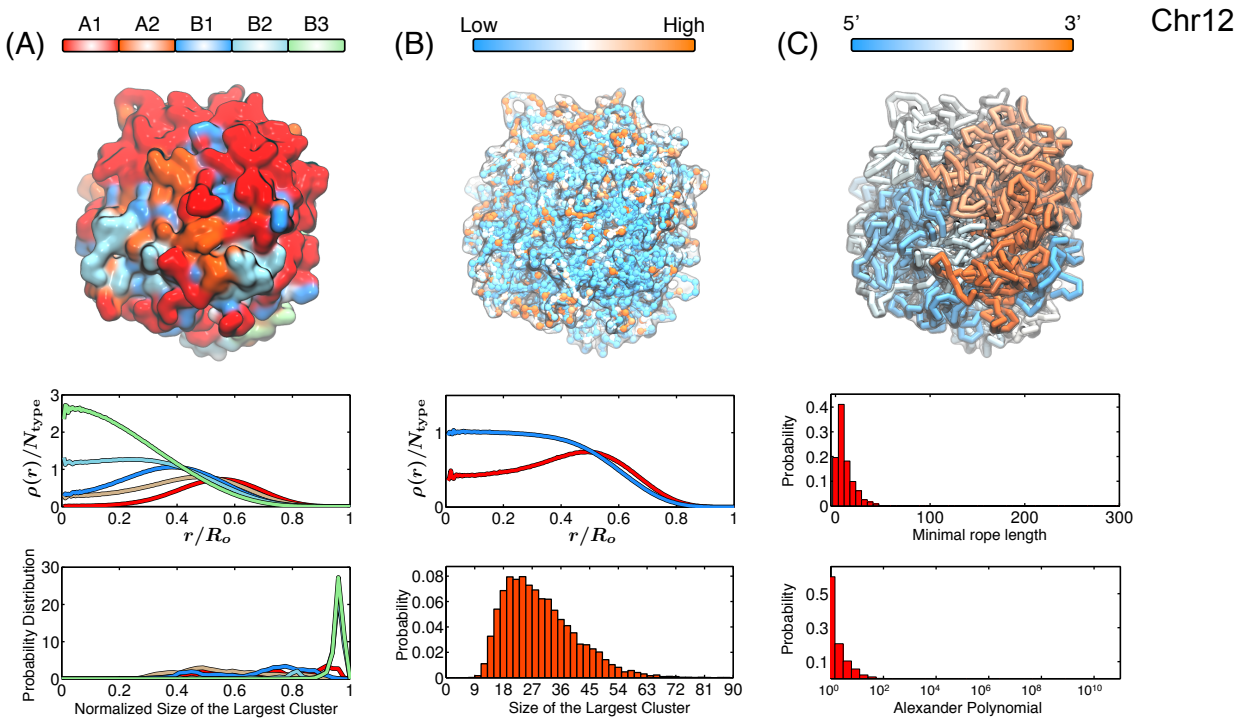

Fig. S27. Structural characterization of the simulated conformations for chromosome 12. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

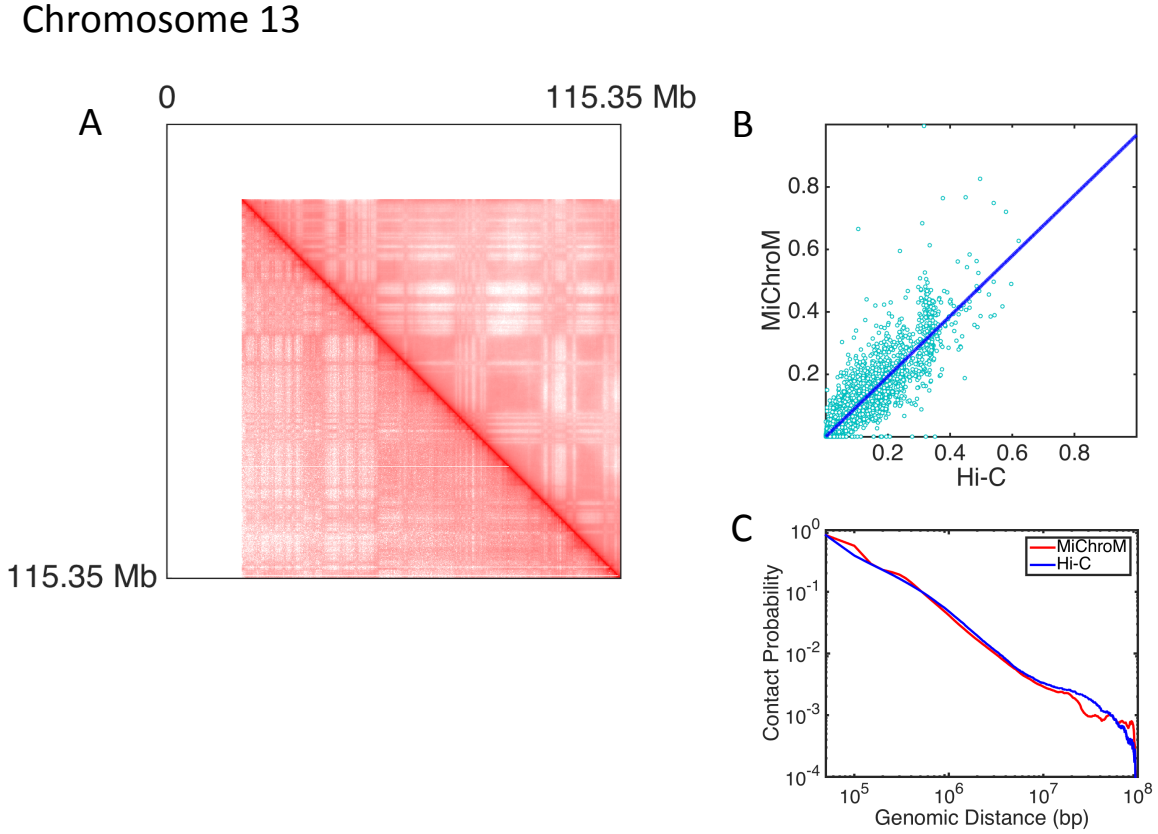

Fig. S28. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 13 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.975 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 10% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

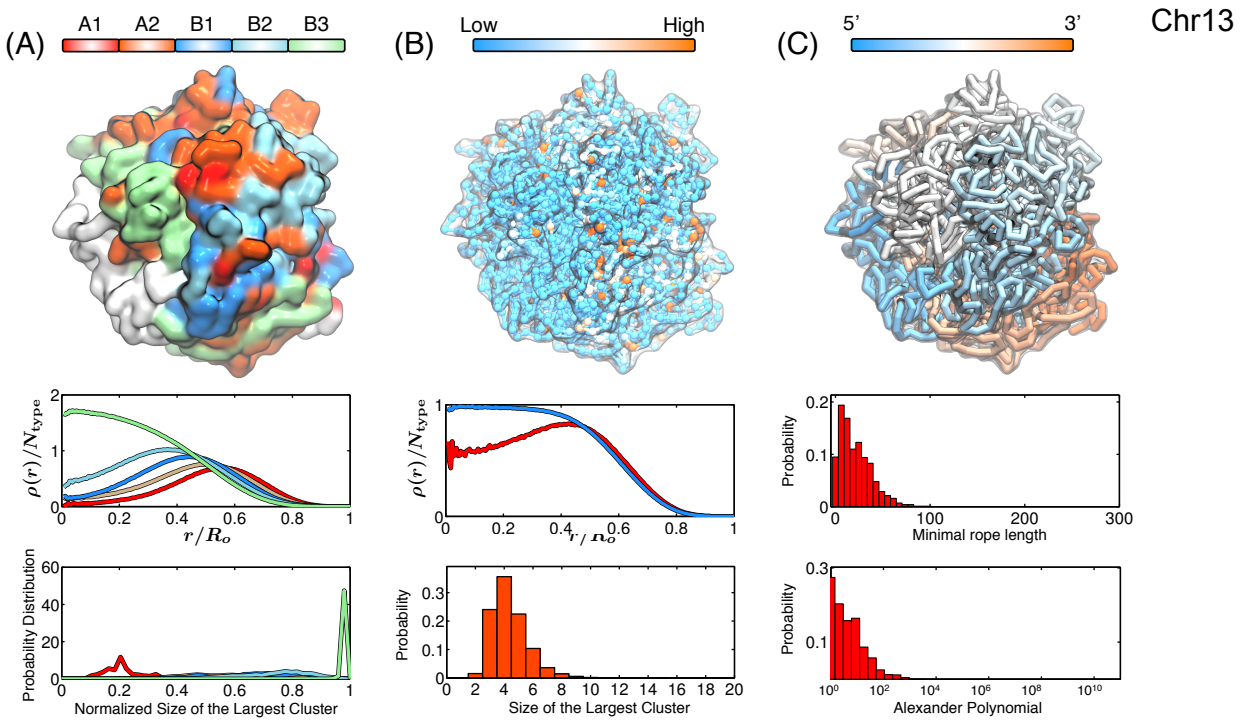

Fig. S29. Structural characterization of the simulated conformations for chromosome 13. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

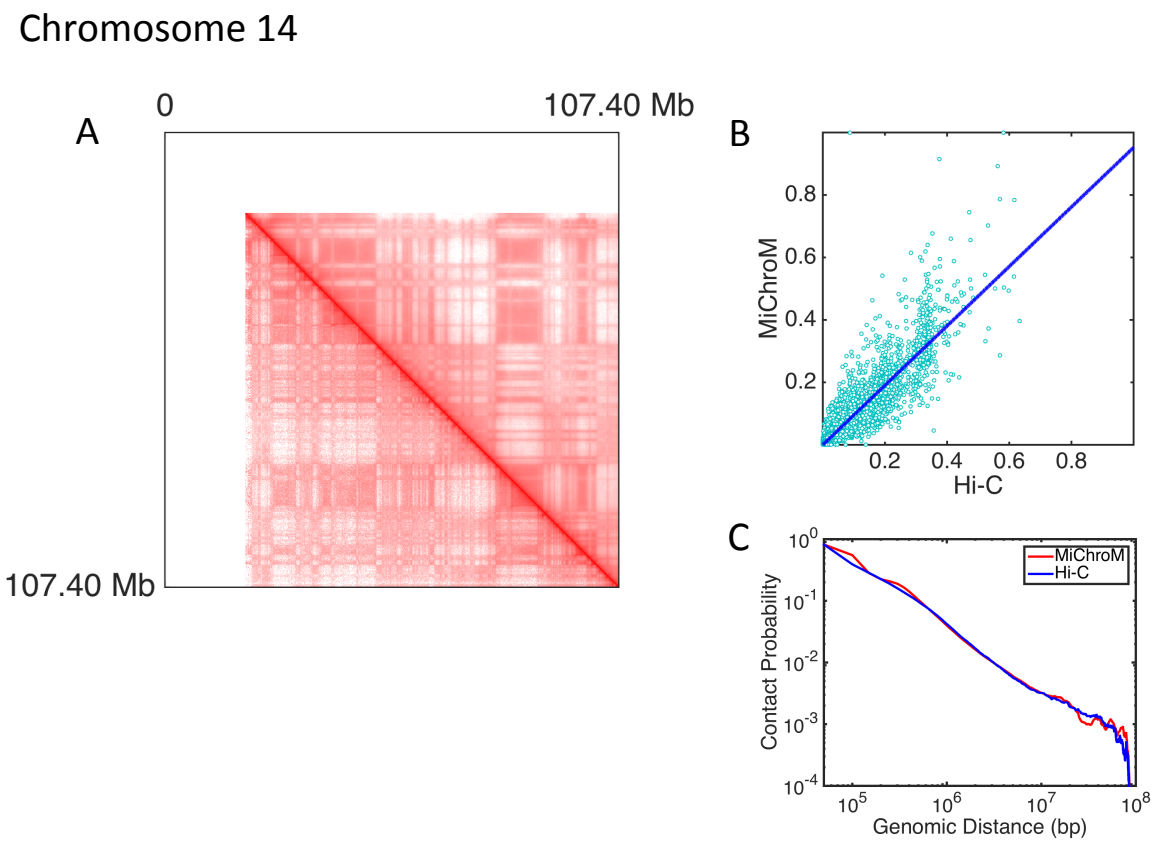

Fig. S30. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 14 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.975 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 10% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

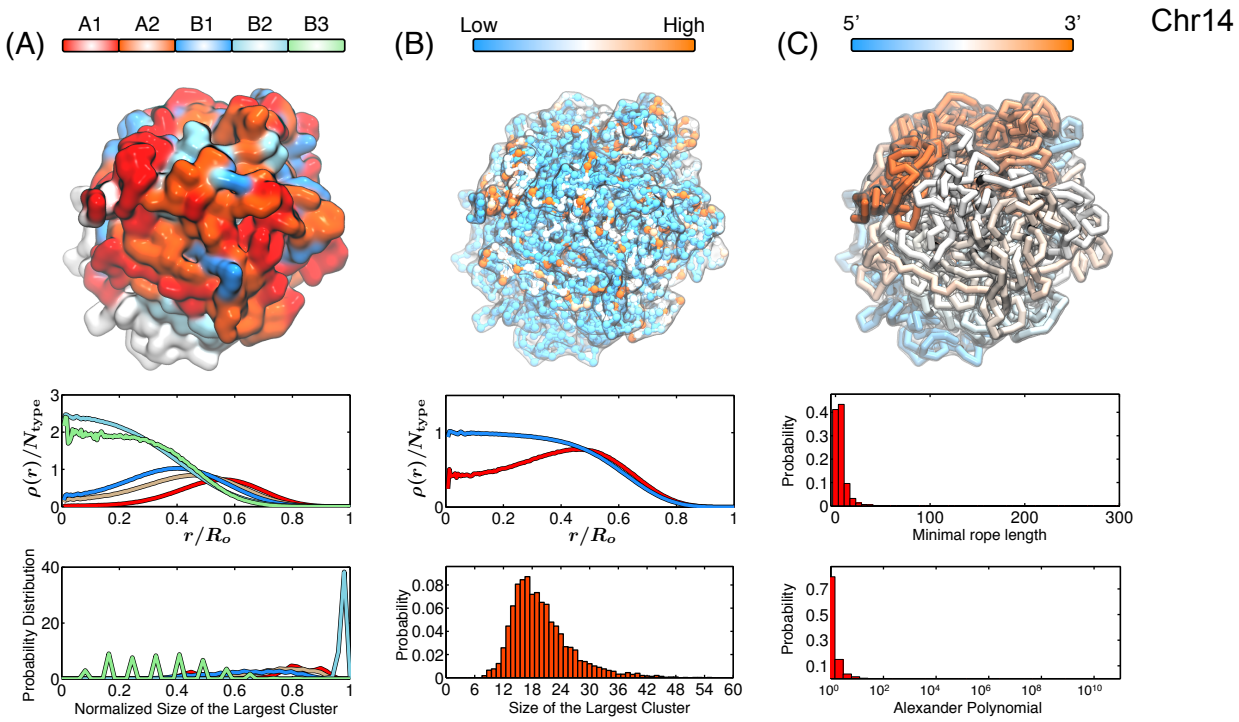

Fig. S31. Structural characterization of the simulated conformations for chromosome 14. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

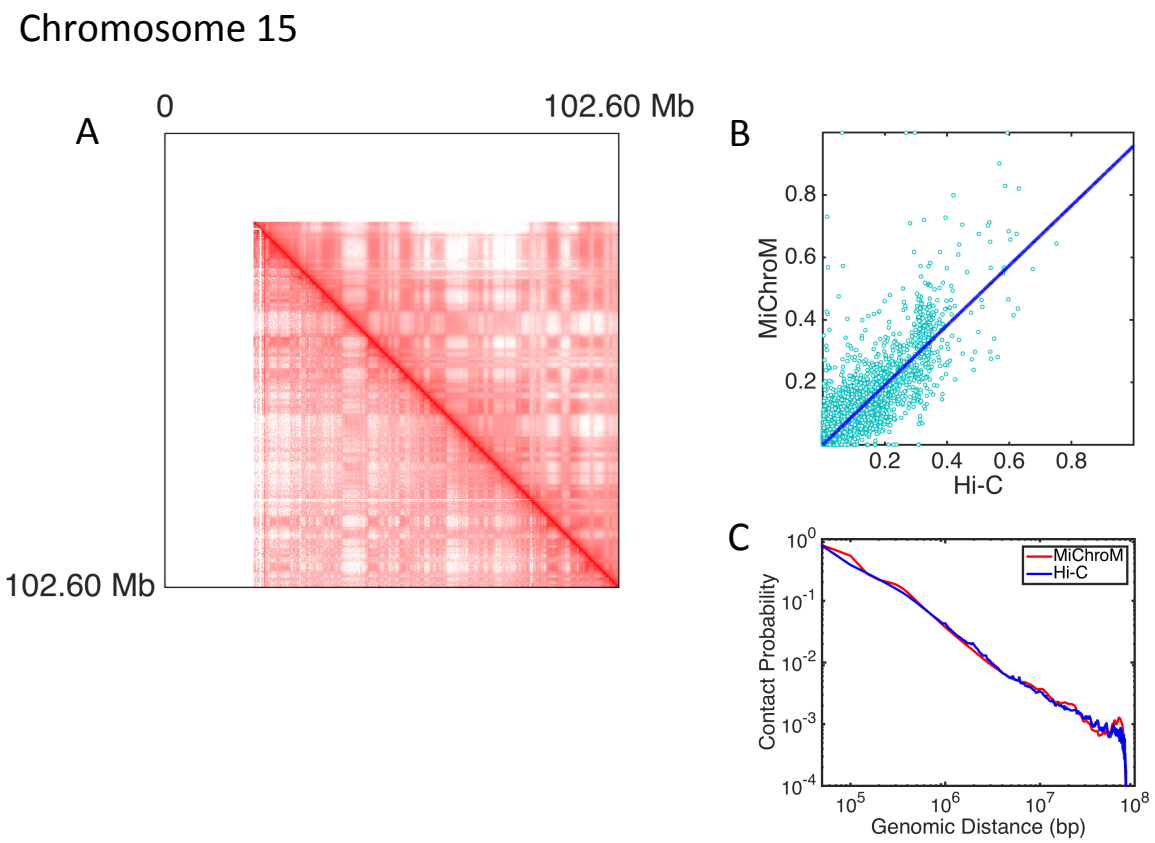

Fig. S32. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 15 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.945 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 10% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

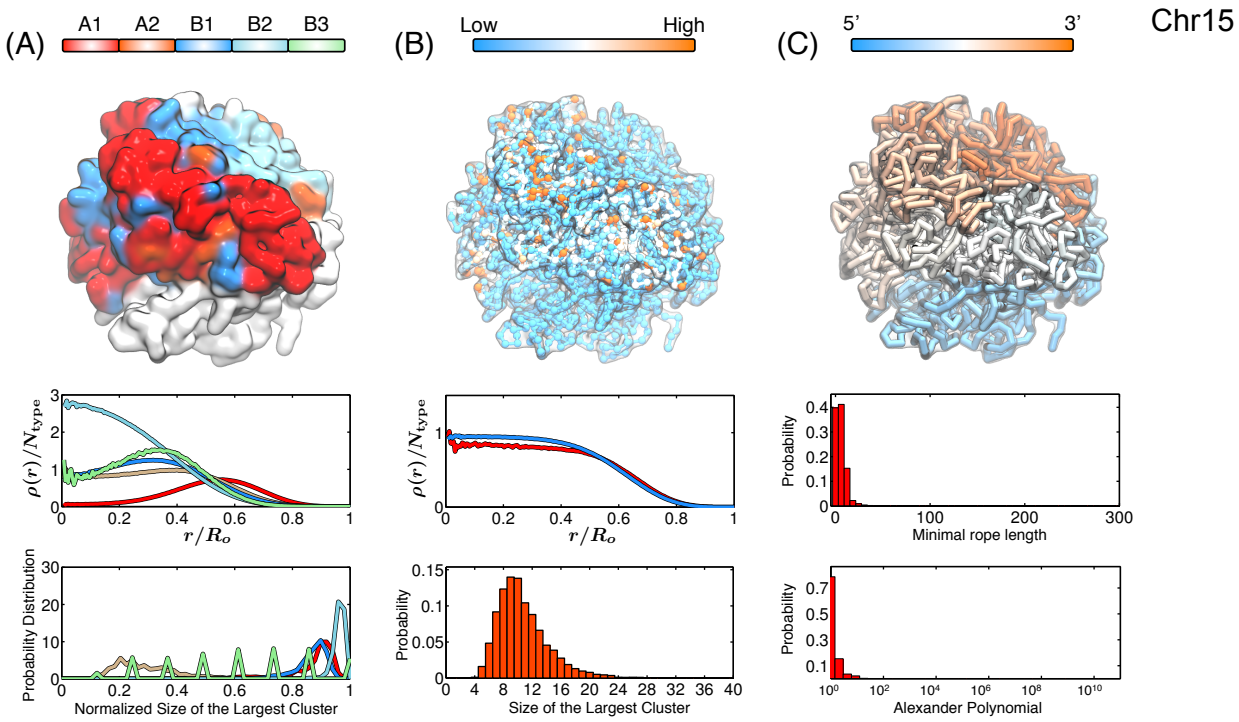

Fig. S33. Structural characterization of the simulated conformations for chromosome 15. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

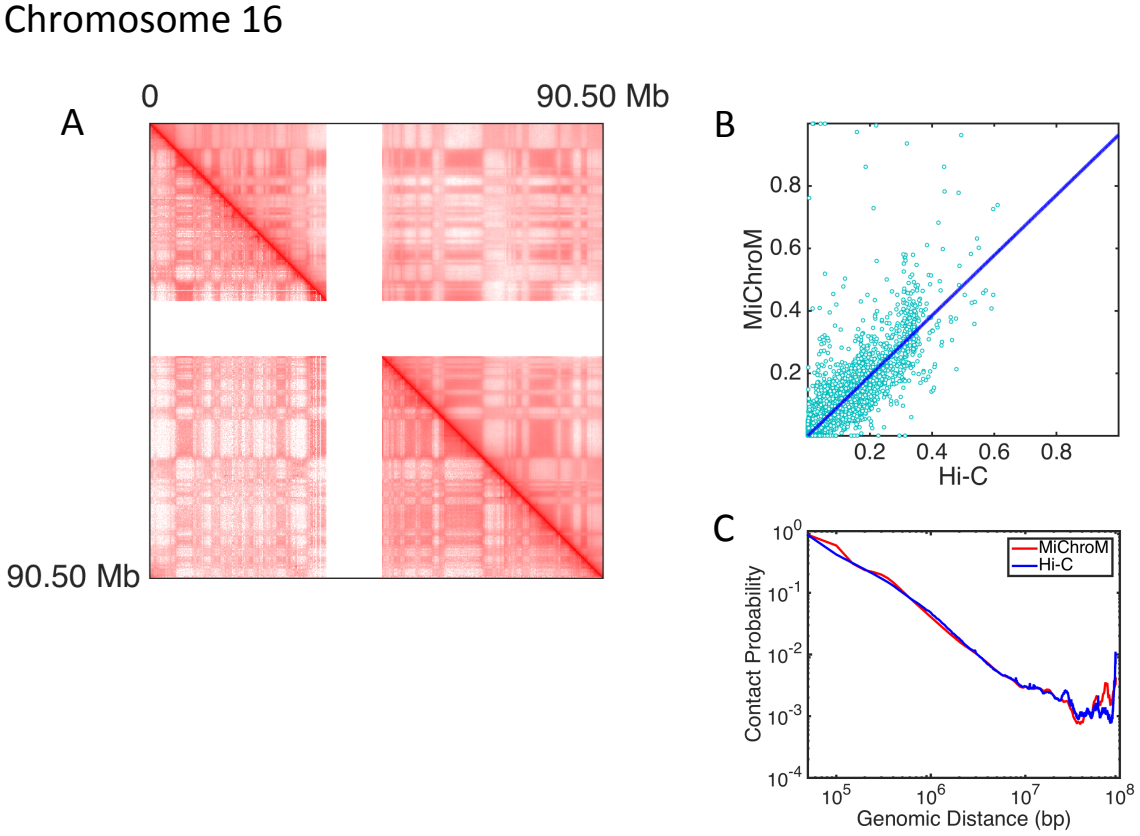

Fig. S34. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 16 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.956 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 10% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

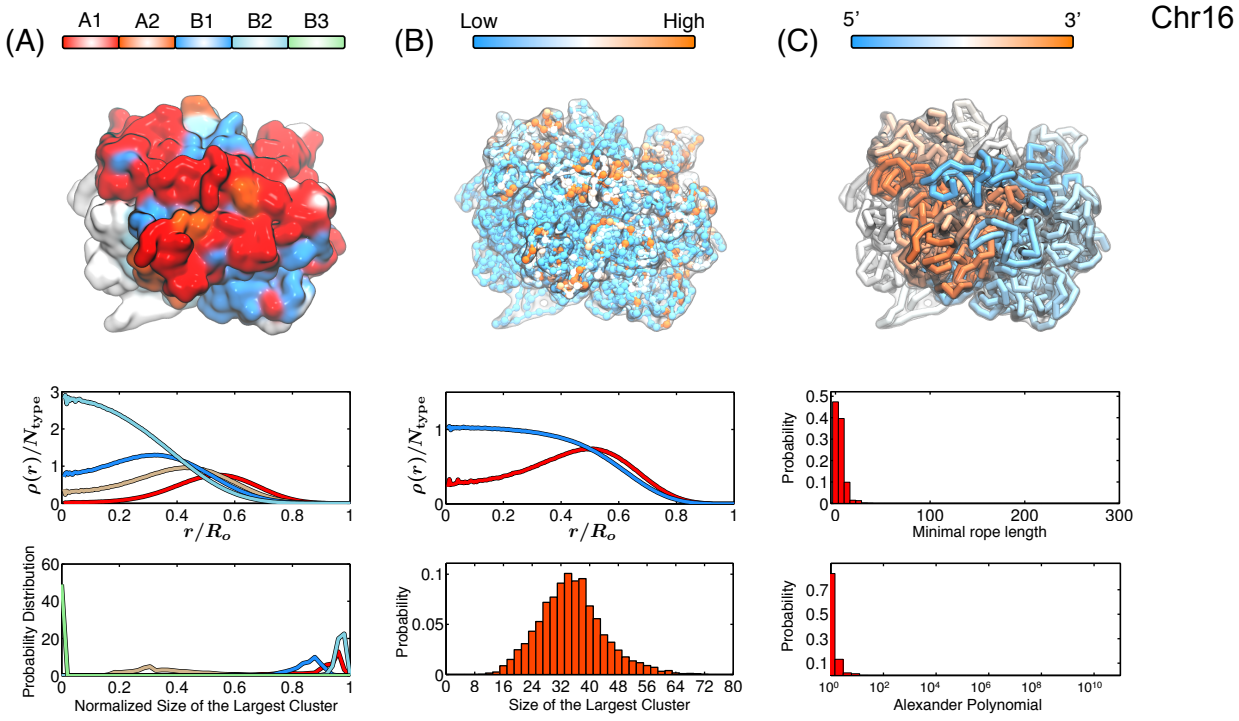

Fig. S35. Structural characterization of the simulated conformations for chromosome 16. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

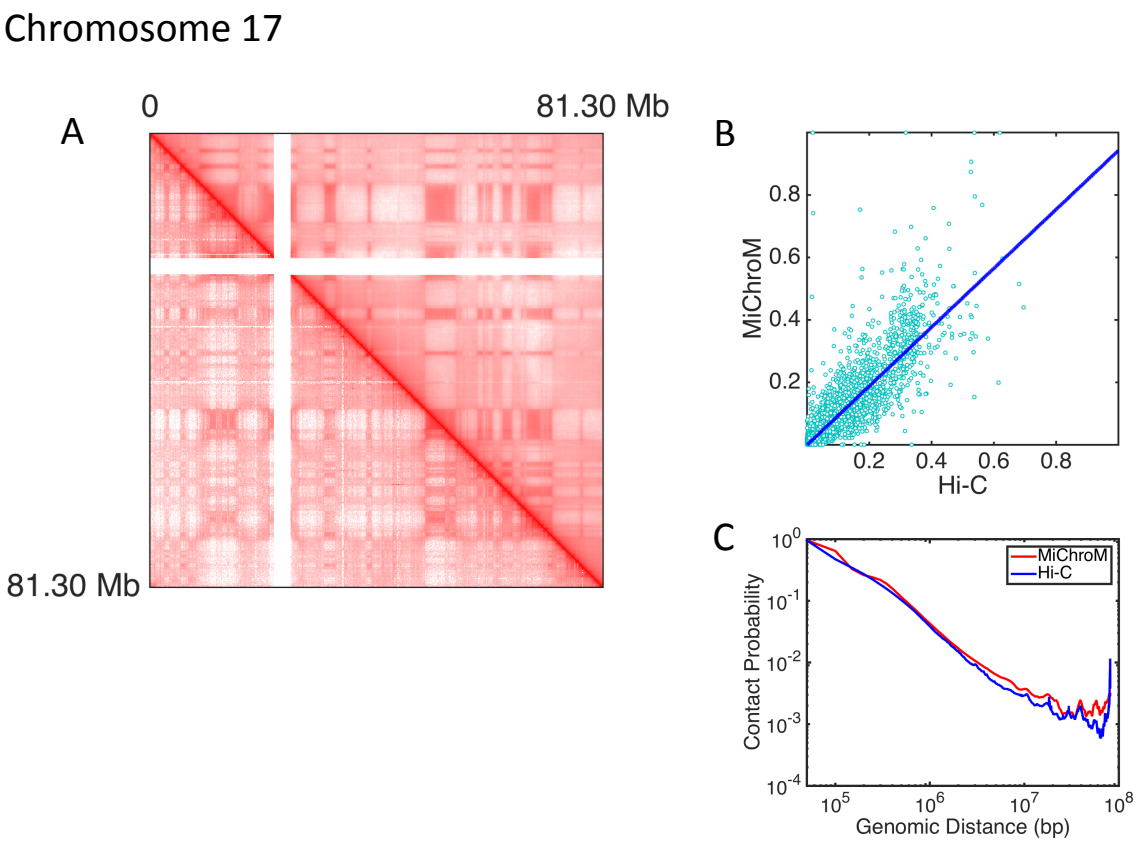

Fig. S36. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 17 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.966 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 10% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

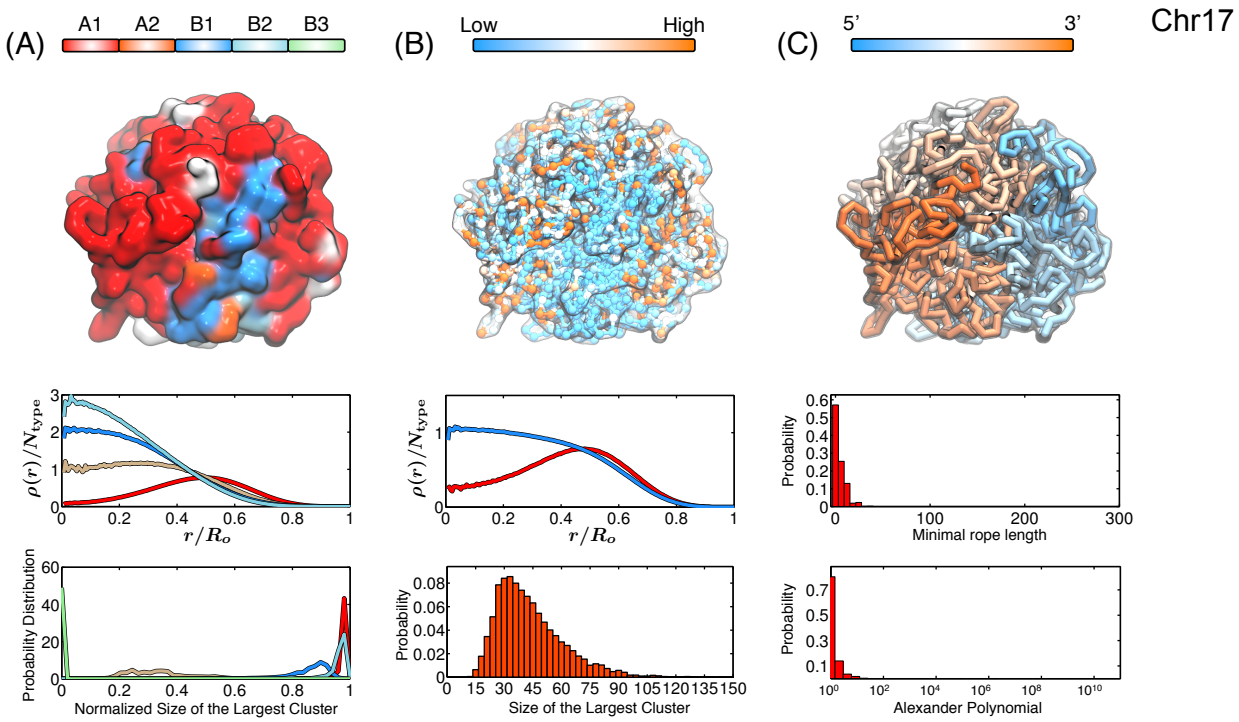

Fig. S37. Structural characterization of the simulated conformations for chromosome 17. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

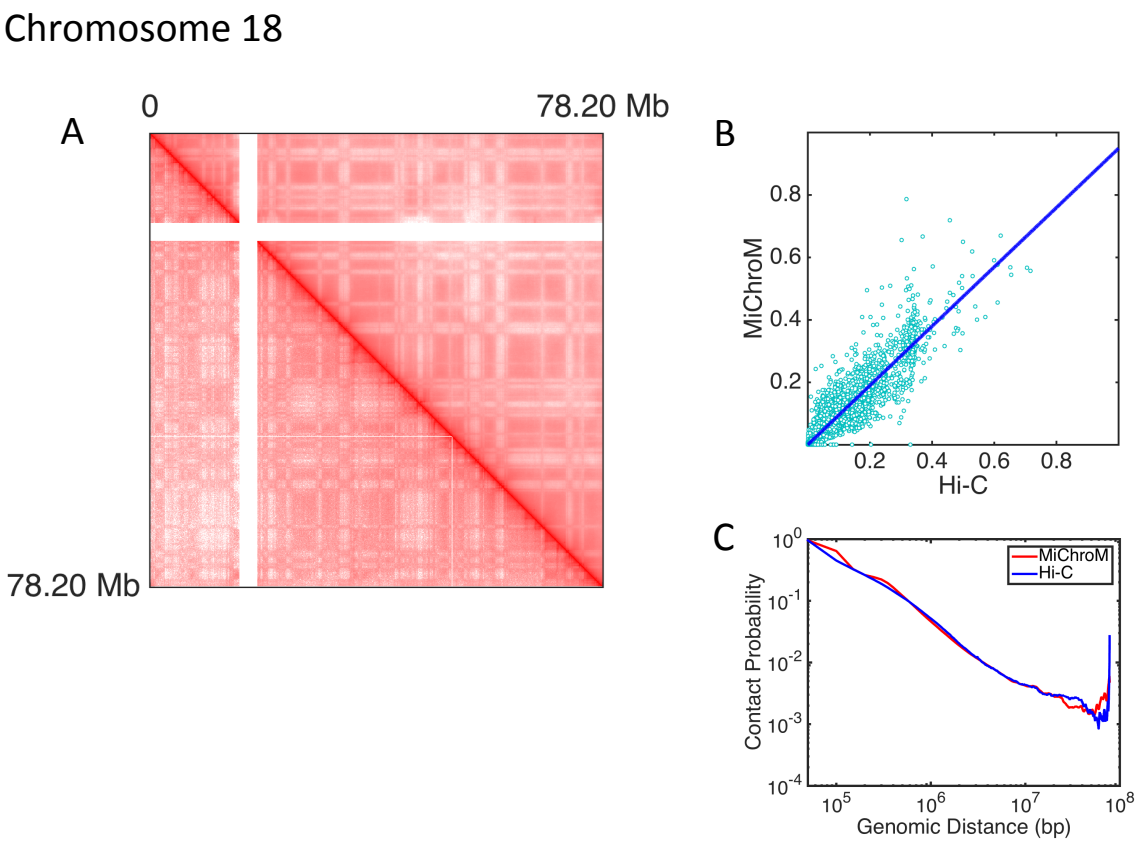

Fig. S38. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 18 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.977. B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 10% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

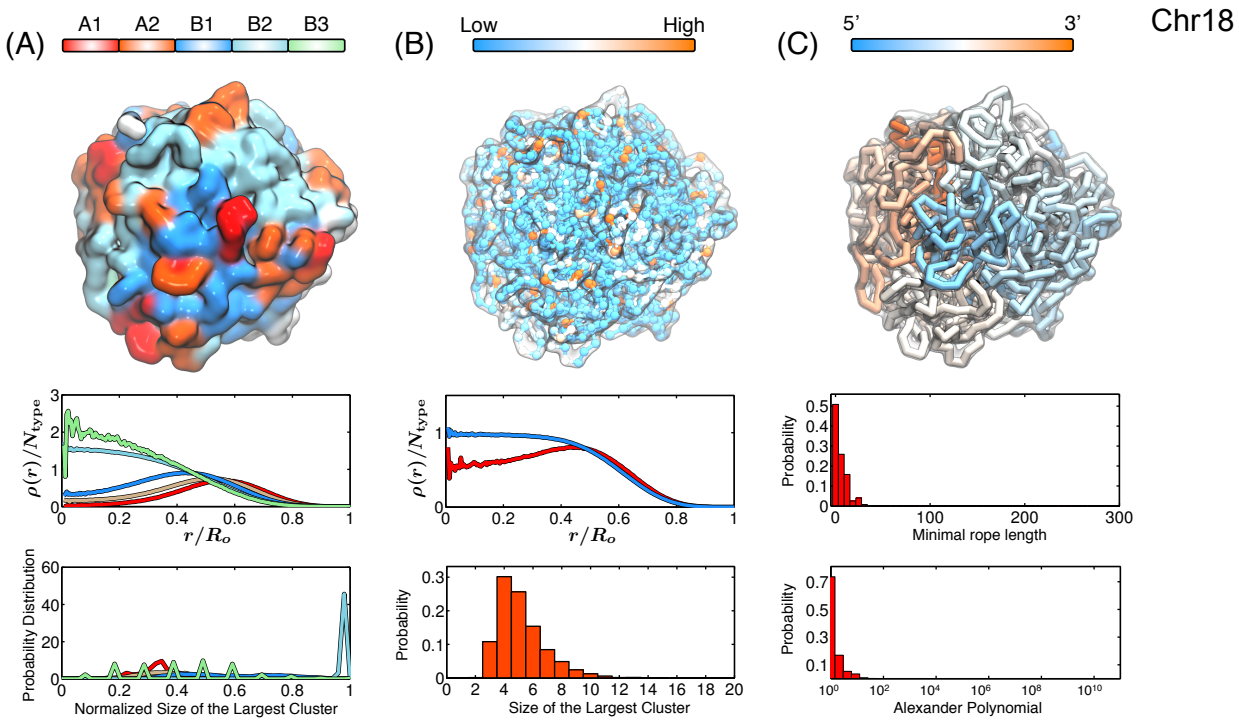

Fig. S39. Structural characterization of the simulated conformations for chromosome 18. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

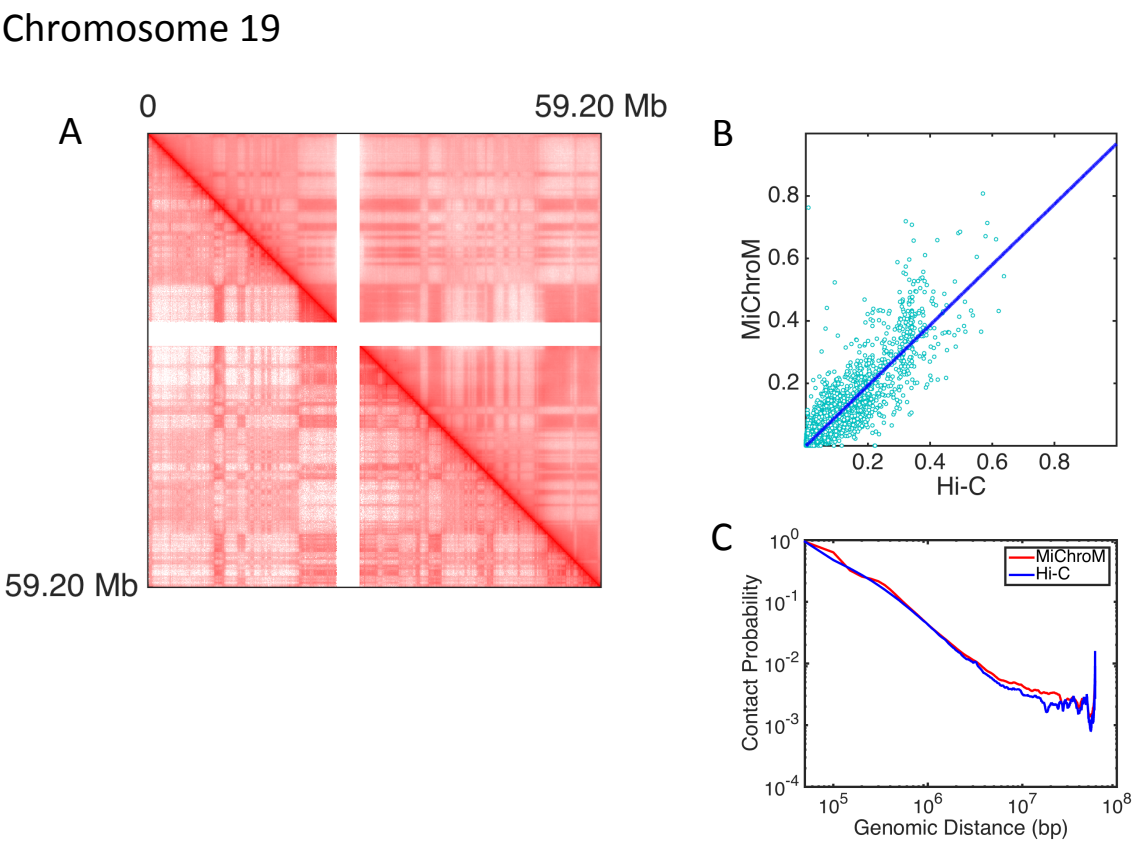

Fig. S40. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 19 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.969 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 10% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

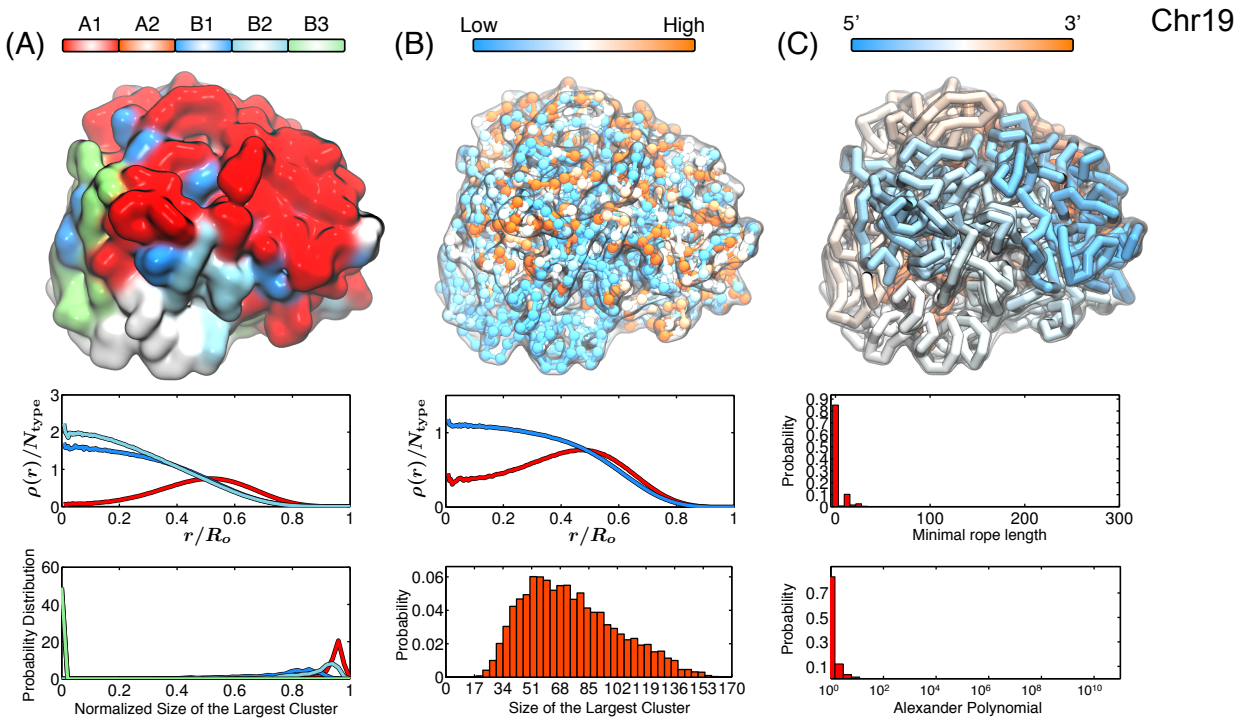

Fig. S41. Structural characterization of the simulated conformations for chromosome 19. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

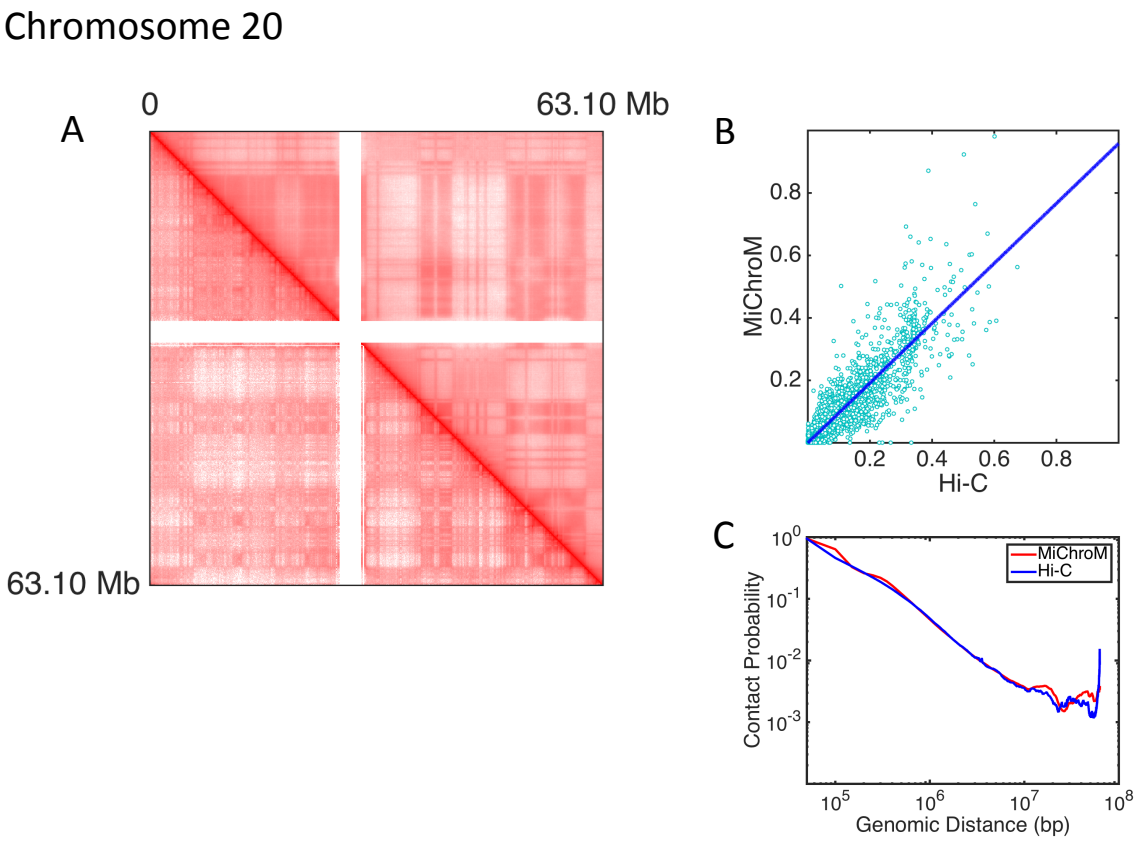

Fig. S42. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 20 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.971 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 10% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

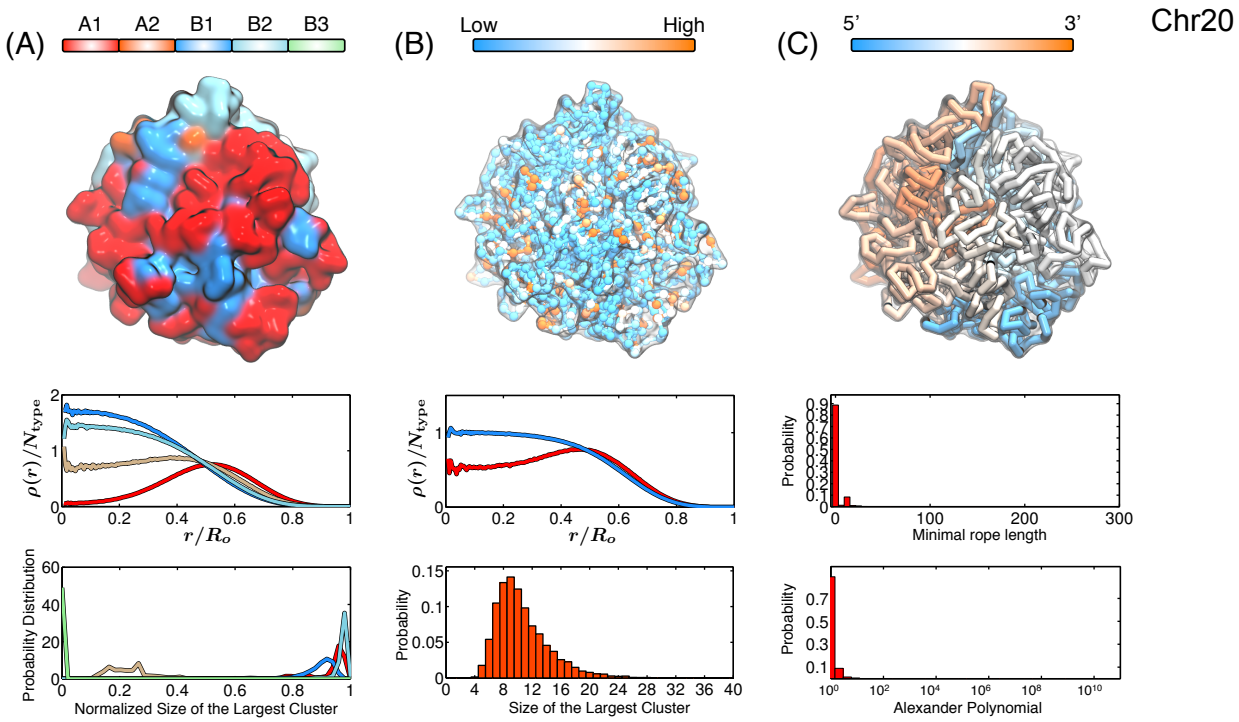

Fig. S43. Structural characterization of the simulated conformations for chromosome 20. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

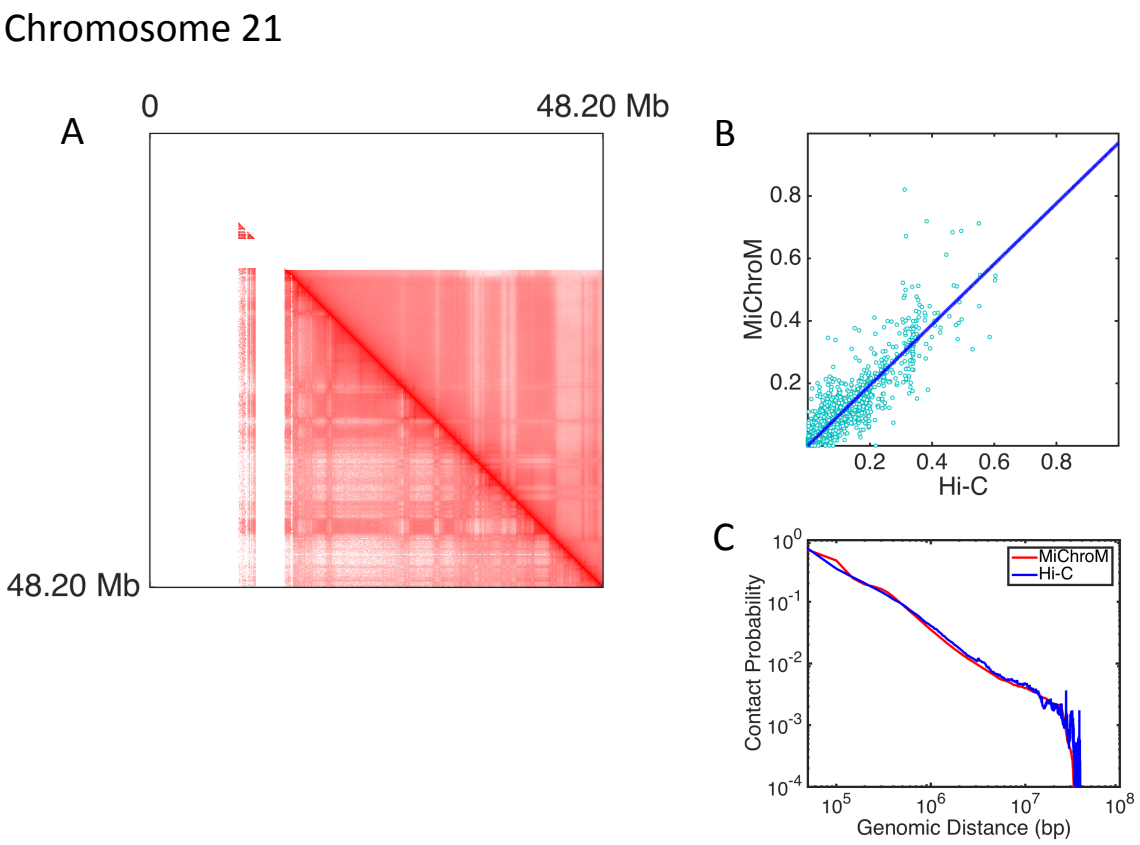

Fig. S44. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 21 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.936 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 10% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

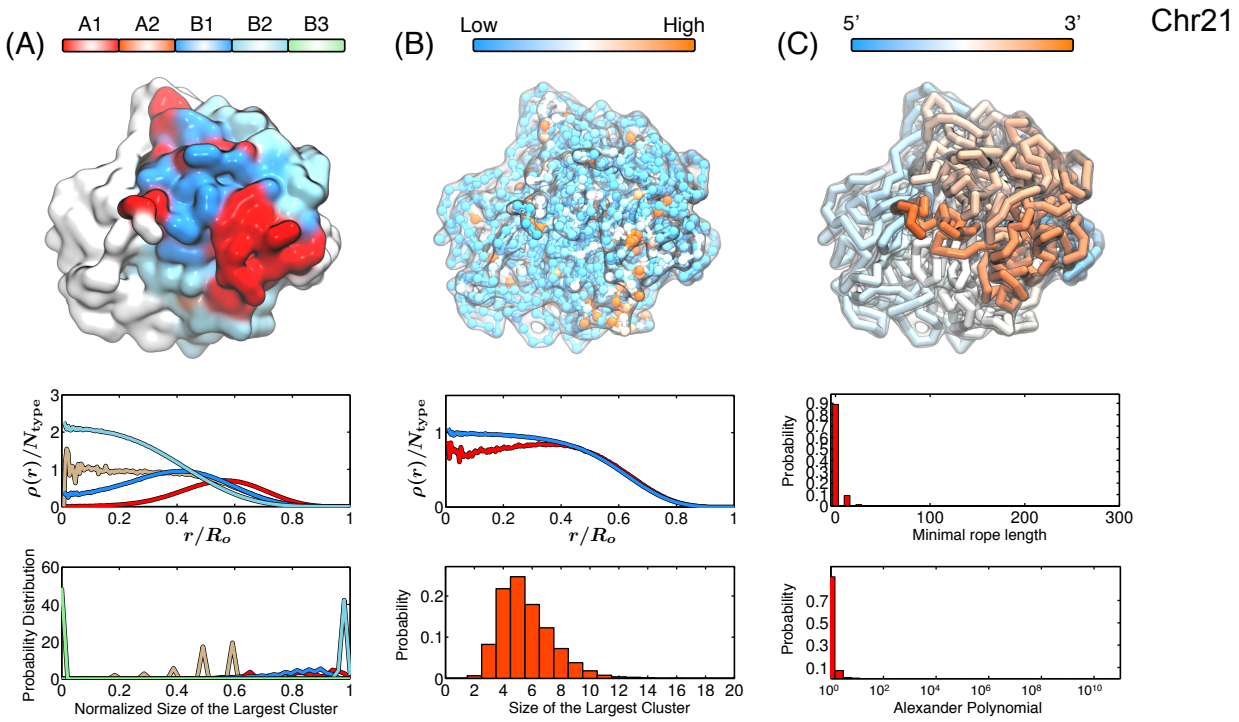

Fig. S45. Structural characterization of the simulated conformations for chromosome 21. A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type.
B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations.
C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

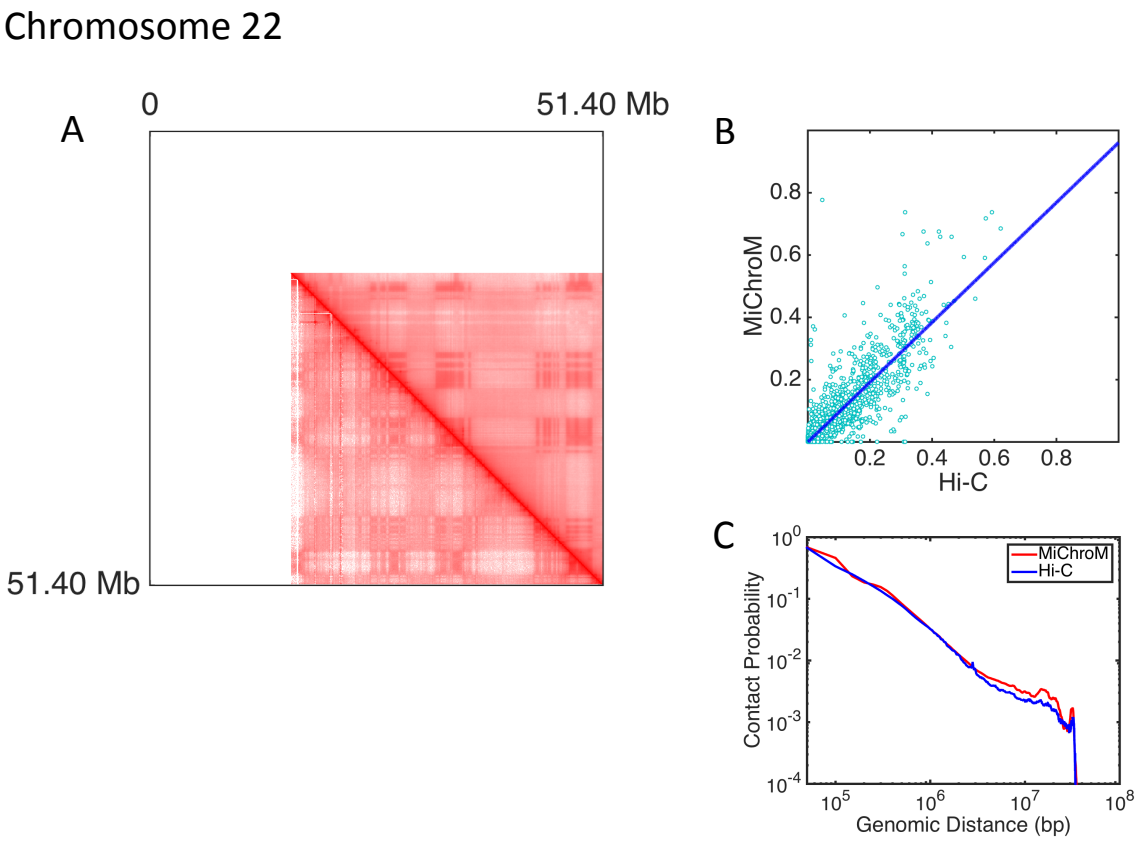

Fig. S46. MiChroM generates 3D structures for all autosomal chromosomes of B-lymphoblastoid cells (GM12878) that are in close agreement with experimental observations. A) Contact map of chromosomes 22 represented in log scale; upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The quality of the generated contact map is high, as shown by the symmetry of the map. Pearson's correlation between the two datasets is 0.953 . B) Scatter plots of Hi-C vs MiChroM datasets (for better rendering only 10% of contacts are shown on graphics) together with a linear fit of the data obtained by using the least squares method. C) The probability of contacts as a function of genomic distance in both measured and modeled maps.

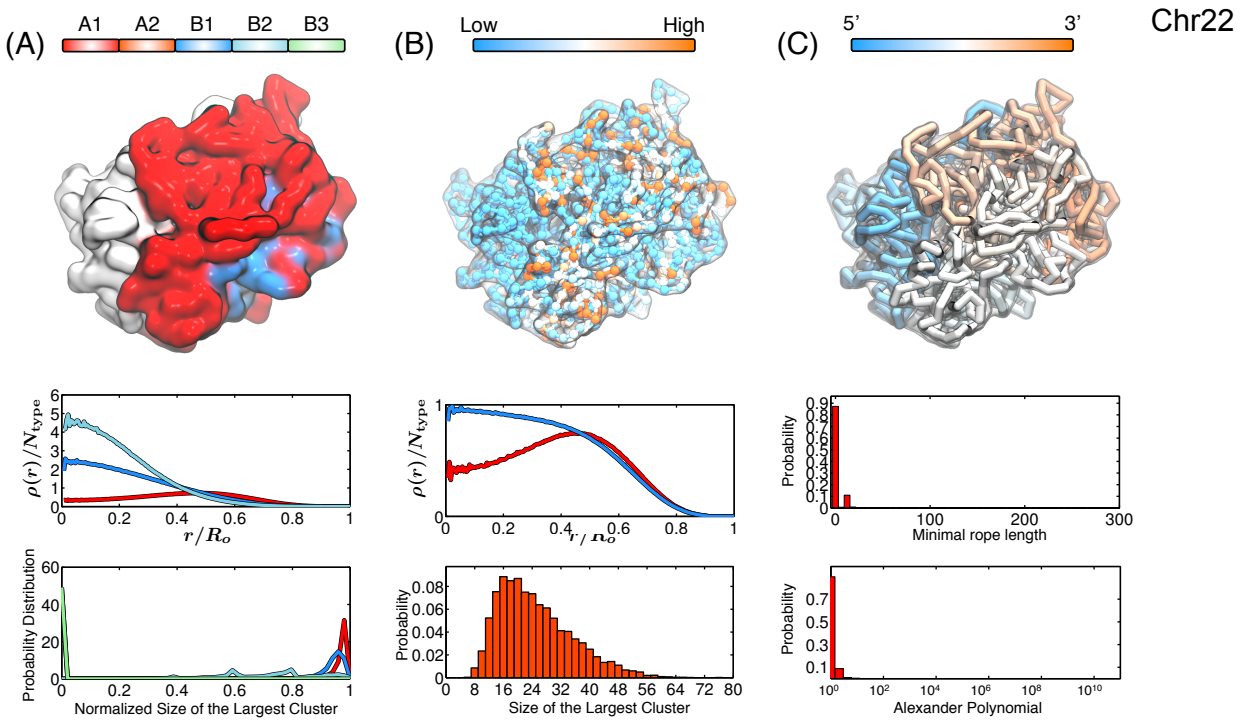

Fig. S47. Structural characterization of the simulated conformations for chromosome 22. (A) Chromatin of different types phase separate, with A types localizing at the surface and B types residing in the interior. (*Top*) Surface plot for the chromosome colored by chromatin types, with the coloring scheme shown at the side. (*Middle*) Radial density profiles for different chromatin types calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions for the size of the largest cluster found for each chromatin type from the ensemble of simulated chromosome conformations. The cluster sizes for different types are normalized by the total number of genomic loci for that type. (B) Genomic loci with high gene expression spatially colocalize at the exterior of the chromosome. (*Top*) A chromosome structure colored by gene expression with red and blue representing high and low expression, respectively. (*Middle*) Radial density profiles for genomic loci with high and low gene expression calculated from the ensemble of simulated chromosome conformations. (*Bottom*) Probabilities for finding the largest cluster of size *N* for highly expressed genomic loci from the ensemble of simulated chromosome conformations. (C) Simulated chromosomes adopt knot-free conformations. (*Top*) A chromosome structure colored by genomic distances, with one end of the chromosome shown in blue and the other end in red. (*Middle*) Probability distributions of the knot invariant measured as minimal rope length for the ensemble of simulated chromosome conformations. (*Bottom*) Probability distributions of the knot invariant measured as Alexander polynomial for the ensemble of simulated chromosome conformations.

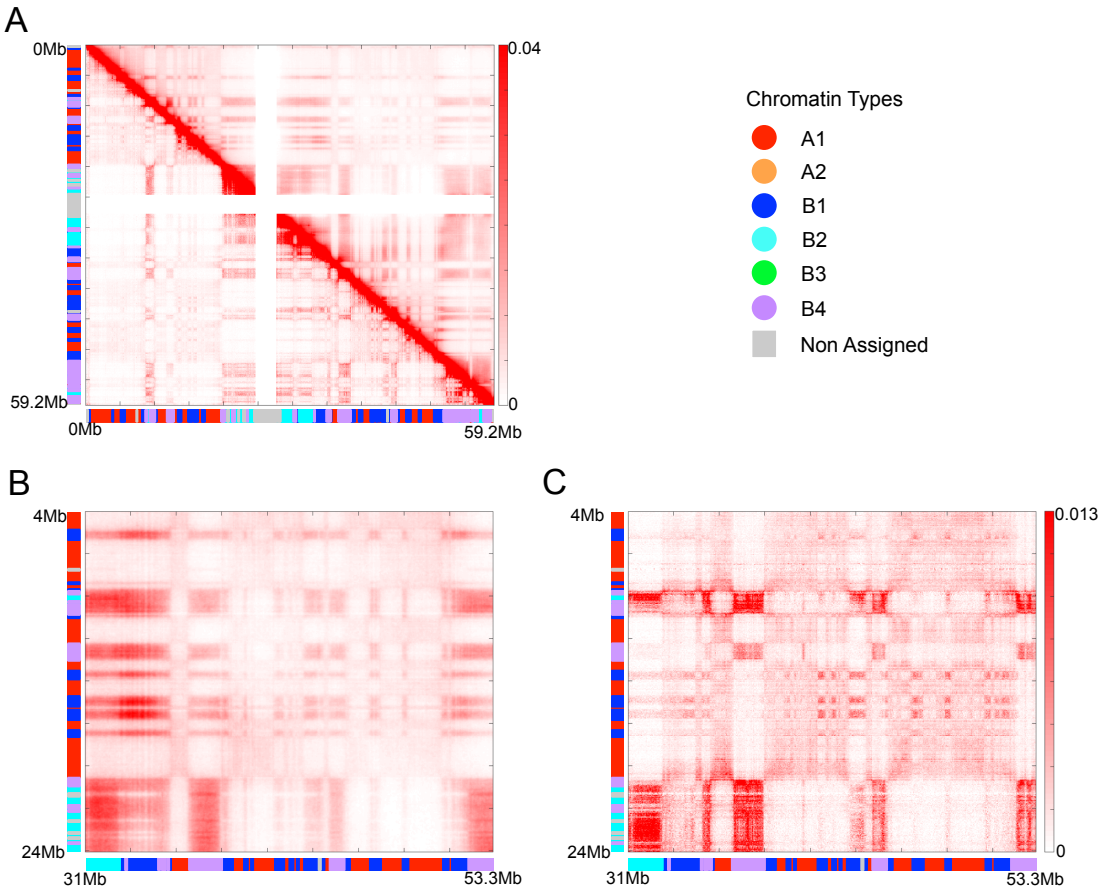

Fig. S48. Subcompartments in chromosome 19. Chromosome 19 contains a large amount of chromatin of type B4 that was treated in MiChroM as B3, which in turn is absent in this chromosome. For this reason here we assess the quality of the contact map obtained *in silico* A) Contact map for chromosome 19, upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. The map appears symmetric at first inspection indicating that the two data sets are similar despite the wrong assignment of chromatin type. Differently from the map in figure S40 this map is not in log scale. B) and C) show (with linear color scale) a magnification of the same region of simulated and experimental contact maps of chromosome 19. In this region of the map the subcompartments structure of type B is clearly visible. It is evident that the behavior of type B4 is only partially reproduced.

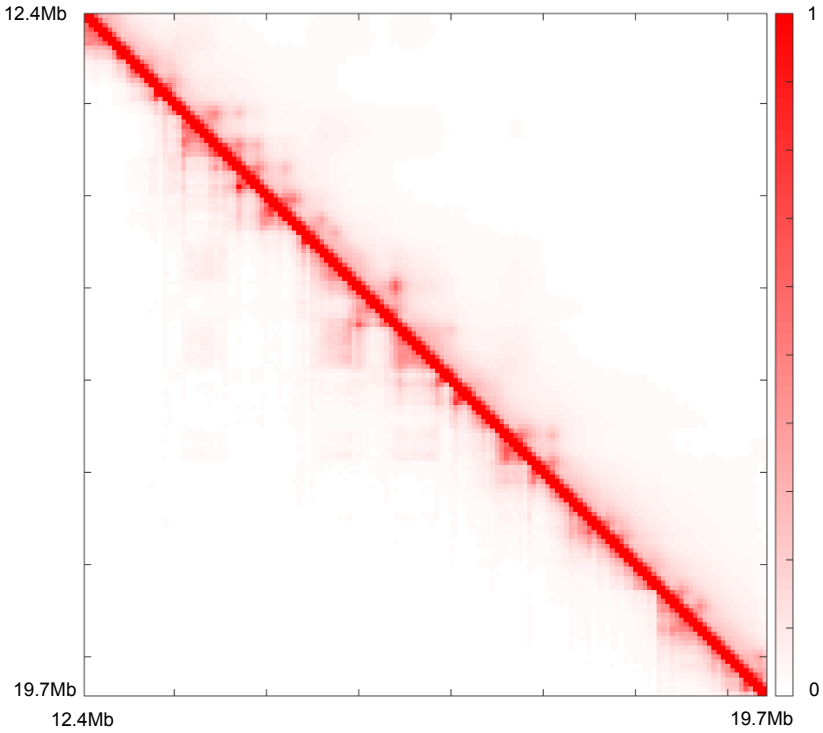

Fig. S49. Chromatin looping in MiChroM. Figure shows a magnification of the contact map of chromosome 19, upper diagonal regions show results from computational modeling and lower diagonal regions show maps obtained using Hi-C. Loops are visible in both the experimental and simulated map as local peaks in contact probability. The peaks characterizing the loops appear to extend over larger regions in the simulated map, which in general shows smoother transitions between high and low contact probabilities. We attribute the lower resolution of the simulated maps to the coarse graining introduced in MiChroM by using a single bead to represent 50 kb of DNA.

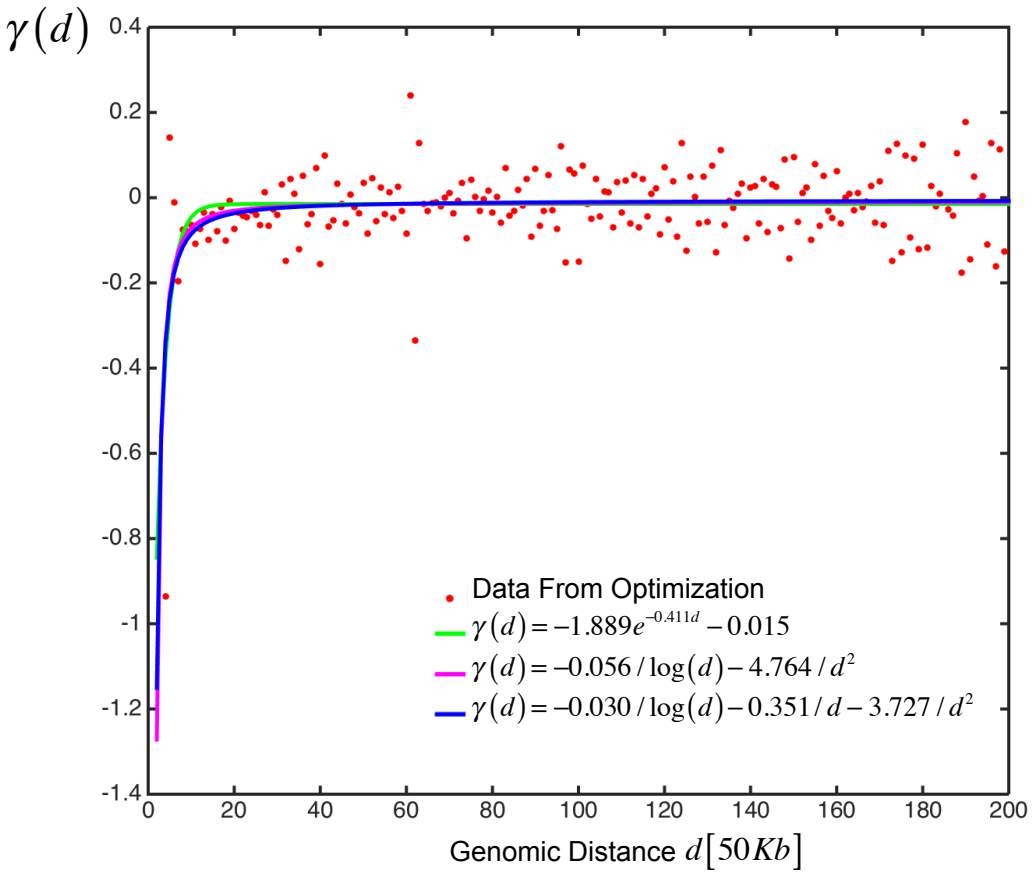

Fig. S50. Several different functional forms can recapitulate the ideal chromosome potential with similar accuracy. The red dots represent the data set $\{\gamma_d, d = 3,...,200\}$ as obtained from the optimization on chromosome 10 of B-lymphoblastoid cells (GM12878). Different possible fitting functions are shown in solid lines. The function used in the sampling phase of all the chromosomes is shown in blue.

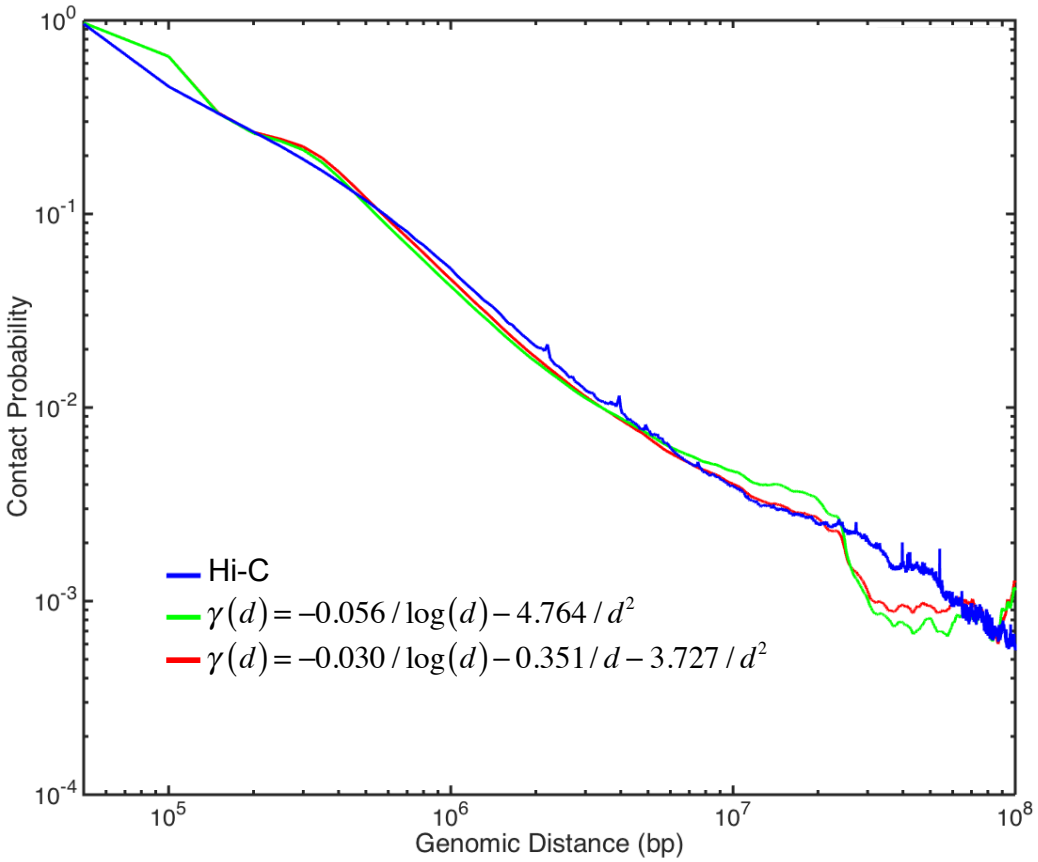

Fig. S51. The contact probability as a function of genomic distance for simulated data from two distinct functional forms of the ideal chromosome potential are shown together the one extrapolated from experimental data. Figure shows that the power law scaling of contact probabilities has little to no sensitivity toward the precise functional form of the ideal chromosome potential; every functional form consistent with the contact energies extracted from the optimization is equally valid.

References

1. Jaynes ET (1957) Information Theory and Statistical Mechanics. *Phys Rev* 106(4):620-630.
2. Agmon N, Alhassid Y, & Levine RD (1979) Algorithm for Finding the Distribution of Maximal Entropy. *J Comput Phys* 30(2):250-258.
3. Mead LR & Papanicolaou N (1984) Maximum-Entropy in the Problem of Moments. *J Math Phys* 25(8):2404-2417.
4. Kremer K & Grest GS (1990) Dynamics of Entangled Linear Polymer Melts - a Molecular-Dynamics Simulation. *J Chem Phys* 92(8):5057-5086.
5. Naumova N*, et al.* (2013) Organization of the Mitotic Chromosome. *Science* 342(6161):948-953.
6. Rosa A & Everaers R (2008) Structure and Dynamics of Interphase Chromosomes. *Plos Comput Biol* 4(8).
7. Rao SSP*, et al.* (2014) A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* 159(7):1665-1680.
8. Plimpton S (1995) Fast Parallel Algorithms for Short-Range Molecular-Dynamics. *J Comput Phys* 117(1):1-19.
9. Mortazavi A, Williams BA, Mccue K, Schaeffer L, & Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5(7):621-628.
10. Rycroft CH (2009) VORO plus plus : A three-dimensional Voronoi cell library in C plus. *Chaos* 19(4).
11. Hopcroft J & Tarjan R (1973) Efficient Algorithms for Graph Manipulation. *Commun Acm* 16(6):372-378.
12. Zhang B & Wolynes PG (2015) Topology, structures, and energy landscapes of human chromosomes. *P Natl Acad Sci USA* 112(19):6062-6067.
13. Pieranski P (1998) In search of ideal knots. *Ideal knots*, eds Stasiak A, Katritch V, & Kauffman LH (World Scientific, Singapore ; River Edge, NJ), pp 20-41.
14. Lua RC (2012) PyKnot: a PyMOL tool for the discovery and analysis of knots in proteins. *Bioinformatics* 28(15):2069-2071.
15. Vologods.Av, Lukashin AV, Frankkam.Md, & Anshelev.Vv (1974) Knot Problem in Statistical-Mechanics of Polymer-Chains. *Zh Eksp Teor Fiz+* 66(6):2153-2163.
16. Kolesov G, Virnau P, Kardar M, & Mirny LA (2007) Protein knot server: detection of knots in protein structures. *Nucleic Acids Res* 35:W425-W428.