De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture

Michele Di Pierro[a,1,2], Ryan R. Cheng[a,1], Erez Lieberman Aiden[a,b], Peter G. Wolynes[a,c,d], and José N. Onuchic[a,d,2]

[a]Center for Theoretical Biological Physics, Rice University, Houston, TX 77005; [b]Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030; [c]Department of Chemistry, Rice University, Houston, TX 77005; and [d]Department of Physics & Astronomy, Rice University, Houston, TX 77005

Contributed by José N. Onuchic, October 4, 2017 (sent for review August 24, 2017; reviewed by Jie Liang and Tamar Schlick)

Inside the cell nucleus, genomes fold into organized structures that are characteristic of cell type. Here, we show that this chromatin architecture can be predicted de novo using epigenetic data derived from chromatin immunoprecipitation-sequencing (ChIP-Seq). We exploit the idea that chromosomes encode a 1D sequence of chromatin structural types. Interactions between these chromatin types determine the 3D structural ensemble of chromosomes through a process similar to phase separation. First, a neural network is used to infer the relation between the epigenetic marks present at a locus, as assayed by ChIP-Seq, and the genomic compartment in which those loci reside, as measured by DNA-DNA proximity ligation (Hi-C). Next, types inferred from this neural network are used as an input to an energy landscape model for chromatin organization [Minimal Chromatin Model (MiChroM)] to generate an ensemble of 3D chromosome conformations at a resolution of 50 kilobases (kb). After training the model, dubbed Maximum Entropy Genomic Annotation from Biomarkers Associated to Structural Ensembles (MEGABASE), on odd-numbered chromosomes, we predict the sequences of chromatin types and the subsequent 3D conformational ensembles for the even chromosomes. We validate these structural ensembles by using ChIP-Seq tracks alone to predict Hi-C maps, as well as distances measured using 3D fluorescence in situ hybridization (FISH) experiments. Both sets of experiments support the hypothesis of phase separation being the driving process behind compartmentalization. These findings strongly suggest that epigenetic marking patterns encode sufficient information to determine the global architecture of chromosomes and that de novo structure prediction for whole genomes may be increasingly possible.

epigenetics | machine learning | energy landscape theory | genomic architecture | Hi-C

In the nucleus of eukaryotic cells, the 1D information of the genome is organized in three dimensions (1, 2). It is increasingly evident that genomic spatial organization is a key element of transcriptional regulation (1, 3, 4). During interphase, the 3D arrangement of chromatin brings into close spatial proximity sections of DNA separated by great genomic distance, introducing interactions between genes and regulatory elements. These folding patterns are cell type-specific (5, 6), and their disruption can lead to disease (7–10).

The use of high-resolution contact mapping experiments (Hi-C) has revealed that, at the large scale, genome structure is dominated by the segregation of human chromatin into compartments. Initial analysis of Hi-C experiments revealed that loci typically exhibited one of two long-range contact patterns, suggesting the presence of two spatial neighborhoods, dubbed the A and B compartments (11). Subsequently, higher resolution experiments have shown the presence of six distinct long-range patterns, indicating the presence of six subcompartments (A1, A2, B1, B2, B3, and B4) in human lymphoblastoid cells (GM12878) (6). The compartmentalization of the genome has been observed in many organisms [including mouse (6, 12) and *Drosophila* (13–15)], and has been confirmed by microscopy experiments (16). Crucially, the long-range contact pattern seen at a locus is cell type-specific, and is strongly associated with particular chromatin marks.

To model this structure, we recently introduced an effective energy landscape model for chromatin structure called the Minimal Chromatin Model (MiChroM) (17). This model combines a generic polymer potential with additional interaction terms governing compartment formation, as well as other processes involved in chromatin organization (4, 18–22) [i.e., the local helical structural tendency of the chromatin filament (17, 23–25) and the chromatin loops associated with the presence of CCCTC-binding factor (CTCF) (6, 26–28)]. The formation of compartments (as well as any other interaction in the MiChroM) is assumed to operate only through direct protein-mediated contacts bringing about segregation of chromatin types through a process of phase separation (17, 29). The MiChroM shows that the compartmentalization patterns that Hi-C maps reveal can be transformed into 3D models of genome structure at 50-kb resolution.

Here, we extend the earlier work by demonstrating that the structure of chromosomes can be predicted, de novo, by inferring chromatin types from chromatin immunoprecipitation-sequencing (ChIP-Seq) data and then using these inferences as

Significance

In the nucleus of eukaryotic cells, the genome is organized in three dimensions in an architecture that depends on cell type. This organization is a key element of transcriptional regulation, and its disruption often leads to disease. We demonstrate that it is possible to predict how a genome will fold based on the epigenetic marks that decorate chromatin. Epigenetic marking patterns are used to predict the corresponding ensemble of 3D structures by leveraging both energy landscape theory and neural network-based machine learning. These predictions are extensively validated by the results of DNA-DNA ligation assays and fluorescence microscopy, which are found to be in exceptionally good agreement with theory.

Author contributions: M.D.P., R.R.C., E.L.A., P.G.W., and J.N.O. designed research; M.D.P. and R.R.C. performed research; M.D.P. and R.R.C. contributed new reagents/analytic tools; M.D.P., R.R.C., E.L.A., P.G.W., and J.N.O. analyzed data; and M.D.P., R.R.C., E.L.A., P.G.W., and J.N.O. wrote the paper.

Reviewers: J.L., University of Illinois at Chicago; and T.S., New York University.

The authors declare no conflict of interest.

[1]M.D.P. and R.R.C. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: michele.dipierro@rice.edu or jonuchic@rice.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1714980114/-/DCSupplemental.

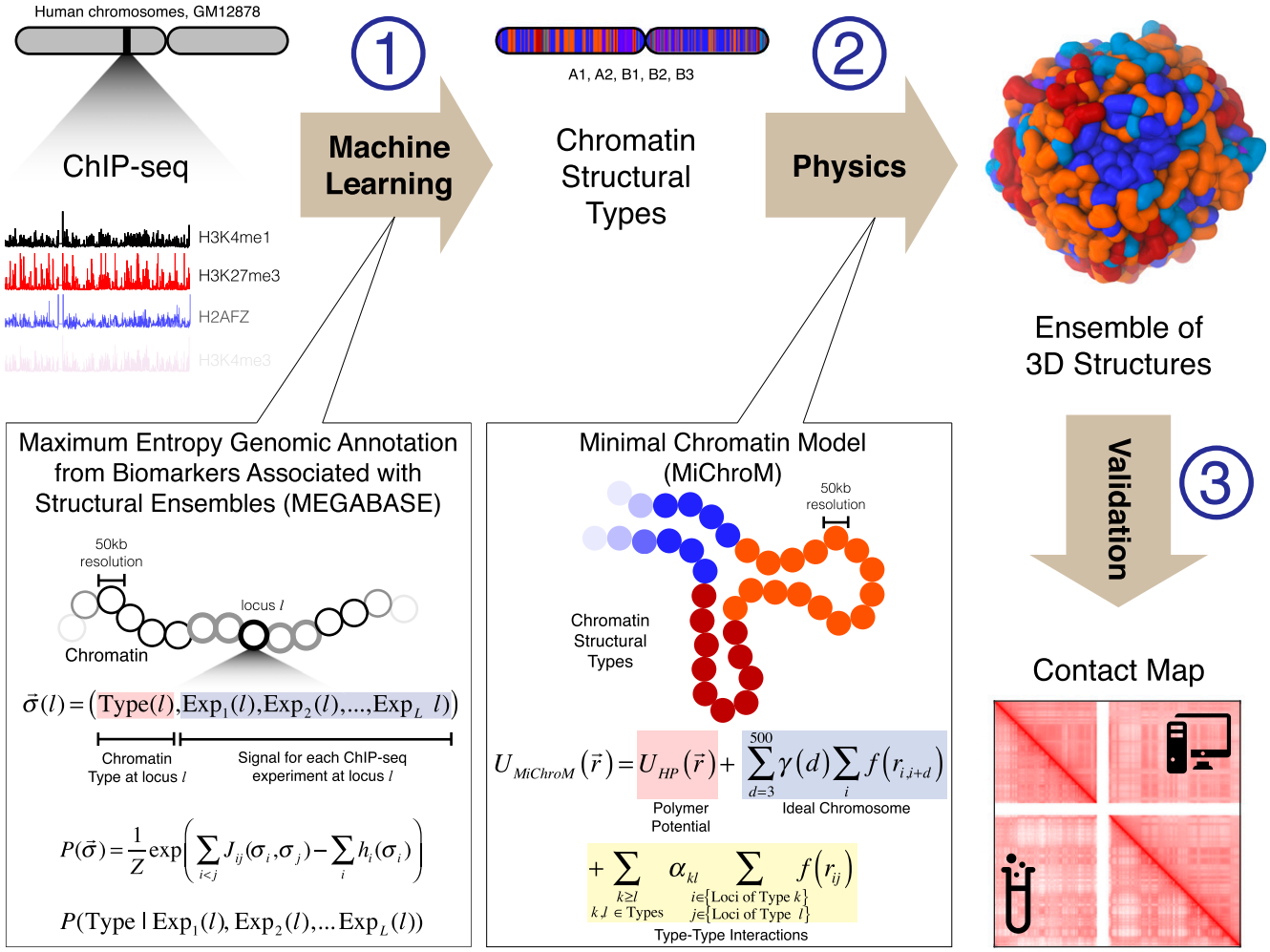

Fig. 1. Schematic illustration of the MEGABASE + MiChroM computational pipeline. (1) ChIP-Seq data constitute the only input to our pipeline. ChIP-Seq tracks obtained from a publicly available resource (ENCODE) are converted into a sequence of chromatin structural types using a neural network dubbed MEGABASE. The neural network encodes the relationship between compartmentalization and the biochemical state of each locus along the genome. (2) Sequences of chromatin structural types are used as input to a physical model for chromatin folding (MiChroM) to obtain the ensembles of 3D structures of specific chromosomes (17). MiChroM is an effective energy landscape model consisting of a generic polymer with chromatin-type interactions and a translational invariant local ordering term (Ideal Chromosome). (3) Ensembles of 3D structures are validated by comparing the predicted contact maps with those experimentally determined by using Hi-C.

an input into an effective energy landscape model. The work flow behind this approach is broadly described in Fig. 1.

Although the compartments and subcompartments visible in Hi-C maps correlate with a handful of specific epigenetic modifications present at those loci (also ref. 6), the distributions of epigenetic markers found in each compartment are broad and largely overlap. It is therefore impossible to assign any given locus correctly to a specific compartment using the frequency of any single epigenetic modification. To overcome this difficulty, we use a machine learning approach to extract information from the raw chromatin immunoprecipitation (ChIP-Seq) data. We first obtained ChIP-Seq profiles available from the Encyclopedia of DNA Elements (ENCODE) project for the GM12878 lymphoblastoid cell line, encompassing 84 protein-binding experiments and 11 histone marks. Next, we discretized each of these profiles, partitioning them into 50-kb loci, each of which is assigned a value from 1 (weakest signal) to 20 (strongest signal). We then constructed a neural network to uncover the relationship between compartment annotations and epigenetic markings. We use a neural network in which each data type available at a given locus corresponds to a single neuron (30). The state of the network is represented by the state vector $\vec{\sigma}(l) = (C(l), \mathrm{Exp}_1(l), \mathrm{Exp}_2(l), \ldots, \mathrm{Exp}_L(l))$, which represents all of the data available at locus l, with C being the subcompartment annotation and Exp_i being the result of the ith ChIP-Seq experiment. The data at each locus are further assumed to be distributed according to a Boltzmann distribution for a Potts model:

$$H(\vec{\sigma}) = -\sum_{i<j} J_{ij}(\sigma_i, \sigma_j) - \sum_i h_i(\sigma_i),$$

$$P(\vec{\sigma}) = \frac{1}{z}\exp(-H(\vec{\sigma})),$$

where the $P(\vec{\sigma})$ indicates the probability of observing the state vector $\vec{\sigma}$ at any given locus l, the J_{ij} interactions capture local pairwise correlations between epigenetic marks or between marks and chromatin types, and h_i determines the individual frequencies of chromatin types and markers. This procedure is equivalent to training a Boltzmann machine to encode the information contained in the dataset. The learning strategy is based on the idea that the parameters of the neural network should

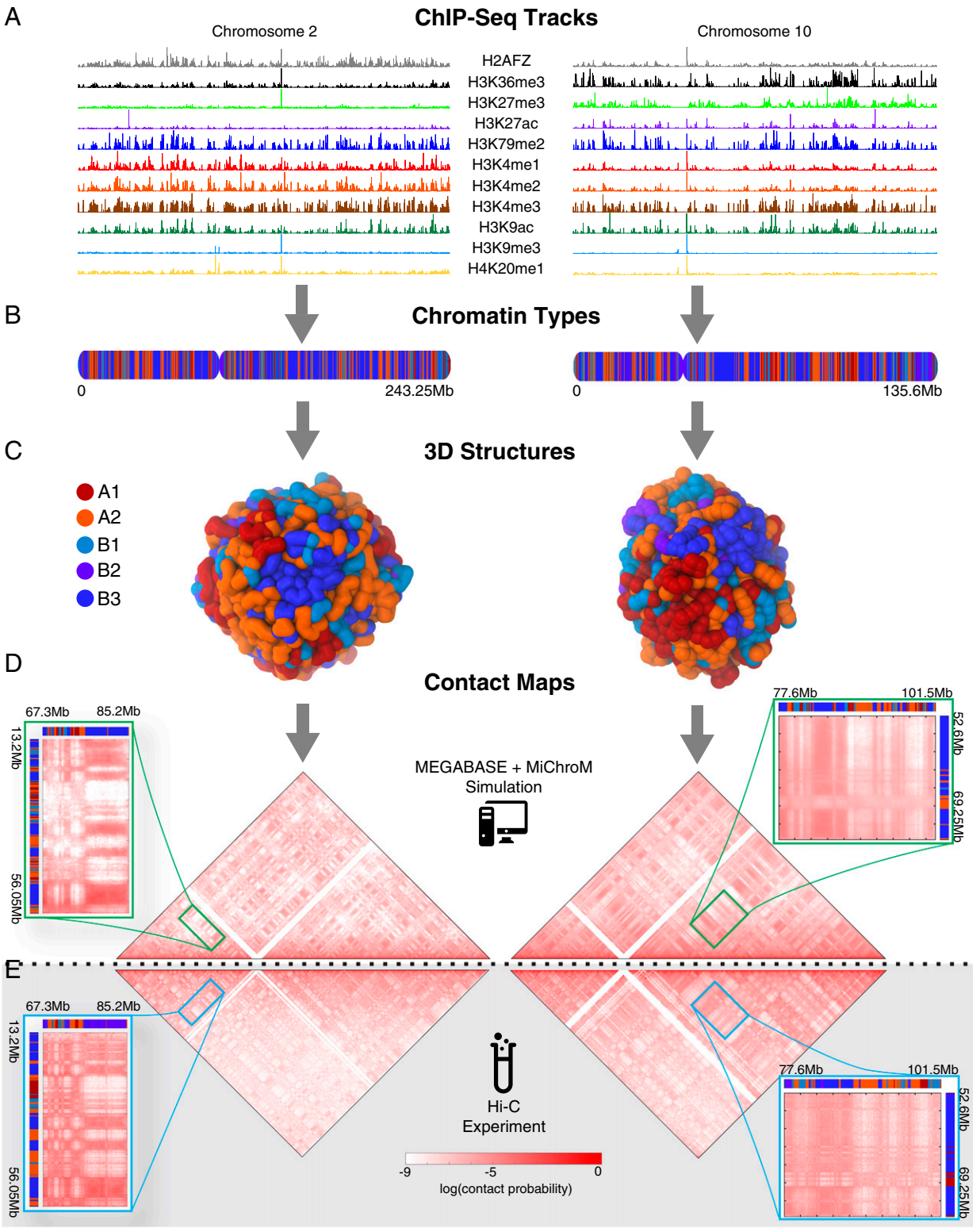

Fig. 2. Predicting the 1D chromatin sequences, 3D conformations, and 2D contact probabilities of human chromosomes from epigenetic marking patterns. We apply MEGABASE + MiChroM to obtain an ensemble of 3D structures for all of the autosomes of cell line GM12878. For illustrative purposes, predictions for chromosome 2 (*Left*) and chromosome 10 (*Right*) are shown, respectively. (*A*) Ninety-five ChIP-Seq tracks are downloaded from the ENCODE database and used as input for MEGABASE to predict 1D sequences of chromatin types (shown in *B*). The 3D structure of each chromosome is encoded in its specific 1D sequence of chromatin structural types. (*C*) Typical 3D conformation obtained by MiChroM is shown for chromosomes 2 and 10. (*D*) Approximately 50,000 structures are collected from simulation to generate high-quality contact maps. These contact maps are compared with the Hi-C maps shown in *E*. The simulations correctly predict the long-range contact probability patterns that are observed in Hi-C maps, as seen in the magnified regions.

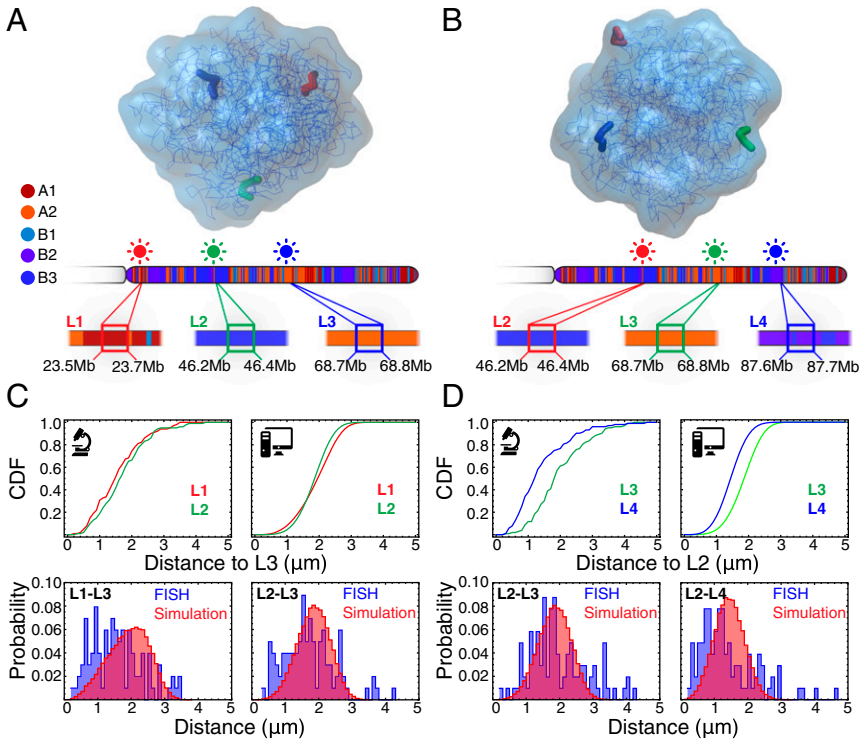

Fig. 3. Simulated conformational ensembles predict the distances measured by 3D FISH experiments. Simulations and 3D FISH experiments support the idea that the compartmentalization observed in Hi-C maps emerges from the phase separation of chromatin structural types. (*A* and *B*) Cartesian distances between four loci (L1, L2, L3, and L4) in chromosome 14 (cell line GM06990) were measured in two distinct 3D FISH experiments reported by Lieberman-Aiden et al. (11). The same distances were measured using the MEGABASE + MiChroM pipeline. The positions of the fluorescent probes are illustrated in representative 3D configurations from simulations, as well as along the chromosome. As illustrated by the annotations from MEGABASE shown in the figure, the four loci are composed of chromatin of alternating types: L1 and L3 composed of type A chromatin and L2 and L4 composed of type B chromatin. (*C* and *D*) Cumulative distribution functions (CDF) show that loci composed of chromatin belonging to the same type tend to be closer in space than otherwise, despite the interlaced order and despite lying at greater genomic distances. This phenomenon is observed in FISH experiments, and it is correctly predicted by our ChIP-Seq–based modeling. The comparison between the predicted and measured probability distributions shows excellent agreement for both the average distance and the distance fluctuations (more examples of validation with FISH data are provided in *SI Appendix*). The average ratio between simulated distances and FISH-measured distances has been used to calibrate the length scale of simulation. One unit of length in simulation corresponded to a length of 0.17 μm, which also implies the size of a simulated chromosomal territory being ∼2–3 μm across, which is consistent with what was previously reported by Cremer and Cremer (2).

maximize the likelihood of observing the set of state vectors representing a particular training set. A similar strategy has been previously introduced to quantify the correlated mutational patterns observed in amino acid sequence data of protein families occurring under natural selection to aid protein structure prediction (31, 32).

The quality of compartment prediction is improved when we include in the Potts model interactions that do not just refer to a single 50-kb locus but also to interactions encoding correlations between markings and annotations of nearest neighbors and next nearest neighbors (i.e., the neural network correlates information from loci $l-2, l-1, l, l+1, l+2$). Through these couplings, the probability of observing a specific state vector at a given locus is correlated with the states of the adjacent segments, thus minimizing the effect of uncorrelated noise. This strategy is analogous to the construction of secondary structure predictors in protein folding using helix–coil models (33).

The inferred probabilistic model is then marginalized to predict the most probable chromatin type for a given locus l when given the experimental ChIP-Seq measurements of loci $(l-2,\ l-1,\ l,\ l+1,\ l+2)$:

$$CST(l) = \arg\max P\left(C|\mathrm{Exp}_{1,\ldots,L}(l-2,l-1,l,l+1,l+2)\right).$$

We refer to the resulting probabilistic predictor of chromatin structural types (CST) as the Maximum Entropy Genomic Annotation from Biomarkers Associated to Structural Ensembles (MEGABASE). Once trained for a given new input sequence of epigenetic marks, the model can then find the most probable sequence of corresponding compartment annotations.

The state vectors of every locus of the odd-numbered chromosomes comprise the training set. The state vectors of the even-numbered chromosomes then provide a test set to quantify the performance of the trained model.

After training on the odd-numbered chromosomes, we used our statistical model to predict the chromatin types for the independent set of the even chromosomes of the cell line GM12878 from their epigenetic marking profiles. For the test set, the predicted type assignments are in broad agreement with the experimentally determined structural annotations in the study by Rao et al. (6). Specifically, the model is very accurate in predicting the assignments to compartments (A vs. B), while producing a larger number of mismatches between the predicted chromatin types and the published subcompartment annotations, which are more fine-grained (A1 vs. A2, B1 vs. B2 vs. B3) (*SI Appendix*, Fig. S1).

Once predicted sequences of type annotations are available, we use our earlier MiChroM to sample the predicted conformational ensembles of 3D structures. To highlight the relationship between chromatin types and compartmentalization, we use the MiChroM Hamiltonian with the same parameters that had already been determined, but omit the term in that

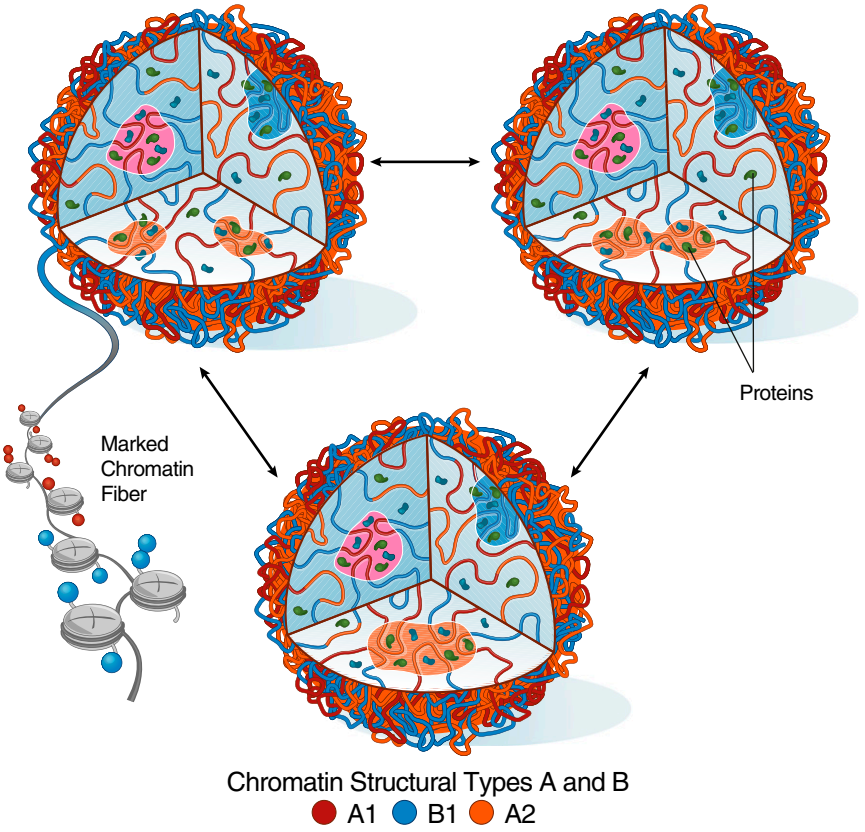

Fig. 4. Process of microphase separation explains compartmentalization in chromosomes. The MEGABASE + MiChroM hypothesizes that chromatin characterized by homogeneous epigenetic markings undergoes a process similar to phase separation under the action of the proteome present in the nucleus. In simulations, we observe that segments of chromatin belonging to the same structural type tend to segregate, forming liquid droplets, which rearrange dynamically by splitting and fusing. This simple process of phase separation is sufficient to explain the emergence of compartmentalization in genomes as observed in DNA-DNA ligation assays and microscopy experiments.

energy function that models the CTCF-mediated looping interactions. These looping interactions seem to arise from a distinct process from compartmentalization, and omitting such interactions does not disrupt the large-scale architecture of chromosomes (17) (the results of additional simulations, including also the CTCF-mediated looping interactions, are provided in *SI Appendix*, Fig. S2).

The simulations all start from a random collapsed polymer having the proper length confined in a spherical region at correct density (*SI Appendix*). After equilibration, we collect an ensemble of 3D structures representing the chromosome-specific energy landscape as shaped by the inferred chromatin-type sequences (used as input) and by the MiChroM effective interactions.

From the ensemble of equilibrium conformations, we calculate the contact probabilities between any pair of loci within each chromosome. We compare the resulting contact maps from the simulated ensemble of 3D structures with the experimental Hi-C maps reported by Rao et al. (6). The overall agreement between the experimental and simulated contact probabilities is visually evident. The comparison between the simulated and experimental contact maps is shown in Fig. 2 for representative chromosomes in the test set (i.e., the even autosomes). The Pearson's coefficient is ∼0.9 or higher for all of the chromosomes whether in the training set or test set, and the analysis of the Pearson's coefficient as a function of genomic distance (*SI Appendix*, Figs. S3–S24) confirms that the two sets of maps are correlated exceptionally well. The power law scaling of the contact probability between two loci as a function of their genomic distance is reproduced well at all genomic distances in a comparison with Hi-C data (*SI Appendix*, Figs. S3–S24).

Finally, we compare the Cartesian distances between multiple pairs of loci as predicted through the use of our computational model with those measured by using 3D fluorescence in situ hybridization (FISH), and reported by Rao et al. (6) for the cell line GM12878 and by Lieberman-Aiden et al. (11) for the closely related cell line GM06990. FISH experiments in Fig. 3 show that chromatin belonging to the same structural type tends to come into contact more frequently than otherwise, supporting the idea that compartmentalization is induced by a process of phase separation. This behavior is predicted with quantitative accuracy by our ChIP-Seq–based simulation. Remarkably, simulations predict all of the experimentally determined average distances, together with their variances (Fig. 3 and *SI Appendix*, Figs. S25–S27).

Representative predicted 3D conformations for chromosome 2 and chromosome 10 are shown in Fig. 2.

As previously observed by Di Pierro et al. (17), analysis of the conformational ensembles shows the existence of microphase separation between chromatin of different types, leading to the formation of the characteristic patterns of interactions seen in Hi-C maps. Examples of the long-range patterns that are captured by our predictions are shown in Fig. 2. The more transcriptionally active segments of chromatin (compartments A1 and A2 in Fig. 2) are more frequently found on the outer surface, while the inactive segments (compartments B1, B2, and B3 in Fig. 2) typically reside in the core of chromosomes.

The quality of the structural predictions achieved using the chromatin annotation inferred by MEGABASE shows that there exists a clear sequence-to-structure relationship between the sequences of chromatin types predicted from epigenetic marks and genome architecture. The accuracy achieved by using our energy landscape model in predicting the effects of compartmentalization, as seen by Hi-C and 3D FISH, supports the plausibility of microphase separation being the physical process driving compartmentalization in chromosomes (17, 34–36) (Fig. 4).

The success achieved in reliably predicting chromosome architecture indicates that our probabilistic model captures the essential features of epigenetic marks that are associated with compartmentalization. Hence, we further exploit MEGABASE to study this relationship by calculating the content of mutual information shared between markers and compartments, and so quantifying which of the markers are the best predictors of compartmentalization. It is immediately evident that certain biochemical markers share a high content of mutual information with chromatin structural types, while others do not. According to our neural network, histone methylations HK36me3, H3K27me3, H3K4me1, and H4K20me1 and nuclear proteins EED, ZBED1, TRIM22, and HCFC1 carry most of the information associated with identifying the chromatin types (*SI Appendix*, Fig. S28). In contrast, we see that although compartment A, for example, has a very high content of H3K27ac, that marker by itself is a poor predictor owing to its modest mutual information value.

Histone modifications alone carry enough information to predict genome architecture. To illustrate the disproportionate predictive value of histone marks, we created a reduced model by training MEGABASE using only the 11 patterns of histone modifications out of the 95 tracks available in the ENCODE database. The sequences of chromatin types predicted by this reduced model turn out to be only marginally different from those obtained by the full dataset of ChIP-Seq tracks (*SI Appendix*).

Our results demonstrate clearly that it is possible to generate de novo predictions of the genome's 3D structure, as well as specific predictions about the results of Hi-C and FISH experiments, using only ChIP-Seq data on histone modifications as an input. The faithfulness of the predicted conformational ensembles underlines the existence of a sequence-to-structure relationship between patterns of histone modifications and the 3D spatial arrangement of chromosomes.

These findings offer great hope that, like the problem of protein folding before it, the puzzle of genome folding may be amenable to computational predictions (37). However, despite the success of the neural network-based prediction algorithm, the details of the mechanism underlying chromatin folding remain unclear. Does chromatin fold into a specific conformation because of the particular sequence of epigenetic markers or, vice versa, do compartments share similar epigenetic markers because of chromosome architecture? Dynamical studies using Hi-C and other methods will doubtless be essential in addressing these questions.

ACKNOWLEDGMENTS. We thank Erica J. Di Pierro for help in editing the manuscript. This work was supported by the Center for Theoretical Biological Physics sponsored by National Science Foundation (NSF) Grant PHY-1427654. J.N.O. was also supported by the NSF Grant CHE-1614101 and by the Welch Foundation (Grant C-1792). Additional support to P.G.W. was provided by the D. R. Bullard-Welch Chair at Rice University (Grant C-0016). E.L.A. was also supported by an NIH New Innovator Award (1DP2OD008540-01), the National Human Genome Research Institute (NHGRI) Center for Excellence for Genomic Sciences (HG006193), the Welch Foundation (Q-1866), an NVIDIA Research Center Award, an International Business Machines Corporation (IBM) University Challenge Award, a Google Research Award, a Cancer Prevention Research Institute of Texas Scholar Award (R1304), a McNair Medical Institute Scholar Award, an NIH 4D Nucleome Grant (U01HL130010), an NIH Encyclopedia of DNA Elements Mapping Center Award (UM1HG009375), and the President's Early Career Award in Science and Engineering.

1. Bickmore WA (2013) The spatial organization of the human genome. *Annu Rev Genomics Hum Genet* 14:67–84.
2. Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2:292–301.
3. Whalen S, Truty RM, Pollard KS (2016) Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet* 48:488–496.
4. Gürsoy G, Xu Y, Liang J (2017) Spatial organization of the budding yeast genome in the cell nucleus and identification of specific chromatin interactions from multi-chromosome constrained chromatin model. *PLoS Comput Biol* 13:e1005658.
5. Krijger PH, et al. (2016) Cell-of-origin-specific 3D genome structure acquired during somatic cell reprogramming. *Cell Stem Cell* 18:597–610.
6. Rao SSP, et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159:1665–1680.
7. Göndör A (2013) Dynamic chromatin loops bridge health and disease in the nuclear landscape. *Semin Cancer Biol* 23:90–98.
8. Krijger PH, de Laat W (2016) Regulation of disease-associated gene expression in the 3D genome. *Nat Rev Mol Cell Biol* 17:771–782.
9. Fullwood MJ, et al. (2009) An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* 462:58–64.
10. Montefiori L, et al. (2016) Extremely long-range chromatin loops link topological domains to facilitate a diverse antibody repertoire. *Cell Rep* 14:896–906.
11. Lieberman-Aiden E, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326:289–293.
12. Dixon JR, et al. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485:376–380.
13. Eagen KP, Hartl TA, Kornberg RD (2015) Stable chromosome condensation revealed by chromosome conformation capture. *Cell* 163:934–946.
14. Sexton T, et al. (2012) Three-dimensional folding and functional organization principles of the Drosophila genome. *Cell* 148:458–472.
15. Li QJ, et al. (2017) The three-dimensional genome organization of *Drosophila melanogaster* through data integration. *Genome Biol* 18:145.
16. Wang S, et al. (2016) Spatial organization of chromatin domains and compartments in single chromosomes. *Science* 353:598–602.
17. Di Pierro M, Zhang B, Aiden EL, Wolynes PG, Onuchic JN (2016) Transferable model for chromosome architecture. *Proc Natl Acad Sci USA* 113:12168–12173.
18. Barbieri M, et al. (2012) Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci USA* 109:16173–16178.
19. Brackley CA, Johnson J, Kelly S, Cook PR, Marenduzzo D (2016) Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucleic Acids Res* 44:3503–3512.
20. Jost D, Carrivain P, Cavalli G, Vaillant C (2014) Modeling epigenome folding: Formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res* 42:9553–9561.
21. Wong H, et al. (2012) A predictive computational model of the dynamic 3D interphase yeast nucleus. *Curr Biol* 22:1881–1890.
22. Tjong H, Gong K, Chen L, Alber F (2012) Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Res* 22:1295–1305.
23. Zhang B, Wolynes PG (2016) Shape transitions and chiral symmetry breaking in the energy landscape of the mitotic chromosome. *Phys Rev Lett* 116:248101.
24. Zhang B, Wolynes PG (2015) Topology, structures, and energy landscapes of human chromosomes. *Proc Natl Acad Sci USA* 112:6062–6067.
25. Grigoryev SA, et al. (2016) Hierarchical looping of zigzag nucleosome chains in metaphase chromosomes. *Proc Natl Acad Sci USA* 113:1238–1243.
26. Sanborn AL, et al. (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc Natl Acad Sci USA* 112:E6456–E6465.
27. Phillips JE, Corces VG (2009) CTCF: Master weaver of the genome. *Cell* 137:1194–1211.
28. Nichols MH, Corces VG (2015) A CTCF code for 3D genome architecture. *Cell* 162:703–705.
29. Zhang B, Wolynes PG (2017) Genomic energy landscapes. *Biophys J* 112:427–433.
30. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. *Proc Natl Acad Sci USA* 79:2554–2558.
31. Lapedes A, Giraud B, Jarzynski C (2002) Using sequence alignments to predict protein structure and stability with high accuracy. arXiv:1207.2484.
32. Ekeberg M, Lövkvist C, Lan Y, Weigt M, Aurell E (2013) Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models. *Phys Rev E Stat Nonlin Soft Matter Phys* 87:012707.
33. Bryngelson JD, Hopfield JJ, Southard SN (1990) A protein structure predictor based on an energy model with learned parameters. *Tetrahedron Comput Methodol* 3:129–141.
34. Hnisz D, Shrinivas K, Young RA, Chakraborty AK, Sharp PA (2017) A phase separation model for transcriptional control. *Cell* 169:13–23.
35. Larson AG, et al. (2017) Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* 547:236–240.
36. Strom AR, et al. (2017) Phase separation drives heterochromatin domain formation. *Nature* 547:241–245.
37. Wolynes PG (2015) Evolution, energy landscapes and the paradoxes of protein folding. *Biochimie* 119:218–230.