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

Michele Di Pierro[1*], Ryan R. Cheng[1*], Erez Lieberman Aiden[1,2], Peter G. Wolynes[1,3,4], and José N. Onuchic[1,4]

Short Abstract — **We show that this chromatin architecture can be predicted *de novo* using epigenetic data derived from ChIP-Seq. We exploit the idea that chromosomes encode a 1D sequence of chromatin types, analogous to a sequence of amino acids for a protein. A neural network is used to infer the relation between the epigenetic marks present at a locus and the genomic compartment in which those loci reside. The sequence of types inferred from this neural network is used as an input to an energy landscape model for chromatin organization (MiChroM) in order to generate an ensemble of 3D chromosome conformations.**

Keywords **— Epigenetics, Genome Architecture, Machine Learning, Energy Landscape Theory, Hi-C, FISH**

I. PURPOSE

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. Analysis of Hi-C experiments revealed that loci exhibit at least six long-range contact patterns, indicating the presence of at least six sub-compartments (A1, A2, B1, B2, B3, and B4) in human lymphoblastoid cells (GM12878) [1]. Further, the long-range contact pattern seen at a locus is cell-type specific, and is strongly associated with particular chromatin marks.

To model chromosome structure, an effective energy landscape model for chromatin structure called the Minimal Chromatin Model (MiChroM) was previously introduced [2]. This model combines a generic polymer potential with additional interaction terms governing compartment formation, the local helical structural tendency of the chromatin filament, and the chromatin loops associated with the presence of CCCTC-binding factor (CTCF). The formation of compartments (as well as any other interaction in MiChroM) is assumed to operate only through direct protein-mediated contacts bringing about segregation of chromatin types through a process of phase separation. MiChroM shows that the compartmentalization patterns that Hi-C maps reveal can be transformed into 3D models of genome structure at 50 kb resolution.

We extend upon the earlier work [3] by demonstrating that the structure of chromosomes can be predicted, *de novo*, by inferring chromatin types from ChIP-Seq data and then using these inferences as an input into MiChroM.

We first obtained ChIP-Seq profiles available from the ENCODE project for the GM12878 lymphoblastoid cell line, encompassing protein-binding experiments and histone marks. We then constructed a neural network to uncover the relationship between compartment annotations and epigenetic markings. This neural network allowed us to predict the chromatin type of a locus, provided biochemical data for that locus.

The predicted sequence of chromatin types for a chromosome then serves as direct input for molecular dynamics simulations using the MiChroM potential, which generates an ensemble of 3D structures. The de novo prediction of chromosome architecture for human lymphoblastoid cells was extensively validated against DNA-DNA ligation and fluorescence in situ hybridization data, demonstrating that there is sufficient information encoded in the biochemical data to accurately predict chromosomal structures. The broad agreement between theory and experiment point to the existence of a sequence-to-structure relationship between epigenetic modifications and chromosomal structure.

Furthermore, since the MEGABASE annotation is made from biochemical data alone, it supports the idea that phase separation of distinct chromatin types is carried out by proteins and regulated by epigenetics.

REFERENCES

[1] 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.

[2] Di Pierro M, Zhang B, Aiden EL, Wolynes PG, & Onuchic JN (2016) Transferable model for chromosome architecture. *P Natl Acad Sci USA* 113(43):12168-12173.

[3] Di Pierro M, Cheng RR, Aiden EL, Wolynes PG, & Onuchic JN (2017) De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *P Natl Acad Sci USA* 114(46):12126-12131.

[1] Center for Theoretical Biological Physics, Rice University, Houston, Texas 77005.
[2] The Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030.
[3] Department of Chemistry, Rice University, Houston, Texas 77005.
[4] Department of Physics & Astronomy, Rice University, Houston, Texas 77005.
[*] These authors contributed equally.

Nothing should be here on page 2! Please limit your abstract to a single page, and create a one-page .pdf file for submission.