



The three-dimensional architecture of the human genome: understanding the physical mechanisms controlling gene expression

> FAPESP School on Exact, Natural and Life Sciences 7-10 August 2022

Peter Wolynes



Michele Di Pierro



Olga Dudchenko



Ryan Cheng



Erez Aiden



Vinicius Contessoto



Sumitabha Brahmachari



Antonio B. De Oliveira Jr.

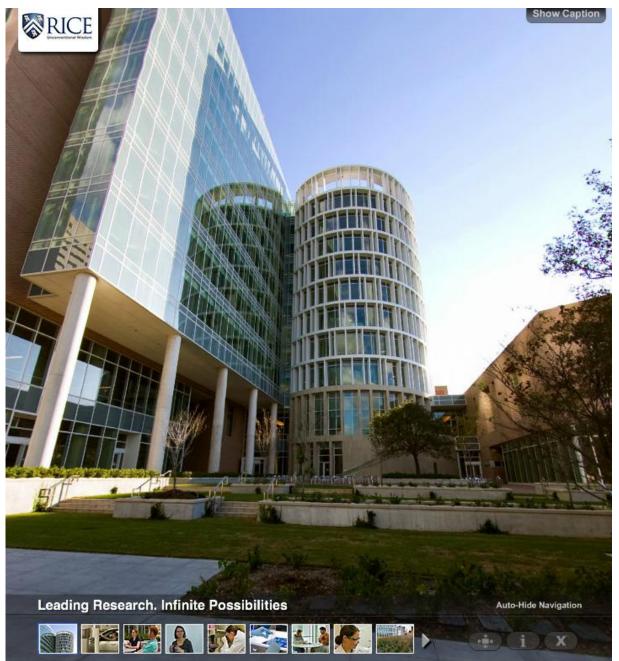


José N. Onuchic Center for Theoretical Biological Physics CTBP Rice University ctbp.rice.edu

Center for Theoretical Biological Physics

10th floor of the BioScience Research Collaborative





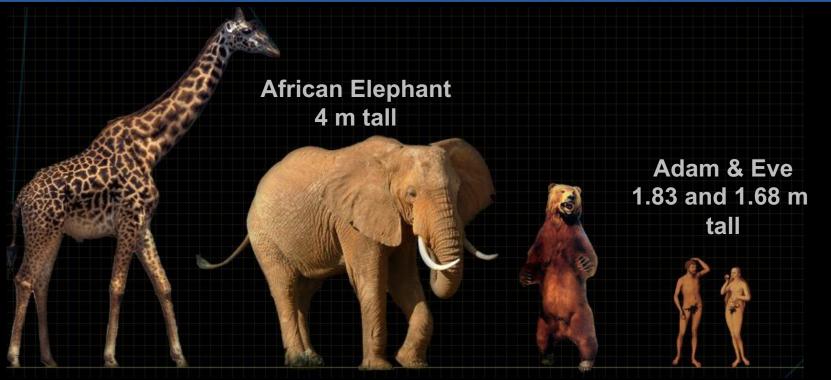
5/22/12 6:19 AM

VIRTUAL TOUR PRODUCED BY VIRTUALLY ANYWHERE



Preface: A Short Survey of Length Scales





Giraffe 6.1 m tall

Kodiak Bear 3 m tall

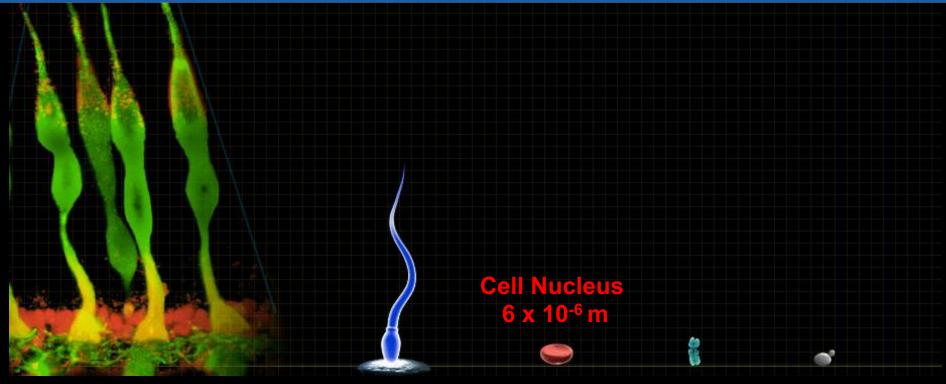
Our Genome 2 m long





Preface: A Short Survey of Length Scales





Photoreceptors 1 x 10⁻⁴ m Sperm 6 x 10⁻⁵ m Metaphase X Chromosome 7 x 10⁻⁶ m

Yeast 3 x 10⁻⁶ m







DNA must be open and knots free



Very little accessible tape!



Knots don't help reading!







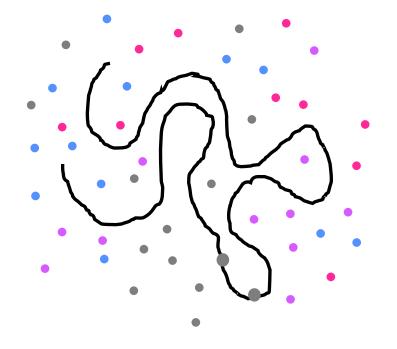






A Polymer In A Complex Environment



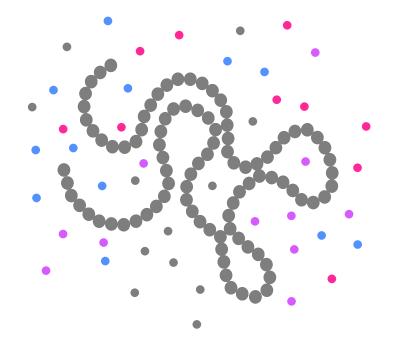






A Complex Polymer In A Complex Environment



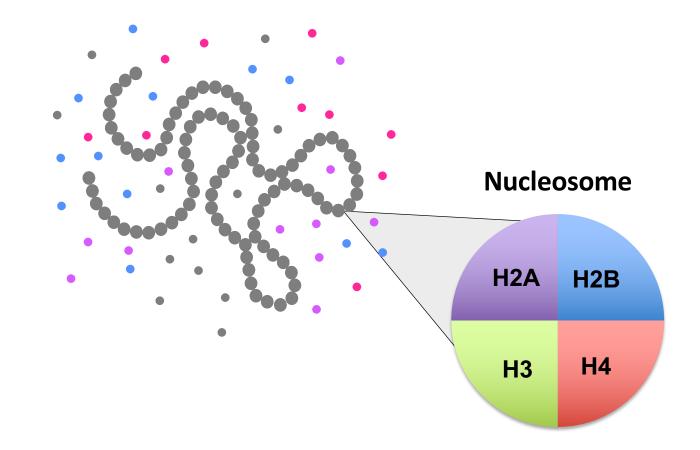






More Than A Polymer: Epigenetics



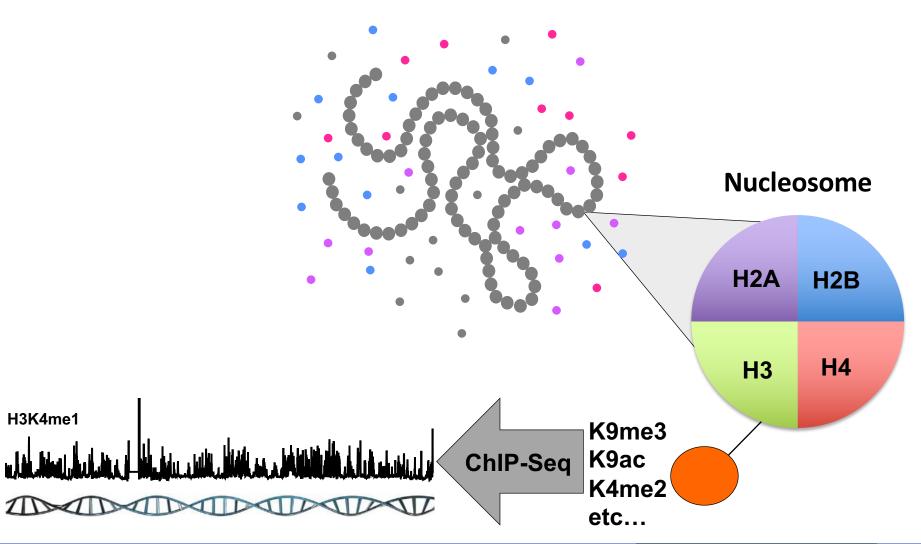






More Than A Polymer: Epigenetics



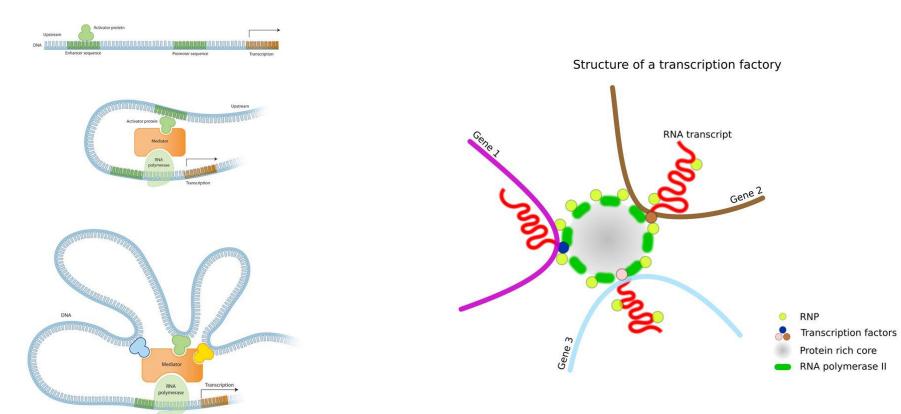








Enhancers



Transcription Factories

© Nature Education

Rieder et al. 2012







Fluorescence *In Situ* Hybridization (FISH)



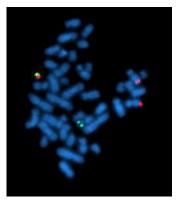
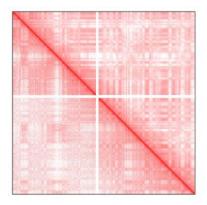


Image from Wikipedia

Chromosome Conformation Capture (3C, 4C, 5C, Hi-C)

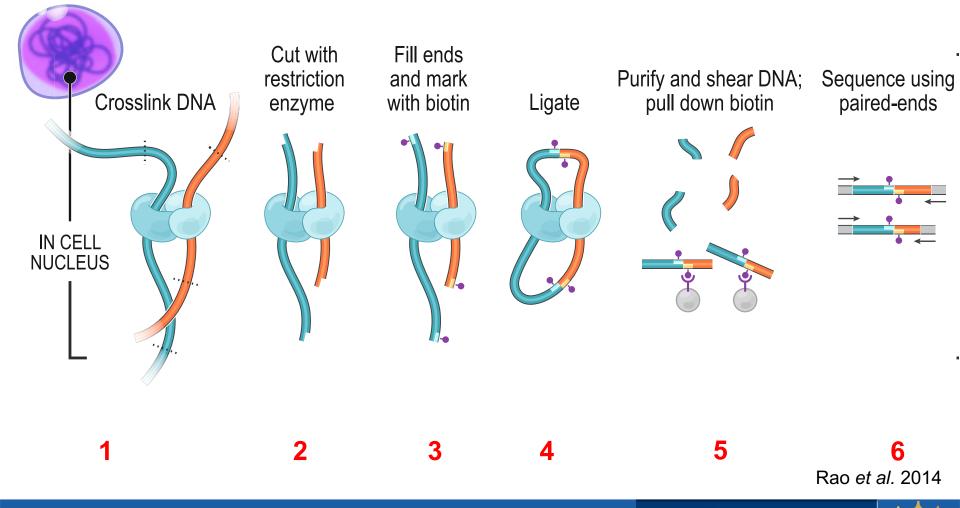








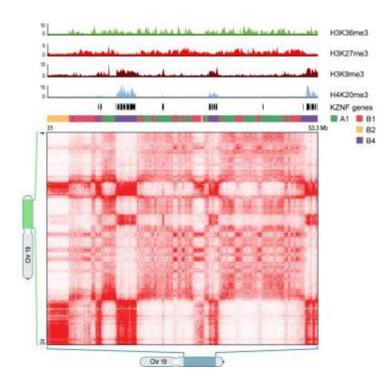


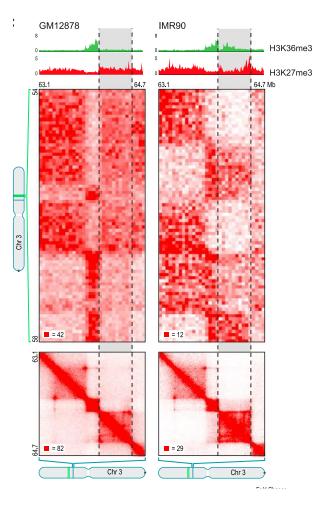




Chromatin Architecture is Specific to the Cell Type





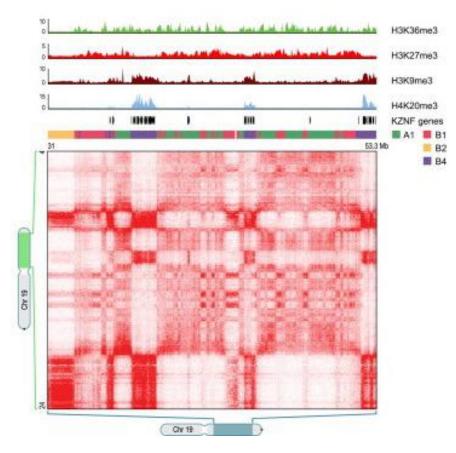


Lieberman Aiden Lab



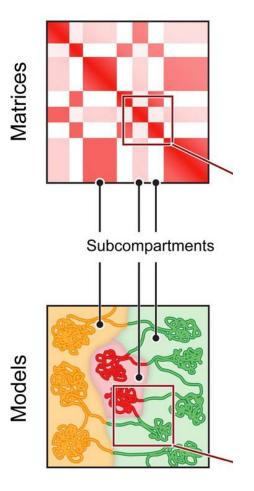


DNA-DNA Ligation Experiments Reveals Compartments which appear to correlate with epigenetic modifications

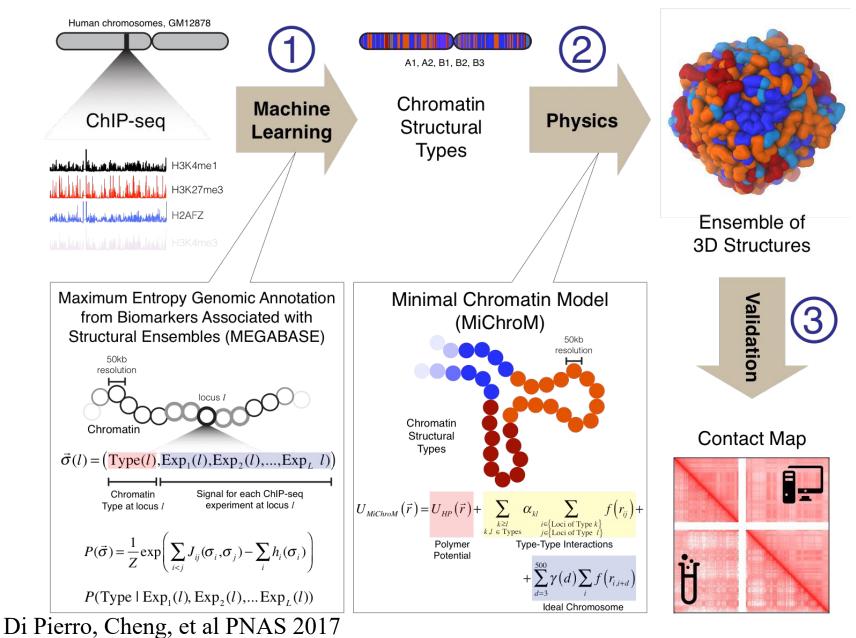


Patterns in contact probabilities identify 6 chromatin types (Sub-compartments): A1, A2, B1, B2, B3, B4

Rao & Huntley et al, Cell (2014)

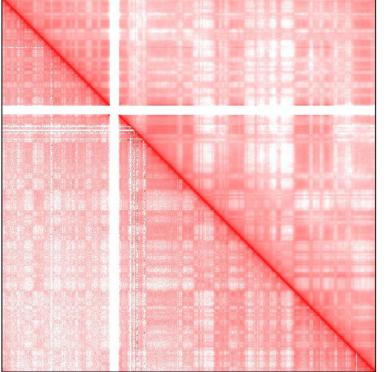


Schematic Illustration of Computational Pipeline

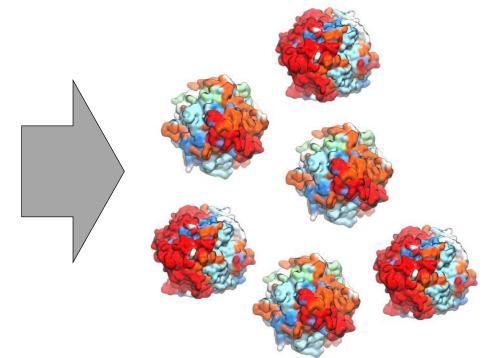




DNA-DNA ligation Assays: Hi-C



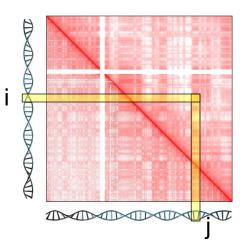
Structural Ensemble











1 Chromosome, 1 Conformation, 1 Experiment



Did *i* and *j* crosslink?

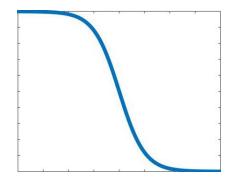
yes/no

How often did *i* and *j* crosslink?

• A lot when they are close, not much when they are far.



$$f_{ij} = f(r_{ij})$$
$$p_{ij} = \left\langle f(r_{ij}) \right\rangle$$





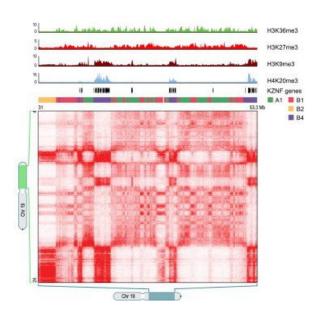


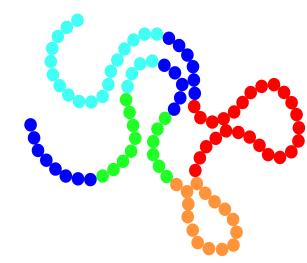
The Minimal Chromatin Model (MiChroM): Phase Separation



Fact 1. Phase Separation of Chromatin Structural Types

If and when two segments of chromatin form a contact the energy of the contact depends only on the type identity of the contact.

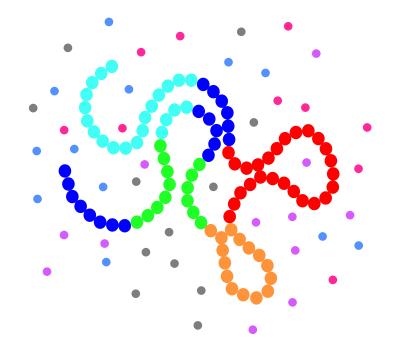






The Minimal Chromatin Model: Some Motivated Physical Assumptions





Implicit Protein Model

Type-to-Type Interactions are mediated by a cloud of proteins that bind with different selectivity to different sections of chromatin







Fact #2:

Chromatin form loops at specific locations related to the activity of the protein CTCF

Phillips and Corces 2009 Rao *et al.* 2014

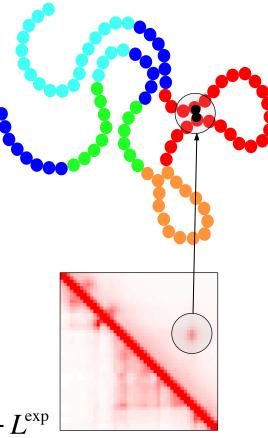
Physical Assumption #2:

If and when the two ends of a loop come into contact, an there is an additional gain in effective free energy

Observable

Constraint

$$L(\vec{r}) = \sum_{(i,j)\in \{\text{Loops Sites}\}} f(r_{ij}) \qquad c_L = \int L(\vec{r}) \pi^{\text{MiChroM}}(\vec{r}) d\vec{r} - L^{\exp}$$







The Minimal Chromatin Model (MiChroM): Ideal Chromosome



Fact #3: The effect of all protein motors acting along the DNA polymer

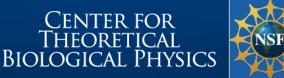
Ideal Chromosome/Lengthwise Compaction. Every time two loci at genomic distance d come into contact there is a gain/loss of $\gamma(d)$ effective free energy.

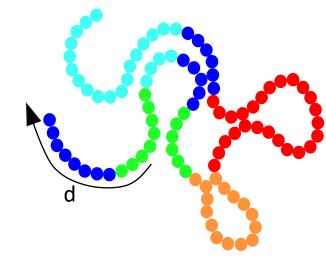
A translational invariant function of the genomic distance consistent with

- \circ the notion of a higher order fiber in chromatin
- liquid crystal behavior

But more general...

M.D.P., B. Zhang, E. Lieberman Aiden, P. G. Wolynes, and J. N. Onuchic, *PNAS 2016* B. Zhang and P. G. Wolynes, *PNAS 2015*



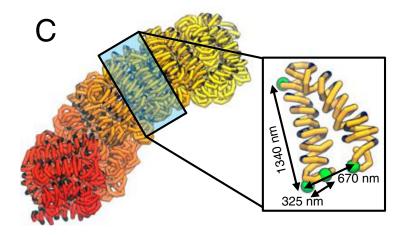


The Ideal Chromosome



- A translational invariant function
 of the genomic distance
- Consistent with
 - the notion of a higher order fiber in chromatin
 - liquid crystal behavior

But more general...



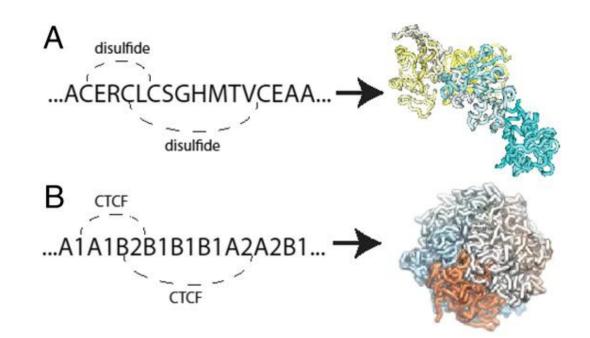
Zhang and Wolynes 2015



The Minimal Chromatin Model (MiChroM): Phase Separation



A one-dimensional sequence encodes the three-dimensional fold of chromosomes



with



M.D.P., B. Zhang, E. Lieberman Aiden, P. G. Wolynes, and J. N. Onuchic, *PNAS* 2016 + Commentary by G. Gürsoy and J. Liang, *PNAS* 2016



The Minimal Chromatin Model: 3 Simple Physical Assumptions



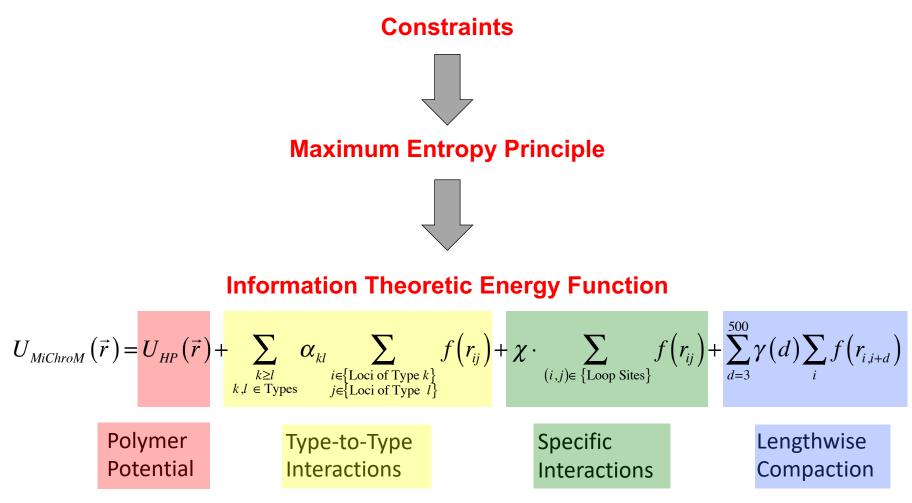
Physical Assumption	Observable	Constraint
 #1 If and when two segments of chromatin form a contact the energy of the contact depends only on the type identity of the contact. Rao et al., 2014, Fillion et al., 2010 	$T_{kl}(\vec{r}) = \sum_{\substack{i \in \{\text{Loci of Type } k\}\\j \in \{\text{Loci of Type } l\}}} f(r_{ij})$	$c_T^{kl} = \int T_{kl}(\vec{r}) \pi^{\text{MiChroM}}(\vec{r}) d\vec{r} - T_{kl}^{\text{exp}}$
#2 If and when the two anchors of a CTCF-mediated loop come into contact, there is an additional gain in effective free energy	$L(\vec{r}) = \sum_{(i,j)\in \{\text{Loops Sites}\}} f(r_{ij})$	$c_L = \int L(\vec{r}) \pi^{\text{MiChroM}}(\vec{r}) d\vec{r} - L^{\text{exp}}$
Phillips and Corces, 2009, Rao <i>et al.,</i> 2014		
#3 Ideal Chromosome/Local Compaction. Every time two loci at genomic distance d come into contact there is a gain/loss of $\gamma(d)$ effective free energy.	$G_d(\vec{r}) = \sum_i f(r_{i,i+d})$	$c_G^d = \int G_d(\vec{r}) \pi^{\text{MiChroM}}(\vec{r}) d\vec{r} - G_d^{\text{exp}}$





The Minimal Chromatin Model (MiChroM): the Effective Energy Function

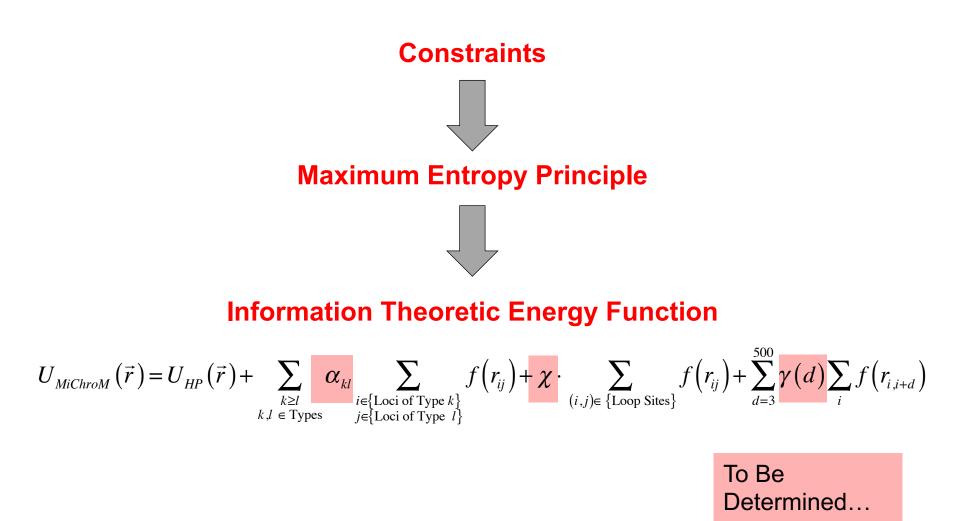




M.D.P., B. Zhang, E. Lieberman Aiden, P. G. Wolynes, and J. N. Onuchic, *PNAS 2016* + Commentary by G. Gürsoy and J. Liang, *PNAS 2016*







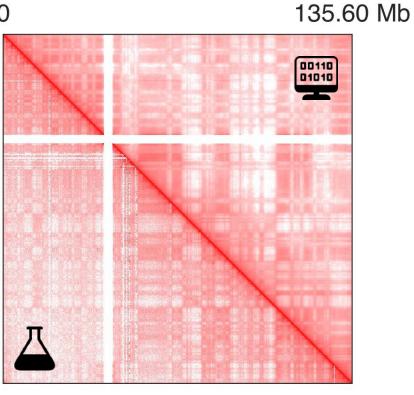


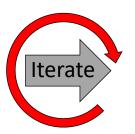


Human Chromosome 10 B-lymphoblastoid cells (GM12878)

0

135.60 Mb





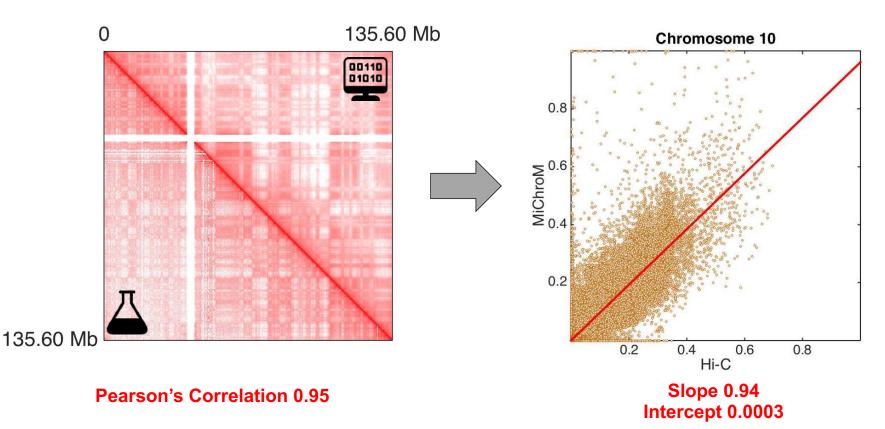
 $\left(oldsymbol{lpha}_{_{kl}} \;, \gamma(d) \;, oldsymbol{\chi}
ight)$







Human Chromosome 10 B-lymphoblastoid cells (GM12878)

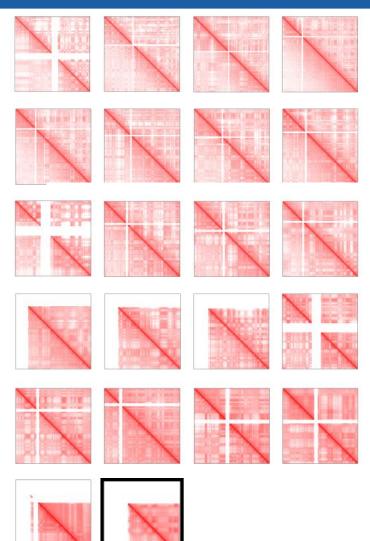


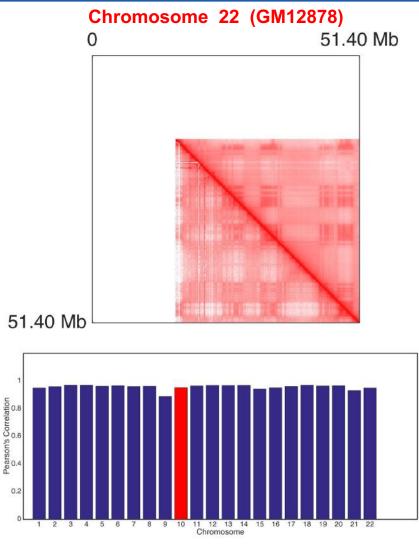




The Minimal Chromatin Model: Predictivity





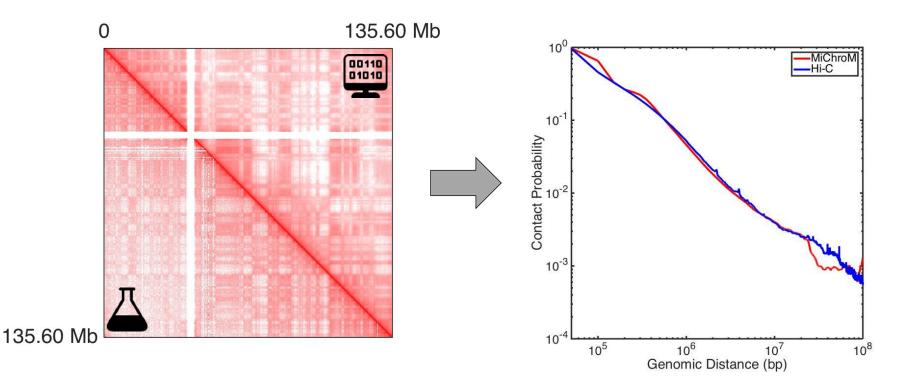








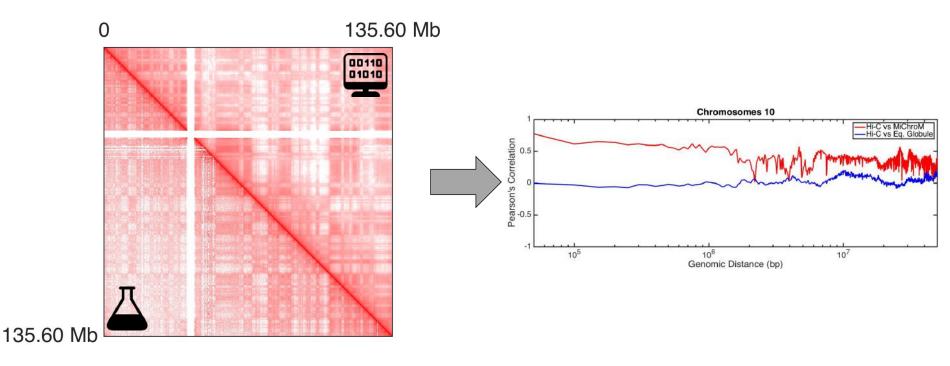
Human Chromosome 10 B-lymphoblastoid cells (GM12878)







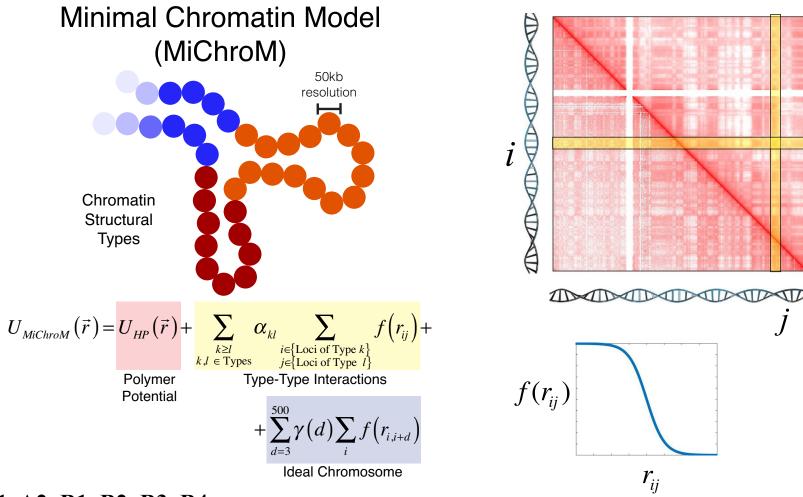
Human Chromosome 10 B-lymphoblastoid cells (GM12878)







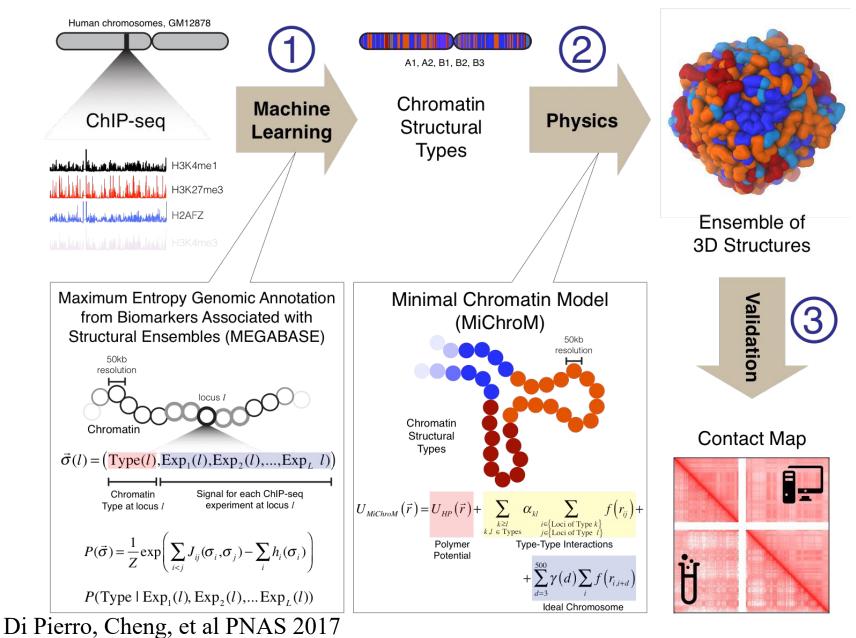
A physical polymer model with discrete chromatin types captures compartmentalization



A1, A2, B1, B2, B3, B4

Di Pierro, Zhang, Lieberman Aiden, Wolynes, Onuchic, PNAS 2016

Schematic Illustration of Computational Pipeline

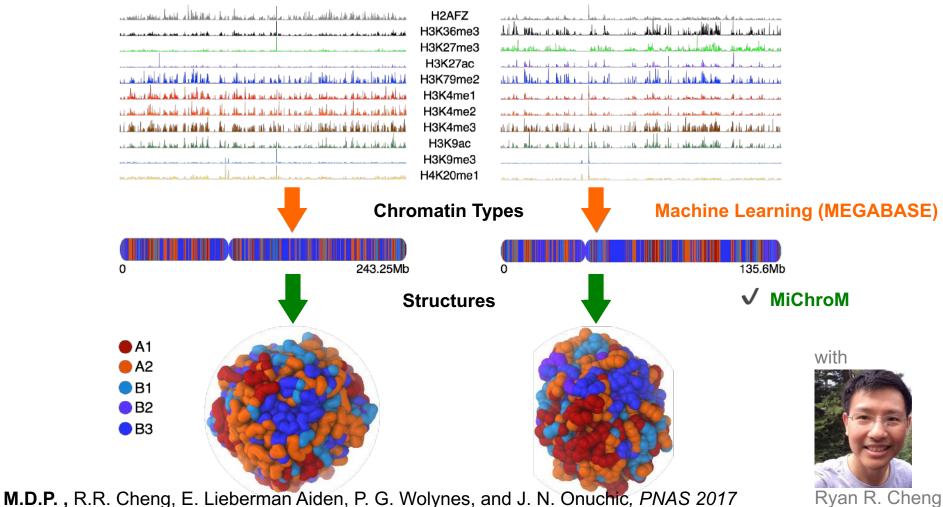


Where Is the Blueprint? From Epigenetics To Chromatin Types (MEGABASE)

Chromosome 2

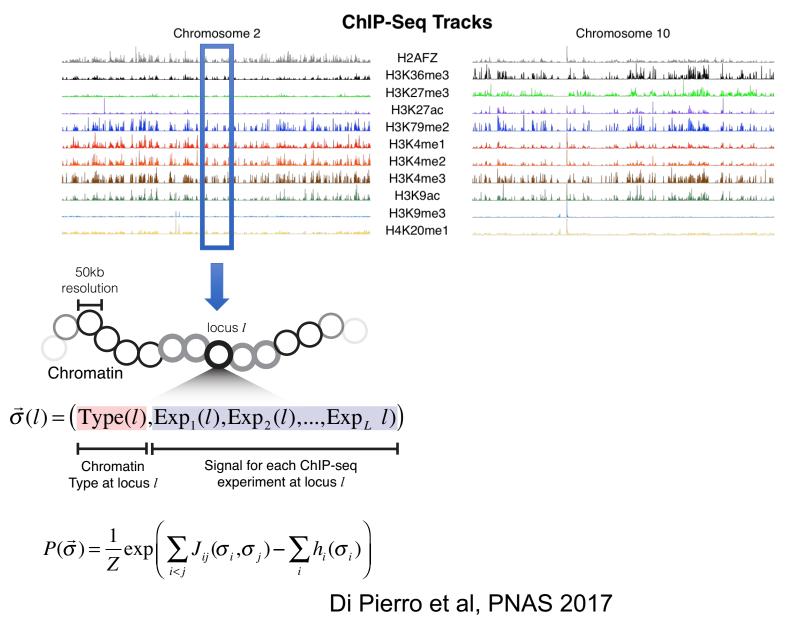
Epigenetic Markers



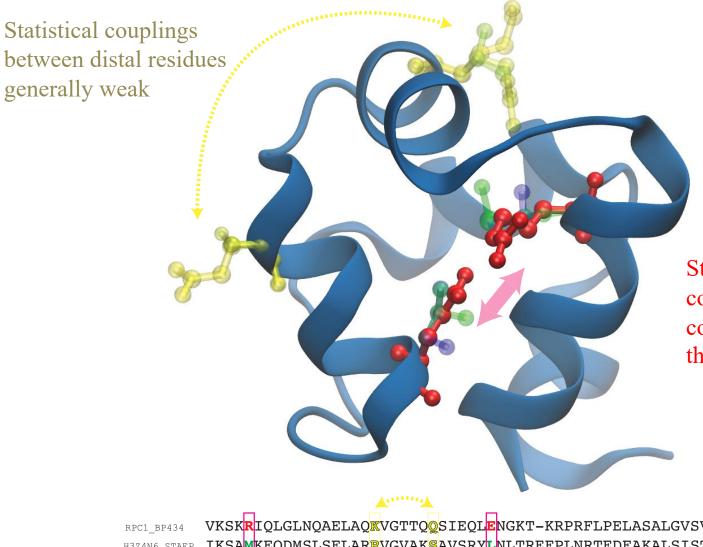




De Novo Structure Prediction of Human Chromosomes



Amino acid coevolution in proteins



Since 1998

Strong statistical couplings between coevolving residues that are in contact

RPC1_BP434 H3Z4N6_STAEP C2X6S5_BACCE

VKSKRIQLGLNQAELAQKVGTTQQSIEQLENGKT-KRPRFLPELASALGVSVDWL IKSAMKEQDMSLSELARRVGVAKSAVSRYLNLTREFPLNRTEDFAKALSISTEYL IKKLLKERALSMRQLGILTNIDPATVSRIINGKQPPKQKHLQKFAECLQVPPQLL Sequence probability distribution depends on pairwise and single site parameters

AAKAPSARGHATKPRAPKDAQHEAA AAKAP ARGHATKPRA KDAQHEAA SAKEK EKMKIVKN-L DKGKKSGS TELET LAPSG TALATAKKKE TDRTDDPV TELET FTLDOVKPRA KDGKKRSS

$$P_i(A_i) \equiv f_i(A_i)$$

$$P_{ij}(A_i, A_j) \equiv f_{ij}(A_i, A_j)$$

$$i = 6$$
 $j = 17$
 $f_{6,17}(S_6, P_{17}) = 2/6$

Using maximum entropy principle to model the joint probability distribution

 $e_{ij}(A,B) \approx -(C^{-1})_{ij}(A,B)$

$$C_{ij}(A,B) = f_{ij}(A,B) - f_i(A)f_j(B)$$

 $P(A_1, \dots, A_L) = \frac{1}{Z} \exp\{\sum_{i < j} e_{ij}(A_i, A_j) + \sum_i h_i(A_i)\}$

Faruck Morcos Martin Weight

Direct Information Metric

Direct Probabilities are defined as:

-1

$$P_{ij}^{(dir)}(A,B) = \frac{1}{Z} \exp\{e_{ij}(A,B) + \hat{h}_i(A) + \hat{h}_i(B)\}$$

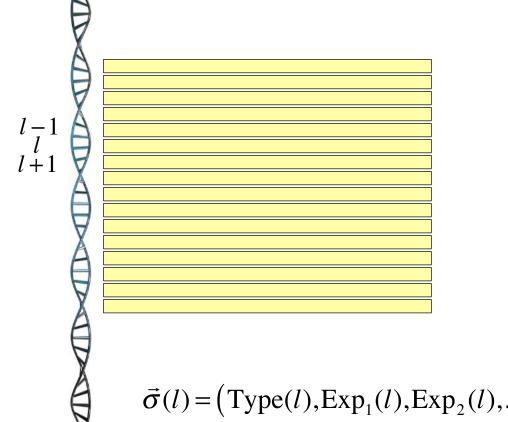
The probabilities for residue couplets are ranked using Direct Information

$$DI_{ij} = \sum_{A,B=1}^{q} P_{ij}^{(dir)}(A,B) \ln \frac{P_{ij}^{(dir)}(A,B)}{f_i(A) f_j(B)}$$
Select top couplets
$$True \text{ positive contacts (<8A) are } \int_{|\mathbf{i}|>4}^{r} \int_{|\mathbf{i}|>4}^{\mathsf{solution}} \int_{\mathbf{i}|\mathbf{i}|>4}^{\mathsf{solution}} \int_{\mathbf{i}|\mathbf{i}|=1}^{\mathsf{solution}} \int_{\mathbf{i}|\mathbf{i}|=1}^{\mathsf{solutio$$

True Positive (TP) rates

A Statistical Model for ChIP-Seq Data





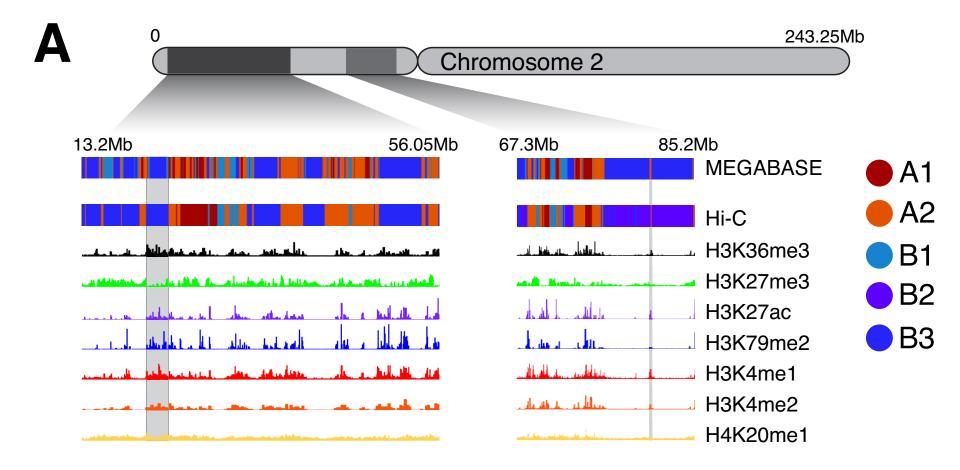
 $P(\vec{\sigma}) = \frac{1}{Z} \exp(-H(\vec{\sigma}))$ $H(\vec{\sigma}) = -\sum_{i < j} J_{ij}(\sigma_i, \sigma_j) - \sum_i h_i(\sigma_i)$

$$\vec{\sigma}(l) = (\text{Type}(l), \text{Exp}_1(l), \text{Exp}_2(l), \dots, \text{Exp}_L(l))$$

Training set: odd-numbered chromosomes Test set: even-numbered chromosomes

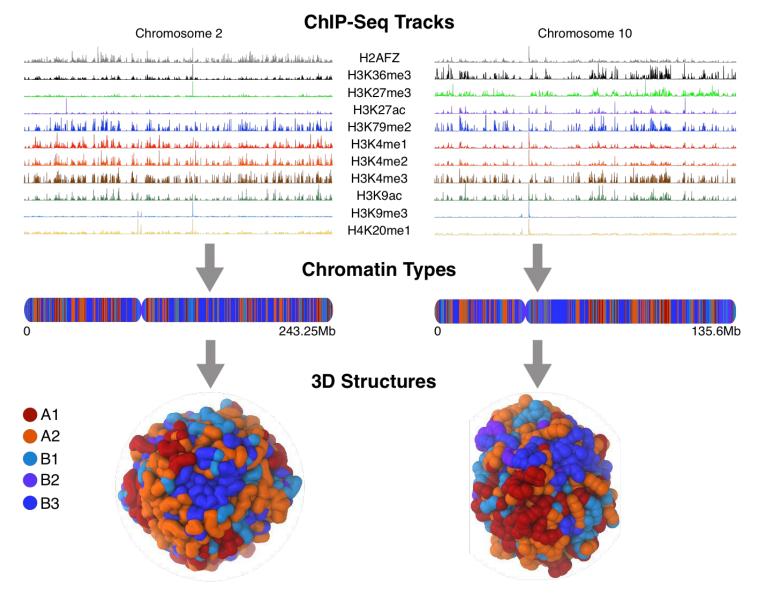






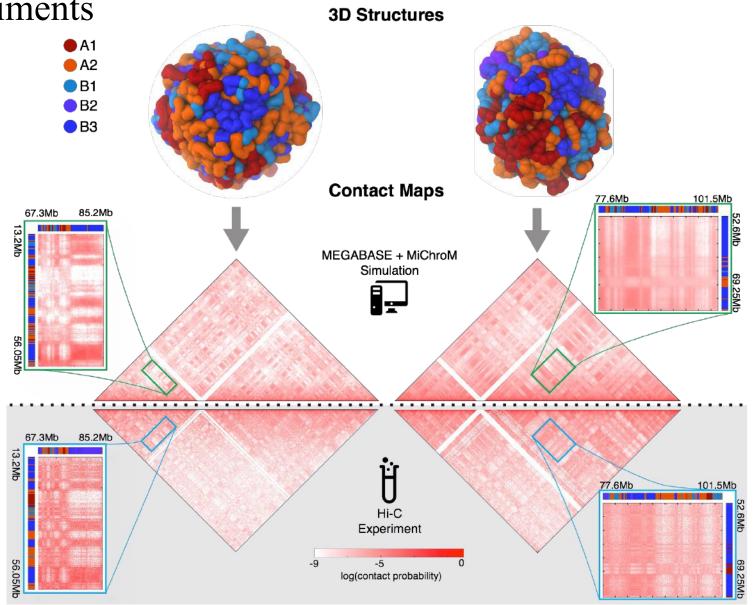


De Novo Structure Prediction of Human Chromosomes



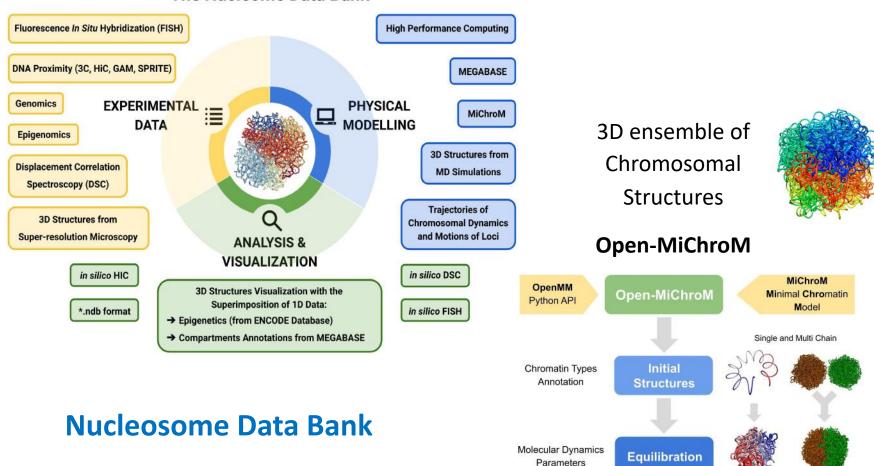
Di Pierro, Cheng, et al PNAS 2017

Comparison of simulated structures with DNA-DNA Ligation Experiments 3D Structures



Di Pierro, Cheng, et al PNAS 2017

ndb.rice.edu



Ensemble of 3D

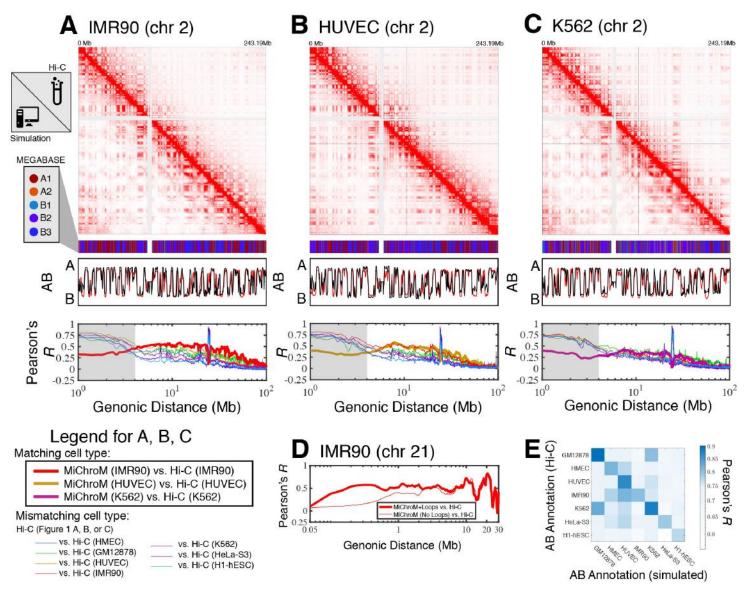
Structures

Production

A web platform to simulate and browse the three-dimensional architecture of genomes

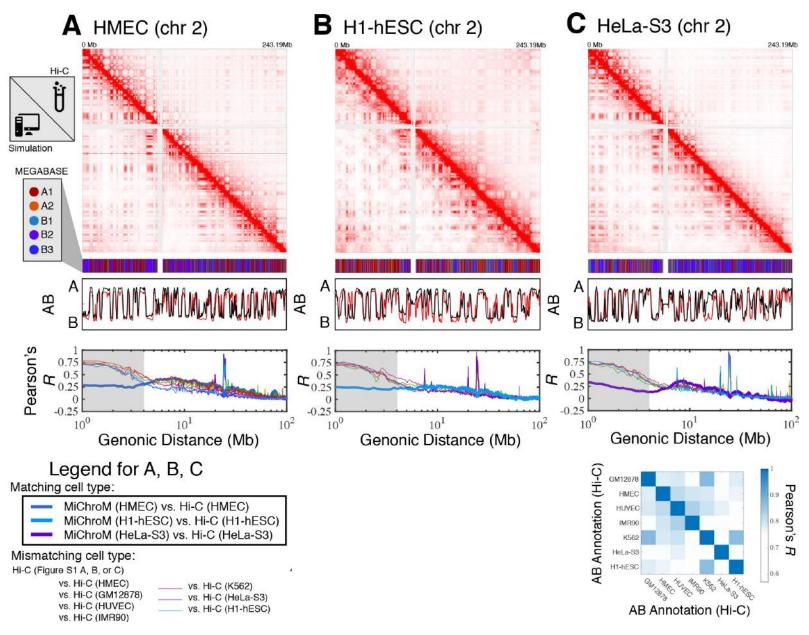
The Nucleome Data Bank

Prediction of chromosome structures for differentiated cell lines and for immortalized leukemia cells.

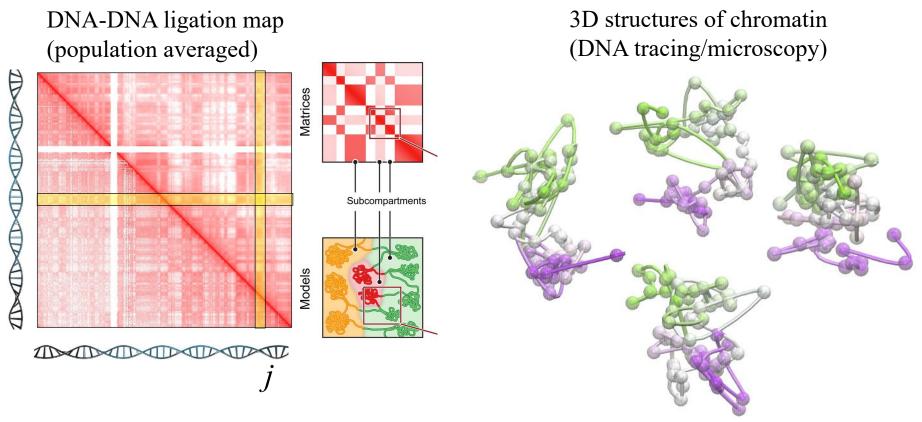


Cheng et al. eLife 2020

Prediction of chromosome structures for HMEC, H1-hESC, and HeLa-S3.



Chromosome Structural Heterogeneity: No two structures are identical

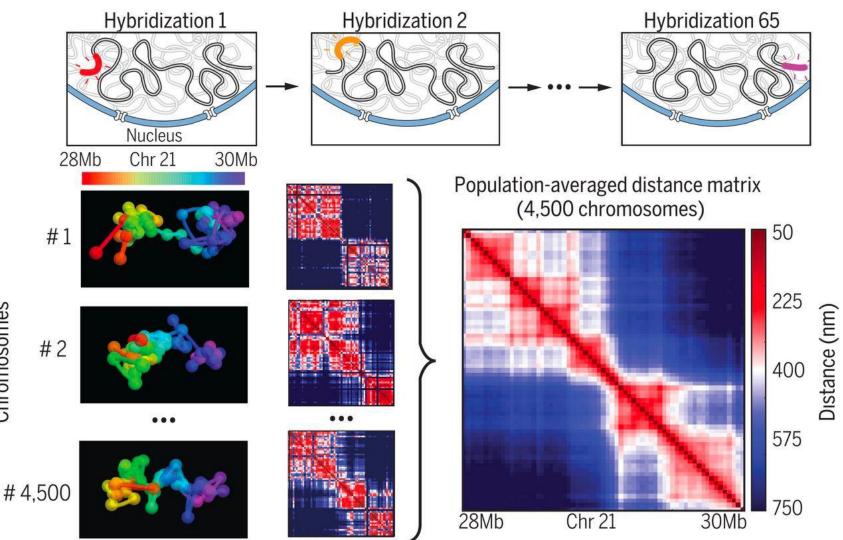


Rao & Huntley et al, Cell 2014

Bintu et al, Science 2018

Super-resolution imaging of chromatin

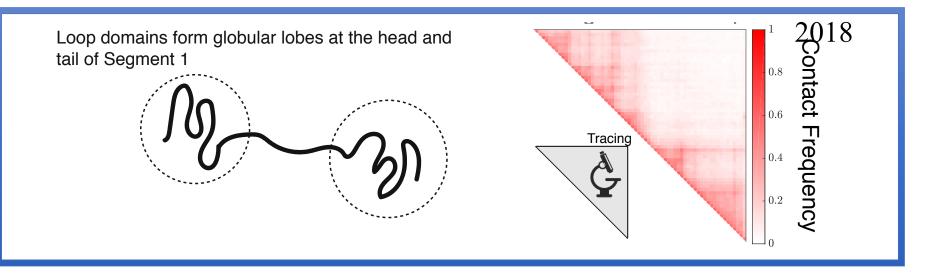
B. Bintu and L. Mateo..... A. Boettiger, Xiaowei Zhuang, Science



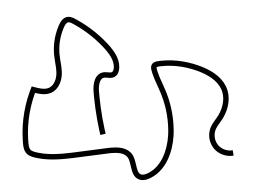
Chromosomes

Super-resolution imaging of chromatin

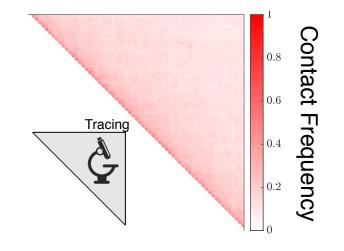
B. Bintu and L. Mateo..... A. Boettiger, Xiaowei Zhuang, Science



Segment 1: chr21 29.37-31.32 Mb



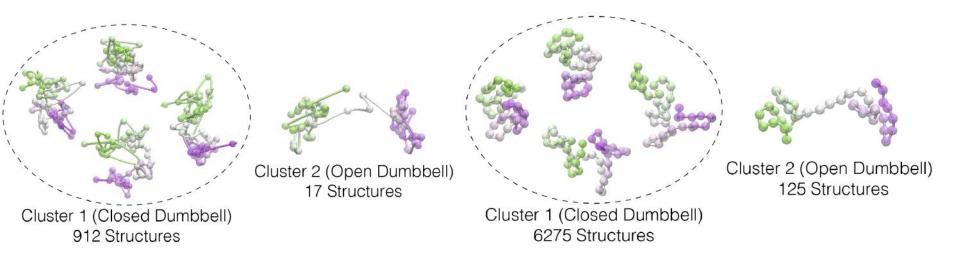
Segment 2: chr21 20.0-21.9 Mb



ANALYSIS OF CHROMATIN STRUCTURES REVEAL STRUCTURAL TRANSITIONS: OPEN VS. CLOSED

Experimentally Traced Structures (Bintu et al *Science* 2018) IMR90 chromosome 21: 29.37-31.32Mb Simulated Structures (MiChroM)

IMR90 chromosome 21: 29.37-31.32Mb

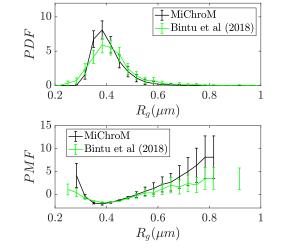


Structural similarity order parameter:

$$Q_{AB} = \frac{1}{N} \sum_{i,j} \exp\left(-\frac{\left(r_{ij}^{A} - r_{ij}^{B}\right)^{2}}{2\delta^{2}}\right)$$
$$\delta = 0.165 \mu m$$

Free energy difference between open and closed structures (experiment & simulation):

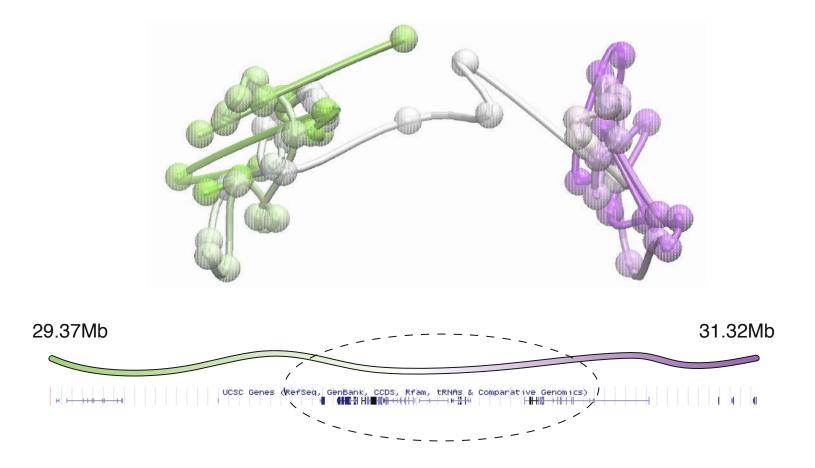
$$\log(N_{closed} / N_{open}) = E_{open} - E_{closed} \sim 4k_B T$$



Cheng et al, eLife 2020.

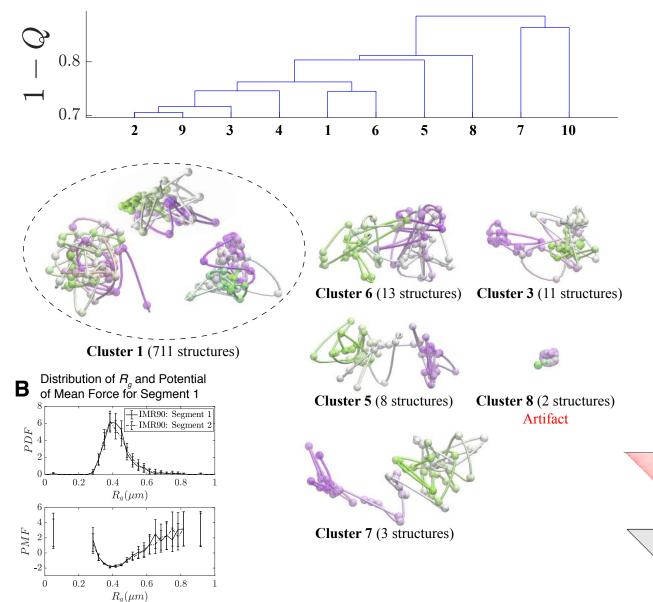
A relationship between gene expression and chromosome structure?

Genes primarily located in linker region



Analysis of Experimental Traced Chromatin Structures of

A Hierarchical Clustering of Traced Structures of IMR90 Segment 2 (chr21 20.0-21.9 Mb)

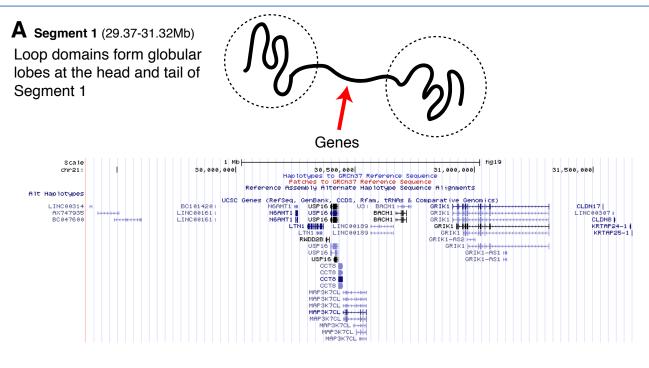


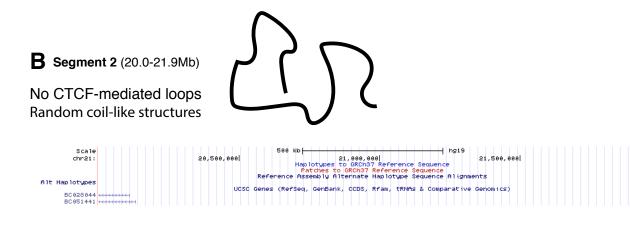
Contact Frequency

Tracing

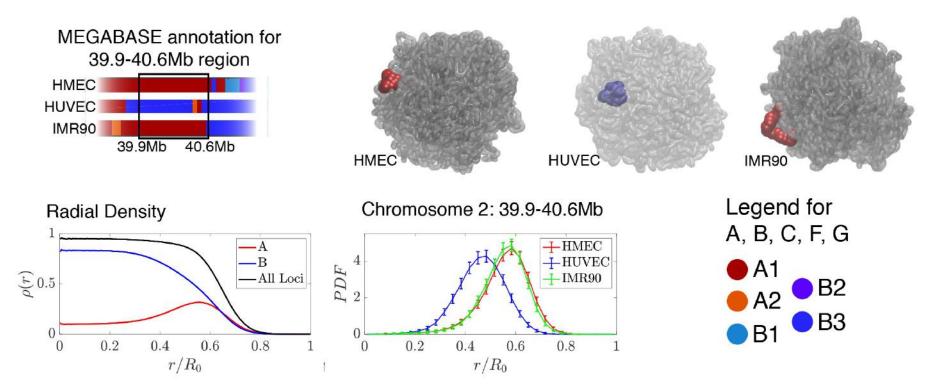
G

A relationship between gene expression and chromosome structure?



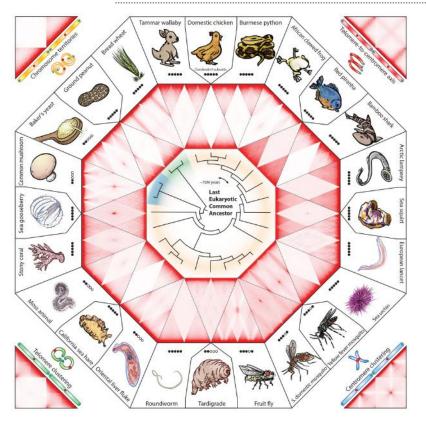


A-compartment chromatin moves to territorial surface



Consistent with experimental observations: Nagano, T., et al., Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. Nature, 2013. 502(7469): p. 59-64.

CHROMOSOME ARCHITECTURE ACROSS EVOLUTION CORRELATES WITH CONDENSIN II ACTIVITY



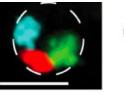
Hoencamp et al., Science (2021)

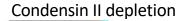
Type I architecture		Type II architecture	
Centromere clustering	\succ	Chromosome	۶
Telomere clustering	\triangleright	territories	
Telo-to-centromere axis		All condensin II subunits present	\succ
One or more condensin II subunits absent			

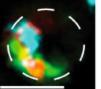
Condensin II (loop-extruding enzyme) confers longitudinal rigidity to chromosomes and defines chromosome

territories (Houlard et al., *Nat. Cell Biol.* 2015, Bauer et al. *PLoS Genet.* 2012; Rosin et al. *PLoS Genet.* 2018, ...)

Wild Type



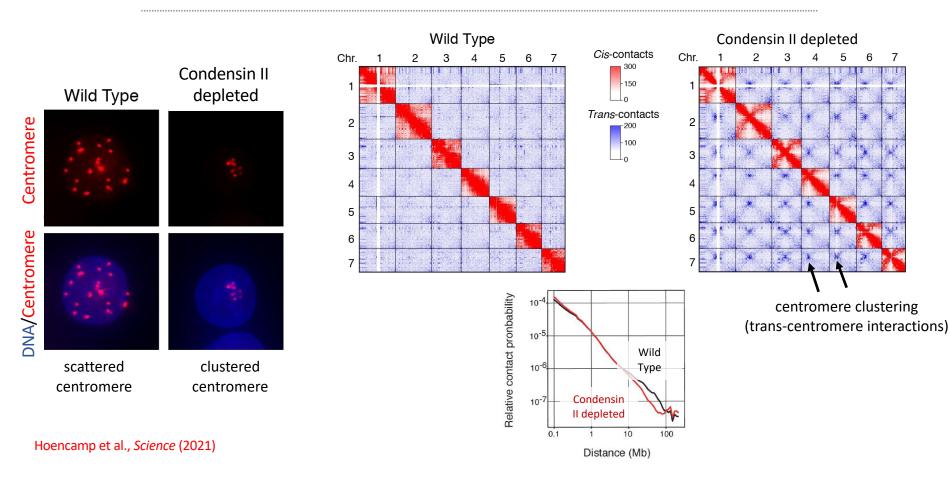




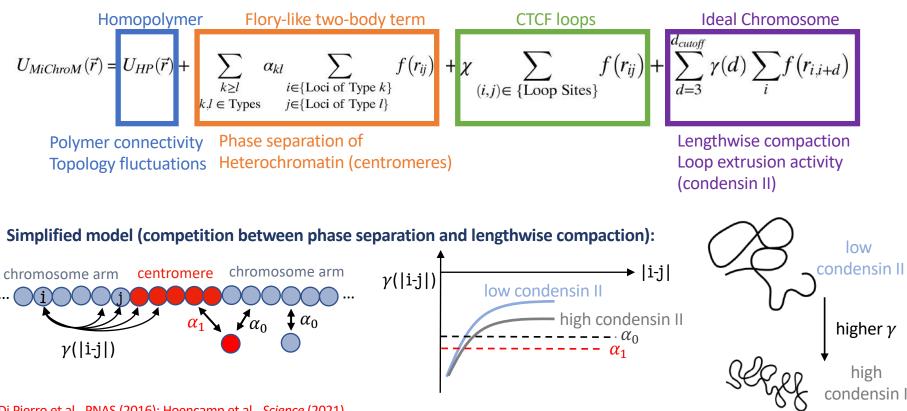


Rosin et al. PLoS Genet. 2018

CONDENSIN II DEPLETION LEADS TO A PHENOTYPE WITH CLUSTERED-CENTROMERES AND LOWER CIS-COMPACTION

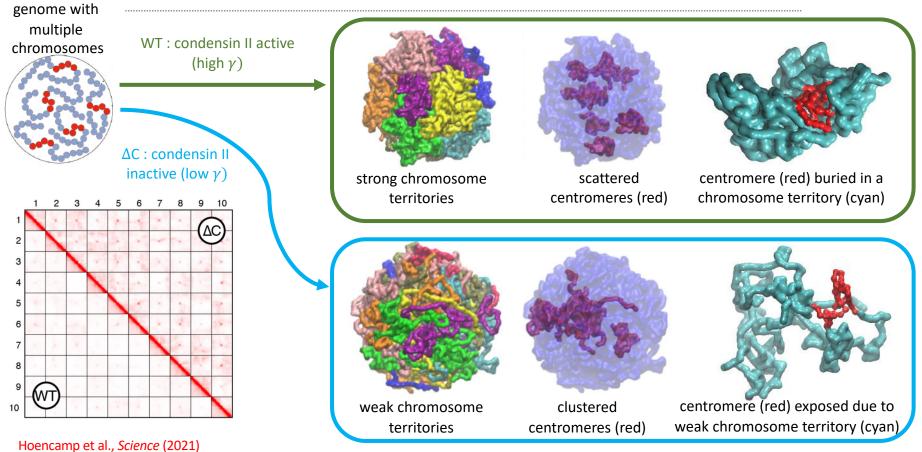


MICHROM INTERACTIONS AND LENGTHWISE COMPACTION ACTIVITY OF CONDENSIN II



Di Pierro et al., PNAS (2016); Hoencamp et al., Science (2021)

CONDENSIN II ACTIVITY (LENGTHWISE COMPACTION) COUNTERACTS CENTROMERE CLUSTERING



Ryan Cheng., Sumithaba Brahmachari, Vinicius Contessoto, Antonio Oliveira Jr. Dongya Jia - NIH Susmita Roy – IISER – India Fang Bai – Shanghai Tech, China * Michele Di Pierro NorthEastern Ellinor Haglund – Hawaii * Heiko Lammert - Karlsruher Faruck Morcos, - University of Texas at Dallas Mingyang Lu - Northeastern * Jeff Noel - Max Delbruck Center, Berlin Alex Schug - Karlsruher Institut für Technologie (KIT) Biman Jana – Indian Assoc. Cultivation of Science - India Joanna Sulkowska - University of Warsaw Koby Levy - Weizmann Institute * Changbong Hyeon - KIAS, Korea Cecilia Clementi – Rice University * Joan-Emma Shea – UCSB Adelia Aquino – Texas Tech * Steve Plotkin – UBC, Canada Osamu Miyashita – RIKEN, Japan John Finke - University of Washington, Tacoma Antitsa Stoycheva - Gilead Sciences, Inc. Nick Socci – Sloan Katering Yoko Suzuki - Meisei University - Tokyo Chigusa Kobayashi - RIKEN, Japan Eric Nelson – UTSWMC, Texas Jian Liu - NIH Shachi Gosavi - NCBS - India Dana Krepel – Biotech Israel Daniel Schultz – Harvard Leandro Oliveira and Jorge Chahine - UNESP, Brazil Marcio de Mello Cardoso – UFRJ, Brasil Kareen Mehrabiani, Weiqi Lu, Madelin Galbraith, Esteban Rojas, Erdong Ding, Mattheus Mello, Ragini Mahajan Federico Bocci – UC Iirvine Xingcheng Lin – MIT Bin Huang – private company Nathan Eddy - Indigo Lis Sun - China Marcelo Boareto - ETH Zurich, Switzerland Ricardo Santos – University of Campinas, Brazil Ryan Heyes – University of Michigan Mohit Raghunathan Amazon Seattle Paul Whitford – Northeastern University Hugh Nymeyer – Florida State University Marcos Betancourt - BMCC New York Vitor Leite - UNESP, Brazil Leslie Chavez - LANL Margaret Cheung - University of Houston Sichun Yang – Case Western University Peter Leopold – BioAnalyte Inc.

Collaborators

Peter Wolynes, Herbie Levine, Erez Lieberman Aiden Ben Rowland - Netherlands Cancer Institute Jianpeng Ma – Baylor College of Medicine Angel Garcia and Karissa Sanbonmatsu - LANL Pat Jennings and Terry Hwa- UCSD Martin Weigt – Univ. Paris, France Samuel Flores – Univ. of Uppsala, Sweden Zan Luthey-Schulten – UIUC Glaucius Oliva and Adriano Andricopulo - USP, Brazil Hualiang Jiang – SIMM Chins Charles Brooks – U. Michigan Stefan Klumpp – MPIKG, Germany Shoji Takada - Kobe, Japan Ashok Deniz – TSRI Mikael Oliverberg – Stockholm Antonio Francisco Pereira de Araujo – Brasilia Yoshitaka Tanimura - Kyoto, Japan Yuko Okamoto – Nagoya University, Japan Ulrich Hansmann – University of Arkansas Rachel Nechushtai – Hebrew University Ron Mittler – Univ. North Texas Jeff Tabor – Rice Univ.



National Science Foundation WHERE DISCOVERIES BEGIN

