3DGenomics

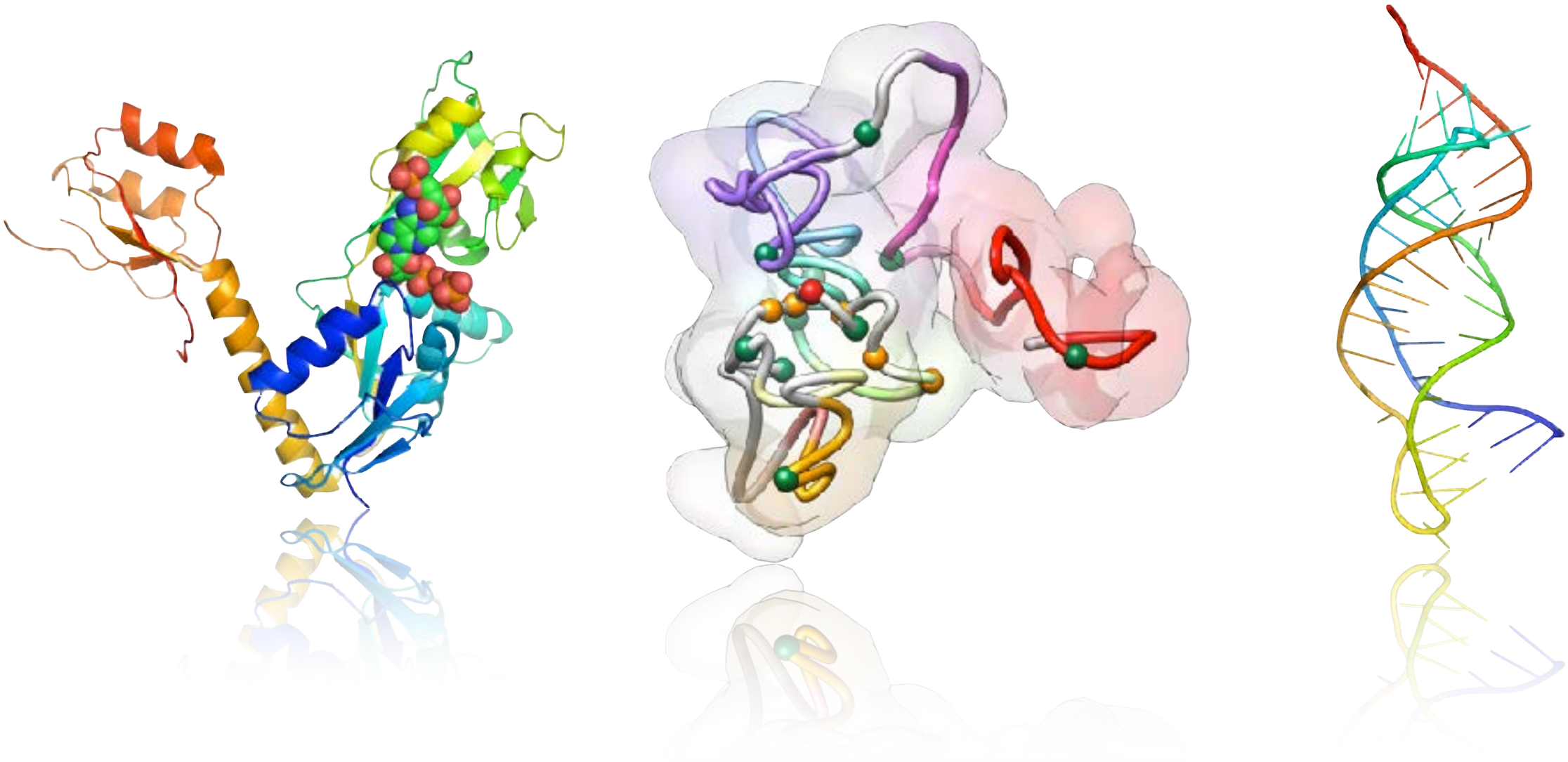
Marc A. Marti-Renom
Structural Genomics Group
CNAG-CRG

ICREA

cnag CRG Centre for Genomic Regulation

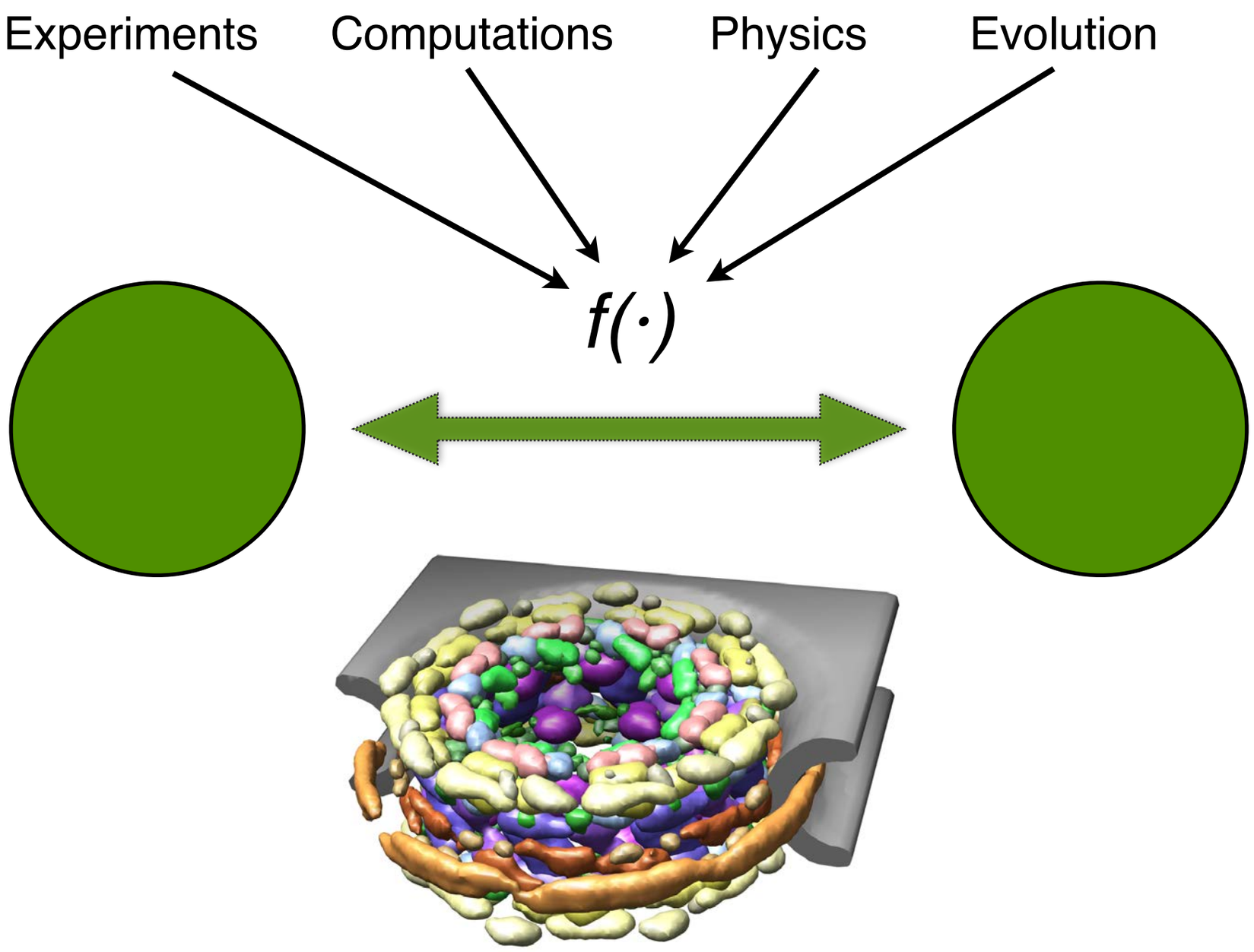
Structural Genomics Group

http://www.marciuslab.org



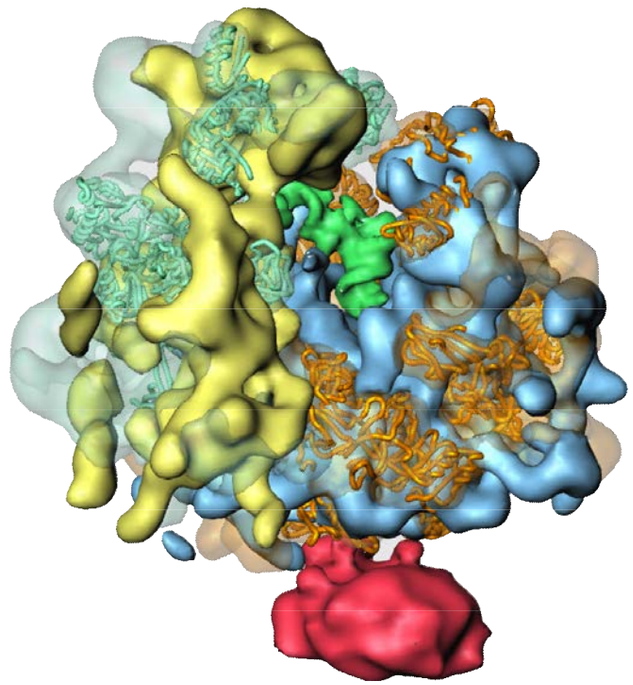
Integrative Modeling Platform

http://www.integrativemodeling.org

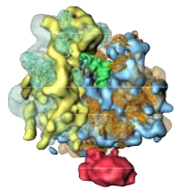


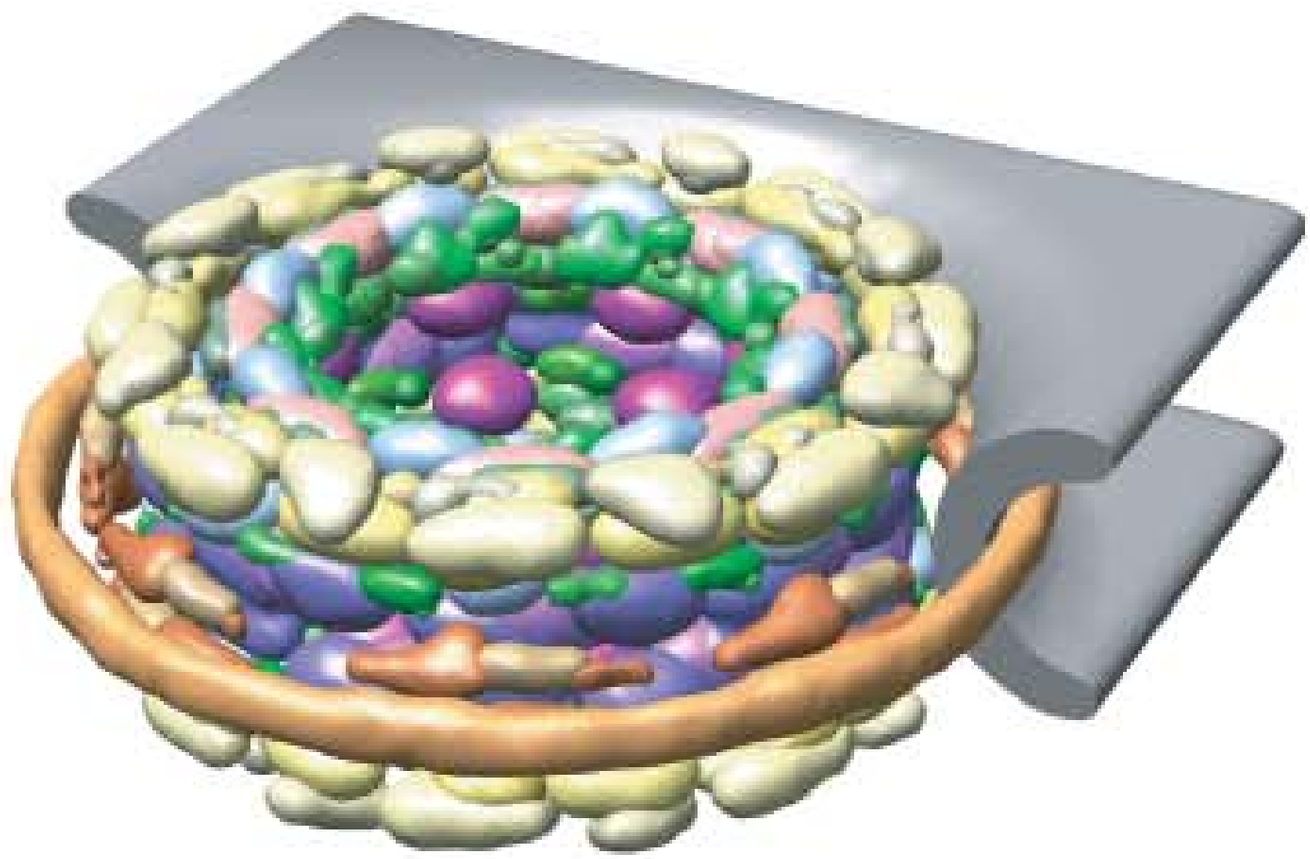
From: Russel, D. et al. PLOS Biology 10, e1001244 (2012).

Data Integration

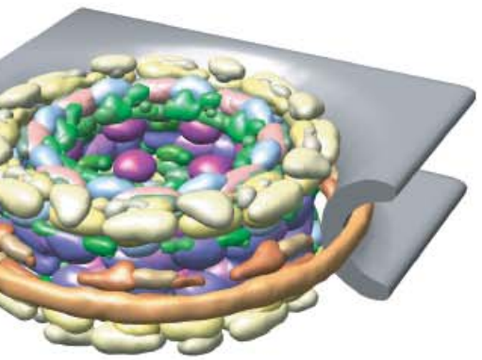


Data Integration

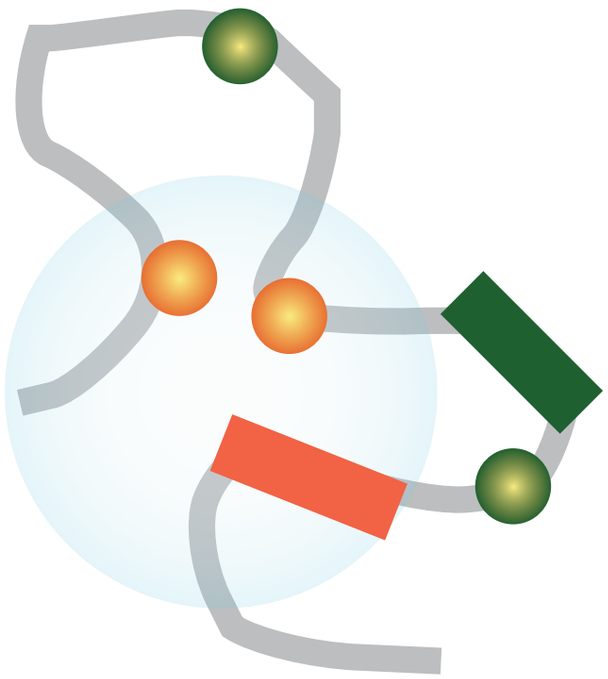


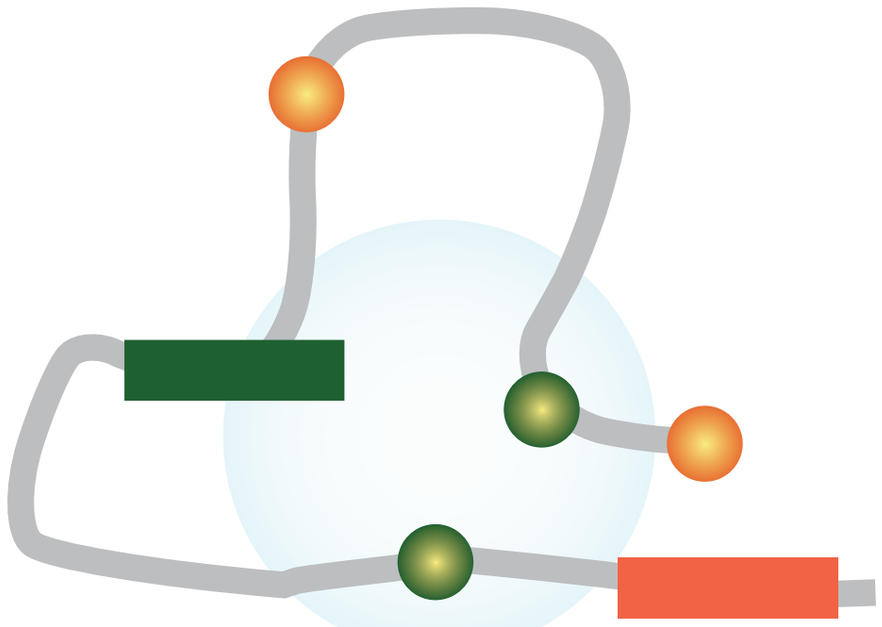


Data Integration



Complex genome organization

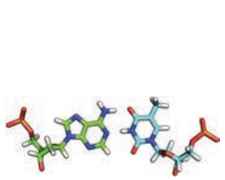


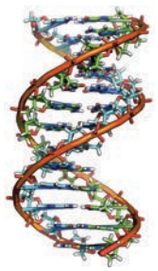


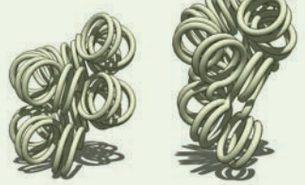
Resolution Gap

Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol 7, e1002125 (2011)

Knowledge

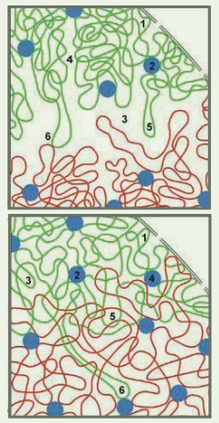


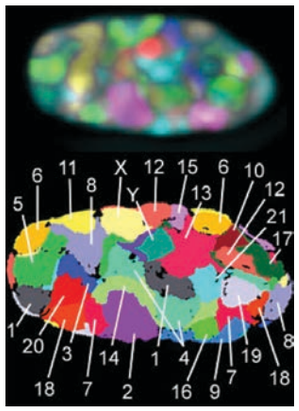




IDM

INM





DNA length

10^0 10^3 10^6 10^9 nt

Volume

10^{-9} 10^{-6} 10^{-3} 10^0 10^3 μm^3

Time

10^{-10} 10^{-8} 10^{-6} 10^{-4} 10^{-2} 10^0 10^2 10^3 s

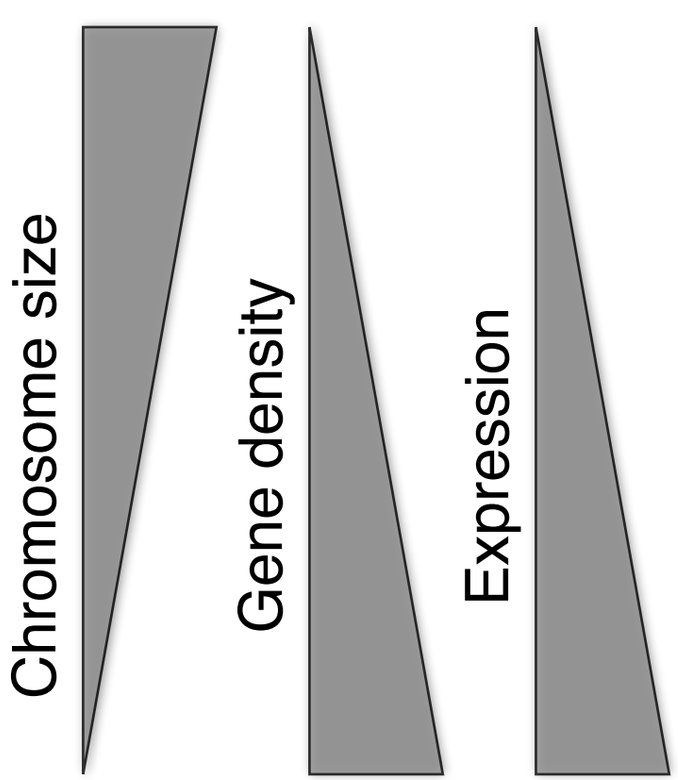
Resolution

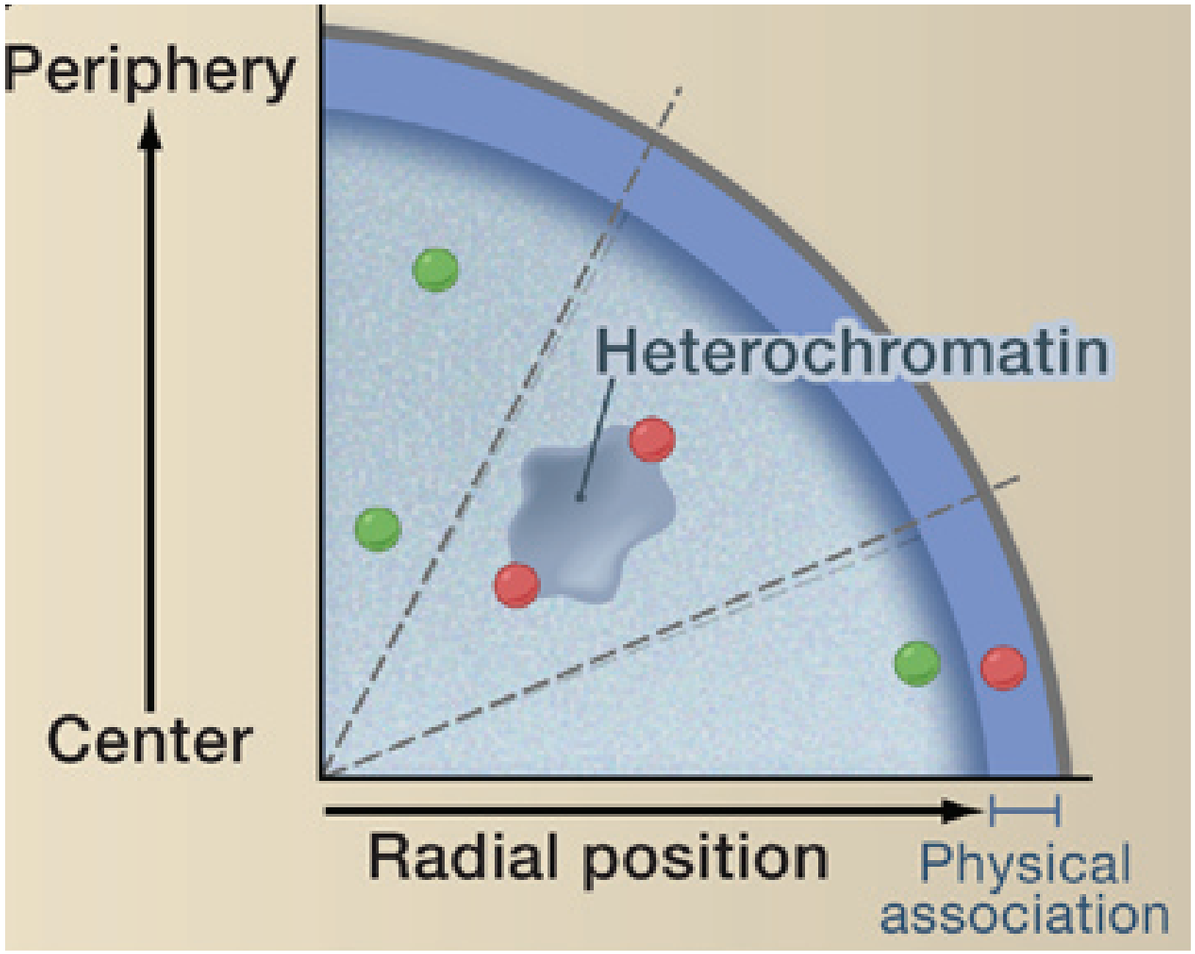
10^{-3} 10^{-2} 10^{-1} μ

cnag CRG Centre for Genomic Regulation

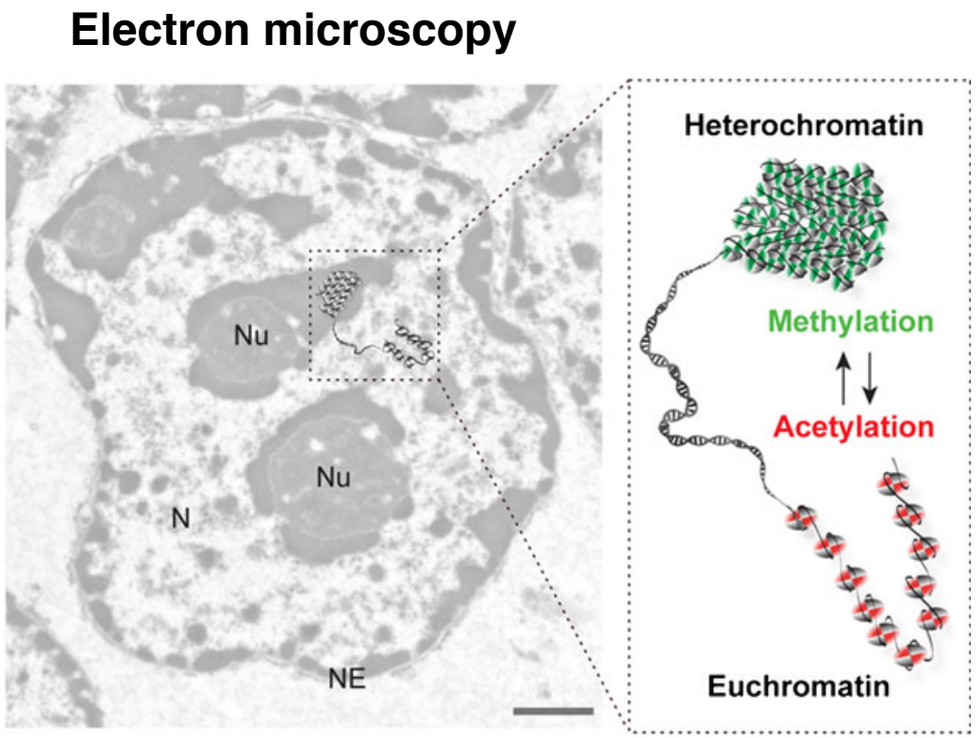
Level I: Radial genome organization

Takizawa, T., Meaburn, K. J. & Misteli, T. The meaning of gene positioning. Cell 135, 9–13 (2008).





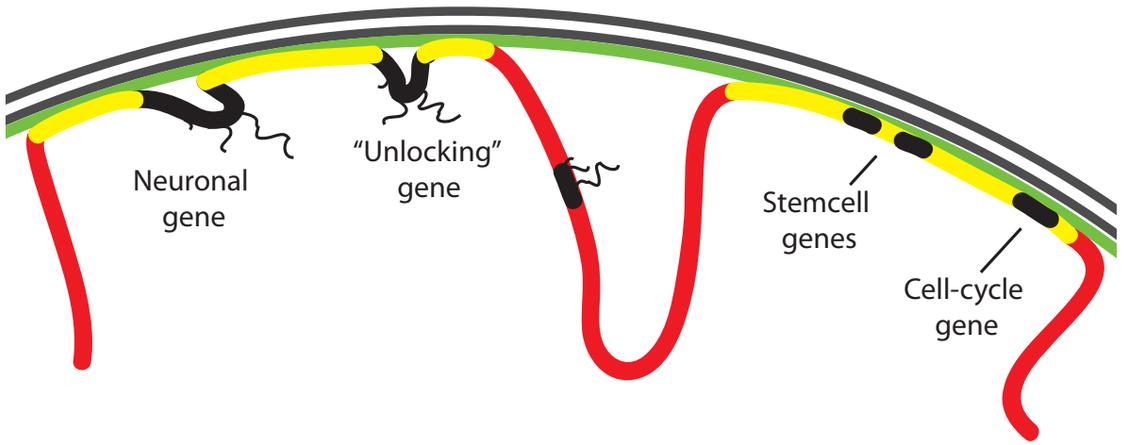
Level II: Euchromatin vs heterochromatin



Euchromatin:
chromatin that is located away from the nuclear lamina, is generally less densely packed, and contains actively transcribed genes

Heterochromatin:
chromatin that is near the nuclear lamina, tightly condensed, and transcriptionally silent

Level III: Lamina-genome interactions

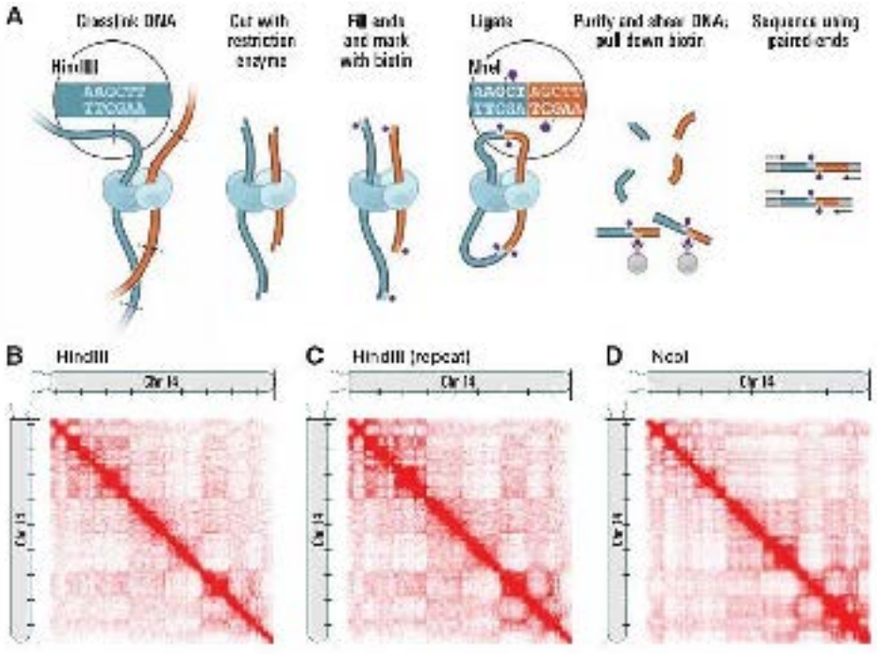


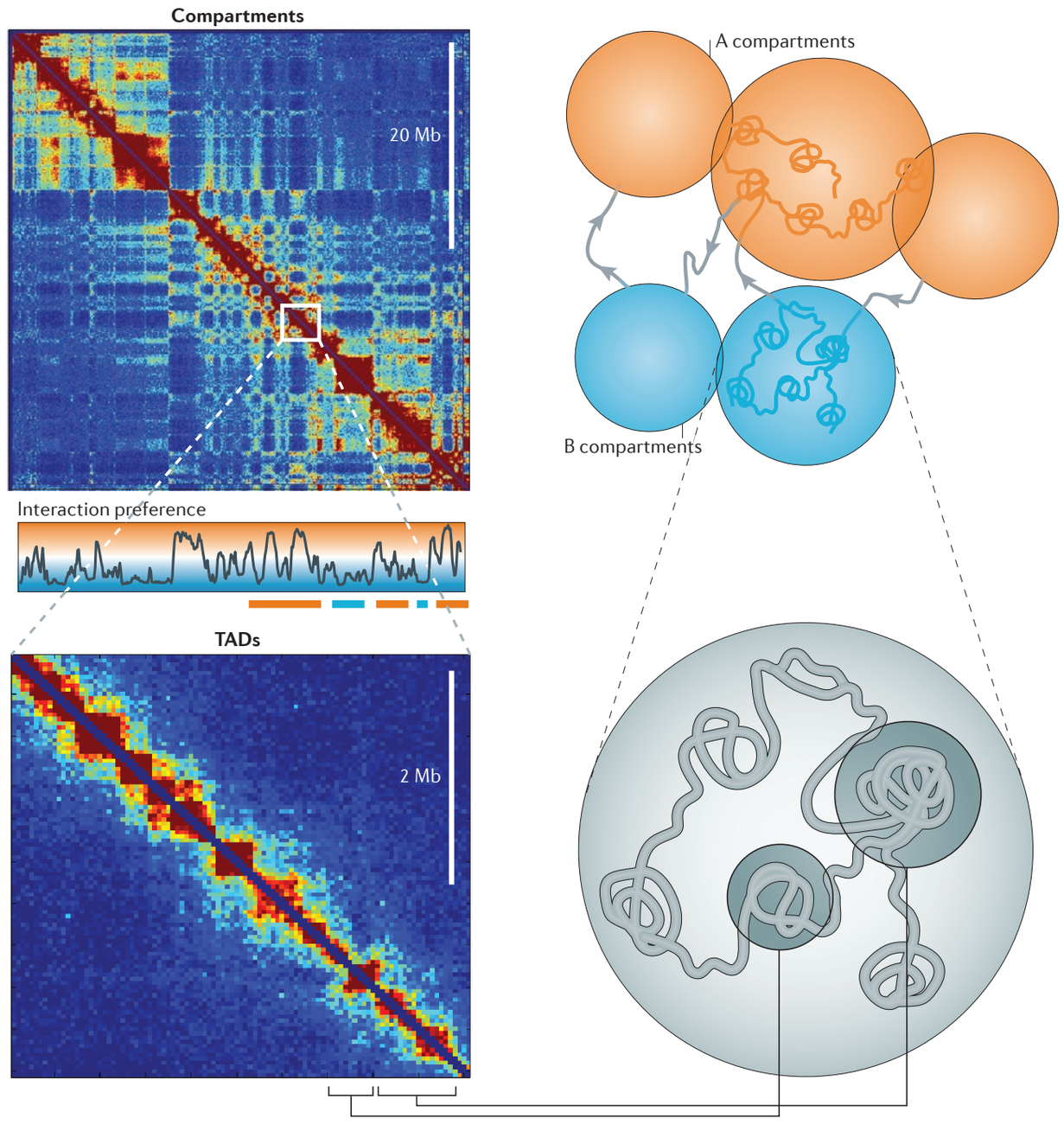
Most genes in Lamina Associated Domains are transcriptionally silent, suggesting that **lamina-genome interactions** are widely involved in the control of **gene expression**

Adapted from Molecular Cell 38, 603-613, 2010

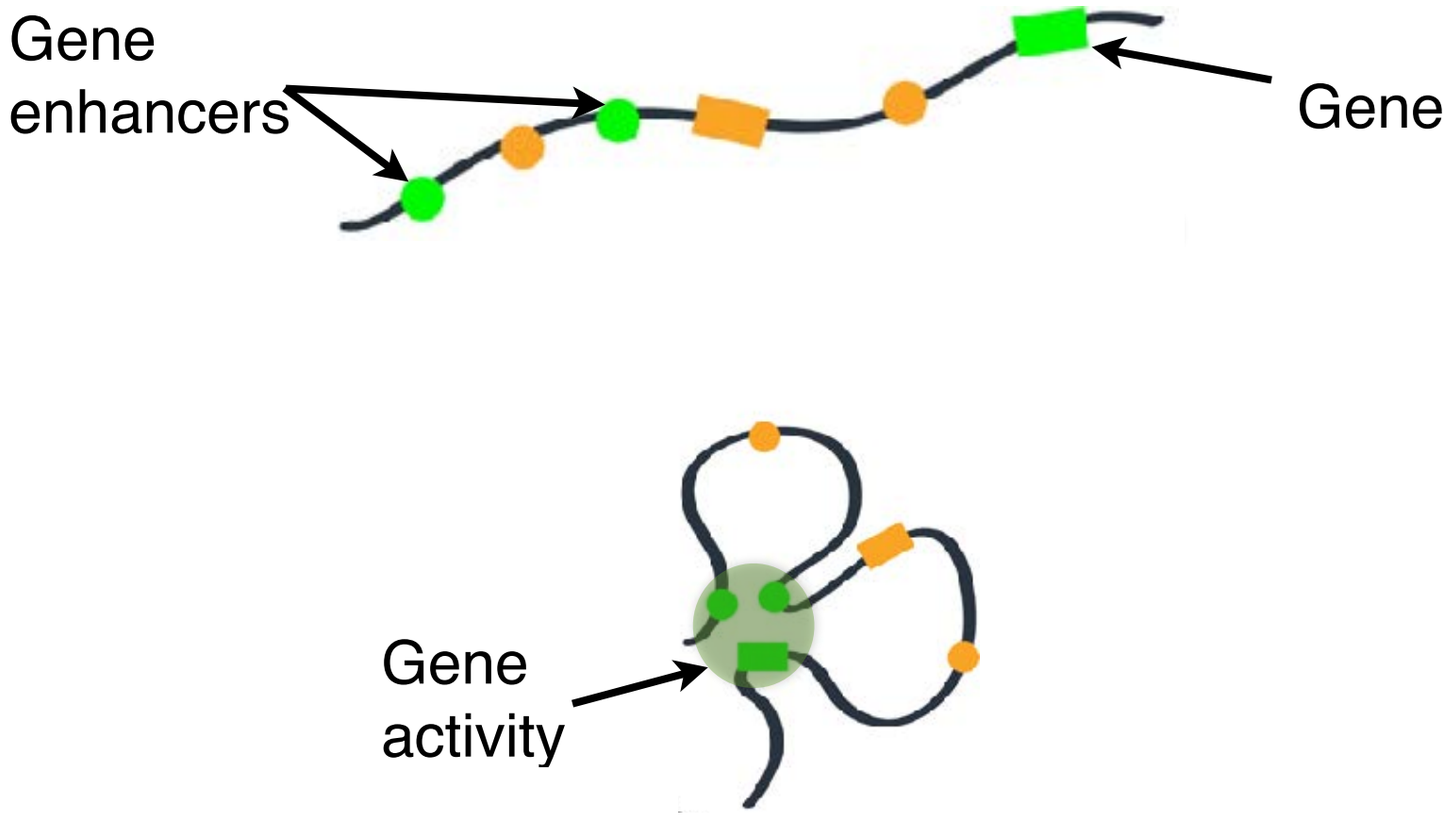
Level IV: Higher-order organization

Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet 14, 390–403 (2013).





Level V: Chromatin loops



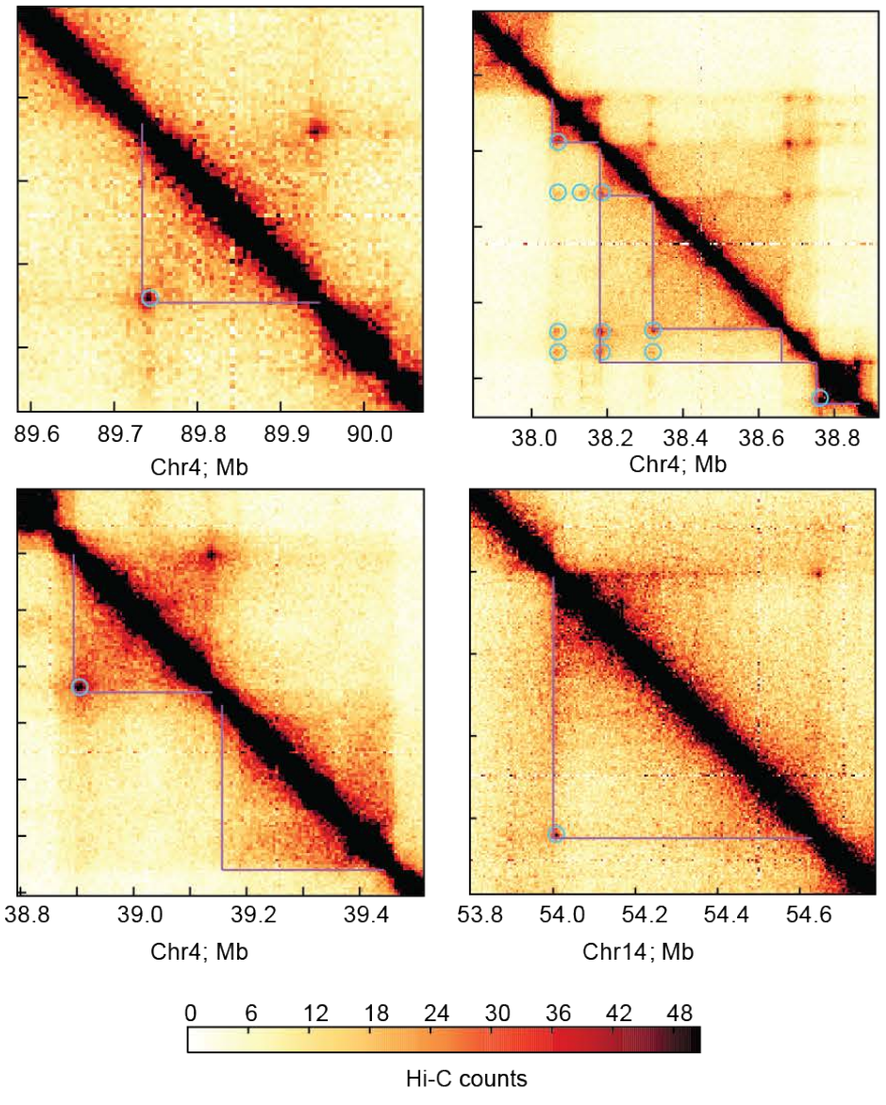
Loops bring distal genomic regions in close proximity to one another

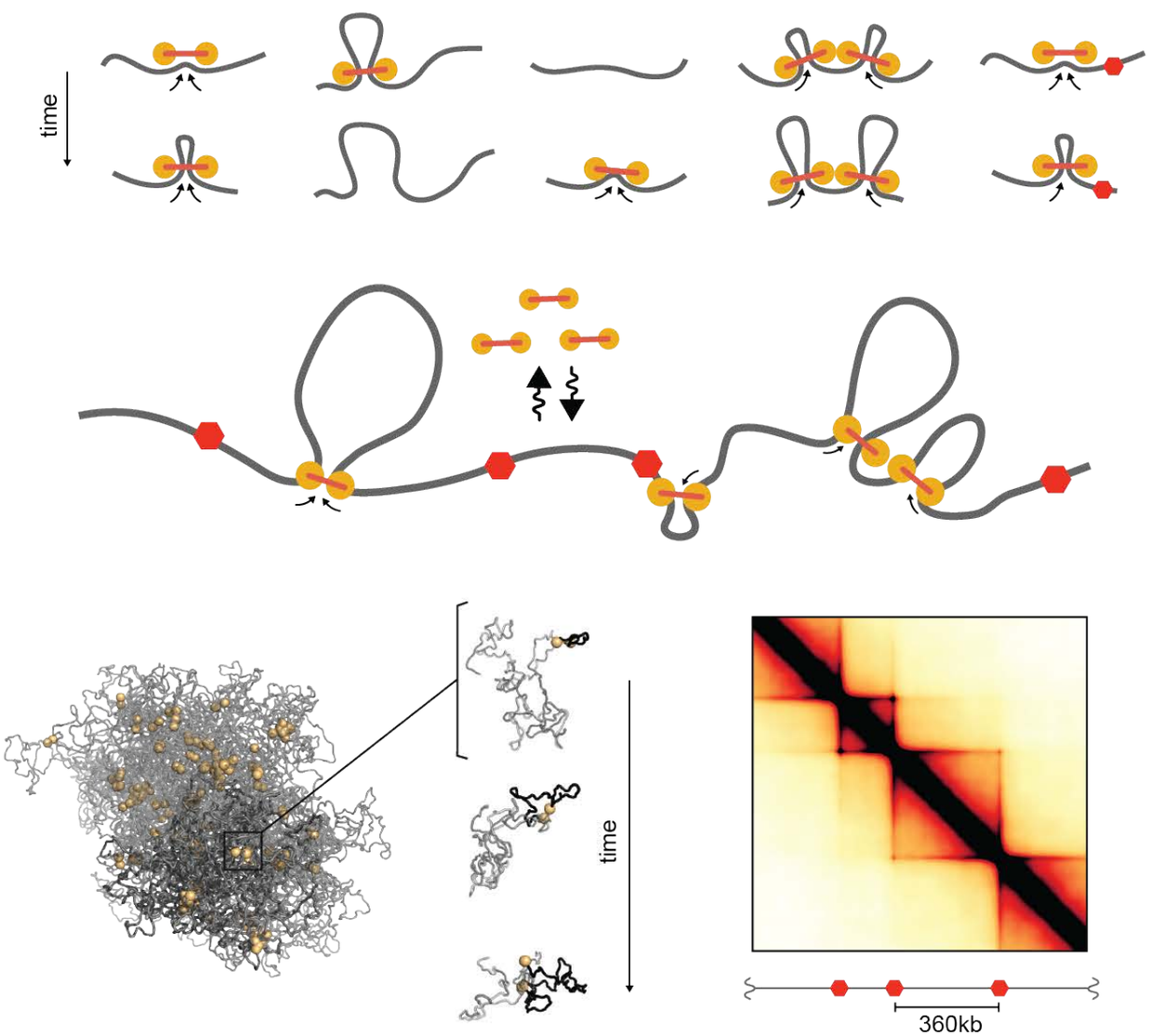
This in turn can have profound effects on gene transcription

Enhancers can be thousands of kilobases away from their target genes in any direction (or even on a separate chromosome)

Level V: Loop-extrusion as a driving force

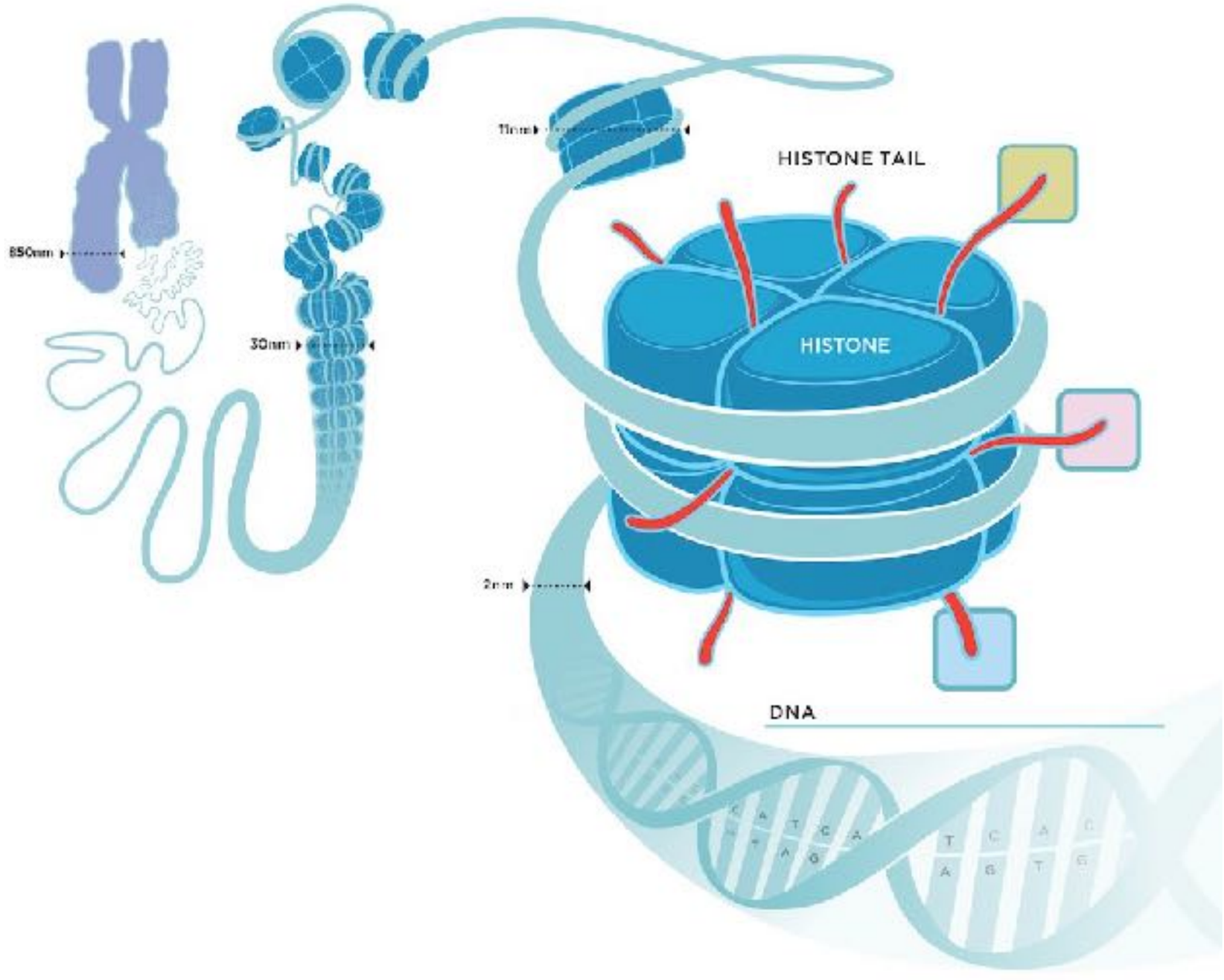
Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2015).
Formation of Chromosomal Domains by Loop Extrusion. bioRxiv.





Level VI: Nucleosome

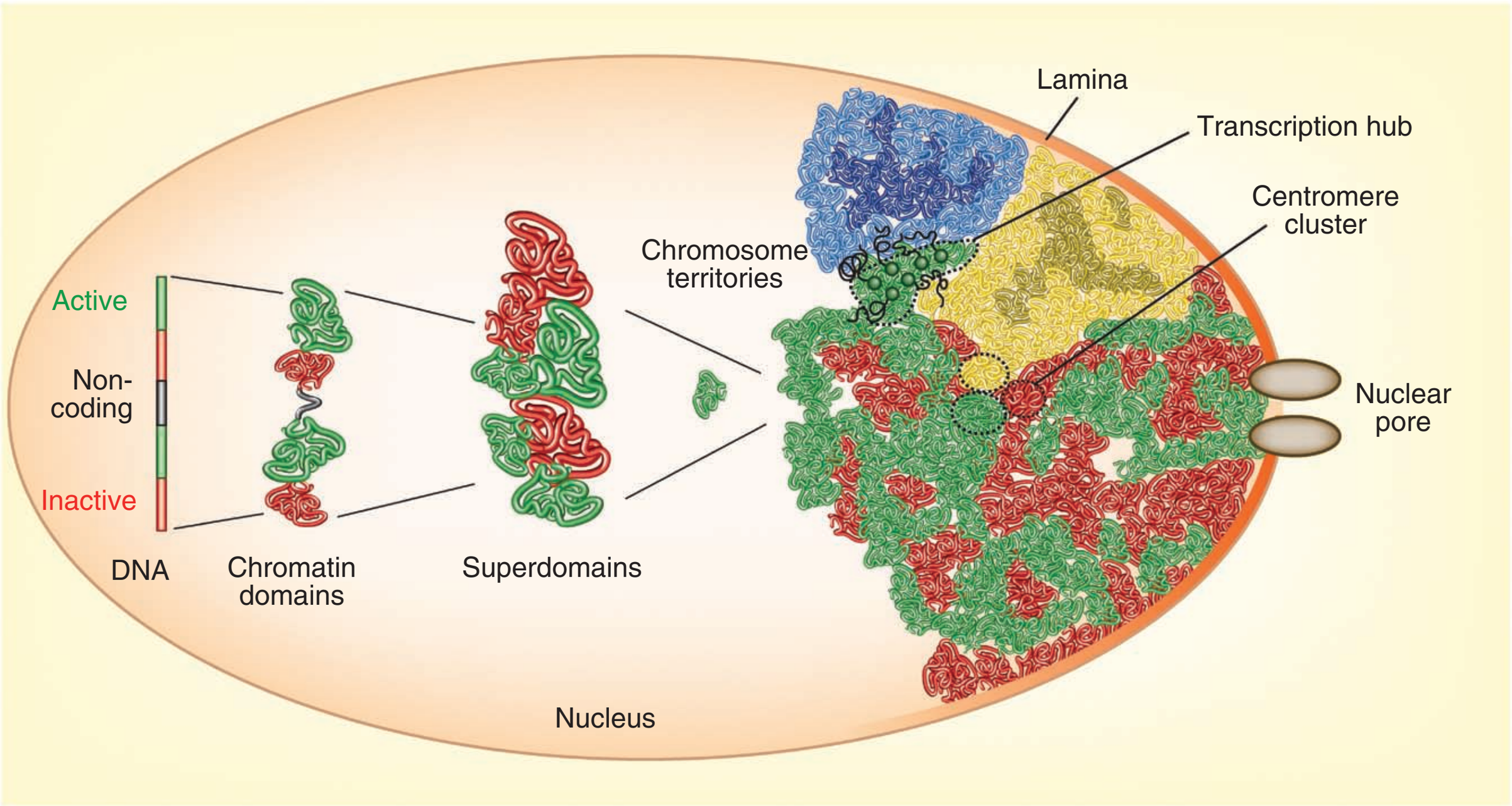
Chromosome **Chromatin fibre** **Nucleosome**



Adapted from Richard E. Ballermann, 2012

Complex genome organization

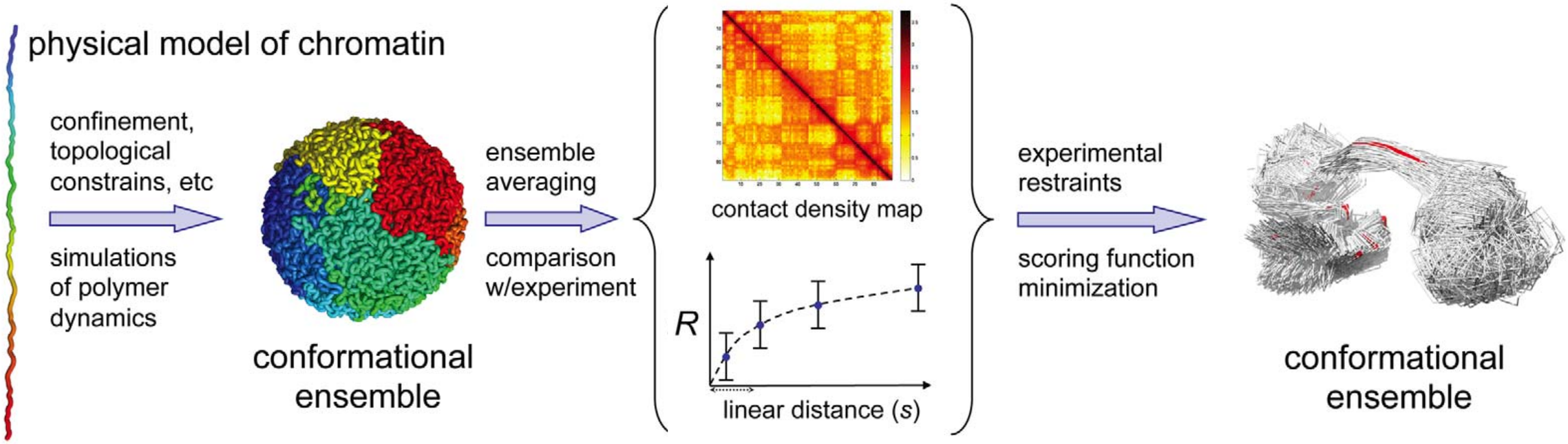
Cavalli, G. & Misteli, T. Functional implications of genome topology. Nat Struct Mol Biol 20, 290–299 (2013).



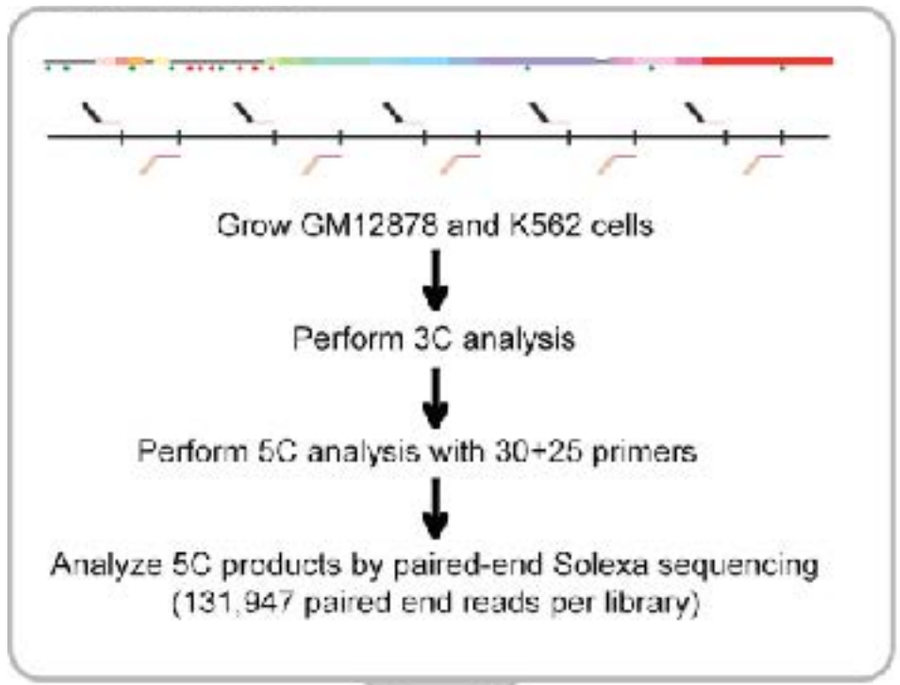
Marina Corral

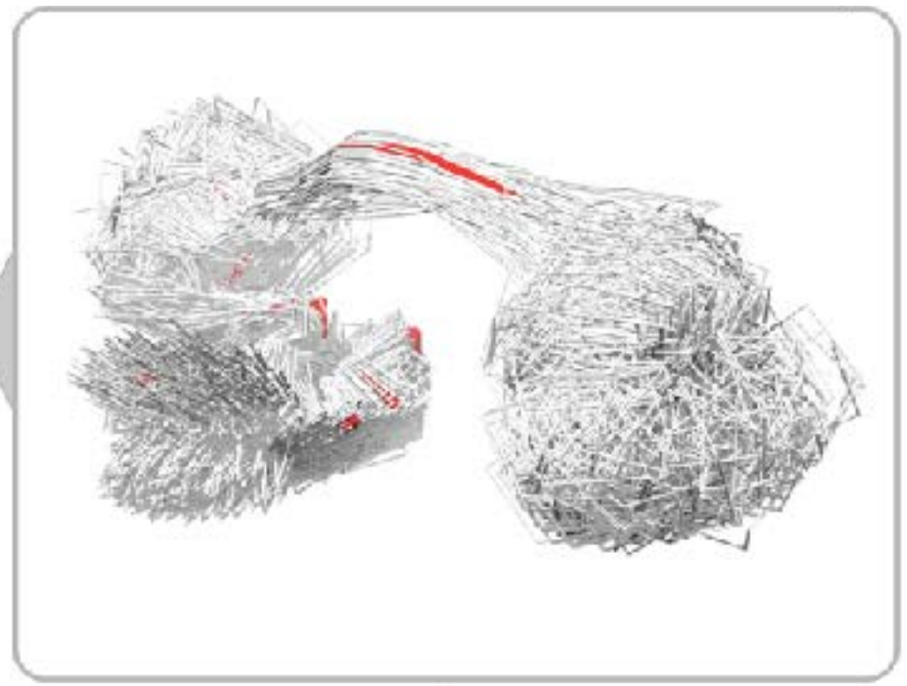
Modeling Genomes

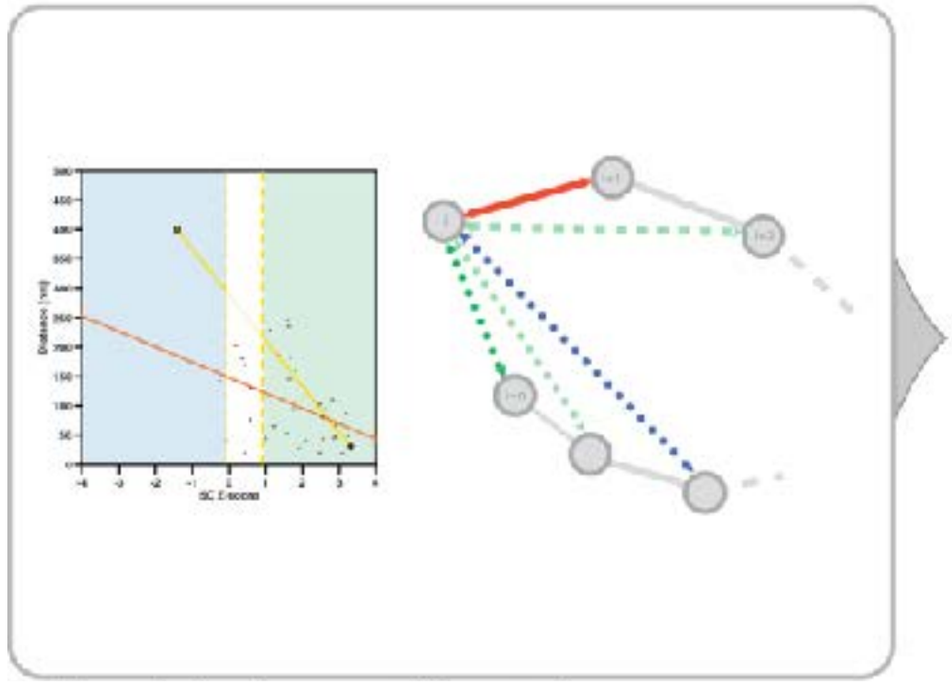
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol 7, e1002125 (2011)

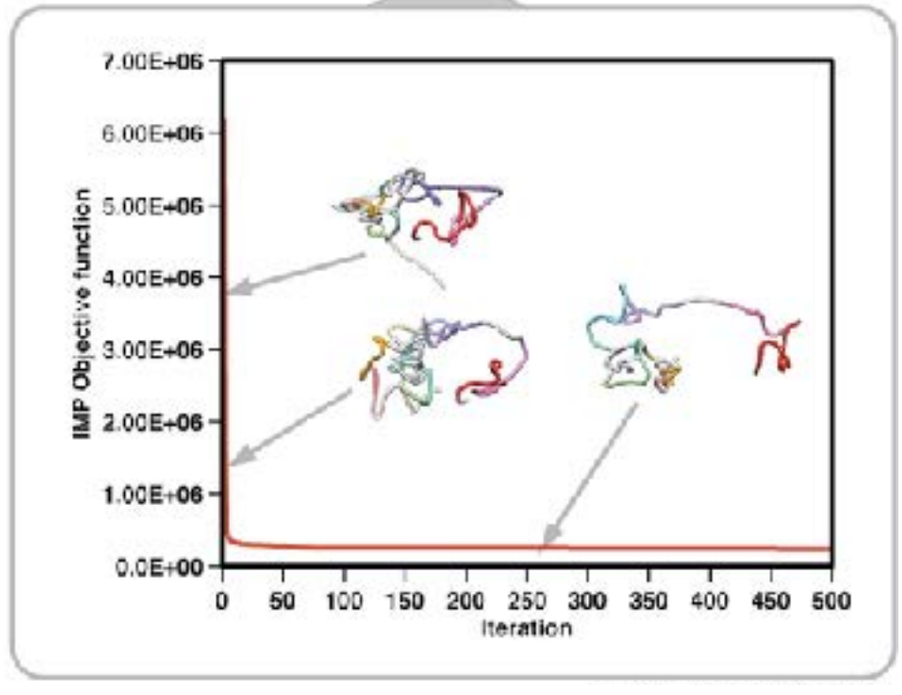


Experiments

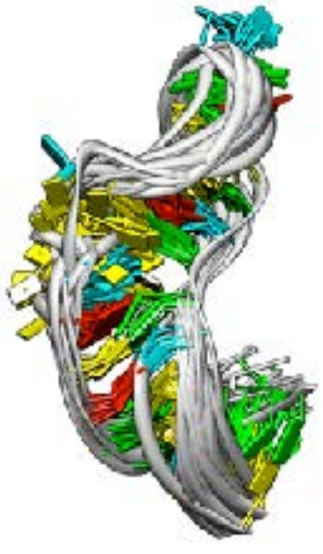


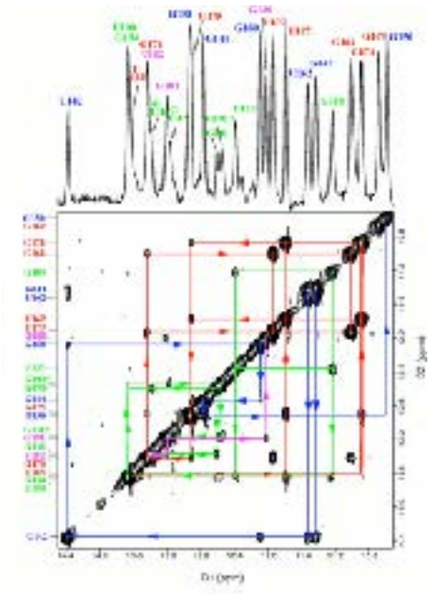






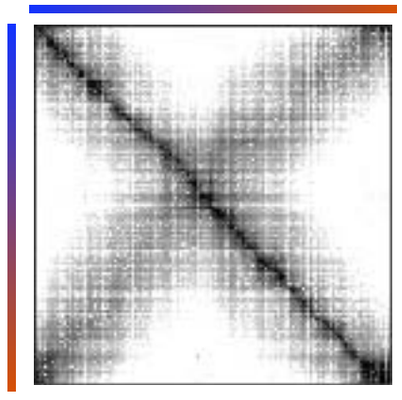
Computation





Biomolecular structure determination

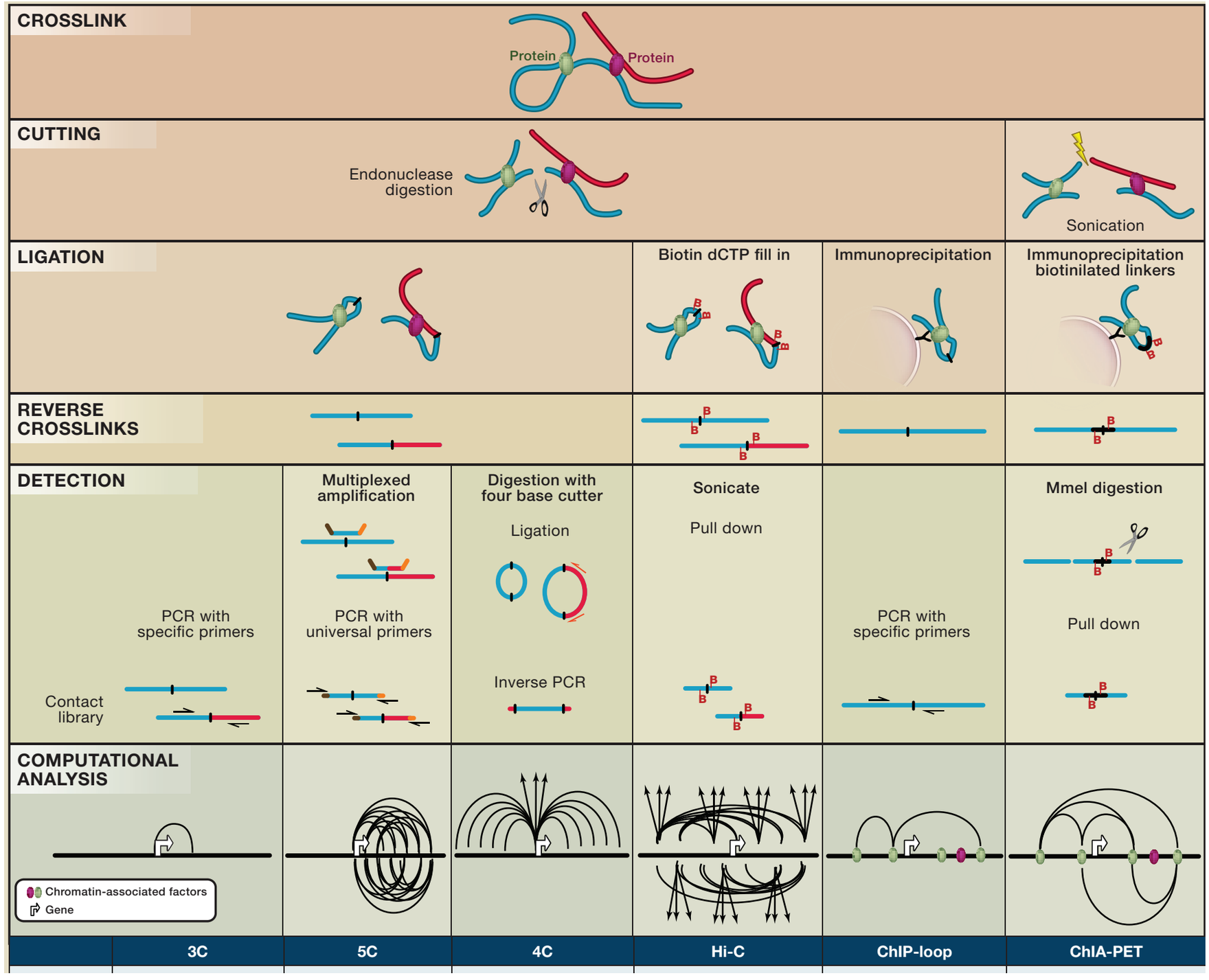
2D-NOESY data



Chromosome structure determination

5C data

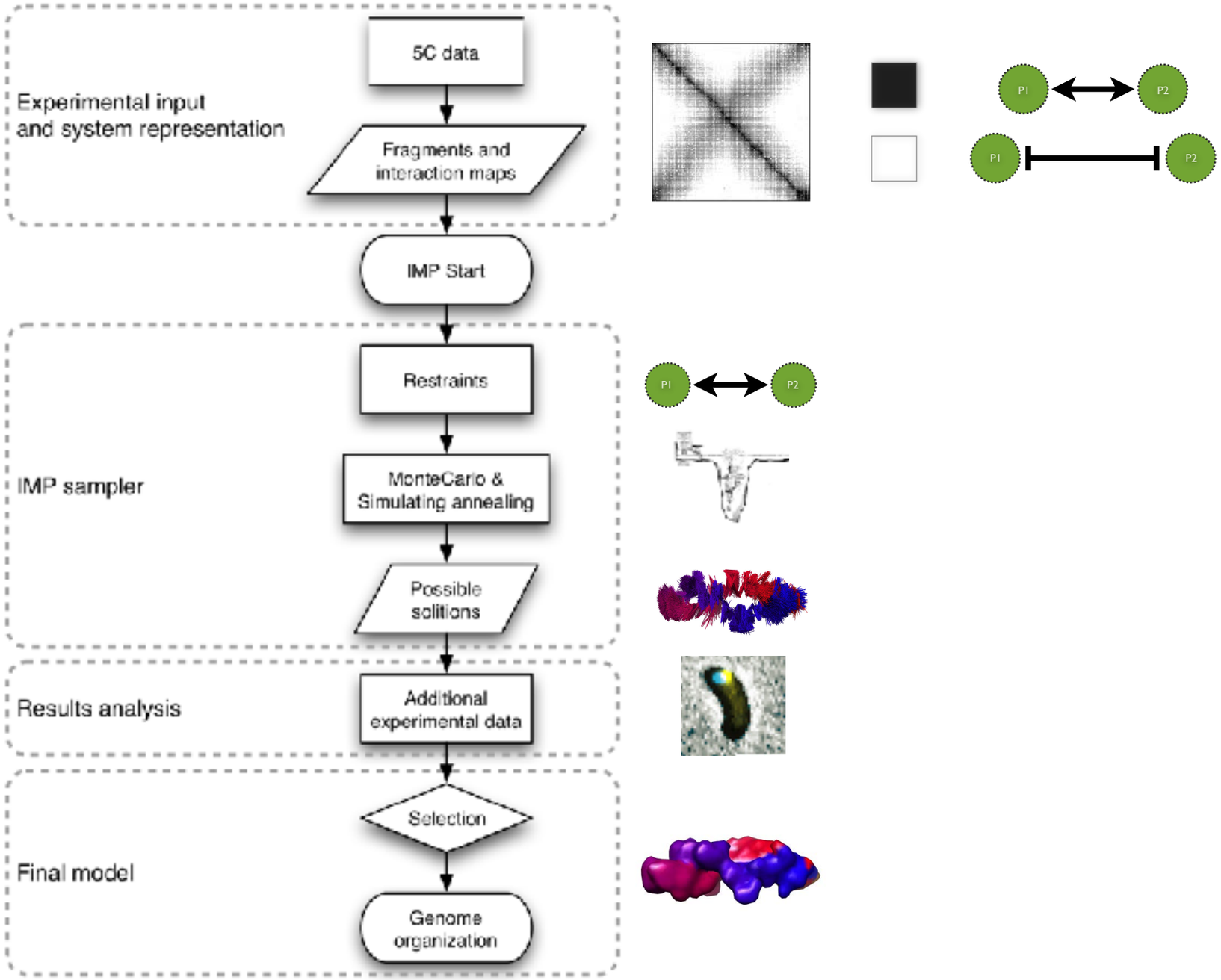
Chromosome Conformation Capture



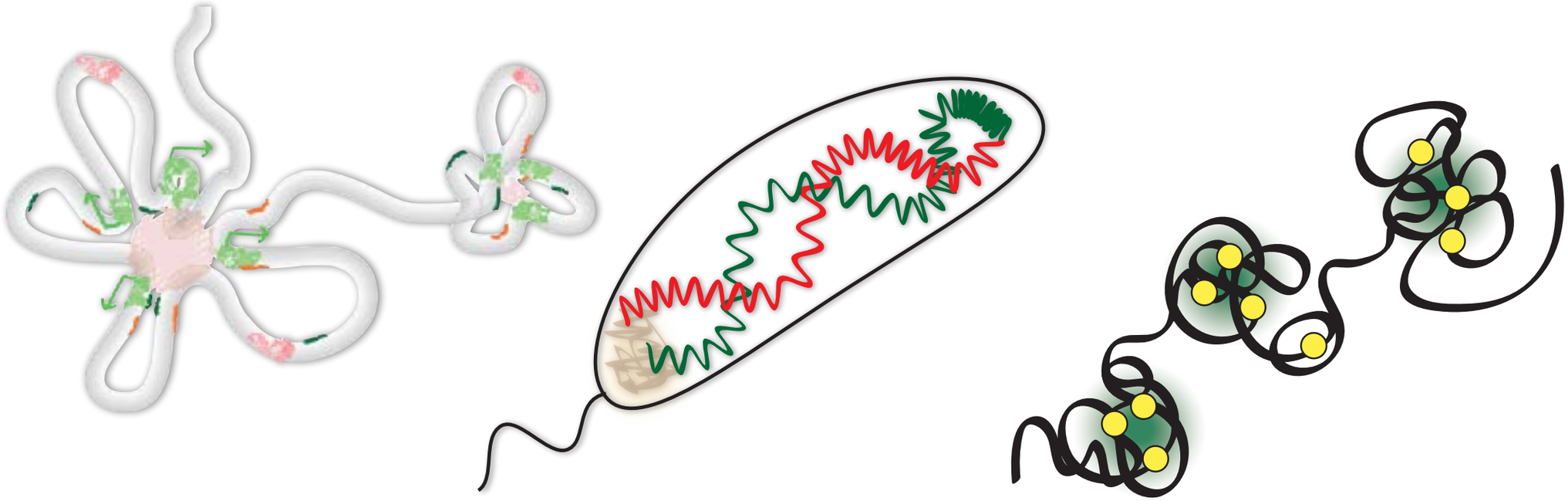
Hakim, O., & Misteli, T. (2012). SnapShot: Chromosome Confirmation Capture. Cell, 148(5), 1068–1068.e2.

Modeling 3D Genomes

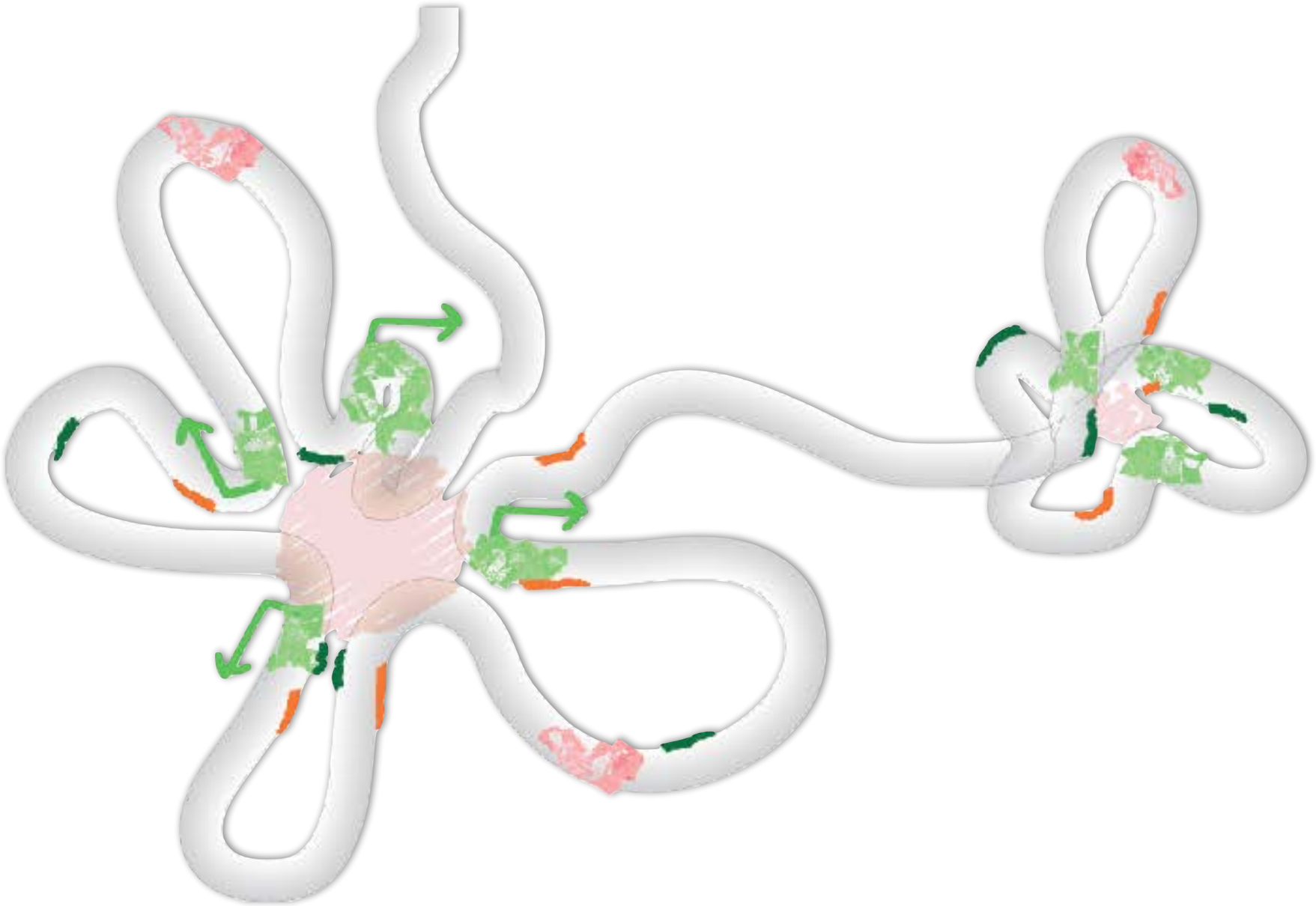
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).



Examples...



Human α-globin domain



Human α-globin domain

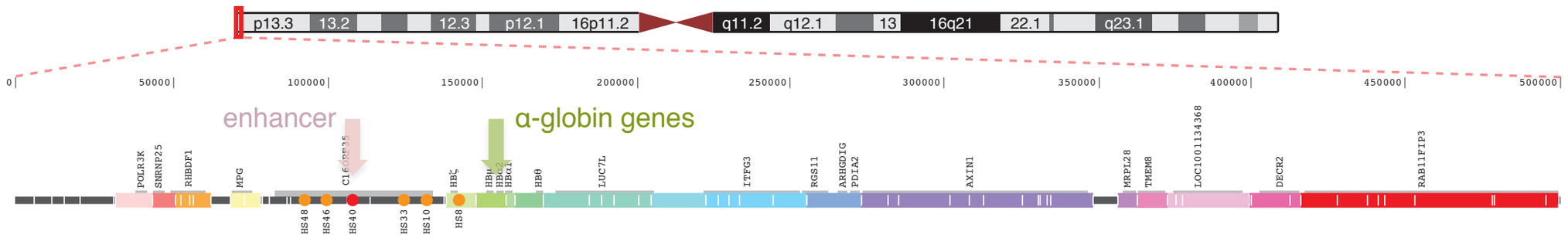
ENm008 genomic structure and environment

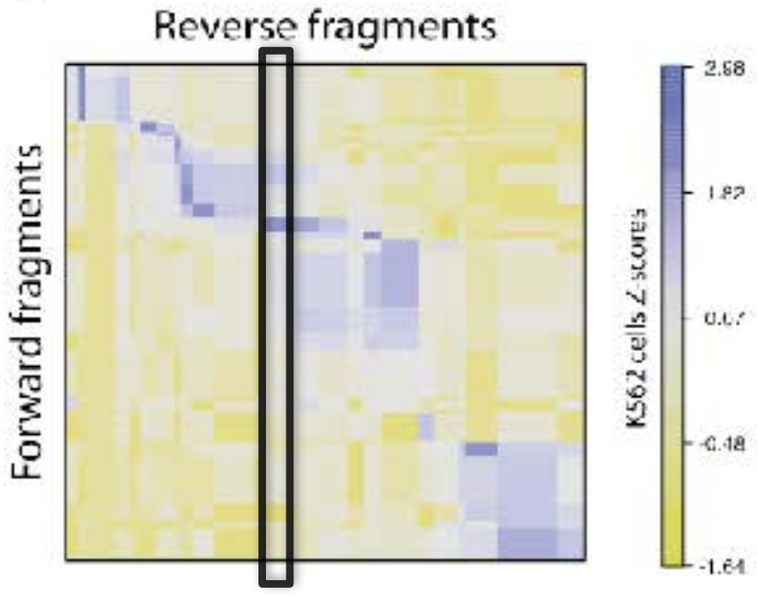
The ENCODE data for ENm008 region was obtained from the UCSC Genome Browser tracks for: RefSeq annotated genes, Affymetrix/CSHL expression data (Gingeras Group at Cold Spring Harbor), Duke/NHGRI DNaseI Hypersensitivity data (Crawford Group at Duke University), and Histone Modifications by Broad Institute ChIP-seq (Bernstein Group at Broad Institute of Harvard and MIT).

ENCODE Consortium. Nature (2007) vol. 447 (7146) pp. 799-816

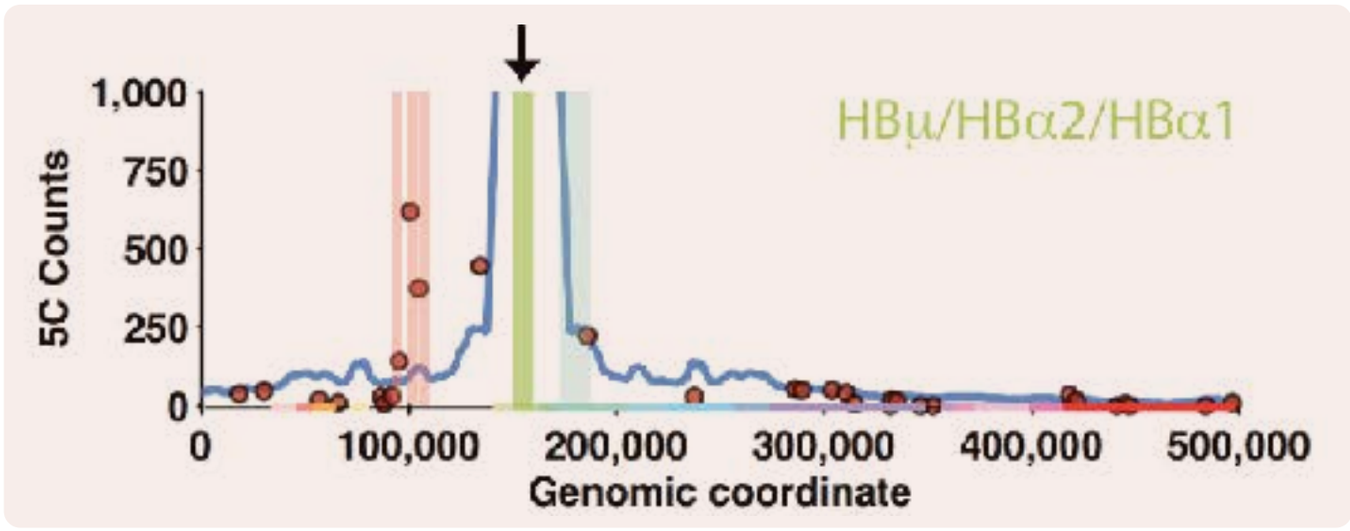
Human α-globin domain

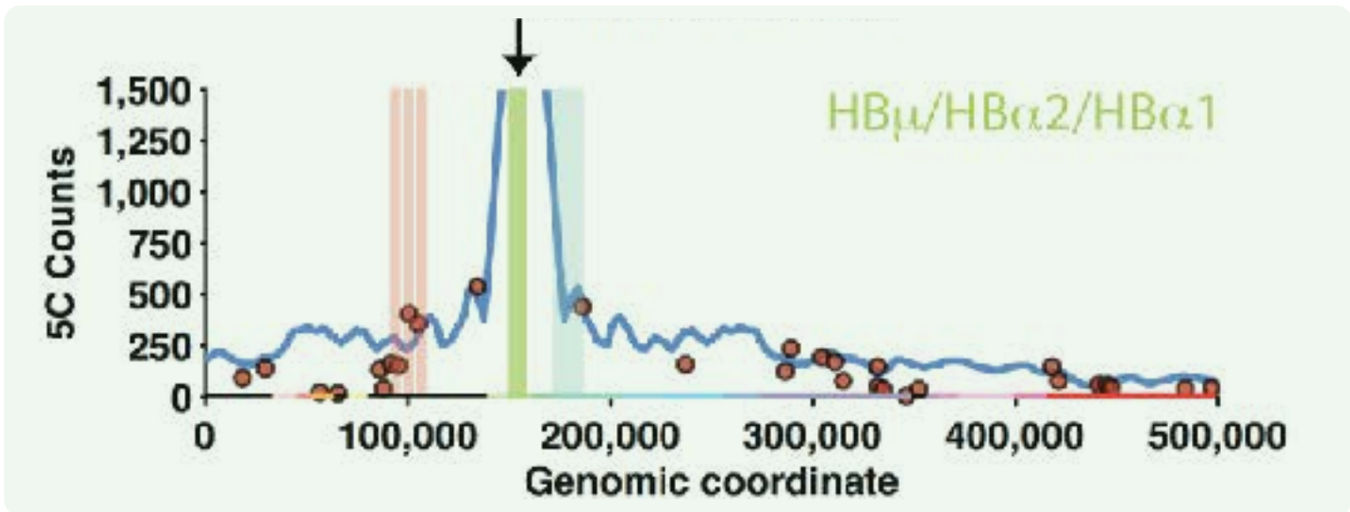
ENm008 genomic structure and environment





K562 cells:
α-globin genes active

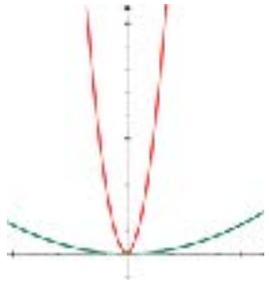




Representation

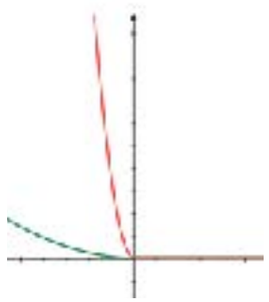
Harmonic

$$H_{i,j} = k\left(d_{i,j} - d^0_{i,j}\right)^2$$



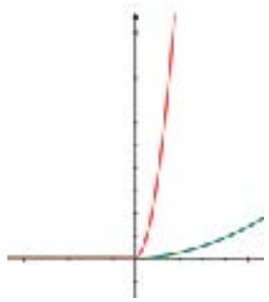
Harmonic Lower Bound

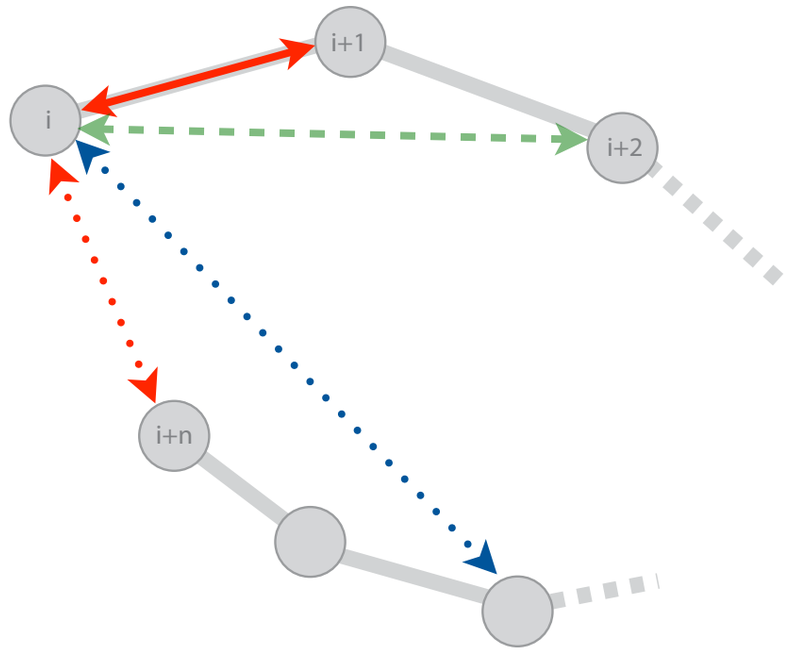
$$\begin{cases} if\ \ d_{i,j} \le d^0_{i,j}; \quad lbH_{i,j} = k\left(d_{i,j} - d^0_{i,j}\right)^2 \\ if\ \ d_{i,j} > d^0_{i,j}; \quad lbH_{i,j} = 0 \end{cases}$$



Harmonic Upper Bound

$$\begin{cases} if\ \ d_{i,j} \ge d^0_{i,j}; \quad ubH_{i,j} = k\left(d_{i,j} - d^0_{i,j}\right)^2 \\ if\ \ d_{i,j} < d^0_{i,j}; \quad ubH_{i,j} = 0 \end{cases}$$

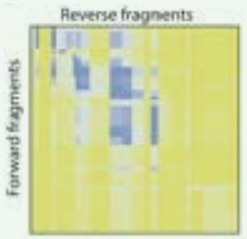


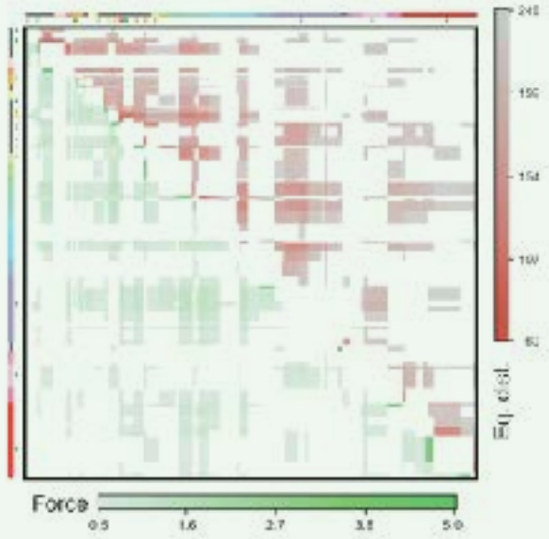


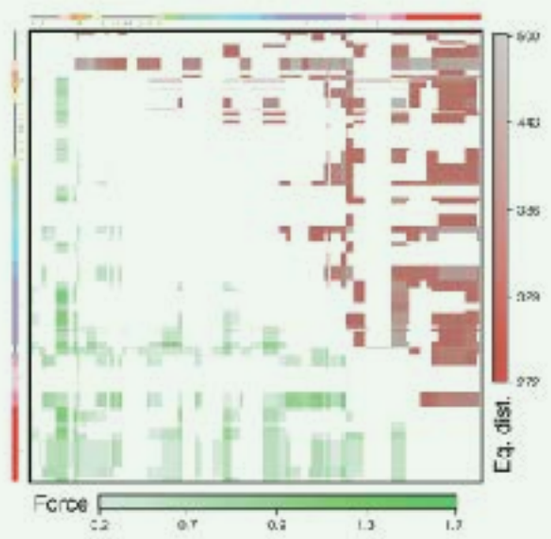
Scoring

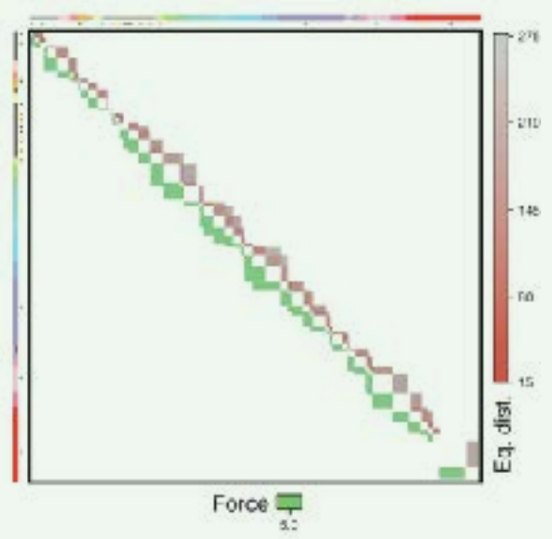
GM12878

70 fragments
1,520 restraints





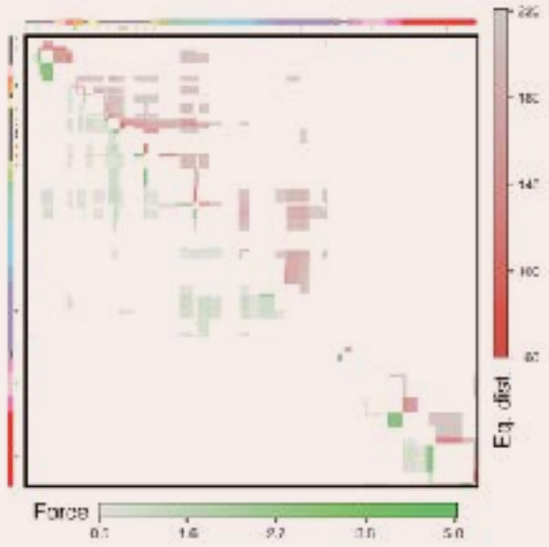


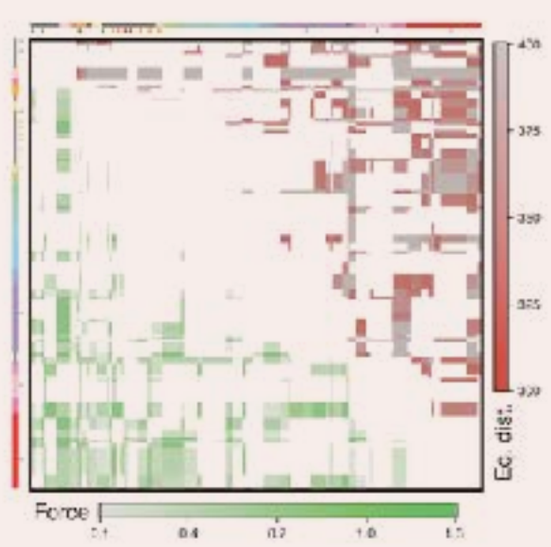


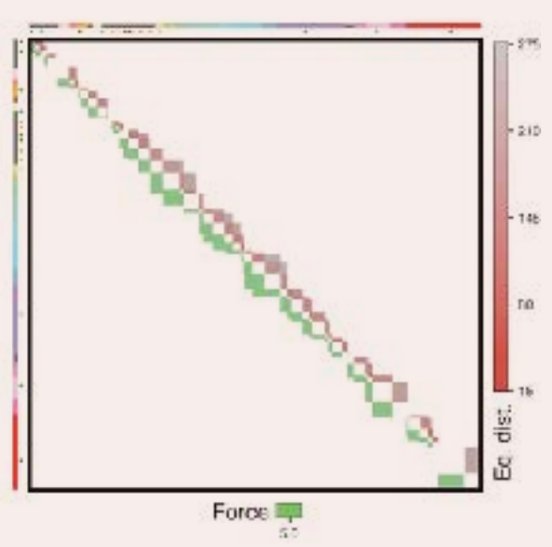
Harmonic | Harmonic Lower Bound | Harmonic Upper Bound

K562

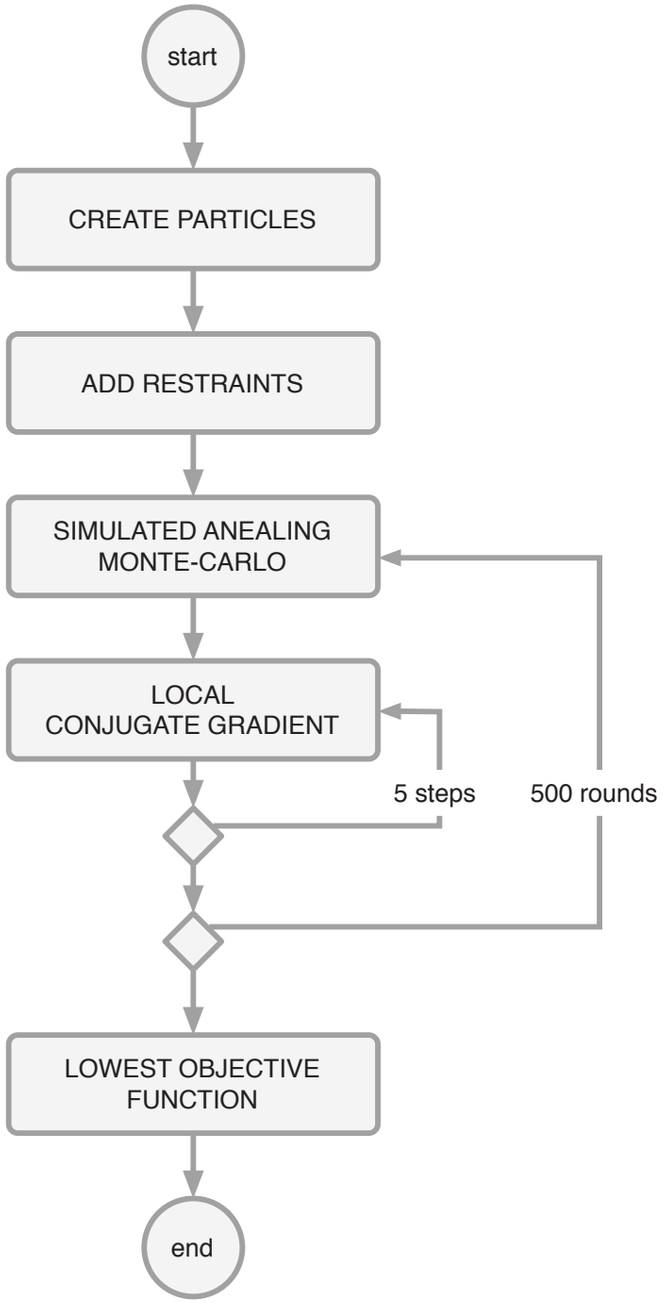
70 fragments
1,049 restraints

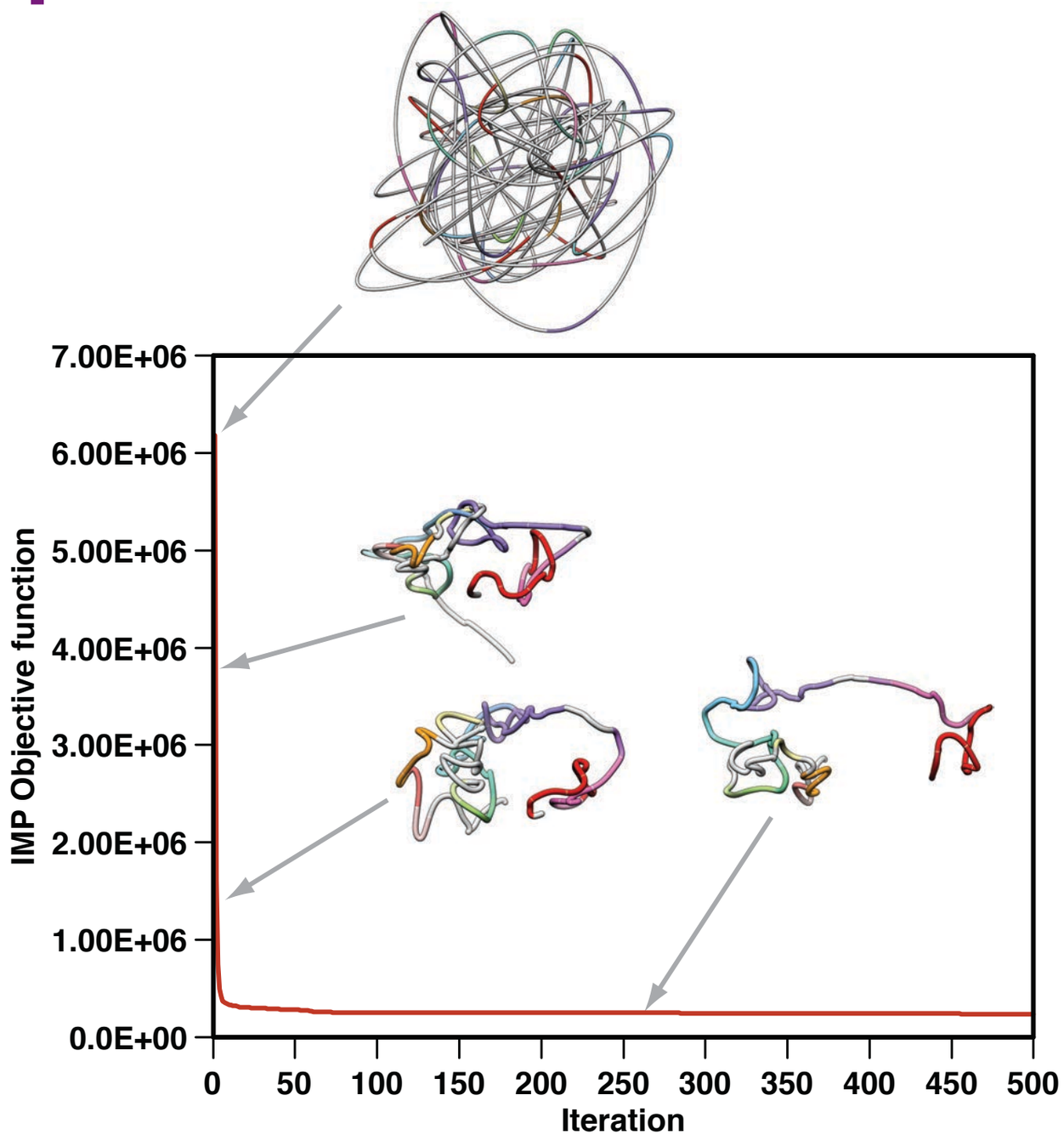




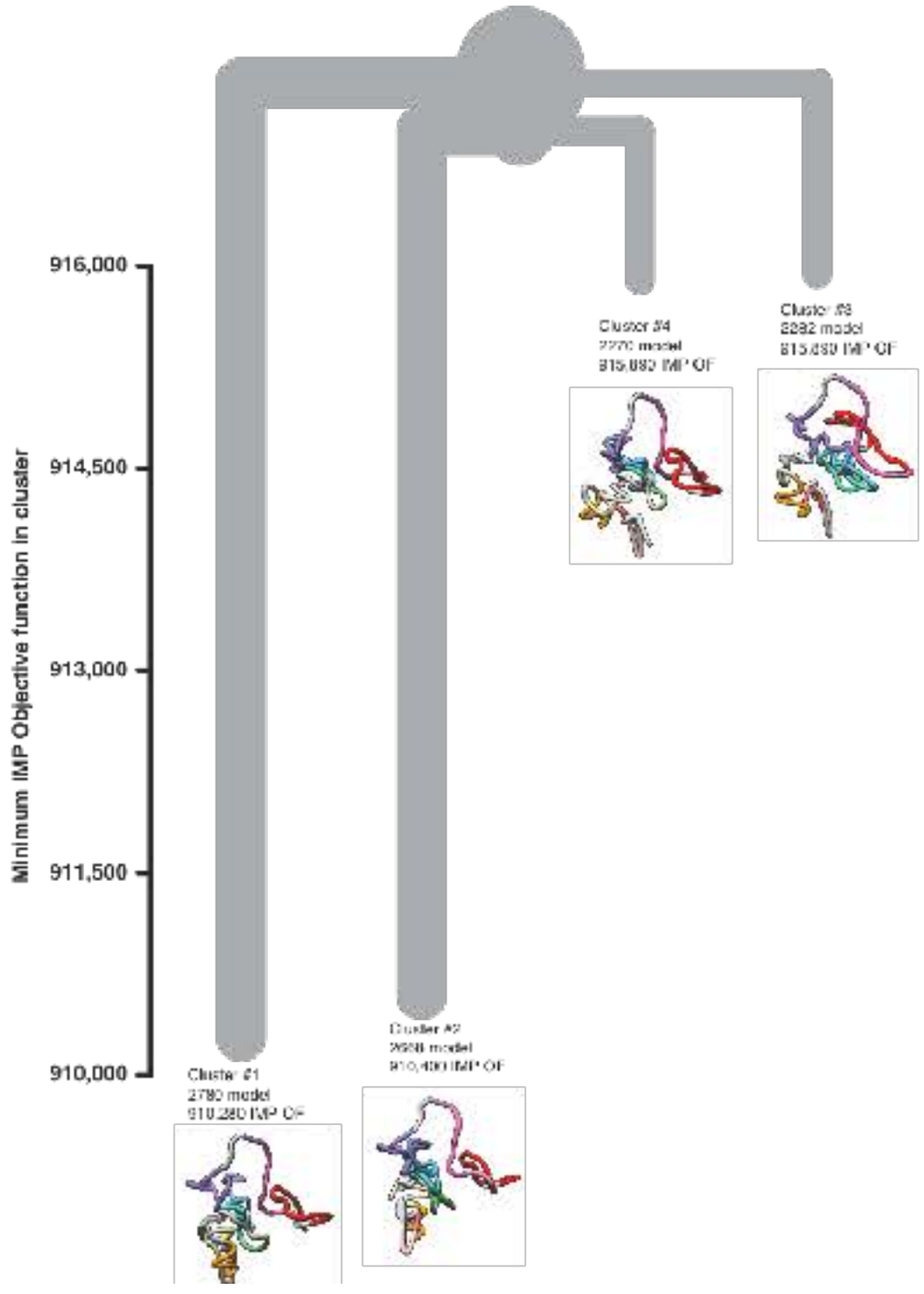


Optimization



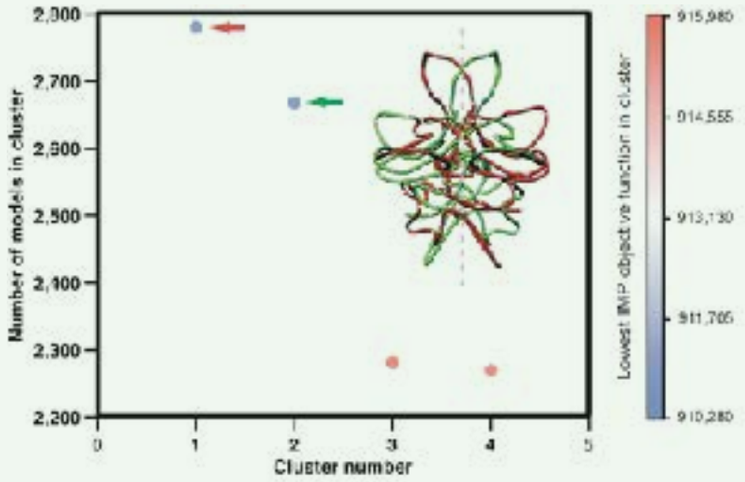


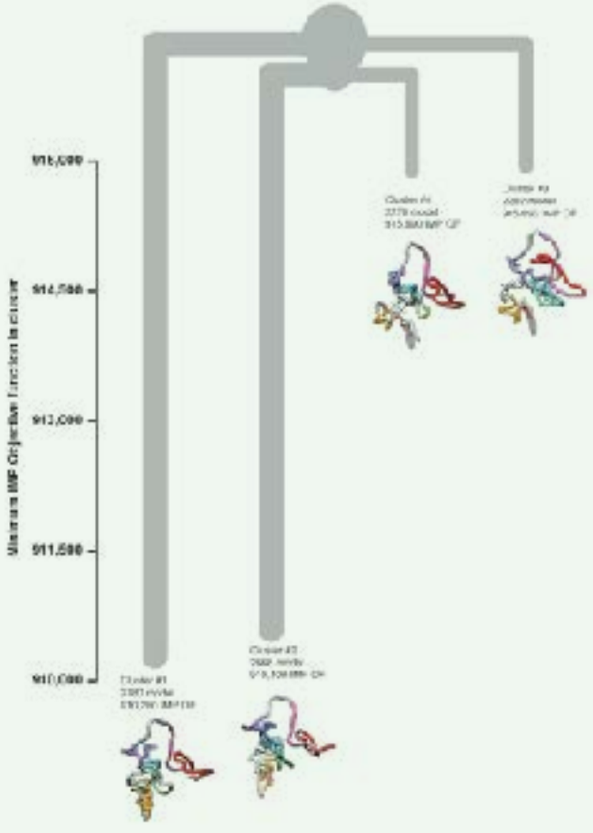
Clustering


916,000
914,500
913,000
911,500
910,000
Minimum IMP Objective function in cluster
Cluster #4
2770 model
915,082 IMP OF
Cluster #3
2282 model
915,089 IMP OF
Cluster #1
2780 model
910,280 IMP OF
Cluster #2
2658 model
910,480 IMP OF

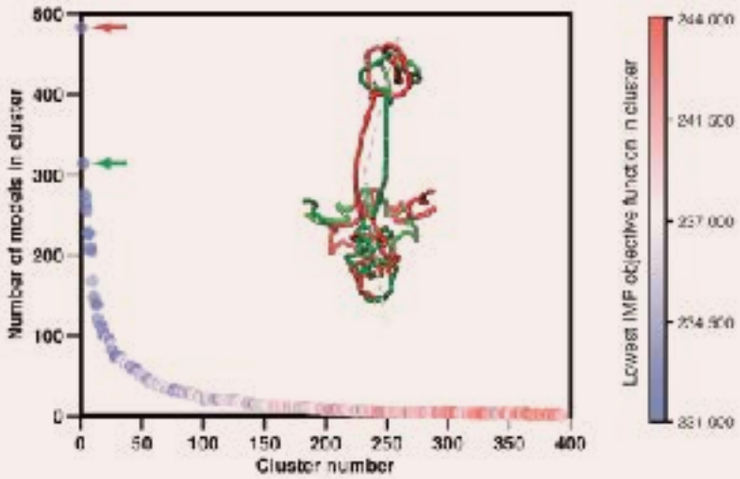
Not just one solution

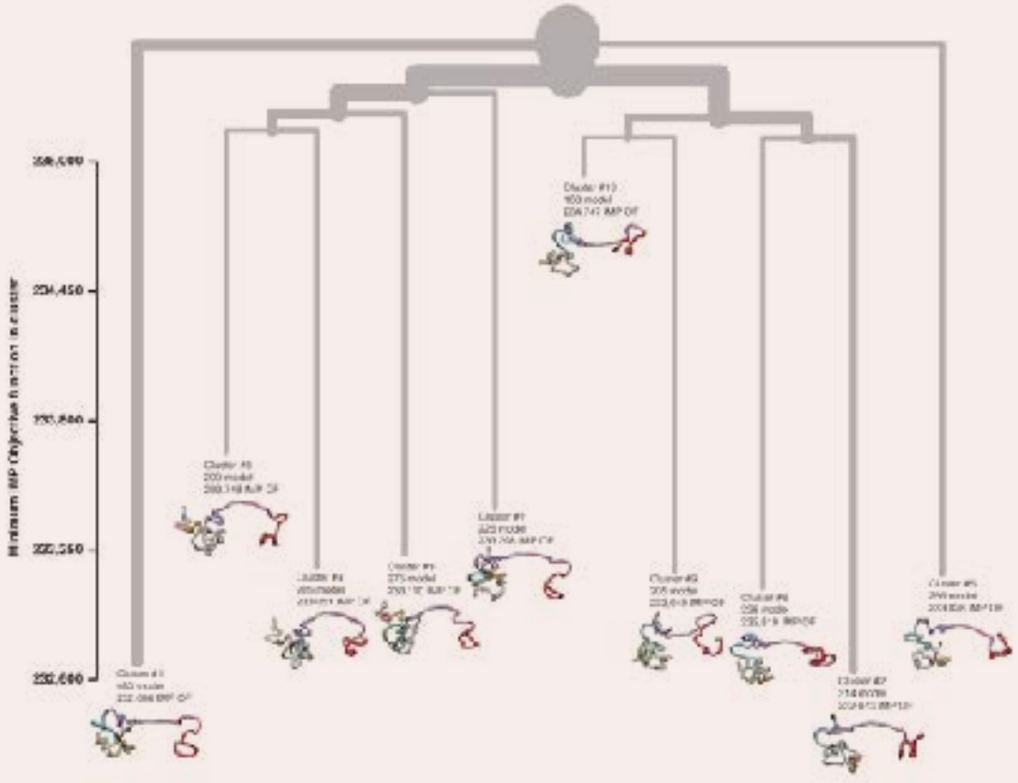
GM12878





K562

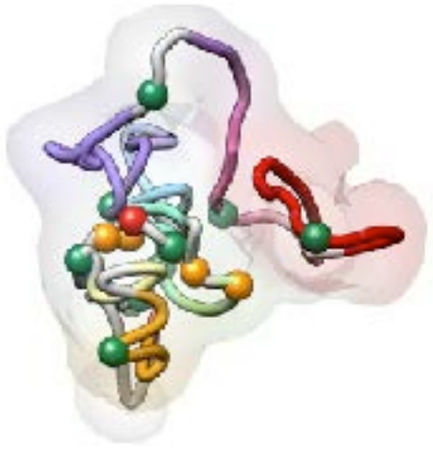




Regulatory elements

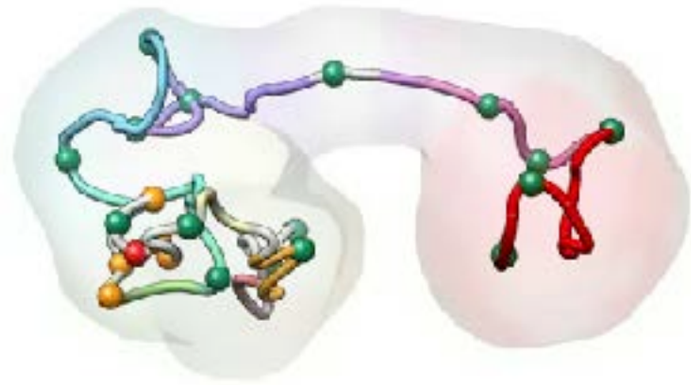
GM12878

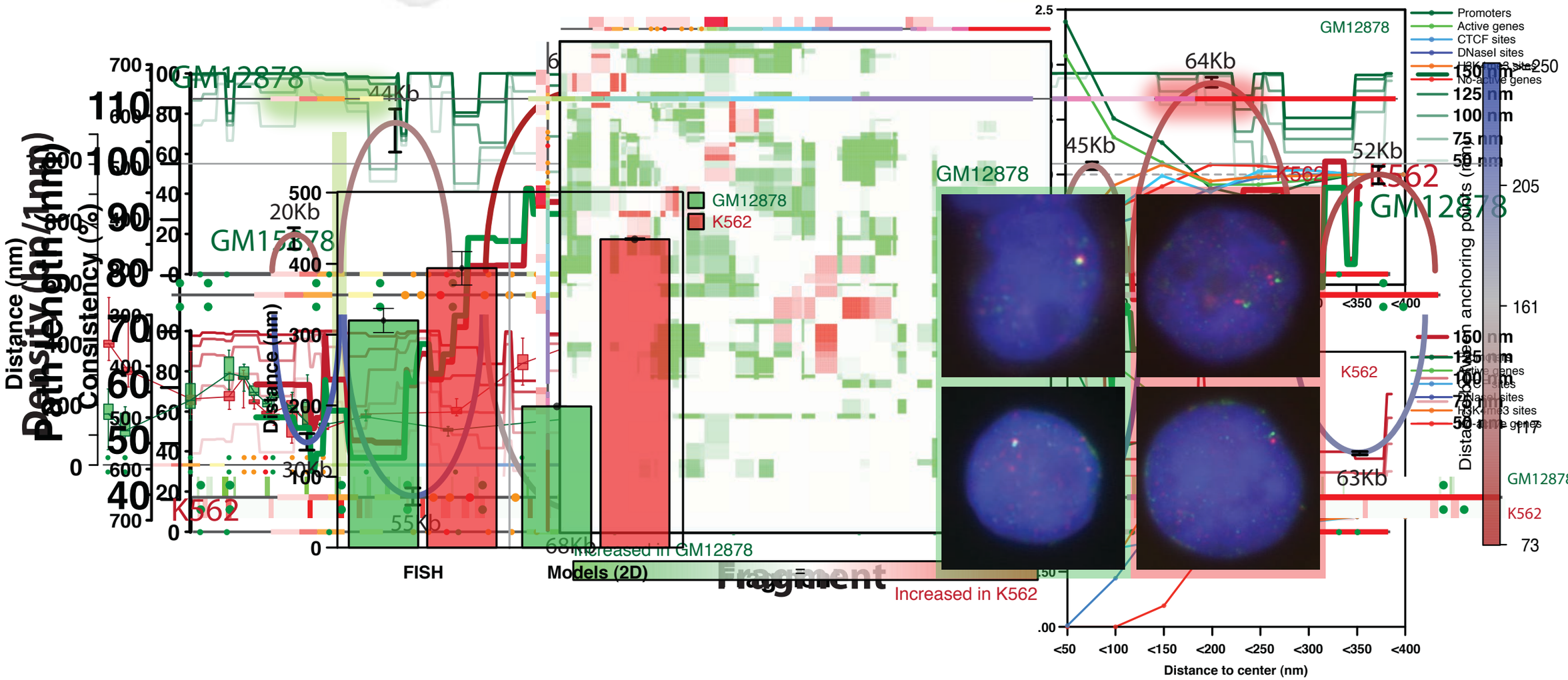
Cluster #1
2780 model



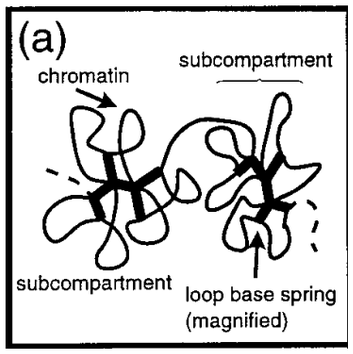
K562

Cluster #2
314 model

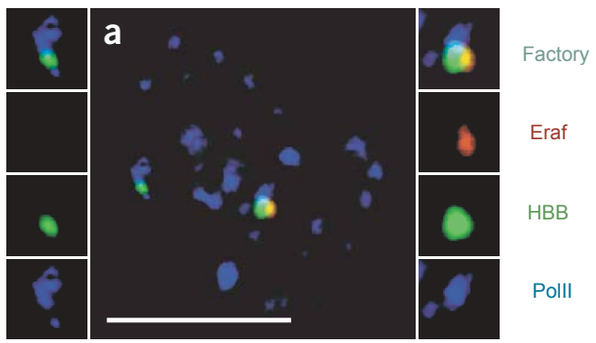




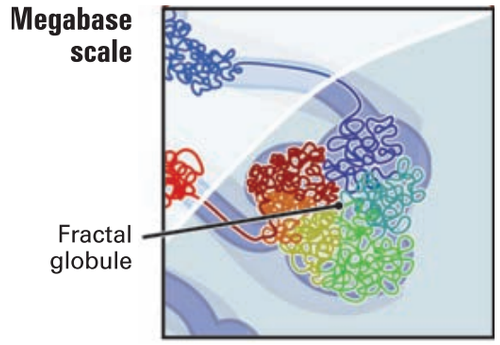
The "Chromatin Globule" model



Münkel et al. JMB (1999)



Osborne et al. Nat Genet (2004)

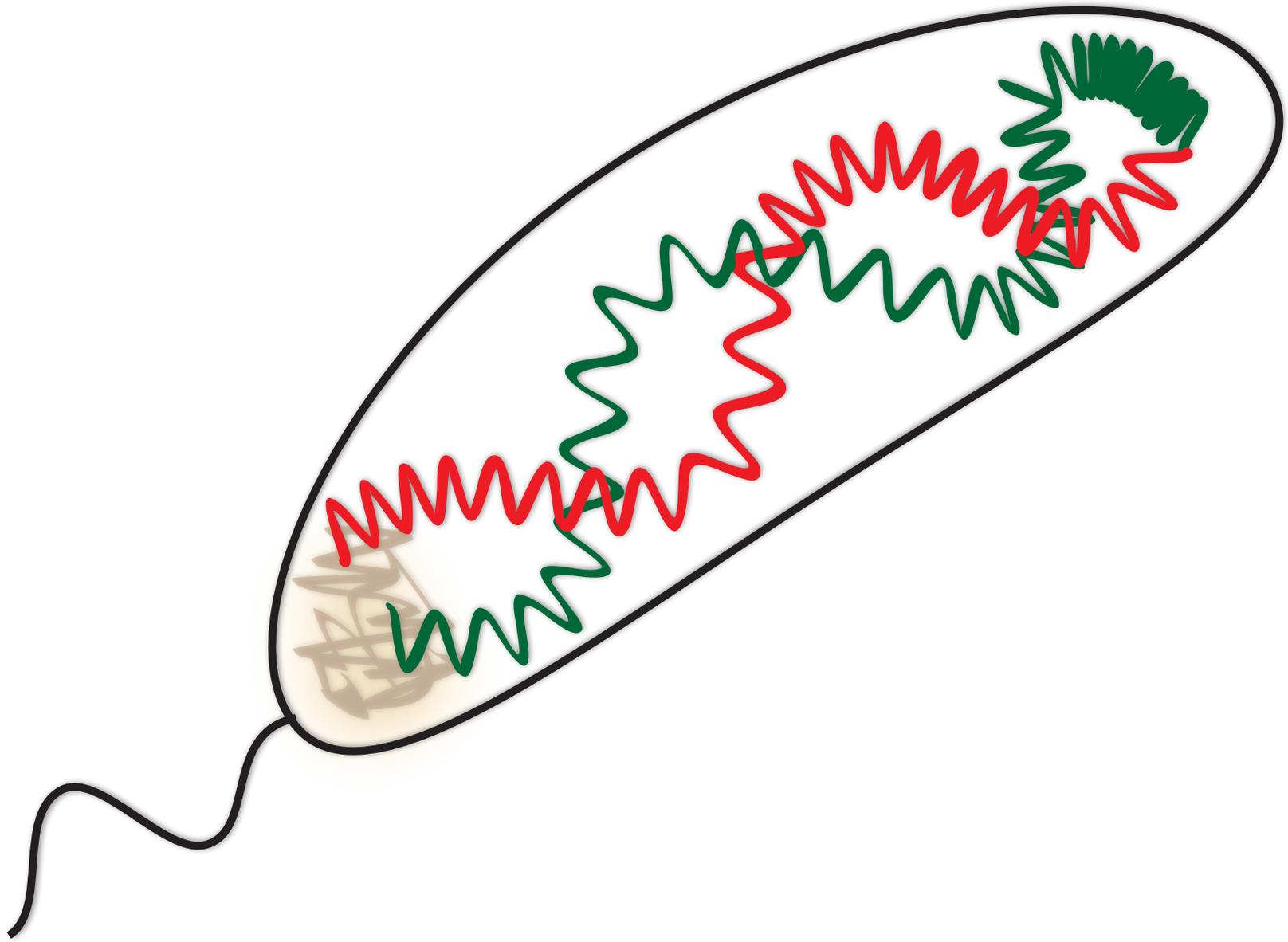


Lieberman-Aiden et al. Science (2009)

D. Baù *et al.* **Nat Struct Mol Biol** (2011) 18:107-14

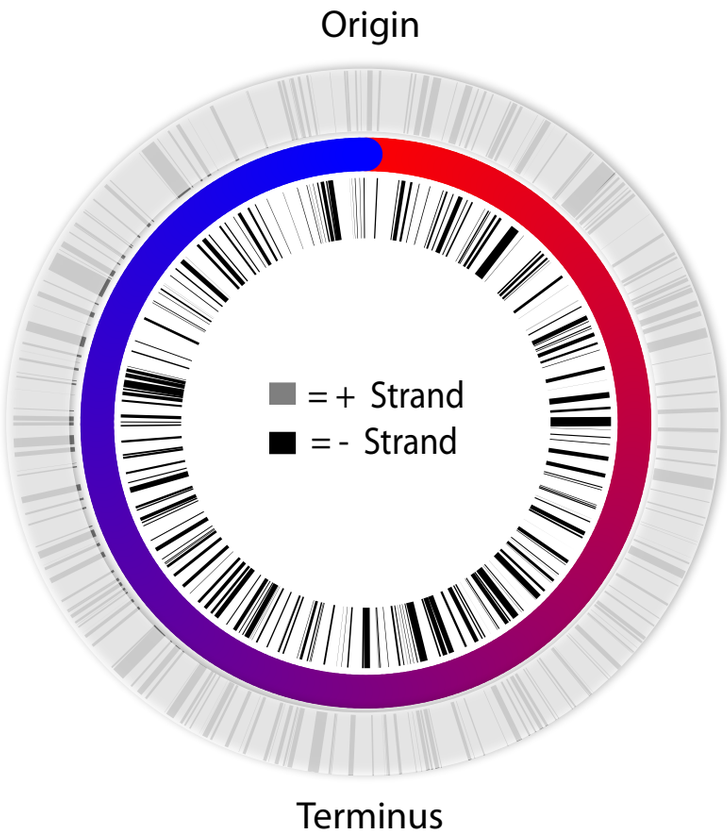
A. Sanyal *et al.* **Current Opinion in Cell Biology** (2011) 23:325–33.

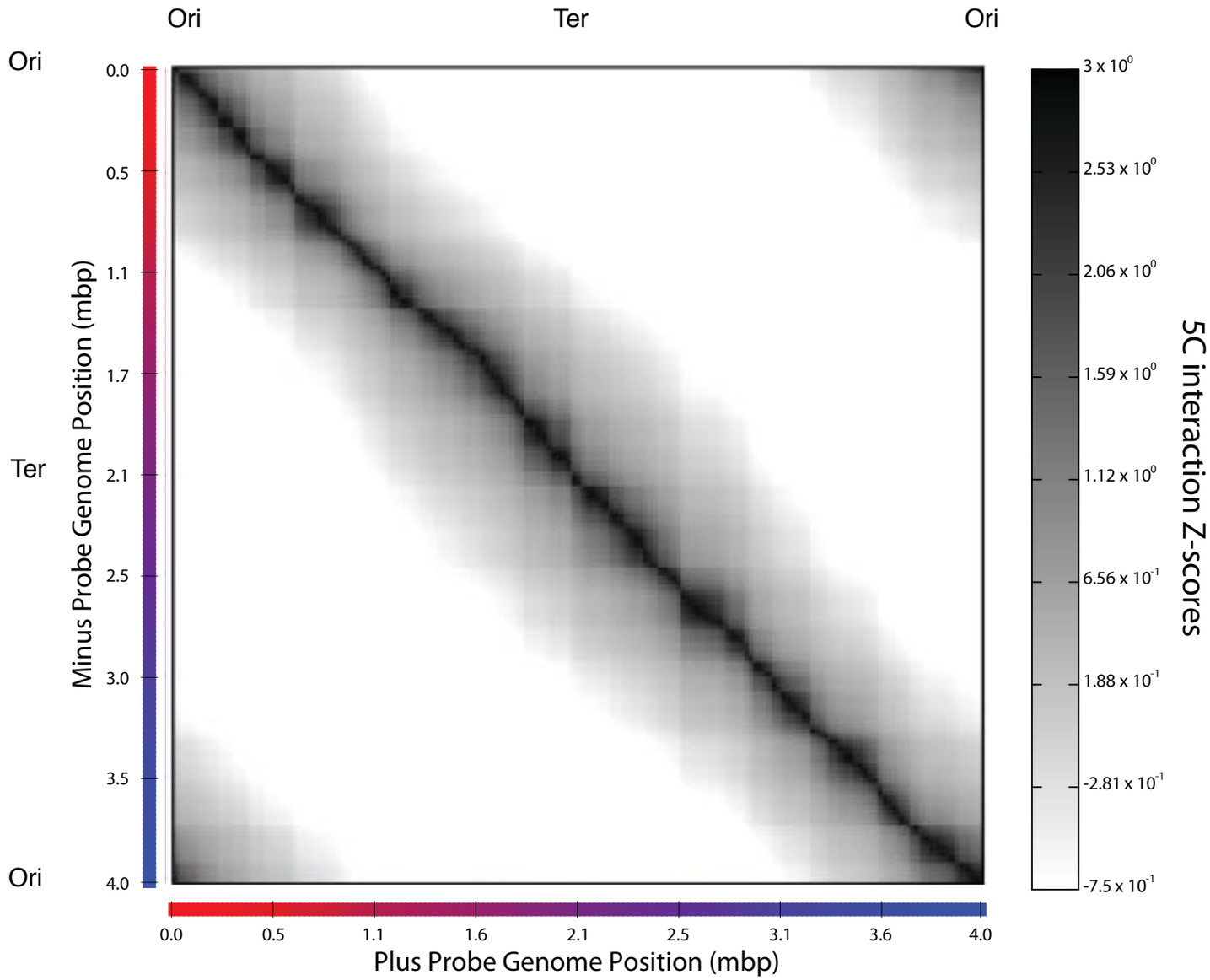
Caulobacter crescentus genome



The 3D architecture of Caulobacter Crescentus

4,016,942 bp & 3,767 genes





169 5C primers on + strand

170 5C primers on – strand

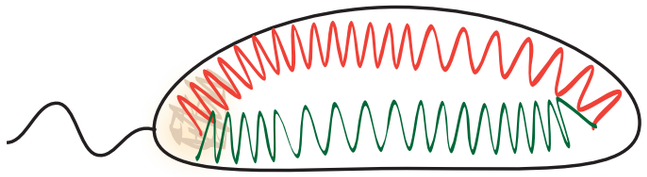
28,730 chromatin interactions

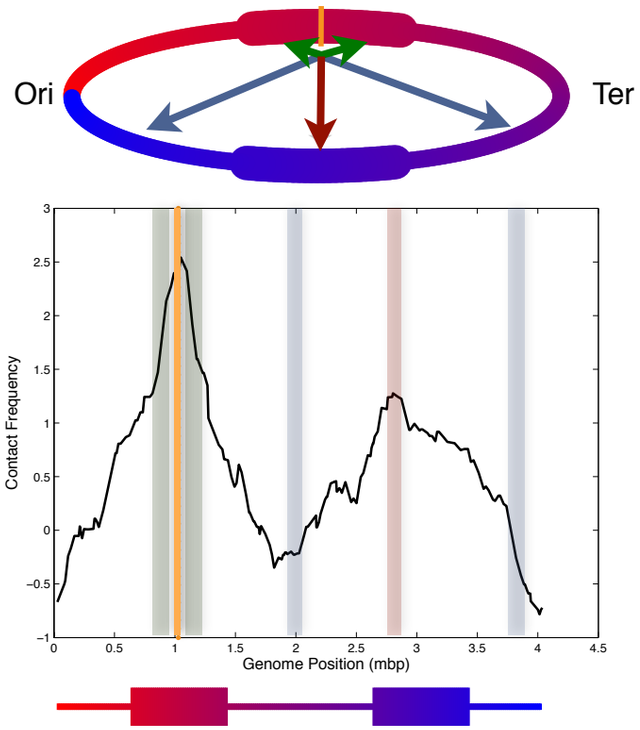
~13Kb

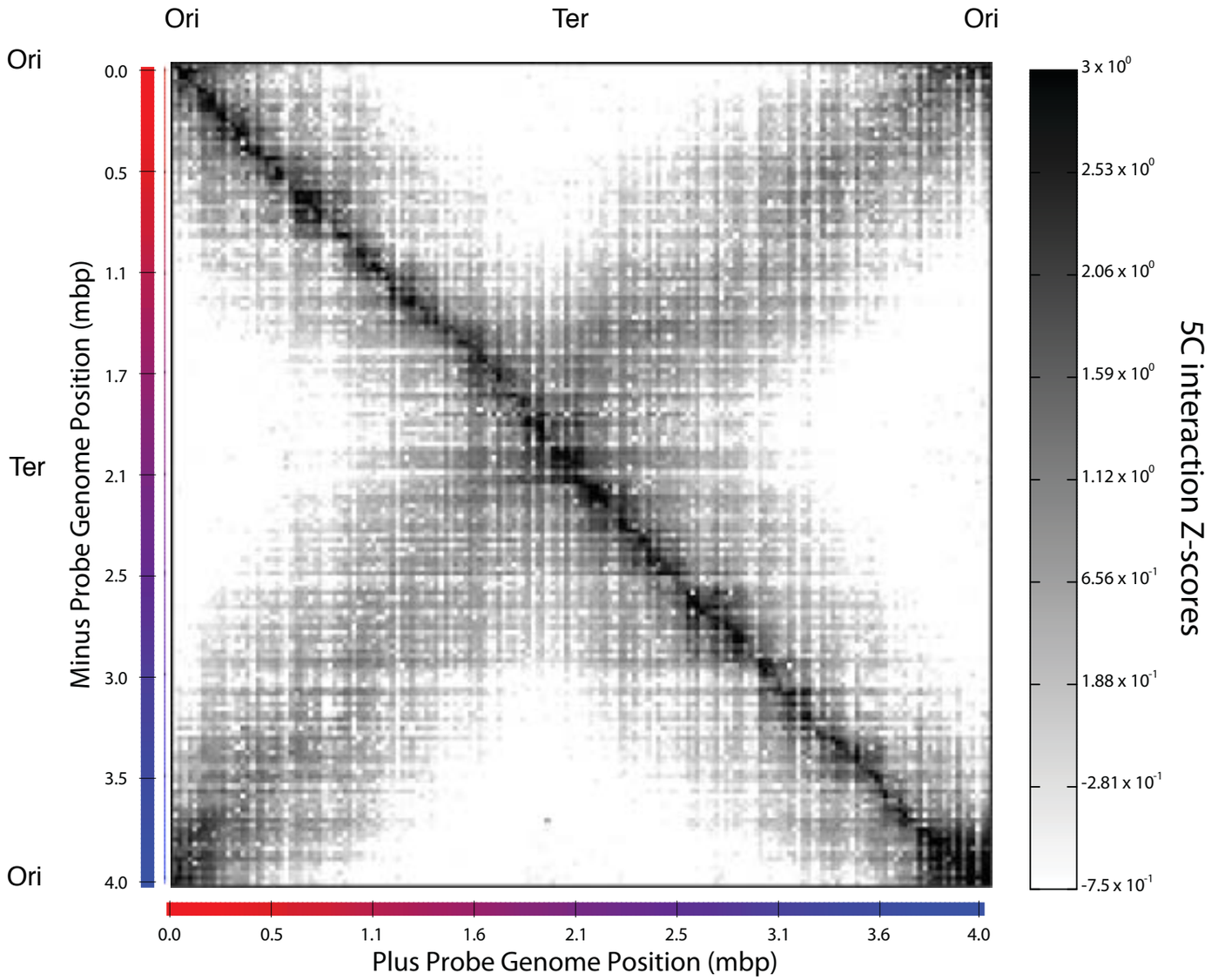
cnag CRG Centre for Genomic Regulation

5C interaction matrix

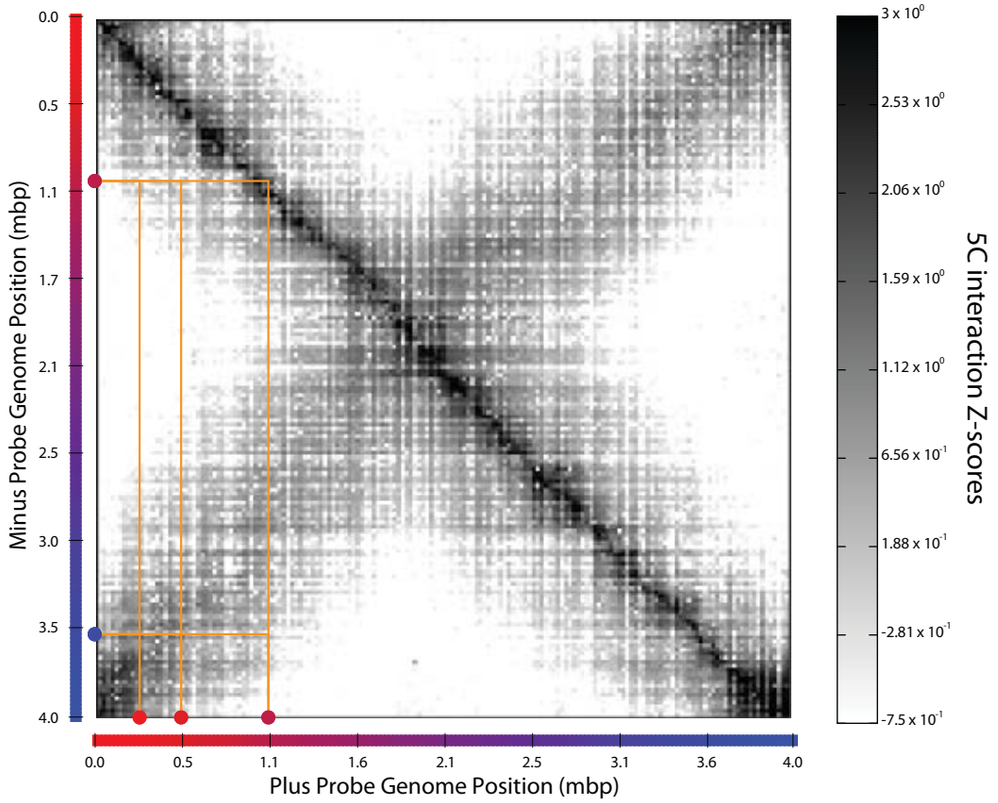
ELLIPSOID for Caulobacter cresentus

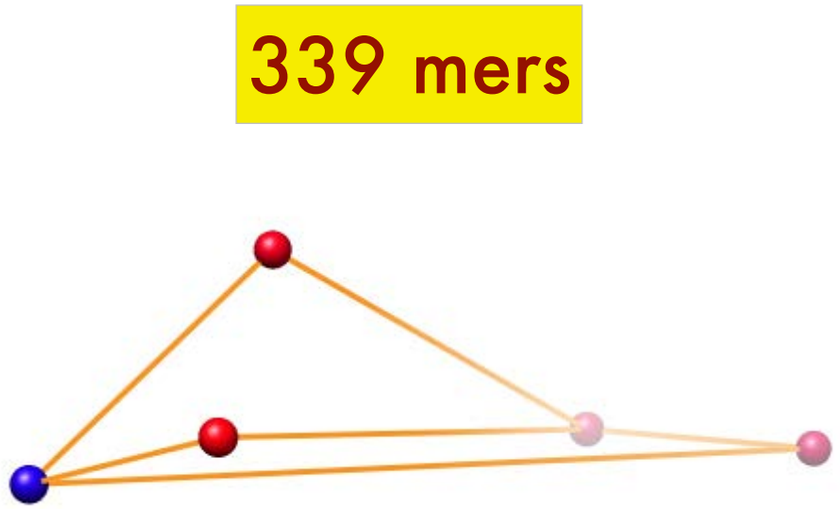


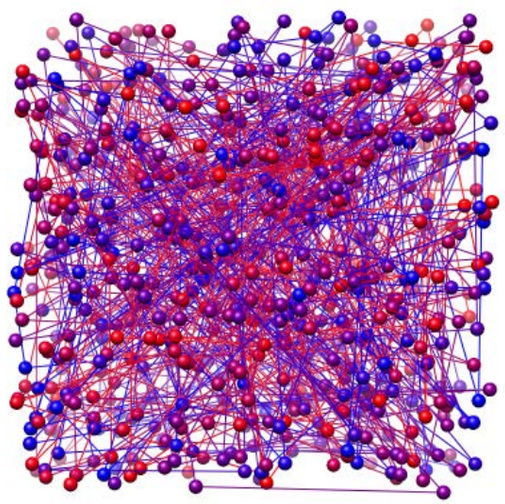


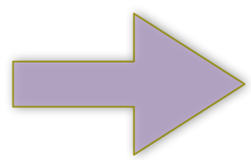


3D model building with the 5C + IMP approach







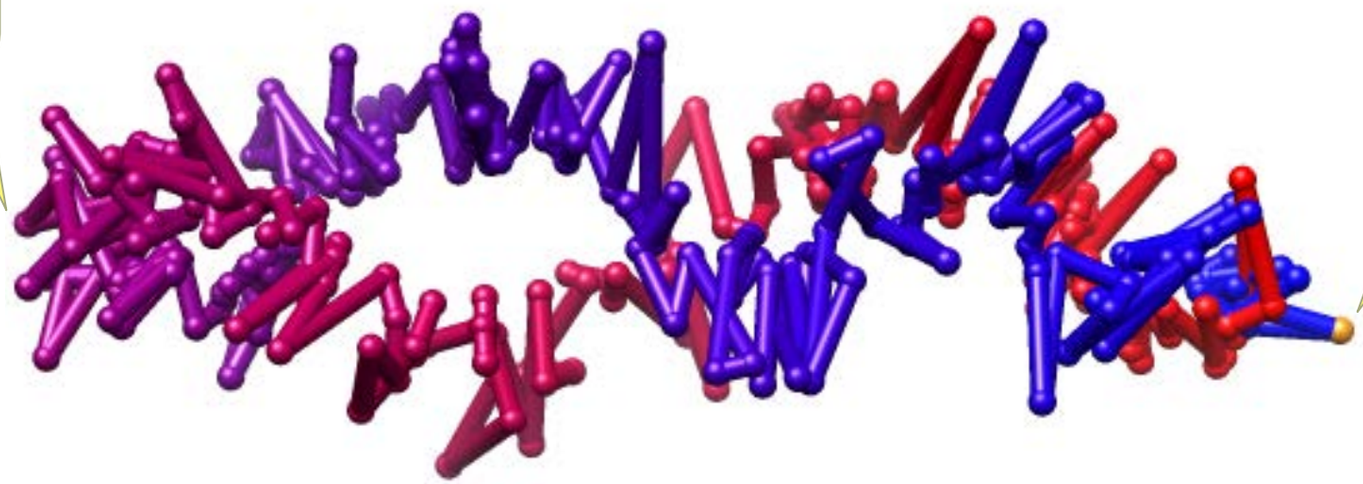


Genome organization in Caulobacter crescentus

Arms are helical

Resolution

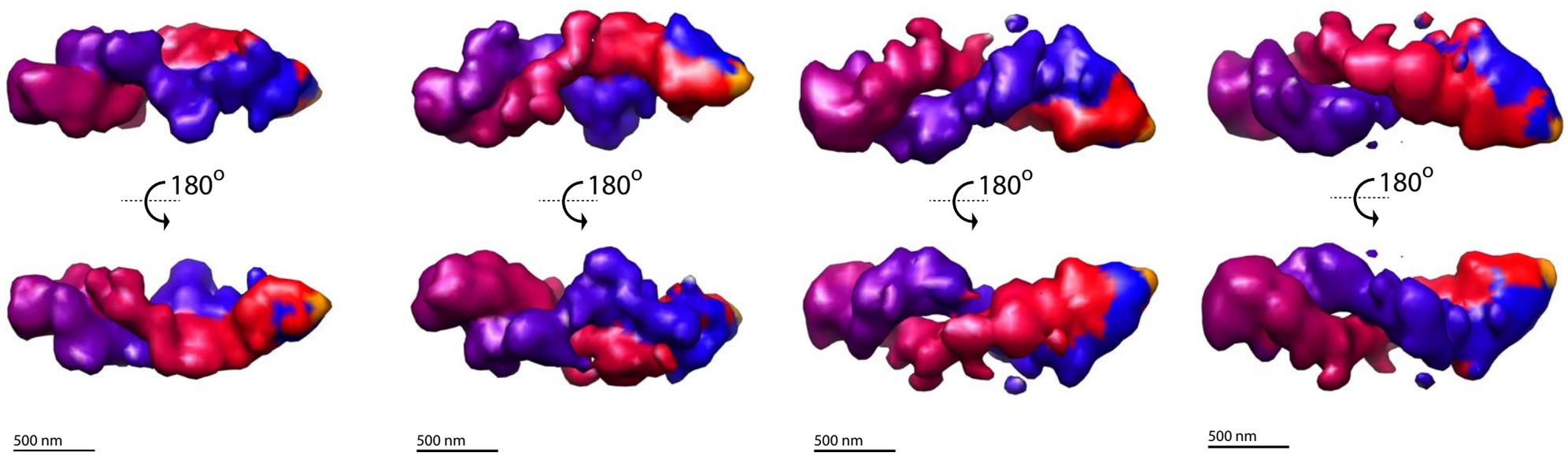
dif site 47±17Kb from Ter



Centromer-like

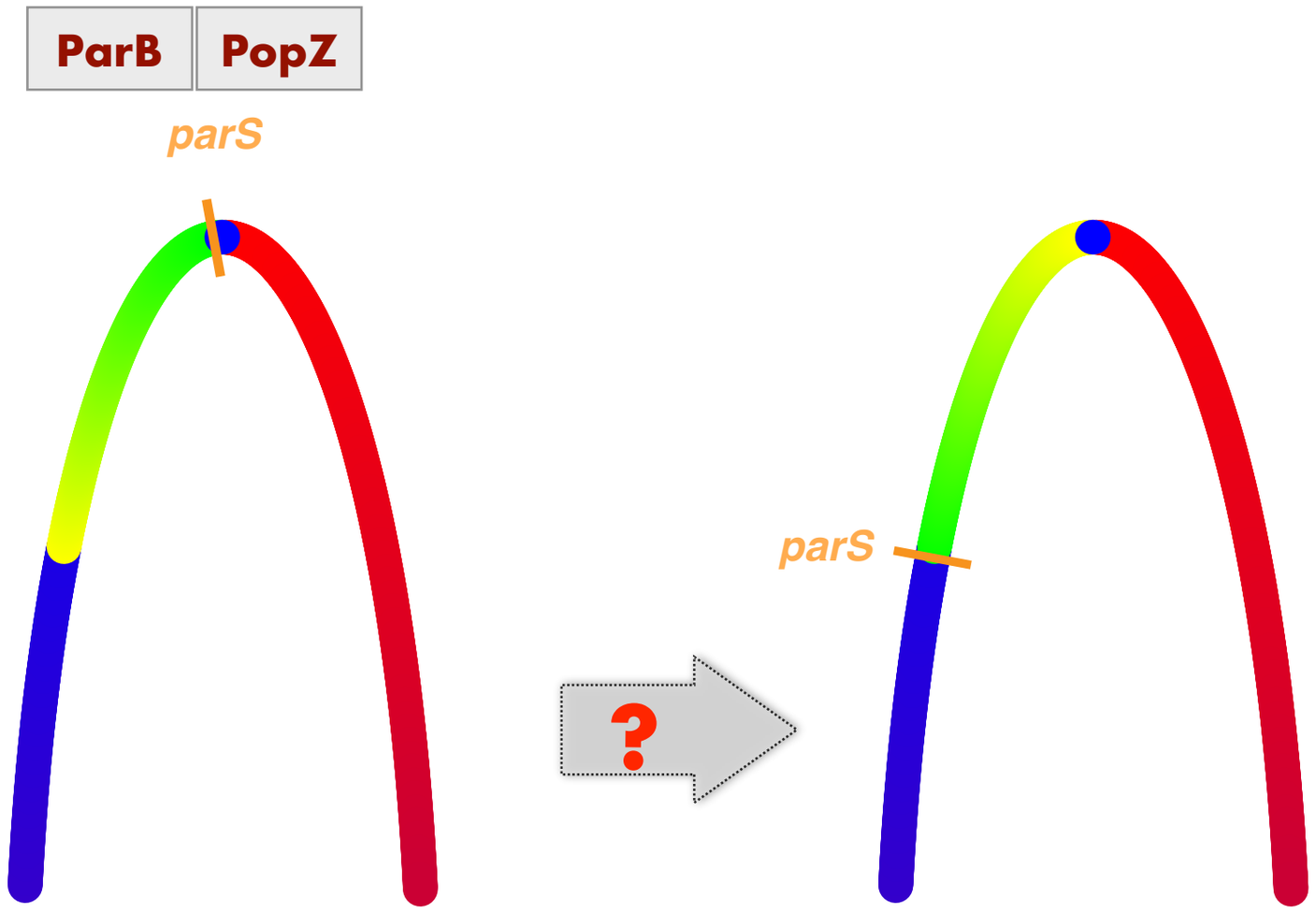
parS sites 25±17Kb from Ori

Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4



MIRRORS!

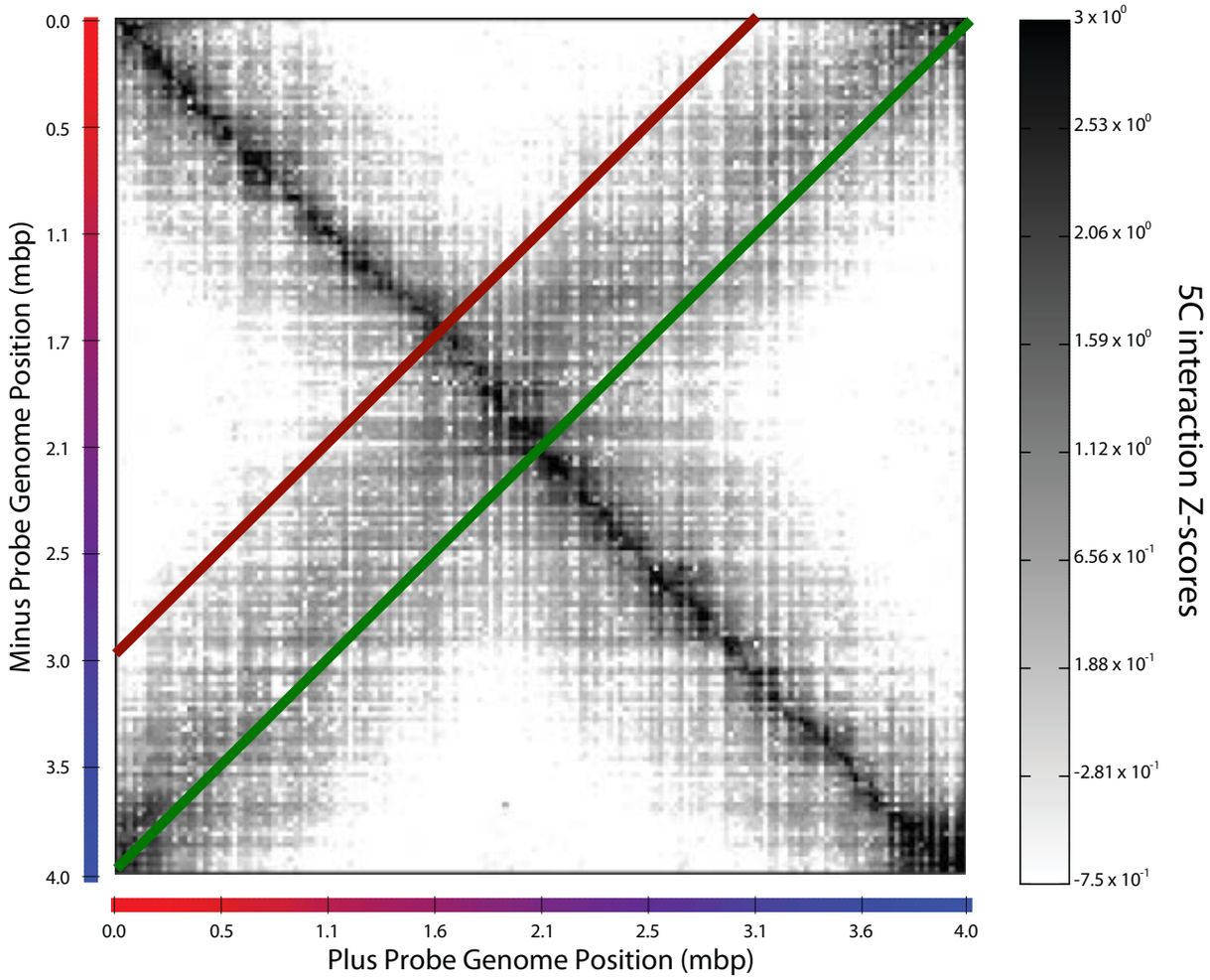
Moving the parS sites 400 Kb away from Ori

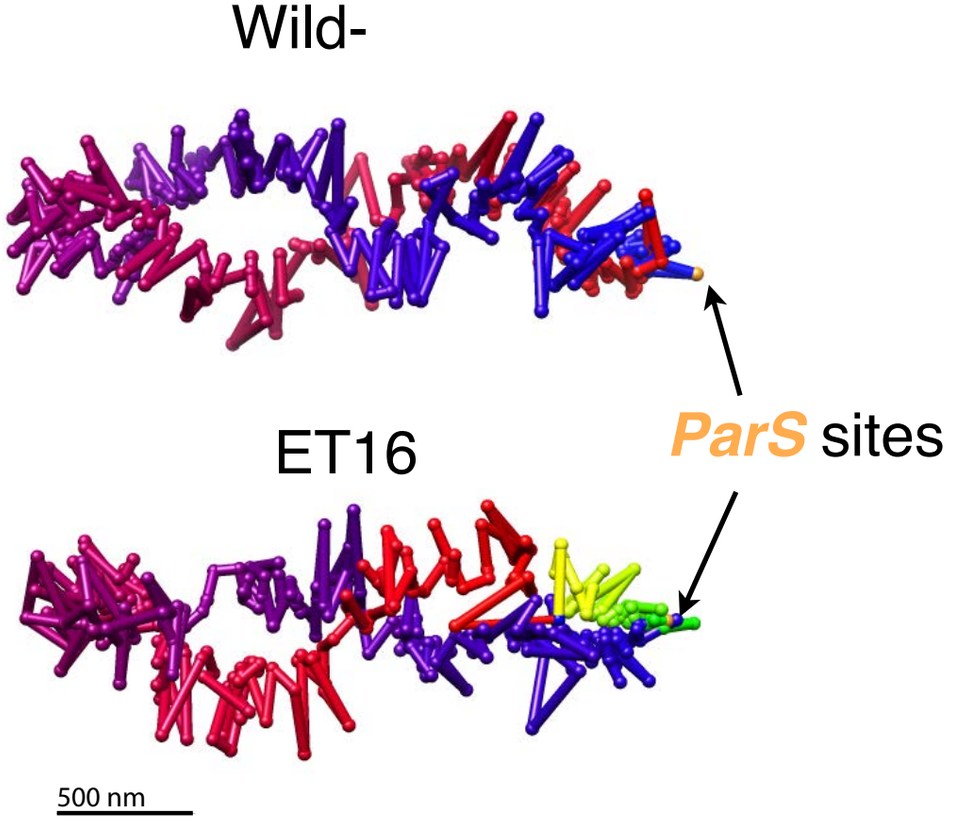


Wild-type

ET166

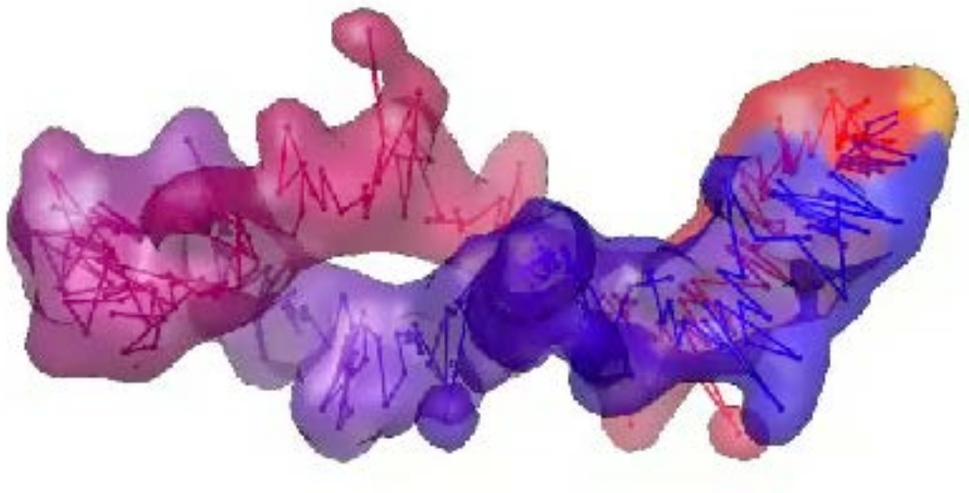
Moving the parS sites results in whole genome rotation!

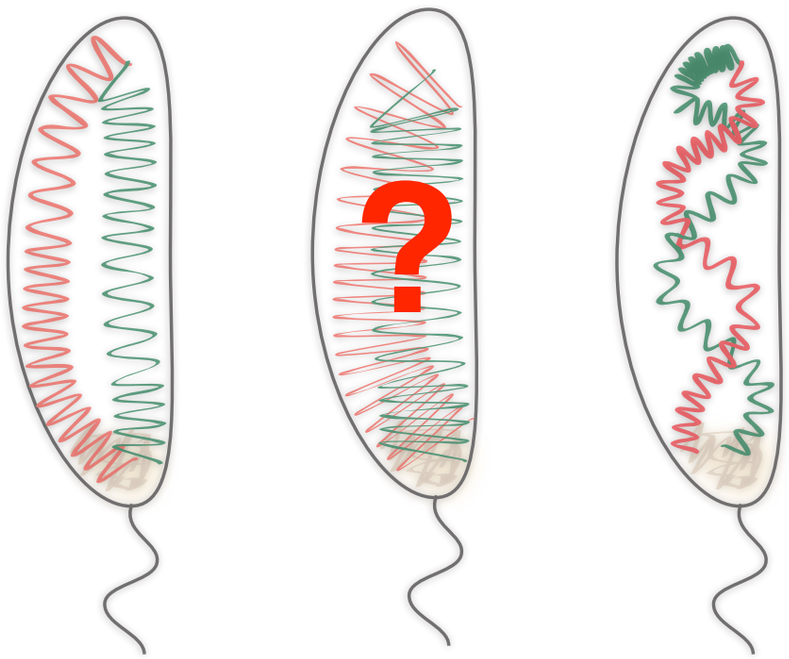


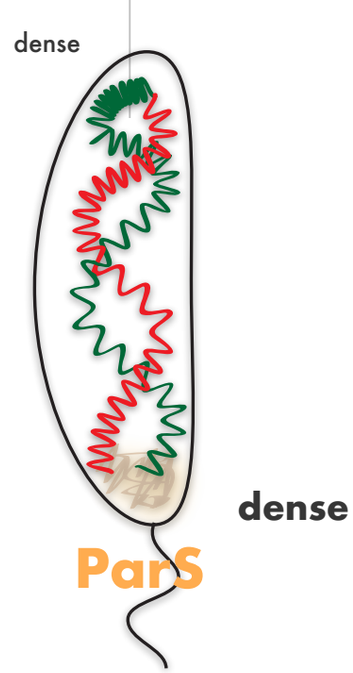


Arms are **STILL**

Genome architecture in Caulobacter





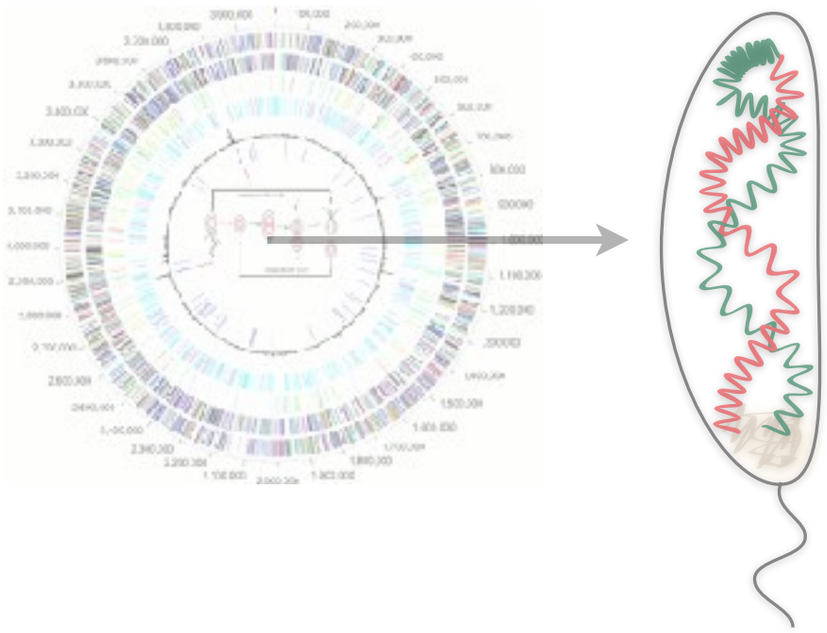


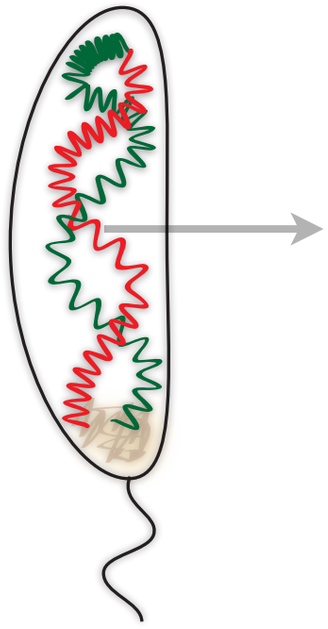
M.A. Umbarger, *et al.* **Molecular Cell** (2011) **44**:252–264

From Sequence to Function

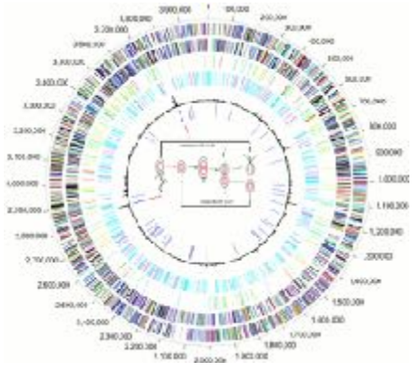
5C + IMP

Technology





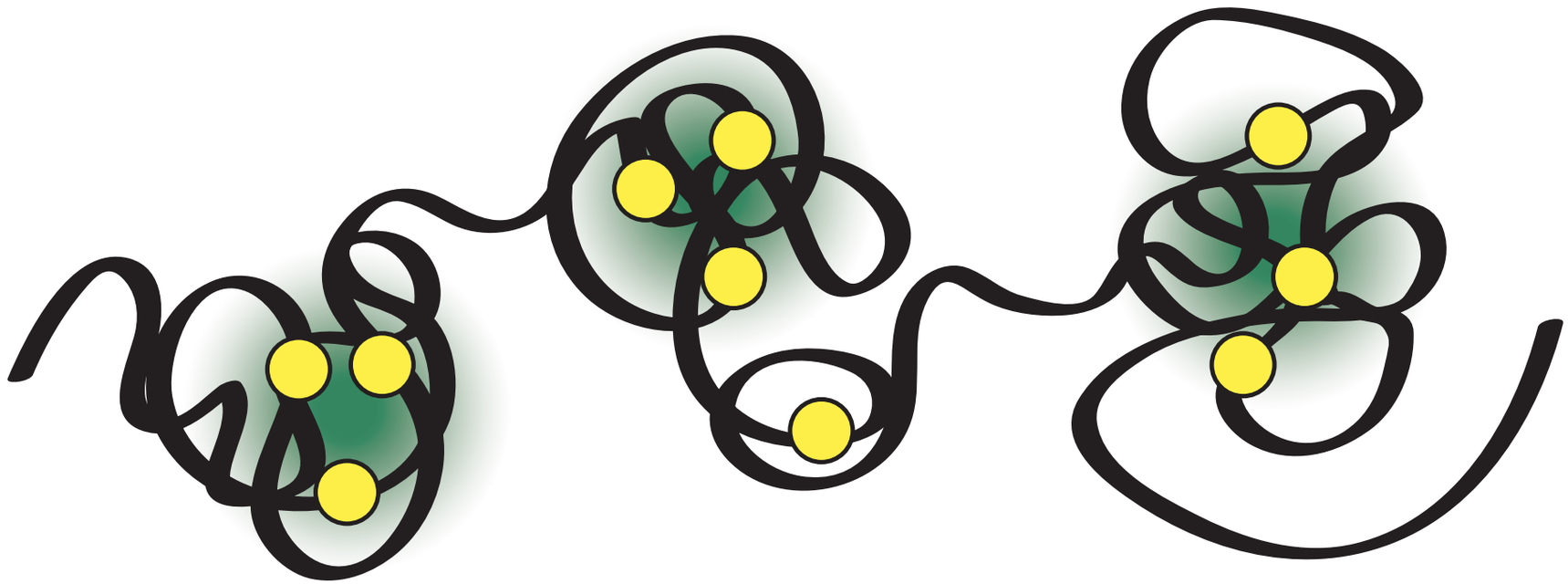
Hypothesis



Function!

D. Baù and M.A. Marti-Renom **Chromosome Res** (2011) 19:25-35.

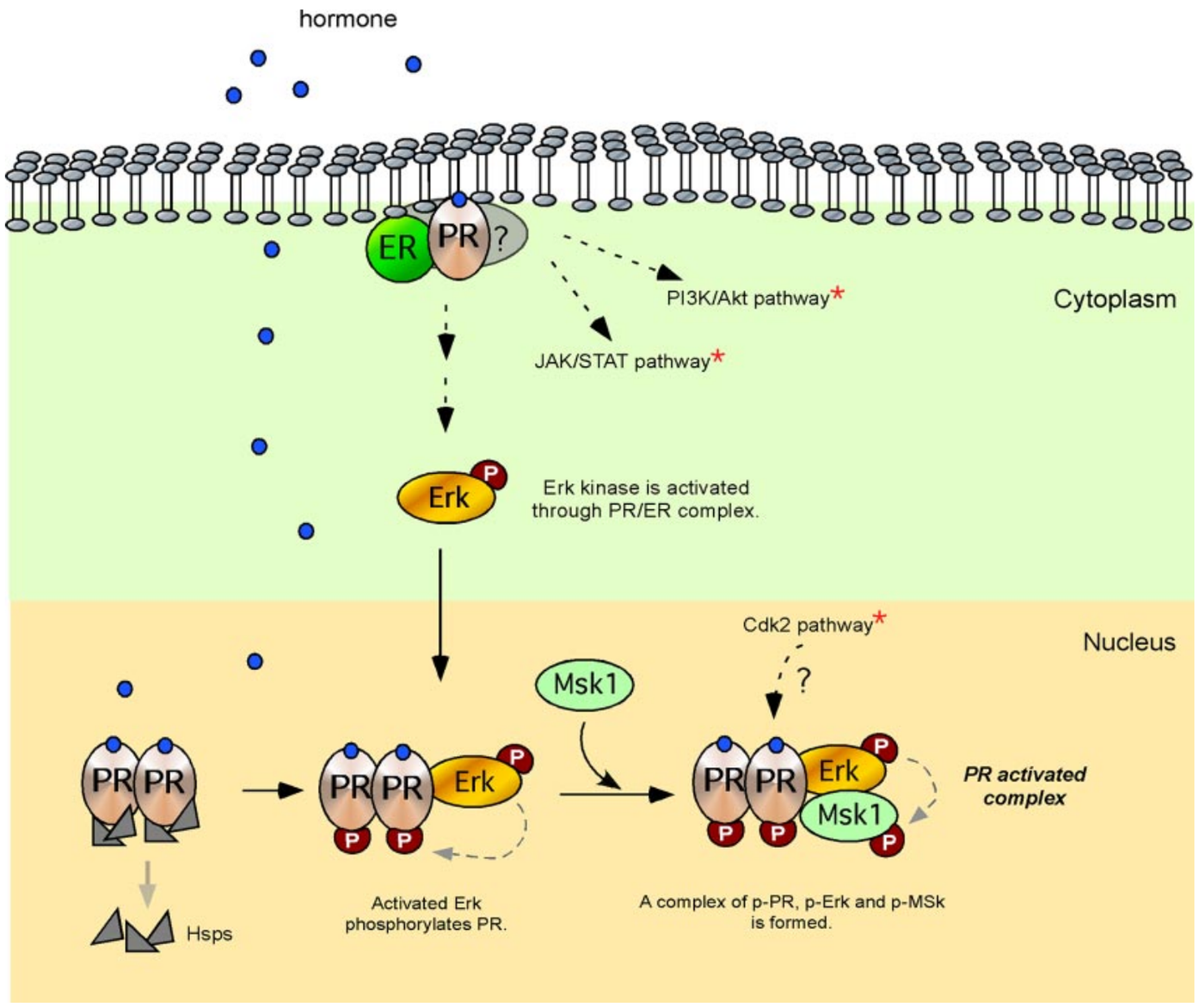
On TADs and hormones



Davide Baù

François le Dily

Progesterone-regulated transcription in breast cancer



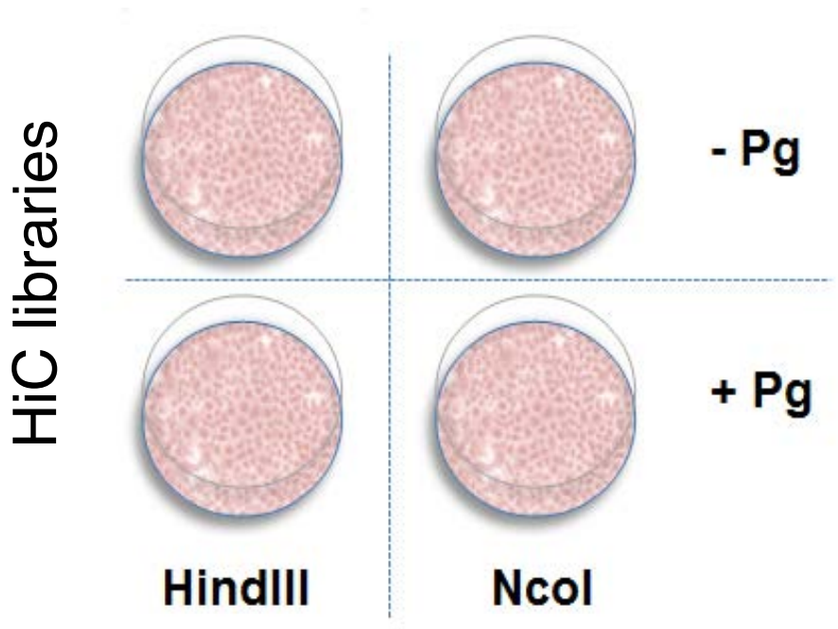
> 2,000 genes **Up**-regulated

> 2,000 genes **Down**-

Regulation in 3D?

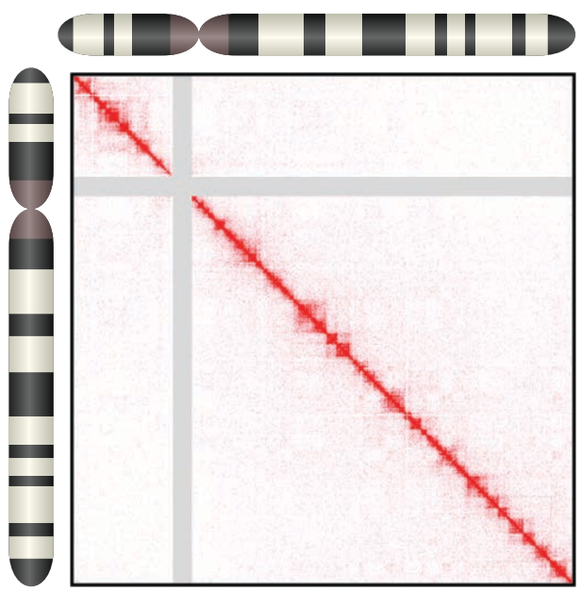
Vicent *et al* 2011, Wright *et al* 2012, Ballare *et al* 2012

Experimental design

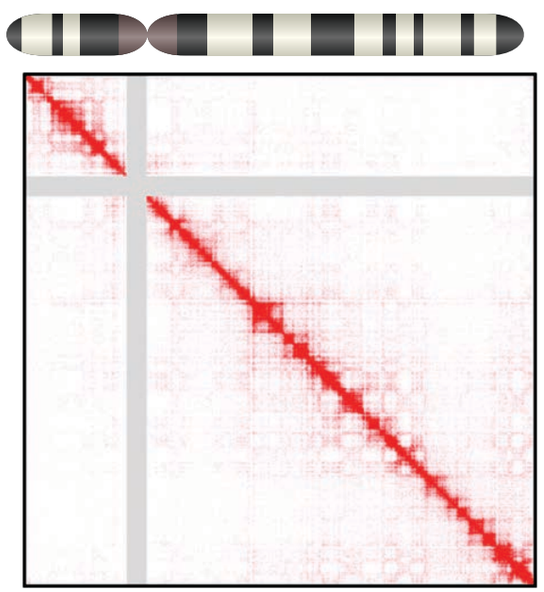


ChIP-Seq
RNA-Seq
Hi-C

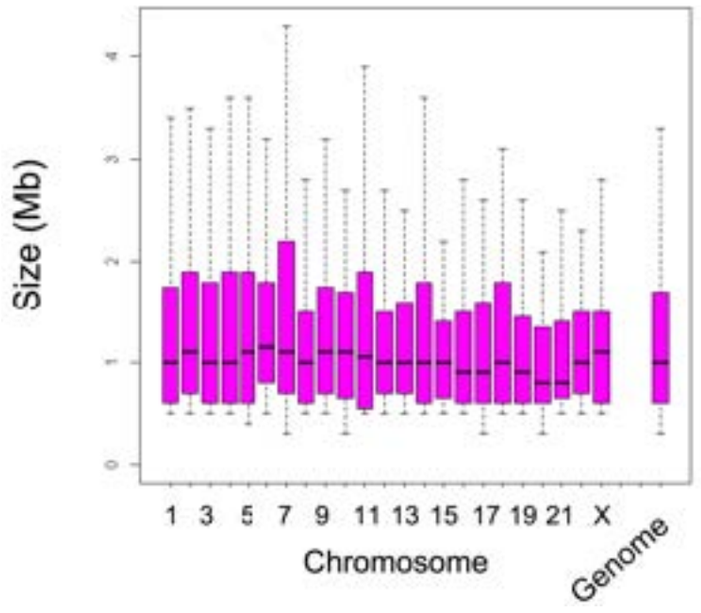
Chr.18 (Hind III)



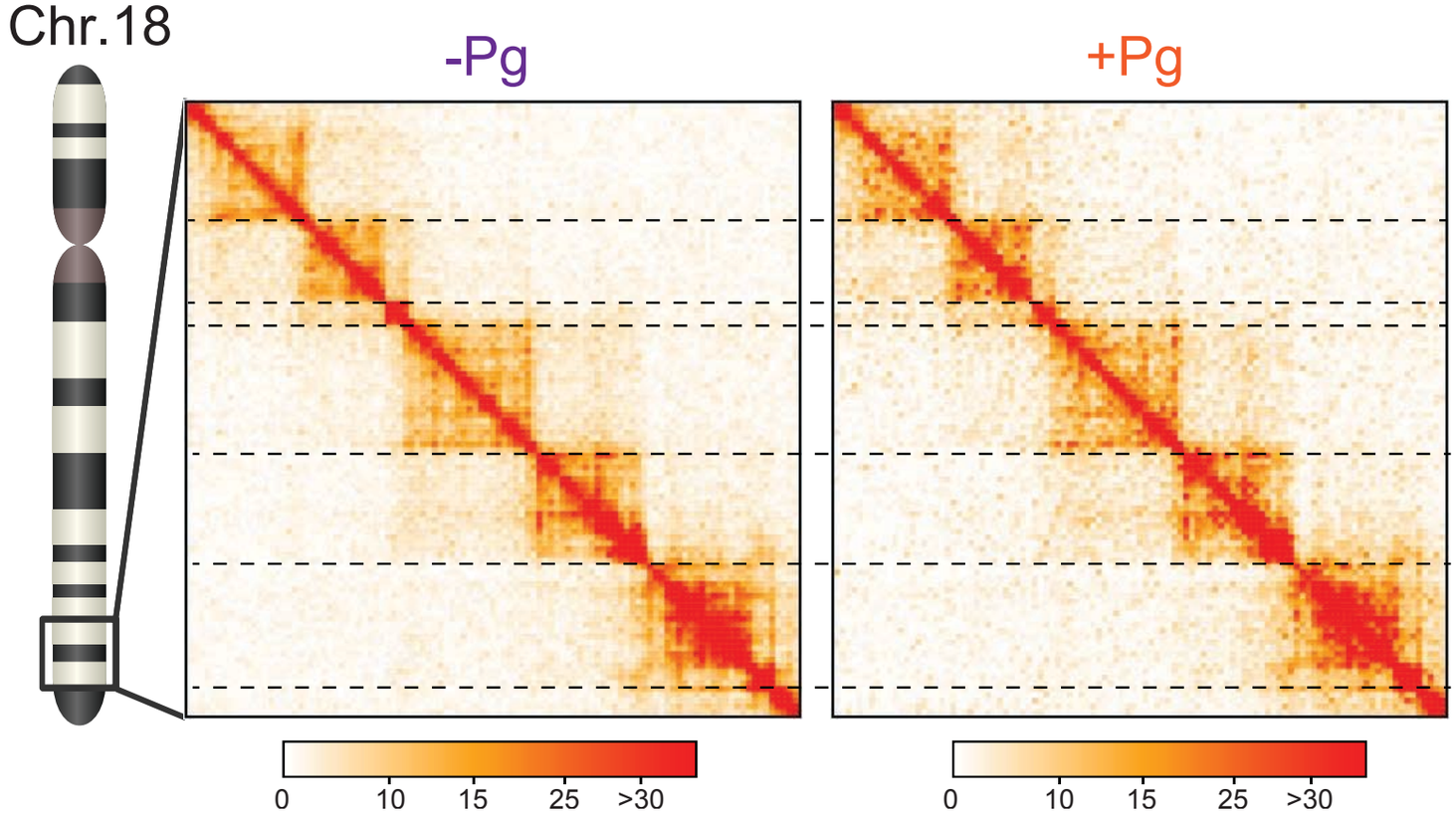
Chr.18 (NcoI)

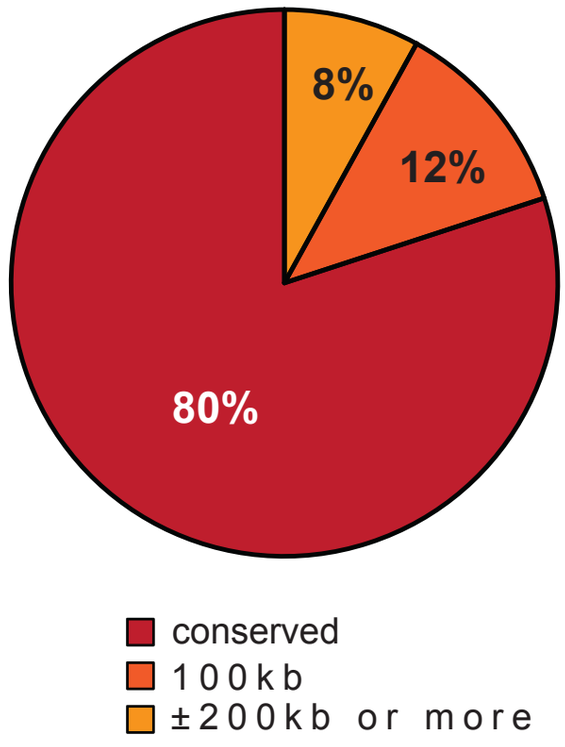


Are there TADs? how robust?

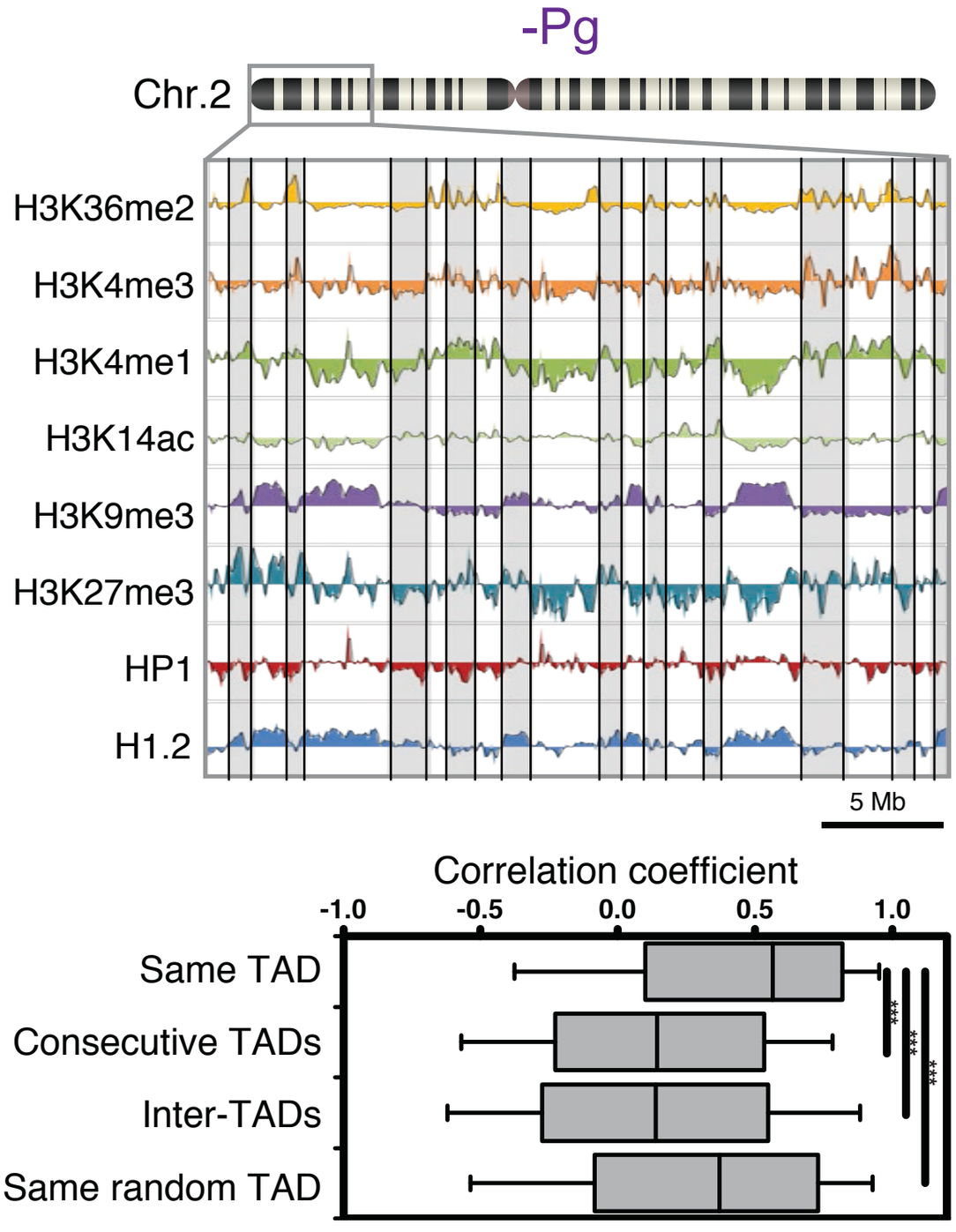


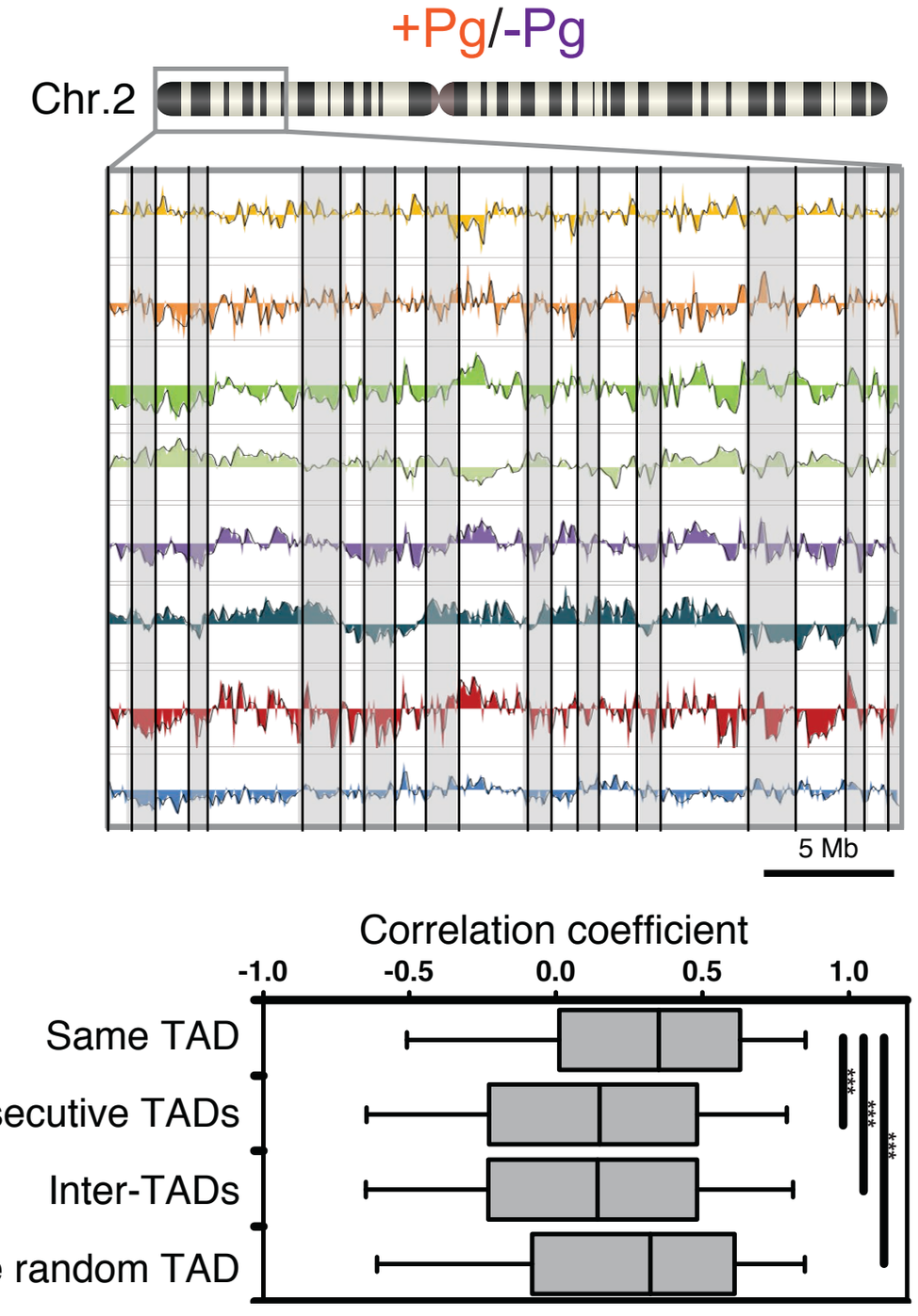
>2,000 detected TADs



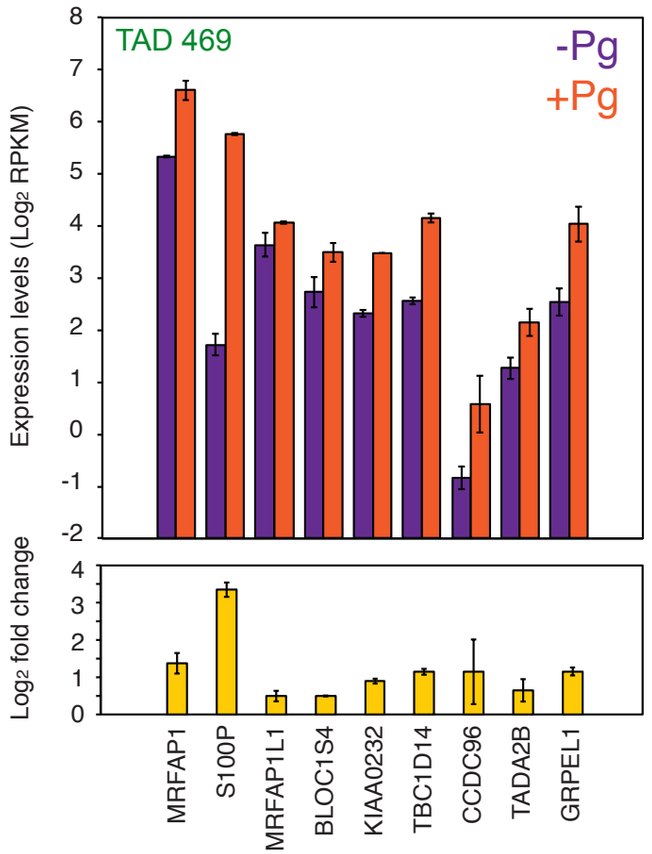


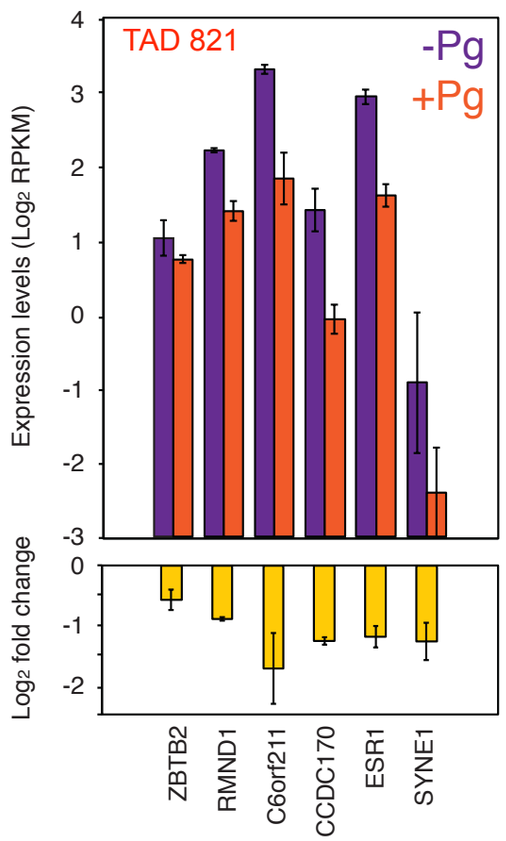
Are TADs homogeneous?

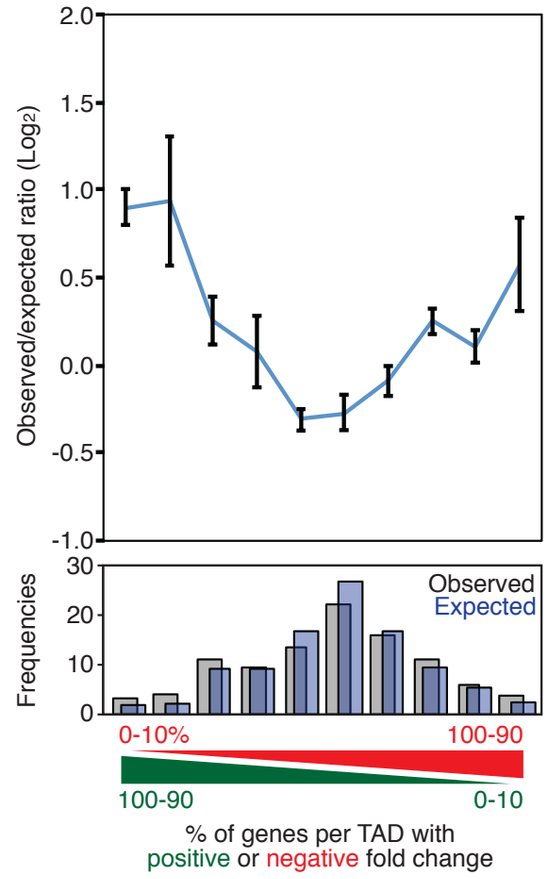




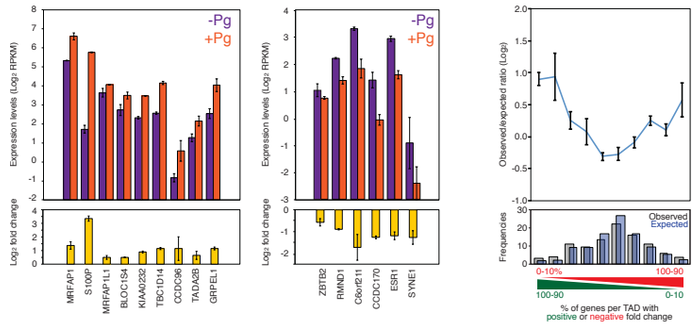
Do TADs respond differently to Pg treatment?



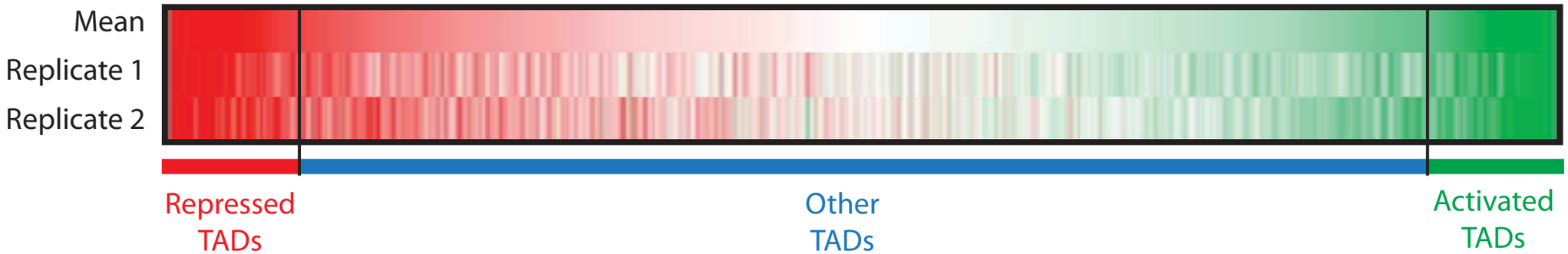




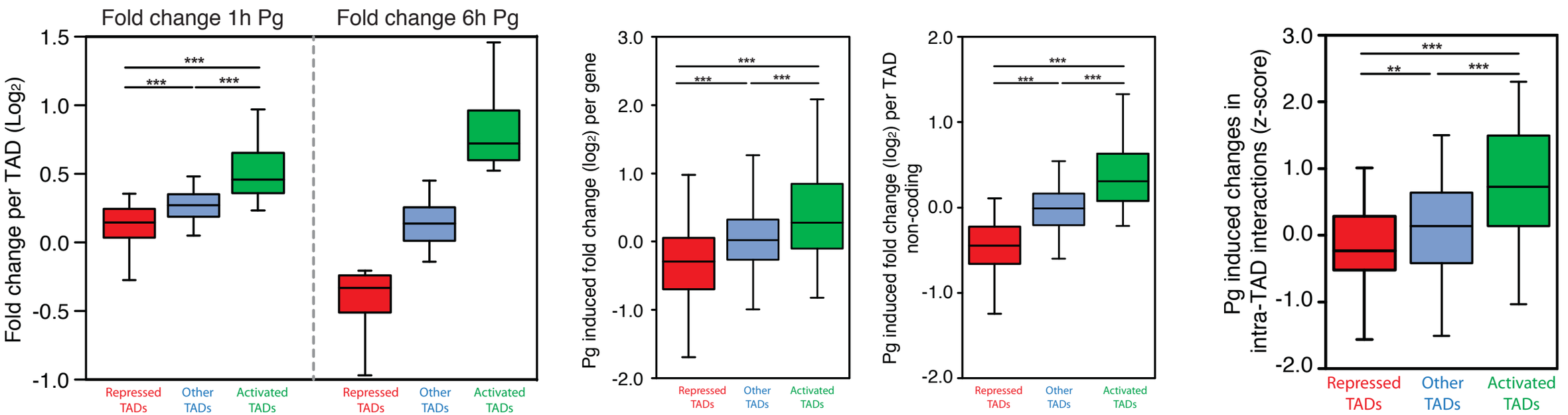
Do TADs respond differently to Pg treatment?



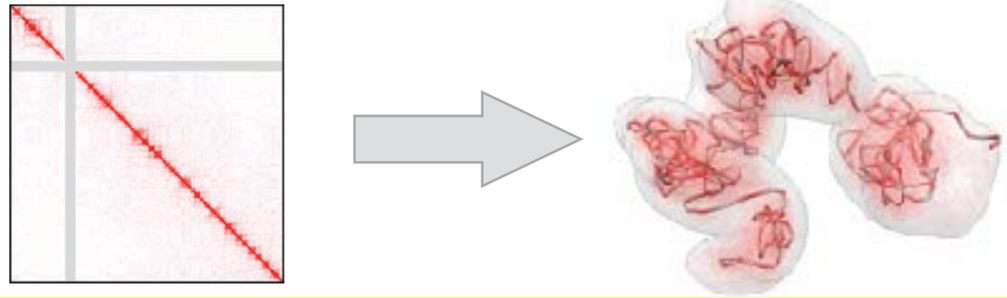
Pg induced fold change per TAD (6h)



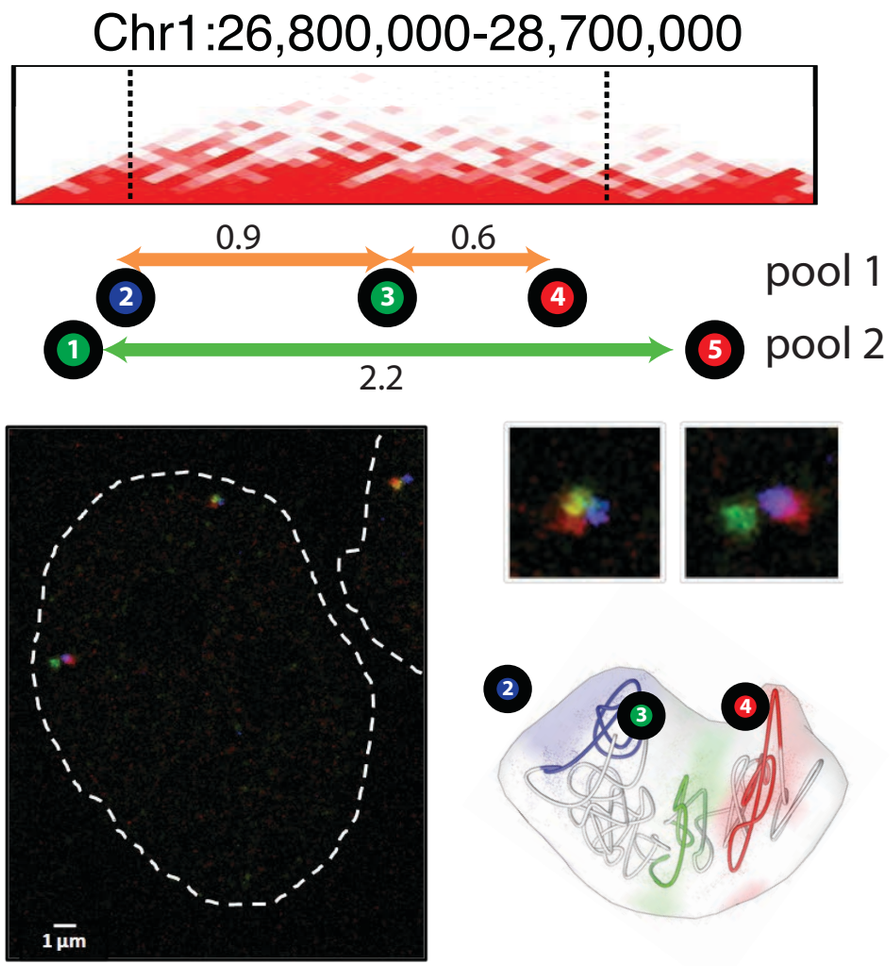
Repressed TADs | Other TADs | Activated TADs

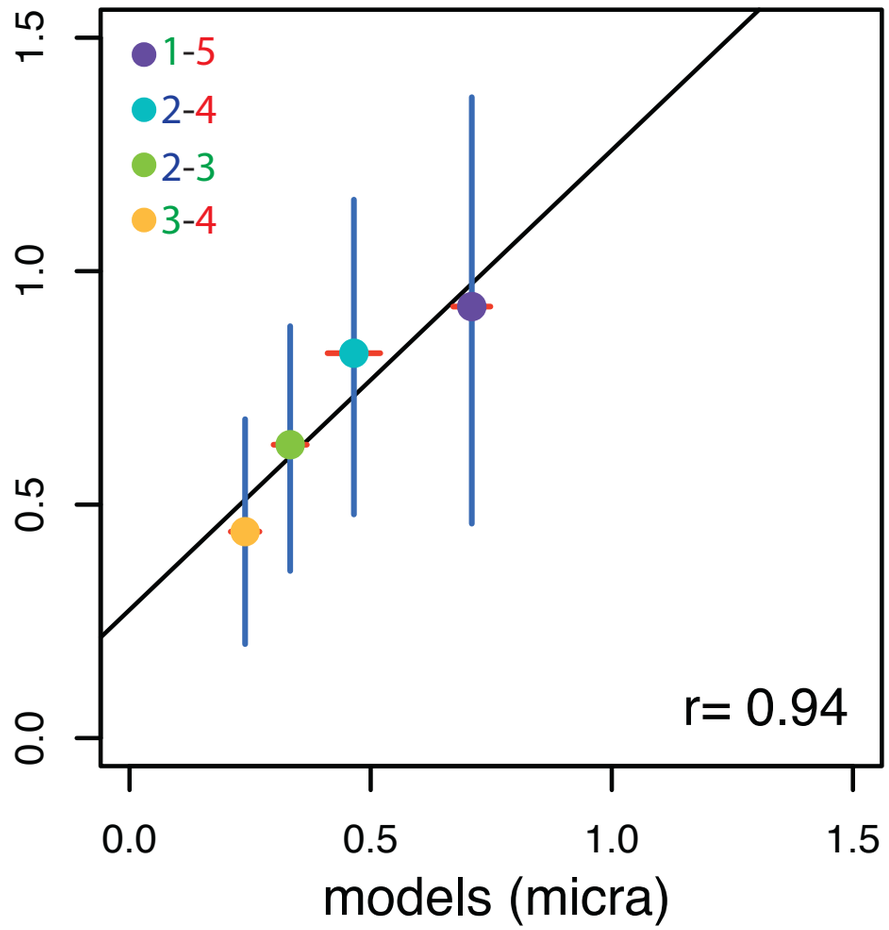


Modeling 3D TADs



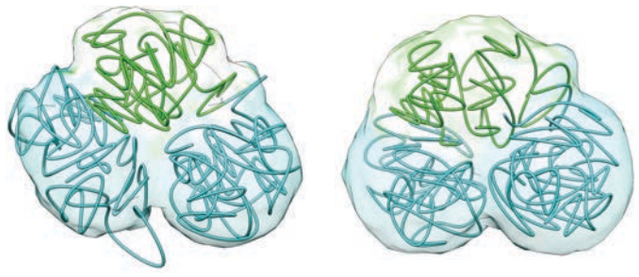
61 genomic regions containing 209 TADs covering 267Mb

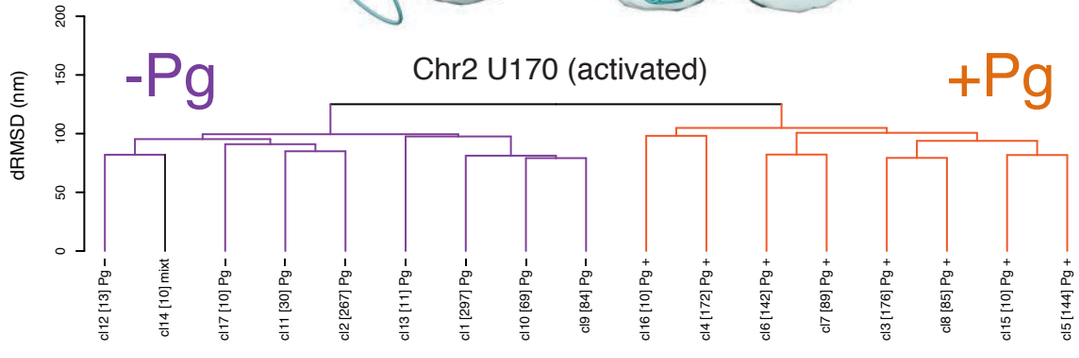




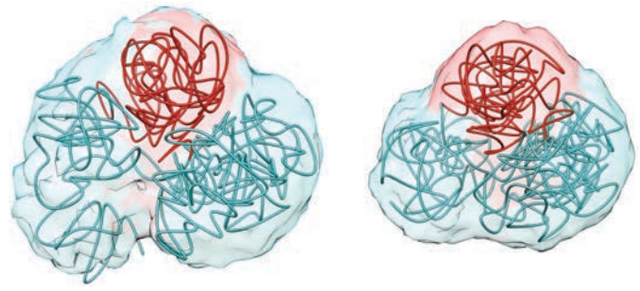
How TADs respond structurally to Pg?

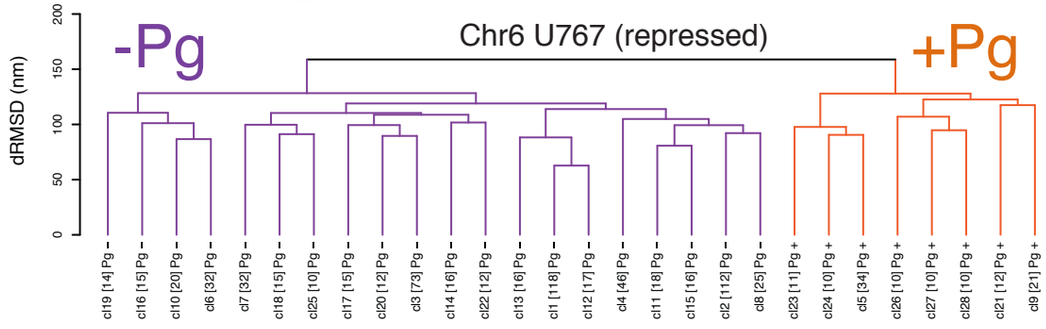
Chr2:9,600,000-13,200,000

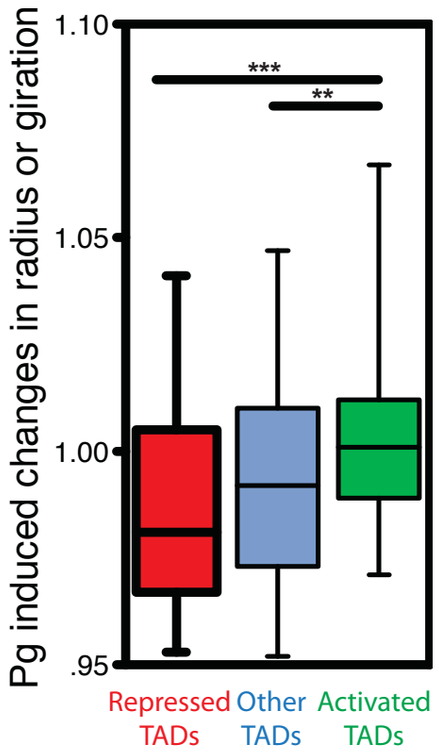


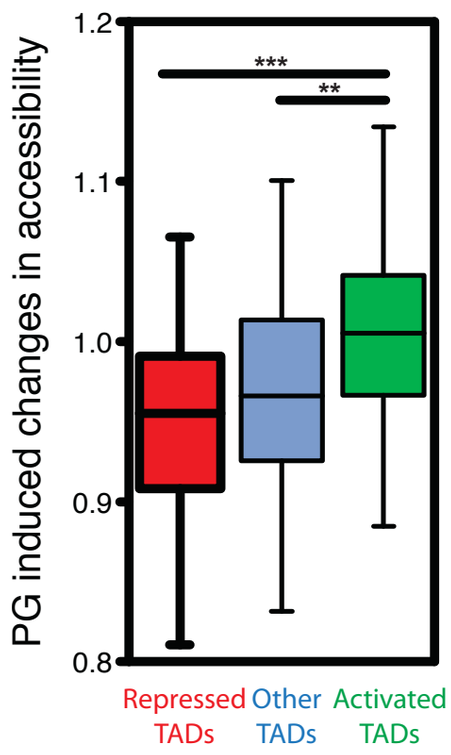


Chr6:71,800,000-76,500,000

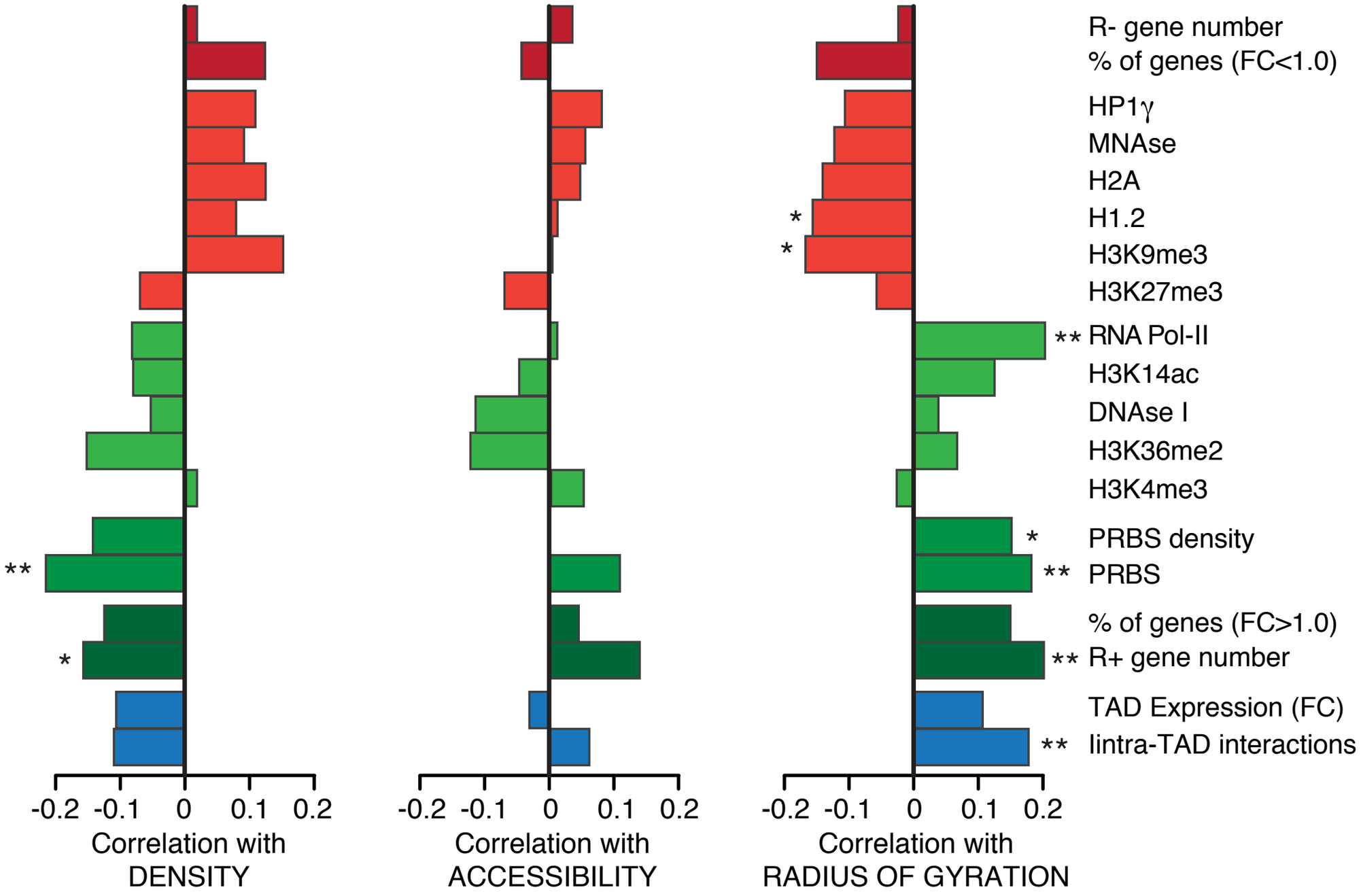


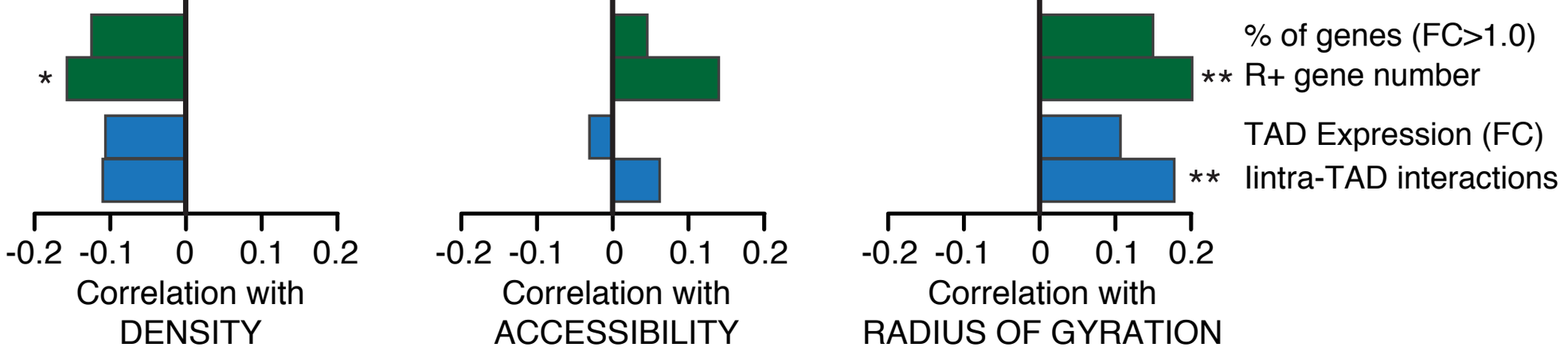


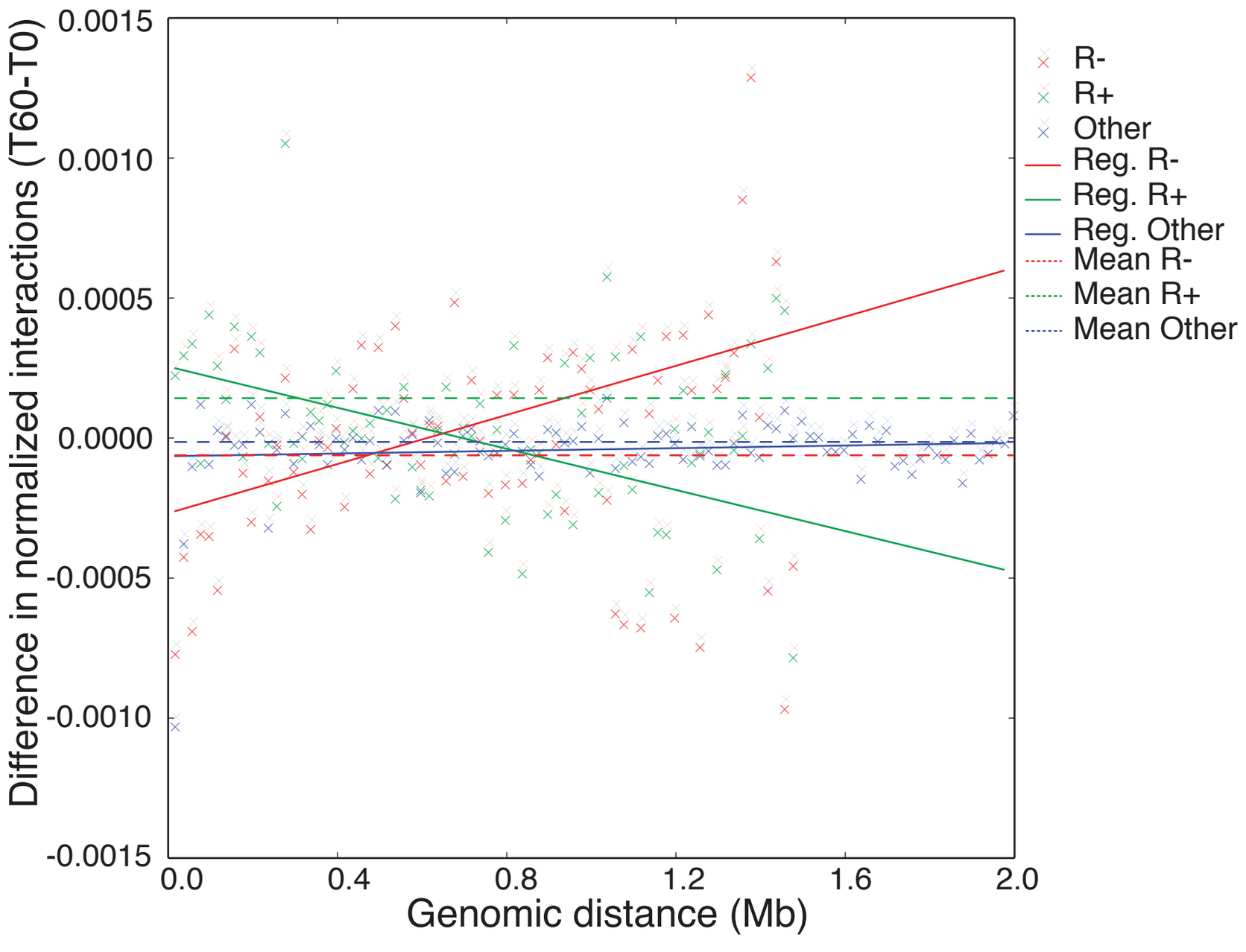




How TADs respond structurally to Pg?



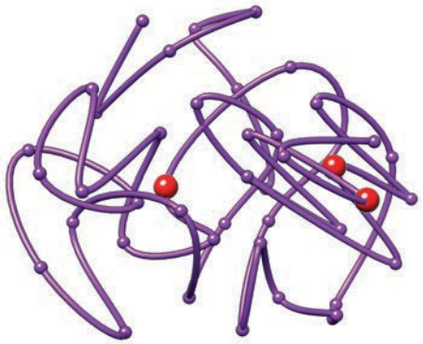

*
-0.2 -0.1 0 0.1 0.2
Correlation with
DENSITY
-0.2 -0.1 0 0.1 0.2
Correlation with
ACCESSIBILITY
-0.2 -0.1 0 0.1 0.2
Correlation with
RADIUS OF GYRATION
% of genes (FC>1.0)
** R+ gene number
TAD Expression (FC)
** Iintra-TAD interactions

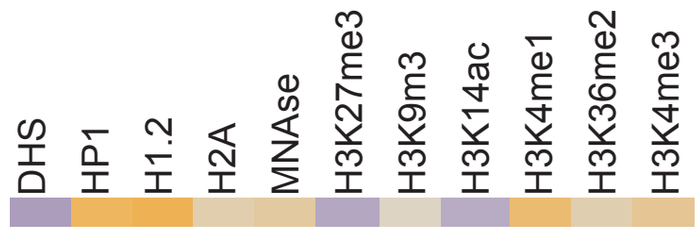

Difference in normalized interactions (T60-T0)
0.0015
0.0010
0.0005
0.0000
-0.0005
-0.0010
-0.0015
0.0
0.4
0.8
1.2
1.6
2.0
Genomic distance (Mb)
R-
R+
Other
Reg. R-
Reg. R+
Reg. Other
Mean R-
Mean R+
Mean Other

Model for TAD regulation

Repressed TAD

chr1 U41

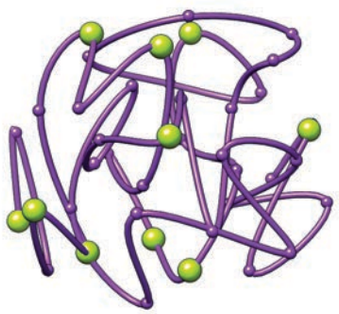


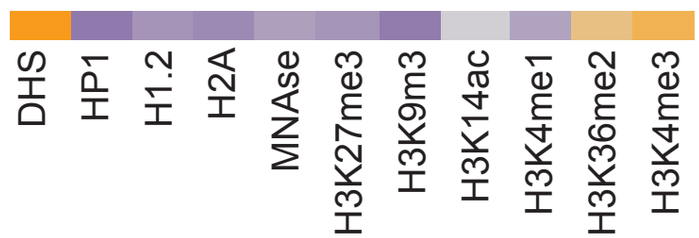


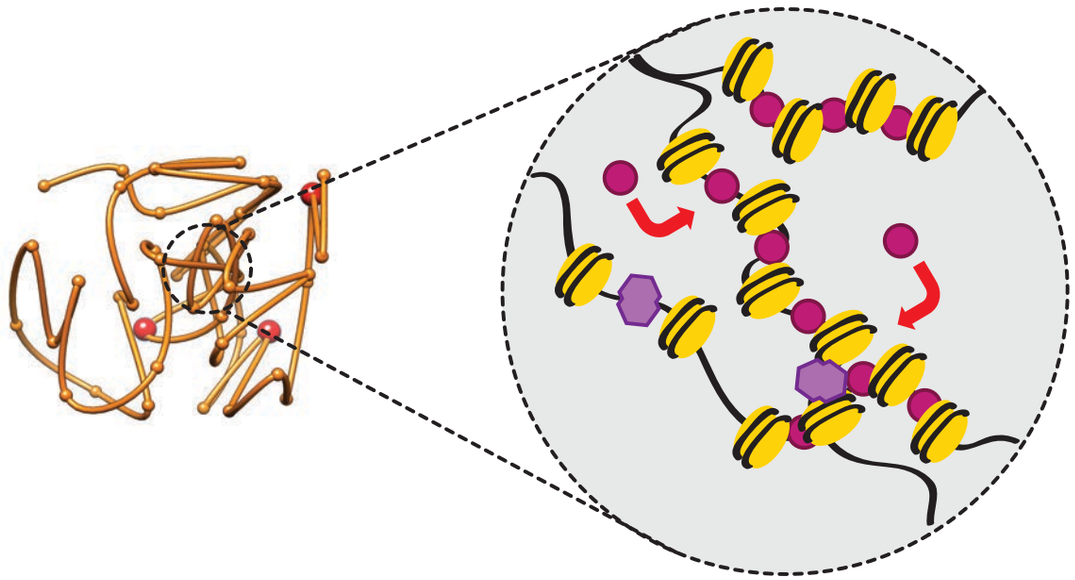
Structural transition **+Pg**

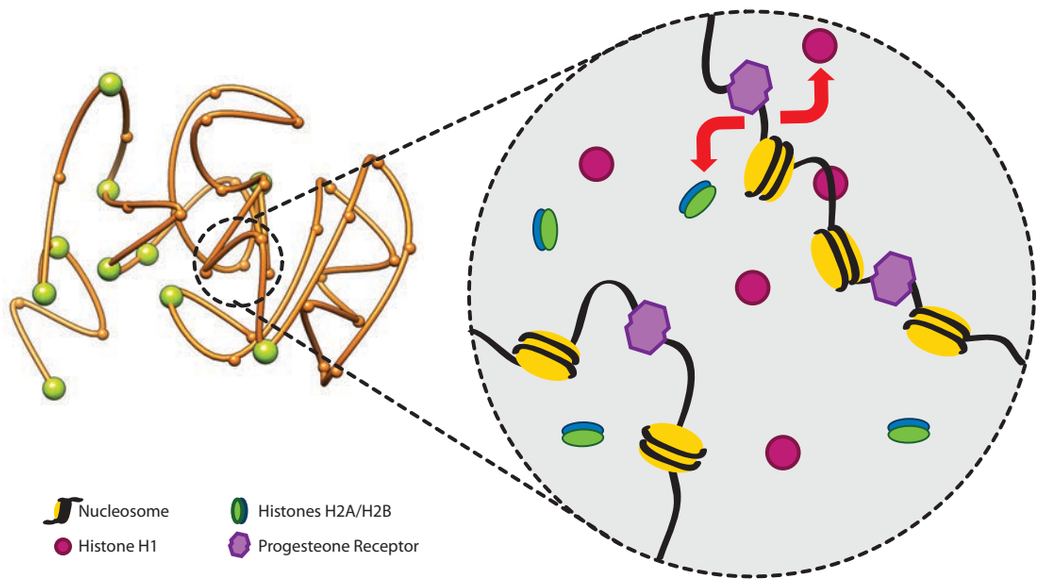
Activated TAD

chr2 U207









PLoS CB Outlook

Marti-Renom MA, Mirny LA (2011) PLoS Comput Biol 7(7): e1002125.

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Review

Bridging the Resolution Gap in Structural Modeling of 3D Genome Organization

Marc A. Marti-Renom[1]*, Leonid A. Mirny[2]

1 Structural Genomics Laboratory, Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe, Valencia, Spain, **2** Harvard-MIT Division of Health Sciences and Technology, and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America

Abstract: Over the last decade, and especially after the advent of fluorescent *in situ* hybridization imaging and chromosome conformation capture methods, the availability of experimental data on genome three-dimensional organization has dramatically increased. We now have access to unprecedented details of how genomes organize within the interphase nucleus. Development of new computational approaches to leverage this data has already resulted in the first three-dimensional structures of genomic domains and genomes. Such approaches expand our knowledge of the chromatin folding principles, which has been classically studied using polymer physics and molecular simulations. Our outlook describes computational approaches for integrating experimental data with polymer physics, thereby bridging the resolution gap for structural determination of genomes and genomic domains.

This is an "Editors' Outlook" article for *PLoS Computational Biology*

Recent experimental and computational advances are resulting in an increasingly accurate and detailed characterization of how genomes are organized in the three-dimensional (3D) space of the nucleus (Figure 1) [1]. At the lowest level of chromatin organization, naked DNA is packed into nucleosomes, which forms the so-called chromatin fiber composed of DNA and proteins. However, this initial packing, which reduces the length of the DNA by about seven times, is not sufficient to explain the higher-order folding of chromosomes during interphase and metaphase. It is now accepted that chromosomes and genes are non-randomly and dynamically positioned in the cell nucleus during the interphase, which challenges the classical representation of genomes as linear static sequences. Moreover, compartmentalization, chromatin organization, and spatial location of genes are associated with gene expression and the functional status of the cell. Despite the importance of 3D genomic architecture, we have a limited understanding of the molecular mechanisms that determine the higher-order organization of genomes and its relation to function. Computational biology plays an important role in the plethora of new technologies aimed at addressing this knowledge gap [2]. Indeed, Thomas Cremer, a pioneer in studying nuclear organization using light microscopy, recently highlighted the importance of computational science in complementing and leveraging experimental observations of genome organization [2]. Therefore, computational approaches to integrate experimental observations with chromatin physics are needed to determine the architecture (3D) and dynamics (4D) of genomes.

We present two complementary approaches to address this challenge: (i) the first approach aims at developing simple polymer models of chromatin and determining relevant interactions (both physical and biological) that explain experimental observations; (ii) the second approach aims at integrating diverse experimental observations into a system of spatial restraints to be satisfied, thereby constraining possible structural models of the chromatin. The goal of both approaches is dual: to obtain most accurate 3D and 4D representation of chromatin architecture and to understand physical constraints and biological phenomena that determine its organization. These approaches are reminiscent of the protein-folding field where the first strategy was used for characterizing protein "foldability" and the second was implemented for modeling the structure of proteins using nuclear magnetic resonance and other experimental constraints. In fact, our outlook consistently returns to the many connections between the two fields.

What Does Technology Show Us?

Today, it is possible to quantitatively study structural features of genomes at diverse scales that range from a few specific loci, through chromosomes, to entire genomes (Table 1) [3]. Broadly, there are two main approaches for studying genomic organization: light microscopy and cell/molecular biology (Figure 2). Light microcopy [4], both with fixed and living cells, can provide images of a few loci within individual cells [5,6], as well as their dynamics as a function of time [7] and cell state [8]. On a larger scale, light microscopy combined with whole-chromosome staining reveals chromosomal territories during interphase and their reorganization upon cell division. Immunofluorescence with fluorescent antibodies in combination with RNA, and DNA fluorescence *in situ* hybridization (FISH) has been used to determine the co-localization of loci and nuclear substructures.

Using cellular and molecular biology, novel chromosome conformation capture (3C)-based methods such 3C [9], 3C-on-chip or circular 3C (the so-called 4C) [10,11], 3C carbon copy (5C) [12], and Hi-C [13] quantitatively measure frequencies of spatial contacts between genomic loci averaged over a large

Citation: Marti-Renom MA, Mirny LA (2011) Bridging the Resolution Gap in Structural Modeling of 3D Genome Organization. PLoS Comput Biol 7(7): e1002125. doi:10.1371/journal.pcbi.1002125

Editor: Philip E. Bourne, University of California San Diego, United States of America

Published July 14, 2011

Funding: MAM-R acknowledges support from the Spanish Ministry of Science and Innovation (BFU2010-19310). LM is acknowledging support of the NCI-funded MIT Center for Physics Sciences in Oncology. The funders had no role in decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mmarti@cipf.es

Davide Baù
Gireesh K. Bogu
Yasmina Cuartero
François le Dily
David Dufour
Irene Farabella
Silvia Galan
Francesca di Giovanni
Mike Goodstadt
Francisco Martínez-Jiménez
François Serra
Paula Soler
Yannick Spill
Marco di Stefano
Marie Trussart

http://marciuslab.org
http://3DGenomes.org
http://cnag.crg.eu