

# High Performance Computing in Biological Physics



***IL PROGETTO IBiSCo E LA TRANSIZIONE VERSO IL “CENTRO NAZIONALE DI RICERCA IN HIGH PERFORMANCE COMPUTING, BIG DATA E QUANTUM COMPUTING (ICSC)”***

**Complesso Universitario di Monte S. Angelo  
Università di Napoli "Federico II"  
Dipartimento di Fisica “E. Pancini”**

**Andrea Maria Chiariello, Napoli, 18/04/2024**



# The DNA

Classically, nucleotide sequence  
or “letters”



The sequence identifies the species and  
today can be obtained with  
extreme precision and at low costs  
(hundreds \$)



In humans, the sequence is ~3.5 billions bps long  
organized in 23 chromosomes (x2)

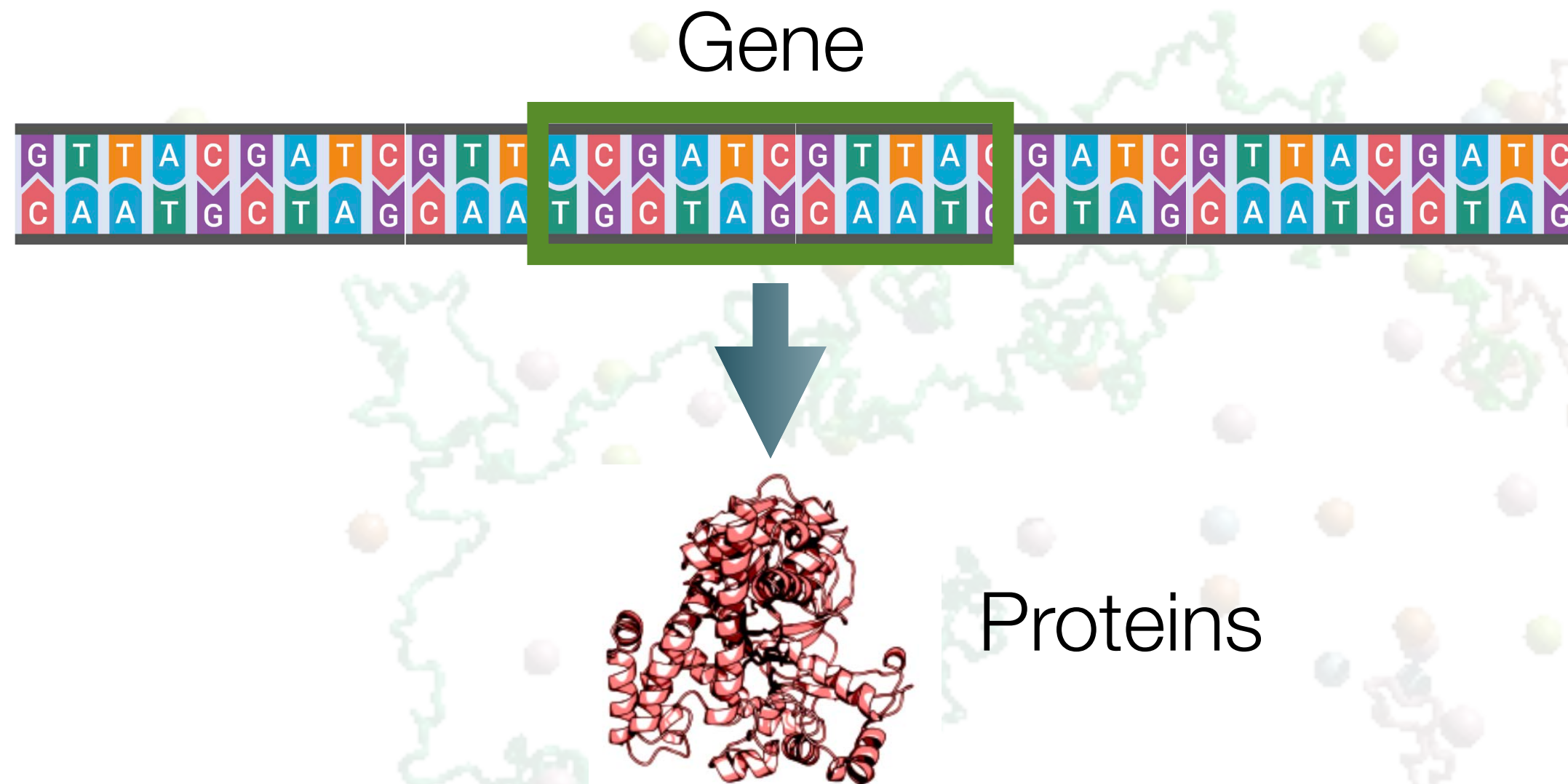
In mouse, it is ~2.7 billions bp long  
and 20 chromosomes





# The gene

---



Regions that encode the “*recipe*” of proteins through the genetic code

Human genome has ~22.000 genes, not so different from mice (~20.000) or *Drosophila M.* (~16.000)

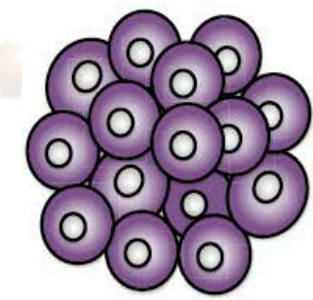
In humans, only 2-5% of DNA sequence encodes for proteins

Actually, the remaining part is fundamental to gene regulation:

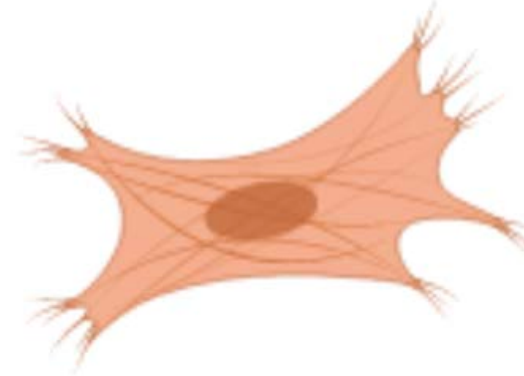
- 1) Contains regulatory elements
- 2) Contributes to 3D architecture

# Genes are regulated by enhancers

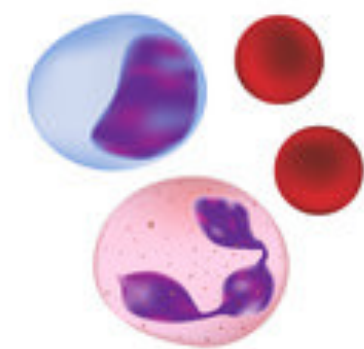
**Enhancers** are regulating sequences that determine the activity of a gene



Stem cells



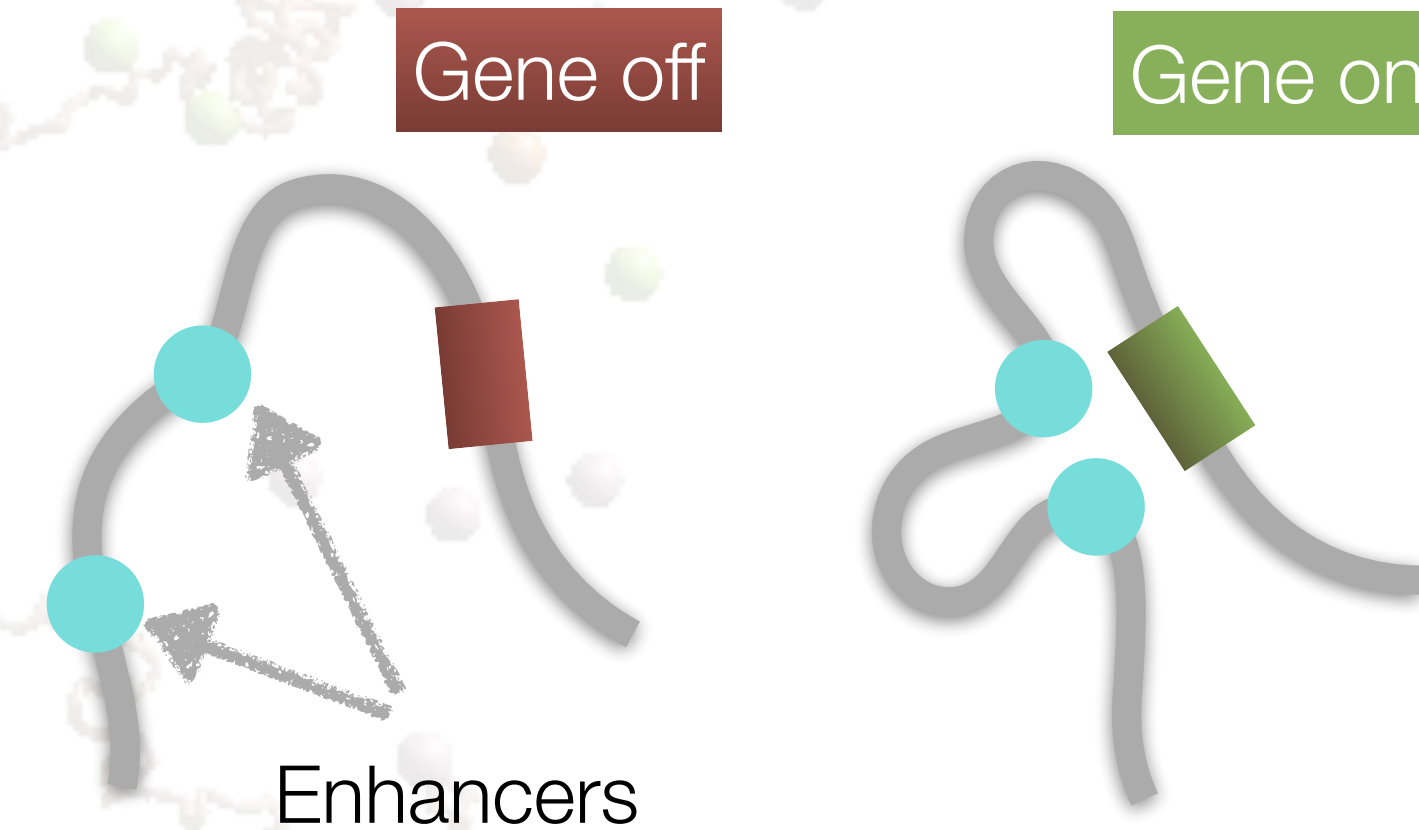
Muscles



Blood cells



Neurons

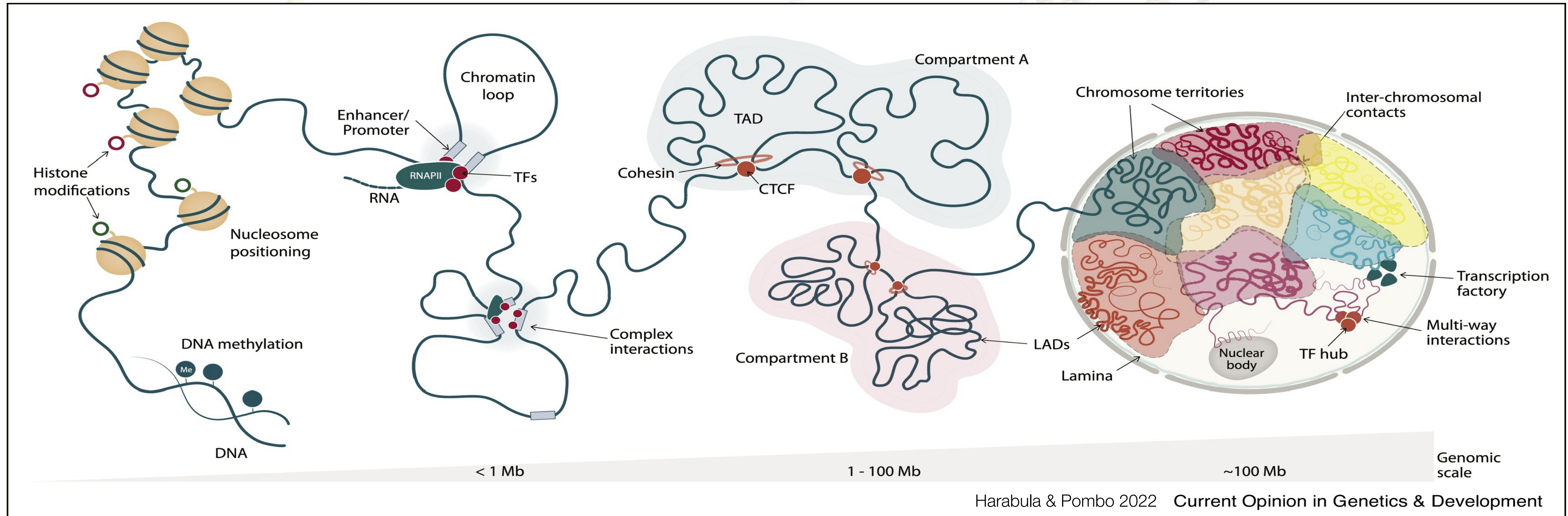


Different activity of single genes determines the identity of different tissues and cell types

An altered activity of genes can lead to severe pathogenic effects  
e.g. cancer and congenital diseases

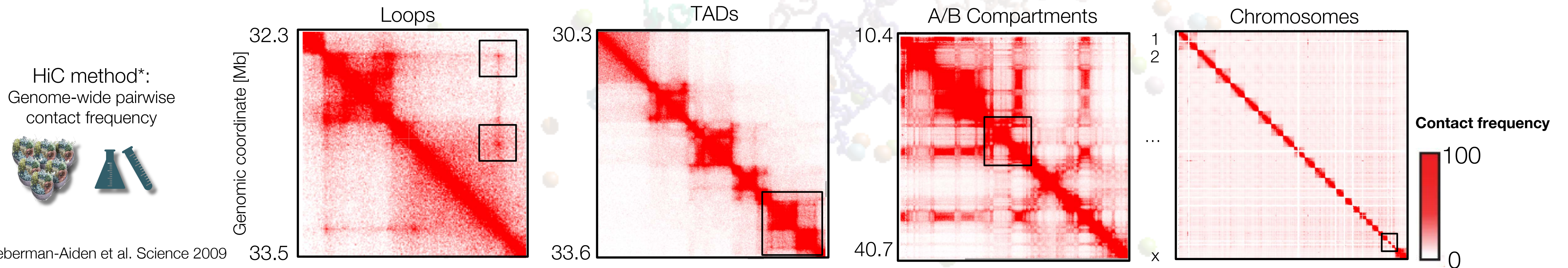
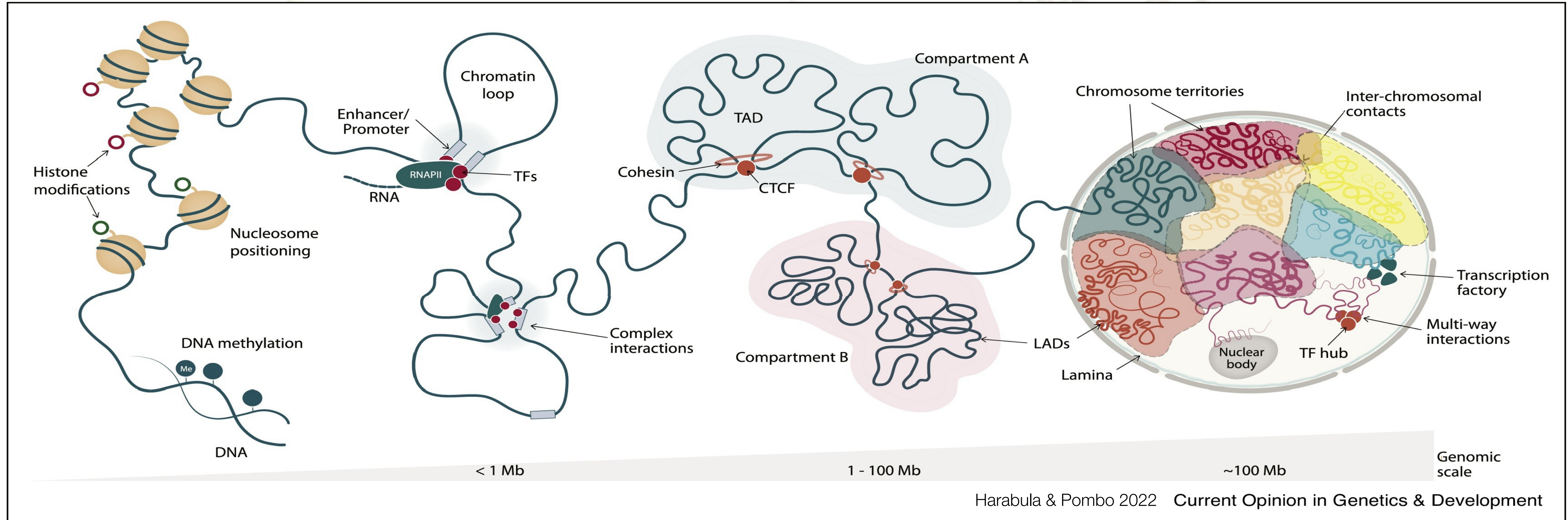


# Genomes exhibit a multiscale spatial organization





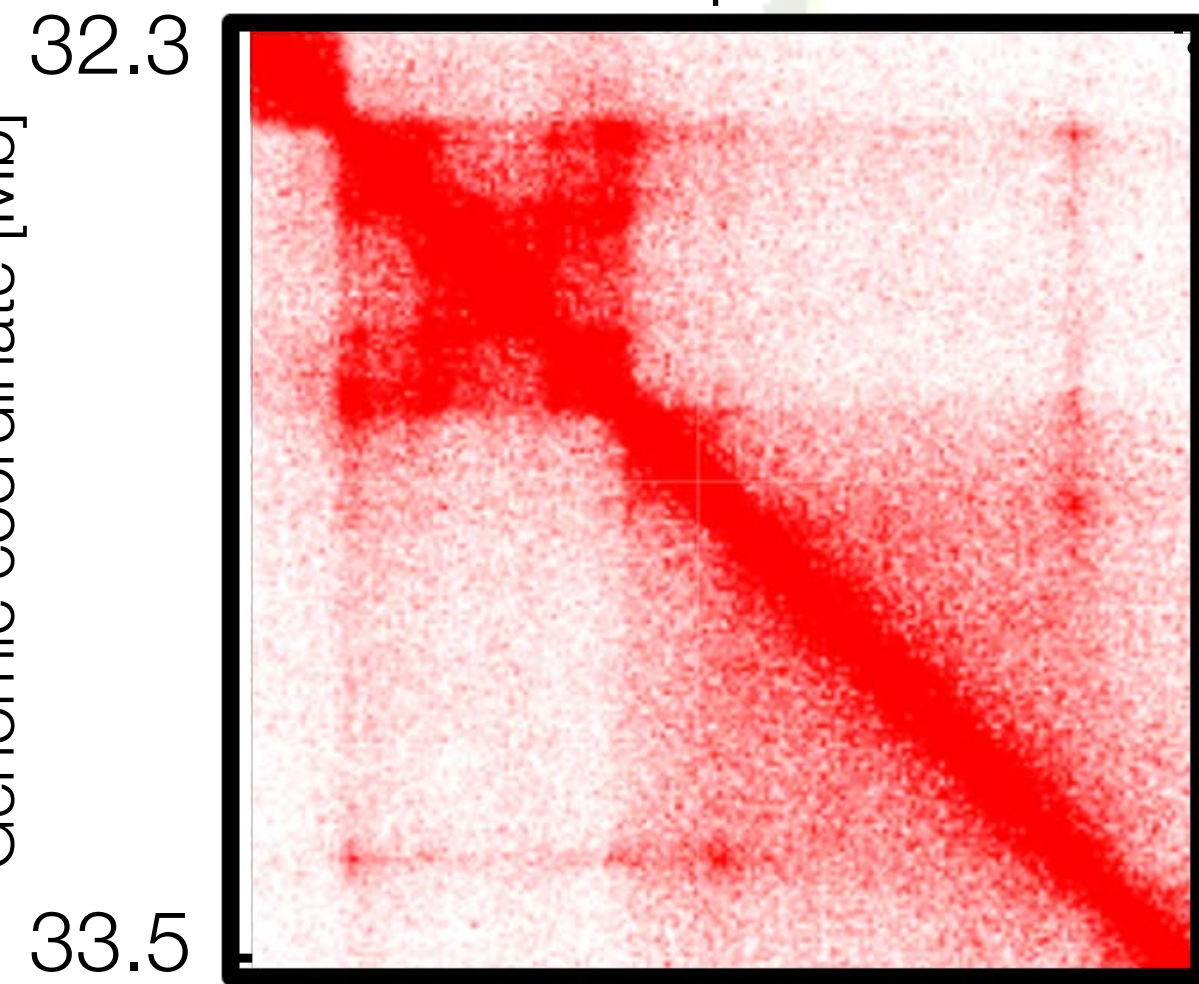
# Genomes exhibit a multiscale spatial organization



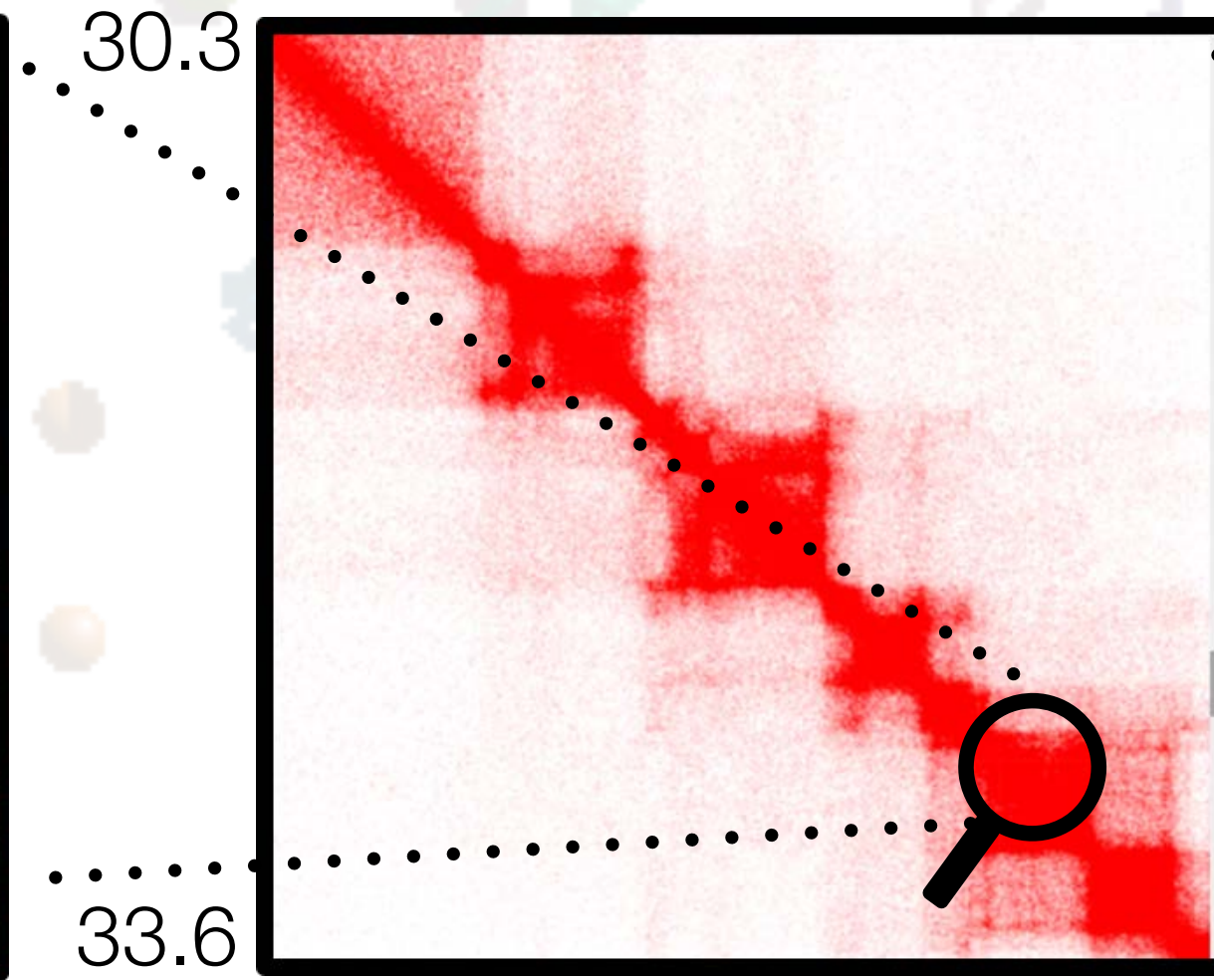


# Genomes exhibit a multiscale spatial organization

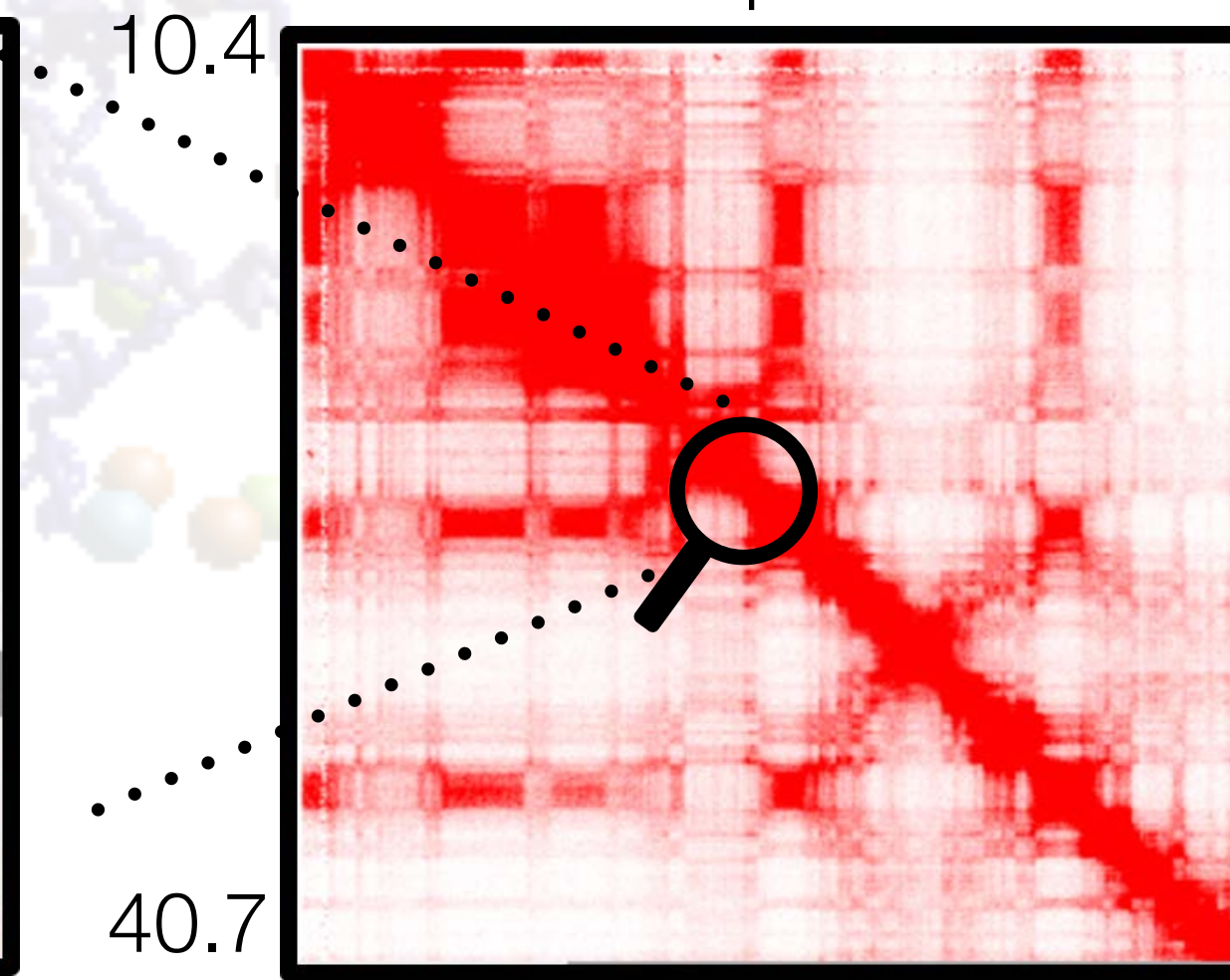
Loops



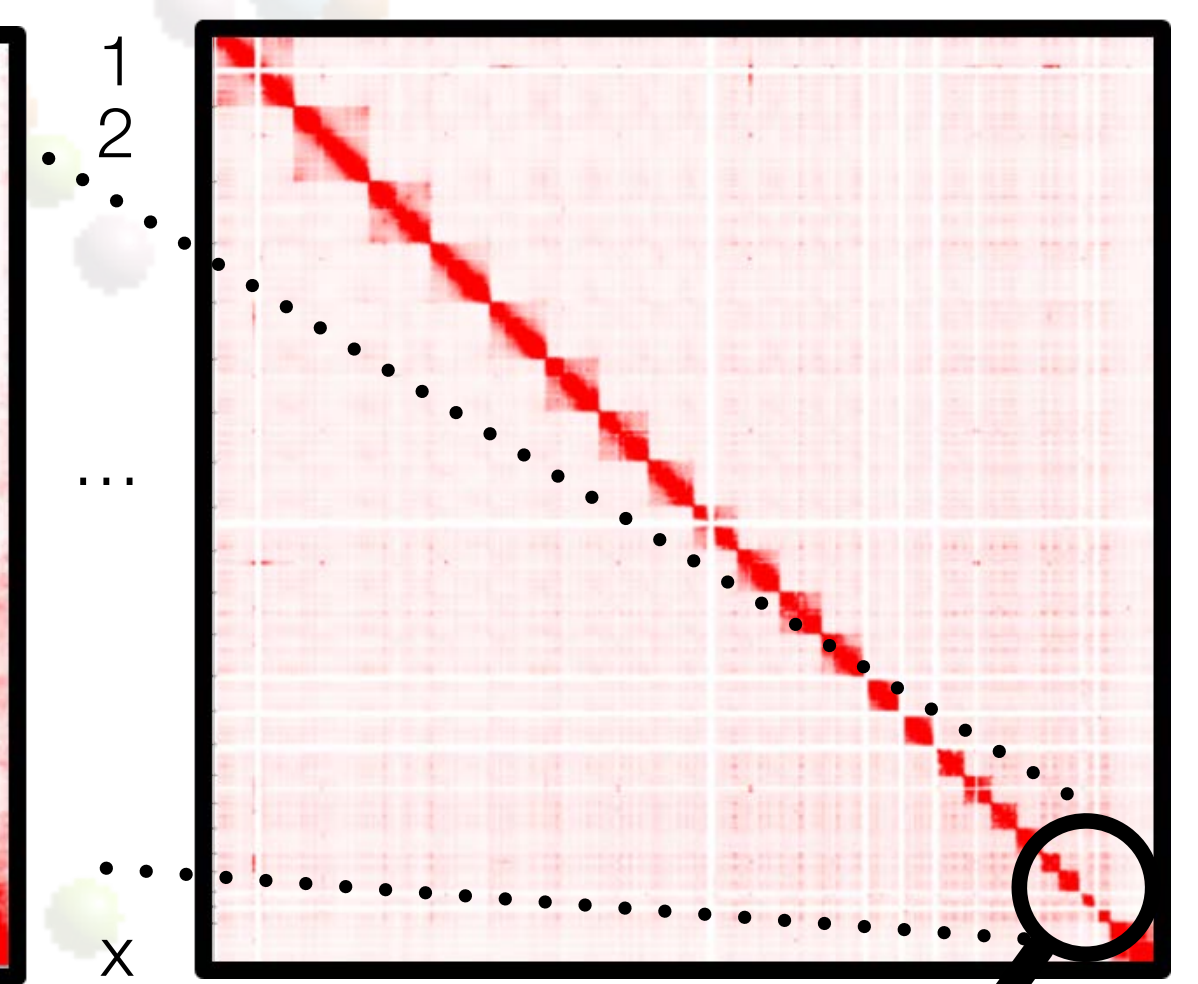
TADs



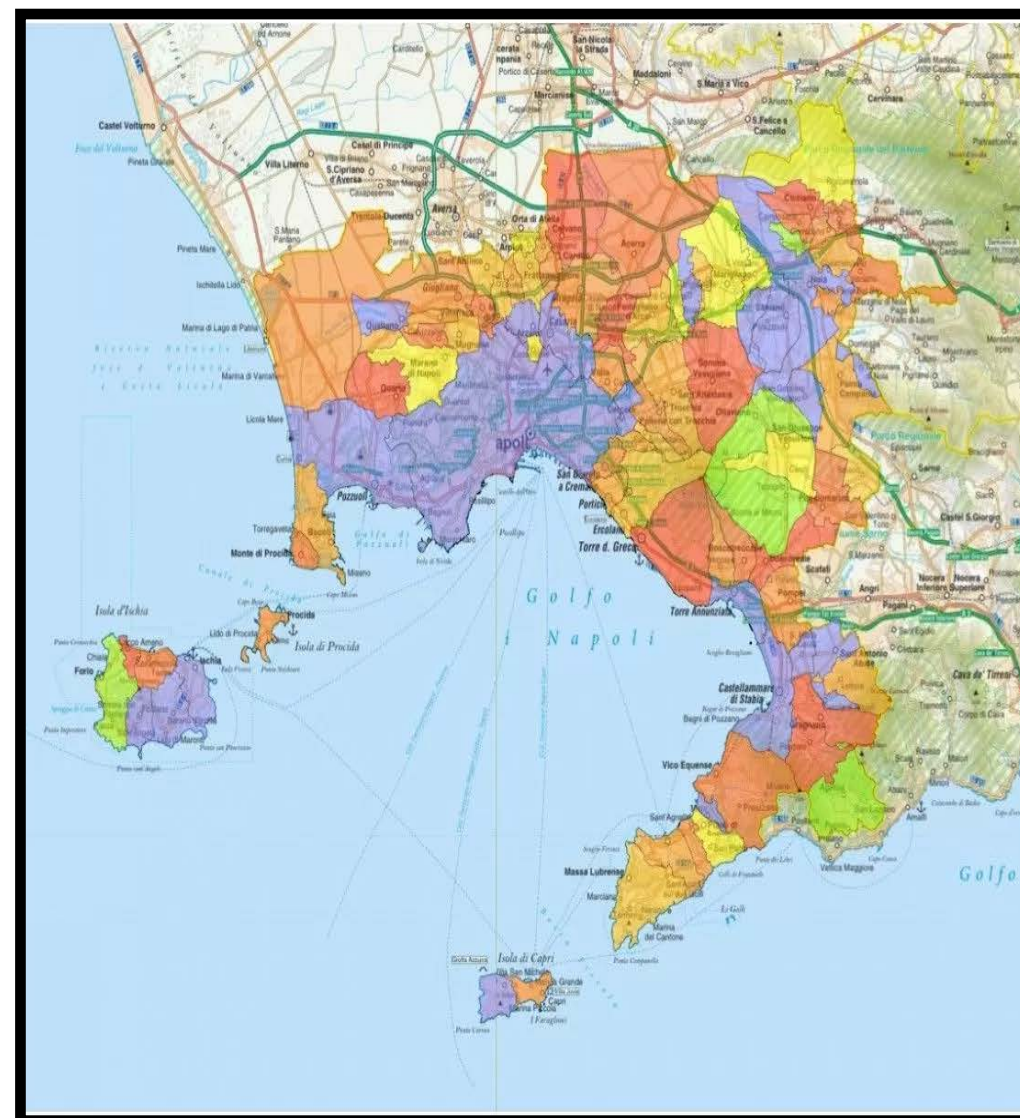
A/B Compartments



Chromosomes



Cities



Regions



Countries



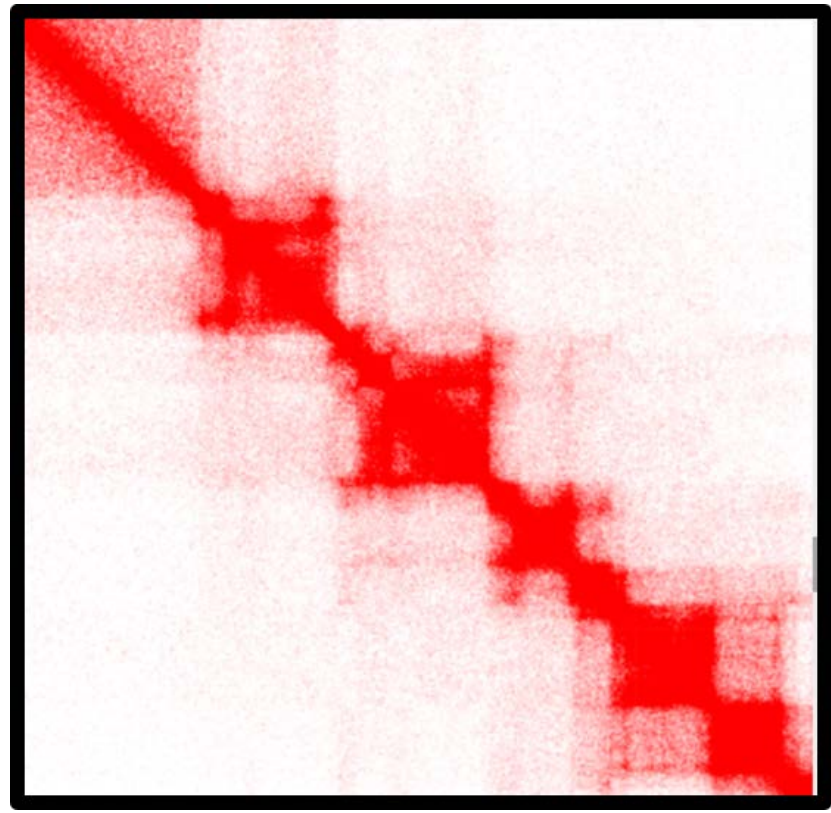
Continent



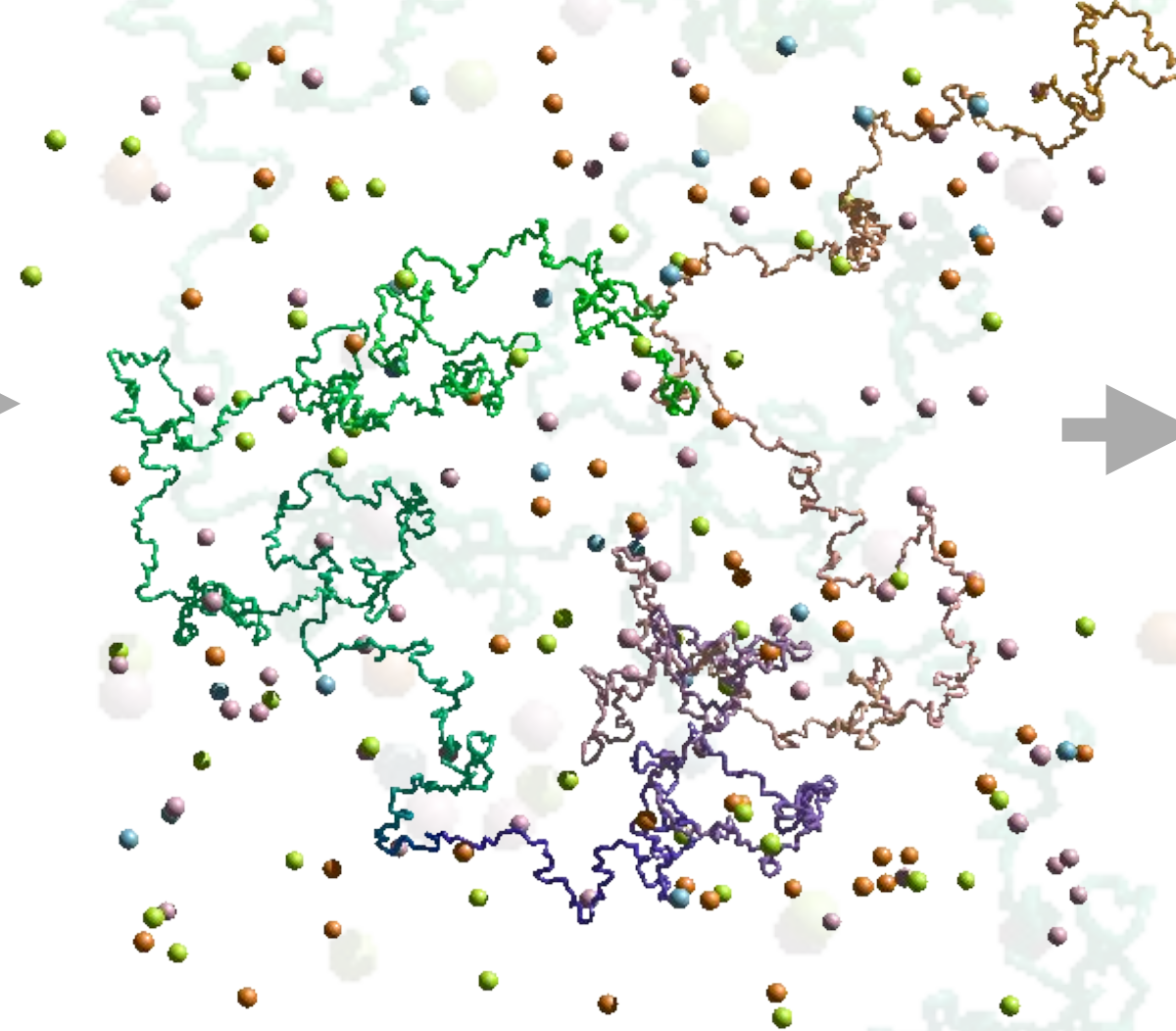


# HPC is a key resource to integrate theory and experiments and simulate real genome

Start from experimental  
(e.g. Hi-C) data



Define a polymer model  
Training with data

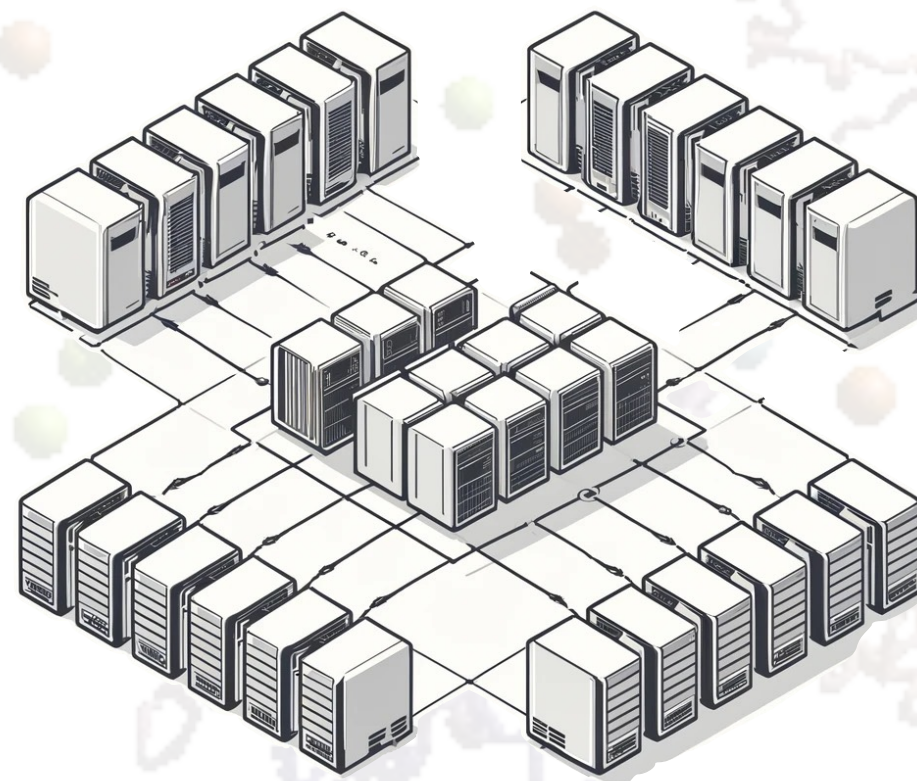


Polymer length N:  $10^2$ - $10^4$   
beads  
Molecules:  $10^1$ - $2 \cdot 10^4$   
Interactions list



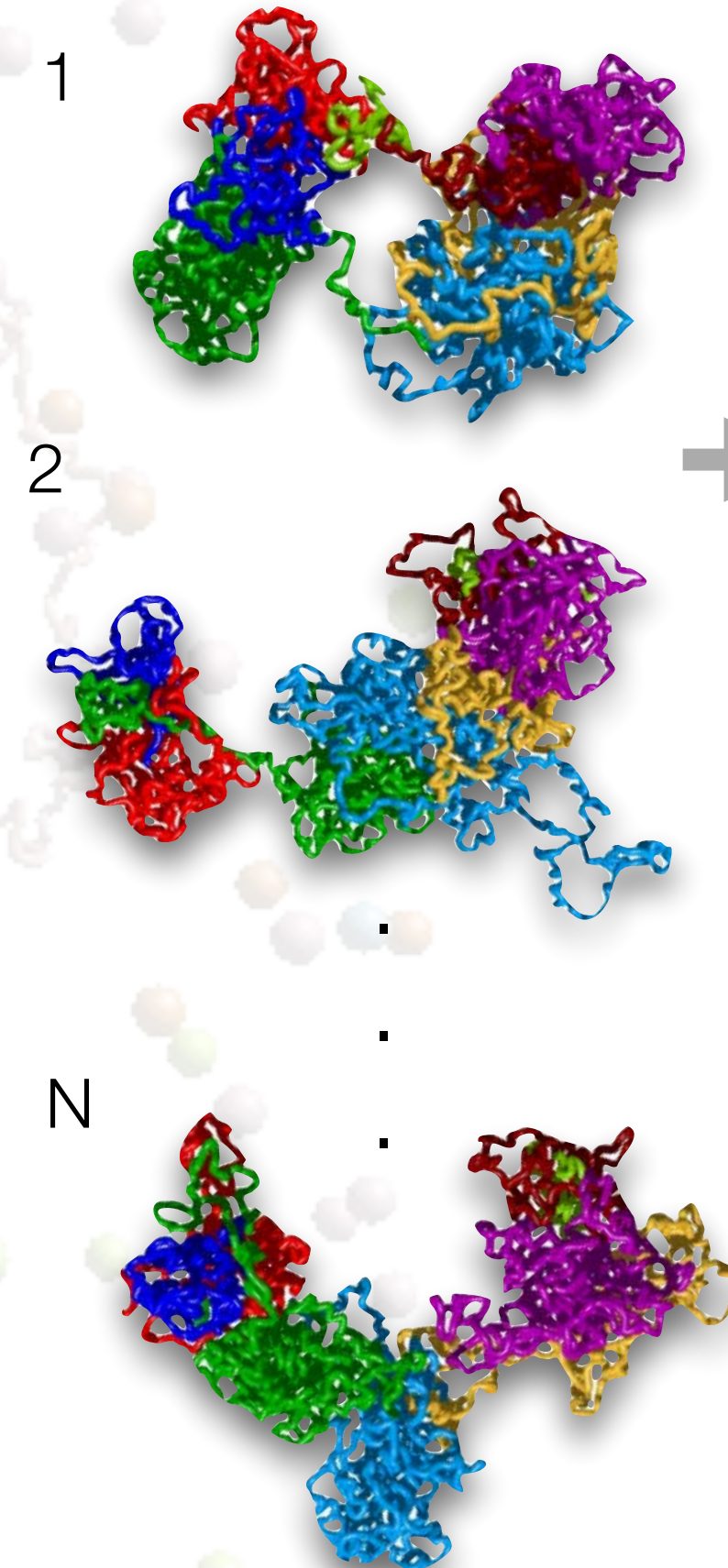
Perform massive  
parallel Molecular  
Dynamics simulations

$$m \frac{d^2 \mathbf{x}_i}{dt^2} = -\nabla_i V - \gamma \frac{d\mathbf{x}_i}{dt} + \sqrt{2k_B T \gamma} \boldsymbol{\eta}(t)$$



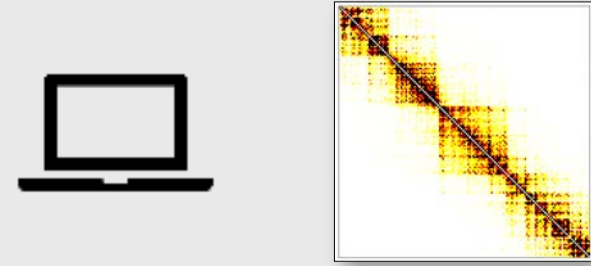
Generate ~50-100 dynamics  
Integrate  $\sim 10^8$  timesteps  
( $\sim 10^5$  core hours)  
Do that for each parameter  
choice and condition

Derive an ensemble  
of 3D structures  
Bona-fide representation of  
real chromosome structures

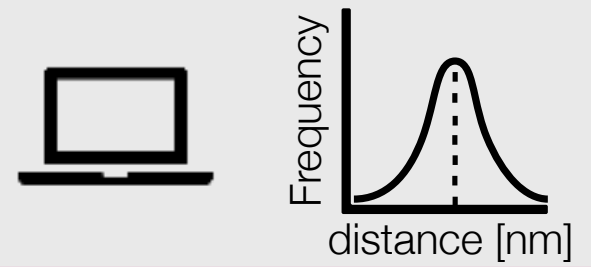


*In-silico* data  
prediction

Testing physical  
mechanisms involved



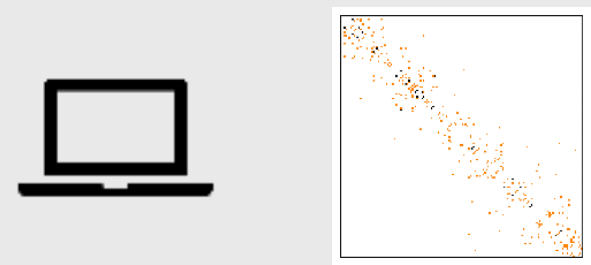
Prediction of genetic  
mutations



*Multi-way* contacts

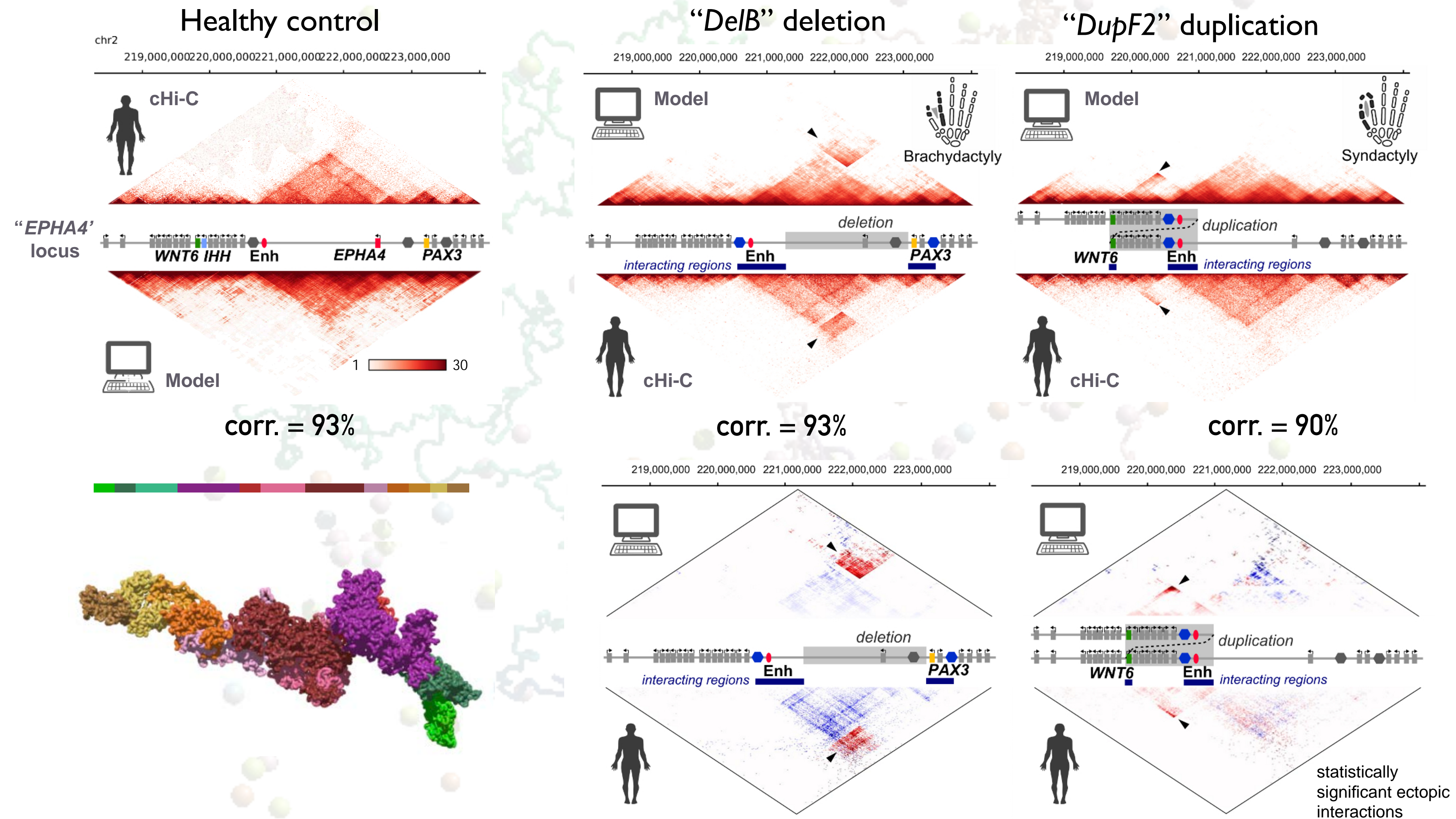


“Single cell” version  
of observables



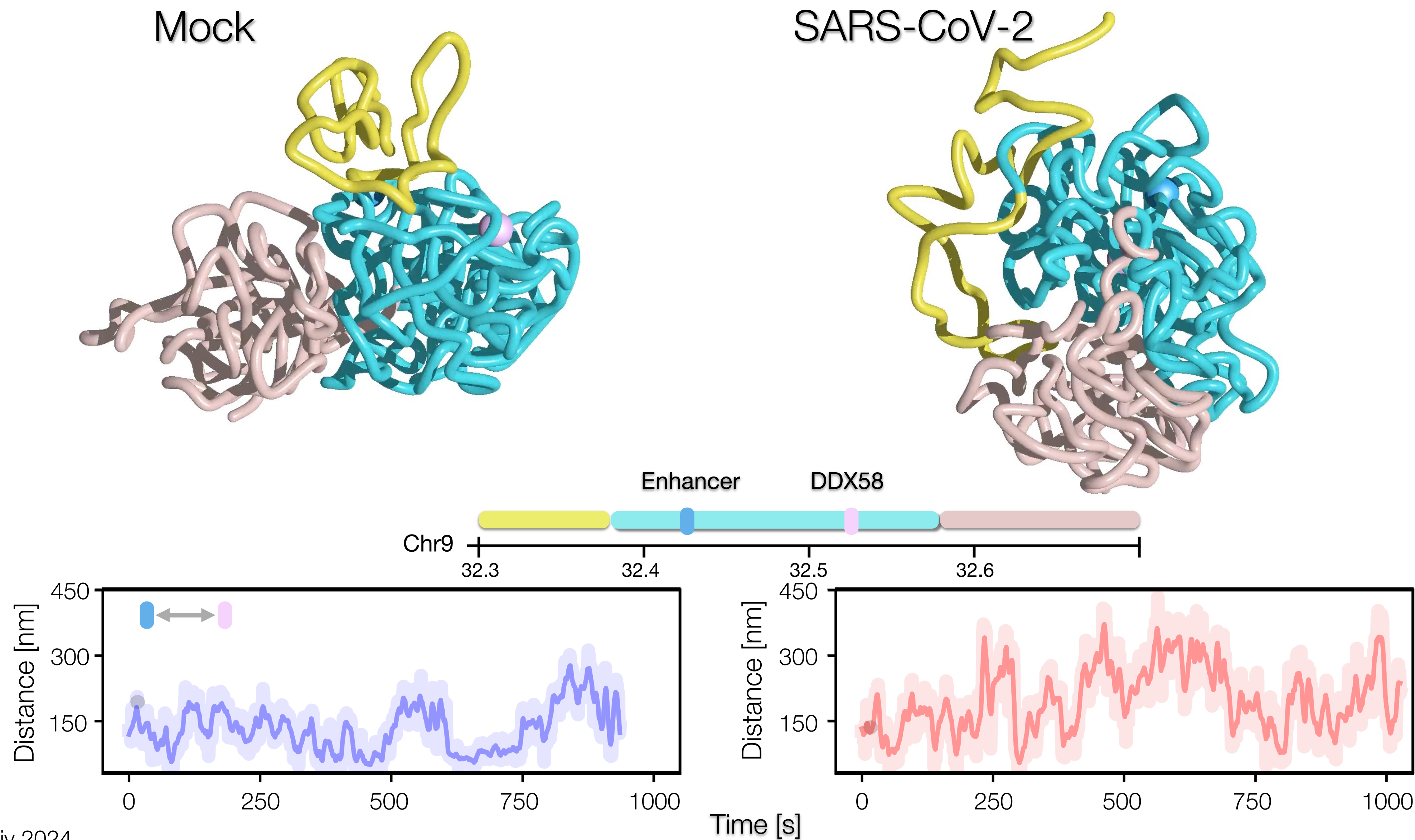


# Predict the impact of pathogenic genetic mutations





# Understand how SARS-CoV-2 changes genome structure





# A synergistic strategy combining theory, HPC and experiments

## nature genetics

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾ [Subscribe](#)

[nature](#) > [nature genetics](#) > [letters](#) > article

Letter | Published: 16 April 2018

**Polymer physics predicts the effects of structural variants on chromatin architecture**

## nature methods

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature methods](#) > [analyses](#) > article

Analysis | [Open access](#) | Published: 07 May 2021

**Comparison of the Hi-C, GAM and SPRITE methods using polymer models of chromatin**

## nature

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [articles](#) > article

Article | [Open access](#) | Published: 17 November 2021

**Cell-type specialization is encoded by specific chromatin topologies**

## nature communications

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature communications](#) > [articles](#) > article

Article | [Open access](#) | Published: 13 July 2022

**Loop-extrusion and polymer phase-separation can co-exist at the single-molecule level to shape chromatin folding**

## nature

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [articles](#) > article

Article | [Open access](#) | Published: 20 December 2023

**RNA-mediated symmetry breaking enables singular olfactory receptor choice**

HPC resources like IbISCo  
fundamental to invent,  
test and validate our models



# Conclusions

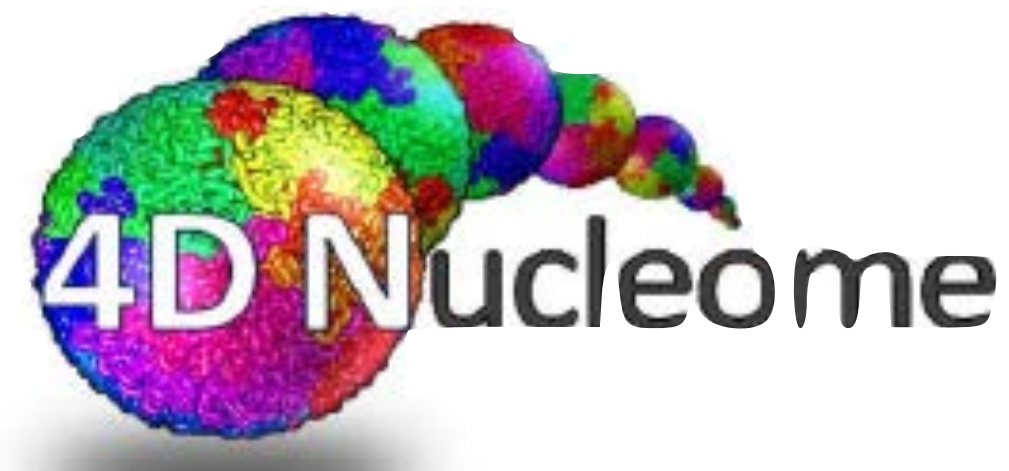
- Polymer physics models including physical mechanisms quantitatively describe genome folding;
- The combination of experiment, theories and computation is essential to make models accurate and predictive;
- HPC is pivotal in using those models to simulate such highly complex systems; all those kind of research would not be possible without resources like IBiSCo.





# Acknowledgements

---



## **Complex Systems Group**

Mario Nicodemi

Andrea Esposito

Alex Abraham

Simona Bianco

Andrea Fontana

Sougata Guha

Mattia Conte

Florinda di Piero

Francesca Vercellone

## Collaborators:

- Dr. Wenbo Li, Texas University, Houston
- Prof. Bing Ren, UC San Diego
- Prof. Ana Pombo, Max Delbrück Center for Molecular Medicine, Berlin
- Prof. Stavros Lomvardas, Columbia University, New York
- Dr. Sarah Kinkley, Max Planck Institute for Molecular Genetics, Berlin