

Decoding Genome Folding with Chromatin Tracing and 3D Modelling Approaches

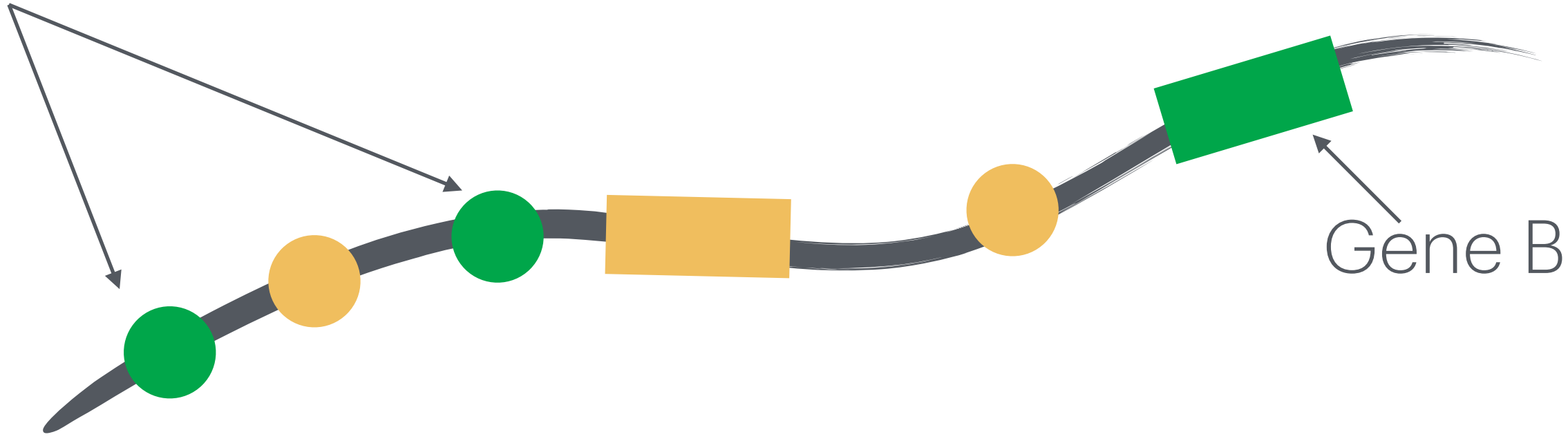
Irene Farabella
irene.farabella@iit.it

Physical sensing and processing - Summer school
Unibo - Virtual
7 July 2025

From Linear Code to Spatial Structure: Understanding Genome Organization

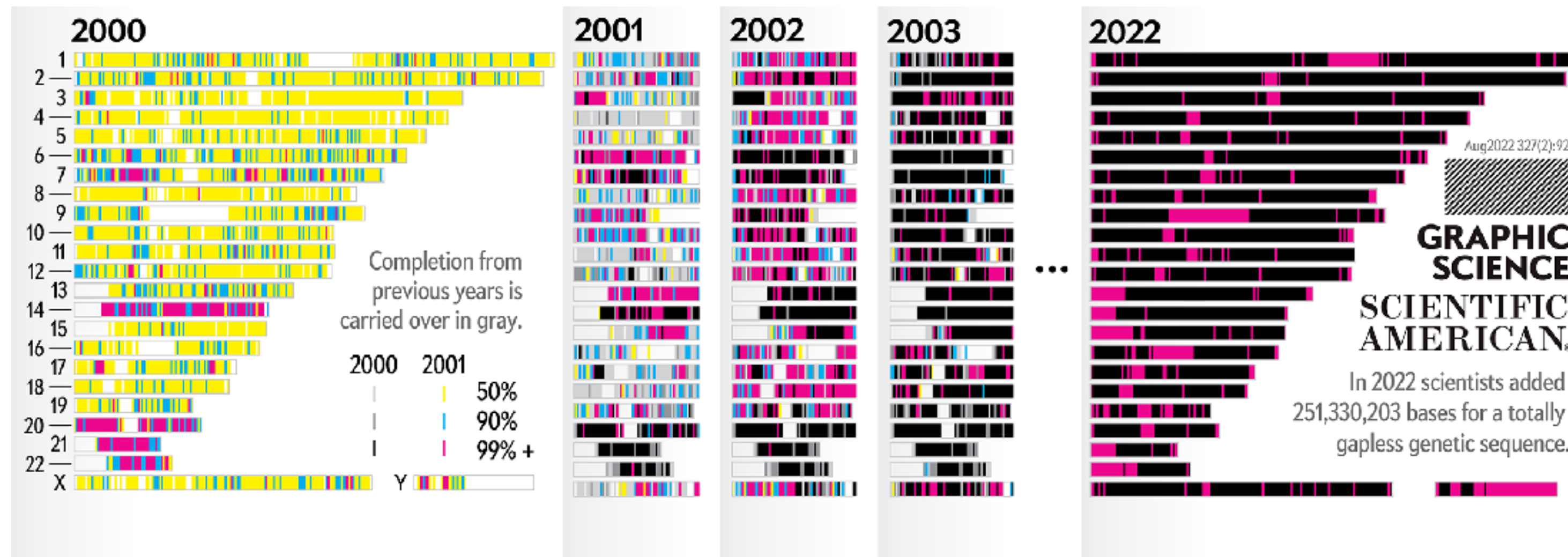


Gene B enhancers

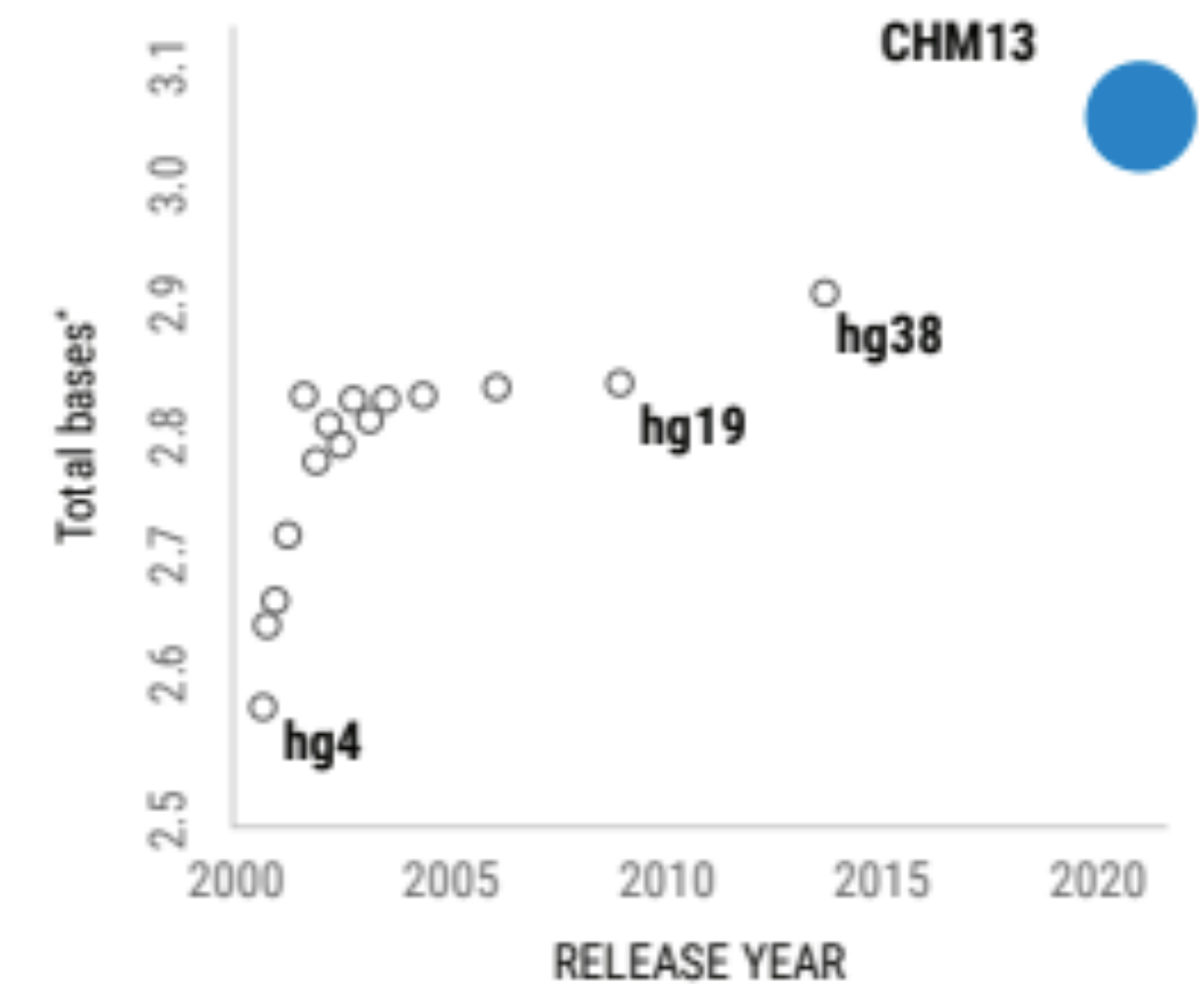


Gene B

From Linear Code to Spatial Structure: Understanding Genome Organization



<https://mk.bcgsc.ca/>

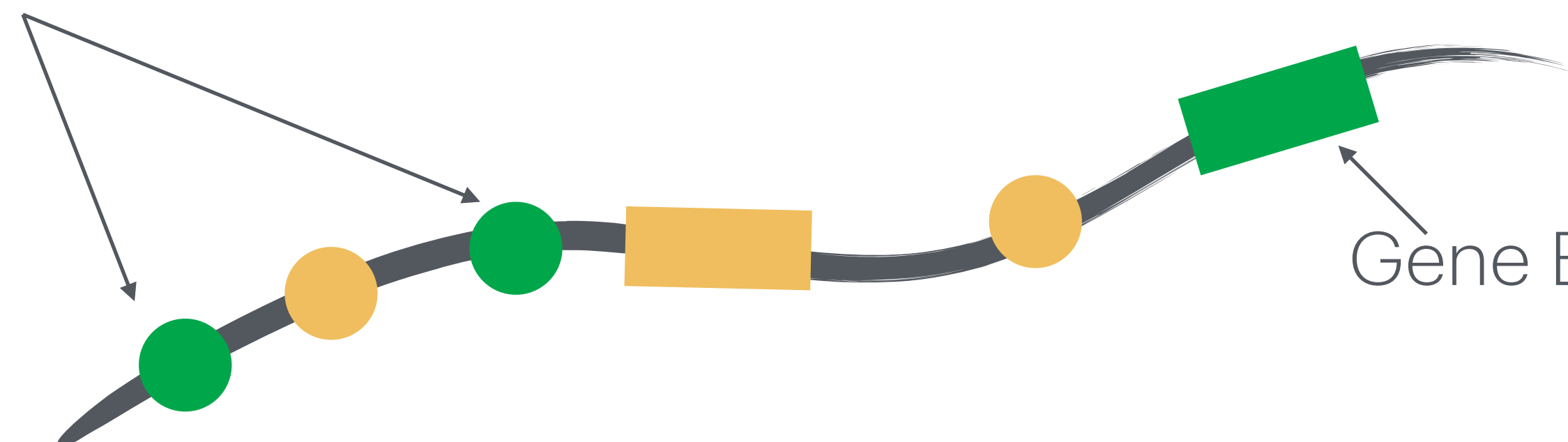


Total non-gap bases in UCSC reference genome releases dating back to September 2000 (hg4) and ending with T2T-CHM13 in 2021

0.34nm 1 bp

$0.34 \times 10^{-9} \times 3.1 \times 10^9 \times 2 \sim 2 \text{ metri DNA}$

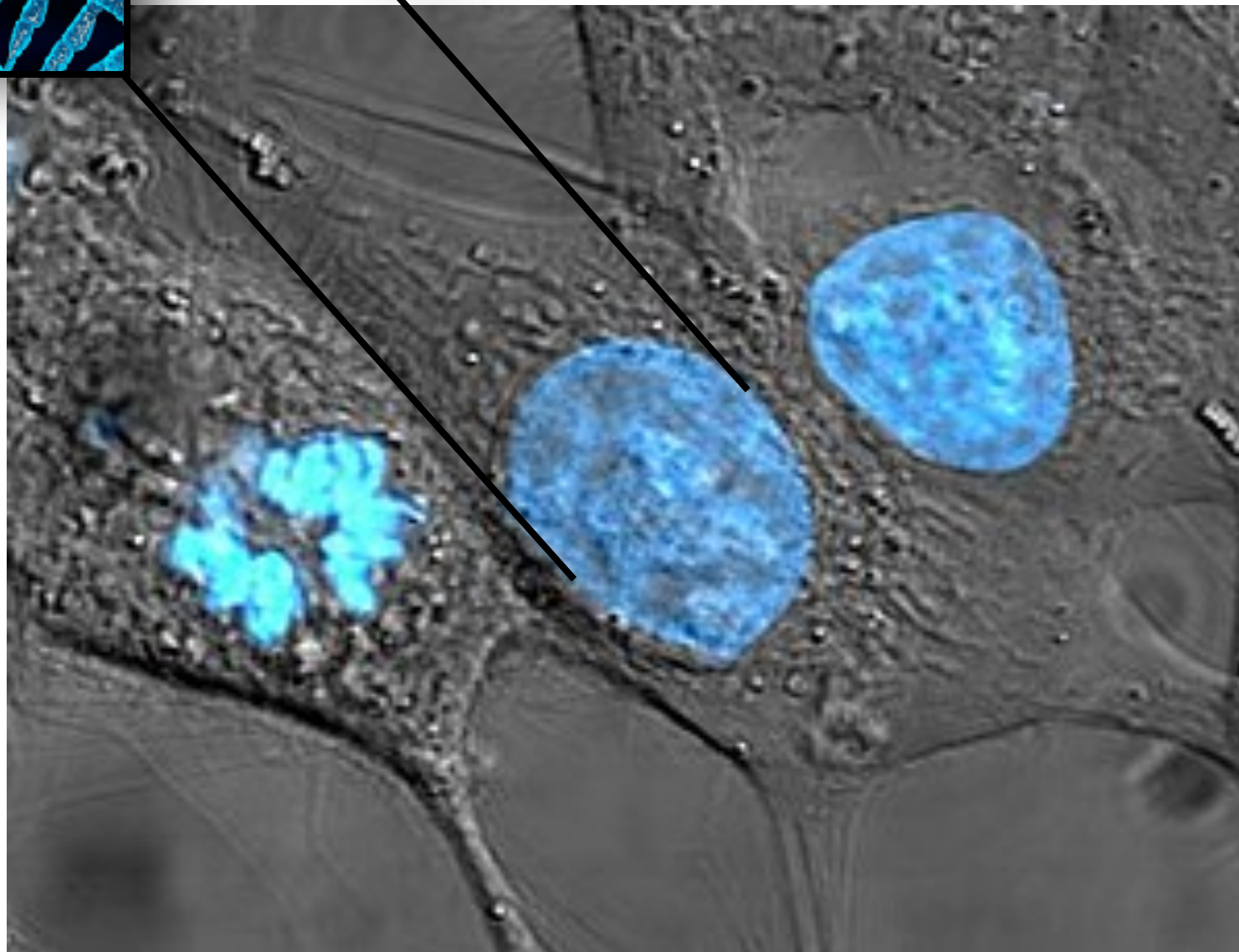
Gene B enhancers



How does the genome fit into a nucleus?



~2 m



~ 6 μm (10^{-6} m)



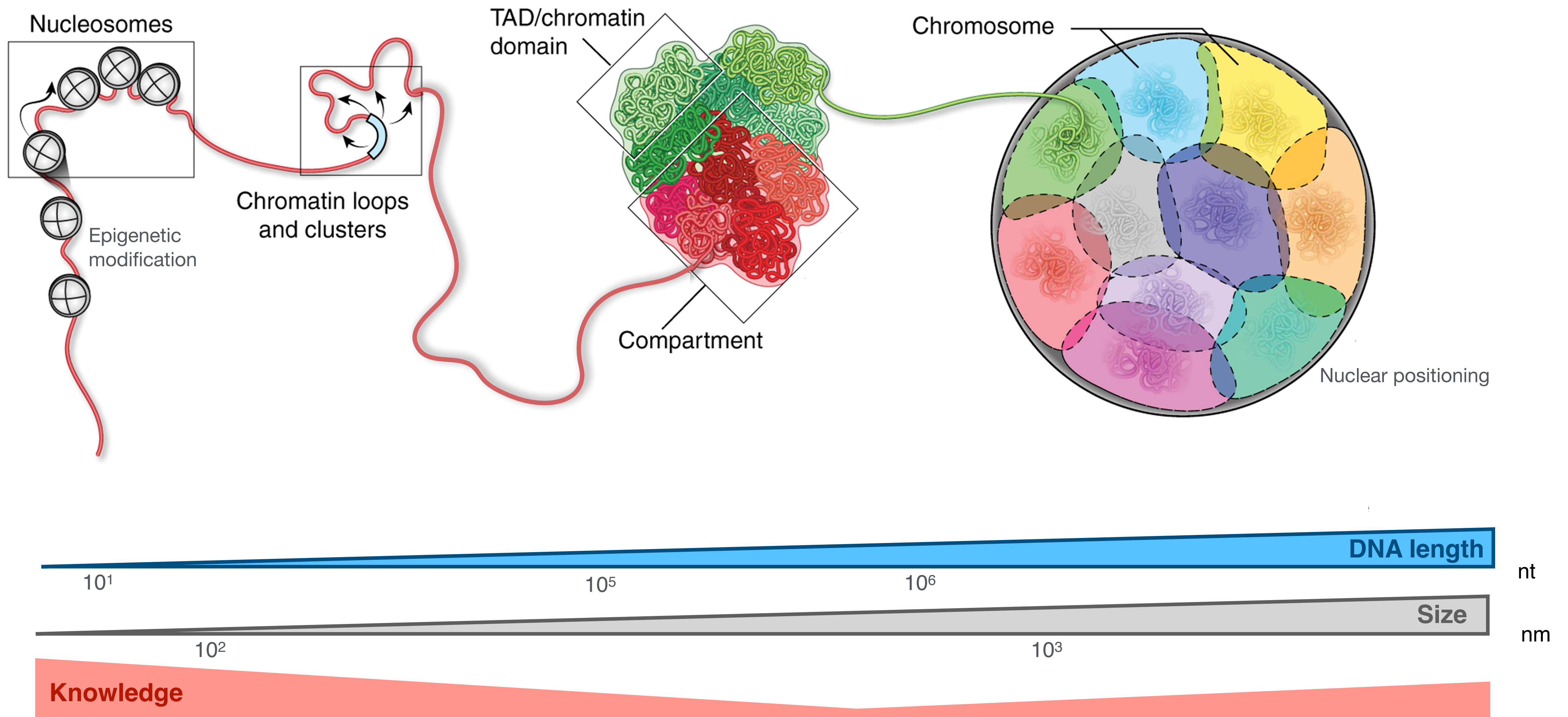
30 000 km



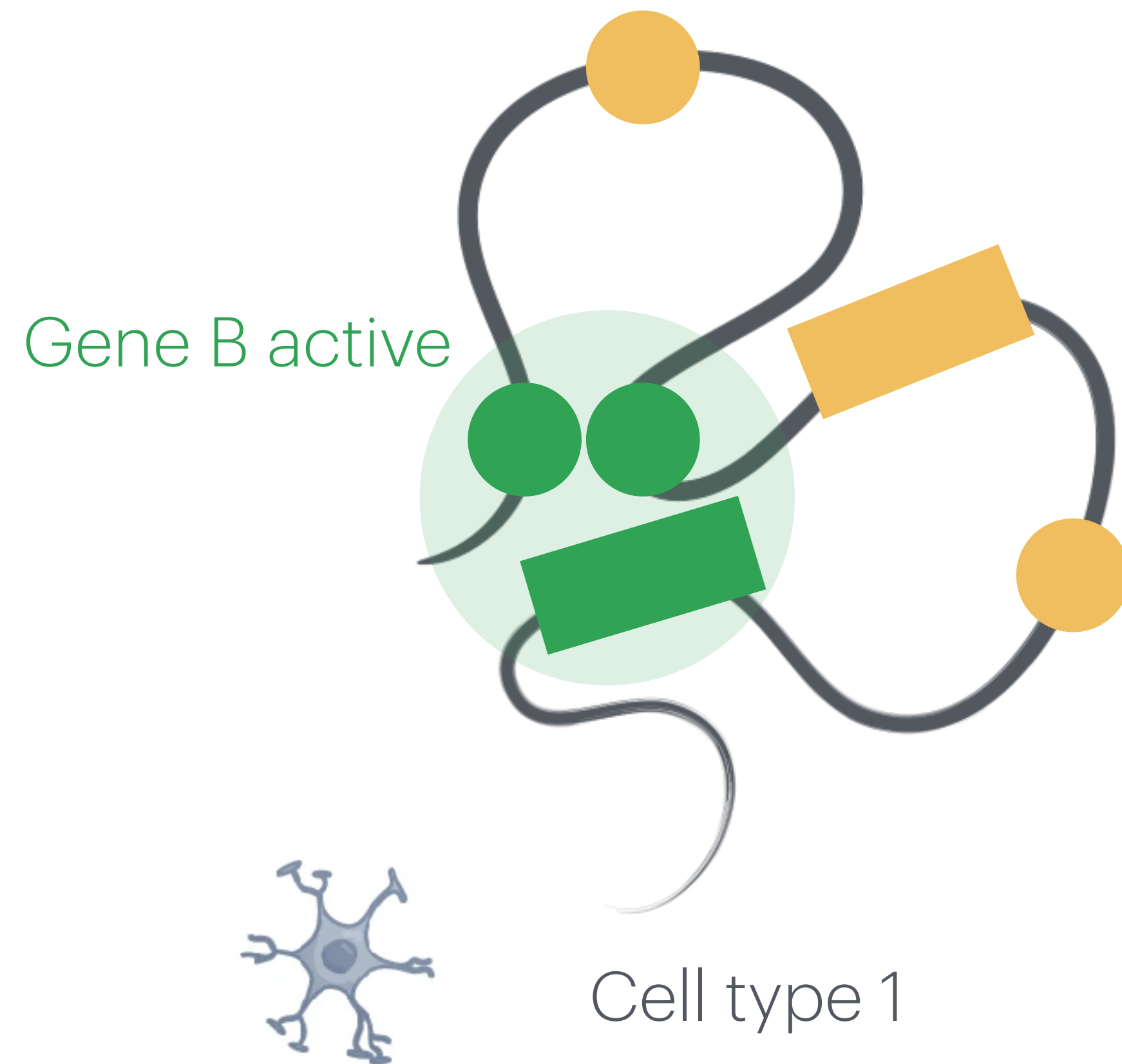
~ 100 m

If the nucleus were the size of a football pitch (≈ 100 metres), then the uncoiled DNA would be over 30,000 km long — almost enough to circle the Earth once!

The genome is organised at multiple scales



The 3D genome organisation is linked to biological function



Studying Genome Topography: Tools and Approaches

Sequence-based technologies

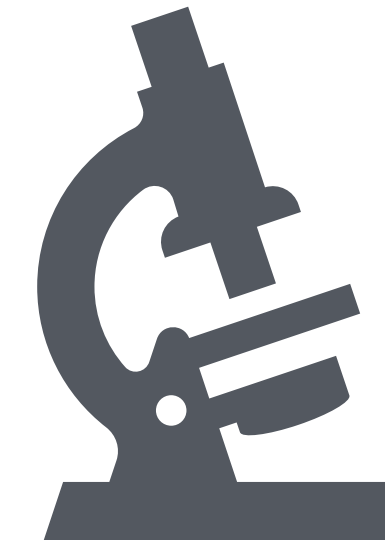
Next-Generation Sequencing (NGS)



Big Data



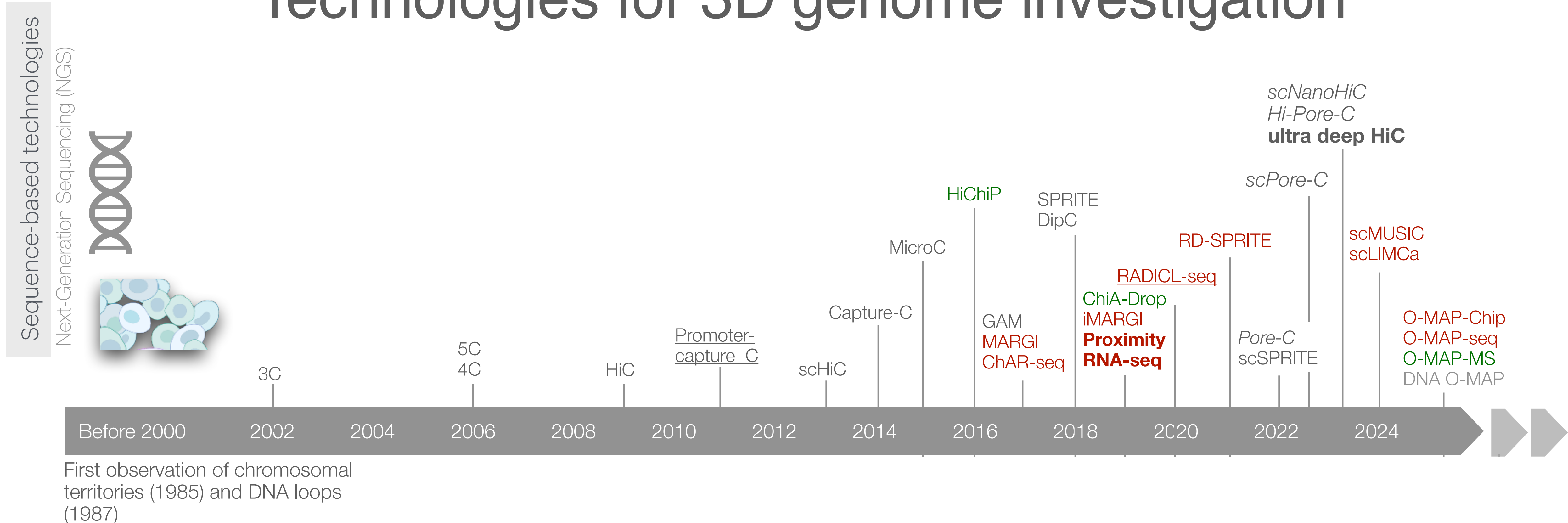
Imaging-based technologies



Computational tools

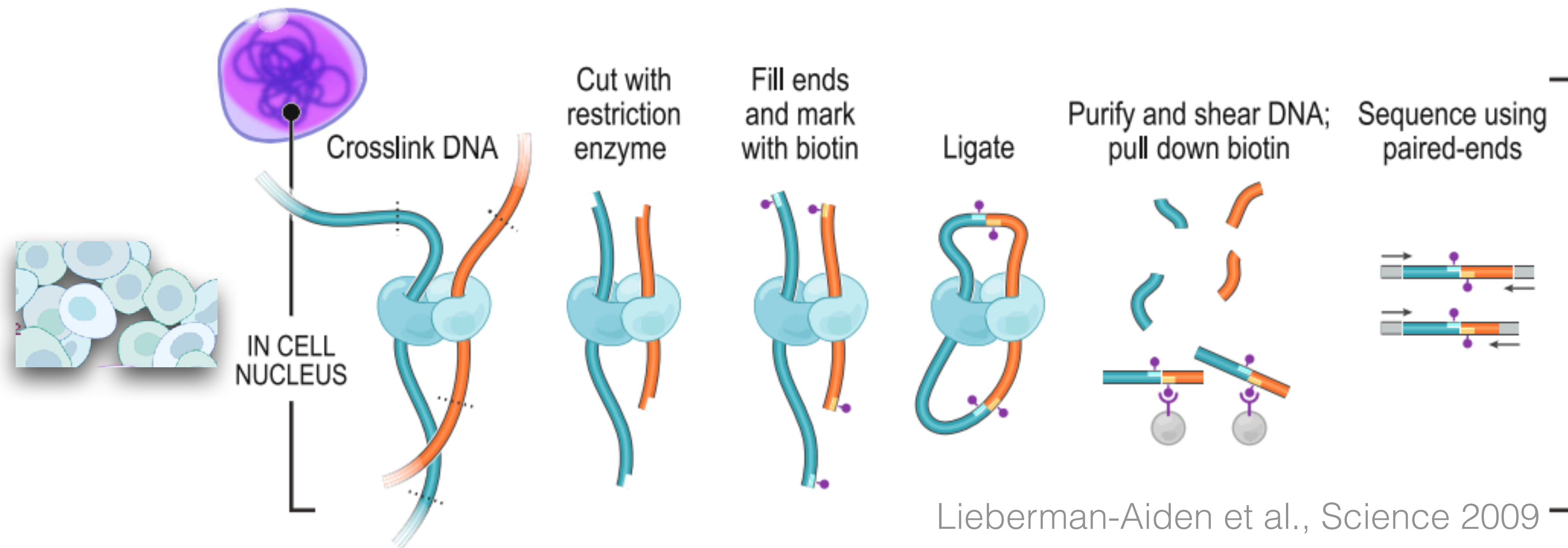


Technologies for 3D genome investigation



Multiple mechanisms of 3D genome organisation

Higher-order organisation



Sequencing-based experiments (cells population information)

- Capture (C)-based methods
e.g. HiC, pHiC, Micro-C, ...

- Ligation-independent methods
e.g. SPRITE, GAM, ..

Multiple mechanisms of 3D genome organisation

Higher-order organisation



Cell type 1

| | ● | ● | ● | ■ | ● | ■ |
|---|----|---|----|---|---|----|
| ● | 12 | 2 | 12 | | 1 | 12 |
| ● | 2 | 6 | 2 | | 1 | |
| ● | 12 | 2 | 12 | | 1 | 10 |
| ■ | | | | 6 | 1 | |
| ● | 1 | 1 | 1 | 1 | 6 | 1 |
| ■ | 12 | | 10 | | 1 | 12 |



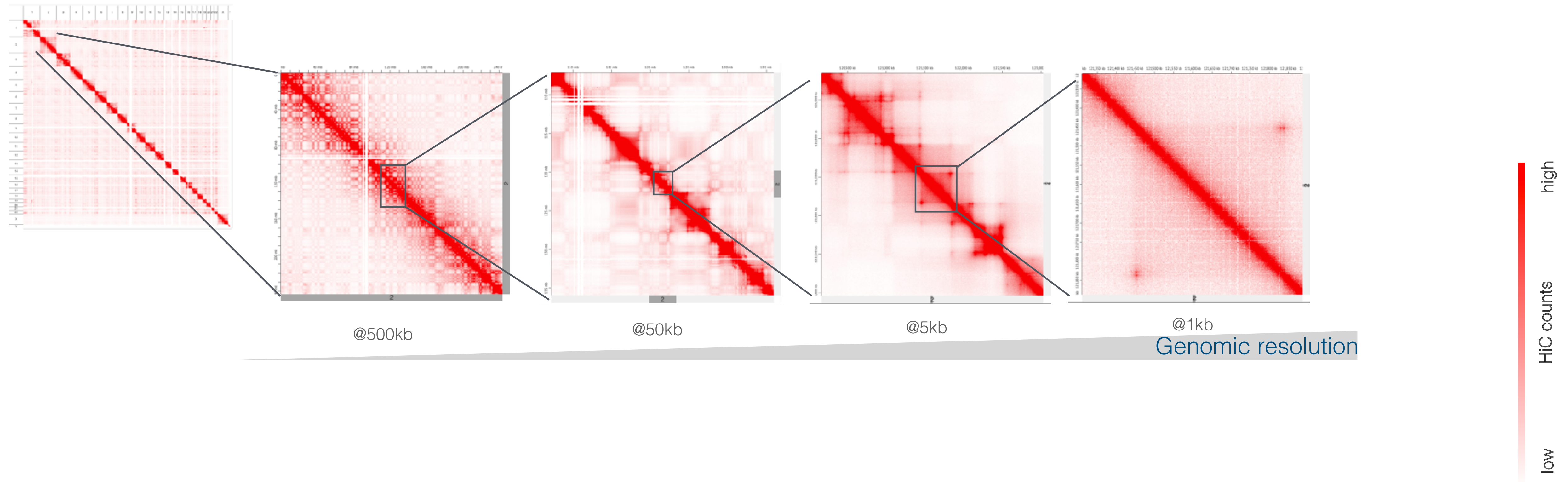
| | ● | ● | ● | ■ | ● | ■ |
|---|---|----|---|----|----|---|
| ● | 6 | 2 | 1 | 2 | 1 | 1 |
| ● | 2 | 12 | | 12 | 12 | |
| ● | 1 | | 6 | | | |
| ■ | 2 | 12 | | 12 | 12 | 1 |
| ● | 1 | 12 | | 12 | 12 | 1 |
| ■ | 1 | | | 1 | 1 | 6 |



Cell type 2

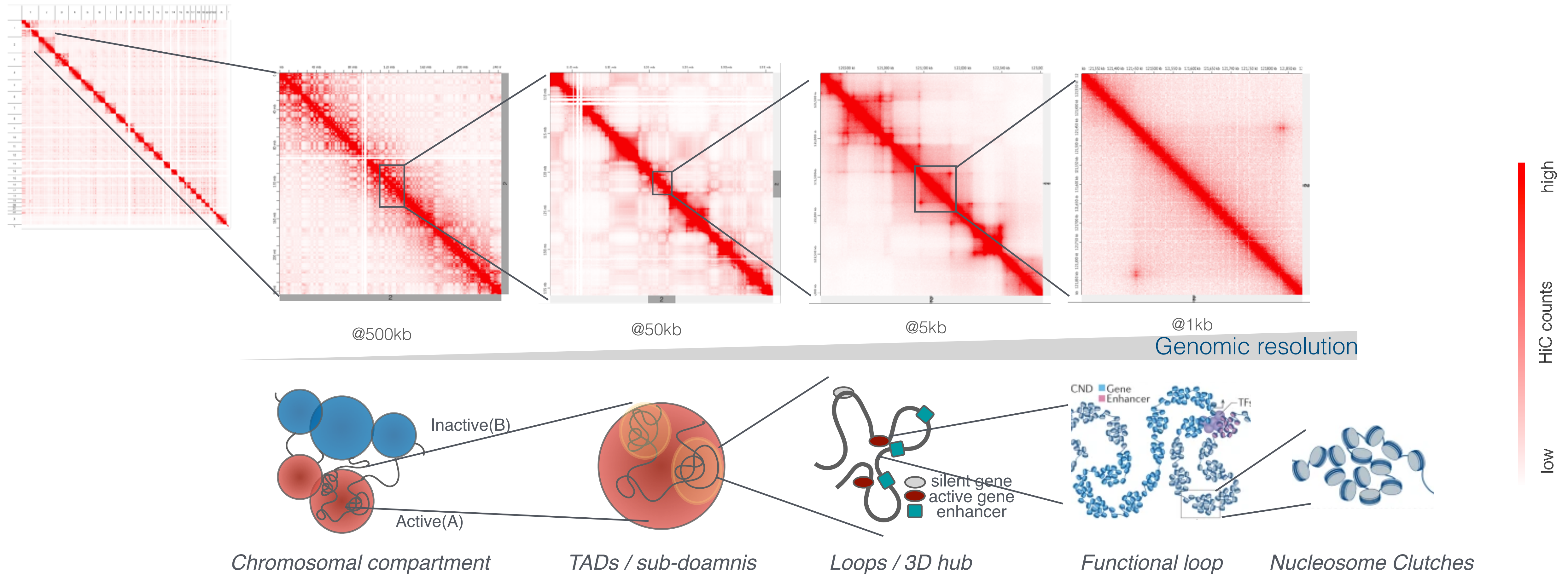
Multiple mechanisms of 3D genome organisation

Higher-order organisation

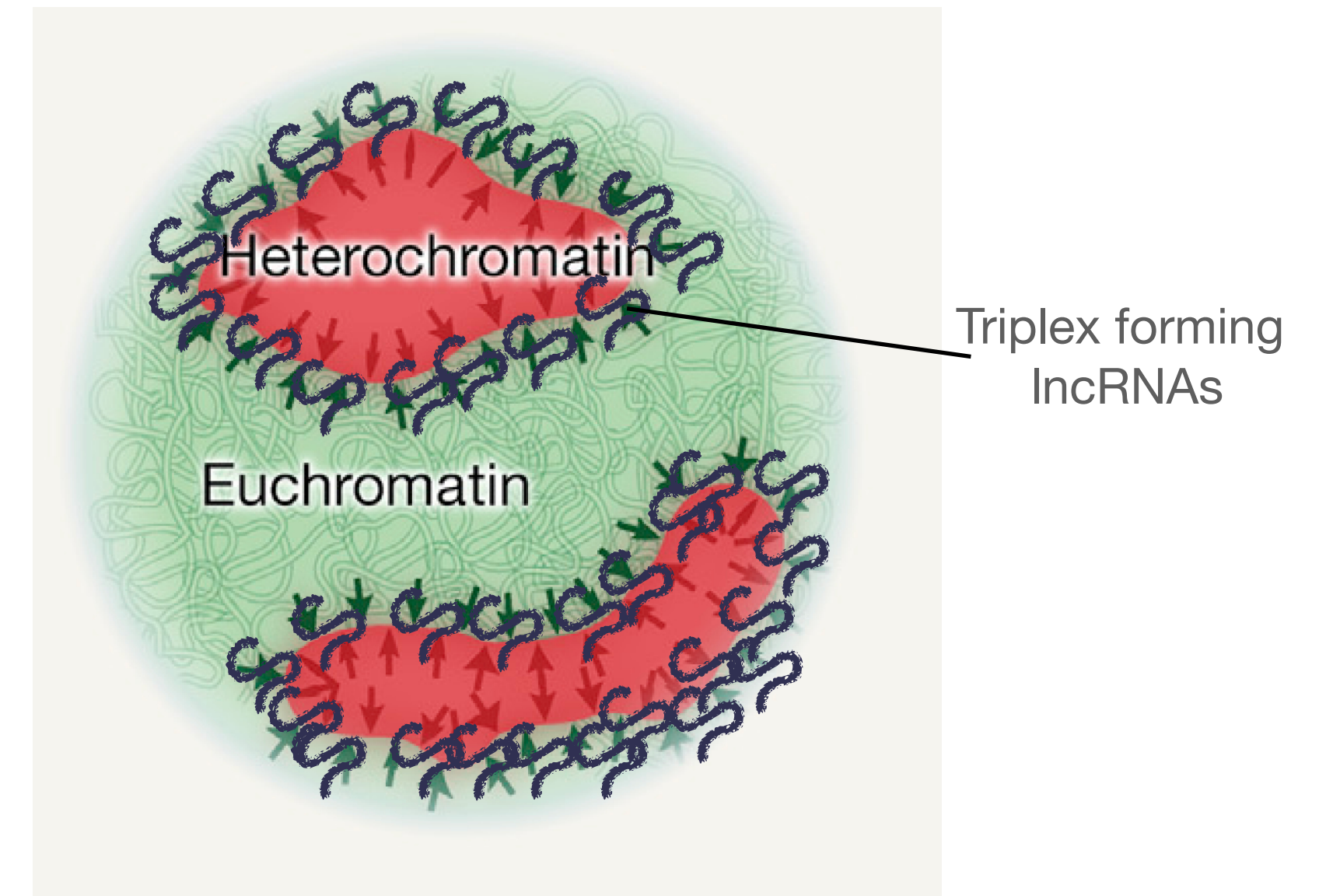
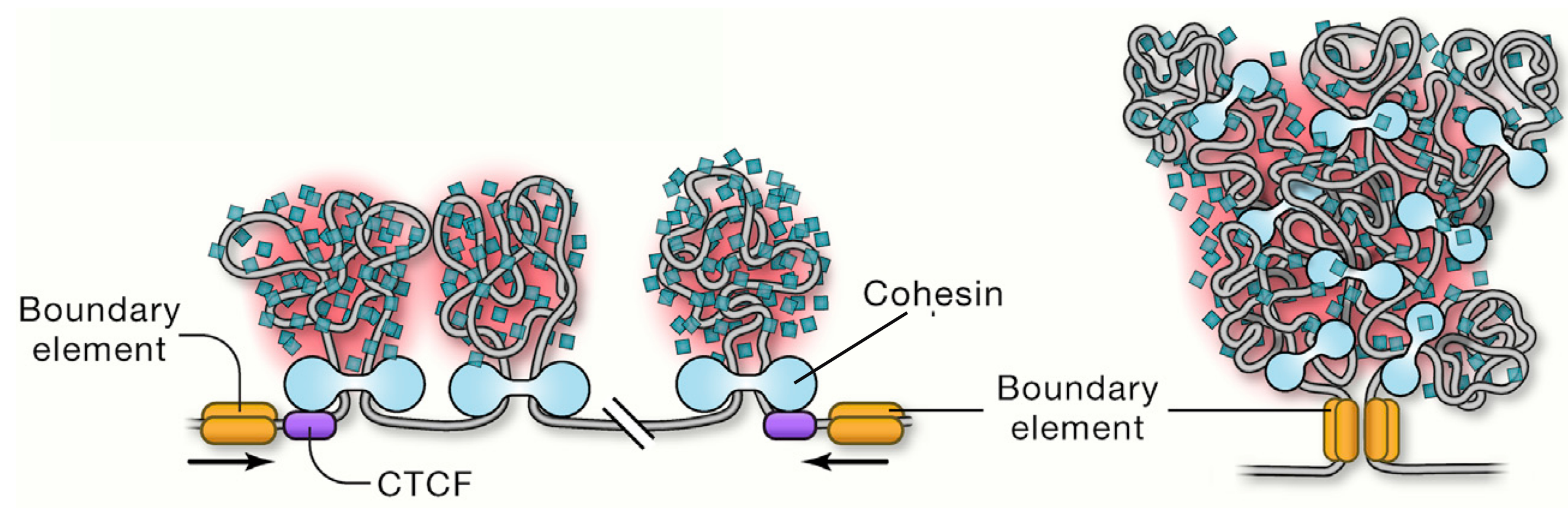


Multiple mechanisms of 3D genome organisation

Higher-order organisation



Multiple mechanisms of chromosome organisation

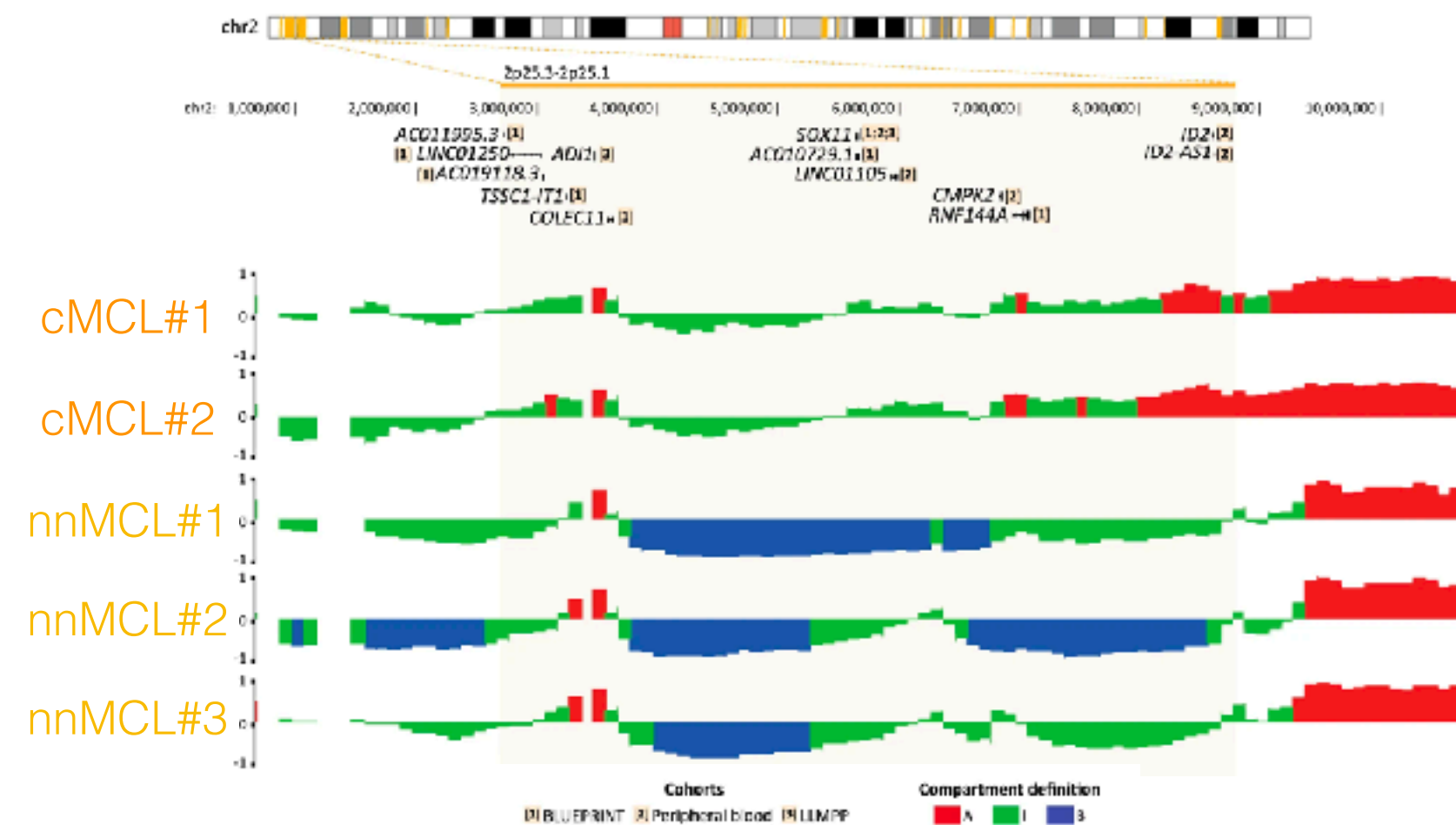
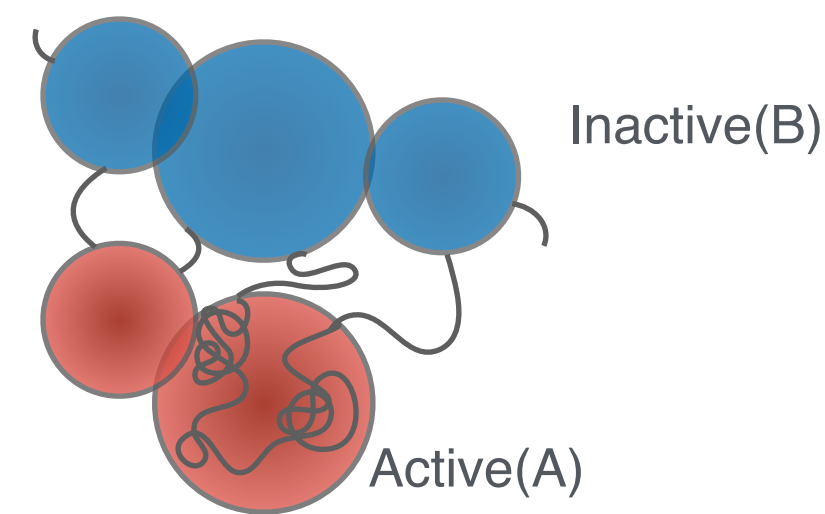


Dixon et al., Nature 2012
Rao et al., 2014 Cell
Nora et al., 2017 Cell
Nuebler et al., 2018 PNAS
Mirny, Imakaev, Abdennur. 2019 Curr Opin Cell Biol.
Morf ... Farabella ..et al., 2019 Nat. Biotech.
Farabella et al. 2021 accepted Nat. Struct. Mol. Biol.

Adapted from Misteli Cell 2020

Genome organisation

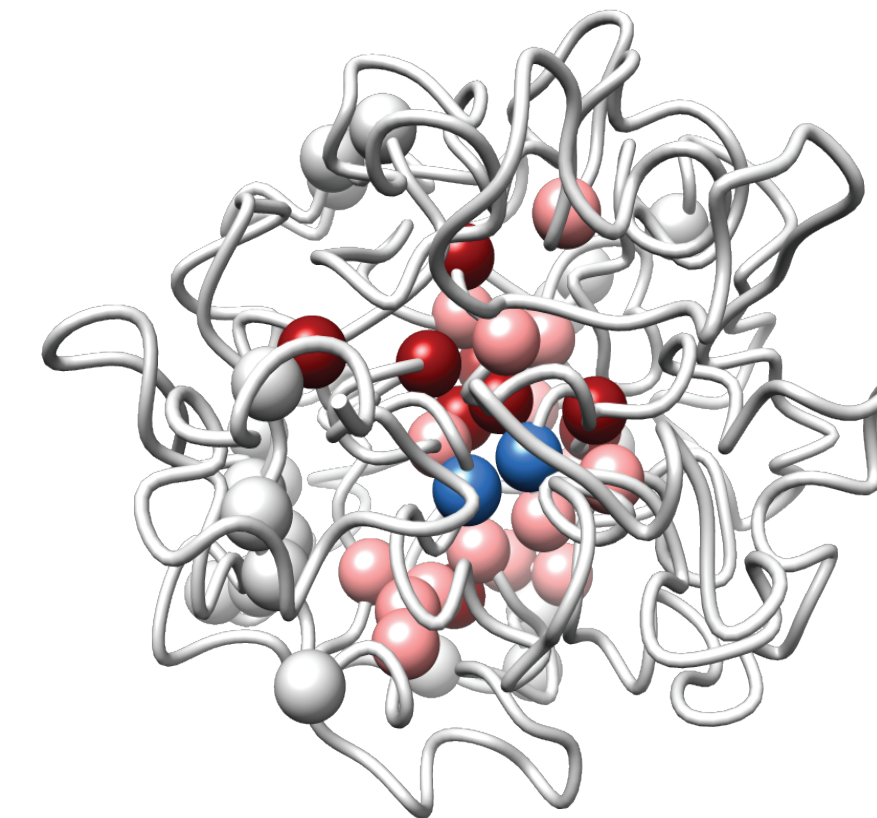
Sequencing-based experiments (cells population information)



SOX11 marker in mantle cell lymphoma (MCL)

cMCL clinically aggressive conventional MCL
nnMCL clinically indolent non-nodal leukemic MCL

Promoter
Enhancer
>200nm



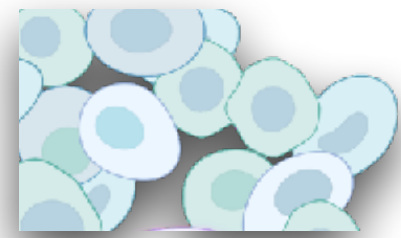
pcHi-C 1.9 Mb ILS1 locus @5kb
in human pancreatic islets cell

3D enhancer hub

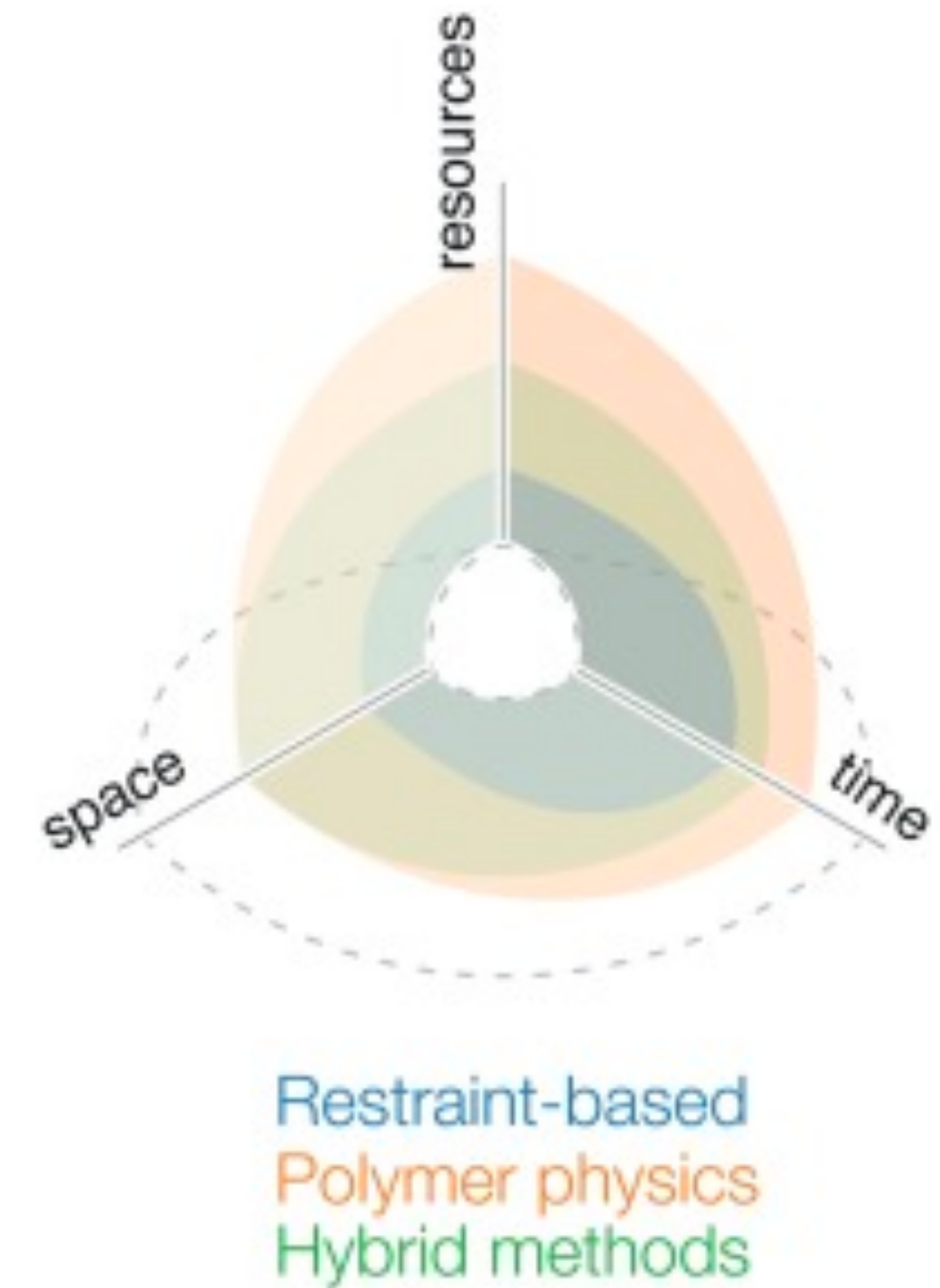
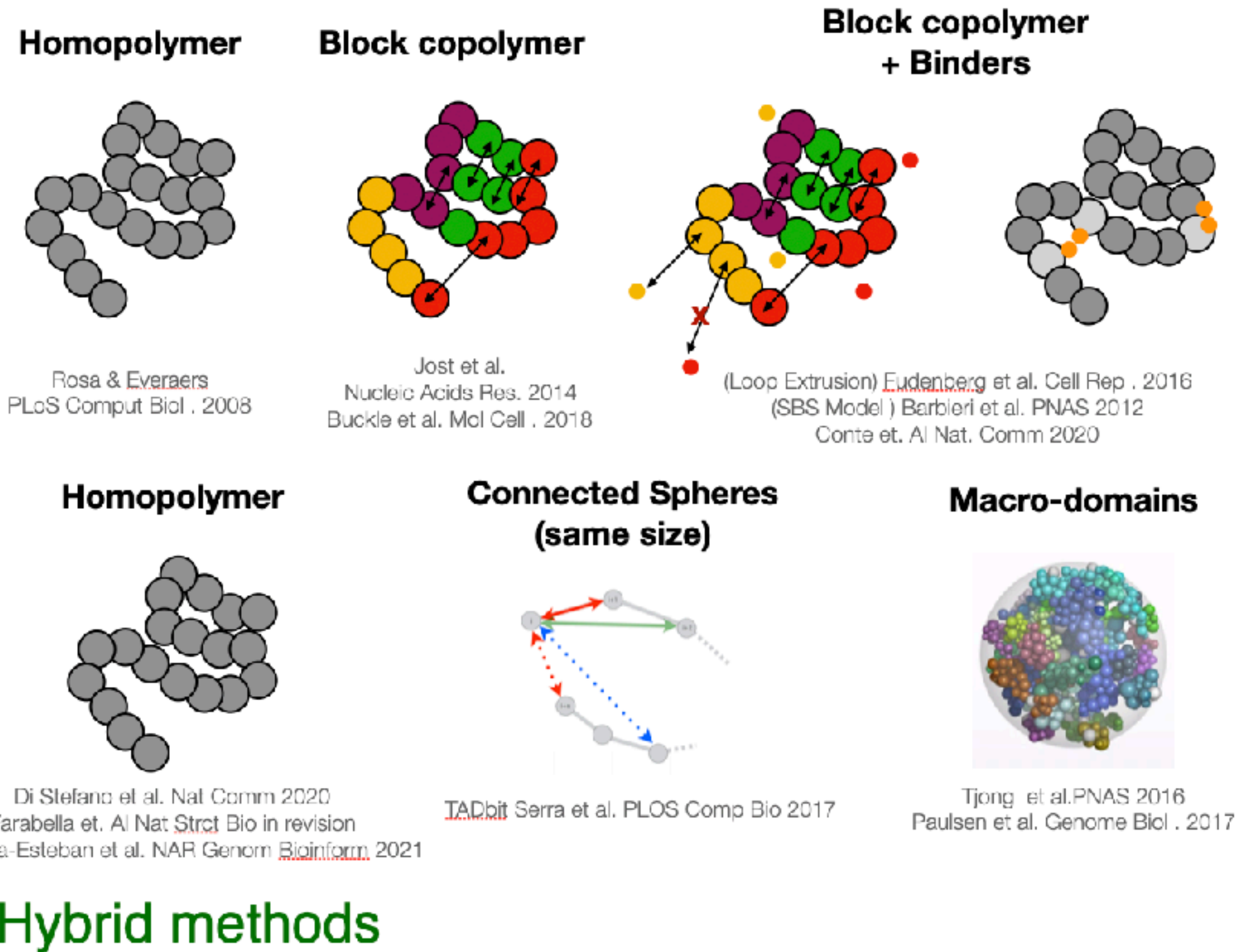
Genetic variation in islet regulatory elements hubs impacts insulin secretion heritability in T2D and hub annotations can be used for polygenic scores that predict T2D risk driven by islet regulatory variants.

From Contact Matrices to 3D Models: Computational Tools to Reconstruct Genome Architecture

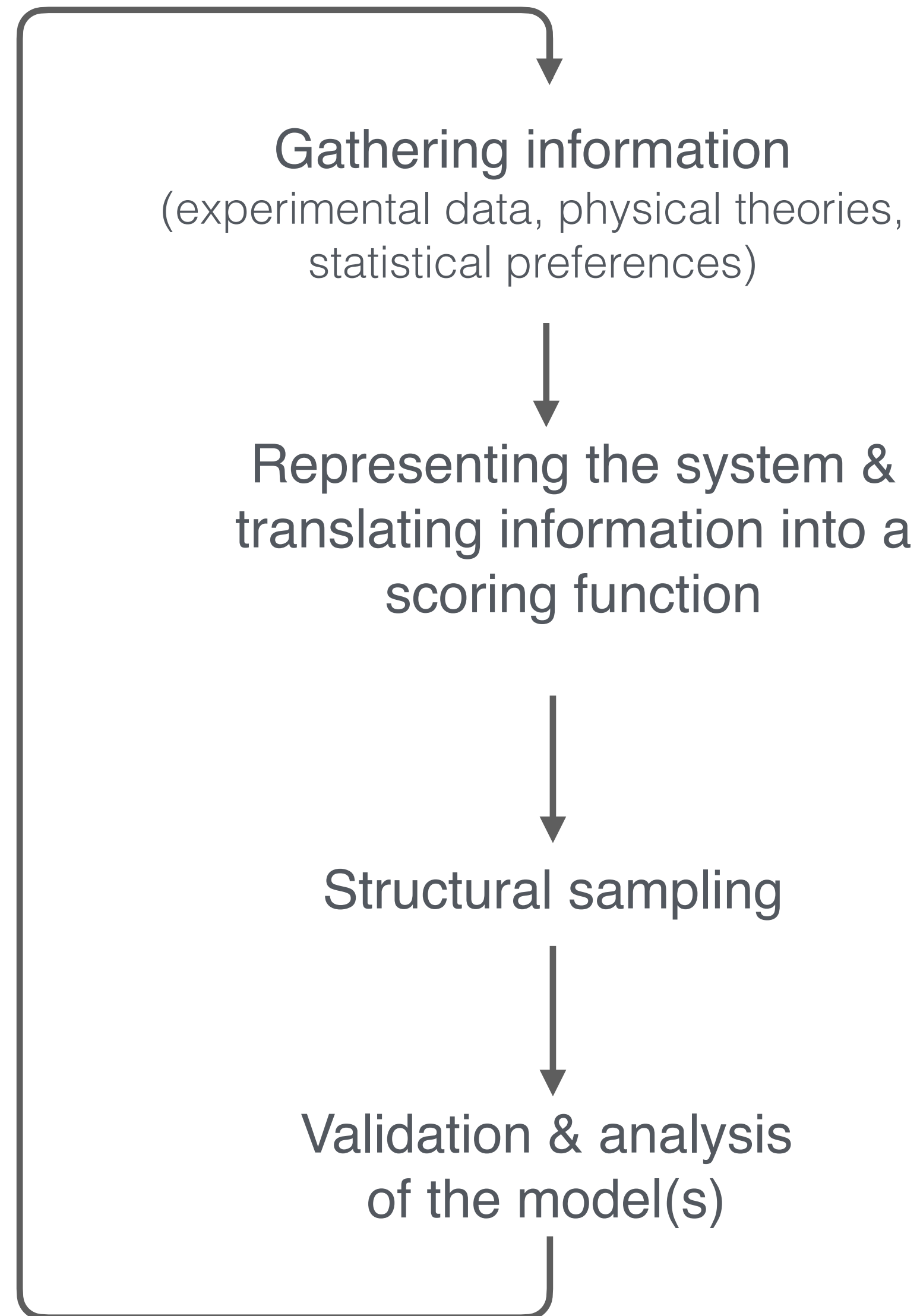
Theory-driven/
Mechanistic/Bottom-up
Modelling



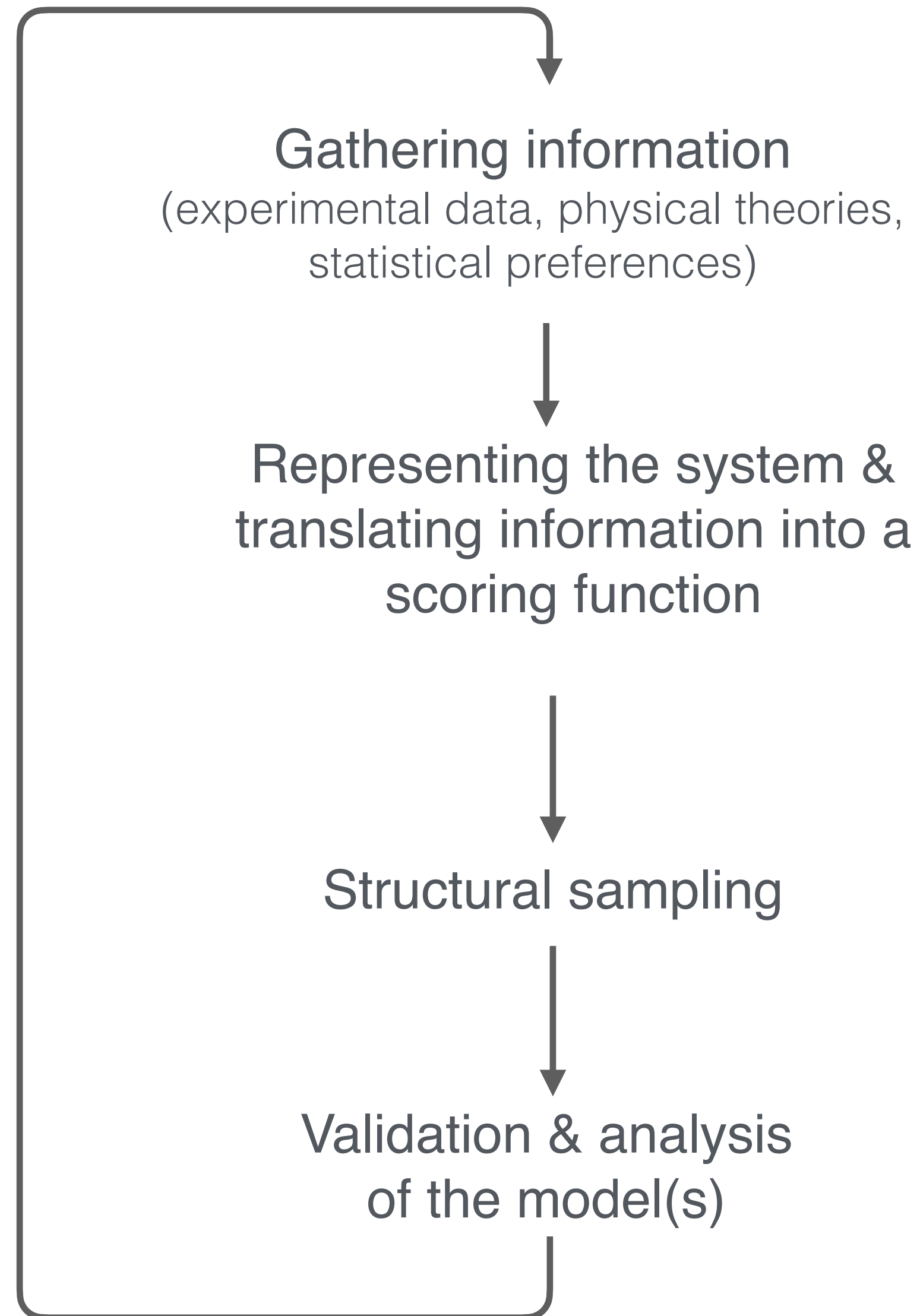
Data-driven/Top-down/
Integrative Modelling



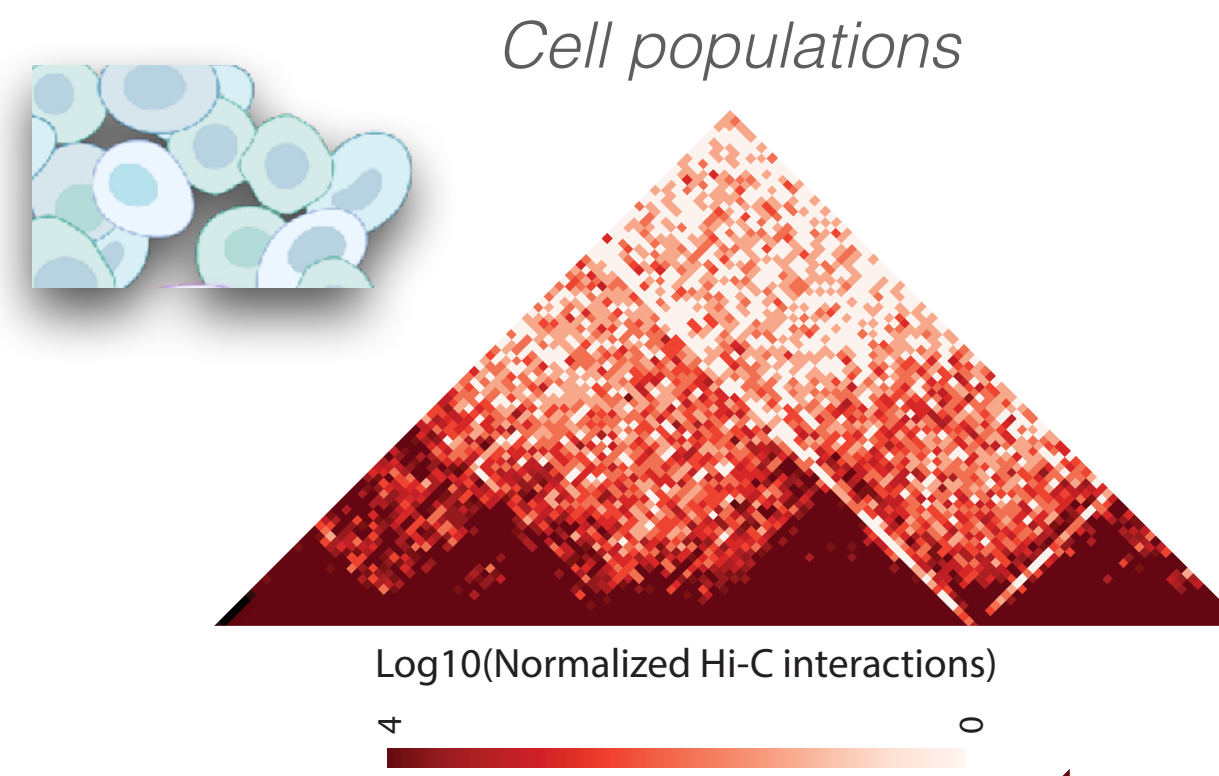
Integrative Modelling



Integrative Modelling



Contact Informations

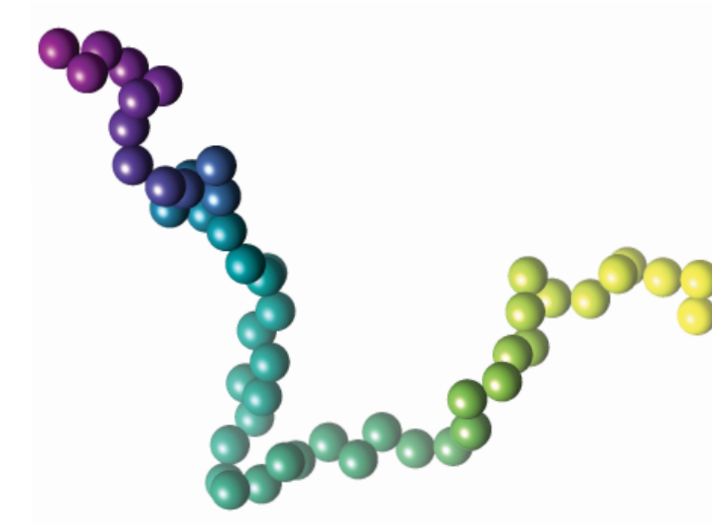


Statistical preferences

Nuclear component binding propensity in cell population



Integrative Modelling



Chromatin fibers, macro domain, beads-on-spring representation

Gathering information
(experimental data, physical
theories, statistical preferences)

**Representing the system
& translating information
into a scoring function**

Structural sampling

Validation & analysis
of the model(s)



Di Stefano et al., Nat. Commun 2020

Serra et al., PLoS Comput Biol. 2017

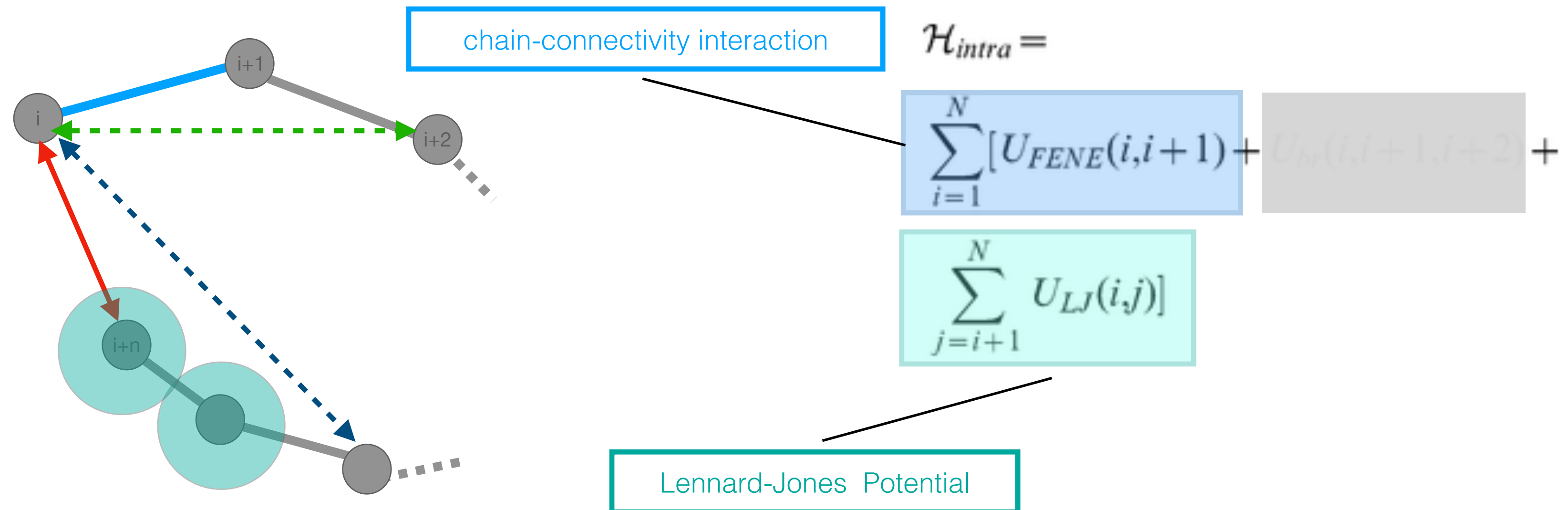
Integrative Modelling

Gathering information
(experimental data, physical
theories, statistical preferences)

Representing the system
& translating information
into a scoring function

Structural sampling

Validation & analysis
of the model(s)



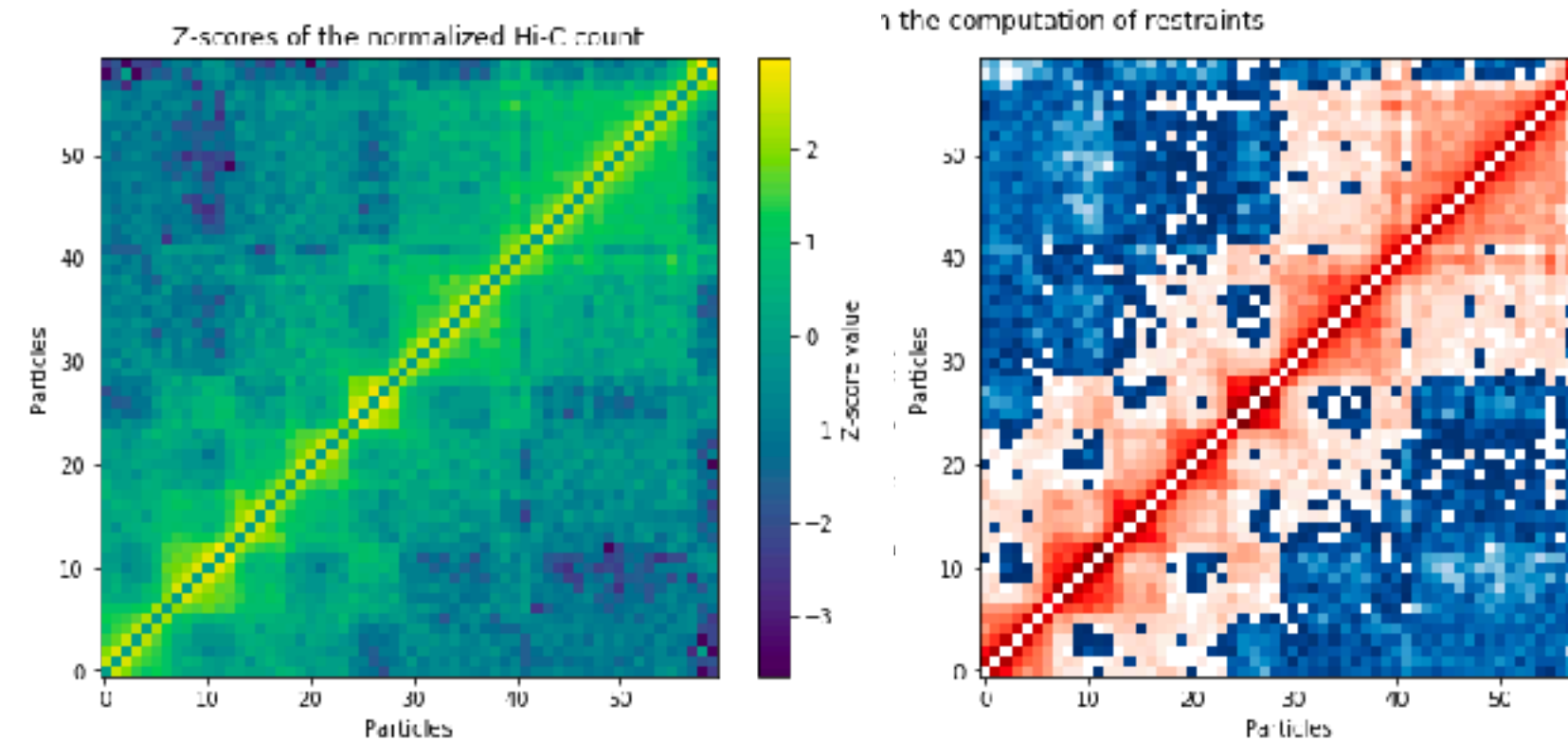
Integrative Modelling

Gathering information
(experimental data, physical theories, statistical preferences)

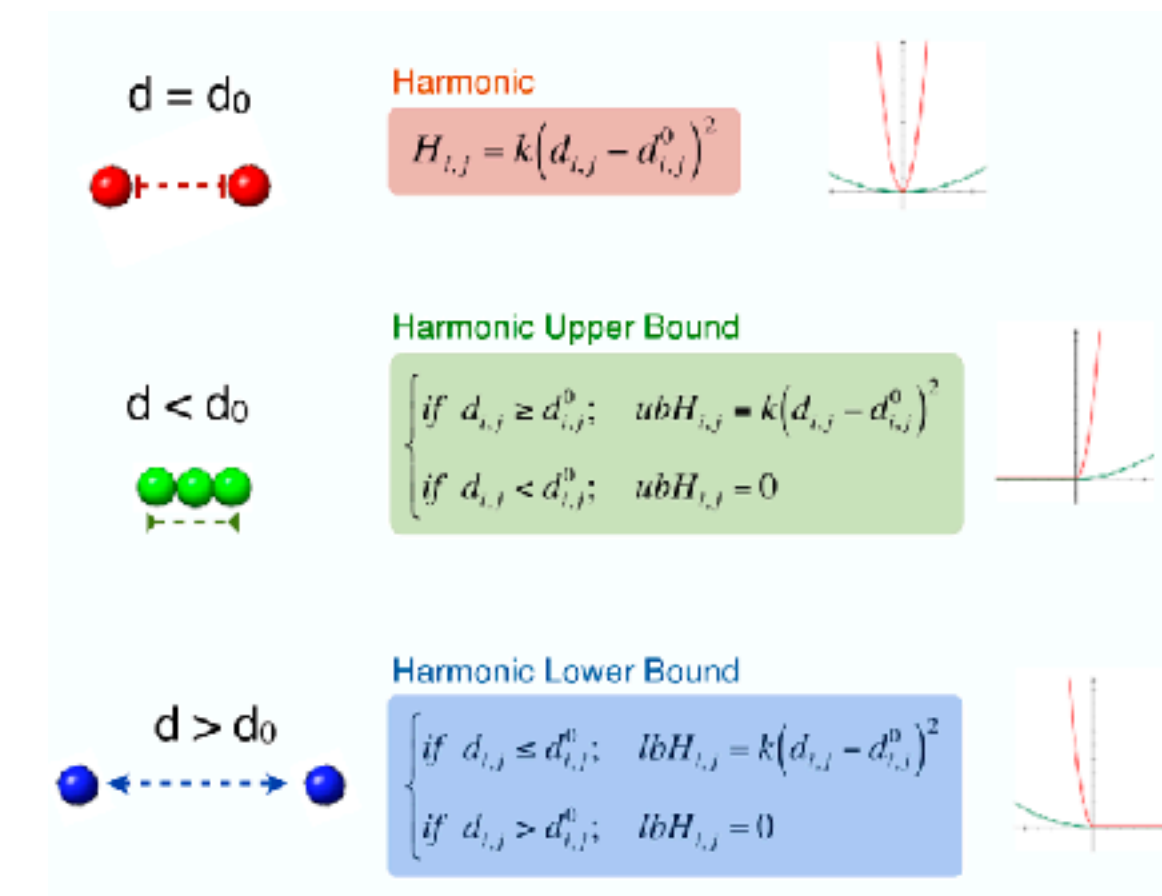
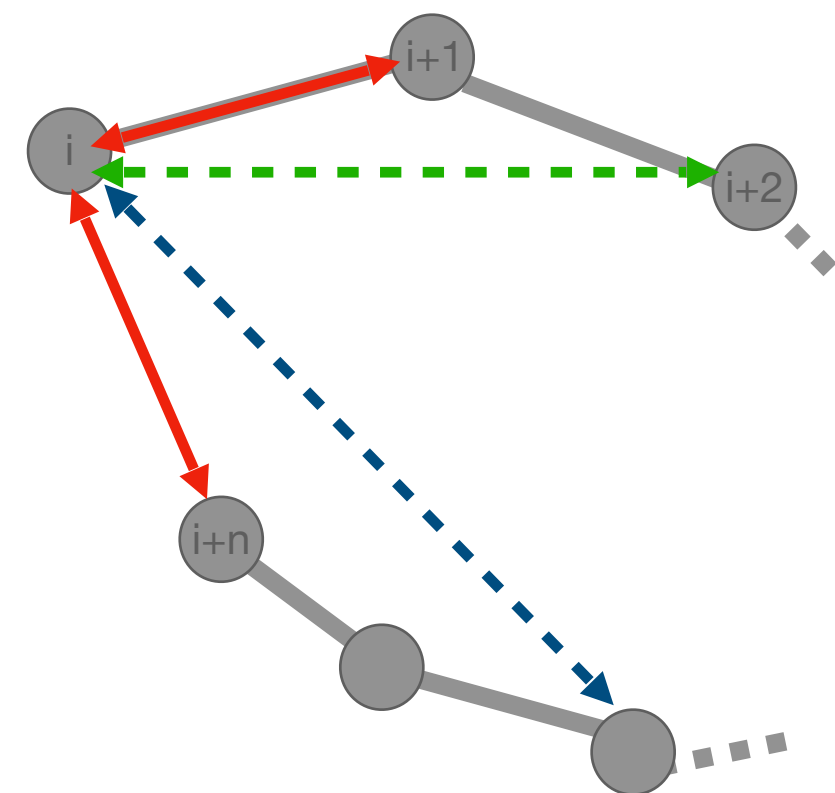
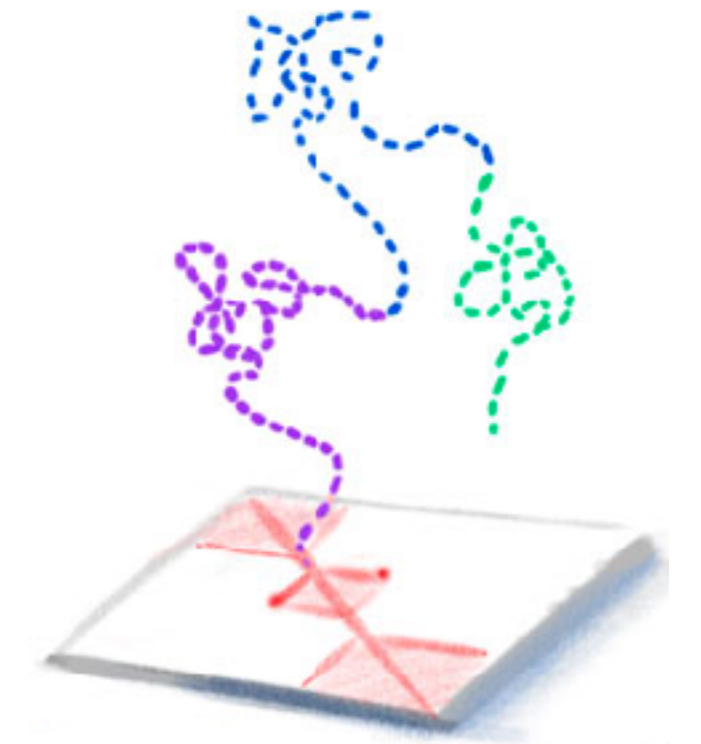
Representing the system
& translating information
into a scoring function

Structural sampling

Validation & analysis
of the model(s)



$$P(|\mathbf{x}_1 - \mathbf{x}_2| < d)$$



Di Stefano et al., Nat. Commun 2020

Serra et al., PLoS Comput Biol. 2017

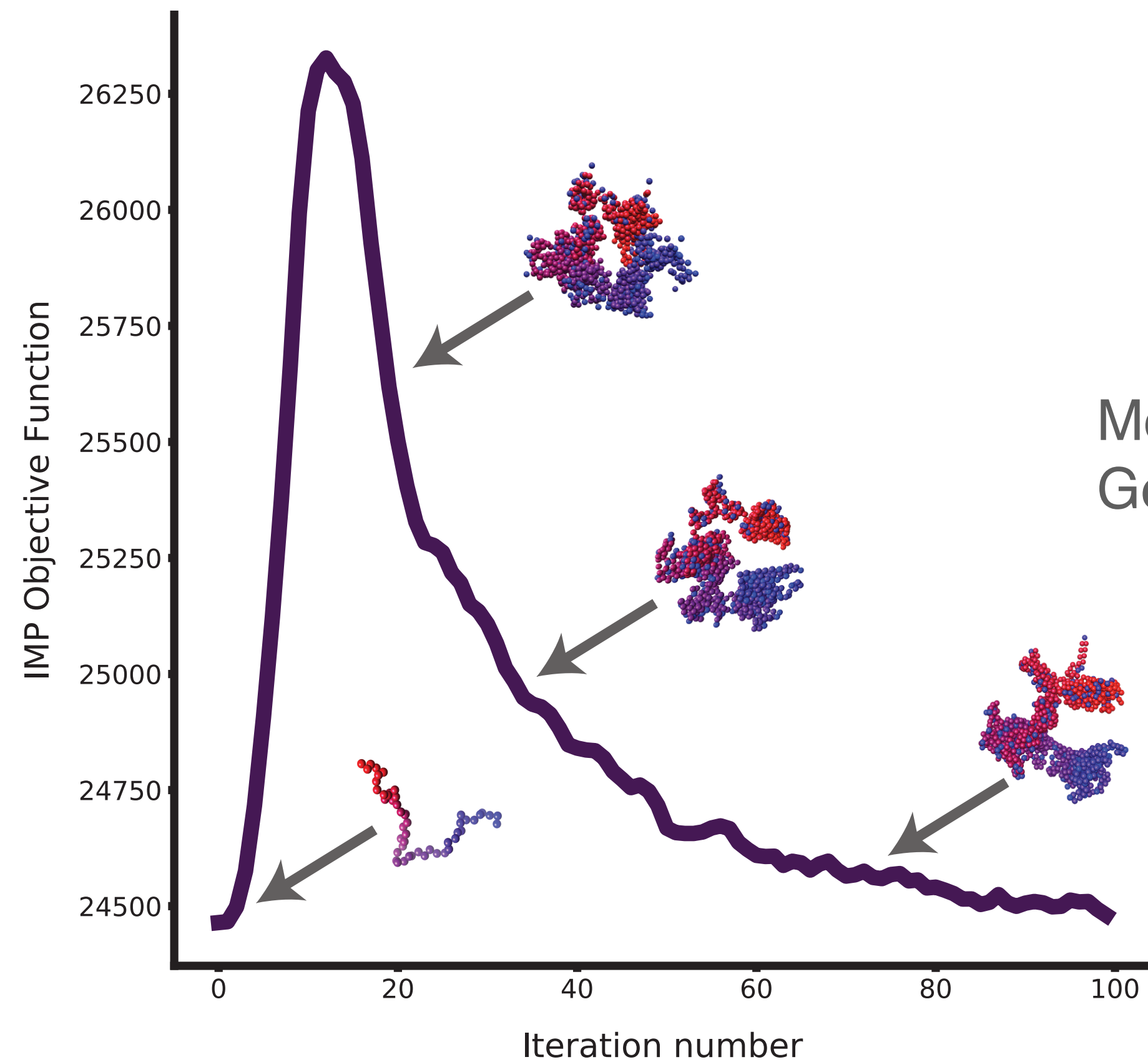
Integrative Modelling

Gathering information
(experimental data, physical
theories, statistical preferences)

Representing the system &
translating information into a
scoring function

Structural sampling

Validation & analysis
of the model(s)



Optimisation

Monte Carlo, Molecular Dynamics,
Genetic Algorithm, Random Search, ...



Di Stefano et al., Nat. Commun 2020

Serra et al., PLoS Comput Biol. 2017

Integrative Modelling

Gathering information
(experimental data, physical
theories, statistical preferences)



Representing the system &
translating information into a
scoring function

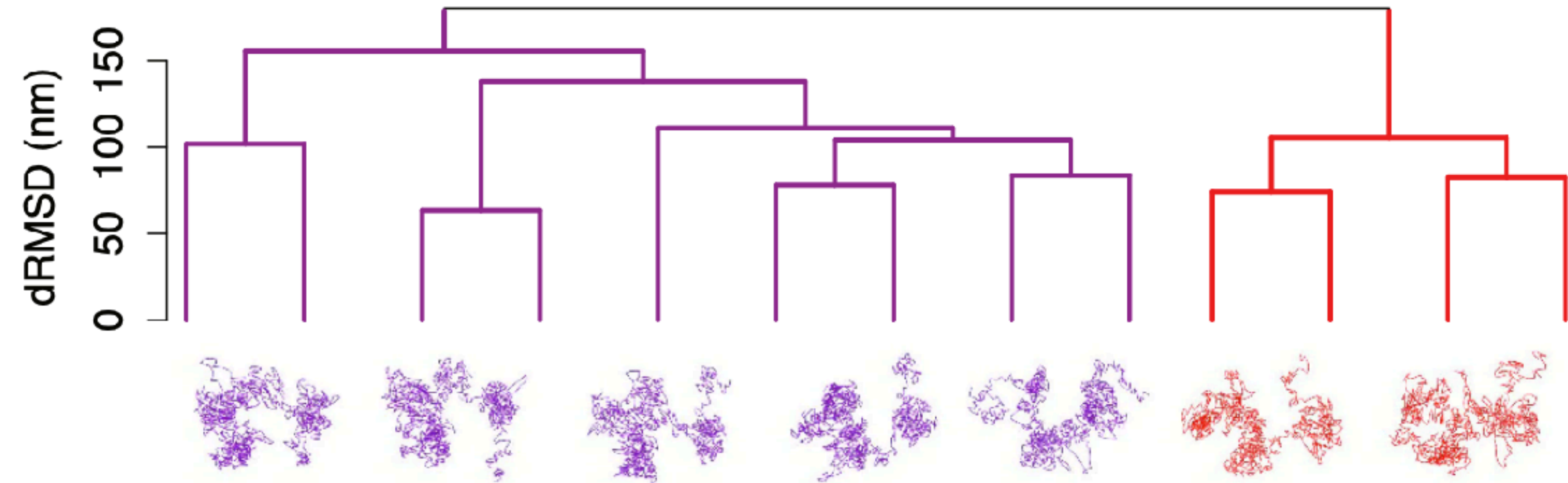


Structural sampling

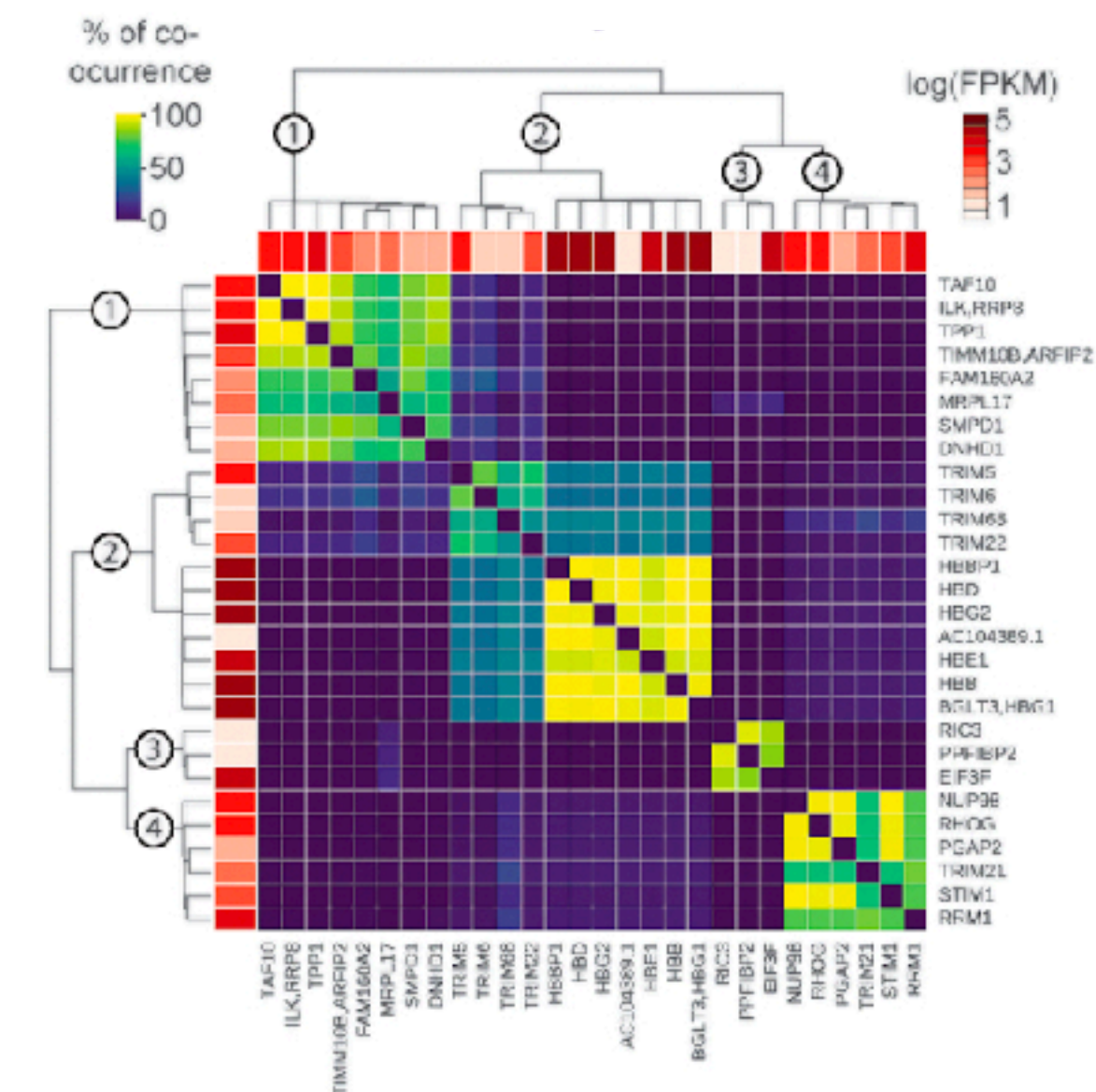


**Validation & analysis
of the model(s)**

3D population variation

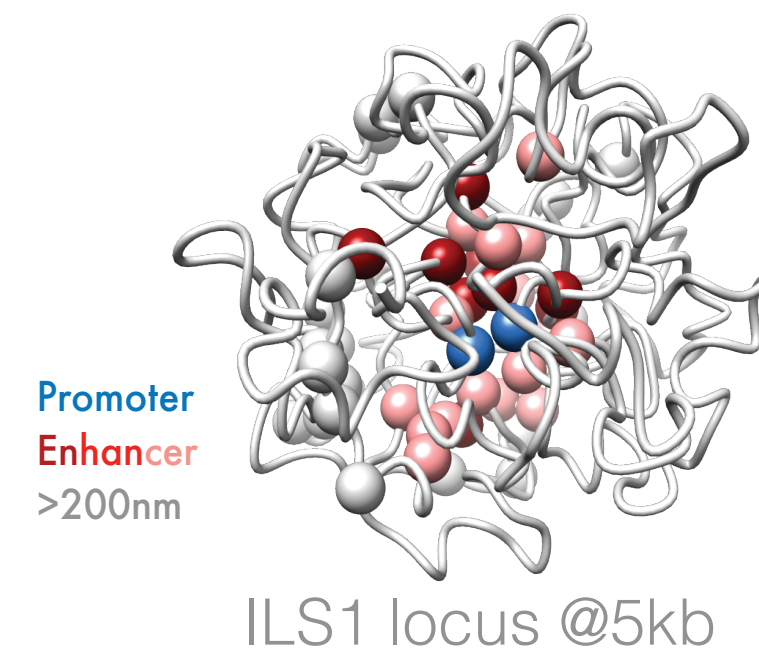


Deconvolute population cell 3D variation
Co-occurrence analysis



Integrative Modelling unveiling genome organisation

3D islet-enhancer hub and type 2 diabetes variants



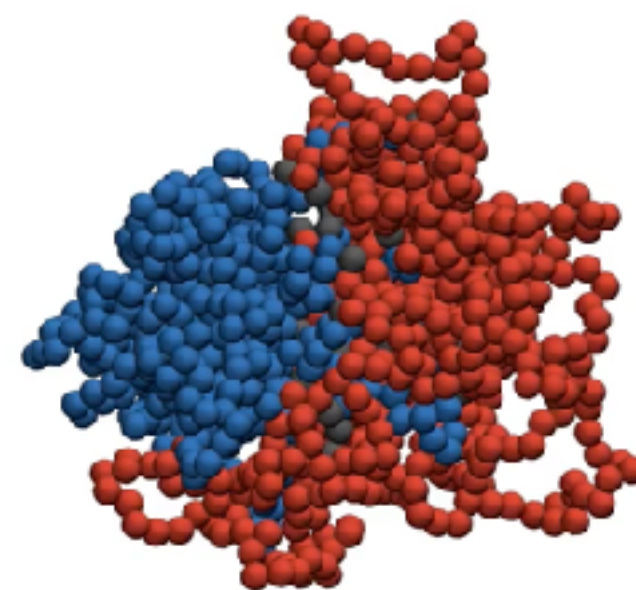
Miguel-Escalada et al., Nat. Gen. 2019
Mendieta-Esteban, ..., **Farabella**[^] et al., NAR Genom Bioinform. 2021

4D in reprogramming



Di Stefano et al., Nat. Commun 2020

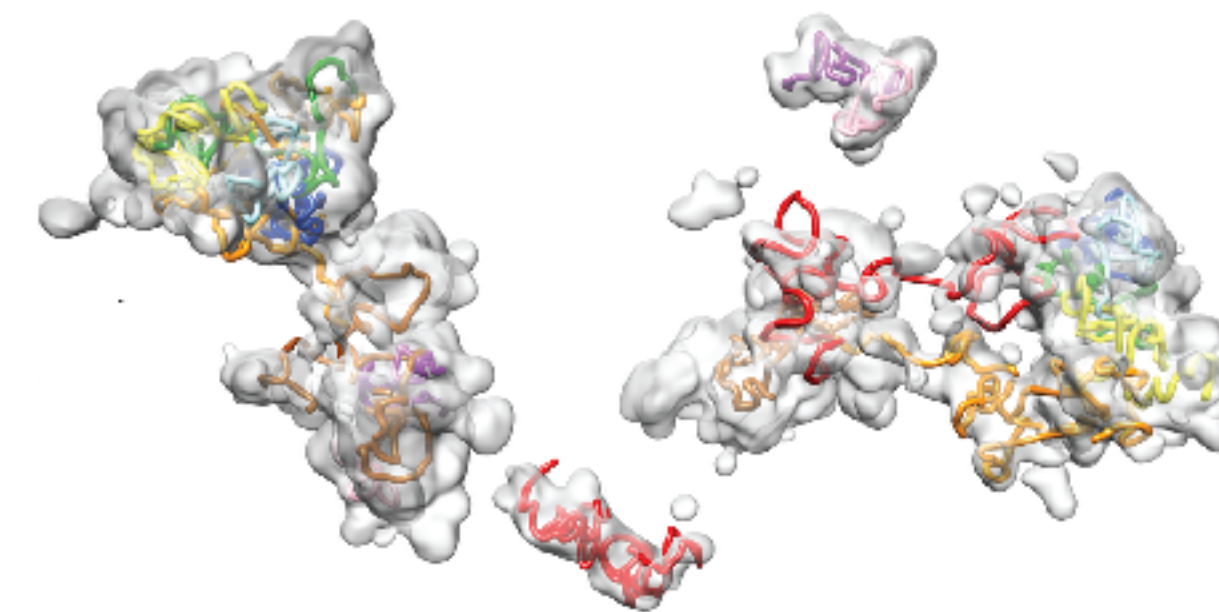
Role of RNA in shaping the 3D genome



Whole chromosomes @50 kb and @100kb

Farabella et al. accepted Nat. Struct. Mol. Biol. 2021
Morf et al., Nat. Biotech. 2019

Homolog-specific integrative model of diploid human cells



PGP1f chr19:7,335,095-15,449,189 (~8Mb) @10kb

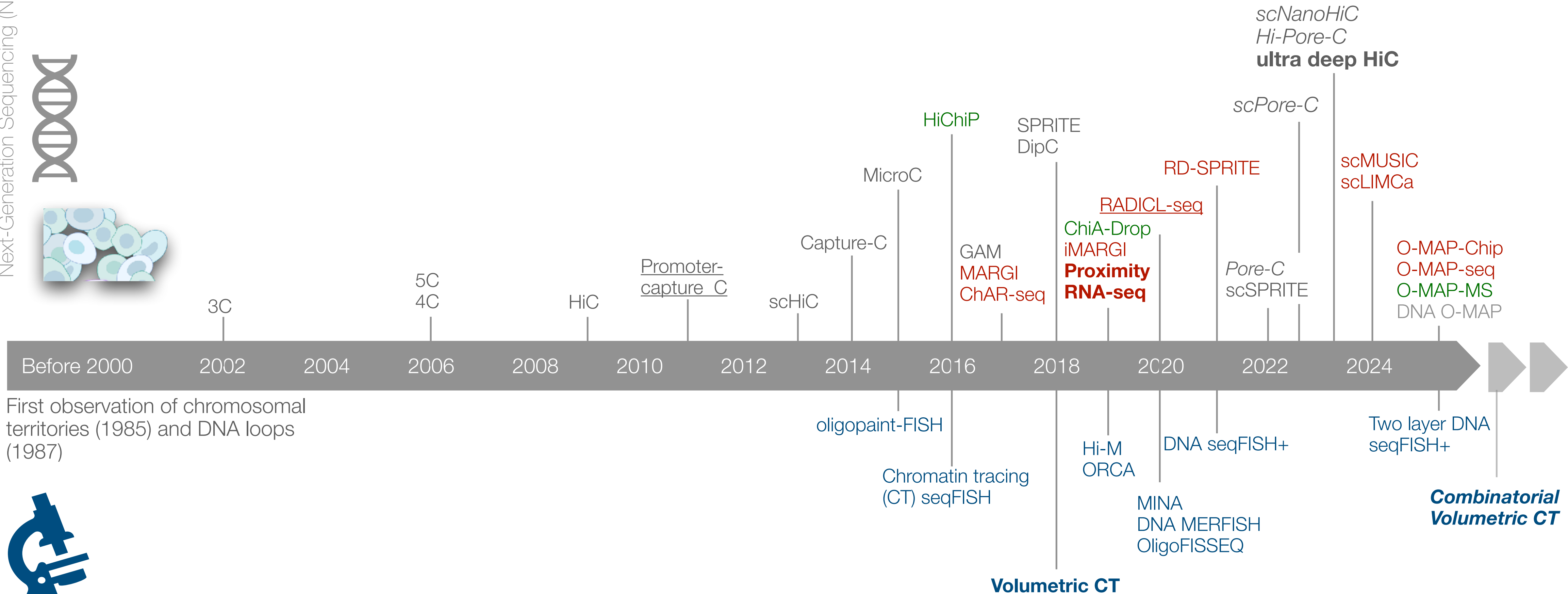
Nir*, **Farabella***, Pérez Estrada*, Ebeling*, et al. PlosGen. 2018
Farabella et al., in preparation

Can We Directly Resolve 3D Genome Structure at Single-Chromosome and Single-Cell Resolution?

Technologies for 3D genome investigation

Sequence-based technologies

Next-Generation Sequencing (NGS)



Imaging-based technologies



Multimodal potential RNA-Protein-genome
RNA-genome or RNA-RNA interactions
Protein-genome interactions

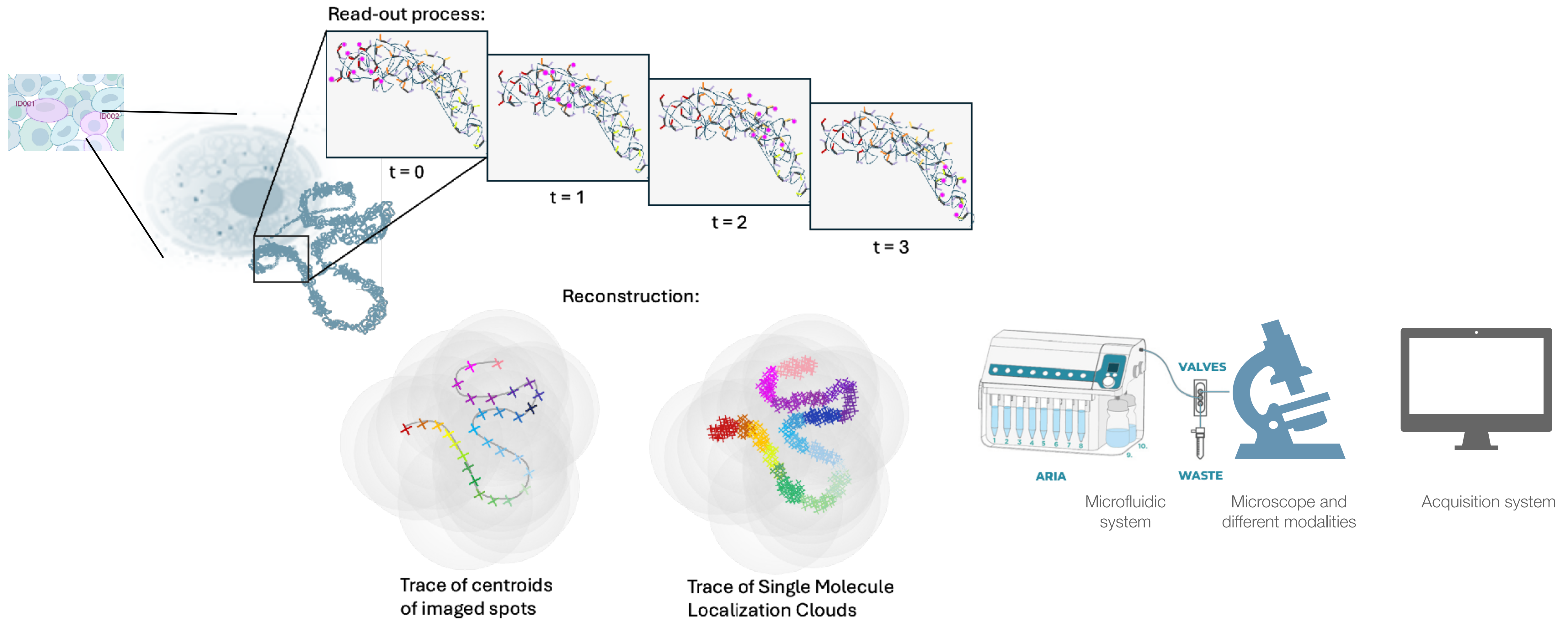
Volumetric CT
Proximity RNA-seq
Ultra deep HiC

Combinatorial Volumetric CT
CT assessment and analysis tools

Nir, Farabella et al, PlosGen 2018
Morf, Wingett, Farabella et al. Nat Biotech. 2019
Harris, Gu, Olshansky, ... Farabella... et al. et al., Nat Commun 2023
Lee, Piacere, ... Farabella, Wu in preparation
Piacere... Farabella CIMA in preparation

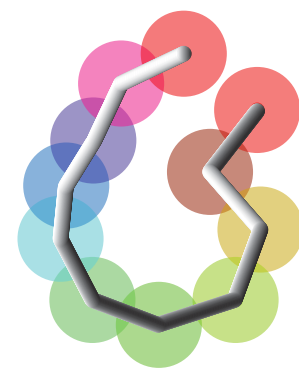
Imaging-based technologies

Oligo-based Chromatin Tracing

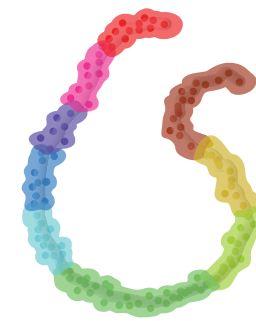


Visualising the genome

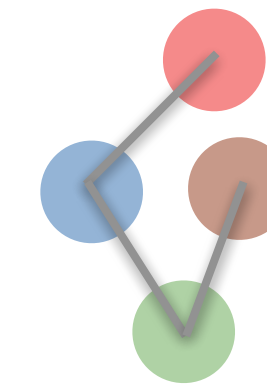
3D Genome organisation in individual cell with oligo-based chromatin tracing



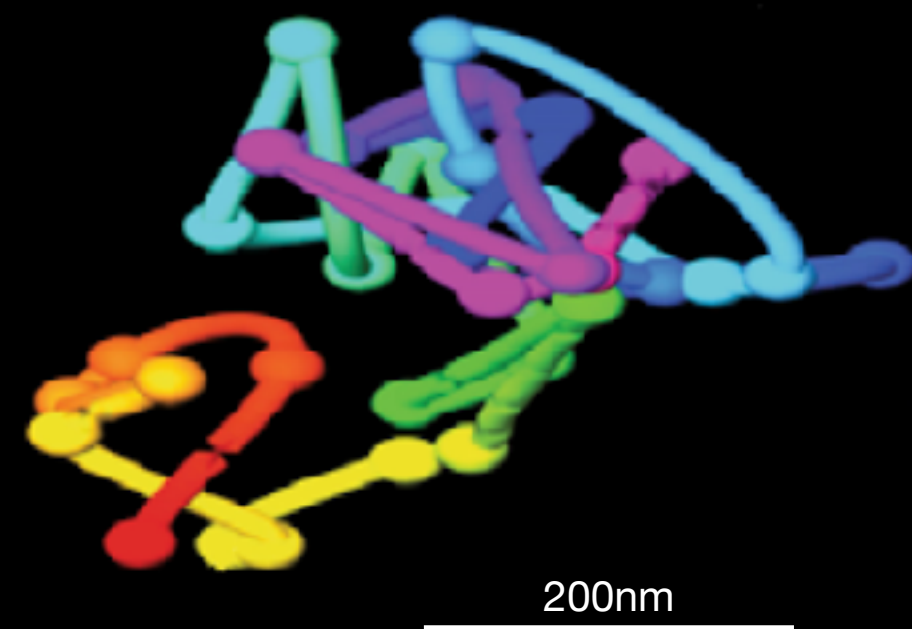
Ball and Stick
Continuous labelling



Volumetric
Continuous and/or discontinuous labelling

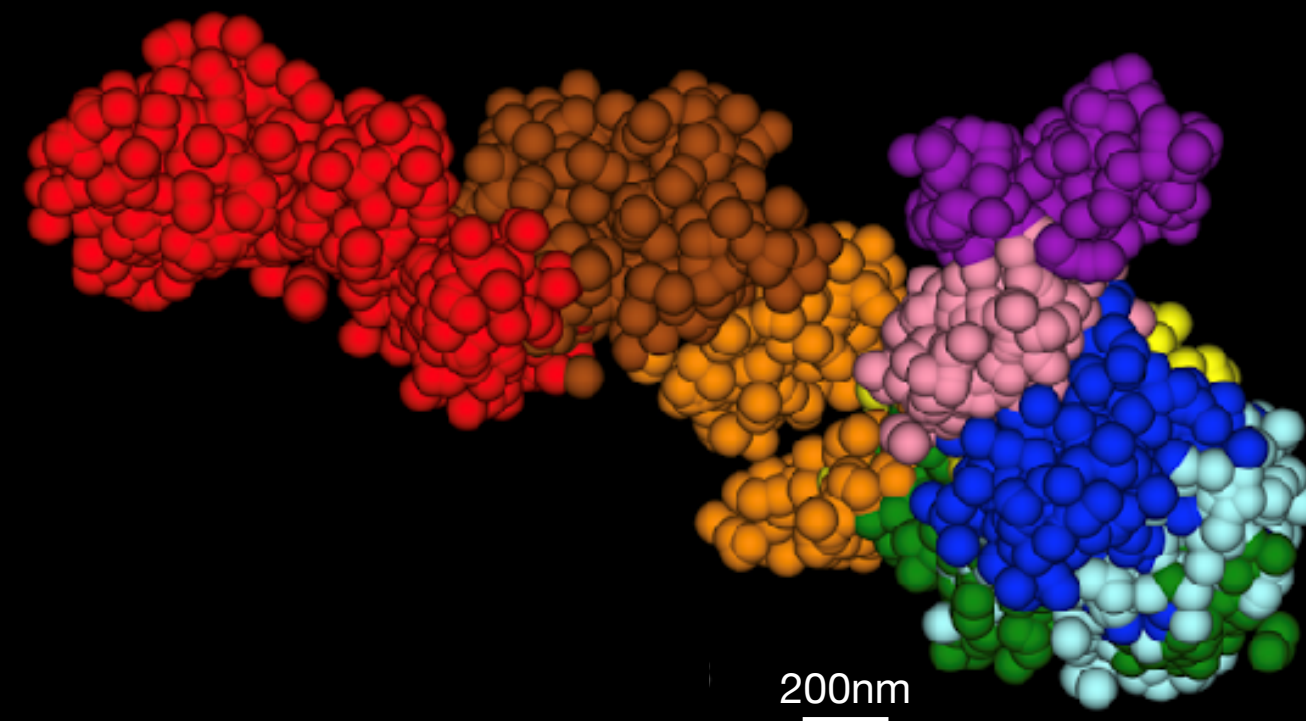


Ball and Stick
Discontinuous labelling



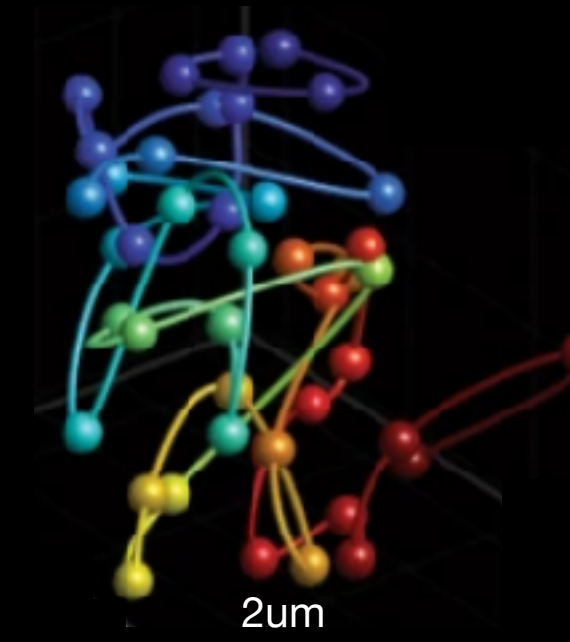
Mateo et al., Nature 2019

Kilobases



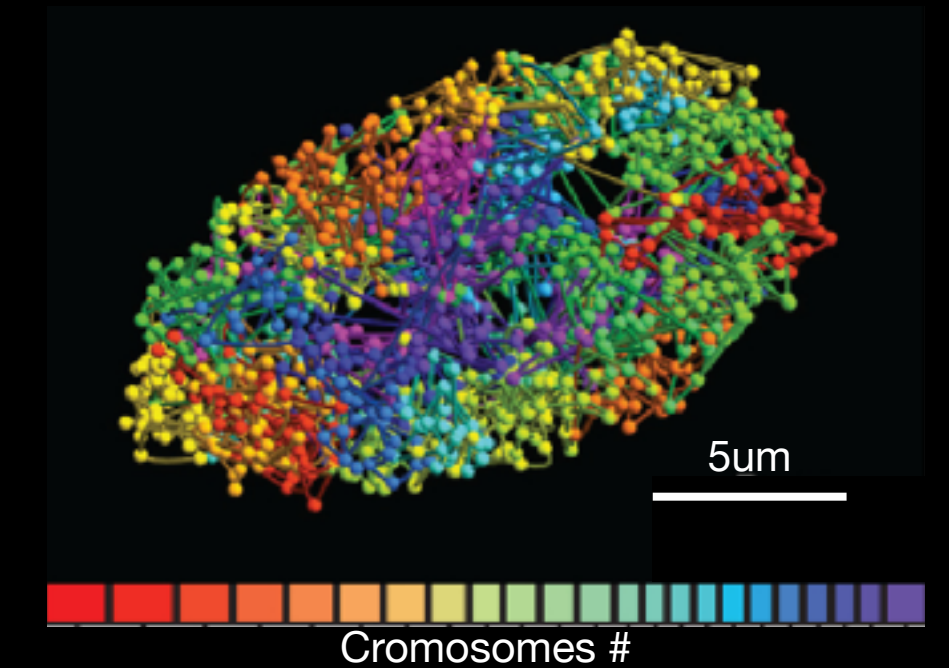
Nir*, **Farabella***, Pérez Estrada*,
Ebeling*, et al. PlosGen 2018

Megababses



Liu et al., Nat Commun. 2020

Chromosomes-wide



Su*, Zheng*, Kinrot*, Bintu*
et al., Cell 2020

Genome-wide

Volumetric Chromatin Tracing

seq-OligoSTORM



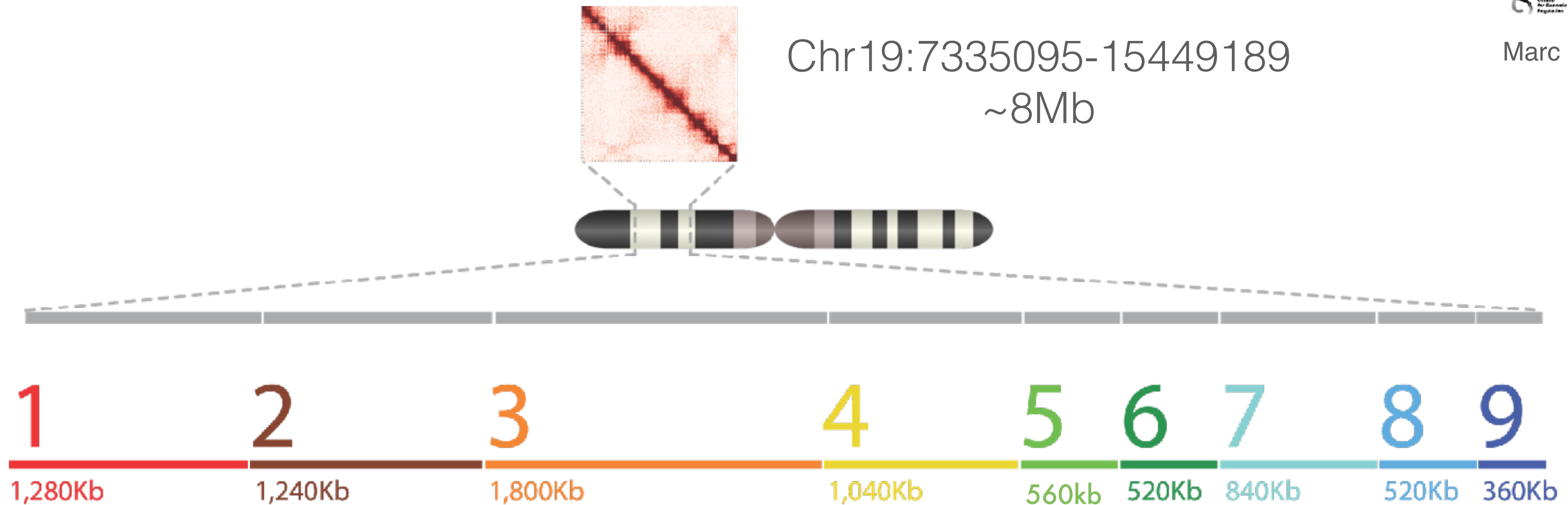
Ting Wu



Guy Nir



Marc Marti Renom



Volumetric Chromatin Tracing

seq-OligoSTORM



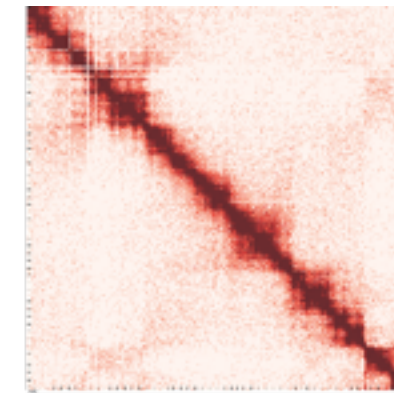
Ting Wu



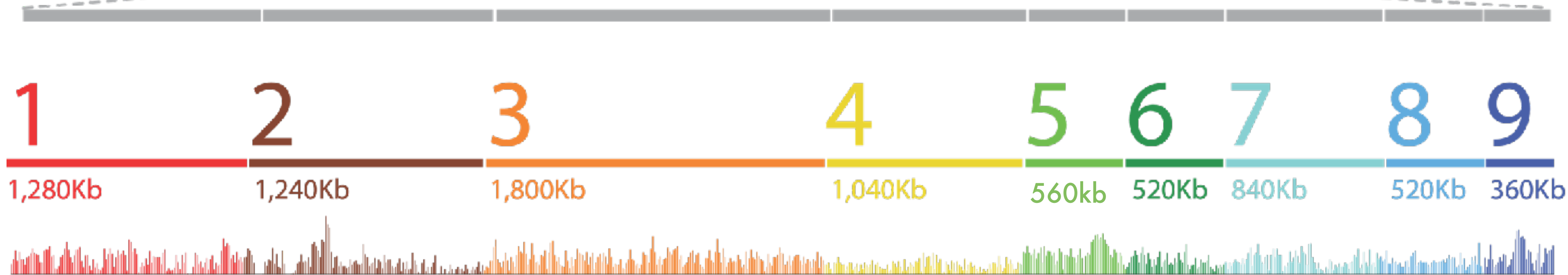
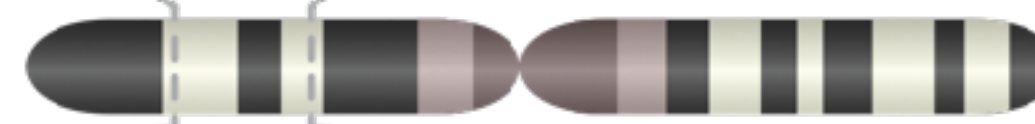
Guy Nir



Marc Marti Renom



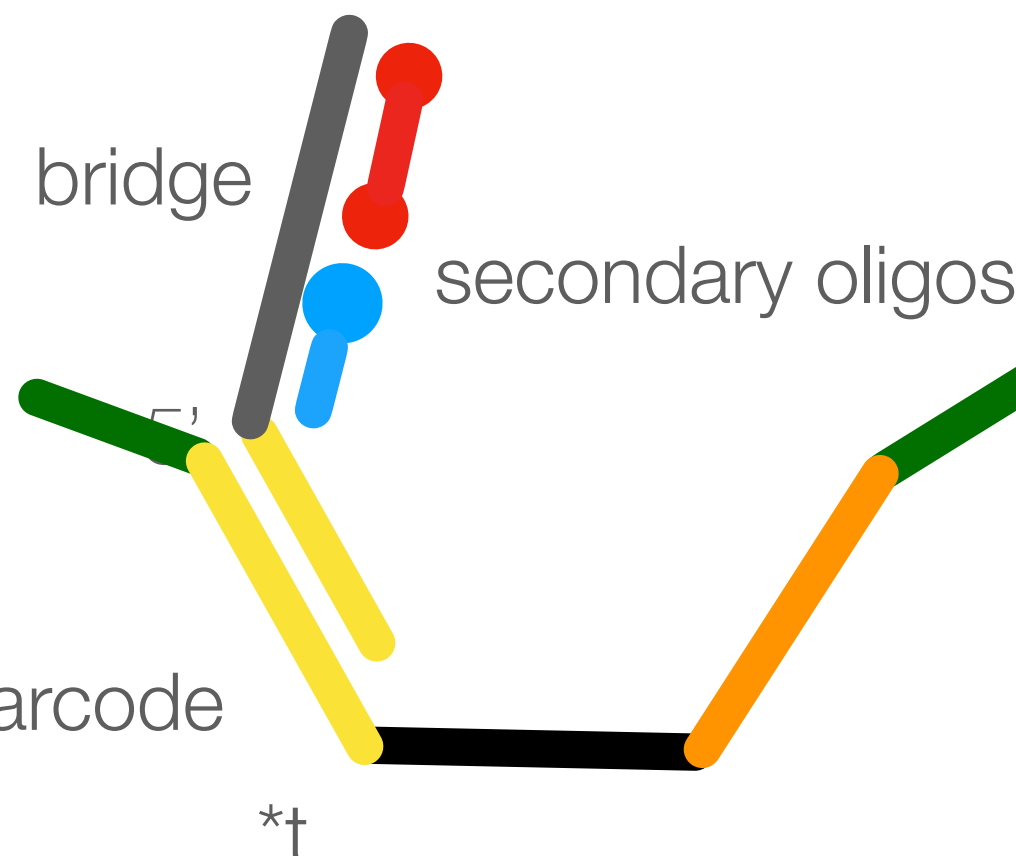
Chr19:7335095-15449189
~8Mb



Oligopaints
(multiplexed FISH probes)

mainstreet
universal primer

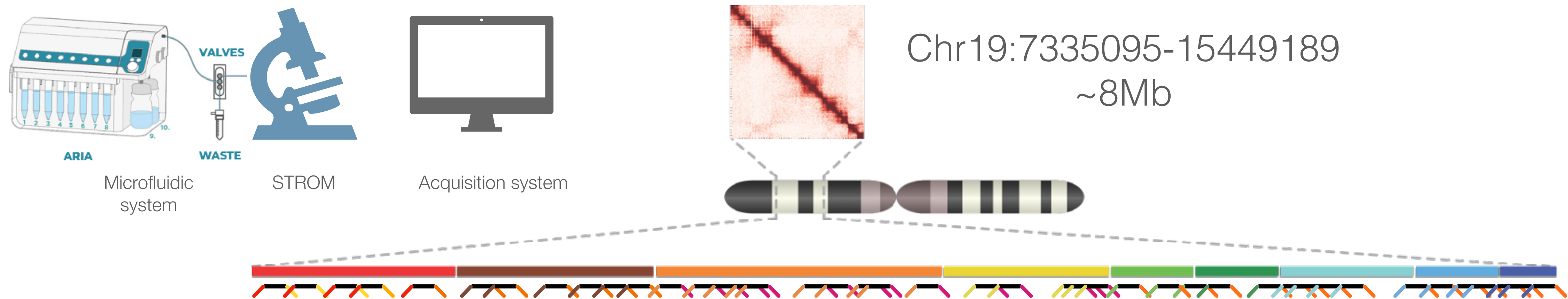
region specific barcode



backstreet
3'

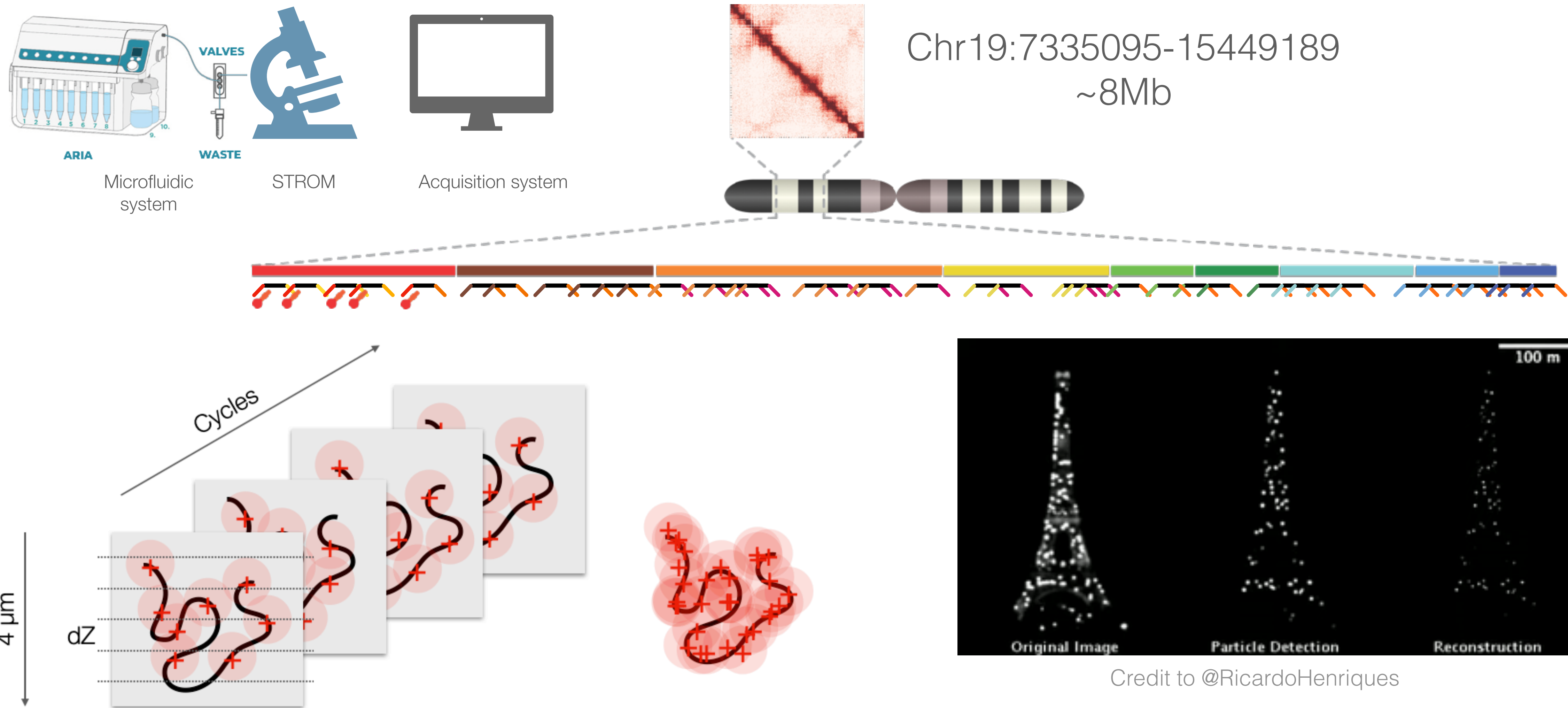
Volumetric Chromatin Tracing

seq-OligoSTORM



Volumetric Chromatin Tracing

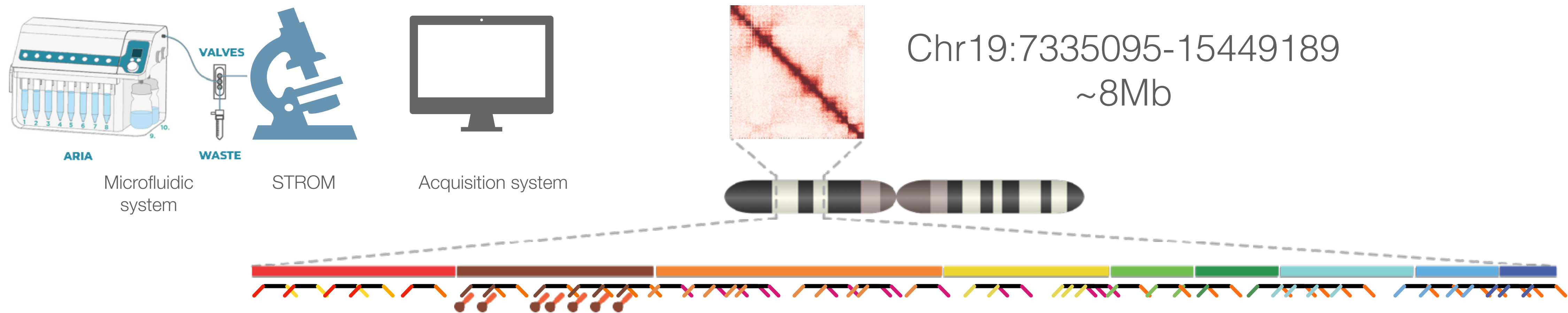
seq-OligoSTORM



Credit to @RicardoHenriques

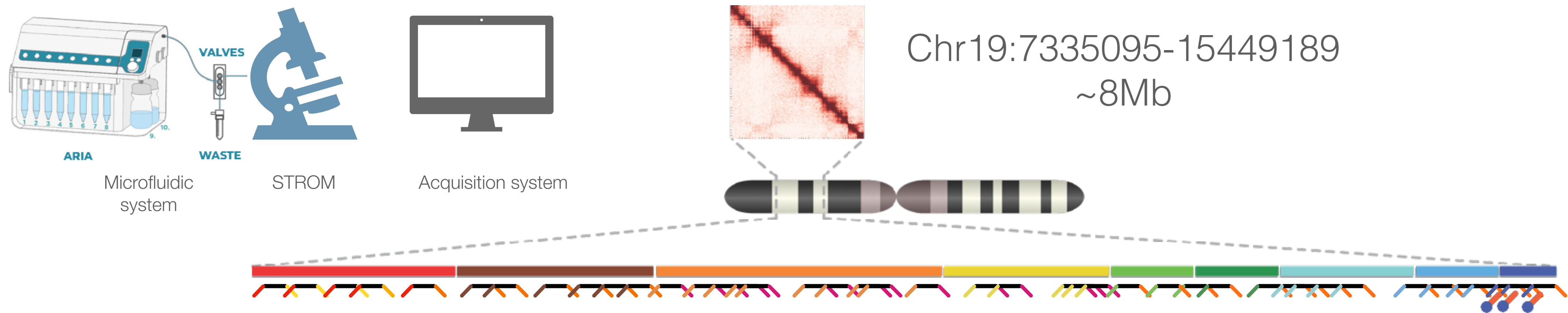
Volumetric Chromatin Tracing

seq-OligoSTORM



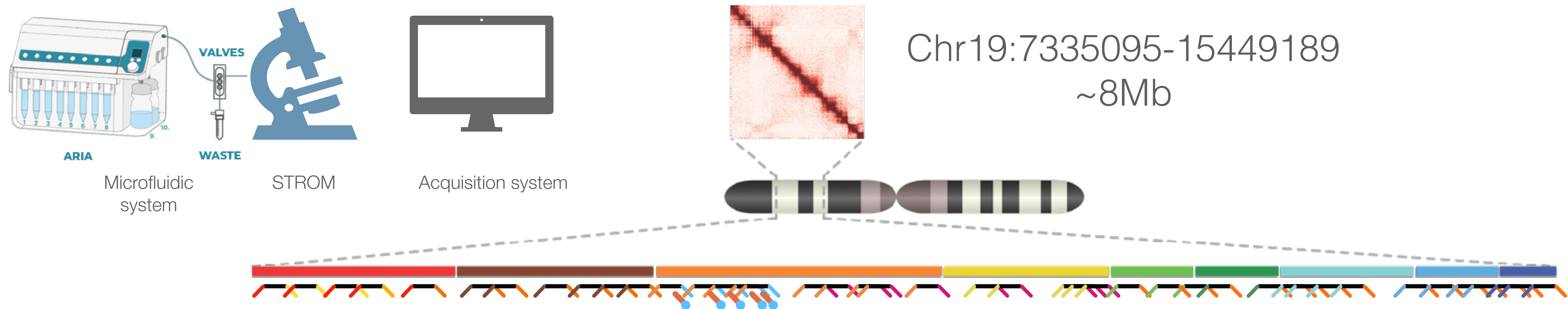
Volumetric Chromatin Tracing

seq-OligoSTORM



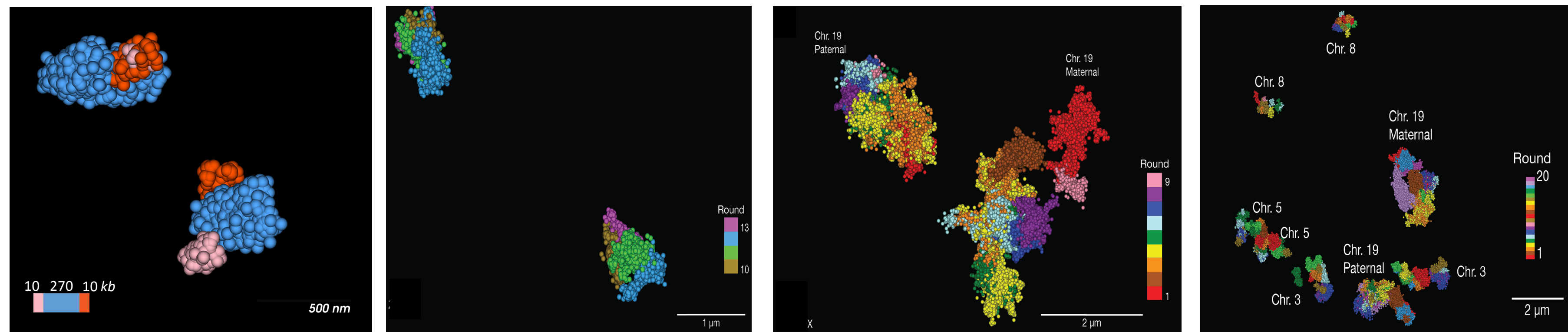
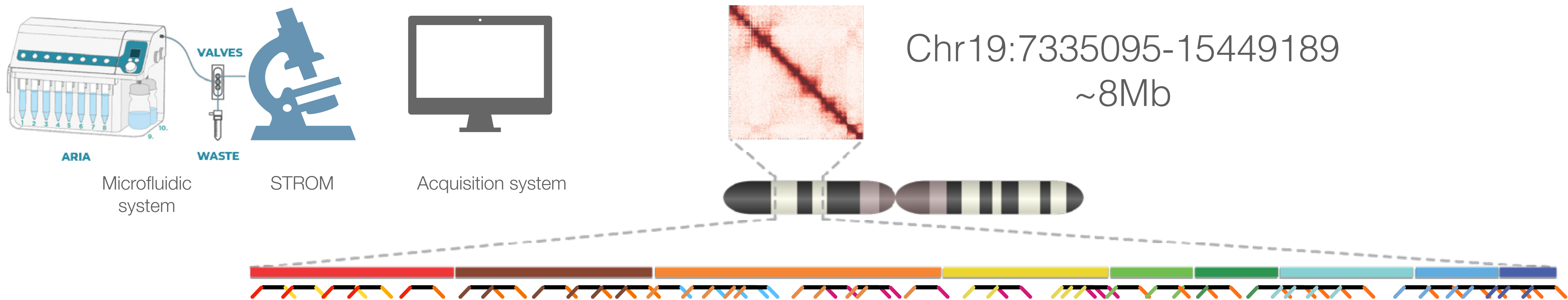
Volumetric Chromatin Tracing

seq-OligoSTORM



Volumetric Chromatin Tracing

seq-OligoSTORM



Chromatin
Loops

Chromatin
domains

Chromosomal
compartments

Multiple
chromosomes

DNA length

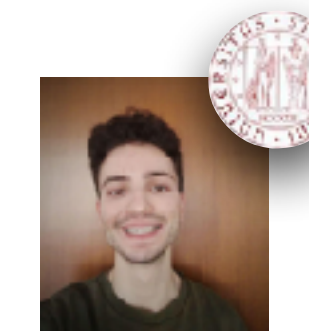


Chromatin IMaging Analysis tool

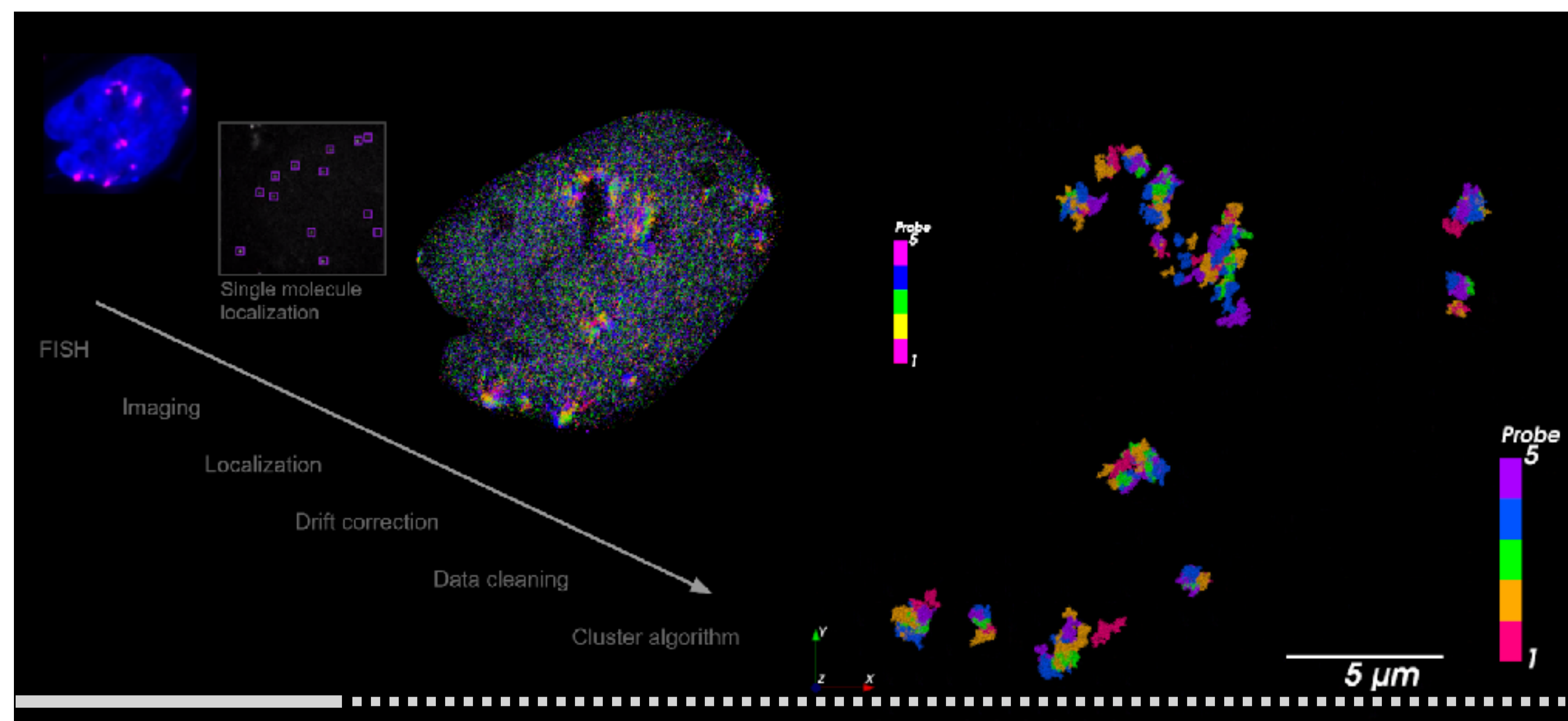


PGP1f Chr21 walk of 7.5Mb and Chr1 walk of 7.5Mb

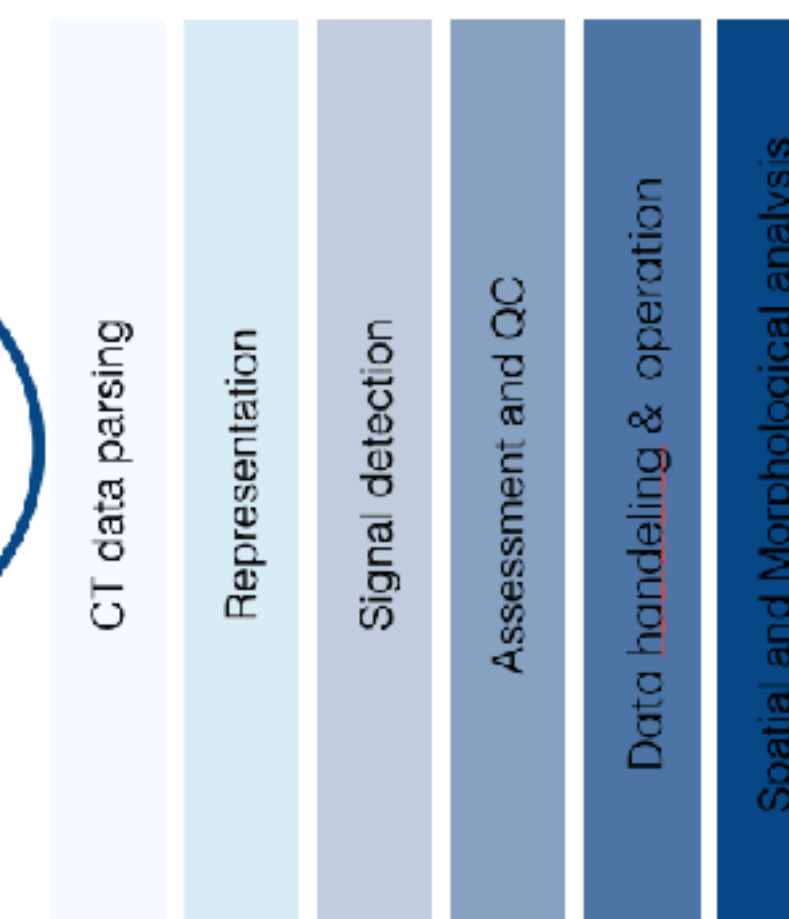
Sarah Aufmkolk and Antonios Lioutas



Farabella Lab: Mattia Roccatello, **Ivan Piacere**,
& Antony Palmieri.



SMLM processing



<https://gitlab.iit.it/ina/CIMA>

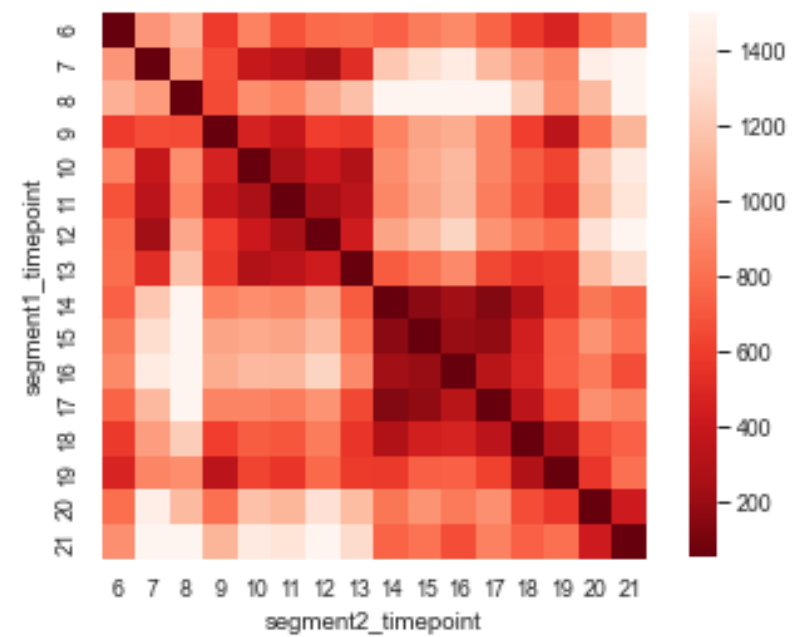


Nir*, **Farabella***, Pérez Estrada*, Ebeling*, et al. PlosGen. 2018
Piacere...Farabella CIMA in preparation

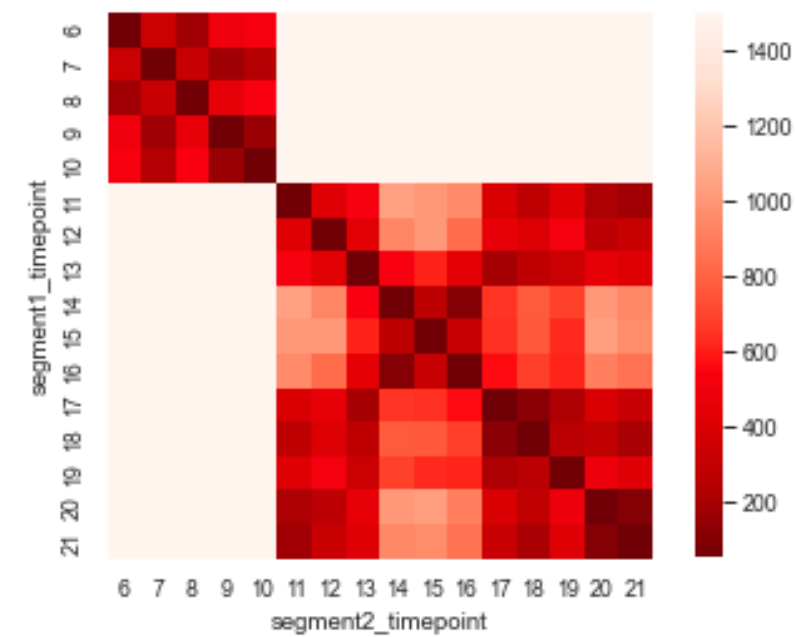
3D genome organisation variability at megabase scale

inter-nuclei structural variability

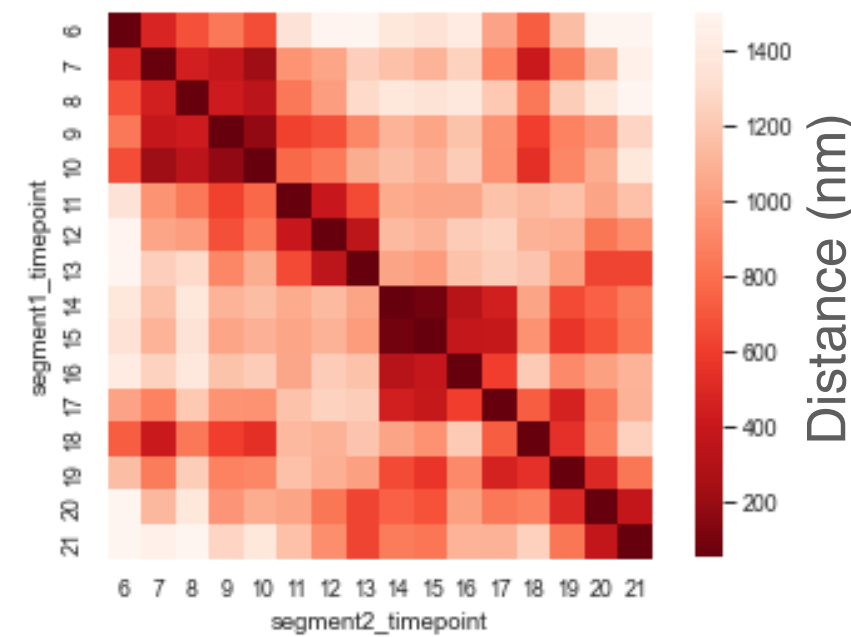
Trace #1



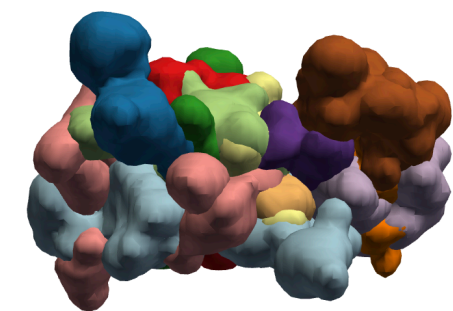
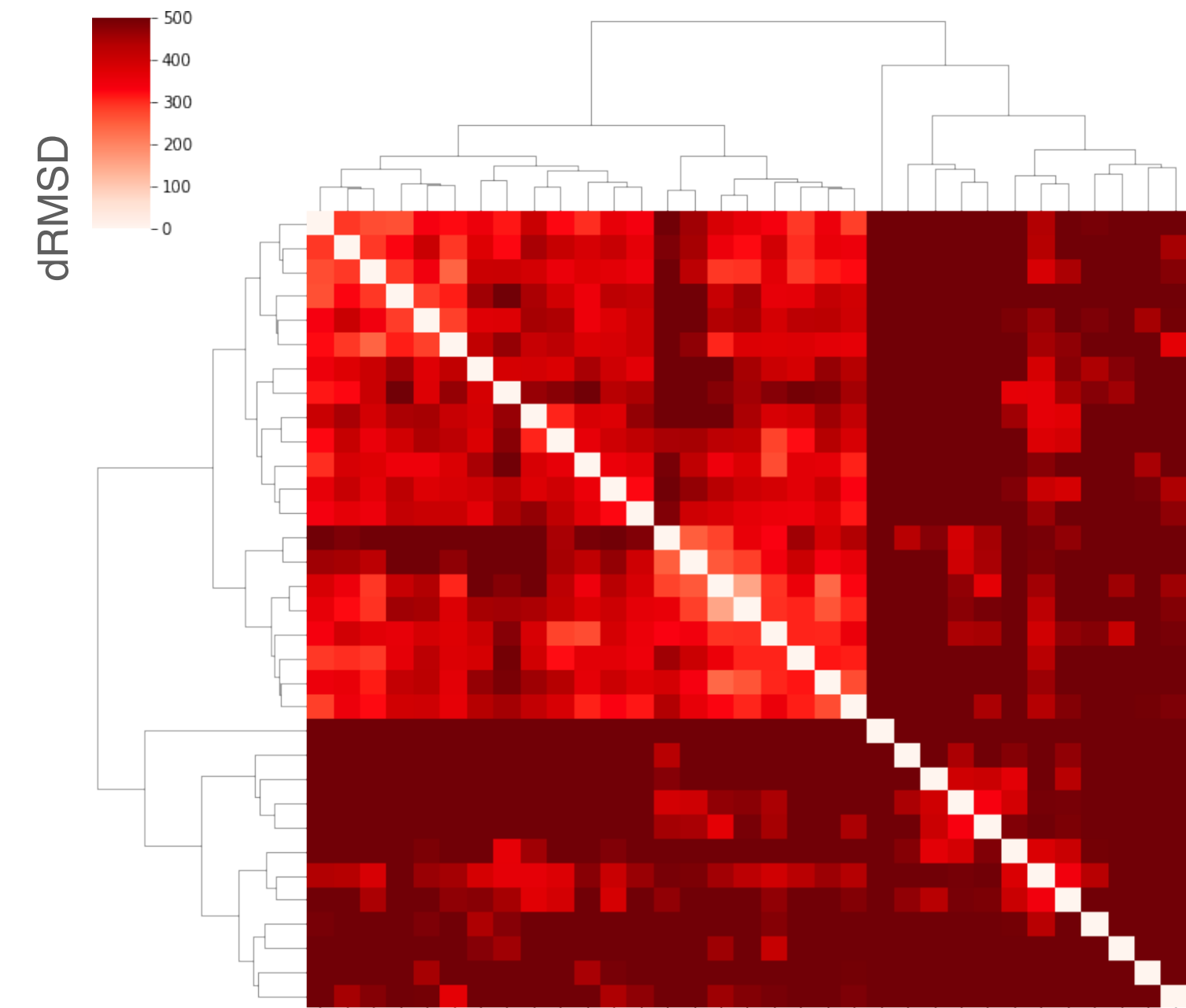
Trace #2



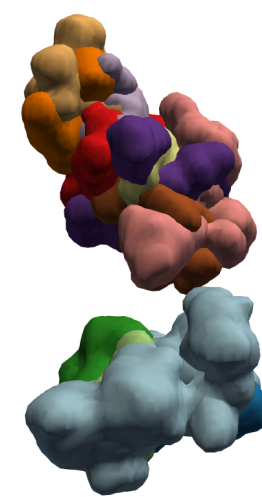
Trace #3



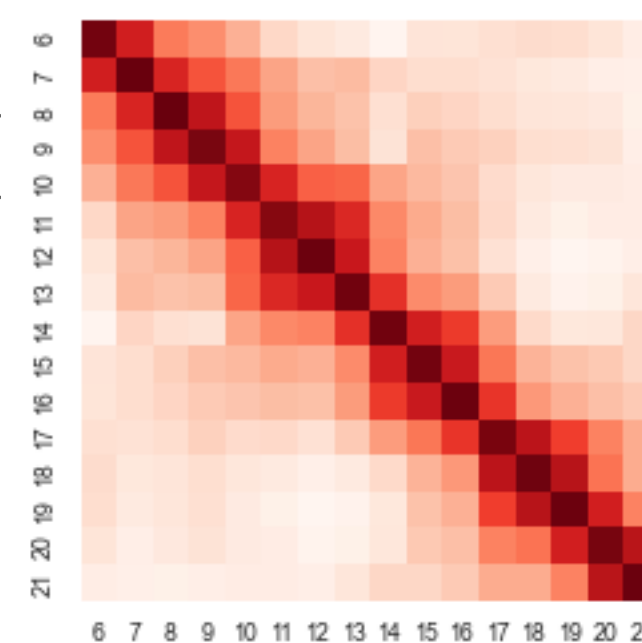
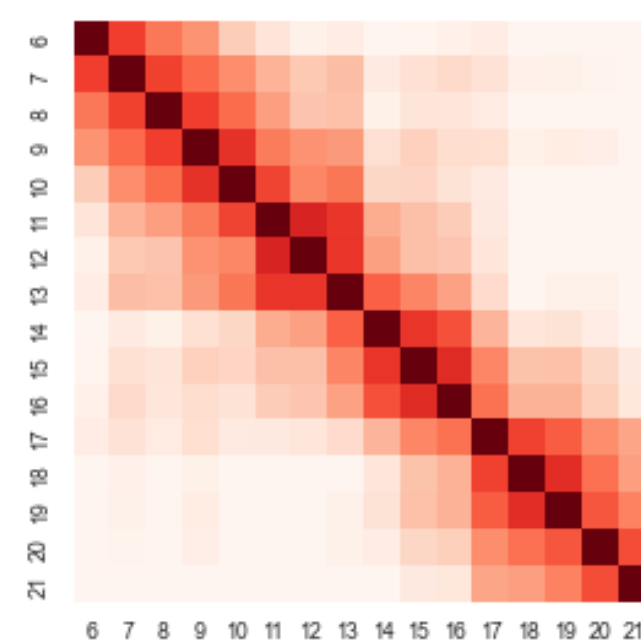
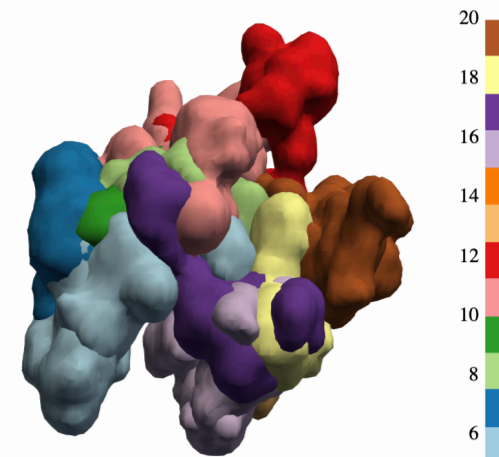
3D population variation



Mean Distance



HiC

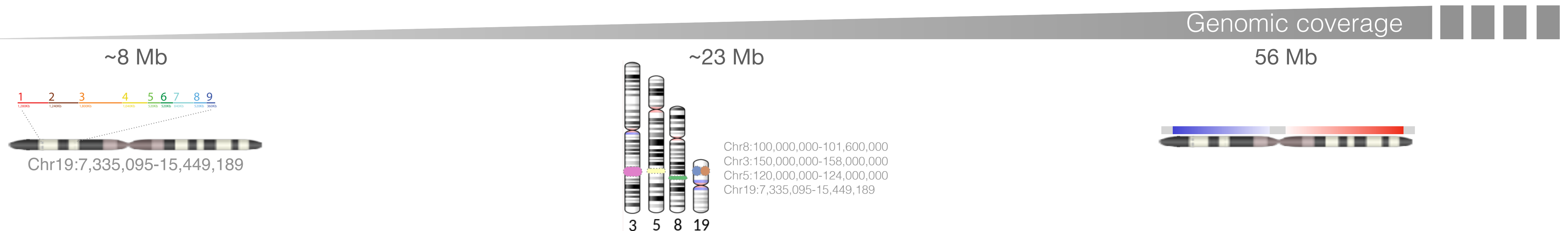
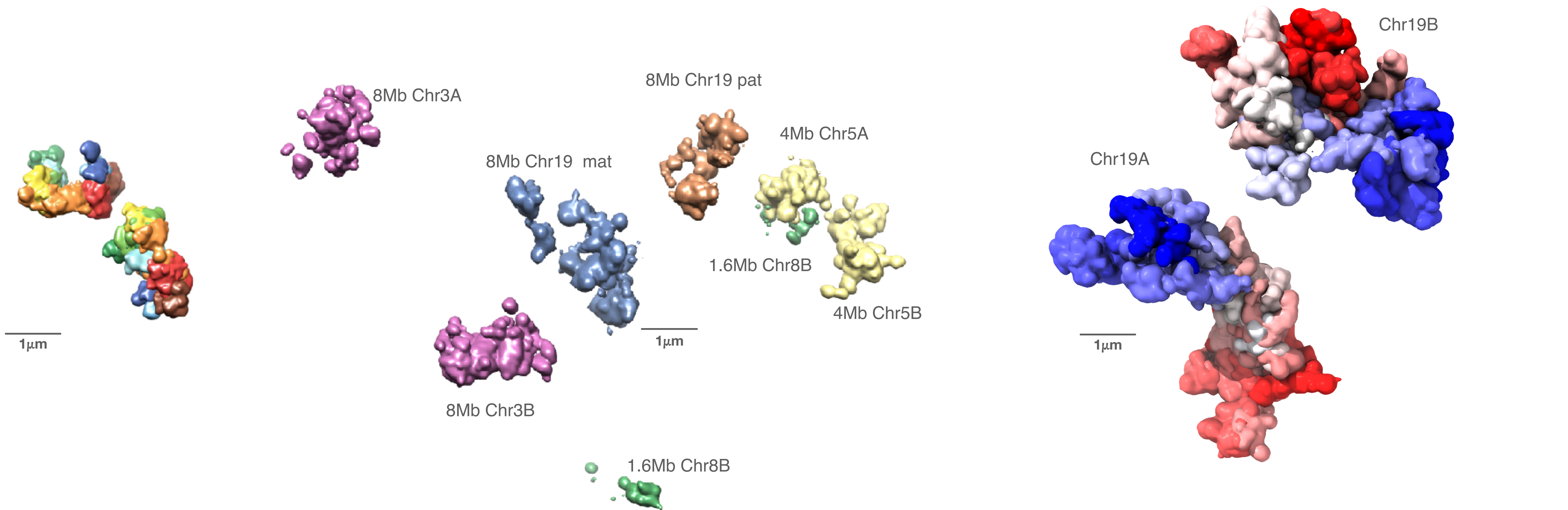


PGP1f chr3:150Mb-158Mb
~8Mb

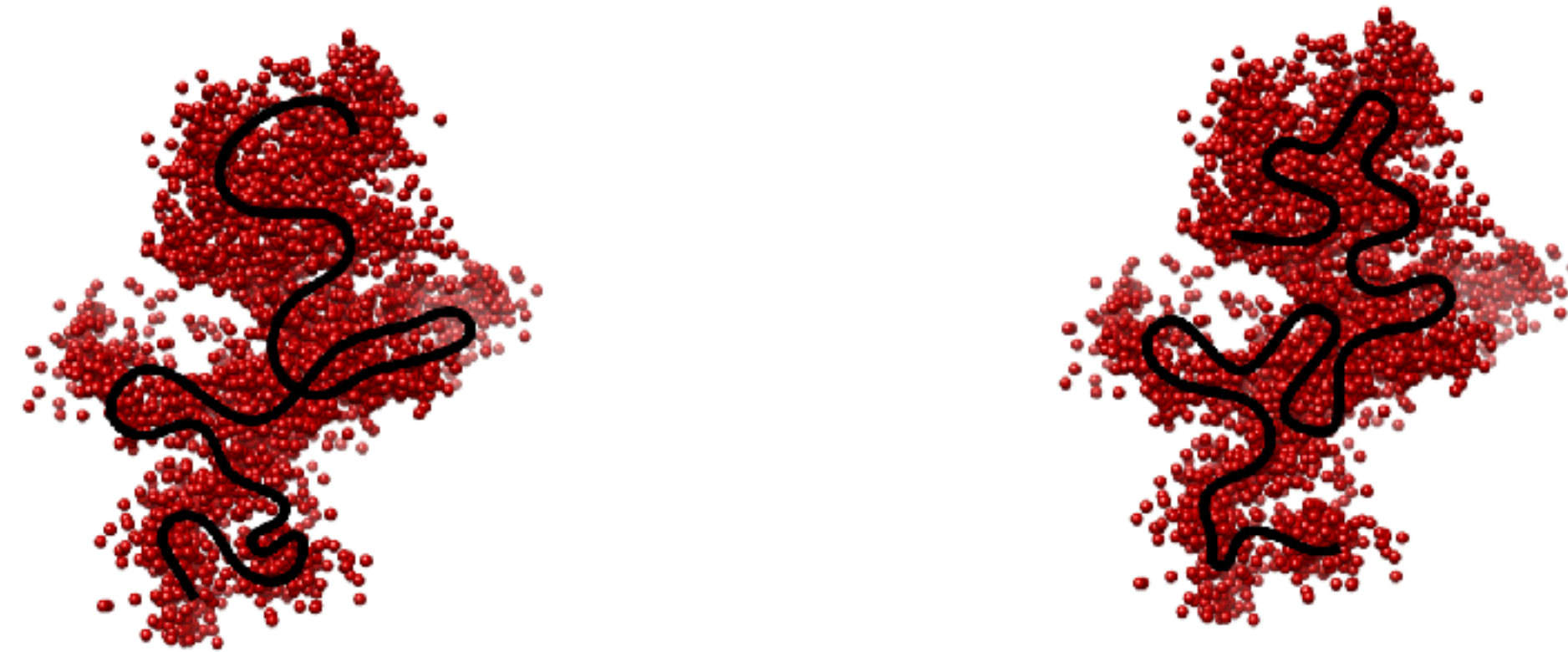
How Can We Scale Up Chromatin Imaging from loci towards
Genome-Wide Coverage?

Towards imaging the whole genome at nanoscale

With seq-OligoSTORM and hierarchical barcoding



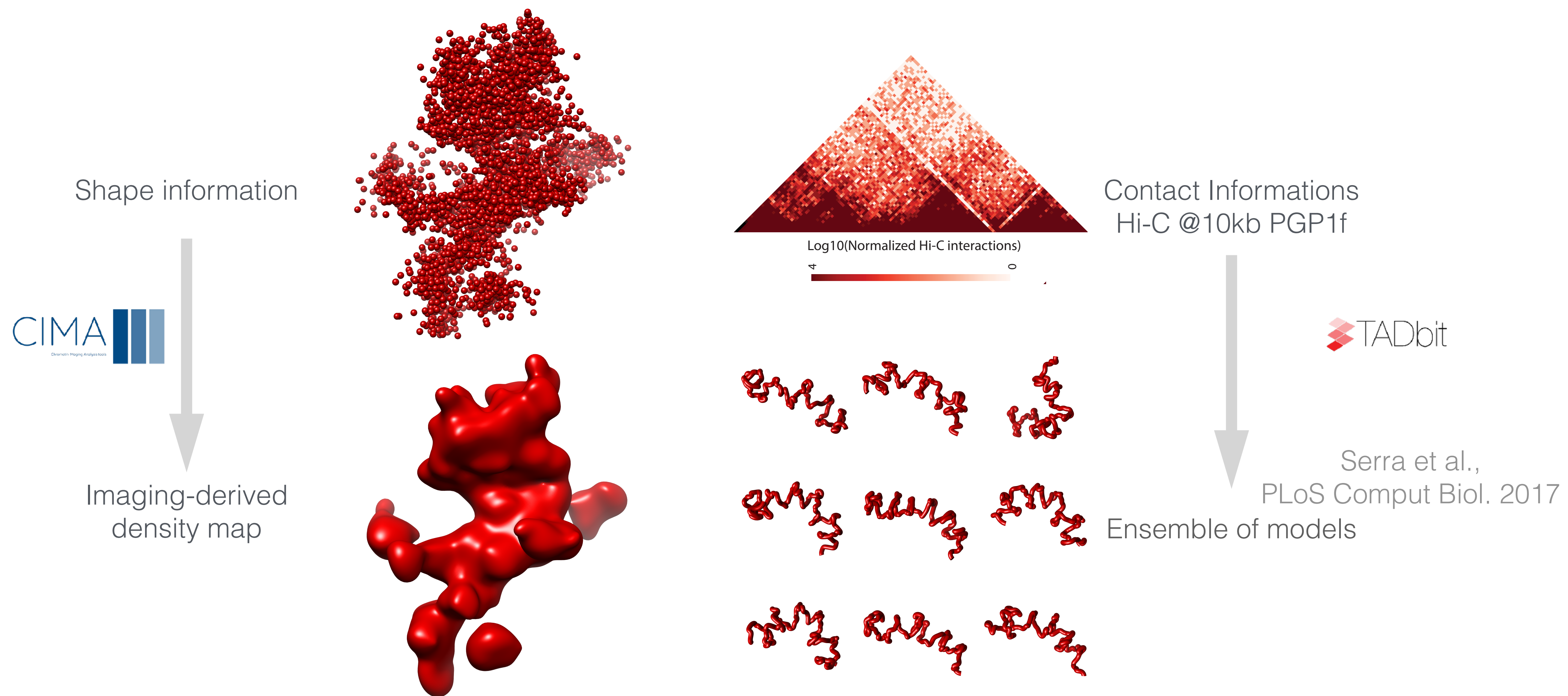
Can we increase the genomic resolution of our data
without loosing coverage?



.... 1Mb —————> 10Kb —————> 1kb

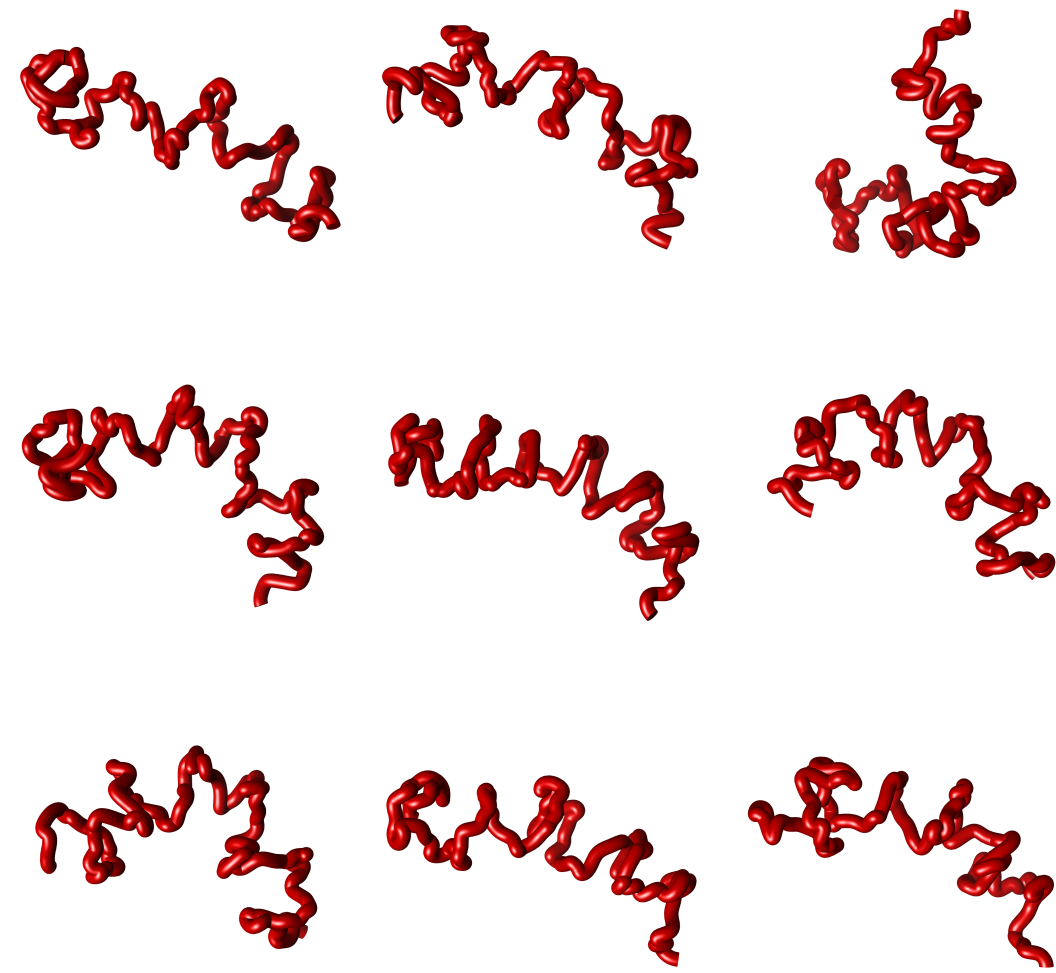
IMGR

Integrative Modelling of Genomic Regions



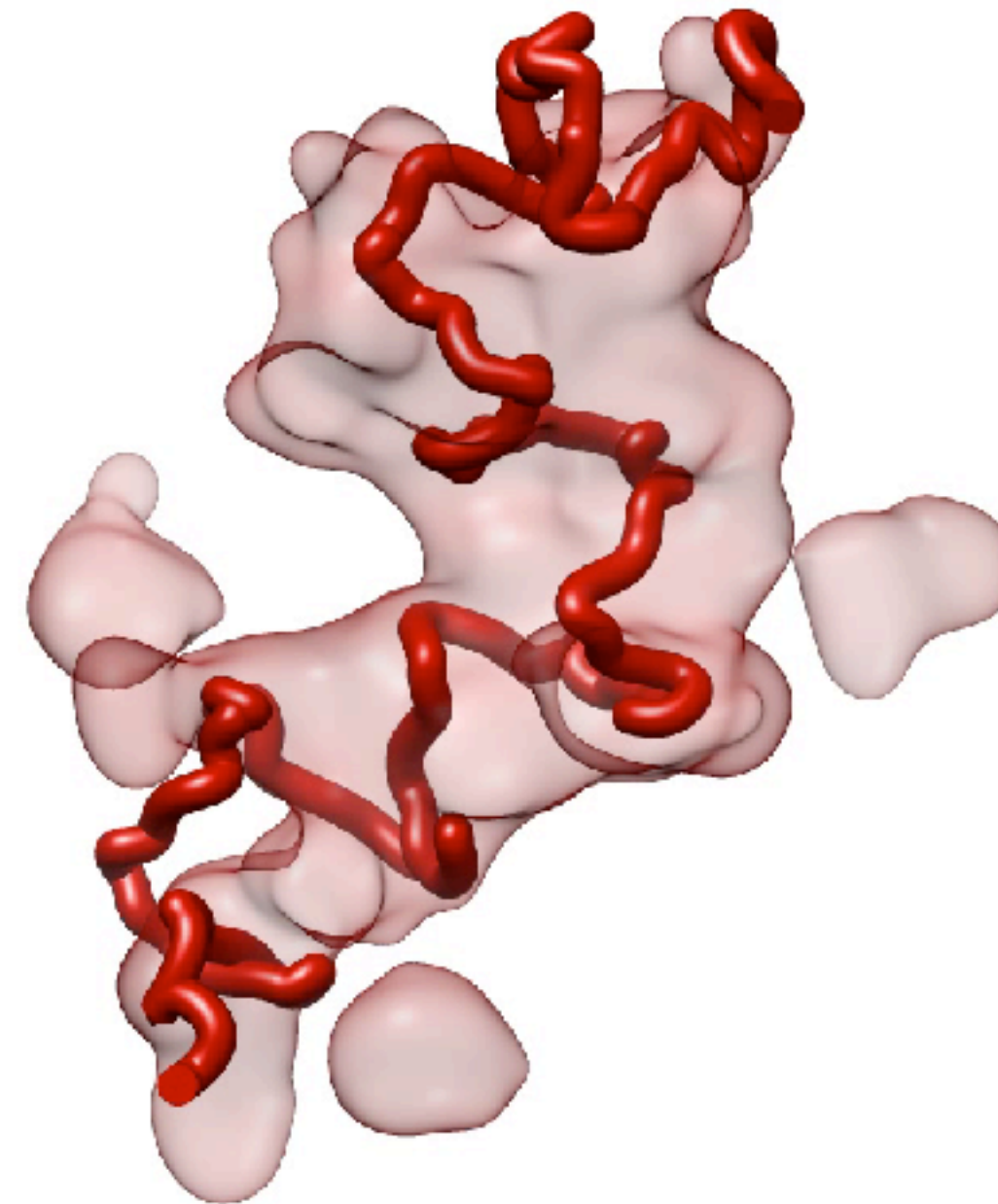
IMGR

Integrative Modelling of Genomic Regions

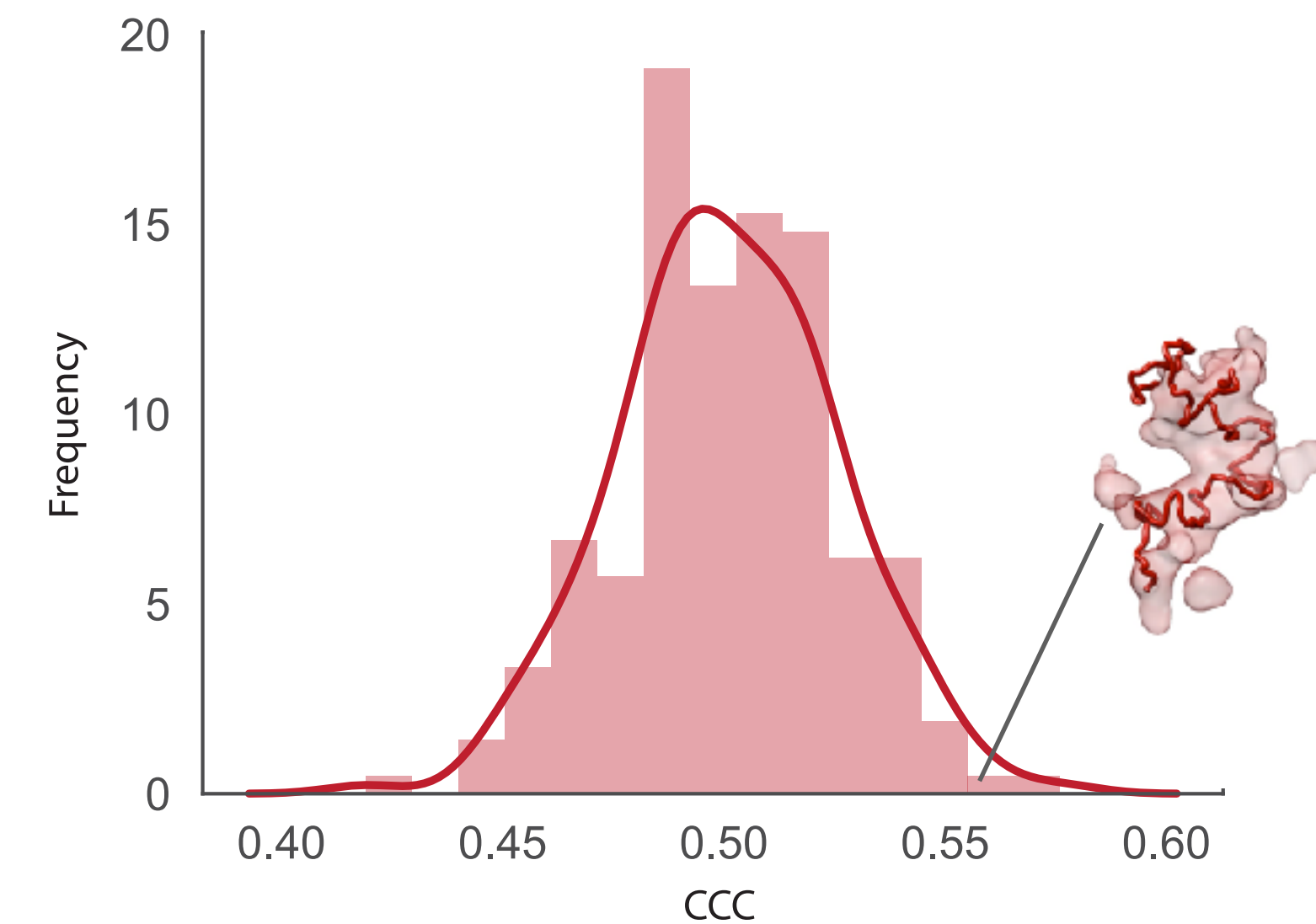


1

Rigid body fitting



$$CCC = \frac{\sum_{i \in M} (\rho_i^P - \bar{\rho}^P)(\rho_i^T - \bar{\rho}^T)}{\sqrt{[\sum_{i \in M} (\rho_i^P - \bar{\rho}^P)^2 \sum_{i \in M} (\rho_i^T - \bar{\rho}^T)^2]}}$$



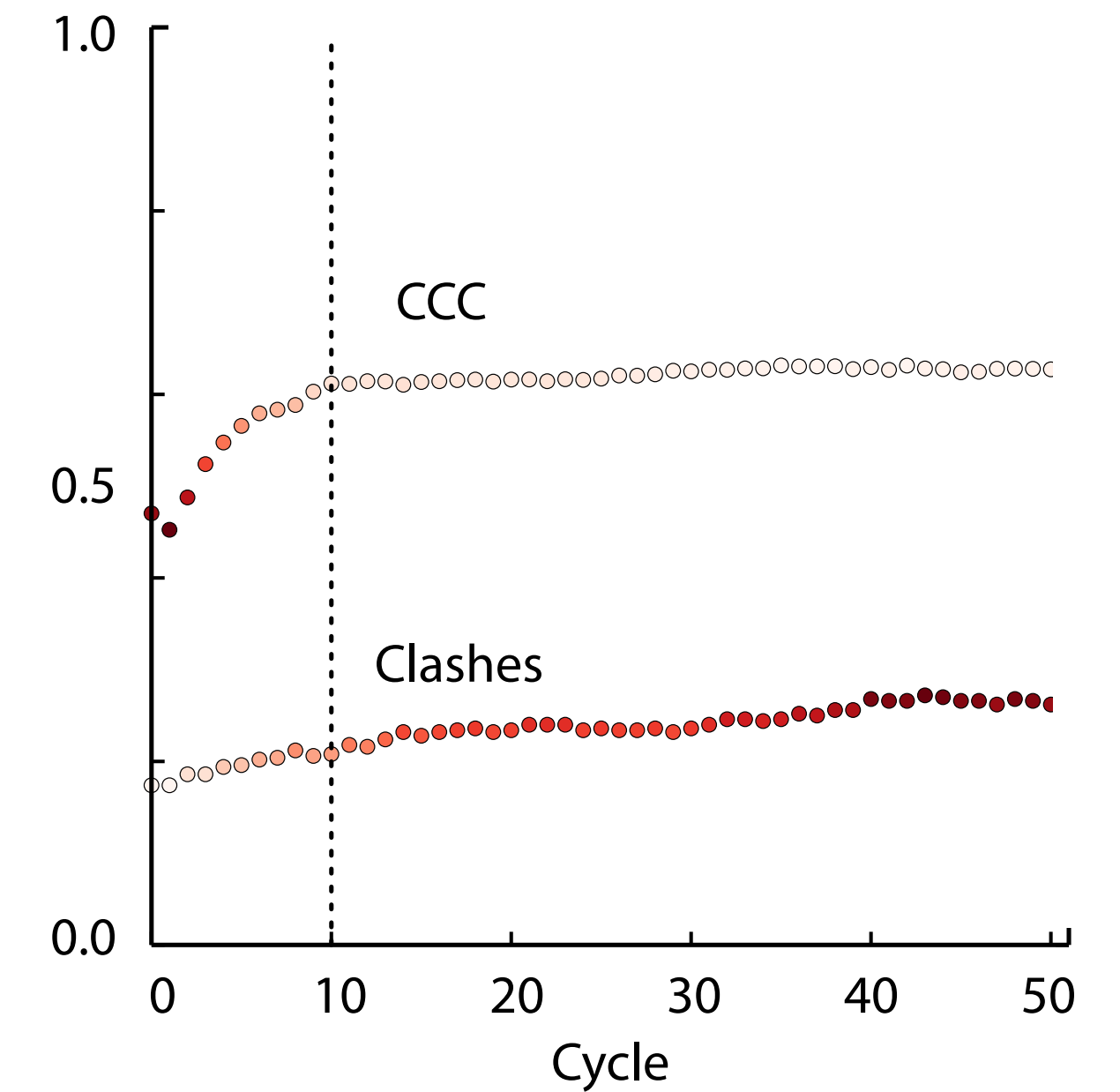
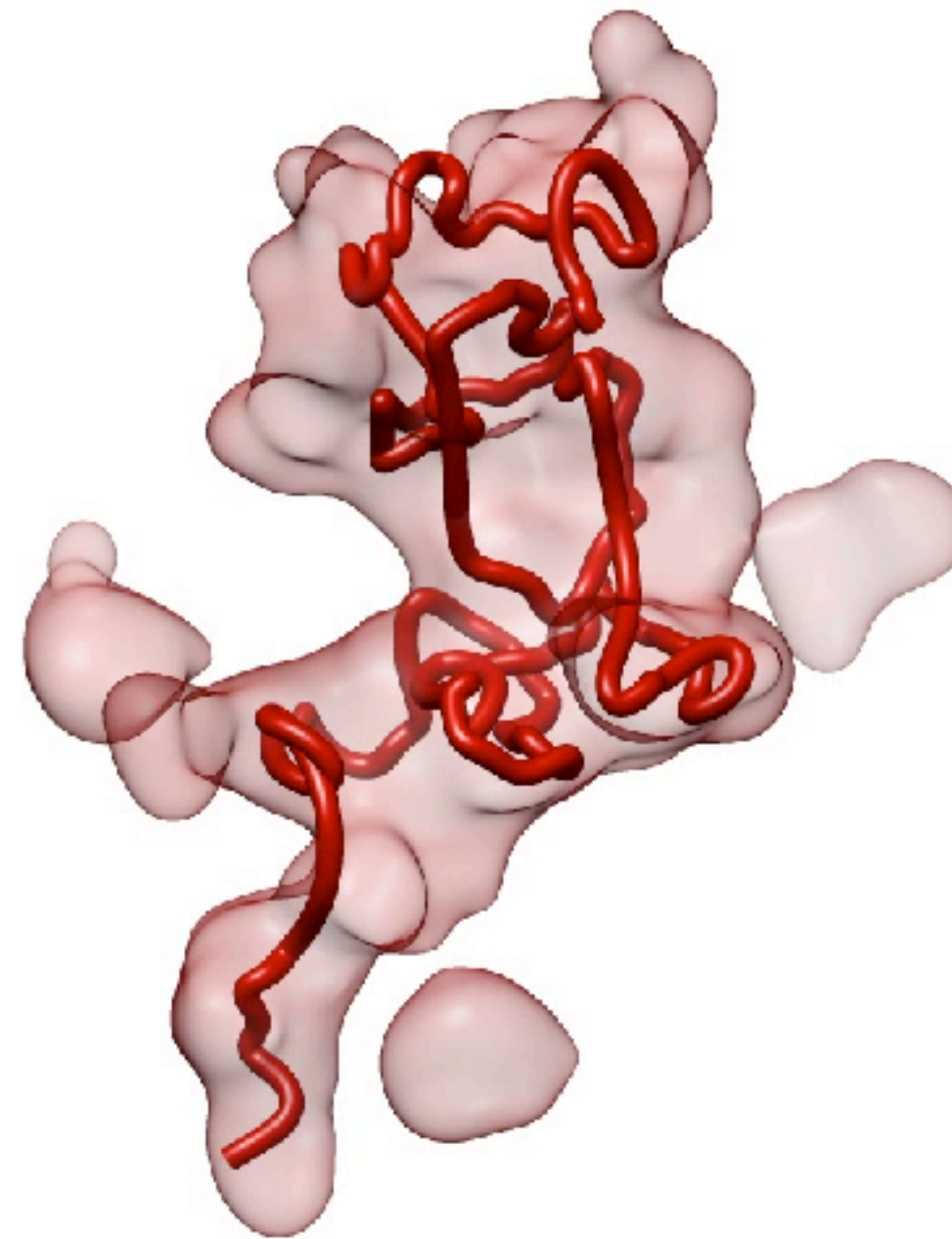
6D random search and optimisation based on goodness-of-fit

IMGR

Integrative Modelling of Genomic Regions

2

Flexible fitting refinement

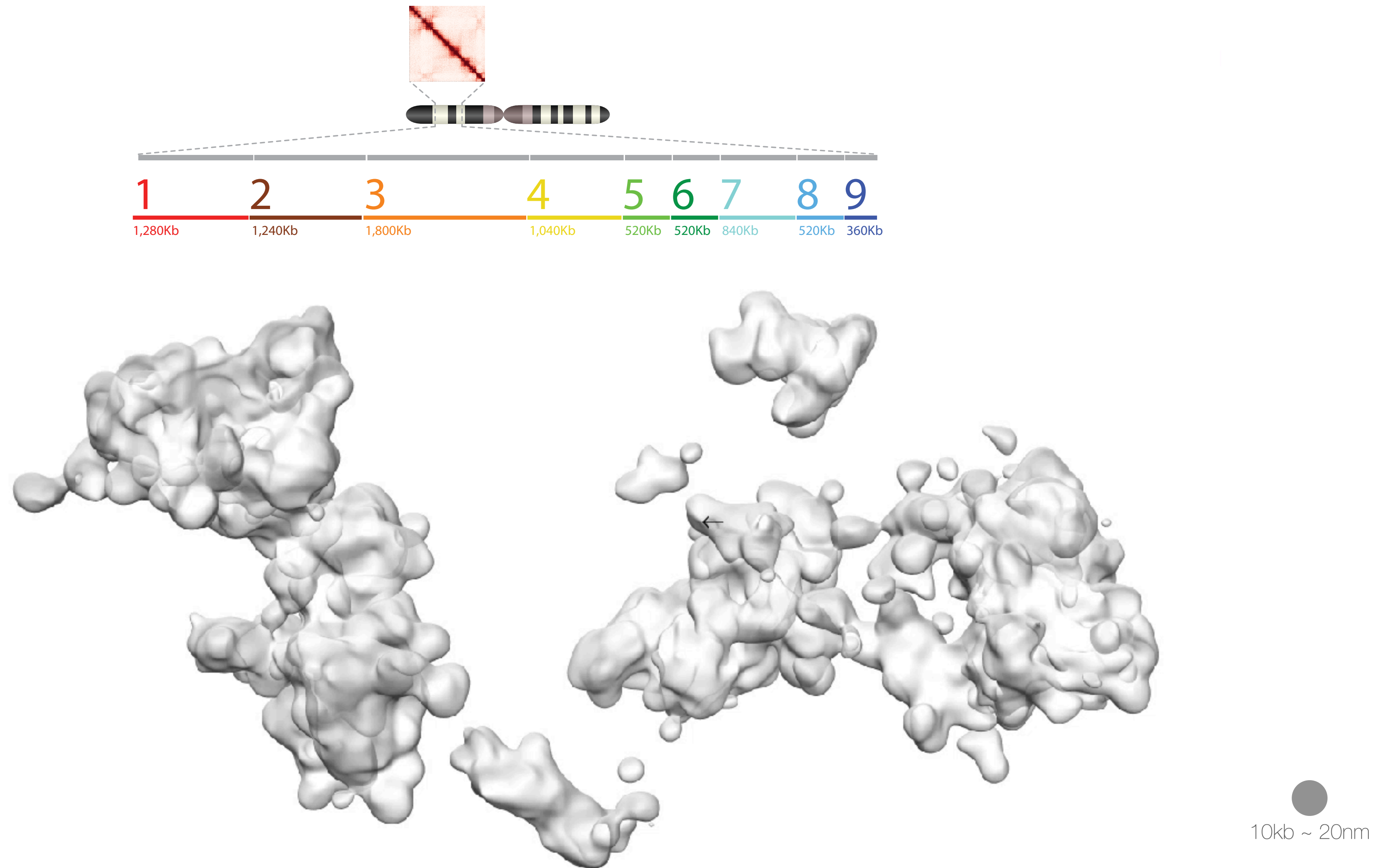


Simulated Annealing Molecular Dynamics / Conjugate Gradient minimisation

$$E = W_1 E^{\text{CCC}}(P) + W_2 E^{\text{H-LB}}(P) + W_2 E^{\text{H-UB}}(P) + W_3 E^{\text{EV}}(P) + W_2 E^{\text{C}}(P)$$

Chromosome walking path @10Kb resolution

8Mb homolog-specific model of diploid human cells



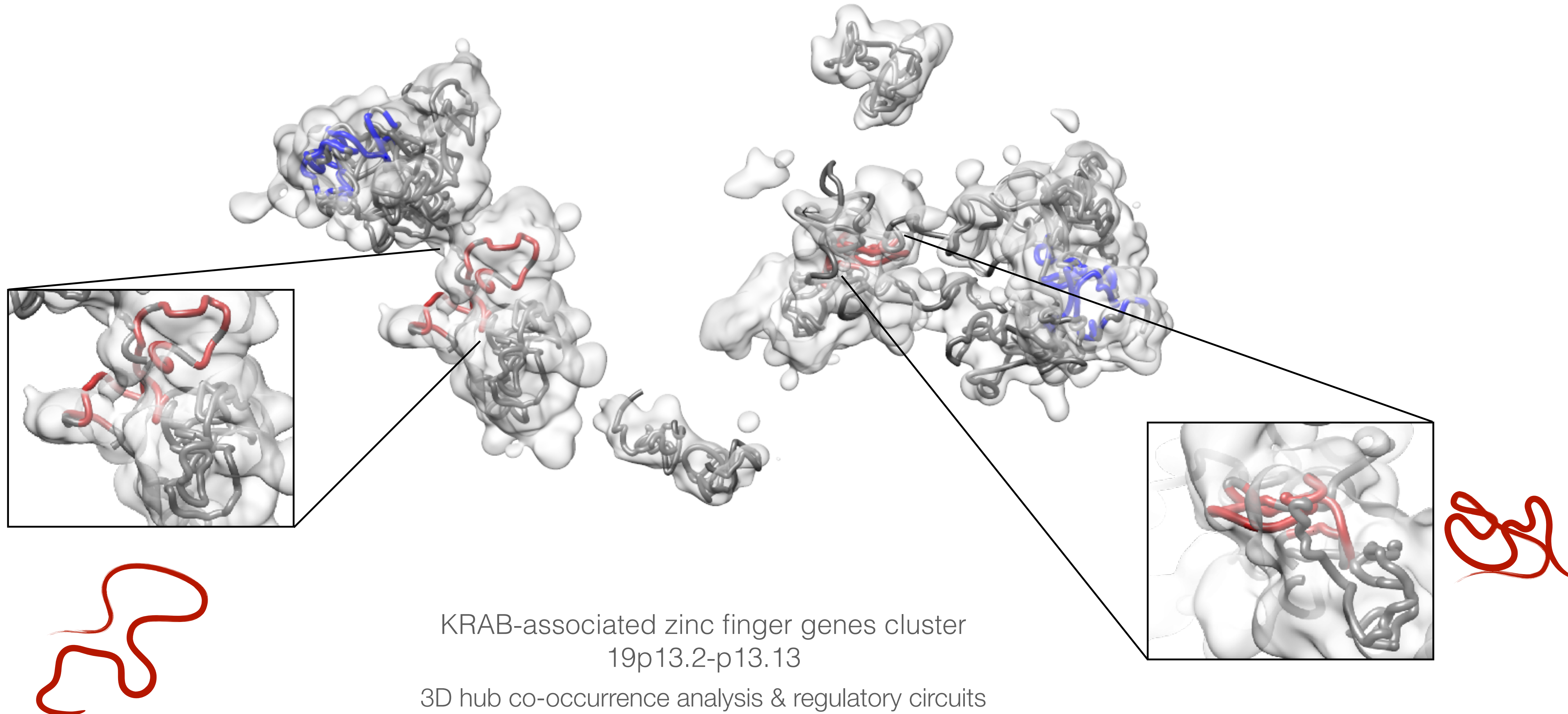
PGP1f chr19:7400000-15560000

Homolog-specific integrative model of diploid human cells @ 10kb

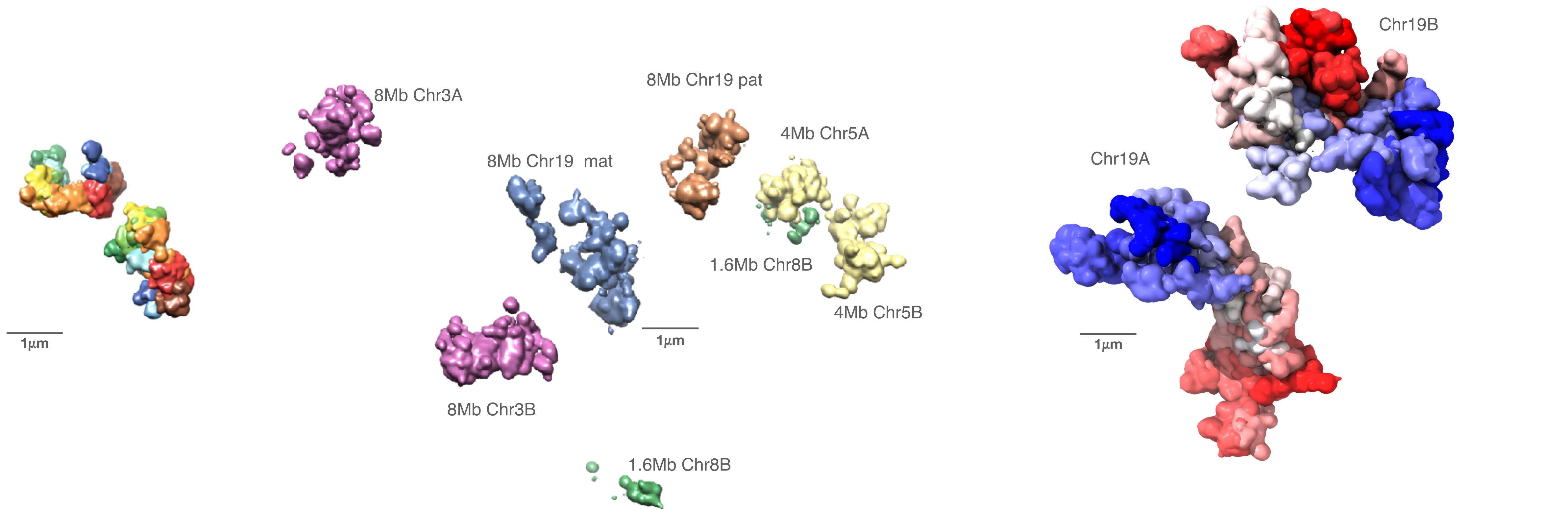
Nir*, Farabella*, Pérez Estrada*, Ebeling*, et al. PlosGen. 2018

Chromosome walking path @10Kb resolution

Connecting sequence and structure across length scales



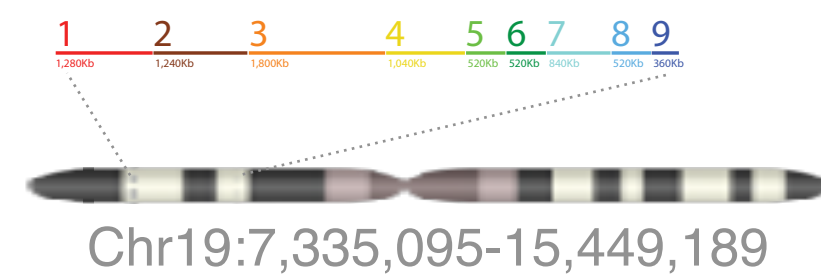
Visualising the Genome across scales



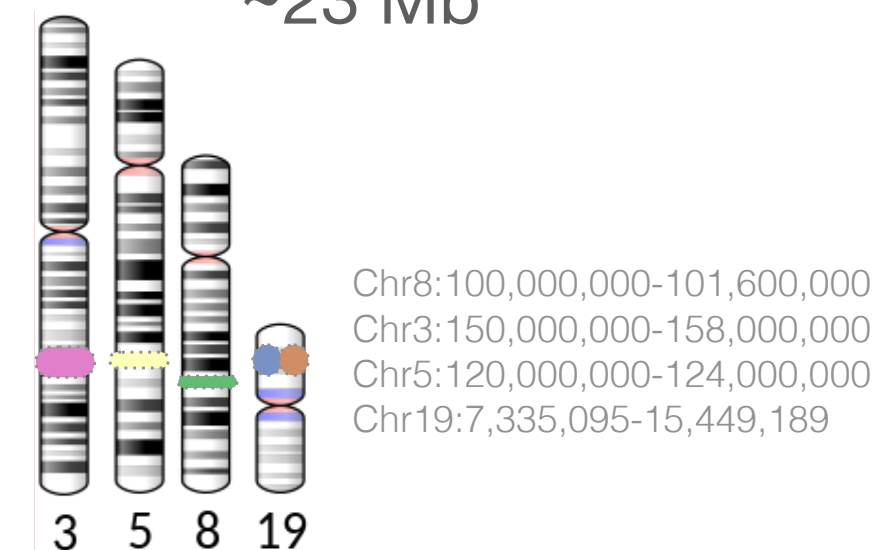
Genomic coverage

56 Mb

~8 Mb



~23 Mb



 iosonoirene.bsky.social

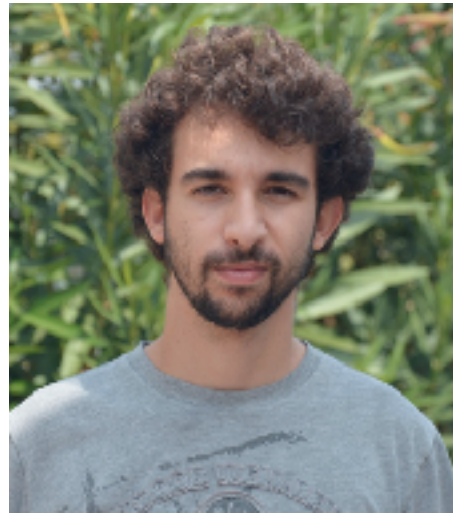
 @iosonoirene

irene.farabella@iit.it

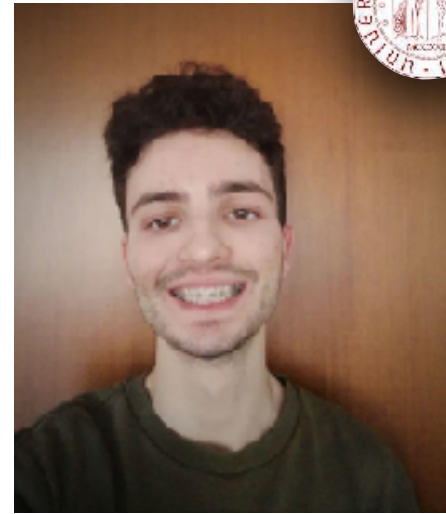
Acknowledgement



Chantal Usai



Ivan Piacere



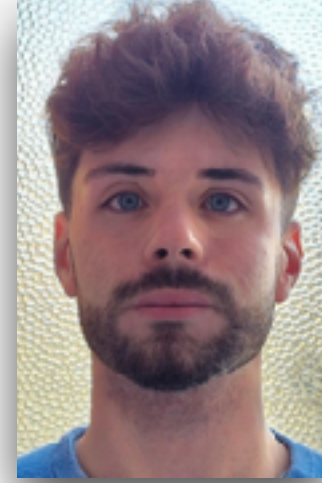
Anthony Palmieri



Campigli Di
Giammartino Lab



Anja Tolic



Mattia Roccatello

www.farabellalab.org

www.ina.iit.it



ISTITUTO ITALIANO
DI TECNOLOGIA
INTEGRATIVE NUCLEAR
ARCHITECTURE

the
giovanni

ARMENISE
HARVARD foundation

Alessandra Merlotti
Nico Curti
Daniel Remondini



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



4DN Imaging WG



Mark Bates
Lenny Reinkensmeier



Francesco Rinaldi



Wu Lab
Boston
hwu@genetics.med.harvard.edu

Aiden Lab
Houston
brez.leberman@bcm.edu

Marti-Renom Lab
Barcelona
martiren@cnag.org.eu

Neretti Lab
Providence
nicola_neretti@brown.edu

Nir Lab

Jumana AlHaj
Abed Lab
Freiburg

Center for Genome Imaging
www.cegs-cgi.org

Simone Civita
Alberto Diaspro
Michele Oneto
Paolo Bianchini

