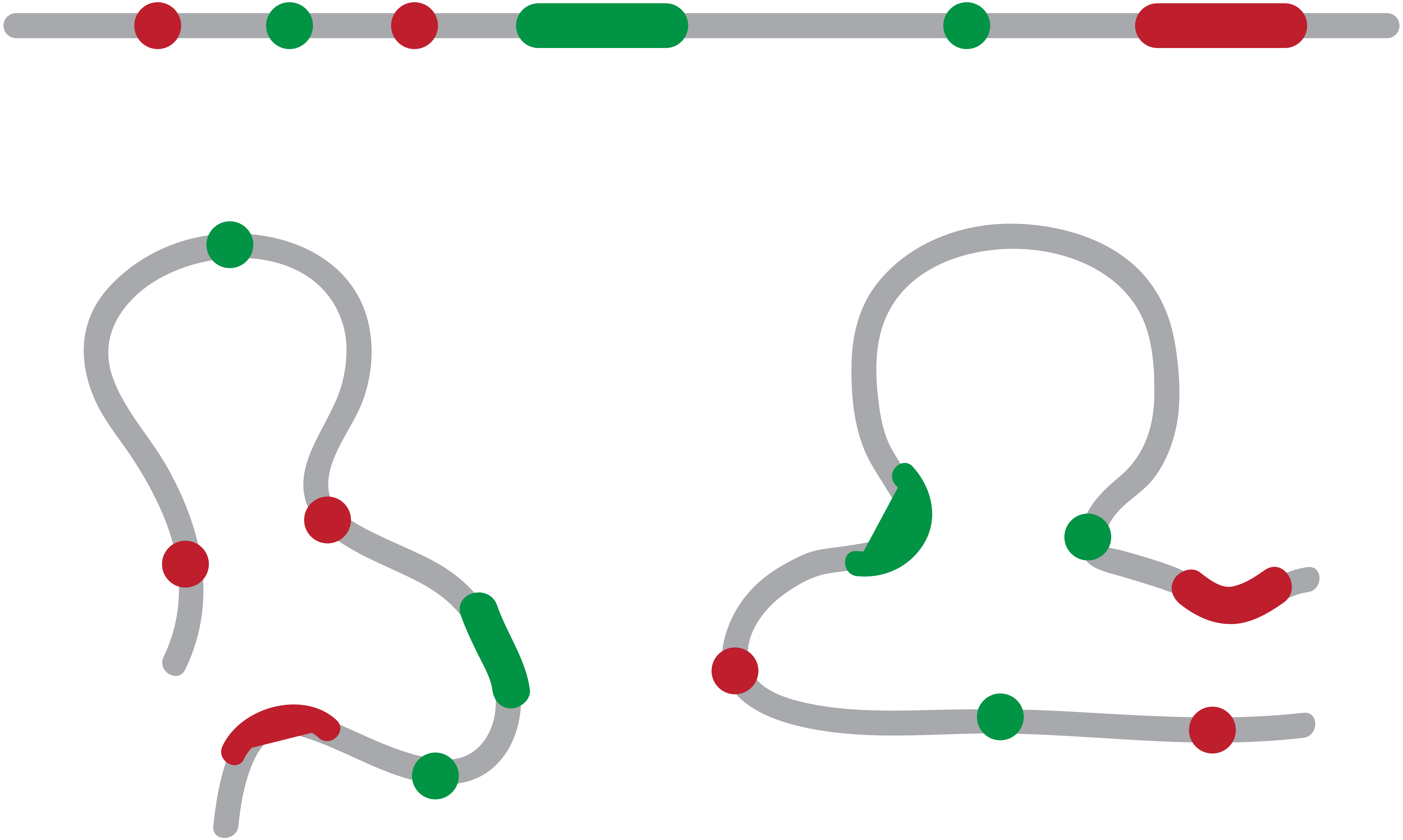

Photo by David Oliete - www.davidoliete.com

Structure determination of genomes and genomic domains by satisfaction of spatial restraints

Marc A. Marti-Renom
CNAG-CRG · ICREA

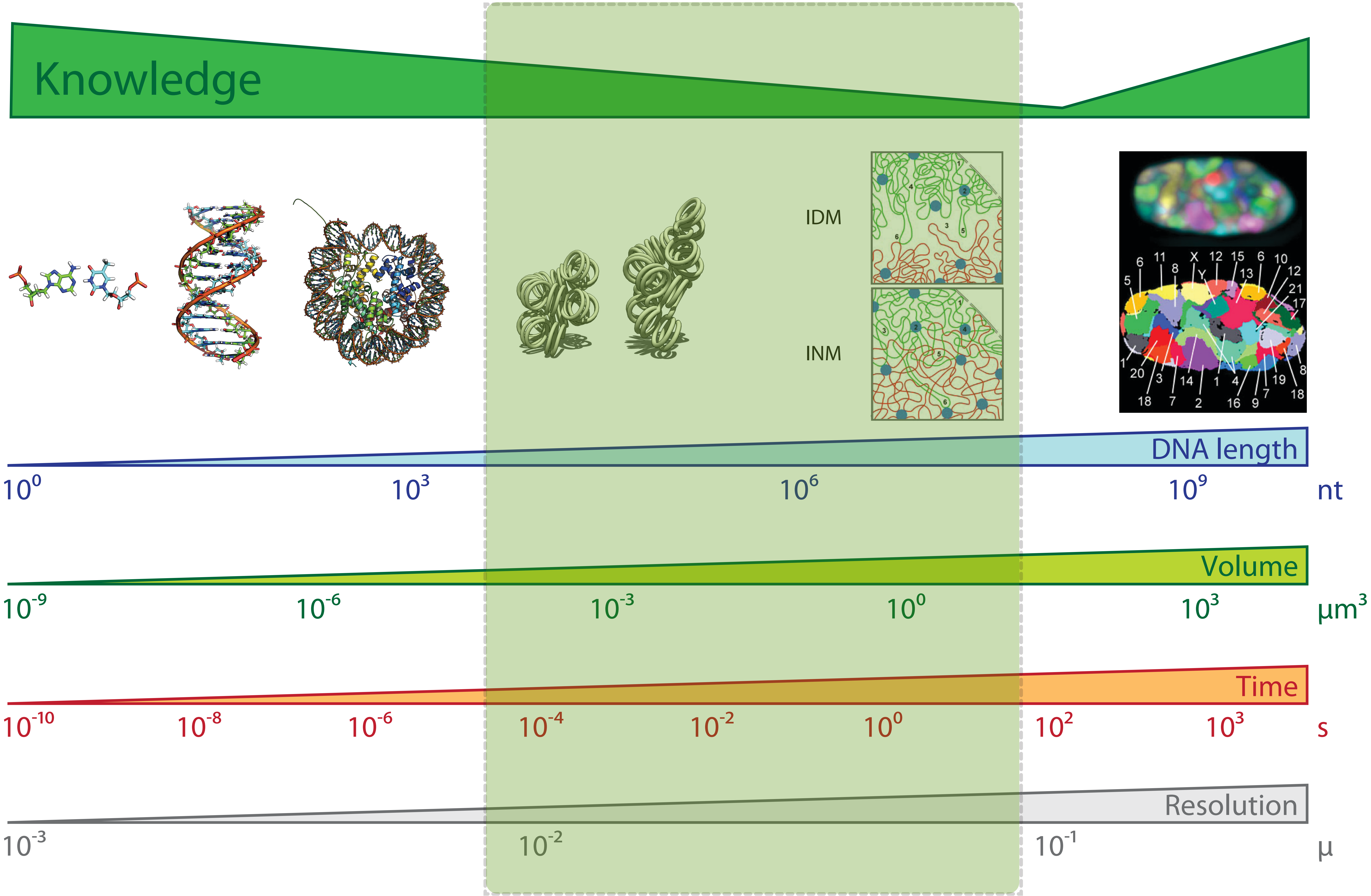
http://marciuslab.org
http://3DGenomes.org
http://cnag.crg.eu

@marciuslab
@mamartirenom



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

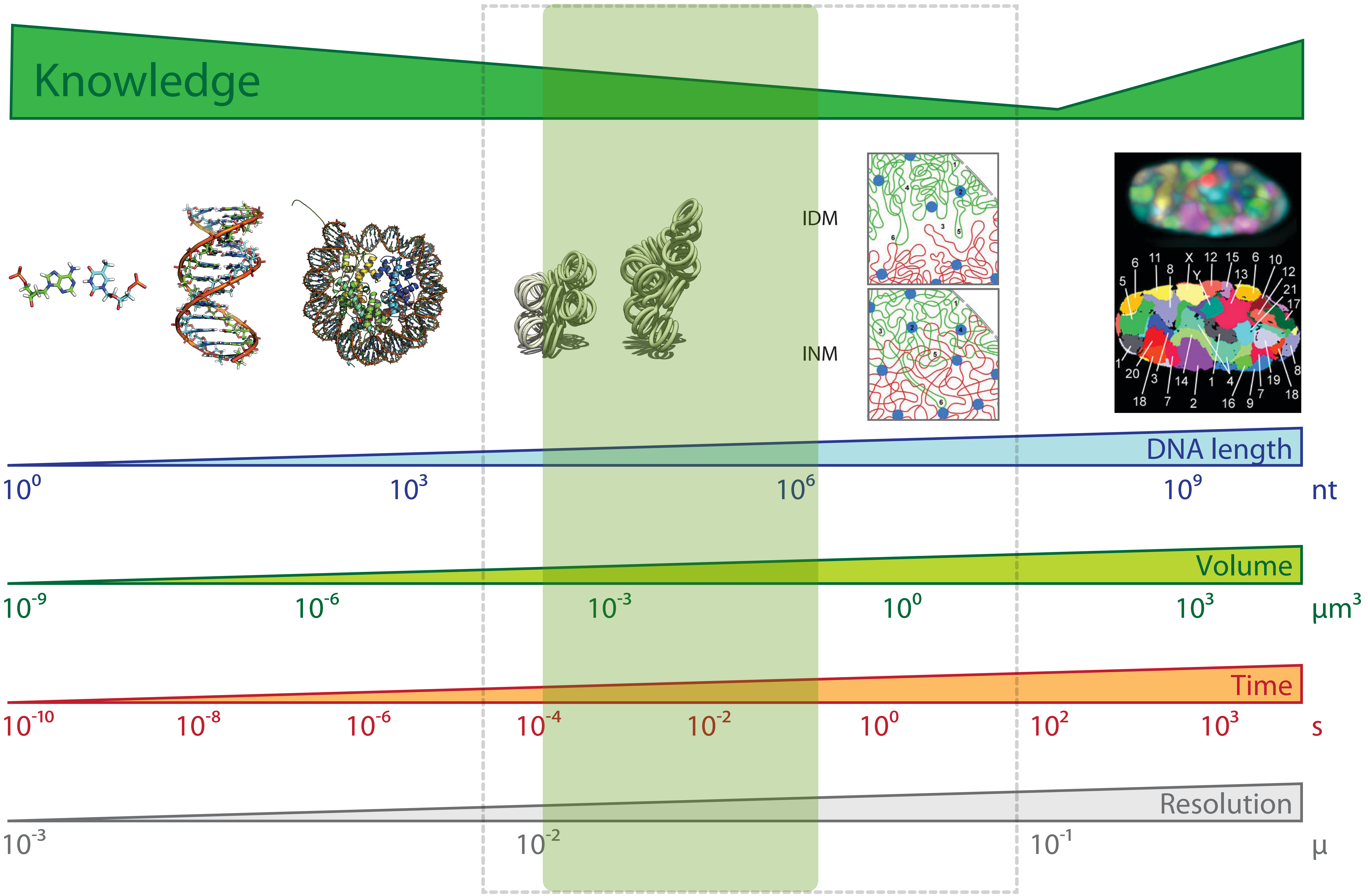
10^{-10} 10^{-8} 10^{-6} 10^{-4} 10^{-2} 10^0 10^2 10^3 s

Resolution

10^{-3} 10^{-2} 10^{-1} μ

Resolution Gap

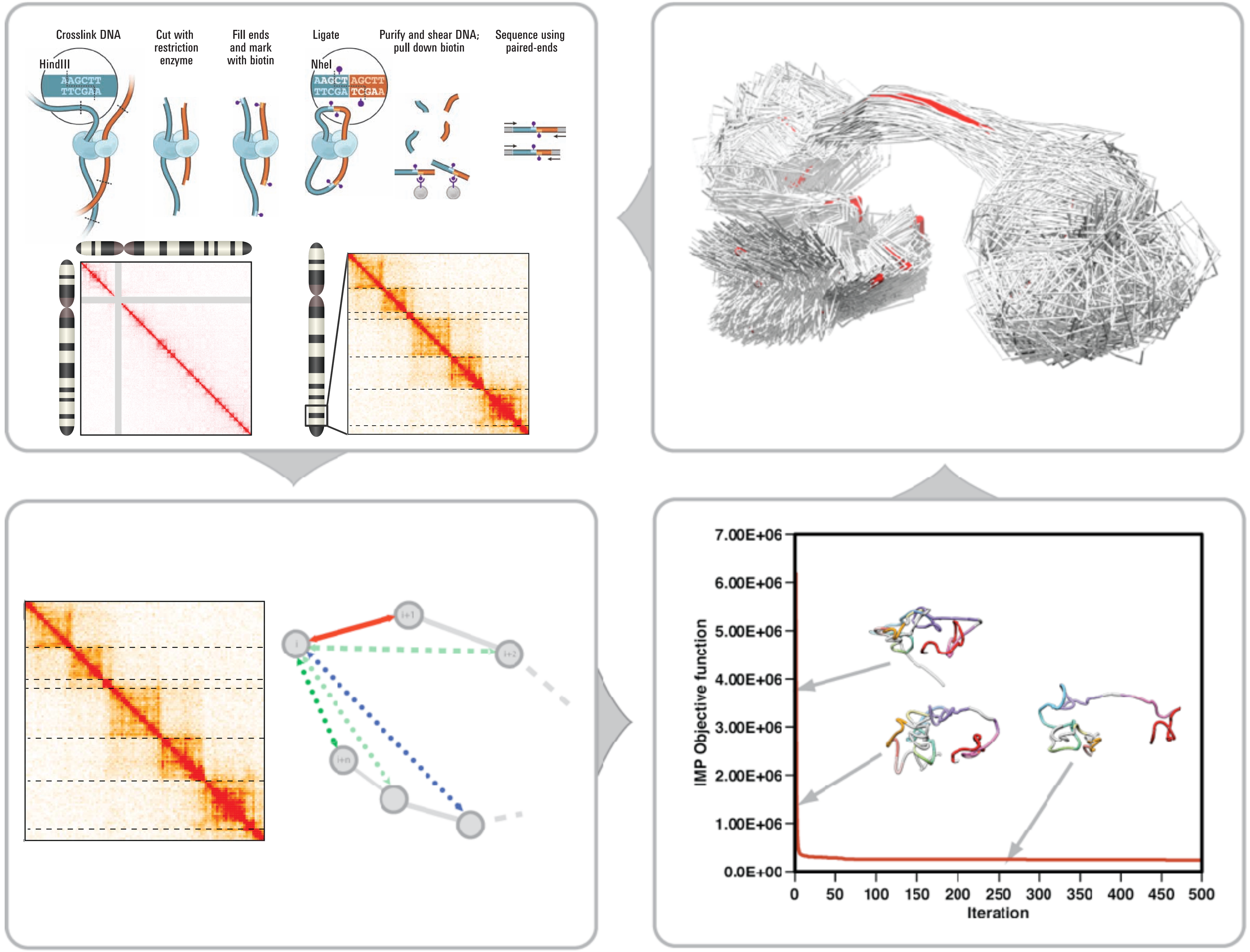
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol 7, e1002125 (2011)



Hybrid Method

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

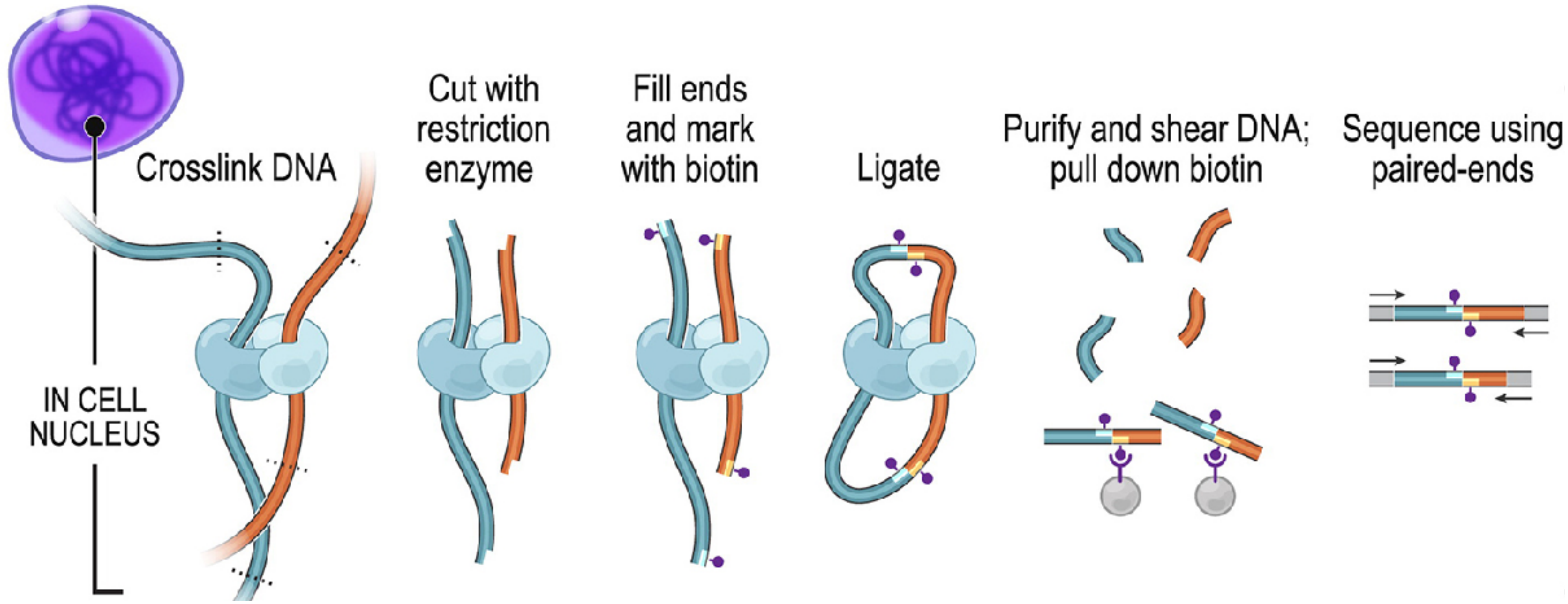
Experiments



Computation

Chromosome Conformation Capture

Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Science, 295(5558), 1306–1311.
Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.



IN CELL NUCLEUS

Crosslink DNA

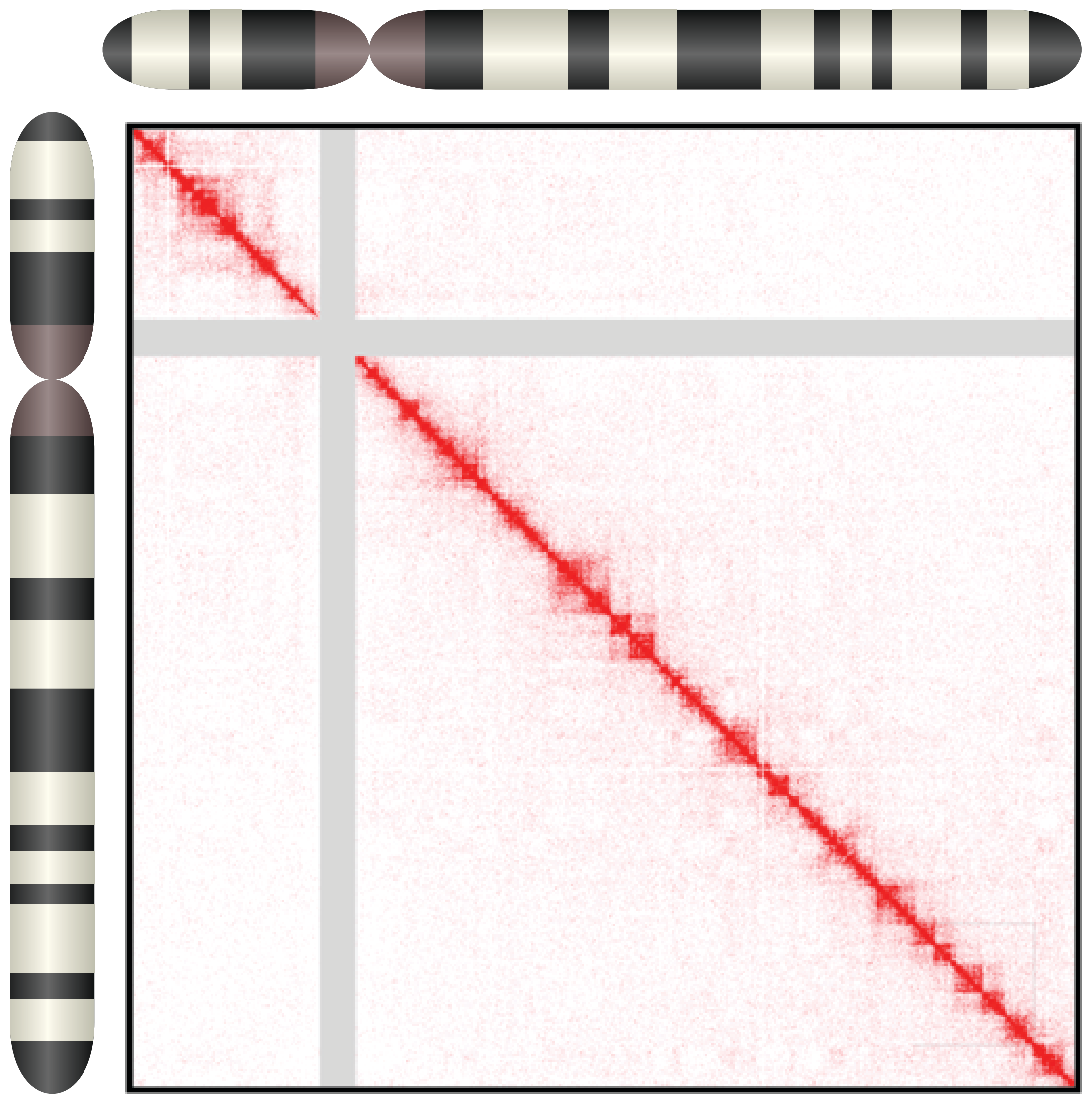
Cut with restriction enzyme

Fill ends and mark with biotin

Ligate

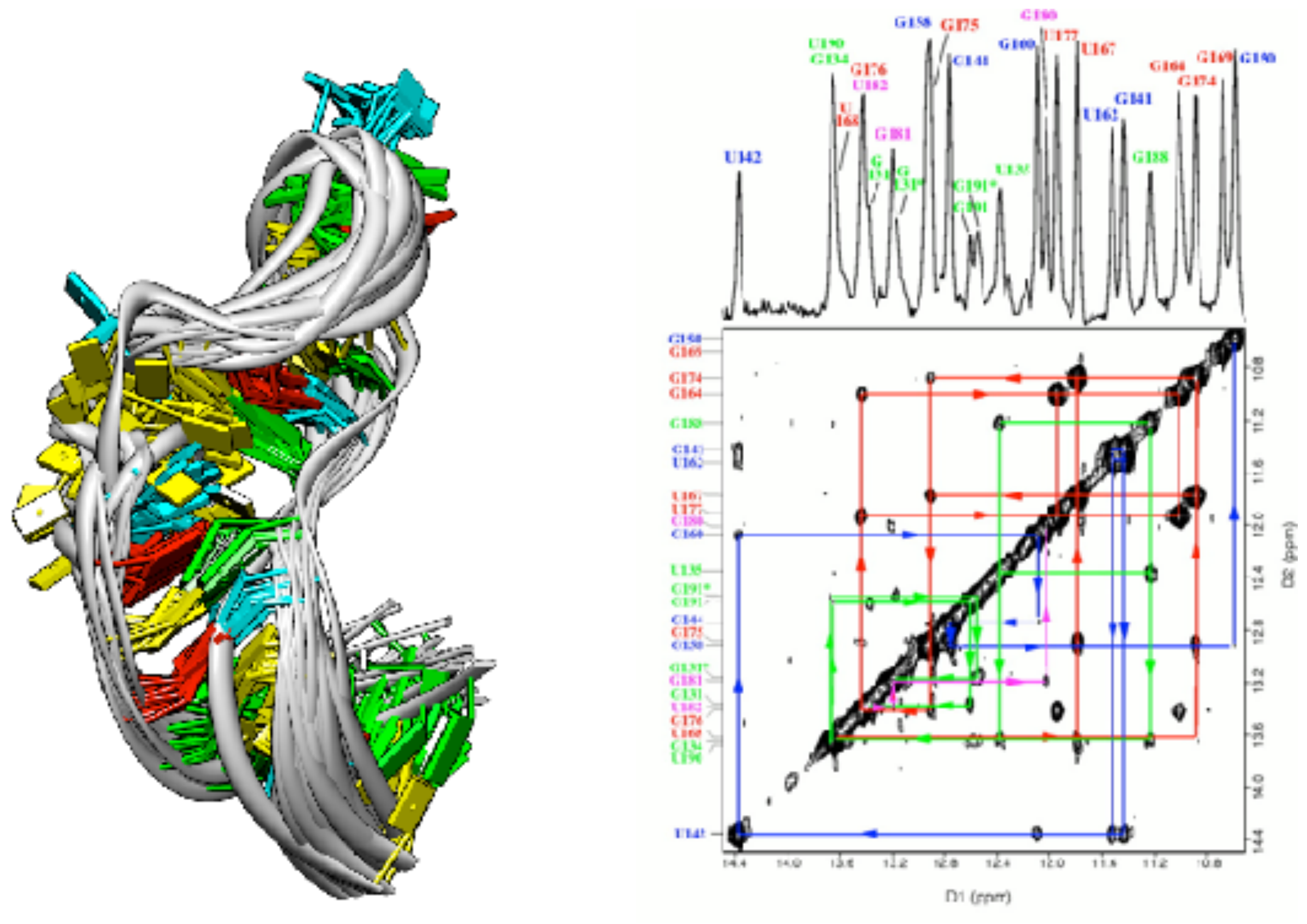
Purify and shear DNA; pull down biotin

Sequence using paired-ends



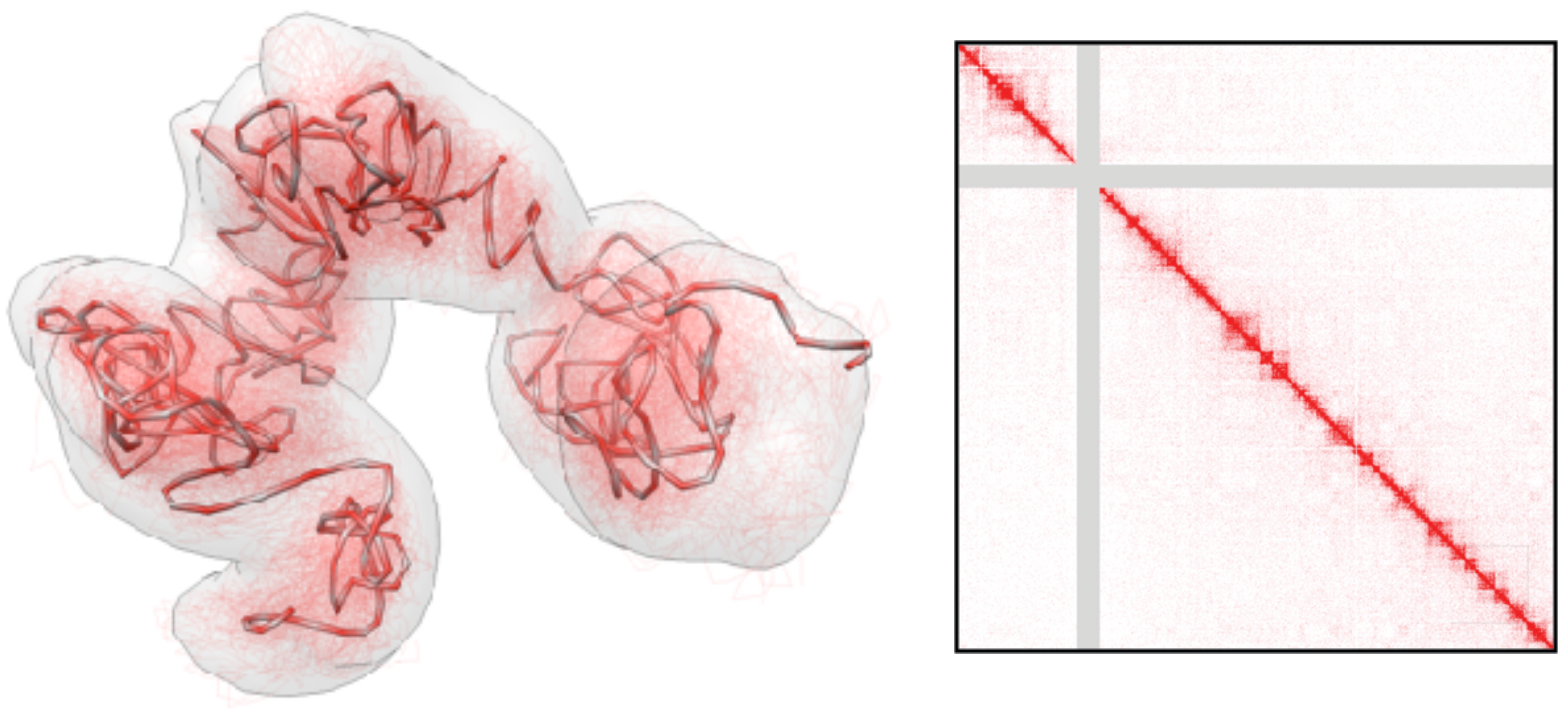
Restraint-based Modeling

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



Biomolecular structure determination

2D-NOESY data

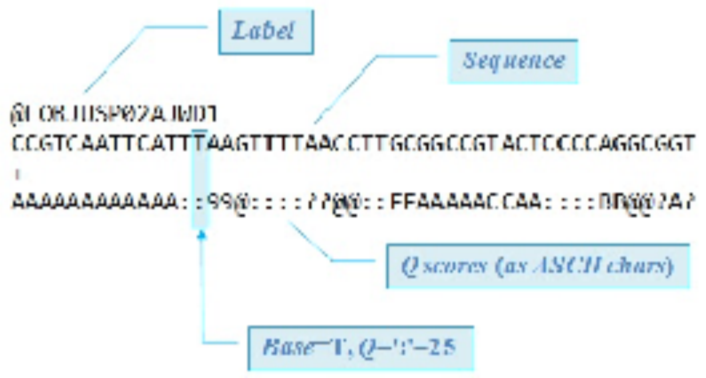


Chromosome structure determination

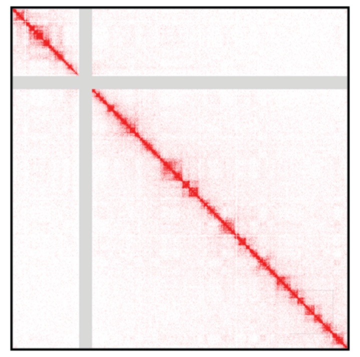
3C-based data

TADbit

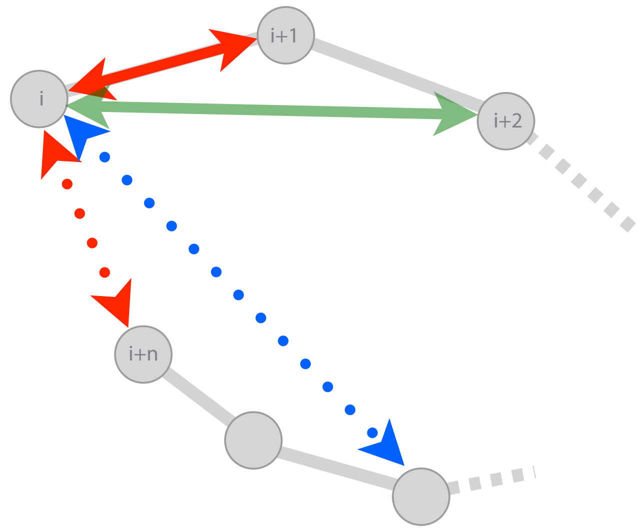
http://3DGenomes.org



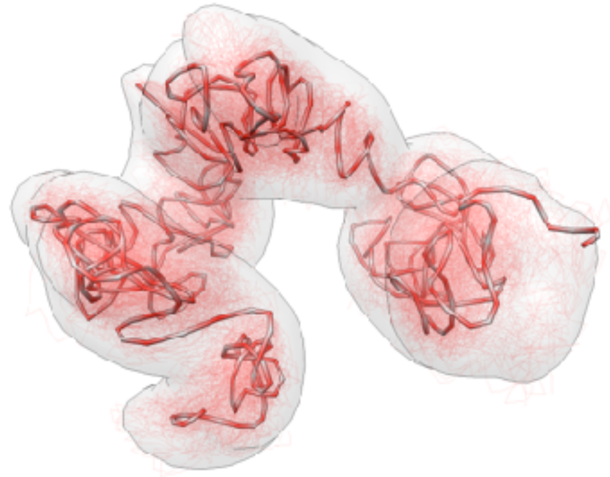
FastQ files to Maps



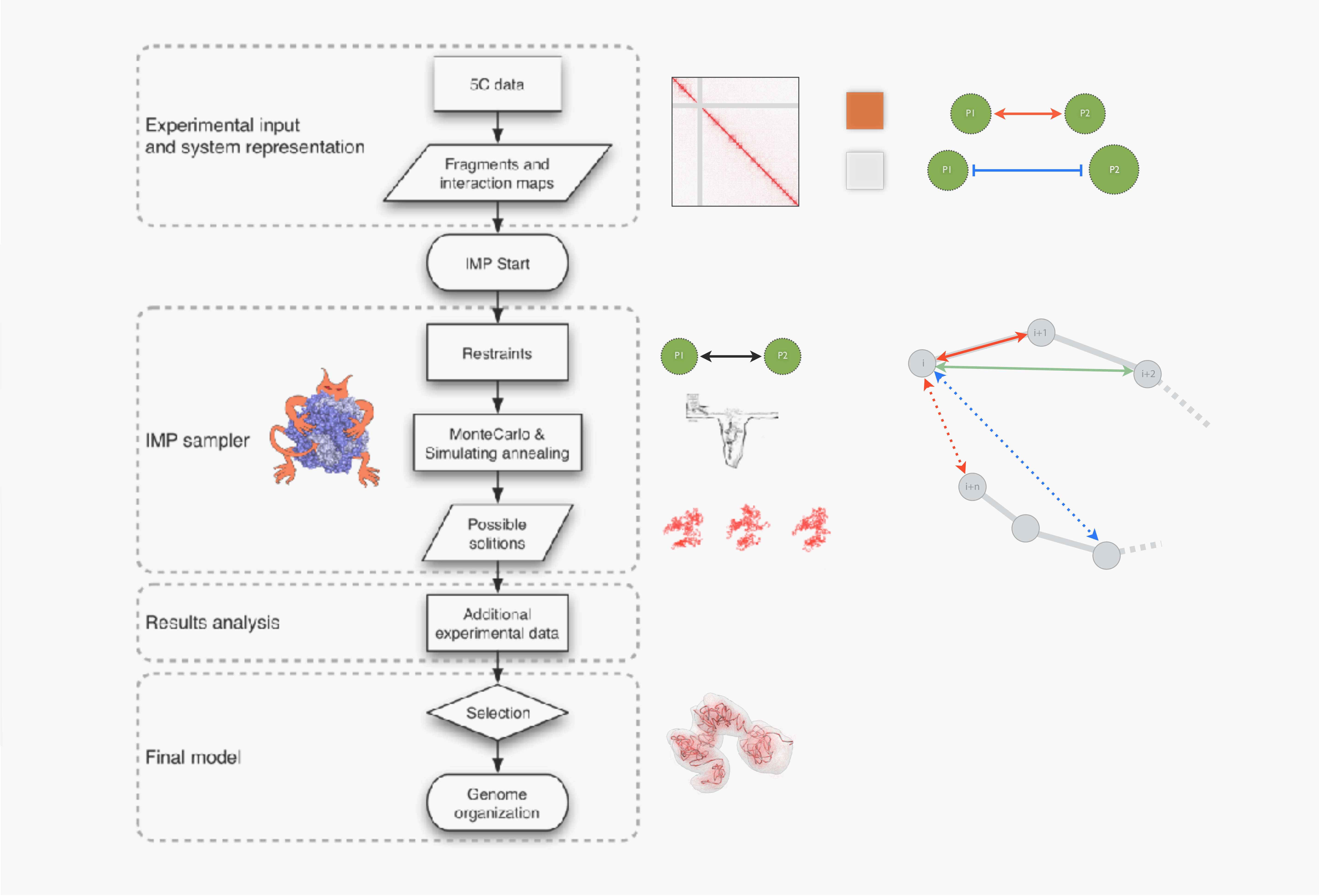
Map analysis



Model building



Model analysis



Experimental input and system representation

5C data

Fragments and interaction maps

IMP Start

IMP sampler

Restraints

MonteCarlo & Simulating annealing

Possible solutions

Results analysis

Additional experimental data

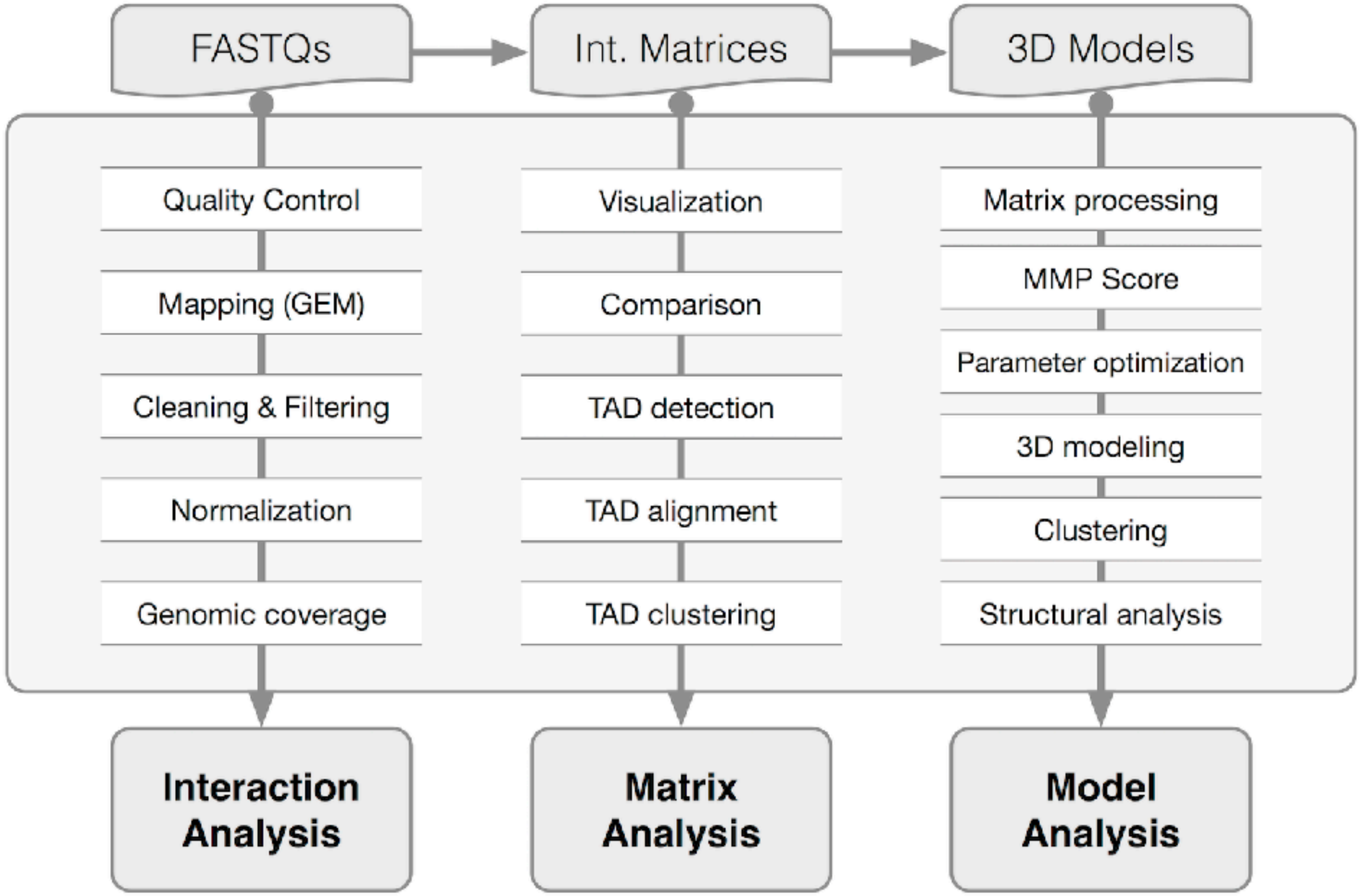
Final model

Selection

Genome organization

TADbit

Serra, Baù, et al. (2017). PLOS CompBio



FASTQs → Int. Matrices → 3D Models

FASTQs
- Quality Control
- Mapping (GEM)
- Cleaning & Filtering
- Normalization
- Genomic coverage

Interaction Analysis

Int. Matrices
- Visualization
- Comparison
- TAD detection
- TAD alignment
- TAD clustering

Matrix Analysis

3D Models
- Matrix processing
- MMP Score
- Parameter optimization
- 3D modeling
- Clustering
- Structural analysis

Model Analysis

- Baù, D. et al. Nat Struct Mol Biol (2011)
- Umbarger, M. A. et al. Mol Cell (2011)
- Le Dily, F. et al. Genes & Dev (2014)
- Belton, J.M. et al. Cell Reports (2015)
- Trussart M. et al. Nature Communication (2017)
- Cattoni, D. et al. Nature Communication (2017)
- Stadhouders R. et al. Nature Genetics (2018)
- Kojic, A., Cuadrado, A. et al. Nat Struct Mol Biol (2018)
- Beekman R. et al. Nature Medicine (2018)
- Mas, G. et al. Nature Genetics (2018)
- Pascual-Reguant, L. et al. Nature Communication (2018)
- Nir, Farabella, Perez-Estrada, et al. PLOS Genetics (2018)
- Cuadrado, Giménez-Llorente et al. Cell Reports (2019)
- Vara et al. Cell Reports (2019)
- Miguel-Escalada et al. Nature Genetics (2019)
- Morf et al. Nature Biotechnology (2019)

Nature Structural & Molecular Biology, 25(9), 766-777, 2018
Cell, 173(7), 1796-1809.e17, 2018
Structure, 26(6), 894-904.e2, 2018
Genome Research, 29(1), 29-39, 2019
Genome Research, 29(1), gr.238527.118, 2019

Chromosome walking with super-resolution imaging and modeling

Photo by David Oliete - www.davidoliete.com

Guy Nir
Irene Farabella
Cynthia Perez-Estrada
with Wu Lab (HMS, Boston) & Aiden Lab (UT, Texas)

PLOS Genetics (2018) 14(12) e1007872

Can we walk the chromatin path in the nucleus?

by

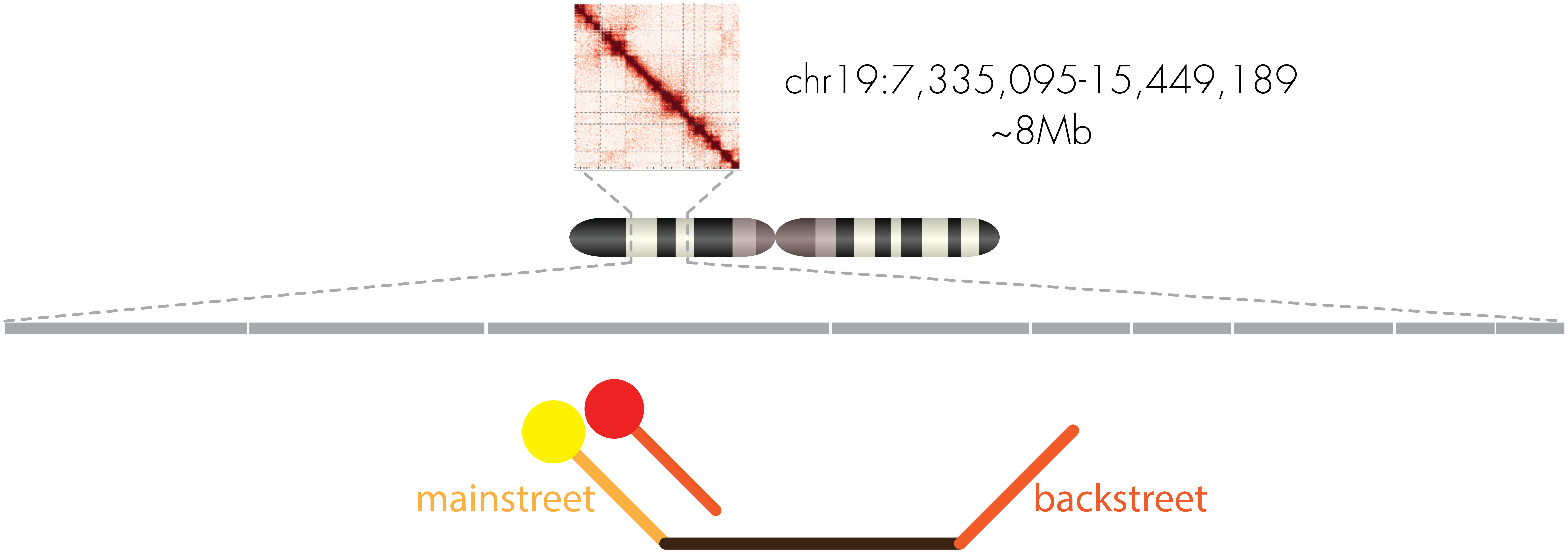
Integrating imaging and Hi-C maps with modeling.

by developing a method for

Oligopaint-based modeling of genomes

High-resolution imaging

Tracing chromosomes with OligoSTORM & fluidics cycles in PGP1 cells



chr19:7,335,095-15,449,189

~8Mb

mainstreet

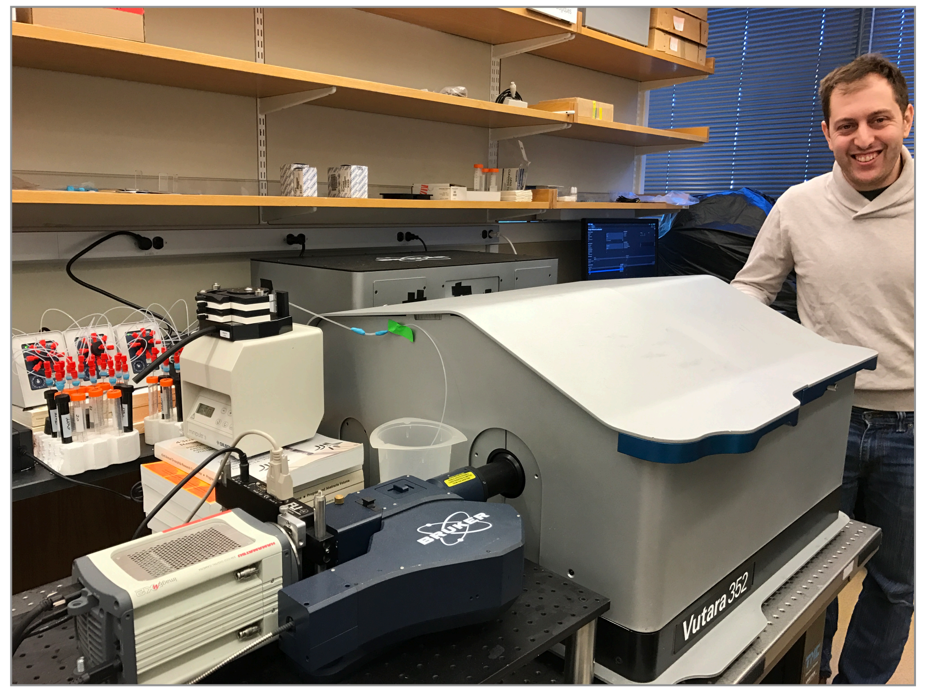
backstreet

homologous 32-42bp

Beliveau et al. Nat. Comm. 2015

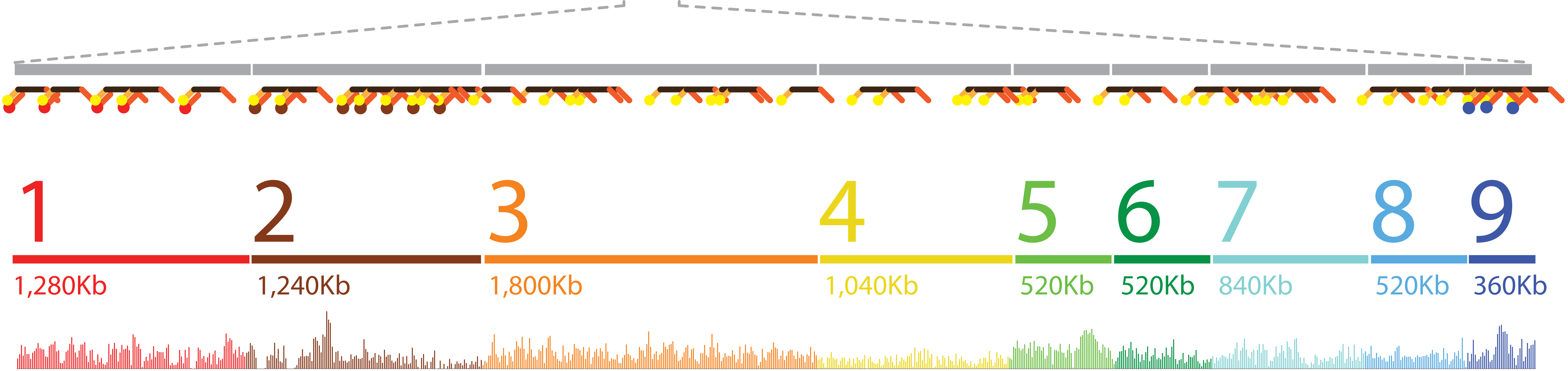
High-resolution imaging

Tracing chromosomes with OligoSTORM & fluidics cycles in PGP1 cells



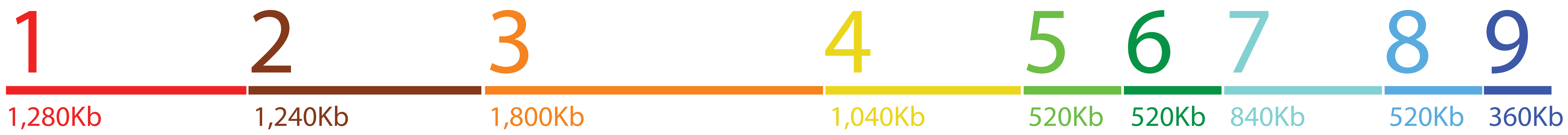
Guy Nir Harvard Med School

Bodgan Bintu Harvard
Carl Ebeling Bruker
Jeff Stuckey Bruker
John Schreiner Zero Epsilon
Steve Callahan Zero Epsilon

High-resolution imaging

Tracing chr19:7,335,095-15,449,189 ~8Mb



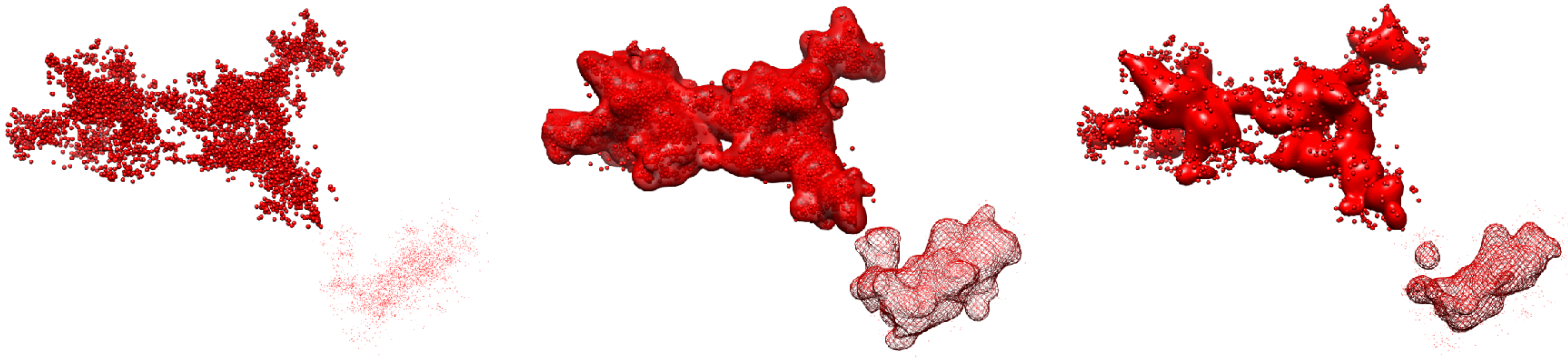
1	2	3	4	5	6	7	8	9
1,280Kb	1,240Kb	1,800Kb	1,040Kb	520Kb	520Kb	840Kb	520Kb	360Kb

Cell-02

High-resolution imaging

XYZ points convolution into a density map

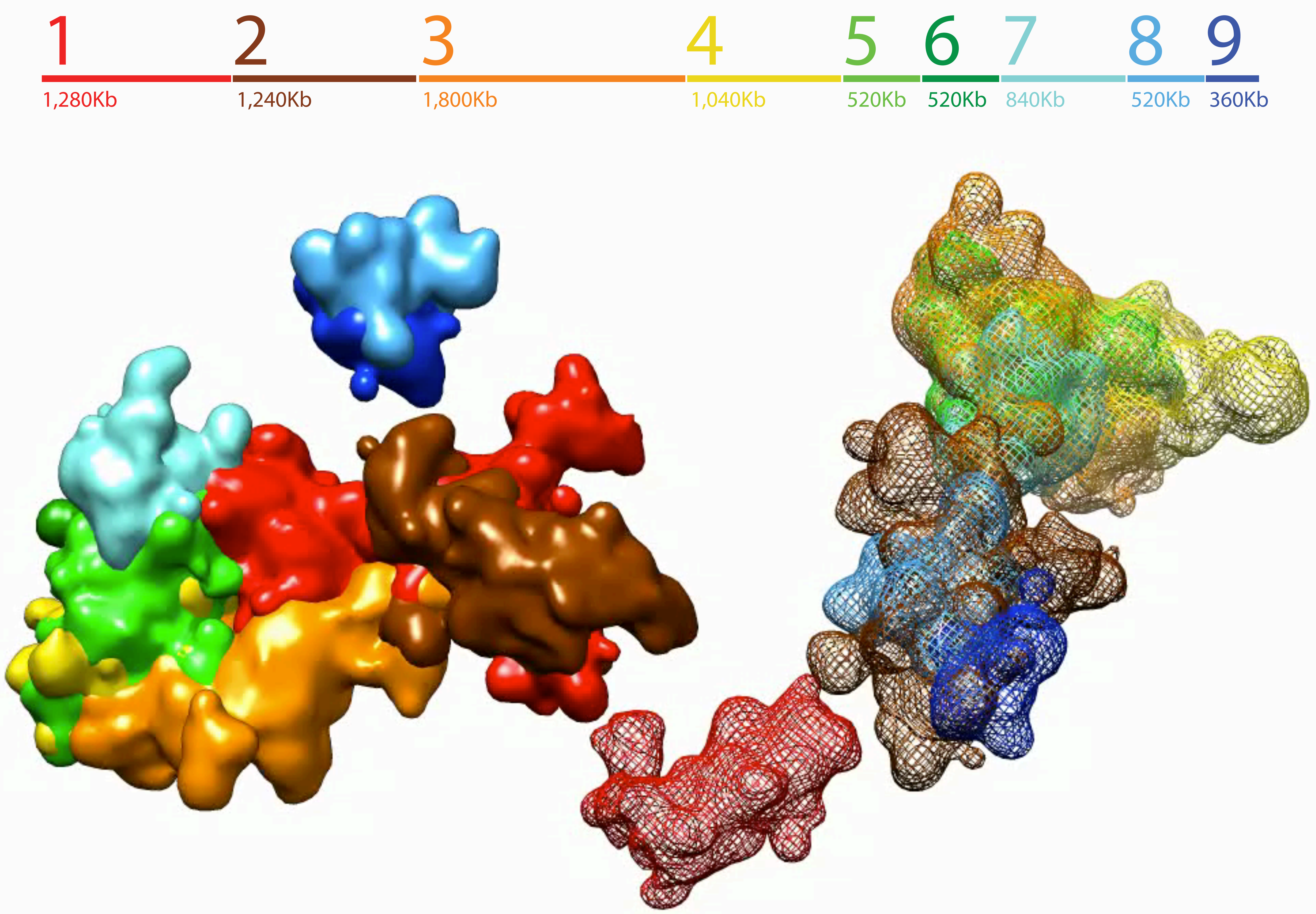
$$\rho(x,y,z)=\sum_N \frac{Z_N}{\left(\sigma\sqrt{2\pi}\right)^3} e^{-\frac{(x-x_n)^2+(y-y_n)^2+(z-z_n)^2}{2\sigma^2}}$$



Farabella et al, J Appl Crystallogr. 2015

Density maps

Cell-02 · Density map @ 50nm

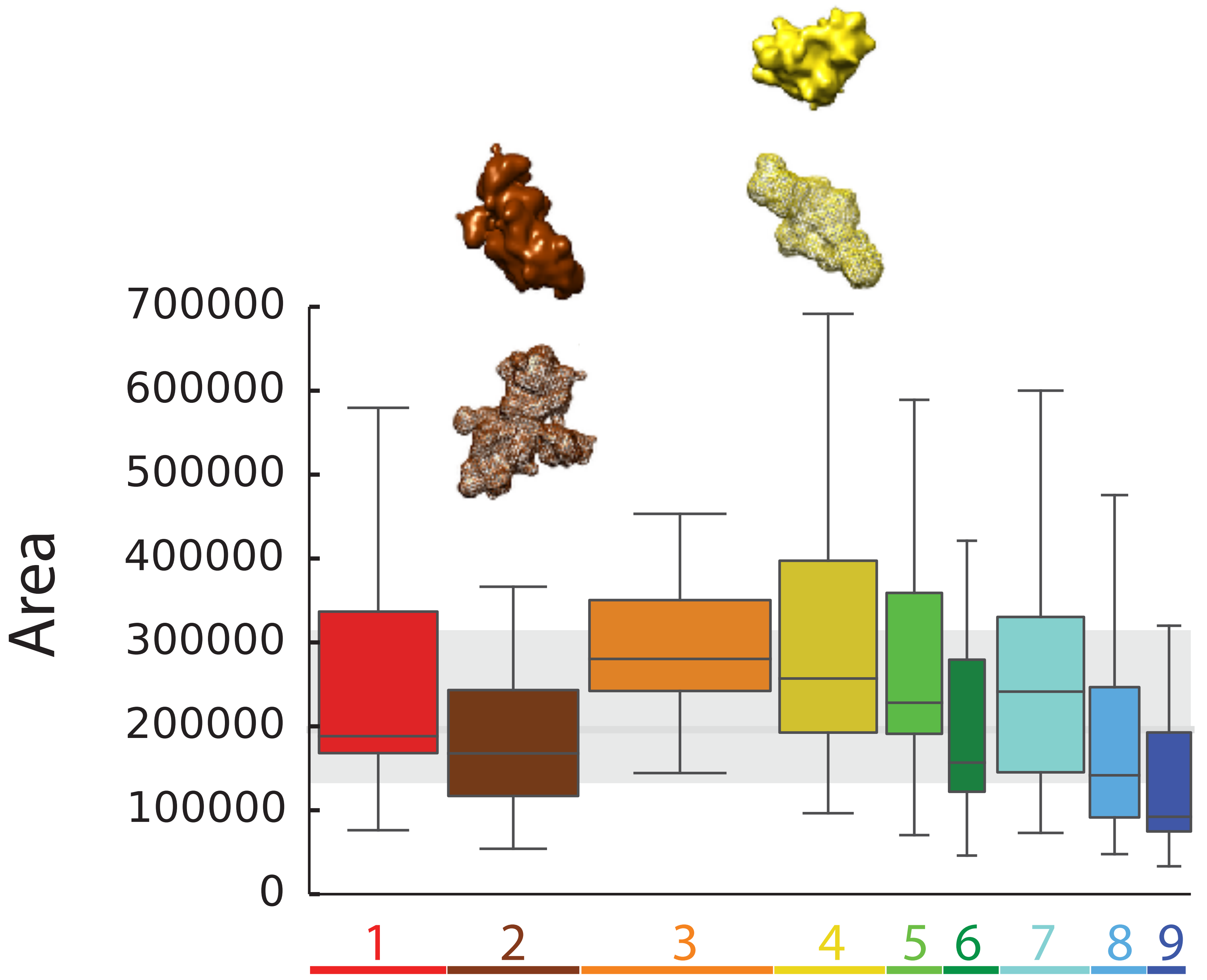


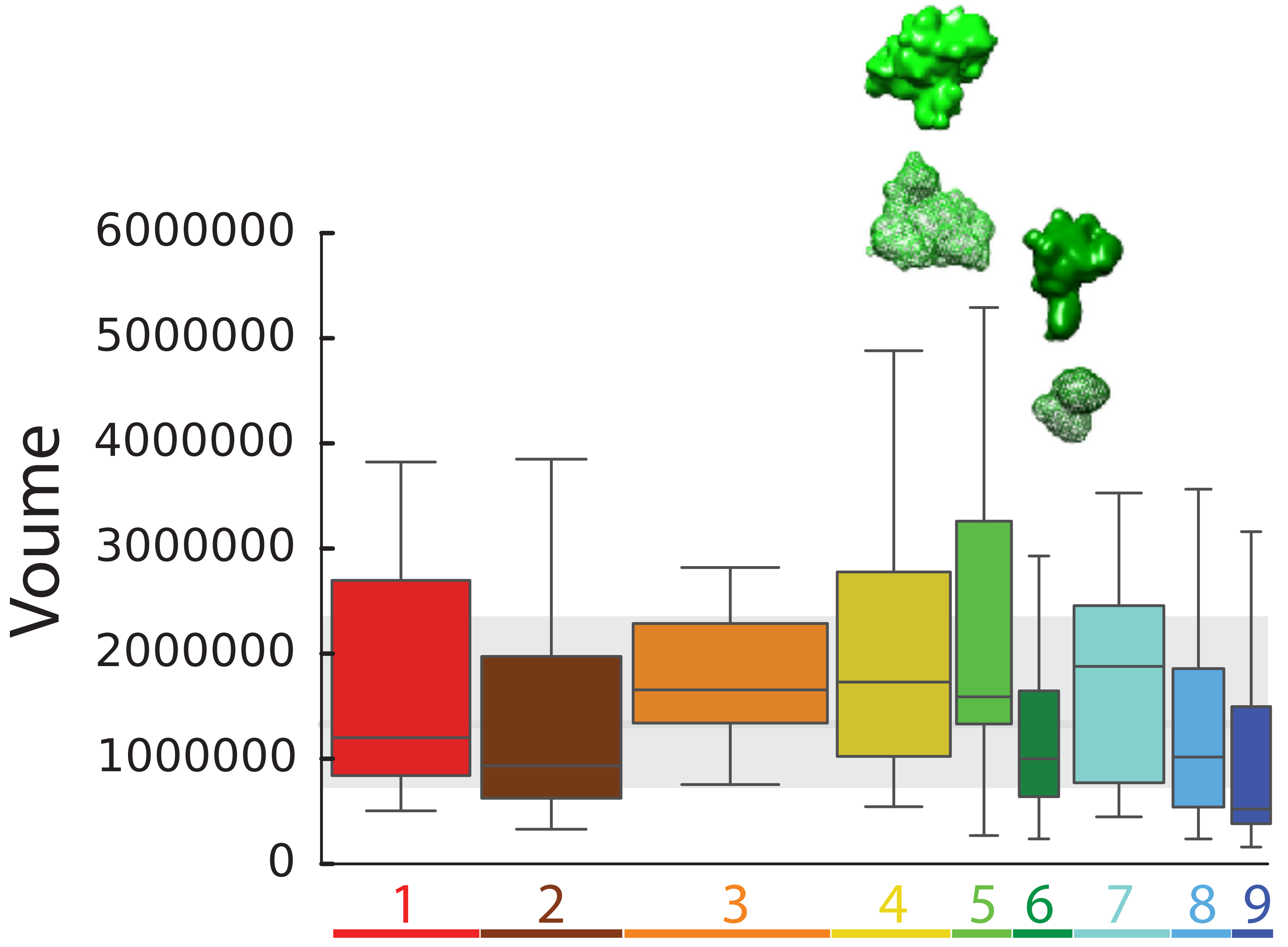
Area (nm^2)
Volume (nm^3)
Sphericity
Overlap (%)
Distance (nm)

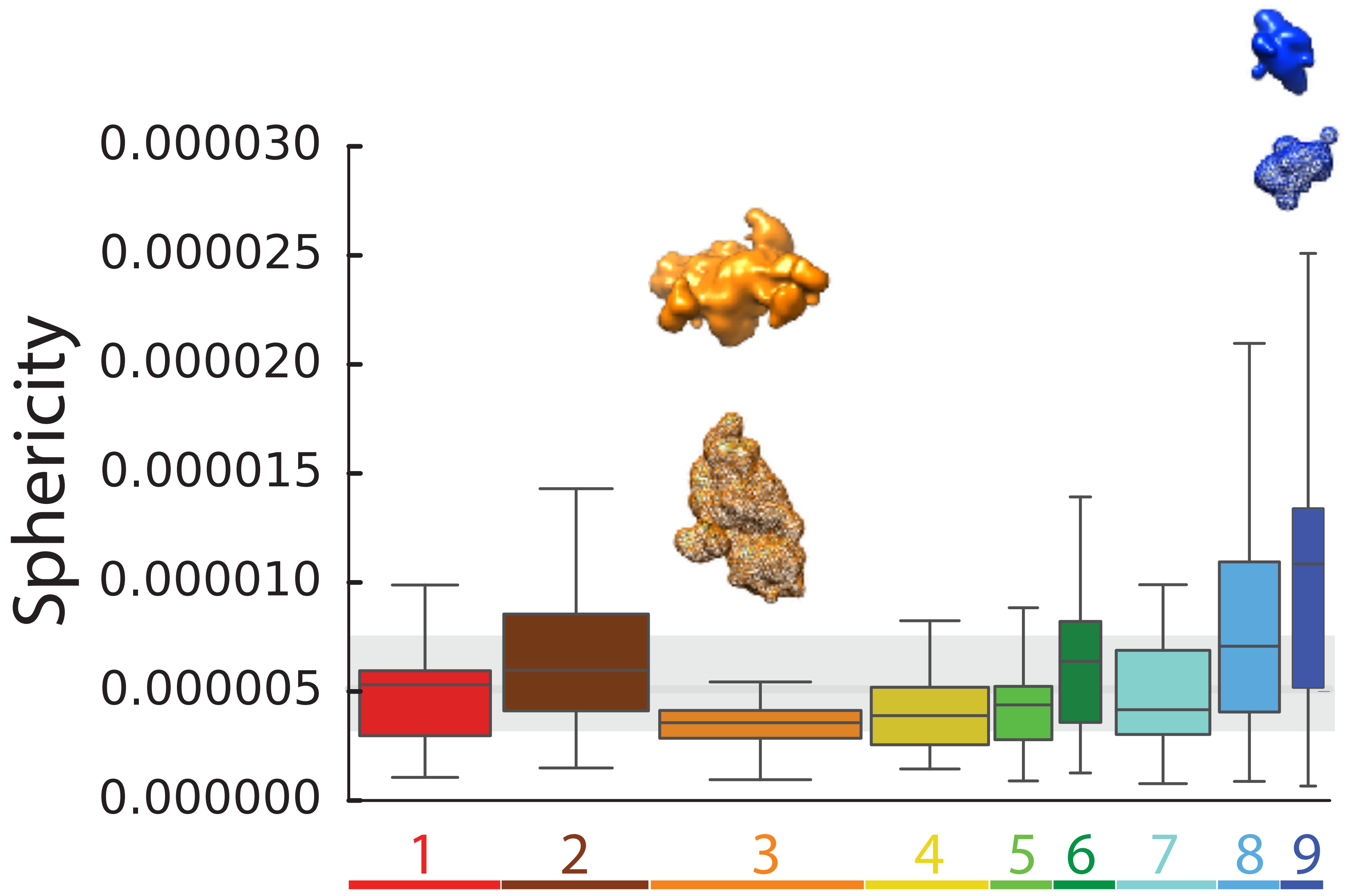
Farabella et al, J Appl Crystallogr. 2015

Structural features

Area, Volume and Sphericity of 19 cells each with 2 homologous resolved

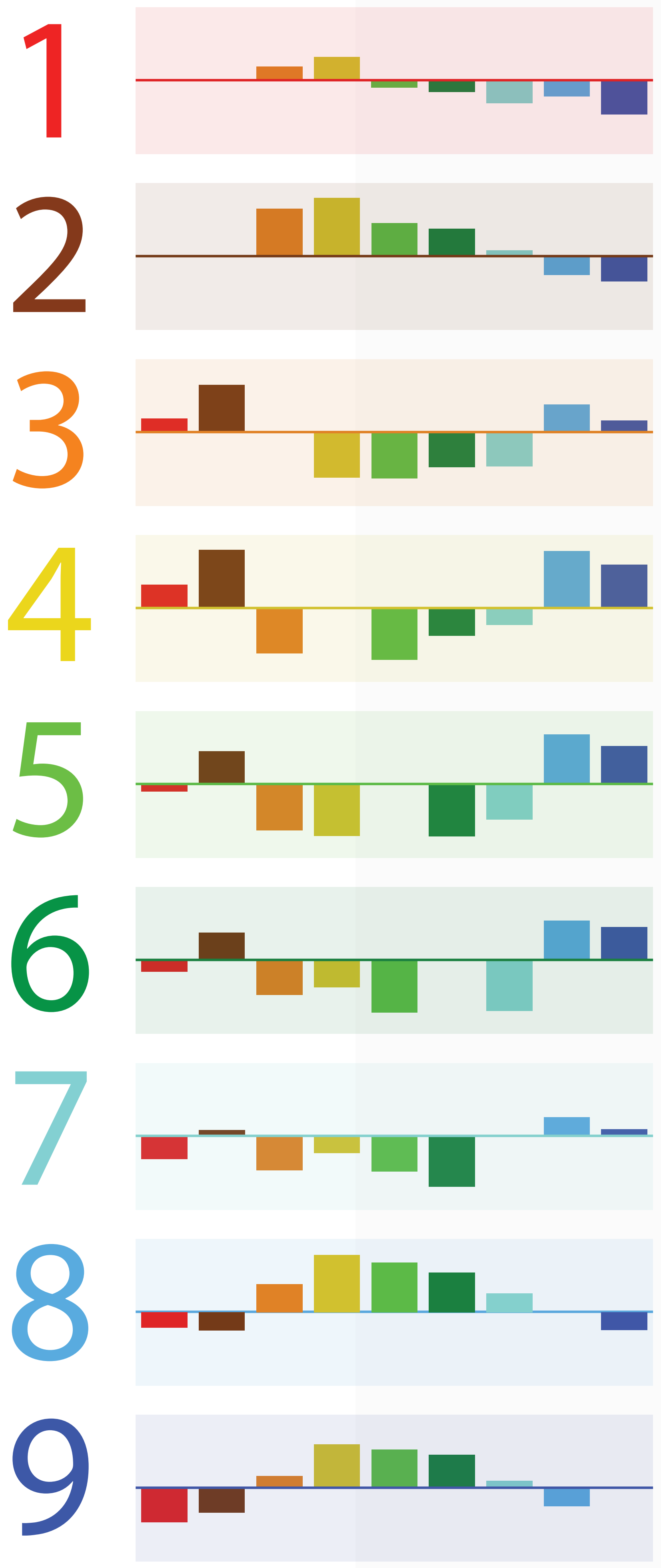


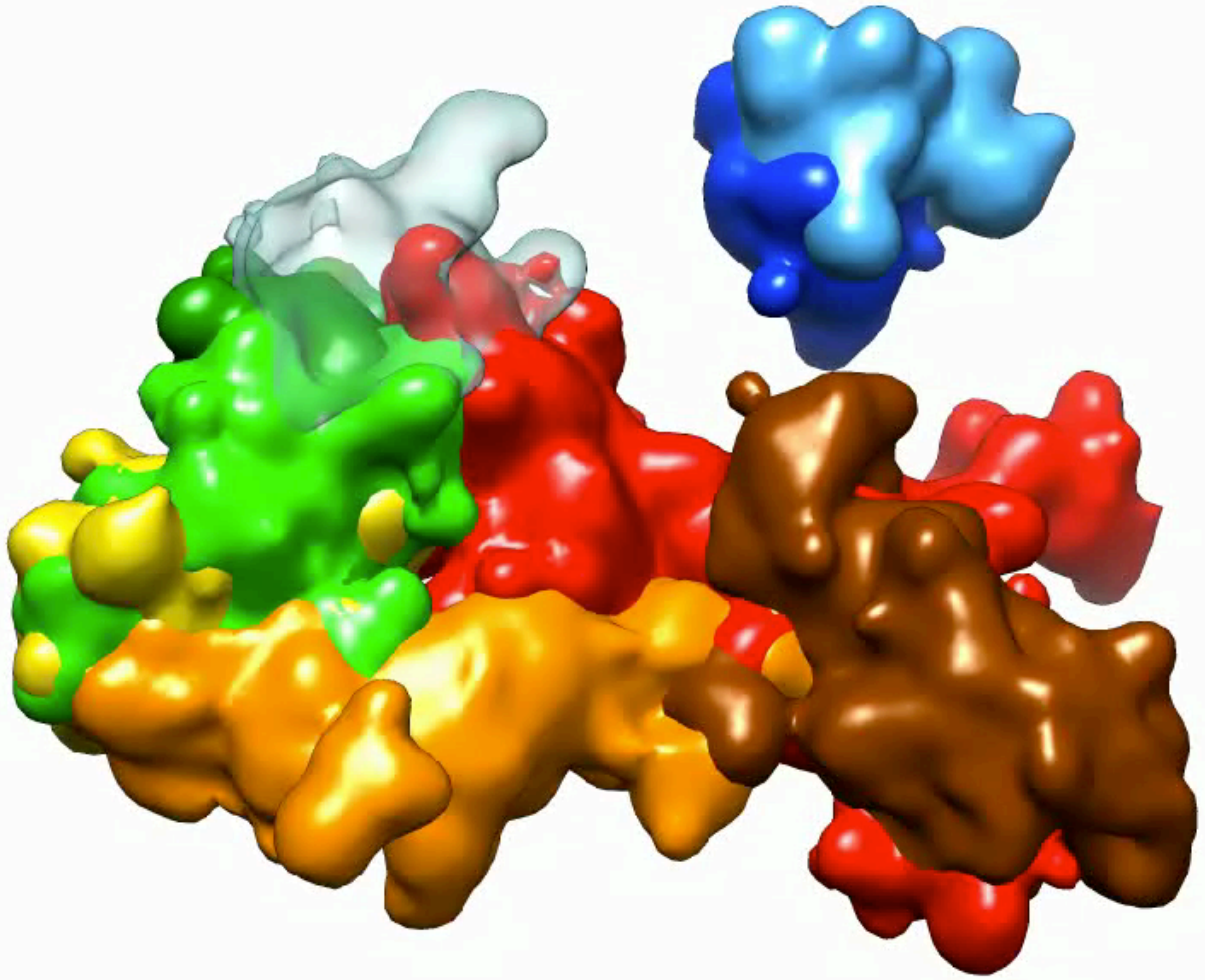


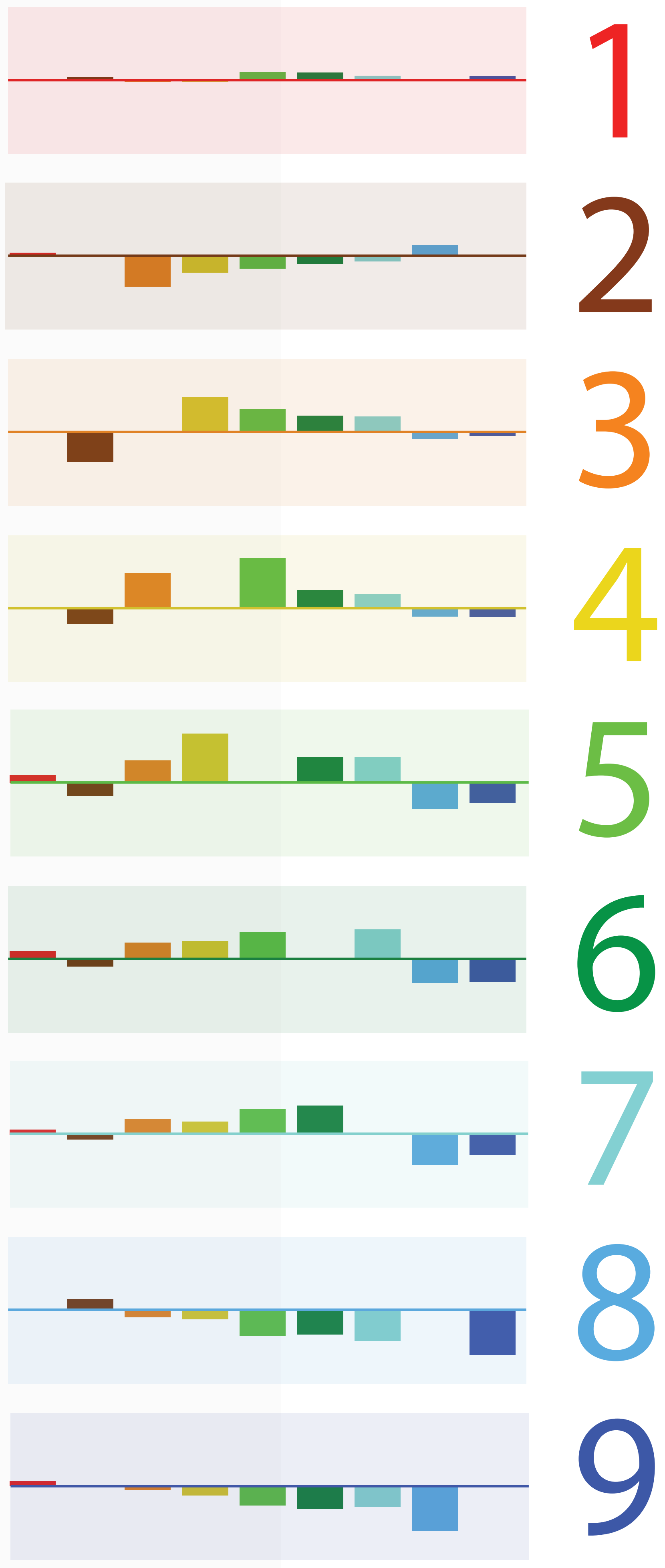


Spatial arrangement

Distance and overlap of 19 cells each with 2 homologous resolved

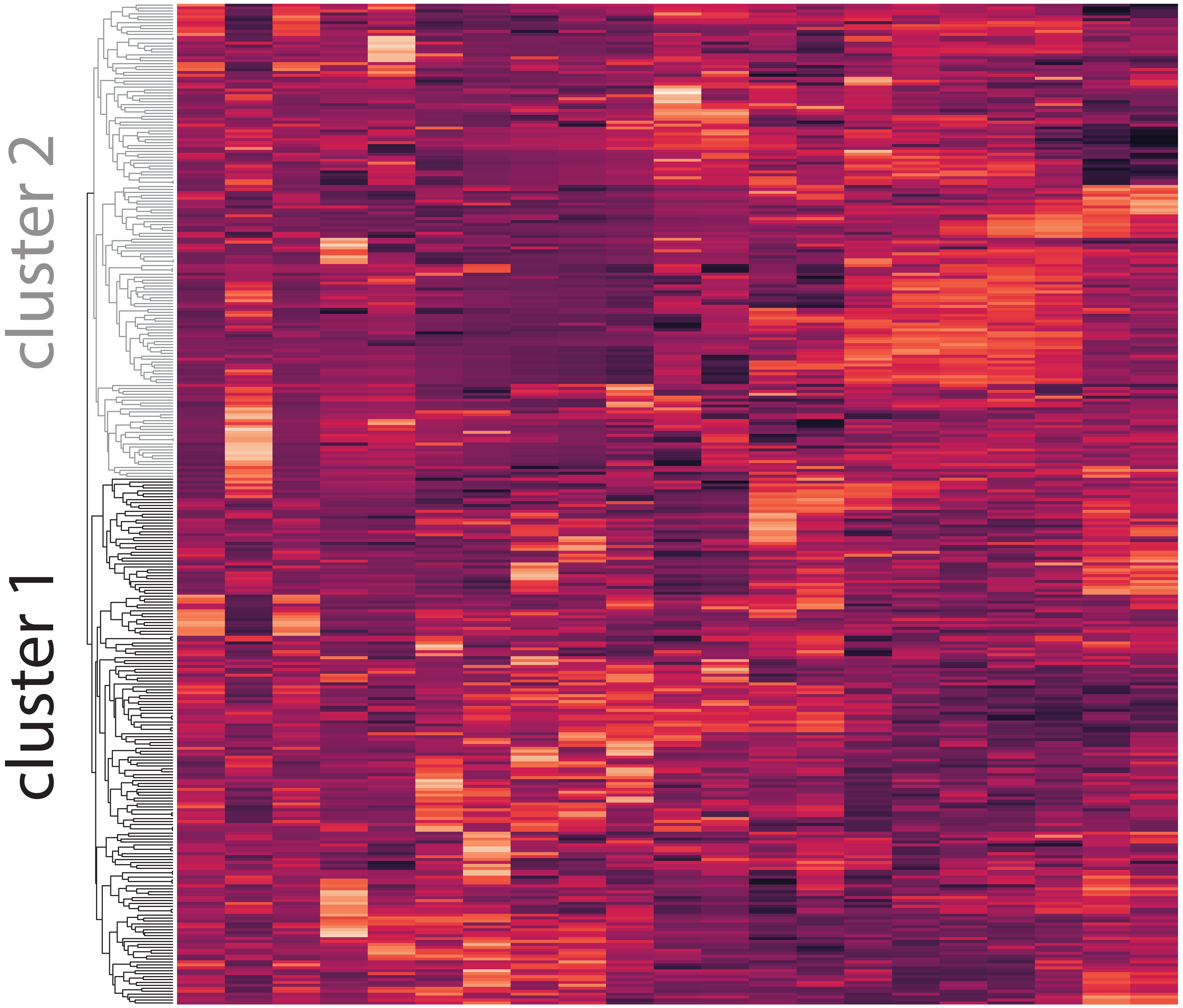

Diff. distance

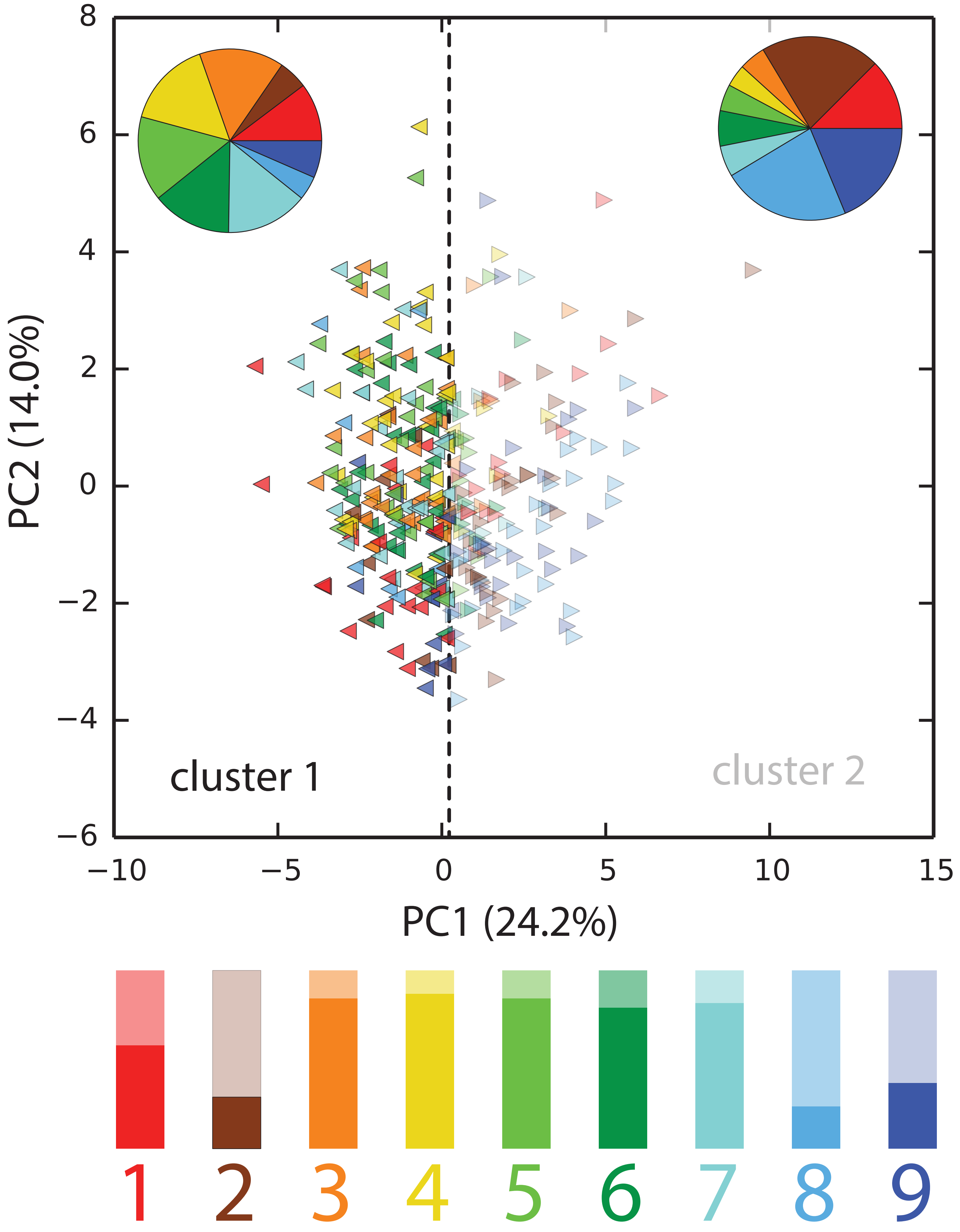



Diff. overlap

Structural clustering

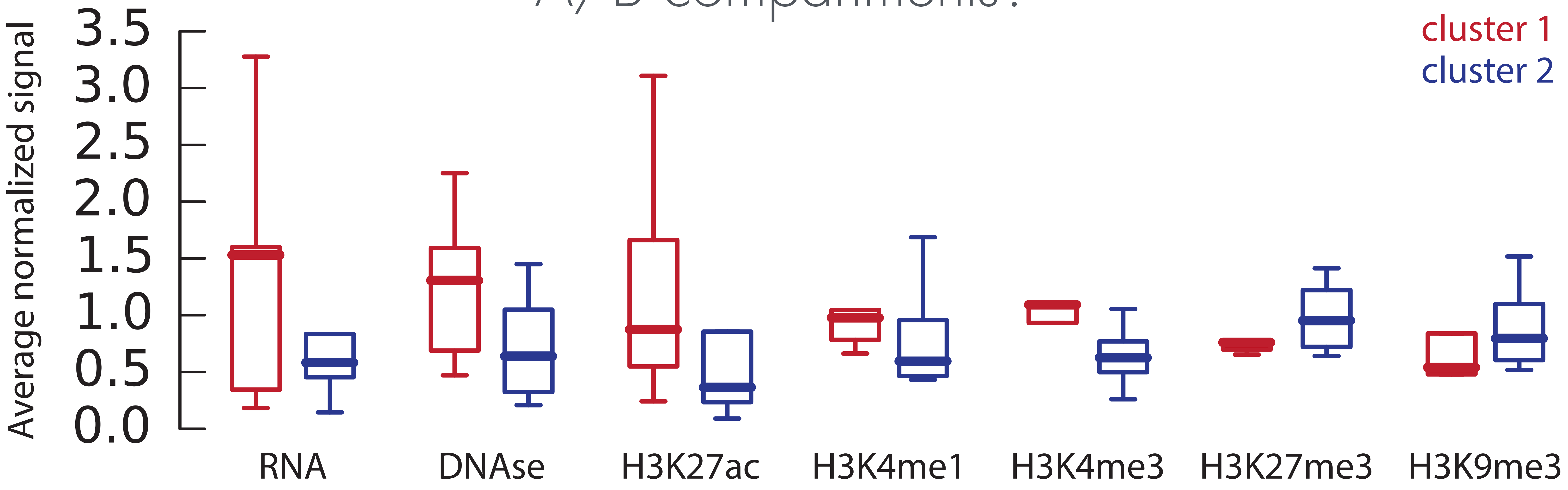
19 cells each with 2 homologous and 9 segments each (342)

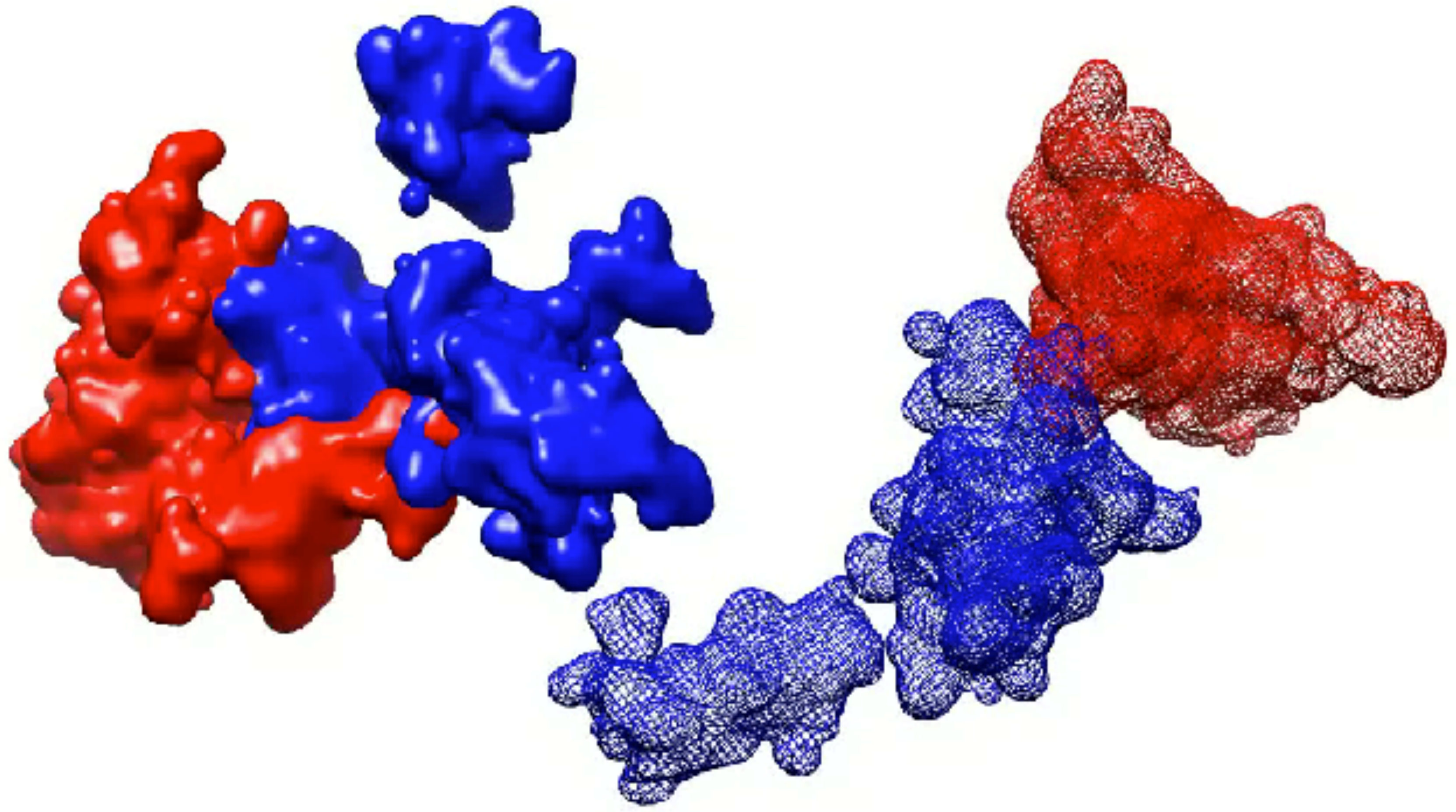


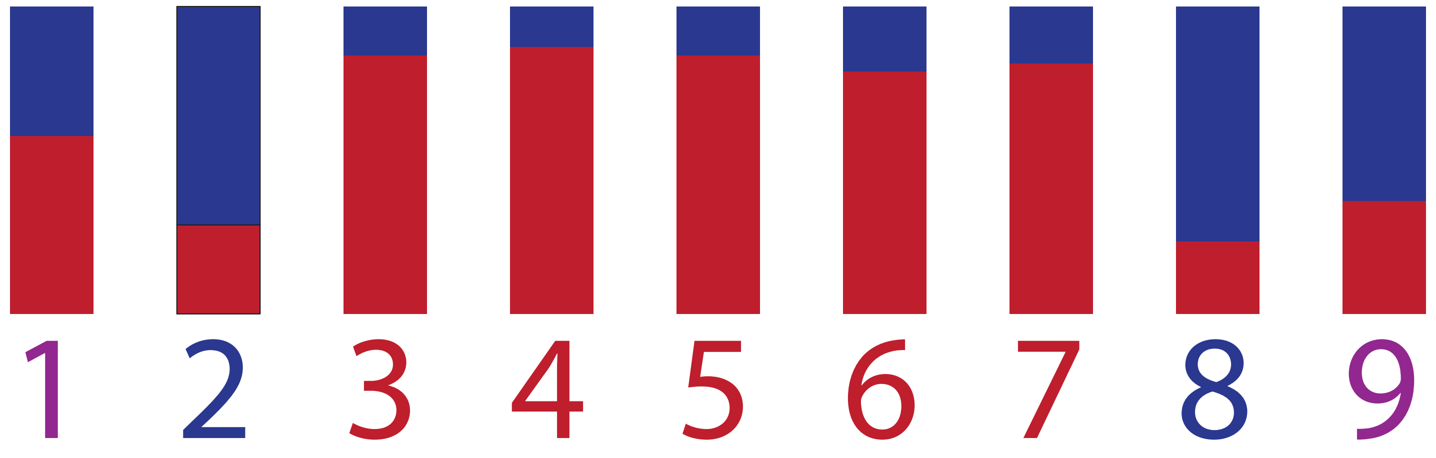


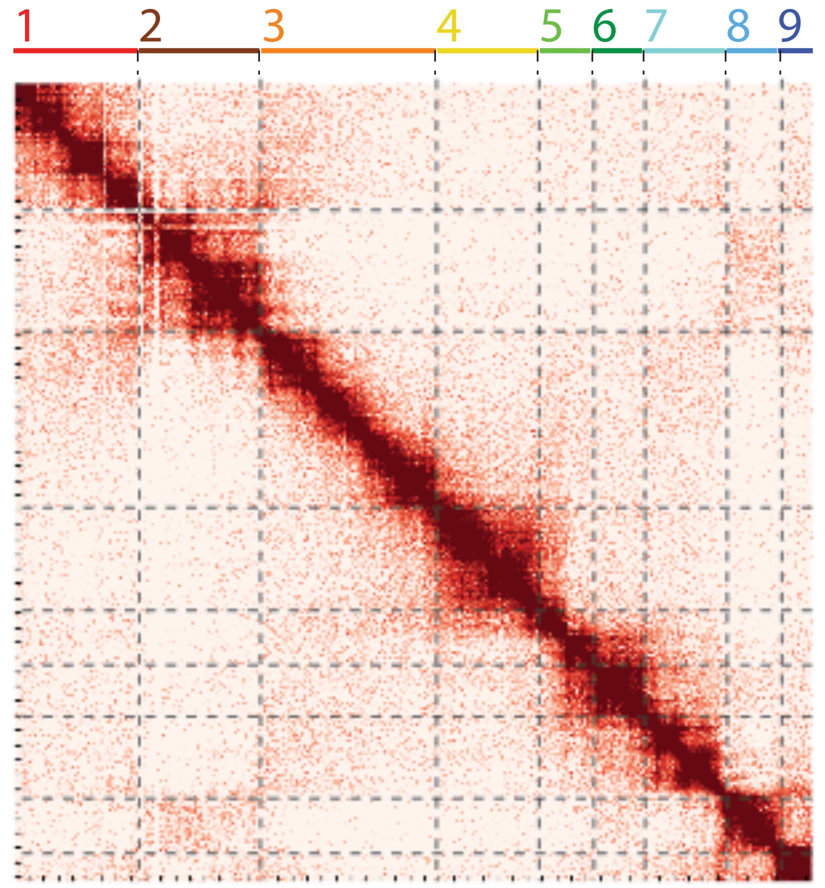
Cluster properties

A/B compartments?





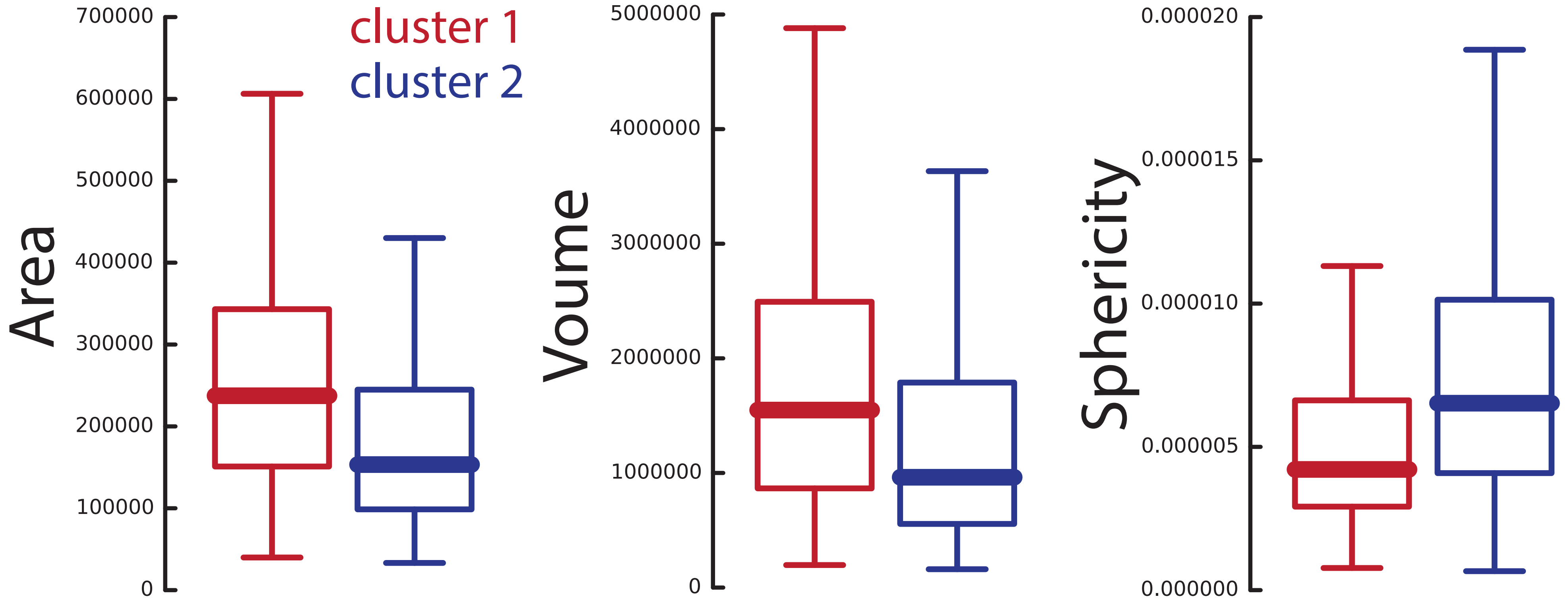


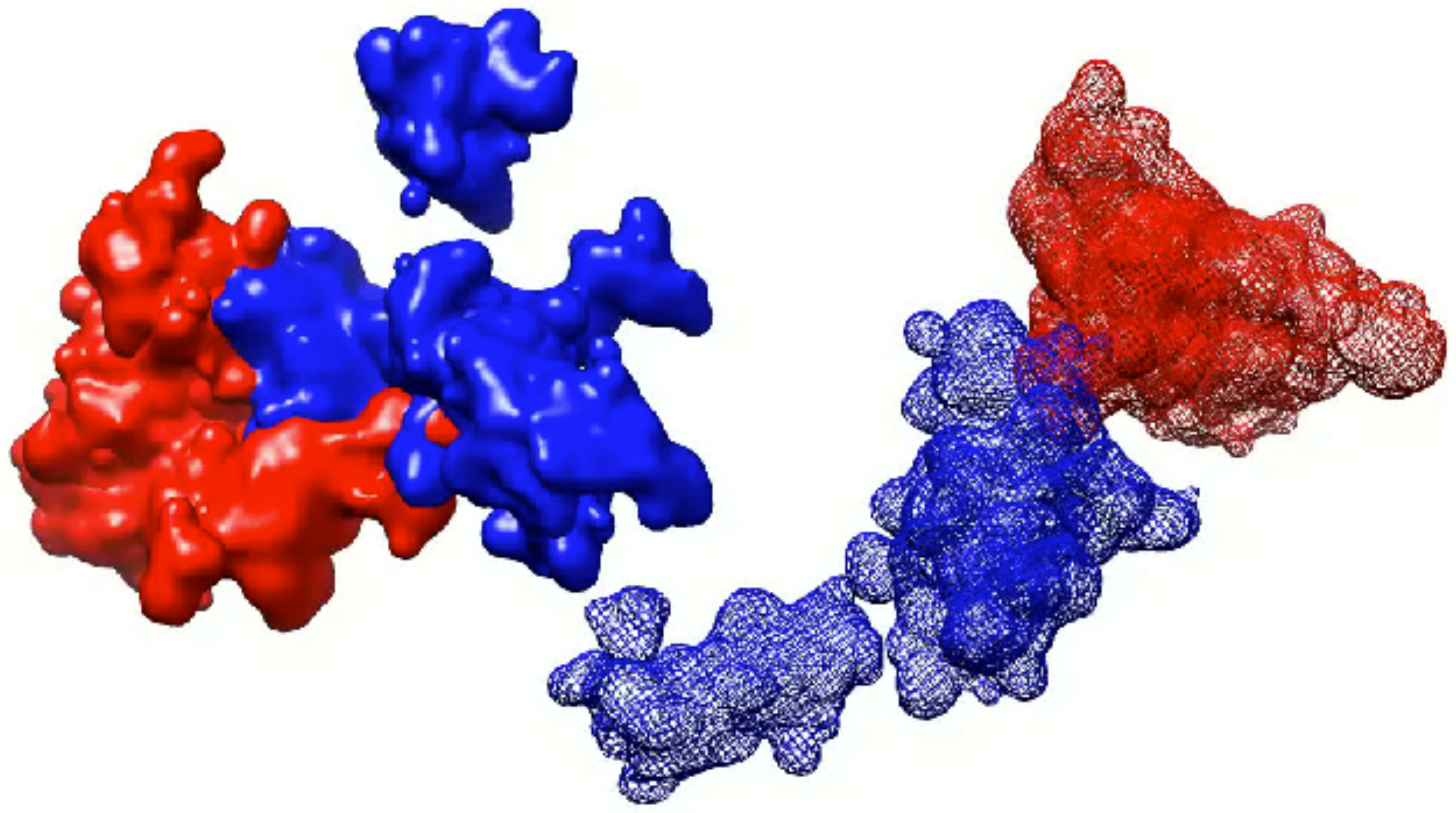


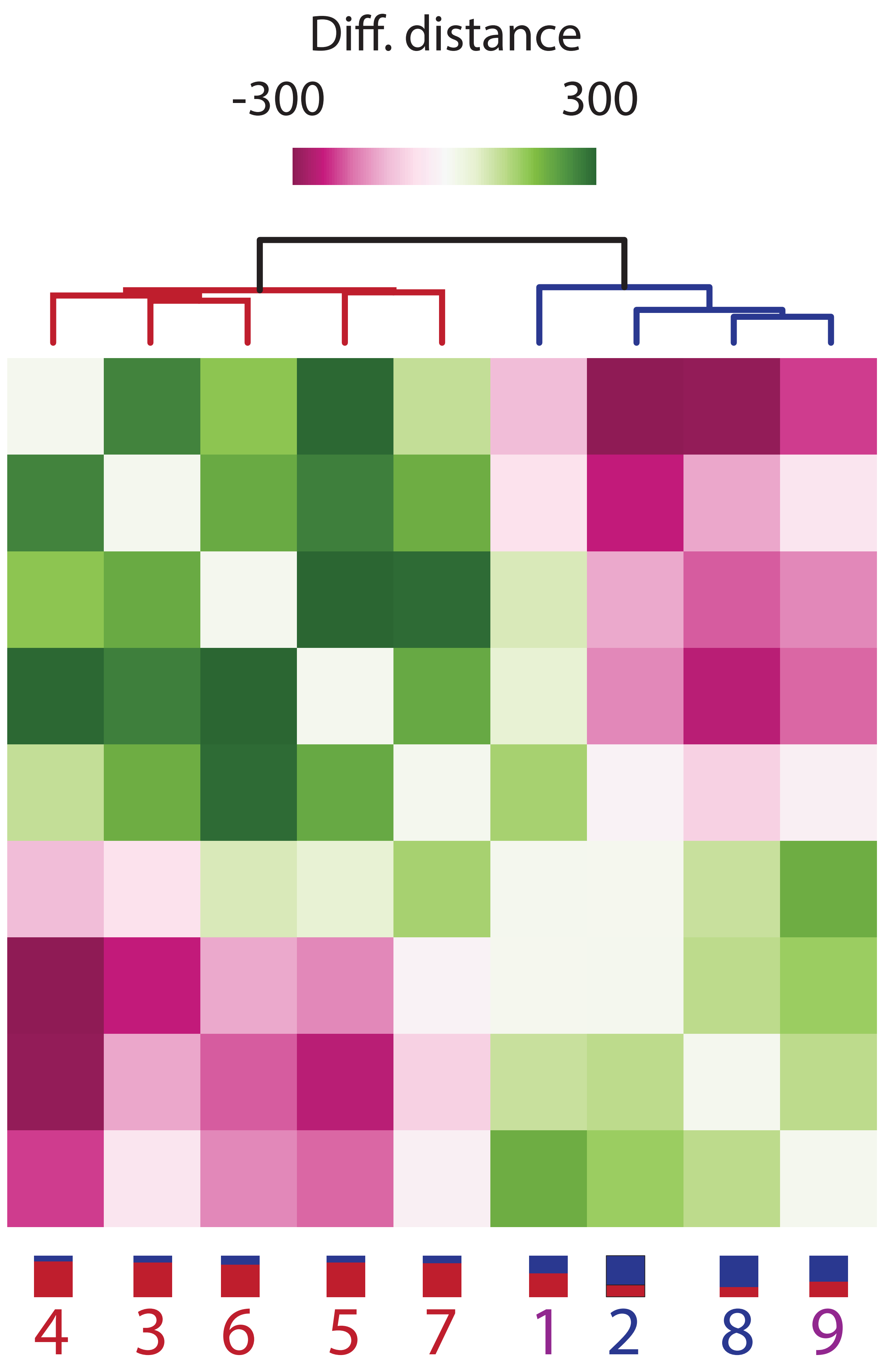
PGP1 ChIP-seq and Hi-C data from ENCODE and Lieberman-Aiden Lab, respectively

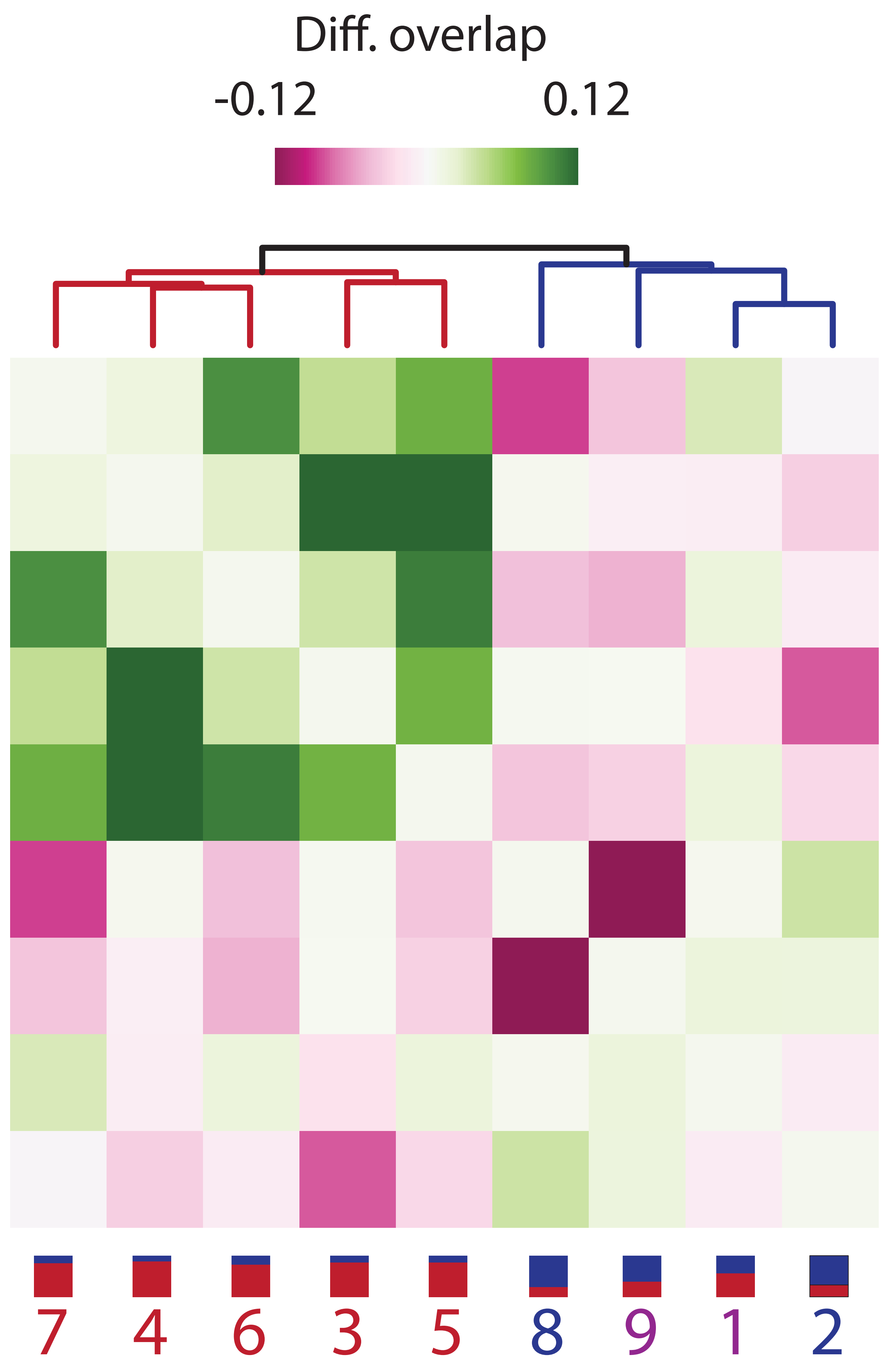
Cluster properties

A/B compartment properties









Can we walk the chromatin path in the nucleus?

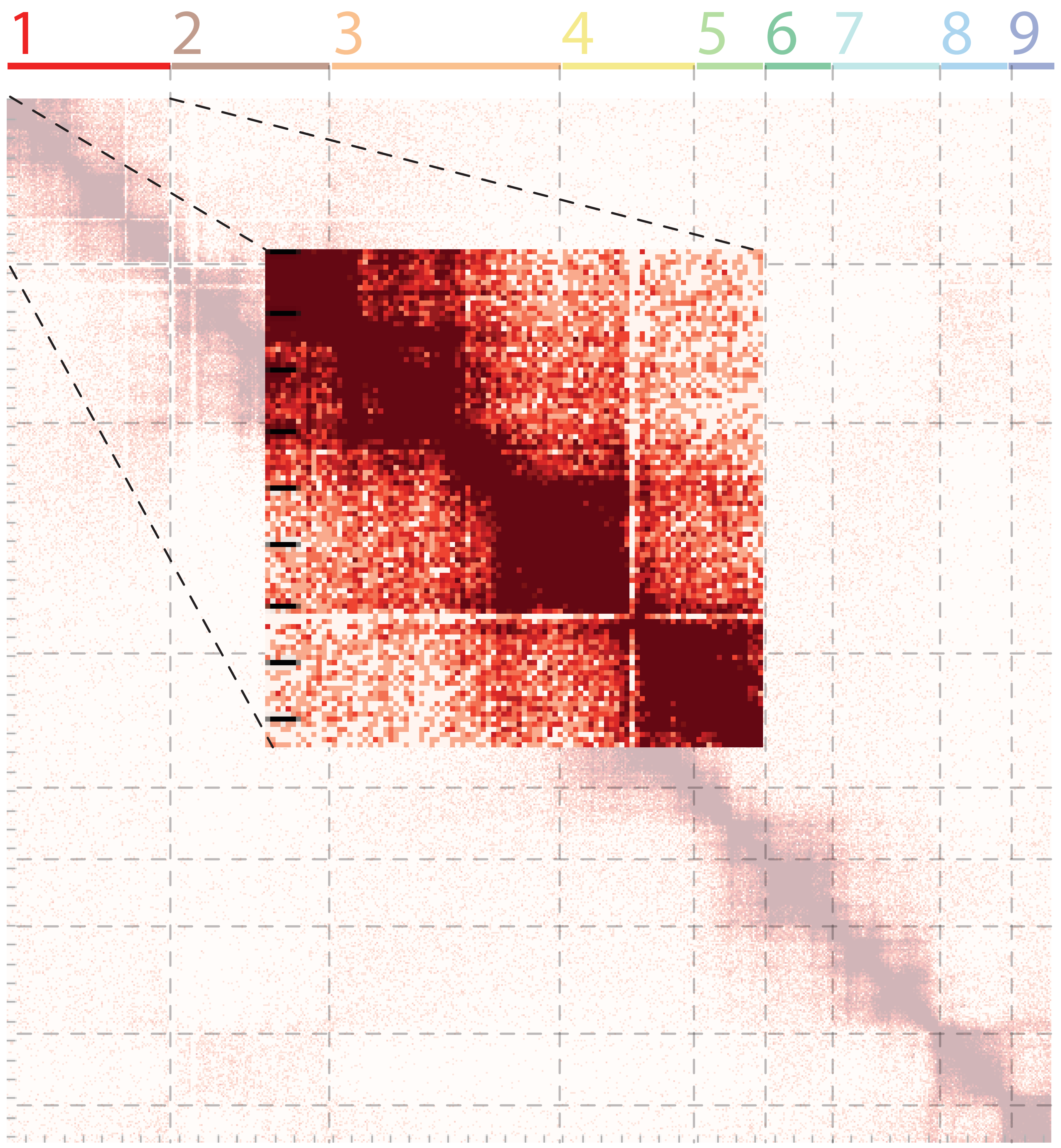
YES!

Can we increase the resolution of our data?

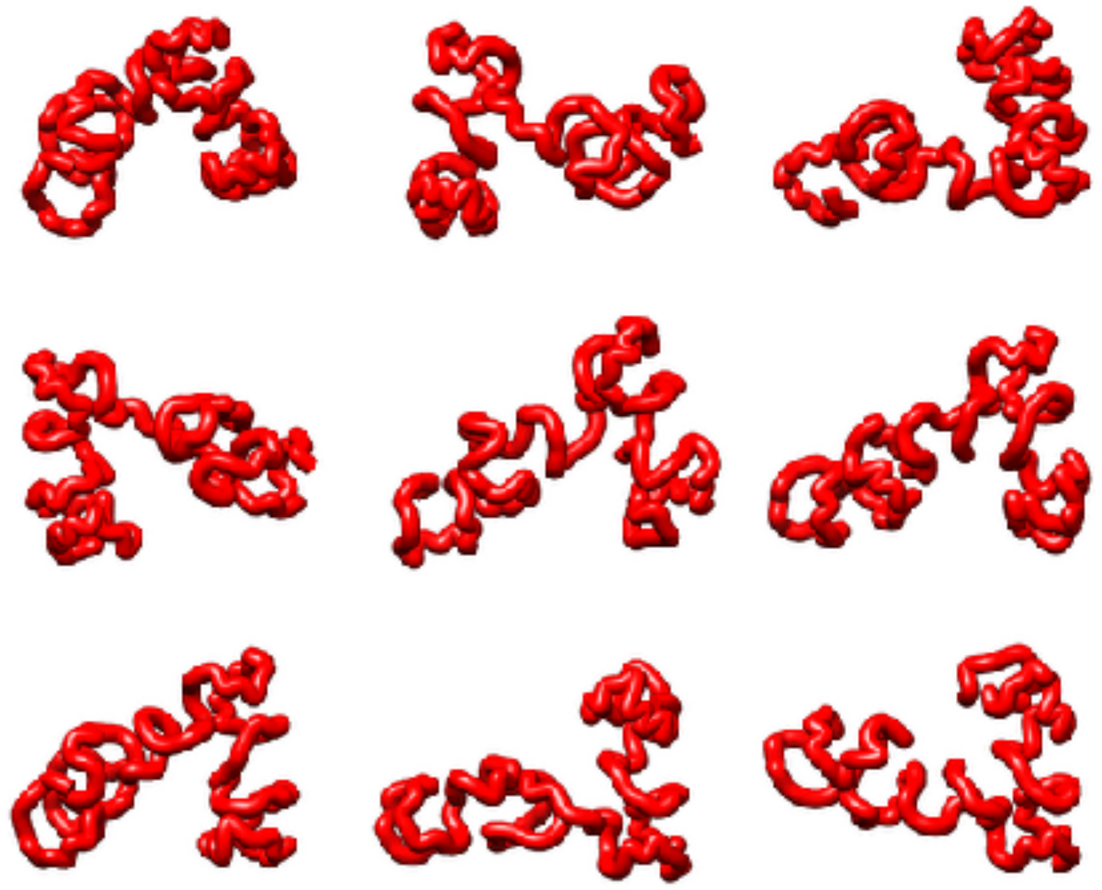
by fitting 3D models based on Hi-C interaction maps

Increasing resolution

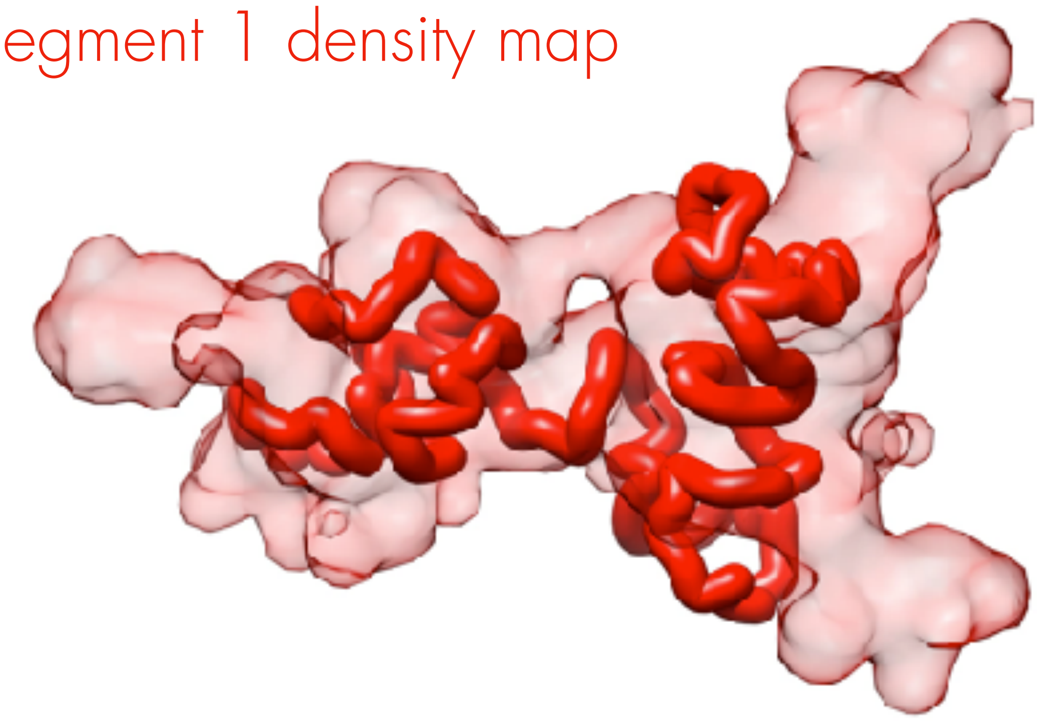
Rigid body fitting 3D structures based on Hi-C data



Segment 1 3D models

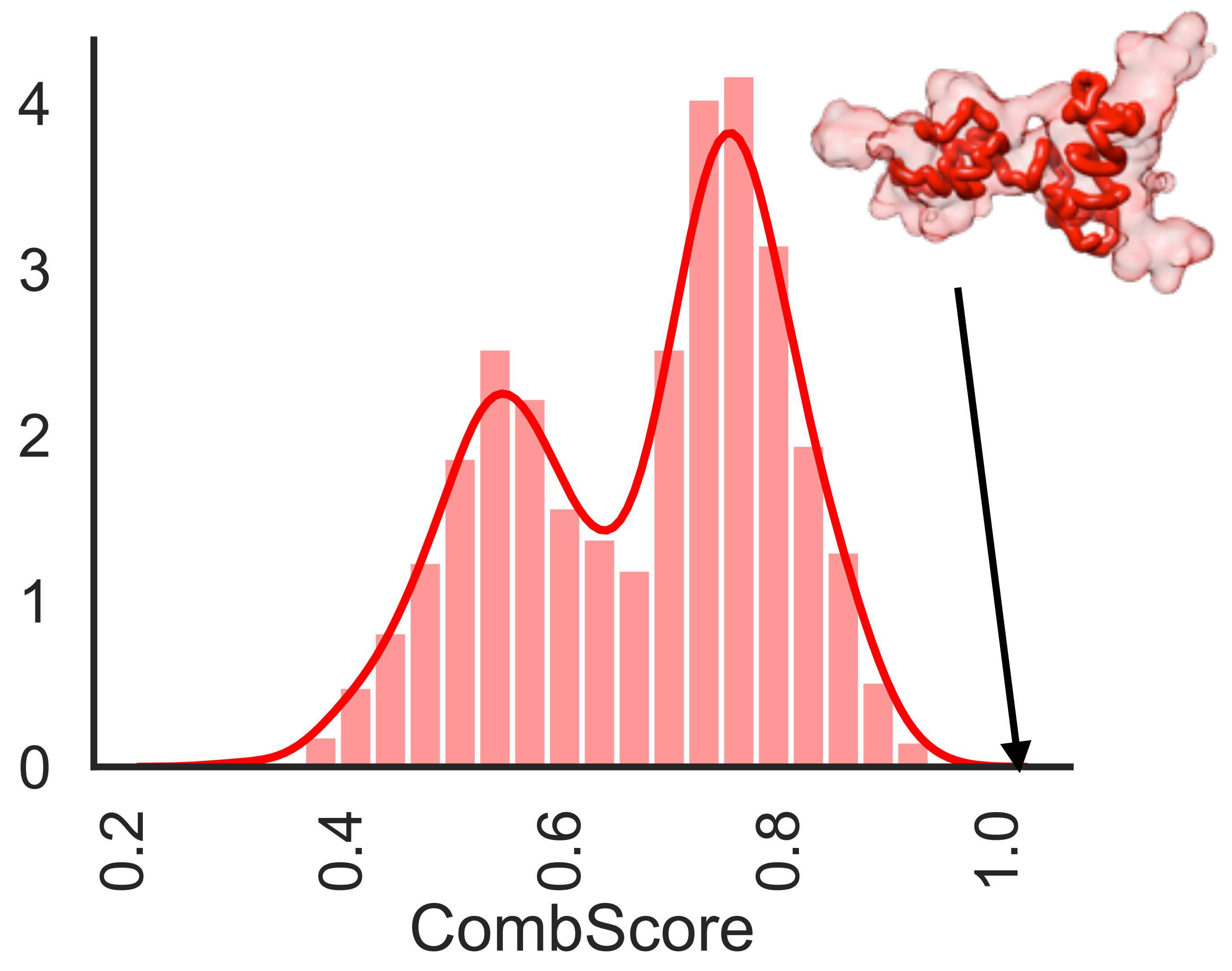


Segment 1 density map



$$\mathrm{ConS} = 1 - \frac{d_{P,COM}}{\max\left(d_{P,COM}\right)}$$

$$CCC = \frac{\sum_{i=1}^{M}\left[\rho_i^{EM} - \bar{\rho}^{EM}\right]\left[\rho_i^{P} - \bar{\rho}^{P}\right]}{\sqrt{\sum_{i=1}^{M}\left[\rho_i^{EM} - \bar{\rho}^{EM}\right]^2 \sum_{i=1}^{M}\left[\rho_i^{P} - \bar{\rho}^{P}\right]^2}}$$



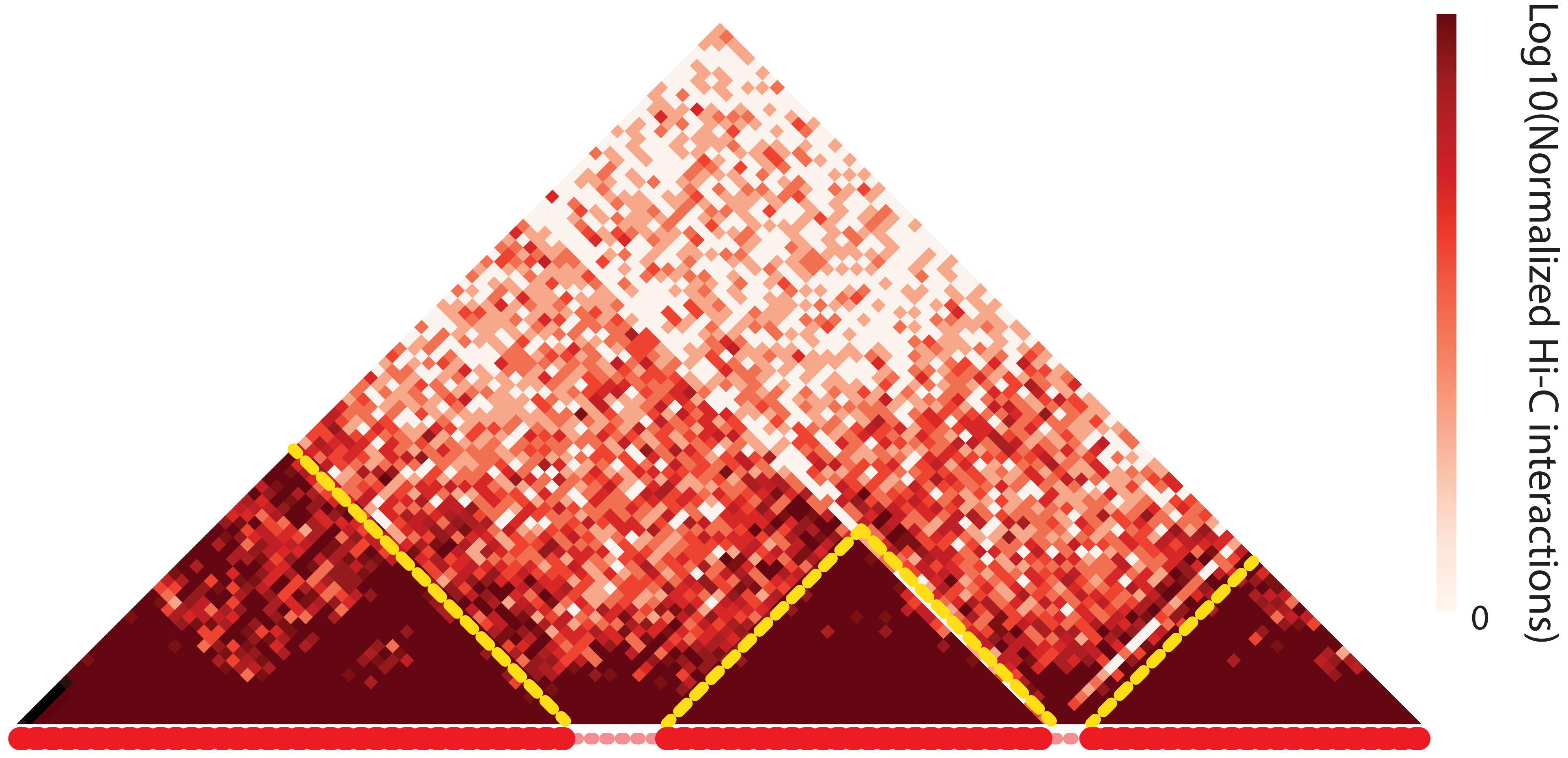
TADbit

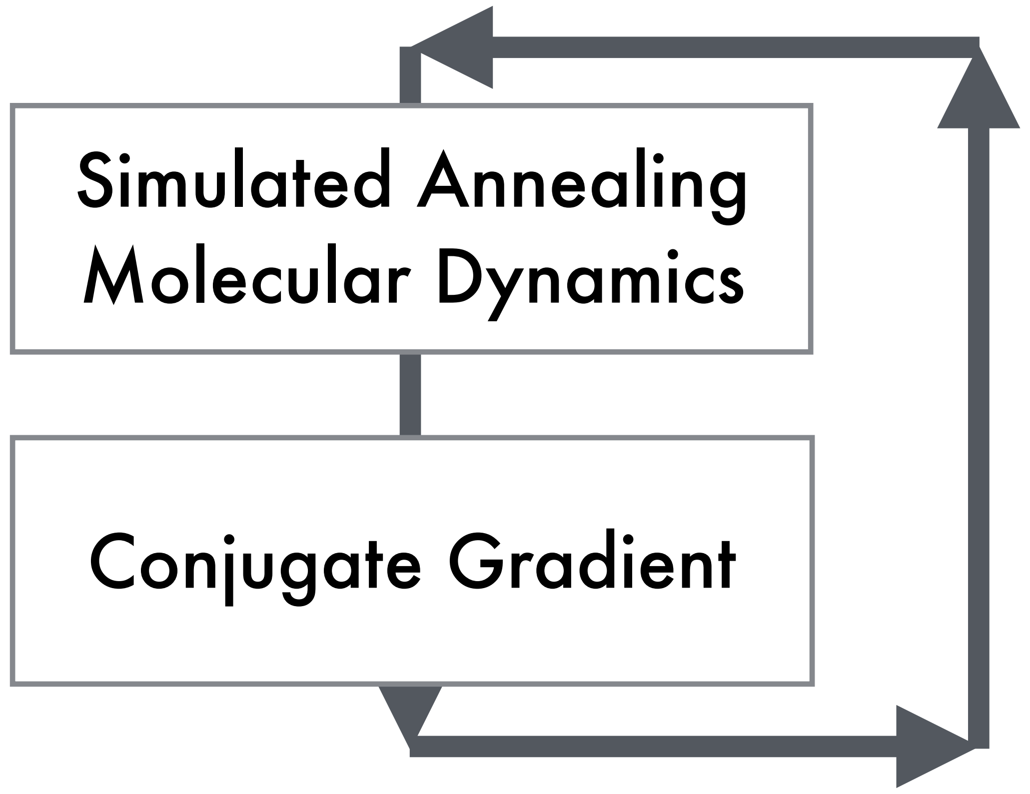
Serra, Baù, et al. PLOS CB 2017
http://www.3DGenomes.org

Farabella et al, J Appl Crystallogr. 2015
Roseman, 2000; Wriggers & Chacon, Structure 2001

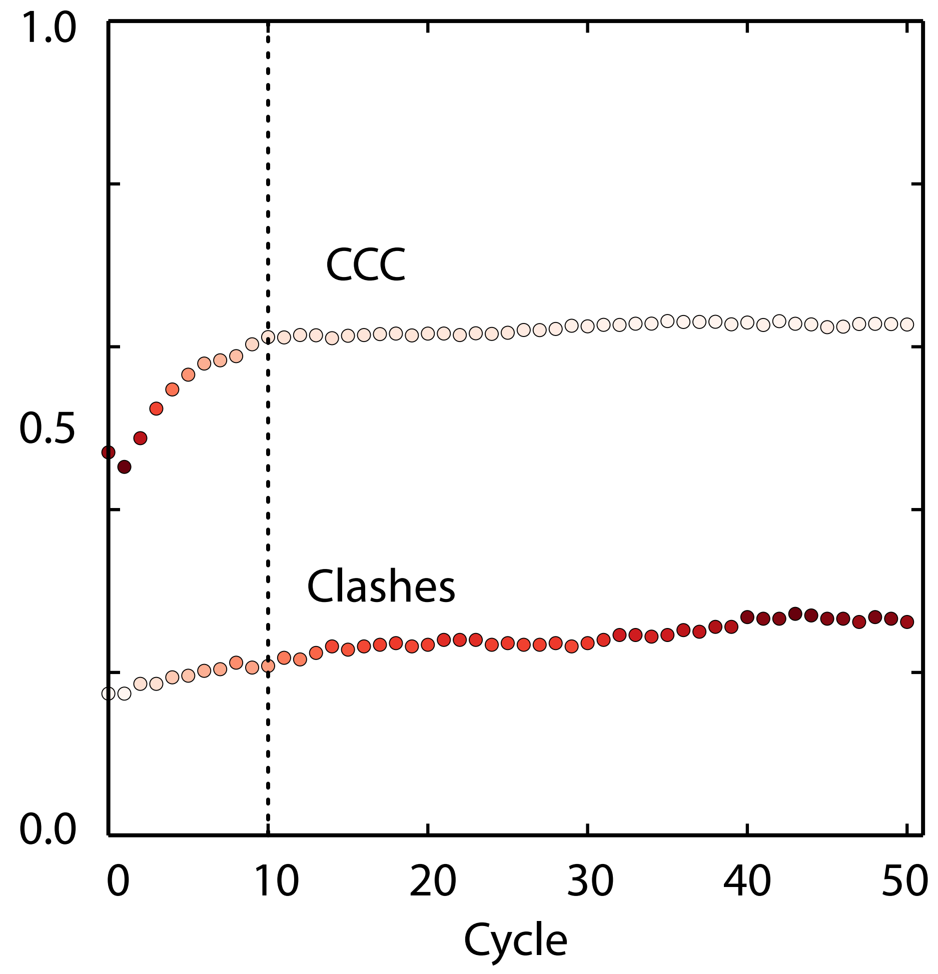
Increasing resolution

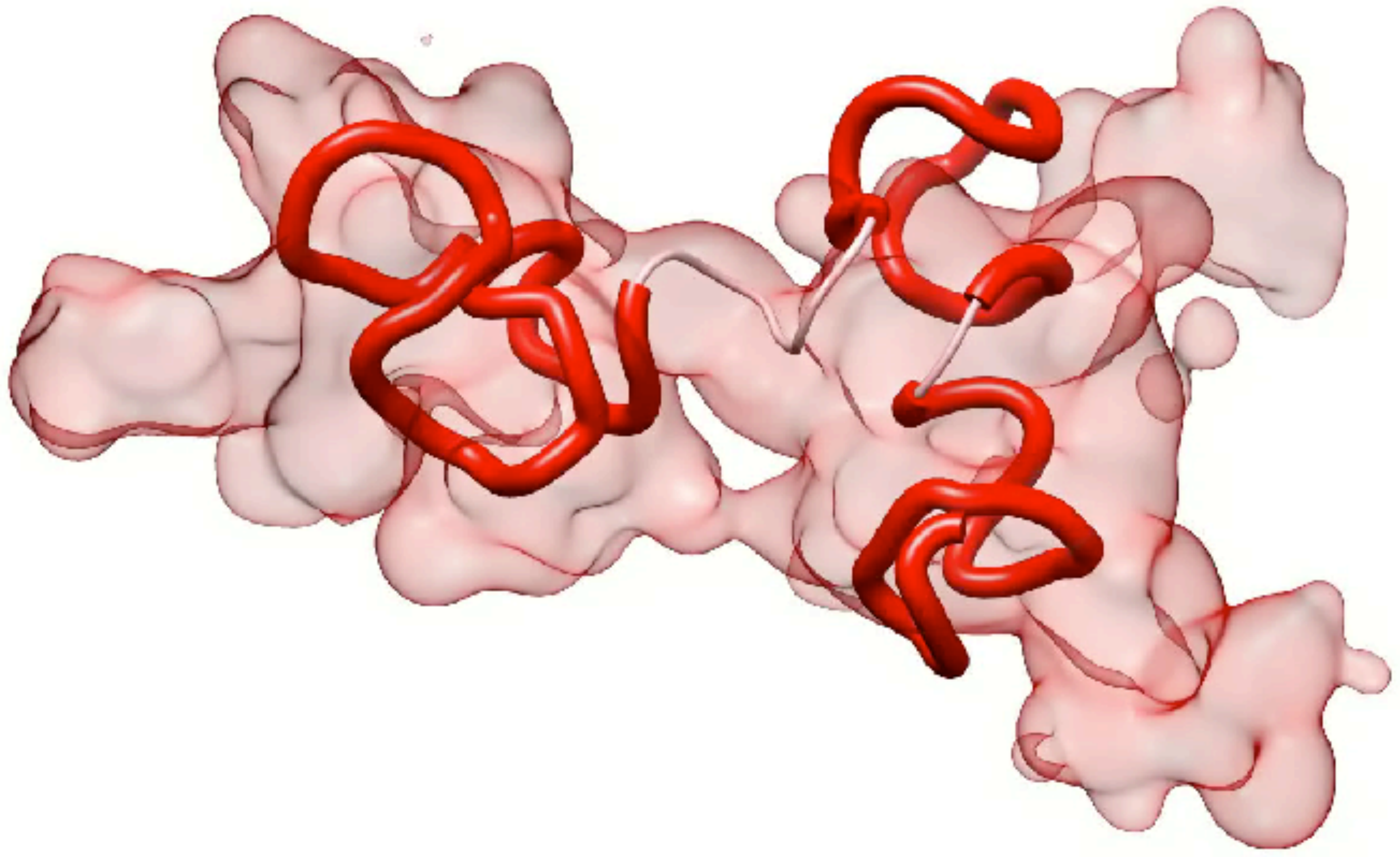
Flexible fitting 3D structures based on Hi-C data




Simulated Annealing
Molecular Dynamics

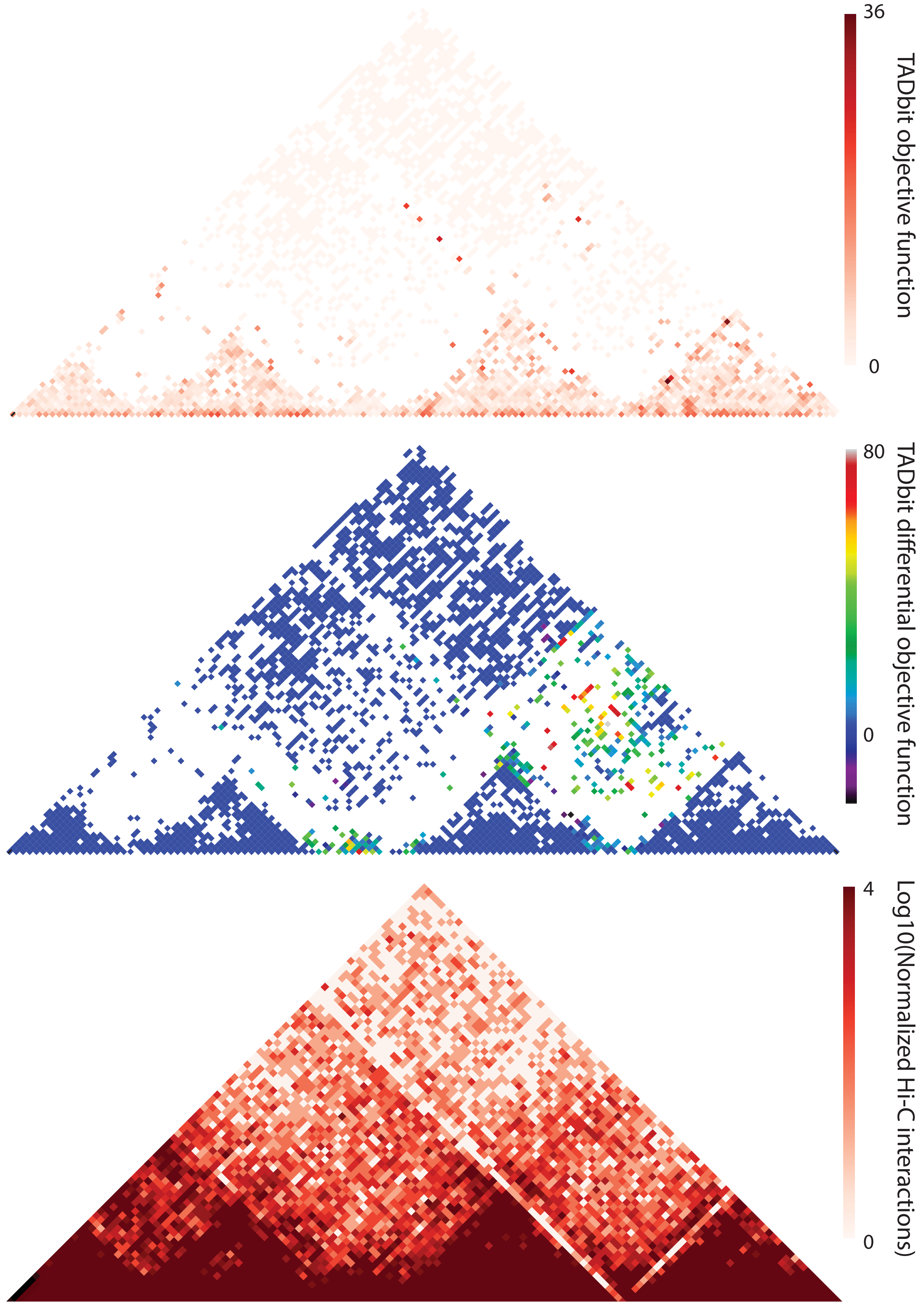
Conjugate Gradient



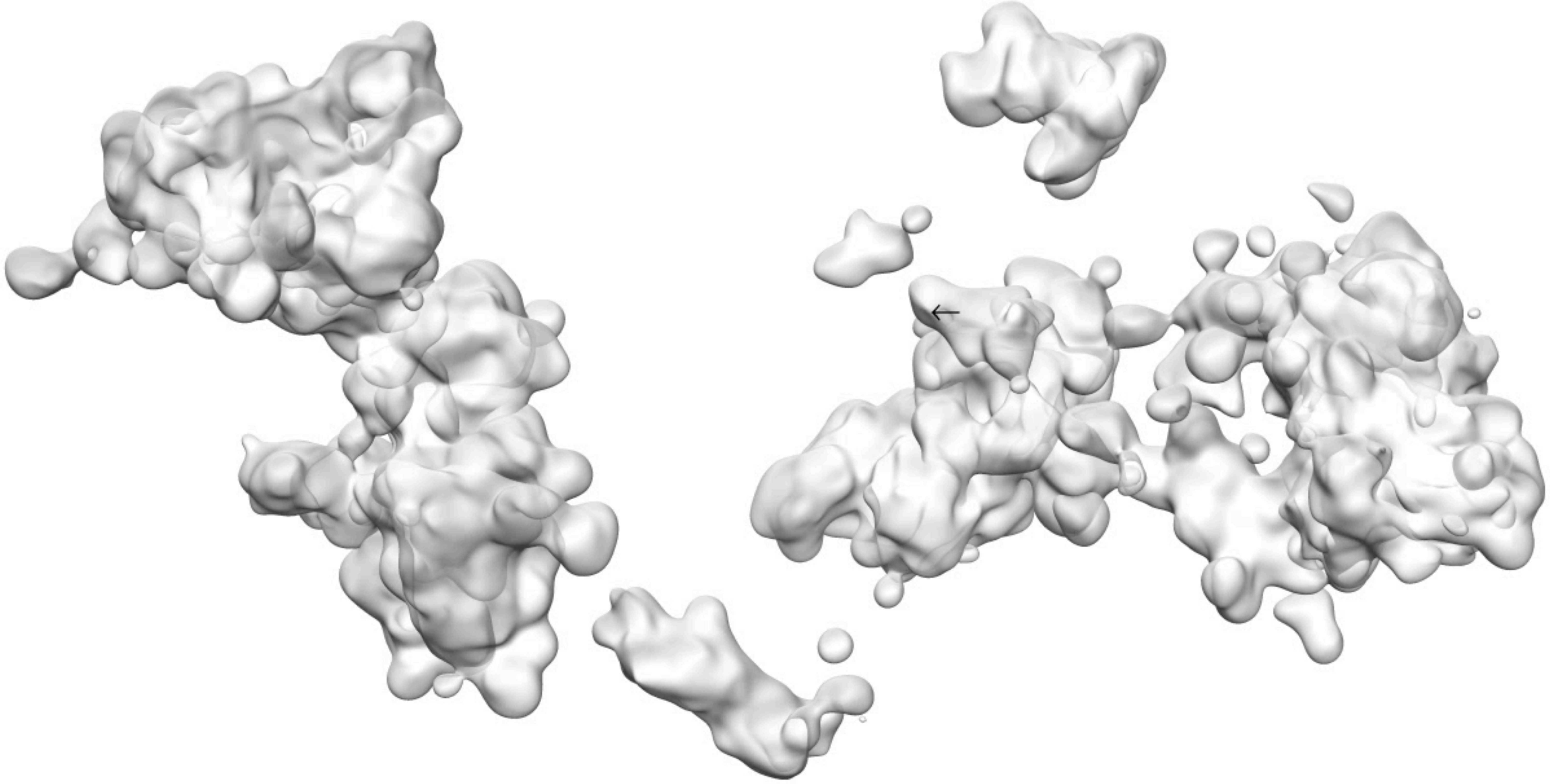


Increasing resolution

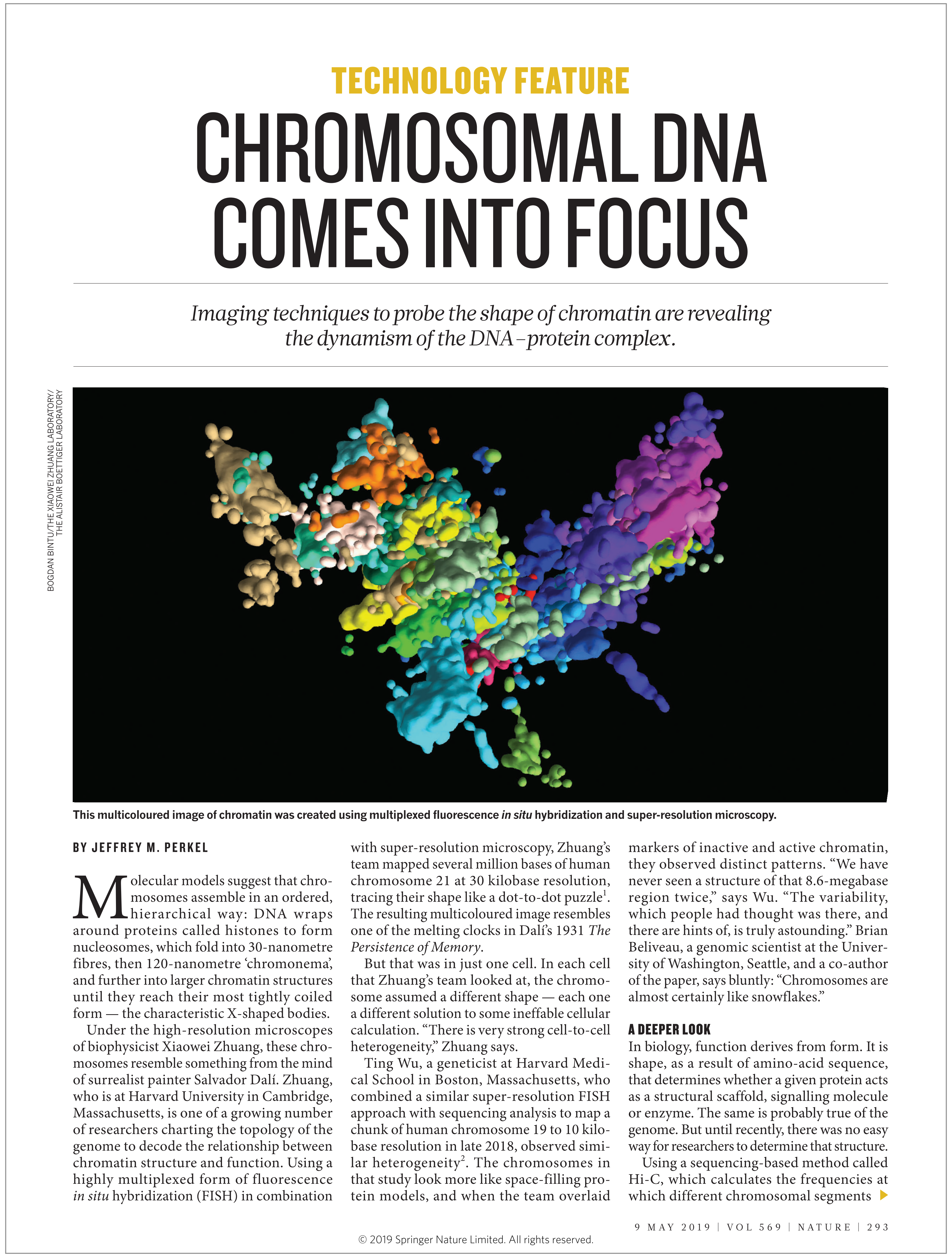
Flexible fitting 3D structures based on Hi-C data



Chromosome walking path @10Kb resolution



Jeffrey M. Perkel Nature 569, 293-294 (2019)



TECHNOLOGY FEATURE

CHROMOSOMAL DNA COMES INTO FOCUS

Imaging techniques to probe the shape of chromatin are revealing the dynamism of the DNA–protein complex.

BOGDAN BINTU/THE XIAOWEI ZHUANG LABORATORY/
THE ALISTAIR BOETTIGER LABORATORY

This multicoloured image of chromatin was created using multiplexed fluorescence *in situ* hybridization and super-resolution microscopy.

BY JEFFREY M. PERKEL

Molecular models suggest that chromosomes assemble in an ordered, hierarchical way: DNA wraps around proteins called histones to form nucleosomes, which fold into 30-nanometre fibres, then 120-nanometre ‘chromonema’, and further into larger chromatin structures until they reach their most tightly coiled form — the characteristic X-shaped bodies.

Under the high-resolution microscopes of biophysicist Xiaowei Zhuang, these chromosomes resemble something from the mind of surrealist painter Salvador Dalí. Zhuang, who is at Harvard University in Cambridge, Massachusetts, is one of a growing number of researchers charting the topology of the genome to decode the relationship between chromatin structure and function. Using a highly multiplexed form of fluorescence *in situ* hybridization (FISH) in combination with super-resolution microscopy, Zhuang's team mapped several million bases of human chromosome 21 at 30 kilobase resolution, tracing their shape like a dot-to-dot puzzle[1]. The resulting multicoloured image resembles one of the melting clocks in Dalí's 1931 *The Persistence of Memory*.

But that was in just one cell. In each cell that Zhuang's team looked at, the chromosome assumed a different shape — each one a different solution to some ineffable cellular calculation. “There is very strong cell-to-cell heterogeneity,” Zhuang says.

Ting Wu, a geneticist at Harvard Medical School in Boston, Massachusetts, who combined a similar super-resolution FISH approach with sequencing analysis to map a chunk of human chromosome 19 to 10 kilobase resolution in late 2018, observed similar heterogeneity[2]. The chromosomes in that study look more like space-filling protein models, and when the team overlaid markers of inactive and active chromatin, they observed distinct patterns. “We have never seen a structure of that 8.6-megabase region twice,” says Wu. “The variability, which people had thought was there, and there are hints of, is truly astounding.” Brian Beliveau, a genomic scientist at the University of Washington, Seattle, and a co-author of the paper, says bluntly: “Chromosomes are almost certainly like snowflakes.”

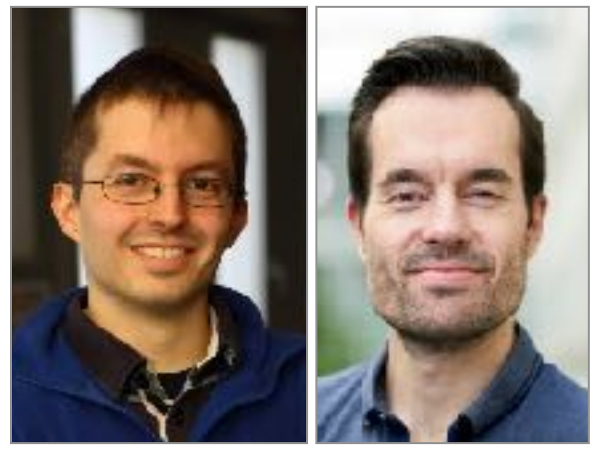
A DEEPER LOOK

In biology, function derives from form. It is shape, as a result of amino-acid sequence, that determines whether a given protein acts as a structural scaffold, signalling molecule or enzyme. The same is probably true of the genome. But until recently, there was no easy way for researchers to determine that structure.

Using a sequencing-based method called Hi-C, which calculates the frequencies at which different chromosomal segments ▶


Photo by David Oliete - www.davidoliete.com

3D structural dynamics of the SOX2 locus activation

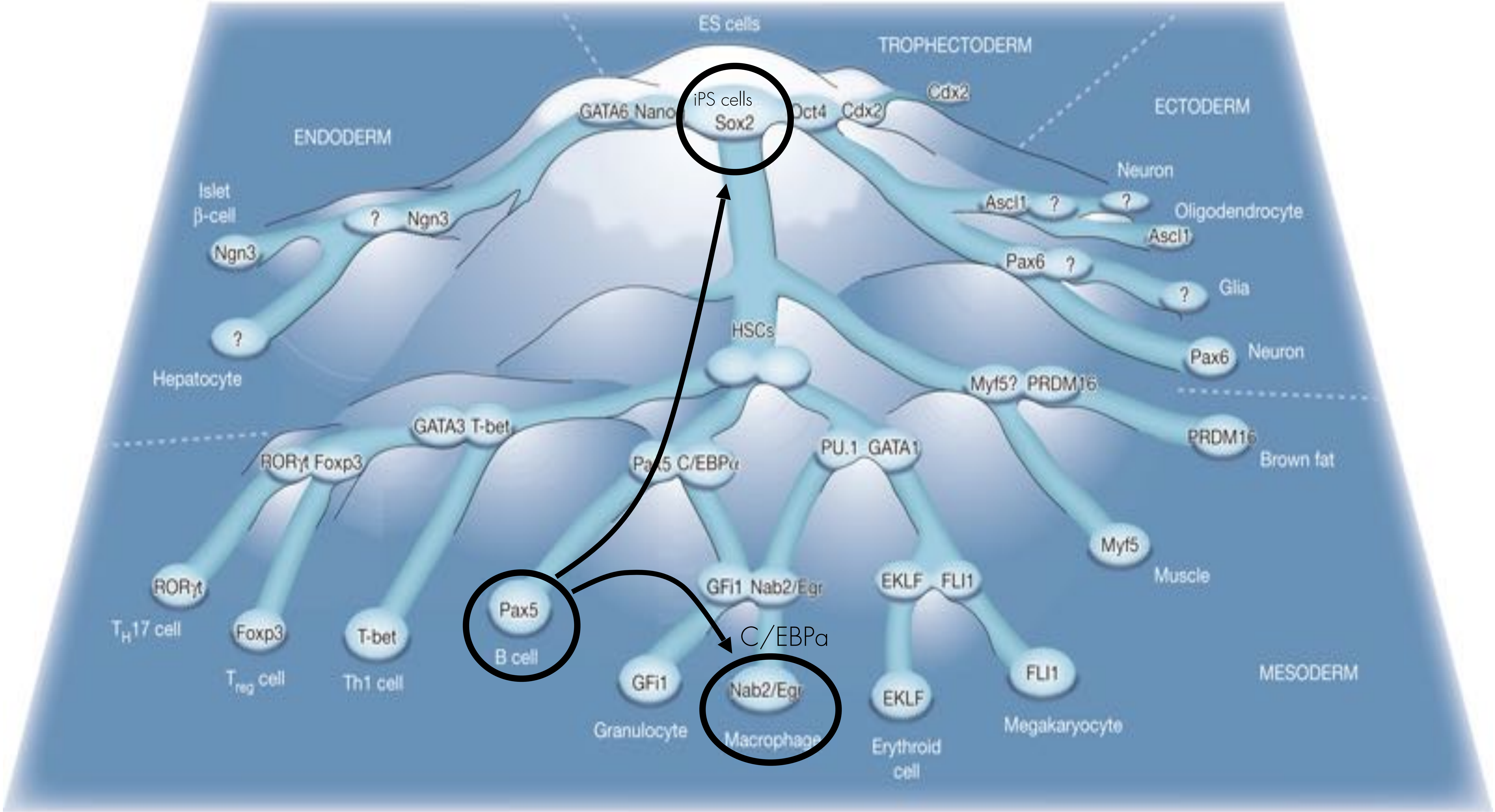


Marco di Stefano
Ralph Stadhouders
with Graf Lab (CRG, Barcelona)

Nature Genetics (2018) 50 238–249 & BioRxived

Transcription factors dictate cell fate

Graf & Enver (2009) Nature



Transcription factors (TFs) determine cell identity through gene regulation

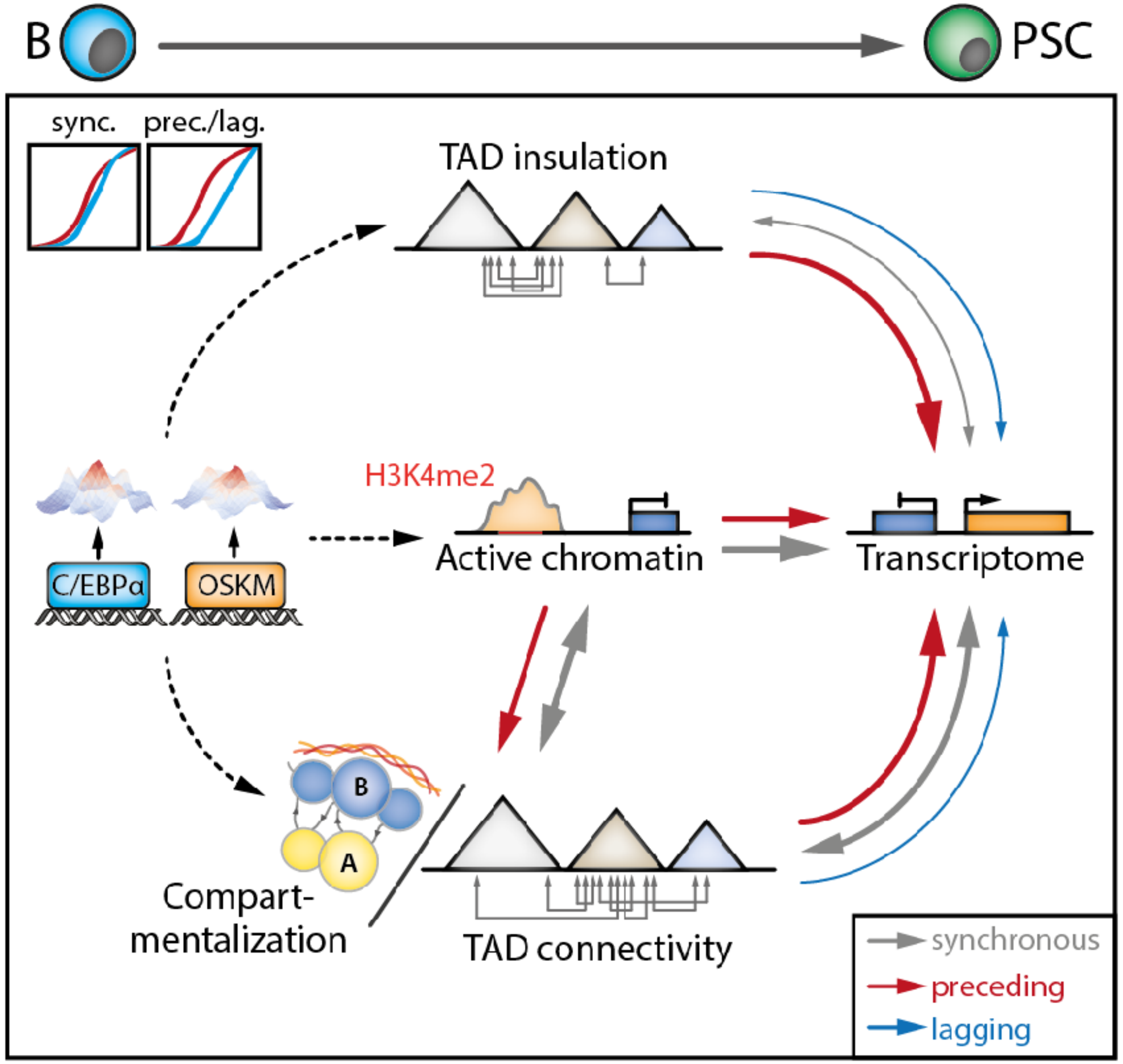
Normal 'forward' differentiation

Cell fates can be converted by enforced TF expression

Transdifferentiation or reprogramming

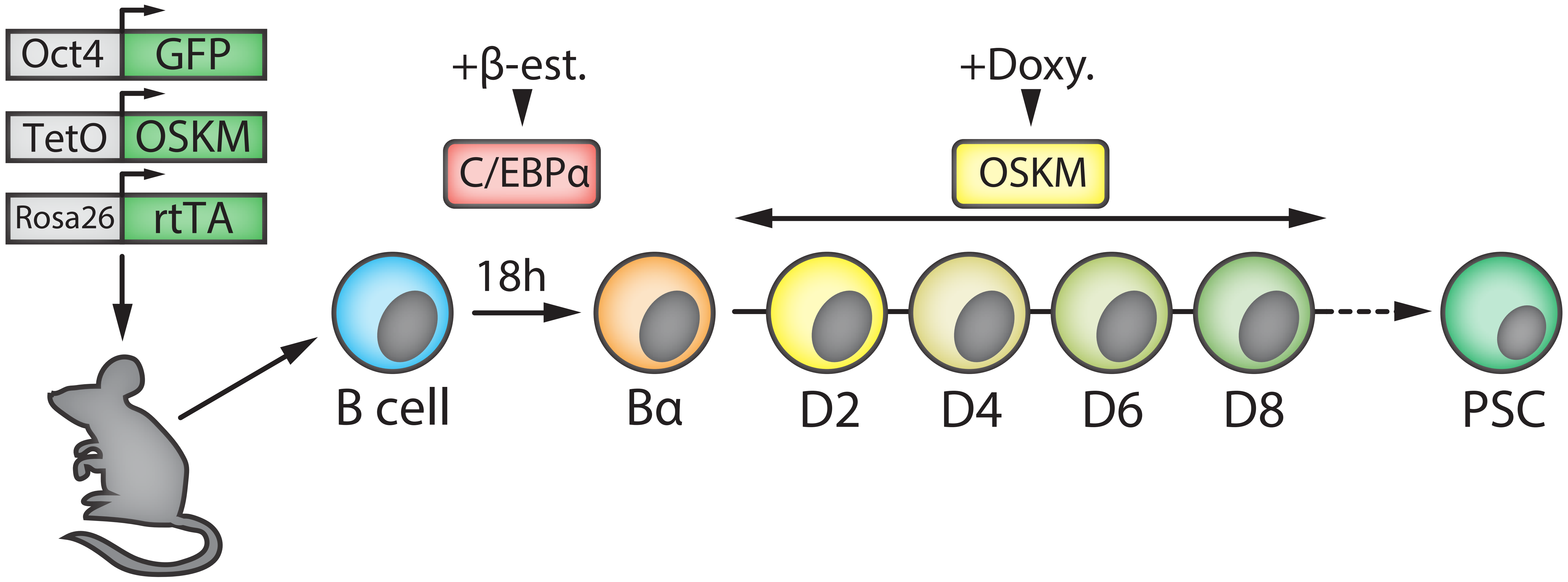
Interplay: topology, gene expression & chromatin

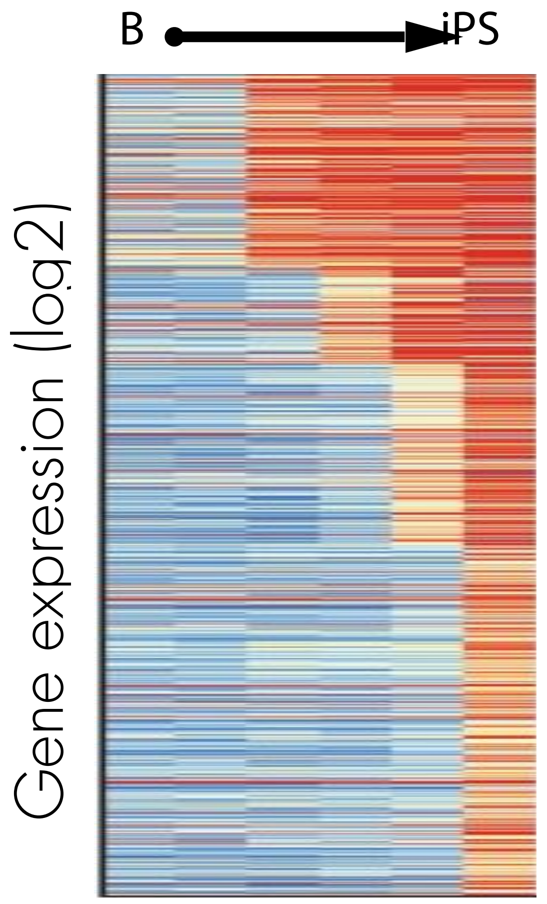
Stadhouders, R., Vidal, E. et al. (2018) Nature Genetics

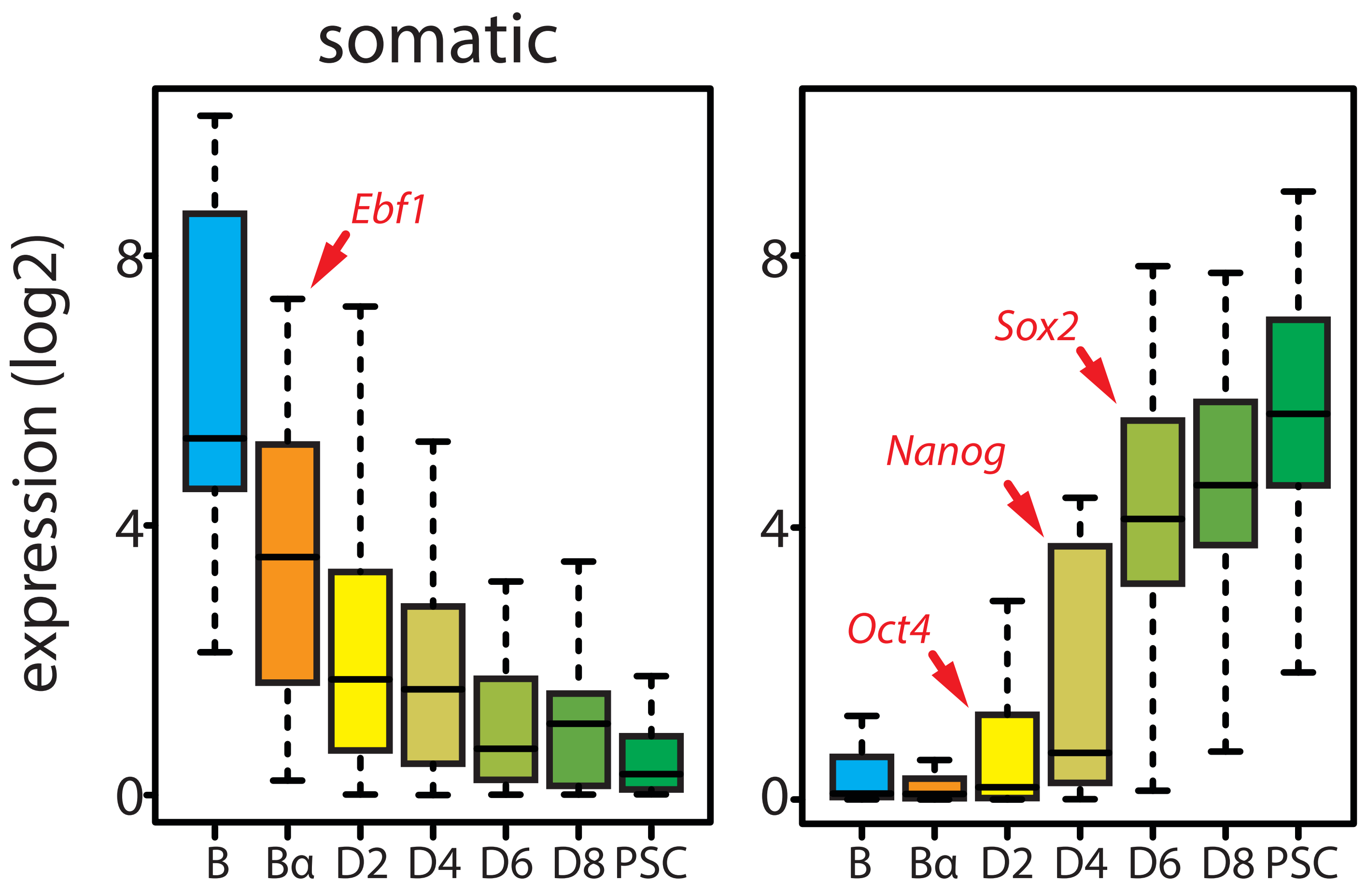


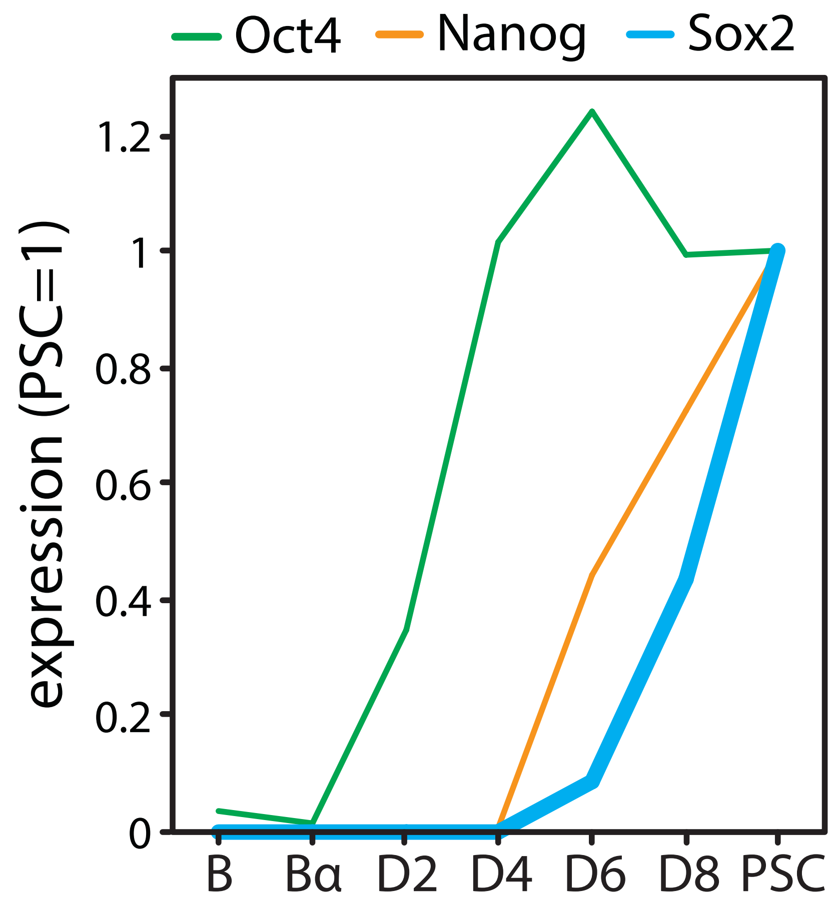
Reprogramming from B to PSC

Stadhouders, R., Vidal, E. et al. (2018) Nature Genetics



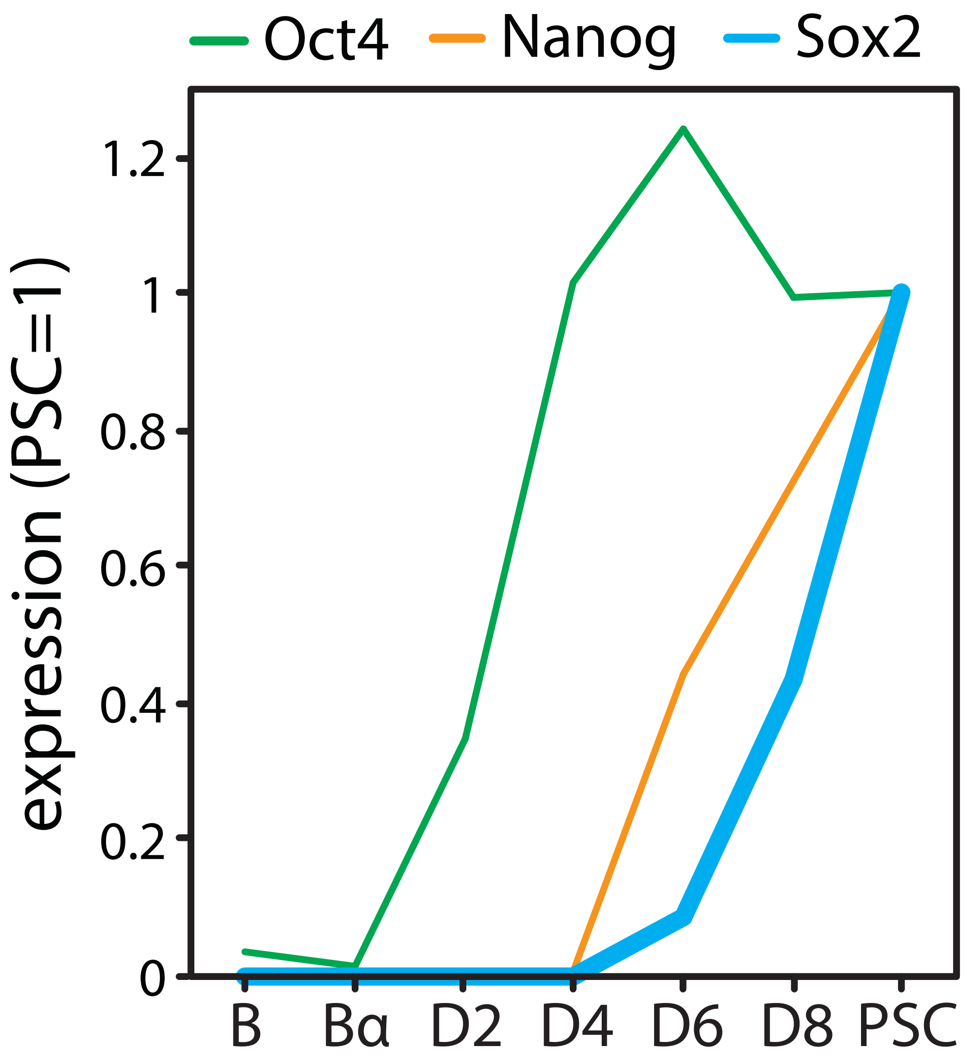


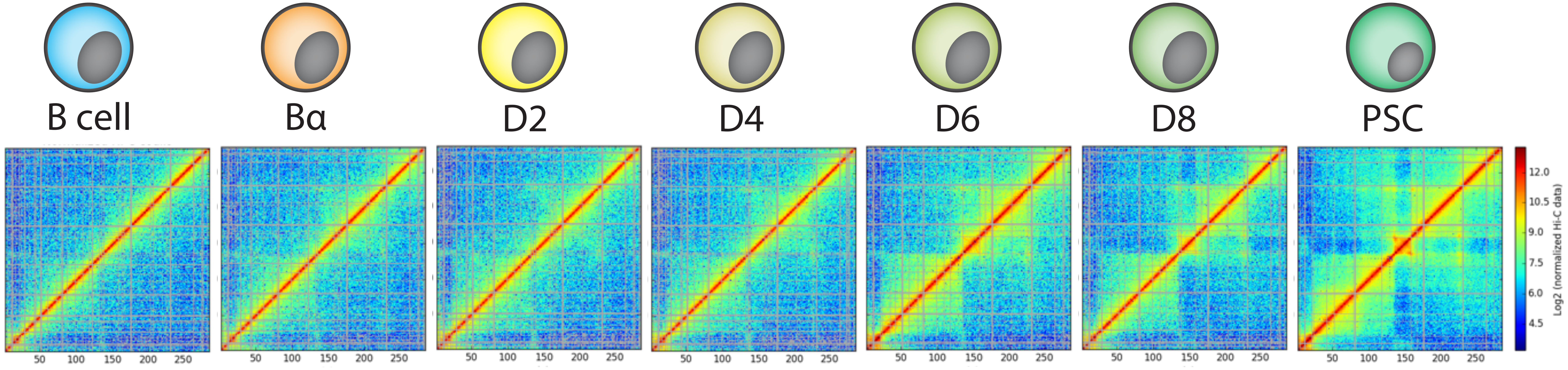




Hi-C maps of reprogramming from B to PSC

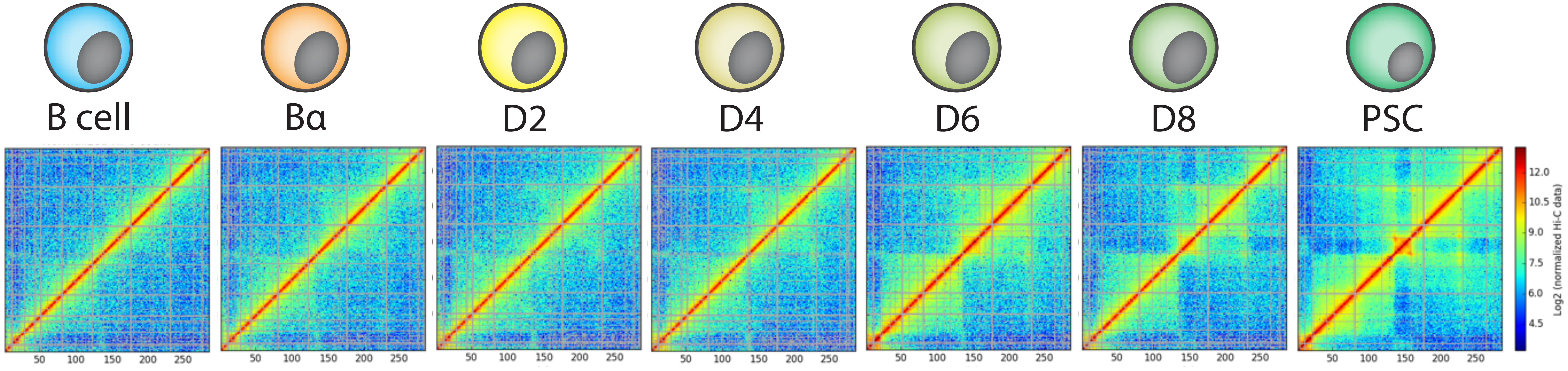
The SOX2 locus





Hi-C maps of reprogramming from B to PSC

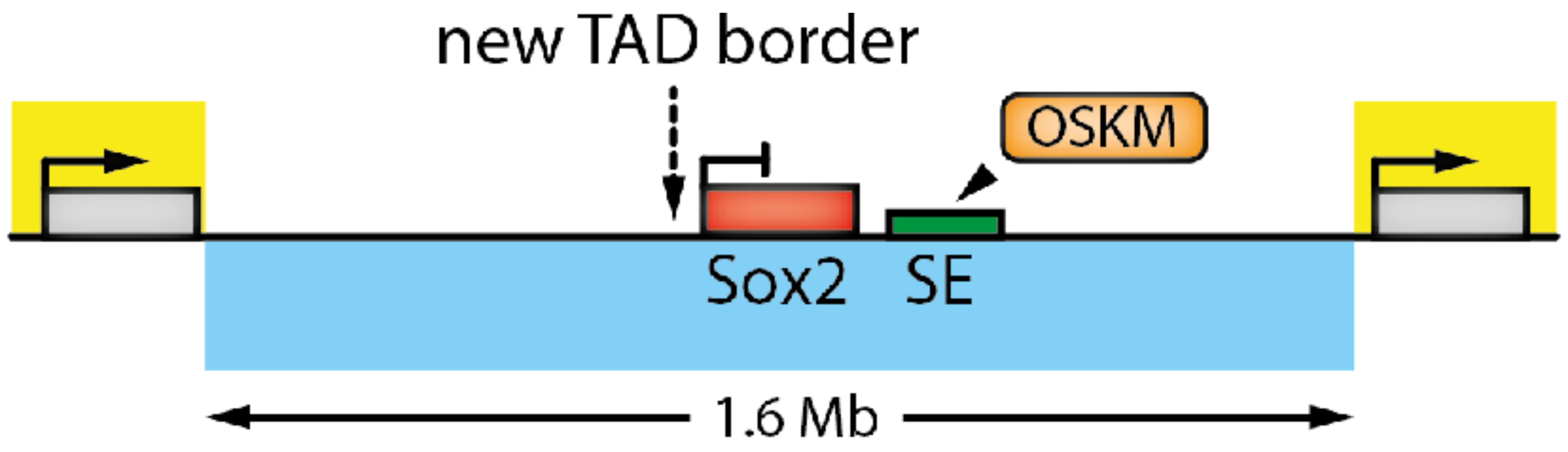
The SOX2 locus

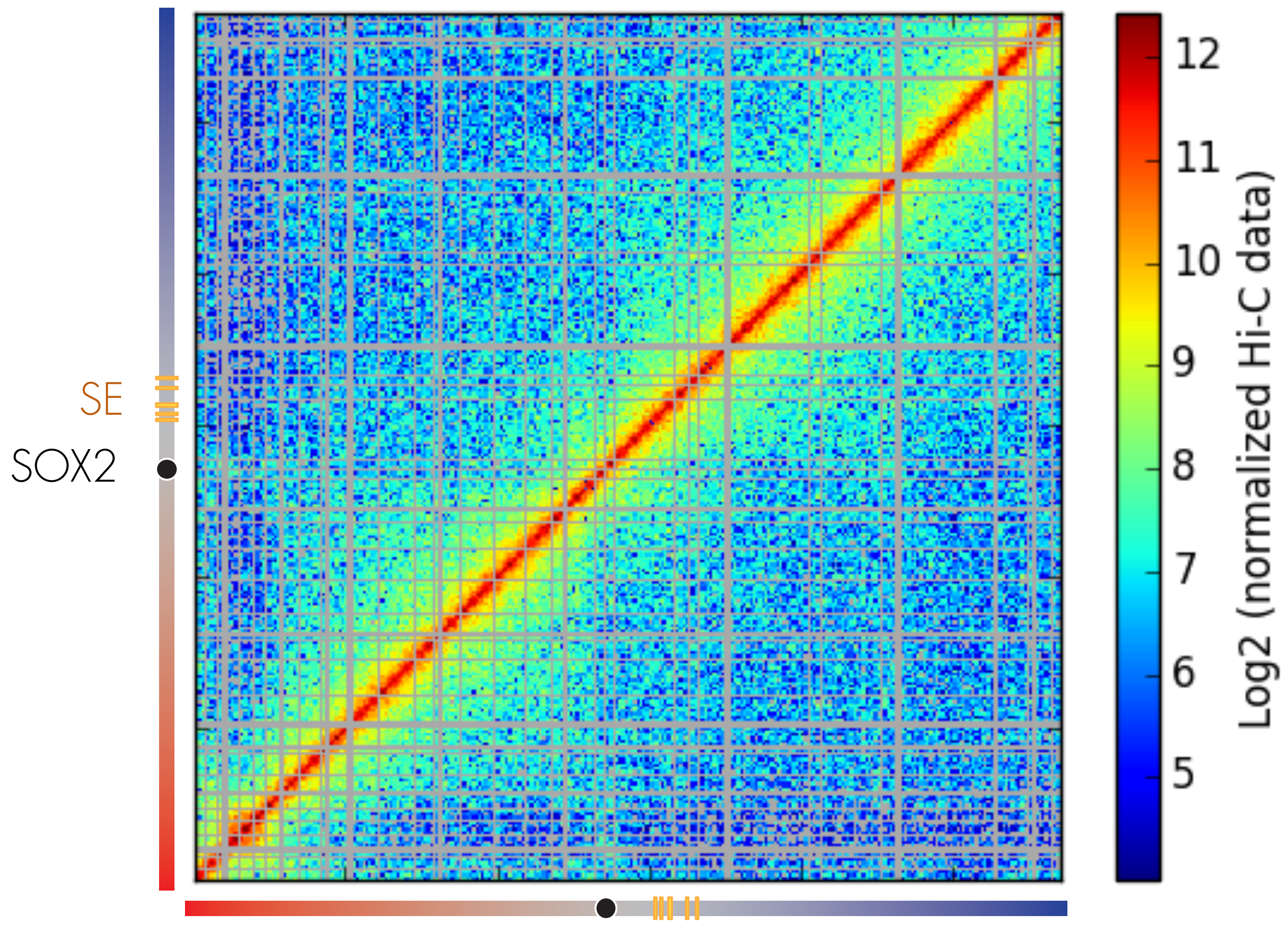


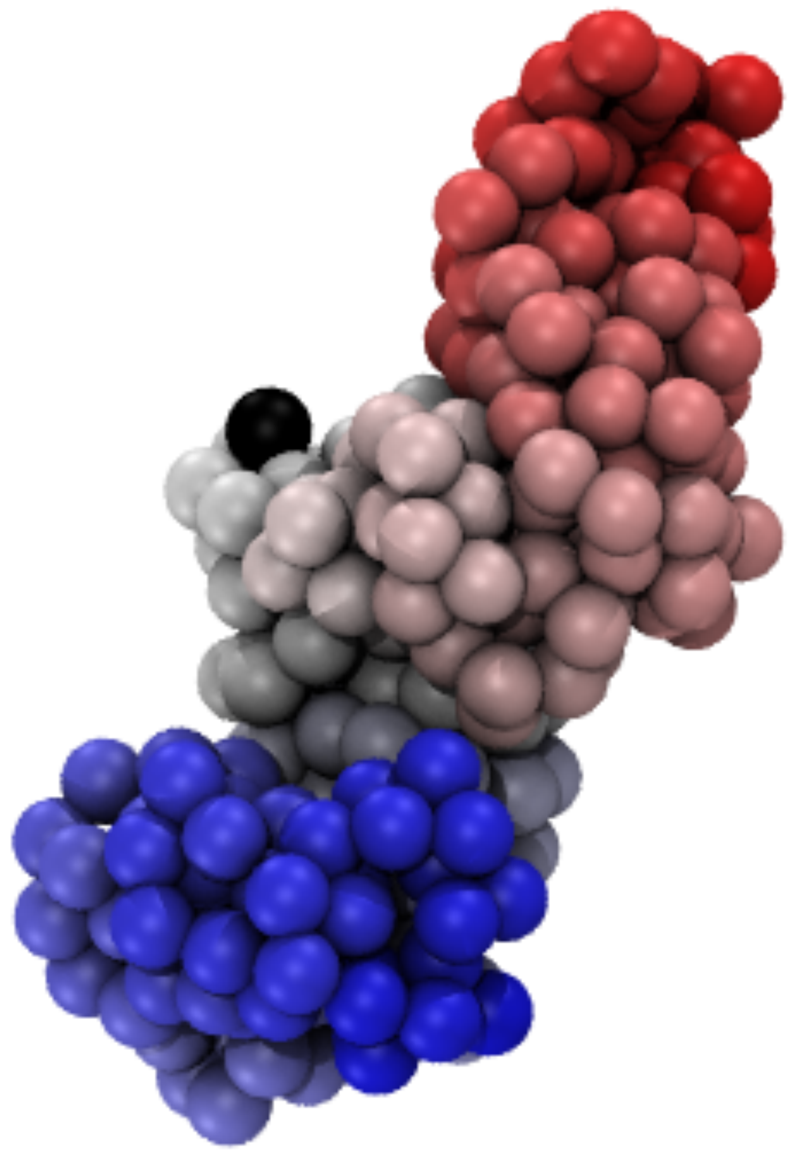
How does these structural rearrangements interplay with the transcription activity?

What are the main drivers of structural transitions?

TADbit modeling of SOX2 from B cells Hi-C





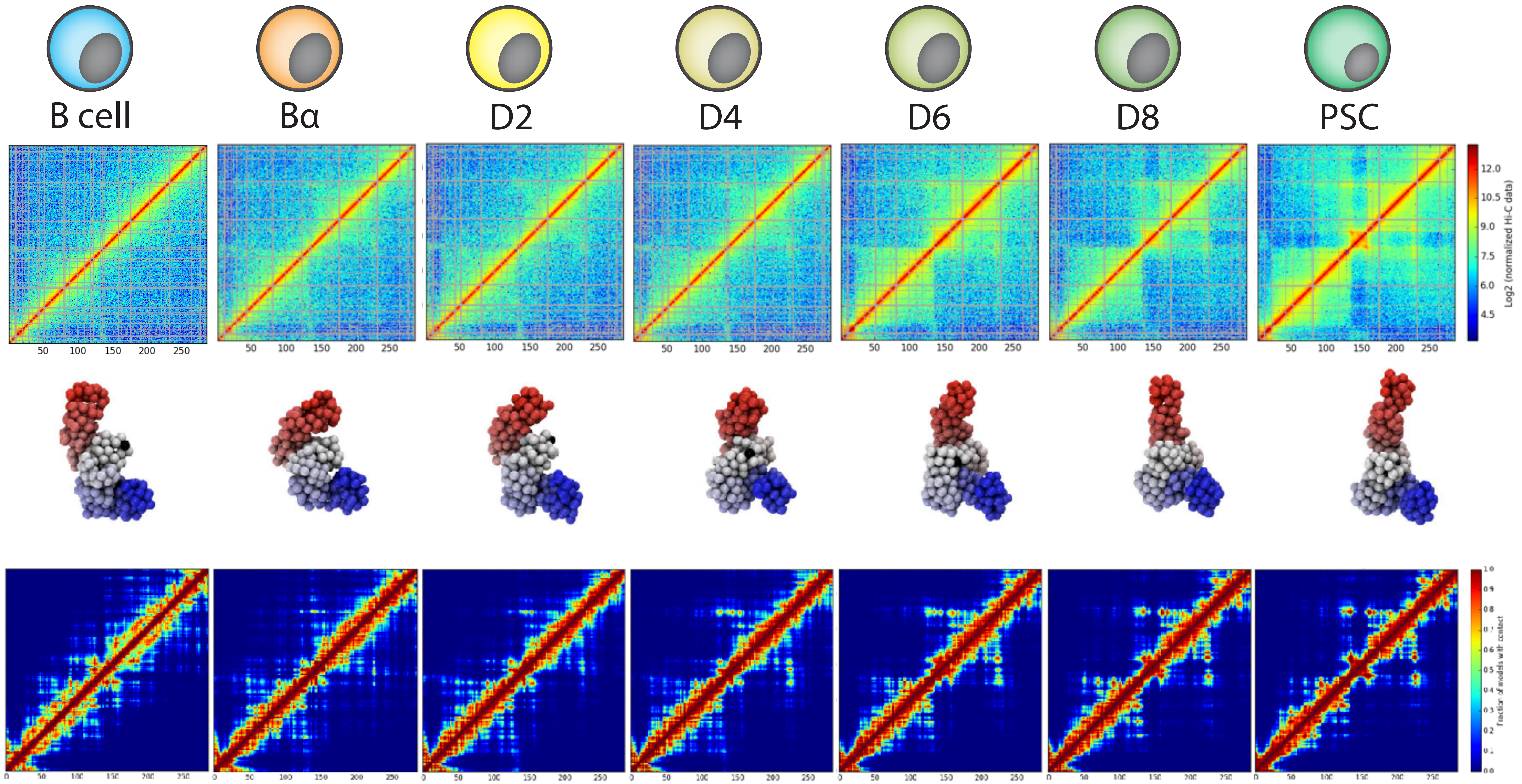


Optimal IMP parameters

lowfreq=0 , upfreq=1 , maxdist=200nm, dcutoff=125nm, particle size=50nm (5kb)

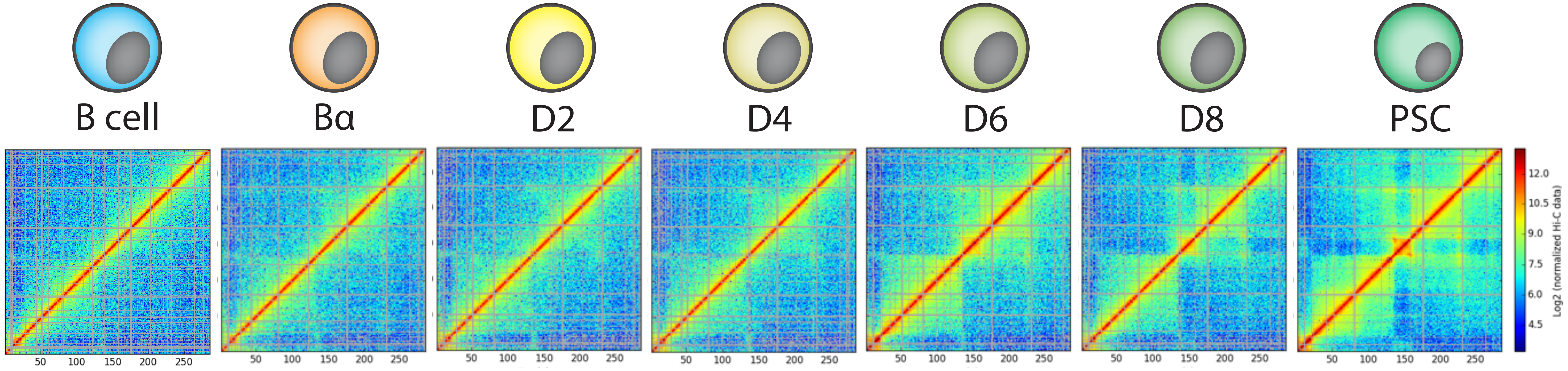
Models of reprogramming from B to PSC

The SOX2 locus



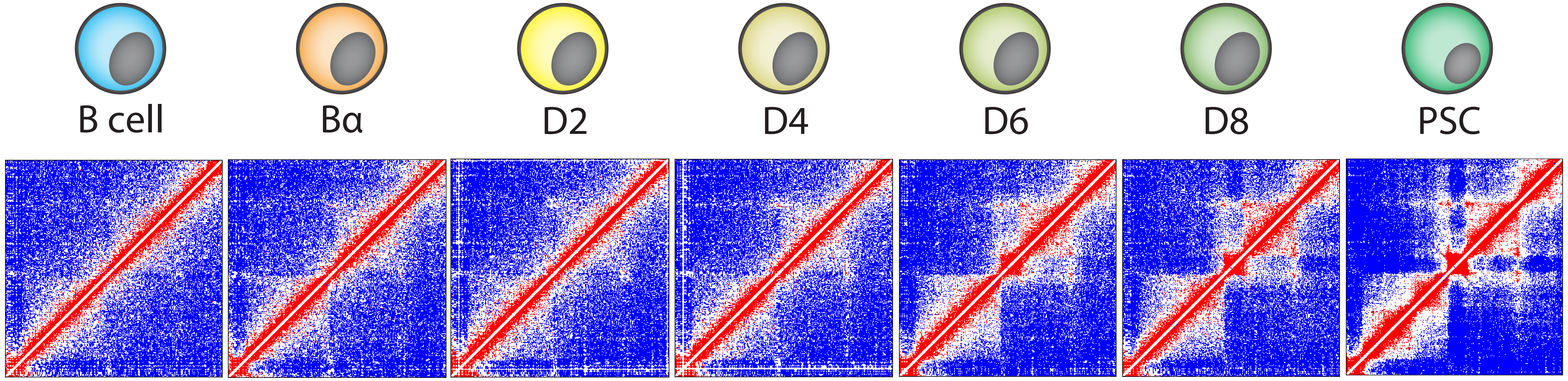
TADdyn: from time-series Hi-C maps to dynamic restraints

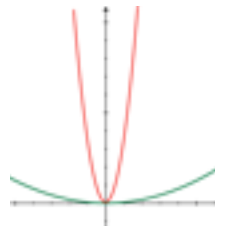
The SOX2 locus



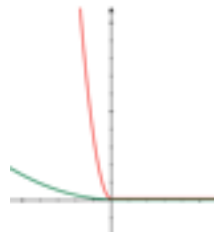
TADdyn: from time-series Hi-C maps to dynamic restraints

The SOX2 locus





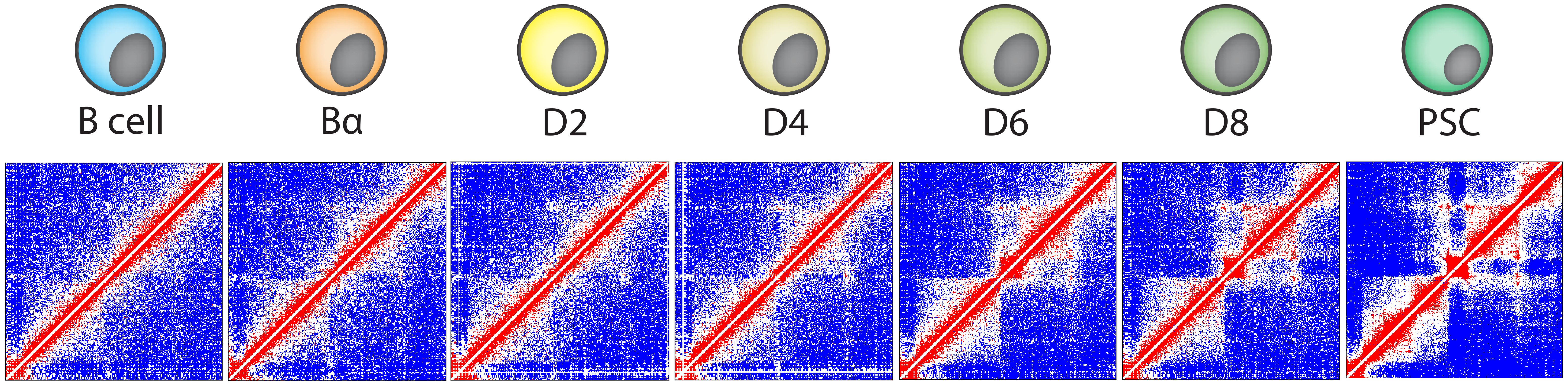
🟥 Harmonic

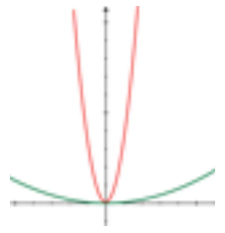


🟦 HarmonicLowerBound

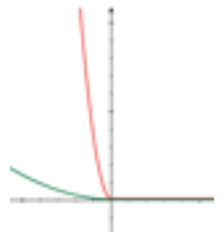
TADdyn: from time-series Hi-C maps to dynamic restraints

The SOX2 locus

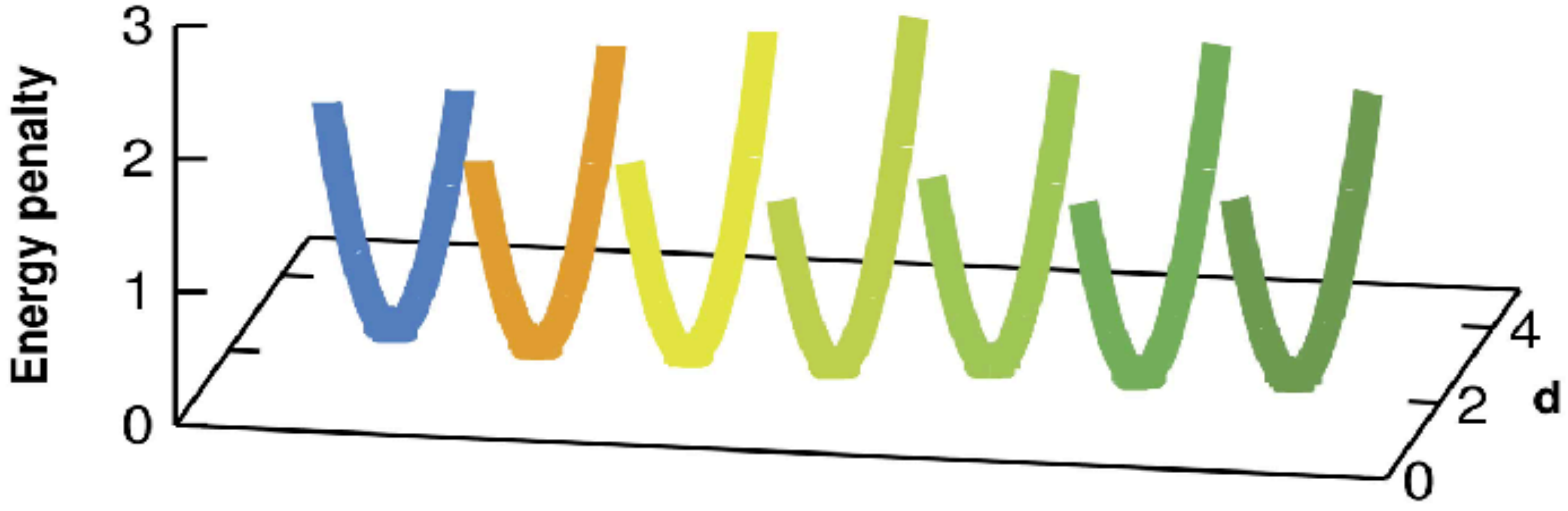




🟥 Harmonic



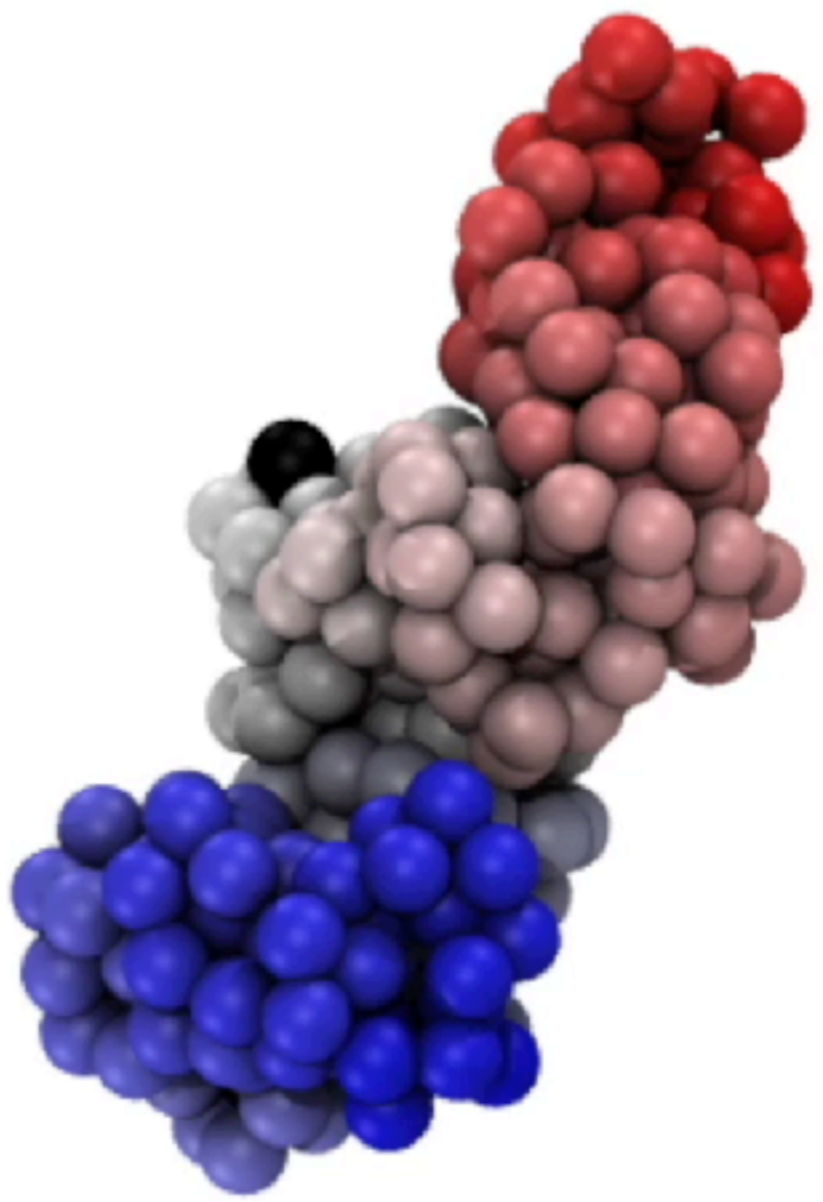
🟦 HarmonicLowerBound

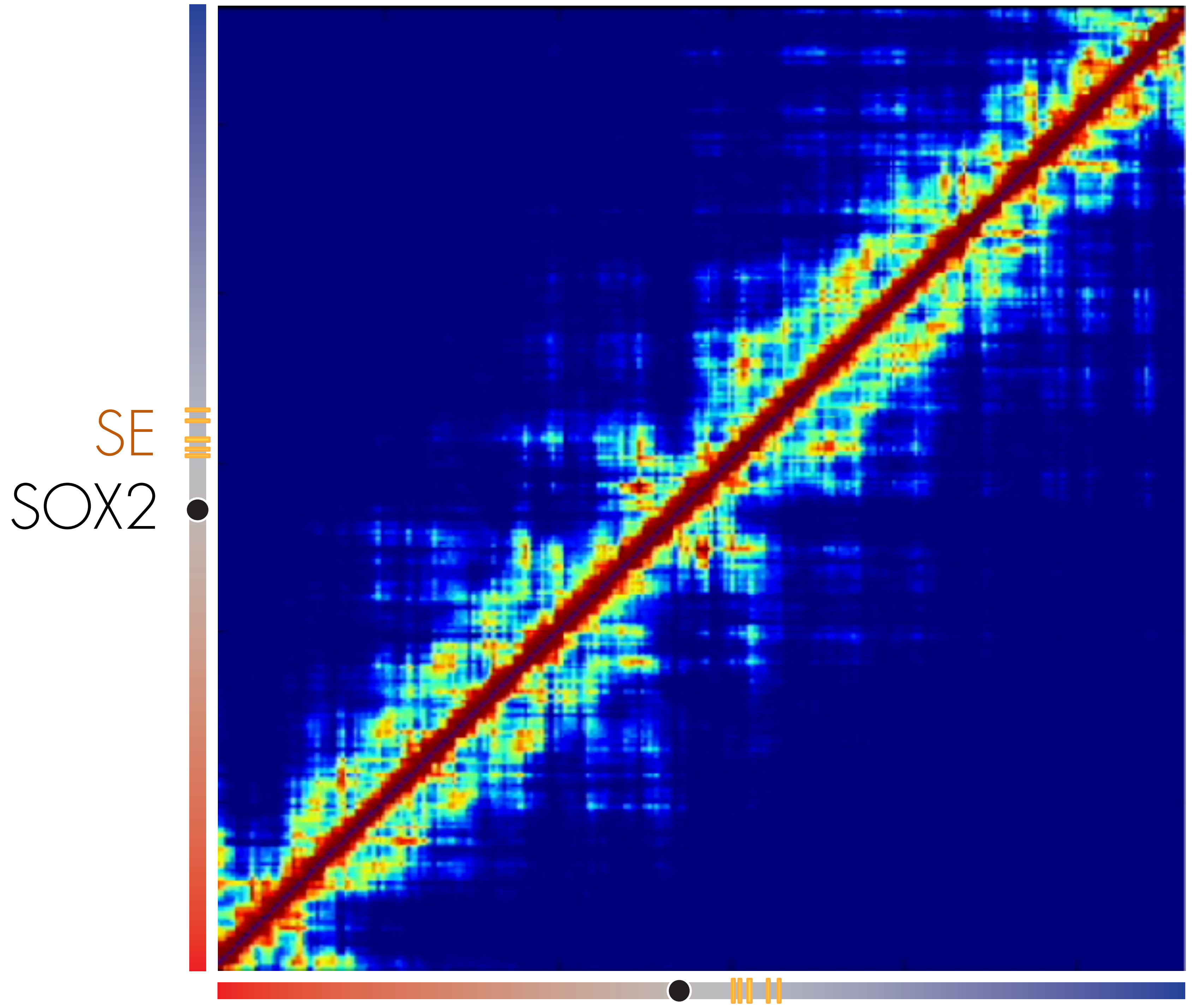


Transition	Stable	Vanishing	Raising
B -> Bα	18,612	6,984	7,290
Bα -> D2	18,512	7,390	6,687
D2 -> D4	18,369	6,830	6,893
D4 -> D6	18,971	6,291	7,289
D6 -> D8	20,167	6,093	6,250
D8 -> ES	20,679	5,738	6,173

SOX2 locus structural changes from B to PSC

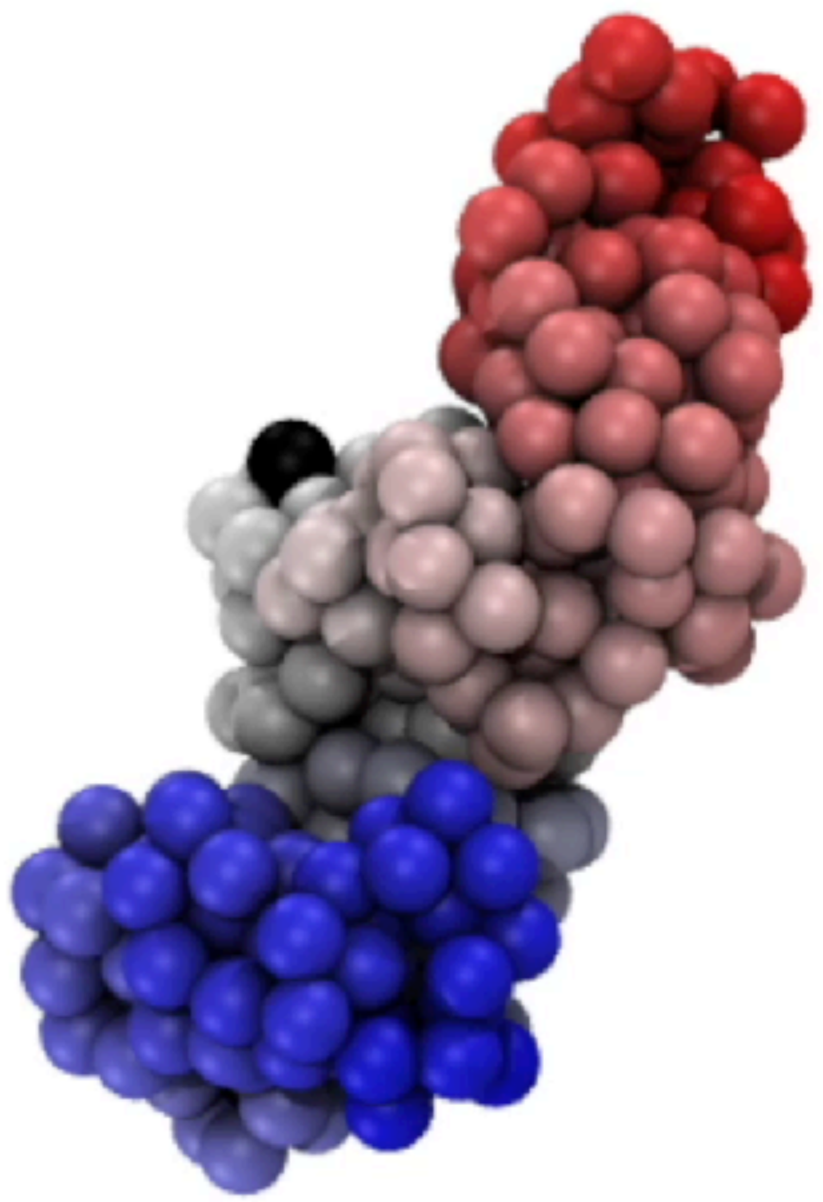
Contacts

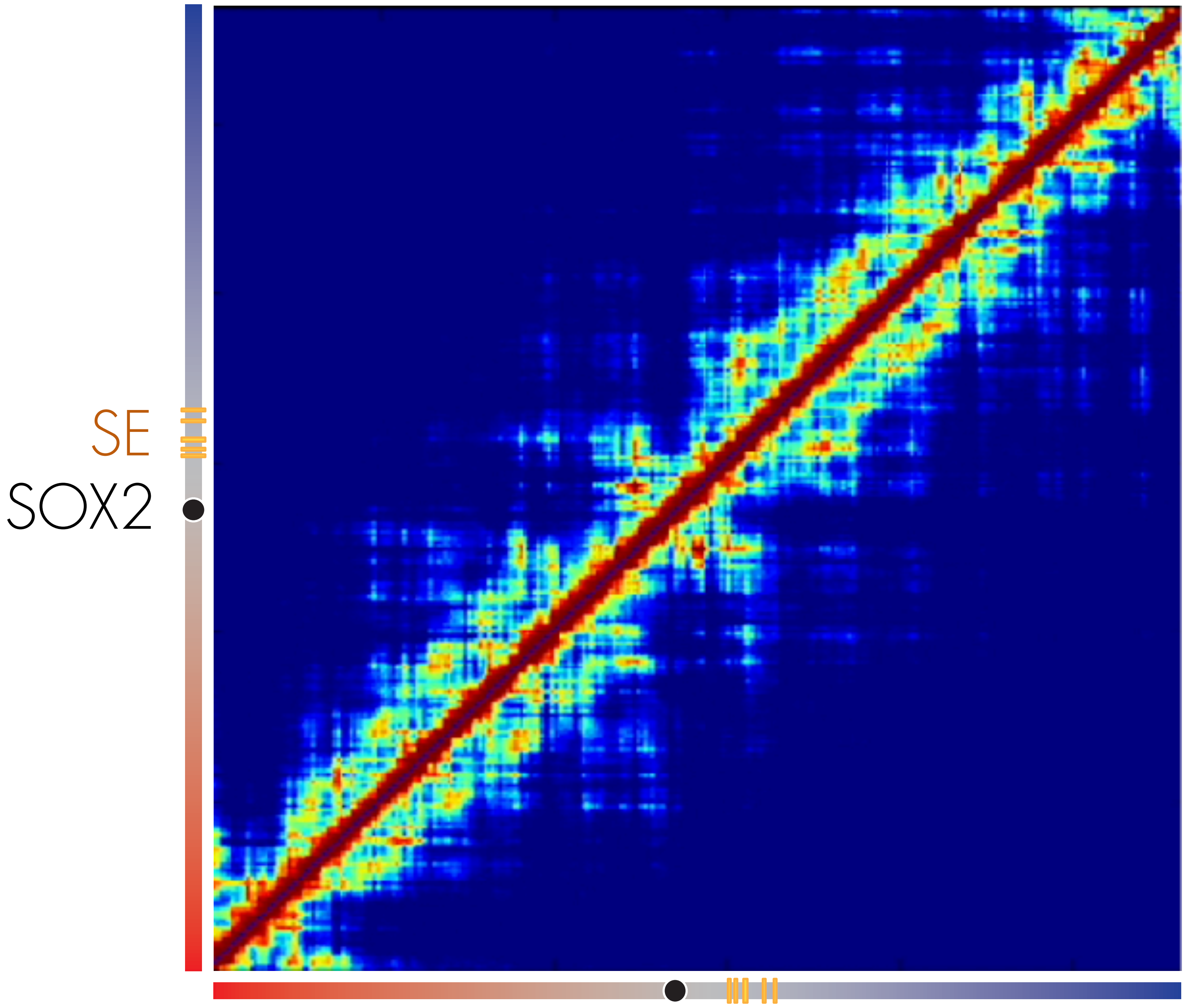




SOX2 locus structural changes from B to PSC

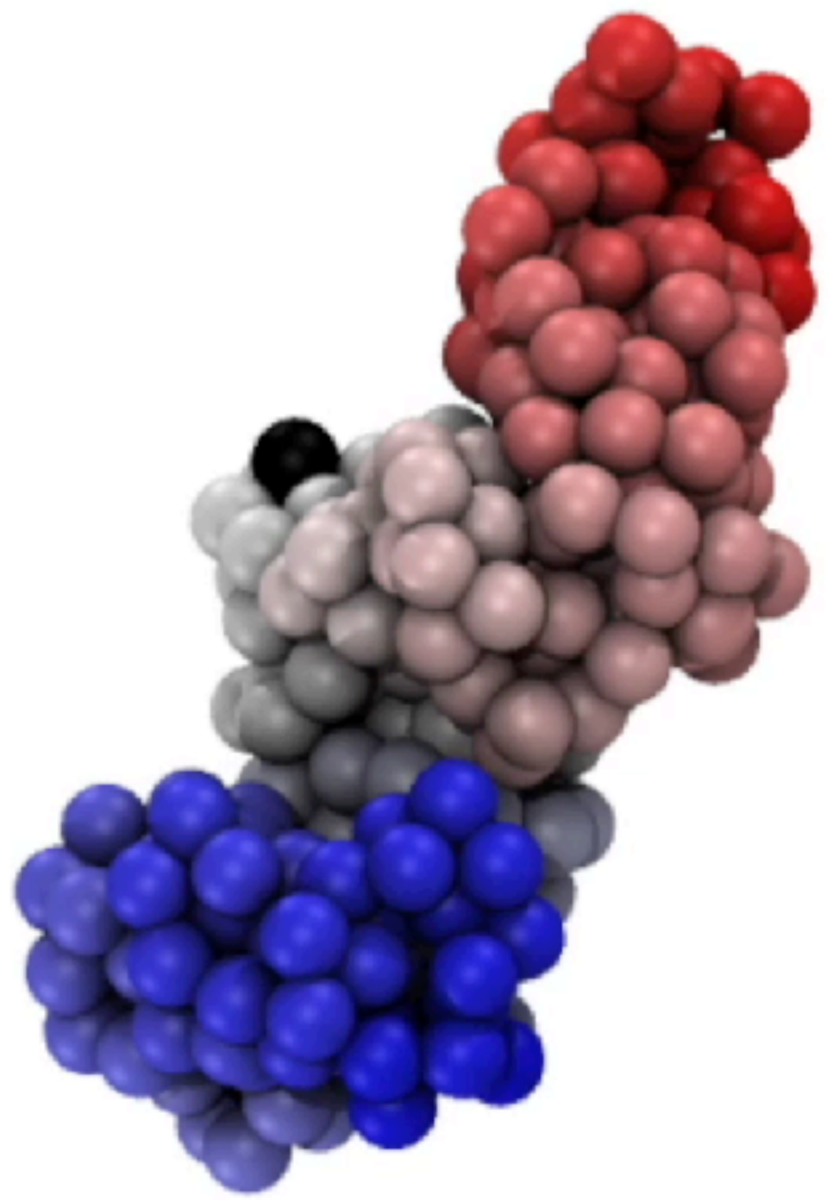
Contacts

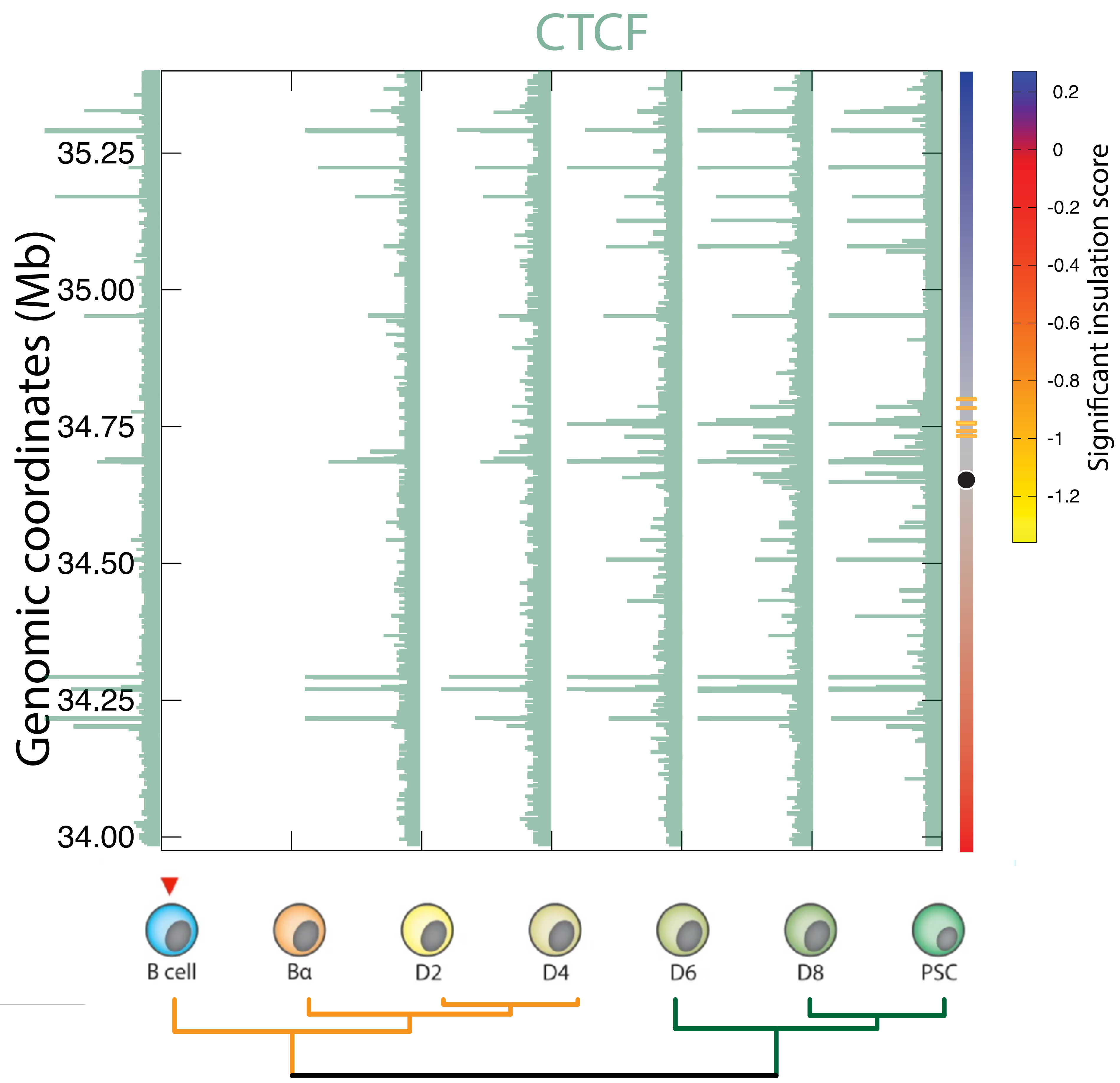




SOX2 locus structural changes from B to PSC

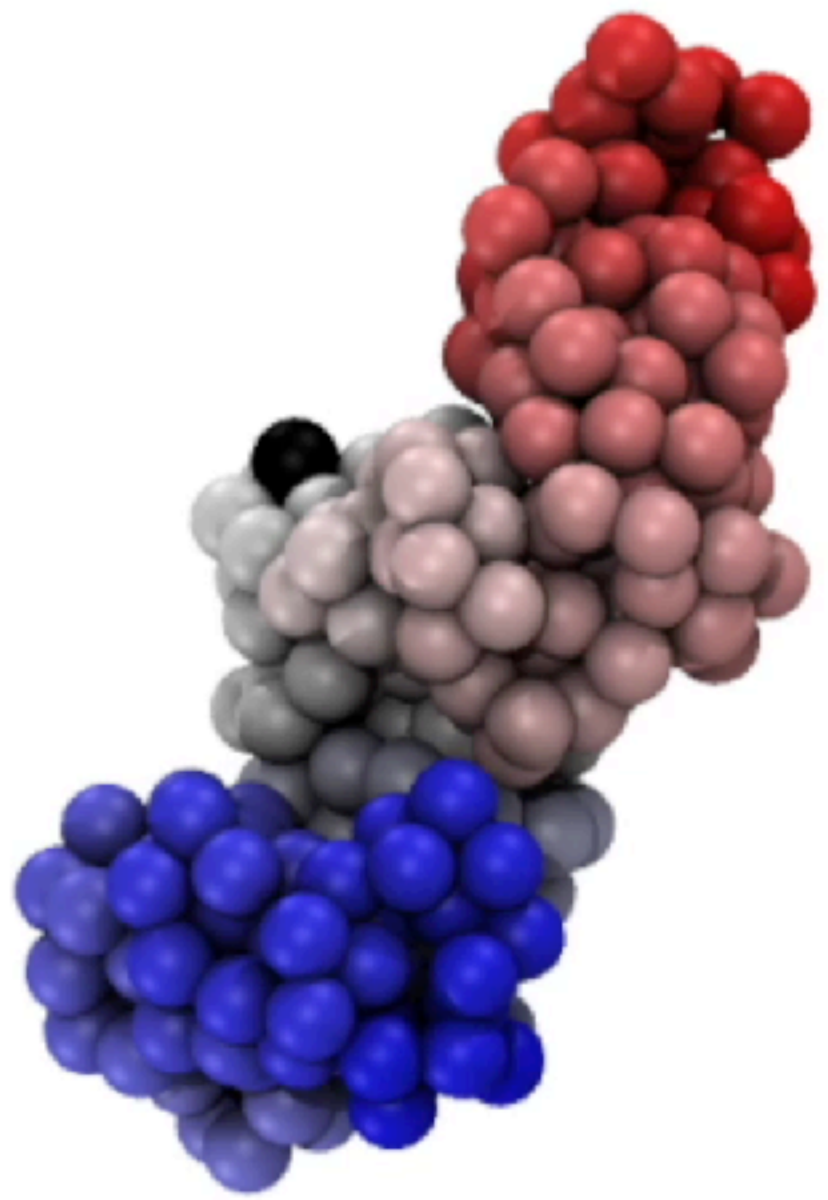
TAD borders

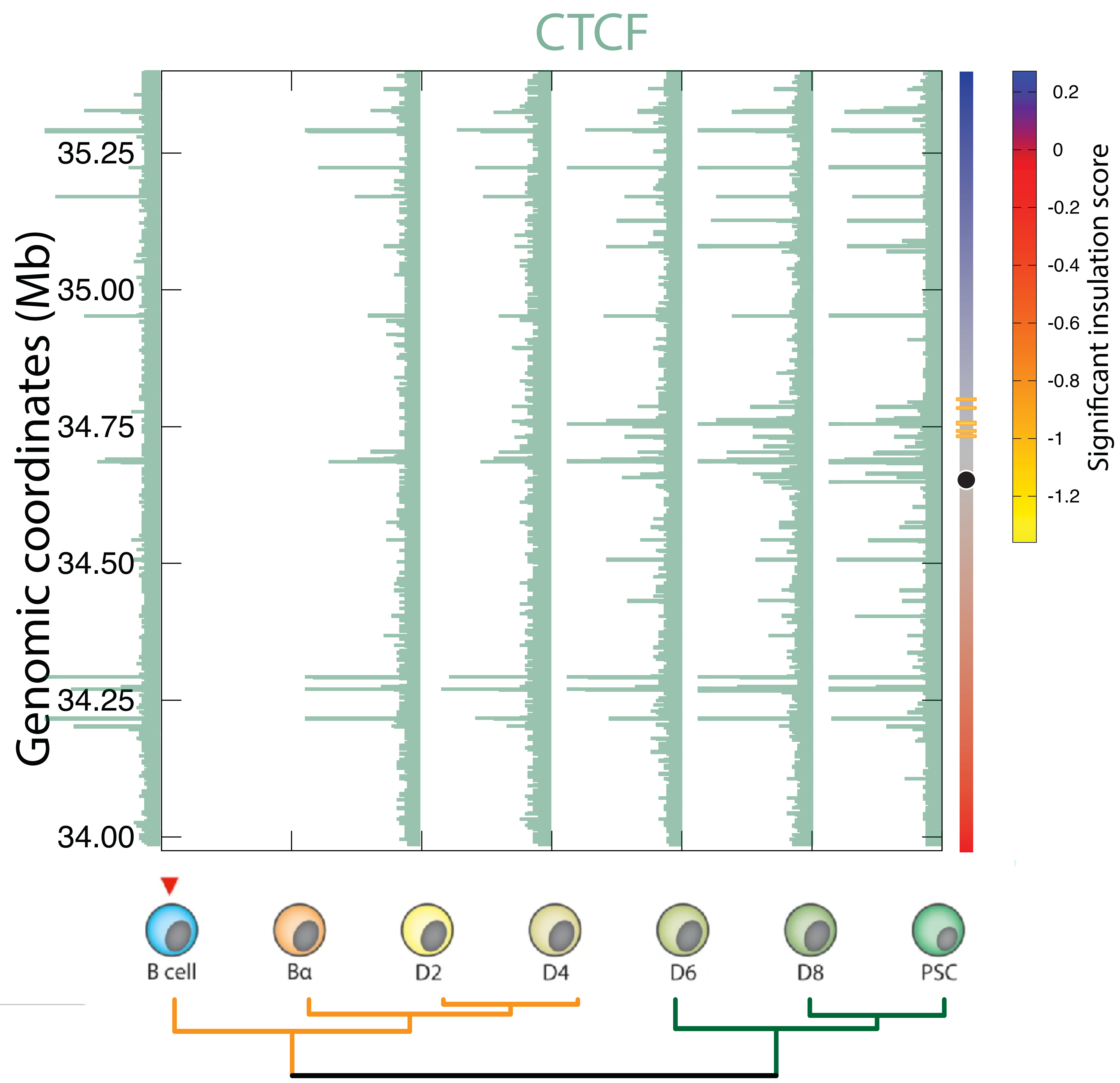




SOX2 locus structural changes from B to PSC

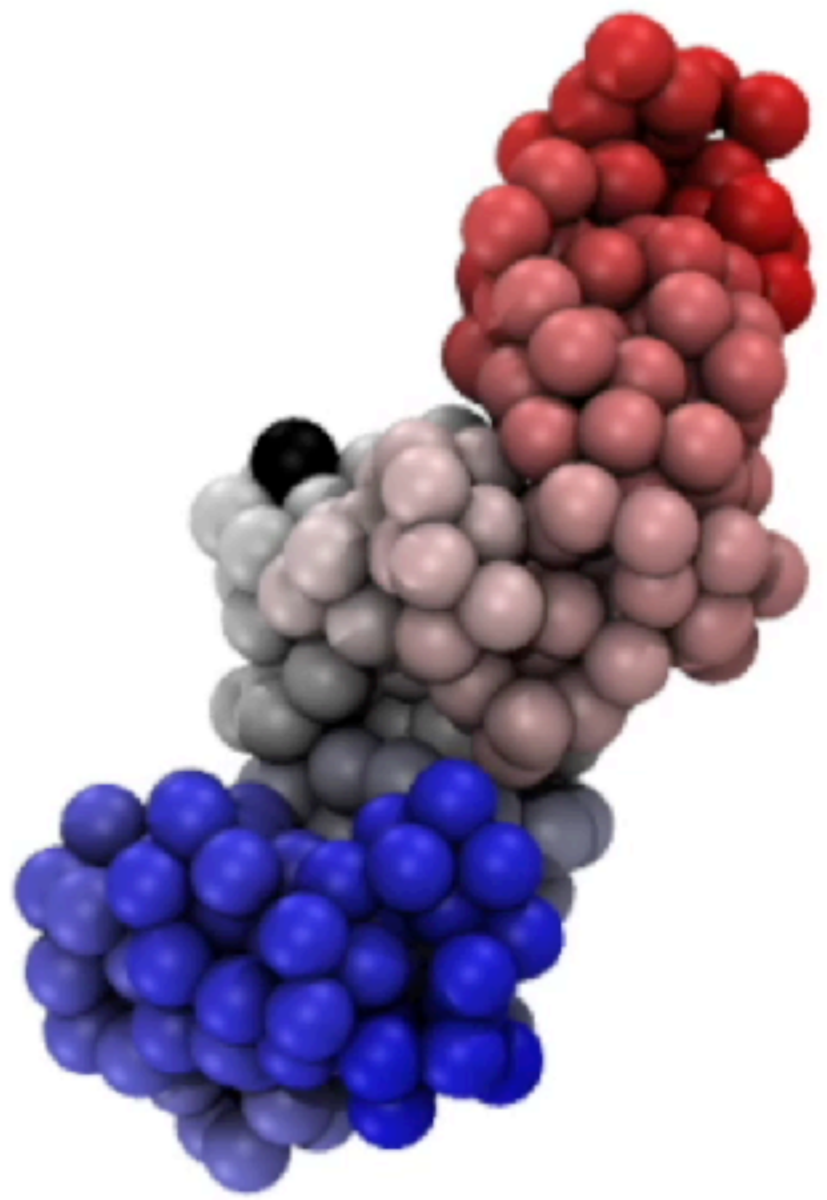
TAD borders

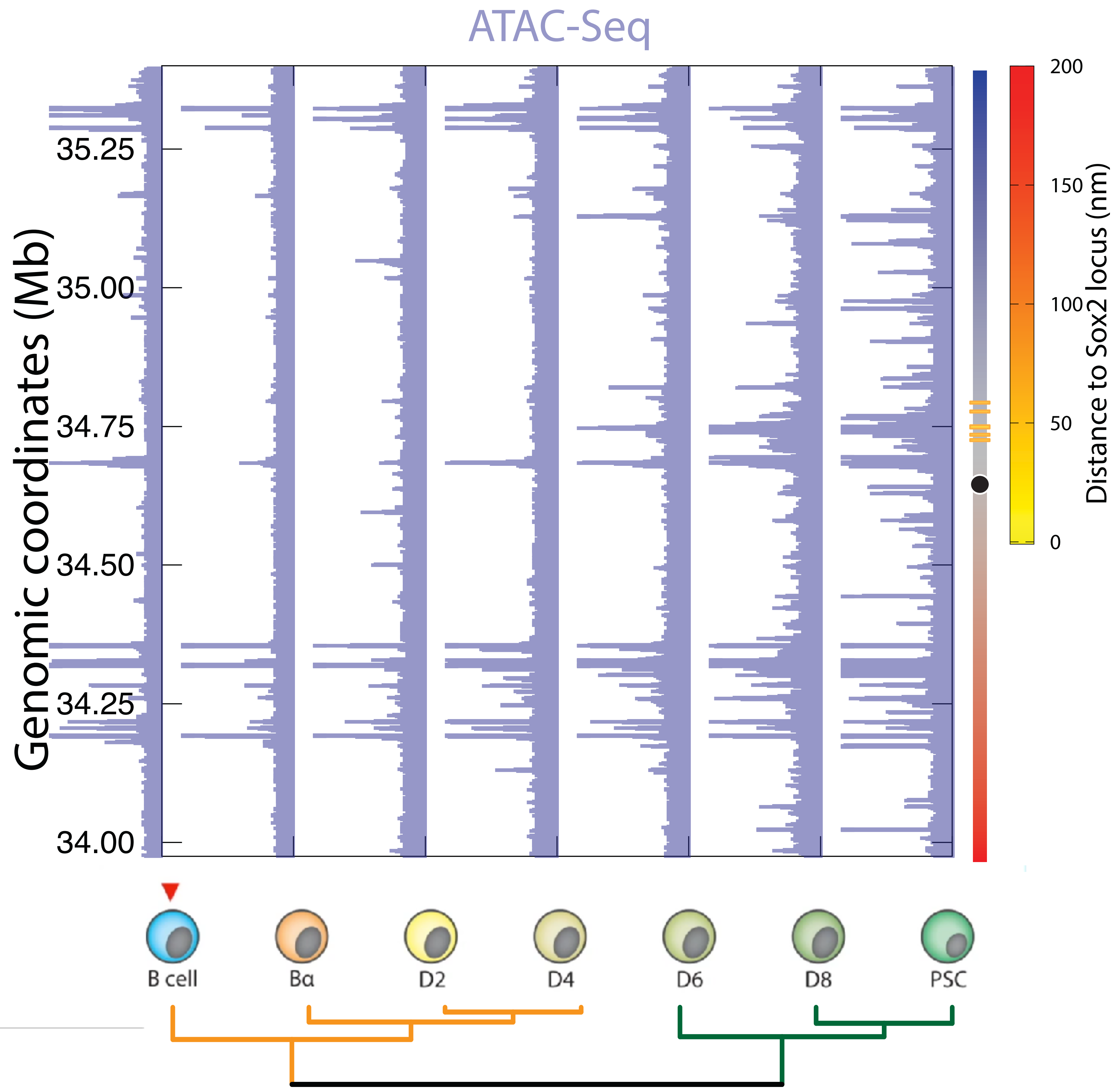




SOX2 locus structural changes from B to PSC

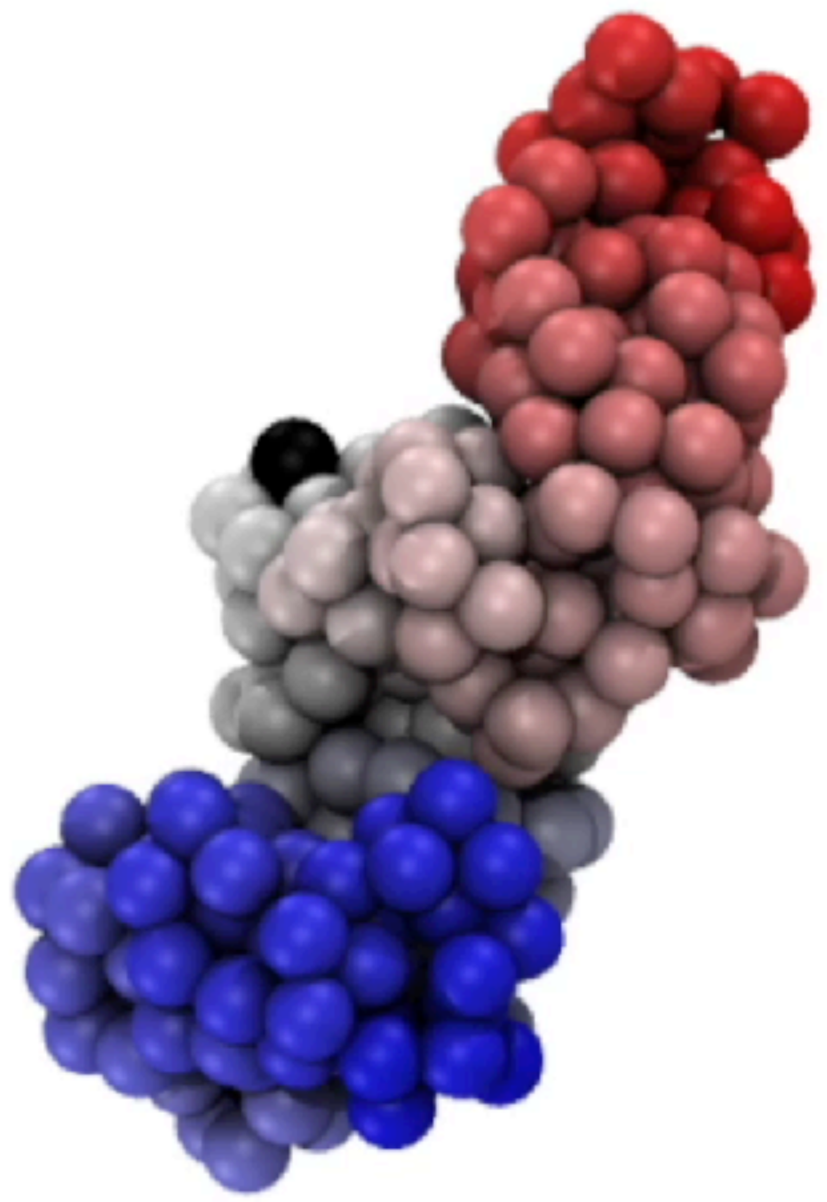
Distance to regulatory elements

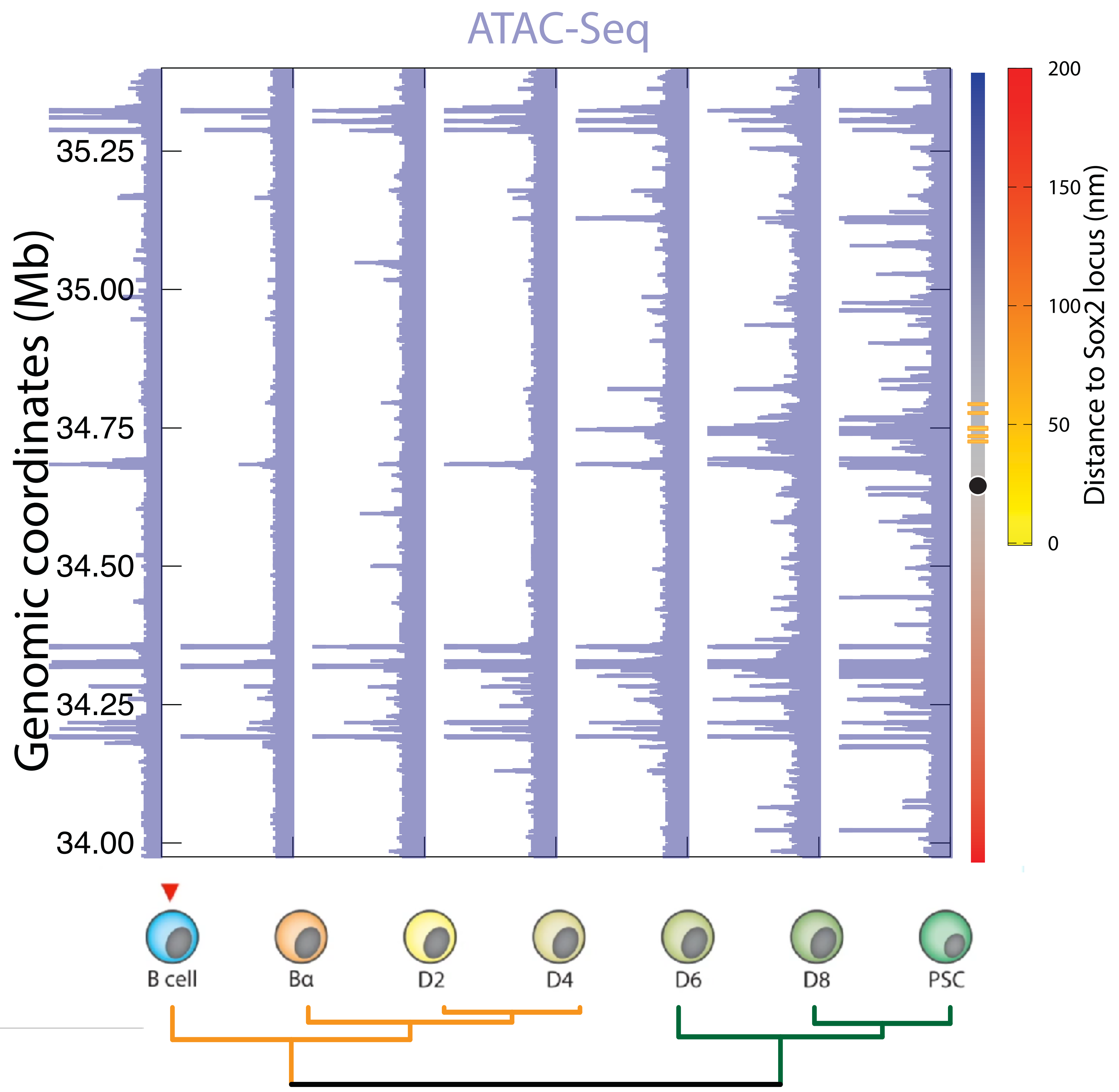




SOX2 locus structural changes from B to PSC

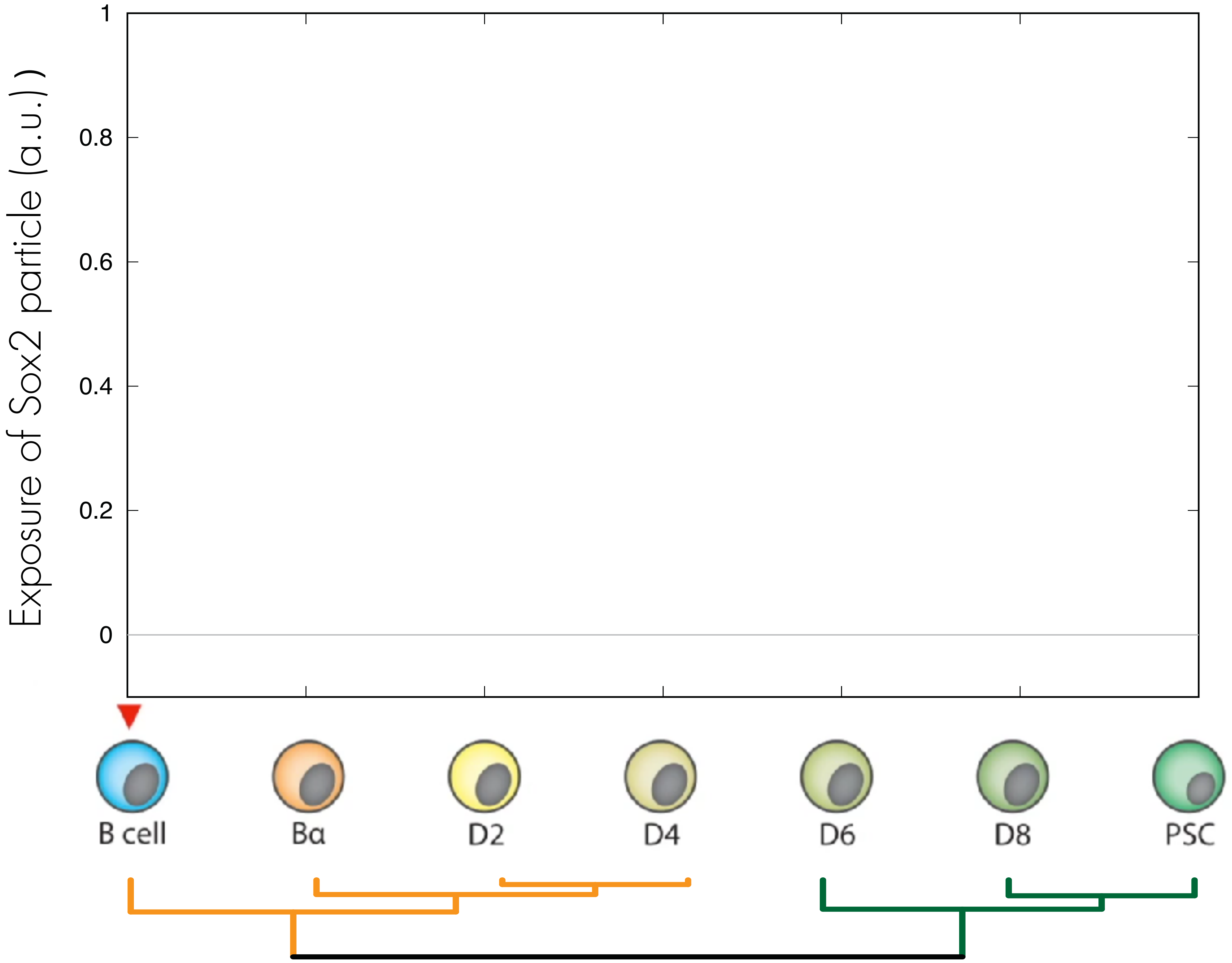
Distance to regulatory elements





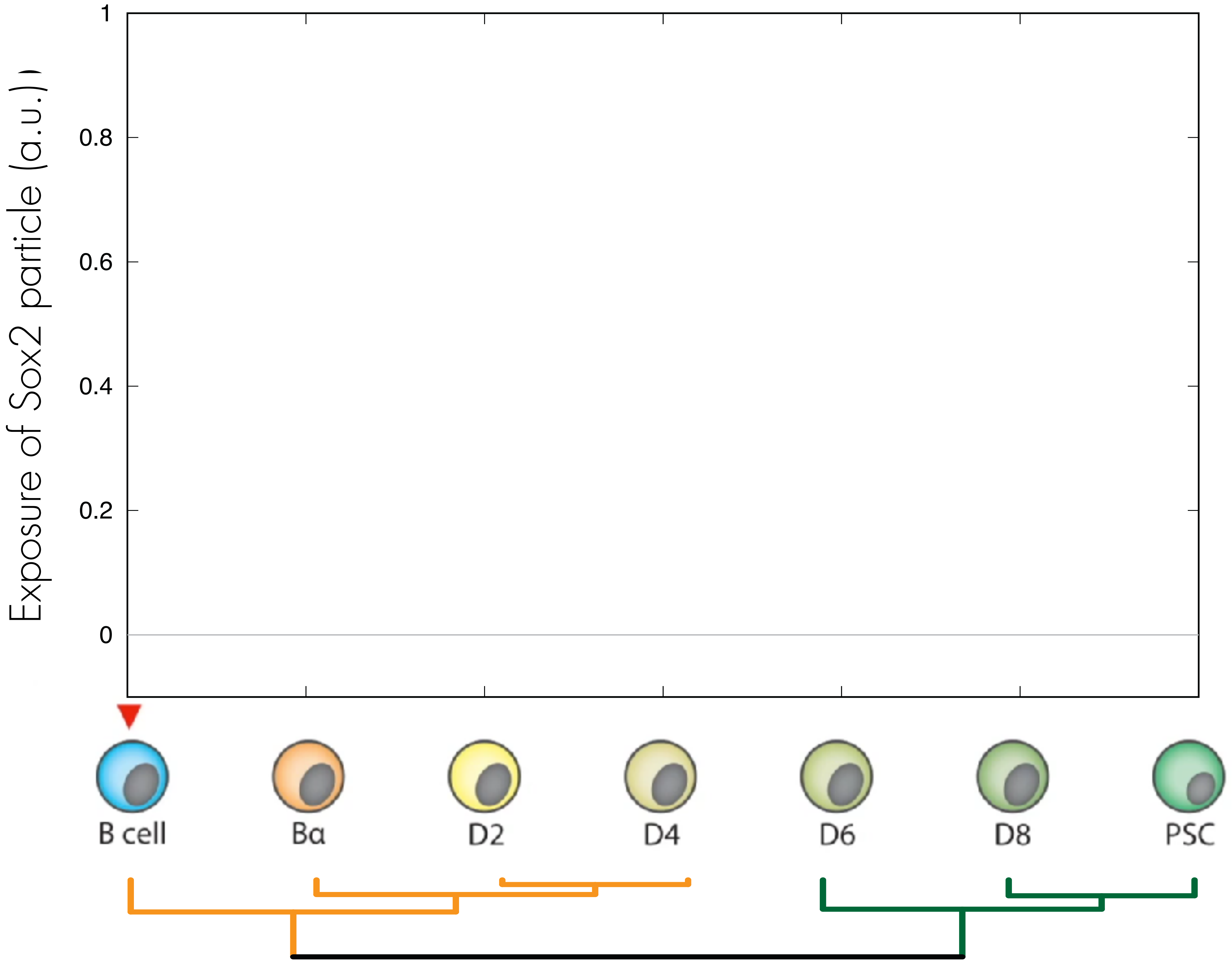
SOX2 locus structural changes from B to PSC

Structural exposure



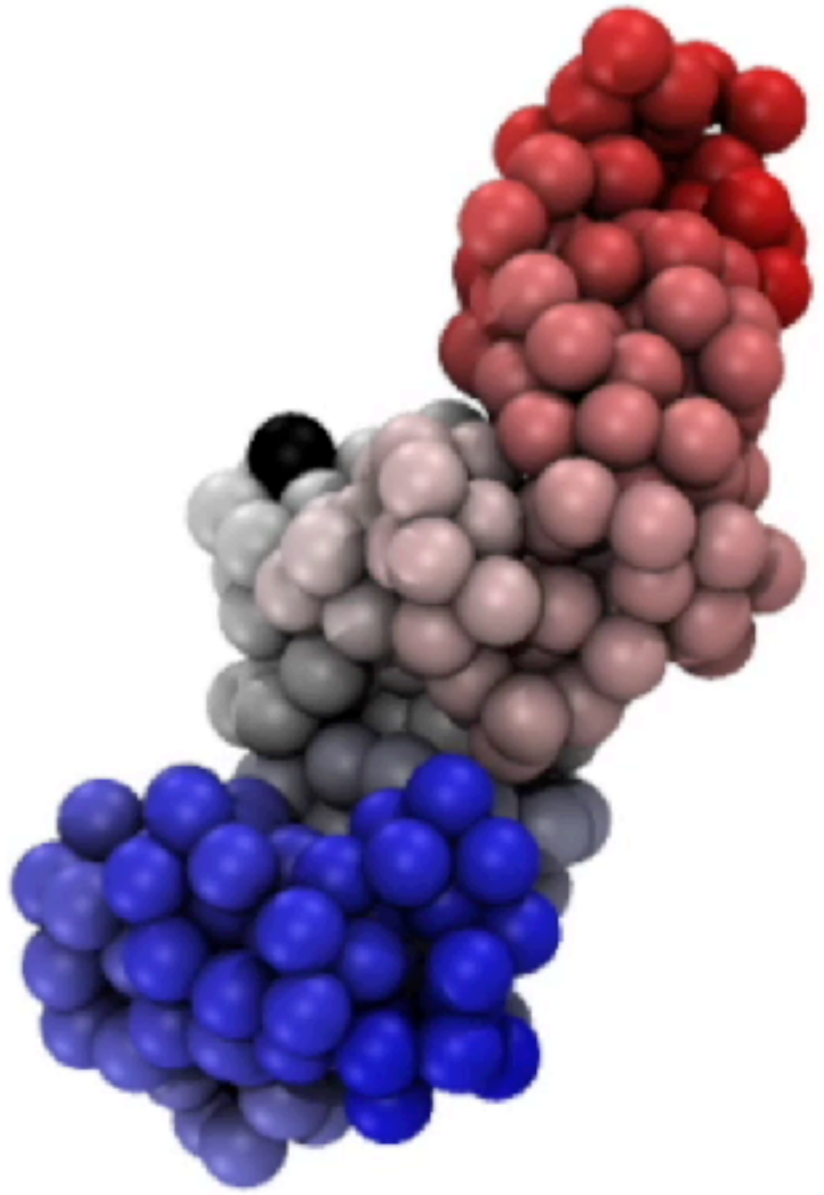
SOX2 locus structural changes from B to PSC

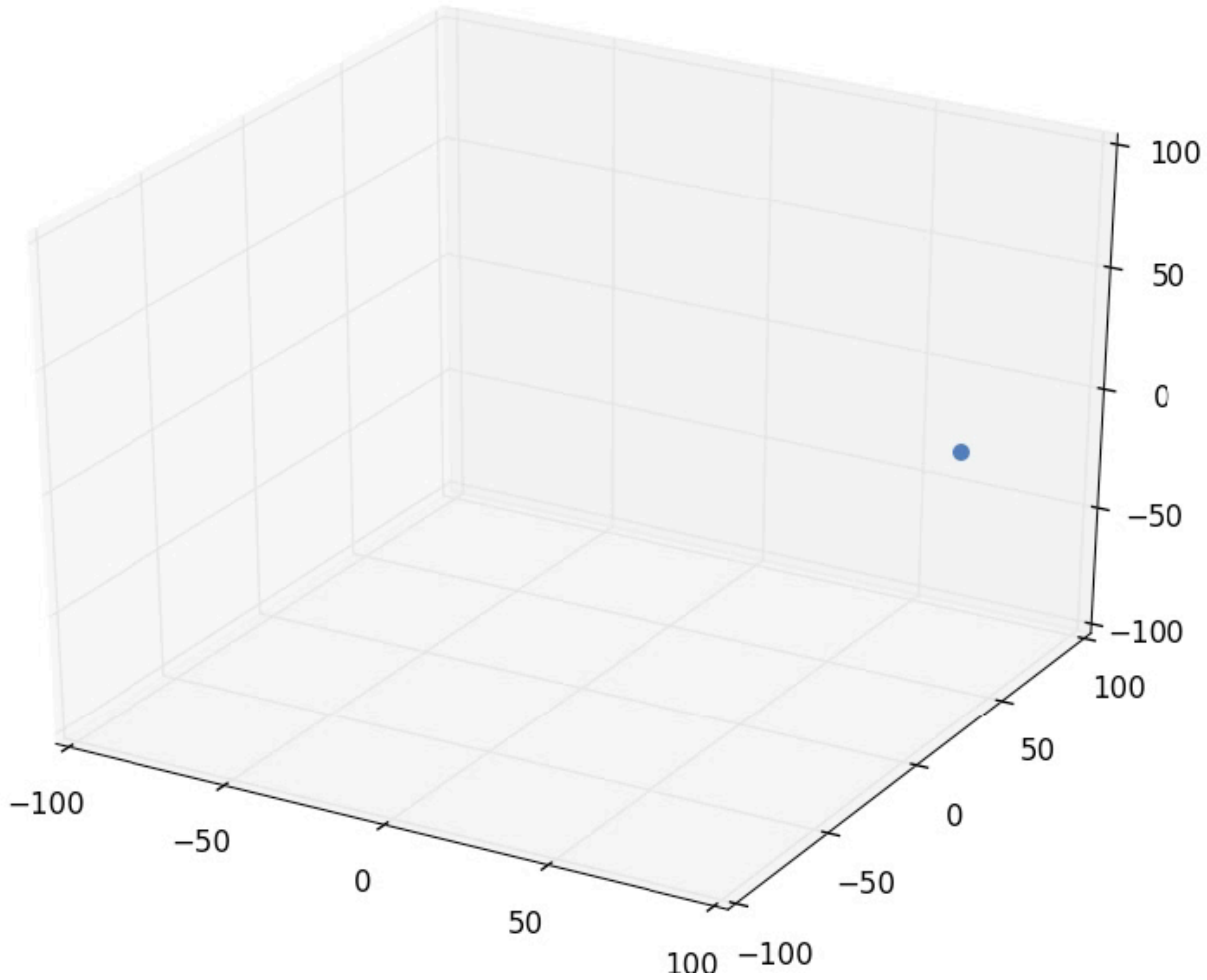
Structural exposure

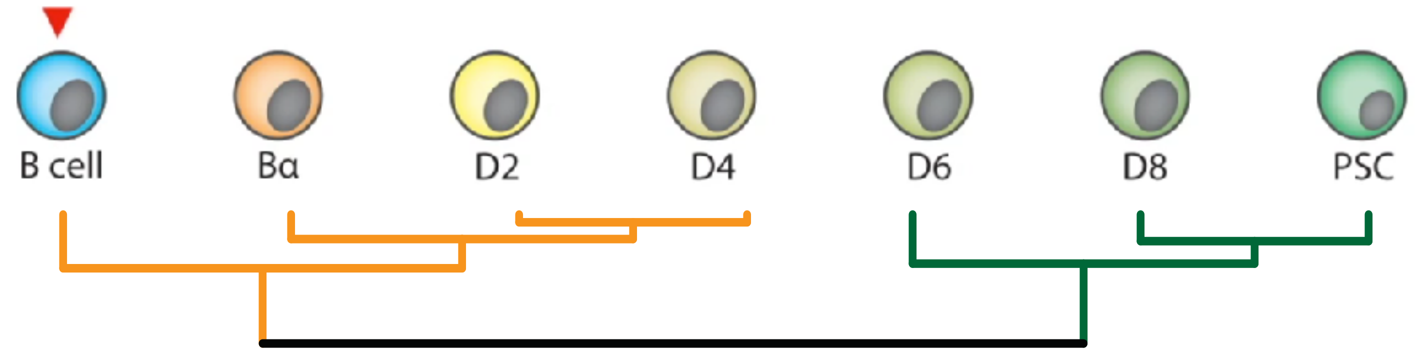


SOX2 locus dynamics changes from B to PSC

SOX2 displacement

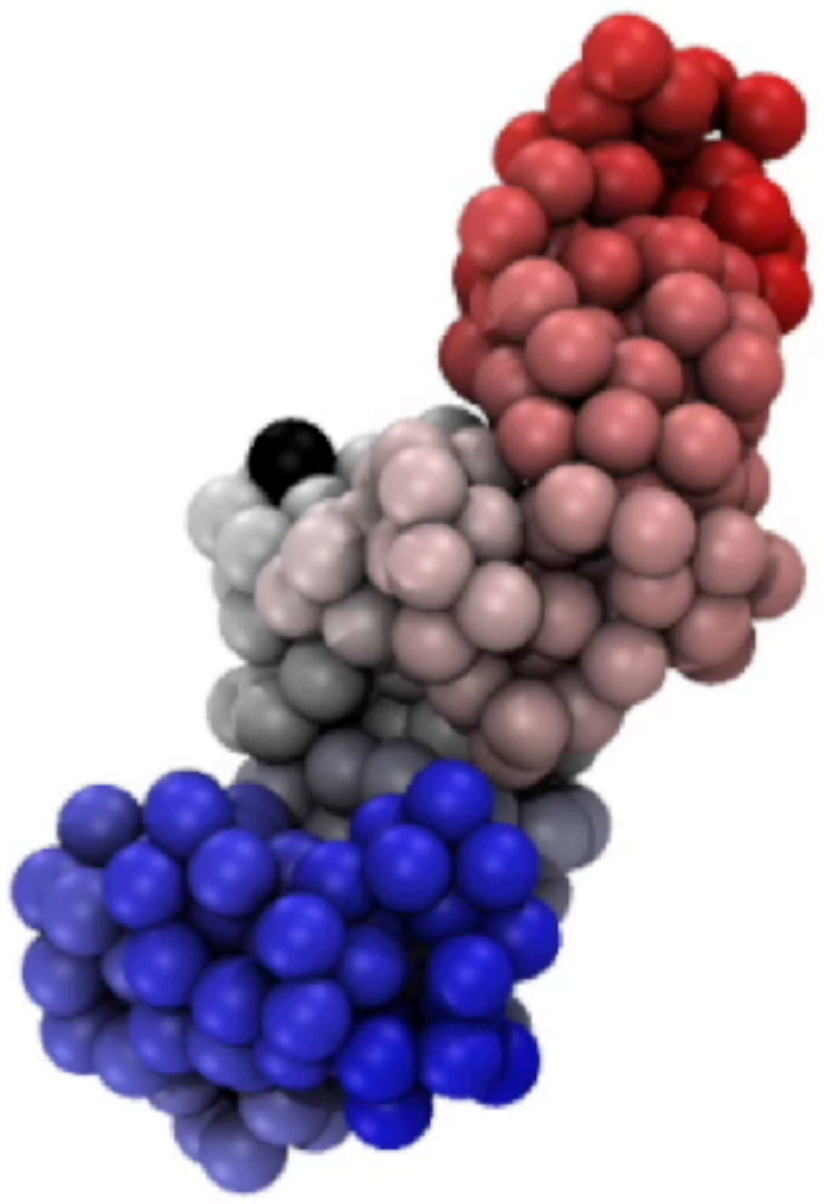


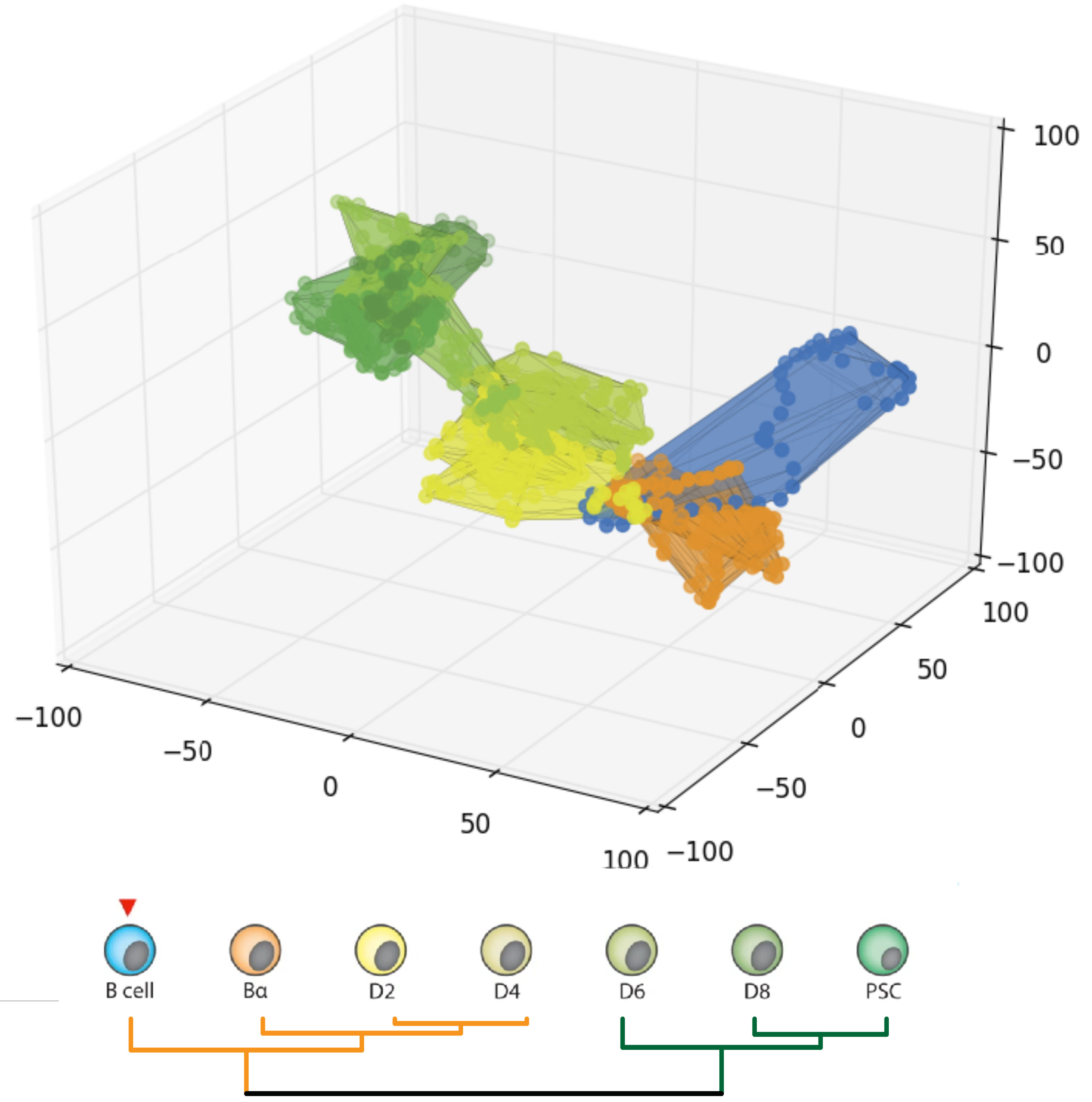




SOX2 locus dynamics changes from B to PSC

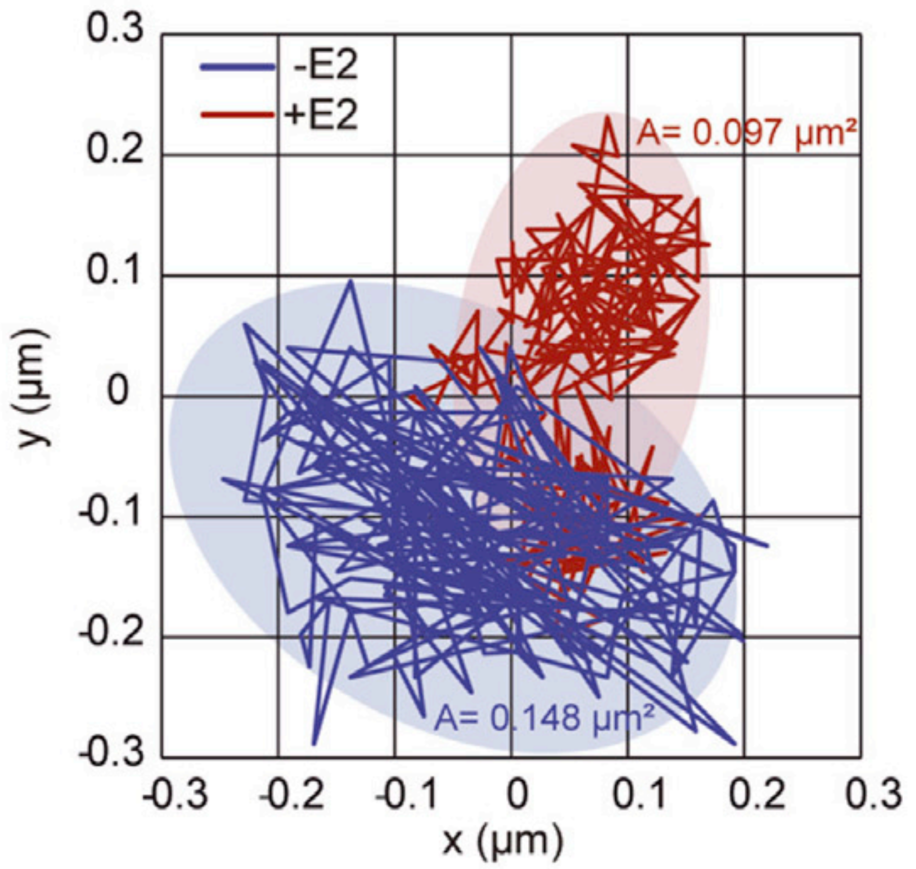
SOX2 displacement





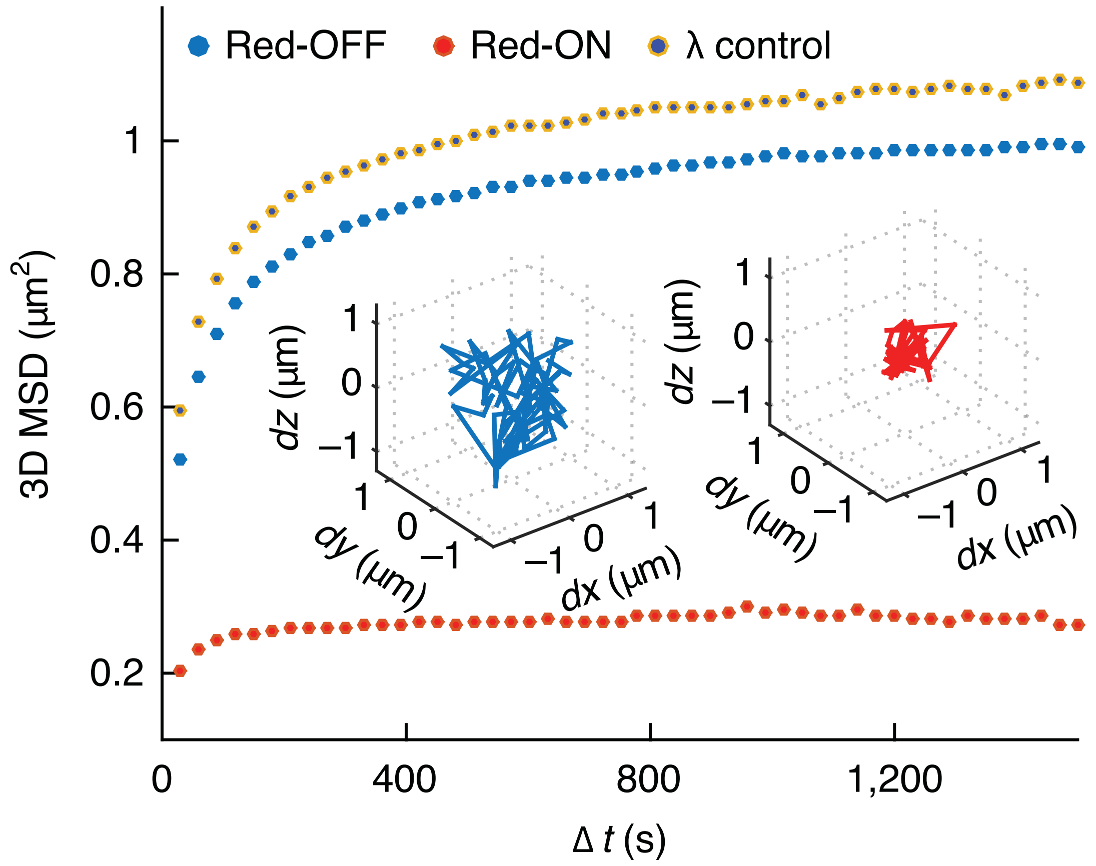
SOX2 locus dynamics changes from B to PSC

SOX2 displacement



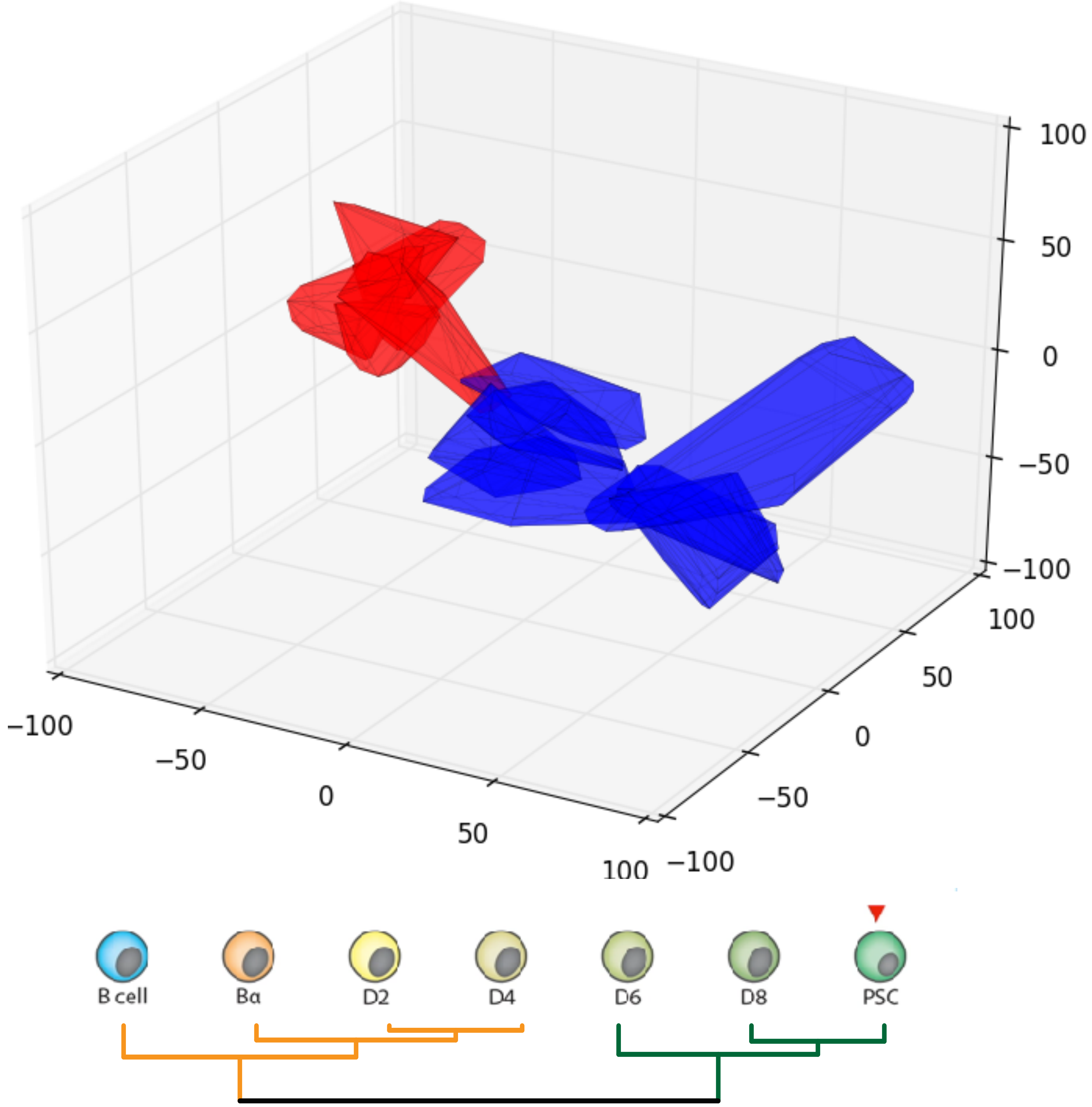
Two dimensional trajectories and area explored over 50s of the CCND1 locus recored before -E2 and after +E2 activation.

Germier ,T., et al, (2017) Blophys J.



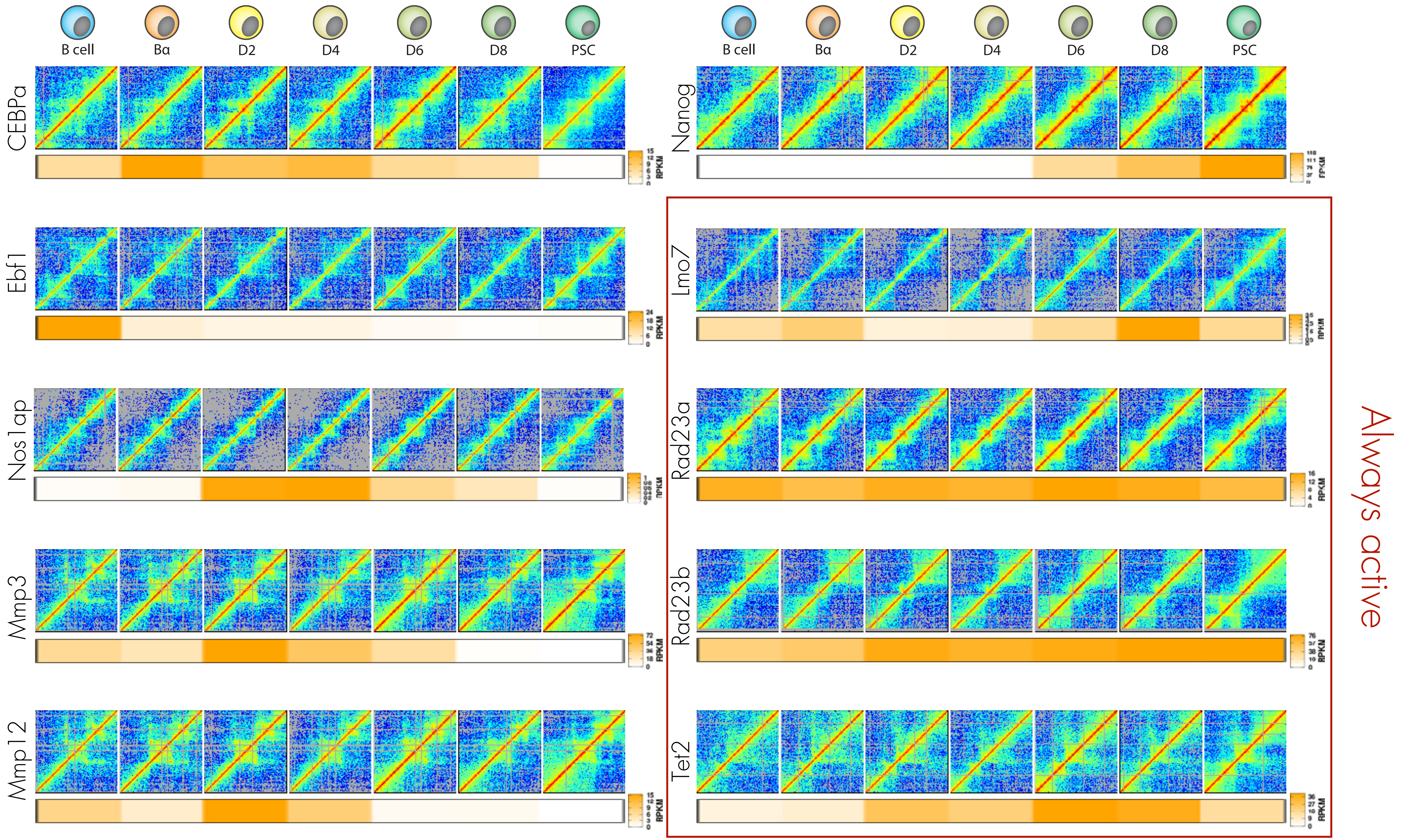
Transcription affects the 3D topology of the enhancer-promoted enhancing its temporal stability and is associated with further spatial compaction.

Chen ,T., et al, (2018) Nat. Genetics

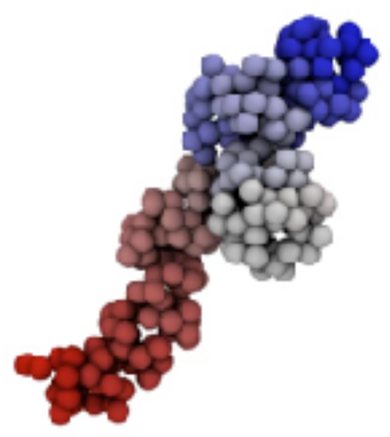


Structural changes from B to PSC

Other 10 loci

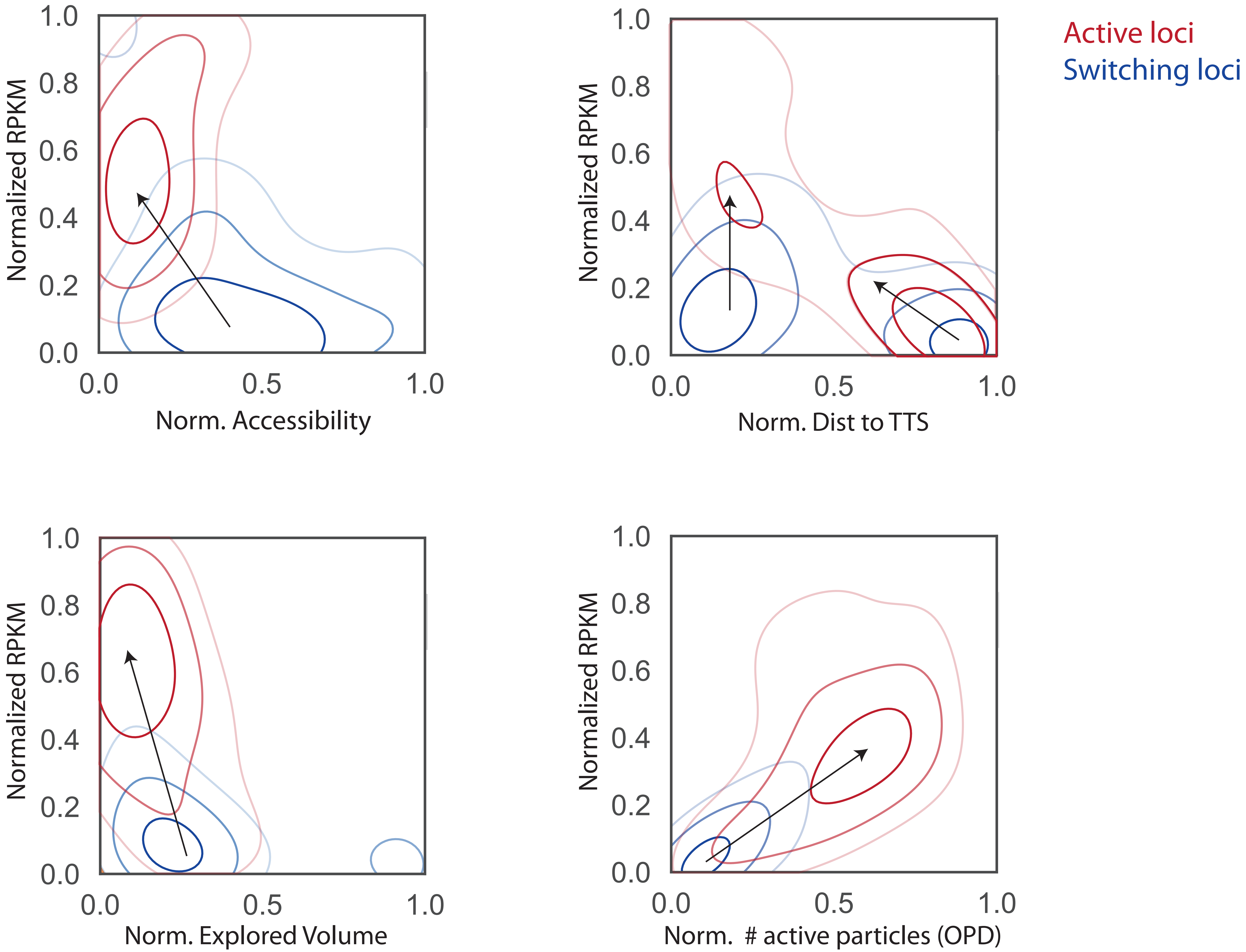


Switch

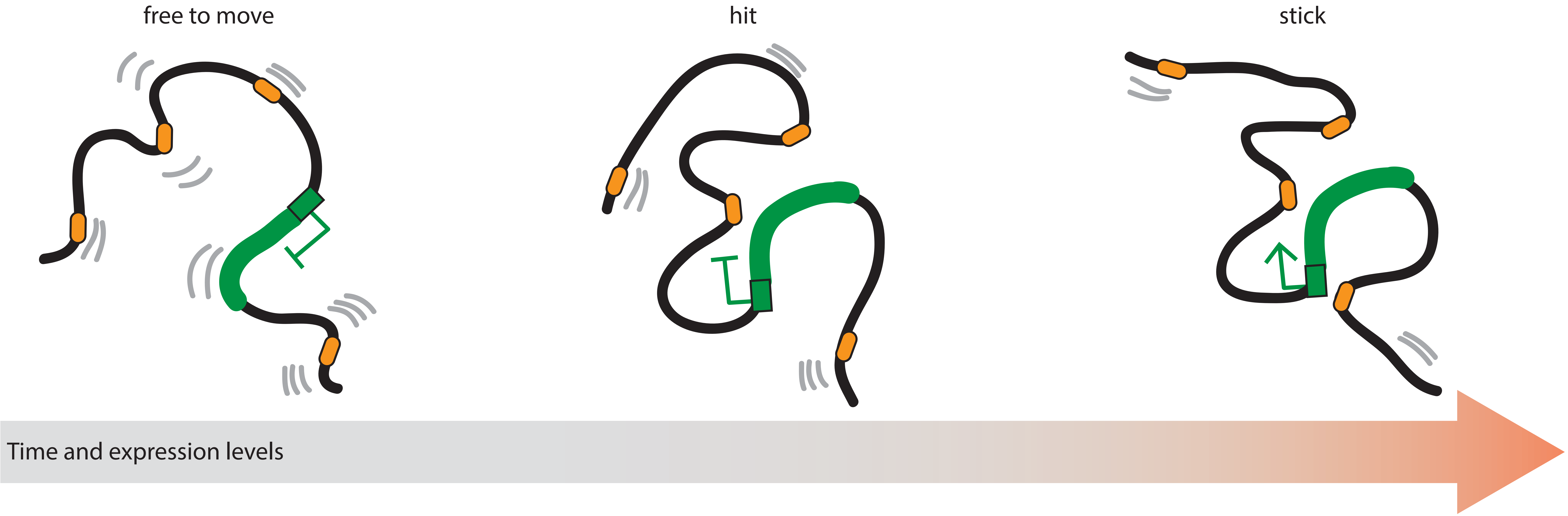


Dynamics of gene activation

Trends in all 11 loci



A "hit-and-stick" model for transcriptional activation



http://marciuslab.org
http://3DGenomes.org

@marciuslab
@mamartirenom

cnag
CRG Centre for Genomic Regulation
ICREA

David Castillo
Yasmina Cuartero
Marco Di Stefano
Irene Farabella
Silvia Galan
Mike Goodstadt
Maria Marti-Marimon
Francesca Mugianesi
Julen Mendieta
Juan Rodriguez
Paula Soler
Aleksandra Sparavier

.: Our current sponsors :.

erc
GOBIERNO DE ESPAÑA
MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Generalitat de Catalunya
Fundació La Marató de TV3
Fundación Bancaria "la Caixa"