
HUMAN "The Movie" by Yann Arthus-Bertrand

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

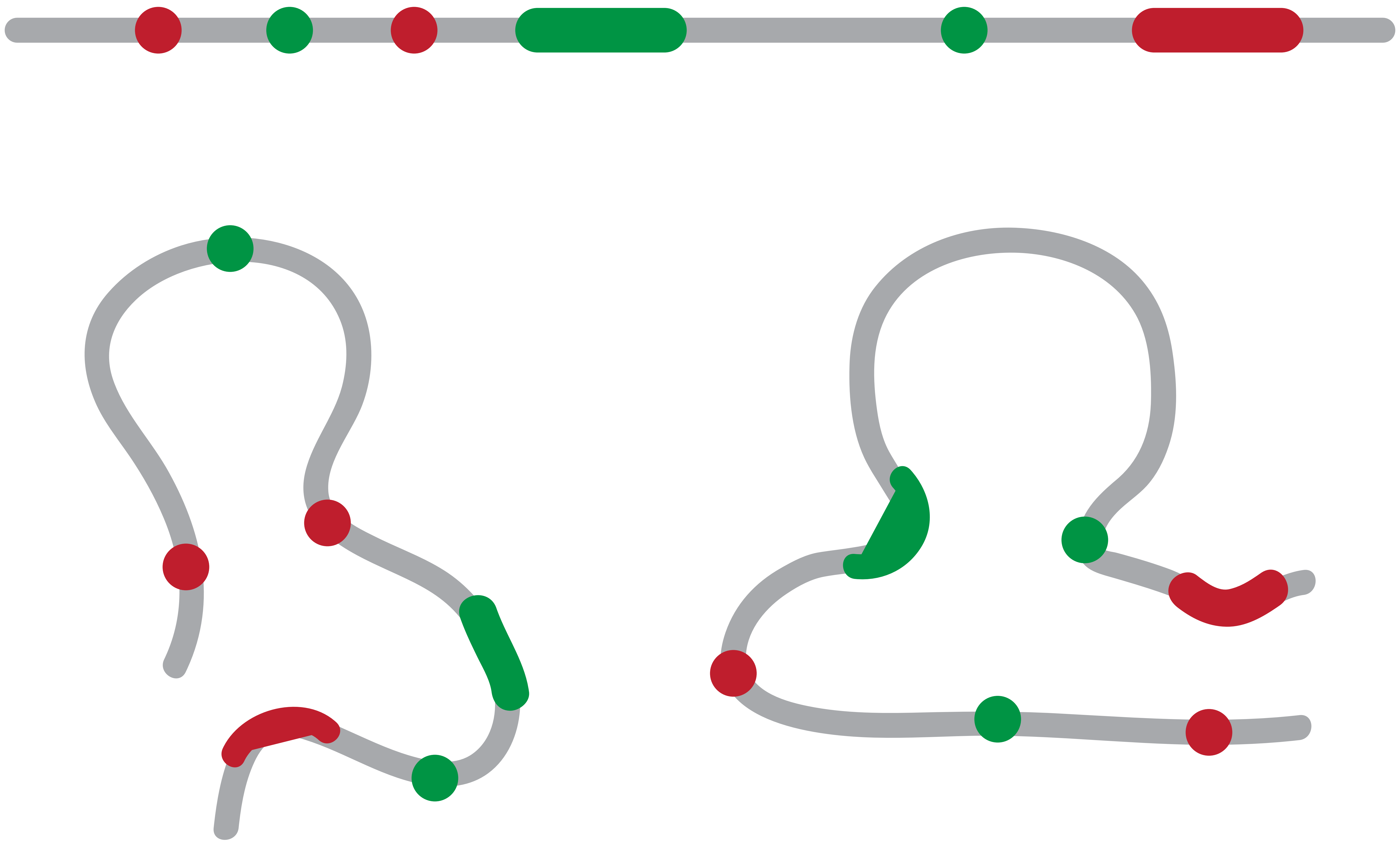
All you will see in the screen is here:

http://sgt.cnag.cat/www/presentations/files/slides/20190919_CD_SummerSchool.pdf

I encourage you to:

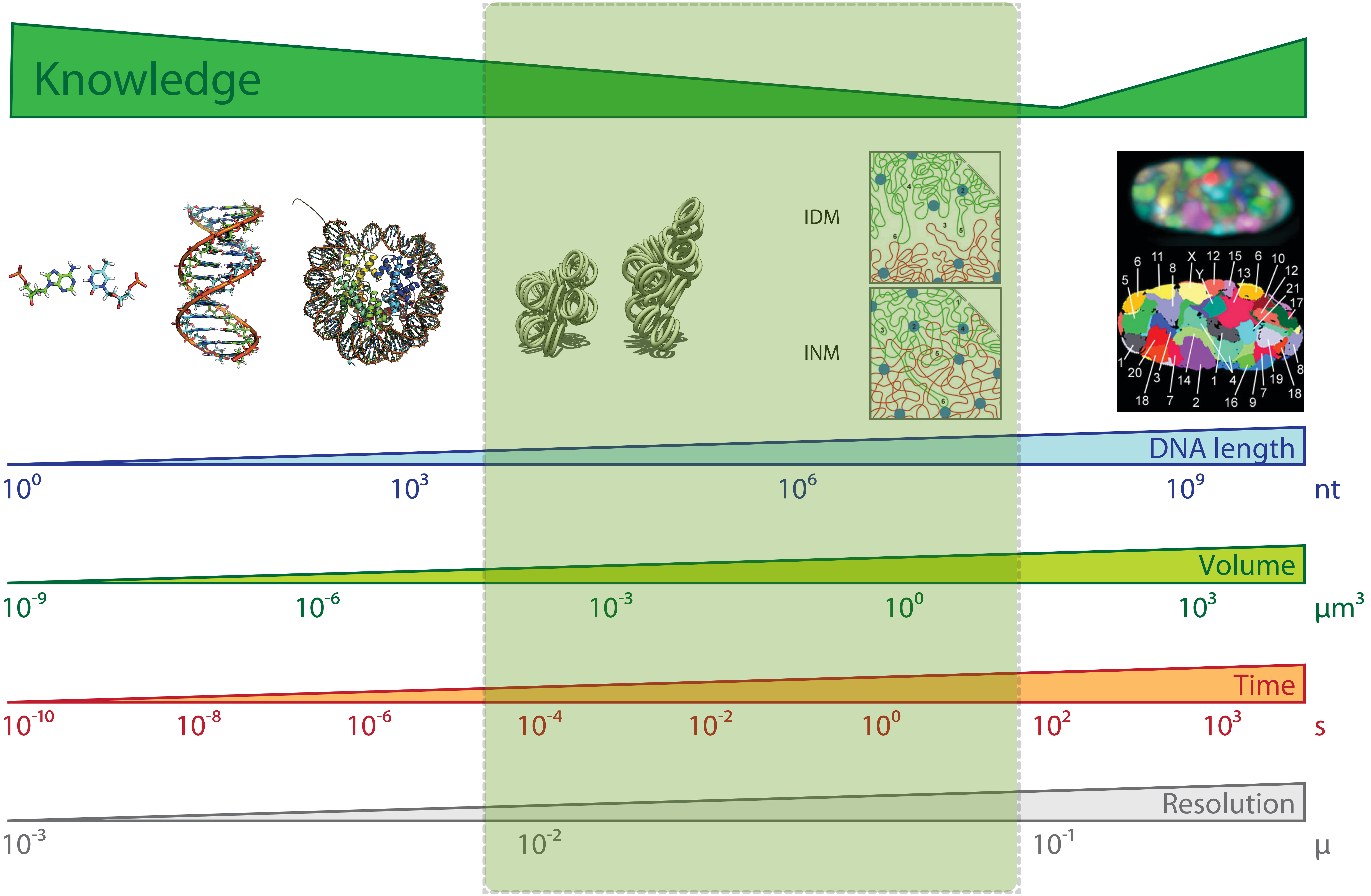
listen AND speak

not necessarily in this order… 😁



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

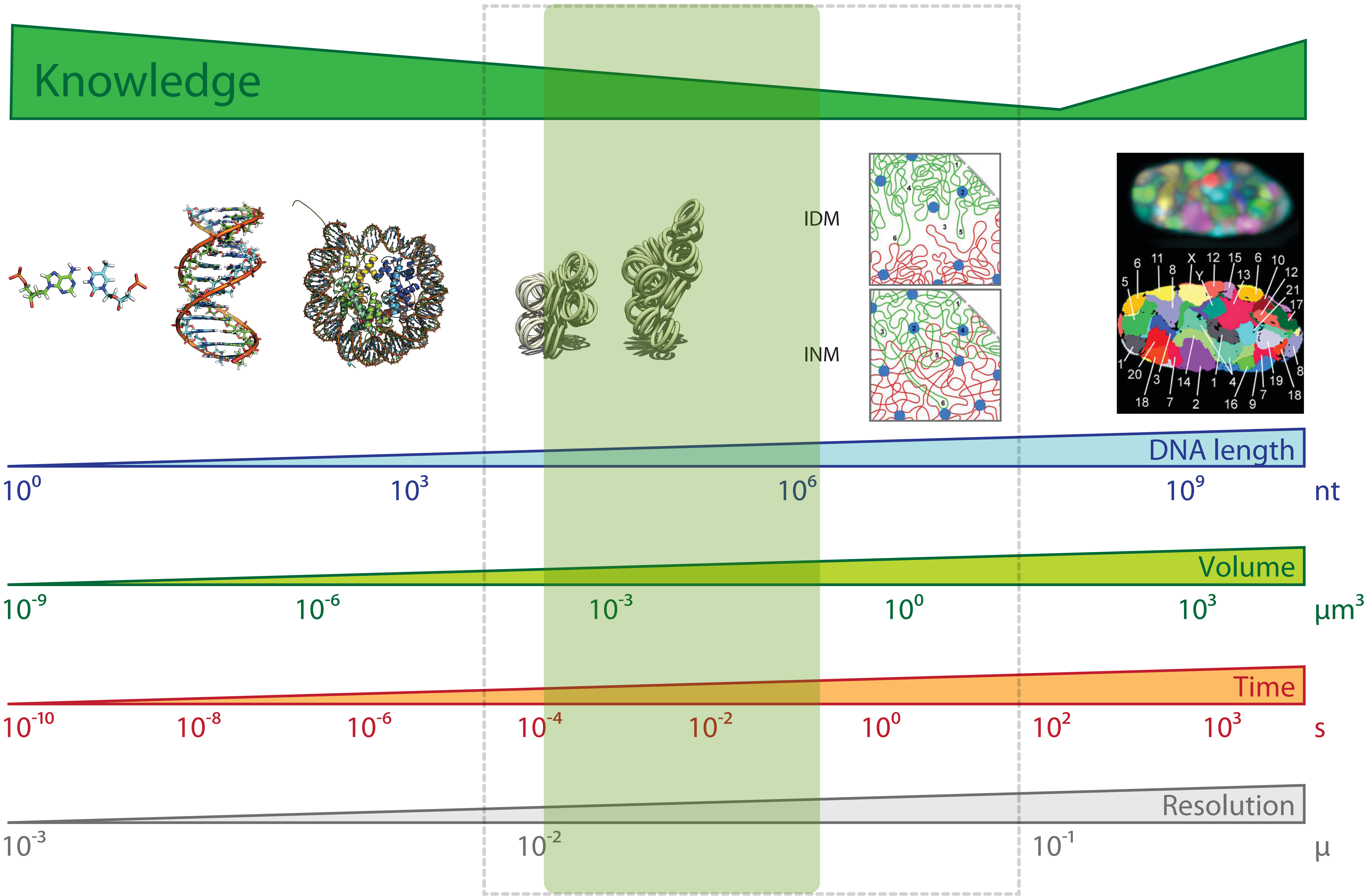
10^{-10} 10^{-8} 10^{-6} 10^{-4} 10^{-2} 10^0 10^2 10^3 s

Resolution

10^{-3} 10^{-2} 10^{-1} μ

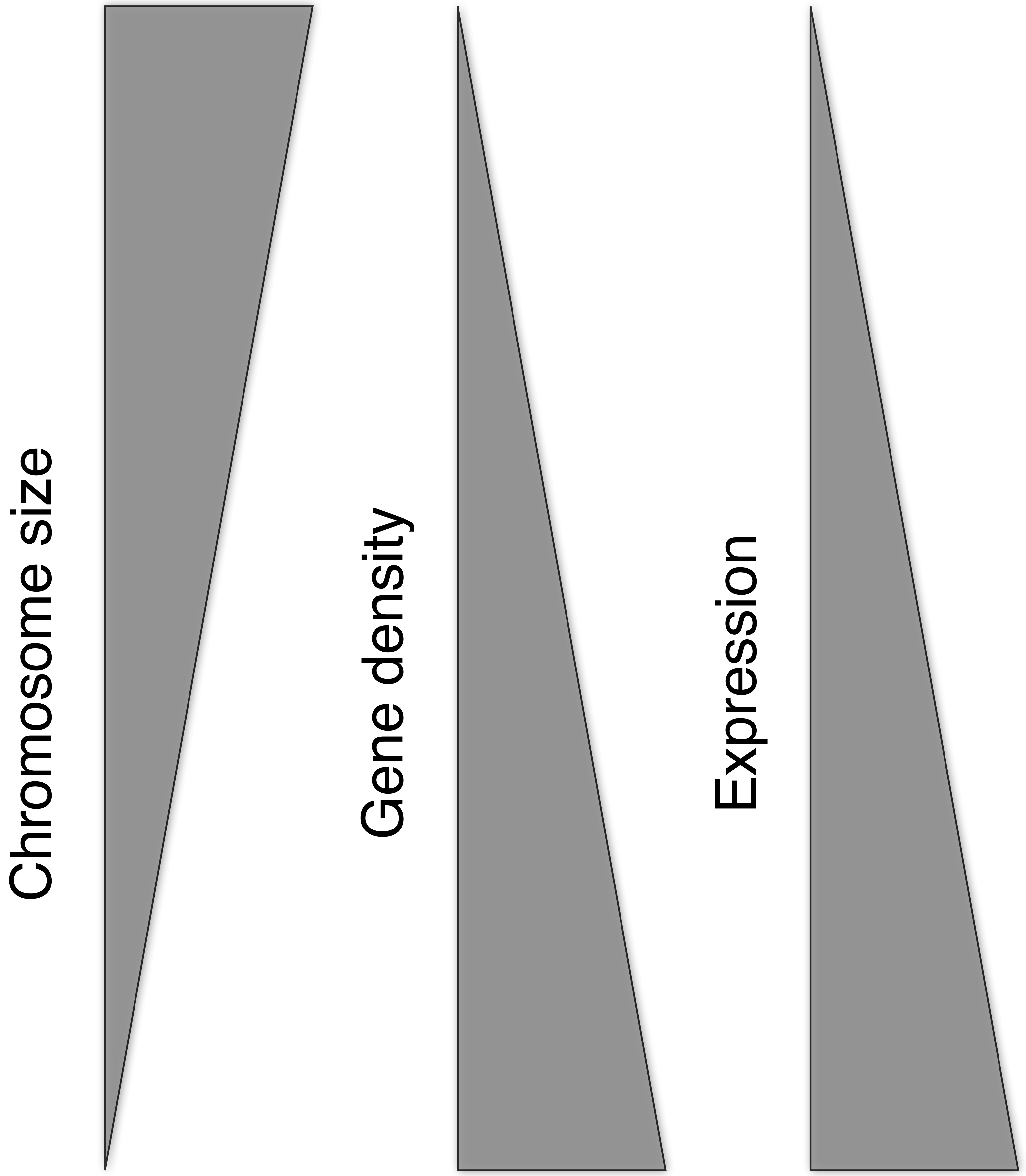
Resolution Gap

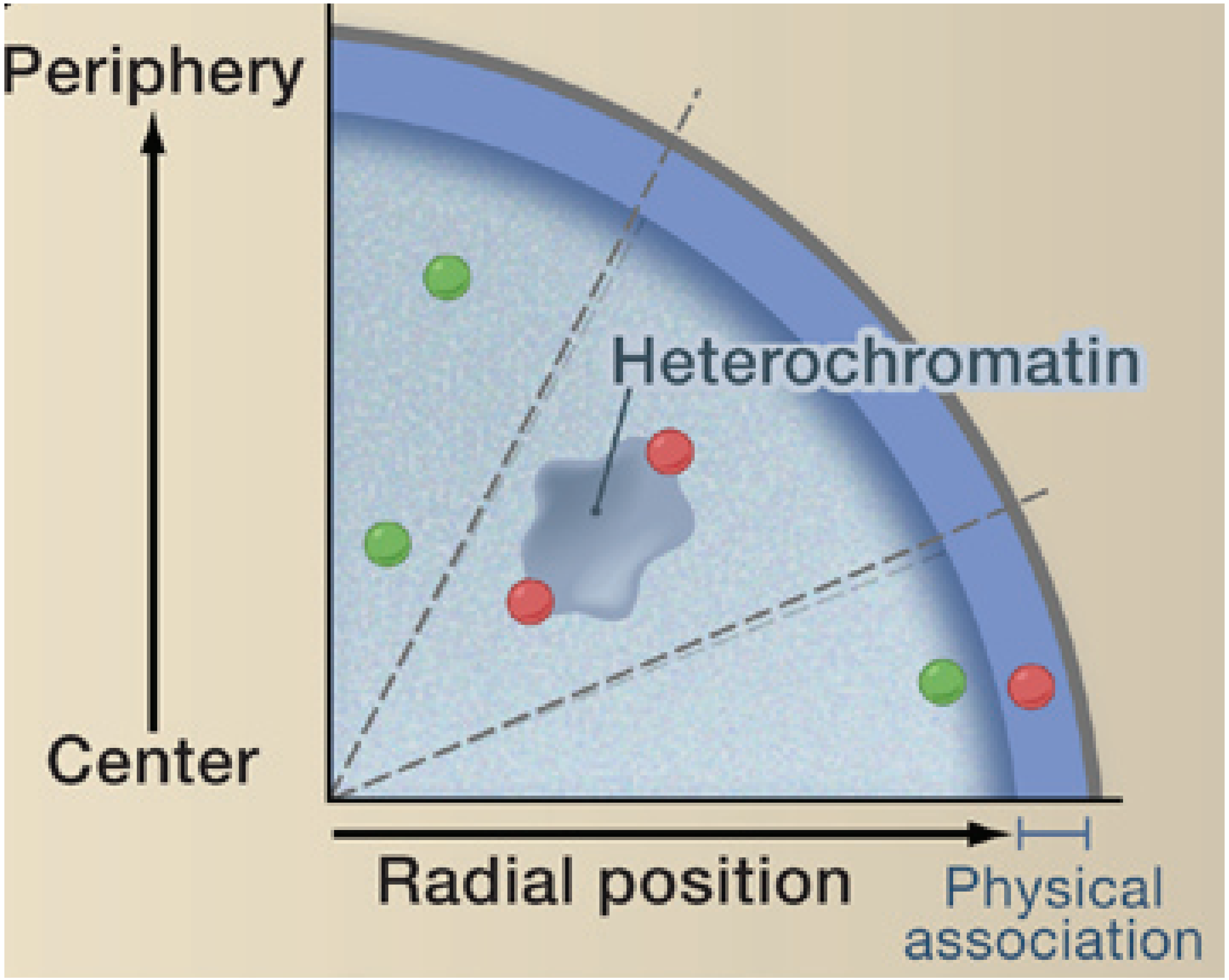
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol 7, e1002125 (2011)



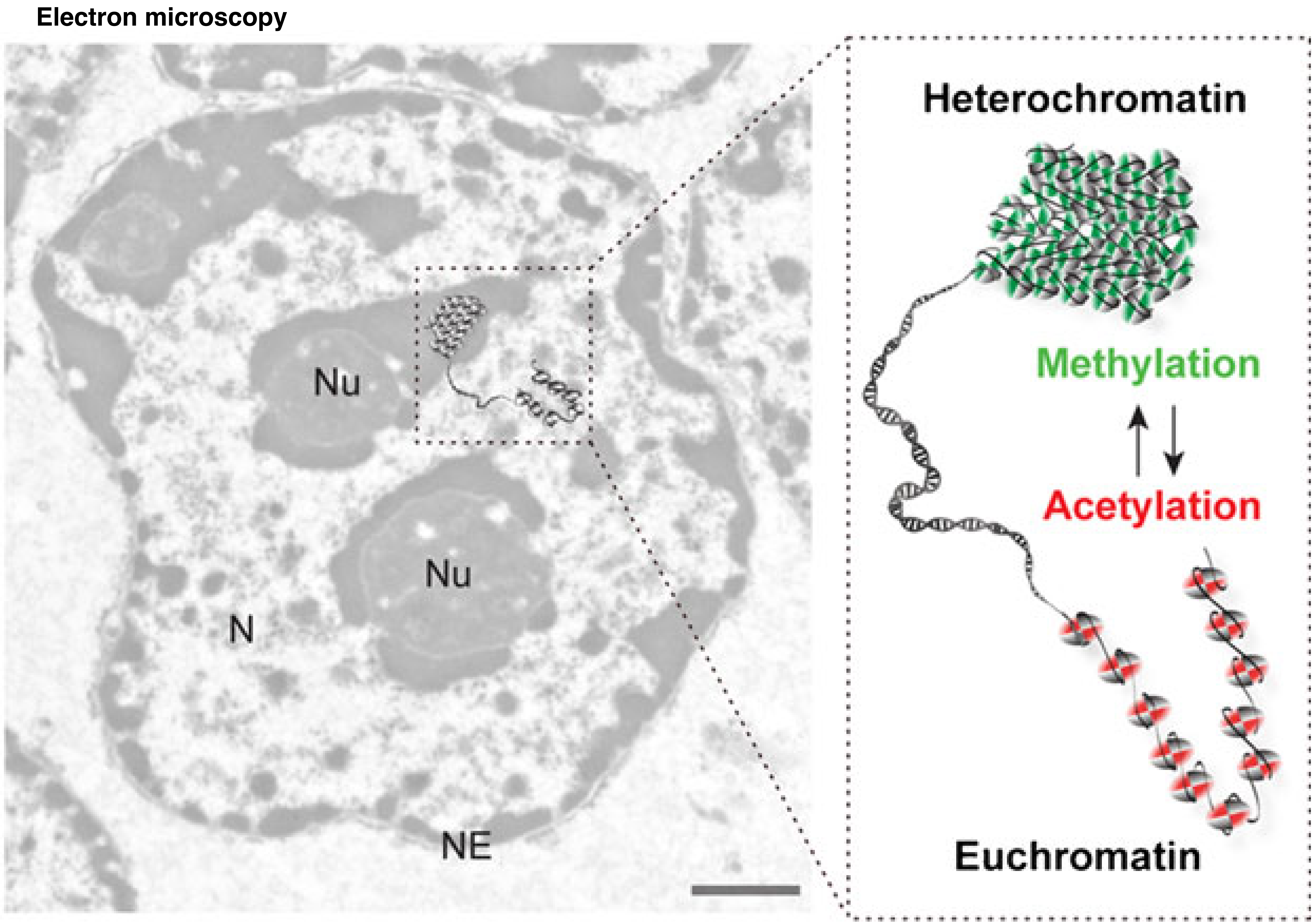
Level I: Radial genome organization

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

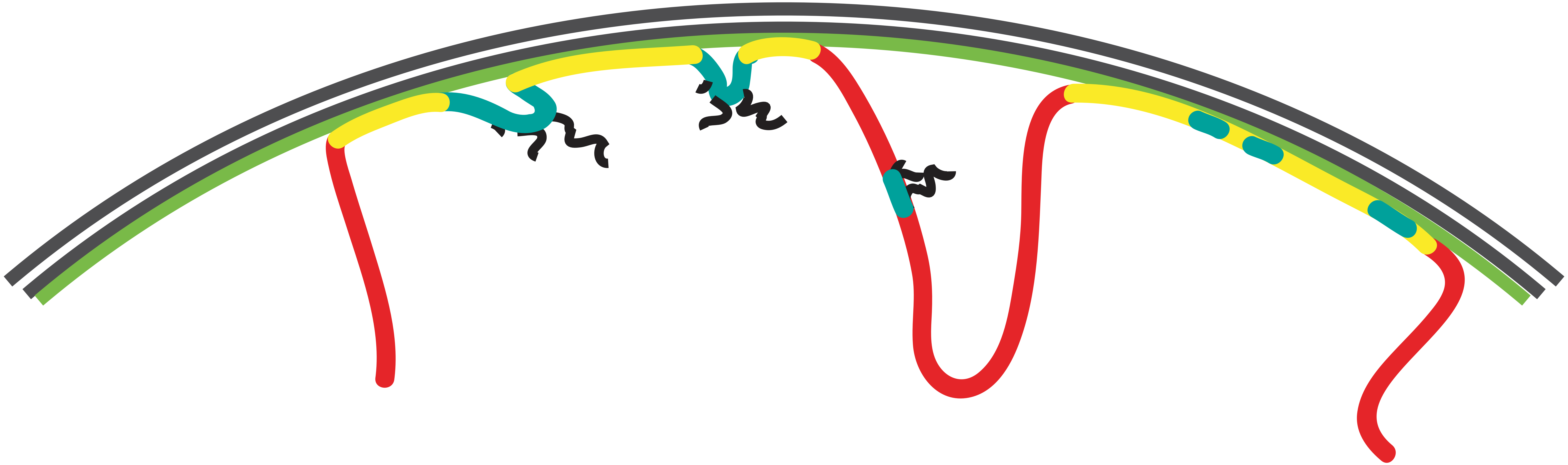




Level II: Euchromatin vs heterochromatin



Level III: Lamina-genome interactions



⬛ nuclear membrane
🟩 nuclear lamina

🟥 internal chromatin (mostly active)
🟨 lamina-associated domains (repressed)

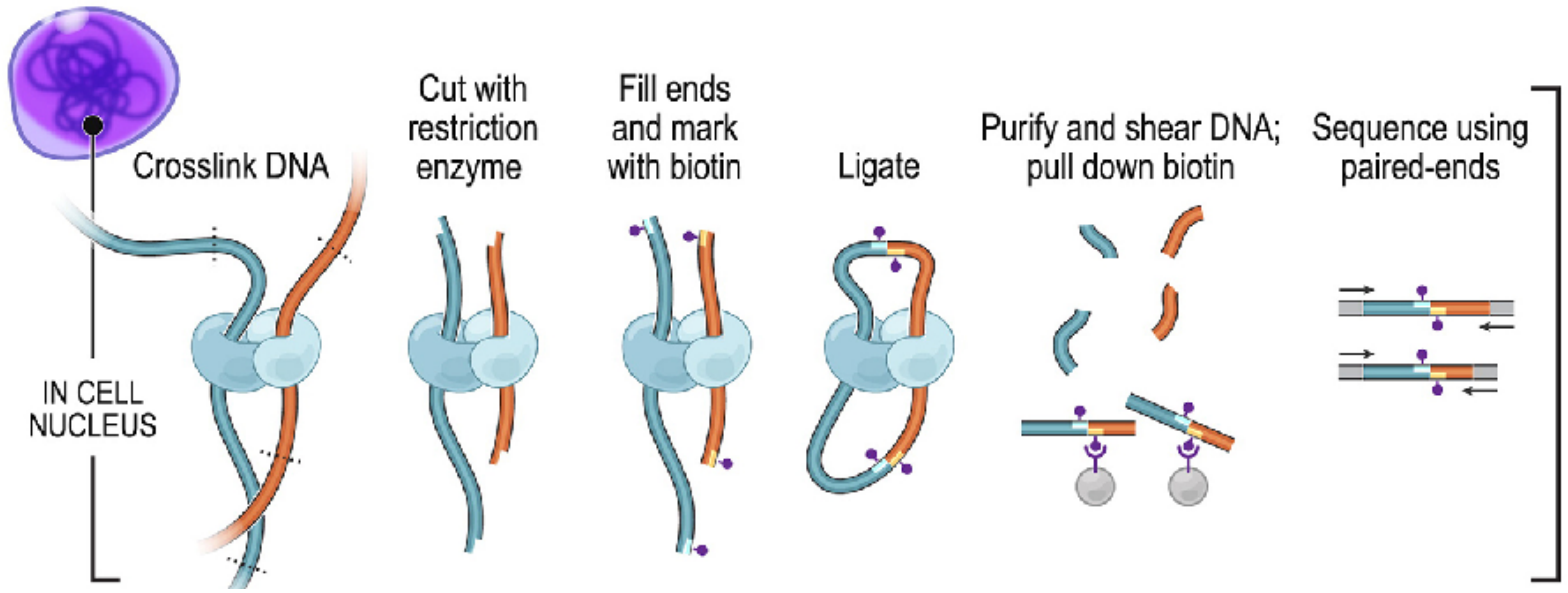
🟦 Genes

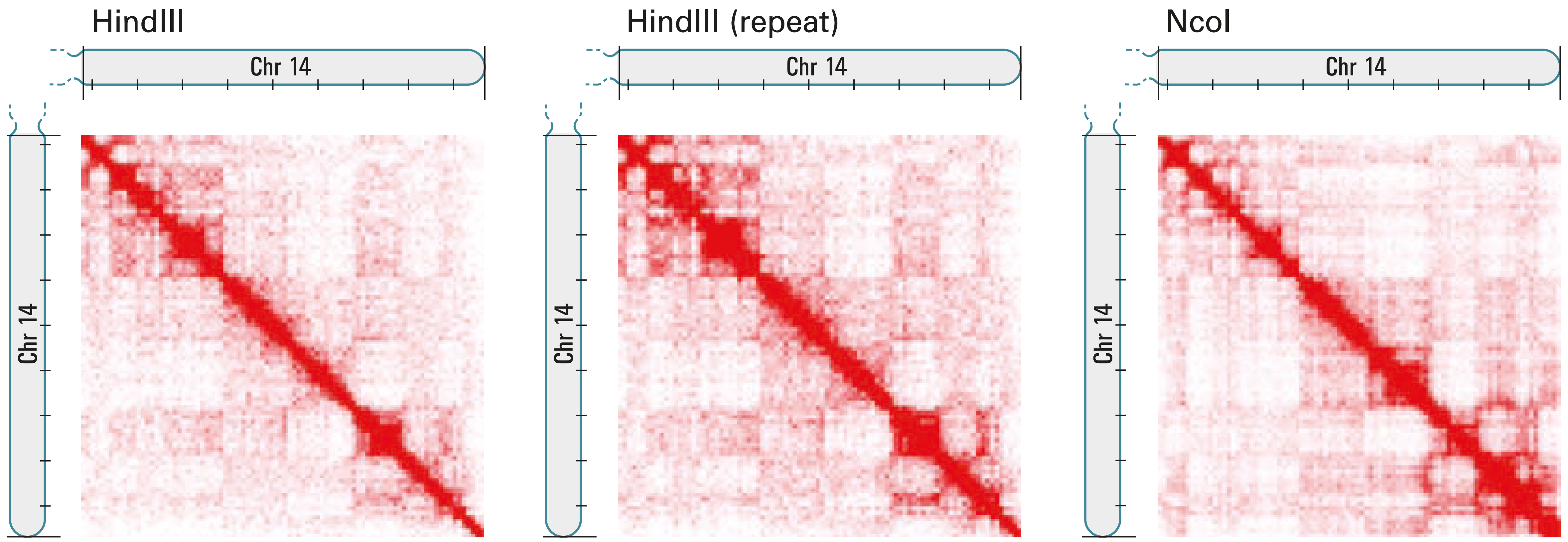
〰 mRNA

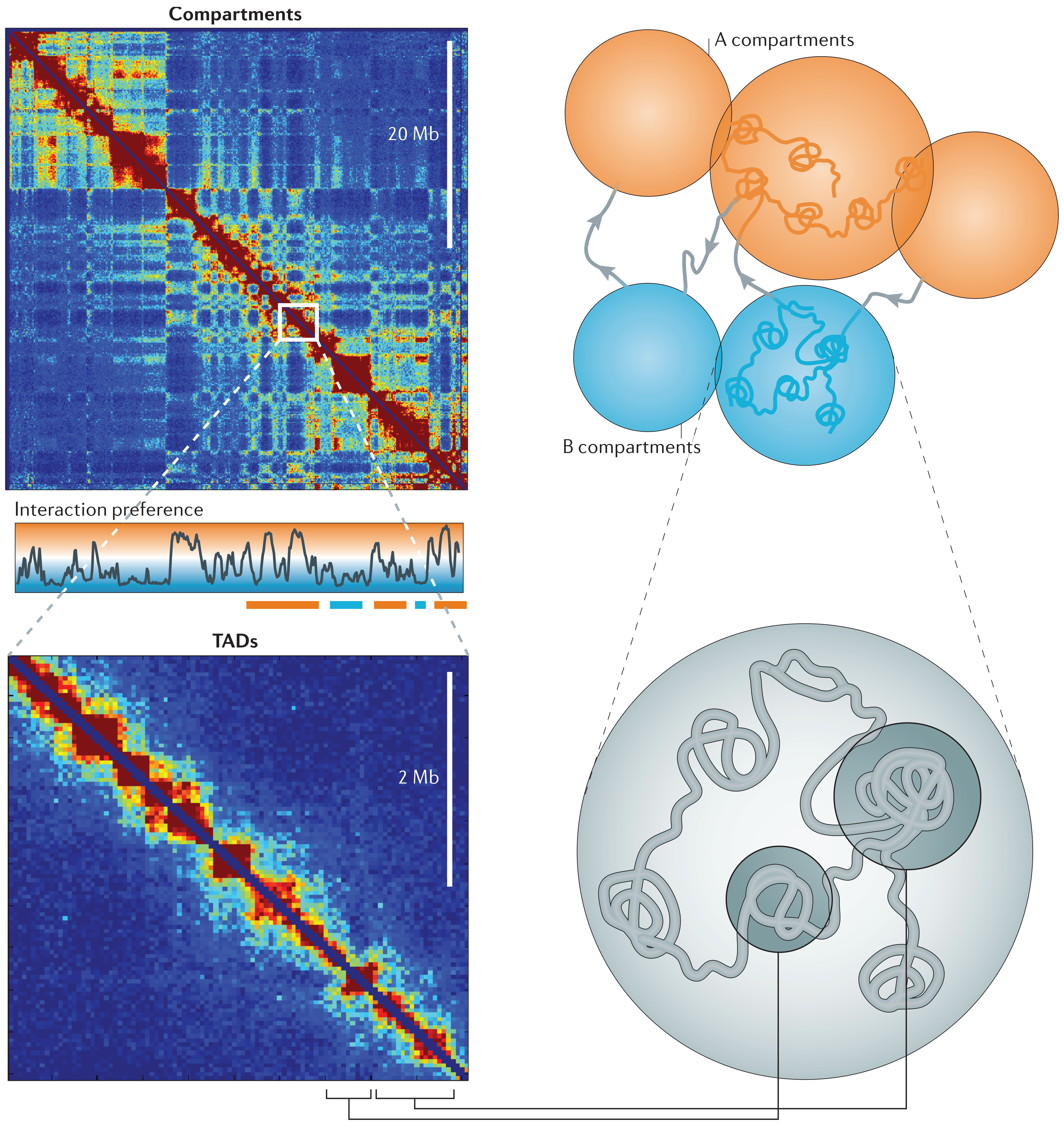
Adapted from Molecular Cell 38, 603-613, 2010

Level IV: Higher-order organization

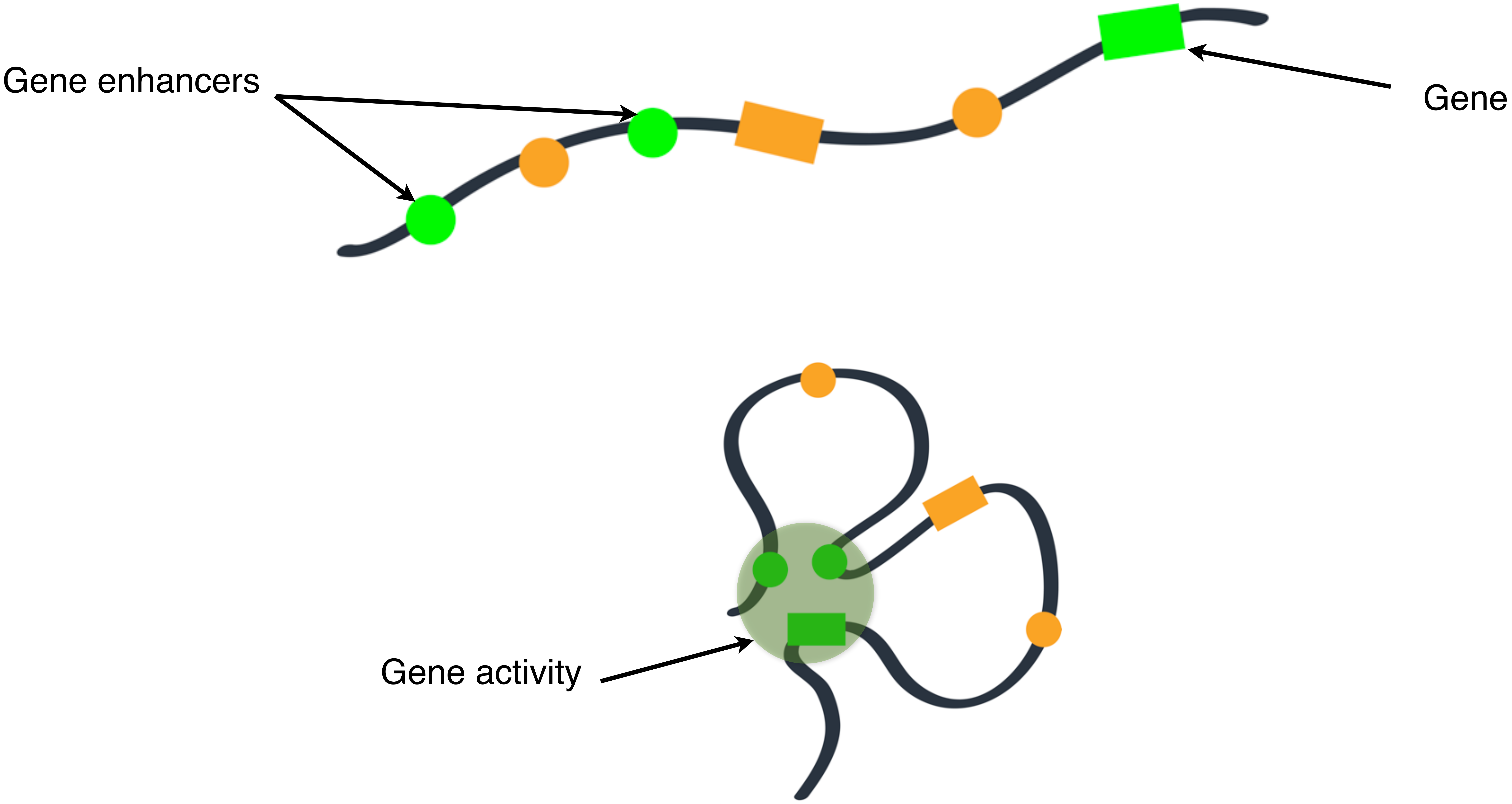
Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Nat Rev Genet 14, 390–403 (2013).







Level V: Chromatin loops



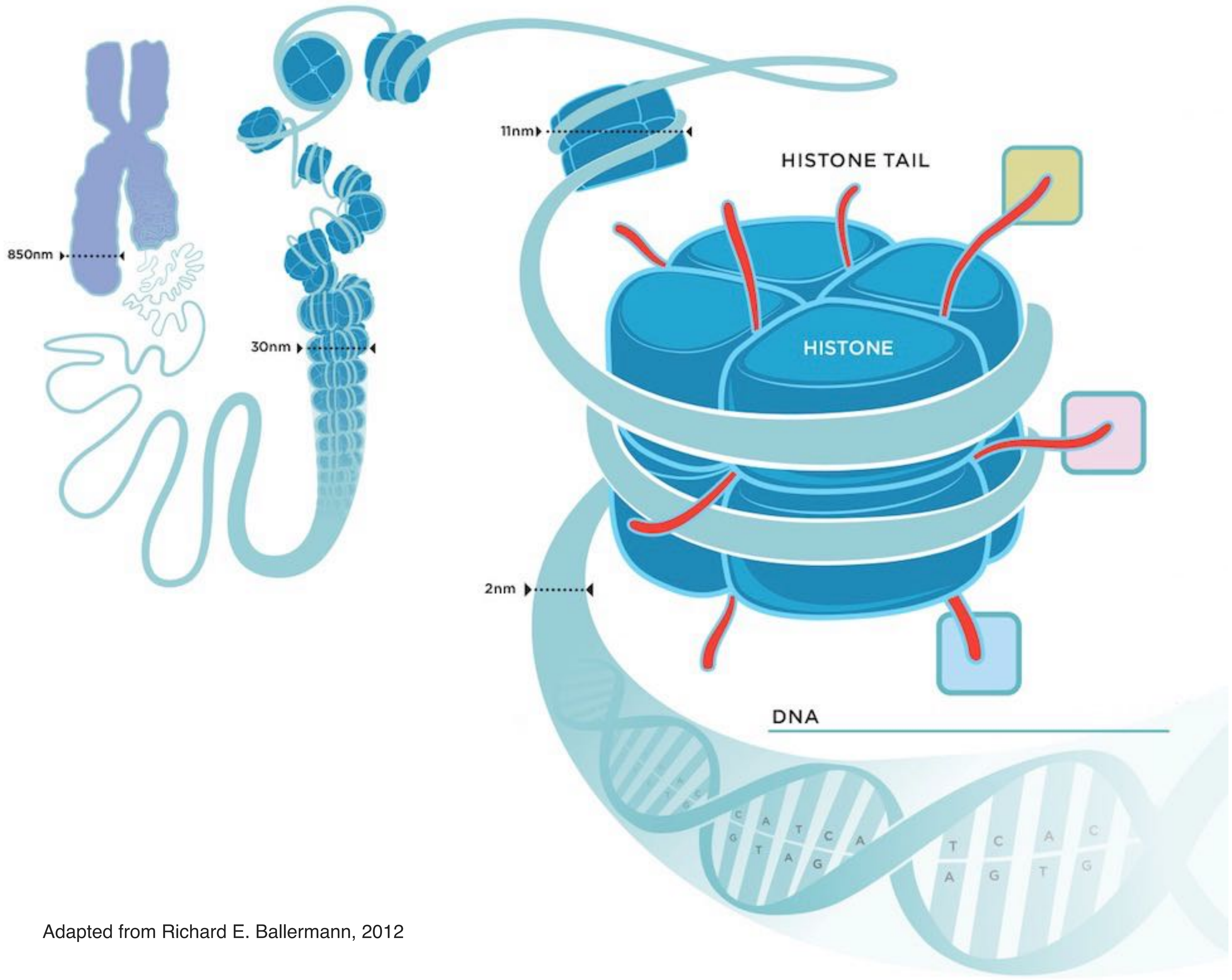
Gene enhancers

Gene

Gene activity

Level VI: Nucleosome

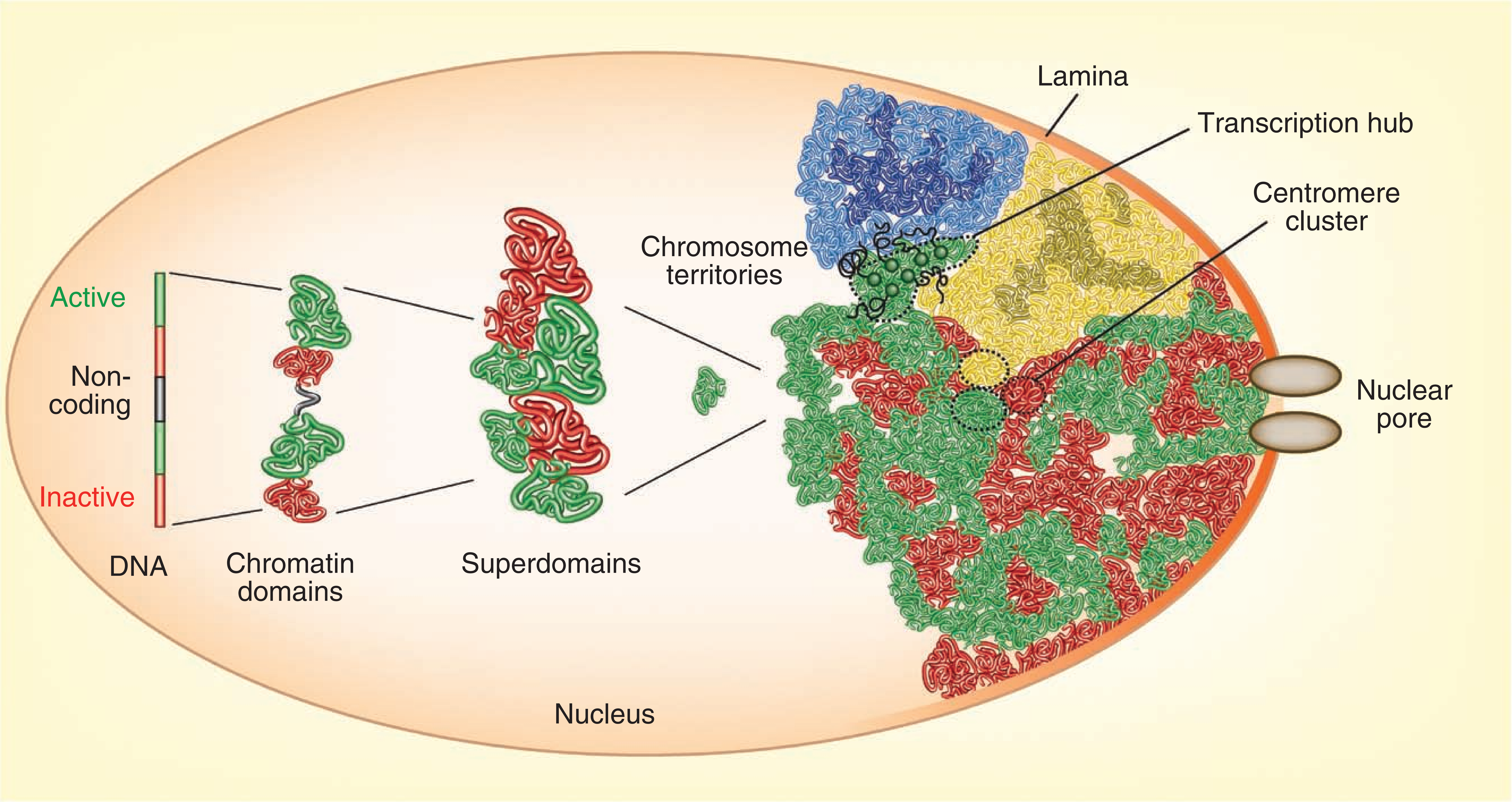
Chromosome **Chromatin fibre** **Nucleosome**



Adapted from Richard E. Ballermann, 2012

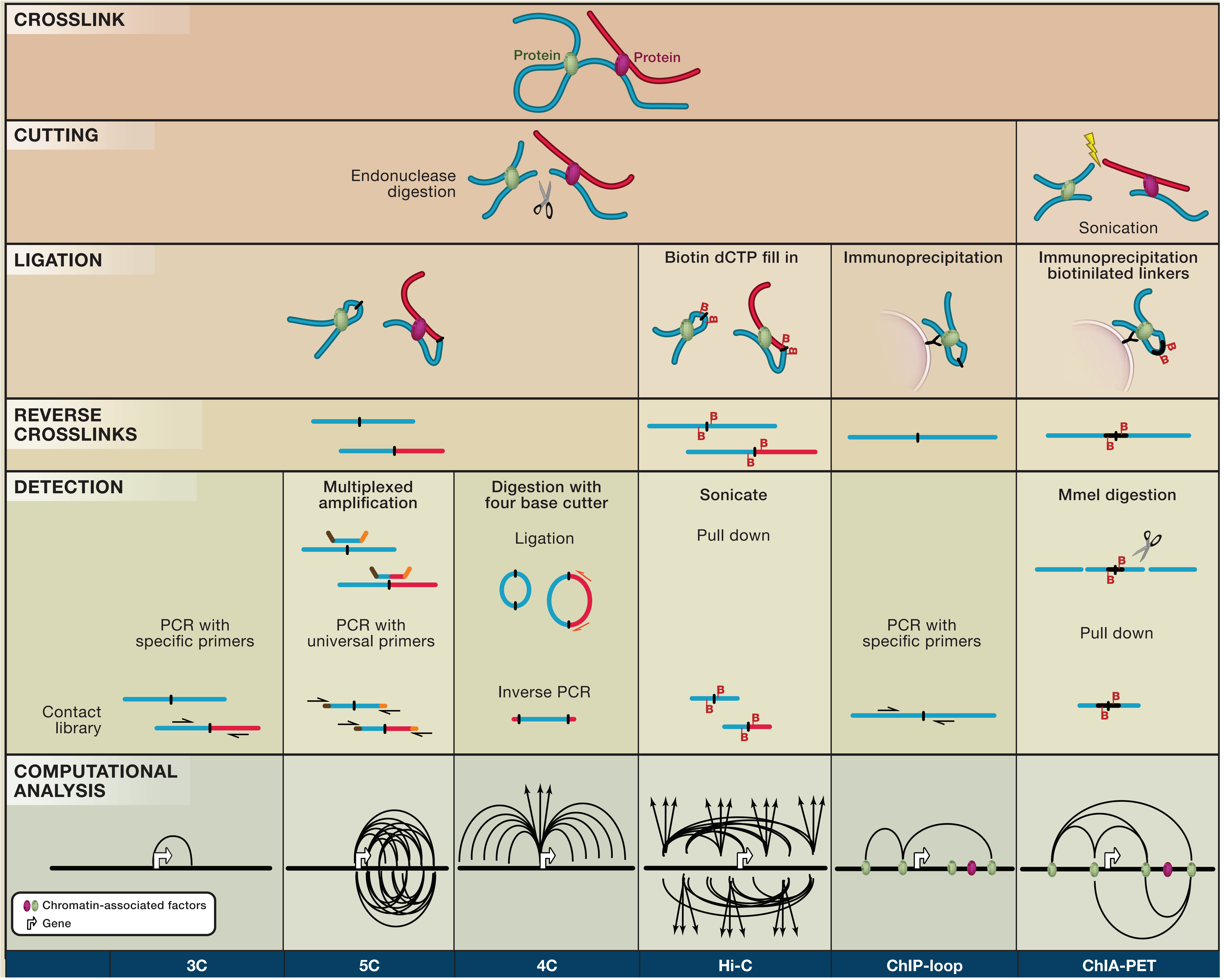
Complex genome organization

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

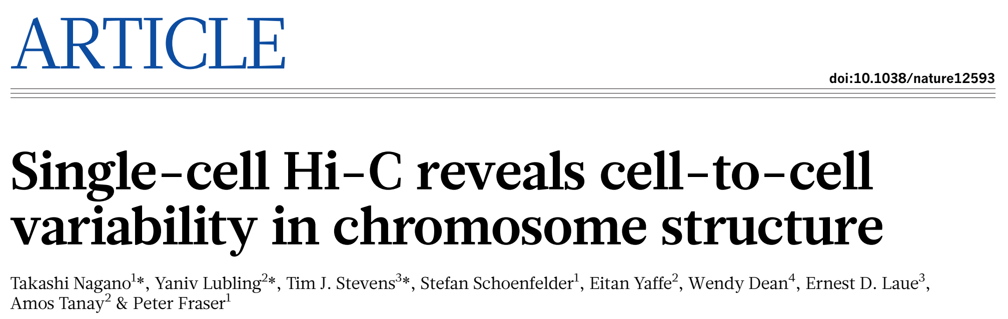


Marina Corral

Chromosome Conformation Capture



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



ARTICLE

doi:10.1038/nature12593

Single-cell Hi-C reveals cell-to-cell variability in chromosome structure

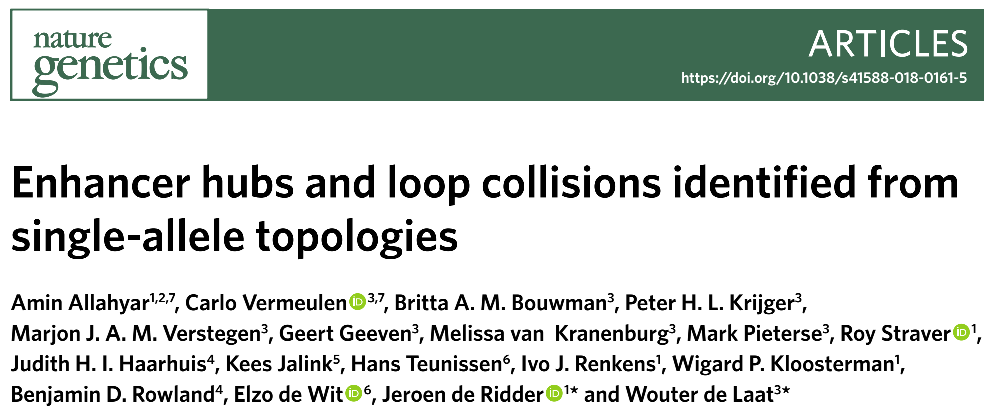
Takashi Nagano[1]*, Yaniv Lubling[2]*, Tim J. Stevens[3]*, Stefan Schoenfelder[1], Eitan Yaffe[2], Wendy Dean[4], Ernest D. Laue[3], Amos Tanay[2] & Peter Fraser[1]

LETTER

doi:10.1038/nature20158

Capturing pairwise and multi-way chromosomal conformations using chromosomal walks

Pedro Olivares-Chauvet[1], Zohar Mukamel[1], Aviezer Lifshitz[1], Omer Schwartzman[1], Noa Oded Elkayam[1], Yaniv Lubling[1], Gintaras Deikus[2], Robert P. Sebra[2] & Amos Tanay[1]



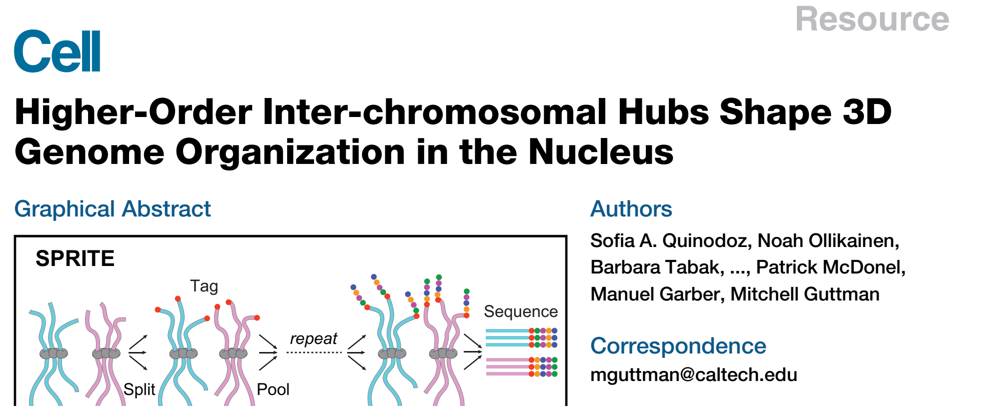
nature genetics

ARTICLES

https://doi.org/10.1038/s41588-018-0161-5

Enhancer hubs and loop collisions identified from single-allele topologies

Amin Allahyar[1,2,7], Carlo Vermeulen[3,7], Britta A. M. Bouwman[3], Peter H. L. Krijger[3], Marjon J. A. M. Verstegen[3], Geert Geeven[3], Melissa van Kranenburg[3], Mark Pieterse[3], Roy Straver[1], Judith H. I. Haarhuis[4], Kees Jalink[5], Hans Teunissen[6], Ivo J. Renkens[1], Wigard P. Kloosterman[1], Benjamin D. Rowland[4], Elzo de Wit[6], Jeroen de Ridder[1]* and Wouter de Laat[3]*



Cell

Resource

Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus

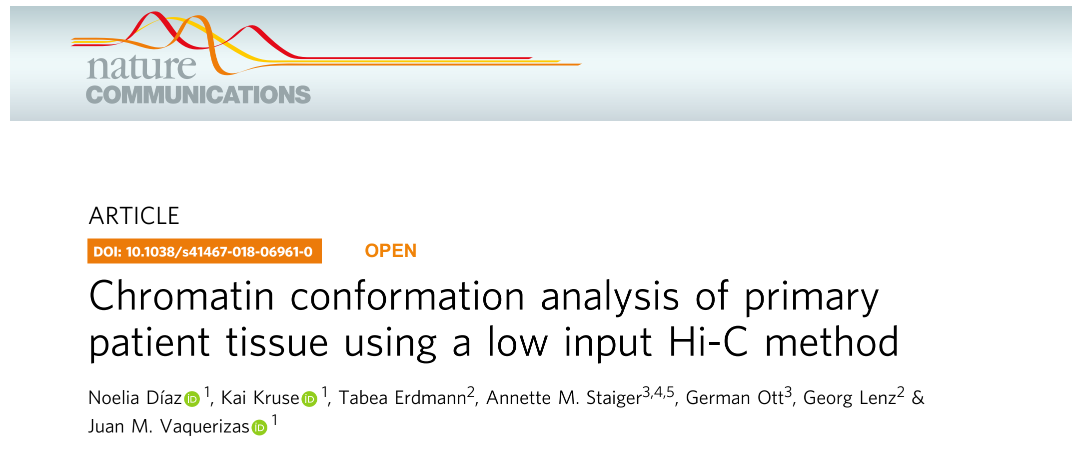
Graphical Abstract

Authors

Sofia A. Quinodoz, Noah Ollikainen, Barbara Tabak, ..., Patrick McDonel, Manuel Garber, Mitchell Guttman

Correspondence

mguttman@caltech.edu



nature COMMUNICATIONS

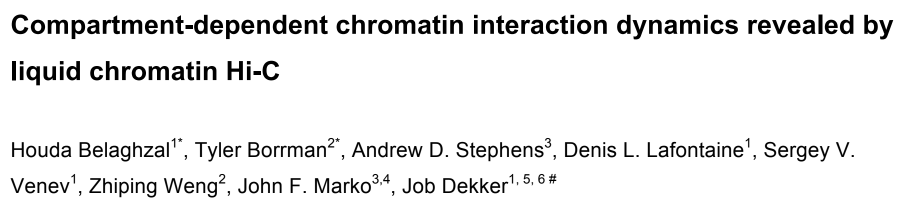
ARTICLE

DOI: 10.1038/s41467-018-06961-0

OPEN

Chromatin conformation analysis of primary patient tissue using a low input Hi-C method

Noelia Díaz[1], Kai Kruse[1], Tabea Erdmann[2], Annette M. Staiger[3,4,5], German Ott[3], Georg Lenz[2] & Juan M. Vaquerizas[1]

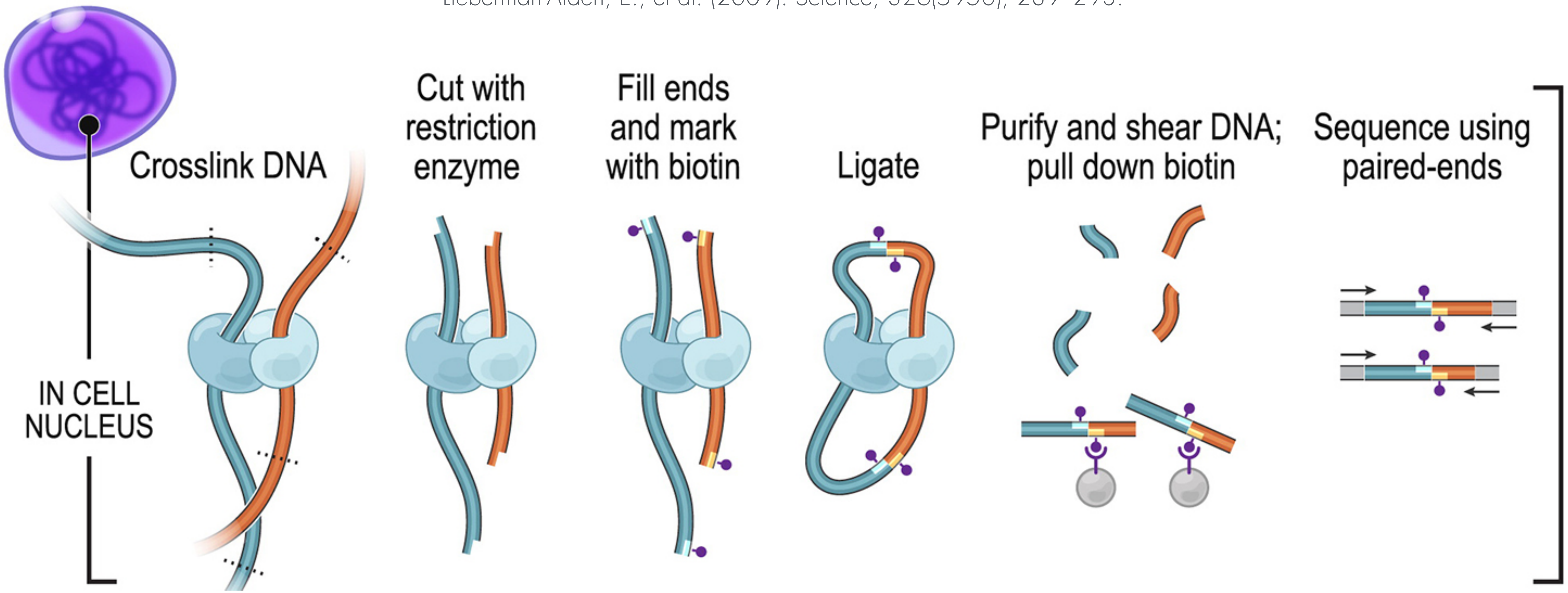


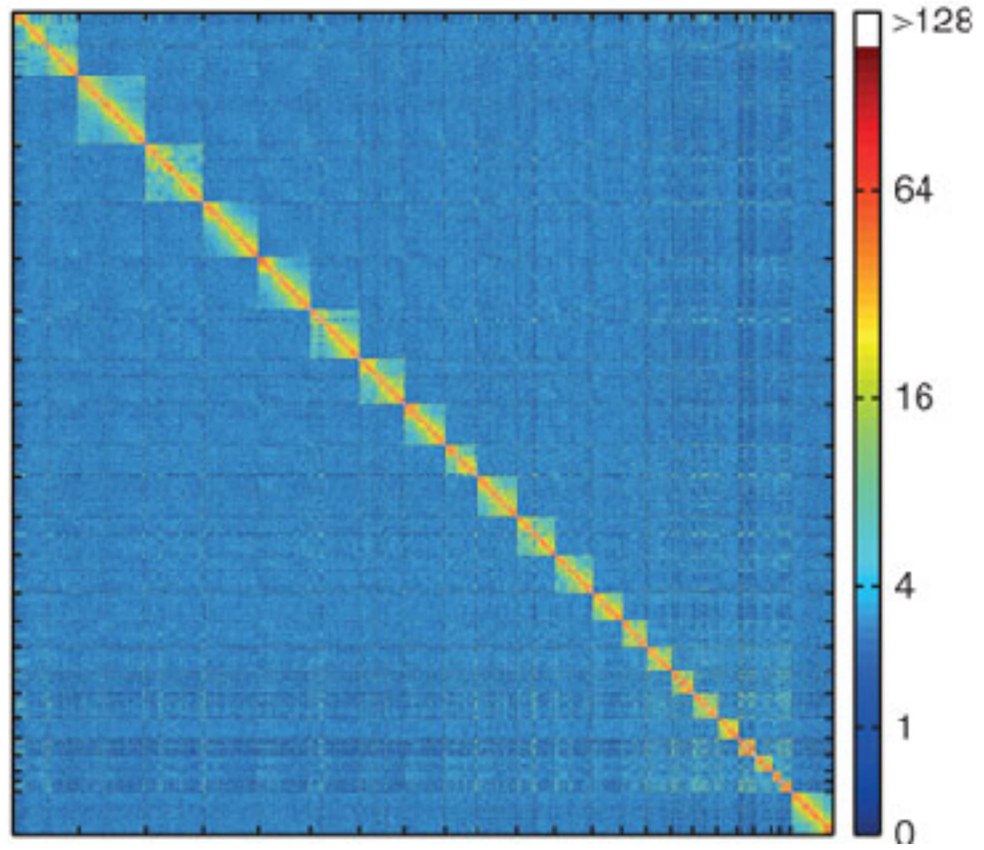
Compartment-dependent chromatin interaction dynamics revealed by liquid chromatin Hi-C

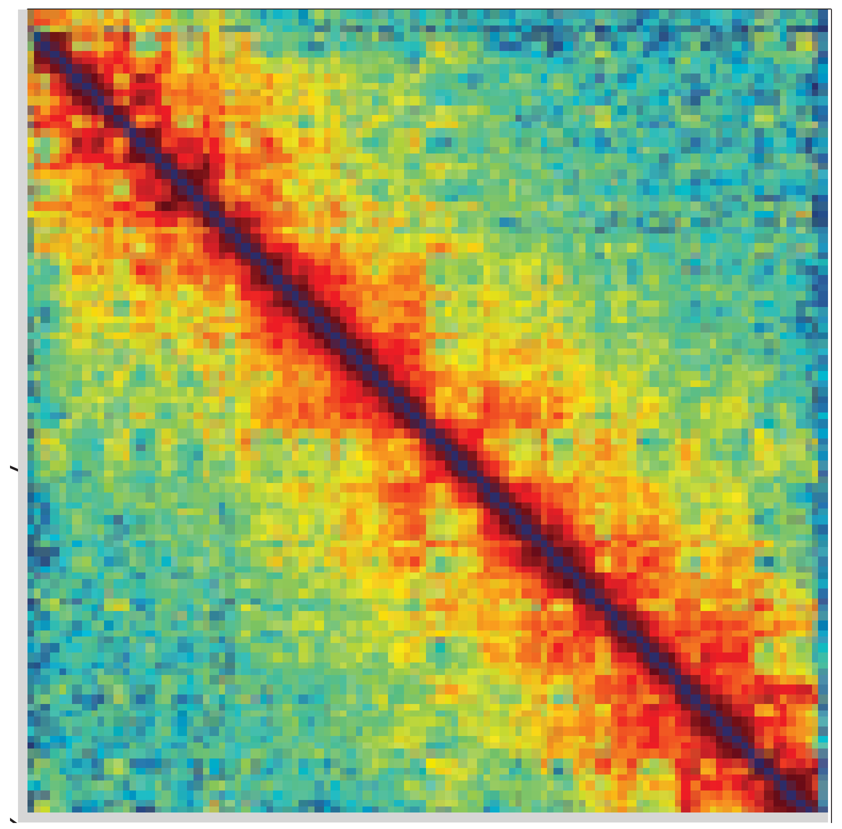
Houda Belaghzal[1*], Tyler Borrman[2*], Andrew D. Stephens[3], Denis L. Lafontaine[1], Sergey V. Venev[1], Zhiping Weng[2], John F. Marko[3,4], Job Dekker[1, 5, 6 #]

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.

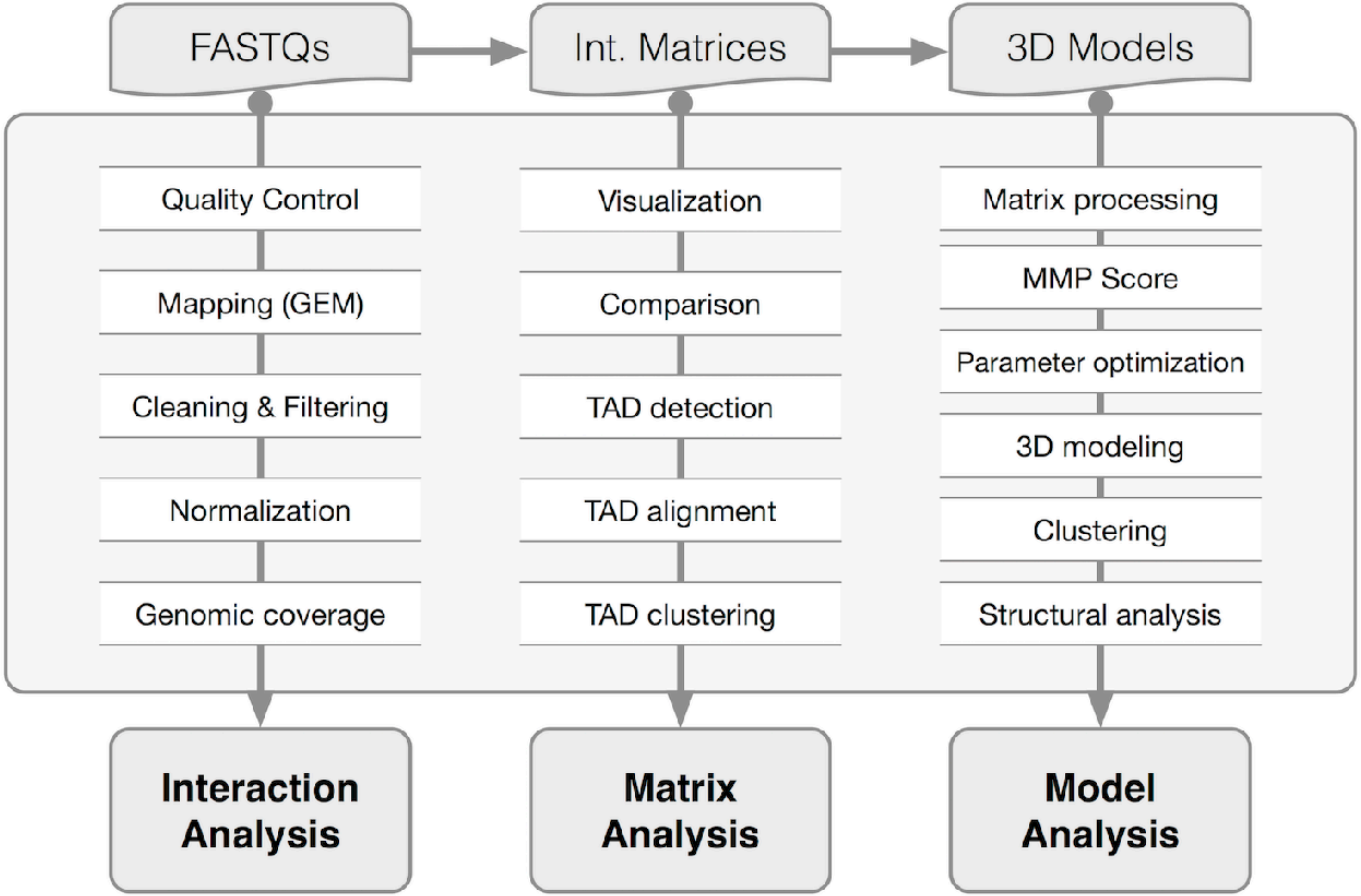






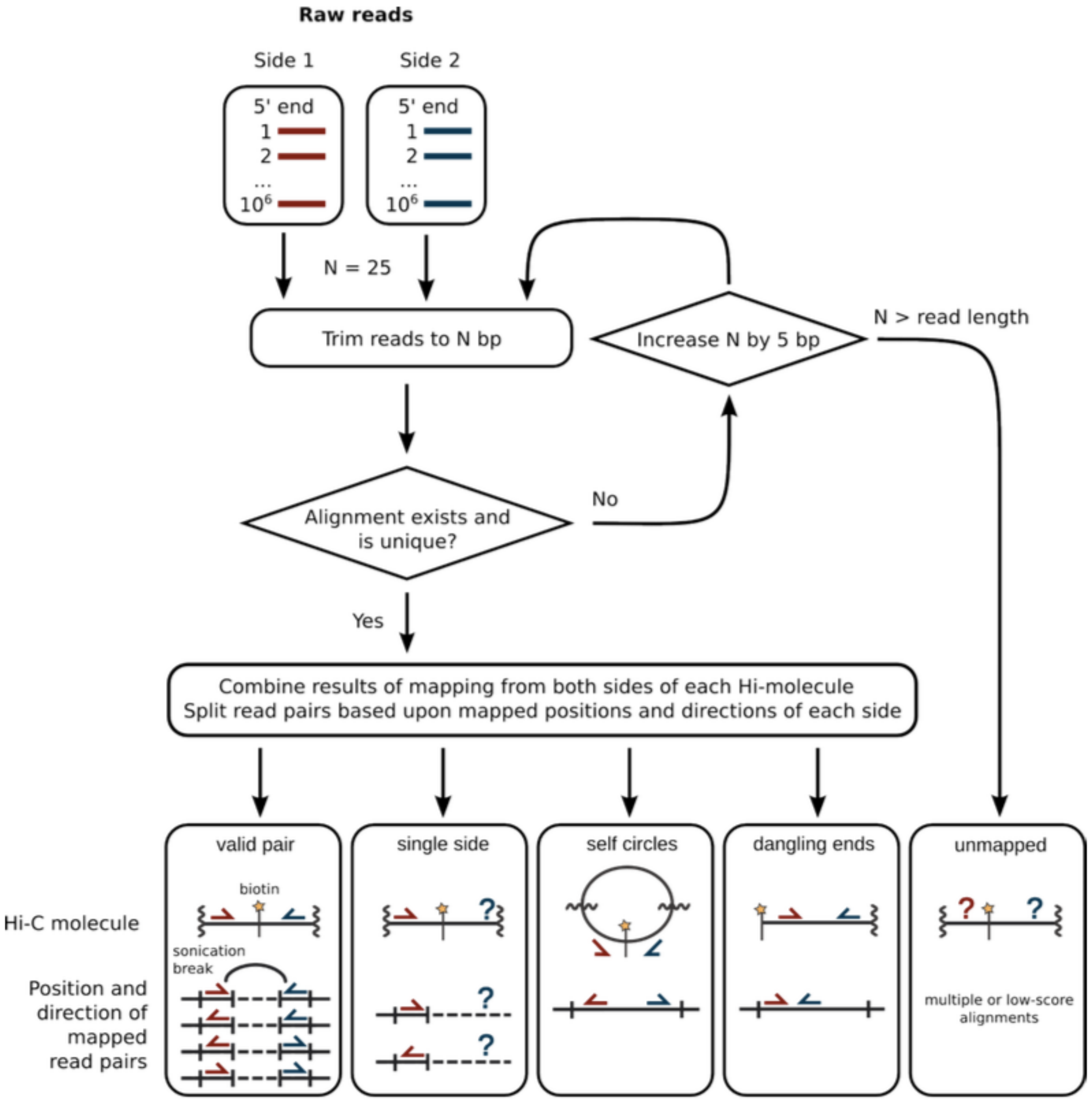
Analyzing 3C-based data (mostly Hi-C)

Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Nat Rev Genet 14, 390–403 (2013)



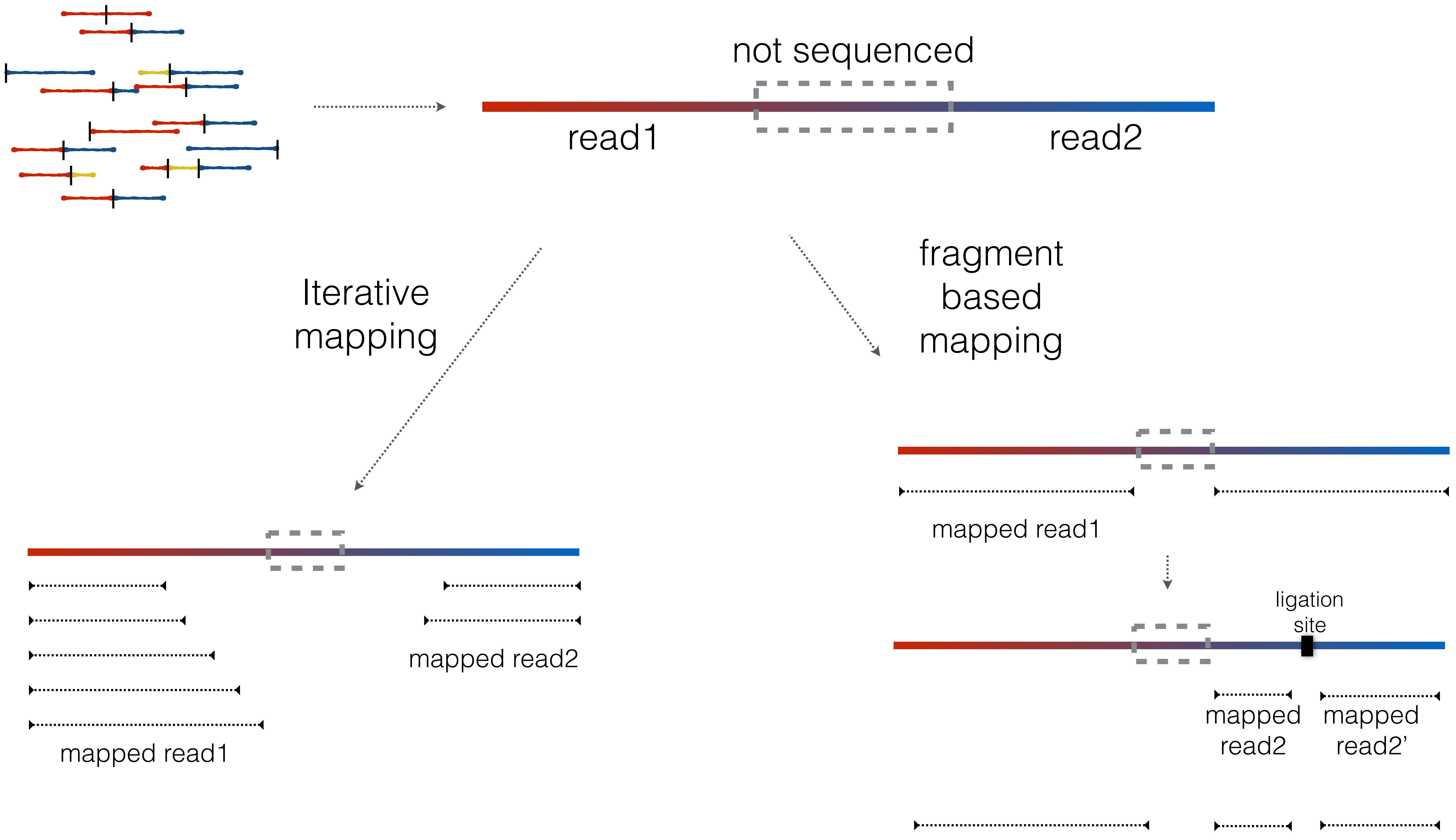
Mapping & Filtering

Imakaev, M. V et al. (2012). Nature Methods, 9(10), 999–1003.



Mapping @TADbit

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



not sequenced

read1

read2

Iterative mapping

fragment based mapping

mapped read1

mapped read2

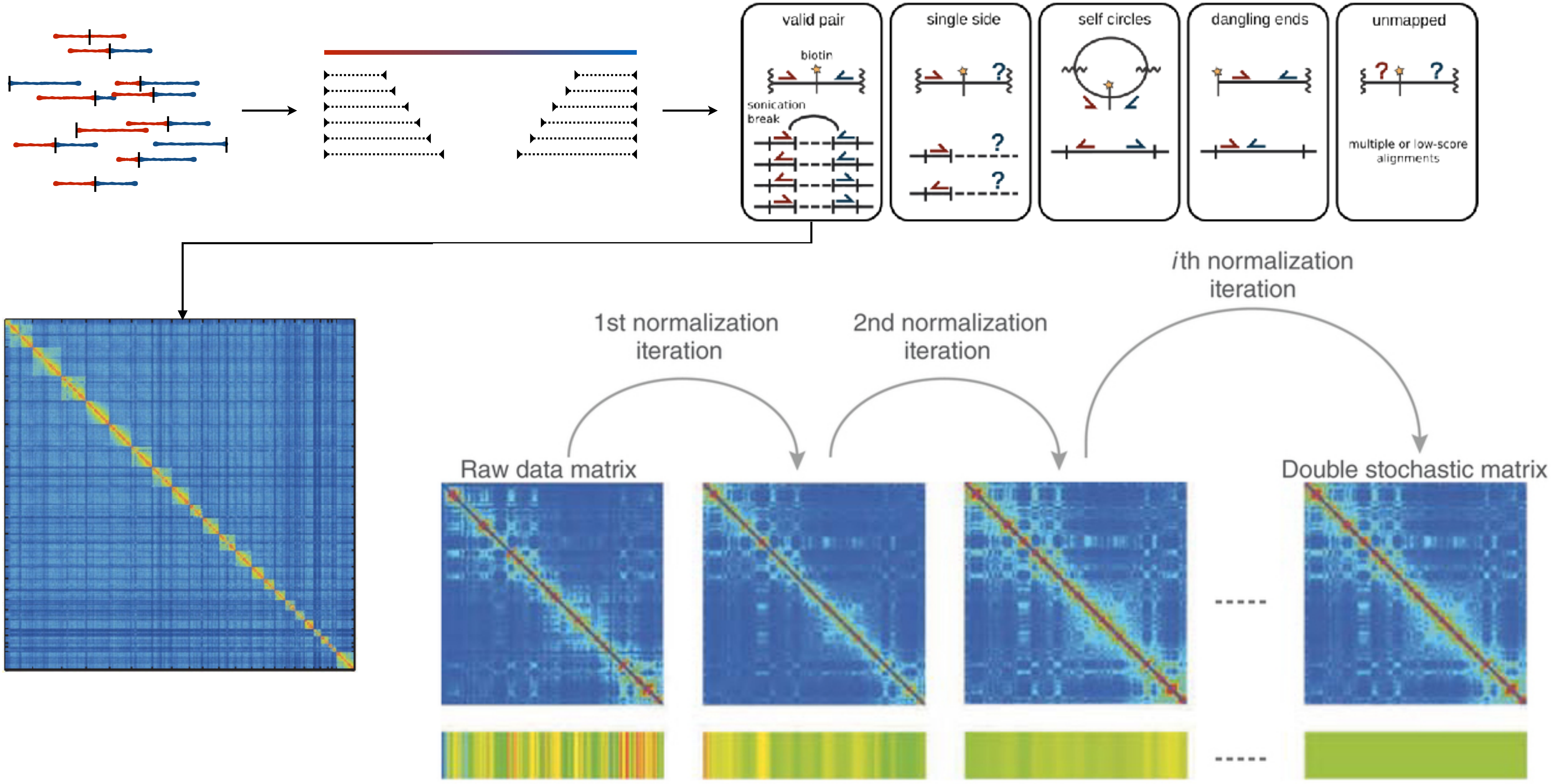
mapped read1

ligation site

mapped read2

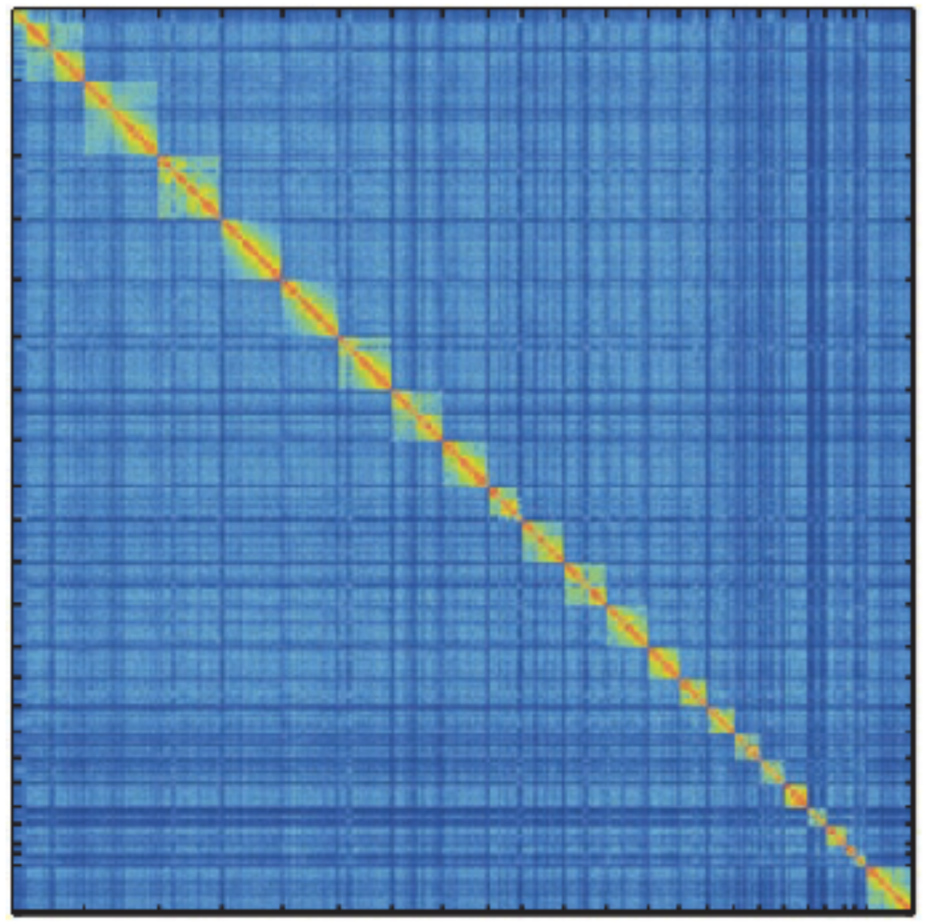
mapped read2’

Interaction matrices



Zooming in on genome organization. Zhou, X. J., & Alber, F. Nature Methods (2012)

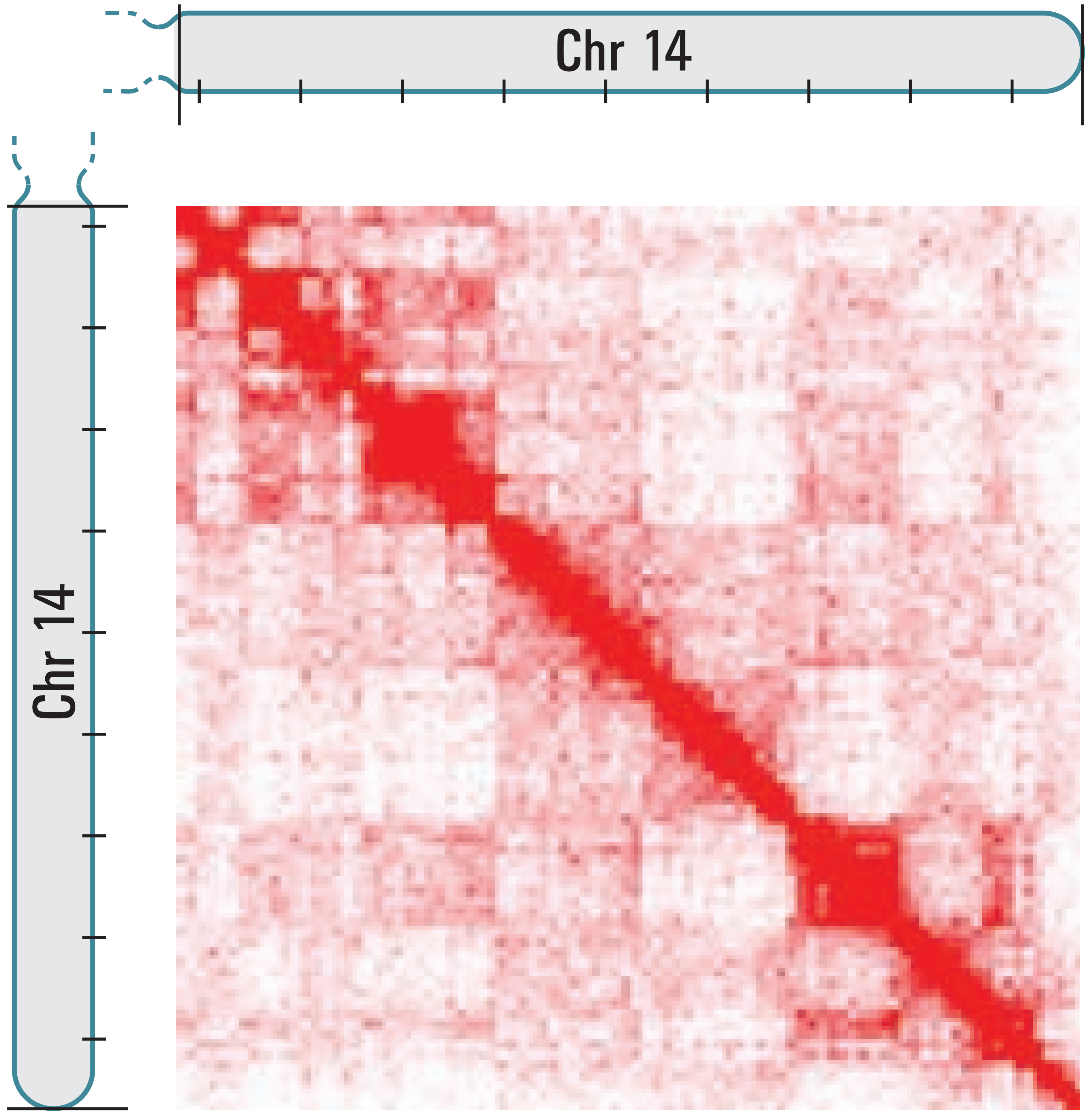
How much you normally map?

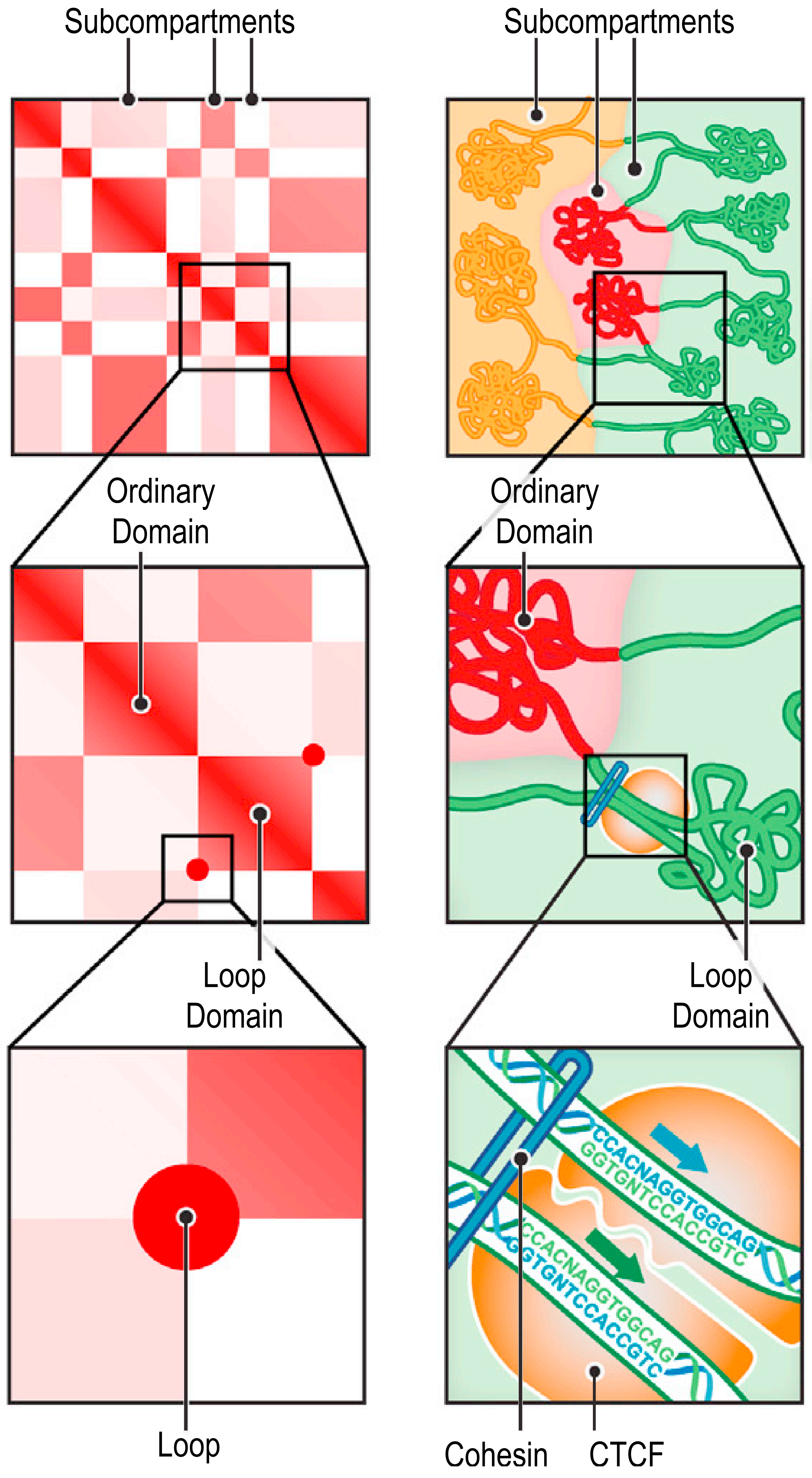


- 80-90% each end => 60-80% intersection
- ~1% multiple contacts
- Many of intersecting pairs will be lost in filtering…
- Final 40-60% of valid pairs
- One measure of quality is the CIS/TRANS ration (70-80% good)

Hierarchical genome organisation

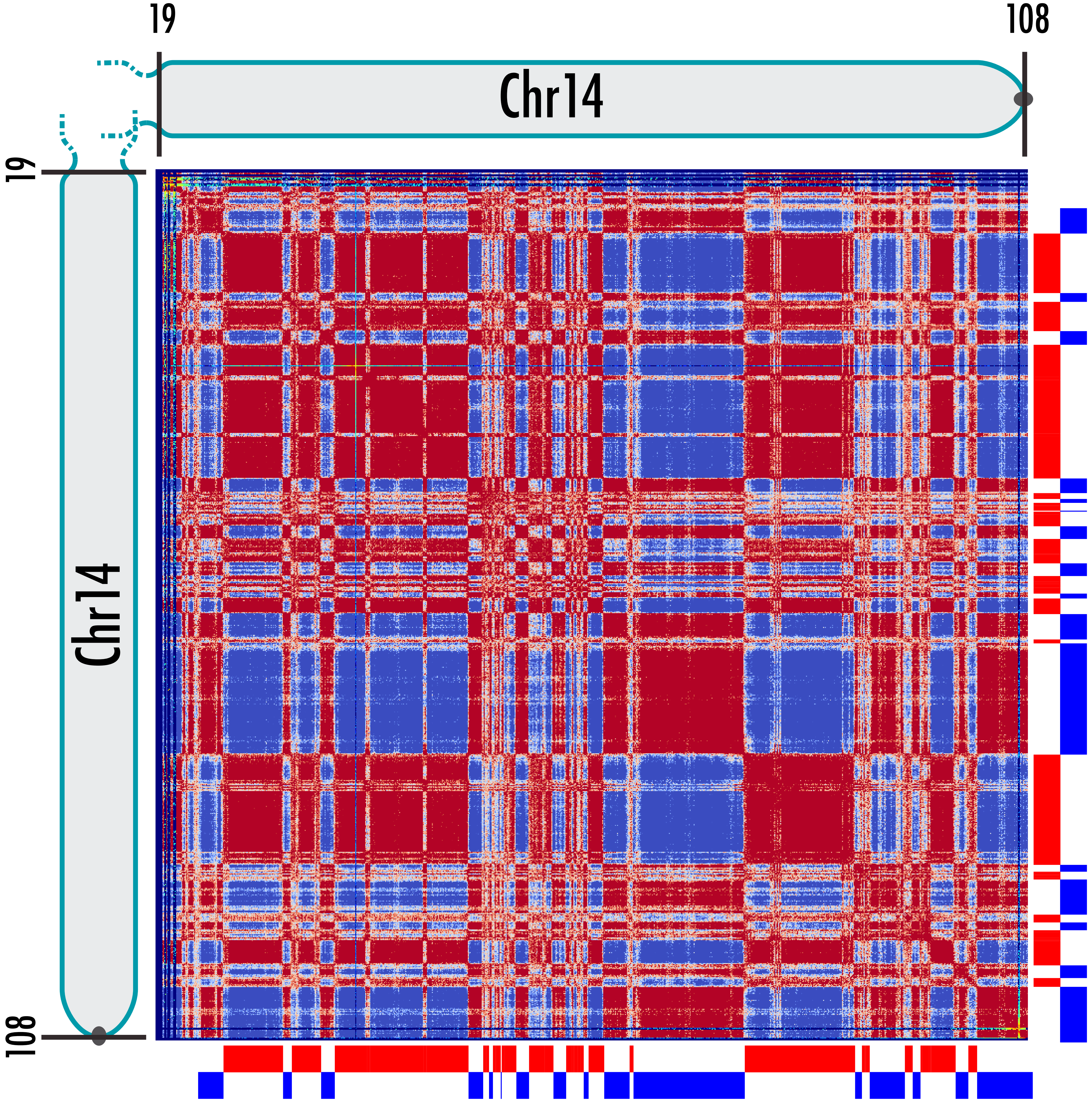
Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.
Rao, S. S. P., et al. (2014). Cell, 1–29.

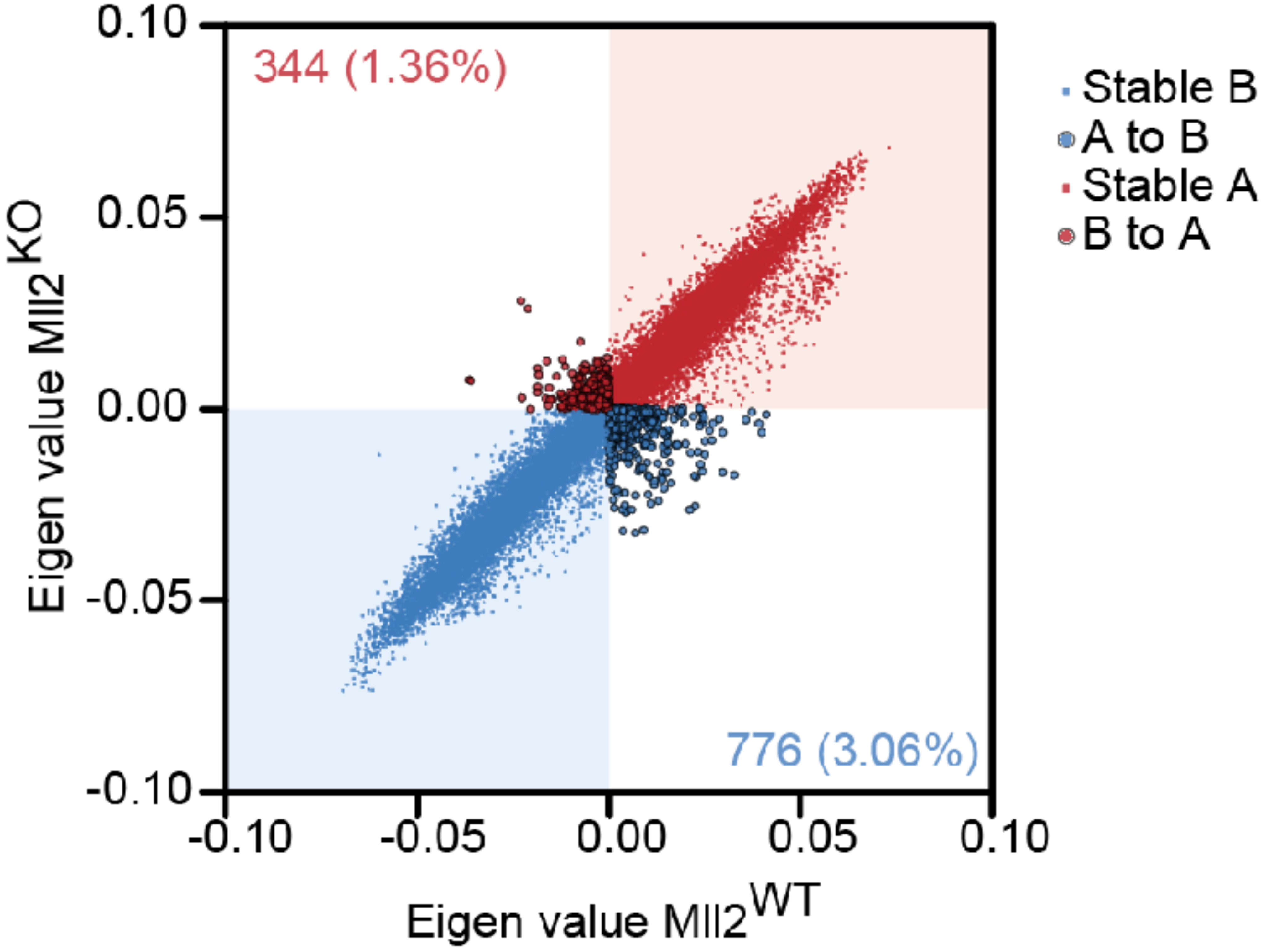




A/B Compartment

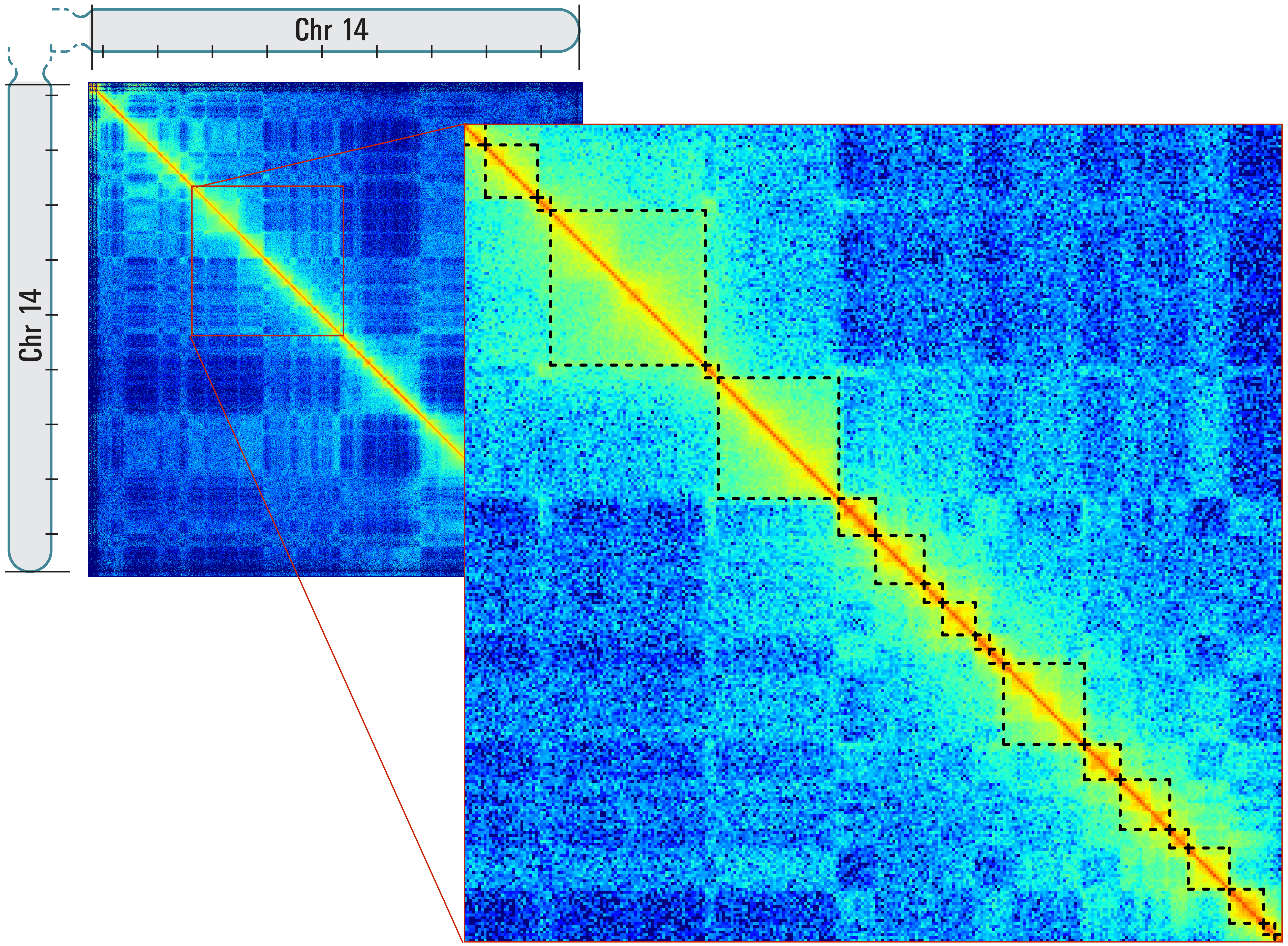
Chromosome 14

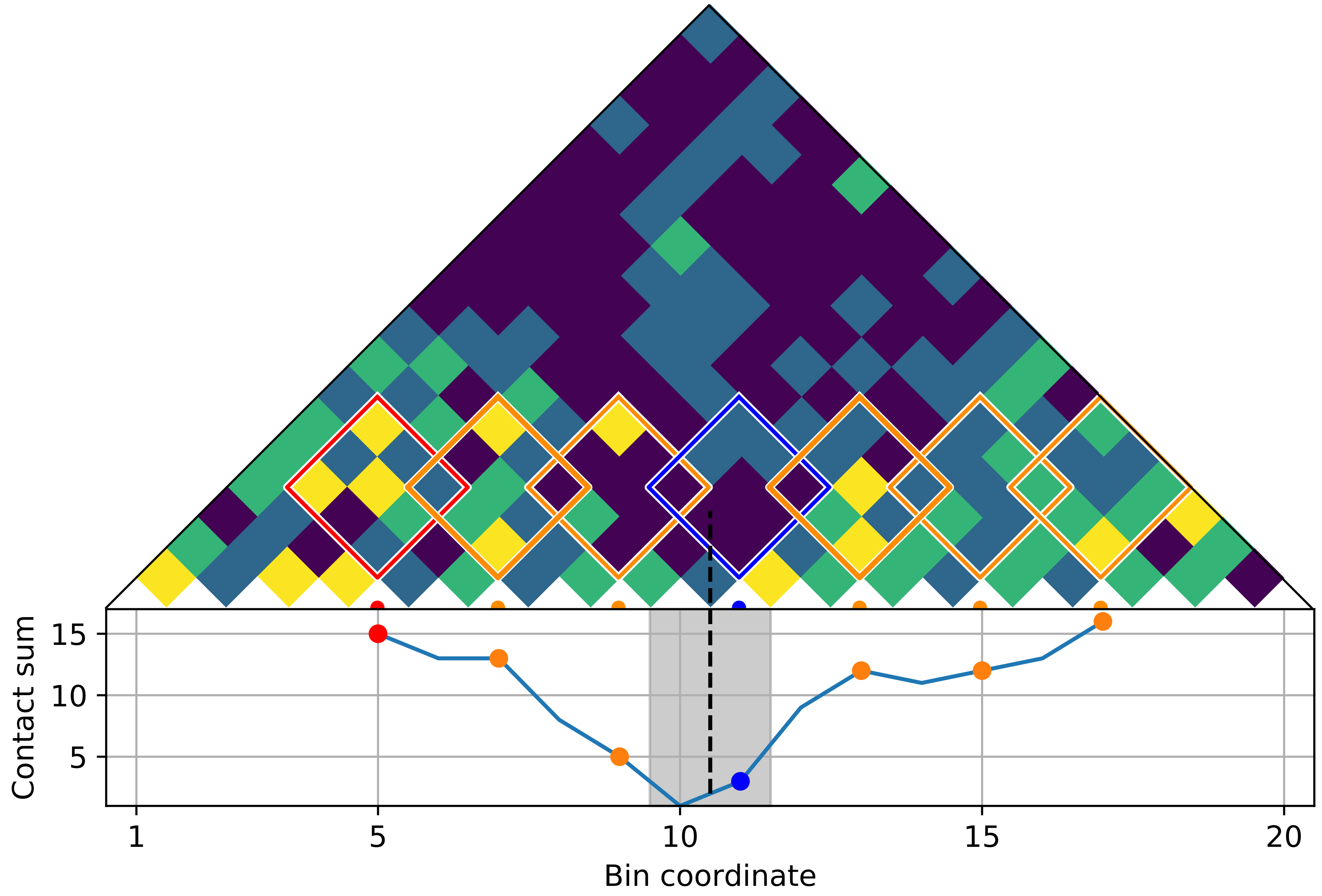




TADs

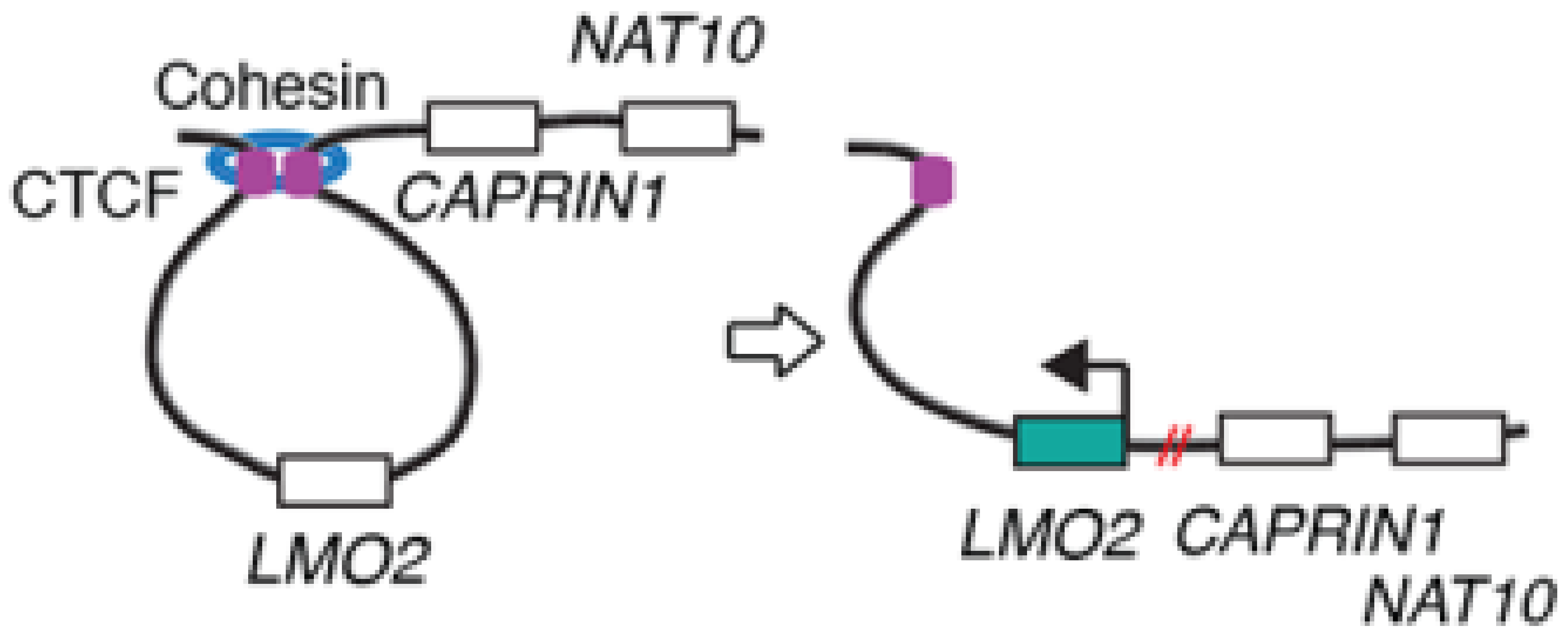
Chromosome 14

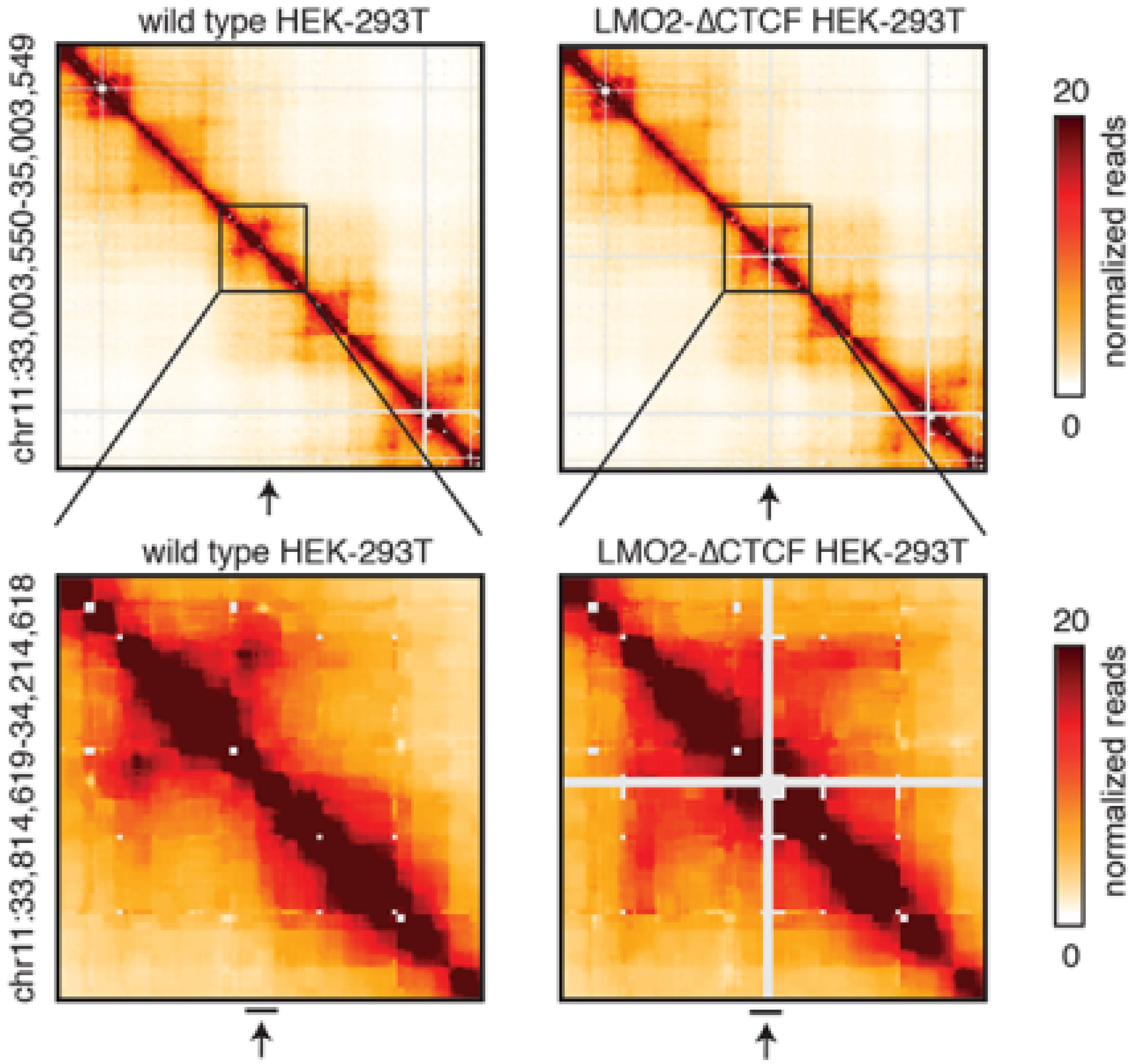




TADs are functional units

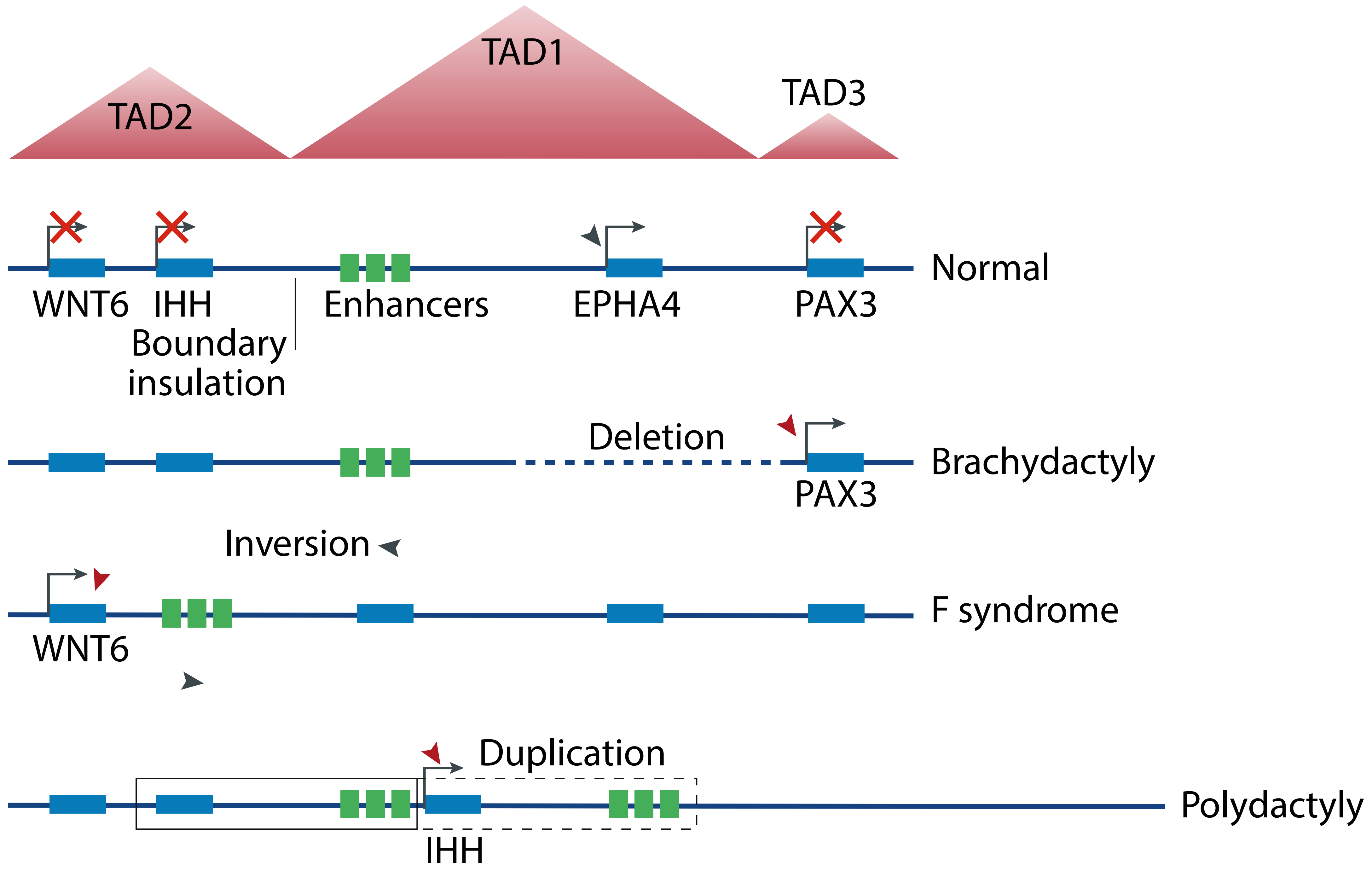
Hnisz, D., et al. (2016). Science



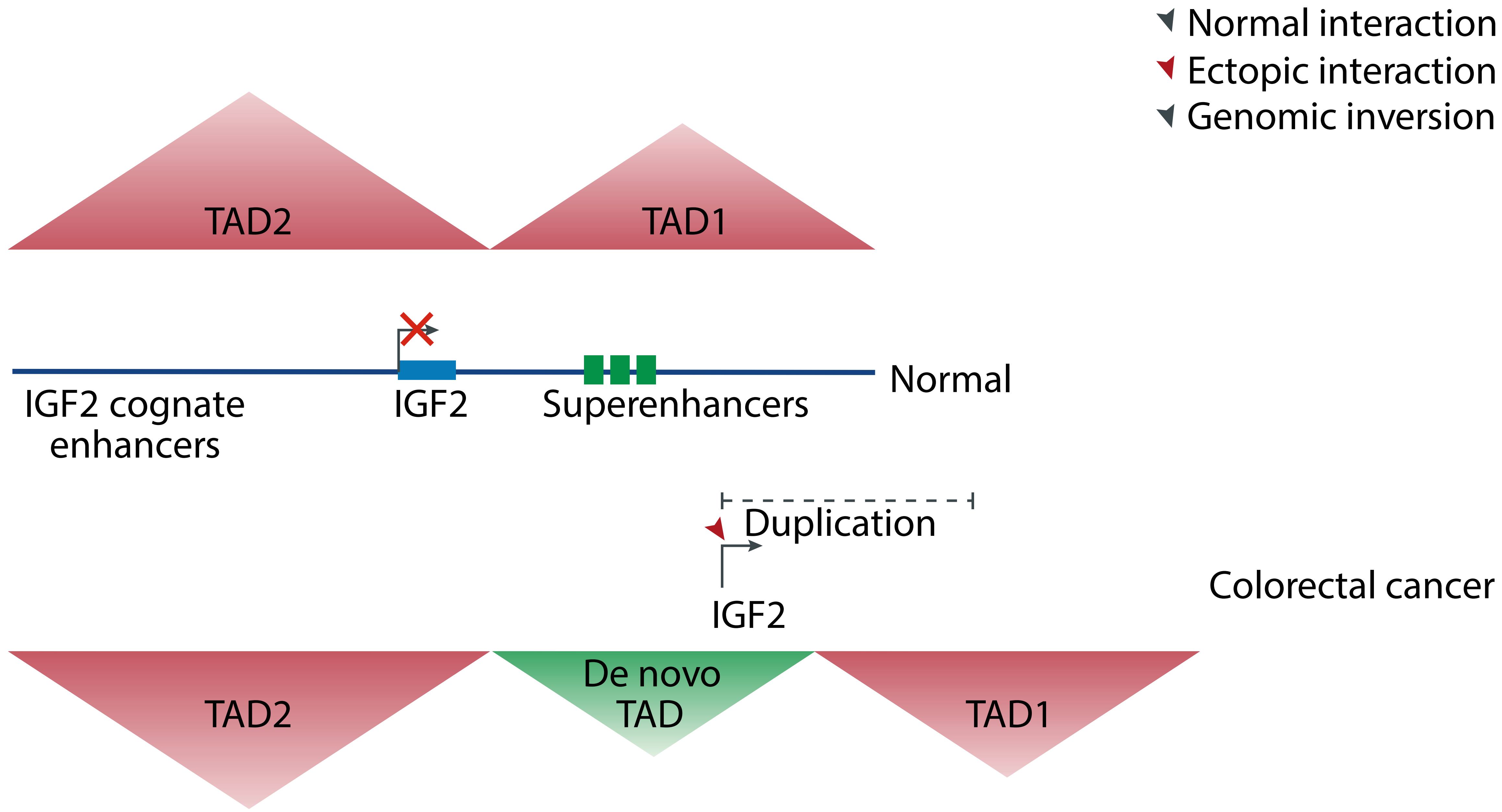


TADs are functional units

Figure adapted from Hui Zheng and Wei Xie. Nature Reviews Molecular Cell Biology (2019)



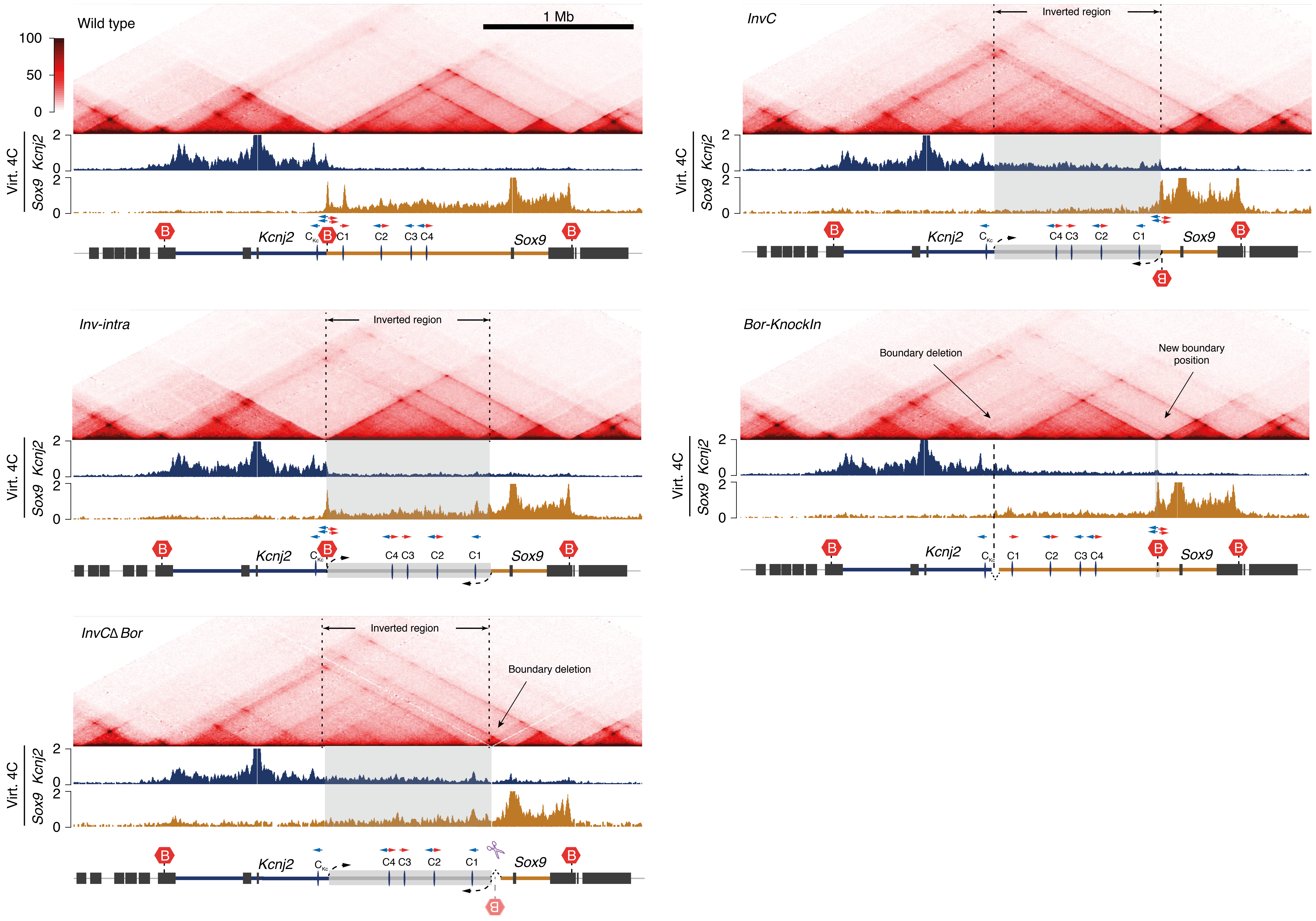
Lupianez, D. G. et al. Cell 161, 1012–1025 (2015)



Flavahan, W. A. et al. Nature 529, 110–114 (2016).

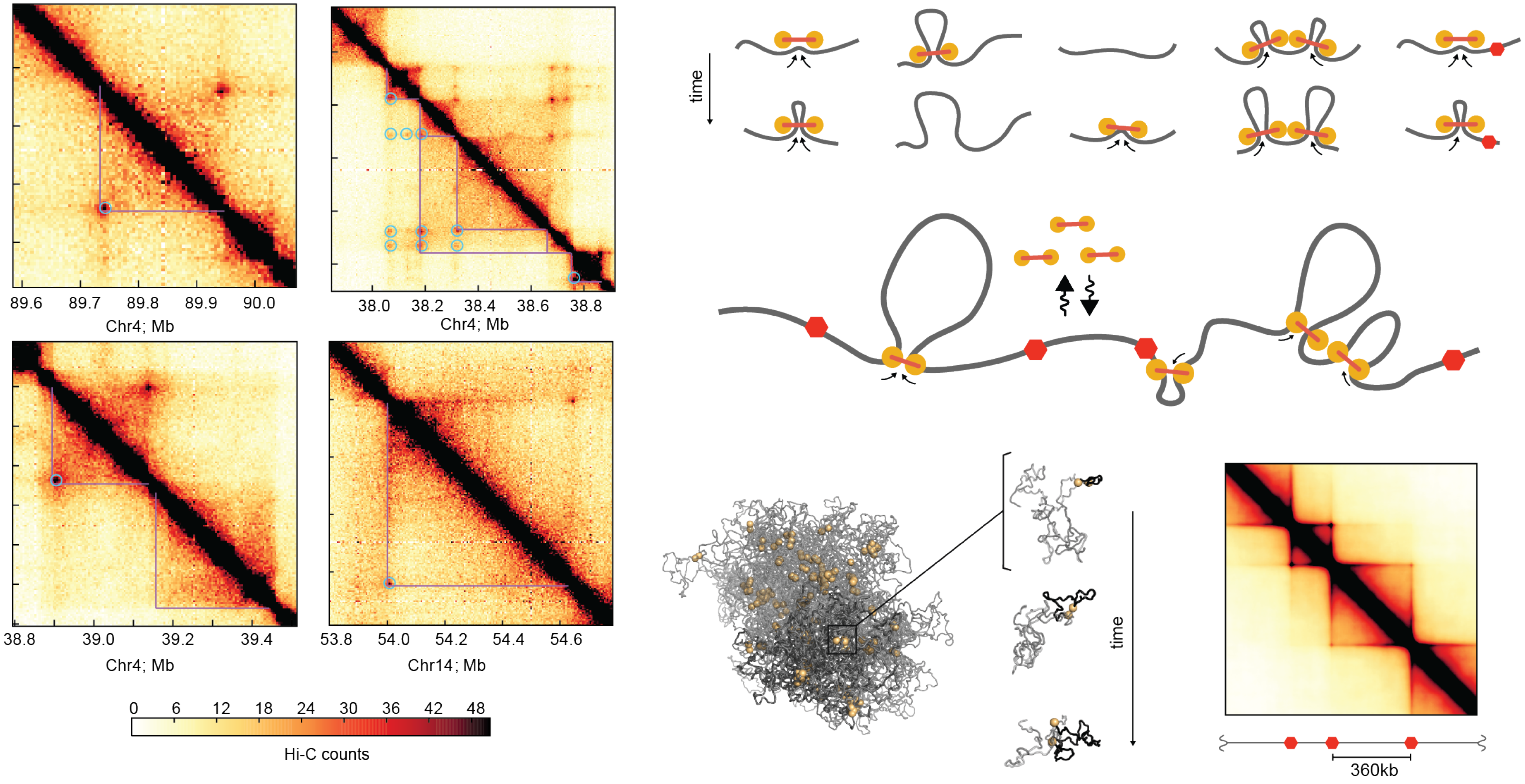
TADs are functional units

Despang, et al. (2019). Nature Genetics 51,1263–1271 (2019)



Loop-extrusion as a TAD forming mechanism

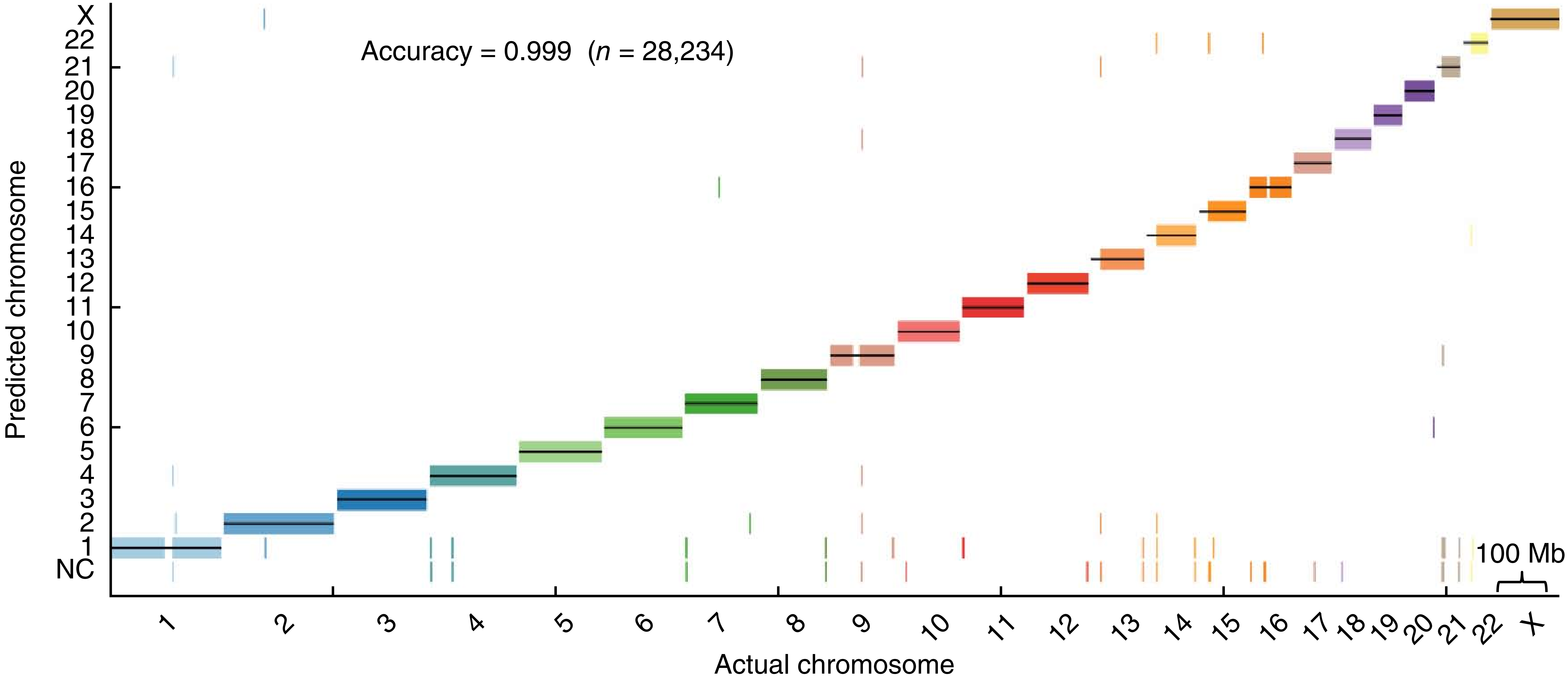
Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2018).
Cold Spring Harb Symp Quant Biol 2017. 82: 45-55




SIDE EFFECTS

Hi-C for de-novo assembly

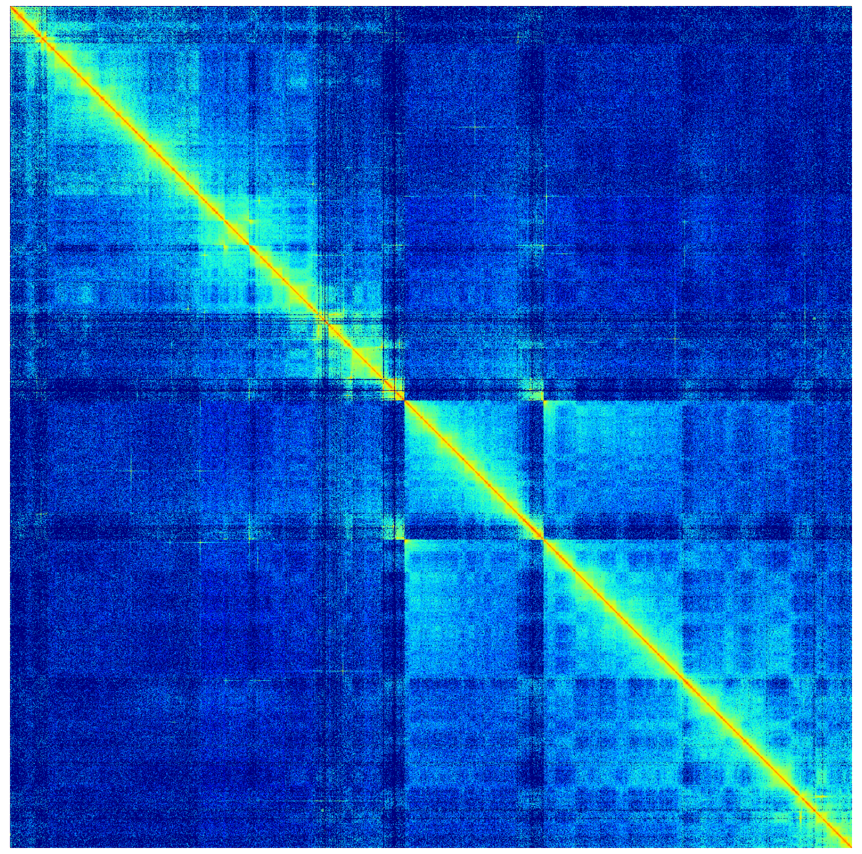
Kaplan, N., & Dekker, J. (2013). Nature Biotechnology, 31(12), 1143–1147.

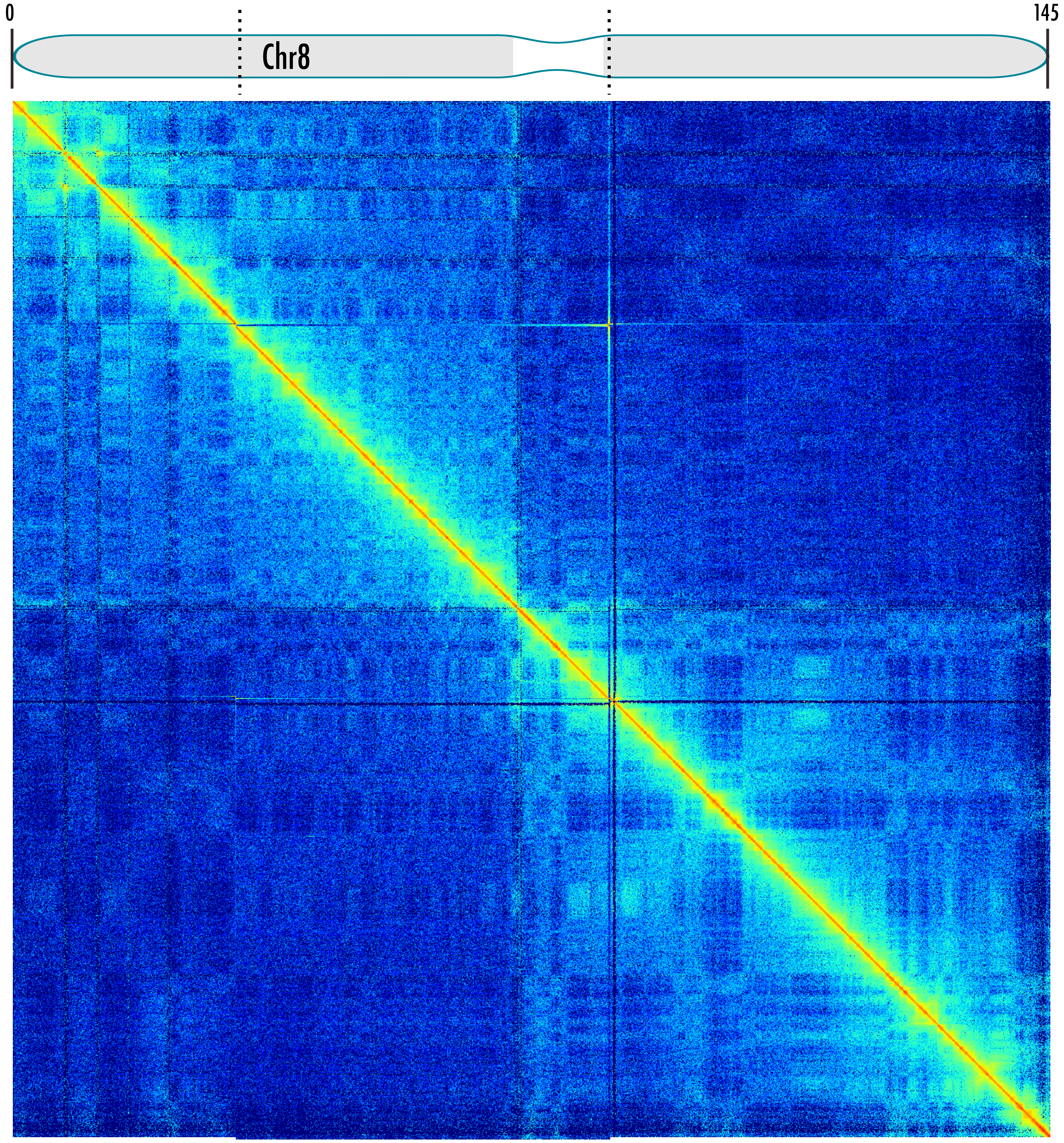


Assembly error detection

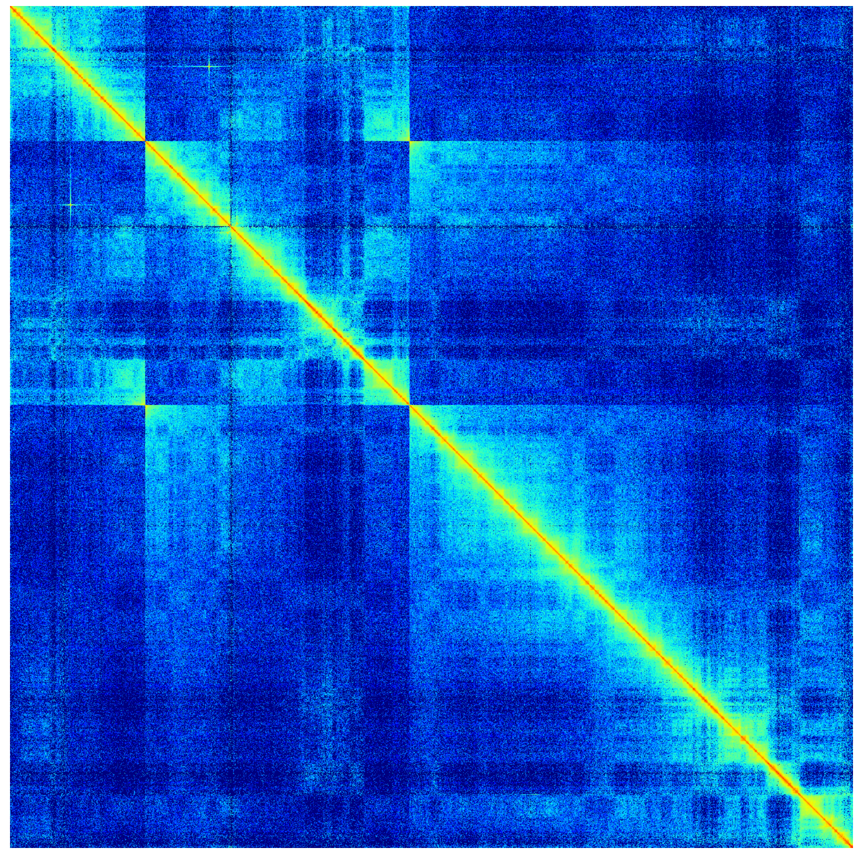
Chromosome 8 Gorilla

Chr 7





Chr 12

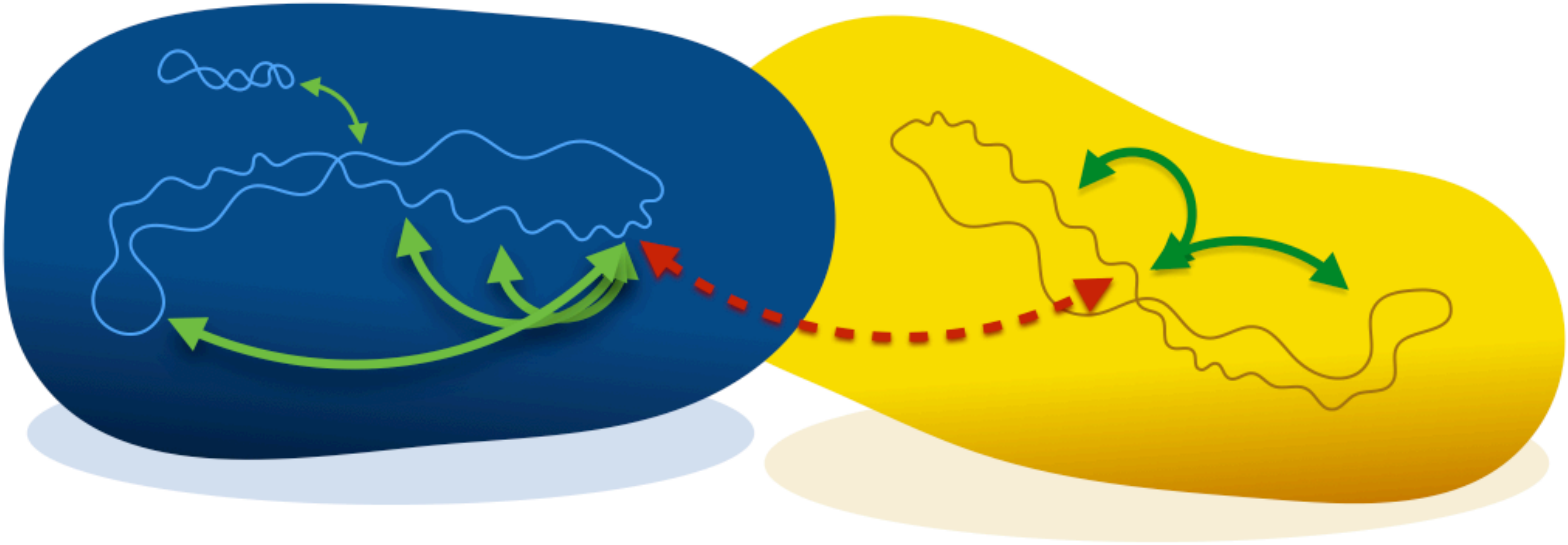


GGO8 has an inversion of the region corresponding to HSA8:30.0-86.9Mb
Aylwyn Scally (Department of Genetics, University of Cambridge)

Hi-C for meta genomics

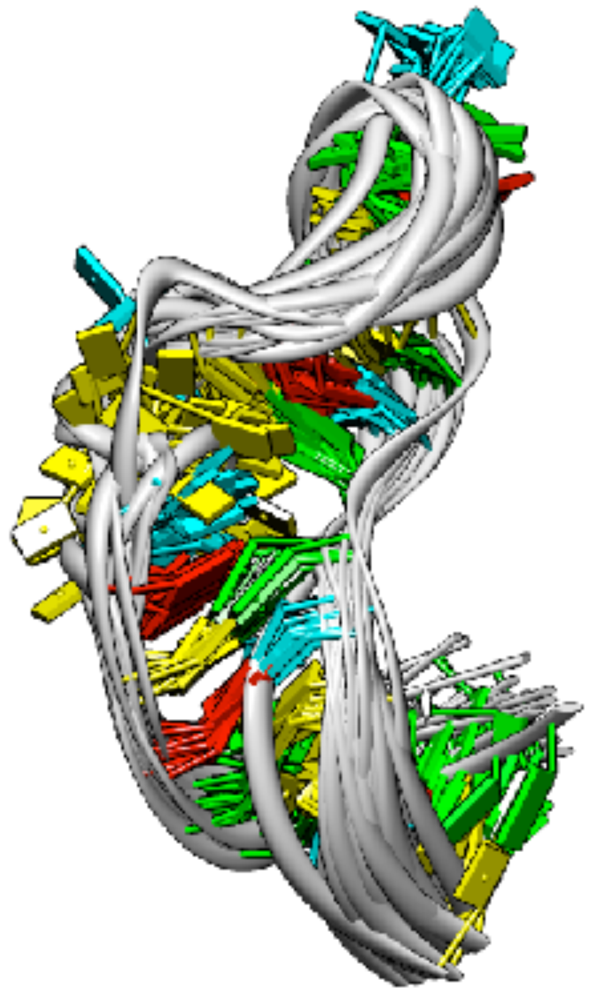
Beitel, C. W., Froenicke, L., Lang, J. M., Korf, I. F., Michelmore, R. W., Eisen, J. A., & Darling, A. E. (2014). Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. doi:10.7287/peerj.preprints.260v1

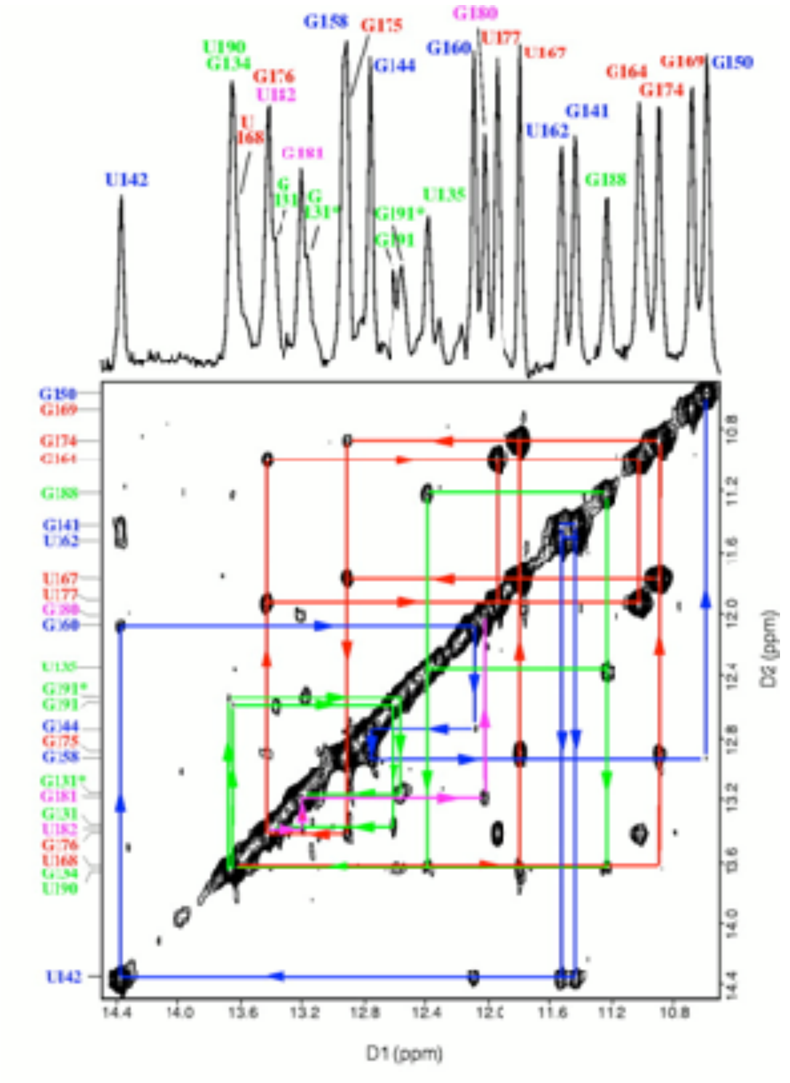
Romain Koszul



Restraint-based Modeling

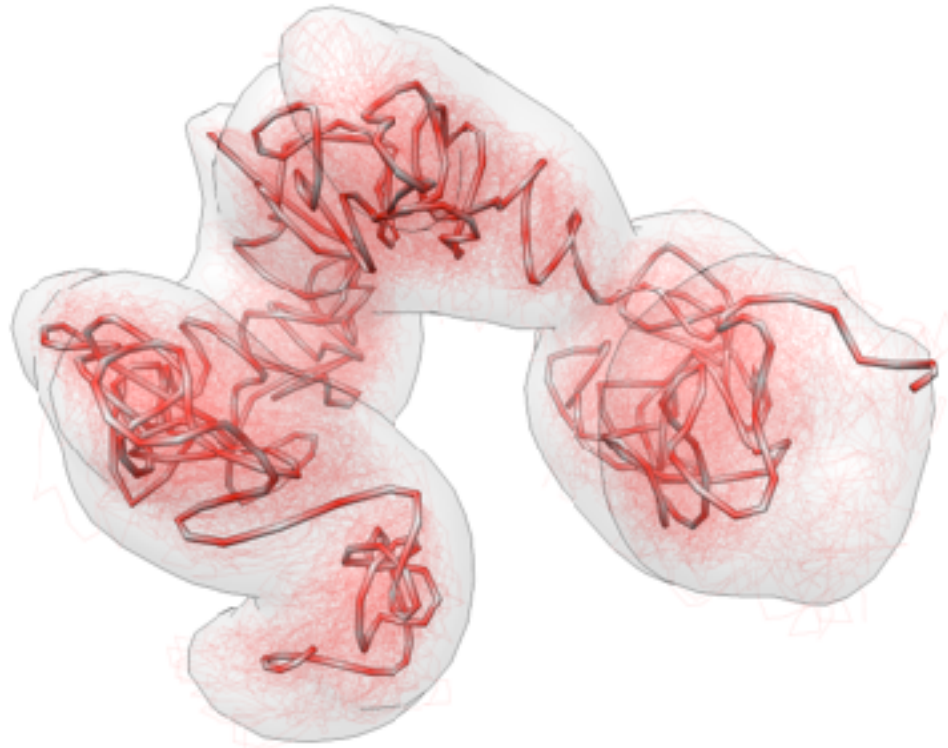
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).

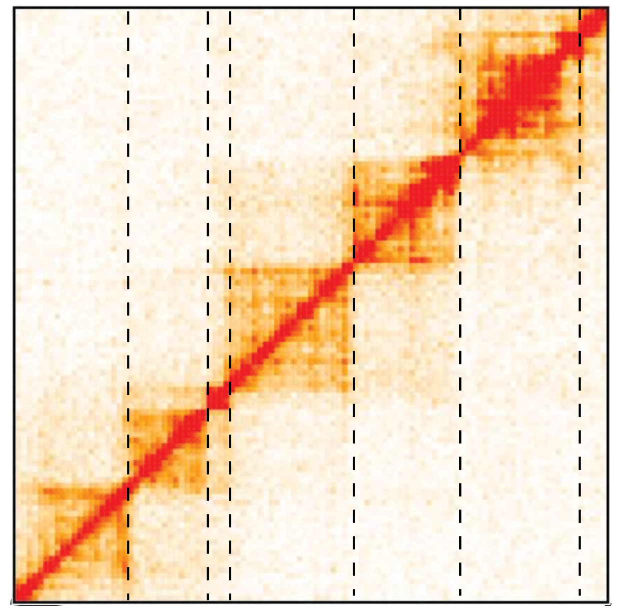




Biomolecular structure determination

2D-NOESY data





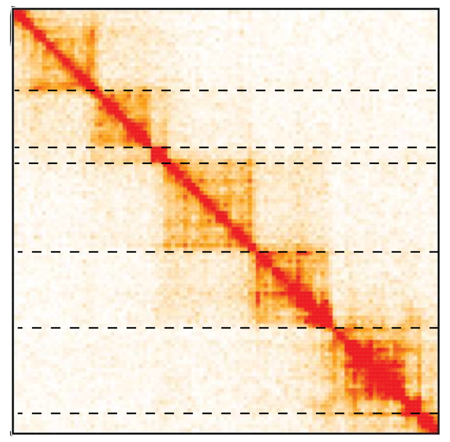
Chromosome structure determination

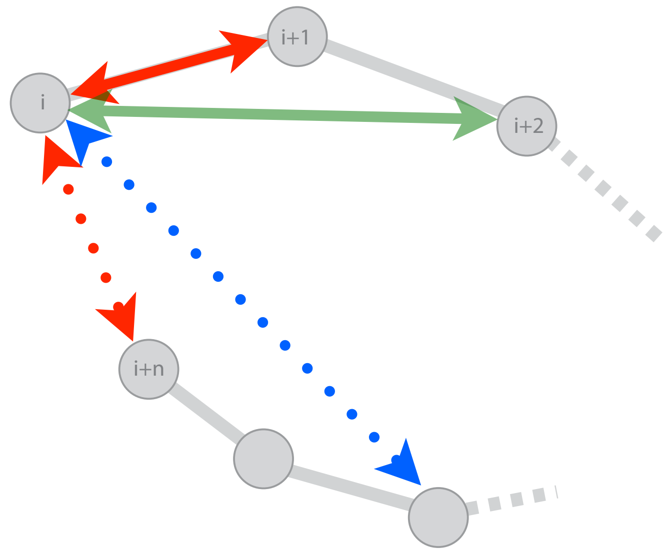
3C-based data

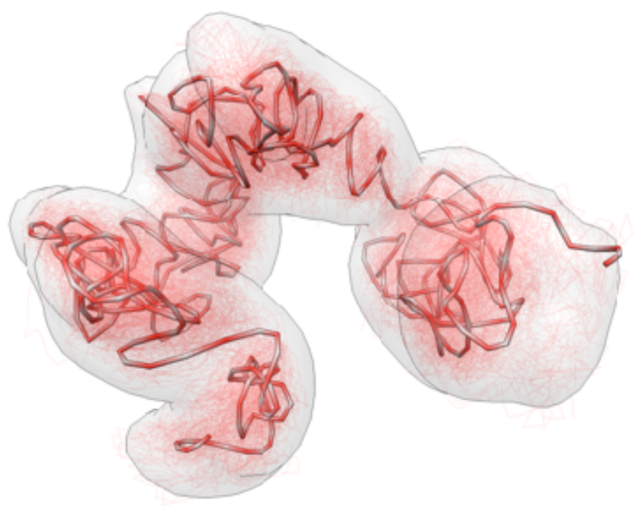
TADbit

http://3DGenomes.org

Serra, F., Baù, D. et al. PLOS CB (2017)





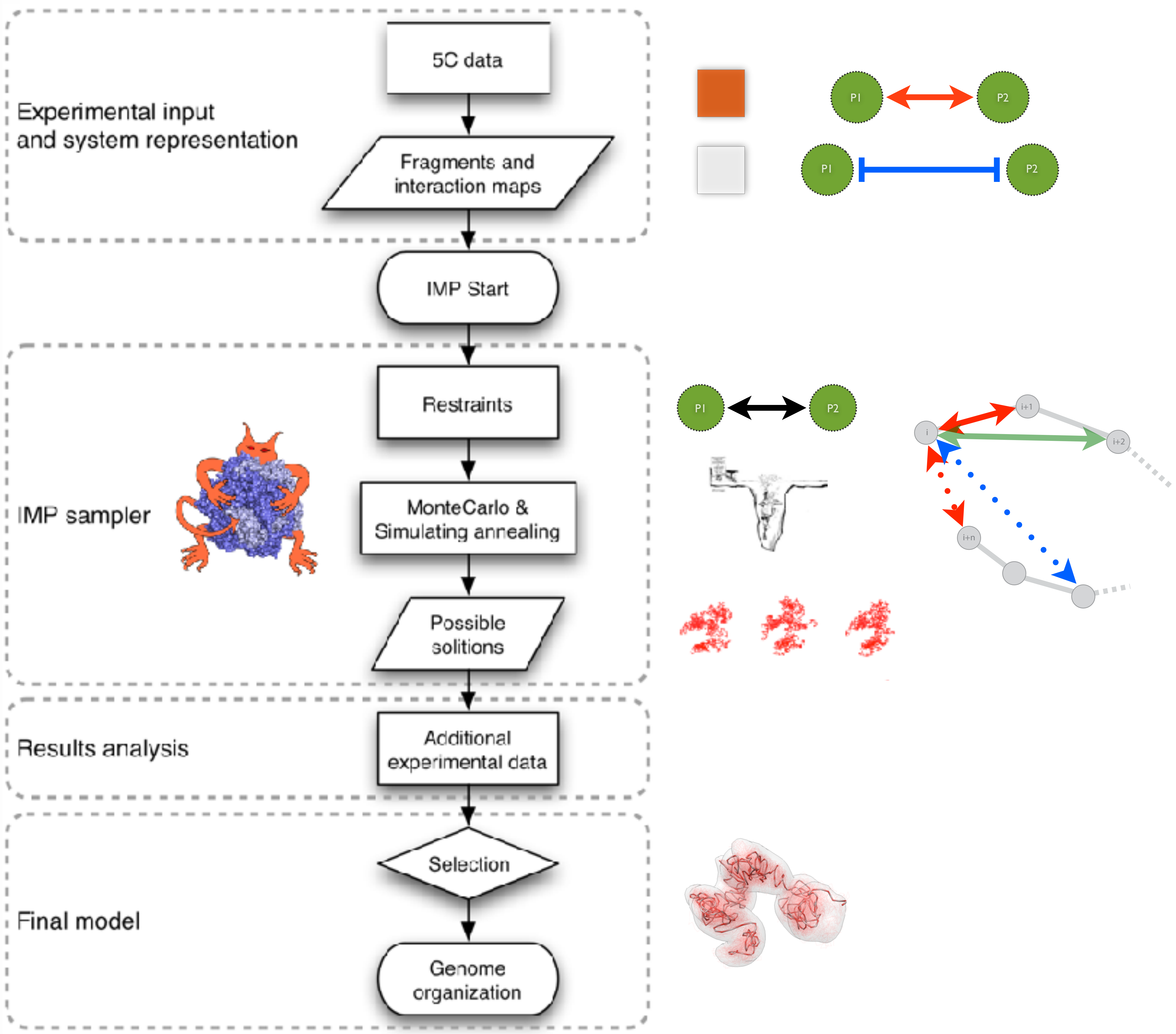


FastQ files to Maps

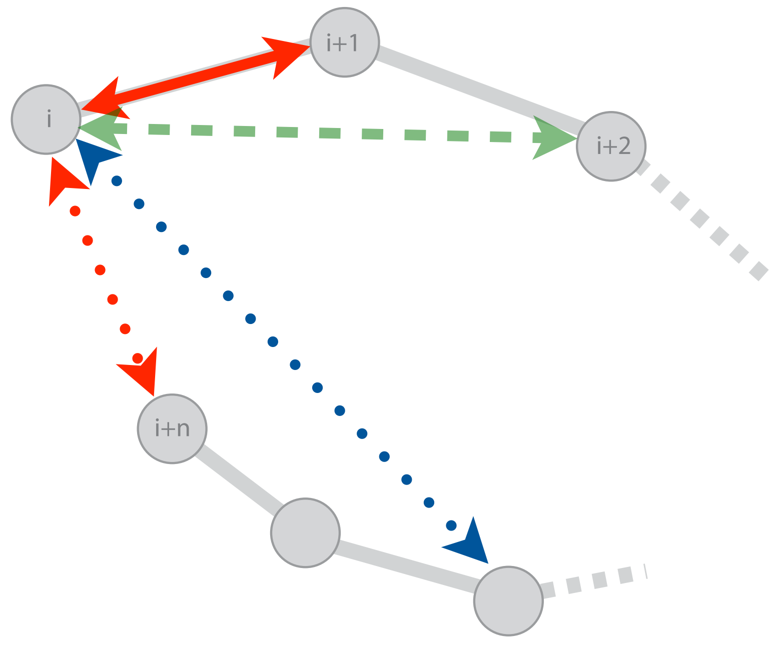
Map analysis

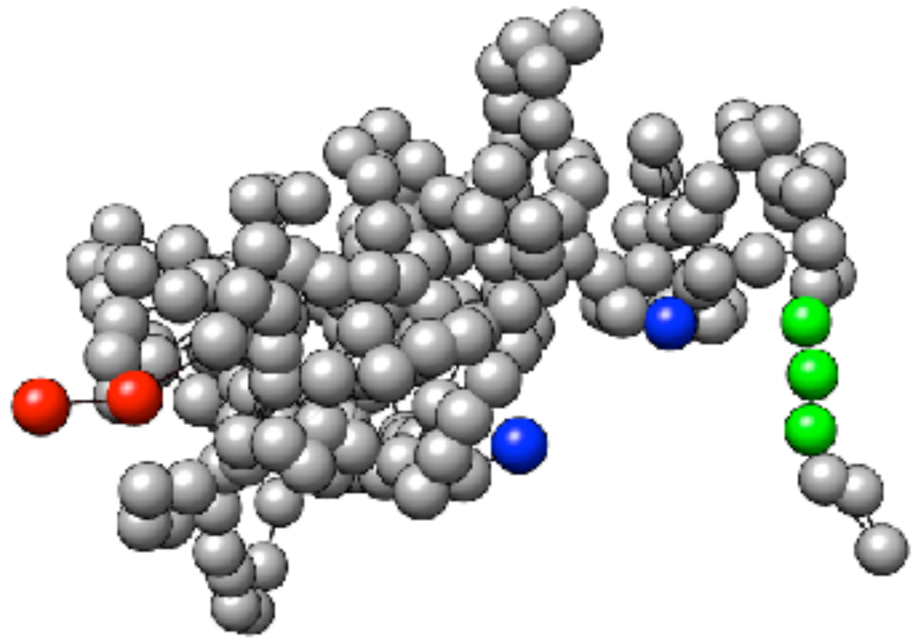
Model building

Model analysis

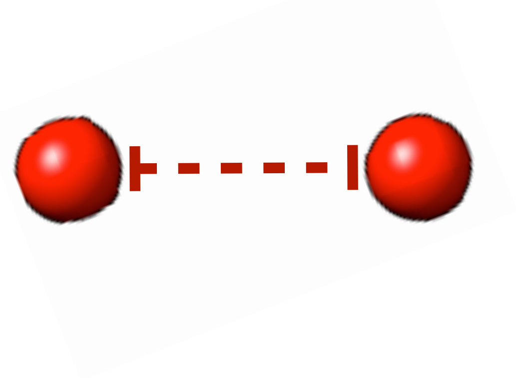


Model representation and scoring



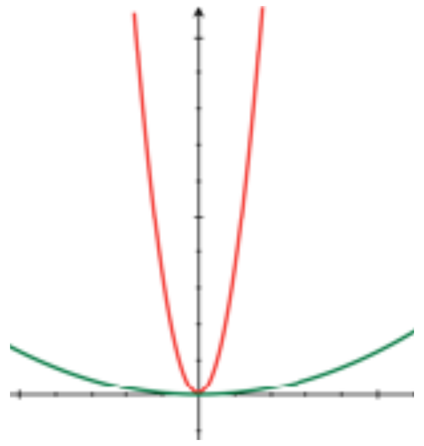


d = d0

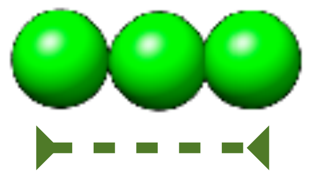


Harmonic

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

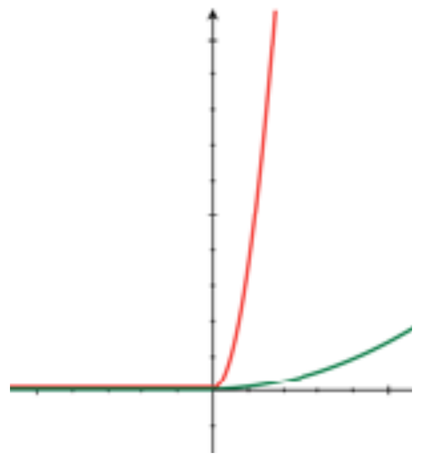


d < d0

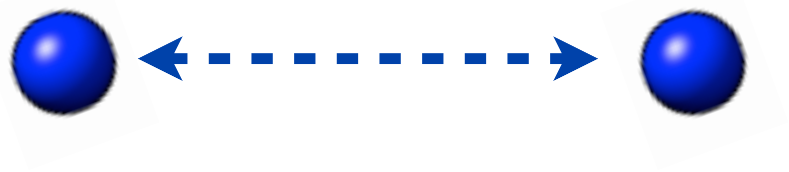


Harmonic Upper Bound

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

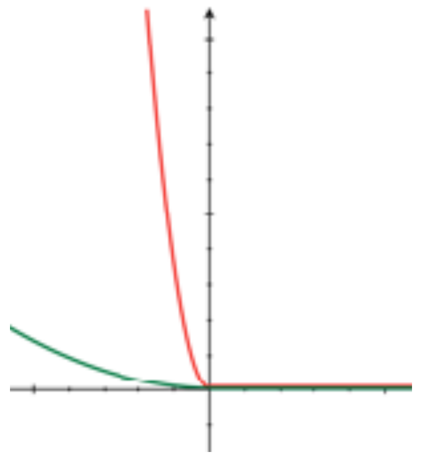


d > d0

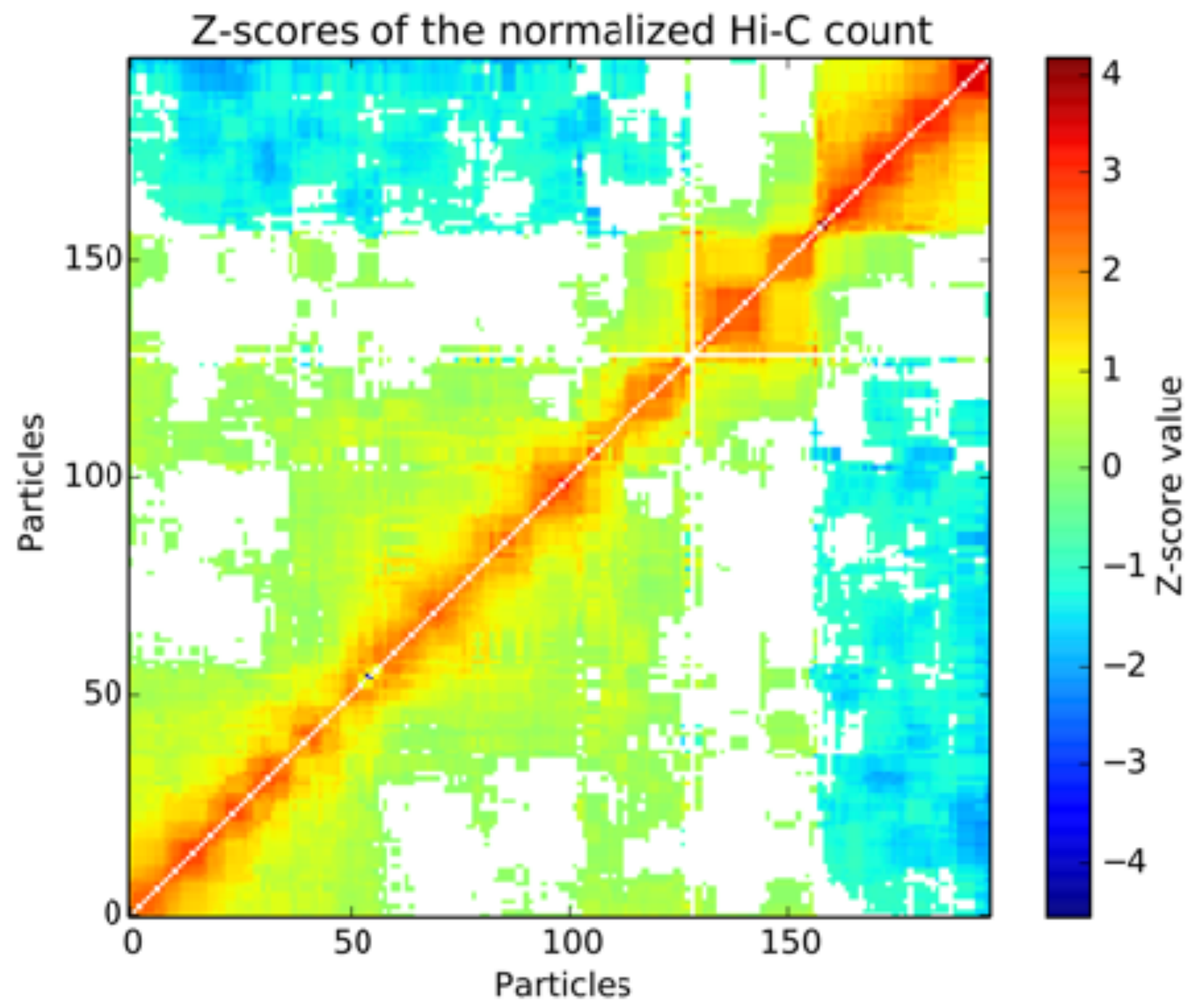


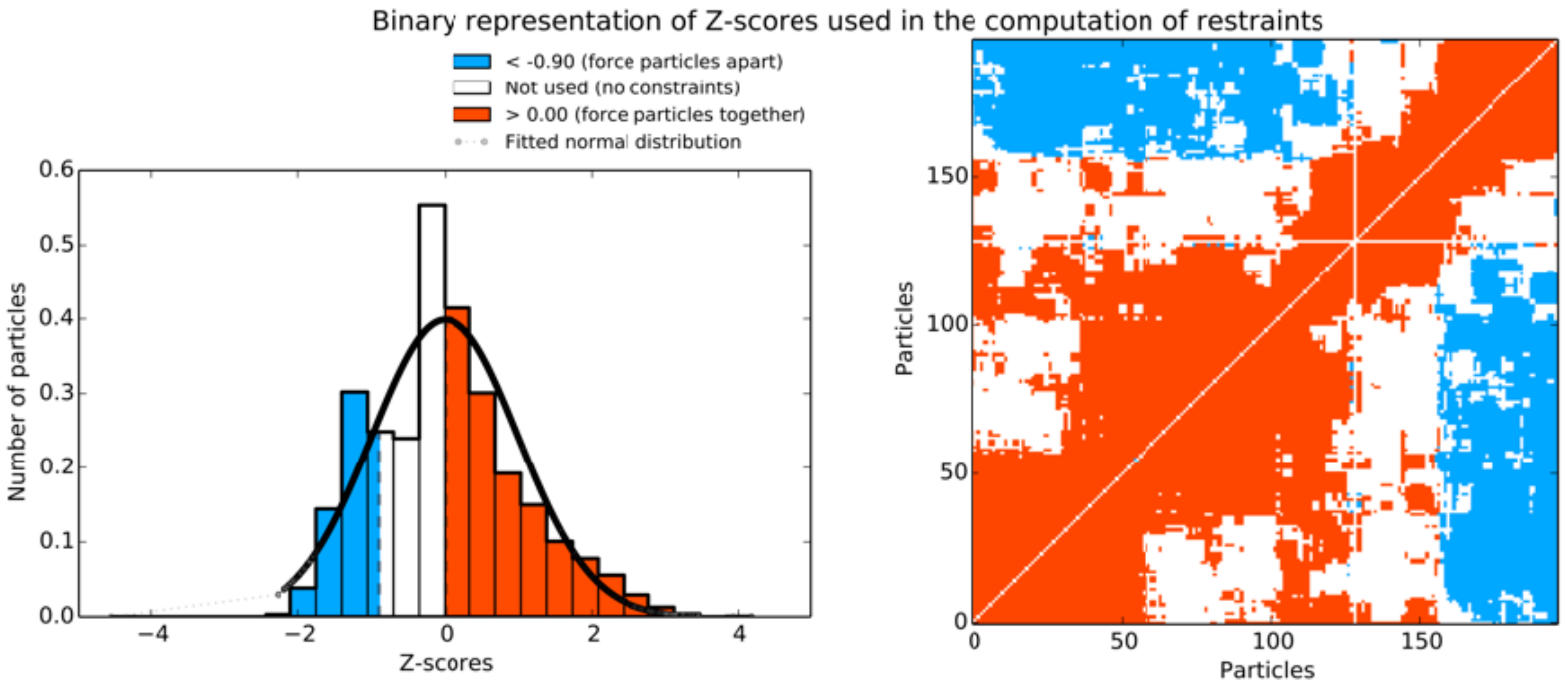
Harmonic Lower Bound

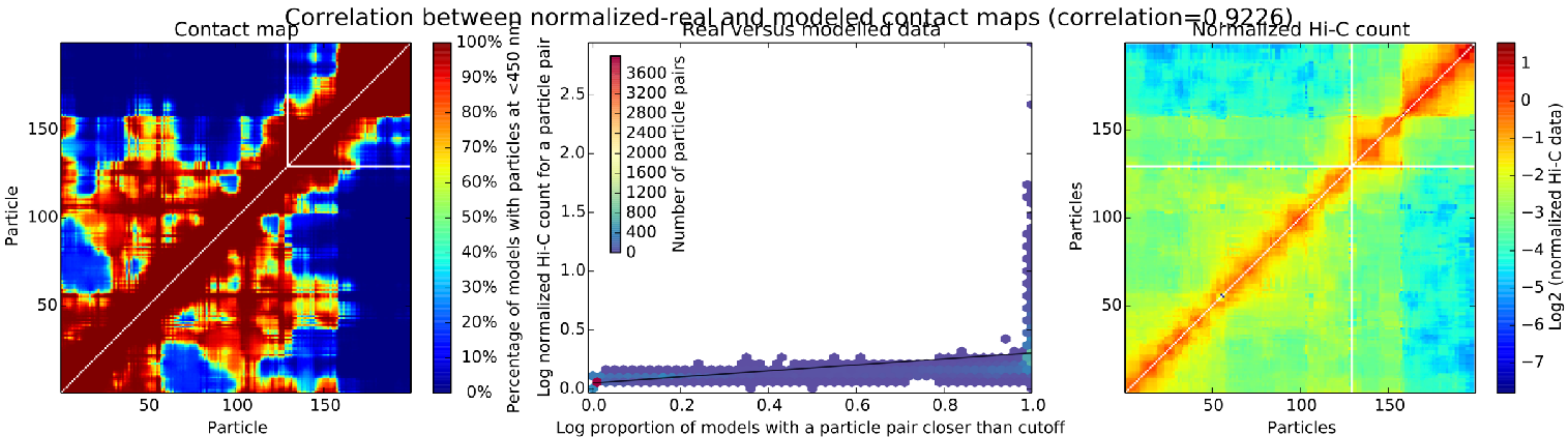
$$\begin{cases} if\ \ d_{i,j} \leq d^0_{i,j}; & lbH_{i,j} = k\left(d_{i,j} - d^0_{i,j}\right)^2 \\ if\ \ d_{i,j} > d^0_{i,j}; & lbH_{i,j} = 0 \end{cases}$$



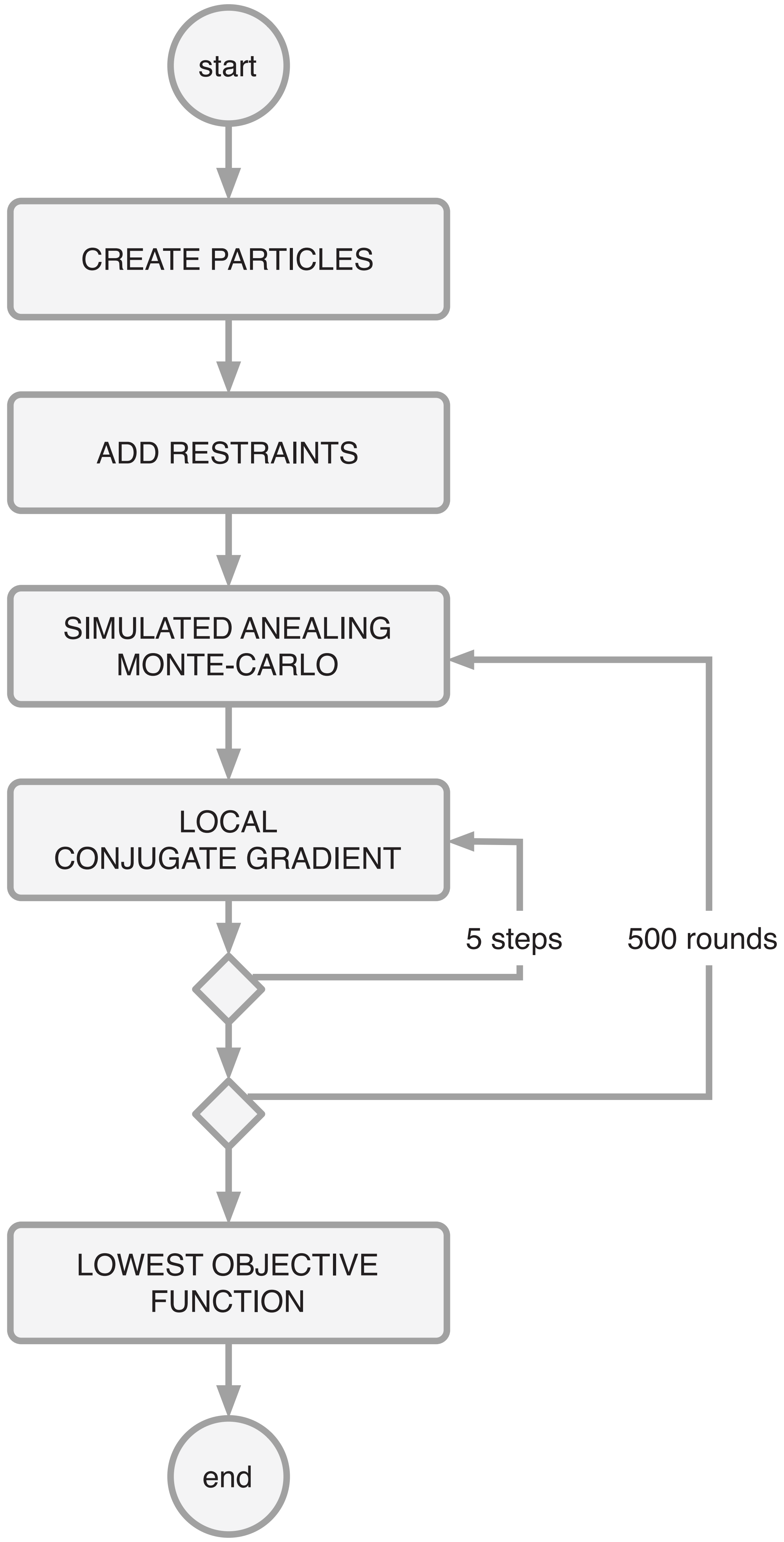
Parameter optimization


Z-scores of the normalized Hi-C count


Binary representation of Z-scores used in the computation of restraints


Correlation between normalized-real and modeled contact maps (correlation=0.9226)

Optimization of the scoring function



start

CREATE PARTICLES

ADD RESTRAINTS

SIMULATED ANEALING MONTE-CARLO

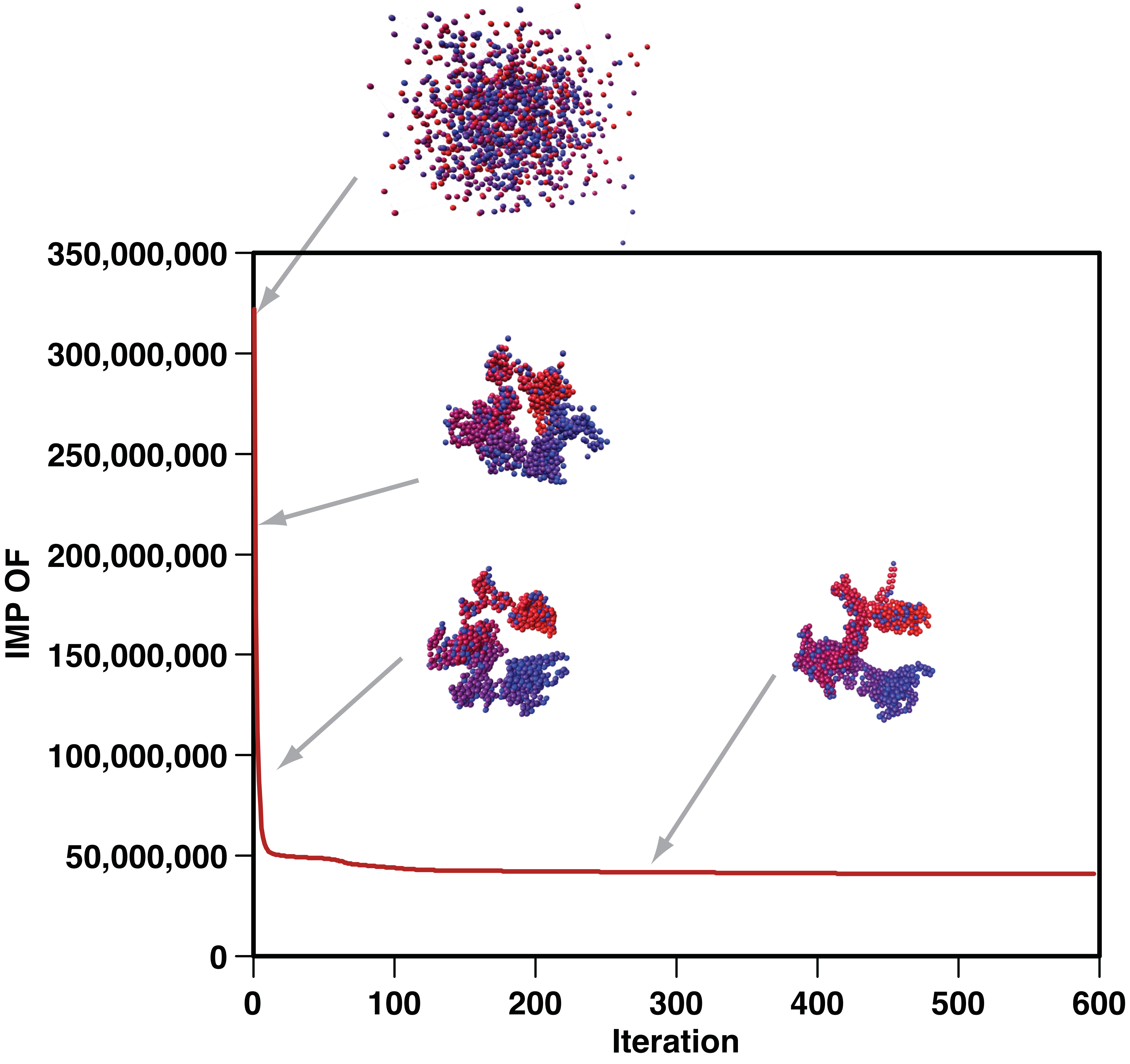
LOCAL CONJUGATE GRADIENT

5 steps

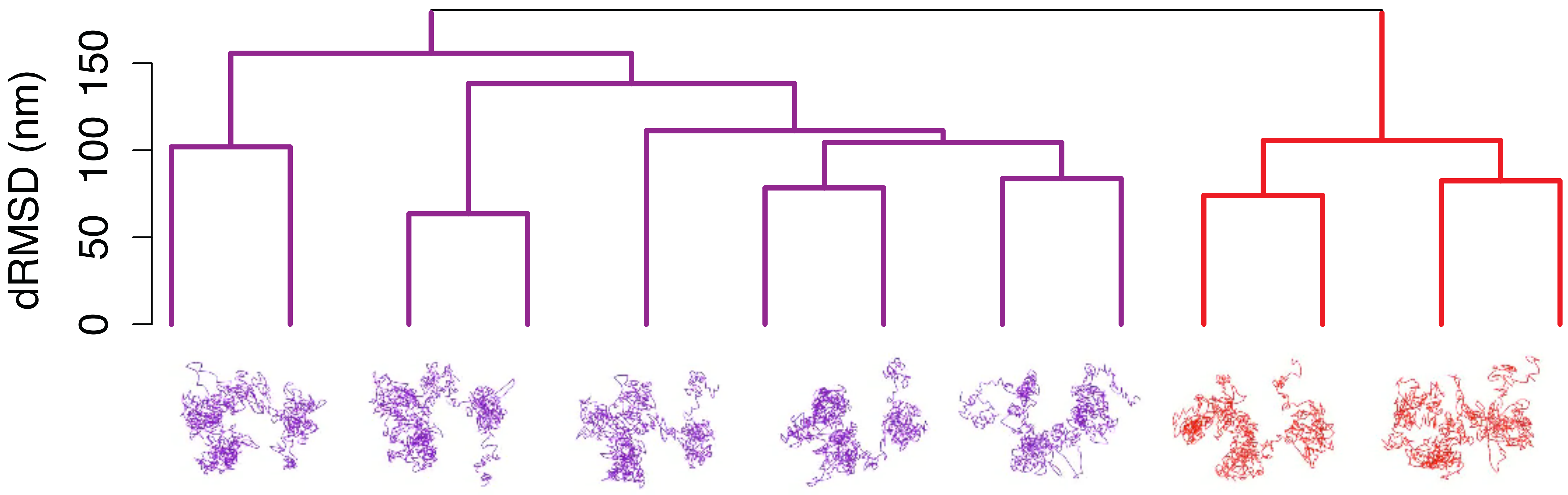
500 rounds

LOWEST OBJECTIVE FUNCTION

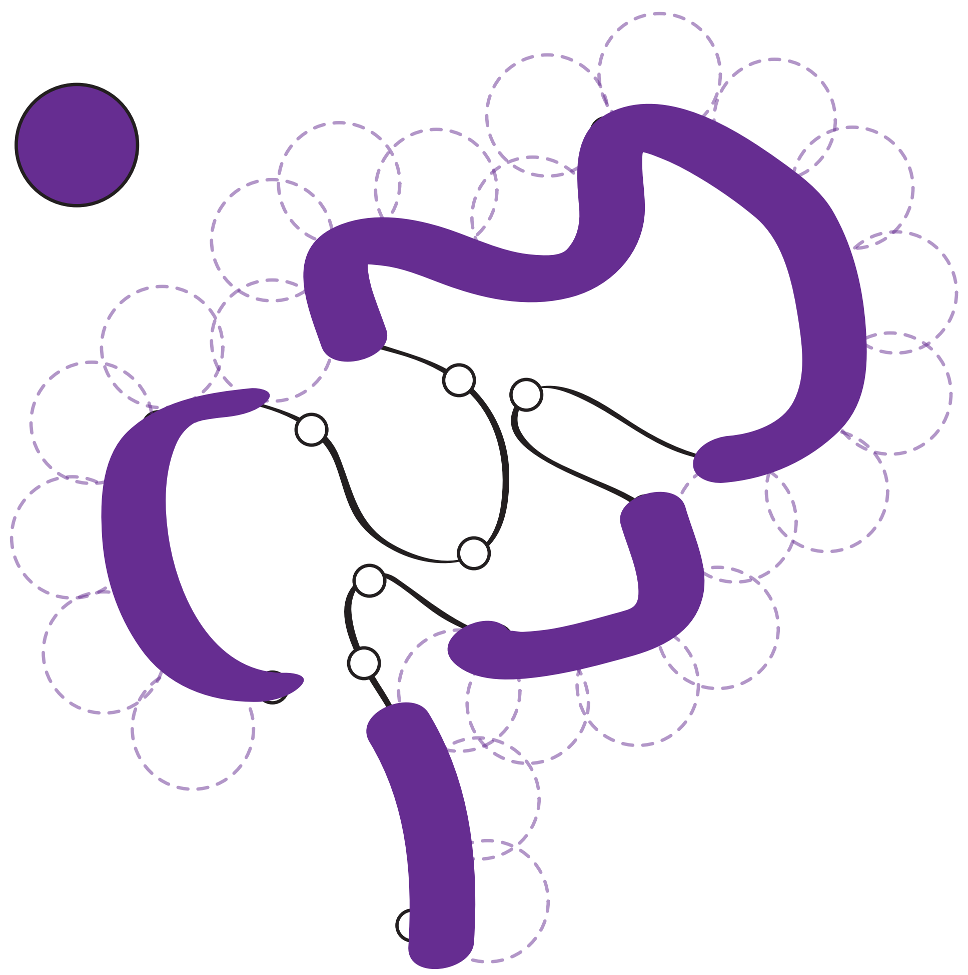
end



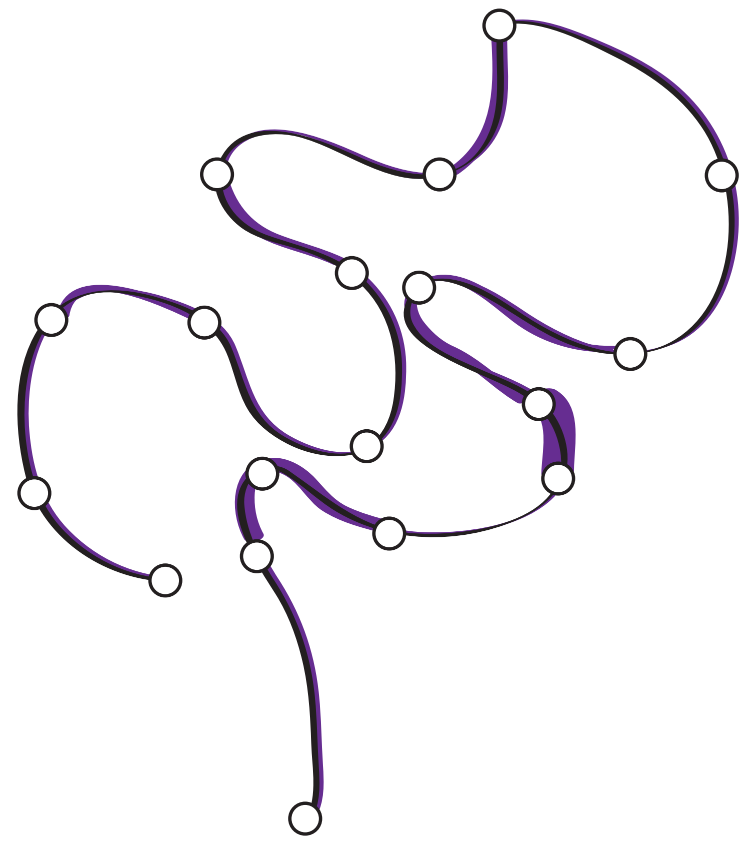
Model analysis: clustering and structural features



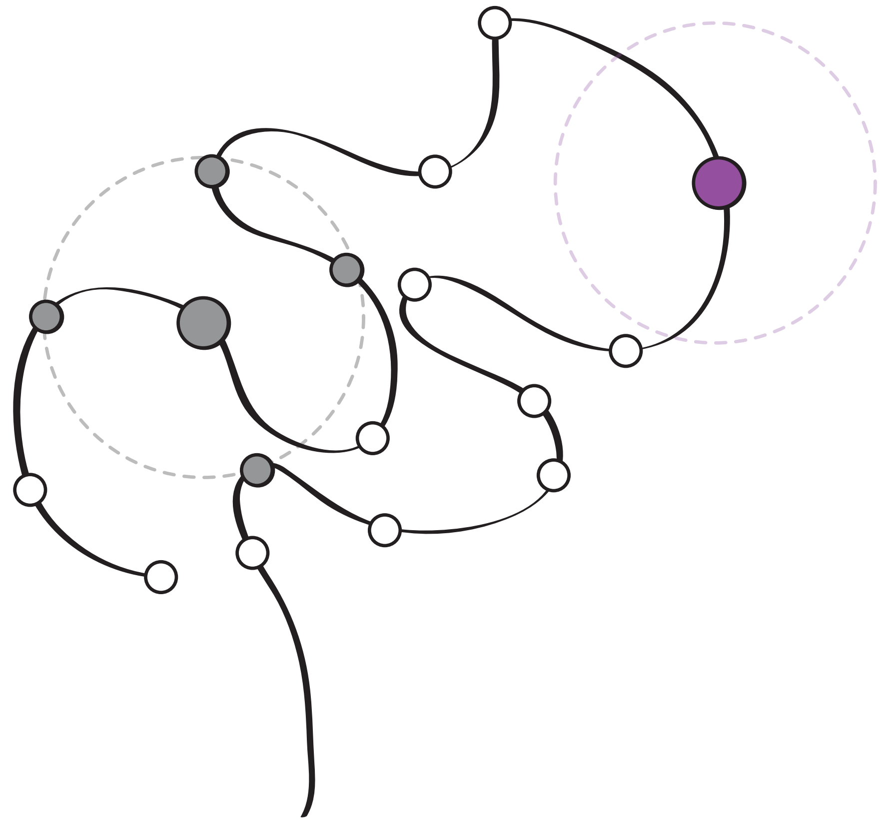
Accessibility (%)



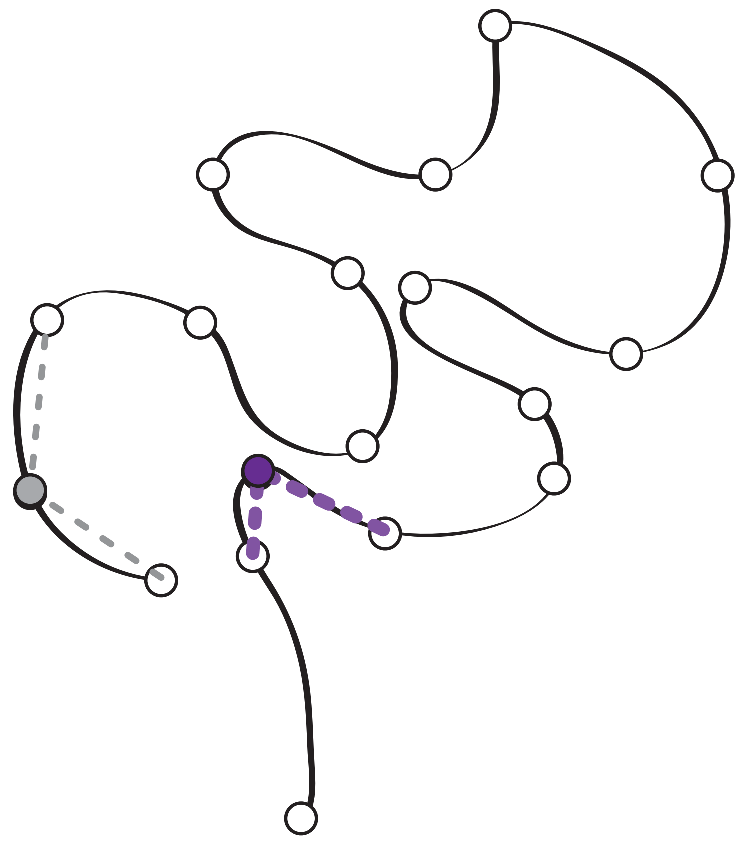
Density (bp/nm)



Interactions

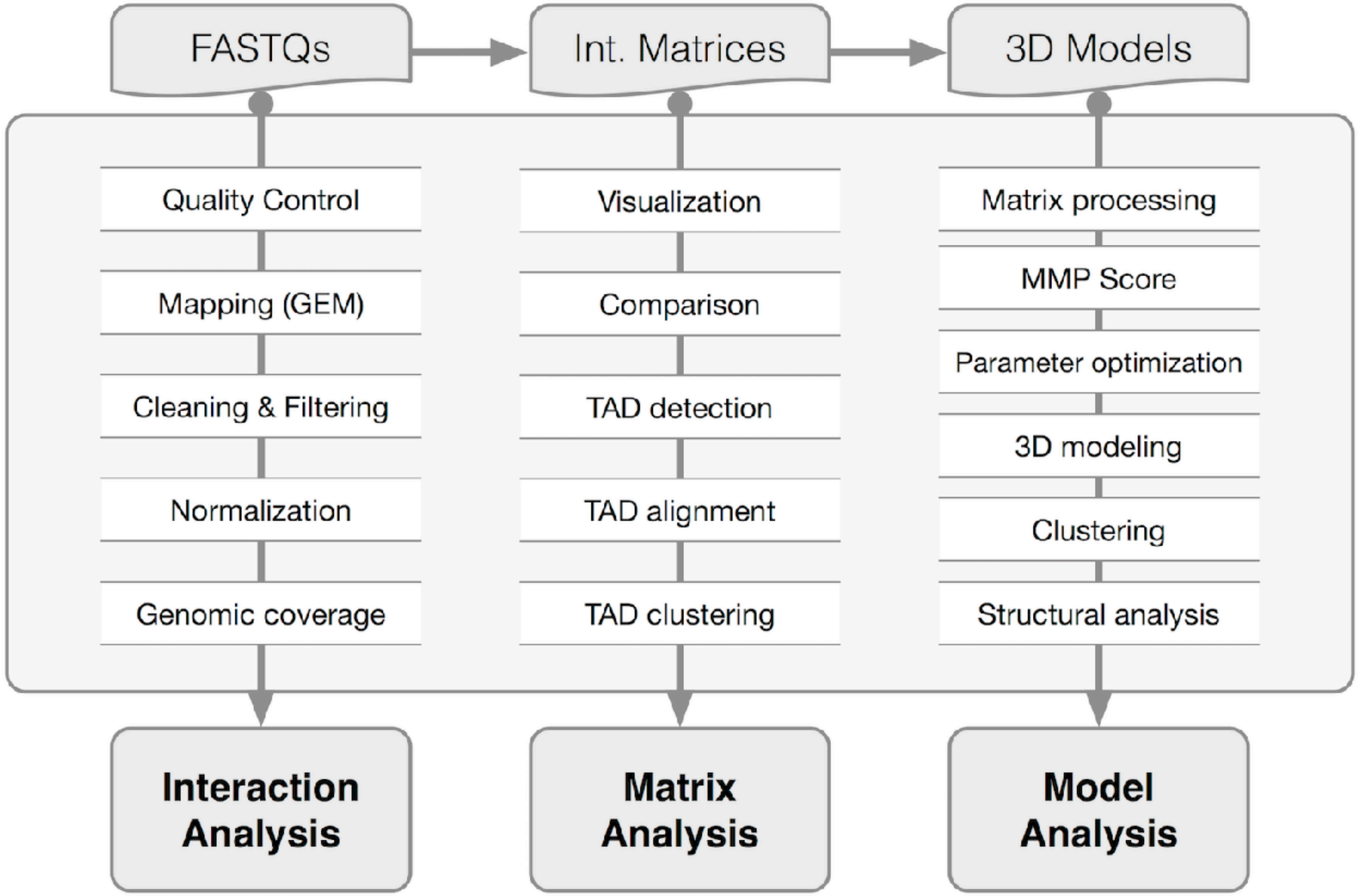


Angle



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

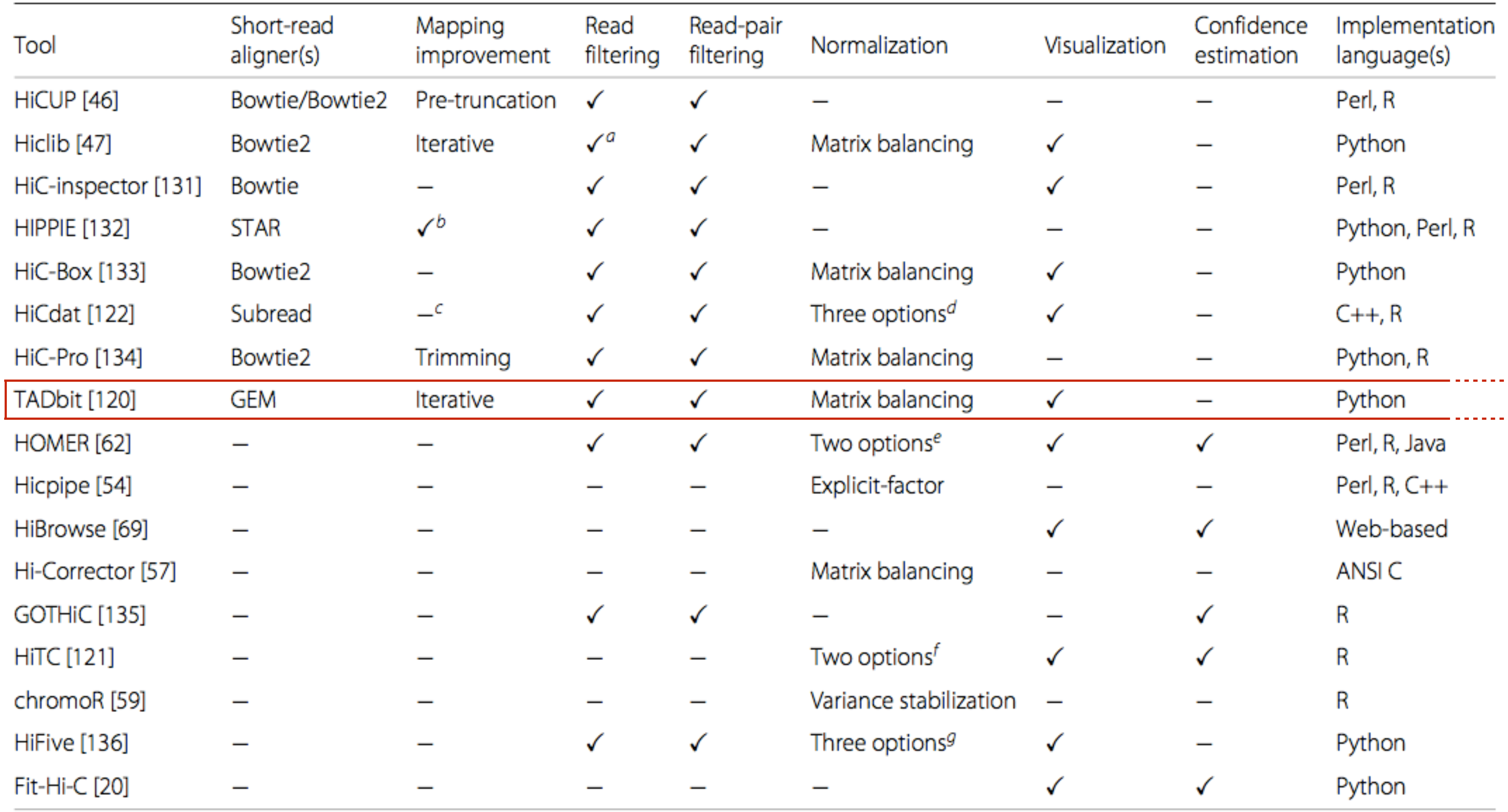
Model Analysis

- 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
BMC Biology 17(1), 55, 2019

https://github.com/3DGenomes/tadbit
https://github.com/3DGenomes/MethodsMolBiol

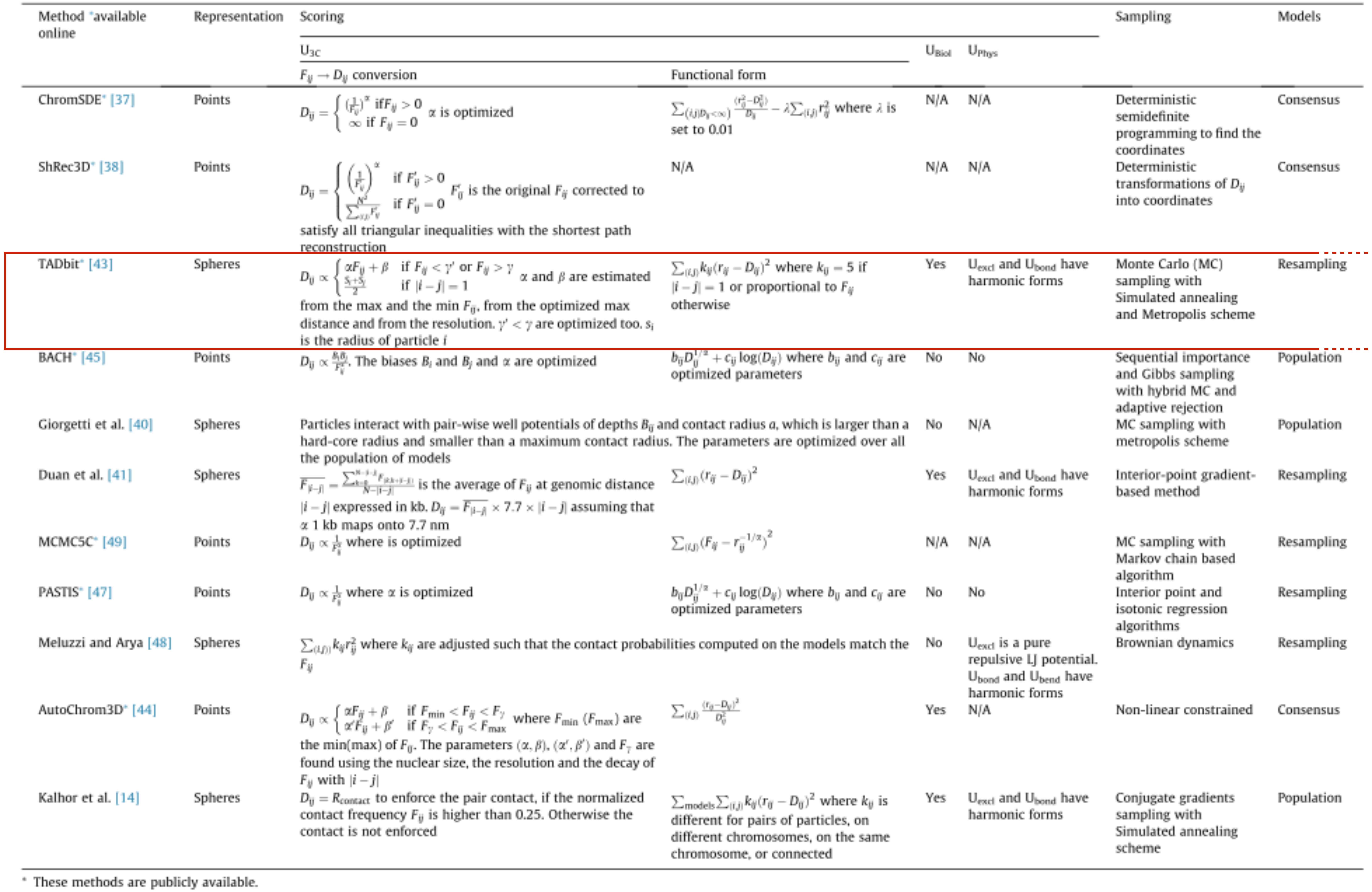
DISCLAIMER — Many alternatives



Tool	Short-read aligner(s)	Mapping improvement	Read filtering	Read-pair filtering	Normalization	Visualization	Confidence estimation	Implementation language(s)
HiCUP [46]	Bowtie/Bowtie2	Pre-truncation	✓	✓	–	–	–	Perl, R
Hiclib [47]	Bowtie2	Iterative	✓[a]	✓	Matrix balancing	✓	–	Python
HiC-inspector [131]	Bowtie	–	✓	✓	–	✓	–	Perl, R
HIPPIE [132]	STAR	✓[b]	✓	✓	–	–	–	Python, Perl, R
HiC-Box [133]	Bowtie2	–	✓	✓	Matrix balancing	✓	–	Python
HiCdat [122]	Subread	–[c]	✓	✓	Three options[d]	✓	–	C++, R
HiC-Pro [134]	Bowtie2	Trimming	✓	✓	Matrix balancing	–	–	Python, R
TADbit [120]	GEM	Iterative	✓	✓	Matrix balancing	✓	–	Python
HOMER [62]	–	–	✓	✓	Two options[e]	✓	✓	Perl, R, Java
Hicpipe [54]	–	–	–	–	Explicit-factor	–	–	Perl, R, C++
HiBrowse [69]	–	–	–	–	–	✓	✓	Web-based
Hi-Corrector [57]	–	–	–	–	Matrix balancing	–	–	ANSI C
GOTHiC [135]	–	–	✓	✓	–	–	✓	R
HiTC [121]	–	–	–	–	Two options[f]	✓	✓	R
chromoR [59]	–	–	–	–	Variance stabilization	–	–	R
HiFive [136]	–	–	✓	✓	Three options[g]	✓	–	Python
Fit-Hi-C [20]	–	–	–	–	–	✓	✓	Python

Analysis methods for studying the 3D architecture of the genome
Ay, F. & Noble, W. S. Genome Biol. 16, 183 (2015).

DISCLAIMER — Many alternatives



Method *available online	Representation	Scoring				Sampling	Models
		U_{3C}		U_{Biol}	U_{Phys}		
		$F_{ij} \rightarrow D_{ij}$ conversion	Functional form				
ChromSDE* [37]	Points	$D_{ij} = \begin{cases} (\frac{1}{F_{ij}})^{\alpha} \text{ if} F_{ij} > 0 \\ \infty \text{ if } F_{ij} = 0 \end{cases}$ α is optimized	$\sum_{(i,j)\vert D_{ij}<\infty)} \frac{(r_{ij}^2 - D_{ij}^2)}{D_{ij}} - \lambda \sum_{(i,j)} r_{ij}^2$ where λ is set to 0.01	N/A	N/A	Deterministic semidefinite programming to find the coordinates	Consensus
ShRec3D* [38]	Points	$D_{ij} = \begin{cases} \left(\frac{1}{F'_{ij}}\right)^{\alpha} & \text{if } F'_{ij} > 0 \\ \frac{N^2}{\sum_{(i,j)} F'_{ij}} & \text{if } F'_{ij} = 0 \end{cases}$ F'_{ij} is the original F_{ij} corrected to satisfy all triangular inequalities with the shortest path reconstruction	N/A	N/A	N/A	Deterministic transformations of D_{ij} into coordinates	Consensus
TADbit* [43]	Spheres	$D_{ij} \propto \begin{cases} \alpha F_{ij} + \beta & \text{if } F_{ij} < \gamma' \text{ or } F_{ij} > \gamma \\ \frac{S_i + S_j}{2} & \text{if } \vert i-j \vert = 1 \end{cases}$ α and β are estimated from the max and the min F_{ij}, from the optimized max distance and from the resolution. $\gamma' < \gamma$ are optimized too. s_i is the radius of particle i	$\sum_{(i,j)} k_{ij}(r_{ij} - D_{ij})^2$ where $k_{ij} = 5$ if $\vert i-j \vert = 1$ or proportional to F_{ij} otherwise	Yes	U_{excl} and U_{bond} have harmonic forms	Monte Carlo (MC) sampling with Simulated annealing and Metropolis scheme	Resampling
BACH* [45]	Points	$D_{ij} \propto \frac{B_i B_j}{F_{ij}^{\alpha}}$. The biases B_i and B_j and α are optimized	$b_{ij} D_{ij}^{1/\alpha} + c_{ij} \log(D_{ij})$ where b_{ij} and c_{ij} are optimized parameters	No	No	Sequential importance and Gibbs sampling with hybrid MC and adaptive rejection	Population
Giorgetti et al. [40]	Spheres	Particles interact with pair-wise well potentials of depths B_{ij} and contact radius a, which is larger than a hard-core radius and smaller than a maximum contact radius. The parameters are optimized over all the population of models		No	N/A	MC sampling with metropolis scheme	Population
Duan et al. [41]	Spheres	$\overline{F_{\vert i-j \vert}} = \frac{\sum_{k=0}^{N-\vert i-j \vert} F_{(k,k+\vert i-j \vert)}}{N-\vert i-j \vert}$ is the average of F_{ij} at genomic distance $\vert i-j \vert$ expressed in kb. $D_{ij} = \overline{F_{\vert i-j \vert}} \times 7.7 \times \vert i-j \vert$ assuming that α 1 kb maps onto 7.7 nm	$\sum_{(i,j)} (r_{ij} - D_{ij})^2$	Yes	U_{excl} and U_{bond} have harmonic forms	Interior-point gradient-based method	Resampling
MCMC5C* [49]	Points	$D_{ij} \propto \frac{1}{F_{ij}^{\alpha}}$ where is optimized	$\sum_{(i,j)} (F_{ij} - r_{ij}^{-1/\alpha})^2$	N/A	N/A	MC sampling with Markov chain based algorithm	Resampling
PASTIS* [47]	Points	$D_{ij} \propto \frac{1}{F_{ij}^{\alpha}}$ where α is optimized	$b_{ij} D_{ij}^{1/\alpha} + c_{ij} \log(D_{ij})$ where b_{ij} and c_{ij} are optimized parameters	No	No	Interior point and isotonic regression algorithms	Resampling
Meluzzi and Arya [48]	Spheres	$\sum_{(i,j)} k_{ij} r_{ij}^2$ where k_{ij} are adjusted such that the contact probabilities computed on the models match the F_{ij}		No	U_{excl} is a pure repulsive LJ potential. U_{bond} and U_{bend} have harmonic forms	Brownian dynamics	Resampling
AutoChrom3D* [44]	Points	$D_{ij} \propto \begin{cases} \alpha F_{ij} + \beta & \text{if } F_{min} < F_{ij} < F_{\gamma} \\ \alpha' F_{ij} + \beta' & \text{if } F_{\gamma} < F_{ij} < F_{max} \end{cases}$ where F_{min} (F_{max}) are the min(max) of F_{ij}. The parameters (α, β), (α', β') and F_{γ} are found using the nuclear size, the resolution and the decay of F_{ij} with $\vert i-j \vert$	$\sum_{(i,j)} \frac{(r_{ij} - D_{ij})^2}{D_{ij}^2}$	Yes	N/A	Non-linear constrained	Consensus
Kalhor et al. [14]	Spheres	$D_{ij} = R_{contact}$ to enforce the pair contact, if the normalized contact frequency F_{ij} is higher than 0.25. Otherwise the contact is not enforced	$\sum_{models} \sum_{(i,j)} k_{ij}(r_{ij} - D_{ij})^2$ where k_{ij} is different for pairs of particles, on different chromosomes, on the same chromosome, or connected	Yes	U_{excl} and U_{bond} have harmonic forms	Conjugate gradients sampling with Simulated annealing scheme	Population

* These methods are publicly available.

Restraint-based three-dimensional modeling of genomes and genomic domains.
Serra F, Di Stefano M, Spill YG, Cuartero Y, Goodstadt M, Baù D, Marti-Renom MA. FEBS Lett 589: 2987–2995 (2015)

What next?

Can we see the path of chromatin @10Kb resolution? images + Hi-C + modeling

Is there a dynamic coupling between structure and gene activity?

Is genome structure more conserved than sequence?

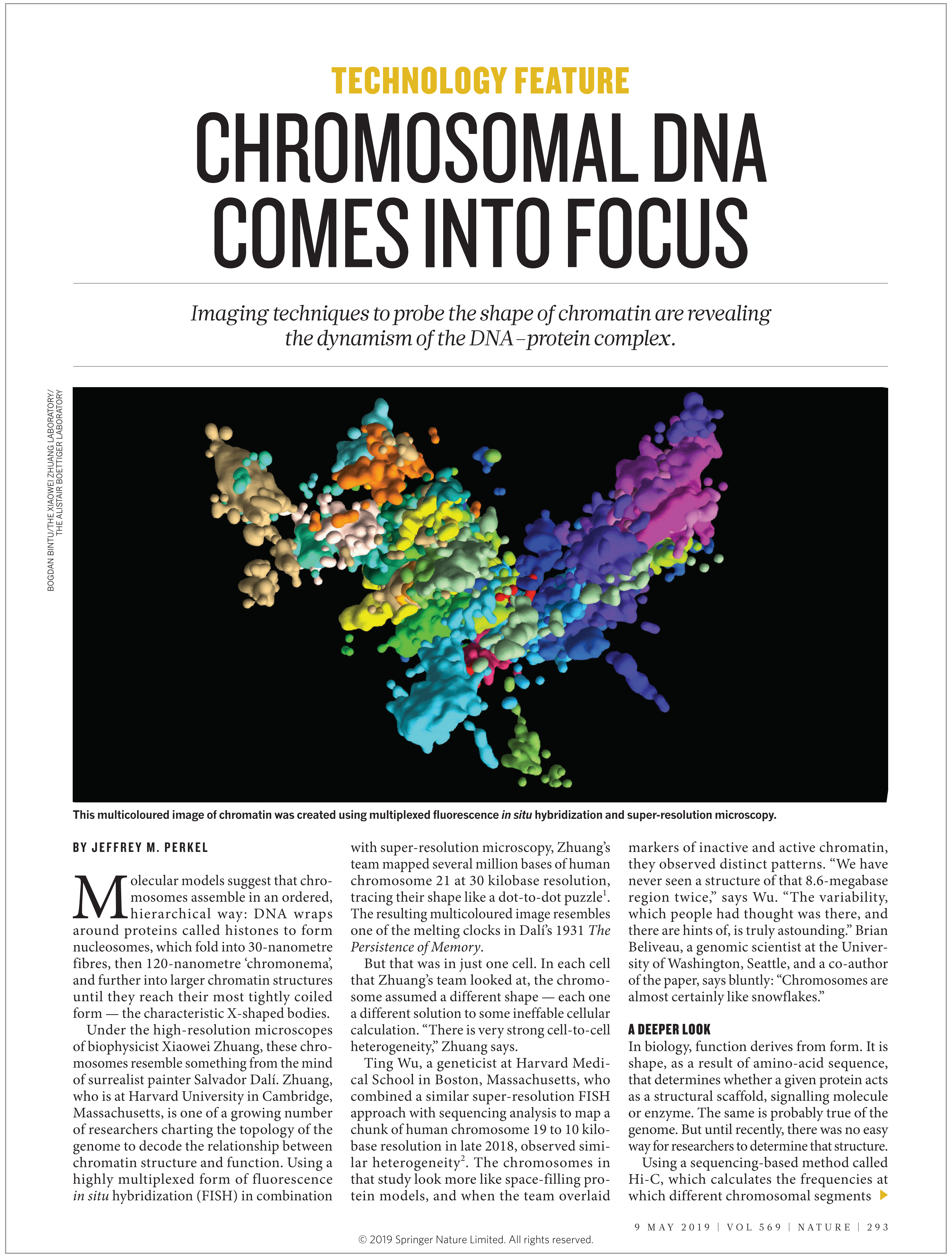
Photo by David Oliete - www.davidoliete.com

Chromosome walking with super-resolution imaging and modeling

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

PLOS Genetics (2018) 14(12) e1007872

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

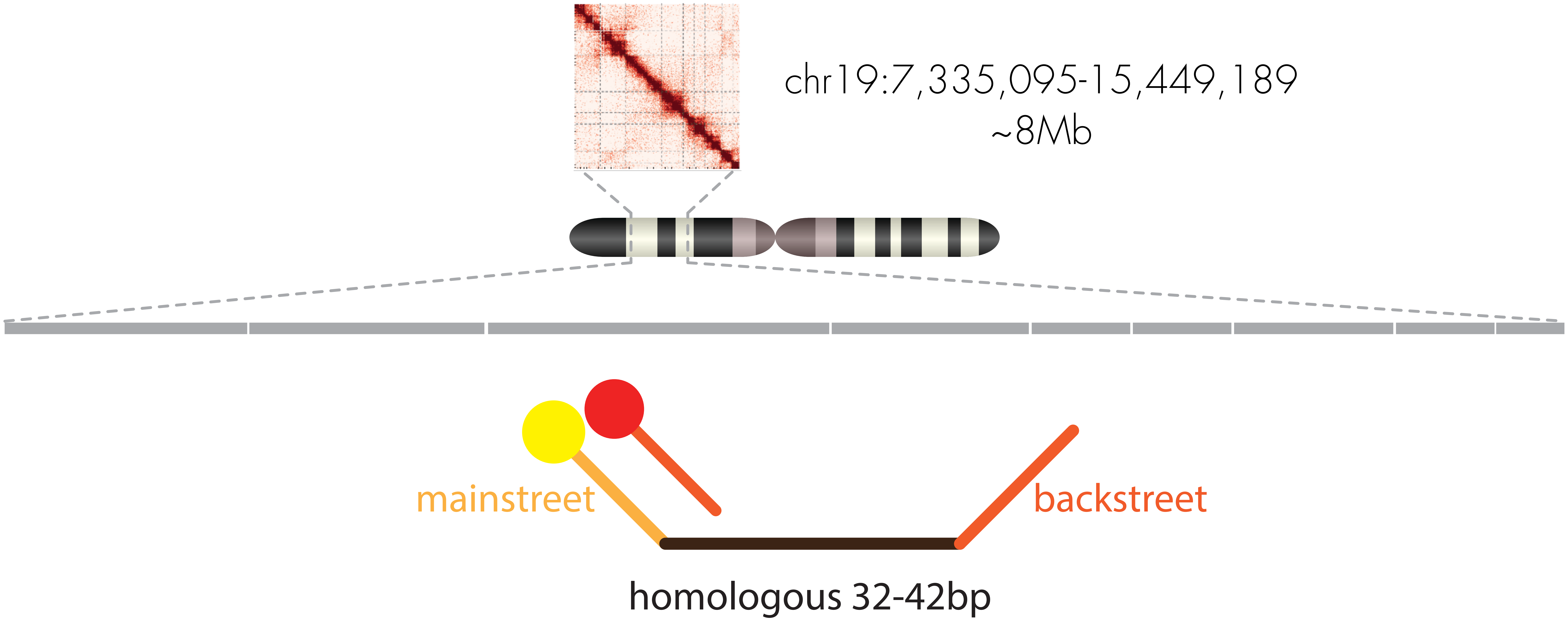
In biology, function derives from form. It is shape, as a result of amino-acid sequence, that determines whether a given protein acts as a structural scaffold, signalling molecule or enzyme. The same is probably true of the genome. But until recently, there was no easy way for researchers to determine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

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

1. **Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells.** Bintu B, Mateo LJ, Su JH, Sinnott-Armstrong NA, Parker M, Kinrot S, Yamaya K, Boettiger AN, Zhuang X.
   Science. 2018 10 26; 362(6413)
   https://doi.org/10.1126/science.aau1783 PMID: 30361340
2. **Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling.** Nir G, Farabella I, Pérez Estrada C, Ebeling CG, Beliveau BJ, Sasaki HM, Lee SD, Nguyen SC, McCole RB, Chattoraj S, Erceg J, AlHaj Abed J, Martins NMC, Nguyen HQ, Hannan MA, Russell S, Durand NC, Rao SSP, Kishi JY, Soler-Vila P, Di Pierro M, Onuchic JN, Callahan SP, Schreiner JM, Stuckey JA, Yin P, Aiden EL, Marti-Renom MA, Wu CT.
   PLoS Genet. 2018 12; 14(12):e1007872
   https://doi.org/10.1371/journal.pgen.1007872 PMID: 30586358
3. **Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms.** Cardozo Gizzi AM, Cattoni DI, Fiche JB, Espinola SM, Gurgo J, Messina O, Houbron C, Ogiyama Y, Papadopoulos GL, Cavalli G, Lagha M, Nollmann M.
   Mol Cell. 2019 Feb 12;
   https://doi.org/10.1016/j.molcel.2019.01.011 PMID: 30795893
4. **Visualizing DNA folding and RNA in embryos at single-cell resolution.** Mateo LJ, Murphy SE, Hafner A, Cinquini IS, Walker CA, Boettiger AN.
   Nature. 2019 Mar 18;
   https://doi.org/10.1038/s41586-019-1035-4 PMID: 30886393

High-resolution imaging

Tracing chromosomes with OligoSTORM & fluidics cycles in PGP1 cells



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

mainstreet

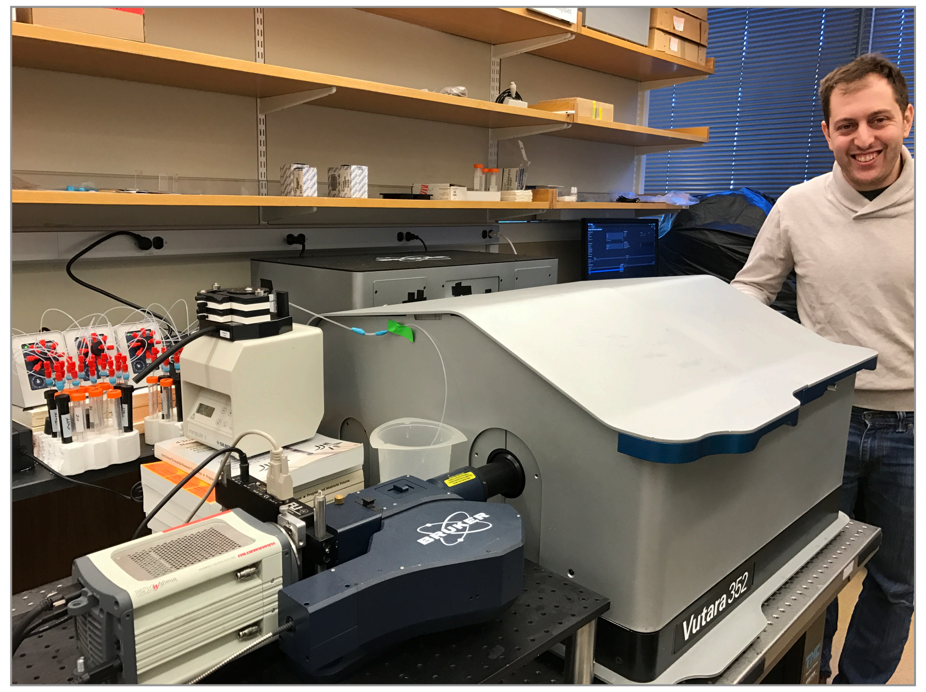
backstreet

homologous 32-42bp

Beliveau et al. Nat. Comm. 2015

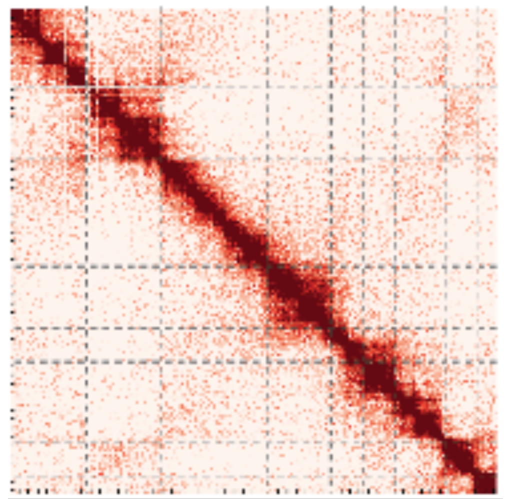
High-resolution imaging

Tracing chromosomes with OligoSTORM & fluidics cycles in PGP1 cells

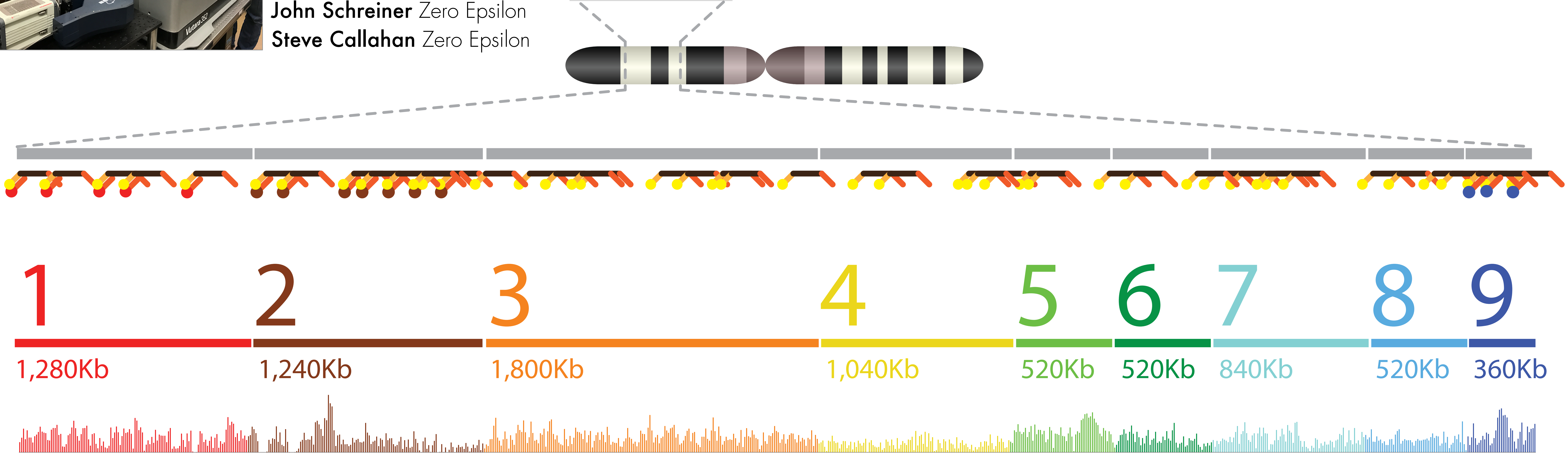


Guy Nir Harvard Med School

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



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



High-resolution imaging

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

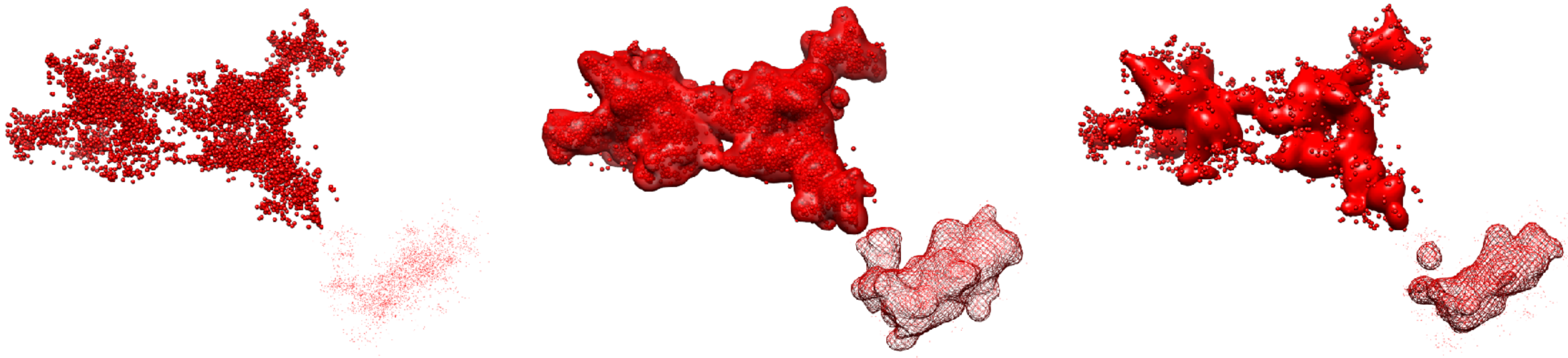
1	2	3	4	5	6	7	8	9
1,280Kb	1,240Kb	1,800Kb	1,040Kb	520Kb	520Kb	840Kb	520Kb	360Kb

Cell-02

High-resolution imaging

XYZ points convolution into a density map

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

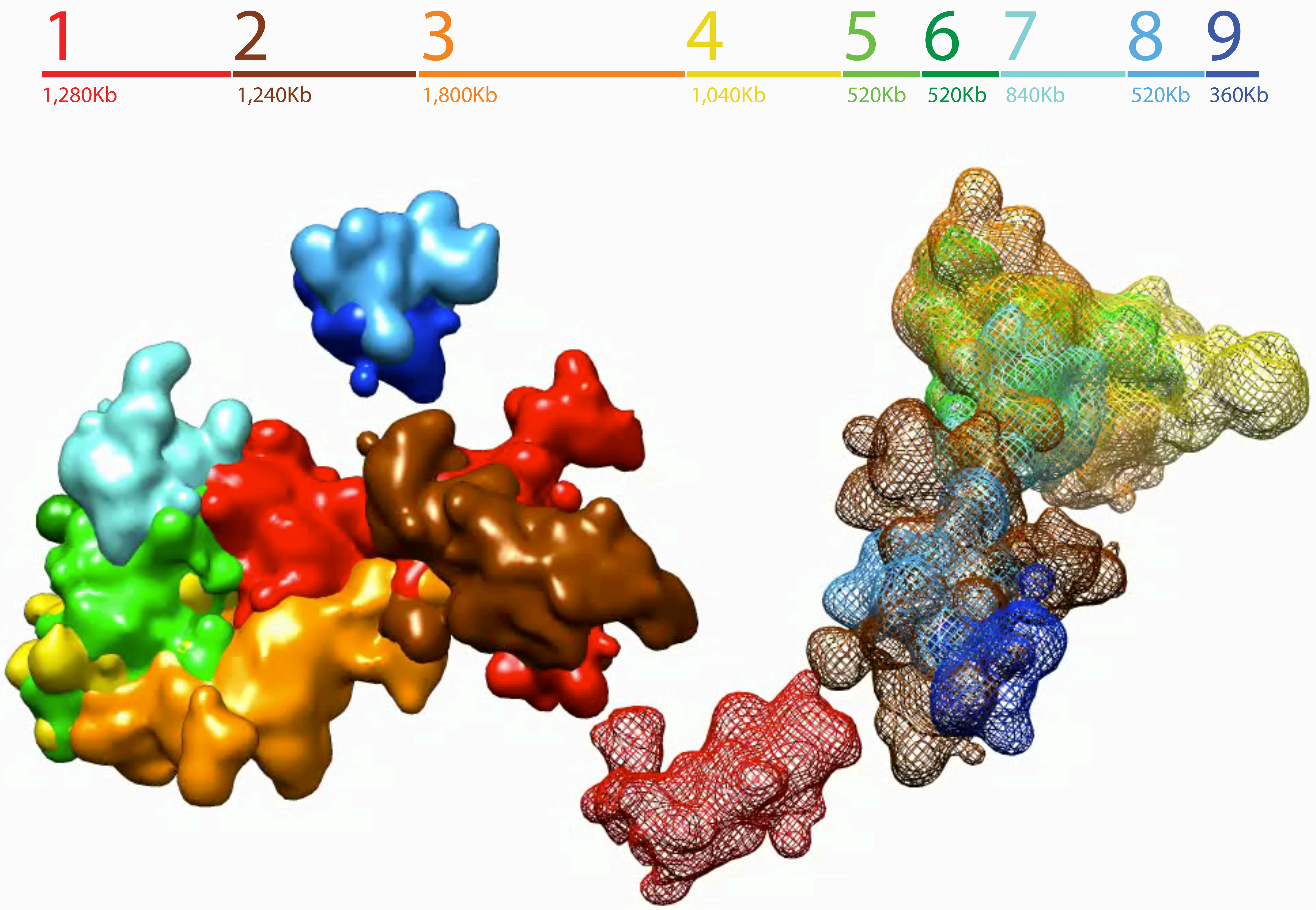


Cell-02 · Segment 1

Farabella et al, J Appl Crystallogr. 2015

Density maps

Cell-02 · Density map @ 50nm



Area (nm²)

Volume (nm³)

Sphericity

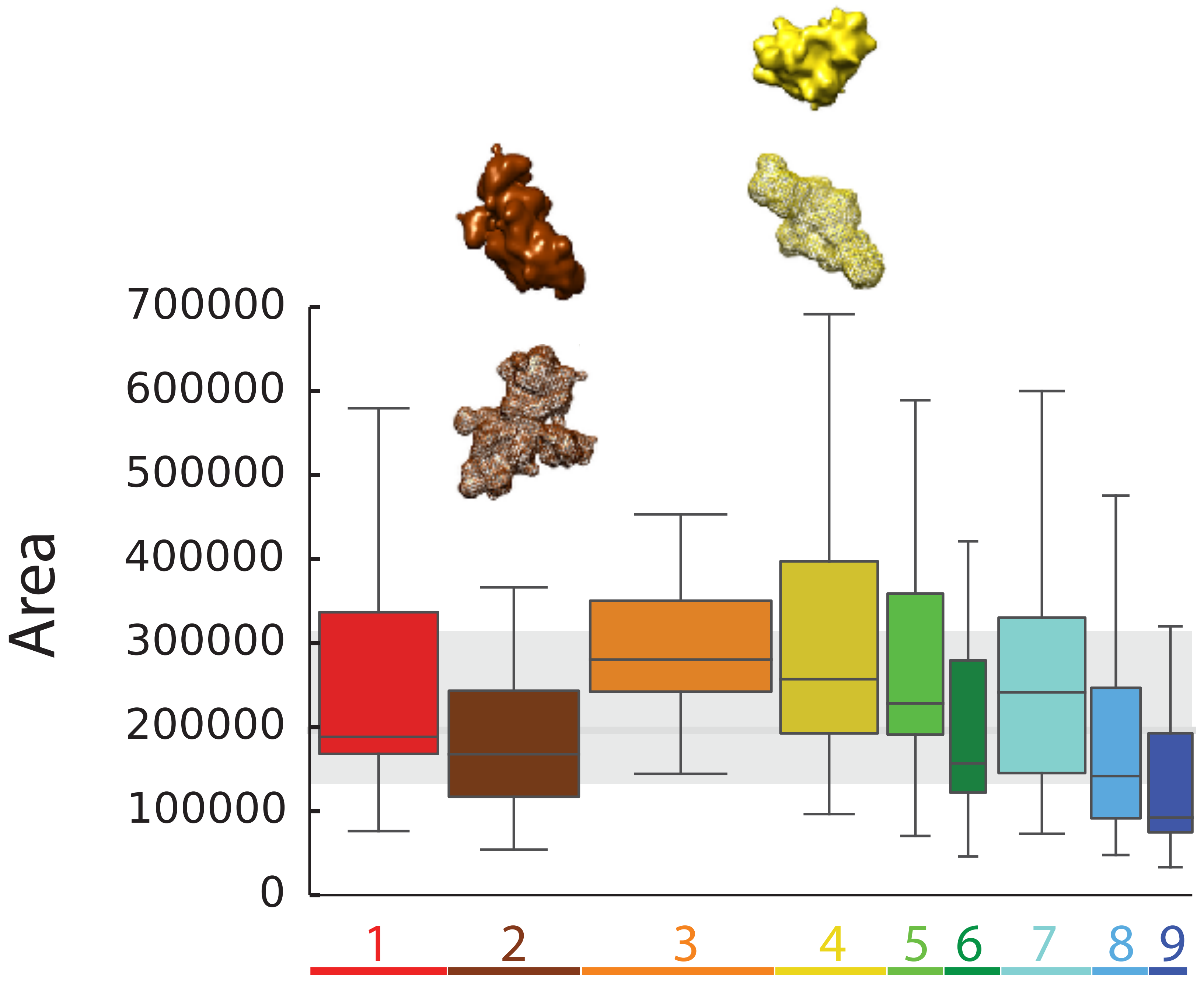
Overlap (%)

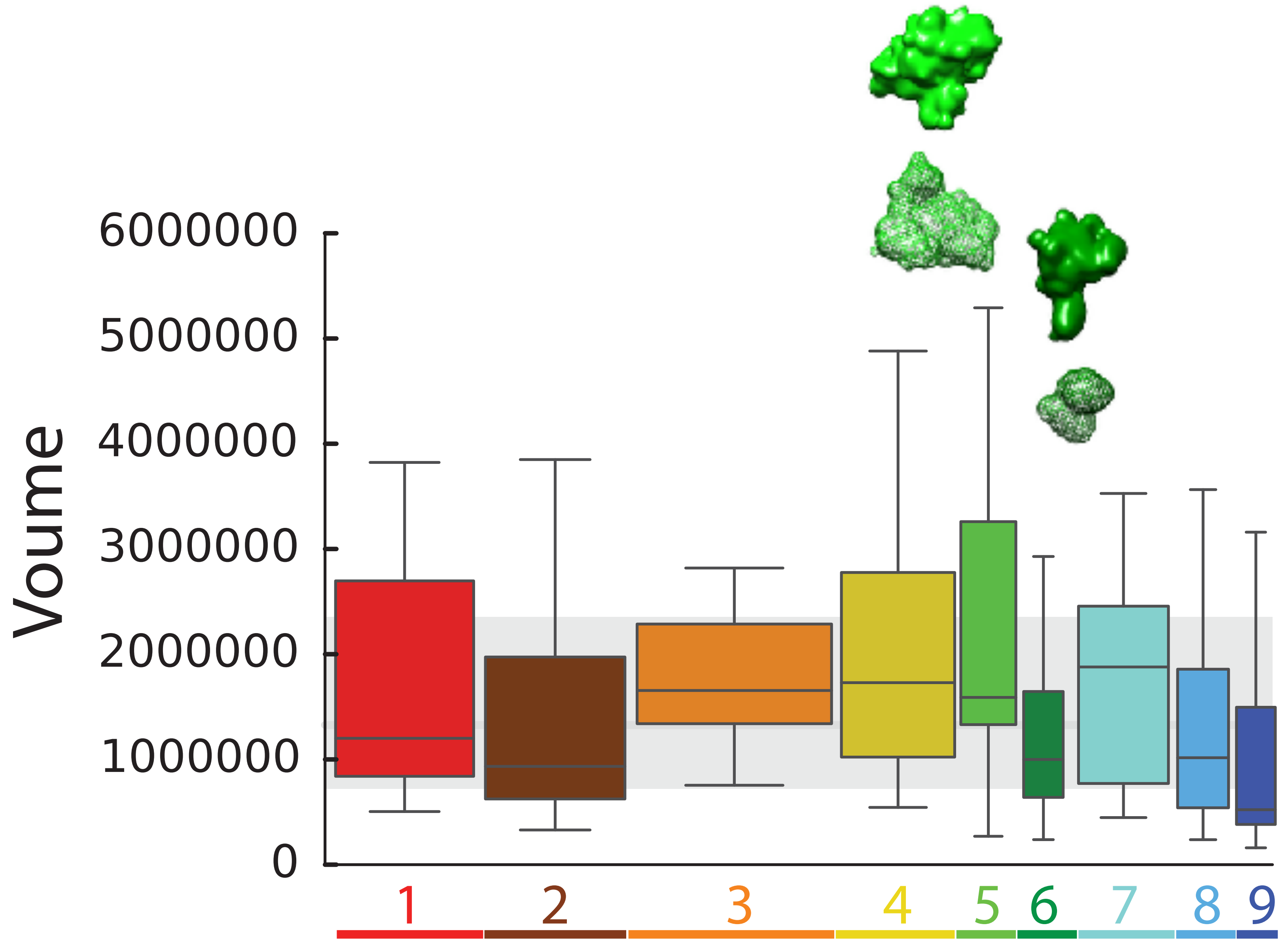
Distance (nm)

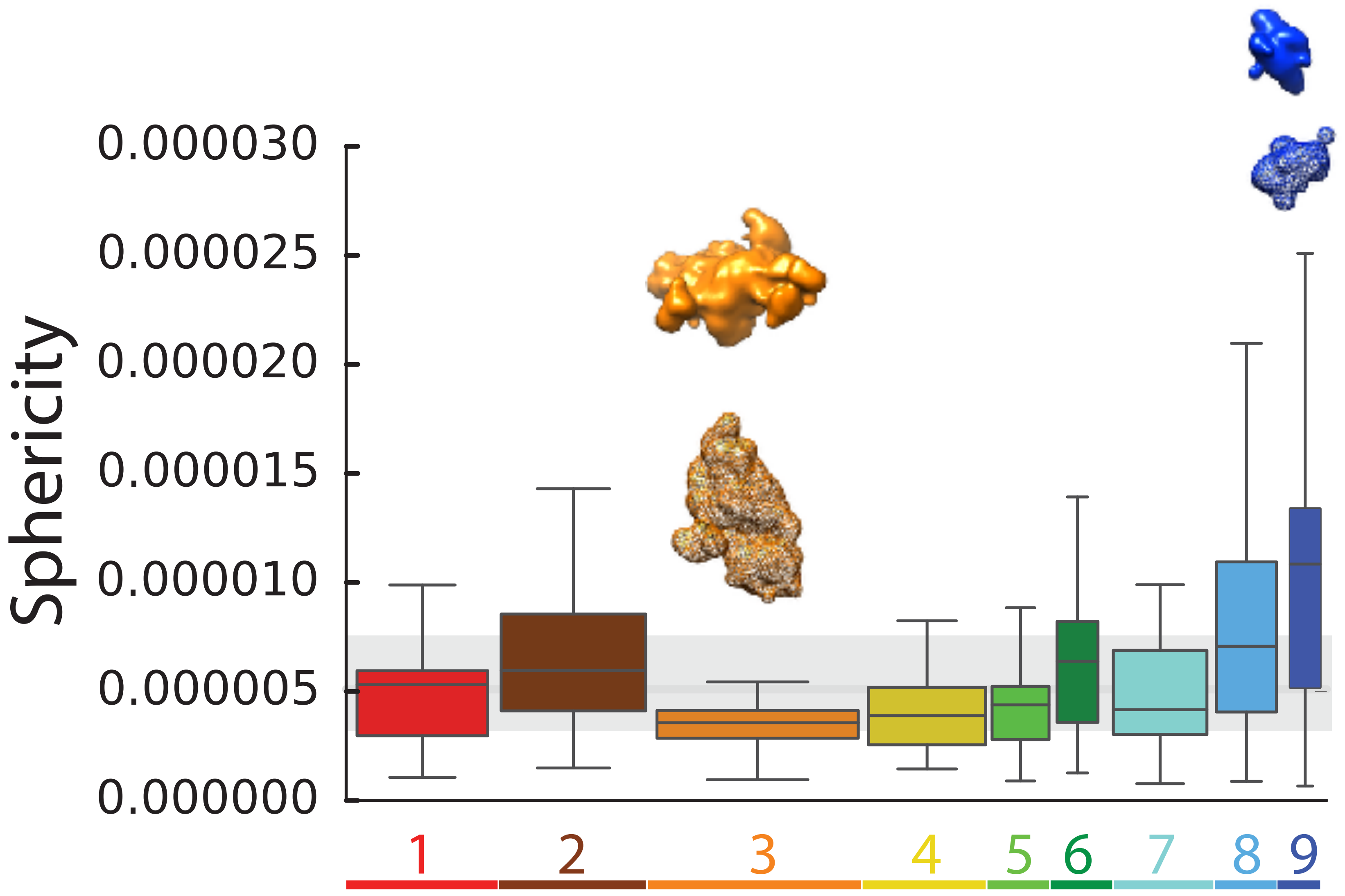
Farabella et al, J Appl Crystallogr. 2015

Structural features

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

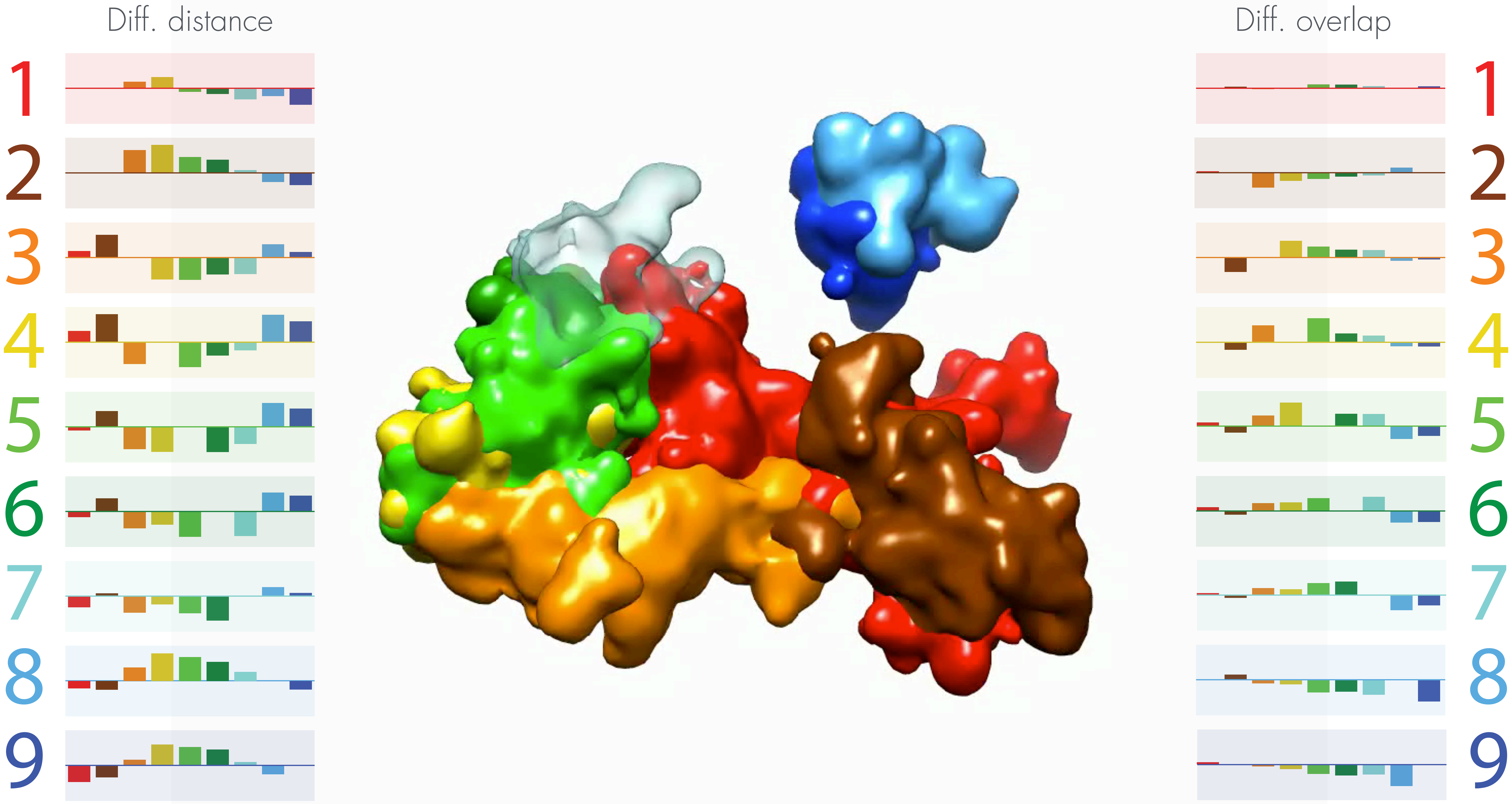






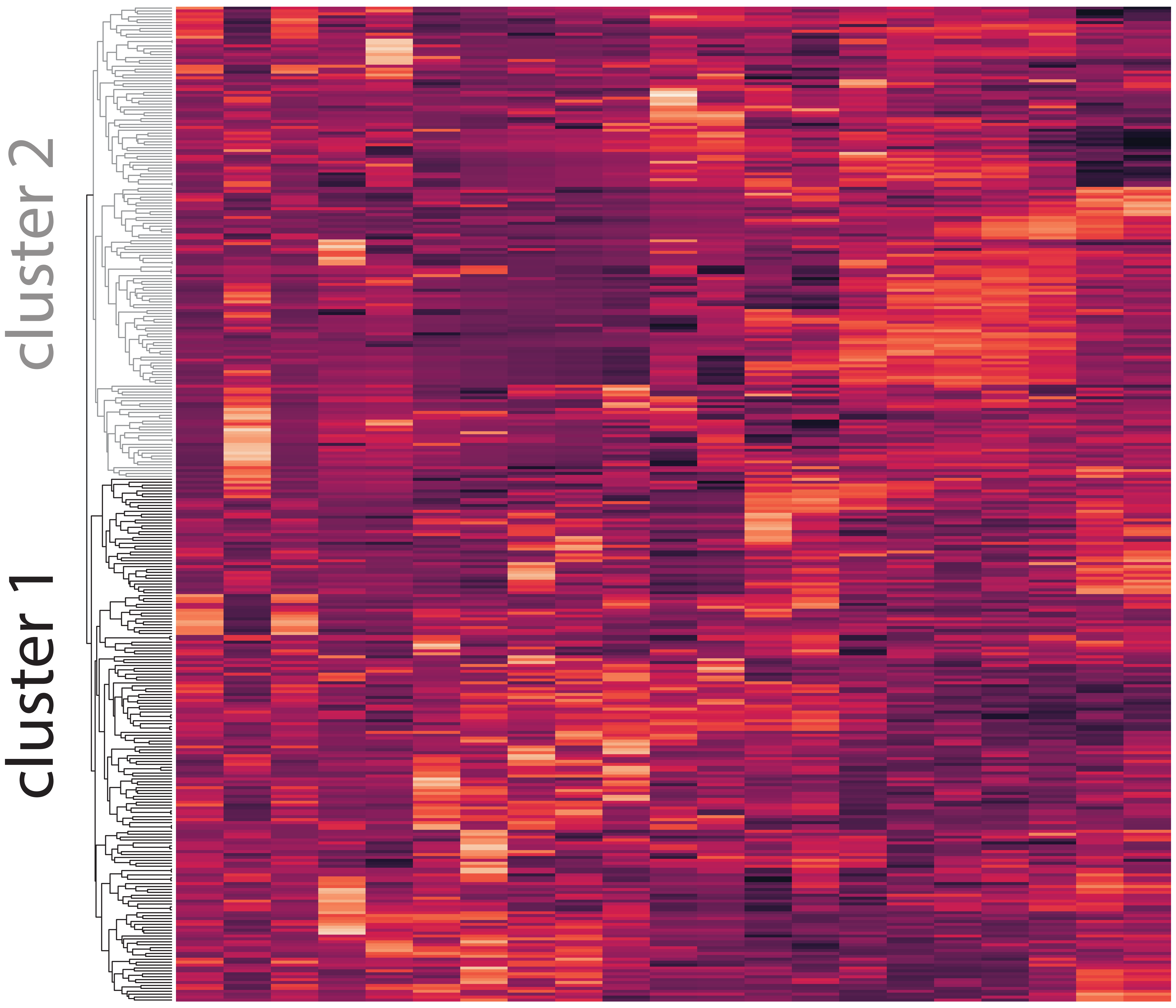
Spatial arrangement

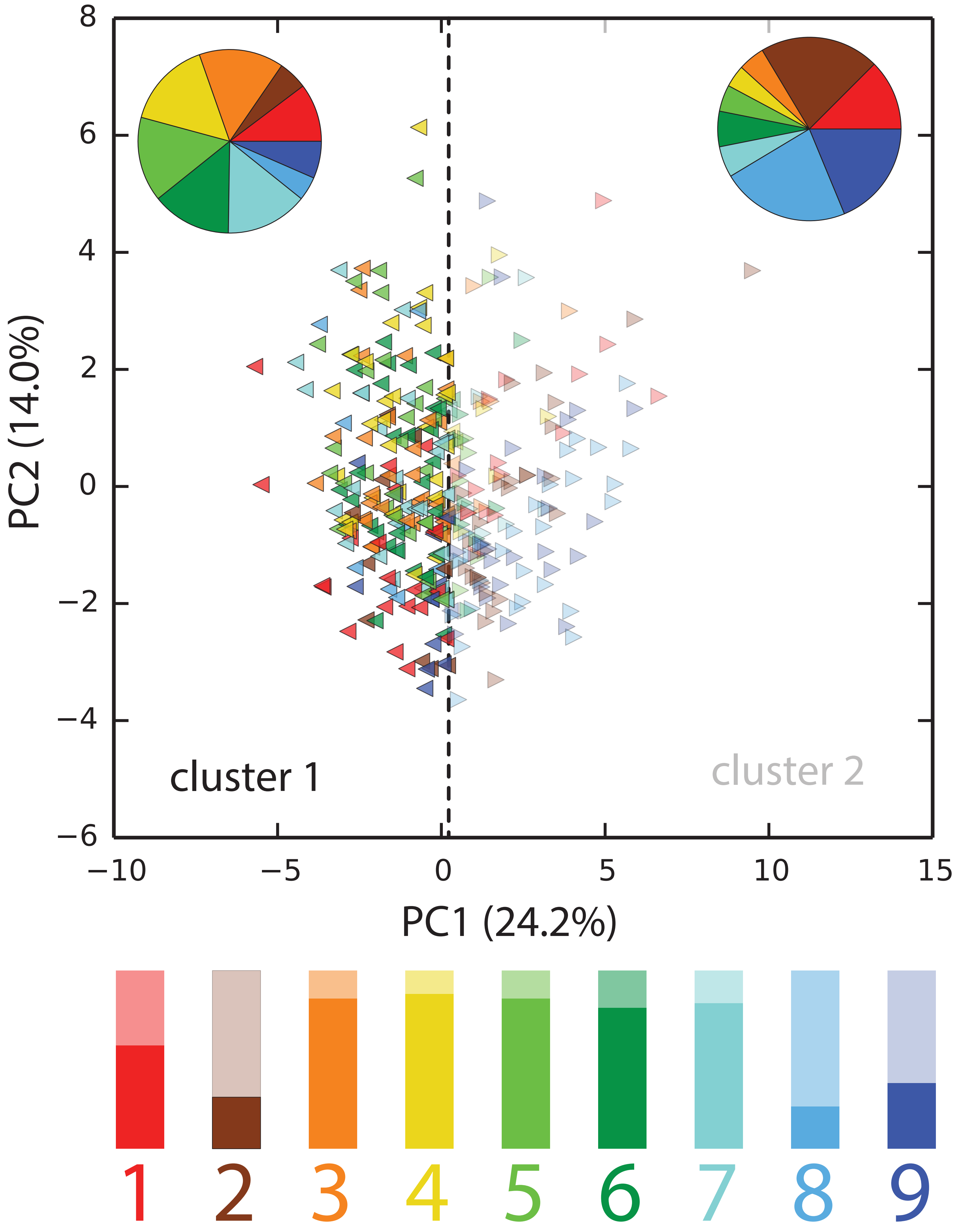
Distance and overlap of 19 cells each with 2 homologous resolved



Structural clustering

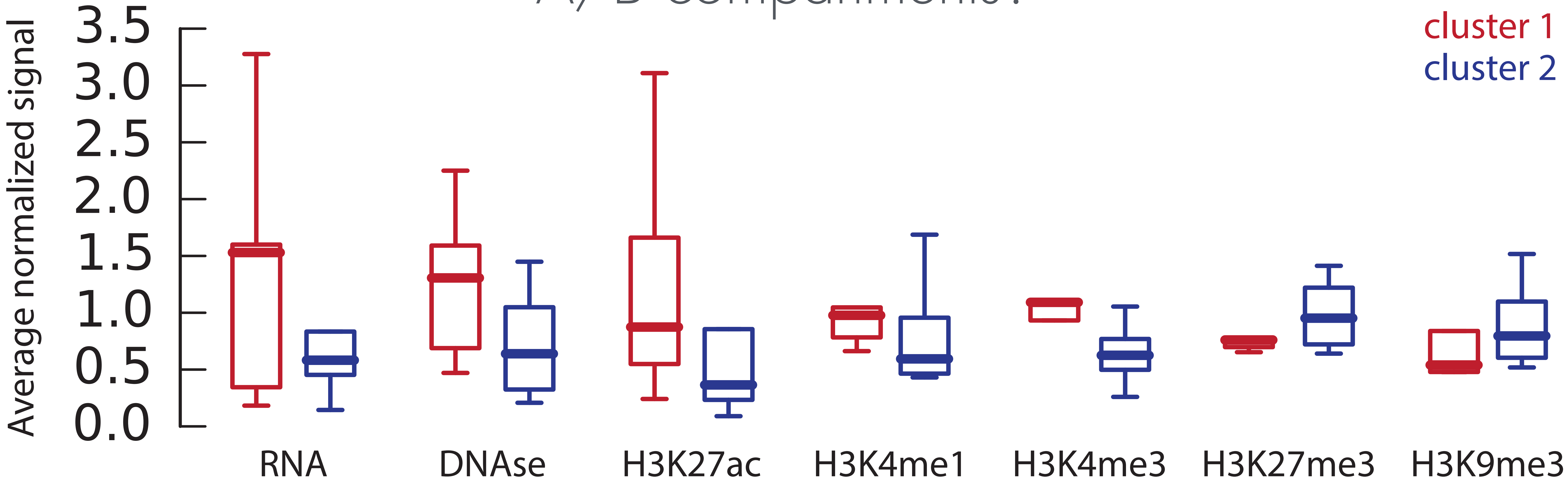
19 cells each with 2 homologous and 9 segments each (342)

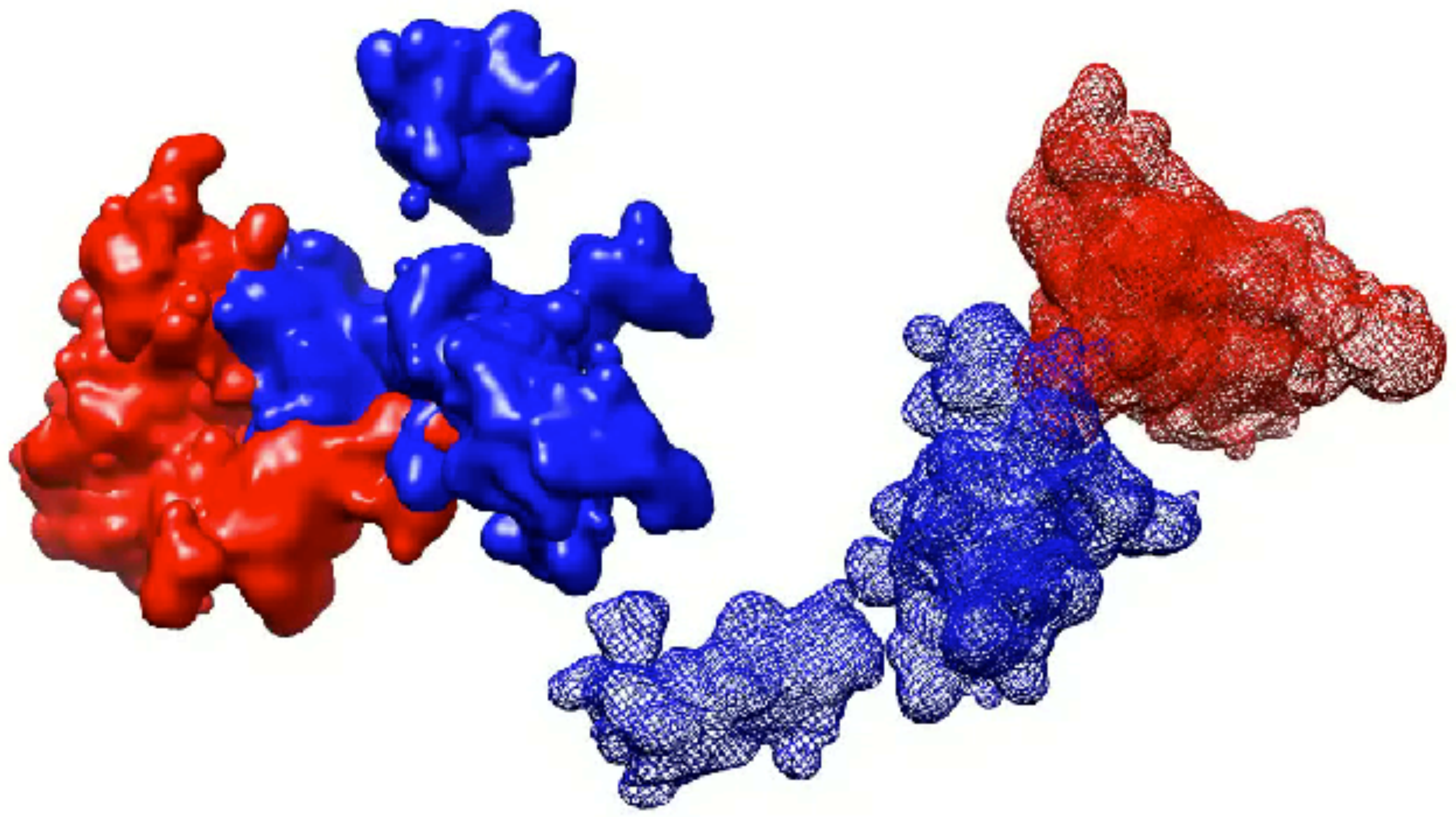


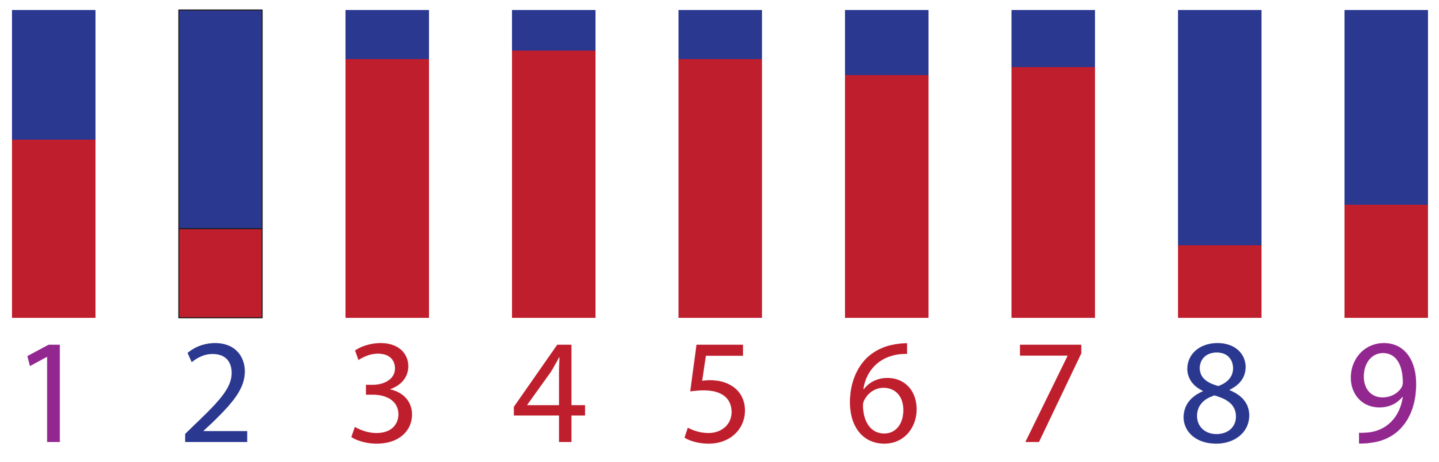


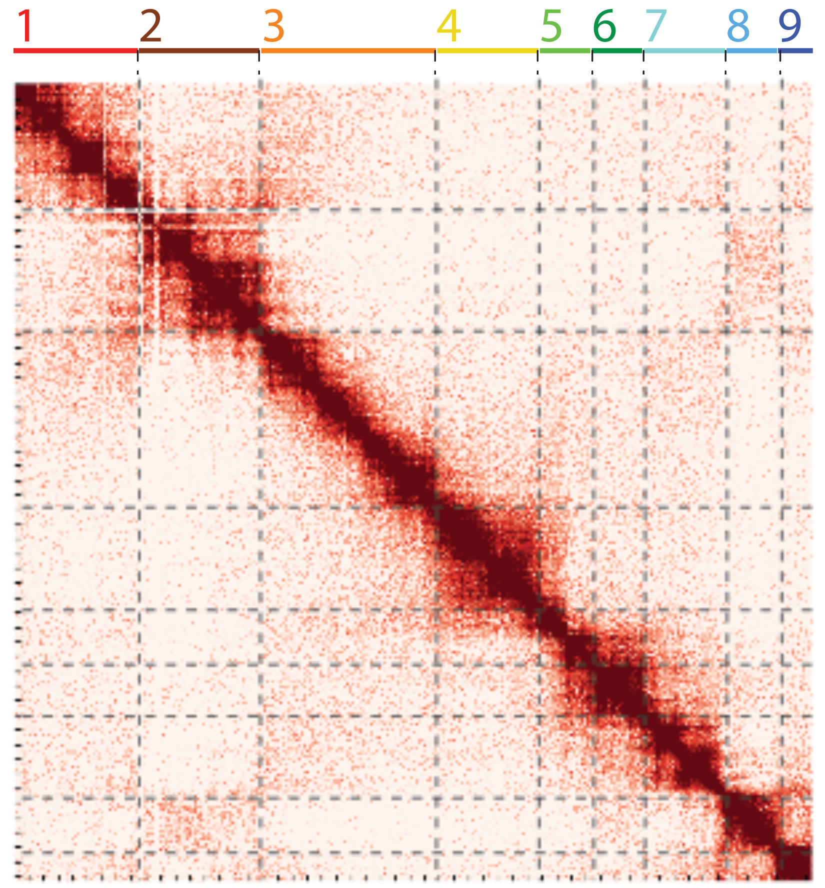
Cluster properties

A/B compartments?





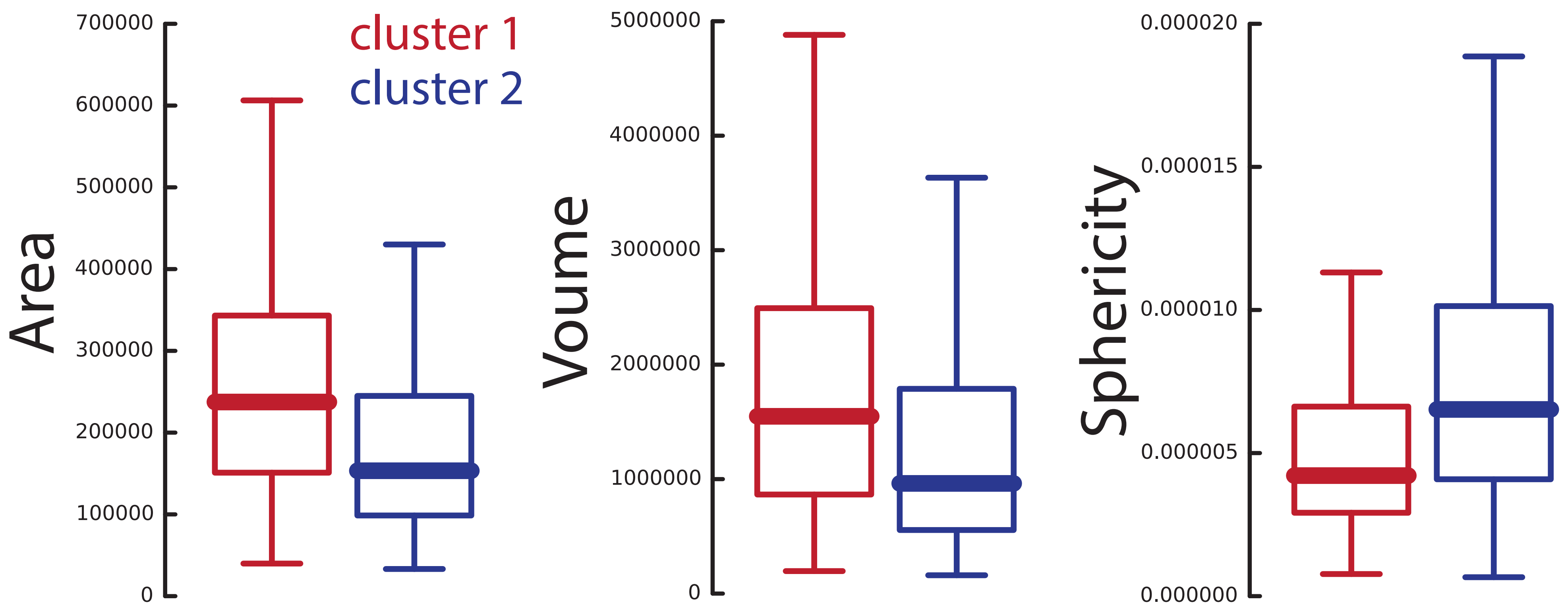


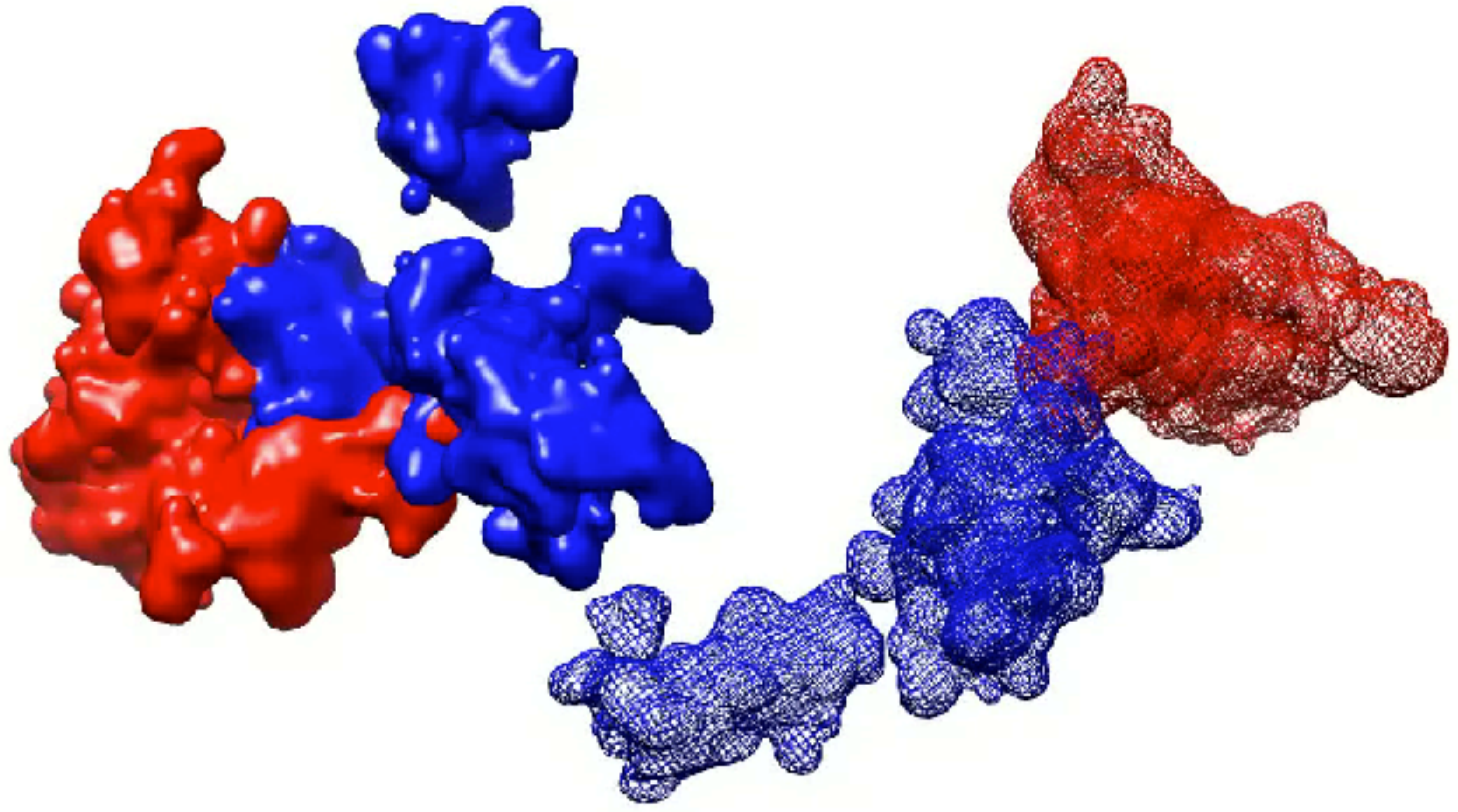


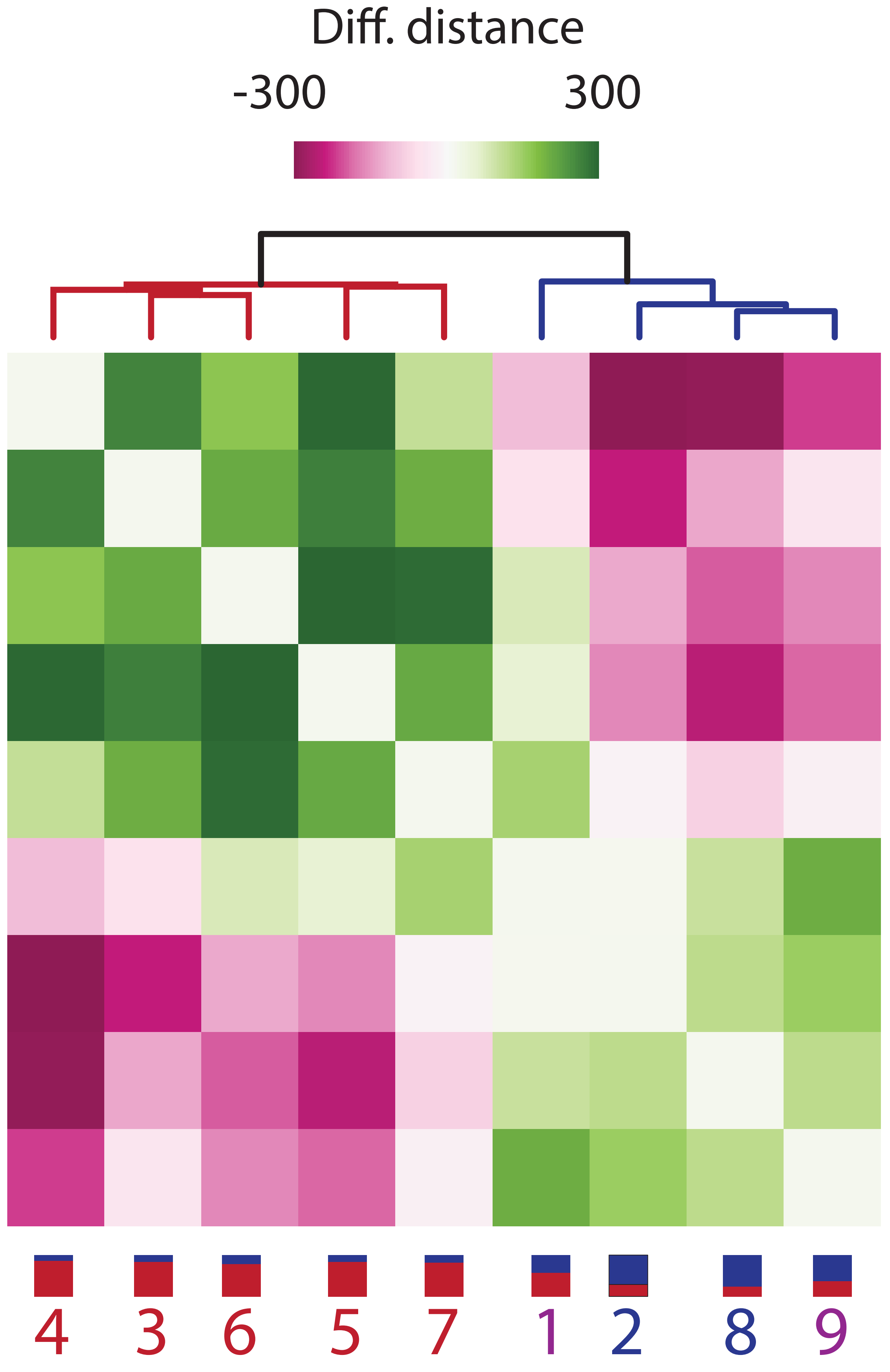
PGP1 ChIP-seq and Hi-C data from ENCODE and Lieberman-Aiden Lab, respectively

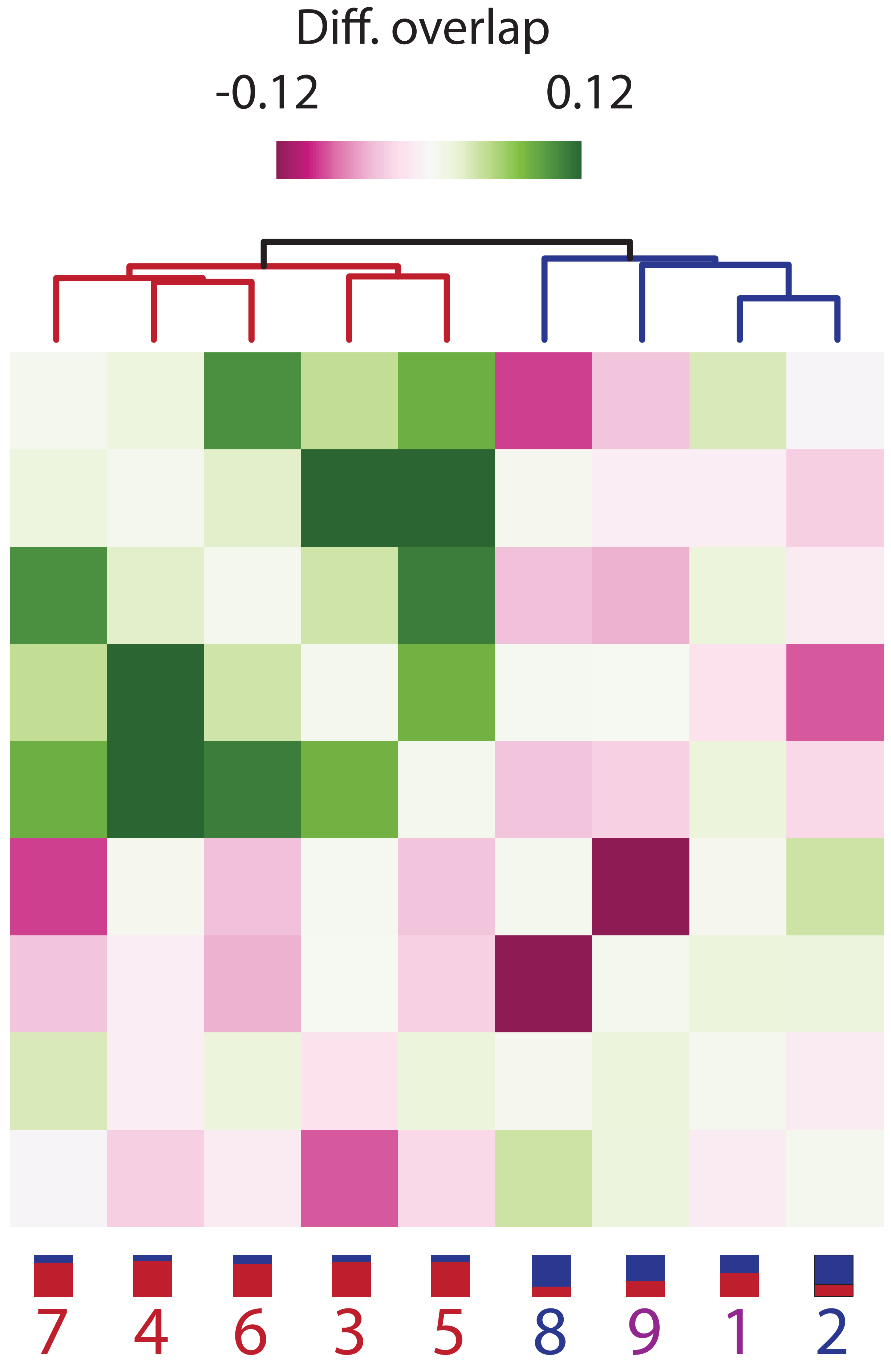
Cluster properties

A/B compartment properties



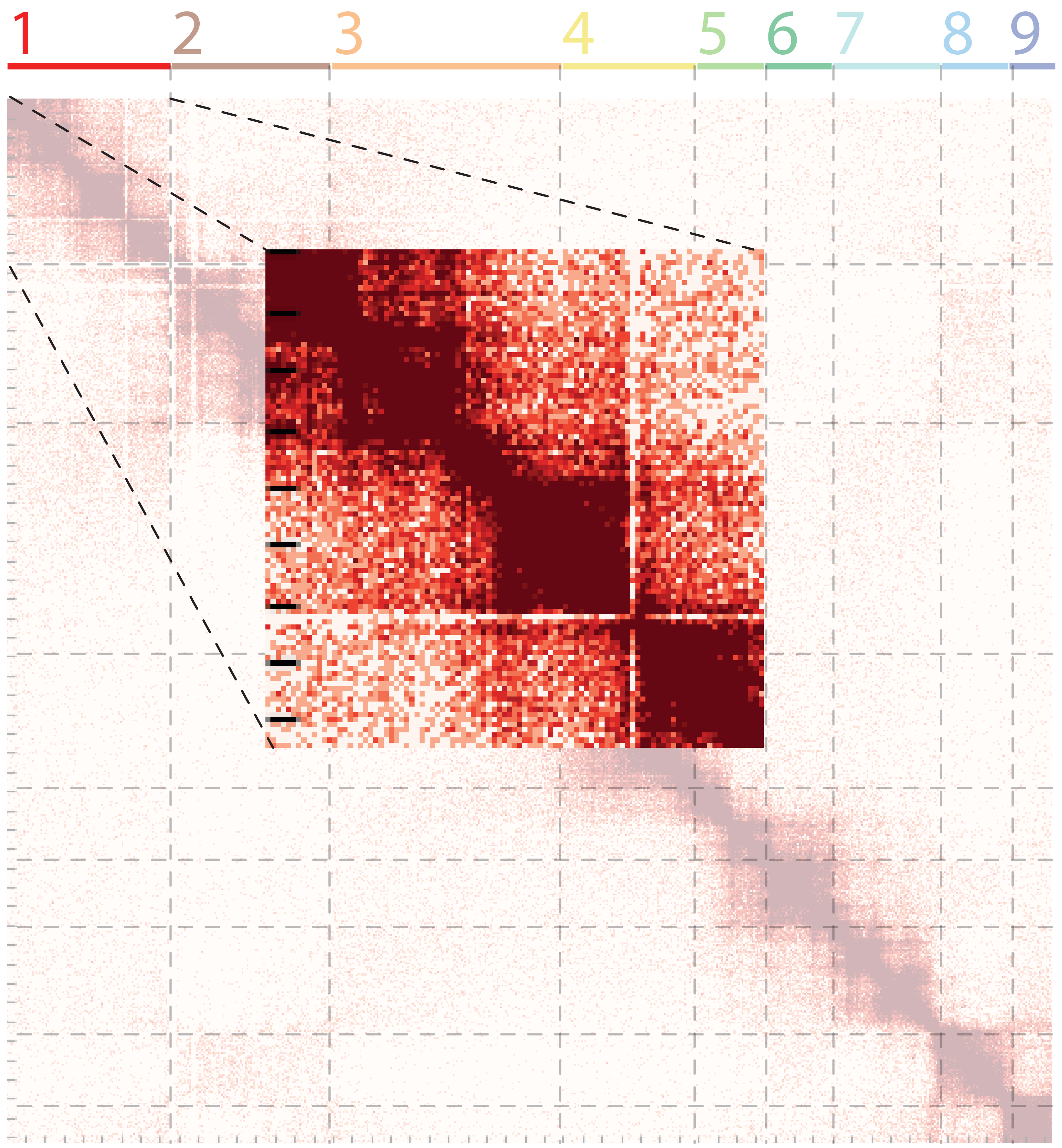




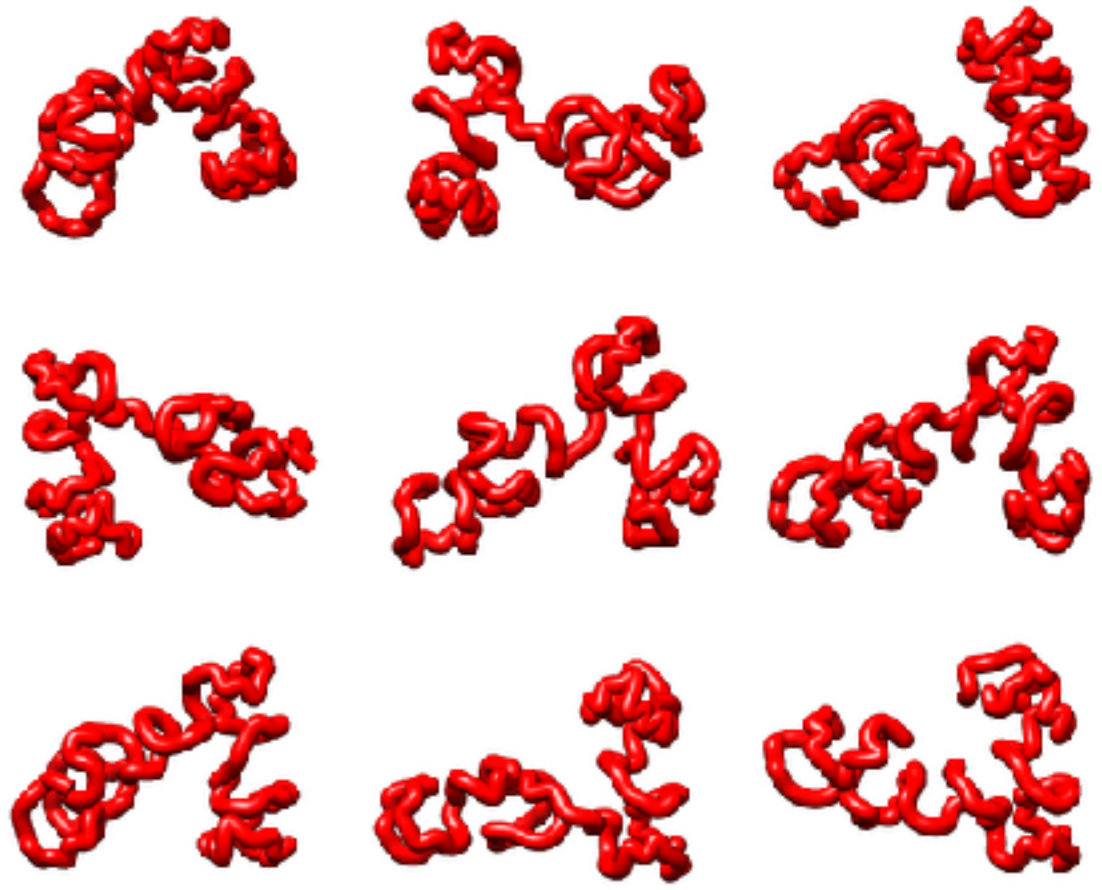


Increasing resolution

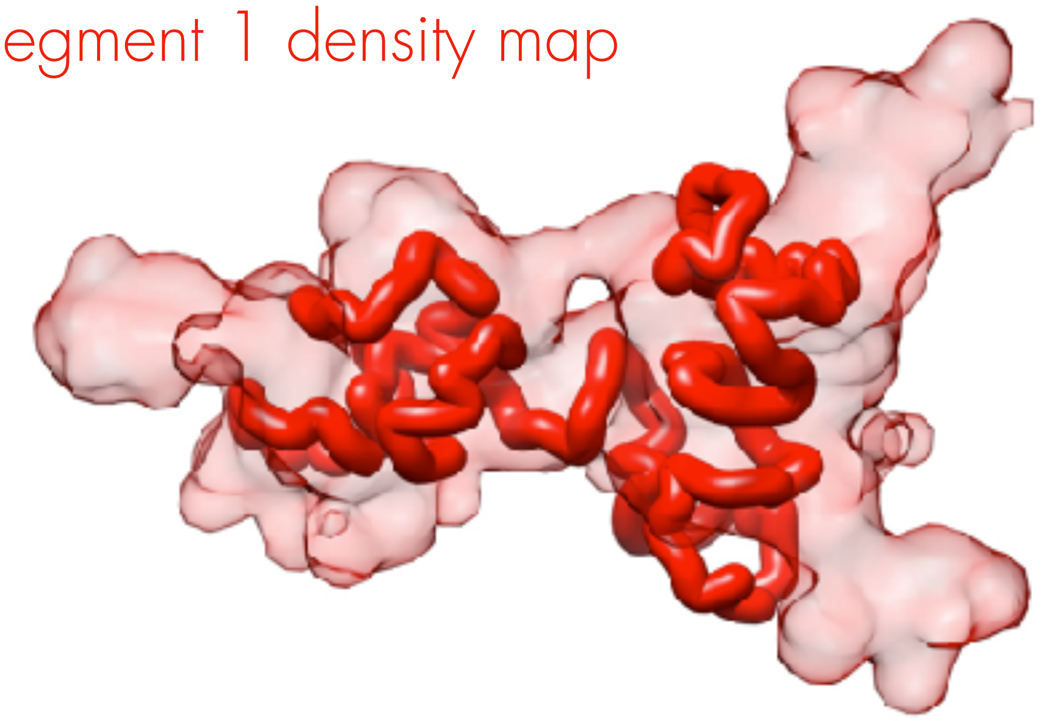
Rigid body fitting 3D structures based on Hi-C data



Segment 1 3D models

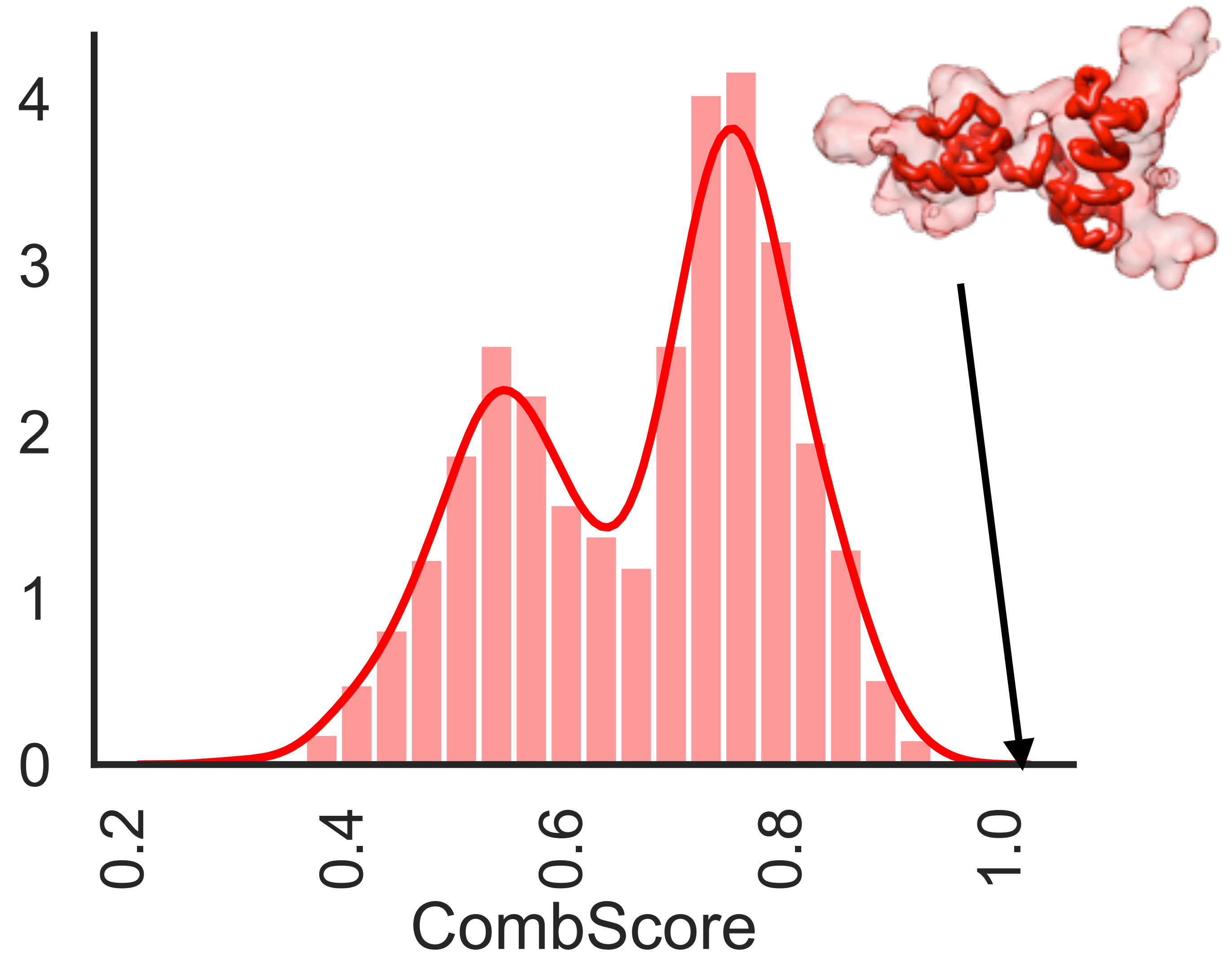


Segment 1 density map



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

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



TADbit

Serra, Baù, et al. PLOS CB 2017

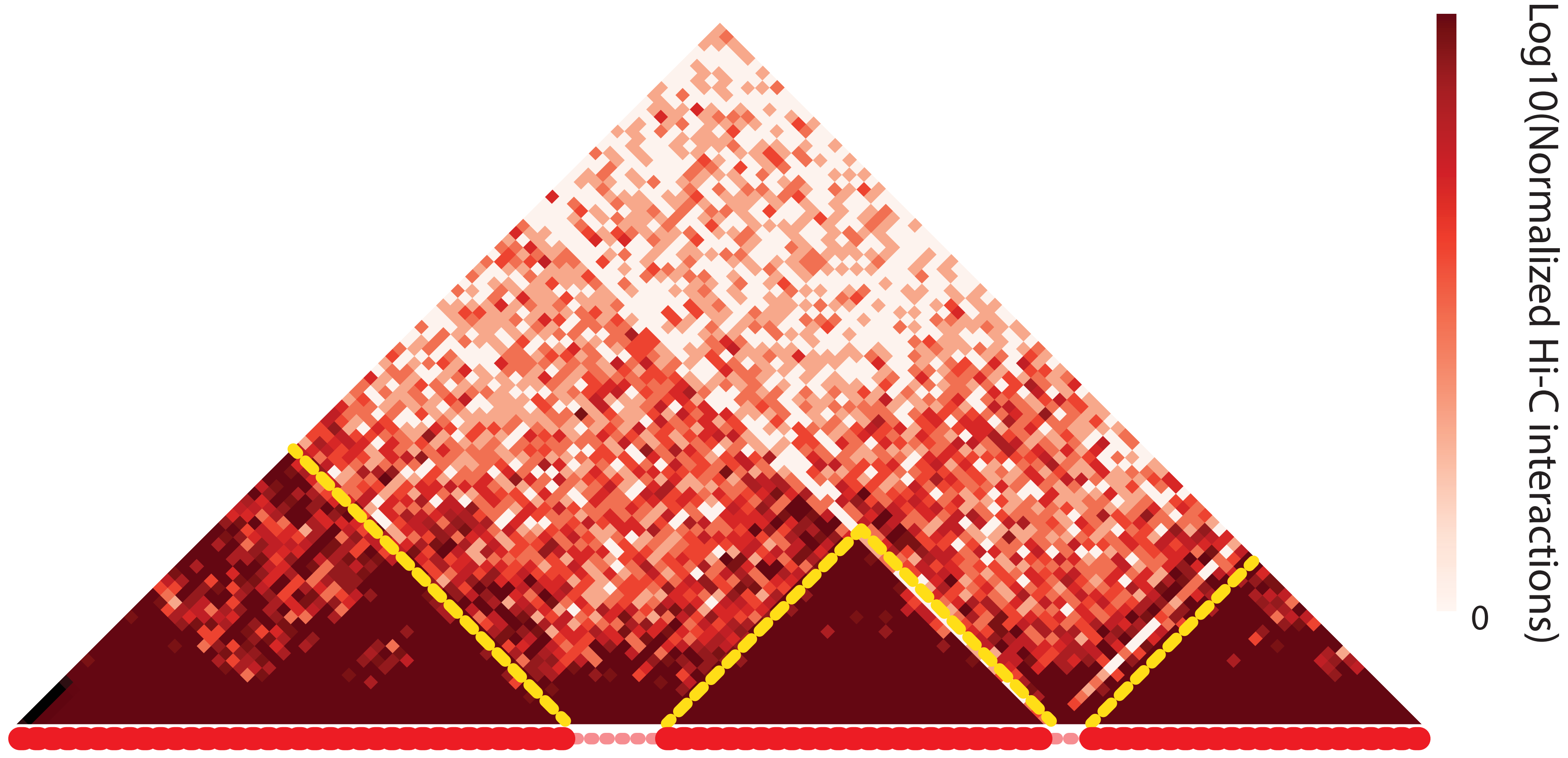
http://www.3DGenomes.org

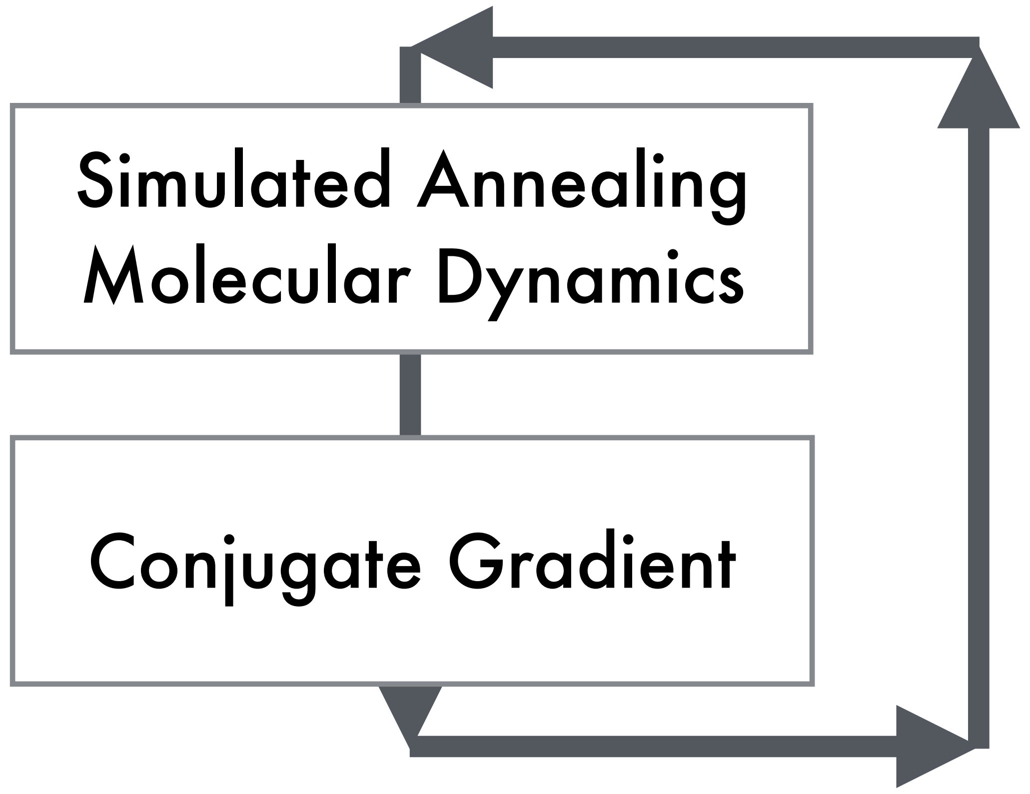
Farabella et al, J Appl Crystallogr. 2015

Roseman, 2000; Wriggers & Chacon, Structure 2001

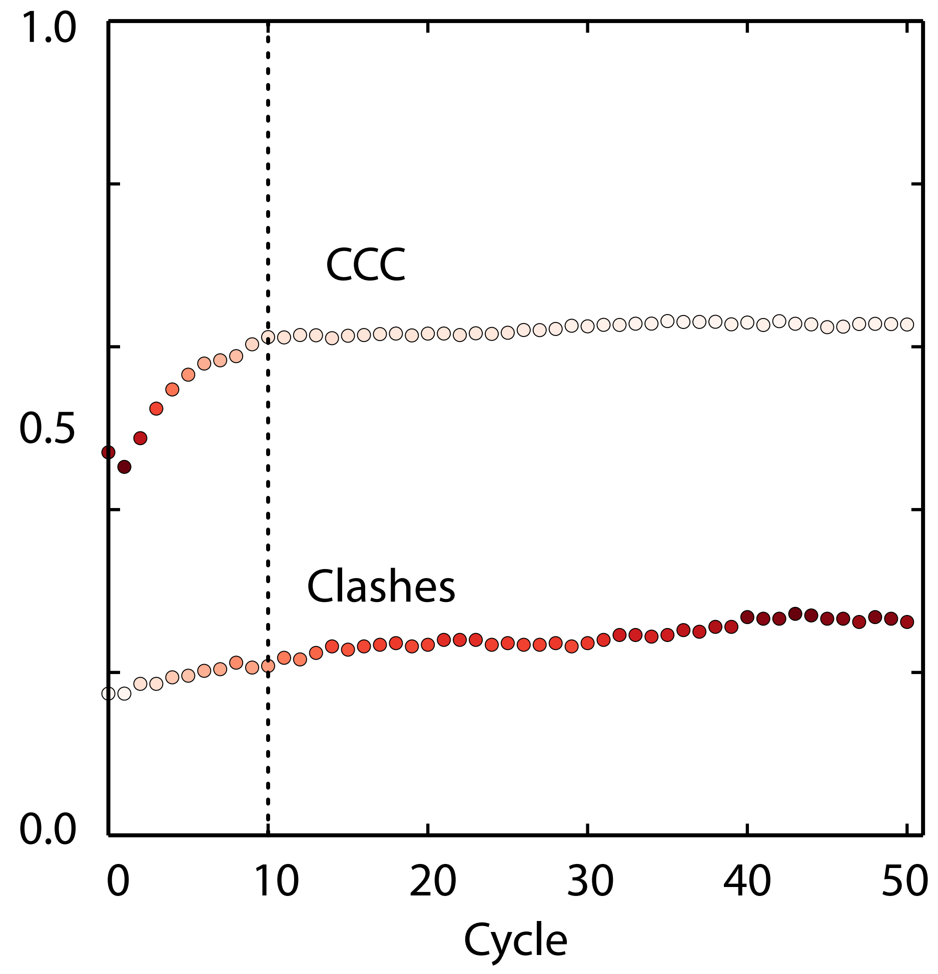
Increasing resolution

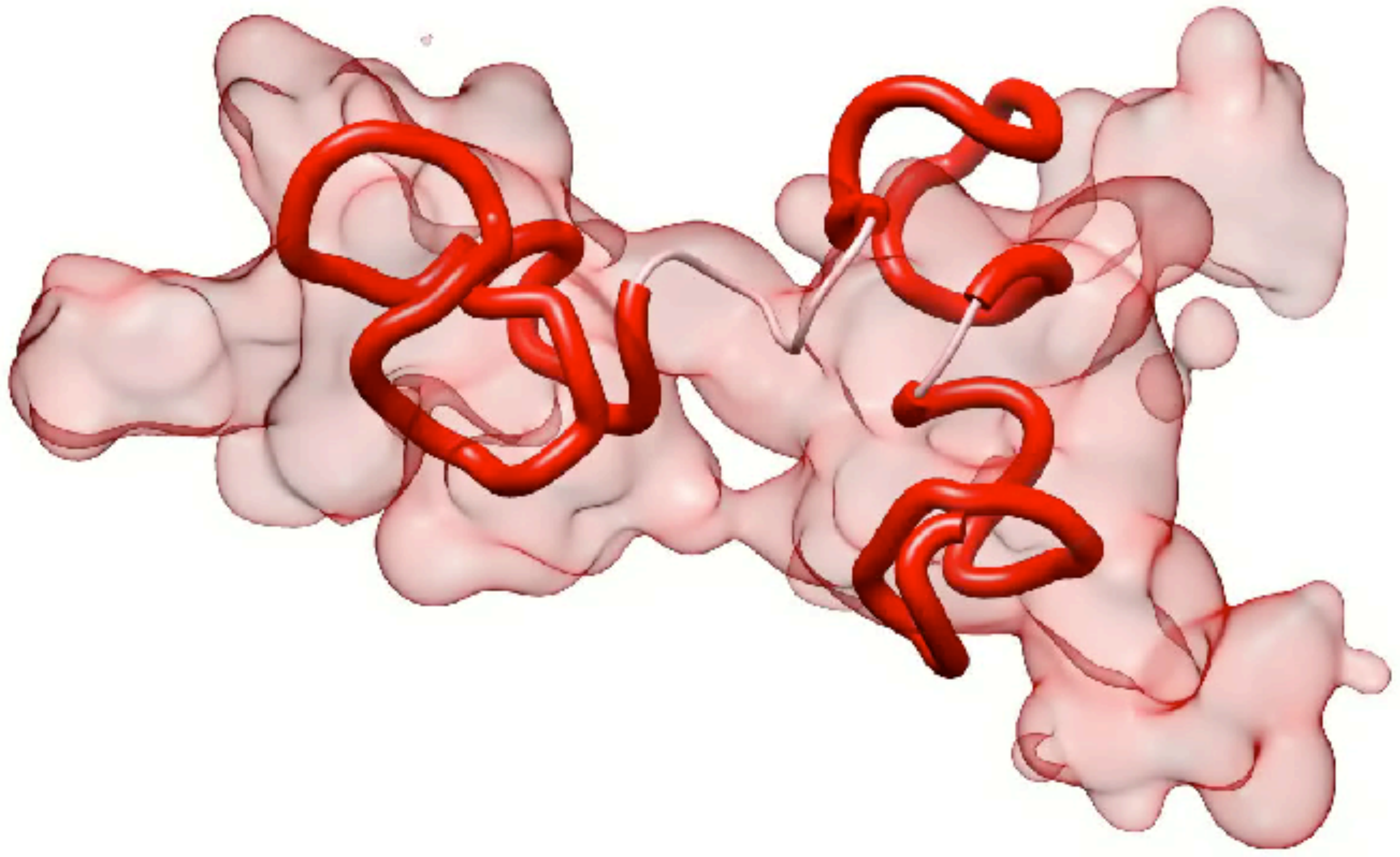
Flexible fitting 3D structures based on Hi-C data



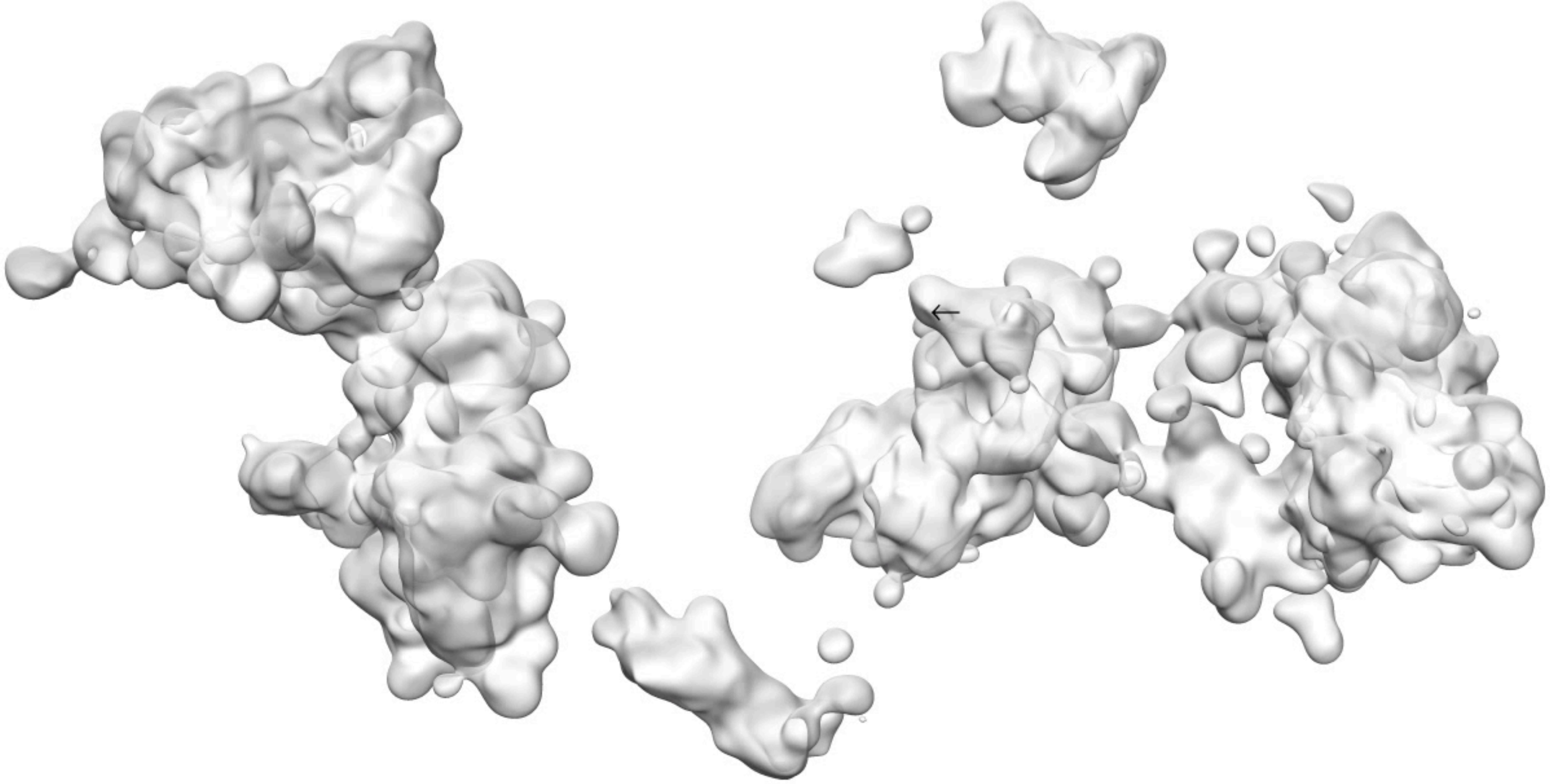

Simulated Annealing
Molecular Dynamics

Conjugate Gradient





Chromosome walking path @10Kb resolution



What next?

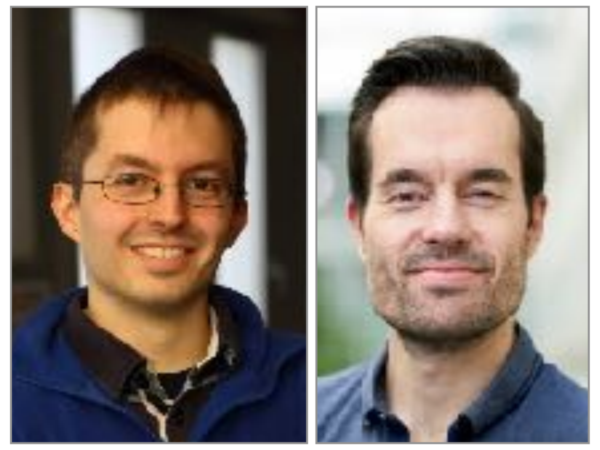
Is there a dynamic coupling between structure and gene activity?

Is genome structure more conserved than sequence?

The End!

Photo by David Oliete - www.davidoliete.com

Dynamics of gene activation

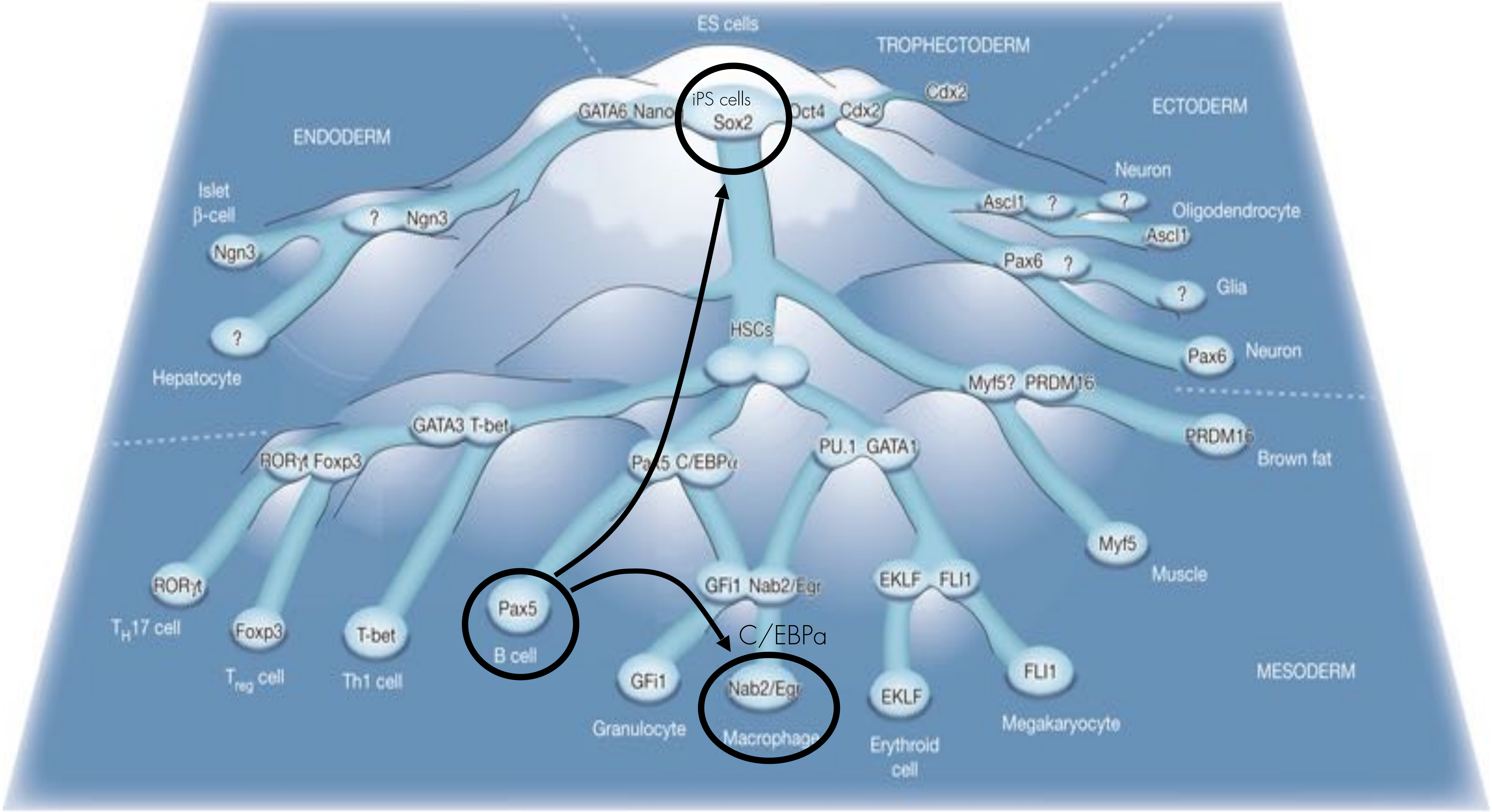


Marco di Stefano
Ralph Stadhouders
with Graf Lab (CRG, Barcelona)

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

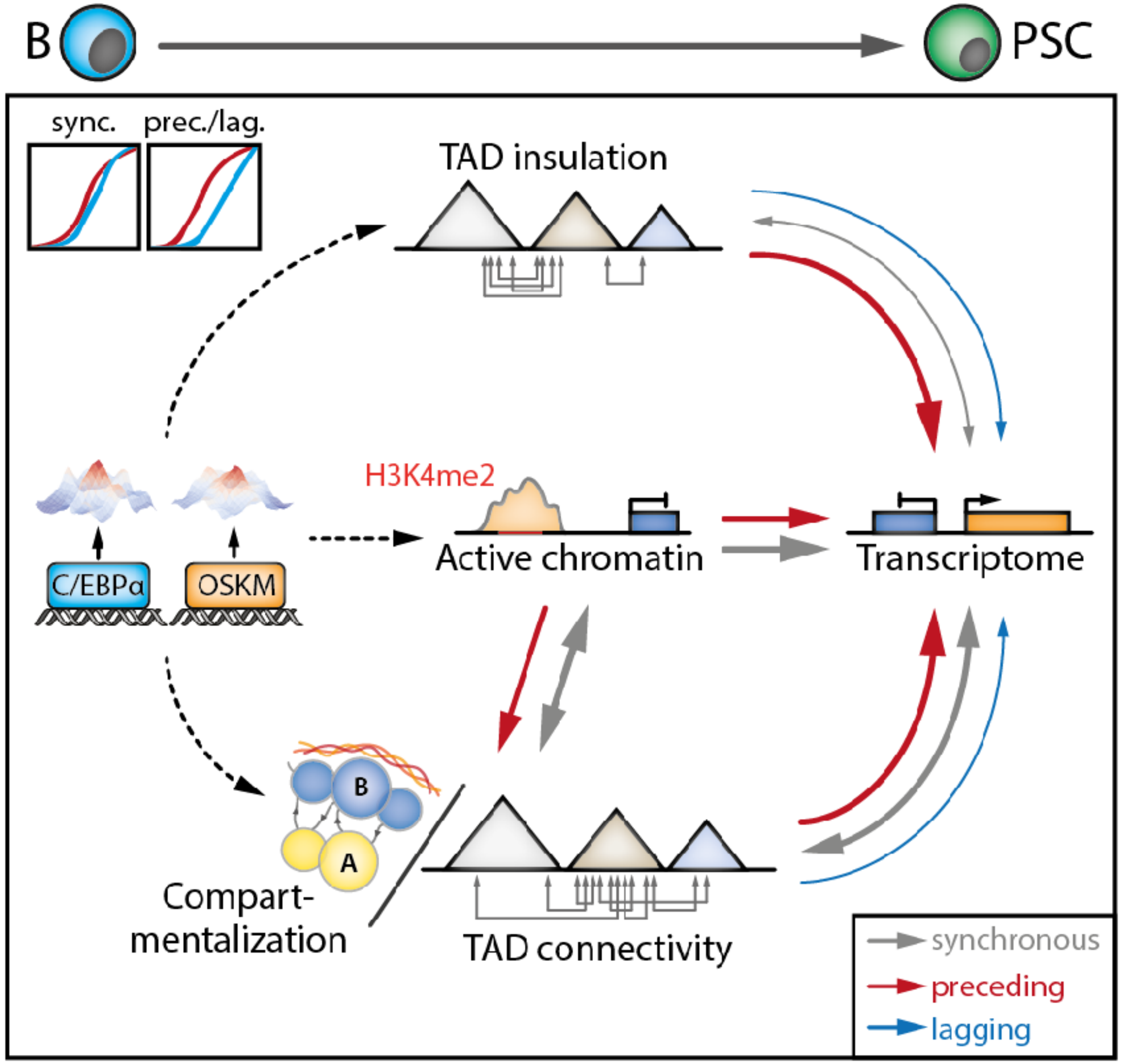
Normal 'forward' differentiation

Cell fates can be converted by enforced TF expression

Transdifferentiation or reprogramming

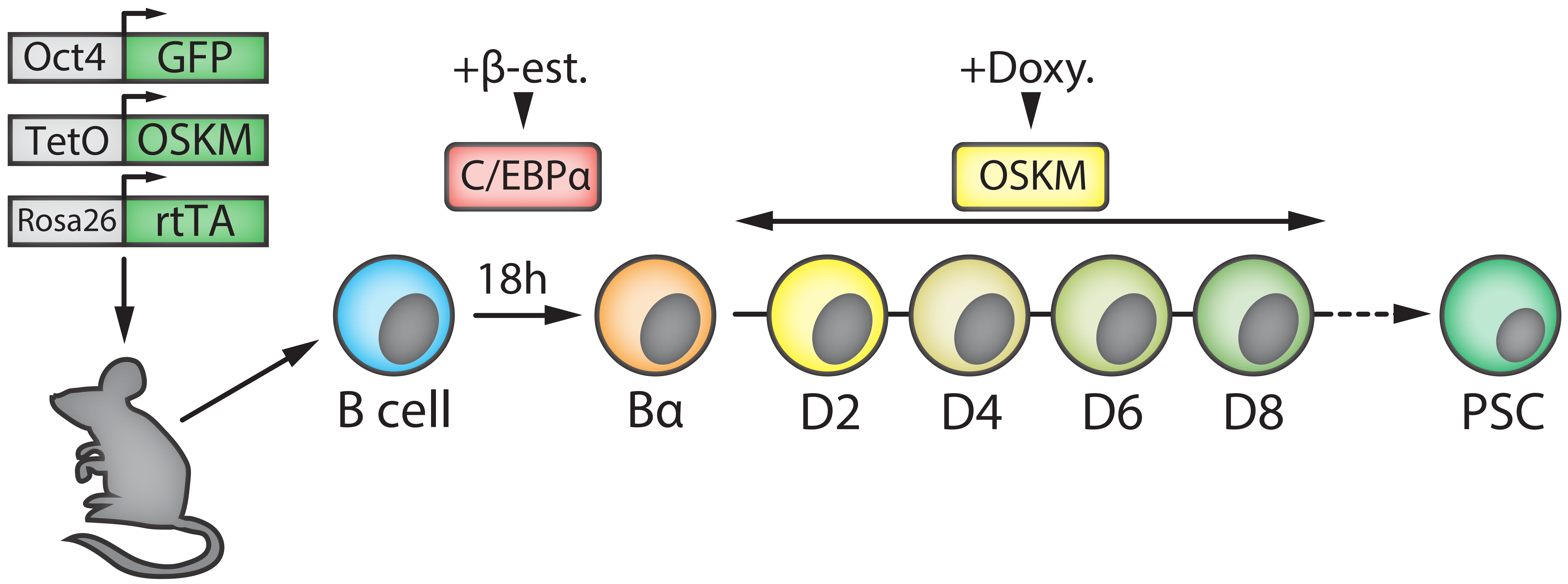
Interplay: topology, gene expression & chromatin

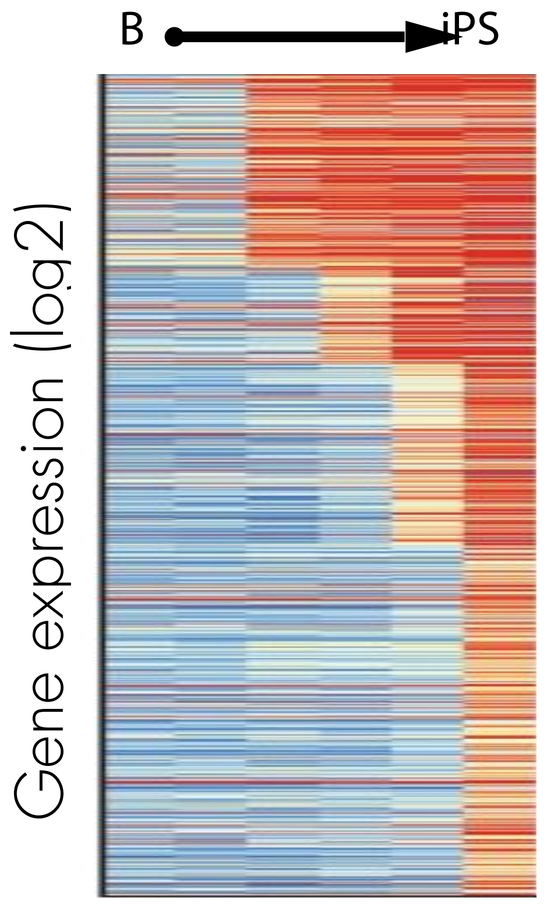
Stadhouders, R., Vidal, E. et al. (2018) Nature Genetics

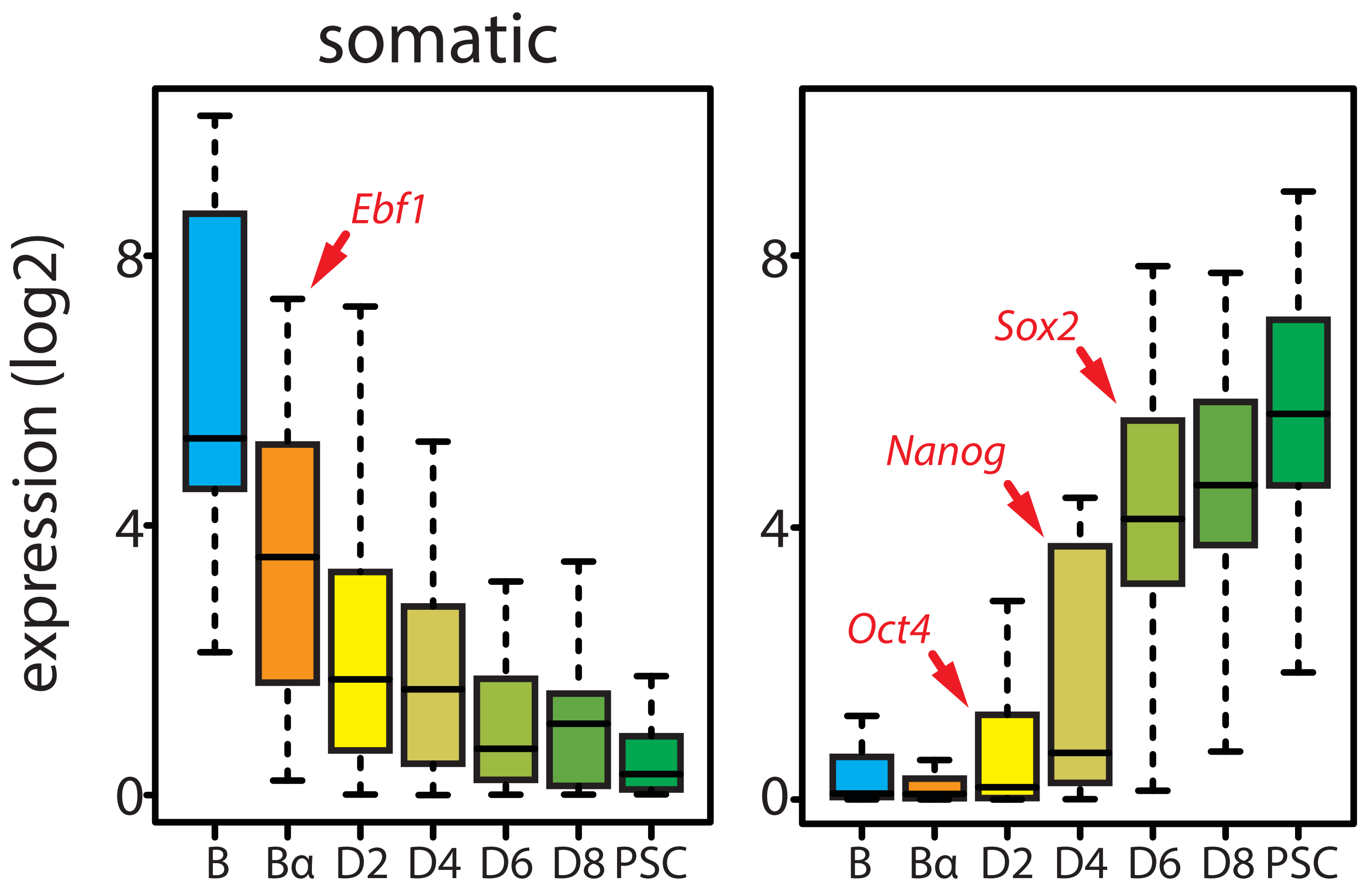


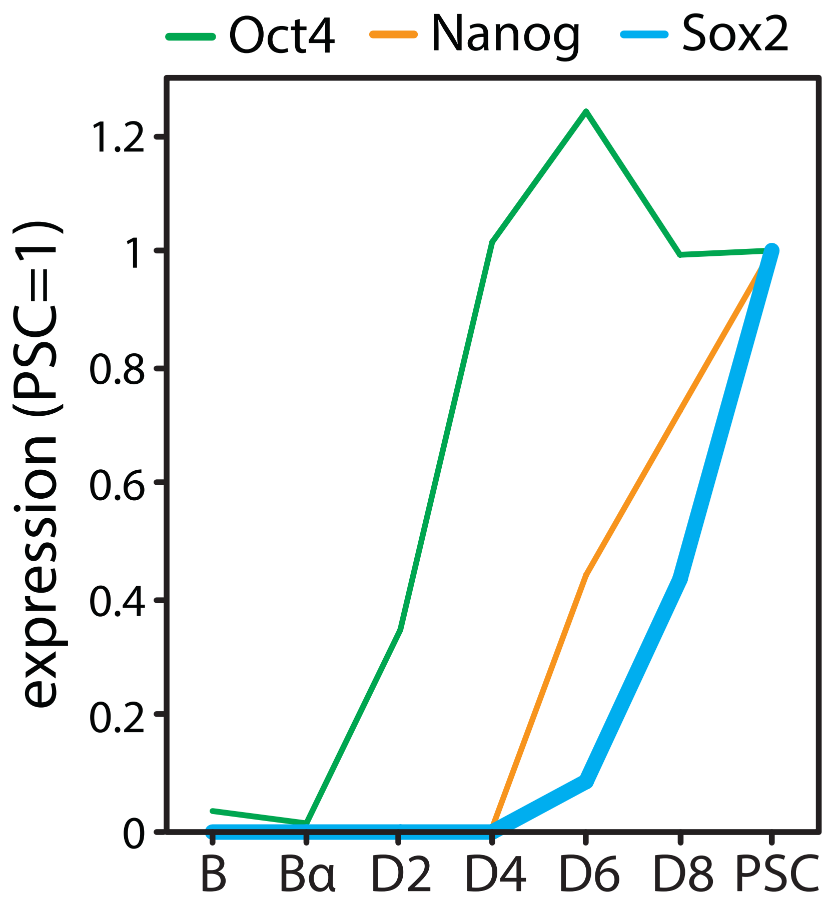
Reprogramming from B to PSC

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



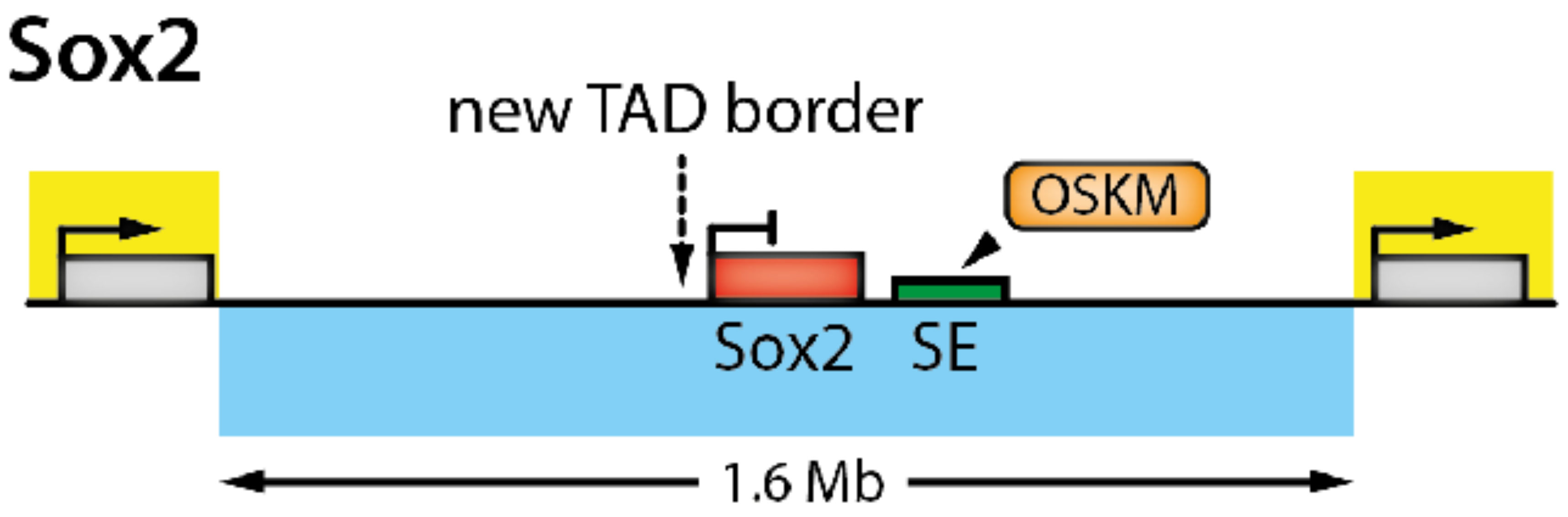


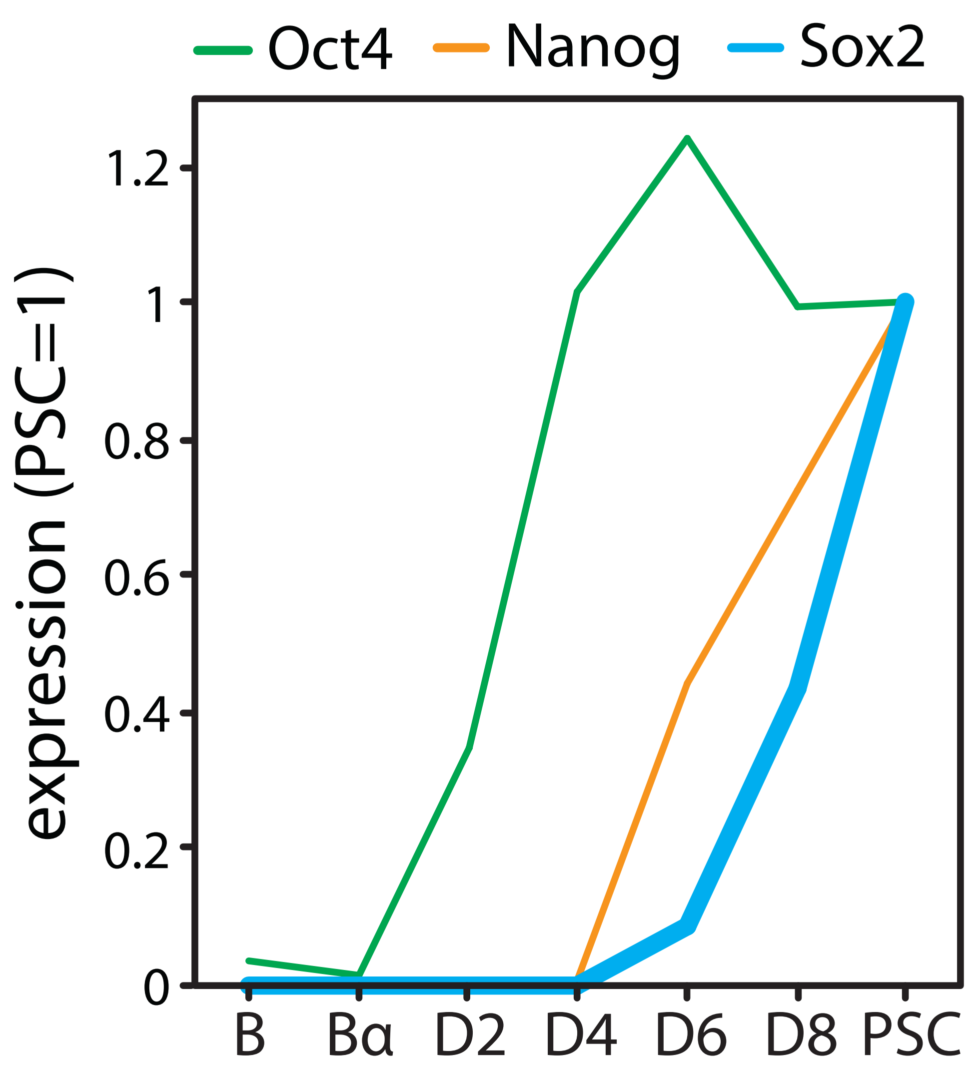


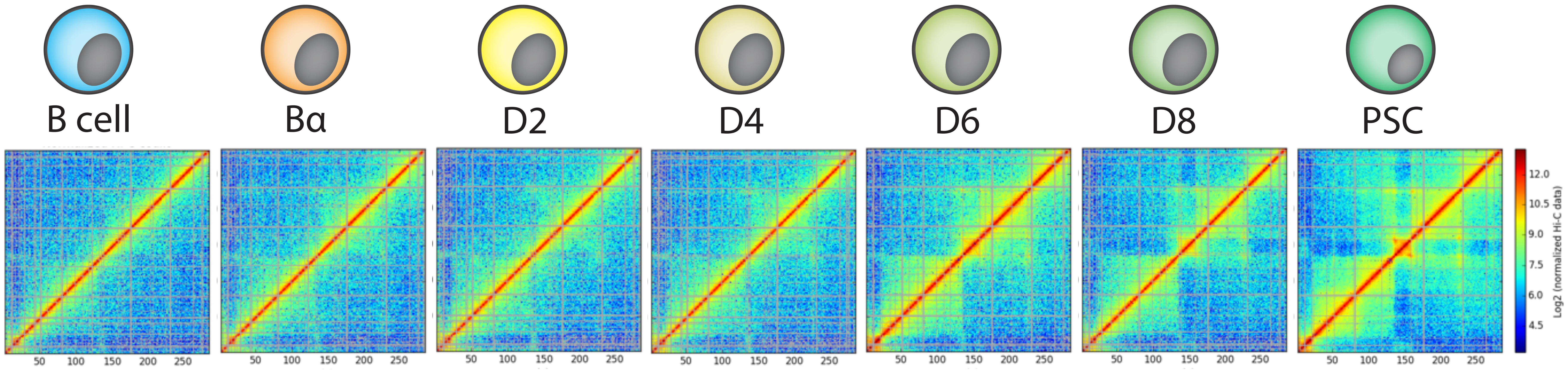


Hi-C maps of reprogramming from B to PSC

The SOX2 locus

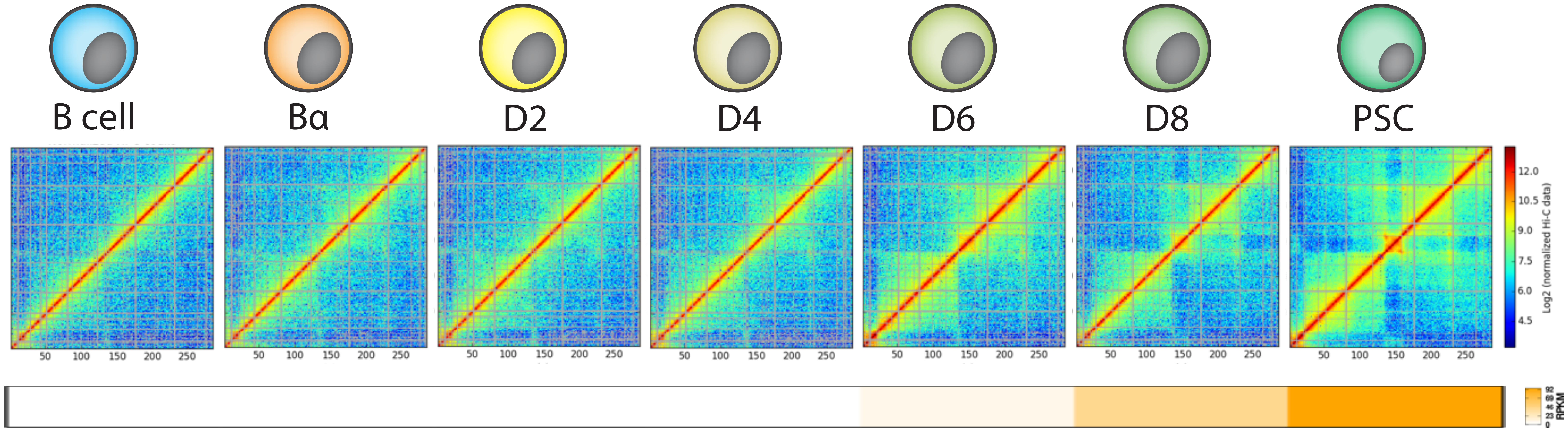






Hi-C maps of reprogramming from B to PSC

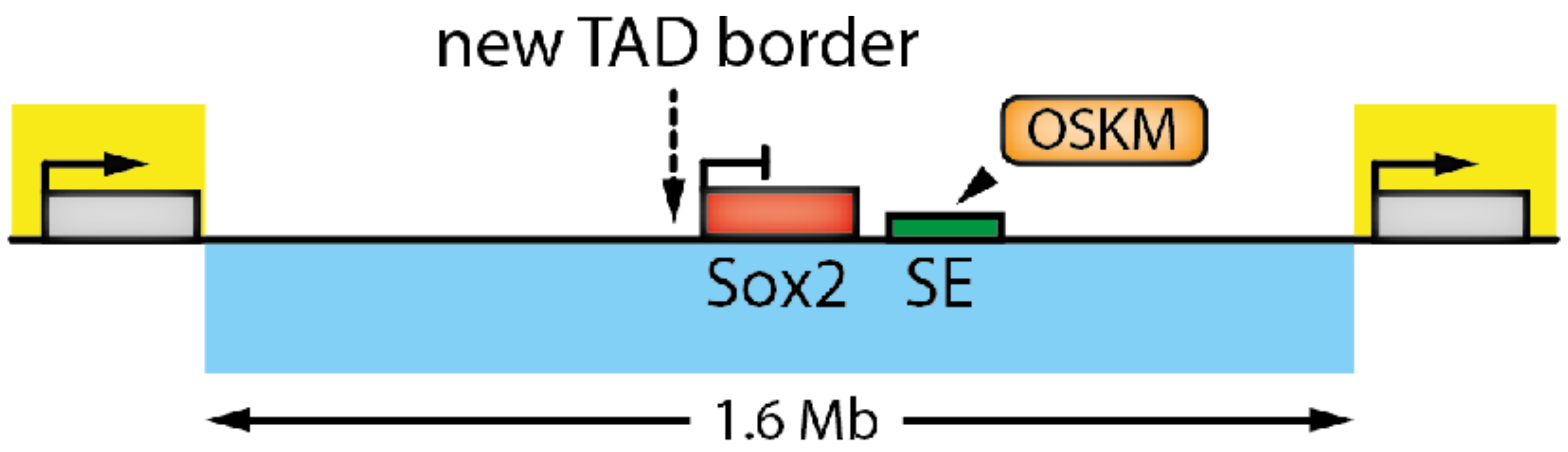
The SOX2 locus

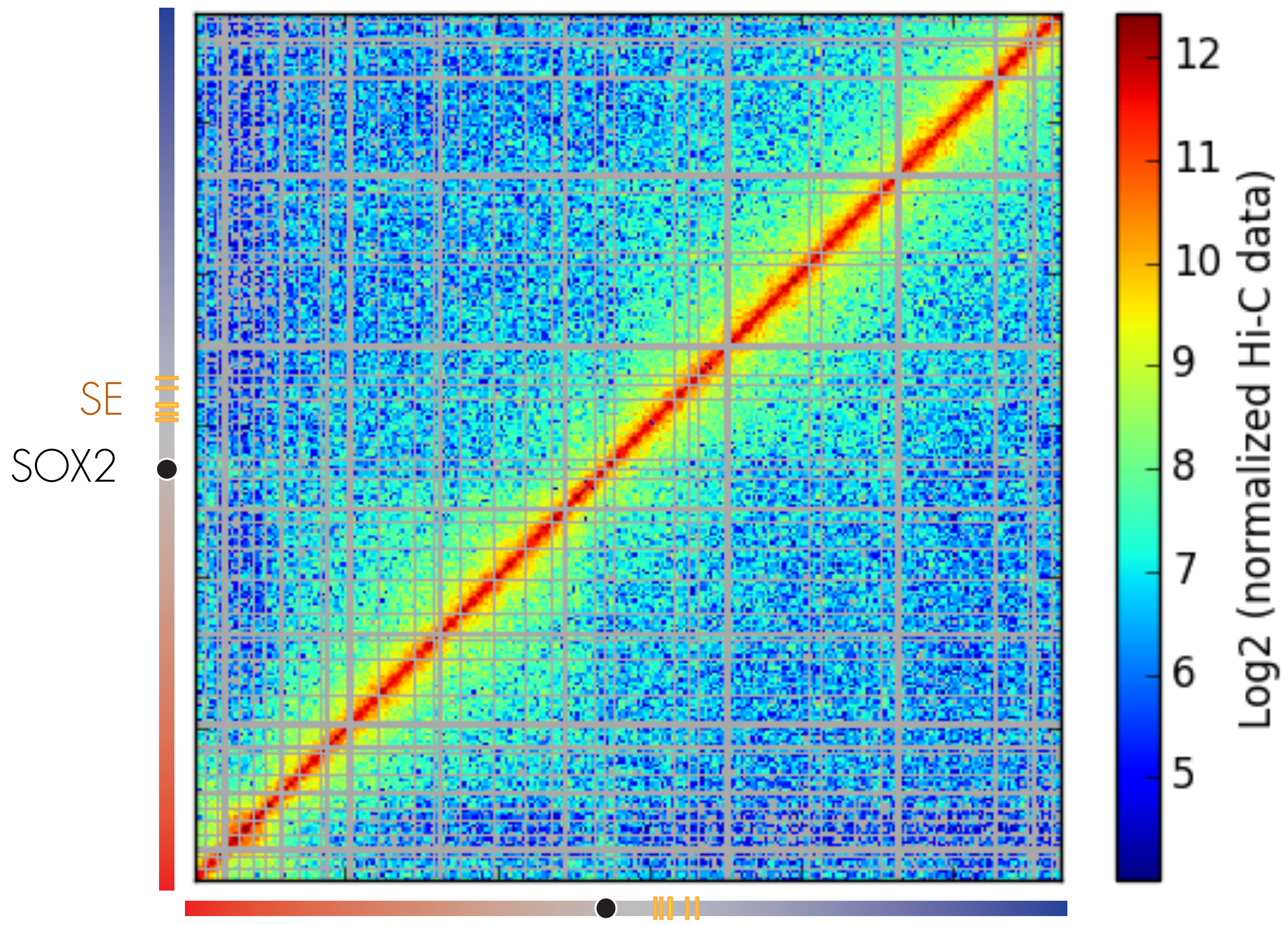


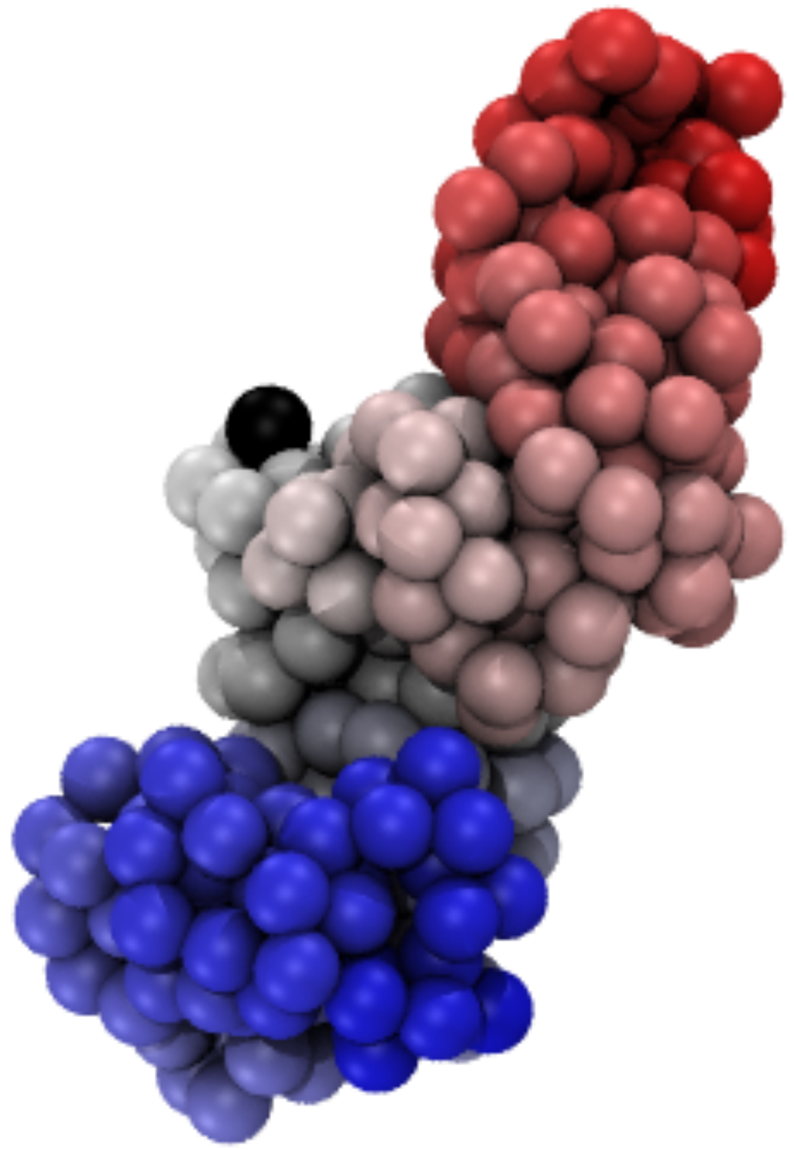
How does these structural rearrangements interplay with the transcription activity?

What are the main drivers of structural transitions?

TADbit modeling of SOX2 from B cells Hi-C





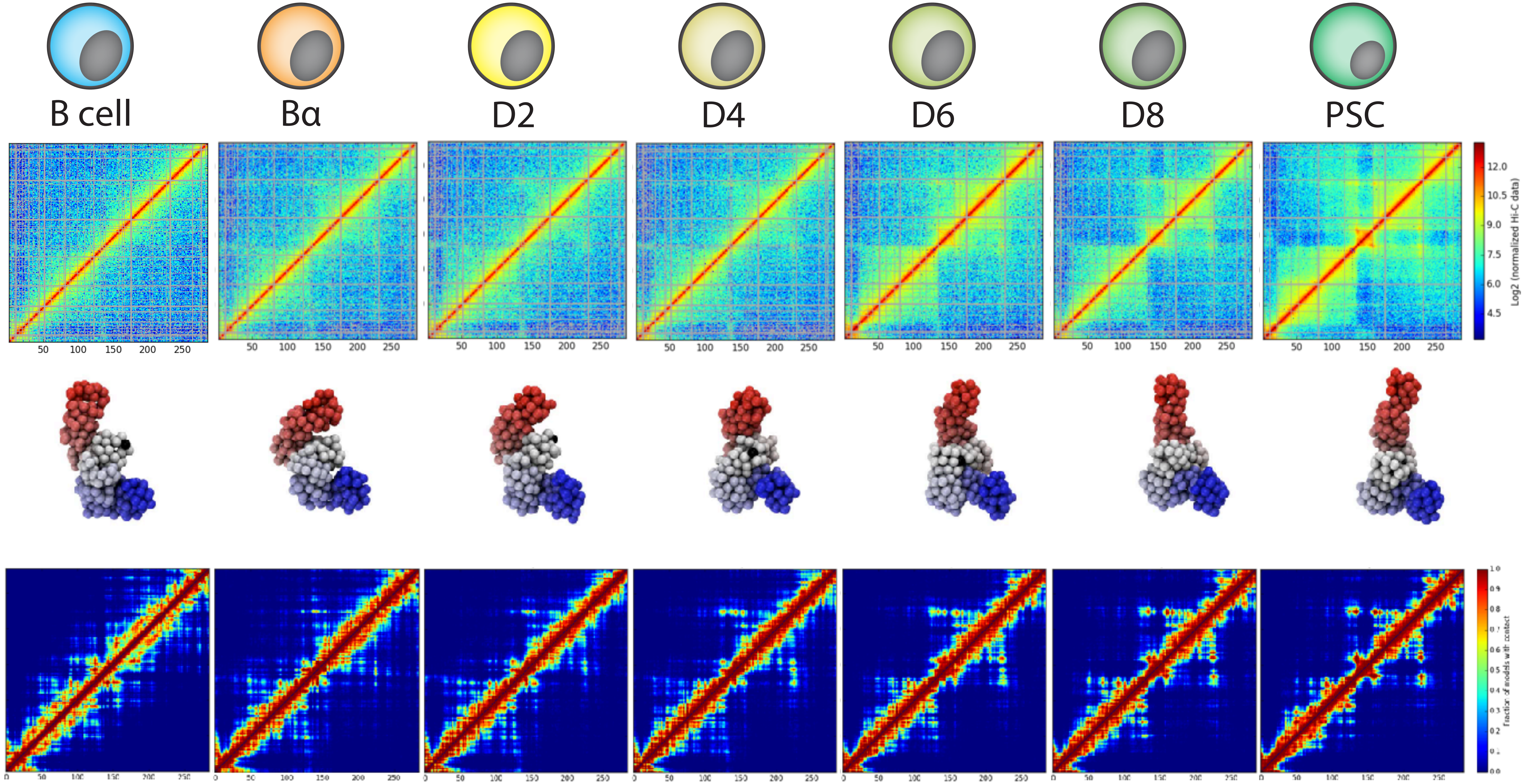


Optimal IMP parameters

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

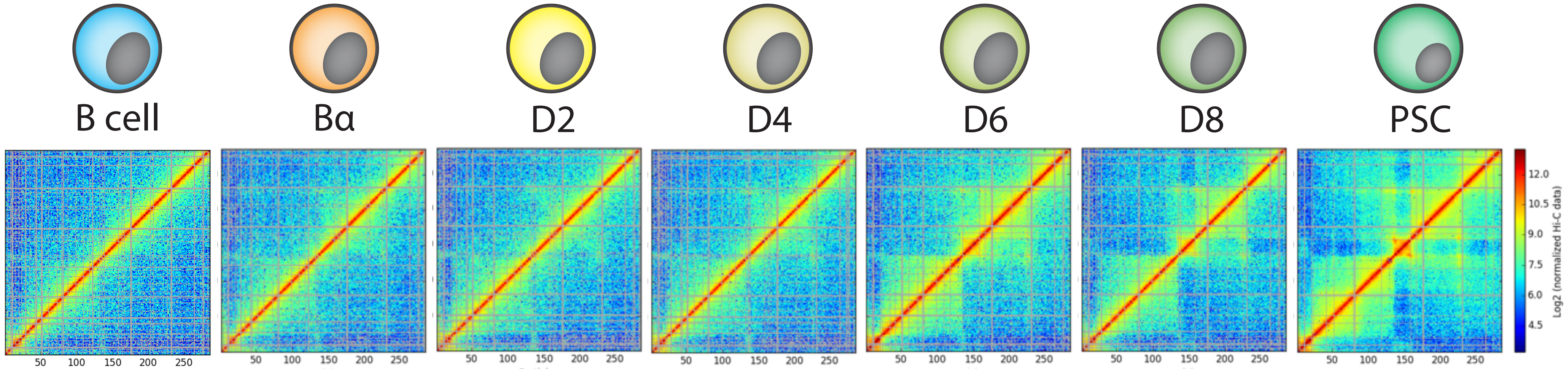
Models of reprogramming from B to PSC

The SOX2 locus



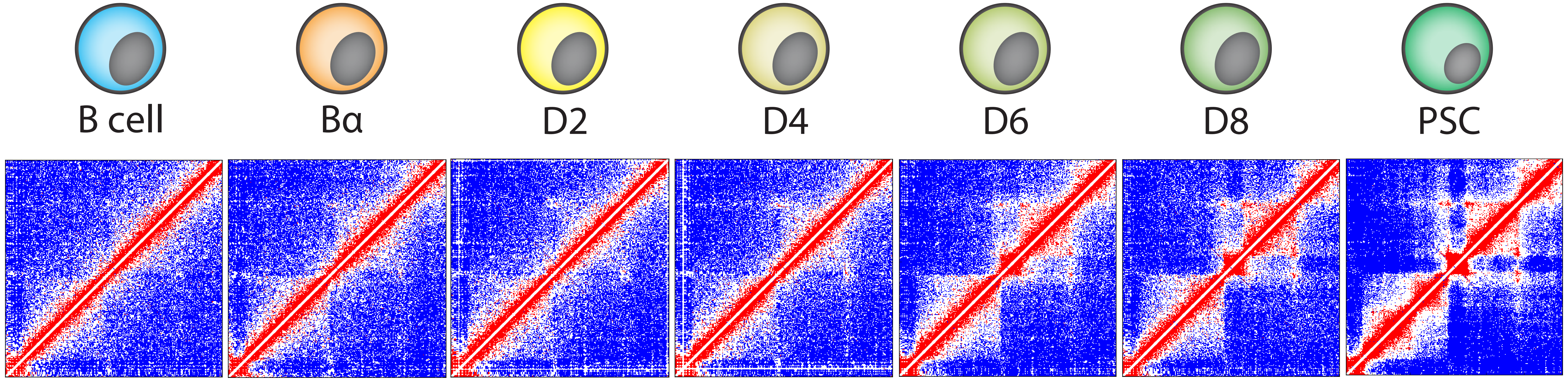
TADdyn: from time-series Hi-C maps to dynamic restraints

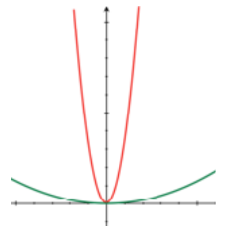
The SOX2 locus



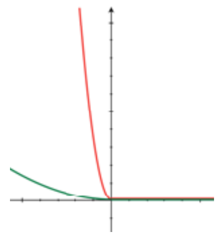
TADdyn: from time-series Hi-C maps to dynamic restraints

The SOX2 locus





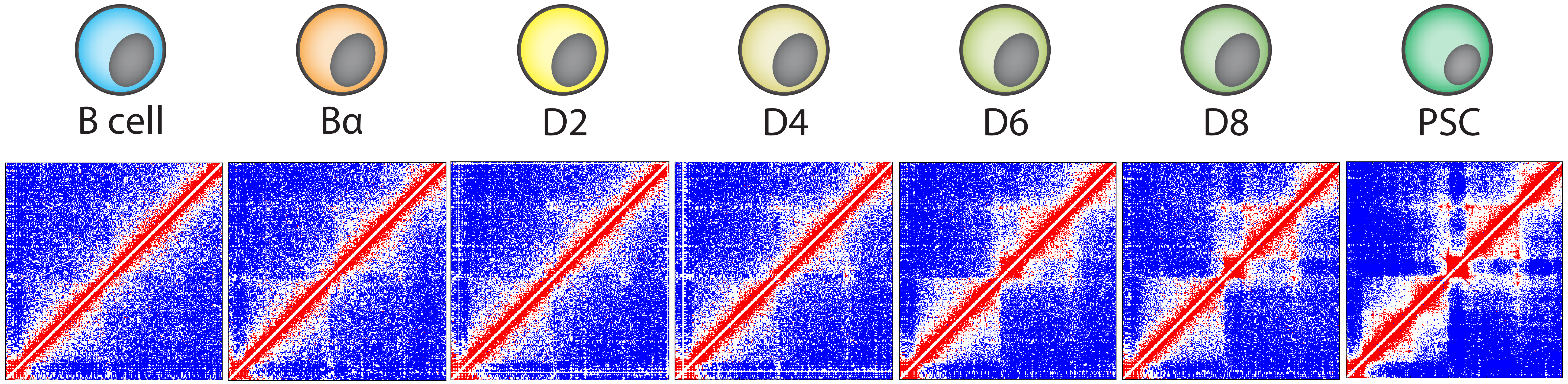
🟥 Harmonic

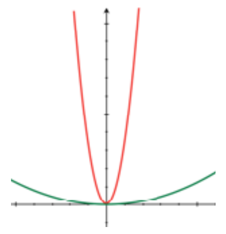


🟦 HarmonicLowerBound

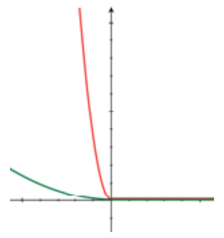
TADdyn: from time-series Hi-C maps to dynamic restraints

The SOX2 locus

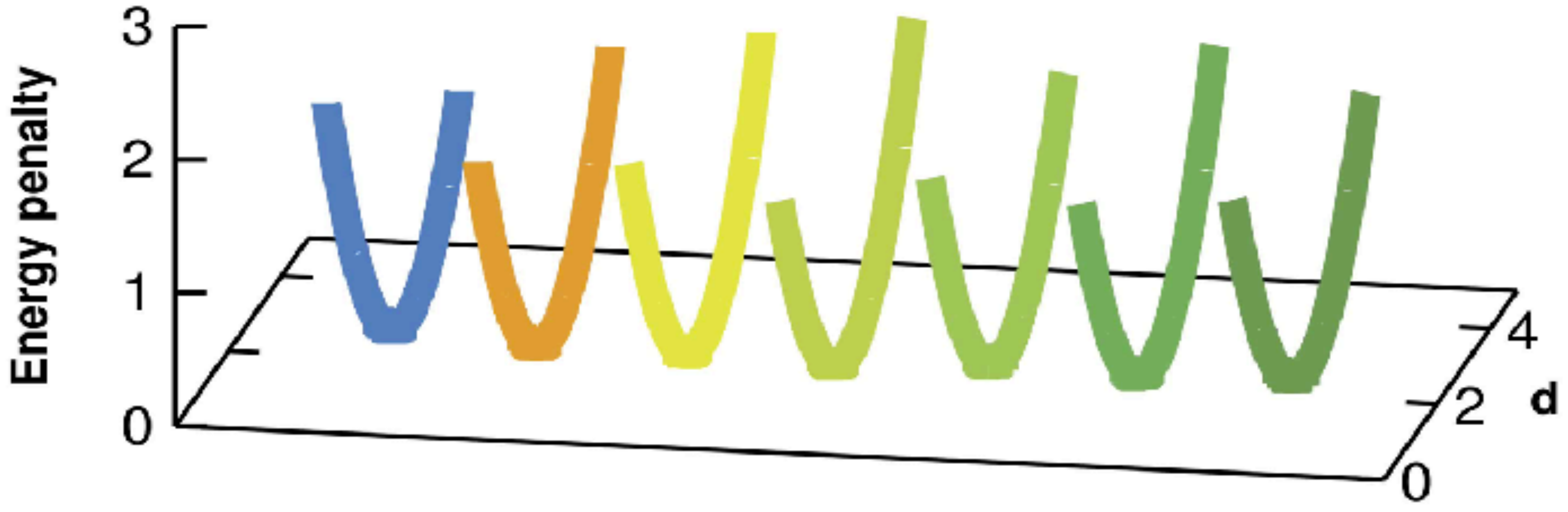




🟥 Harmonic



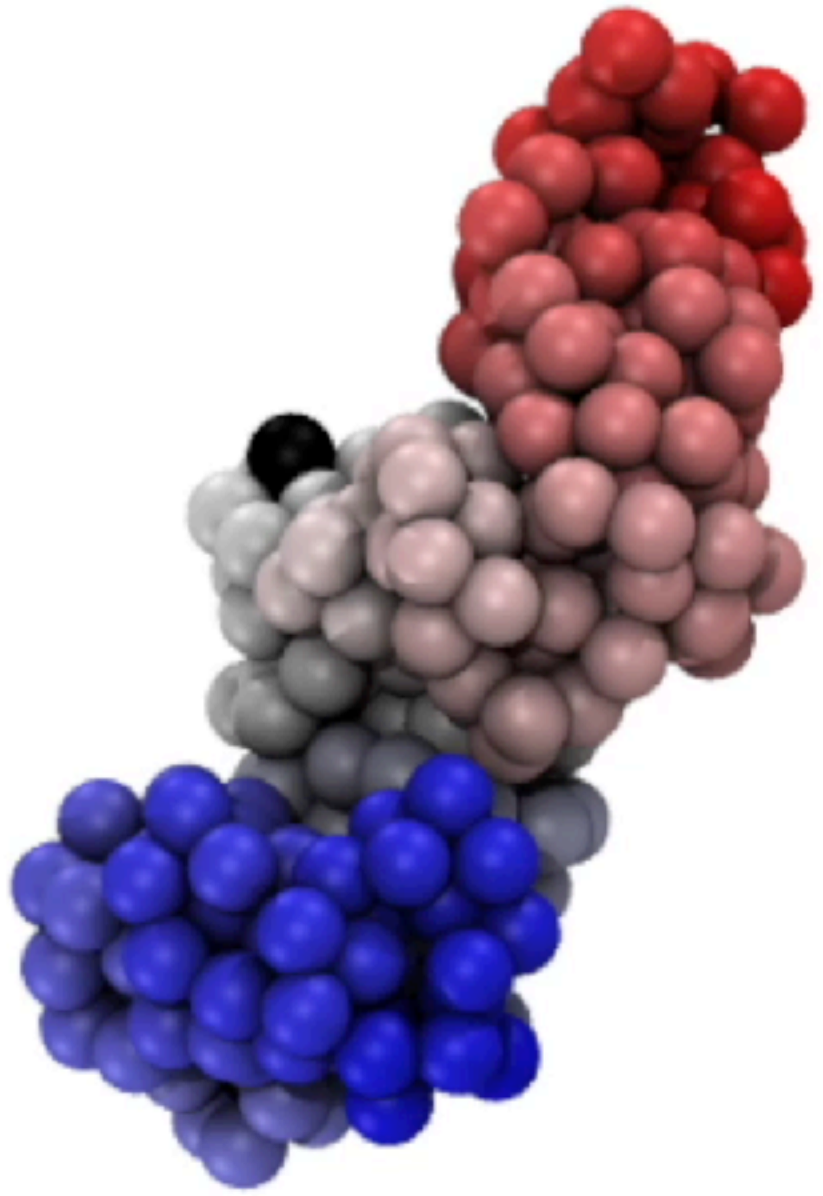
🟦 HarmonicLowerBound

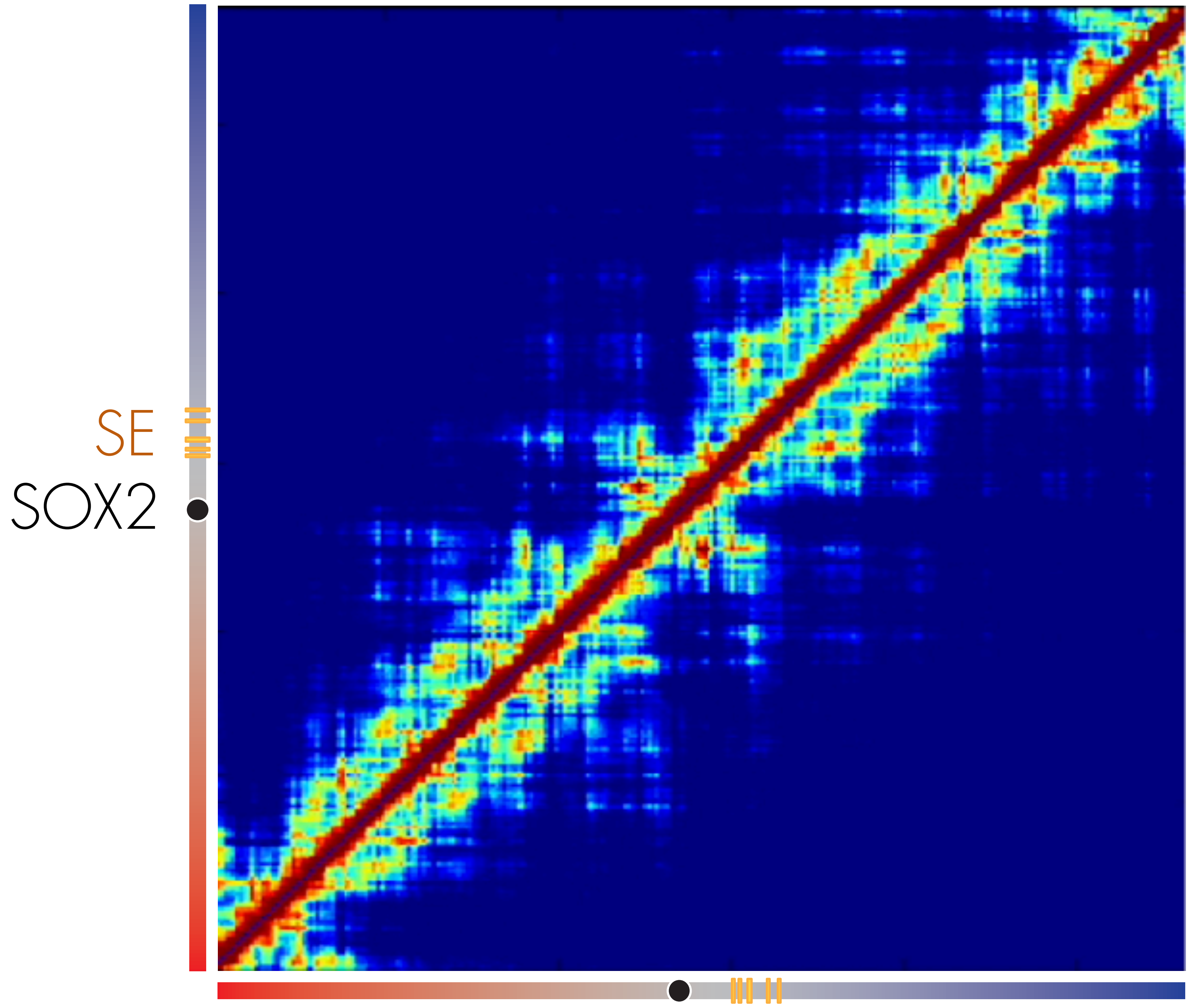


Transition	Stable	Vanishing	Raising
B -> Bα	18,612	6,984	7,290
Bα -> D2	18,512	7,390	6,687
D2 -> D4	18,369	6,830	6,893
D4 -> D6	18,971	6,291	7,289
D6 -> D8	20,167	6,093	6,250
D8 -> ES	20,679	5,738	6,173

SOX2 locus structural changes from B to PSC

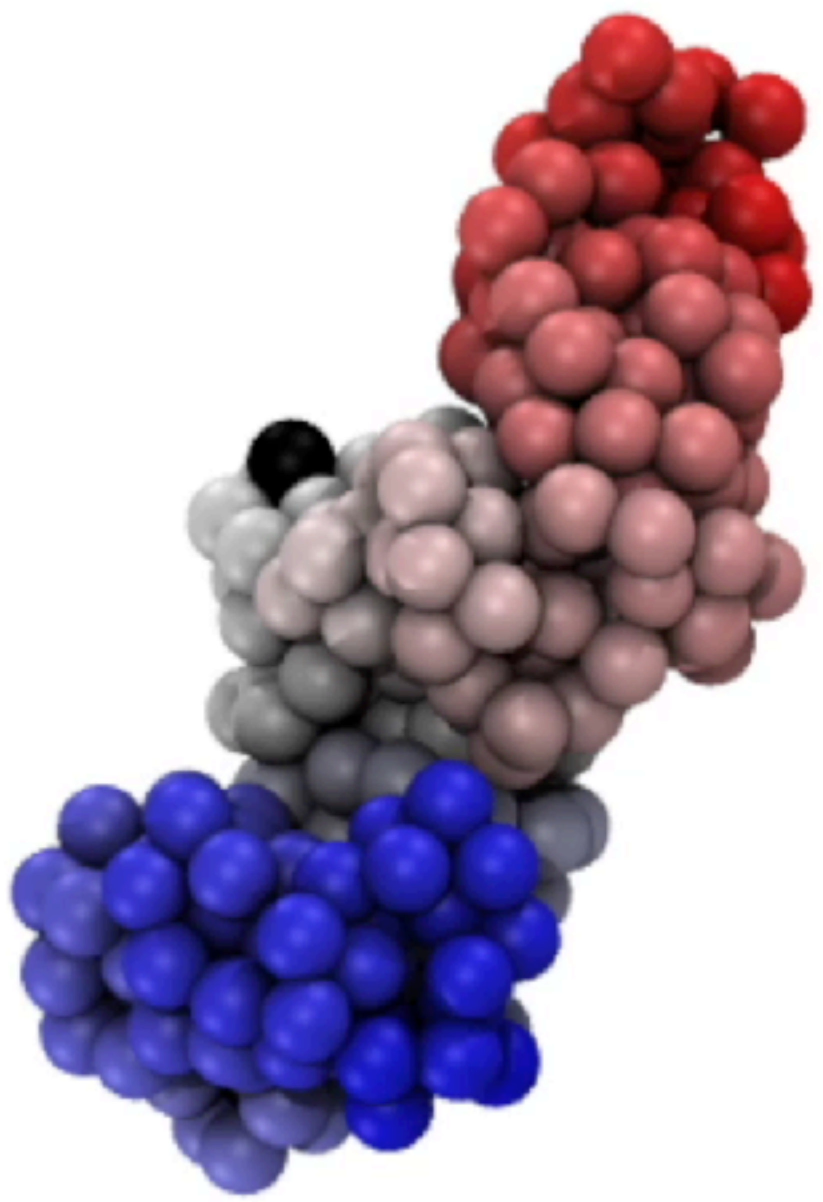
Contacts

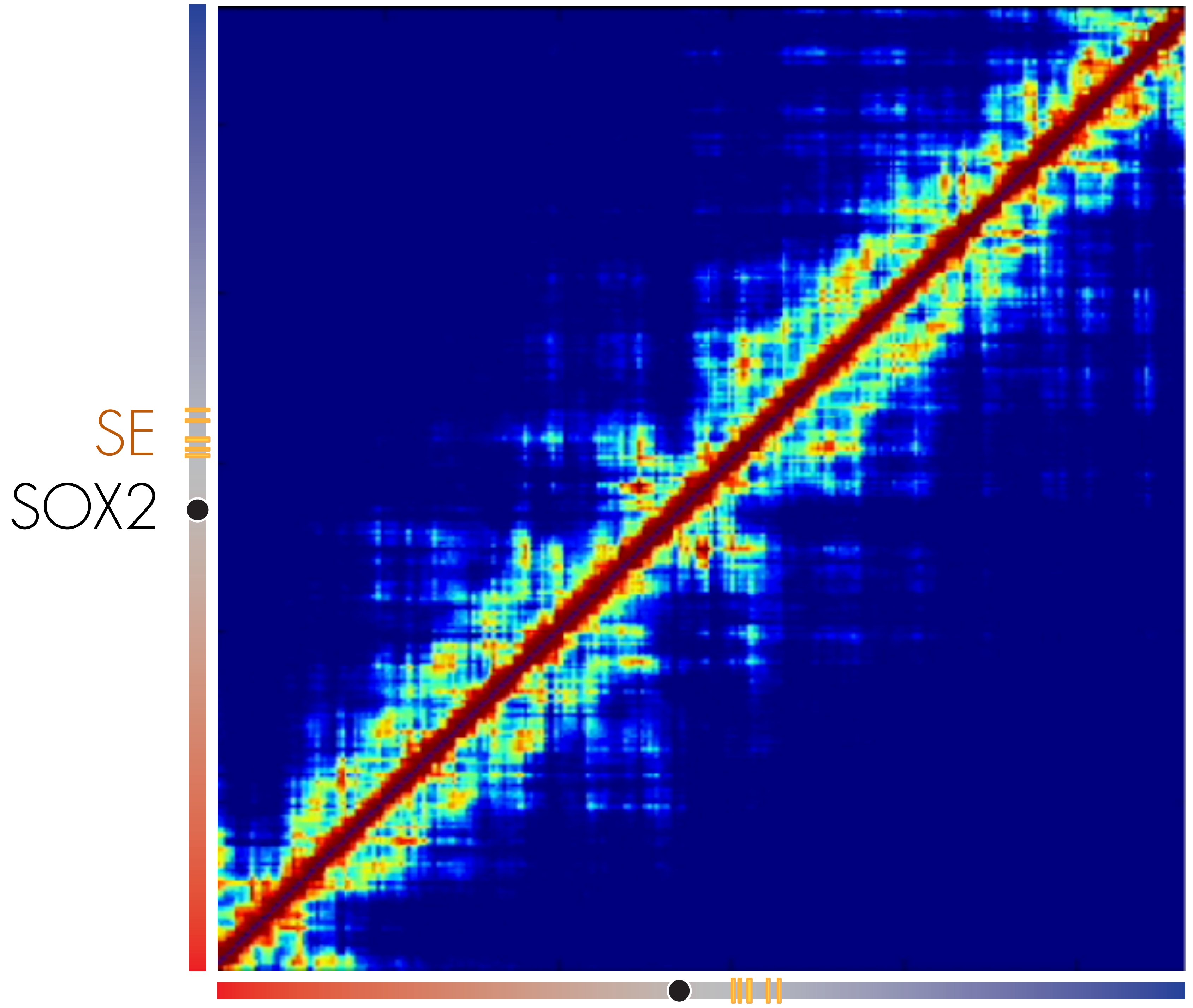




SOX2 locus structural changes from B to PSC

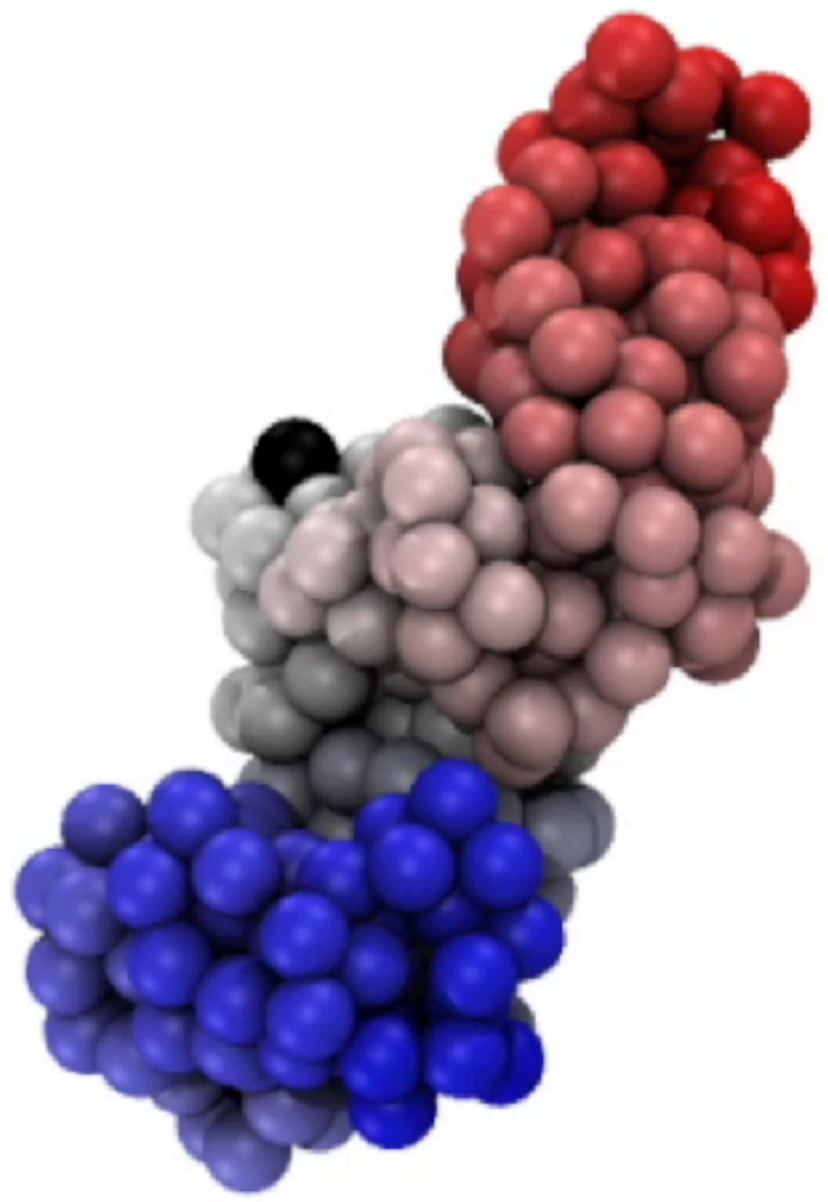
Contacts

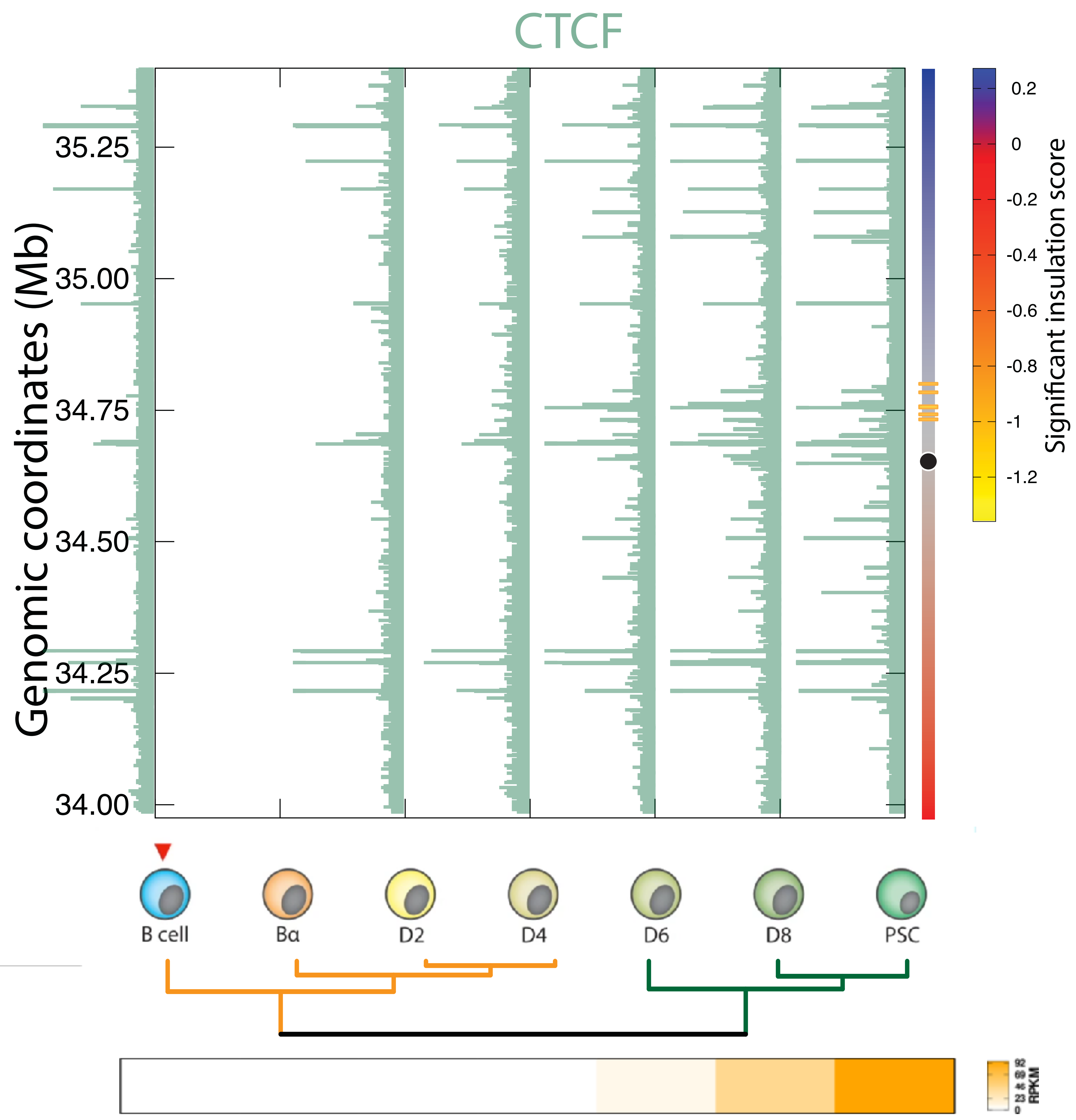




SOX2 locus structural changes from B to PSC

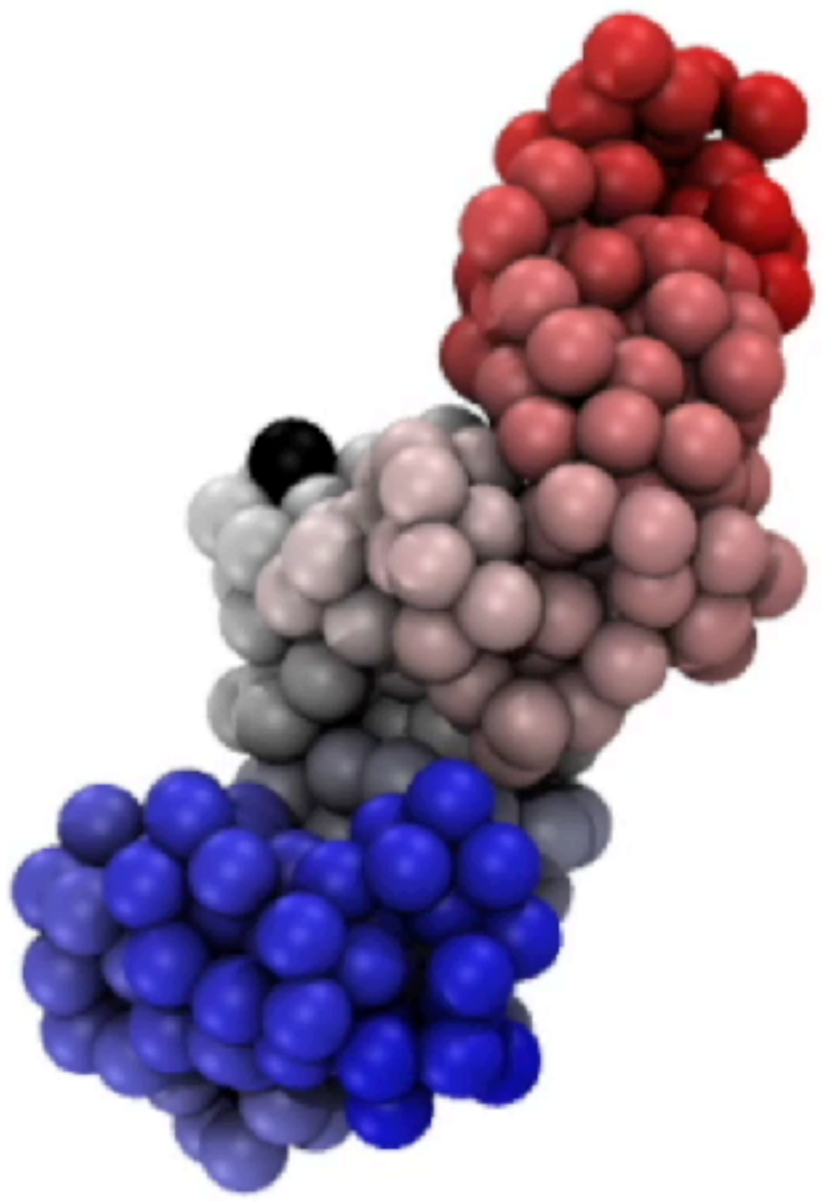
TAD borders

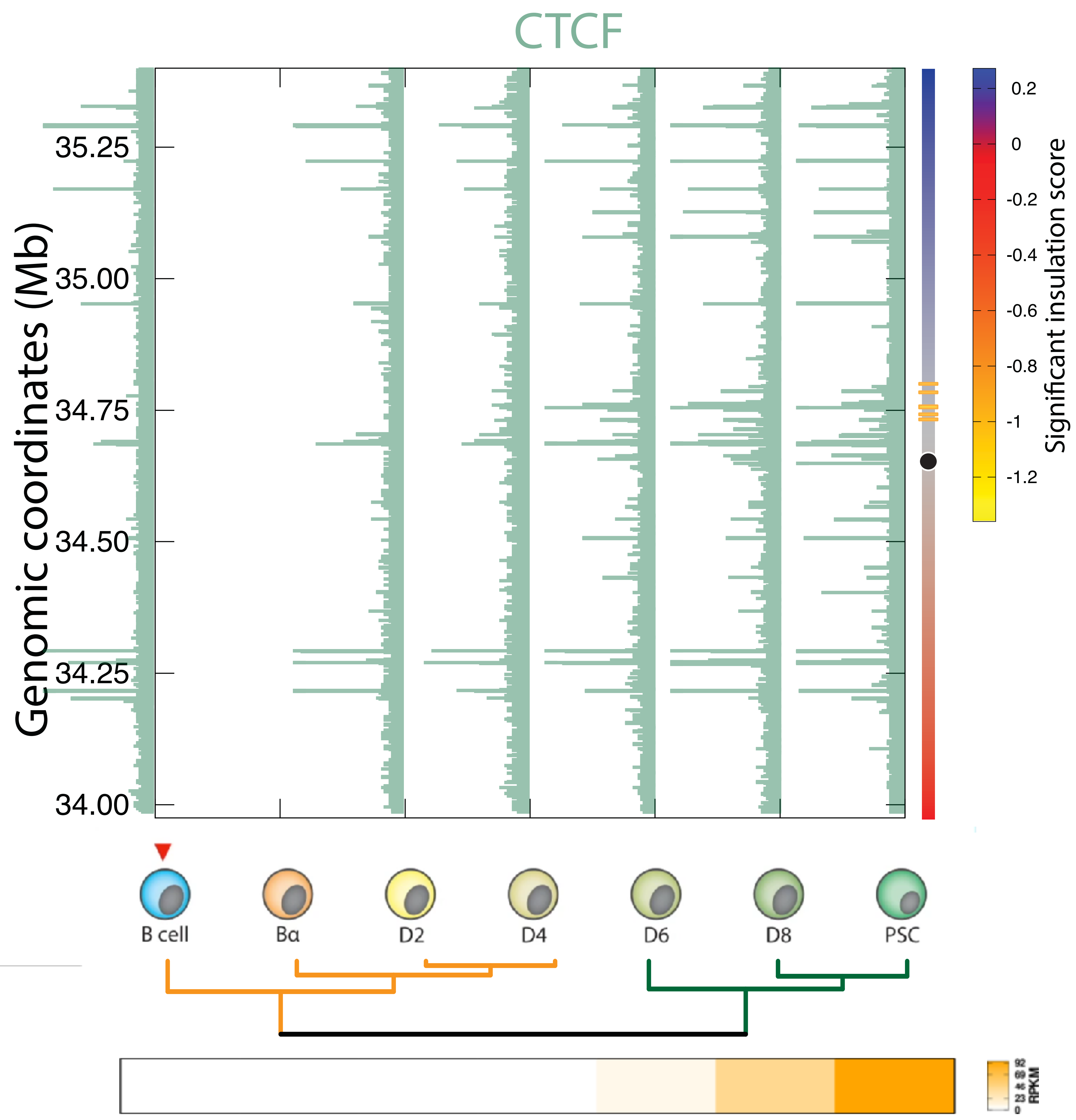




SOX2 locus structural changes from B to PSC

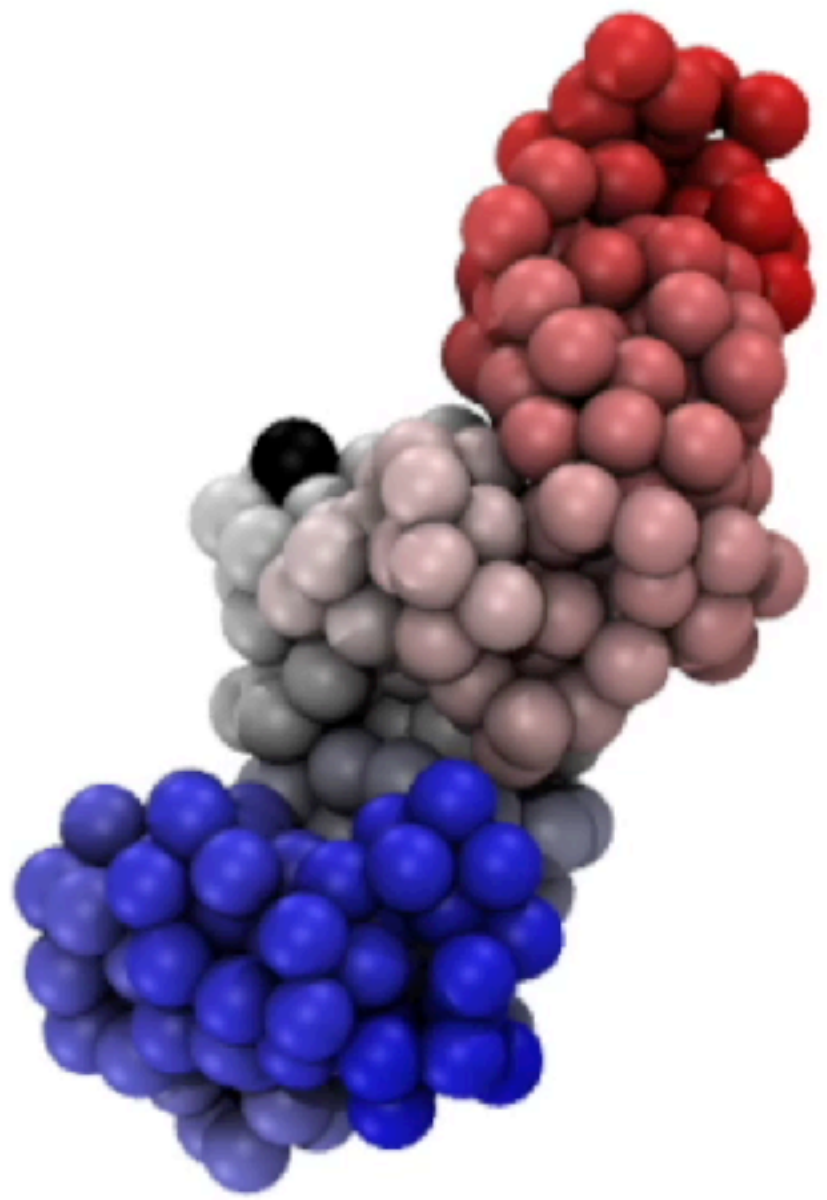
TAD borders

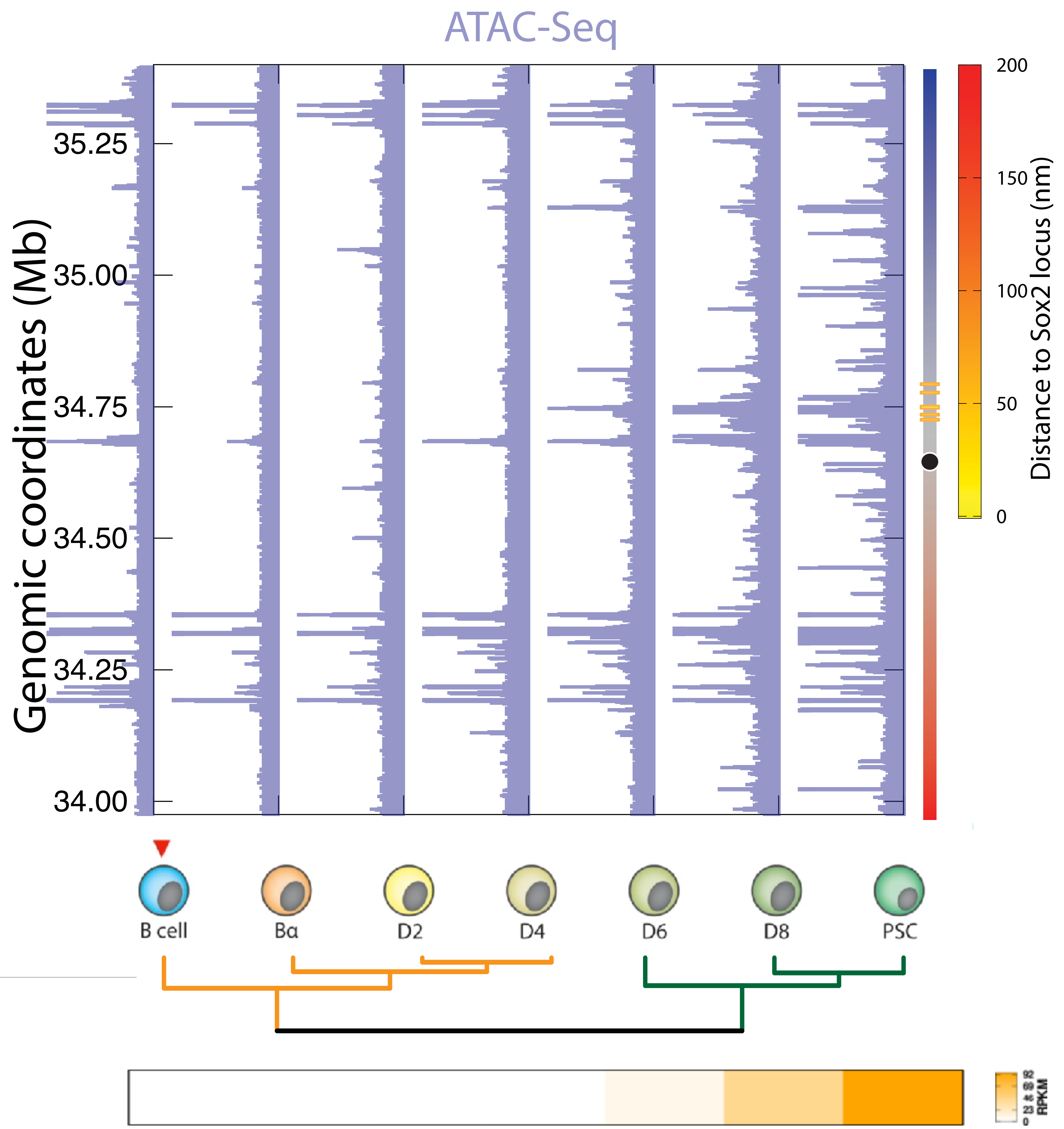




SOX2 locus structural changes from B to PSC

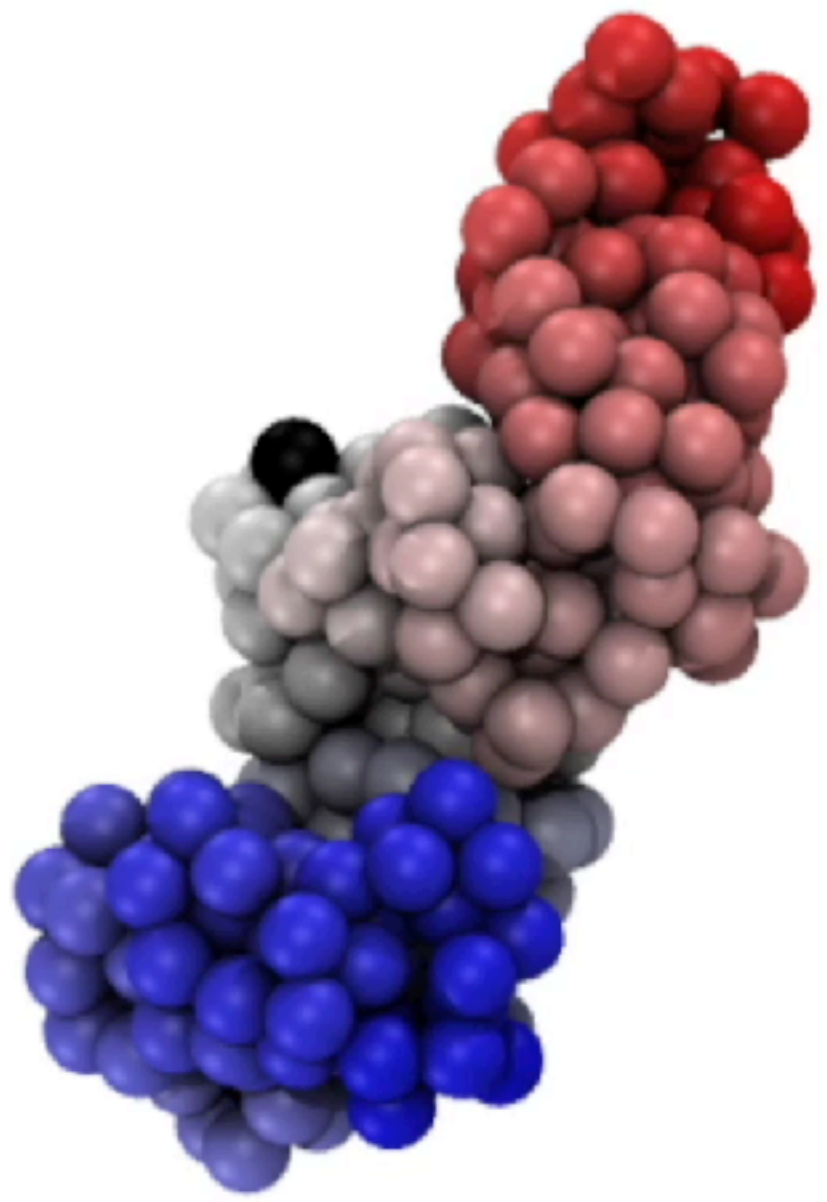
Distance to regulatory elements

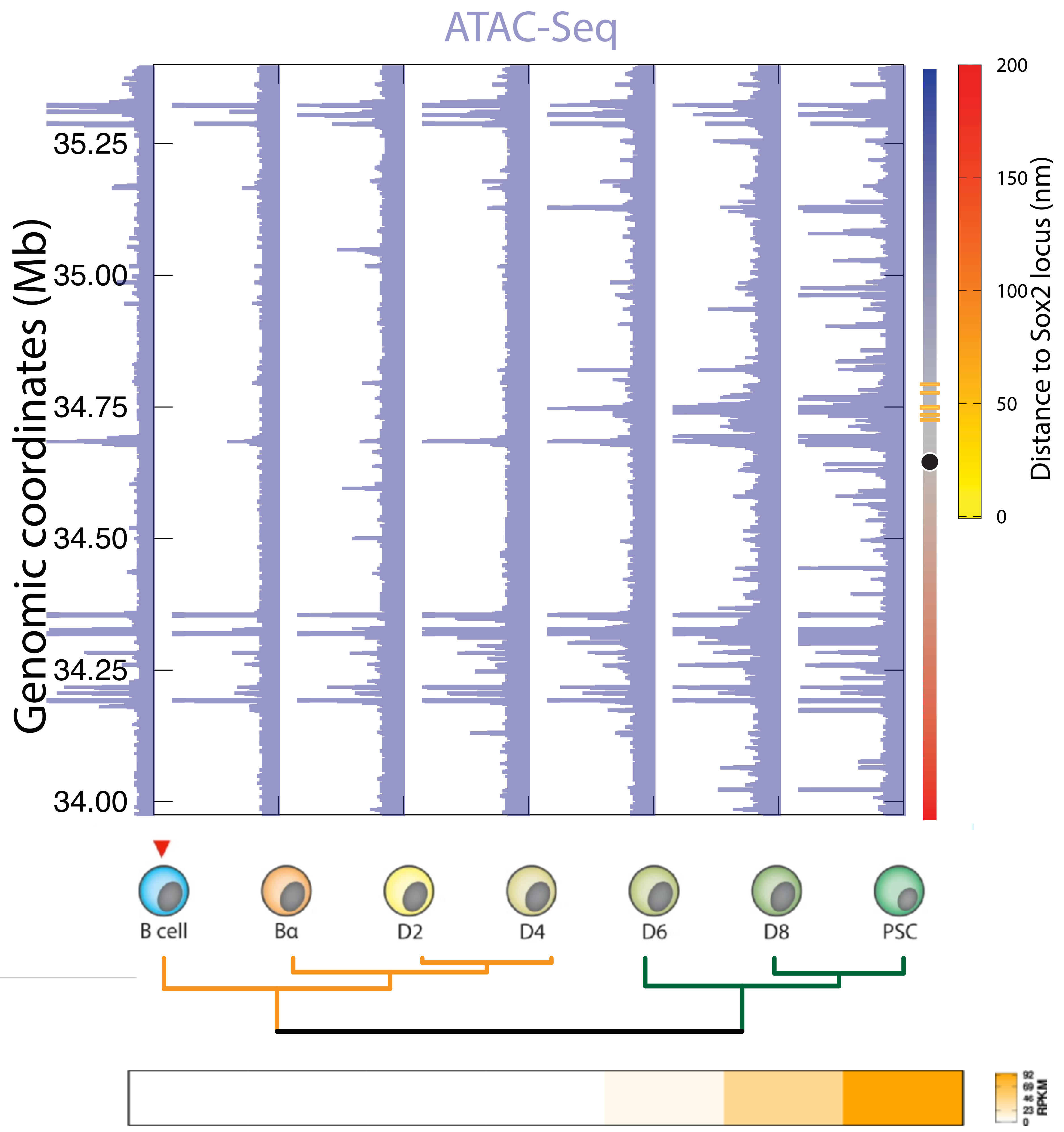




SOX2 locus structural changes from B to PSC

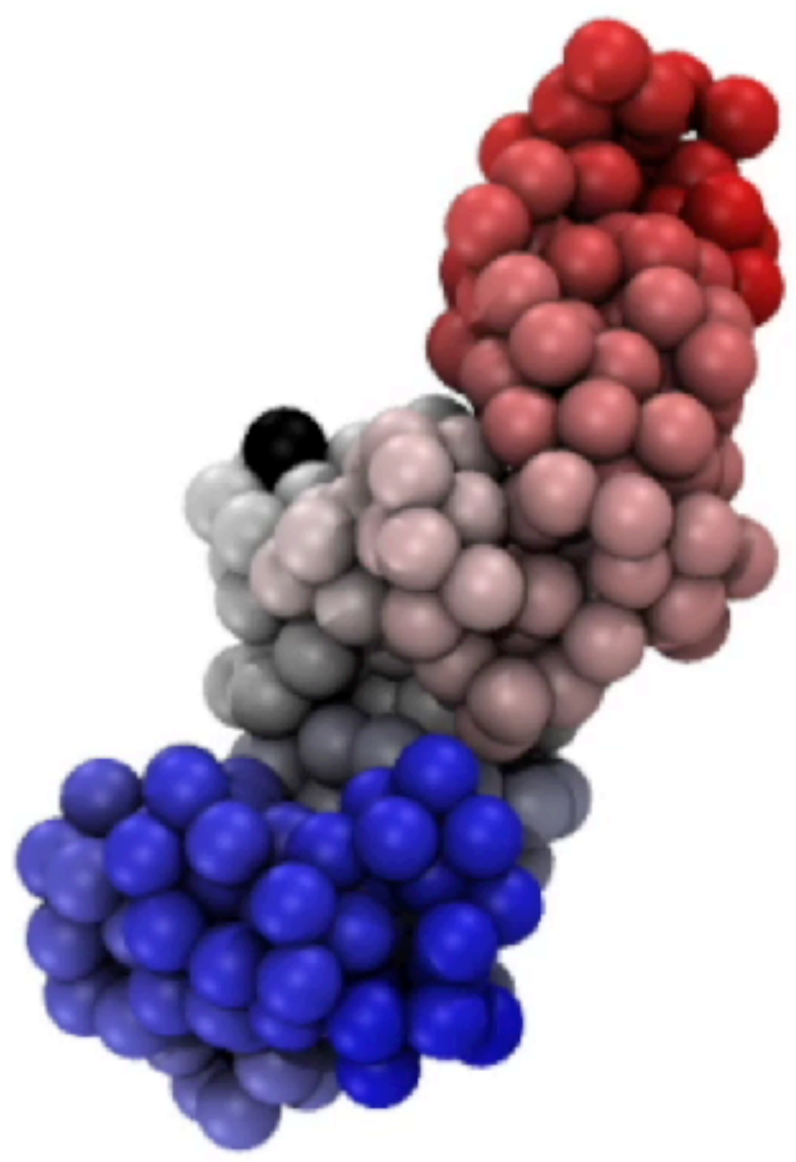
Distance to regulatory elements

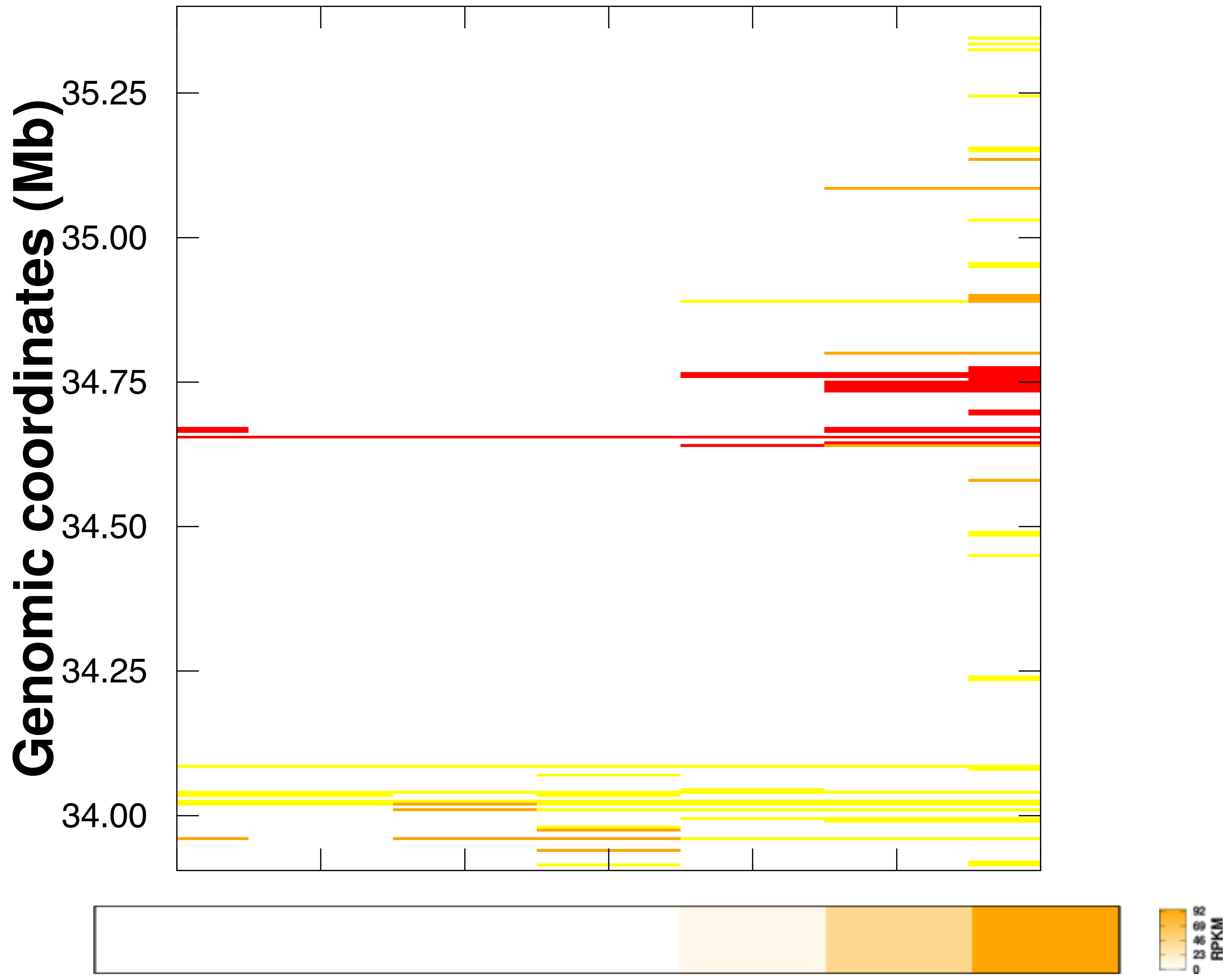




SOX2 locus structural changes from B to PSC

Chromatin Activity

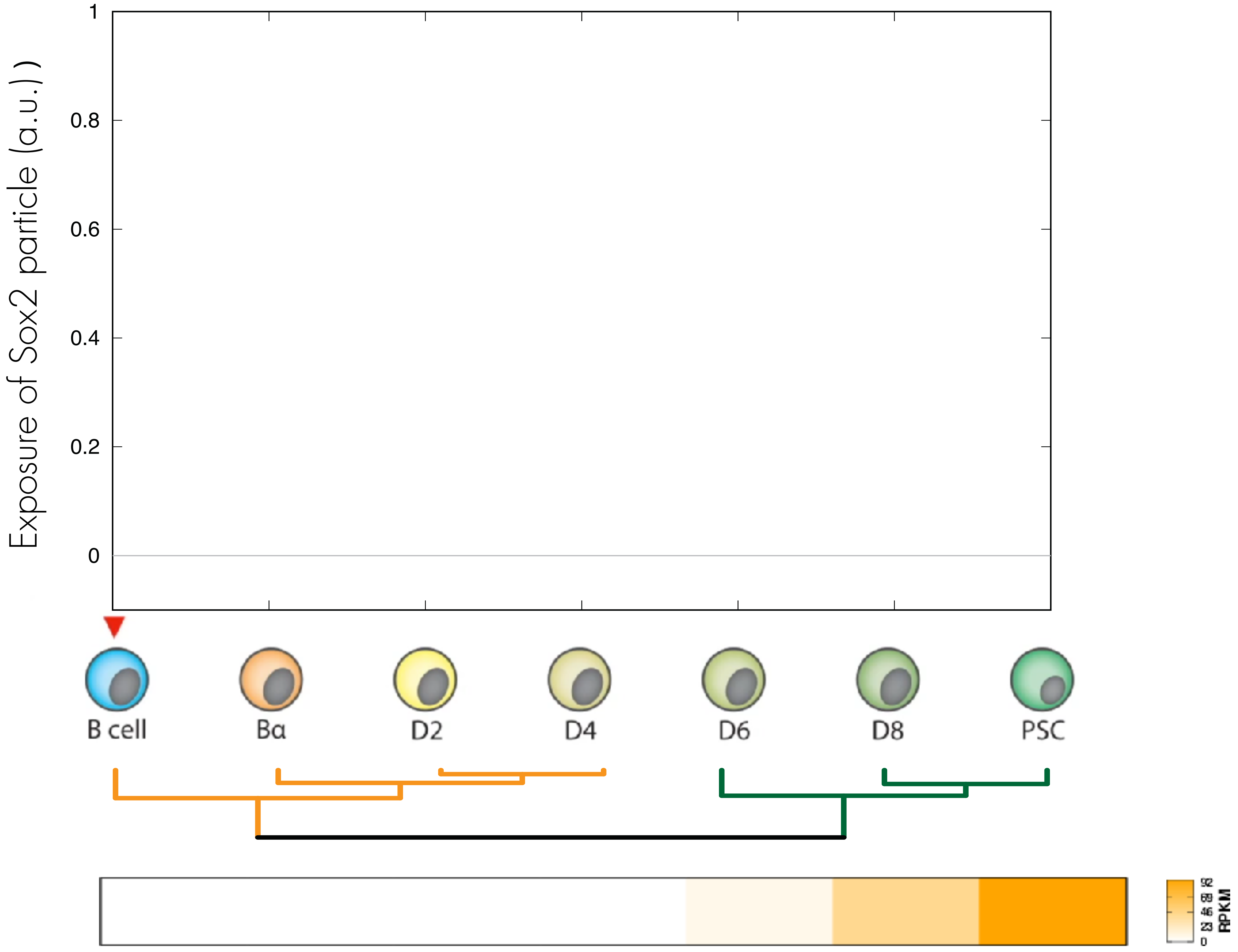




	B	Ba	D2	D4	D6	D8	PSC
A	9	6	7	13	13	22	48
AP	4	1	4	4	4	13	23
APD	3	1	1	1	4	10	15

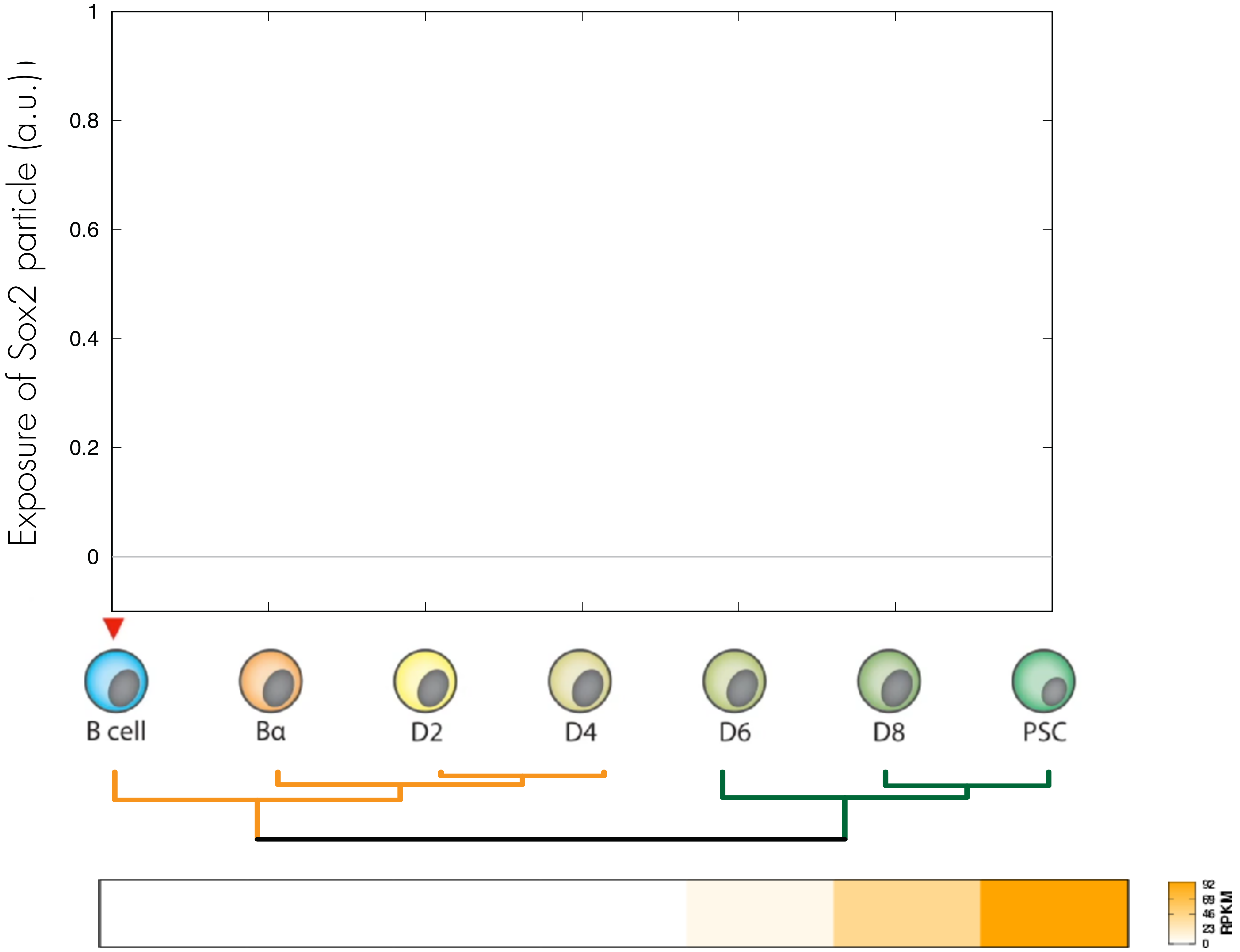
SOX2 locus structural changes from B to PSC

Structural exposure



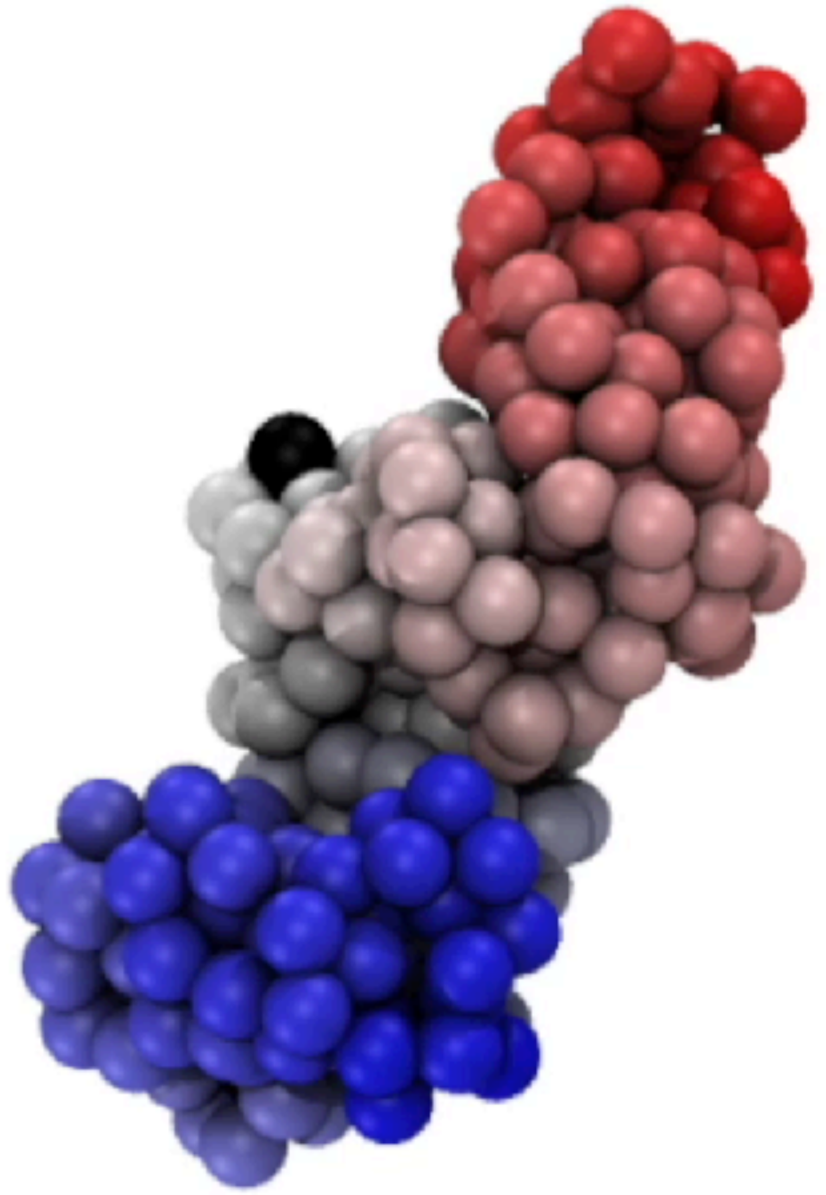
SOX2 locus structural changes from B to PSC

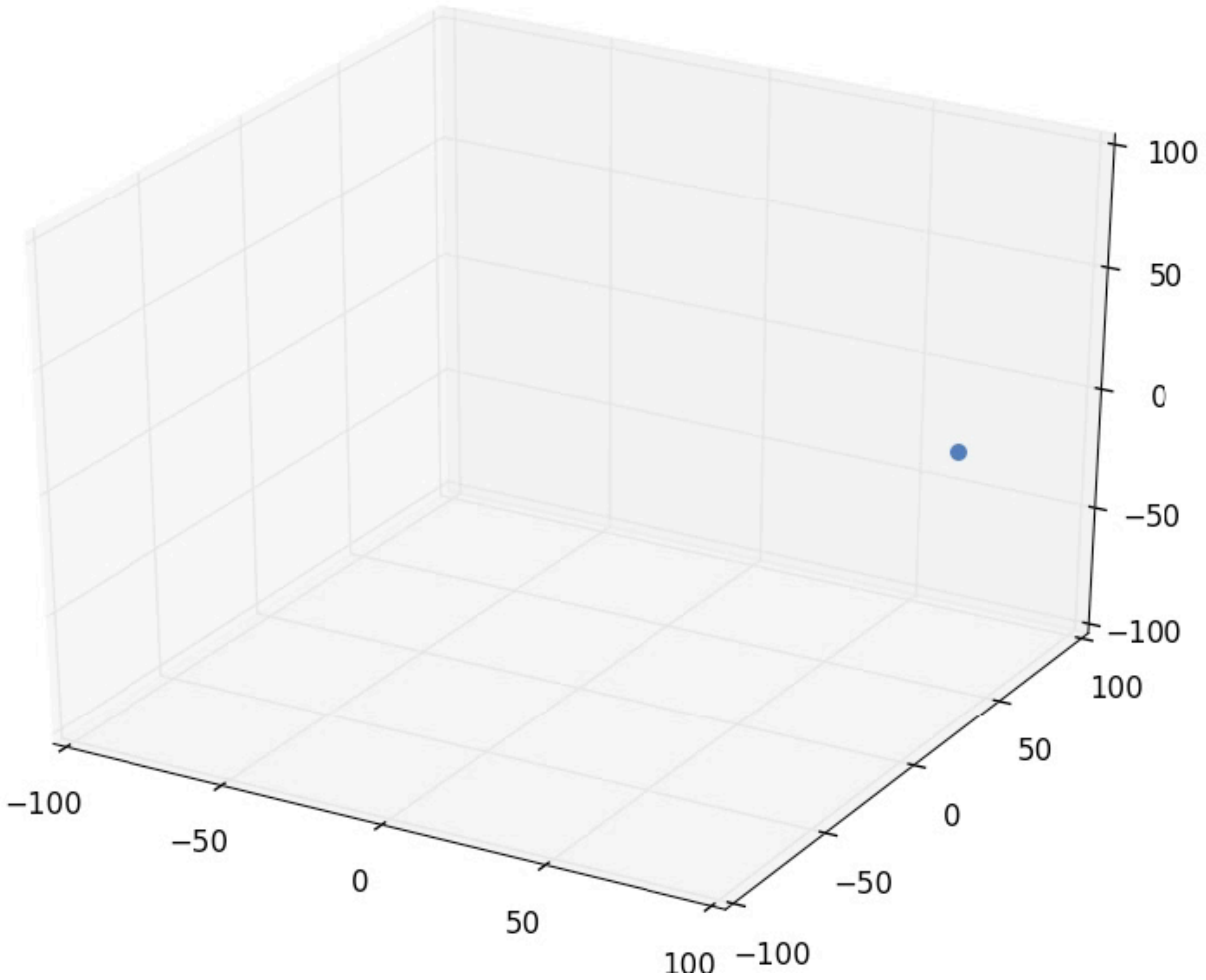
Structural exposure

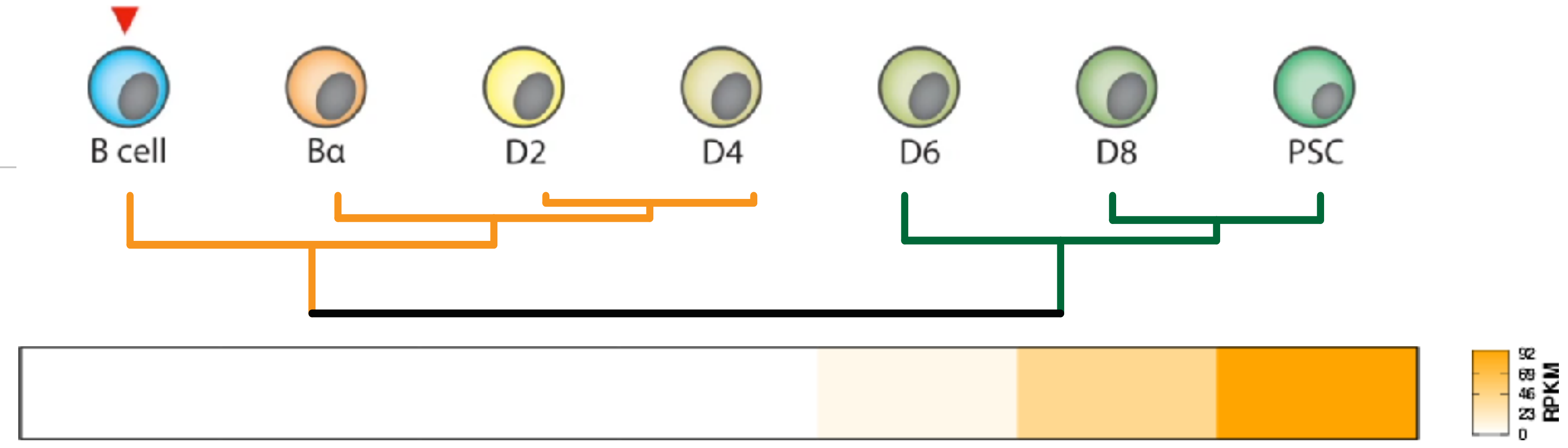


SOX2 locus dynamics changes from B to PSC

SOX2 displacement

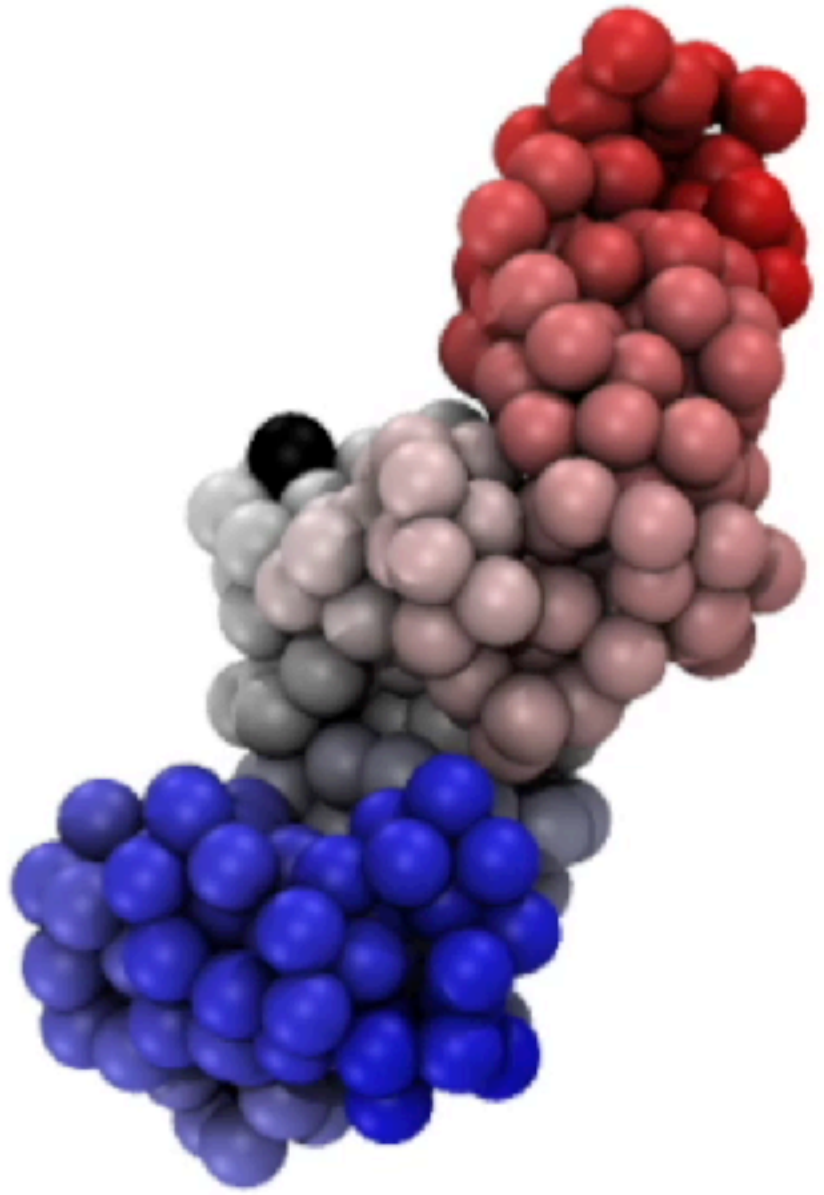


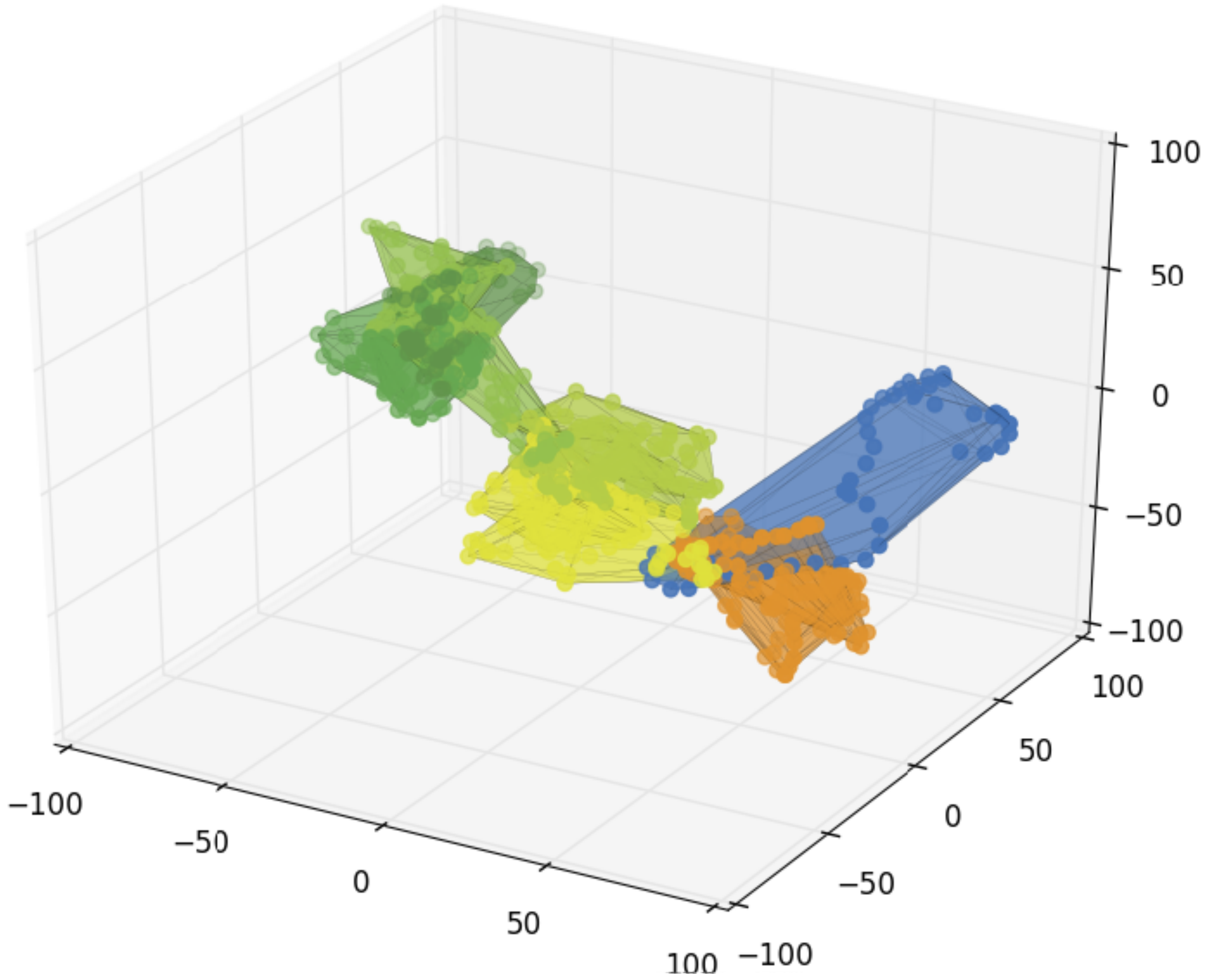


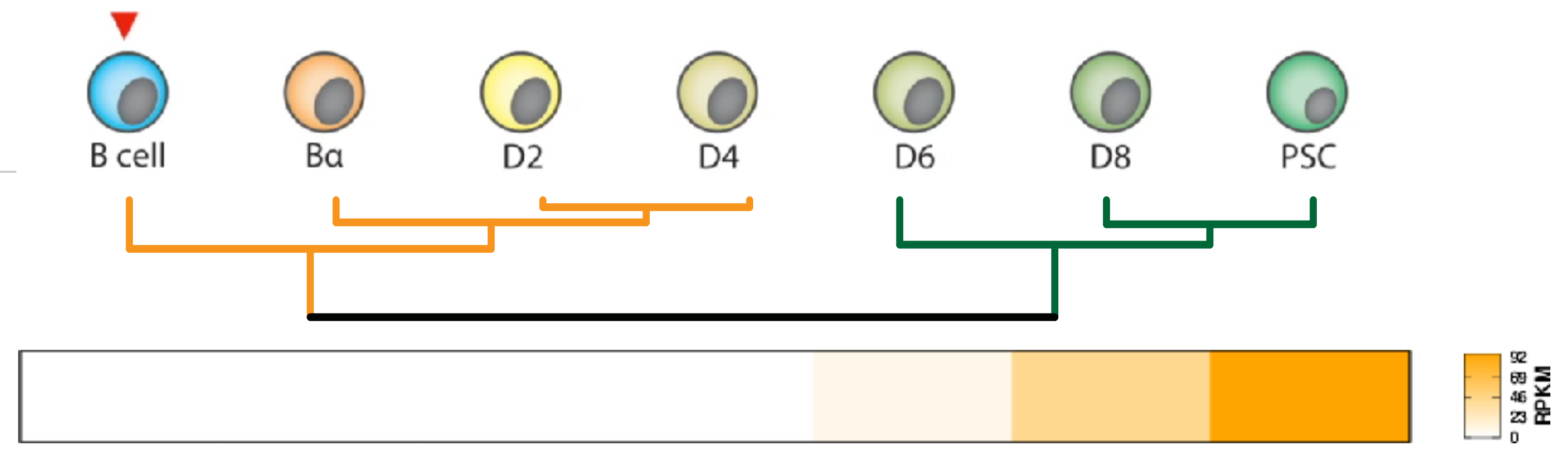


SOX2 locus dynamics changes from B to PSC

SOX2 displacement

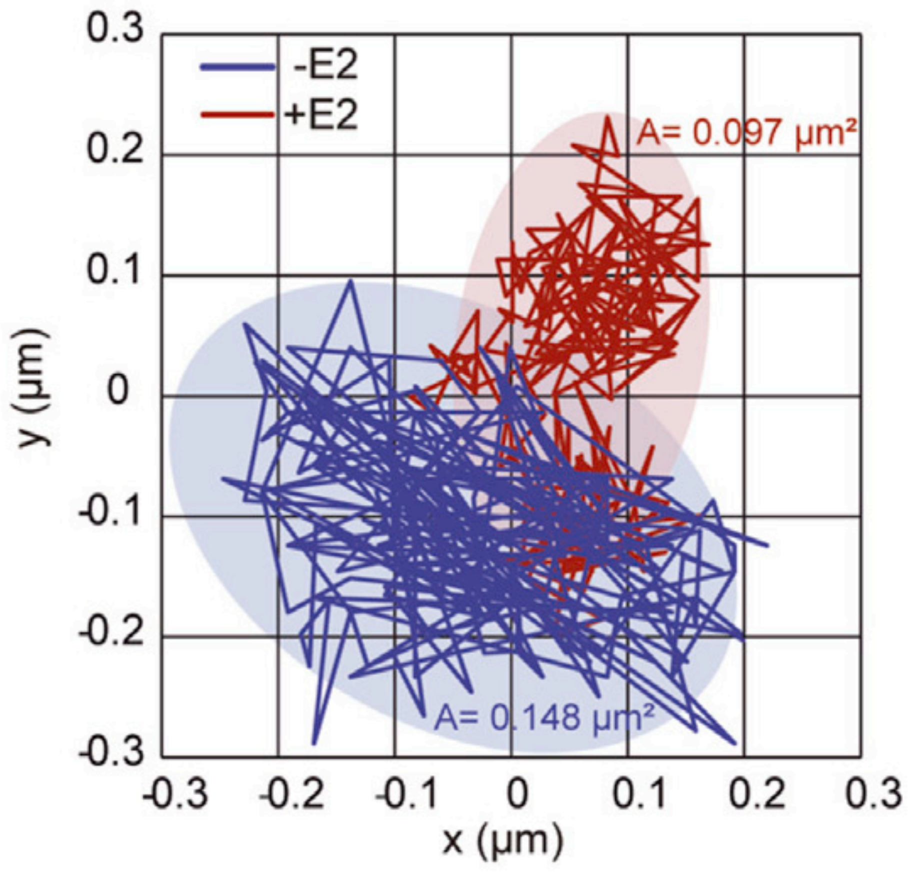






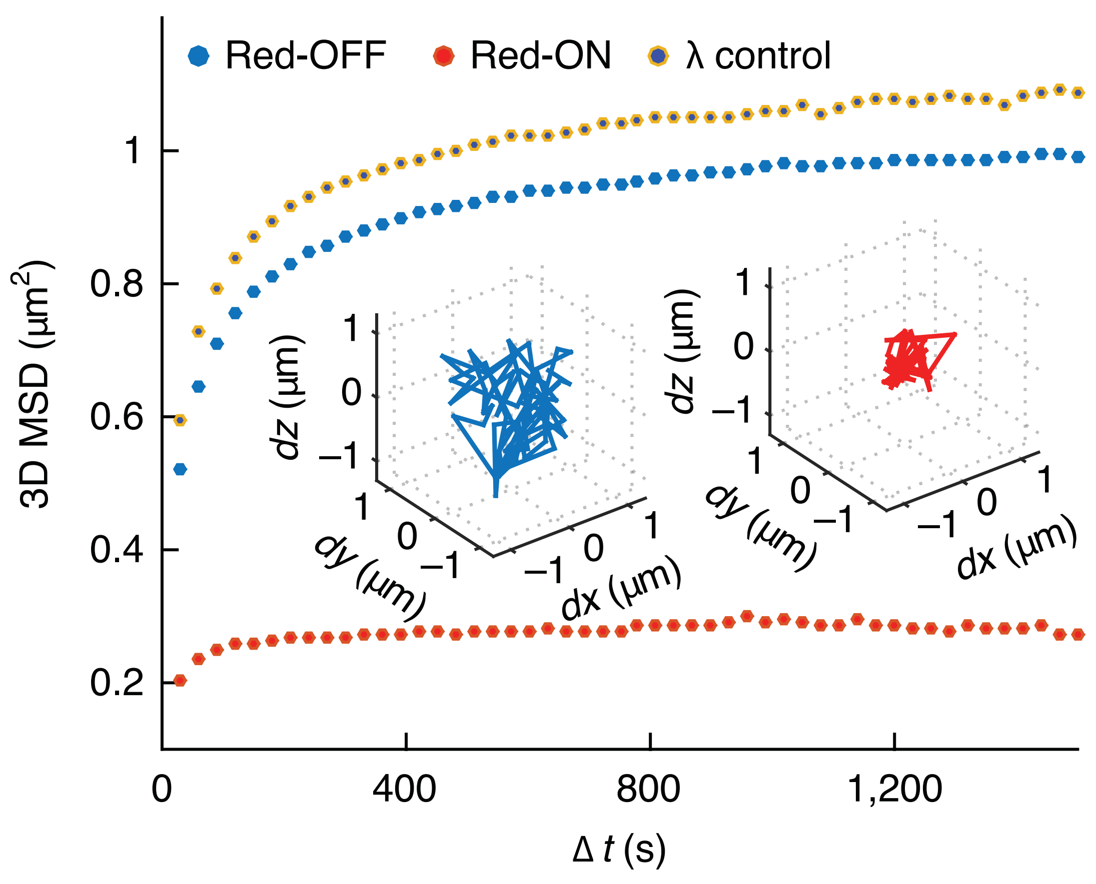
SOX2 locus dynamics changes from B to PSC

SOX2 displacement



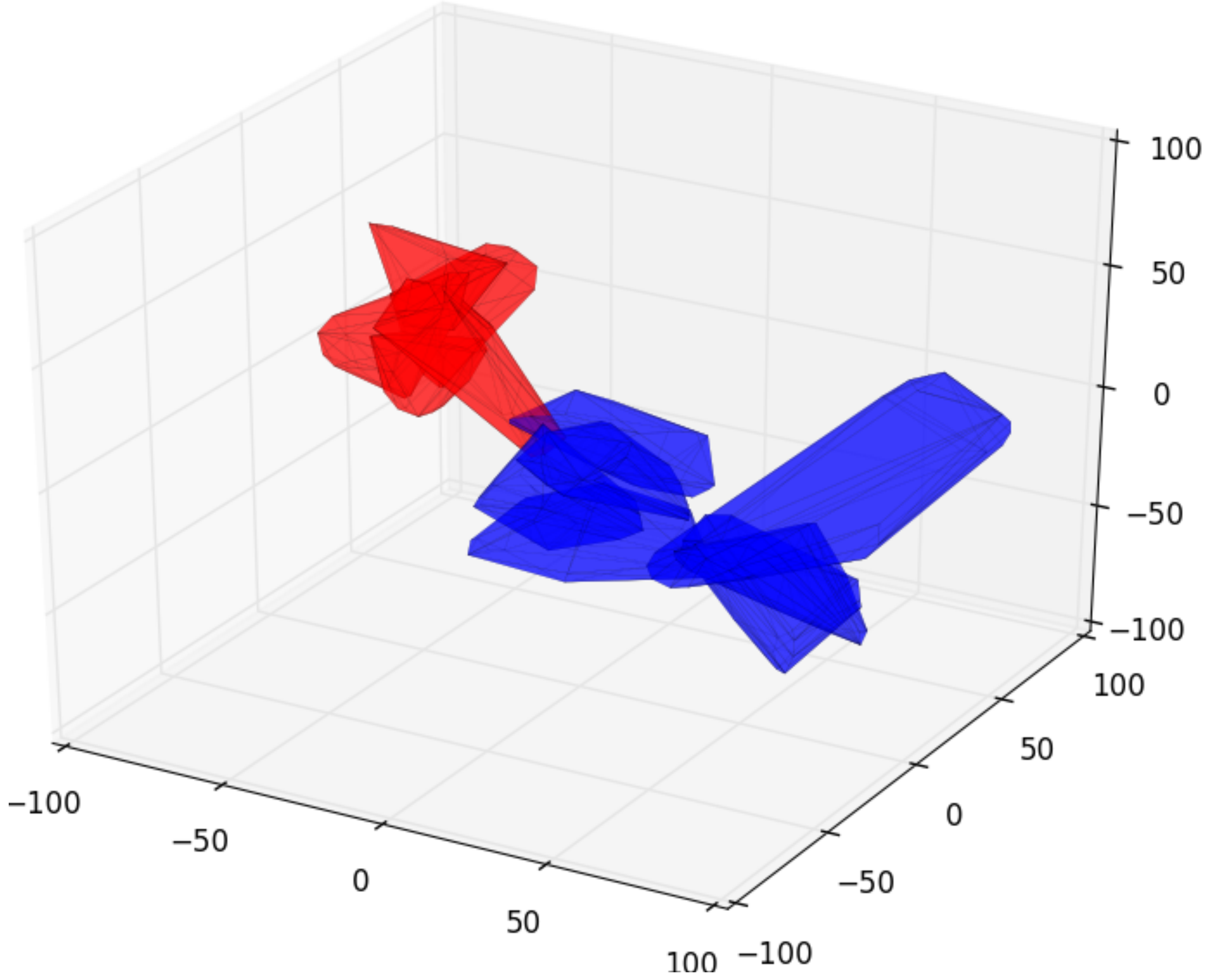
Two dimensional trajectories and area explored over 50s of the CCND1 locus recored before -E2 and after +E2 activation.

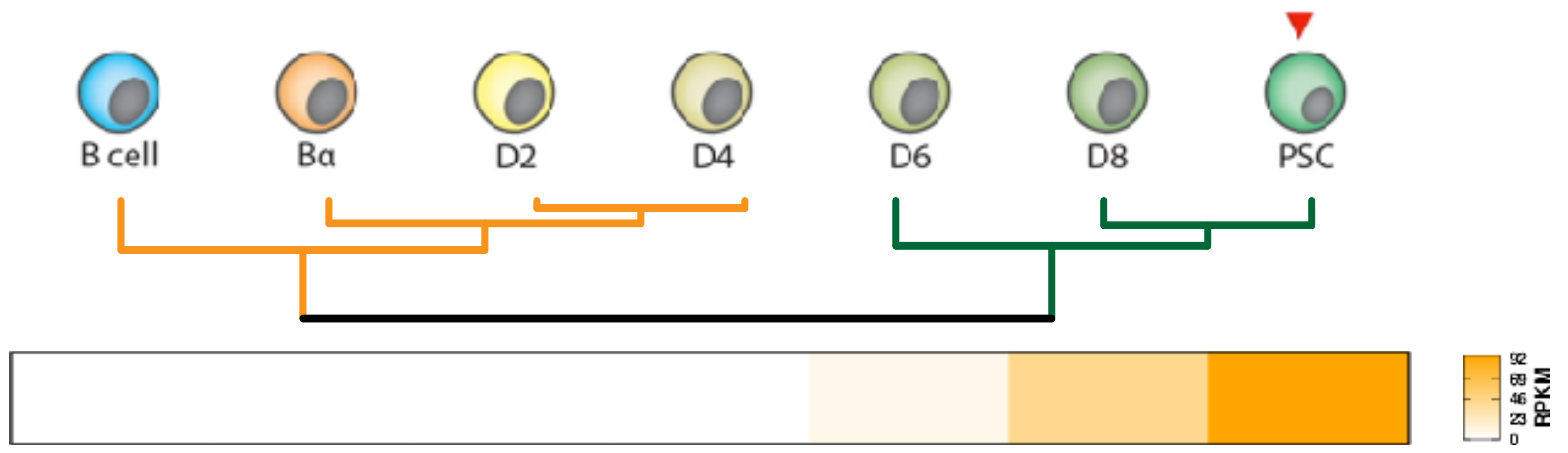
Germier ,T., et al, (2017) Blophys J.



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

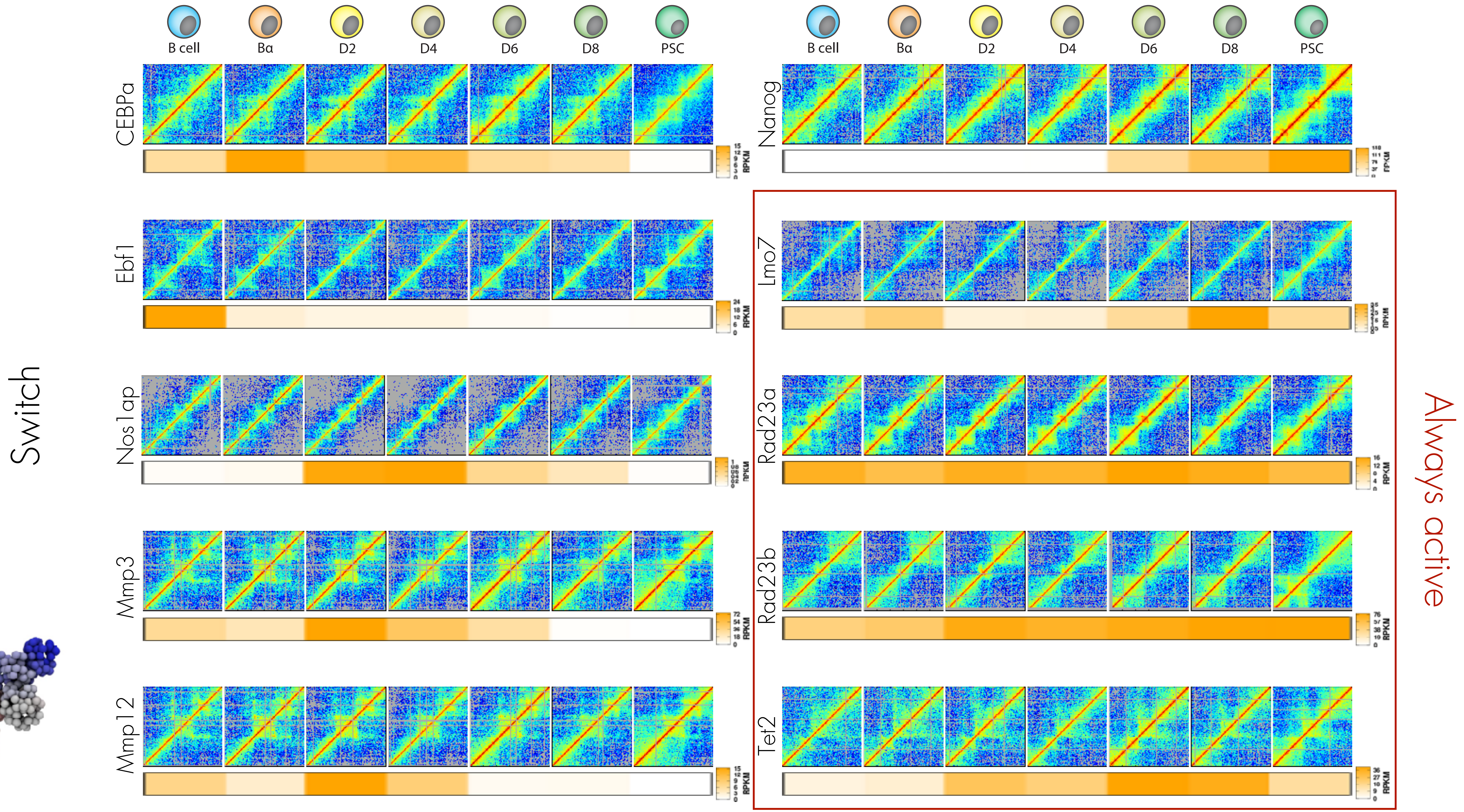
Chen ,T., et al, (2018) Nat. Genetics

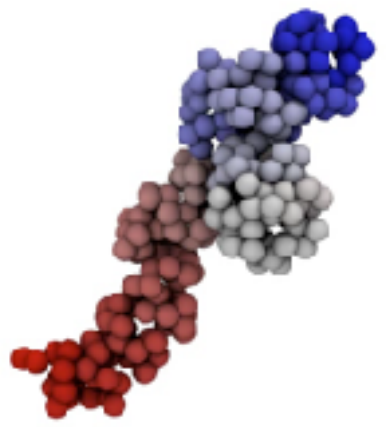




Structural changes from B to PSC

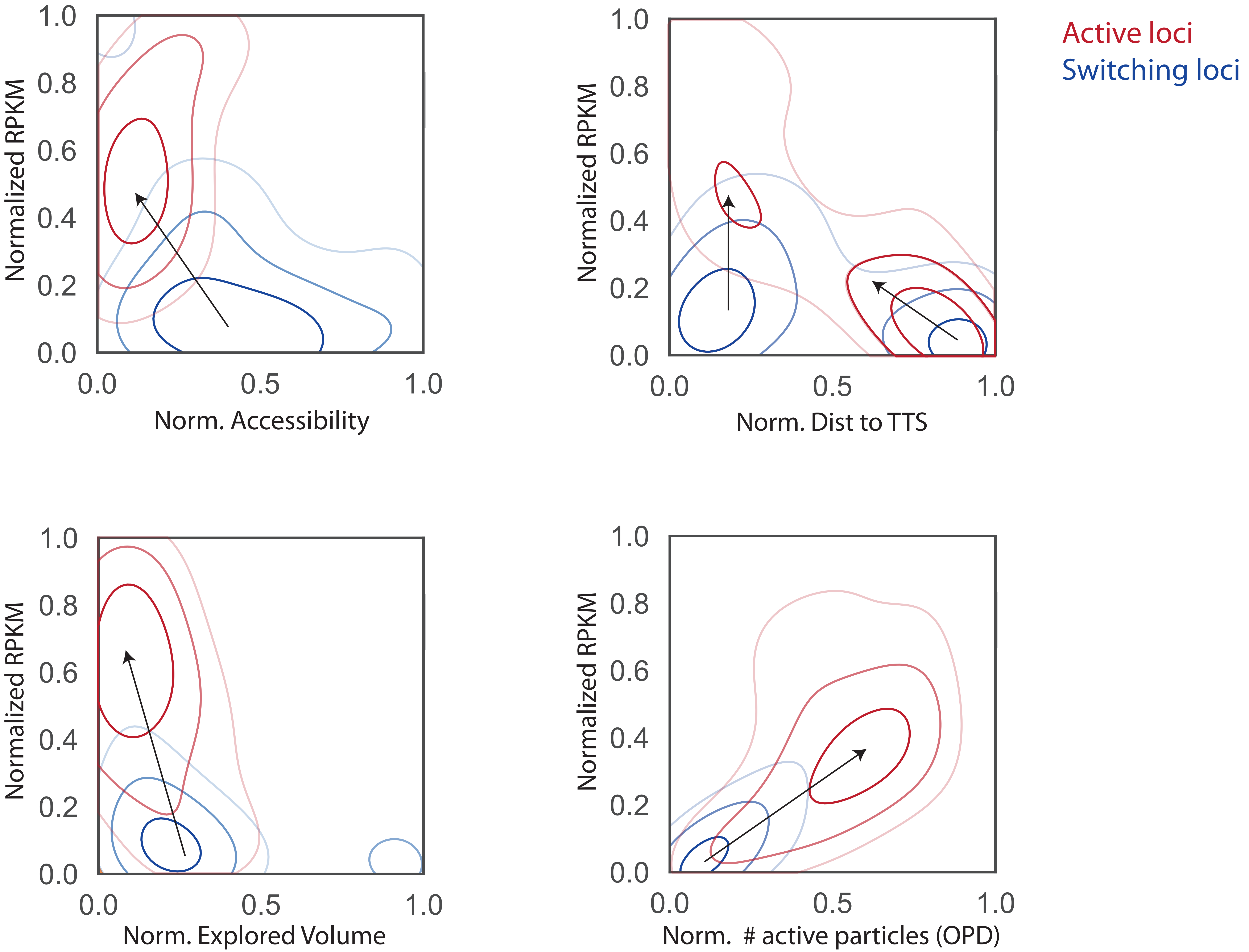
Other 10 loci



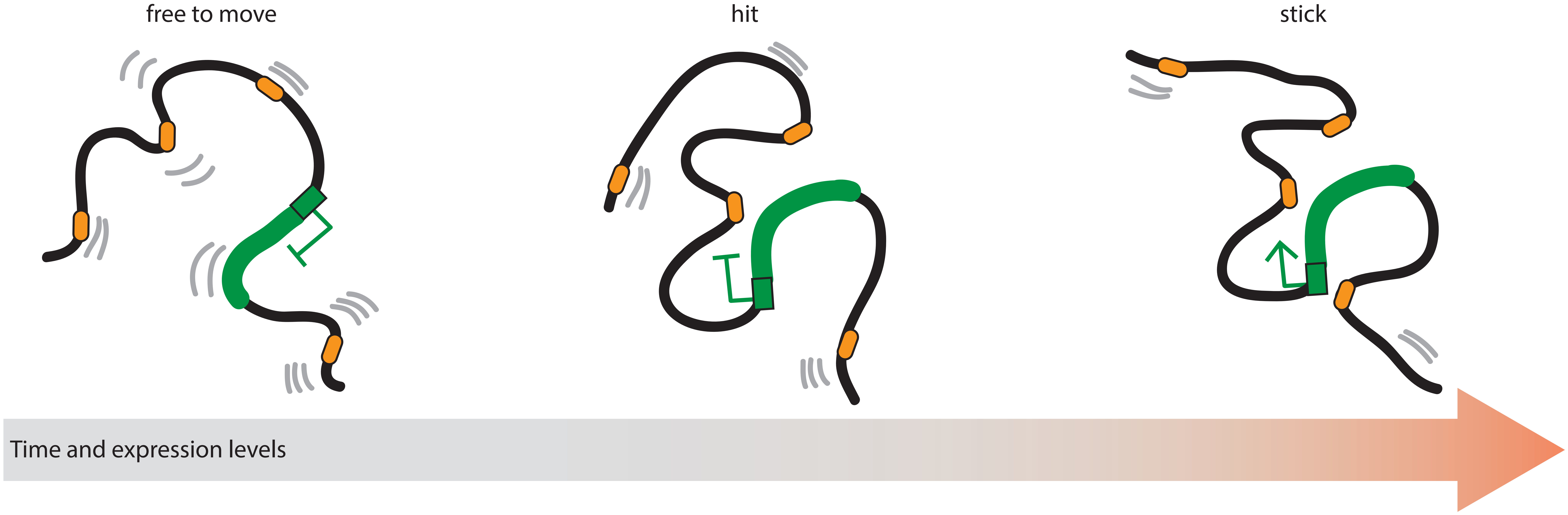


Dynamics of gene activation

Trends in all 11 loci



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



free to move

hit

stick

Time and expression levels

What next?

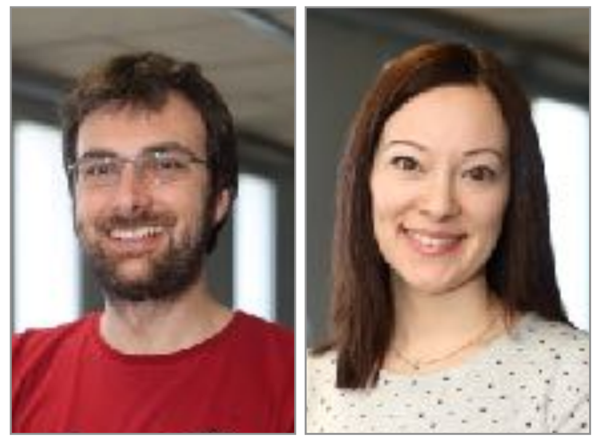
Can we see the path of chromatin @10Kb resolution? images + Hi-C + modeling

Is genome structure more conserved than sequence?

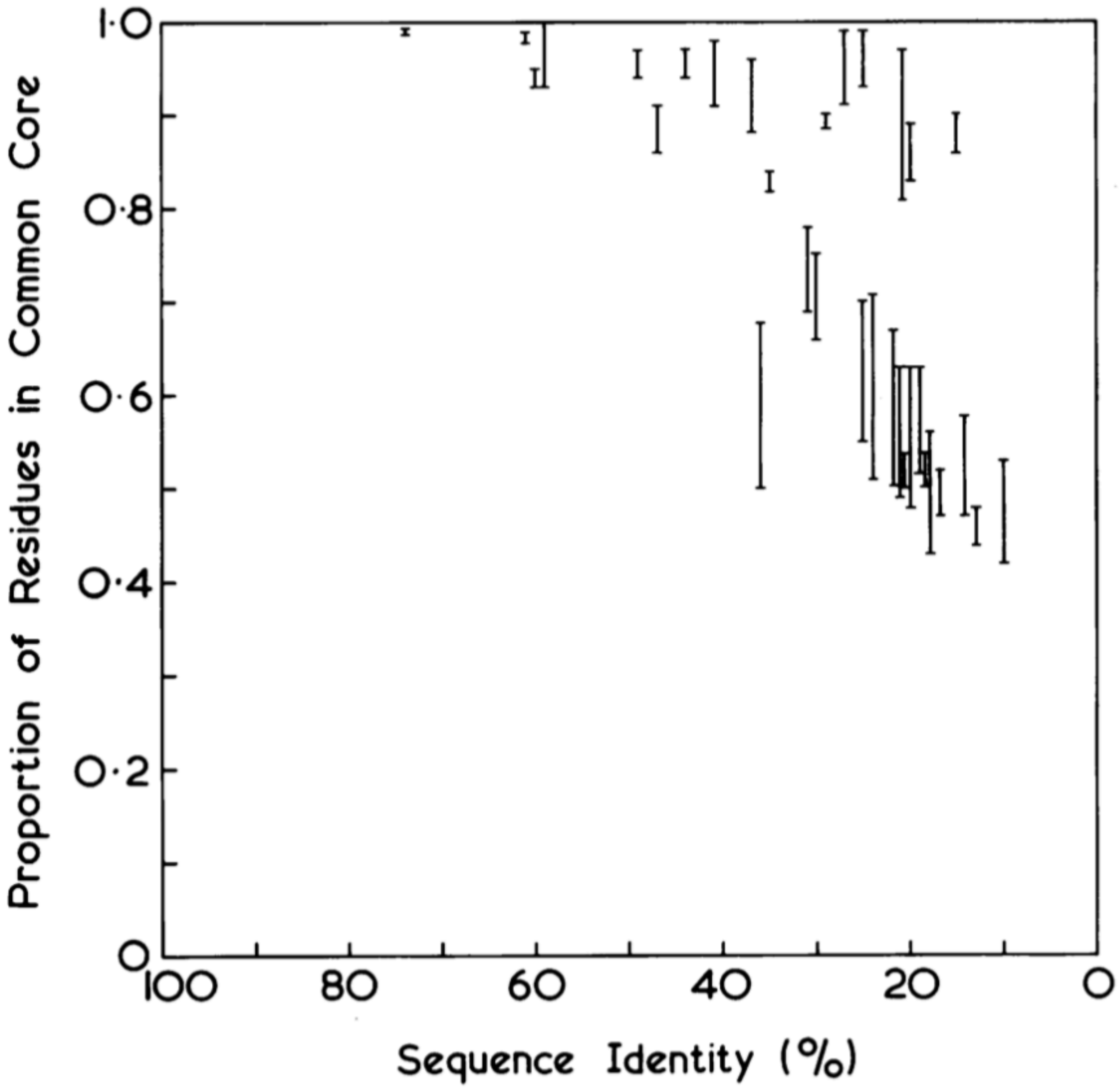
The End!


Photo by David Oliete - www.davidoliete.com

Genome 3D structure is more conserved than 1D in primates

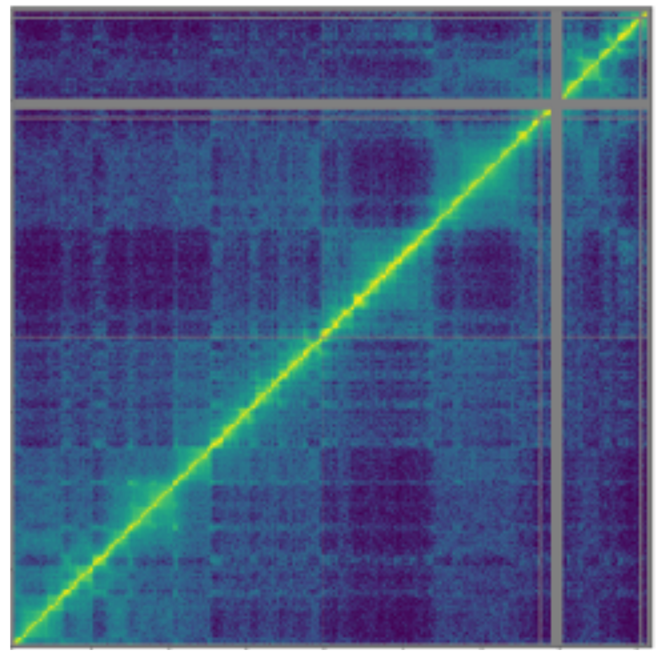

Unpublished

François Serra
Yasmina Cuartero
with Marquès Lab (UPF, Barcelona)



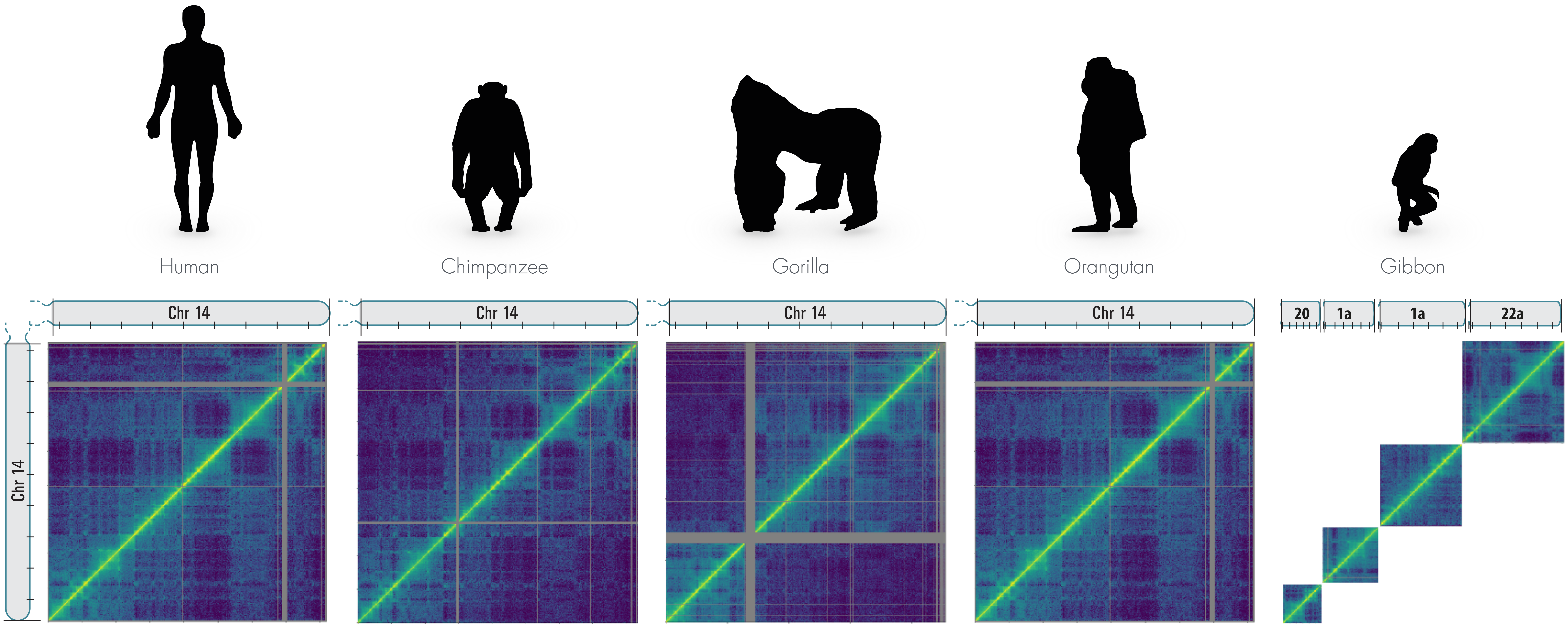
C. Chothia & A. Lesk (1986) EMBO J. 5(4):823-826

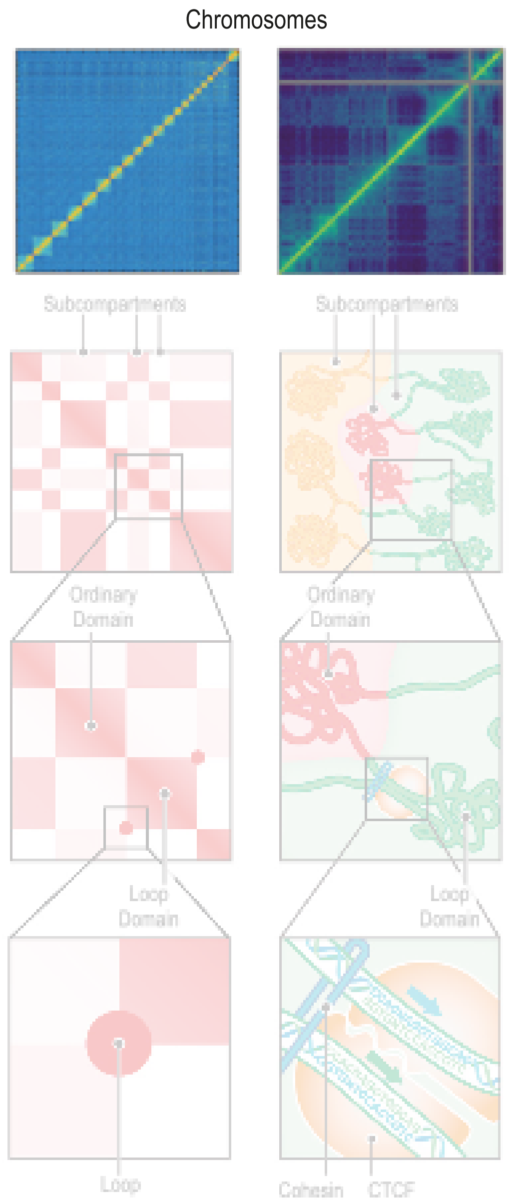
- Proteome is 10sM AAs in 10,000s of proteins
- 20 AA in average ~500 AA length
- ~150 structural AA for ~5 AA for function
- MM+QM
- No confinement
- Dynamics



- Genome is 1,000sM NTs in 10s of chroms
- 4 NT in average ~100M NTs length
- ~??? structural NTs for ~??? NTs for function
- Packing + Phase Separation + Modelers
- Confinement
- ++++Dynamics

Hi-C matrices from lymphoblasts in seven primates

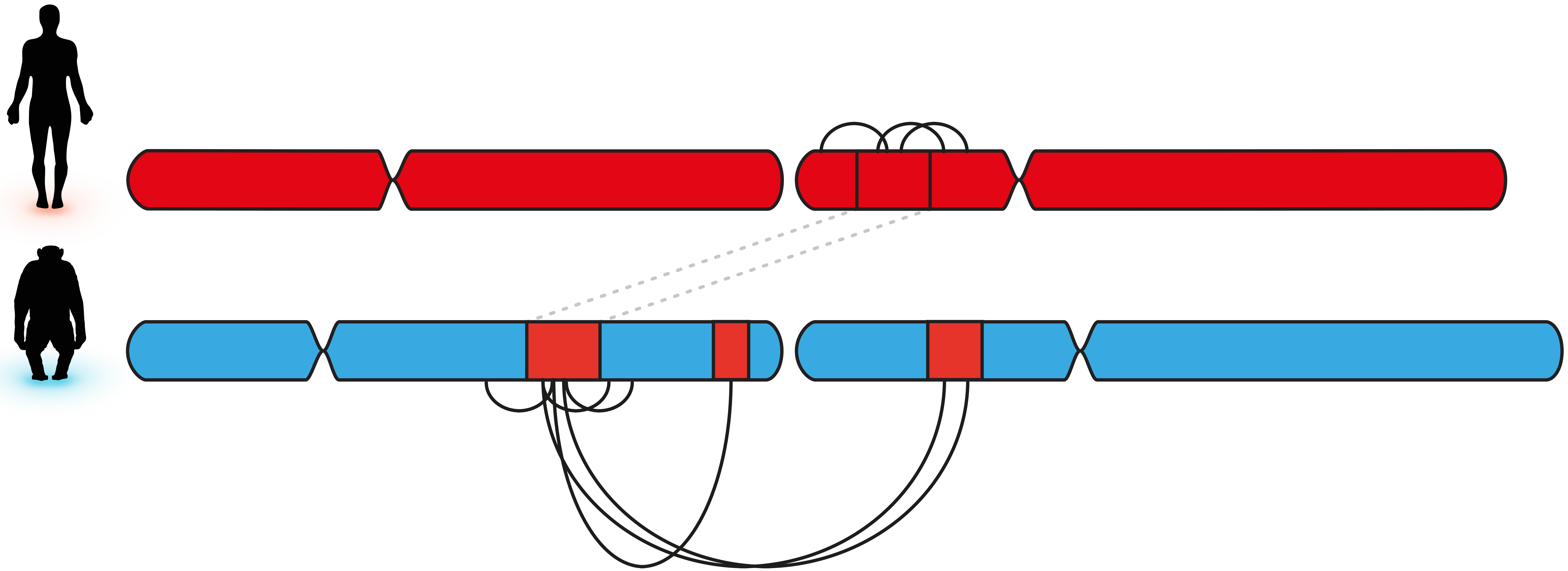


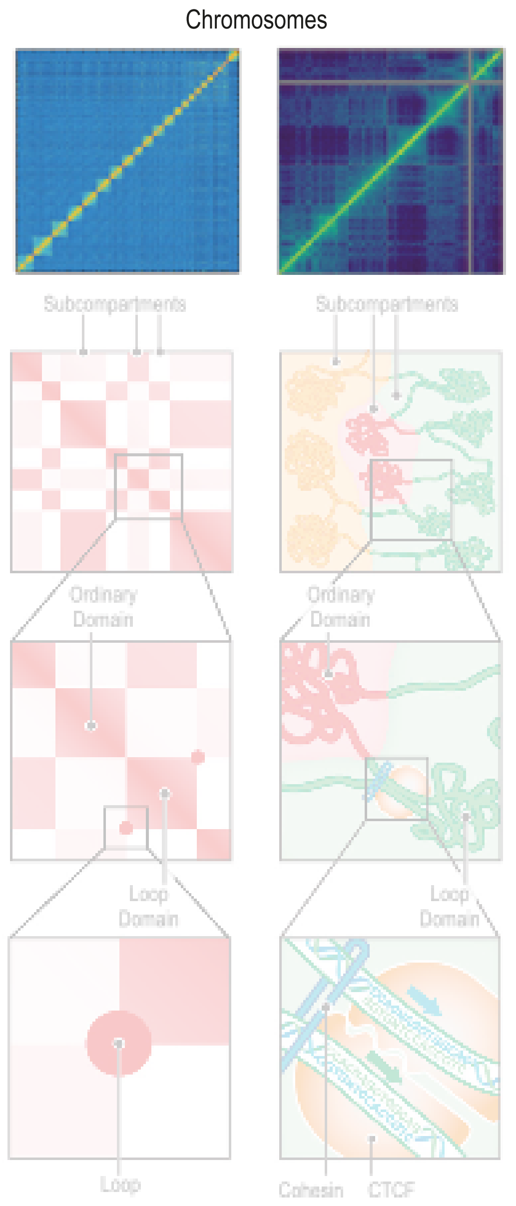


Synteny breakpoints in 3D

Two regions of at least 100 kb separated by more than 750 kb (including trans chromosomal)

Common detections from Ruiz-Herrera's Lab (@300Kb res) and our lab (ENSEMBLE @1Kb res)

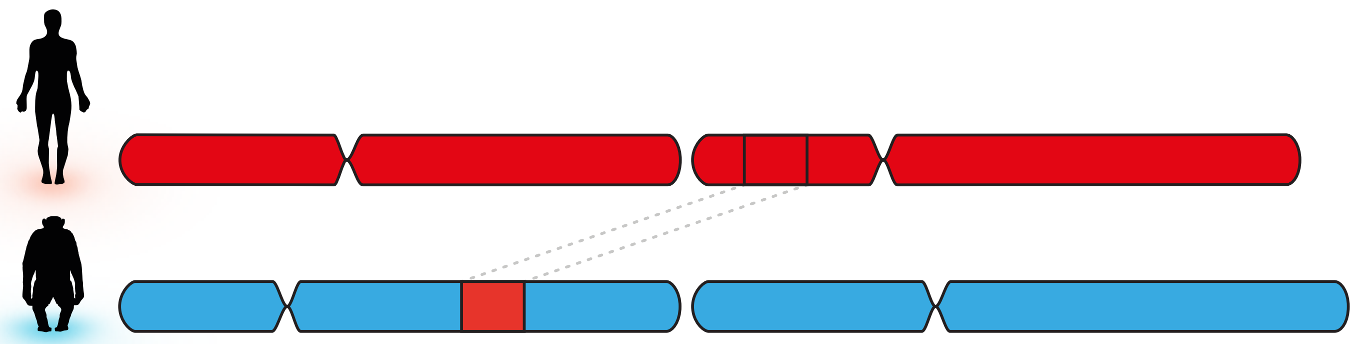


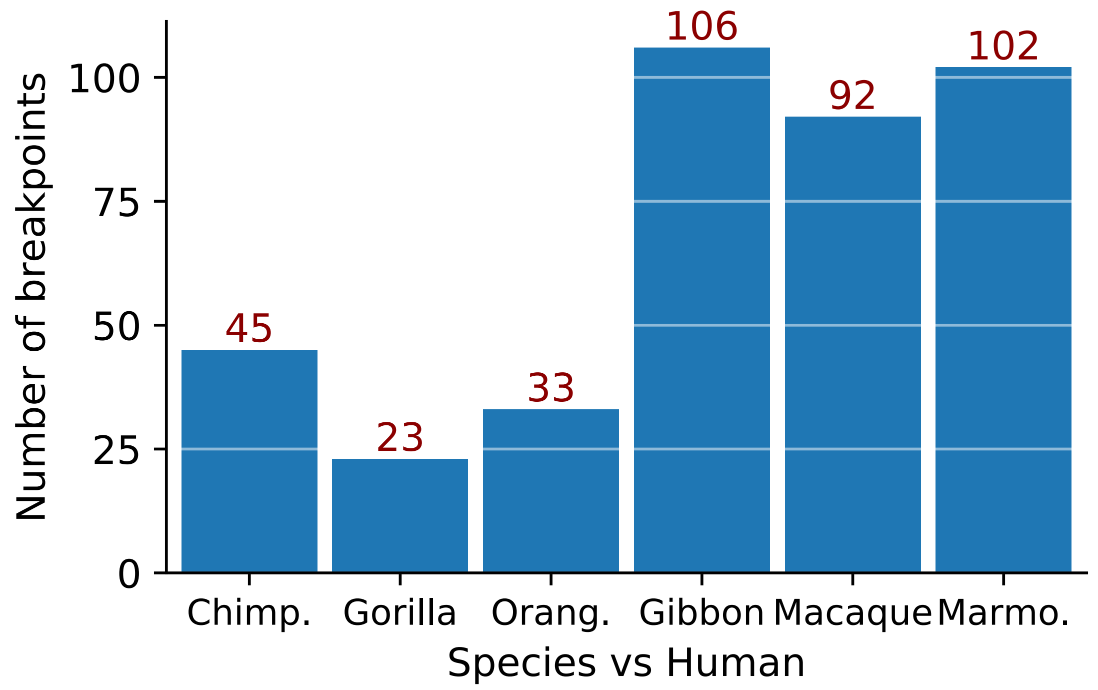


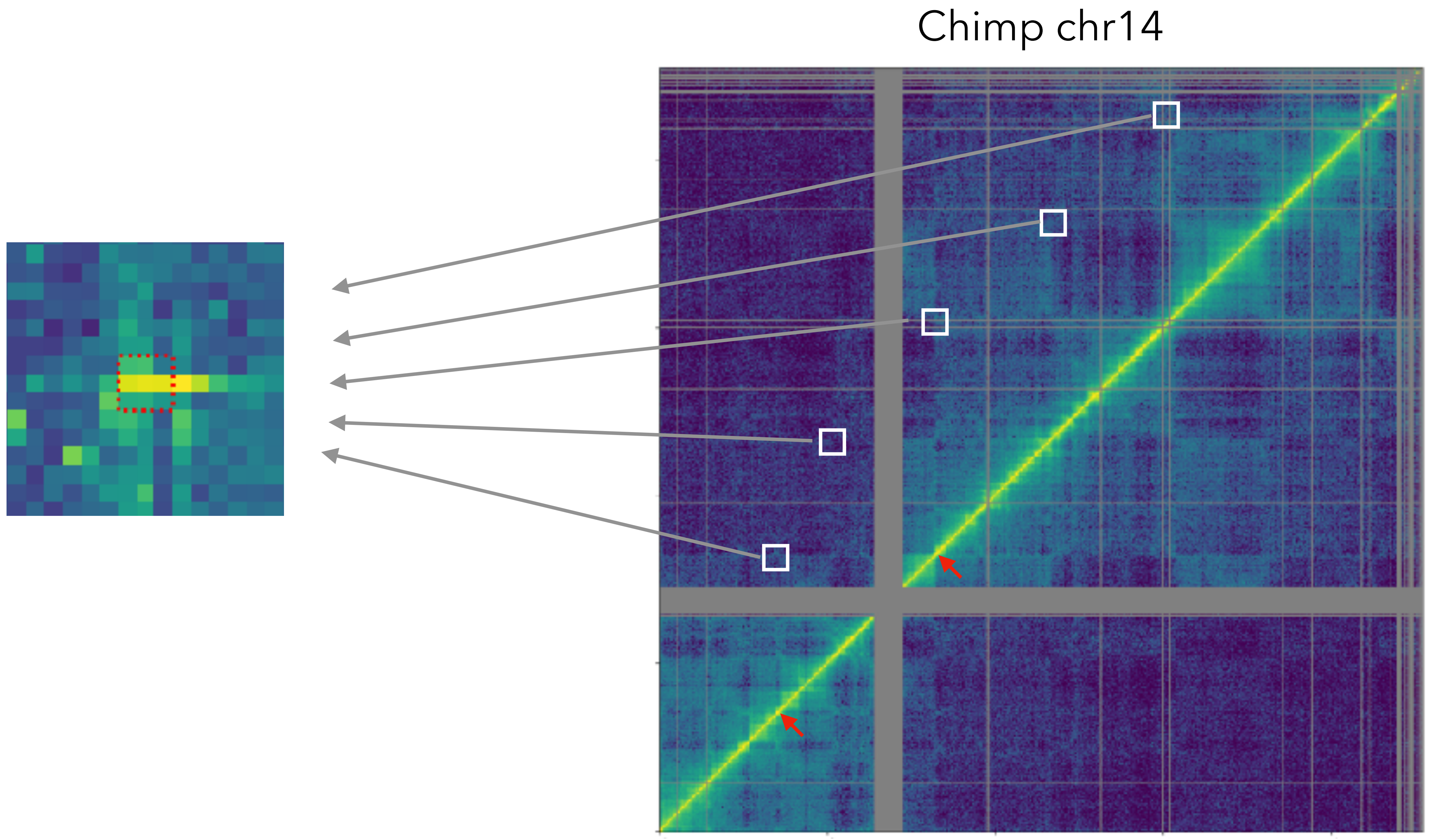
Synteny breakpoints in 3D

Two regions of at least 100 kb separated by more than 750 kb (including trans chromosomal)

Common detections from Ruiz-Herrera's Lab (@300Kb res) and our lab (ENSEMBLE @1Kb res)

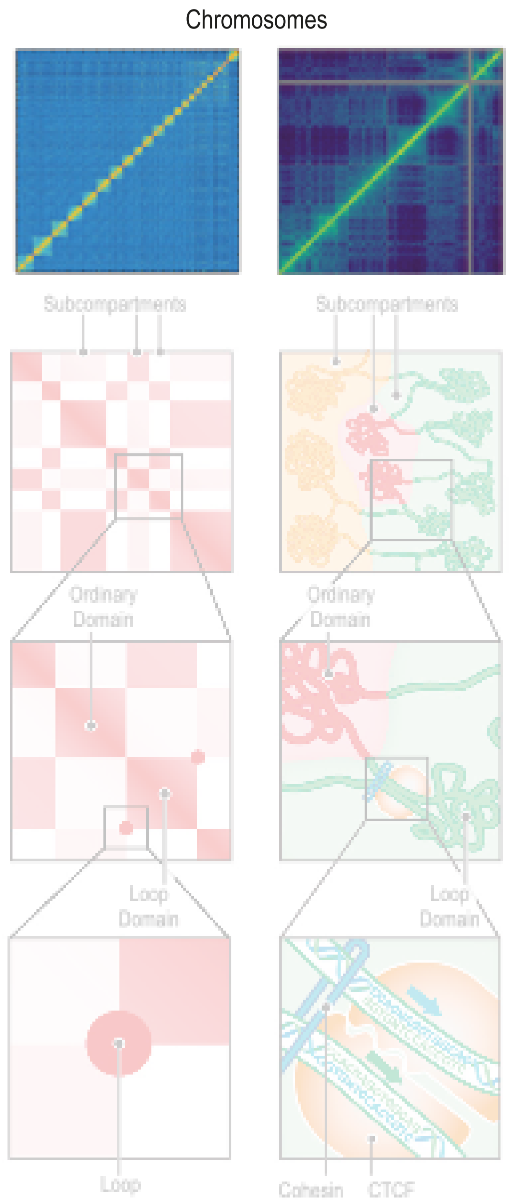


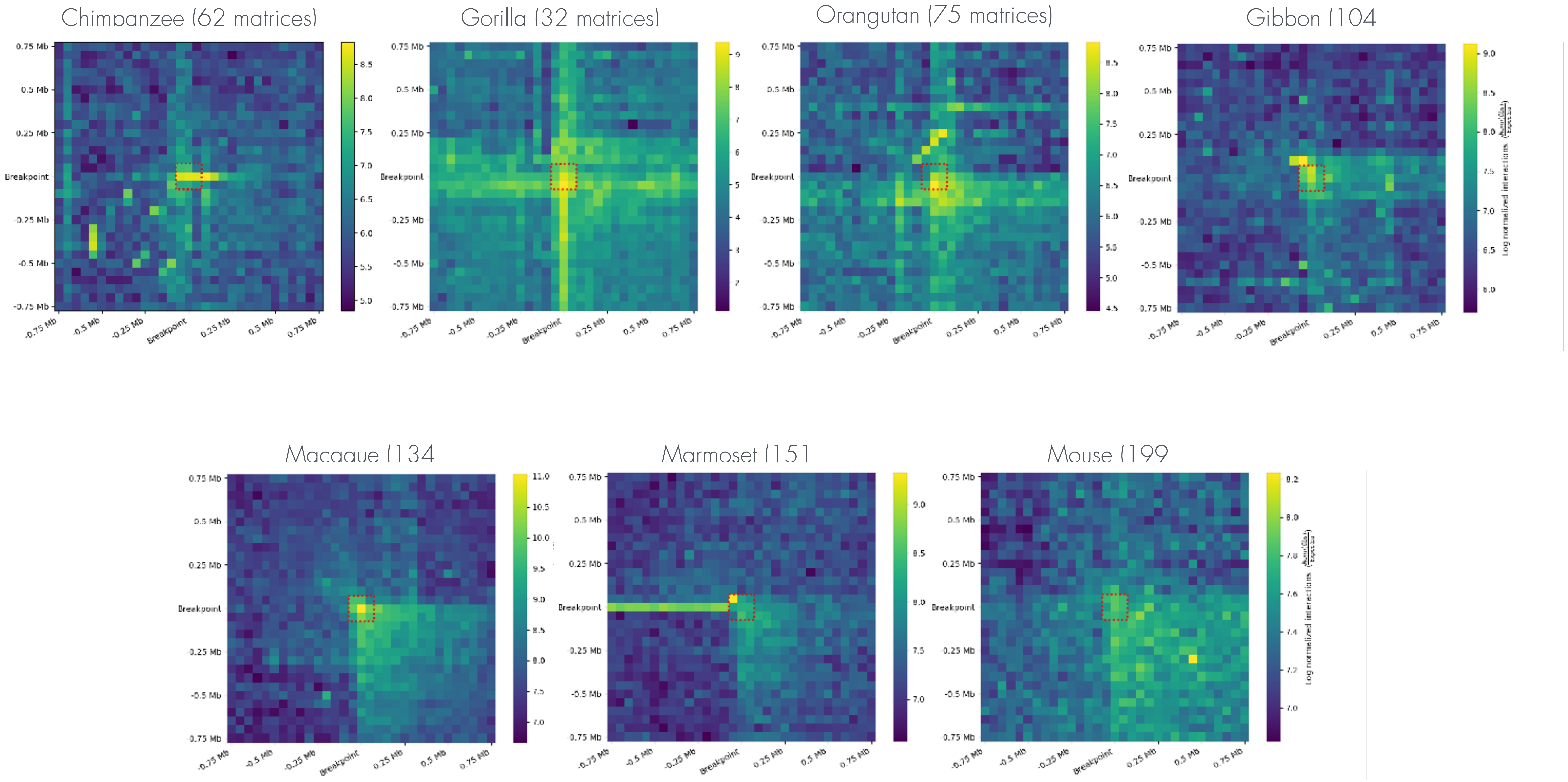




Synteny breakpoints in 3D

Two regions of at least 100 kb separated by more than 750 kb (including trans chromosomal)




Chimpanzee (62 matrices)

Gorilla (32 matrices)

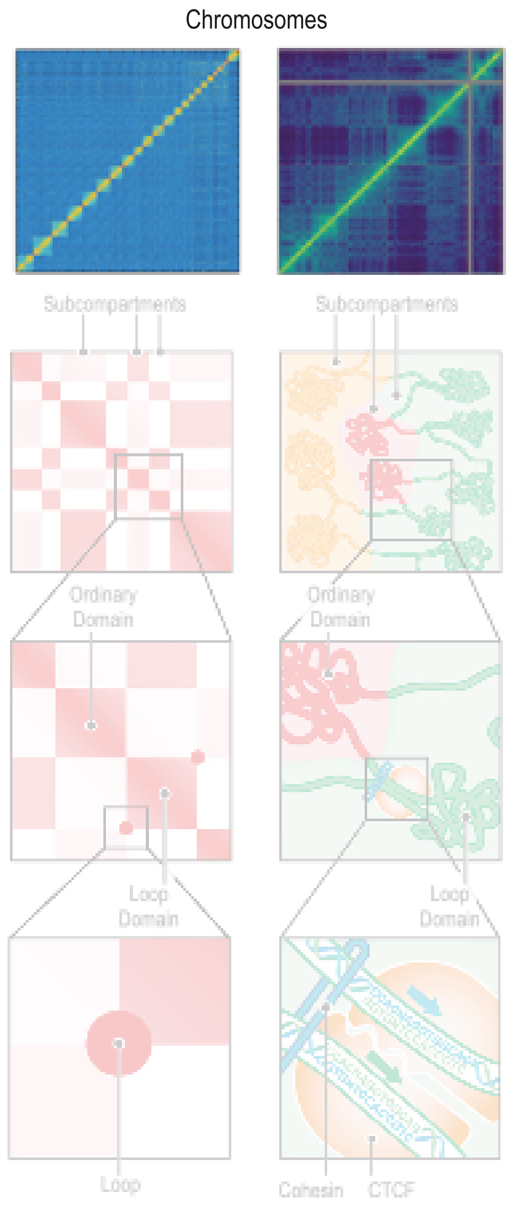
Orangutan (75 matrices)

Gibbon (104

Macaque (134

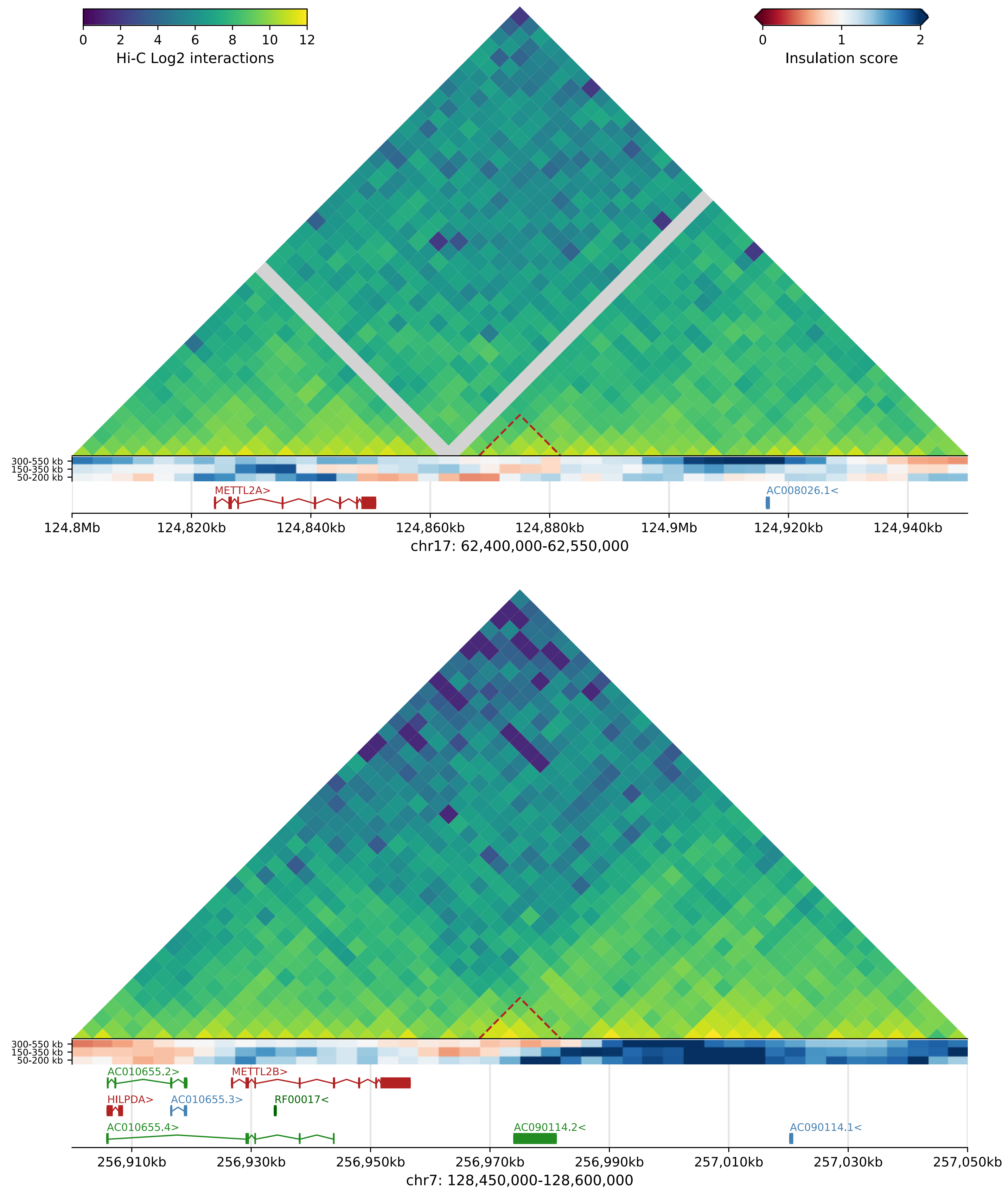
Marmoset (151

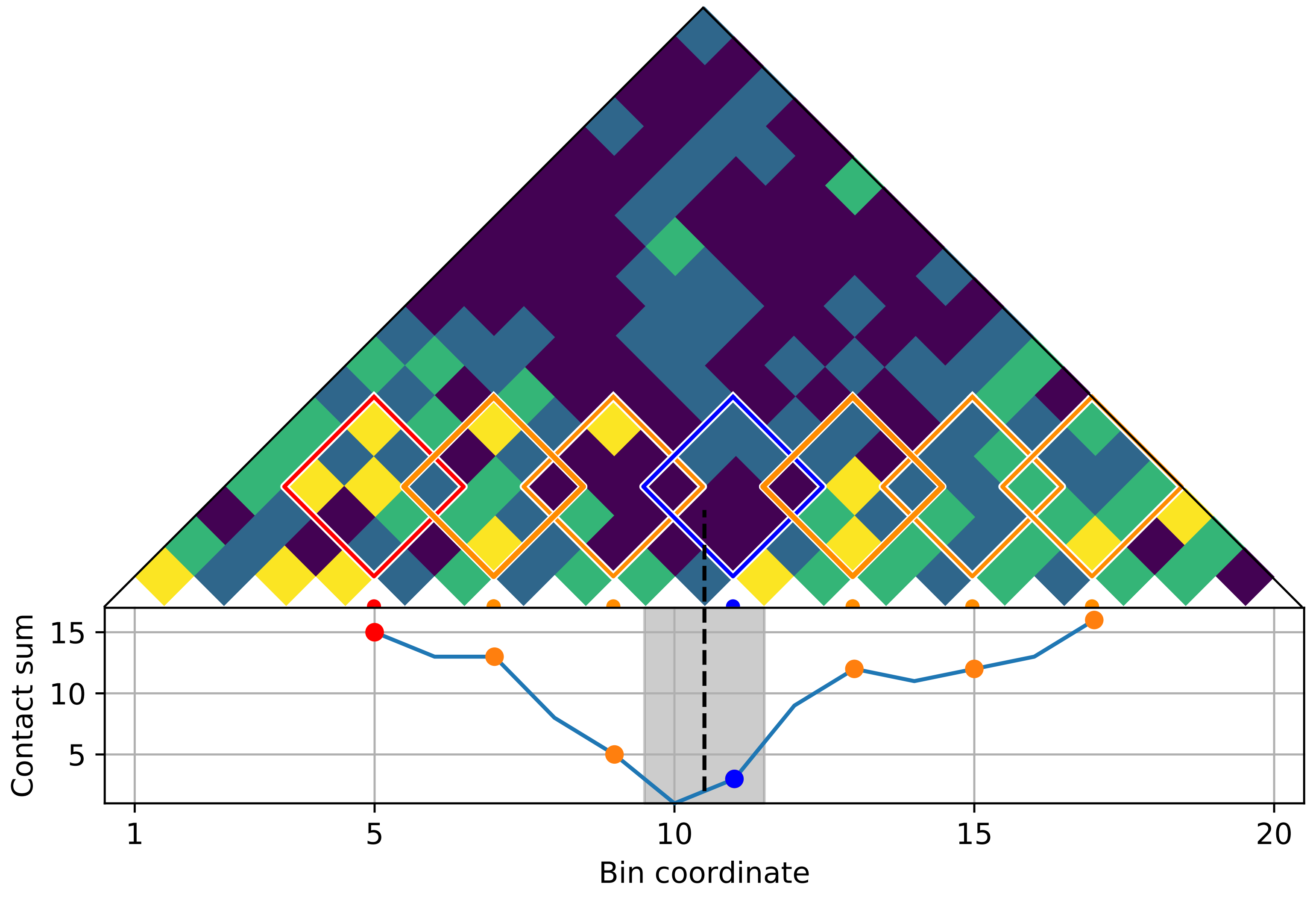
Mouse (199

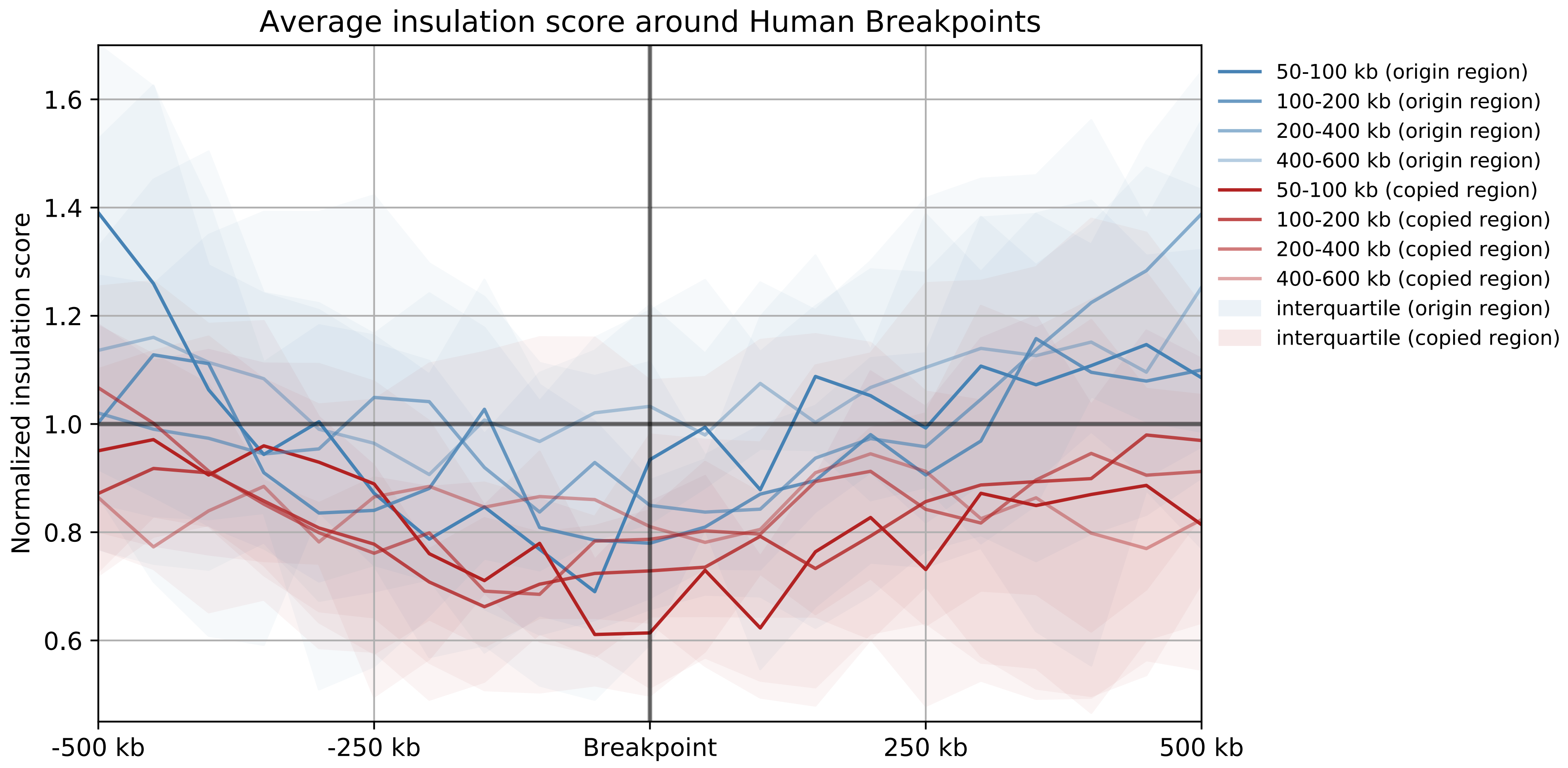


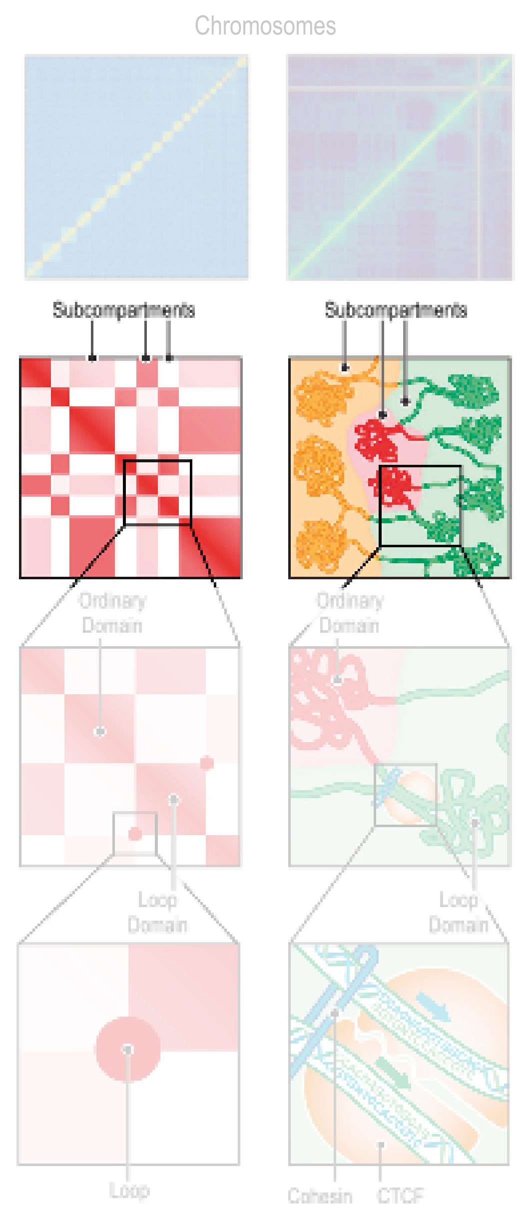
Duplications

Duplicated METTL2 locus in human with respect to primates



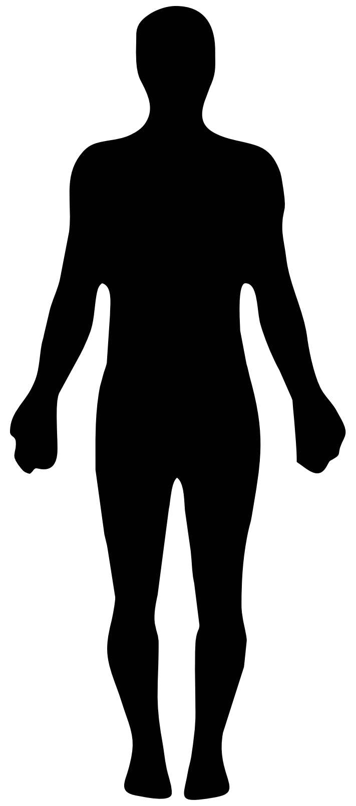


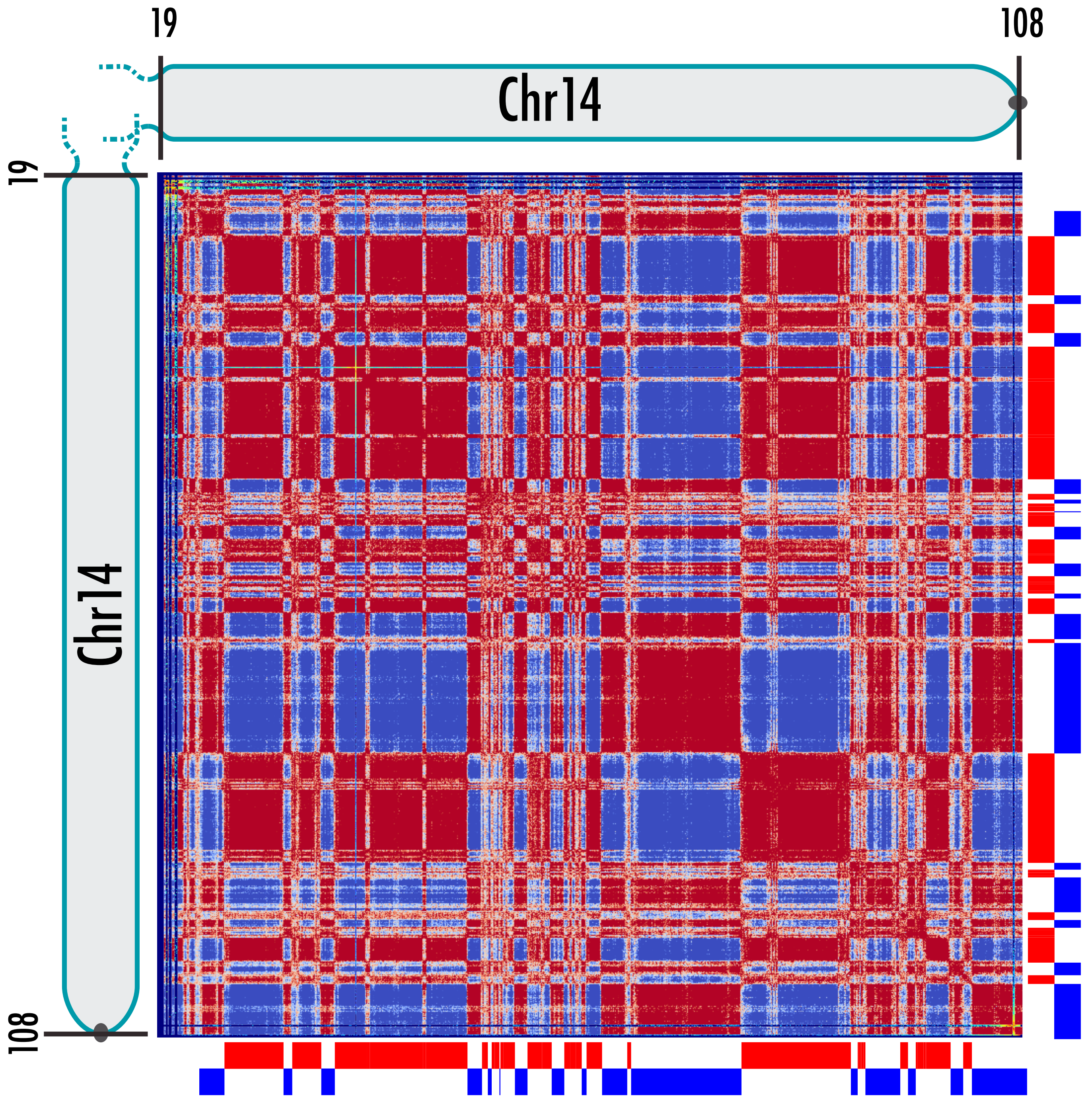


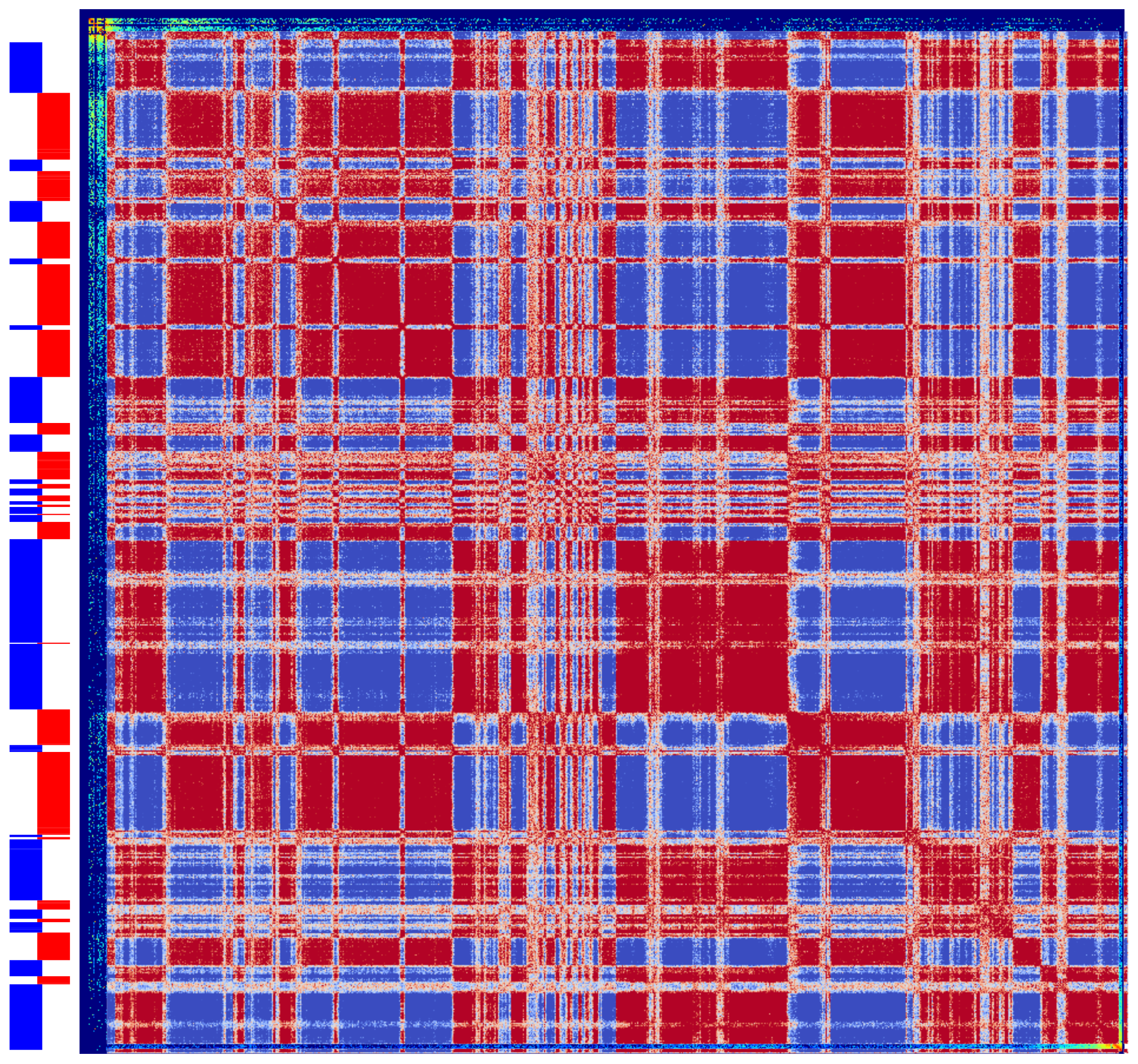


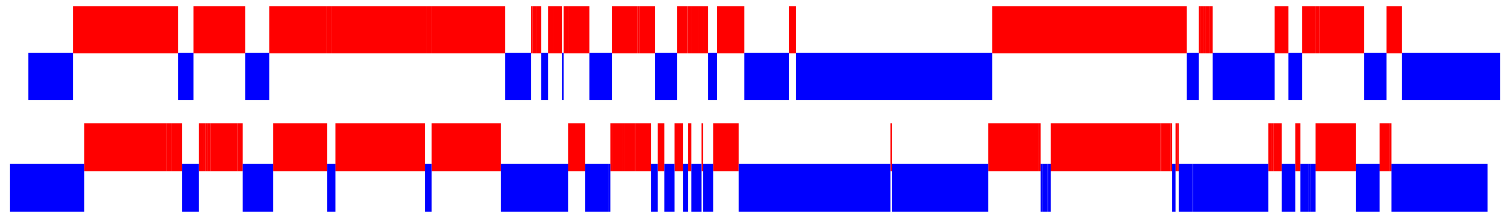
Genome compartments

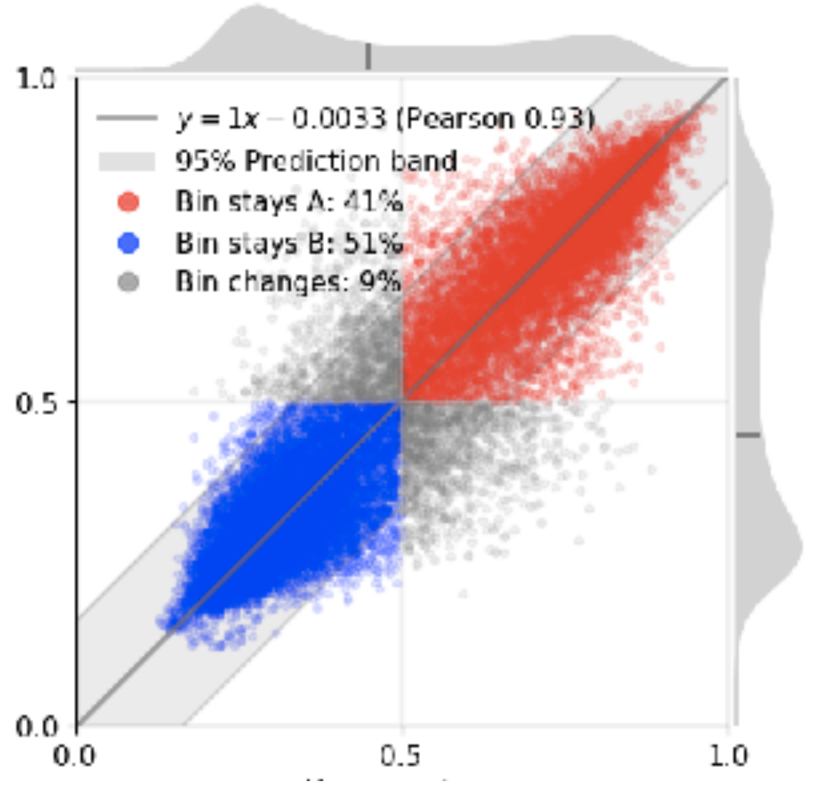
Conservation of the A/B compartments





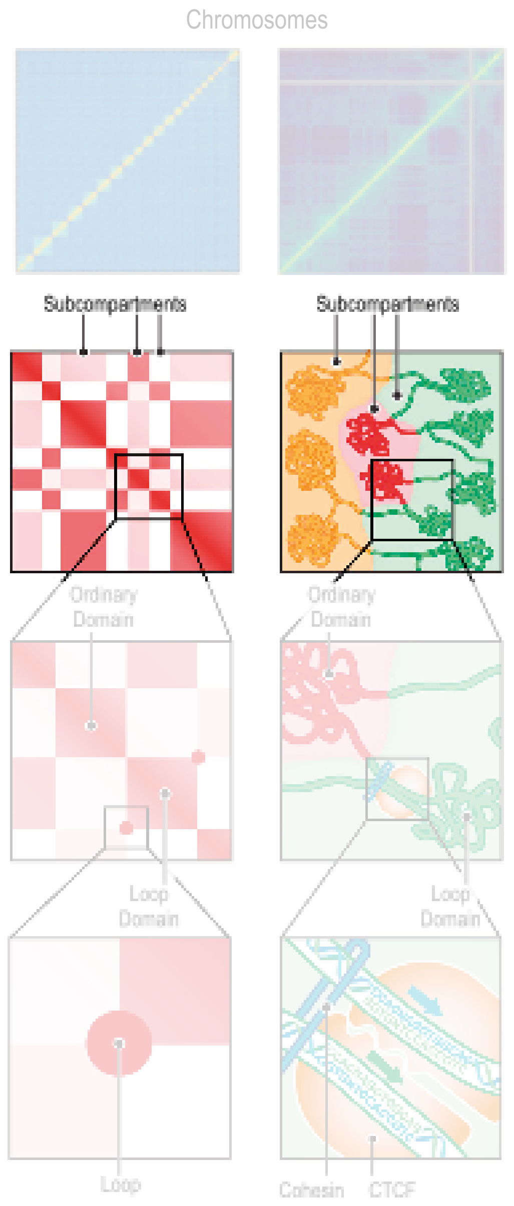


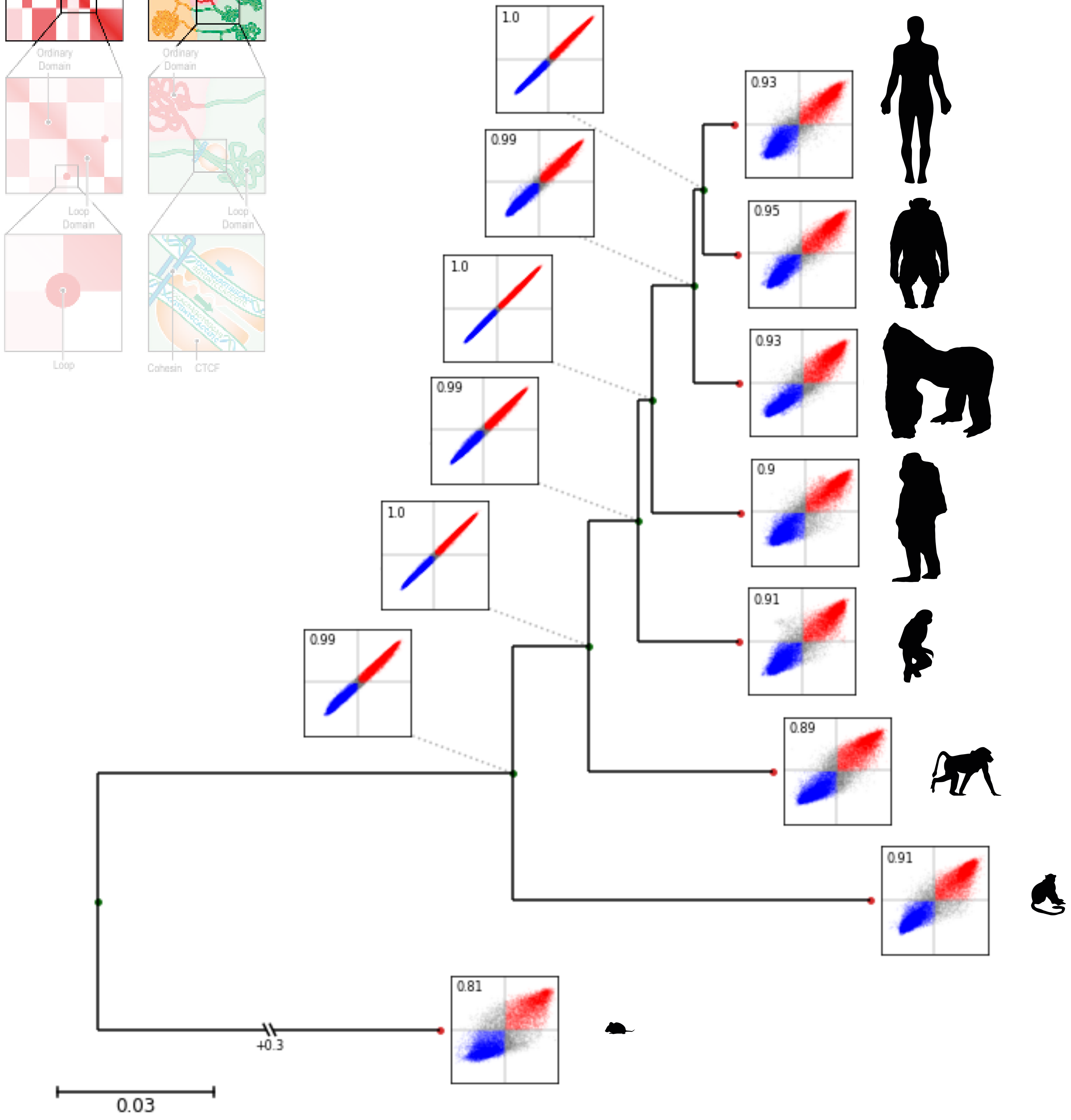


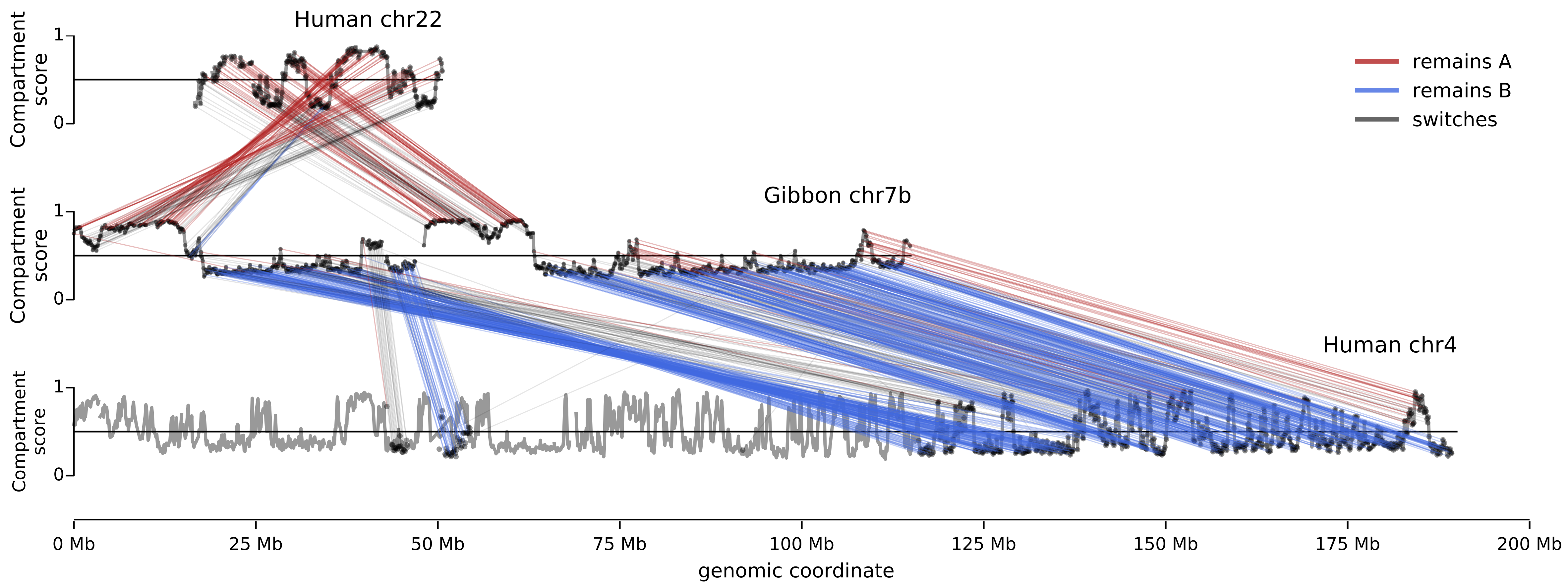


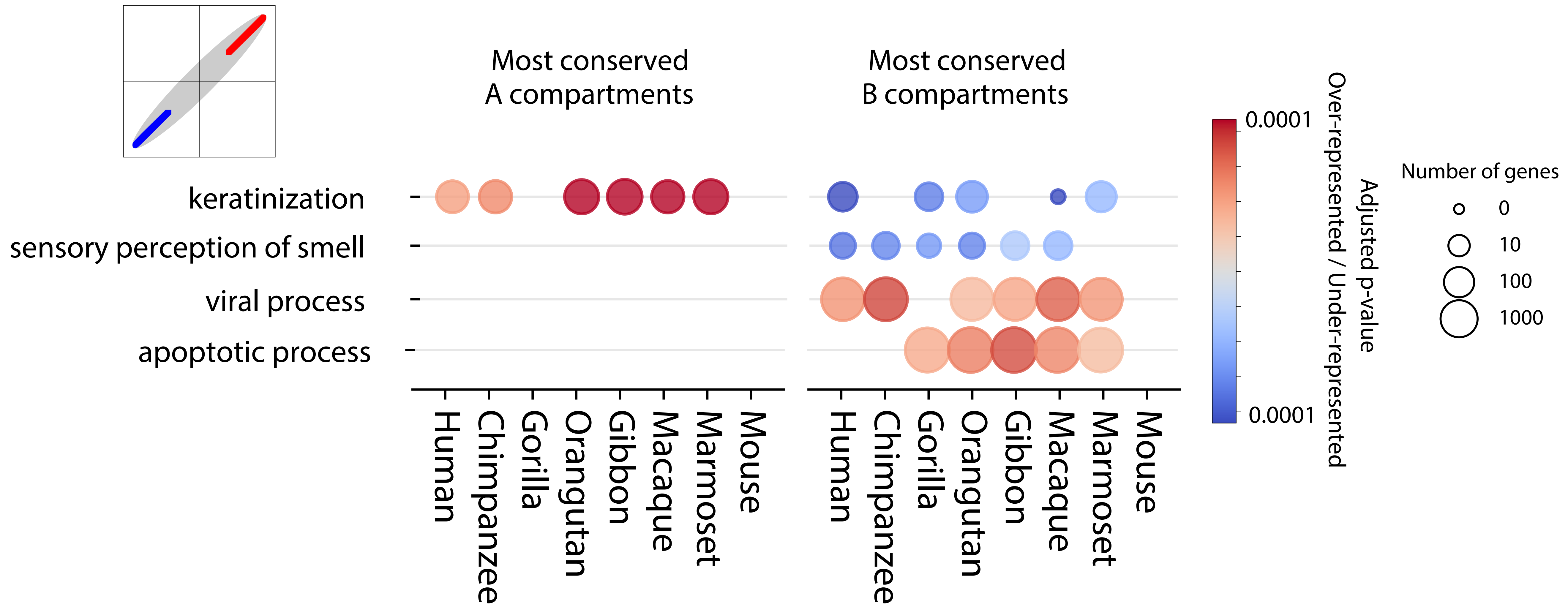
Genome compartments

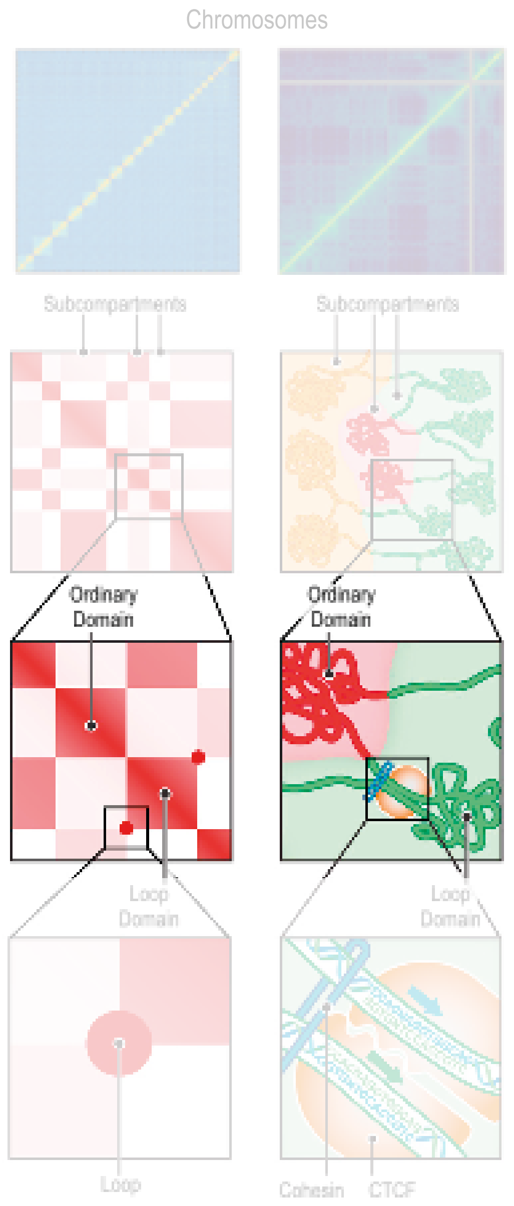
Conservation of the A/B compartments





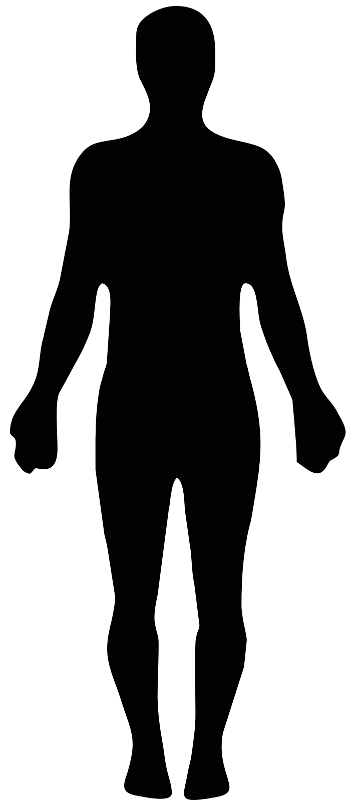


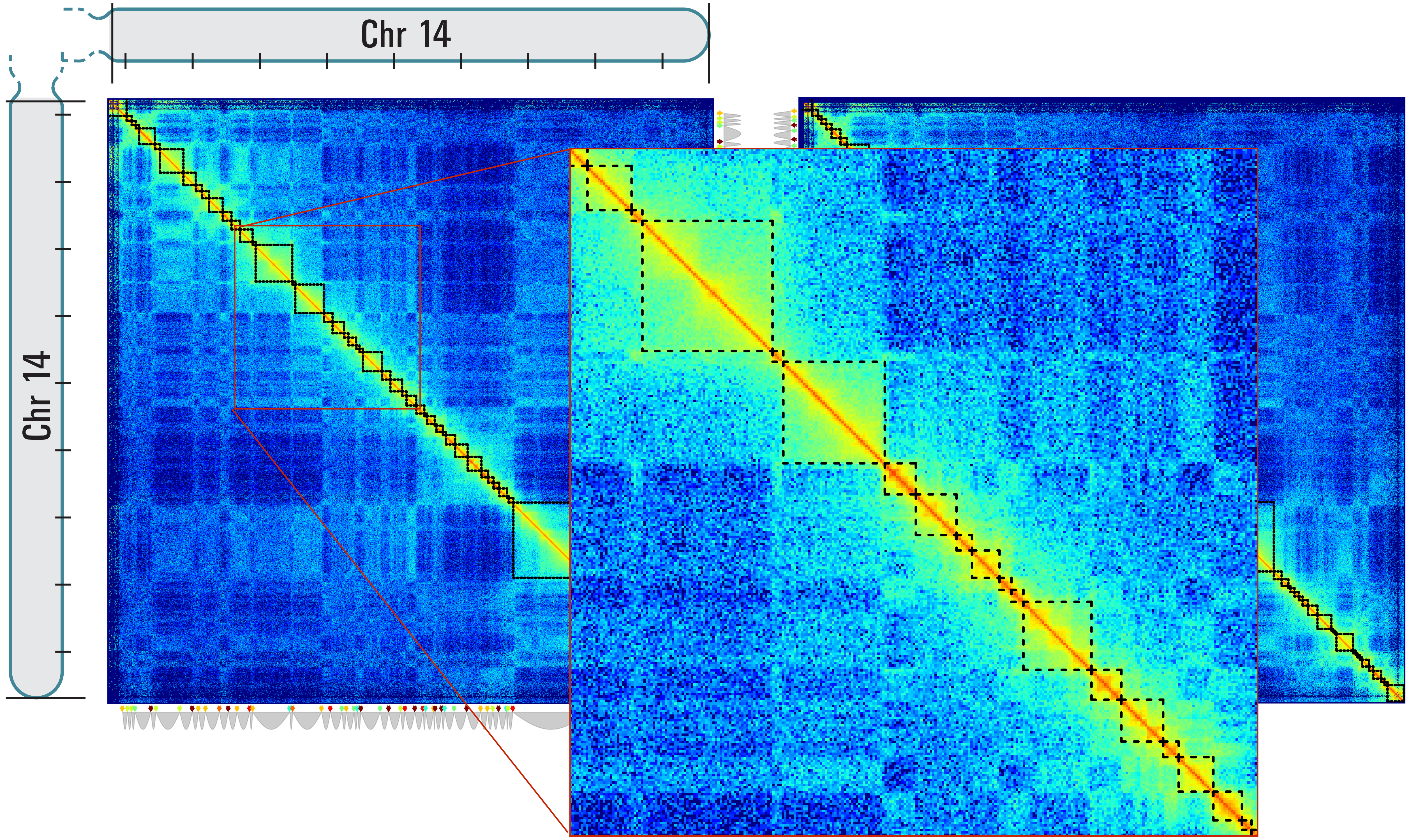


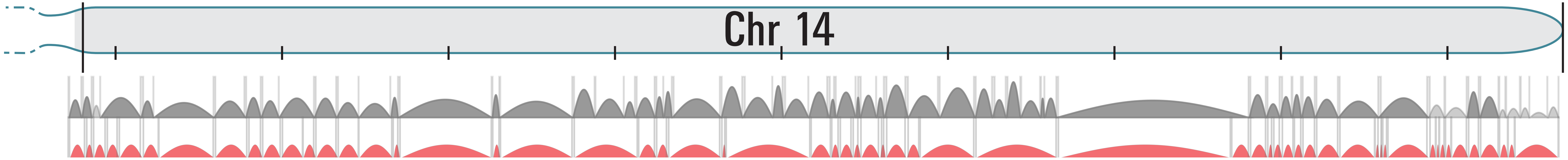


Genome Topologically Associating Domains

Conservation of TADs

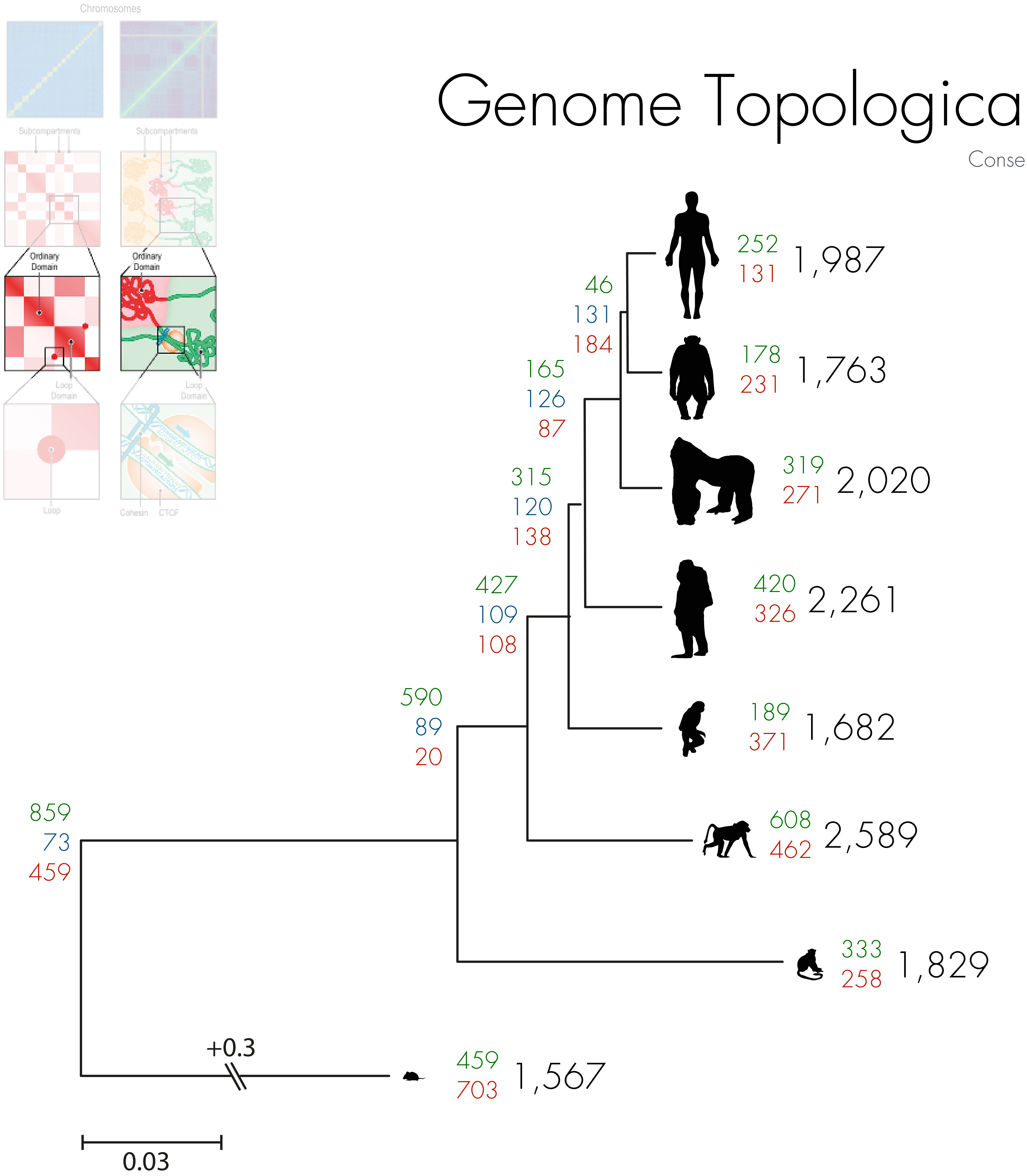


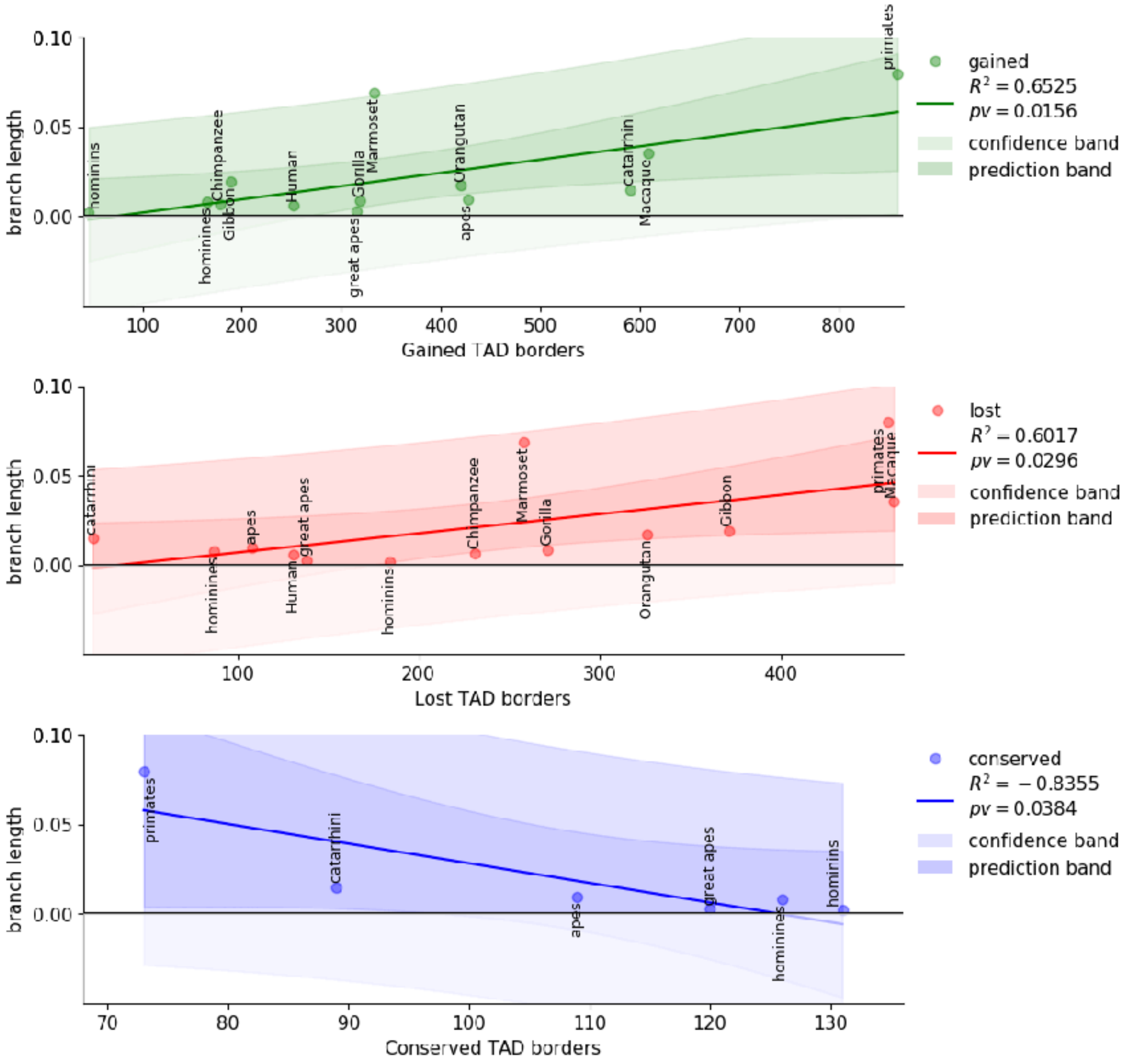


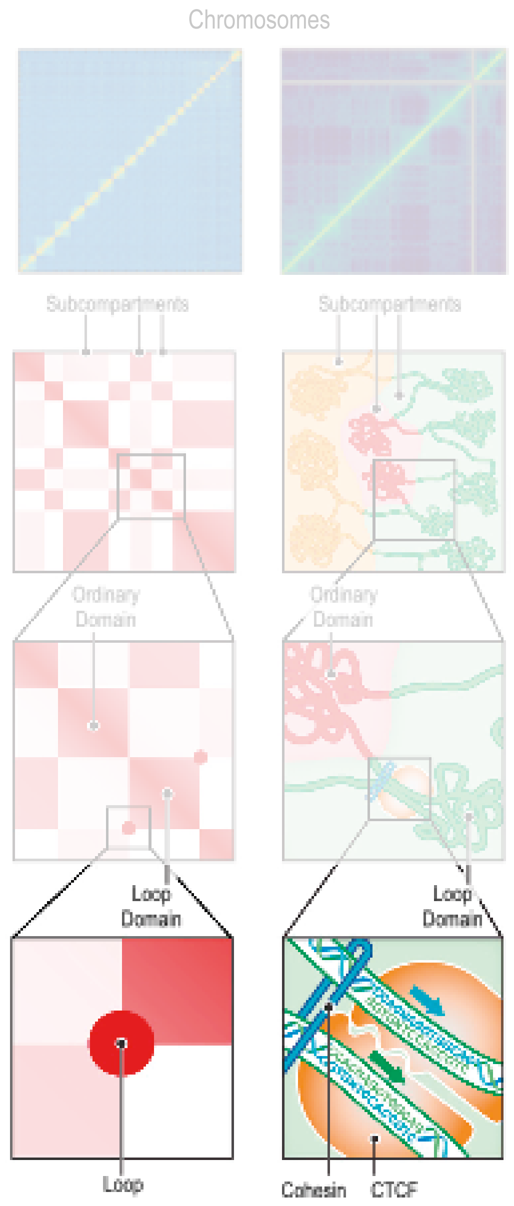


Genome Topologically Associating Domains

Conservation of TADs

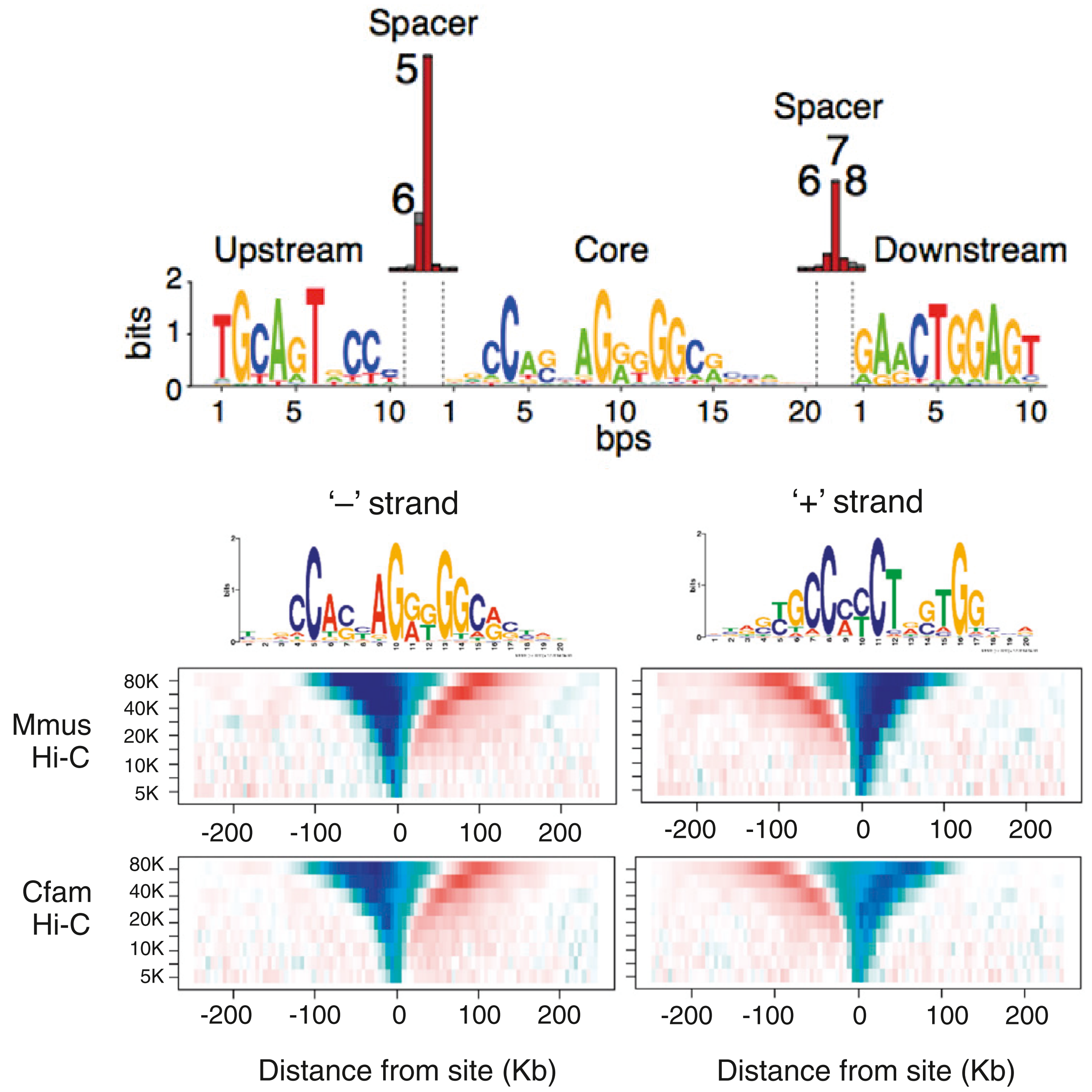






Loops

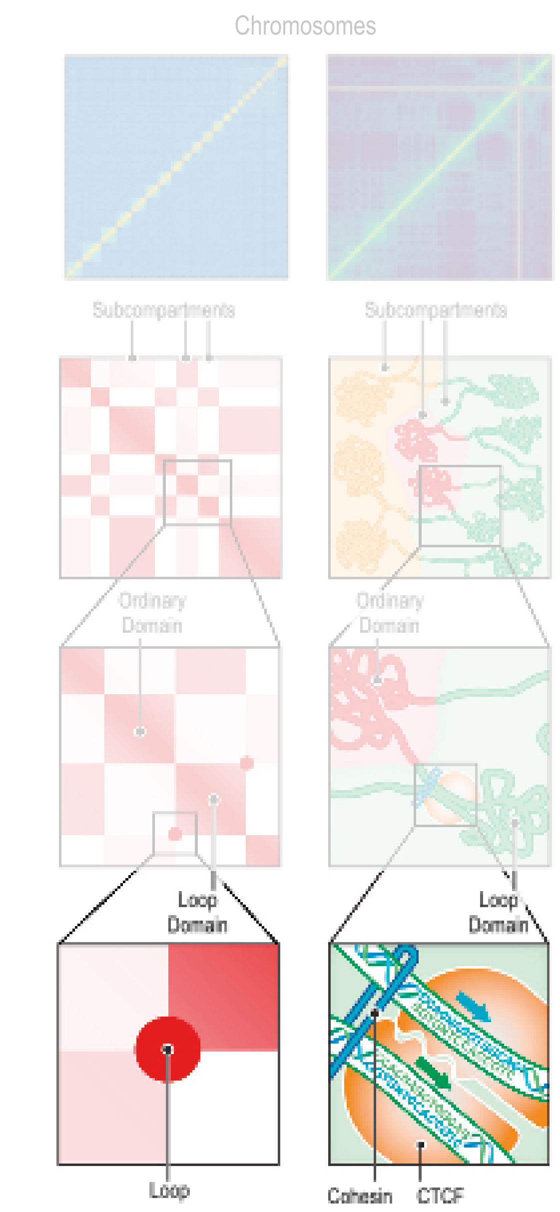
Conservation of CTCF sites

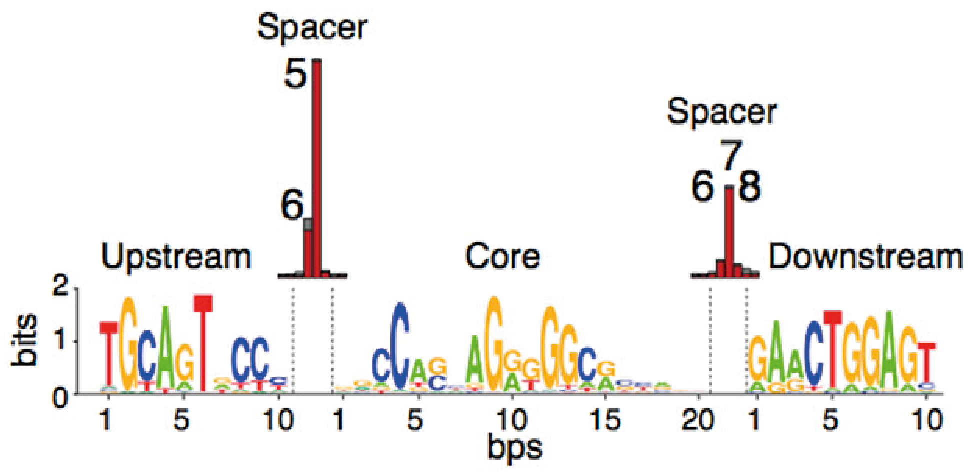


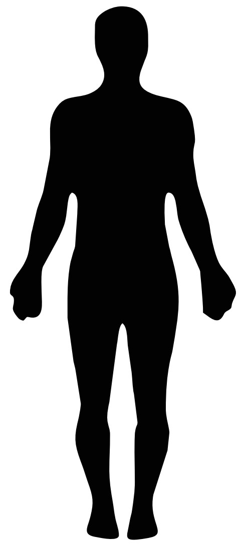
Vietri Rudan, et al. Cell Rep. 2015 Mar 03; 10(8) 1297-1309
Nakahashi et al. Cell Rep. 2013 May 30; 3(5) 1678-1689

Loops

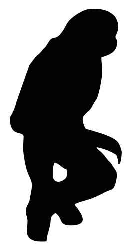
Conservation of CTCF sites





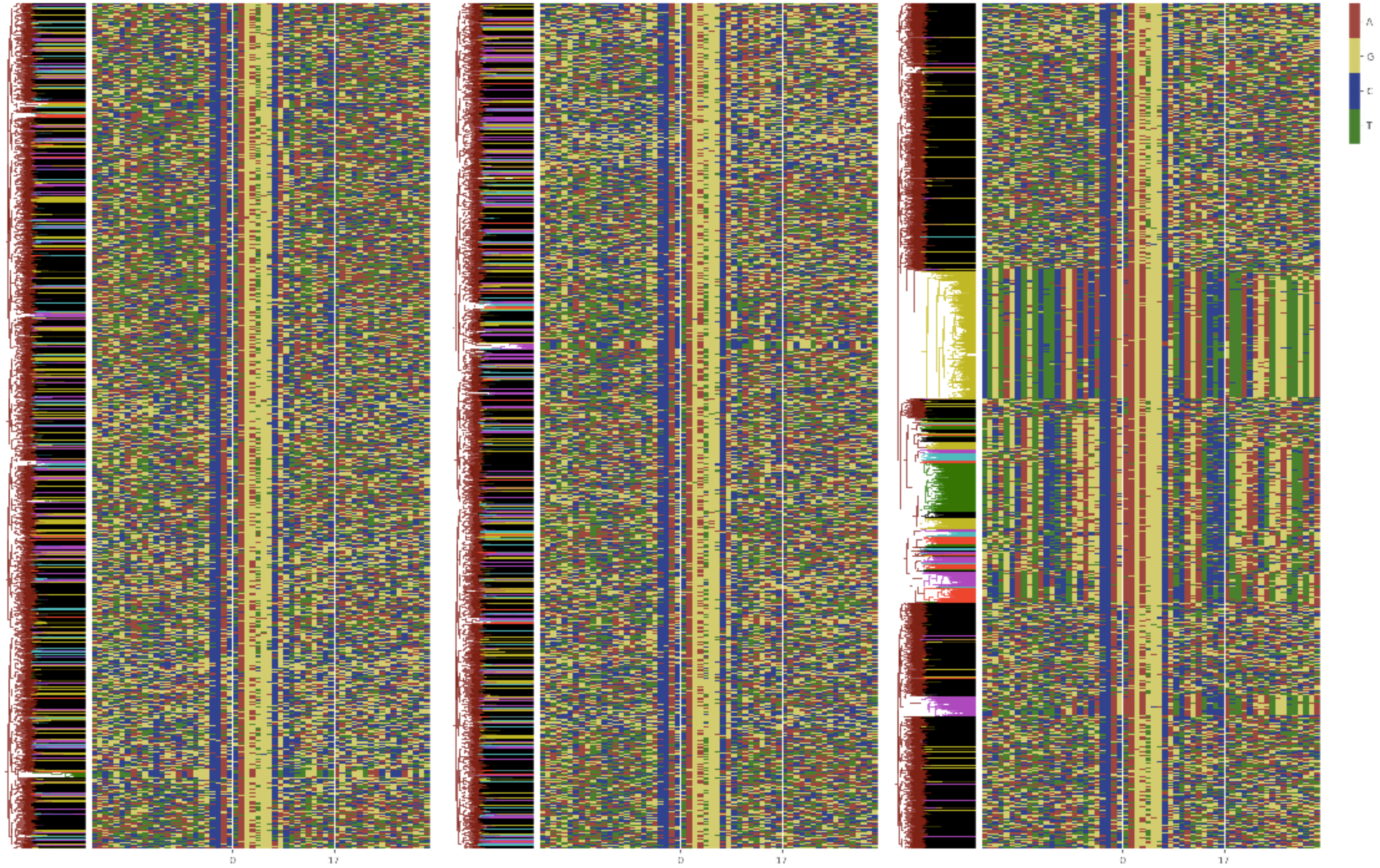


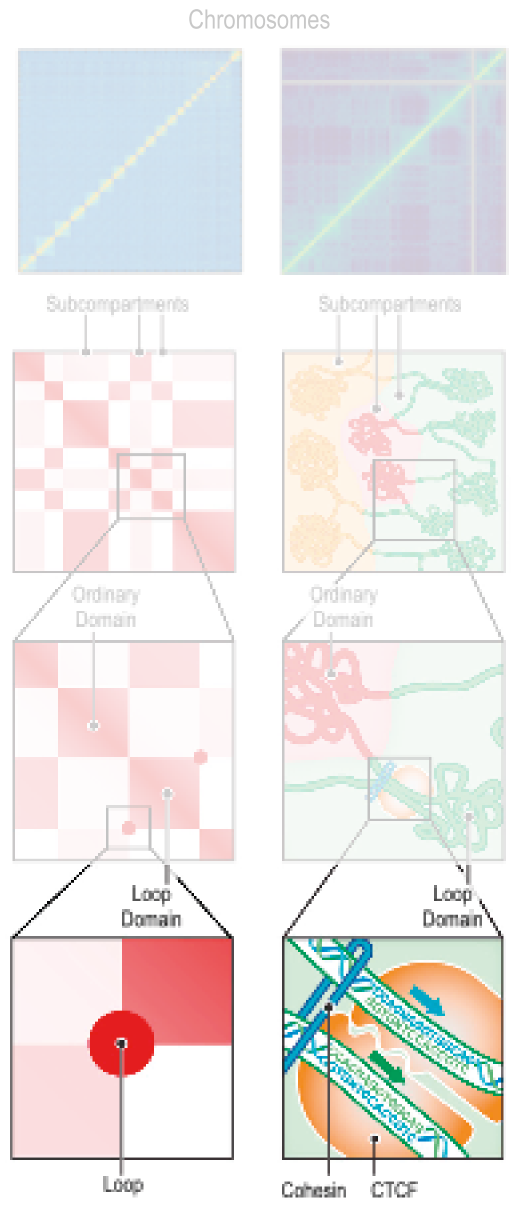
Human



Gibbon

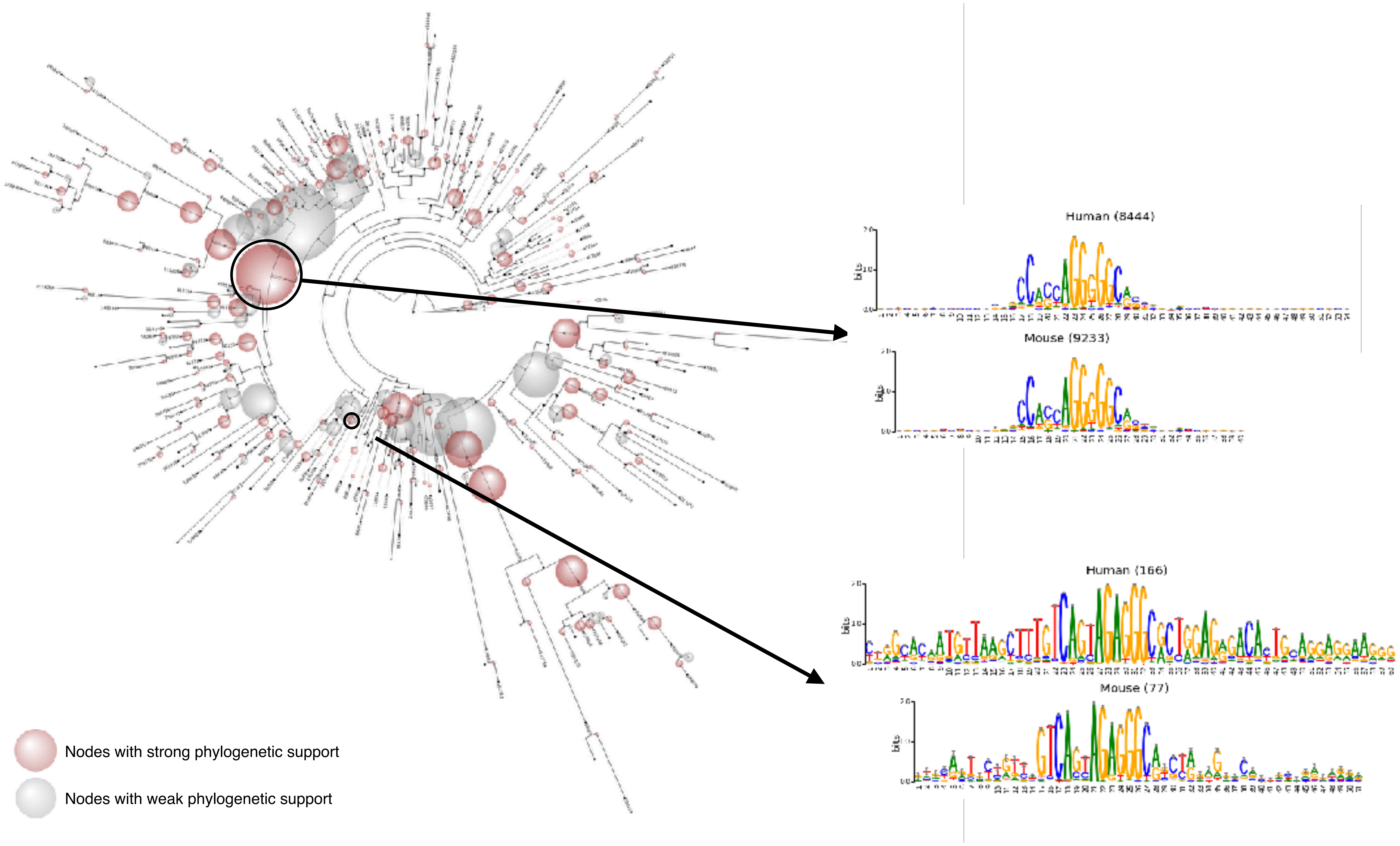
Mouse





Loops

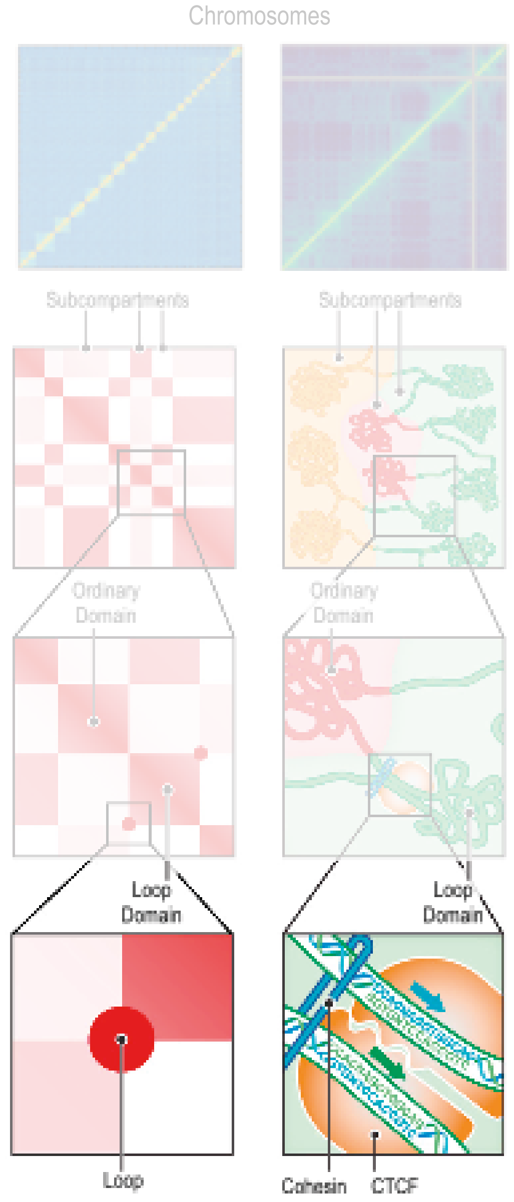
Conservation of CTCF sites

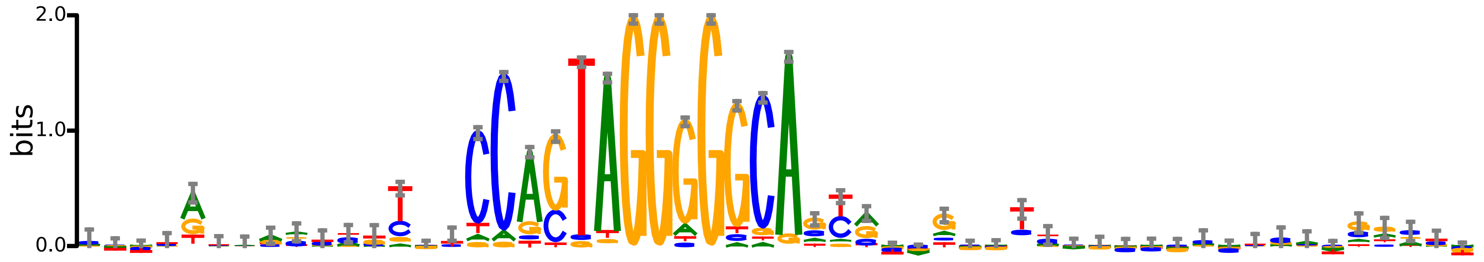


- >150 K sequences to align (all species together)
- Tree reconstruction using FastTree2
- Reduce the complexity filtering by node support
- Selection of 173 nodes with 0.9 bootstrap
- At least 100 CTCF sites (and at least 50 from single species)

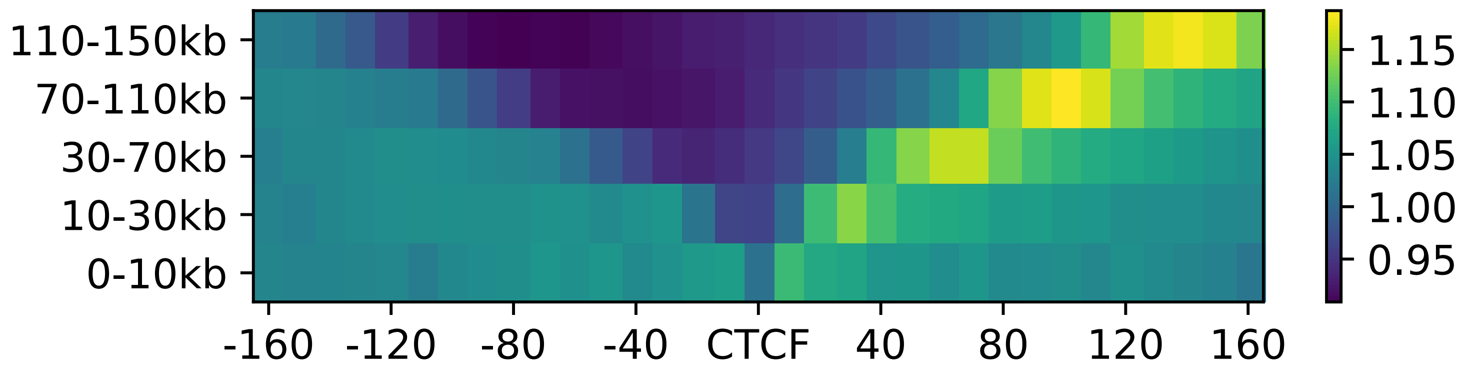
Loops

Conservation of CTCF sites

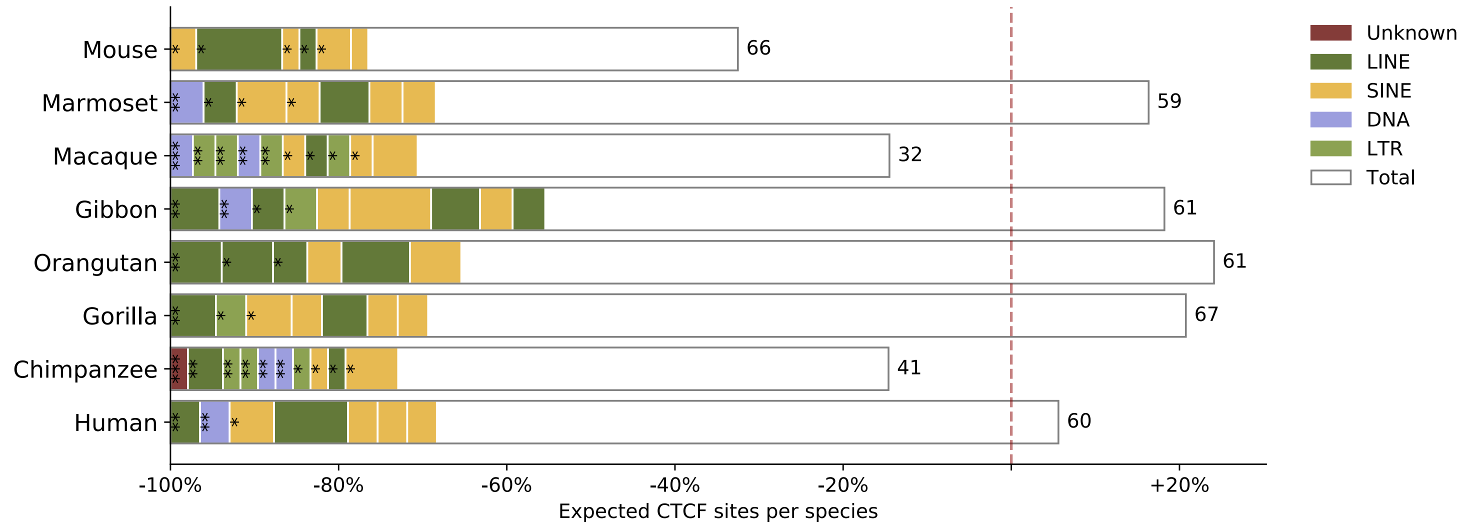




- Motif (nucleotide content)



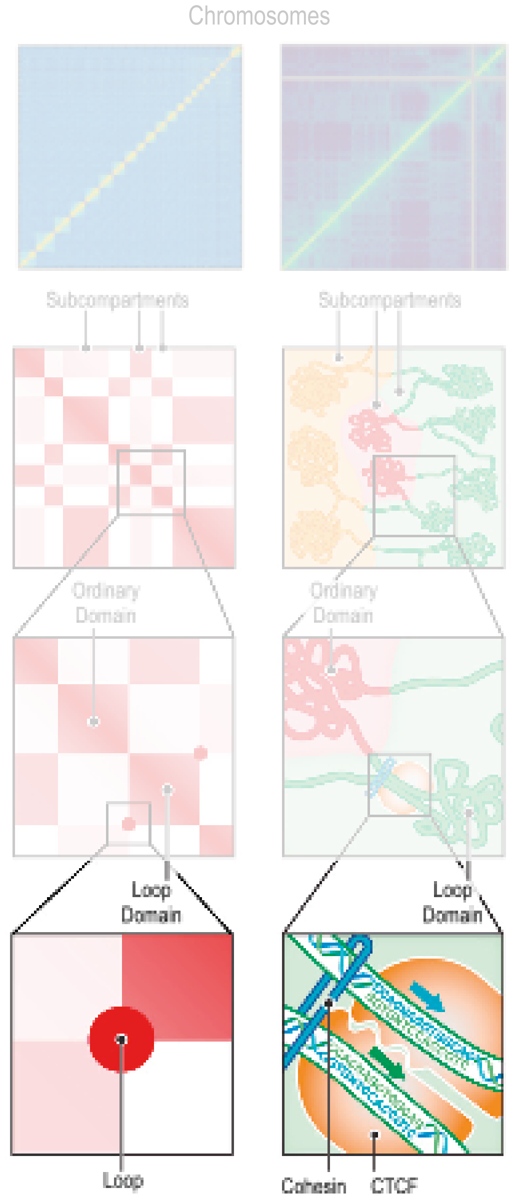
- Insulation/looping (interaction directionality)

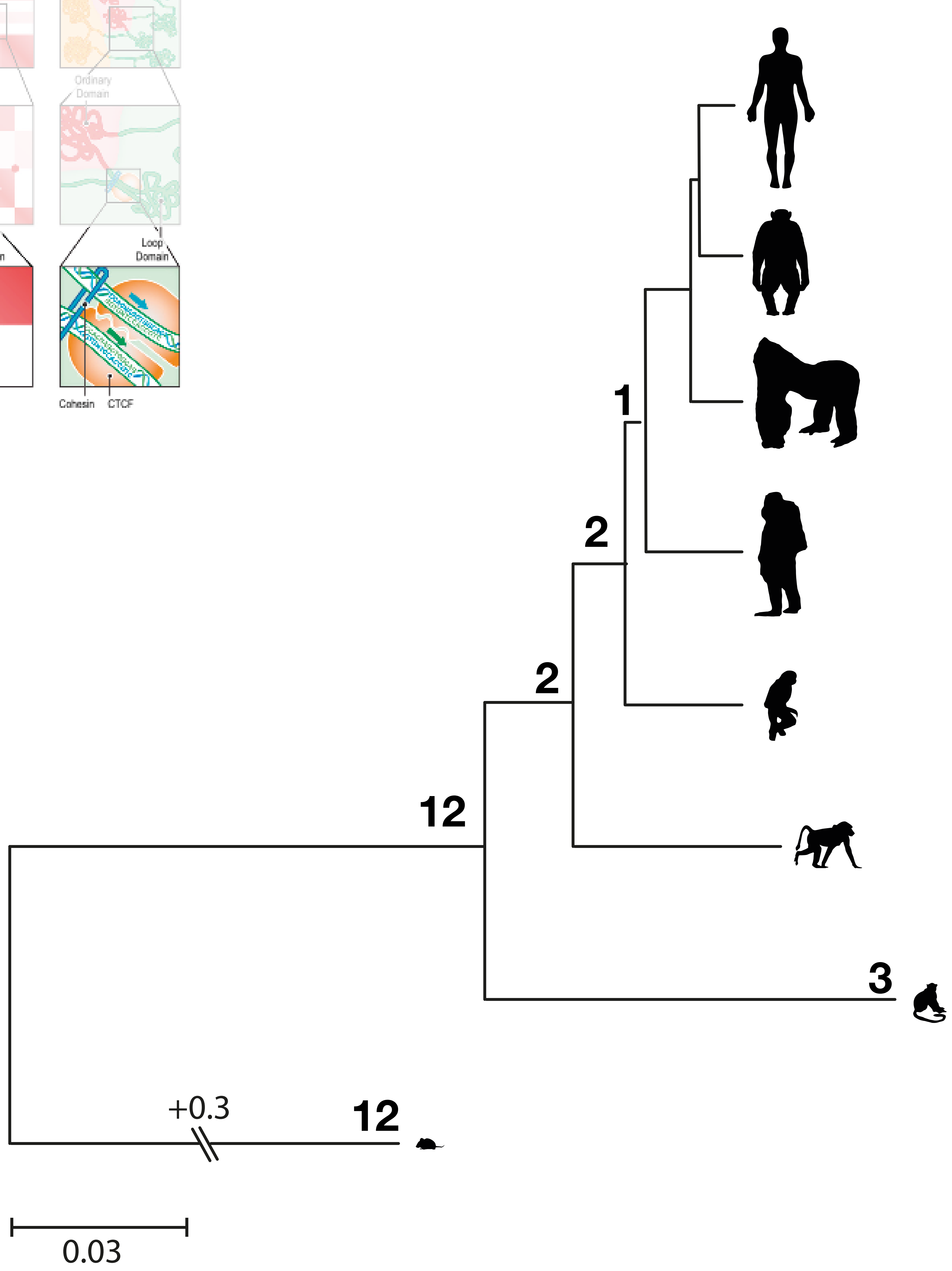


- Enrichment in repetitive elements

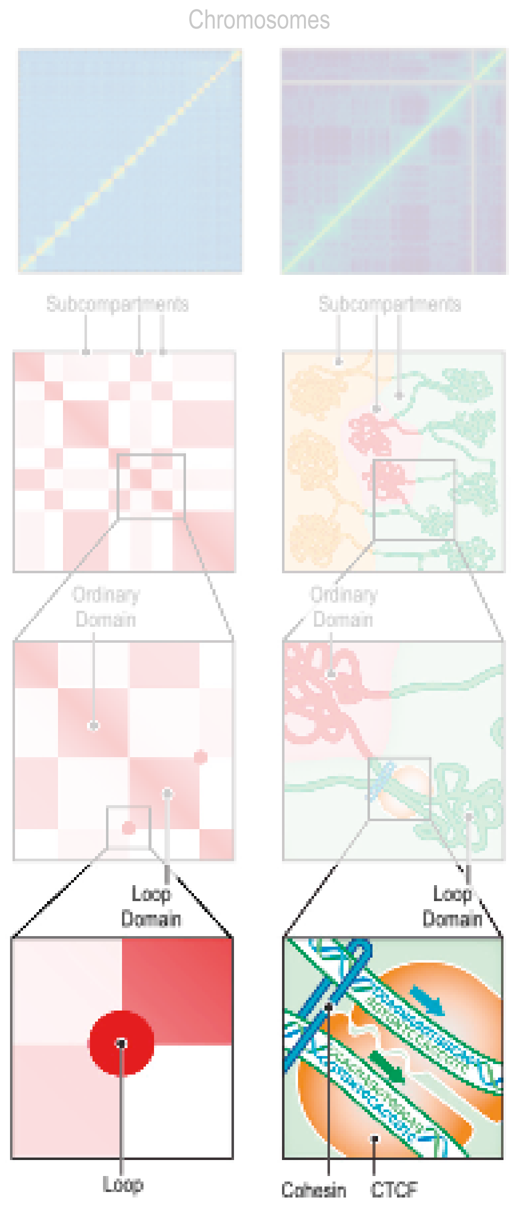
Loops

Conservation of CTCF sites



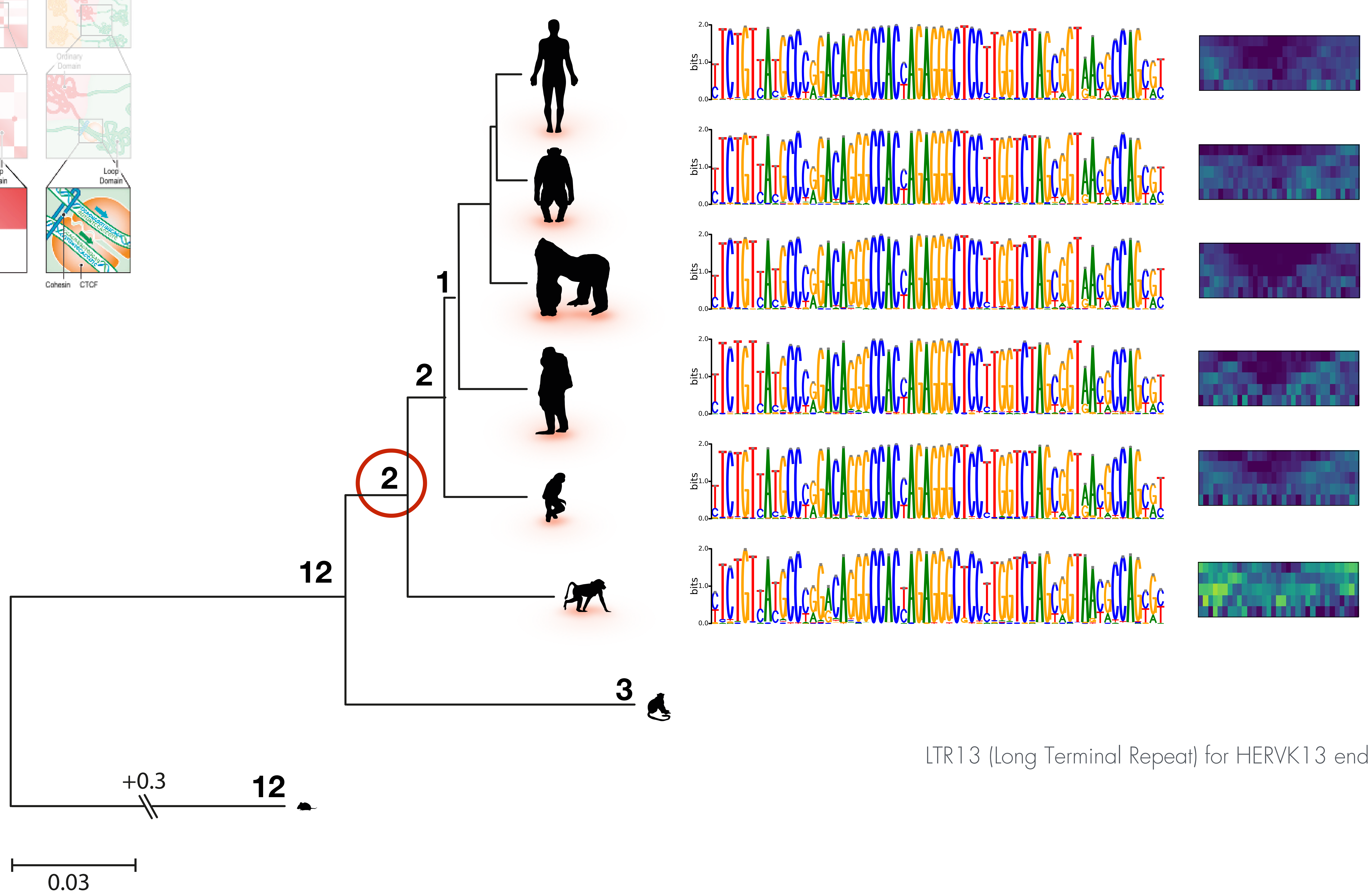


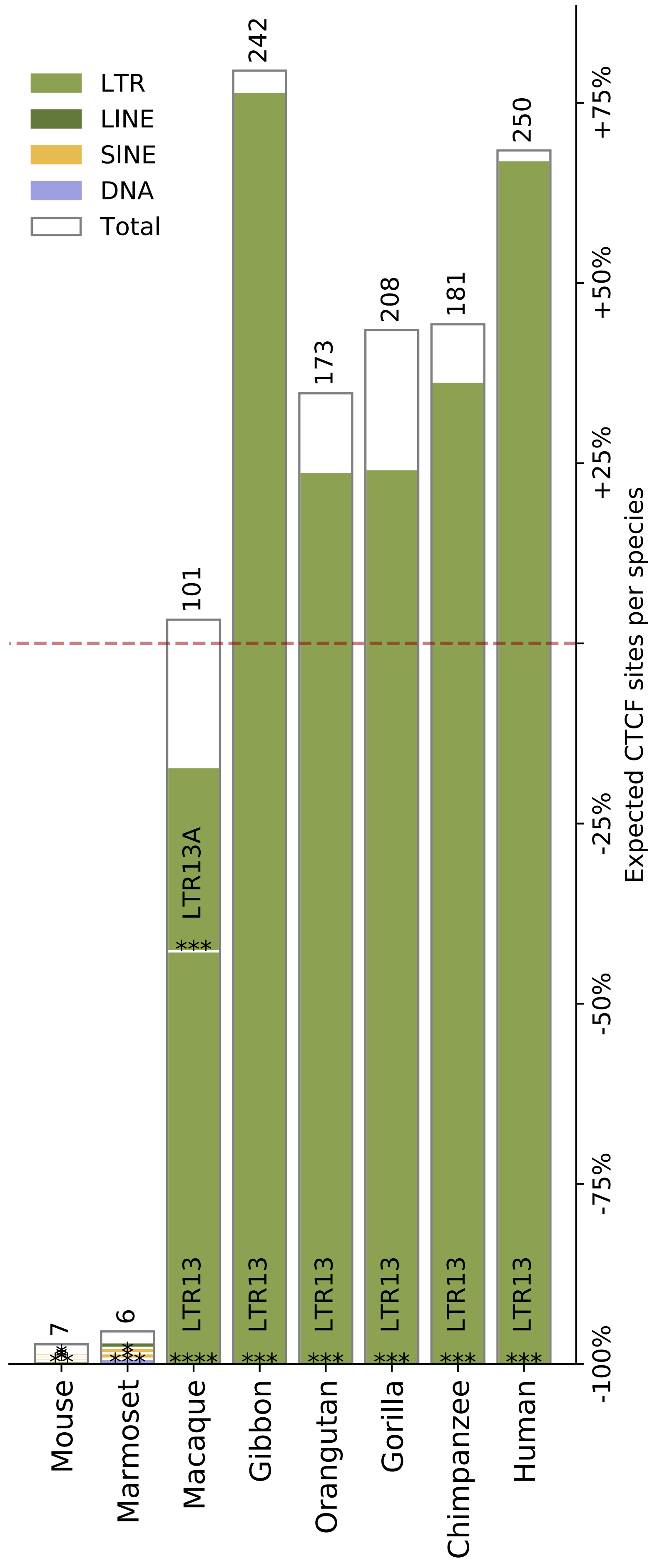
Few events of genome expansion through transposons involving CTCF sites



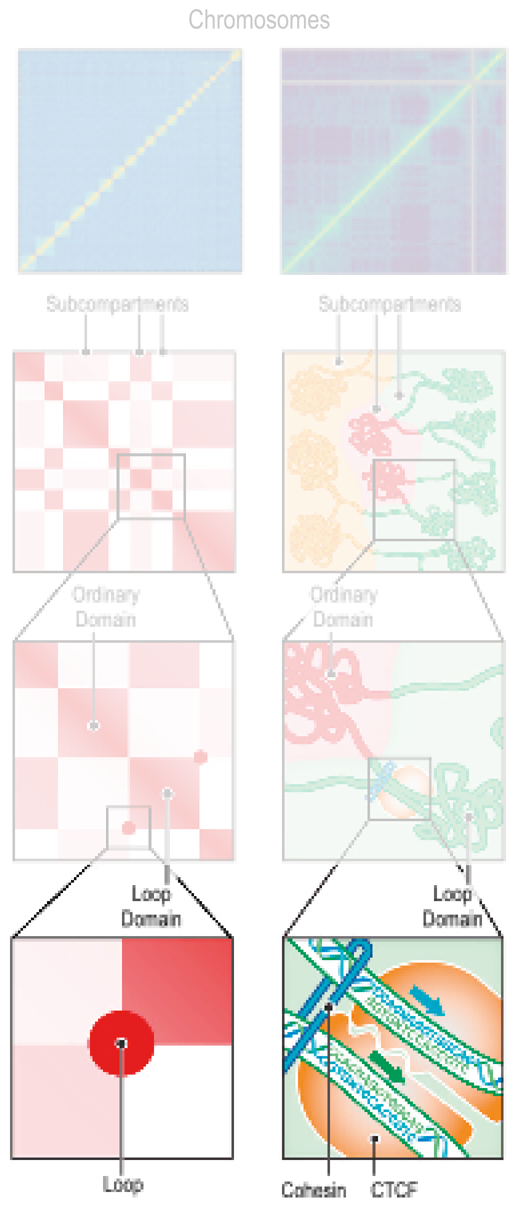
Loops

Conservation of CTCF sites



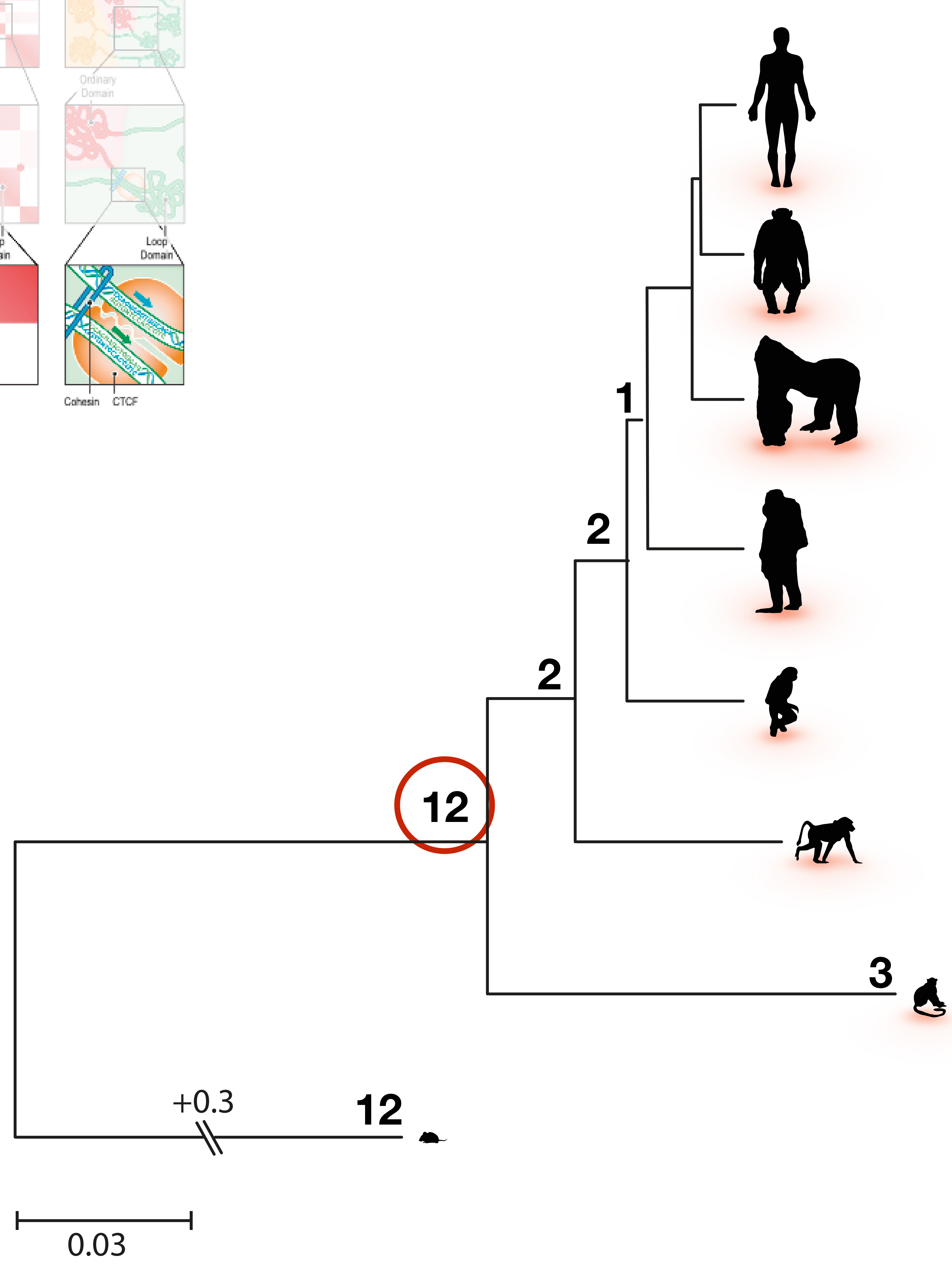


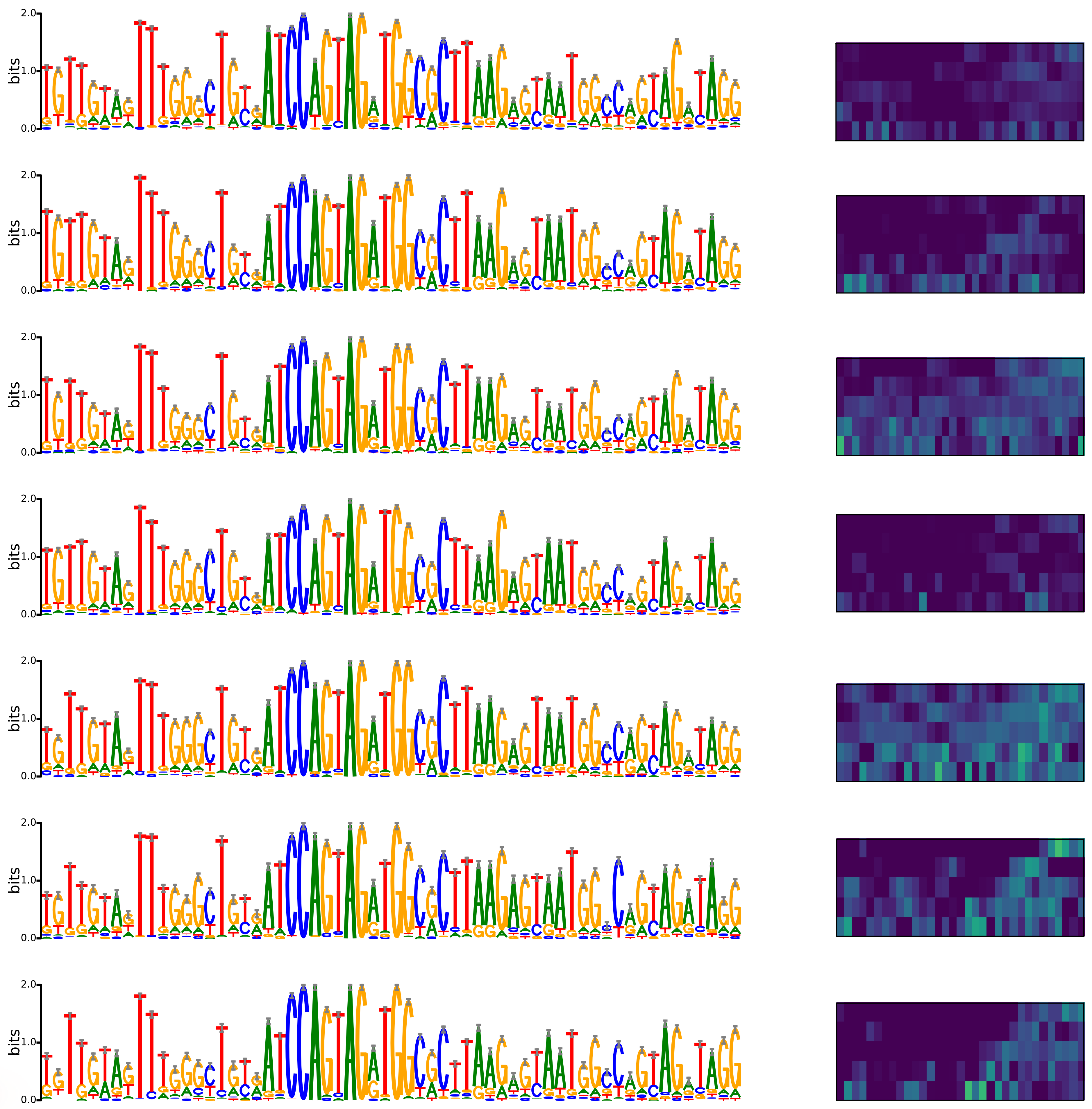
LTR13 (Long Terminal Repeat) for HERVK13 endogenous retrovirus

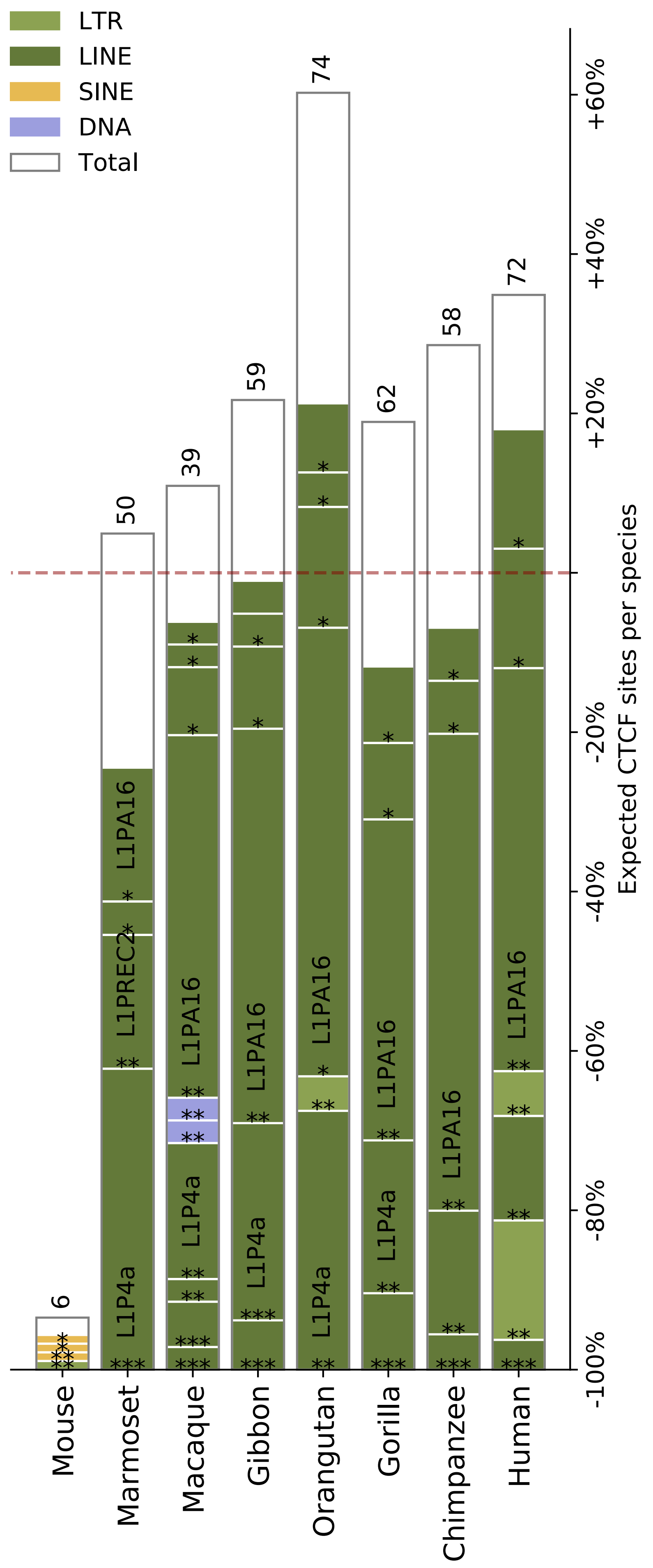


Loops

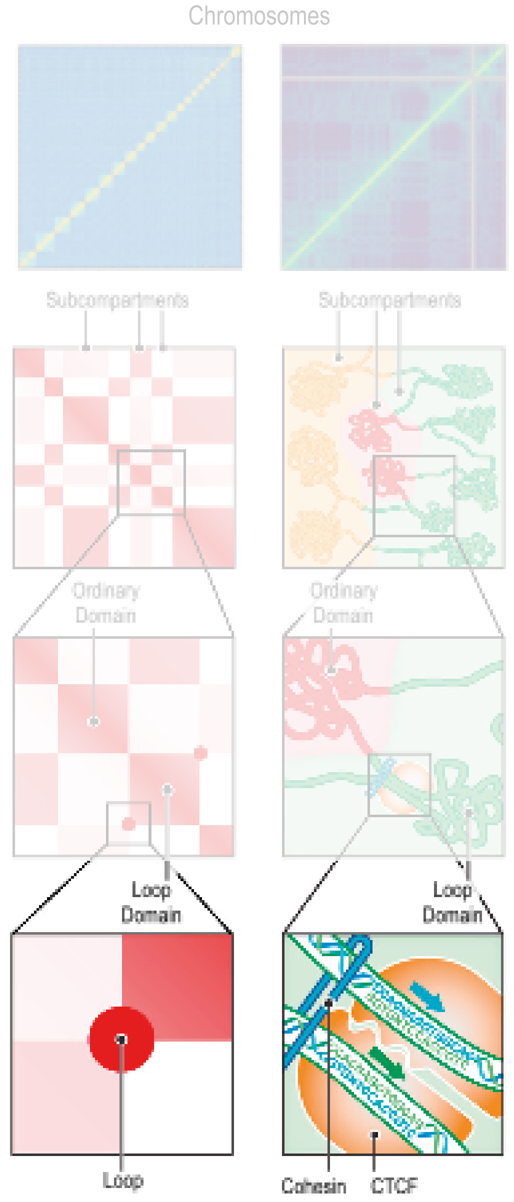
Conservation of CTCF sites





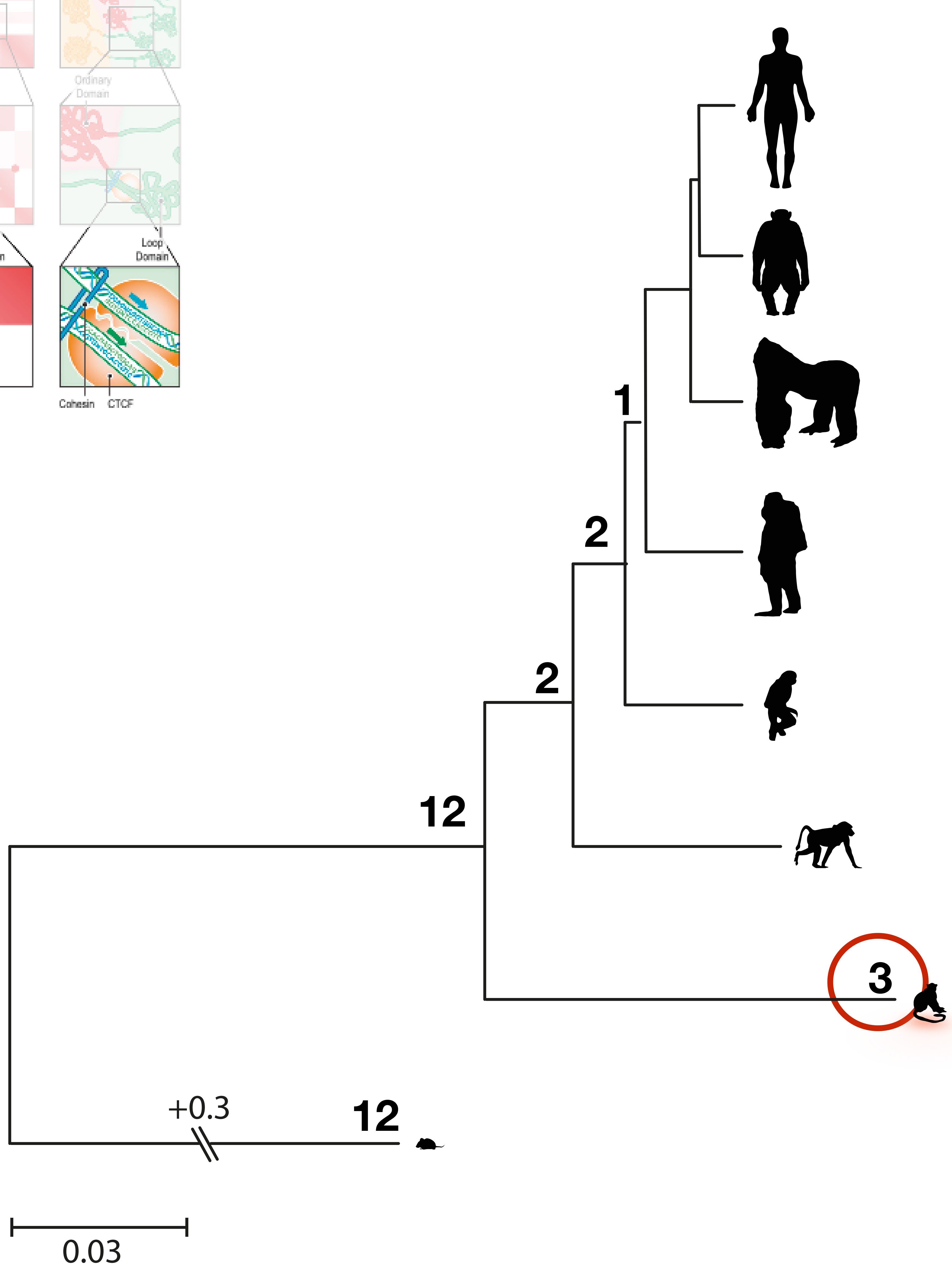


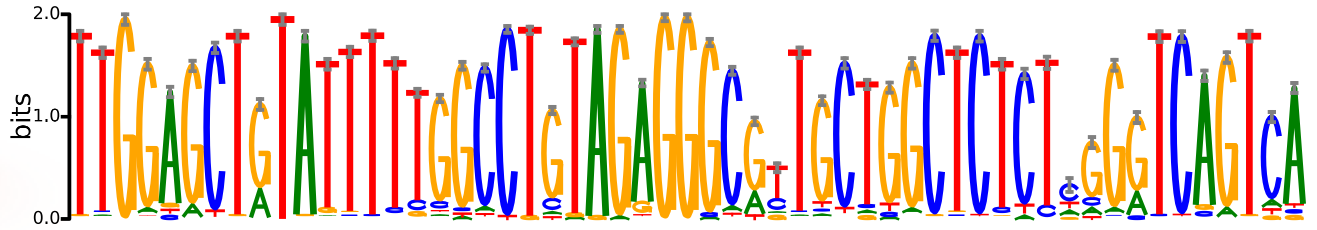
3' end of L1 retrotransposon, L1PA16_3end subfamily

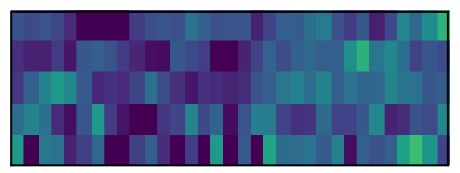


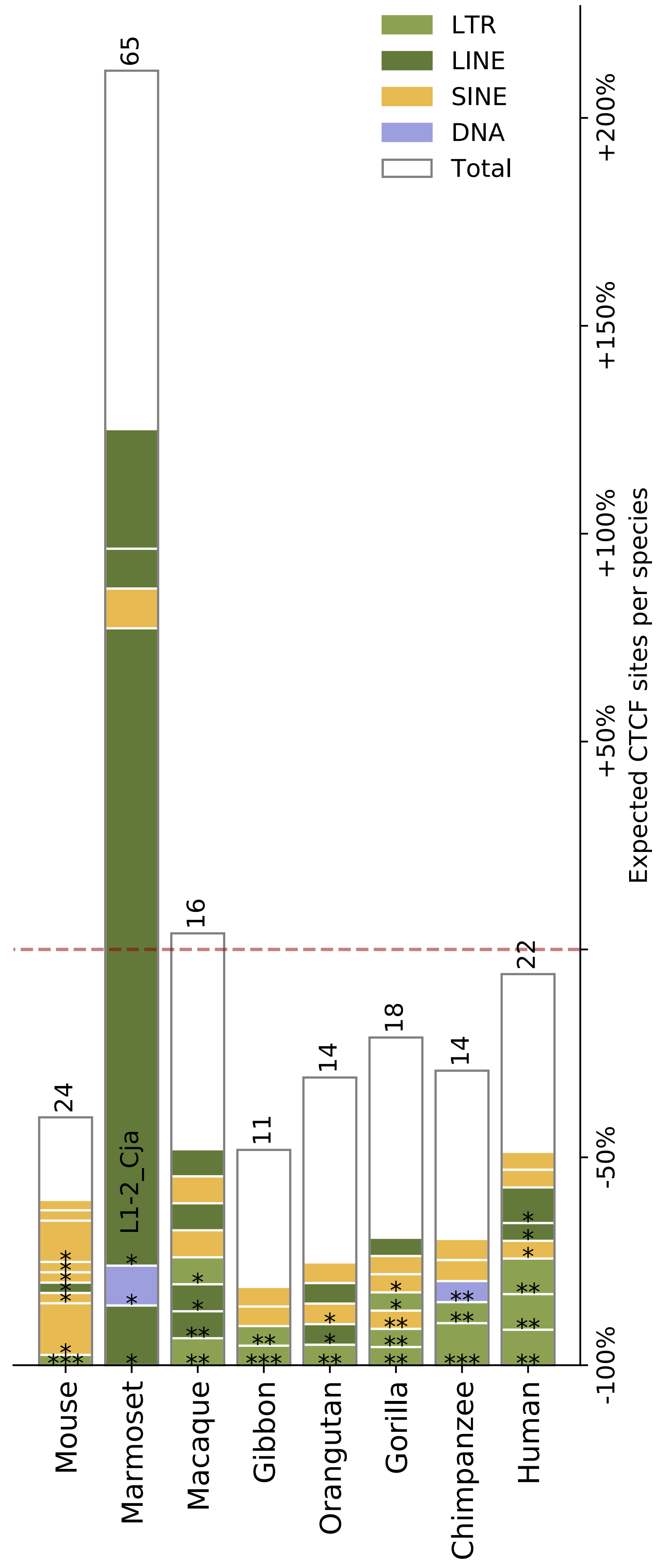
Loops

Conservation of CTCF sites









LINE L1 specific of Callithrix jacchus (Marmoset)

Summary

- Conservation of 3D structure after chromosomic rearrangements.
- Recently duplicated regions are more isolated from surrounding DNA.
- Compartments are very conserved in primates; detectable changes may not be related to cell-specific features.
- TAD borders conserved with similar selective strength against gain and loss
- New, but few primate specific expansions carrying CTCF sites
- Selection against the creation of new and strong CTCF sites

What next?

Can we see the path of chromatin @10Kb resolution? images + Hi-C + modeling

Is there a dynamic coupling between structure and gene activity?

The End!

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
CHROMDESIGN Training the next generation of researches
GOBIERNO DE ESPAÑA MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Generalitat de Catalunya
Fundació La Marató de TV3
Fundación Bancaria "la Caixa"