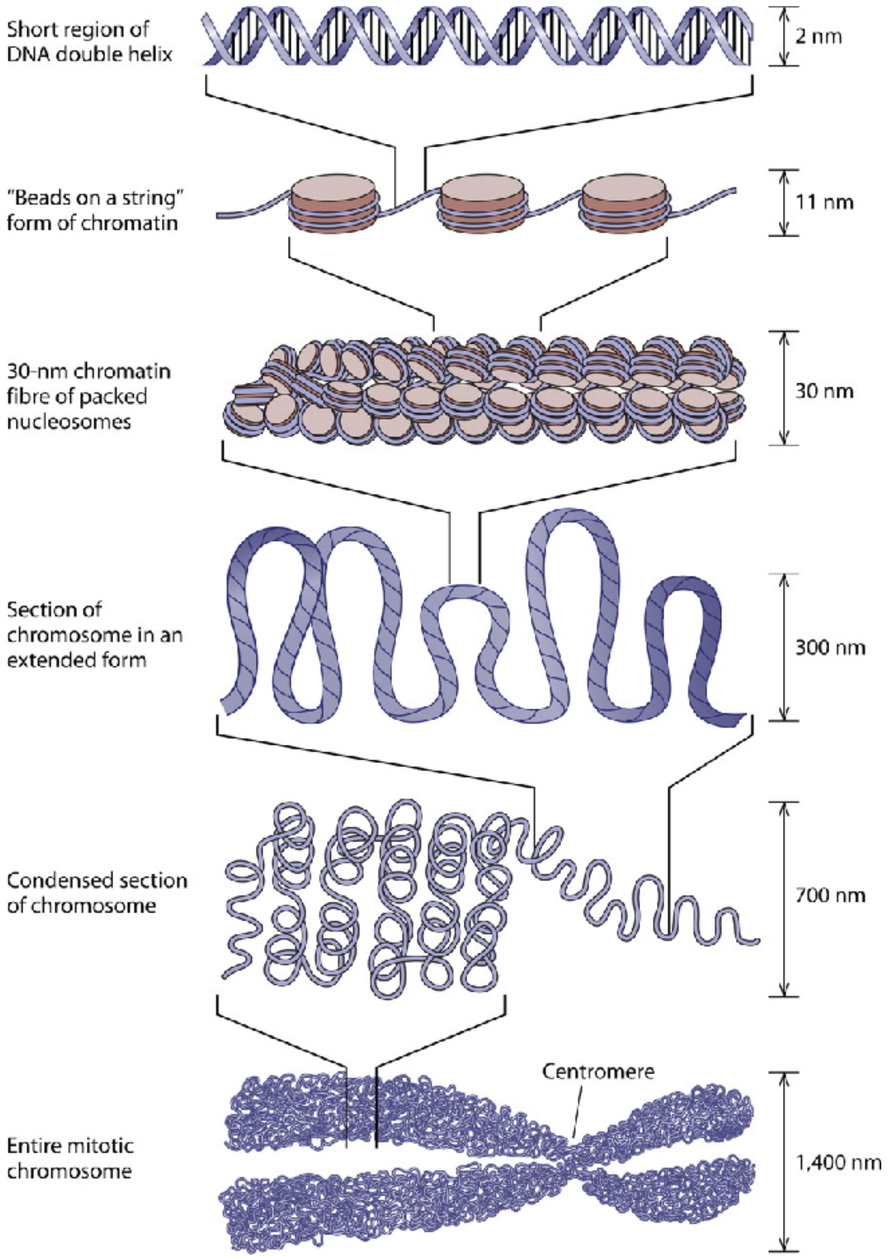
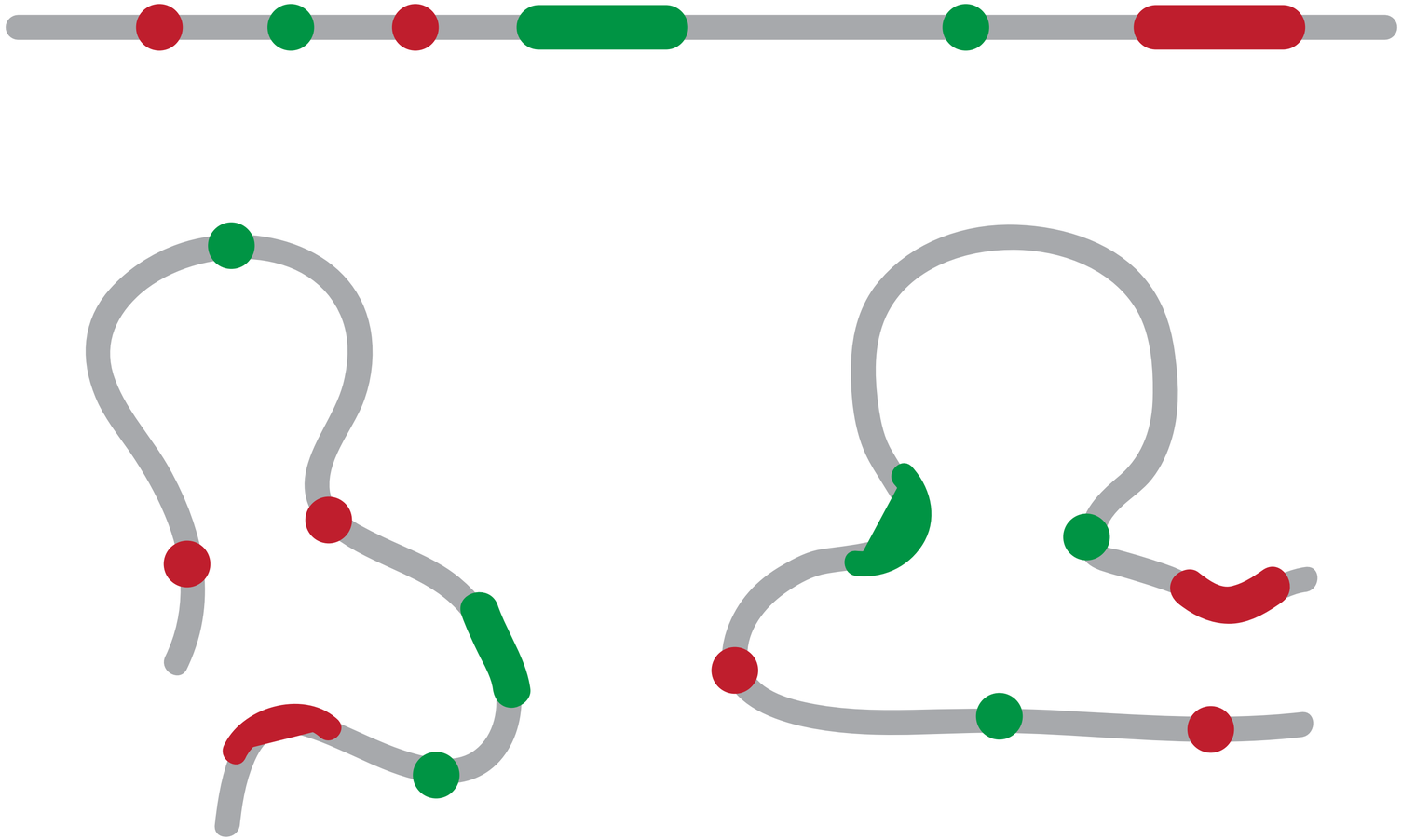
Structure determination of genomes and genomic domains by satisfaction of spatial restraints.

Marc A. Marti-Renom
CNAG-CRG · ICREA

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

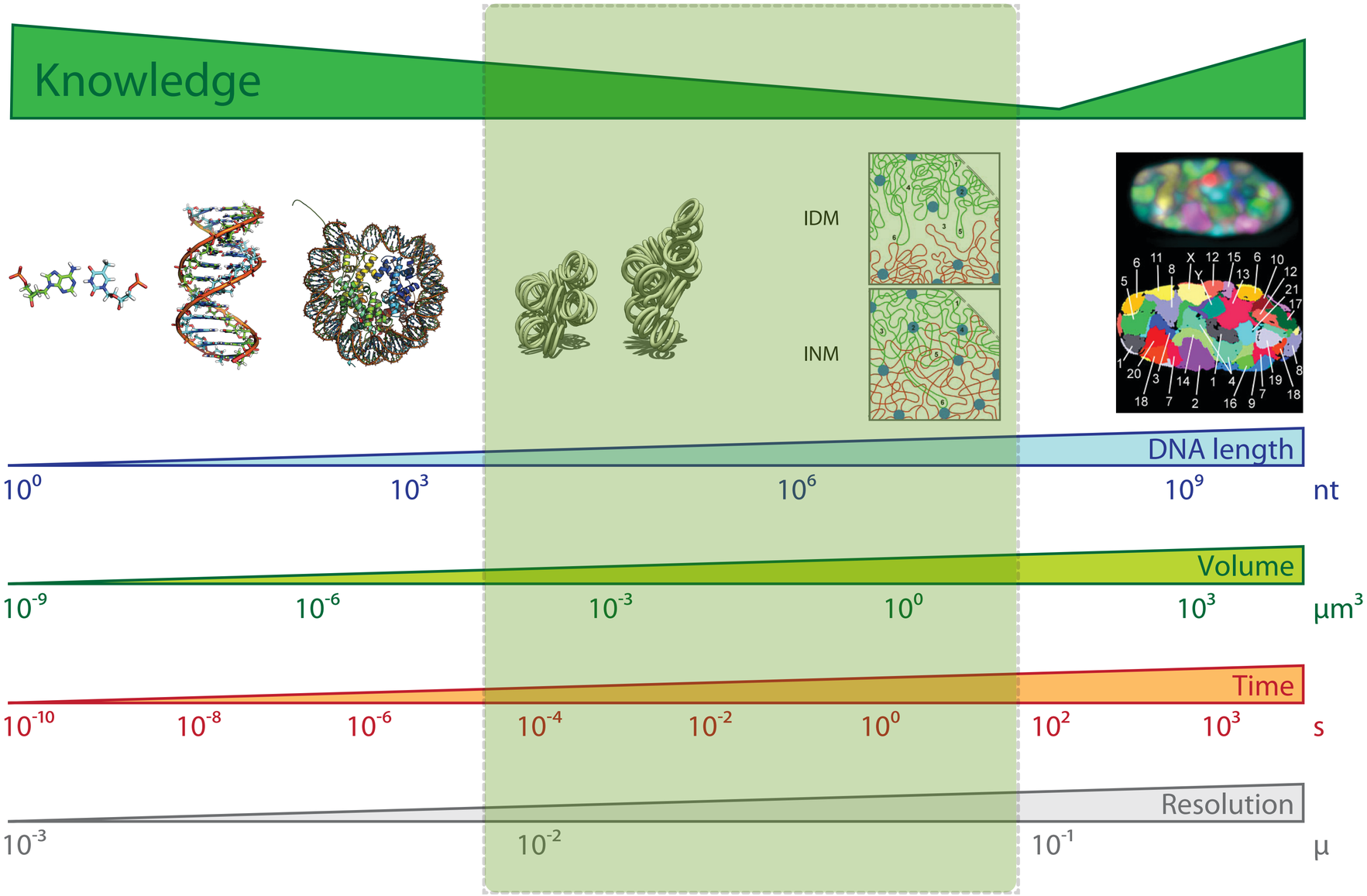
cnag CRG ICREA


Short region of DNA double helix
2 nm
"Beads on a string" form of chromatin
11 nm
30-nm chromatin fibre of packed nucleosomes
30 nm
Section of chromosome in an extended form
300 nm
Condensed section of chromosome
700 nm
Centromere
Entire mitotic chromosome
1,400 nm



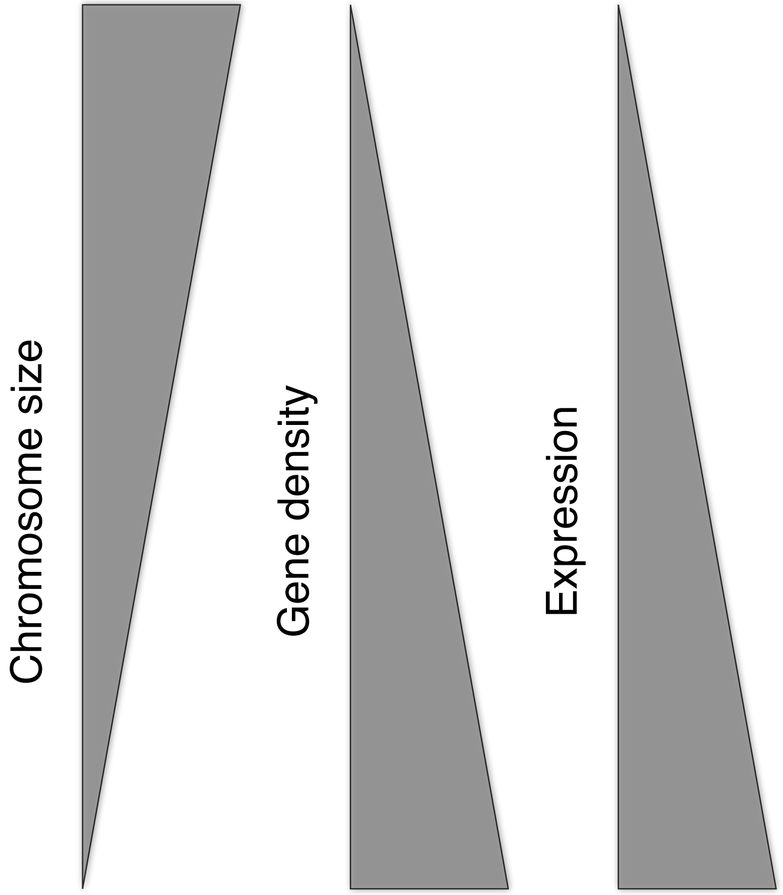
Resolution Gap

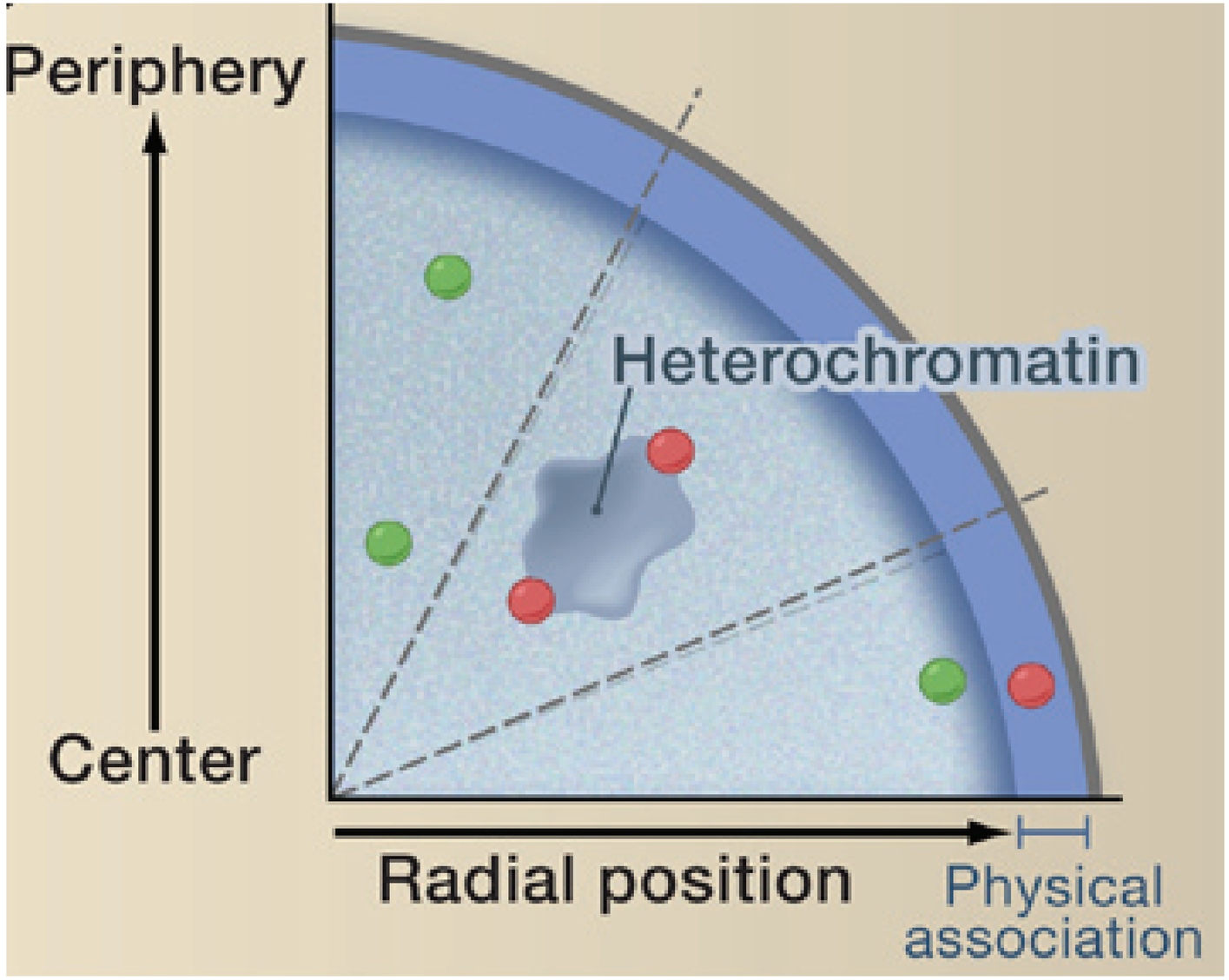
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol *7*, e1002125 (2011)



Level I: Radial genome organization

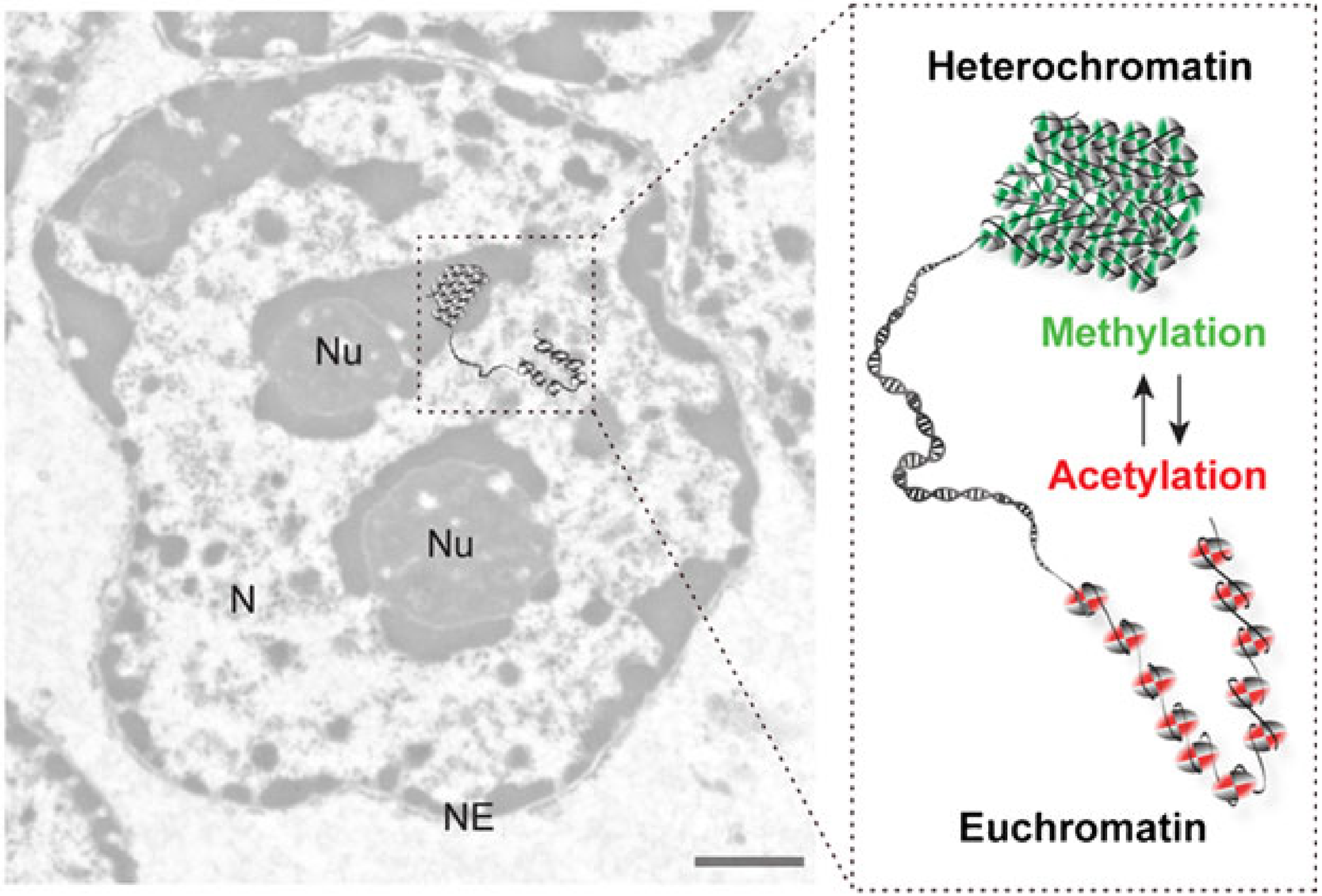
Takizawa, T., Meaburn, K. J. & Misteli, T. The meaning of gene positioning. Cell 135, 9–13 (2008).



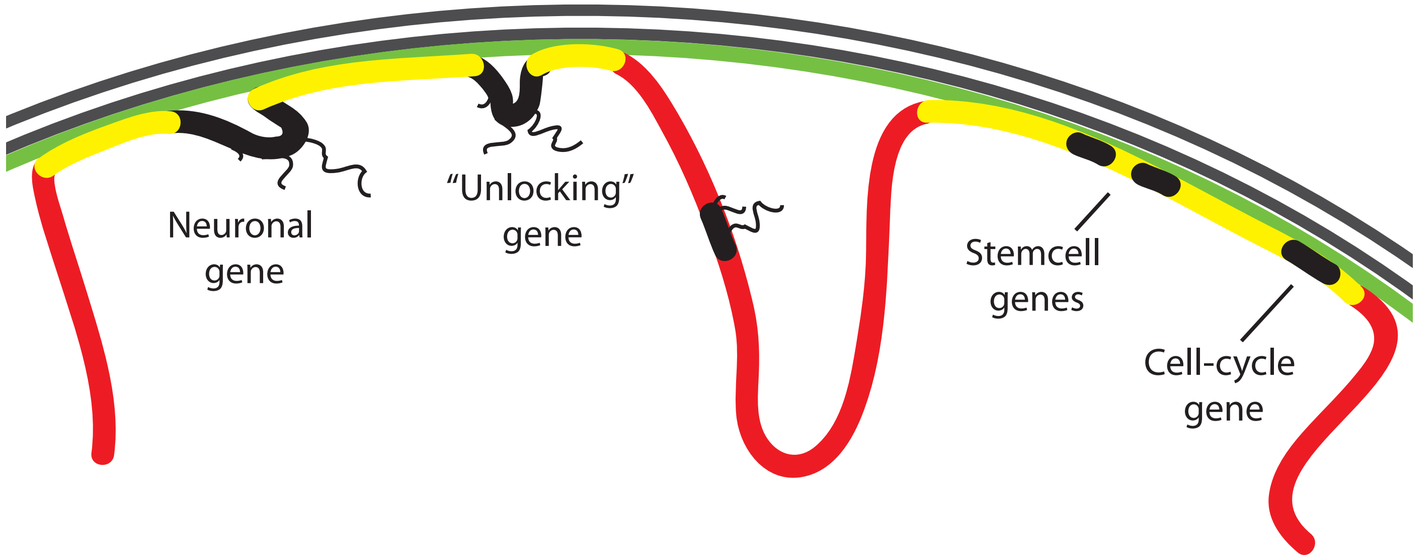


Level II: Euchromatin vs heterochromatin

Electron microscopy



Level III: Lamina-genome interactions



nuclear membrane
nuclear lamina

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

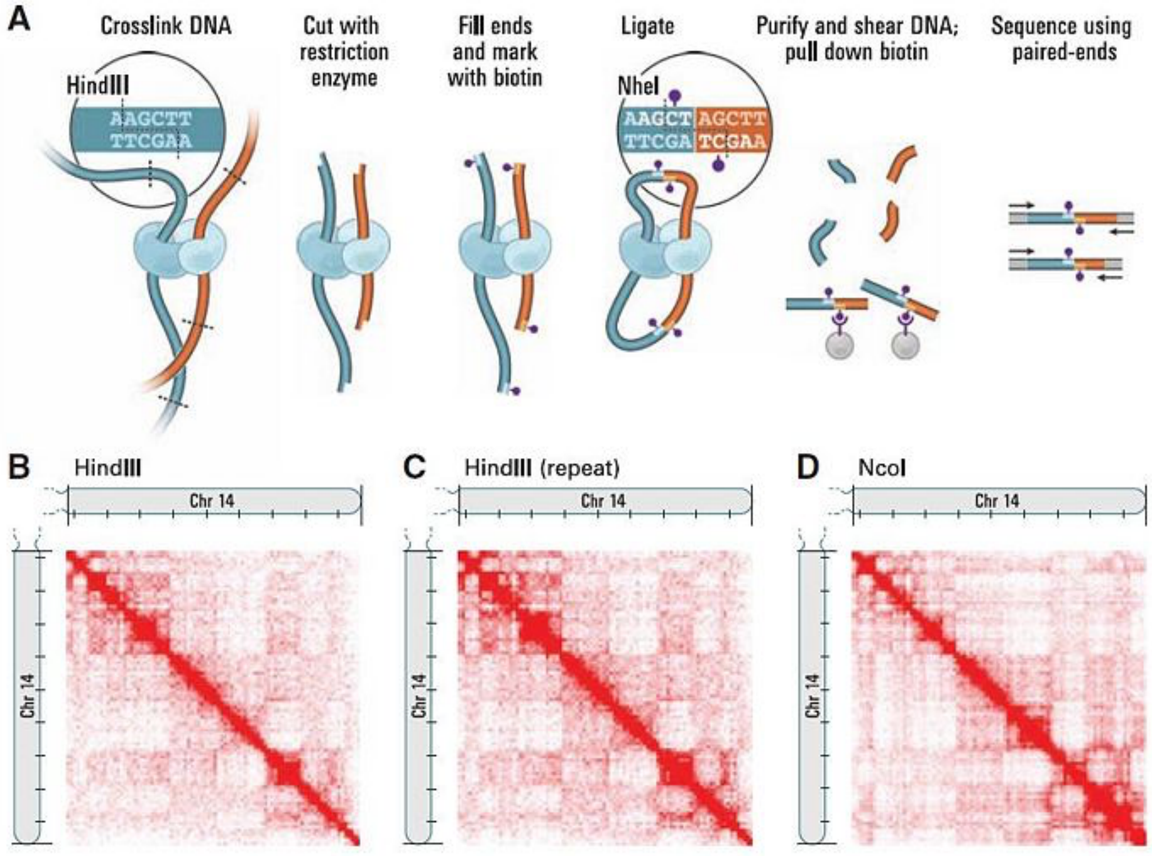
Genes

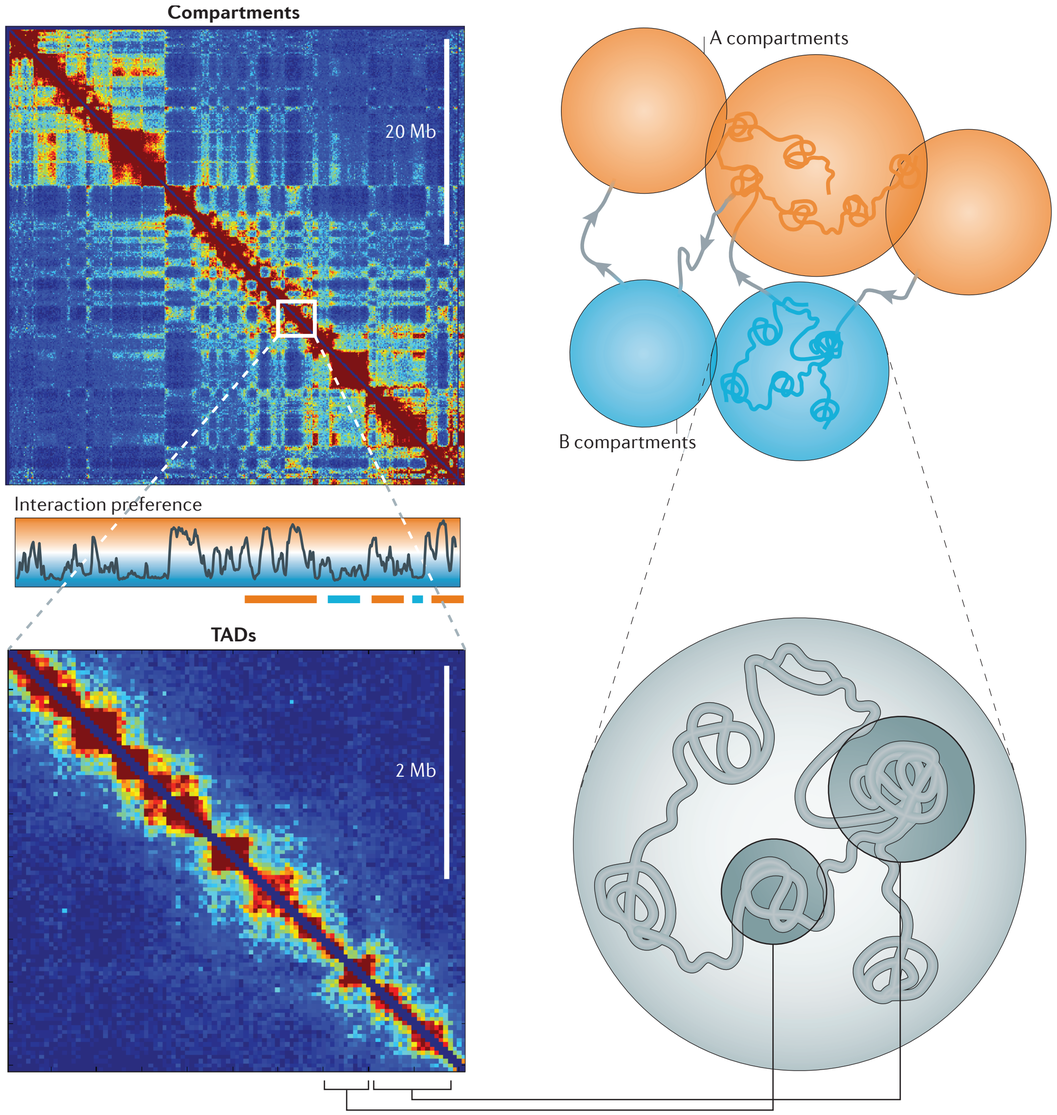
mRNA

Adapted from Molecular Cell 38, 603-613, 2010

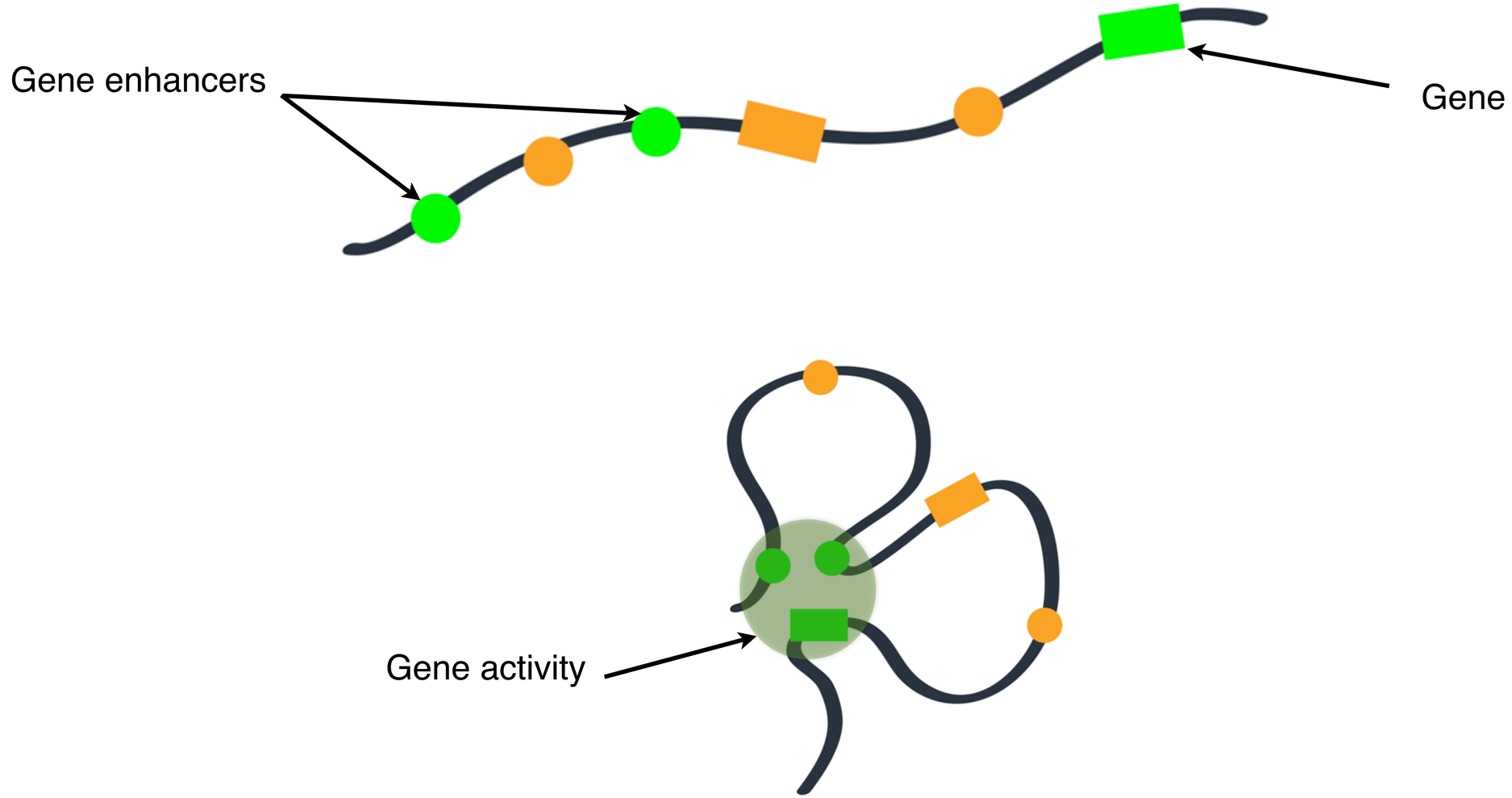
Level IV: Higher-order organization

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



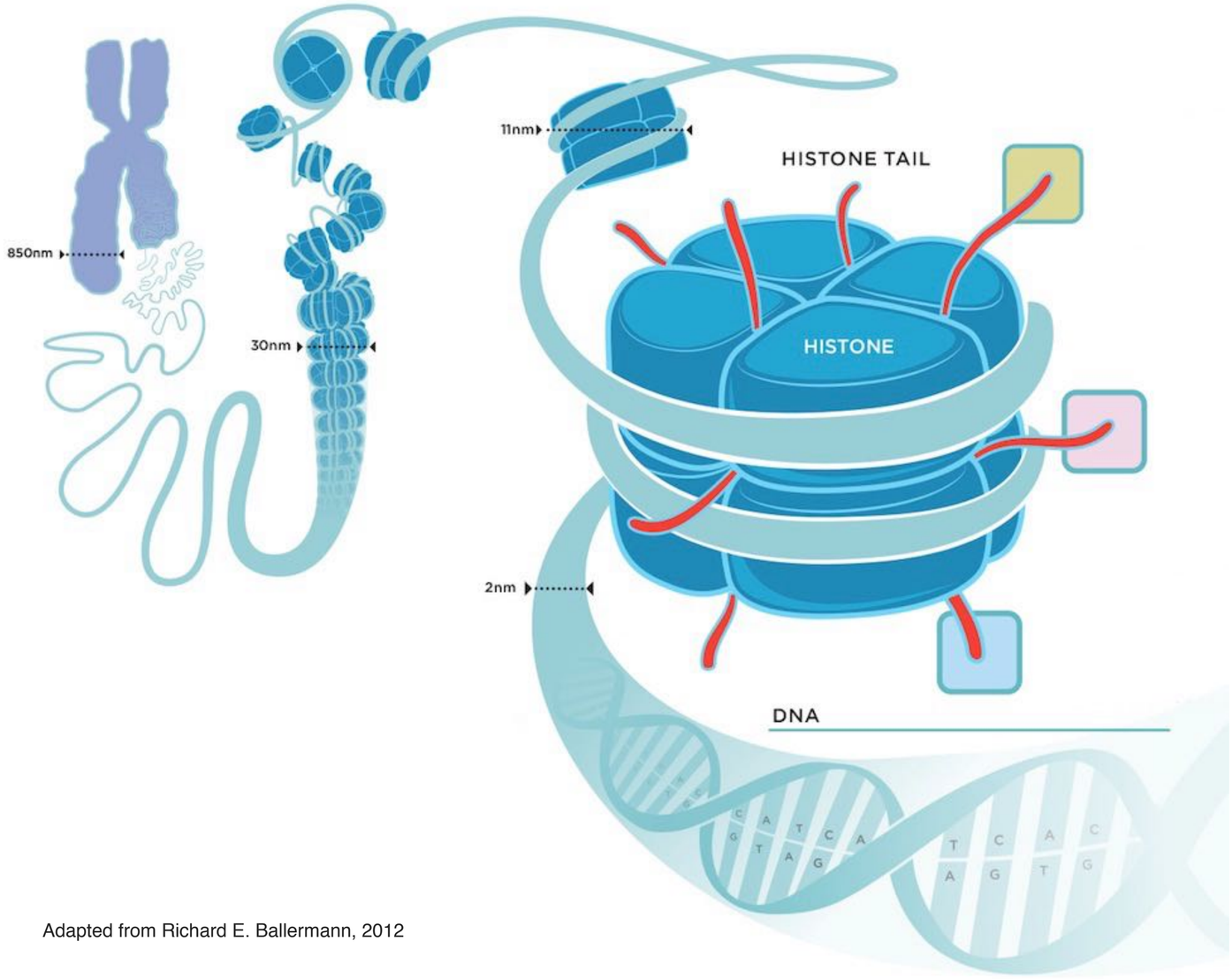


Level V: Chromatin loops



Level VI: Nucleosome

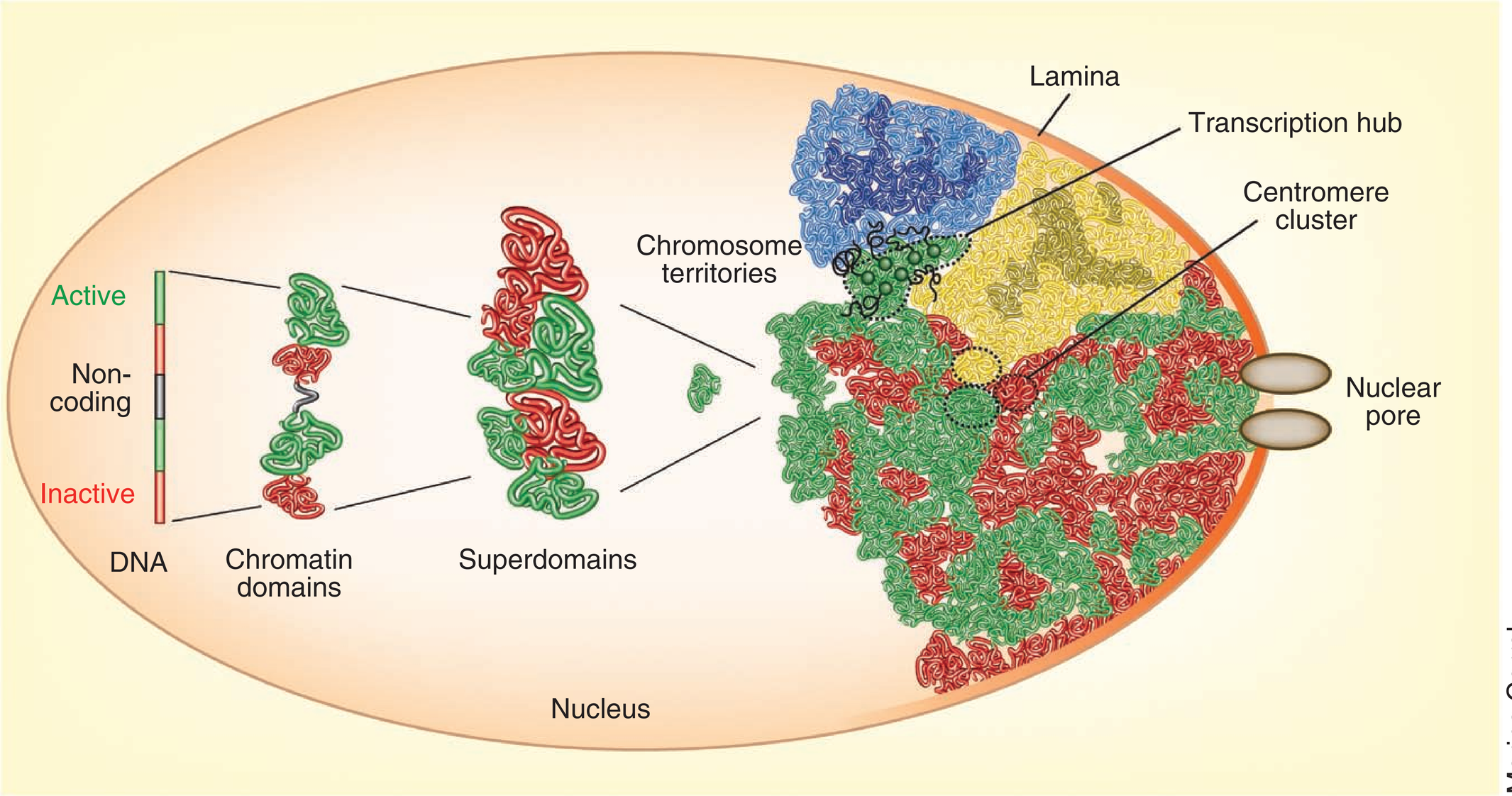
Chromosome **Chromatin fibre** **Nucleosome**



Adapted from Richard E. Ballermann, 2012

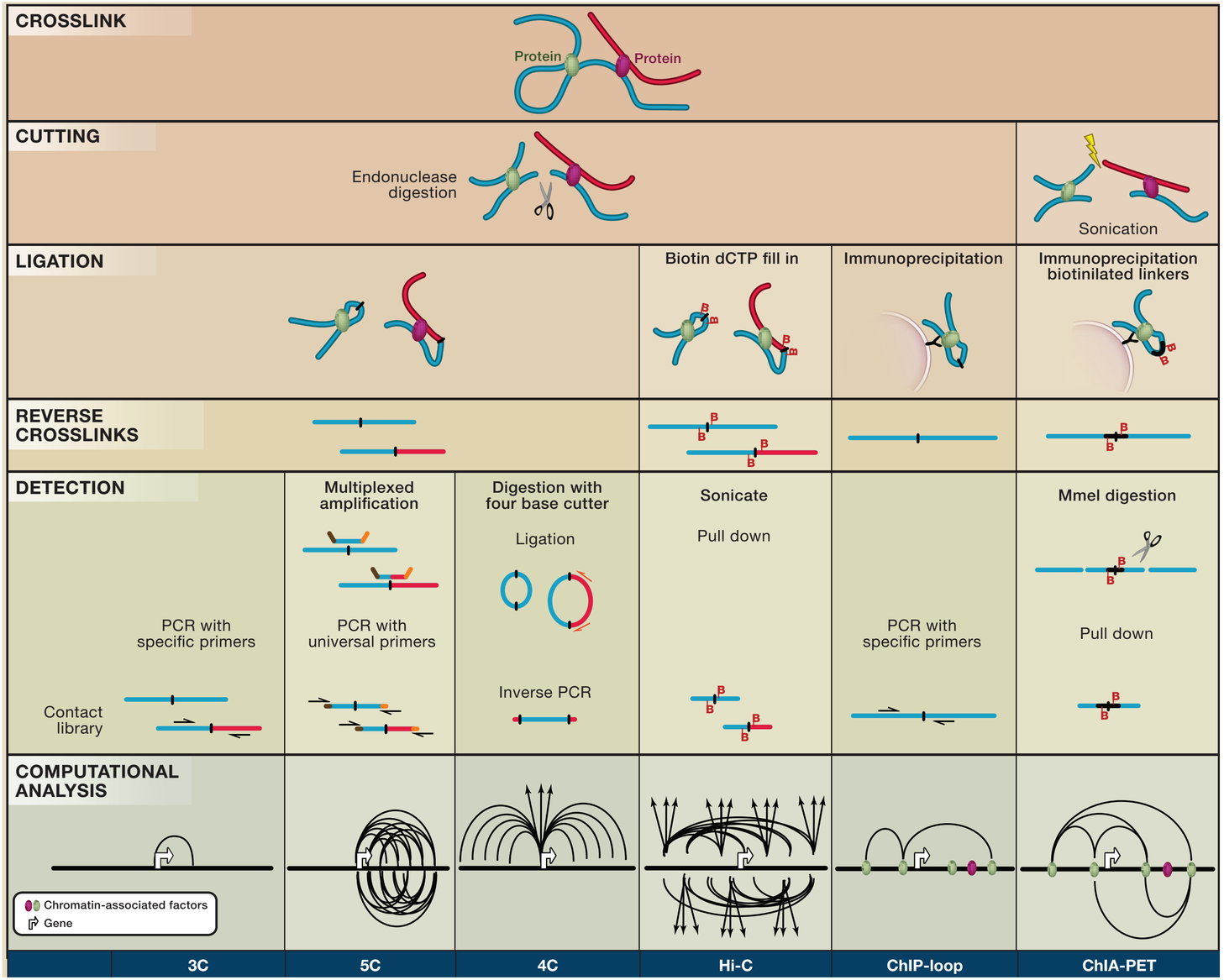
Complex genome organization

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



Marina Corral

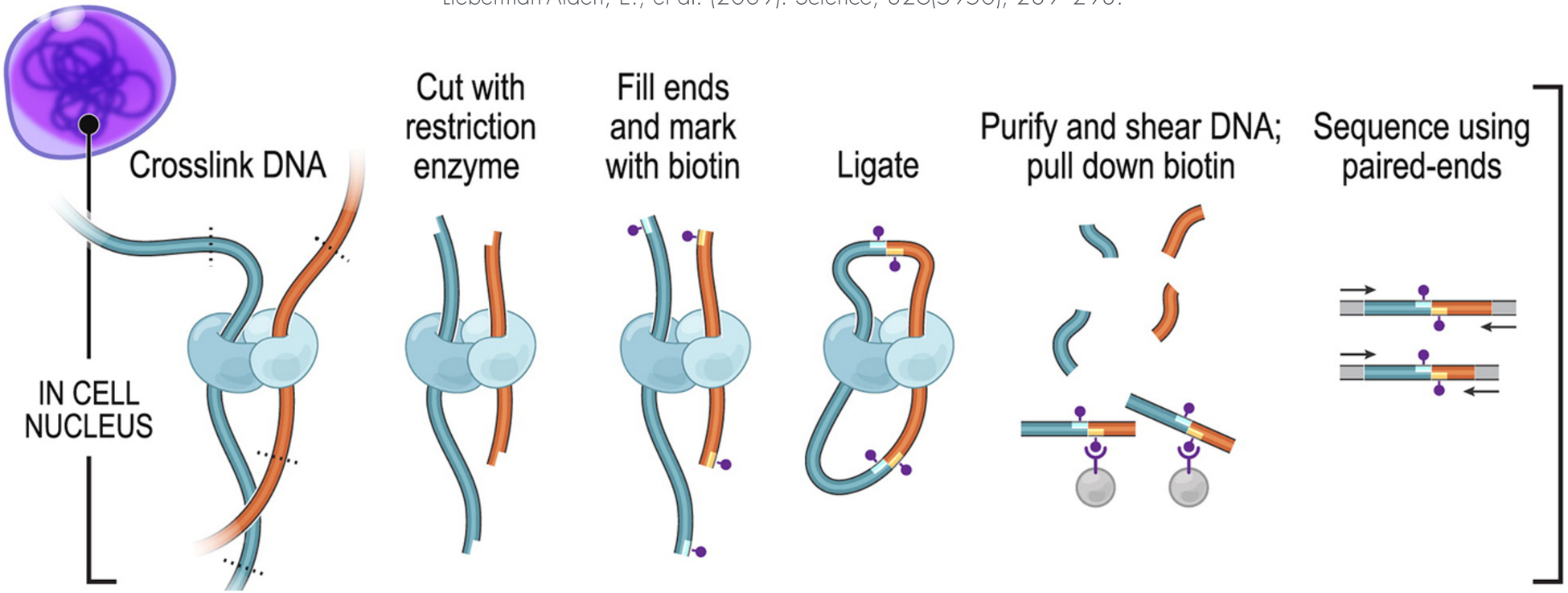
Chromosome Conformation Capture

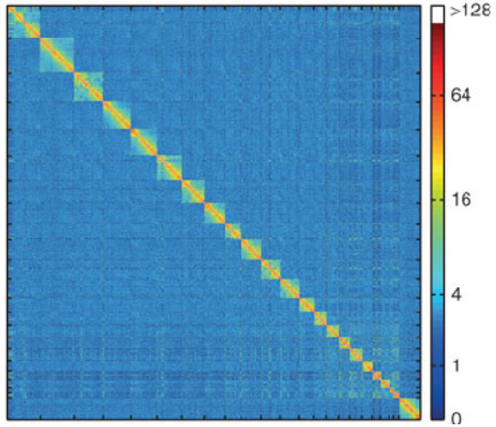


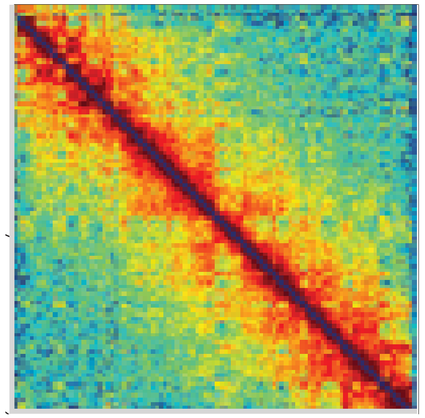
Hakim, O., & Misteli, T. (2012). SnapShot: Chromosome Confirmation Capture. Cell, 148(5), 1068–1068.e2.

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.

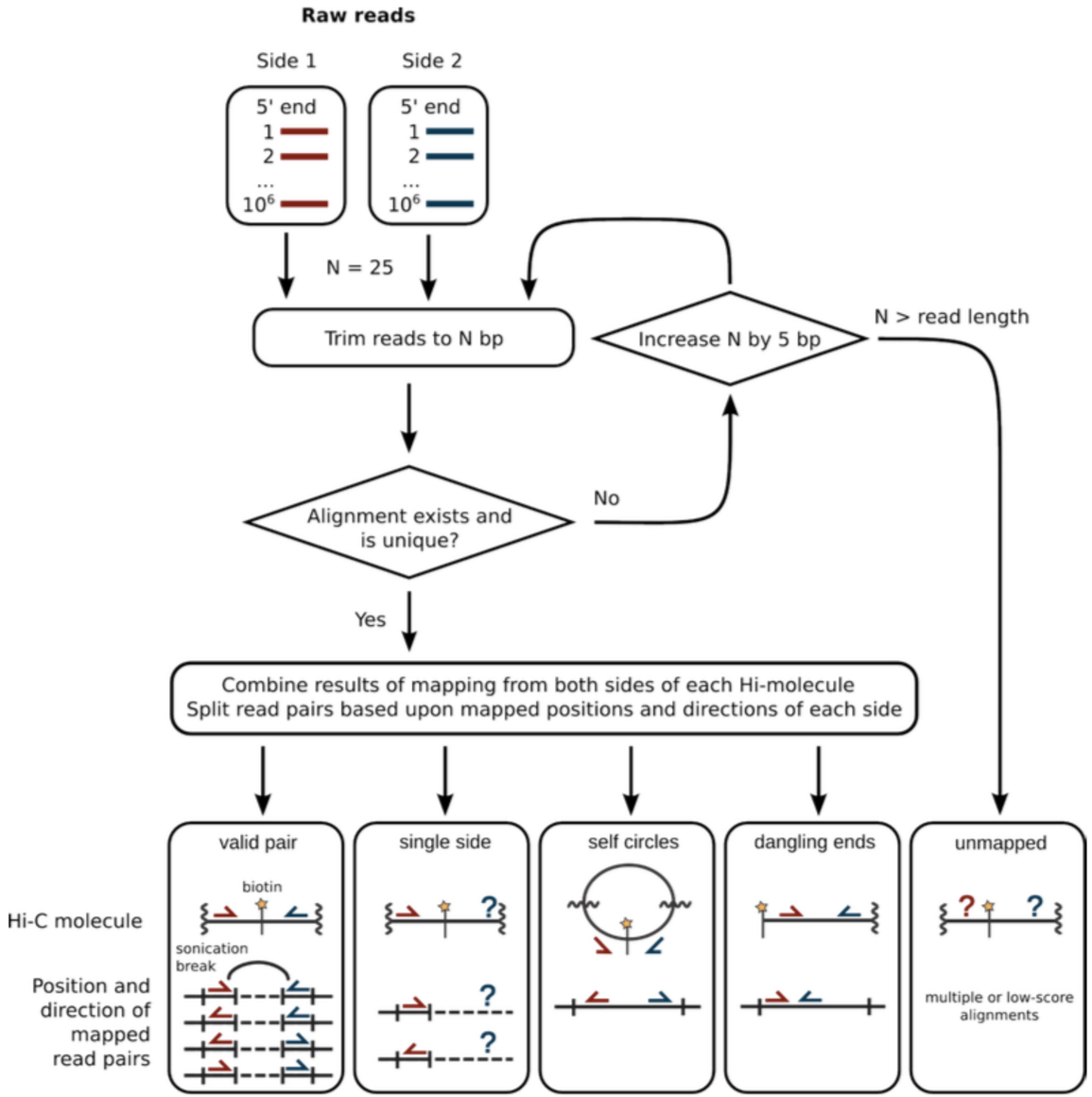






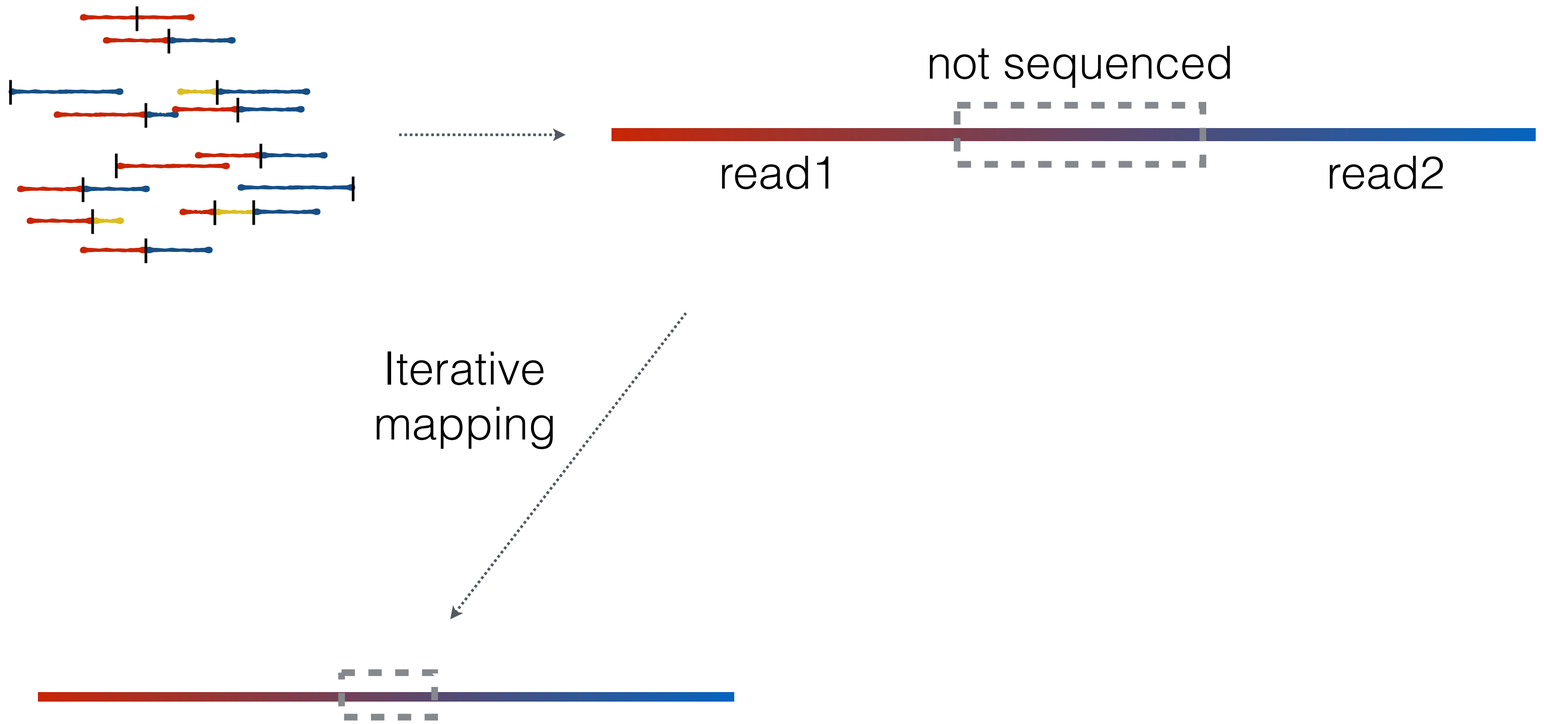
Mapping & Filtering

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



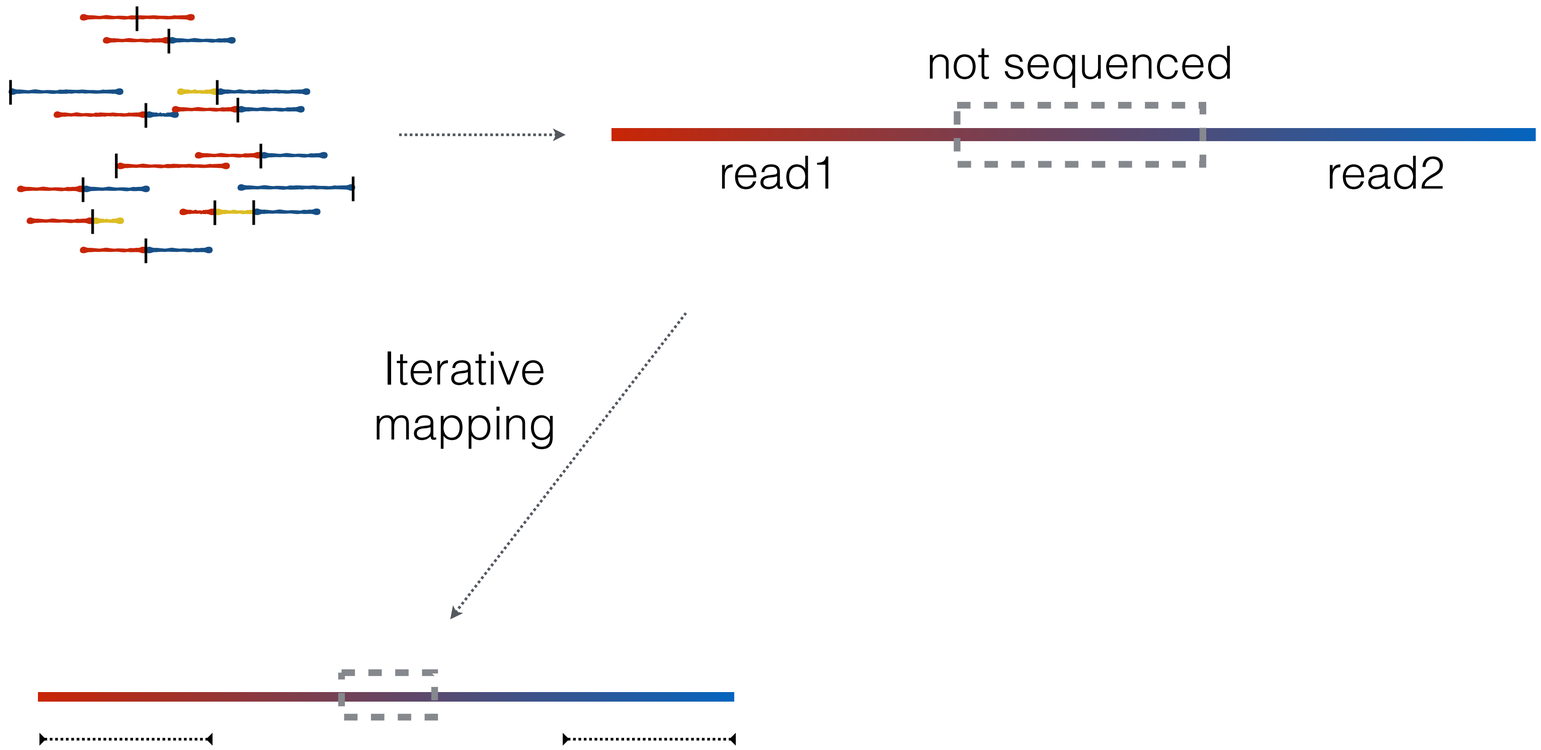
Mapping @TADbit

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



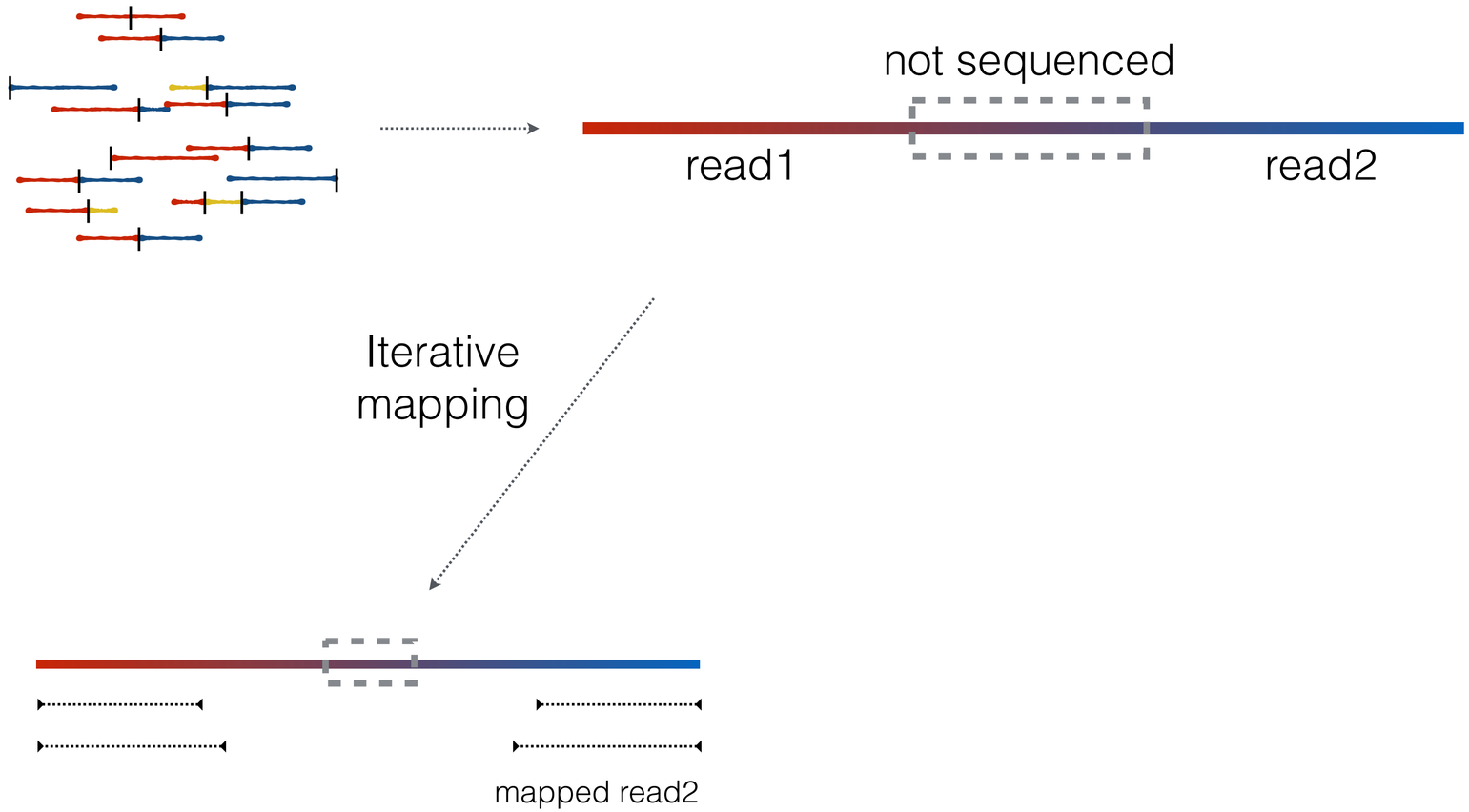
Mapping @TADbit

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



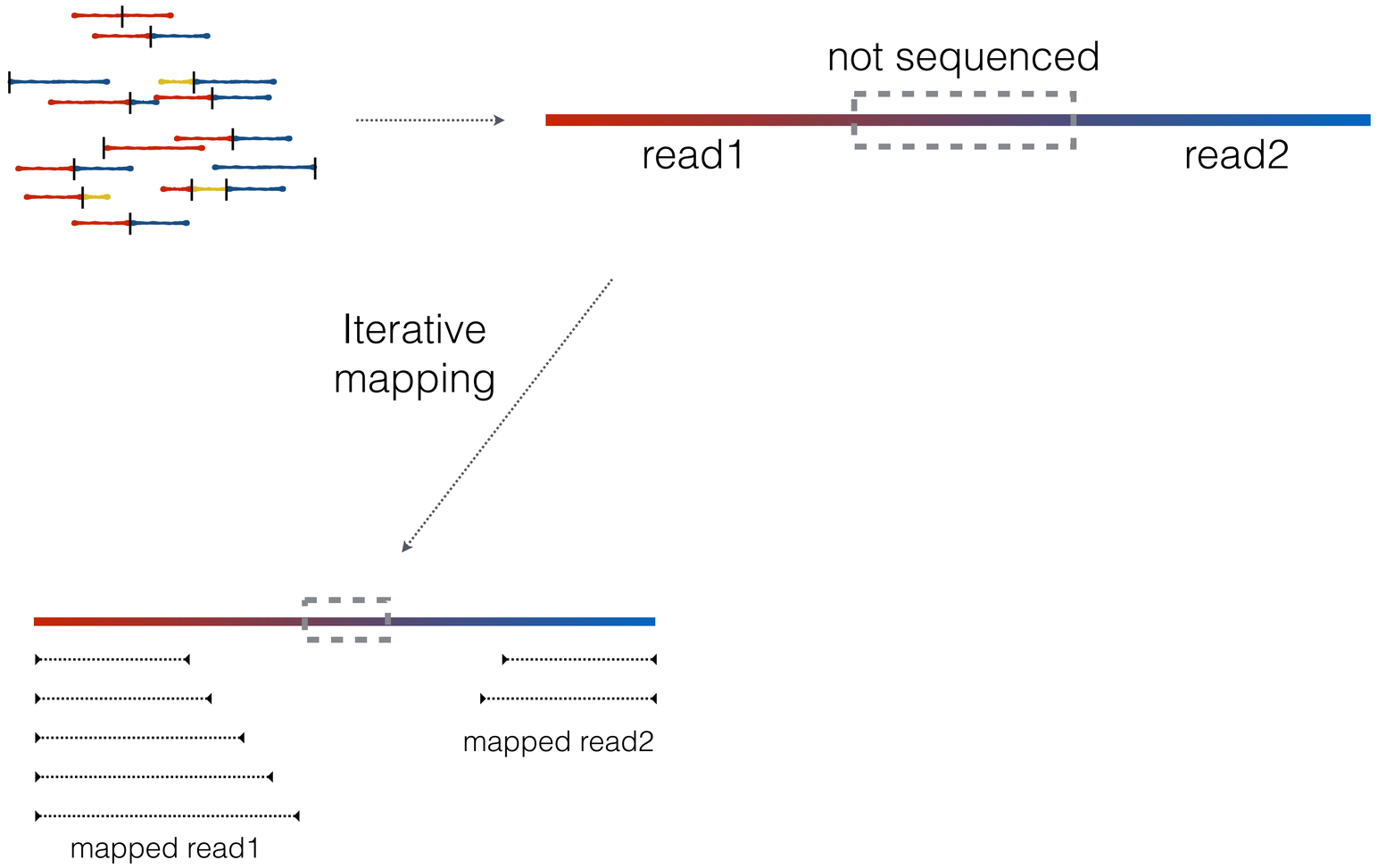
Mapping @TADbit

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



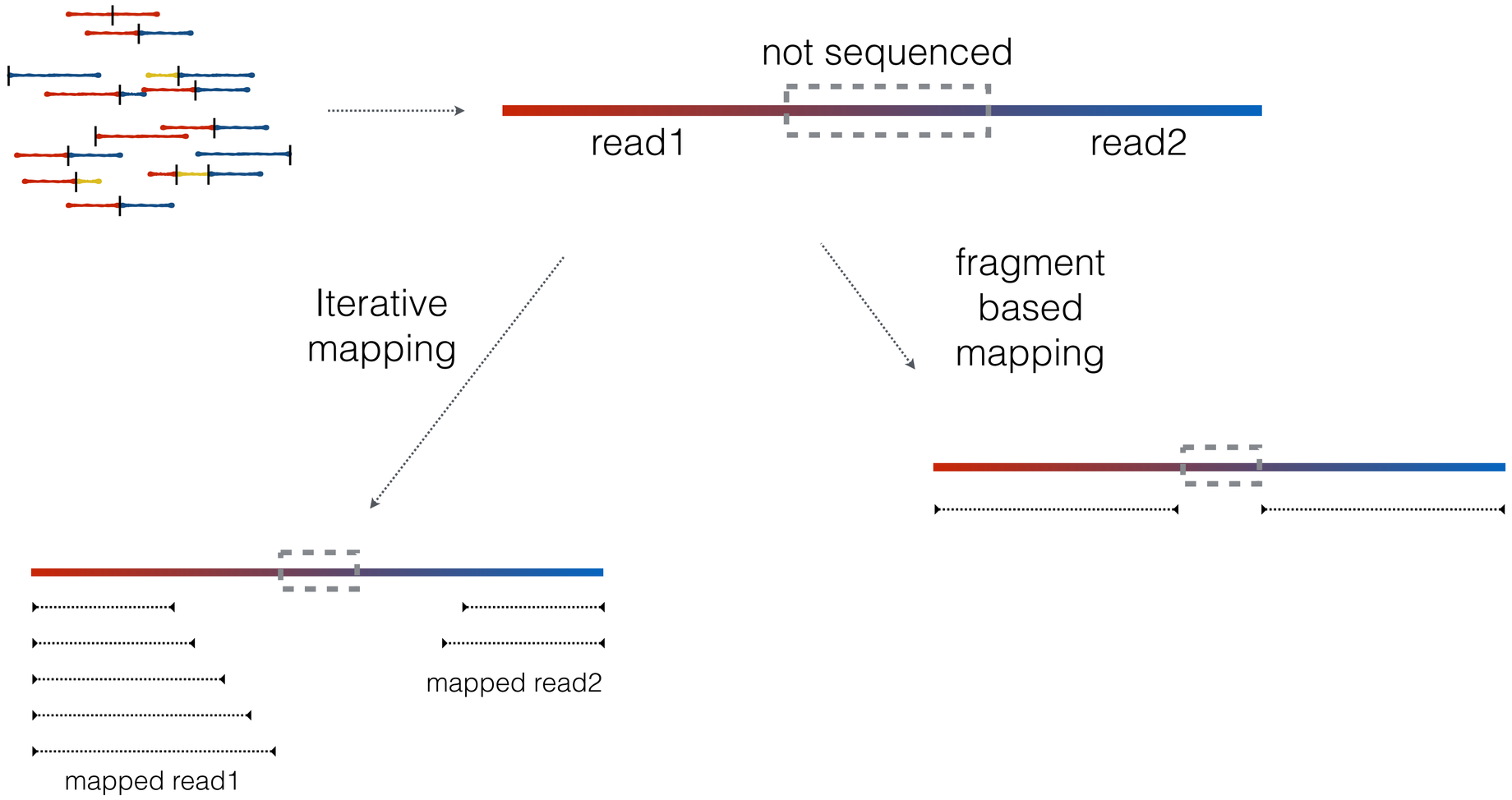
Mapping @TADbit

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



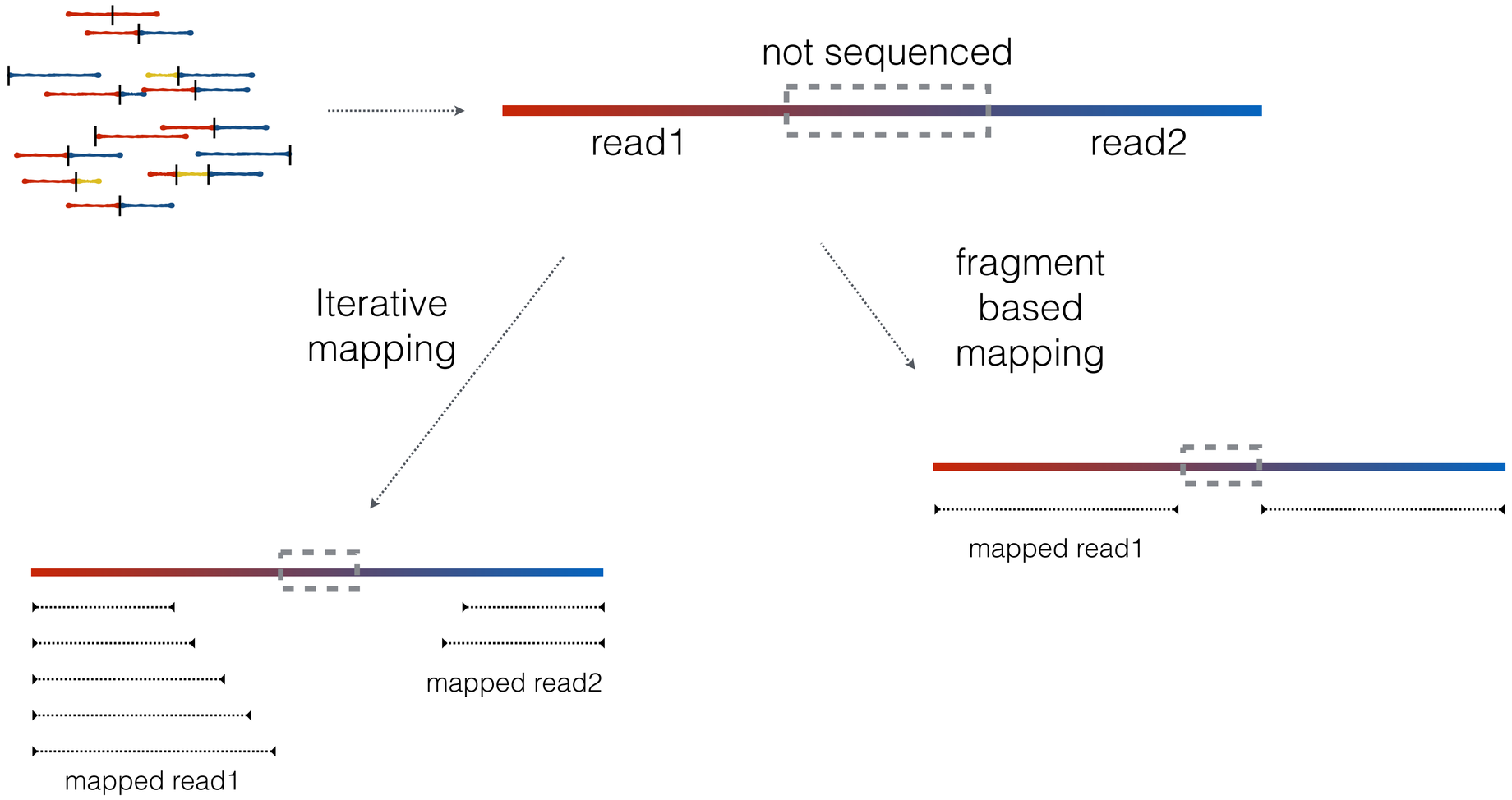
Mapping @TADbit

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



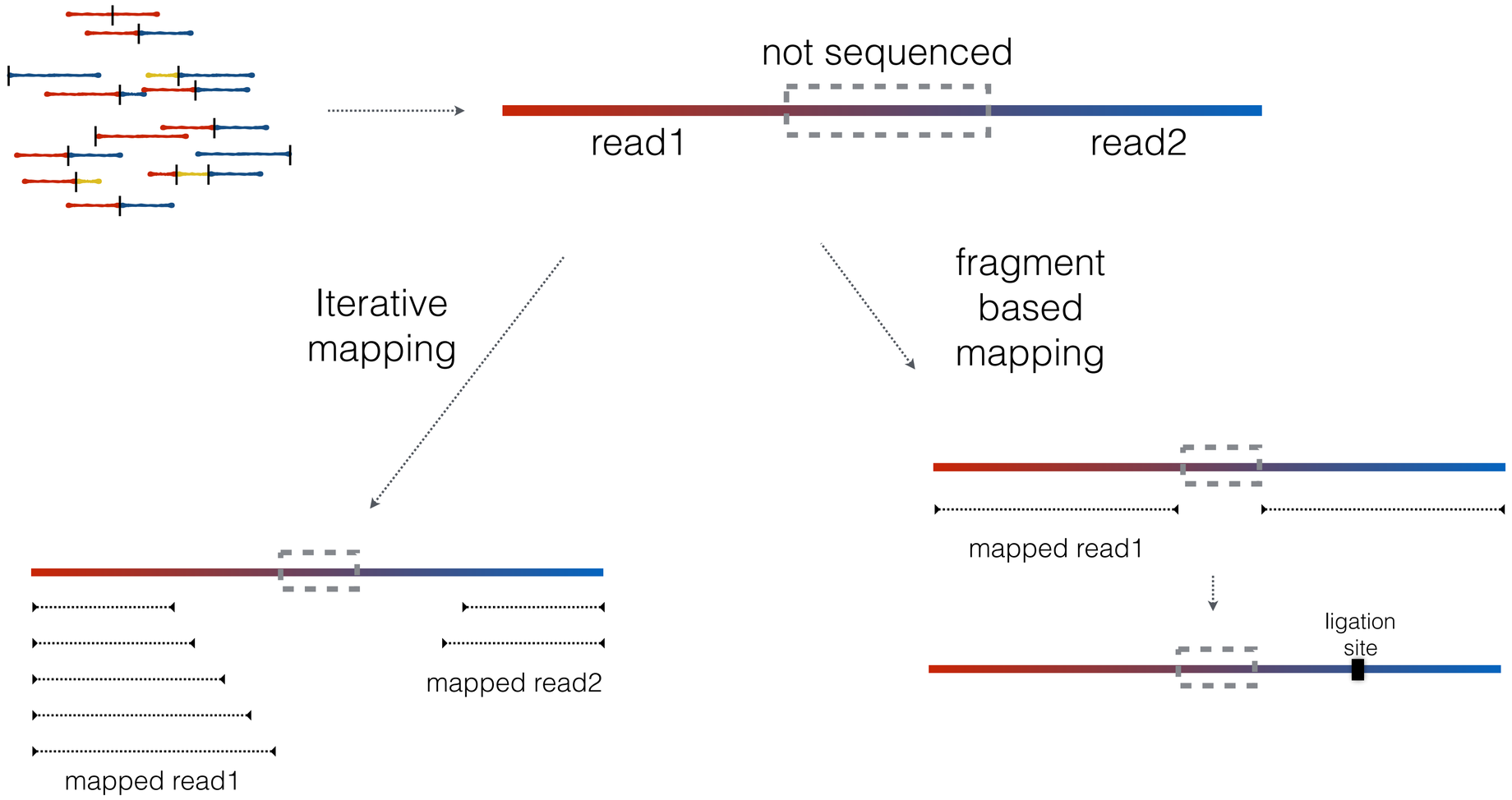
Mapping @TADbit

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



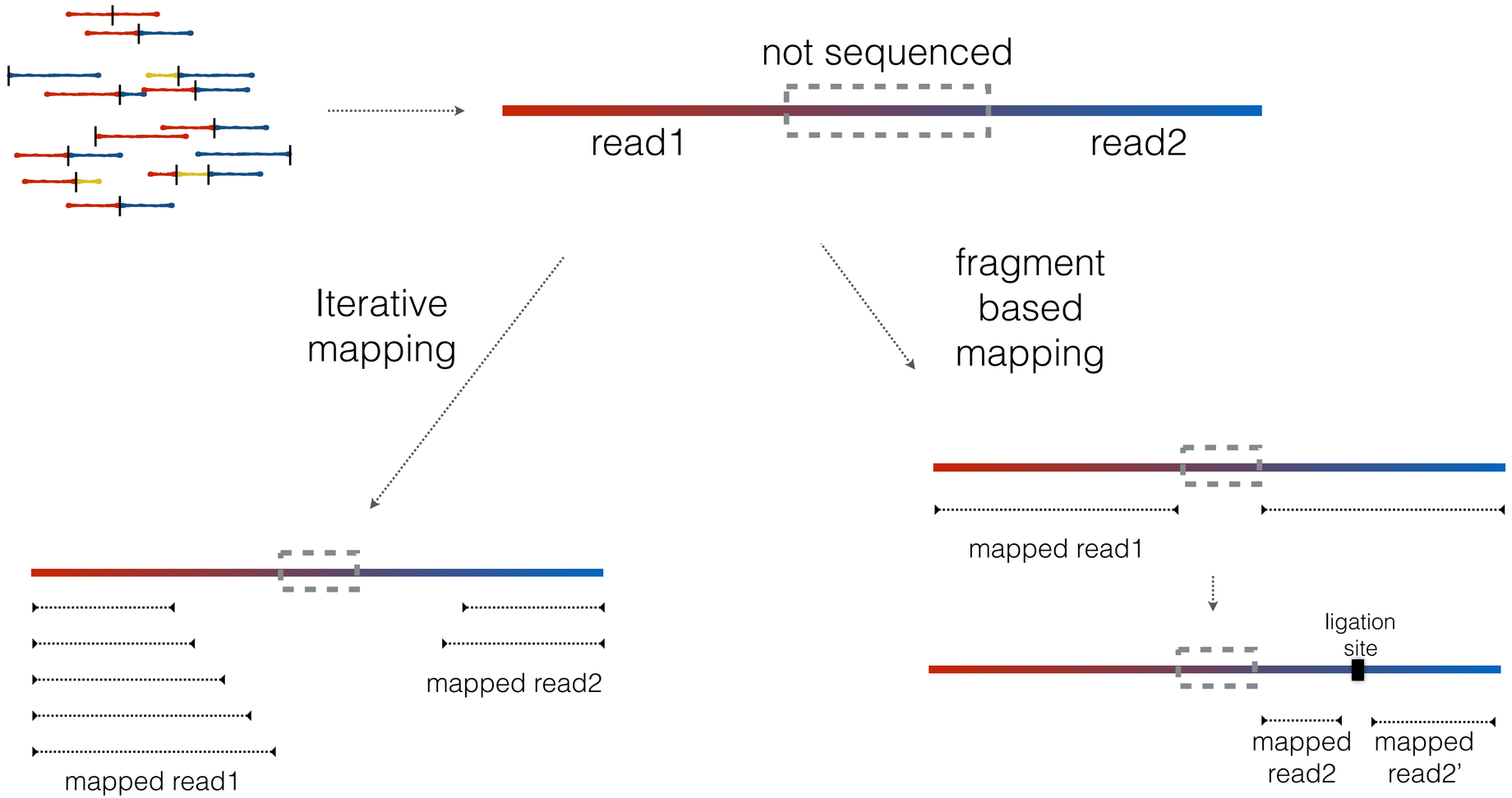
Mapping @TADbit

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



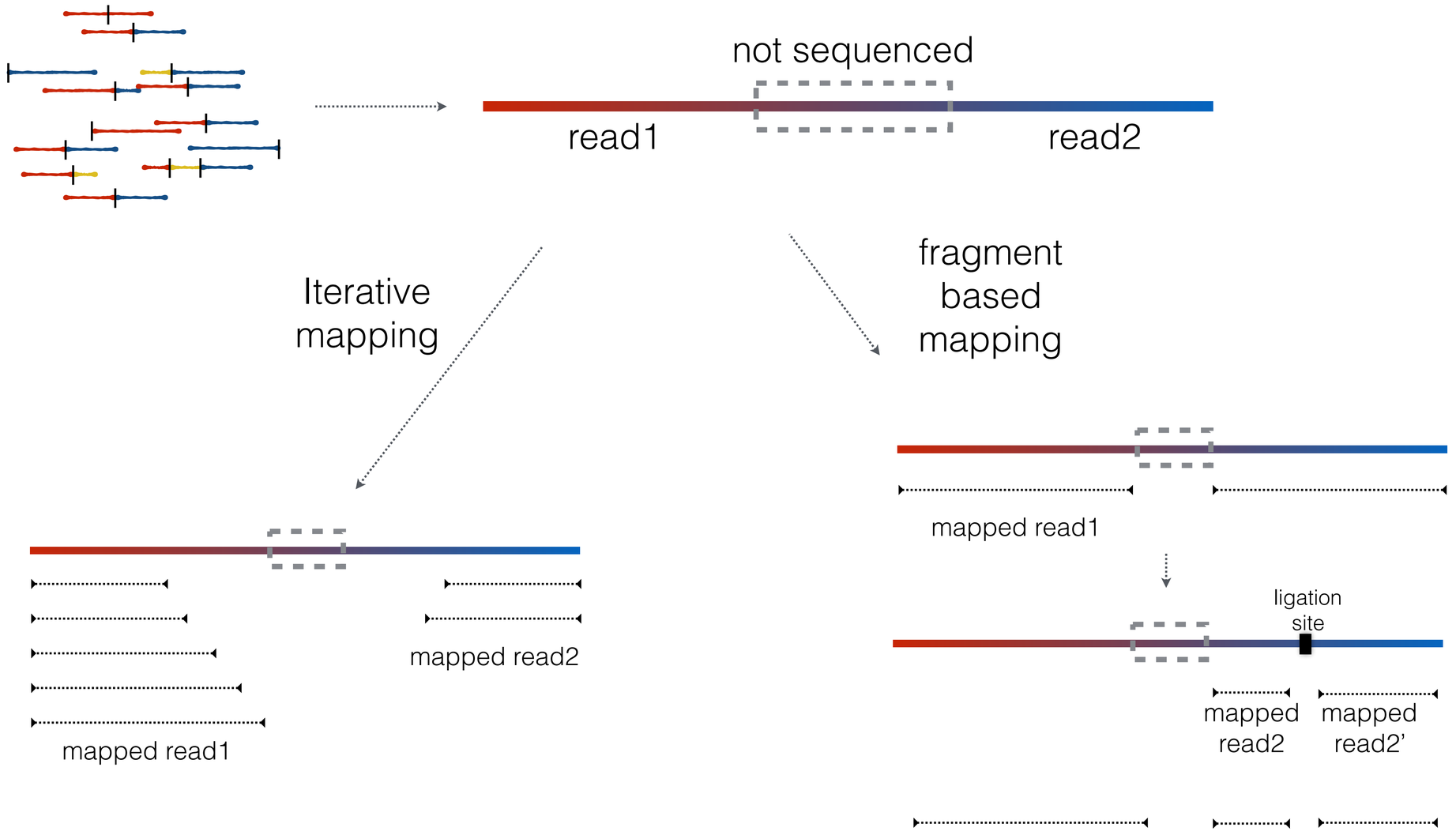
Mapping @TADbit

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

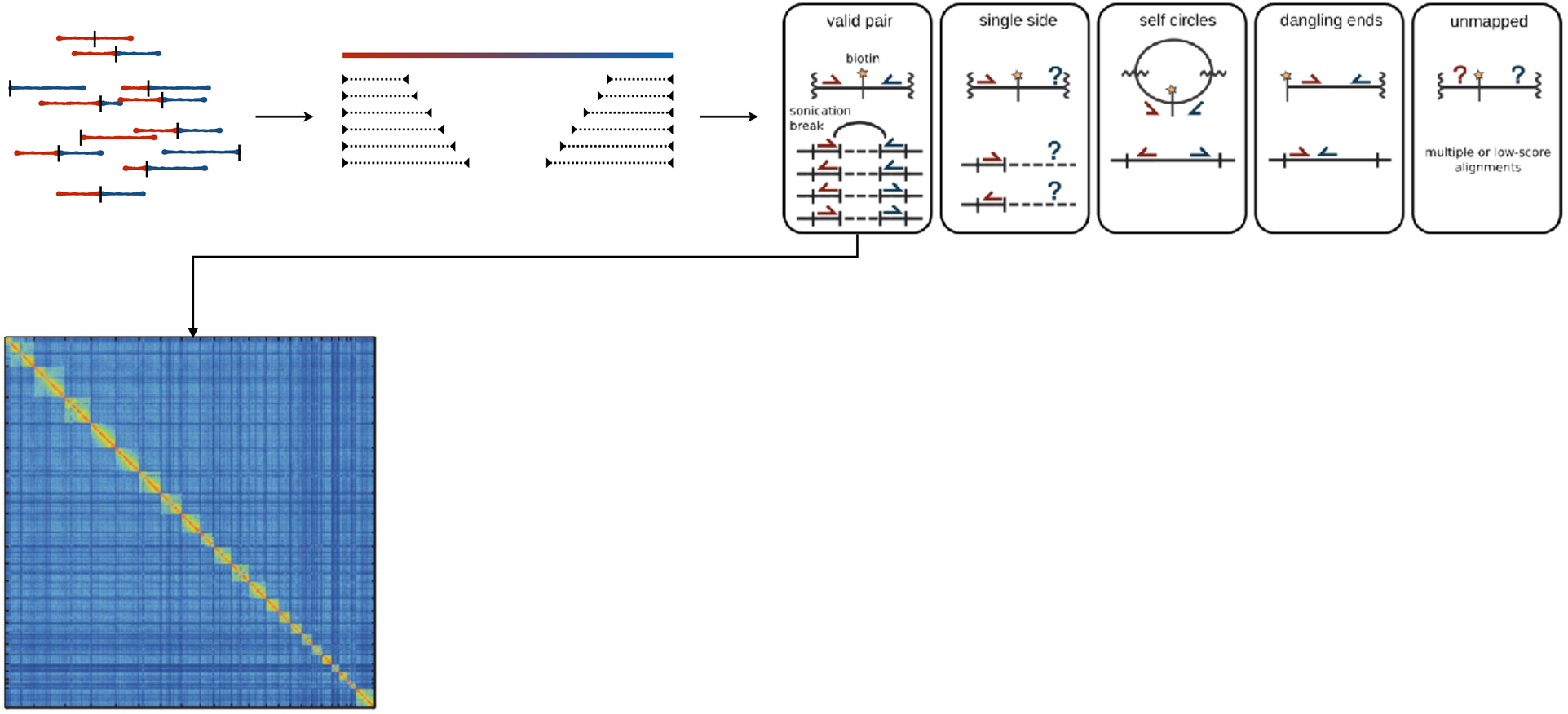


Mapping @TADbit

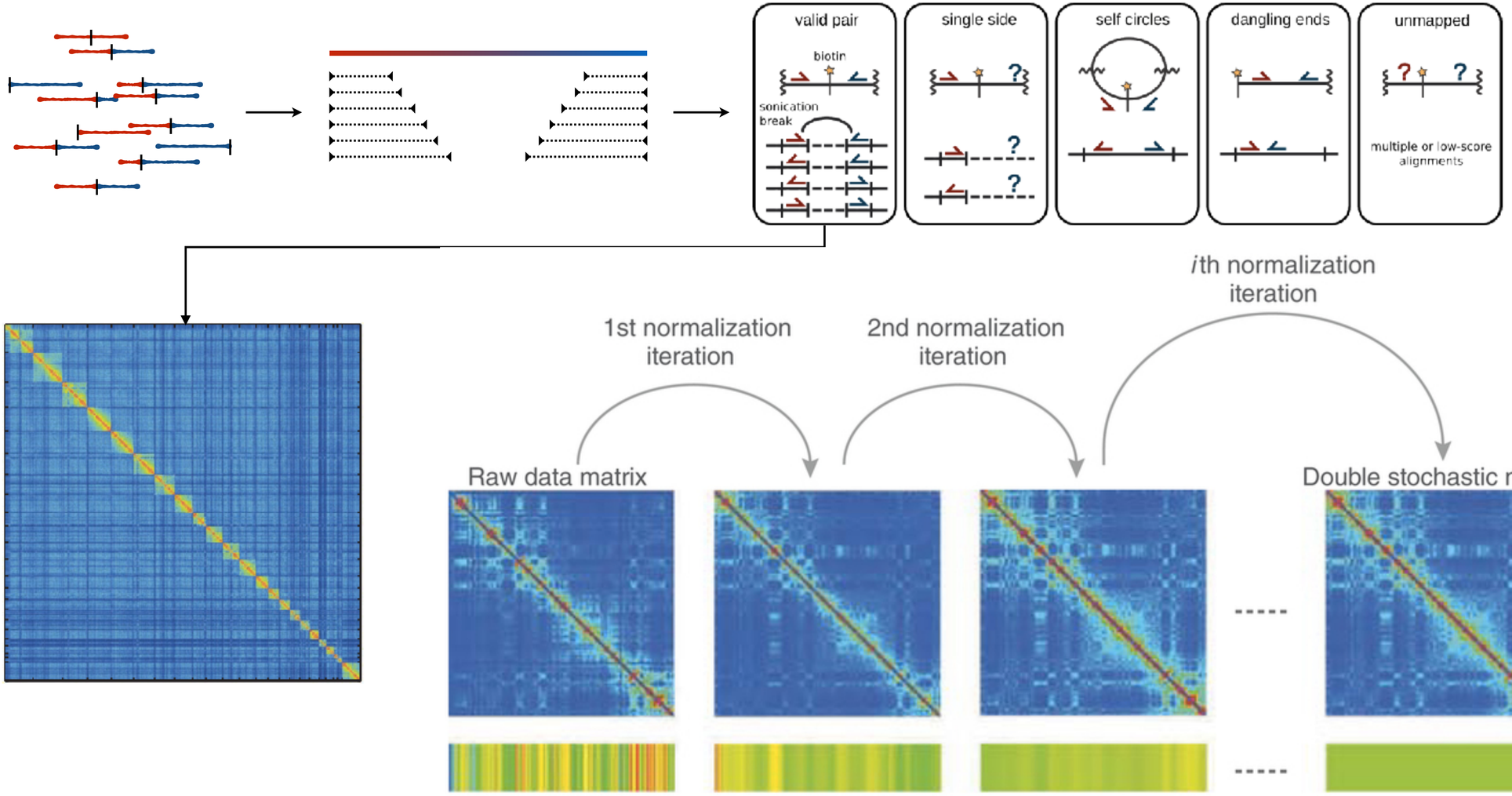
Serra, Baù, et al. (2017). PLOS CompBio



Interaction matrices



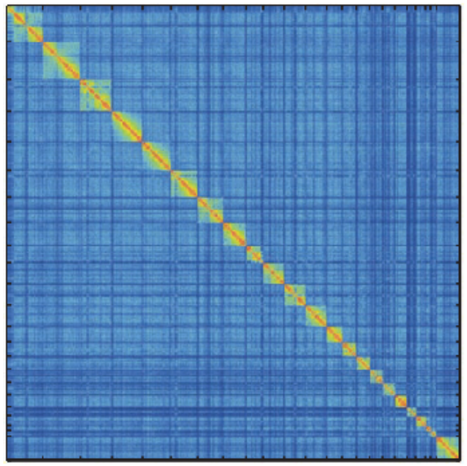
Interaction matrices



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

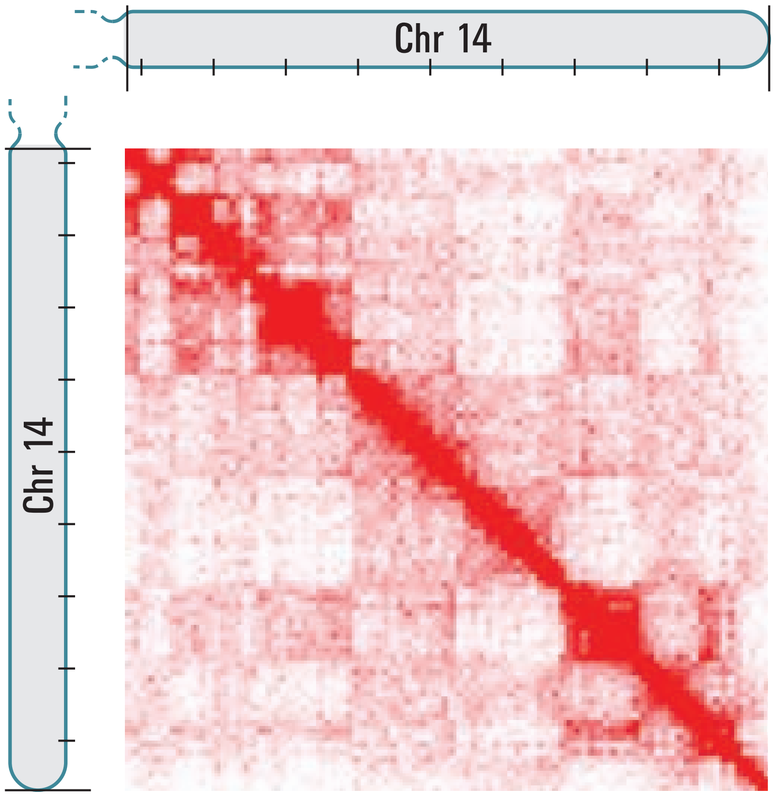
How much you normally map?

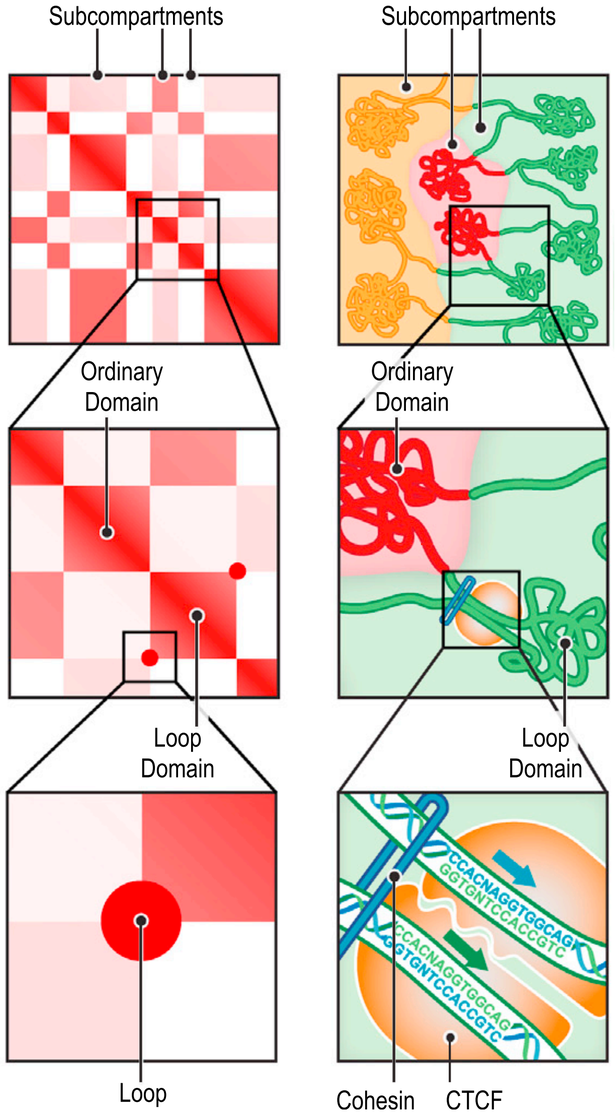
- 80-90% each end => 60-80% intersection
- ~1% multiple contacts
- Many of intersecting pairs will be lost in filtering…
- Final 40-60% of valid pairs
- One measure of quality is the CIS/TRANS ration (70-80% good)



Hierarchical genome organisation

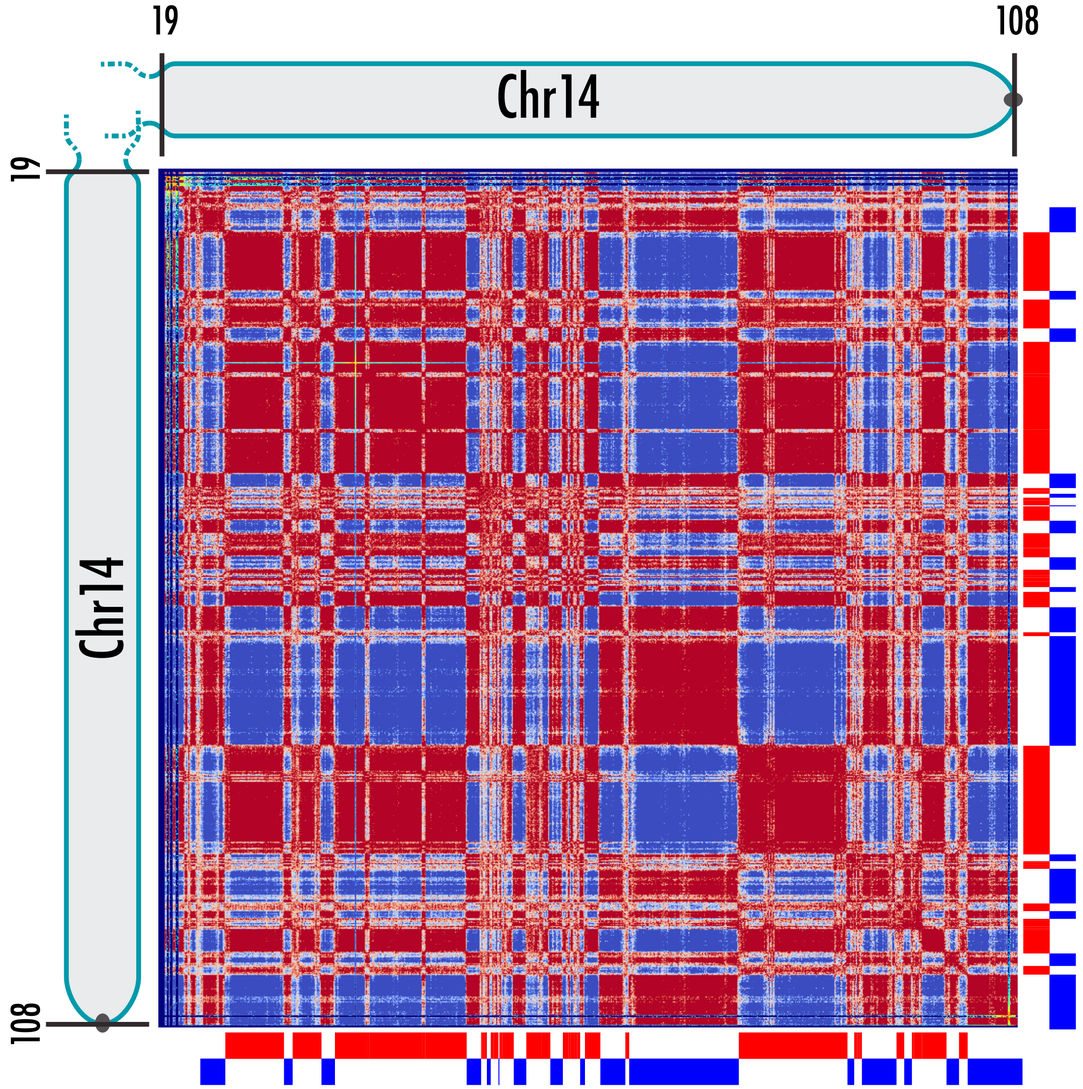
Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.
Rao, S. S. P., et al. (2014). Cell, 1–29.





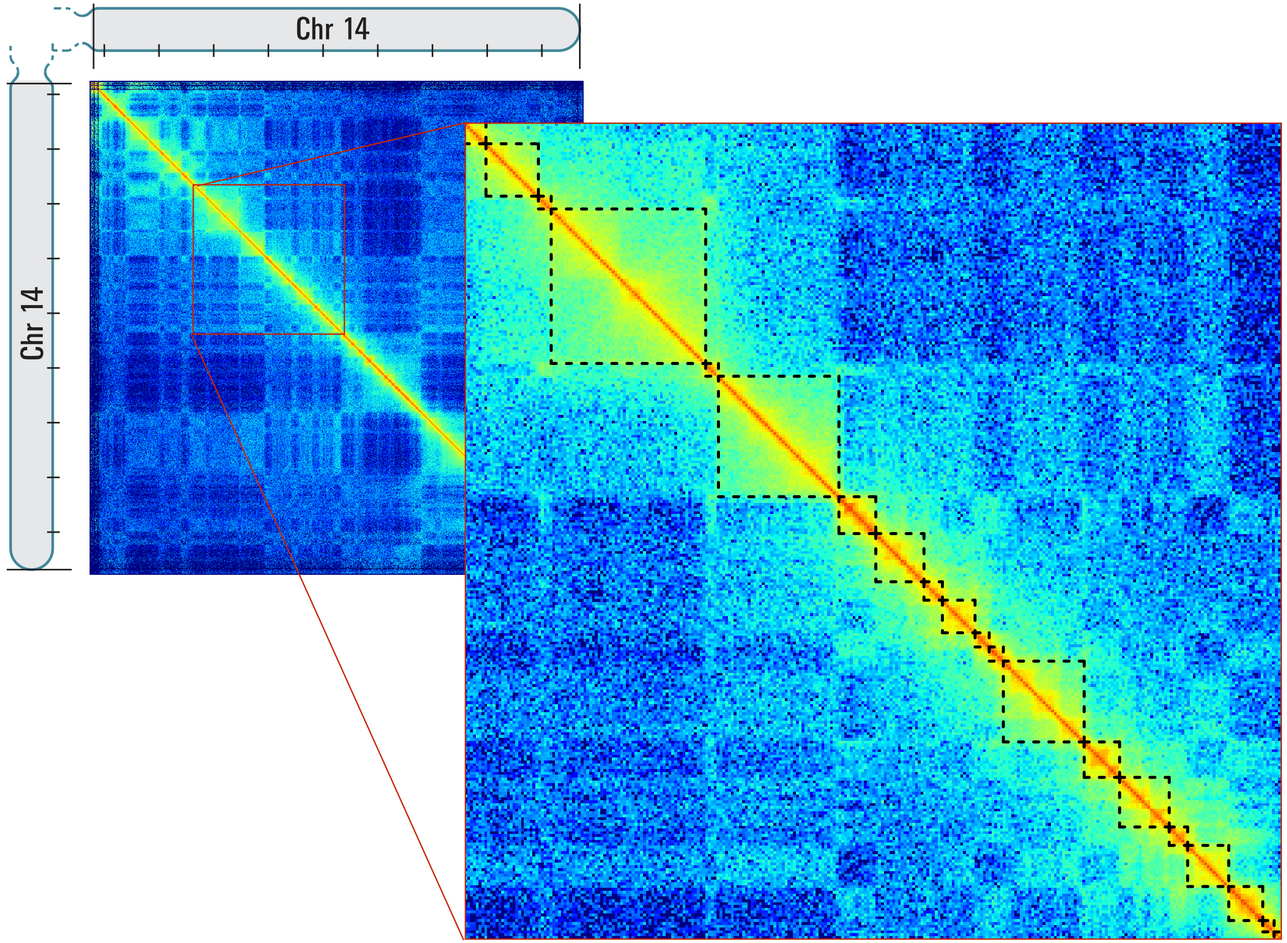
A/B Compartment

Chromosome 14



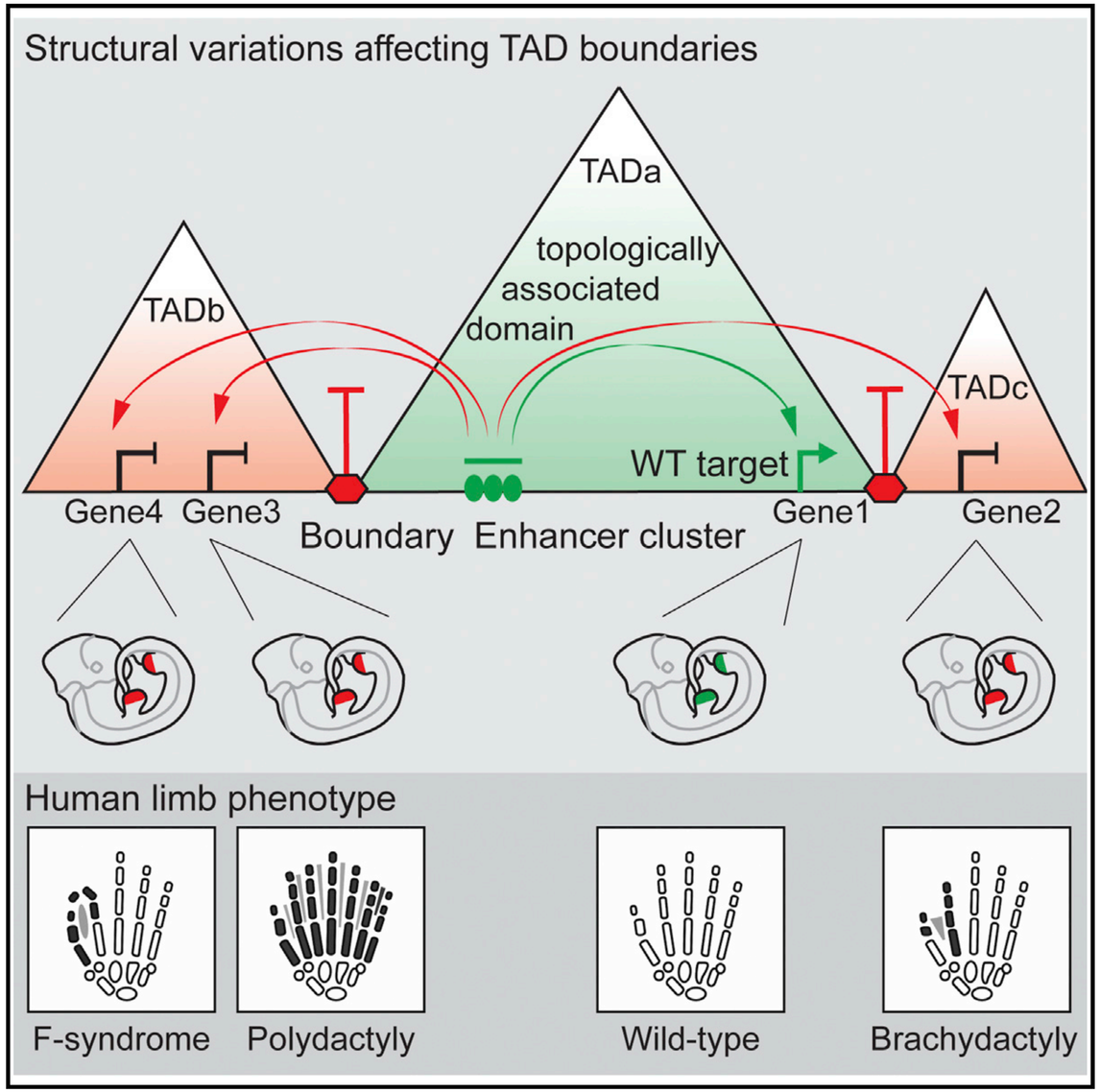
TADs

Chromosome 14



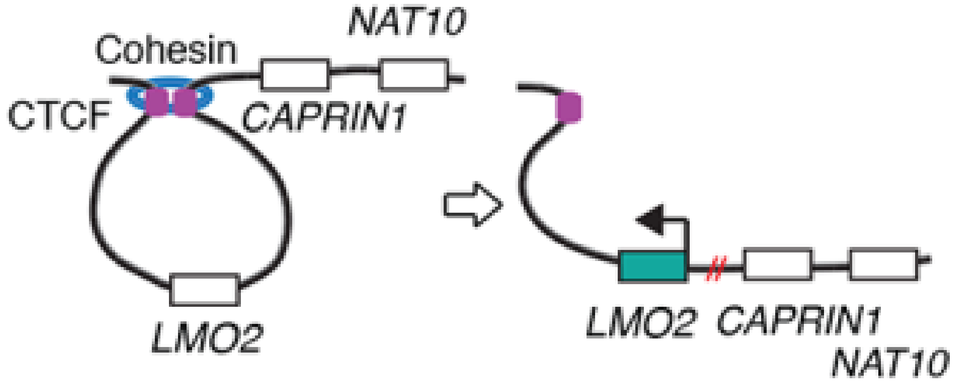
TADs are functional units

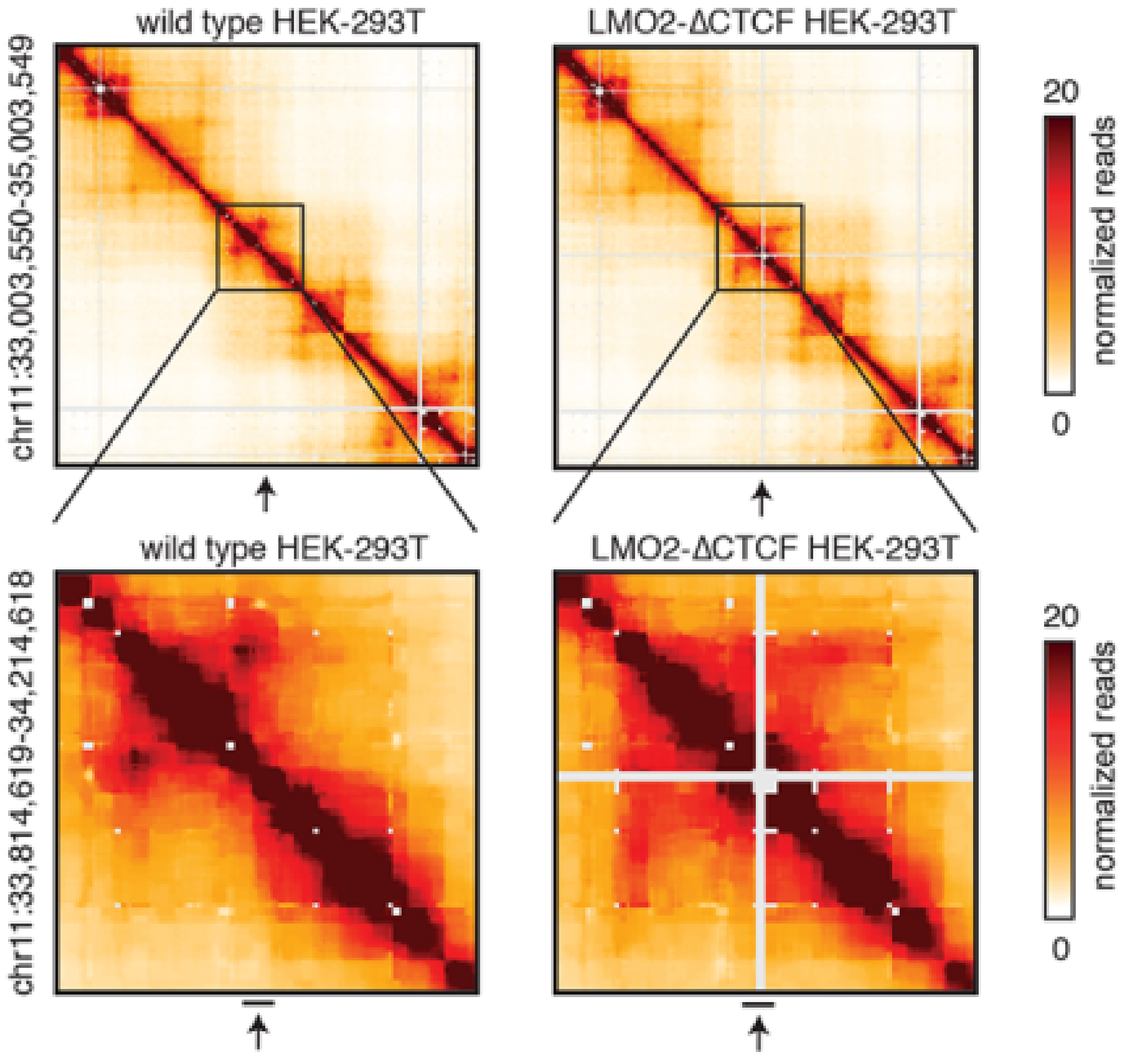
Lupiáñez, et al. (2015). Cell, 1–15.



TADs are functional units

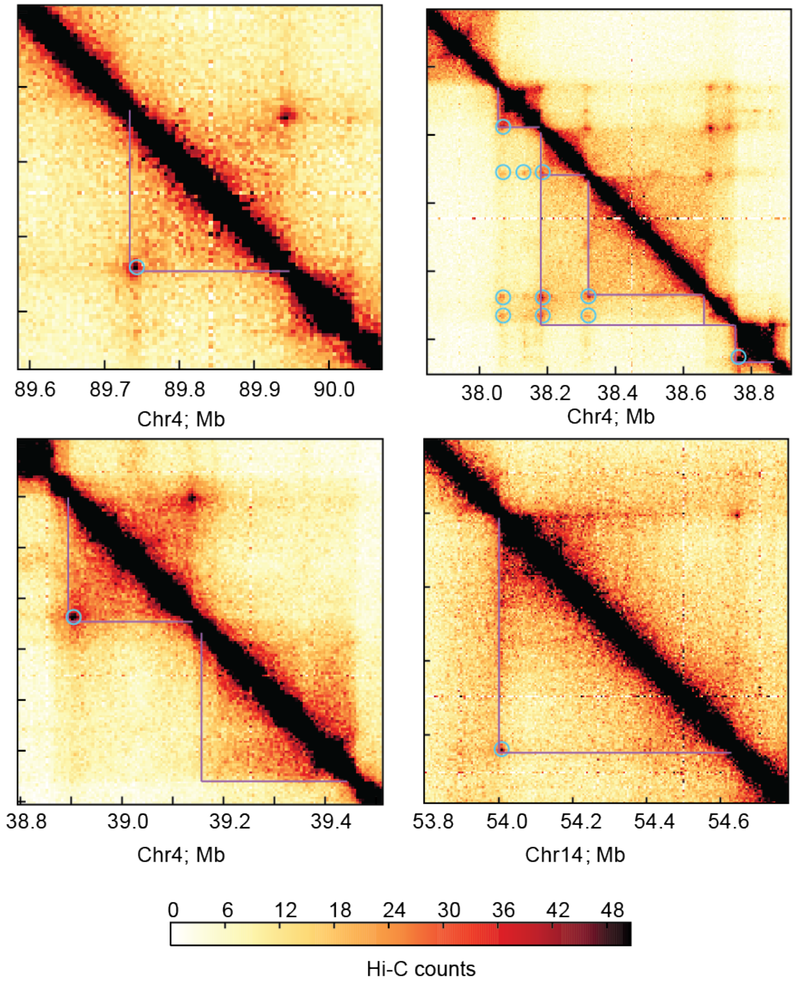
Hnisz, D., et al. (2016). Science

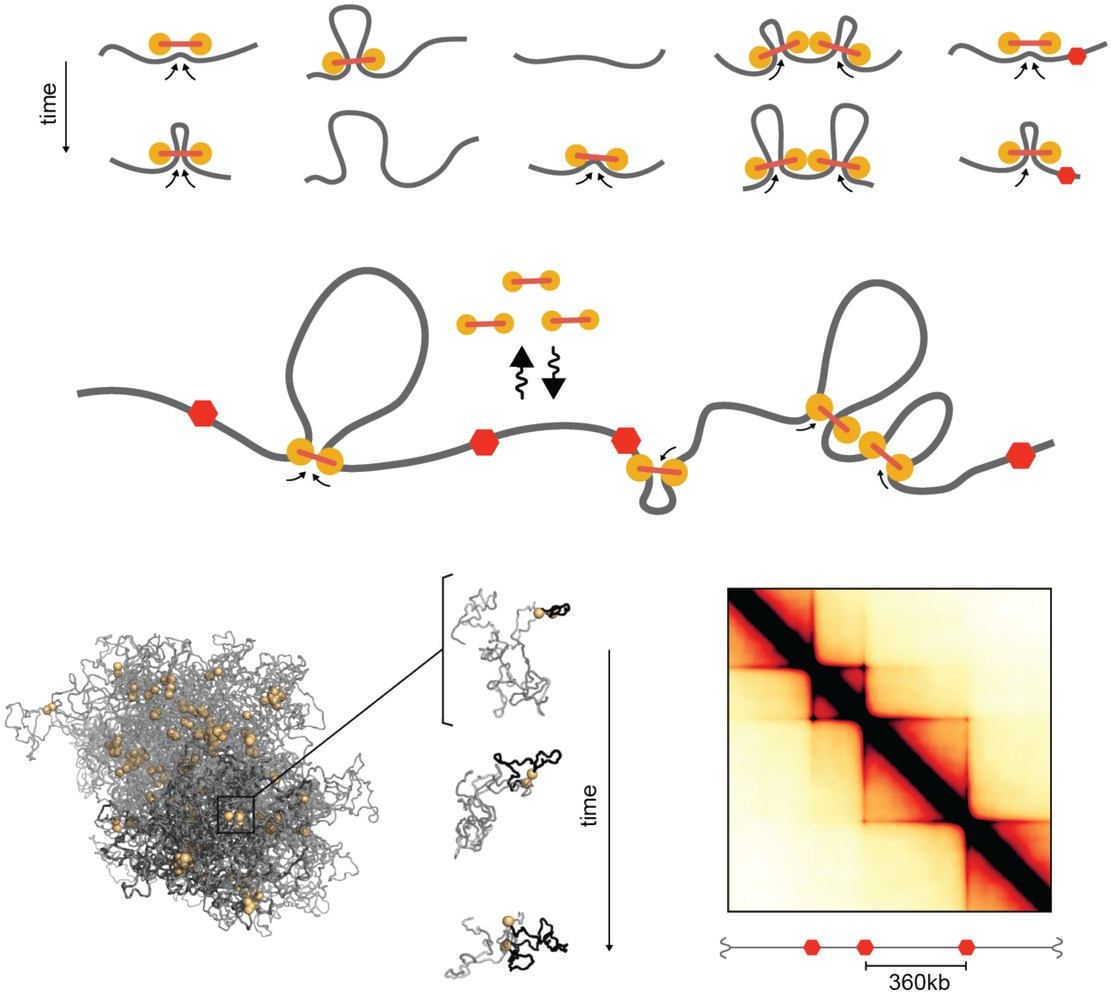




Level V: Loop-extrusion as a driving force

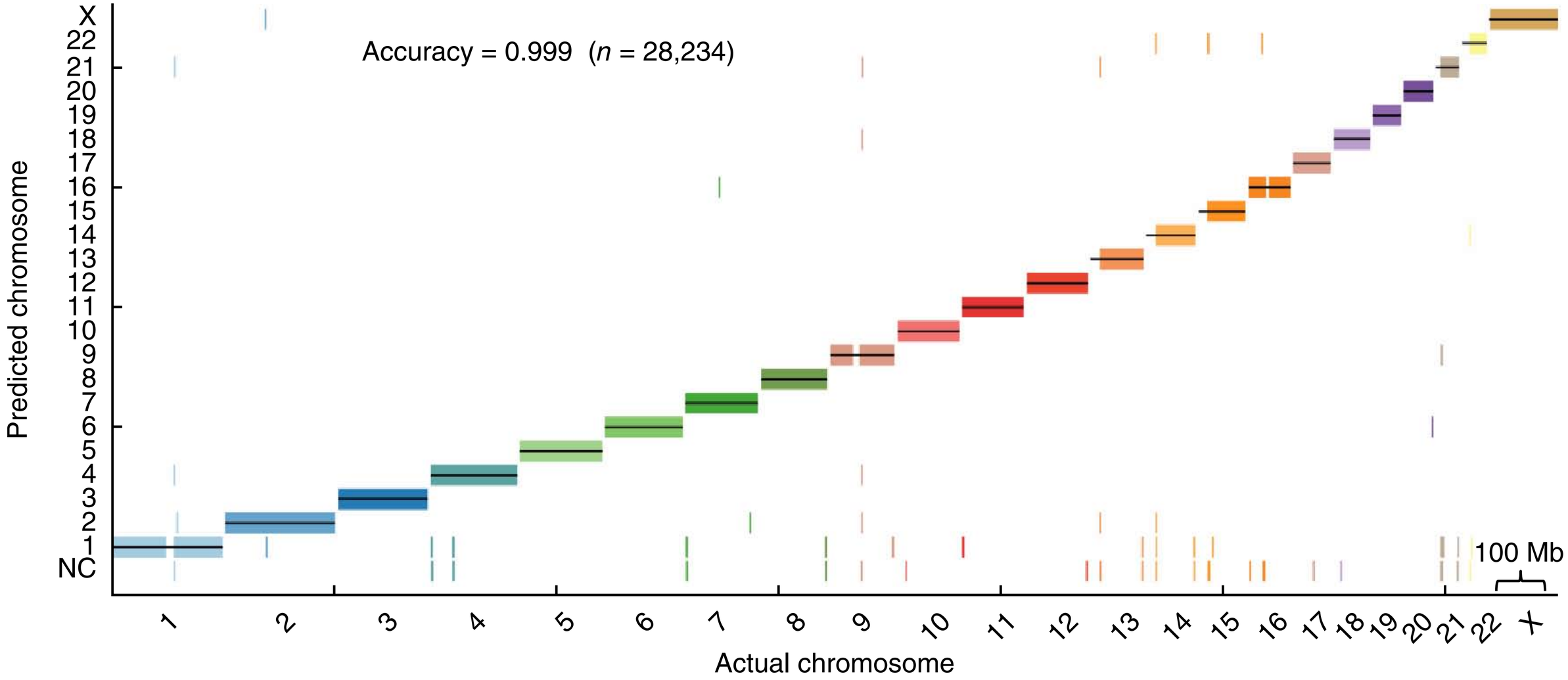
Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2015).
Formation of Chromosomal Domains by Loop Extrusion. bioRxiv.





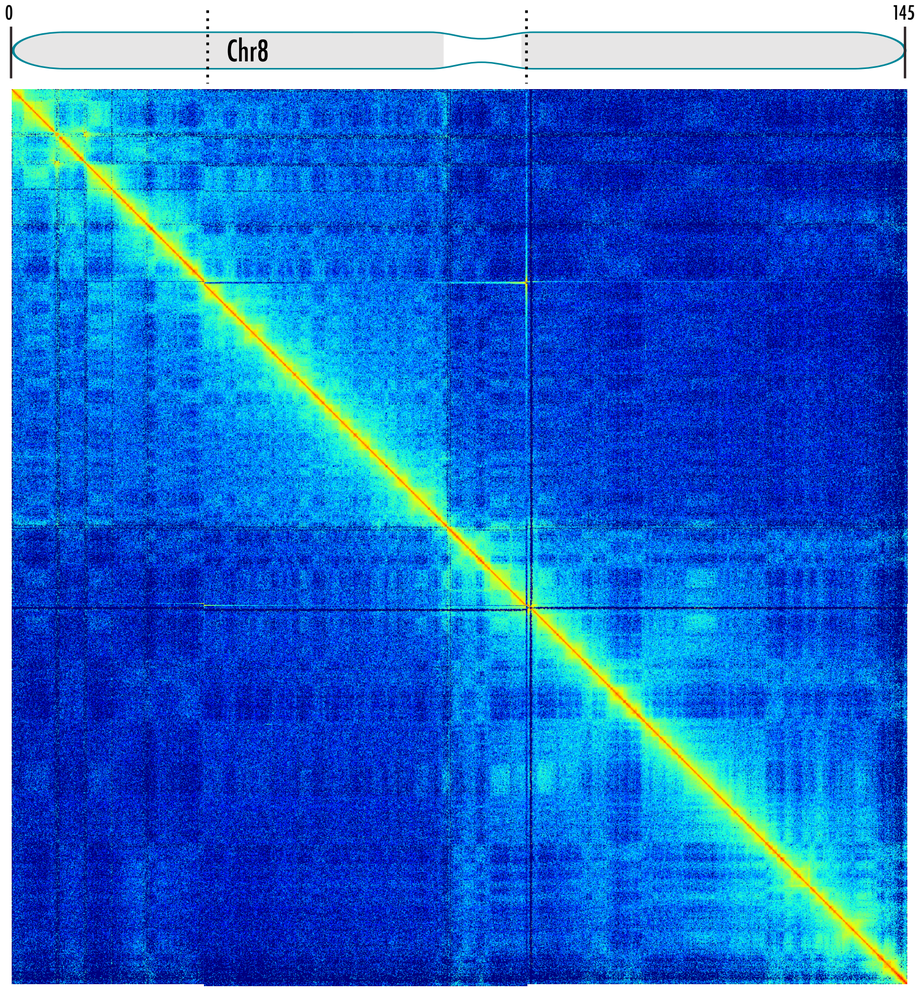
Hi-C for de-novo assembly

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

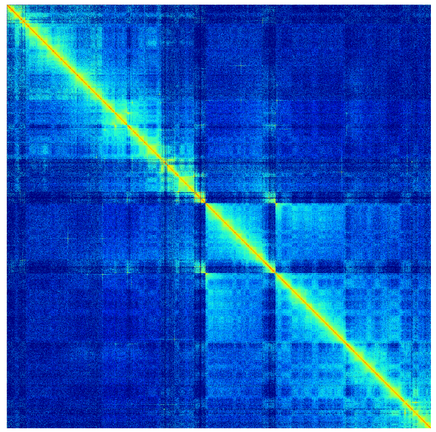


Assembly error detection

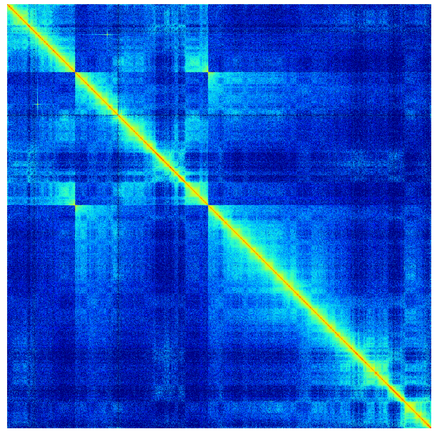
Chromosome 8 Gorilla



Chr 7



Chr 12

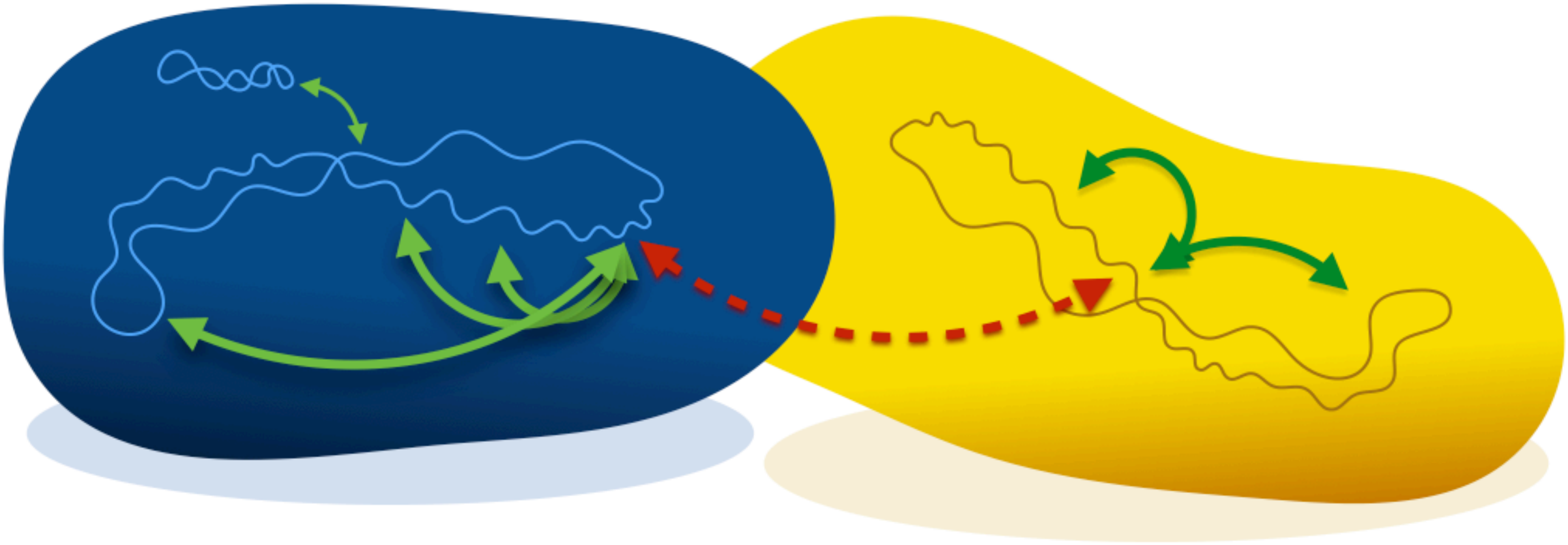


GGO8 has an inversion of the region corresponding to HSA8:30.0-86.9Mb

Aylwyn Scally (Department of Genetics, University of Cambridge)

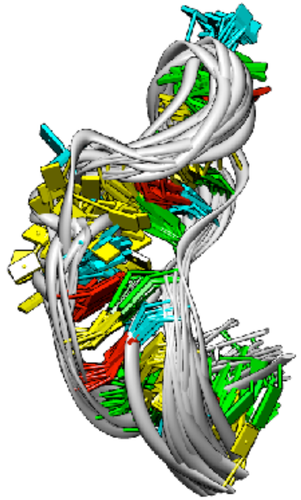
Hi-C for meta genomics

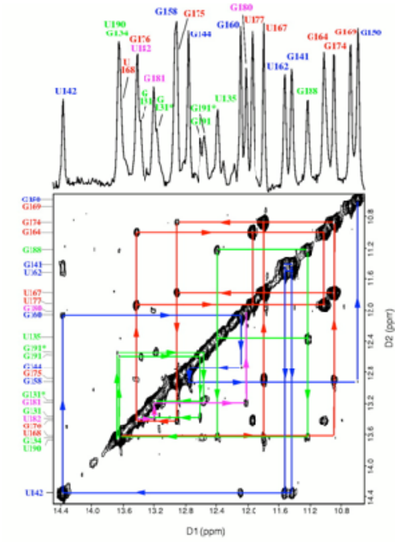
Beitel, C. W., Froenicke, L., Lang, J. M., Korf, I. F., Michelmore, R. W., Eisen, J. A., & Darling, A. E. (2014). Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. doi:10.7287/peerj.preprints.260v1



Restraint-based Modeling

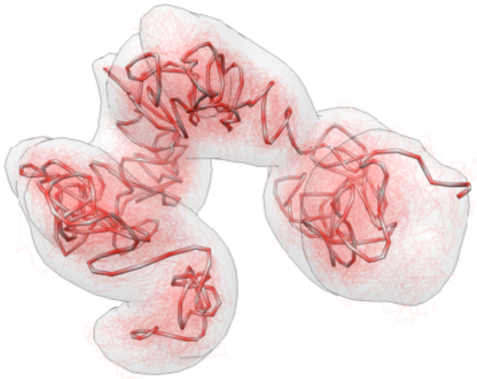
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).

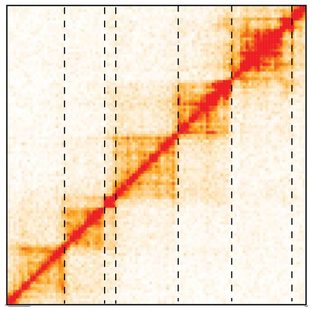




Biomolecular structure determination

2D-NOESY data





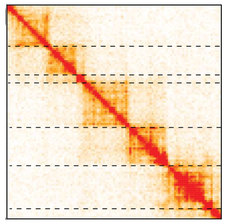
Chromosome structure determination

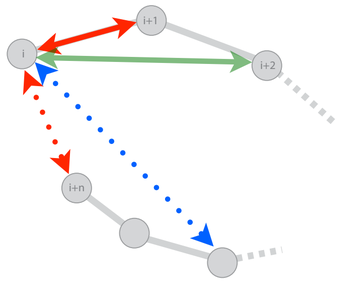
3C-based data

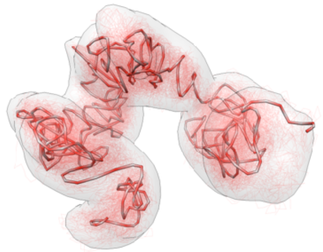
TADbit

http://3DGenomes.org

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





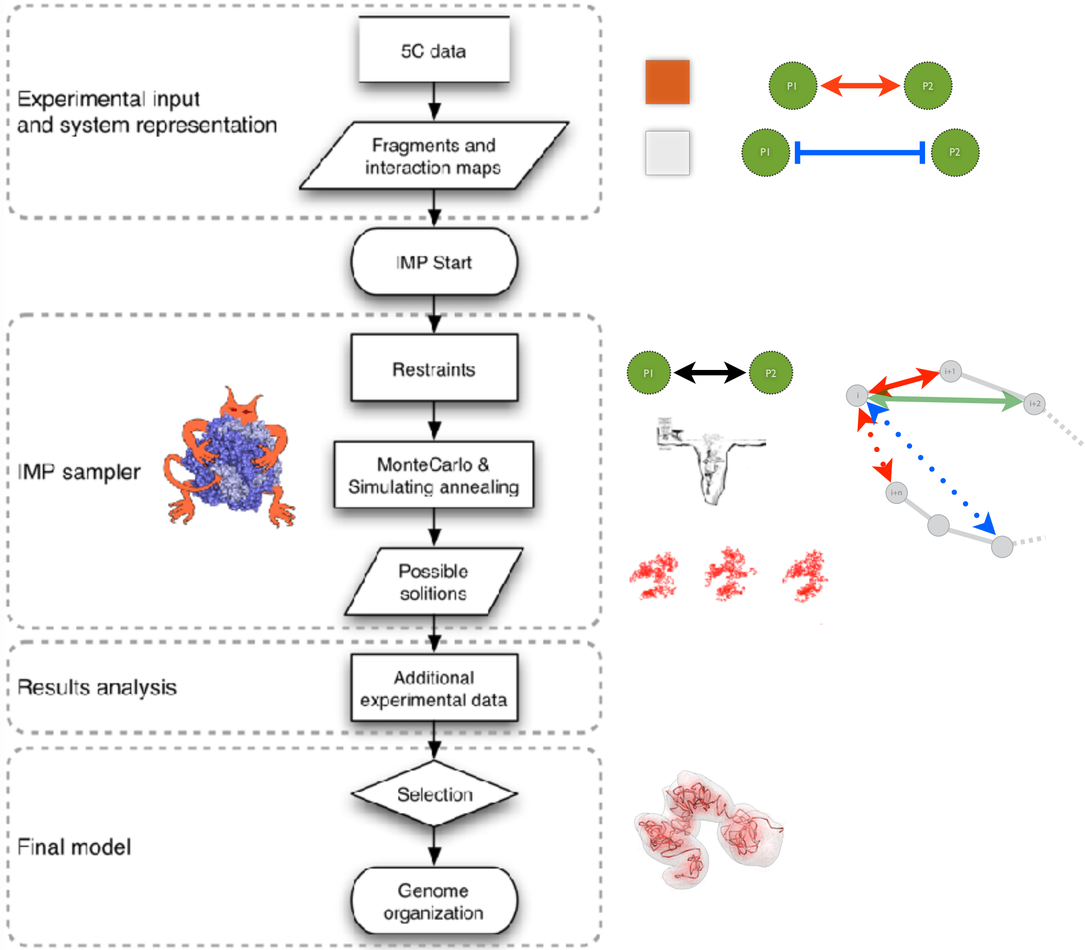


FastQ files to Maps

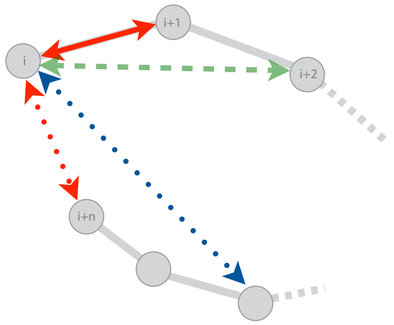
Map analysis

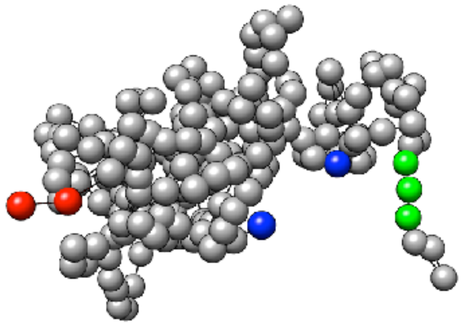
Model building

Model analysis



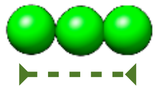
Model representation and scoring





d = d_0

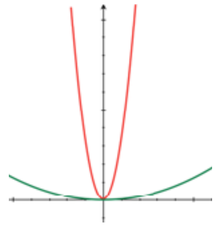
d < d_0



d > d_0

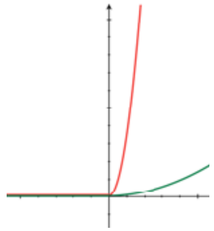
Harmonic

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



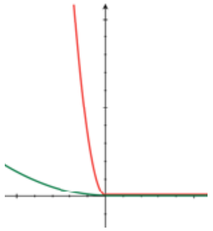
Harmonic Upper Bound

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

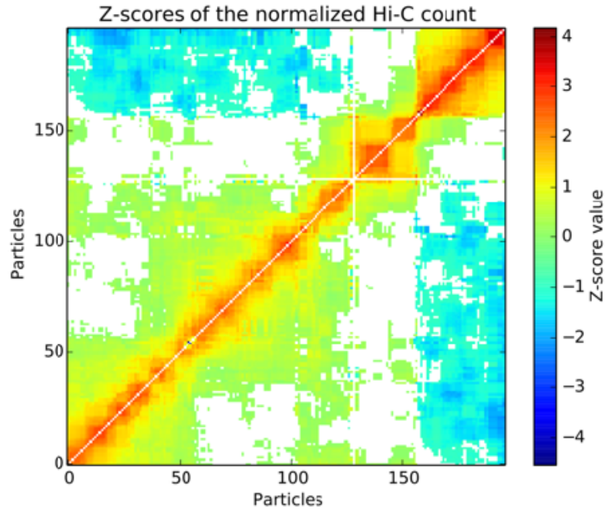


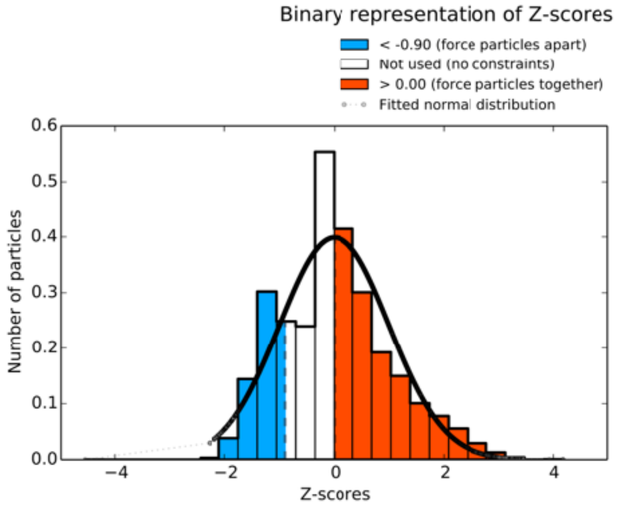
Harmonic Lower Bound

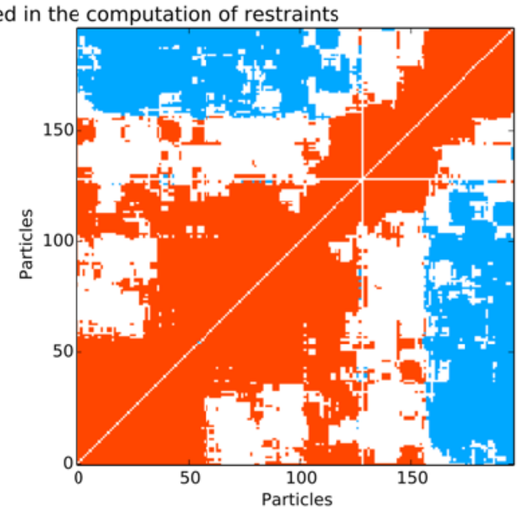
$$\begin{cases} if\ \ d_{i,j} \leq d^0_{i,j}; & lbH_{i,j} = k\left(d_{i,j} - d^0_{i,j}\right)^2 \\ if\ \ d_{i,j} > d^0_{i,j}; & lbH_{i,j} = 0 \end{cases}$$

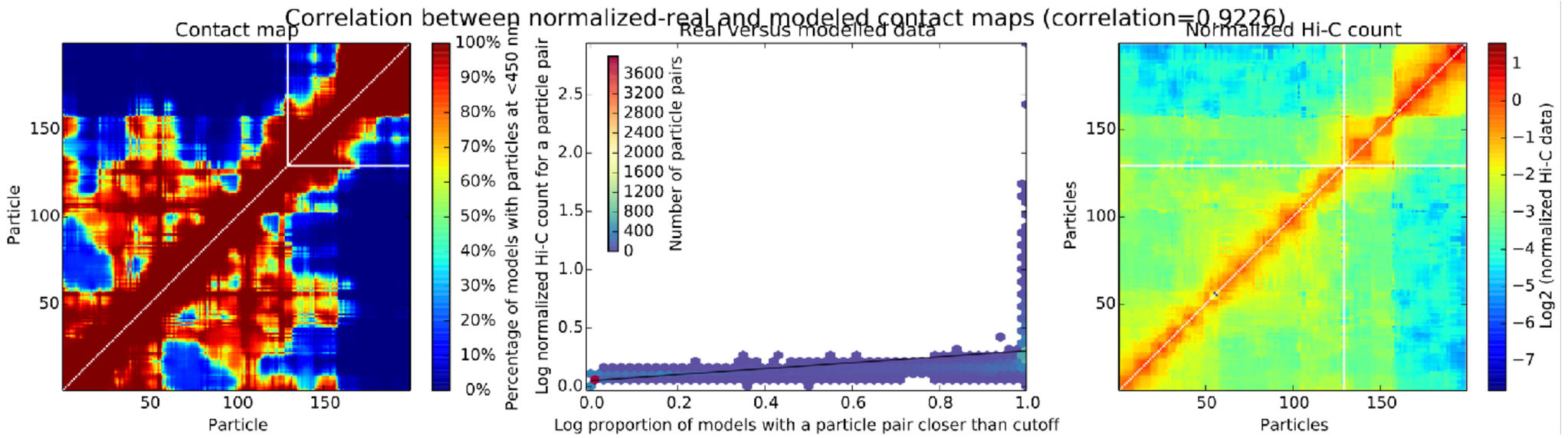


Parameter optimization

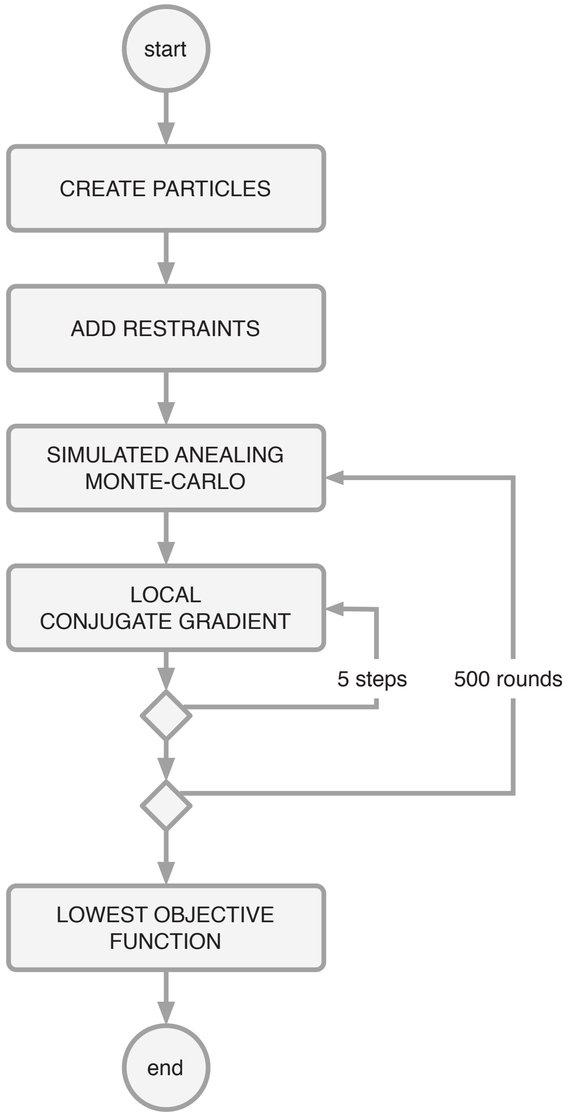


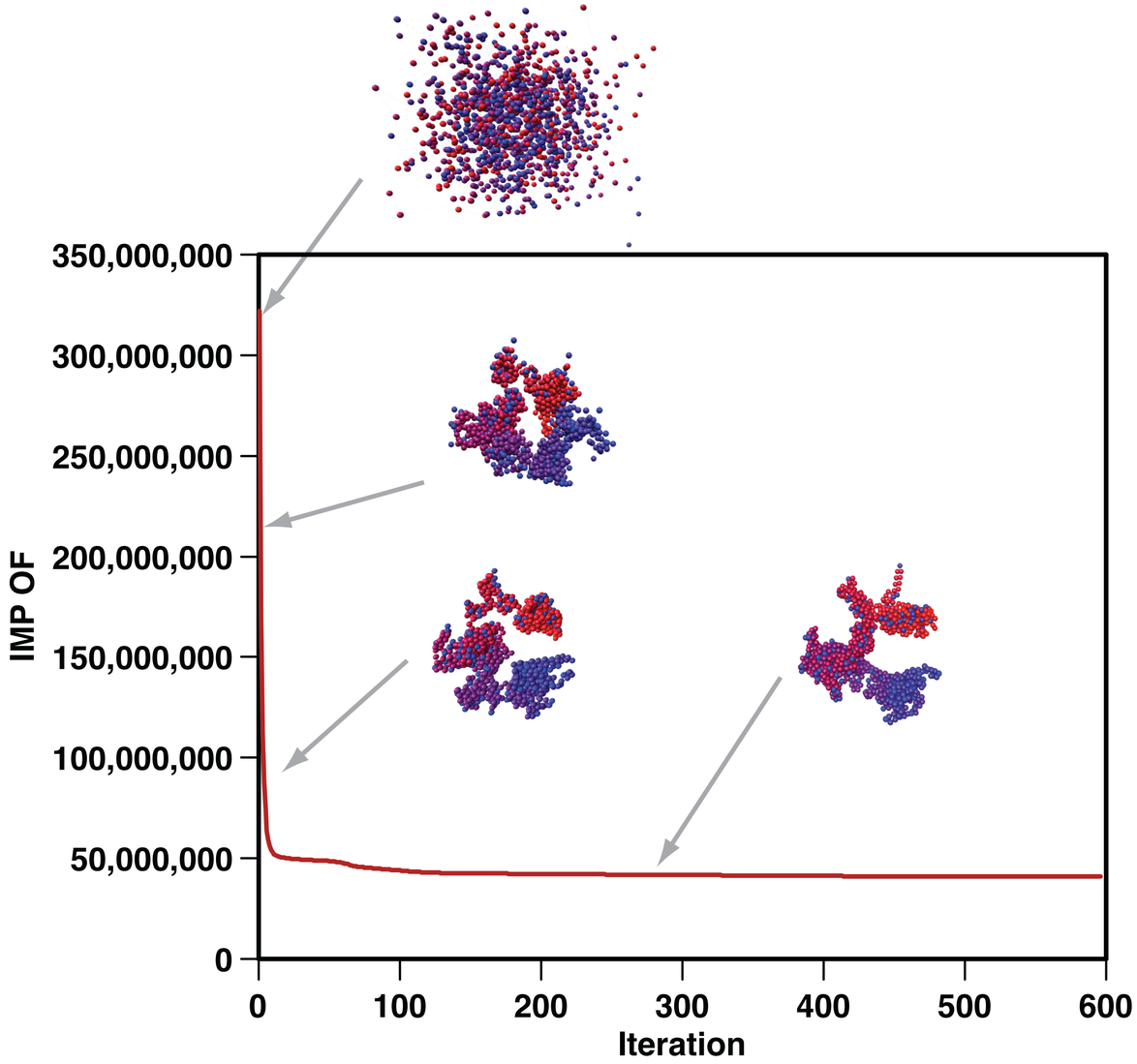




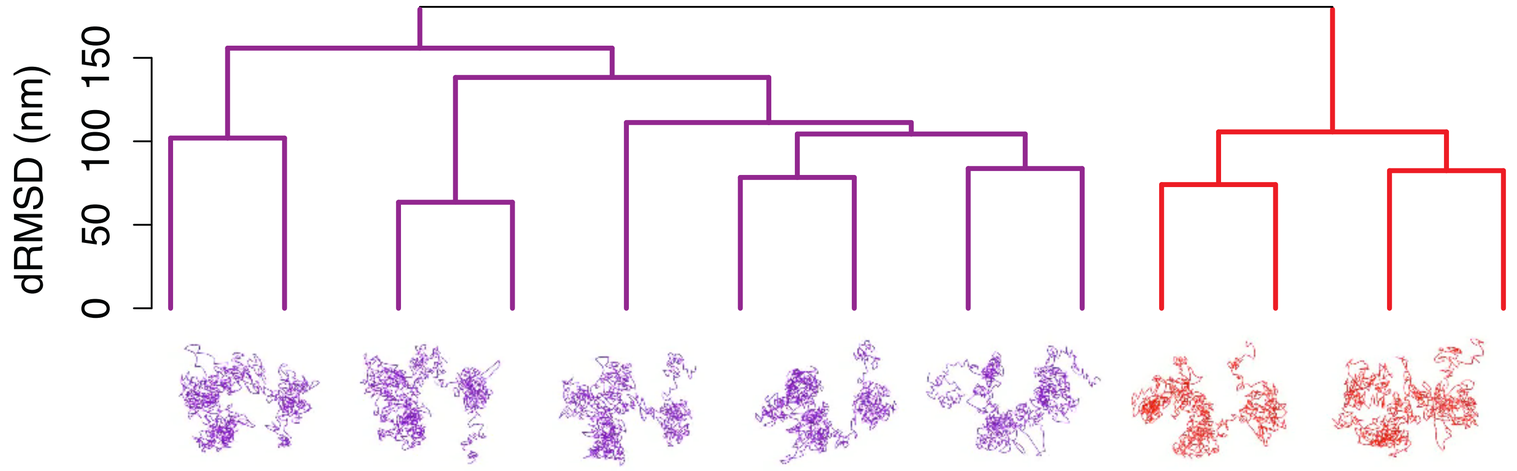


Optimization of the scoring function

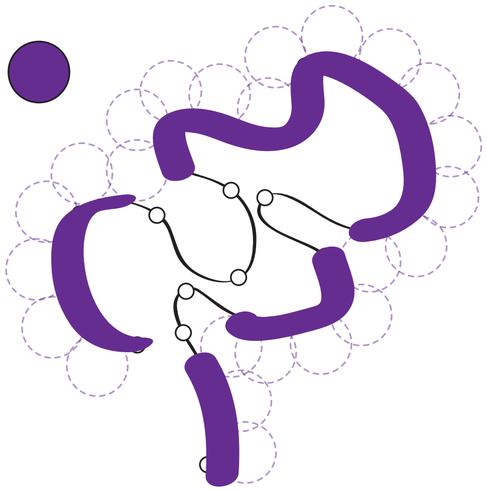




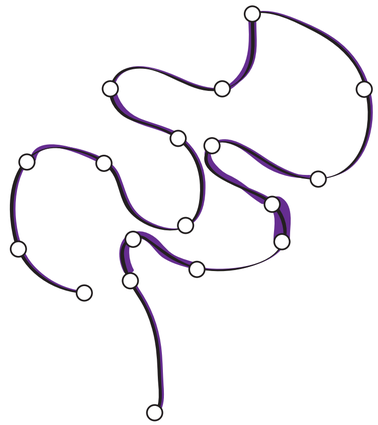
Model analysis: clustering and structural features



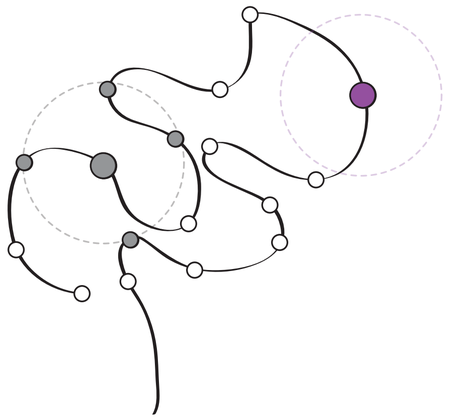
Accessibility (%)



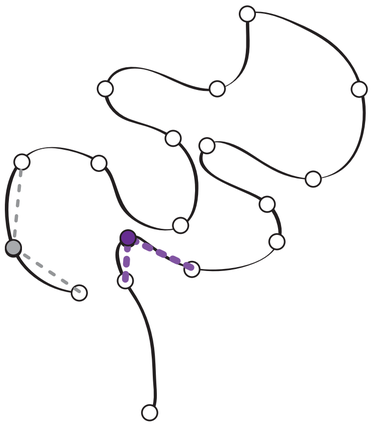
Density (bp/nm)



Interactions

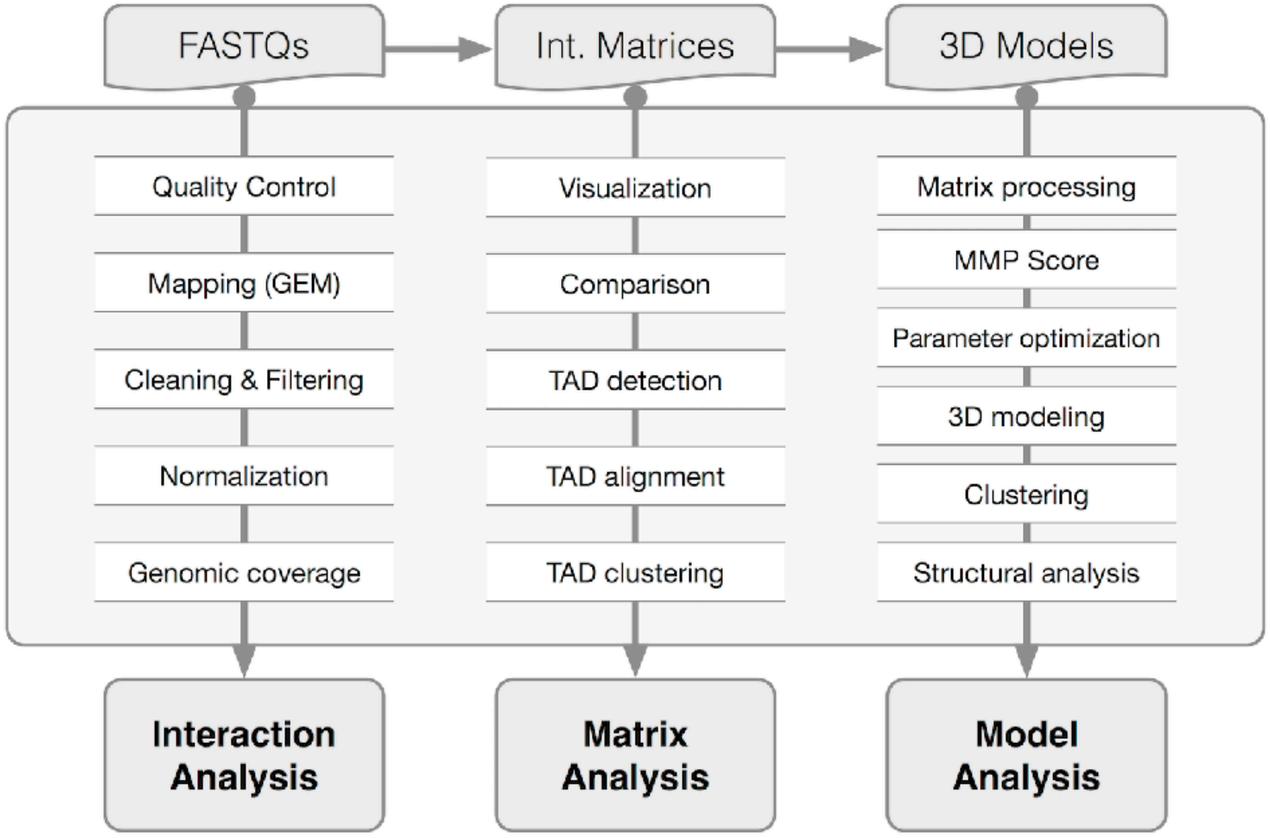


Angle



TADbit

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



- 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) in press
- Pascual-Reguant, L. et al. Nature Comm. (2018) in press

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	$\sum_{(i,j)D_{ij}<\infty)} \frac{(r_{ij}^2 - D_{ij}^2)}{D_{ij}} - \lambda \sum_{(i,j)} r_{ij}^2$ 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	$D_{ij} \propto \begin{cases} \alpha F_{ij} + \beta & \text{if } F_{ij} < \gamma' \text{ or } F_{ij} > \gamma \\ \frac{s_i + s_j}{2} & \text{if } \lvert i-j \rvert = 1 \end{cases}$ 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 $\lvert i-j \rvert = 1$ 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	$\overline{F_{\lvert i-j \rvert}} = \frac{\sum_{k=0}^{N-1-\lvert i-j \rvert} F_{(k,k+\lvert i-j \rvert)}}{N-\lvert i-j \rvert}$ is the average of F_{ij} at genomic distance $\lvert i-j \rvert$ expressed in kb. $D_{ij} = \overline{F_{\lvert i-j \rvert}} \times 7.7 \times \lvert i-j \rvert$ 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 $\lvert i-j \rvert$	$\sum_{(i,j)} \frac{(r_{ij} - D_{ij})^2}{D_{ij}^2}$	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)

Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors

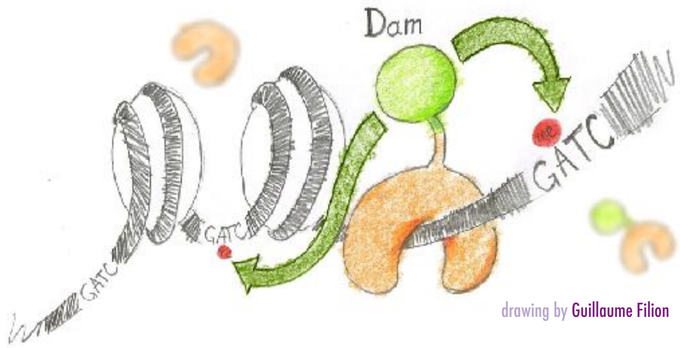
Marc A. Marti-Renom

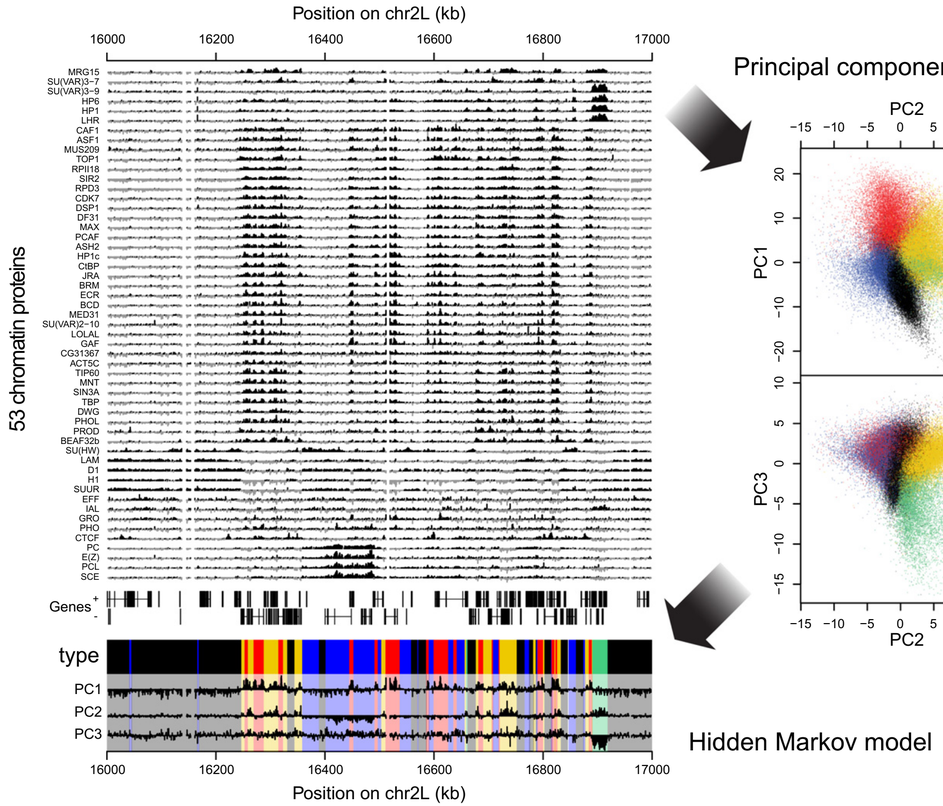
CNAG-CRG · ICREA

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

Fly Chromatin COLORs

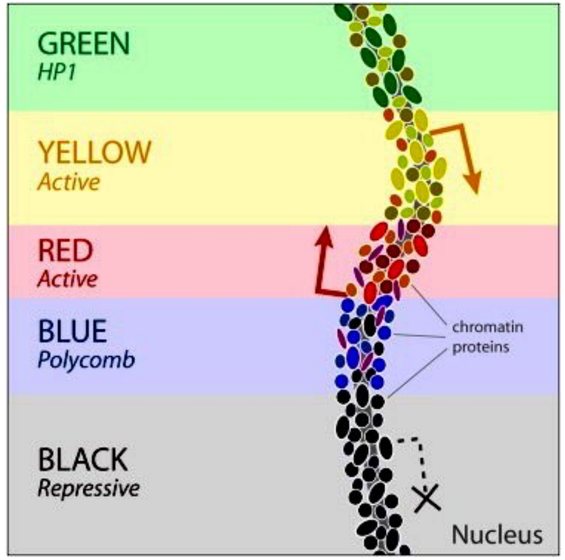
Filion et al. (2010). Cell, 143(2), 212–224.





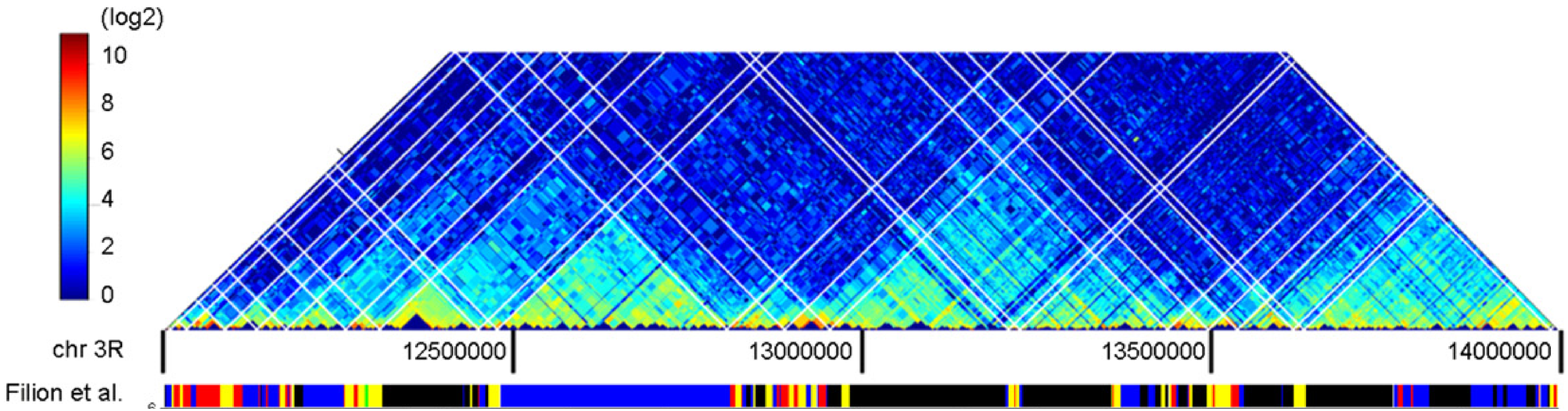
Principal component analysis

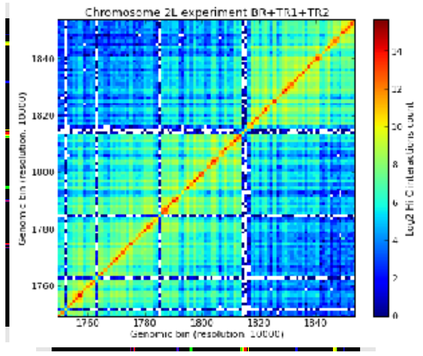
Hidden Markov model

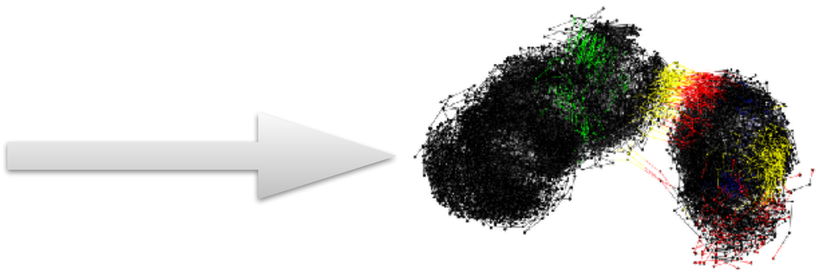


Fly Chromatin COLORs

Hou et al. (2012). Molecular Cell, 48(3), 471–484.





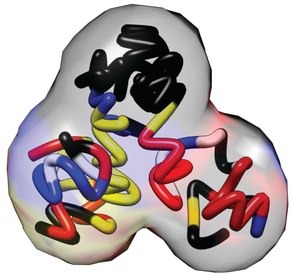


All fly genome
@2Kb resolution

Structural properties

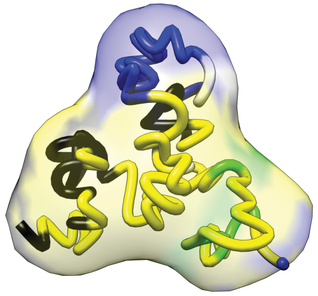
RED dense region
3R:18920000-19920000

22% 17% 0% 11% 45% 6%



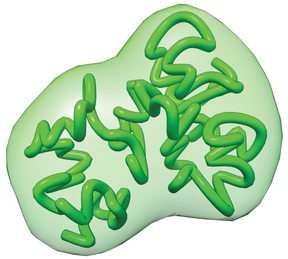
YELLOW dense region
X:15590000-16600000

0% 48% 4% 20% 26% 3%



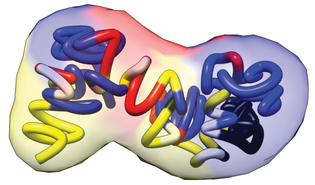
GREEN dense region
2R:510000-1530000

0% 0% 100% 0% 0% 0%



BLUE dense region
3L:210000-1230000

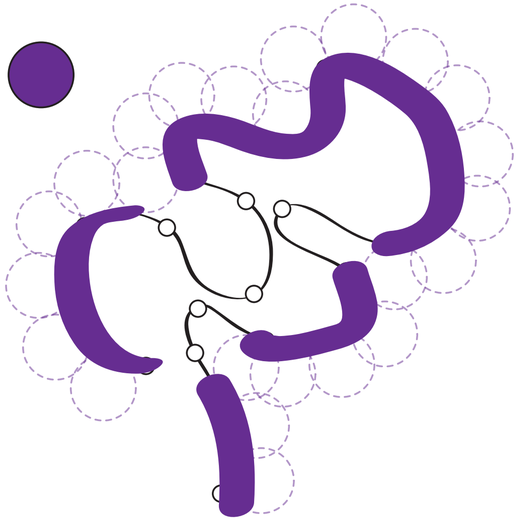
11% 17% 0% 52% 13% 0%



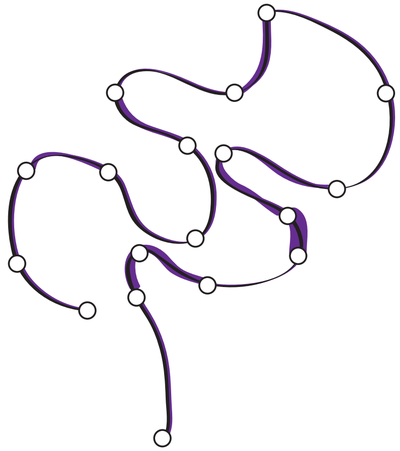
BLACK dense region
2L:17500000-18530000

1% 0% 0% 0% 98% 1%

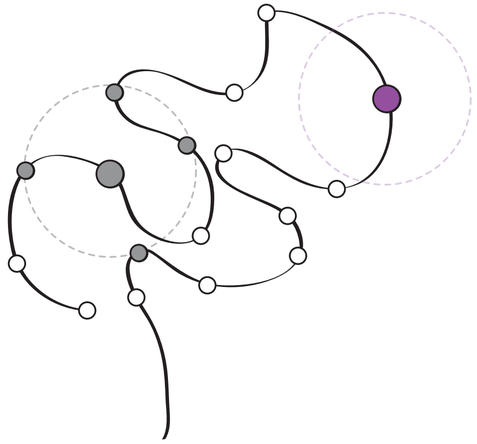
Accessibility (%)



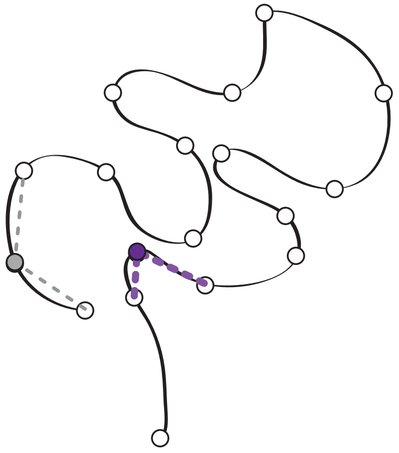
Density (bp/nm)



Interactions



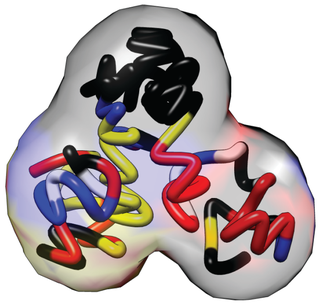
Angle



Structural COLORs

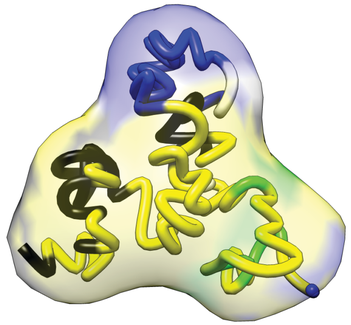
RED dense region
3R:18920000-19920000

22% 17% 0% 11% 45% 6%



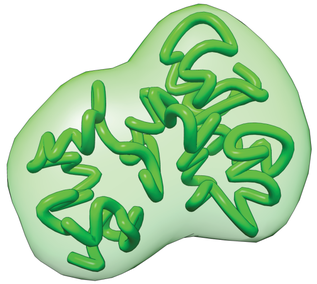
YELLOW dense region
X:15590000-16600000

0% 48% 4% 20% 26% 3%



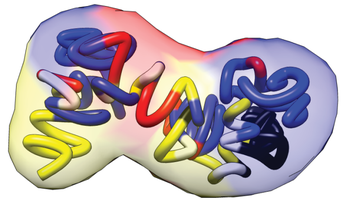
GREEN dense region
2R:510000-1530000

0% 0% 100% 0% 0% 0%



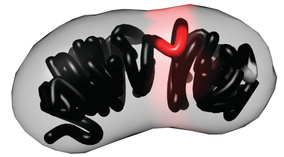
BLUE dense region
3L:210000-1230000

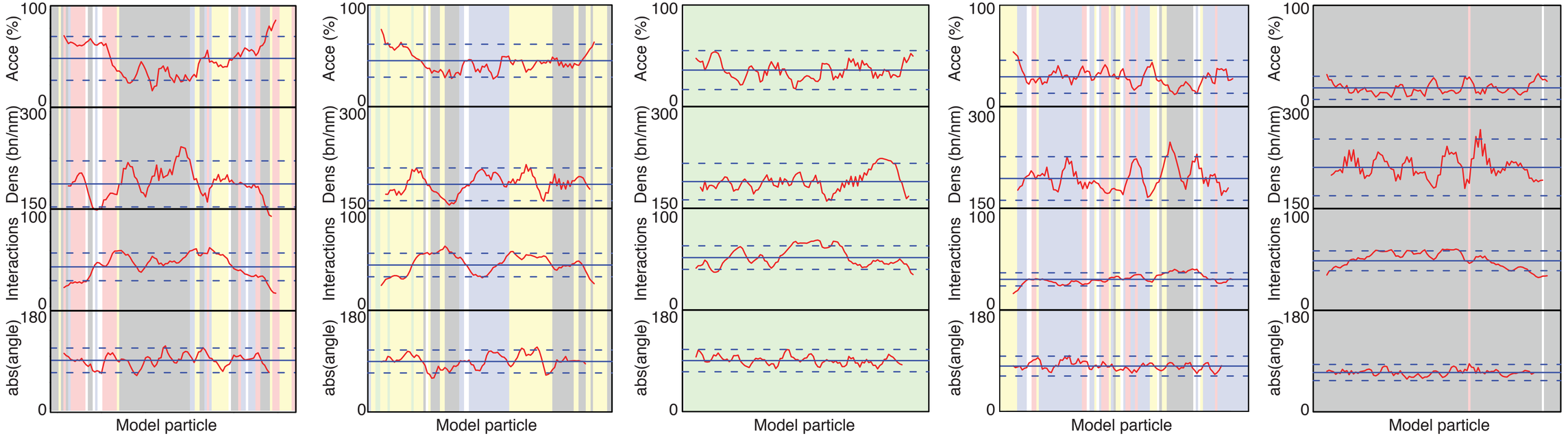
11% 17% 0% 52% 13% 0%



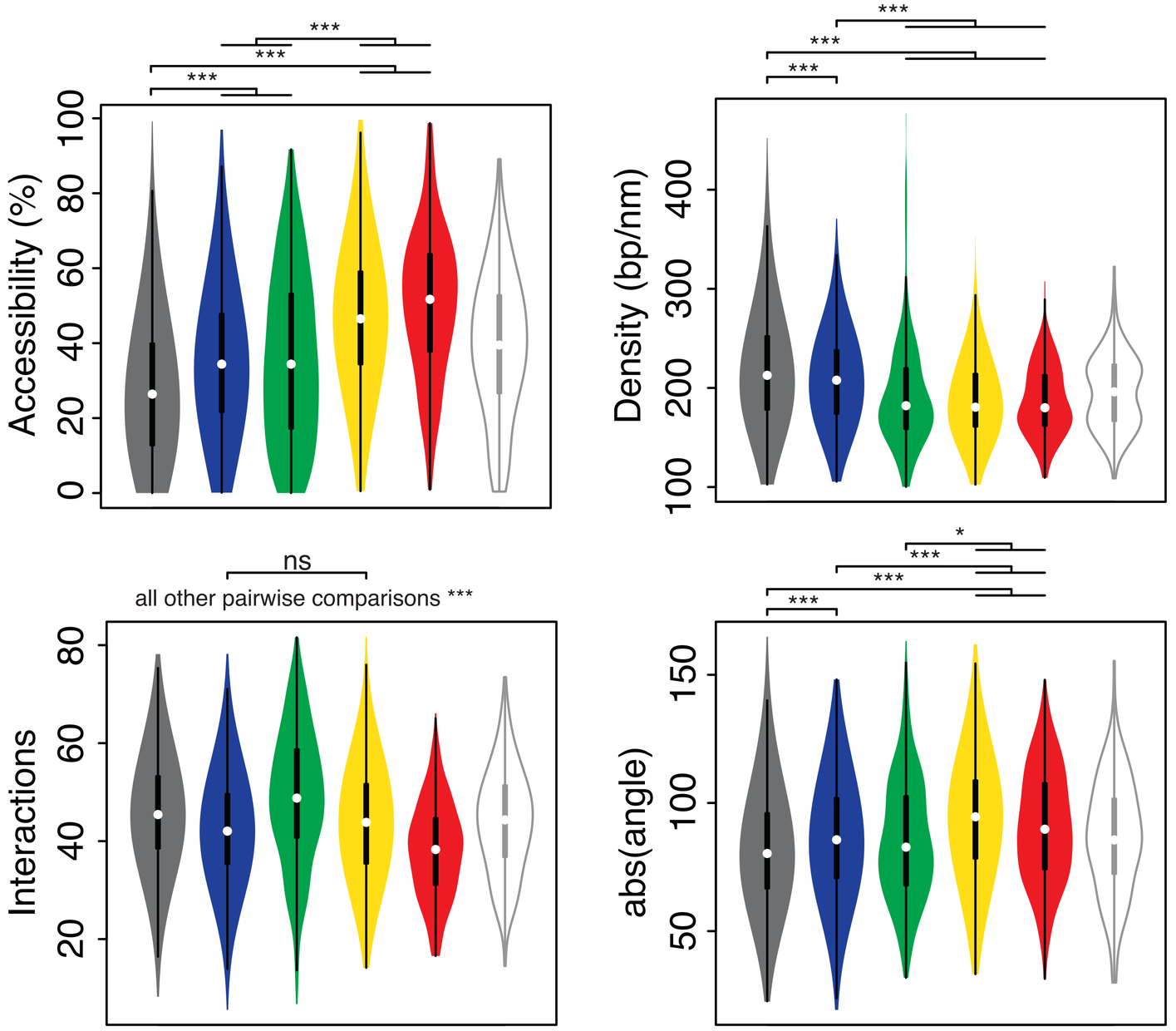
BLACK dense region
2L:17500000-18530000

1% 0% 0% 0% 98% 1%

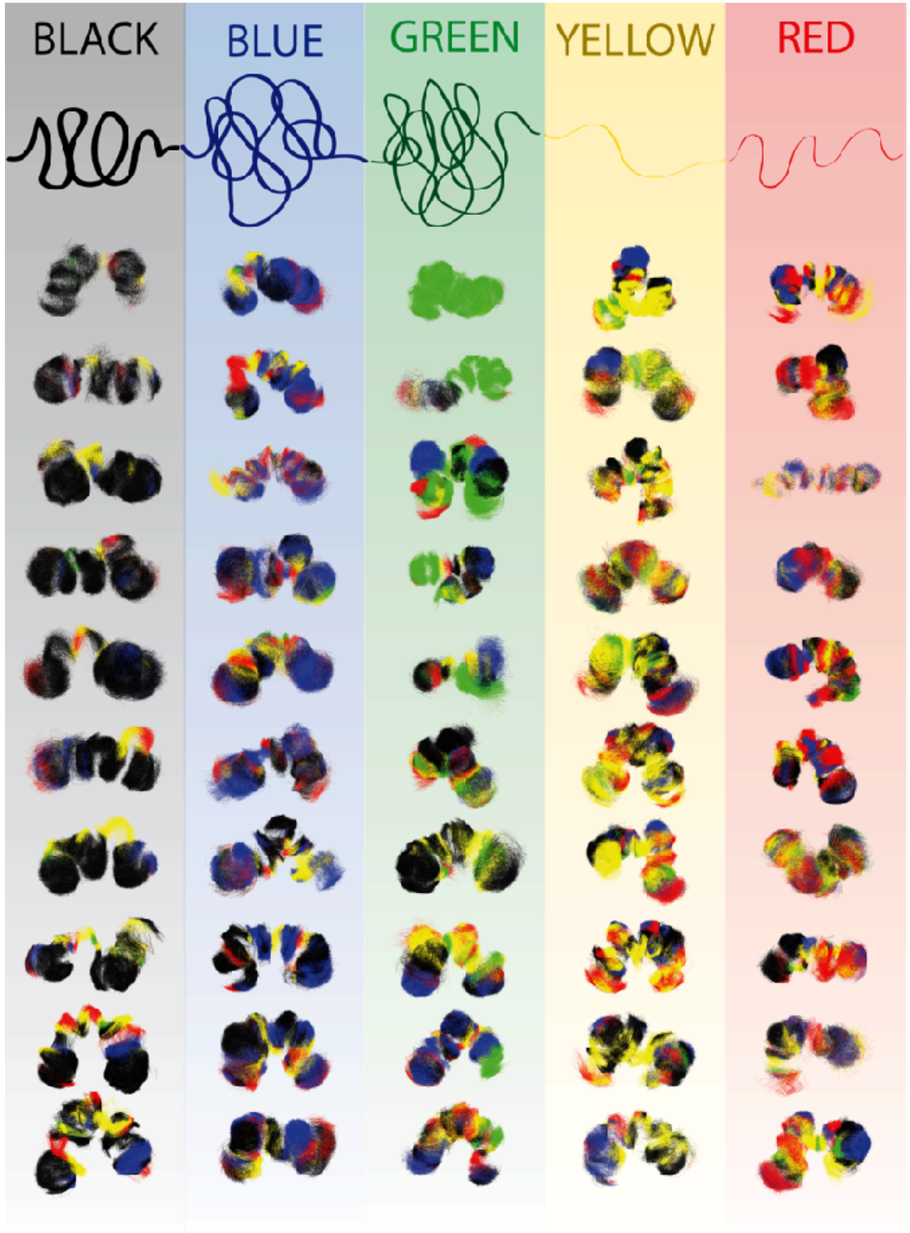


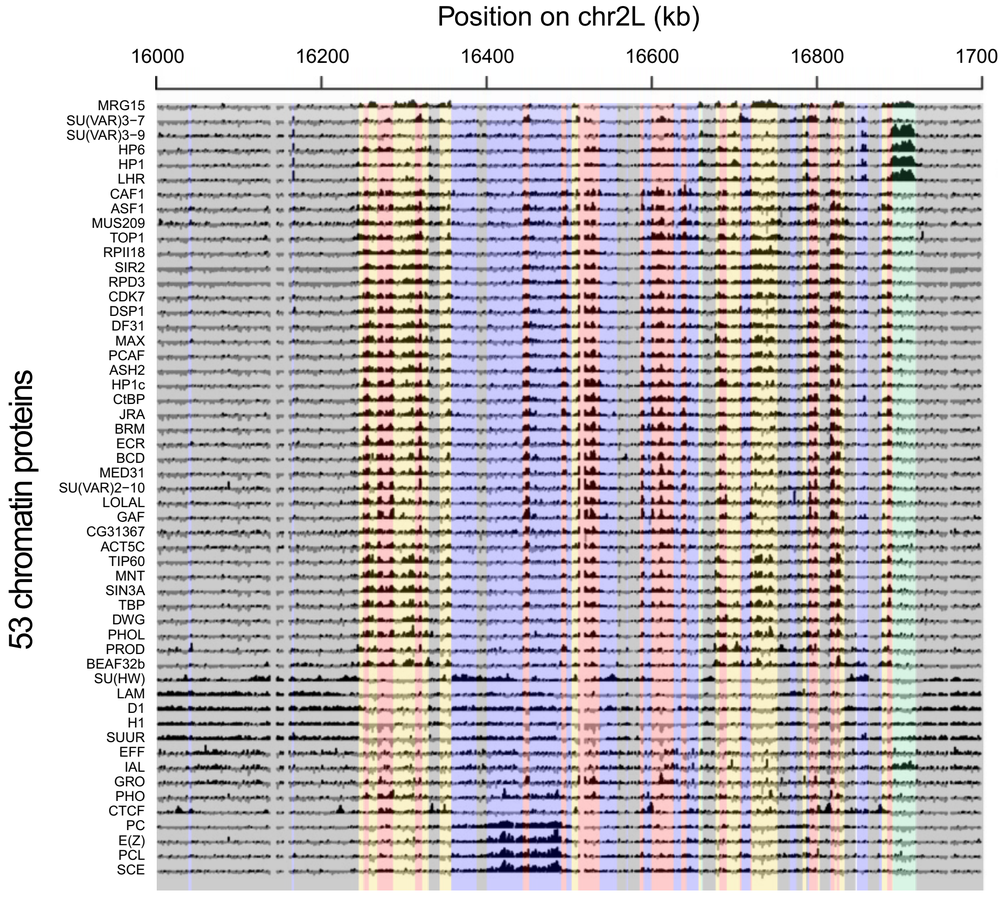


Structural COLORs



Structural COLORs





3D structural dynamics of the SOX2 locus activation

Marc A. Marti-Renom

CNAG-CRG · ICREA

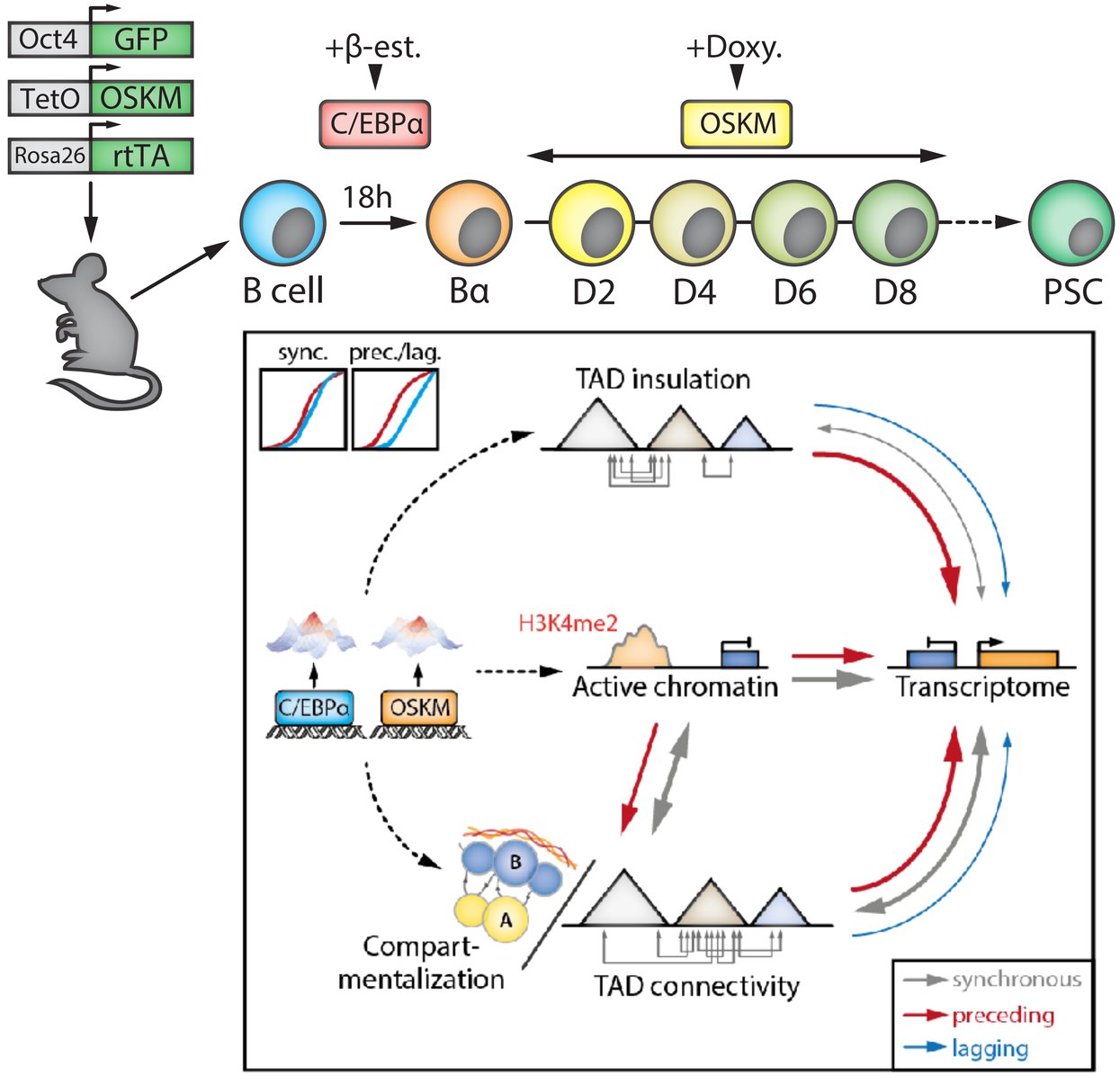
http://marciuslab.org
http://3DGenomes.org
http://cnag.crg.eu

Marco di Stefano
CNAG-CRG

Ralph Stadhouders, Enrique Vidal & Thomas Graf
CRG

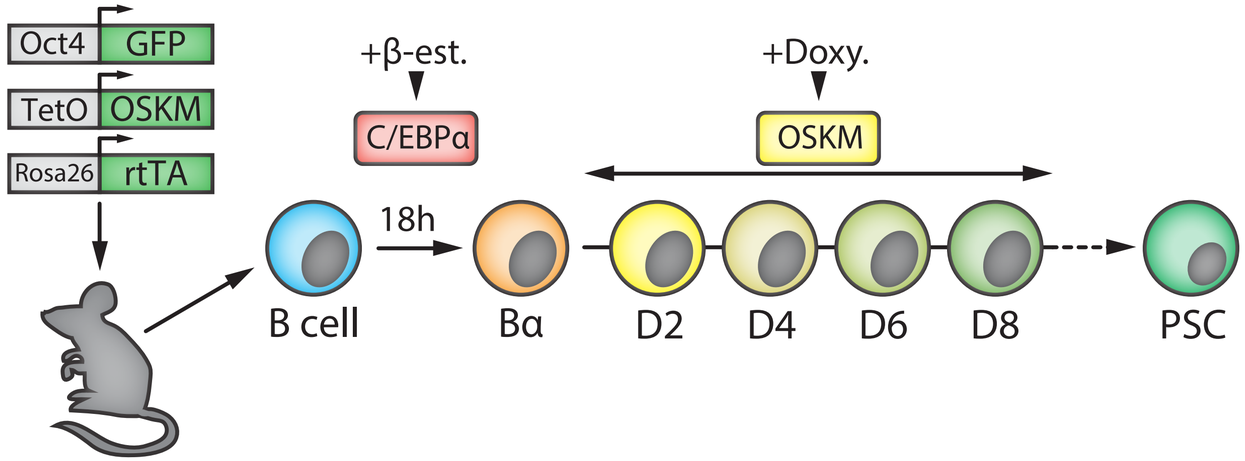
Reprogramming from B to PSC

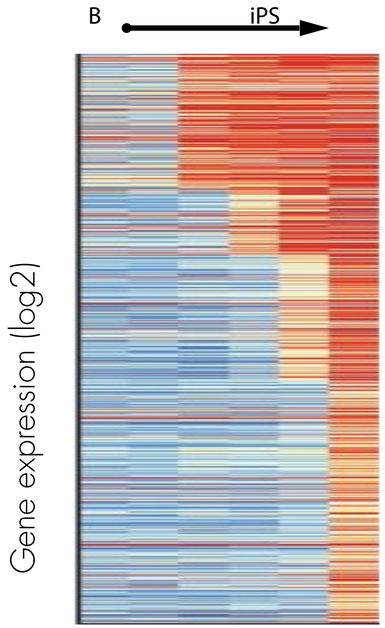
Stadhouders, R., Vidal, E. et al. (2018) Nature Genetics

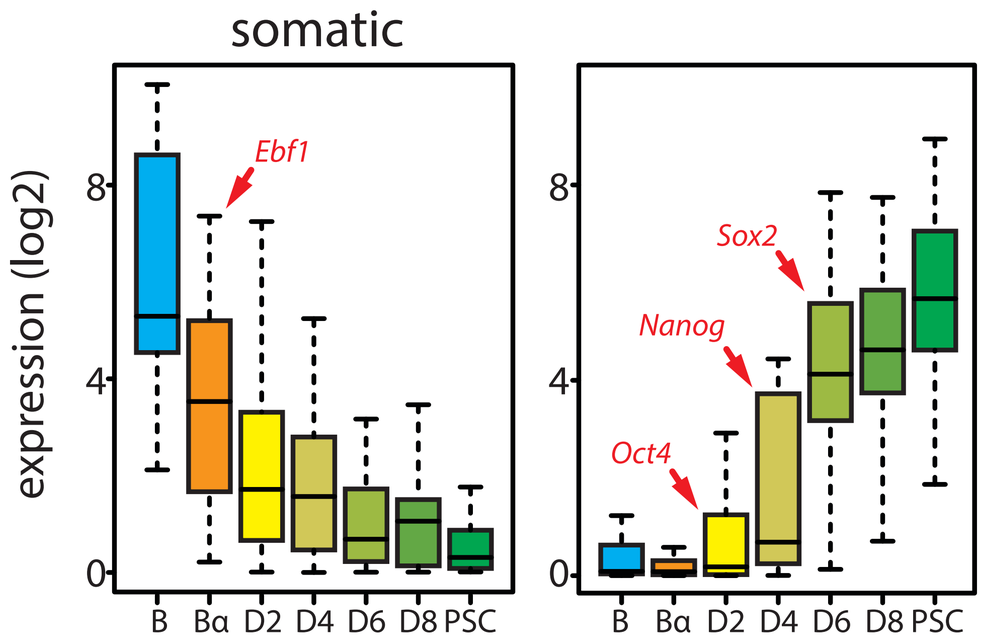


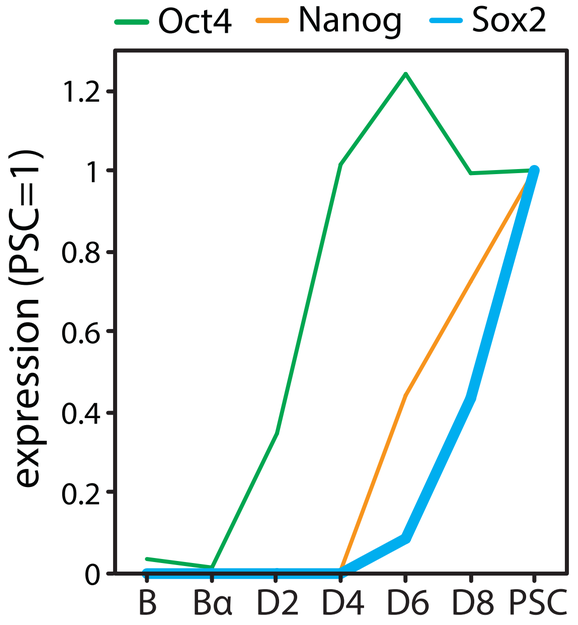
Reprogramming from B to PSC

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



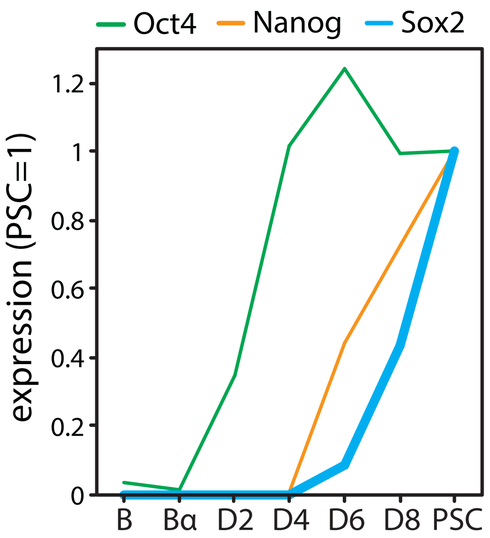


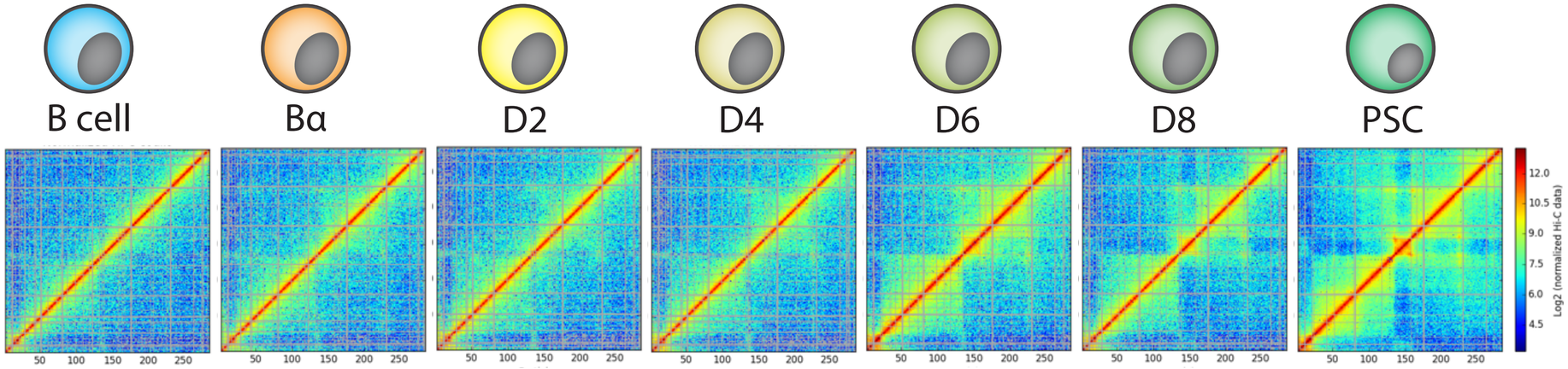




Hi-C maps of reprogramming from B to PSC

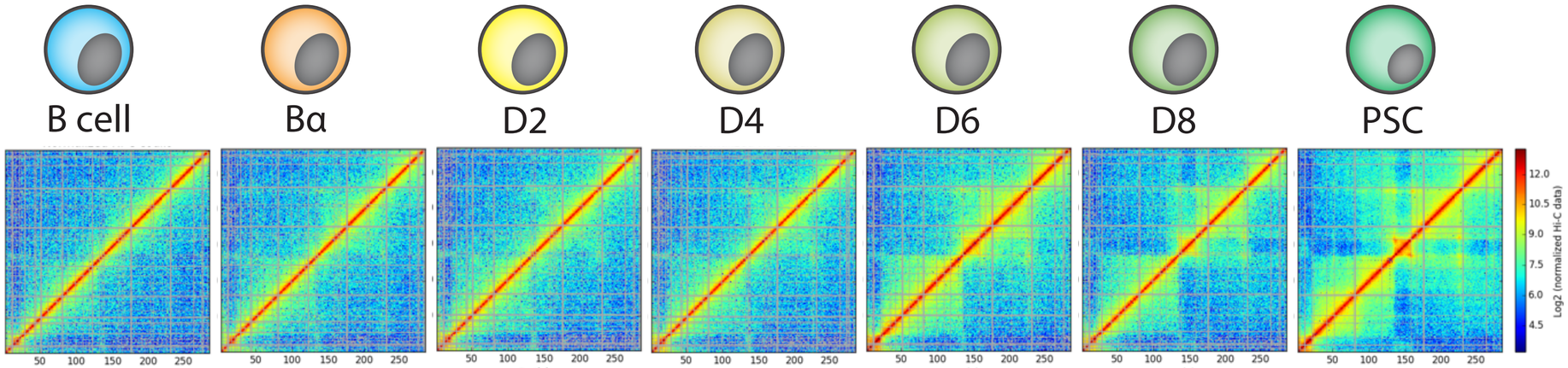
The SOX2 locus





Hi-C maps of reprogramming from B to PSC

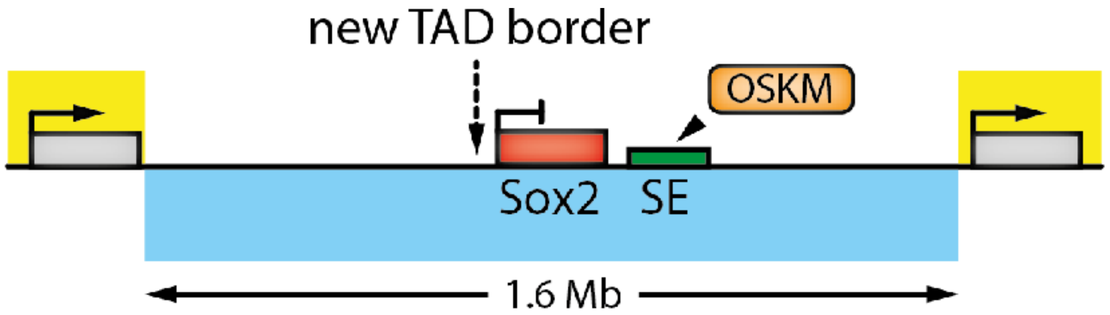
The SOX2 locus

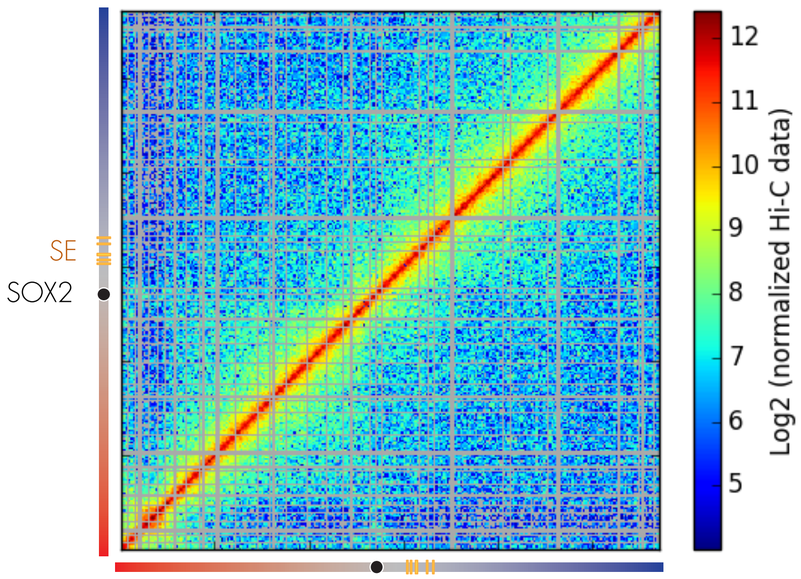


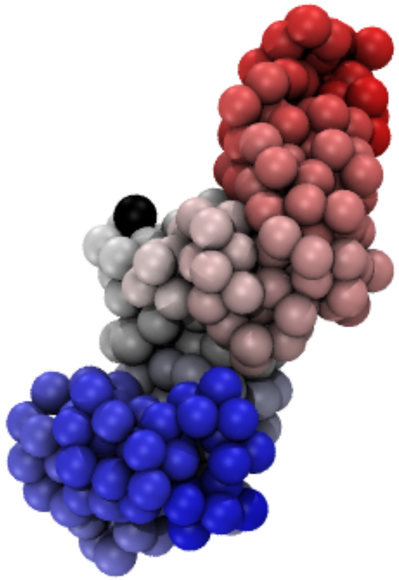
How does these structural rearrangements interplay with the transcription activity?

What are the main drivers of structural transitions?

TADbit modeling of SOX2 from B cells Hi-C





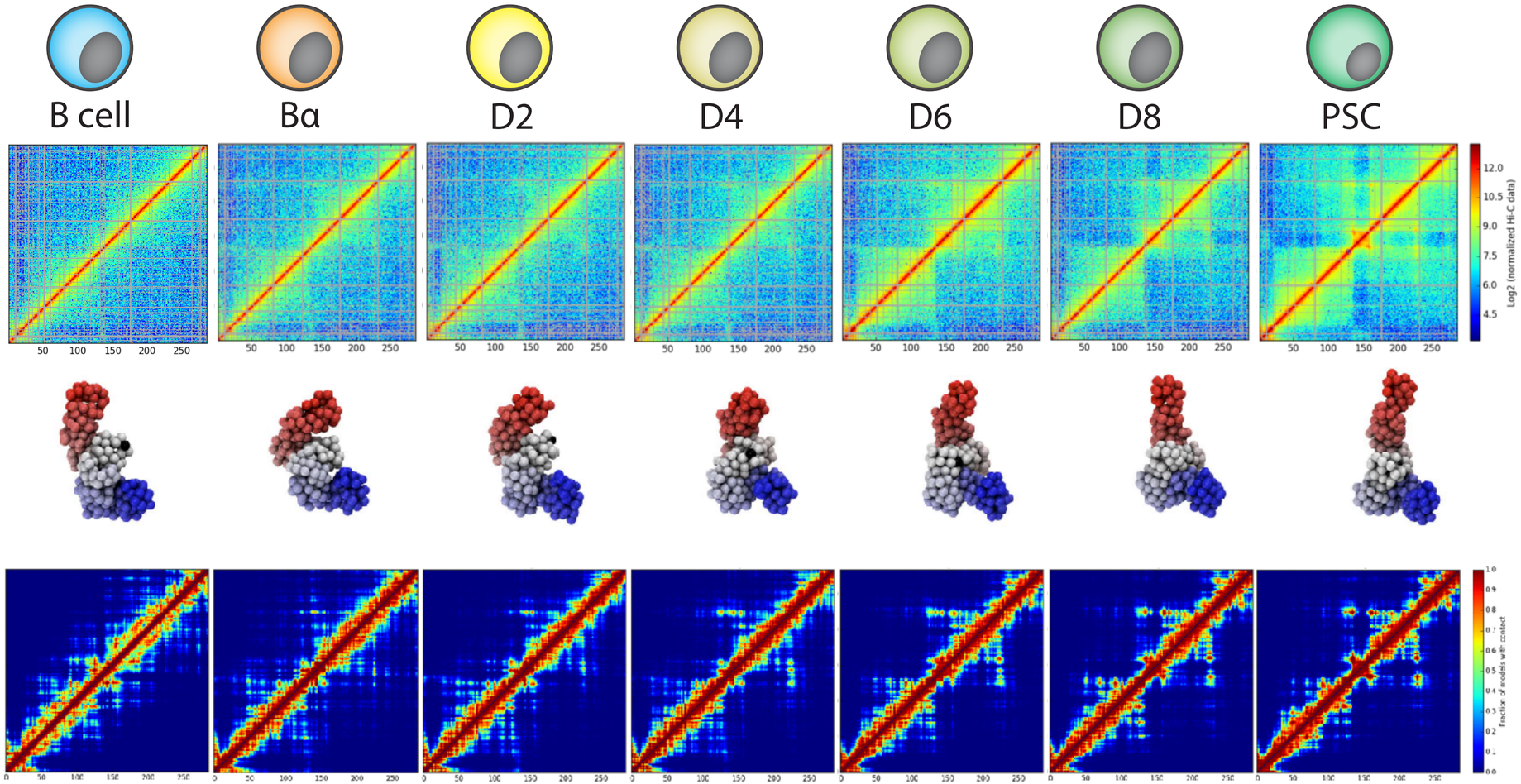


Optimal IMP parameters

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

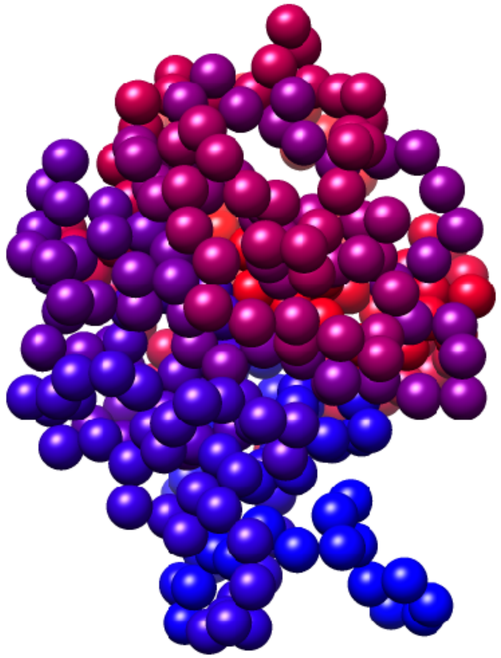
Models of reprogramming from B to PSC

The SOX2 locus



TADdyn. Dynamics of chromatin

Marco Di Stefano

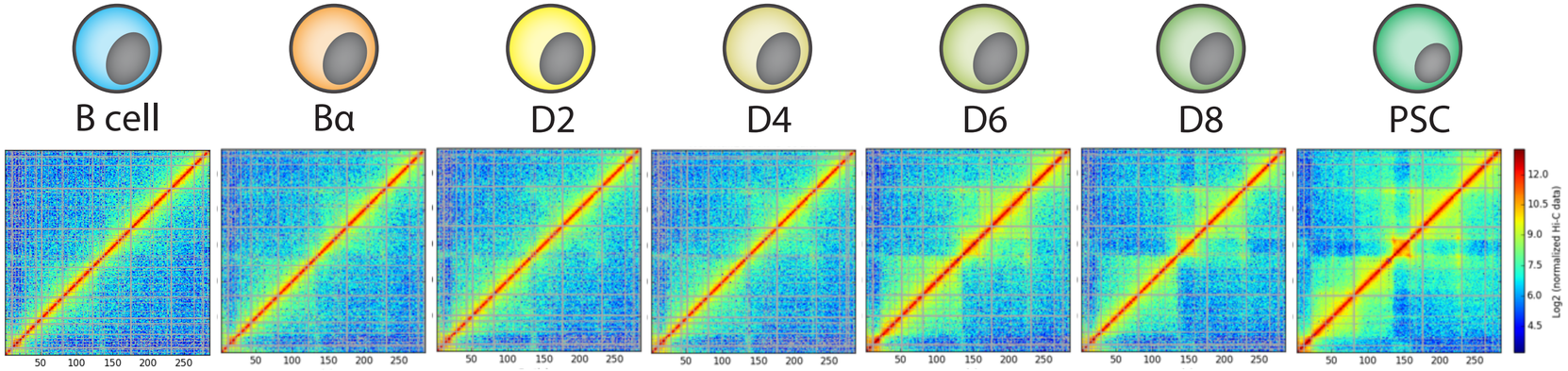


$$\mathcal{H}_{intra} = \sum_{i=1}^{N} U_{FENE}(i, i+1) + U_{br}(i, i+1, i+2) + \sum_{j=i+1}^{N} U_{LJ}(i, j)$$

Chain-connectivity interaction
Bending
Lennard-Jones Potential

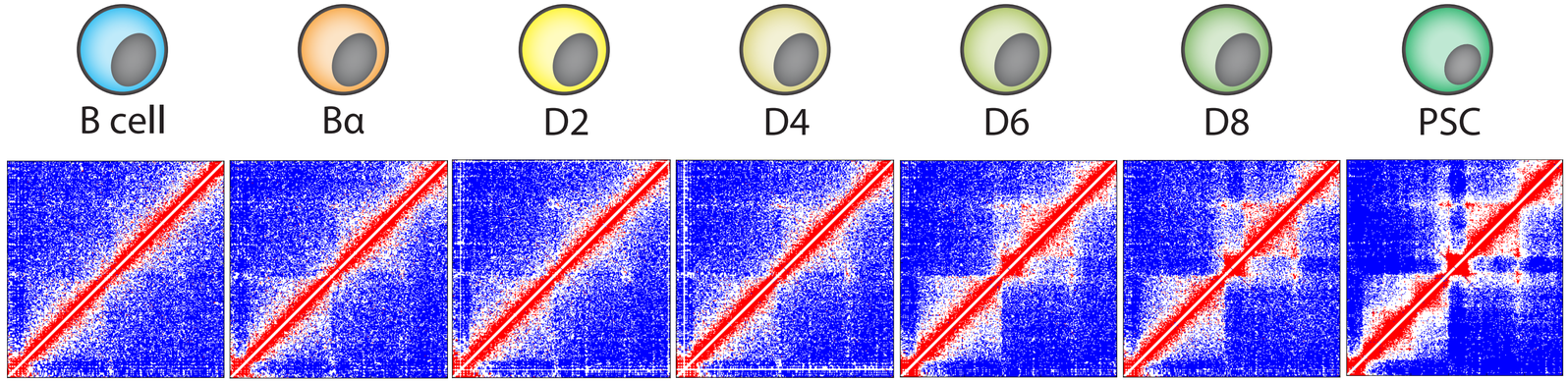
TADdyn: from time-series Hi-C maps to dynamic restraints

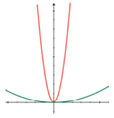
The SOX2 locus



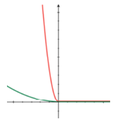
TADdyn: from time-series Hi-C maps to dynamic restraints

The SOX2 locus





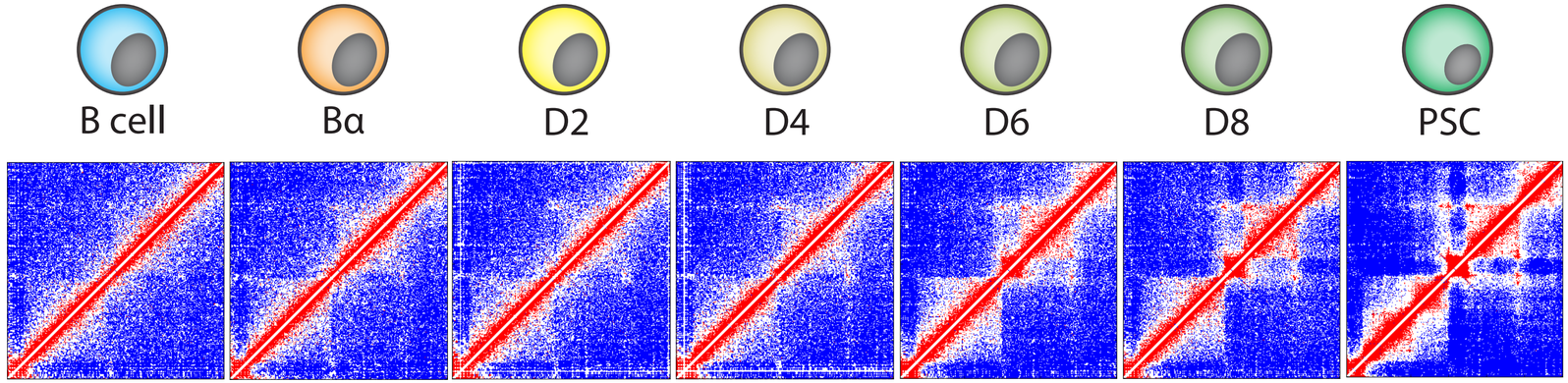
Harmonic



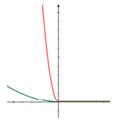
HarmonicLowerBound

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

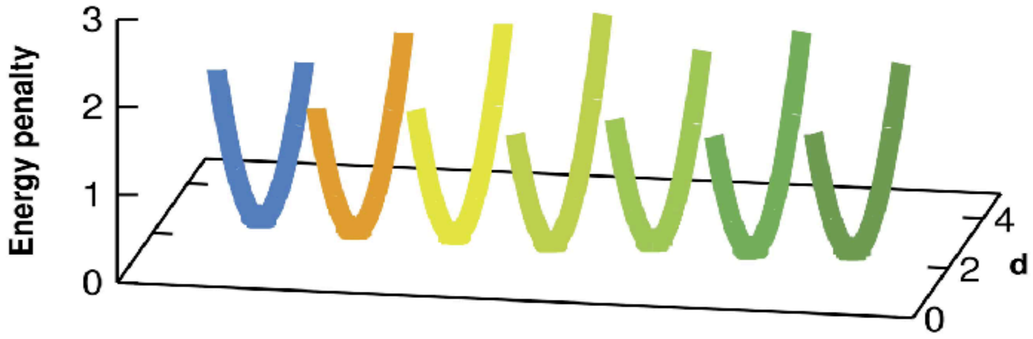
The SOX2 locus



Harmonic



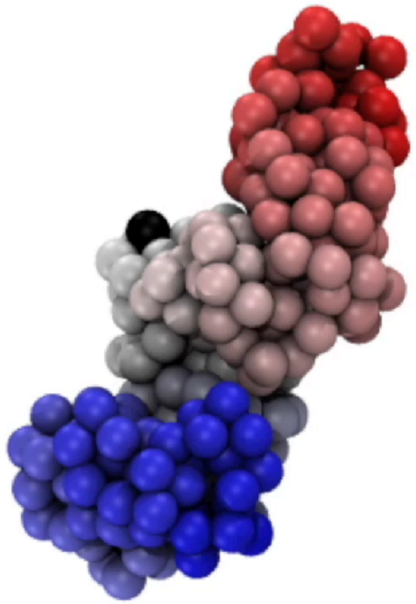
HarmonicLowerBound

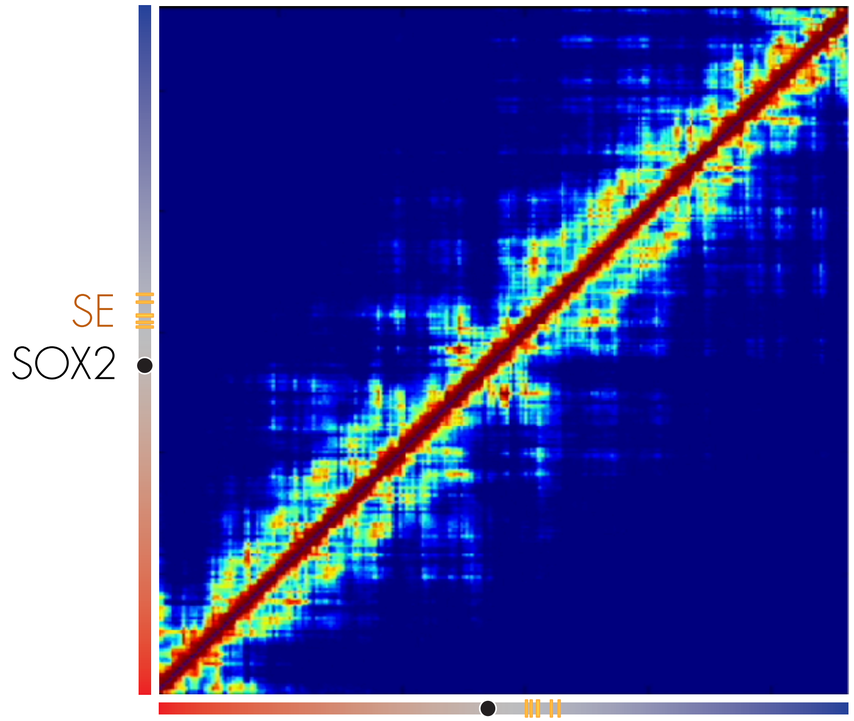


Transition	Stable	Vanishing	Raising
B -> Bα	18,612	6,984	7,290
Bα -> D2	18,512	7,390	6,687
D2 -> D4	18,369	6,830	6,893
D4 -> D6	18,971	6,291	7,289
D6 -> D8	20,167	6,093	6,250
D8 -> ES	20,679	5,738	6,173

SOX2 locus structural changes from B to PSC

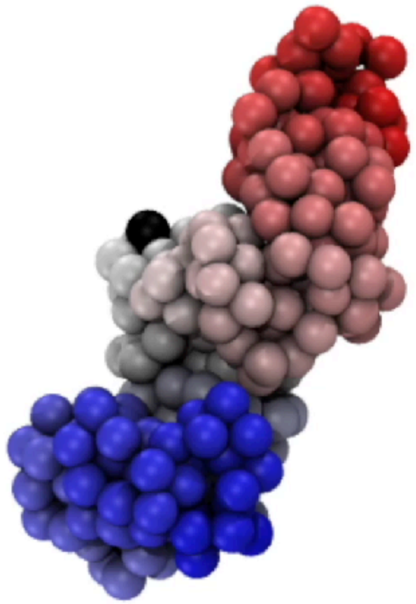
Contacts

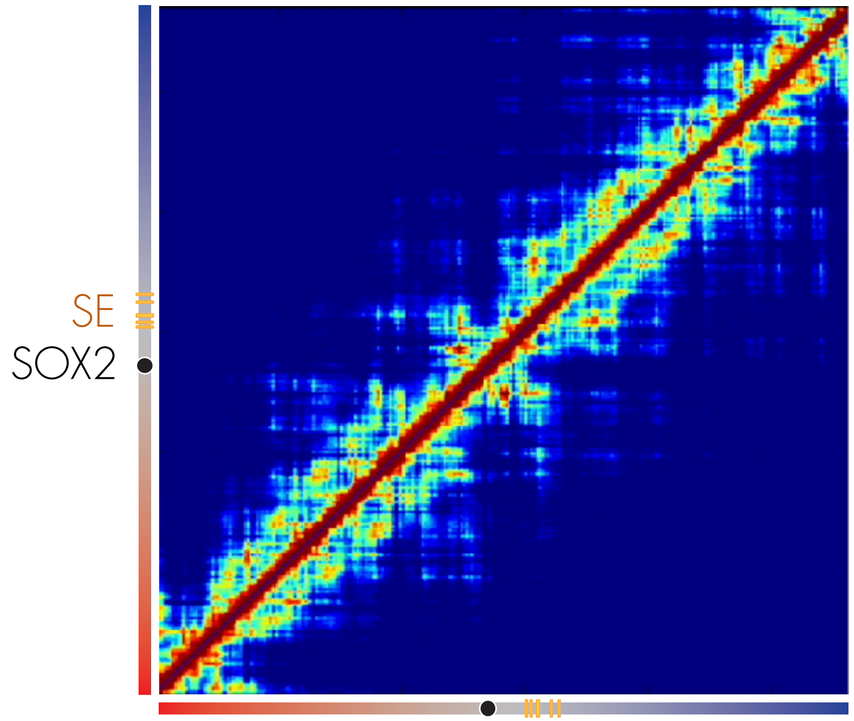




SOX2 locus structural changes from B to PSC

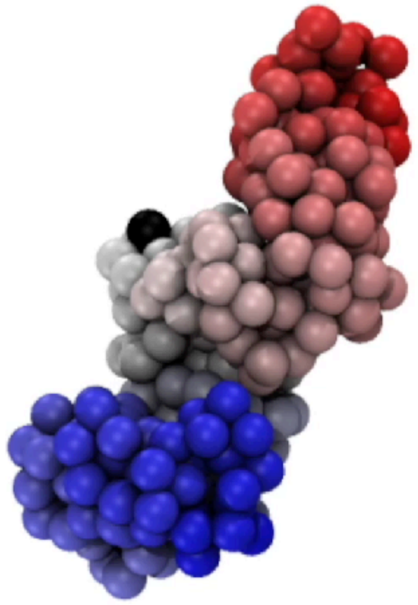
Contacts

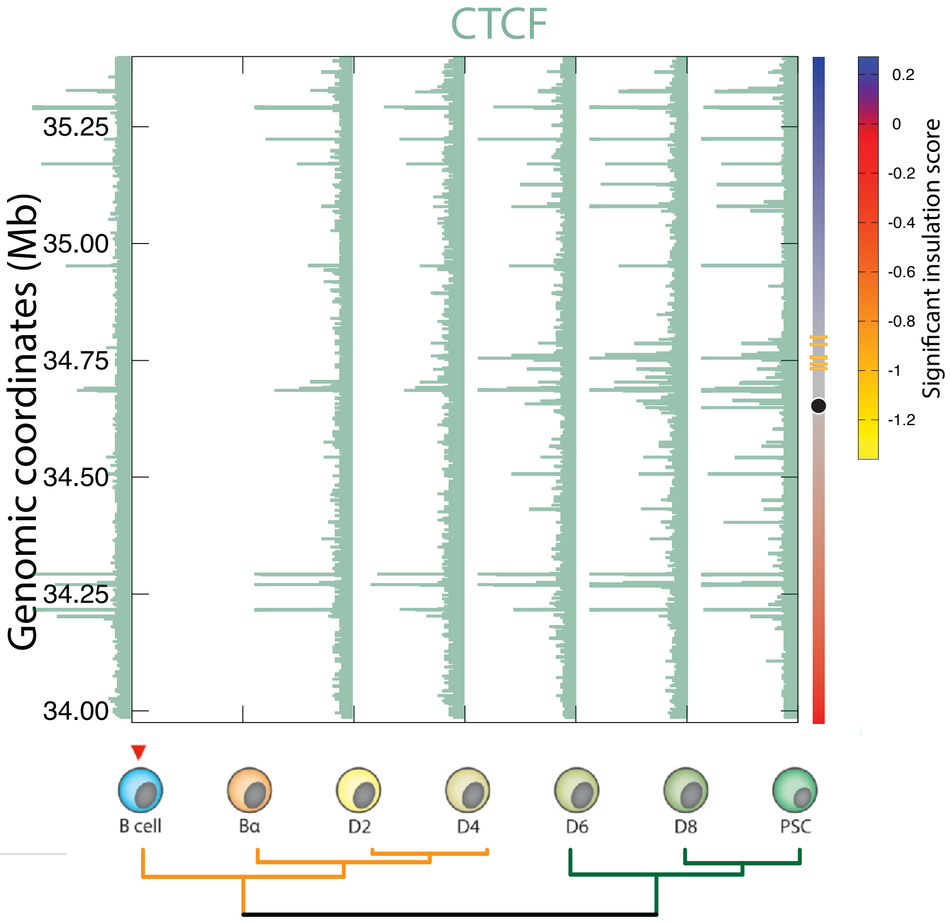




SOX2 locus structural changes from B to PSC

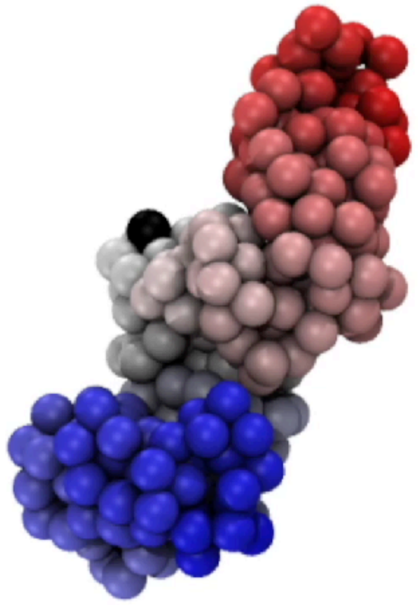
TAD borders

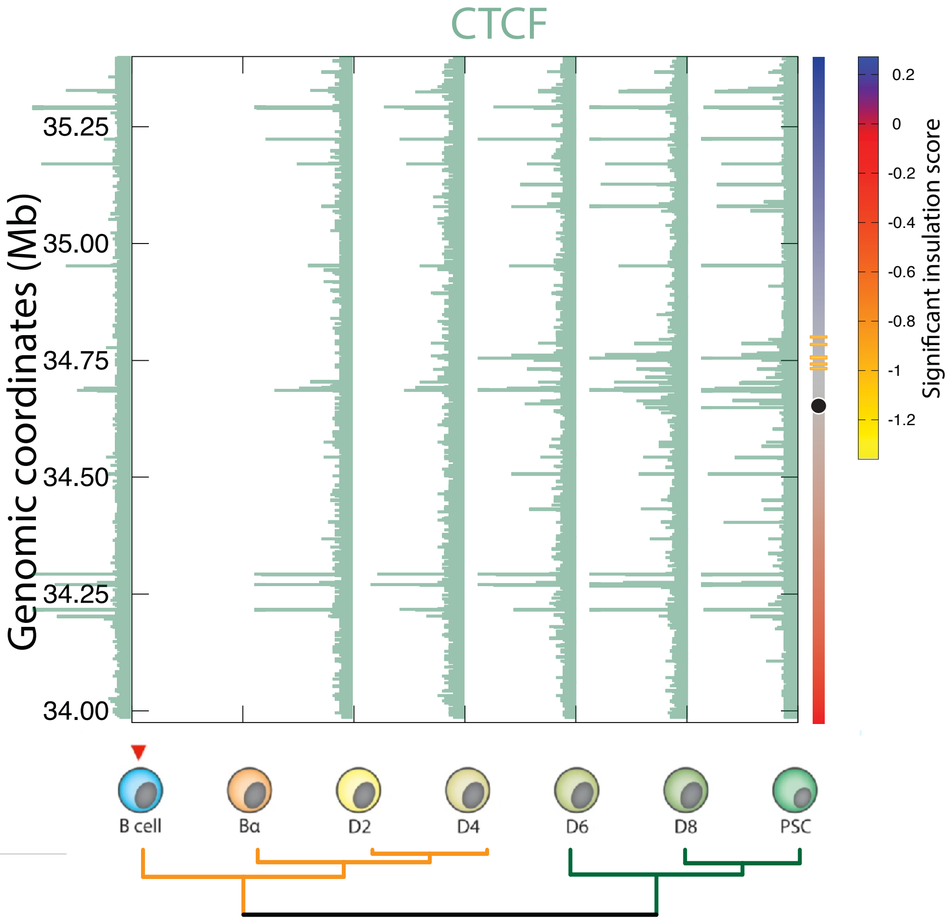




SOX2 locus structural changes from B to PSC

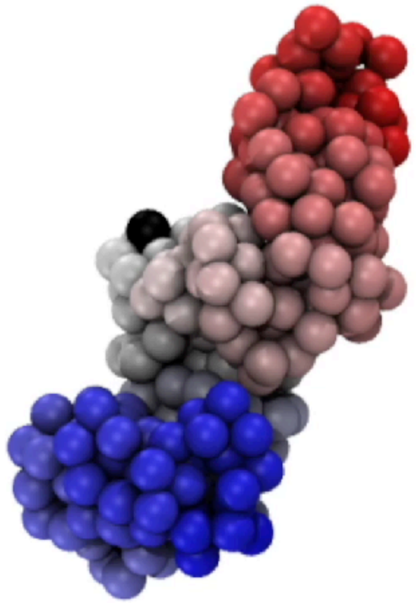
TAD borders

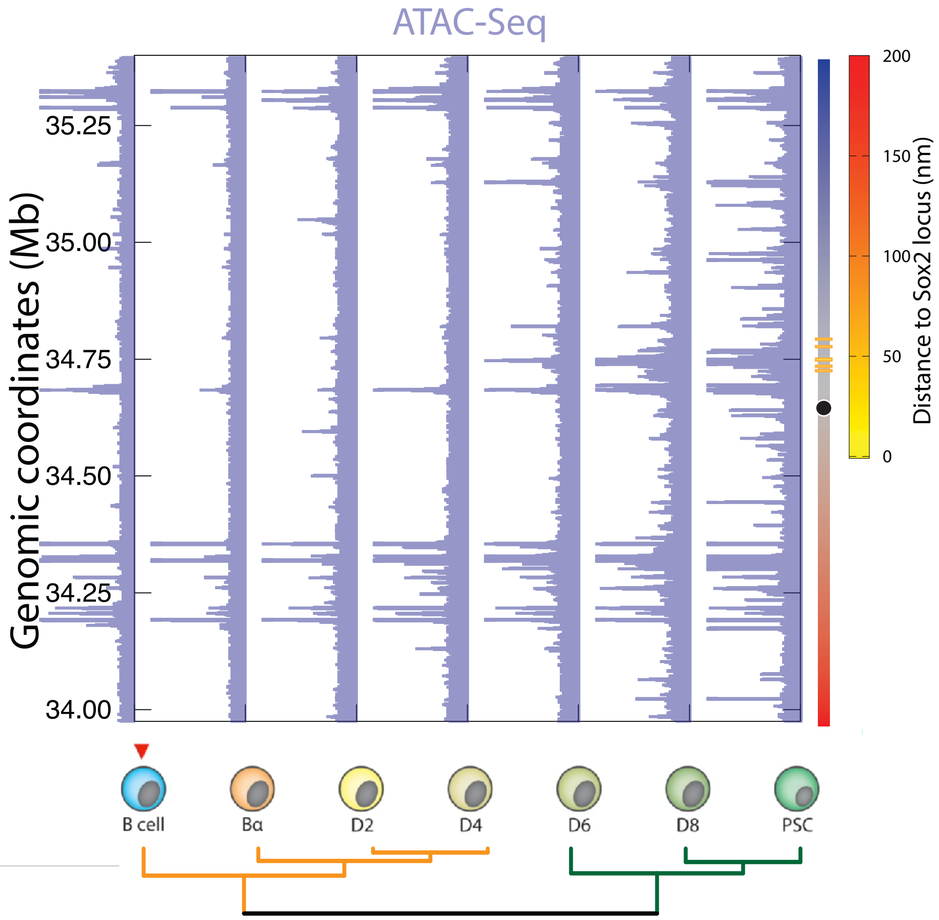




SOX2 locus structural changes from B to PSC

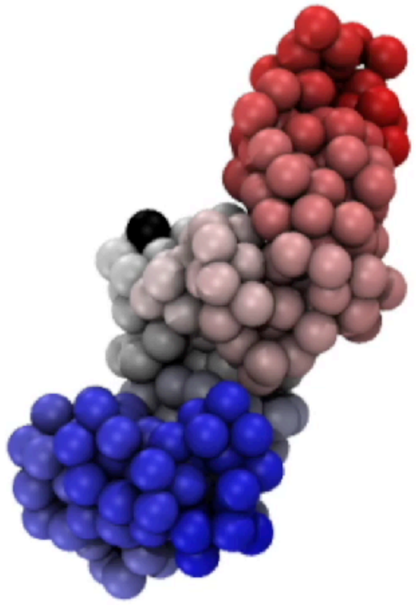
Distance to regulatory elements

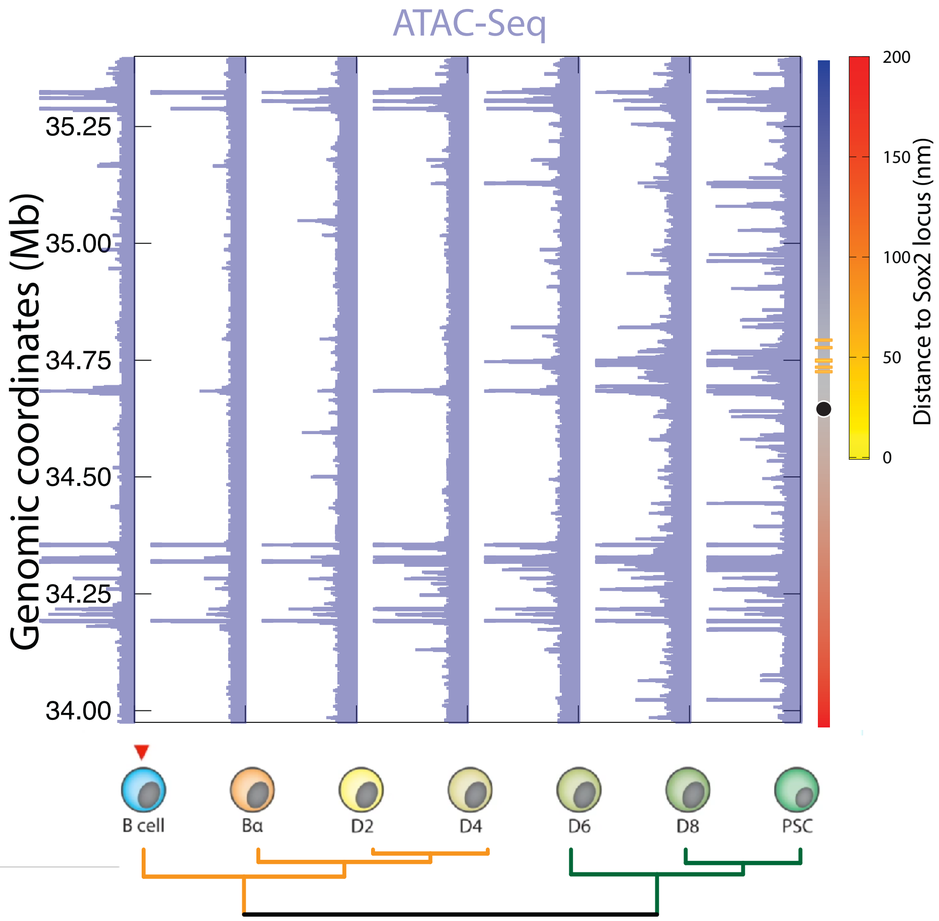




SOX2 locus structural changes from B to PSC

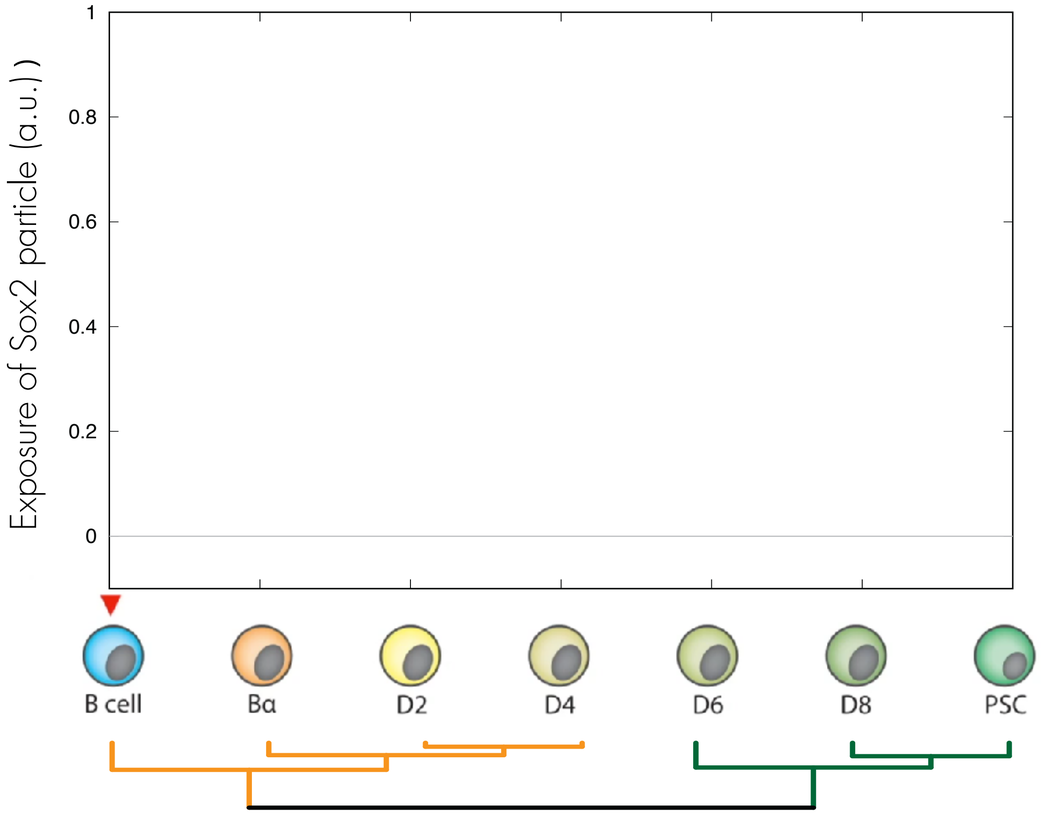
Distance to regulatory elements





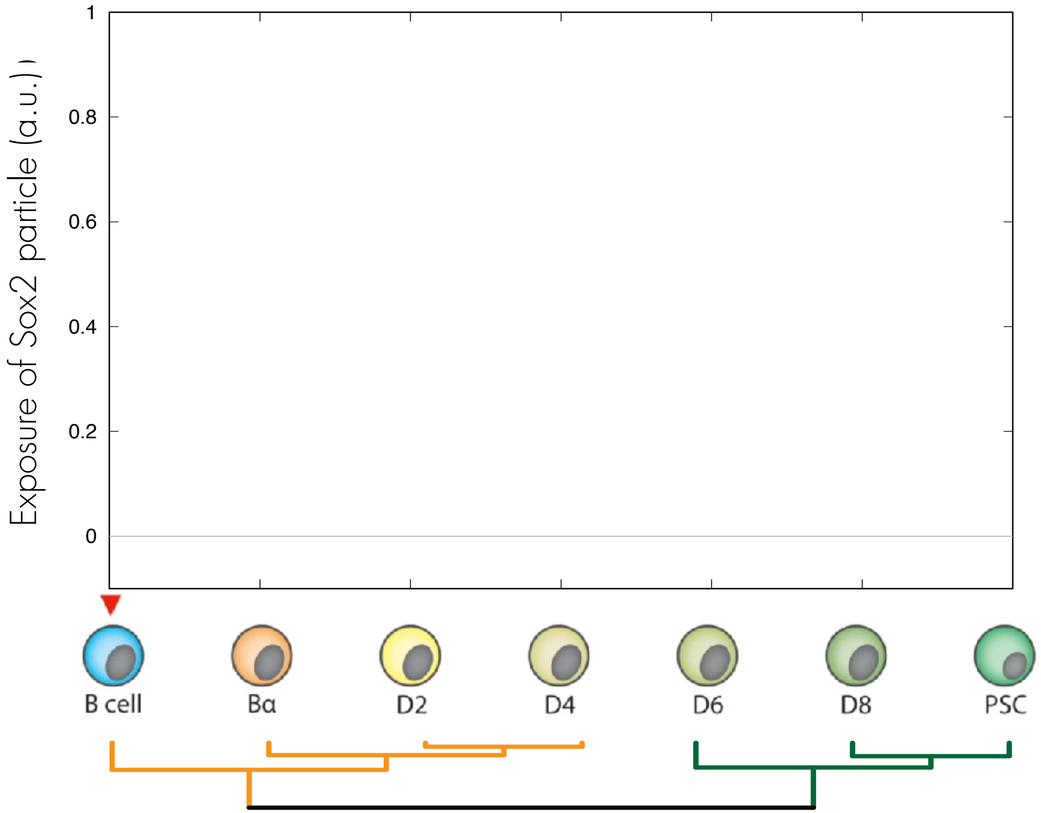
SOX2 locus structural changes from B to PSC

Structural exposure



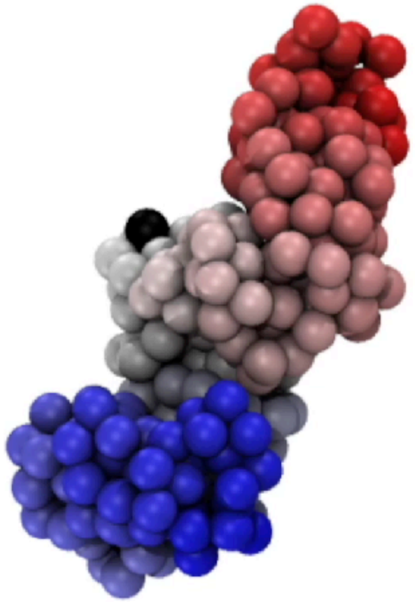
SOX2 locus structural changes from B to PSC

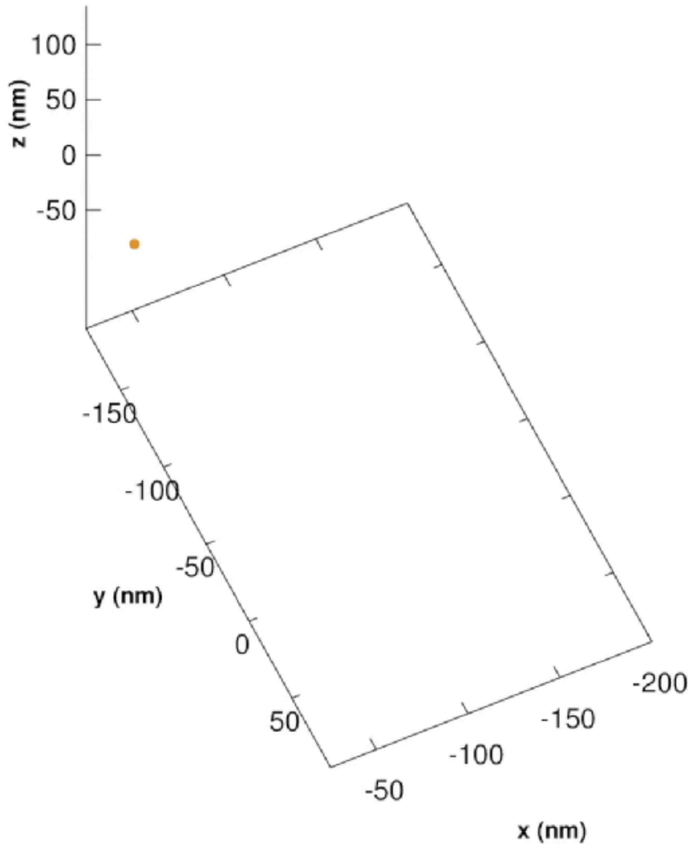
Structural exposure

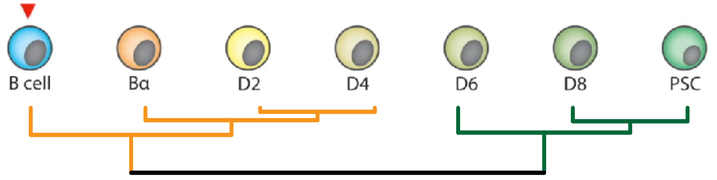


SOX2 locus dynamics changes from B to PSC

SOX2 displacement

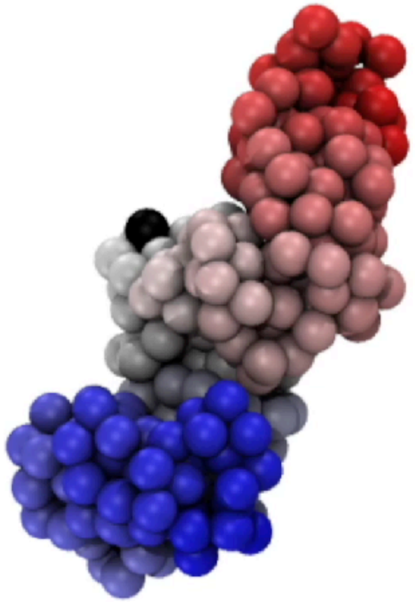


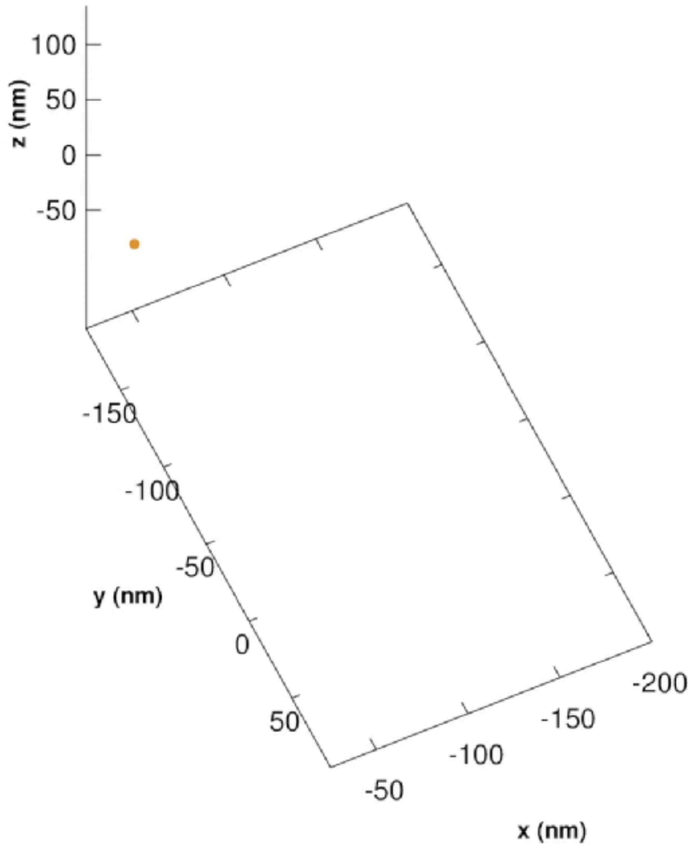


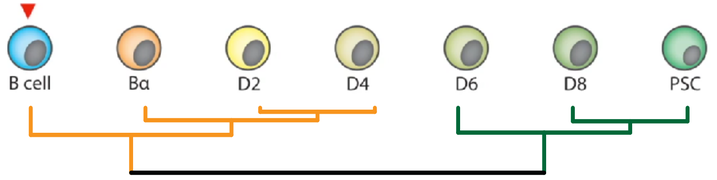


SOX2 locus dynamics changes from B to PSC

SOX2 displacement

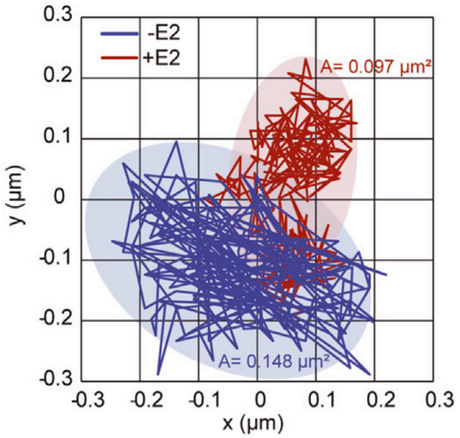






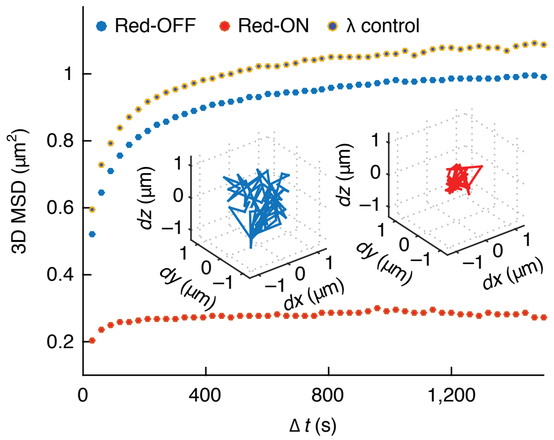
SOX2 locus dynamics changes from B to PSC

SOX2 displacement



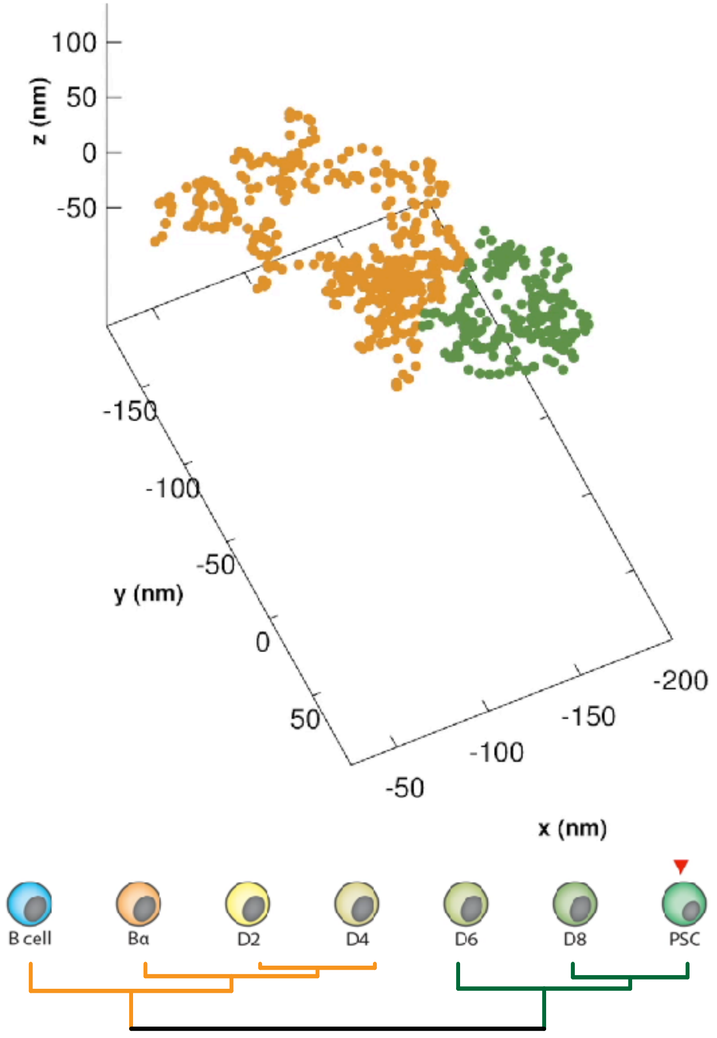
Two dimensional trajectories and area explored over 50s of the CCND1 locus recored before -E2 and after +E2 activation.

Germier ,T., et al, (2017) Biophys 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



A “cage” model for transcriptional activation

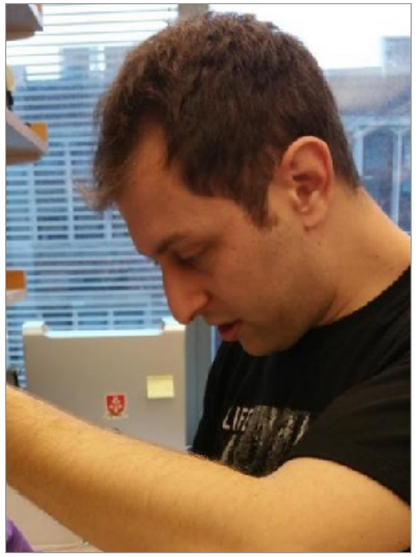
Chromosome walking with super-resolution imaging and modeling

Marc A. Marti-Renom
CNAG-CRG · ICREA

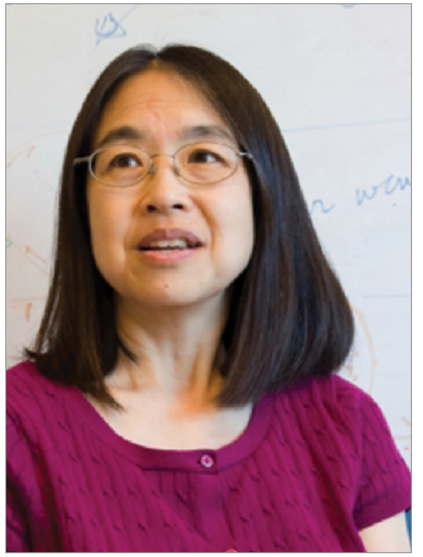
http://marciuslab.org
http://3DGenomes.org
http://cnag.crg.eu

cnag CRG ICREA

Irene Farabella
CNAG-CRG



Guy Nir
Harvard Med School



Ting Wu
Harvard Med School

Can we walk the chromatin path in the nucleus?

by

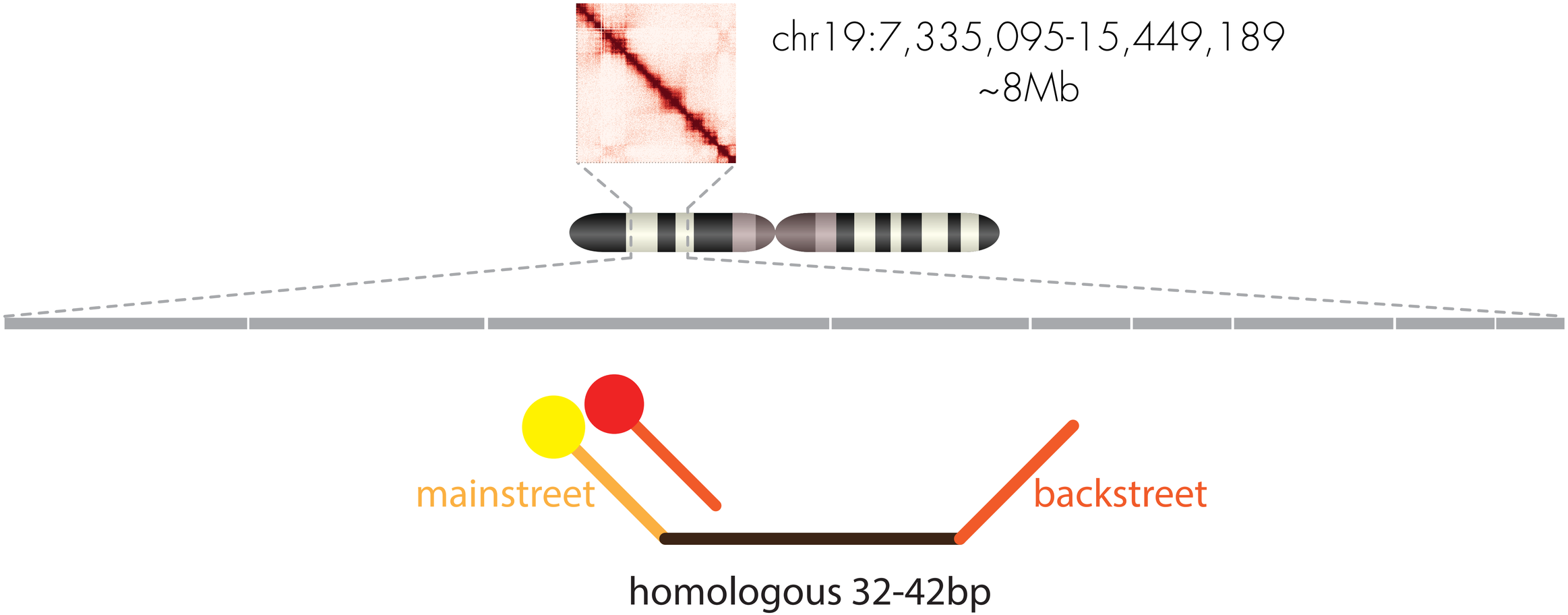
Integrating imaging and Hi-C maps with modeling.

by developing a method for

Oligopaint-based modeling of genomes

High-resolution imaging

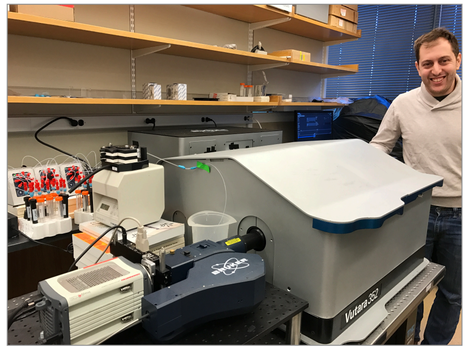
Tracing chromosomes with OligoSTROM & fluidics cycles in PGP1 cells



Beliveau et al. Nat. Comm. 2015

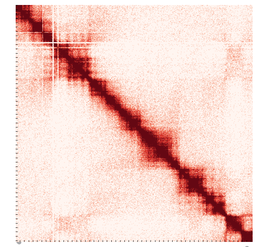
High-resolution imaging

Tracing chromosomes with OligoSTROM & fluidics cycles in PGP1 cells

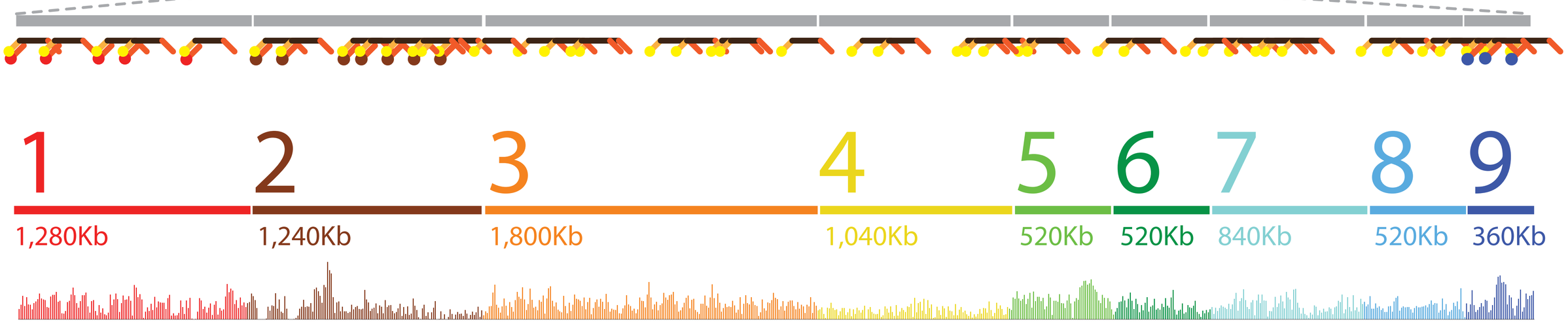


Guy Nir Harvard Med School

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

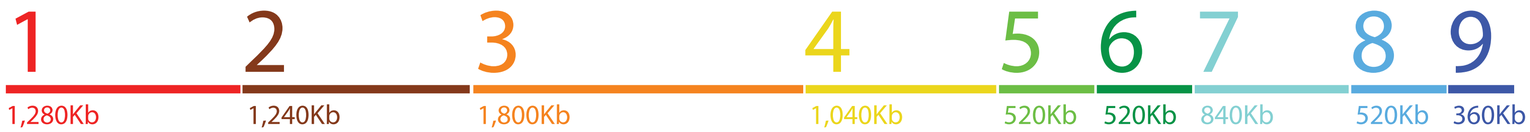


chr19:7,335,095-15,449,189
~8Mb



High-resolution imaging

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

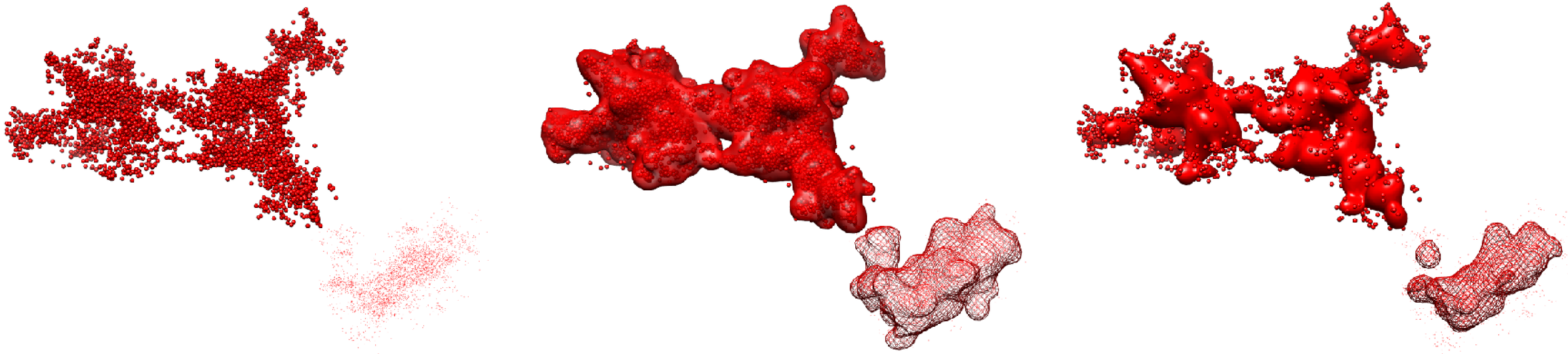


Cell-02

High-resolution imaging

XYZ points convolution into a density map

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

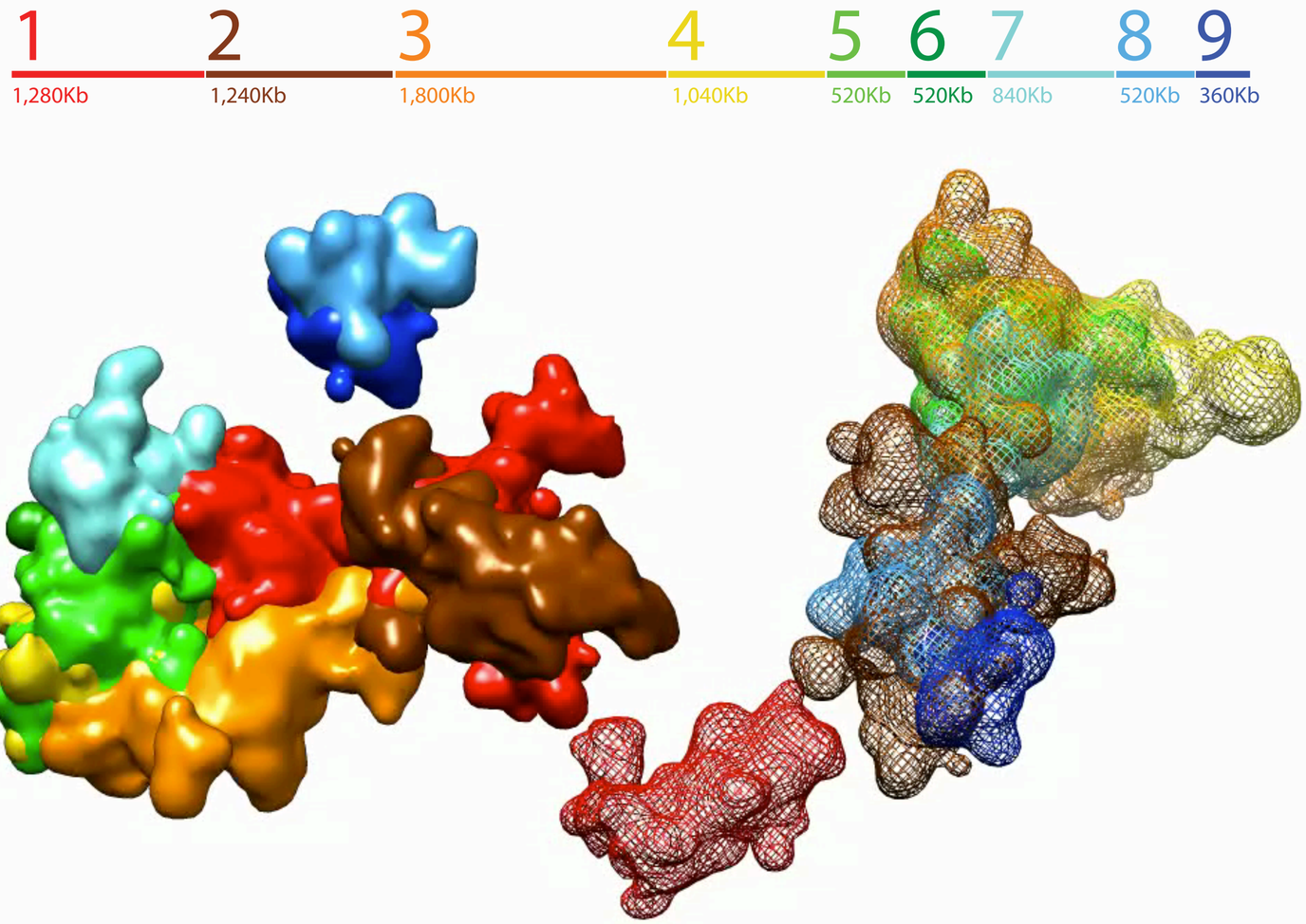


Cell-02 · Segment 1

Farabella et al, J Appl Crystallogr. 2015

Density maps

Cell-02 · Density map @ 50nm

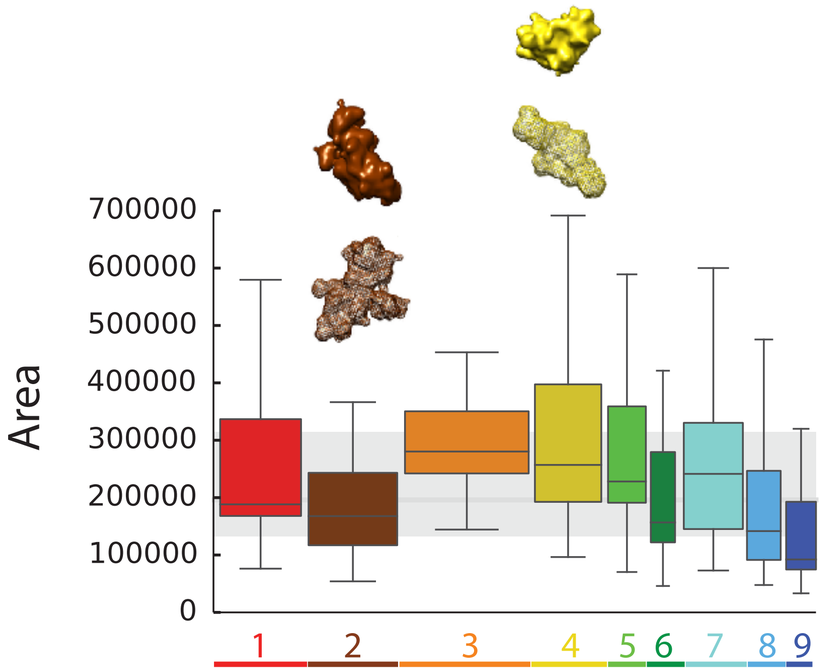


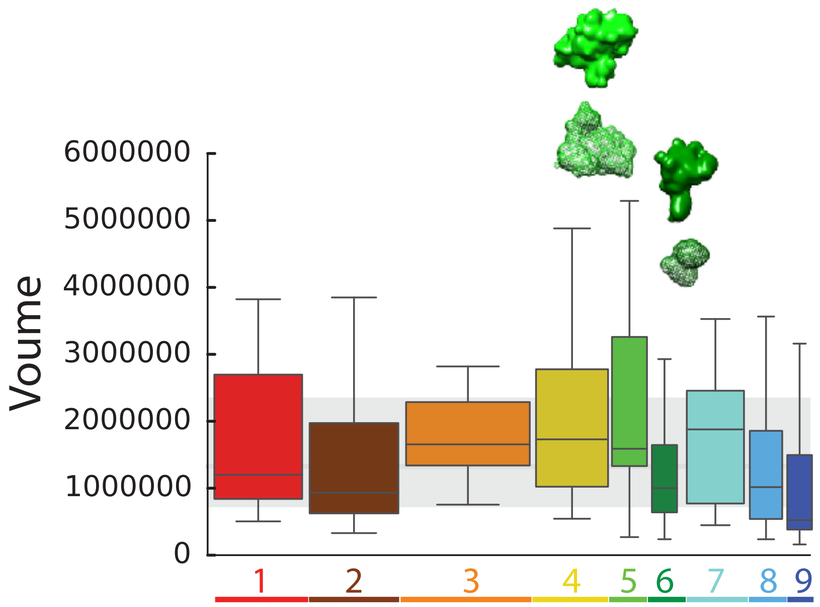
Area (nm^2)
Volume (nm^3)
Sphericity
Overlap (%)
Distance (nm)

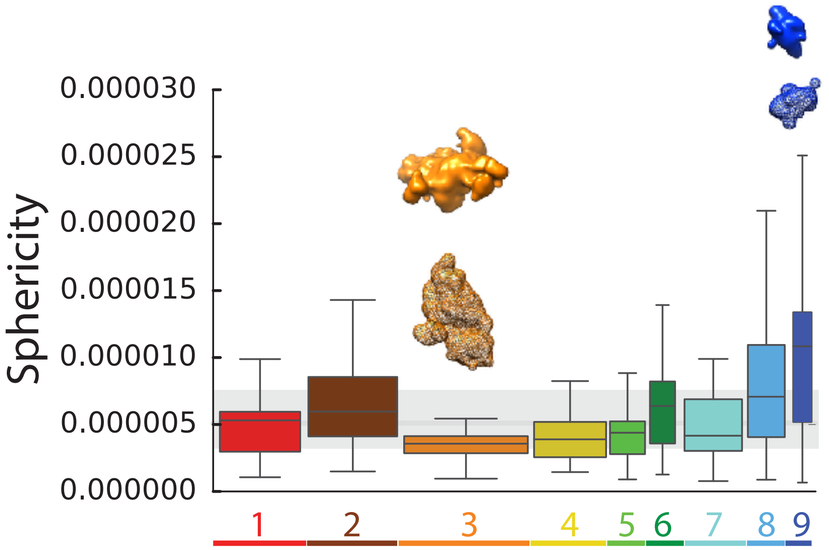
Farabella et al, J Appl Crystallogr. 2015

Structural features

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



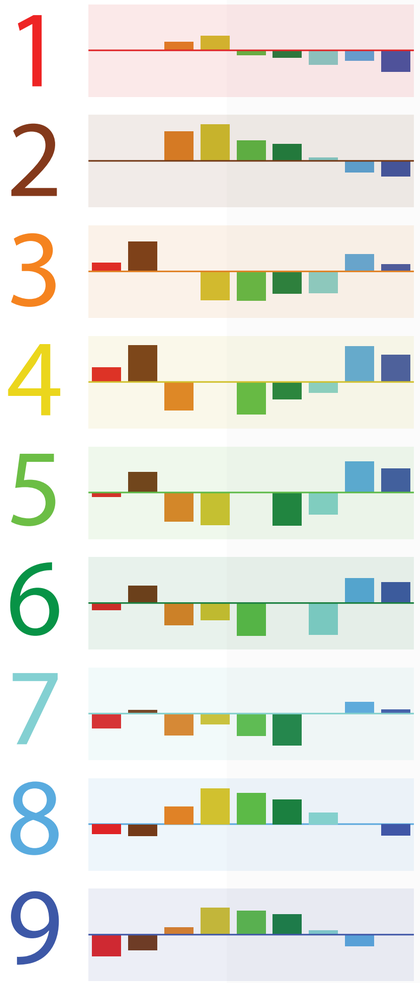


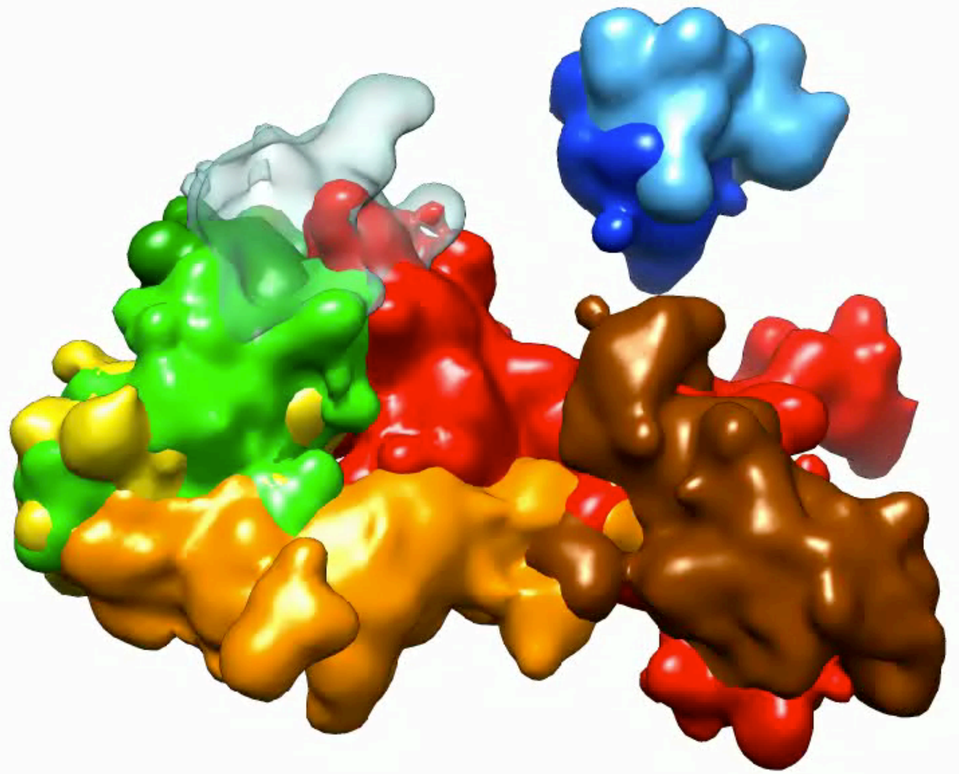


Spatial arrangement

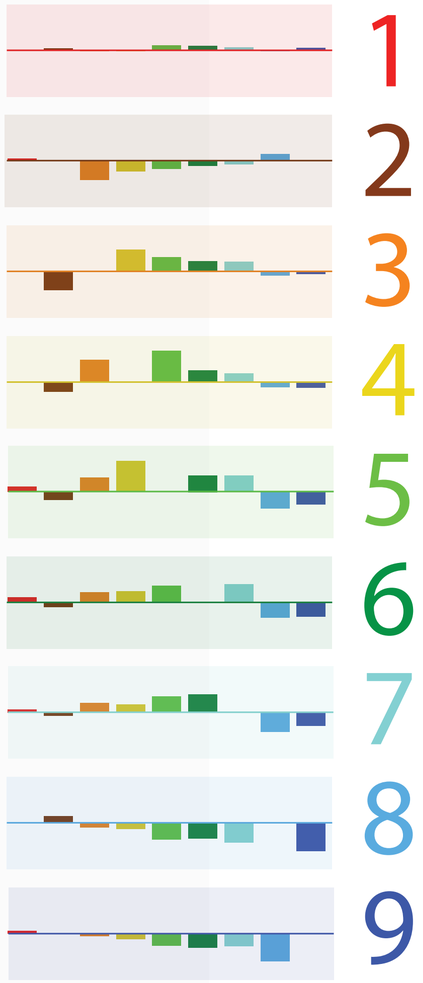
Distance and overlap of 19 cells each with 2 homologous resolved

Diff. distance



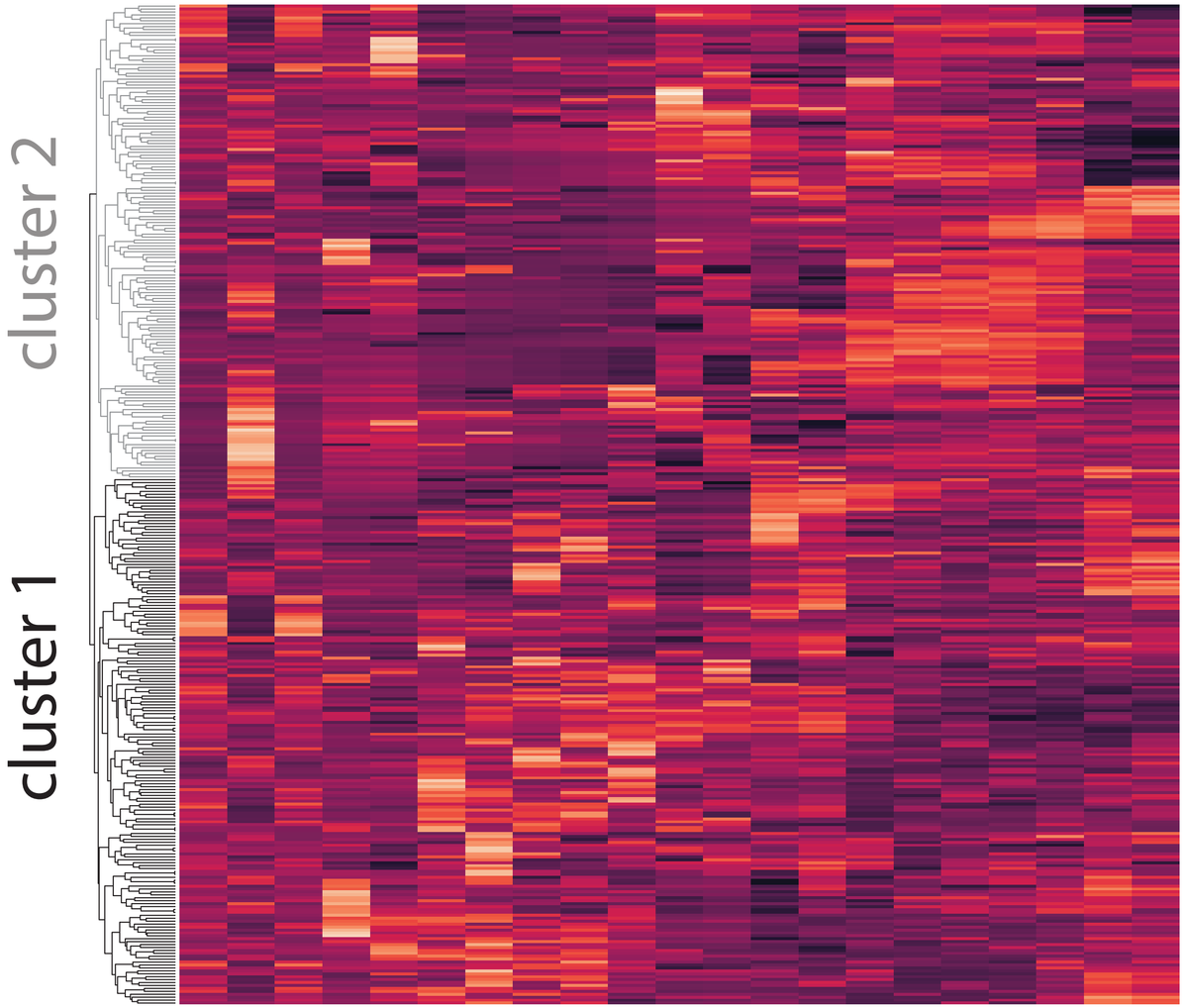


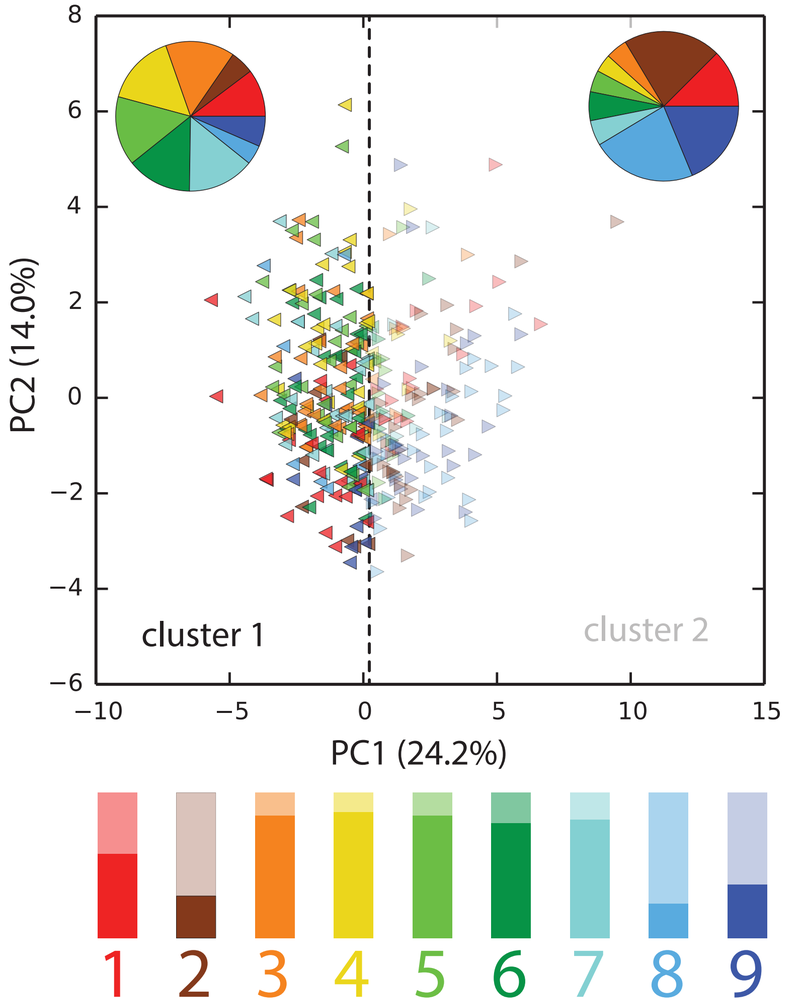
Diff. overlap



Structural clustering

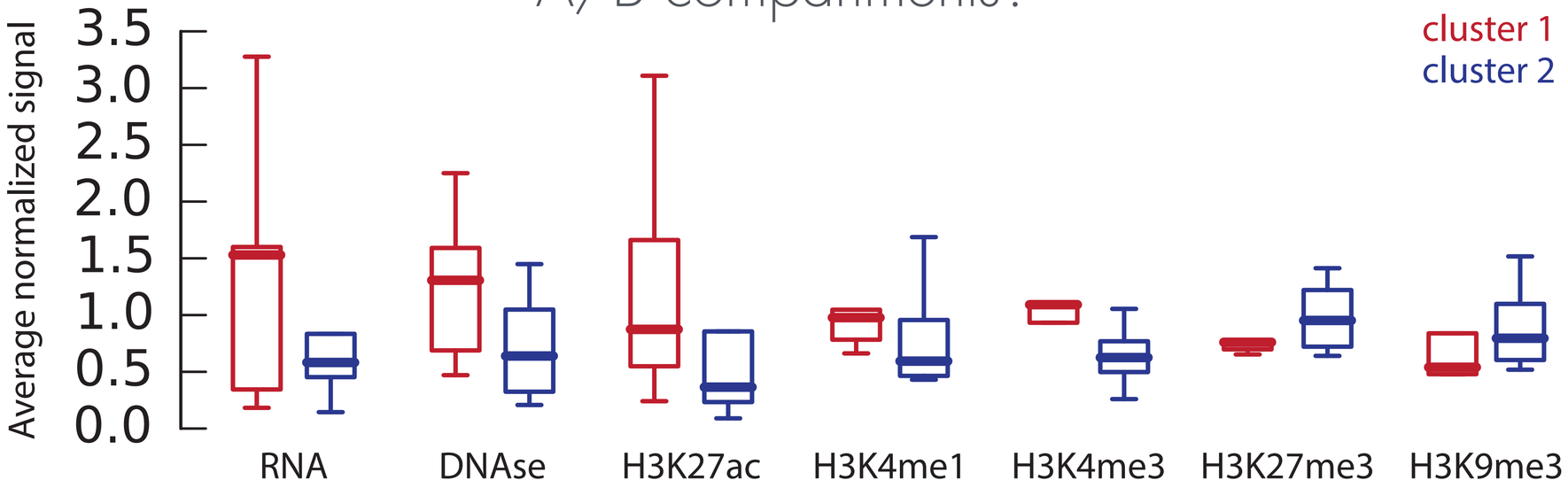
19 cells each with 2 homologous and 9 segments each (342)

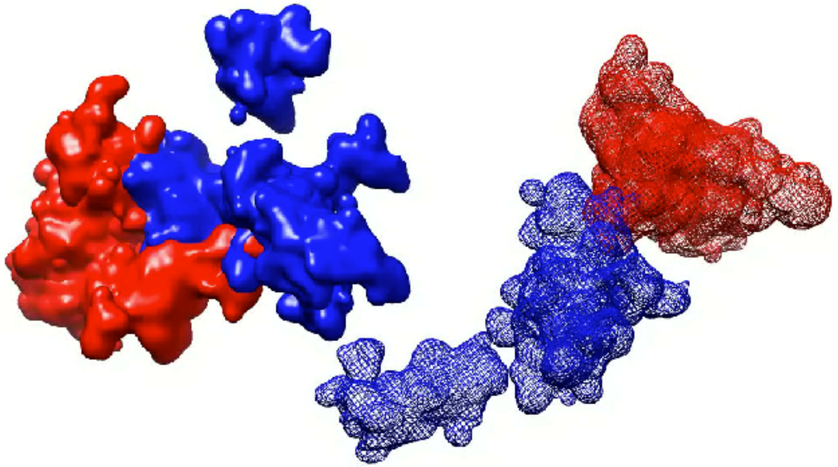


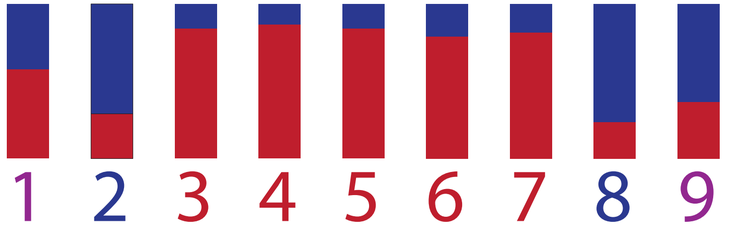


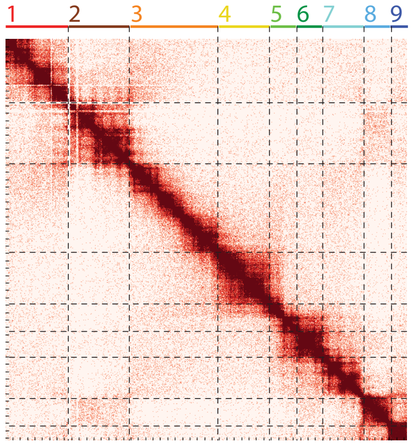
Cluster properties

A/B compartments?





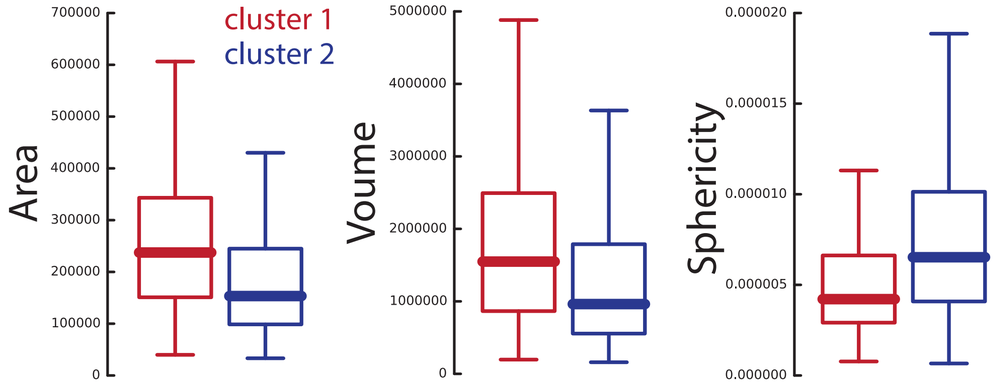


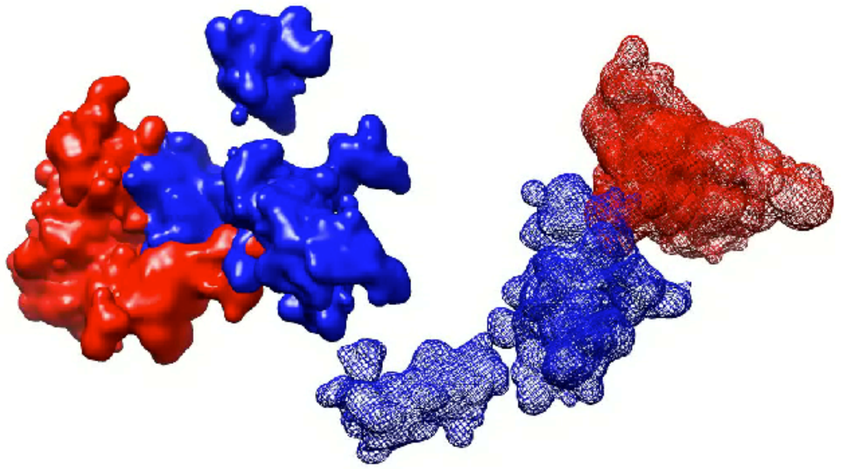


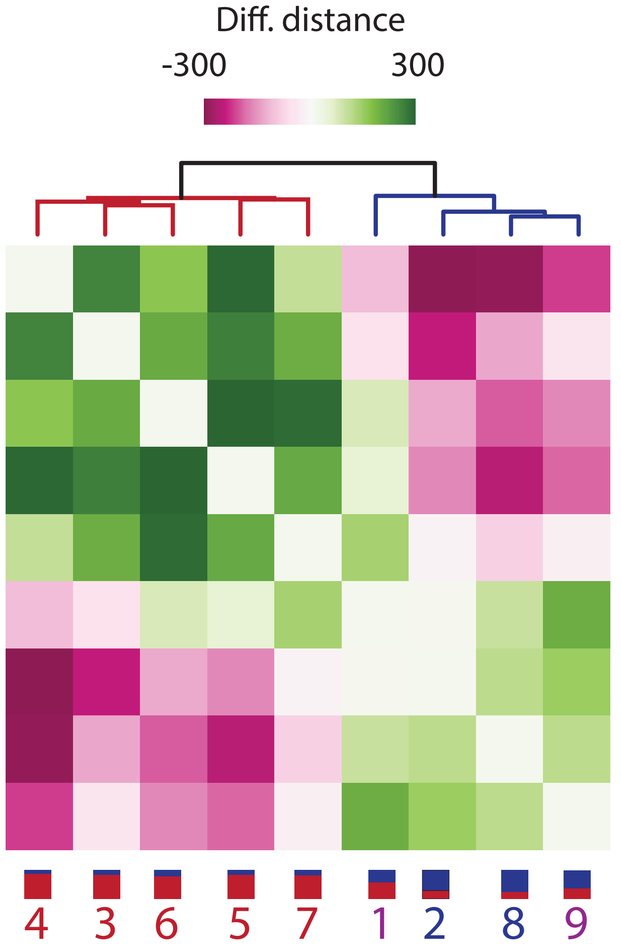
PGP1 ChIP-seq and Hi-C data from ENCODE and Lieberman-Aiden Lab, respectively

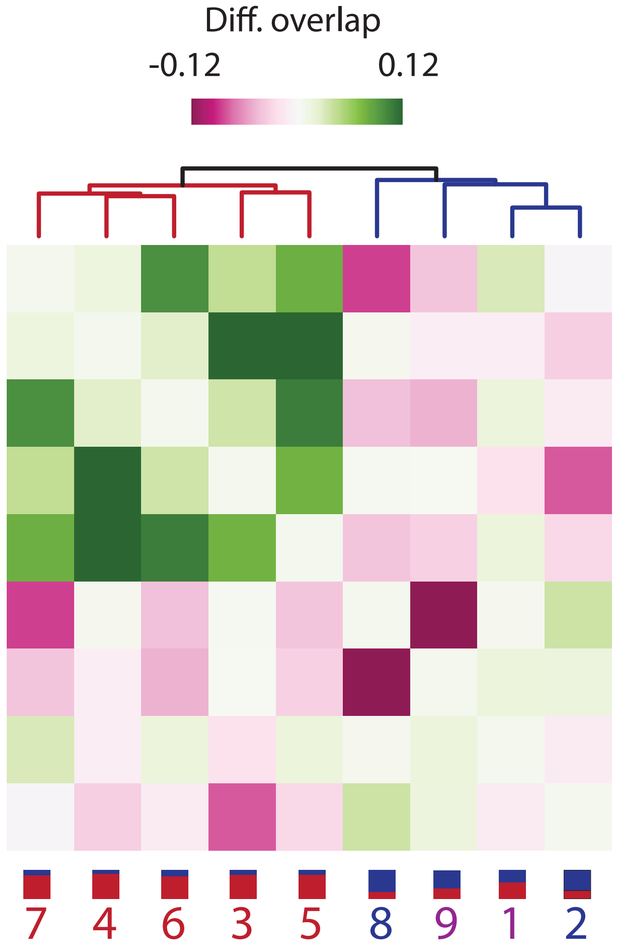
Cluster properties

A/B compartment properties









Can we walk the chromatin path in the nucleus?

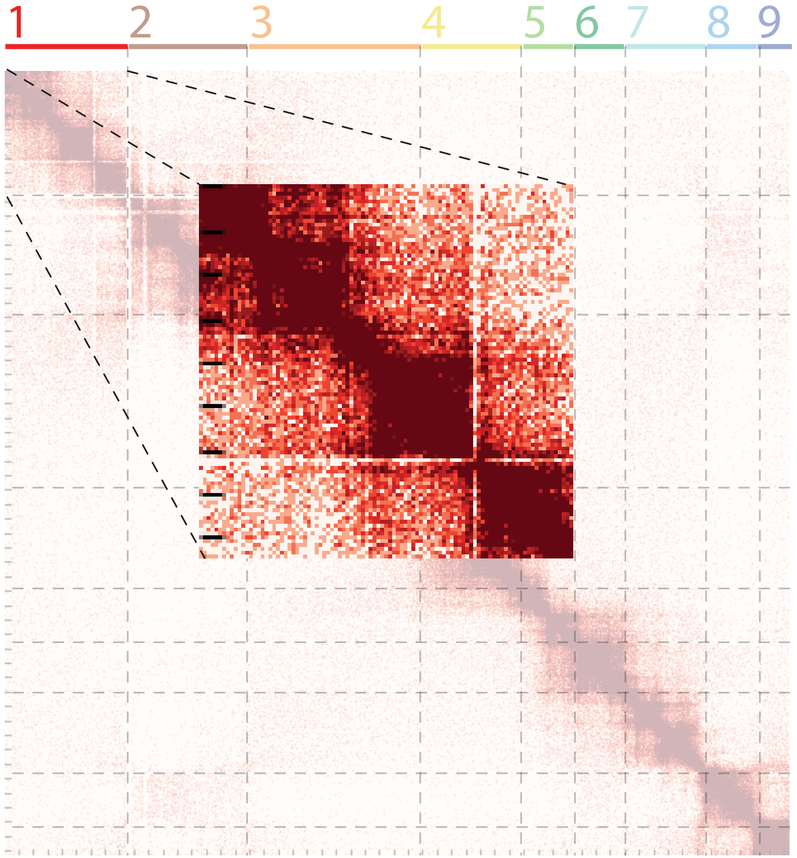
YES!

Can we increase the resolution of our data?

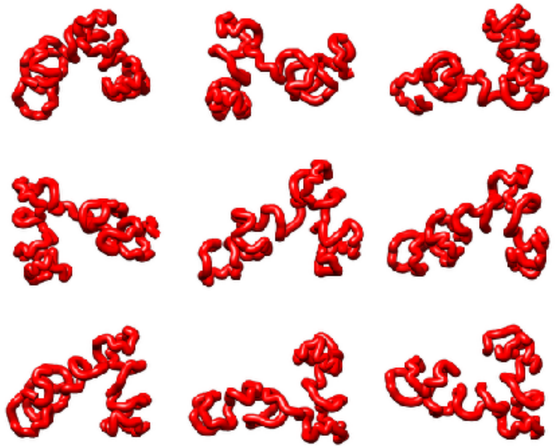
by fitting 3D models based on Hi-C interaction maps

Increasing resolution

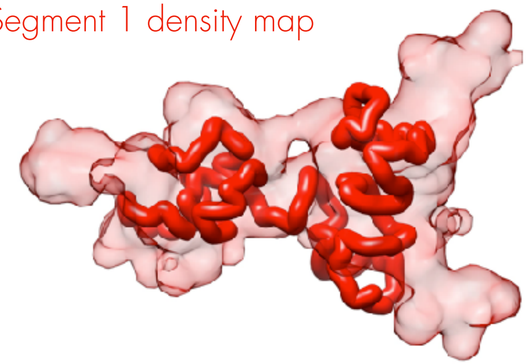
Rigid body fitting 3D structures based on Hi-C data



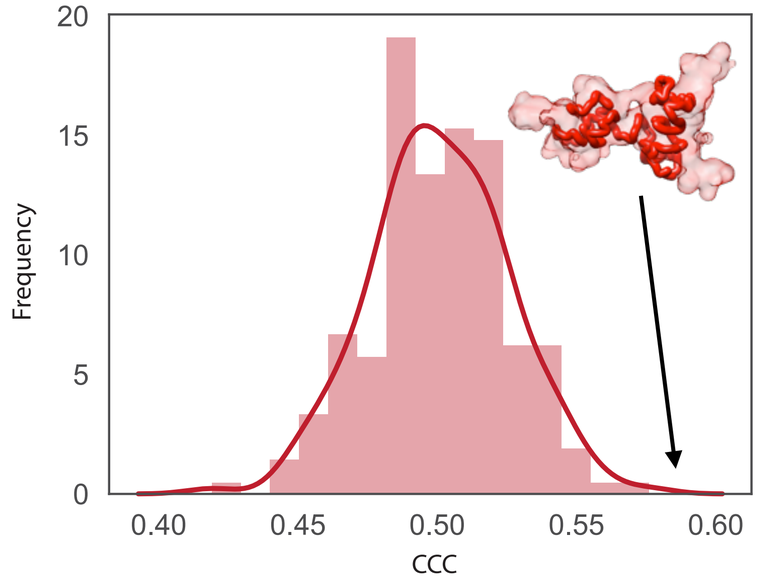
Segment 1 3D models



Segment 1 density map



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



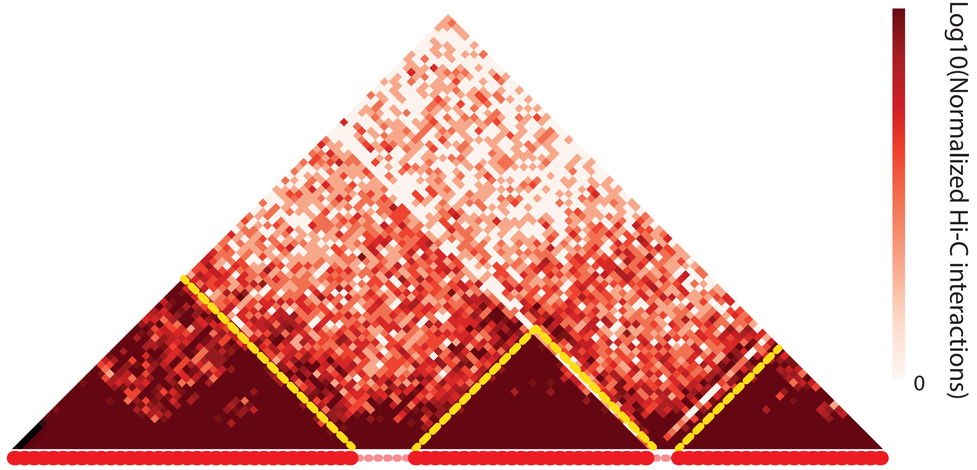
TADbit

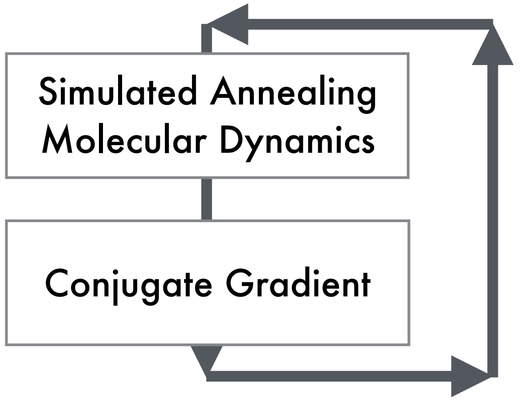
Serra, Baù, et al. PLOS CB 2017
http://www.3DGenomes.org

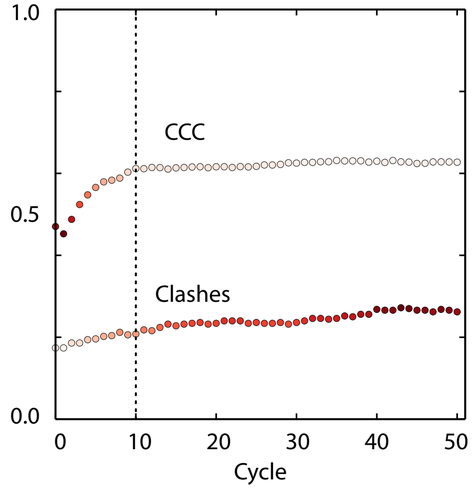
Farabella et al, J Appl Crystallogr. 2015
Roseman, 2000; Wriggers & Chacon, Structure 2001

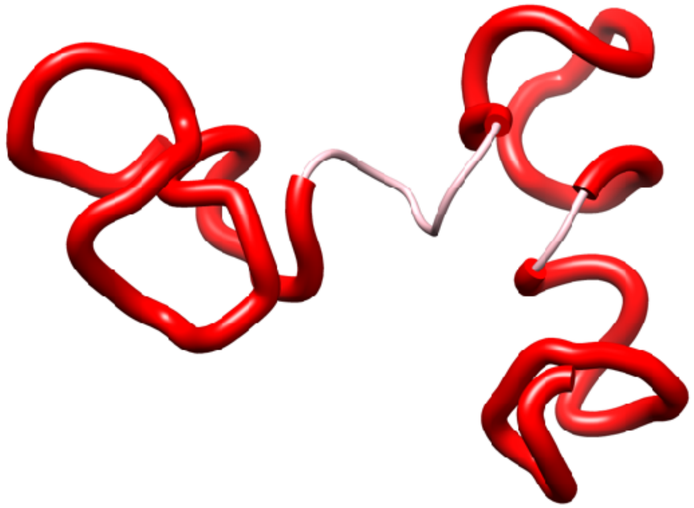
Increasing resolution

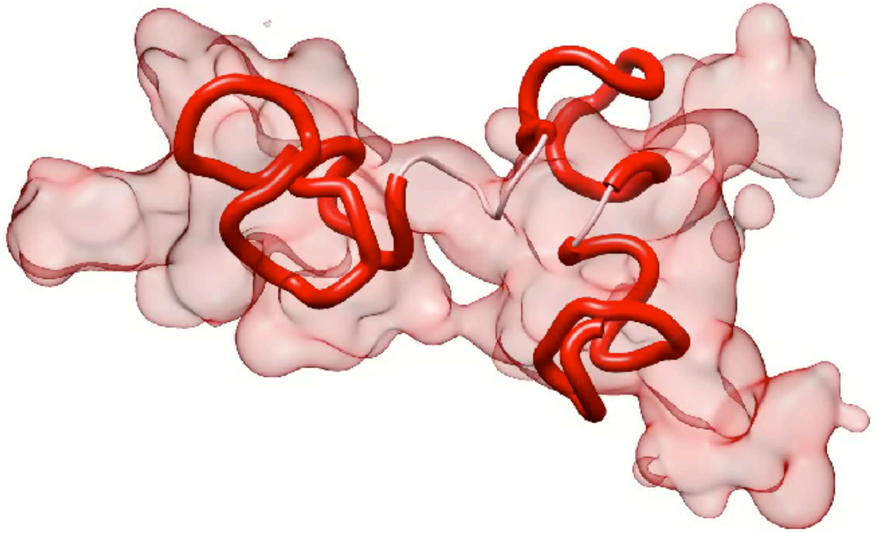
Flexible fitting 3D structures based on Hi-C data





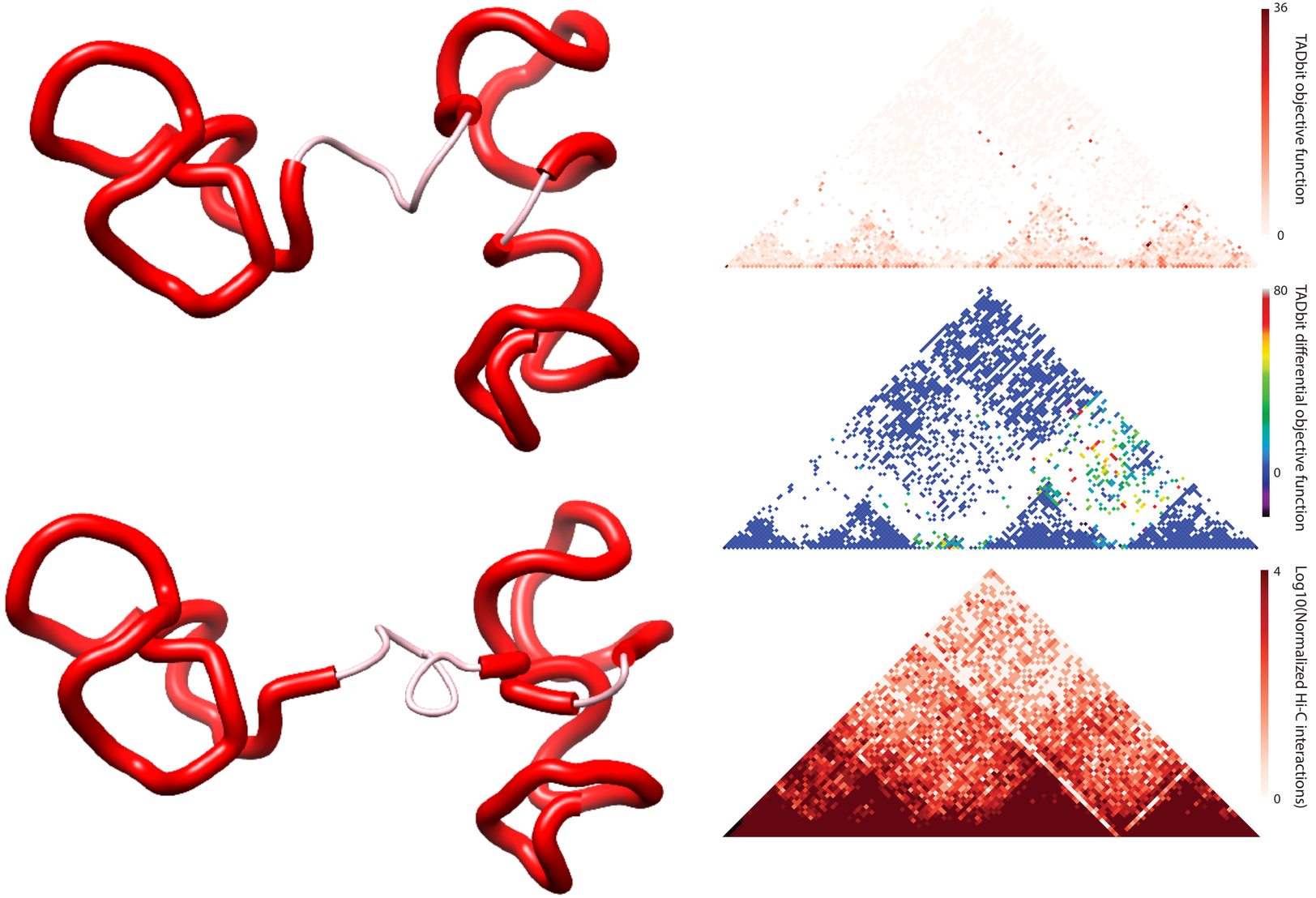




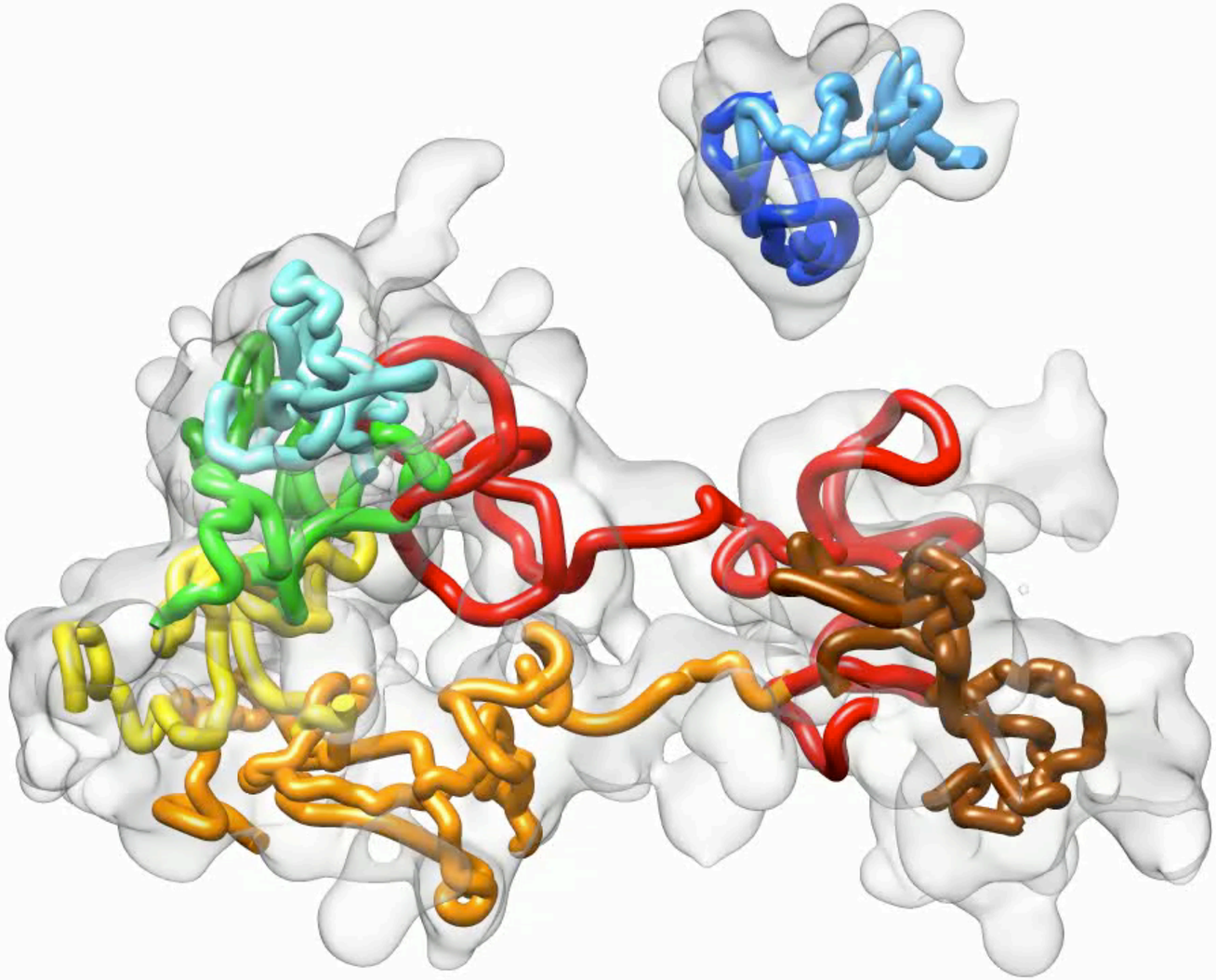


Increasing resolution

Flexible fitting 3D structures based on Hi-C data



Chromosome walking path @10Kb resolution



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

David Castillo
Yasmina Cuartero
Marco Di Stefano
Irene Farabella
Silvia Galan
Mike Goodstadt
Francesca Mugianesi
Julen Mendieta
Juan Rodriguez
François Serra
Paula Soler
Aleksandra Sparavier

.: Our current sponsors :.