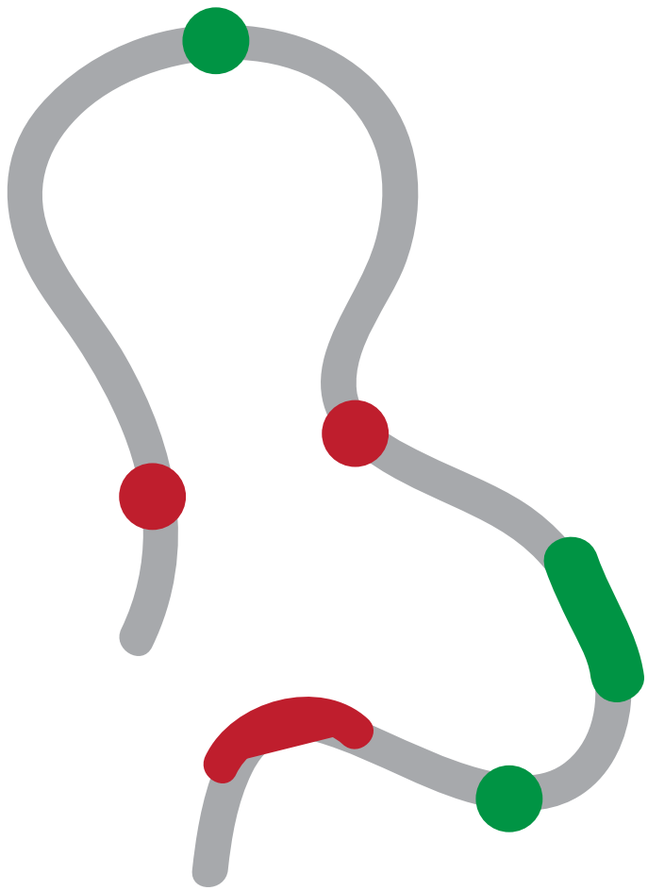
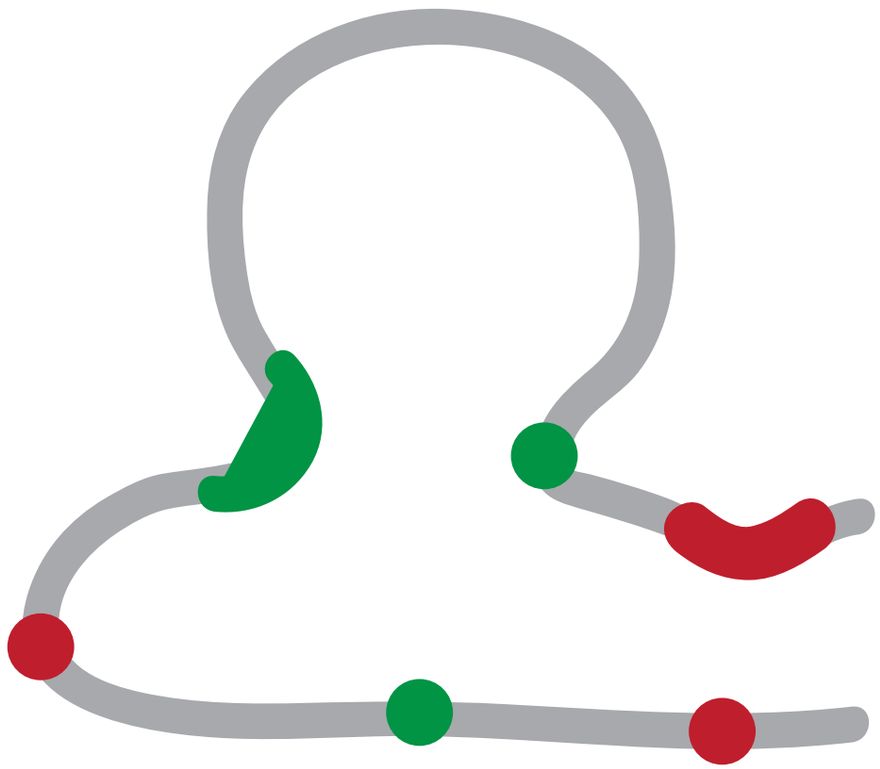
Structure determination of genomes and genomic domains by satisfaction of spatial restraints

Marc A. Marti-Renom
Structural Genomics Group (ICREA, CNAG-CRG)

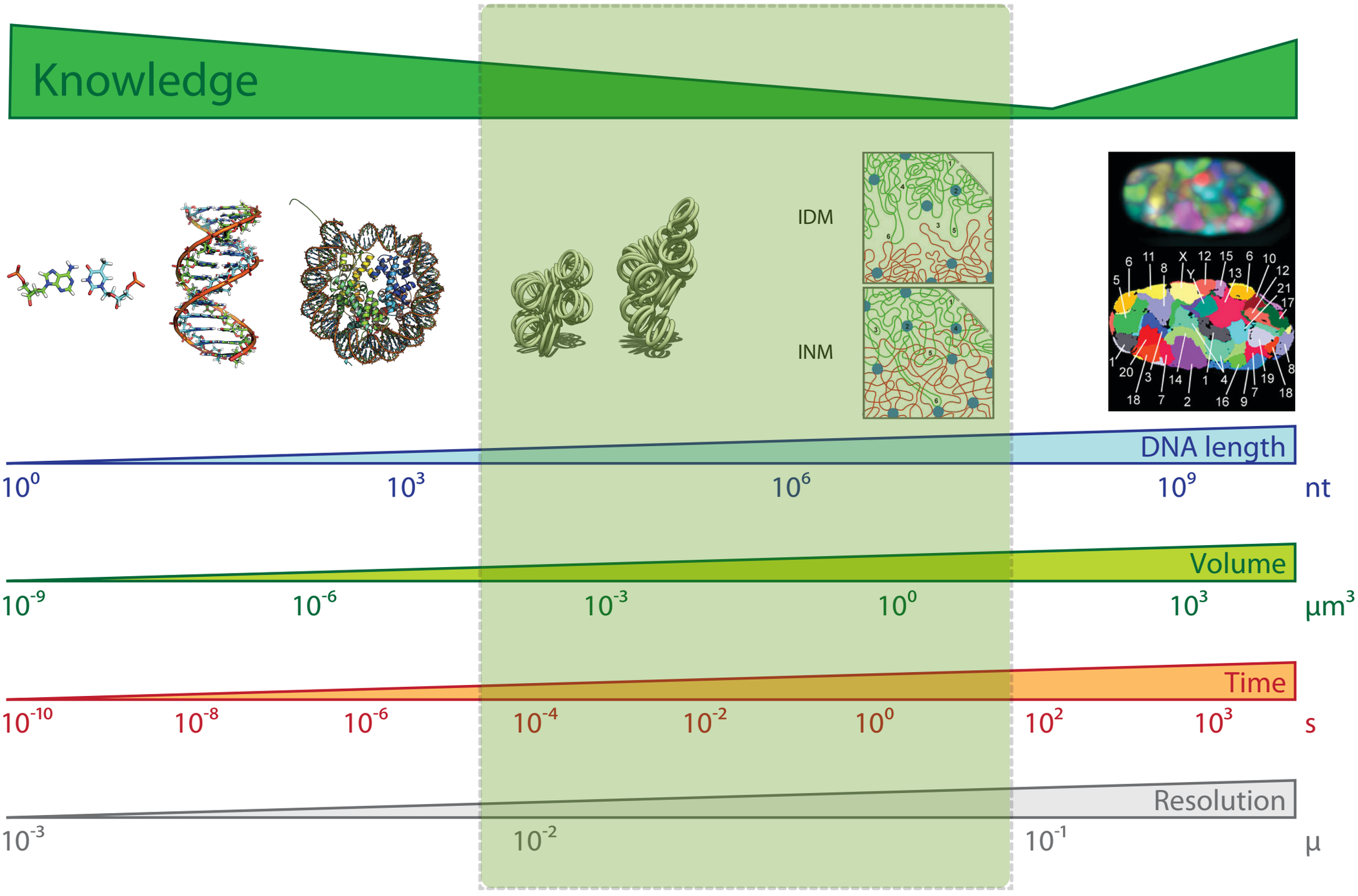
http://marciuslab.org
http://3DGenomes.org
http://cnag.crg.eu





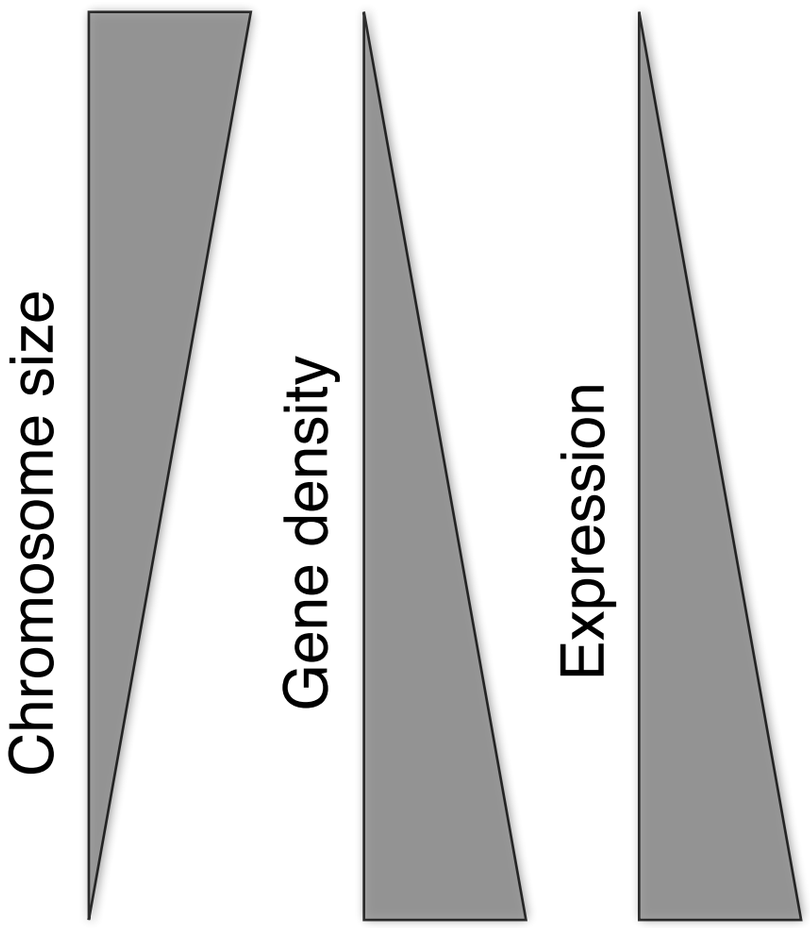
Resolution Gap

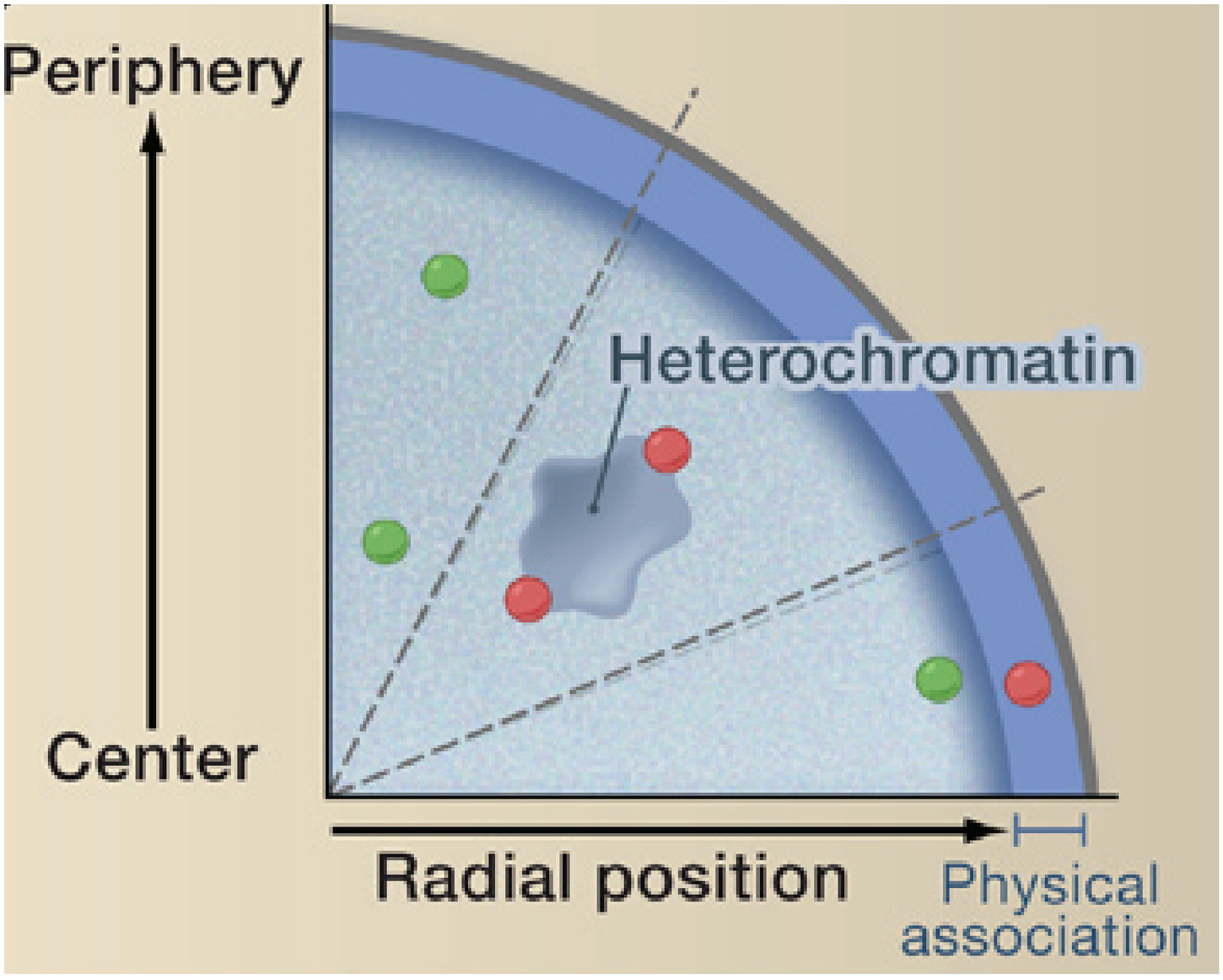
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol *7*, e1002125 (2011)



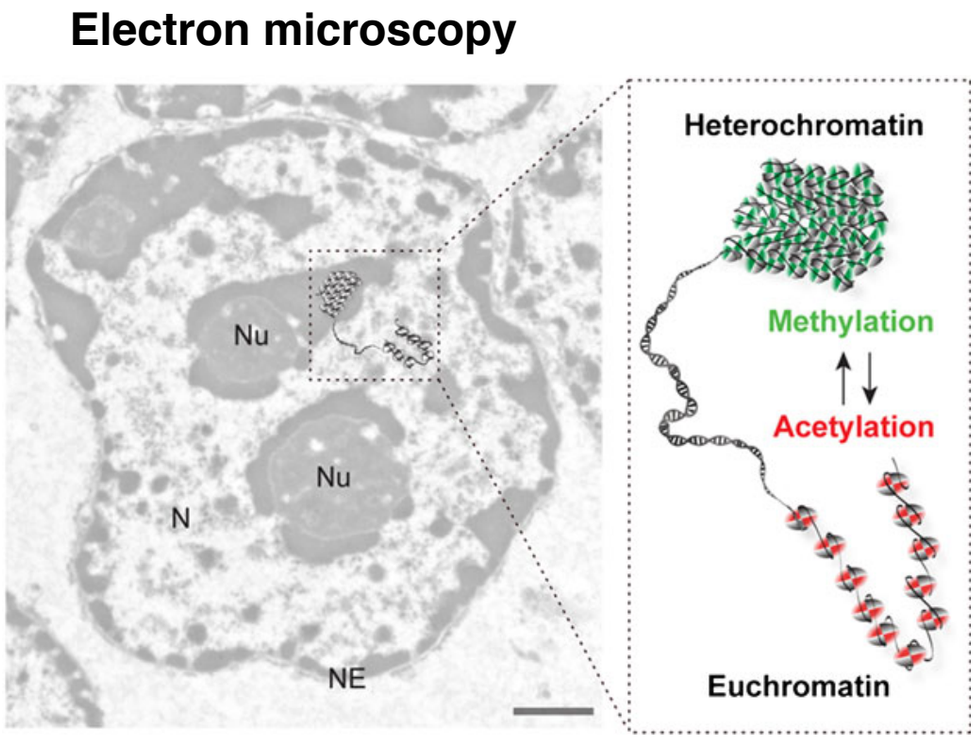
Level I: Radial genome organization

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





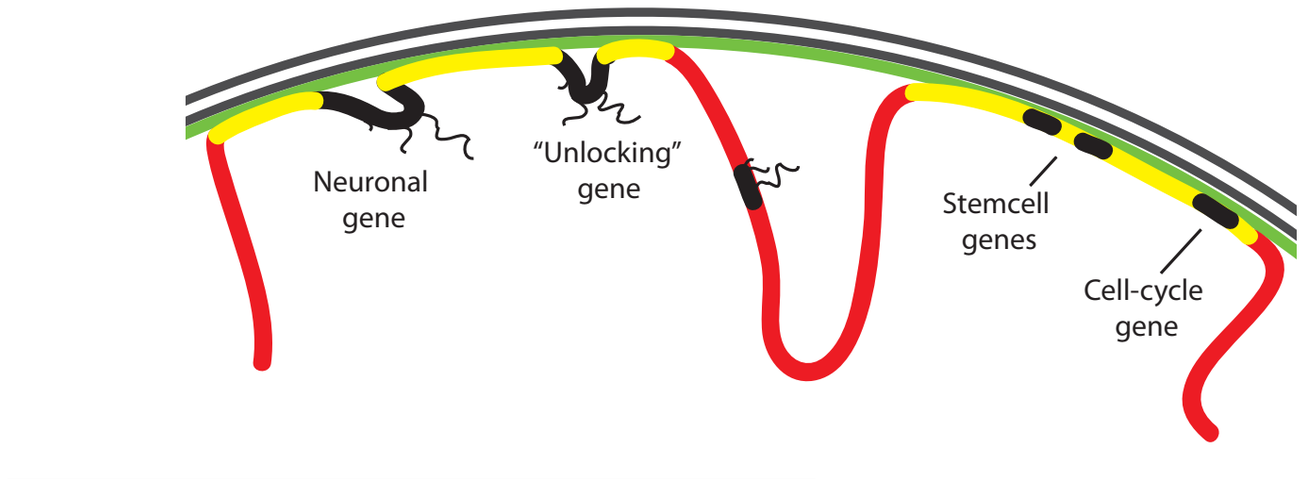
Level II: Euchromatin vs heterochromatin



Euchromatin:
chromatin that is located away from the nuclear lamina, is generally less densely packed, and contains actively transcribed genes

Heterochromatin:
chromatin that is near the nuclear lamina, tightly condensed, and transcriptionally silent

Level III: Lamina-genome interactions

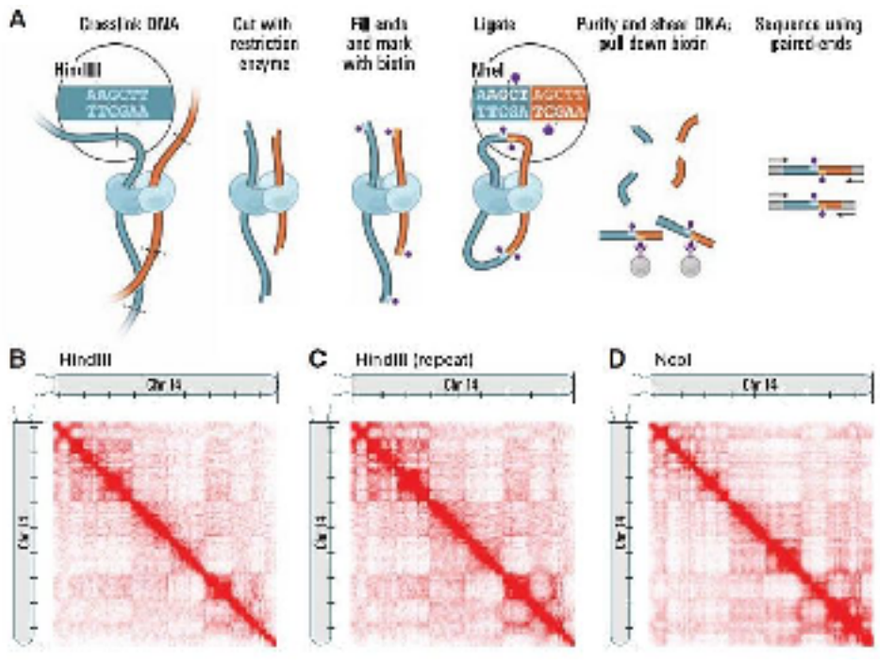


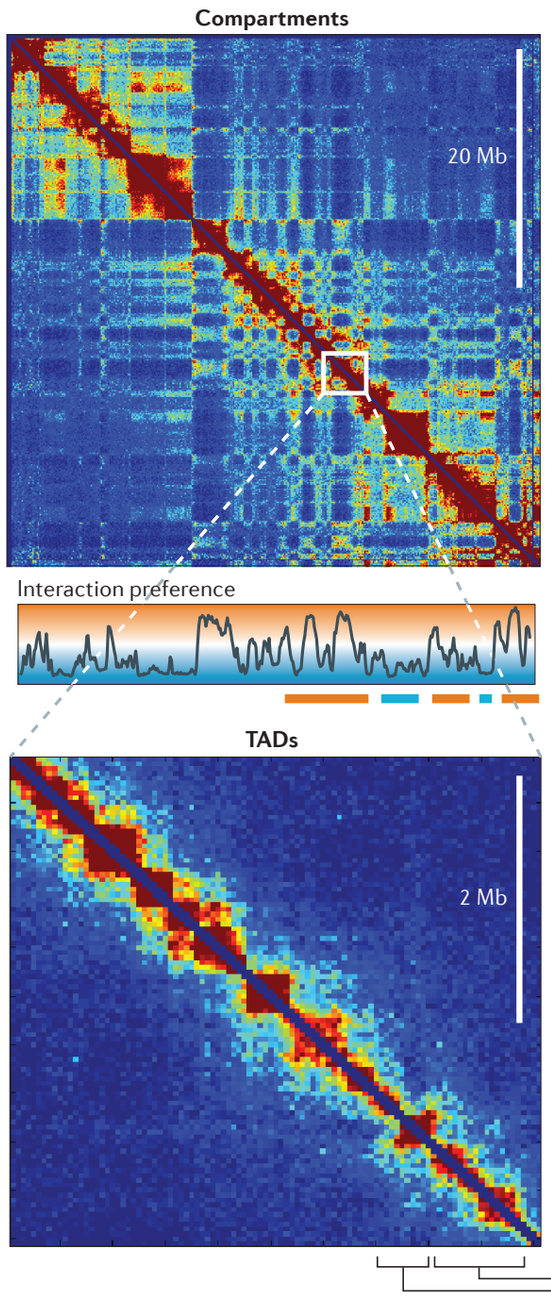
Most genes in Lamina Associated Domains are transcriptionally silent, suggesting that **lamina-genome interactions** are widely involved in the control of **gene expression**

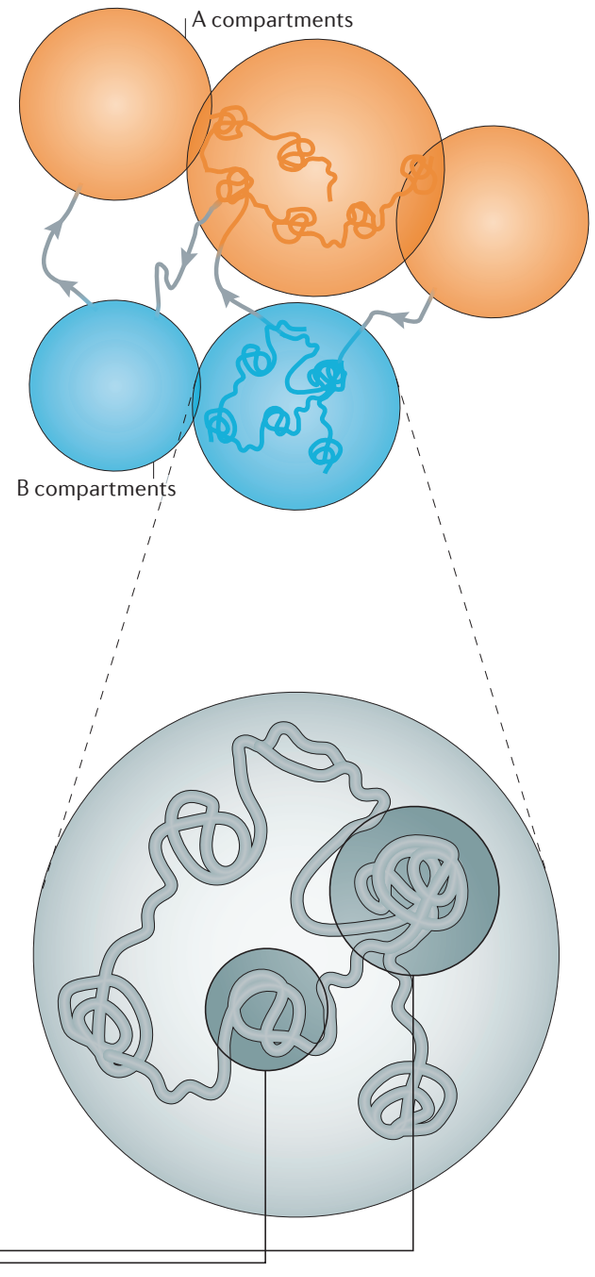
Adapted from Molecular Cell 38, 603-613, 2010

Level IV: Higher-order organization

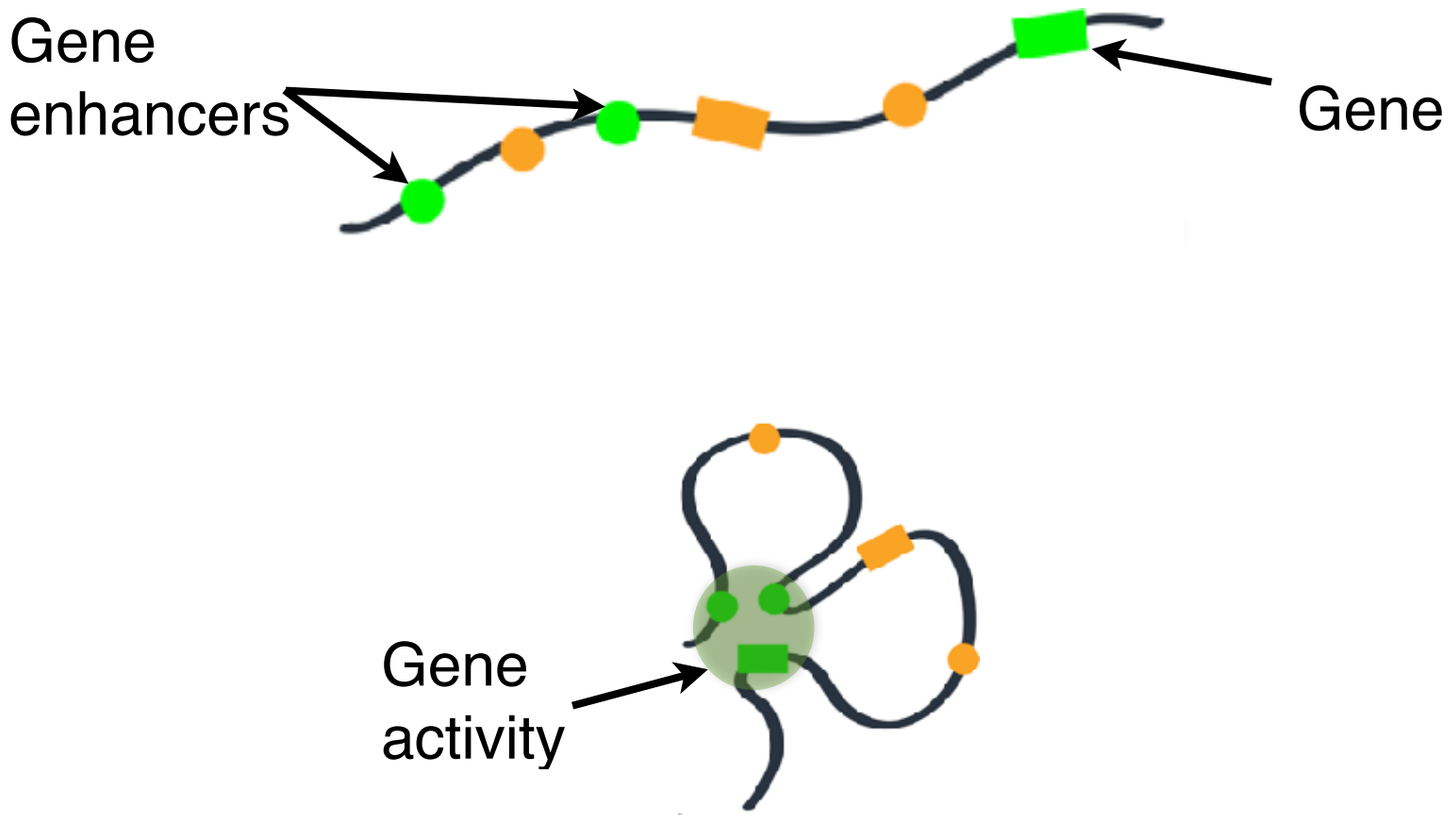
Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Nat Rev Genet 14, 390–403 (2013).







Level V: Chromatin loops



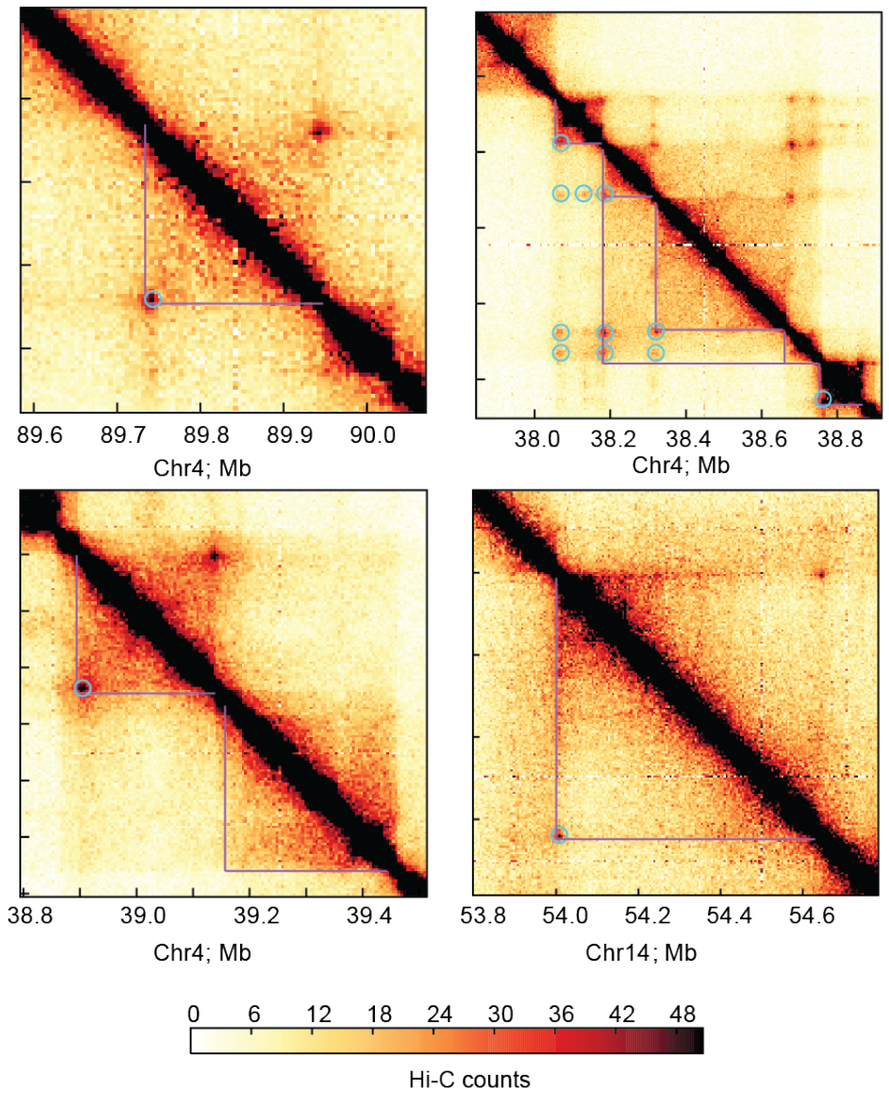
Loops bring distal genomic regions in close proximity to one another

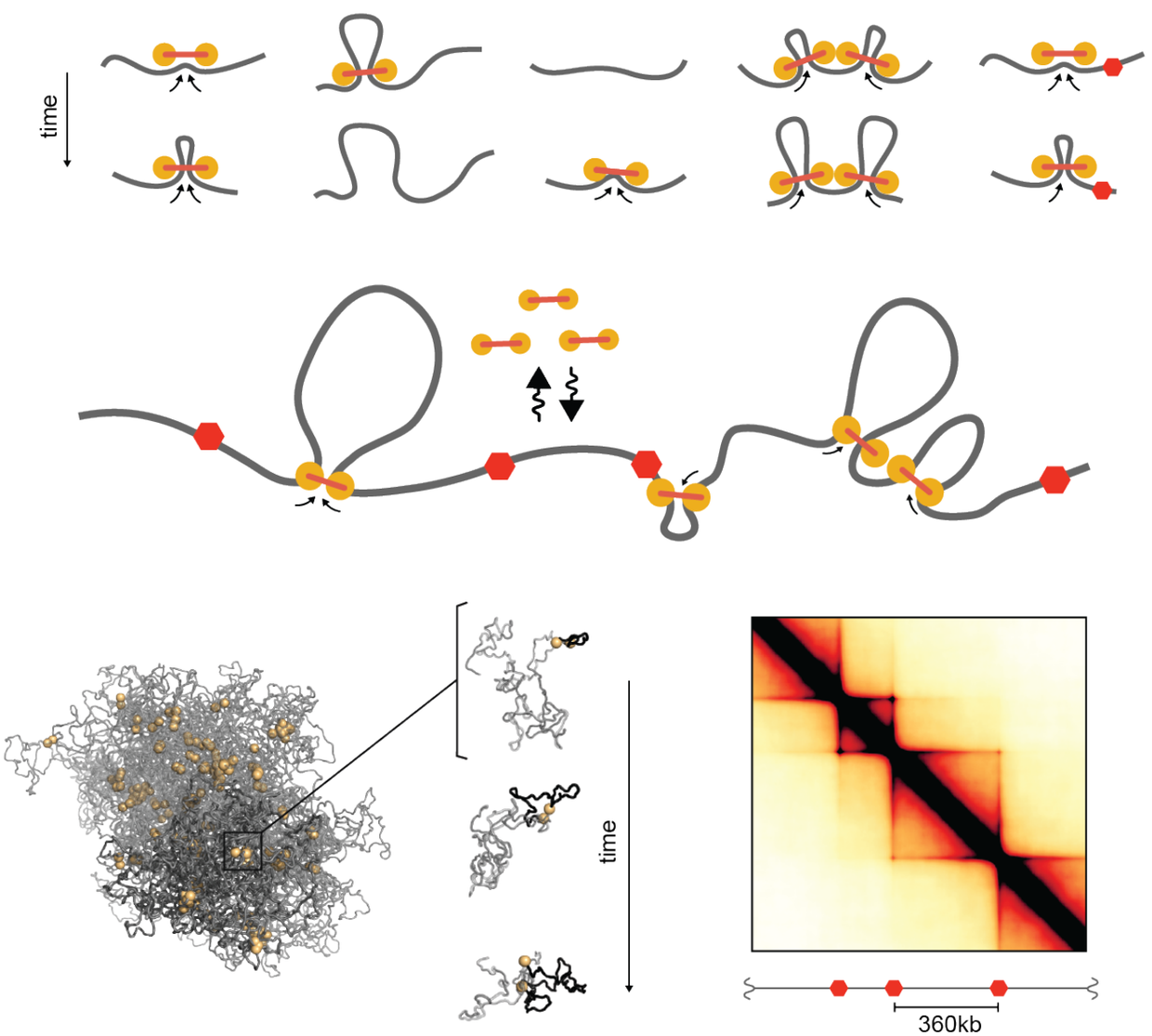
This in turn can have profound effects on gene transcription

Enhancers can be thousands of kilobases away from their target genes in any direction (or even on a separate chromosome)

Level V: Loop-extrusion as a driving force

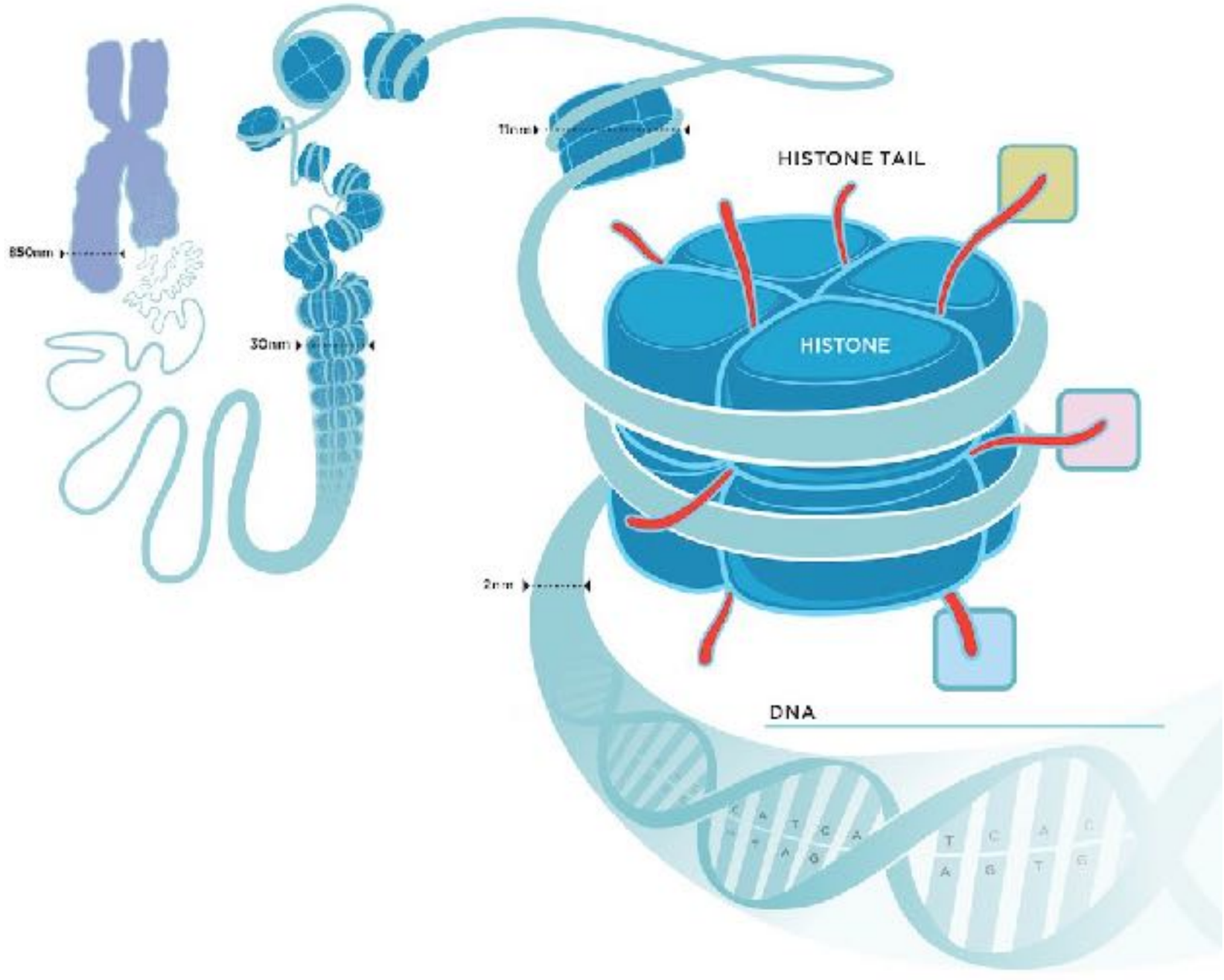
Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2015).
Formation of Chromosomal Domains by Loop Extrusion. bioRxiv.





Level VI: Nucleosome

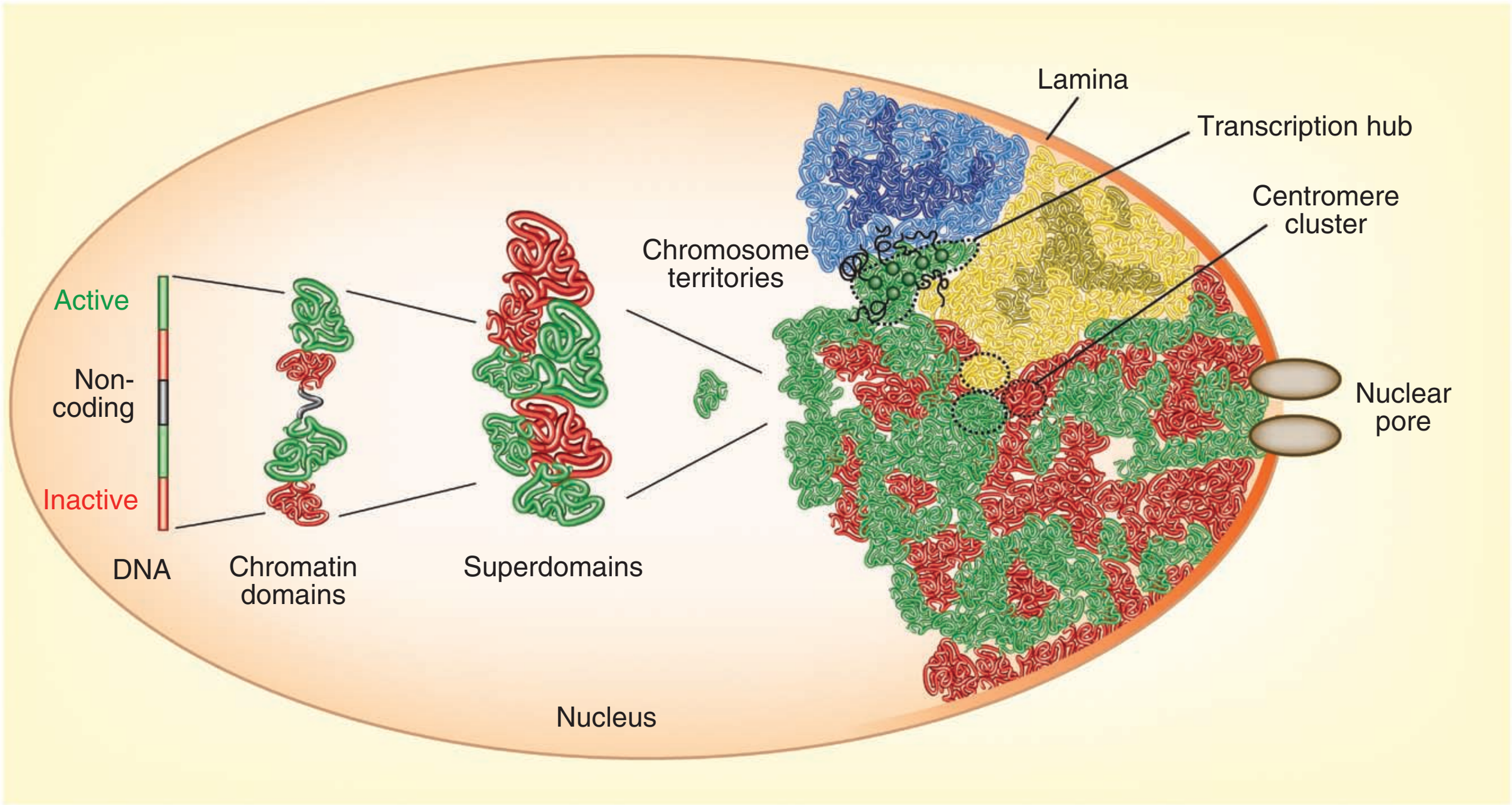
Chromosome **Chromatin fibre** **Nucleosome**



Adapted from Richard E. Ballermann, 2012

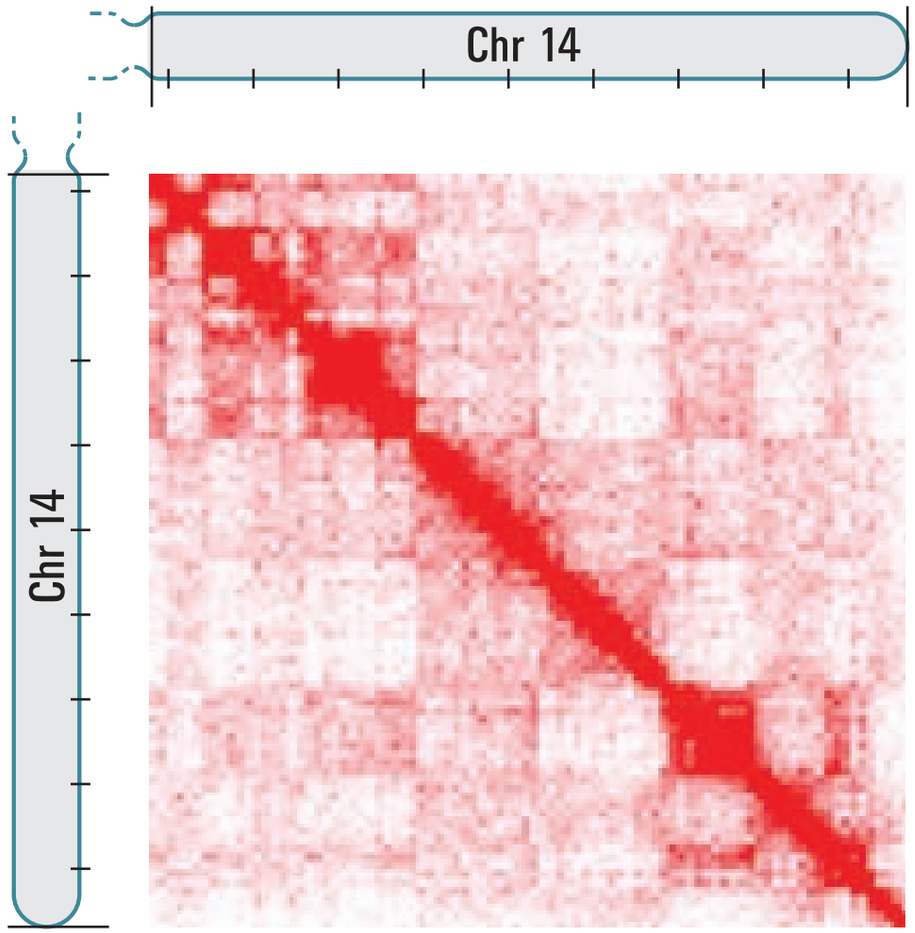
Complex genome organization

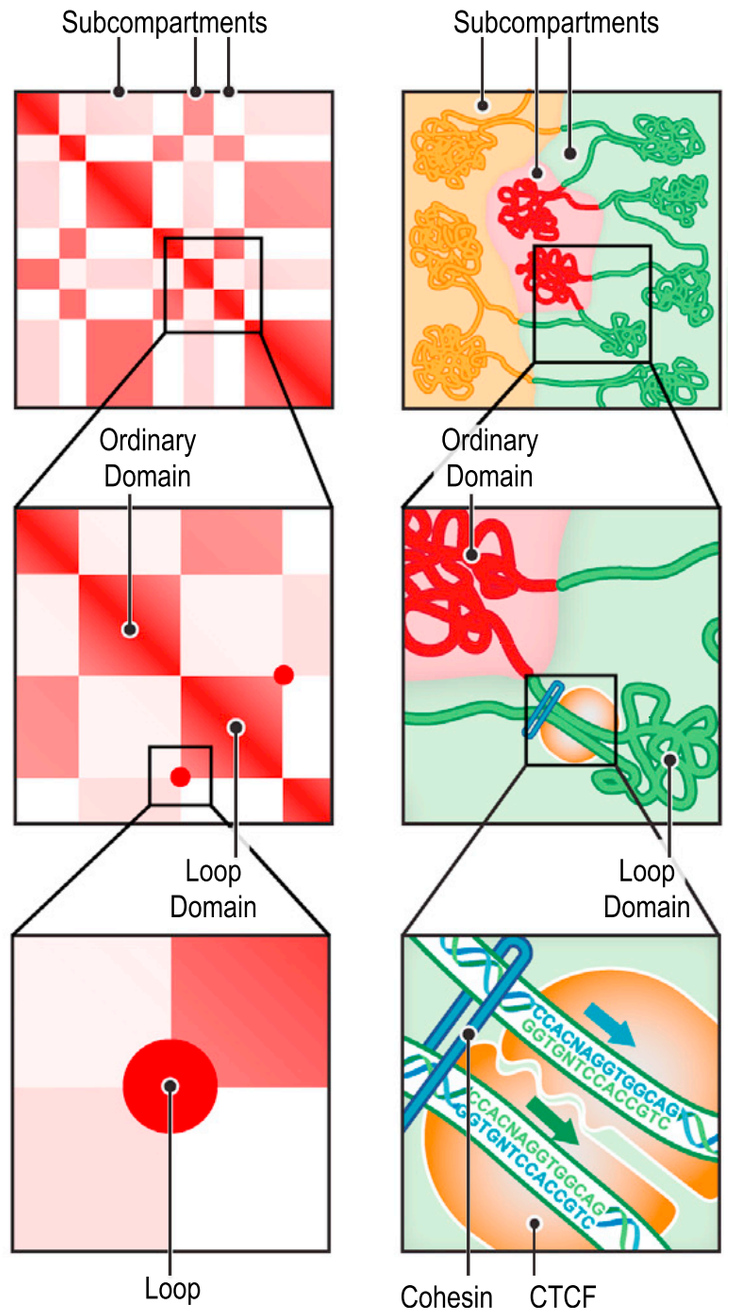
Cavalli, G. & Misteli, T. Functional implications of genome topology. Nat Struct Mol Biol 20, 290–299 (2013).



Marina Corral

Hierarchical genome organisation

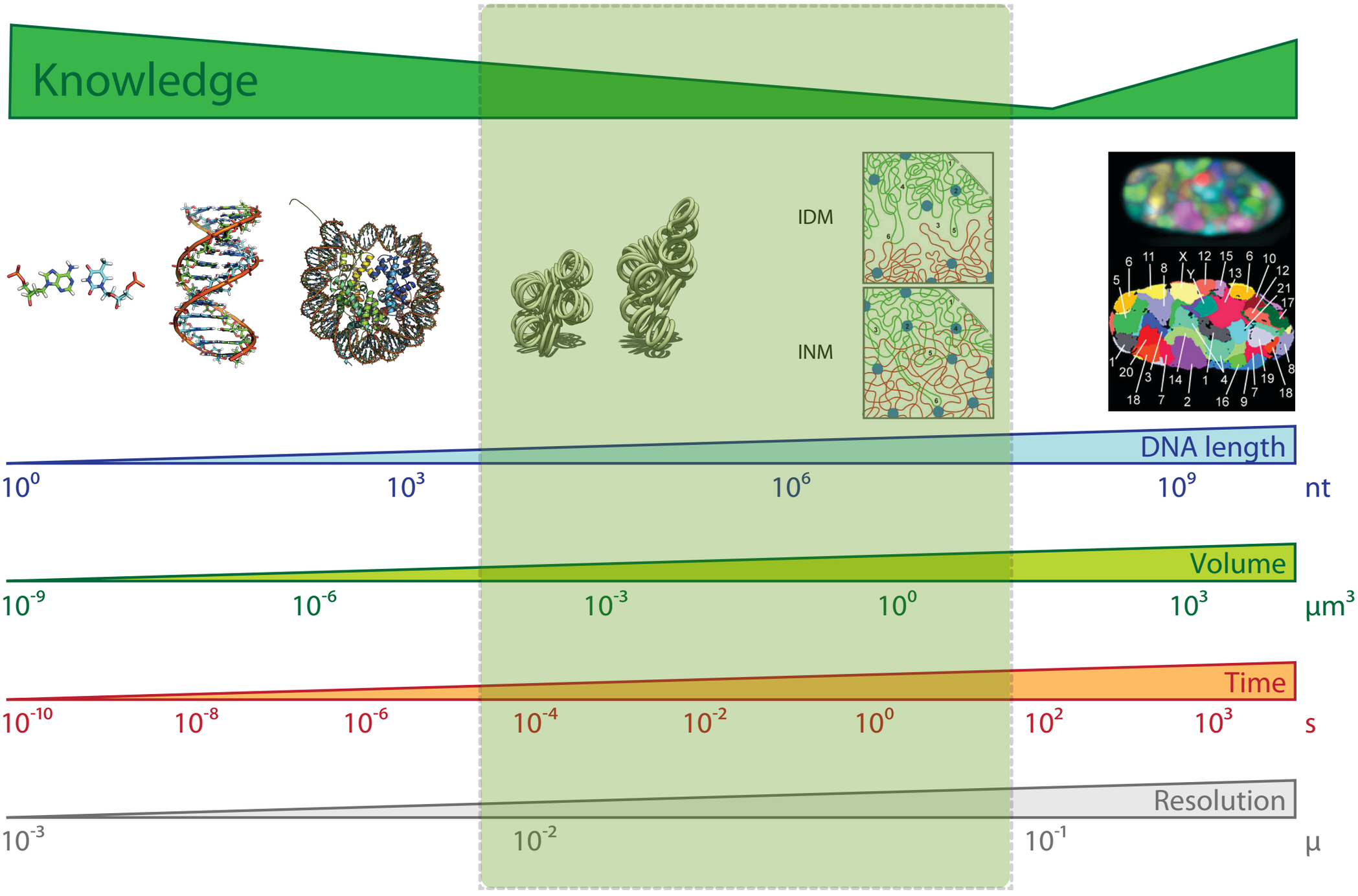




Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.
Rao, S. S. P., et al. (2014). Cell, 1–29.

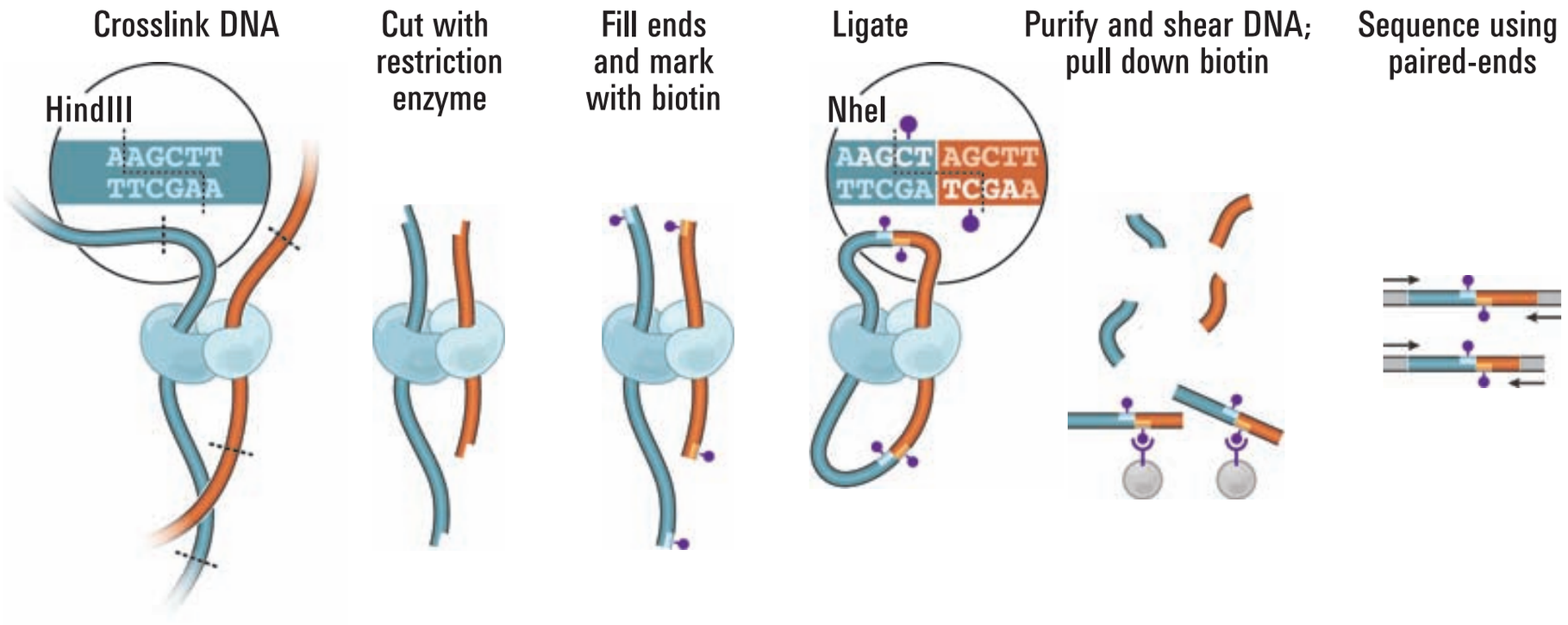
Resolution Gap

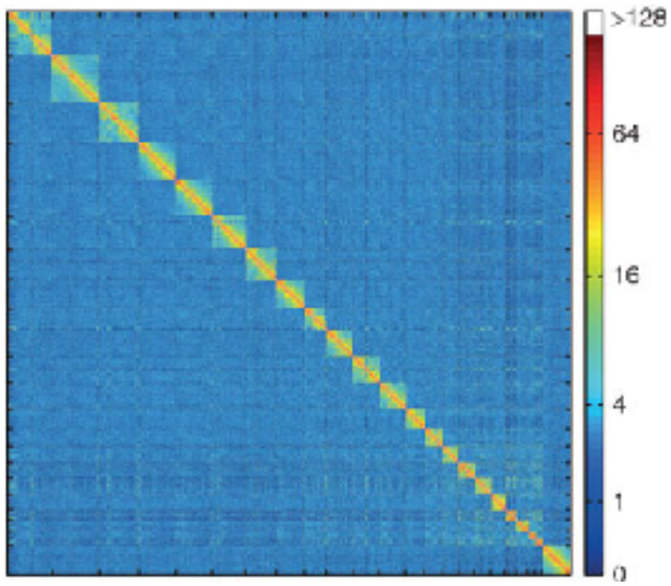
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol *7*, e1002125 (2011)



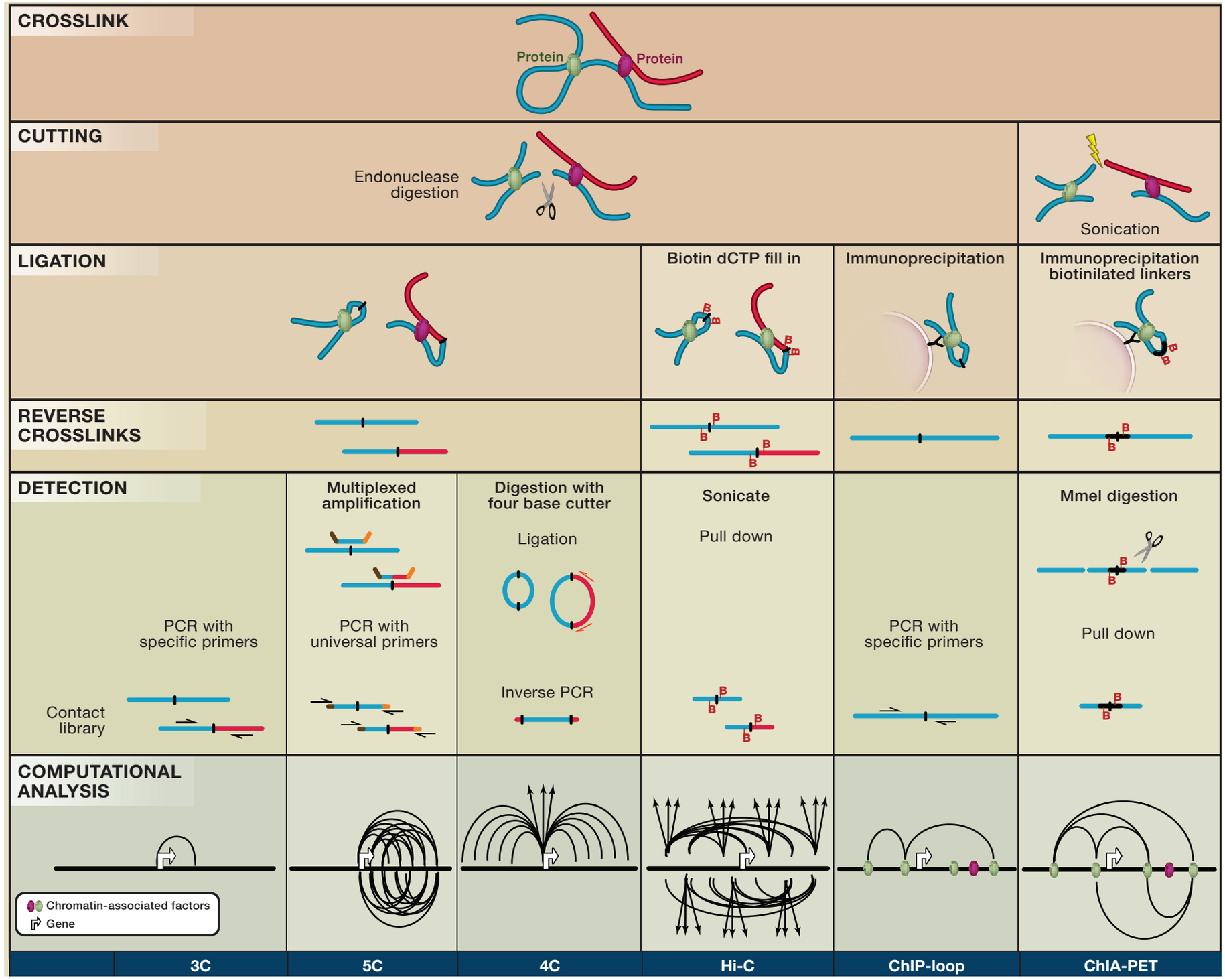
Chromosome Conformation Capture

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.



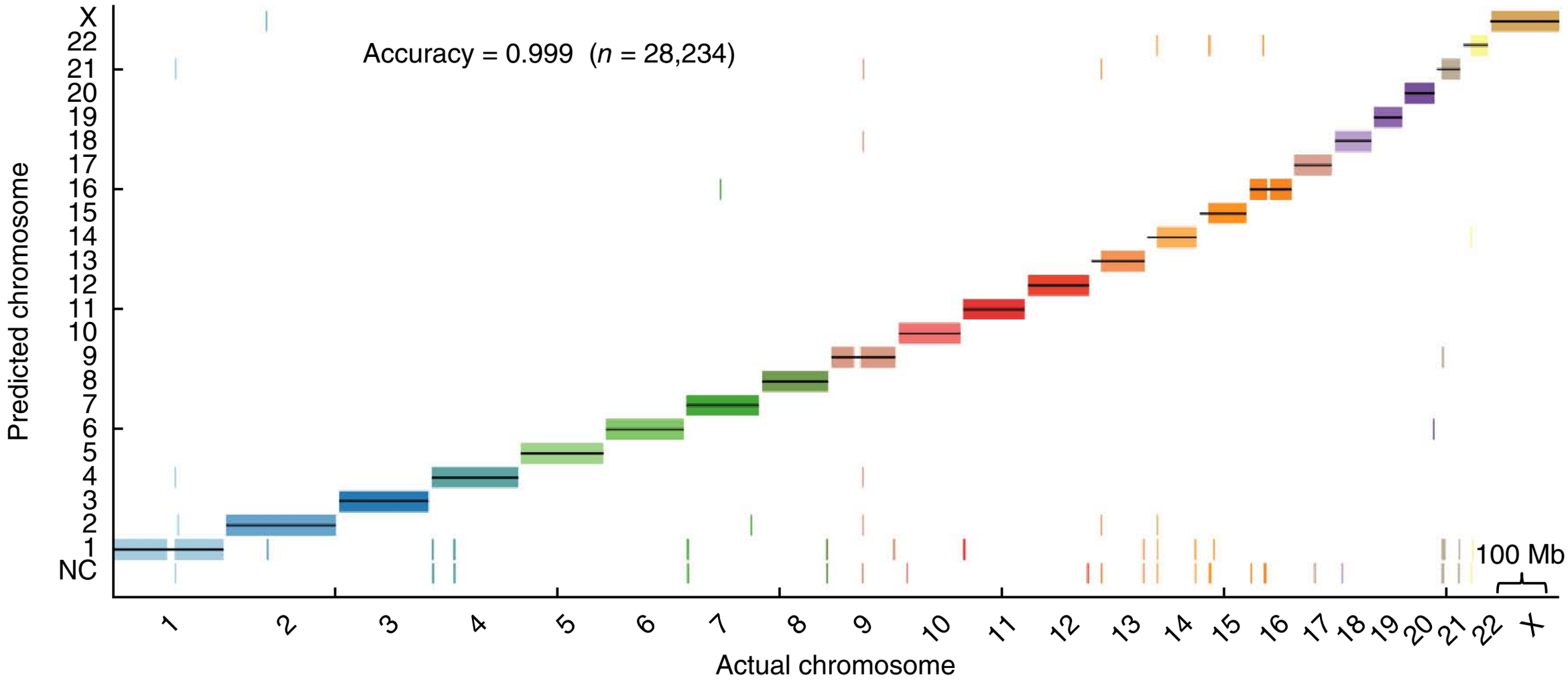


Chromosome Conformation Capture



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

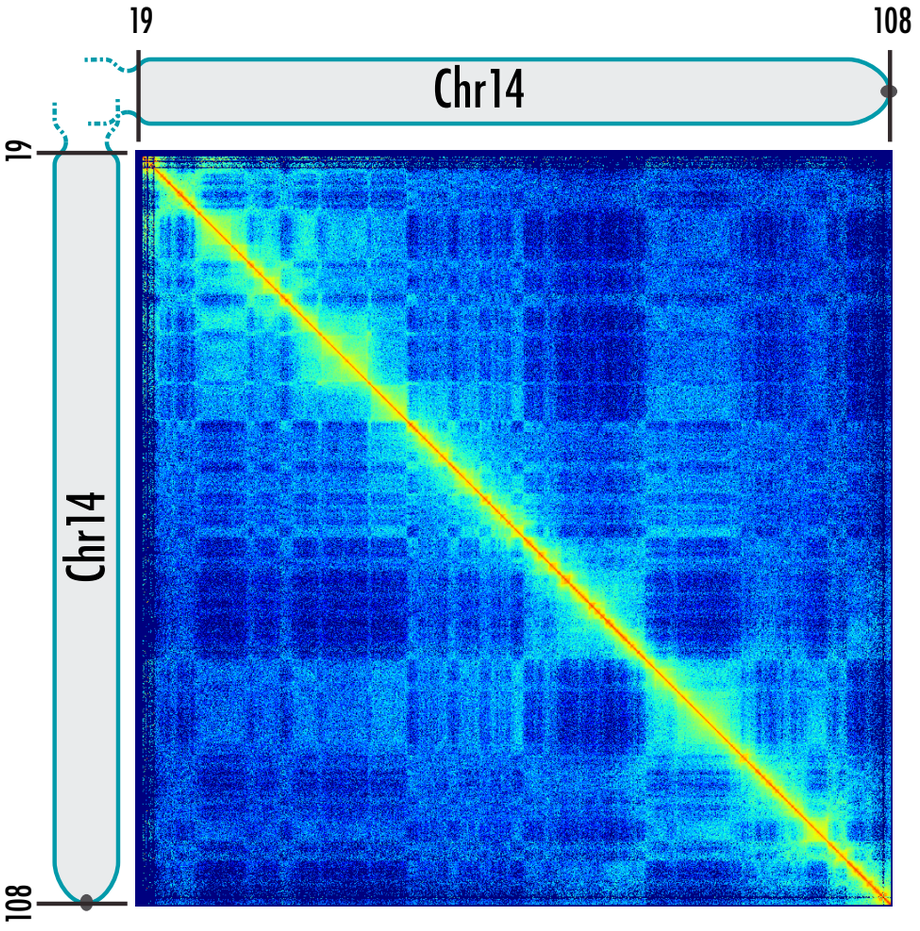
Chromosome Conformation Capture for de-novo assembly

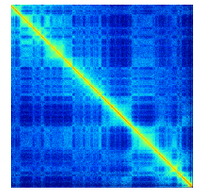


Kaplan, N., & Dekker, J. (2013). High-throughput genome scaffolding from in vivo DNA interaction frequency. Nature Biotechnology, 31(12), 1143–1147.

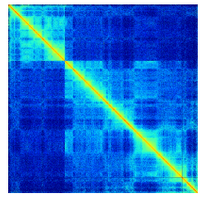
Great apes lymphoblast maps

Chromosome 14

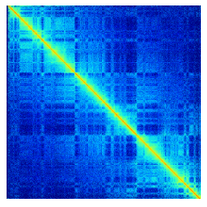




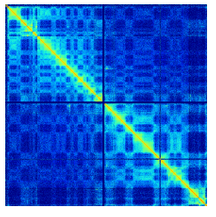
Chimpanzee



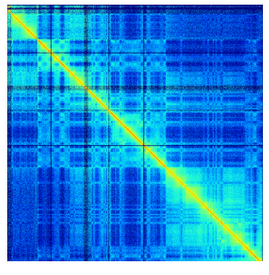
Gorilla



Orangutan



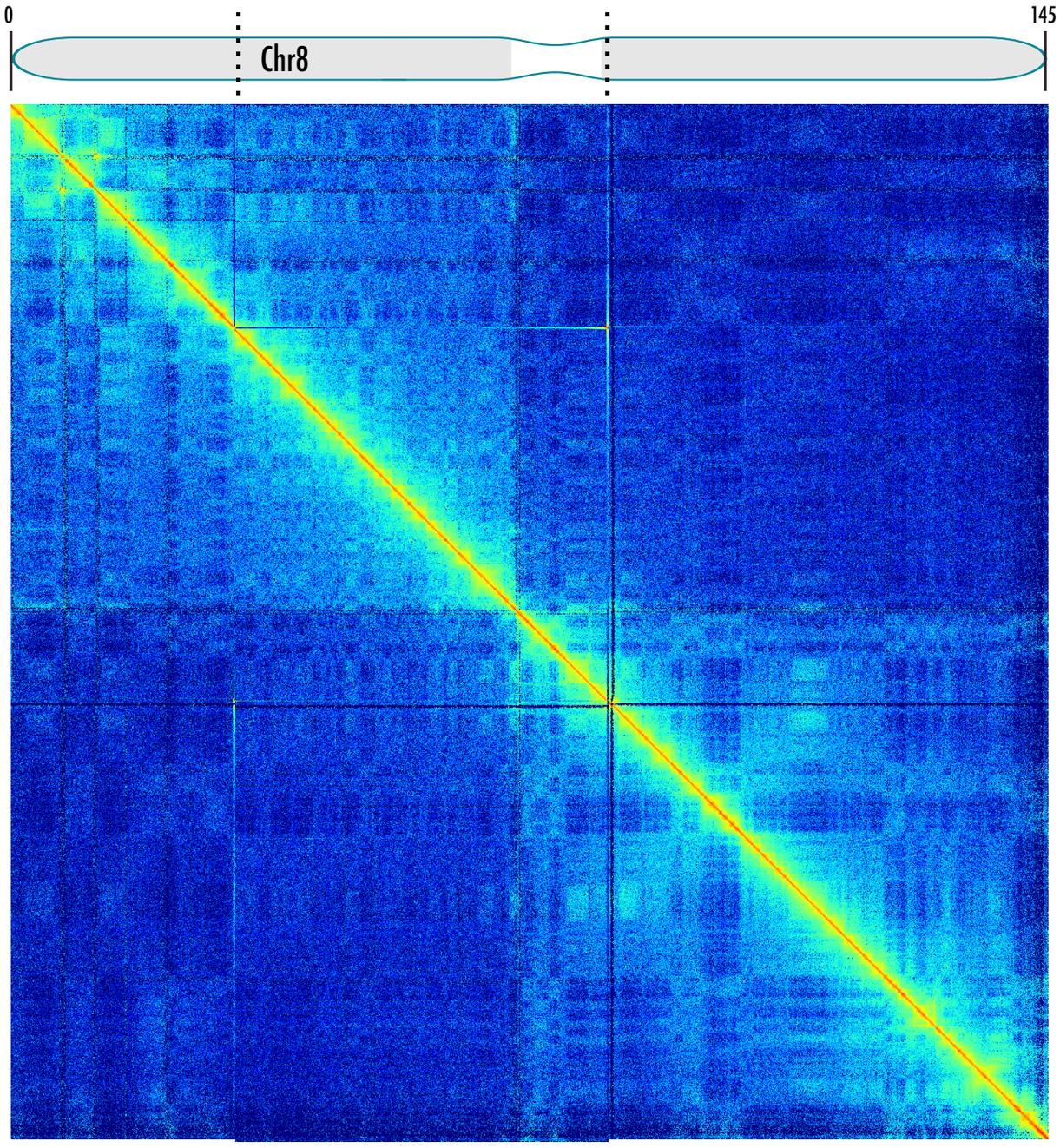
Gibbon



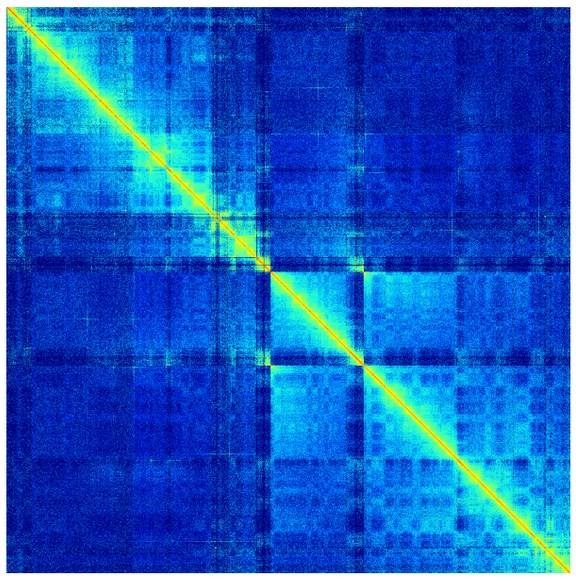
Mouse

Assembly error detection

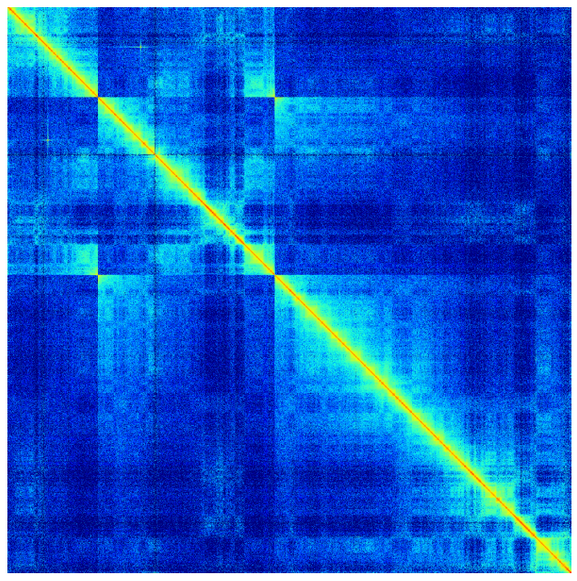
Chromosome 8 Gorilla



Chr 7

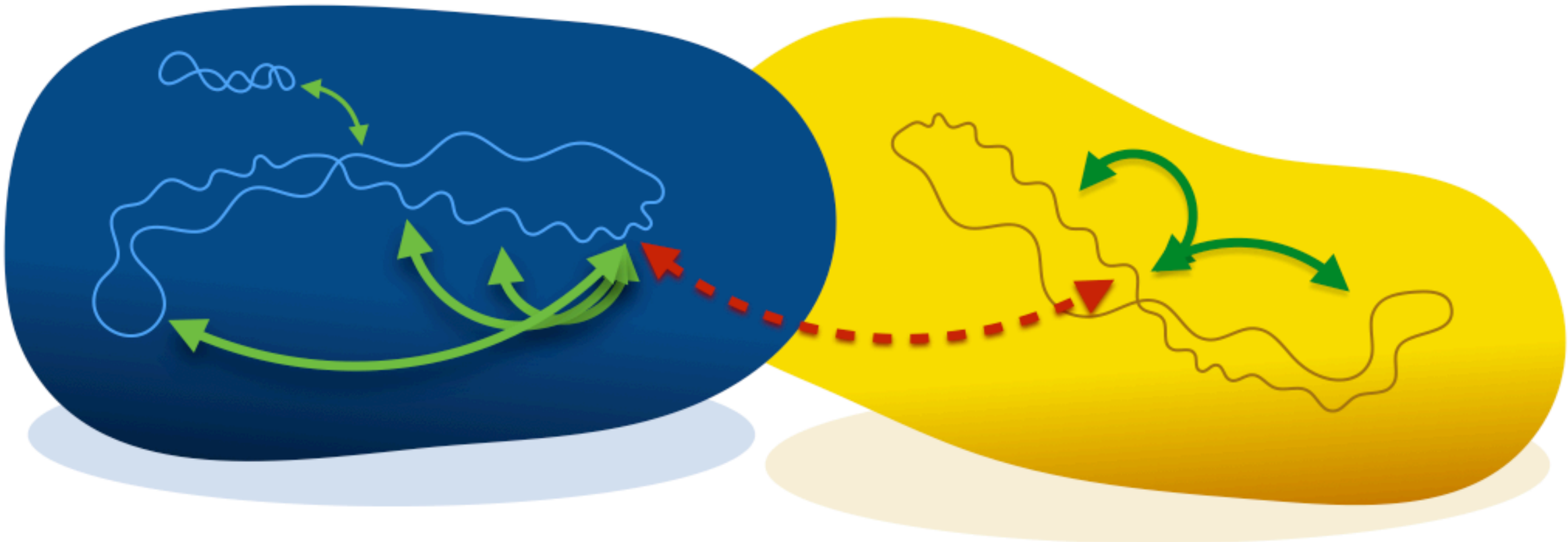


Chr 12



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

Chromosome Conformation Capture for meta genomics

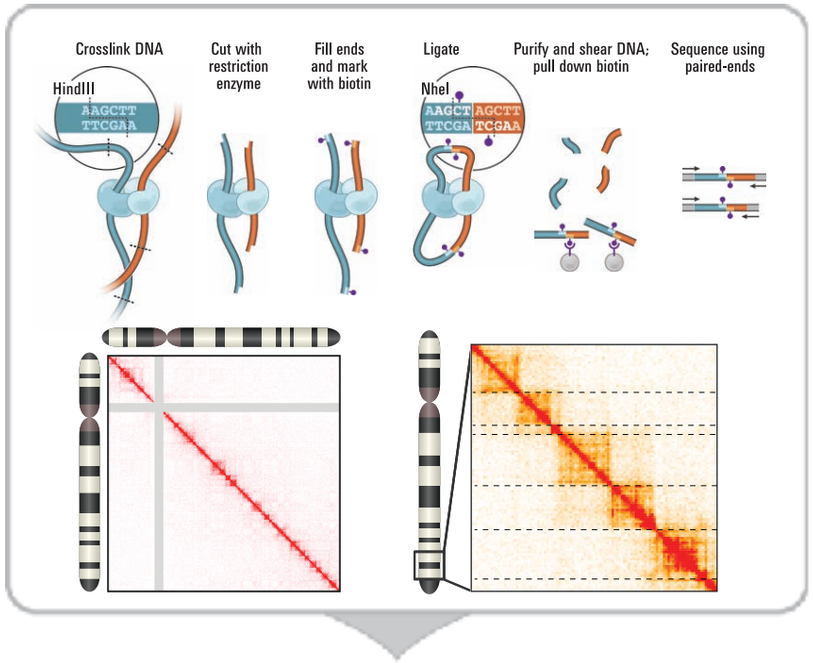


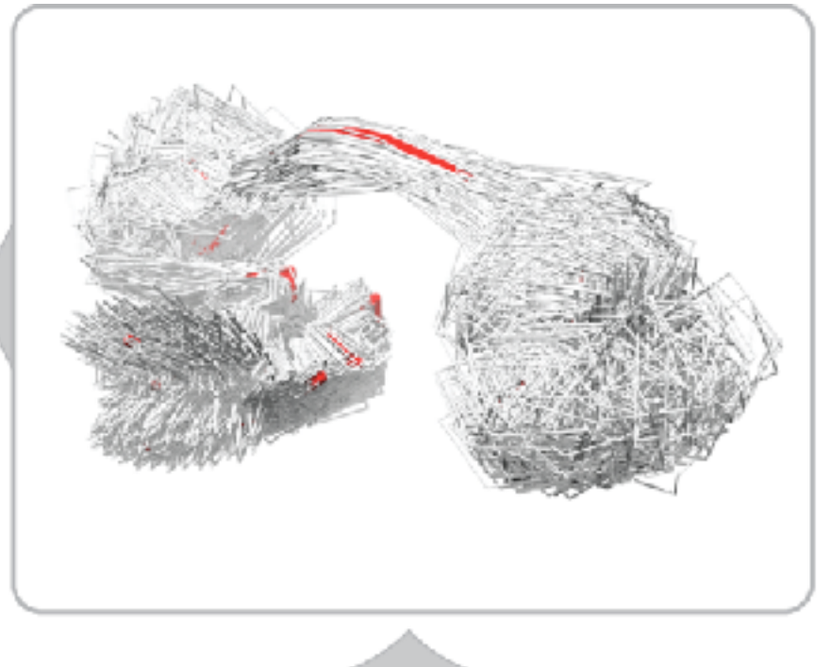
Beitel, C. W., Froenicke, L., Lang, J. M., Korf, I. F., Michelmore, R. W., Eisen, J. A., & Darling, A. E. (2014). Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. doi:10.7287/peerj.preprints.260v1

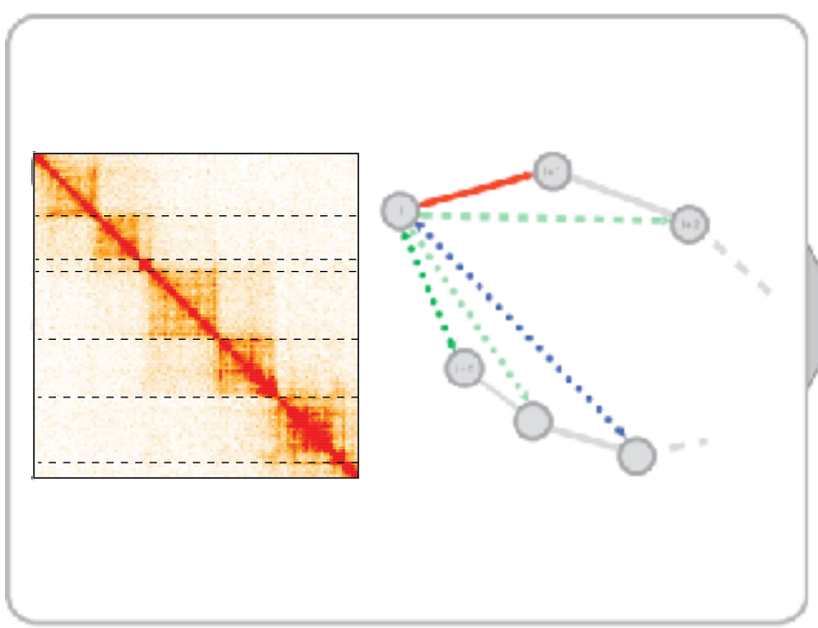
Hybrid Method

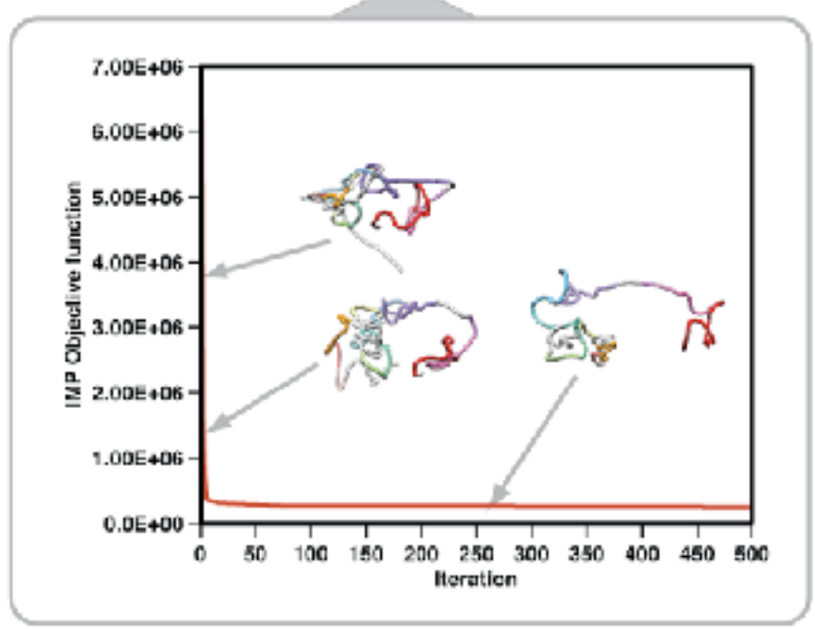
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).

Experiments



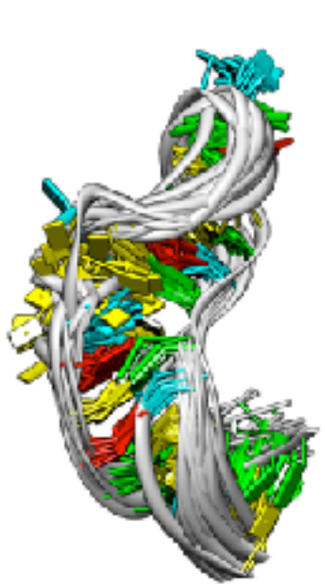


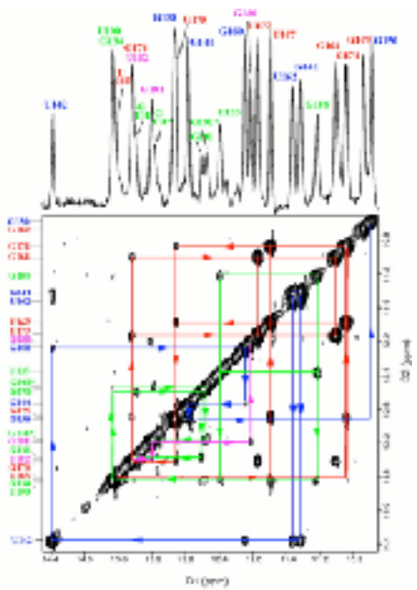




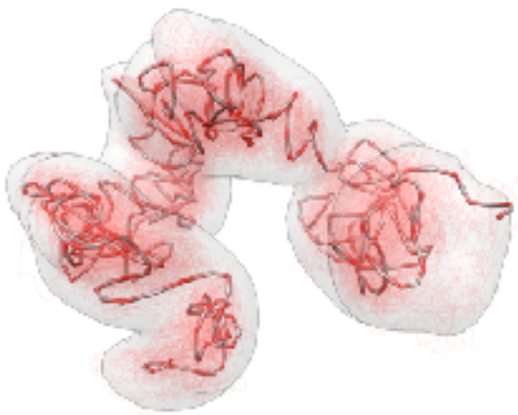
Computation

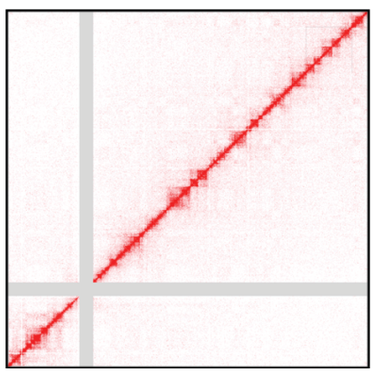
Structure determination using Hi-C data





Biomolecular structure determination
2D-NOESY data



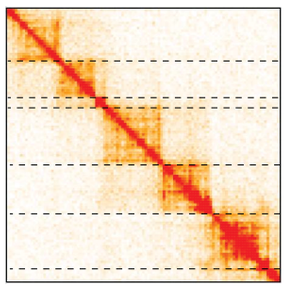


Chromosome structure determination
3C-based data

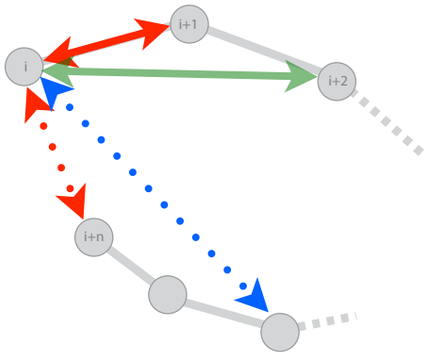
TADbit

http://3DGenomes.org

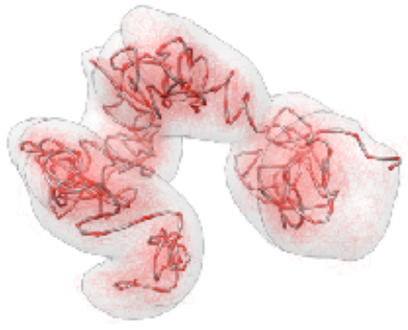
FastQ files to Maps



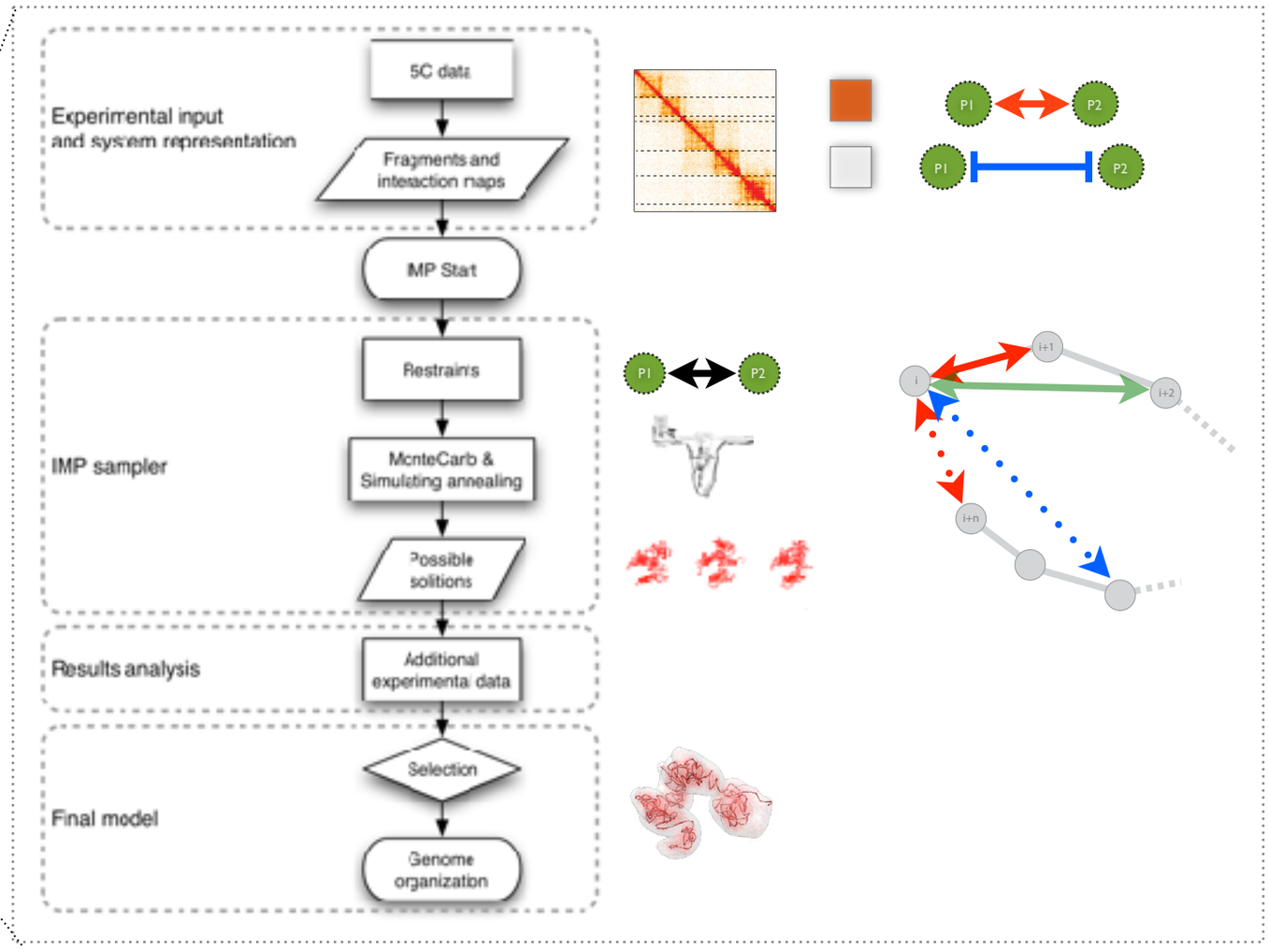
Map analysis



Model building



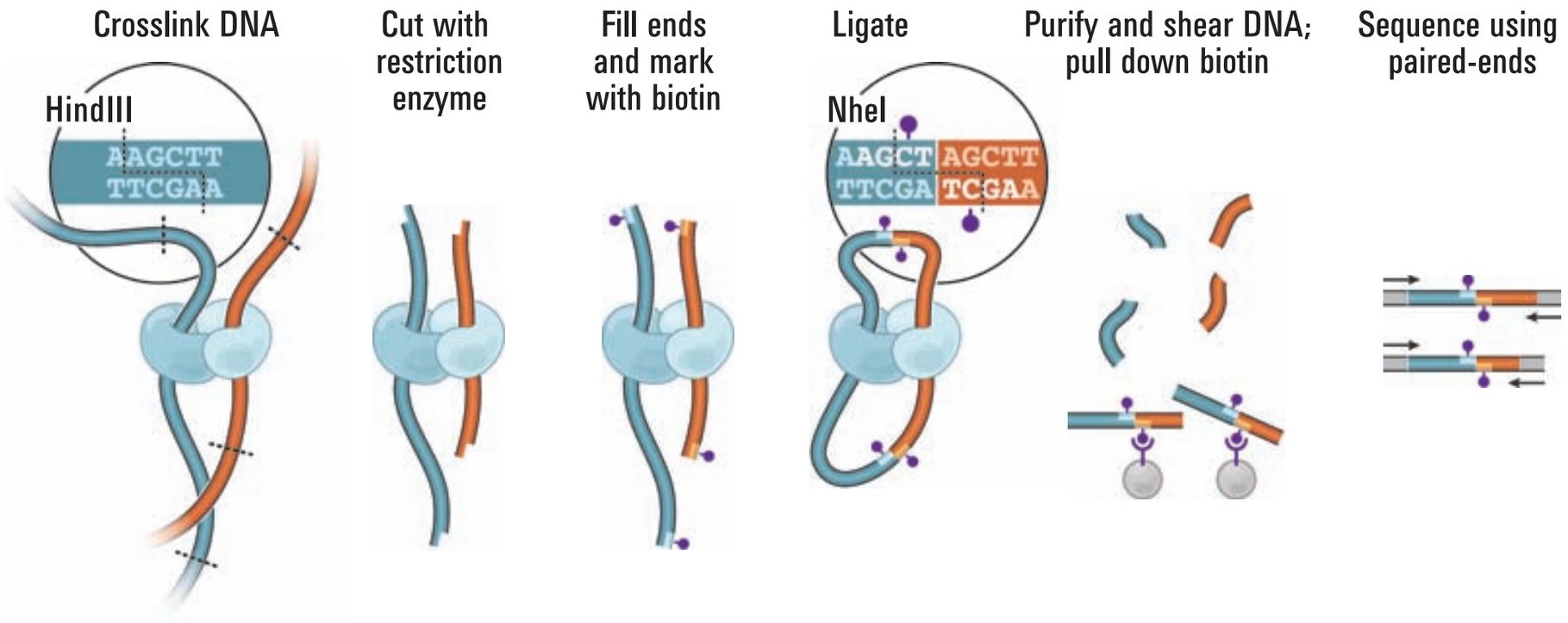
Model analysis

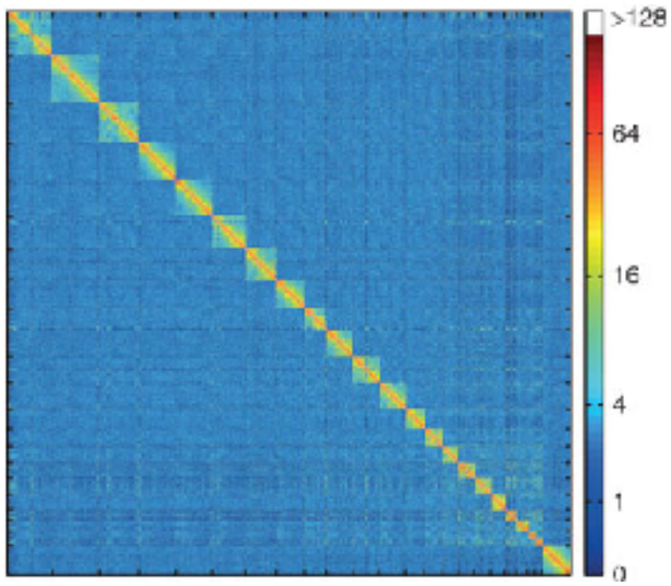



Got FASTQ?

Hi-C experiment

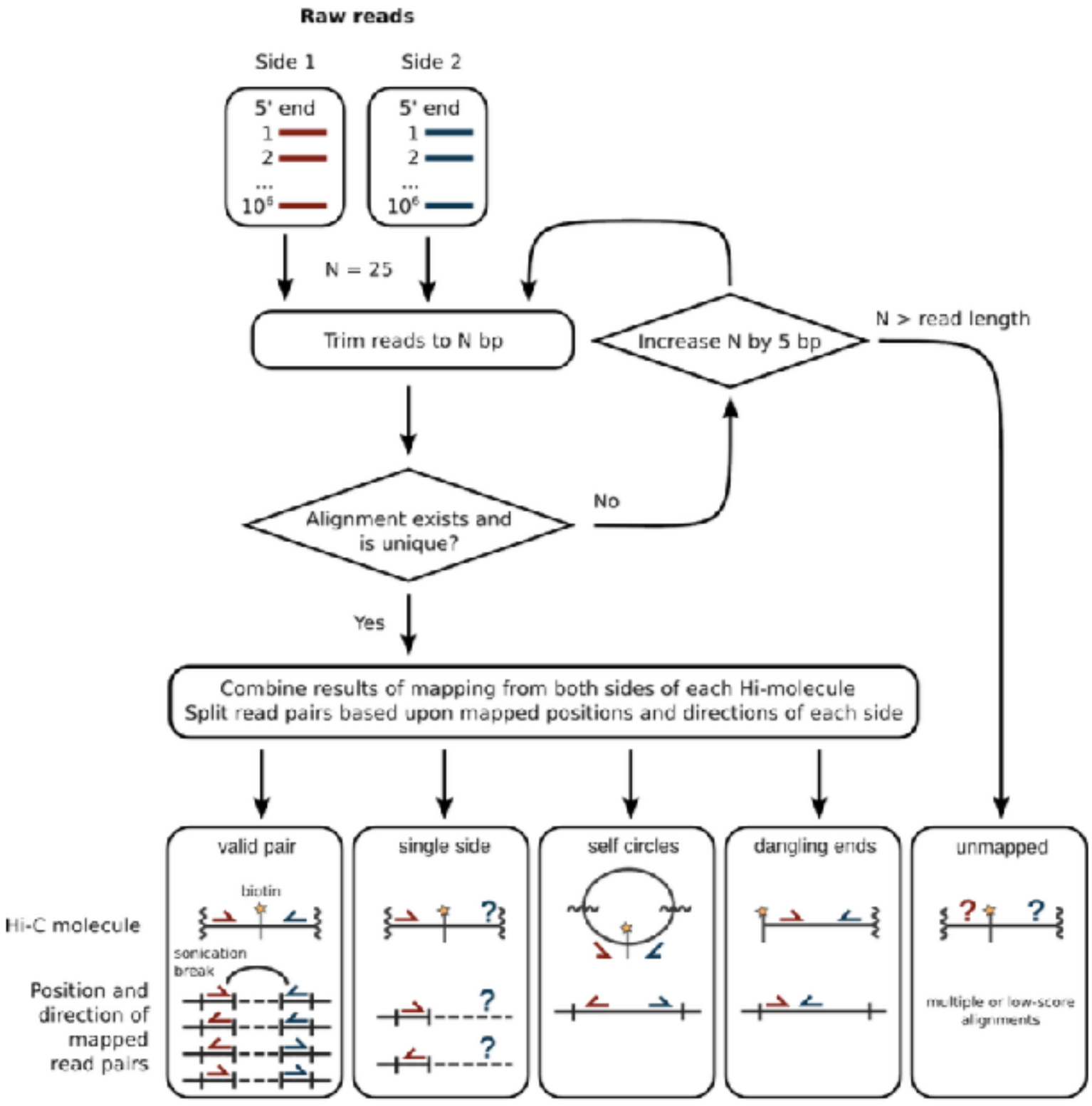
Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.





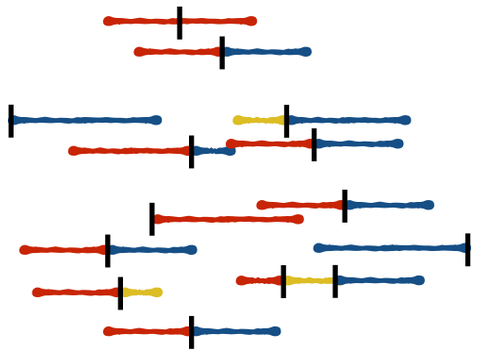
Mapping & Filtering

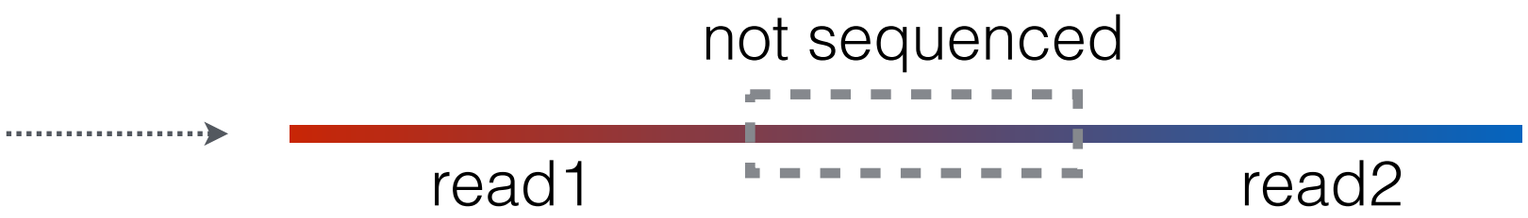
Imakaev, M. V et al. (2012). Nature Methods, 9(10), 999–1003.



Mapping @TADbit

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

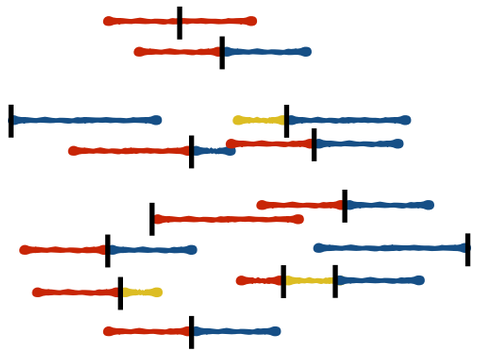




Iterative mapping

Mapping @TADbit

Serra et al. (2017). PLOS CompBio

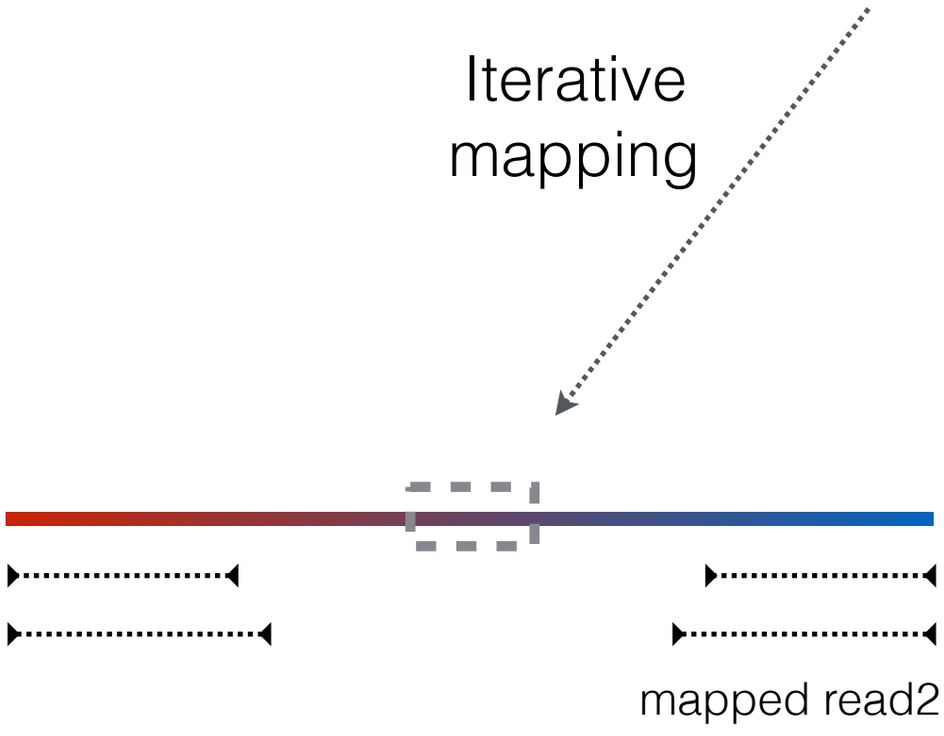


Iterative mapping

Mapping @TADbit

Serra et al. (2017). PLOS CompBio

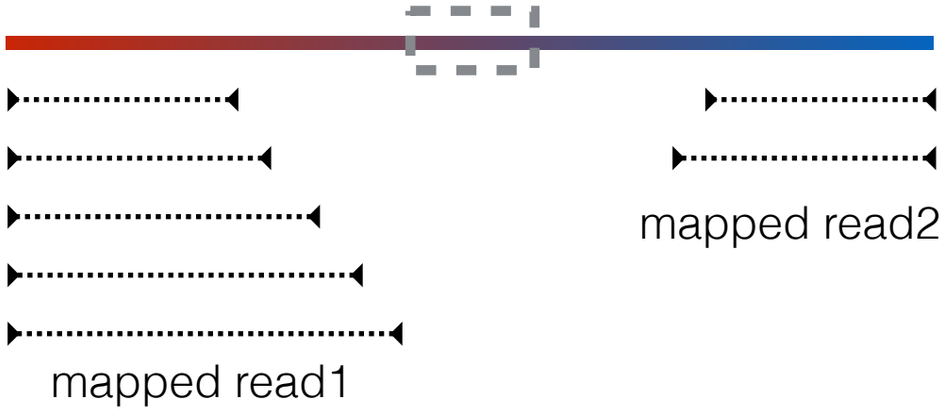
Iterative mapping



Mapping @TADbit

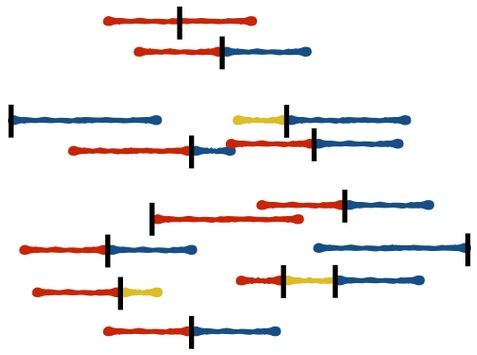
Serra et al. (2017). PLOS CompBio

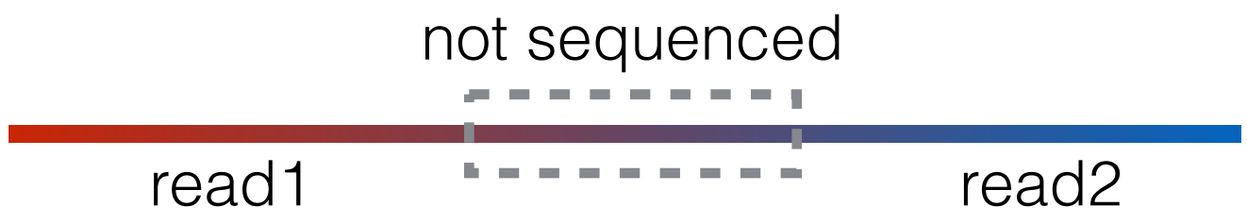
Iterative mapping



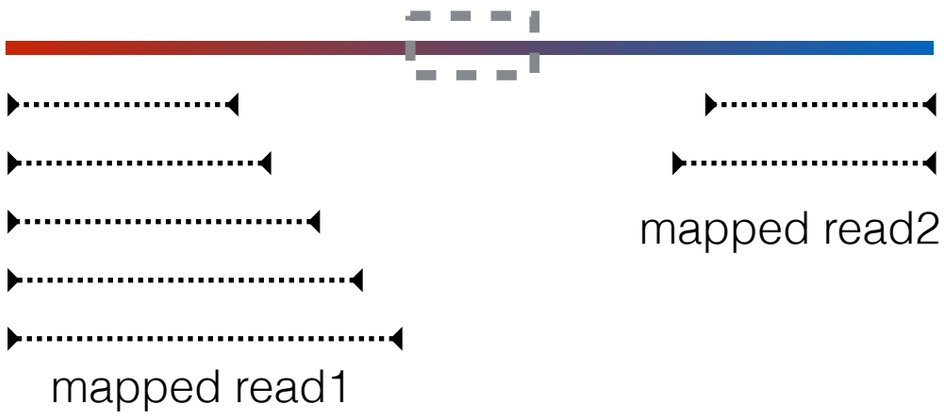
Mapping @TADbit

Serra et al. (2017). PLOS CompBio





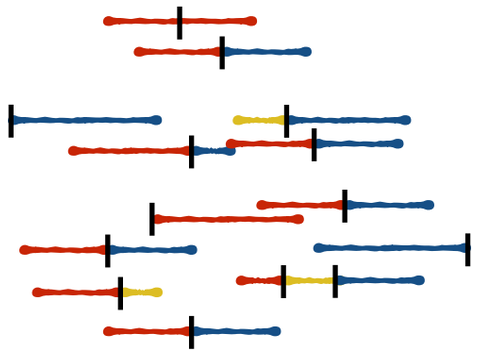
Iterative mapping

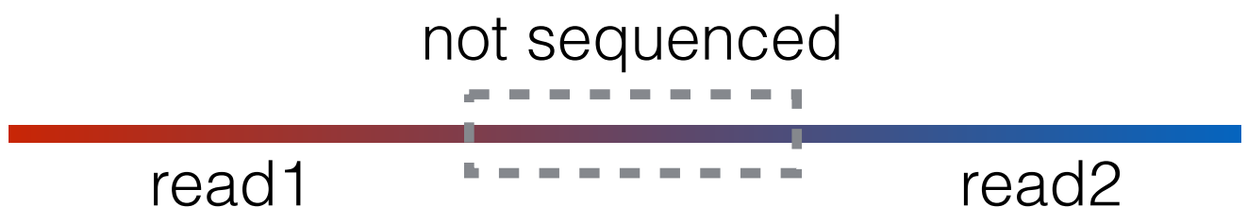


fragment based mapping

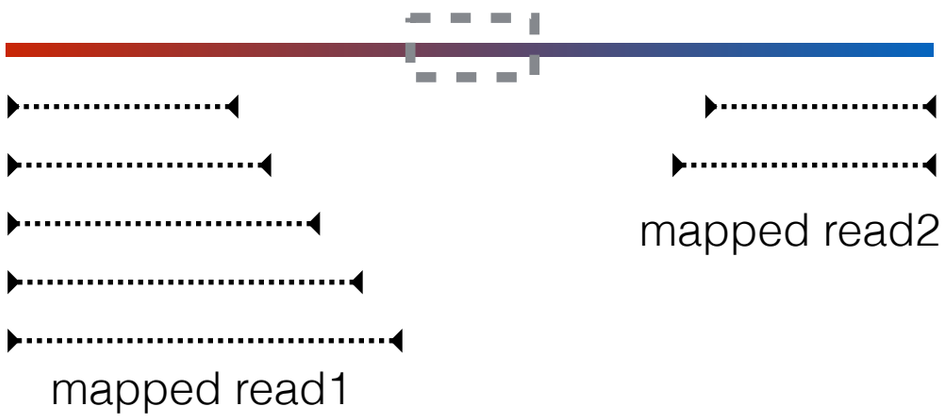
Mapping @TADbit

Serra et al. (2017). PLOS CompBio

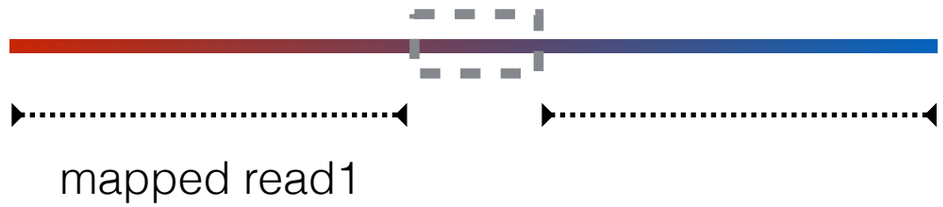




Iterative mapping

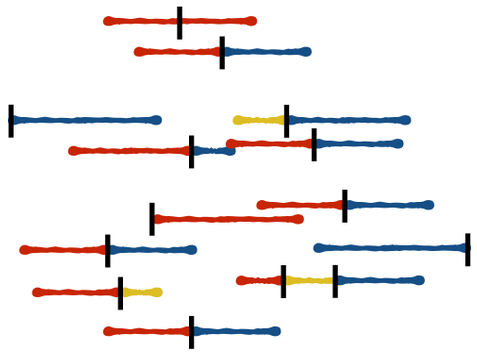


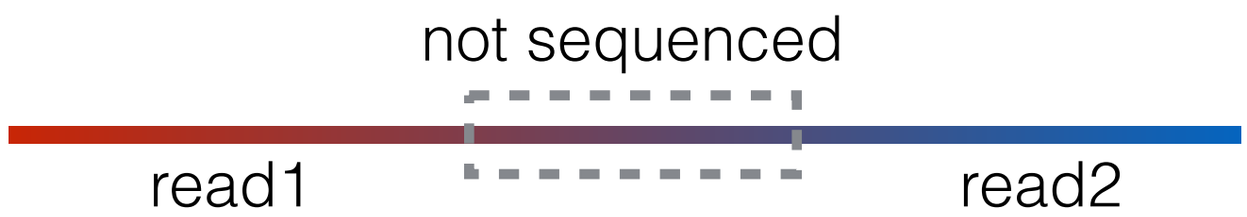
fragment based mapping



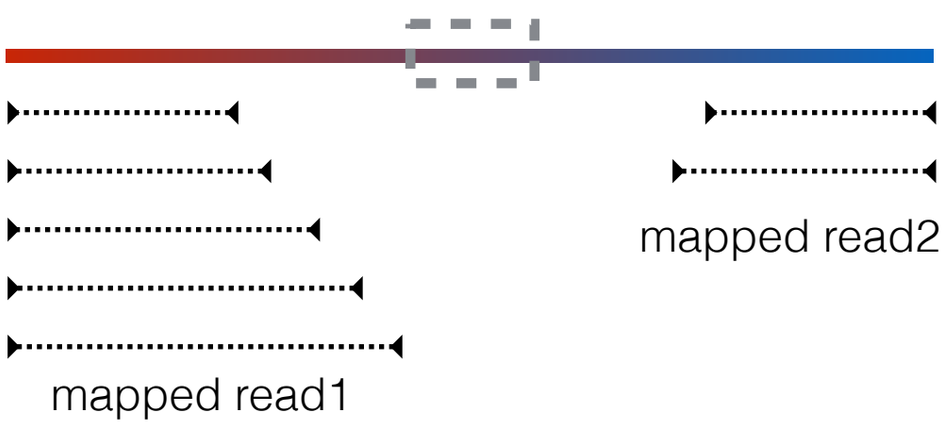
Mapping @TADbit

Serra et al. (2017). PLOS CompBio

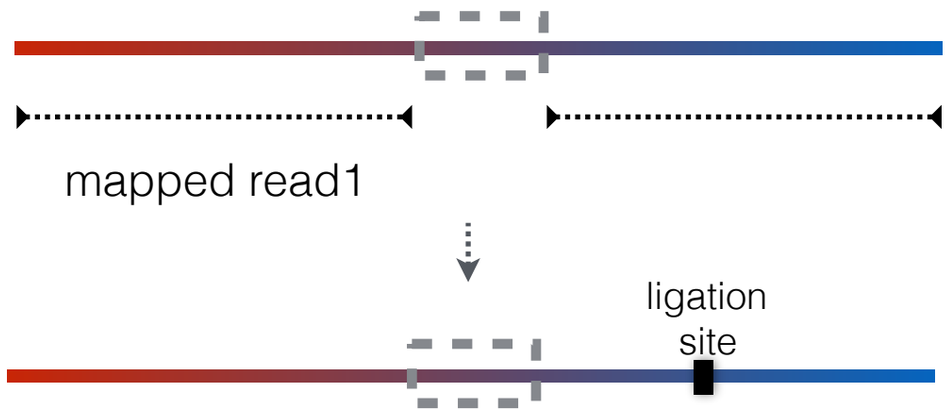




Iterative mapping

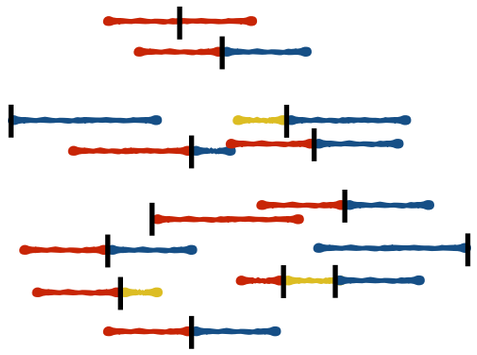


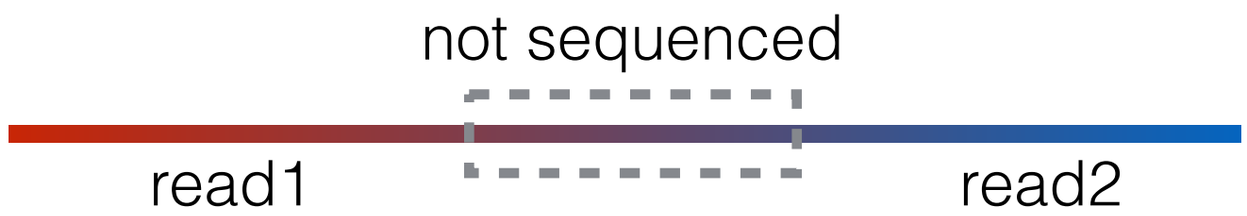
fragment based mapping



Mapping @TADbit

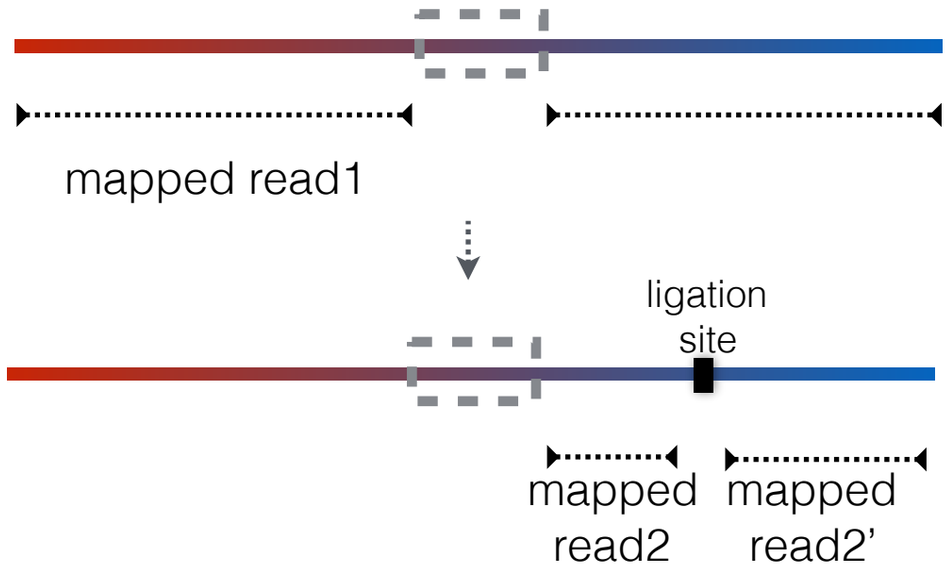
Serra et al. (2017). PLOS CompBio





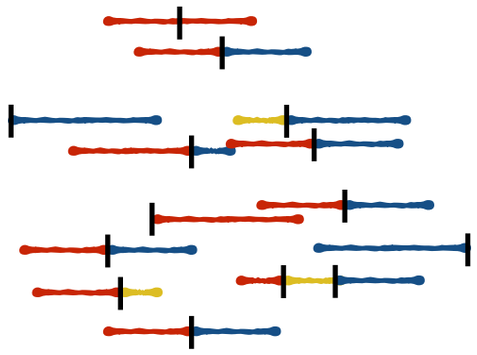
Iterative mapping

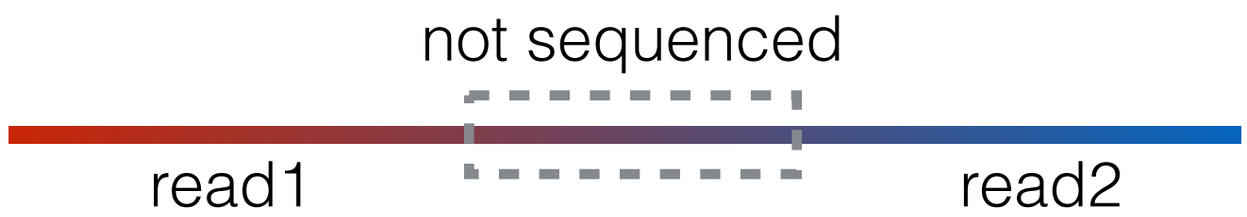
fragment based mapping



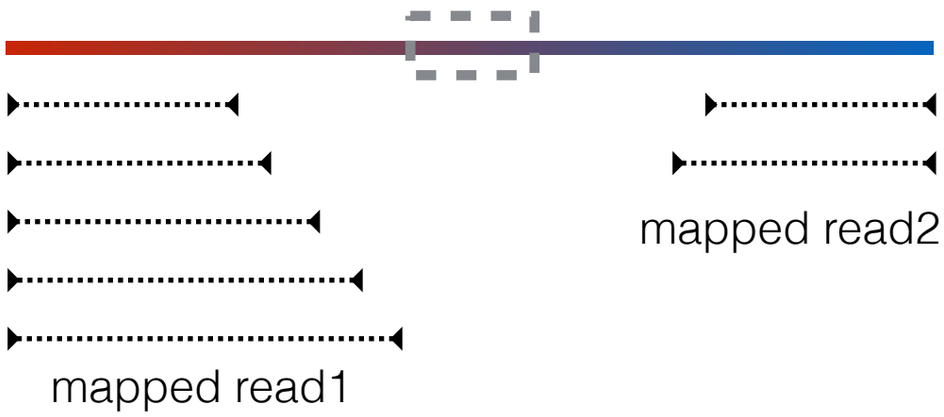
Mapping @TADbit

Serra et al. (2017). PLOS CompBio

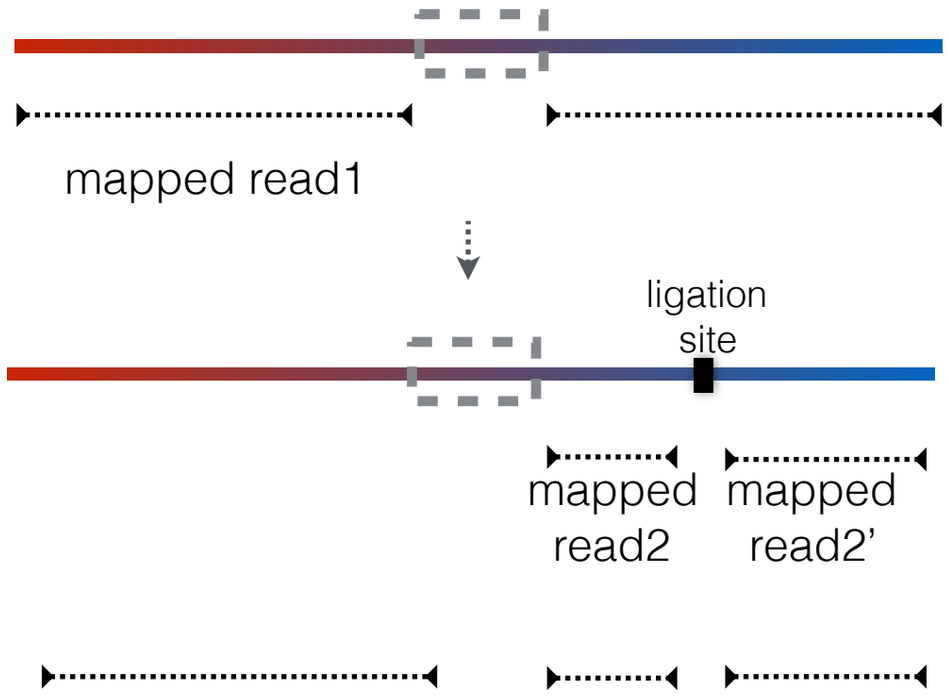




Iterative mapping



fragment based mapping

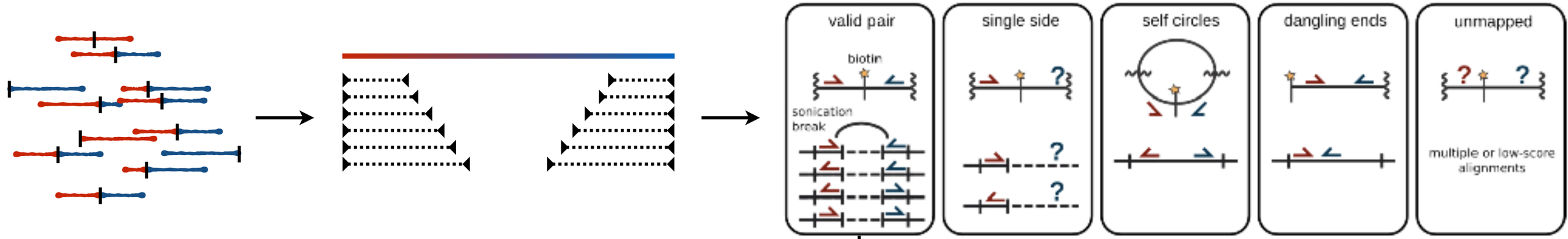


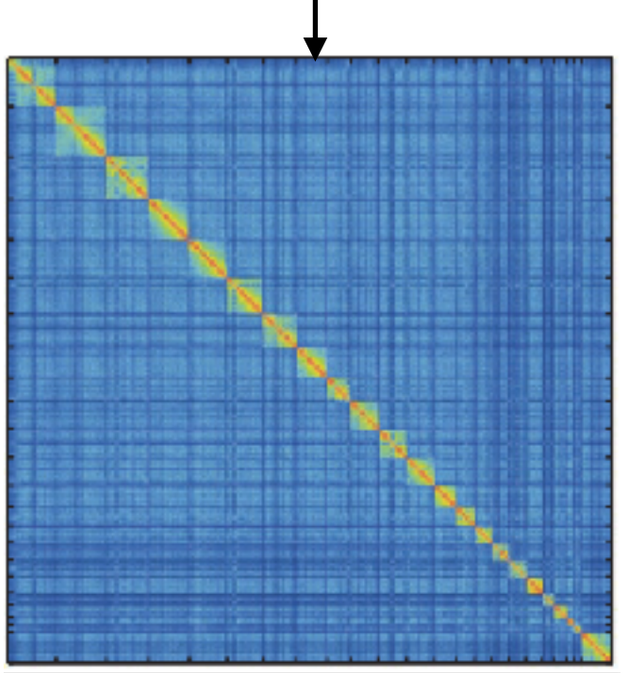
How much you normally map?

- 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)

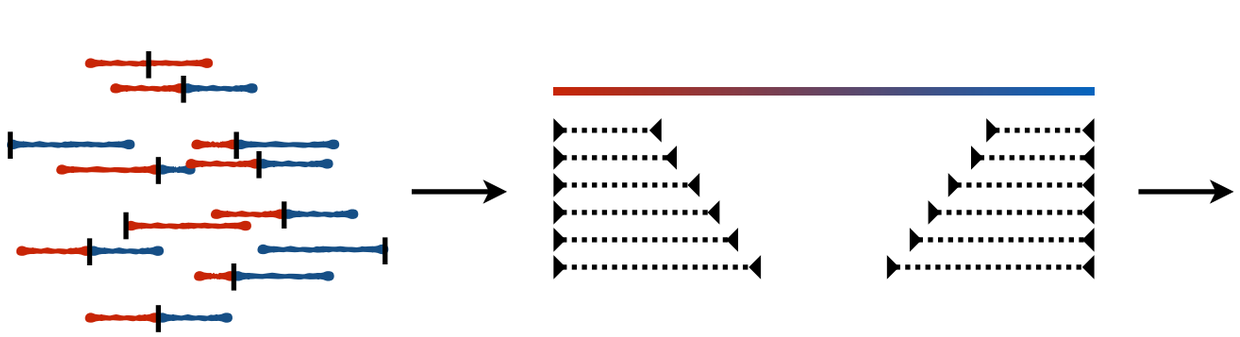

Got mapped
reads?

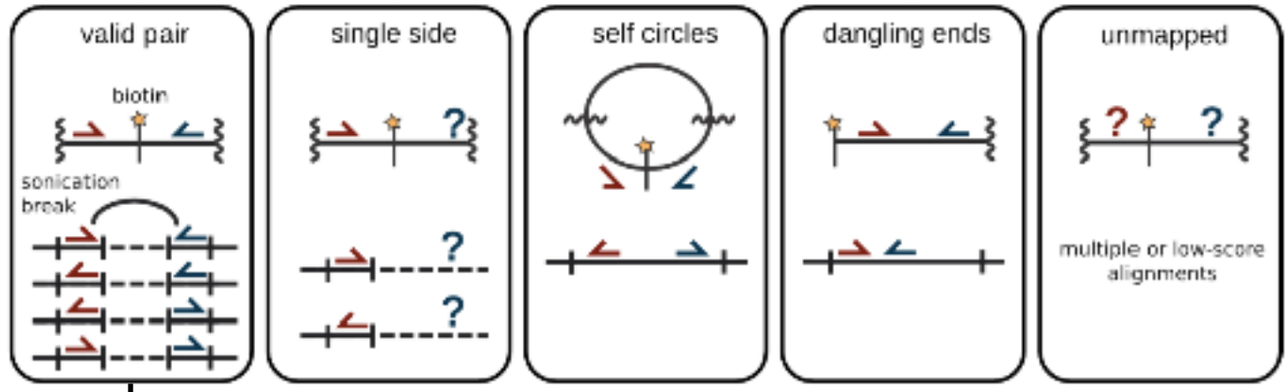
Interaction matrices

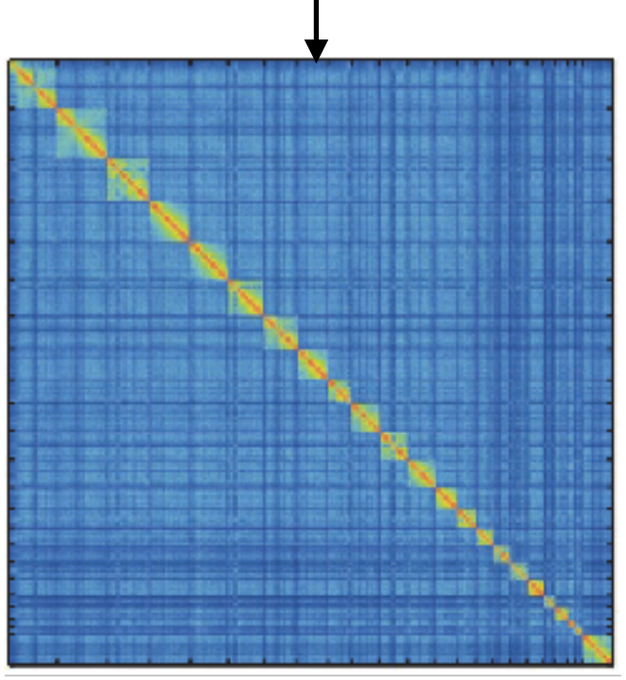


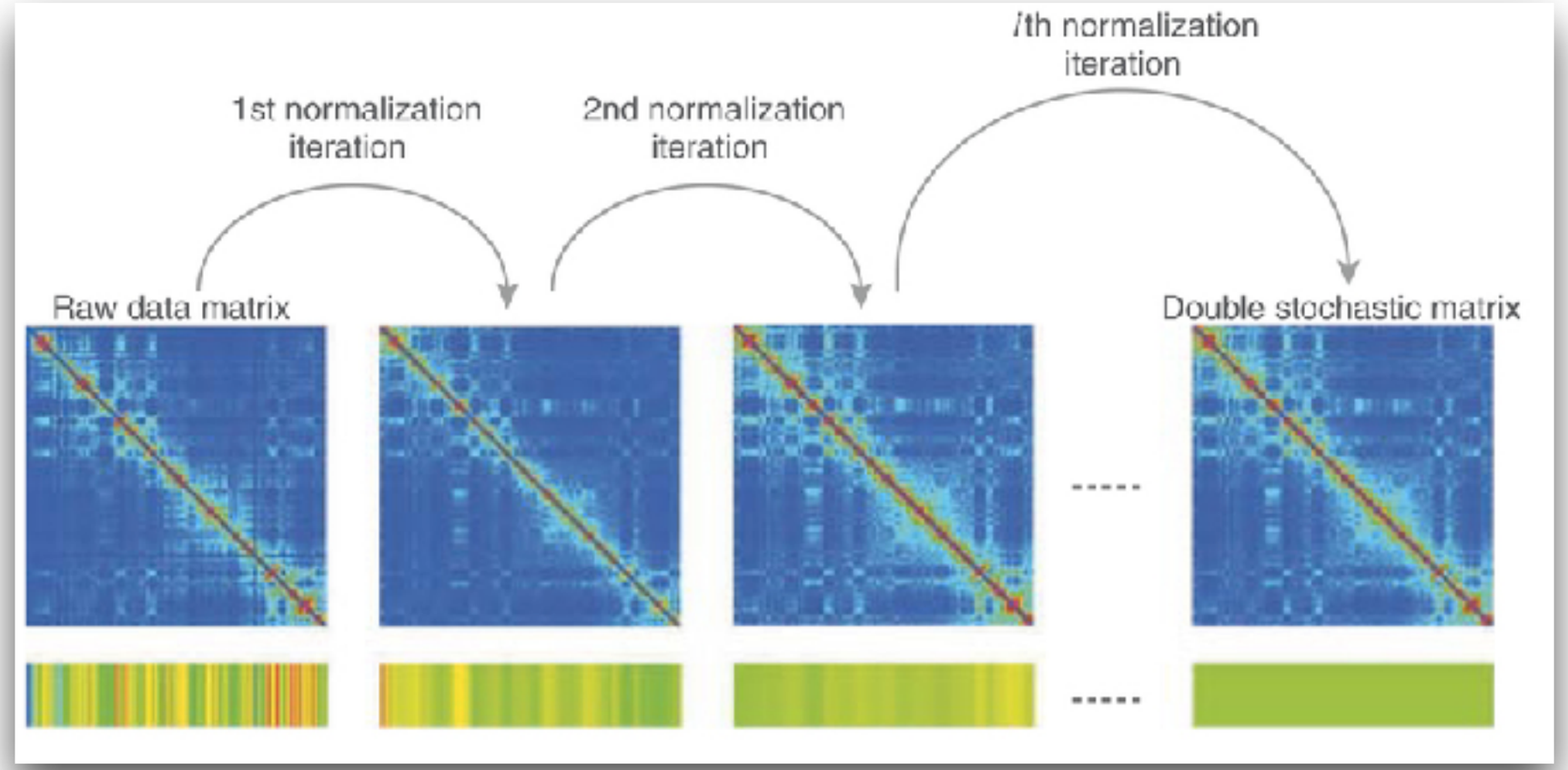


Interaction matrices



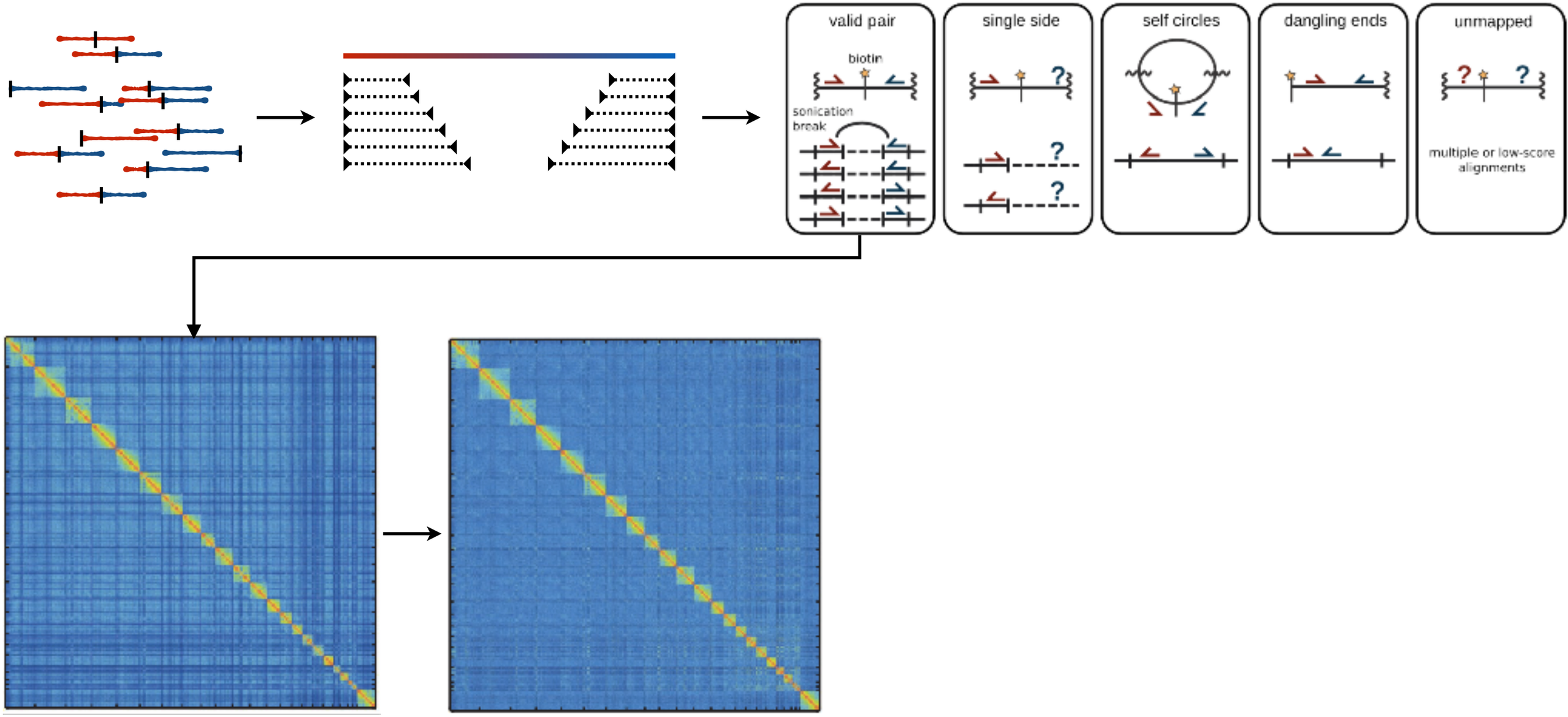




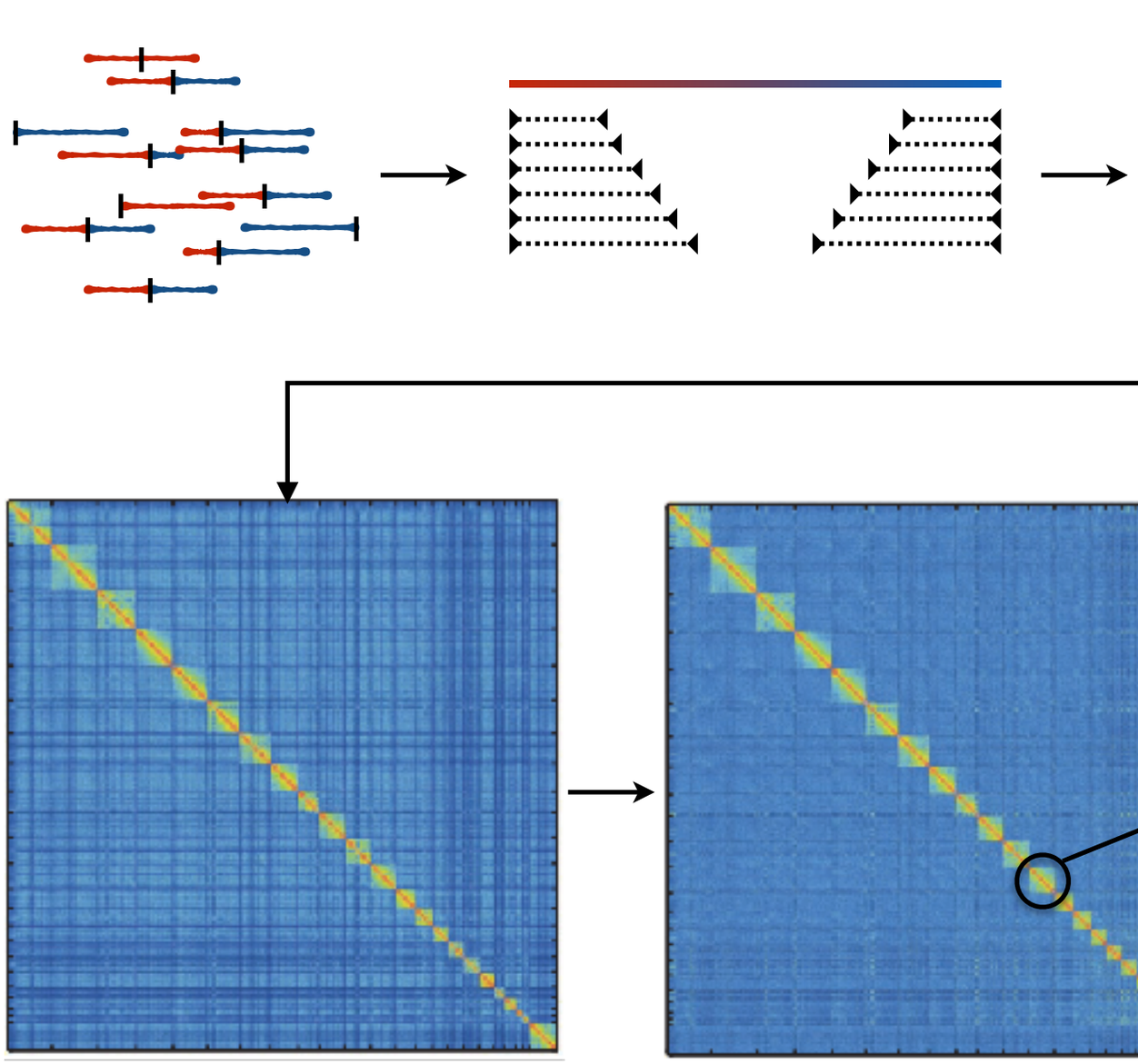


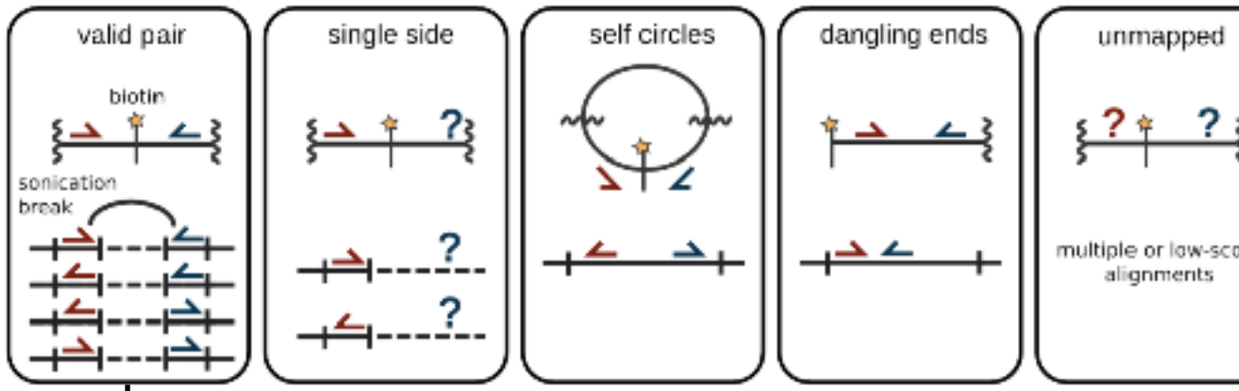
Zooming in on genome organization.
Zhou, X. J., & Alber, F. Nature Methods (2012)

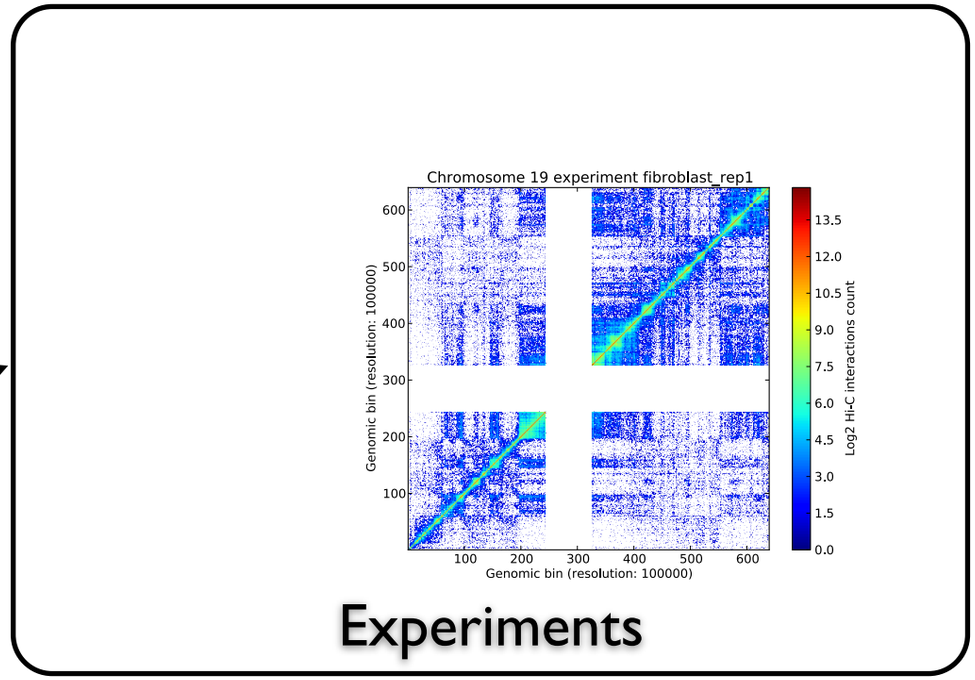
Interaction matrices



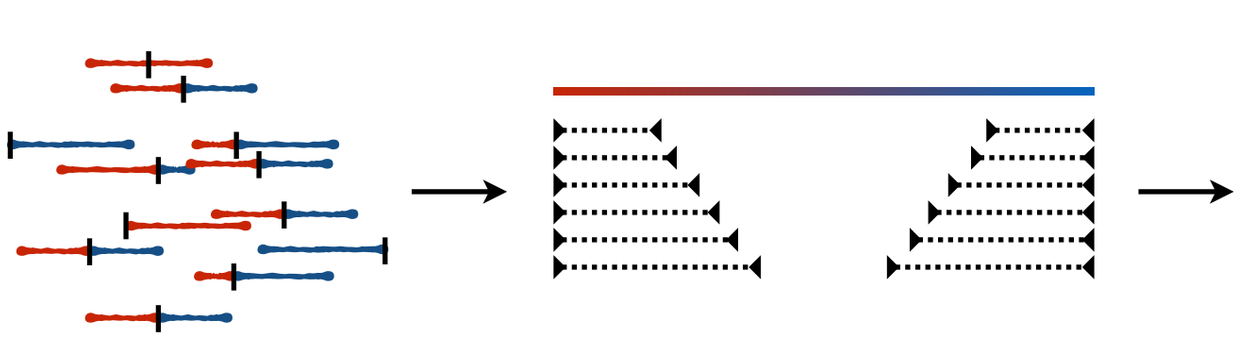
Interaction matrices

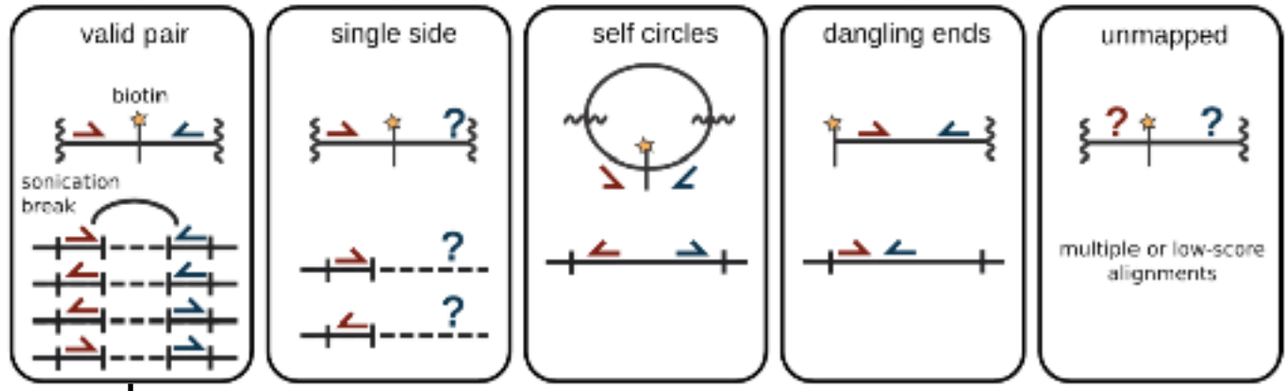


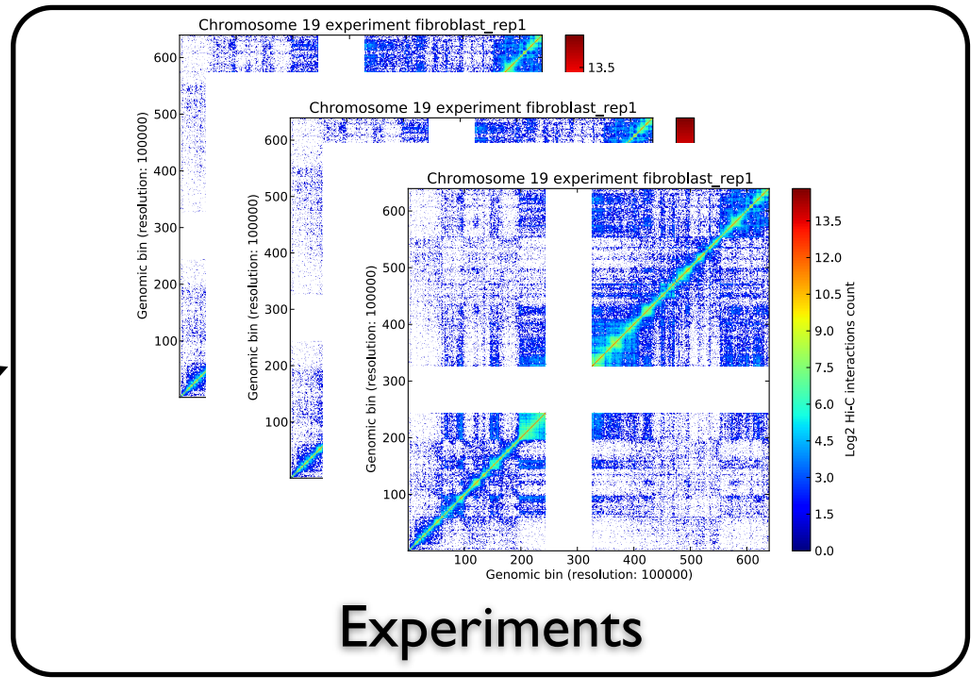




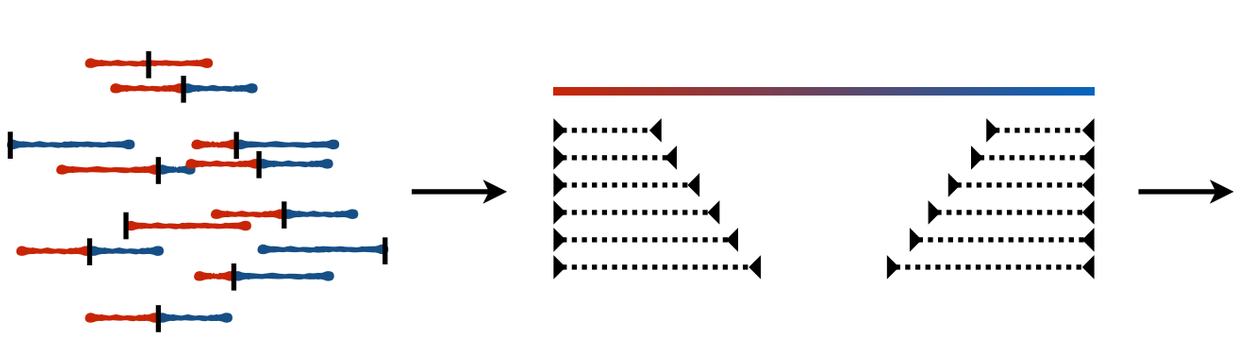
Interaction matrices

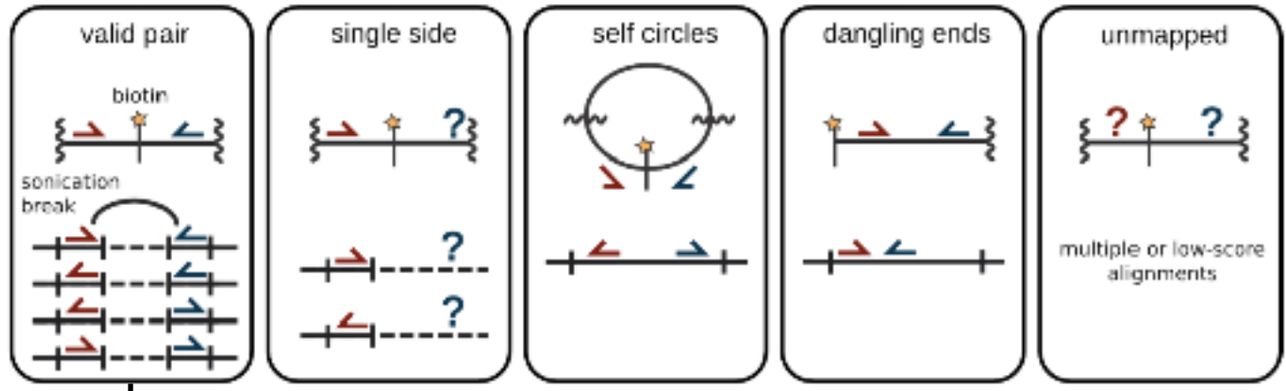


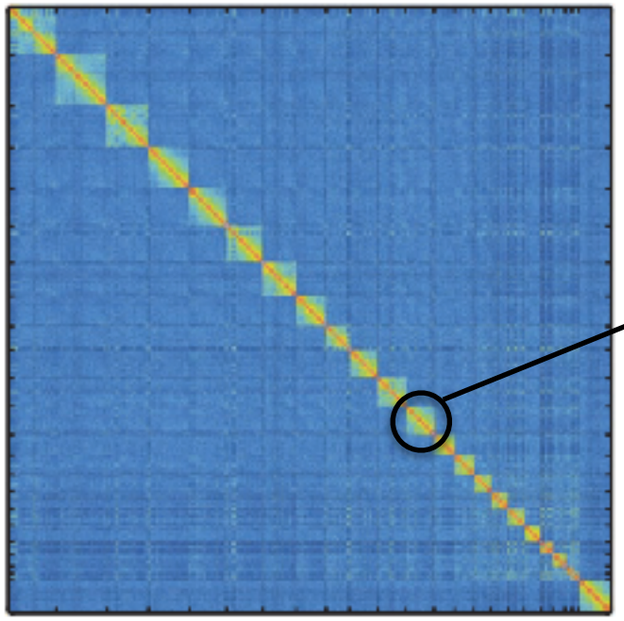


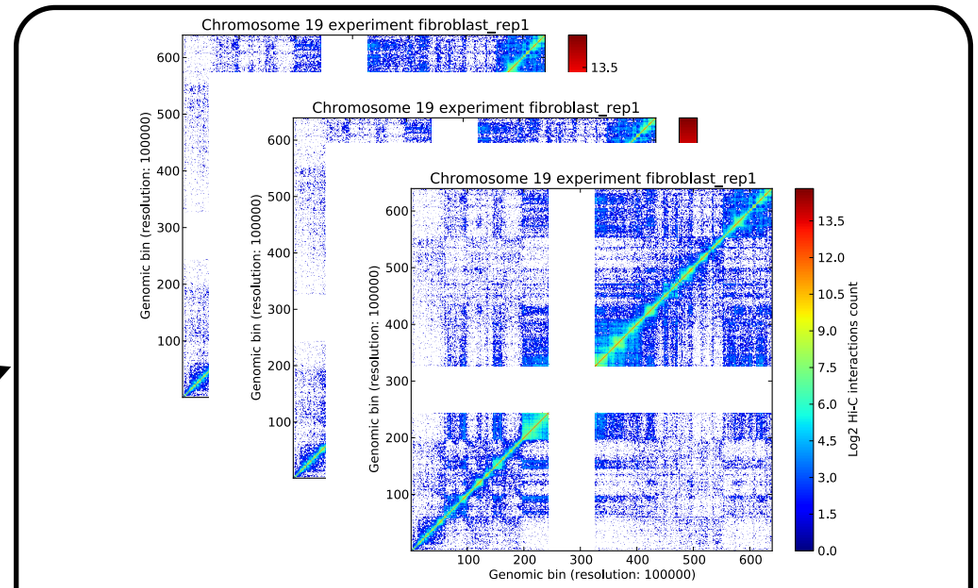


Interaction matrices









3D modeling

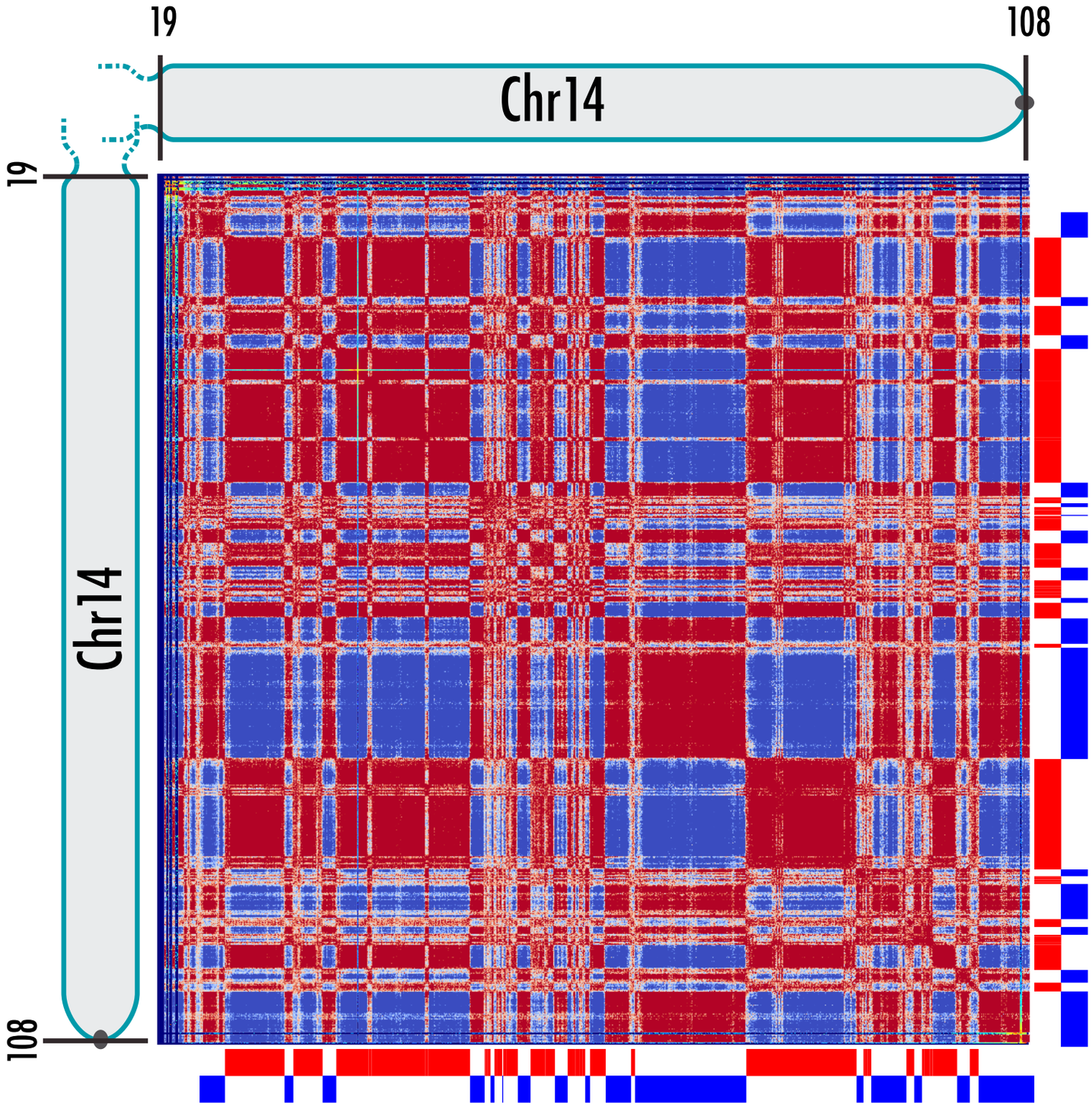
Matrix analysos

Structural Models

TADs & A/B

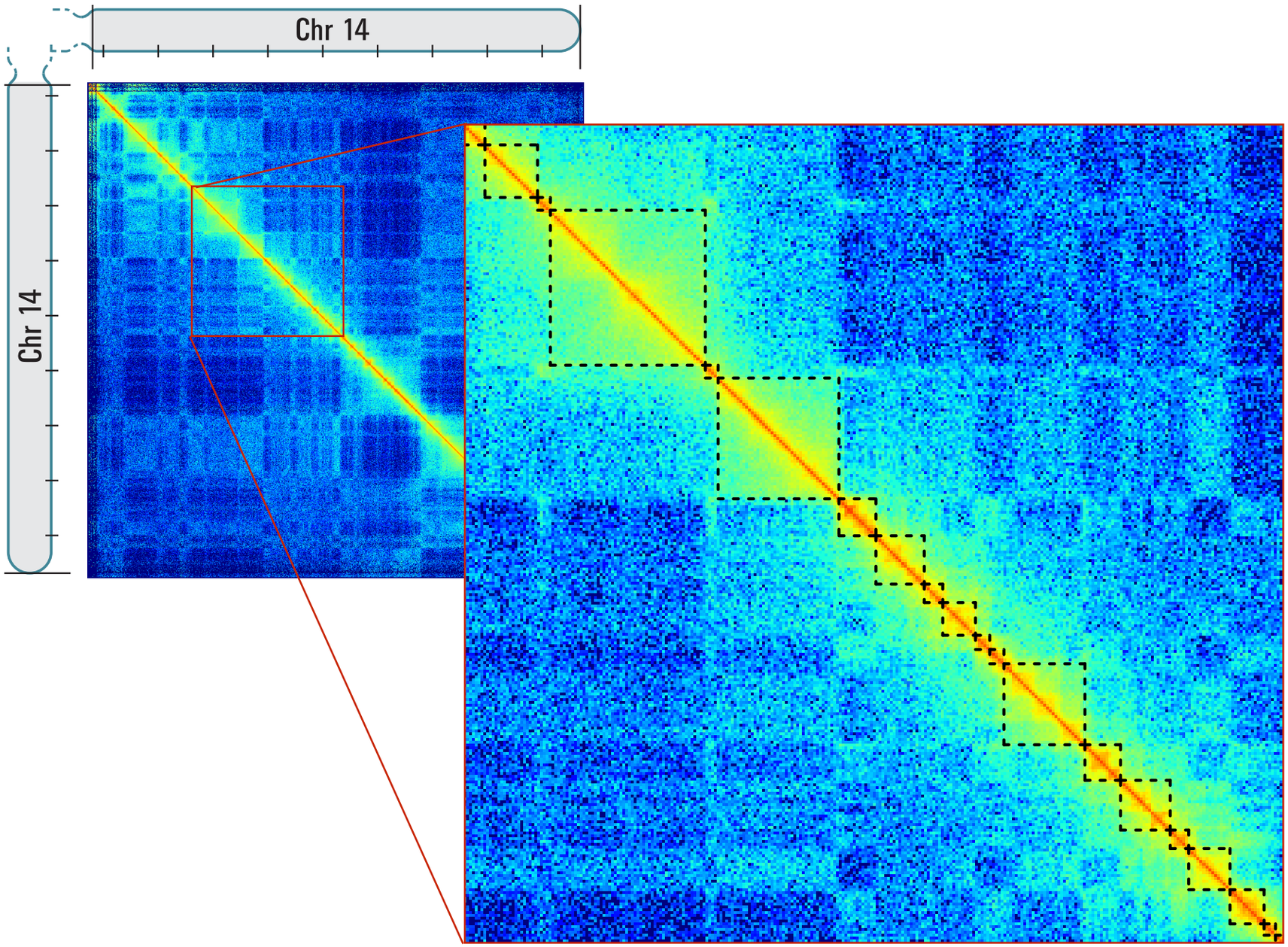
A/B Compartment

Chromosome 14



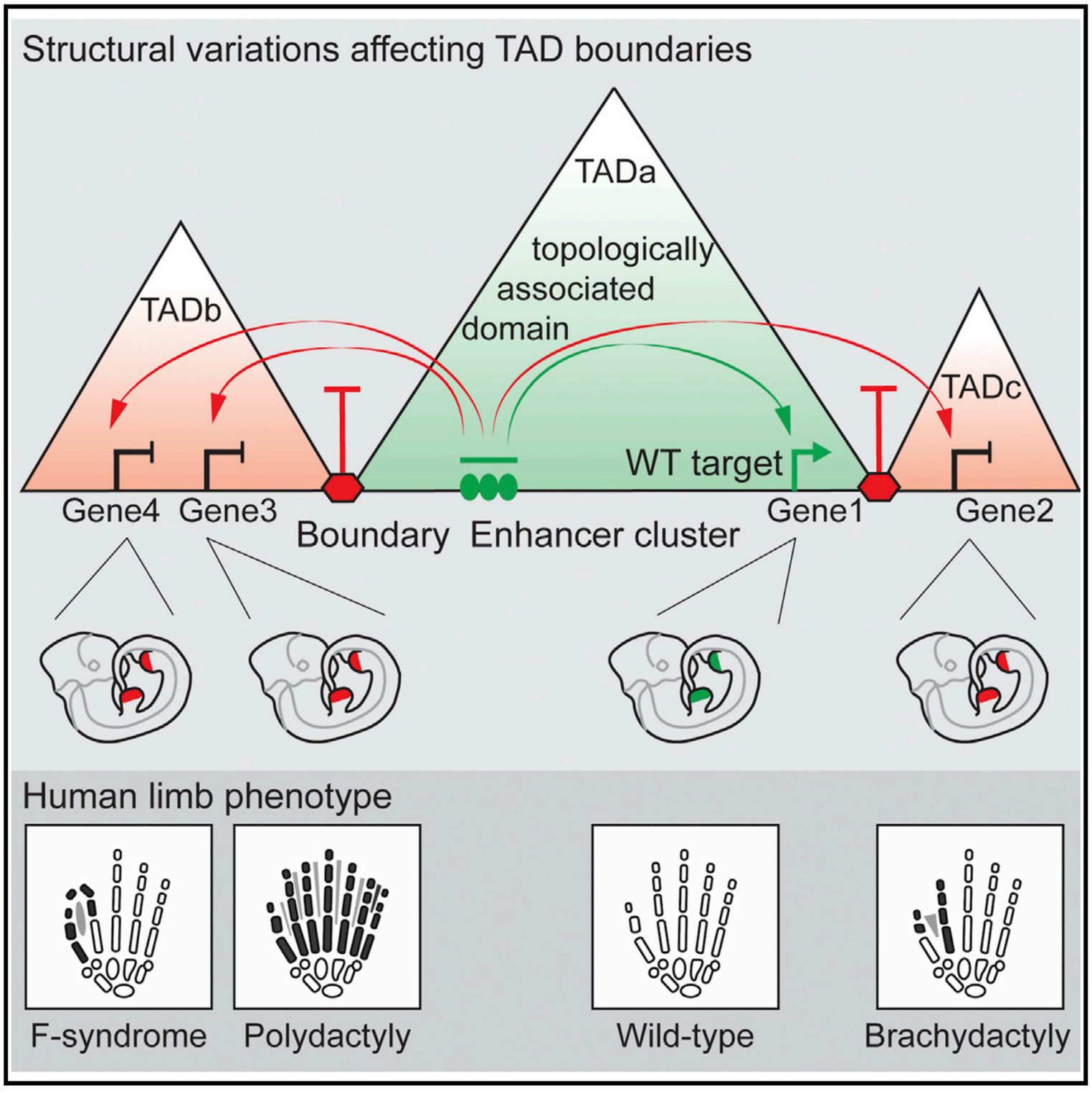
TADs

Chromosome 14



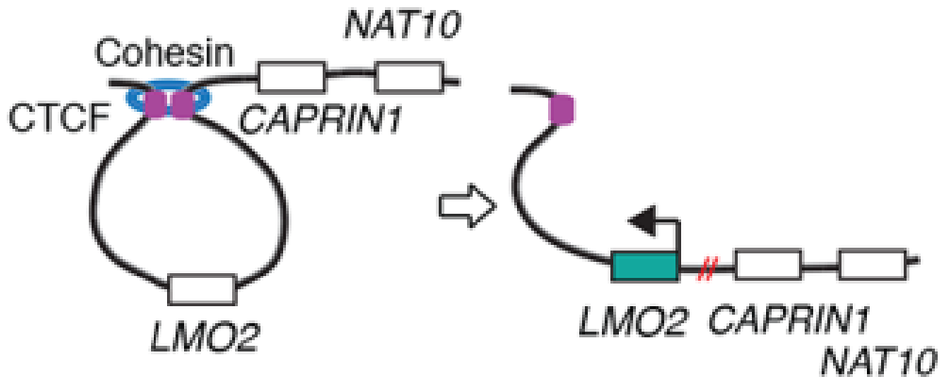
TADs are functional units

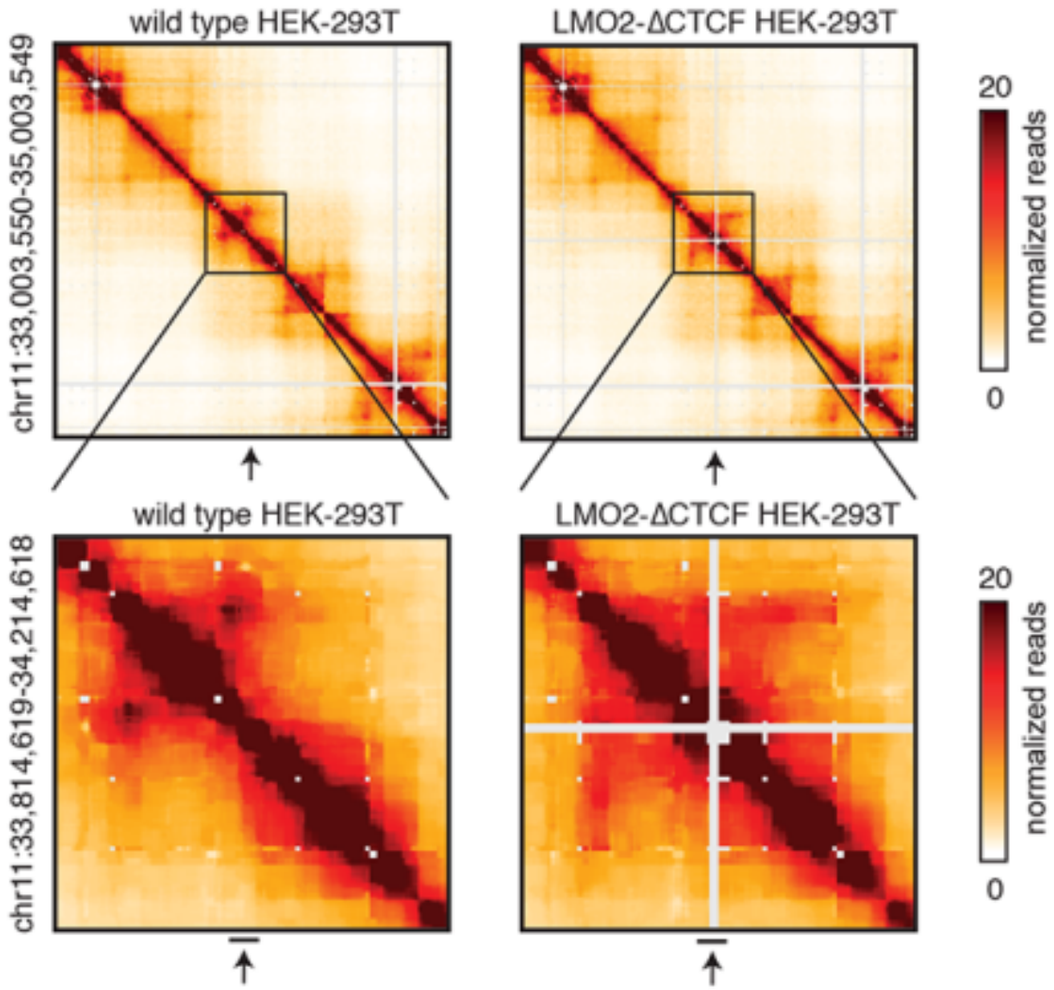
Lupiáñez, et al. (2015). Cell, 1–15.



TADs are functional units

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





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
HTC [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).

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

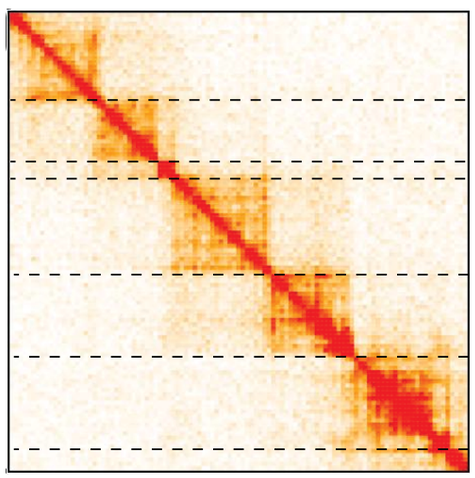
* These methods are publicly available.

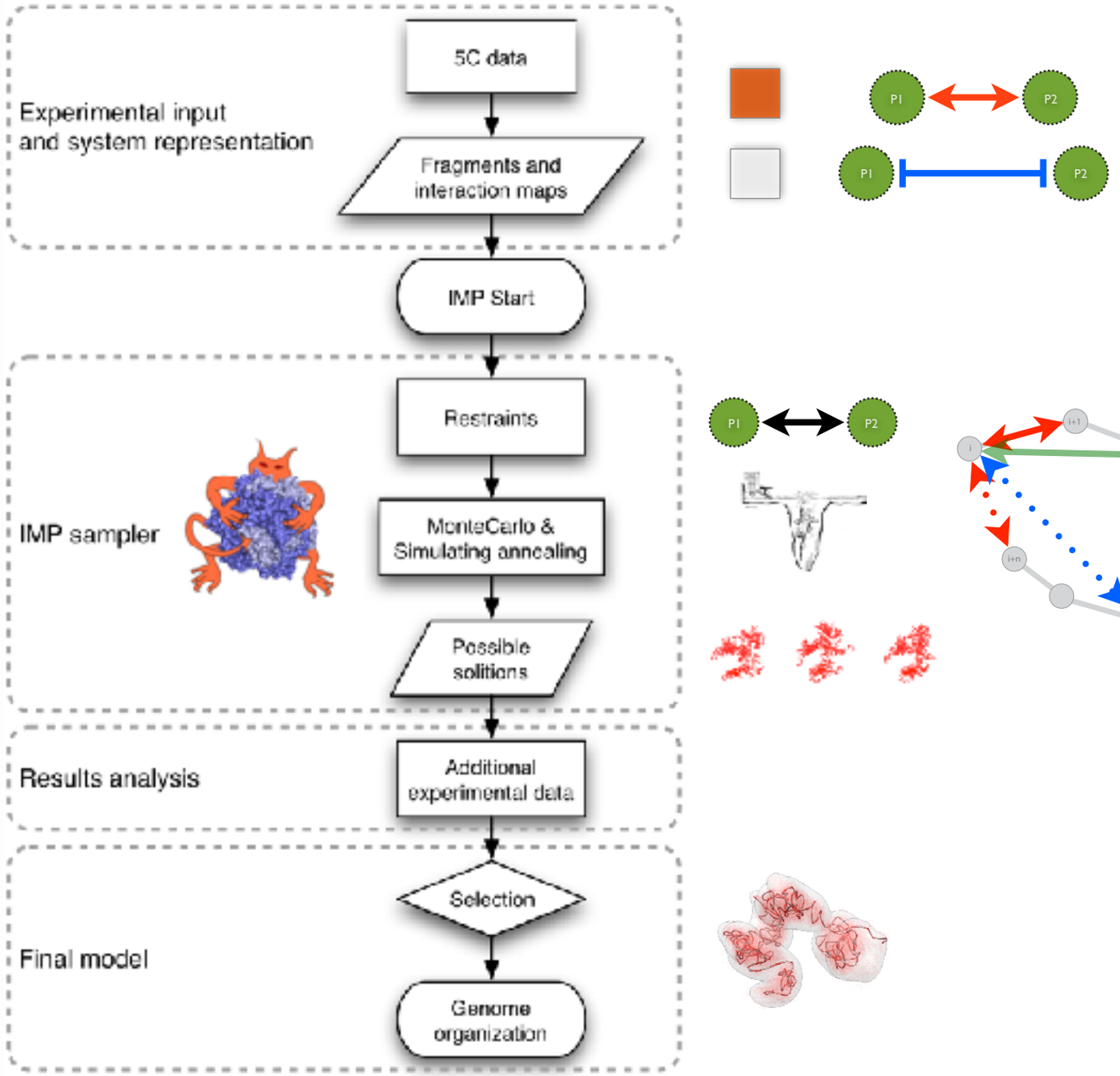
Restraint-based three-dimensional modeling of genomes and genomic domains.
Serra F, Di Stefano M, Spill YG, Cuartero Y, Goodstadt M, Baù D, Marti-Renom MA. FEBS Lett 589: 2987–2995 (2015)


Got normalized
Hi-C maps?

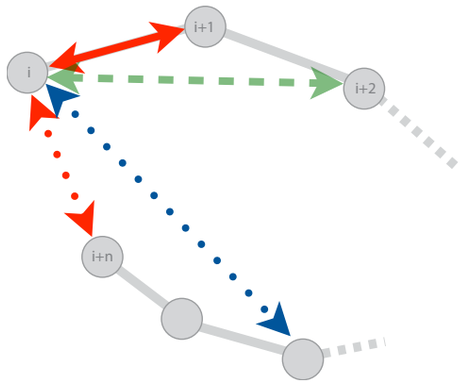
TADbit

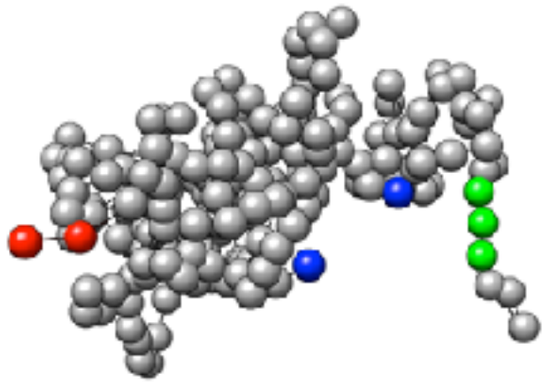
http://3DGenomes.org
http://www.integrativemodeling.org





Model representation and scoring





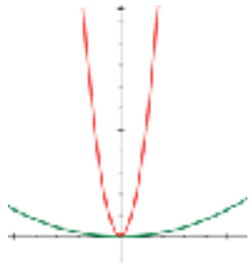
$d = d_0$

$d < d_0$

$d > d_0$

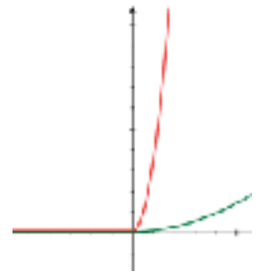
Harmonic

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



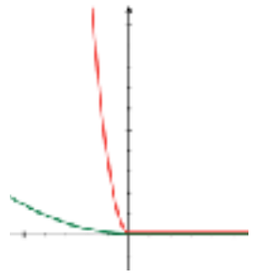
Harmonic Upper Bound

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

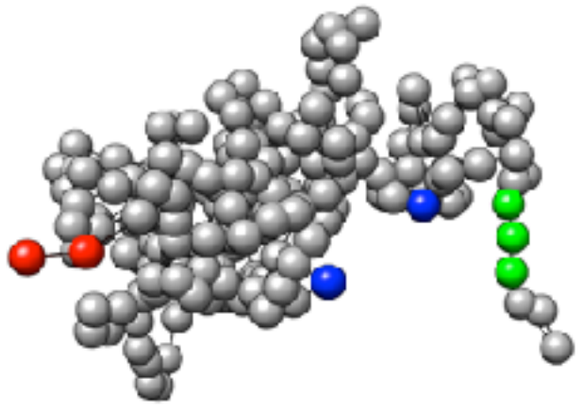


Harmonic Lower Bound

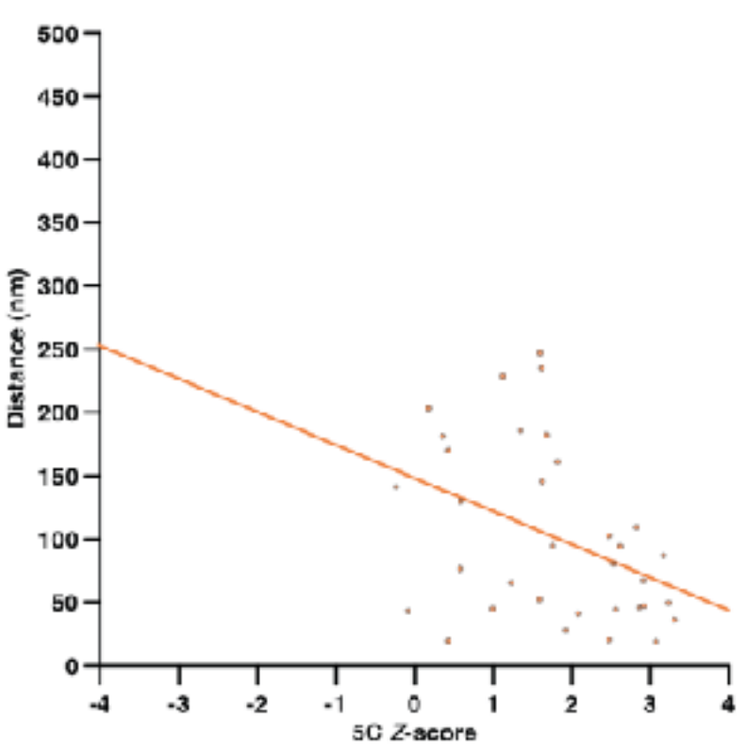
$$\begin{cases} if\ \ d_{i,j} \leq d^0_{i,j}; \quad lbH_{i,j} = k\left(d_{i,j} - d^0_{i,j}\right)^2 \\ if\ \ d_{i,j} > d^0_{i,j}; \quad lbH_{i,j} = 0 \end{cases}$$



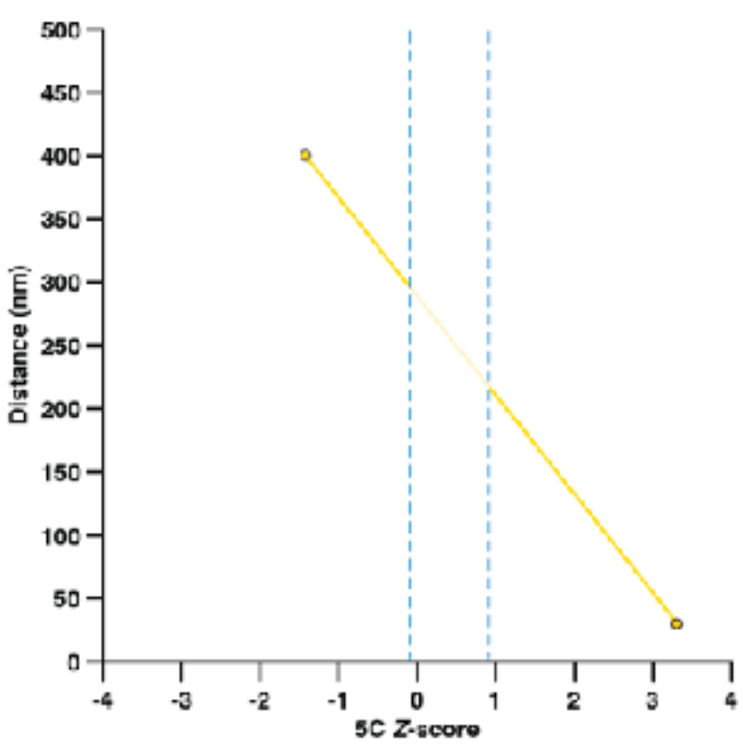
From 3C data to spatial distances



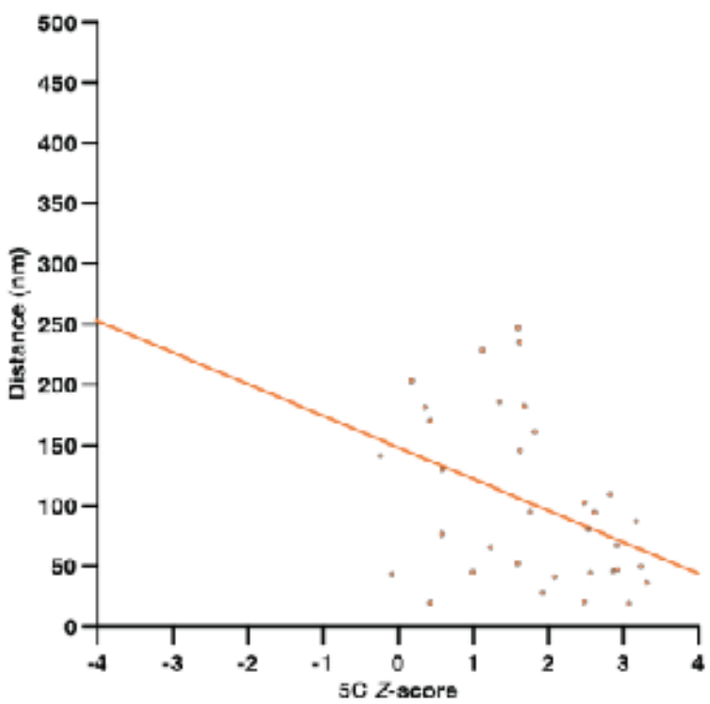
Neighbor fragments

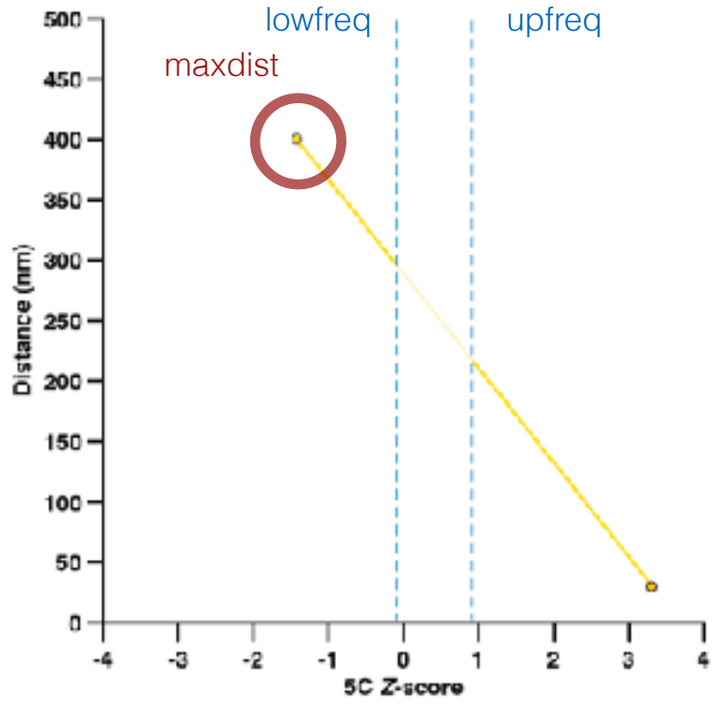


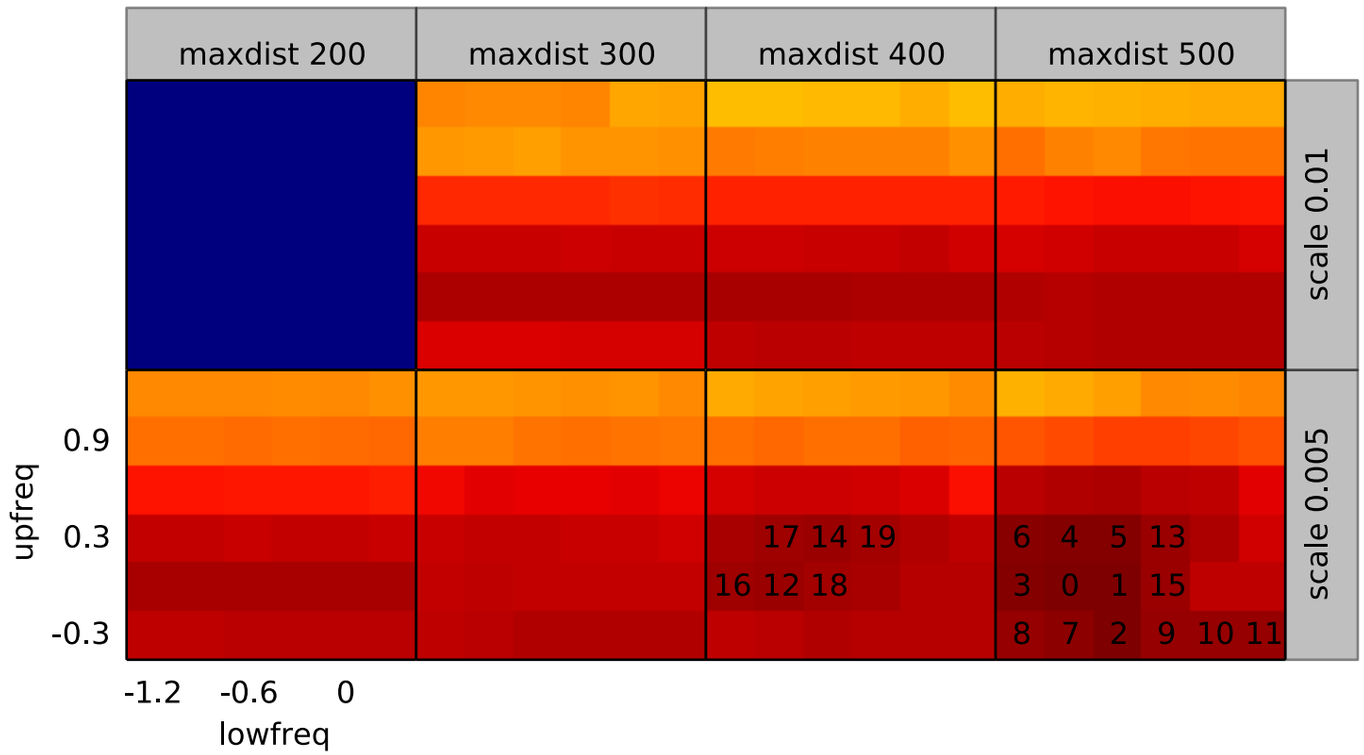
Non-Neighbor fragments



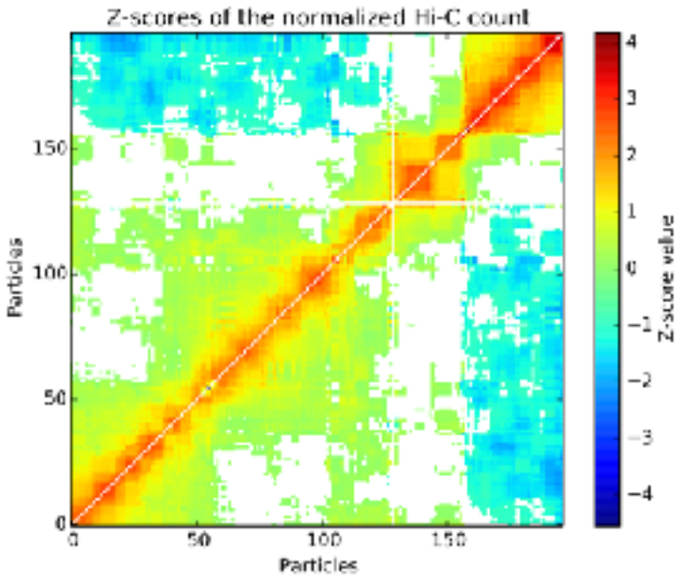
Parameter optimization

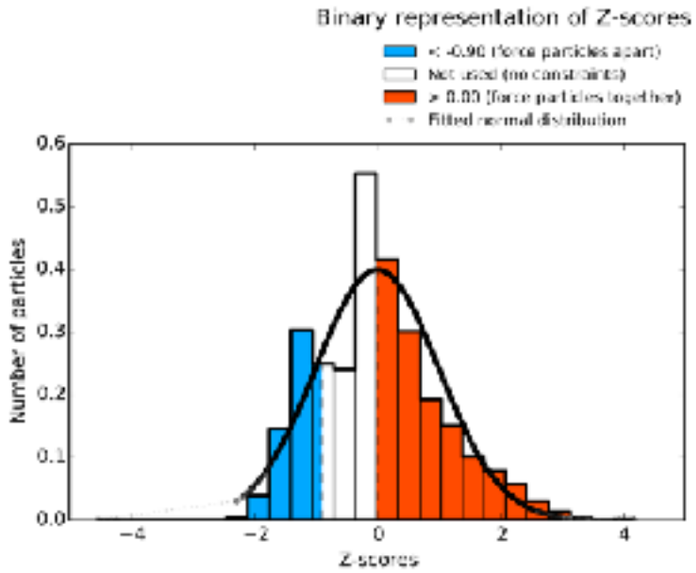


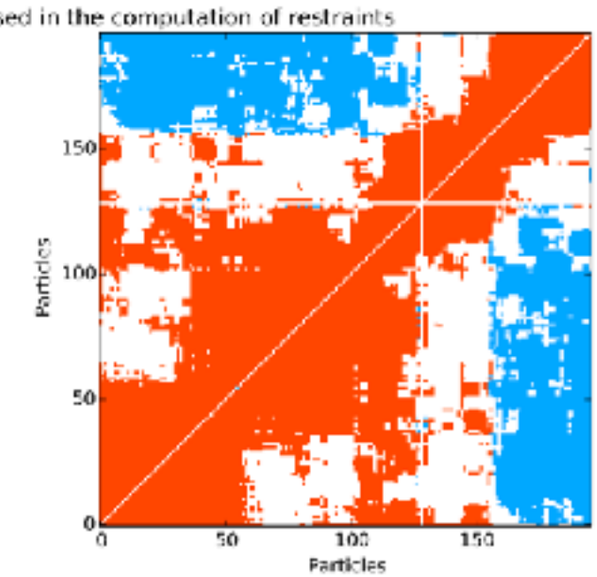


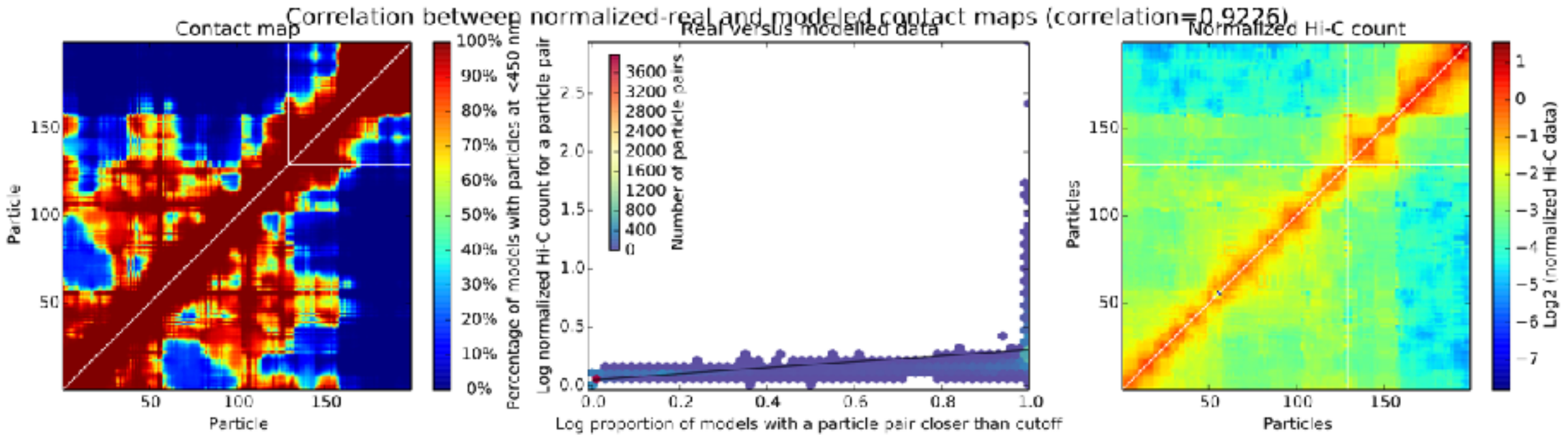


Parameter optimization

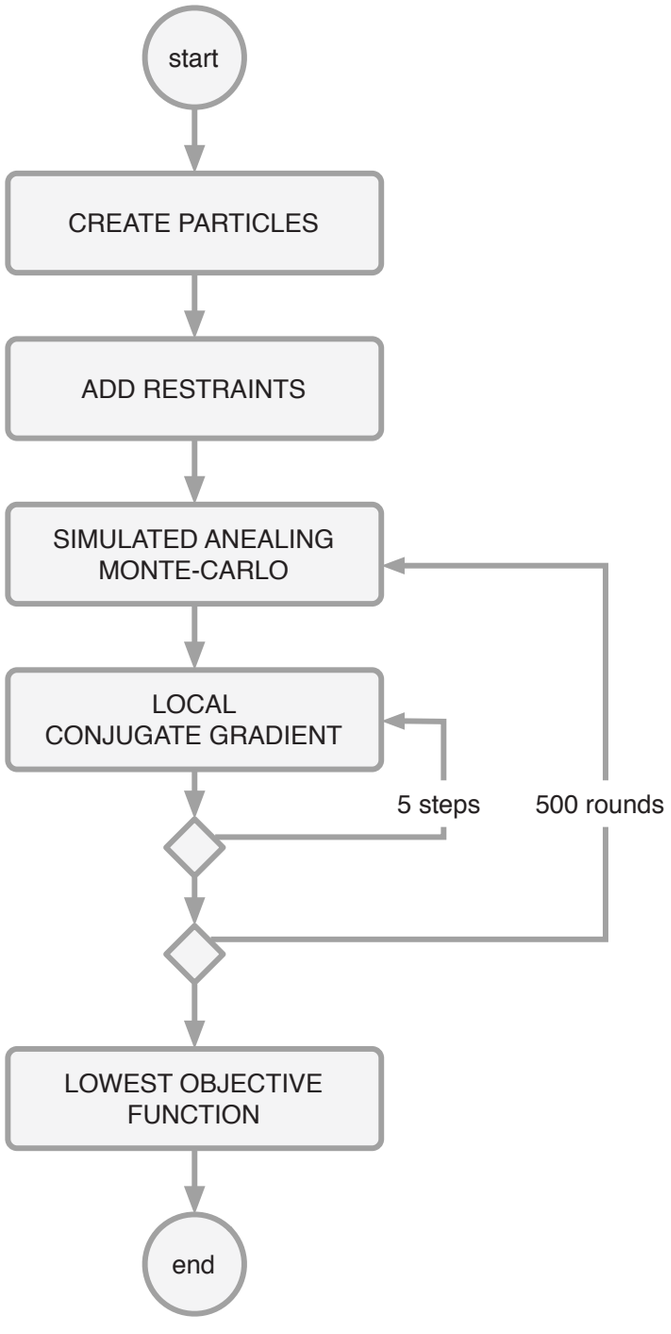


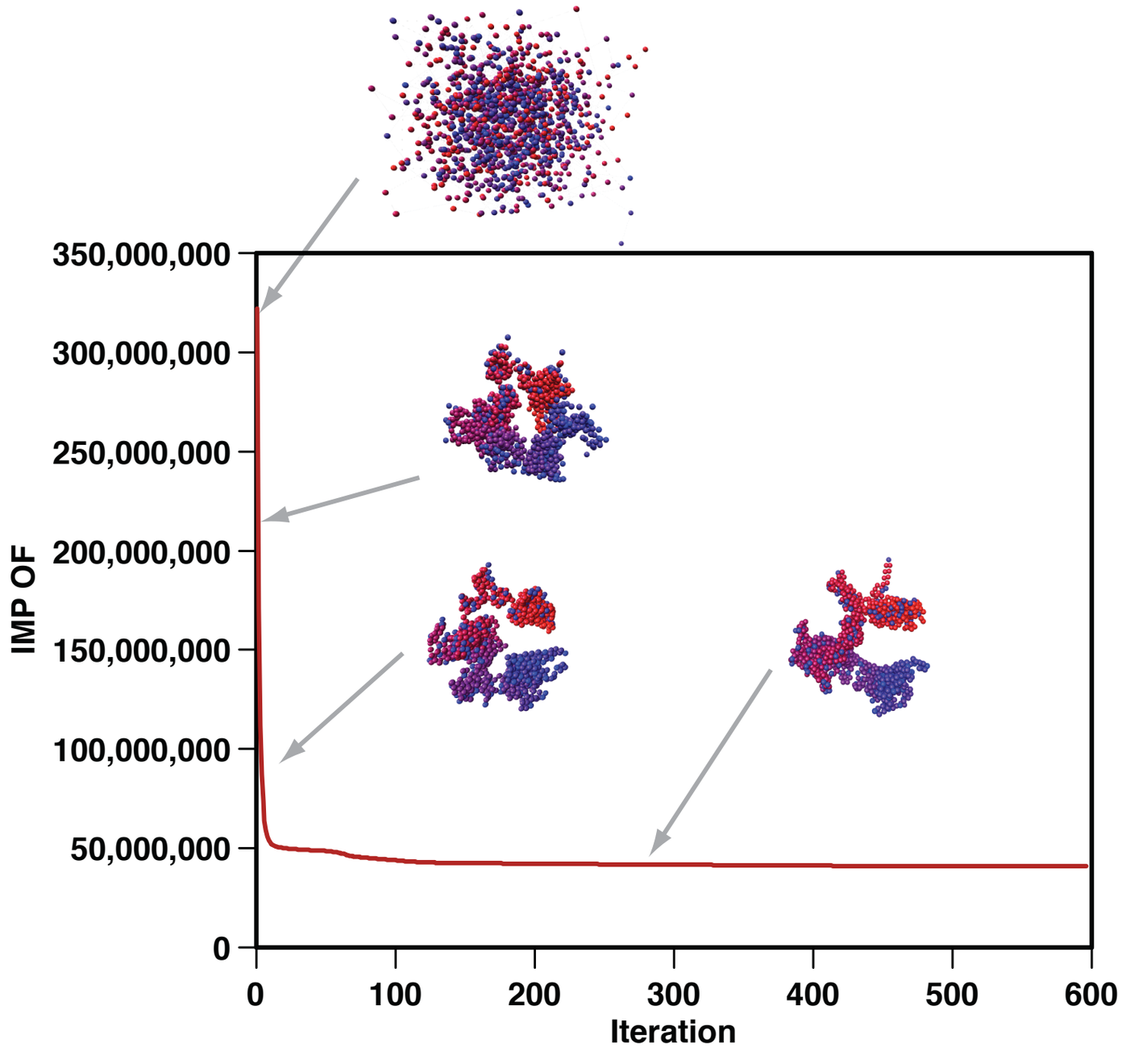




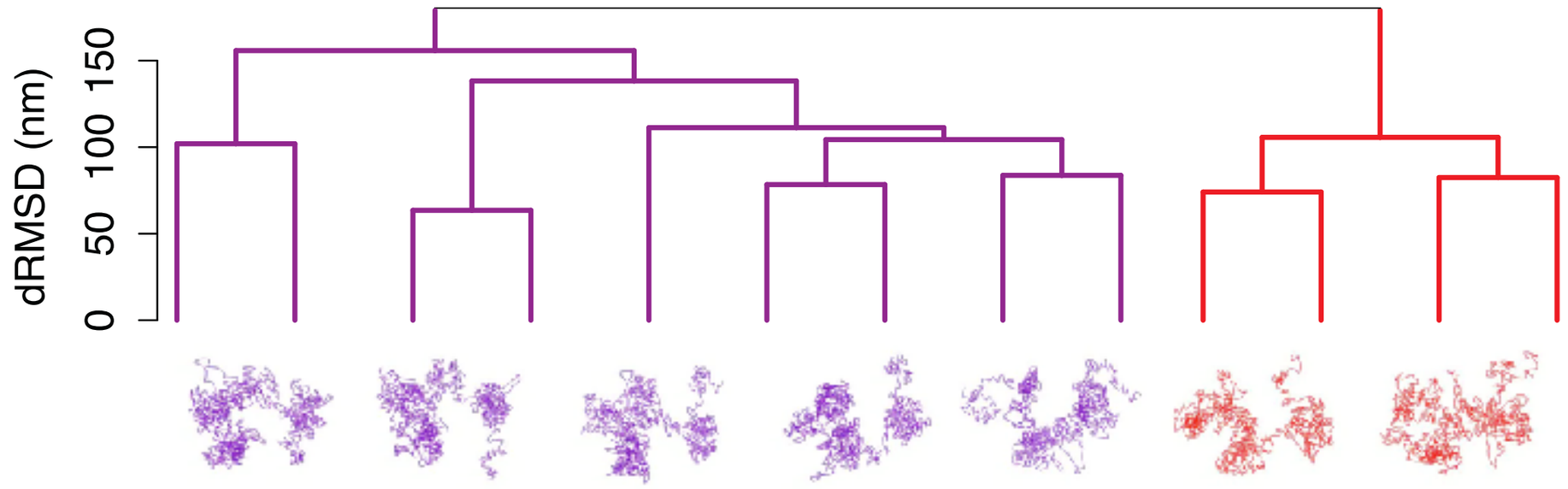


Optimization of the scoring function

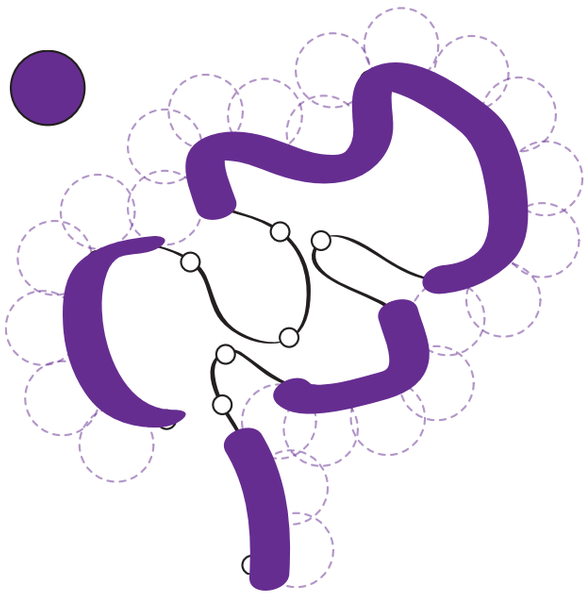




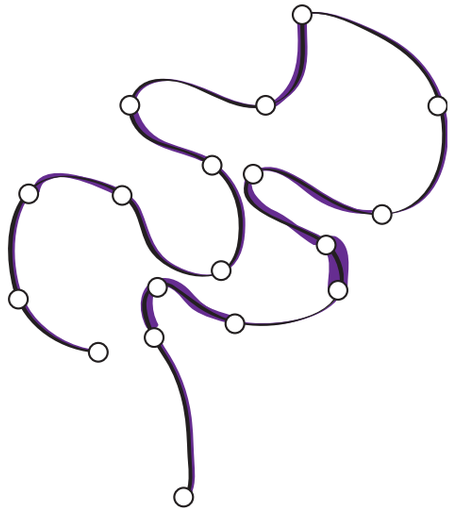
Model analysis: clustering and structural features



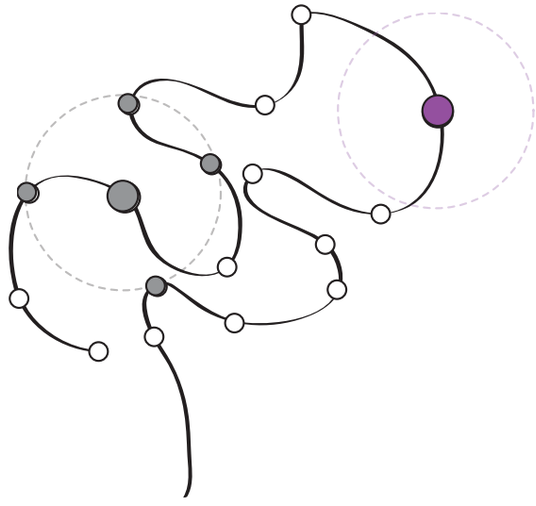
Accessibility (%)



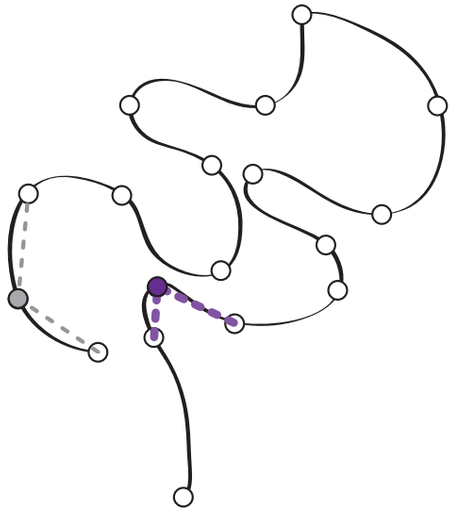
Density (bp/nm)

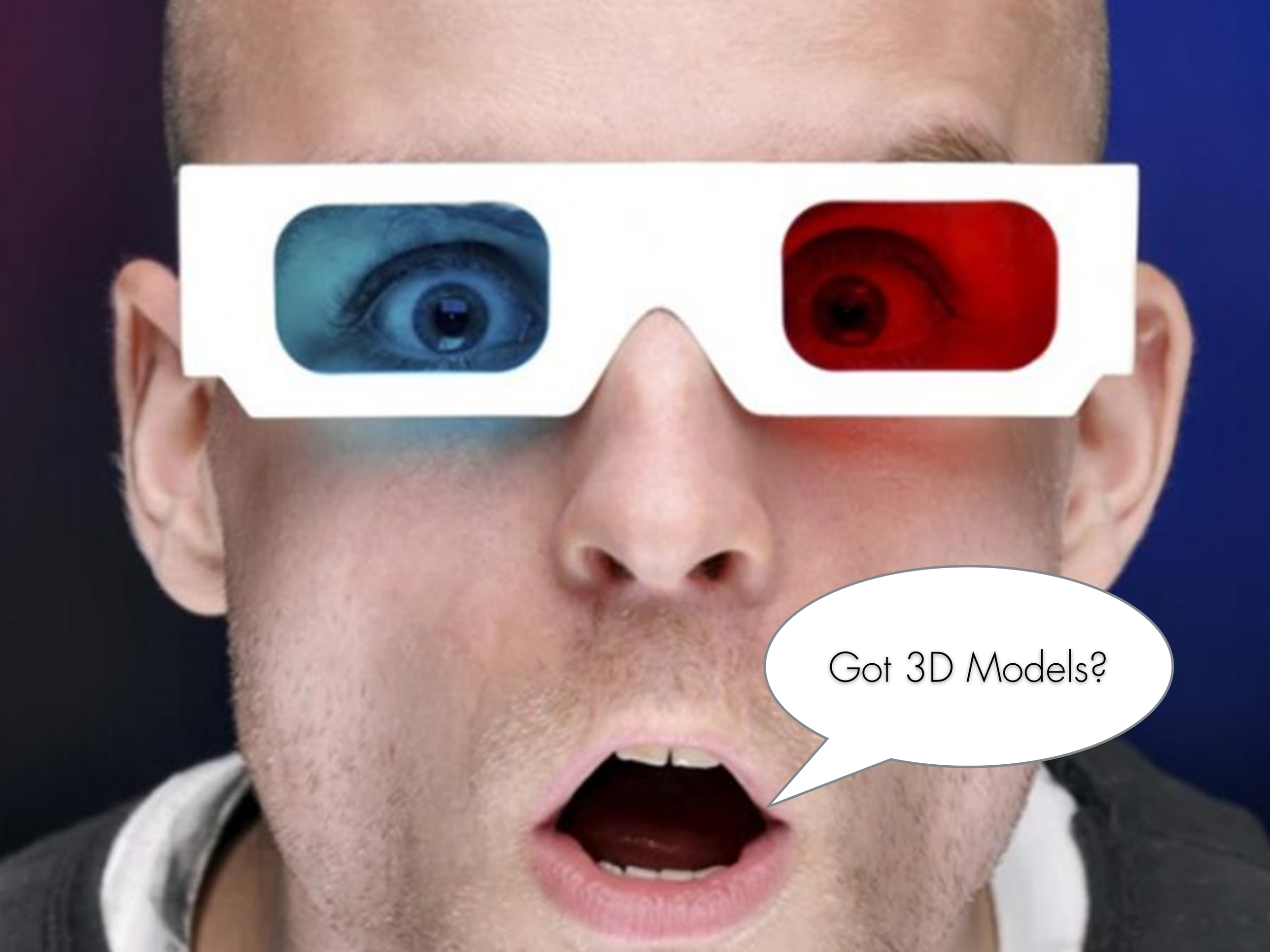


Interactions



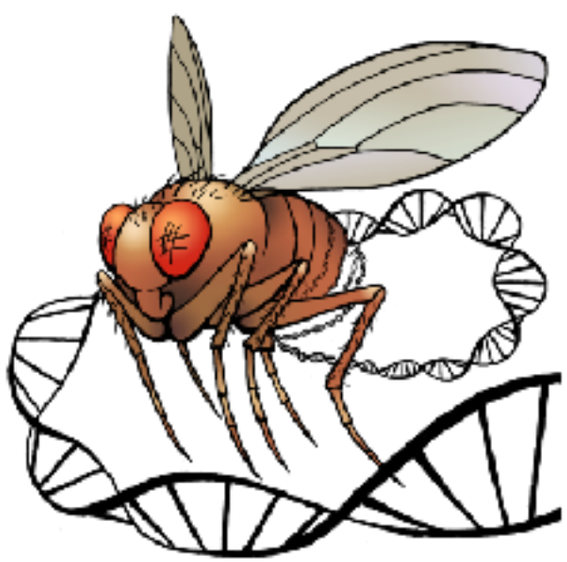
Angle




Got 3D Models?

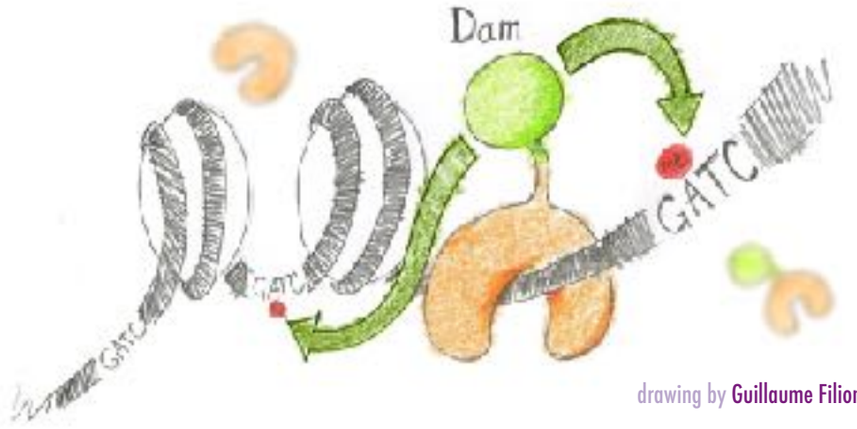
Structuring the COLORs of chromatin

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

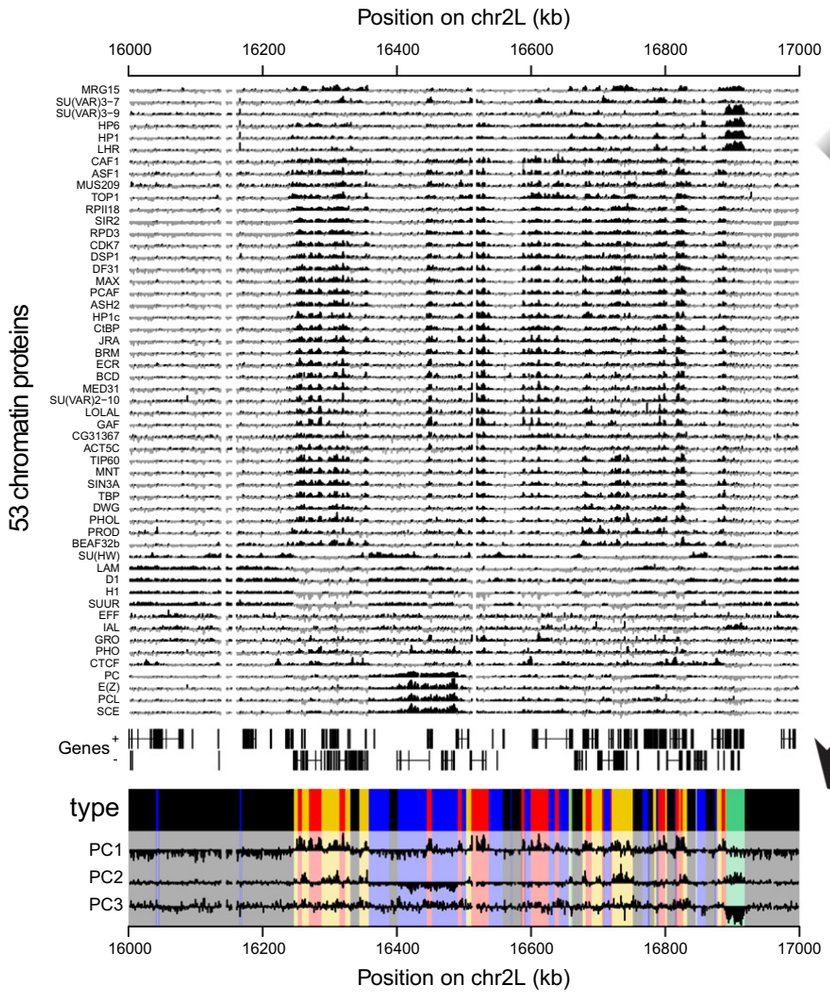


Fly Chromatin COLORs

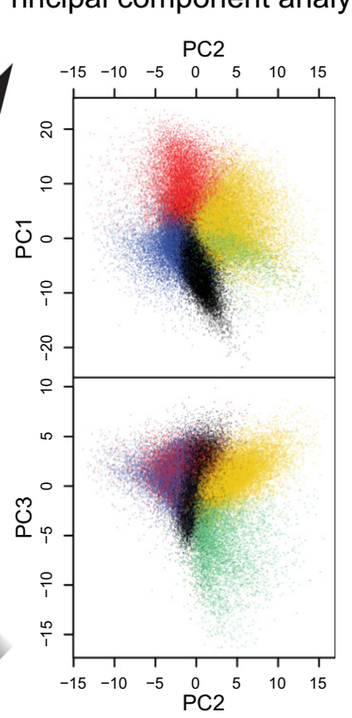
Filion et al. (2010). Cell, 143(2), 212–224.



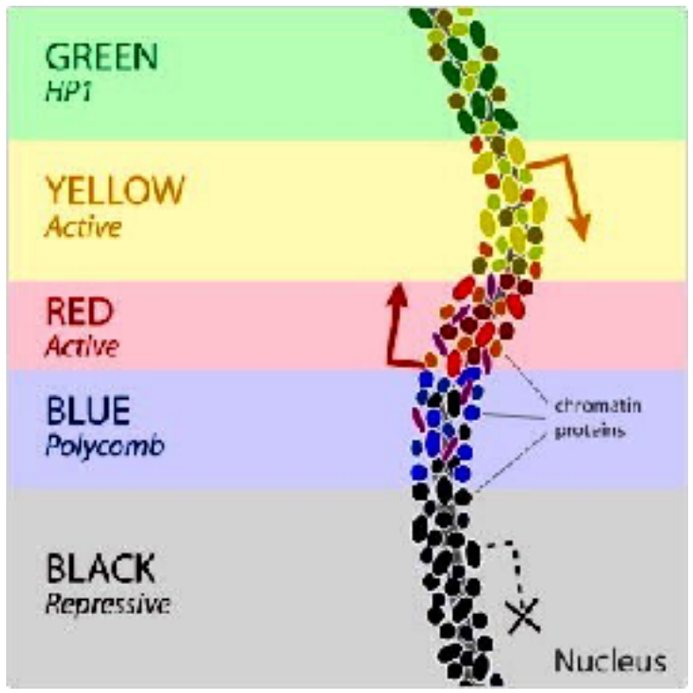
drawing by Guillaume Filion



Principal component analysis

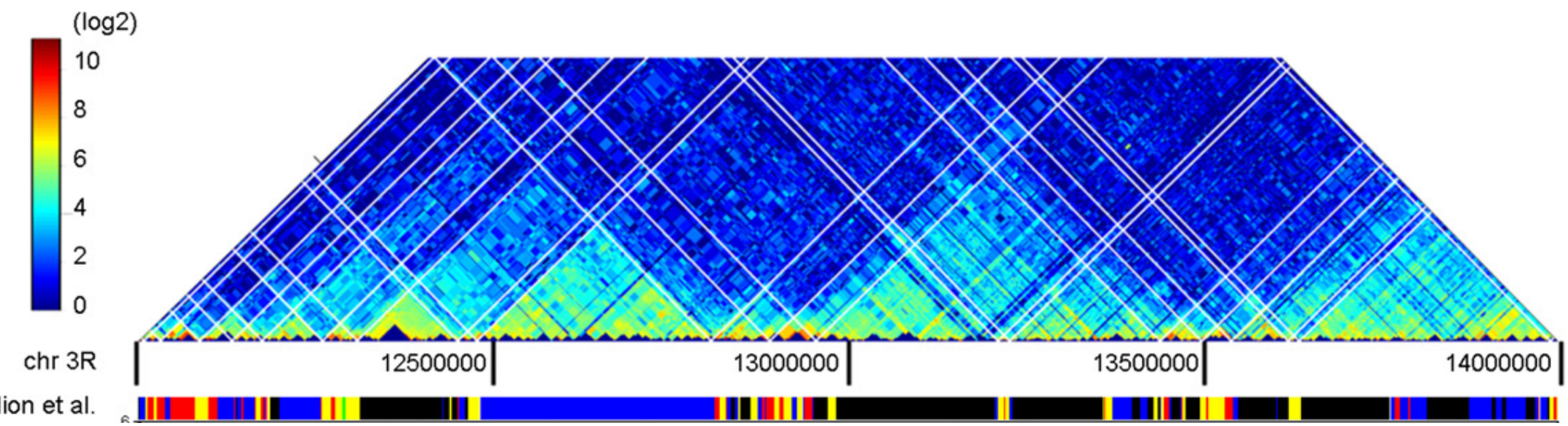


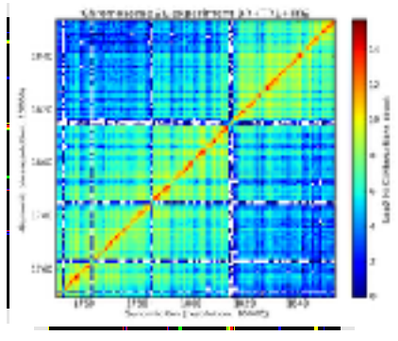
Hidden Markov model

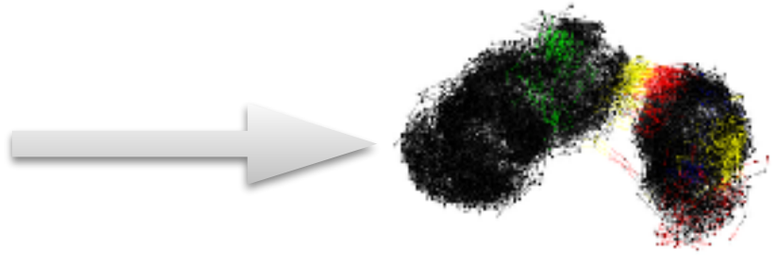


Fly Chromatin COLORs

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

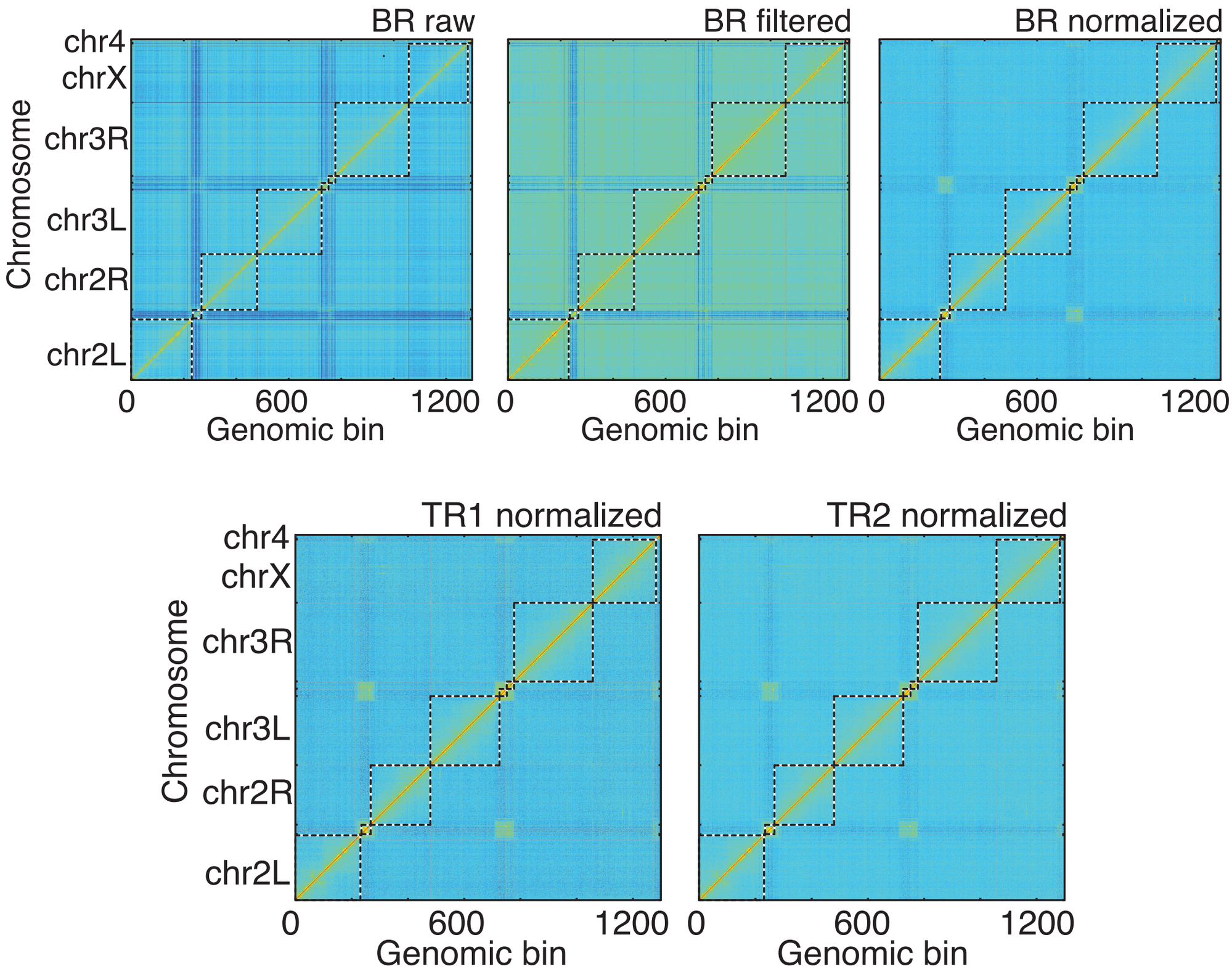




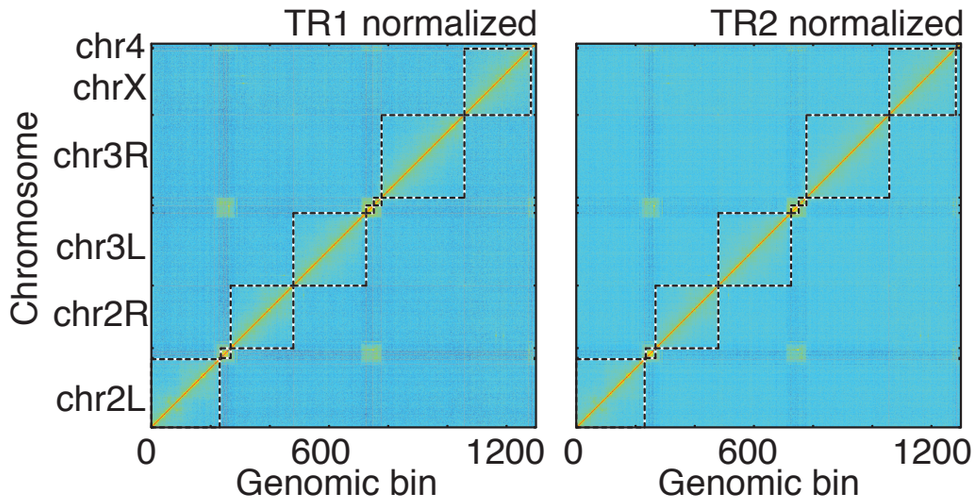


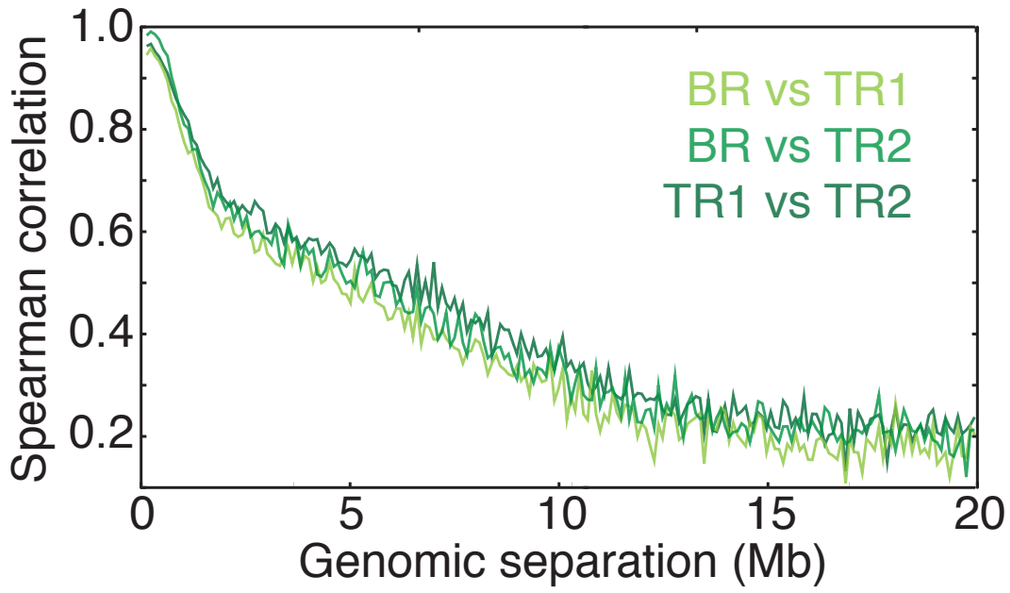
~200 regions of ~5Mb each
2Kb resolution

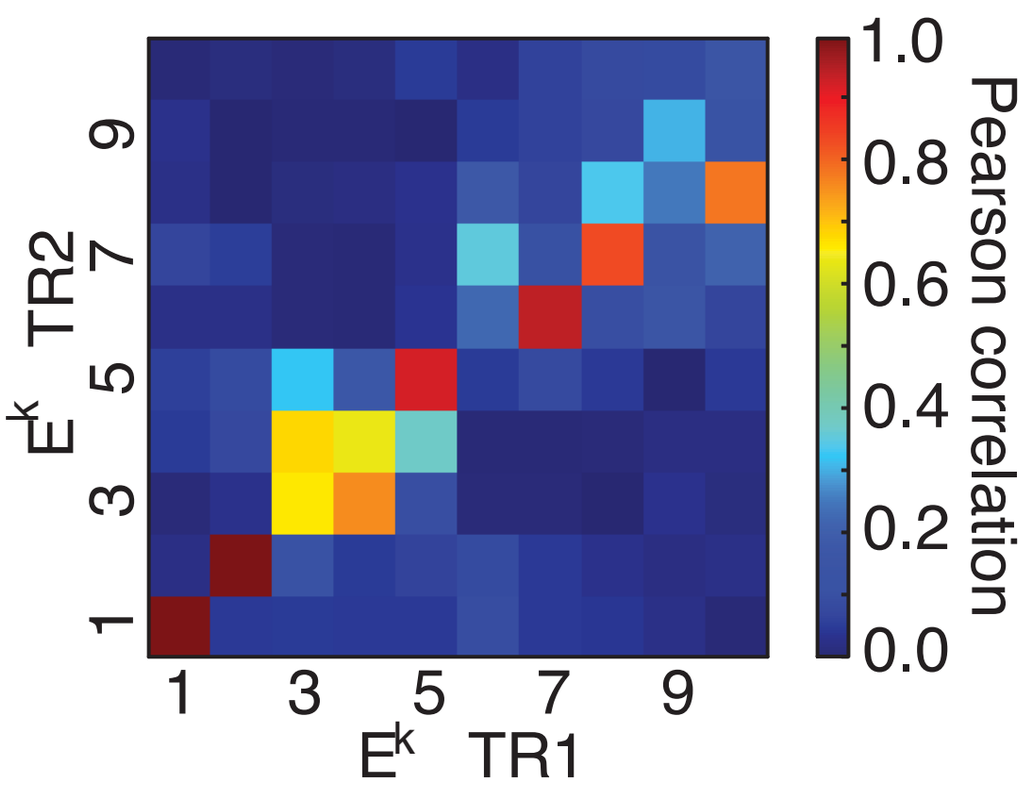
Mapping · Filtering · Normalizing



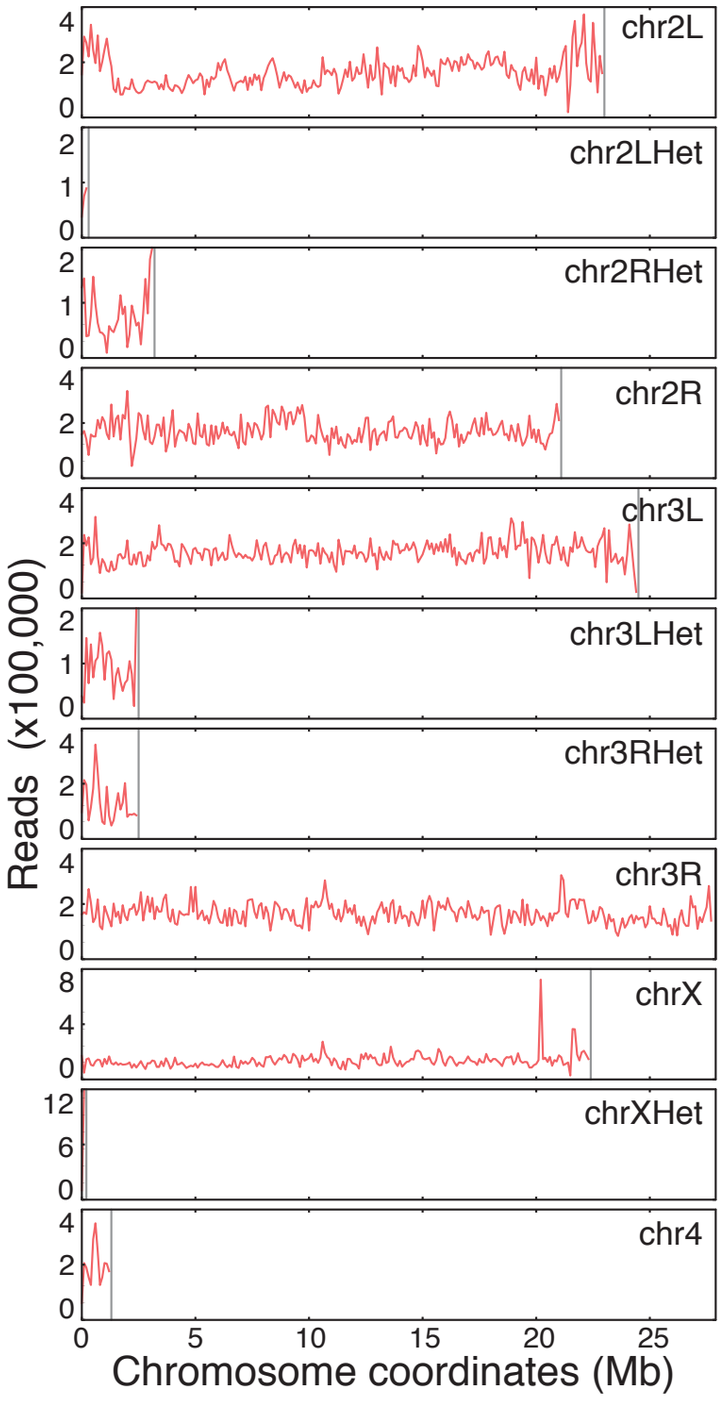
Matrix comparison

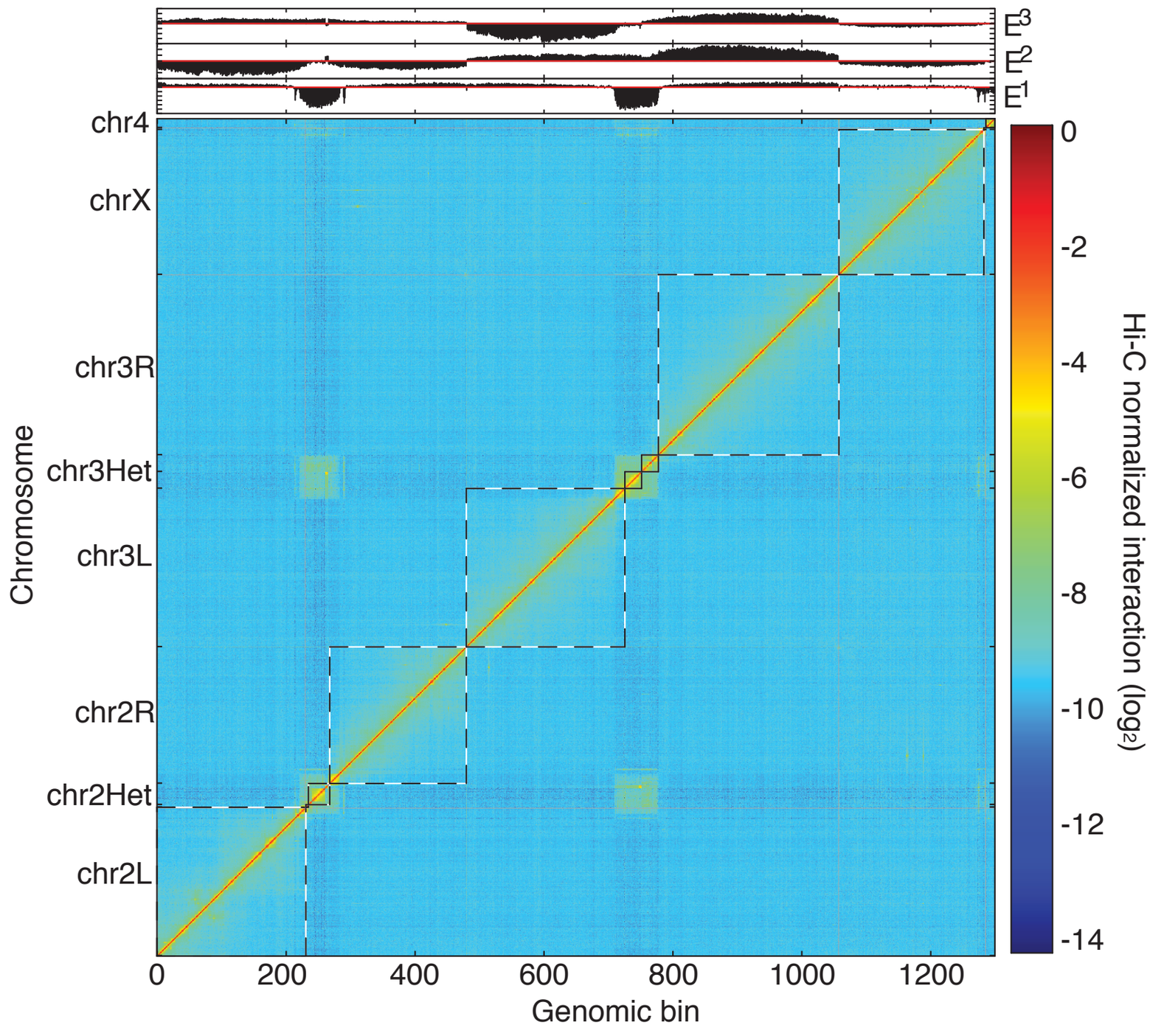




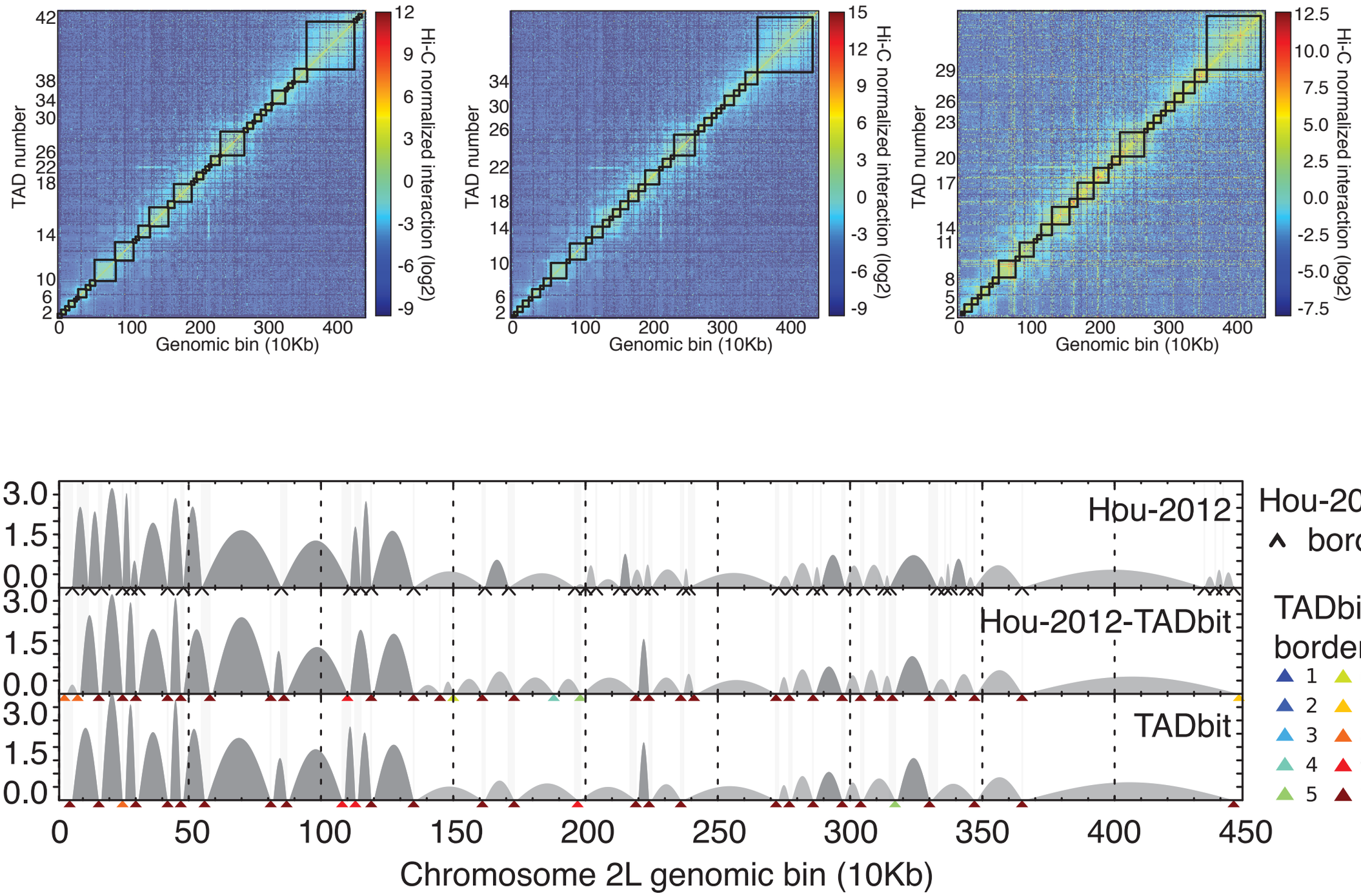


Matrix merging





TAD detection · comparison

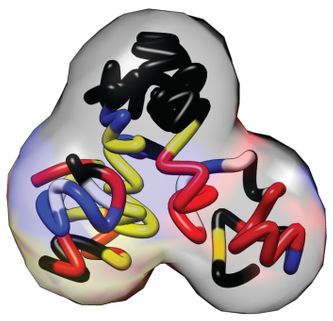


Structural properties

50 1Mb regions. 10 enriched for each color.

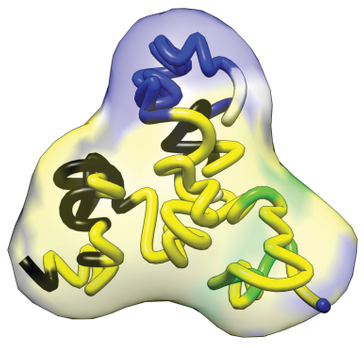
RED dense region
3R:18920000-19920000

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



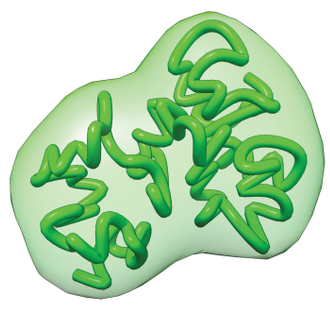
YELLOW dense region
X:15590000-16600000

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



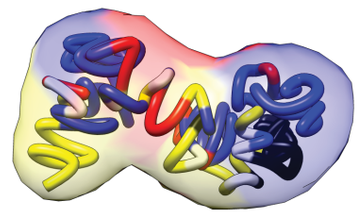
GREEN dense region
2R:510000-1530000

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



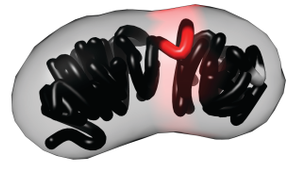
BLUE dense region
3L:210000-1230000

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

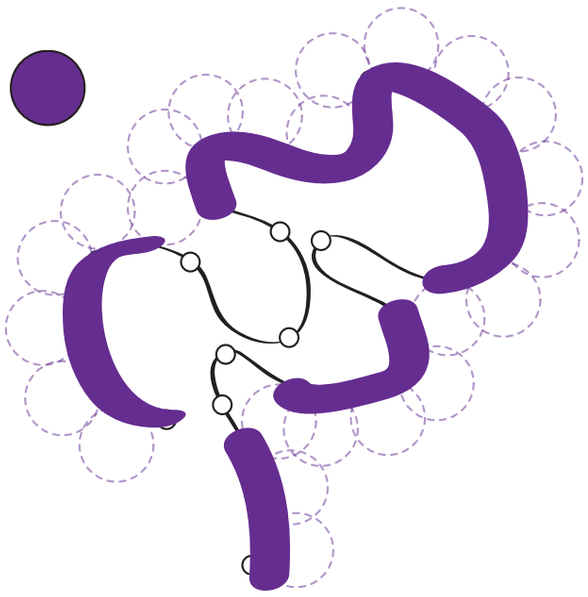


BLACK dense region
2L:17500000-18530000

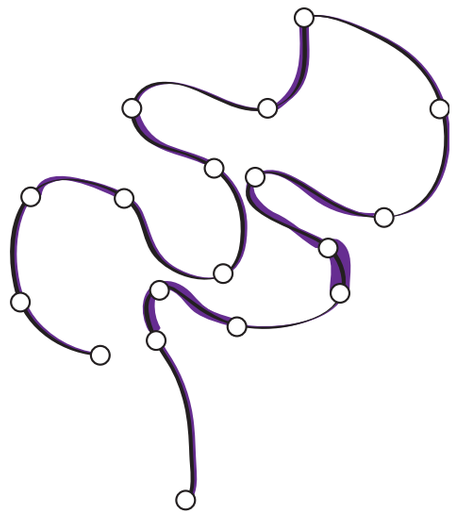
1% 0% 0% 0% 98% 1%



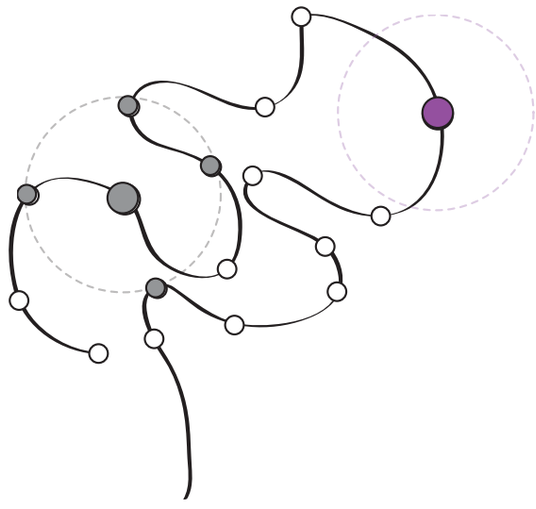
Accessibility (%)



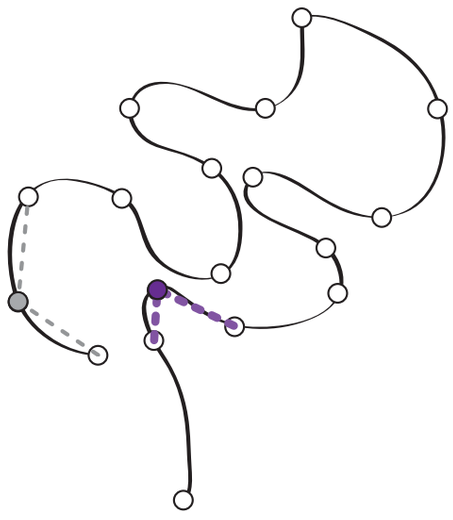
Density (bp/nm)



Interactions



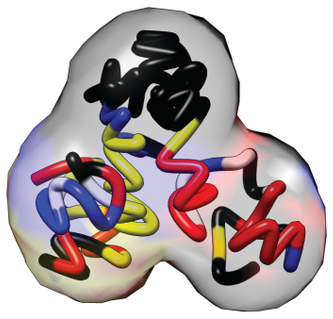
Angle



Structural COLORs

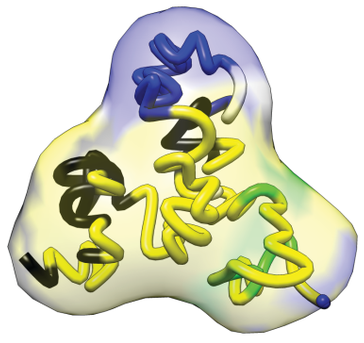
RED dense region
3R:18920000-19920000

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



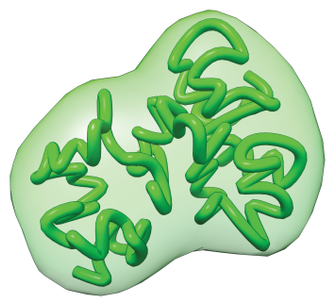
YELLOW dense region
X:15590000-16600000

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



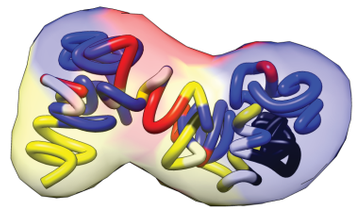
GREEN dense region
2R:510000-1530000

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



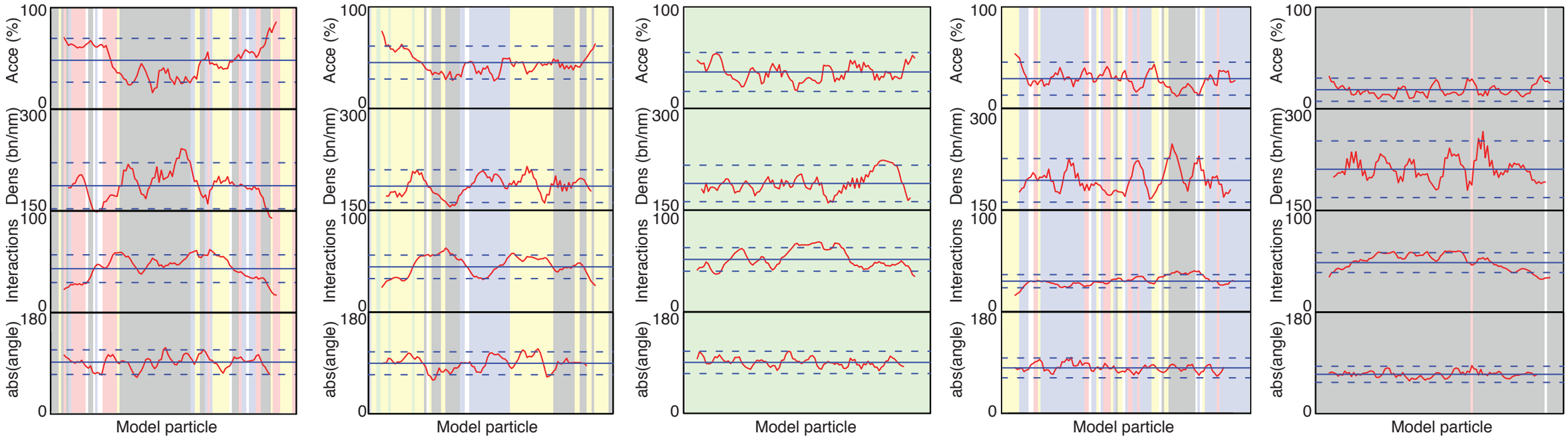
BLUE dense region
3L:210000-1230000

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

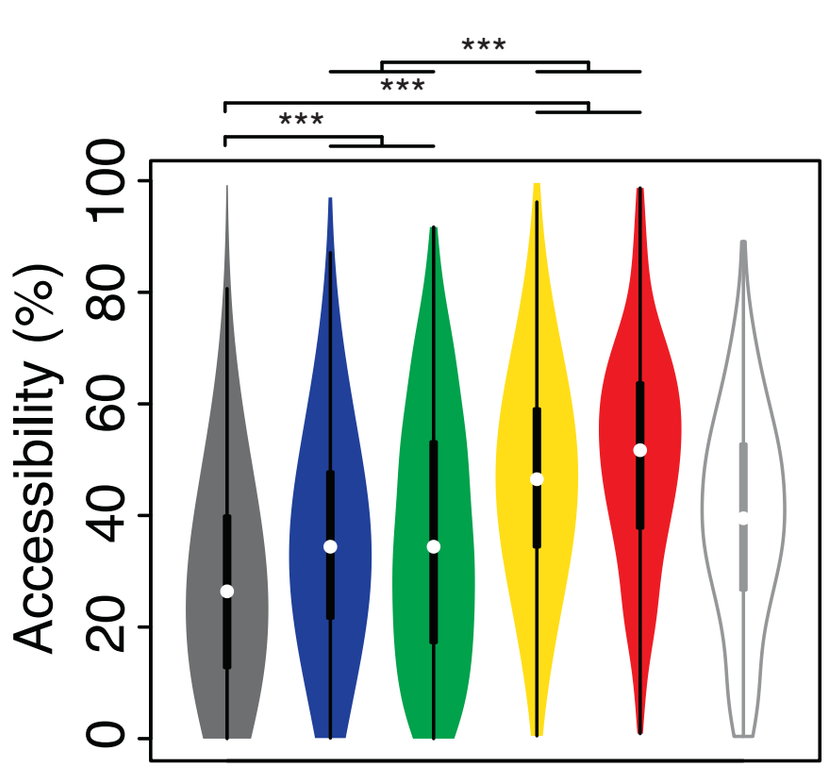


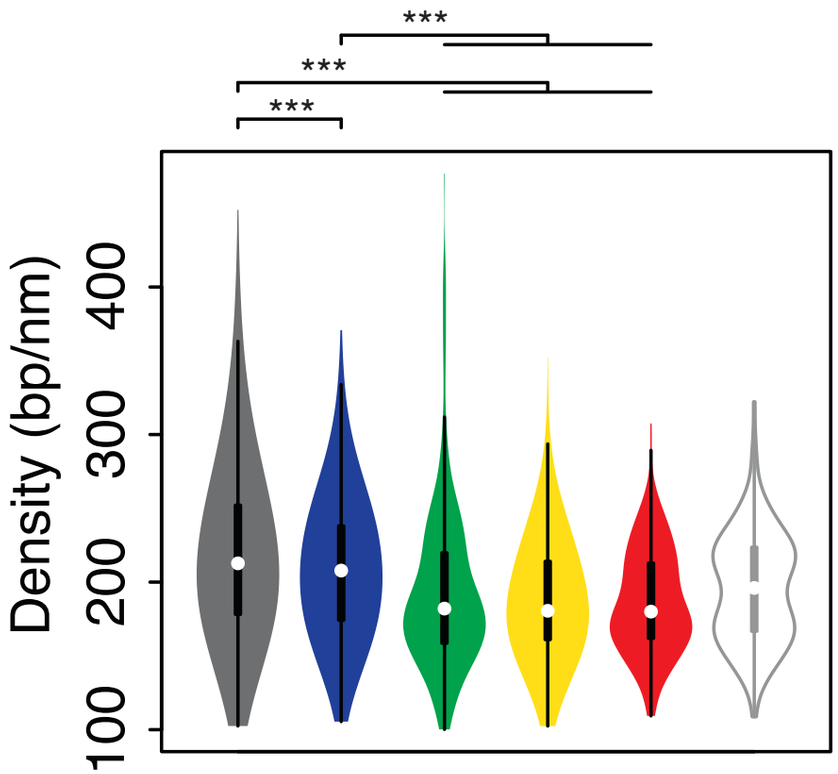
BLACK dense region
2L:17500000-18530000

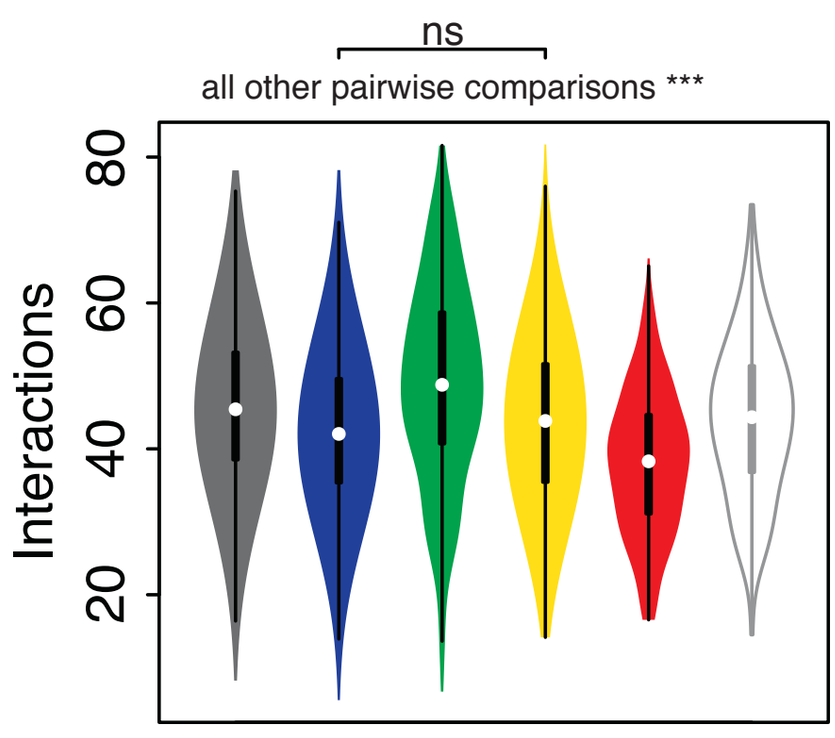
1% 0% 0% 0% 98% 1%

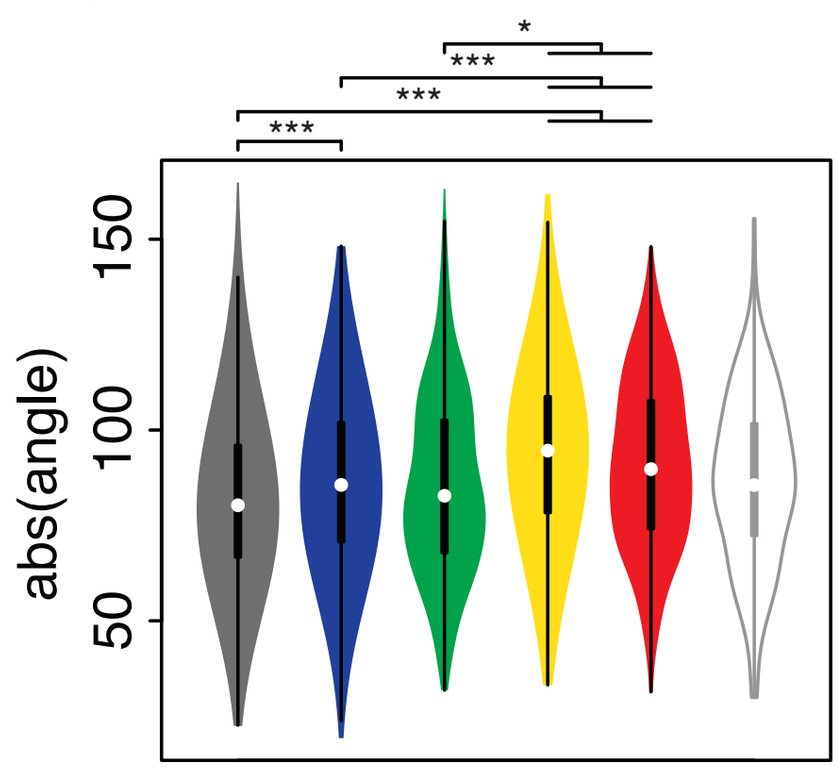


Structural COLORs

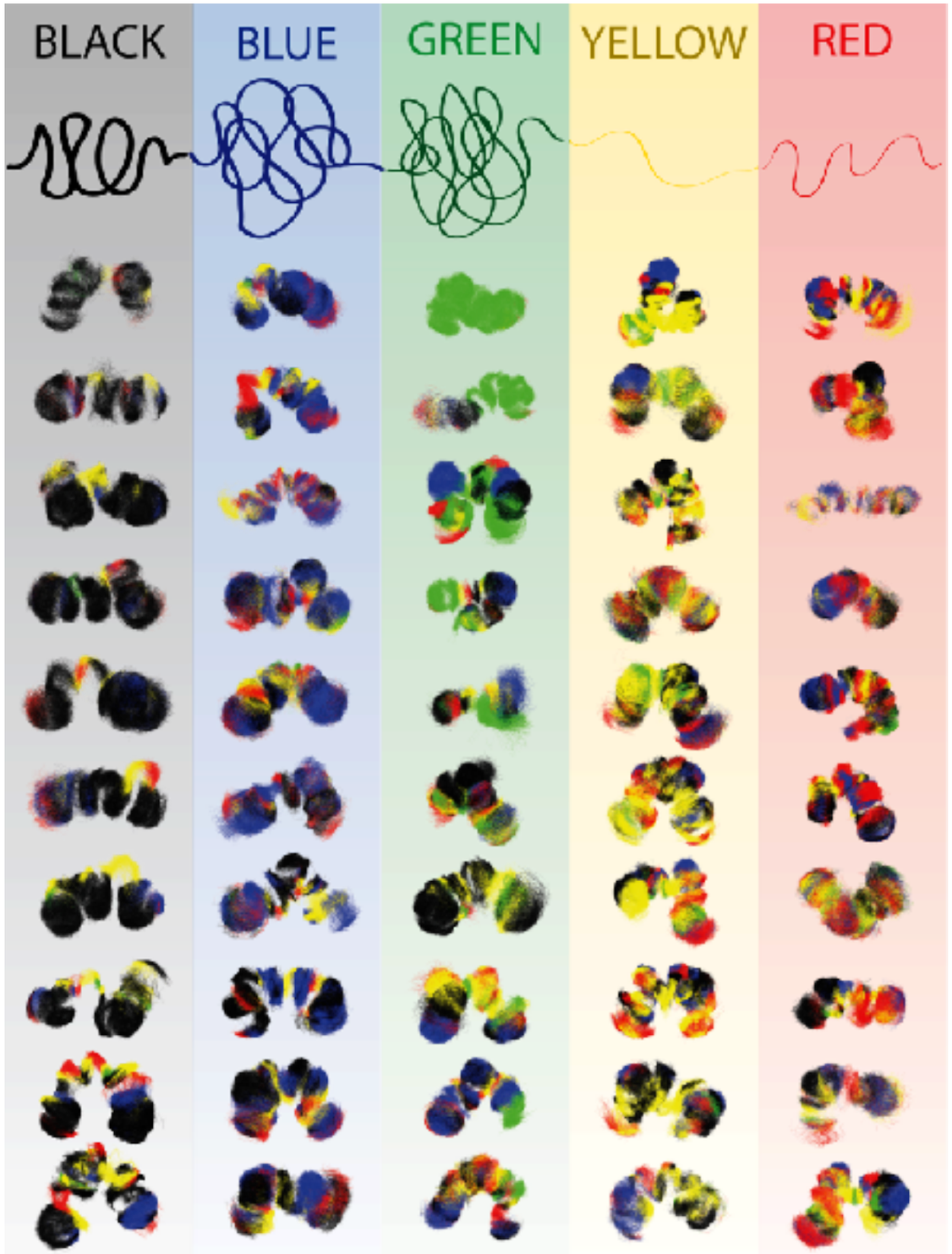


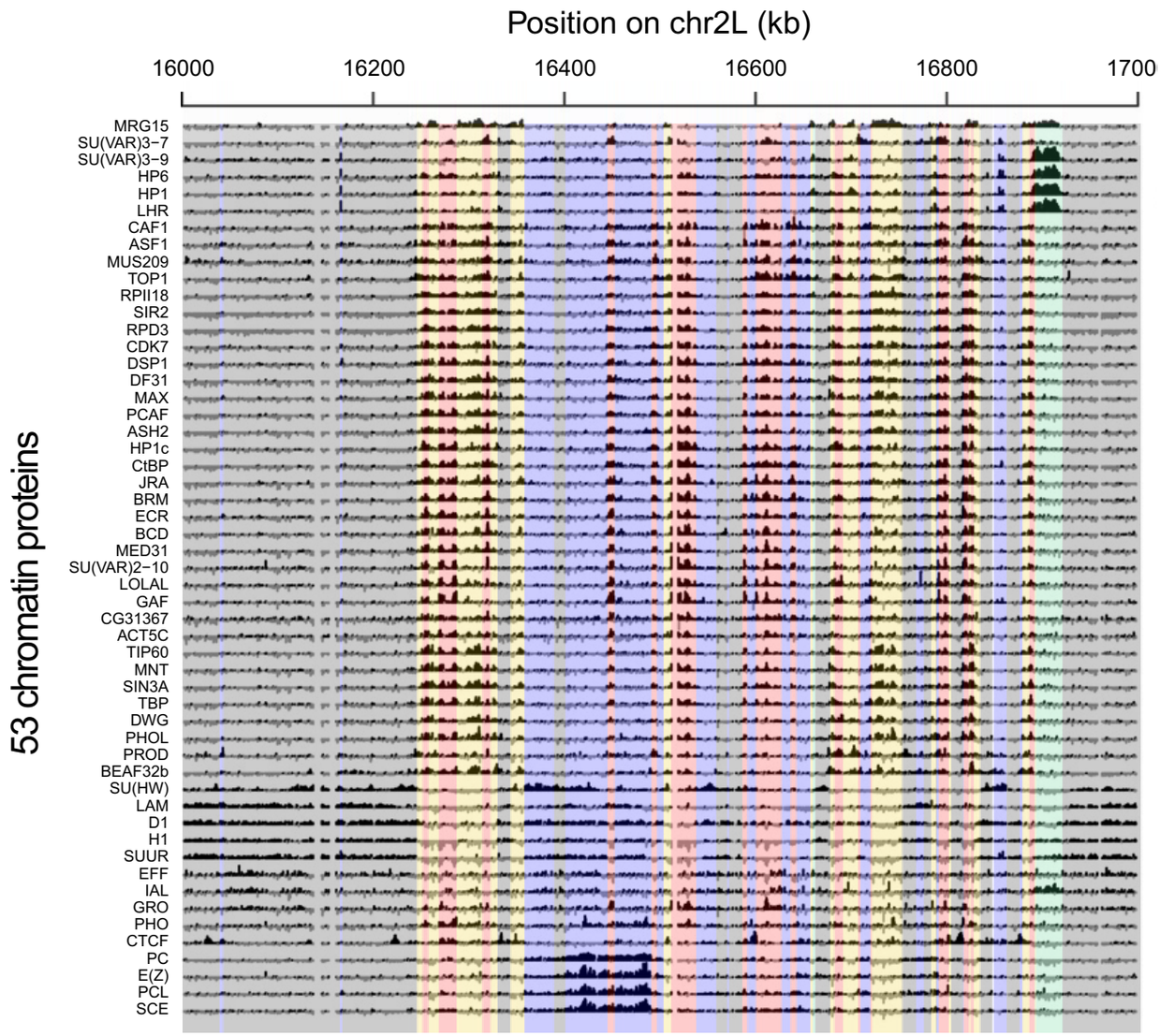






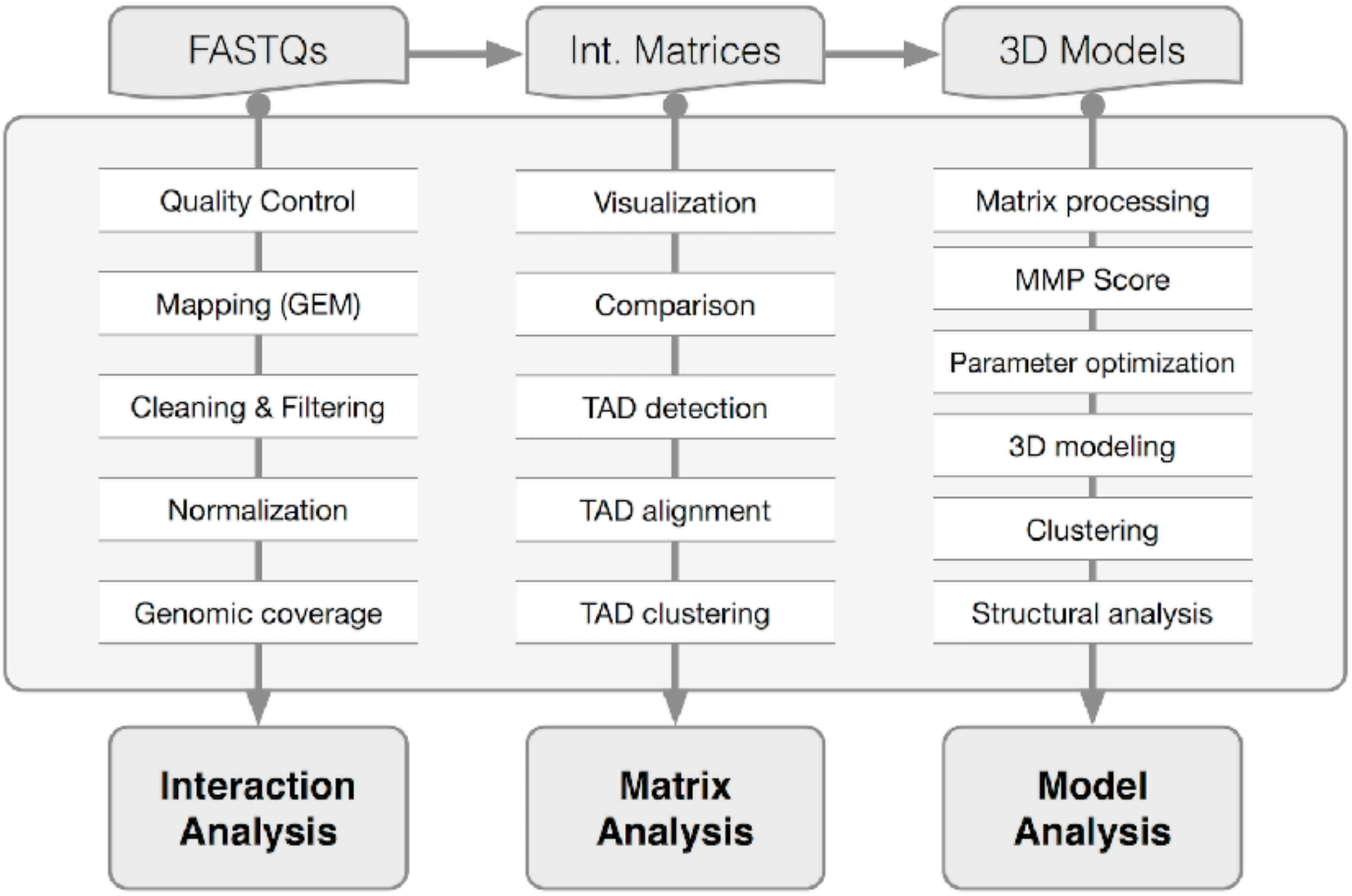
Structural COLORs





TADbit

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



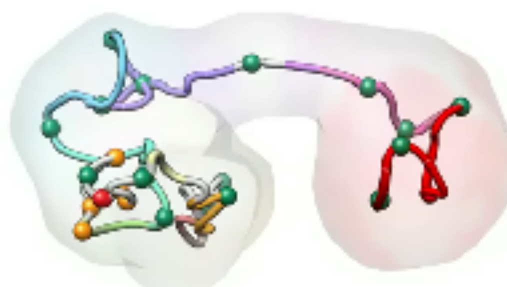
TADbit applications...

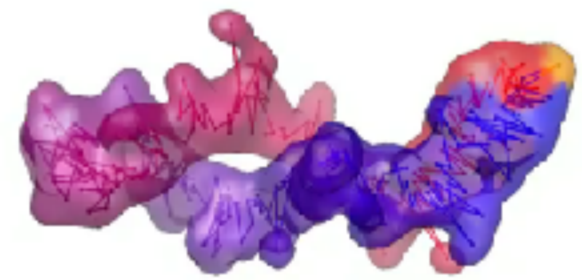
Baù, D. et al. Nat Struct Mol Biol (2011)
Umbarger, M. A. et al. Mol Cell (2011)
Le Dily, F. et al. Genes & Dev (2014)
Trussart et al. Nature Comm. (2017)
Cattoni et al. Nature Communication (2017)


TECHNICAL REPORTS

nature structural & molecular biology

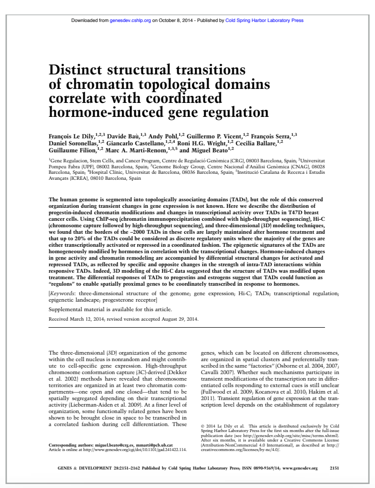
The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules

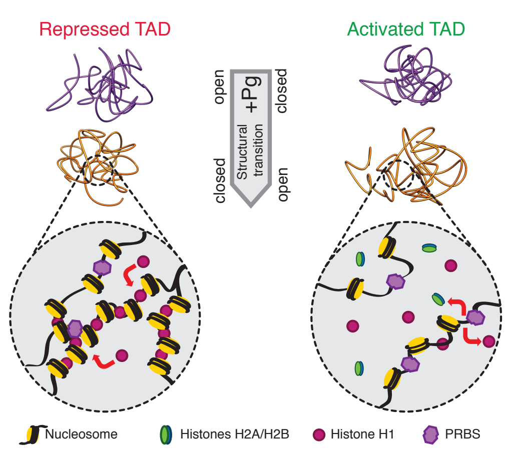


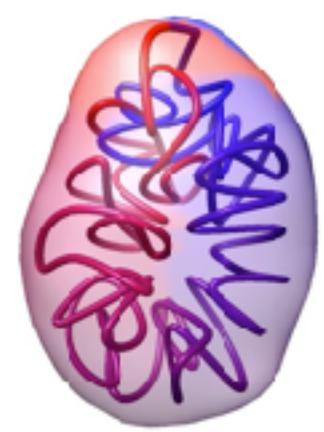



Molecular Cell
Article

The Three-Dimensional Architecture of a Bacterial Genome and Its Alteration by Genetic Perturbation


Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation






nature COMMUNICATIONS

ARTICLE

Defined chromosome structure in the genome-reduced bacterium *Mycoplasma pneumoniae*

David Castillo
Yasmina Cuartero
Irene Farabella
Silvia Galan
Mike Goodstadt
Julen Mendieta
Francesca Mugianesi
Juan Rodríguez
François Serra
Paula Soler
Aleksandra Sparavier
Yannick Spill
Marco di Stefano

Former members: Davide Baù & Marie Trussart

http://sgt.cnag.cat/www/presentations/

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