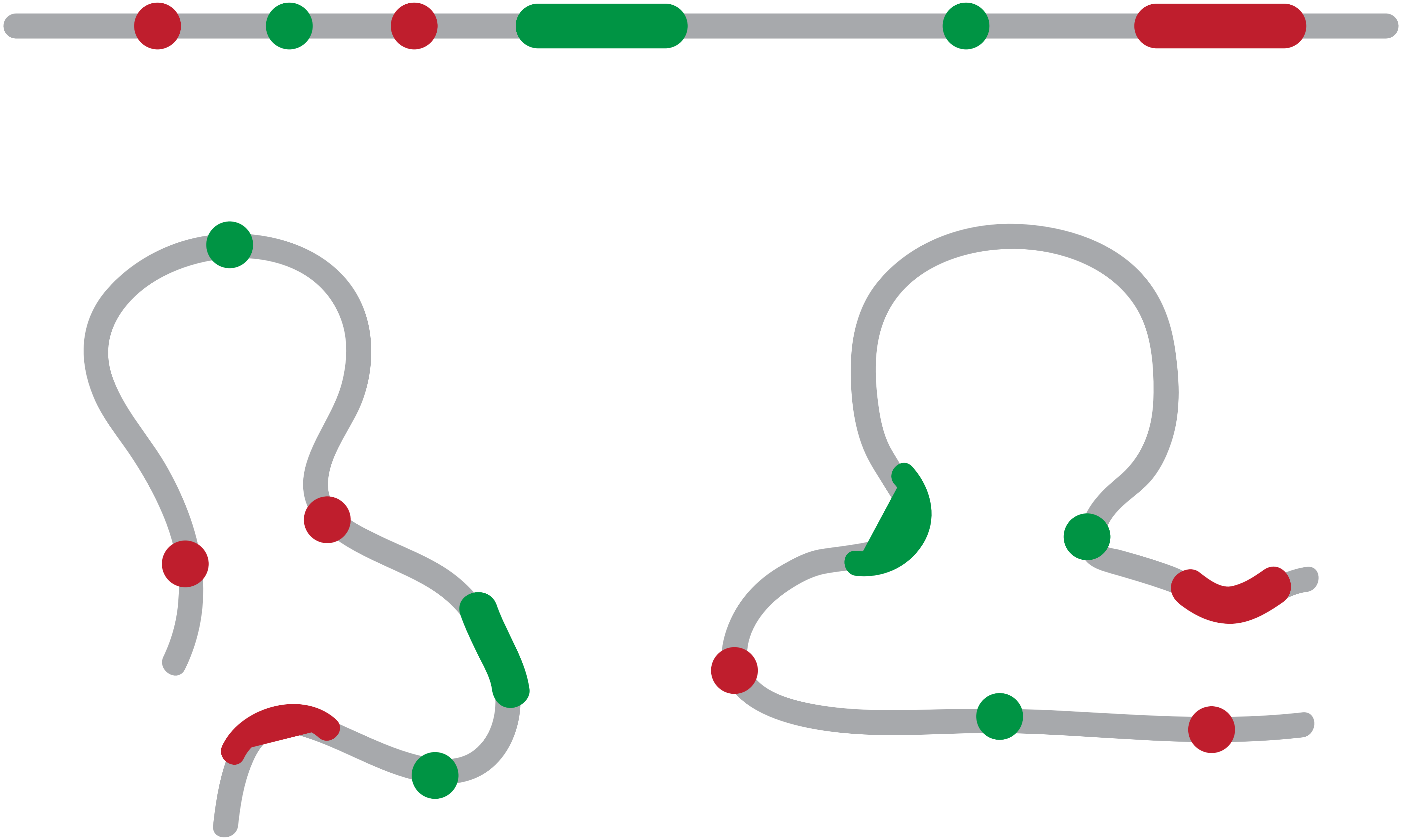
Structure determination of genomes and genomic domains by satisfaction of spatial restraints

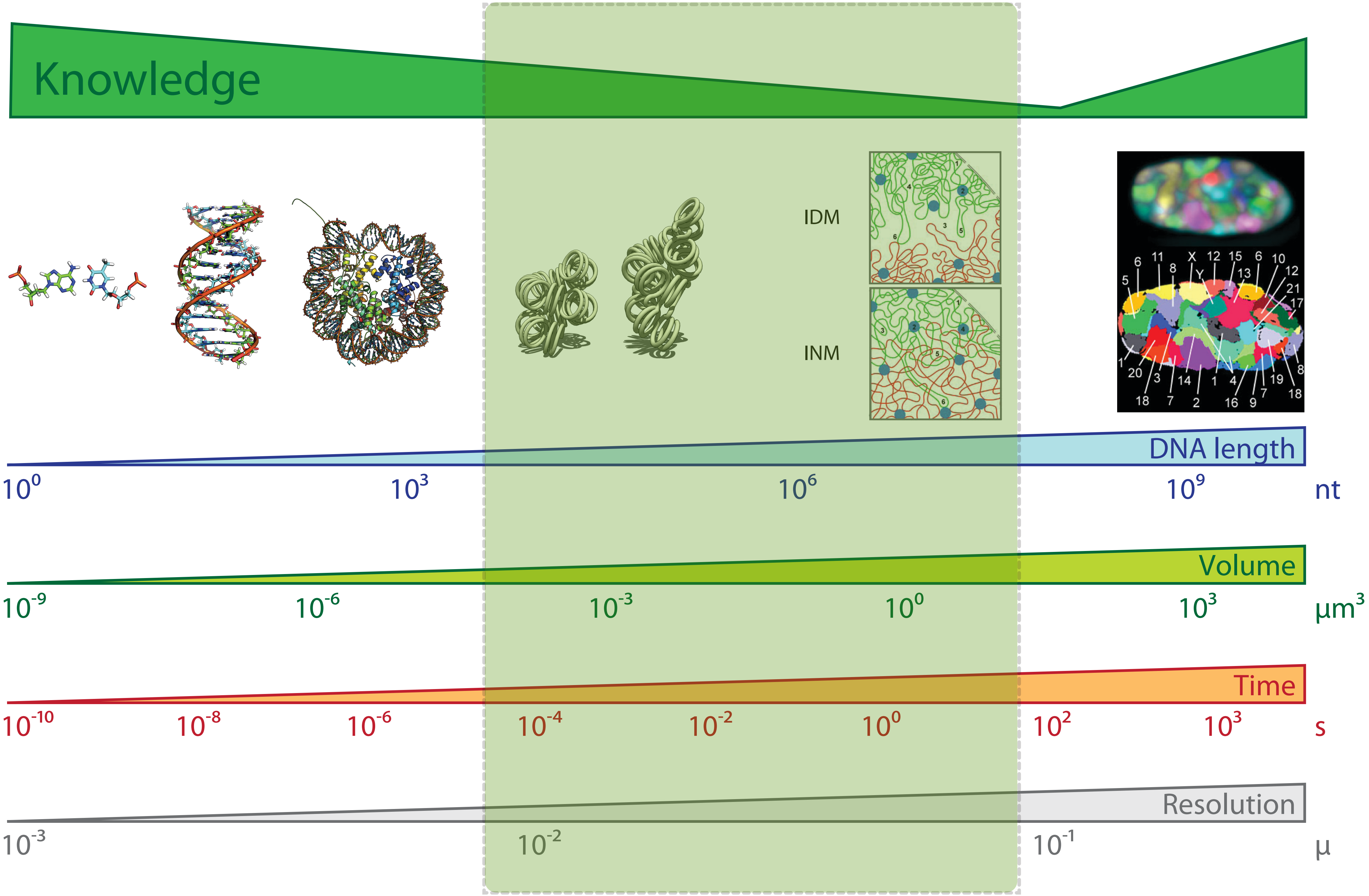
Marc A. Marti-Renom
Structural Genomics Group (ICREA, CNAG-CRG)

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



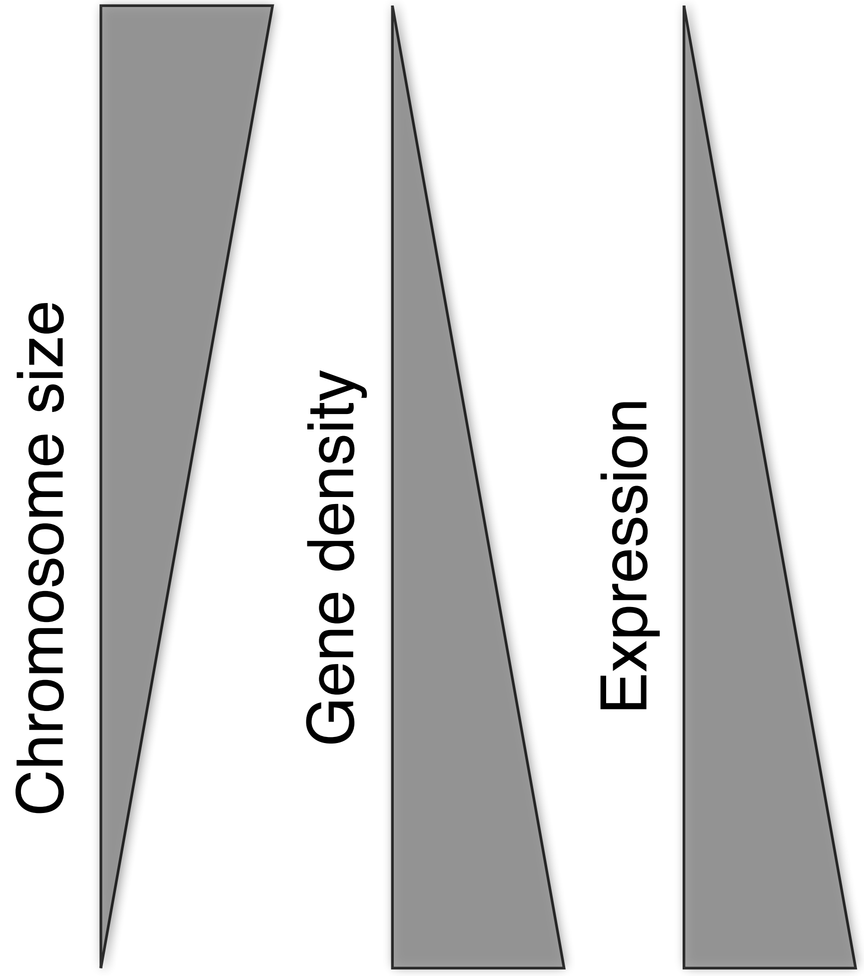
Resolution Gap

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



Level I: Radial genome organization

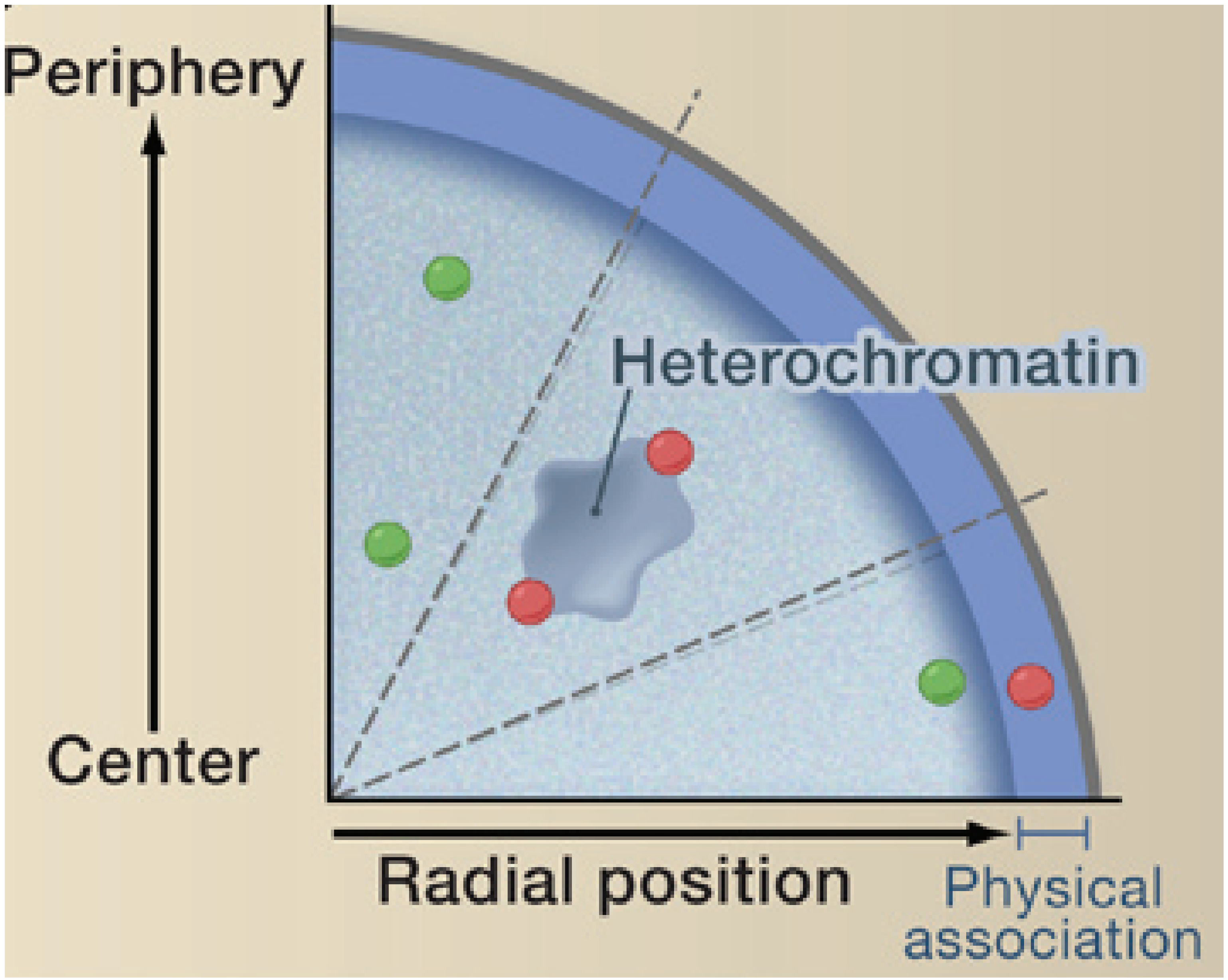
Takizawa, T., Meaburn, K. J. & Misteli, T. The meaning of gene positioning. Cell 135, 9–13 (2008).



Chromosome size

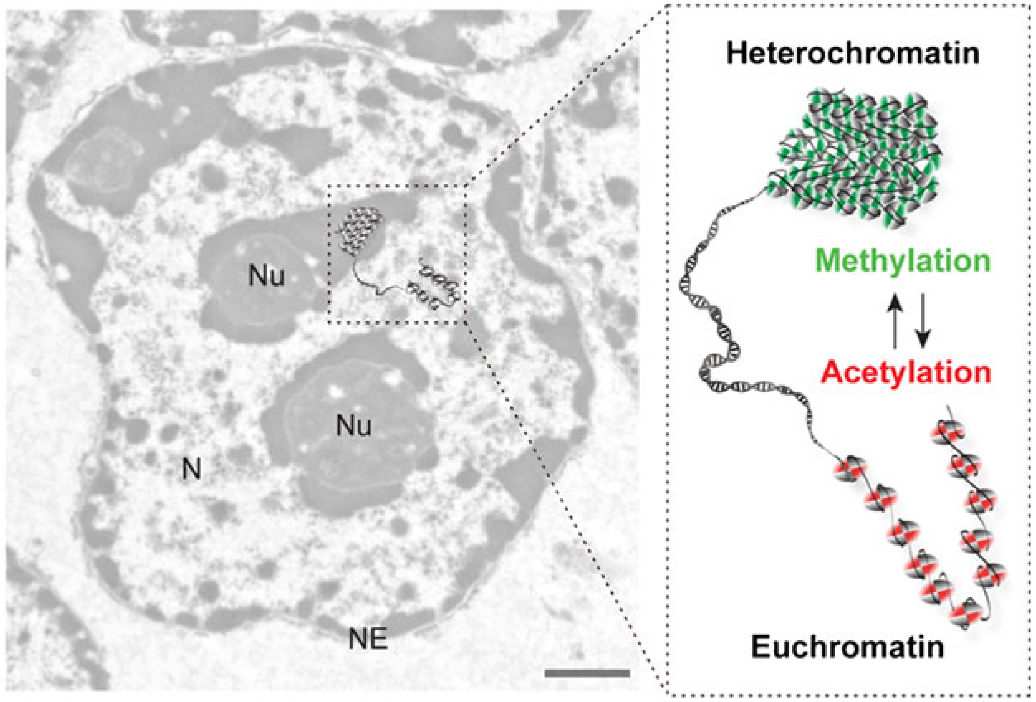
Gene density

Expression



Level II: Euchromatin vs heterochromatin

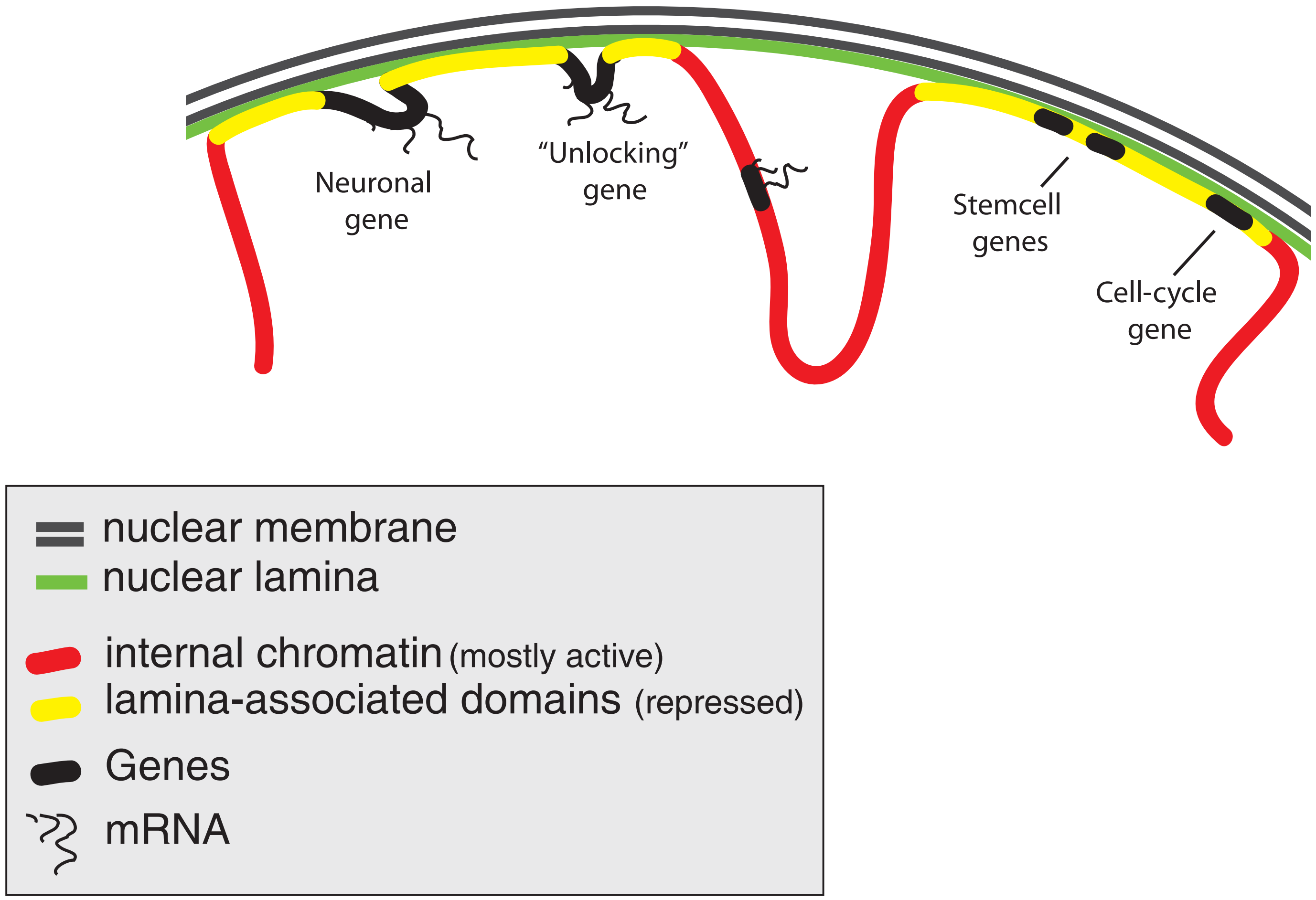
Electron microscopy



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

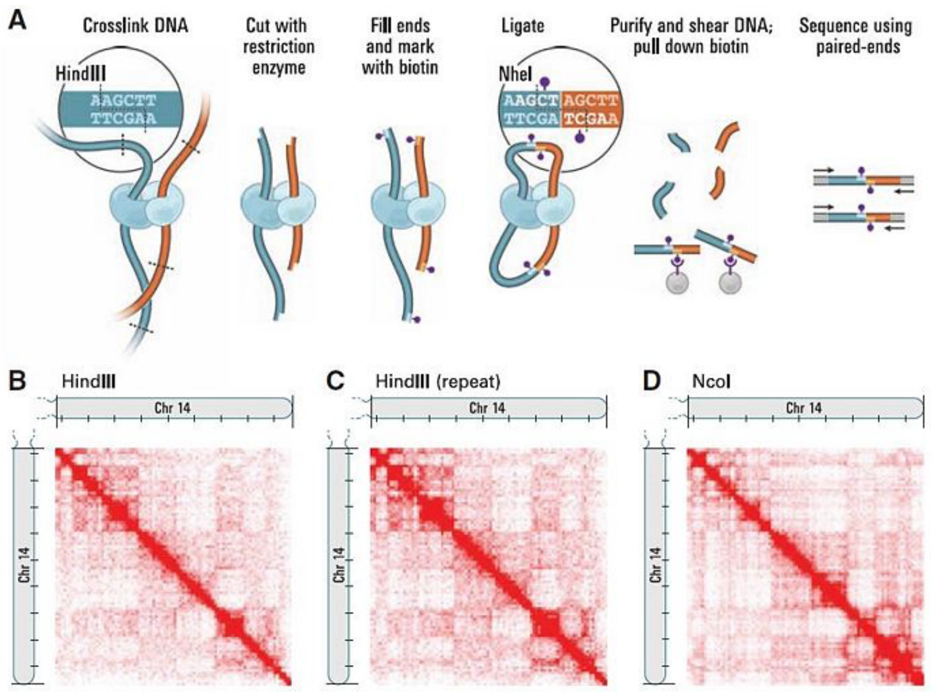


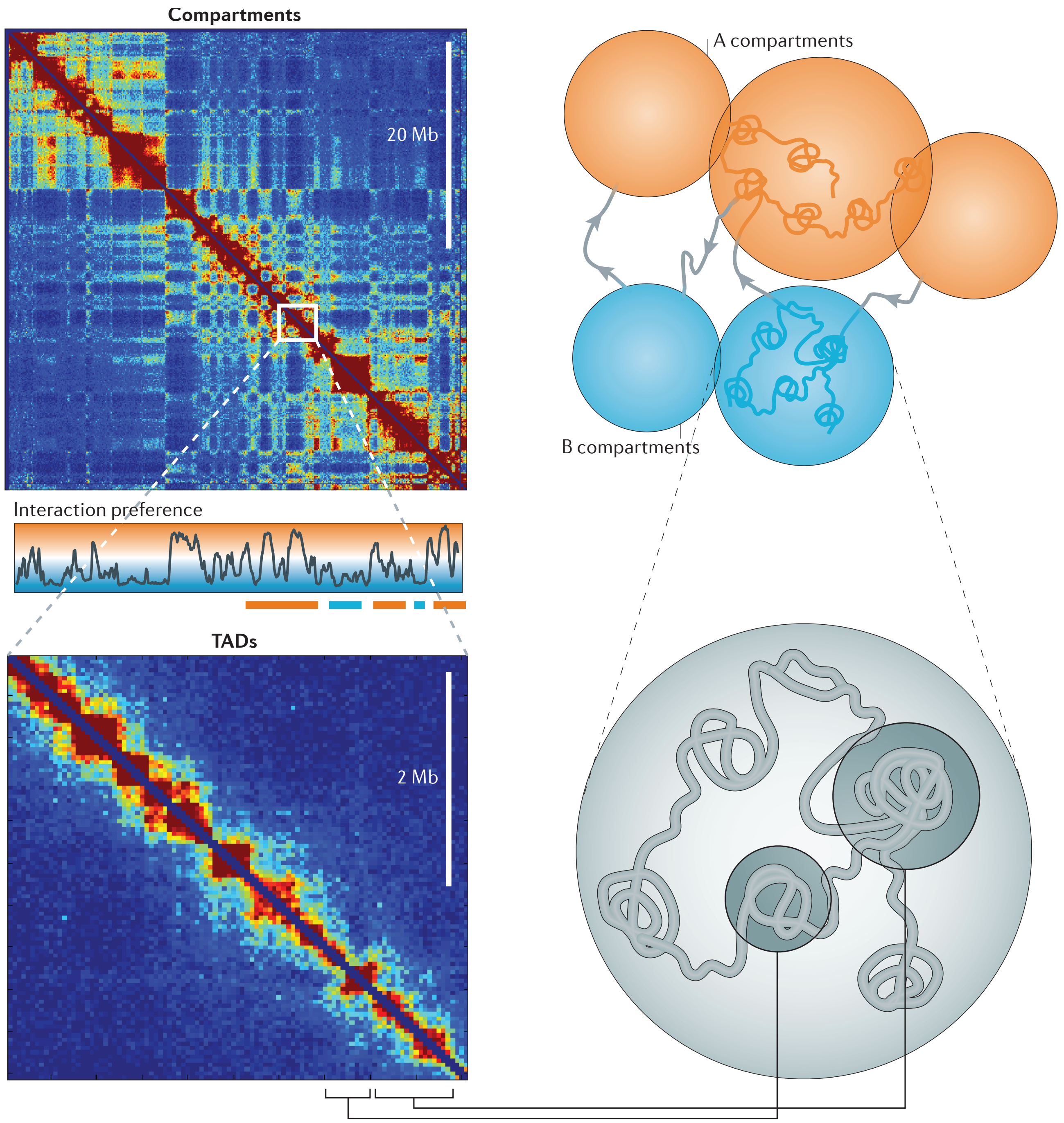
Most genes in Lamina Associated Domains are transcriptionally silent, suggesting that **lamina-genome interactions** are widely involved in the control of **gene expression**

Adapted from Molecular Cell 38, 603-613, 2010

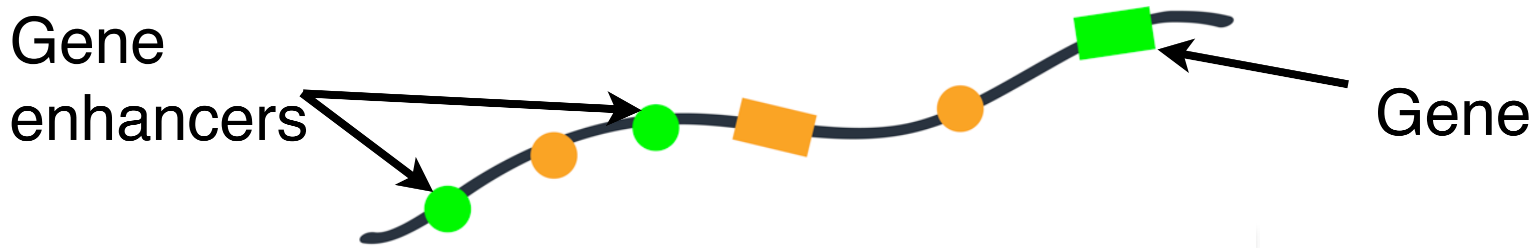
Level IV: Higher-order organization

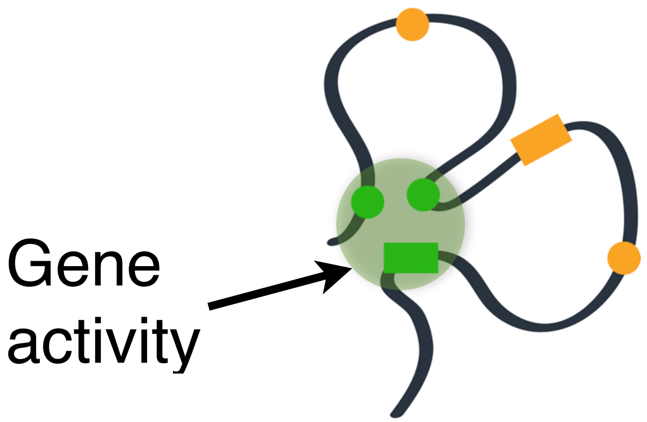
Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Nat Rev Genet 14, 390–403 (2013).





Level V: Chromatin loops





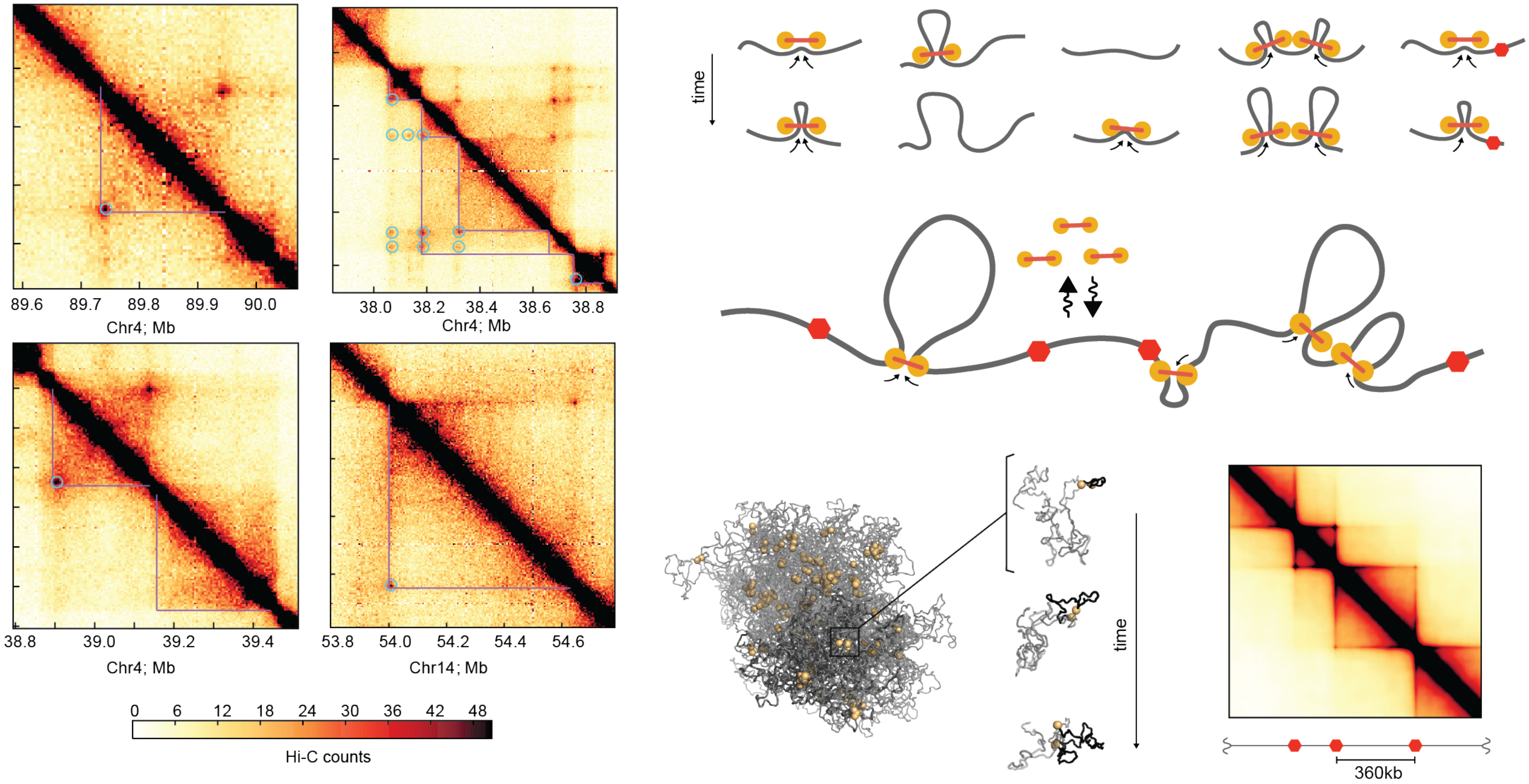
Loops bring distal genomic regions in close proximity to one another

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)

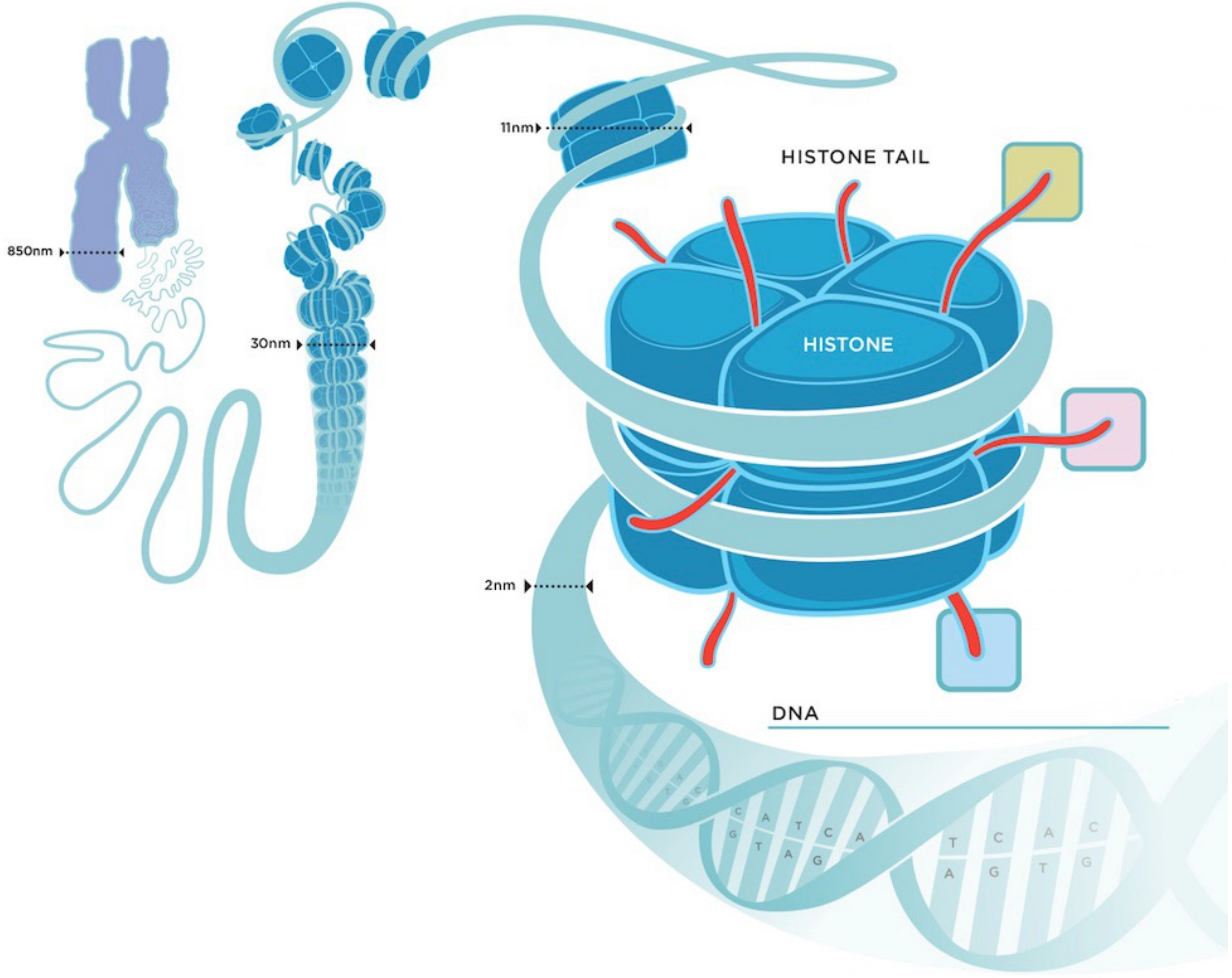
Level V: Loop-extrusion as a driving force

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

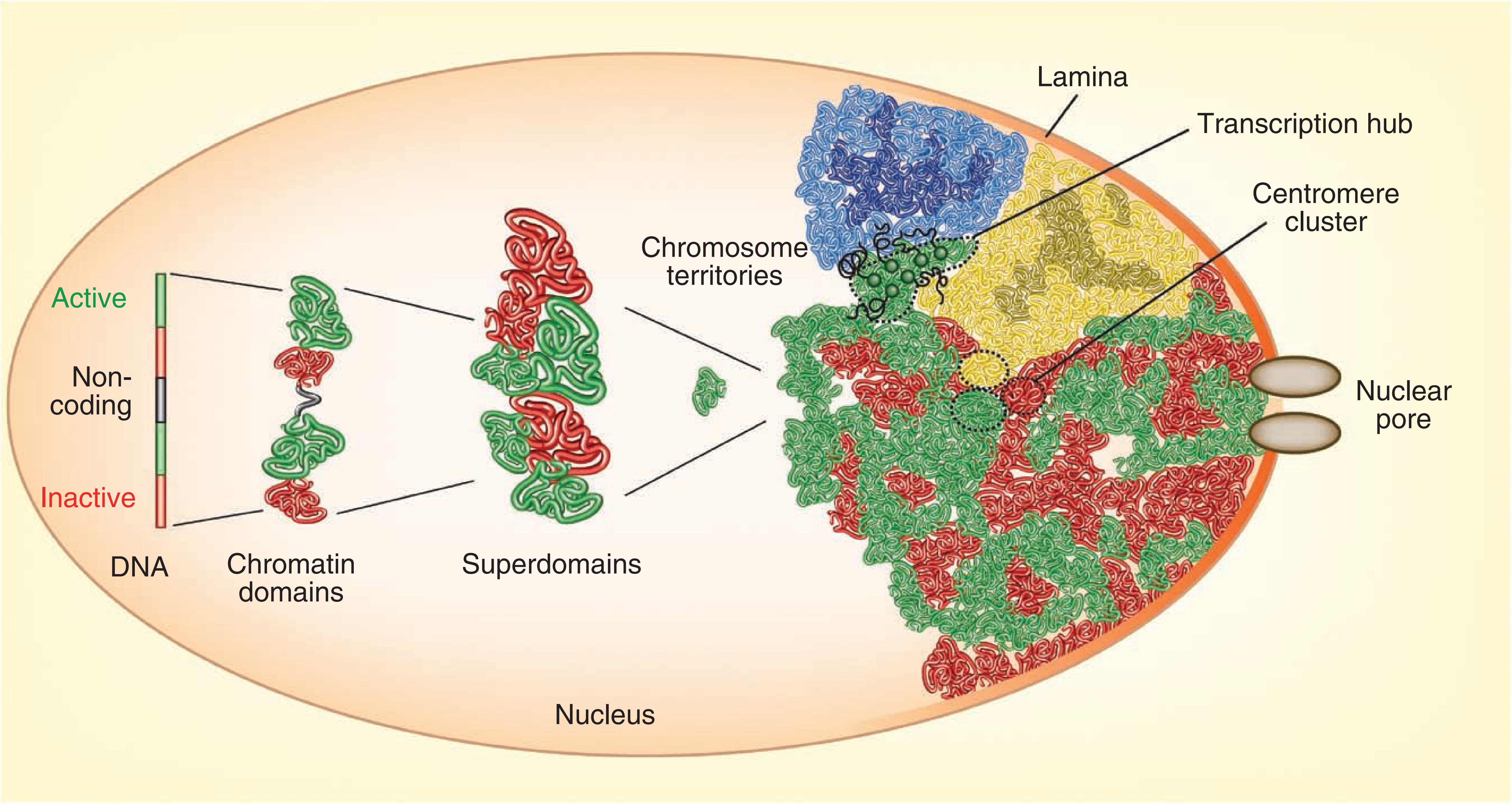
Chromosome **Chromatin fibre** **Nucleosome**



Adapted from Richard E. Ballermann, 2012

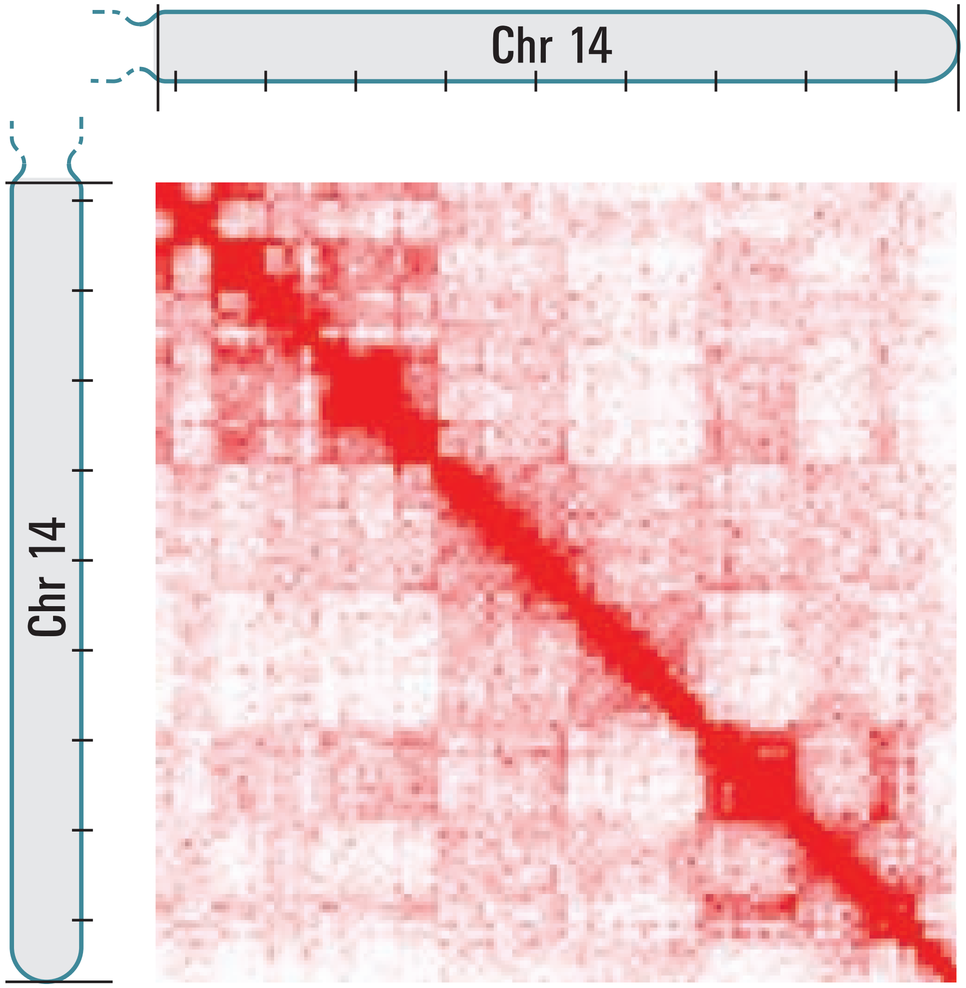
Complex genome organization

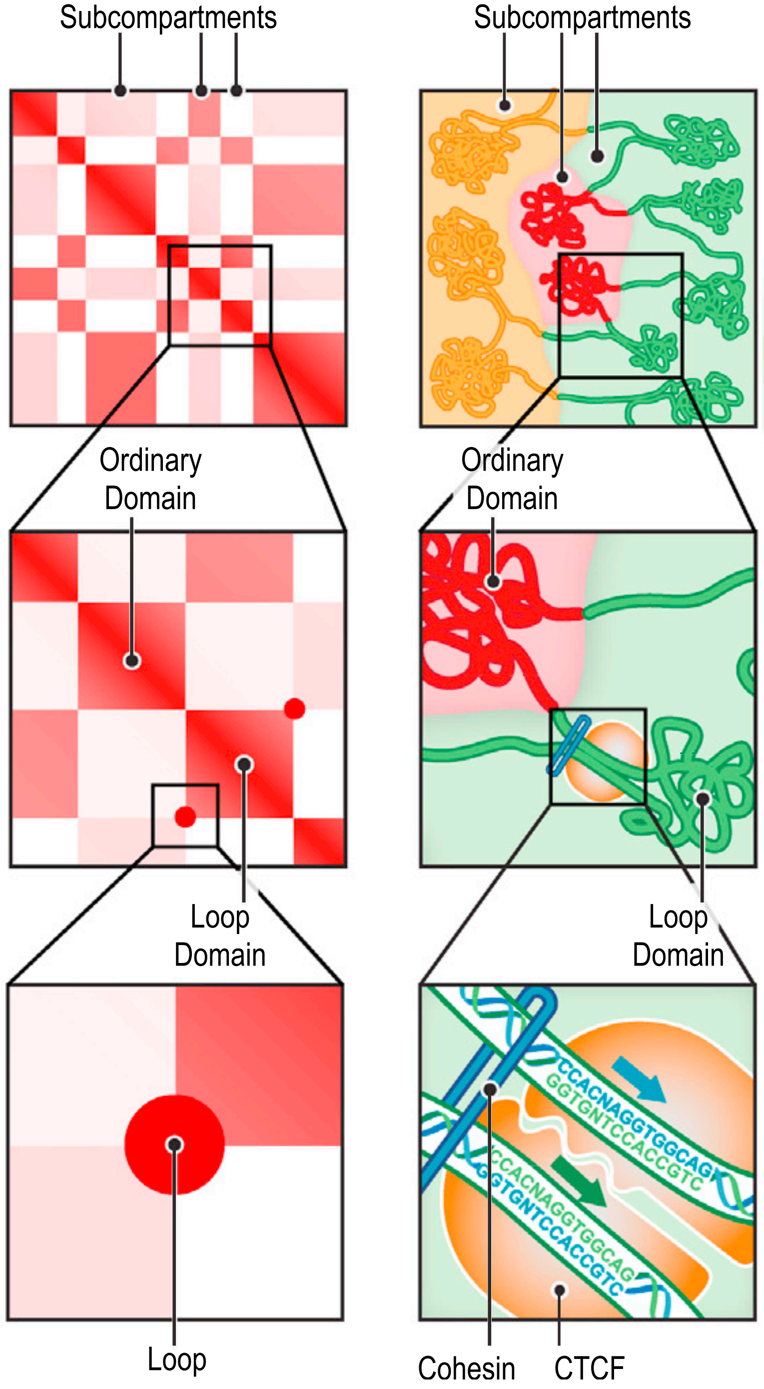
Cavalli, G. & Misteli, T. Functional implications of genome topology. Nat Struct Mol Biol 20, 290–299 (2013).



Marina Corral

Hierarchical genome organisation

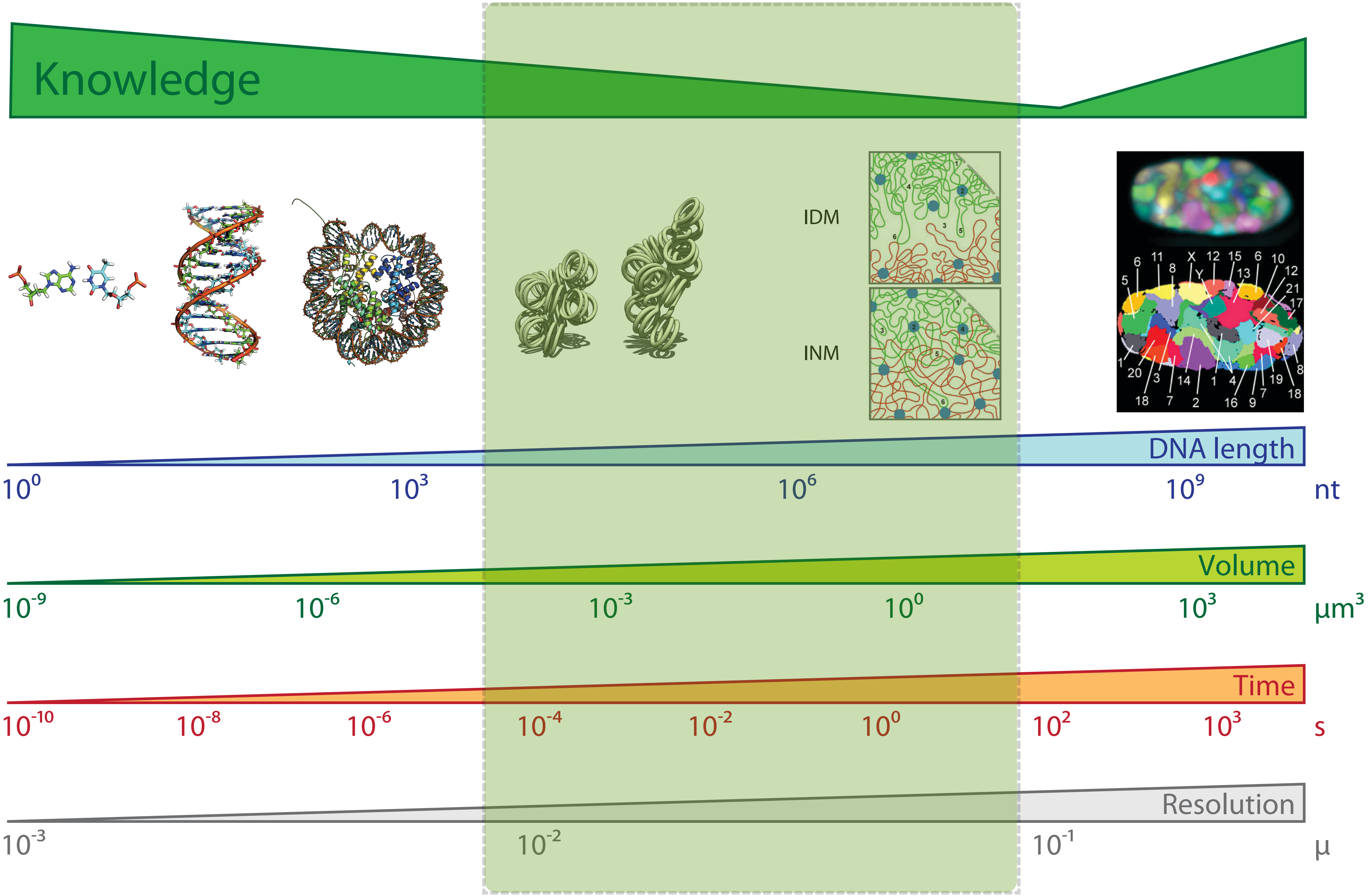




Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.
Rao, S. S. P., et al. (2014). Cell, 1–29.

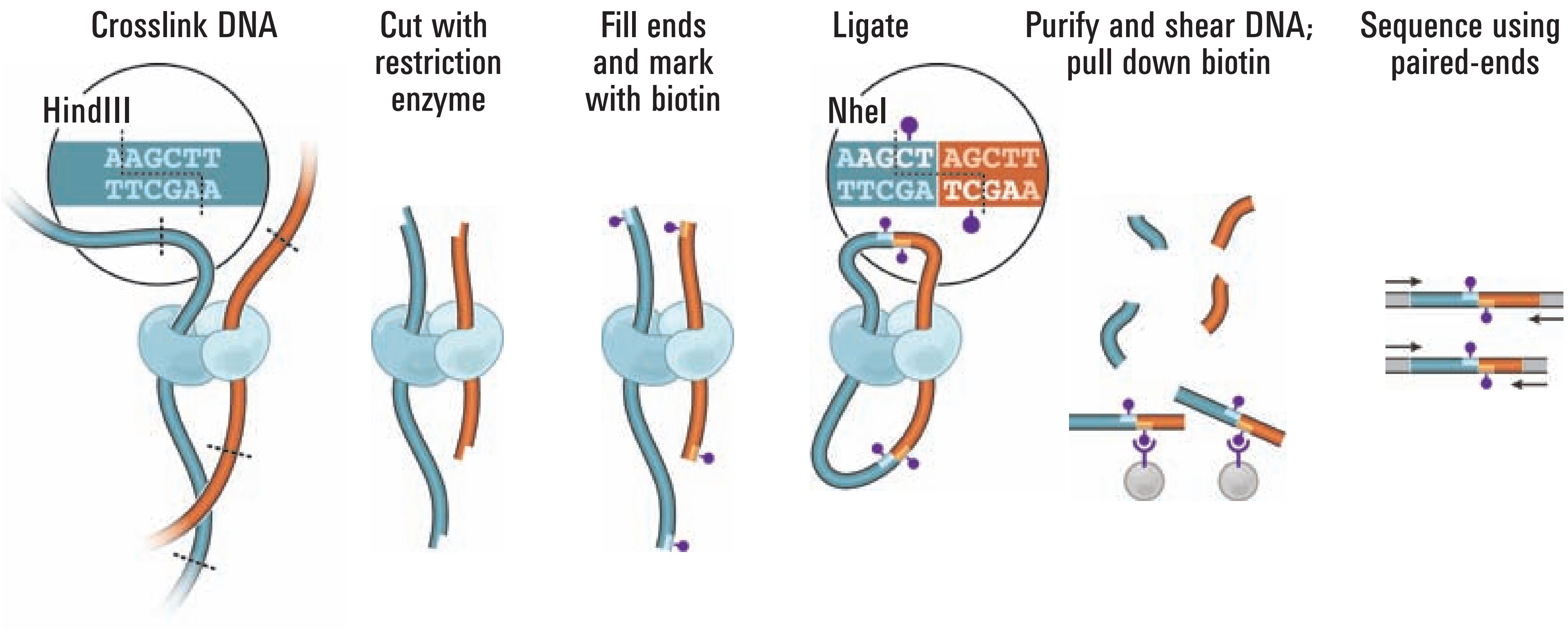
Resolution Gap

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.


Crosslink DNA

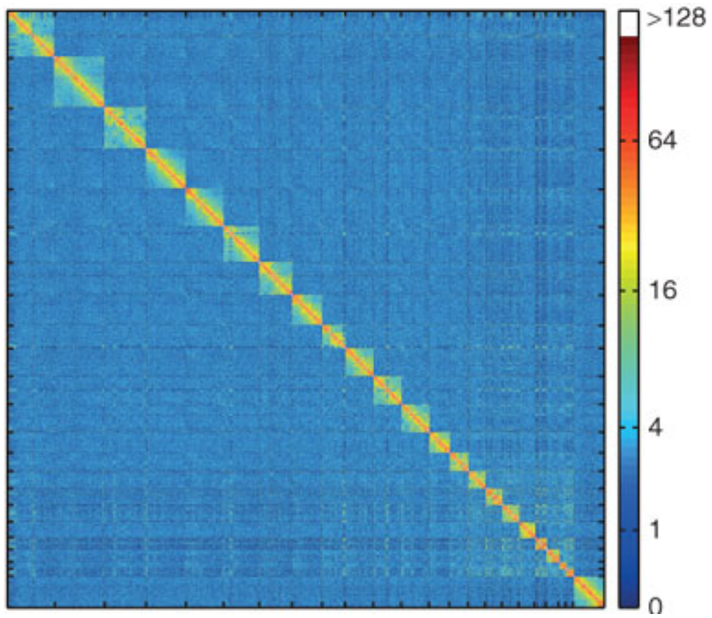
Cut with restriction enzyme

Fill ends and mark with biotin

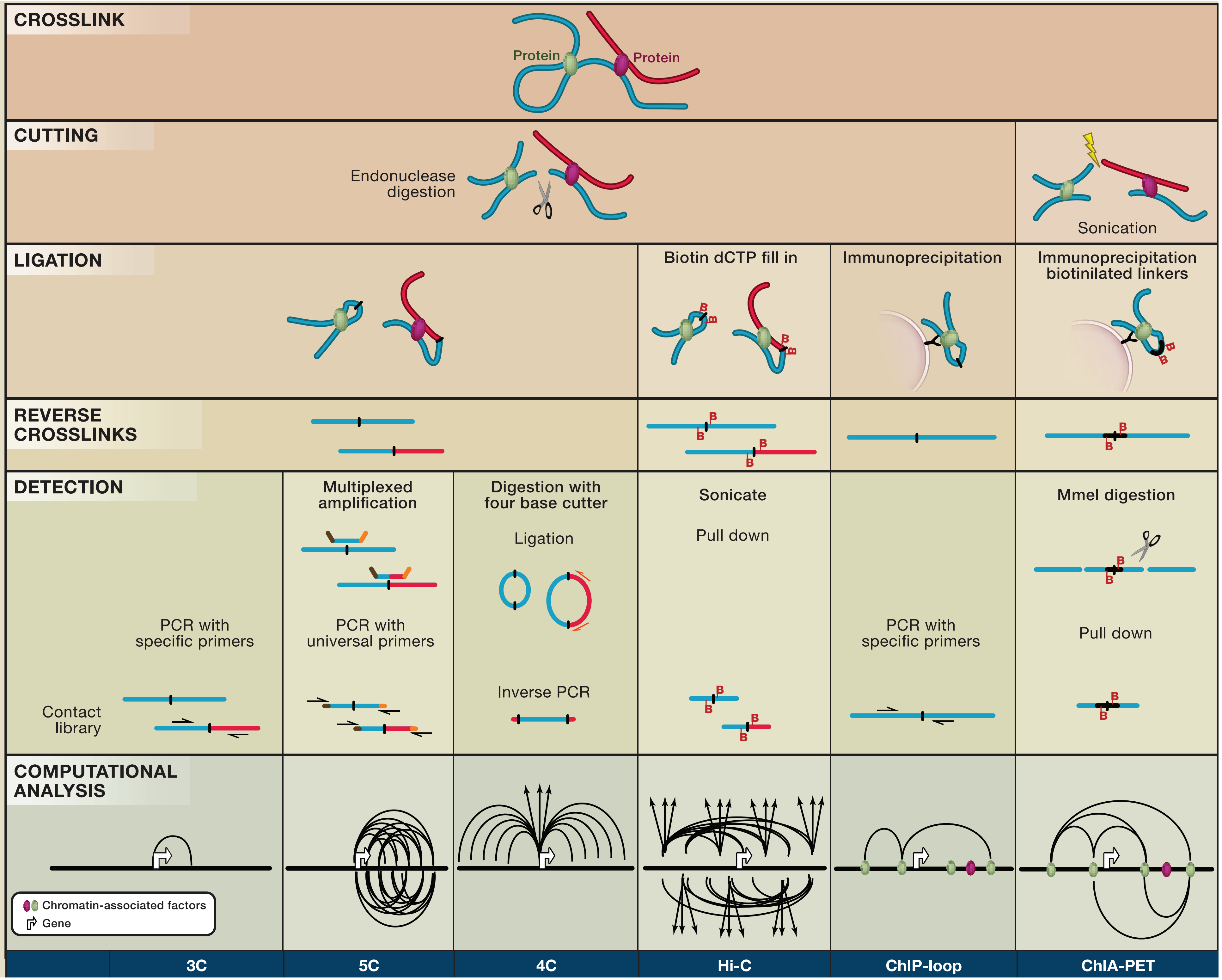
Ligate

Purify and shear DNA; pull down biotin

Sequence using paired-ends

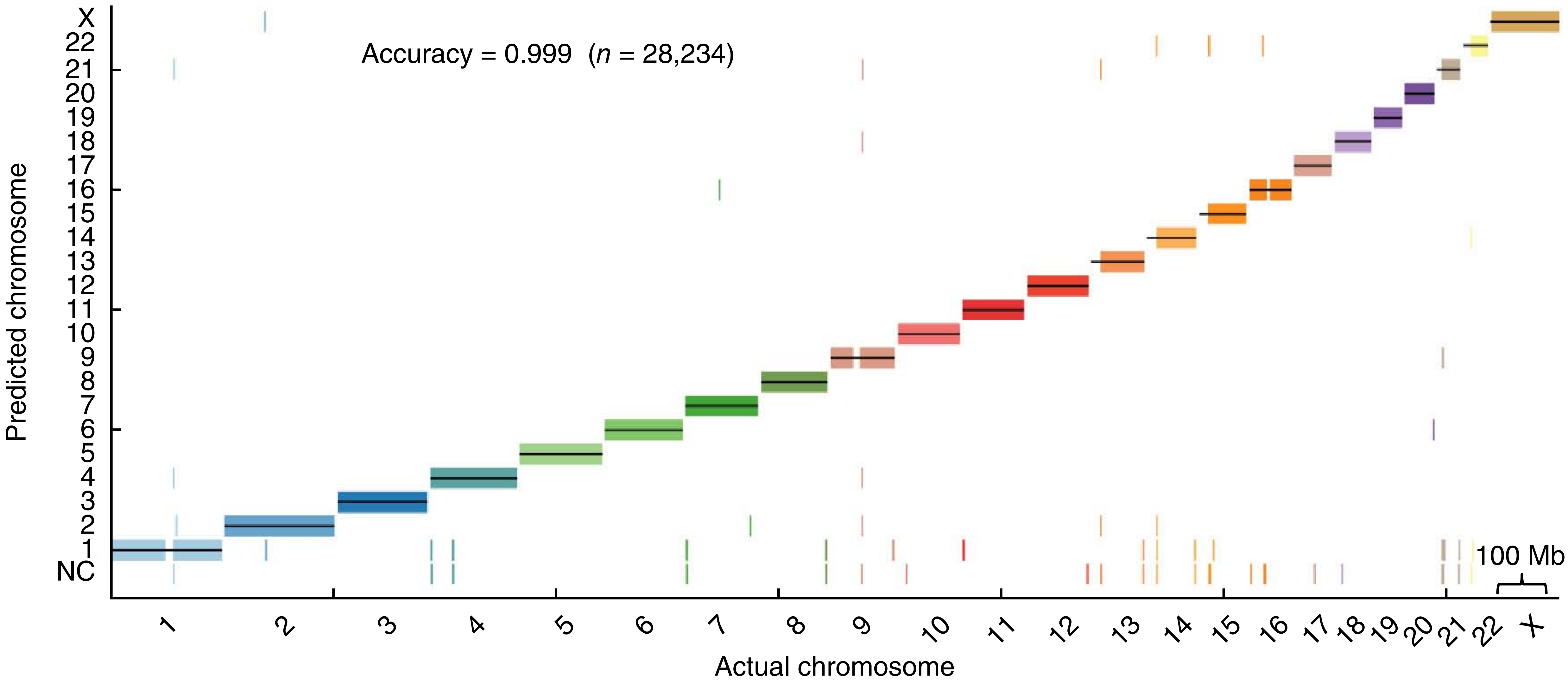


Chromosome Conformation Capture



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

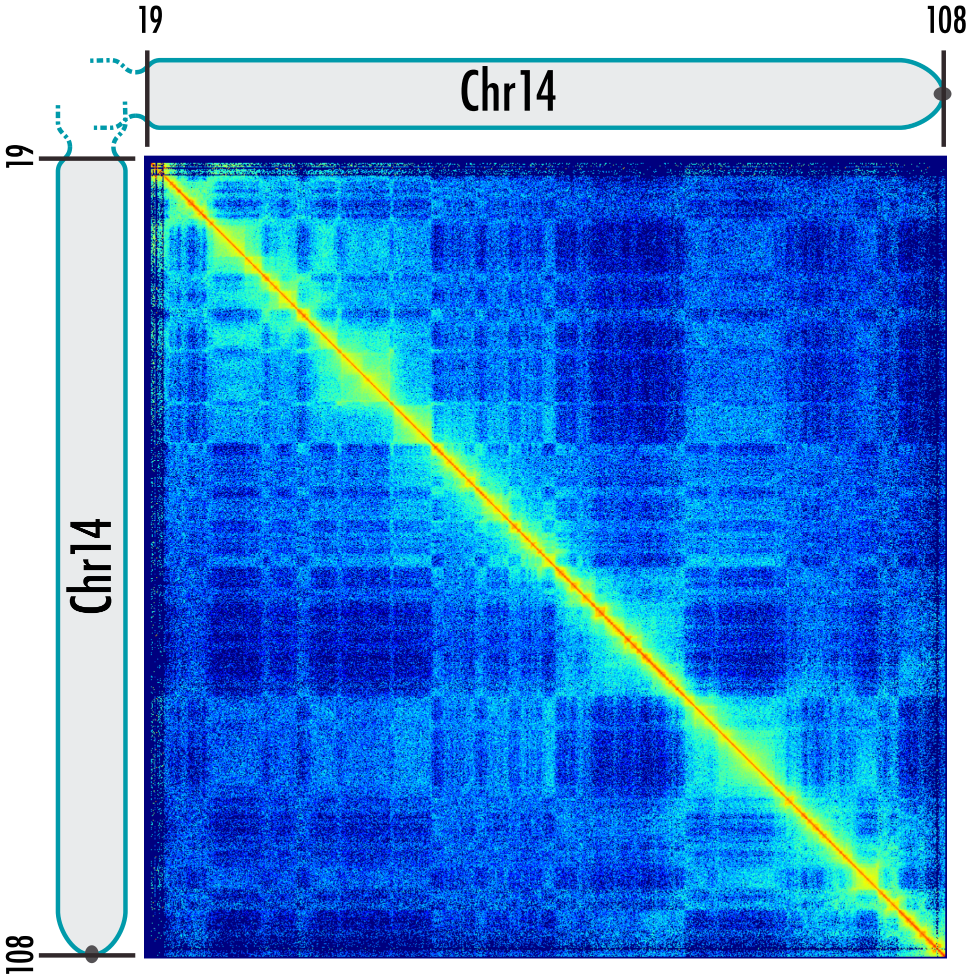
Chromosome Conformation Capture for de-novo assembly

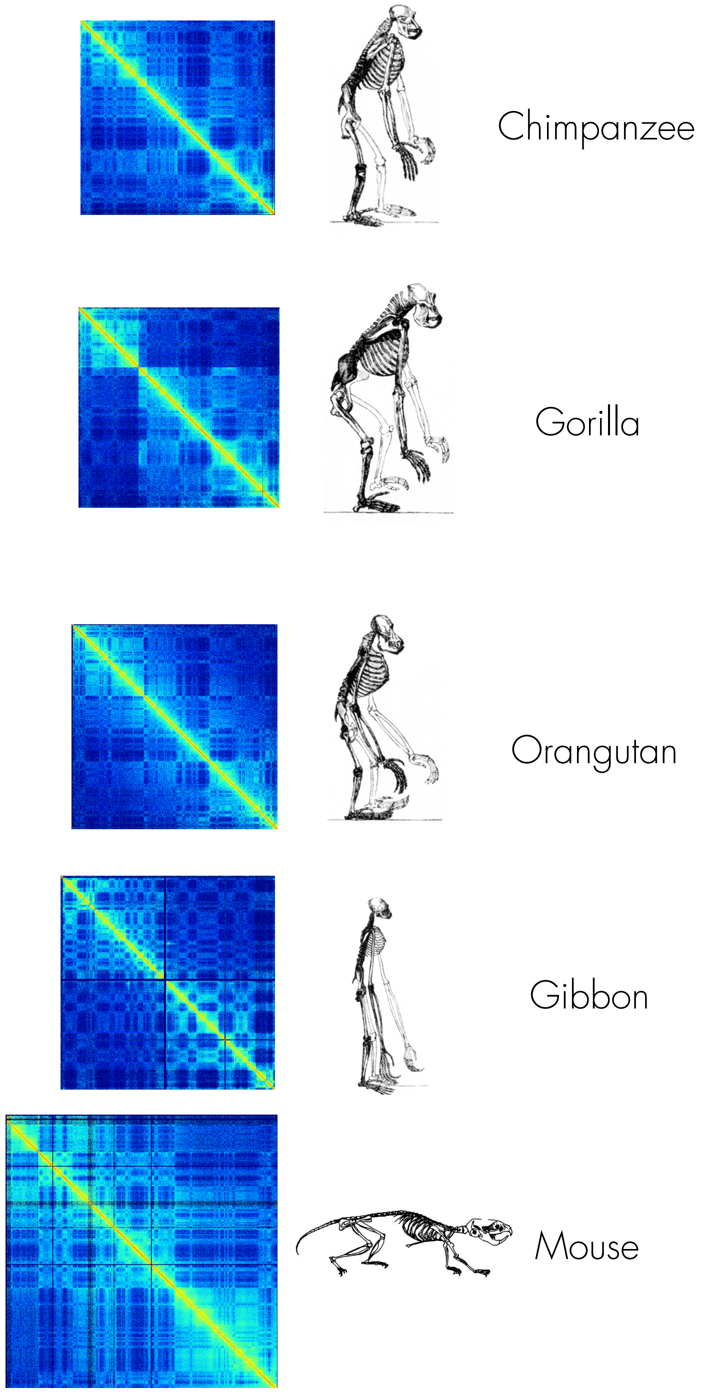


Kaplan, N., & Dekker, J. (2013). High-throughput genome scaffolding from in vivo DNA interaction frequency. Nature Biotechnology, 31(12), 1143–1147.

Great apes lymphoblast maps

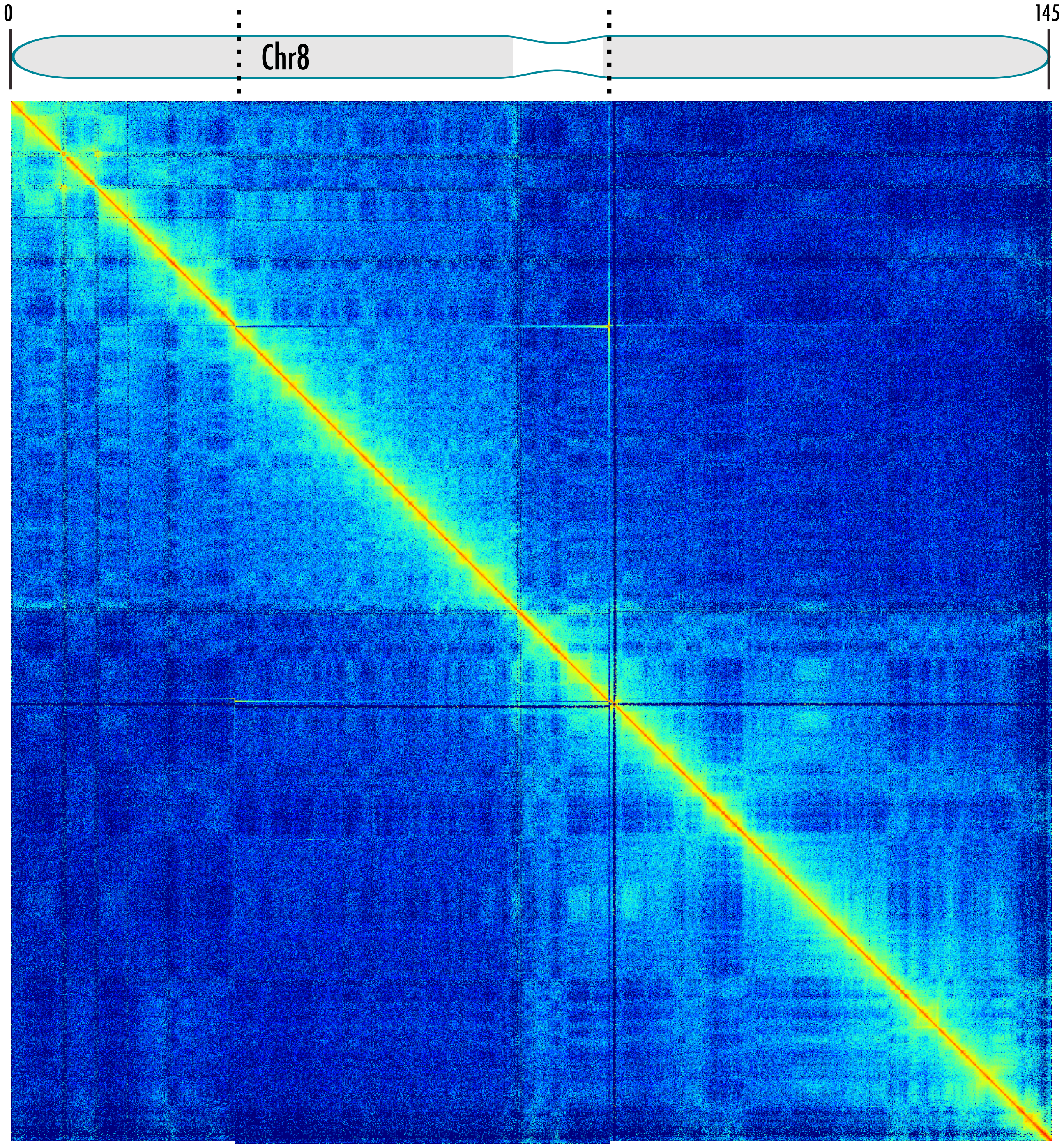
Chromosome 14



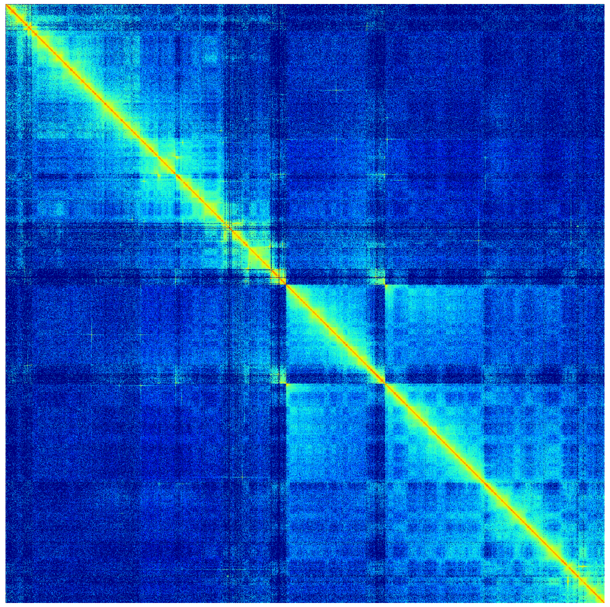


Assembly error detection

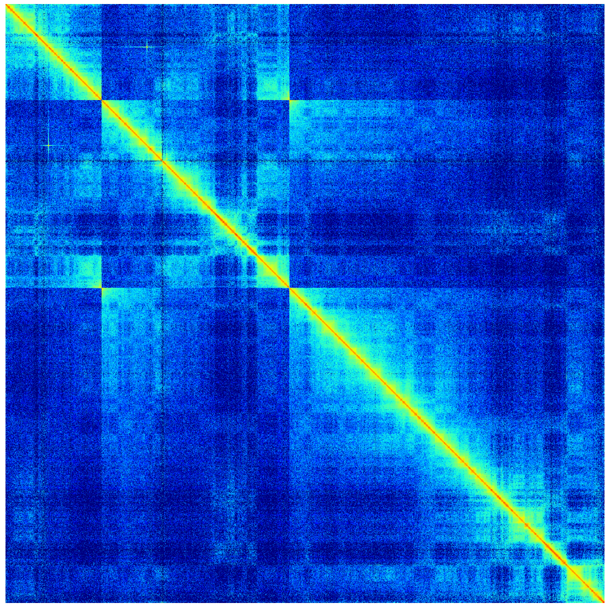
Chromosome 8 Gorilla



Chr 7

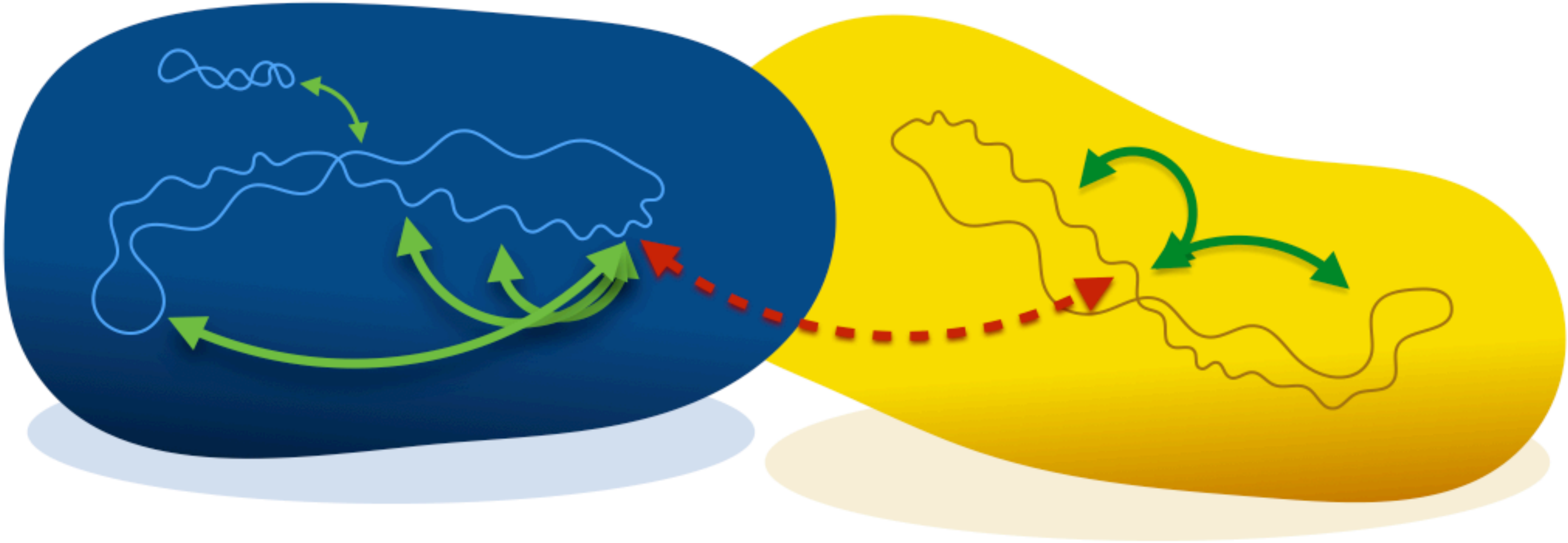


Chr 12



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

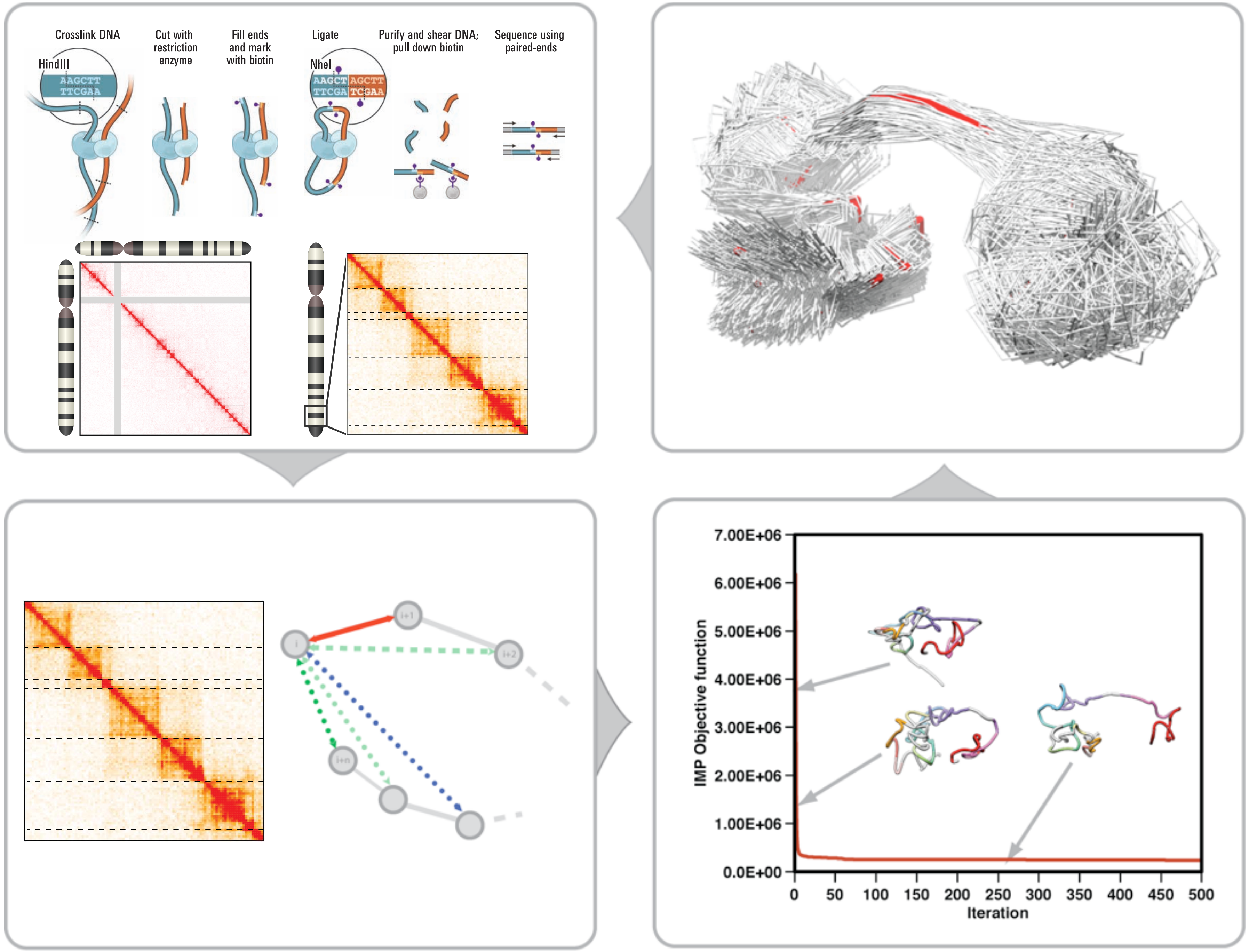


Beitel, C. W., Froenicke, L., Lang, J. M., Korf, I. F., Michelmore, R. W., Eisen, J. A., & Darling, A. E. (2014). Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. doi:10.7287/peerj.preprints.260v1

Hybrid Method

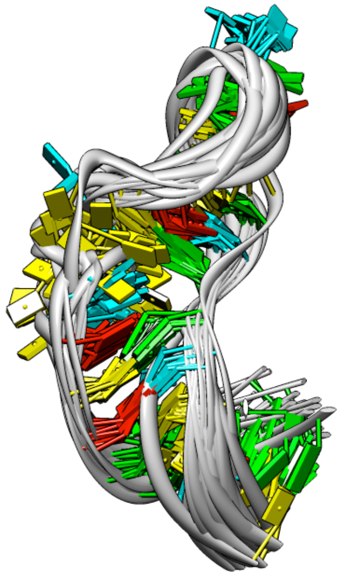
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).

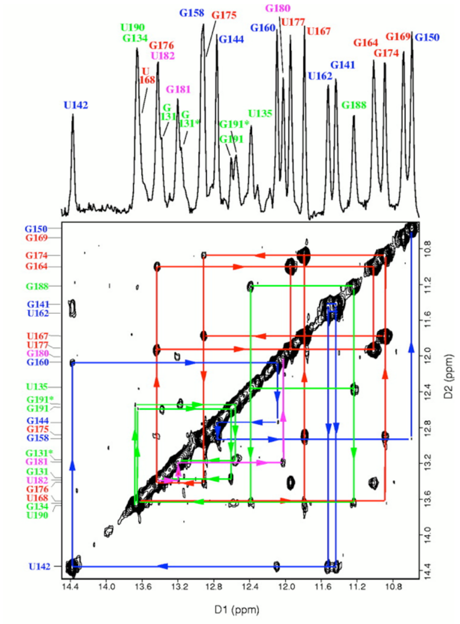
Experiments



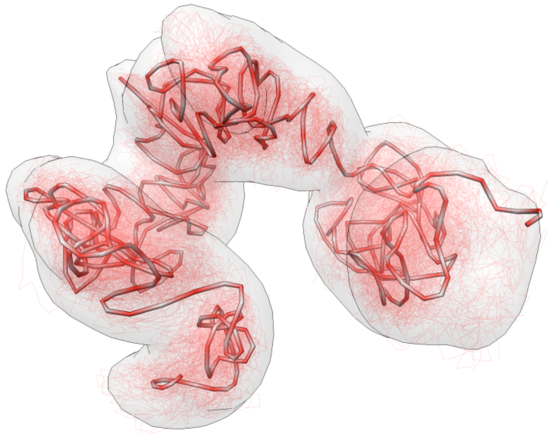
Computation

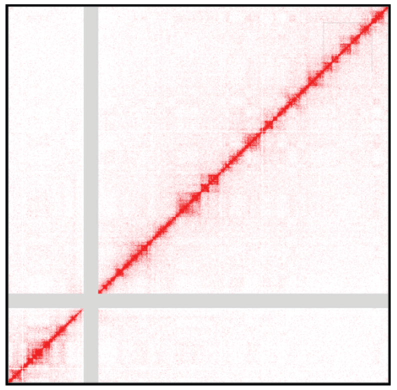
Structure determination using Hi-C data





Biomolecular structure determination
2D-NOESY data

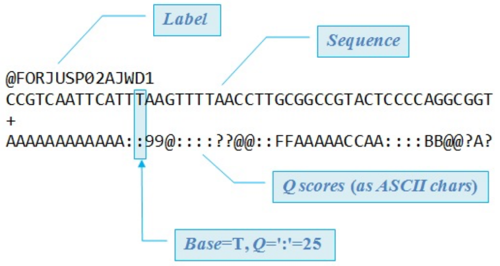




Chromosome structure determination
3C-based data

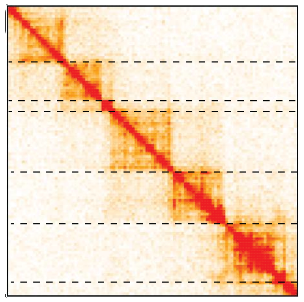
TADbit

http://3DGenomes.org

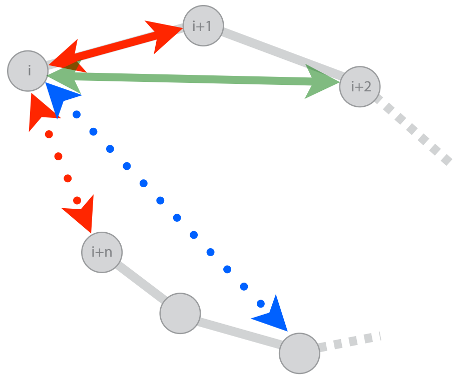


```
@FORJUSP02AJWD1
CCGTCAATTCATTTAAGTTTTAACCTTGCGGCCGTACTCCCCAGGCGGT
+
AAAAAAAAAAAA::99@::::??@@::FFAAAAACCAA::::BB@@?A?
```

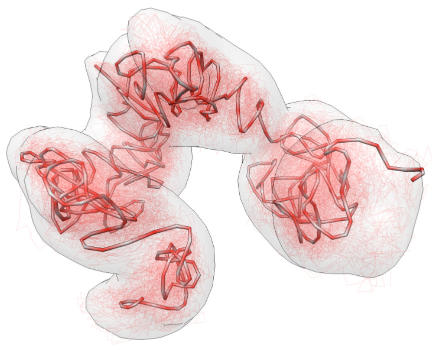
FastQ files to Maps



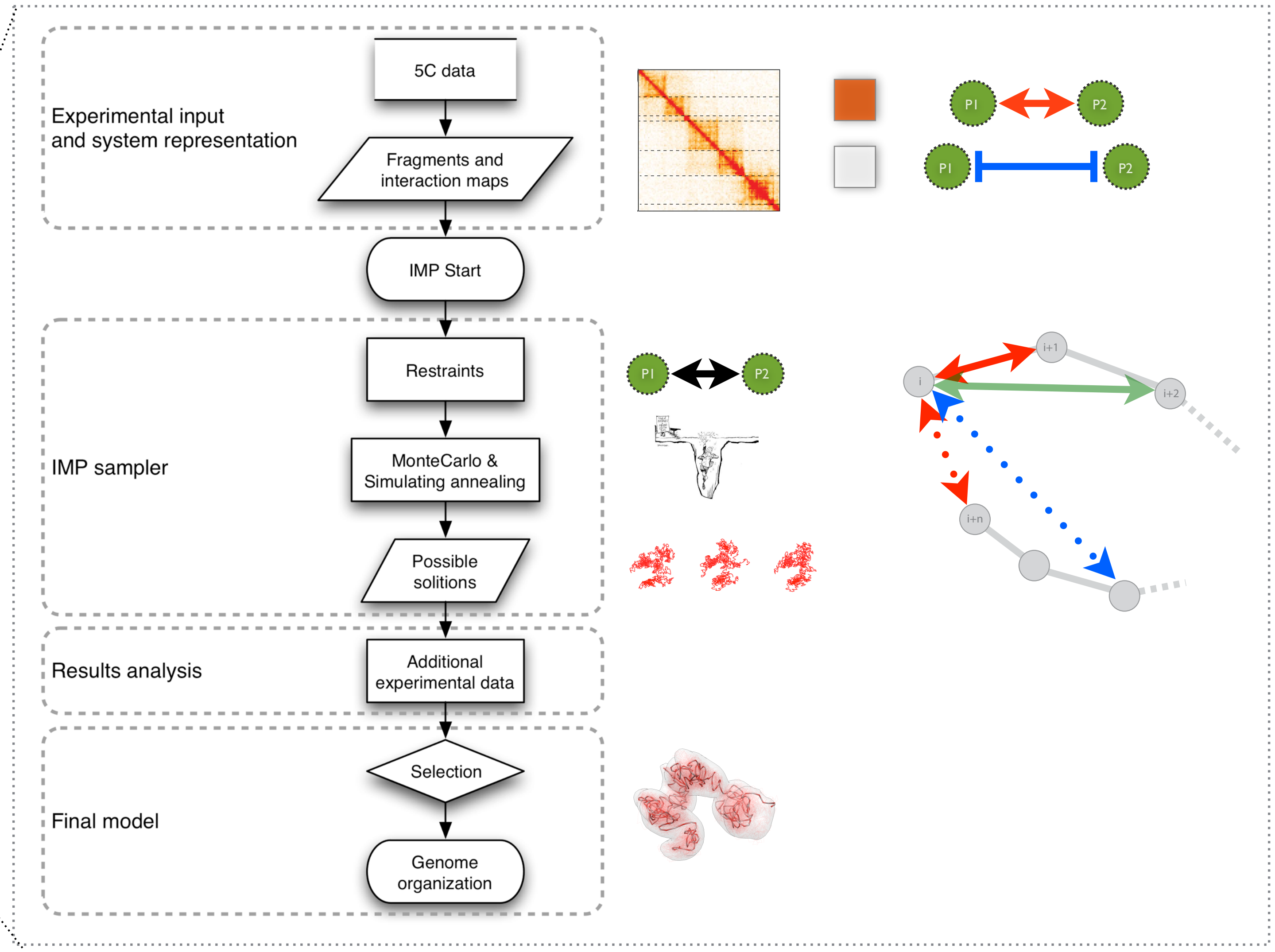
Map analysis



Model building



Model analysis



Experimental input and system representation

5C data

Fragments and interaction maps

IMP Start

IMP sampler

Restraints

MonteCarlo & Simulating annealing

Possible solitions

Results analysis

Additional experimental data

Final model

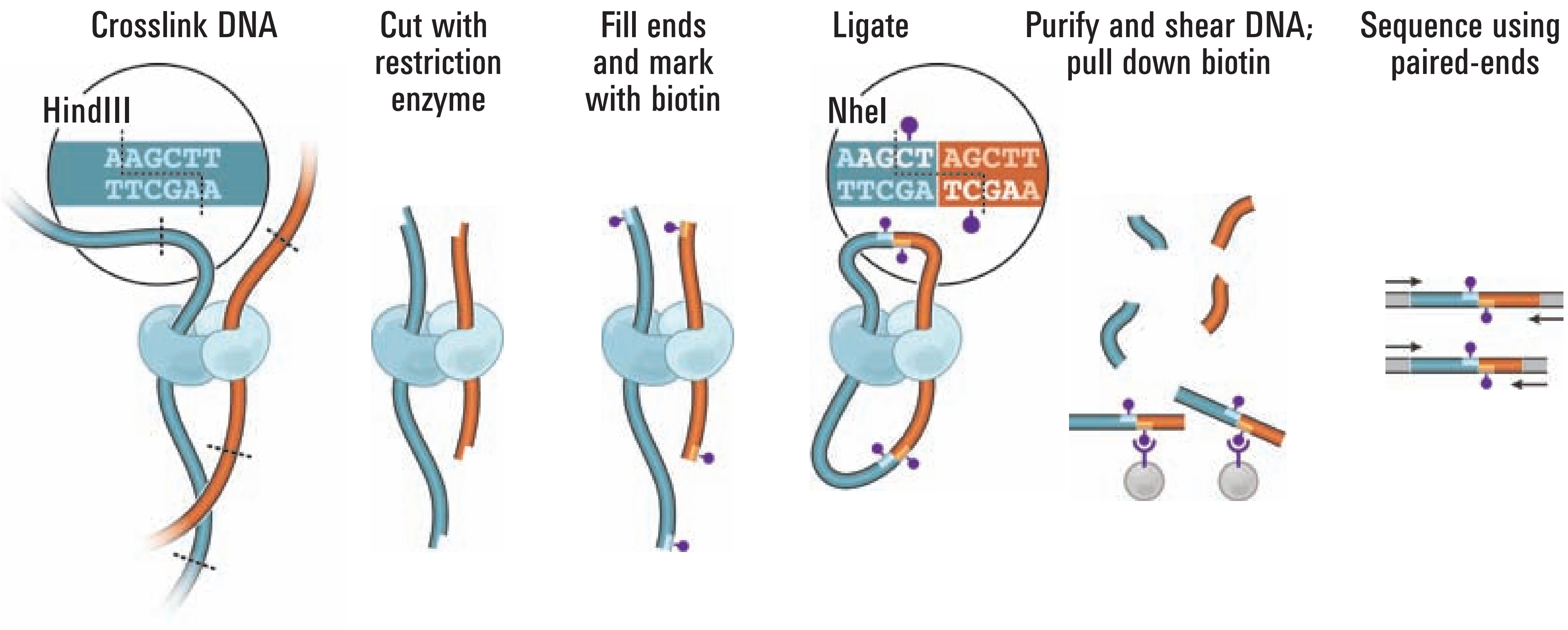
Selection

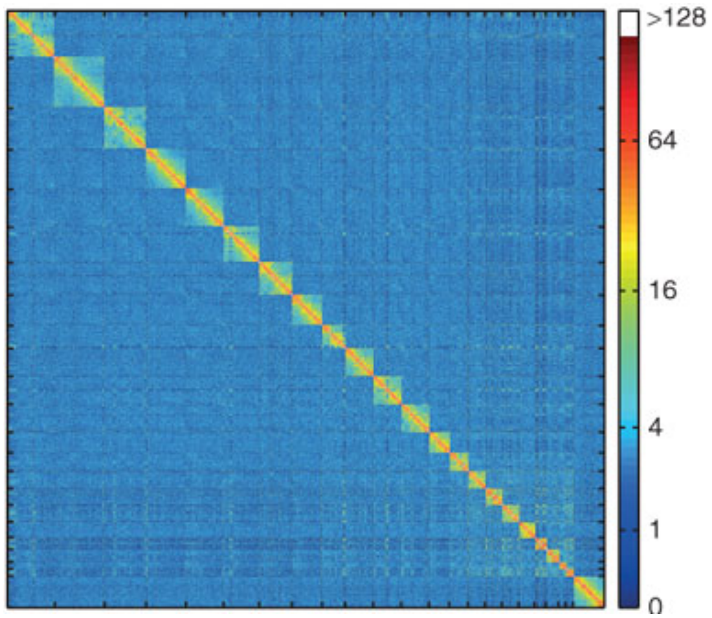
Genome organization


Got FASTQ?

Hi-C experiment

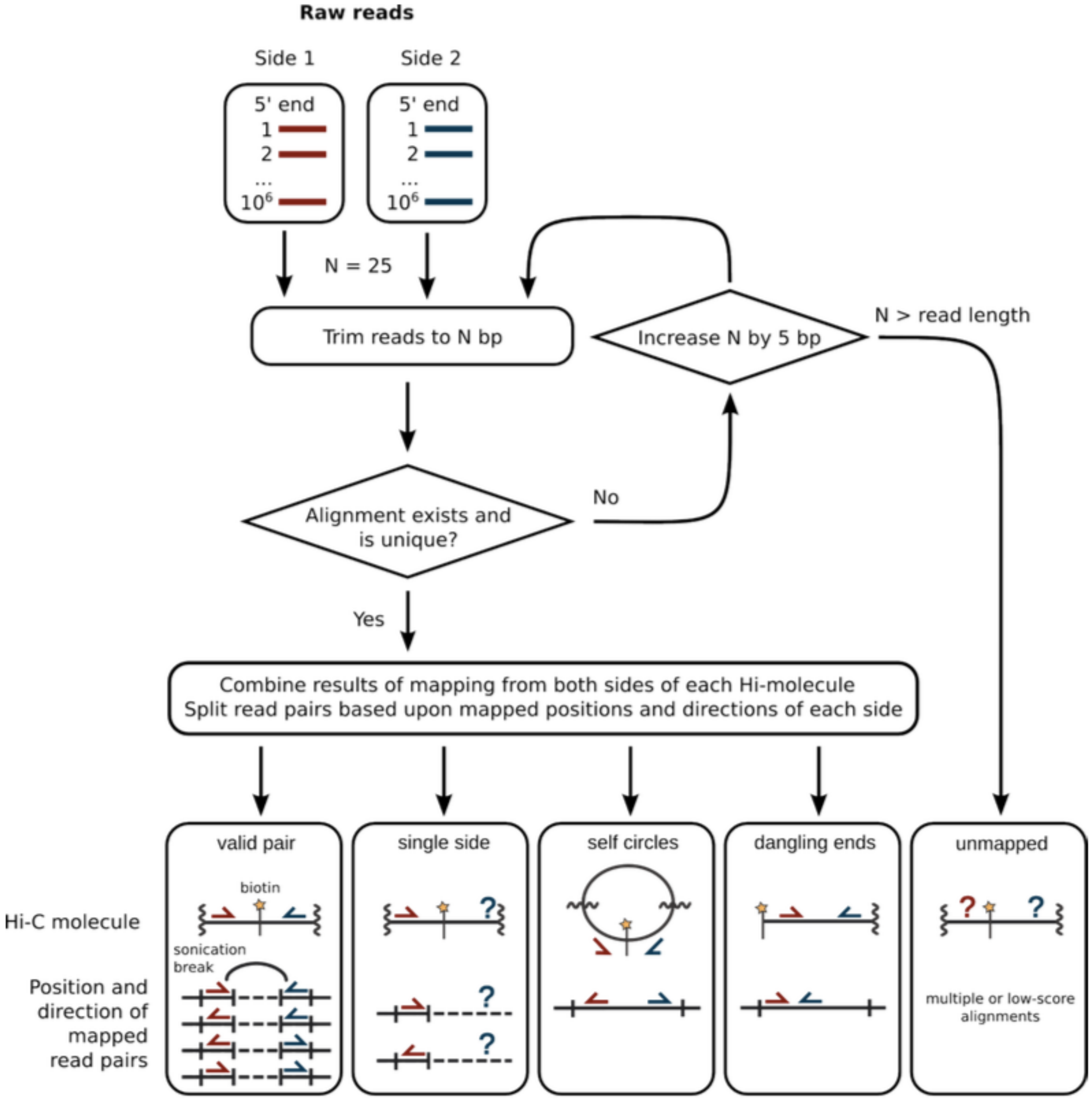
Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.





Mapping & Filtering

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

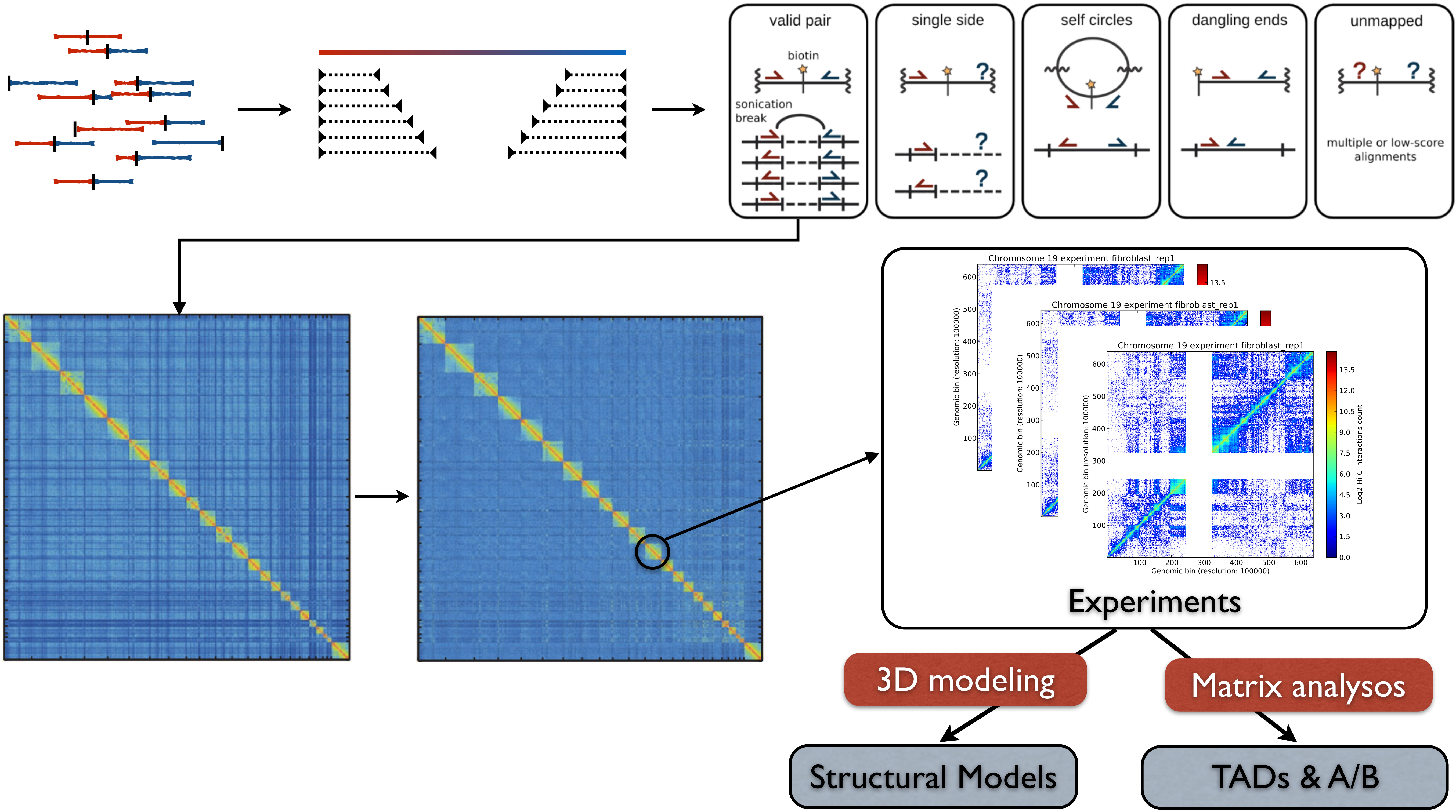


How much you normally map?

- 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



Experiments

3D modeling

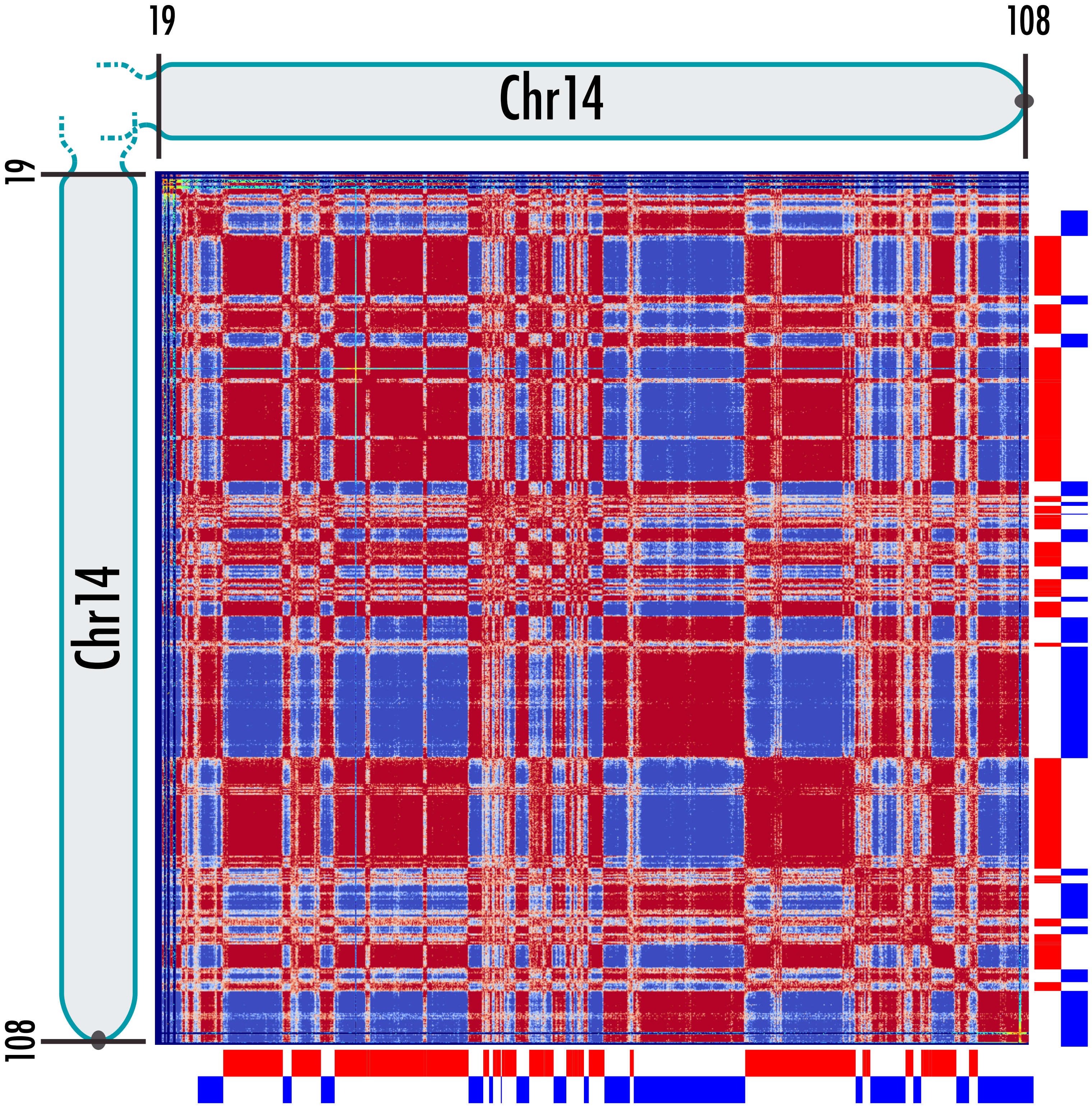
Matrix analysos

Structural Models

TADs & A/B

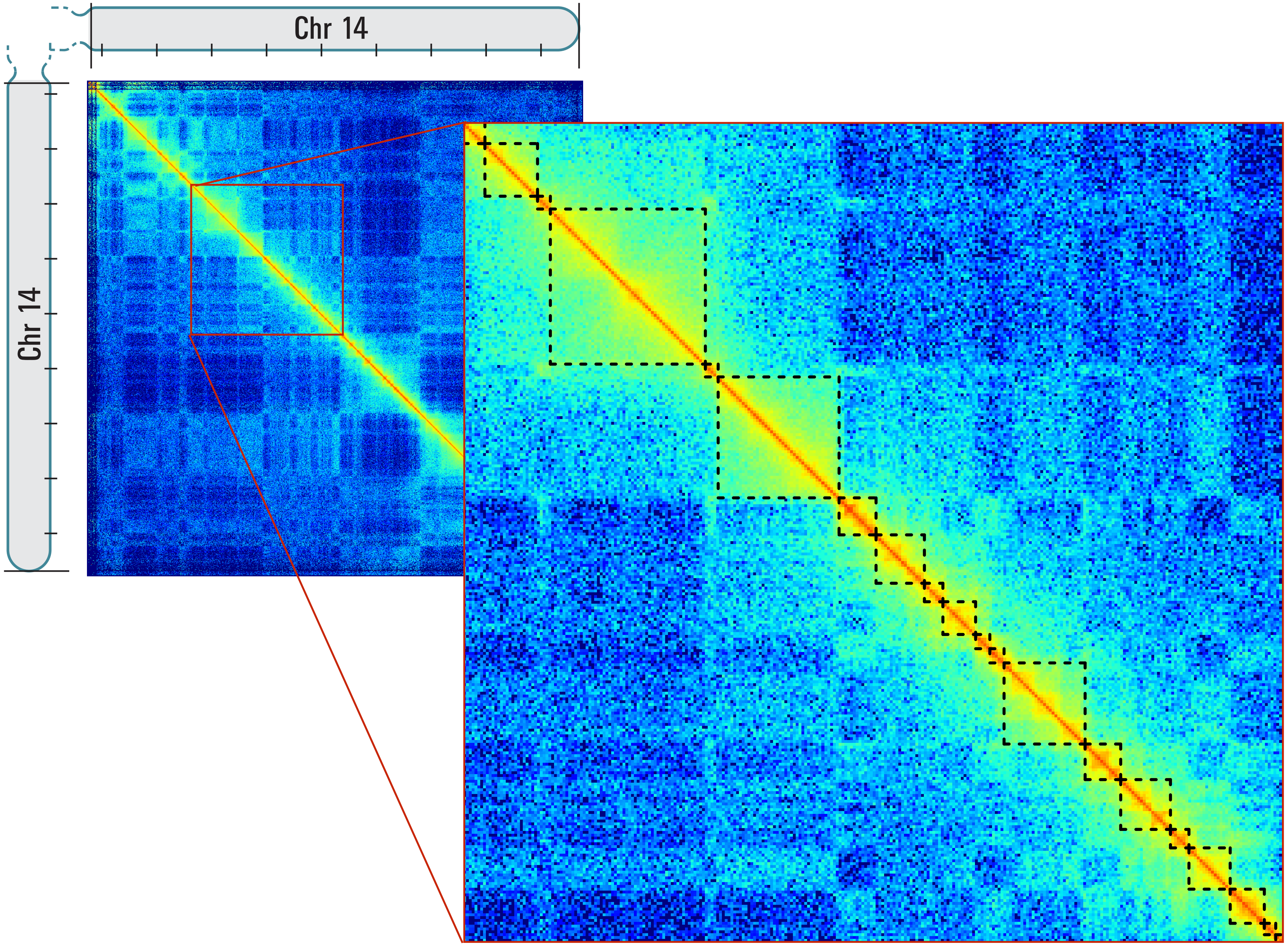
A/B Compartment

Chromosome 14



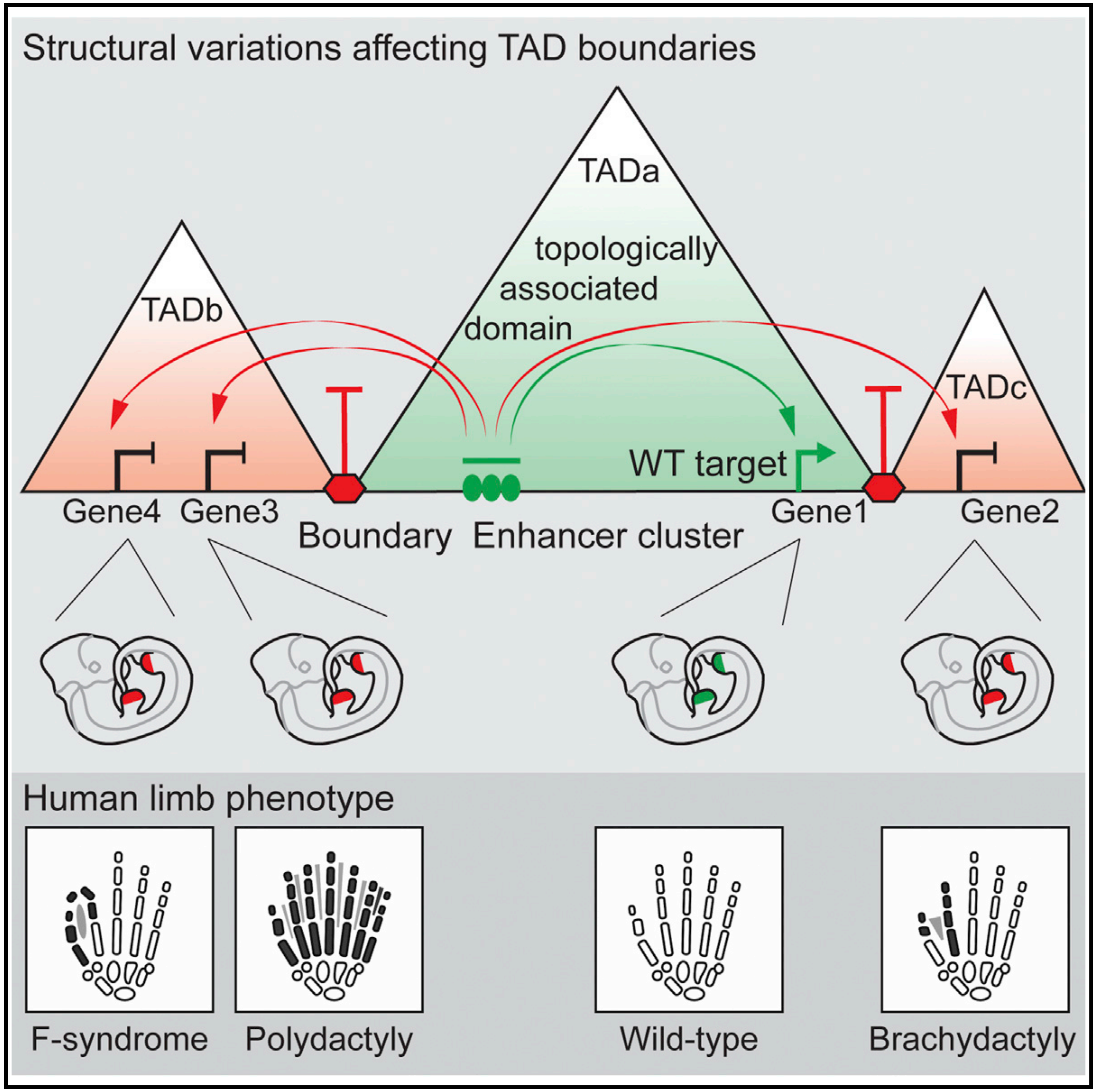
TADs

Chromosome 14



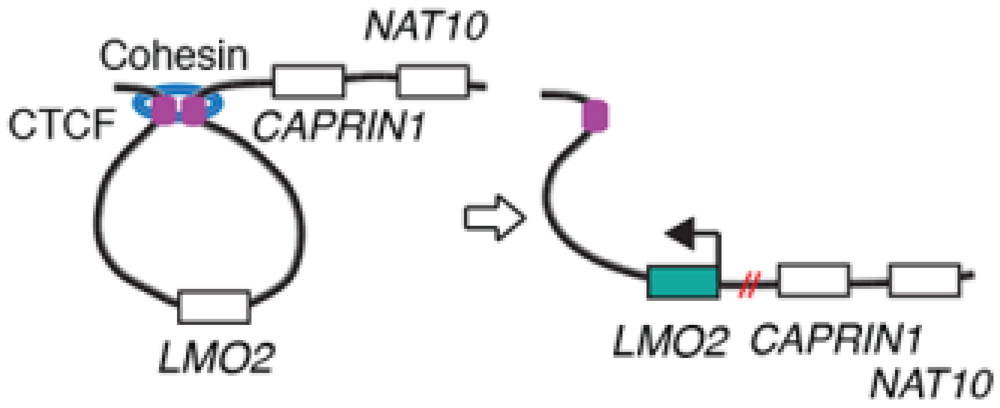
TADs are functional units

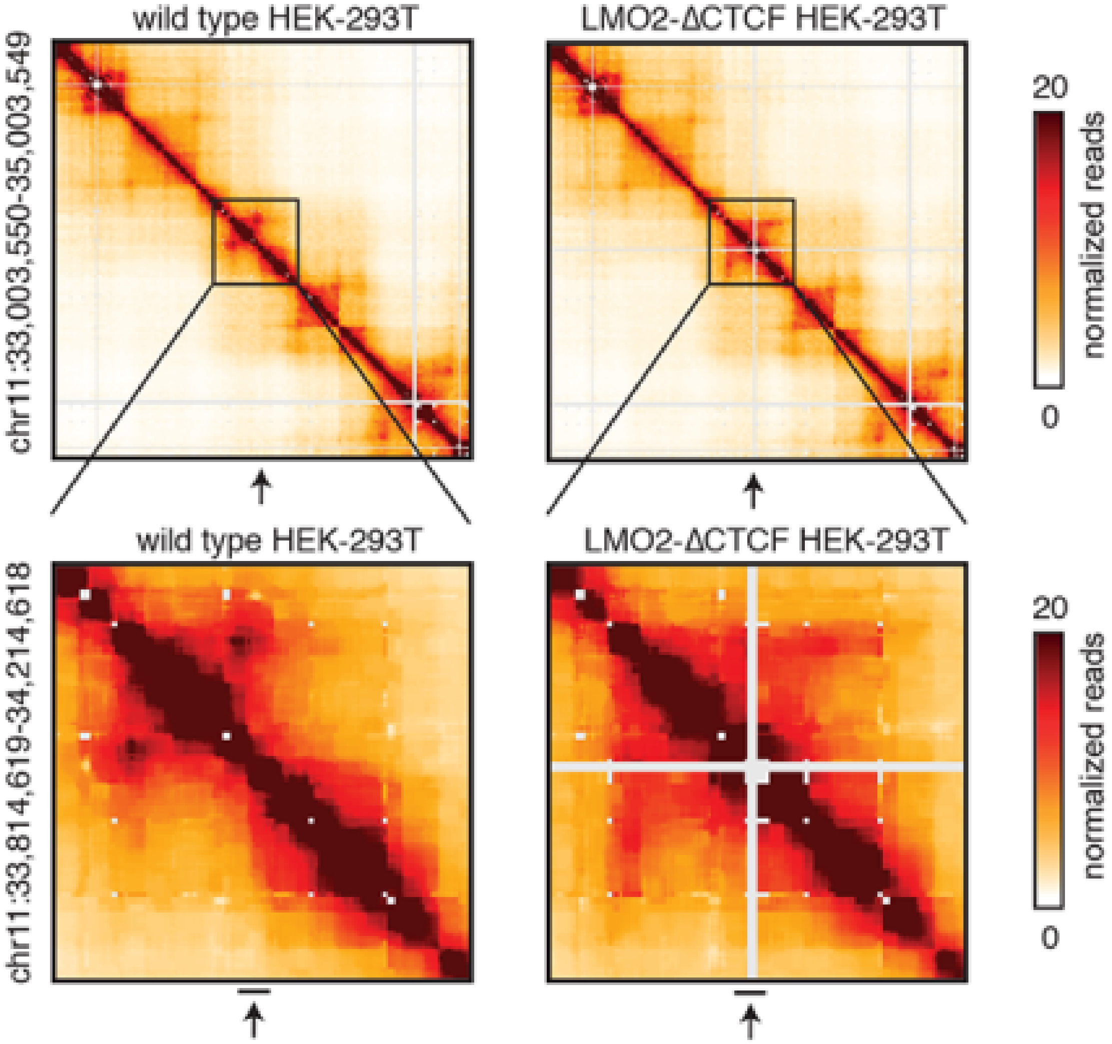
Lupiáñez, et al. (2015). Cell, 1–15.



TADs are functional units

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

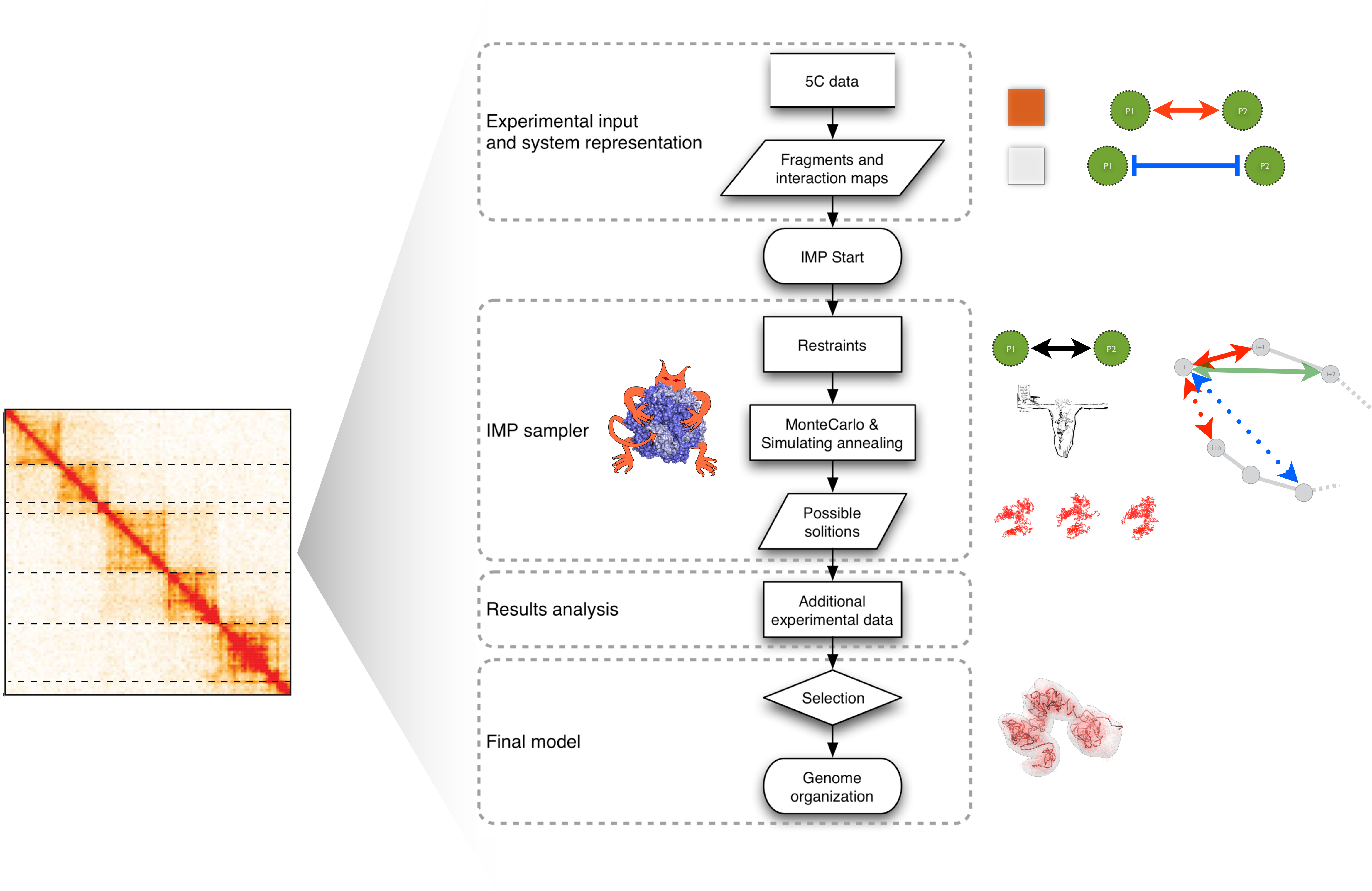





Got normalized Hi-C maps?

TADbit

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



Experimental input and system representation

5C data

Fragments and interaction maps

IMP Start

IMP sampler

Restraints

MonteCarlo & Simulating annealing

Possible solitions

Results analysis

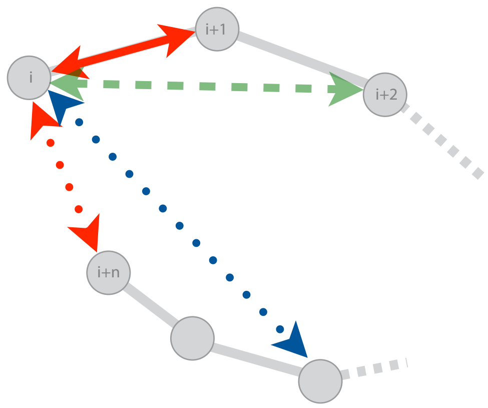
Additional experimental data

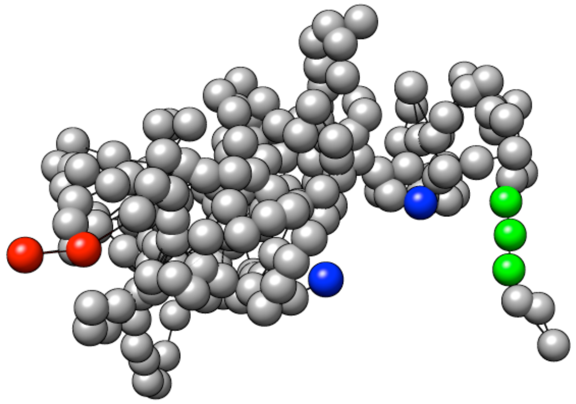
Final model

Selection

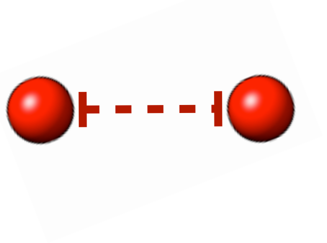
Genome organization

Model representation and scoring



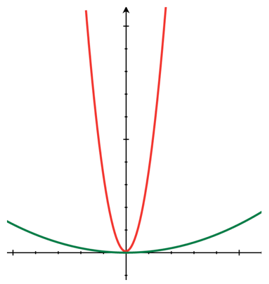


$d = d_0$

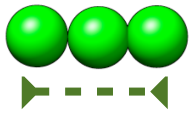


Harmonic

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

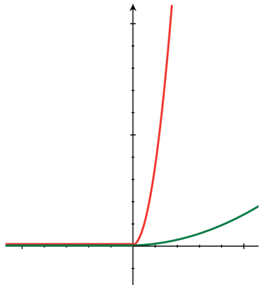


$d < d_0$

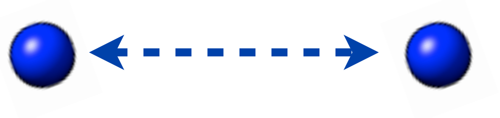


Harmonic Upper Bound

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

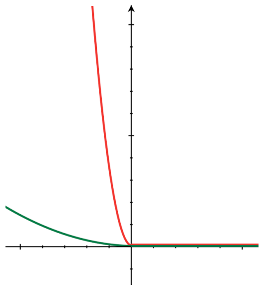


$d > d_0$

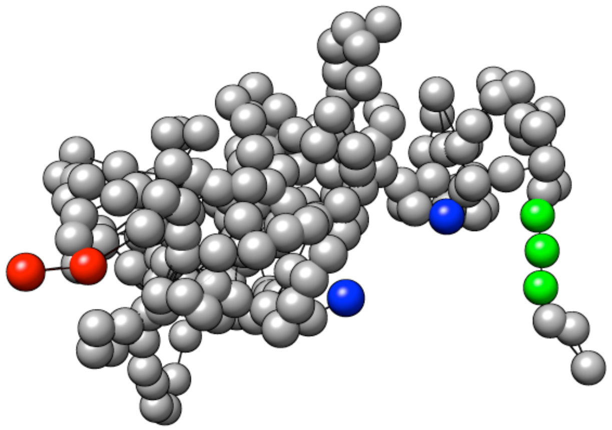


Harmonic Lower Bound

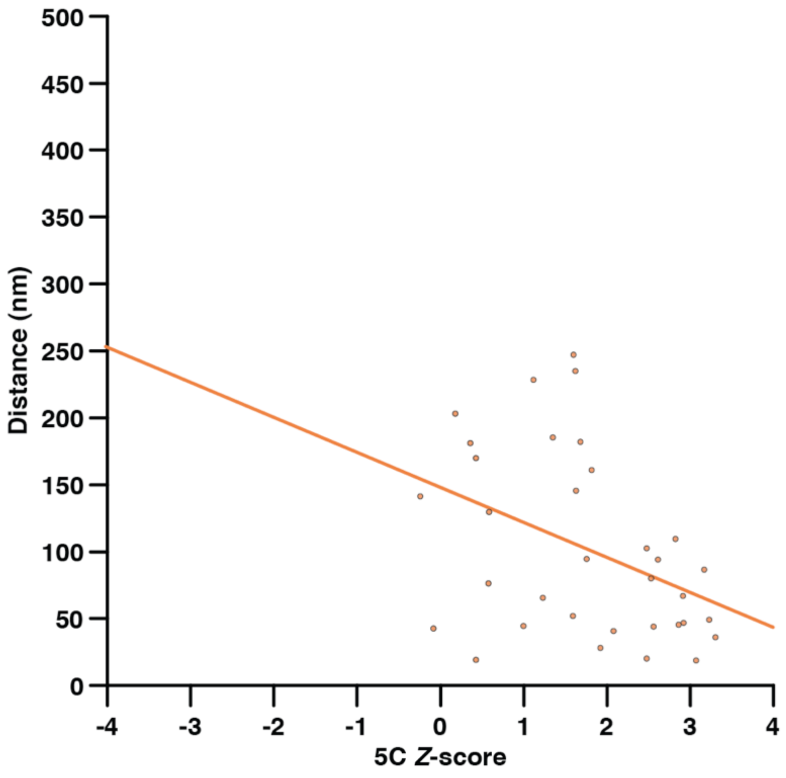
$$\begin{cases} if \;\; d_{i,j} \leq d_{i,j}^{0}; & lbH_{i,j} = k\left(d_{i,j} - d_{i,j}^{0}\right)^2 \\ if \;\; d_{i,j} > d_{i,j}^{0}; & lbH_{i,j} = 0 \end{cases}$$



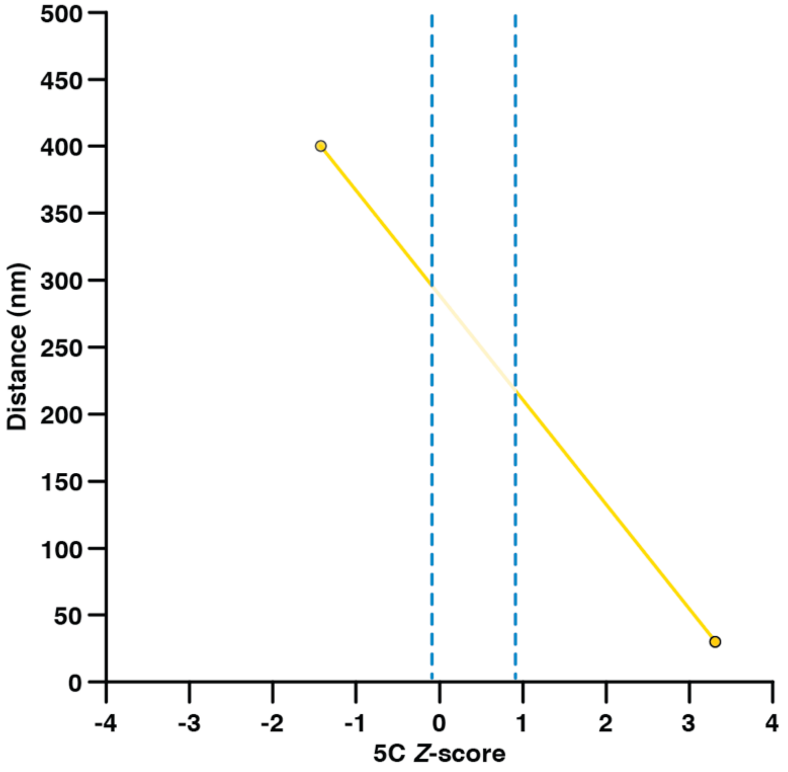
From 3C data to spatial distances



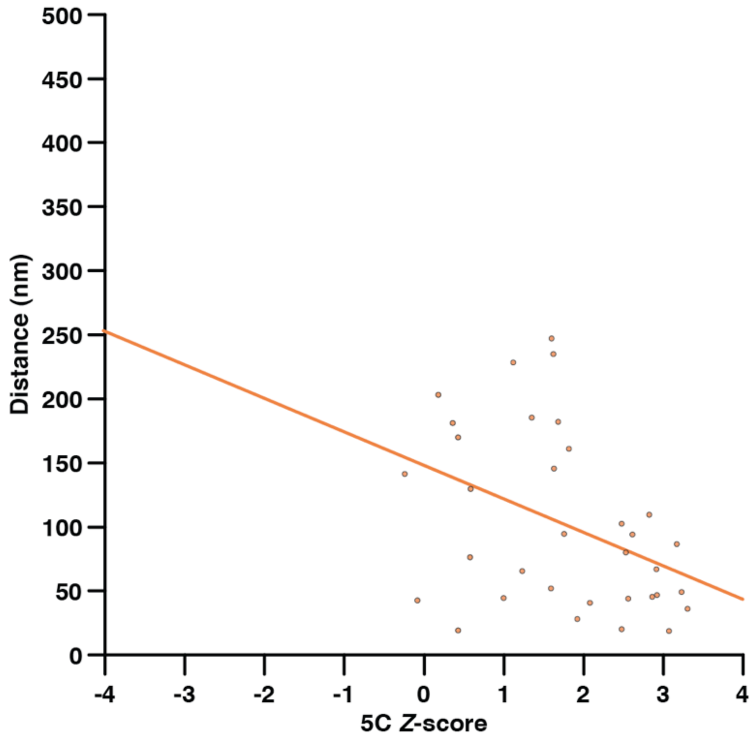
Neighbor fragments

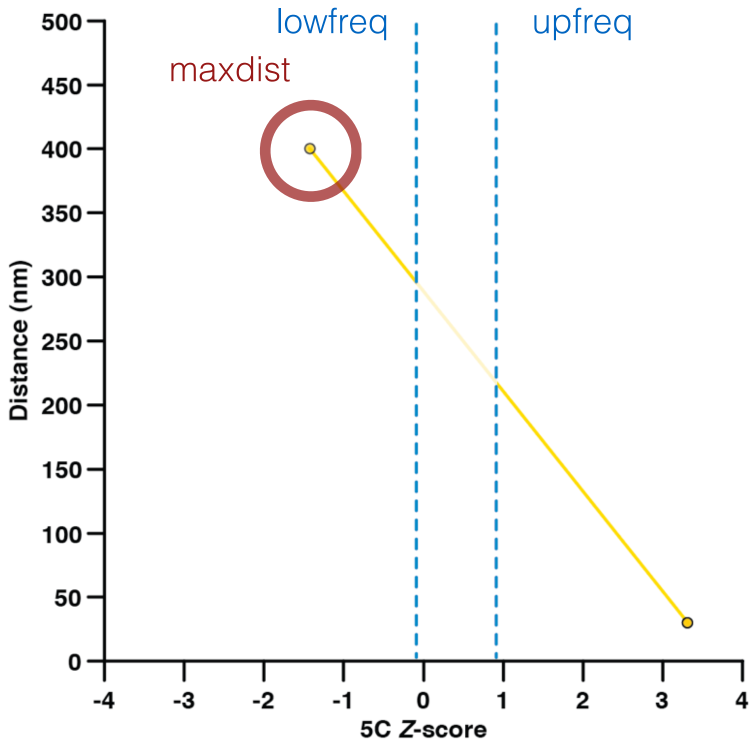


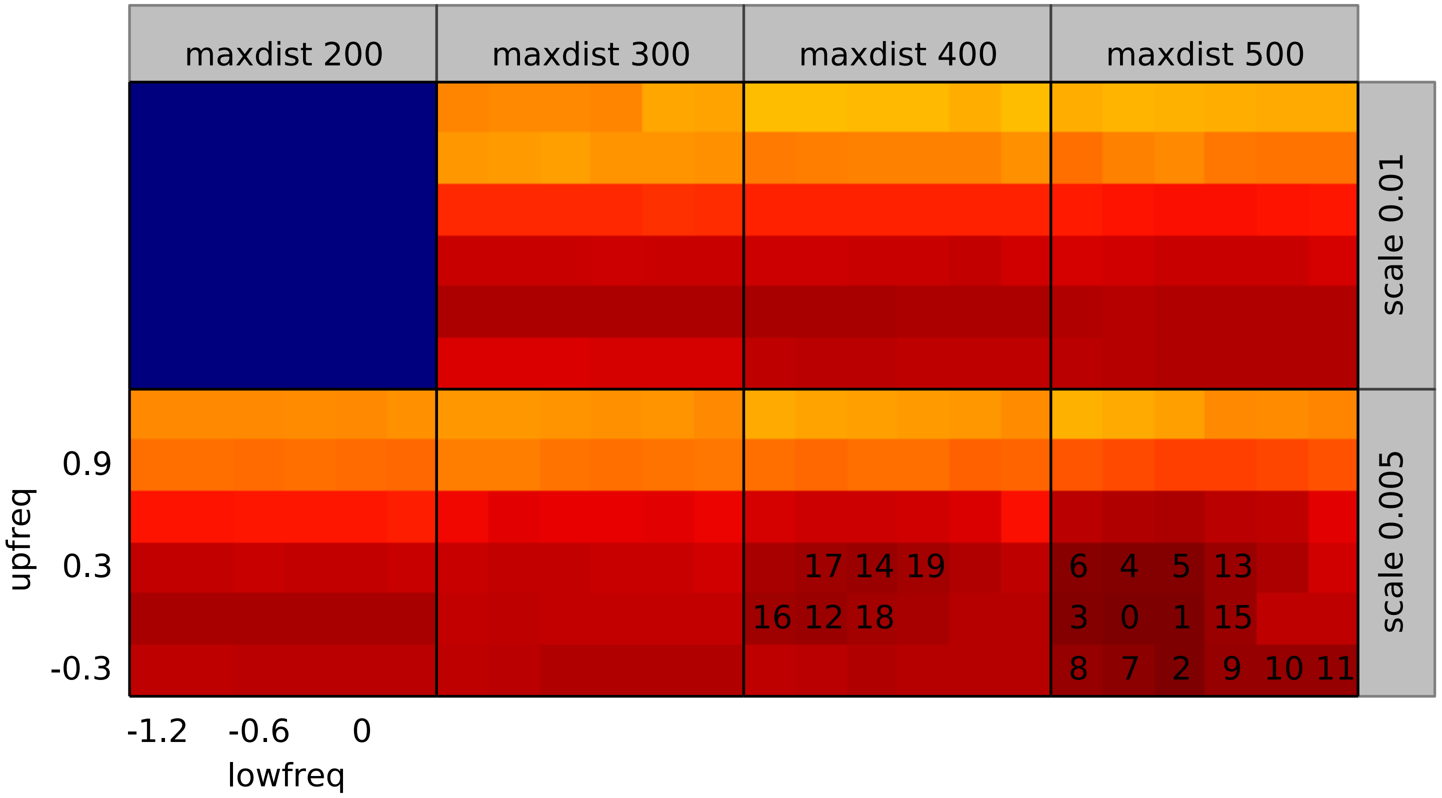
Non-Neighbor fragments



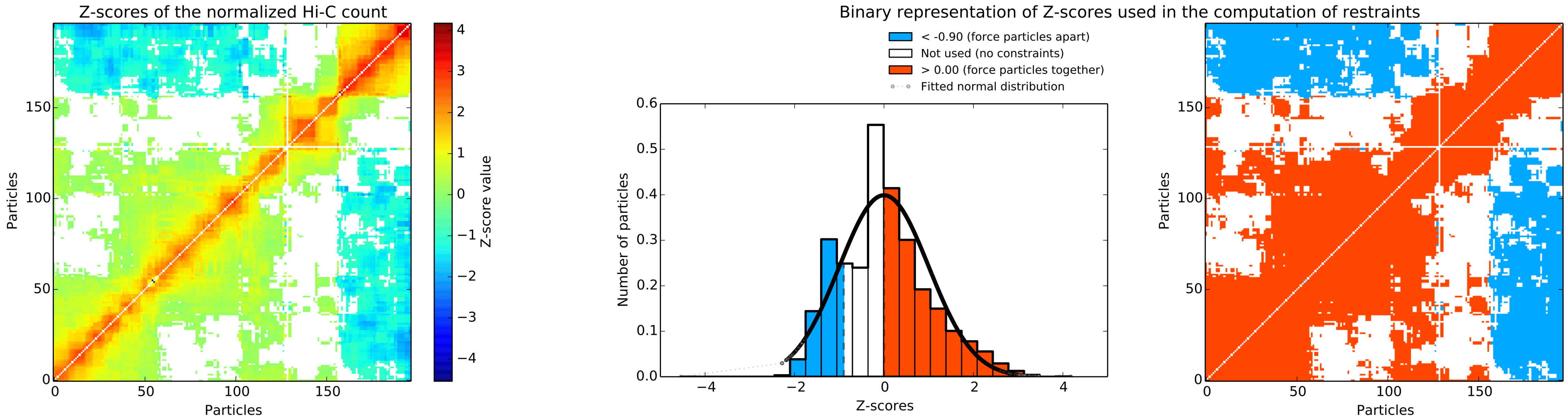
Parameter optimization

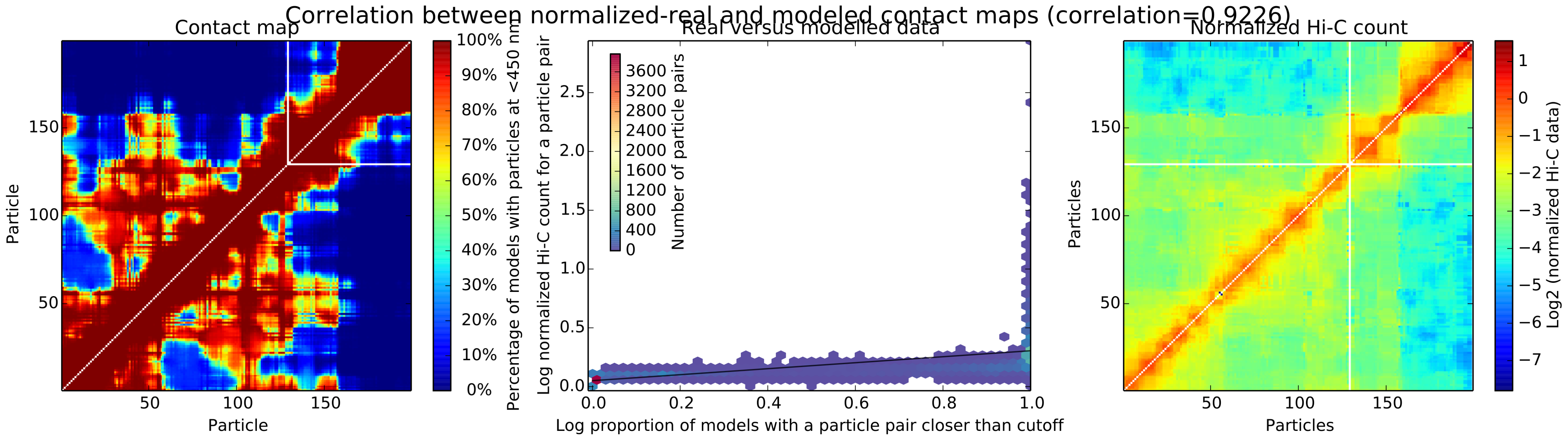




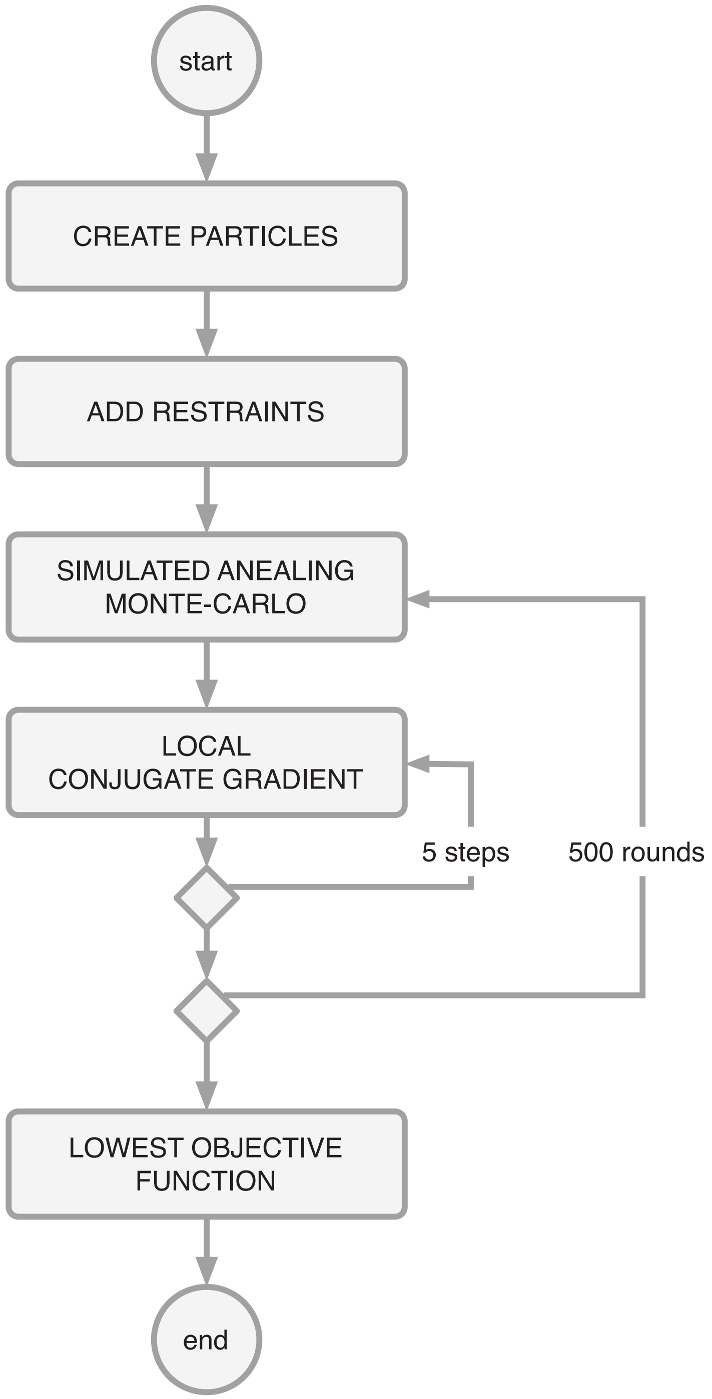


Parameter optimization




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

Optimization of the scoring function


start

CREATE PARTICLES

ADD RESTRAINTS

SIMULATED ANEALING MONTE-CARLO

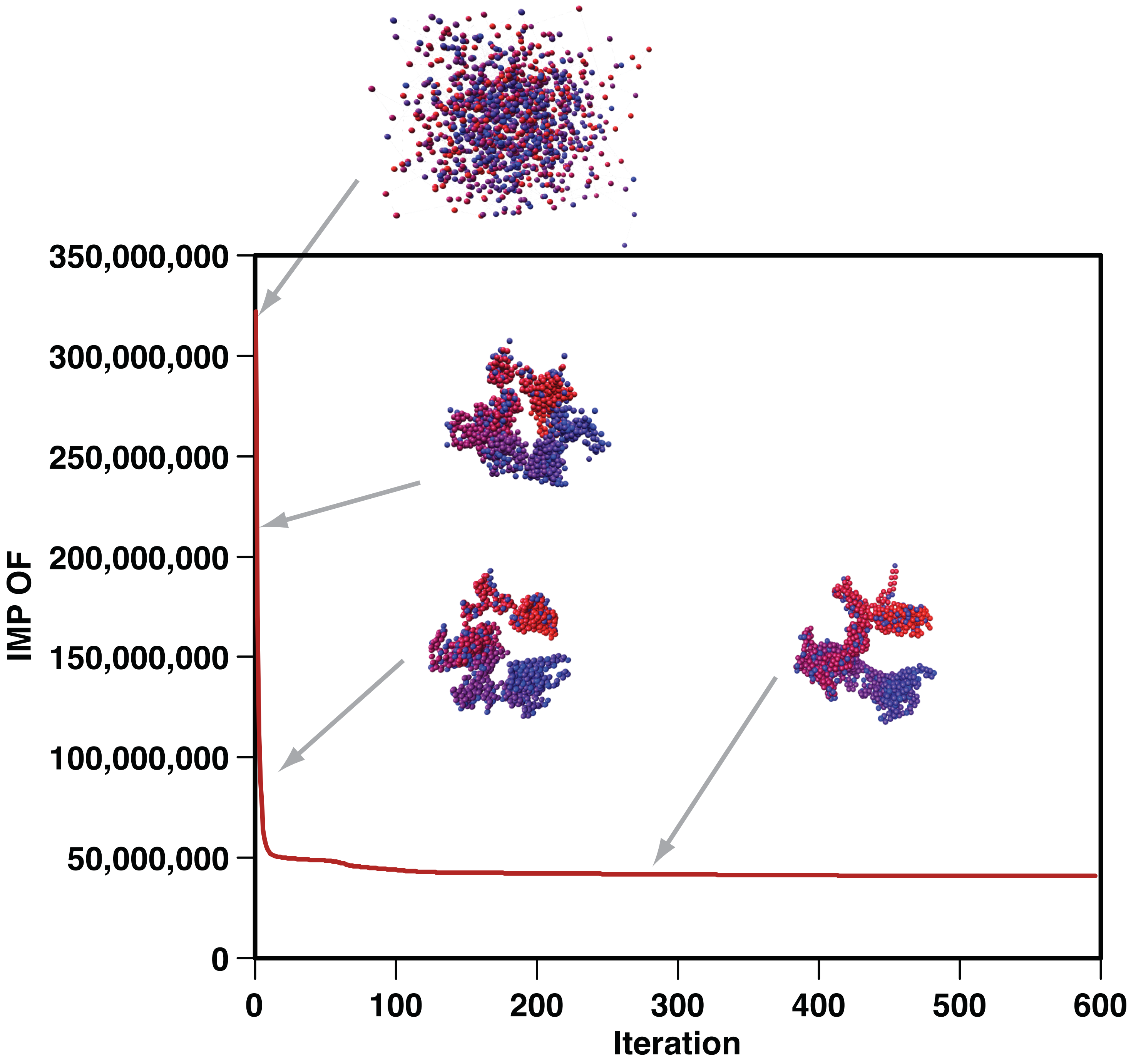
LOCAL CONJUGATE GRADIENT

5 steps

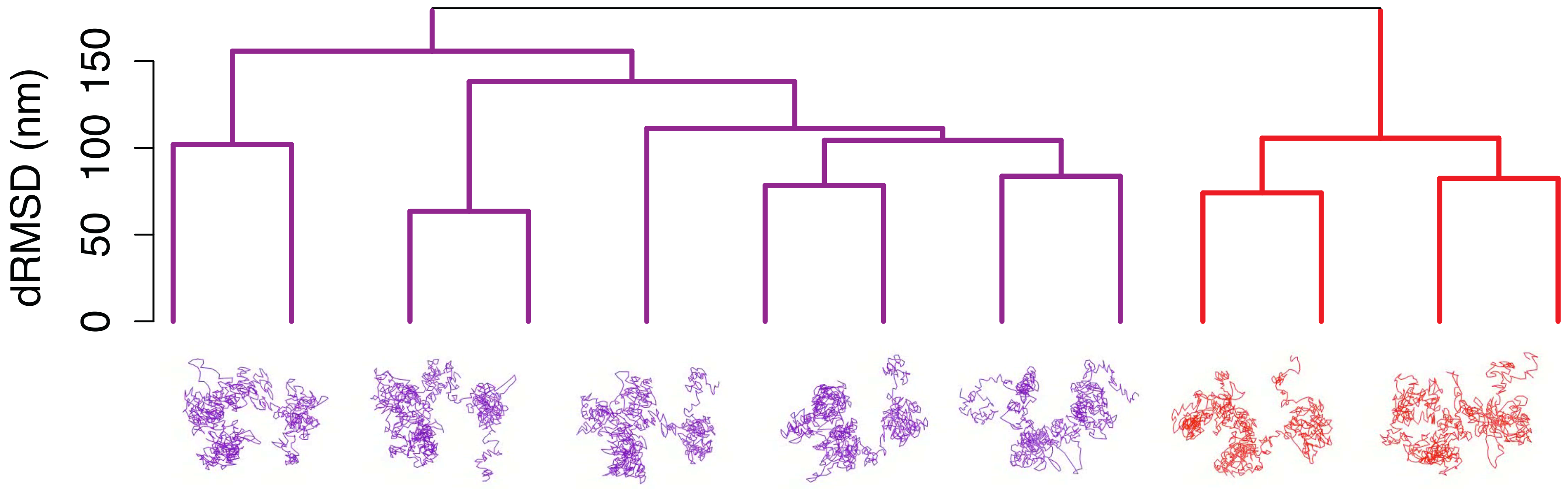
500 rounds

LOWEST OBJECTIVE FUNCTION

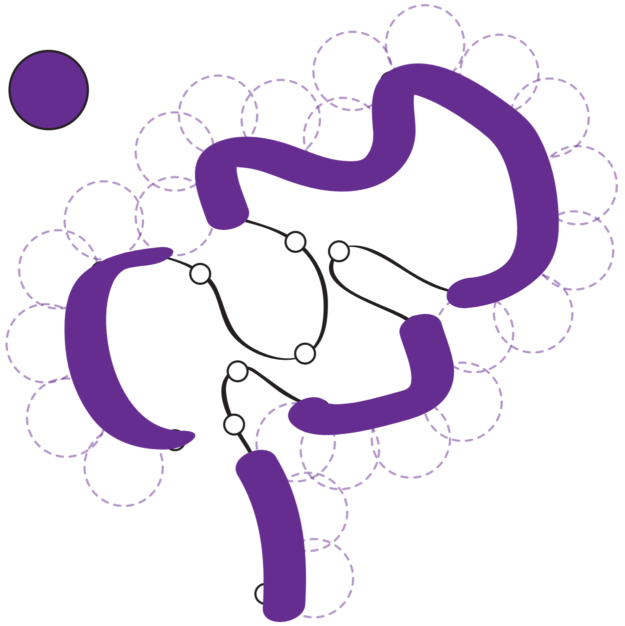
end



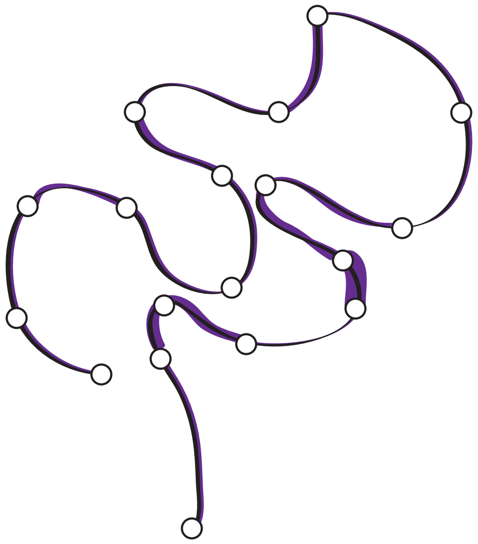
Model analysis: clustering and structural features



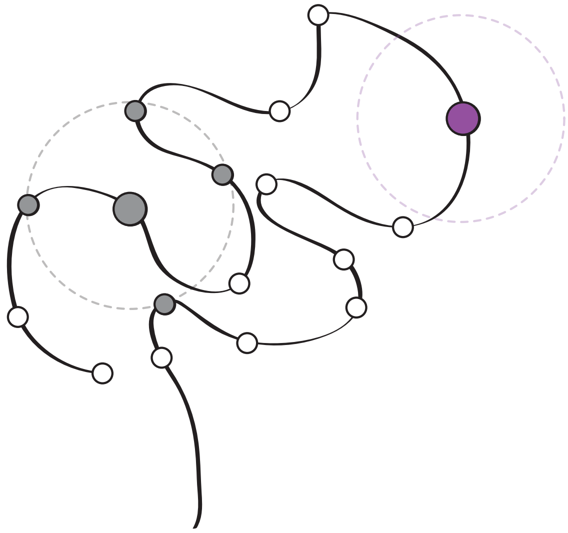
Accessibility (%)



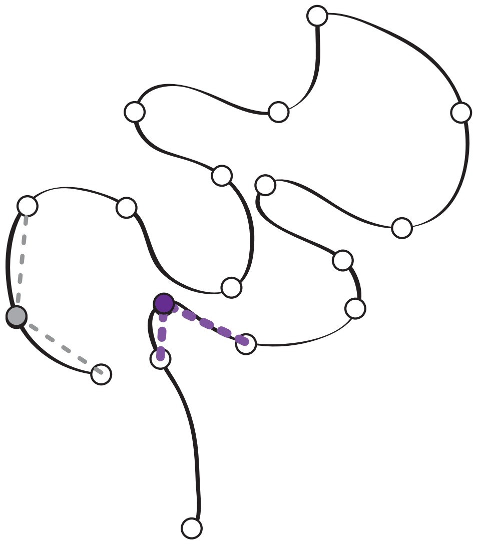
Density (bp/nm)

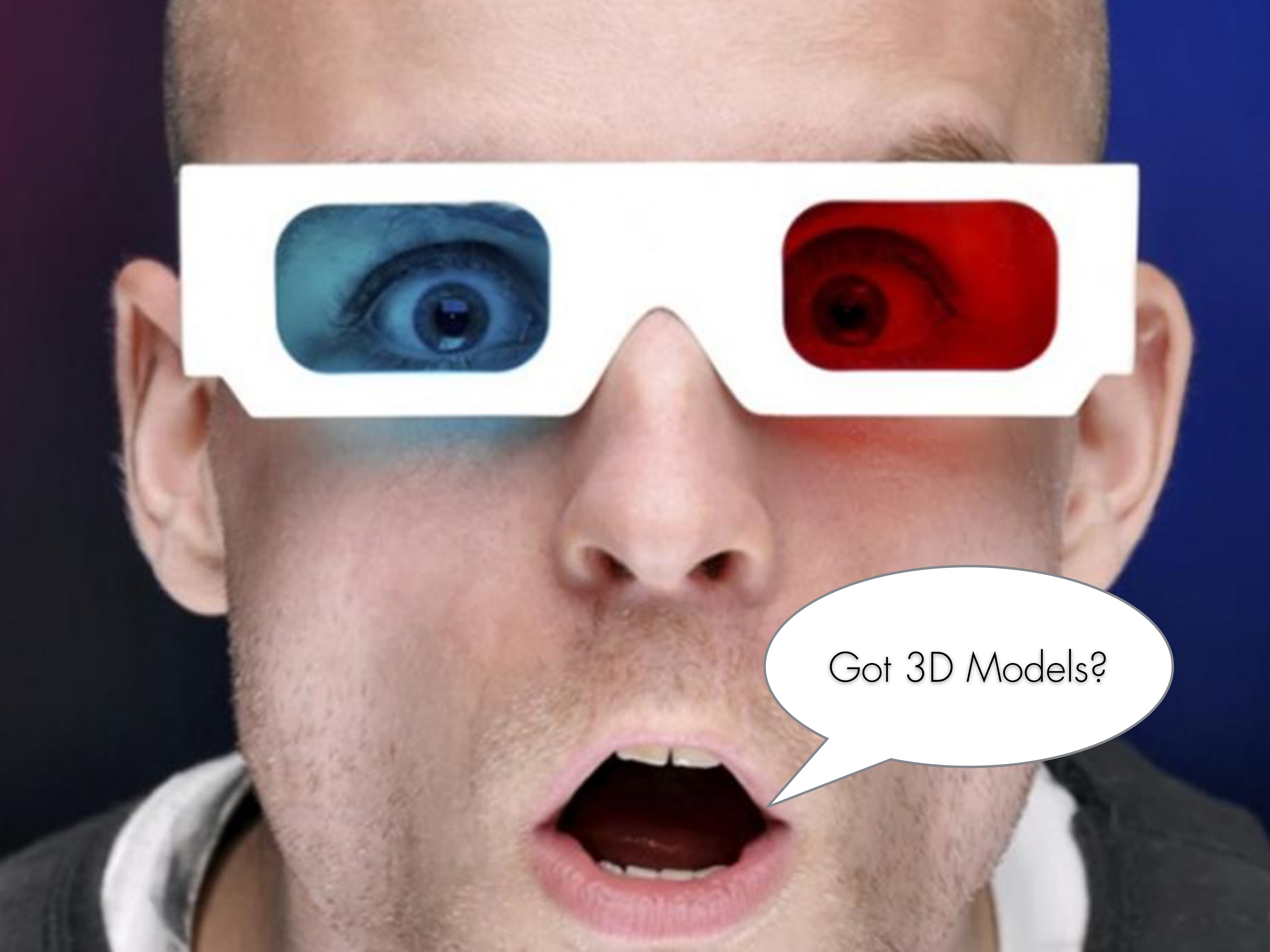


Interactions



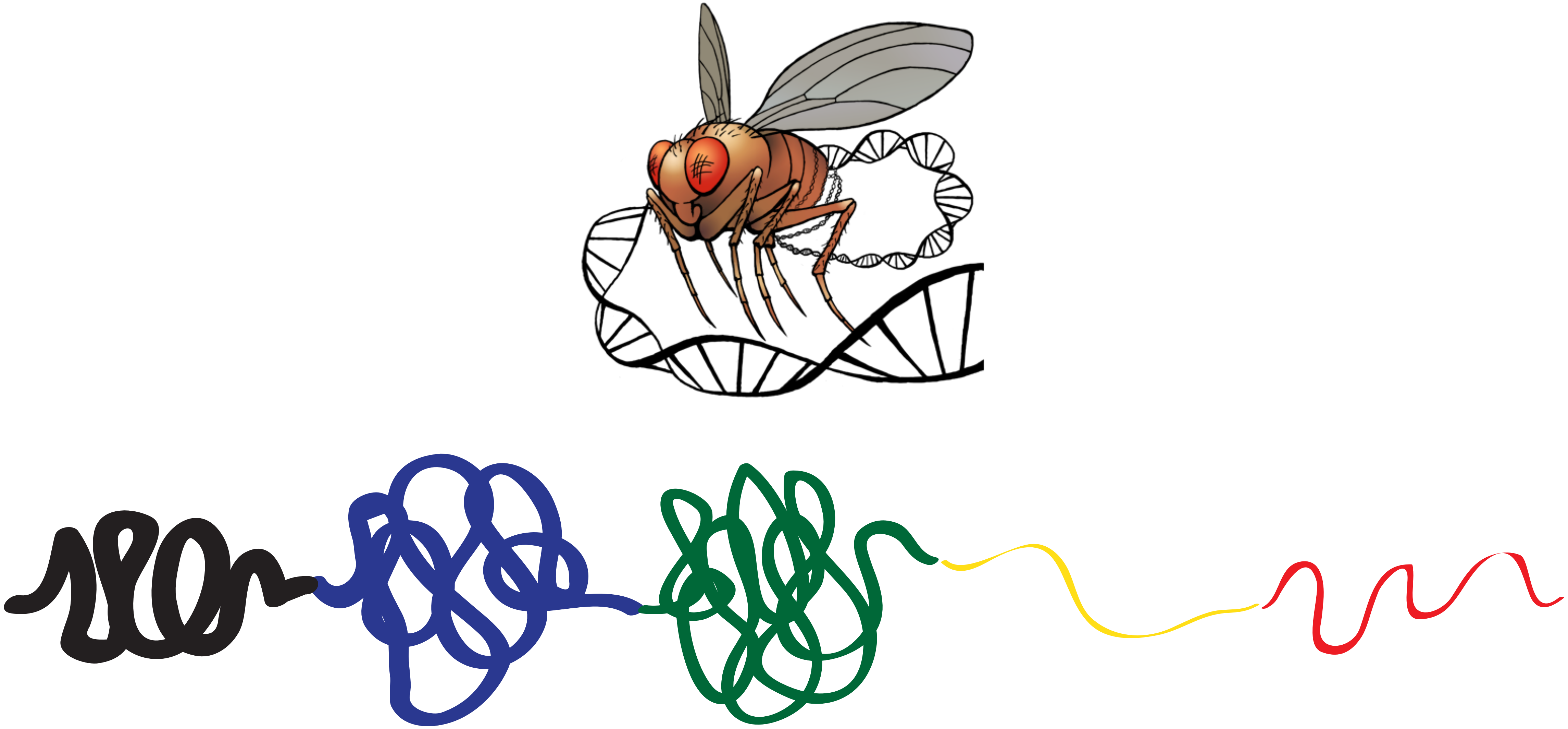
Angle




Got 3D Models?

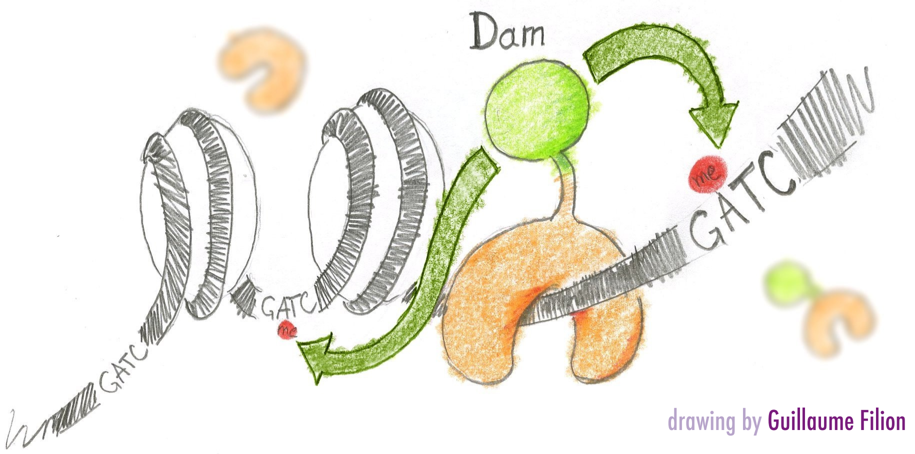
Structuring the COLORs of chromatin

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

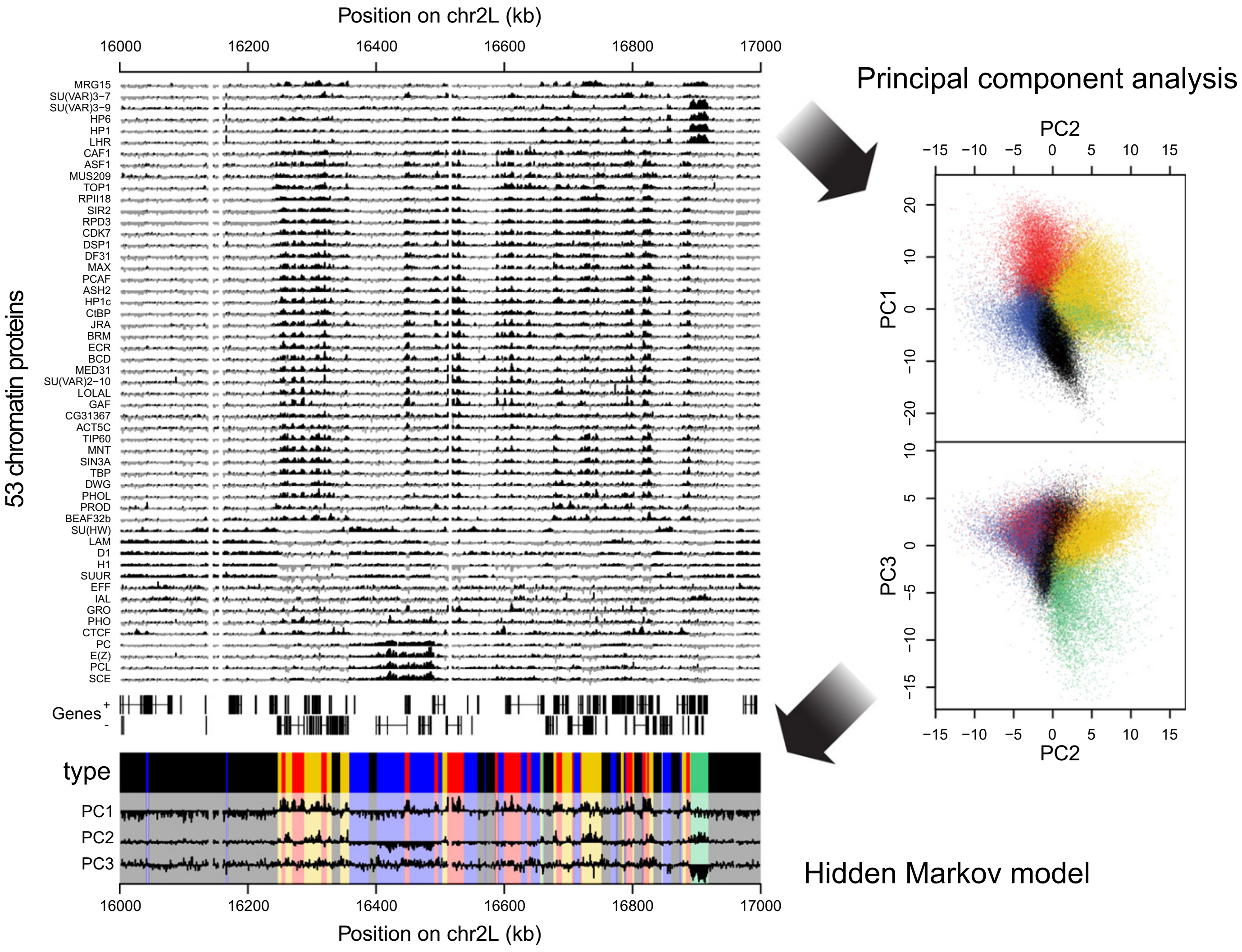


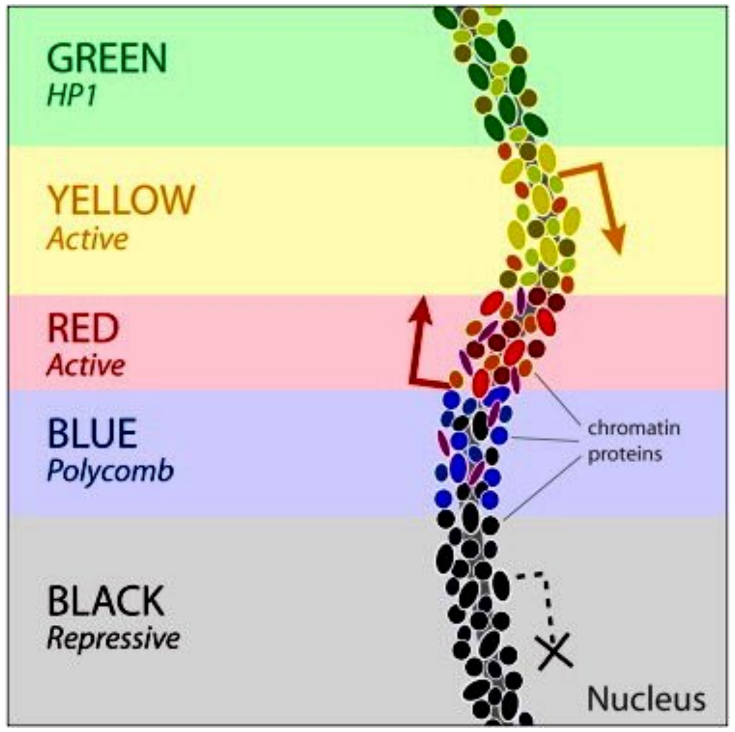
Fly Chromatin COLORs

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



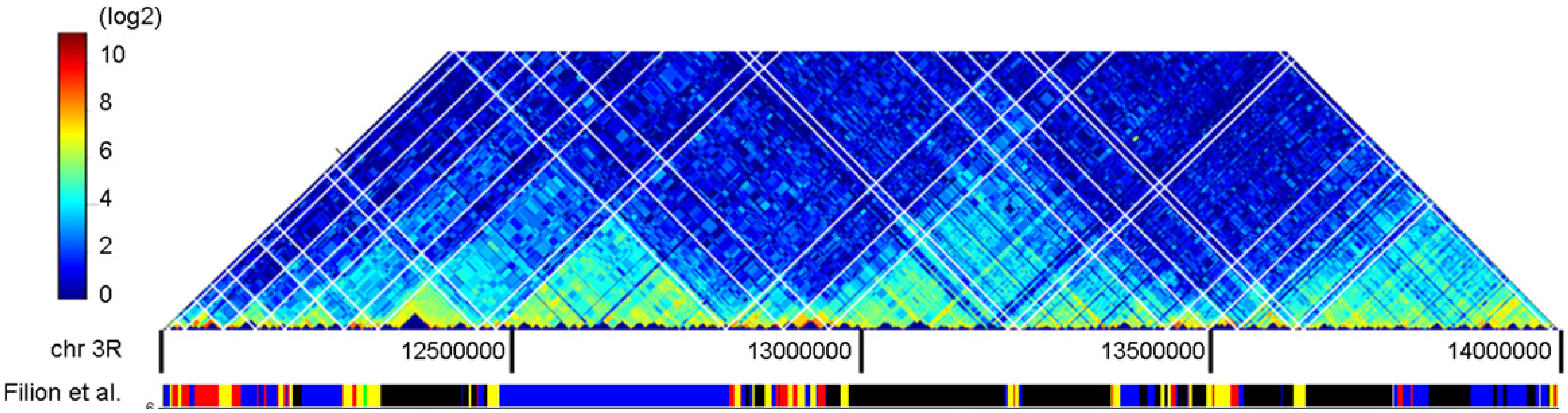
drawing by Guillaume Filion

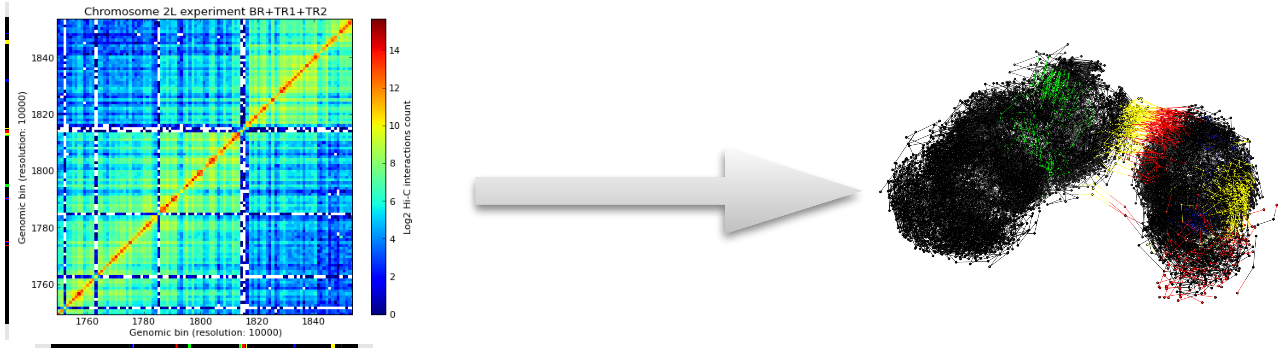




Fly Chromatin COLORs

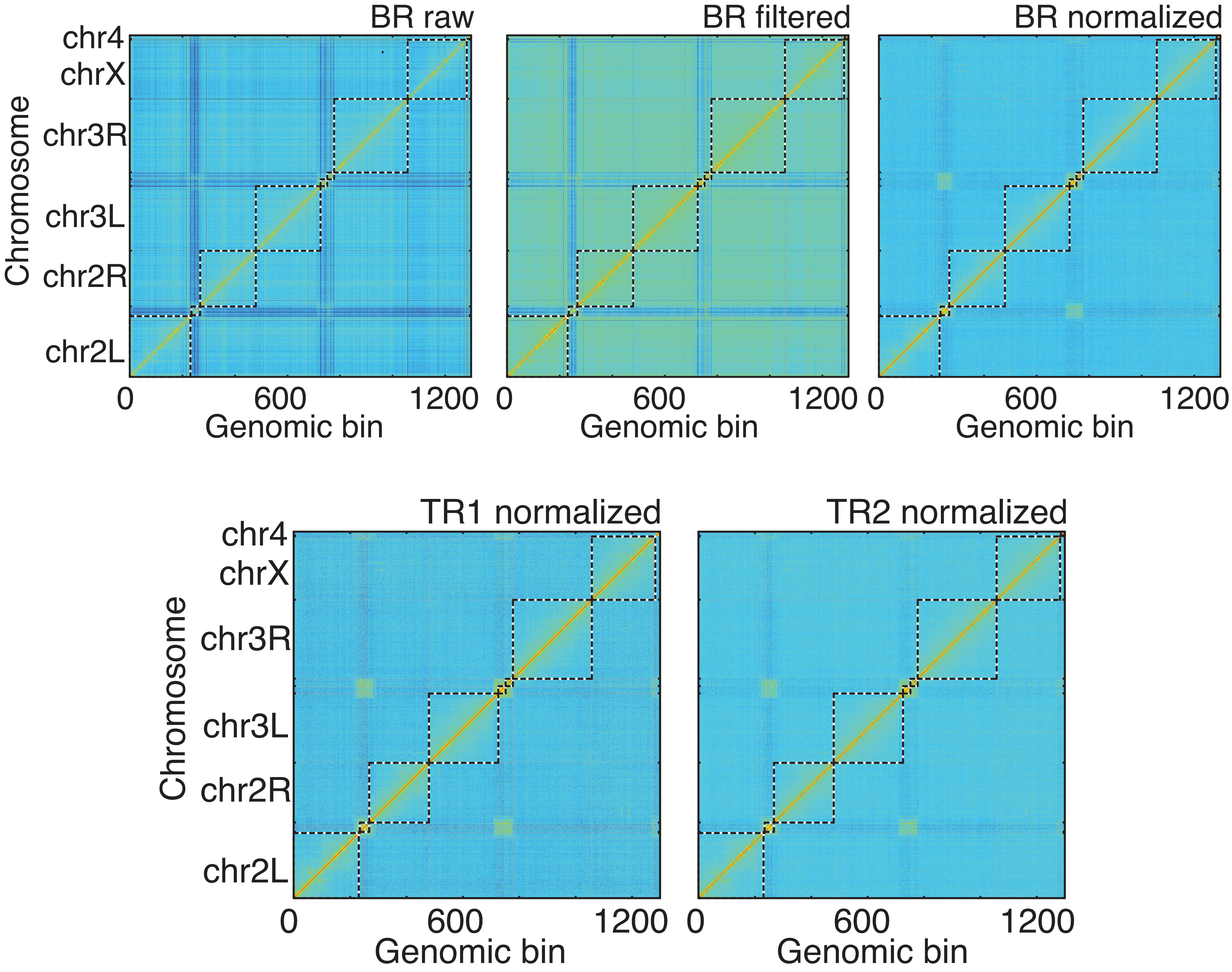
Hou et al. (2012). Molecular Cell, 48(3), 471–484.



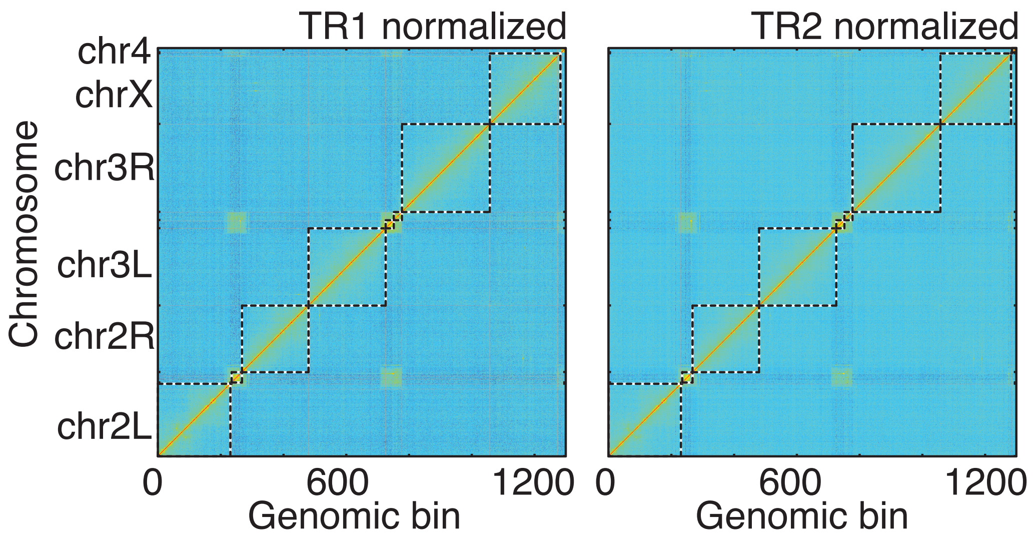


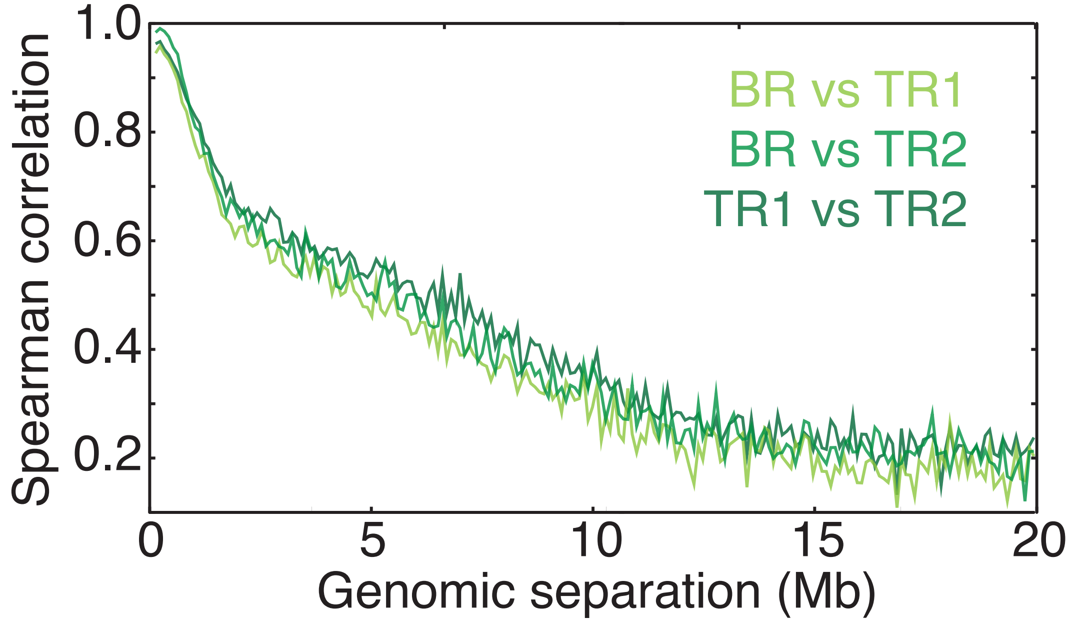
~200 regions of ~5Mb each
2Kb resolution

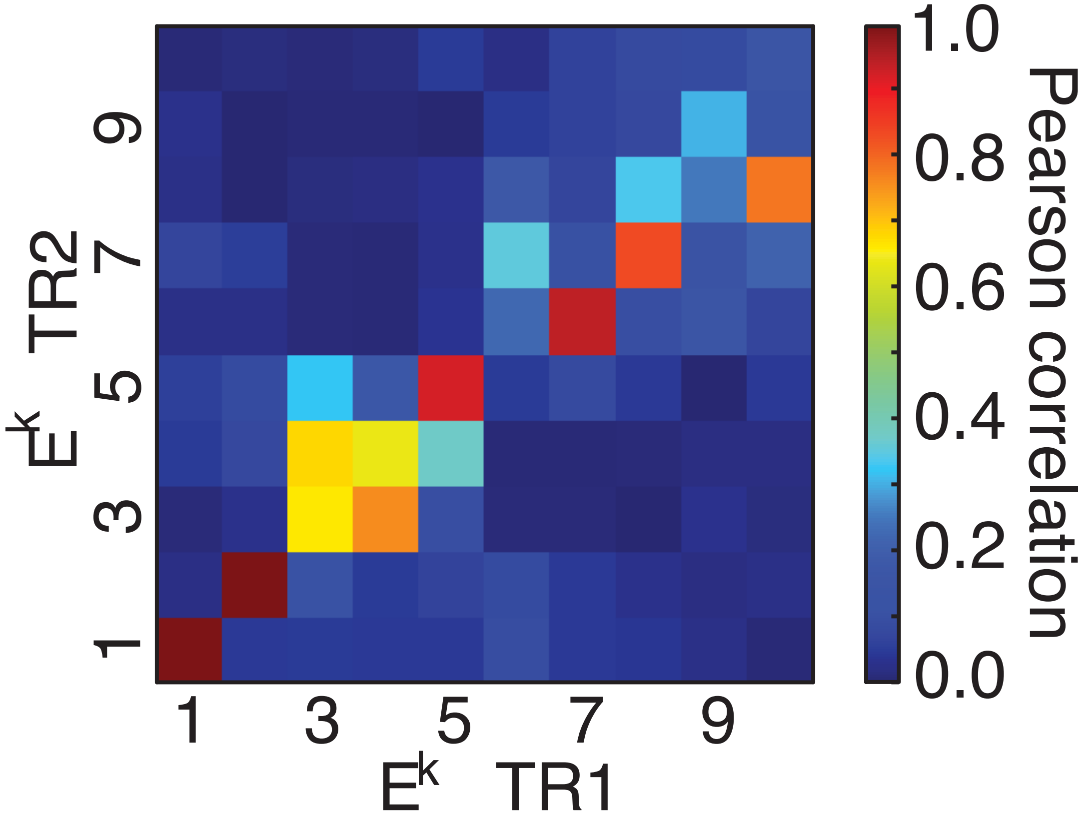
Mapping · Filtering · Normalizing



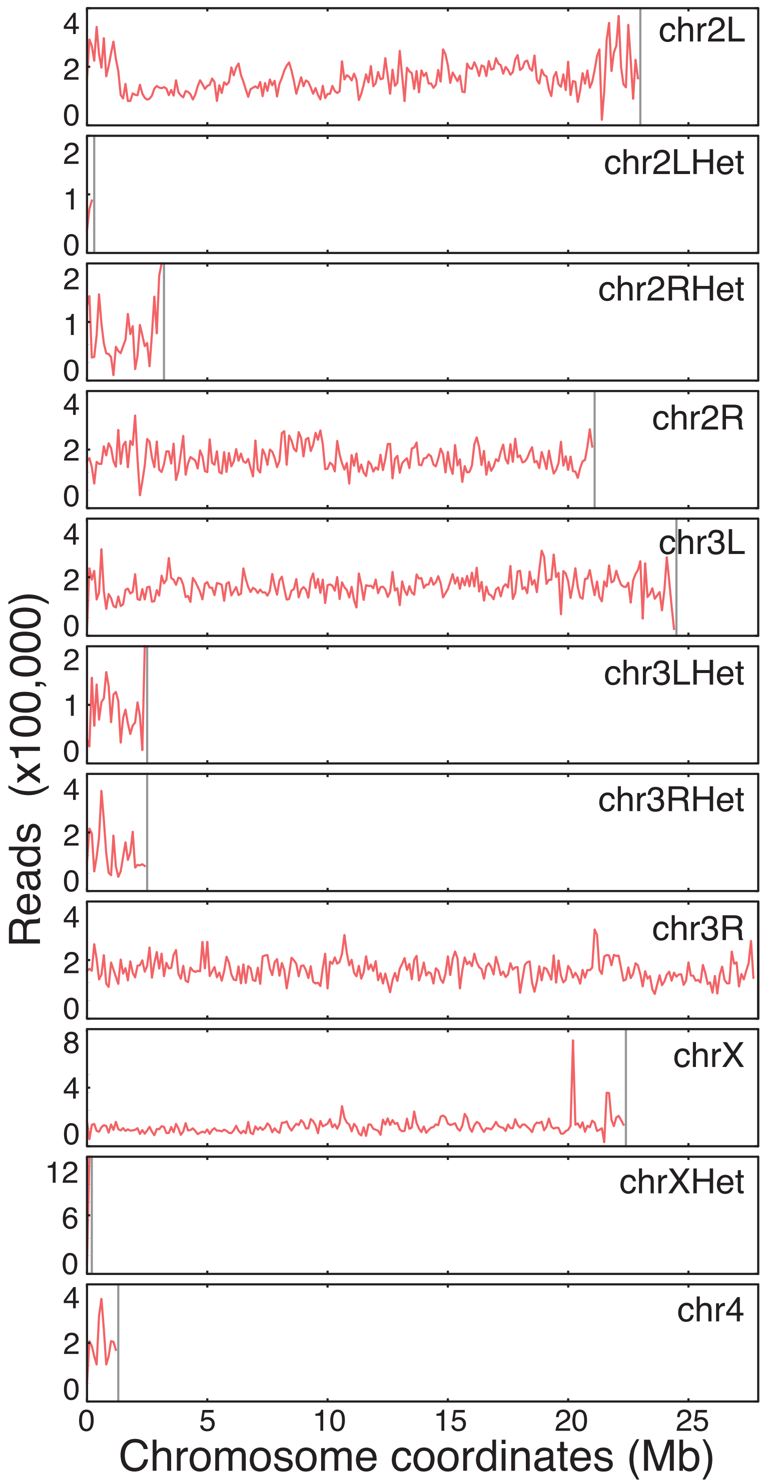
Matrix comparison

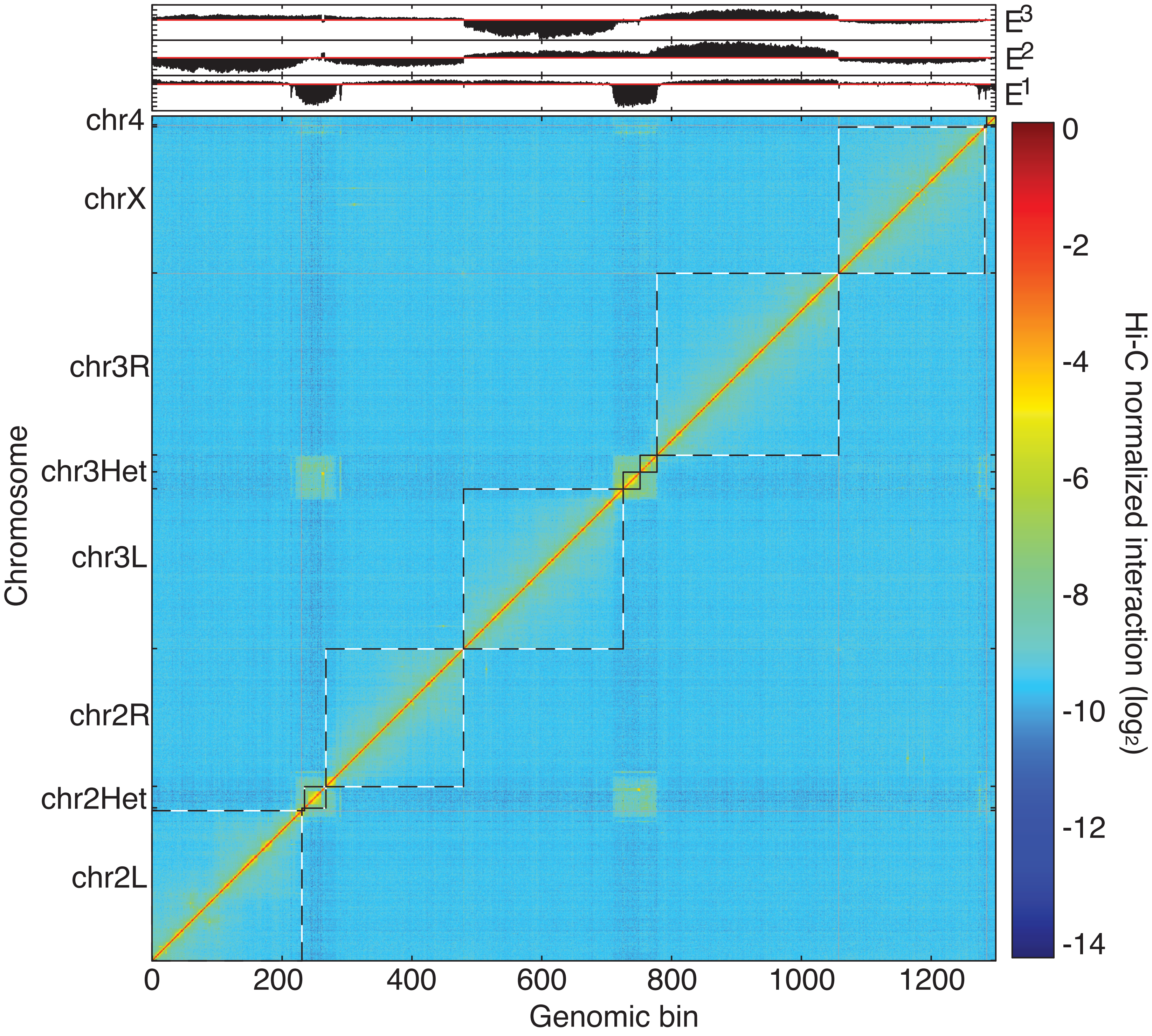




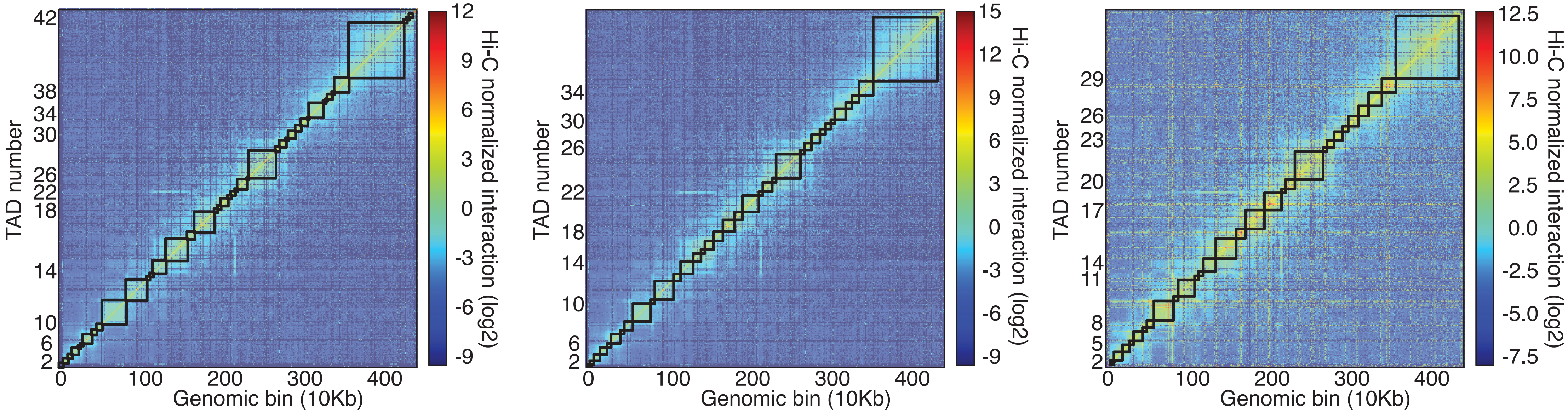


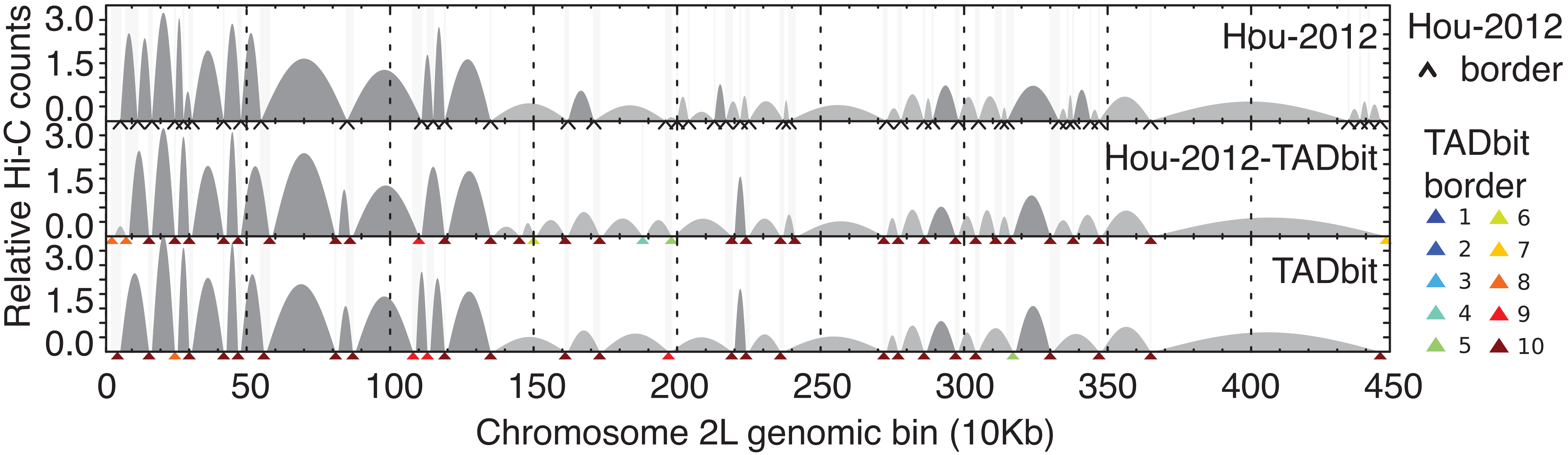
Matrix merging





TAD detection · comparison

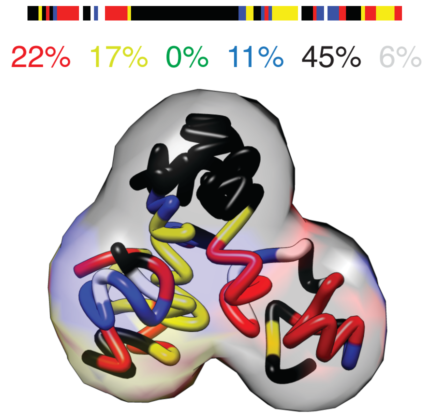




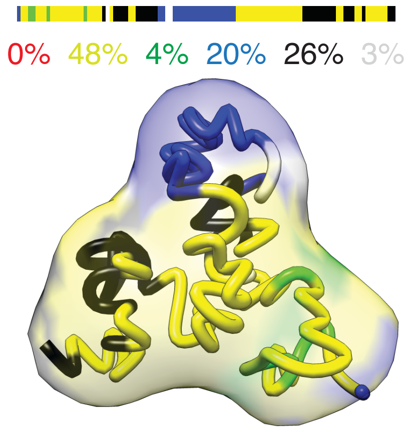
Structural properties

50 1Mb regions. 10 enriched for each color.

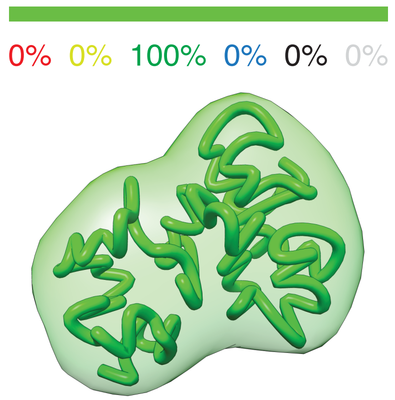
RED dense region
3R:18920000-19920000


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

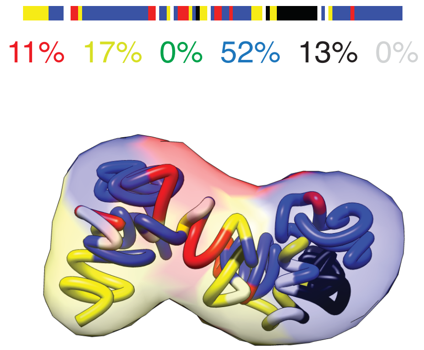
YELLOW dense region
X:15590000-16600000


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

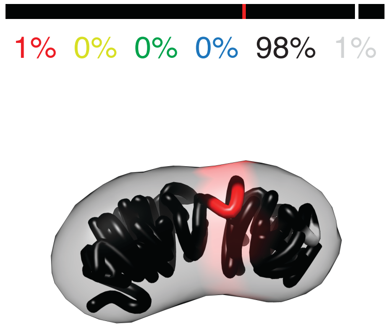
GREEN dense region
2R:510000-1530000


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

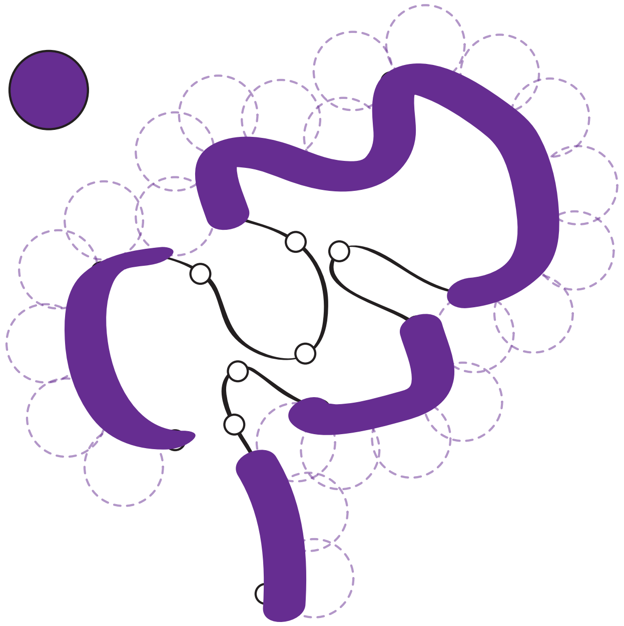
BLUE dense region
3L:210000-1230000


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

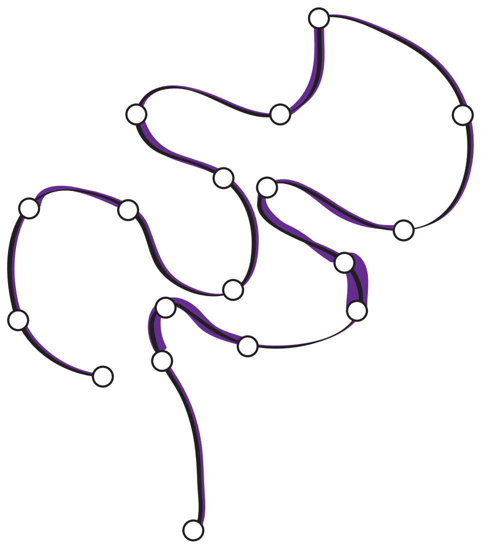
BLACK dense region
2L:17500000-18530000


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

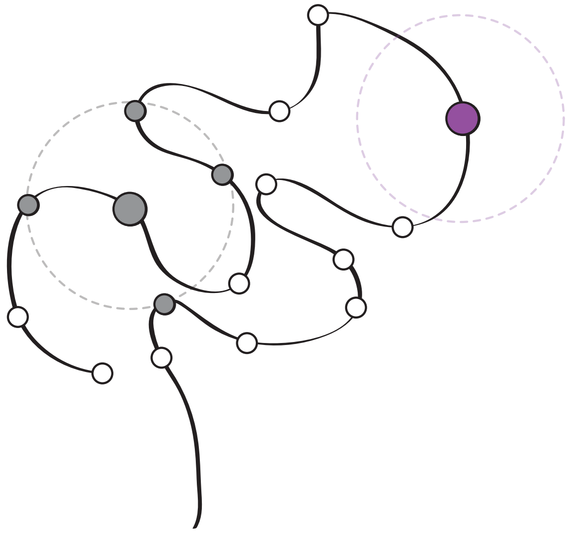
Accessibility (%)



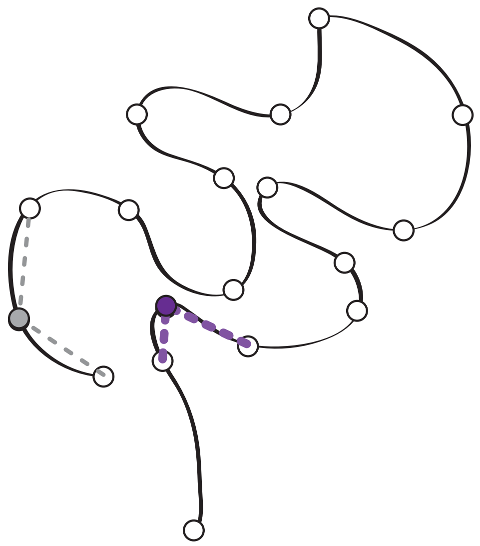
Density (bp/nm)



Interactions



Angle

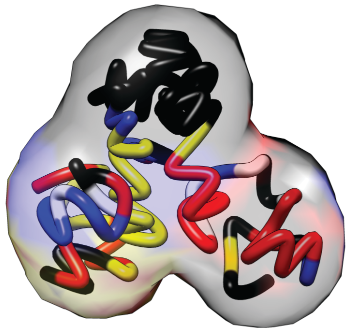


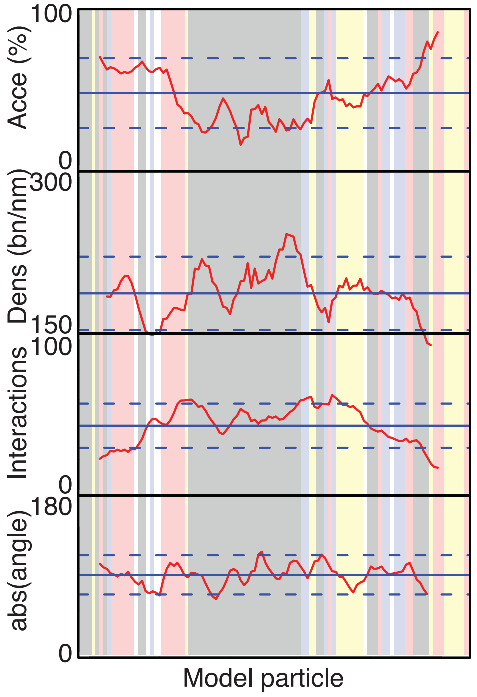
Structural COLORs

RED dense region

3R:18920000-19920000

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

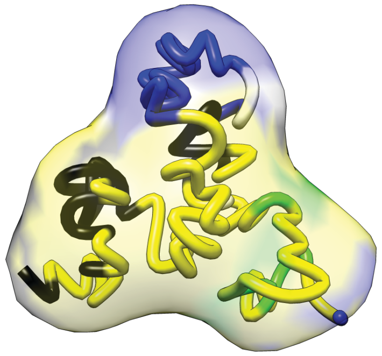


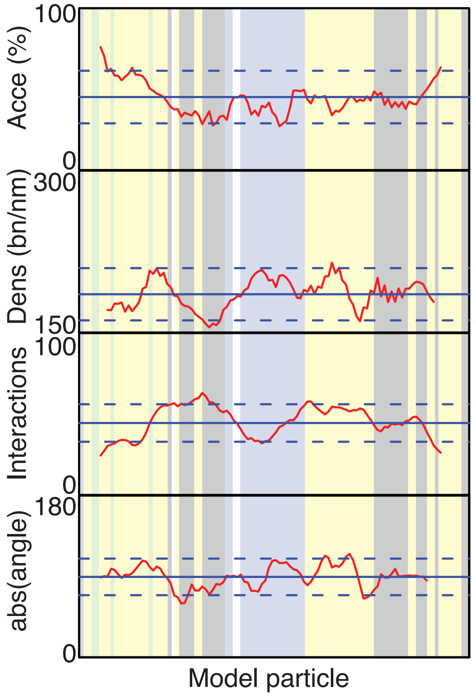


YELLOW dense region

X:15590000-16600000

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

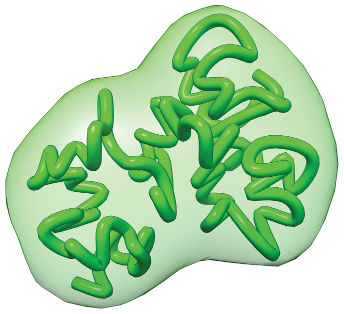


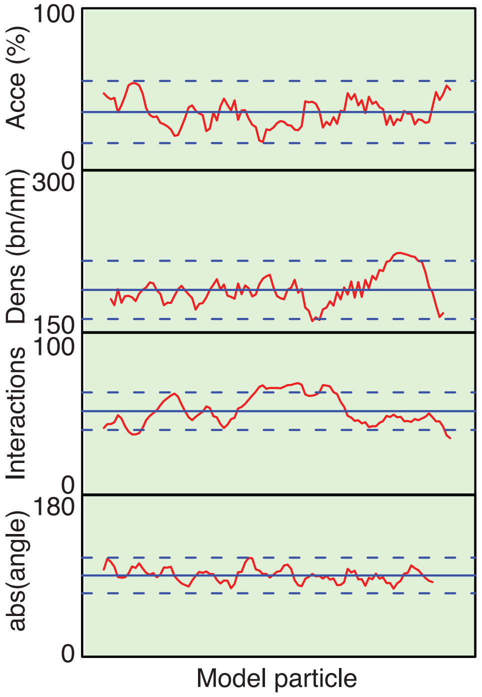


GREEN dense region

2R:510000-1530000

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

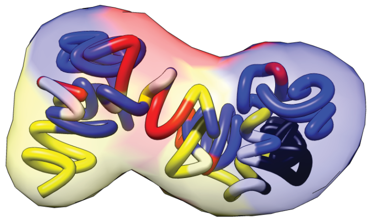


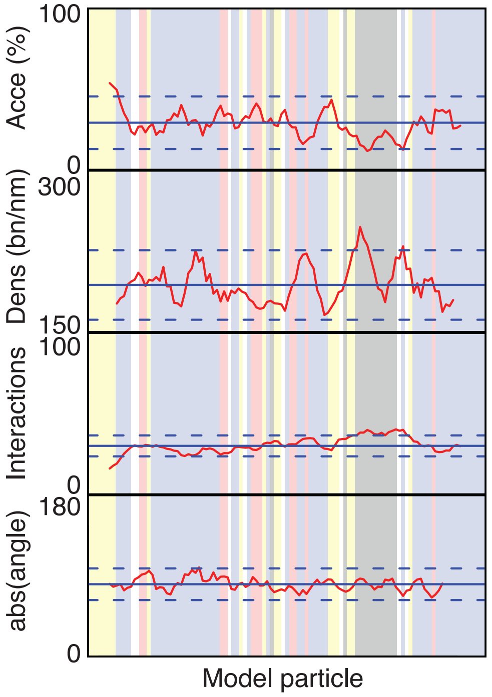


BLUE dense region

3L:210000-1230000

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

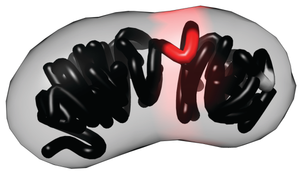


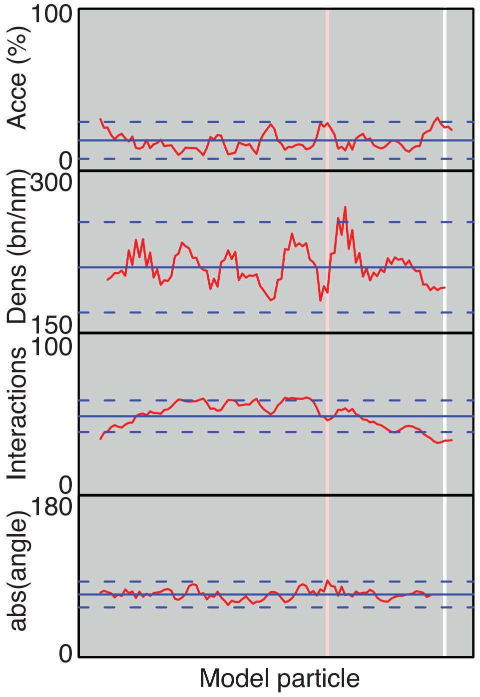


BLACK dense region

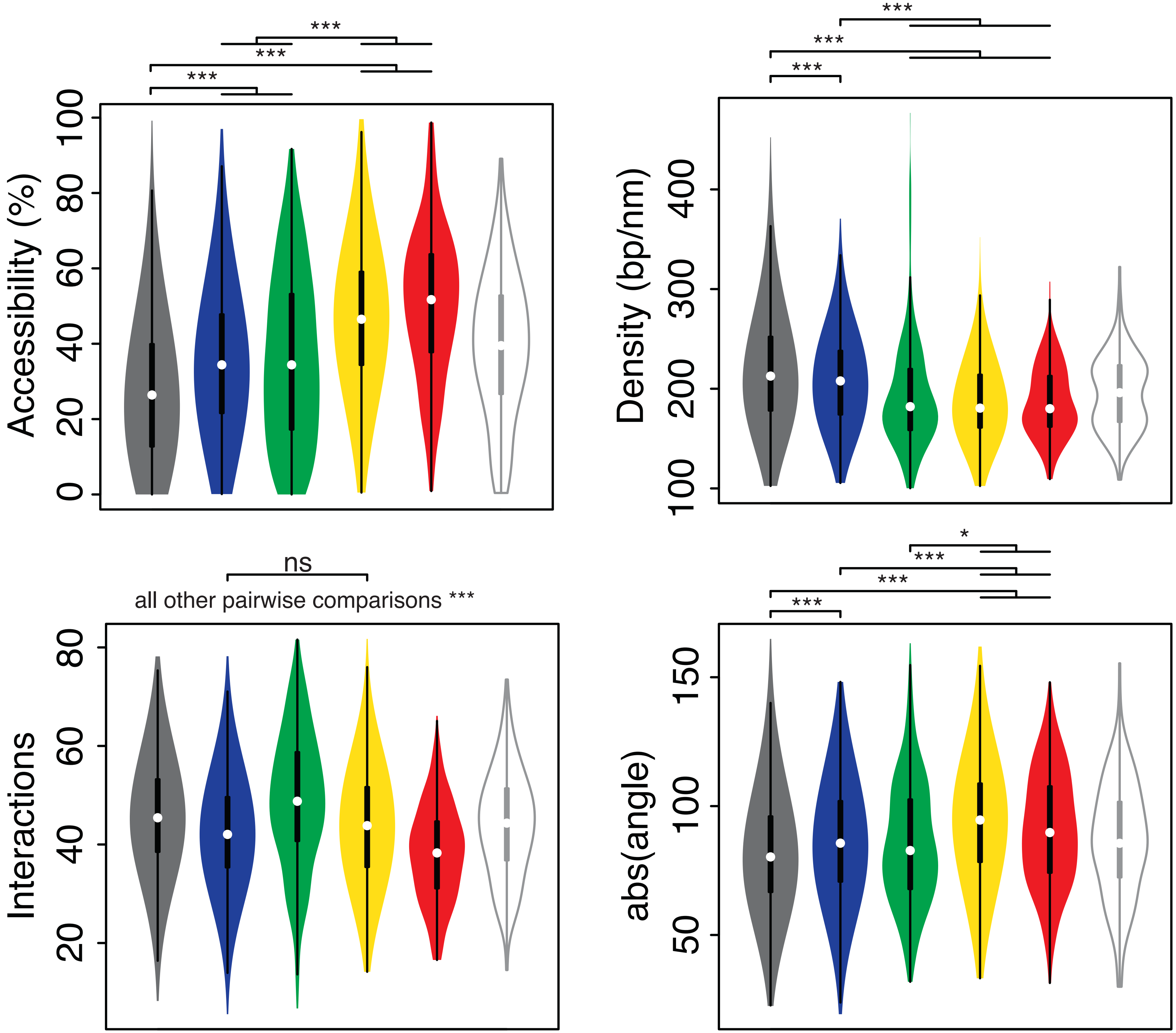
2L:17500000-18530000

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

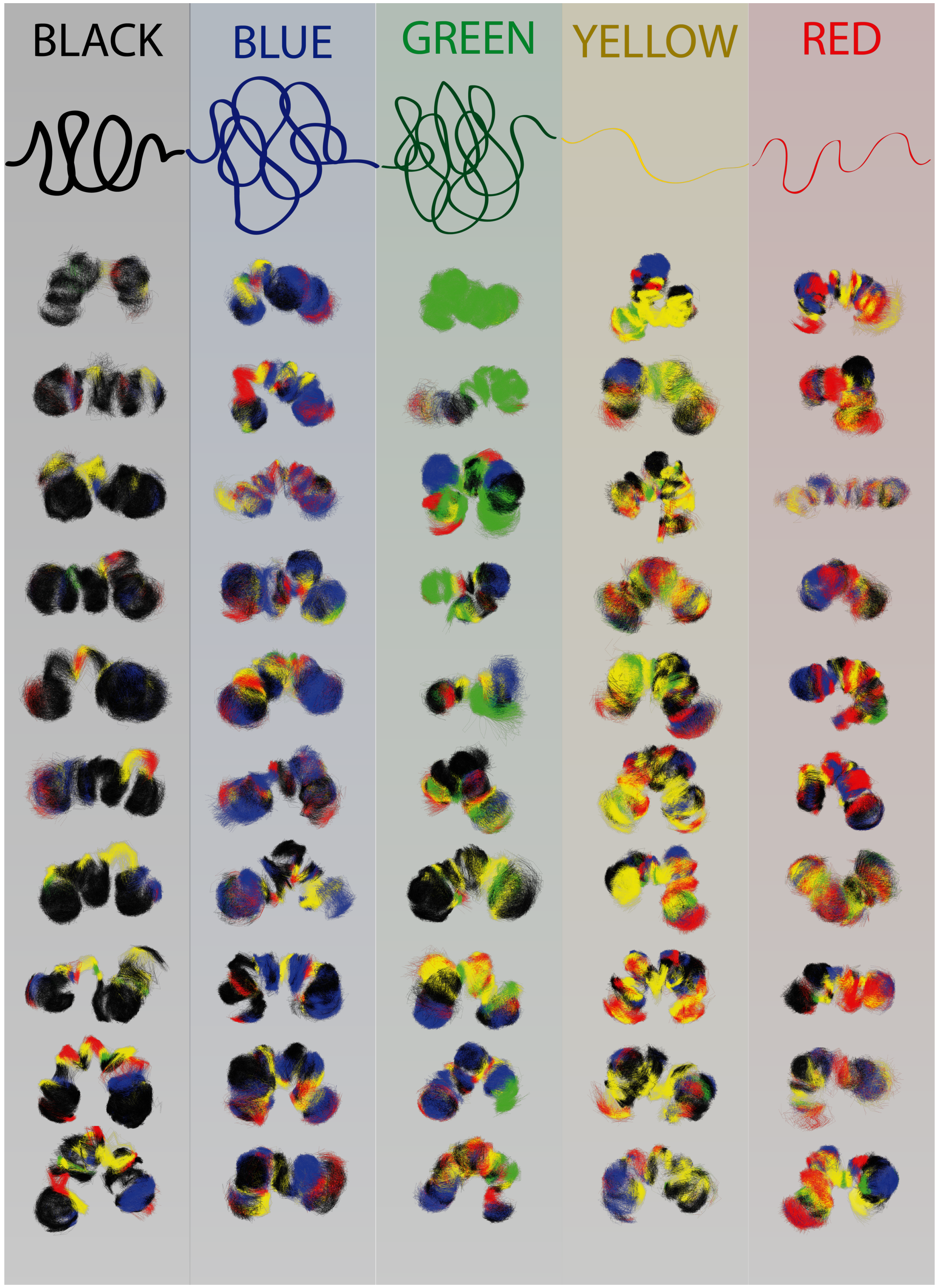


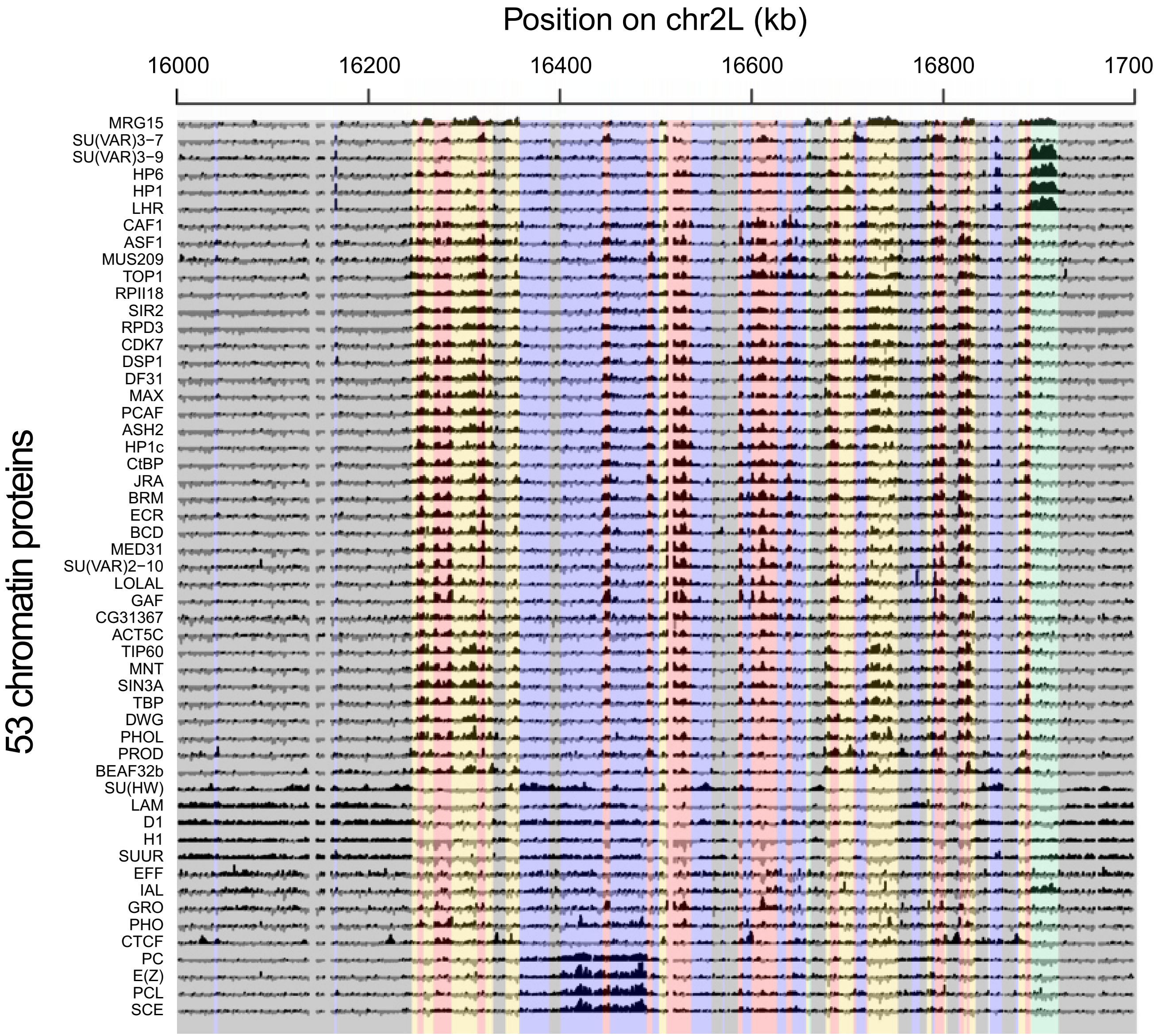


Structural COLORs



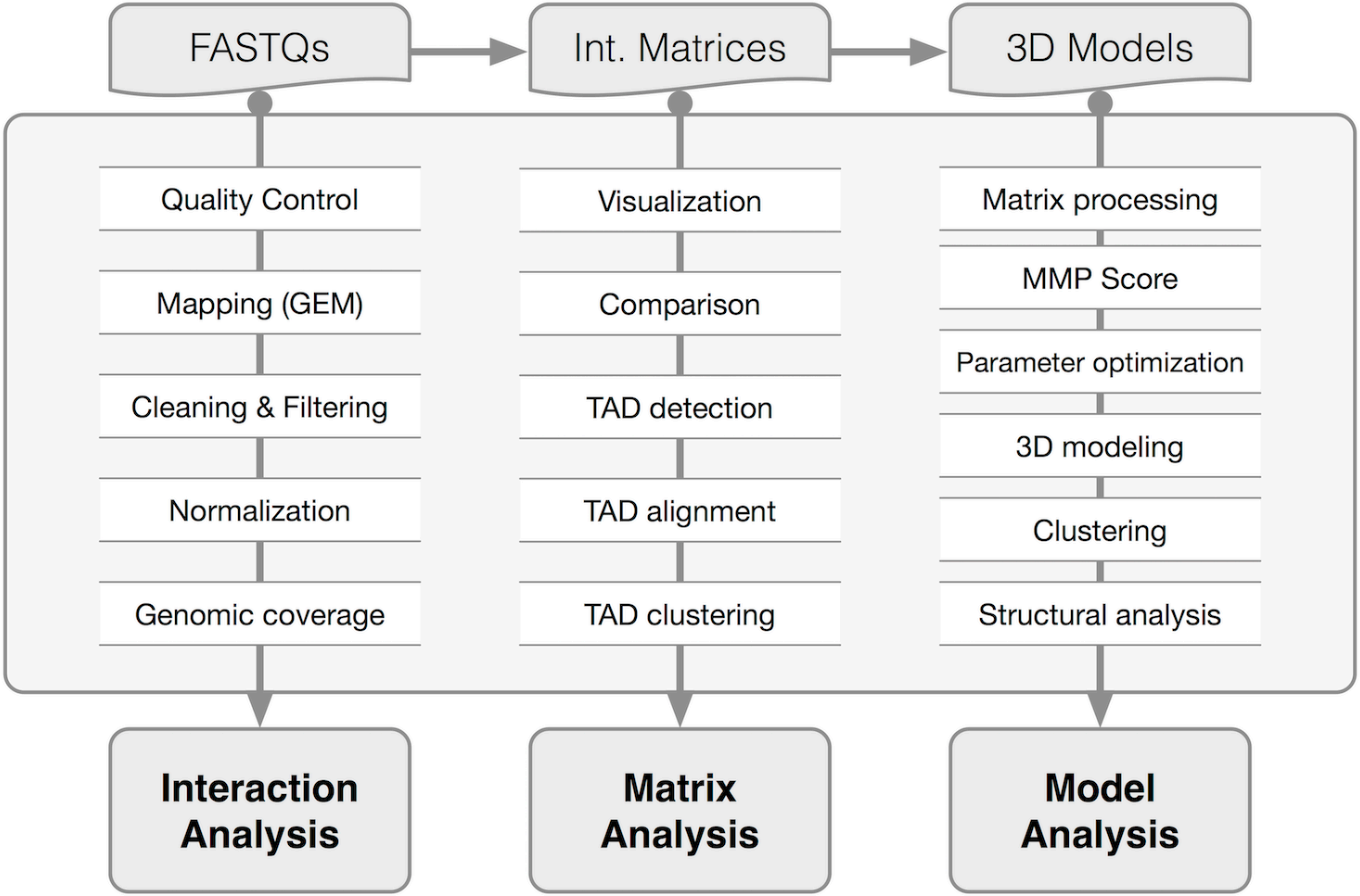
Structural COLORs





TADbit

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



FASTQs → Int. Matrices → 3D Models

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

Interaction Analysis

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

Matrix Analysis

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

Model Analysis

TADbit other applications…

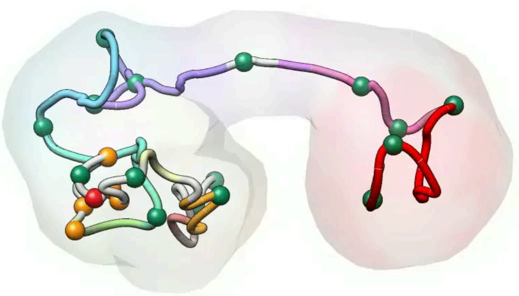
Baù, D. et al. Nat Struct Mol Biol (2011)
Umbarger, M. A. et al. Mol Cell (2011)
Le Dily, F. et al. Genes & Dev (2014)
Trussart et al. Nature Comm. (2017)

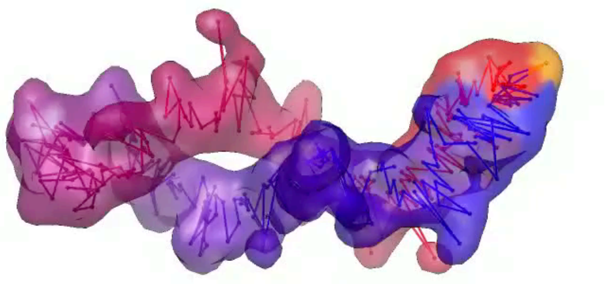
TECHNICAL REPORTS

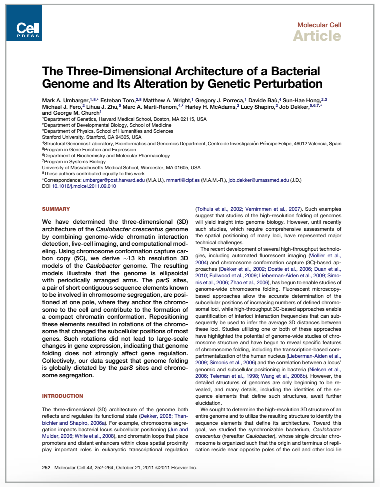
nature structural & molecular biology

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

Davide Baù[1,4], Amartya Sanyal[2,4], Bryan R Lajoie[2,4], Emidio Capriotti[1], Meg Byron[3], Jeanne B Lawrence[3], Job Dekker[2] & Marc A Marti-Renom[1]





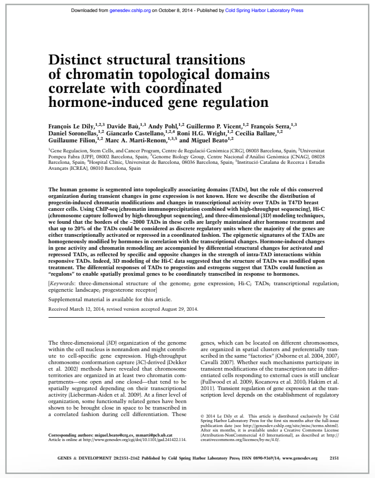


Molecular Cell
Article

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

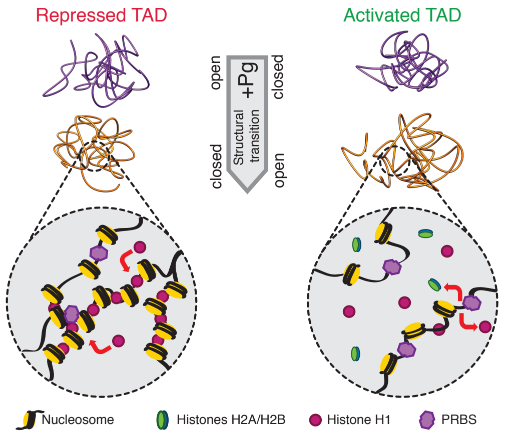
Mark A. Umbarger,[1,8,*] Esteban Toro,[2,8] Matthew A. Wright,[1] Gregory J. Porreca,[1] Davide Baù,[4] Sun-Hae Hong,[2,3] Michael J. Fero,[2] Lihua J. Zhu,[5] Marc A. Marti-Renom,[4,*] Harley H. McAdams,[2] Lucy Shapiro,[2] Job Dekker,[5,6,7,*] and George M. Church[1]

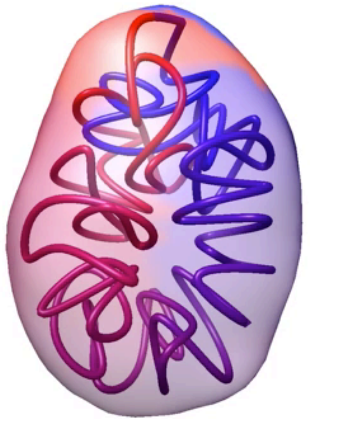
DOI 10.1016/j.molcel.2011.09.010



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

François Le Dily,[1,2,3] Davide Baù,[1,3] Andy Pohl,[1,2] Guillermo P. Vicent,[1,2] François Serra,[1,3] Daniel Soronellas,[1,2] Giancarlo Castellano,[1,2,4] Roni H.G. Wright,[1,2] Cecilia Ballare,[1,2] Guillaume Filion,[1,2] Marc A. Marti-Renom,[1,3,5] and Miguel Beato[1,2]





nature COMMUNICATIONS

ARTICLE

Received 4 Nov 2016 | Accepted 20 Jan 2017 | Published 8 Mar 2017 DOI: 10.1038/ncomms14665 OPEN

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

Marie Trussart[1,2], Eva Yus[1,2], Sira Martinez[1], Davide Baù[3,4], Yuhei O. Tahara[5,6], Thomas Pengo[1,7], Michael Widjaja[8], Simon Kretschmer[9], Jim Swoger[1,2], Steven Djordjevic[8], Lynne Turnbull[8], Cynthia Whitchurch[8], Makoto Miyata[5,6], Marc A. Marti-Renom[2,3,4,10], Maria Lluch-Senar[1,2] & Luís Serrano[1,2,10]

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

cnag
CRG Centre for Genomic Regulation
ICREA

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

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

erc
Multiscale Complex Genomics
GOBIERNO DE ESPAÑA
MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD
Generalitat de Catalunya
Fundació La Marató de TV3
Fundación Bancaria "la Caixa"