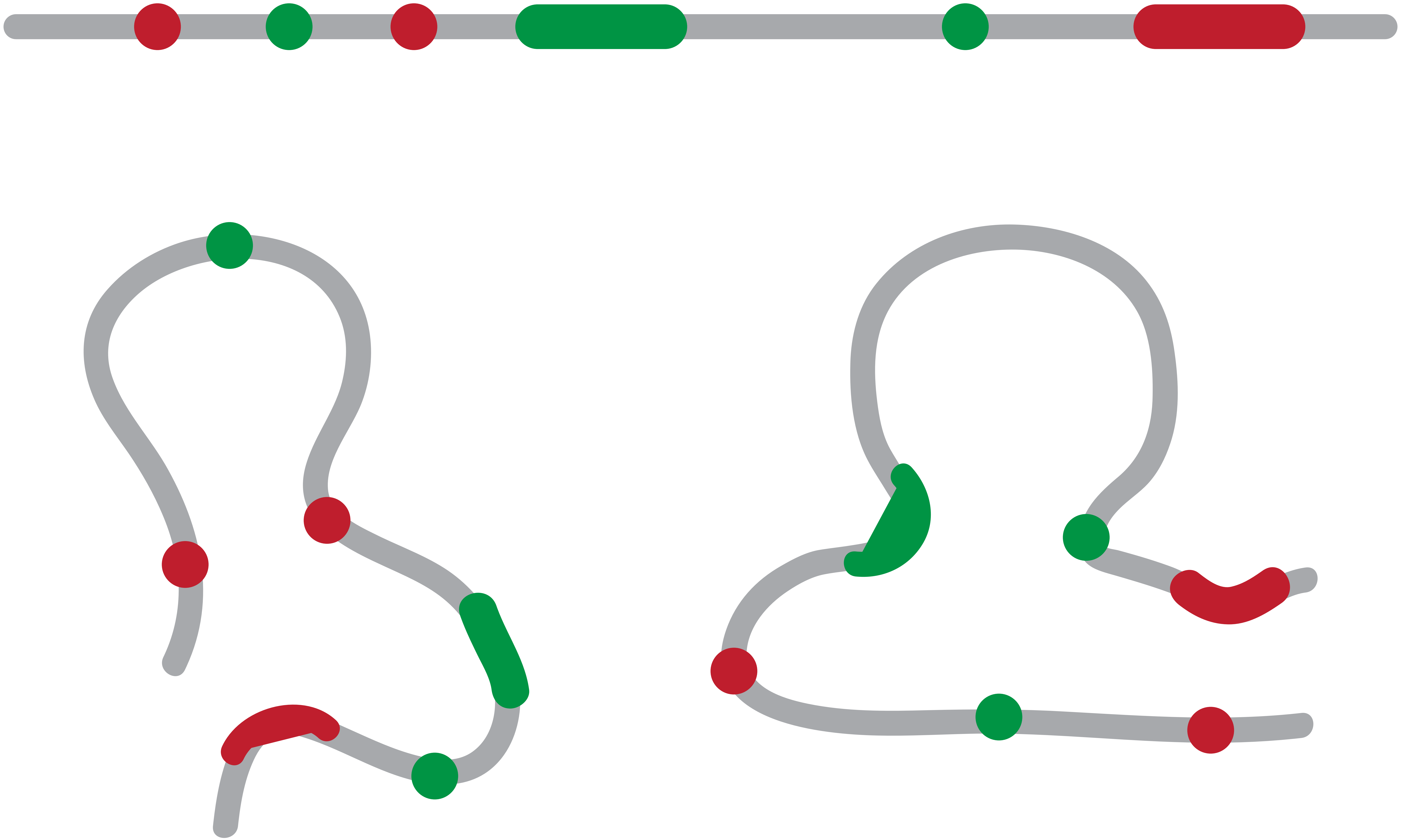

HUMAN "The Movie" by Yann Arthus-Bertrand

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

Marc A. Marti-Renom

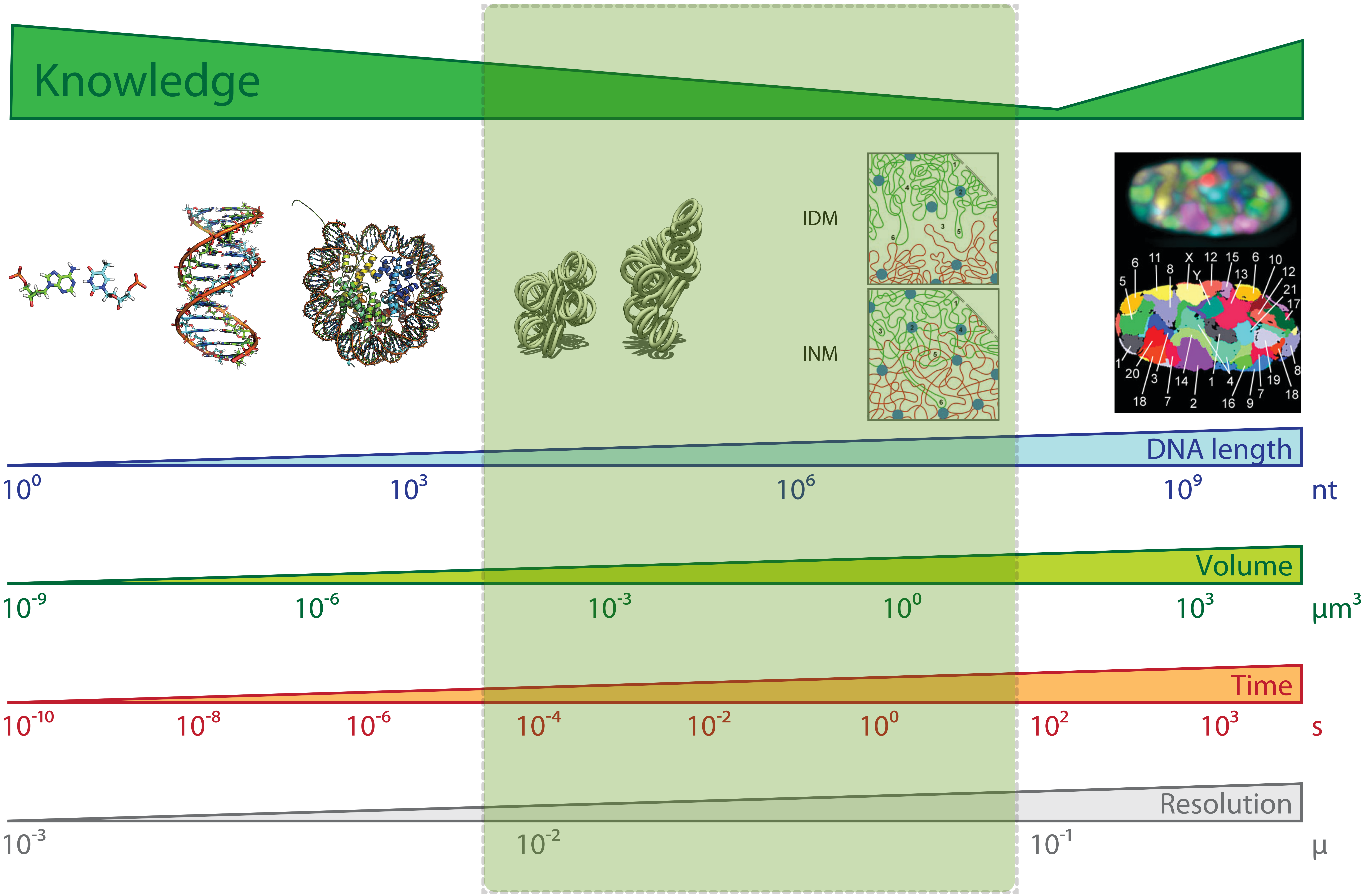
CNAG-CRG · ICREA

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



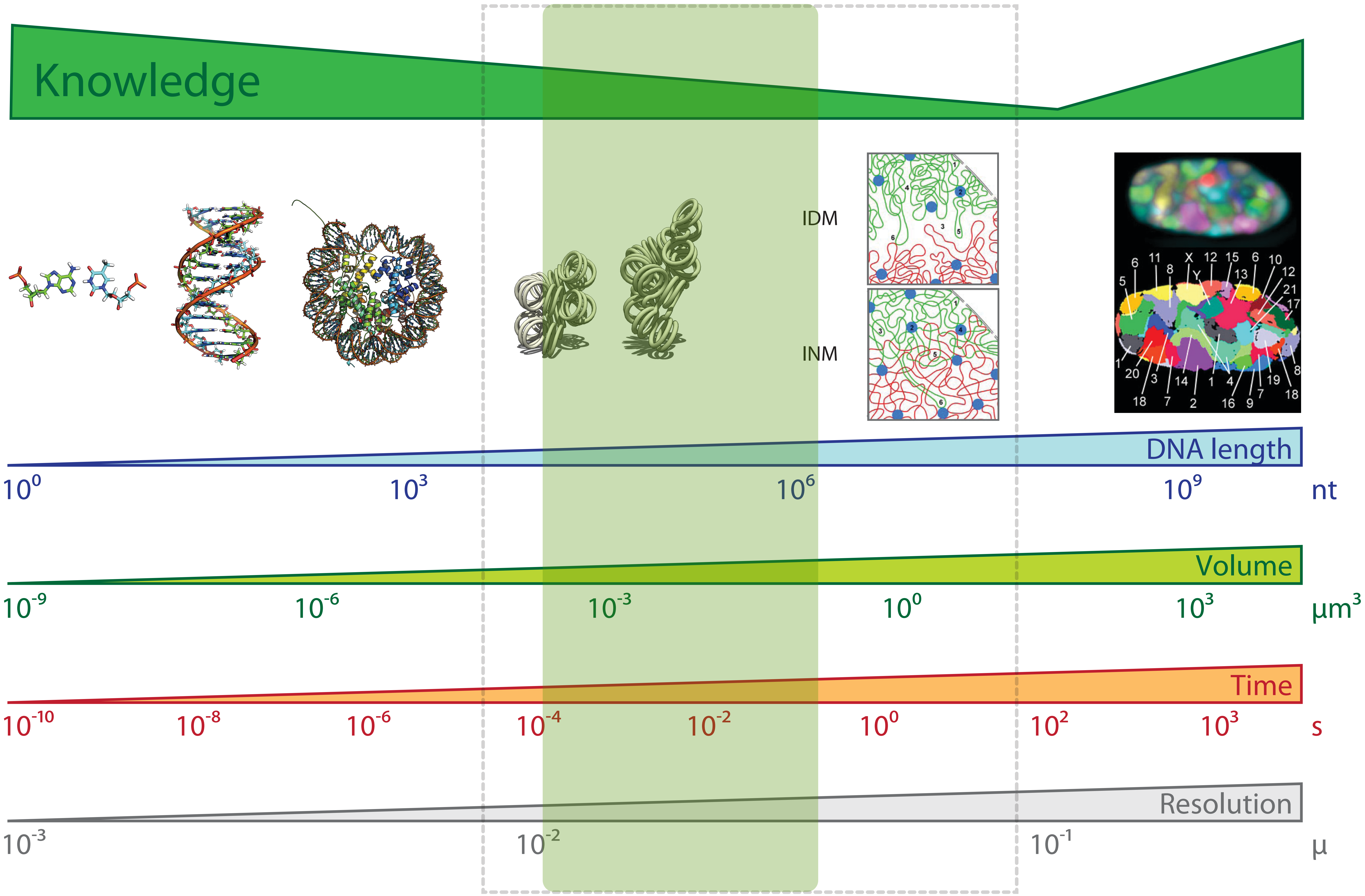
Resolution Gap

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



Resolution Gap

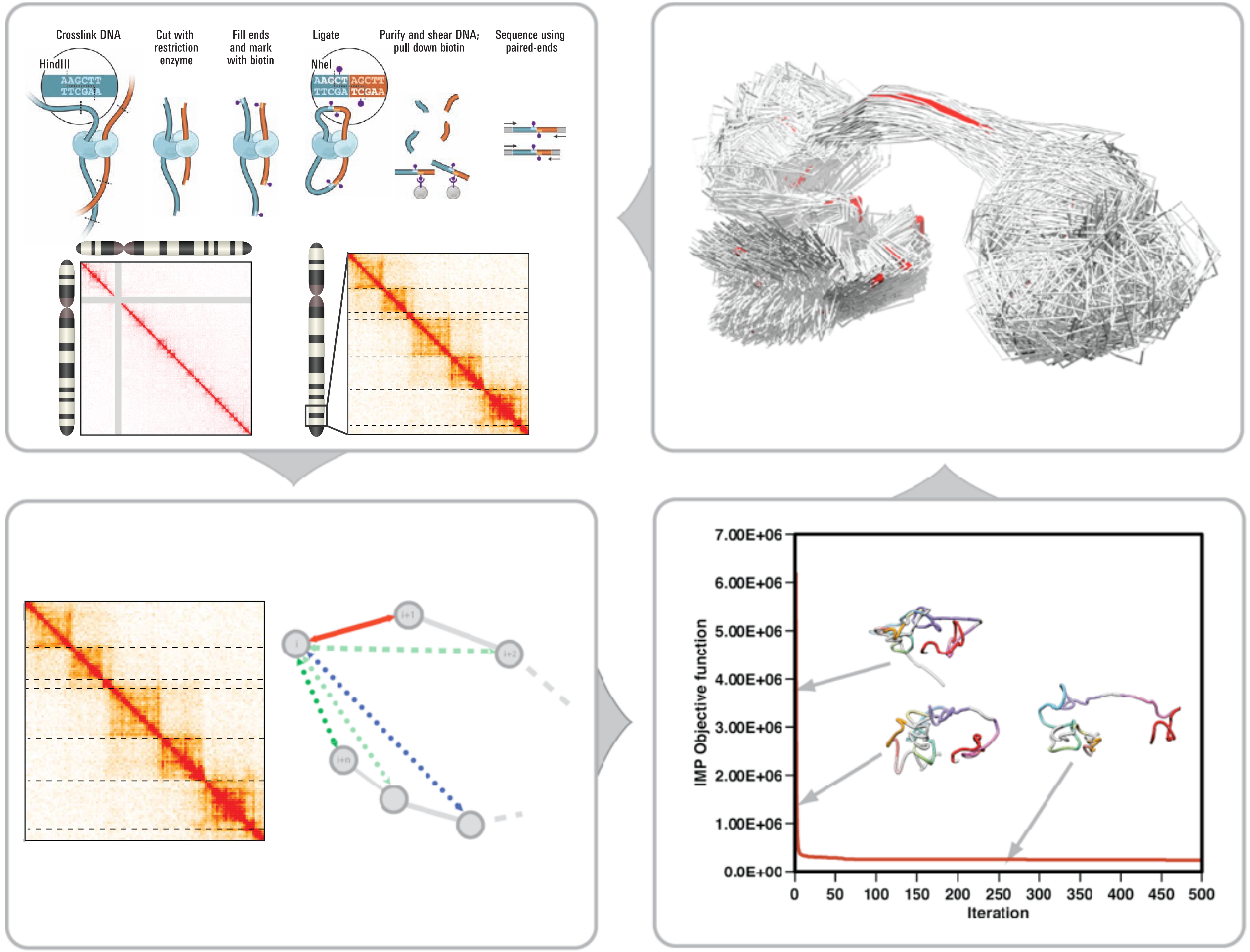
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol 7, e1002125 (2011)



Hybrid Method

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

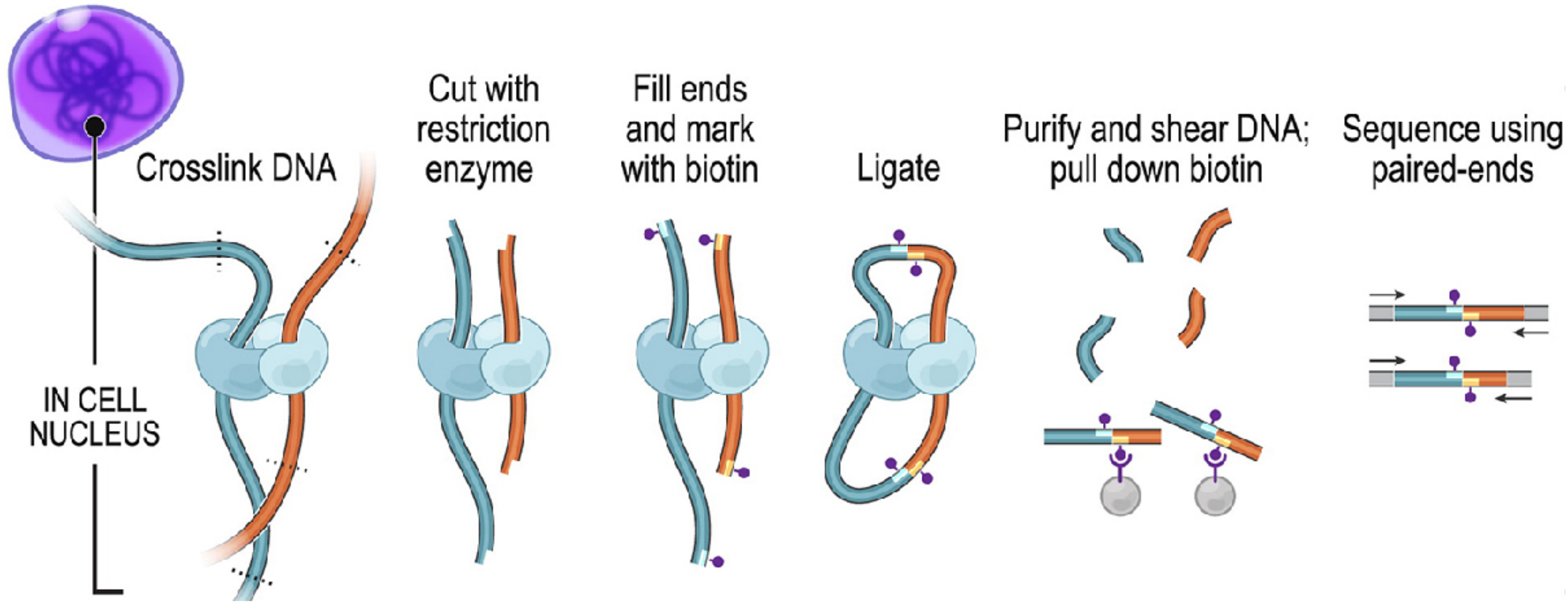
Experiments

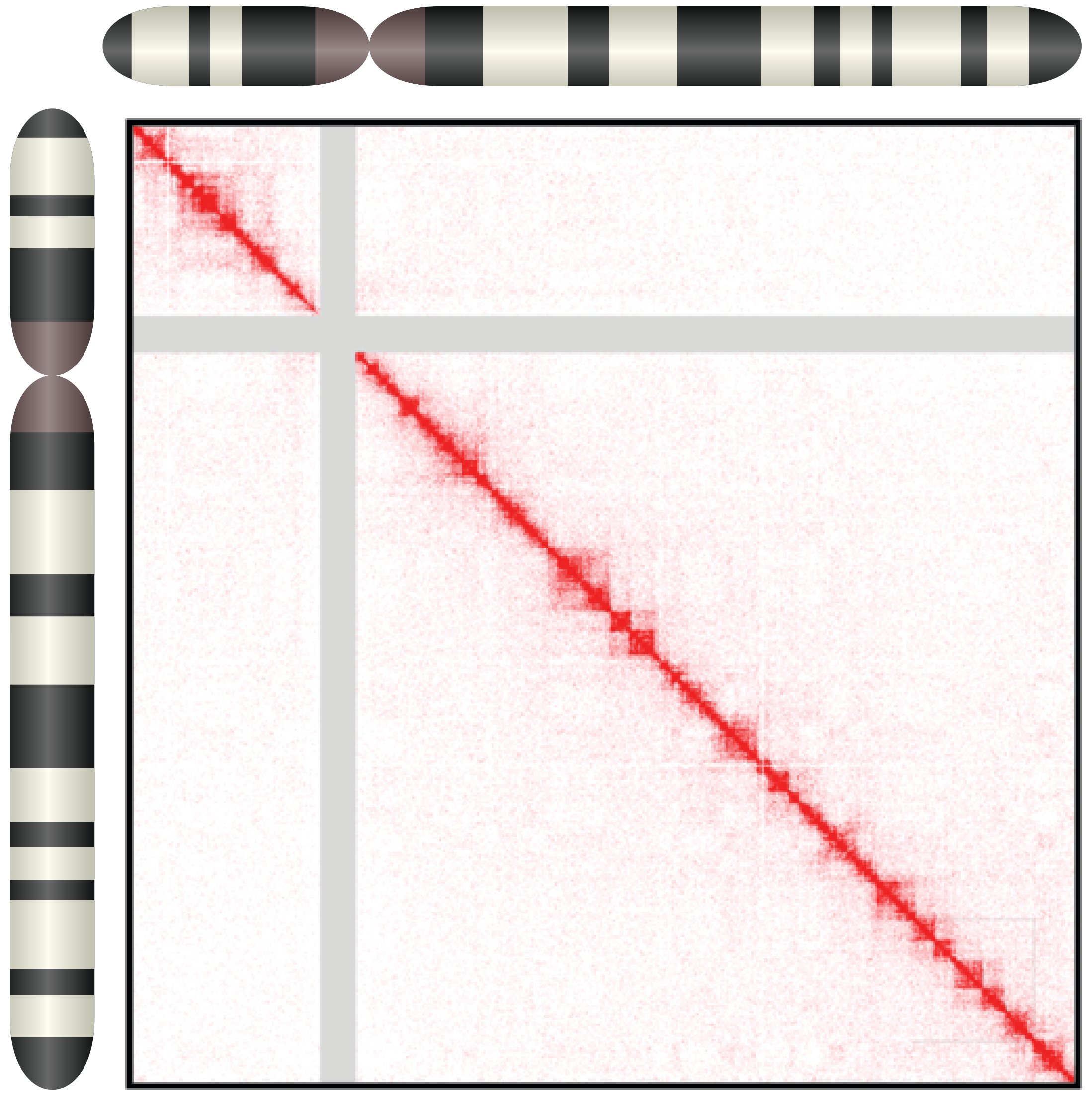


Computation

Chromosome Conformation Capture

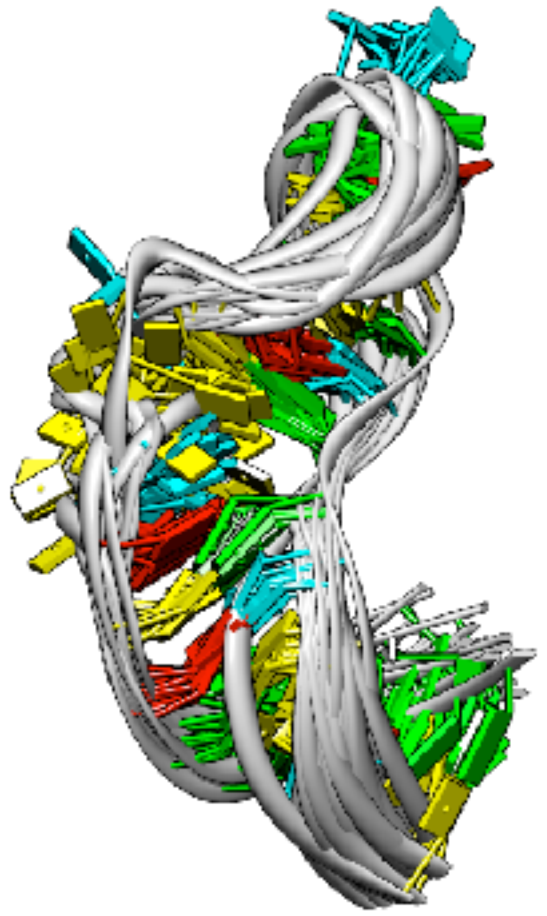
Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Science, 295(5558), 1306–1311.
Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.

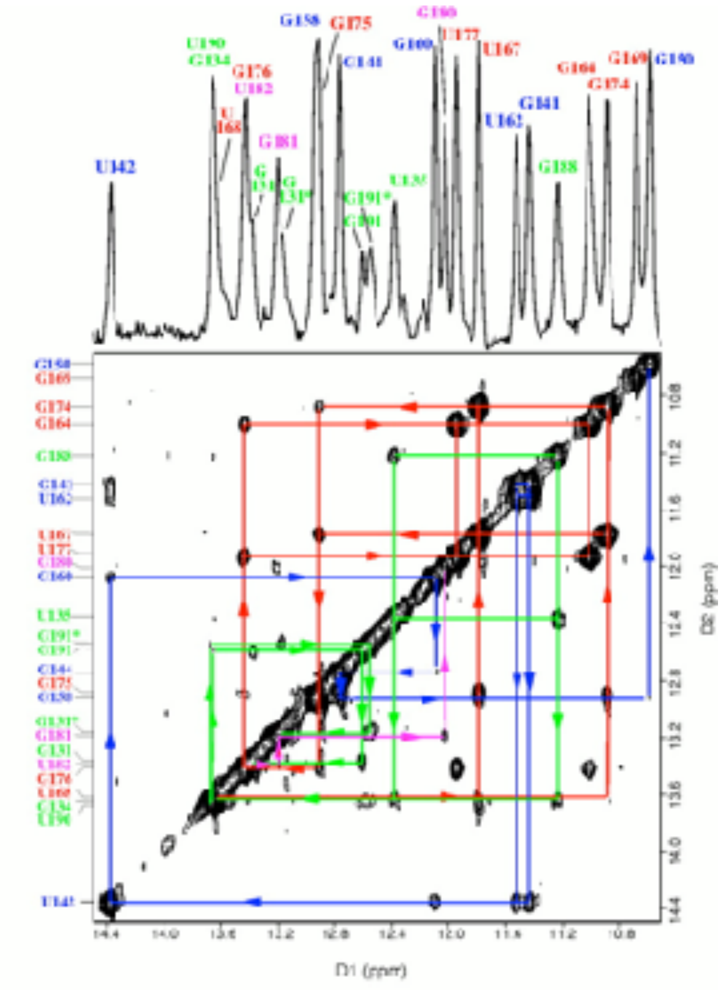




Restraint-based Modeling

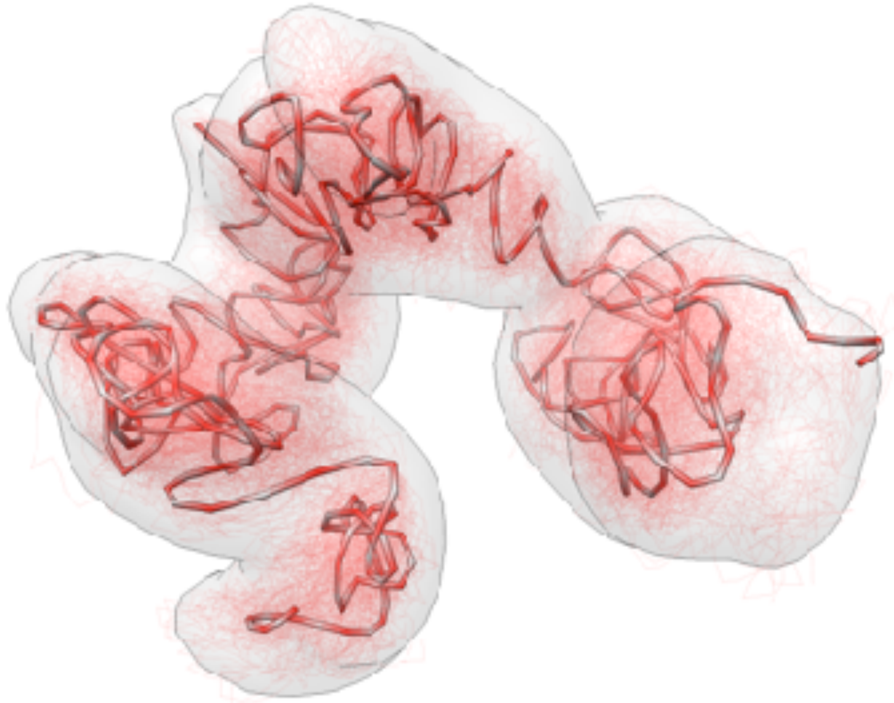
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).

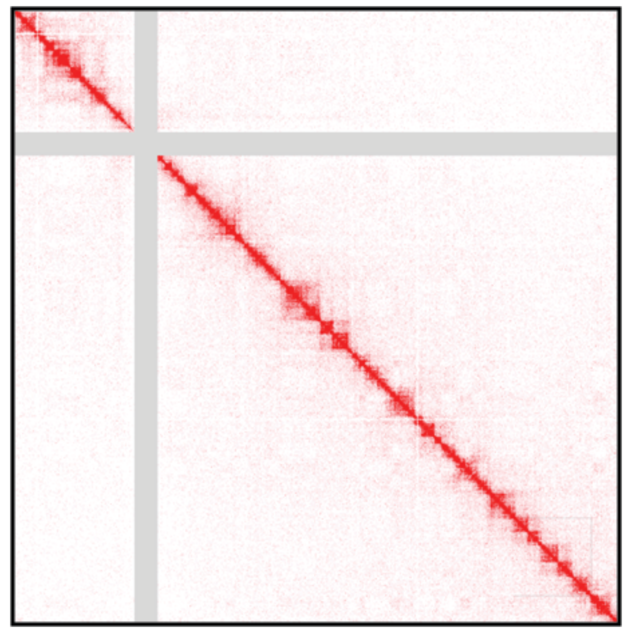




Biomolecular structure determination

2D-NOESY data



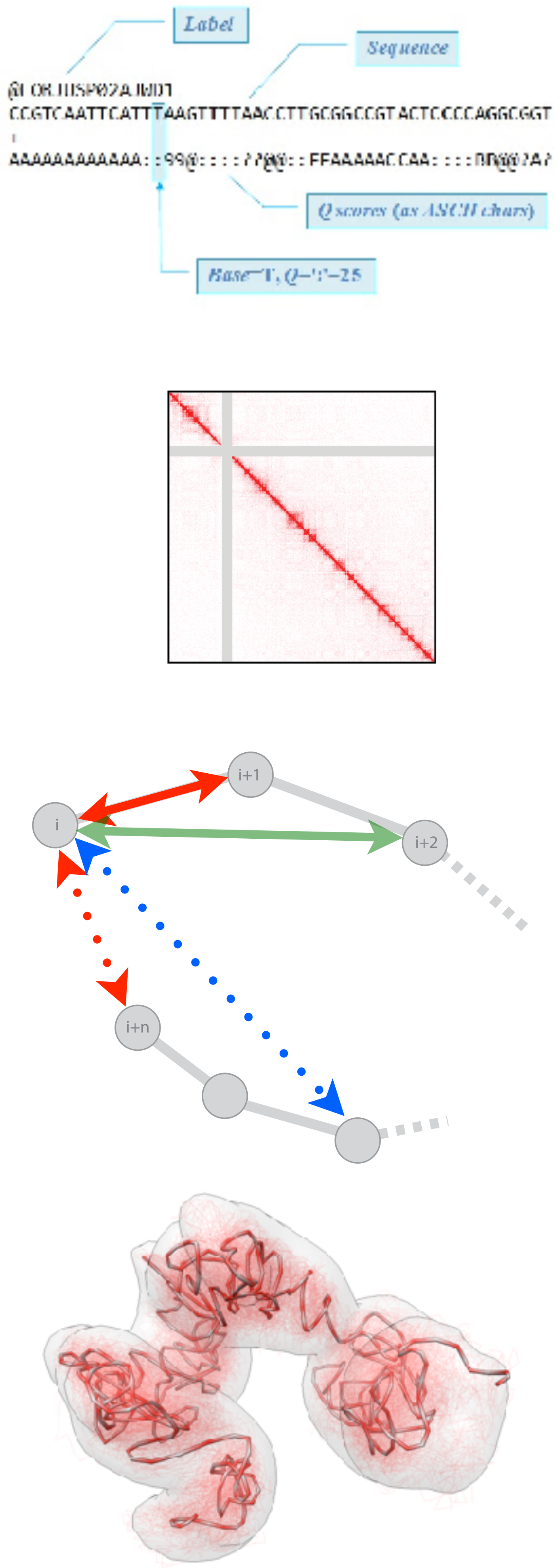


Chromosome structure determination

3C-based data

TADbit

http://3DGenomes.org

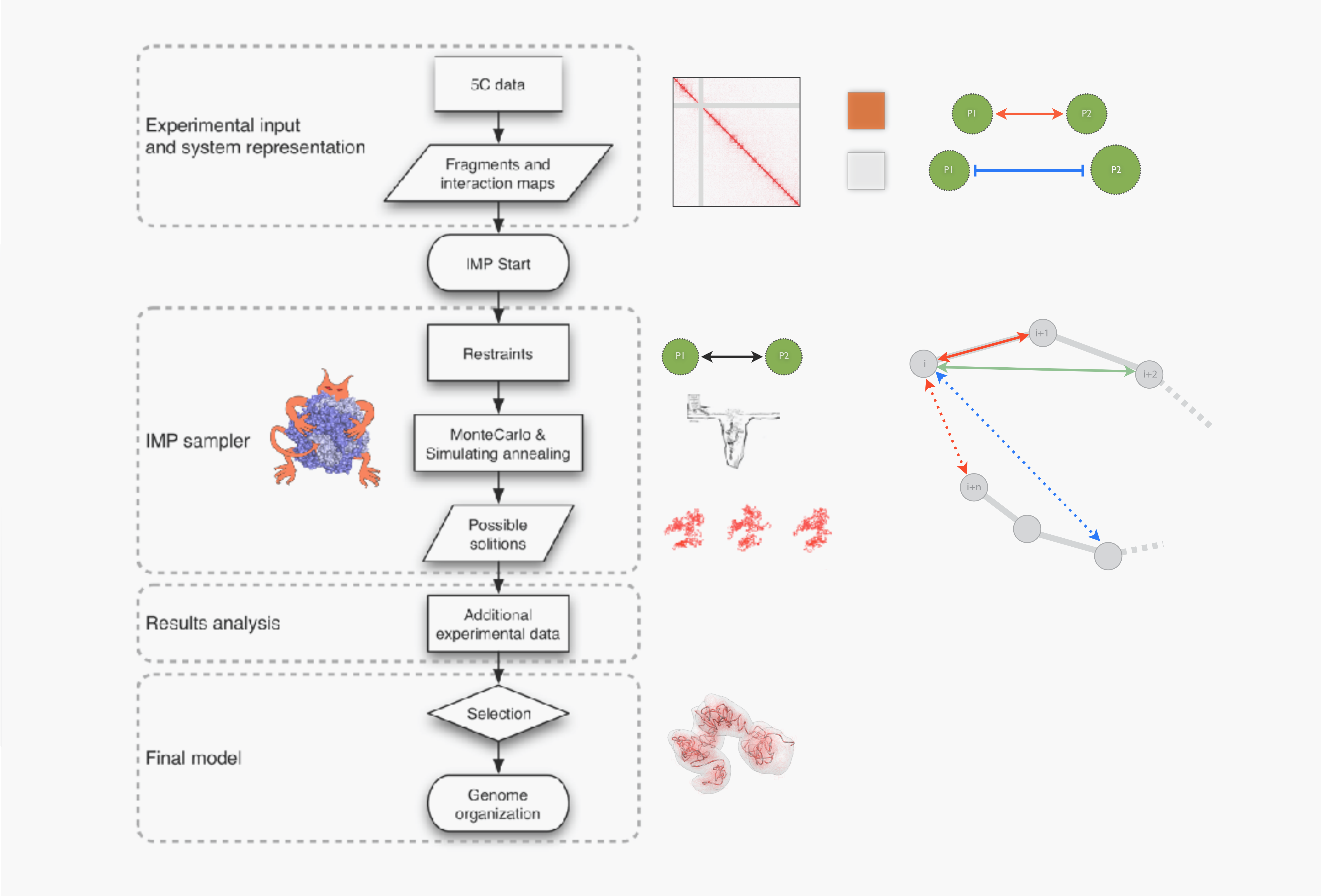


FastQ files to Maps

Map analysis

Model building

Model analysis



Experimental input and system representation

5C data

Fragments and interaction maps

IMP Start

IMP sampler

Restraints

MonteCarlo & Simulating annealing

Possible solutions

Results analysis

Additional experimental data

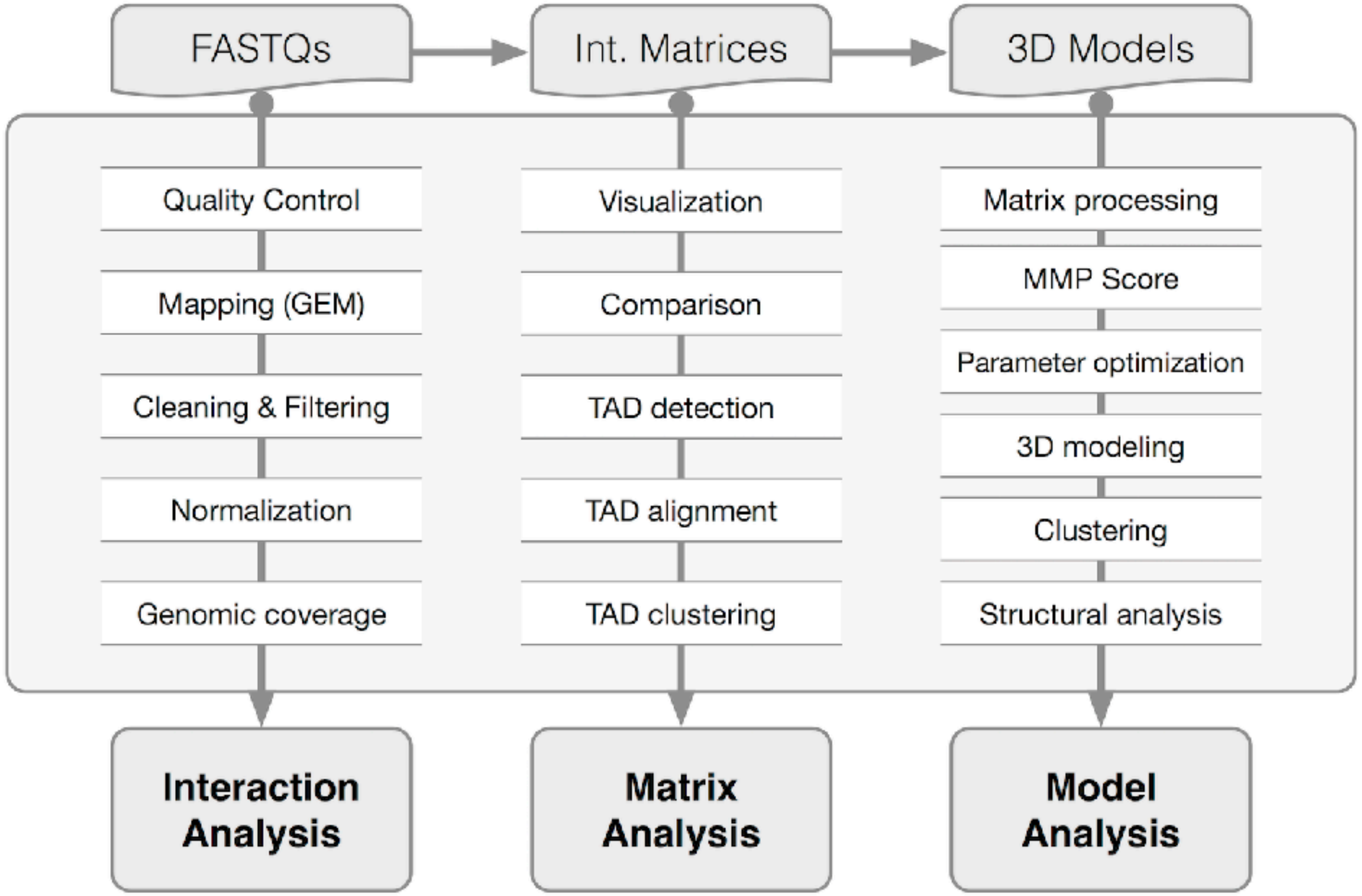
Final model

Selection

Genome organization

TADbit

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



FASTQs → Int. Matrices → 3D Models

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

Interaction Analysis

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

Matrix Analysis

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

Model Analysis

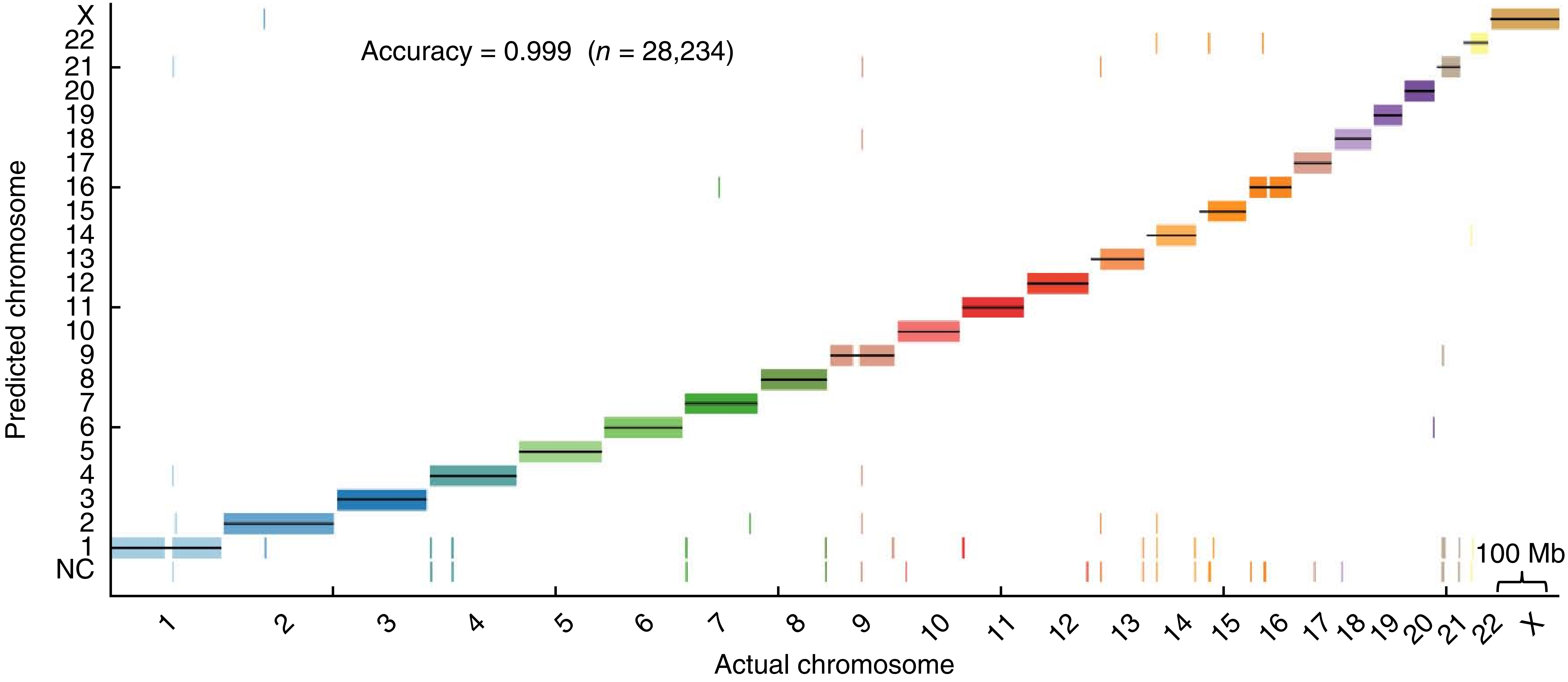
- Baù, D. et al. Nat Struct Mol Biol (2011)
- Umbarger, M. A. et al. Mol Cell (2011)
- Le Dily, F. et al. Genes & Dev (2014)
- Belton, J.M. et al. Cell Reports (2015)
- Trussart M. et al. Nature Communication (2017)
- Cattoni, D. et al. Nature Communication (2017)
- Stadhouders R. et al. Nature Genetics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

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


SIDE EFFECTS

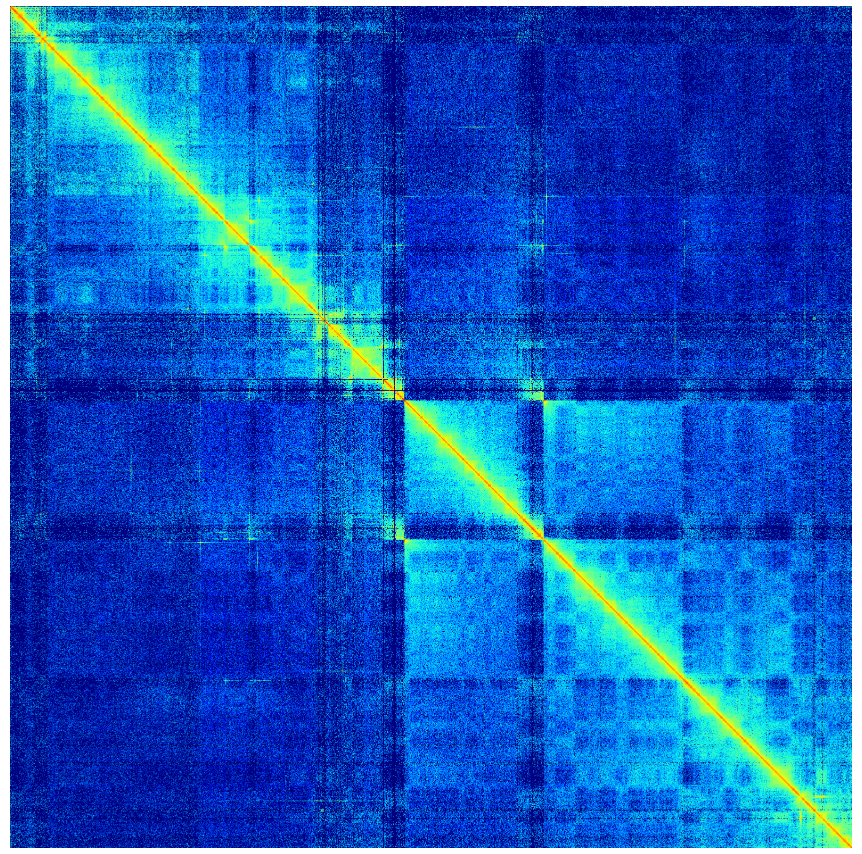
Hi-C for de-novo assembly

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

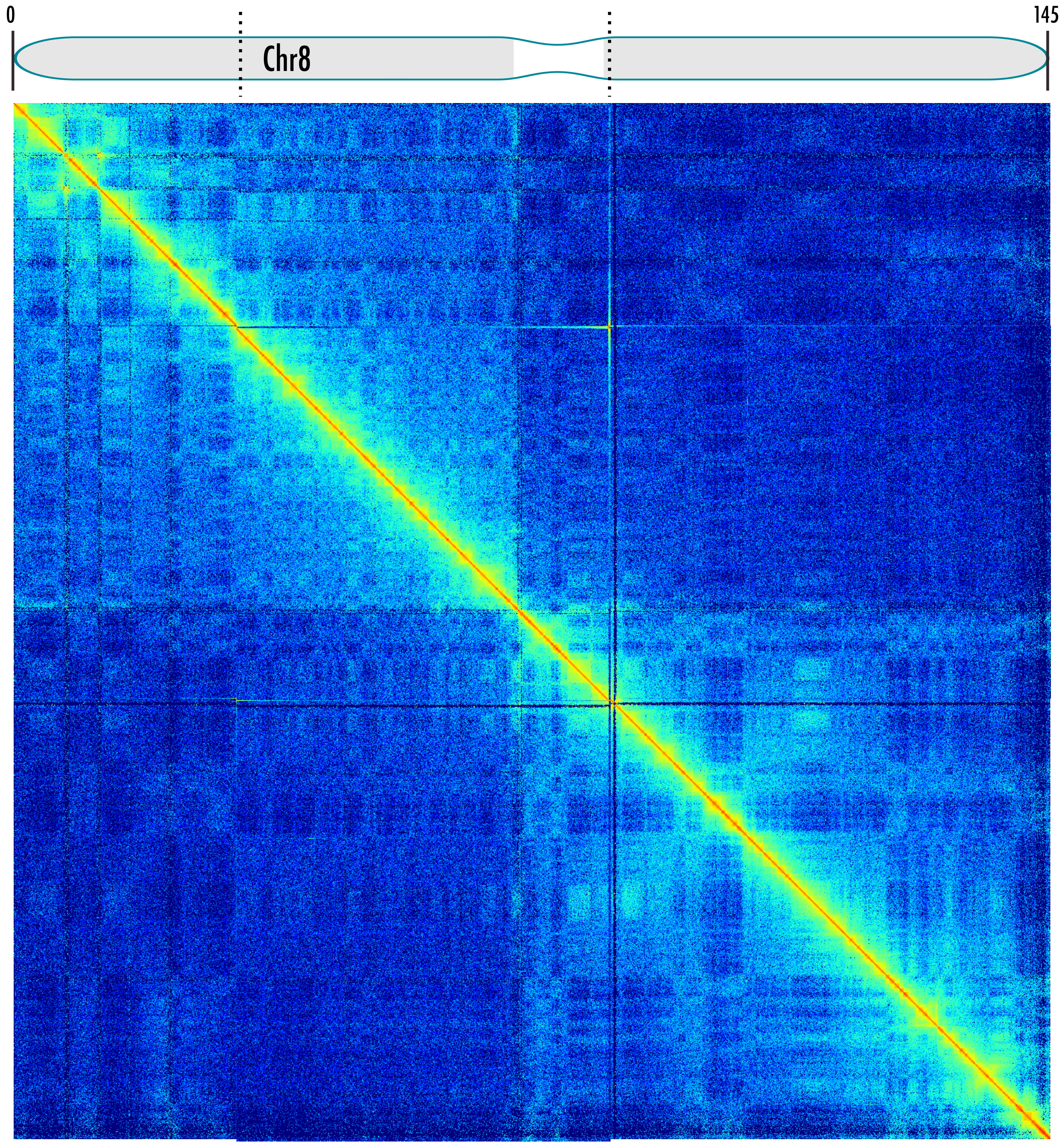


Assembly error detection

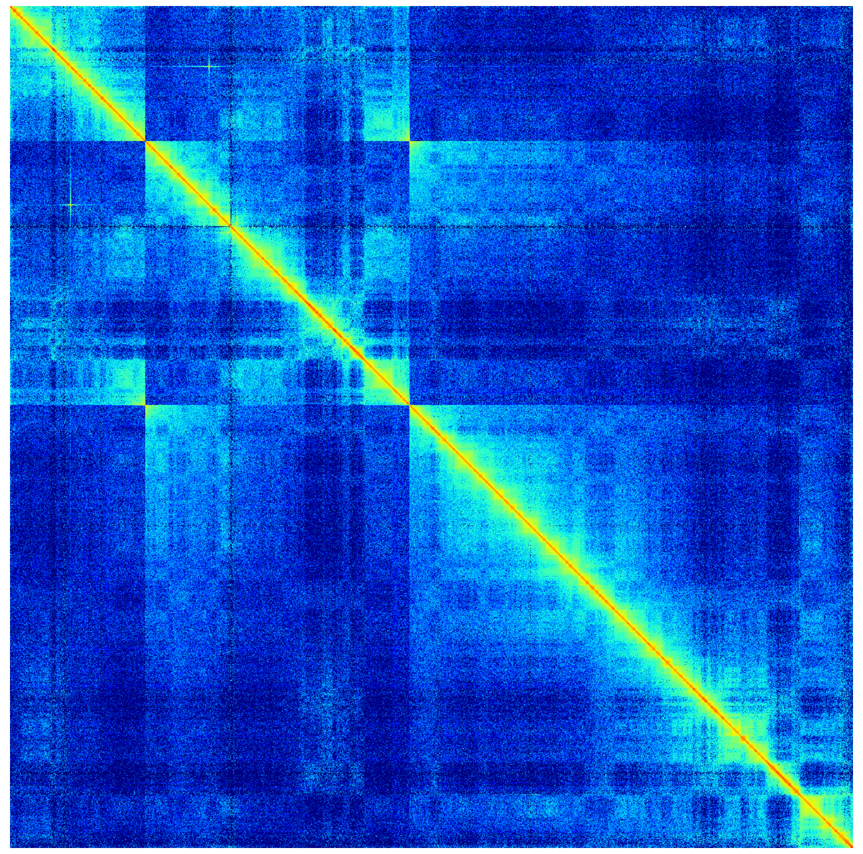
Chr 7



Chromosome 8 Gorilla



Chr 12



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

Hi-C for meta genomics

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

Romain Koszul

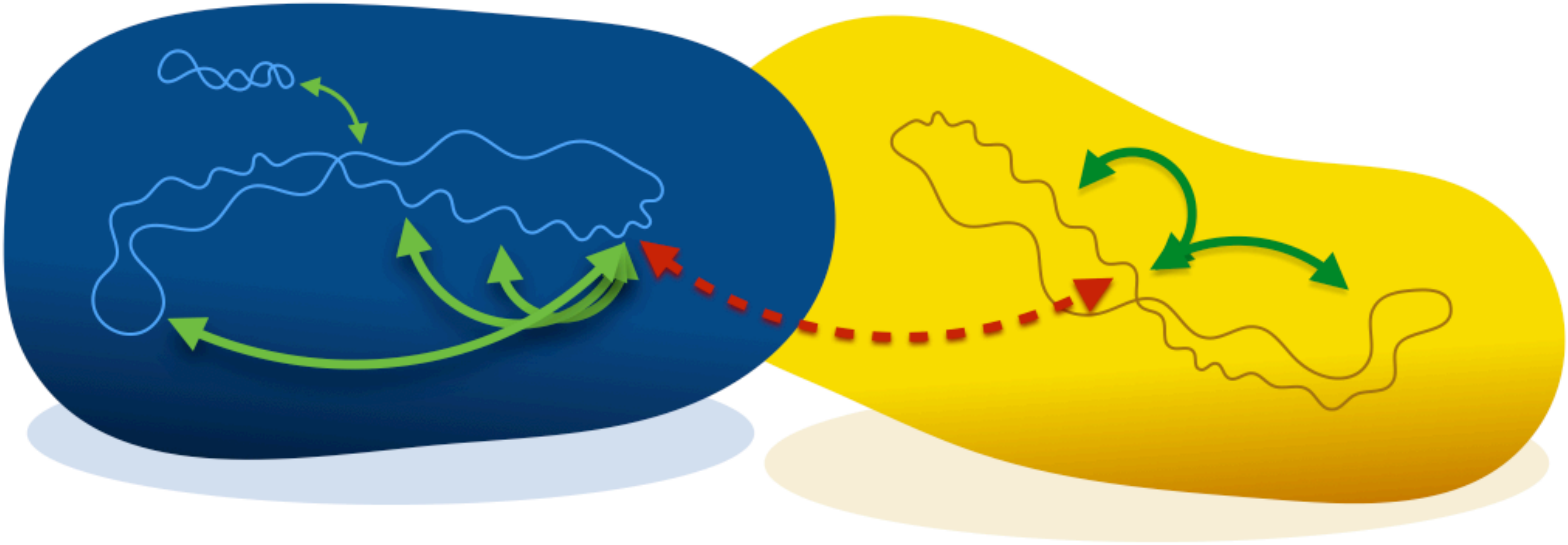
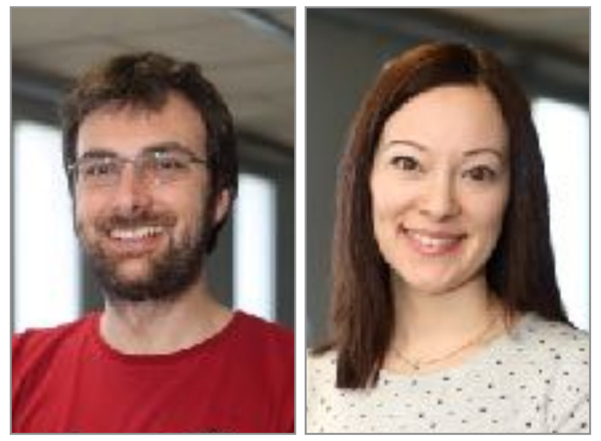


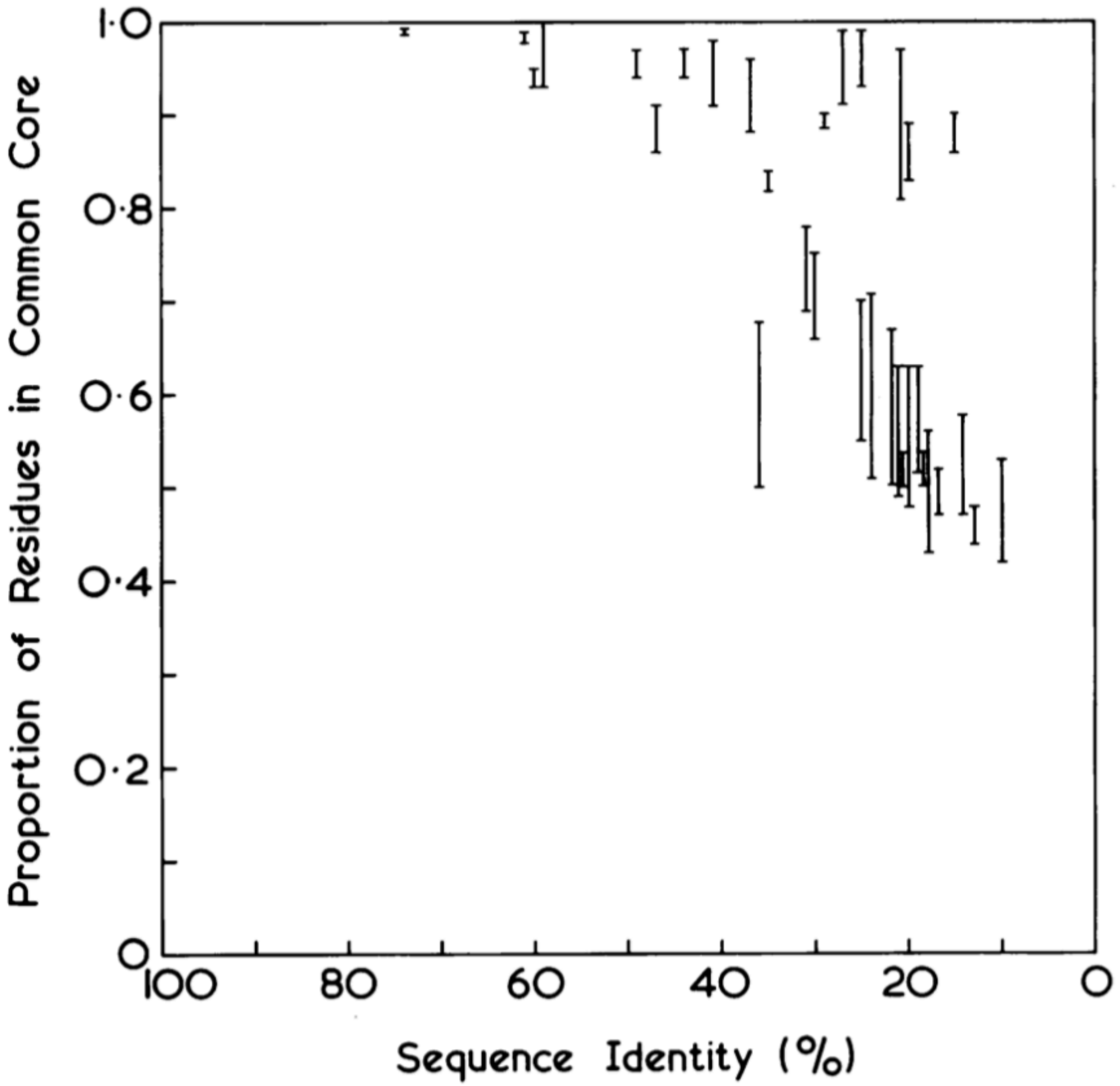

Photo by David Oliete - www.davidoliete.com

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

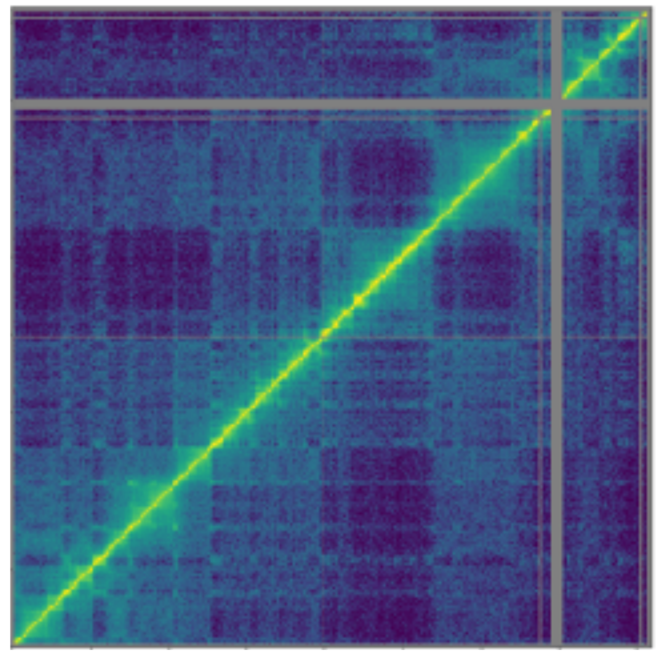


François Serra
Yasmina Cuartero
with Marquès Lab (UPF, Barcelona)

Unpublished



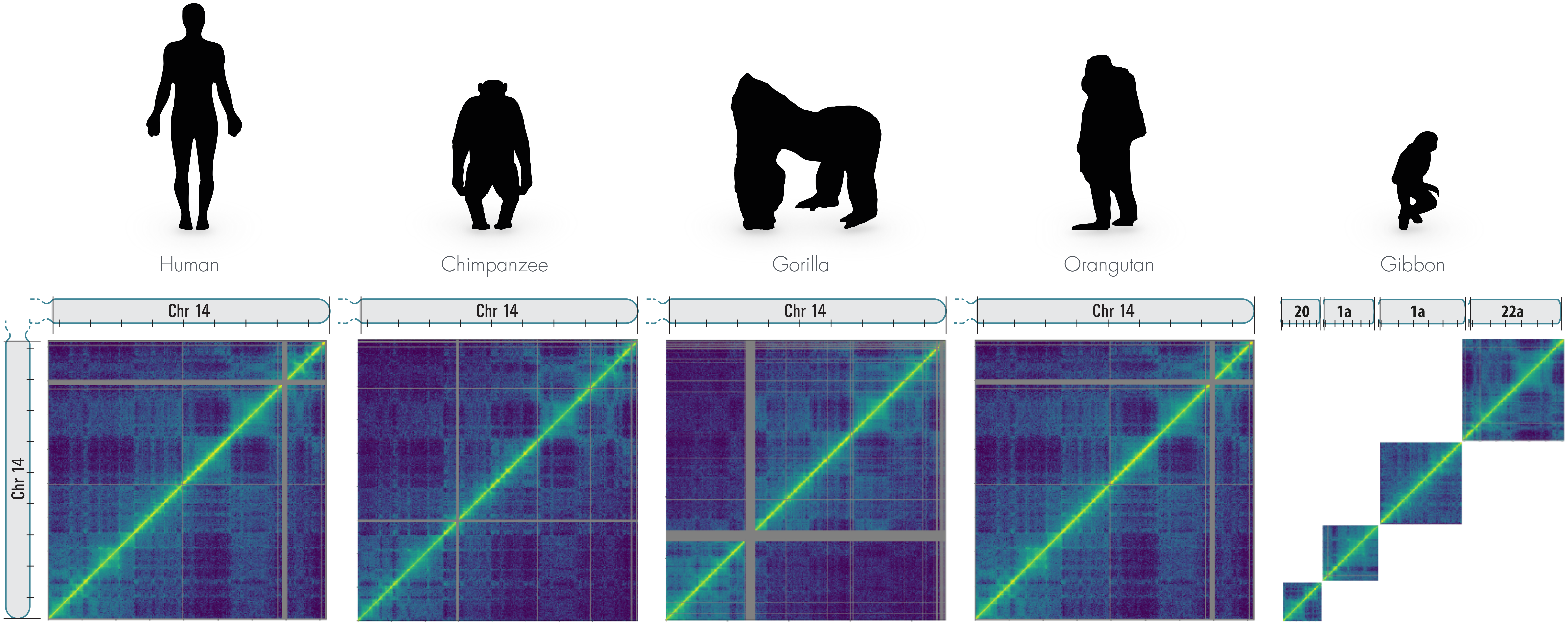
C. Chothia & A. Lesk (1986) EMBO J. 5(4):823-826

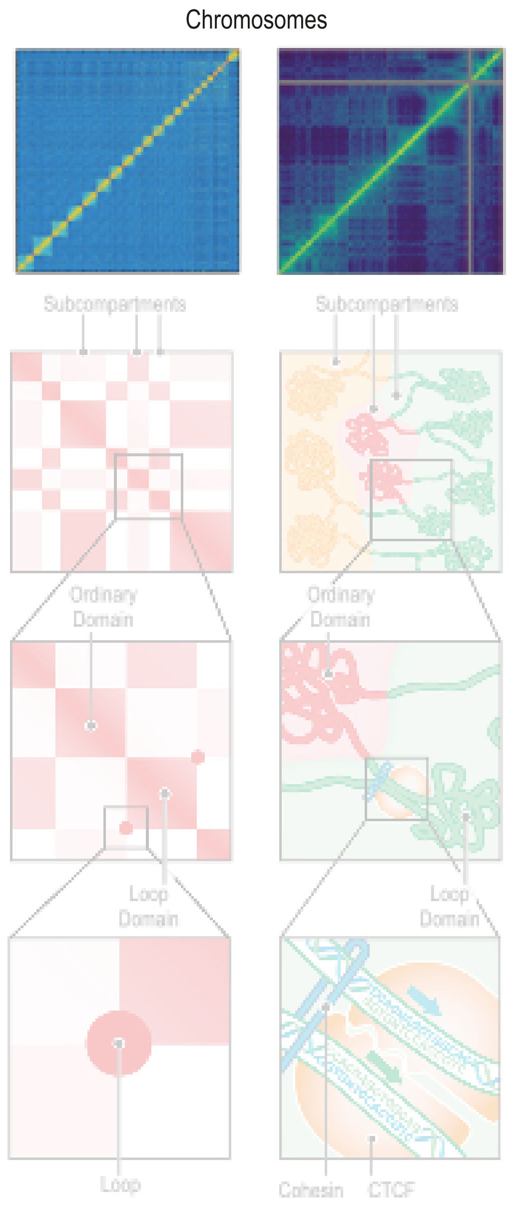


- Proteome is 10sM AAs in 10,000s of proteins
- 20 AA in average ~500 AA length
- ~150 structural AA for ~5 AA for function
- MM+QM
- No confinement
- Dynamics

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

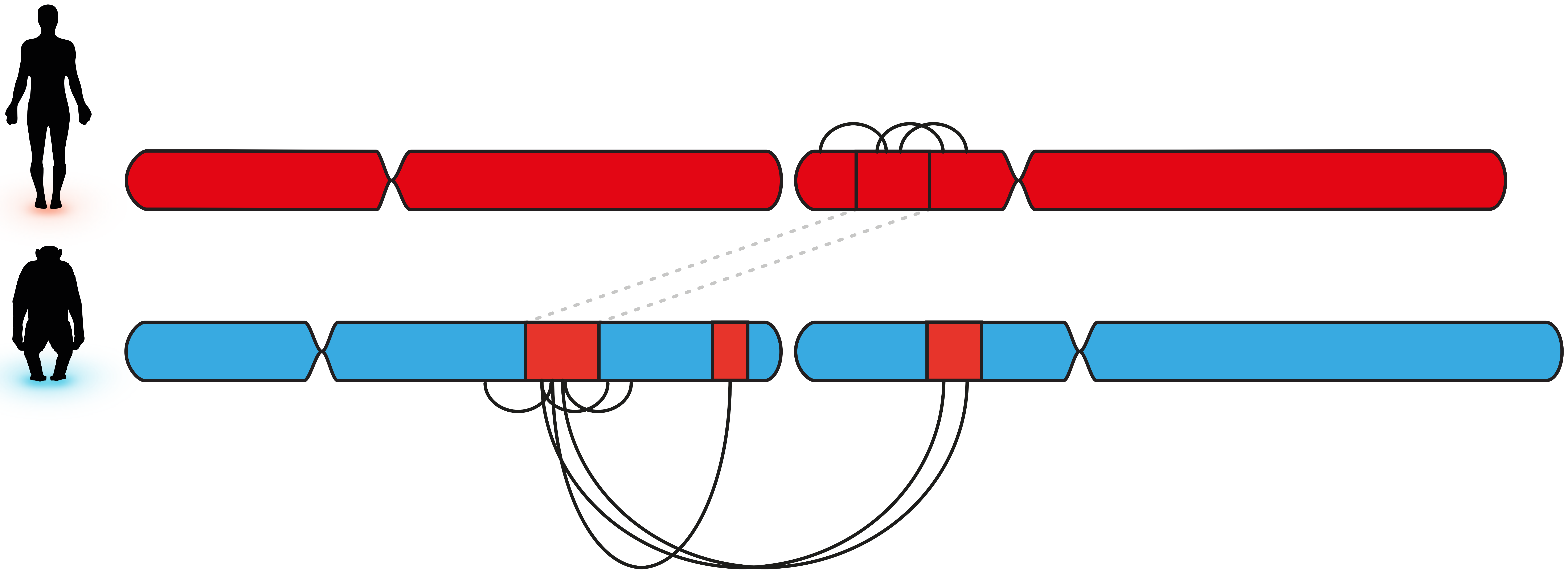
Hi-C matrices from lymphoblasts in seven primates





Synteny breakpoints in 3D

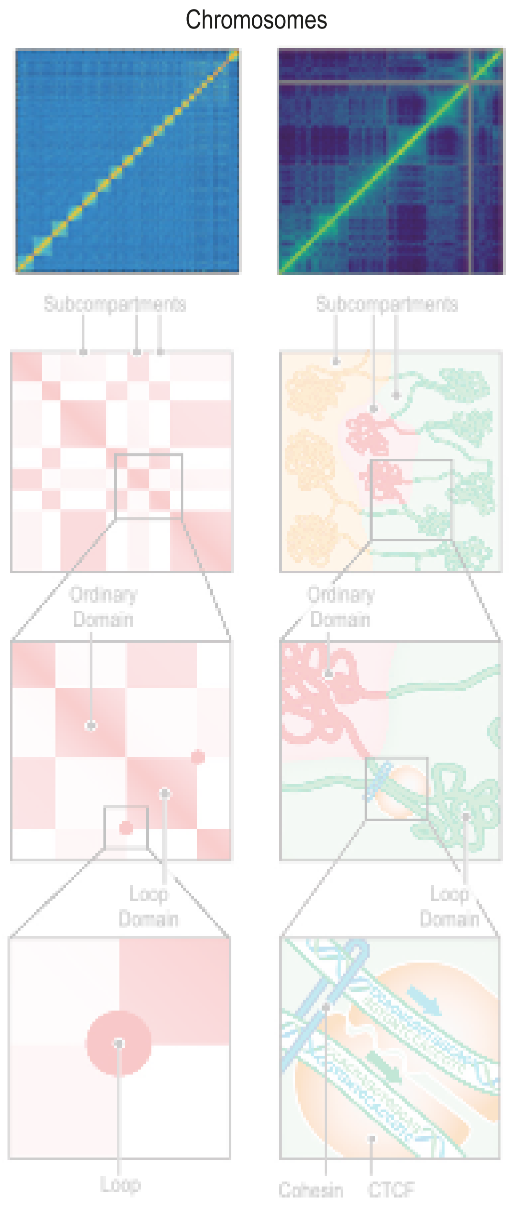
Two regions of at least 100 kb separated by more than 750 kb (including trans chromosomal)
Common detections from Ruiz-Herrera's Lab (@300Kb res) and our lab (ENSEMBLE @1Kb res)

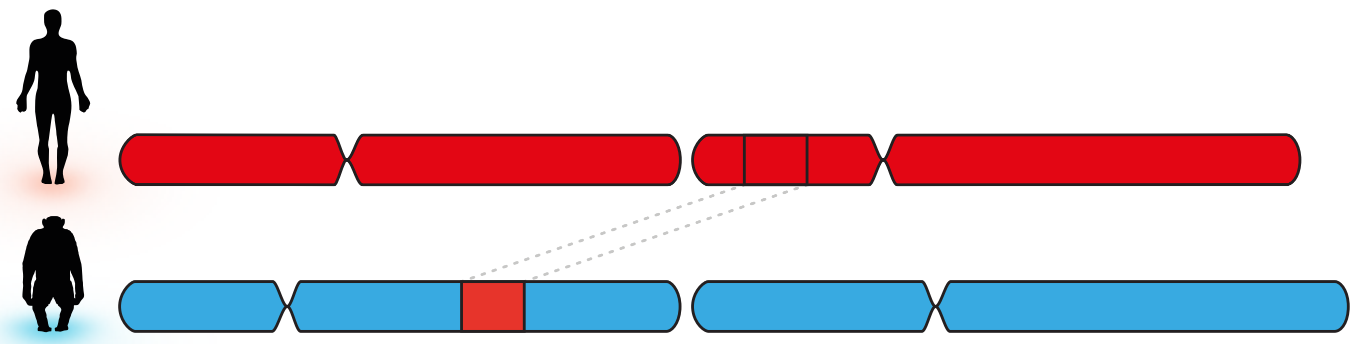


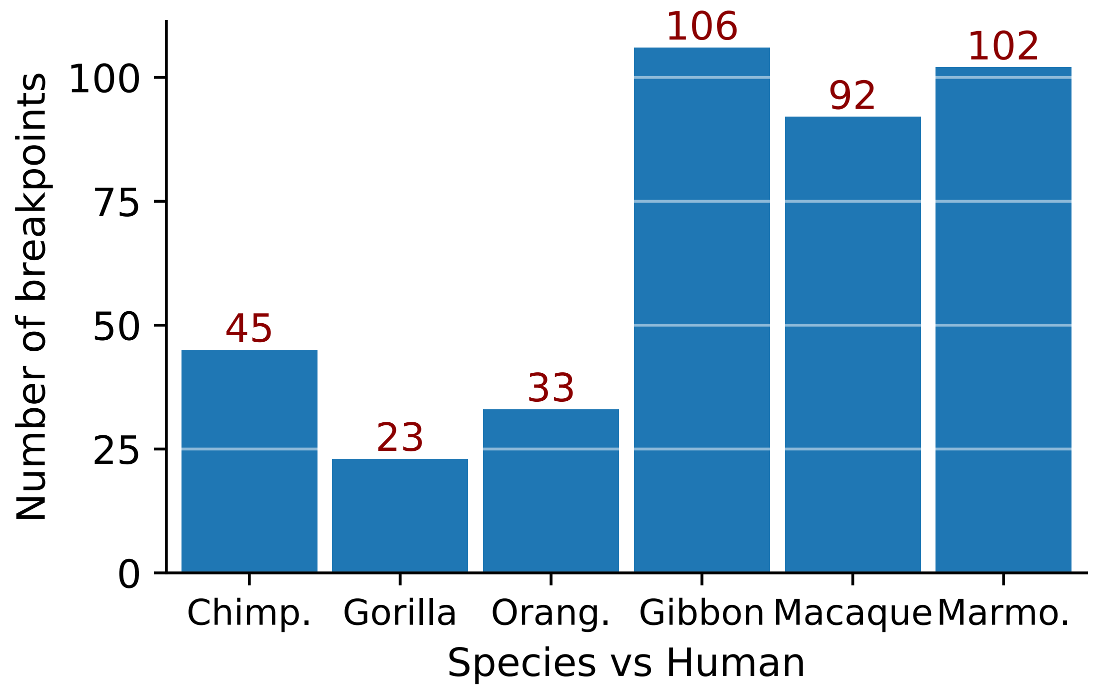
Synteny breakpoints in 3D

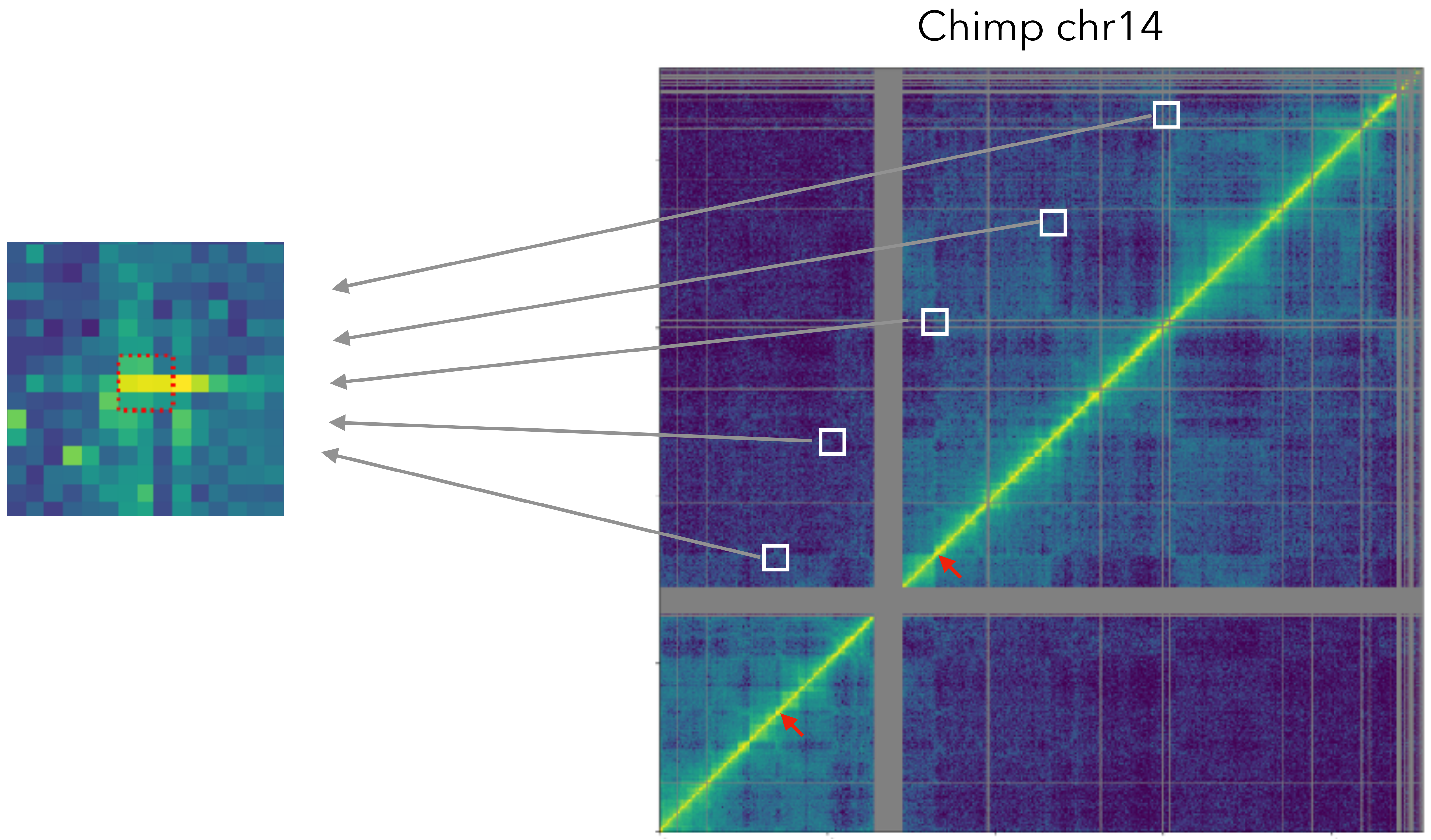
Two regions of at least 100 kb separated by more than 750 kb (including trans chromosomal)

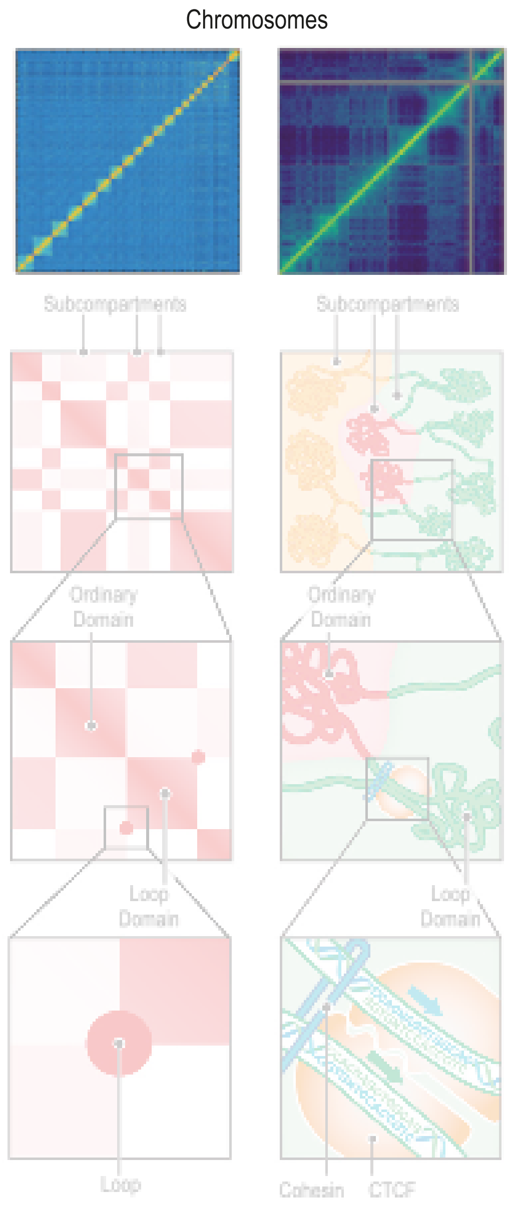
Common detections from Ruiz-Herrera's Lab (@300Kb res) and our lab (ENSEMBLE @1Kb res)







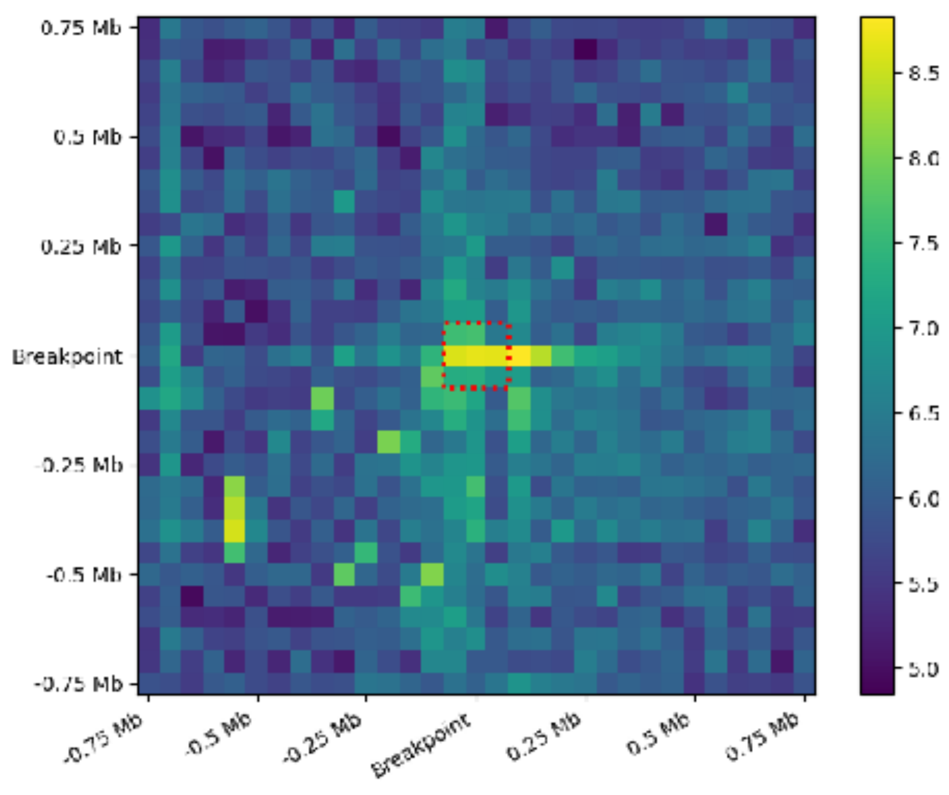

Chimp chr14



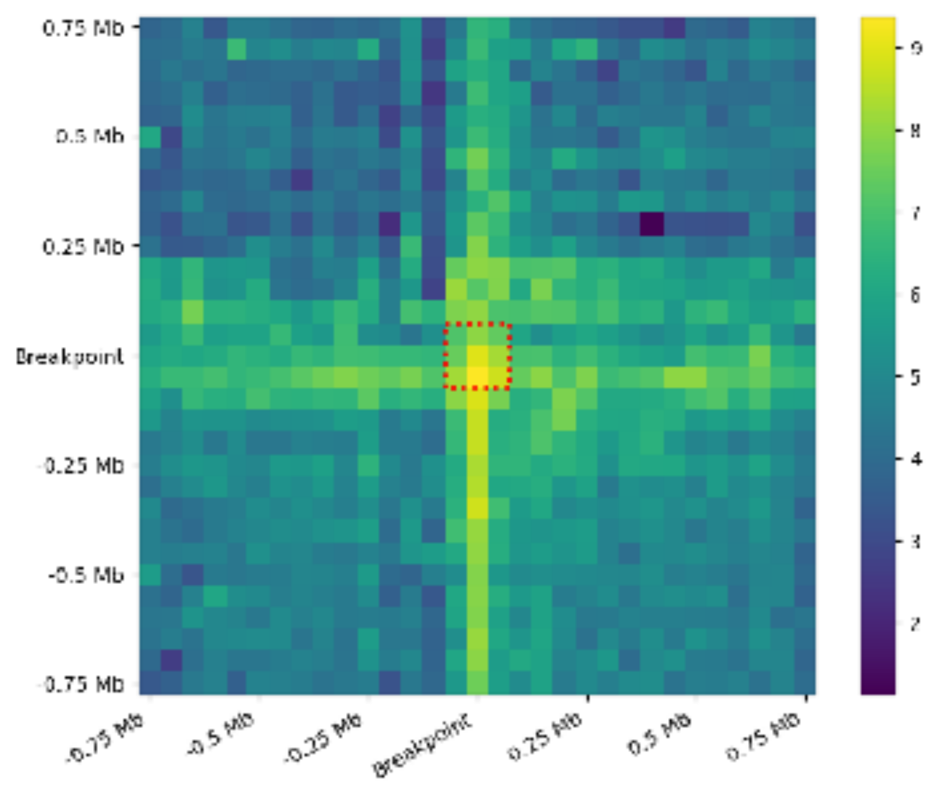
Synteny breakpoints in 3D

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

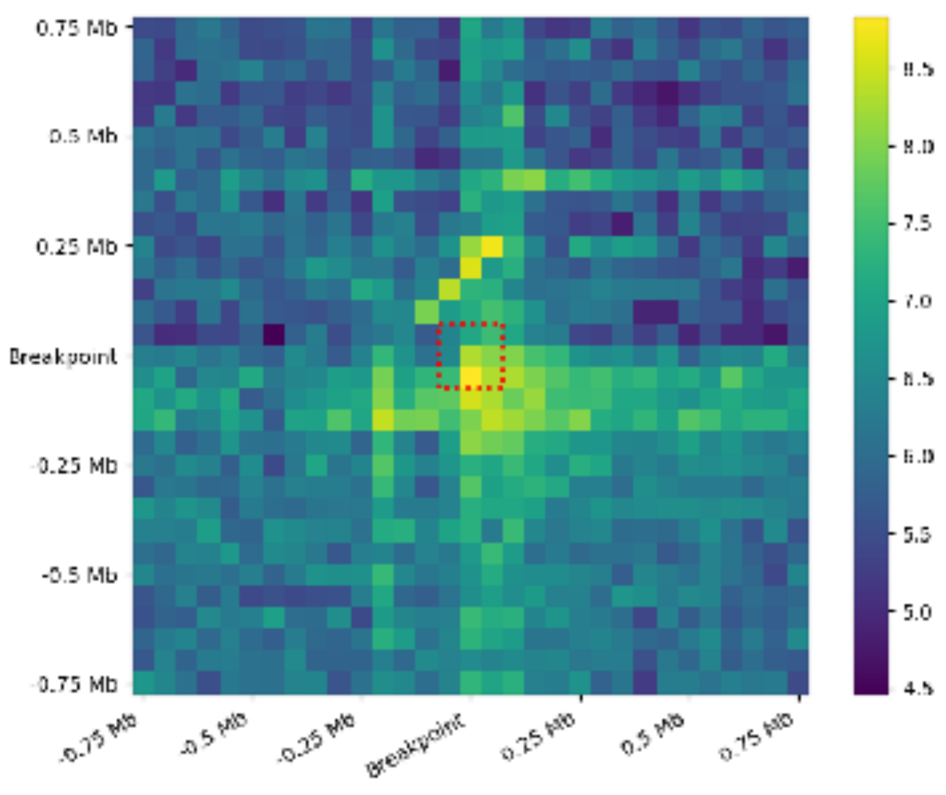
Chimpanzee (62 matrices)



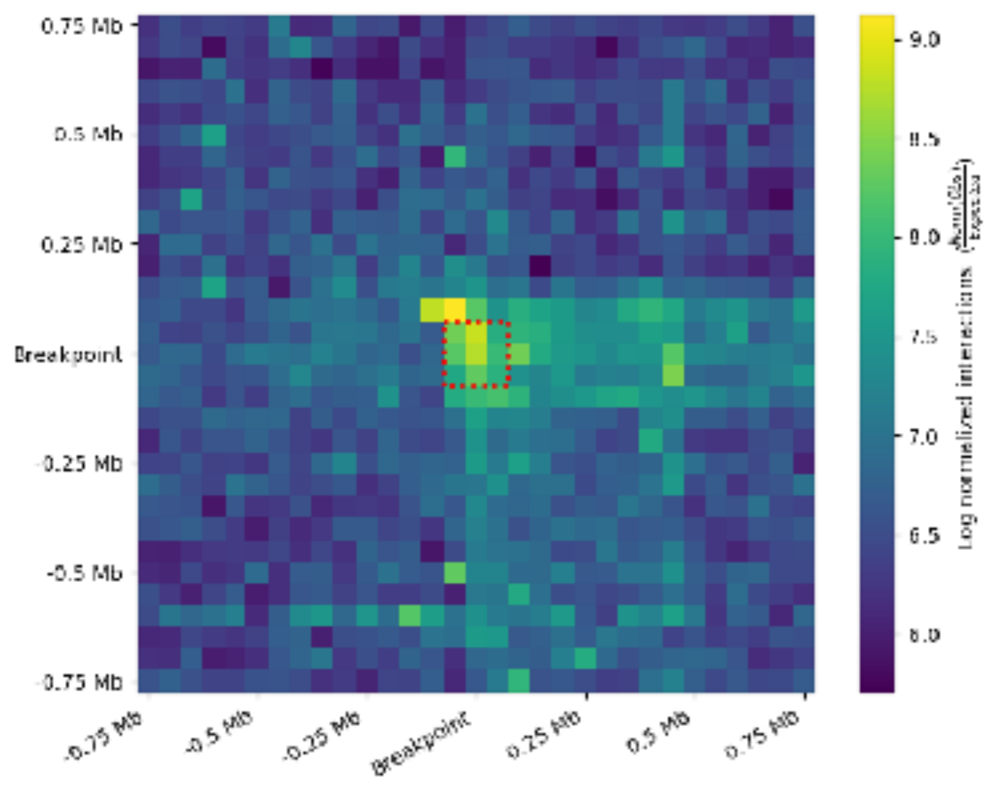
Gorilla (32 matrices)



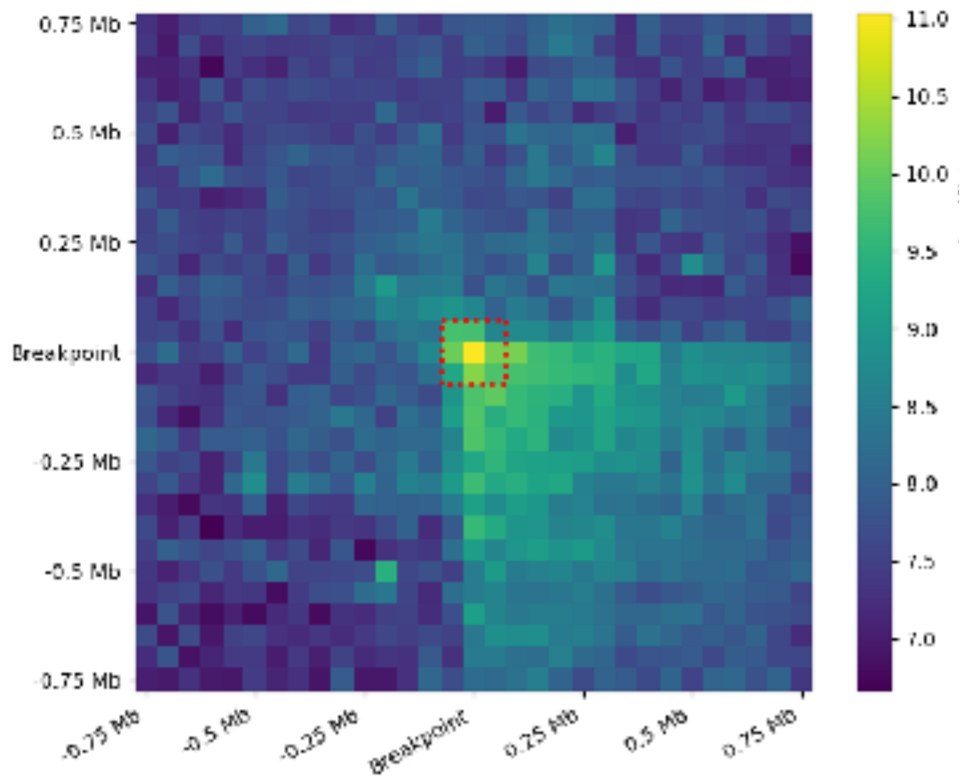
Orangutan (75 matrices)



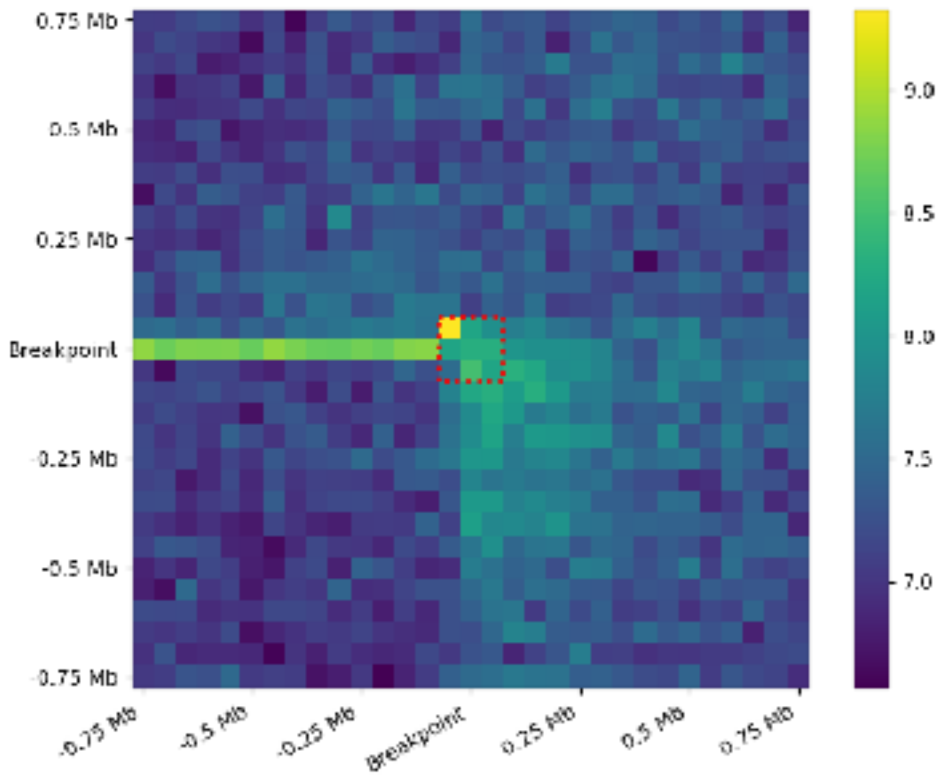
Gibbon (104



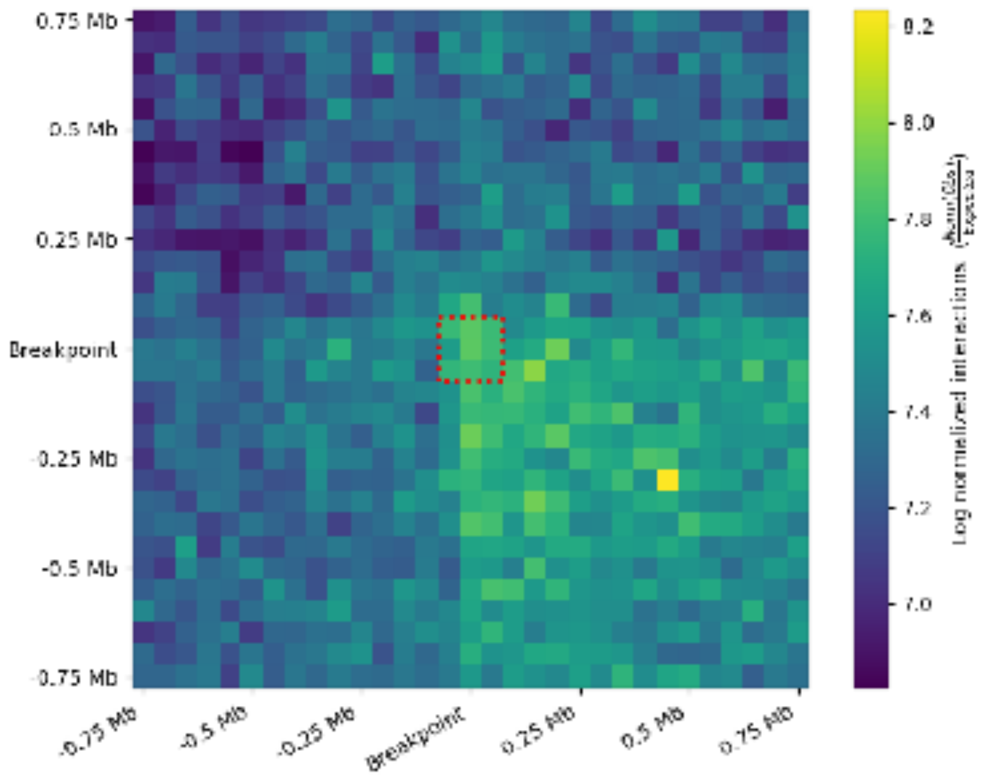
Macaque (134

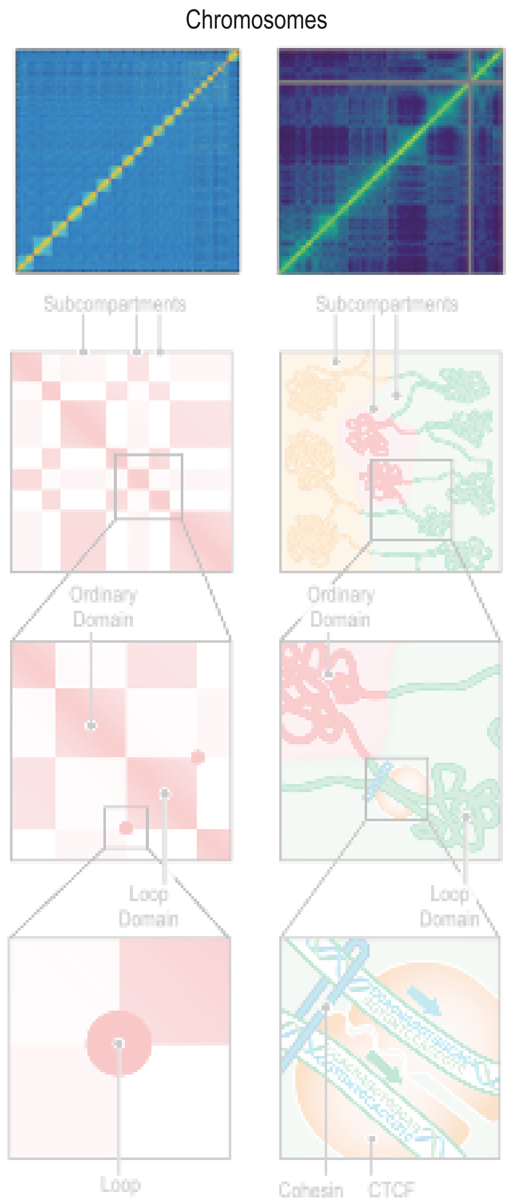


Marmoset (151



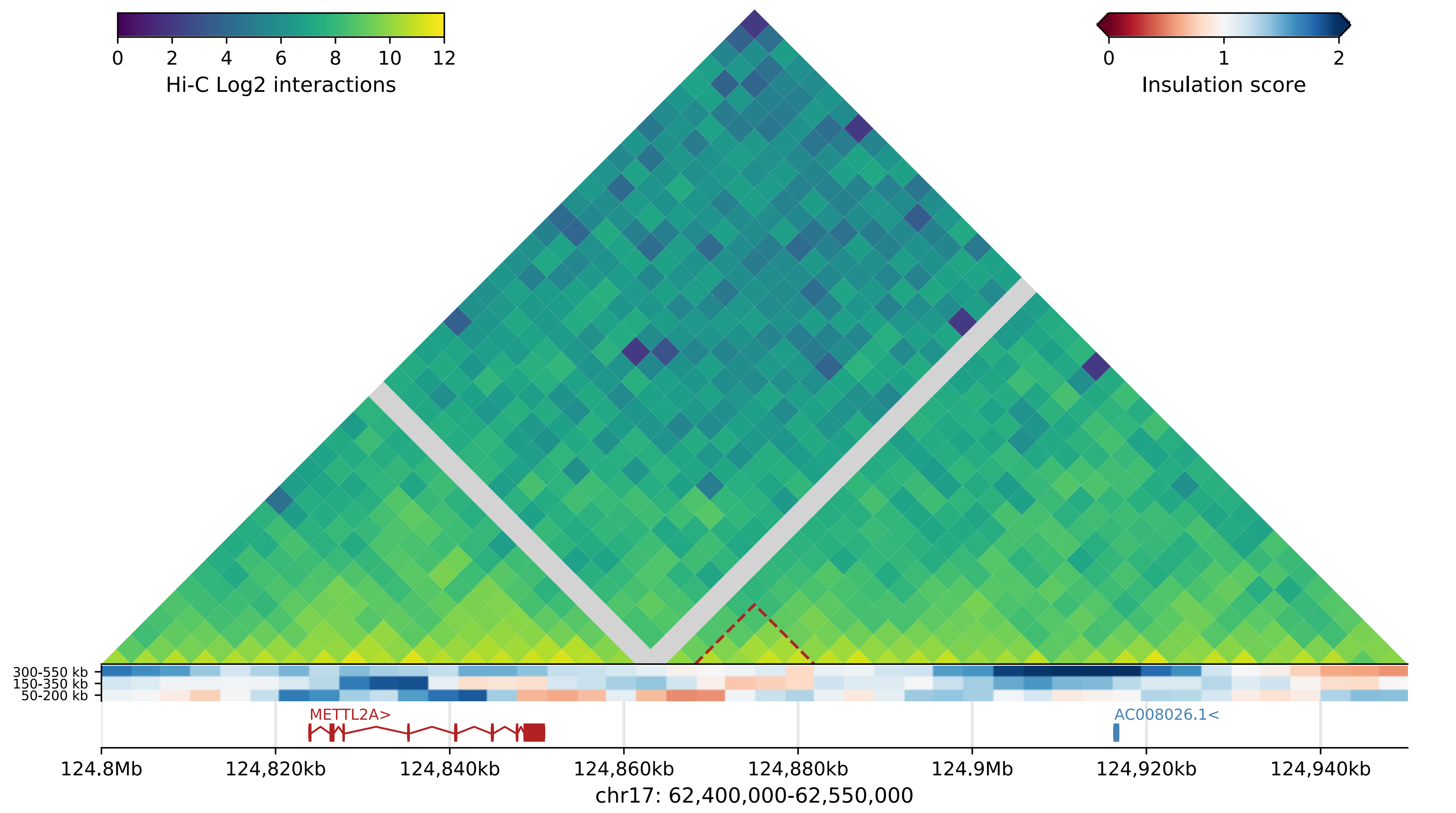
Mouse (199

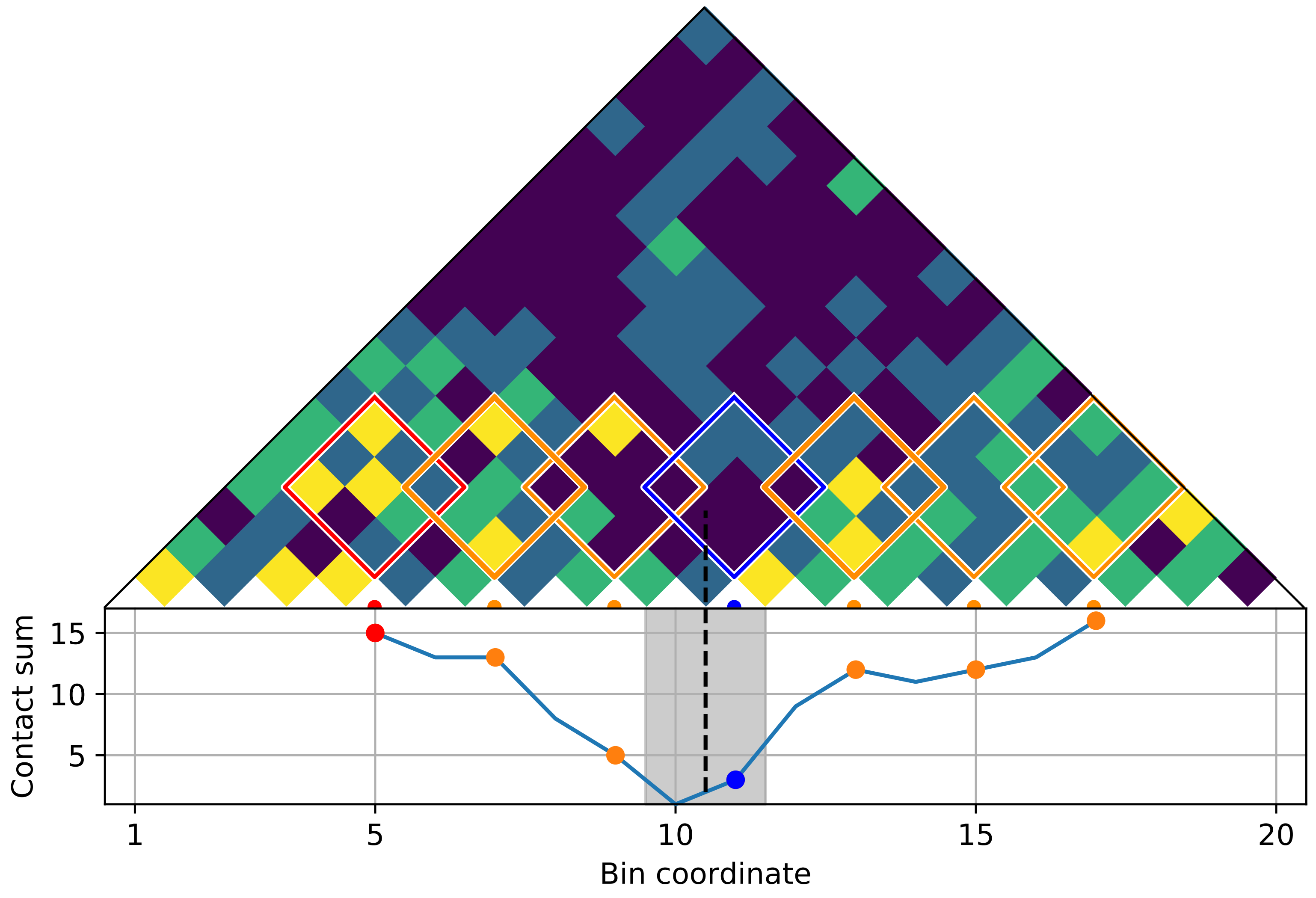


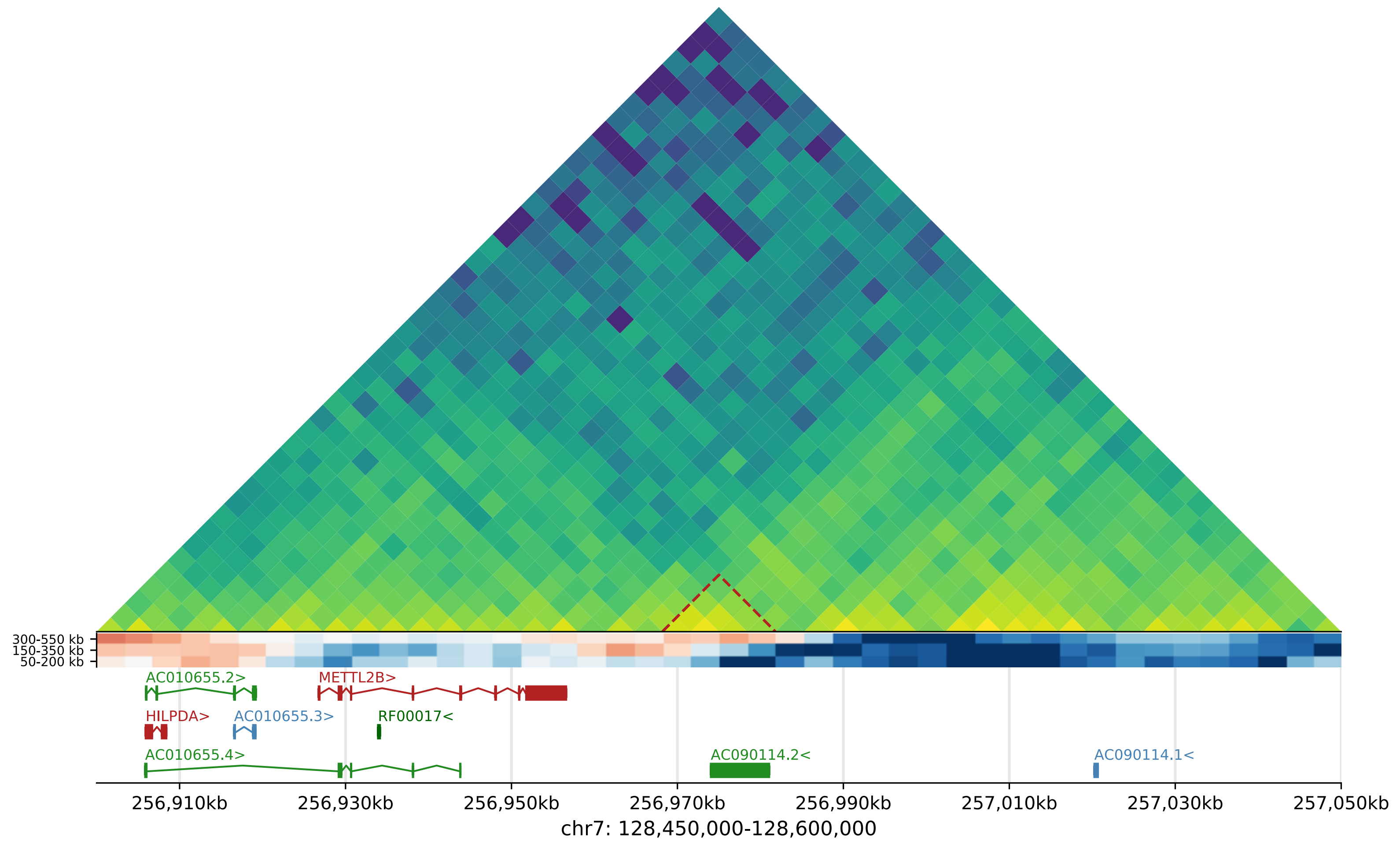


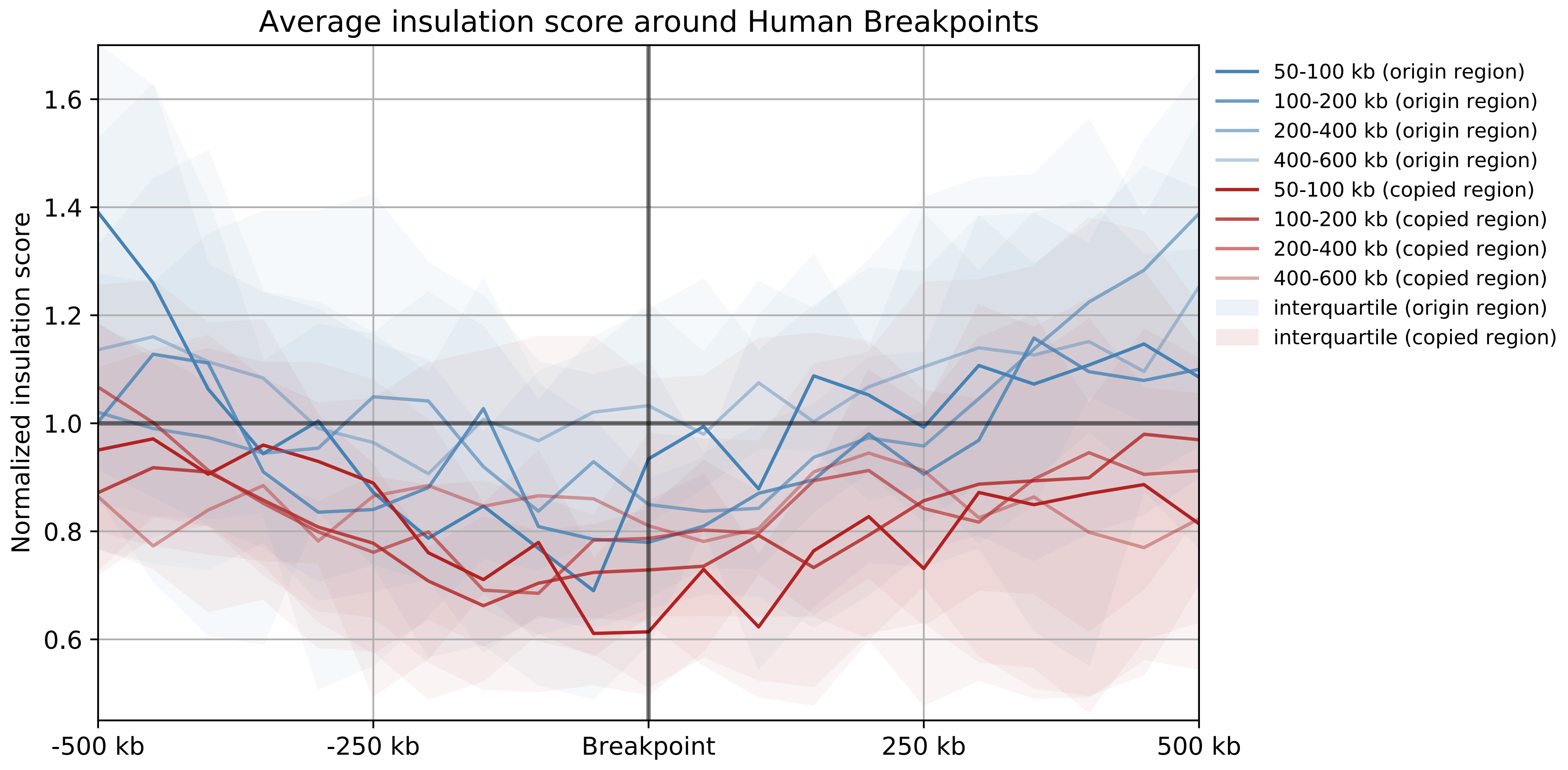
Duplications

Duplicated METTL2 locus in human with respect to primates



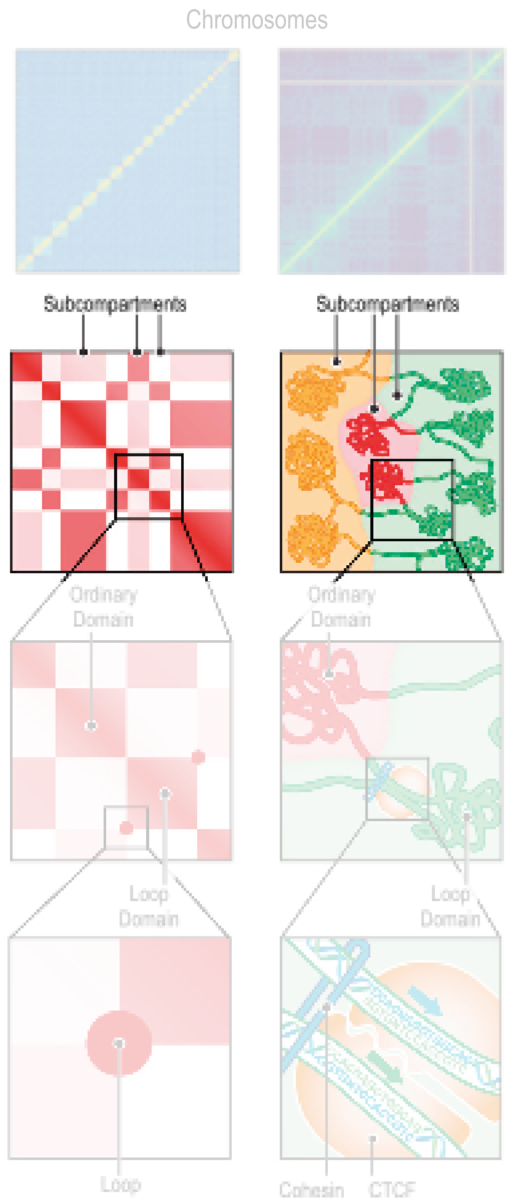


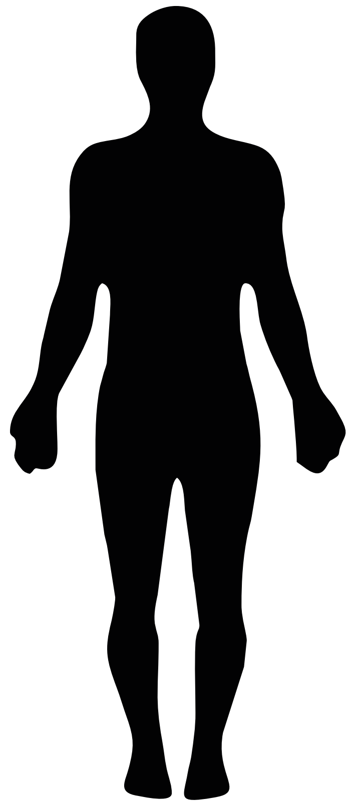


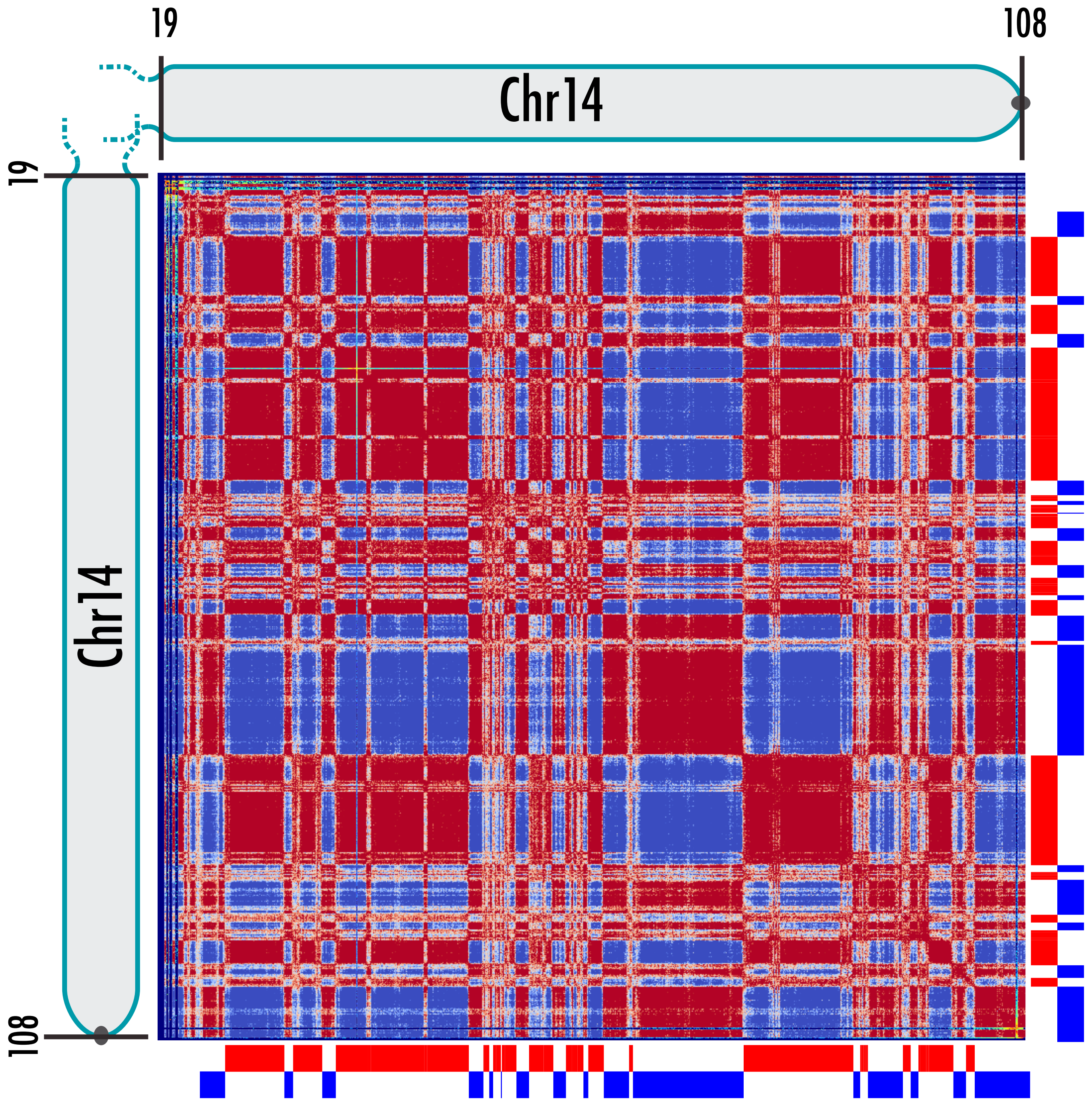


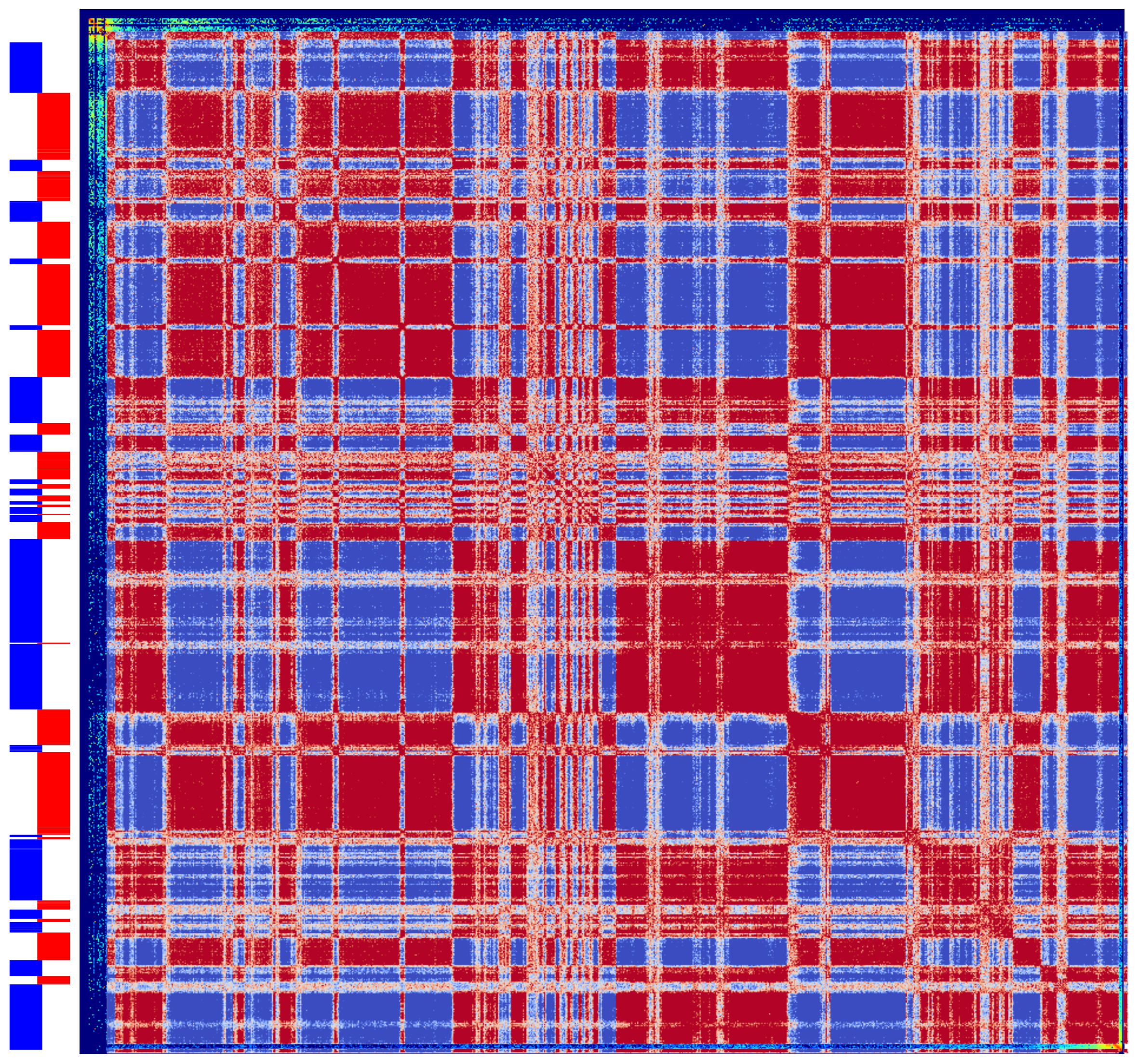
Genome compartments

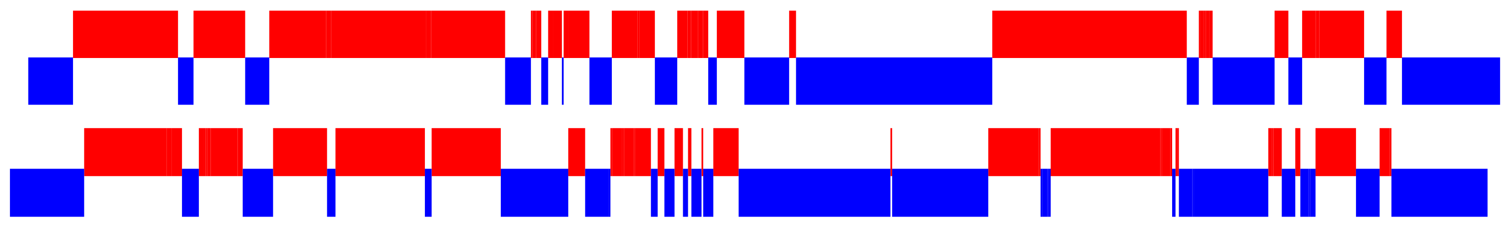
Conservation of the A/B compartments

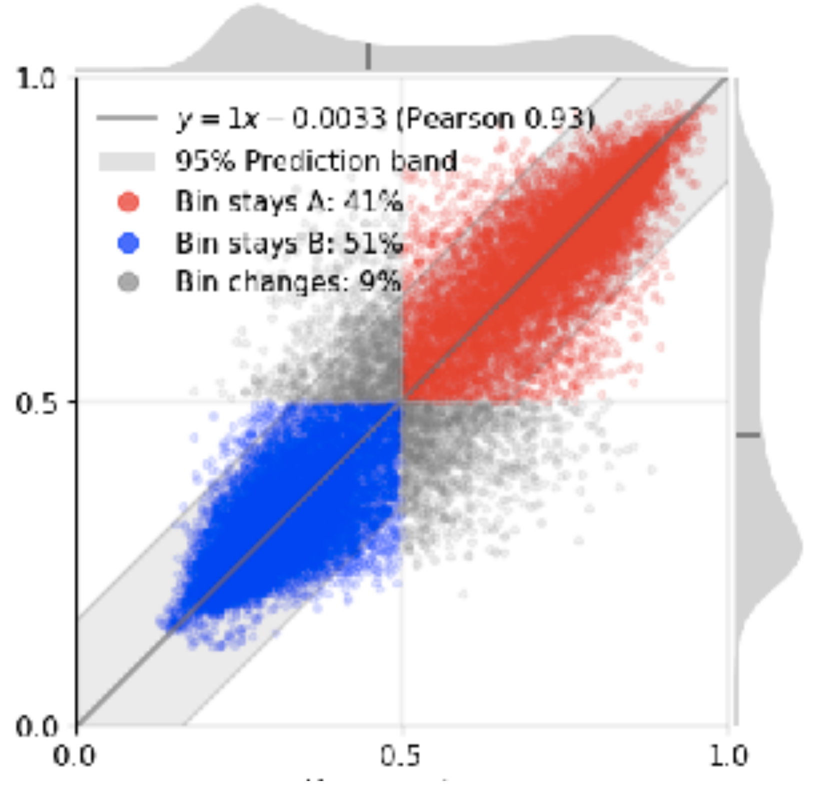


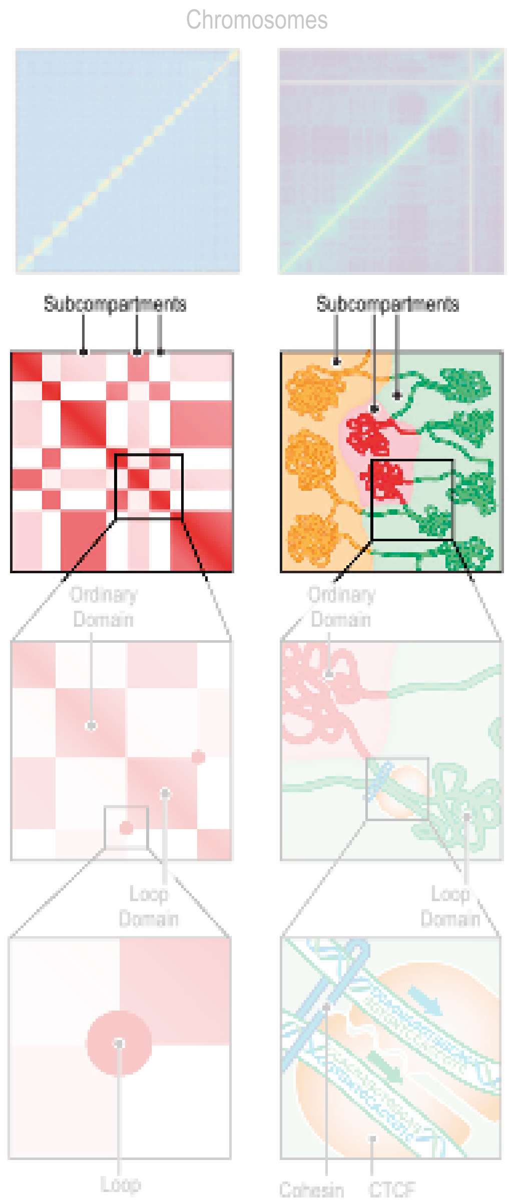






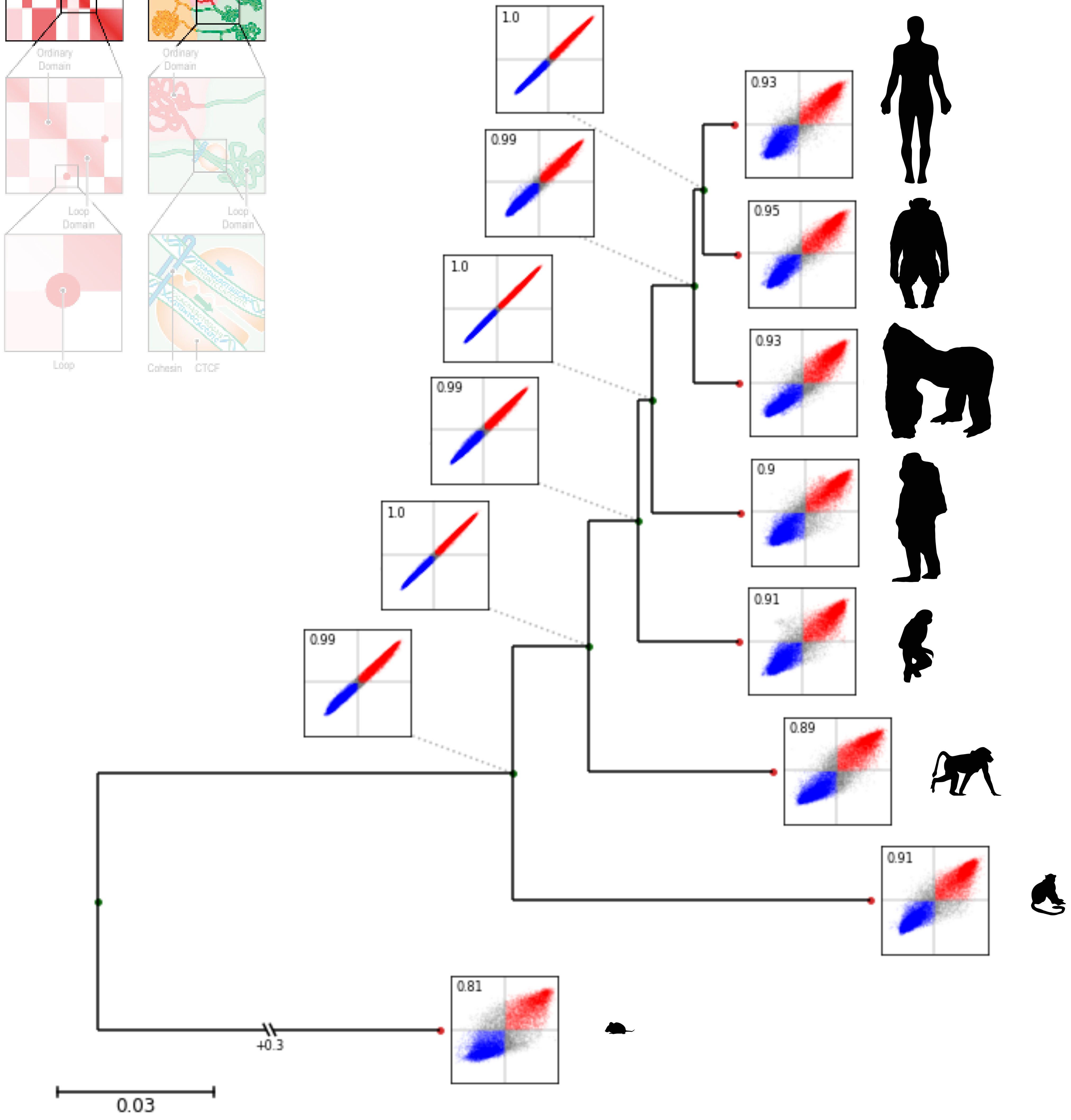


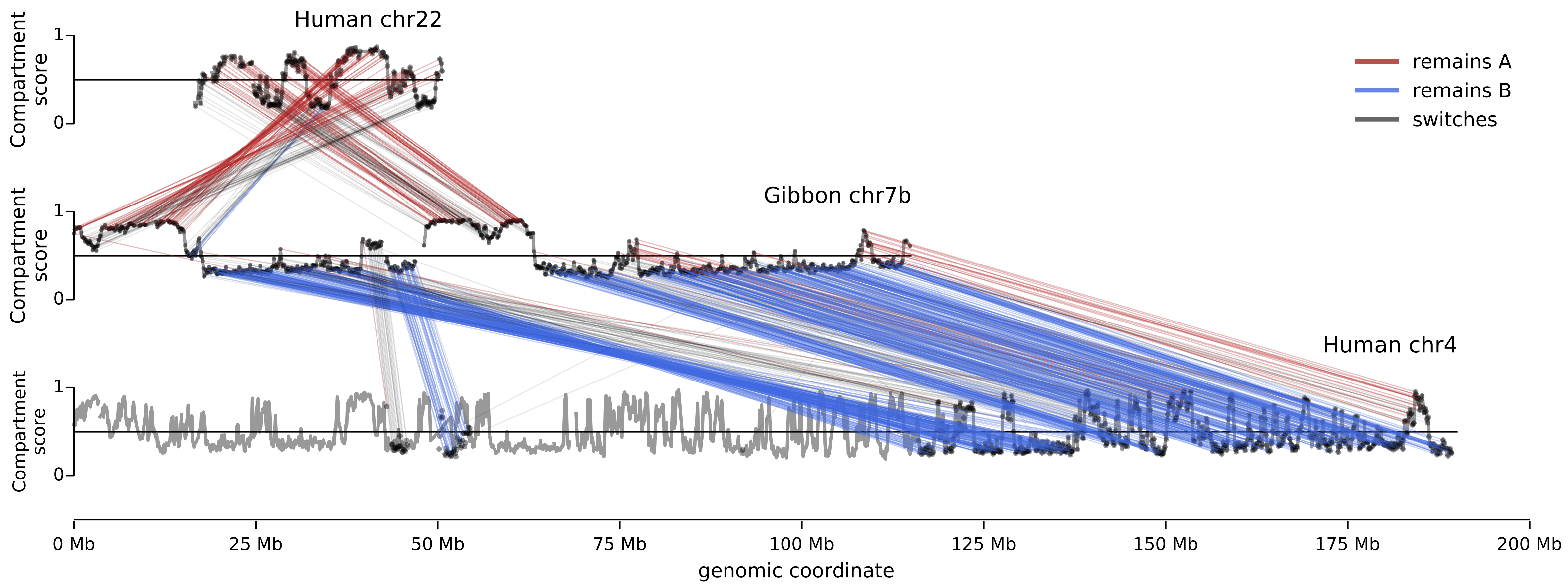


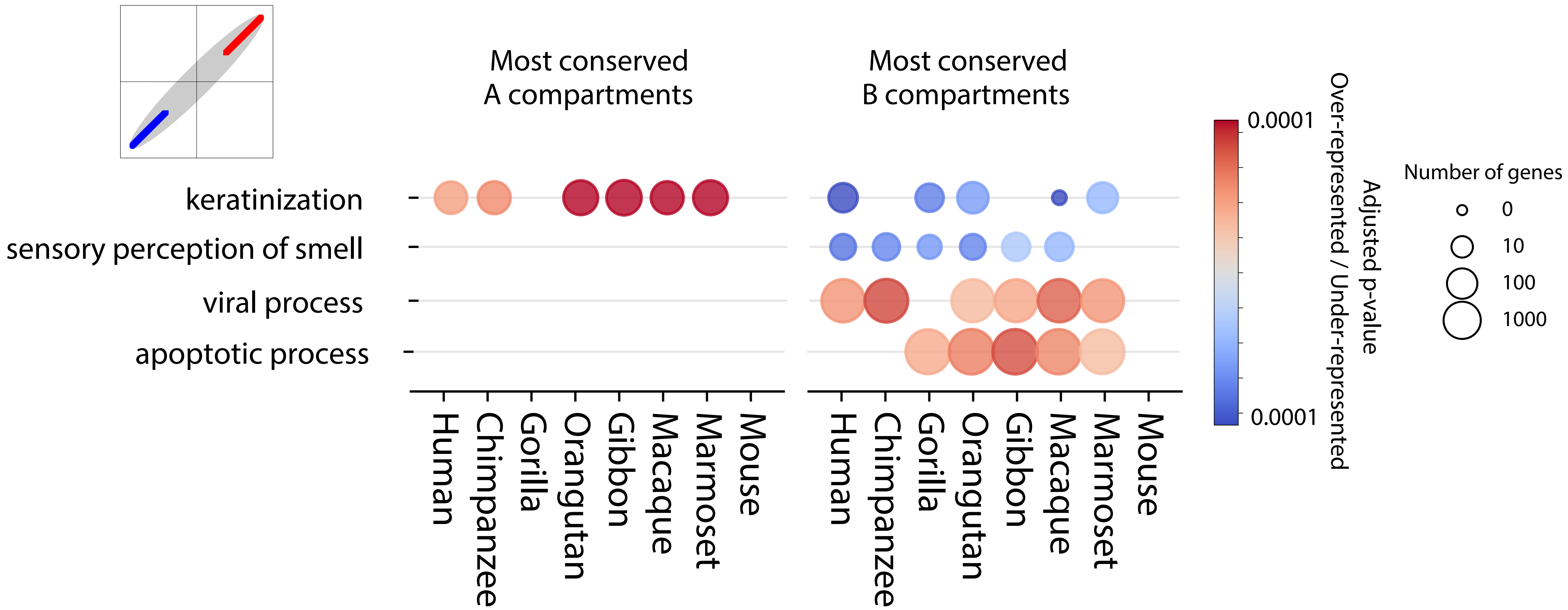


Genome compartments

Conservation of the A/B compartments

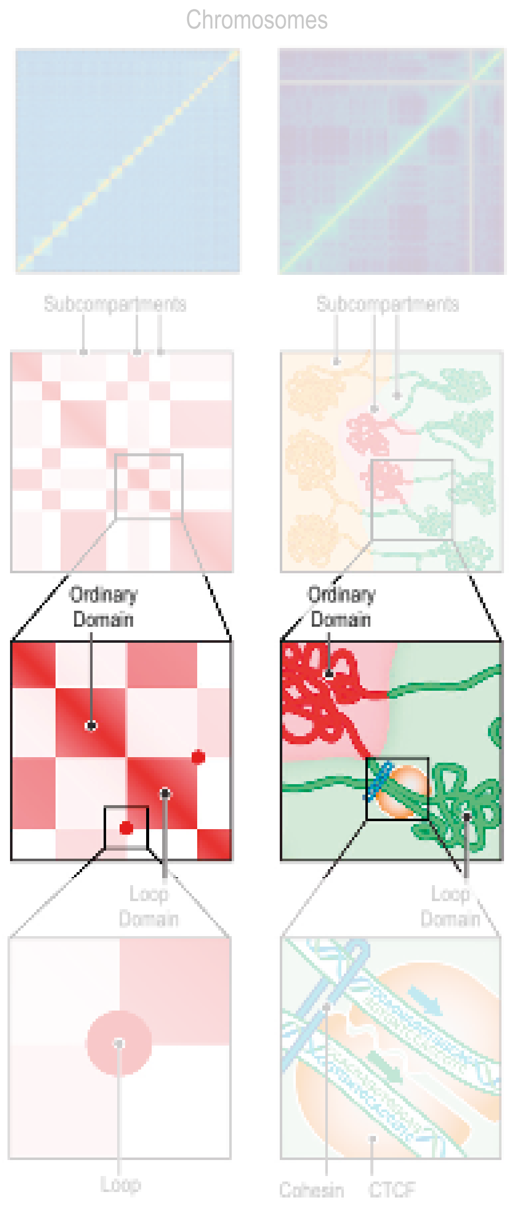


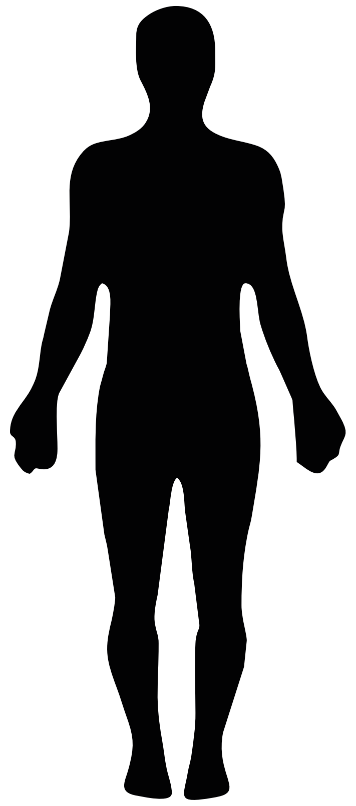


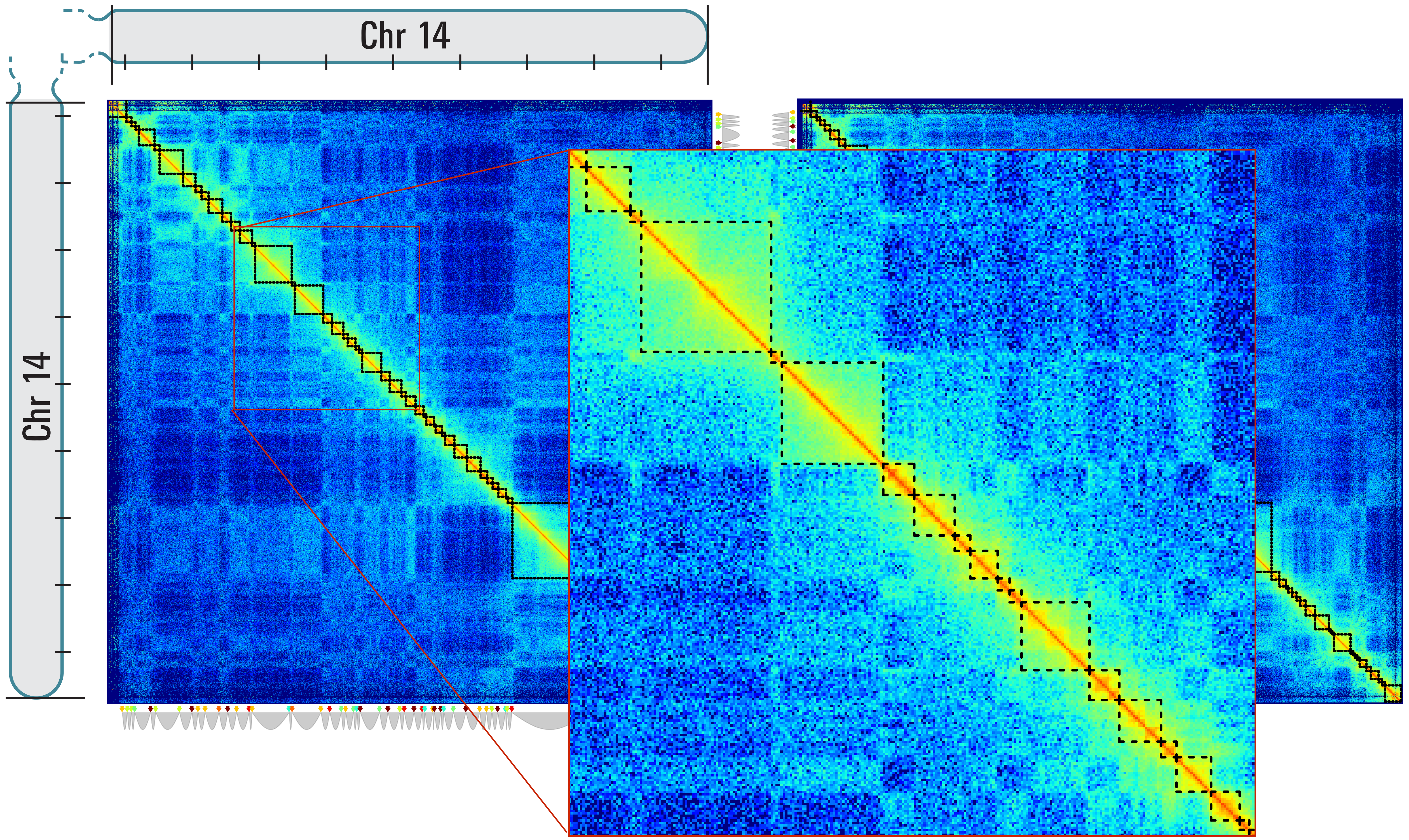


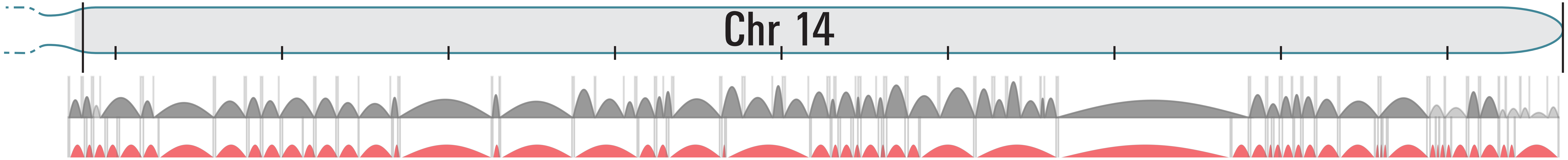
Genome Topologically Associating Domains

Conservation of TADs



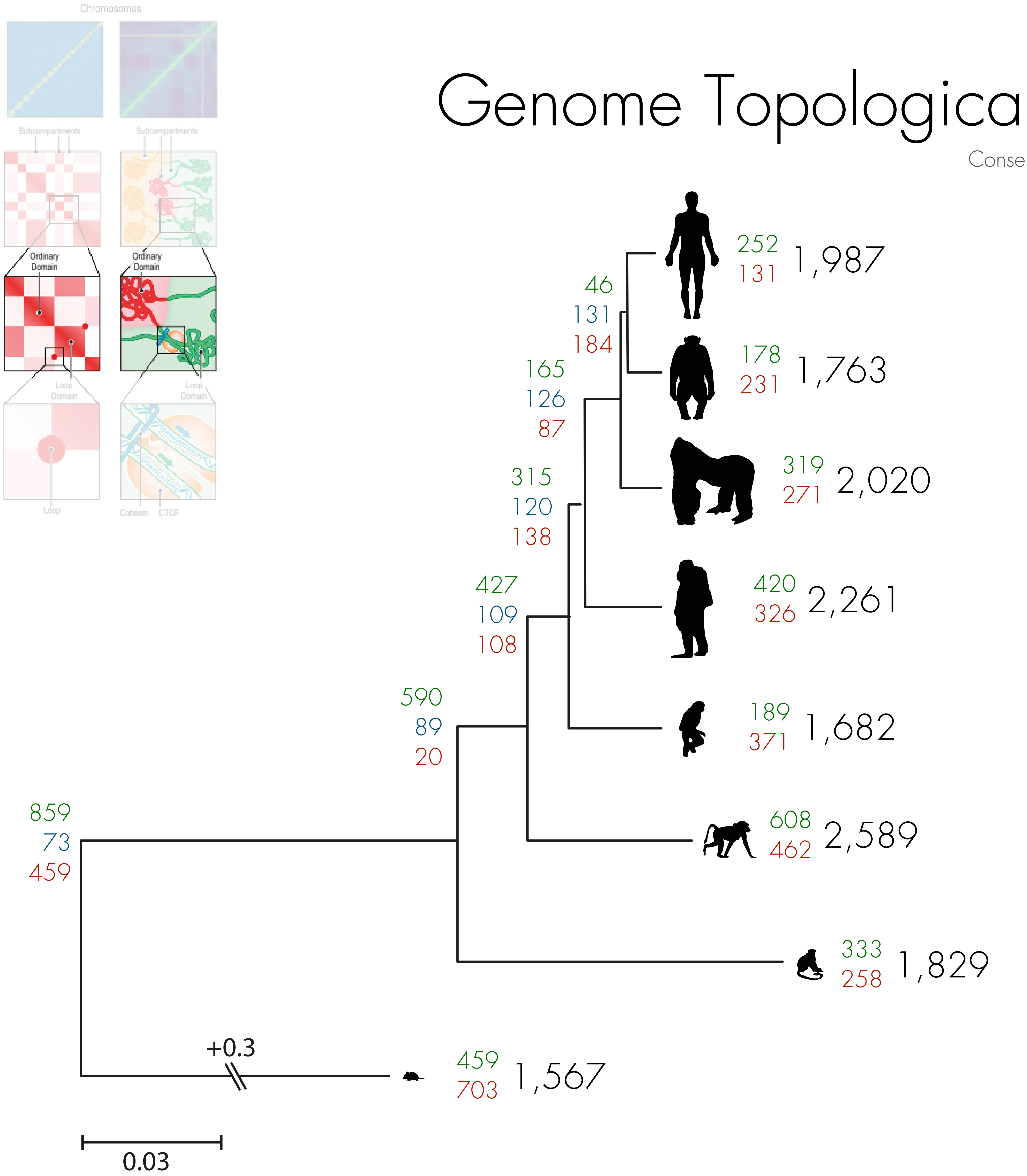


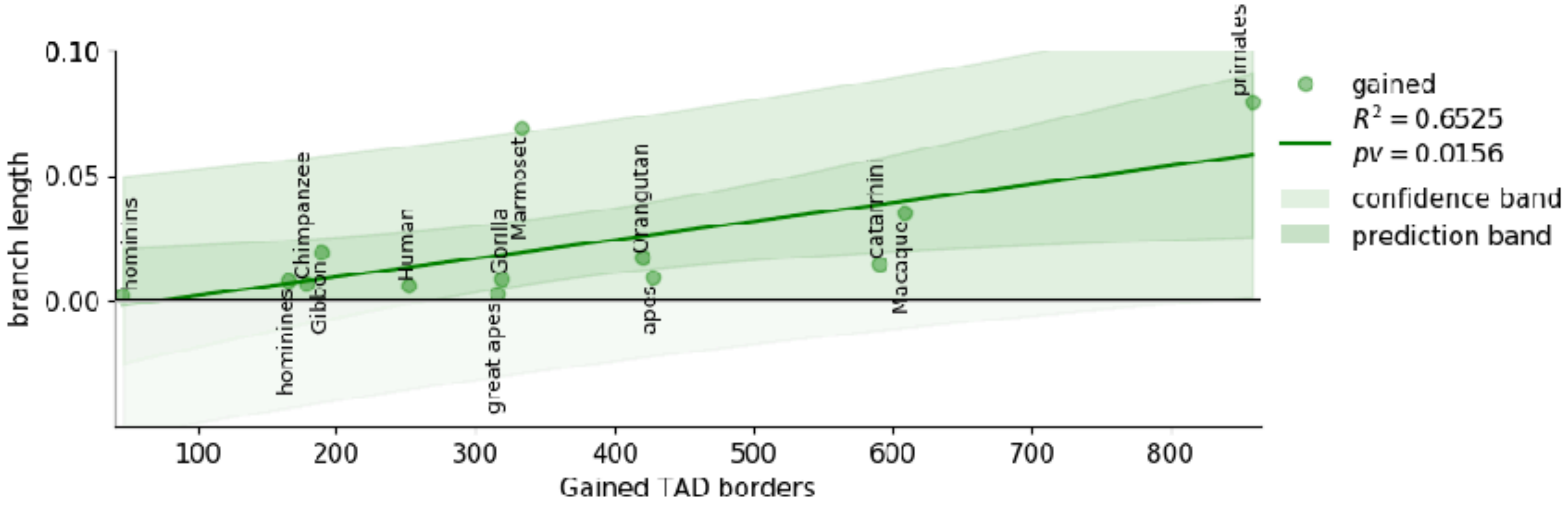


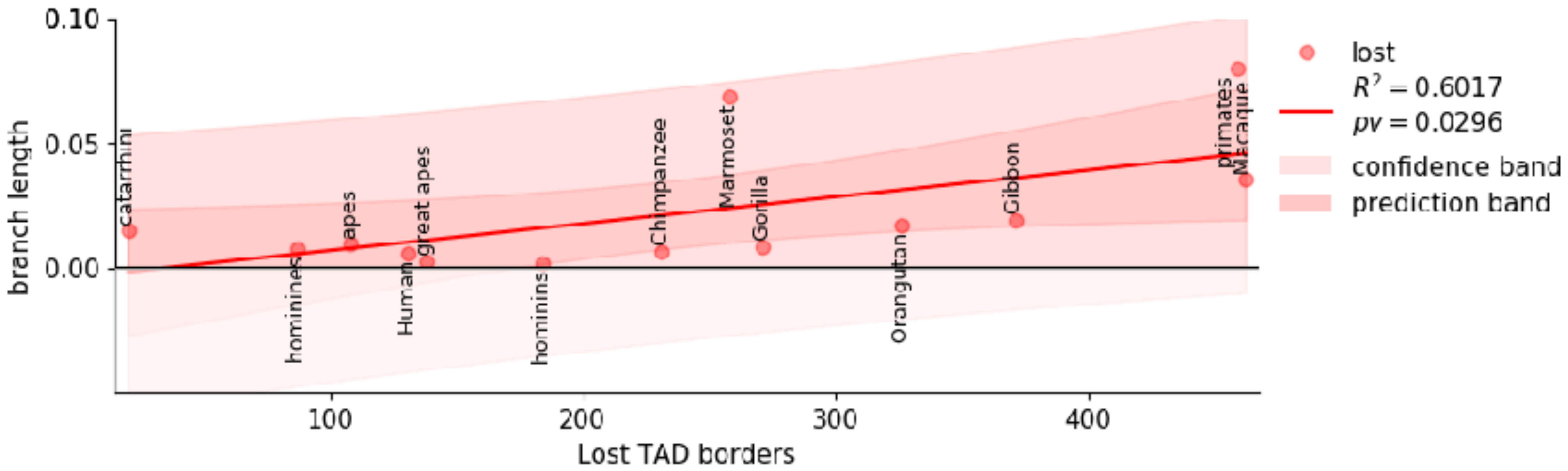


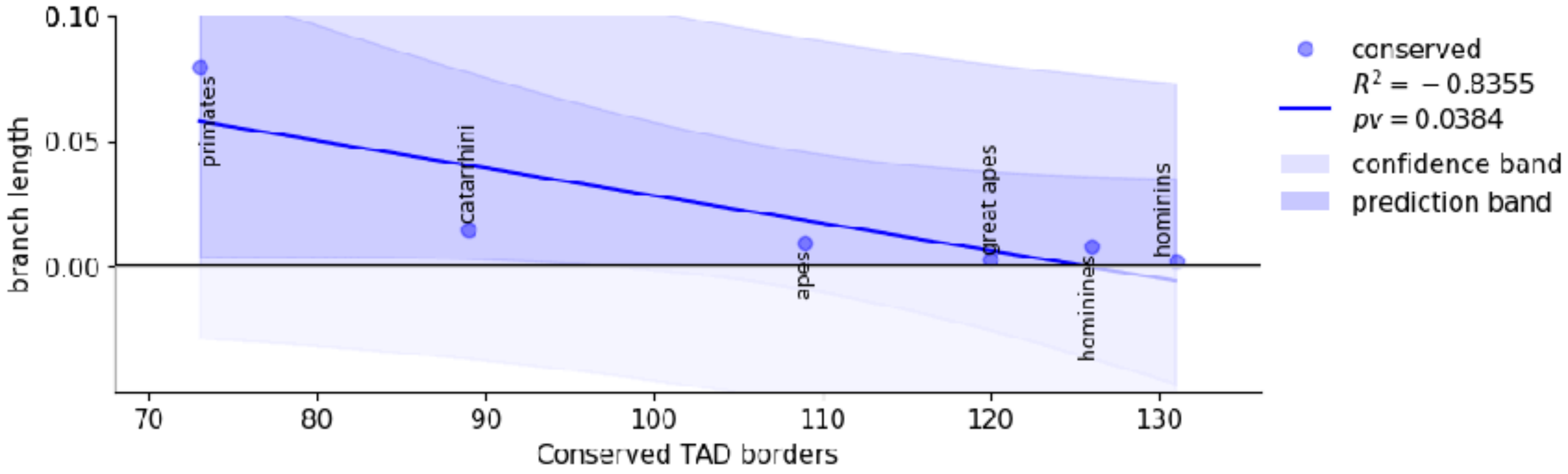
Genome Topologically Associating Domains

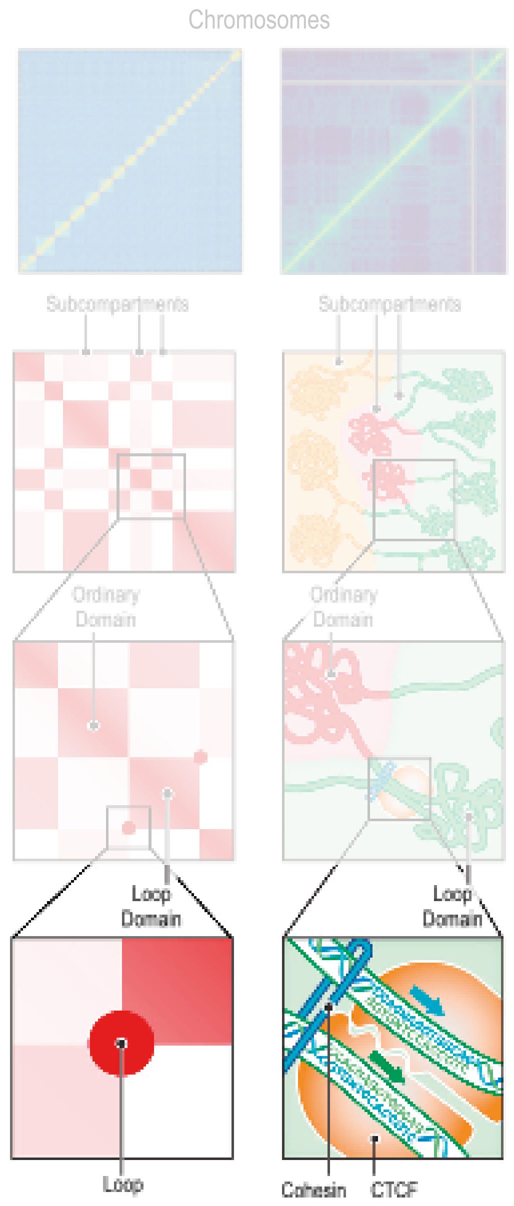
Conservation of TADs





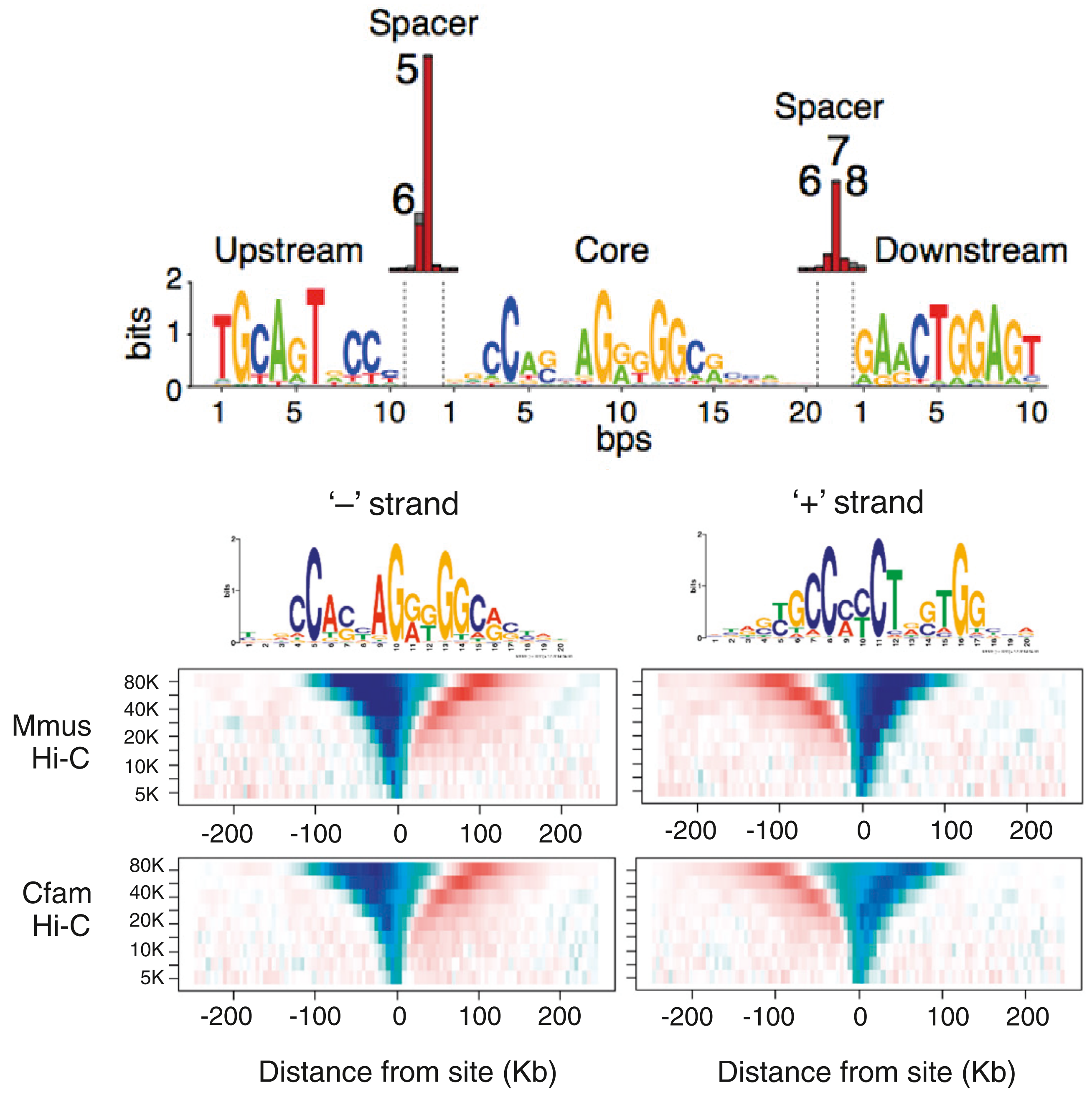






Loops

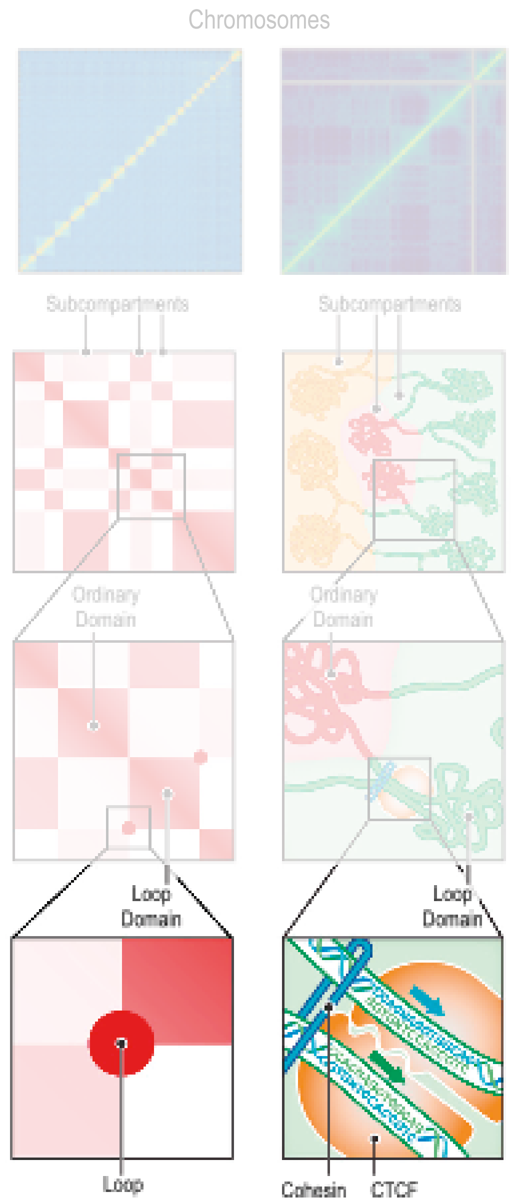
Conservation of CTCF sites

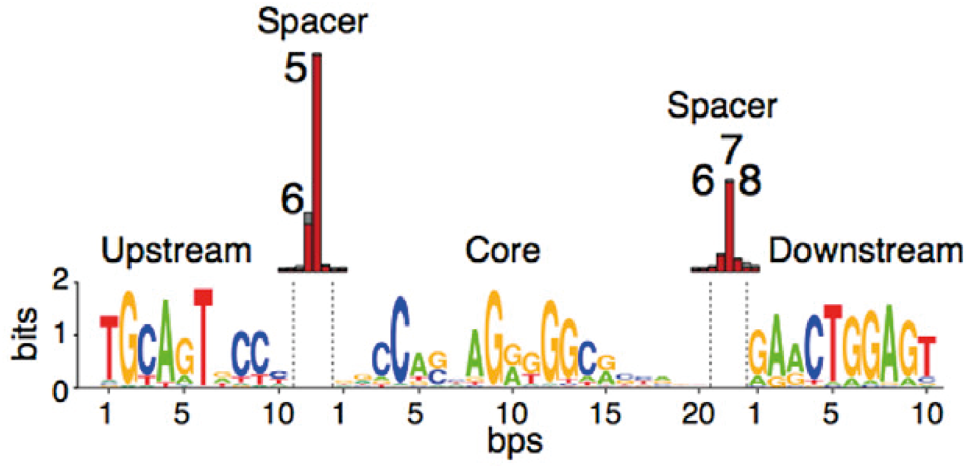


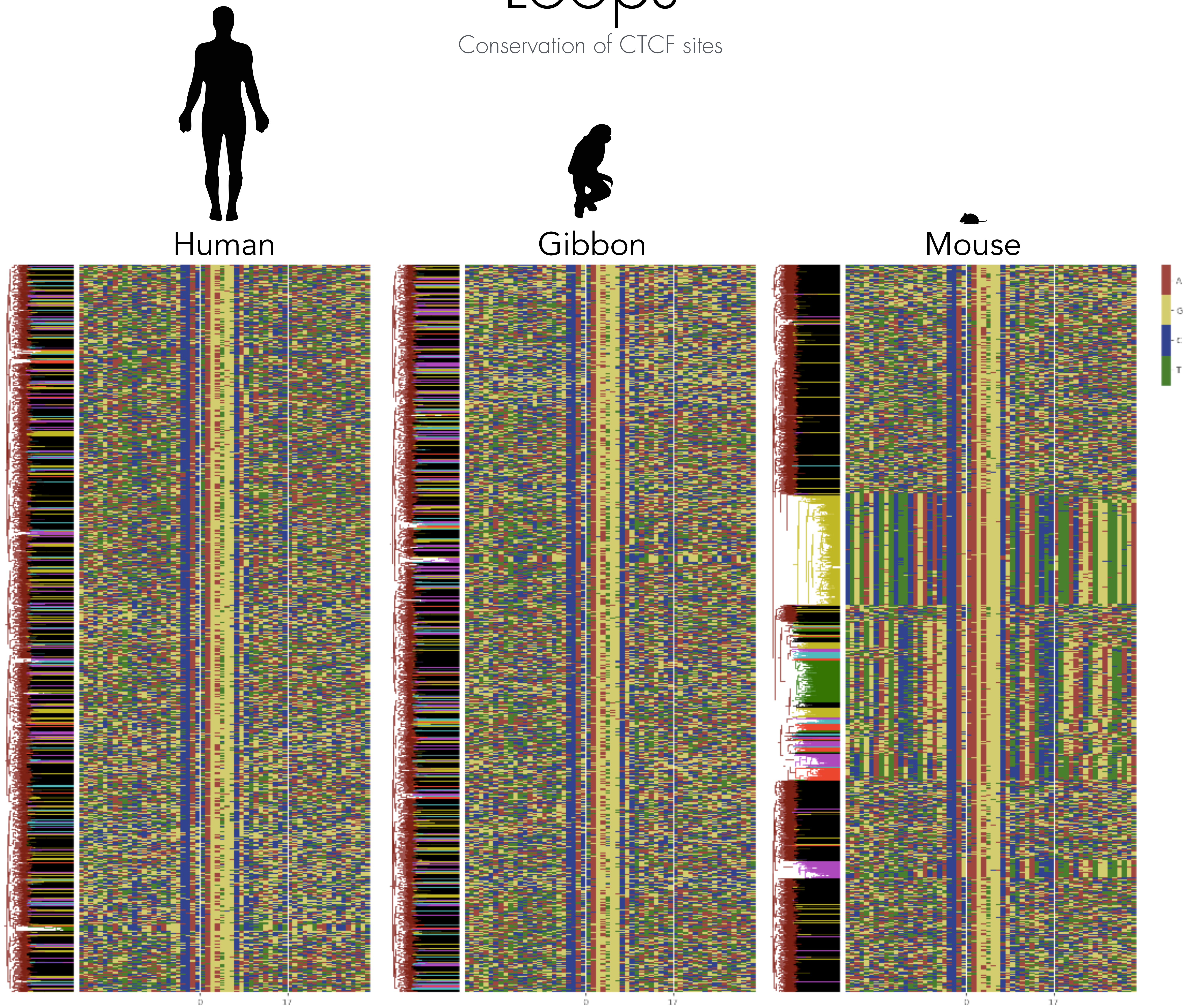
Vietri Rudan, et al. Cell Rep. 2015 Mar 03; 10(8) 1297-1309
Nakahashi et al. Cell Rep. 2013 May 30; 3(5) 1678-1689

Loops

Conservation of CTCF sites

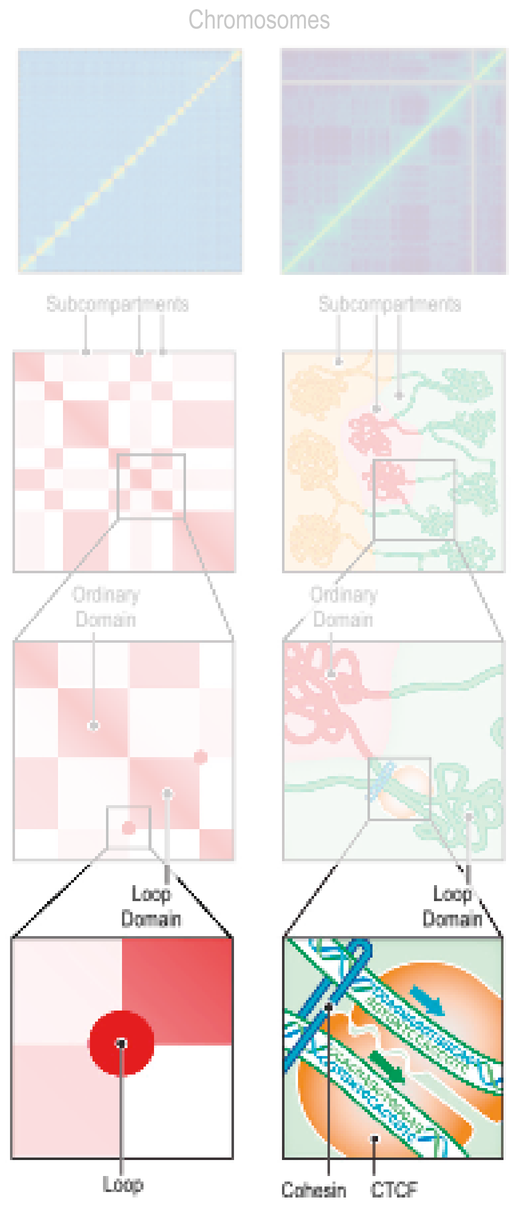




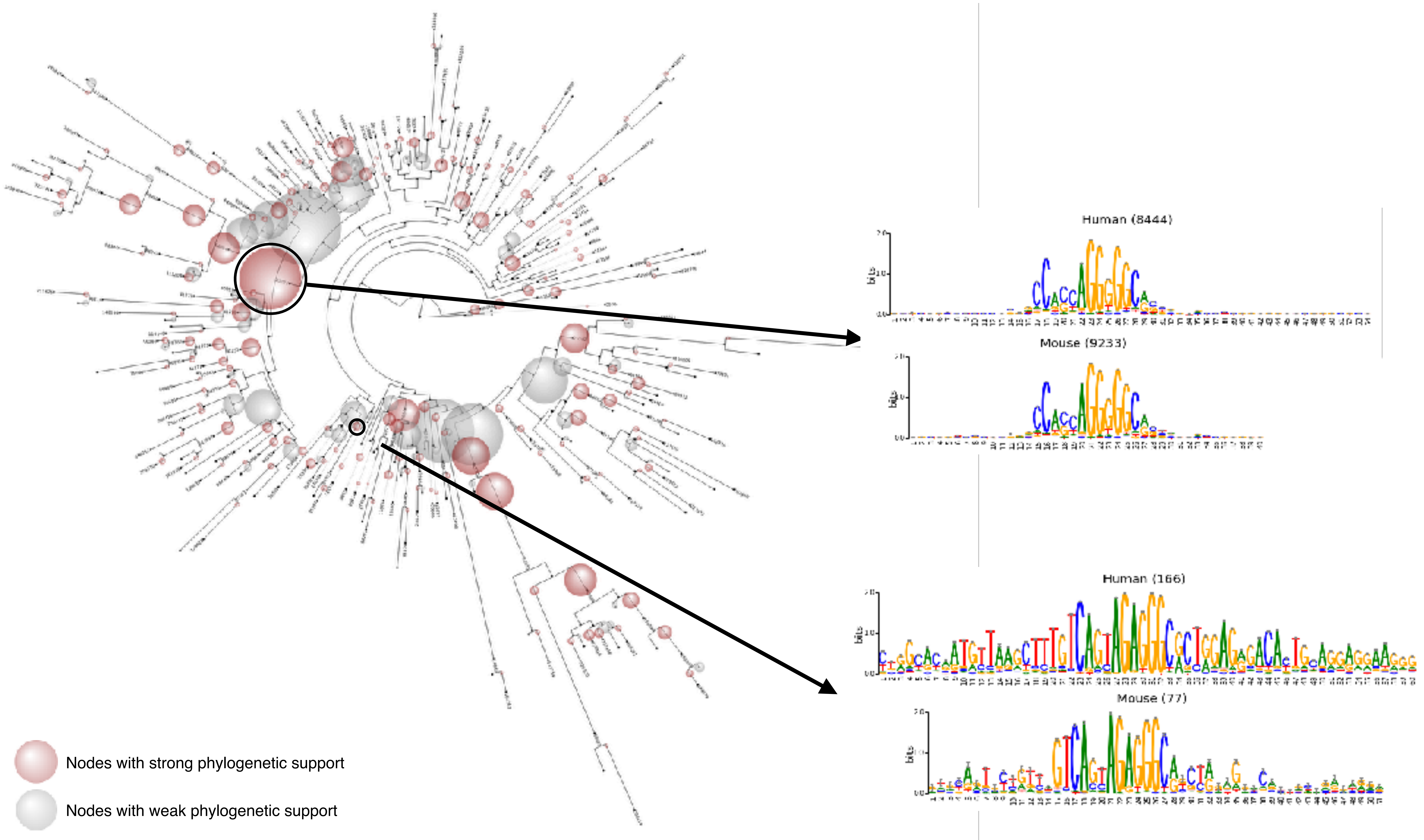


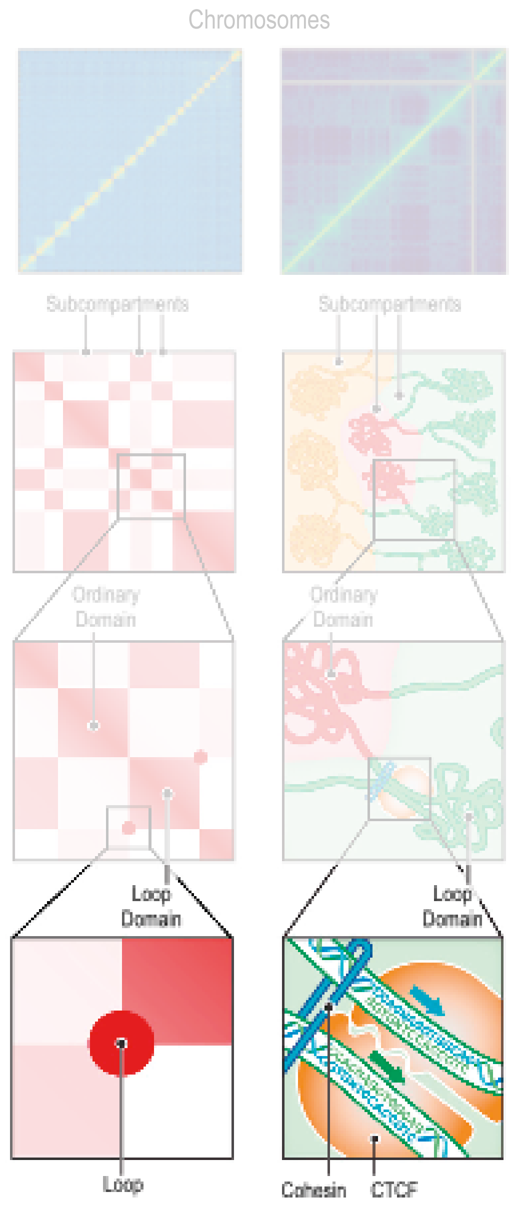
Loops

Conservation of CTCF sites



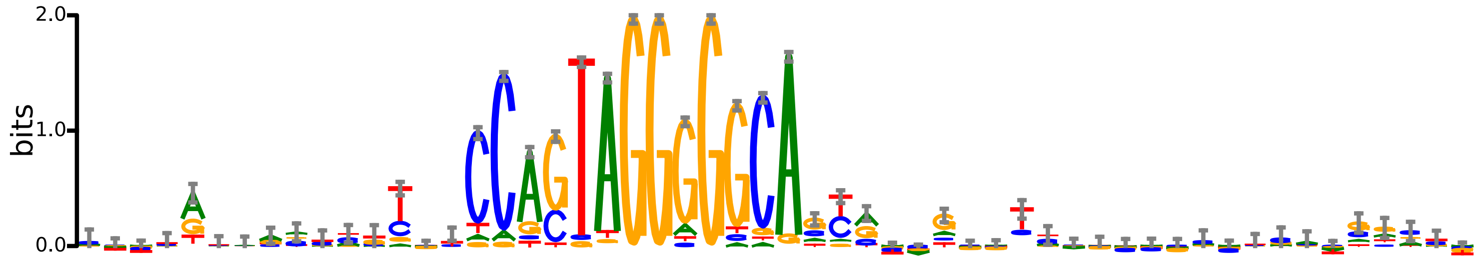
- >150 K sequences to align (all species together)
- Tree reconstruction using FastTree2
- Reduce the complexity filtering by node support
- Selection of 173 nodes with 0.9 bootstrap
- At least 100 CTCF sites (and at least 50 from single species)



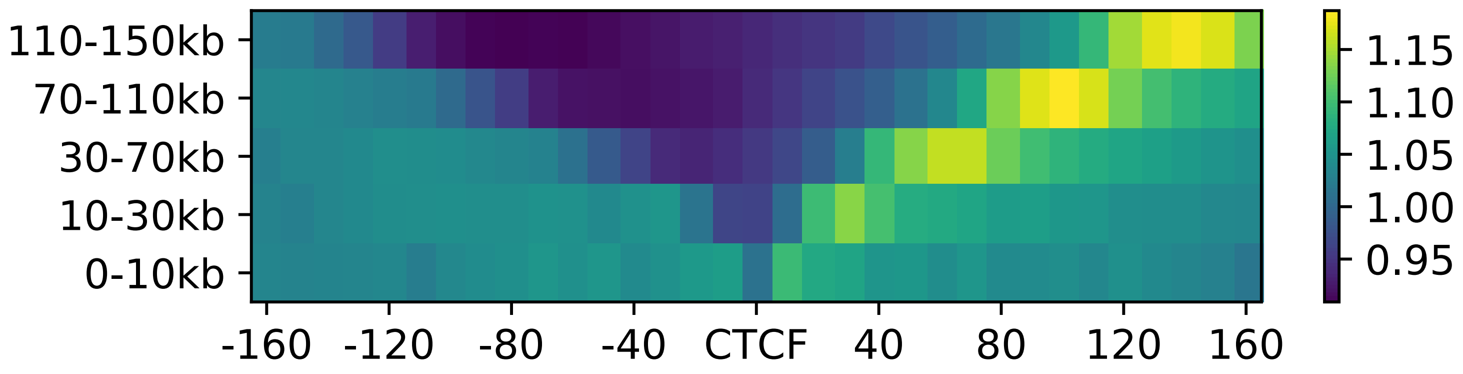


Loops

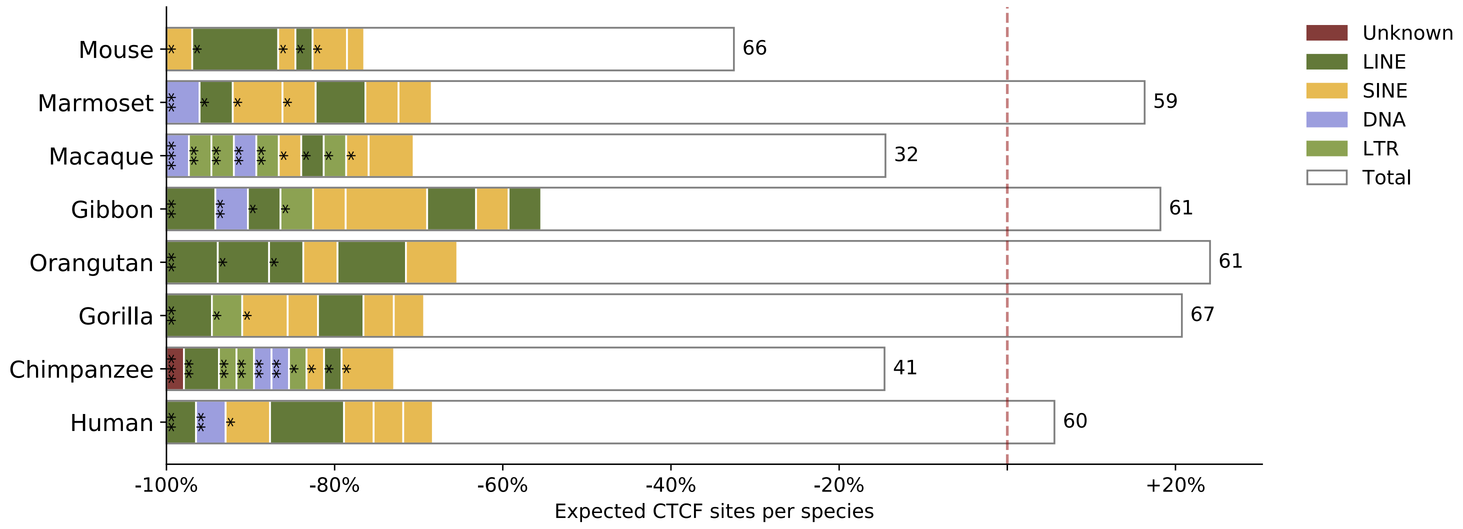
Conservation of CTCF sites



- Motif (nucleotide content)



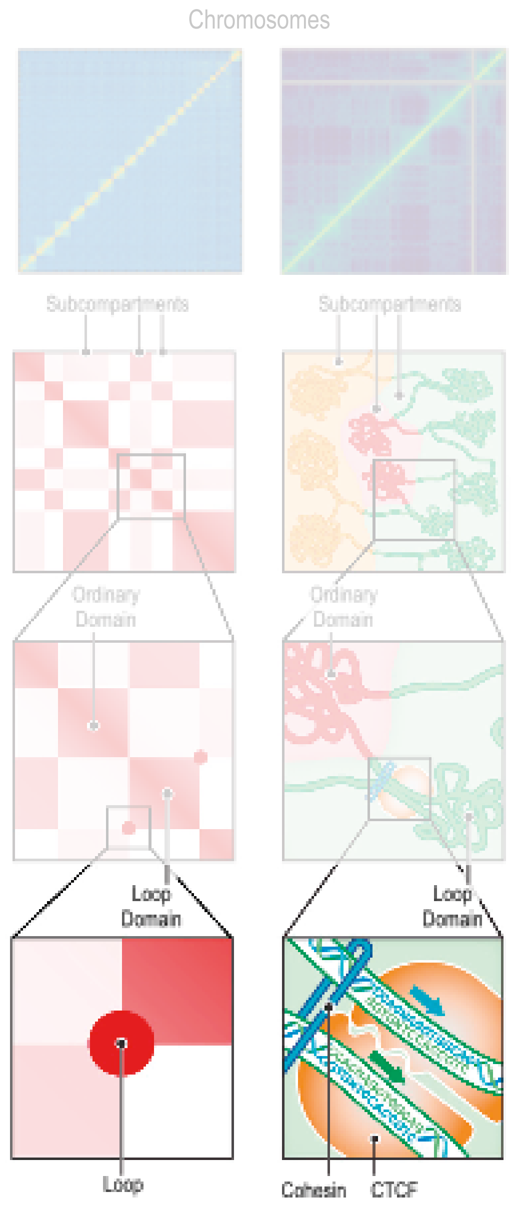
- Insulation/looping (interaction directionality)

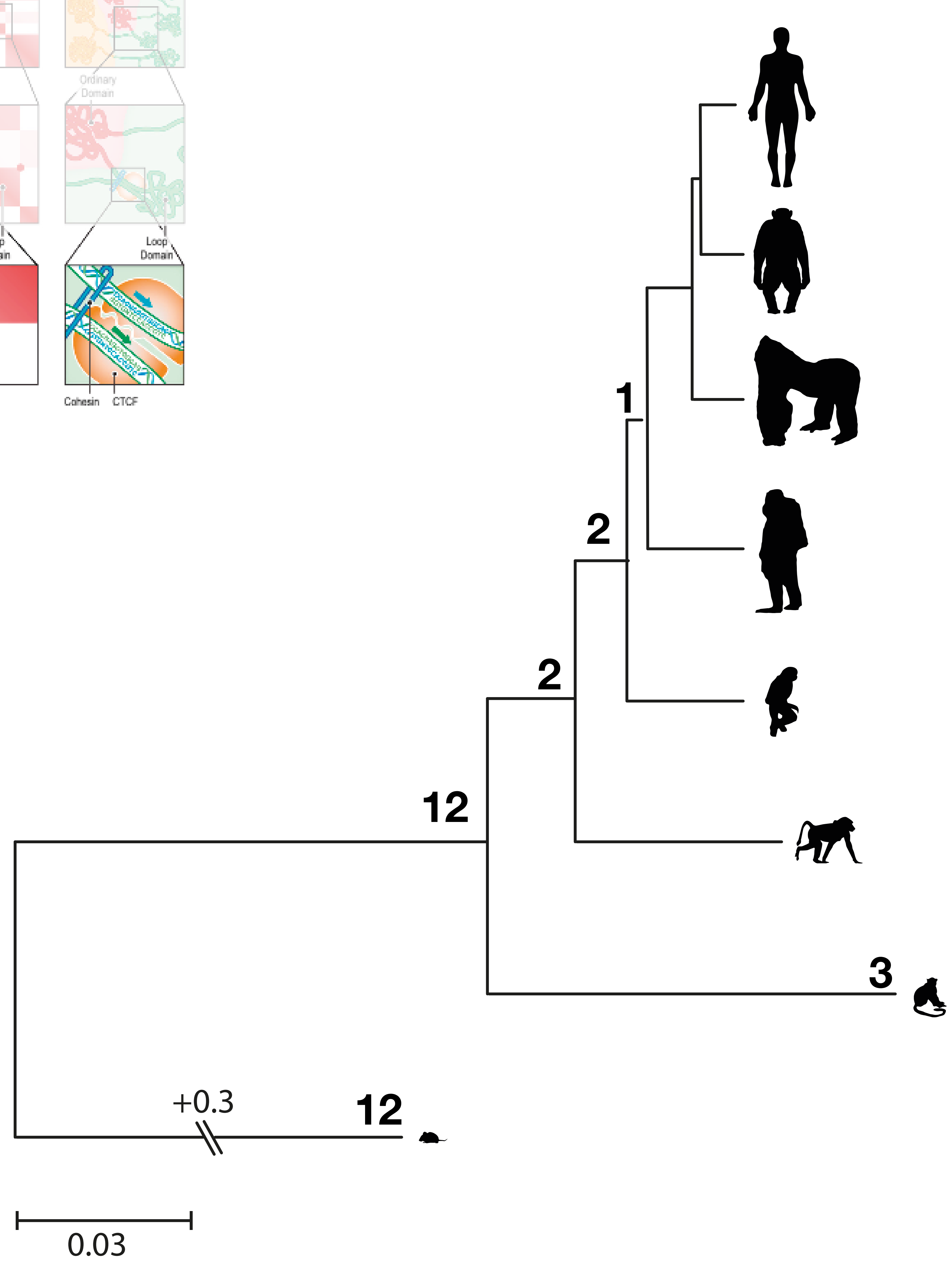


- Enrichment in repetitive elements

Loops

Conservation of CTCF sites

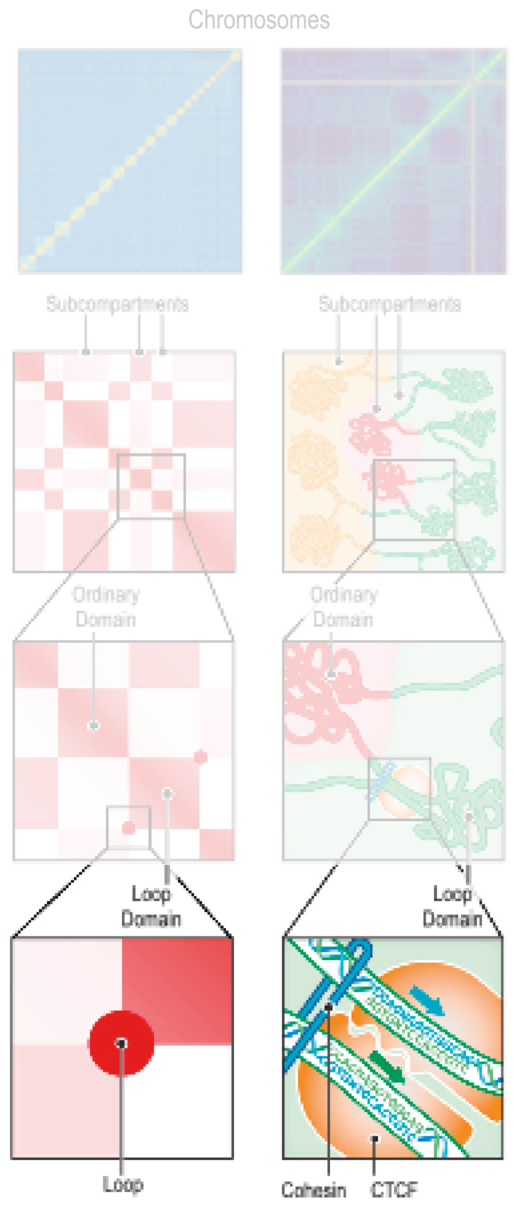


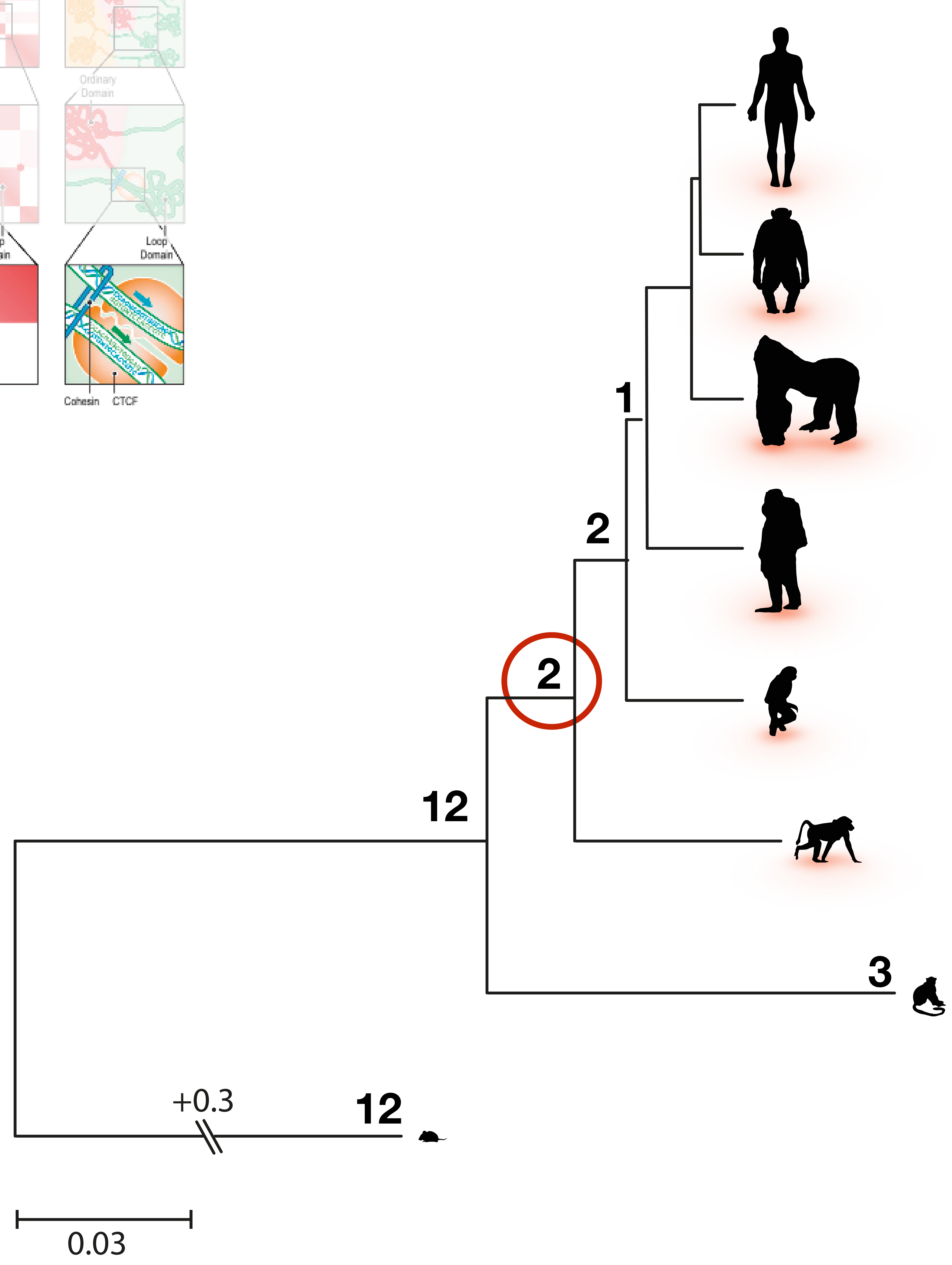


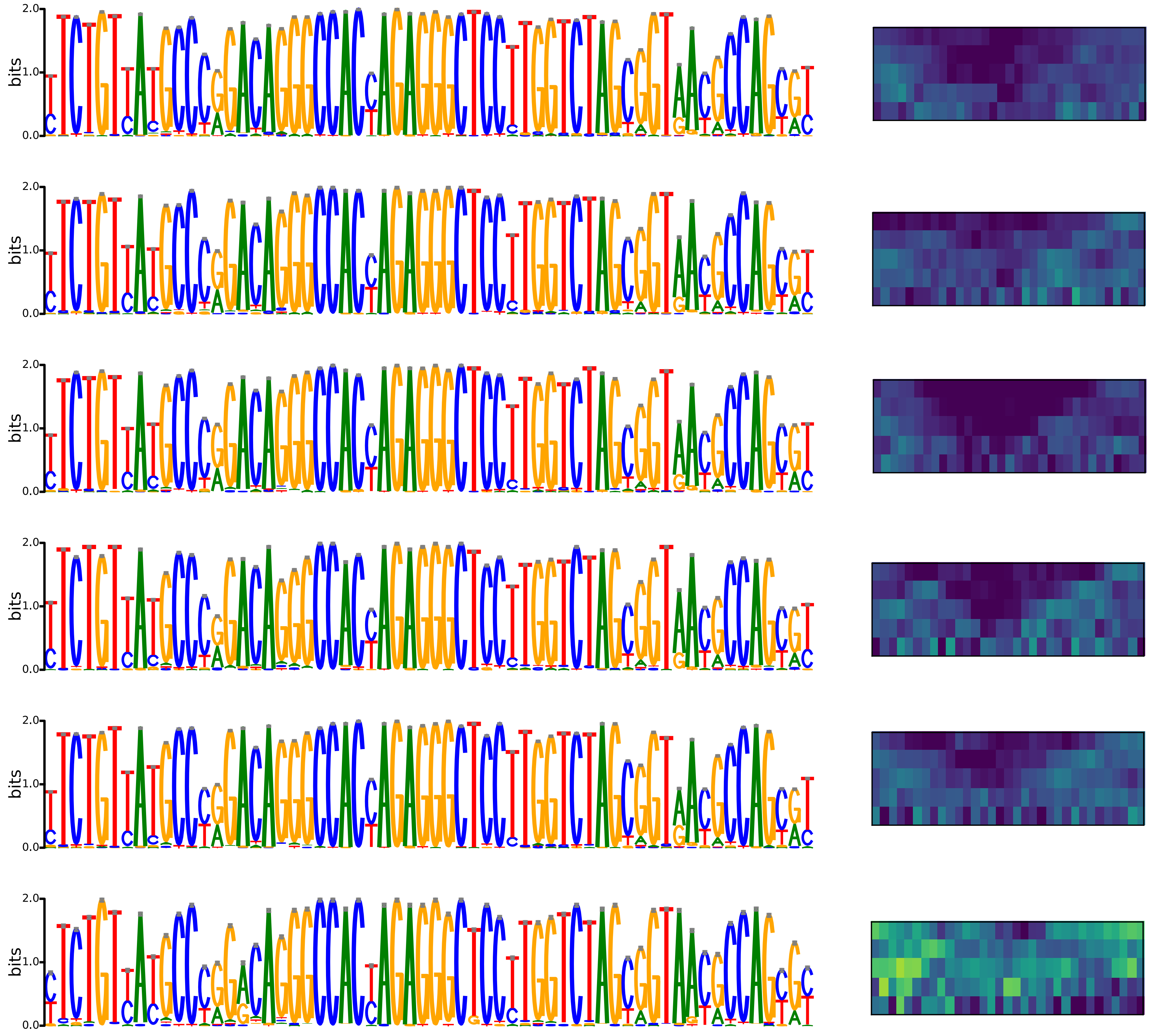
Few events of genome expansion through transposons involving CTCF sites

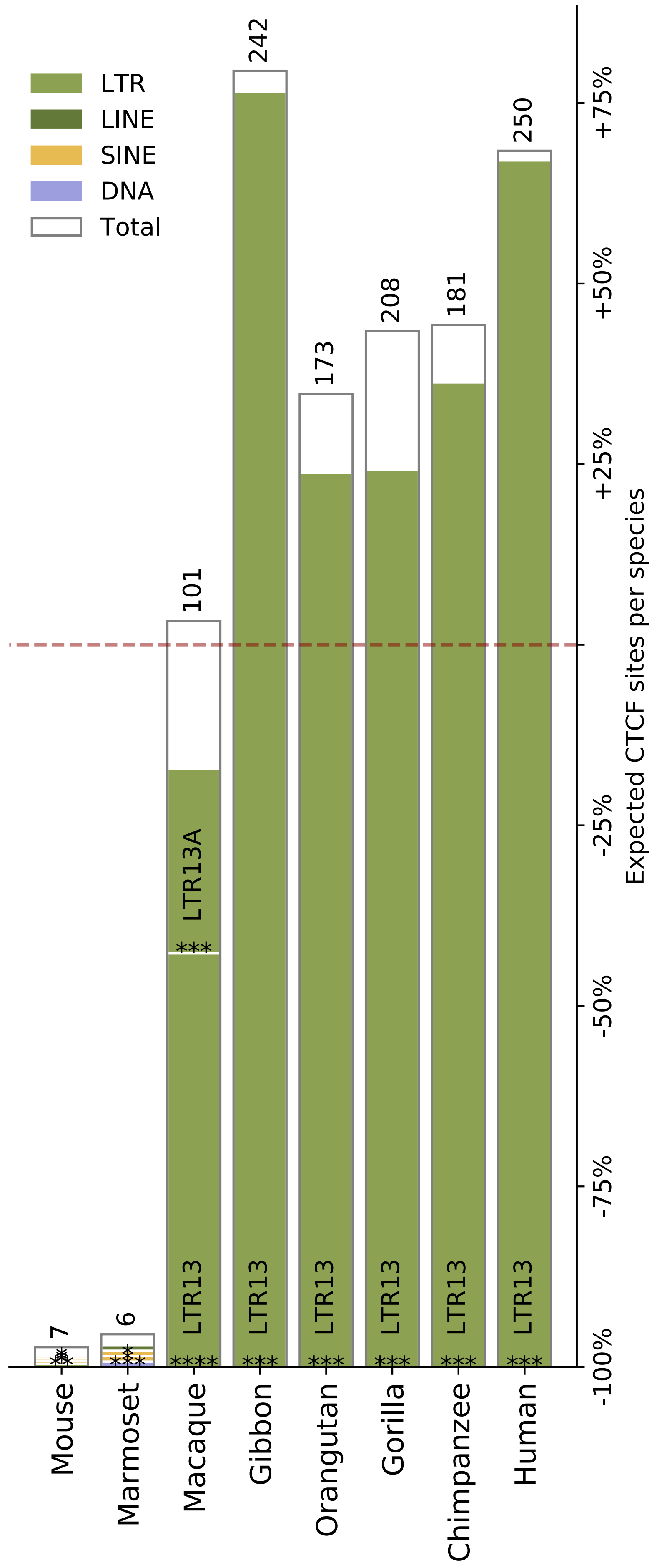
Loops

Conservation of CTCF sites





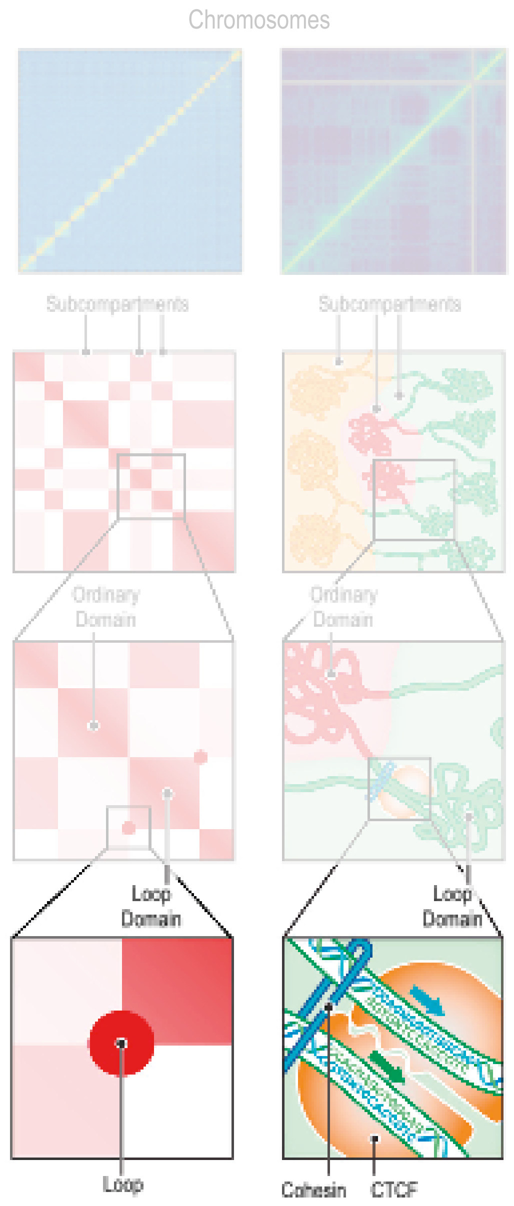


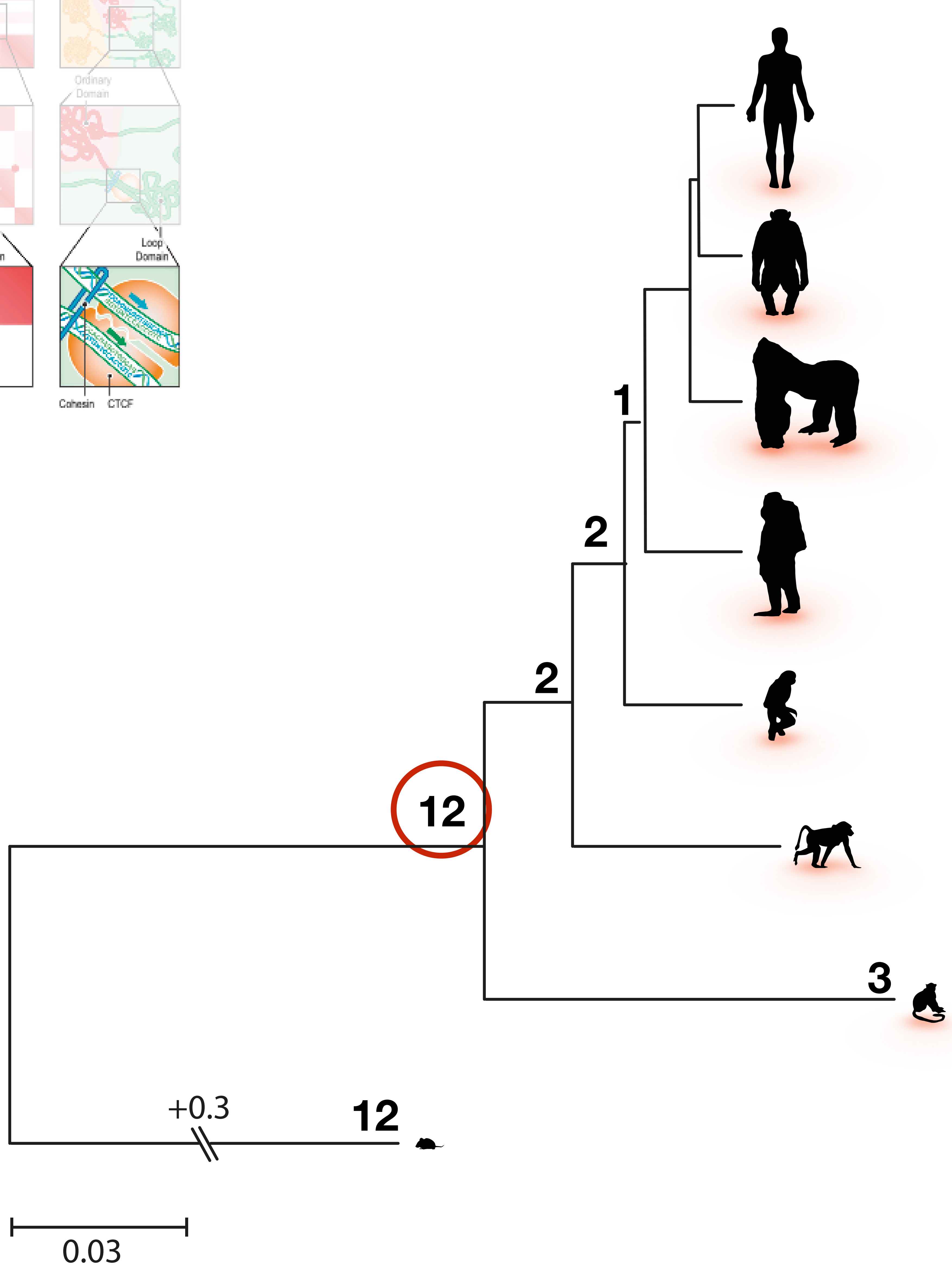


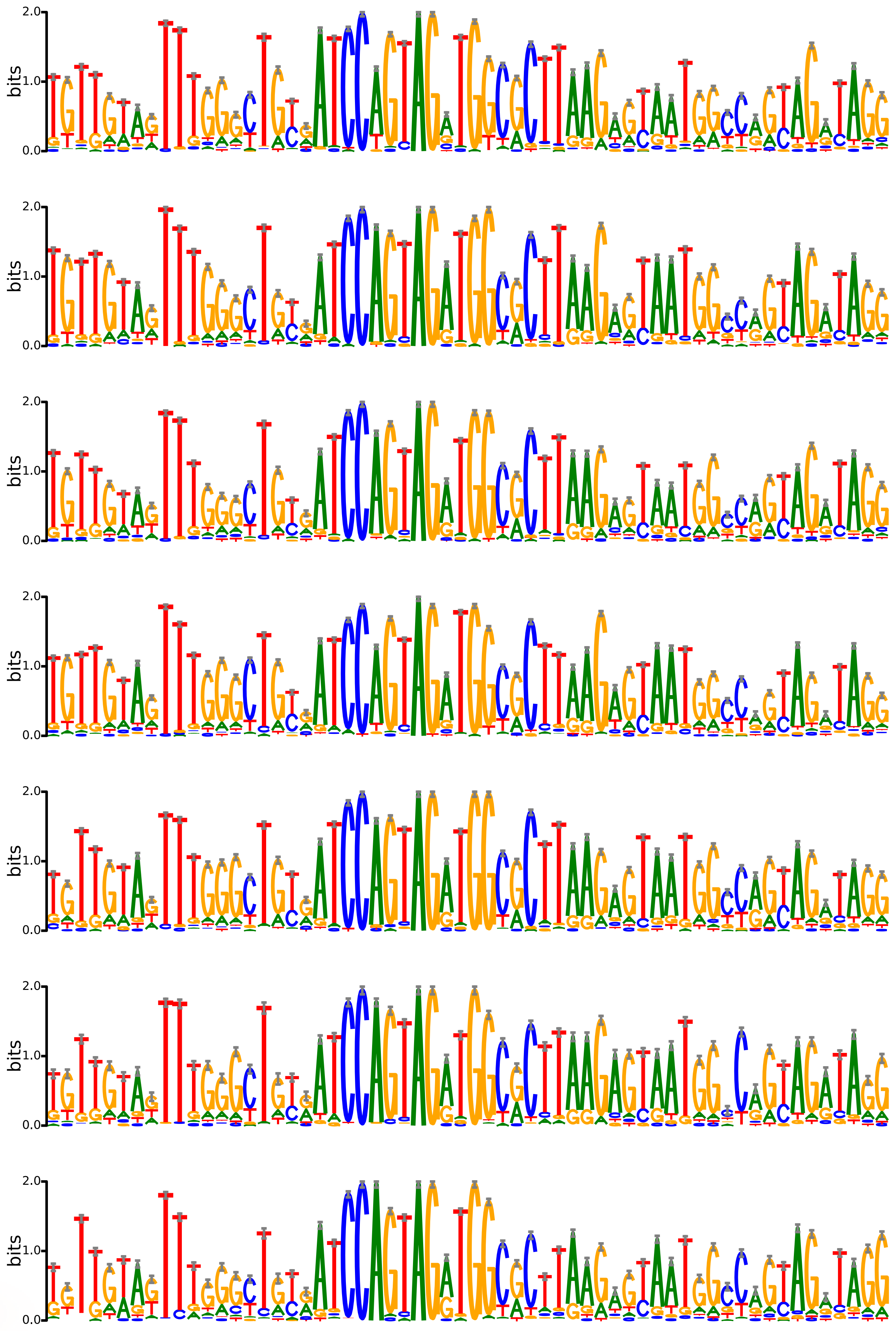
LTR13 (Long Terminal Repeat) for HERVK13 endogenous retrovirus

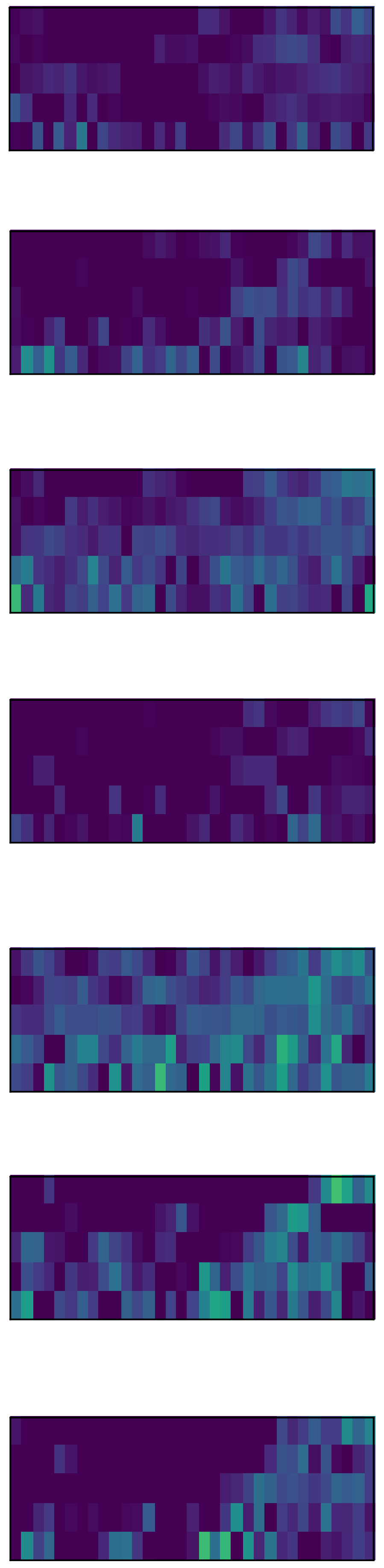
Loops

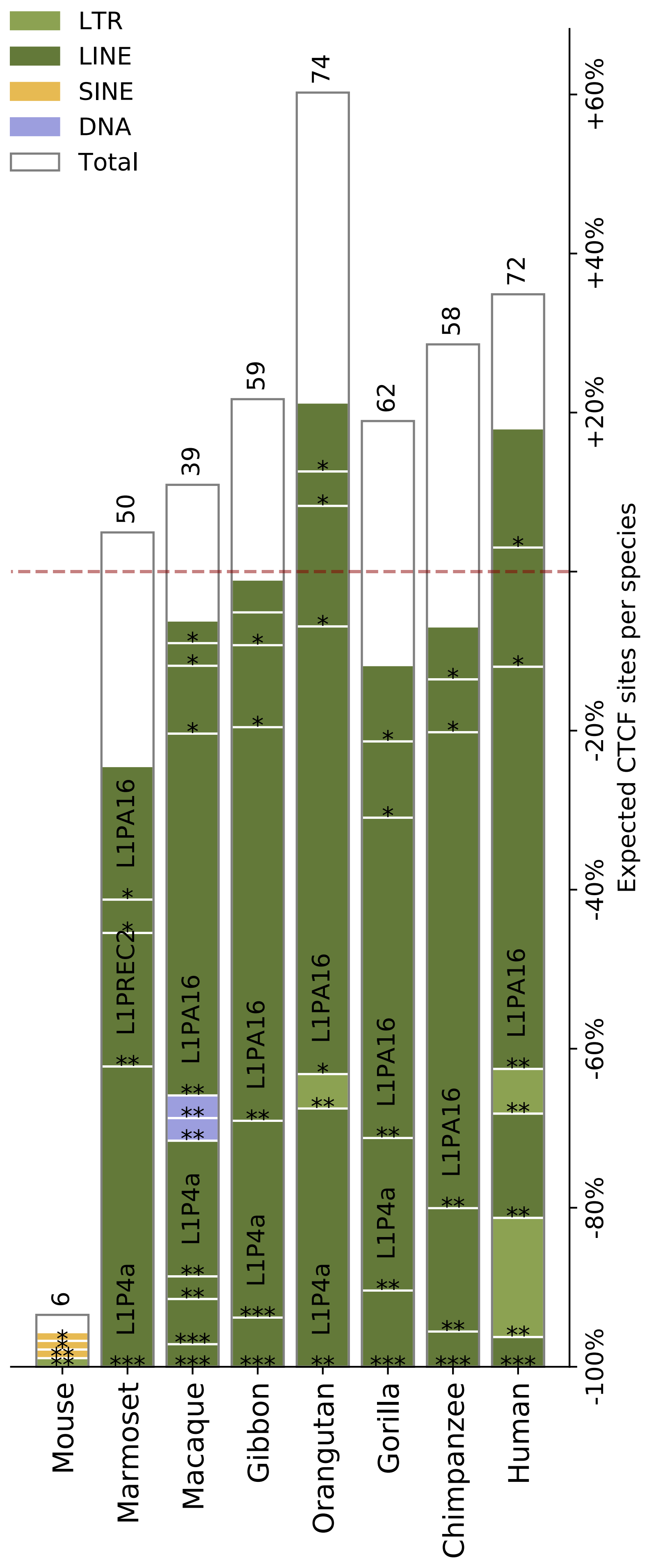
Conservation of CTCF sites







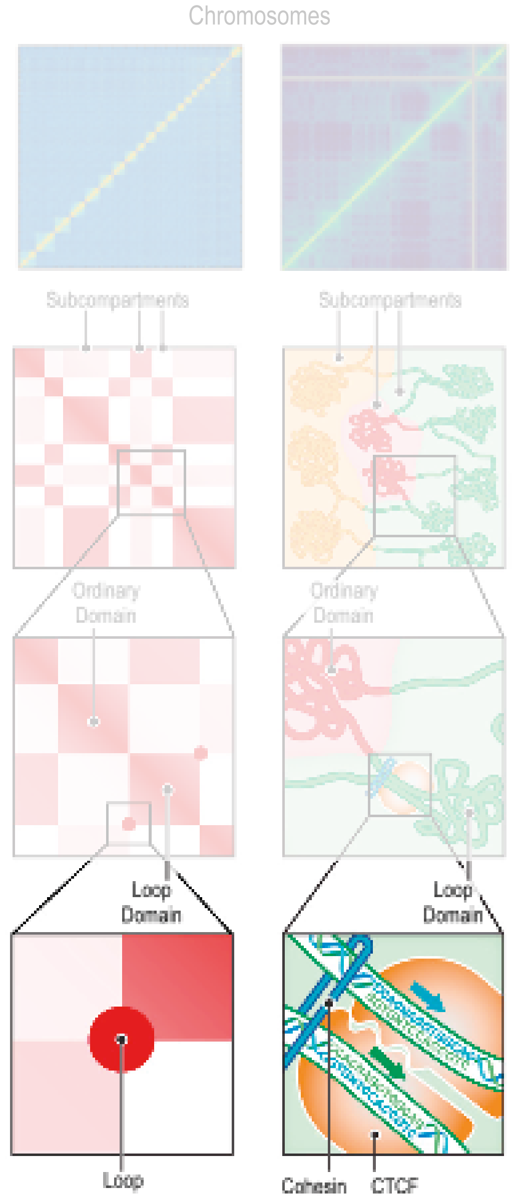


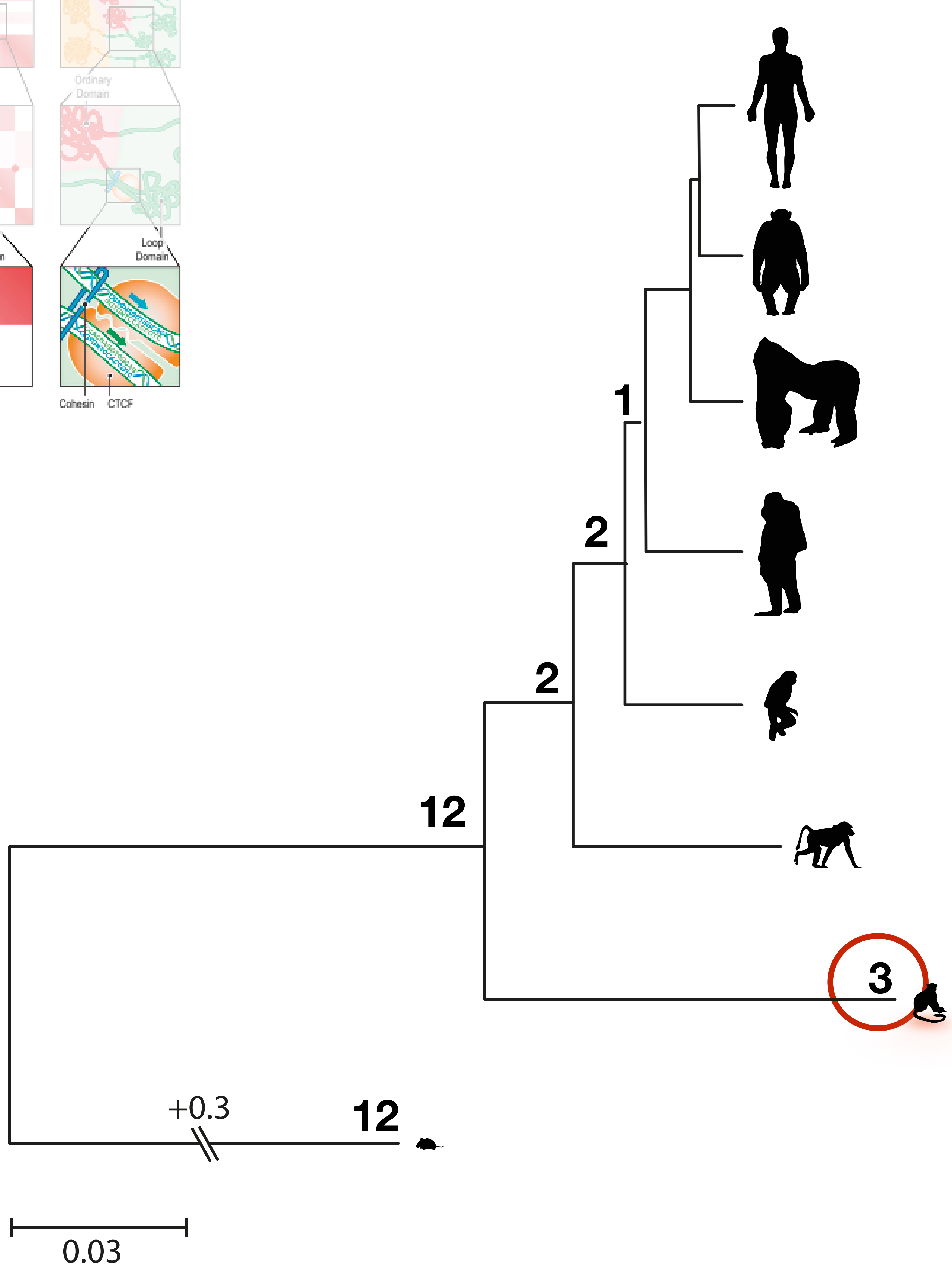


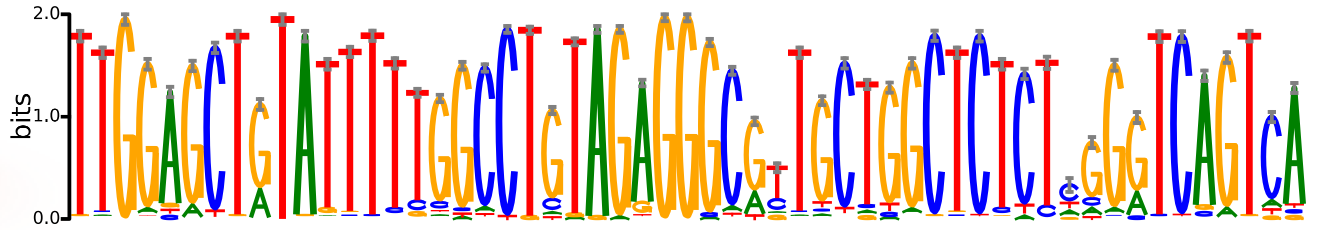
3' end of L1 retrotransposon, L1PA16_3end subfamily

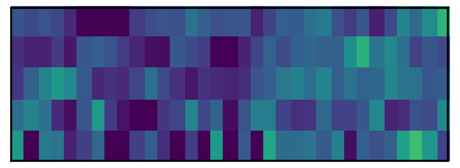
Loops

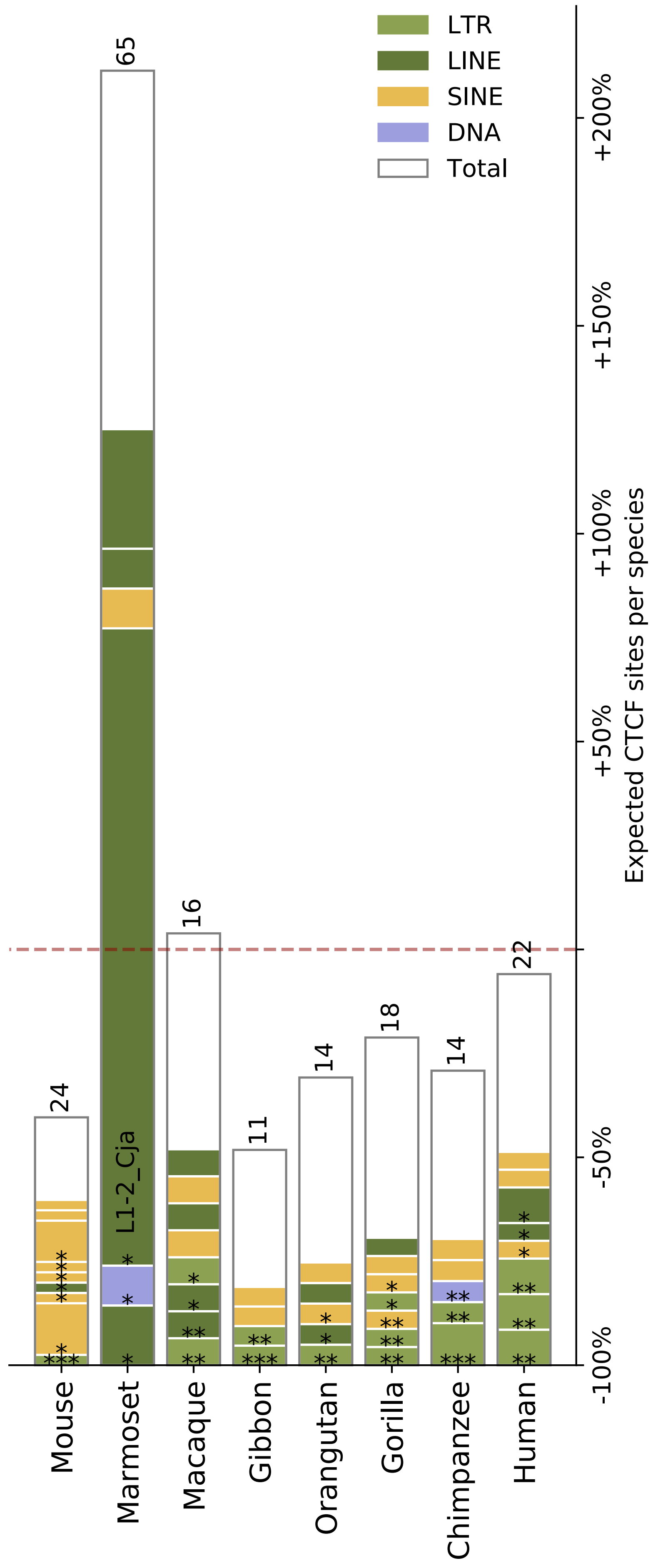
Conservation of CTCF sites











LINE L1 specific of Callithrix jacchus (Marmoset)

Summary

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

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

@marciuslab
@mamartirenom

cnag
CRG Centre for Genomic Regulation
ICREA

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

With François Serra (currently @BSC) and Tomàs Marquès Lab (UPF-IBE)

.: Our current sponsors :.

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