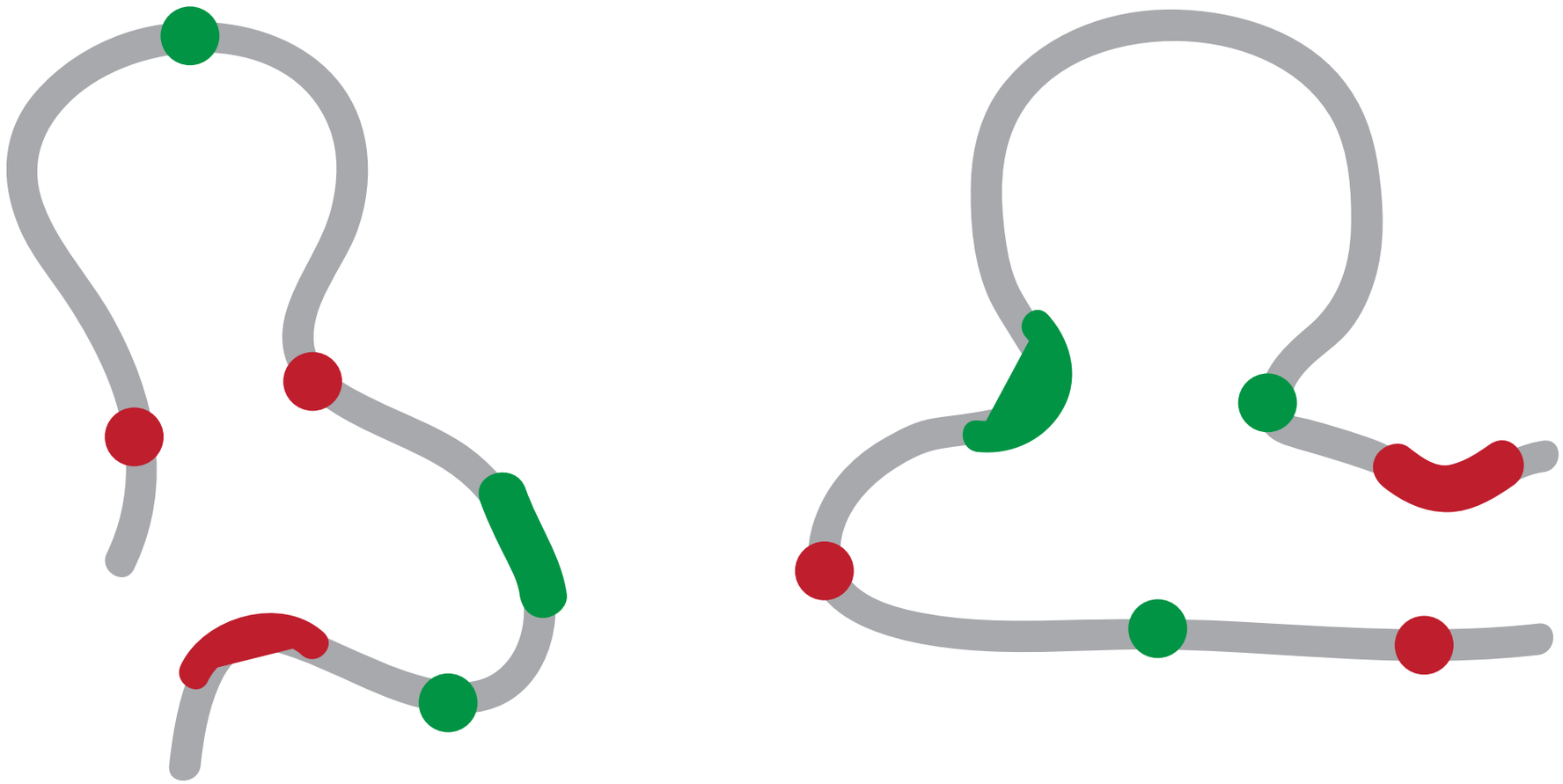
Structure determination of genomes and genomic domains by satisfaction of spatial restraints

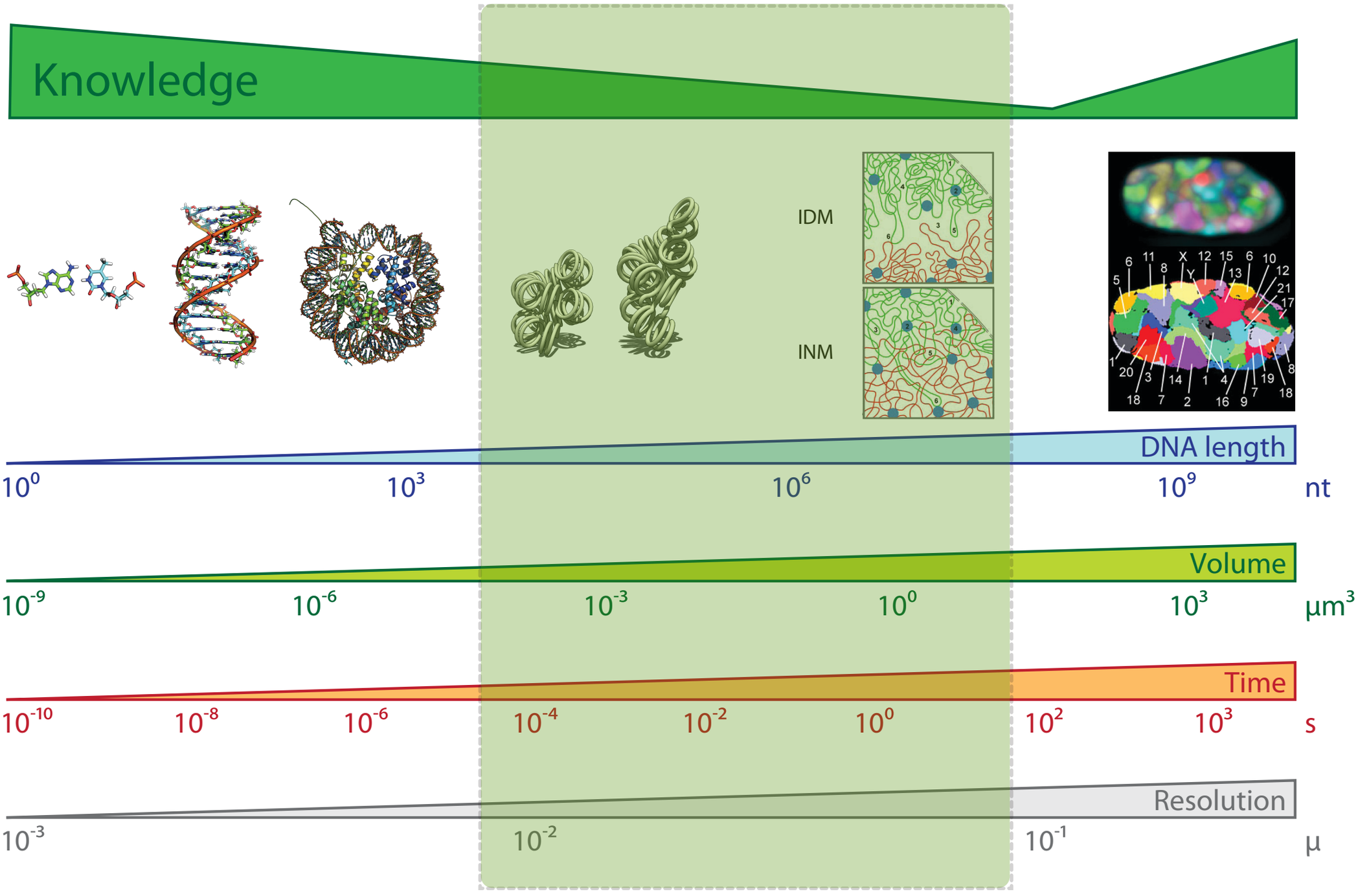
Marc A. Marti-Renom
Structural Genomics Group (ICREA, CNAG-CRG)

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



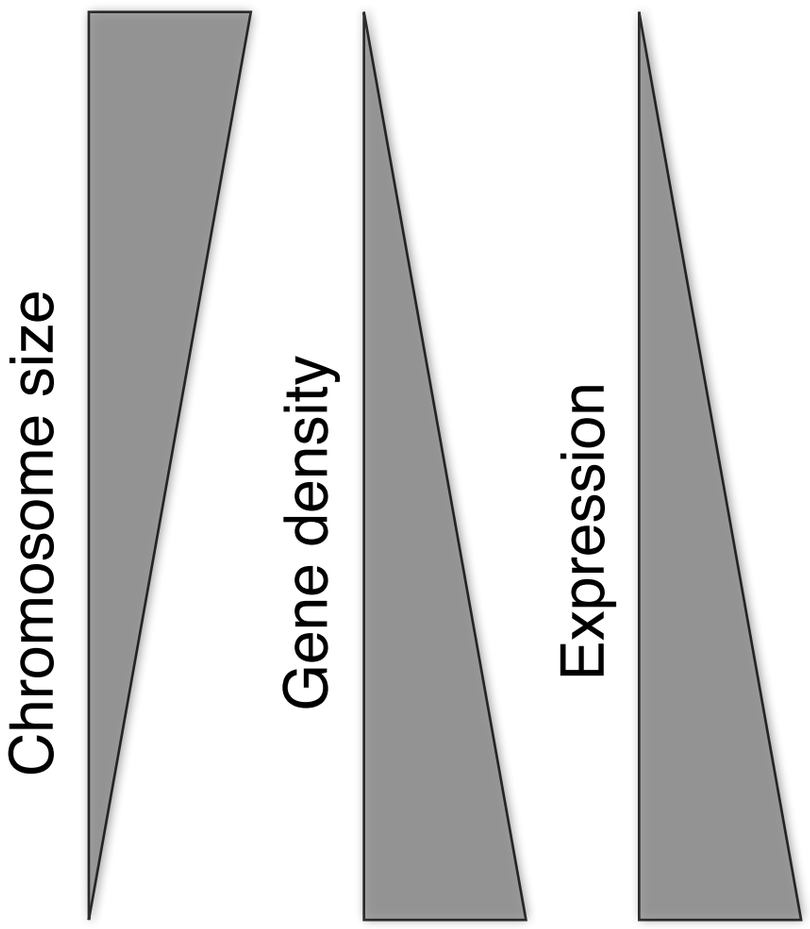
Resolution Gap

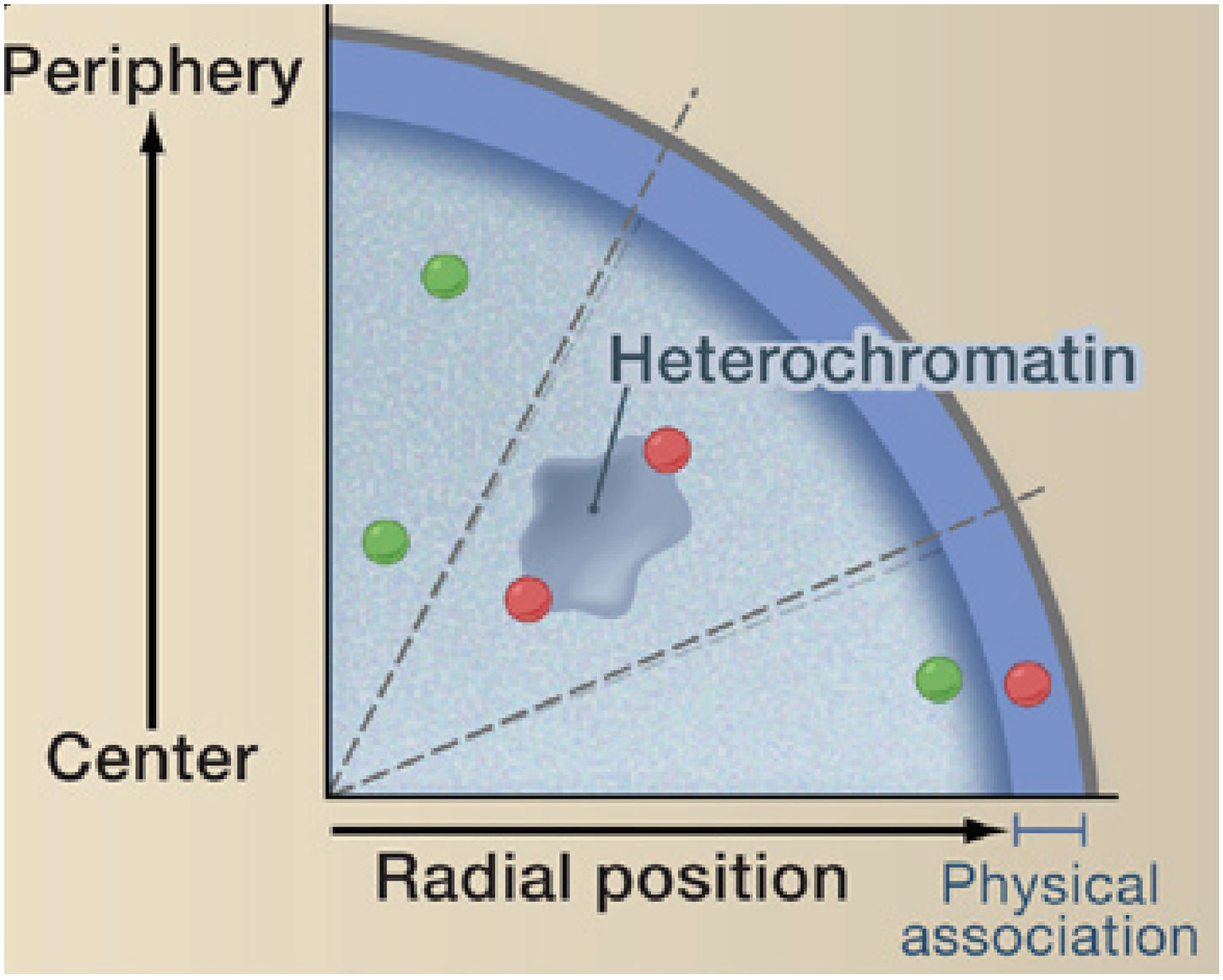
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol *7*, e1002125 (2011)



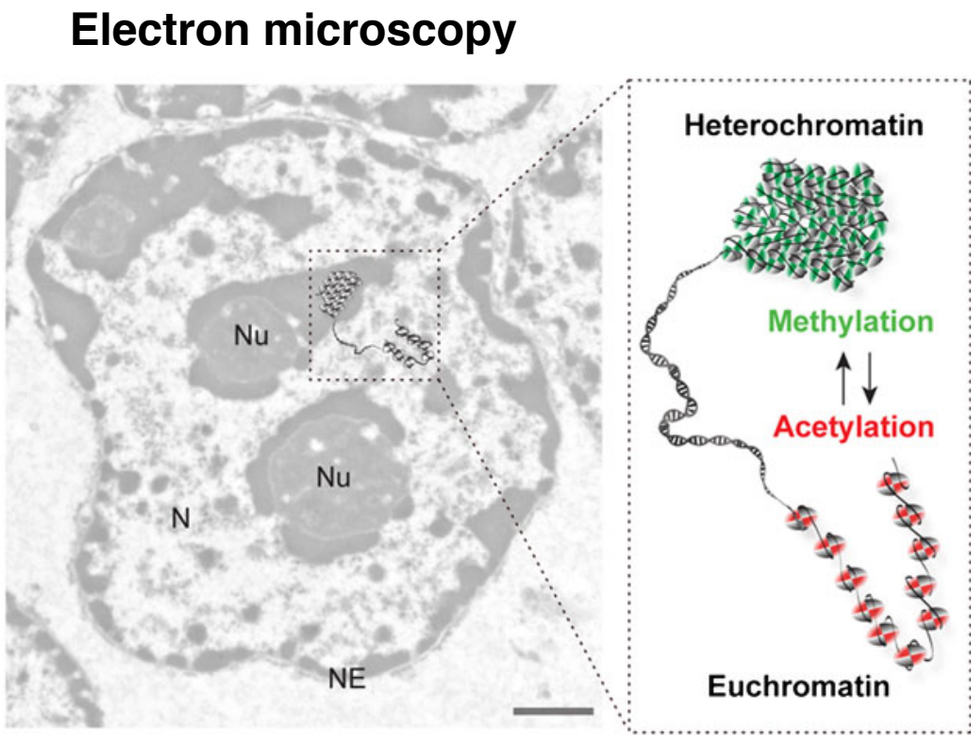
Level I: Radial genome organization

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





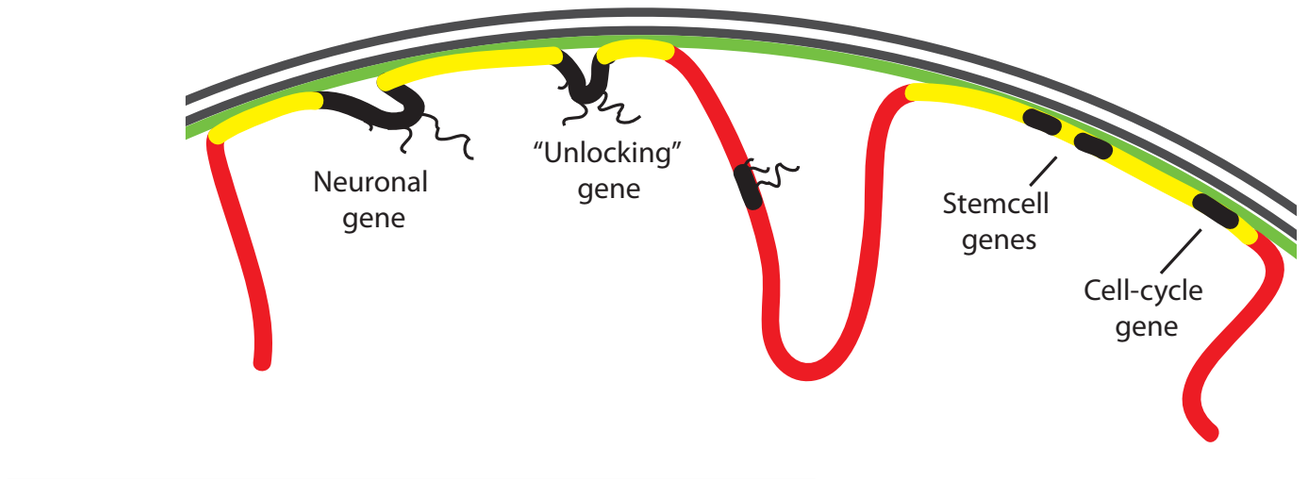
Level II: Euchromatin vs heterochromatin



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

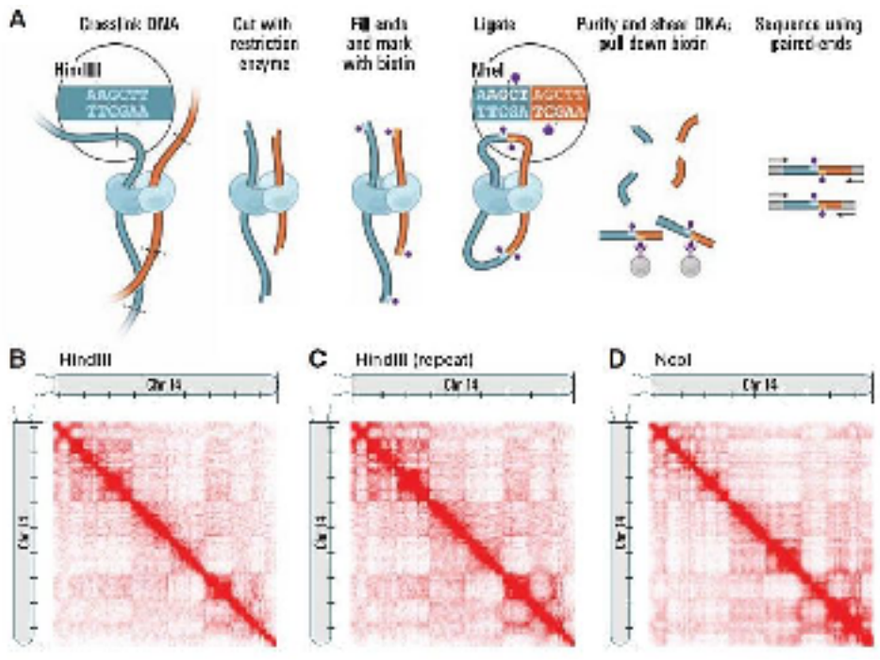


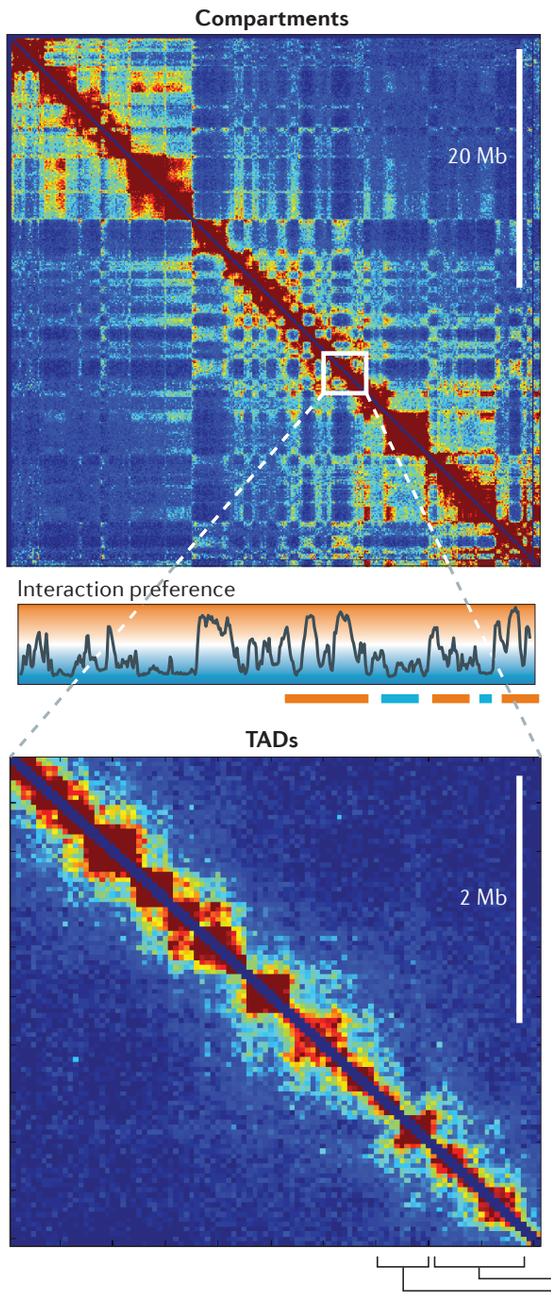
Most genes in Lamina Associated Domains are transcriptionally silent, suggesting that **lamina-genome interactions** are widely involved in the control of **gene expression**

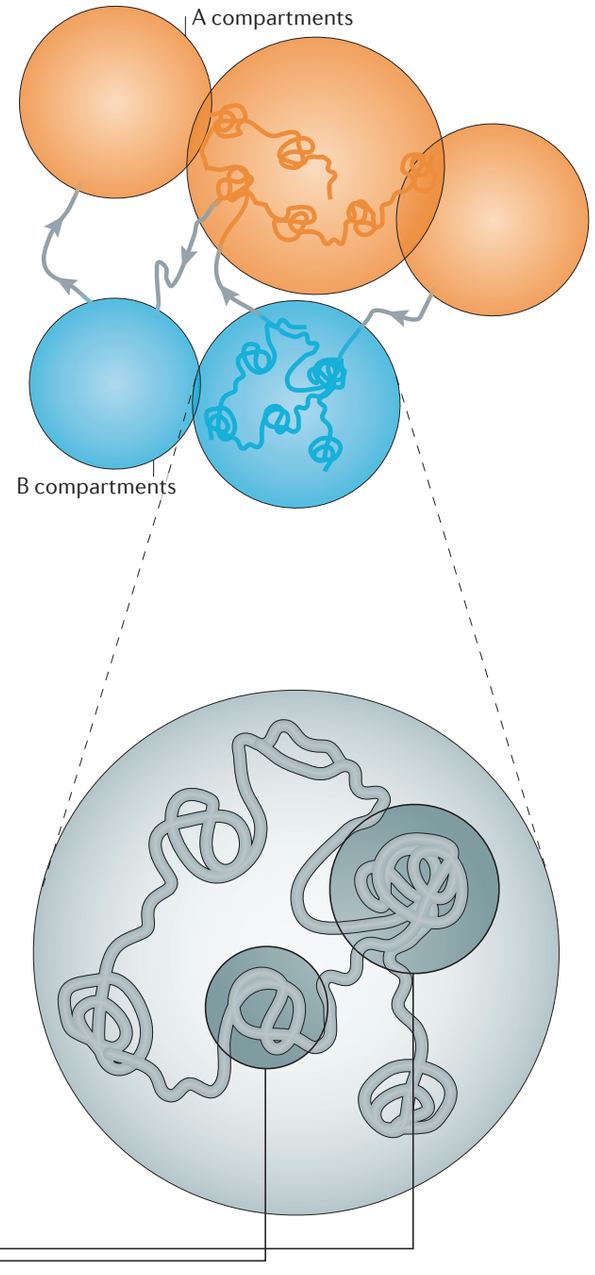
Adapted from Molecular Cell 38, 603-613, 2010

Level IV: Higher-order organization

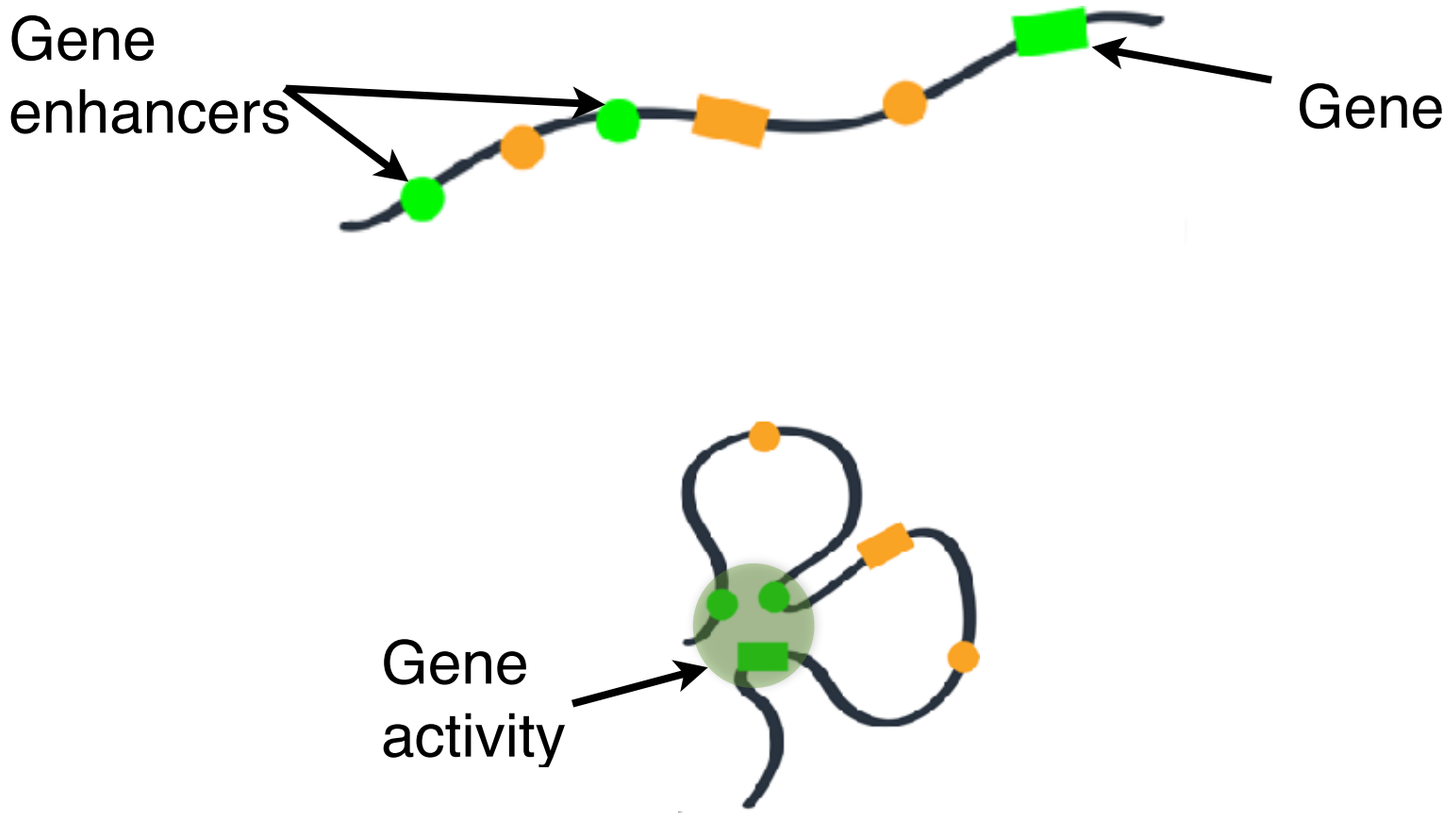
Dekker, J., Marti-Renom, M. A. & Mirny, L. A. Nat Rev Genet 14, 390–403 (2013).







Level V: Chromatin loops



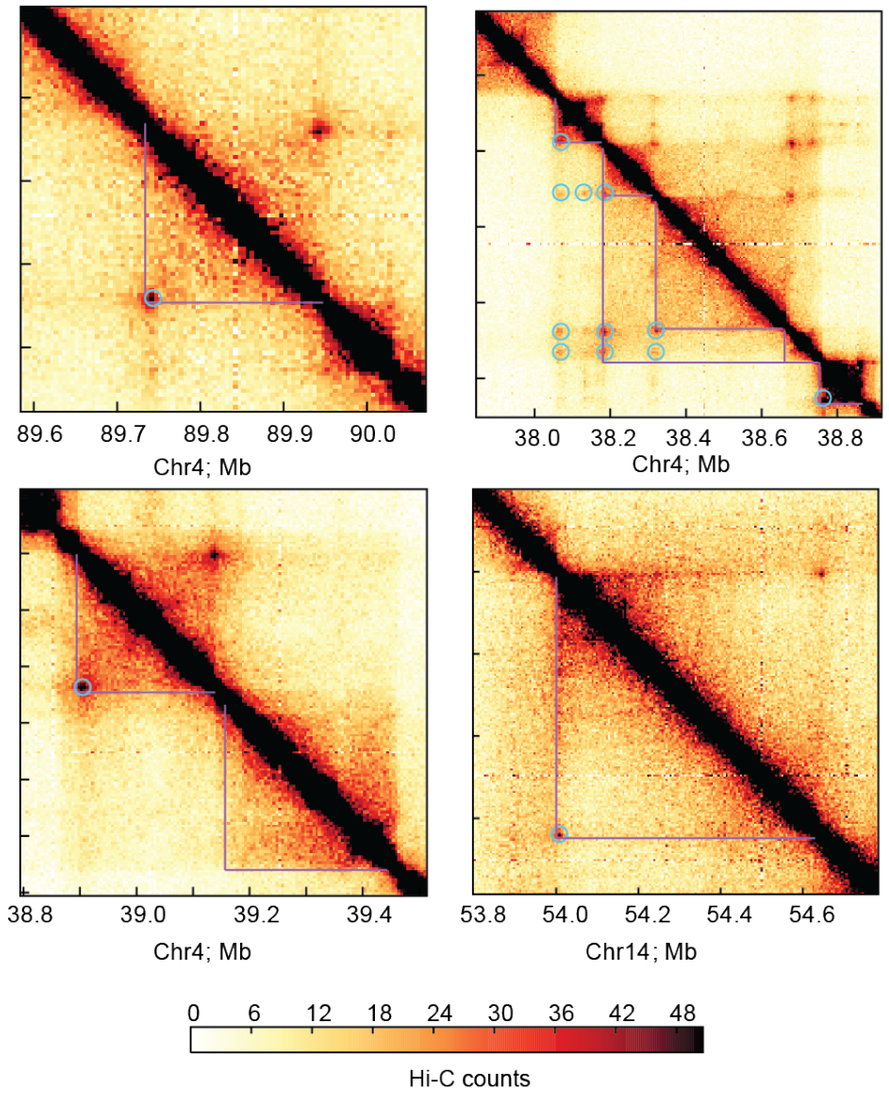
Loops bring distal genomic regions in close proximity to one another

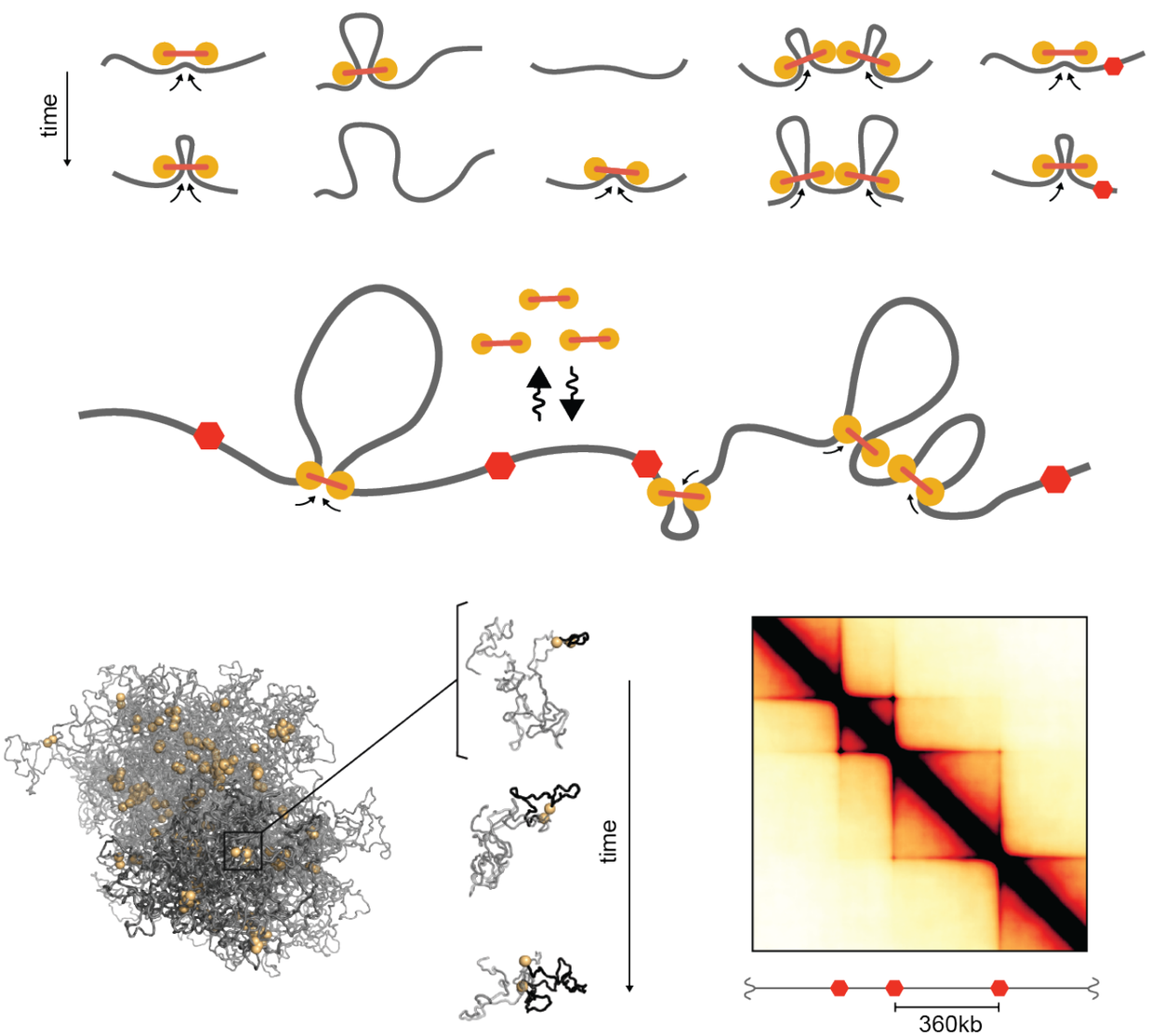
This in turn can have profound effects on gene transcription

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

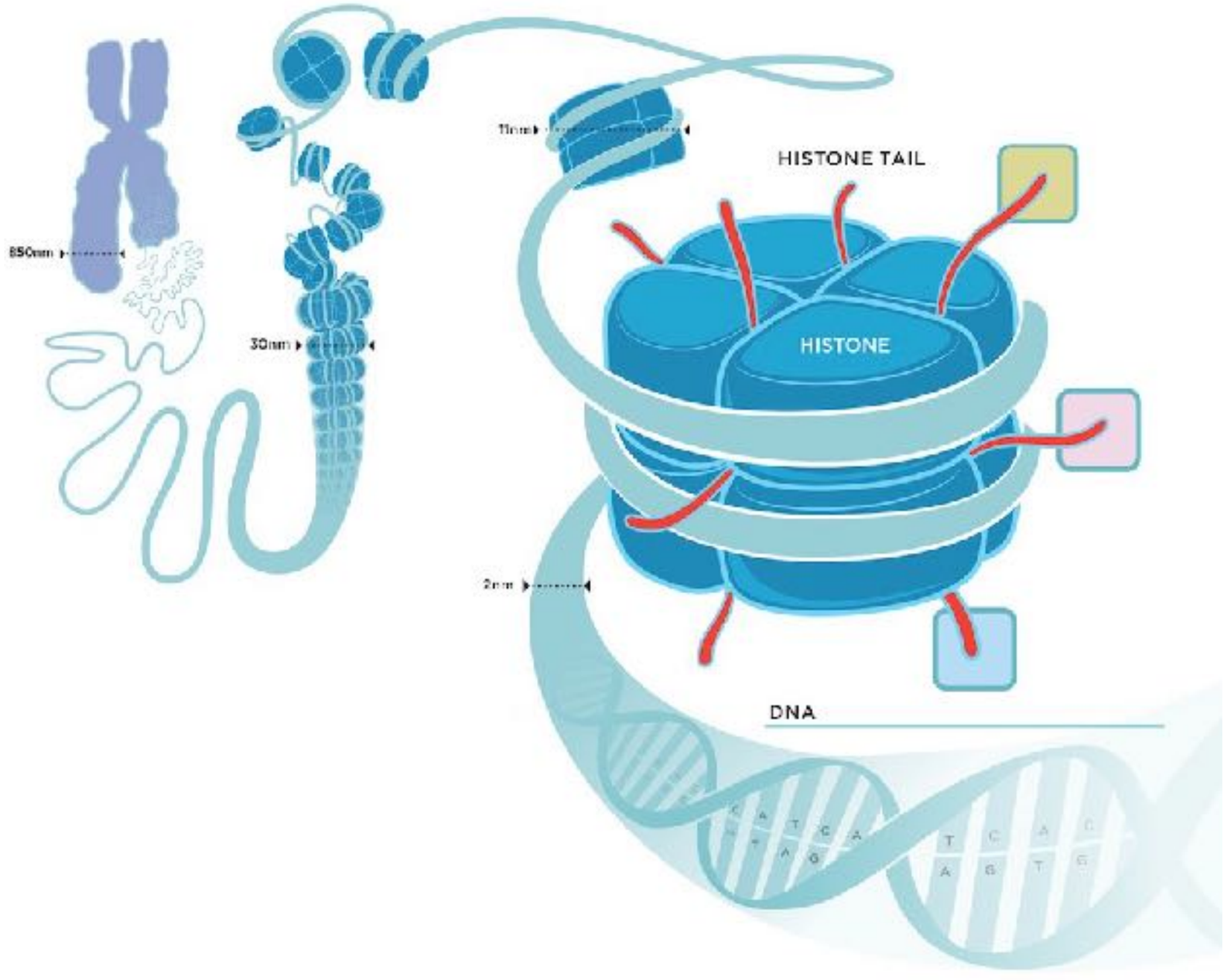
Fudenberg, G., Imakaev, M., Lu, C., Goloborodko, A., Abdennur, N., & Mirny, L. A. (2015).
Formation of Chromosomal Domains by Loop Extrusion. bioRxiv.





Level VI: Nucleosome

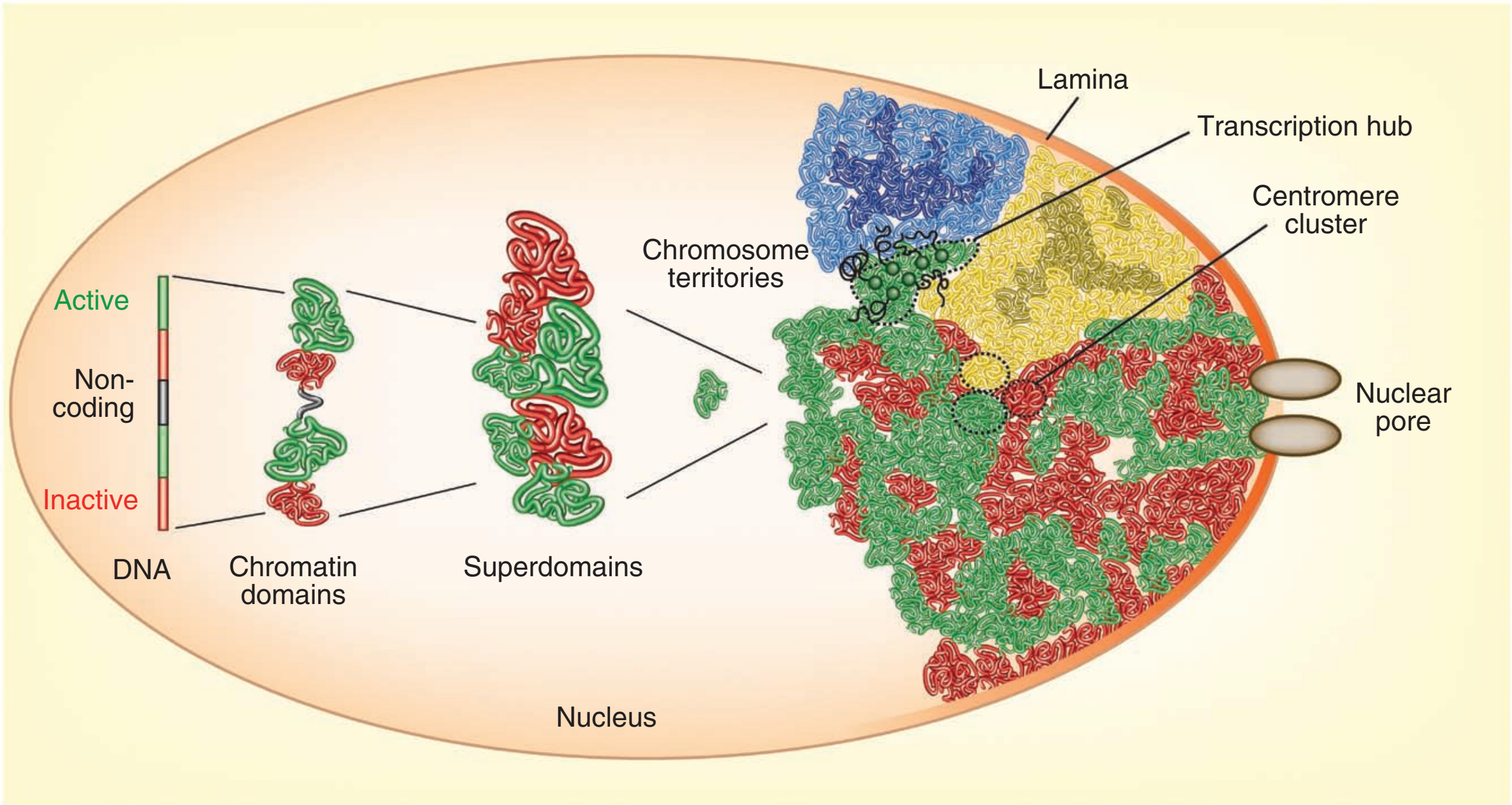
Chromosome **Chromatin fibre** **Nucleosome**



Adapted from Richard E. Ballermann, 2012

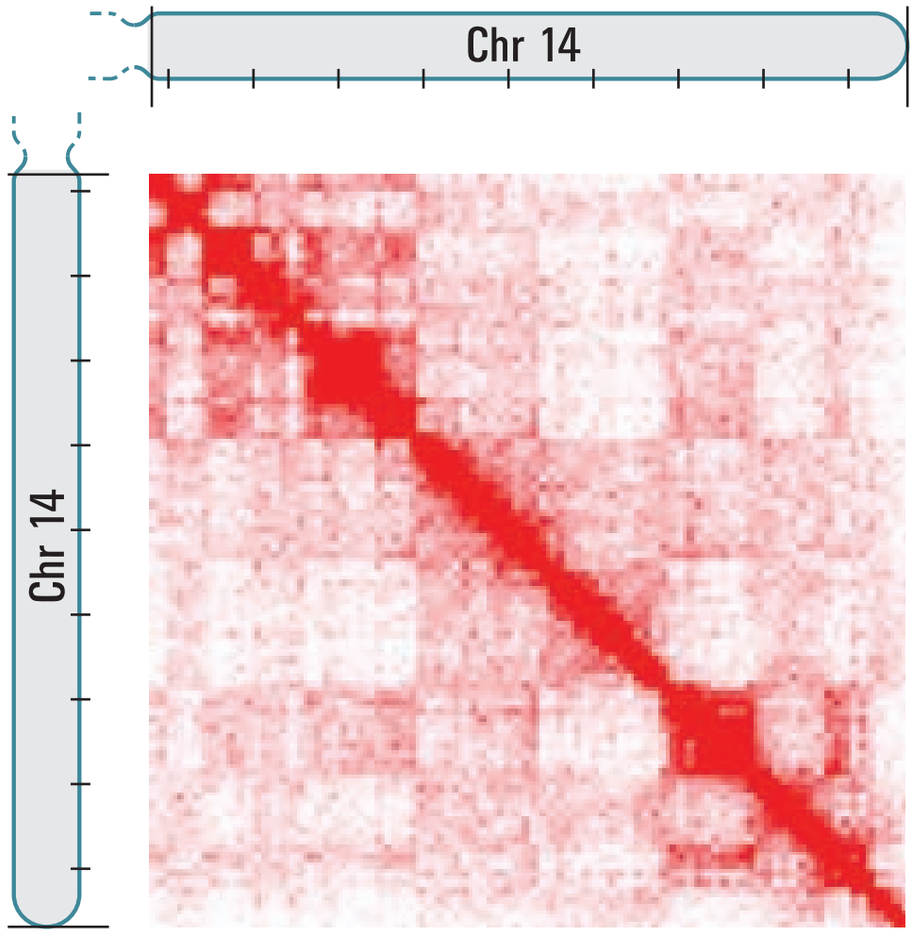
Complex genome organization

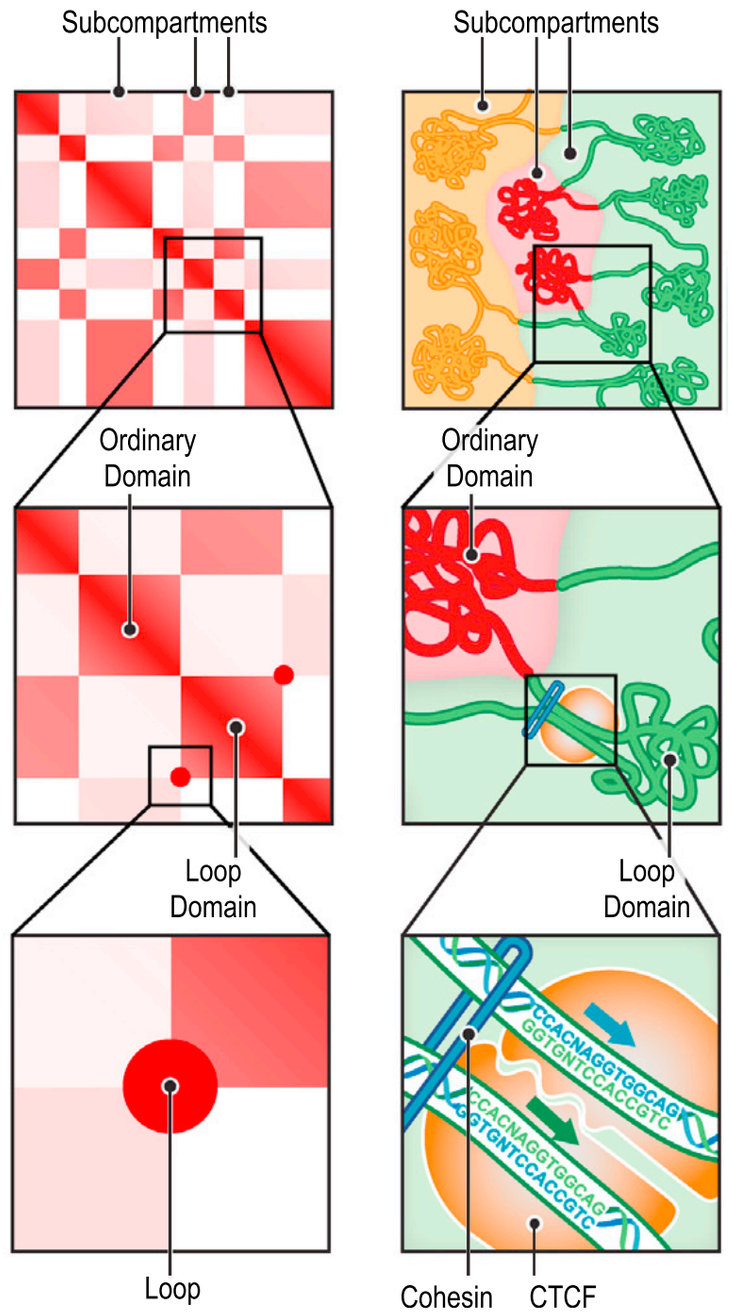
Cavalli, G. & Misteli, T. Functional implications of genome topology. Nat Struct Mol Biol 20, 290–299 (2013).



Marina Corral

Hierarchical genome organisation

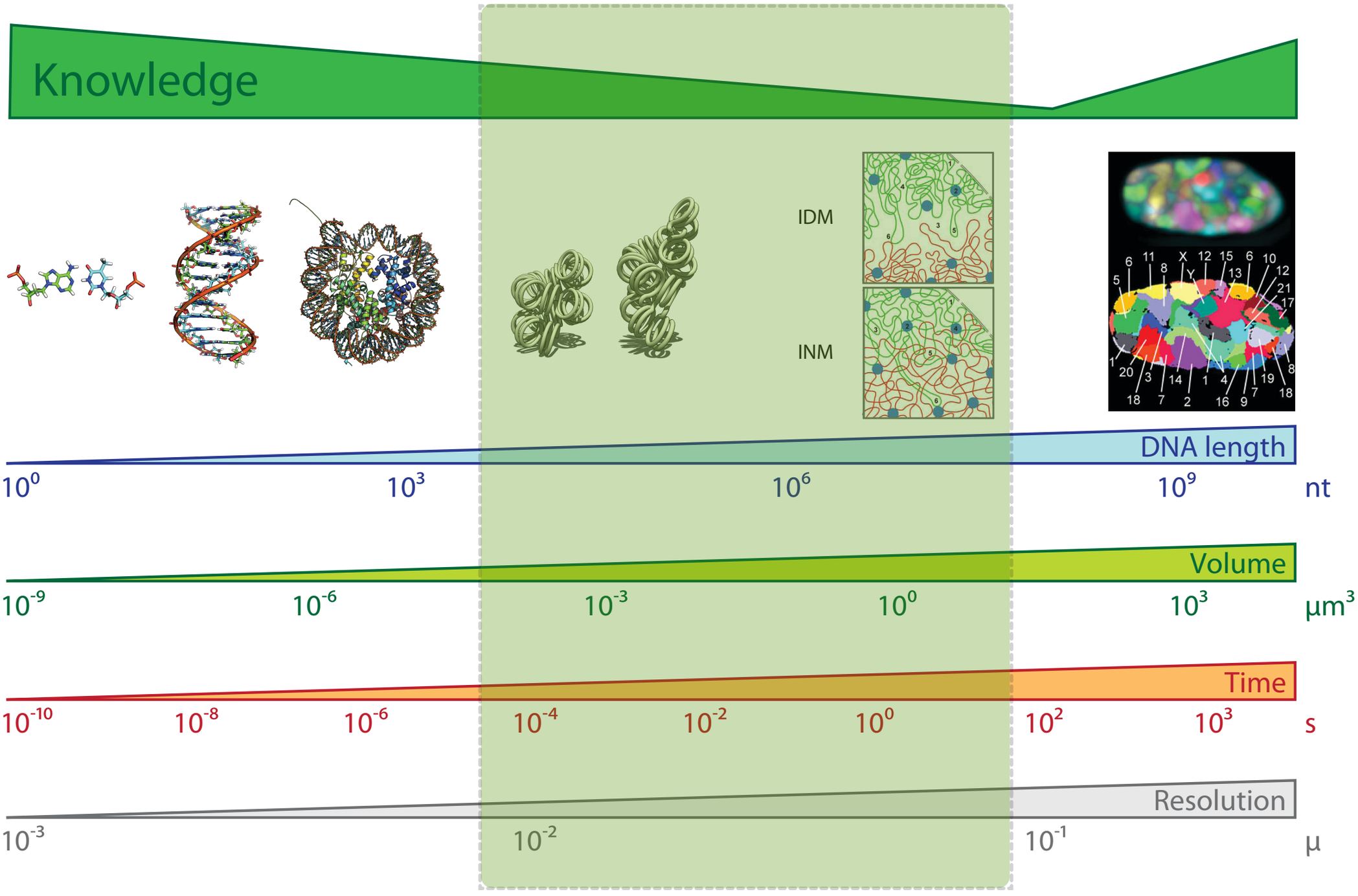




Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.
Rao, S. S. P., et al. (2014). Cell, 1–29.

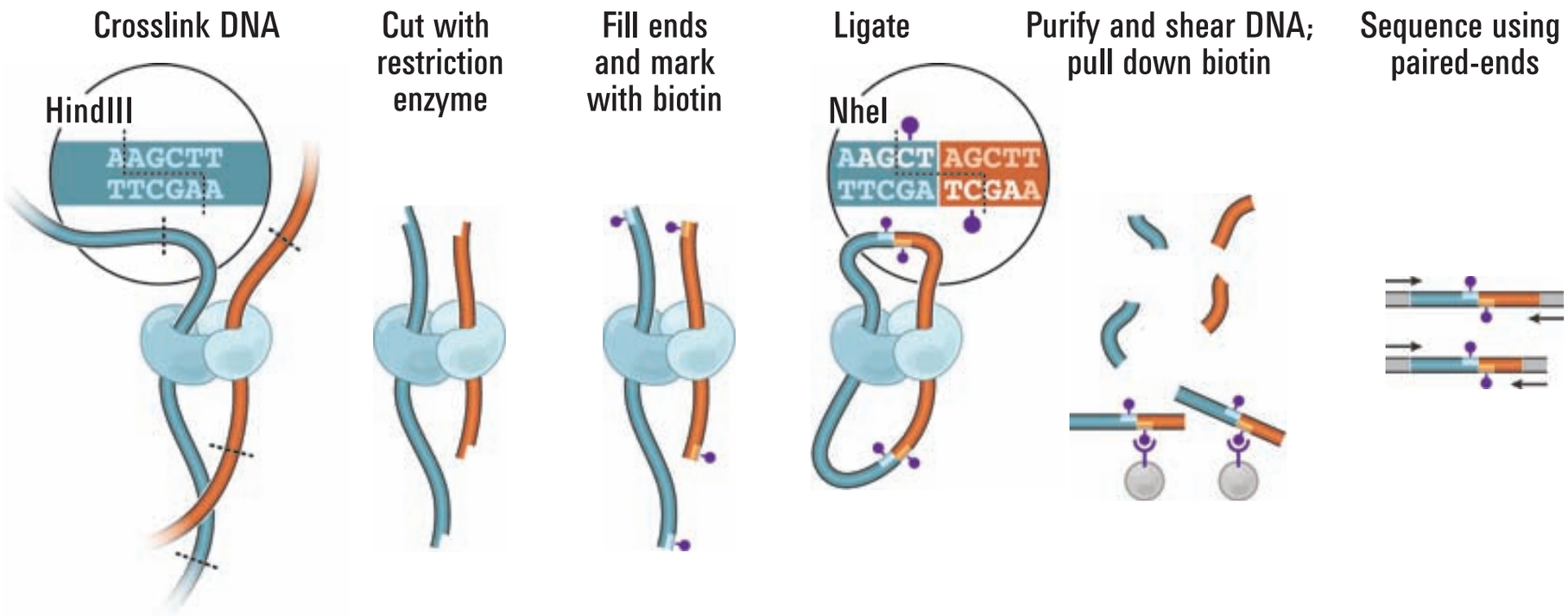
Resolution Gap

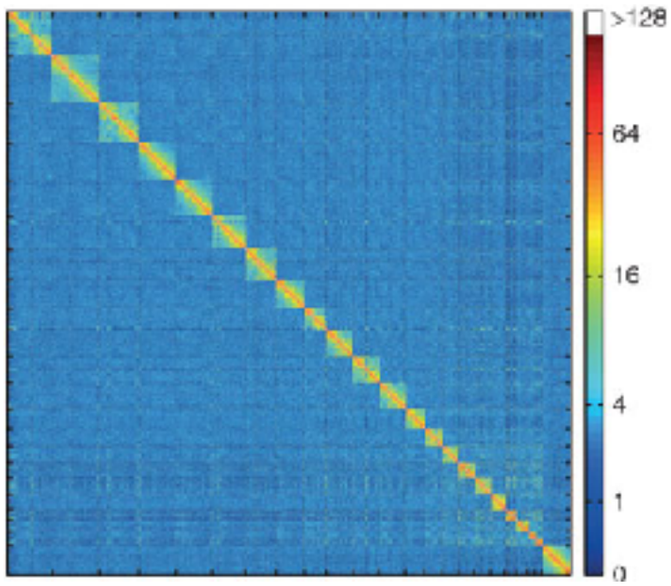
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol *7*, e1002125 (2011)



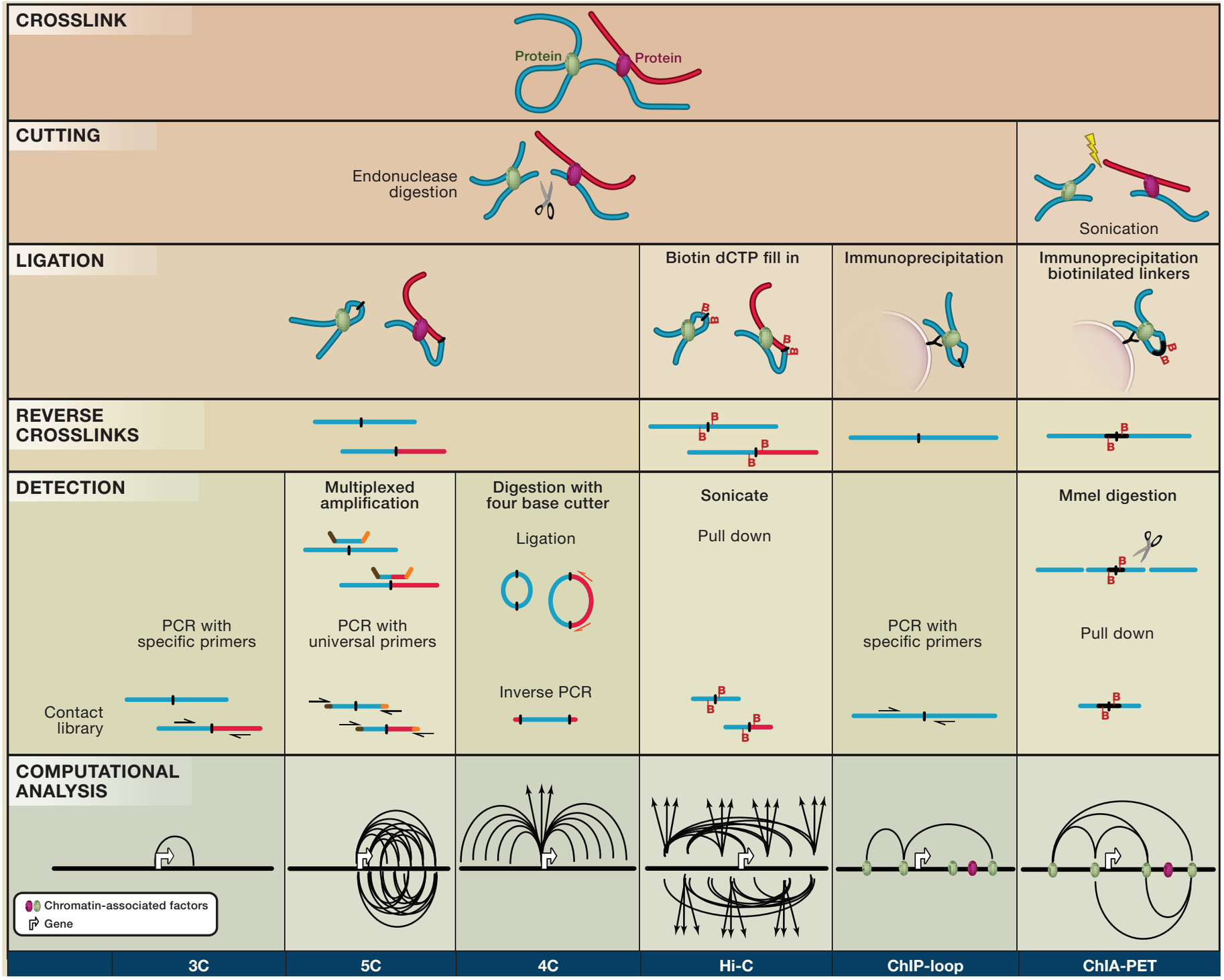
Chromosome Conformation Capture

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.



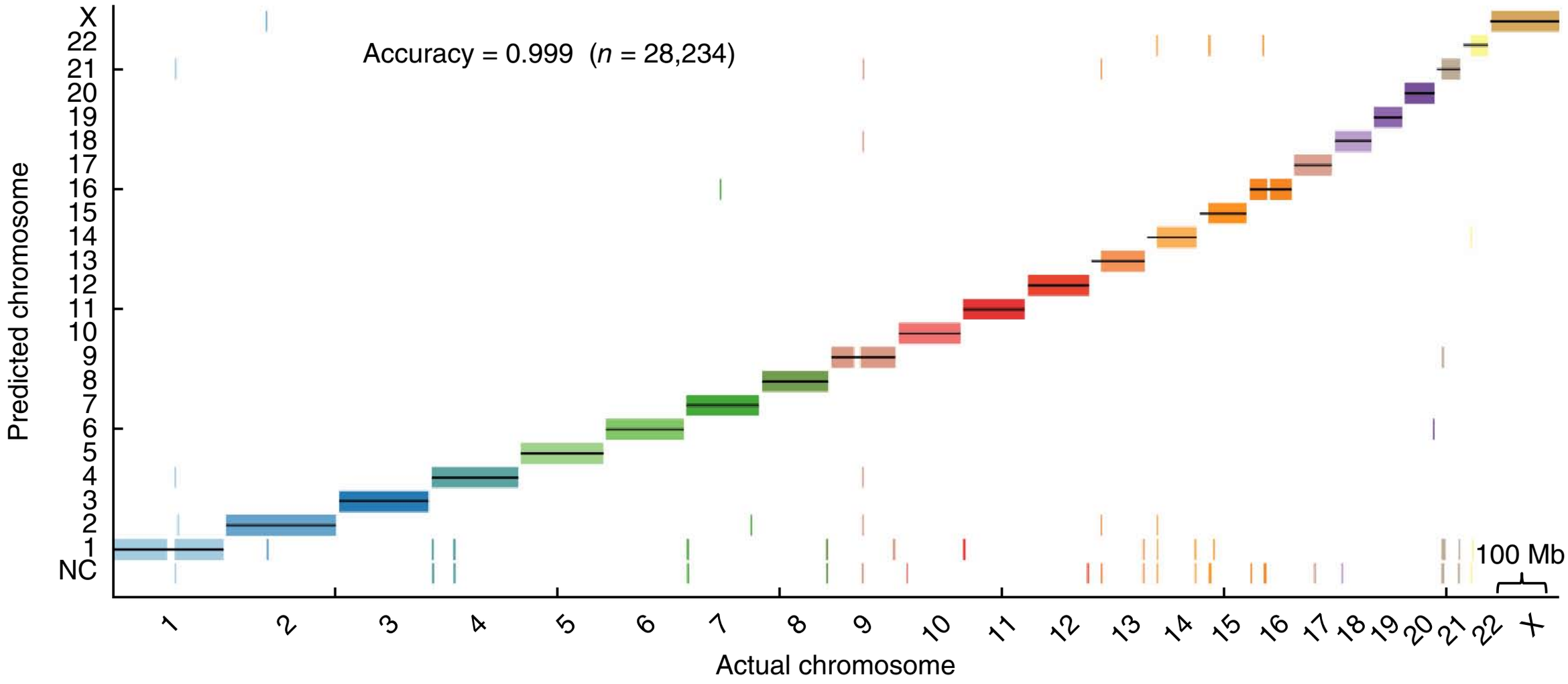


Chromosome Conformation Capture



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

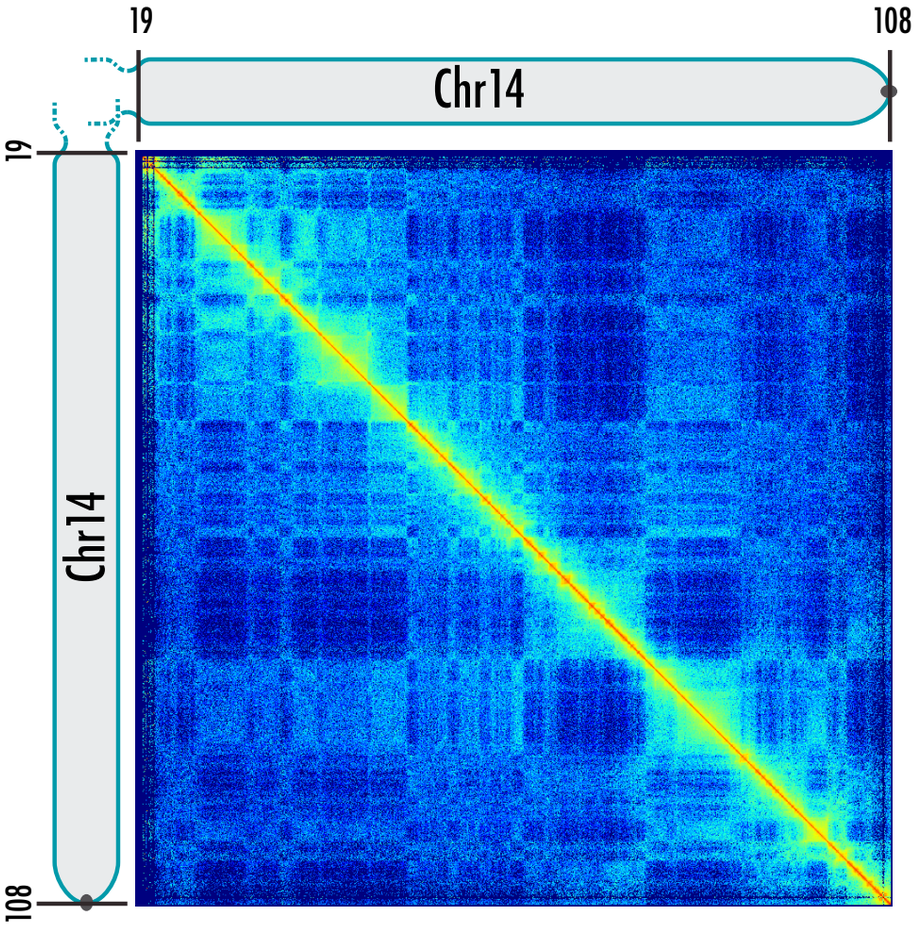
Chromosome Conformation Capture for de-novo assembly

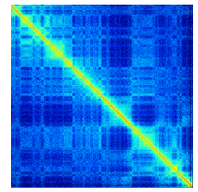


Kaplan, N., & Dekker, J. (2013). High-throughput genome scaffolding from in vivo DNA interaction frequency. Nature Biotechnology, 31(12), 1143–1147.

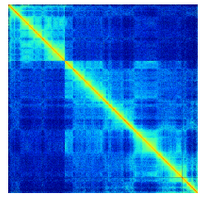
Great apes lymphoblast maps

Chromosome 14

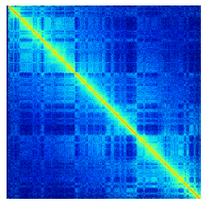




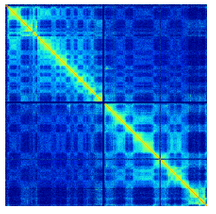
Chimpanzee



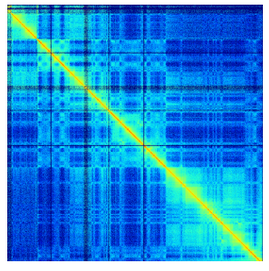
Gorilla



Orangutan



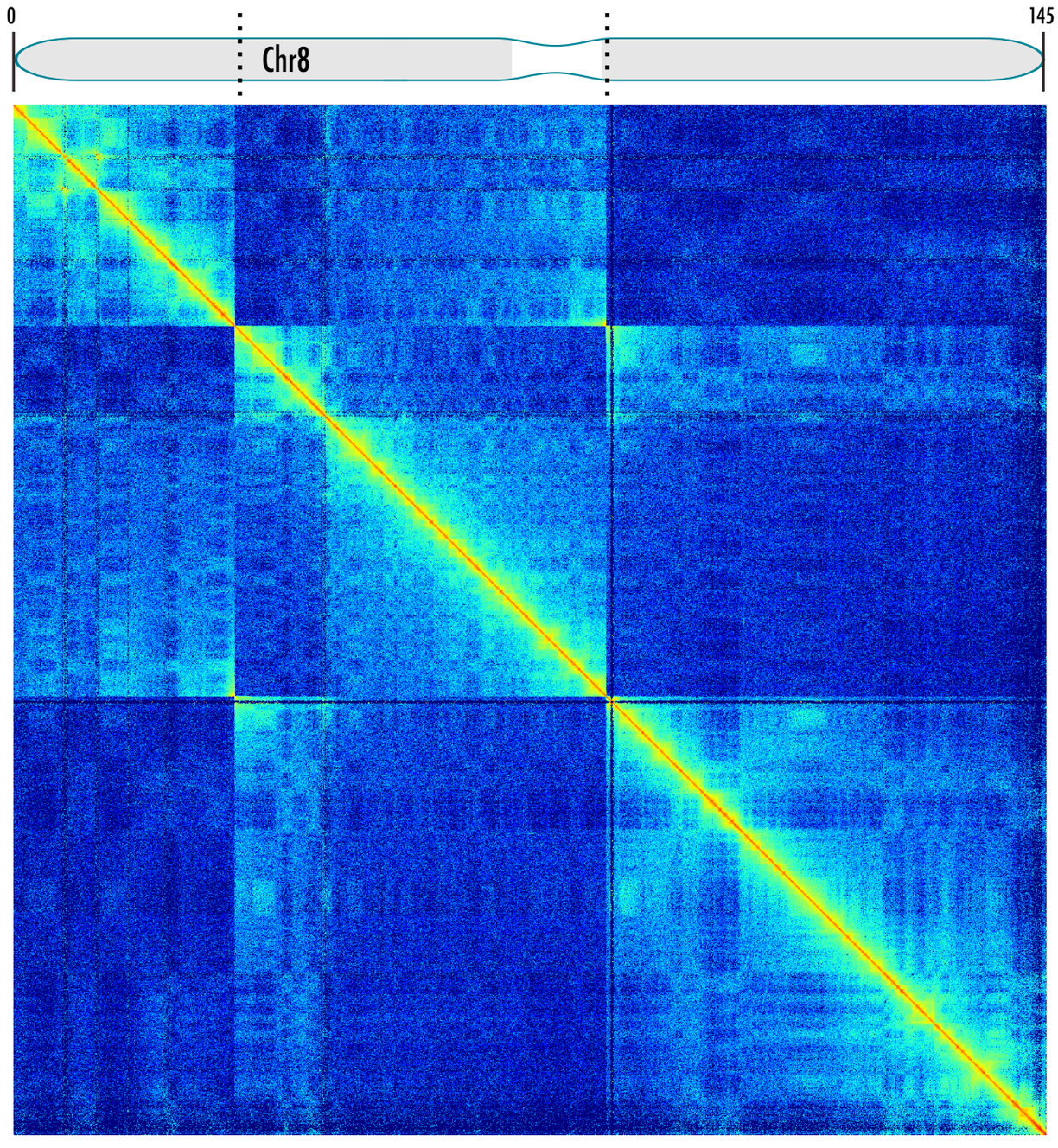
Gibbon



Mouse

Assembly error detection

Chromosome 8 Gorilla

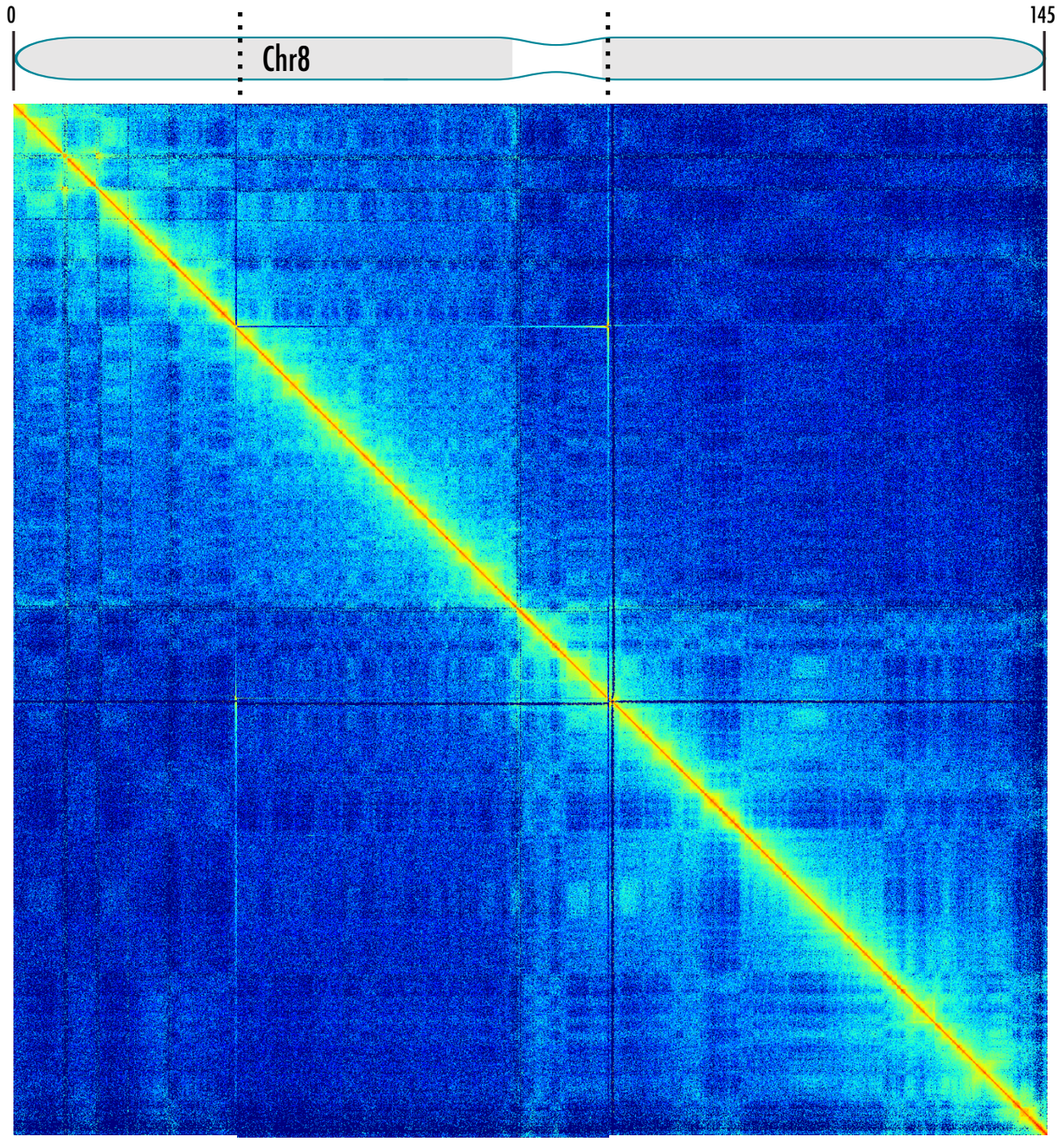


GGO8 has an inversion of the region corresponding to HSA8:30.0-86.9Mb

Aylwyn Scally (Department of Genetics, University of Cambridge)

Assembly error detection

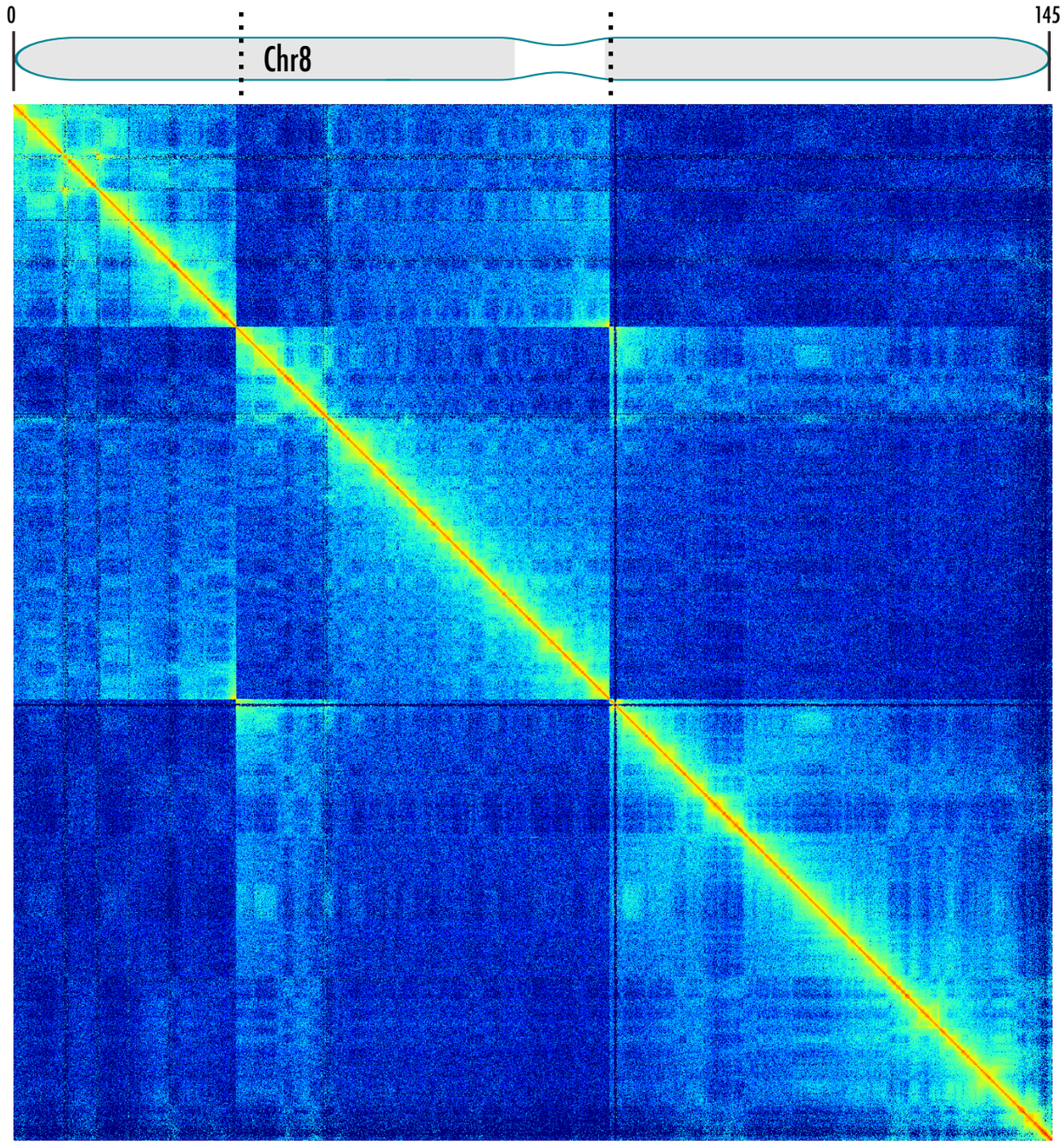
Chromosome 8 Gorilla



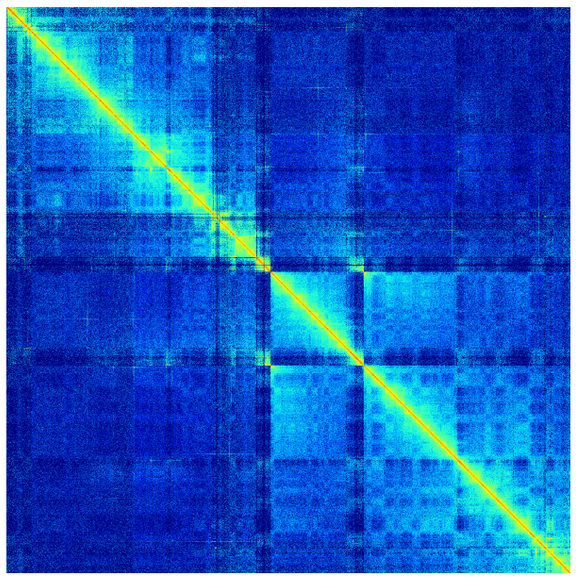
GGO8 has an inversion of the region corresponding to HSA8:30.0-86.9Mb
Aylwyn Scally (Department of Genetics, University of Cambridge)

Assembly error detection

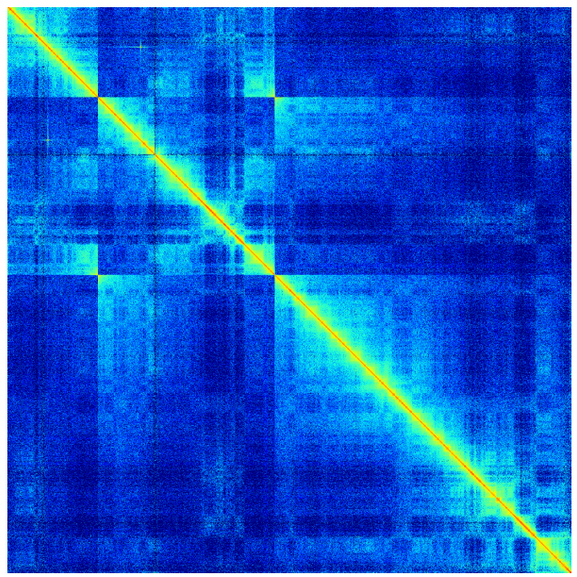
Chromosome 8 Gorilla



Chr 7

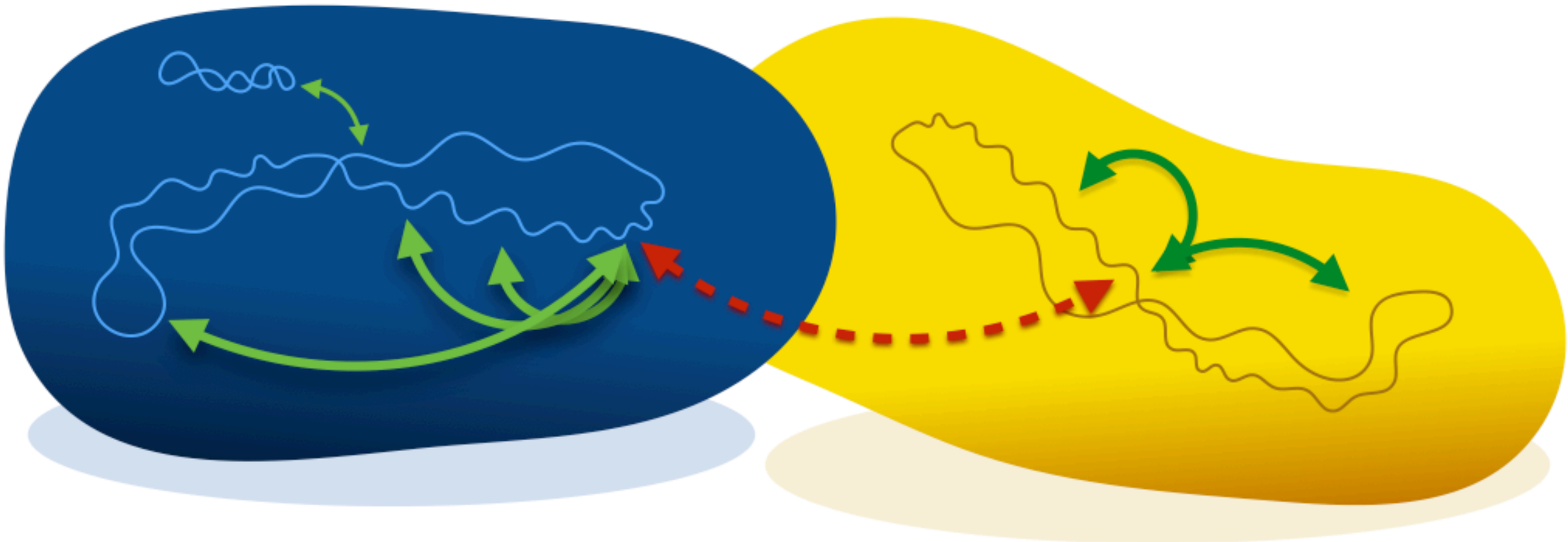


Chr 12



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

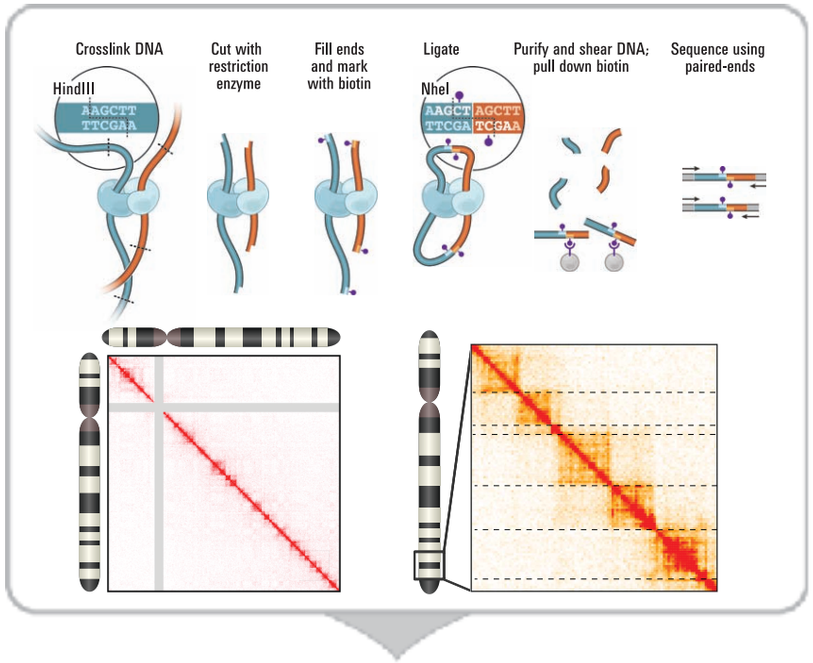


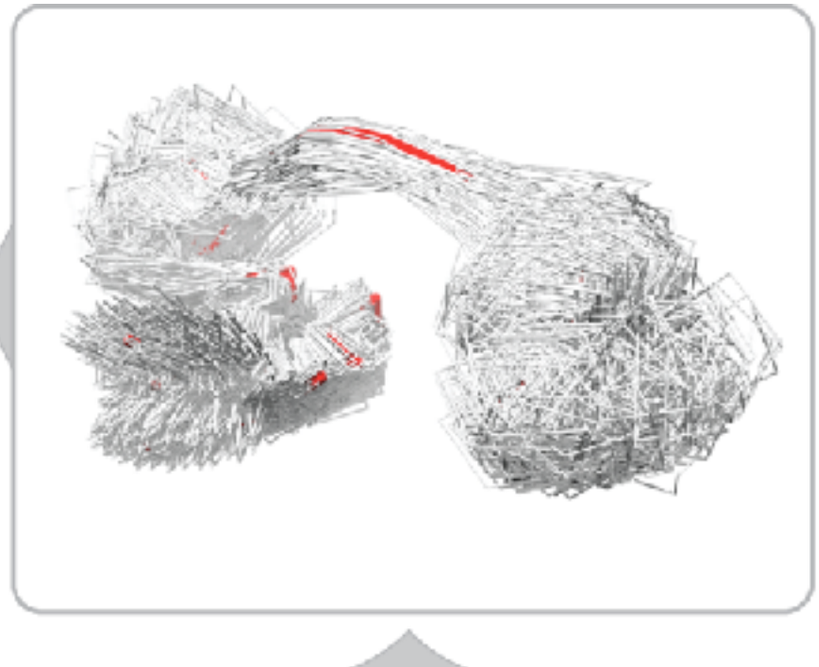
Beitel, C. W., Froenicke, L., Lang, J. M., Korf, I. F., Michelmore, R. W., Eisen, J. A., & Darling, A. E. (2014). Strain- and plasmid-level deconvolution of a synthetic metagenome by sequencing proximity ligation products. doi:10.7287/peerj.preprints.260v1

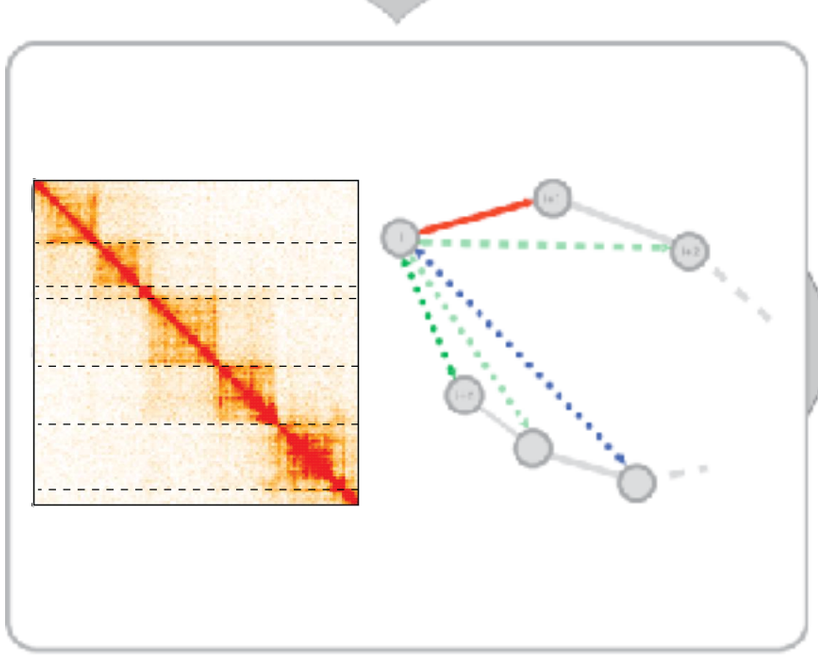
Hybrid Method

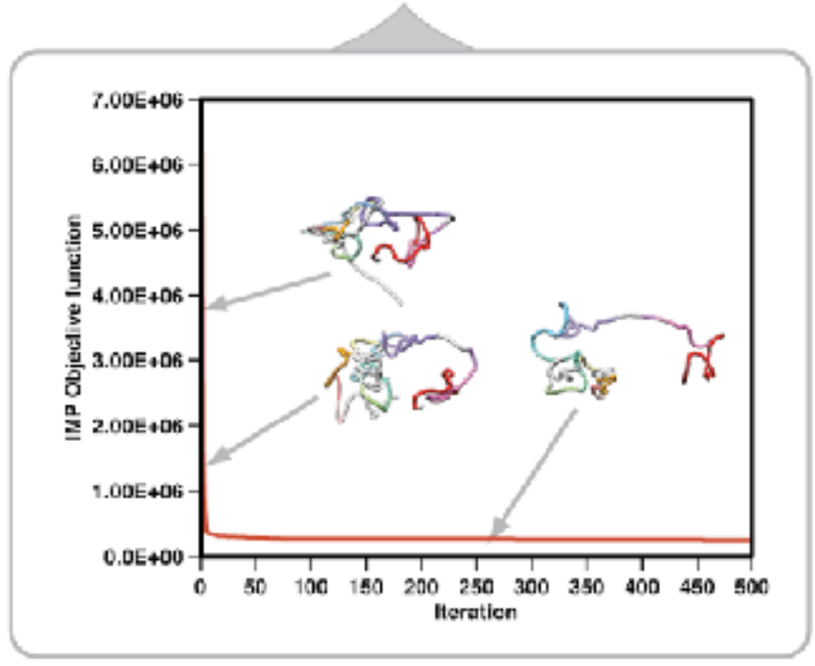
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).

Experiments



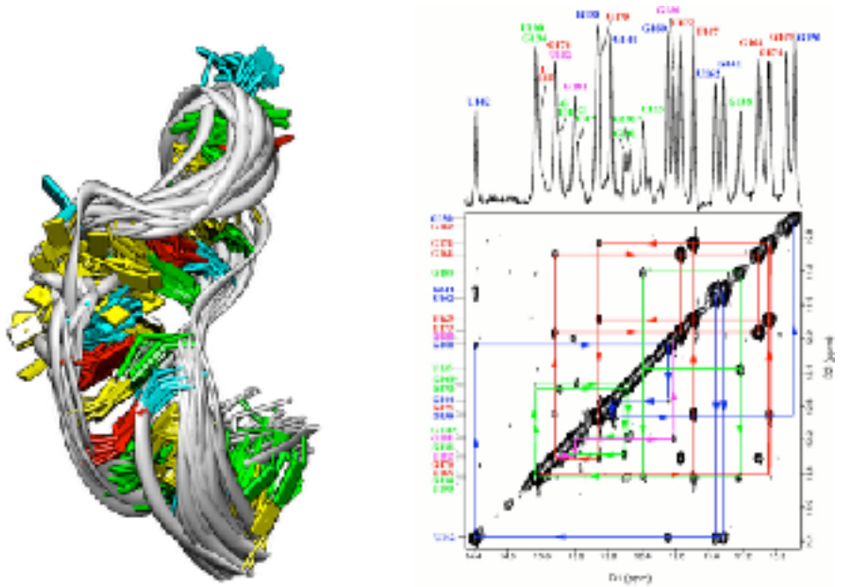




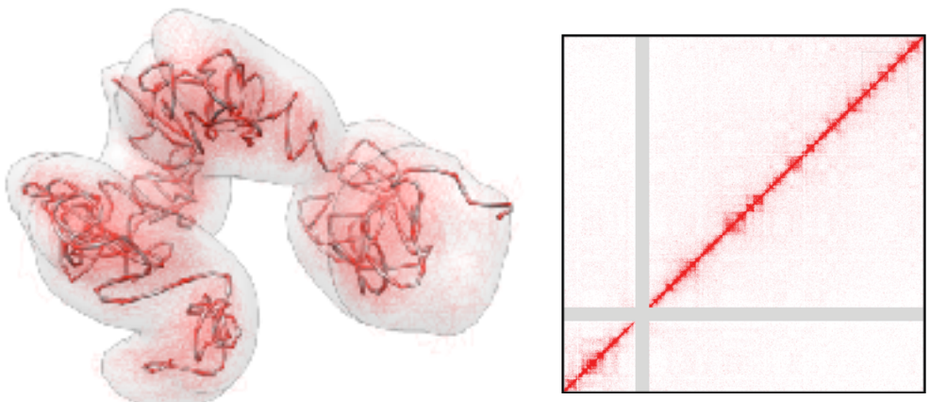


Computation

Structure determination using Hi-C data



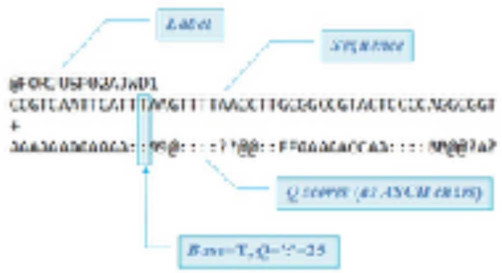
Biomolecular structure determination
2D-NOESY data



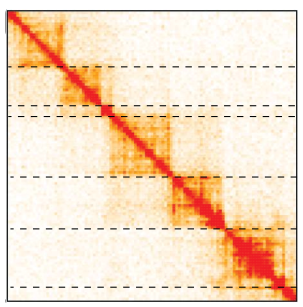
Chromosome structure determination
3C-based data

TADbit

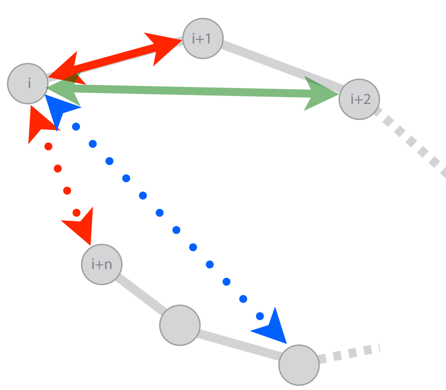
http://3DGenomes.org



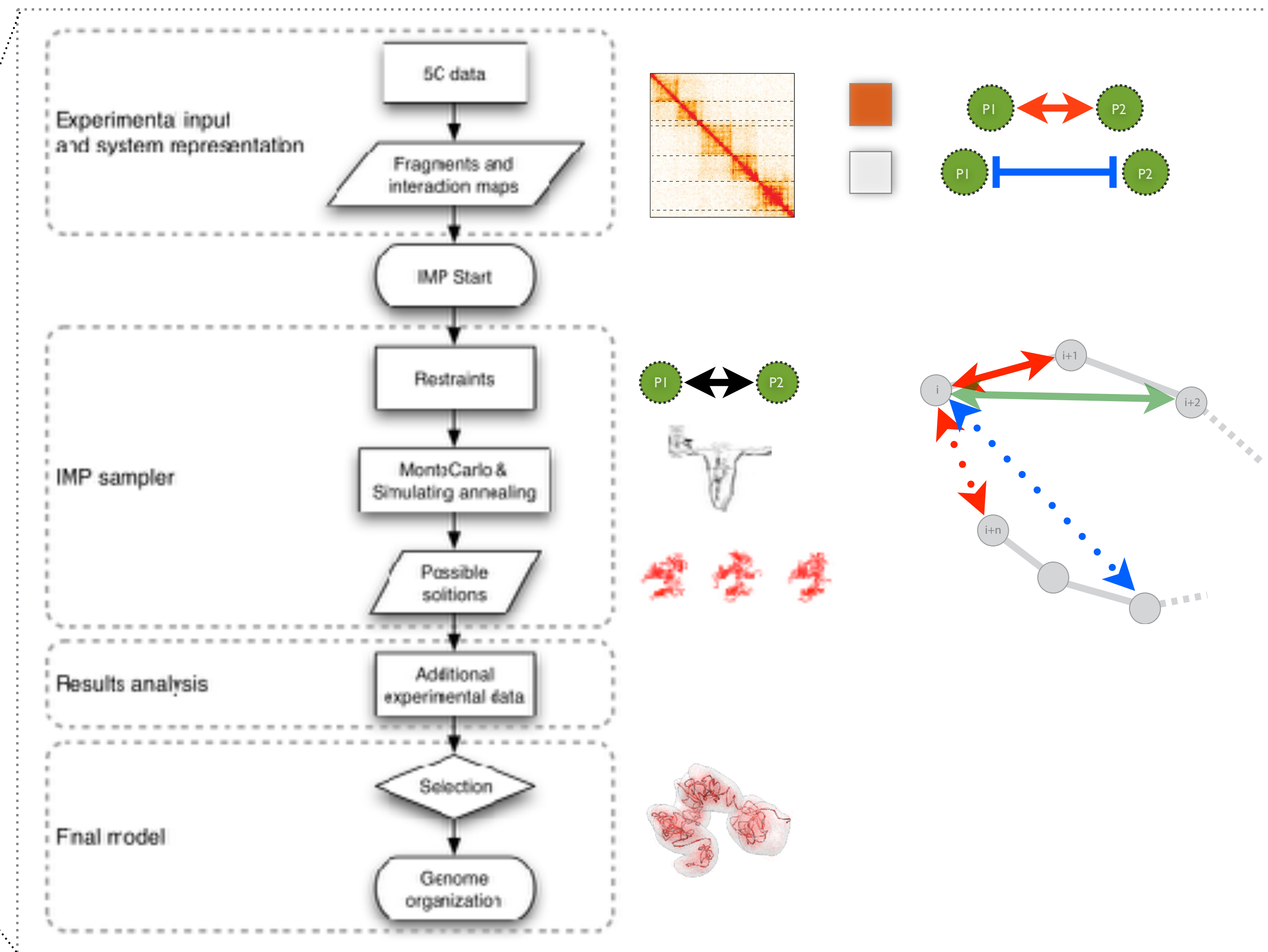
FastQ files to Maps

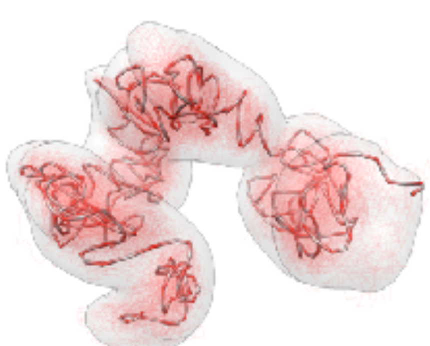


Map analysis



Model building



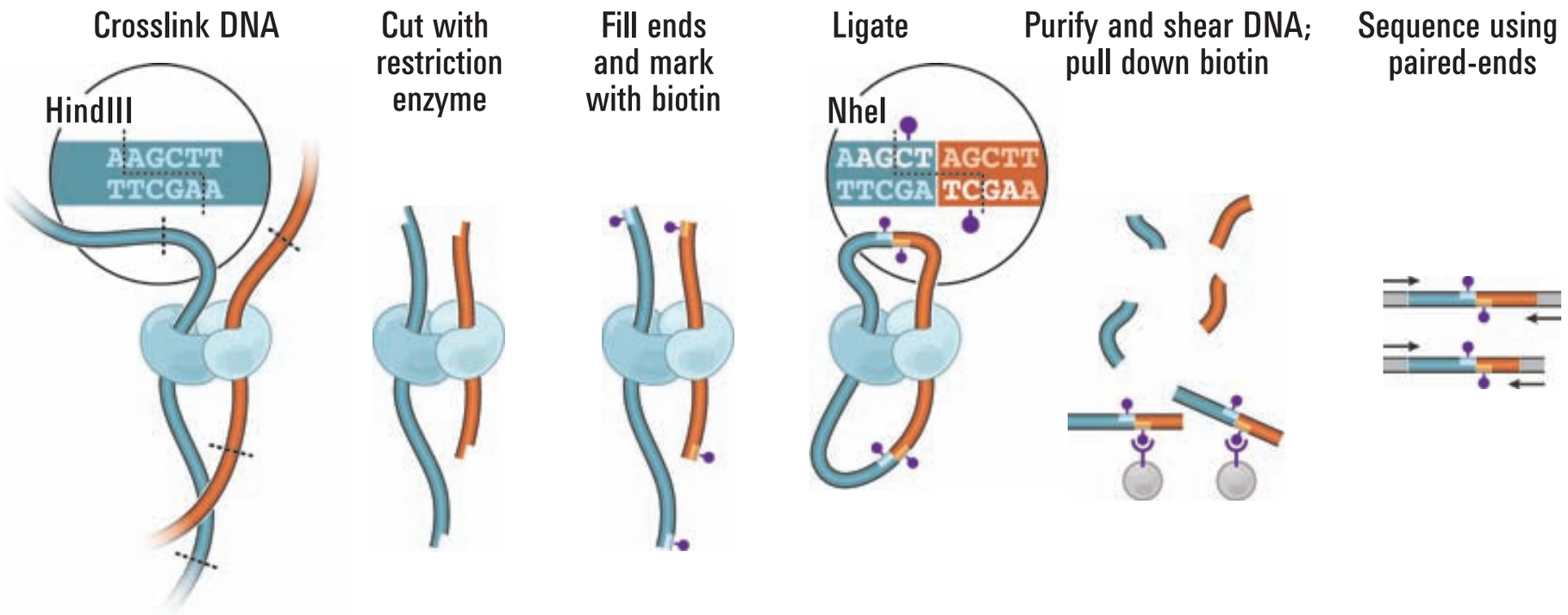


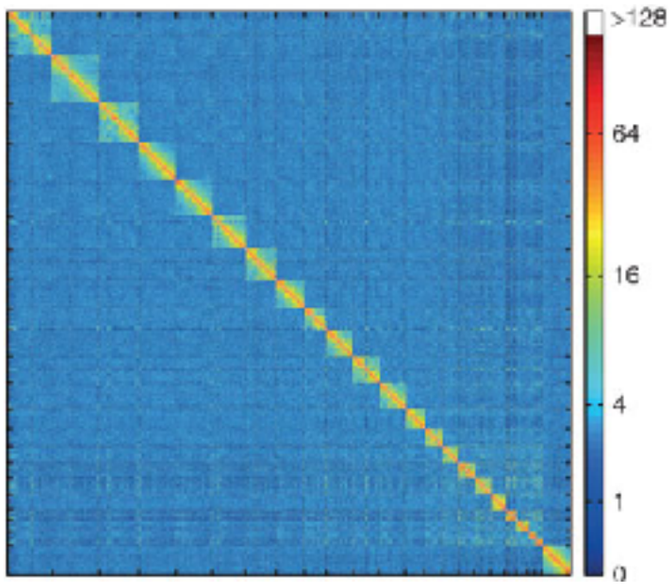
Model analysis


Got FASTQ?

Hi-C experiment

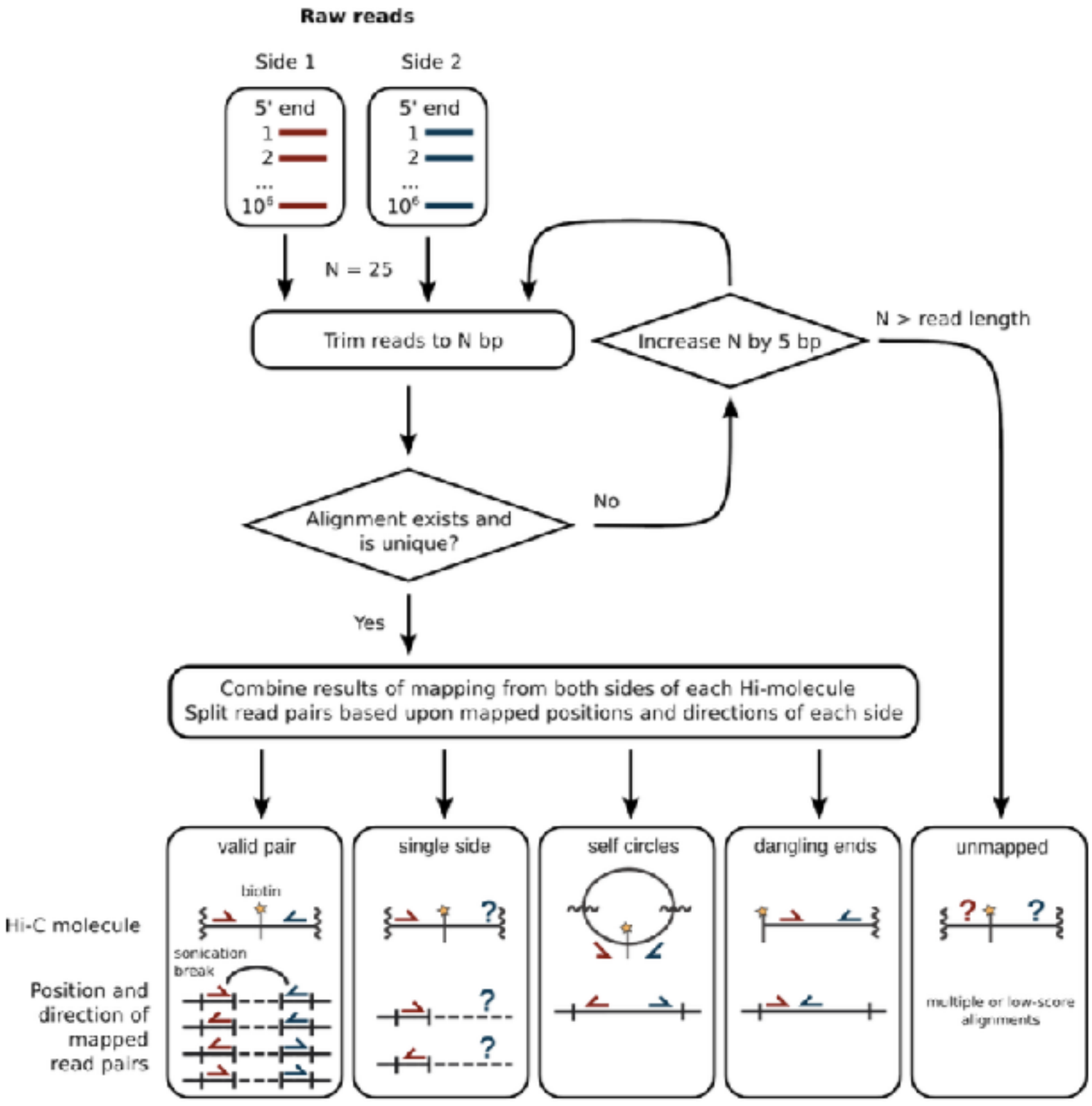
Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.





Mapping & Filtering

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

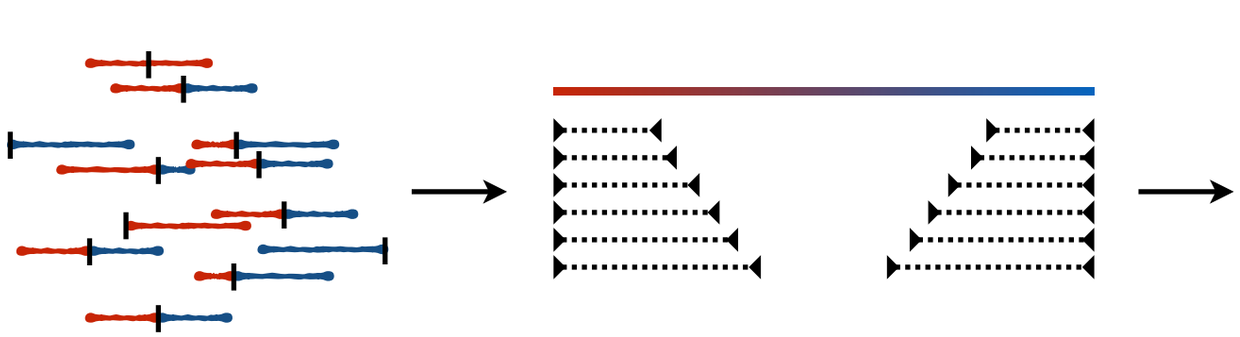


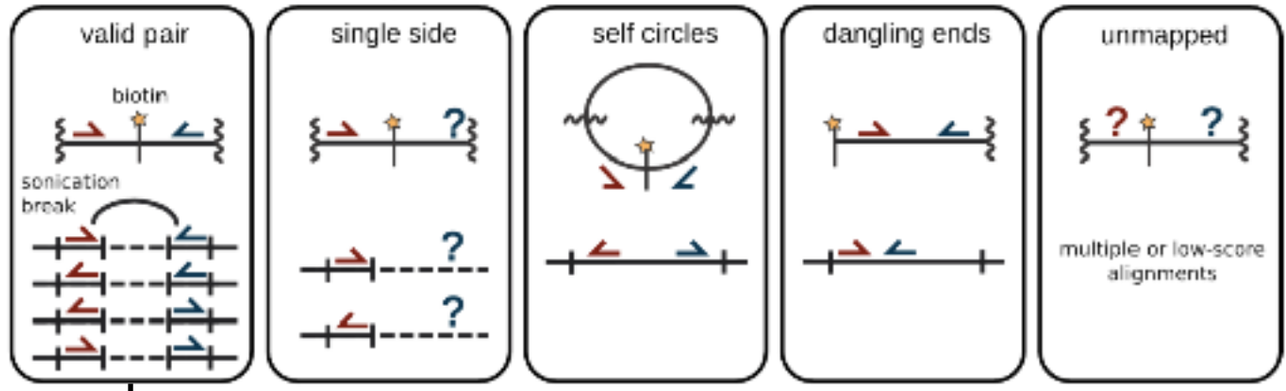
How much you normally map?

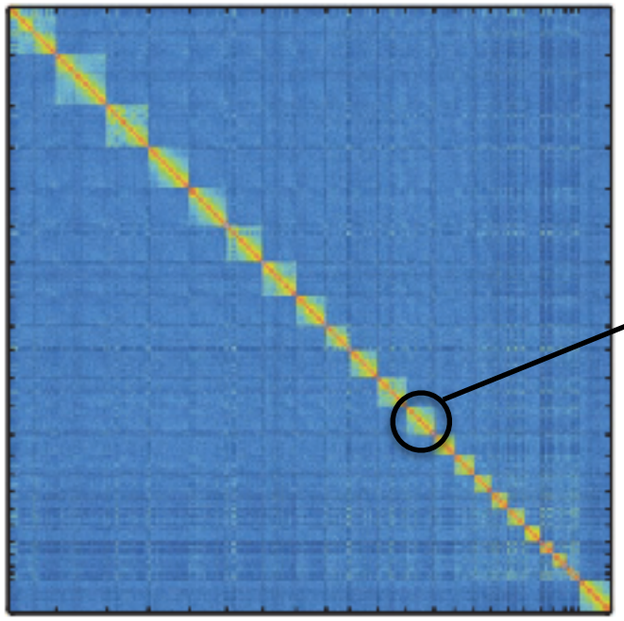
- 80-90% each end => 60-80% intersection
- ~1% multiple contacts
- Many of intersecting pairs will be lost in filtering…
- Final 40-60% of valid pairs
- One measure of quality is the CIS/TRANS ration (70-80% good)

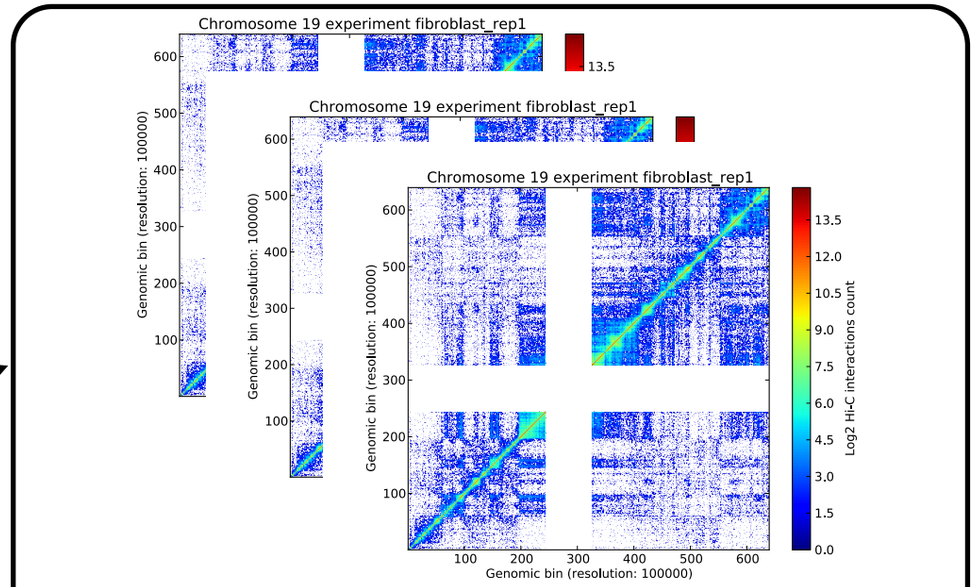

Got mapped reads?

Interaction matrices









3D modeling

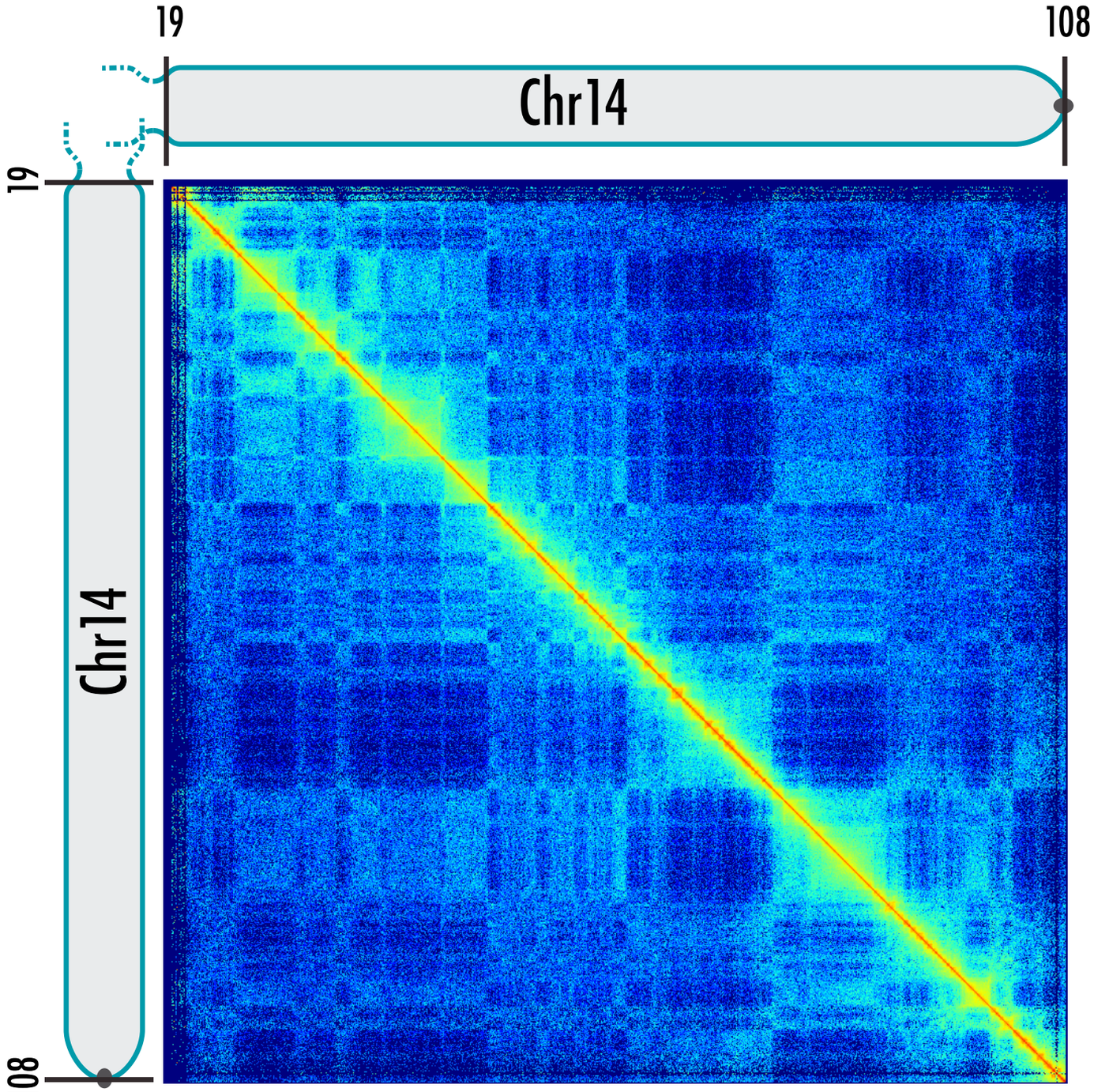
Matrix analysos

Structural Models

TADs & A/B

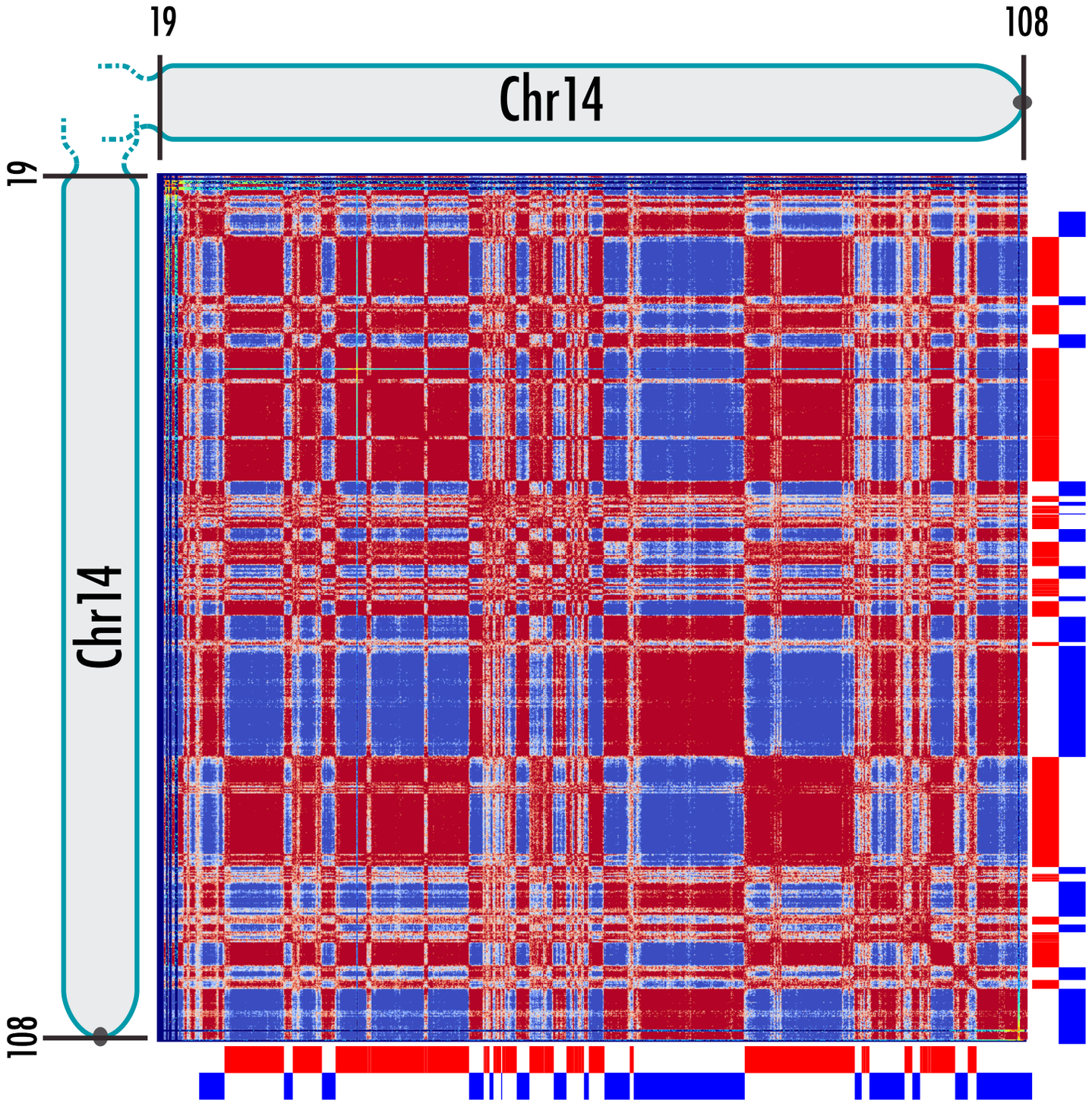
A/B Compartment

Chromosome 14



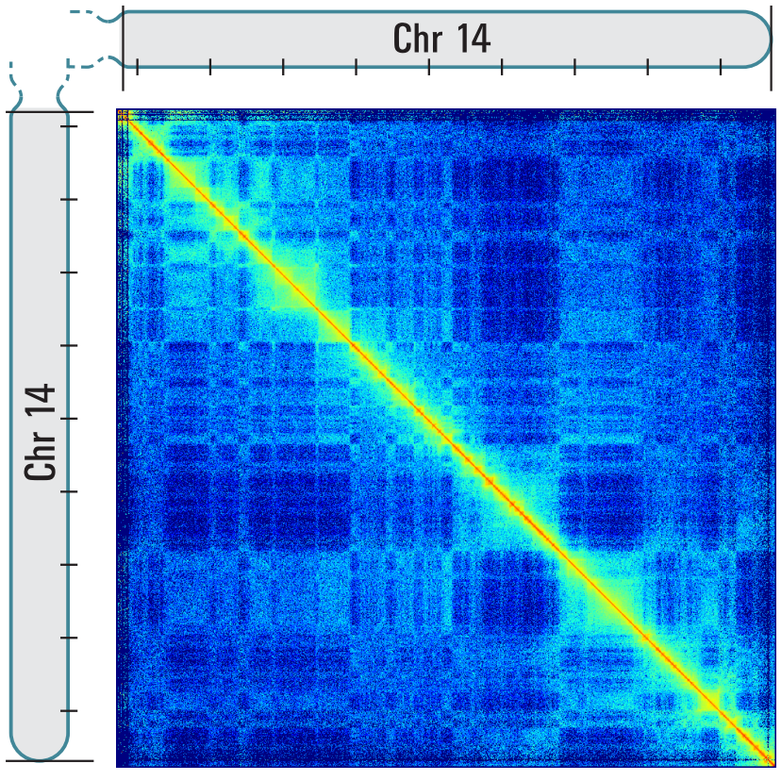
A/B Compartment

Chromosome 14



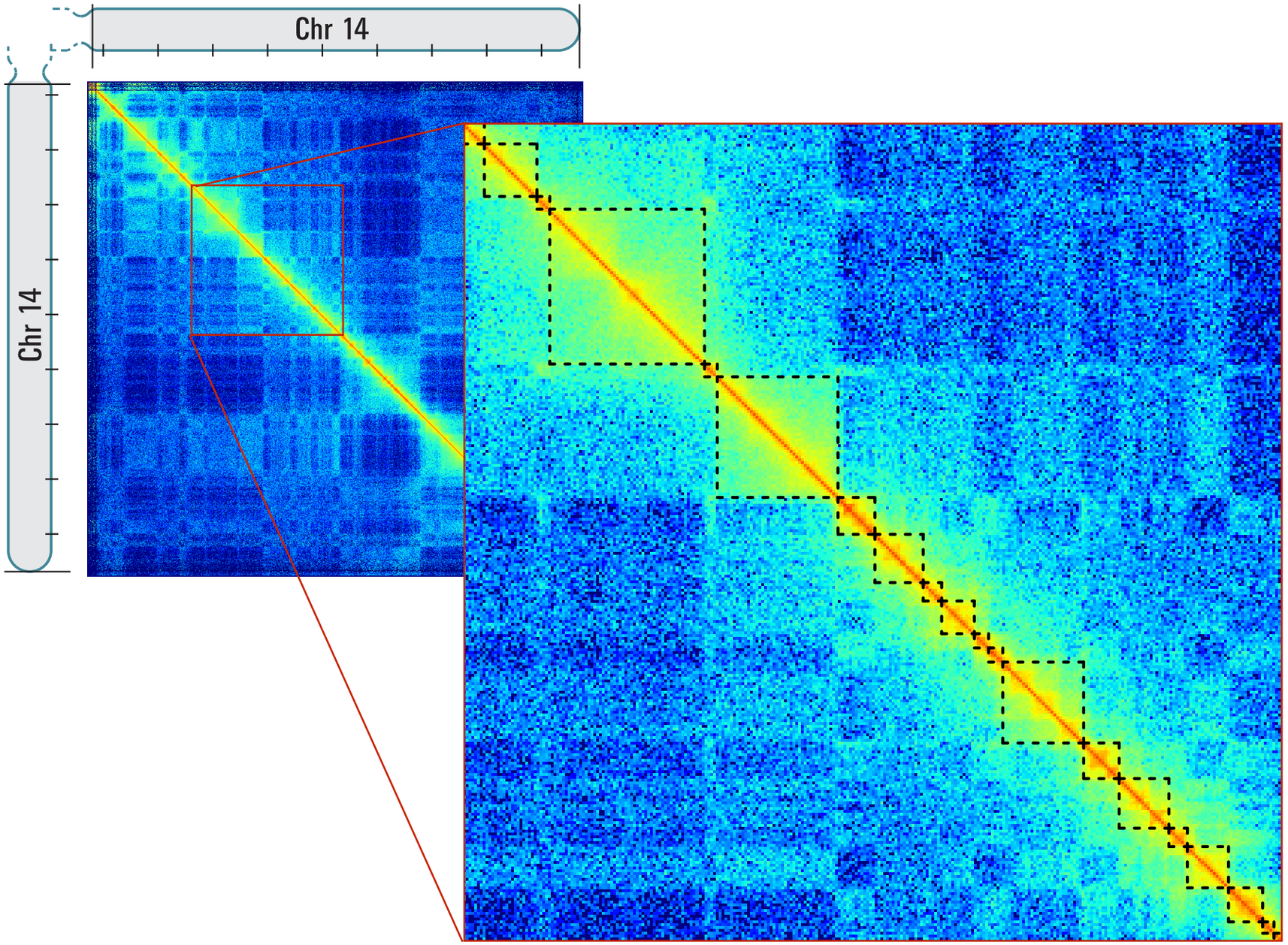
TADs

Chromosome 14



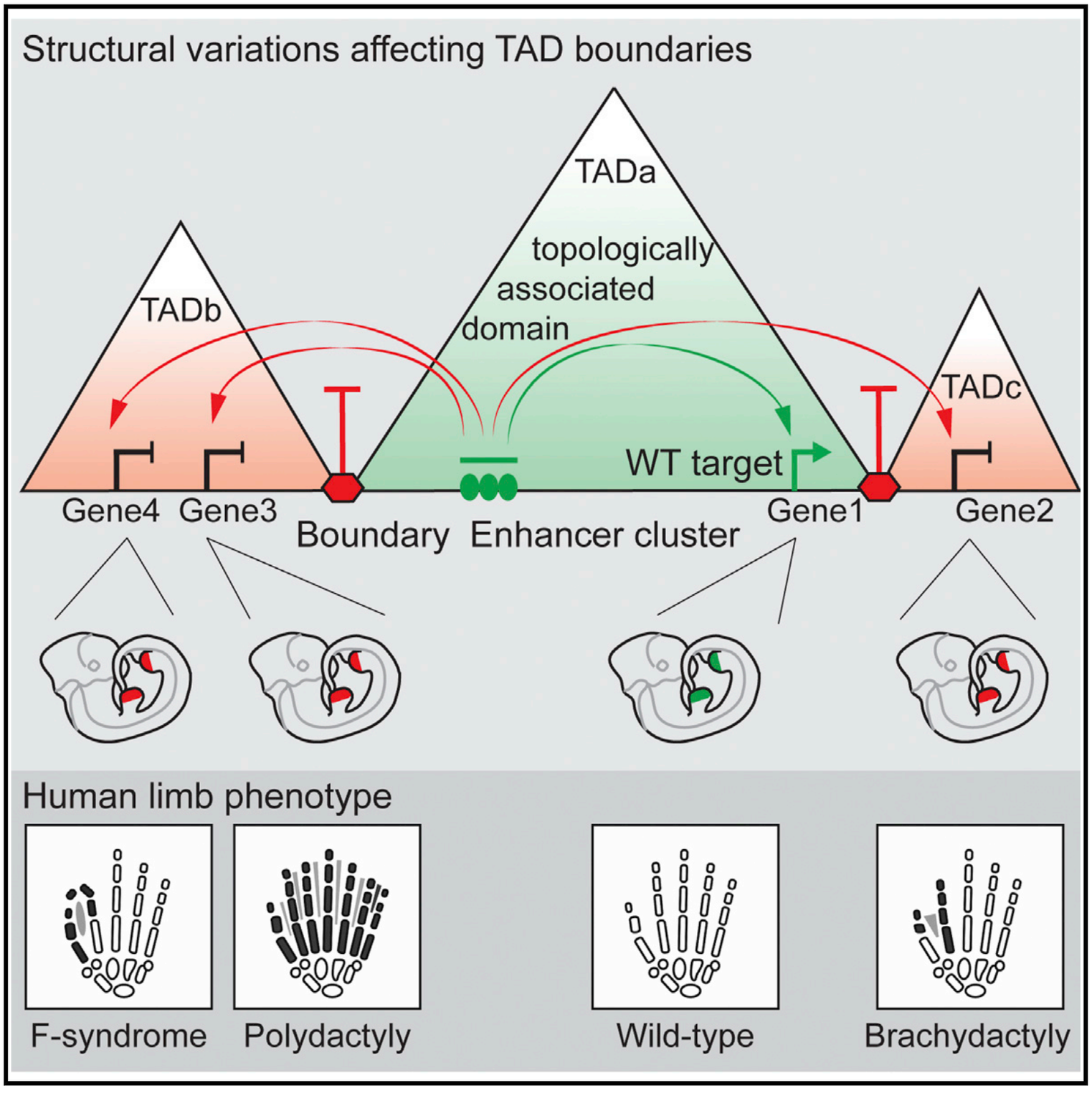
TADs

Chromosome 14



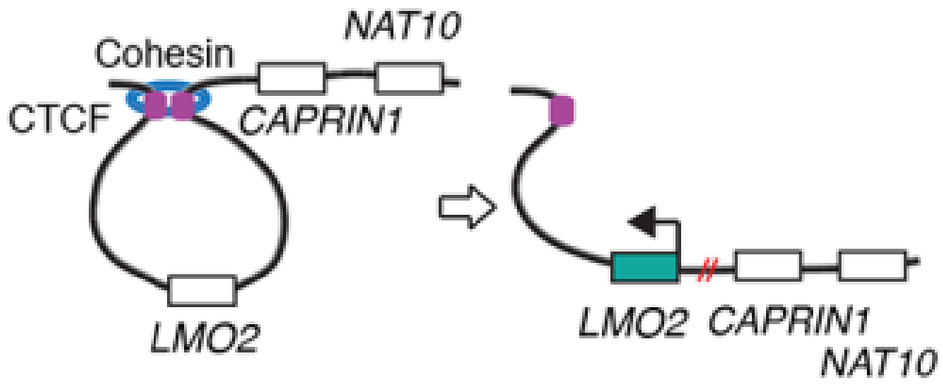
TADs are functional units

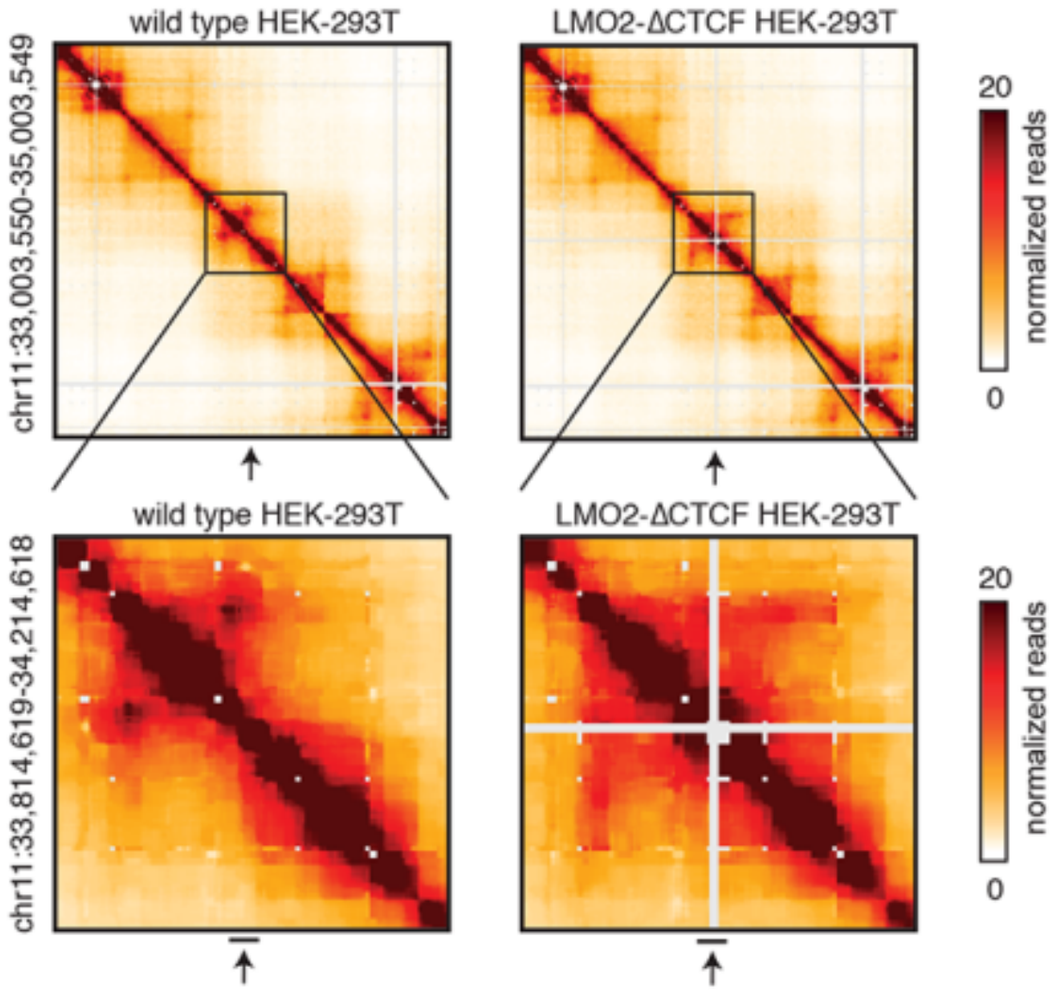
Lupiáñez, et al. (2015). Cell, 1–15.



TADs are functional units

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

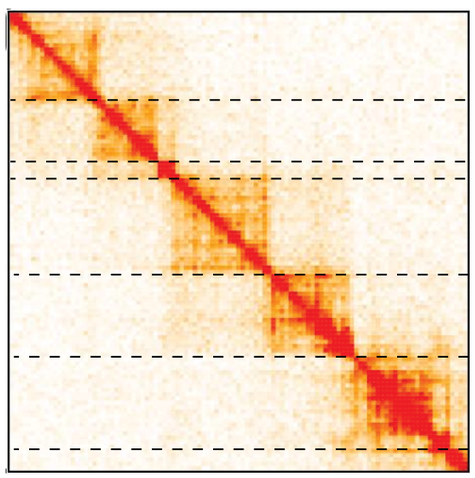


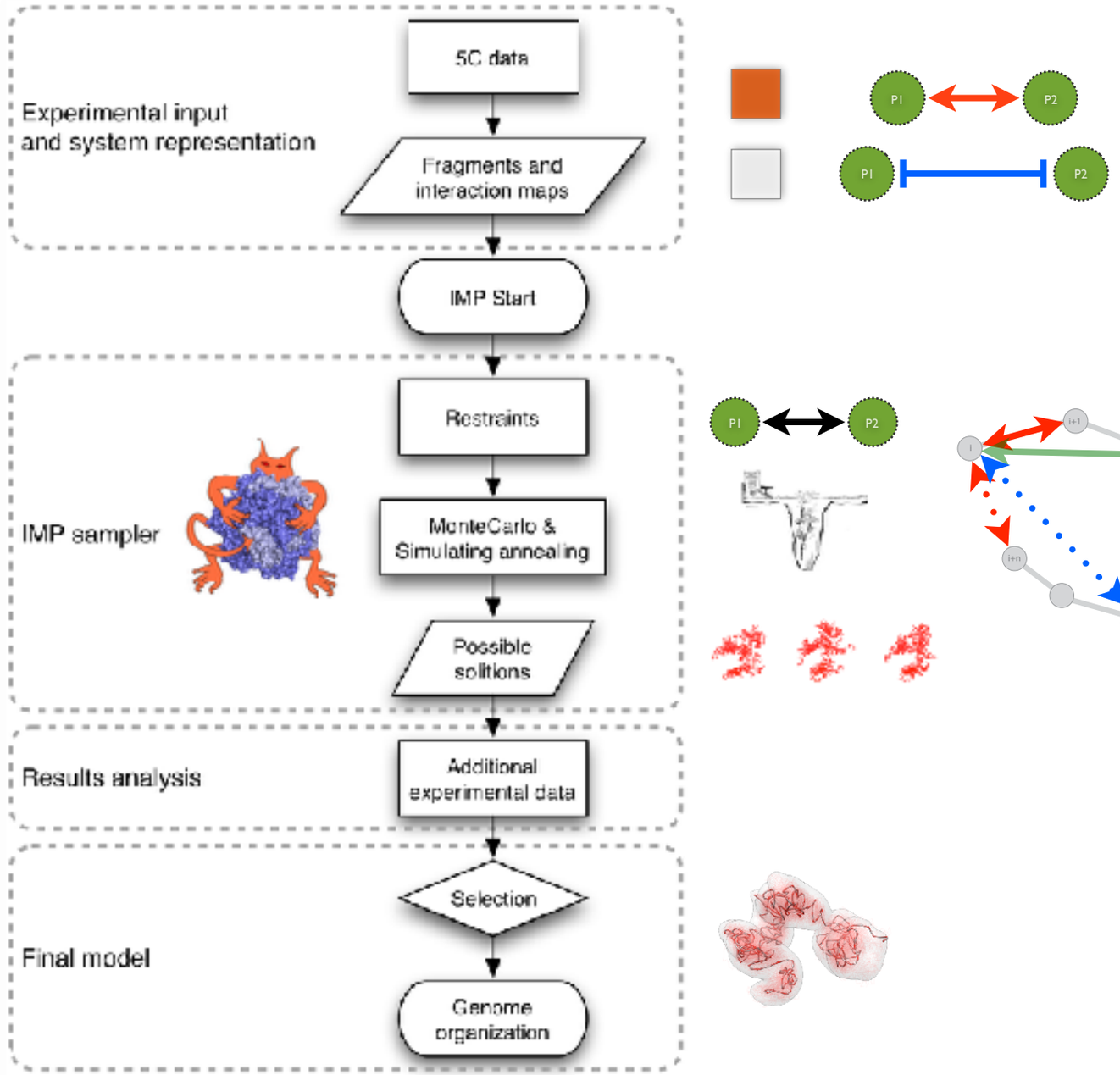



Got normalized
Hi-C maps?

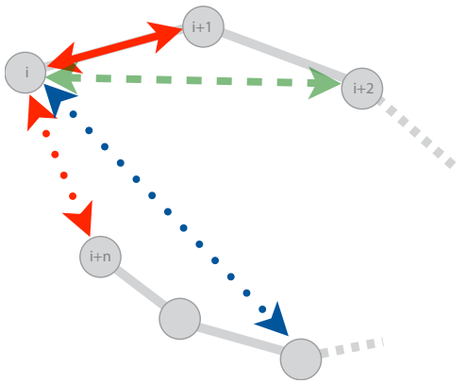
TADbit

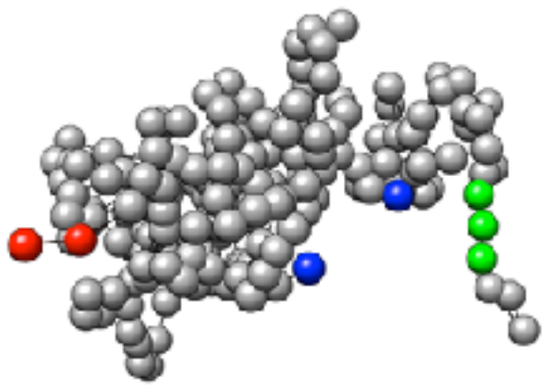
http://3DGenomes.org
http://www.integrativemodeling.org





Model representation and scoring





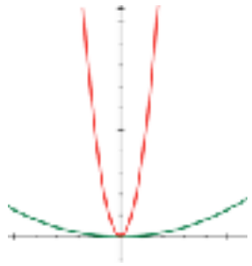
d = d0

d < d0

d > d0

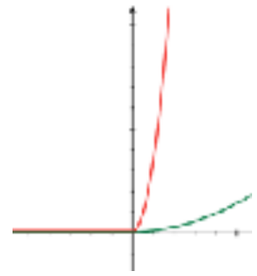
Harmonic

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



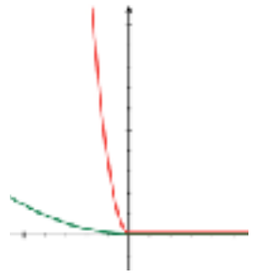
Harmonic Upper Bound

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

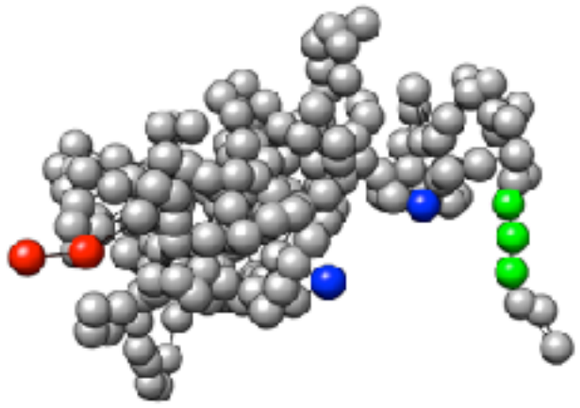


Harmonic Lower Bound

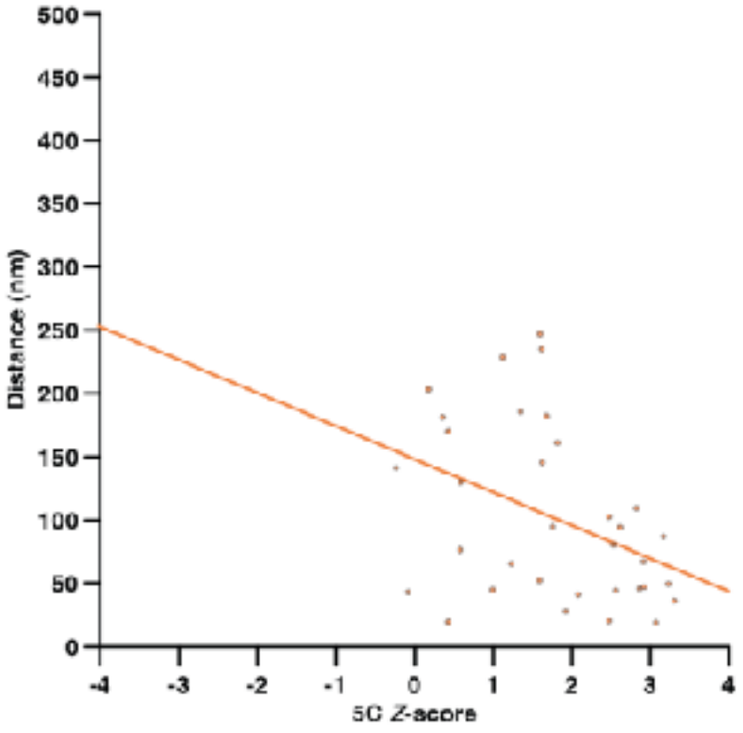
$$\begin{cases} if\ \ d_{i,j} \leq d^0_{i,j}; & lbH_{i,j} = k\left(d_{i,j} - d^0_{i,j}\right)^2 \\ if\ \ d_{i,j} > d^0_{i,j}; & lbH_{i,j} = 0 \end{cases}$$



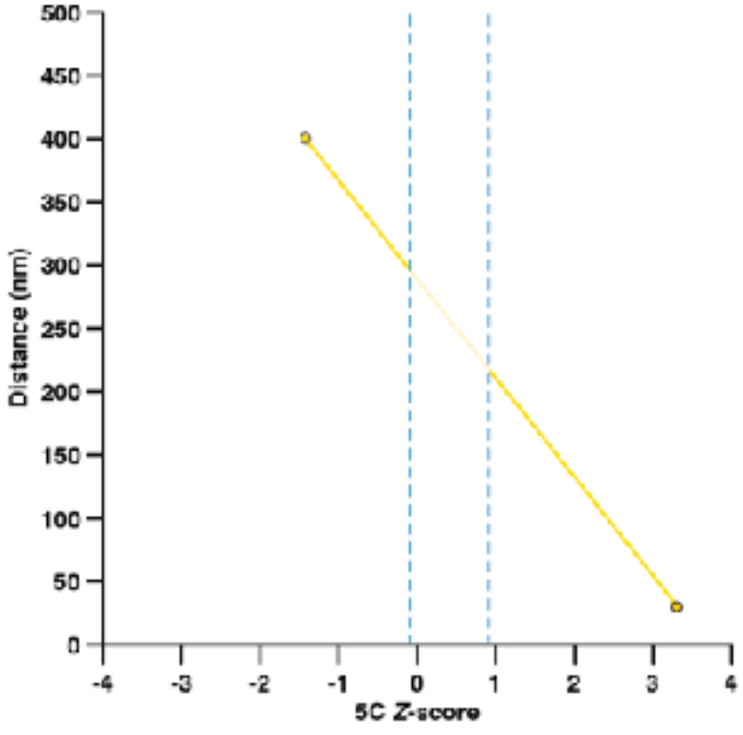
From 3C data to spatial distances



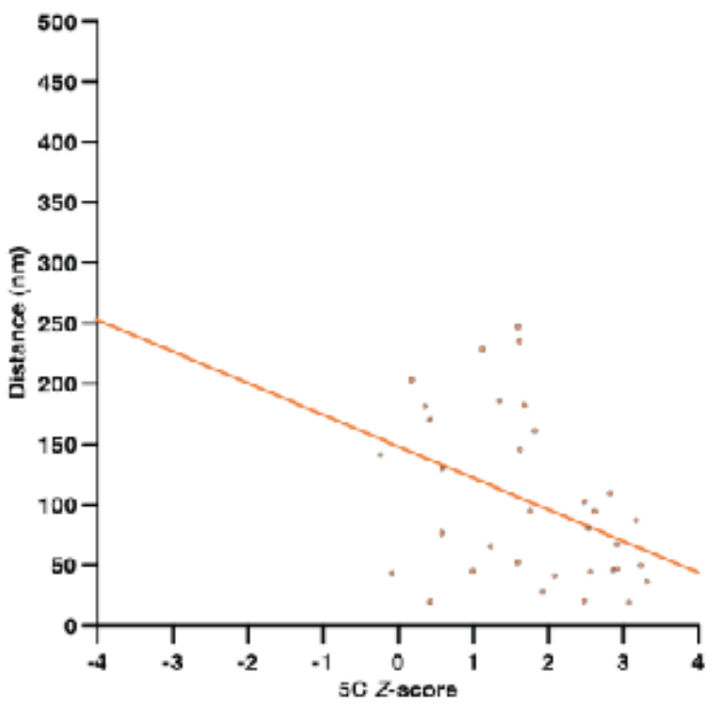
Neighbor fragments

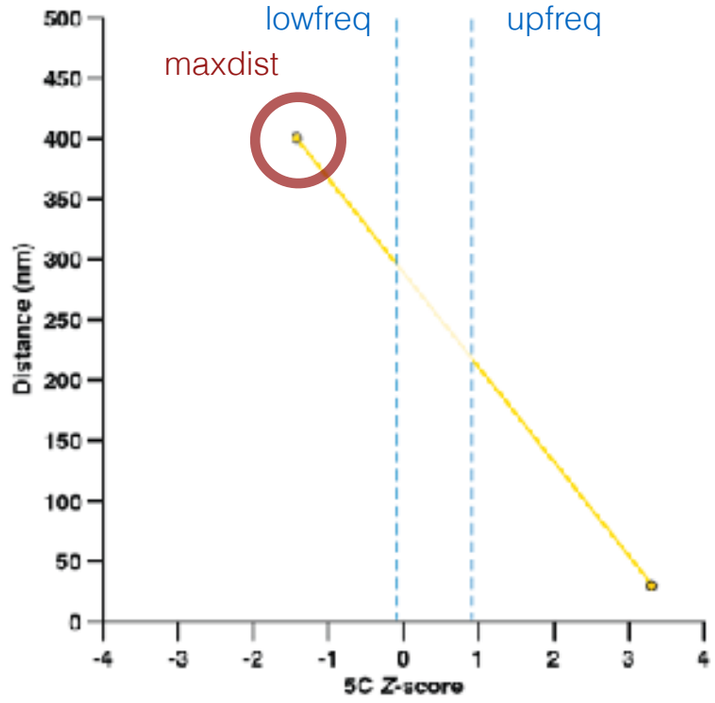


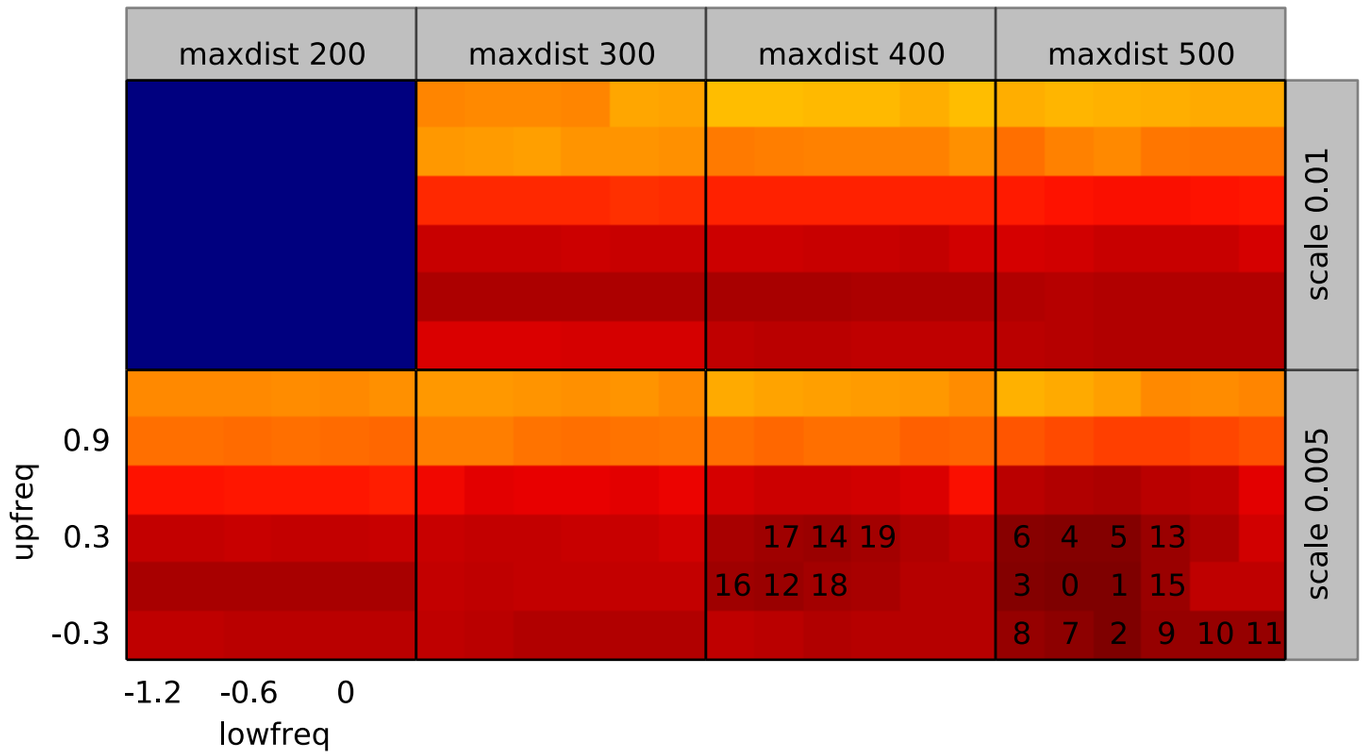
Non-Neighbor fragments



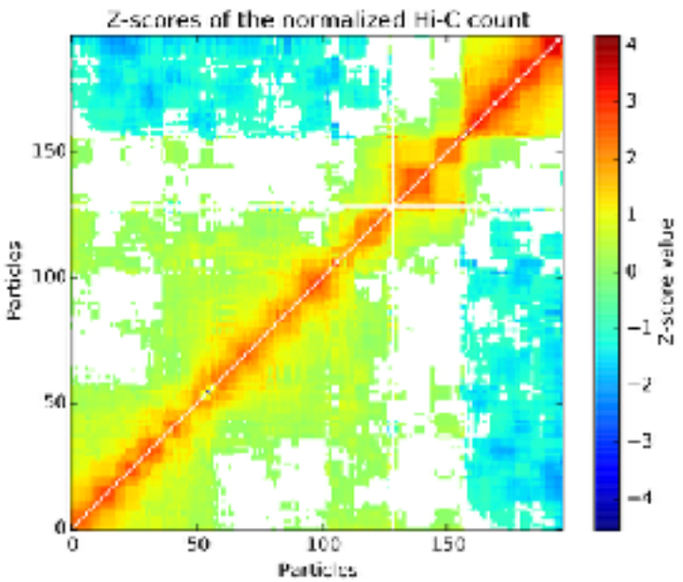
Parameter optimization

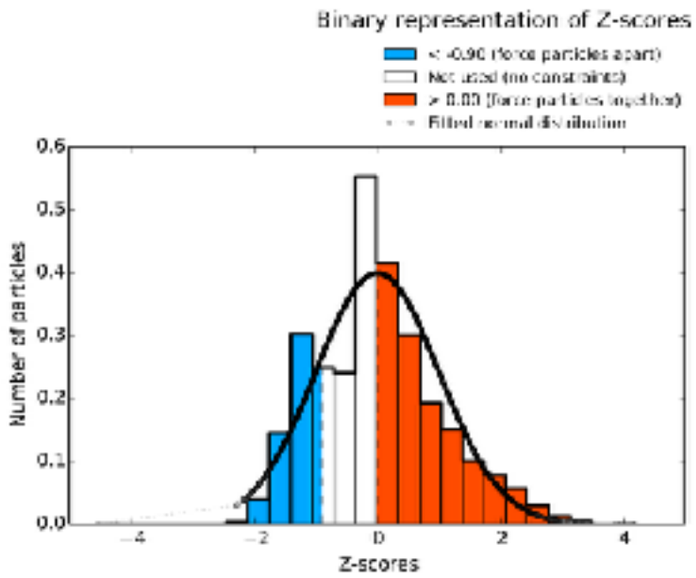


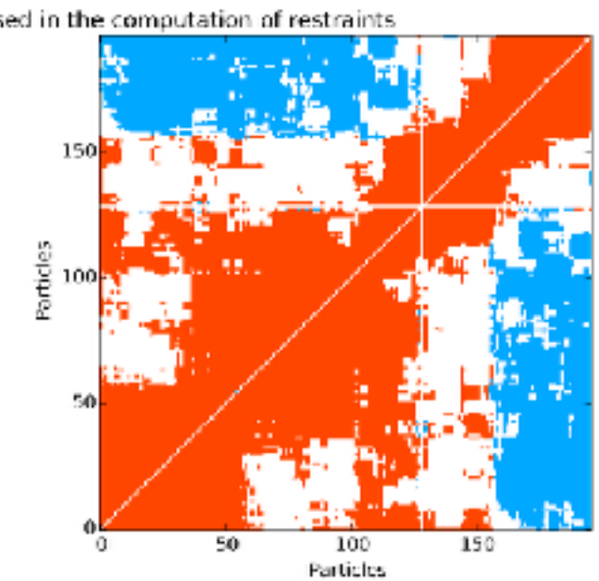


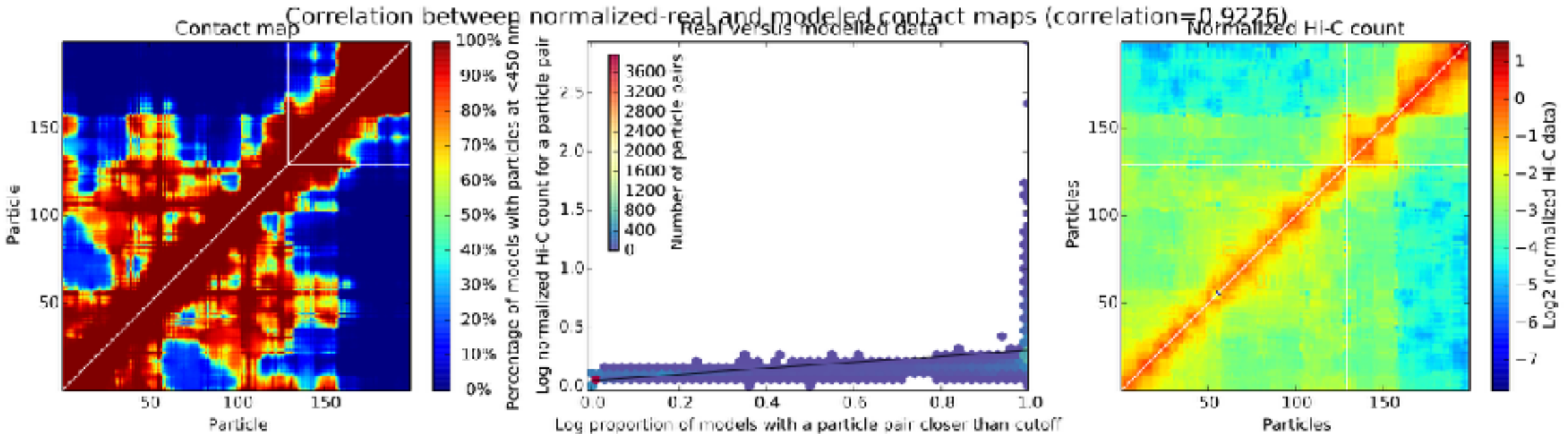


Parameter optimization

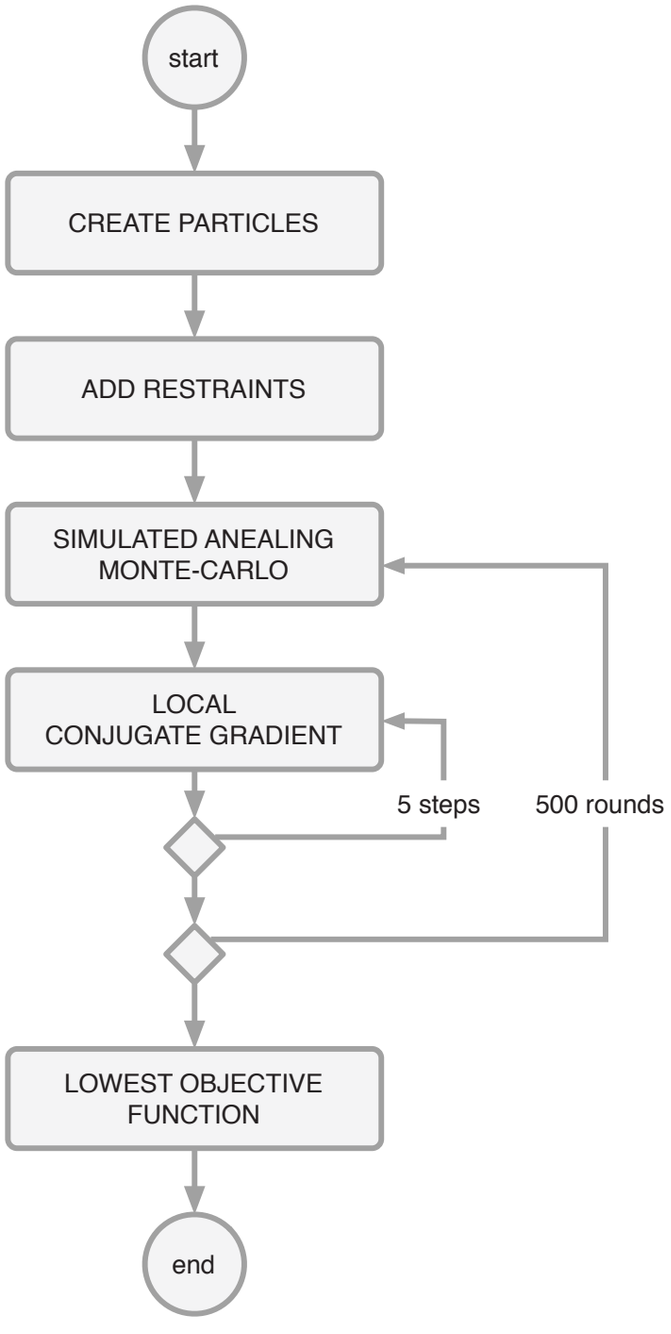


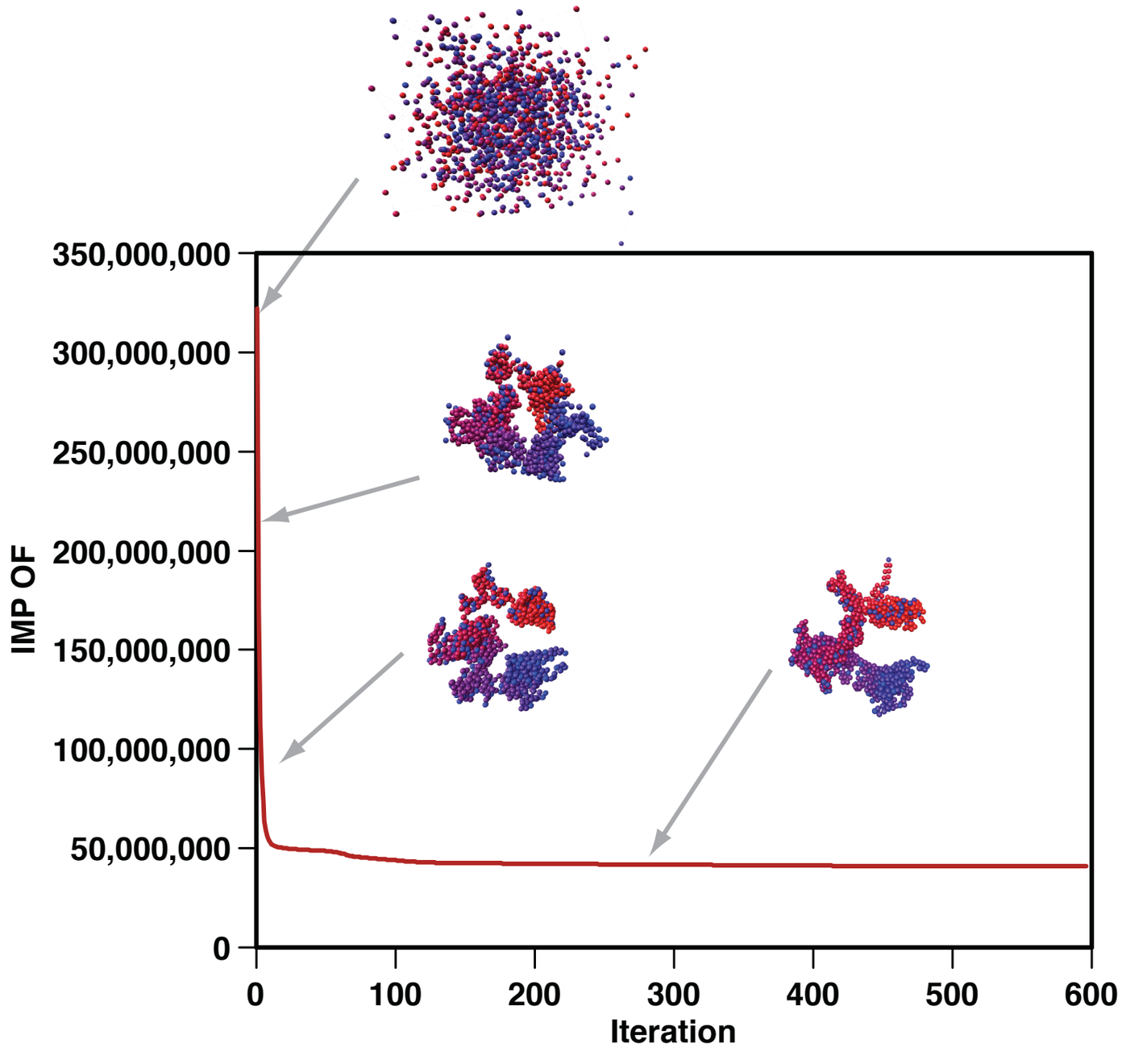




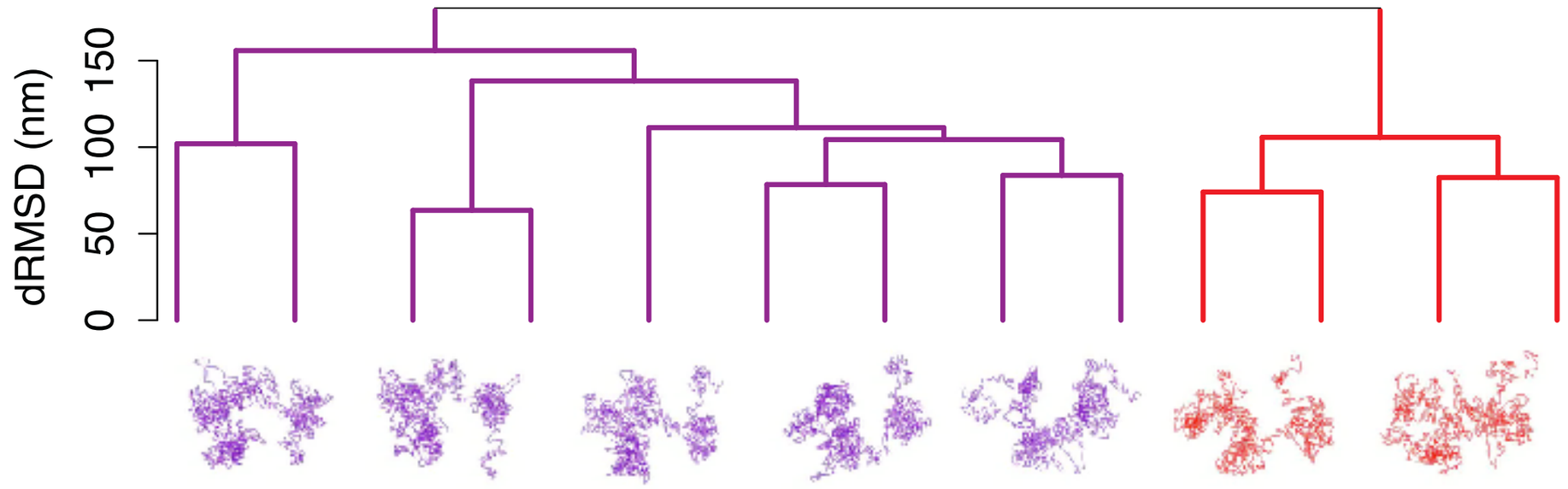


Optimization of the scoring function

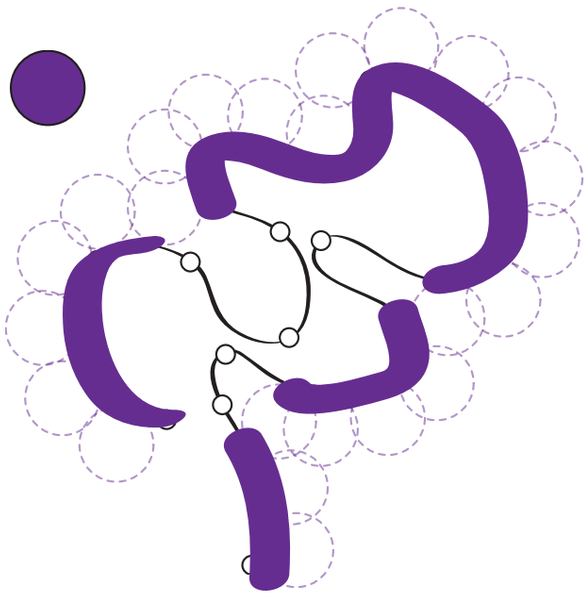




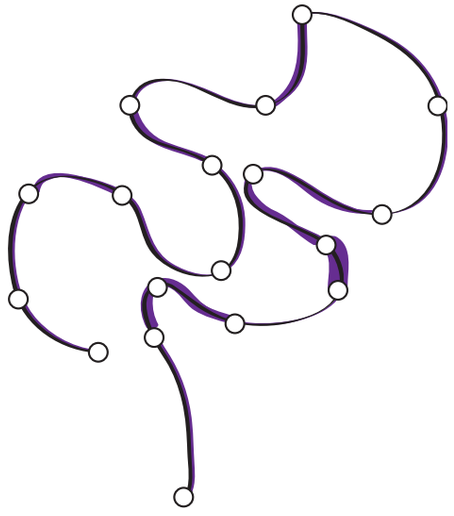
Model analysis: clustering and structural features



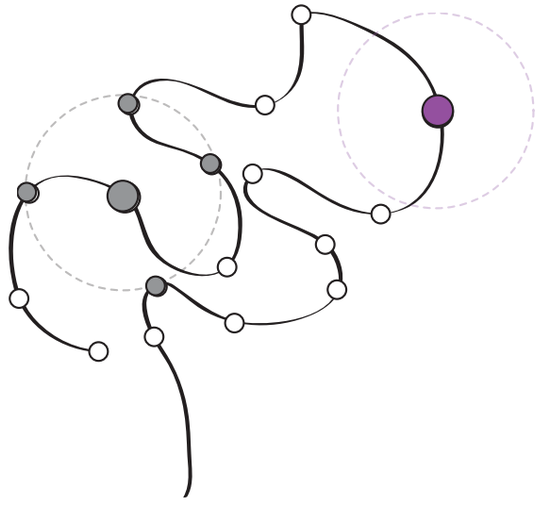
Accessibility (%)



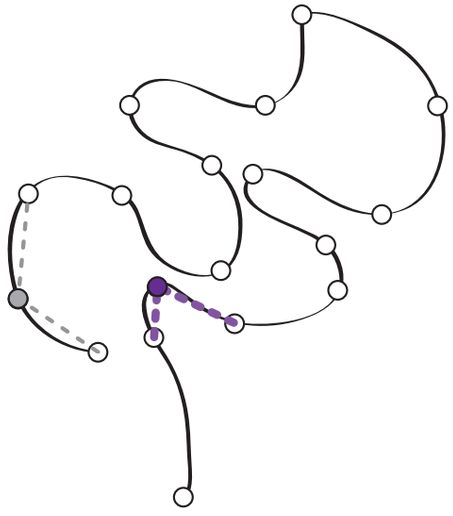
Density (bp/nm)

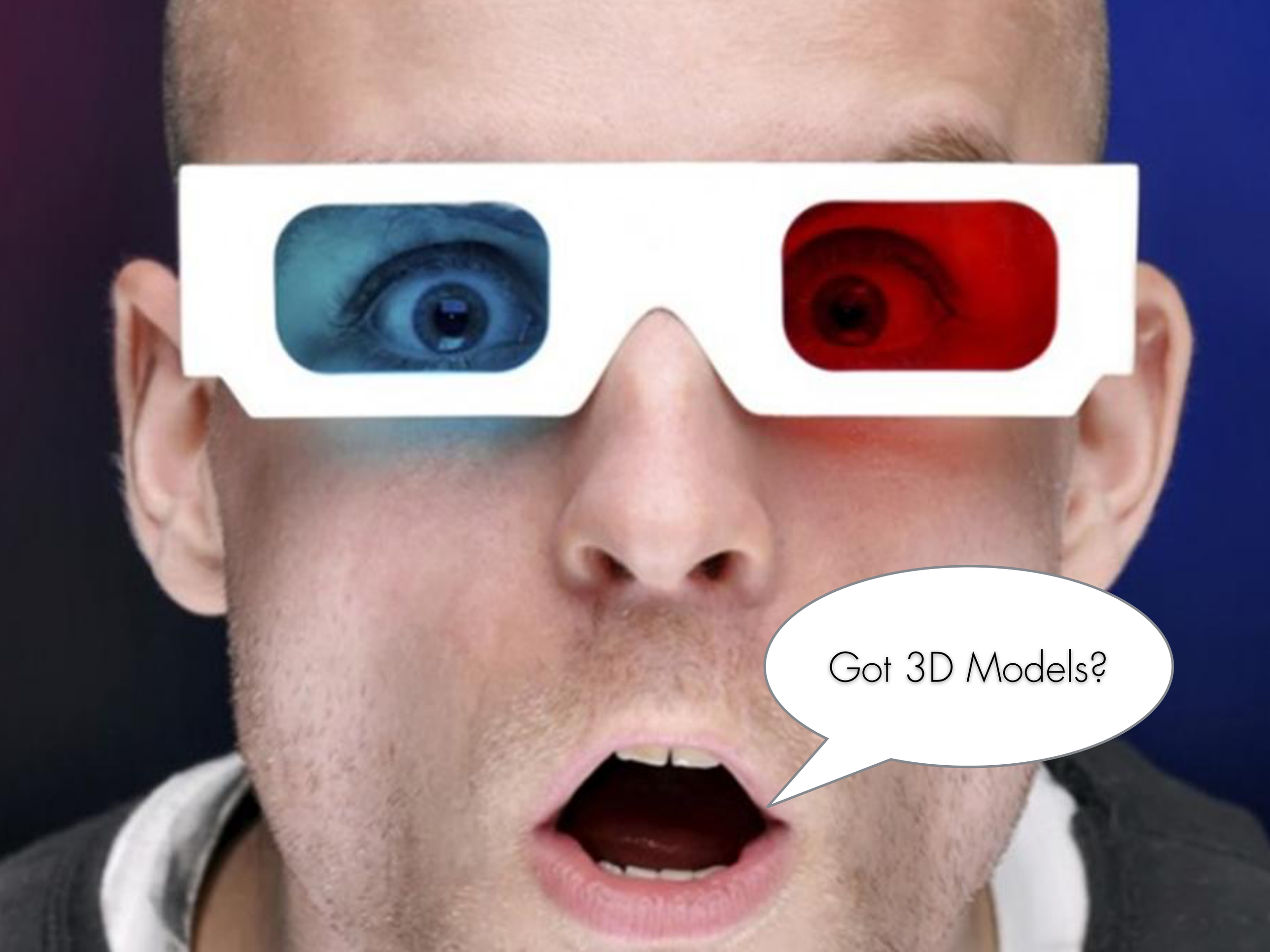


Interactions

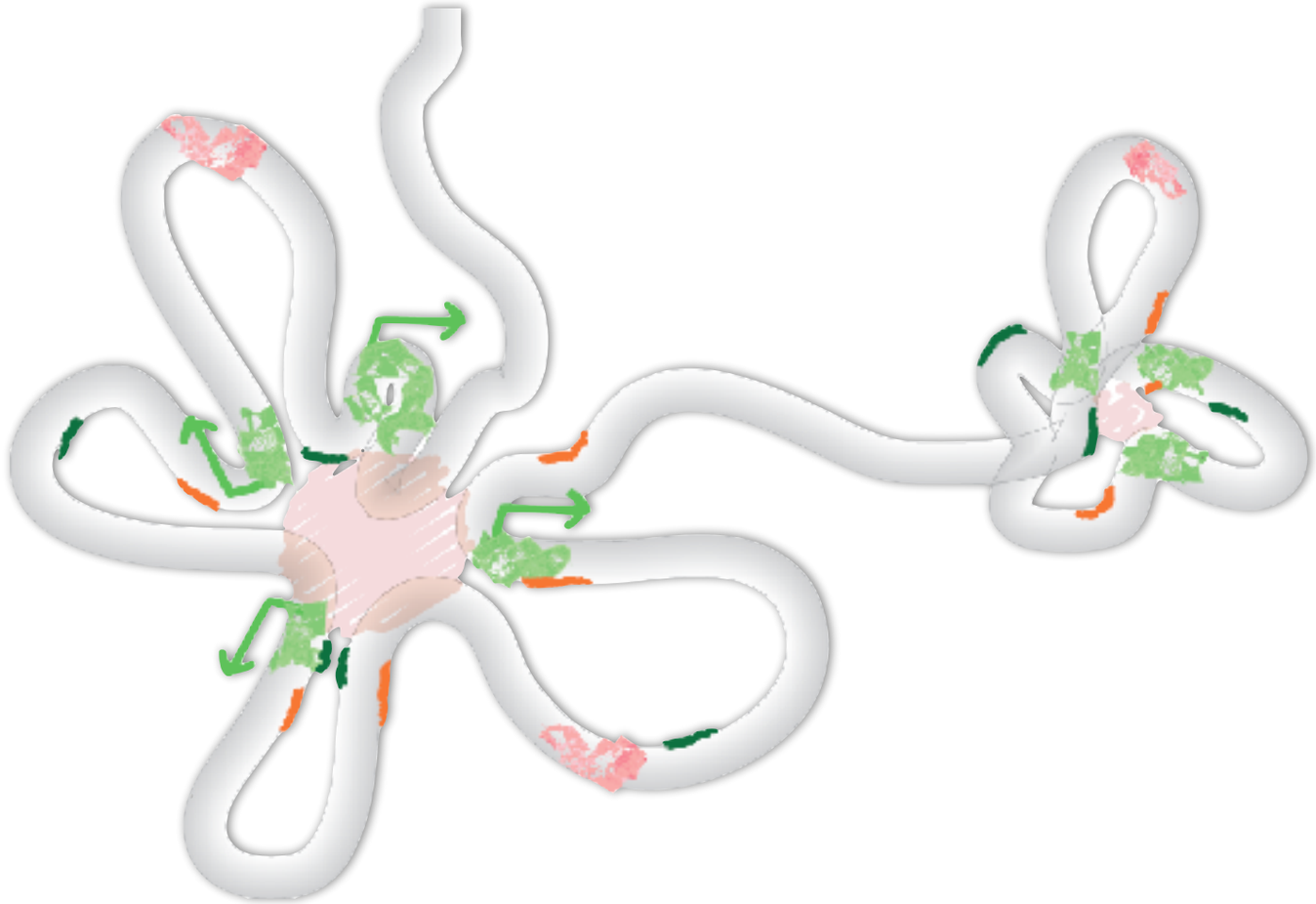


Angle




Got 3D Models?

Human α-globin domain



Davide Baù

Bryan R Lajoie

Amartya Sanyal

Meg Byron

UMASS

Job Dekker

Program in Systems Biology
Department of Biochemistry and Molecular Pharmacology
University of Massachusetts Medical School
Worcester, MA, USA

cnag CRG

Human α-globin domain

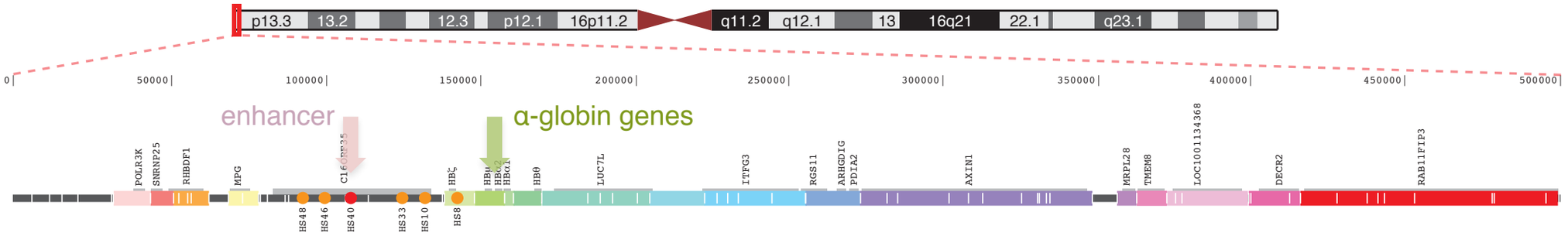
ENm008 genomic structure and environment

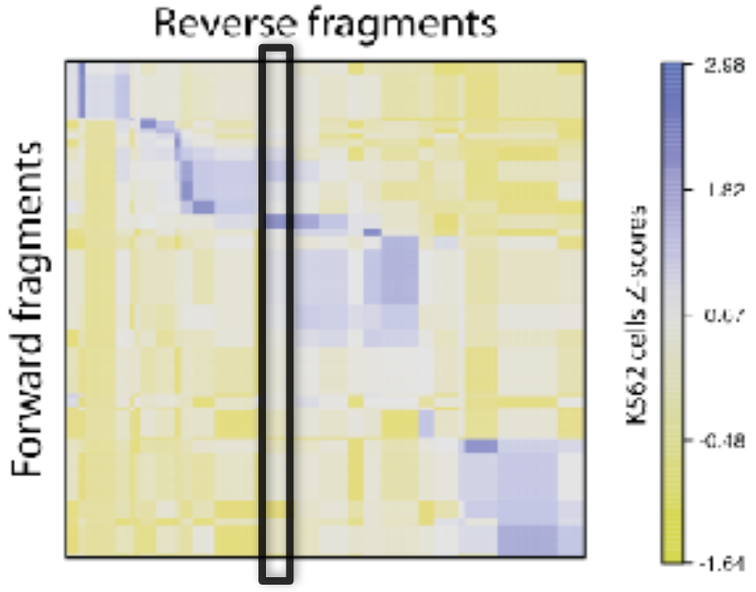
The ENCODE data for ENm008 region was obtained from the UCSC Genome Browser tracks for: RefSeq annotated genes, Affymetrix/CSHL expression data (Gingeras Group at Cold Spring Harbor), Duke/NHGRI DNaseI Hypersensitivity data (Crawford Group at Duke University), and Histone Modifications by Broad Institute ChIP-seq (Bernstein Group at Broad Institute of Harvard and MIT).

ENCODE Consortium. Nature (2007) vol. 447 (7146) pp. 799-816

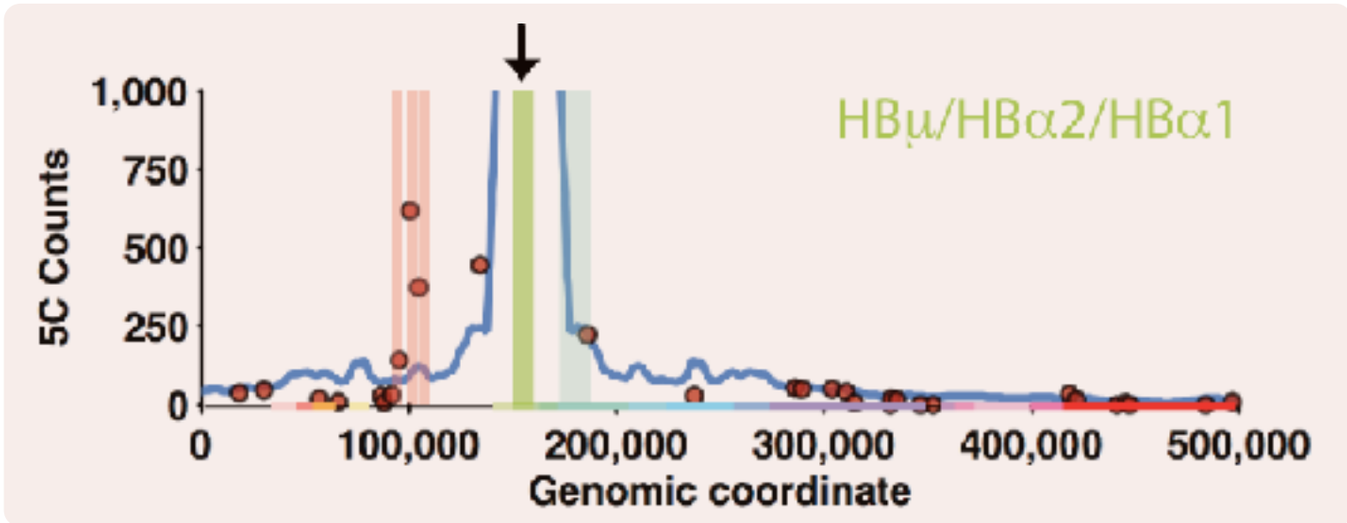
Human α-globin domain

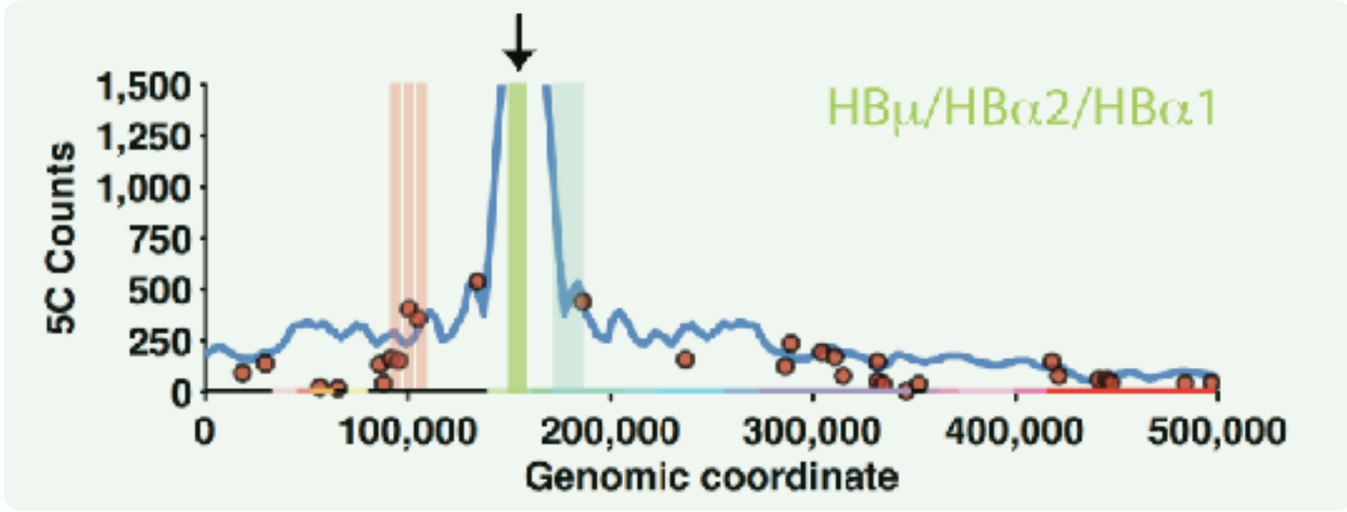
ENm008 genomic structure and environment





K562 cells:
α-globin genes active

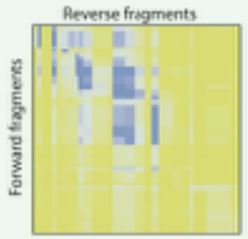


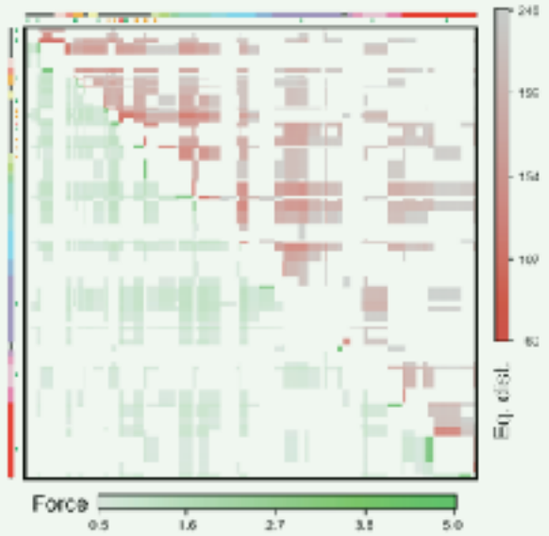


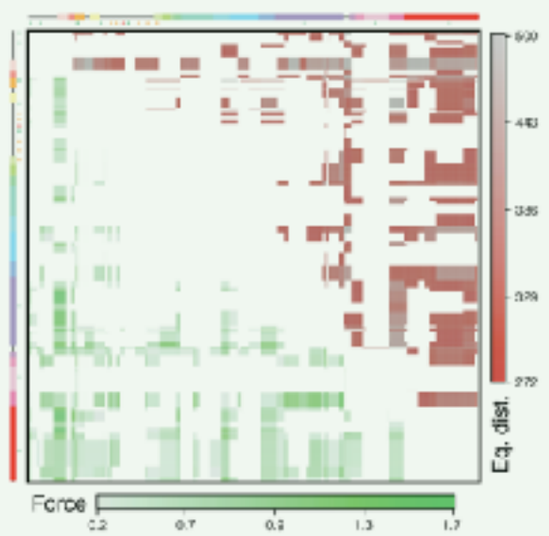
Scoring

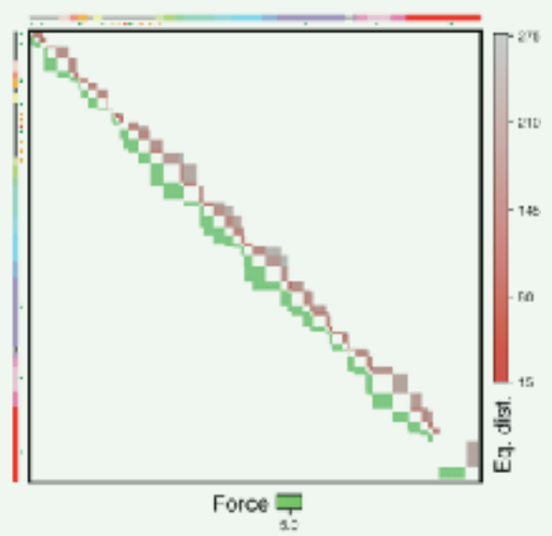
GM12878

70 fragments
1,520 restraints





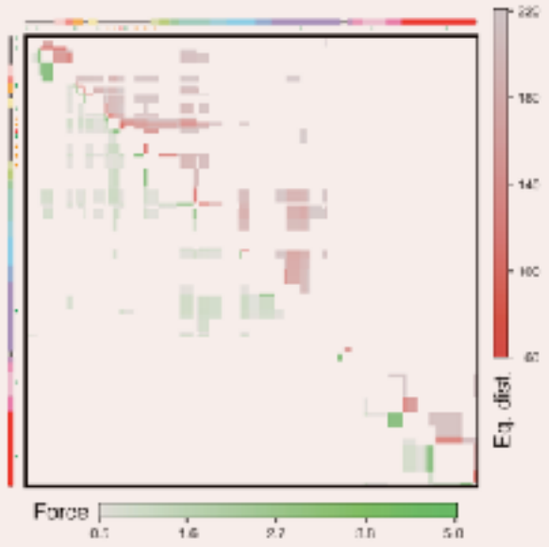


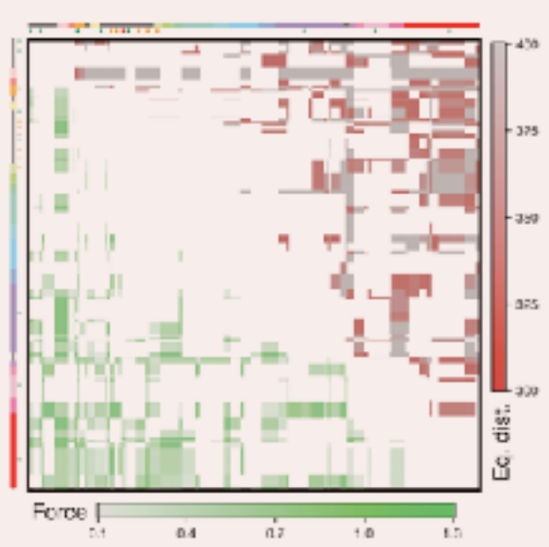


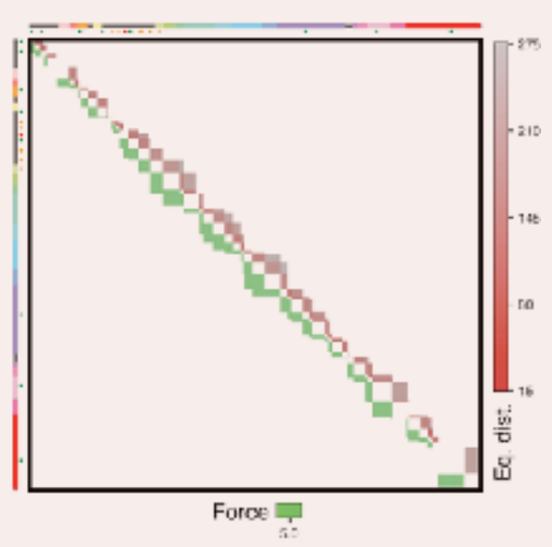
Harmonic | Harmonic Lower Bound | Harmonic Upper Bound

K562

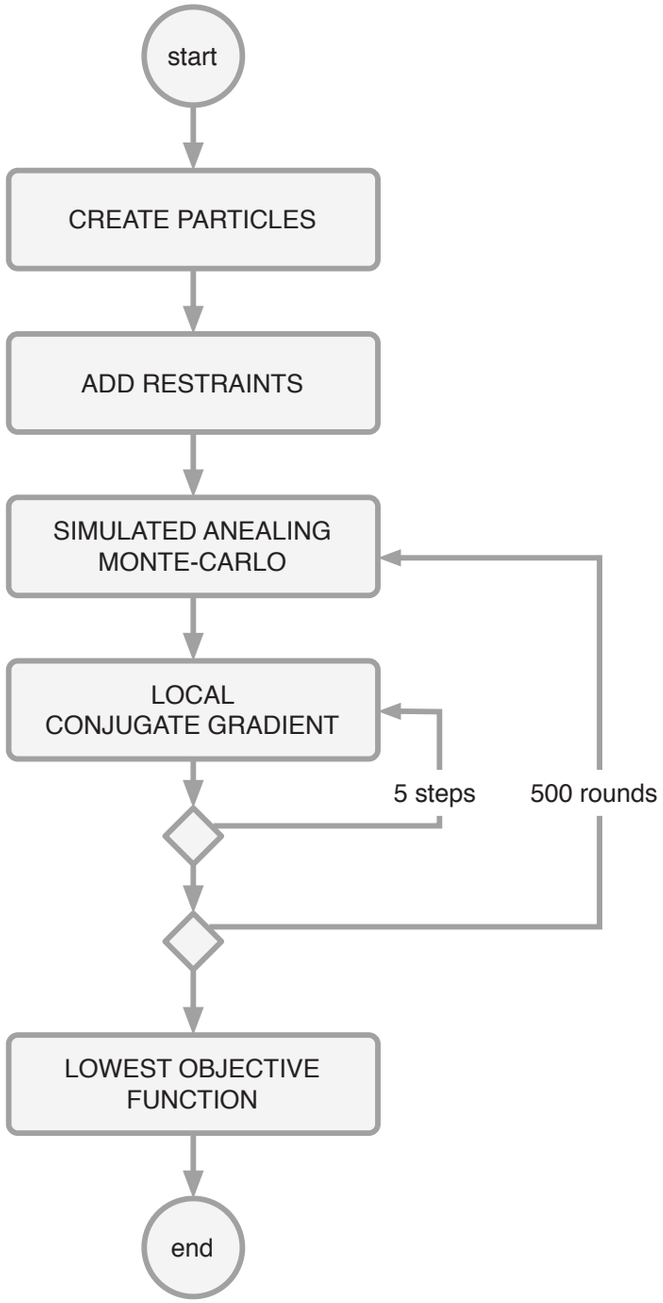
70 fragments
1,049 restraints

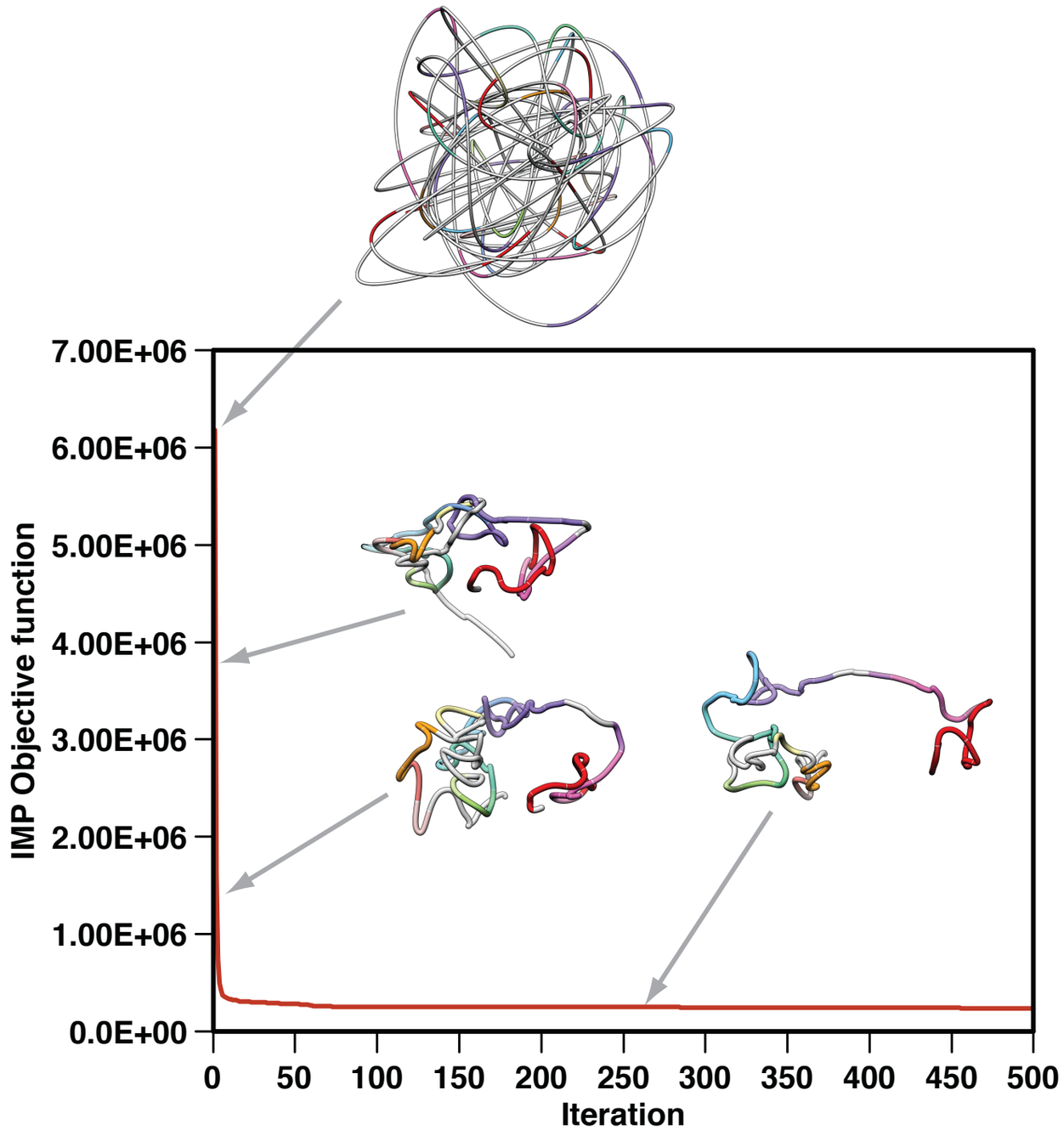




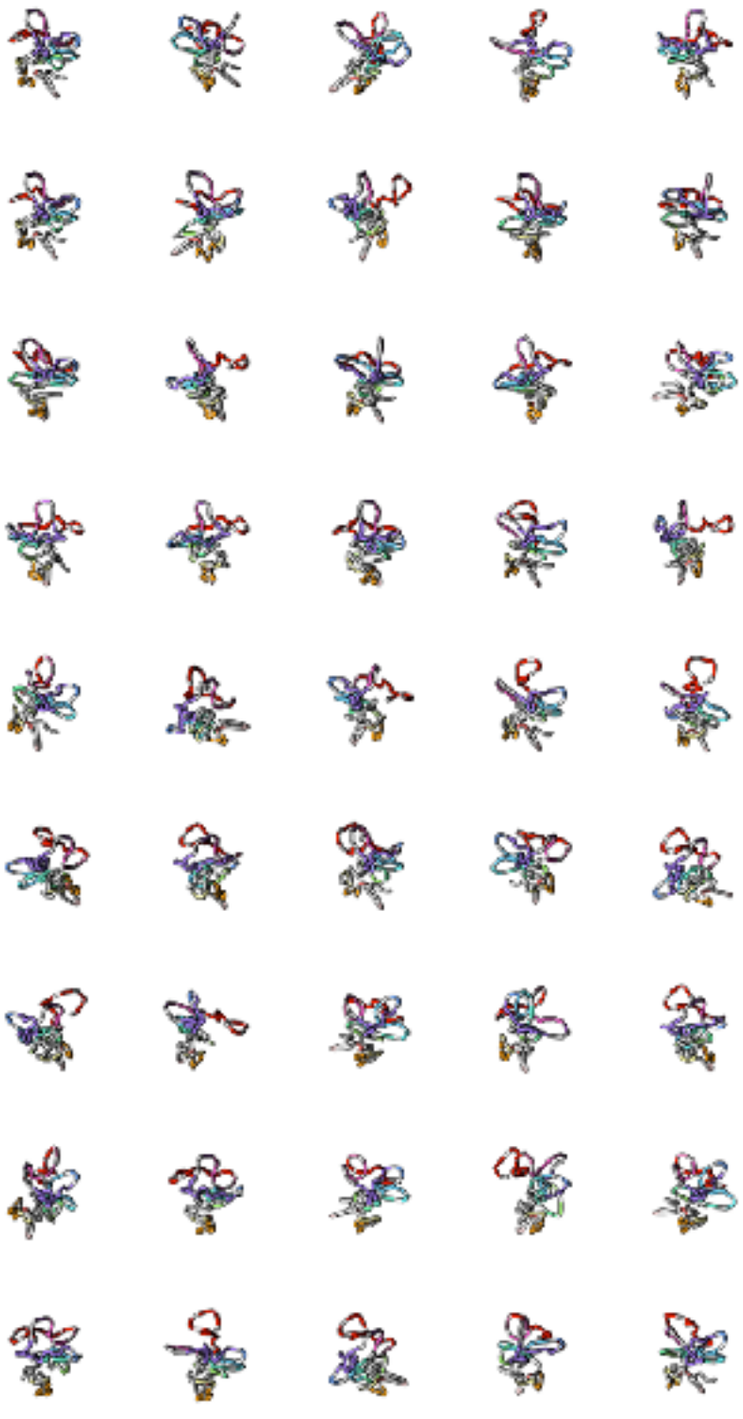


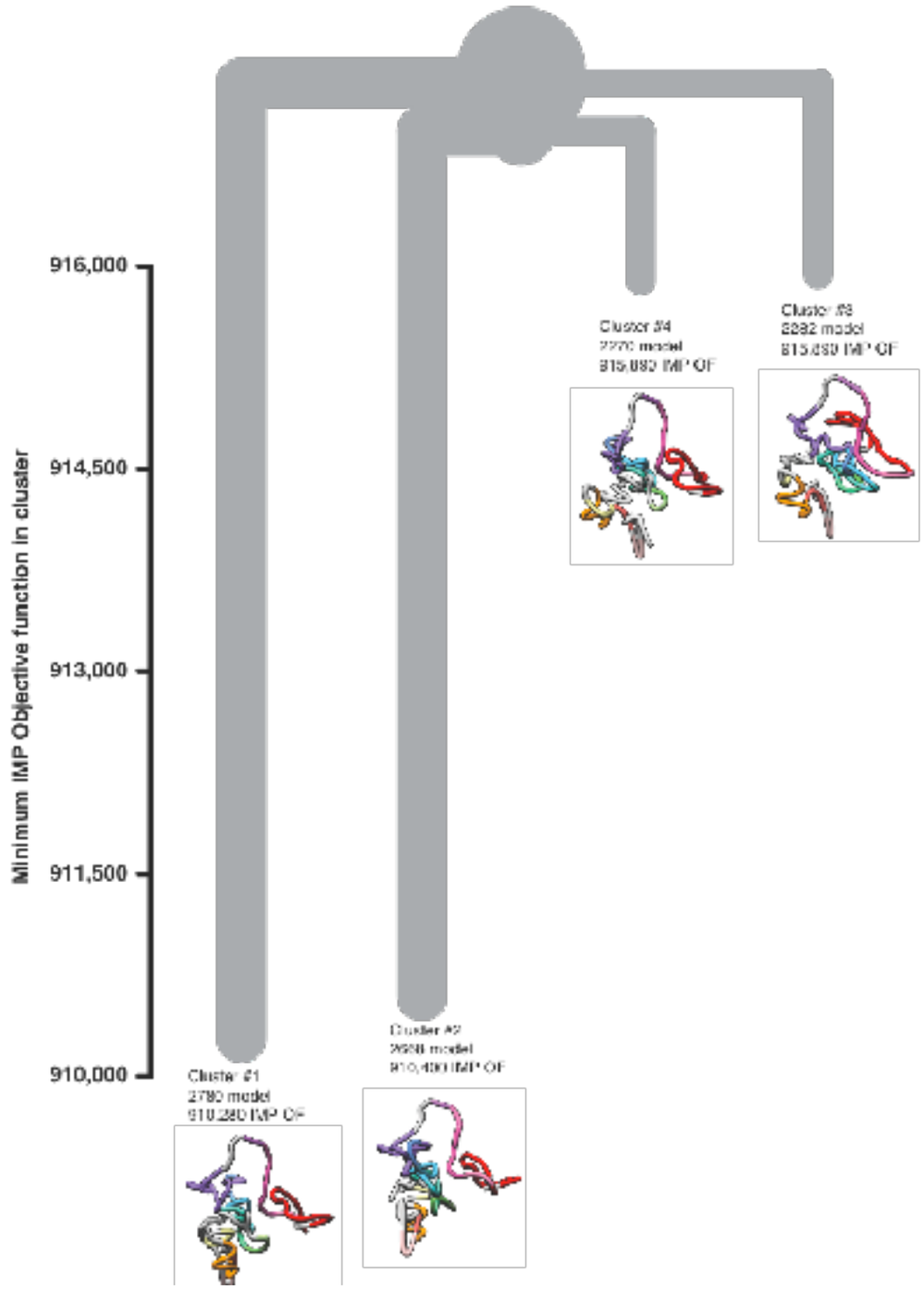
Optimization





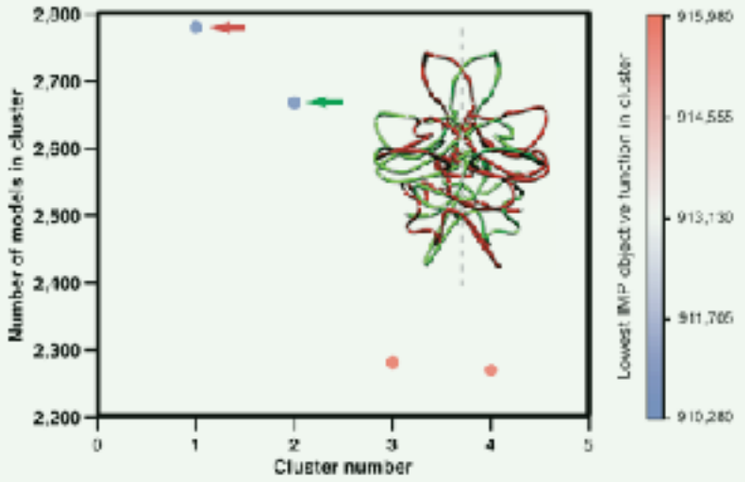
Clustering

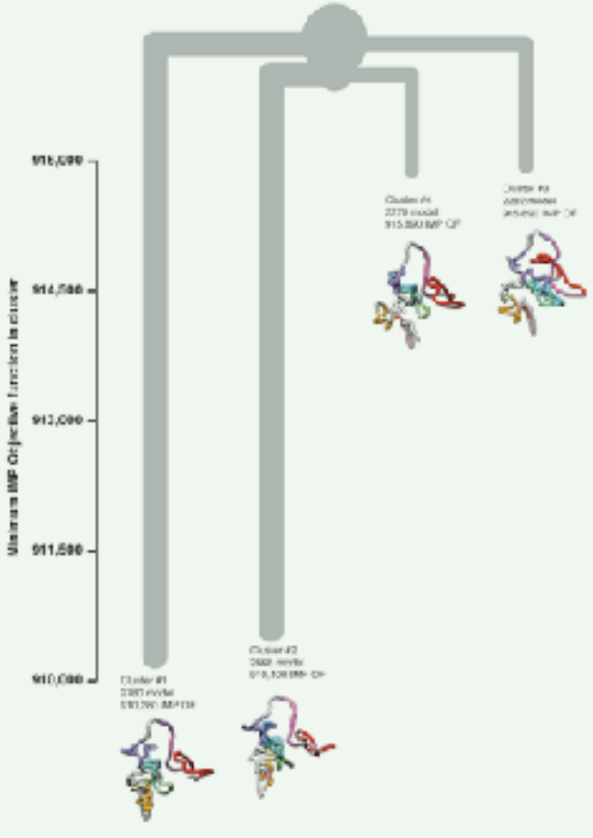



916,000
914,500
913,000
911,500
910,000
Minimum IMP Objective function in cluster
Cluster #4
2770 model
915,082 IMP OF
Cluster #3
2282 model
915,059 IMP OF
Cluster #1
2780 model
910,280 IMP OF
Cluster #2
2658 model
910,480 IMP OF

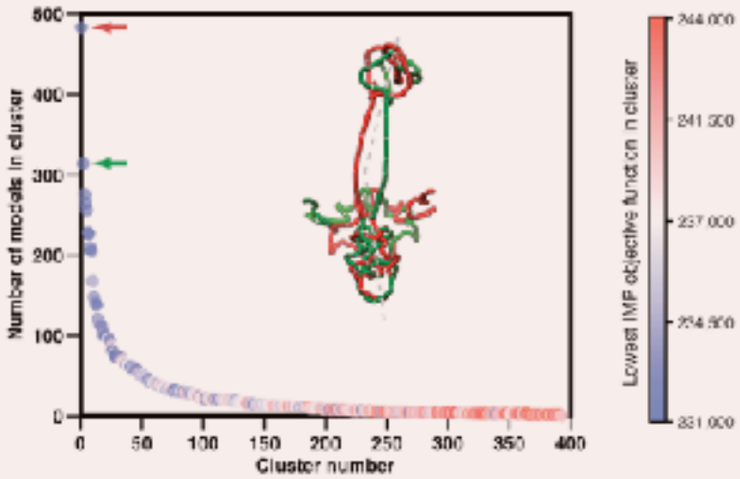
Not just one solution

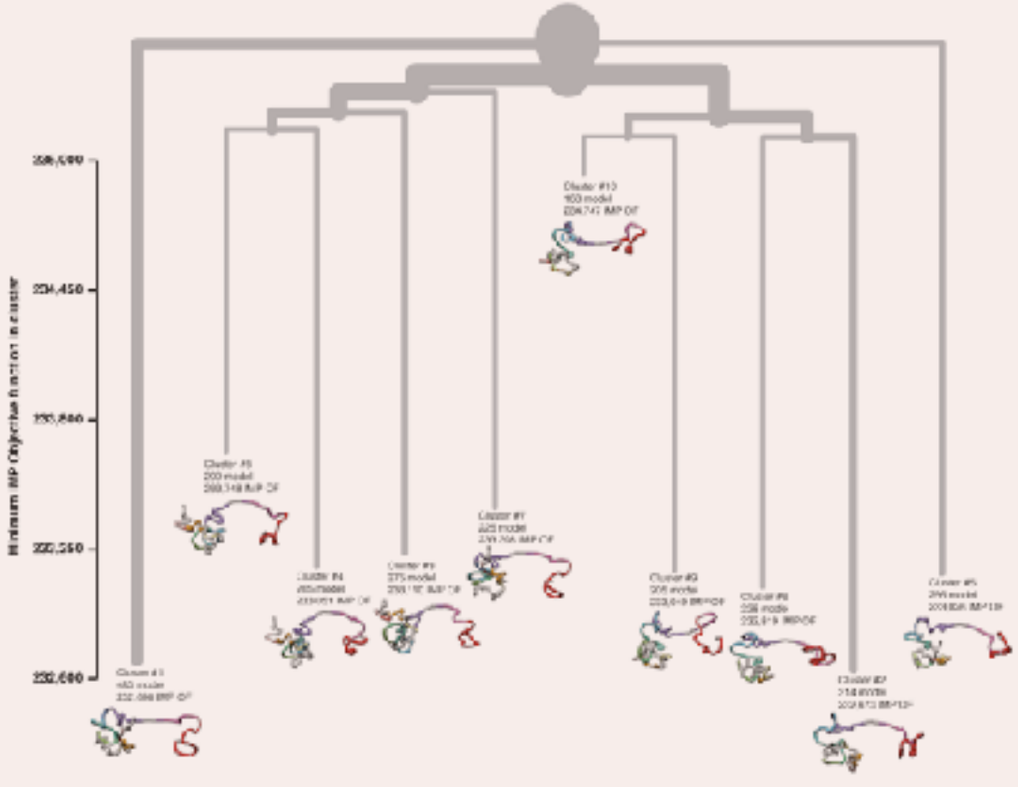
GM12878





K562

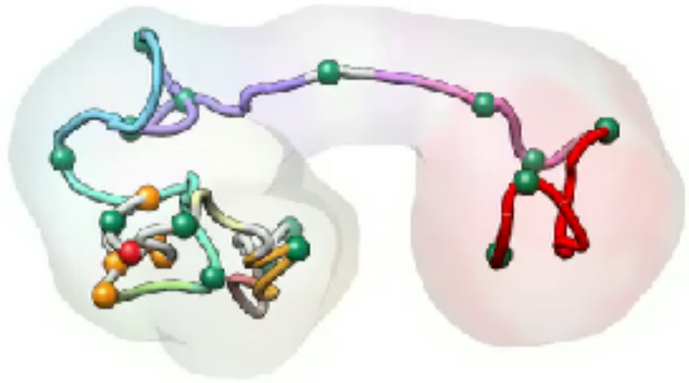




Consistency

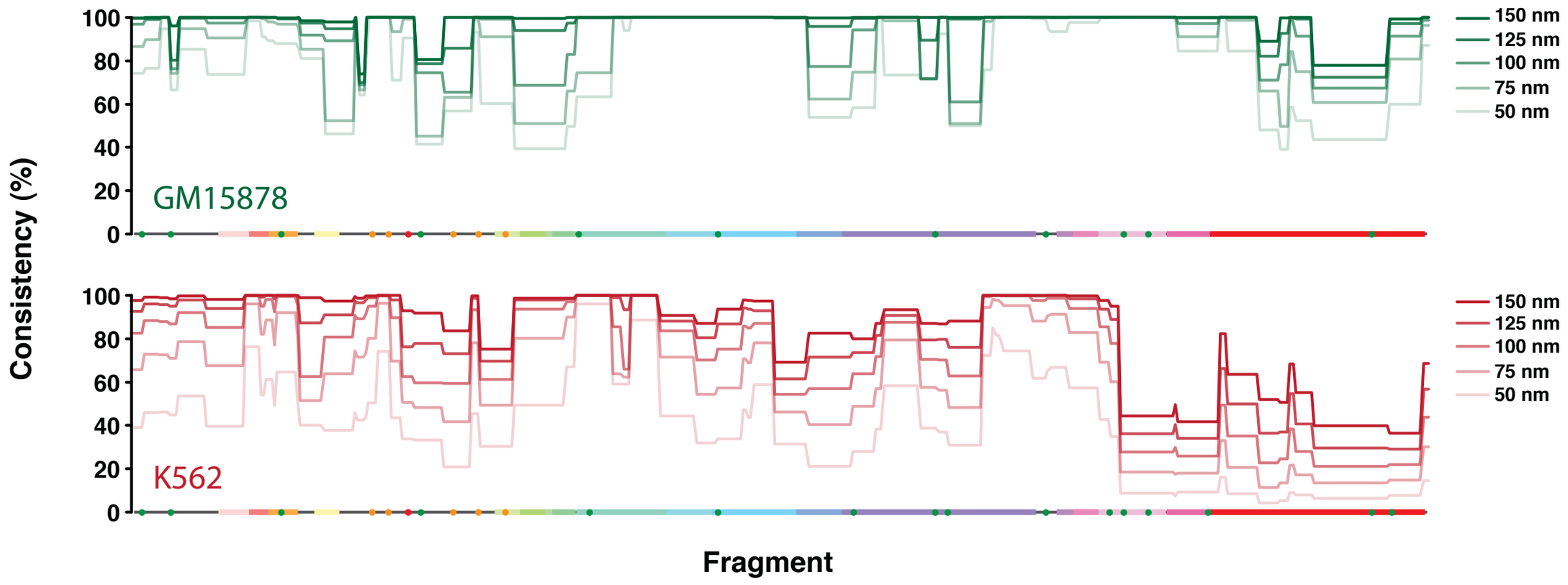
GM12878

Cluster #1
2780 model



K562

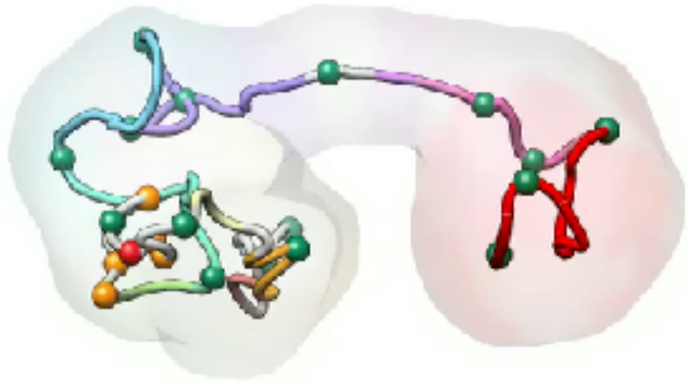
Cluster #2
314 model



Consistency

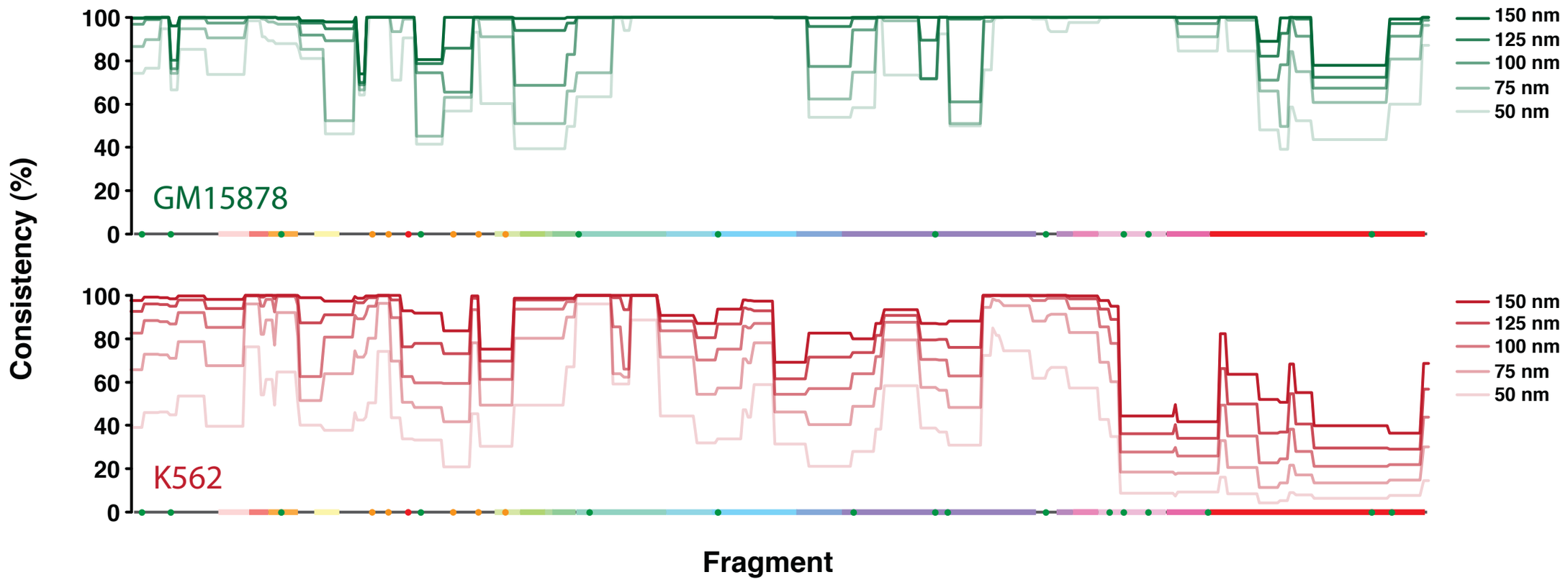
GM12878

Cluster #1
2780 model



K562

Cluster #2
314 model



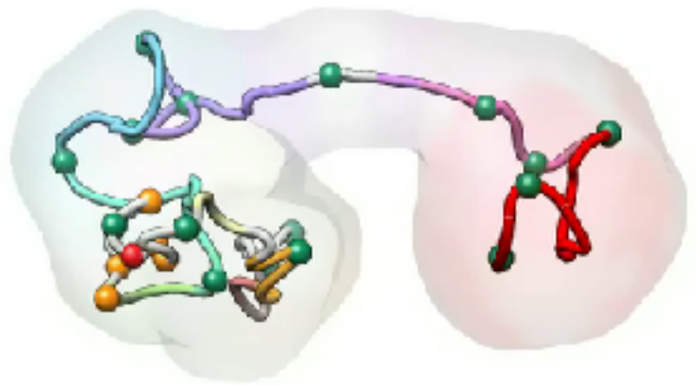
Regulatory elements

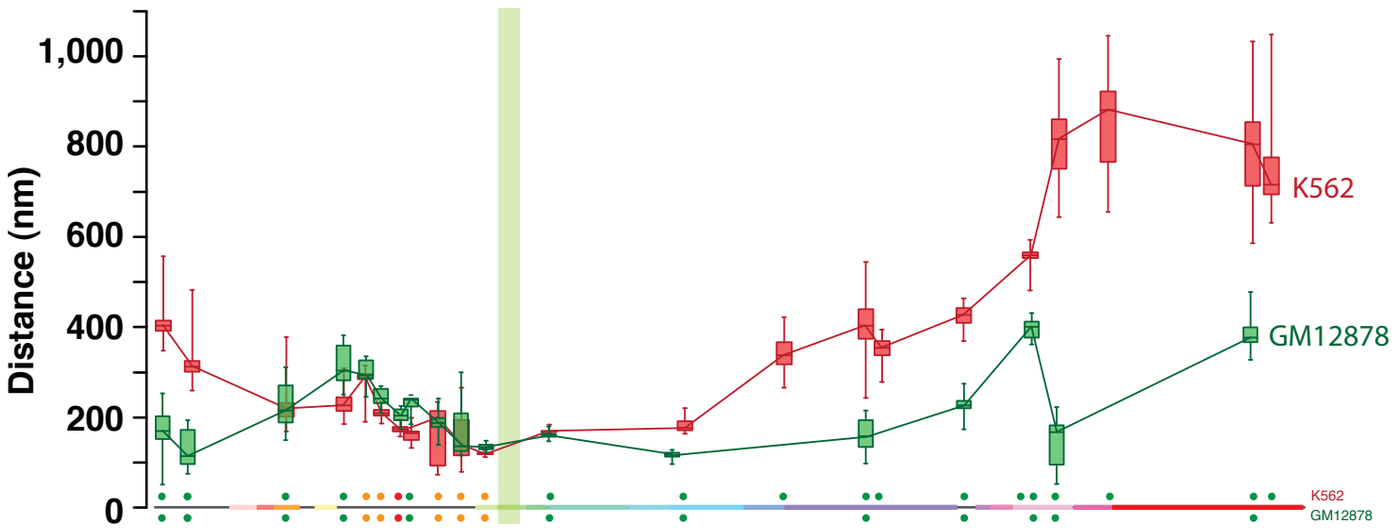
GM12878

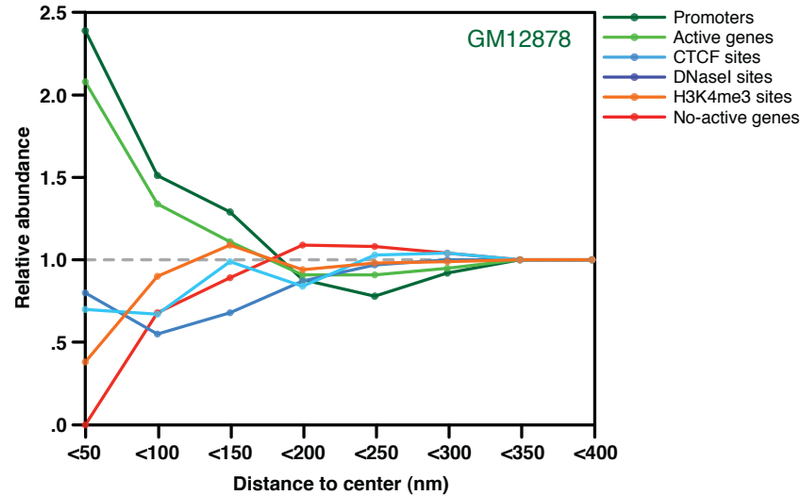
Cluster #1
2780 model

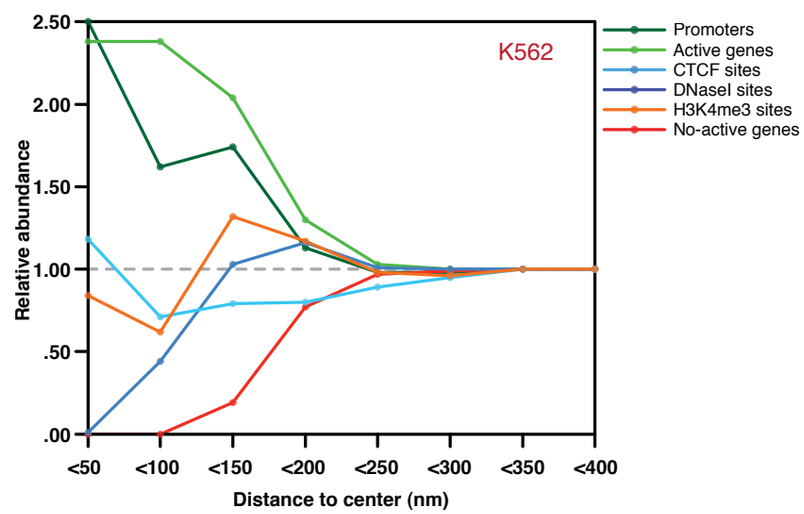
K562

Cluster #2
314 model





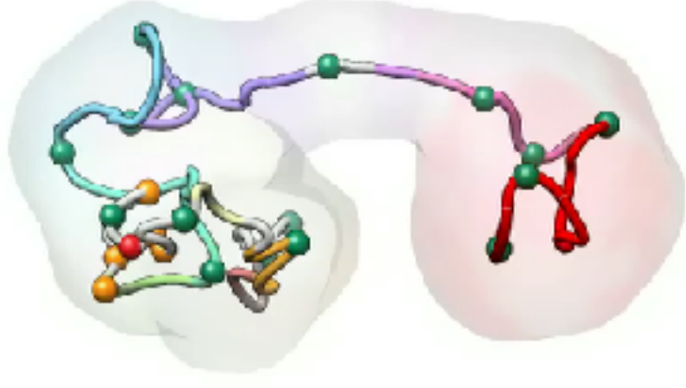




Compactness

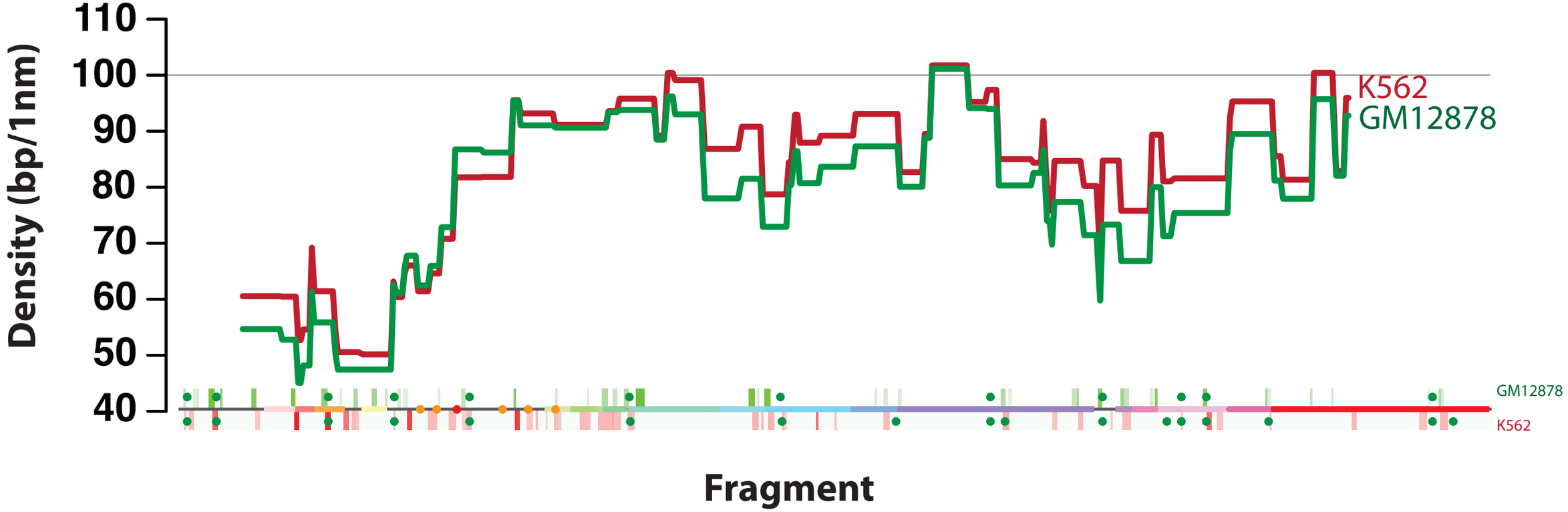
GM12878

Cluster #1
2780 model



K562

Cluster #2
314 model

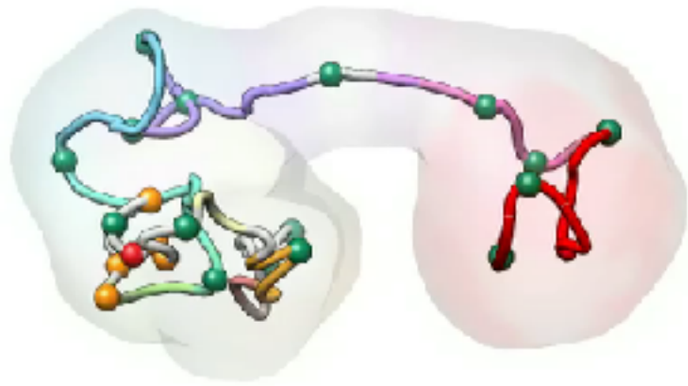


cnag CRG

Multi-loops

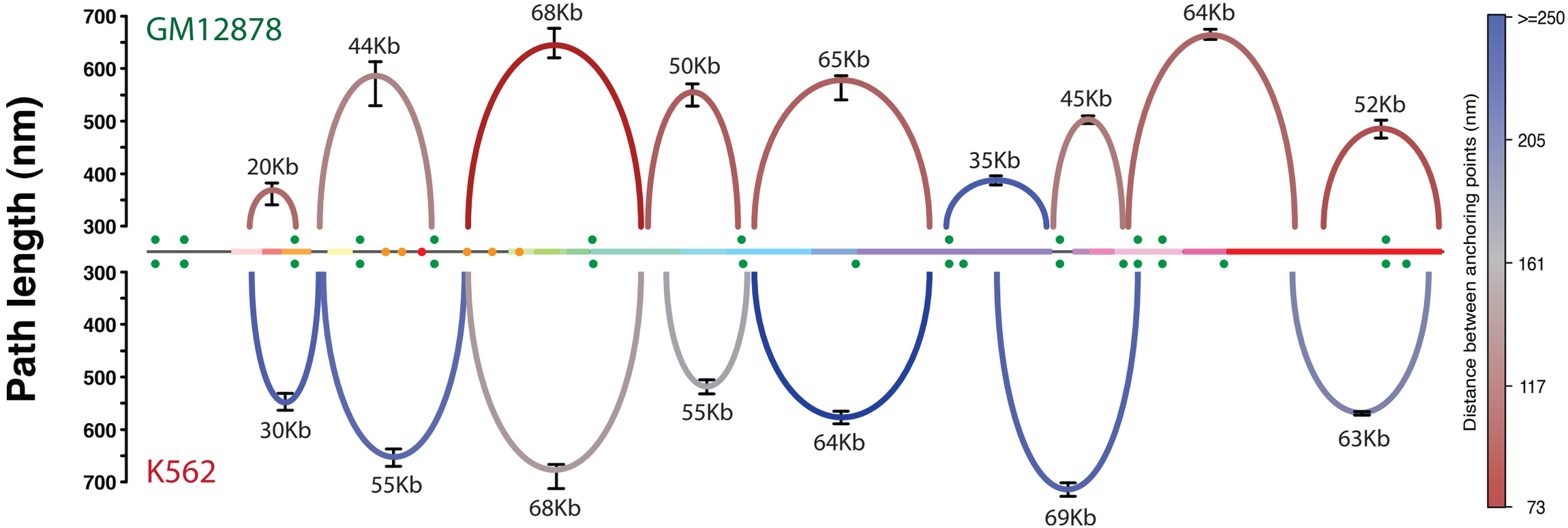
GM12878

Cluster #1
2780 model



K562

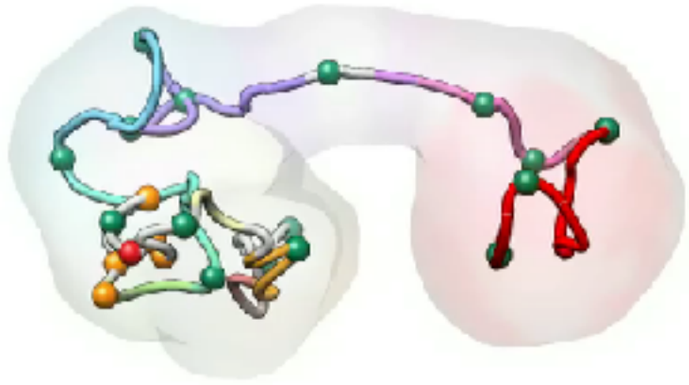
Cluster #2
314 model



Expression

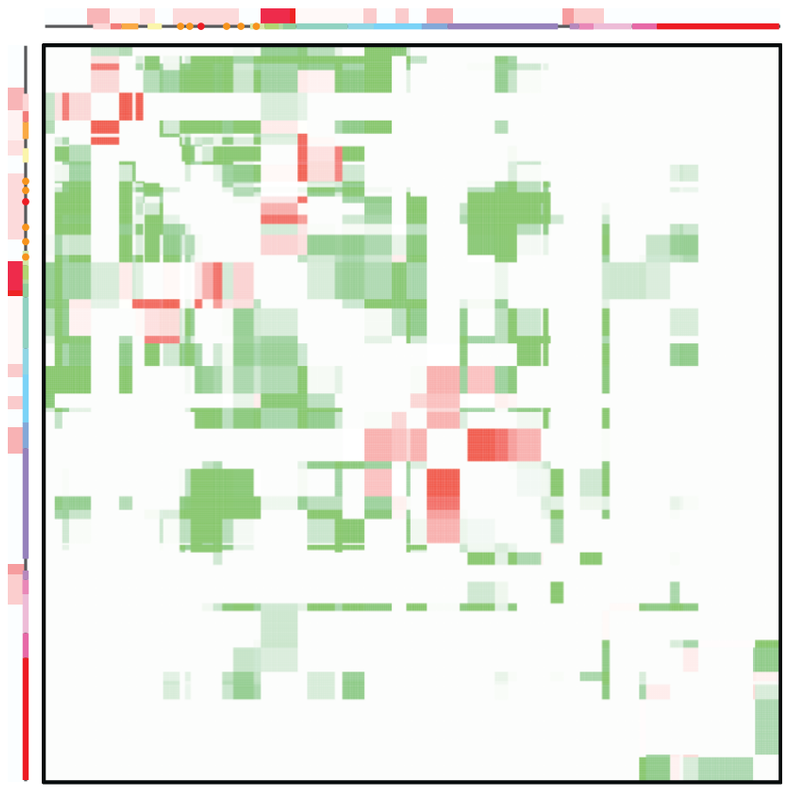
GM12878

Cluster #1
2780 model



K562

Cluster #2
314 model



Increased in GM12878

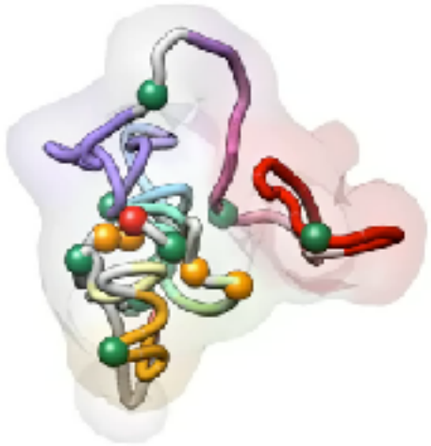
=

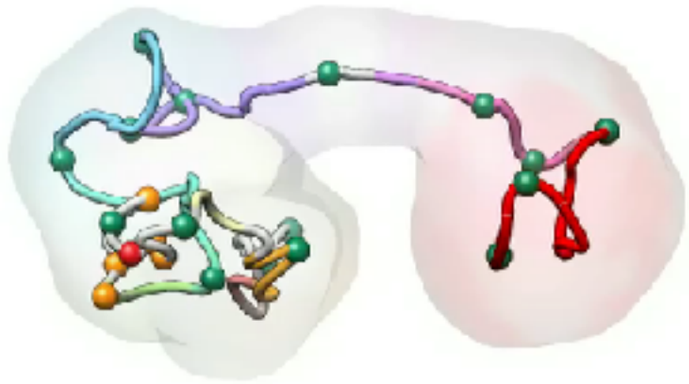
Increased in K562

FISH validation

GM12878

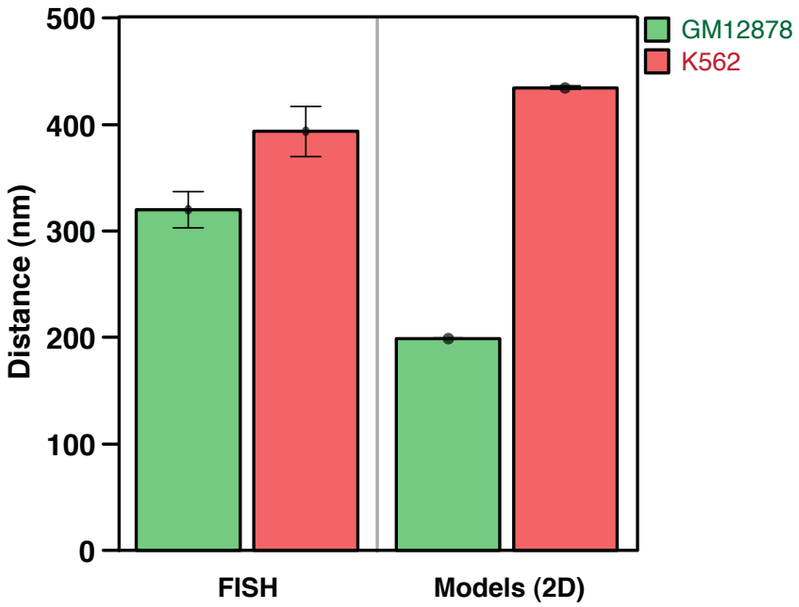
Cluster #1
2780 model

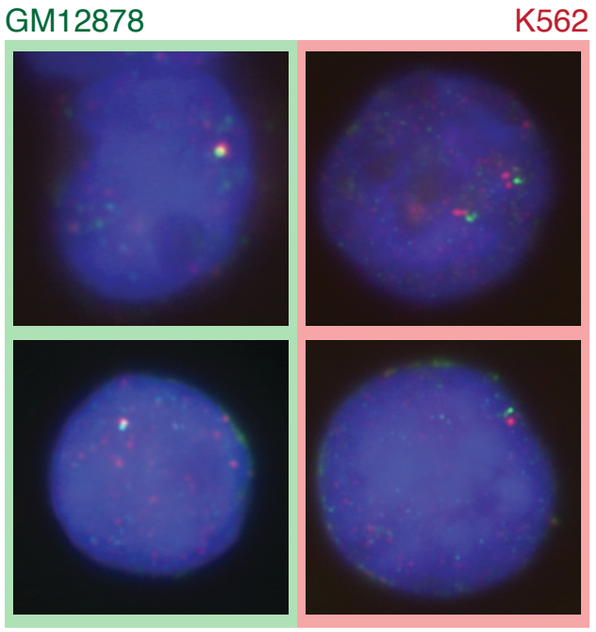




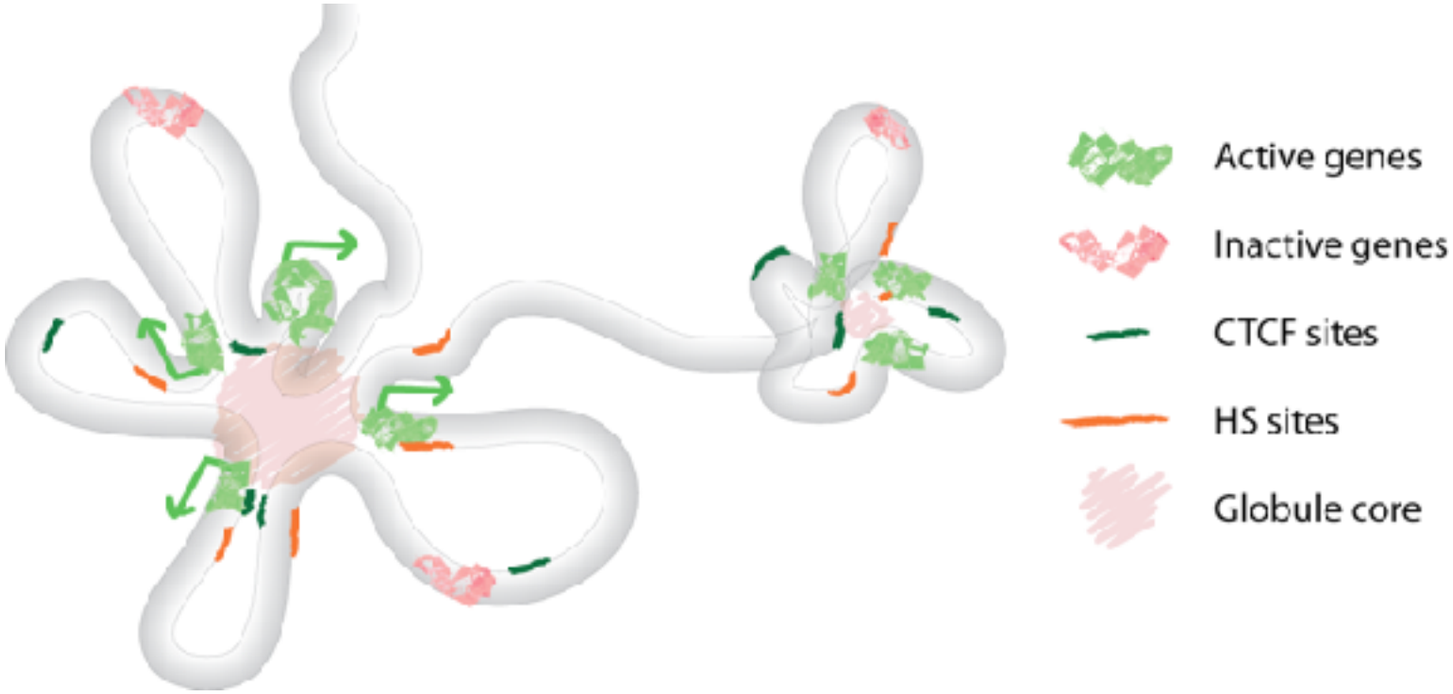
K562

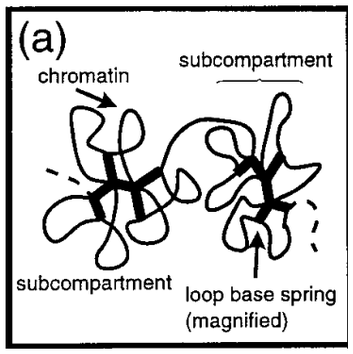
Cluster #2
314 model



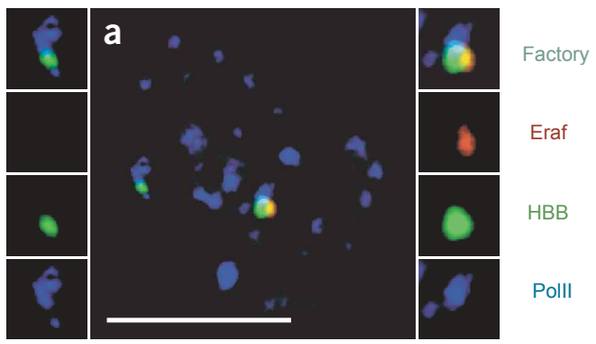


The "Chromatin Globule" model

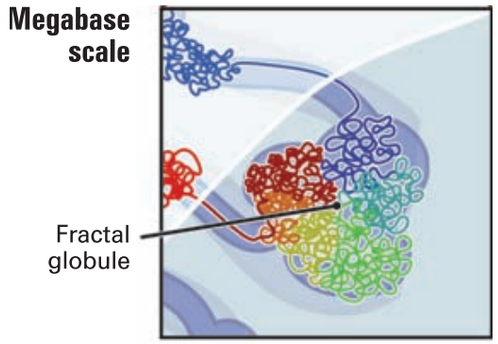




Münkel et al. JMB (1999)



Osborne et al. Nat Genet (2004)



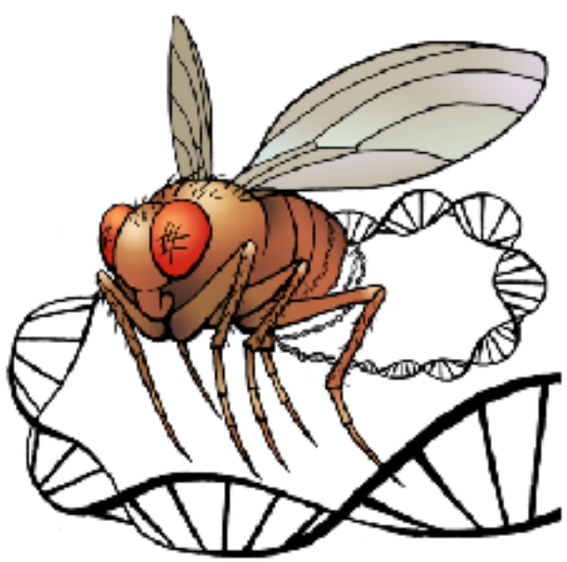
Lieberman-Aiden et al. Science (2009)

D. Baù *et al.* **Nat Struct Mol Biol** (2011) 18:107-14

A. Sanyal *et al.* **Current Opinion in Cell Biology** (2011) 23:325–33.

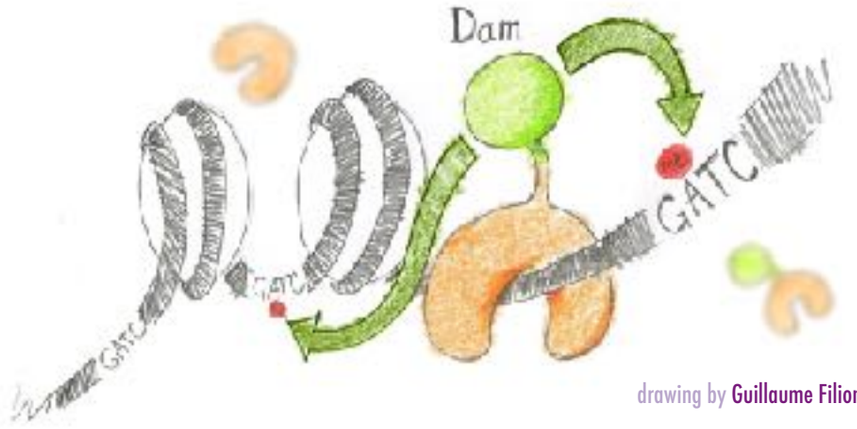
Structuring the COLORs of chromatin

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

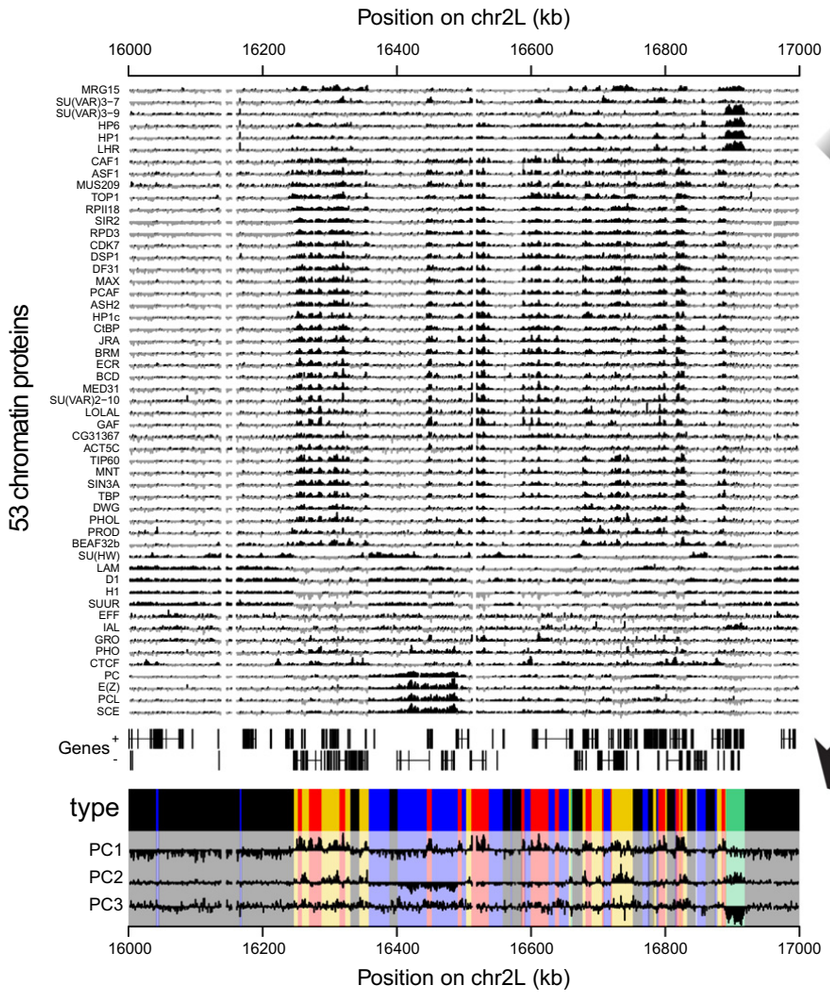


Fly Chromatin COLORs

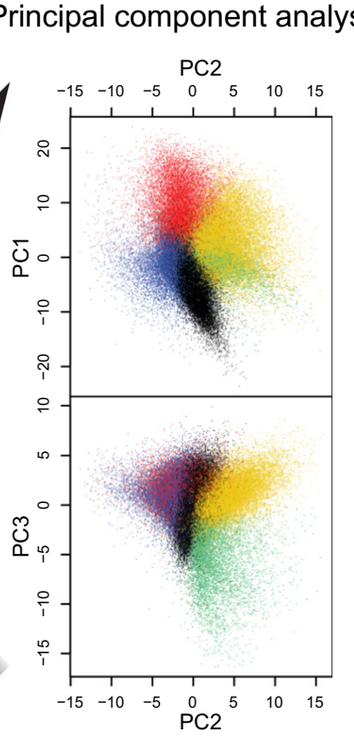
Filion et al. (2010). Cell, 143(2), 212–224.



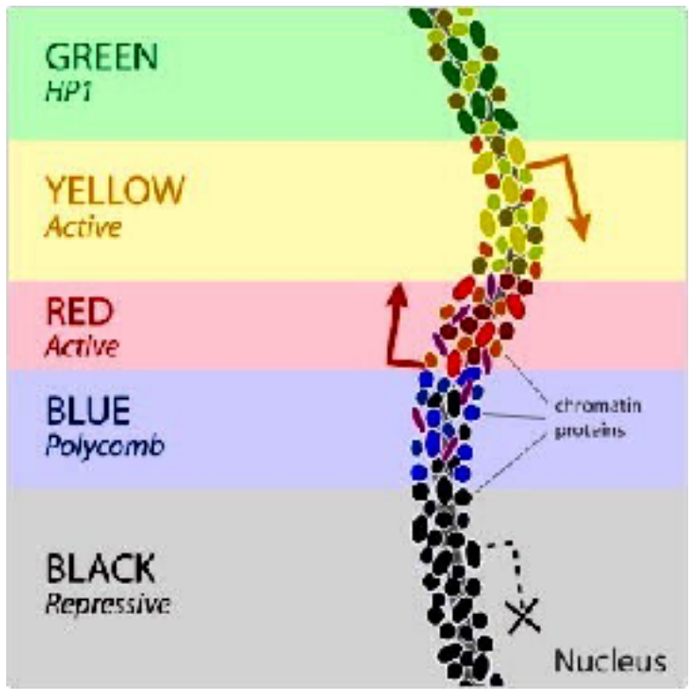
drawing by Guillaume Filion



Principal component analysis

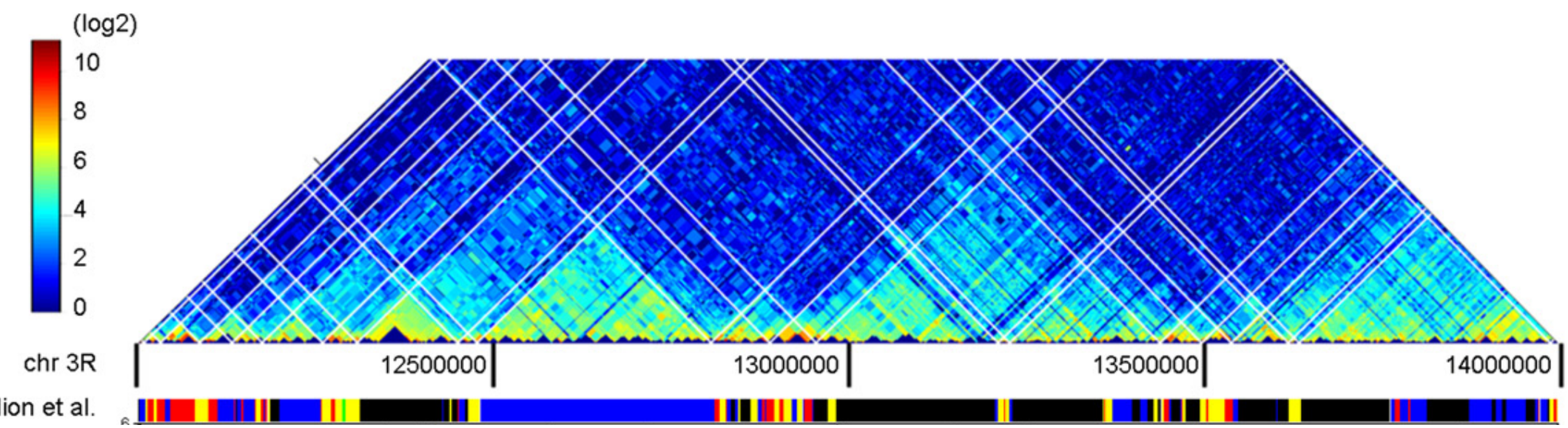


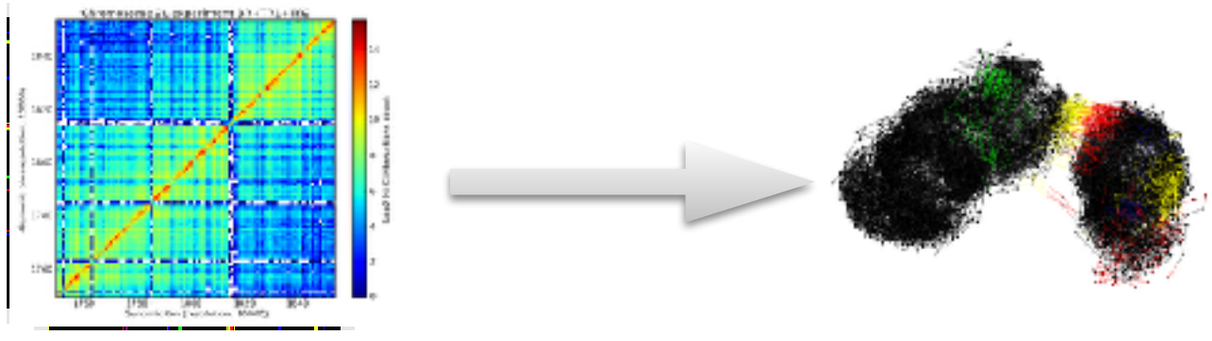
Hidden Markov model



Fly Chromatin COLORs

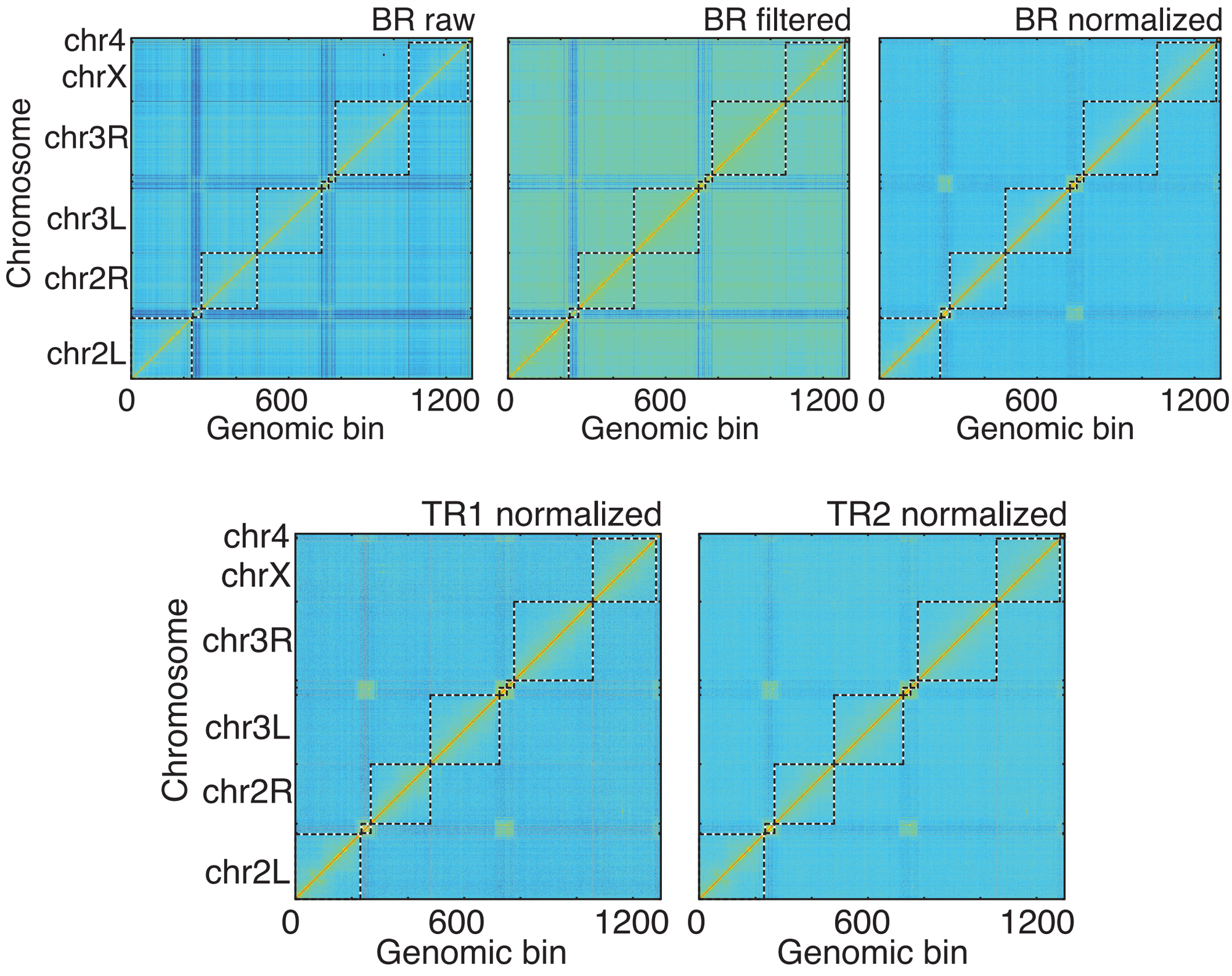
Hou et al. (2012). Molecular Cell, 48(3), 471–484.



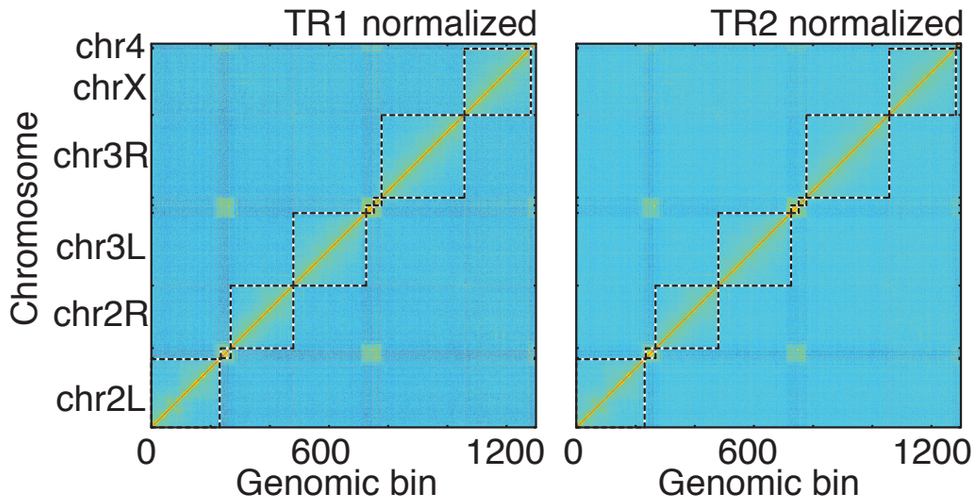


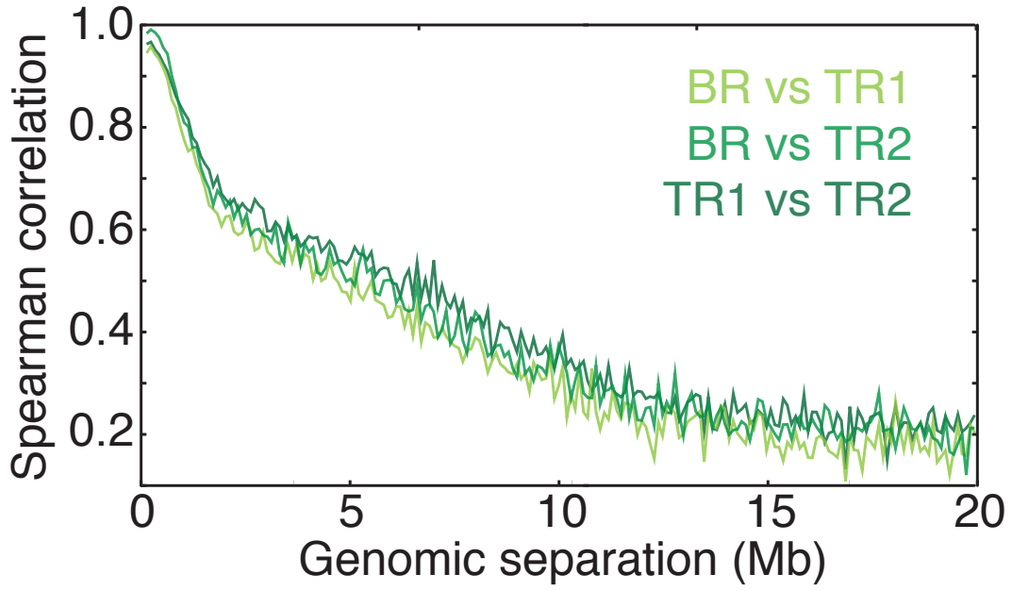
~200 regions of ~5Mb each
2Kb resolution

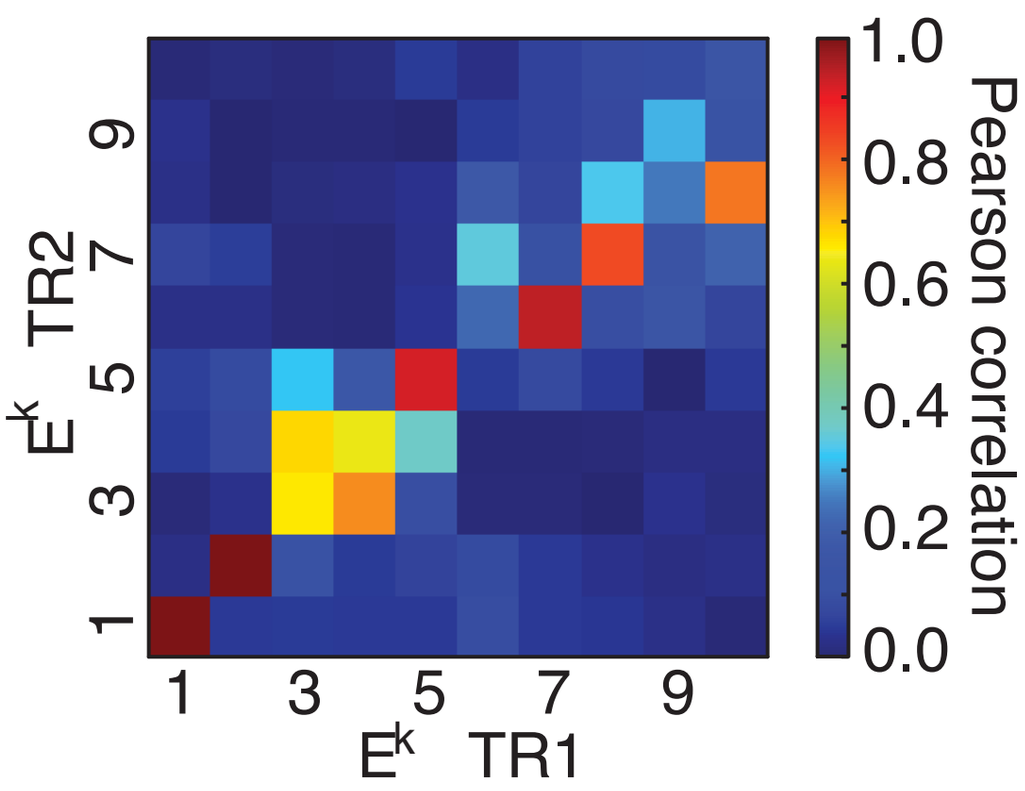
Mapping · Filtering · Normalizing



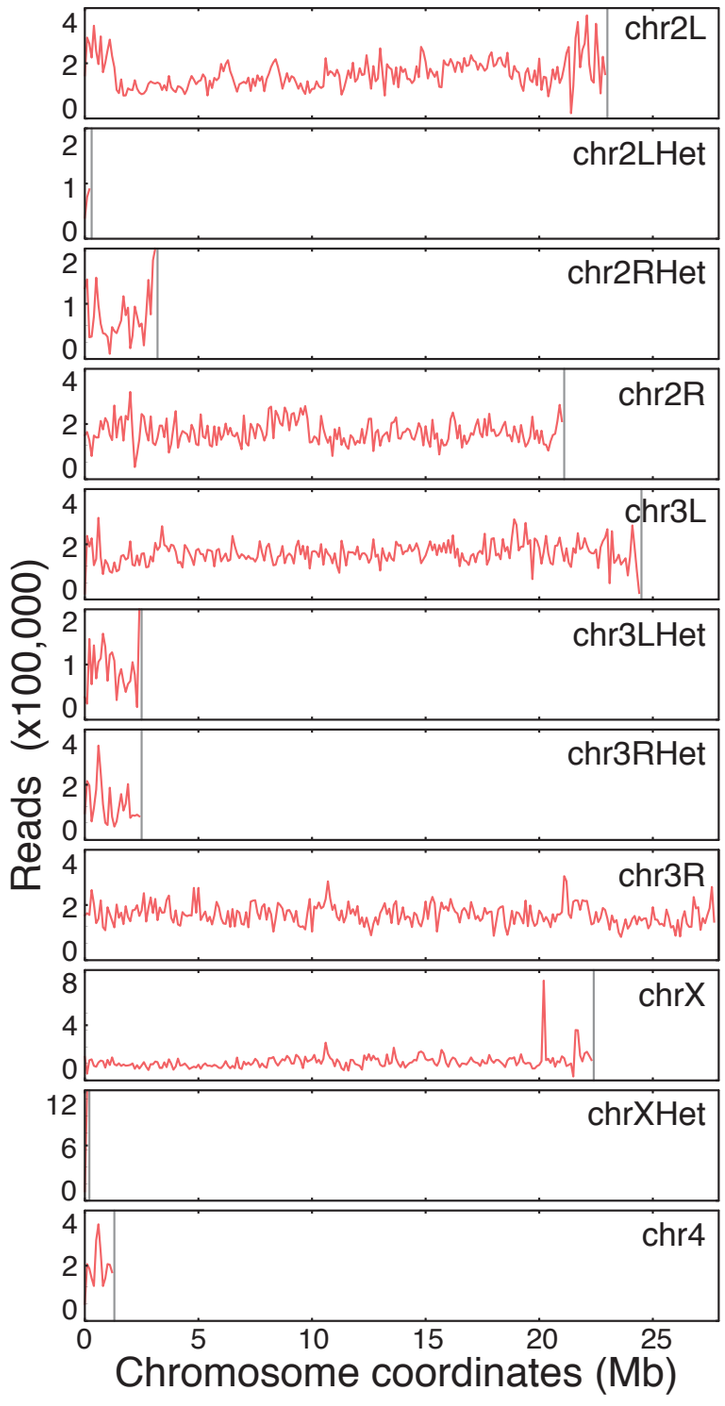
Matrix comparison

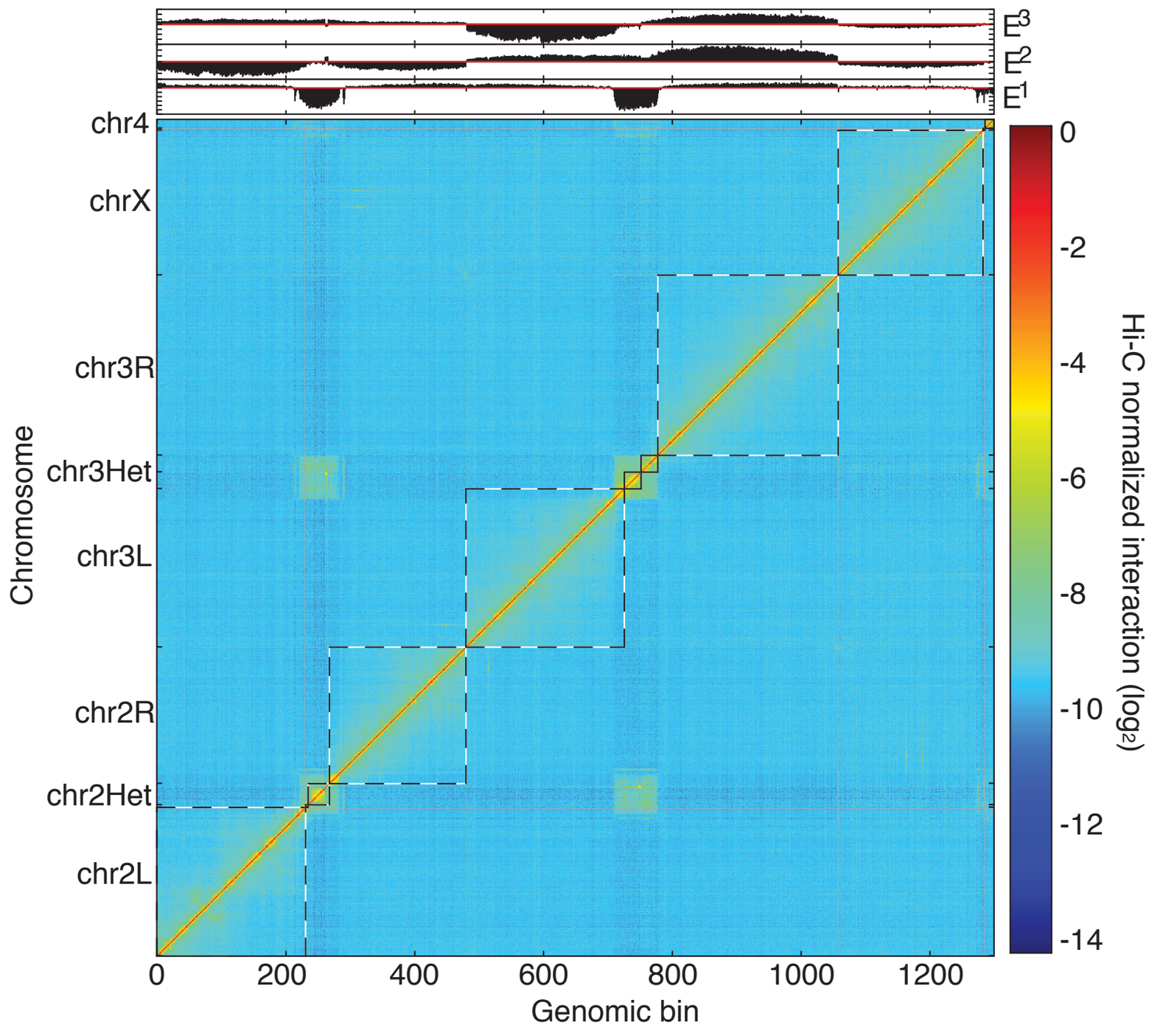




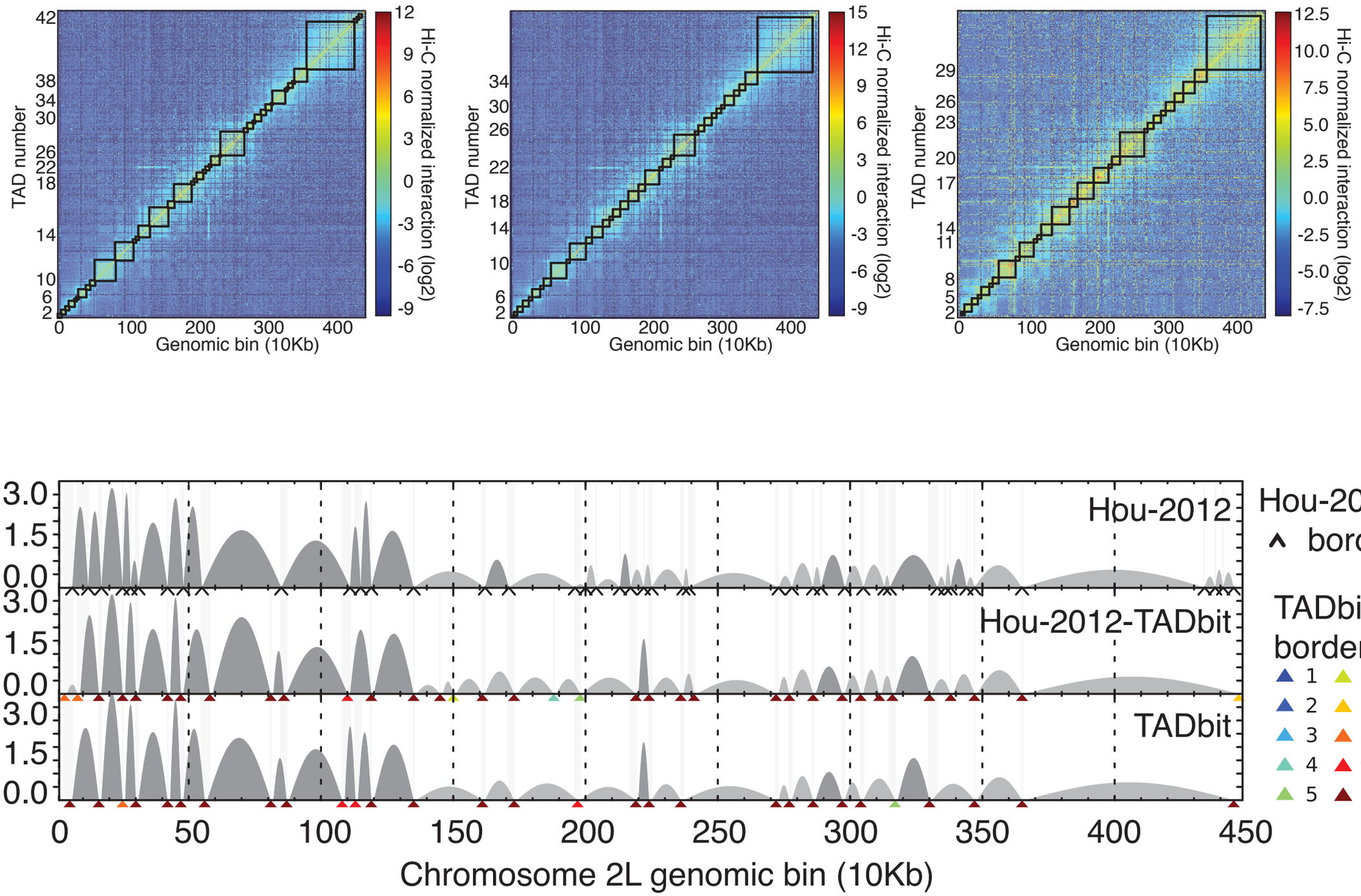


Matrix merging





TAD detection · comparison

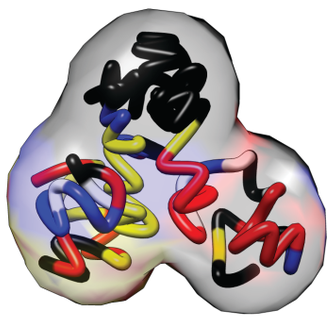


Structural properties

50 1Mb regions. 10 enriched for each color.

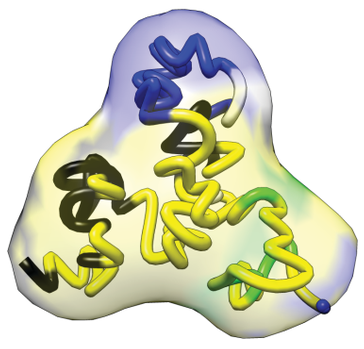
RED dense region
3R:18920000-19920000

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



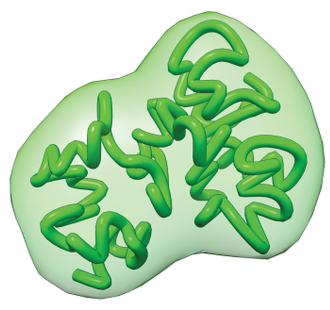
YELLOW dense region
X:15590000-16600000

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



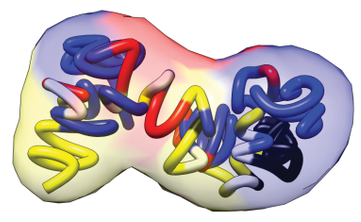
GREEN dense region
2R:510000-1530000

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



BLUE dense region
3L:210000-1230000

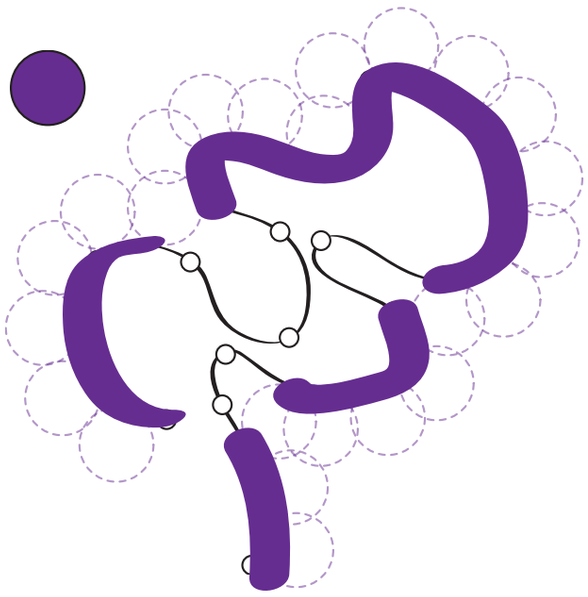
11% 17% 0% 52% 13% 0%



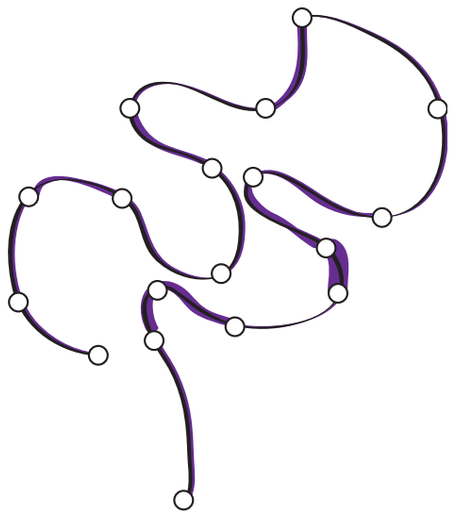
BLACK dense region
2L:17500000-18530000

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

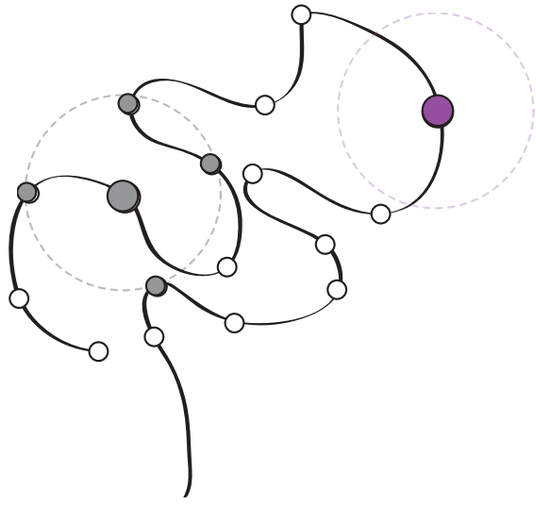
Accessibility (%)



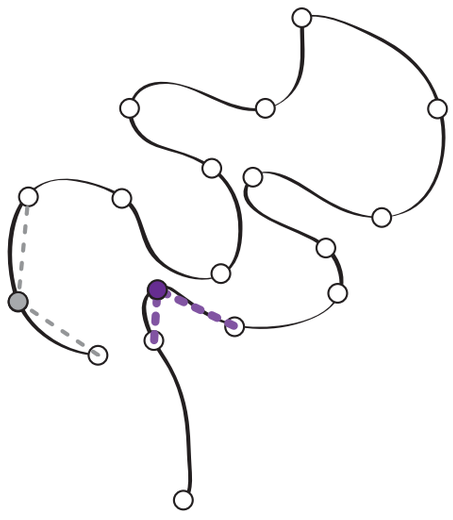
Density (bp/nm)



Interactions



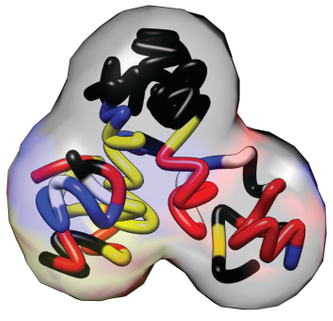
Angle



Structural COLORs

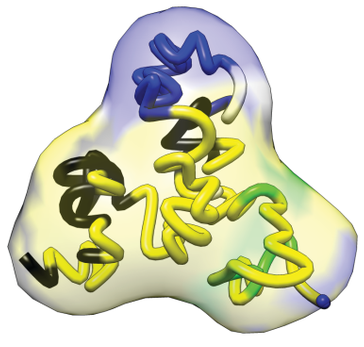
RED dense region
3R:18920000-19920000

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



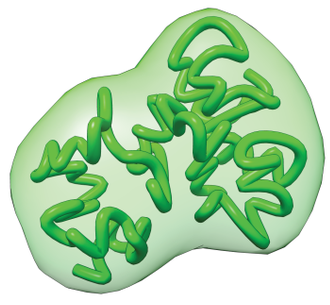
YELLOW dense region
X:15590000-16600000

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



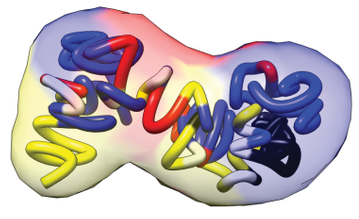
GREEN dense region
2R:510000-1530000

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



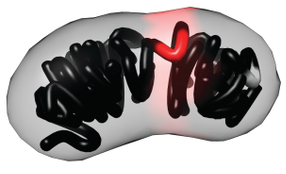
BLUE dense region
3L:210000-1230000

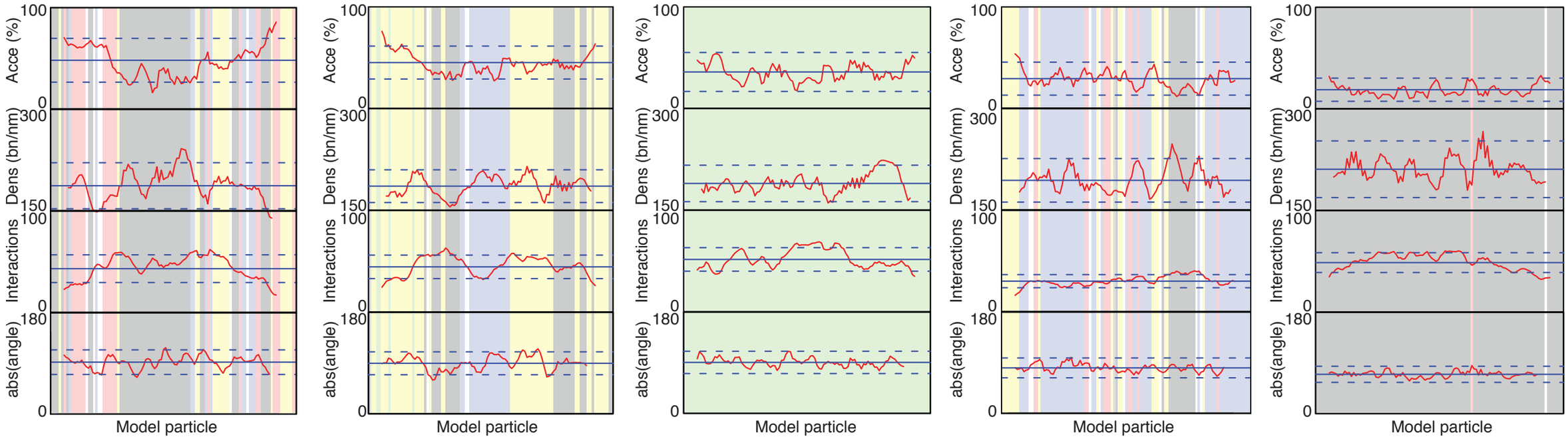
11% 17% 0% 52% 13% 0%



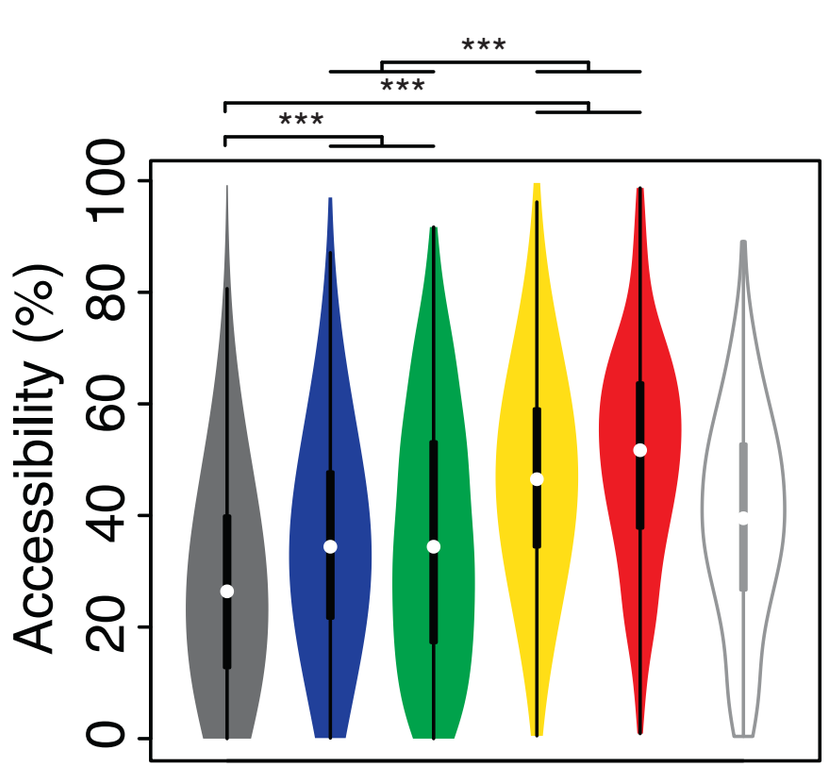
BLACK dense region
2L:17500000-18530000

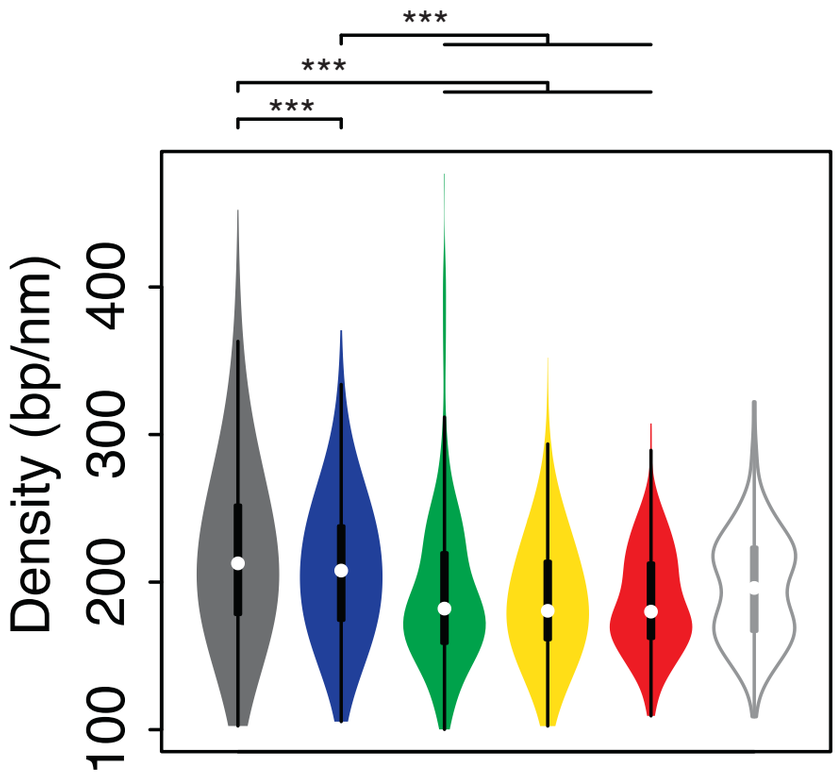
1% 0% 0% 0% 98% 1%

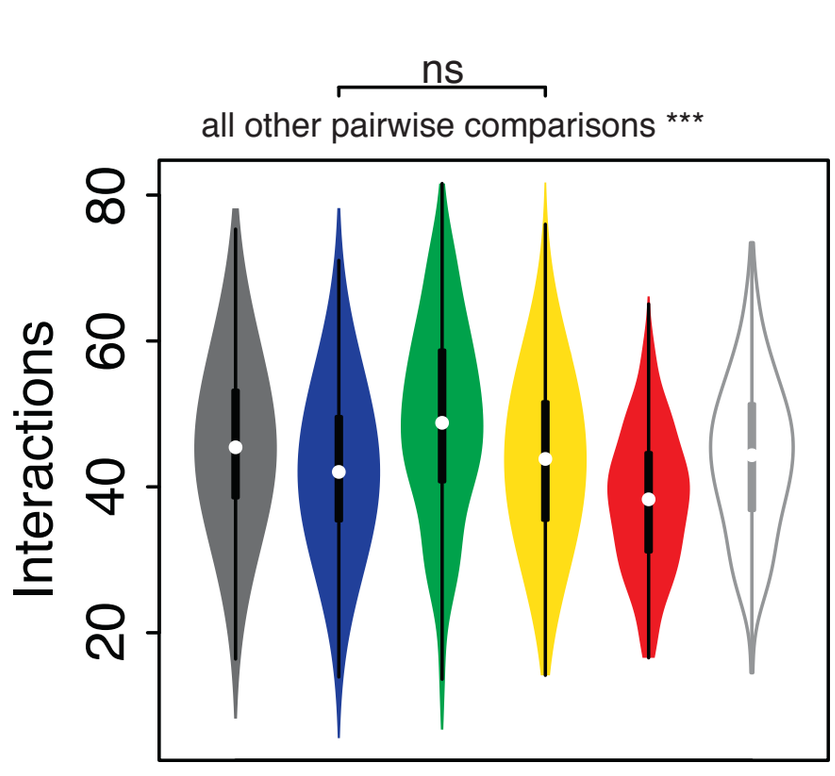


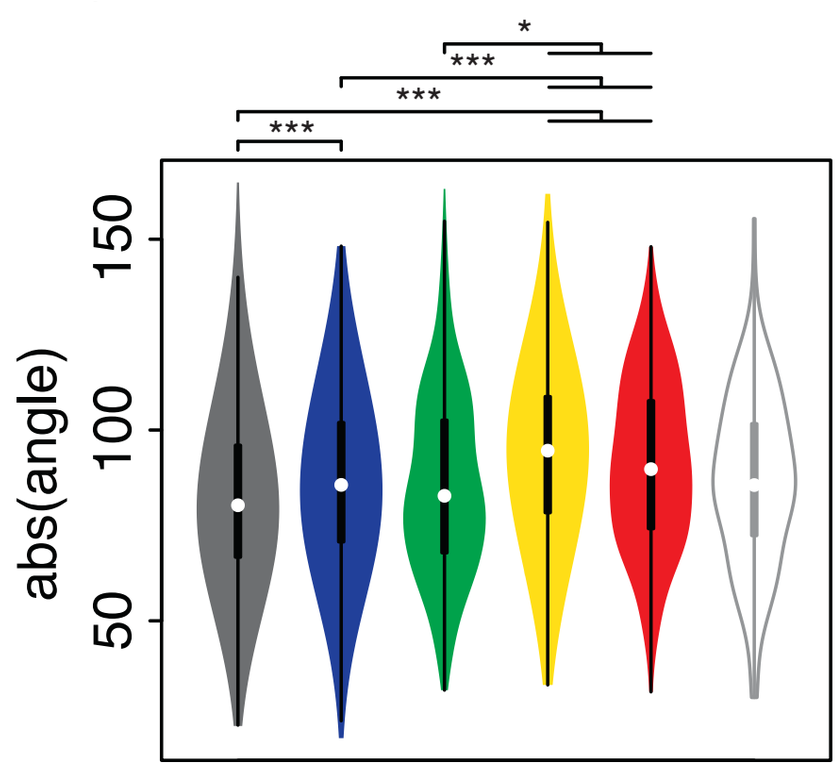


Structural COLORs

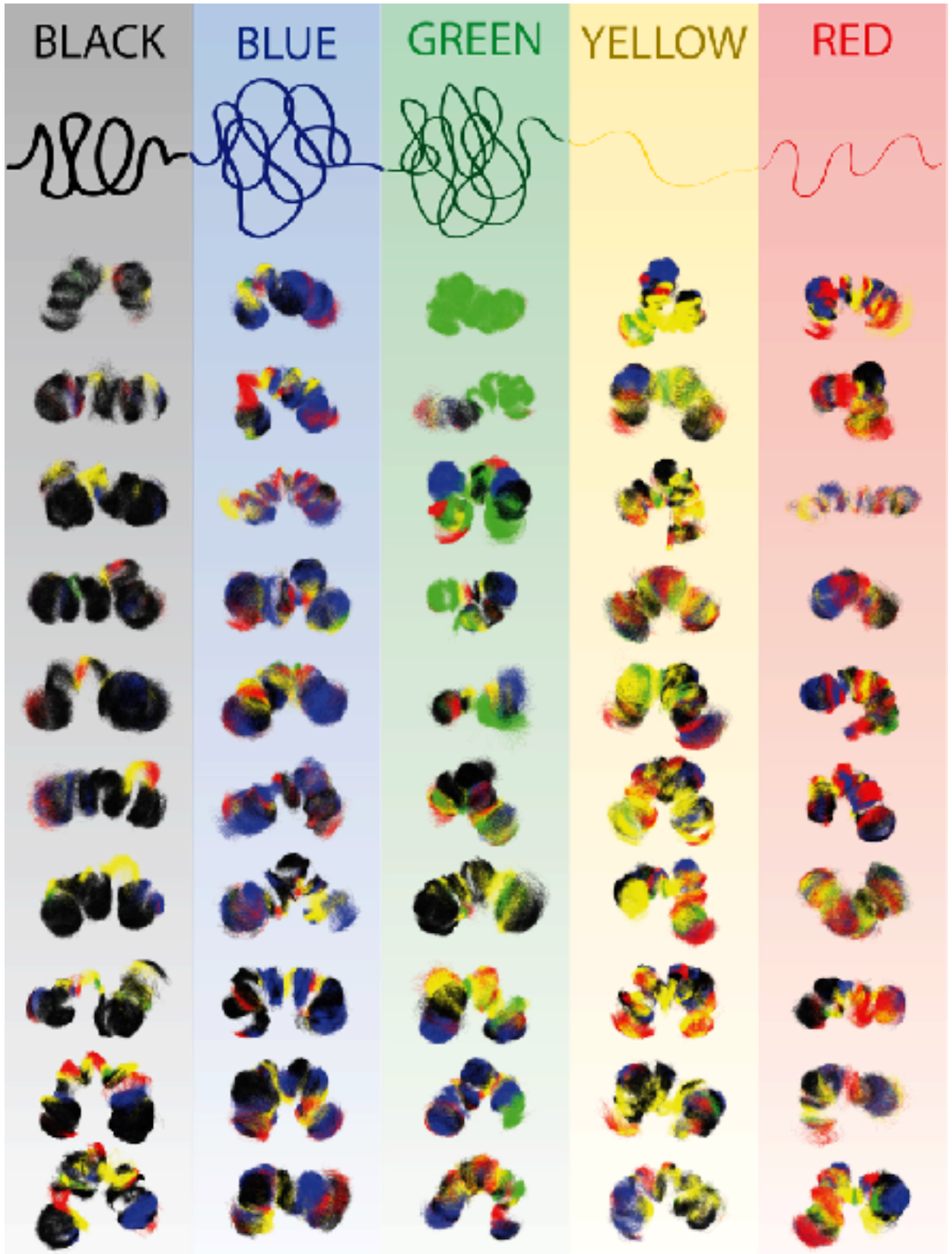


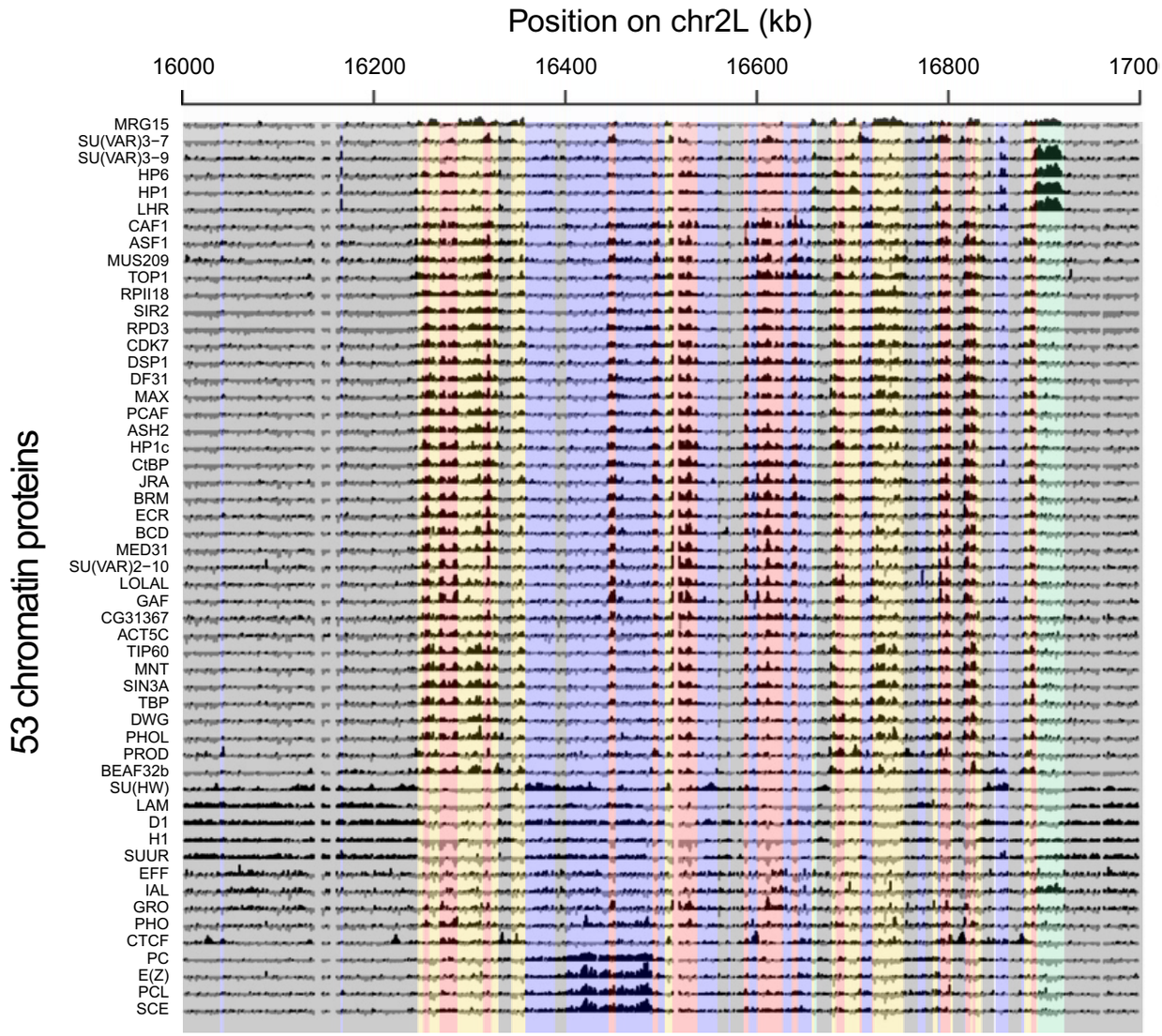






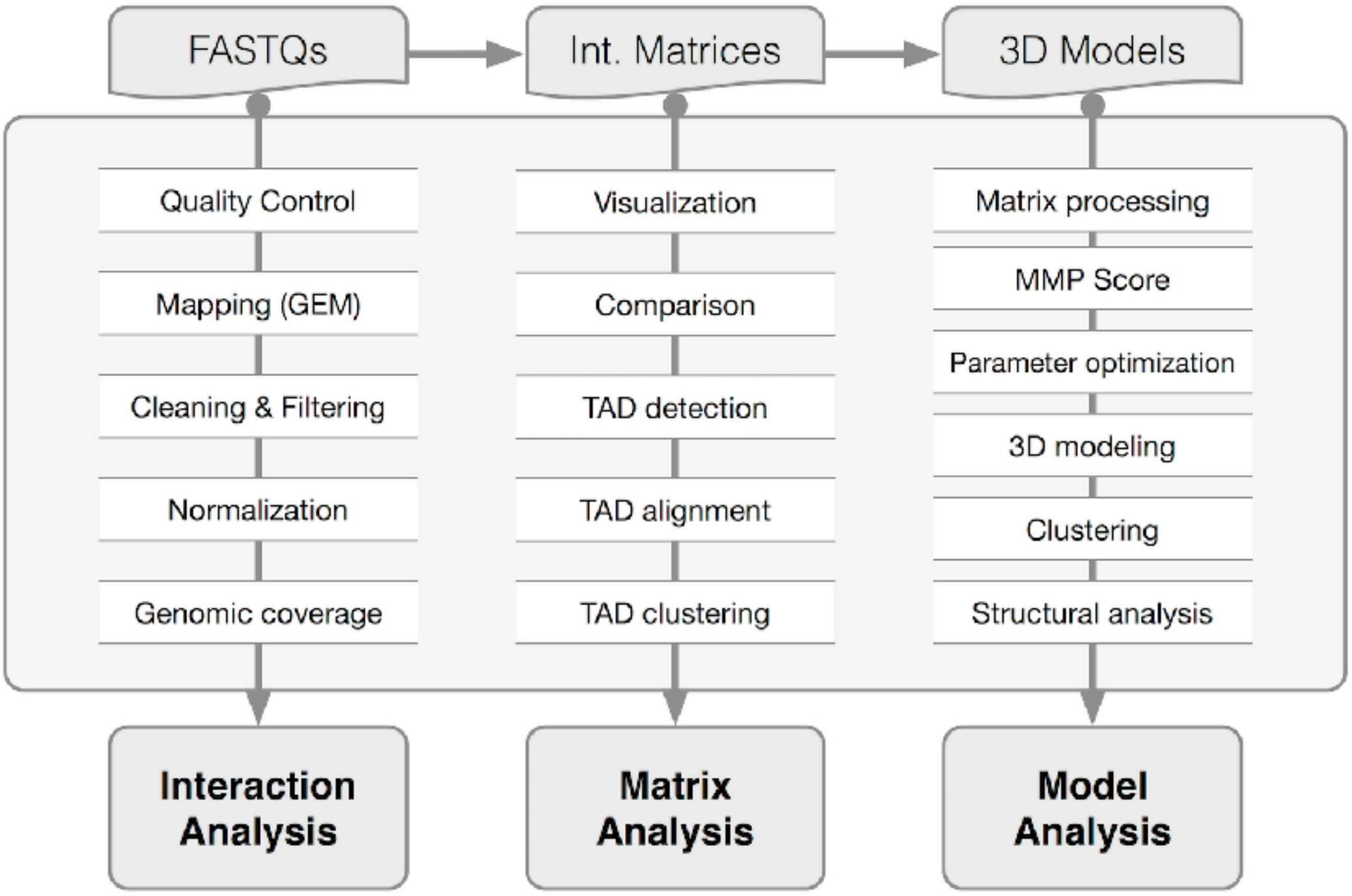
Structural COLORs





TADbit

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



TADbit 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[3], Meg Byron[3], Jeanne B Lawrence[3], Job Dekker[2] & Marc A Marti-Renom[1]

We developed a general approach that combines chromosome conformation capture–carbon copy (5C) with the Integrated Modeling Platform (IMP) to generate high-resolution three-dimensional (3D) models of chromosomes in the nucleus. We applied this approach to the ENm008 domain on human chromosome 16, containing the α-globin locus, which is expressed in K562 cells and silenced in lymphoblastoid cells (GM12878). The models accurately reproduce the known looping interactions between the α-globin genes and their distal regulatory elements. Further, we find using our approach that the domain folds into a single globular conformation in GM12878 cells, whereas two globules are formed in K562 cells. The central cores of these globules are enriched for transcribed genes, whereas nontranscribed chromatin is more peripheral. Our results suggest that transcription mediates a higher-order folding state related to clustering of transcribed genes around shared transcription machineries, as previously observed by microscopy.

Currently, efforts are directed at producing high-resolution genome annotations in which the positions of functional elements or specific chromatin states are mapped onto the linear genome sequence[1]. However, these linear representations do not indicate functional or structural relationships between distant elements. For example, recent studies have implicated distal enhancer elements in the control of regulate gene expression by engaging in long-range chromatin looping interactions. The three-dimensional (3D) organization of chromosomes is thought to facilitate compartmentalization[2,3], chromatin organization[4] and spatial sequestration of genes and their regulatory elements[5–7], all of which may modulate the output and functional state of the genome. A general approach for determining the spatial organization of chromatin can aid in the identification of long-range relationships between genes and distant regulatory elements as well as in the identification of higher-order folding principles of chromatin in general.

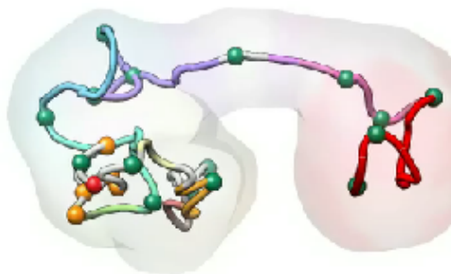
Chromosome conformation capture (3C)–based assays use formaldehyde cross-linking followed by restriction digestion and intramolecular ligation to study chromatin looping interactions[7–11]. 3C-based assays have been used to show that specific elements such as promoters, enhancers and insulators are involved in the formation of chromatin loops[12–14], to identify changes in chromatin conformation during folding[15–17] and, when comprehensive chromatin interaction data are can help researchers build spatial models of chromatin.

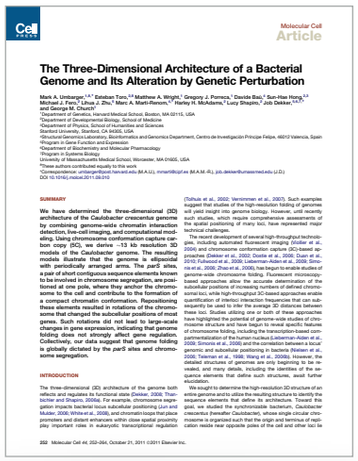
Previously, chromatin conformation has been modeled using polymer models[18,19] and molecular-dynamics simulations[20], which have proven valuable for understanding general features of chromatin fibers, including flexibility and compaction[20,21]. However, such methods only partially leverage the current wealth of experimental data on chromatin folding. Recently, experimentally driven approaches, in combination with computational modeling, have resulted in low-resolution models for the topological conformation of the immunoglobulin heavy-chain locus[22] and the *Hox* gene clusters[23]. However, both of these methods were limited by the resolution and completeness of the input experimental data[22], by insufficient model representation, scoring and optimization[23], or by limited analysis of the 3D models[23].

To overcome such limitations, we developed a new approach that couples high-throughput 3C experiments[24] with the IMP[25]. We applied this approach to determine the higher-order spatial organization of a 500-kilobase (kb) gene-dense domain located near the left telomere of human chromosome 16 (**Fig. 1a**). Embedded in this cluster of globin genes, expressed housekeeping genes in the forms specific of globin synthesis. Our results were validated only in erythroid cells. This 500-kb domain corresponds to the ENm008 region extensively studied by the ENCODE pilot project (**Fig. 1b**)[26].

The α-globin locus has been used widely as a model to study the mechanisms of long-range and tissue-specific gene regulation[27,28]. The α-globin genes are upregulated by a set of functional elements characterized by the presence of DNase I–hypersensitive sites (HSs) located 33 to 48 kb upstream of the 5′ gene. One of these elements, HS40, is considered to be of particular importance[29,30]. This element can act as an enhancer in reporter constructs and its deletion greatly affects activation of the α-globin genes[31]. HS40 is bound by several erythroid transcription factors including GATA factors and NF-E2 (ref. 34). Notably, previous 3C studies have demonstrated direct long-range

[1]Structural Genomics Unit, Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe, Valencia, Spain. [2]Program in Gene Function and Expression, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [3]Department of Cell Biology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to J.D. (job.dekker@umassmed.edu) or M.A.M.-R. (mmarti@cnio.es).
Received 22 November 2009; accepted 29 September 2010; published online 5 December 2010; doi:10.1038/nsmb.1936





Molecular Cell
Article

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

Mark A. Umbarger,[1,4,7] Esteban Toro,[2,8] Matthew A. Wright,[1] Gregory J. Porreca,[1] Daivide Baù,[3,5] Sun-Hae Hong,[3,5] Michael J. Fero,[1] Lihua J. Zhu,[1] Marc A. Marti-Renom,[3,6] Harley H. McAdams,[1] Lucy Shapiro,[2] Job Dekker,[3,4,5] and George M. Church[1]
[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[2]Department of Developmental Biology, Stanford University, Stanford, CA 94305, USA
[3]Department of Physics, School of Humanities and Sciences
Stanford University, Stanford, CA 94305, USA
[4]Structural Genomics Laboratory, Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe, 46012 Valencia, Spain
[5]Program in Gene Function and Expression
[6]Program in Bioinformatics and Integrative Biology and Molecular Pharmacology
[7]Program in Systems Biology
University of Massachusetts Medical School, Worcester, MA 01605, USA
[8]These authors contributed equally to this work
*Correspondence: umbarger@post.harvard.edu (M.A.U.), mmarti@cnio.es (M.A.M.-R.), job.dekker@umassmed.edu (J.D.)
DOI 10.1016/j.molcel.2011.09.010

SUMMARY

We have determined the three-dimensional (3D) architecture of the *Caulobacter crescentus* genome by combining genome-wide chromatin interaction detection, live-cell imaging, and computational modeling. Using chromosome conformation capture carbon copy (5C), we derive ~13 kb resolution 3D models of the *Caulobacter* genome. The resulting models illustrate that the genome is ellipsoidal with periodically arranged arms. The *parS* sites, a pair of short contiguous sequence elements known to be involved in chromosome segregation, are positioned at one pole, where they anchor the chromosome to the cell and contribute to the formation of a compact chromatin conformation. Repositioning these elements resulted in rotations of the chromosome that changed the subcellular positions of most genes. Such rotations did not lead to large-scale changes in gene expression, indicating that genome folding does not strongly affect gene regulation. Collectively, our data suggest that genome folding is globally dictated by the *parS* sites and chromosome segregation.

INTRODUCTION

The three-dimensional (3D) architecture of the genome both reflects and regulates its functional state (Dekker, 2008; Misteli and Soutoglou, 2009). For example, chromosome conformation capture (3C)–based analyses have revealed tissue-specific positioning (Ioudinkova et al., 2006; Kurukuti et al., 2006; Ling et al., 2006; Sandhu et al., 2009; Schoenfelder et al., 2010; Simonis et al., 2006; Tolhuis et al., 2002; Umlauf et al., 2004; Zhao et al., 2006; Zhou et al., 2006) and chromosome conformation capture–based studies of chromosome structure and have begun to reveal specific features of genome organization, including the presence of insulated domains (Dekker et al., 2002; Dostie et al., 2006; Duan et al., 2010; Fullwood et al., 2009; Lieberman-Aiden et al., 2009; Simonis et al., 2006; Tolhuis et al., 2006; Yaffe and Tanay, 2011). However, the detailed structures of genomes are only beginning to be revealed, and many details, including the identities of the sequence elements that define such structures, await further elucidation.

We sought to determine the high-resolution 3D structure of an entire genome and to define the specific structure to identify the elements that define its architecture. Toward this goal, we studied the synchronizable bacterium, *Caulobacter crescentus* (hereafter *Caulobacter*), whose single circular chromosome is organized such that the origin and terminus of replication reside near opposite poles of the cell and other loci lie

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,[1,2,8] Roni H.G. Wright,[1,2] Cecilia Ballare,[1,2] Guillaume Filion,[1,3] Marc A. Marti-Renom,[1,3,5] and Miguel Beato[1,2]
[1]Gene Regulation, Stem Cells, and Cancer Program, Centro de Regulació Genomica (CRG), 08003 Barcelona, Spain; [2]Universitat Pompeu Fabra (UPF), 08002 Barcelona, Spain; [3]Centre for Biology Cell Cycle, Centre Nacional d'Analisis Genomica (CNAG), 08028 Barcelona, Spain; [4]Hospital Clinic, Universitat de Barcelona, 08036 Barcelona, Spain; [5]Institucio Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

The human genome is segmented into topologically associating domains (TADs), but the role of this conserved organization during transient changes in gene expression is not known. Here we describe the distribution of progestin-induced chromatin modifications and changes in transcriptional activity over TADs in T47D breast cancer cells. Using ChIP-seq (chromatin immunoprecipitation combined with high-throughput sequencing), Hi-C (chromosome capture followed by high-throughput sequencing), and three-dimensional (3D) modeling techniques, we found that the borders of the ~3000 TADs in these cells are largely maintained after hormone treatment and that most of the TADs (~2000) are either transcriptionally activated or repressed. Remarkably, the genes within either transcriptionally activated or repressed TADs in a coordinated fashion. The epigenetic signatures of the TADs are homogeneously modified by hormones in correlation with the transcriptional changes. Hormone-induced changes in gene activity and chromatin remodeling are accompanied by differential structural changes for activated and repressed TADs, as reflected by specific and opposite changes in the strength of inter-TAD interactions within responsive TADs. Indeed, 3D modeling of the Hi-C data suggested that the structure of TADs was modified upon treatment. The differential responses of TADs to progestins and estrogens suggest that TADs could function as "regulons" to enable spatially proximal genes to be coordinately transcribed in response to hormones.

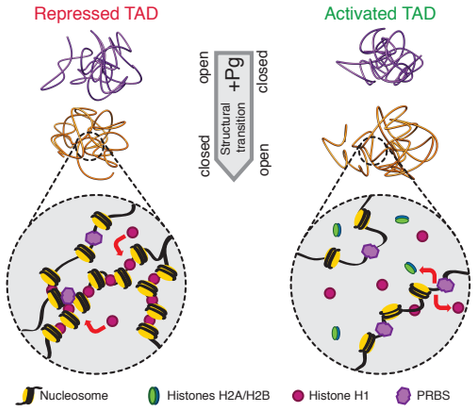
[*Keywords*: three-dimensional structure of the genome; gene expression; Hi-C; TADs; transcriptional regulation; epigenetic landscape; progesterone receptor]

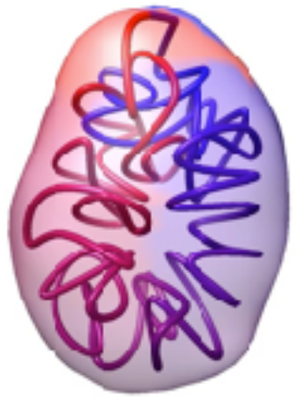
Supplemental material is available for this article.

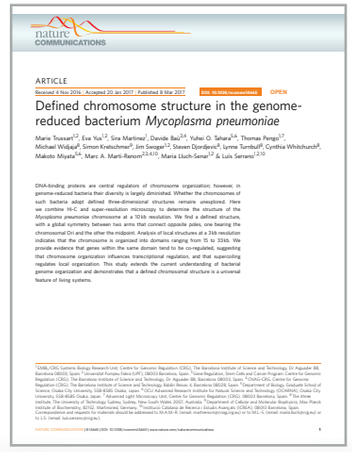
Received March 11, 2014; revised version accepted August 19, 2014.

The three-dimensional (3D) organization of the genome within the cell nucleus is non-random and might contribute to cell-specific gene expression. High-throughput chromosome conformation capture (3C)–derived (Dekker et al. 2002) methods have revealed that chromosome folding is organized in a hierarchy of two compartments—one open and one closed—that tend to be spatially segregated depending on their transcriptional activity (Lieberman-Aiden et al. 2009). At a finer level of organization, some functionally related genes have been shown to be brought close in space to be transcribed in a correlated fashion during cell differentiation. These genes, which can be located on different chromosomes, are organized in spatial clusters and preferentially transcribed in the same "factories" (Osborne et al. 2004, 2007; Cavalli 2007). Whether such mechanisms participate in transient modifications of the transcription rate in differentiated cells responding to physiological stimuli remains unclear (Fullwood et al. 2009; Kocanova et al. 2010; Hakim et al. 2011). Transient regulation of gene expression at the transcription level depends on the establishment of regulatory

GENES & DEVELOPMENT 28:2151–2162 Published by Cold Spring Harbor Laboratory Press; ISSN 0890-9369/14; www.genesdev.org 2151







nature COMMUNICATIONS

ARTICLE

Received 4 May 2016 | Accepted 20 Jan 2017 | Published 8 Mar 2017 DOI: 10.1038/ncomms14852 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 Swanger[1,2], Steven Djordjevic[8], Lynne Turnbull[8], Cynthia Whitchurch[8], Makoto Miyata[5,6], Marc A. Marti-Renom[3,4,6,10], Maria Lluch-Senar[1,2] & Luis Serrano[1,2,10]

DNA-binding proteins are central regulators of chromosome organization; however, in genome-reduced bacteria their diversity is largely diminished. Whether the chromosomes of such bacteria adopt defined three-dimensional structures remains unexplored. Here we combine Hi-C and super-resolution microscopy to determine the structure of the *Mycoplasma pneumoniae* chromosome at a 10 kb resolution. We find a defined structure, with a global symmetry between two arms that connect opposite poles, one bearing the chromosomal Ori and the other the midpoint. Analysis of local structures at a 2 kb resolution indicates that the chromosome is organized into domains ranging from 75 to 334 kb. We provide evidence that genes within the same domain tend to be co-regulated, suggesting that chromosome organization influences transcriptional regulation, and that supercoiling regulates local organization. This study extends the current understanding of bacterial genome organization and demonstrates that a defined chromosomal structure is a universal feature of living systems.

[1]EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr Aiguader 88, Barcelona 08003, Spain. [2]Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain. [3]Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain. [4]CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Baldiri Reixac 4, Barcelona 08028, Spain. [5]Department of Biology, Graduate School of Science, Osaka City University, SUI-BIKEN, Osaka, Japan. [6]OCU Advanced Research Institute for Natural Science and Technology (OCARINA), Osaka City University, Osaka, Japan. [7]Australian Centre for Microscopy & Microanalysis, The University of Sydney, Madsen Building F09, Sydney, New South Wales 2006, Australia. [8]Australian Centre for Microscopy & Microanalysis, Core Biotechnology Services, Molecular Horizons and School of Biomedical Sciences, University of Wollongong, Wollongong, New South Wales 2522, Australia. [9]Department of Cellular and Molecular Biophysics, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany. [10]Institució Catalana de Recerca i Estudis Avançats, (ICREA), 08010 Barcelona, Spain. Correspondence and requests for materials should be addressed to M.A.M.-R. (email: mmarti@cnio.es) or to L.S. (email: luis.serrano@crg.eu).

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[3], Meg Byron[3], Jeanne B Lawrence[3], Job Dekker[2] & Marc A Marti-Renom[1]

We developed a general approach that combines chromosome conformation capture–carbon copy (5C) with the Integrated Modeling Platform (IMP) to generate high-resolution three-dimensional (3D) models of chromosomes in the nucleus. We applied this approach to the ENm008 domain on human chromosome 16, containing the α-globin locus, which is expressed in K562 cells and silenced in lymphoblastoid cells (GM12878). The models accurately reproduce the known looping interactions between the α-globin genes and their distal regulatory elements. Further, we find using our approach that the domain folds into a single globular conformation in GM12878 cells, whereas two globules are formed in K562 cells. The central cores of these globules are enriched for transcribed genes, whereas nontranscribed chromatin is more peripheral. Our results suggest that transcription mediates a higher-order folding state related to clustering of transcribed genes around shared transcription machineries, as previously observed by microscopy.

Currently, efforts are directed at producing high-resolution genome annotations in which the positions of functional elements or specific chromatin states are mapped onto the linear genome sequence[1]. However, these linear representations do not indicate functional or structural relationships between distant elements. For example, recent studies have implicated distal enhancer elements in the control of regulate gene expression by engaging in long-range chromatin looping interactions. The three-dimensional (3D) organization of chromosomes is thought to facilitate compartmentalization[2,3], chromatin organization[4] and spatial sequestration of genes and their regulatory elements[5–7], all of which may modulate the output and functional state of the genome. A general approach for determining the spatial organization of chromatin can aid in the identification of long-range relationships between genes and distant regulatory elements as well as in the identification of higher-order folding principles of chromatin in general.

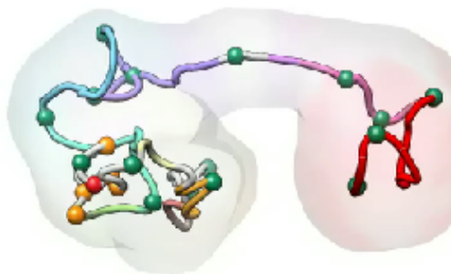
Chromosome conformation capture (3C)–based assays use formaldehyde cross-linking followed by restriction digestion and intramolecular ligation to study chromatin looping interactions[7–11]. 3C-based assays have been used to show that specific elements such as promoters, enhancers and insulators are involved in the formation of chromatin loops[12–14], to identify changes in chromatin conformation during folding[15–17] and, when comprehensive chromatin interaction data are can help researchers build spatial models of chromatin.

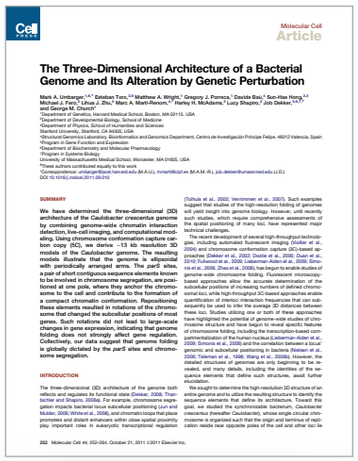
Previously, chromatin conformation has been modeled using polymer models[18,19] and molecular-dynamics simulations[20], which have proven valuable for understanding general features of chromatin fibers, including flexibility and compaction[20,21]. However, such methods only partially leverage the current wealth of experimental data on chromatin folding. Recently, experimentally driven approaches, in combination with computational modeling, have resulted in low-resolution models for the topological conformation of the immunoglobulin heavy-chain locus[22] and the *Hox* gene clusters[23]. However, both of these methods were limited by the resolution and completeness of the input experimental data[22], by insufficient model representation, scoring and optimization[23], or by limited analysis of the 3D models[23].

To overcome such limitations, we developed a new approach that couples high-throughput 3C experiments[24] with the IMP[25]. We applied this approach to determine the higher-order spatial organization of a 500-kilobase (kb) gene-dense domain located near the left telomere of human chromosome 16 (**Fig. 1a**). Embedded in this cluster of globin genes, expressed housekeeping genes in the forms specific of globin synthesis. Our results were validated only in erythroid cells. This 500-kb domain corresponds to the ENm008 region extensively studied by the ENCODE pilot project (**Fig. 1b**)[26].

The α-globin locus has been used widely as a model to study the mechanisms of long-range and tissue-specific gene regulation[27,28]. The α-globin genes are upregulated by a set of functional elements characterized by the presence of DNase I–hypersensitive sites (HSs) located 33 to 48 kb upstream of the 5′ gene. One of these elements, HS40, is considered to be of particular importance[29,30]. This element can act as an enhancer in reporter constructs and its deletion greatly affects activation of the α-globin genes[31]. HS40 is bound by several erythroid transcription factors including GATA factors and NF-E2 (ref. 34). Notably, previous 3C studies have demonstrated direct long-range

[1]Structural Genomics Unit, Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe, Valencia, Spain. [2]Program in Gene Function and Expression, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [3]Department of Cell Biology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to J.D. (job.dekker@umassmed.edu) or M.A.M.-R. (mmarti@cnio.es).
Received 22 November 2009; accepted 29 September 2010; published online 5 December 2010; doi:10.1038/nsmb.1936





Molecular Cell
Article

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

Mark A. Umbarger,[1,4,7] Esteban Toro,[2,8] Matthew A. Wright,[1] Gregory J. Porreca,[1] Daivide Baù,[3,5] Sun-Hae Hong,[3,5] Michael J. Fero,[1] Lihua J. Zhu,[1] Marc A. Marti-Renom,[3,6] Harley H. McAdams,[1] Lucy Shapiro,[2] Job Dekker,[3,4,5] and George M. Church[1]
[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[2]Department of Developmental Biology, Stanford University, Stanford, CA 94305, USA
[3]Department of Physics, School of Humanities and Sciences
Stanford University, Stanford, CA 94305, USA
[4]Structural Genomics Laboratory, Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe, 46012 Valencia, Spain
[5]Program in Gene Function and Expression
[6]Program in Bioinformatics and Integrative Biology and Molecular Pharmacology
[7]Program in Systems Biology
University of Massachusetts Medical School, Worcester, MA 01605, USA
[8]These authors contributed equally to this work
*Correspondence: umbarger@post.harvard.edu (M.A.U.), mmarti@cnio.es (M.A.M.-R.), job.dekker@umassmed.edu (J.D.)
DOI 10.1016/j.molcel.2011.09.010

SUMMARY

We have determined the three-dimensional (3D) architecture of the *Caulobacter crescentus* genome by combining genome-wide chromatin interaction detection, live-cell imaging, and computational modeling. Using chromosome conformation capture carbon copy (5C), we derive ~13 kb resolution 3D models of the *Caulobacter* genome. The resulting models illustrate that the genome is ellipsoidal with periodically arranged arms. The *parS* sites, a pair of short contiguous sequence elements known to be involved in chromosome segregation, are positioned at one pole, where they anchor the chromosome to the cell and contribute to the formation of a compact chromatin conformation. Repositioning these elements resulted in rotations of the chromosome that changed the subcellular positions of most genes. Such rotations did not lead to large-scale changes in gene expression, indicating that genome folding does not strongly affect gene regulation. Collectively, our data suggest that genome folding is globally dictated by the *parS* sites and chromosome segregation.

INTRODUCTION

The three-dimensional (3D) architecture of the genome both reflects and regulates its functional state (Dekker, 2008; Misteli and Soutoglou, 2009). For example, chromosome conformation capture (3C)–based analyses have revealed tissue-specific positioning (Ioudinkova et al., 2006; Kurukuti et al., 2006; Ling et al., 2006; Sandhu et al., 2009; Schoenfelder et al., 2010; Simonis et al., 2006; Tolhuis et al., 2002; Umlauf et al., 2004; Zhao et al., 2006; Zhou et al., 2006) and chromosome conformation capture–based studies of chromosome structure and have begun to reveal specific features of genome organization, including the presence of insulated domains (Dekker et al., 2002; Dostie et al., 2006; Duan et al., 2010; Fullwood et al., 2009; Lieberman-Aiden et al., 2009; Simonis et al., 2006; Tolhuis et al., 2006; Yaffe and Tanay, 2011). However, the detailed structures of genomes are only beginning to be revealed, and many details, including the identities of the sequence elements that define such structures, await further elucidation.

We sought to determine the high-resolution 3D structure of an entire genome and to define the specific structure to identify the elements that define its architecture. Toward this goal, we studied the synchronizable bacterium, *Caulobacter crescentus* (hereafter *Caulobacter*), whose single circular chromosome is organized such that the origin and terminus of replication reside near opposite poles of the cell and other loci lie

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,[1,2,8] Roni H.G. Wright,[1,2] Cecilia Ballare,[1,2] Guillaume Filion,[1,3] Marc A. Marti-Renom,[1,3,5] and Miguel Beato[1,2]
[1]Gene Regulation, Stem Cells, and Cancer Program, Centro de Regulació Genomica (CRG), 08003 Barcelona, Spain; [2]Universitat Pompeu Fabra (UPF), 08002 Barcelona, Spain; [3]Centre for Biology Cell Cycle, Centre Nacional d'Analisis Genomica (CNAG), 08028 Barcelona, Spain; [4]Hospital Clinic, Universitat de Barcelona, 08036 Barcelona, Spain; [5]Institucio Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

The human genome is segmented into topologically associating domains (TADs), but the role of this conserved organization during transient changes in gene expression is not known. Here we describe the distribution of progestin-induced chromatin modifications and changes in transcriptional activity over TADs in T47D breast cancer cells. Using ChIP-seq (chromatin immunoprecipitation combined with high-throughput sequencing), Hi-C (chromosome capture followed by high-throughput sequencing), and three-dimensional (3D) modeling techniques, we found that the borders of the ~3000 TADs in these cells are largely maintained after hormone treatment and that most of the TADs (~2000) are either transcriptionally activated or repressed. Remarkably, the genes within either transcriptionally activated or repressed TADs in a coordinated fashion. The epigenetic signatures of the TADs are homogeneously modified by hormones in correlation with the transcriptional changes. Hormone-induced changes in gene activity and chromatin remodeling are accompanied by differential structural changes for activated and repressed TADs, as reflected by specific and opposite changes in the strength of inter-TAD interactions within responsive TADs. Indeed, 3D modeling of the Hi-C data suggested that the structure of TADs was modified upon treatment. The differential responses of TADs to progestins and estrogens suggest that TADs could function as "regulons" to enable spatially proximal genes to be coordinately transcribed in response to hormones.

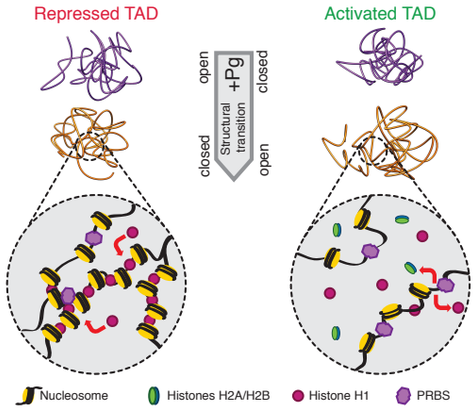
[*Keywords*: three-dimensional structure of the genome; gene expression; Hi-C; TADs; transcriptional regulation; epigenetic landscape; progesterone receptor]

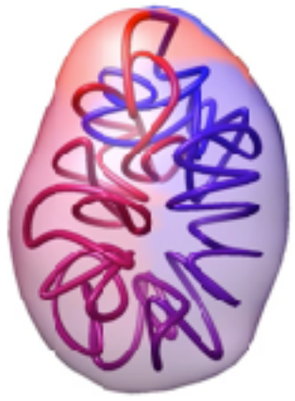
Supplemental material is available for this article.

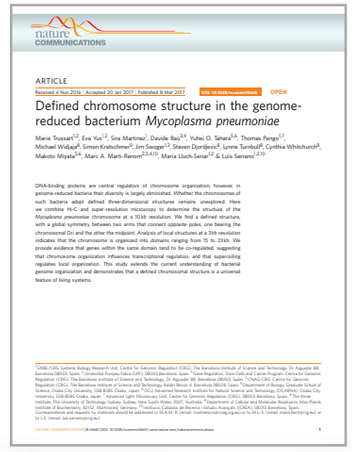
Received March 11, 2014; revised version accepted August 19, 2014.

The three-dimensional (3D) organization of the genome within the cell nucleus is non-random and might contribute to cell-specific gene expression. High-throughput chromosome conformation capture (3C)–derived (Dekker et al. 2002) methods have revealed that chromosome folding is organized in a hierarchy of two compartments—one open and one closed—that tend to be spatially segregated depending on their transcriptional activity (Lieberman-Aiden et al. 2009). At a finer level of organization, some functionally related genes have been shown to be brought close in space to be transcribed in a correlated fashion during cell differentiation. These genes, which can be located on different chromosomes, are organized in spatial clusters and preferentially transcribed in the same "factories" (Osborne et al. 2004, 2007; Cavalli 2007). Whether such mechanisms participate in transient modifications of the transcription rate in differentiated cells responding to physiological stimuli remains unclear (Fullwood et al. 2009; Kocanova et al. 2010; Hakim et al. 2011). Transient regulation of gene expression at the transcription level depends on the establishment of regulatory

GENES & DEVELOPMENT 28:2151–2162 Published by Cold Spring Harbor Laboratory Press; ISSN 0890-9369/14; www.genesdev.org 2151







nature COMMUNICATIONS

ARTICLE

Received 4 May 2016 | Accepted 20 Jan 2017 | Published 8 Mar 2017 DOI: 10.1038/ncomms14852 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 Swanger[1,2], Steven Djordjevic[8], Lynne Turnbull[8], Cynthia Whitchurch[8], Makoto Miyata[5,6], Marc A. Marti-Renom[3,4,6,10], Maria Lluch-Senar[1,2] & Luis Serrano[1,2,10]

DNA-binding proteins are central regulators of chromosome organization; however, in genome-reduced bacteria their diversity is largely diminished. Whether the chromosomes of such bacteria adopt defined three-dimensional structures remains unexplored. Here we combine Hi-C and super-resolution microscopy to determine the structure of the *Mycoplasma pneumoniae* chromosome at a 10 kb resolution. We find a defined structure, with a global symmetry between two arms that connect opposite poles, one bearing the chromosomal Ori and the other the midpoint. Analysis of local structures at a 2 kb resolution indicates that the chromosome is organized into domains ranging from 75 to 334 kb. We provide evidence that genes within the same domain tend to be co-regulated, suggesting that chromosome organization influences transcriptional regulation, and that supercoiling regulates local organization. This study extends the current understanding of bacterial genome organization and demonstrates that a defined chromosomal structure is a universal feature of living systems.

[1]EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr Aiguader 88, Barcelona 08003, Spain. [2]Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain. [3]Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, Barcelona 08003, Spain. [4]CNAG-CRG, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Baldiri Reixac 4, Barcelona 08028, Spain. [5]Department of Biology, Graduate School of Science, Osaka City University, SUI-BIKEN, Osaka, Japan. [6]OCU Advanced Research Institute for Natural Science and Technology (OCARINA), Osaka City University, Osaka, Japan. [7]Australian Centre for Microscopy & Microanalysis, The University of Sydney, Madsen Building F09, Sydney, New South Wales 2006, Australia. [8]Australian Centre for Microscopy & Microanalysis, Core Biotechnology Services, Molecular Horizons and School of Biomedical Sciences, University of Wollongong, Wollongong, New South Wales 2522, Australia. [9]Department of Cellular and Molecular Biophysics, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany. [10]Institució Catalana de Recerca i Estudis Avançats, (ICREA), 08010 Barcelona, Spain. Correspondence and requests for materials should be addressed to M.A.M.-R. (email: mmarti@cnio.es) or to L.S. (email: luis.serrano@crg.eu).

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

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

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