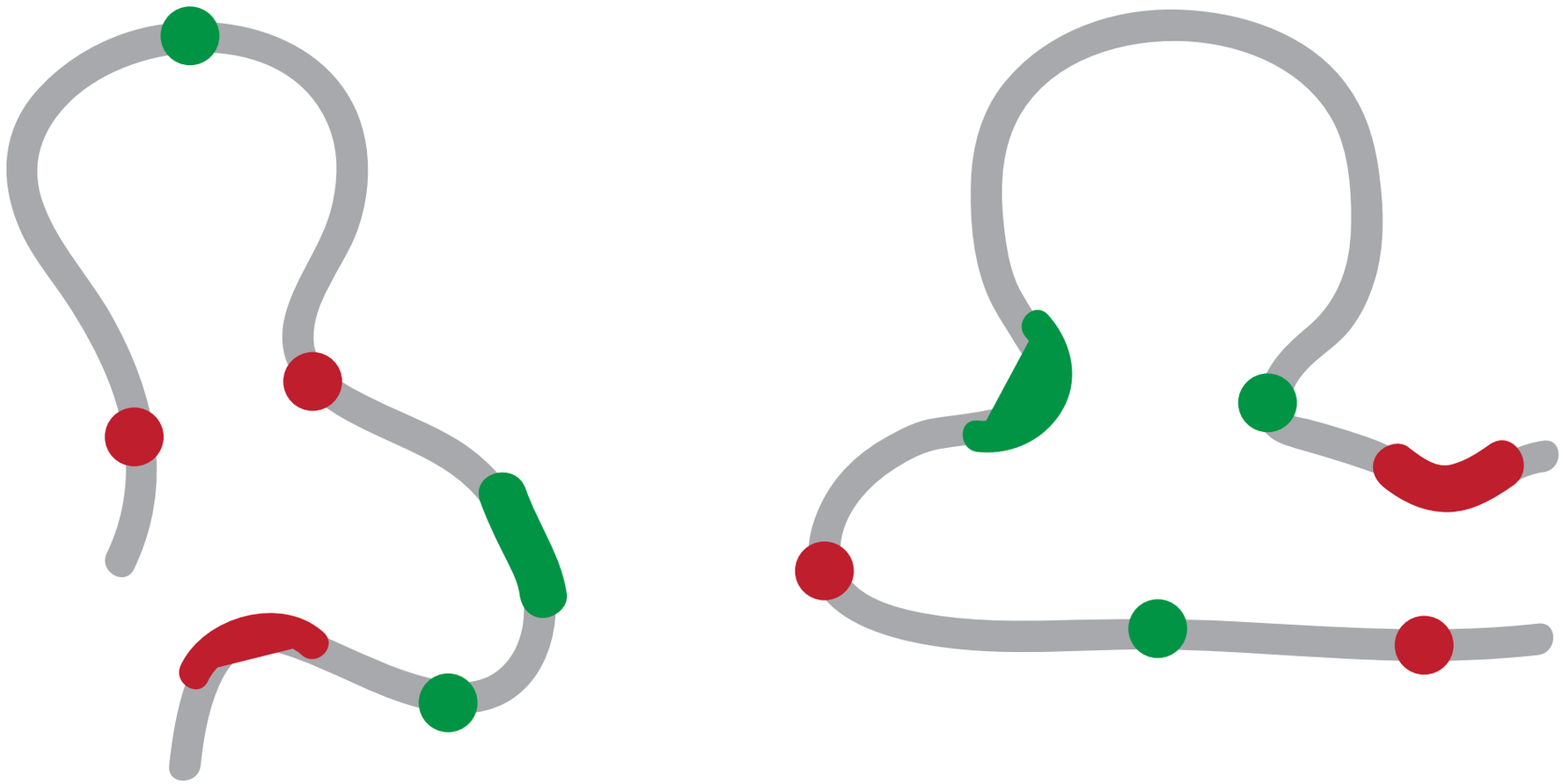
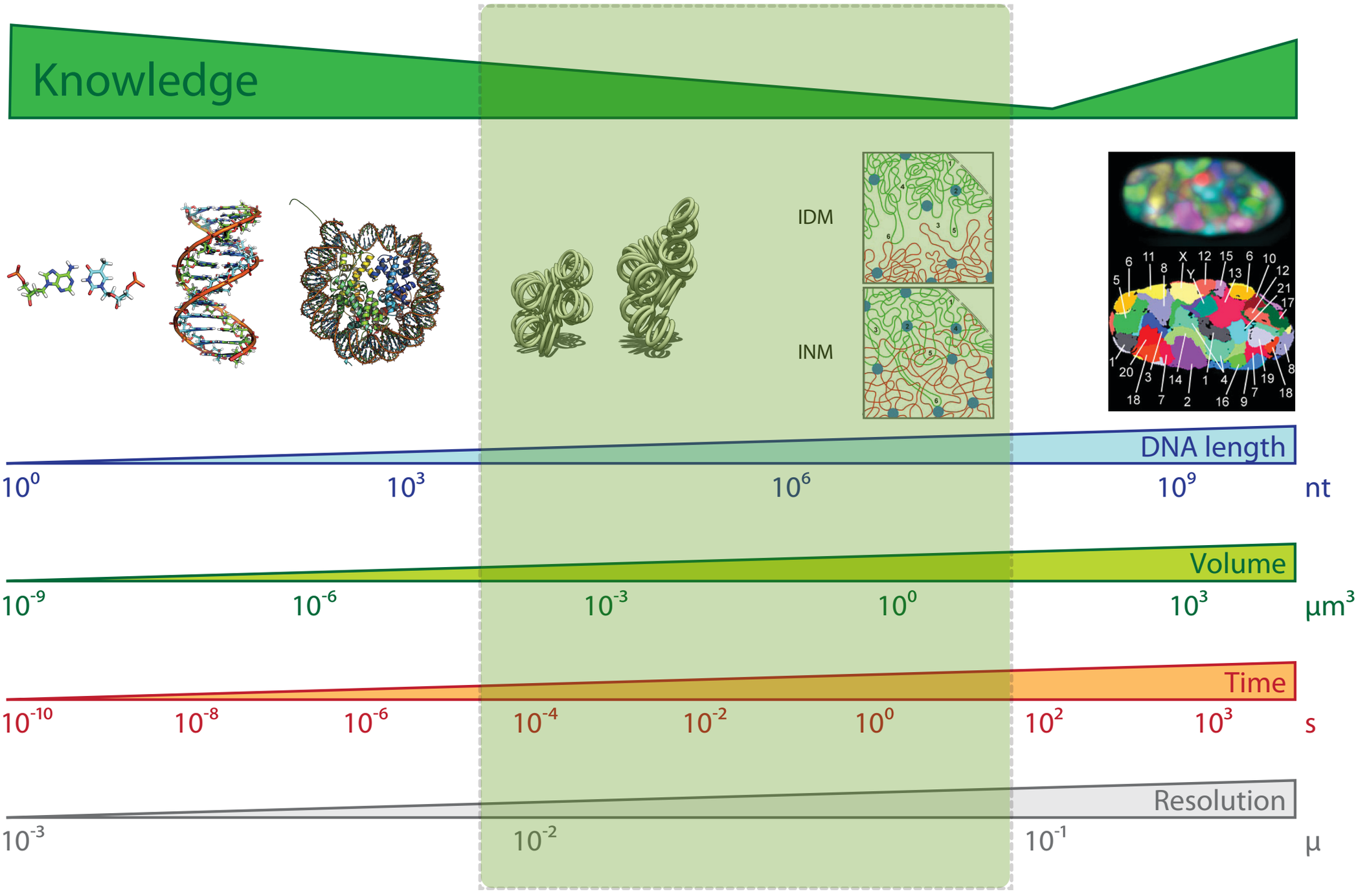
Structure determination of genomes and genomic domains by satisfaction of spatial restraints

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



Resolution Gap

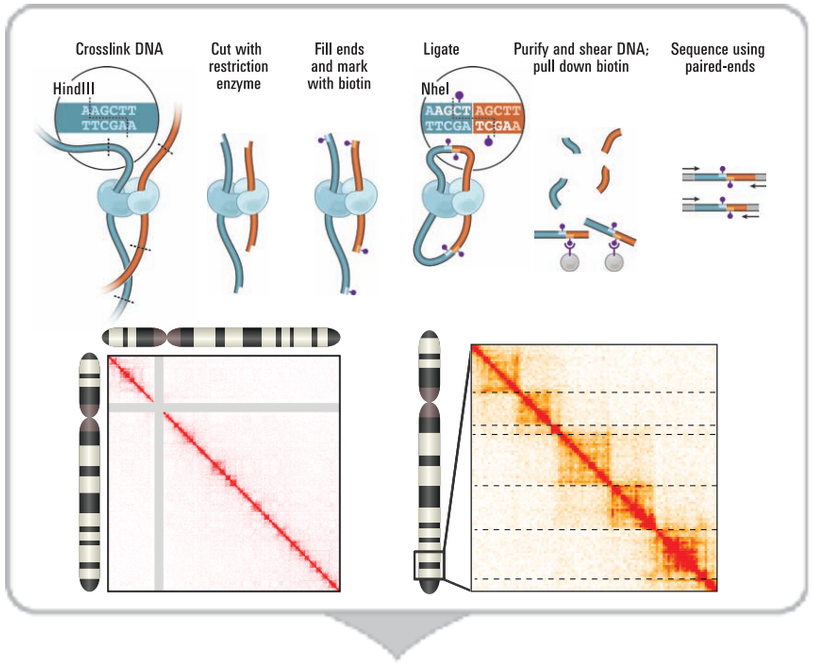
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol *7*, e1002125 (2011)

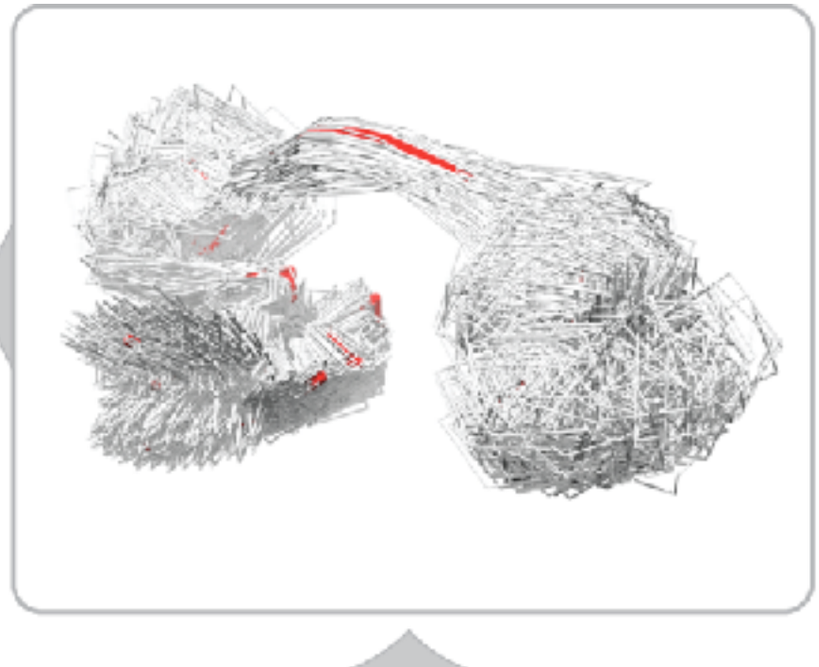


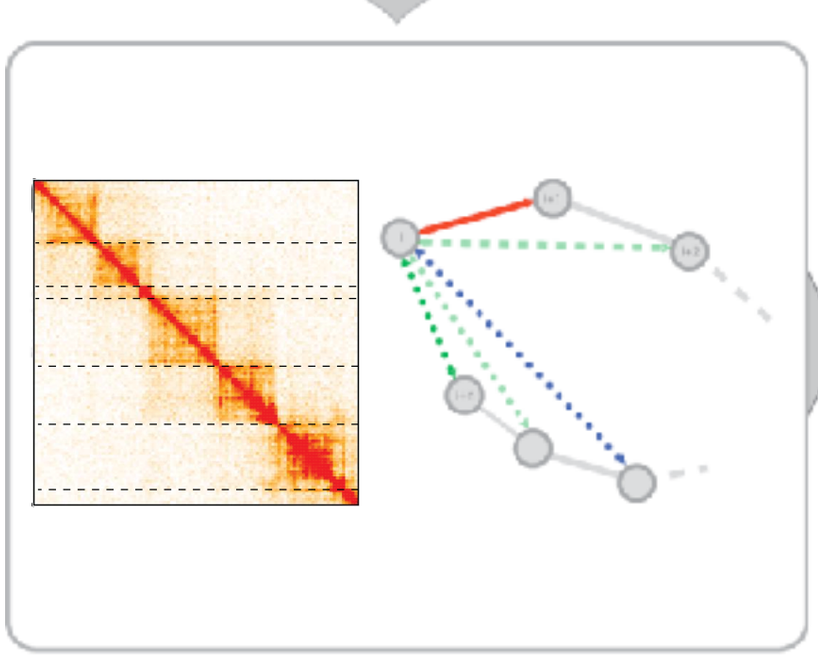
Hybrid Method

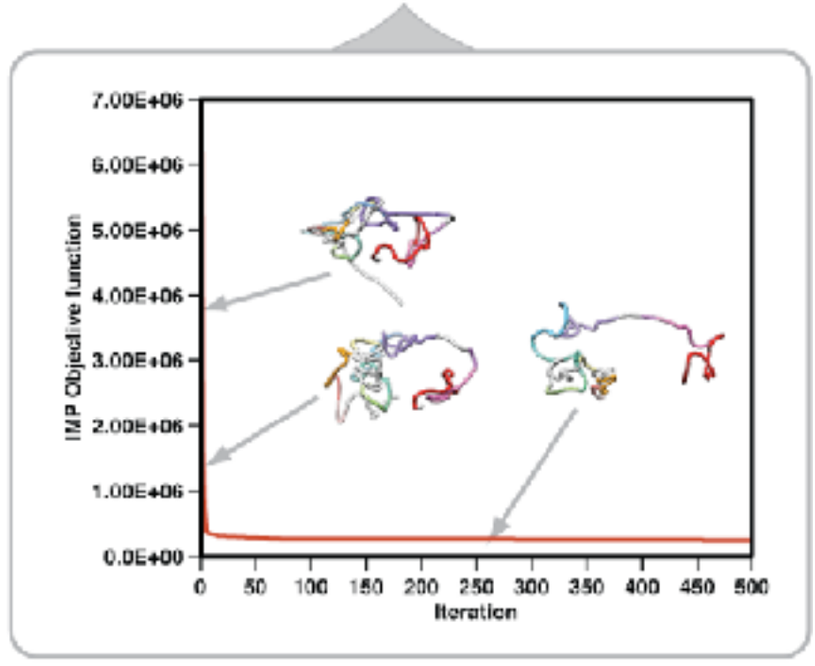
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).

Experiments





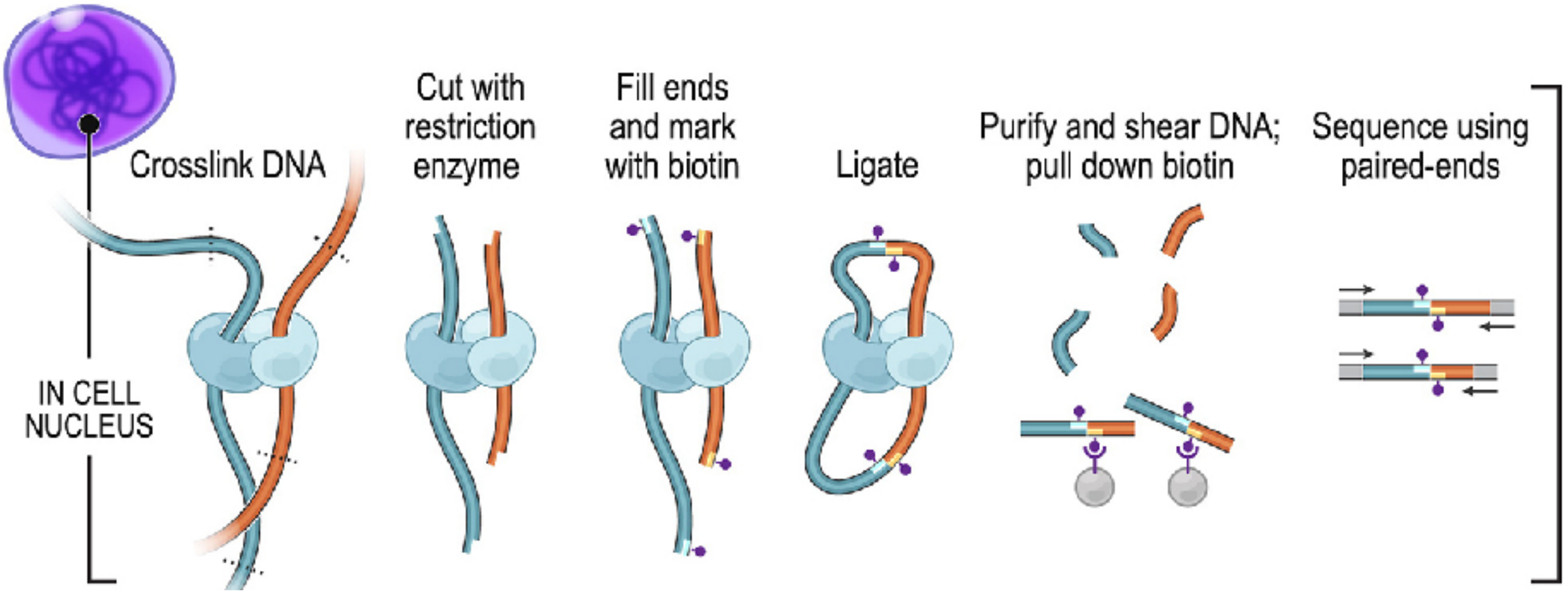




Computation

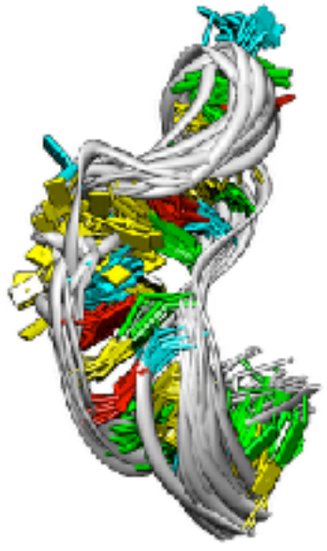
Chromosome Conformation Capture

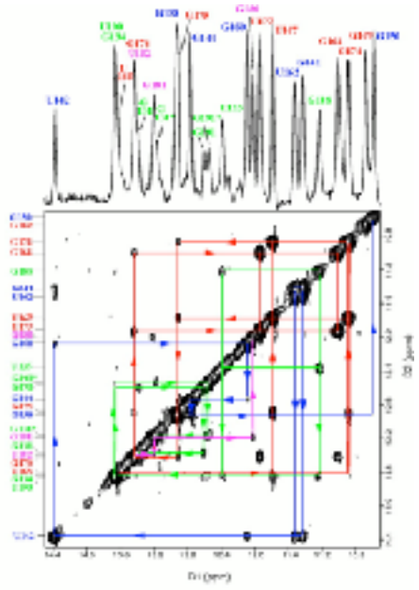
Dekker, J., Rippe, K., Dekker, M., & Kleckner, N. (2002). Science, 295(5558), 1306–1311.
Lieberman-Aiden, E., et al. (2009). Science, 326(5950), 289–293.



Restraint-based Modeling

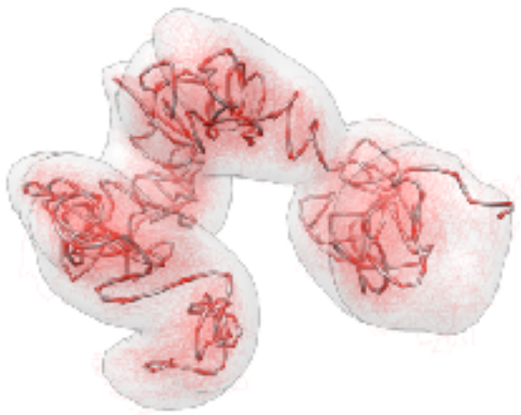
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).

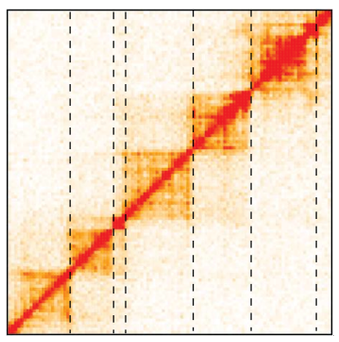




Biomolecular structure determination

2D-NOESY data



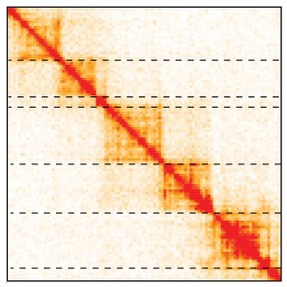


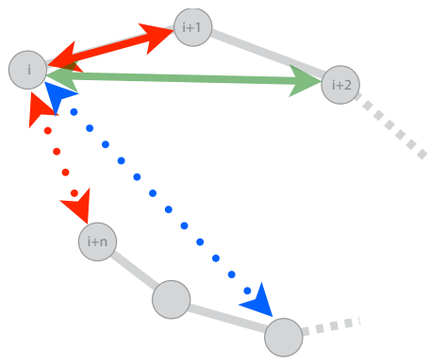
Chromosome structure determination

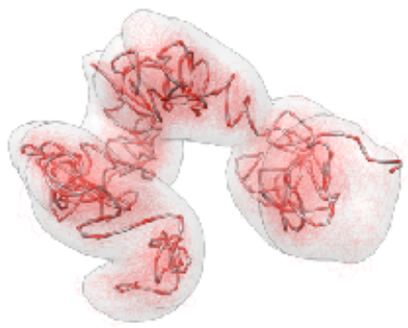
3C-based data

TADbit

http://3DGenomes.org







FastQ files to Maps

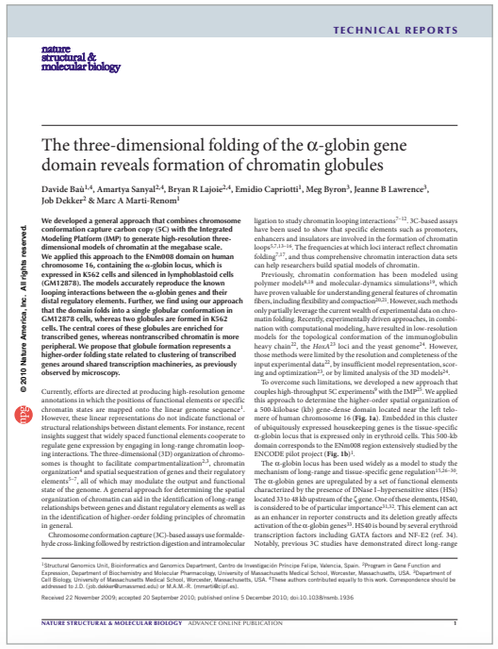
Map analysis

Model building

Model analysis

TADbit previous applications...

Baù, D. et al. Nat Struct Mol Biol (2011)
Umbarger, M. A. et al. Mol Cell (2011)



TECHNICAL REPORTS

nature structural & molecular biology

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

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

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 models of chromatin at the megabase scale. 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. We propose that globule formation represents a higher-order folding state related to clustering of transcribed genes around shared transcription machineries, as previously observed by microscopy.

© 2010 Nature America, Inc. All rights reserved.

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 instance, recent insights suggest that widely spaced functional elements cooperate to 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[7–12]. 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[5,7,13–16]. The frequencies at which loci interact reflect chromatin folding[7,17], and thus comprehensive chromatin interaction data sets can help researchers build spatial models of chromatin.

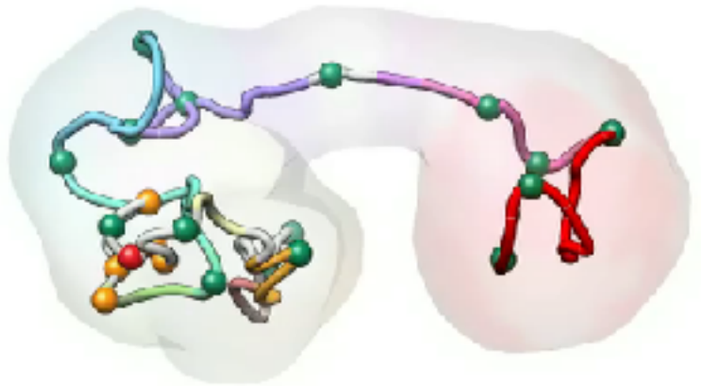
Previously, chromatin conformation has been modeled using polymer models[6,18] and molecular-dynamics simulations[19], which have proven valuable for understanding general features of chromatin fibers, including flexibility and compaction[20,21]. However, such methods only partially leverage the current 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[22], the *Hox*A[23] loci and the yeast genome[24]. However, those 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[24].

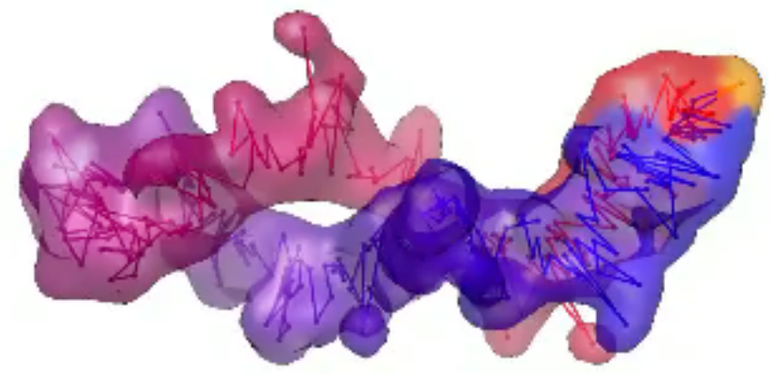
To overcome such limitations, we developed a new approach that couples high-throughput 5C experiments[9] with the IMP[25]. We applied this approach to determine the higher-order spatial organization of a 500-kilobase (kb) gene-dense domain located near the left telomere of human chromosome 16 (**Fig. 1a**). Embedded in this cluster of ubiquitously expressed housekeeping genes is the tissue-specific α-globin locus that is expressed only in erythroid cells. This 500-kb domain corresponds to the ENm008 region extensively studied by the ENCODE pilot project (**Fig. 1b**)[1].

The α-globin locus has been used widely as a model to study the mechanism of long-range and tissue-specific gene regulation[15,26–30]. 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 ζ gene. One of these elements, HS40, is considered to be of particular importance[31,32]. This element can act as an enhancer in reporter constructs and its deletion greatly affects activation of the α-globin genes[33]. 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@cipf.es).

Received 22 November 2009; accepted 20 September 2010; published online 5 December 2010; doi:10.1038/nsmb.1936





Cell PRESS

Molecular Cell
Article

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

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

[1]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[2]Department of Developmental Biology, School of Medicine
[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]Department of Biochemistry and Molecular Pharmacology
[7]Program in Systems Biology
University of Massachusetts Medical School, Worcester, MA 01605, USA
[6]These authors contributed equally to this work
*Correspondence: umbarger@post.harvard.edu (M.A.U.), mmarti@cipf.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.

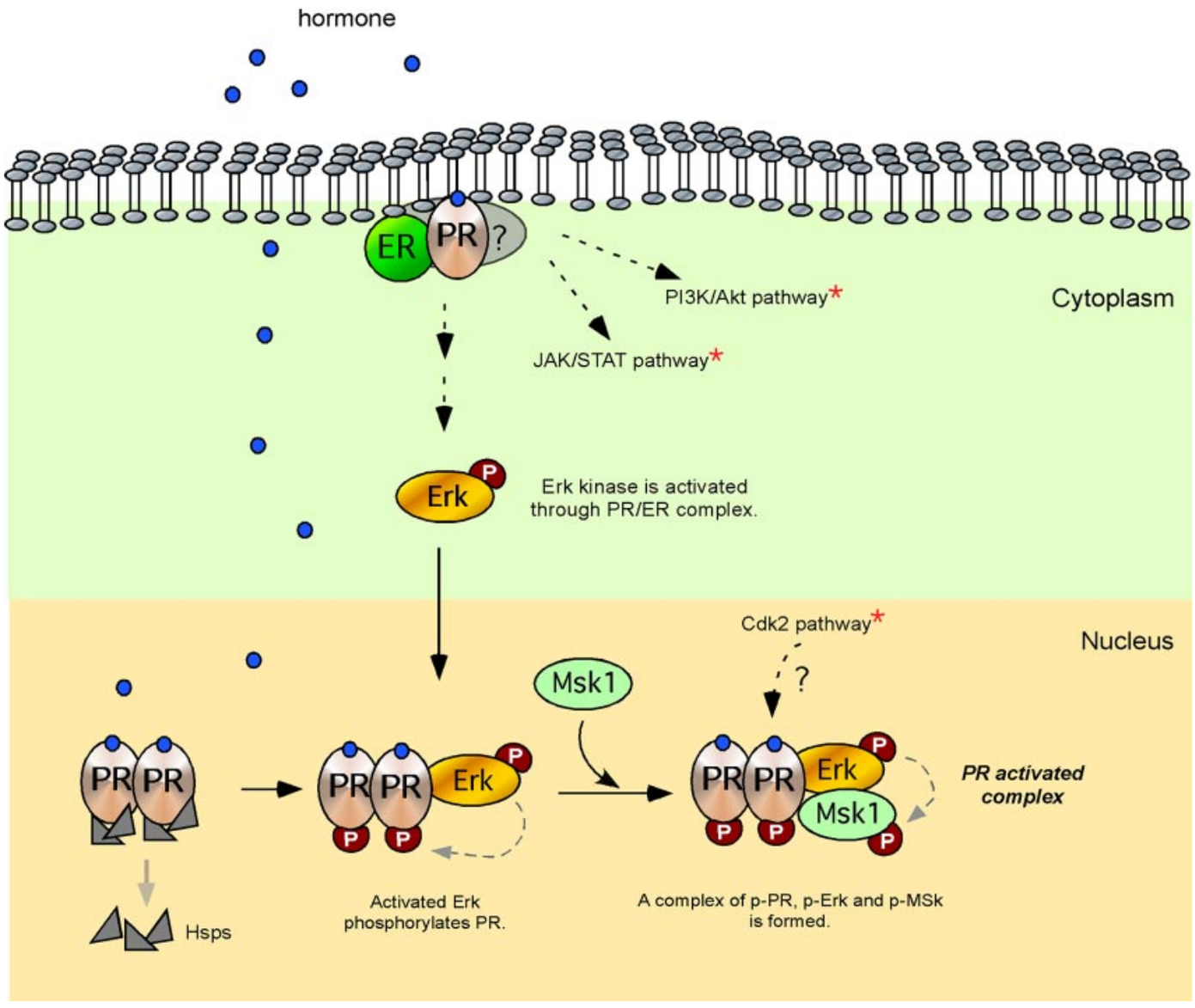
INTRODUCTION

The three-dimensional (3D) architecture of the genome both reflects and regulates its functional state (Dekker, 2008; Thanbichler and Shapiro, 2006a). For example, chromosome segregation impacts bacterial locus subcellular positioning (Jun and Mulder, 2006; White et al., 2008), and chromatin loops that place promoters and distant enhancers within close spatial proximity play important roles in eukaryotic transcriptional regulation (Tolhuis et al., 2002; Vernimmen et al., 2007). Such examples suggest that studies of the high-resolution folding of genomes will yield insight into genome biology. However, until recently such studies, which require comprehensive assessments of the spatial positioning of many loci, have represented major technical challenges.

The recent development of several high-throughput technologies, including automated fluorescent imaging (Viollier et al., 2004) and chromosome conformation capture (3C)-based approaches (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; Zhao et al., 2006), has begun to enable studies of genome-wide chromosome folding. Fluorescent microscopy-based approaches allow the accurate determination of the subcellular positions of increasing numbers of defined chromosomal loci, while high-throughput 3C-based approaches enable quantification of interloci interaction frequencies that can subsequently be used to infer the average 3D distances between these loci. Studies utilizing one or both of these approaches have highlighted the potential of genome-wide studies of chromosome structure and have begun to reveal specific features of chromosome folding, including the transcription-based compartmentalization of the human nucleus (Lieberman-Aiden et al., 2009; Simonis et al., 2006) and the correlation between a locus' genomic and subcellular positioning in bacteria (Nielsen et al., 2006; Teleman et al., 1998; Wang et al., 2006b). 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 utilize the resulting structure to identify the sequence 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

Progesterone-regulated transcription in breast cancer



Vicent *et al* 2011, Wright *et al* 2012, Ballare *et al* 2012

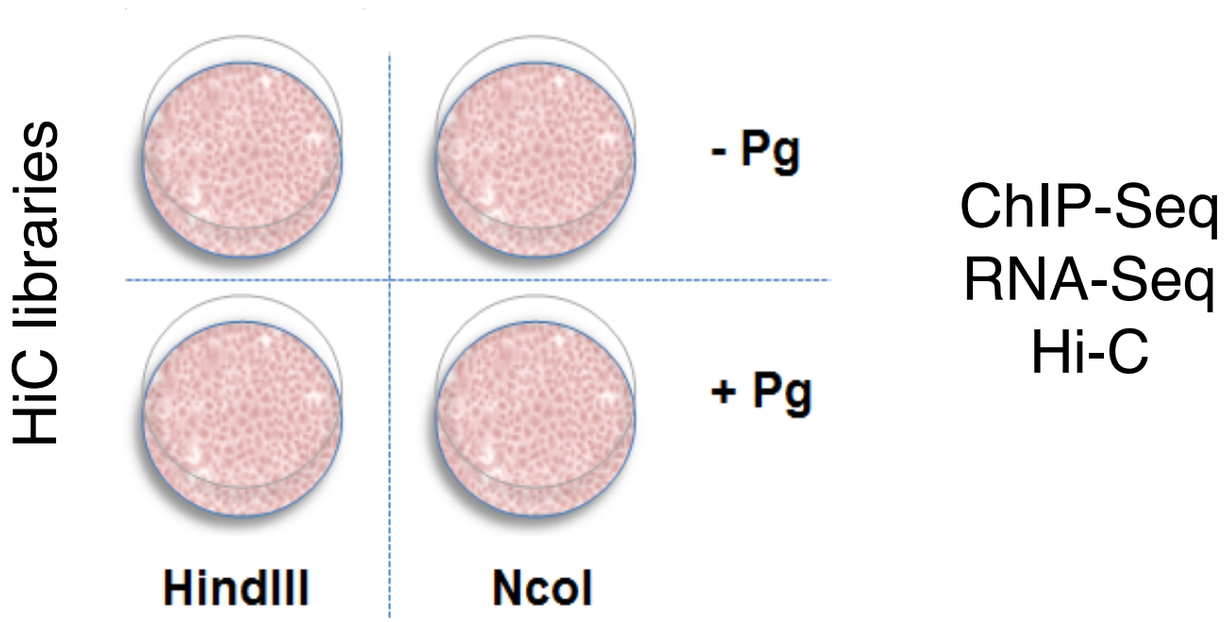
> 2,000 genes Up-regulated

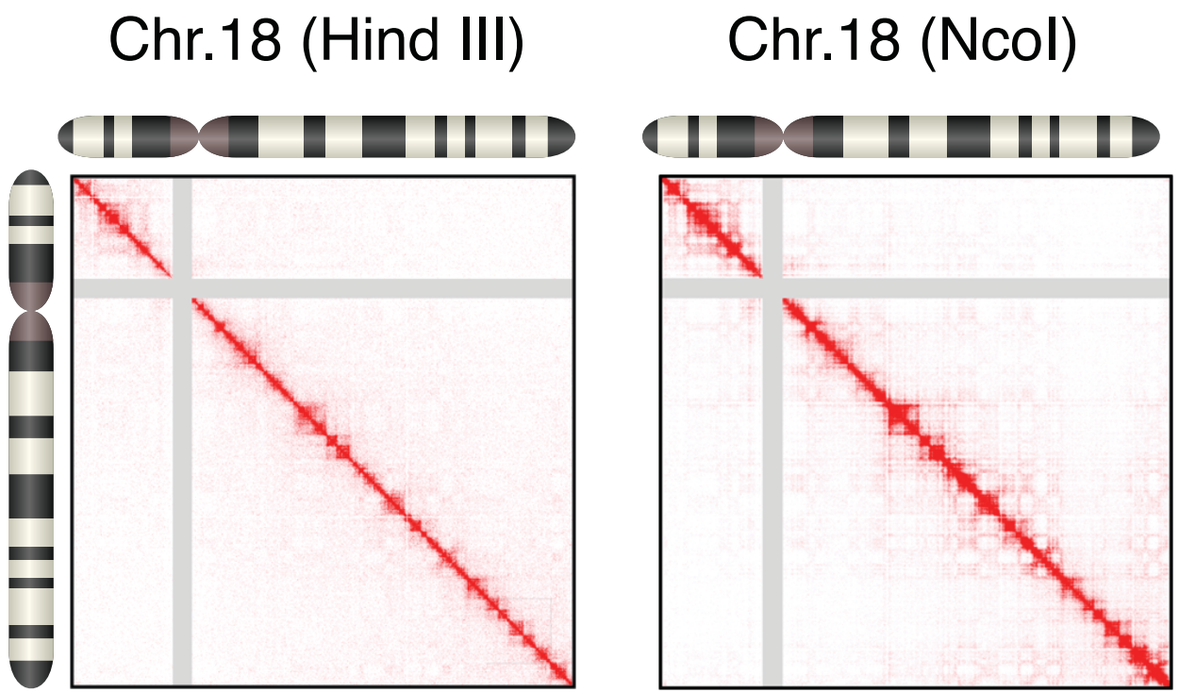
> 2,000 genes Down-regulated

Regulation in 3D?

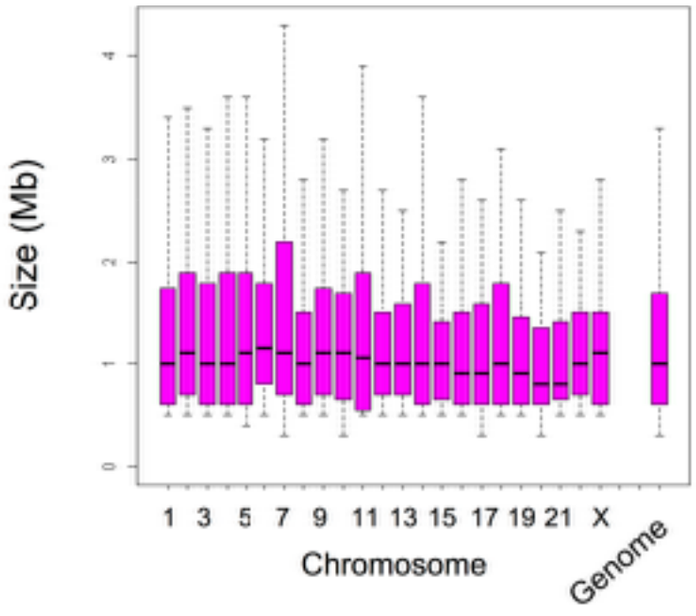
Le Dily, F. et al. Genes & Dev (2014)

Experimental design

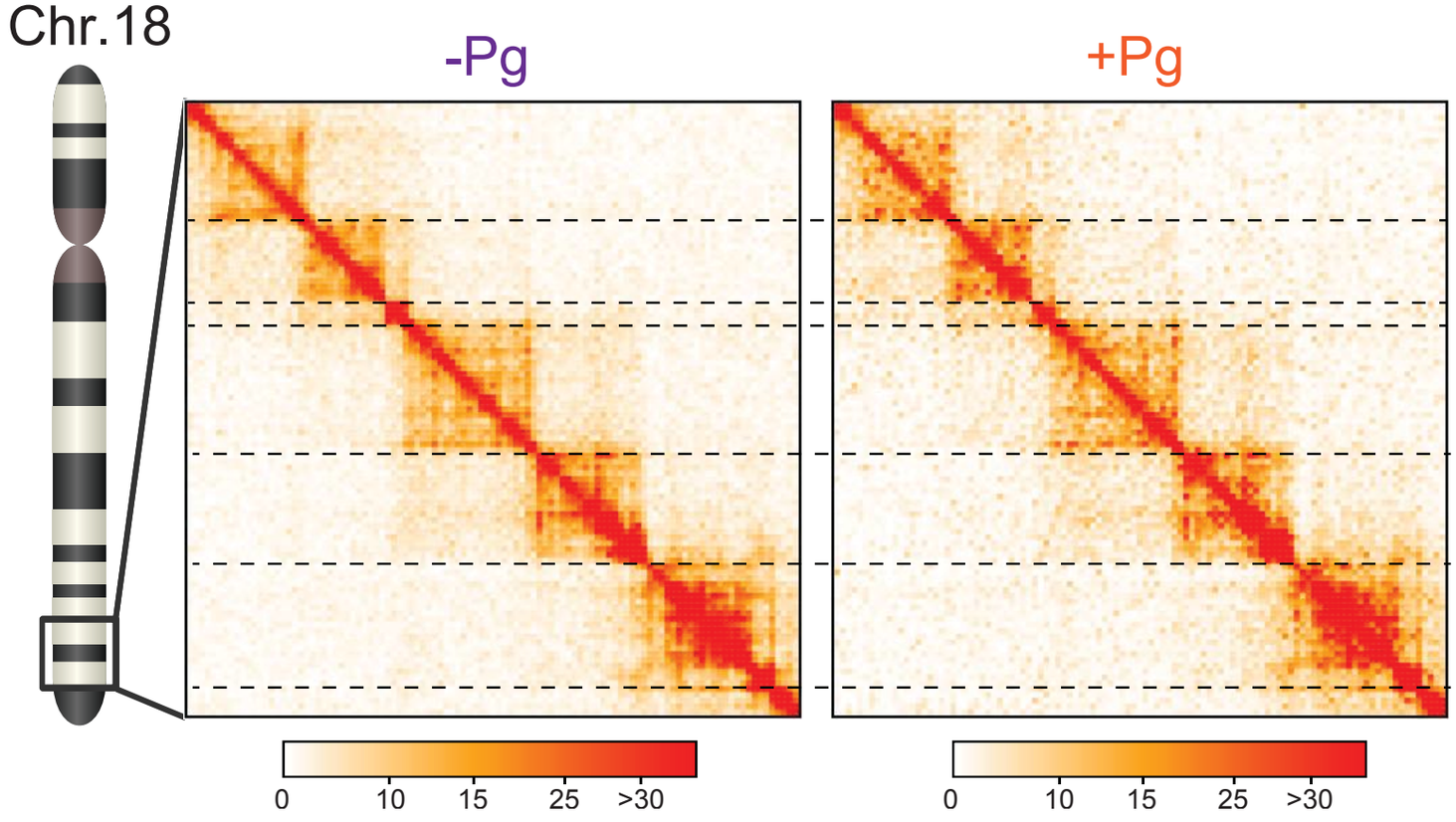


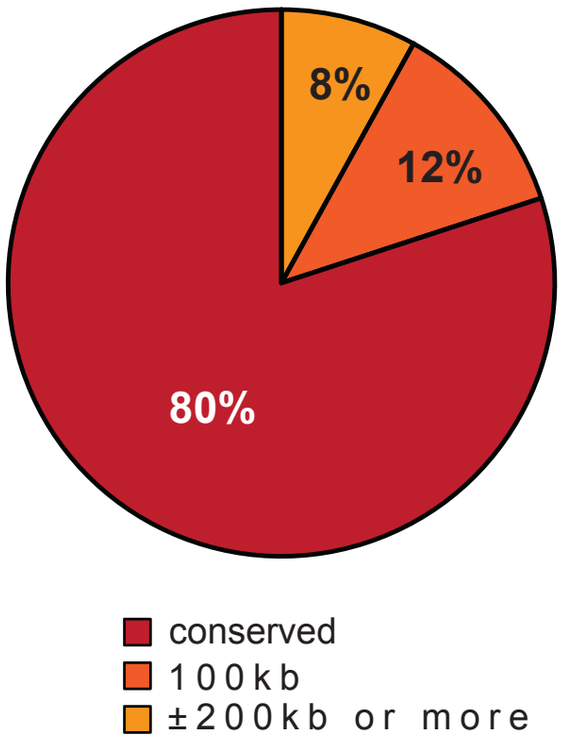


Are there TADs? how robust?

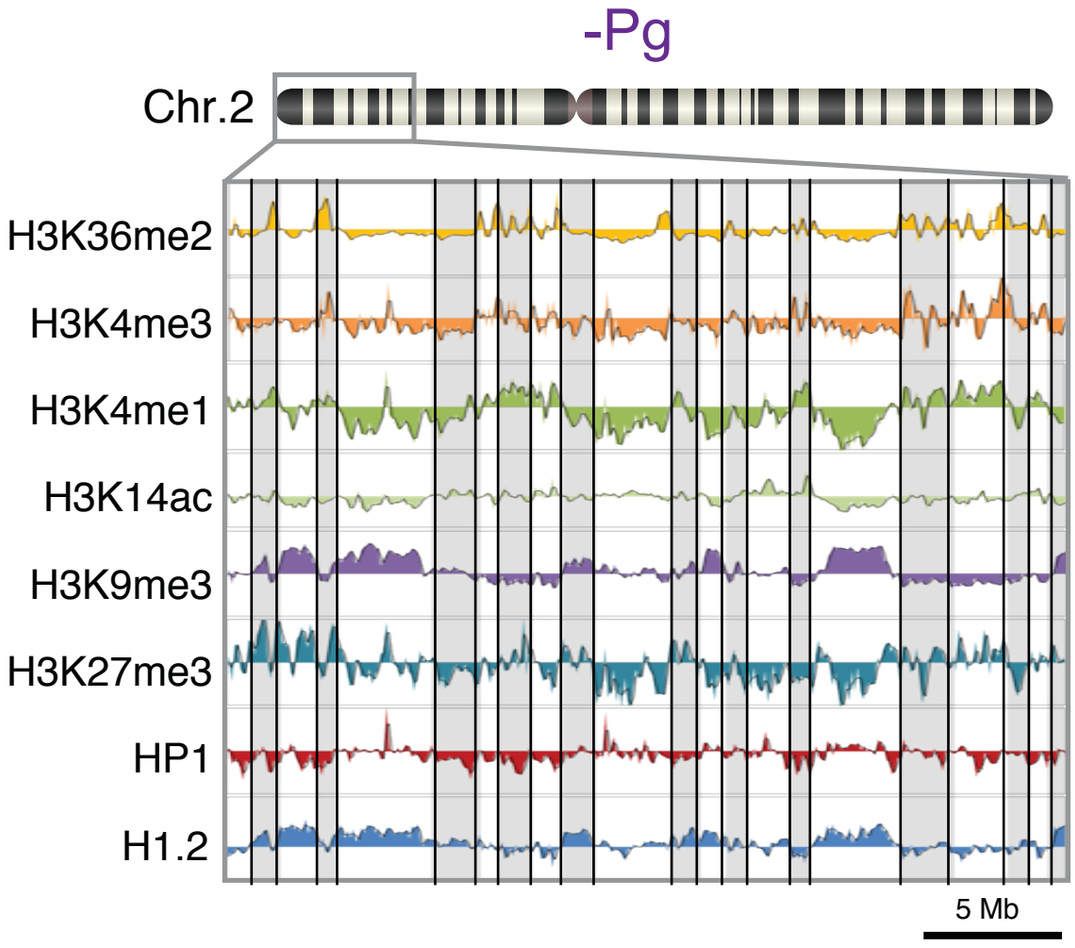


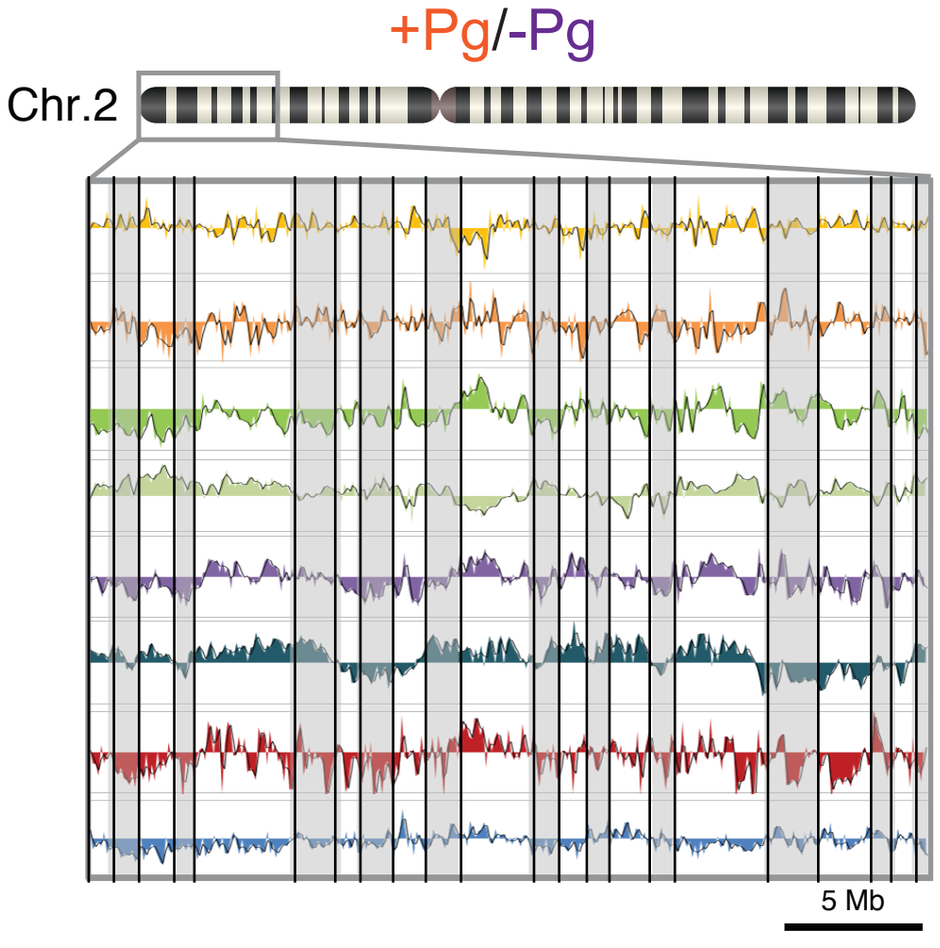
>2,000 detected TADs

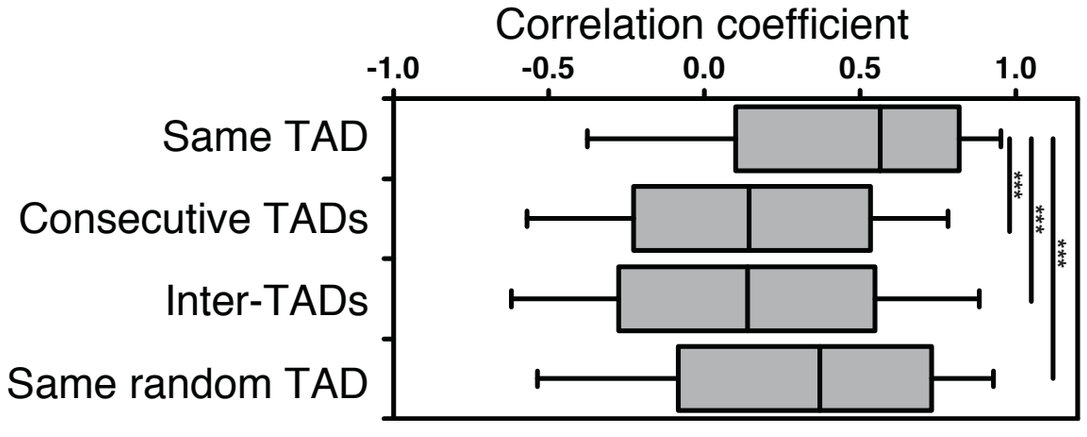


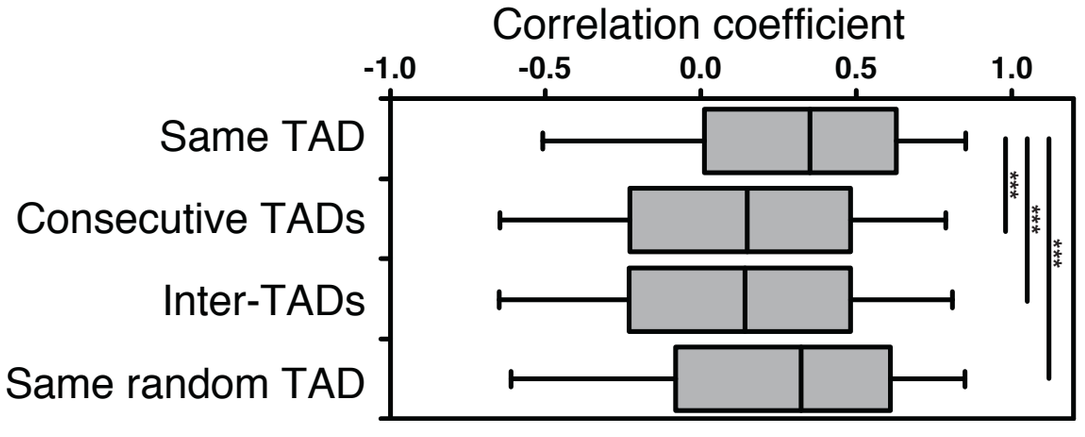


Are TADs homogeneous?

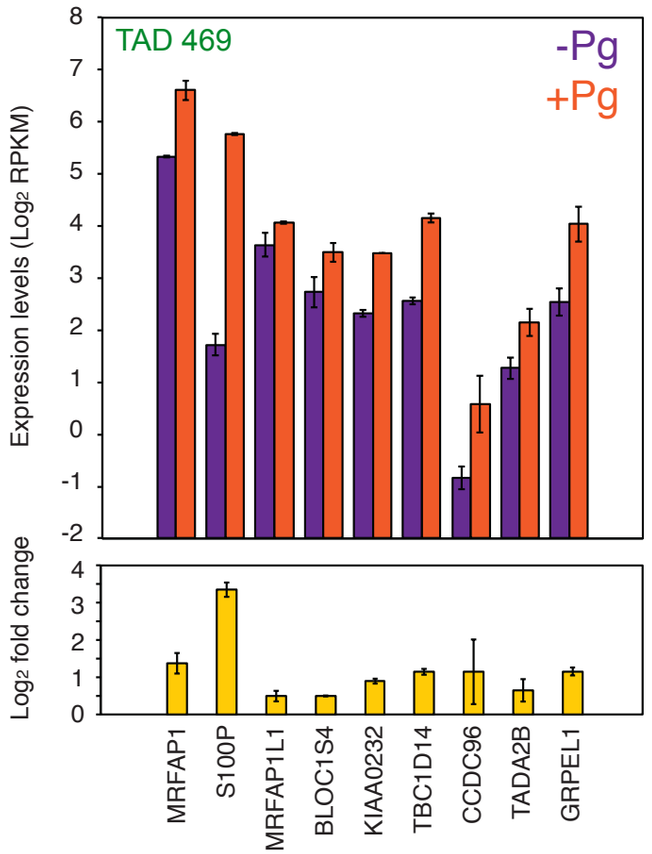


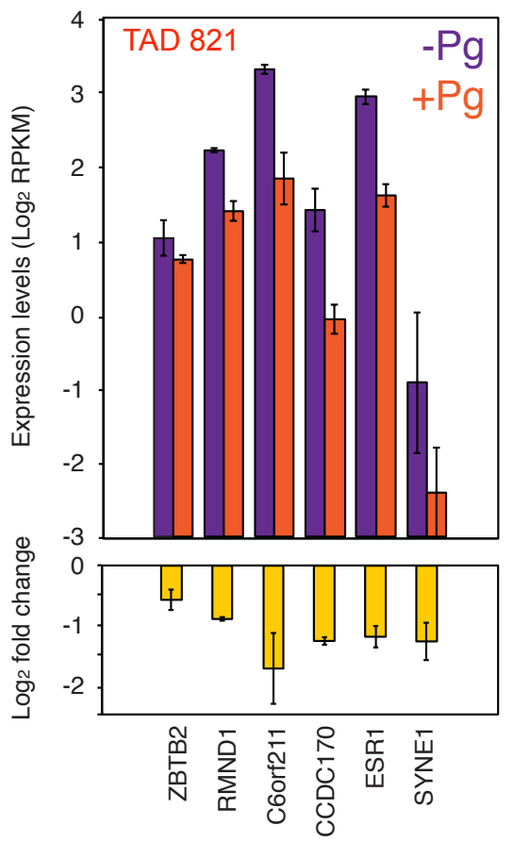


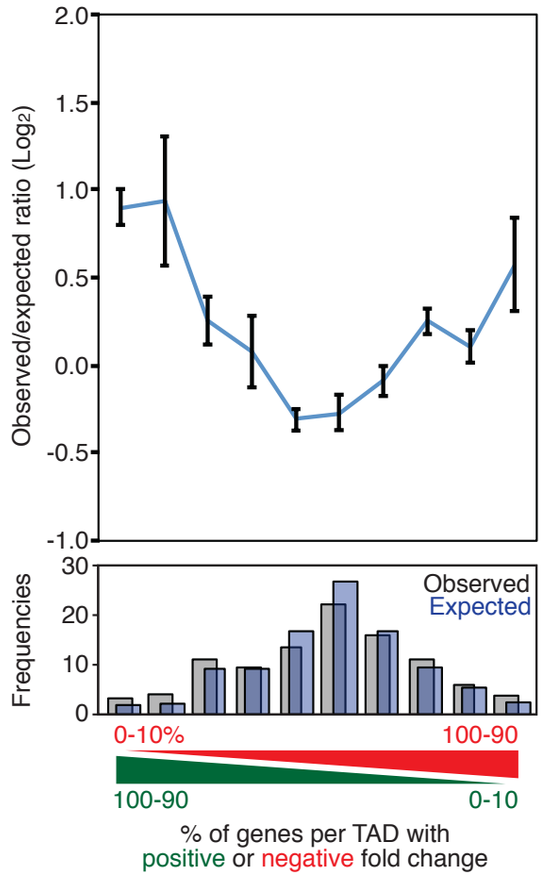




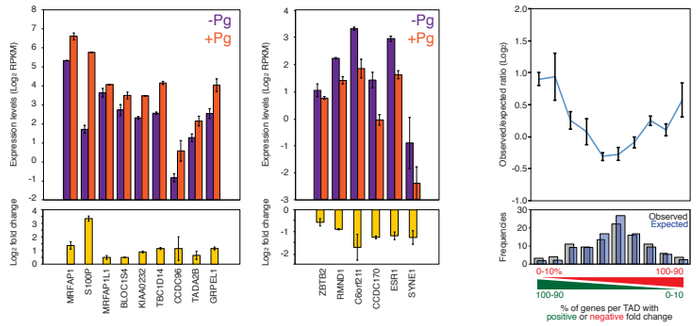
Do TADs respond differently to Pg treatment?





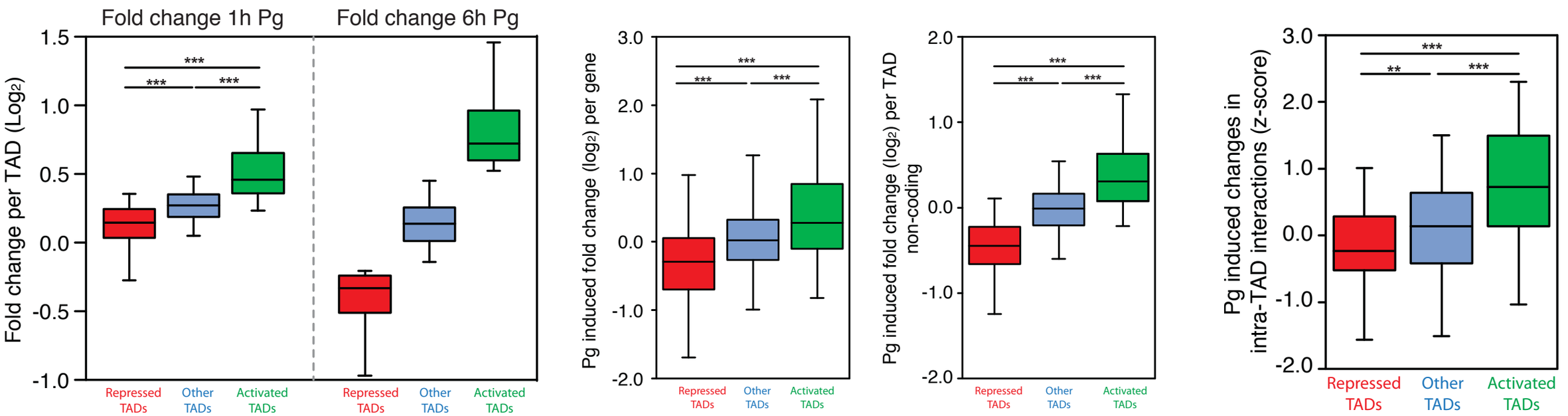


Do TADs respond differently to Pg treatment?

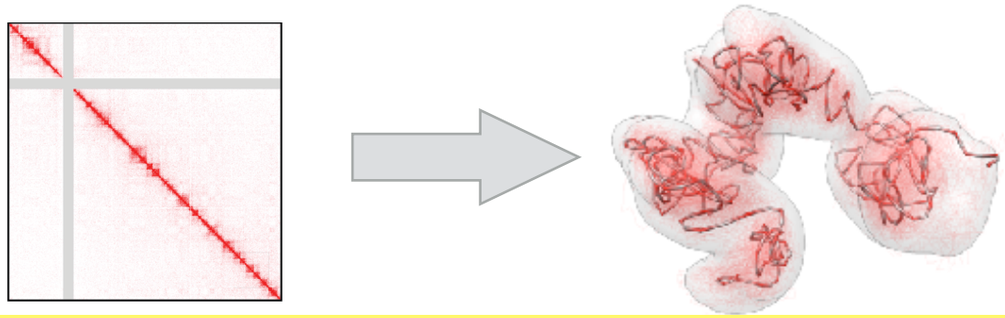


Pg induced fold change per TAD (6h)

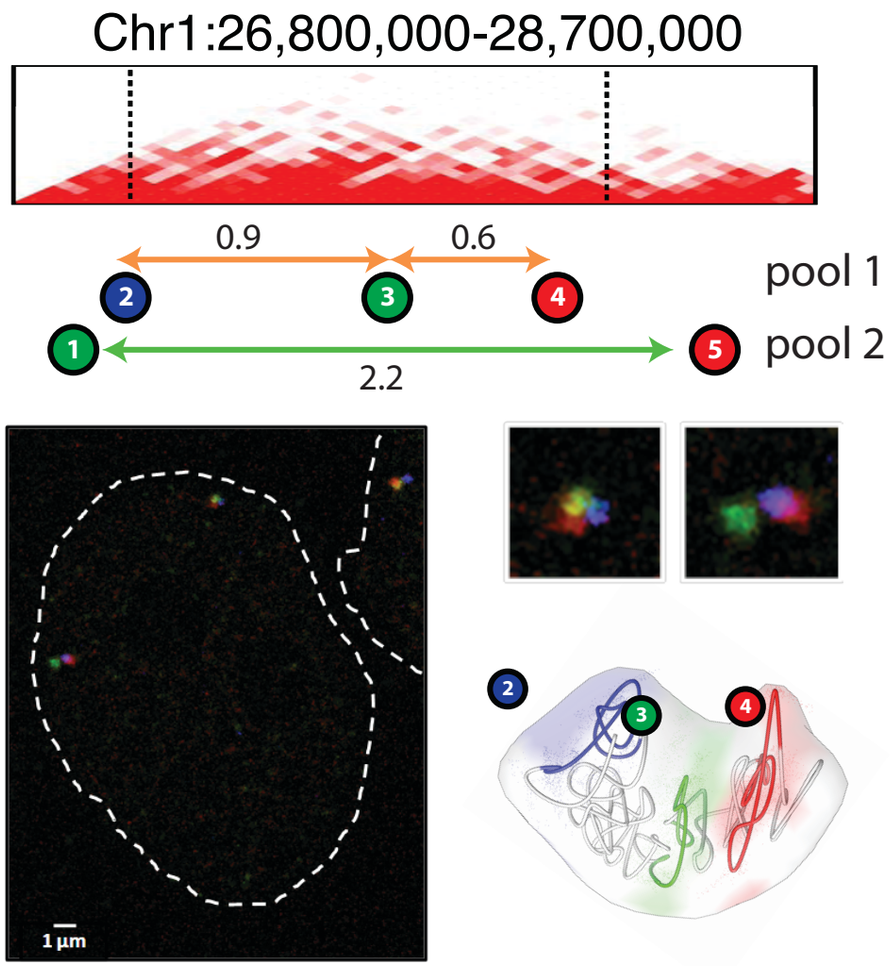
Repressed TADs | Other TADs | Activated TADs

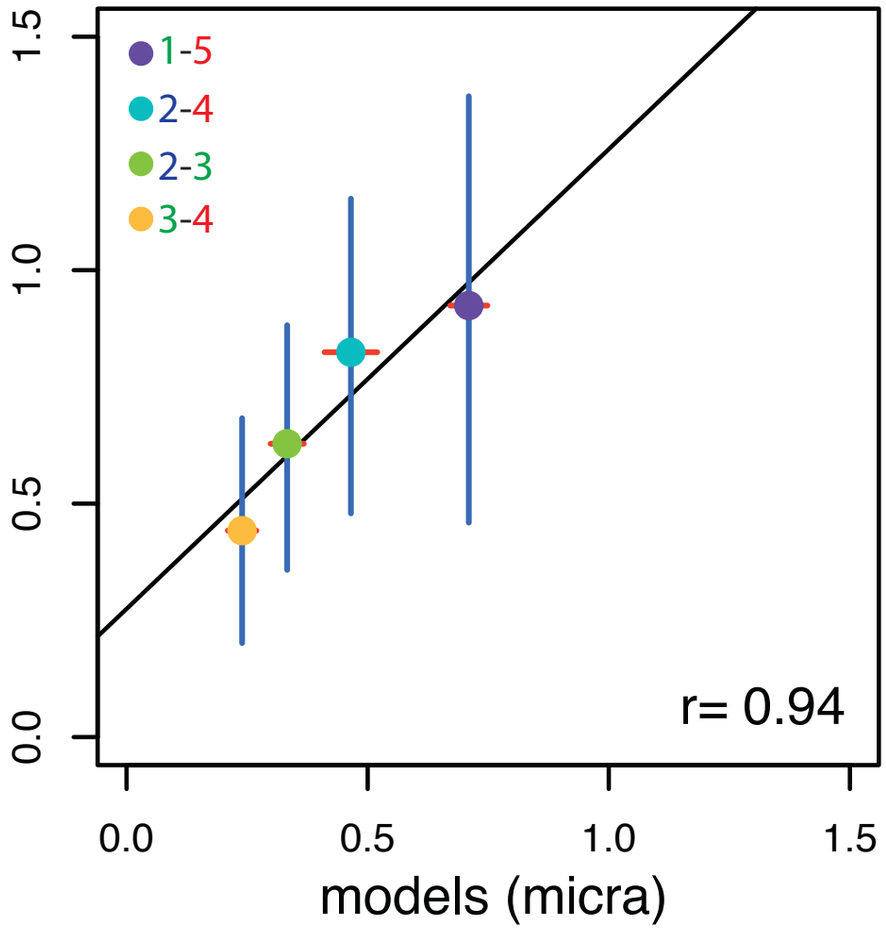


Modeling 3D TADs



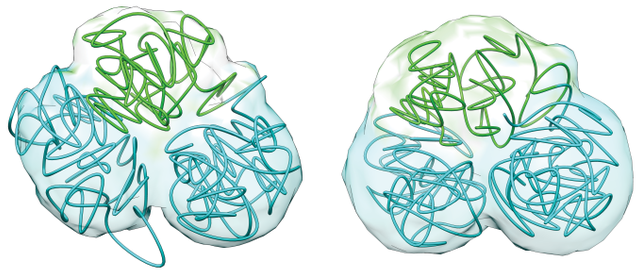
61 genomic regions containing 209 TADs covering 267Mb

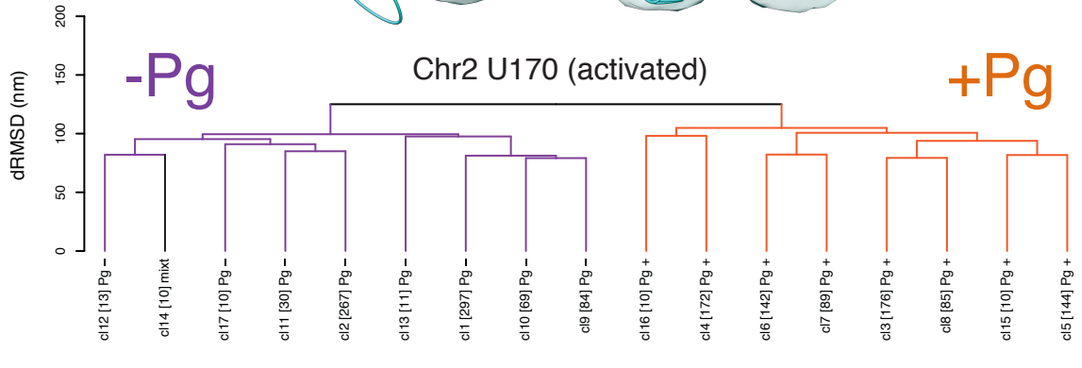




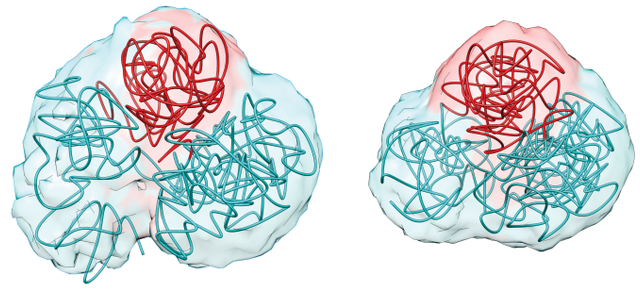
How TADs respond structurally to Pg?

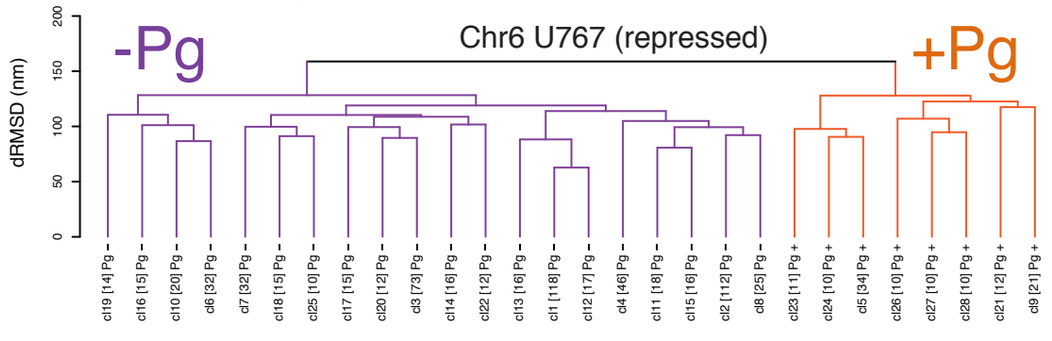
Chr2:9,600,000-13,200,000

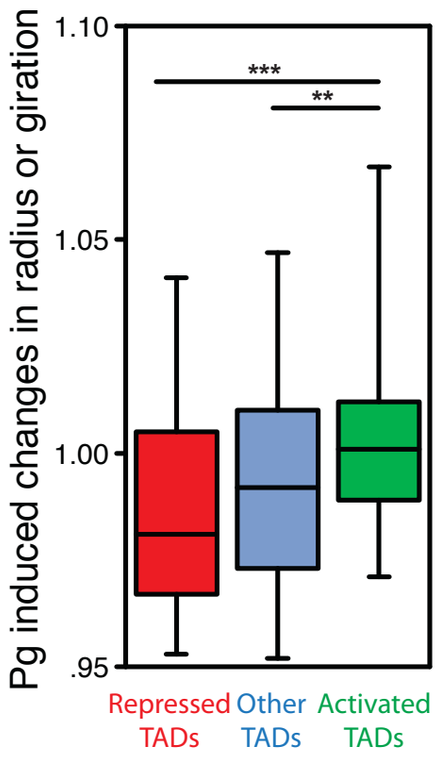


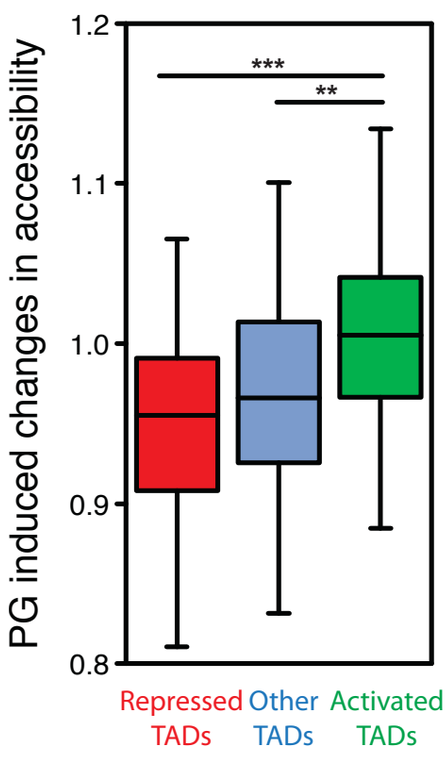


Chr6:71,800,000-76,500,000

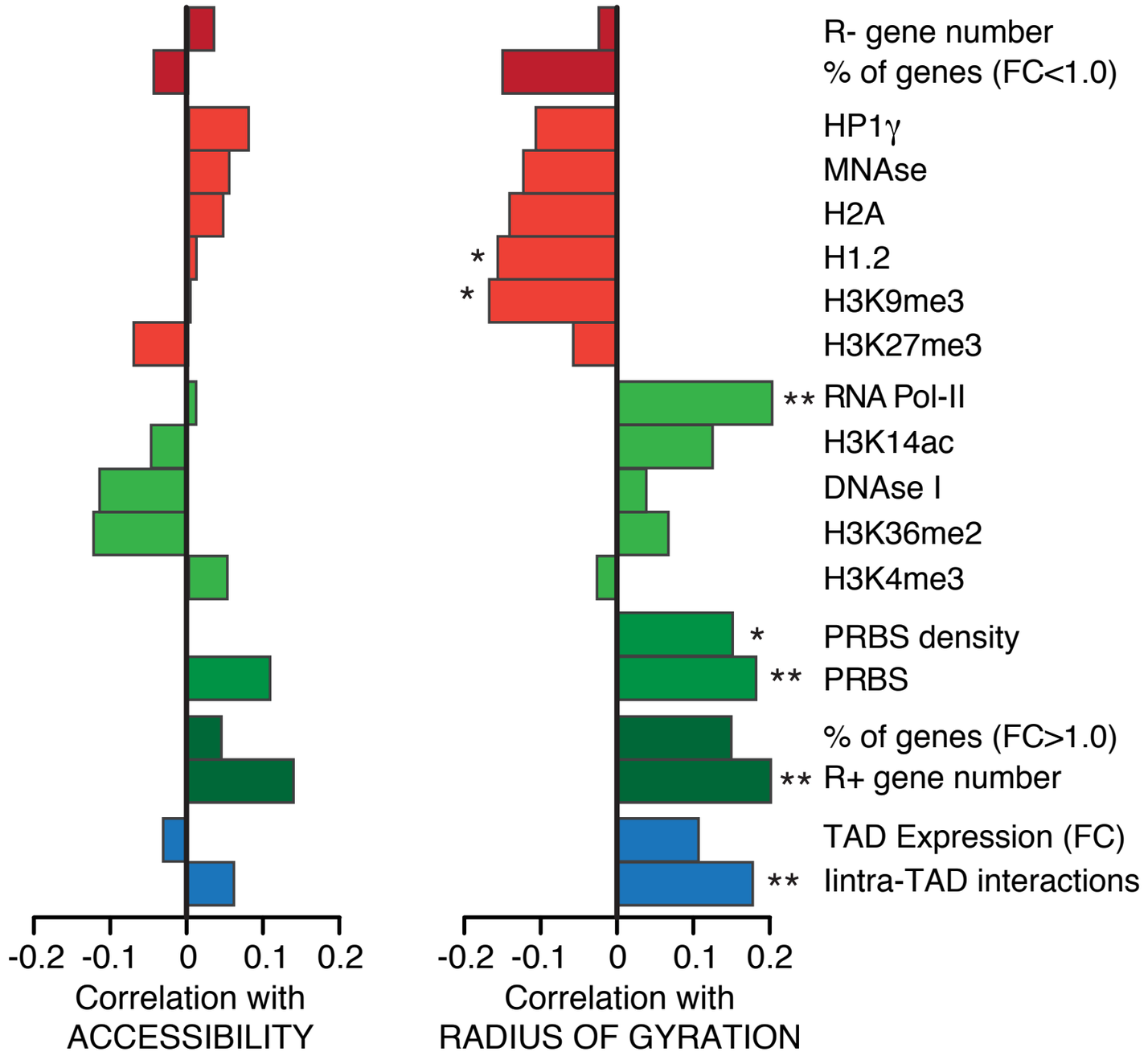




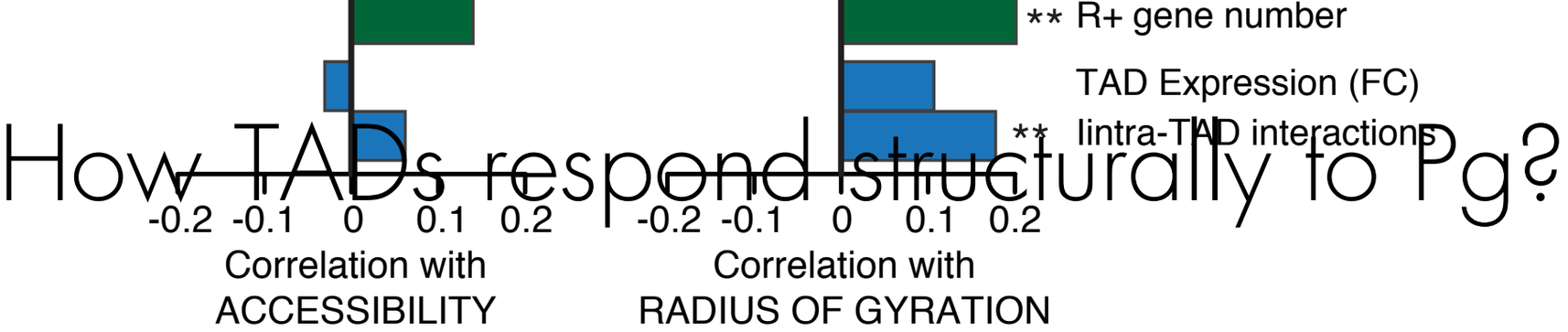


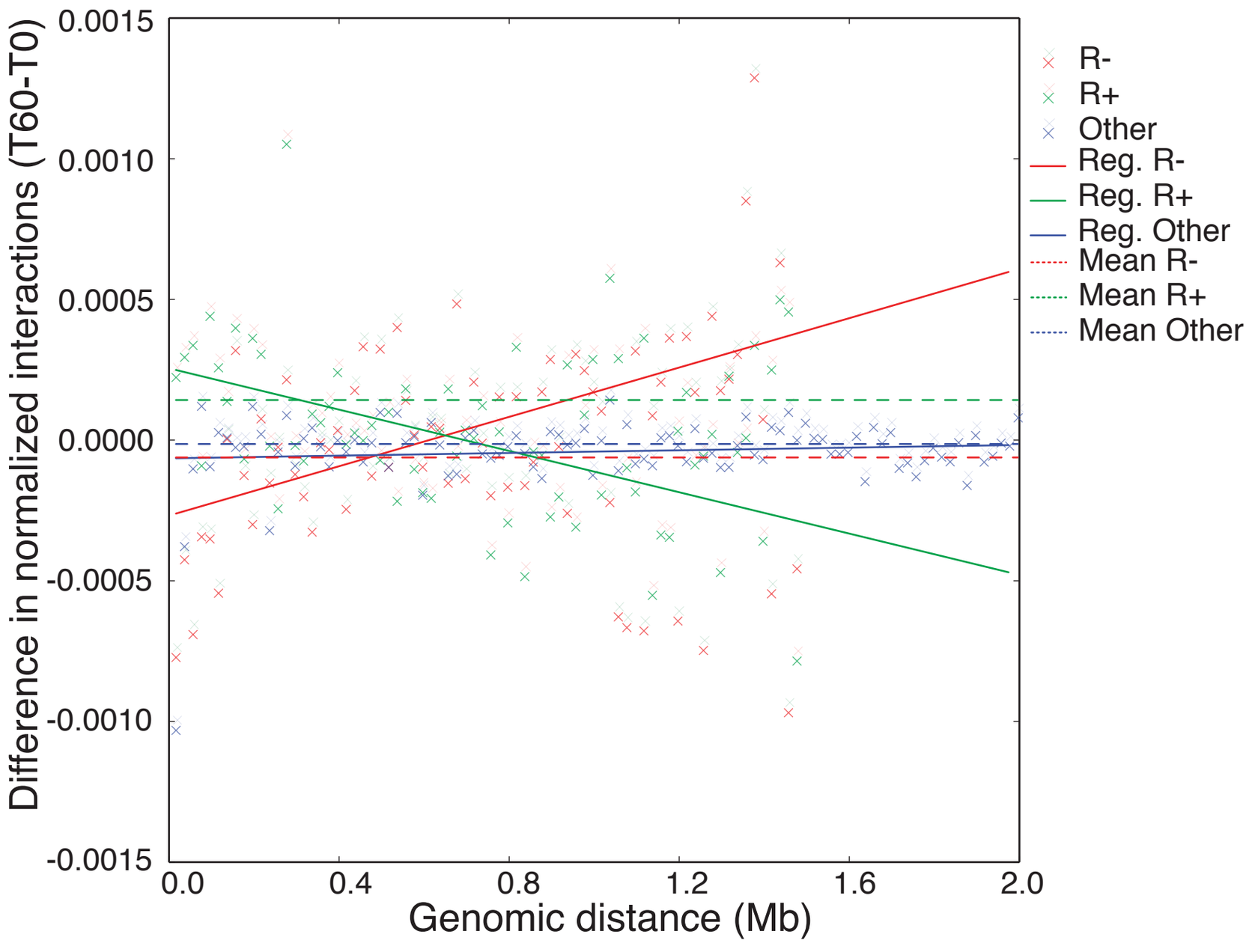


How TADs respond structurally to Pg?



How TADs respond structurally to Pg?

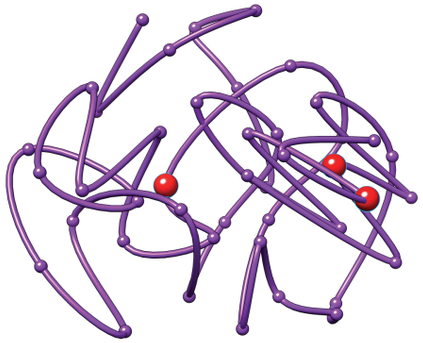


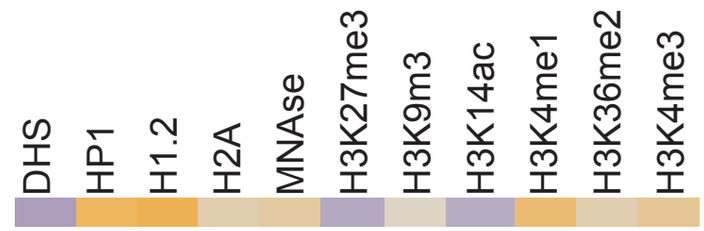


Model for TAD regulation

Repressed TAD

chr1 U41

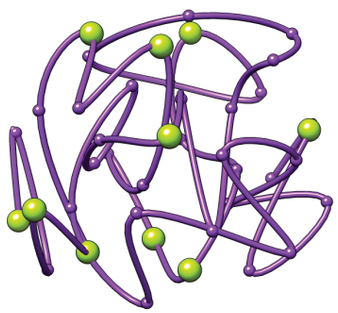


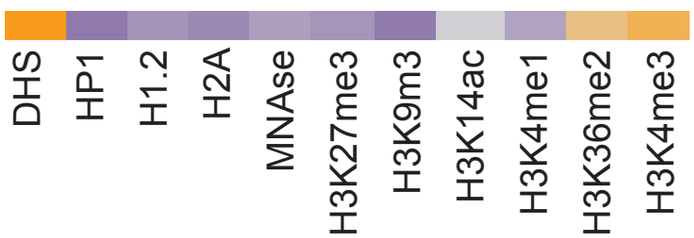


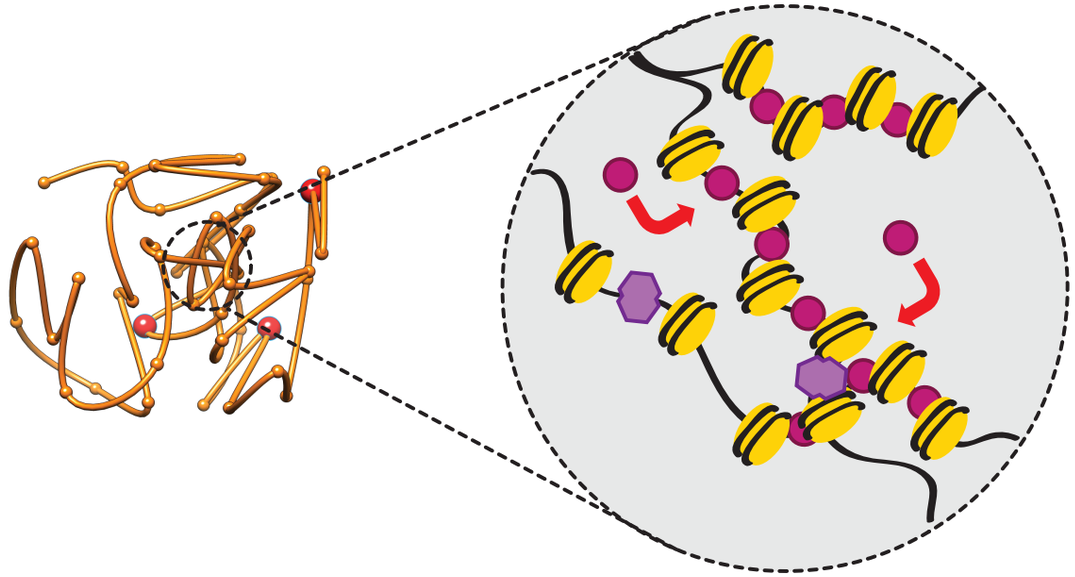
Structural transition **+Pg**

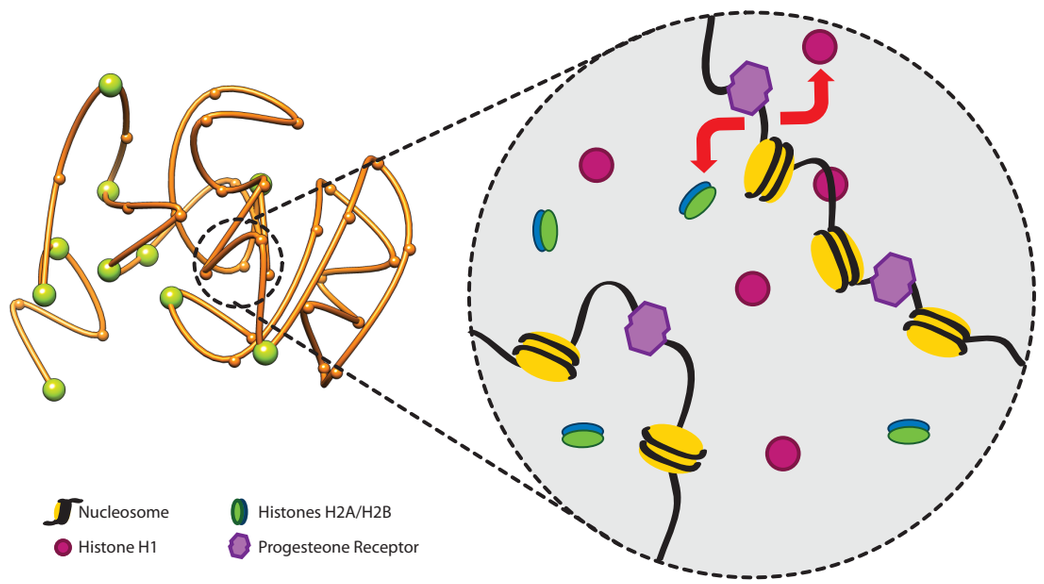
Activated TAD

chr2 U207









..:: Open POSTDOC position to link SNP and 3D structure in disease ::..

François le Dily
François Serra
Davide Baù

Gireesh K. Bogu
Yasmina Cuartero
David Dufour
Irene Farabella
Silvia Galan
Mike Goodstadt
Paula Soler
Yannick Spill
Marco di Stefano

4DGenome Unit - Miguel Beato - Thomas Graf - Guillaume Filion

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