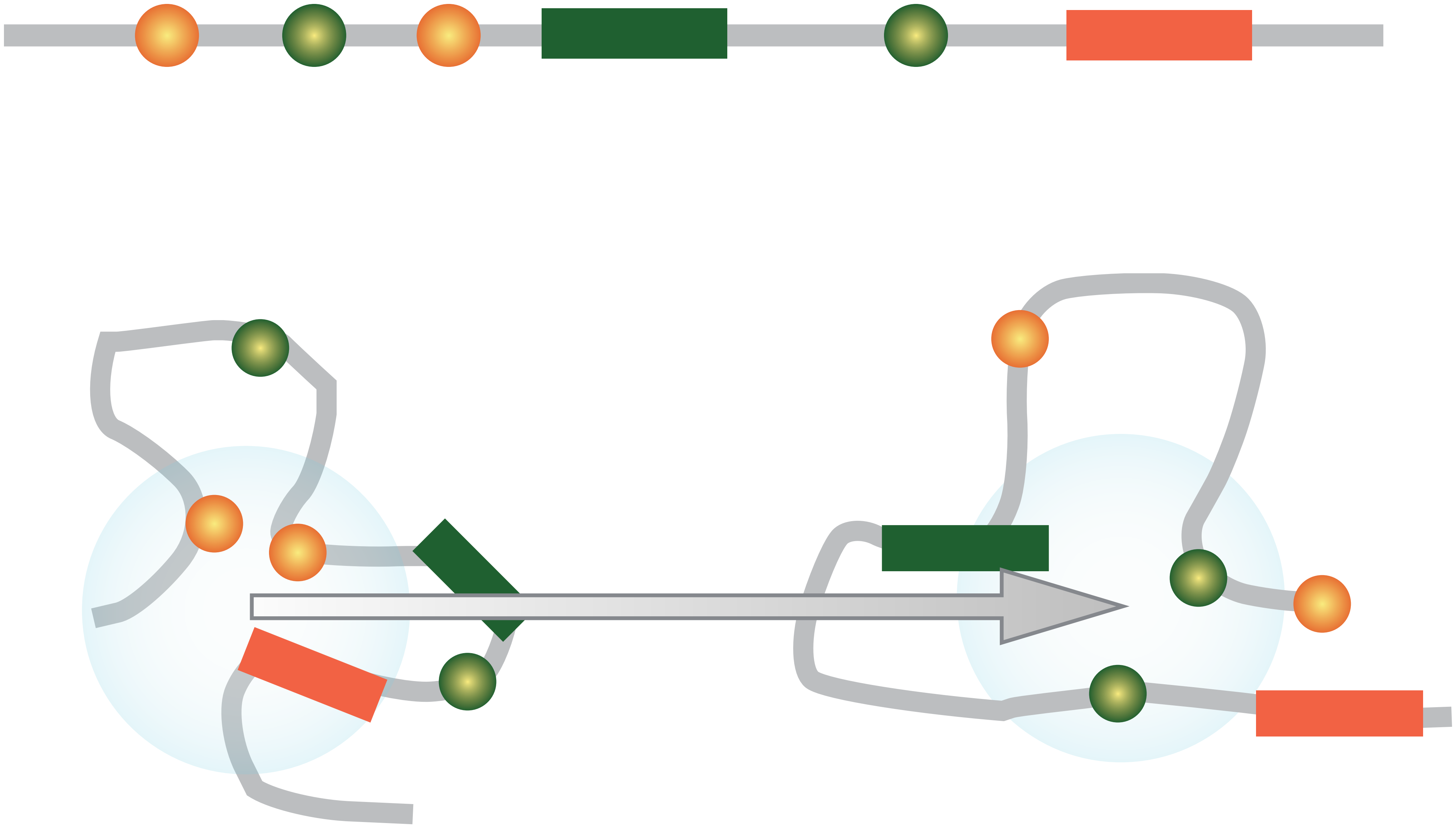
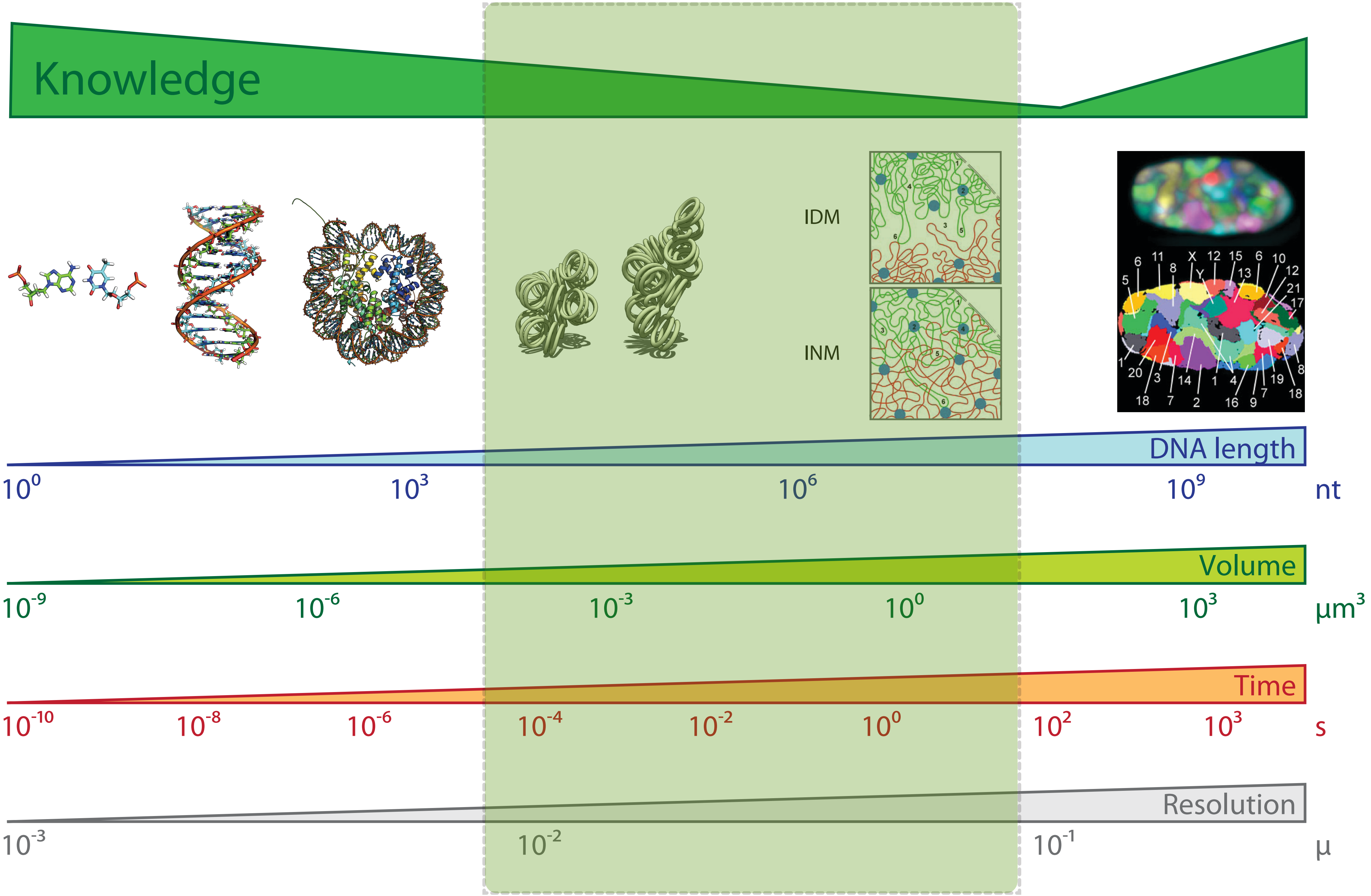
Structure determination of genomes and genomic domains by satisfaction of spatial restraints

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



Resolution Gap

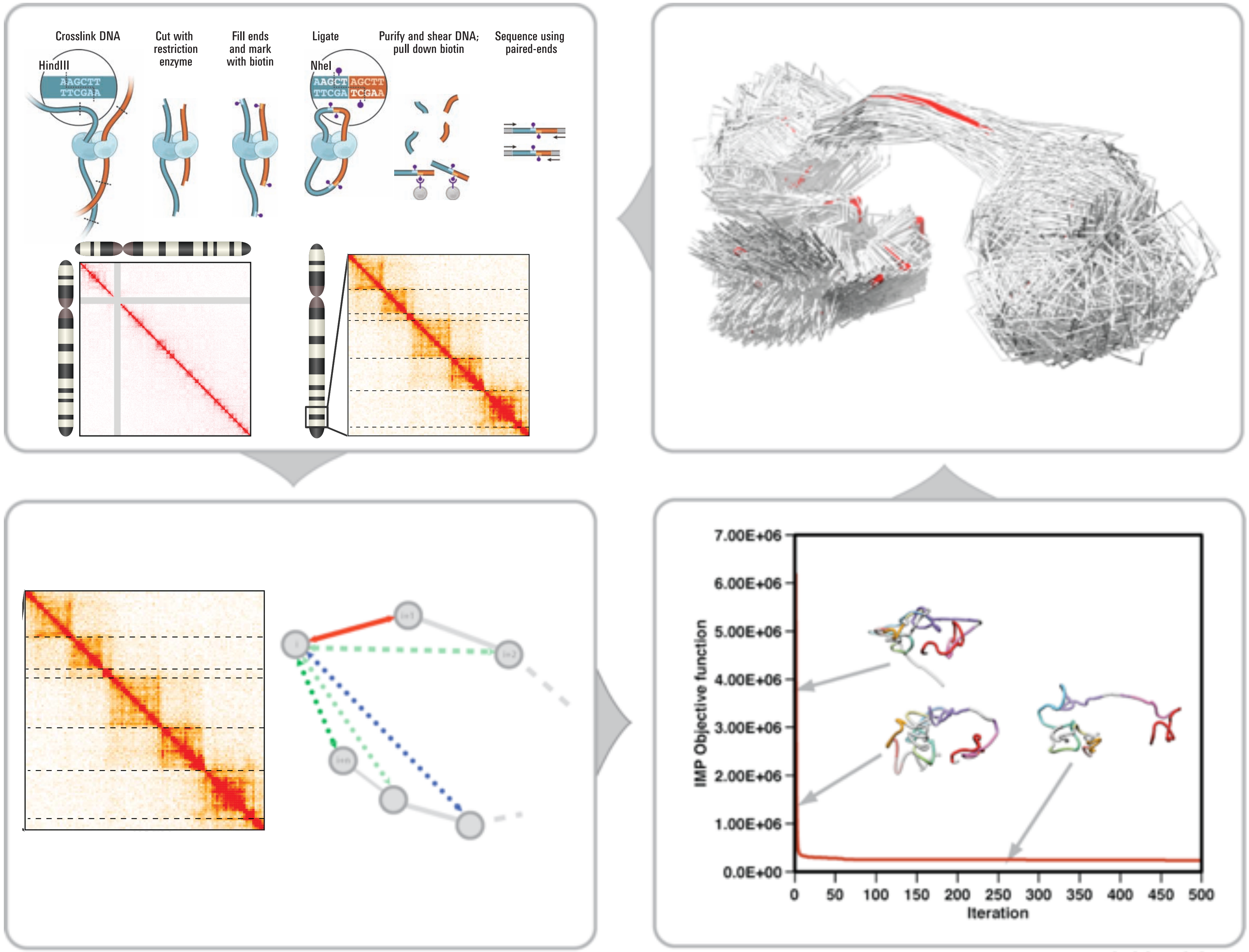
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol 7, e1002125 (2011)



Hybrid Method

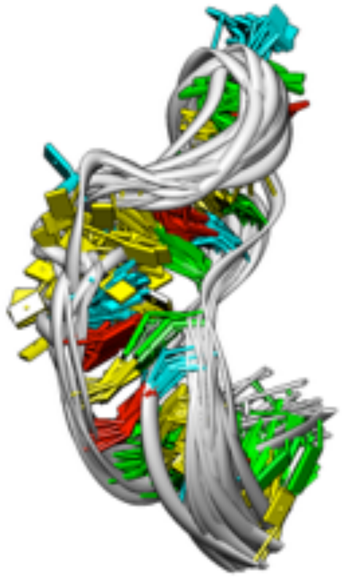
Baù, D. & Marti-Renom, M. A. Methods 58, 300–306 (2012).

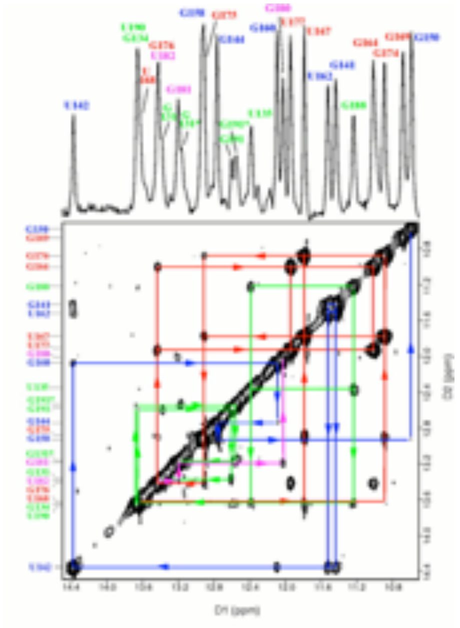
Experiments



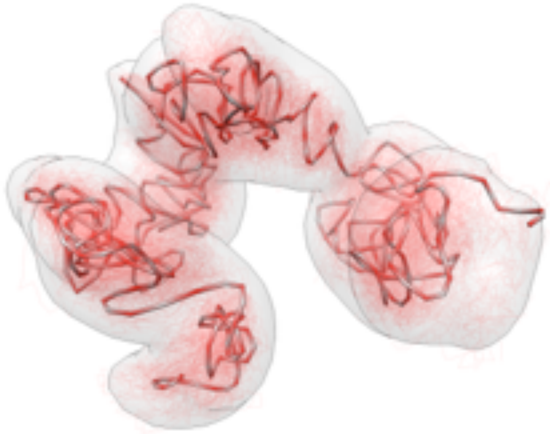
Computation

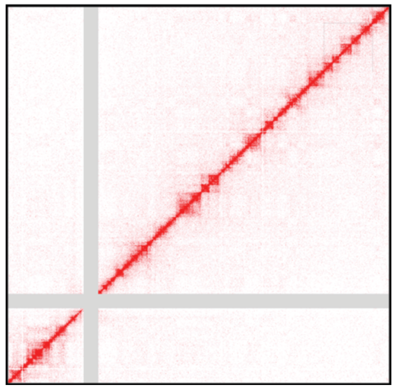
Structure determination by satisfaction of spatial restraints





Biomolecular structure determination
2D-NOESY data





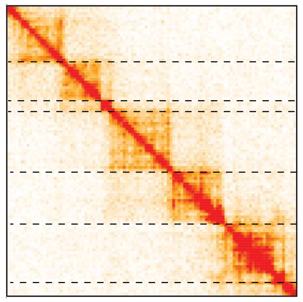
Chromosome structure determination
3C-based data

TADbit

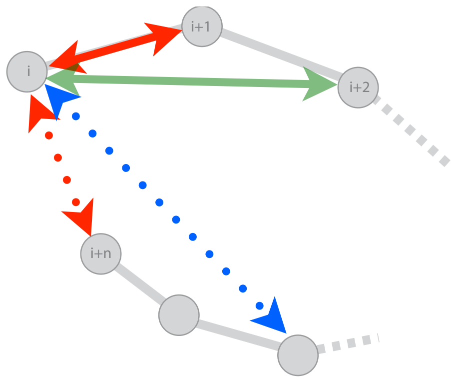
http://3DGenomes.org

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

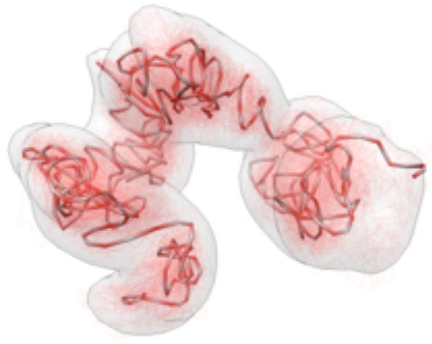
FastQ files to Maps



Map analysis



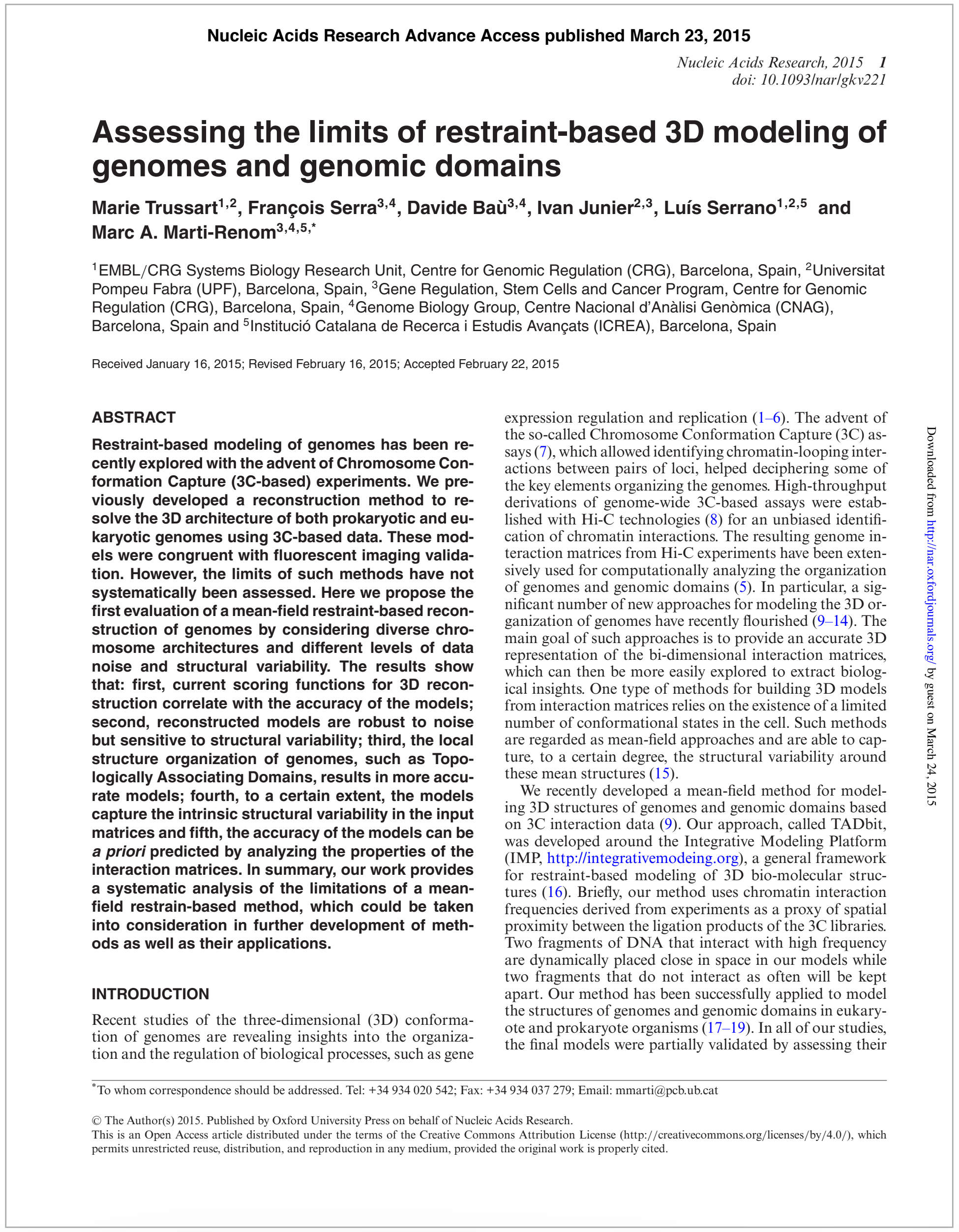
Model building



Model analysis

TADbit accuracy?

Trussart, M. et al. NAR (2015)



Nucleic Acids Research Advance Access published March 23, 2015

Nucleic Acids Research, 2015 **1**
doi: 10.1093/nar/gkv221

Assessing the limits of restraint-based 3D modeling of genomes and genomic domains

Marie Trussart[1,2], François Serra[3,4], Davide Baù[3,4], Ivan Junier[2,3], Luís Serrano[1,2,5] and Marc A. Marti-Renom[3,4,5,*]

[1]EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), Barcelona, Spain, [2]Universitat Pompeu Fabra (UPF), Barcelona, Spain, [3]Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), Barcelona, Spain, [4]Genome Biology Group, Centre Nacional d'Anàlisi Genòmica (CNAG), Barcelona, Spain and [5]Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Received January 16, 2015; Revised February 16, 2015; Accepted February 22, 2015

ABSTRACT

Restraint-based modeling of genomes has been recently explored with the advent of Chromosome Conformation Capture (3C-based) experiments. We previously developed a reconstruction method to resolve the 3D architecture of both prokaryotic and eukaryotic genomes using 3C-based data. These models were congruent with fluorescent imaging validation. However, the limits of such methods have not systematically been assessed. Here we propose the first evaluation of a mean-field restraint-based reconstruction of genomes by considering diverse chromosome architectures and different levels of data noise and structural variability. The results show that: first, current scoring functions for 3D reconstruction correlate with the accuracy of the models; second, reconstructed models are robust to noise but sensitive to structural variability; third, the local structure organization of genomes, such as Topologically Associating Domains, results in more accurate models; fourth, to a certain extent, the models capture the intrinsic structural variability in the input matrices and fifth, the accuracy of the models can be *a priori* predicted by analyzing the properties of the interaction matrices. In summary, our work provides a systematic analysis of the limitations of a mean-field restrain-based method, which could be taken into consideration in further development of methods as well as their applications.

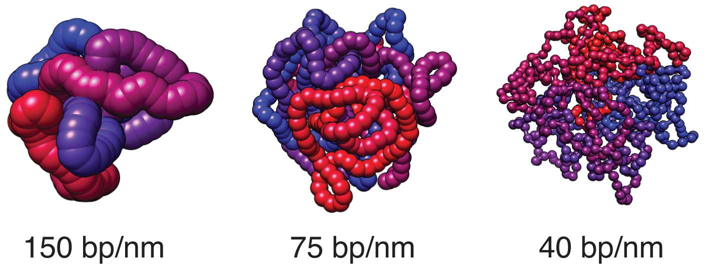
INTRODUCTION

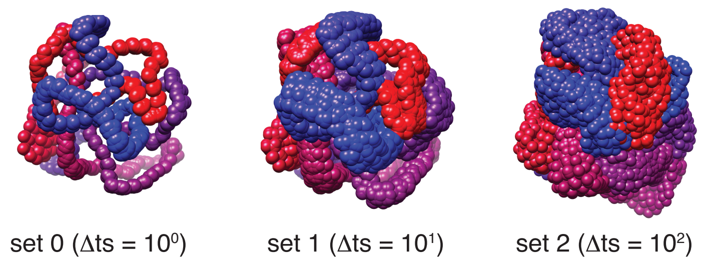
Recent studies of the three-dimensional (3D) conformation of genomes are revealing insights into the organization and the regulation of biological processes, such as gene expression regulation and replication (1–6). The advent of the so-called Chromosome Conformation Capture (3C) assays (7), which allowed identifying chromatin-looping interactions between pairs of loci, helped deciphering some of the key elements organizing the genomes. High-throughput derivations of genome-wide 3C-based assays were established with Hi-C technologies (8) for an unbiased identification of chromatin interactions. The resulting genome interaction matrices from Hi-C experiments have been extensively used for computationally analyzing the organization of genomes and genomic domains (5). In particular, a significant number of new approaches for modeling the 3D organization of genomes have recently flourished (9–14). The main goal of such approaches is to provide an accurate 3D representation of the bi-dimensional interaction matrices, which can then be more easily explored to extract biological insights. One type of methods for building 3D models from interaction matrices relies on the existence of a limited number of conformational states in the cell. Such methods are regarded as mean-field approaches and are able to capture, to a certain degree, the structural variability around these mean structures (15).

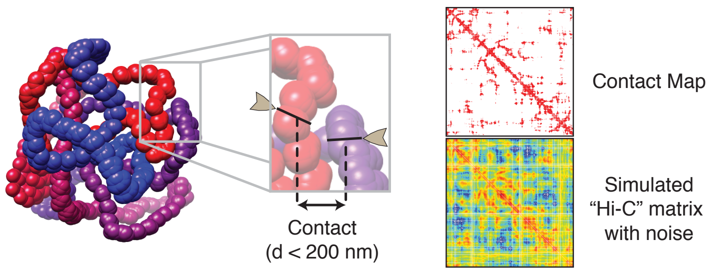
We recently developed a mean-field method for modeling 3D structures of genomes and genomic domains based on 3C interaction data (9). Our approach, called TADbit, was developed around the Integrative Modeling Platform (IMP, http://integrativemodeing.org), a general framework for restraint-based modeling of 3D bio-molecular structures (16). Briefly, our method uses chromatin interaction frequencies derived from experiments as a proxy of spatial proximity between the ligation products of the 3C libraries. Two fragments of DNA that interact with high frequency are dynamically placed close in space in our models while two fragments that do not interact as often will be kept apart. Our method has been successfully applied to model the structures of genomes and genomic domains in eukaryote and prokaryote organisms (17–19). In all of our studies, the final models were partially validated by assessing their

---

[*]To whom correspondence should be addressed. Tel: +34 934 020 542; Fax: +34 934 037 279; Email: mmarti@pcb.ub.cat







chr40_TAD
α=100
Δts=10

<dRMSD>: 32.7 nm
<dSCC>: 0.94

Shameless promotion...

TADbit

November 23th-27th, Lisbon

FUNDAÇÃO CALOUSTE GULBENKIAN

Instituto Gulbenkian de Ciência

http://gtpb.igc.gulbenkian.pt

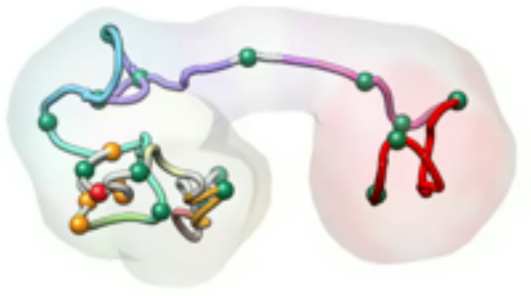
TADbit previous applications...

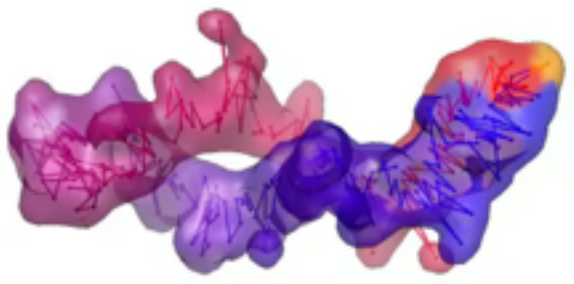
Baù, D. et al. Nat Struct Mol Biol (2011).
Umbarger, M. A. et al. Mol Cell (2011).
Le Dily, F. et al. Genes & Dev (2014)

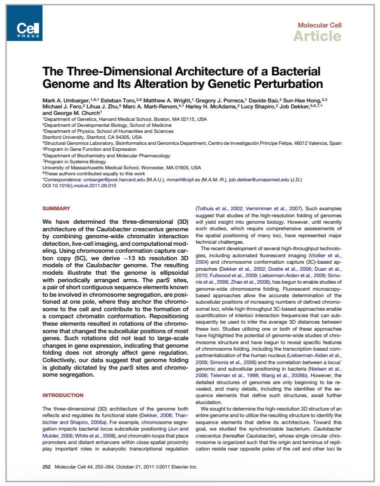

TECHNICAL REPORTS

nature structural & molecular biology

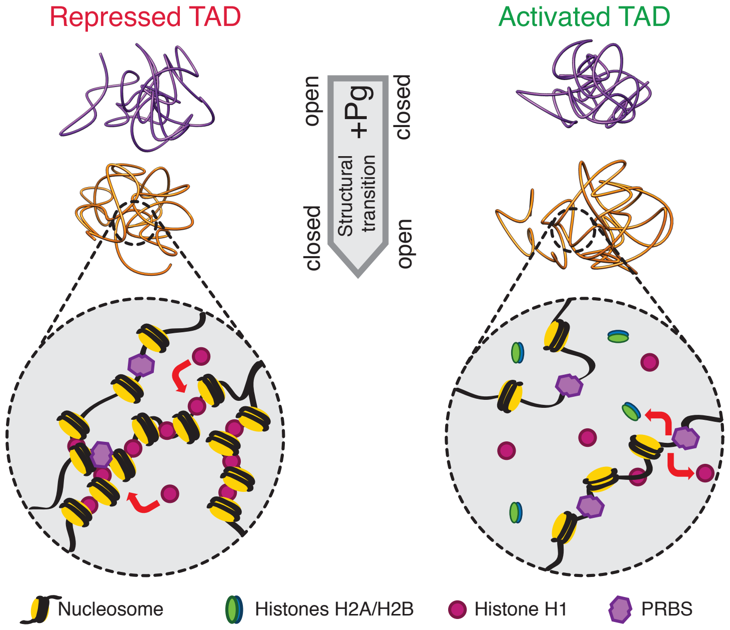
The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules

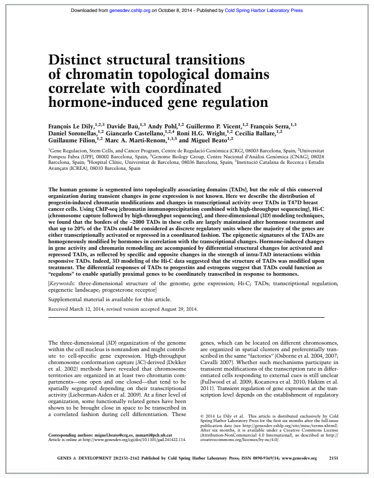




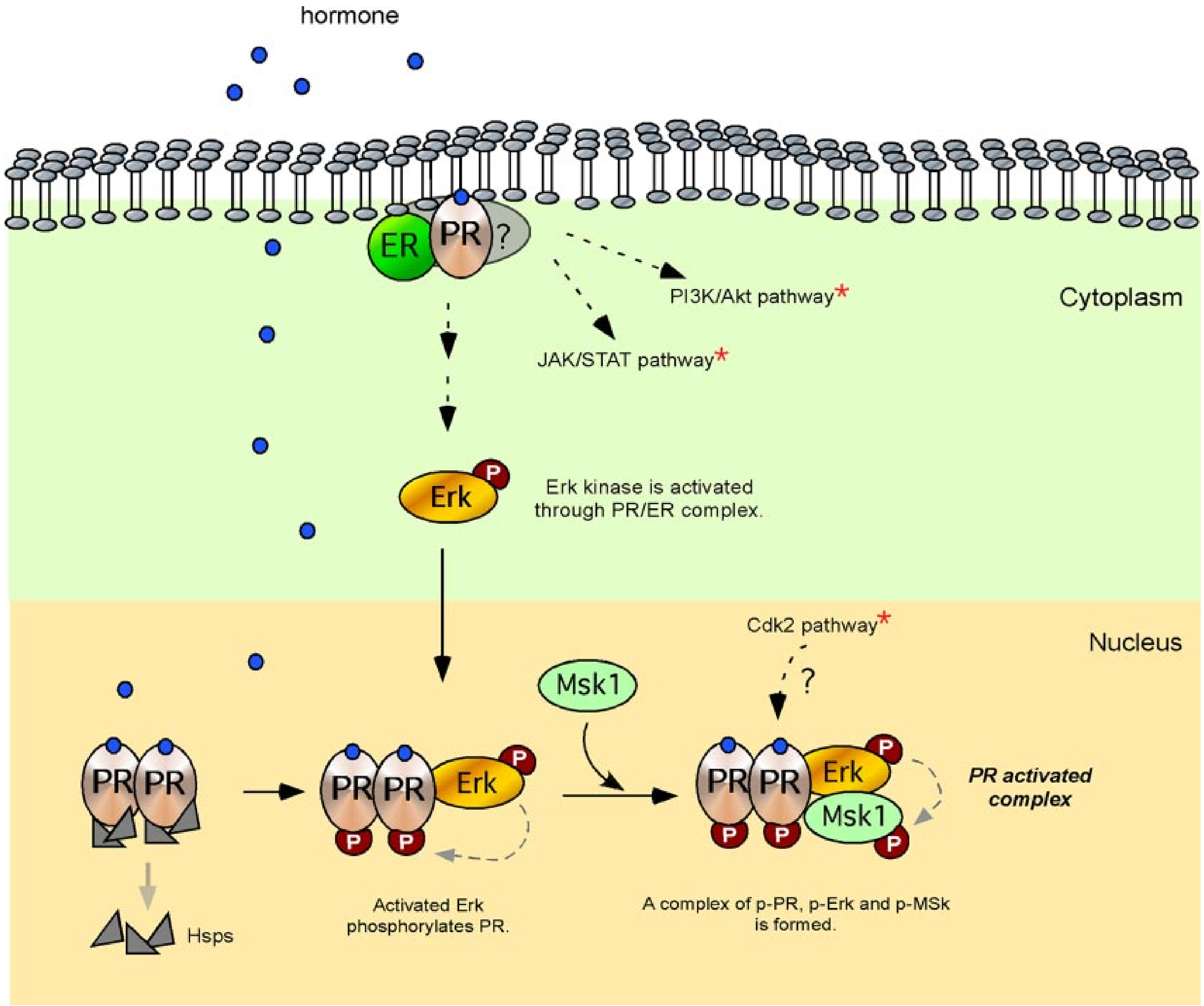

Molecular Cell
Article

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




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

Progesterone-regulated transcription in breast cancer



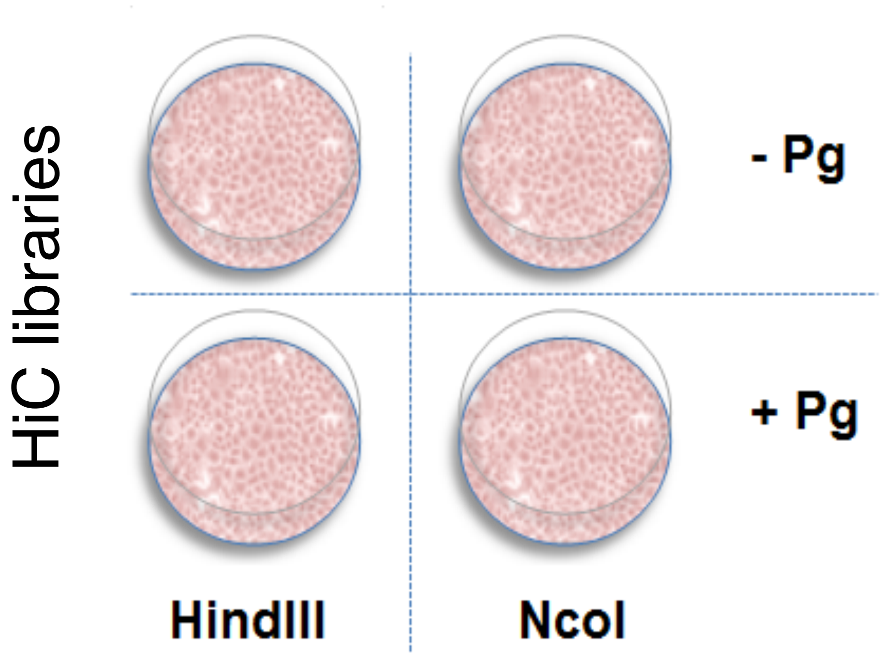
Vicent *et al* 2011, Wright *et al* 2012, Ballare *et al* 2012

> 2,000 genes **Up**-regulated

> 2,000 genes **Down**-regulated

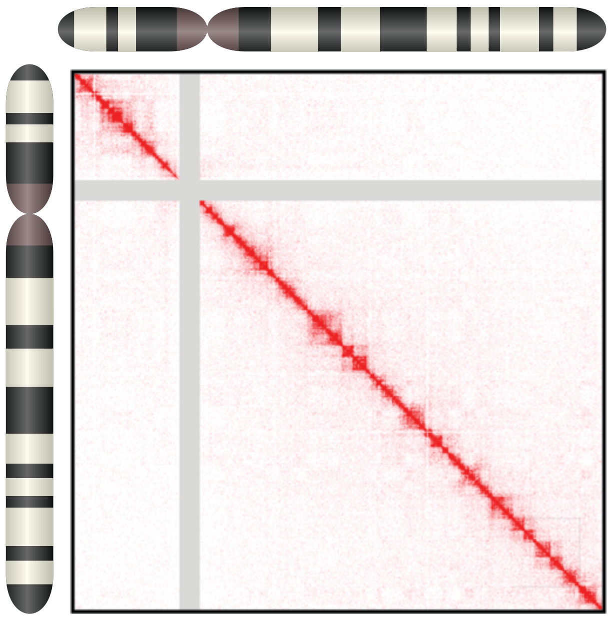
Regulation in 3D?

Experimental design

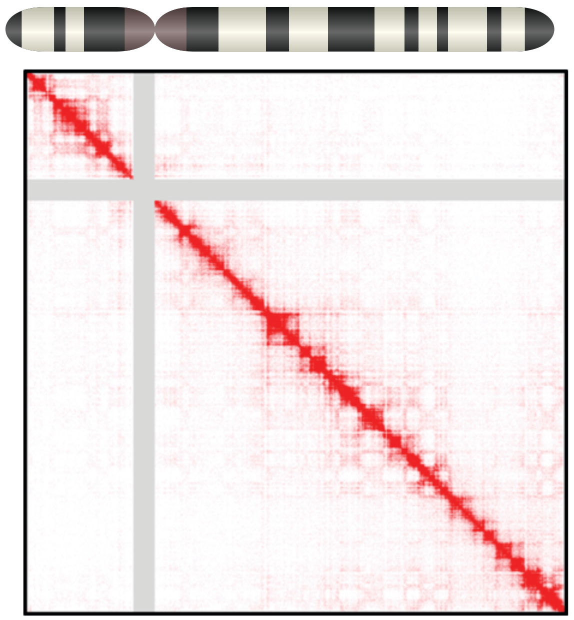


ChIP-Seq
RNA-Seq
Hi-C

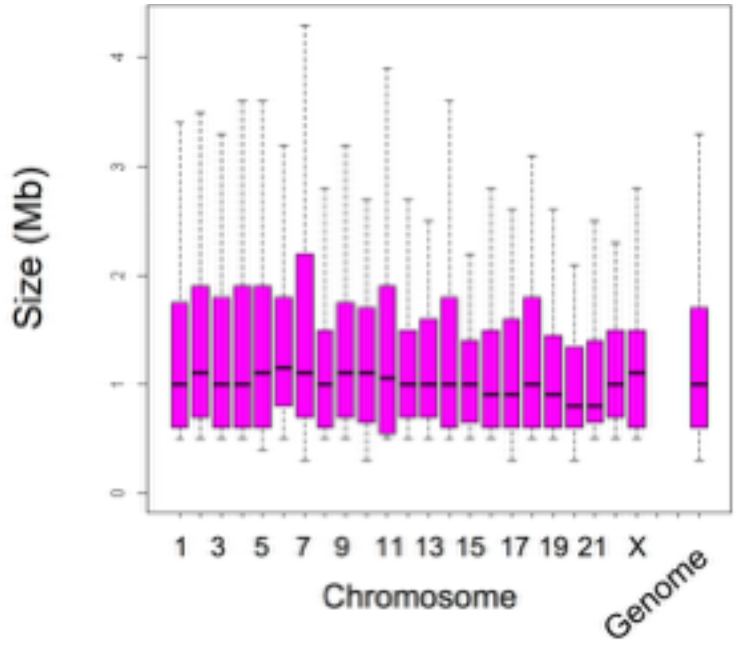
Chr.18 (Hind III)



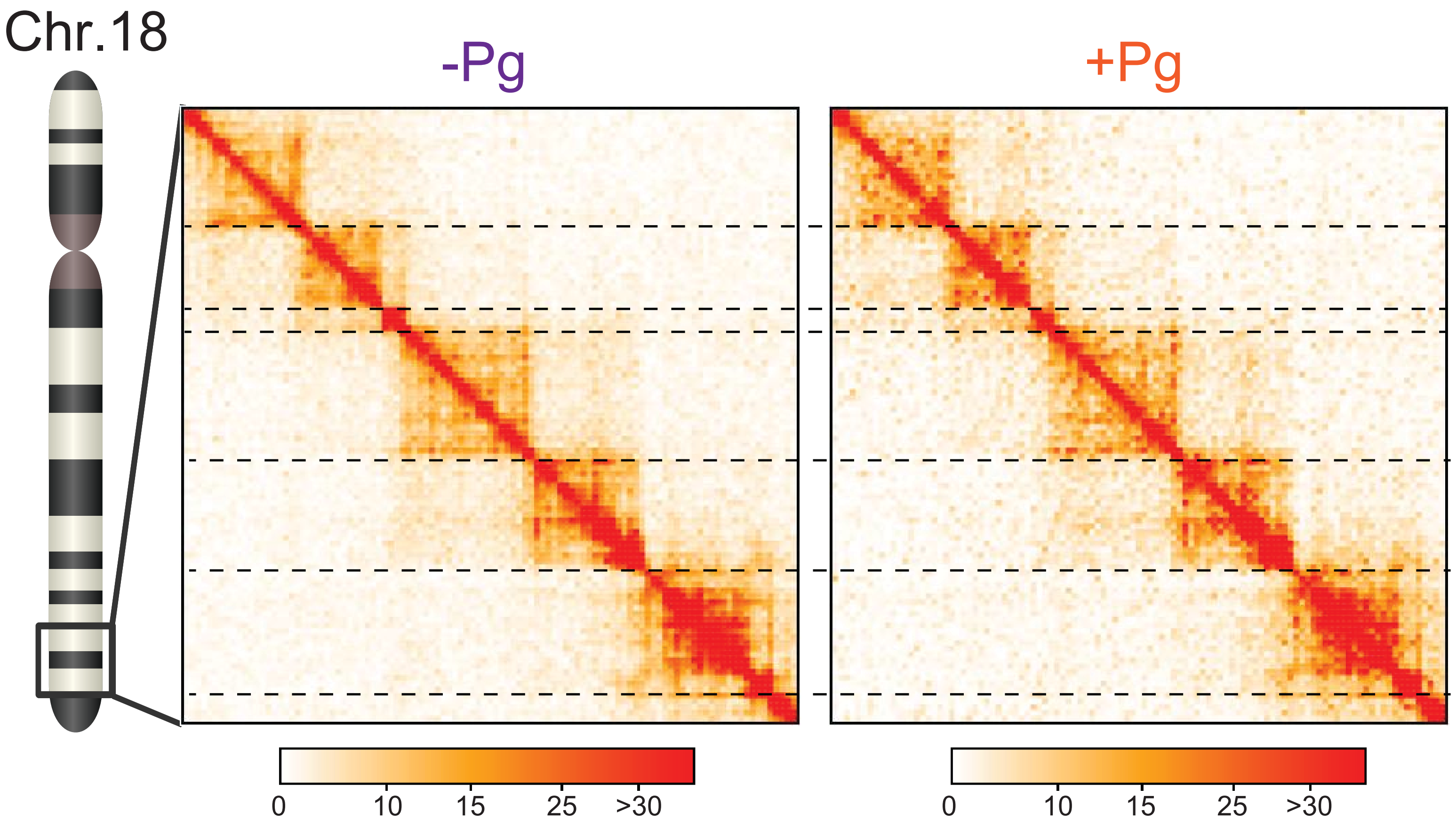
Chr.18 (NcoI)

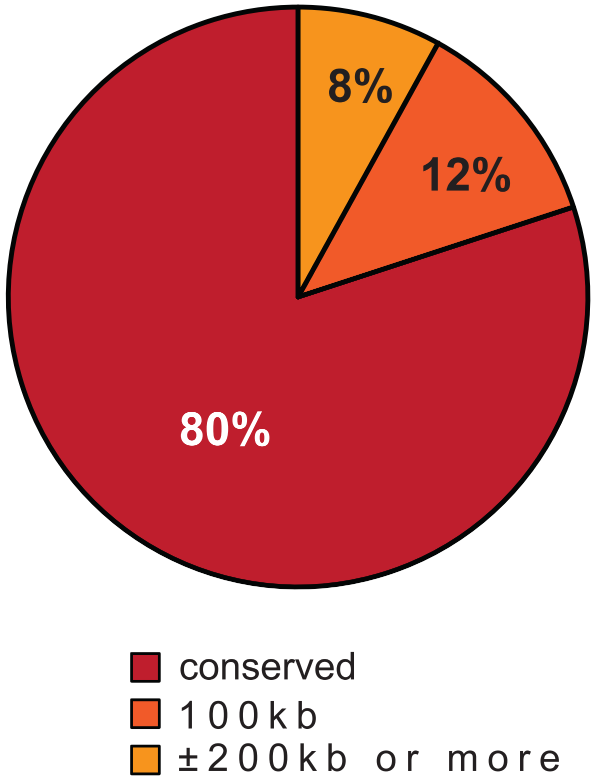


Are there TADs? how robust?

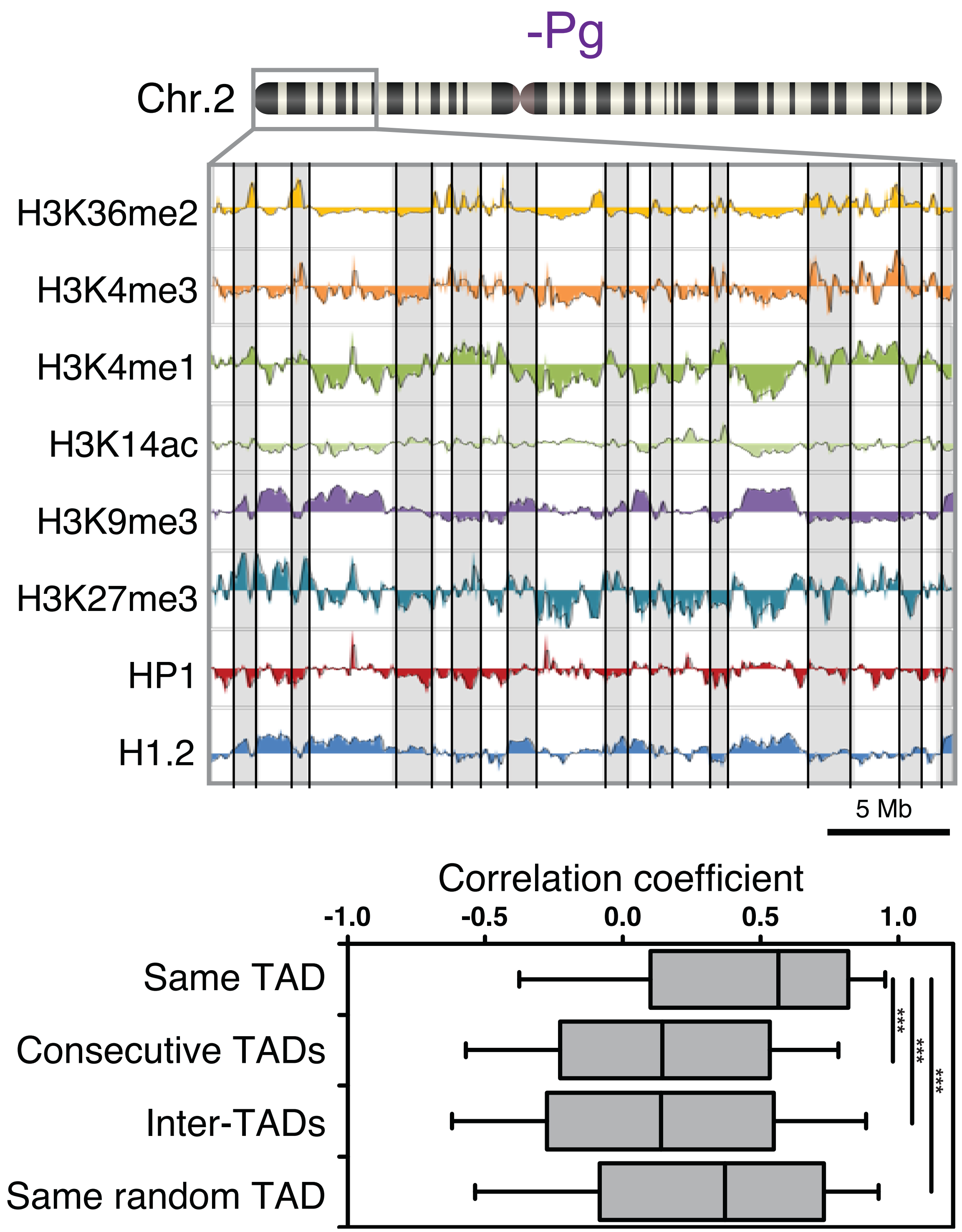


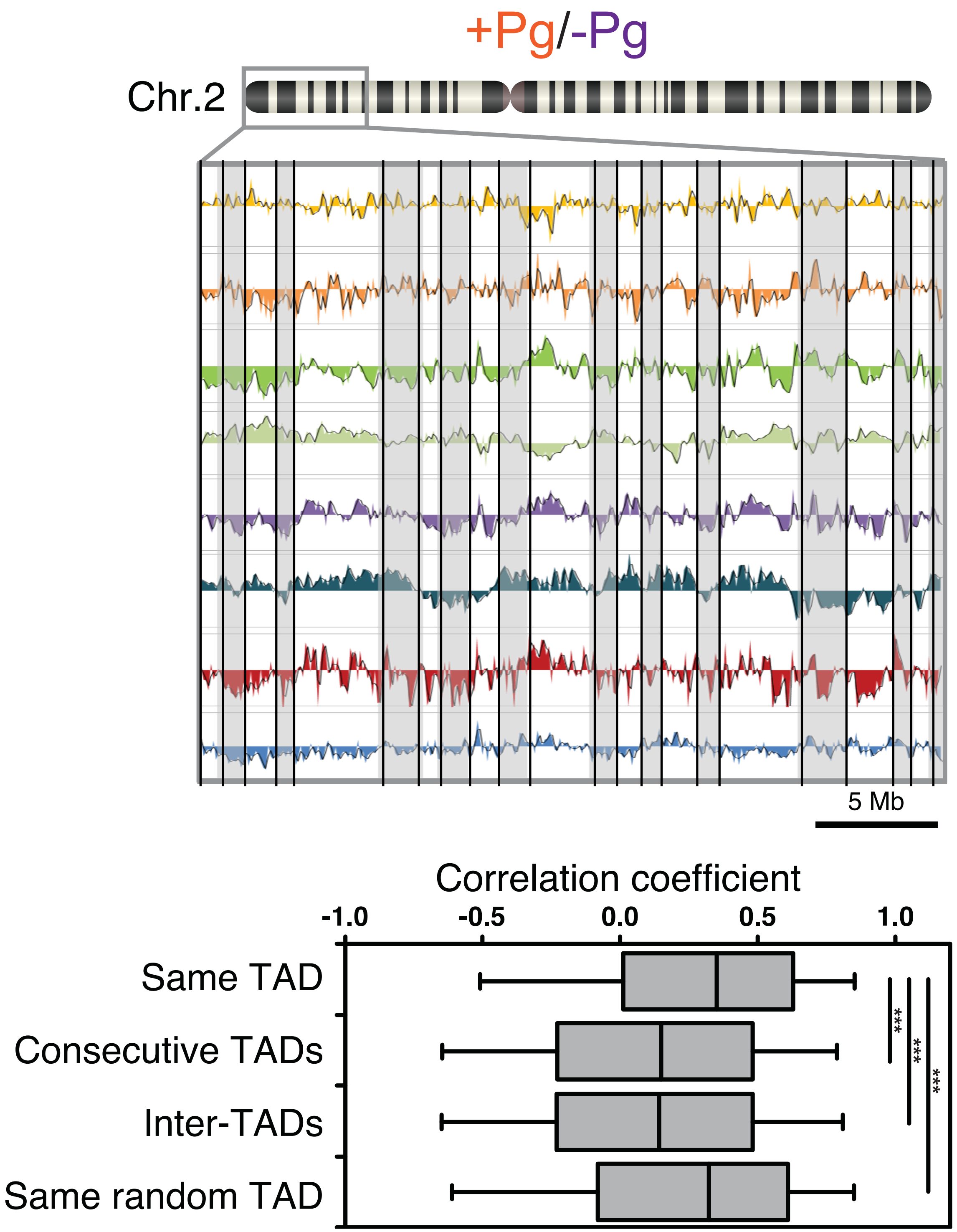
>2,000 detected TADs



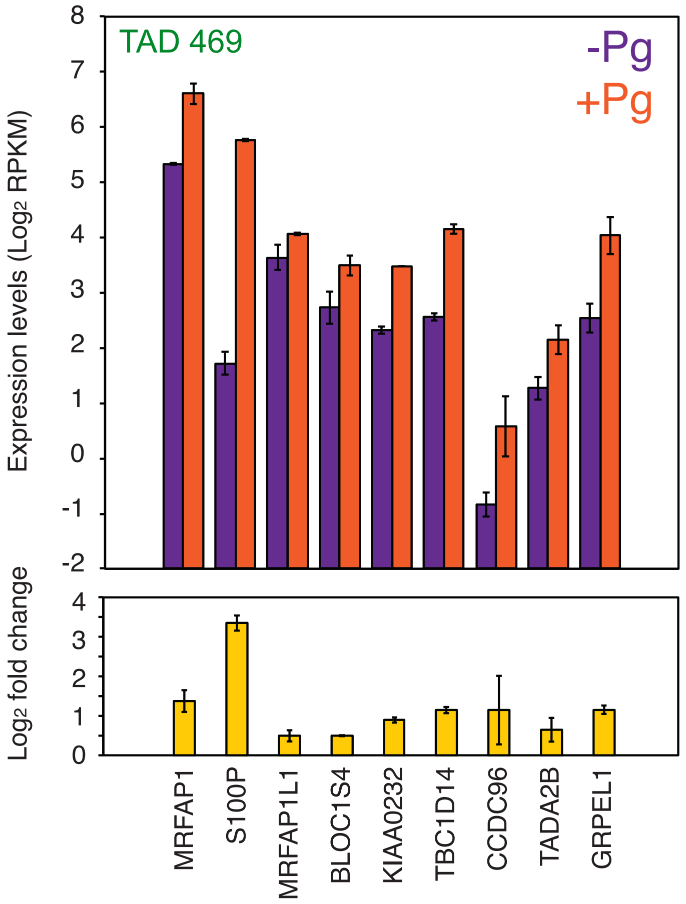


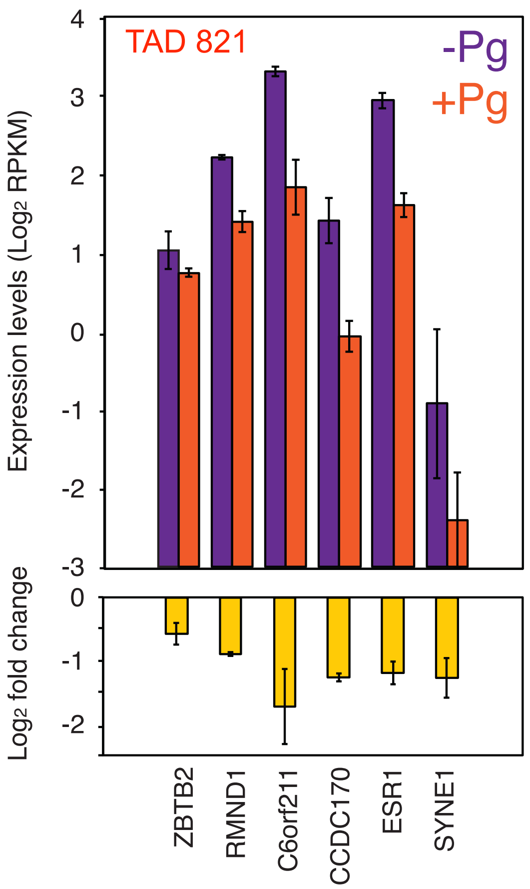
Are TADs homogeneous?

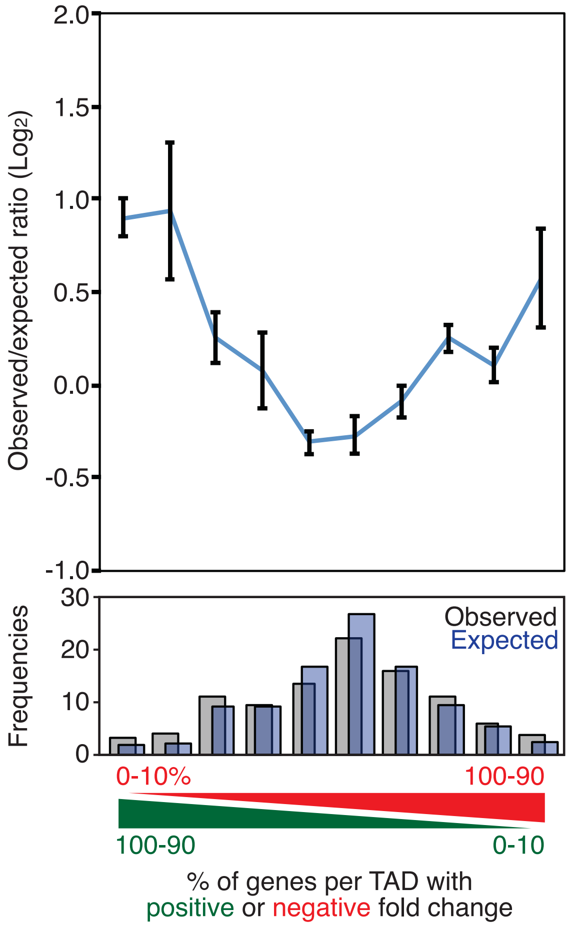




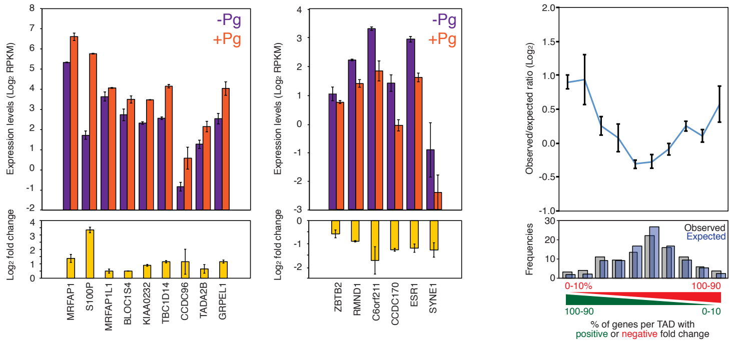
Do TADs respond differently to Pg treatment?

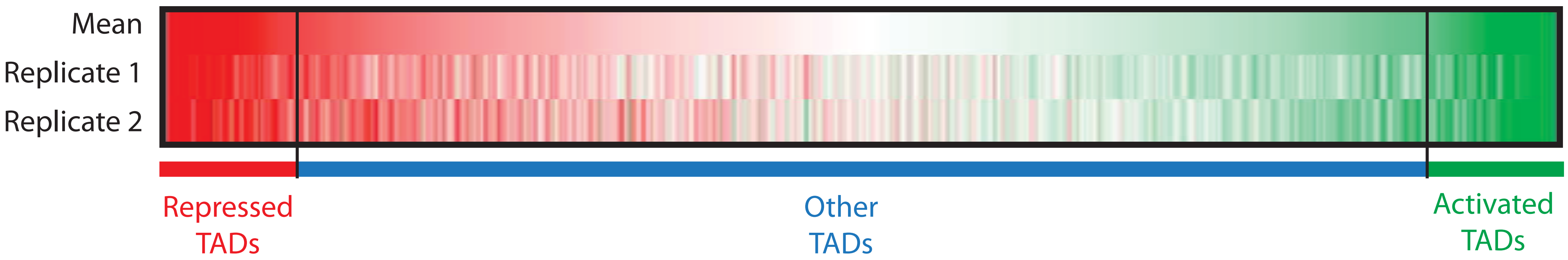


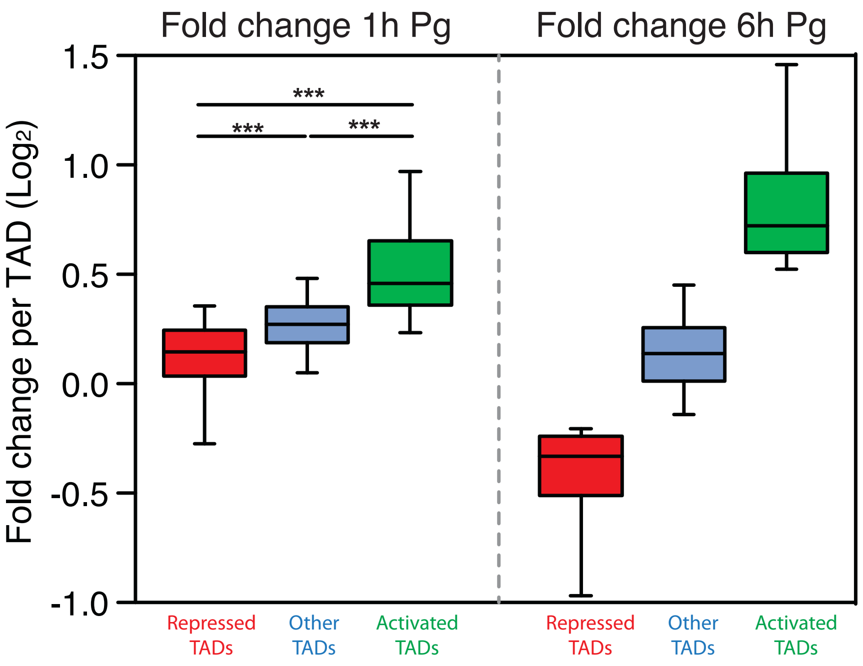


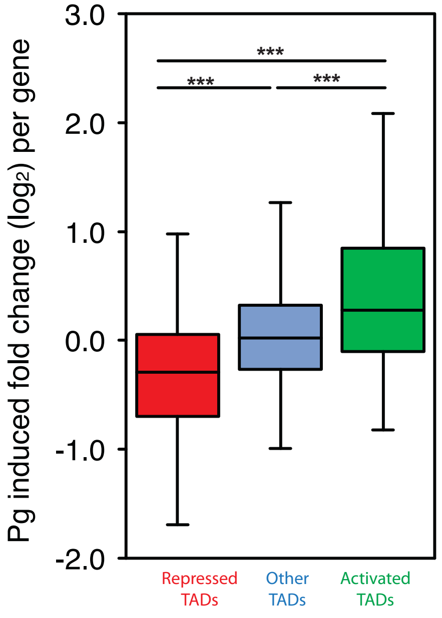


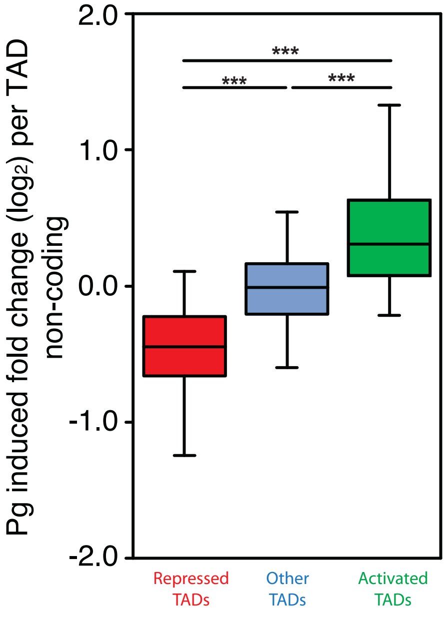
Do TADs respond differently to Pg treatment?

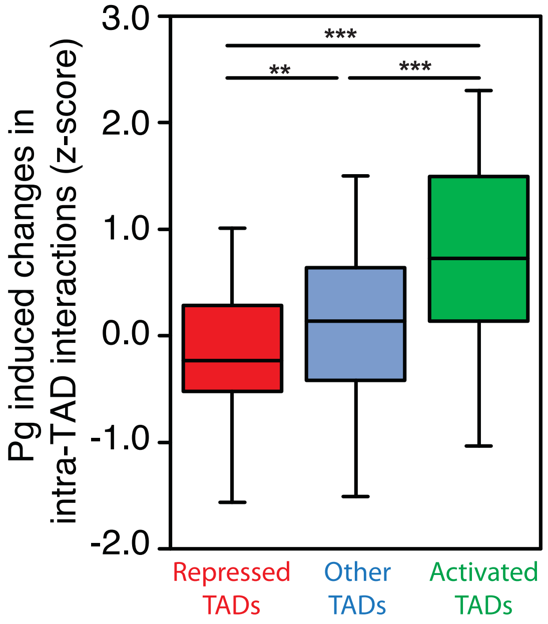



Pg induced fold change per TAD (6h)

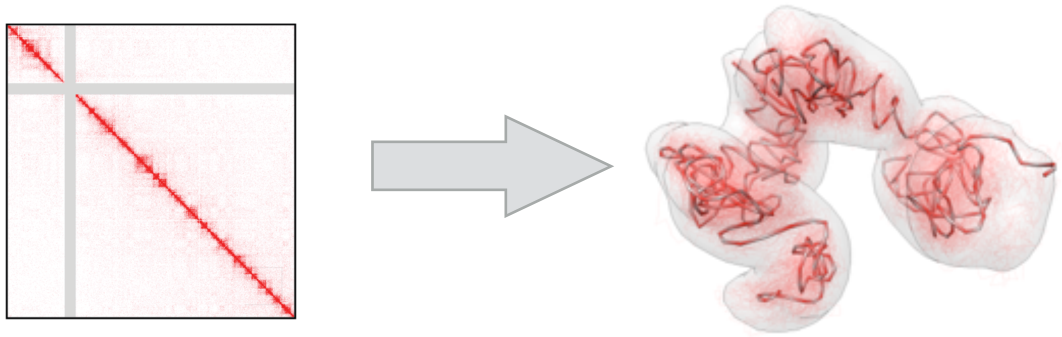




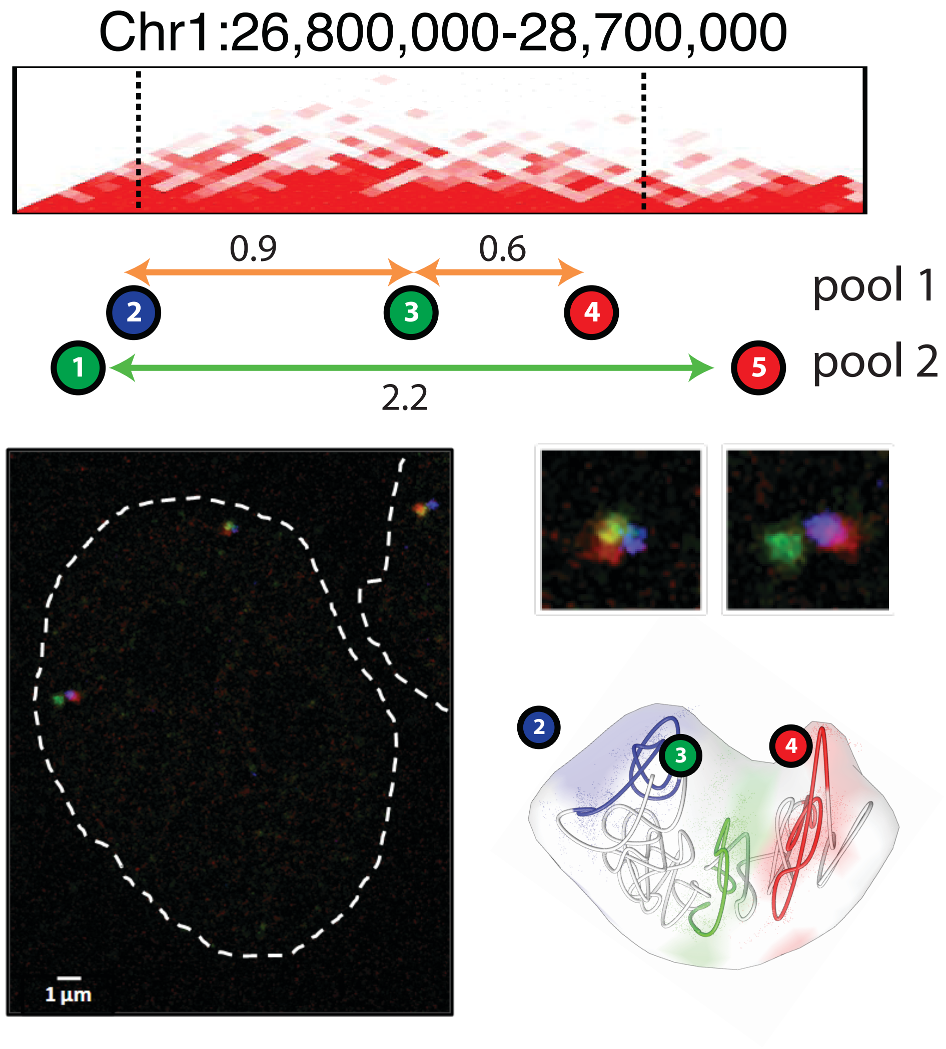


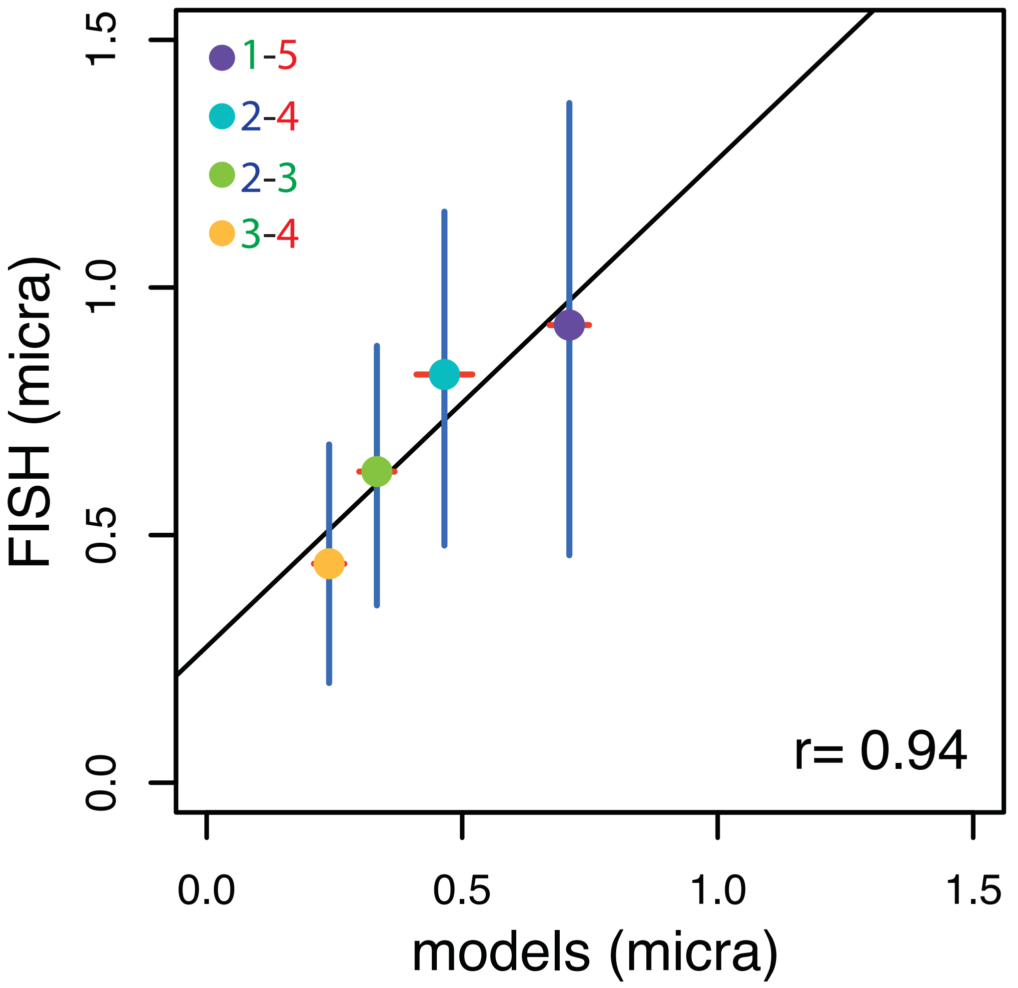


Modeling 3D TADs

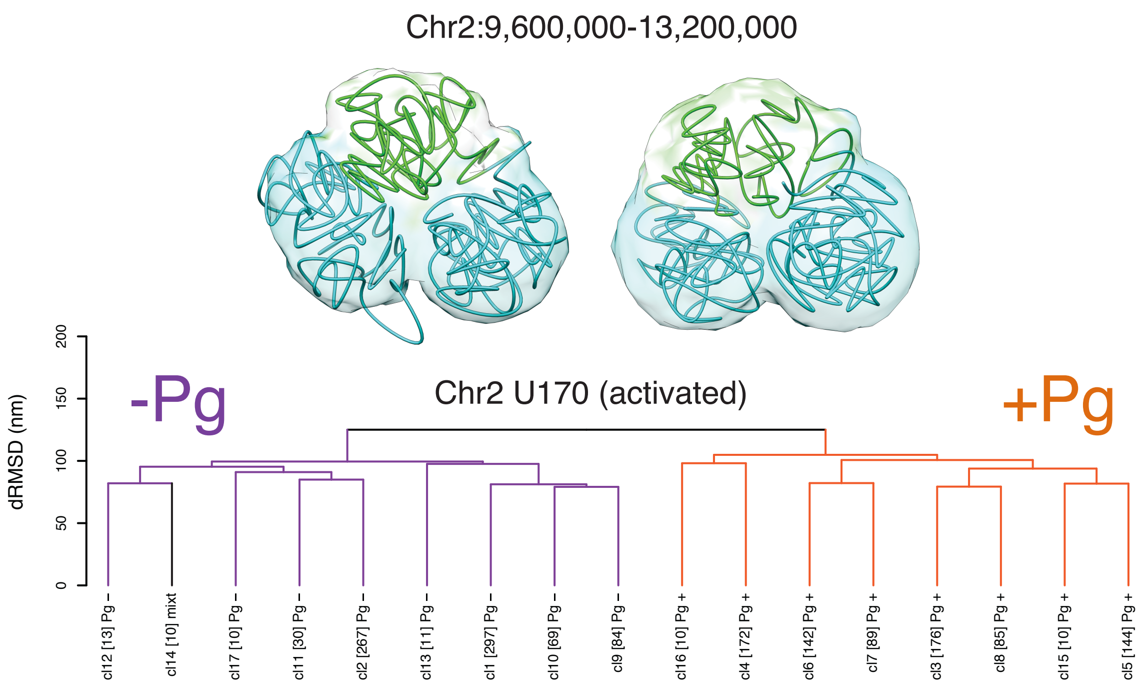


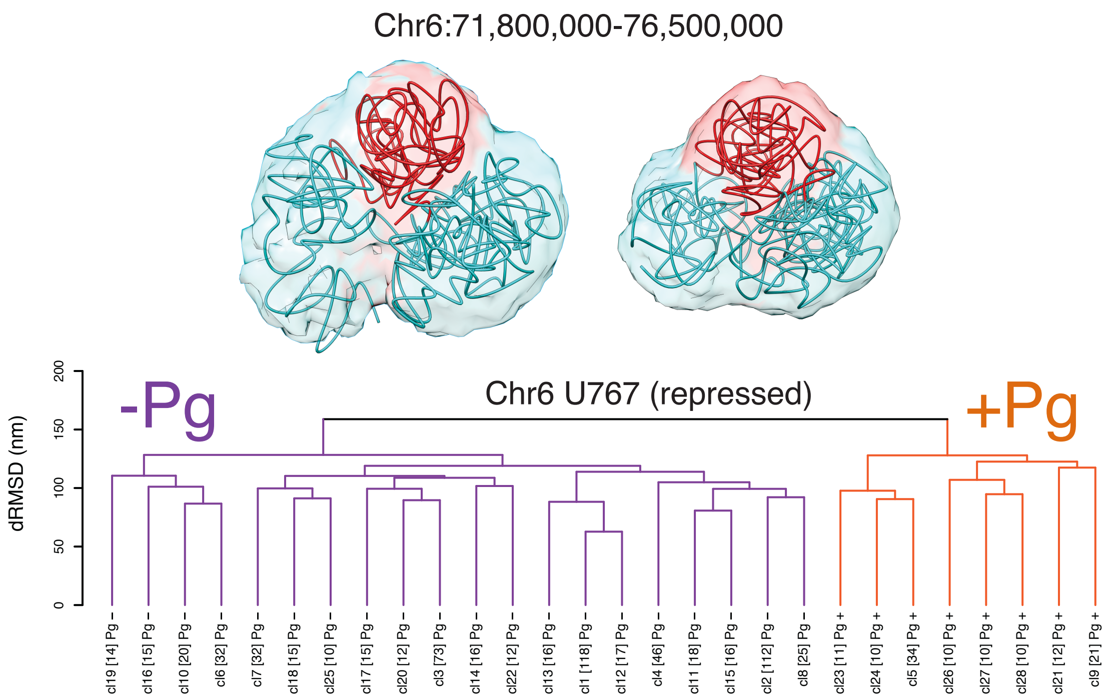
61 genomic regions containing 209 TADs covering 267Mb

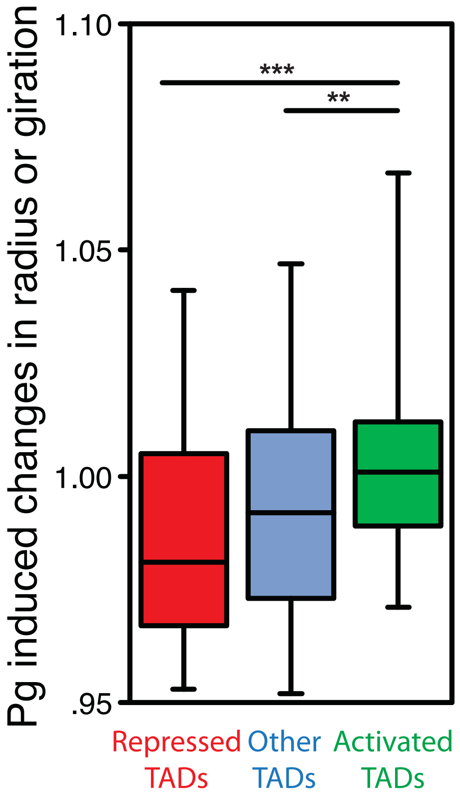


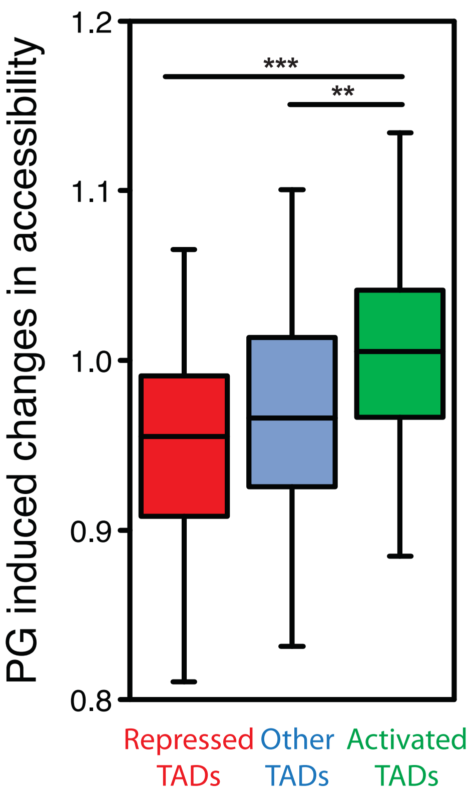


How TADs respond structurally to Pg?

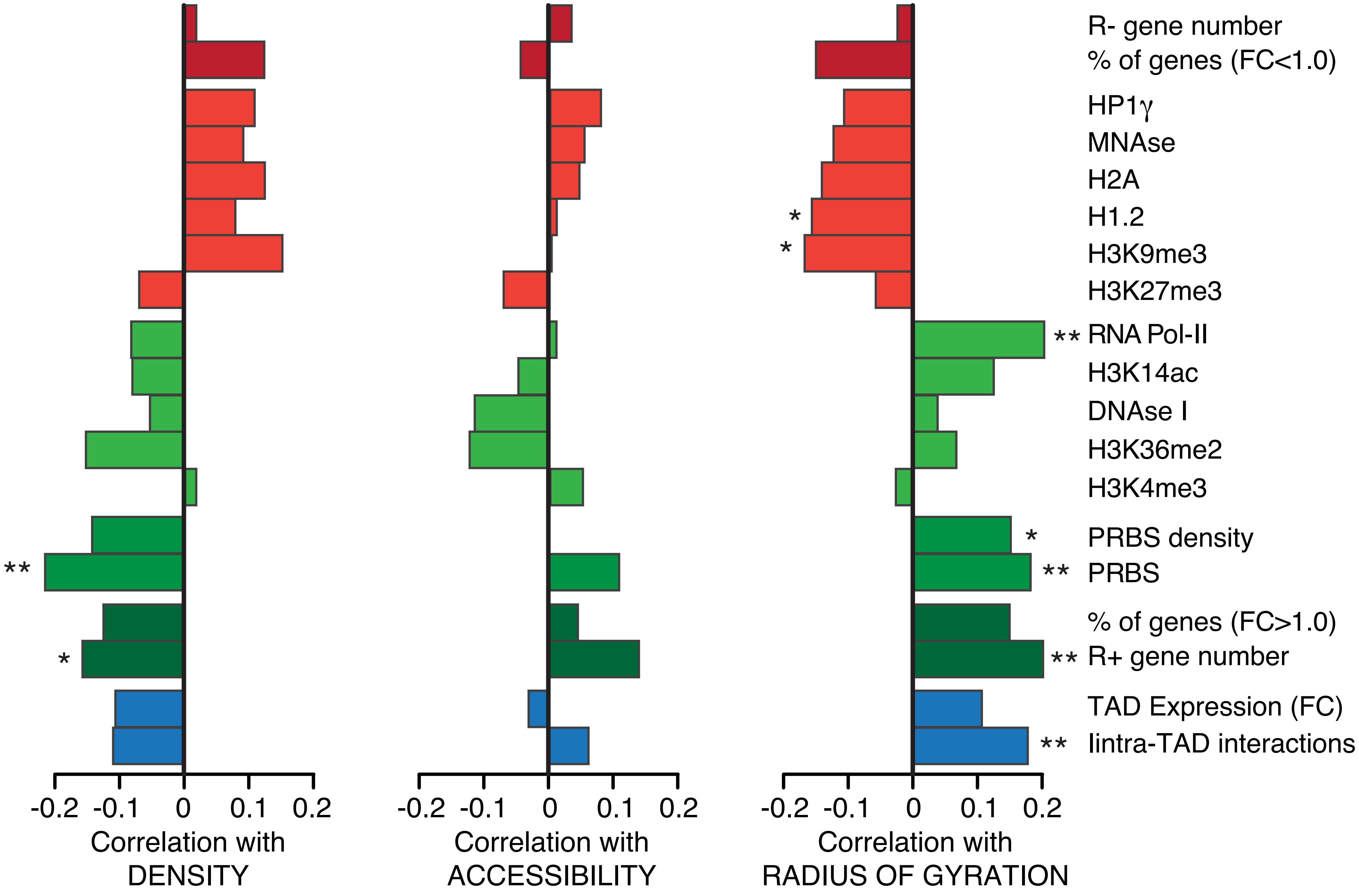


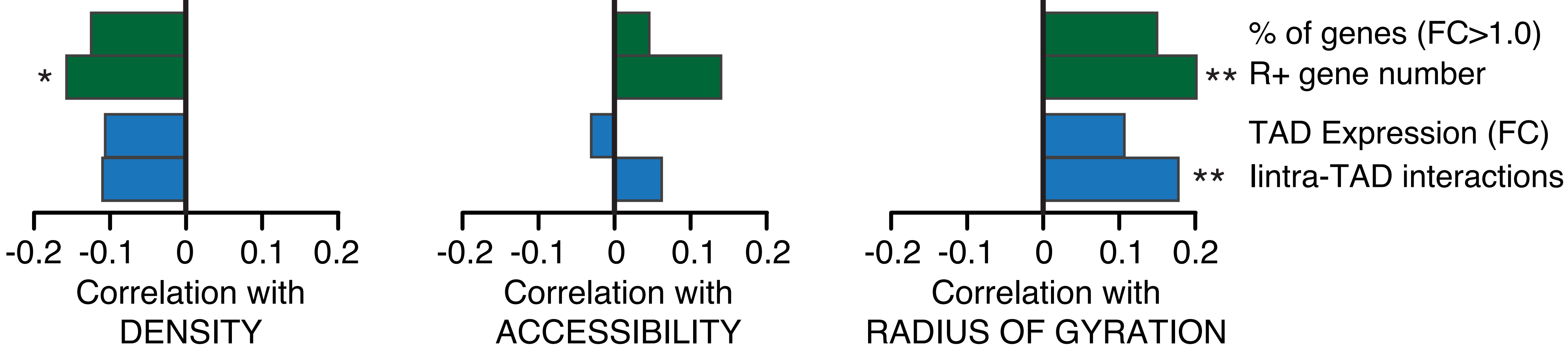


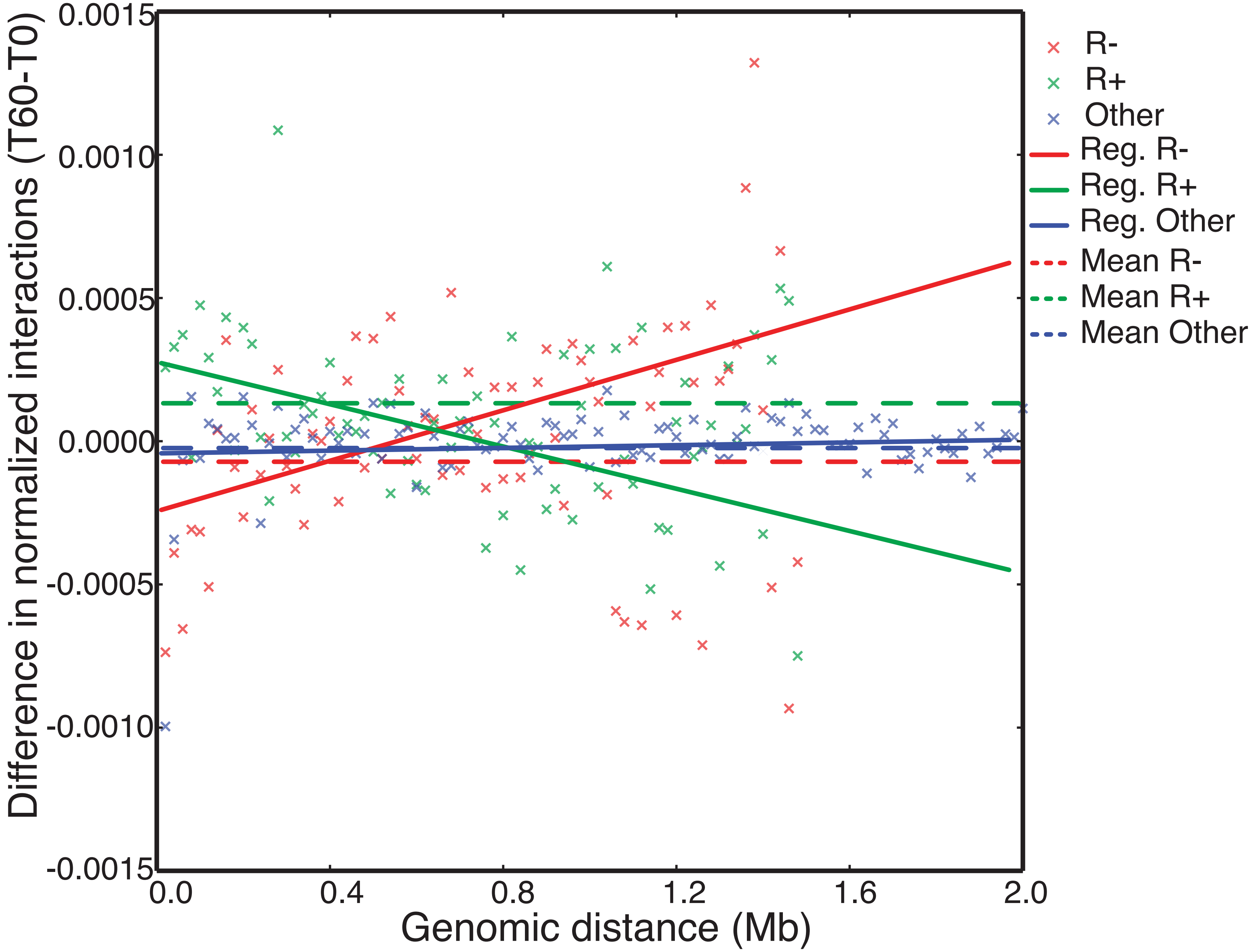




How TADs respond structurally to Pg?

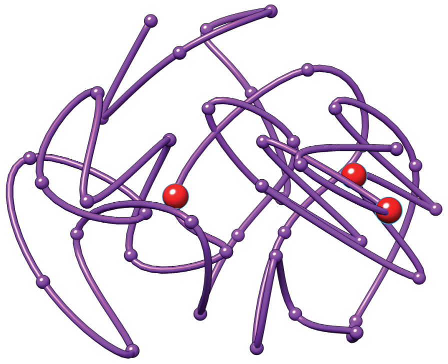


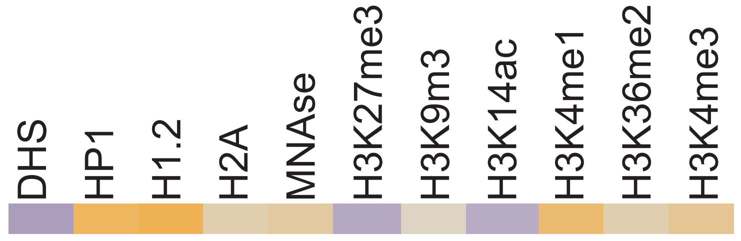

% of genes (FC>1.0)
R+ gene number
TAD Expression (FC)
Iintra-TAD interactions
*
**
**
-0.2 -0.1 0 0.1 0.2
Correlation with DENSITY
Correlation with ACCESSIBILITY
Correlation with RADIUS OF GYRATION

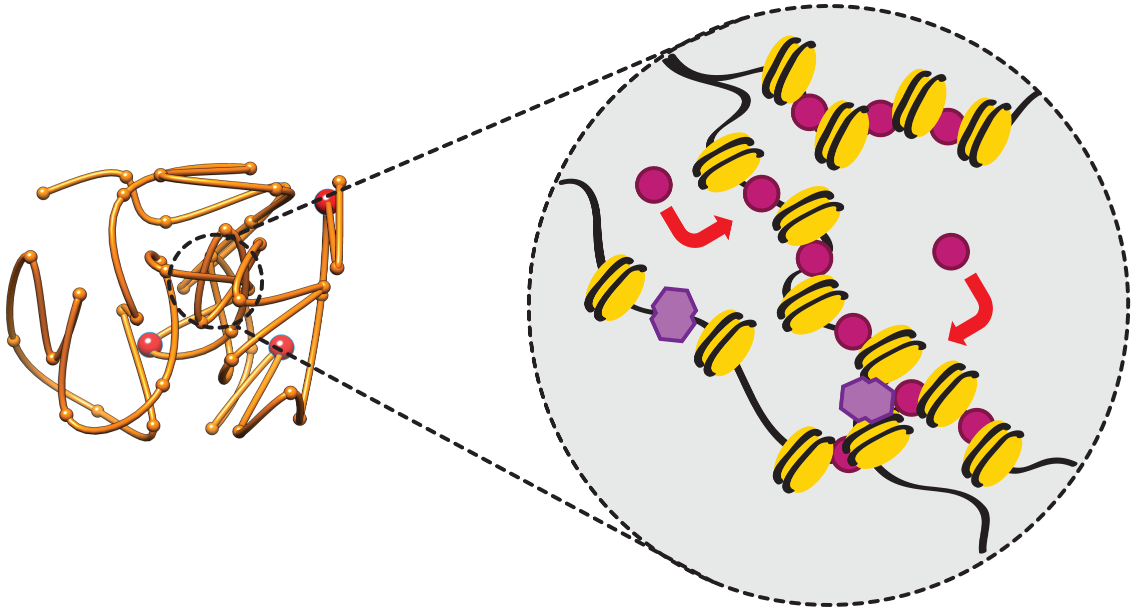

Difference in normalized interactions (T60-T0)
0.0015
0.0010
0.0005
0.0000
-0.0005
-0.0010
-0.0015
0.0
0.4
0.8
1.2
1.6
2.0
Genomic distance (Mb)
R-
R+
Other
Reg. R-
Reg. R+
Reg. Other
Mean R-
Mean R+
Mean Other

Model for TAD regulation

Repressed TAD
chr1 U41

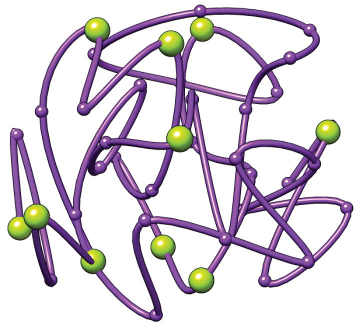


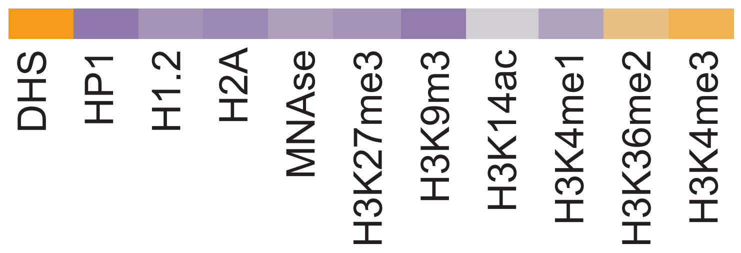


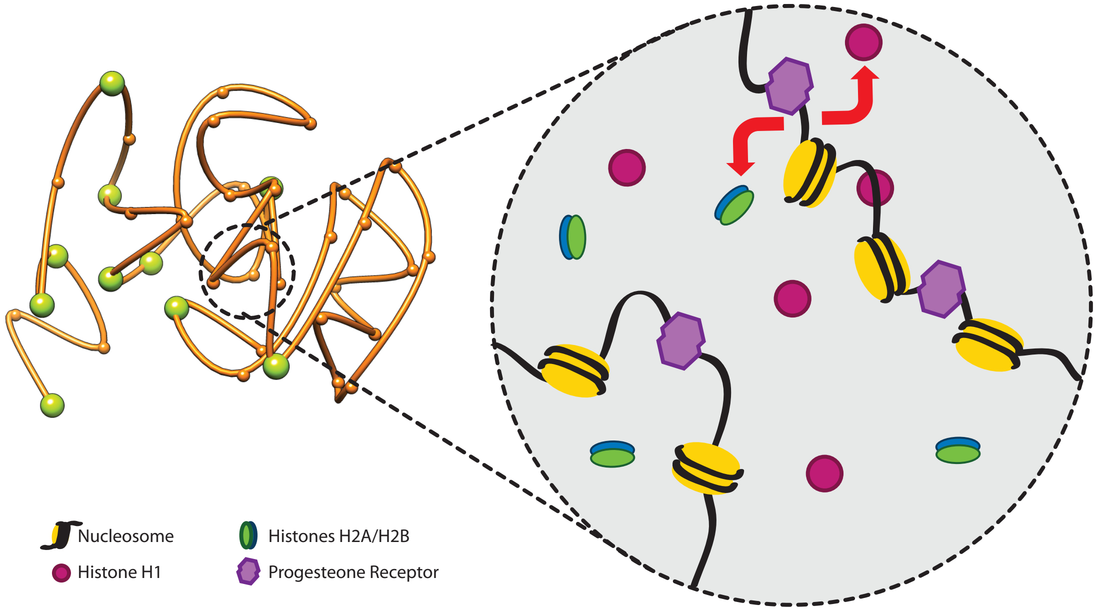


Structural transition **+Pg**

Activated TAD
chr2 U207







Acknowledgments

François le Dily
Davide Baù
François Serra

Mike Goodstadt
Yasmina Cuartero
Francisco Martínez-Jiménez
David Dufour
Gireesh Bogu
Yannick Spill
Marco di Stefano
Irene Farabella
Paula Soler

Miguel Beato & Guillaume Filion
Gene Regulation, Stem Cells and Cancer
Centre de Regulació Genòmica
Barcelona, Spain

http://marciuslab.org
http://3DGenomes.org
http://cnag.cat · http://crg.cat