

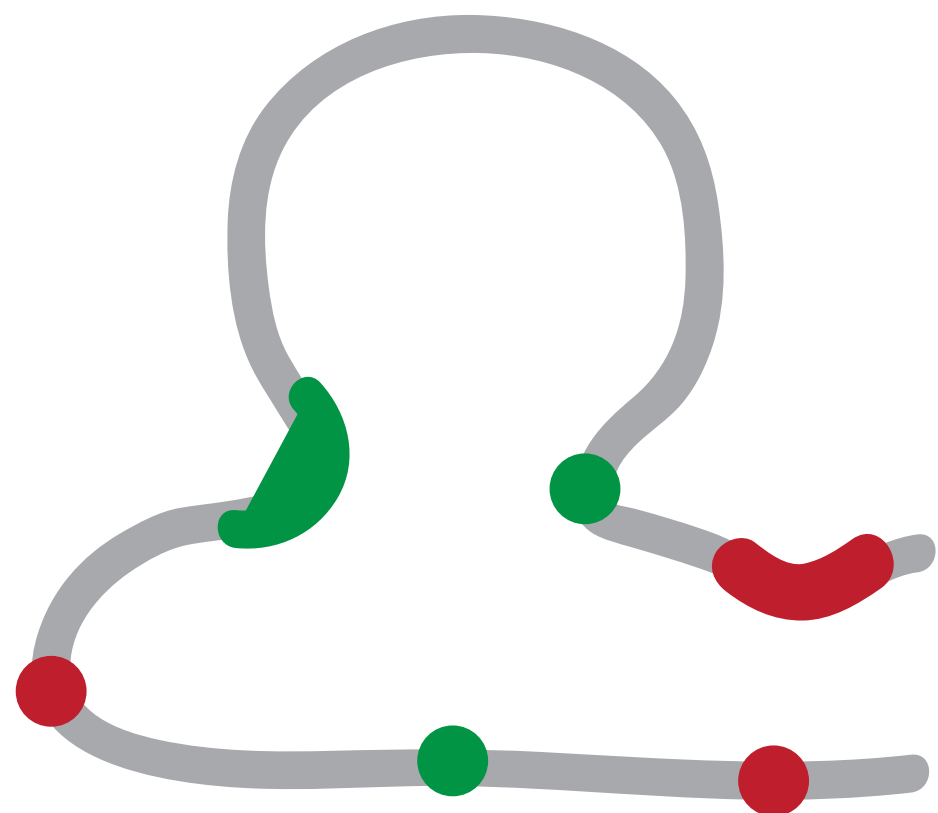
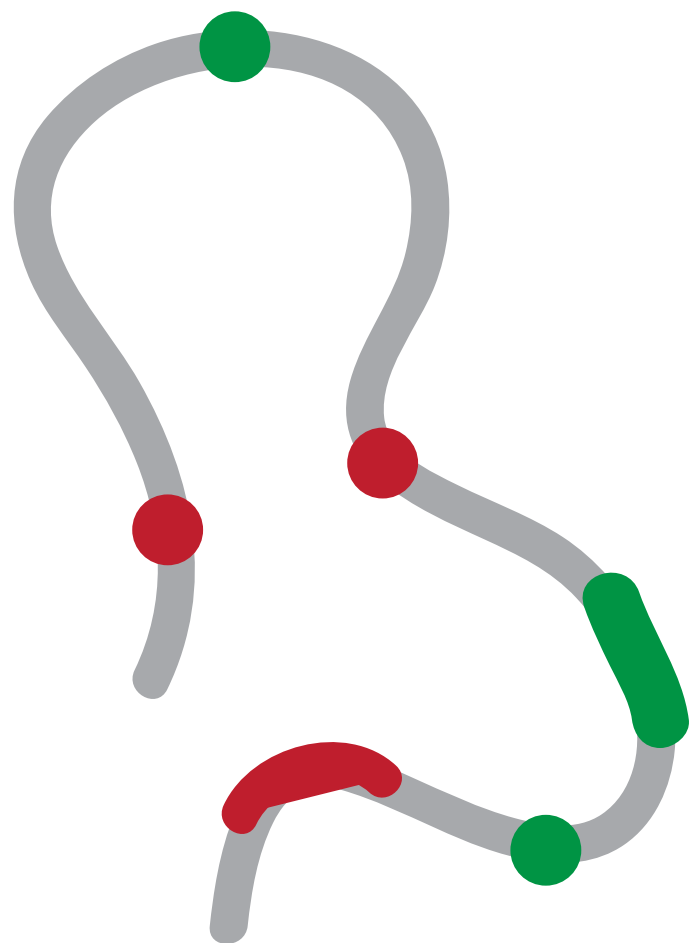
Structure determination of genomes and genomic domains by satisfaction of spatial restraints

Marc A. Marti-Renom

Structural Genomics Group (ICREA, CNAG-CRG)

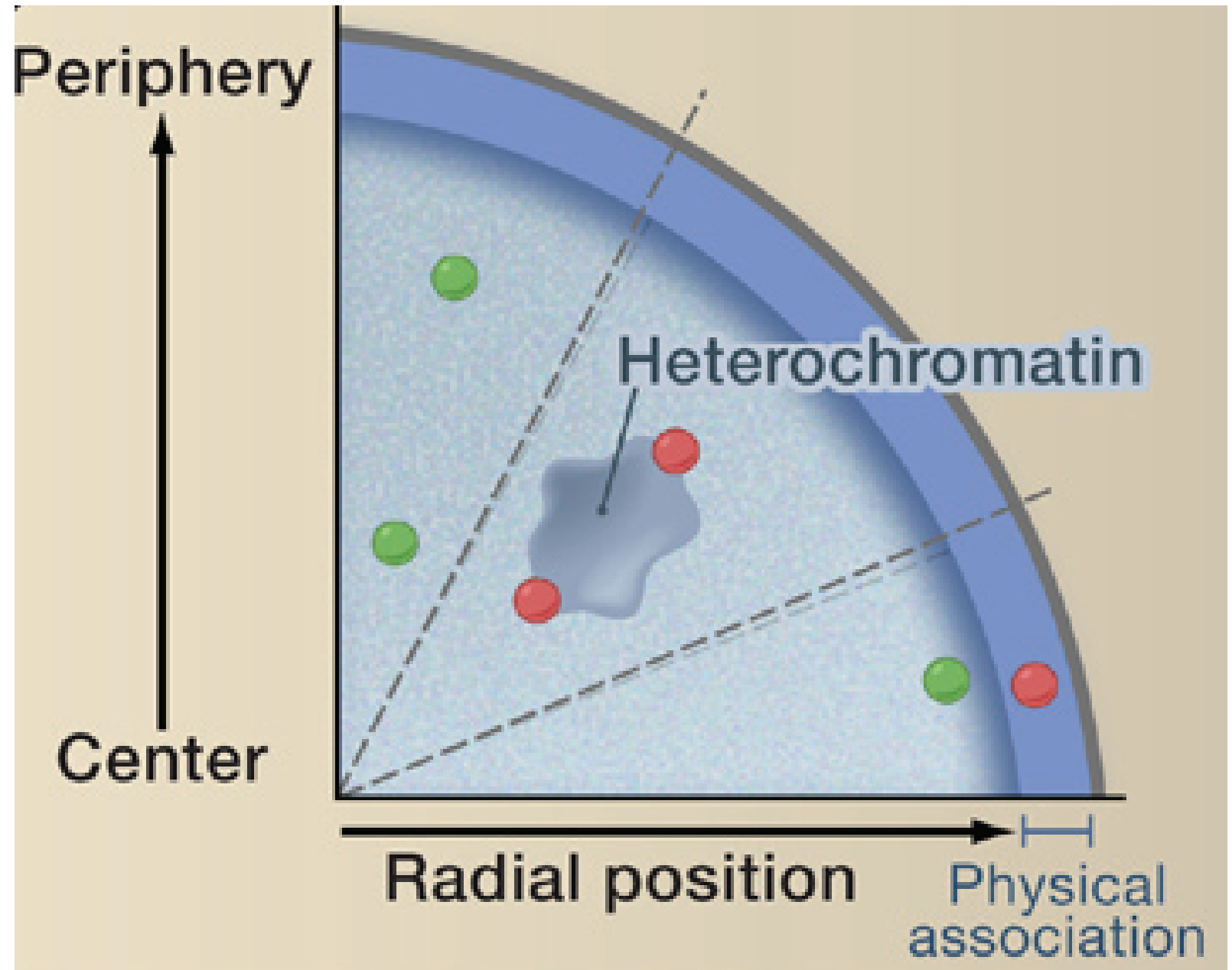
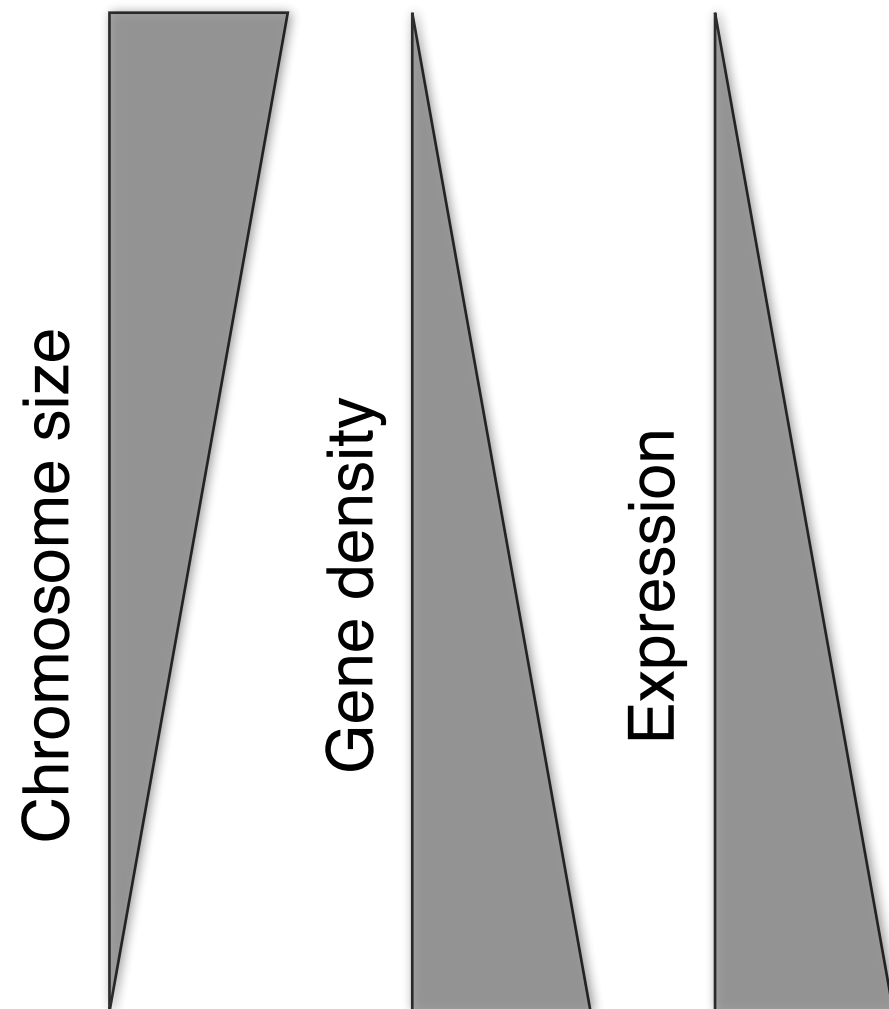
<http://marciuslab.org>
<http://3DGenomes.org>
<http://cnag.crg.eu>

cnag CRG[®]  ICREA



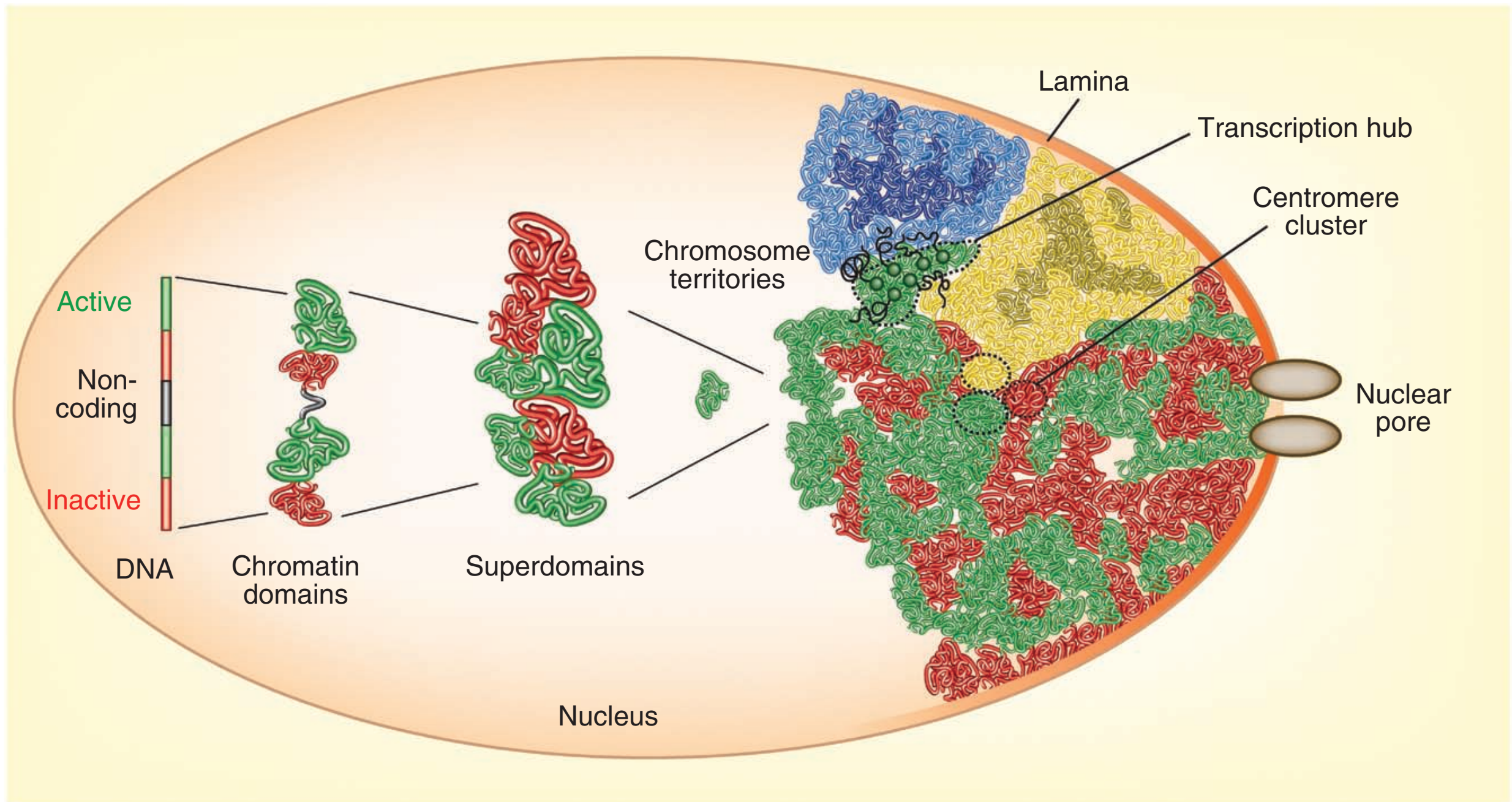
Complex genome organization

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



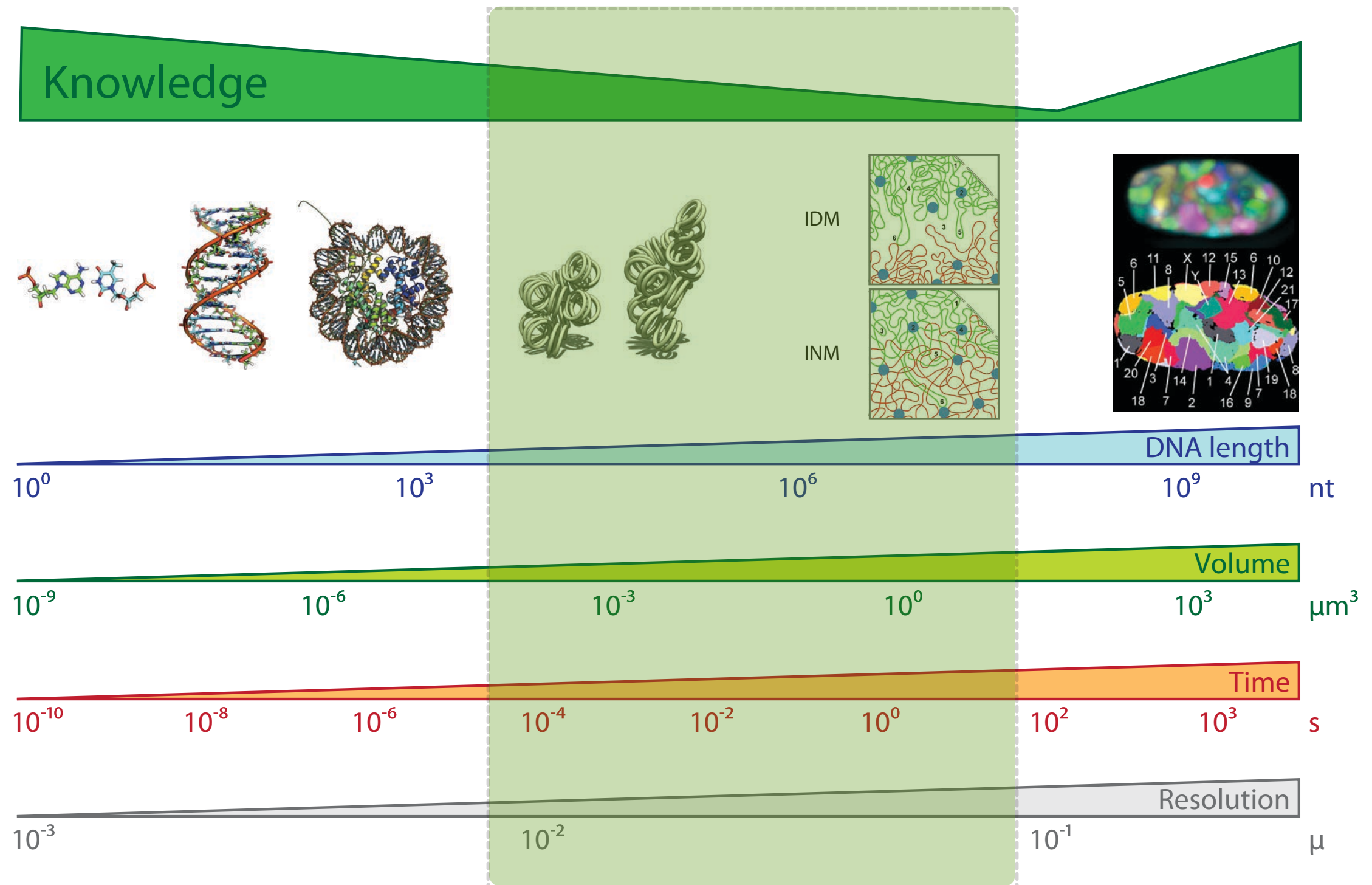
Complex genome organization

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



Resolution Gap

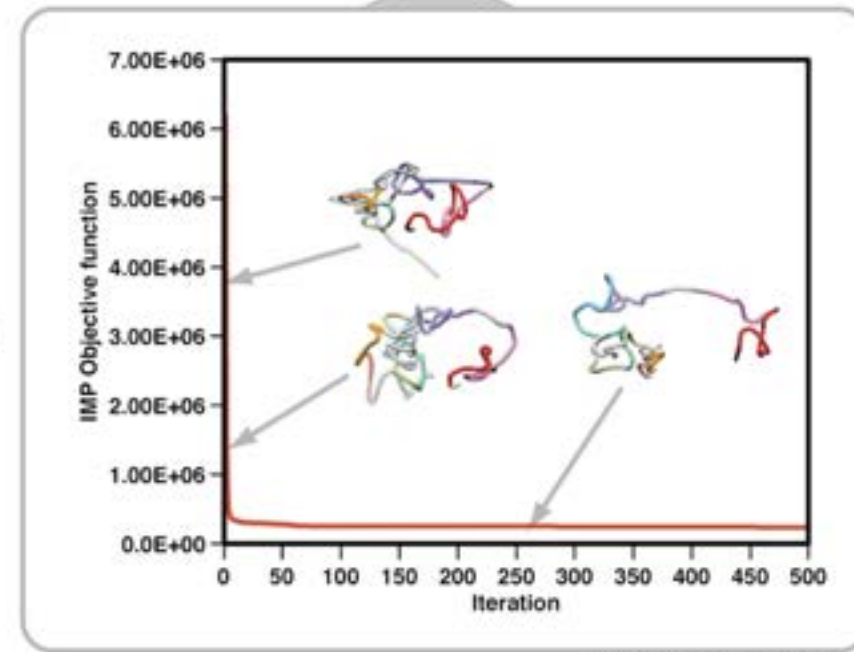
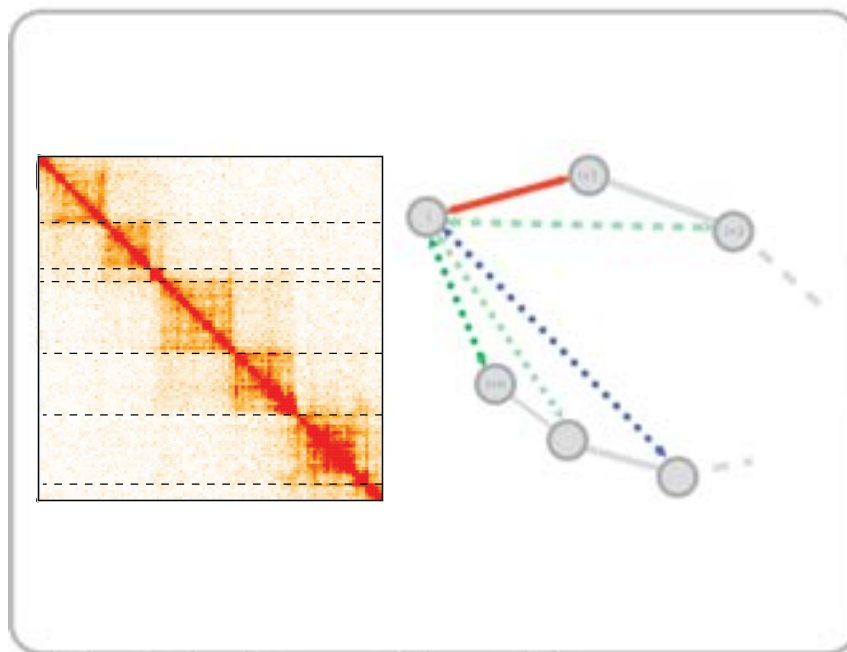
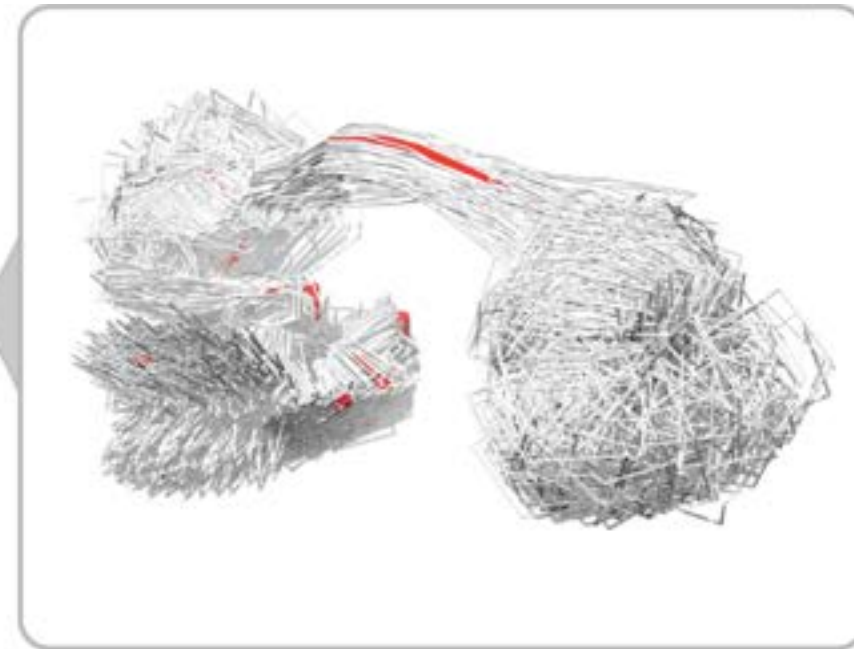
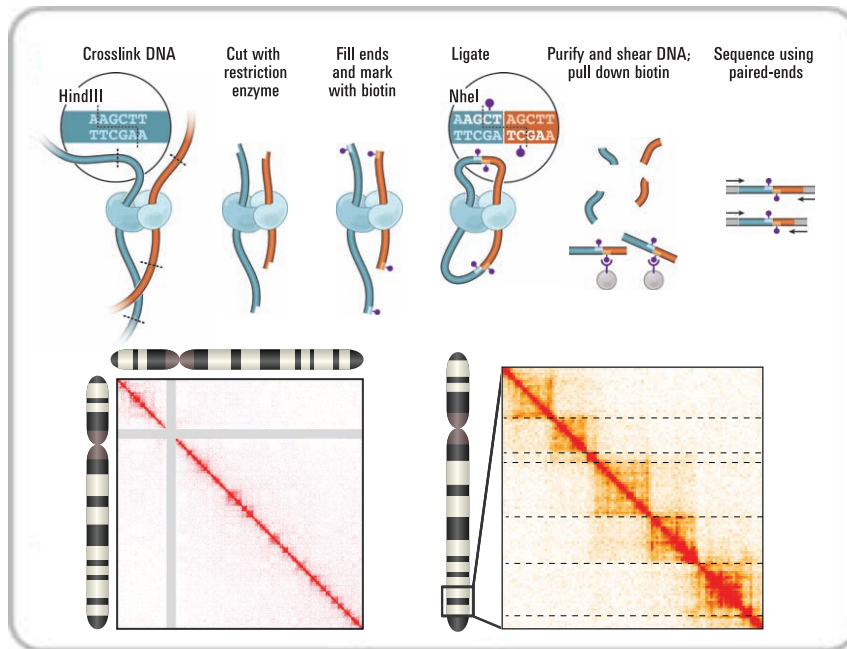
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol 7, e1002125 (2011)



Hybrid Method

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

Experiments

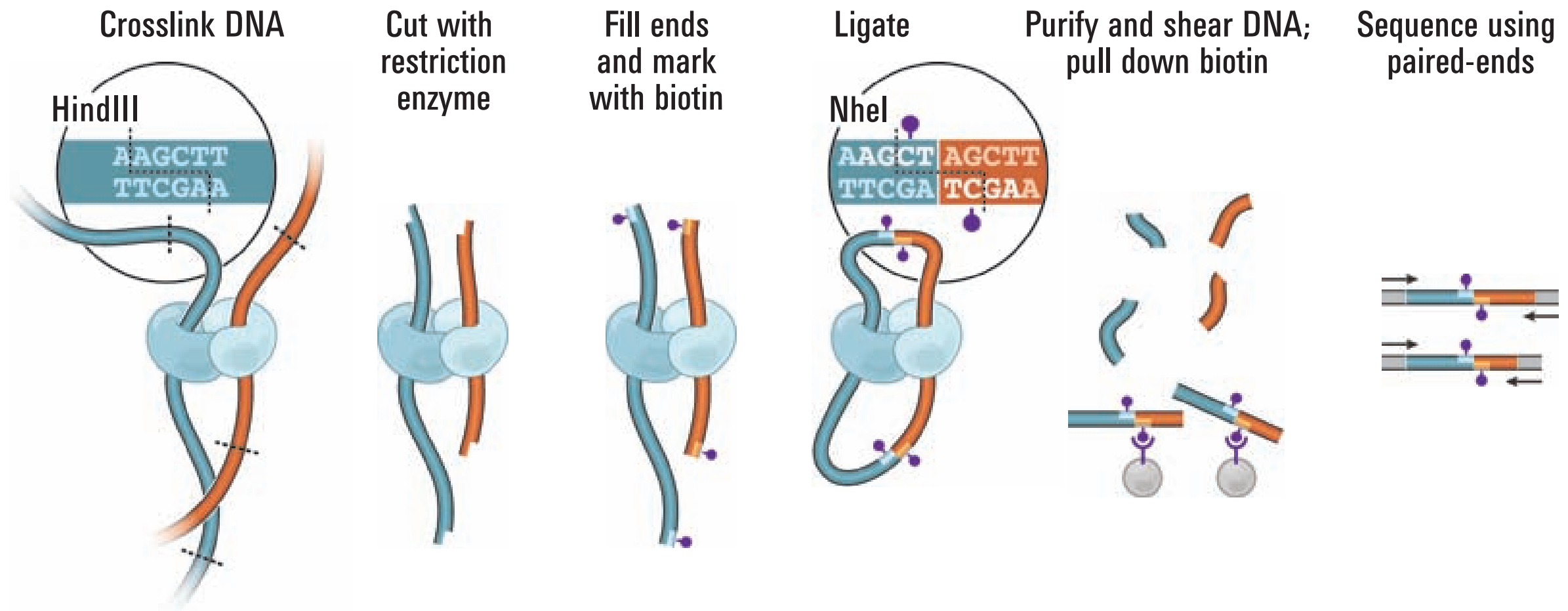


Computation

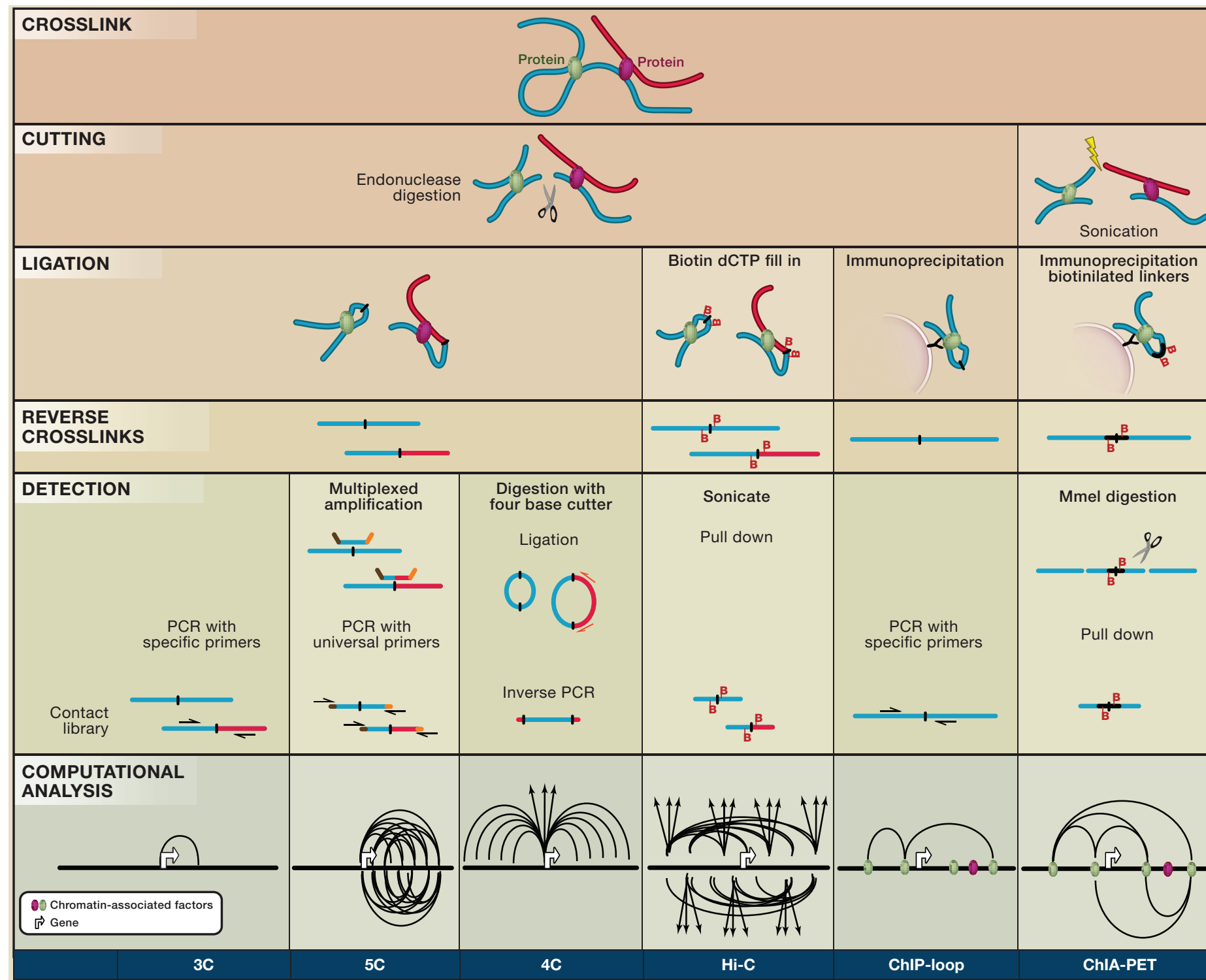
Chromosome Conformation Capture

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.

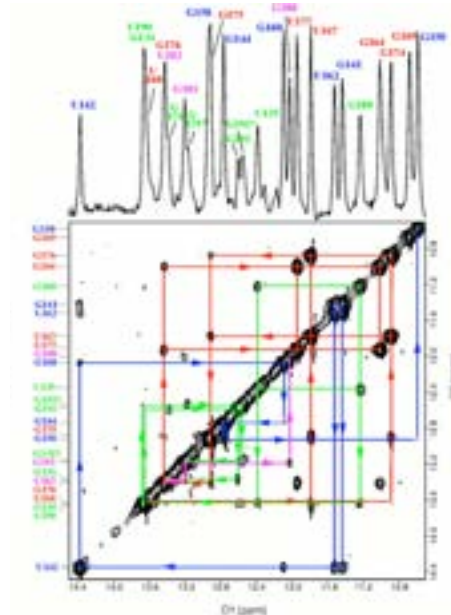
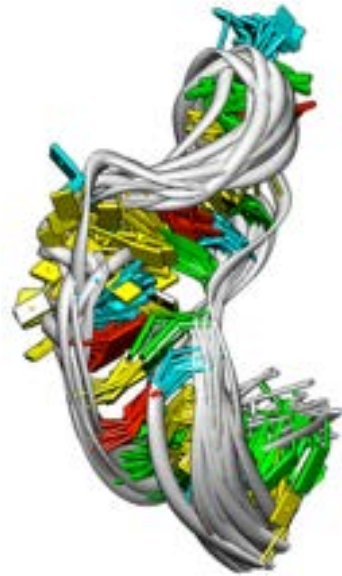


Chromosome Conformation Capture

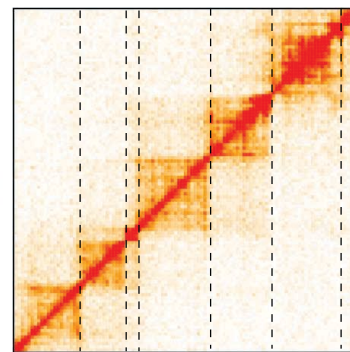
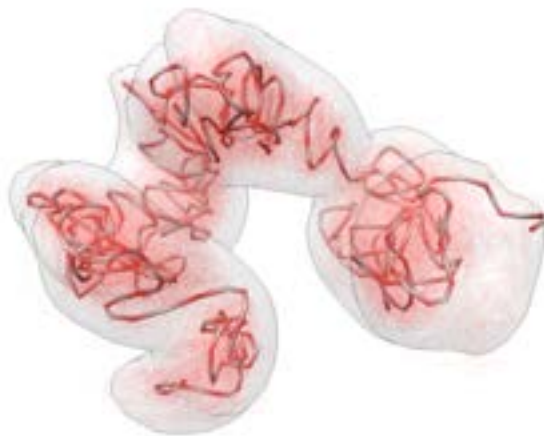


Restraint-based Modeling

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



Biomolecular structure determination
2D-NOESY data

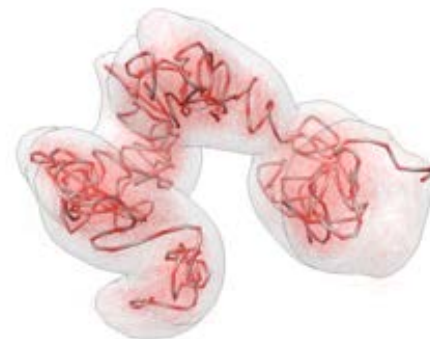
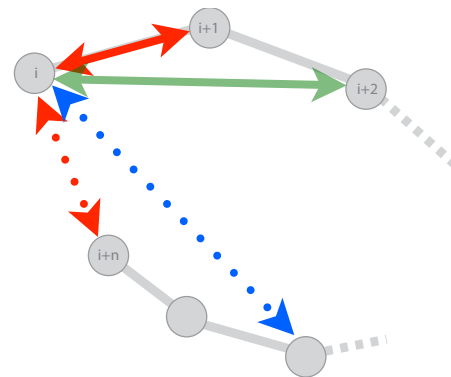
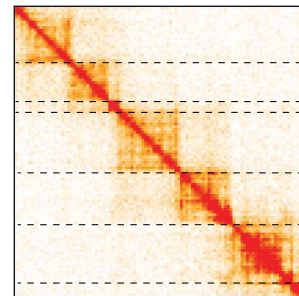


Chromosome structure determination
3C-based data



<http://3DGenomes.org>

Label Sequence
@FORJUSP02A3WD1
CCGTCAATTCAATTAGTTTAACTTGGGGCGTACTCCCAGGCGGT
*
AAAAAAAAA:99@: : ? ? @ : : FF AAAACCA: : : BB @ : ? A ?
Q scores (as ASCII chart)
Base=T, Q=?=25

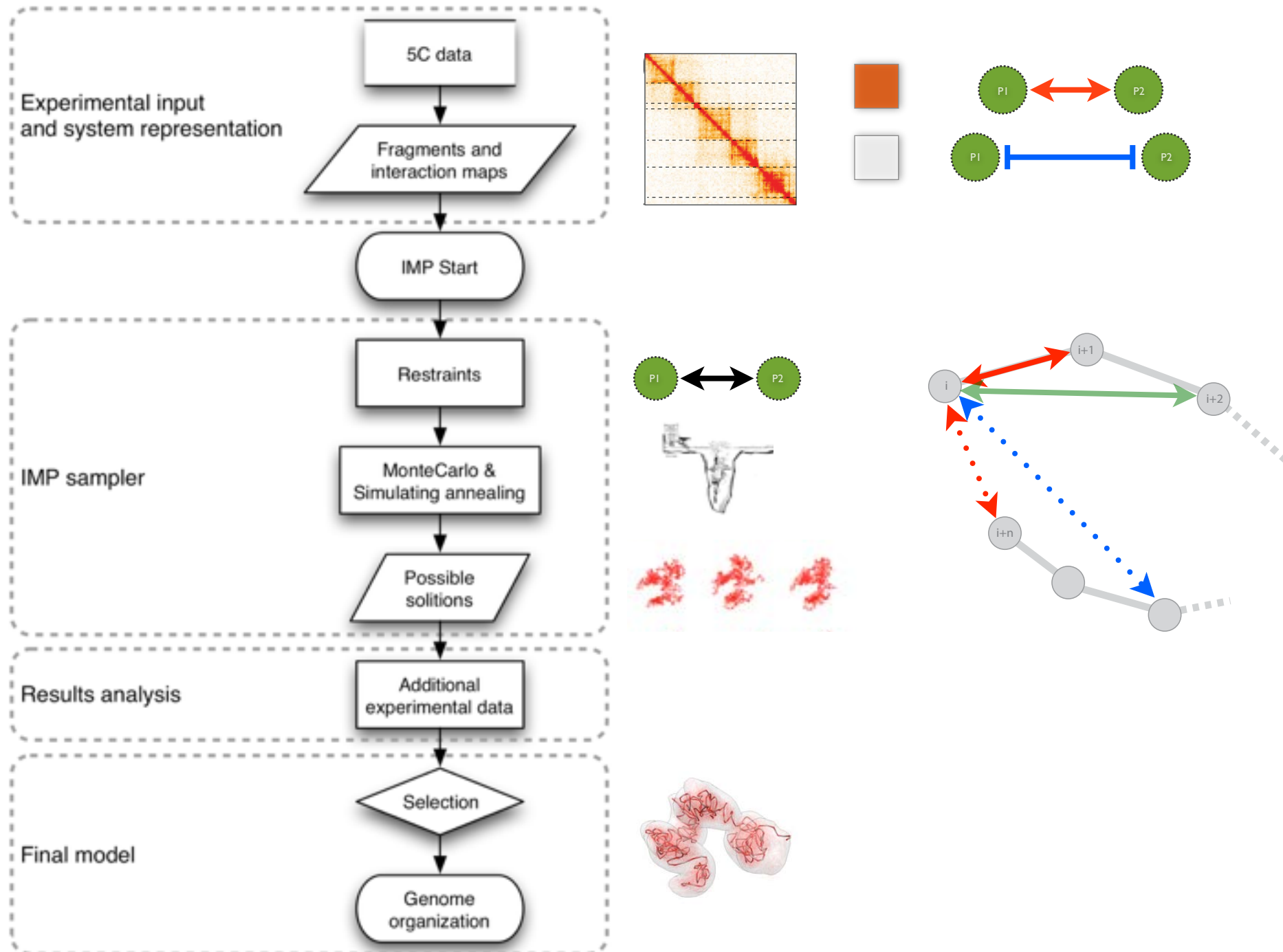


FastQ files to Maps

Map analysis

Model building

Model analysis

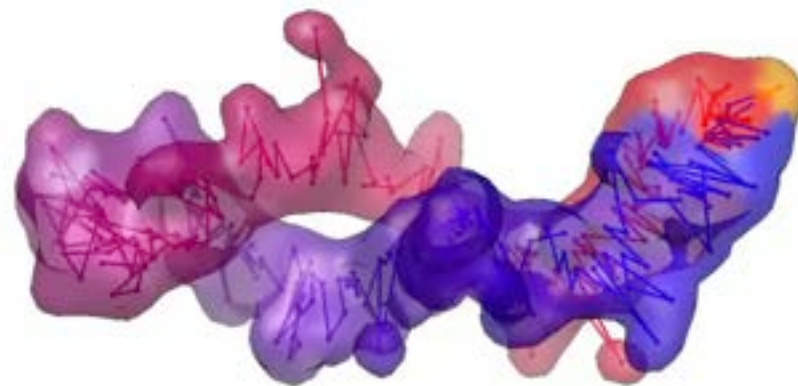
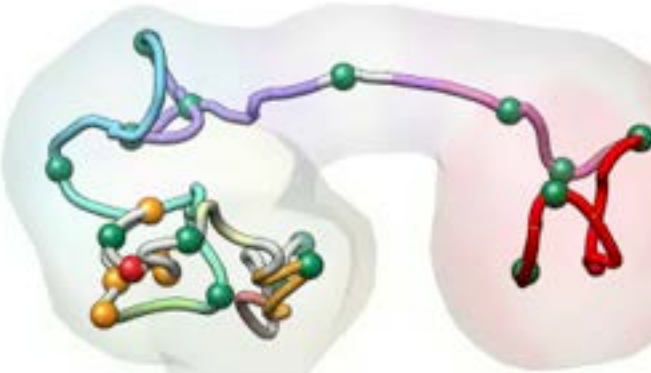
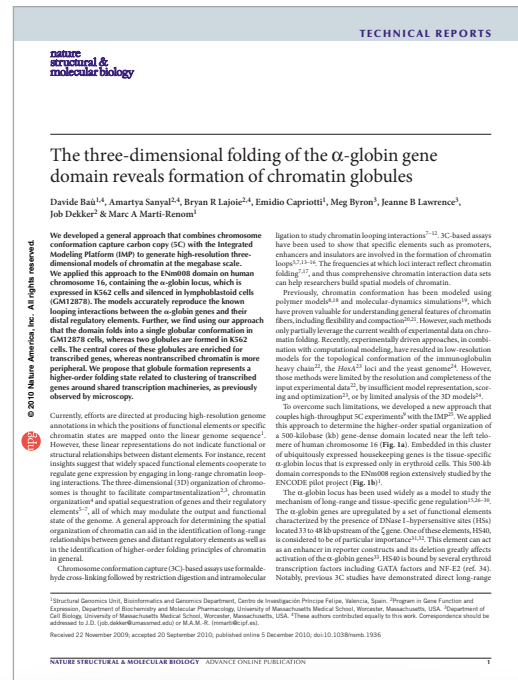


TADbit previous applications...

Baù, D. et al. Nat Struct Mol Biol (2011)

Umbarger, M. A. et al. Mol Cell (2011)

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



Cell

press

Molecular Cell
Article

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

Mark A. Umbarger,^{1,4,5} Esteban Toro,^{2,4} Matthew A. Wright,¹ Gregory J. Porreca,¹ Davide Baù,⁴ Sun-Hae Hong,^{2,4} Michael J. Ferro,³ Lihua J. Zhu,⁴ Marc A. Marti-Renom,^{4,6} Harley H. McAdams,⁷ Lucy Shapiro,⁸ Job Dekker^{1,4,7,9} and George M. Church¹

¹Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
²Department of Developmental Biology, School of Medicine
³Department of Physics, School of Humanities and Sciences
⁴Stanford University, Stanford, CA 94305, USA
⁵Structural Genomics Laboratory, Bioinformatics and Genomics Department, Centro de Investigación Príncipe Felipe, 46102 Valencia, Spain
⁶Program in Gene Function and Expression
⁷University of Massachusetts Medical School, Worcester, MA 01605, USA
⁸Department of Biochemistry and Molecular Pharmacology
⁹Program in Systems Biology

These authors contributed equally to this work.
*Correspondence: umbarger@post.harvard.edu (M.A.U.), mmarti@cpf.es (M.A.M.-R.), job.dekker@umassmed.edu (J.D.)
DOI: 10.1016/j.molcel.2011.08.010

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 (3C), 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

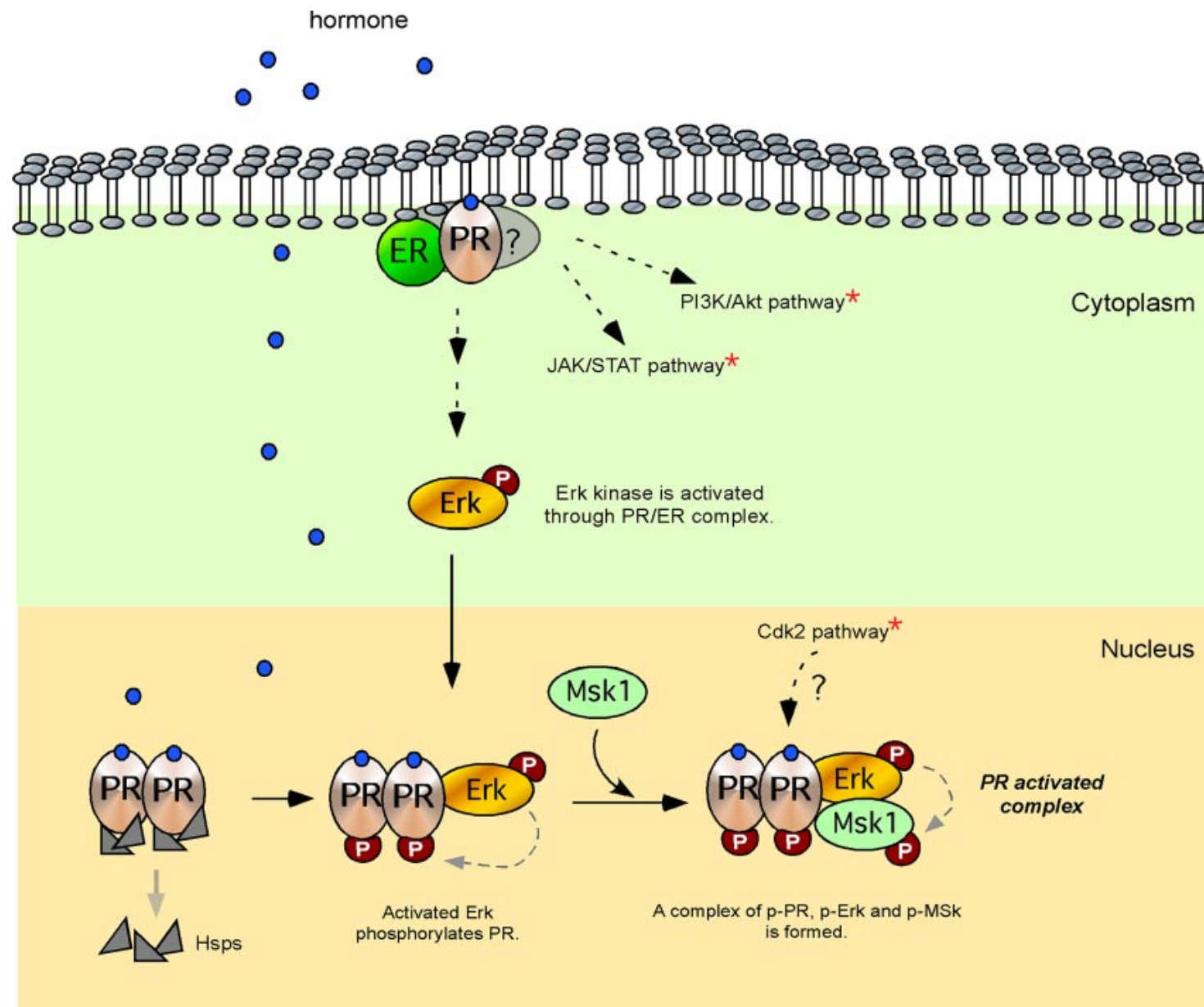
The three-dimensional (3D) architecture of the genome both reflects and regulates its functional state (Dekker, 2006; Thackerier and Shapiro, 2006a). For example, chromosome segregation impacts bacterial locus subcellular positioning (Lin and Mulder, 2008; White et al., 2008), and chromatin loops that place promoters and distant enhancers within close spatial proximity play important roles in eukaryotic transcriptional regulation (Tofais et al., 2002; Vermeulen 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 (Vollmer et al., 2004) and chromosome conformation capture (3C)-based approaches (Dekker et al., 2002; Dostie et al., 2006; Duan et al., 2010; Fulwood et al., 2009; Lieberman-Aiden et al., 2009; Simonis et al., 2008; Zhou et al., 2008), 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 interloca 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., 2008) and the correlation between a local genomic and subcellular positioning in bacteria (Nielsen et al., 2008; Teisman et al., 1998; Wang et al., 2006). 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

252 Molecular Cell 44, 252–264, October 21, 2011 ©2011 Elsevier Inc.

Progesterone-regulated transcription in breast cancer

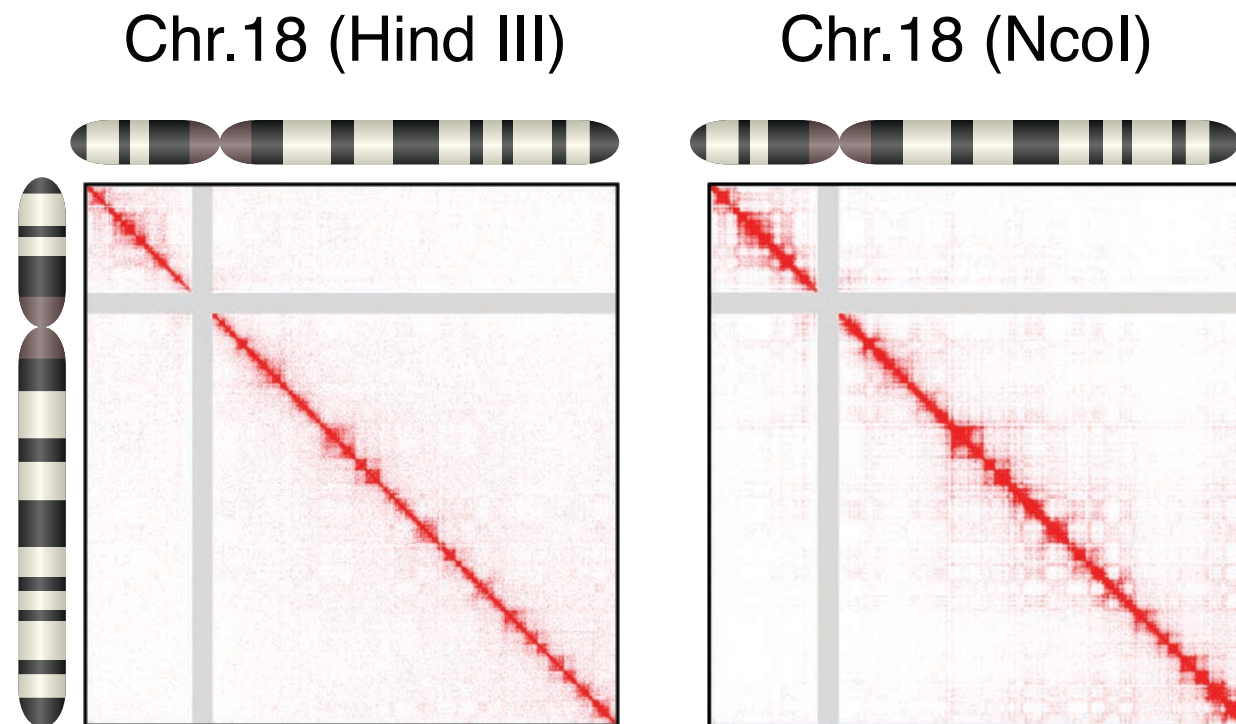
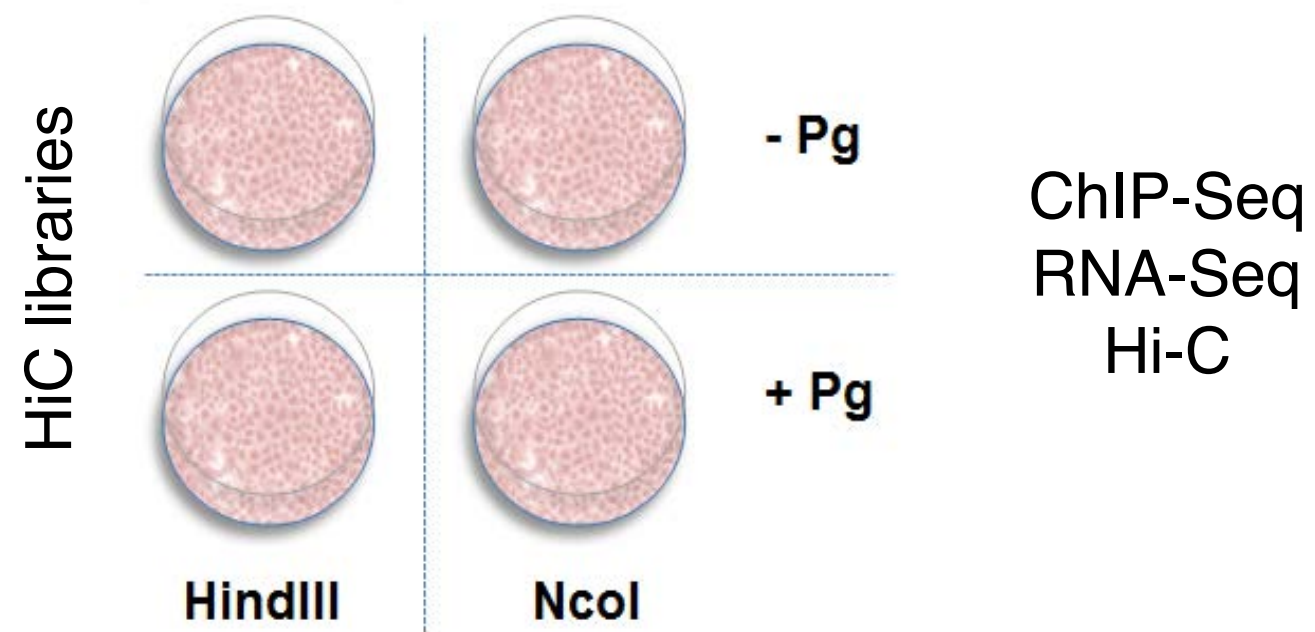


> 2,000 genes **Up**-regulated
> 2,000 genes **Down**-regulated

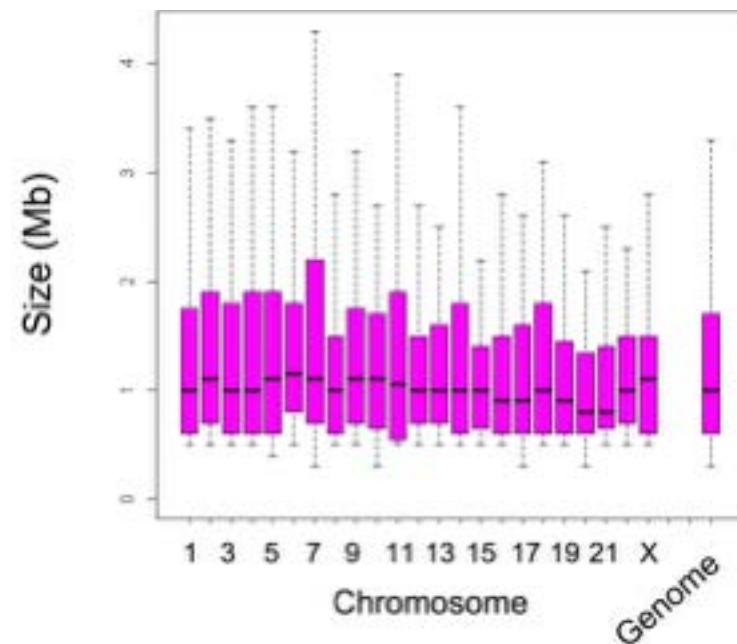
Regulation in 3D?

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

Experimental design



Are there TADs? how robust?



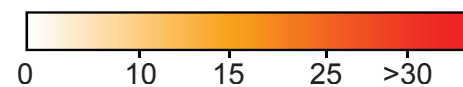
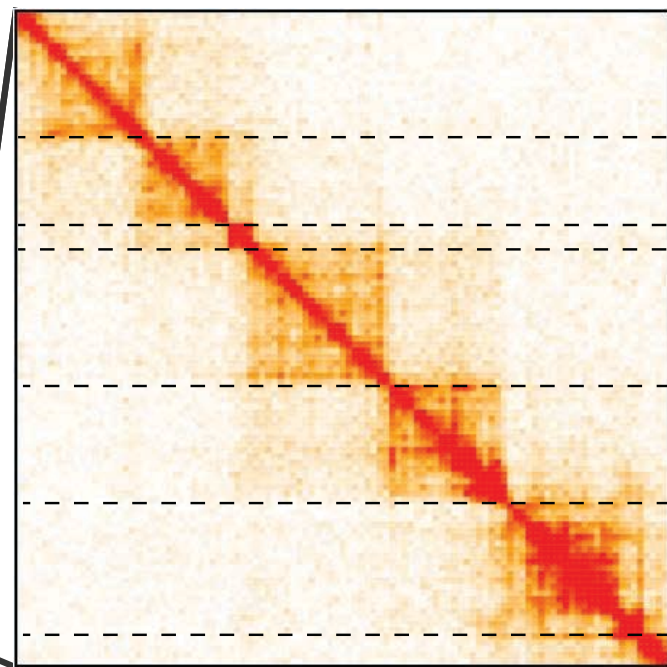
>2,000 detected TADs



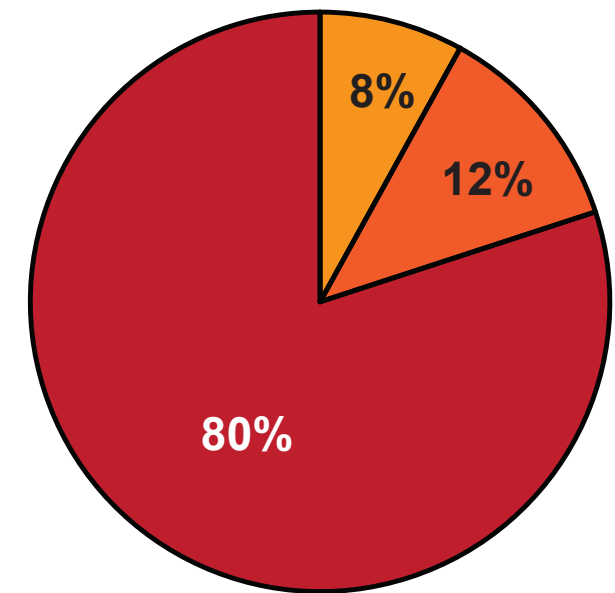
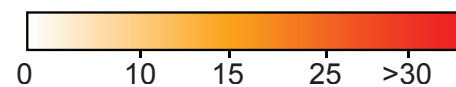
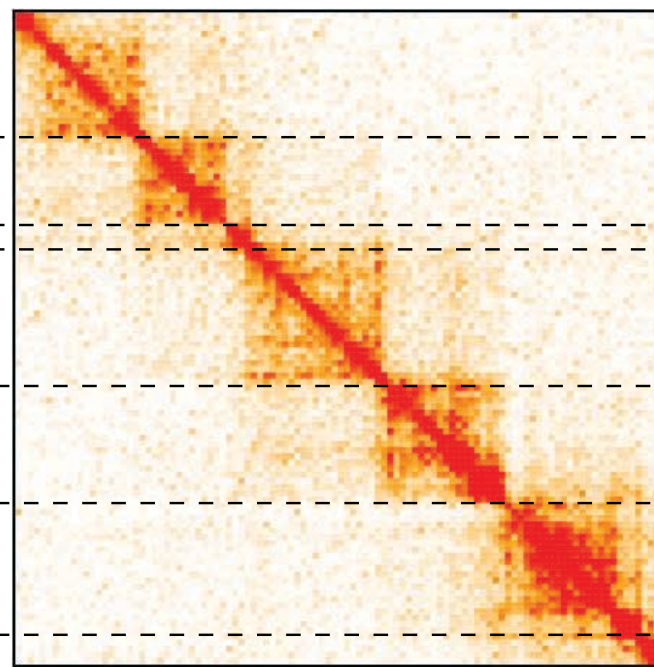
Chr.18



-Pg

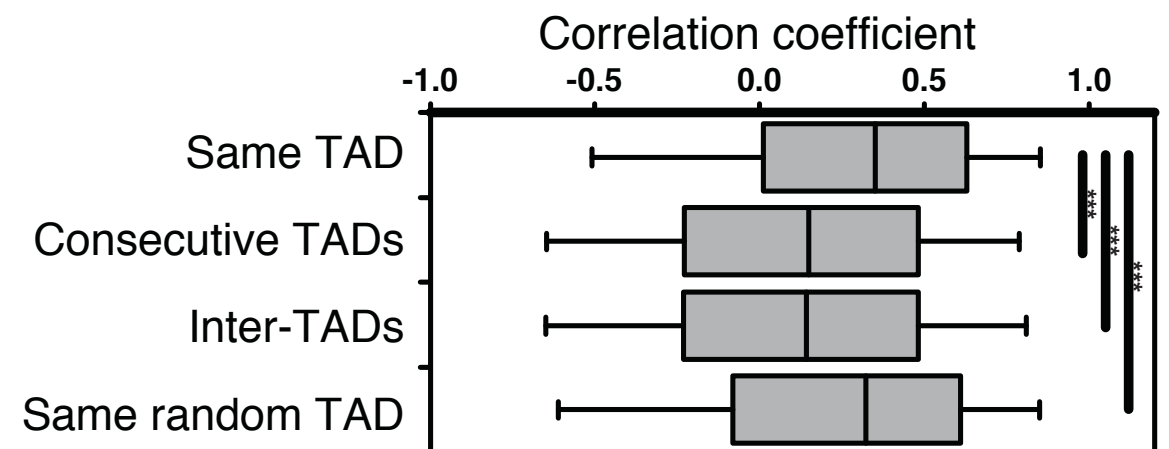
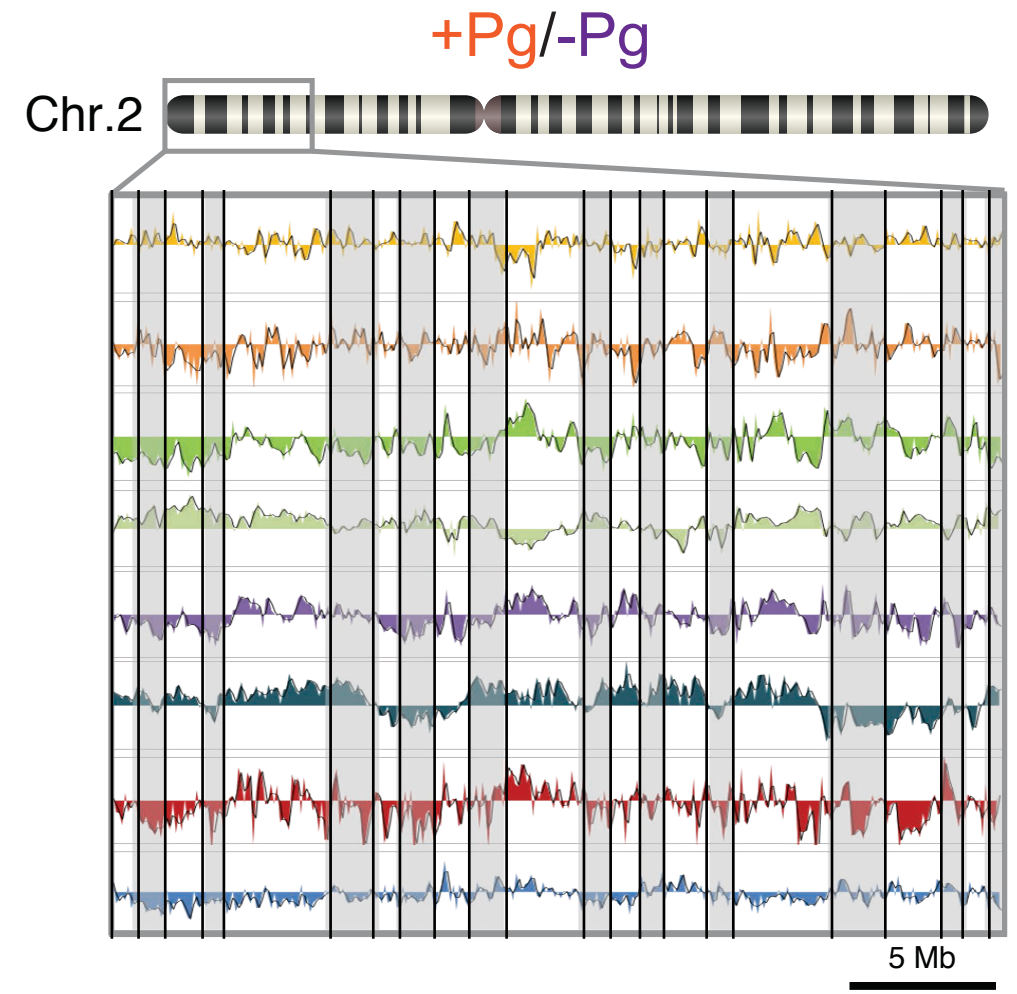
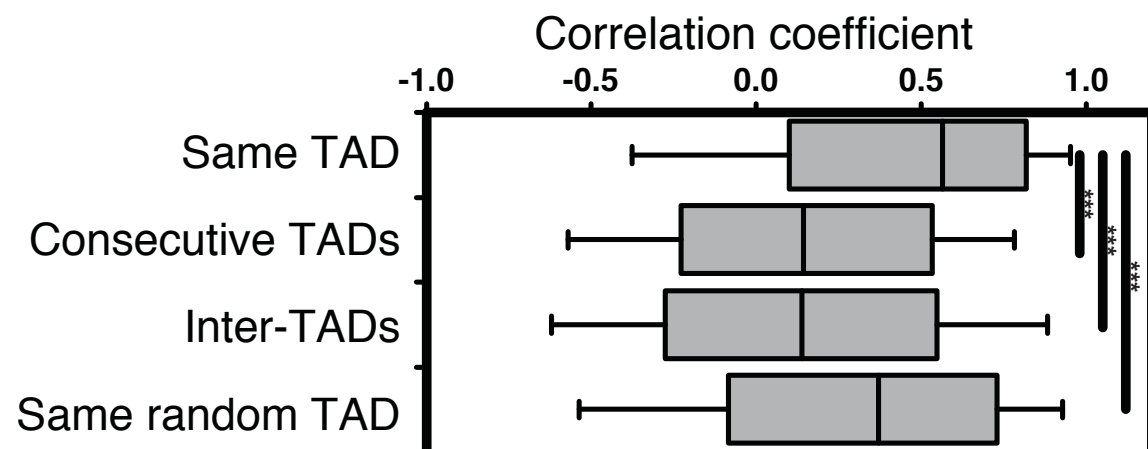
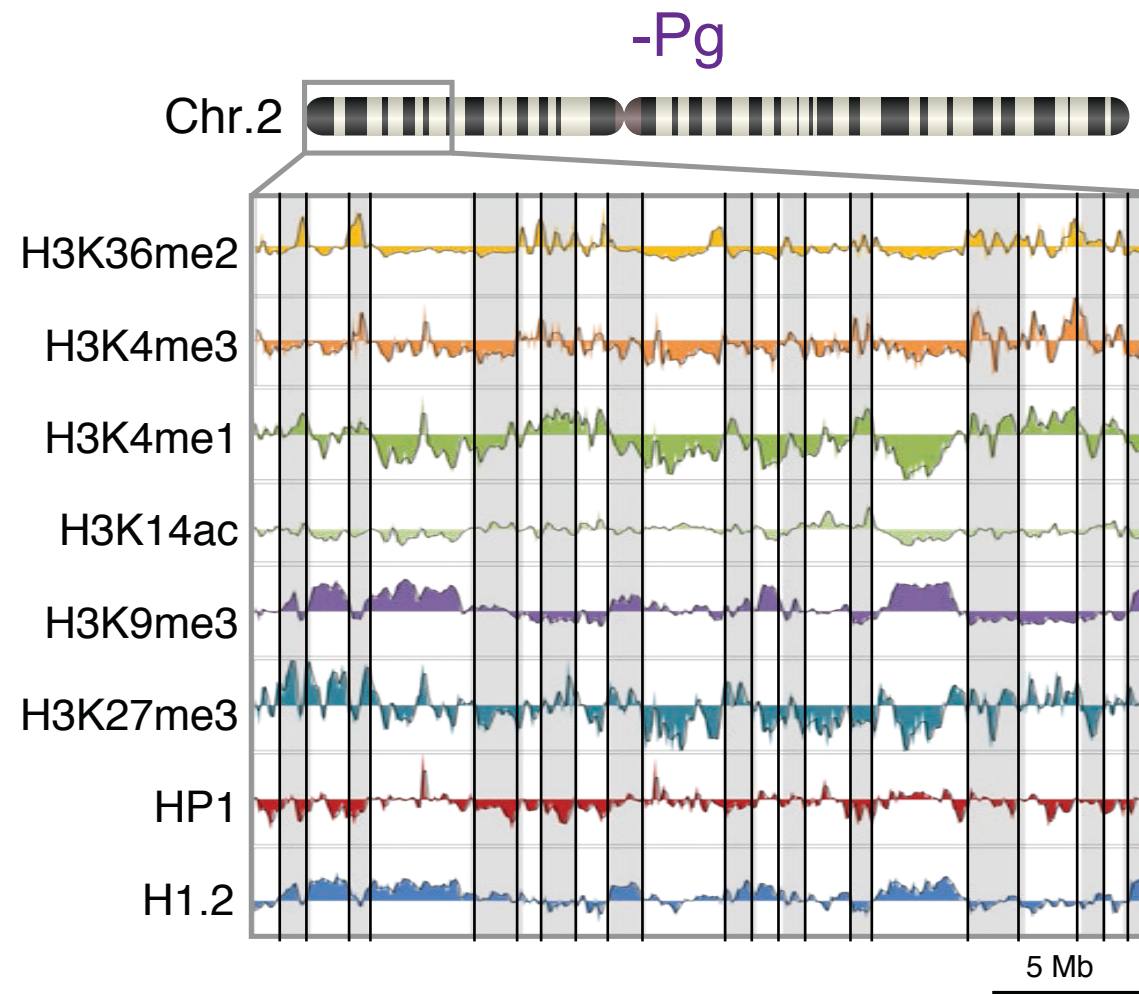


+Pg

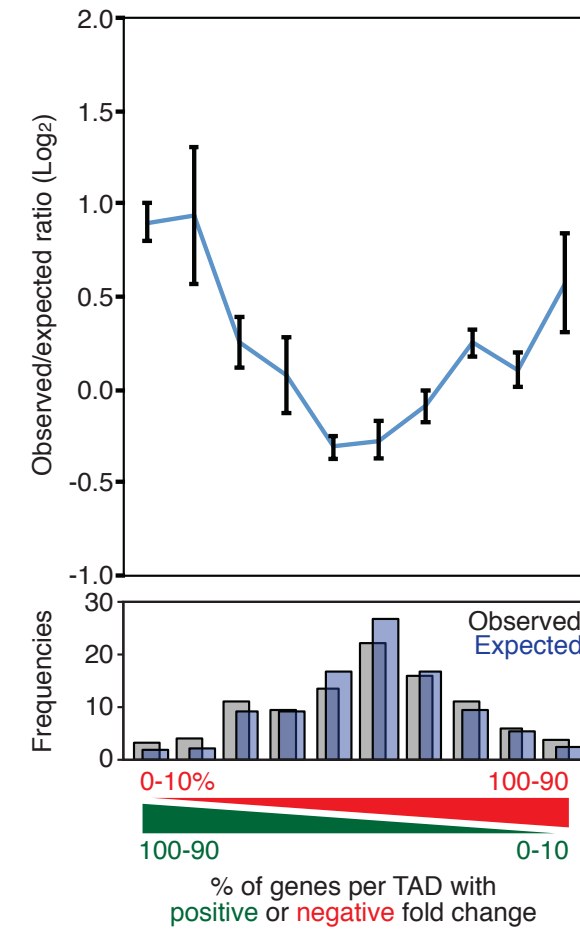
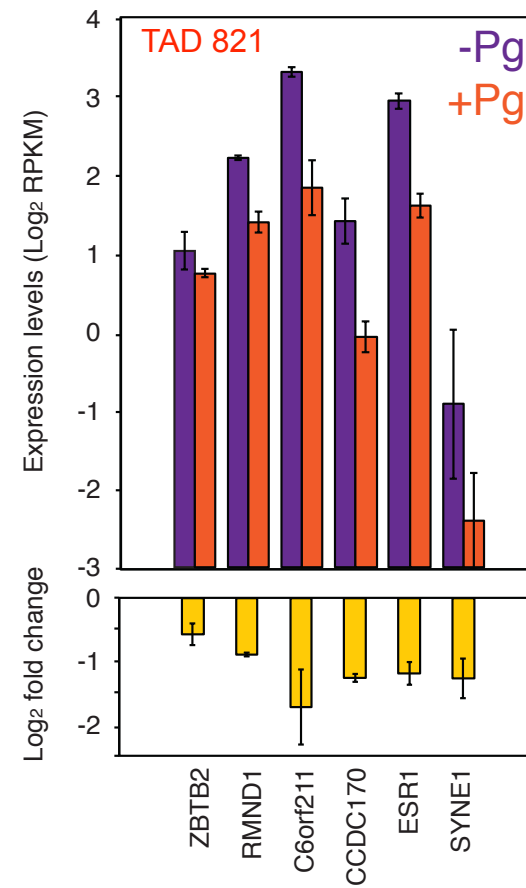
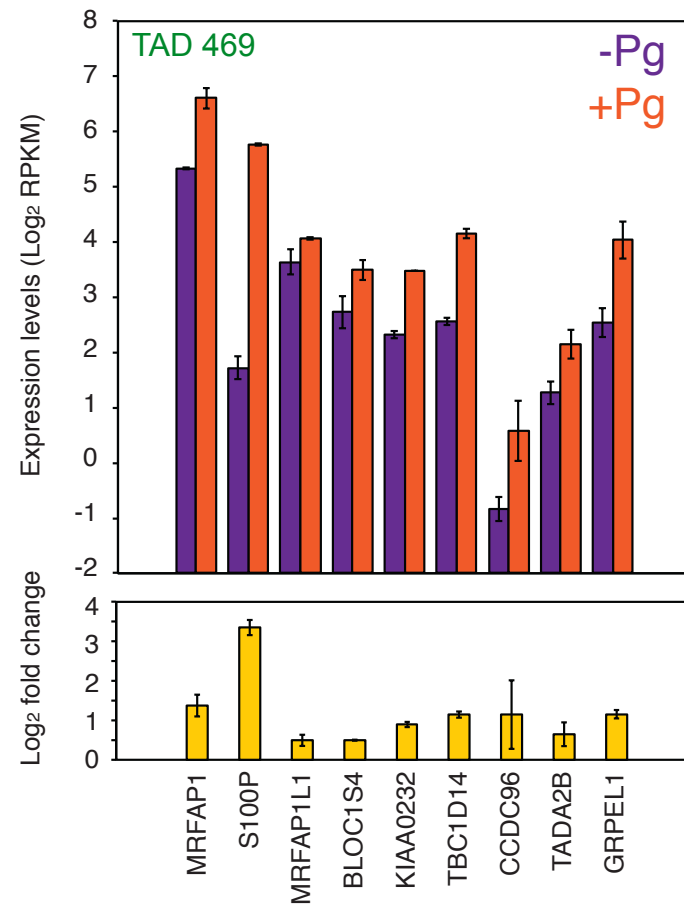


conserved
100 kb
±200 kb or more

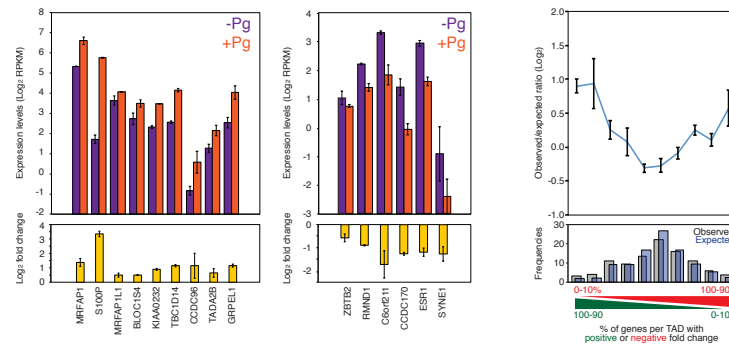
Are TADs homogeneous?



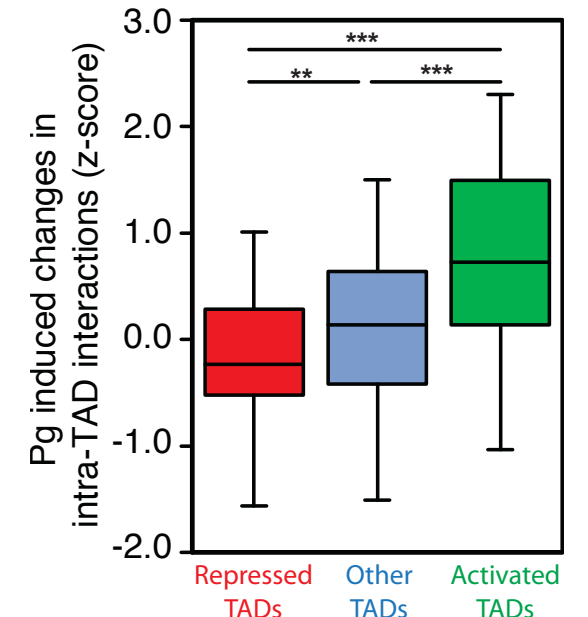
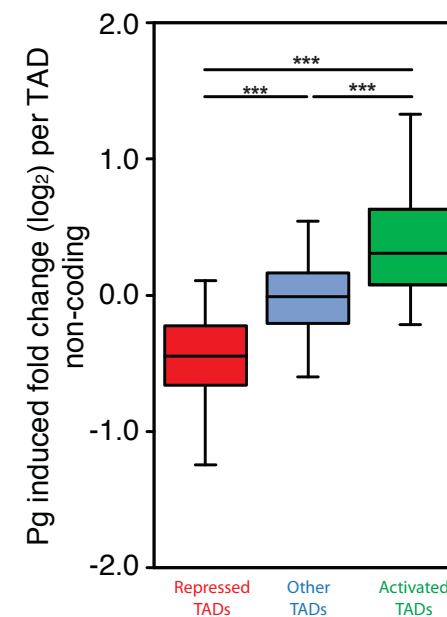
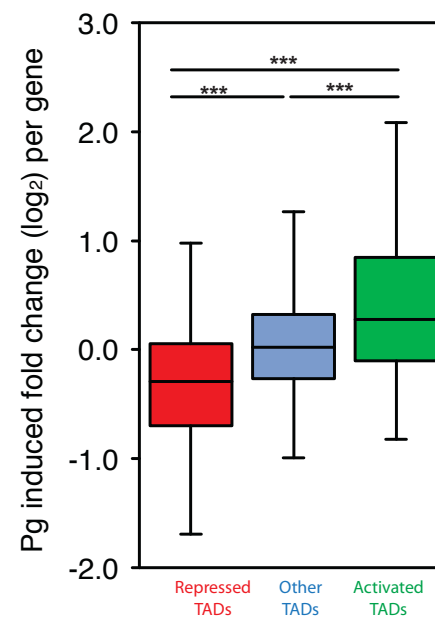
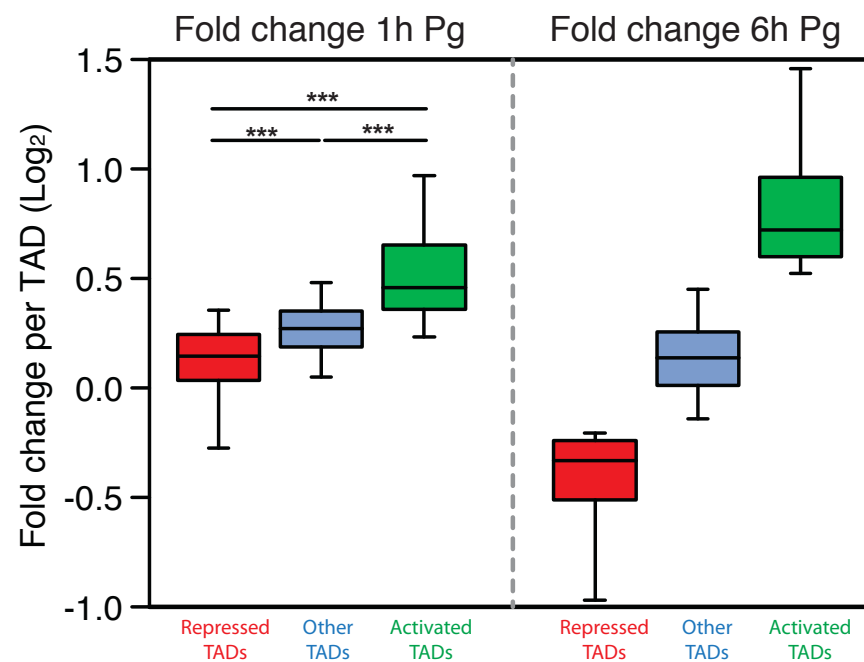
Do TADs respond differently to Pg treatment?



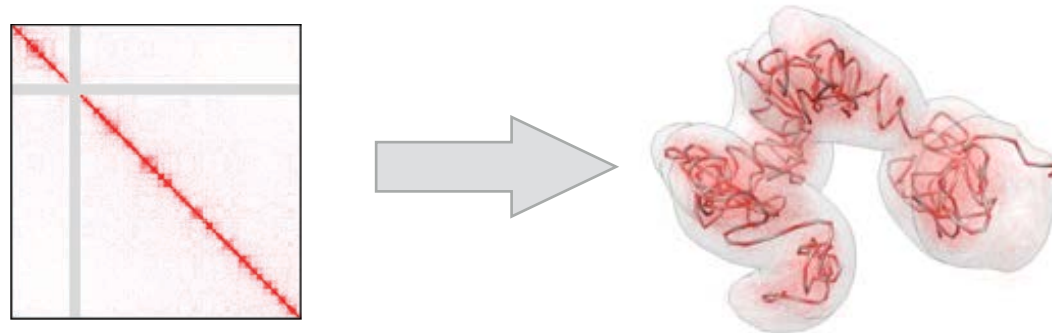
Do TADs respond differently to Pg treatment?



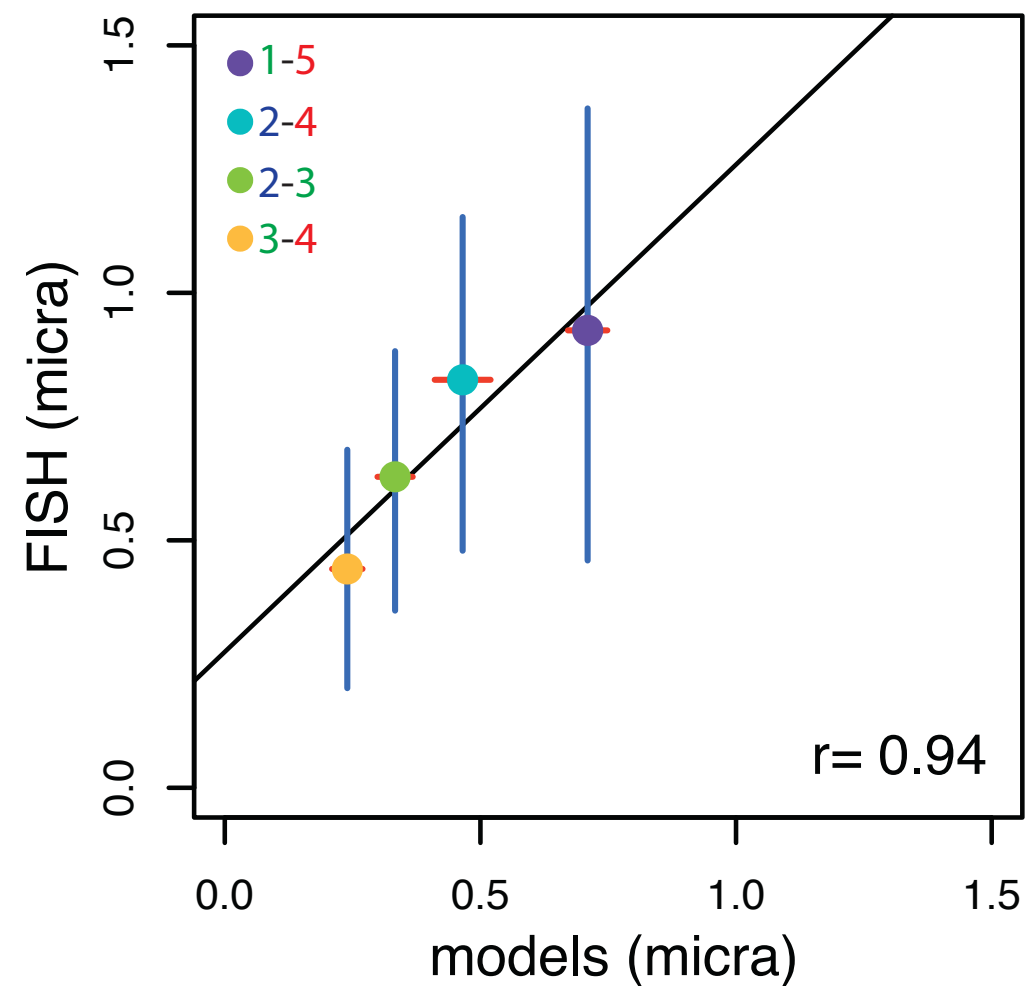
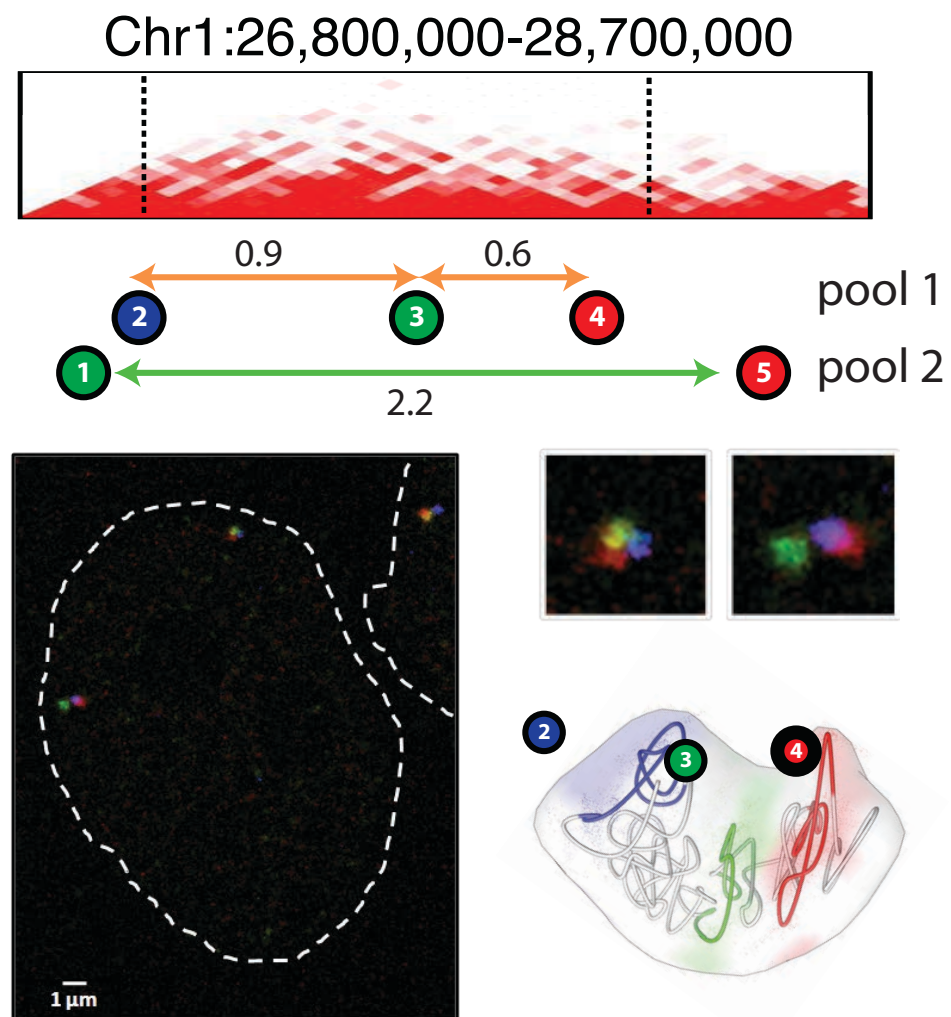
Pg induced fold change per TAD (6h)



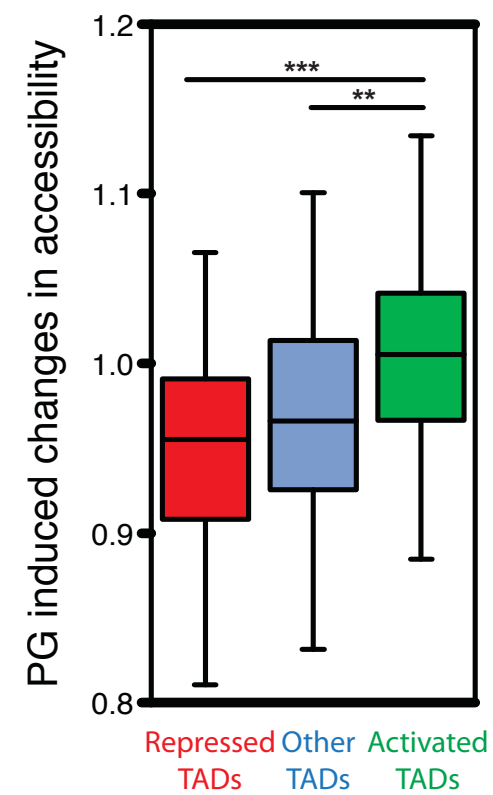
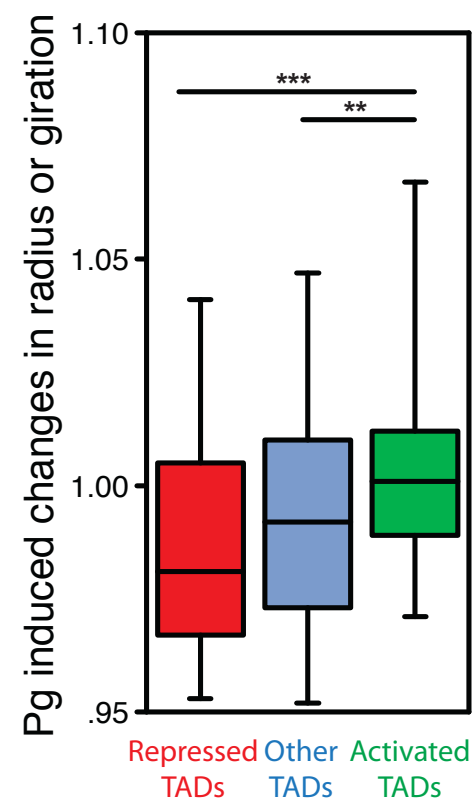
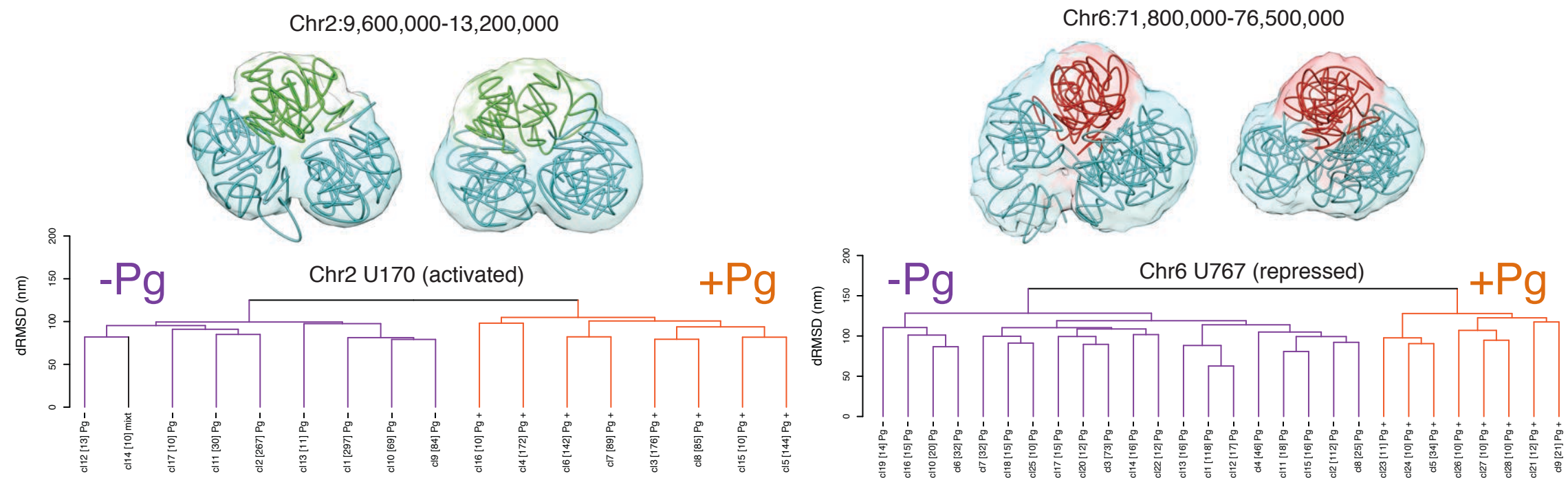
Modeling 3D TADs



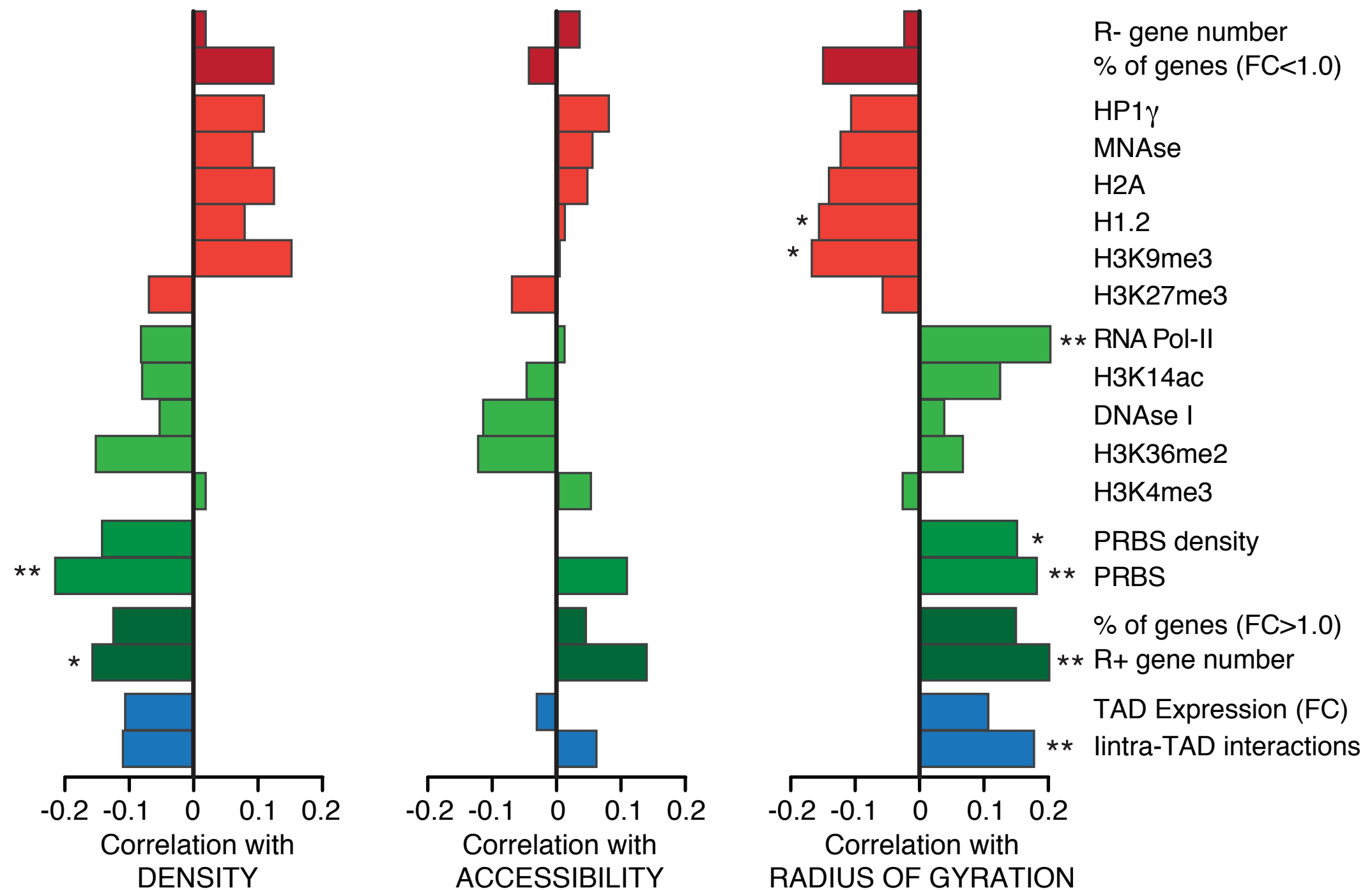
61 genomic regions containing 209 TADs covering 267Mb

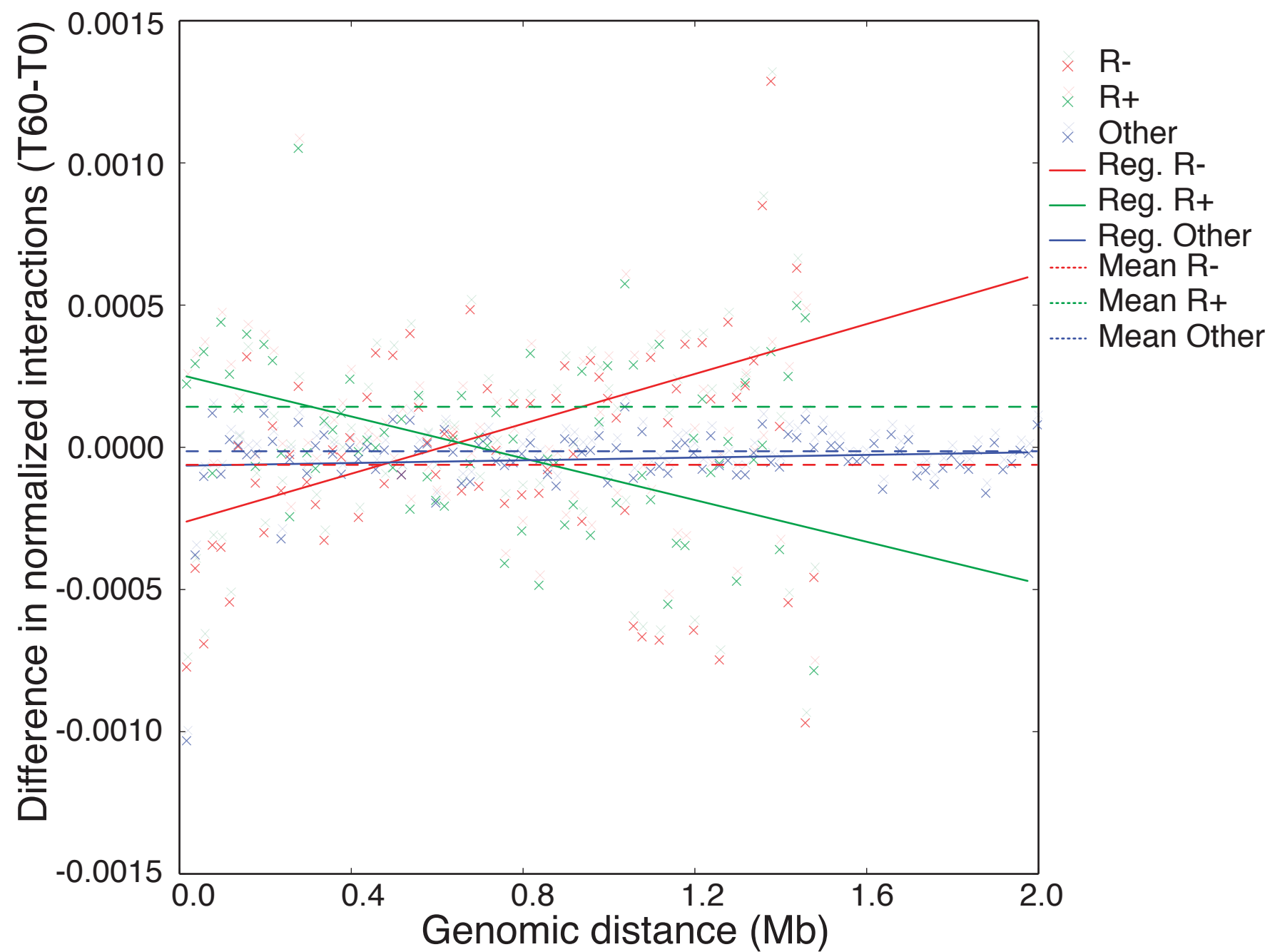
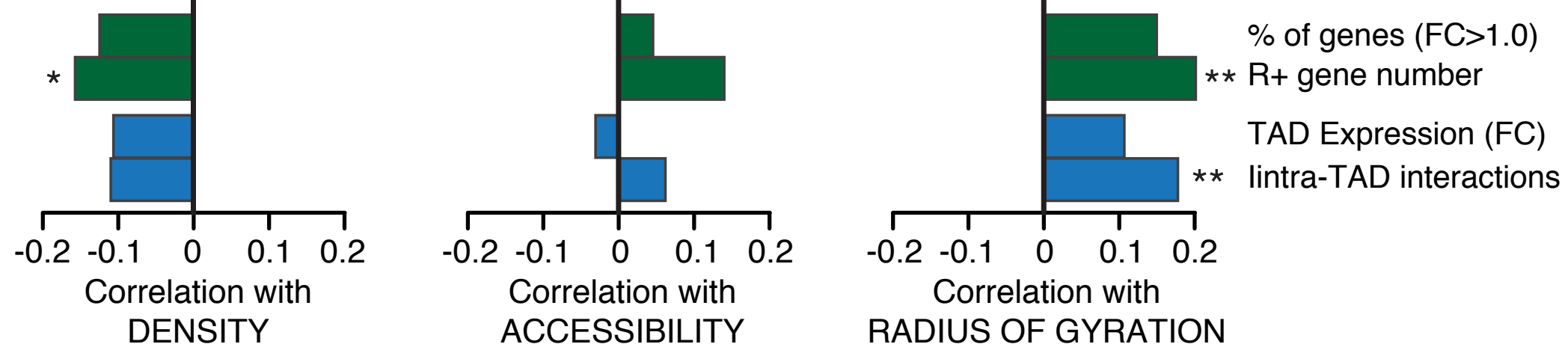


How TADs respond structurally to Pg?



How TADs respond structurally to Pg?

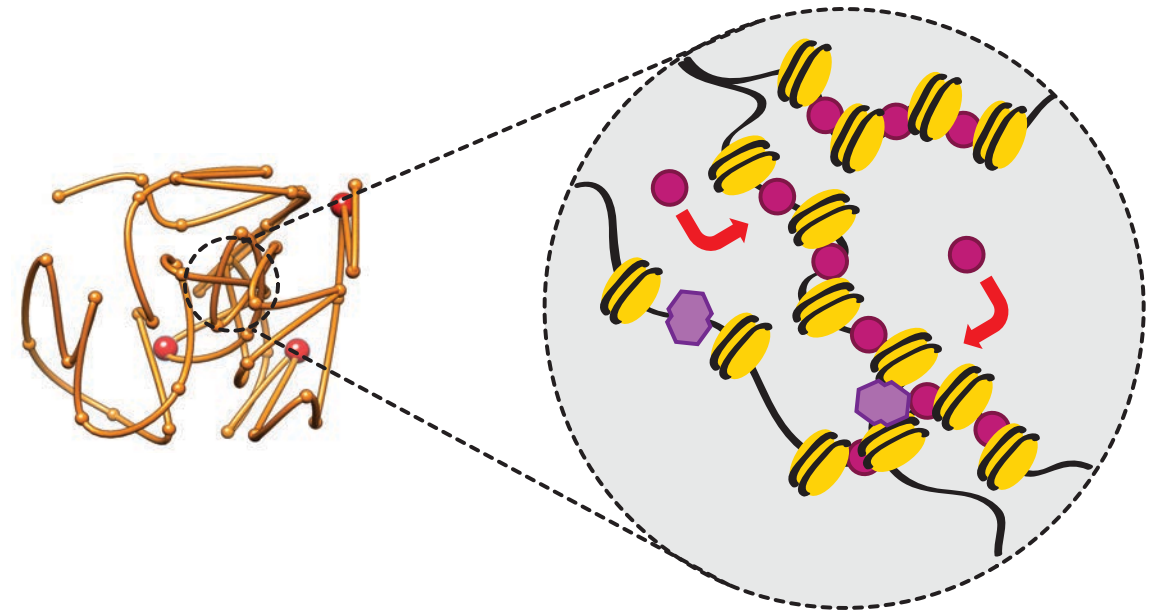
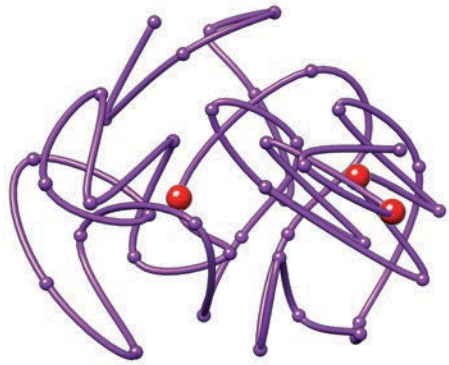




Model for TAD regulation

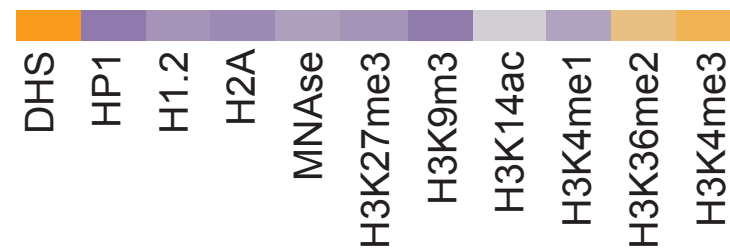
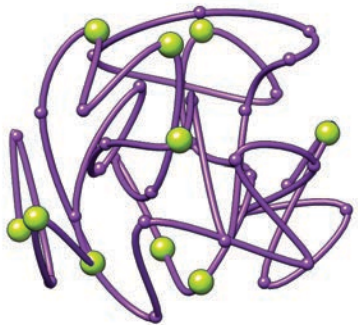
Repressed TAD

chr1 U41



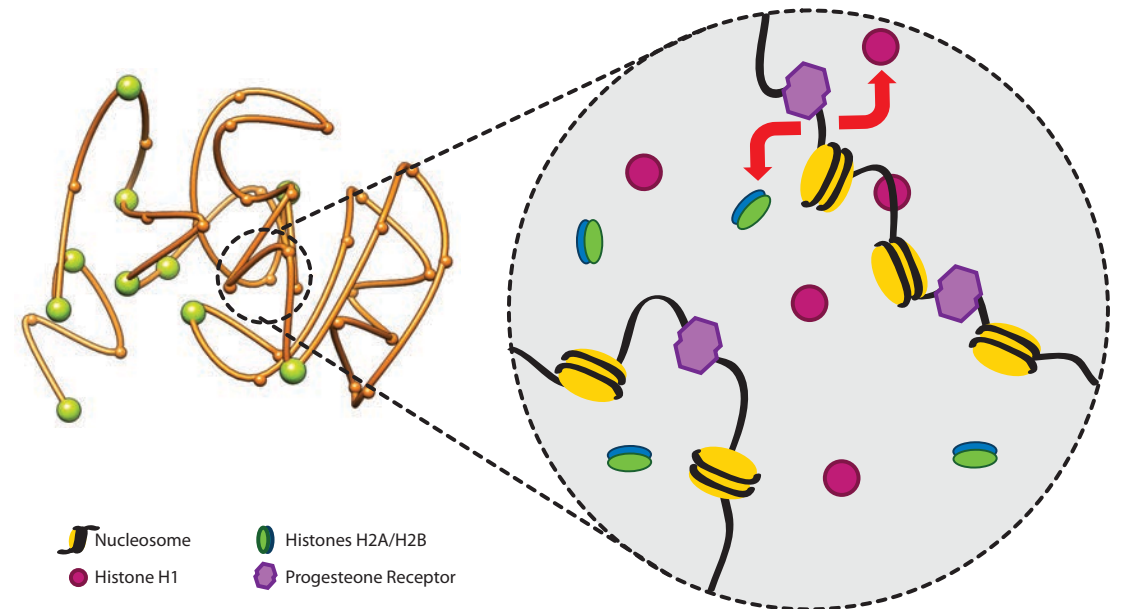
Activated TAD

chr2 U207



Structural
transition

+Pg

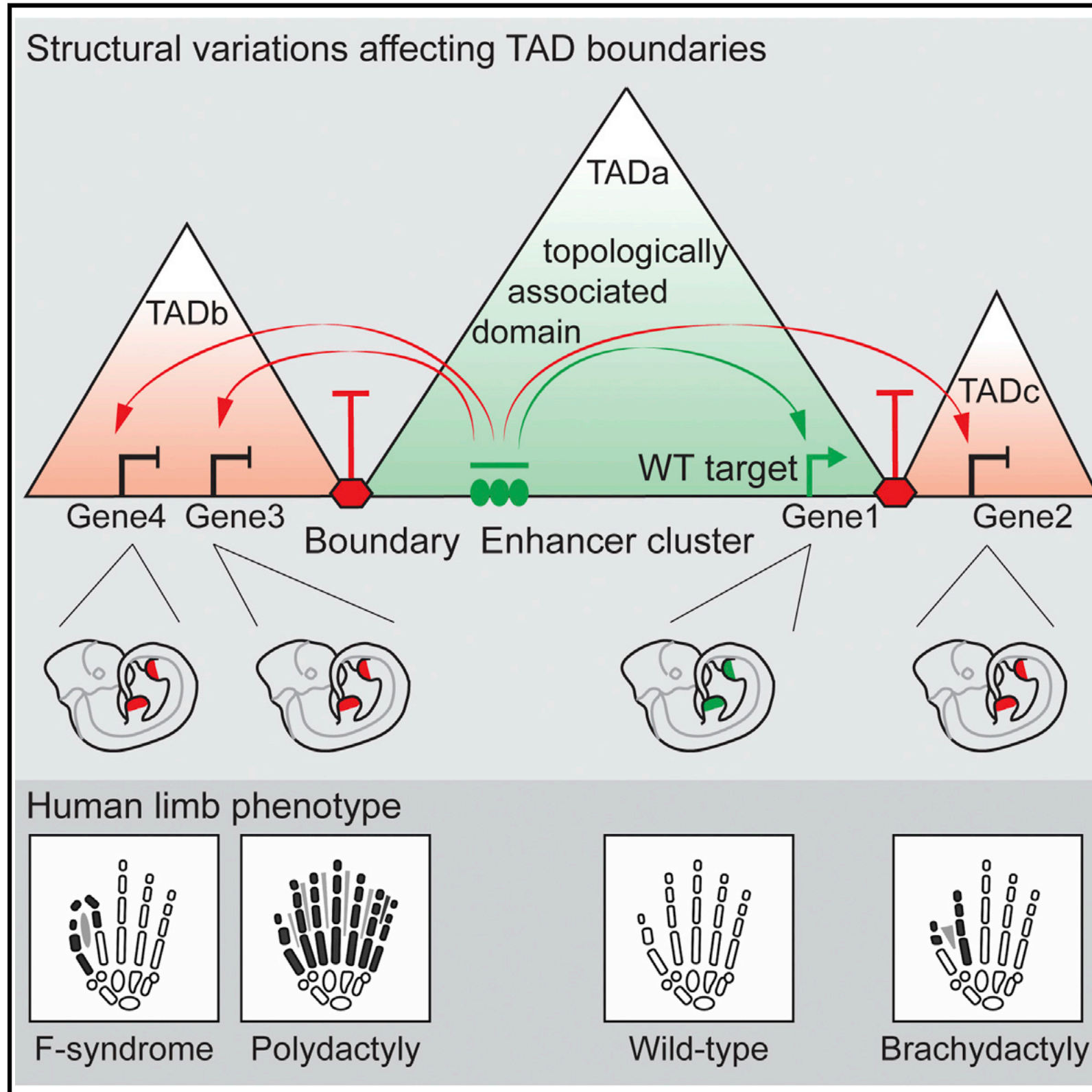


Nucleosome
Histone H1

Histones H2A/H2B
Progesterone Receptor

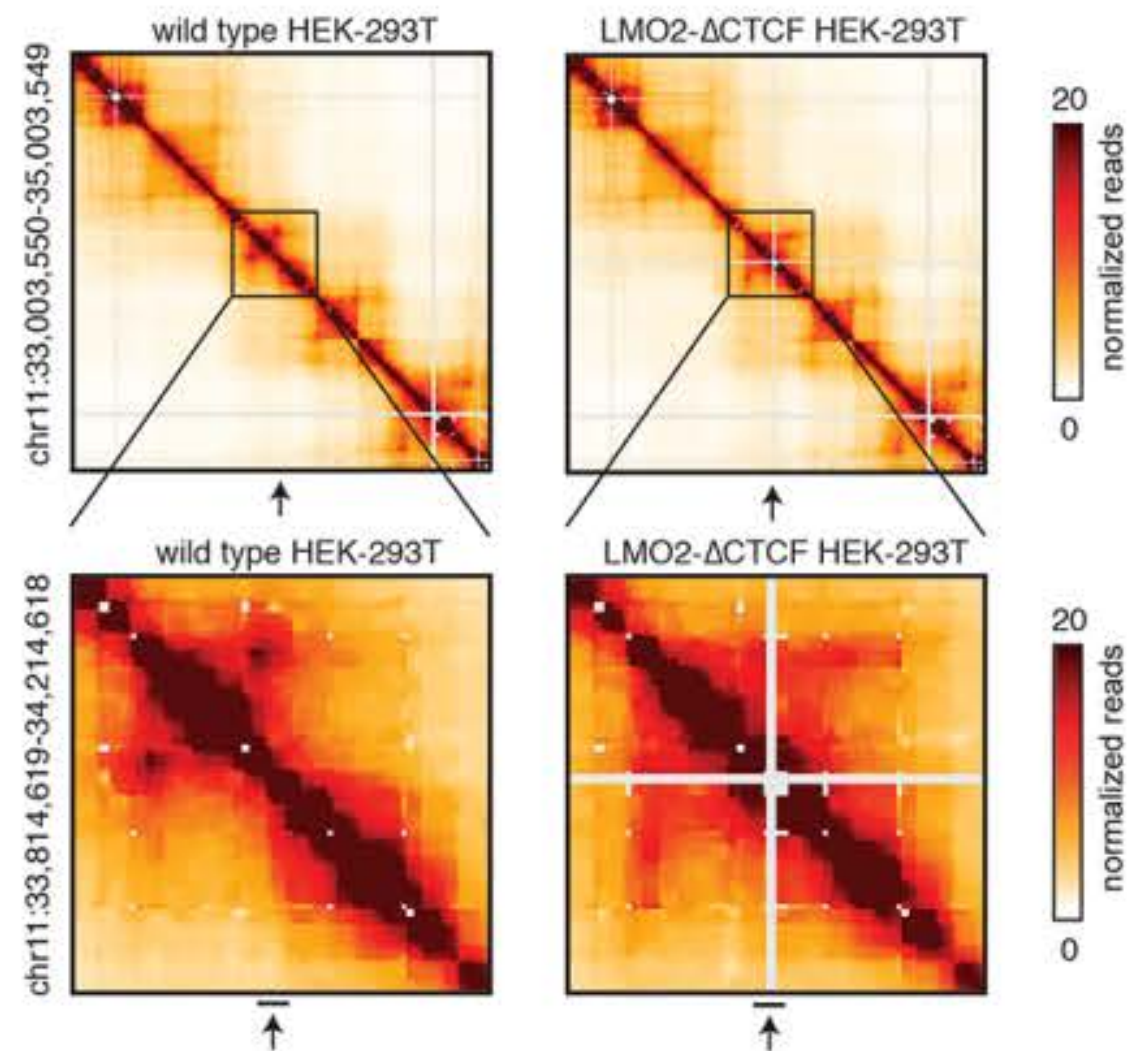
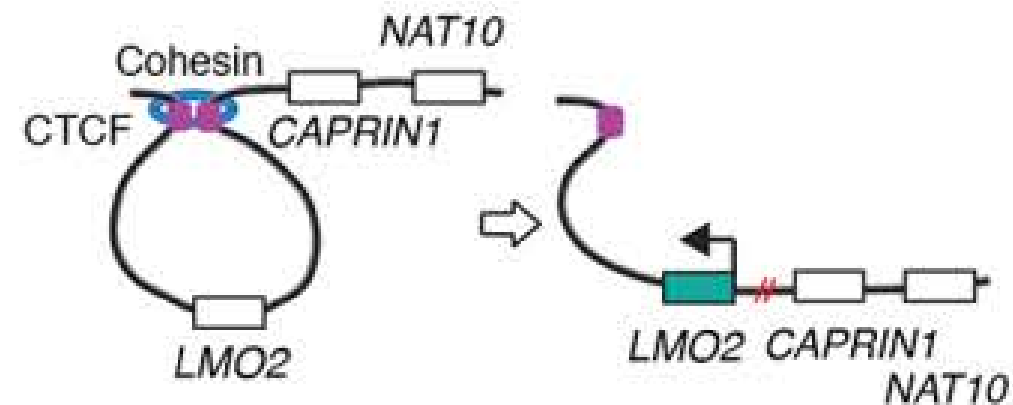
TADs are functional units

Lupiáñez, et al. (2015). Cell, 1–15.

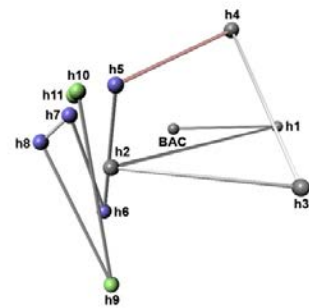


TADs are functional units

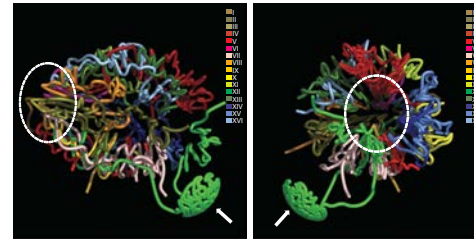
Hnisz, D., et al. (2016). Science, on line



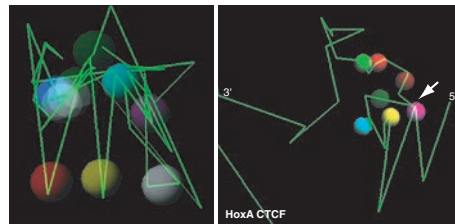
Are the models correct?



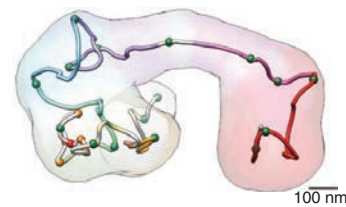
Jhunjhunwala (2008) Cell



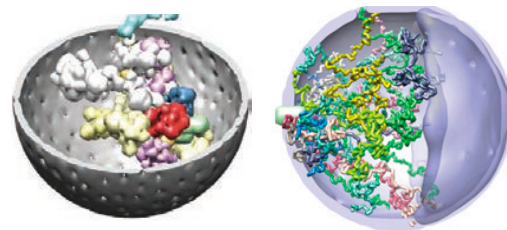
Duan (2010) Nature



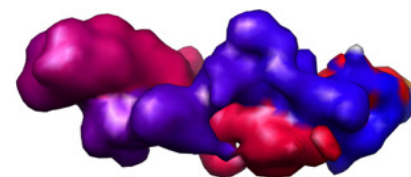
Fraser (2009) Genome Biology
Ferraiuolo (2010) Nucleic Acids Research



Baù (2011) Nature Structural & Molecular Biology



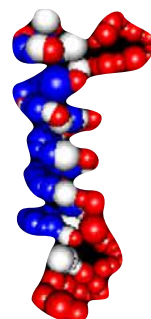
Kalhor (2011) Nature Biotechnology
Tjong (2012) Genome Research



Umbarger (2011) Molecular Cell



Junier (2012) Nucleic Acids Research



Hu (2013) PLoS Computational Biology

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,*}

¹EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), Barcelona, Spain, ²Universitat Pompeu Fabra (UPF), Barcelona, Spain, ³Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), Barcelona, Spain, ⁴Genome Biology Group, Centre Nacional d'Anàlisi Genòmica (CNAG), Barcelona, Spain and ⁵Institució 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 restraint-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

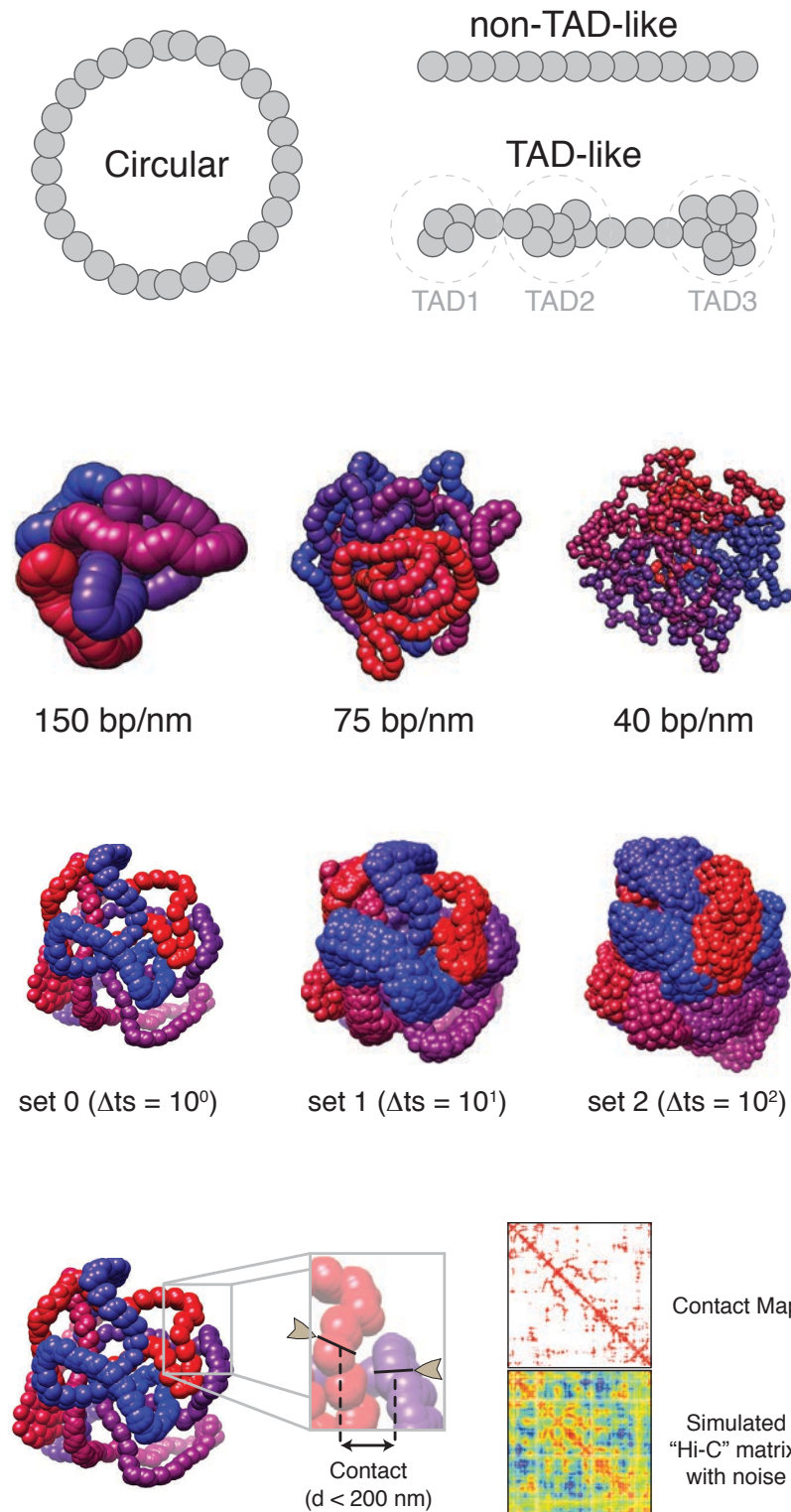
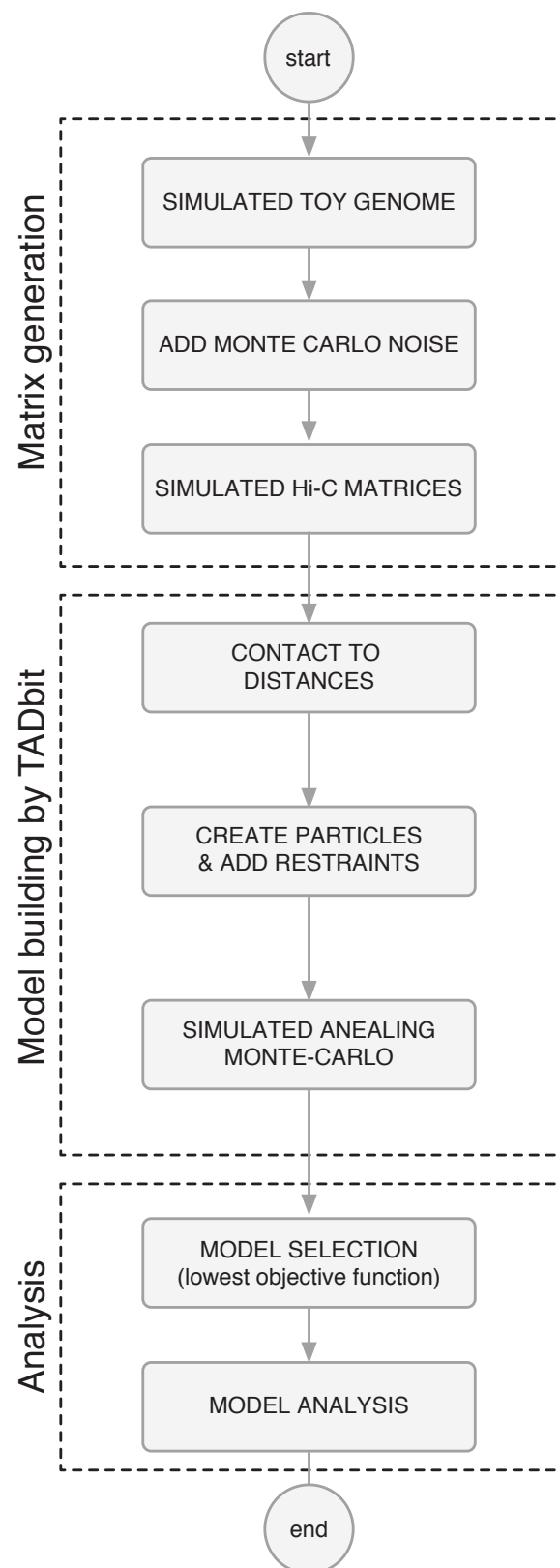
expression regulation and replication (1–6). The advent of the so-called Chromosome Conformation Capture (3C) assays (7), which allowed identifying chromatin-looping interactions between pairs of loci, helped deciphering some of the key elements organizing the genomes. High-throughput derivations of 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://integrativemodelling.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

© The Author(s) 2015. Published by Oxford University Press on behalf of Nucleic Acids Research. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

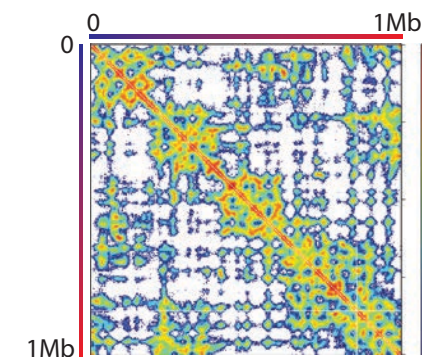
Toy models



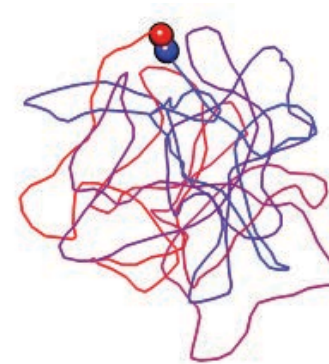
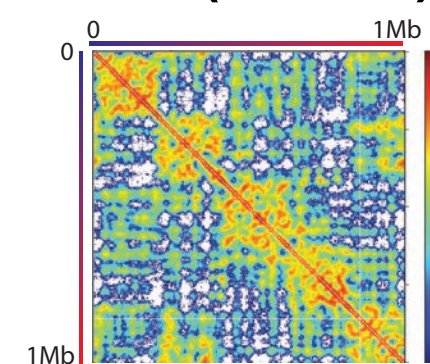
Toy interaction matrices



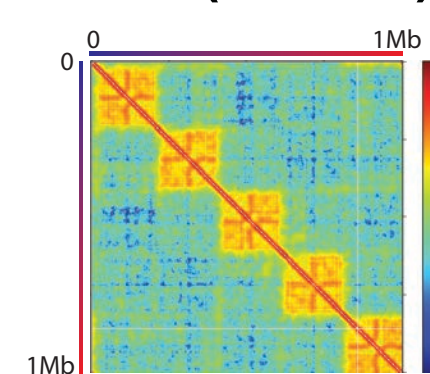
set 0 ($\Delta t_s = 10^0$)



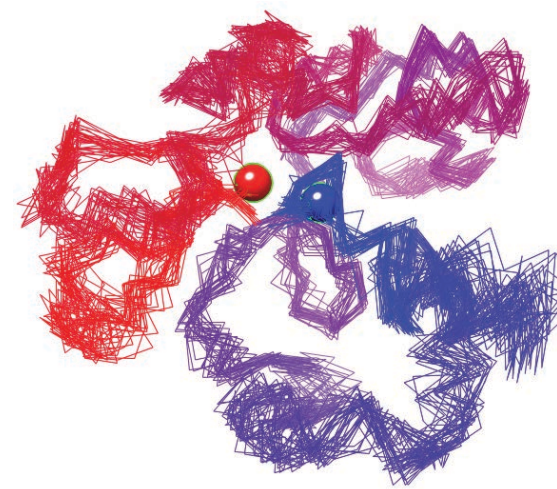
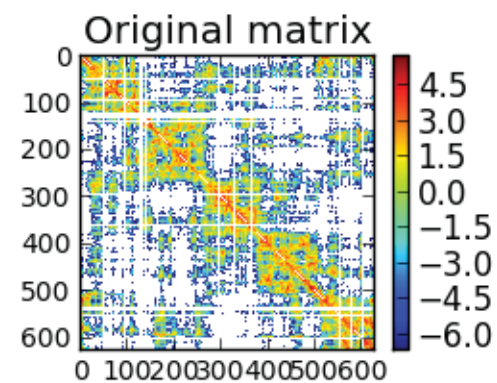
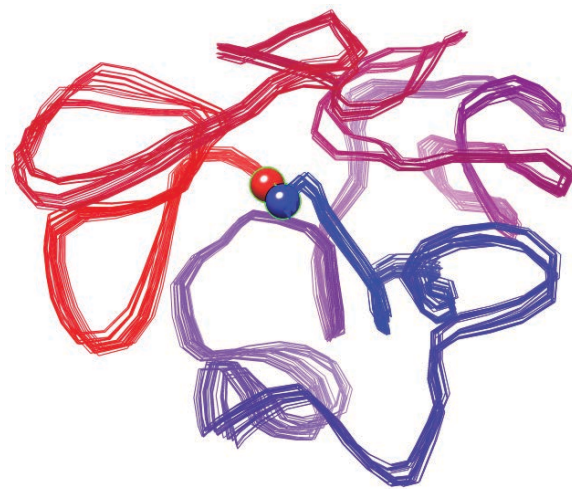
set 4 ($\Delta t_s = 10^4$)



set 6 ($\Delta t_s = 10^6$)



Reconstructing toy models



chr40_TAD

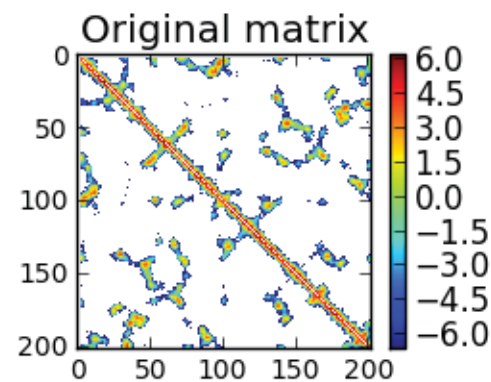
$\alpha=100$

$\Delta t_s=10$

TADbit-SCC: 0.91

<dRMSD>: 32.7 nm

<dSCC>: 0.94



chr150_TAD

$\alpha=50$

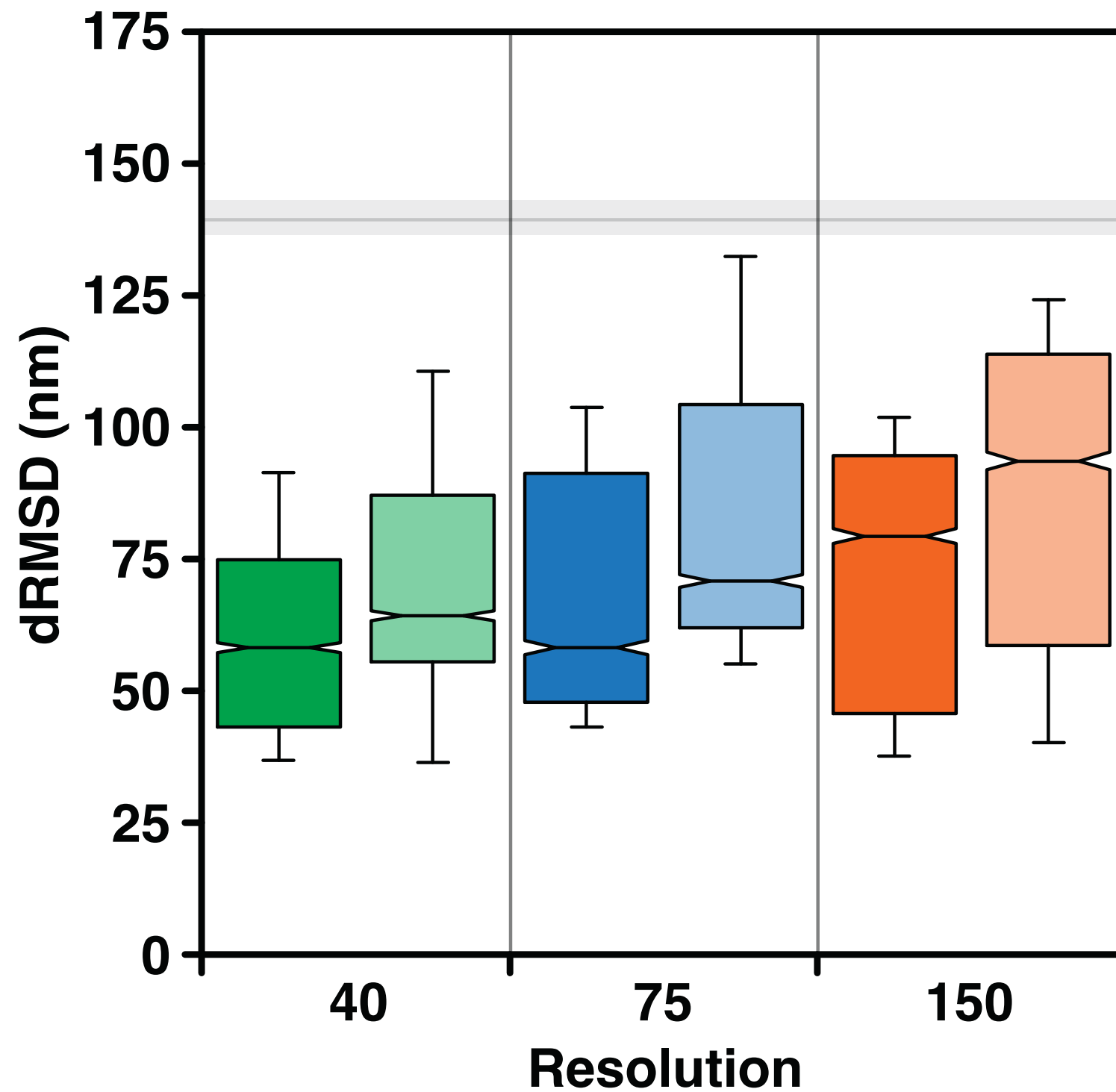
$\Delta t_s=1$

TADbit-SCC: 0.82

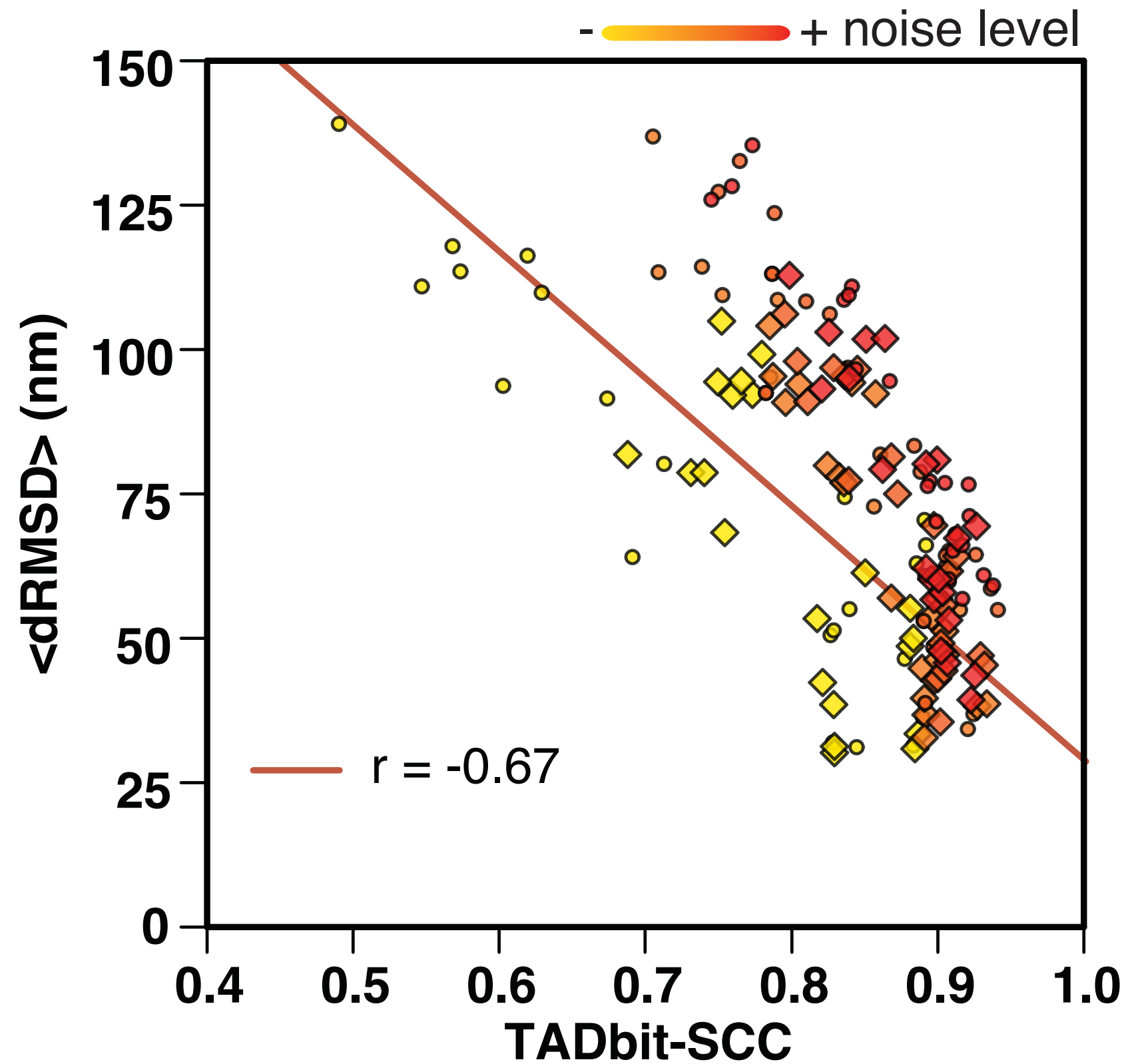
<dRMSD>: 45.4 nm

<dSCC>: 0.86

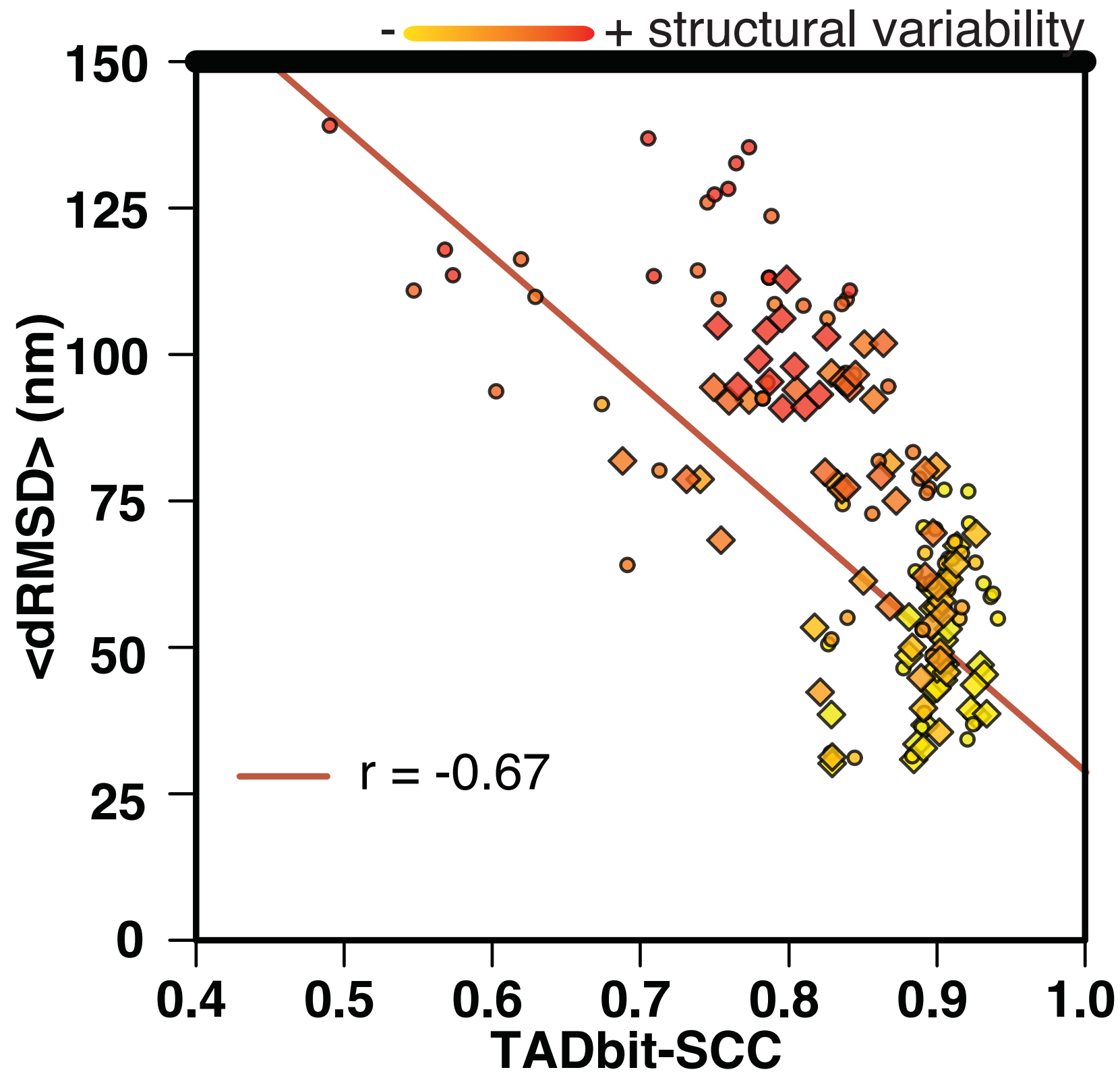
TADs & higher-res are "good"



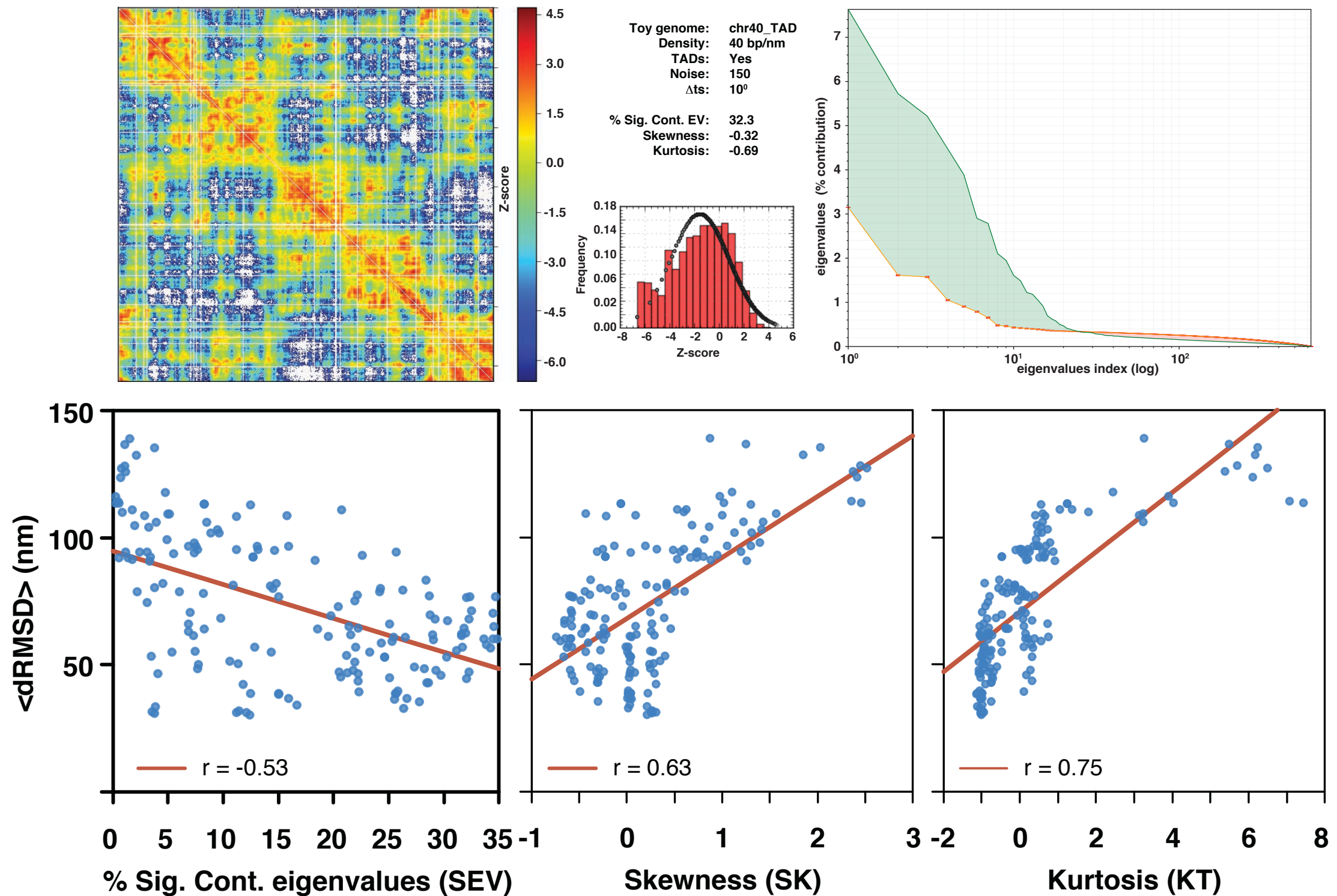
Noise is "OK"



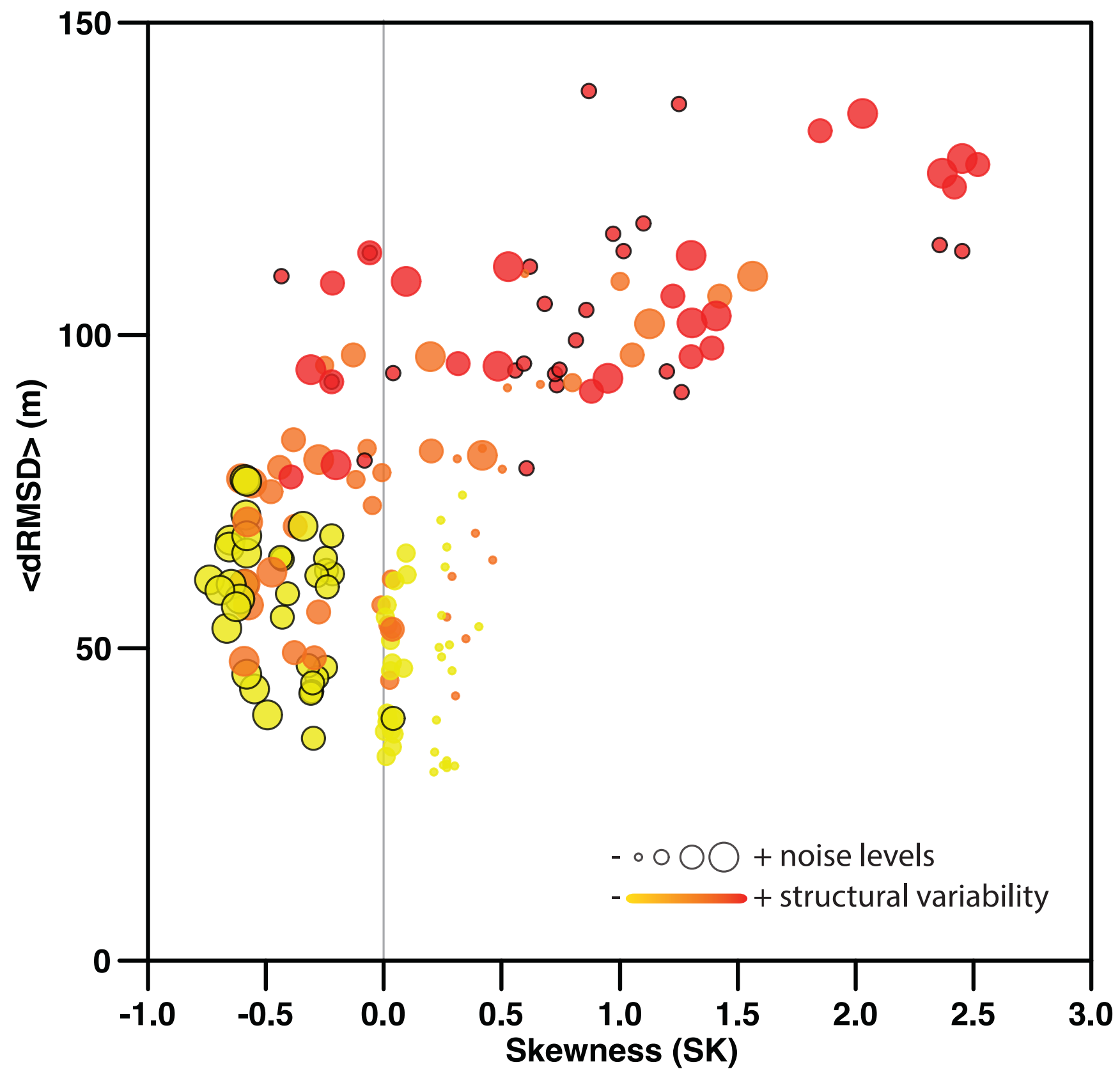
Structural variability is "NOT OK"



Can we predict the accuracy of the models?

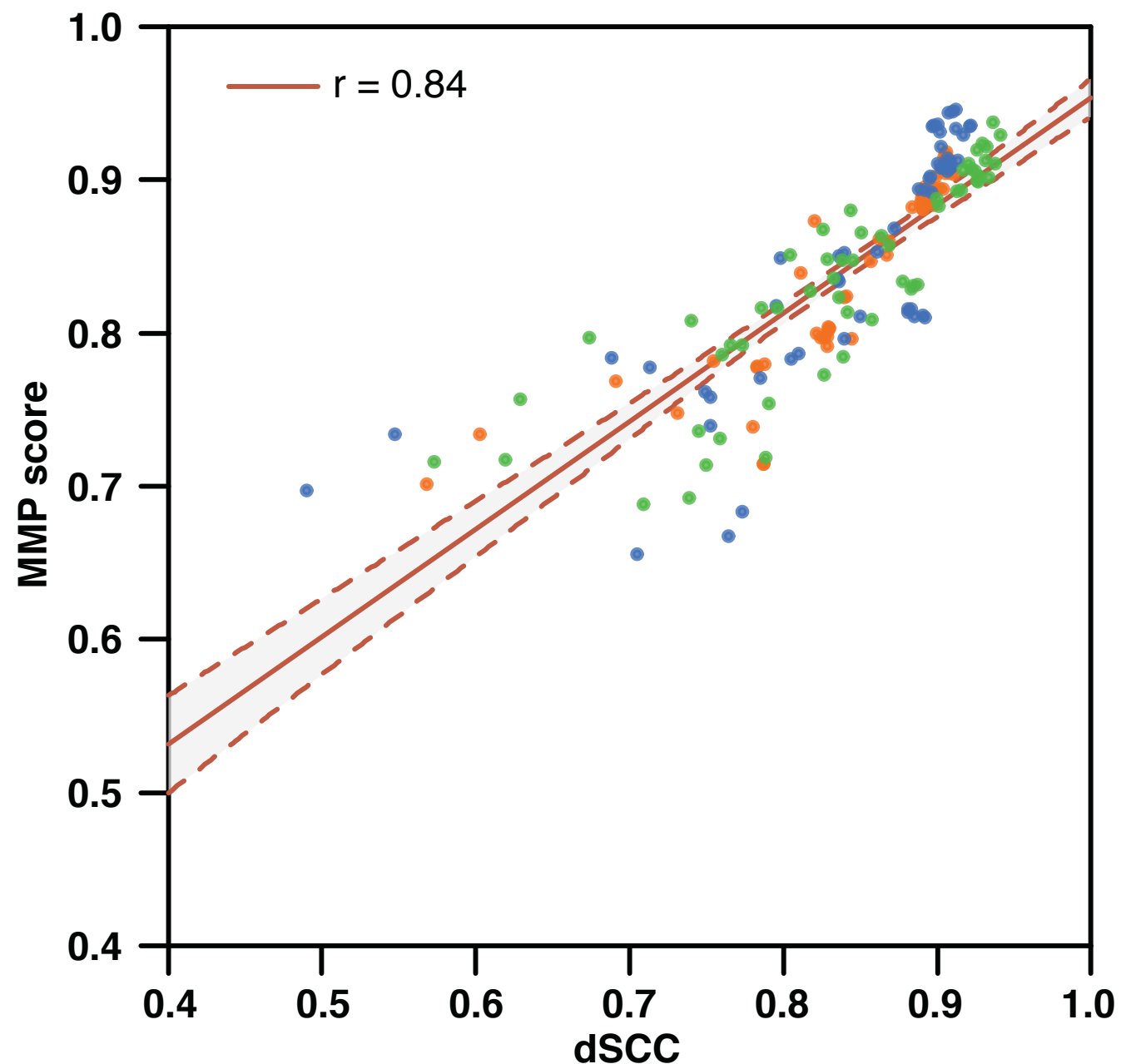


Skewness "side effect"

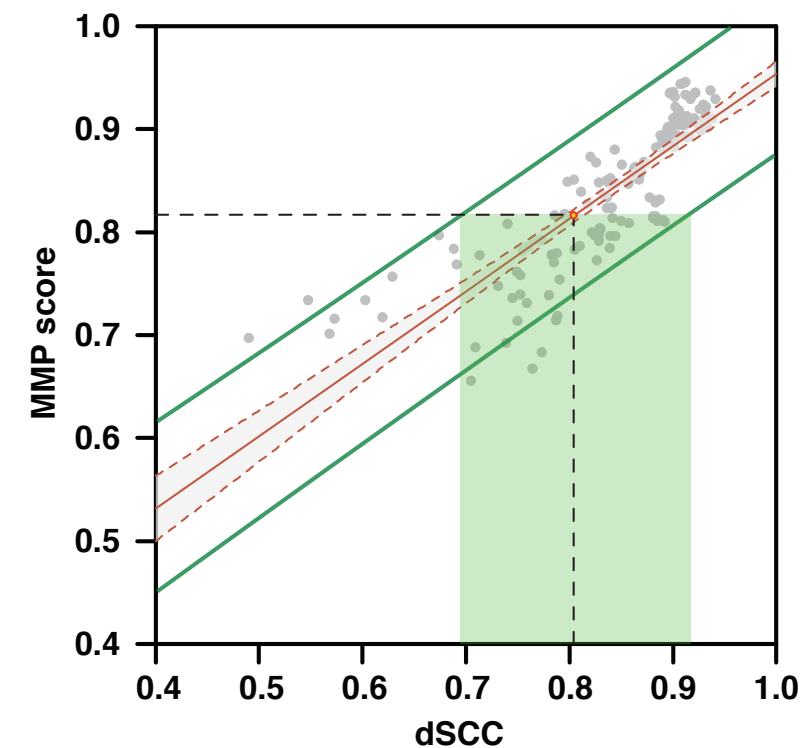
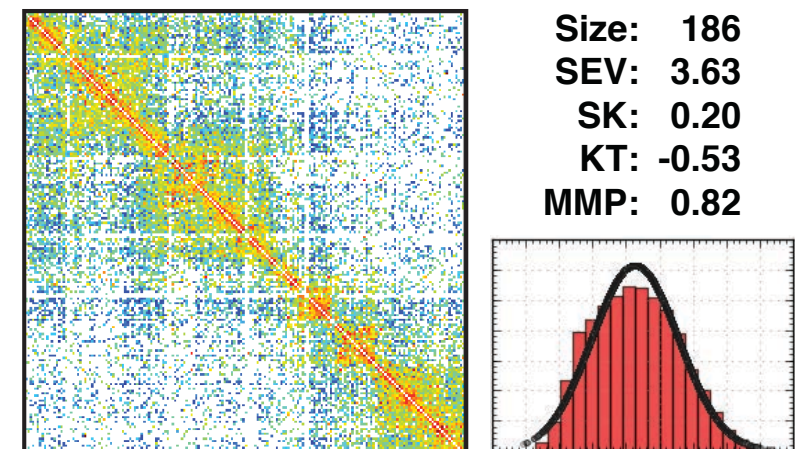


Can we predict the accuracy of the models?

$$\text{MMP} = -0.0002 * \text{Size} + 0.0335 * \text{SK} - 0.0229 * \text{KU} + 0.0069 * \text{SEV} + 0.8126$$



Human Chr1:120,640,000-128,040,000



Higher-res is "good"

put your \$\$ in sequencing

Noise is "OK"

no need to worry much

Structural variability is "NOT OK"

homogenize your cell population!

...but we can differentiate between noise and structural variability

and we can a priori predict the accuracy of the models



François le Dily
Marie Trussart
Davide Baù

Gireesh K. Bogu
Yasmina Cuartero
David Dufour
Irene Farabella
Silvia Galan
Mike Goodstadt
Francisco Martínez-Jiménez
François Serra
Paula Soler
Yannick Spill
Marco di Stefano

in collaboration with Ivan Junier (Université Joseph Fourier), Miguel Beato (CRG) & Luís Serrano (CRG)

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

