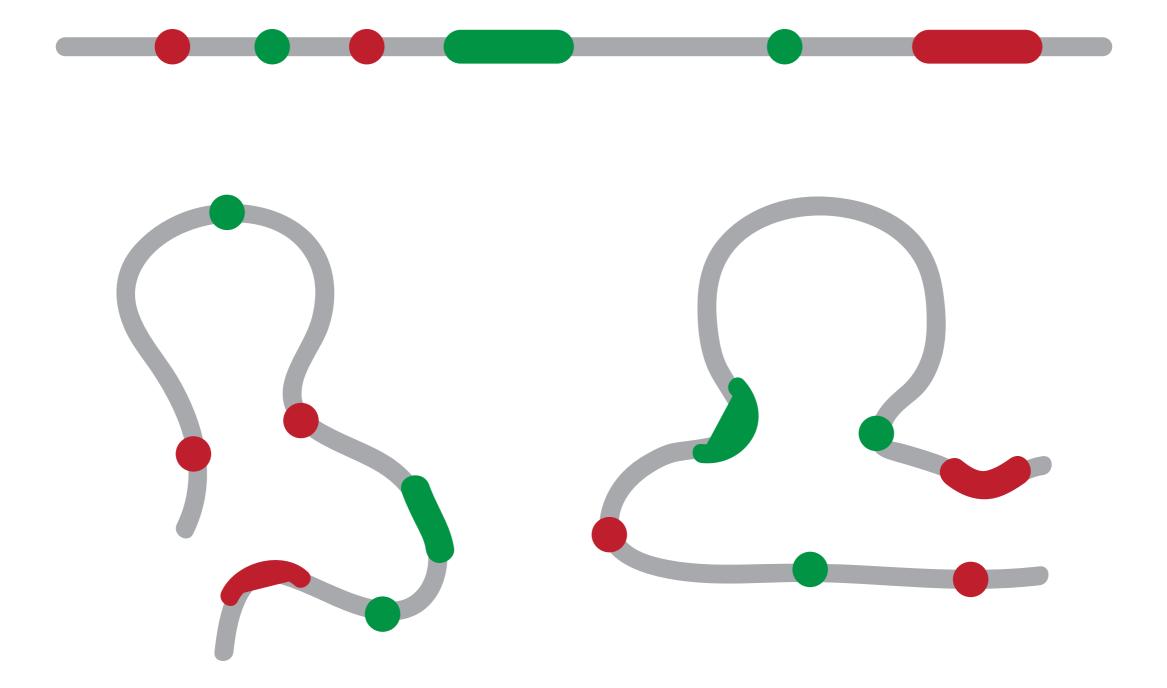
Structure determination of genomes and genomic domains by satisfaction of spatial restraints

Marc A. Marti-Renom

Structural Genomics Group (ICREA, CNAG-CRG)

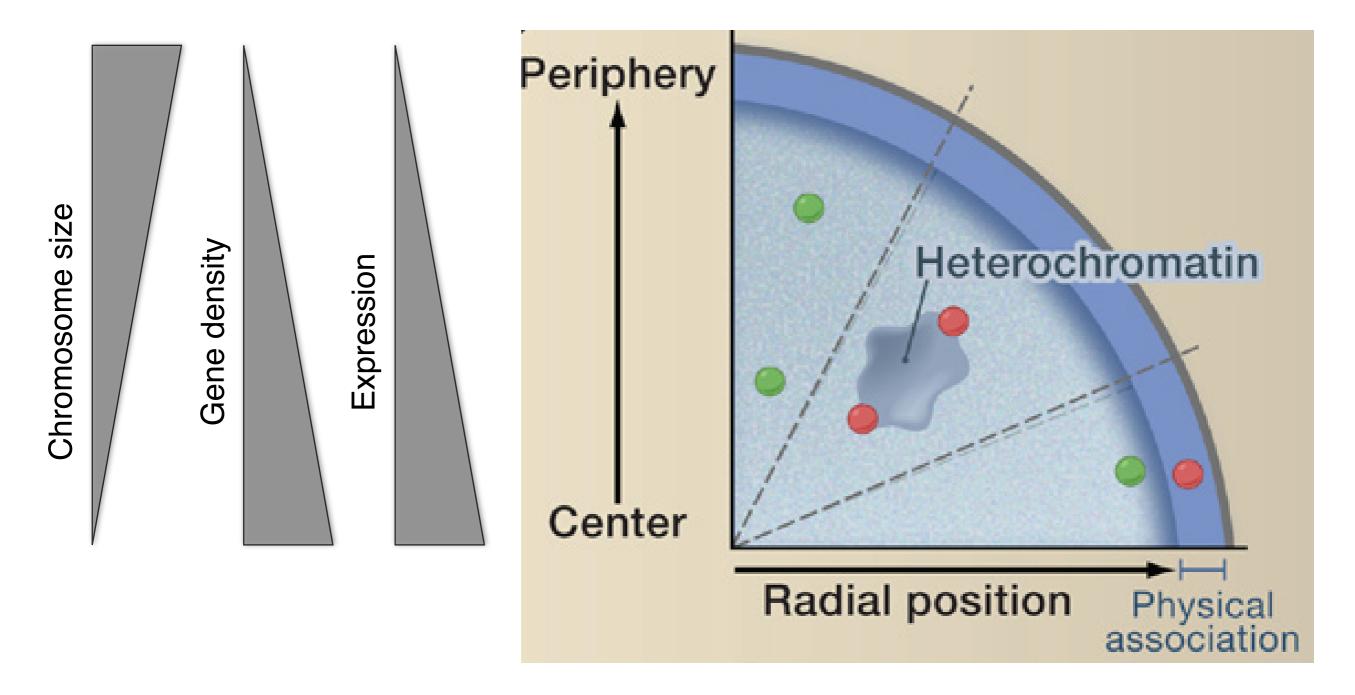
http://marciuslab.org
http://3DGenomes.org
http://cnag.crg.eu





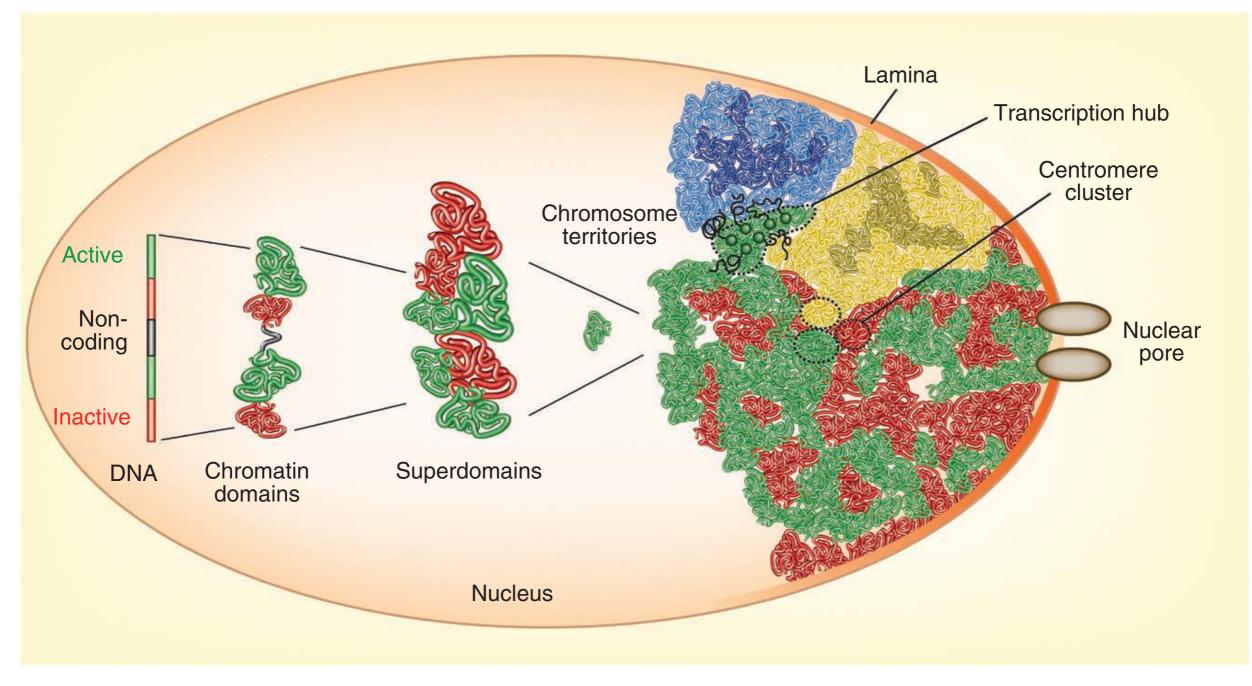
Complex genome organization

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

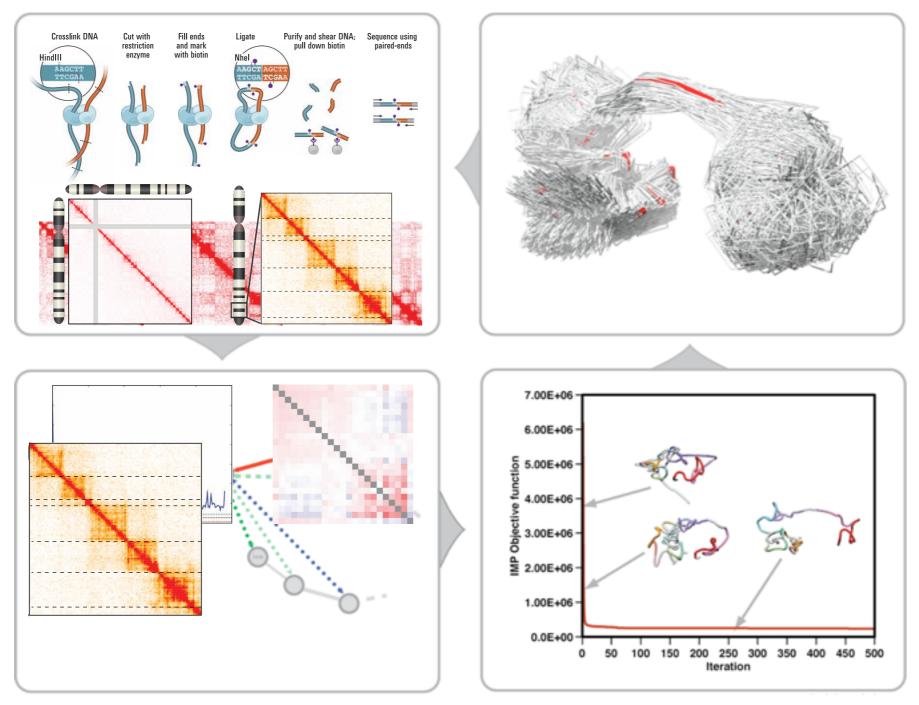
Marti-Renom, M. A. & Mirny, L. A. PLoS Comput Biol 7, e1002125 (2011)

Know	edge								
1 A A A A					IDM			$\begin{array}{c} & 11 \\ & X \\ & 12 \\ & 15 \\ & 6 \\ & 1 \\ & 8 \\ & 1 \\ & 1 \\ & 1 \\ & 20 \\ & 18 \\ & 7 \\ & 2 \\ & 16 \\ & 9 \\ & 18 \end{array}$	7
								DNA length]
10 ⁰		10 ³			10 ⁶			10 ⁹	nt
								Volume	1
10 ⁻⁹		10 ⁻⁶	10 ⁻¹	3		10 ⁰		10 ³	μm ³
10		10	10			10		10	μπ
								Time	1
10 ⁻¹⁰	10 ⁻⁸	10 ⁻⁶	10 ⁻⁴	10 ⁻²		10 ⁰	10 ²	10 ³	S
								Resolution	
10 ⁻³			10 ⁻²				10 ⁻¹		μ

Hybrid Method

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

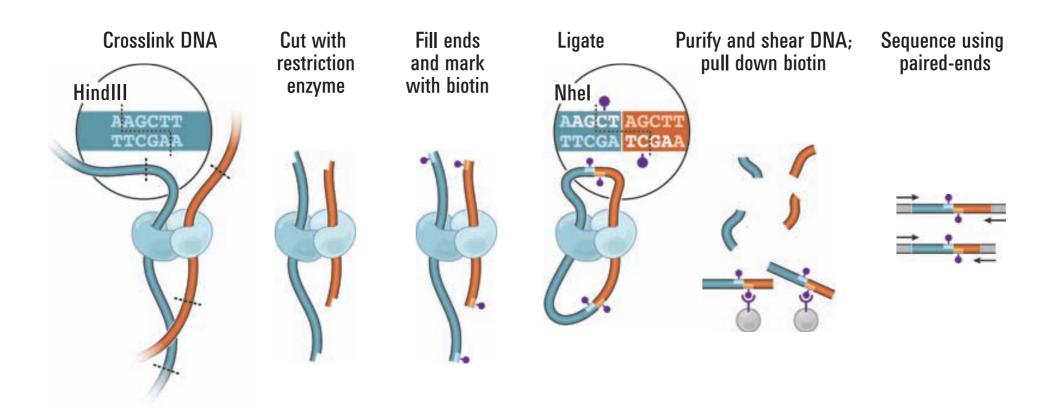
Experiments

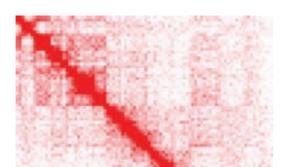


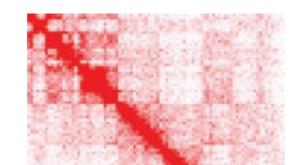
Computation

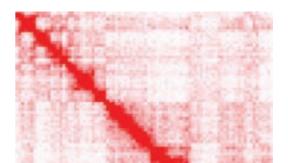
Chromosome Conformation Capture

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.







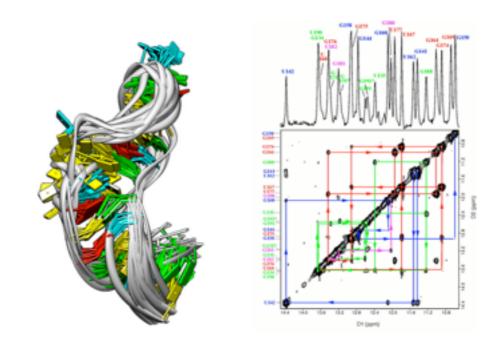


Chromosome Conformation Capture

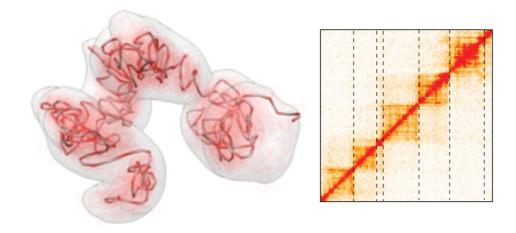
CROSSLINK							
CUTTING	Sonication						
LIGATION		Biotin dCTP fill in	Immunoprecipitation	Immunoprecipitation biotinilated linkers			
REVERSE CROSSLINKS		B B	·	<u> </u>			
DETECTION PCR with specific primers	Multiplexed amplification	Digestion with four base cutter Ligation	Sonicate Pull down	PCR with specific primers	Mmel digestion		
Contact library	-> - ↓₹- ->- ↓ ₹-	Inverse PCR	B B B	<u> </u>	B		
COMPUTATIONAL ANALYSIS							
3C	5C	4C	Hi-C	ChIP-loop	ChIA-PET		

Restraint-based Modeling

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



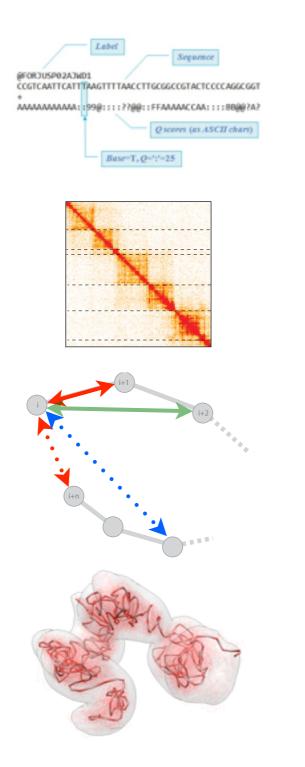
Biomolecular structure determination 2D-NOESY data



Chromosome structure determination 3C-based data



http://3DGenomes.org



FastQ files to Maps

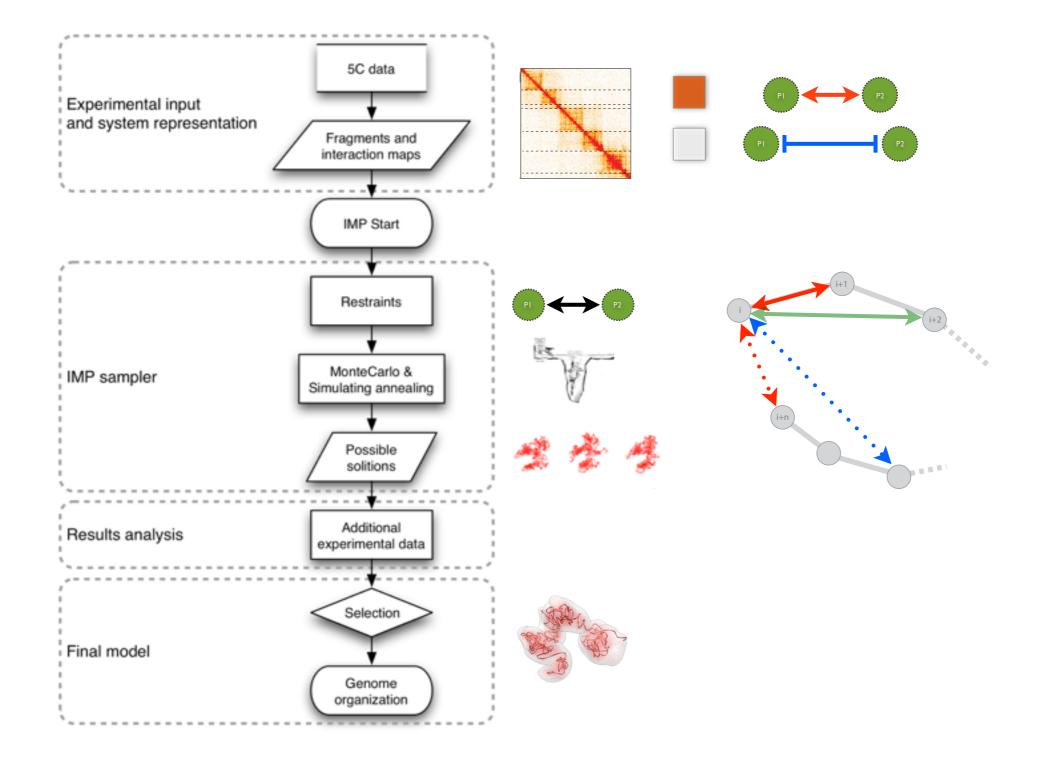
Map analysis

Model building

Model analysis

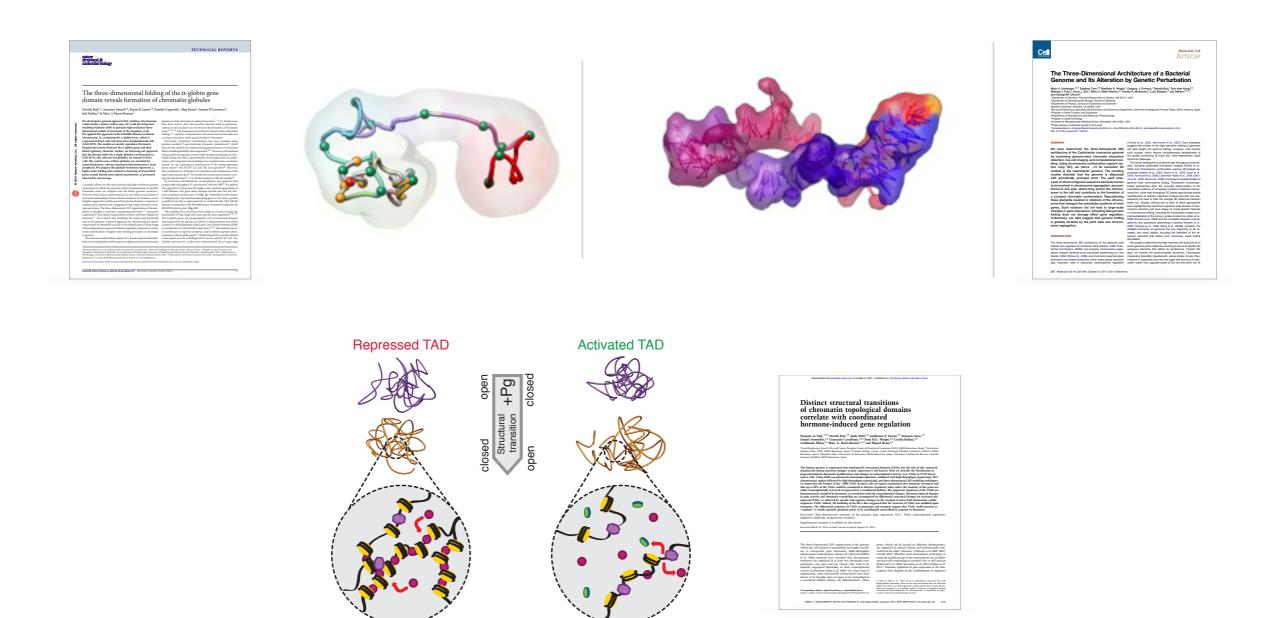


http://3DGenomes.org





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)

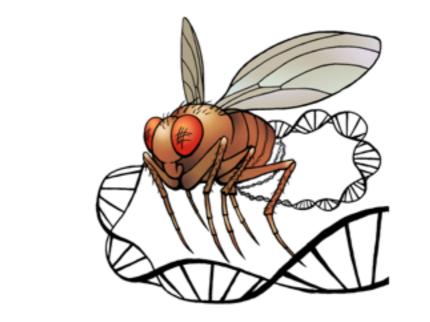


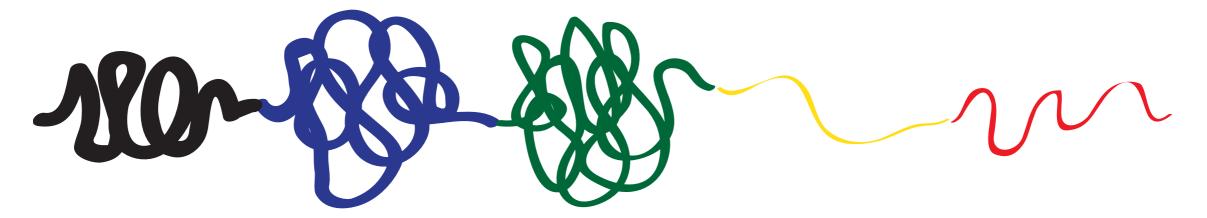
PRBS

JNucleosome

Histones H2A/H2B Histone H1

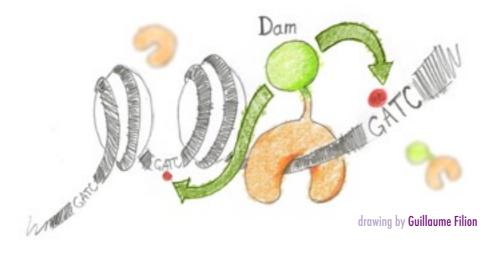
Structuring the **COLORs** of chromatin

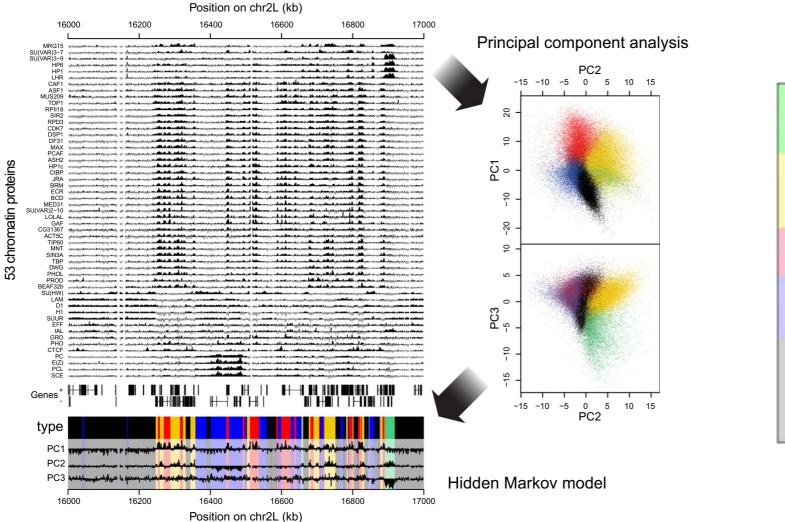


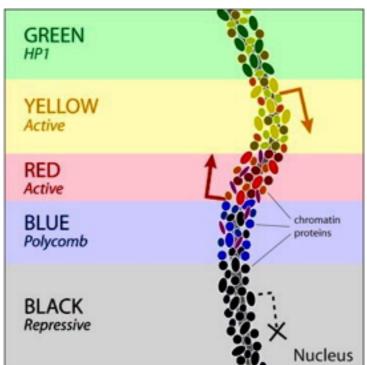


Fly Chromatin **COLORs**

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

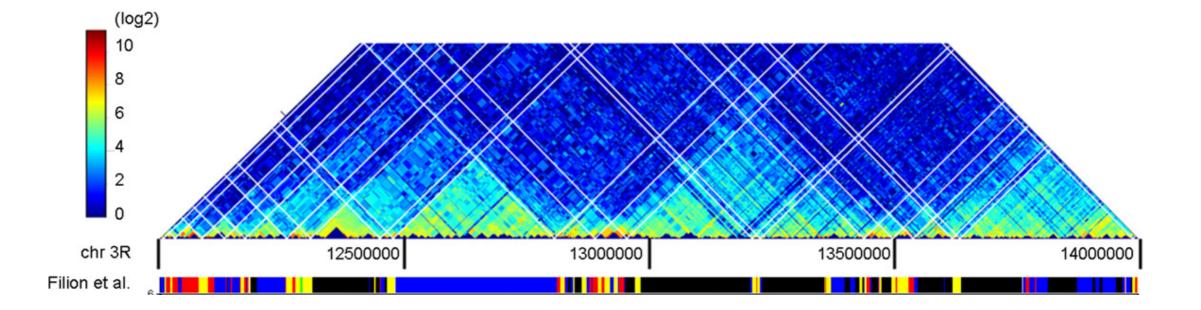


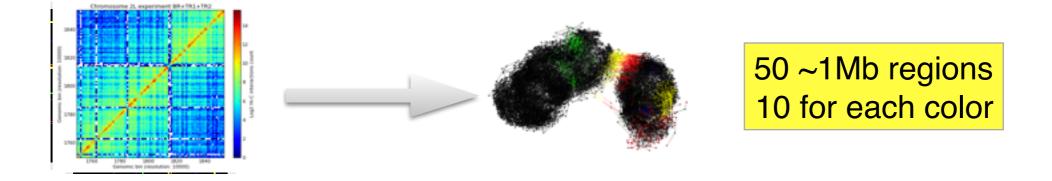




Fly Chromatin **COLORs**

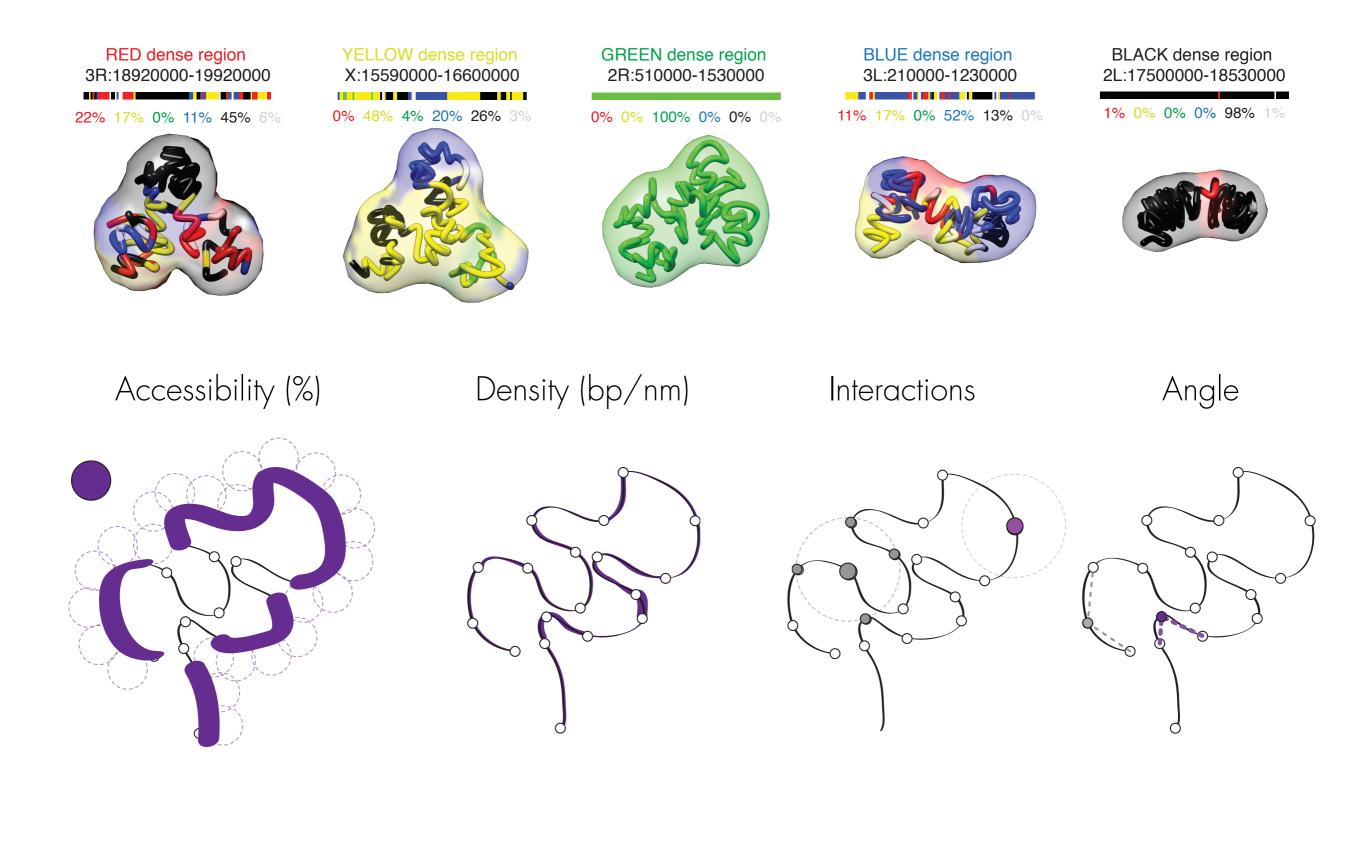
Hou et al. (2012). Molecular Cell, 48(3), 471–484.



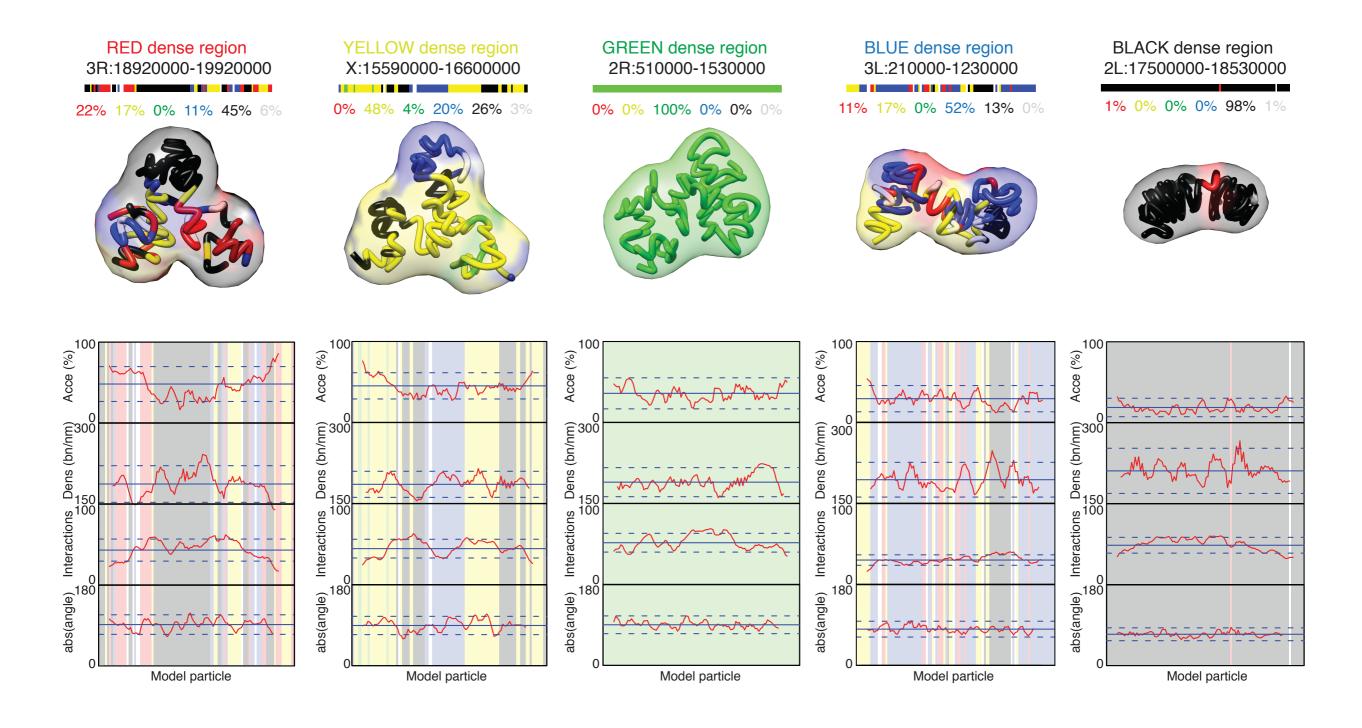


Structural properties

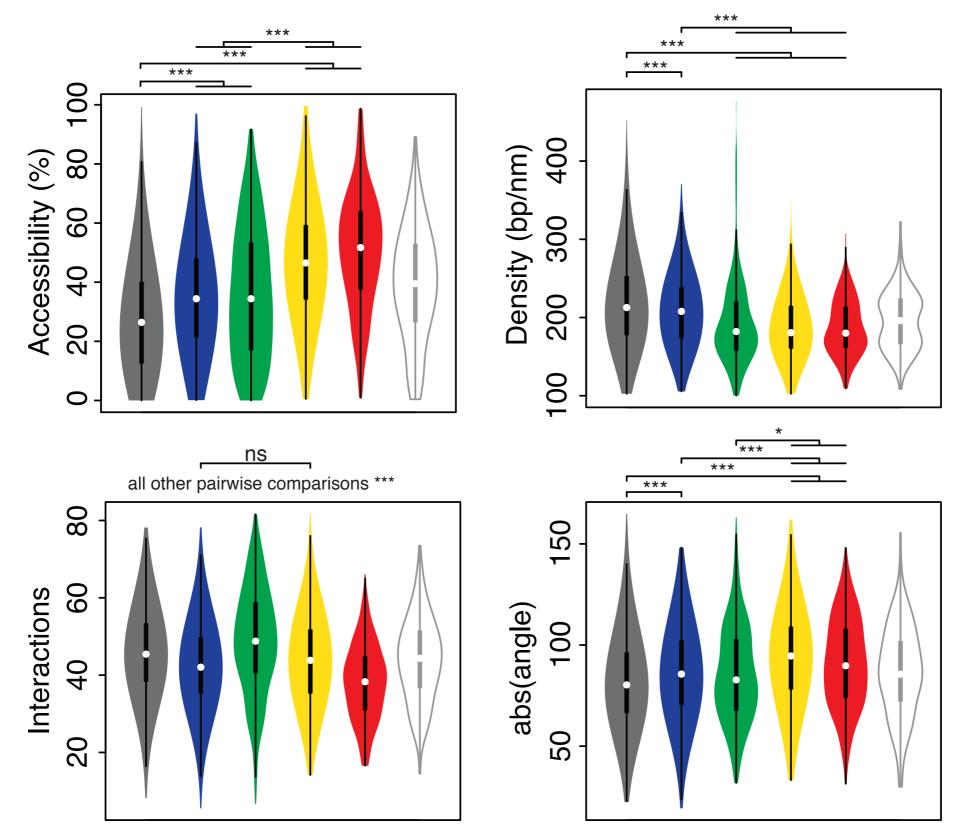
50 1Mb regions. 10 enriched for each color.



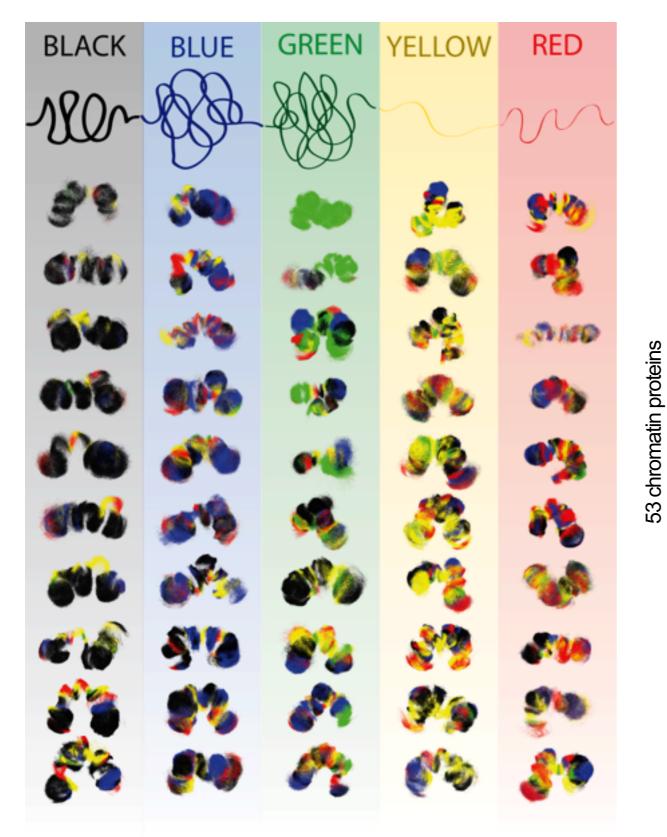
Structural **COLORs**



Structural **COLORs**

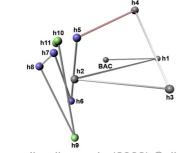


Structural **COLORs**

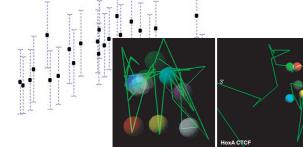


	Position on chr2L (kb)							
16	000	16200	16400	16600	16800	1700		
MRG15		- shitten and the state		An ala and a second sec	And the second state of th			
SU(VAR)3-7	warmen a sound and sound a	n denner maneral de sertes	and the second second second second second		And and the state of the second state	And the state of the		
SU(VAR)3-9		- foresterranterrates		مر بعد المربي بالمعن مريا مند مستري	and the second			
HP6 HP1	Antoin an	- Paulanenanen artan		an dilampan and a balance and a second		and a second sec		
LHR	Charles and a state of the stat	Lauran	able all a second a second a	A second se	and an other state of the state	And a state of the state of the		
CAF1					and the second state and the second second second			
ASF1	man annother and a	a station and and the Mail	A transmission of the second second details		as a state of the state of the second state of	and the second second		
MUS209	gréphisterenterenter e	to plante the second difference	and an iter an about a short of the state of	an all an	Adams Although the second second second	Party Street and Street		
TOP1 RPII18	A contraction of the second of the	in the second seco		al Manager and States and a state		And the state show that		
SIR2	and the second s	-	Addition and a second second			and a subsection of the subsec		
RPD3	minute address of the second s	-	and a second sec		and the second sec	A distant in lattice of		
CDK7	and the second s	or processing and an and	and the second second second	and hits a second state of a second state of the second state of t		and the star of the starting		
DSP1 DF31	-40-40 - 10 - 10 - 10 - 10 - 10 - 10 - 1	to share expression at the second	and the set of a spin of the second state and a	a hite a state of the state of	a deside part and a second state or a first second state or any	Webserge - States		
MAX	Apples	a photographic and a state of the	and as a second second	distance of some desce ba		paring and and		
PCAF	and a subserve of the subserve	- statestand and the back	A 410 10 10 10 10 10 10 10 10 10 10 10 10 1			and a state of the		
ASH2	the subscription of the second		and the second s	of hits and a stand of the stan		Name of the state		
HP1c	entrefficients/strengts/strengts/	a Abstraction and the state	and the second of the second second second	of Management of the second se	an and the surger and the same subscription			
CtBP JRA	months and desired and the second	a Aparellaranting		A Distance of the state of the		supplication contractor		
BRM	distant on a second share a	a selected and a selected at the selected at t	and a second	I Man all the day was		And the second second second		
ECR		a plane and a state	And a second and a second descent	A have made and and and	- Anno and and a star and a star and a star			
BCD	didatificantes qualitationes a		all the second and a second the second	I have a state of the second s	and the second state of the second			
MED31	stanting an an in the second stands of the	- pagesperspective and	and the second s	Jahren Jahren ander	and and the second state of the second state of the	Understand agreeding		
SU(VAR)2-10 LOLAL	and the second s	a susception which and the	and a second s	I have a state of the second		Adaption and a strength		
GAF	shine and a second seco	- Annon Markel	And and and and the same	I the		Contraction of the last		
CG31367	Anaderson and a second design of a		and an an an and the second		and an an an and the set of the set of the set	Angeling the part		
ACT5C	helicited and a second	-	and a market and a start	al have represented to a second day		pre-ares areas		
TIP60	Antoine and a start of		All and a state of the state of	d be a data a calle	an all shares and a shift an and a shift a	Property and the second		
MNT SIN3A	alantalistic and a second s	in protocological and a	Adula	and heles are an end of the second strength of the		Westington and superscriptor		
TBP	description of the second of t	-	4844 - 10	d been and the second	a alles annale all a such as	And a state of the second		
DWG			and the second s		and a second s	And payment and a part		
PHOL	Wardshares and share a		and a second sec	at before a second as she a bear and the	and the destanded and and a second statements	And a strange and a stranger		
PROD	copperation and a special and a		And a state of the	a shake a second second second	the party of the state of the state of the state of the	Name of the second		
BEAF32b SU(HW)	and a state of the second s		and different in the second			and a subscription of the		
LAM		-		and and again a second	a second and and the second second			
D1		-	al a se a des a desta de la companya de la desta de la c					
H1			and the second second second by	of Statement of the statement of the				
SUUR EFF	the second se	a designed and the second	and a state of the	and best best bits bits and an an		And and the surface is		
IAL	and the second states of the s	a the seal or show of the	and a state has been a state	a she had a set the set of the set of the set		Martin And Sharehow		
GRO	alana a state and the second section of	a mare budget and the	and an and the same damas	I be and the second and the second	and the stand of the stand of the stand			
PHO	Shorthan and for the o	a share an an and and	and say a survey of these associations		and the second stand and the second states of the	Contraction and and and		
CTCF		a descent and a sublim		A	an an a star			
PC E(Z)	and a second sec		A Strategy of the strategy of the	a second and a second a second a second	and the party of t	Constantion to be stated on		
PCL	distant description of the second sec	to an advantage of the second s	A DE LE STATE OF THE OWNER	d be-man alter the man	da	Contraction of the same		
SCE			and the second s					

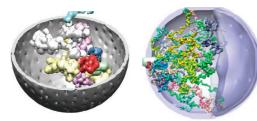
Are the models correct?



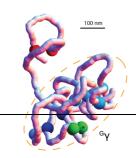




Fraser (2009) Genome Biology Ferraiuolo (2010) Nucleic Acids Research



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



Umbarger (2011) Molecular Cell

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 meanfield restrain-based method, which could be taken into consideration in further development of methods as well as their applications.

INTRODUCTION

the so-called Chromosome Conformation Capture (3C) as-says (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 sig-nificant 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 biolog-ical 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).

expression regulation and replication (1-6). The advent of

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

Recent studies of the three-dimensional (3D) conforma-

tion of genomes are revealing insights into the organization and the regulation of biological processes, such as gene

© 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.

Trussart, et al. (2015), Nucleic Acids Research.

Junier (2012) Nucleic Acids Research

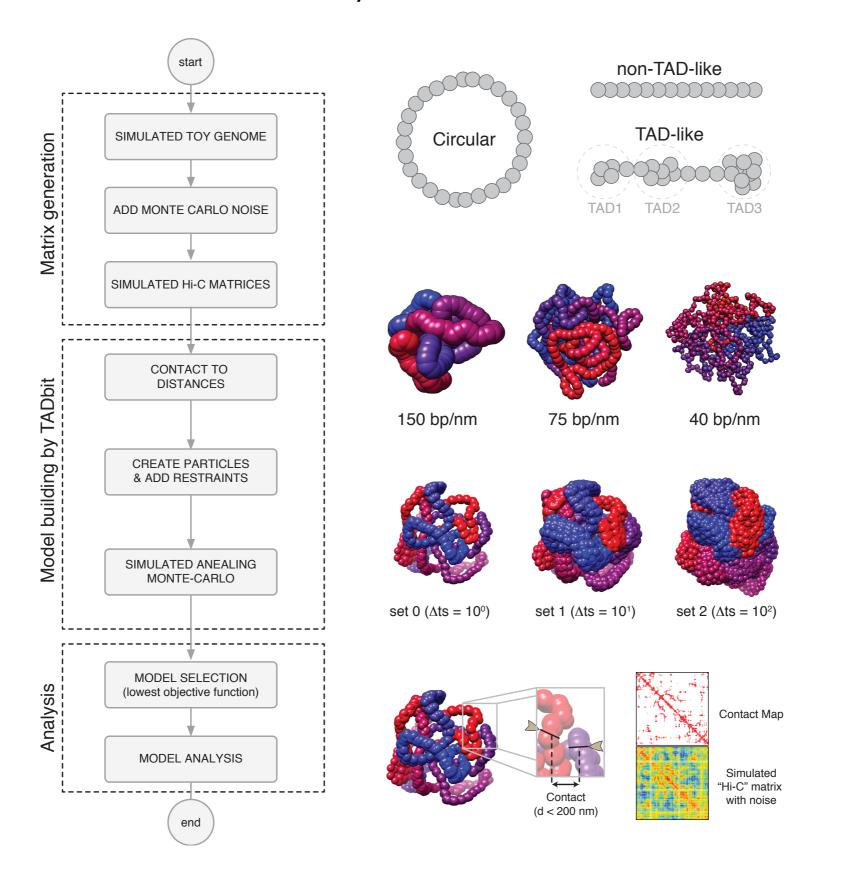
Hu (2013) PLoS Computational Biology

100 nm

Duan (2010) Nature

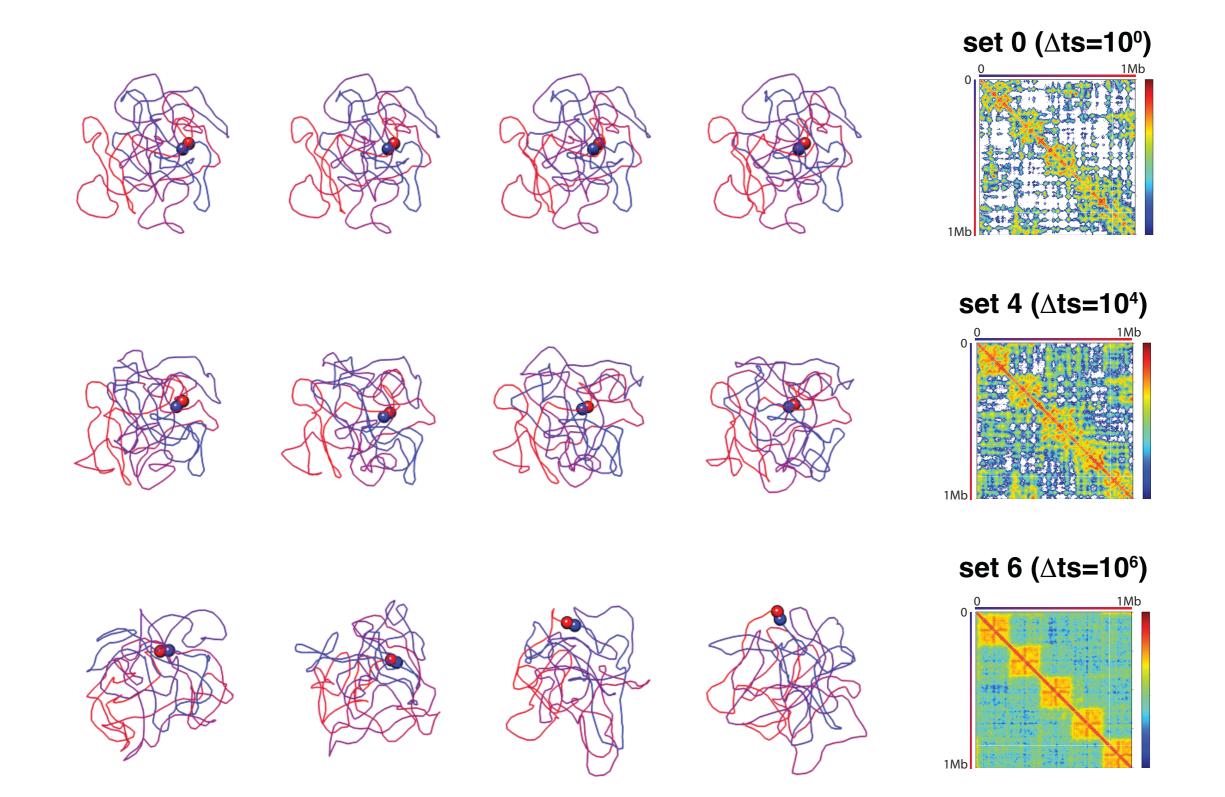
Baù (2011) Nature Structural & Molecular Biology

Toy models

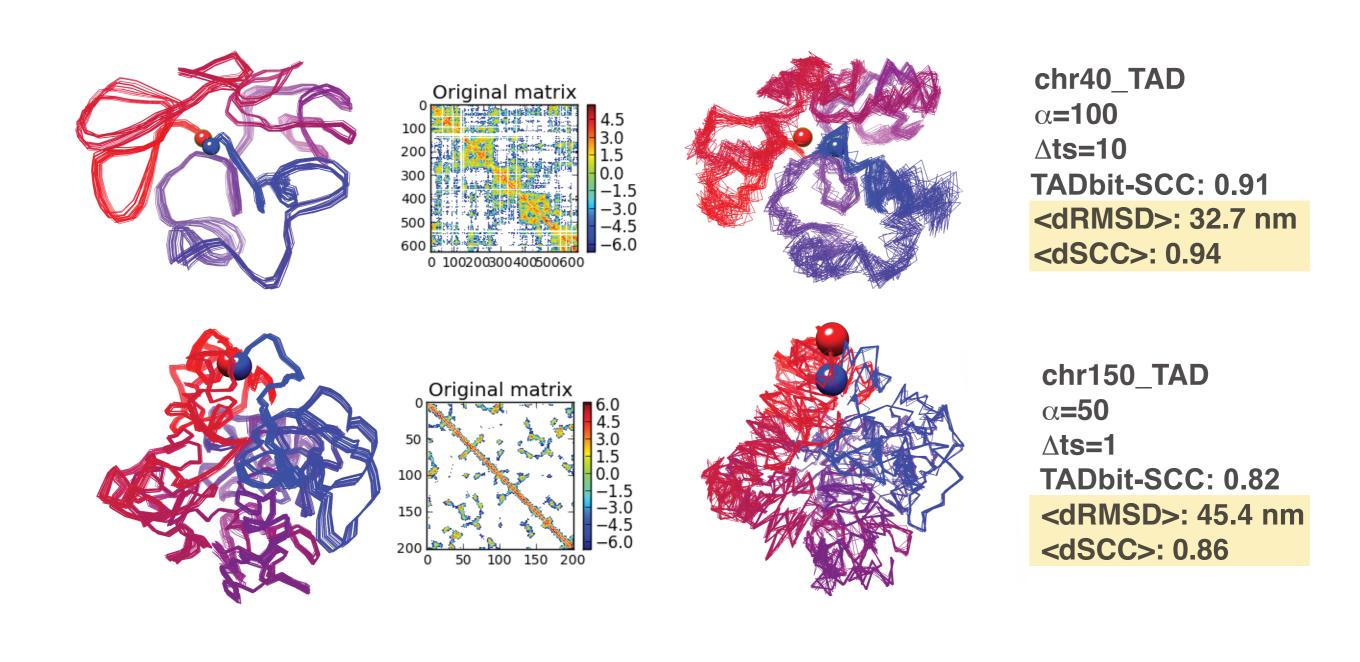


by Ivan Junier

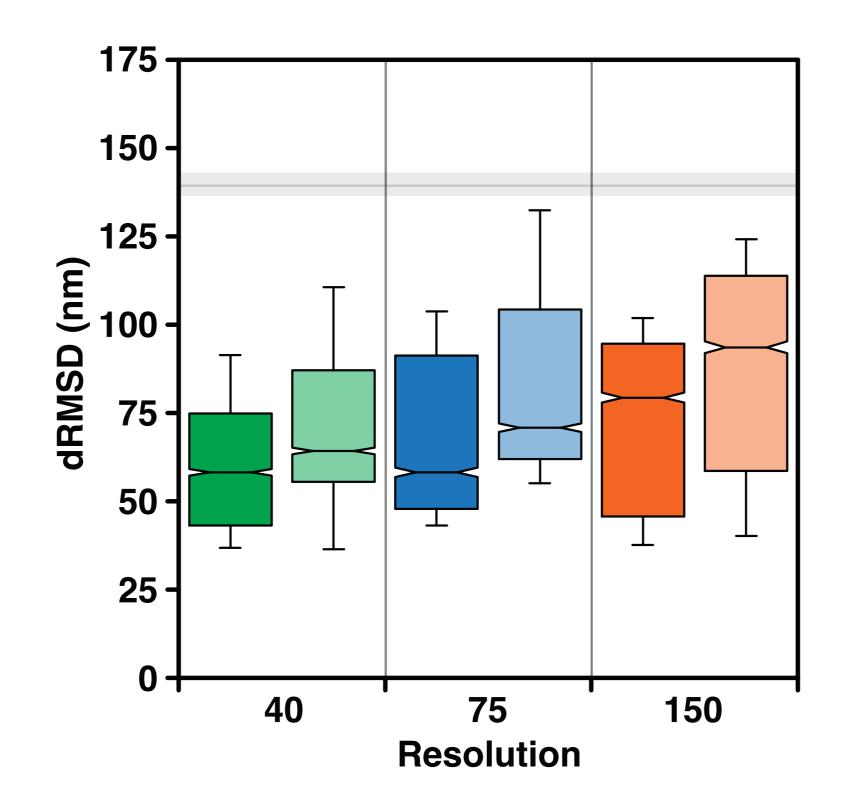
Toy interaction matrices



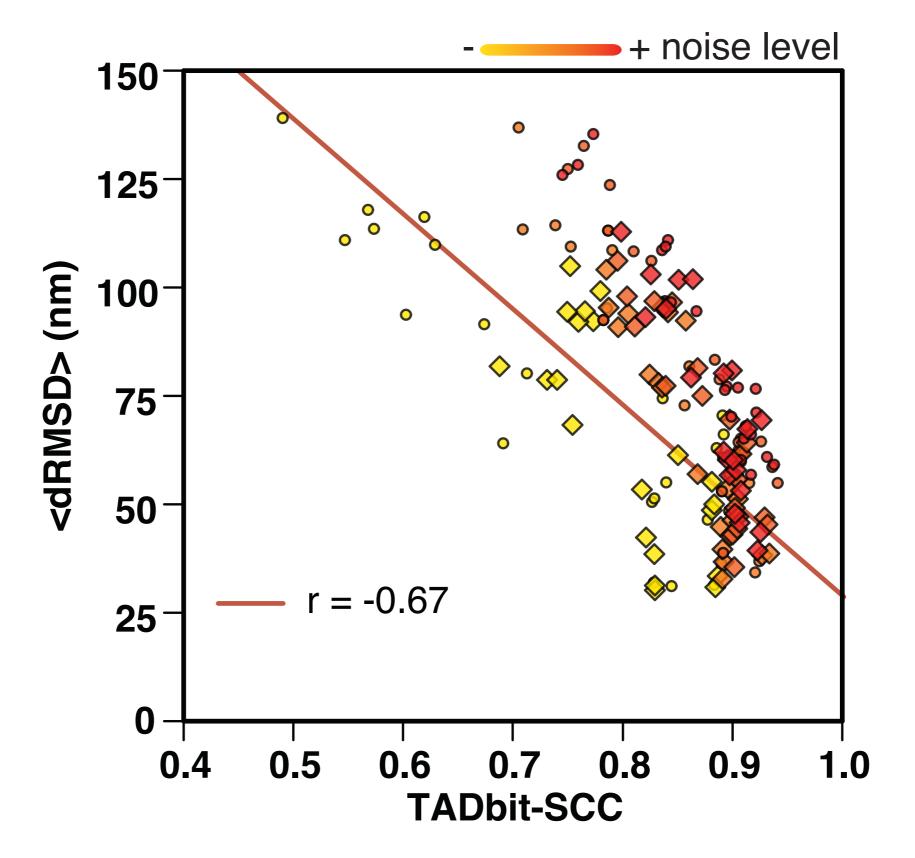
Reconstructing toy models



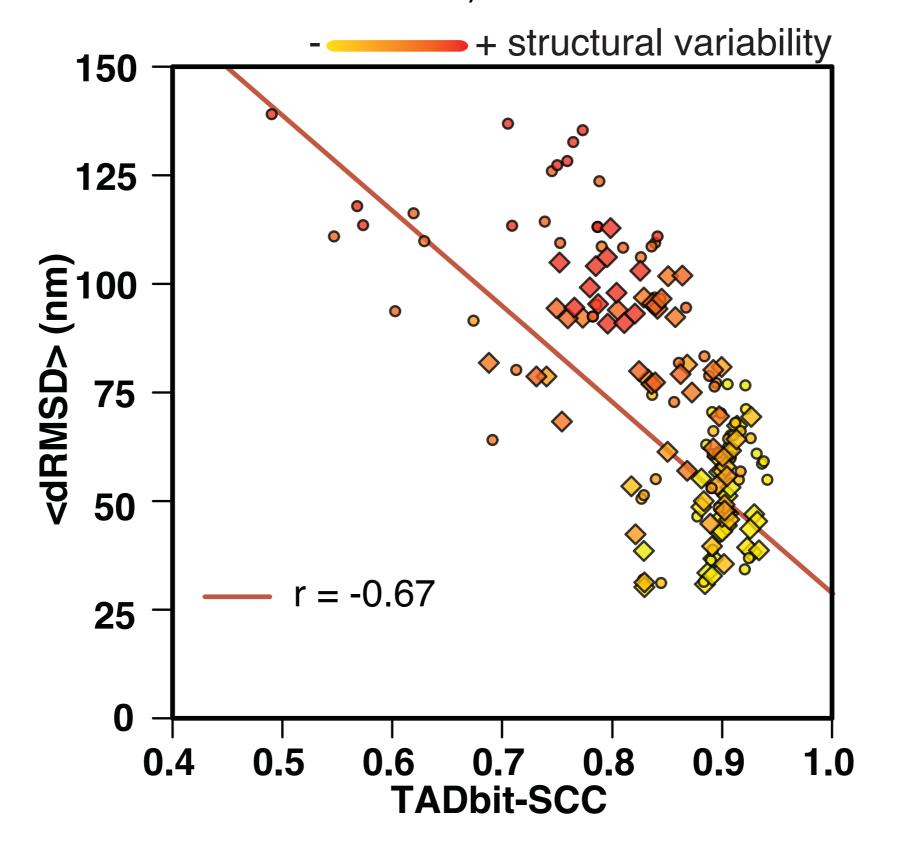
TADs & higher-res are "good"



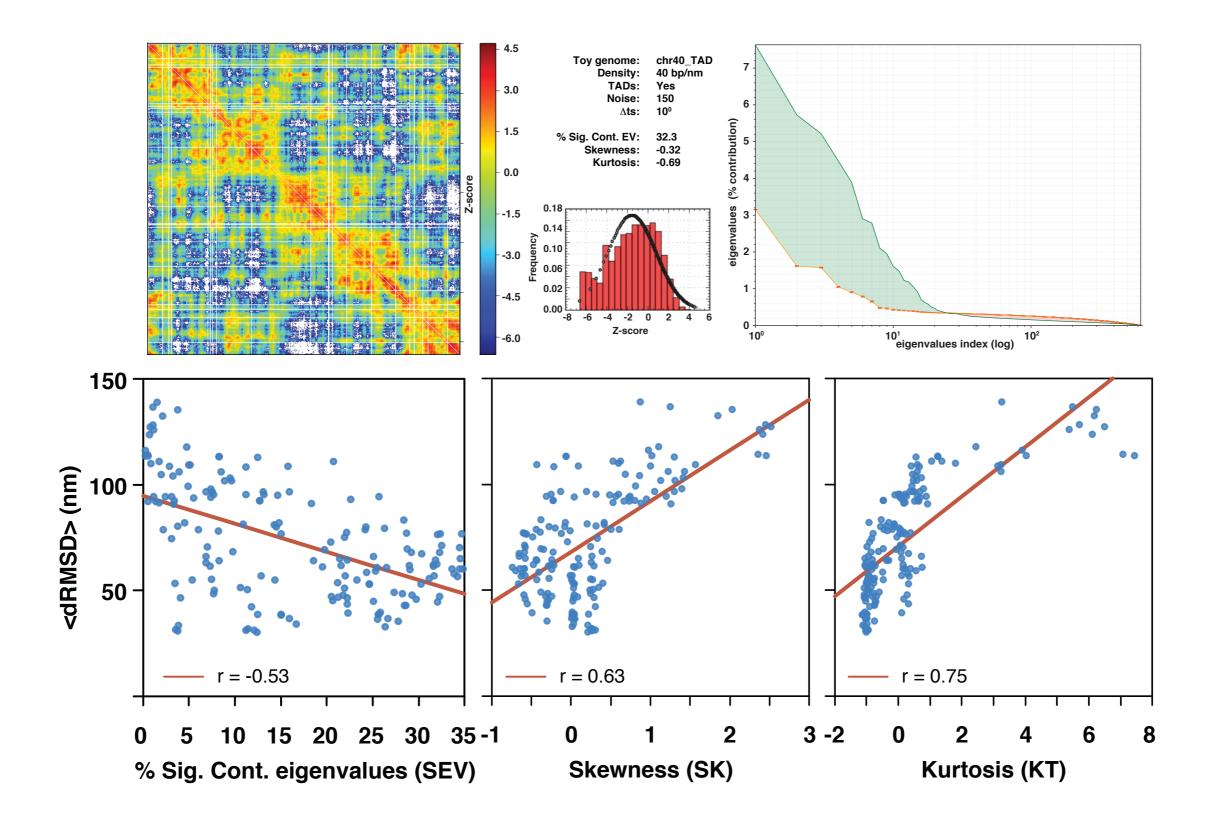
Noise is "OK"



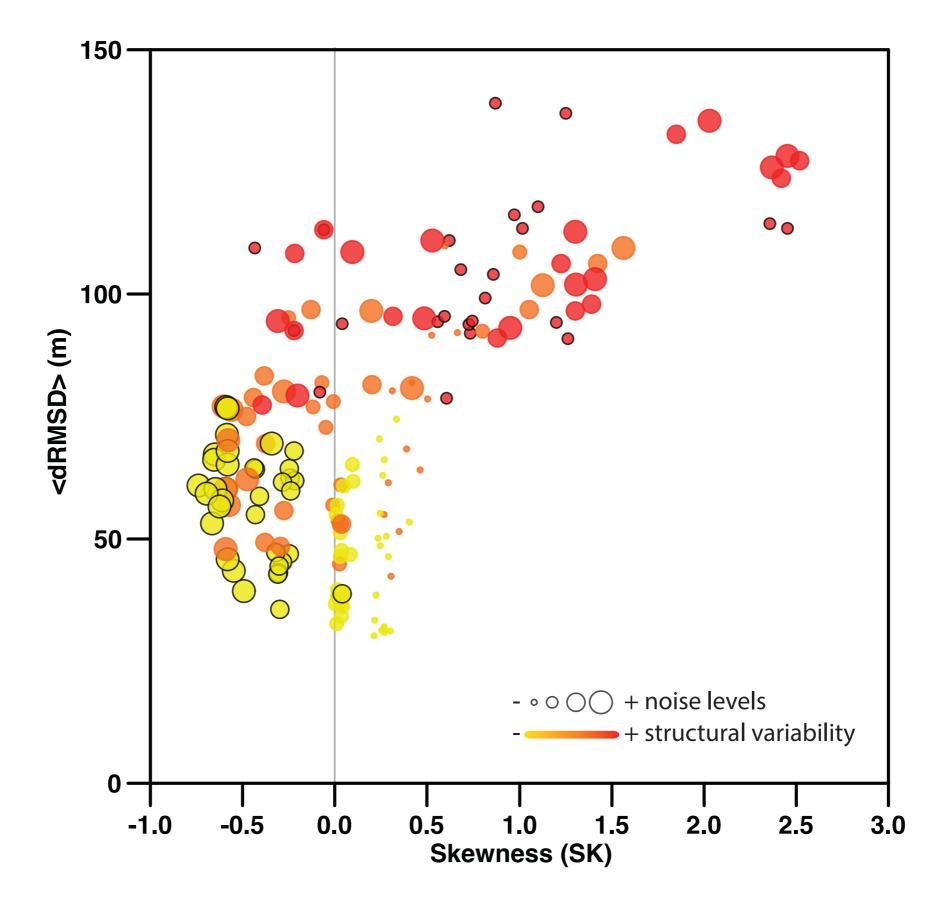
Structural variability is "NOT OK"



Can we predict the accuracy of the models?

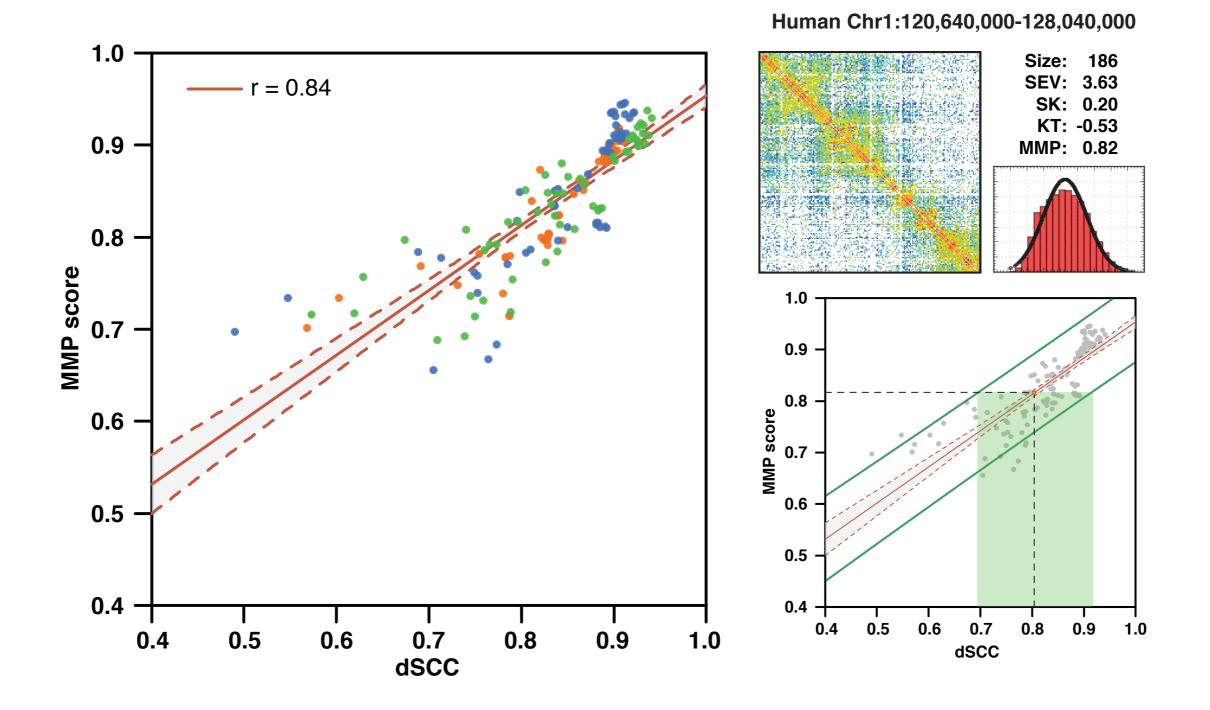


Skewness "side effect"



Can we predict the accuracy of the models?

MMP = -0.0002 * Size + 0.0335 * SK - 0.0229 * KU + 0.0069 * SEV + 0.8126



Higher-res is "good"

put your \$\$ in sequencing

Noise is "OK"

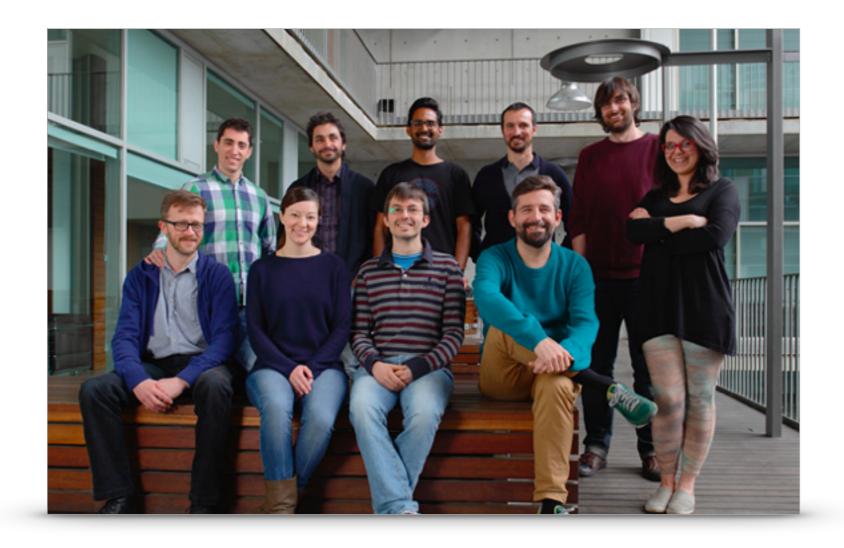
no need to worry much

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



Marie Trussart François Serra Davide Baù

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

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





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