## 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

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)

	IDM			$\begin{array}{c} & & & & \\ & & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & & & \\ & &$	7
				DNA length	
	10 <sup>6</sup>			10 <sup>9</sup>	nt
				Volume	]
10 <sup>-3</sup>		10 <sup>0</sup>		10 <sup>3</sup>	μm³
				Time	
10 <sup>-2</sup>		10 <sup>0</sup>	10 <sup>2</sup>	10 <sup>3</sup>	S
				Resolution	]
			10 <sup>-1</sup>		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



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







### Biomolecular structure determination 2D-NOESY data

Chromosome structure determination 3C-based data





### FastQ files to Maps

Map analysis

Model building

Model analysis





nature structural & molecular biology	
The three-dimensional folding           Davide Balt <sup>A</sup> , Amartya Sanyal <sup>A</sup> , Bryan R.Lajole <sup>A</sup> , End           Davide Salt <sup>A</sup> , Amartya Sanyal <sup>A</sup> , Bryan R.Lajole <sup>A</sup> , End           Davide Balt <sup>A</sup> , Amartya Sanyal <sup>A</sup> , Bryan R.Lajole <sup>A</sup> , End           Davide Salt <sup>A</sup> , Amartya Sanyal <sup>A</sup> , Bryan R.Lajole <sup>A</sup> , End           Davide Balt <sup>A</sup> , Amartya Sanyal <sup>A</sup> , Bryan R.Lajole <sup>A</sup> , End           Davide Salt <sup>A</sup> , Amartya Sanyal <sup>A</sup> , Bryan R.Lajole <sup>A</sup> , Martya Baltan Balta	<text><text><text><text><text><text><text></text></text></text></text></text></text></text>



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

Distinct structural tra of chromatin topologi	cal domains
correlate with coordin hormone-induced gen	
François Le Dily, <sup>1,2,3</sup> Davide Baù, <sup>1,3</sup> Andy Pohl, <sup>1,2</sup> G Daniel Soronellas, <sup>1,2</sup> Giancarlo Castellano, <sup>1,2,4</sup> Roni Guillaume Filion, <sup>1,2</sup> Marc A. Marti-Renom, <sup>1,5,3</sup> and	uillermo P. Vicent, <sup>1,2</sup> François Serra, <sup>1,3</sup> H.G. Wright, <sup>1,2</sup> Cecilia Ballare, <sup>1,2</sup> Miorel Reston <sup>1,2</sup>
<sup>1</sup> Gene Regulacion, Stem Cells, and Cancer Program, Centre de Re Pompeo Fabra (UPF) (8002 Barcelona, Spain, <sup>1</sup> Genome Biology Barcelona, Spain, <sup>1</sup> Hoopital Clinic, Universitat de Barcelona, 080 Avançats (ICREA), 08010 Barcelona, Spain	ulació Genòmica (CRG), 08003 Barcelona, Spain, <sup>2</sup> Universitat Group, Centre Nacional d'Anàlisi Genòmica (CNAG), 08028
(chromosome capture followed by high-throughput sequen	
(chromosome capture followed by high-droughpurit sequence we found that the bodiers of the -2000 TAbs in these cel- that up to 20% of the TADs could be considered as discre- ciclicate transcriptionally avairated or represend in a condi- ing gene activity and chromatin remodeling are accompany represent TADs, netlexed by specification and position that responsive TADs. Indeed, 3D modeling of the IH-C data as treatment. The differential responses of TADs to progesti- "regularma" to enable spatially proximal genes to be coordi (Reyrwork: three-dimensional structure of the genome- tary of the terminal spatial provide the terminal spatial terminal spatial spatial provide the terminal spatial spatial spatial spatial spatial provide the terminal spatial spatial spatial spatial spatial spatial provide the spatial provide the spatial s	ring), and three-dimensional [10] modeling techniques, is are largely maintained after hormour terustnent and e regulatory units where the majority of the genes are ted lashion. The engigenetic signatures of the TADa are the transcriptional changes. Hormone-induced changes of by differential structural changes for activated and ges in the strength of intra-TAD interactions within gested that the structure of TADs was modified upon as and estrogens suggest that TADa could function as antely transcribed in response to hormones.
(chromsome capture followied by high-dramphur segine we imoult durch bounders of the -2000 This in these cell of the second transmission of the second transmission of the chiefer transmission of the second transmission of the inspect of the second second second second transmission in gene activity and dromantin remodeling are accompany represent TADs, as related by postering that and opposite that the second second second second second second second respective transmission of the second second second second regulatory is consistent and the second second second second [ <i>Keyworki</i> , three-dimensional structure of the genome ( <i>fegurentic</i> ) indexego progressiones receptor]	ring), and three-dimensional [10] modeling techniques, is are largely maintained after hormour terustnent and e regulatory units where the majority of the genes are ted lashion. The engigenetic signatures of the TADa are the transcriptional changes. Hormone-induced changes of by differential structural changes for activated and ges in the strength of intra-TAD interactions within gested that the structure of TADs was modified upon as and estrogens suggest that TADa could function as antely transcribed in response to hormones.
(chromosome capture followied by high-throughpuir sequence we found that the borders of the -2000 TAbs in these cel- that up to 20% of the TADs could be considered as discre- cichet transcriptionally activated or presents al a contribu- ingene activity and chromain remeabiling are accompany repressed TADs, as reflected by specific and opposite chan repearsive TADs. Indeed, 3D modeling are accompany treatment. The differential responses of TADs to progesti "reguloms" to enable spatially provide lequidation for the table reguloms" to enable spatially provide lequidation for the table spatial provide lequidation for the spatial provide lequidation for the table spatial provide lequidation for the spatial provide lequidation for the spatial pro	ring), and three-dimensional (DI) modeling techniques, are lengthy maintend after bornsome returners and an employmed and the programmer of the TADs are for flaking. The opigenetic signatures of the TADs are the transcriptional density. However, observed charges generation of the strength of the strength of the generative strength of ismu-TAD interactions within a strength of the strength of the strength of the generative strength of the strength of the strength analytic strength of the strength of the strength of the gene expression, HE-C, TADs, transcriptional regular
(chromsome capitar followich y high-drampjari sogier or lonal thach the sources of the 2,000 TADs index cell cited on the source of the source of the source of the cited transcriptionally activated or represend in a conflict interaction with the source of the source of the represent TADs, an effected by specific and appoint ends in the source of the source of the source of the source represents TADs, and reflected by appointed in the IG-d at an source of the source of the source of the source of the source regulator to enable spatially proximal genes to be could programmer the source of the source of the promo- gengenetic landscape programmers receptor) Supplemental material available for this article.	ring), and three-dimensional (DI) modeling techniques, are lengthy maintend after bornsome returners and an employmed and the programmer of the TADs are for flaking. The opigenetic signatures of the TADs are the transcriptional density. However, observed charges generation of the strength of the strength of the generative strength of ismu-TAD interactions within a strength of the strength of the strength of the generative strength of the strength of the strength analytic strength of the strength of the strength of the gene expression, HE-C, TADs, transcriptional regular



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

# TADbit previous applications...





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





Trussart et al. Nature Communications (2017)

# Structuring the **COLORs** of chromatin





### Fly Chromatin **COLORs** Filion et al. (2010). Cell, 143(2), 212–224.



Position on chr2L (kb)



### Principal component analysis





Hidden Markov model

### Fly Chromatin **COLORs** Hou et al. (2012). Molecular Cell, 48(3), 471–484.









# Structural properties

50 1Mb regions. 10 enriched for each color.



Accessibility (%)

Density (bp/nm)





### BLUE dense region 3L:210000-1230000

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



### BLACK dense region 2L:1750000-18530000

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



### Interactions

Angle



# Structural **COLORs**







# Structural **COLORs**





### Structural **COLORs**



### Position on chr2L (kb)

160	000	16200	16400	16600	16800	1700
		1		1	1	
RG15		an and and an				il operation
R)3-7	al and the second	a dana dana dana dana dana dana dana da				in mailtaining
R)3-9 HP6		a James and a second state				
HP1	Antipatron and a state of		- And a second a state of the second state of			the standard set
LHR		- Jean and a second second second second second				
CAF1		a contact a second stand through the				-
ASF1	and the survey of the second	a statement and the state of a state	an and a sufficiency of the	and a stand of the second stands and	arguestio . Advance . althouse dur.	-
JS209	Anton burden arrentered	a deserves executed and a second and	an and a second second second	an entered all an adda to State and	and the shall and a said a surger of	-
TOP1		a person and the part of the p	and a state of the	and the states of the second states and the	Here and the Marries as many and and	na ndraun sport
RPII18	and the second s	and the second se	Annound the second state of the second state of the	which the second se	and a state of the same of the second state of	-
SIR2	-	- Statement Spanners, Span	and the state of t		and a start of the second start	in little and
RPD3		The state of the s				
CDK7	the second	a prover and the state of the second state of				and the second se
DSP1		to Annenproversetter deliverse and				
DF31	desided a second second second second	a harden and a second sec				and the second sec
MAX	. 1		and the second state of the second state of the second			
PCAF			have a state of the second			
ASH2		an and the first state of the s				
HP1c	The second s	v yperson survey berthat sheet heres			The second s	
CtBP		a glassegurante and a state of the				and the second sec
JRA		an a				
BRM	the second se	a state and a state and be and				
ECR		a provident and the second sec				
BCD IED31	the second	a character and a state of the second state of				
)2-10	the second se	a suma market the and				
OLAL		And and the state of the state				
GAF		And the second s				
31367		a constant of a second distance of the second				
CT5C		a cherrow as a strate that have				
TIP60	Addition to be addition of the		the survey of th		THE REPORT OF THE PARTY OF THE	and the state of t
MNT						
SIN3A						
TBP		ور المعالم الم الله الم الم الم الم الم الم الم				
DWG	undoughter hundraugeneters		and a stand of the second stand of the second		and an address of the second second	ny stantages
PHOL	and the second s	and the second sec	and a second state of a second	Action and a second second second	and the second day in the	a shuman
PROD	and and and and and	a present and the stranger and	and a state of the second	a man market and a second		
F32b		a second and the second second and a second	and the state of the second	In a design of the second second second	and a star a defended a defender and	
J(HW)		to dependence and the same second stands	Barker, Second	and the same a street of a same state	and a second	-
ЪАМ́		and and a state of the second se	and the second state of th	and a second	and the second	
D1	-		where we are a set of the set of			
H1						
SUUR		a deservation and the second sec			a substant of the second se	
EFF		ويستحدر فيجرب وفلي ويرود وصفير فستعار حدودوه عر				
IAL		a demander and a second second second second				
GRO		energian and a survey and				
PHO	and the second se	a warne and a state that the state of the st				
CTCF		a married an and the second and				
PC						
		a comp and a second second a delawarder	annual and the state of the second	a la cara la cara la cara la cara la cara da cara da cara da cara la cara la cara da cara da cara da cara da c	and the state of the second state of the secon	
E(Z)	the second		the second se			
E(Z) PCL SCE	-					

## Are the models correct? Trussart et al. NAR (2015)







Baù (2011) Nature Structural & Molecular Biology





Acemel (2016) Nature Genetics



Jhunjhunwala (2008) Cell



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



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



Giorgetti, (2014) Cell







Research

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<sup>1,2</sup>, François Serra<sup>3,4</sup>, Davide Baù<sup>3,4</sup>, Ivan Junier<sup>2,3</sup>, Luís Serrano<sup>1,2,5</sup> and Marc A. Marti-Renom<sup>3,4,5,\*</sup>

<sup>1</sup>EMBL/CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG), Barcelona, Spain, <sup>2</sup>Universitat Pompeu Fabra (UPF), Barcelona, Spain, <sup>3</sup>Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), Barcelona, Spain, <sup>4</sup>Genome Biology Group, Centre Nacional d'Anàlisi Genòmica (CNAG), Barcelona, Spain and <sup>5</sup>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 meanfield 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 epresentation 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, he 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://creative.com/doc/access/article/ac ons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

# Toy models (168)











150 bp/nm







set 0 ( $\Delta$ ts = 10<sup>0</sup>)



- set 1 ( $\Delta ts = 10^{1}$ )
- set 2 ( $\Delta$ ts = 10<sup>2</sup>)





## Toy interaction matrices























1Mb 🍹

set 4 (∆ts=10<sup>4</sup>)











## Reconstructing toy models











chr40\_TAD **α=100 ∆ts=10** TADbit-SCC: 0.91 <dRMSD>: 32.7 nm <dSCC>: 0.94

chr150\_TAD **α=50** ∆ts=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" + structural variability 150 125 -0 0 0 0 0 ♦ • $\diamond$ 0 r = -0.6725 0 \_ 0.6 0.7 0.8 0.9 1.0 0.4 0.5 **TADbit-SCC**



## 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



## Can we predict the accuracy of the models?

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



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



homogenize your cell population!

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

and we can a priori predict the accuracy of the models

Higher-res is "good" put your \$\$ in sequencing

Noise is "OK"

no need to worry much

Structural variability is "NOT OK"



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

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



### Marie Trussart Davide Baù

Gireesh K. Bogu David Castillo Yasmina Cuartero Irene Farabella Silvia Galan Mike Goodstadt Julen Mendieta François Serra Paula Soler Yannick Spill Marco di Stefano

