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)

Knowl	edge								
1 ANT A					IDM			$\begin{array}{c} 6 & 11 & X & 12 & 15 & 6 & 10 \\ 5 & & & & & & & \\ 5 & & & & & & & \\ 5 & & & &$	
								DNA length	
10 [°]		10 ³			10 ⁶			10 ⁹	nt
								Volume	
10 ⁻⁹		10 ⁻⁶	10 ⁻³	3		10 ⁰		10 ³	μm³
								Time	
10 ⁻¹⁰	10 ⁻⁸	10 ⁻⁶	10 ⁻⁴	10 ⁻²		10 [°]	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





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)





Structuring the **COLORs** of chromatin





Fly Chromatin **COLORs**

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



GREEN

YELLOW

Active

RED

Active

BLUE

Polycomb

BLACK

Repressive

chromatin.

Nucleus

proteins

HP1



Fly Chromatin **COLORs**

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





~200 regions of ~5Mb each 2Kb resolution



Model accuracy Boettiger, A. N., et al. (2016). Nature, 1–15.



Structural properties

50 1Mb regions. 10 enriched for each color.



Structural **COLORs**



Structural **COLORs**



Structural **COLORs**



	Position on chr2L (kb)									
16000		16200	16400	16600	16800	1700				
MRG15	demonstration a	- solution and a second		and the second		printing remarks				
SU(VAR)3-7 SU(VAR)3-9	A day of the day of th	- description of the second		the set out to a the state of the	Ar an Arrange and a set of the set	And the second second				
HP6	and and an and and and and and and and a	- Jammer or a sa			a mar a surplus up a la la contration	And a stand of the stand of the stand				
HP1		. descentioned and			A second as a state of the second sec					
LHR	approach a subsequence of		1	and a state of the state of the state	and the second sec	sector of the se				
CAF1	and have seen as a second s	A COLORIDATION AND DESCRIPTION	and shared a second sec	differences of the lite of the start is		and all of the second sec				
MUS209	Batacologia annual a	. des	and an law or show the start of		the state of the state of the second state of	and a second second				
TOP1		-	and such and an an and an and the second state		And a state of the section of the se	And strategies and strategies				
RPII18	and the second s	a patricker and the second		Block, Concerning allocation are called	an all adopted and a set of the same and the set	Andrew street and				
SIR2		- steinetenent	and the second s	bå anger og at spille bei attender stat.	and the second secon	particular interaction				
RPD3	mpadity addition of the second second	a straige or provide	and the second state of th	I have a second se	and the second second second second second second	Laborate Indiana				
	all and a second se	A second and so that have	the set of a set of a set of the set of the	hits and the day of the	al date of an and all a life of a life	and the state of t				
DF31	what and the second data wants on		and a second sec		and the second state of th	And Andrew Standards				
MAX		-	and a second sec	I between a data and a read bet	and the second the design of the second	Beargedge allow the second				
PCAF	and a state of the	· valeragerardy de la	and the second s	billing opper a det all an er fillen	and the state of the second	and an end of the second				
ASH2	dampine and a second	·· your and a second se	and the subscreen support on the subscreen subscre	hite and a state of the second state of the se	and the second diversity of the second second second	Second Second Second Second				
HP1C C+PD	entreffferensisterensesterense v	- your any and the second	and the second	Man of the Antibility of the A		and the state of the second				
JRA	and the second s		date of the second date of the s	Mar annaddar dar mar		Contraction of the local division of the loc				
BRM	Antonia and a second second second			Mary	and the second design of the second second	And the state of the second				
ECR	designation of the second second second	a sheers and a strategy a	and a second a second second second	hite or a state of the sectors	and the second sec	Warnen manin				
BCD	didatificante especification es	a unananananananan	and the second	har an	and the second second second second second second	Collection of the section of				
MED31	stanting and starting of a	- Incontration		Law of the loss of the second		Analysis areas				
	to be the second s	A designed and a designed and	Andana de anna constate de antes de	Harris Adda and	materia and in the same start in the	advance destant				
GAF		- starsparsers M. M.	here and all and a second and a second	الأسب ومحمد أسأ مرجو ومحمد وال						
CG31367	approximation and a	a house que far a far and a	and any a series any a series of the series and		an and a second second dear first war as a subject of	And the states				
ACT5C	helicitelliterature of the second sec	· States and the second states and and	and the second second second	-	and a second and the second and the second	principal contraction				
TIP60	for the second s	· · · ·	and the second se	Managen Backson and the	and the second	Cardinal agreeding				
SINISA	alantal the second states of the	a particular and the set		han a state of the second day		Marine and a second				
TBP	and the second s		all a second and a second and a second and a second a s		a all a constant and a second and	Andrew watches				
DWG	hephiliterstructure the	· · · ·	and the second s			production and and an				
PHOL	where the second s	a geterneren ander atter the the		Manual and a strate of the str	and the second state and a second state of the second	Analysian descent				
PROD	contration and a stranger of	a streamport stream	And a state of the		and the second	Waterson and a state of the				
SU(HW)	and a state of the second seco		and defined and have been added	the surface of the plant of the surface		Statements Statements				
LAM		-	al a constant subject a su	dant and a company of the	concentrations with as a concentration					
D1		-	and the state of the second second second second second							
H1			and the second data was a second open	Manage States and which						
SUUR	al and the second second side of the		the state of the s	Marine Andrew Andrew Street		Andrea de austra-b				
EFF	filmente and a state of the sta	a familia and a second for the second s	and a start of the start of the start of the start of the	A share been and the ball of the state		Angle Angle and a sharehouse				
GRO	attraction of the second state of the	a ware burgent & burg	all some addition of the second	I destate out to a best of a sector	and a set of the set o	Acres has been the				
PHO	Stratement for the second	- management	and the second s	- when some surplanters and a president	mannel leasensed a war and	the second second				
CTCF			A sample and a second and a second		and a star a					
PC	Conception of the state of the second s		And a second sec							
E(Z)	debrer and a second sec		And the second sec	bert and a same and a link dauget	the sector in the sector of th	and the second second				
SCF			And a state of the second	And and a state of the second se	and the second days of the second days	Annual and a state of the state				
202										

Are the models correct?









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



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



Junier (2012) Nucleic Acids Research



Baù (2011) Nature Structural & Molecular 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 meanfield restrain-based method, which could be taken into consideration in further development of methods as well as their applications.

INTRODUCTION

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

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 prokarvote 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 organiza-

tion 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.

Hu (2013) PLoS Computational Biology

Toy models



by Ivan Junier

Toy interaction matrices



1Mb

Reconstructing toy models



TADs & higher-res are "good"



Noise is "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



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

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

http://sgt.cnag.cat/www/presentations/



