3D folding of chromosomal domains in relation to gene expression



Marc A. Marti-Renom

http://sgu.bioinfo.cipf.es



Structural Genomics Unit Bioinformatics & Genomics Department Prince Felipe Research Center (CIPF), Valencia, Spain





Can we relate structure and expression?



Complex genomes



Resolution

Limited knowledge...

Knowledge



Integrative and iterative approach



Structure determination

Integrative Modeling Platform

http://www.integrativemodeling.org

Alber et al. Nature (2007) vol. 450 (7170) pp. 683-94



Biomolecular structure determination 2D-NOESY data





Chromosome structure determination 5C data

5C technology

Detecting up to millions of interactions in parallel

http://my5C.umassmed.edu

Dostie et al. Genome Res (2006) vol. 16 (10) pp. 1299-309



5C "copies" the 3C library into a 5C library containing only ligation junctions

Performed at high levels of multiplexing:

2,000 primers detect 1,000,000 unique interactions in 1 reaction

Human α -globin domain

ENm008 genomic structure and environment

ENCODE Consortium. Nature (2007) vol. 447 (7146) pp. 799-816



The ENCODE data for ENm008 region was obtained from the UCSC Genome Browser tracks for: RefSeq annotated genes, Affymetrix/CSHL expression data (Gingeras Group at Cold Spring Harbor), Duke/NHGRI DNasel Hypersensitivity data (Crawford Group at Duke University), and Histone Modifications by Broad Institute ChIP-seq (Bernstein Group at Broad Institute of Harvard and MIT).

Human α -globin domain

ENm008 genomic structure and environment

ENCODE Consortium. Nature (2007) vol. 447 (7146) pp. 799-816



Integrative Modeling

http://www.integrativemodeling.org



Representation

Harmonic

$$H_{i,j} = k \left(d_{i,j} - d_{i,j}^0 \right)^2$$

Harmonic Lower Bound

$$\begin{cases} if \ d_{i,j} \le d_{i,j}^{0}; & lbH_{i,j} = k \left(d_{i,j} - d_{i,j}^{0} \right) \\ if \ d_{i,j} > d_{i,j}^{0}; & lbH_{i,j} = 0 \end{cases}$$

Harmonic Upper Bound

$$\begin{cases} if \ d_{i,j} \ge d_{i,j}^{0}; & ubH_{i,j} = k \left(d_{i,j} - d_{i,j}^{0} \right)^{2} \\ if \ d_{i,j} < d_{i,j}^{0}; & ubH_{i,j} = 0 \end{cases}$$



Scoring



Optimization



Clustering



Not just one solution



Not just one solution

and we can de-convolute them!



Consistency

GM12878 Cluster #1 2780 model





Fragment

Regulatory elements

GM12878 Cluster #1 2780 model





K562 Cluster #2 314 model







Compactness

GM12878 Cluster #1 2780 model





Multi-loops

GM12878 Cluster #1 2780 model





69Kb

K562 Cluster #2

73

314 model



Expression

GM12878 Cluster #1 2780 model







FISH validation













Summary

5C data results in comprehensive interaction matrices to build a consistent 3D model





Summary

Models allow for 5C data de-convolution





Models allow for 5C data de-convolution



Summary

Selected models reproduce known (and new) interactions







Large-scale changes in conformation correlate with gene expression of resident genes





The models have been partially validated by FISH



GM12878

K562









Phillips and Corces. Cell (2009)

Acknowledgments



Davide Baù Postdoctoral fellow



Amartya Sanyal Postdoctoral Fellow

University of Massachusetts

UMASS Medical School



Bryan Lajoie Bioinformatician



Emidio Capriotti Postdoctoral fellow



Meg Byron Research Associate

Jeanne Lawrence Department of Cell Biology University of Massachusetts Medical School

Worcester, MA, USA

Job Dekker

Program in Gene Function and Expression Department of Biochemistry and Molecular Pharmacology University of Massachusetts Medical School Worcester, MA, USA



Structural Genomics Unit Bioinformatics and Genomics Department Centro de Investigación Príncipe Felipe Valencia, Spain





D. Baù, A. Sanyal, B. Lajoie, E. Capriotti, M. Byron, J. Lawrence, J. Dekker, and M.A. Marti-Renom. Nature Structural & Molecular Biology (2010) *in press (5th of December)*.

http://sgu.bioinfo.cipf.es http://integrativemodeling.org

Thursday, November 25, 2010