3D folding of chromosomal domains in relation to gene expression

Marc A. Marti-Renom
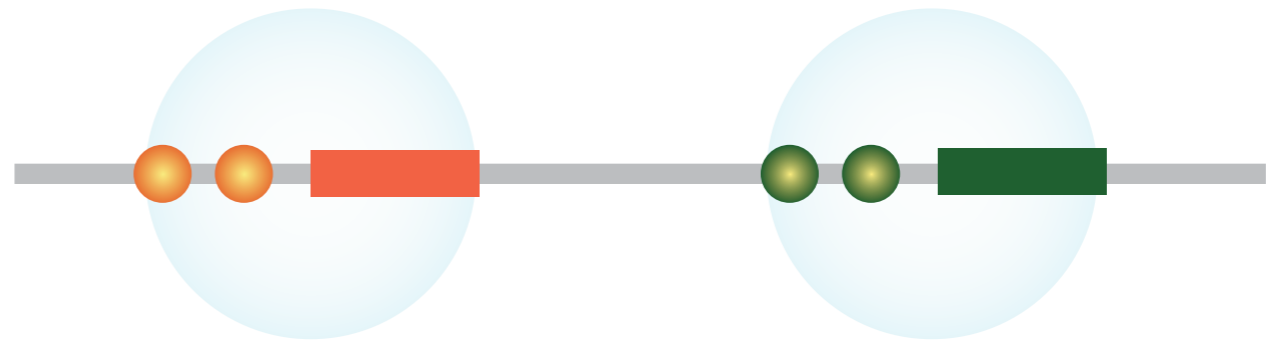http://sgu.bioinfo.cipf.es
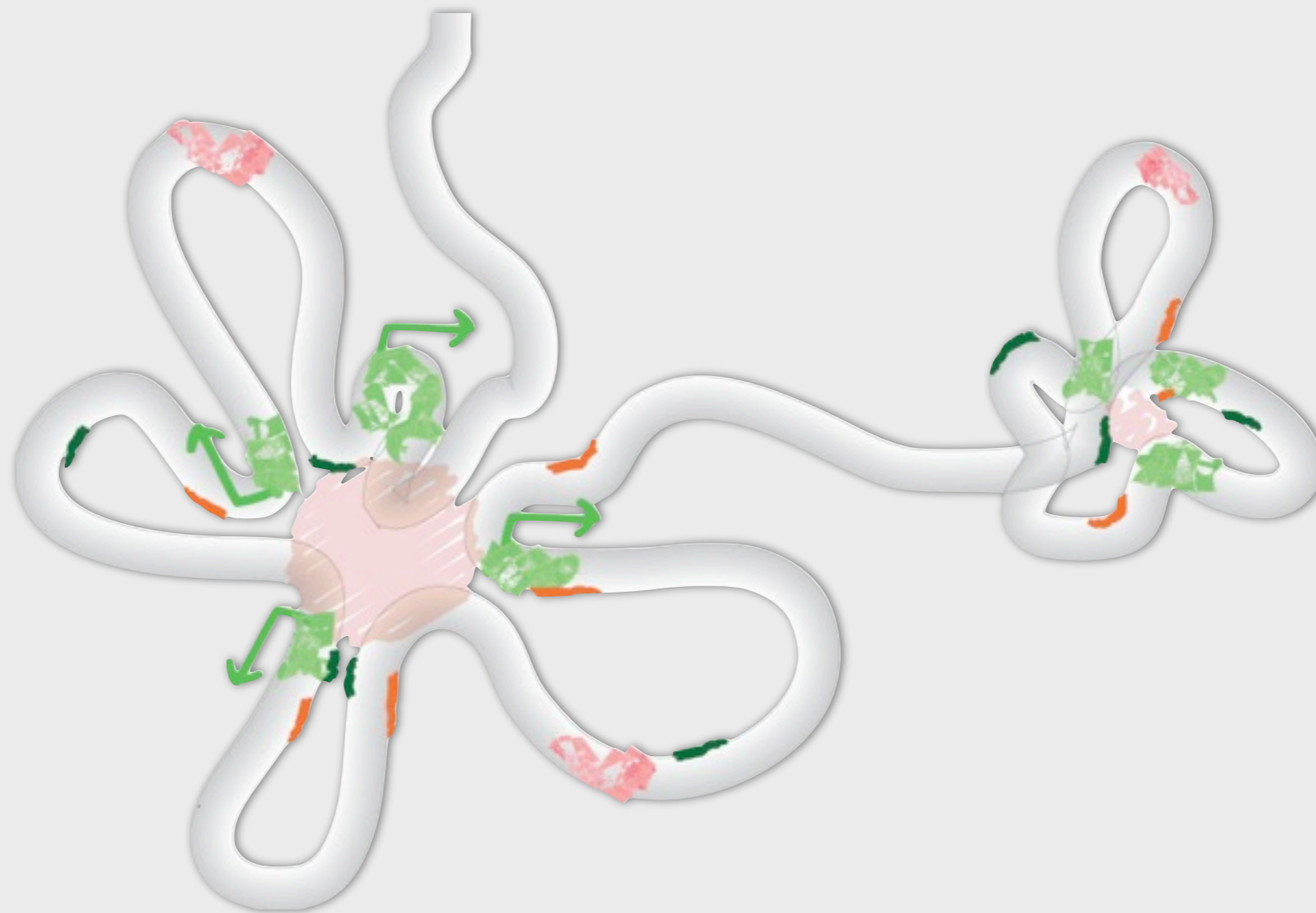
Structural Genomics Unit
Bioinformatics & Genomics Department
Prince Felipe Research Center (CIPF), Valencia, Spain

PRINCIPE FELIPE
CENTRO DE INVESTIGACION
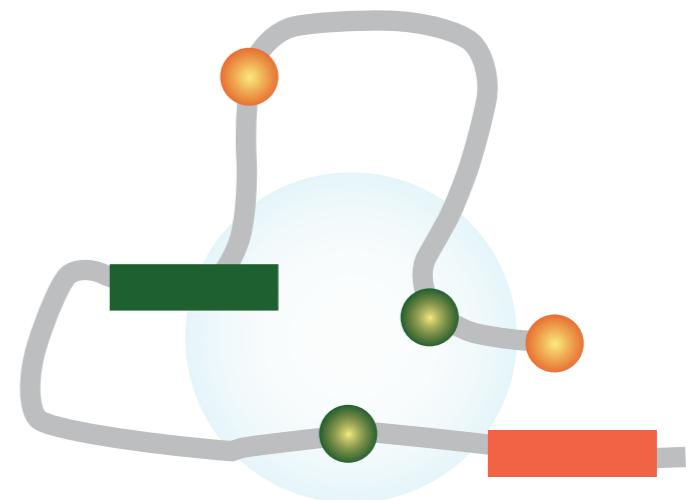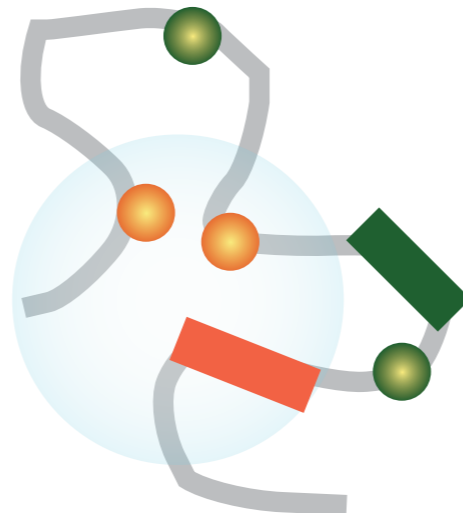
UMASS
MEDICAL
SCHOOL

Wednesday, December 8, 2010

Aim

Can we relate structure and expression?

Simple genomes

Complex genomes

Wednesday, December 8, 2010

Resolution

Limited knowledge...

Knowledge
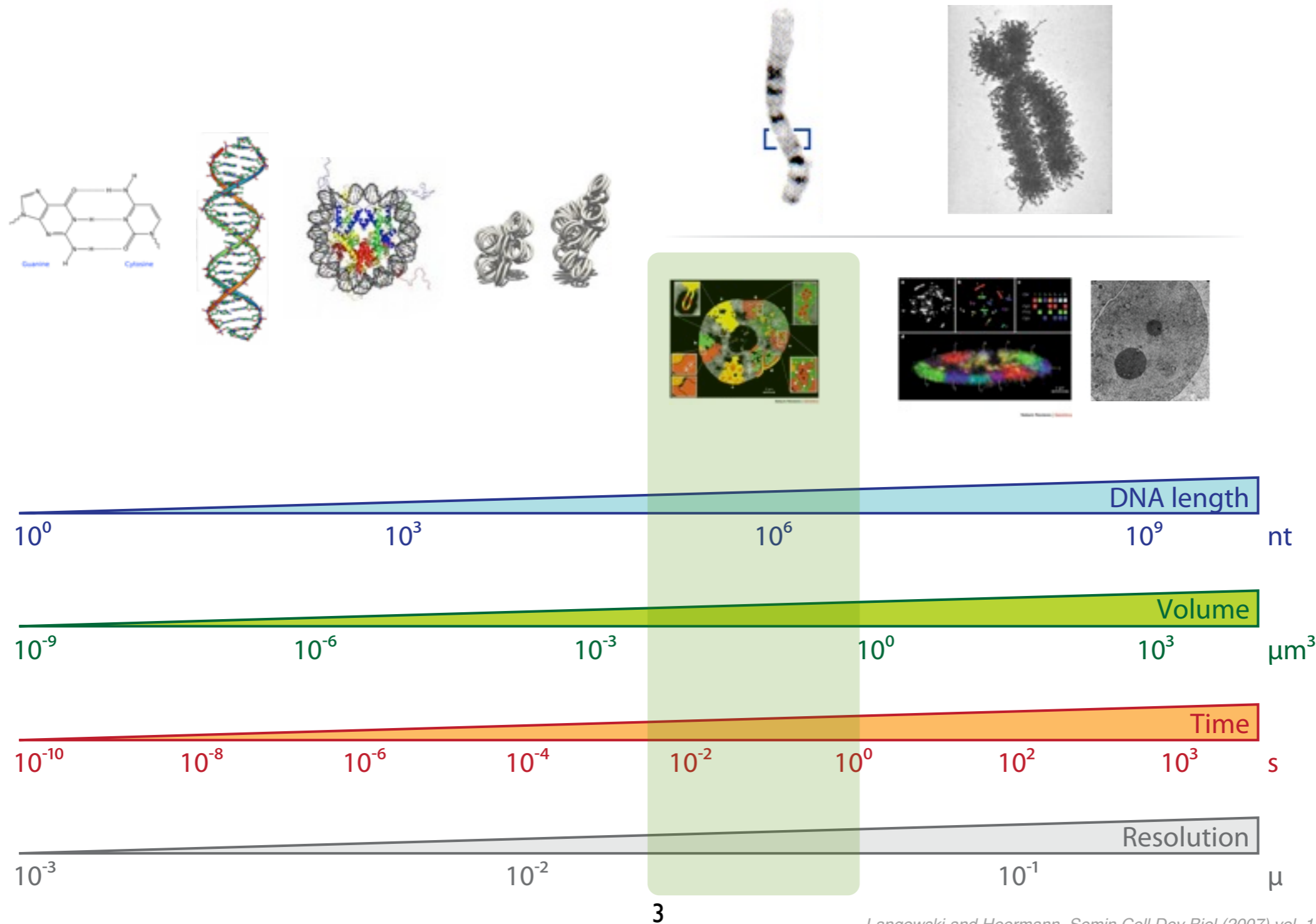
DNA length

| 10^0 | 10^3 | 10^6 | 10^9 | nt |

Volume

| 10^{-9} | 10^{-6} | 10^{-3} | 10^0 | 10^3 | μm^3 |

Time

| 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 |

3

Wednesday, December 8, 2010

Integrative and iterative approach

Experiments

Grow GM12878 and K562 cells

Perform 3C analysis

Perform 5C analysis with 30+25 primers

Analyze 5C products by paired-end Solexa sequencing
(131,947 paired end reads per library)
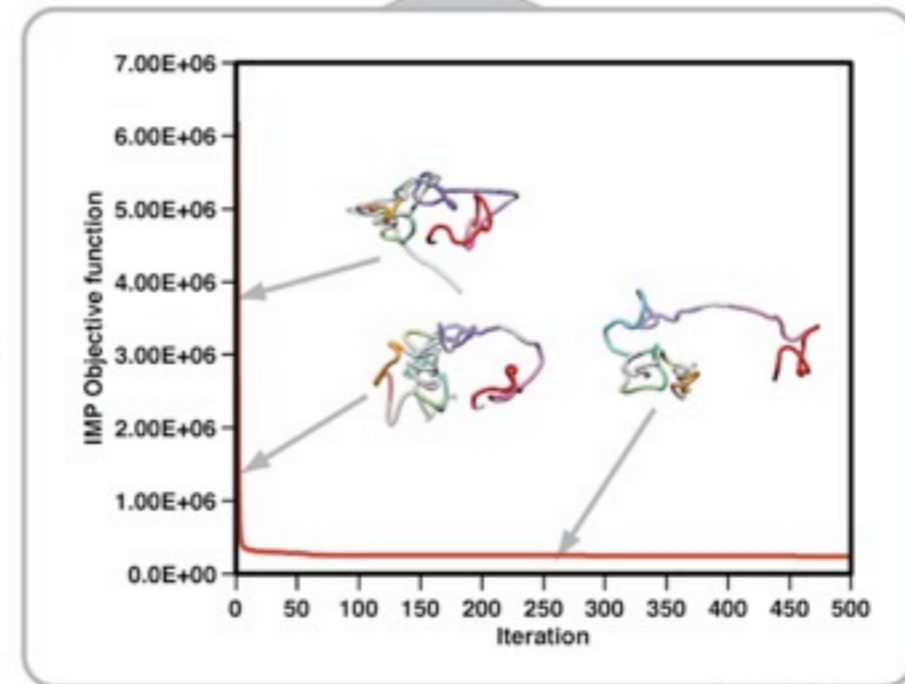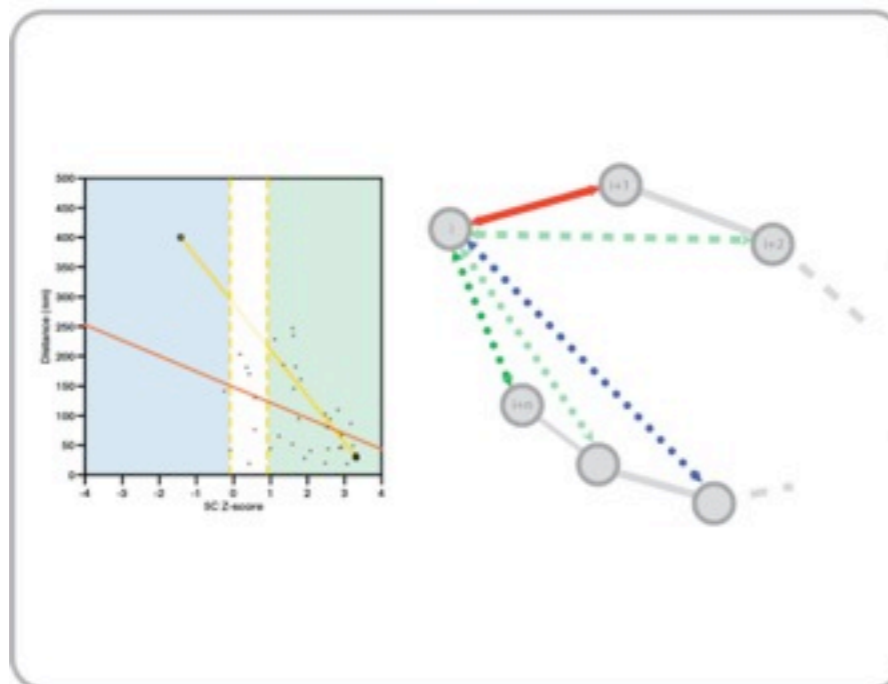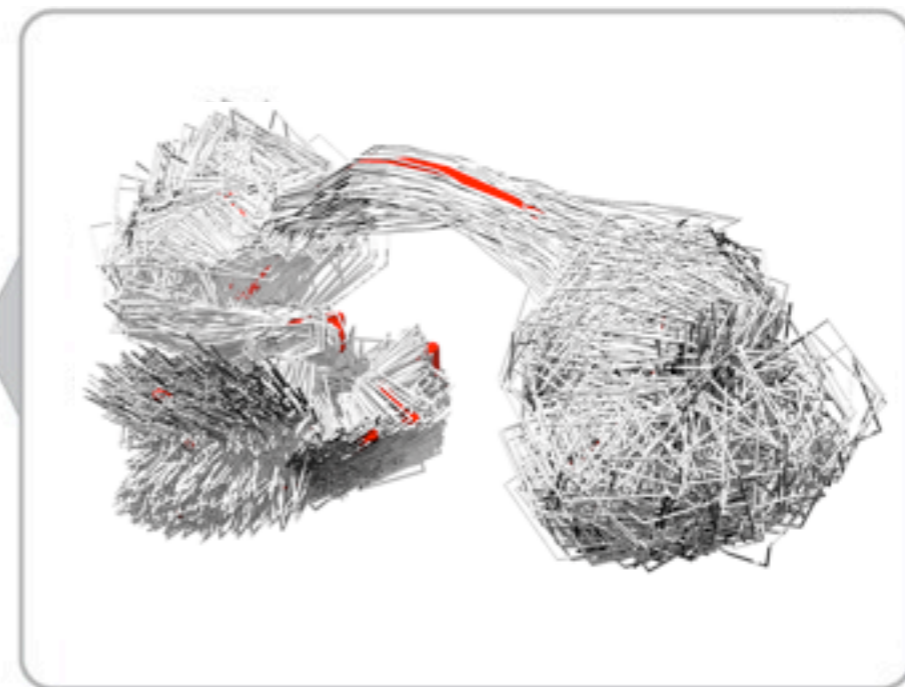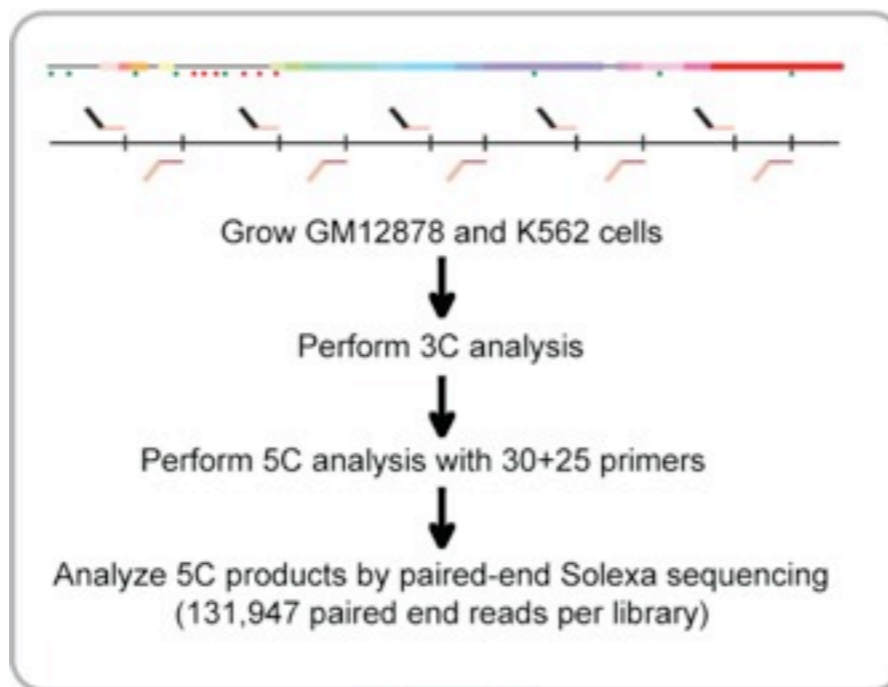
Computation

Structure determination
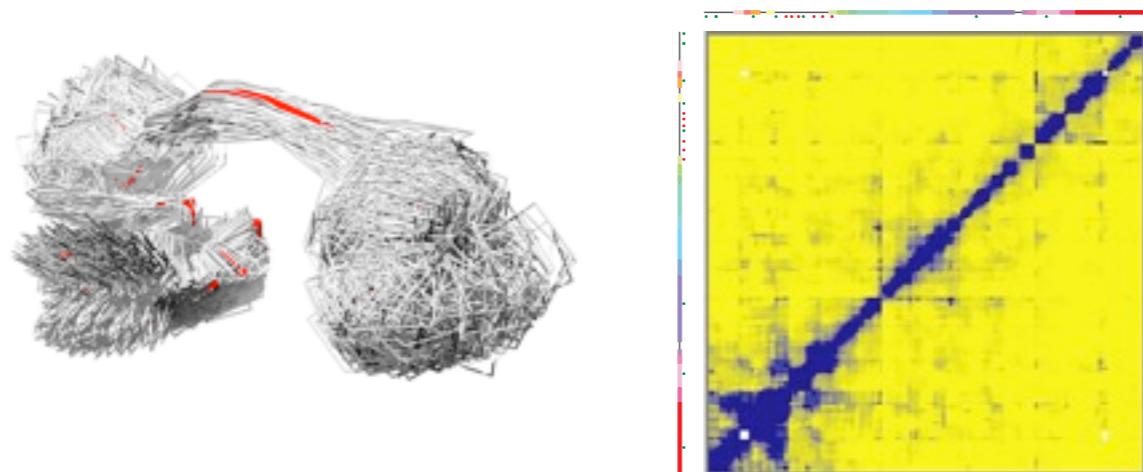
Integrative Modeling Platform

http://www.integrativemodeling.org

Alber et al. Nature (2007) vol. 450 (7170) pp. 683-94
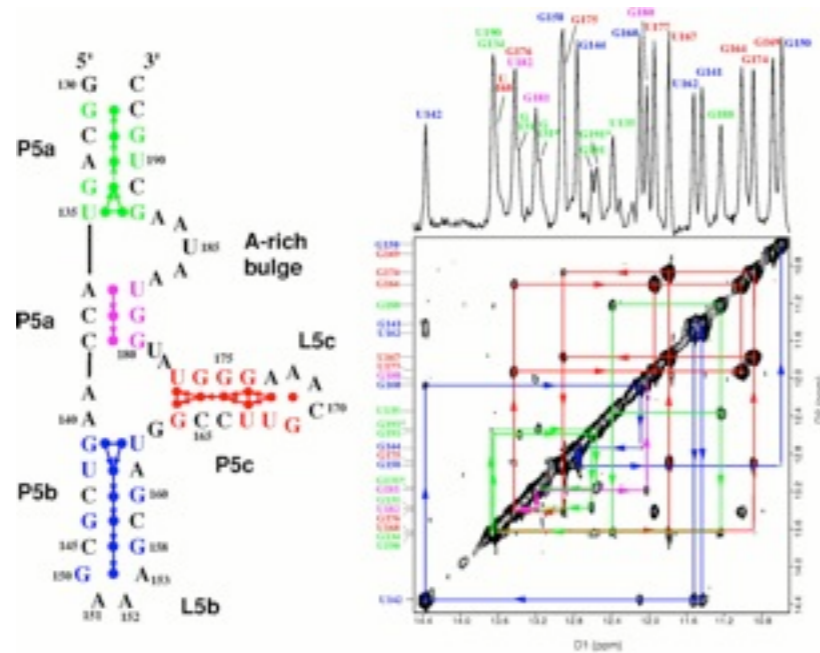
Biomolecular structure determination
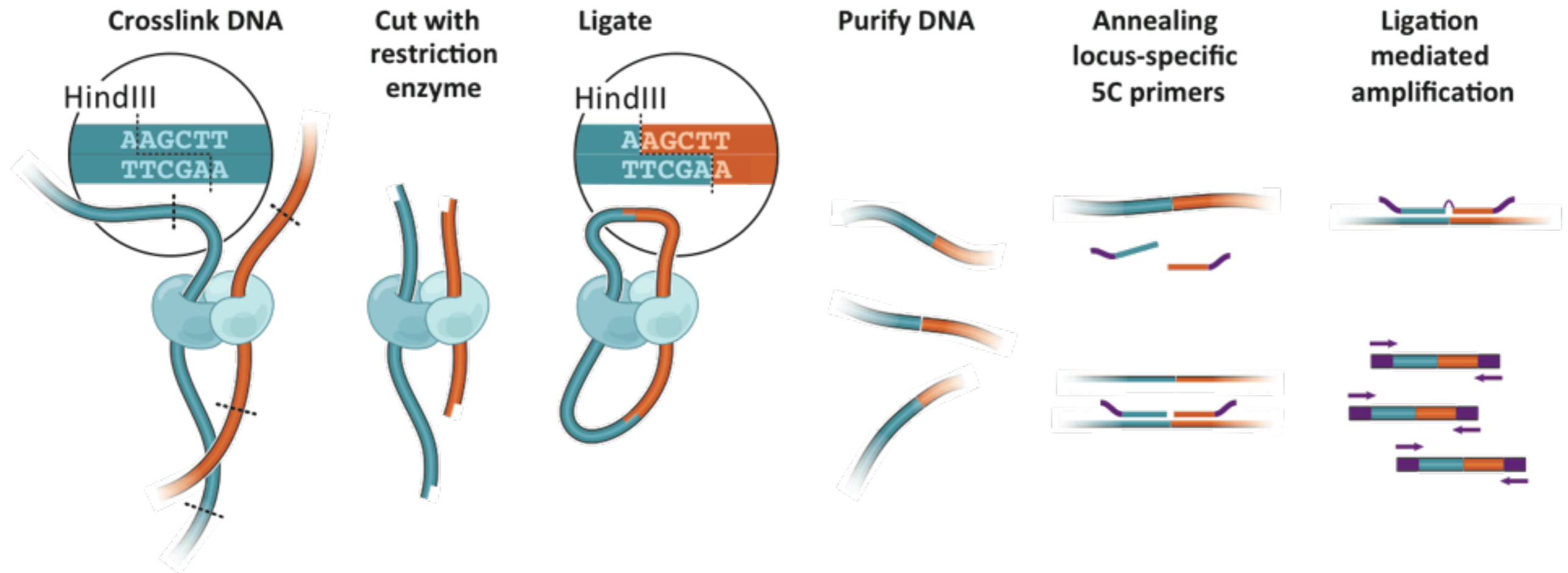*2D-NOESY data*

Chromosome structure determination
5C data

5

5C technology

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

Wednesday, December 8, 2010

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 DNaseI Hypersensitivity data (Crawford Group at Duke University), and Histone Modifications by Broad Institute ChIP-seq (Bernstein Group at Broad Institute of Harvard and MIT).*

Wednesday, December 8, 2010

Human α-globin domain

ENm008 genomic structure and environment

ENCODE Consortium. Nature (2007) vol. 447 (7146) pp. 799-816
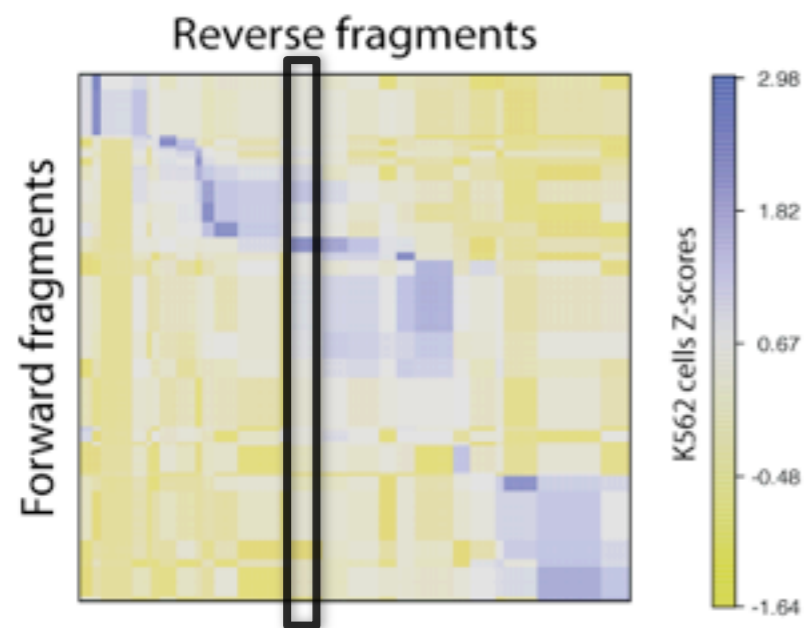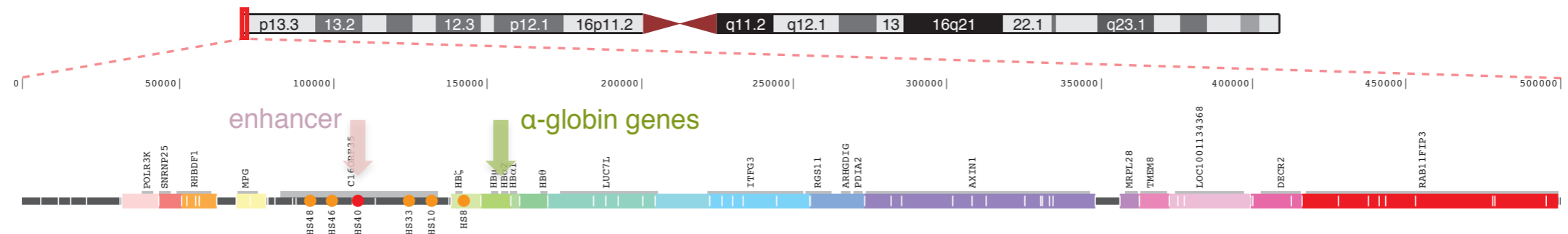
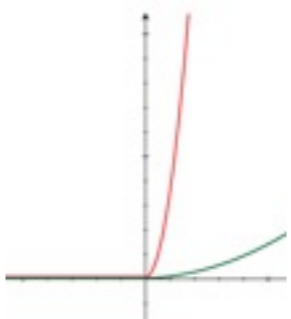K562 cells:
α-globin genes active

Integrative Modeling
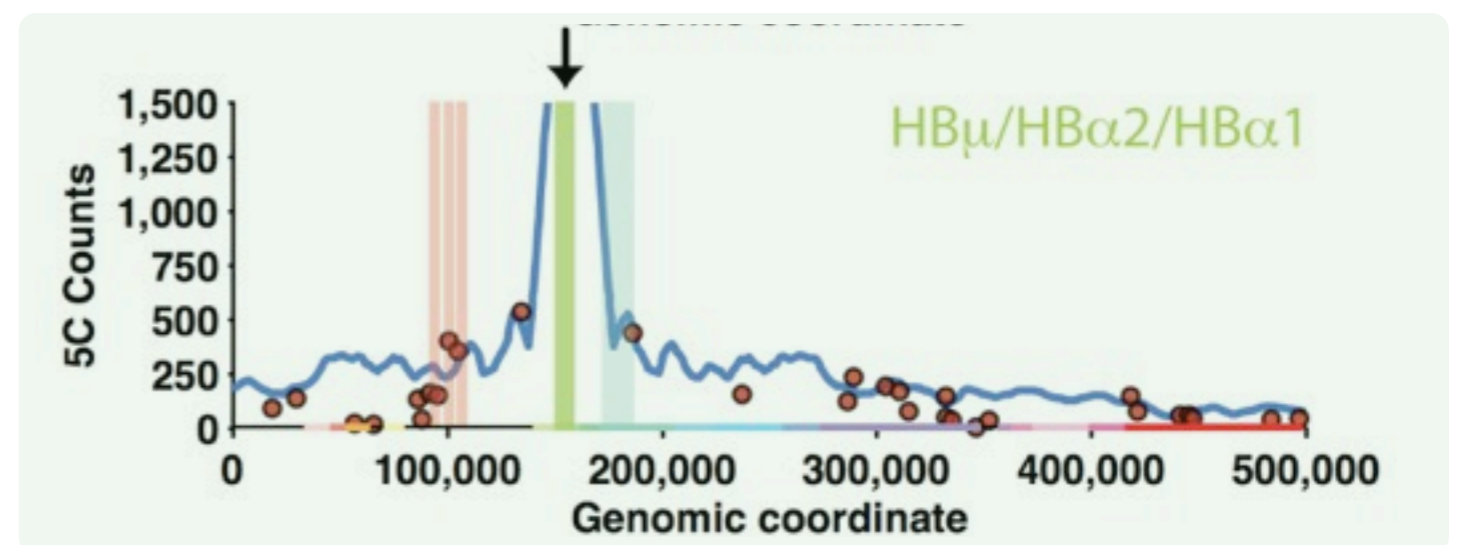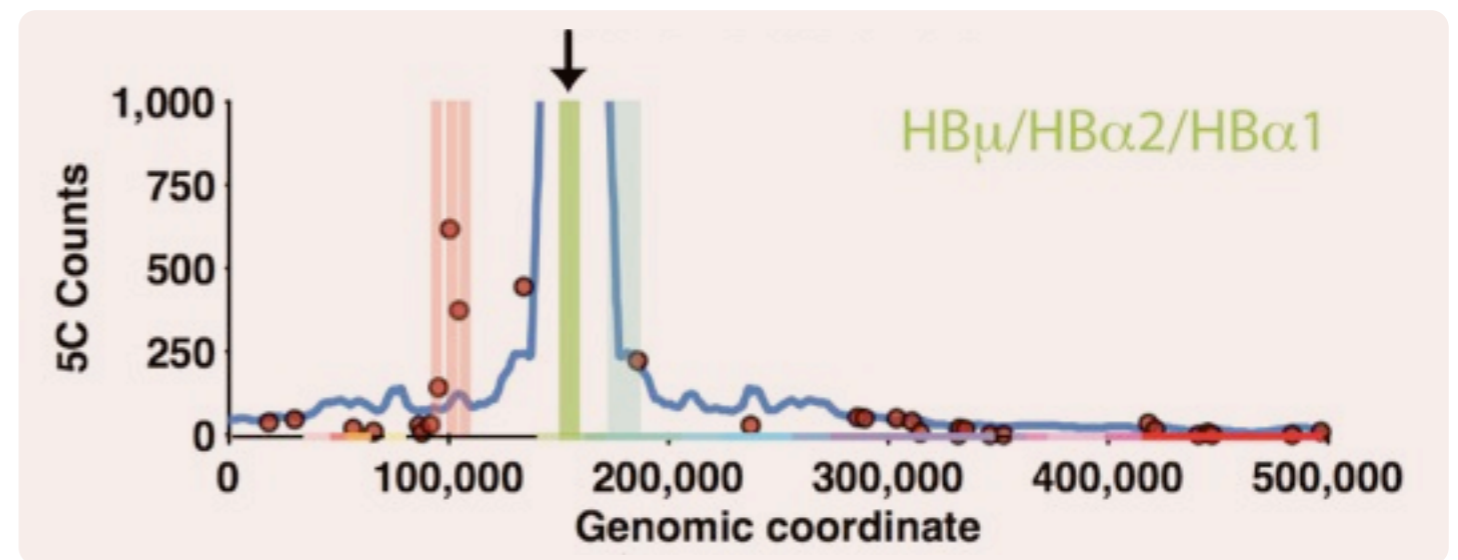
http://www.integrativemodeling.org
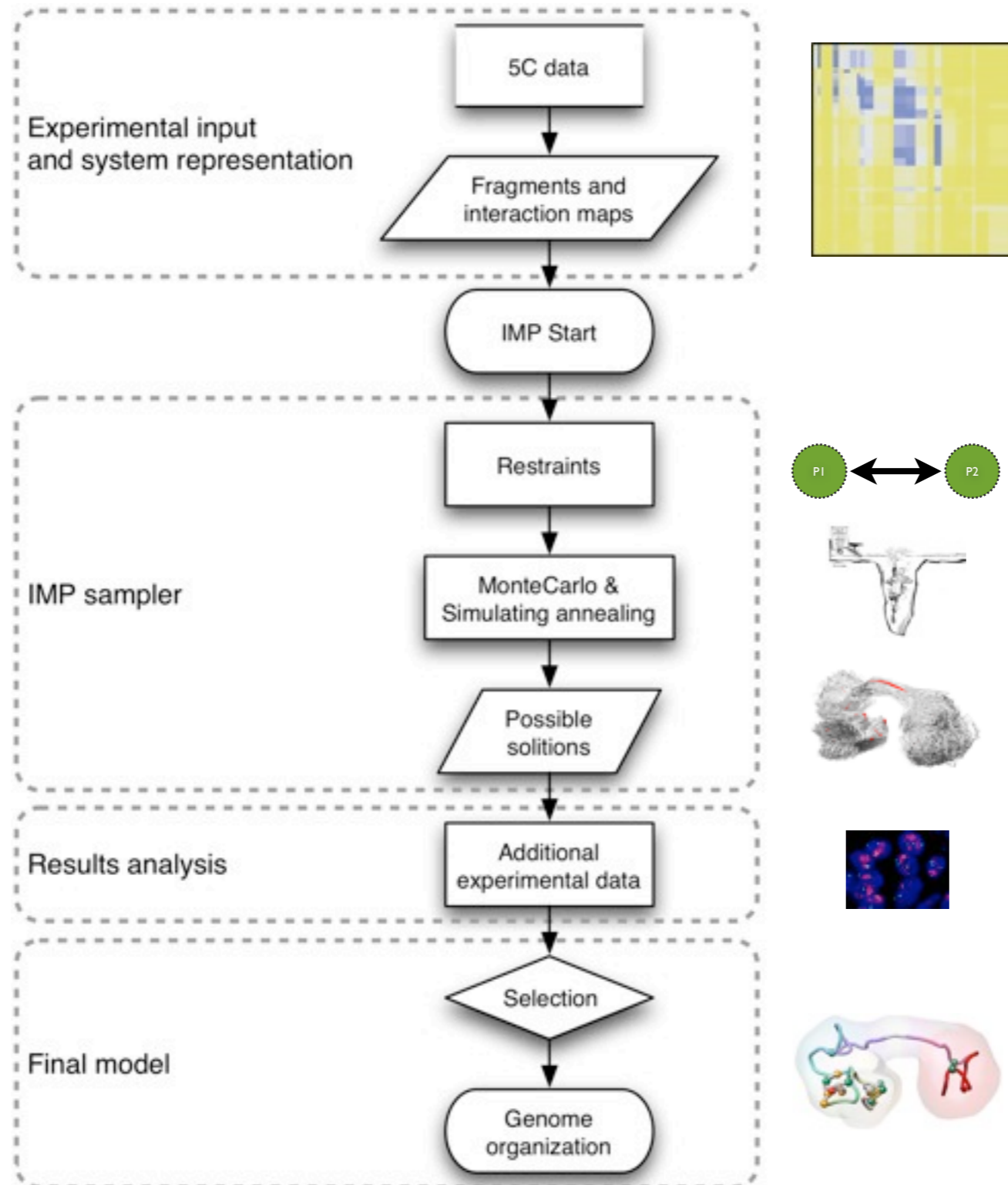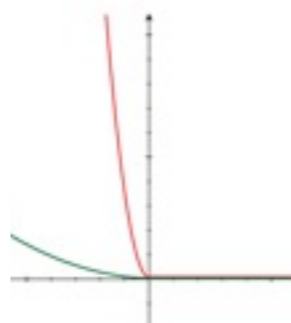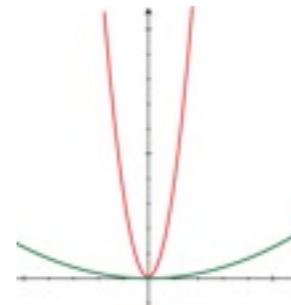
Representation

Harmonic

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

Harmonic Lower Bound

$$\begin{cases} if \;\; d_{i,j} \leq d_{i,j}^0; & lbH_{i,j} = k\left(d_{i,j} - d_{i,j}^0\right)^2 \\[2ex] if \;\; d_{i,j} > d_{i,j}^0; & lbH_{i,j} = 0 \end{cases}$$

Harmonic Upper Bound

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

Wednesday, December 8, 2010
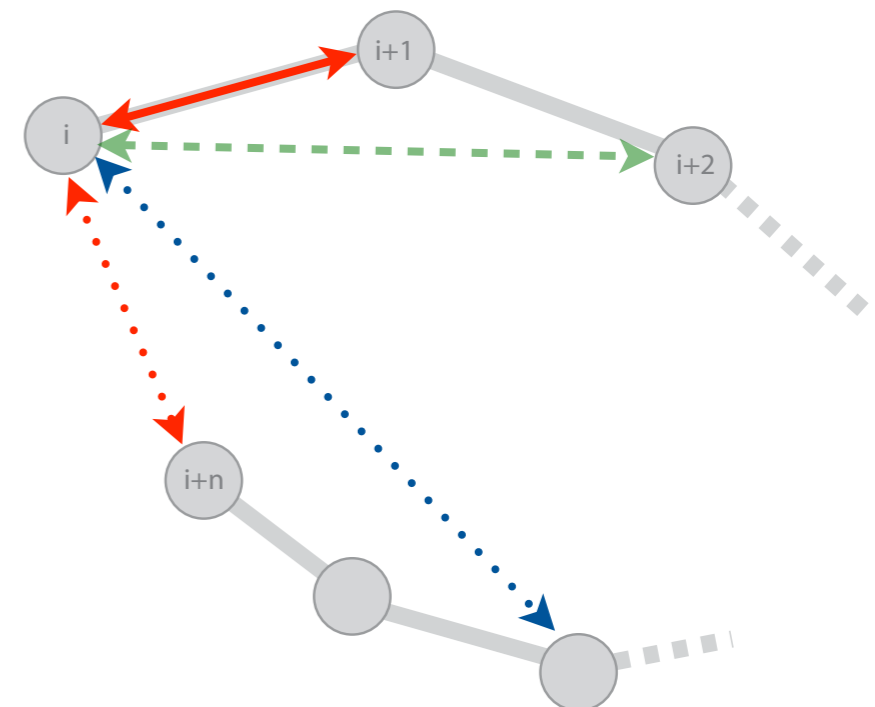
Scoring

GM12878

70 fragments
1,520 restraints
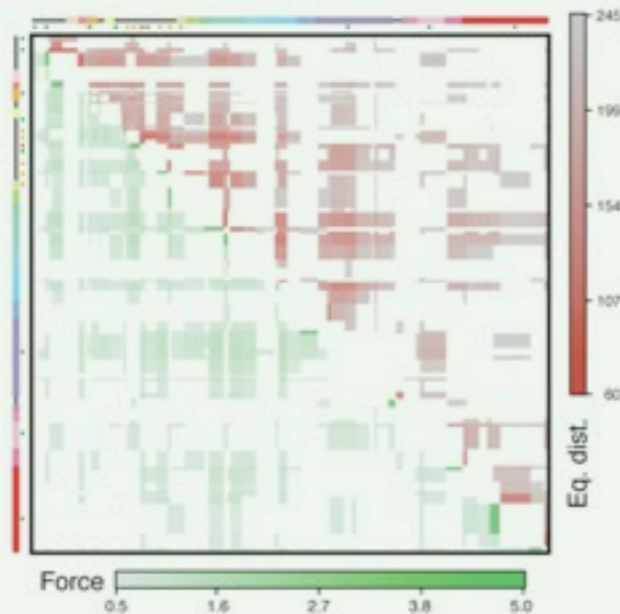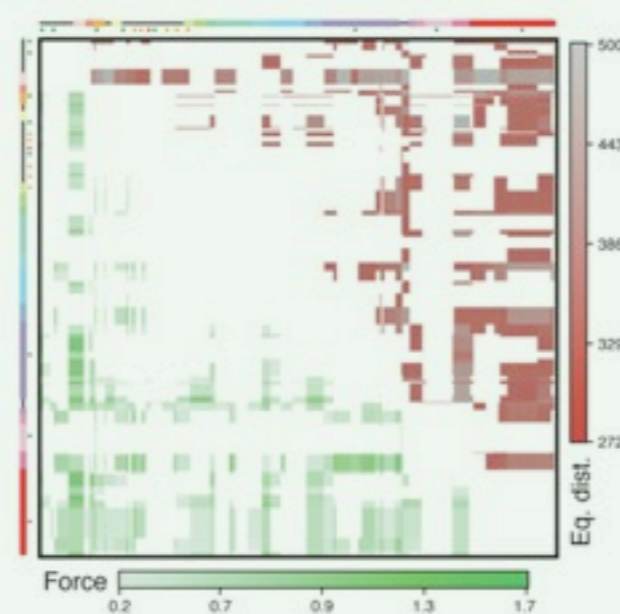
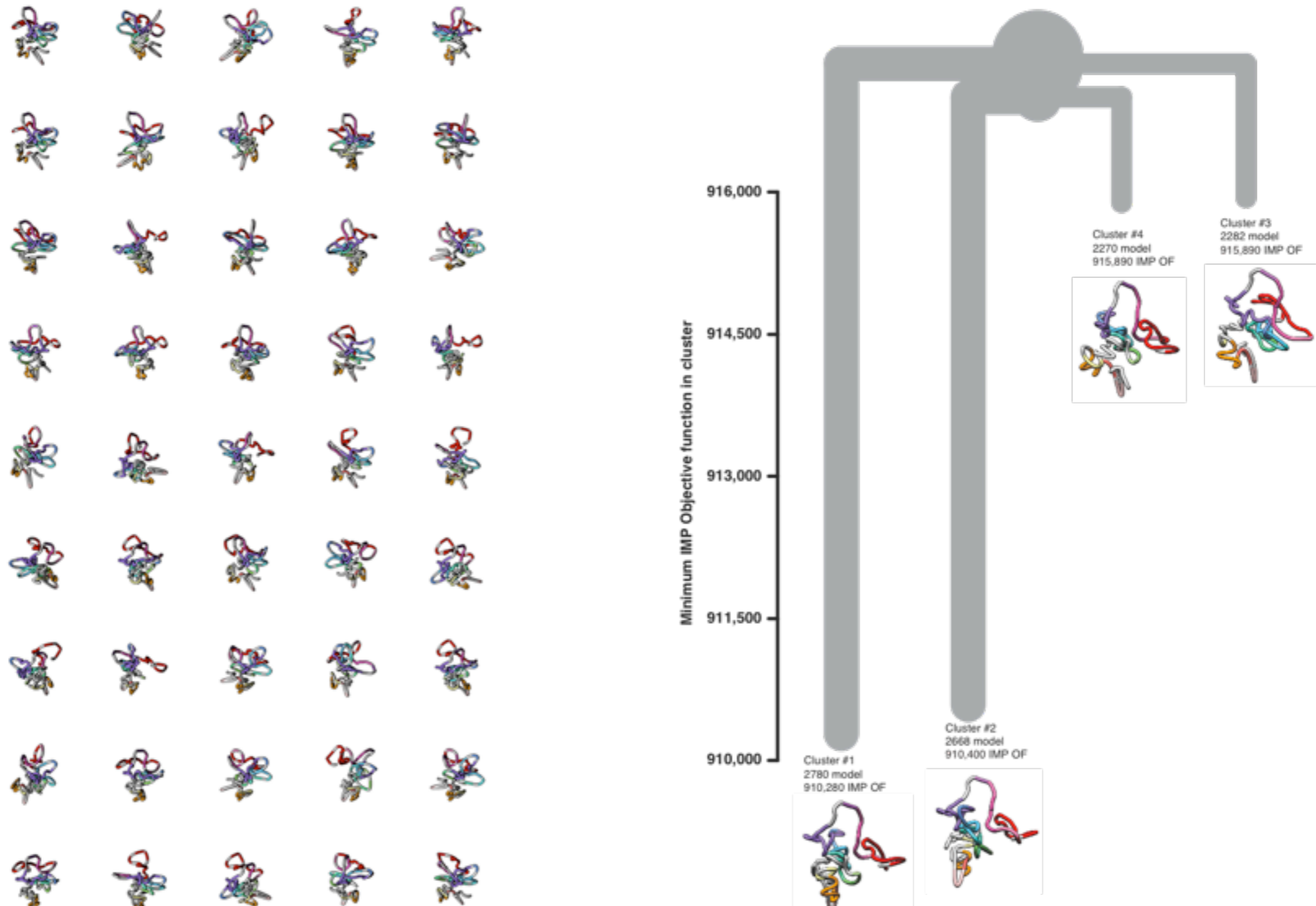Harmonic   Harmonic Lower Bound   Harmonic Upper Bound
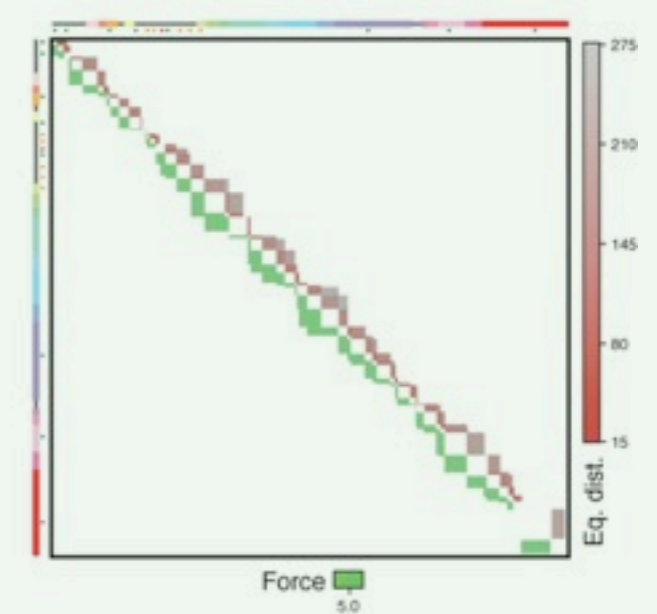
K562
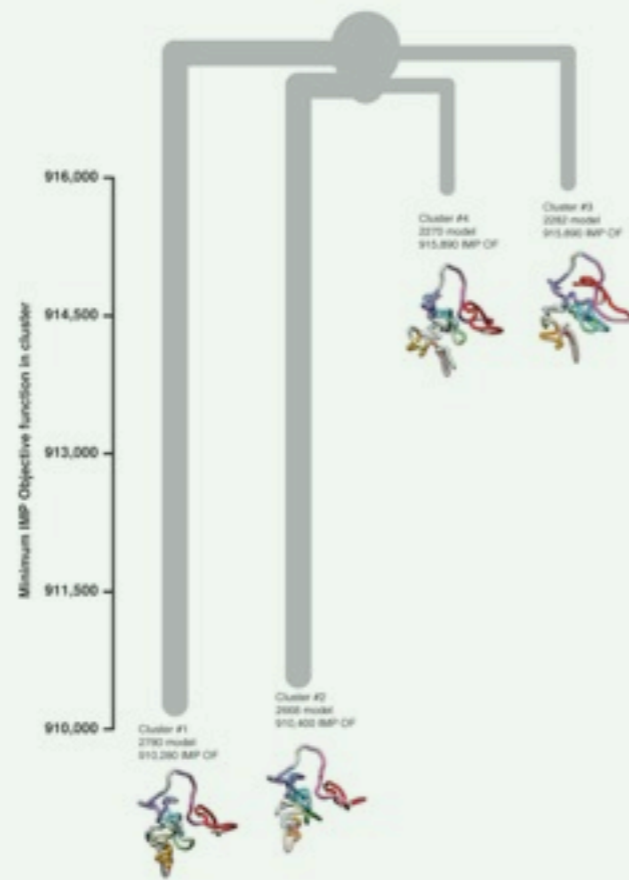
70 fragments
1,049 restraints
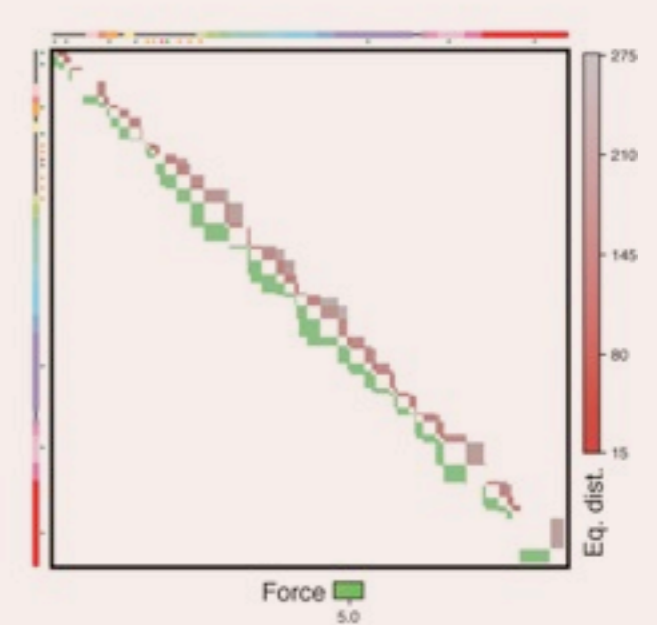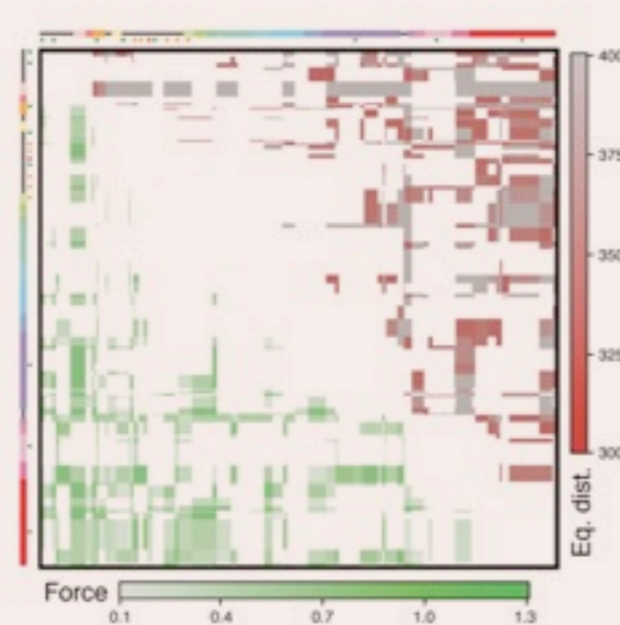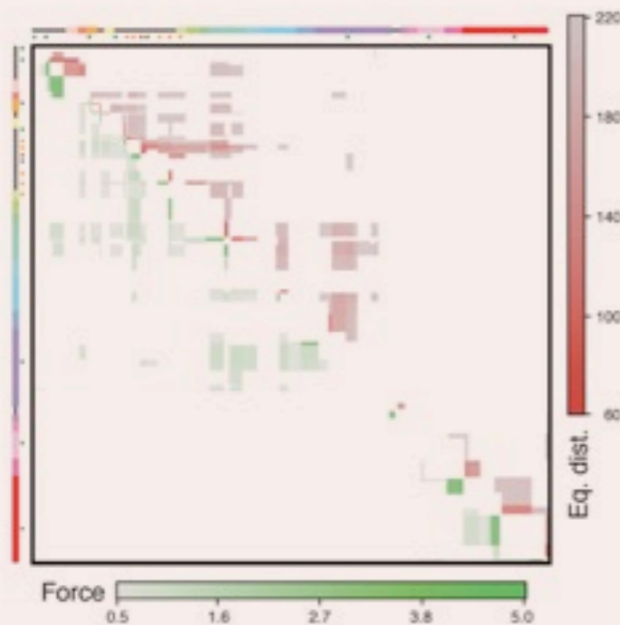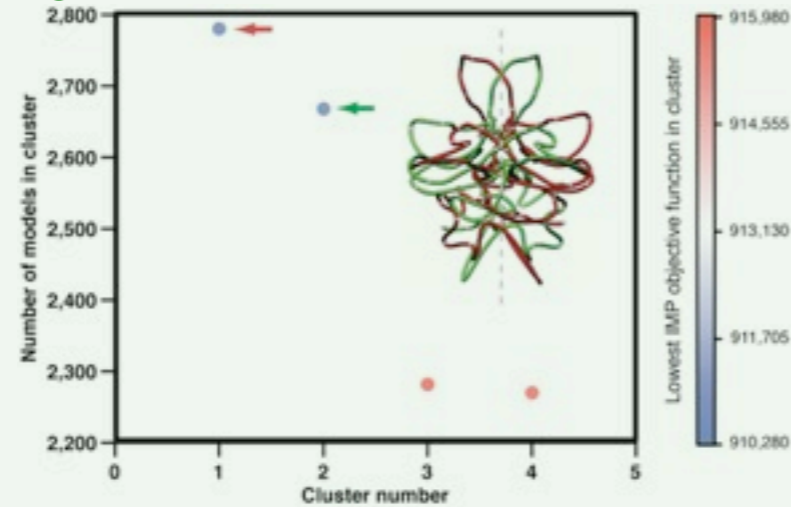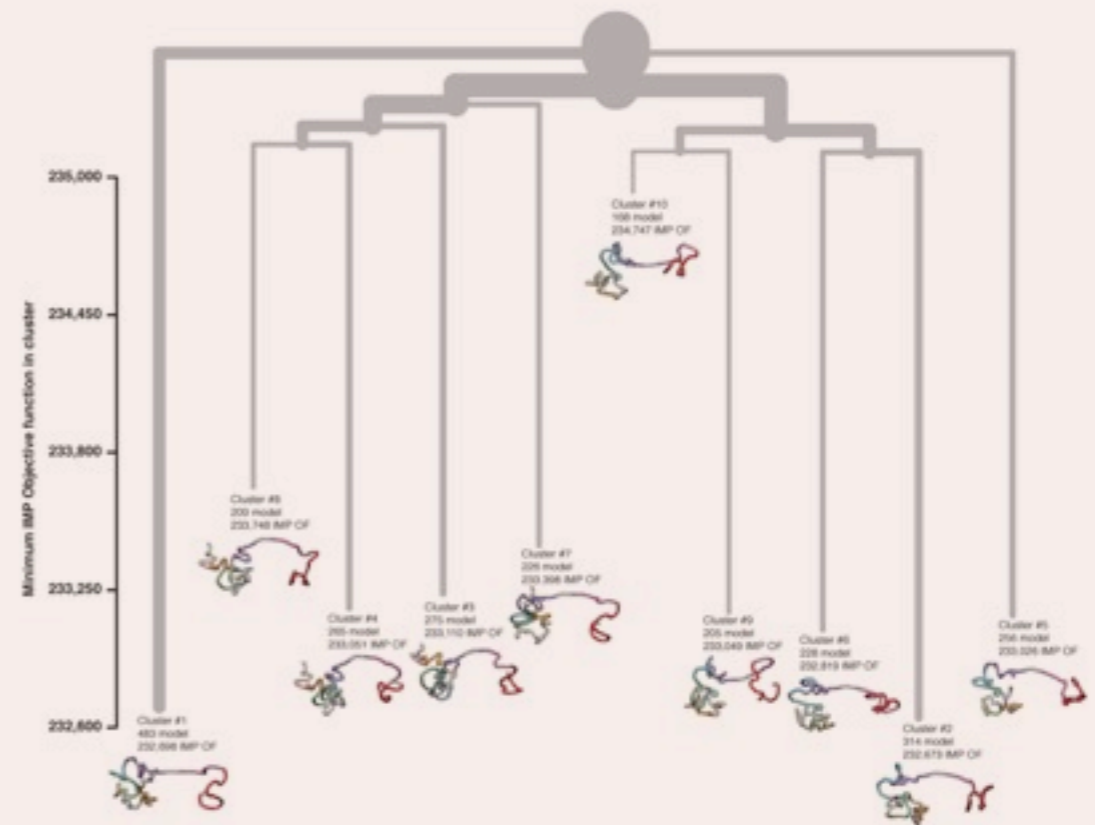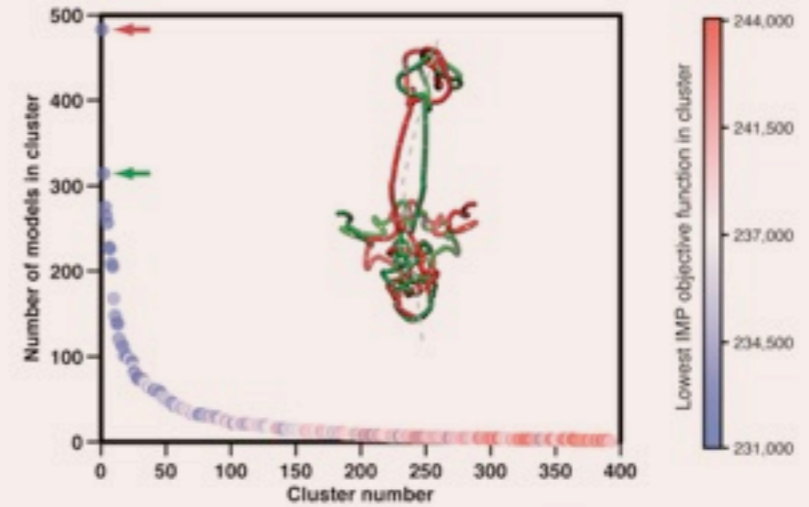
Optimization
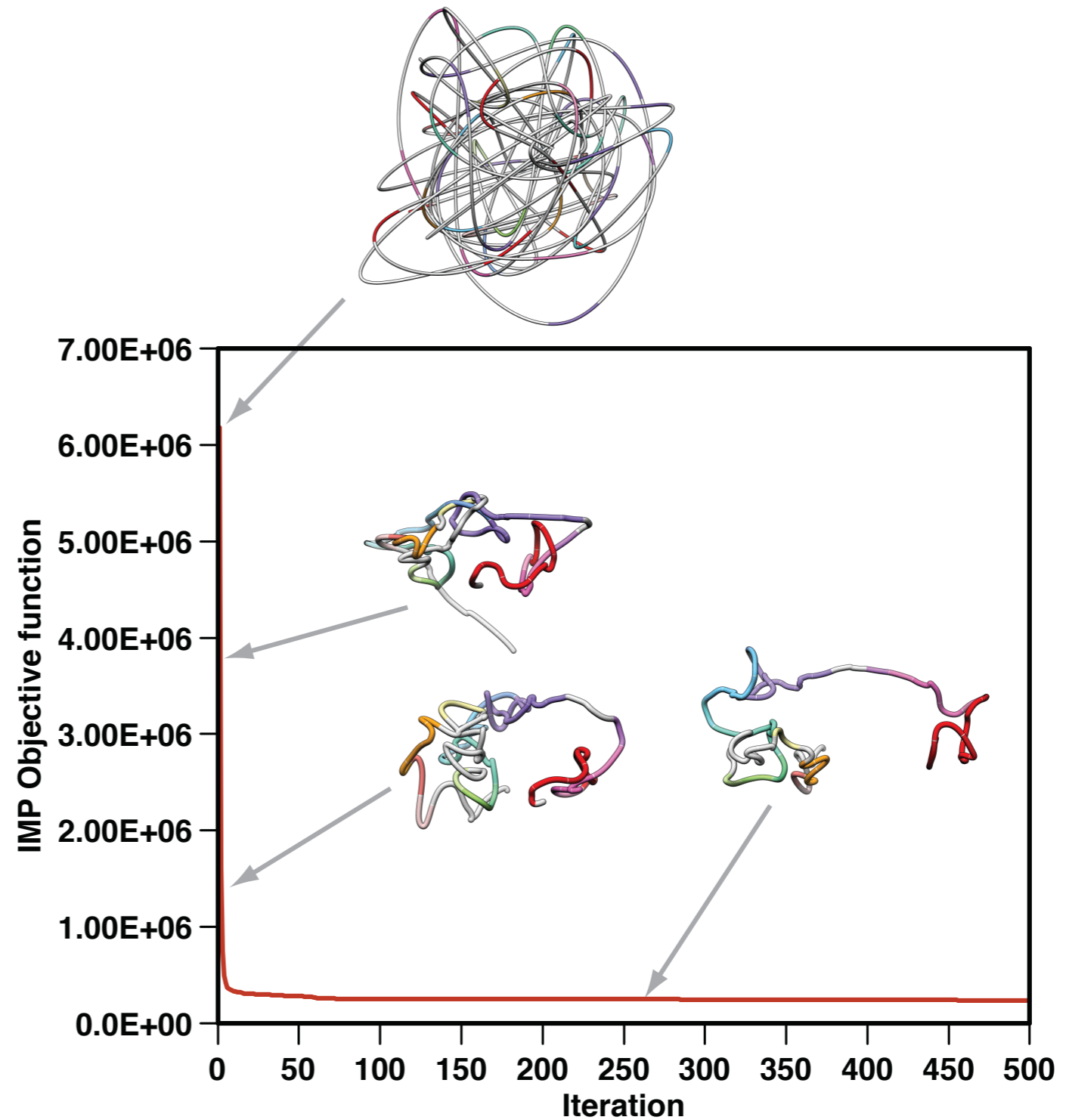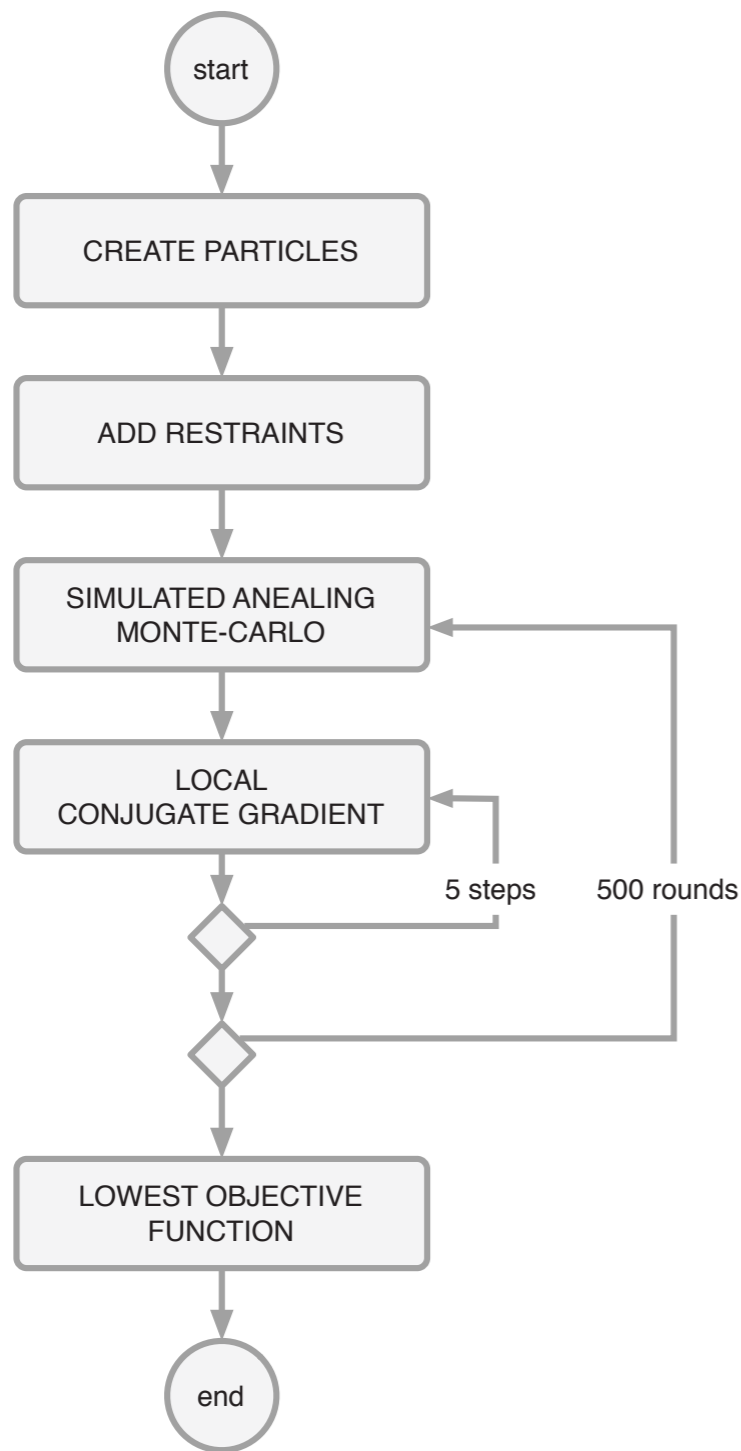
Clustering

Not just *one* solution

GM12878

K562

Not just *one* solution

and we can de-convolute them!

Consistency
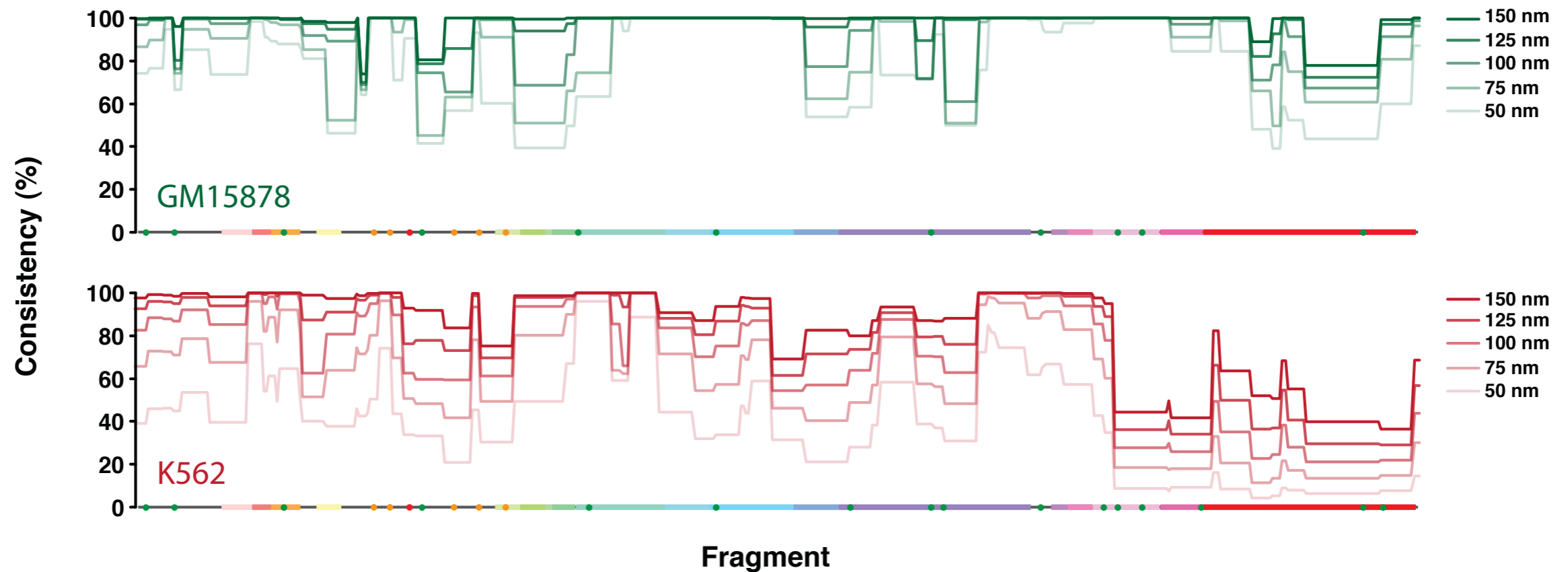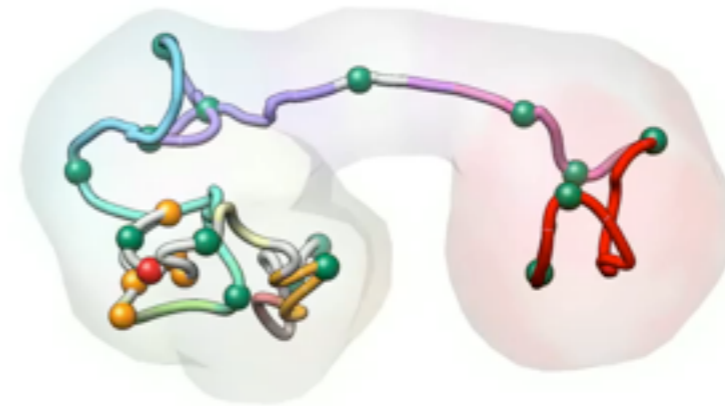
Regulatory elements

Compactness

Multi-loops

Expression
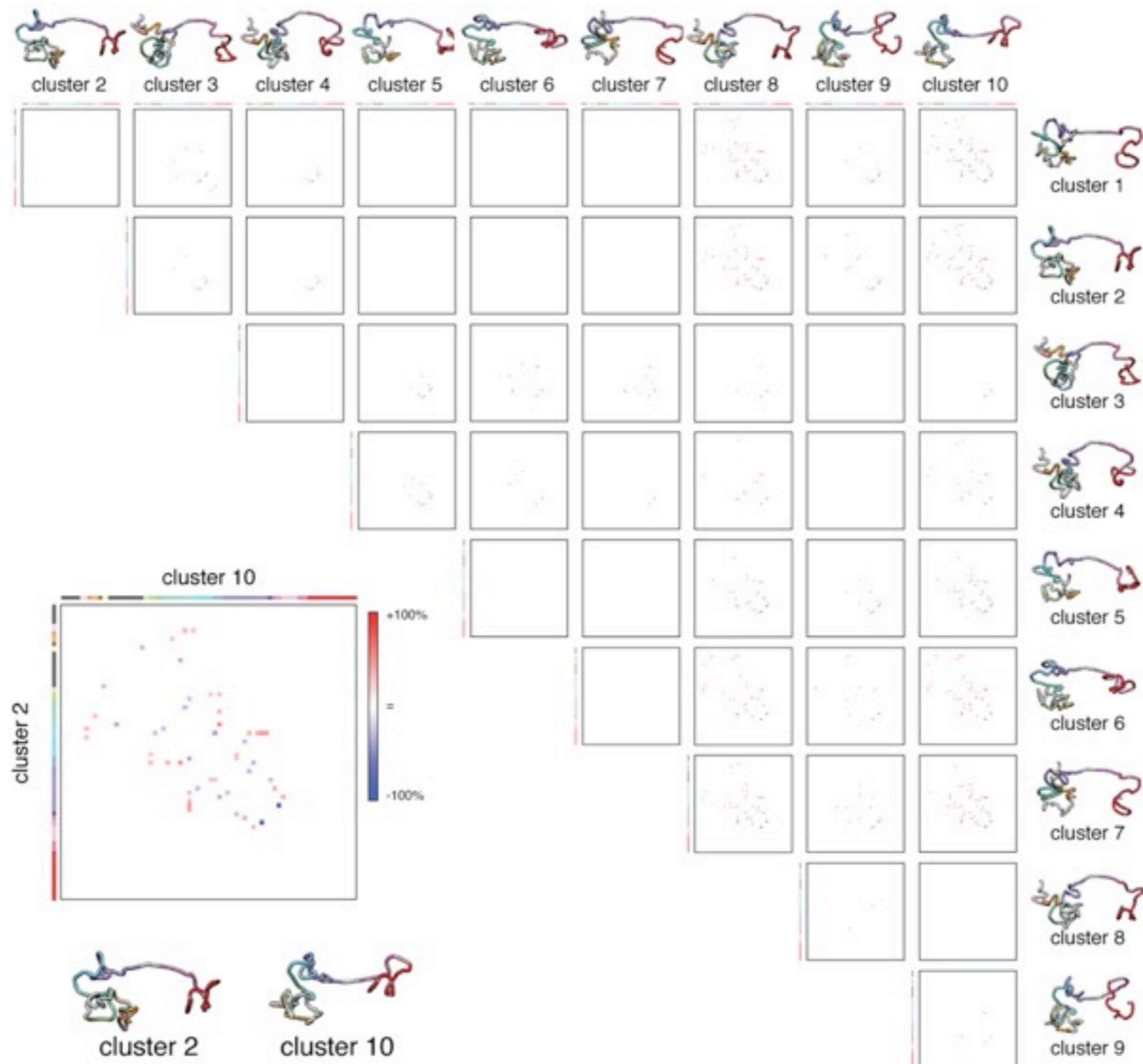
GM12878
Cluster #1
2780 model

K562
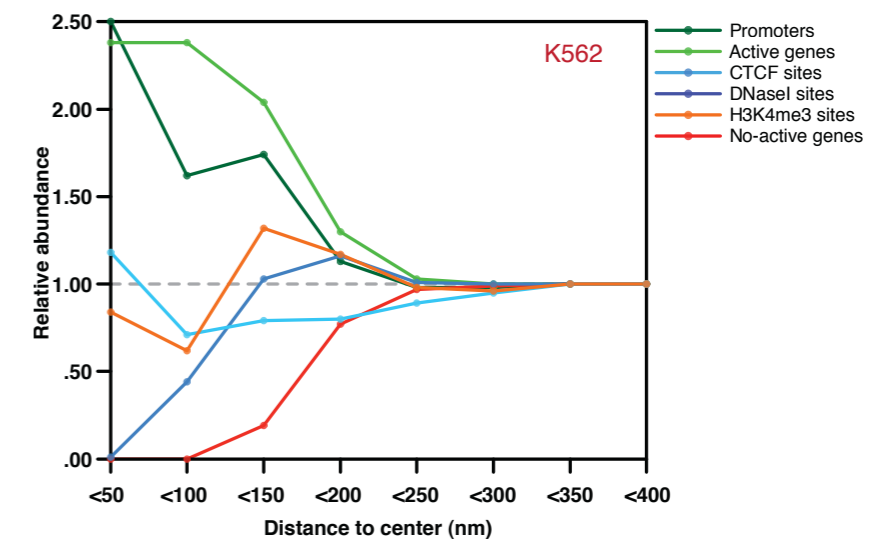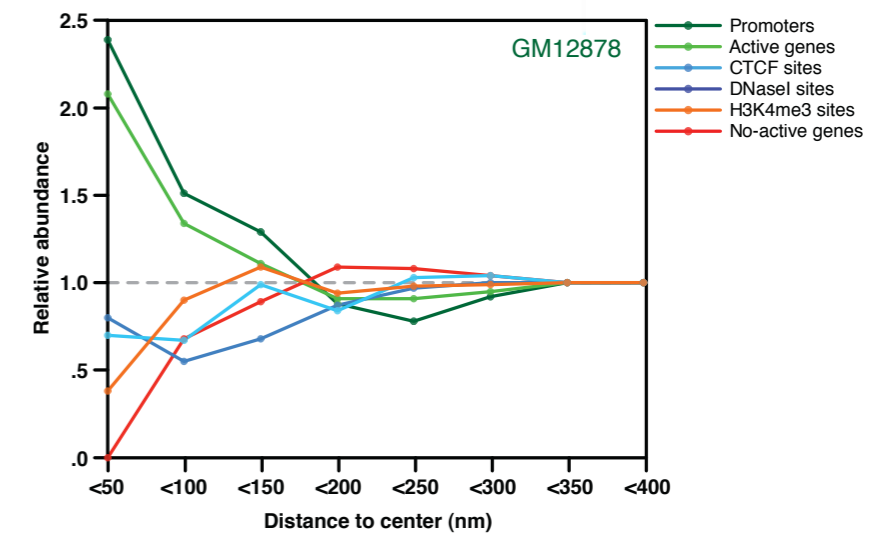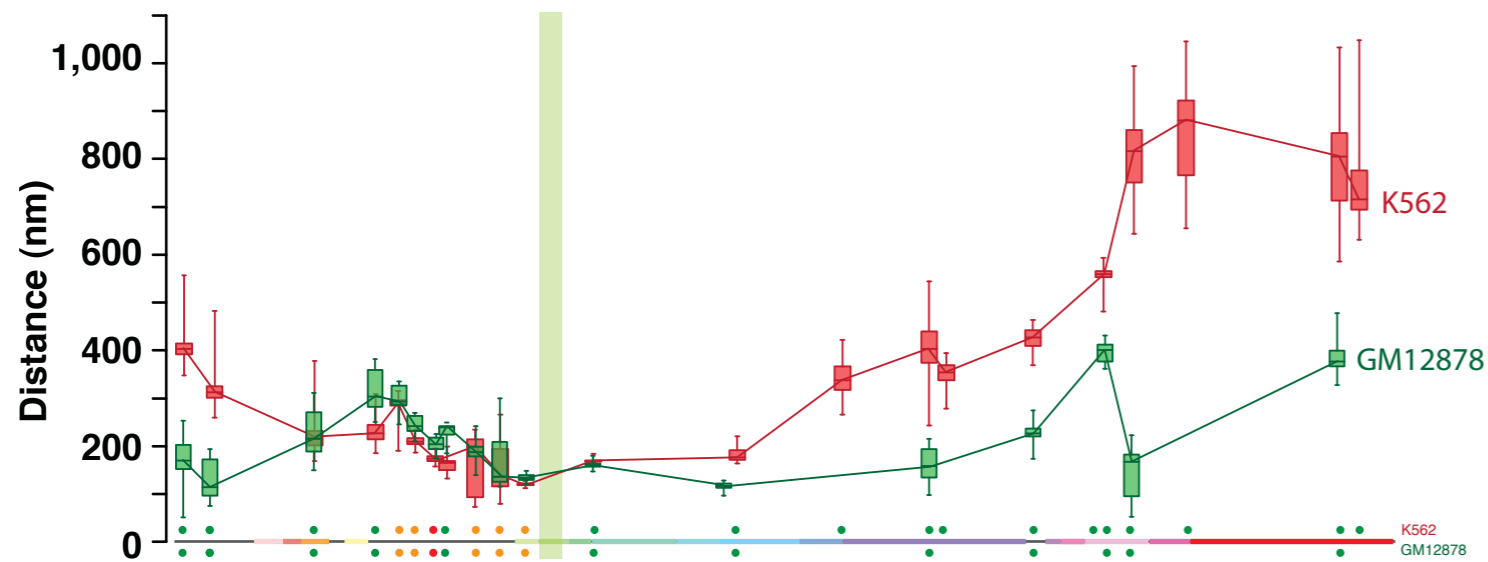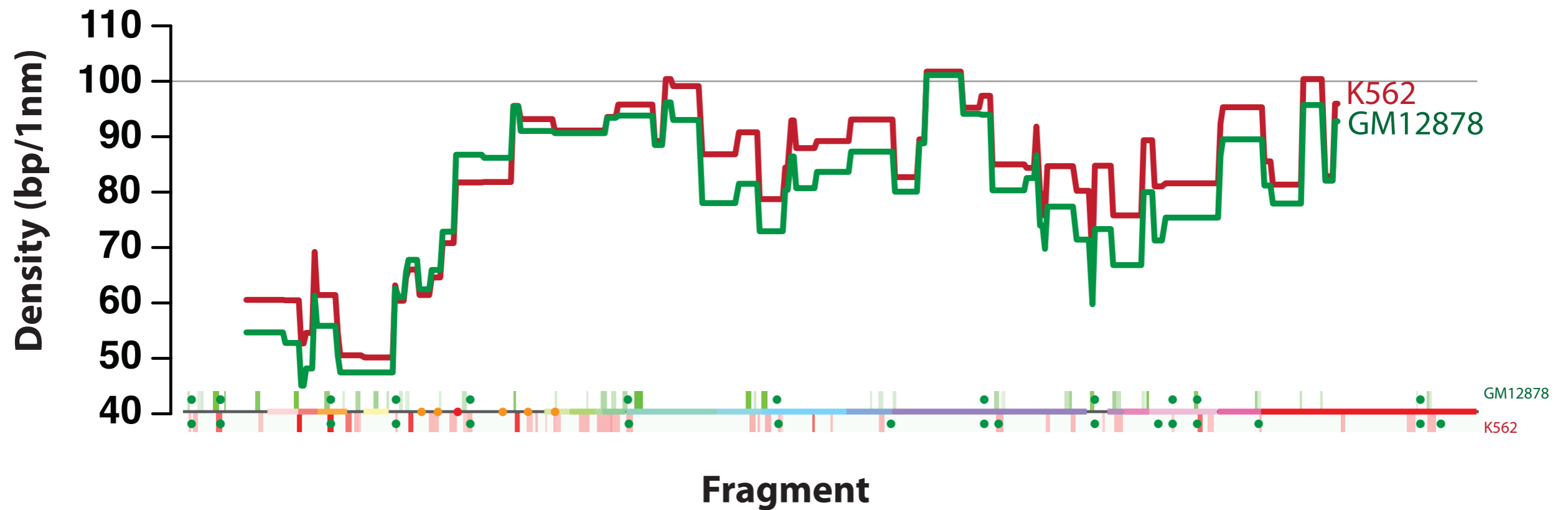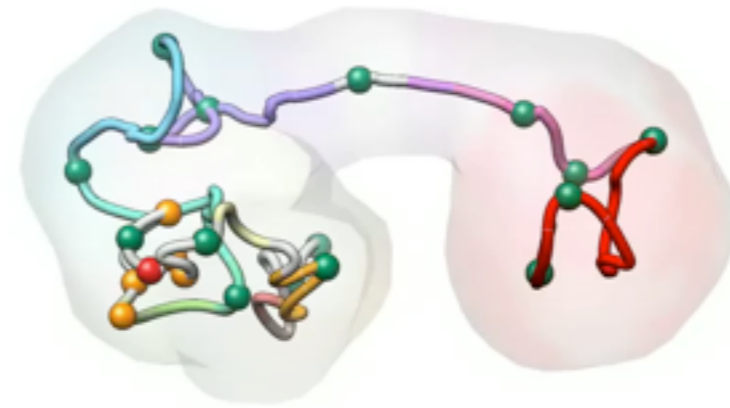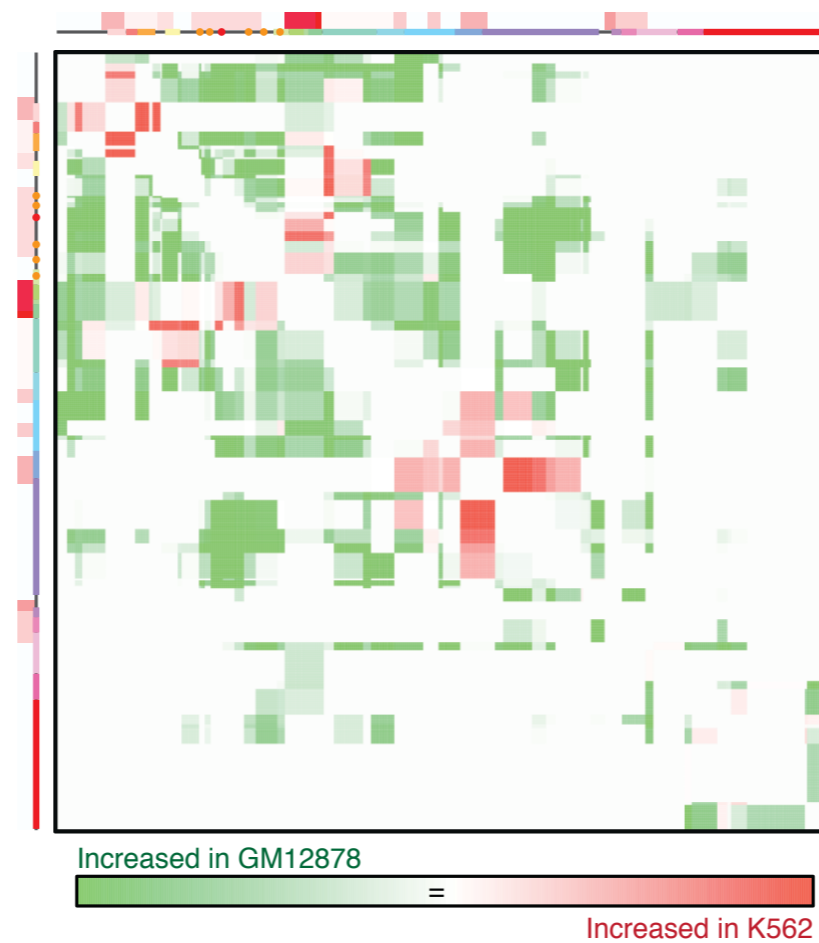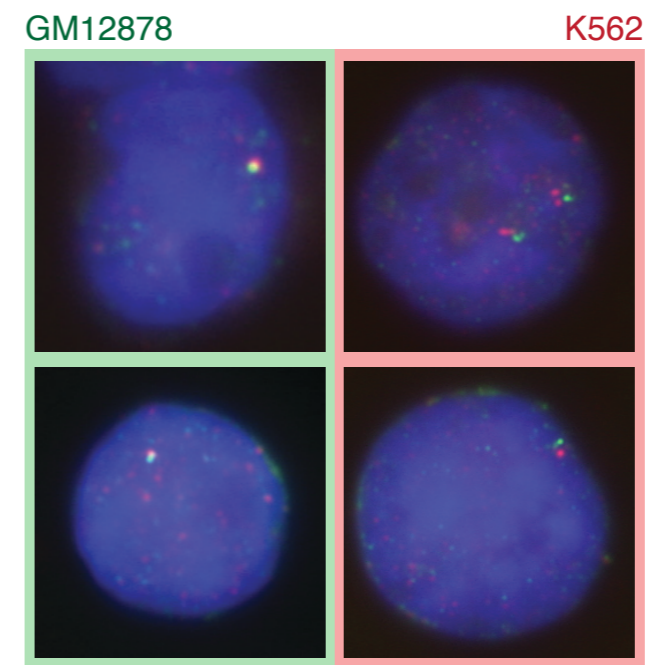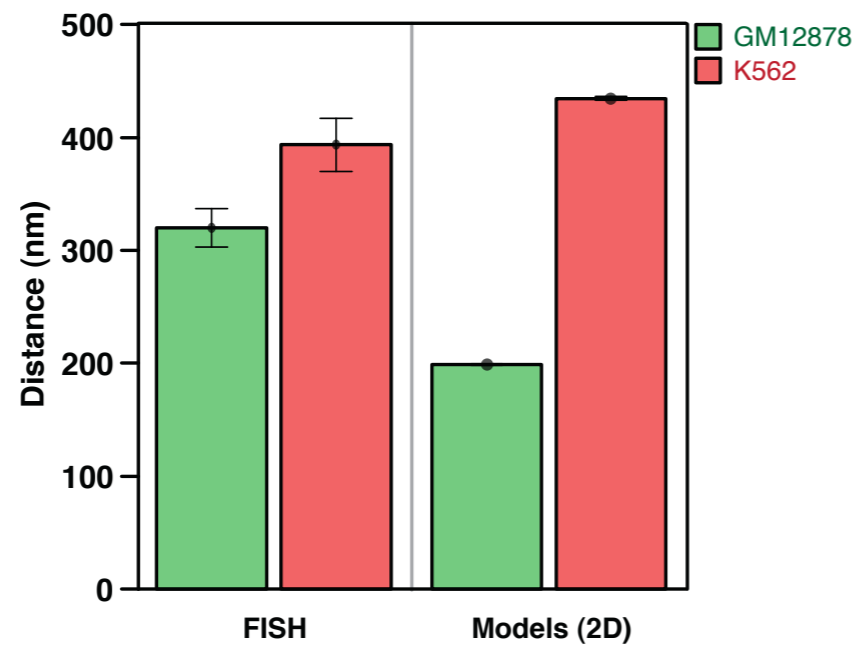Cluster #2
314 model

Increased in GM12878
=
Increased in K562

Wednesday, December 8, 2010

FISH validation
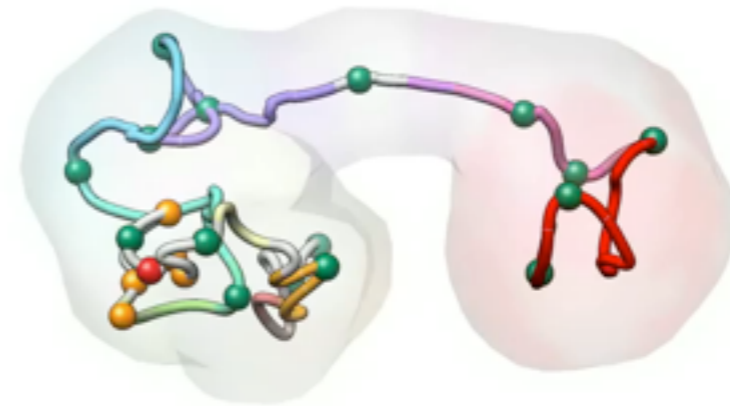
Summary

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

Summary

Models allow for 5C data de-convolution

Summary

Models allow for 5C data de-convolution

Summary

Selected models reproduce known (**and new**) interactions

Wednesday, December 8, 2010
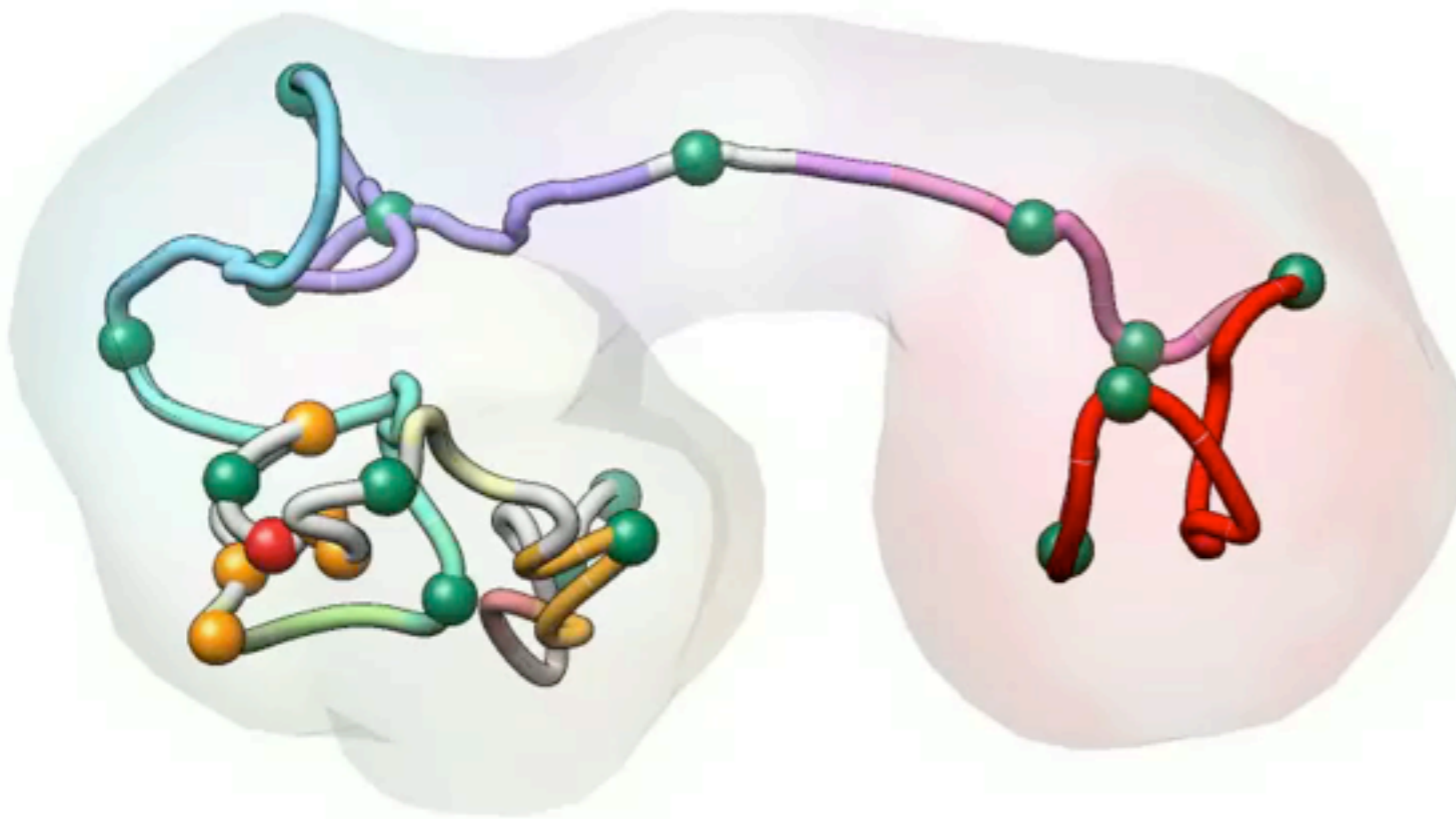
Summary
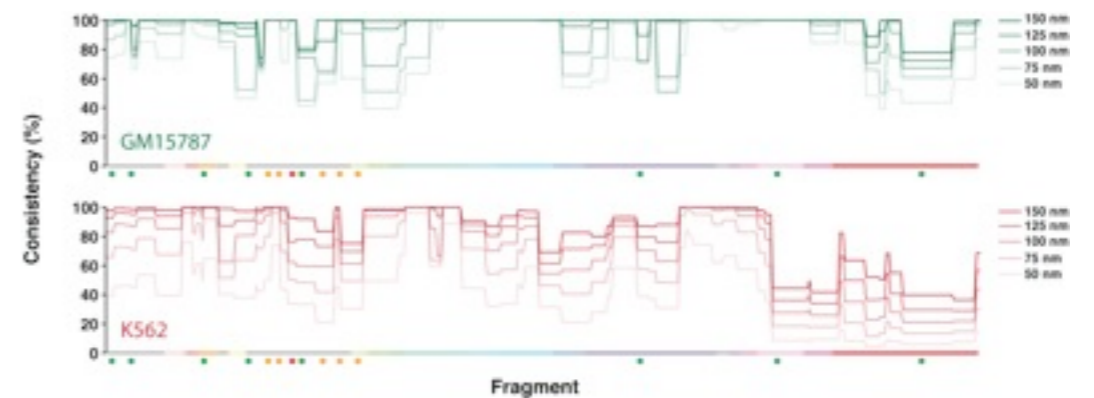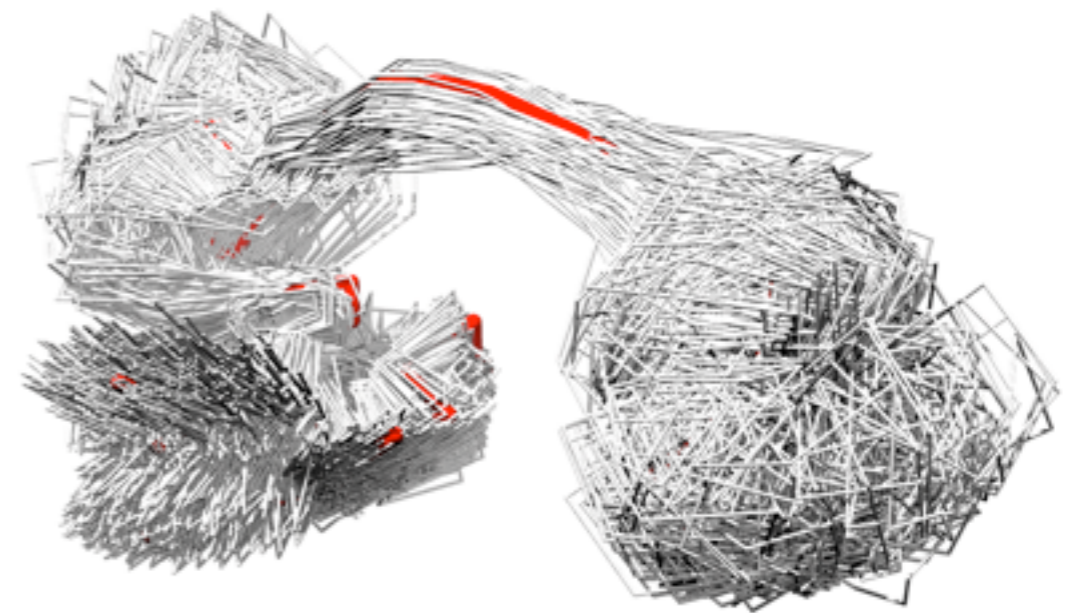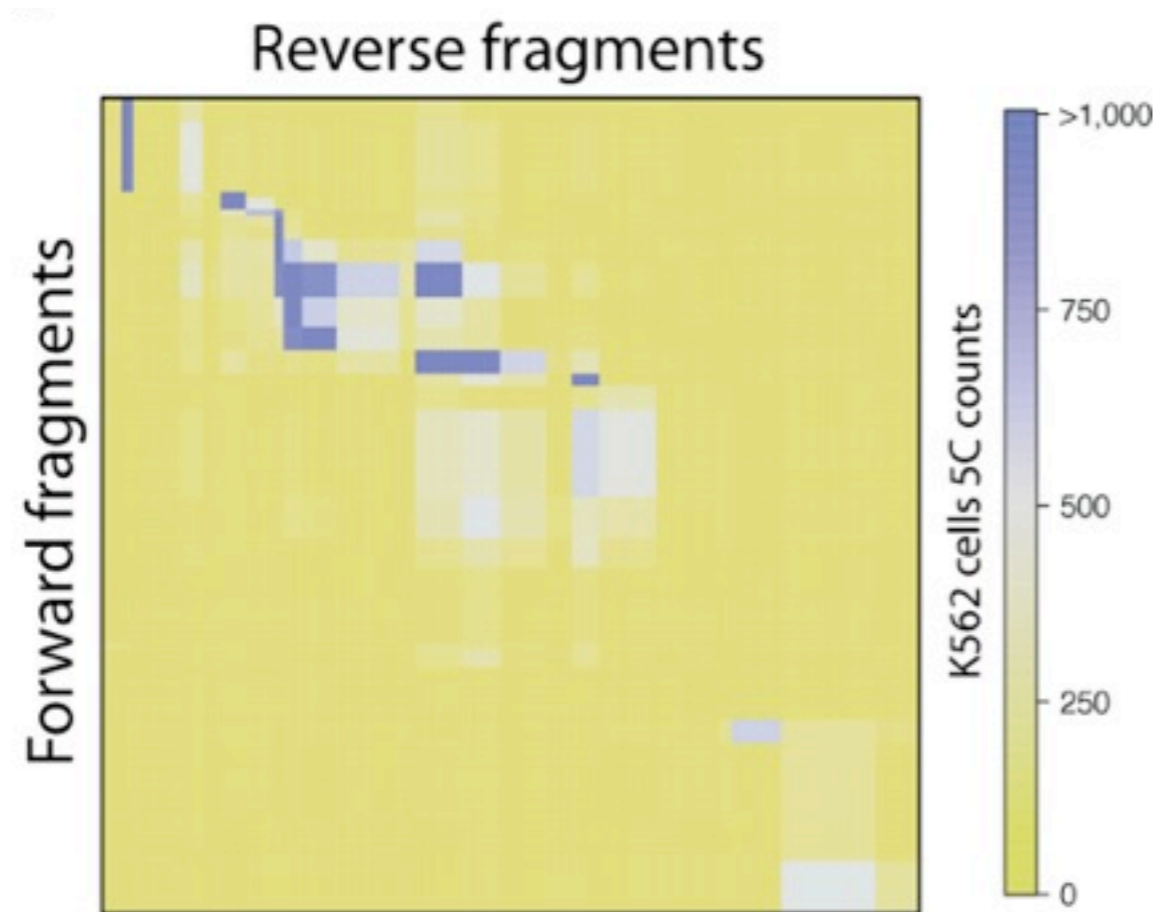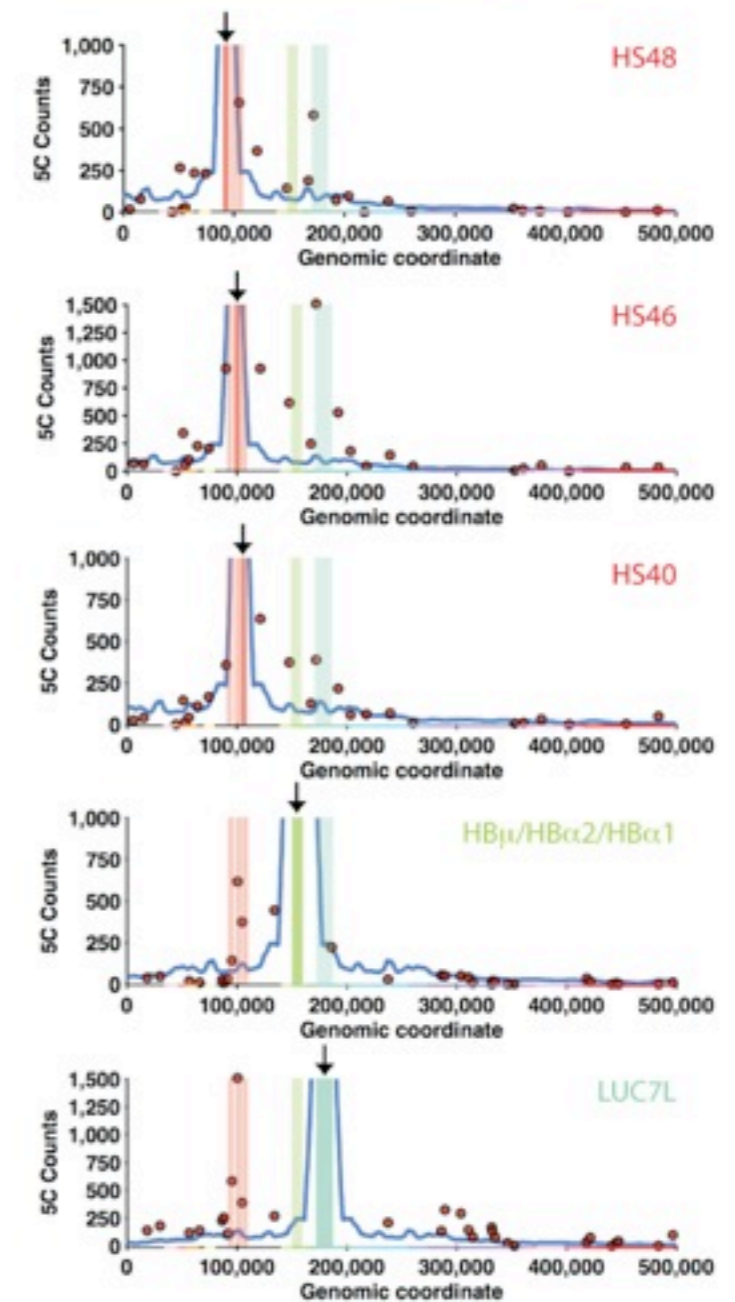
Large-scale changes in conformation correlate with gene expression of resident genes

α-globin-
Enhancer looping
only in K562

Looping interaction
In GM12878 cells

Increased in GM12878

Increased in K562

21
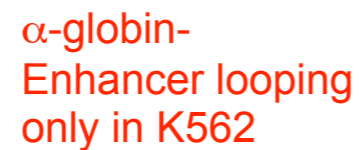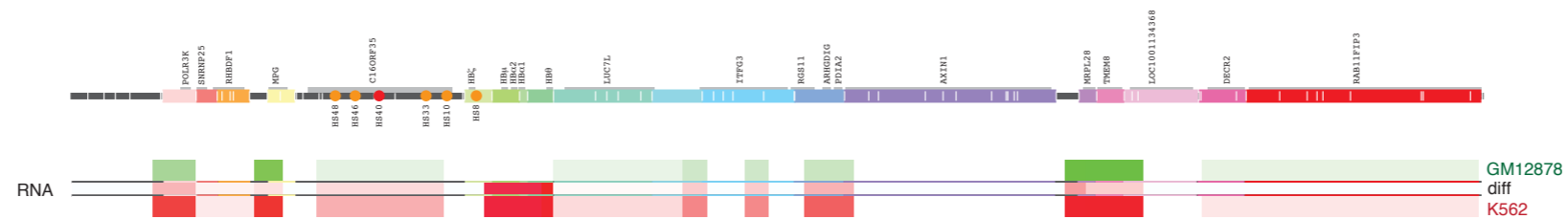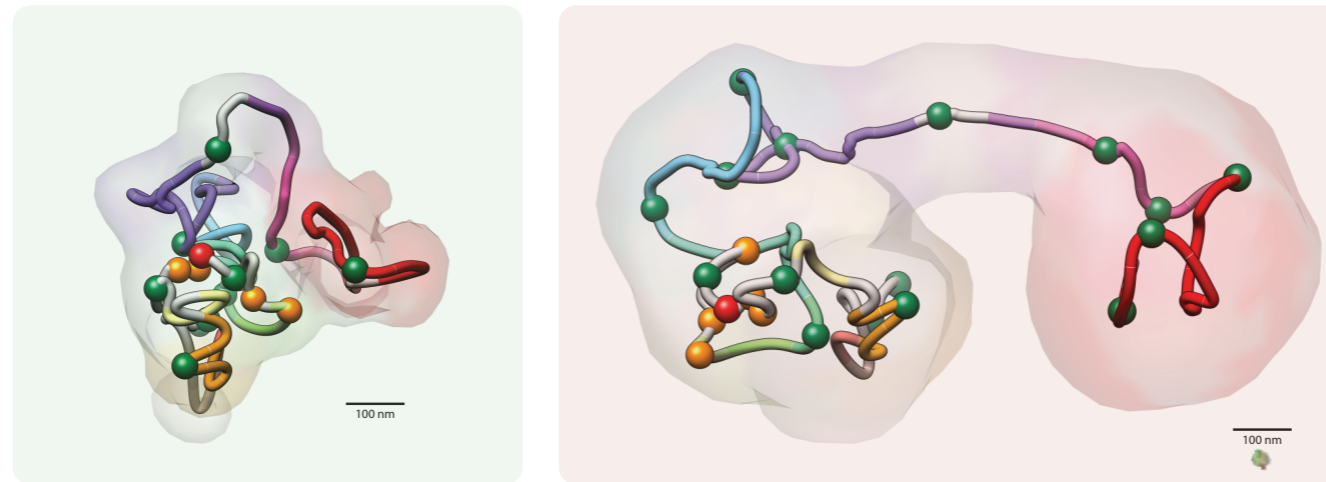
Summary

The models have been partially validated by FISH

GM12878   K562

Active genes
Inactive genes
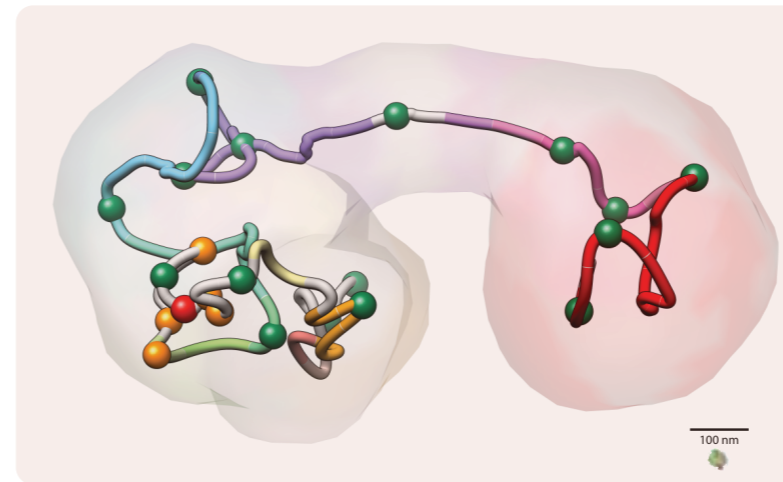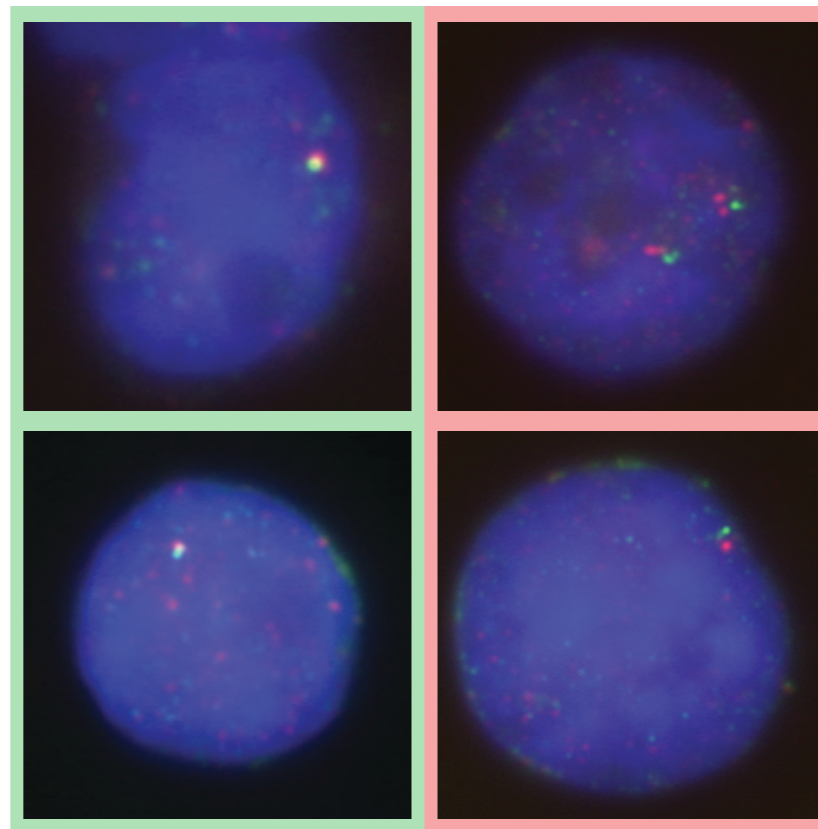CTCF sites
HS sites
Globule core

(a) chromatin — subcompartment — subcompartment — loop base spring (magnified)
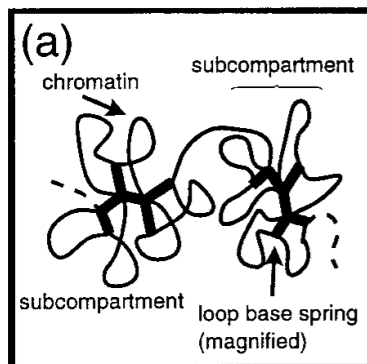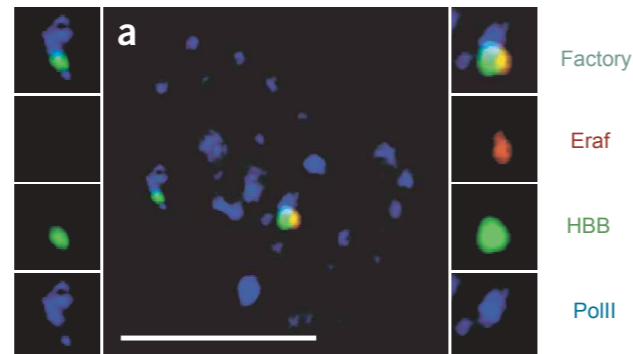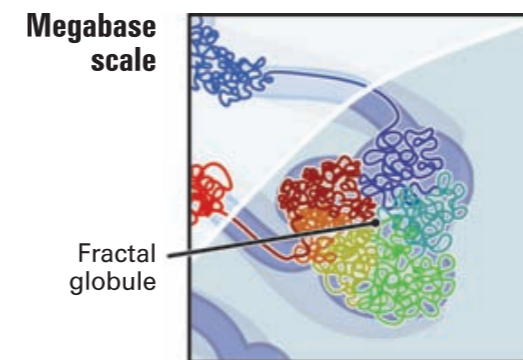
Münkel et al. JMB (1999)

Osborne et al. Nat Genet (2004)

al. Science (2009)

K

L

c-myc — IgH — CTCF
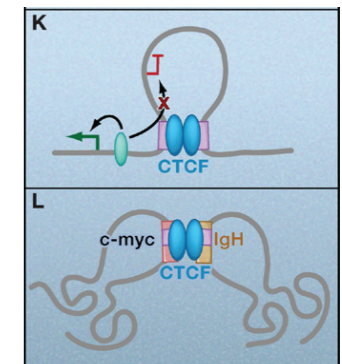
Phillips and Corces. Cell (2009)

Acknowledgments

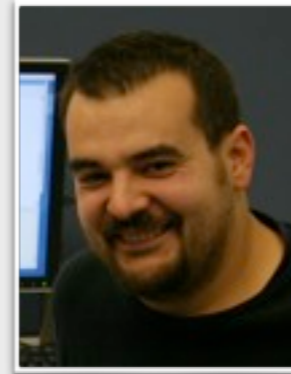**Davide Baù**
Postdoctoral fellow

Amartya Sanyal
Postdoctoral Fellow

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

Marc A. Marti-Renom
Structural Genomics Unit
Bioinformatics and Genomics Department
Centro de Investigación Príncipe Felipe
Valencia, Spain

University of Massachusetts Medical School
UMASS

PRINCIPE FELIPE
CENTRO DE INVESTIGACION

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