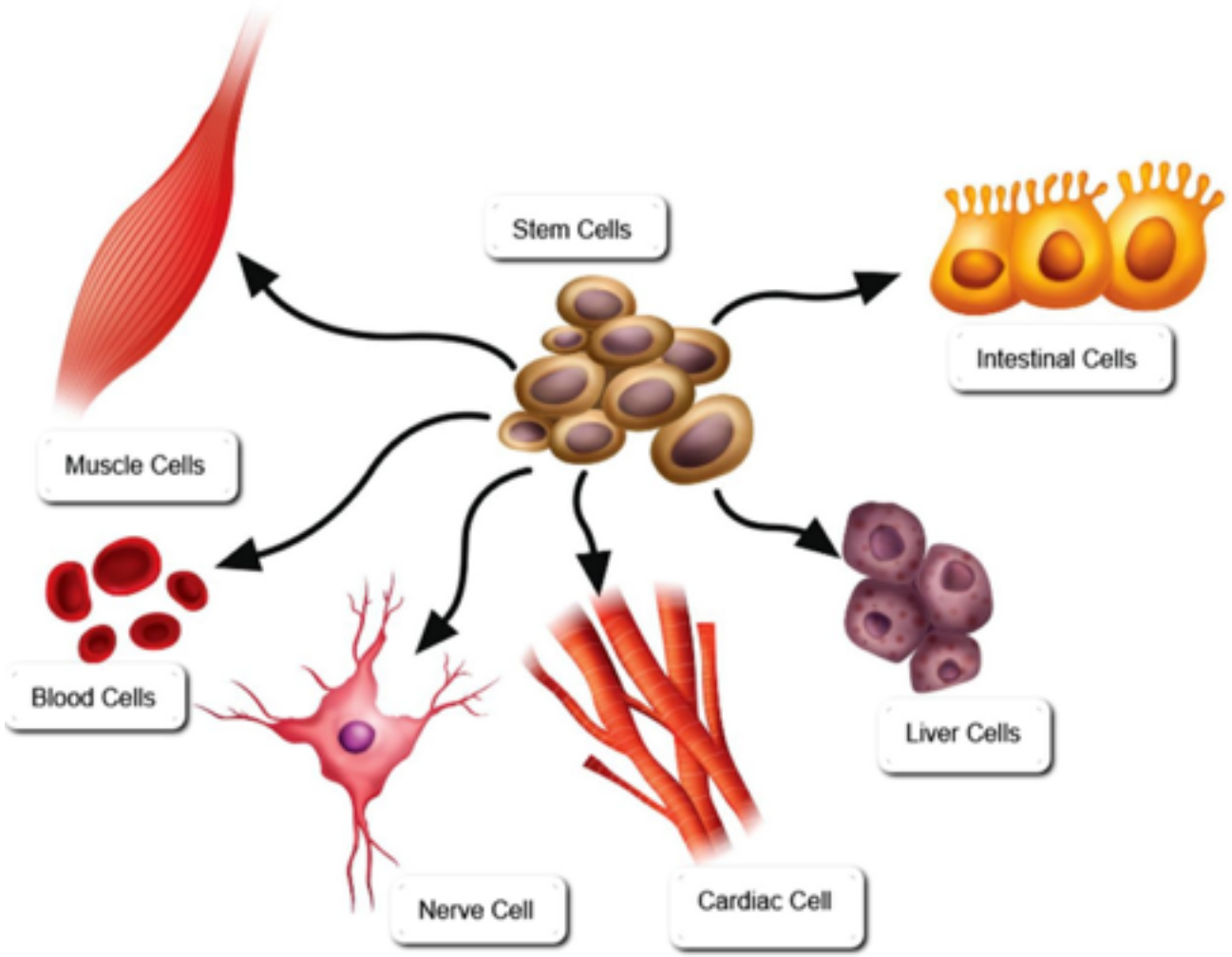
3D genome arcitecture

Brad Krajina
Alex Yoshikawa

The Human Genome Encodes for Rich Phenotypic Diversity

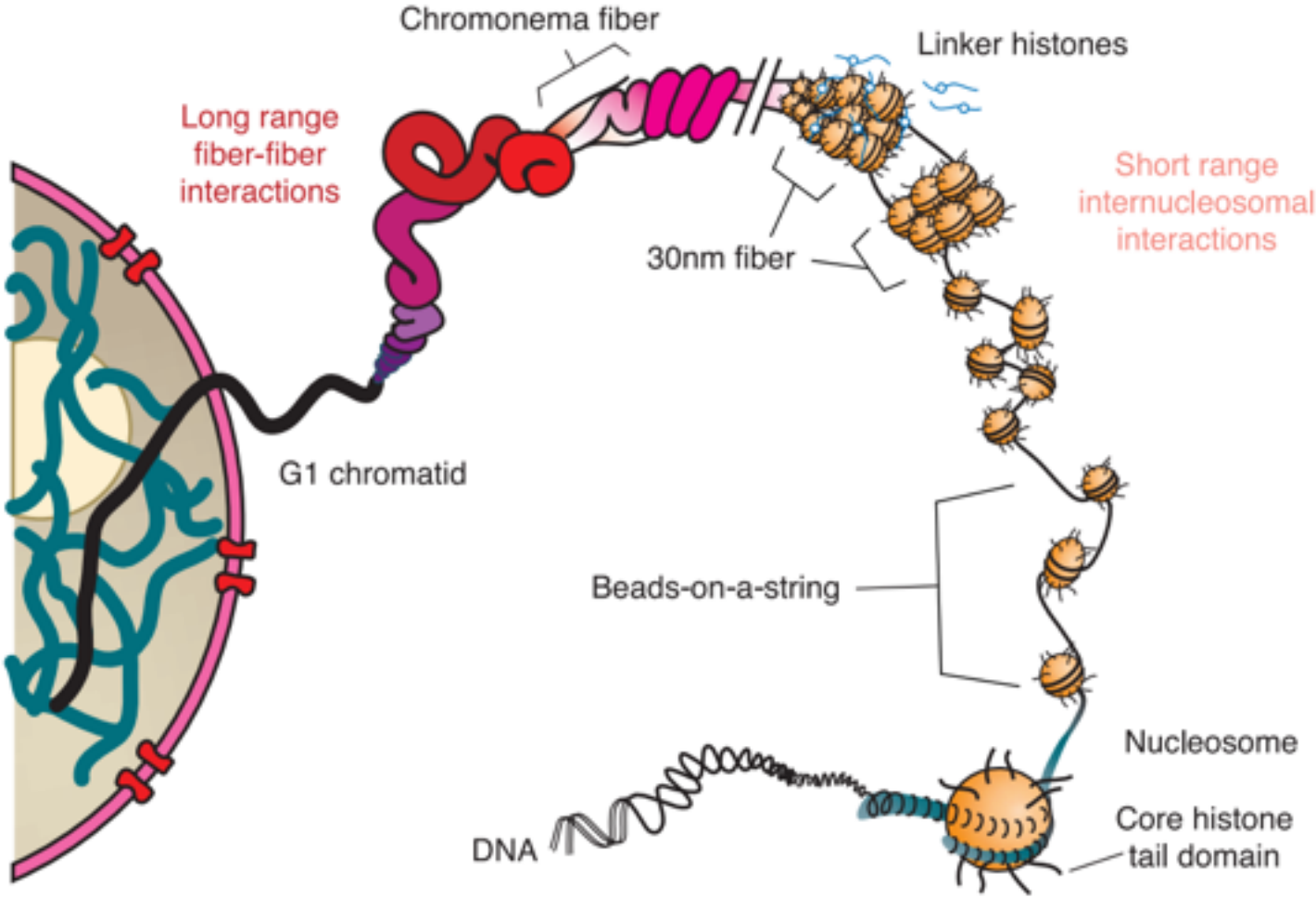


Stephen J. Farenga, Daniel Ness, Michael Hutchinson. *The American Biology Teacher* 77(6). 2015

Unlocking how gene expression is regulated to give rise to a variety of phenotypes from a single genotype may play a central role in understanding human development and disease

Beyond the Central Dogma: 3D Chromosome Organization Regulates Genetic Processes

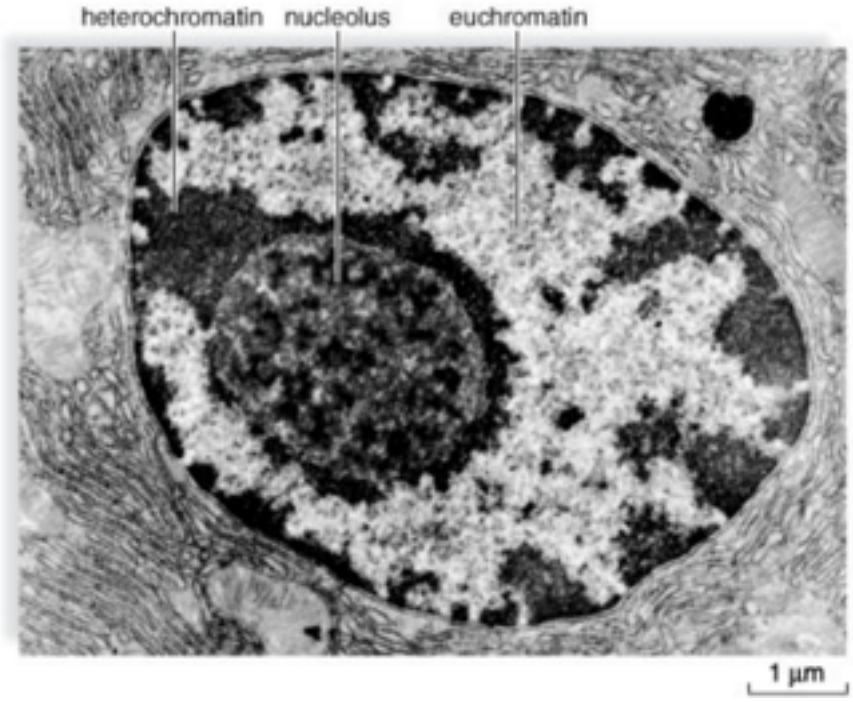
A human cell must compact ~1 m of DNA into a single nucleus, requiring exquisite control of hierarchical 3D organization



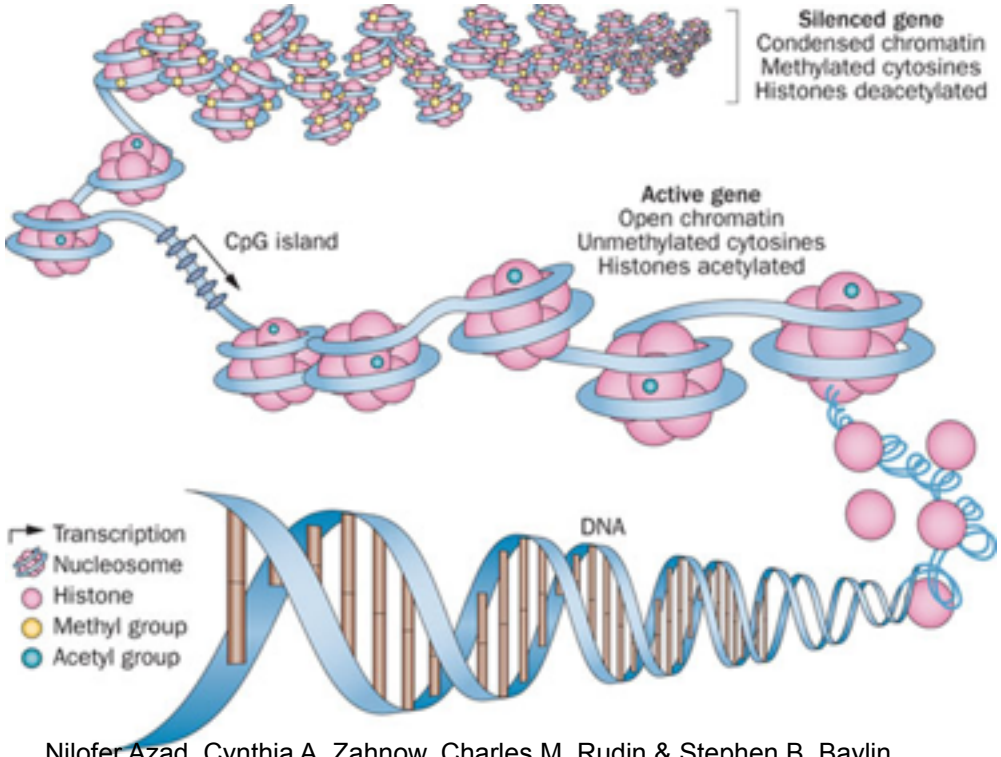
Peter J. Horn, Craig L. Peterson. Science 297(5588),1824. .2002

Eukaryotic Chromosomes are Organized into Transcriptionally Active and Transcriptionally Inactive Domains

Heterochromatin is densely packaged and **transcriptionally silenced**.
Euchromatin is loosely packaged and **transcriptionally active.**



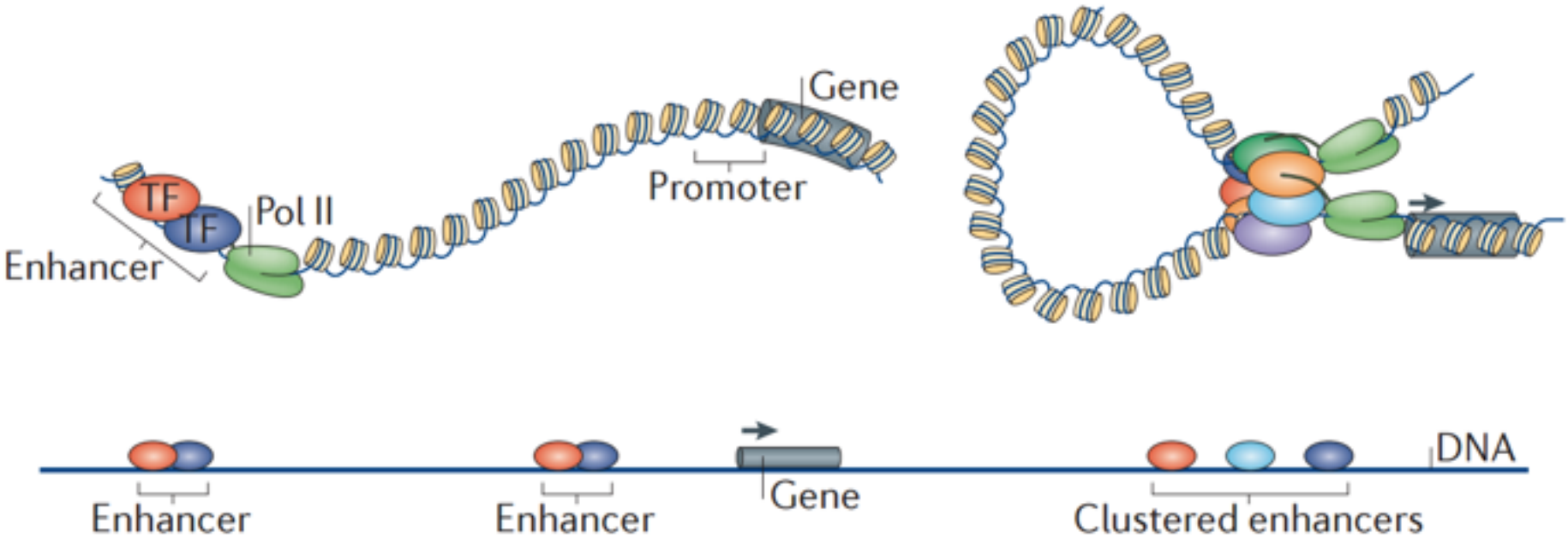
http://classes.midlandstech.edu/carterp/Courses/bio101/chap15/chap15.htm



Nilofer Azad, Cynthia A. Zahnow, Charles M. Rudin & Stephen B. Baylin
Nature Reviews Clinical Oncology 10, 256-266. 2013

DNA looping regulates gene expression

Chromatin looping enables transcriptional activation by contacts between gene promoters and genomically distal enhancer sequences



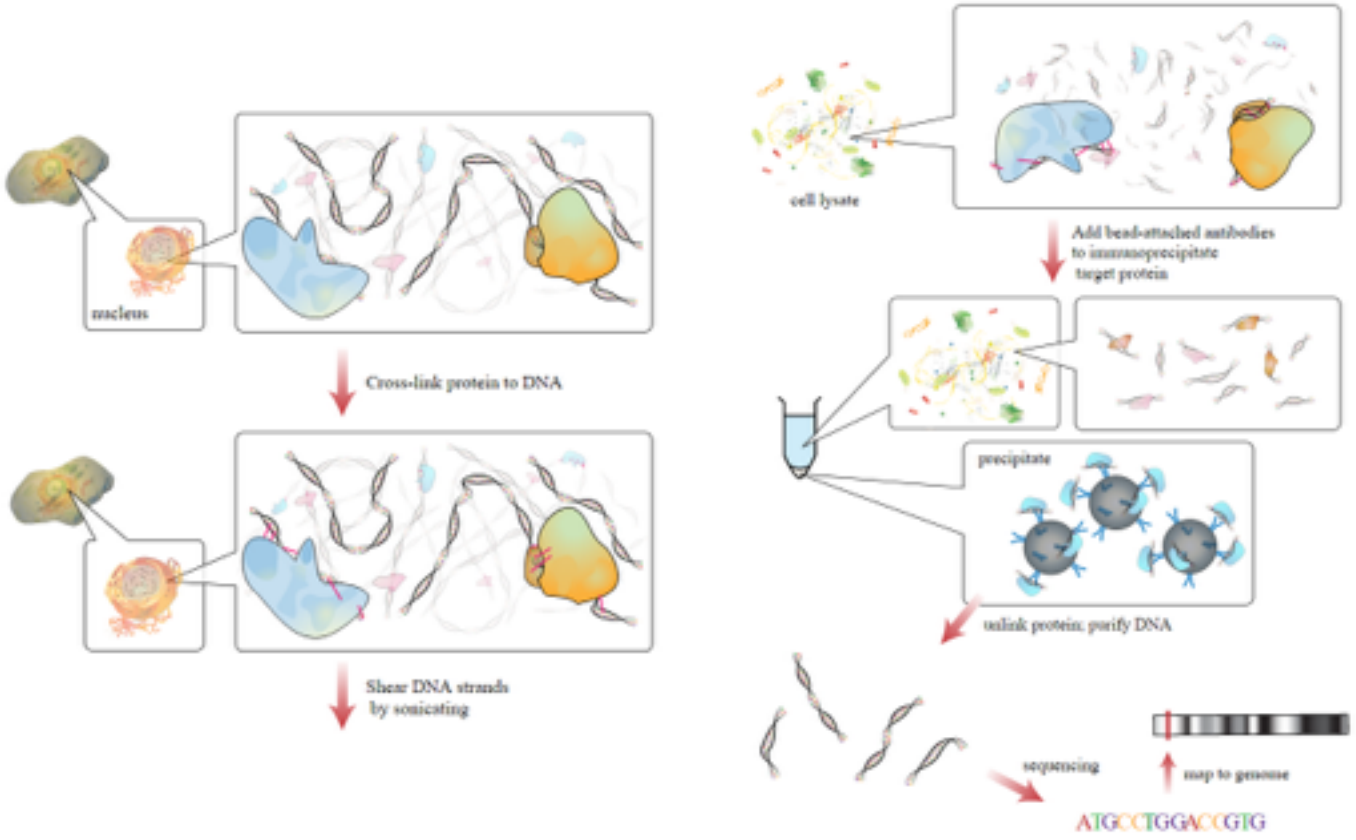
Ana Pombo and Niall Dillon. *Nature Reviews Molecular Cell Biology* 16, 245. 2015

Proteins regulate genome architecture

[Restructuring Chromosomes Youtube Link]

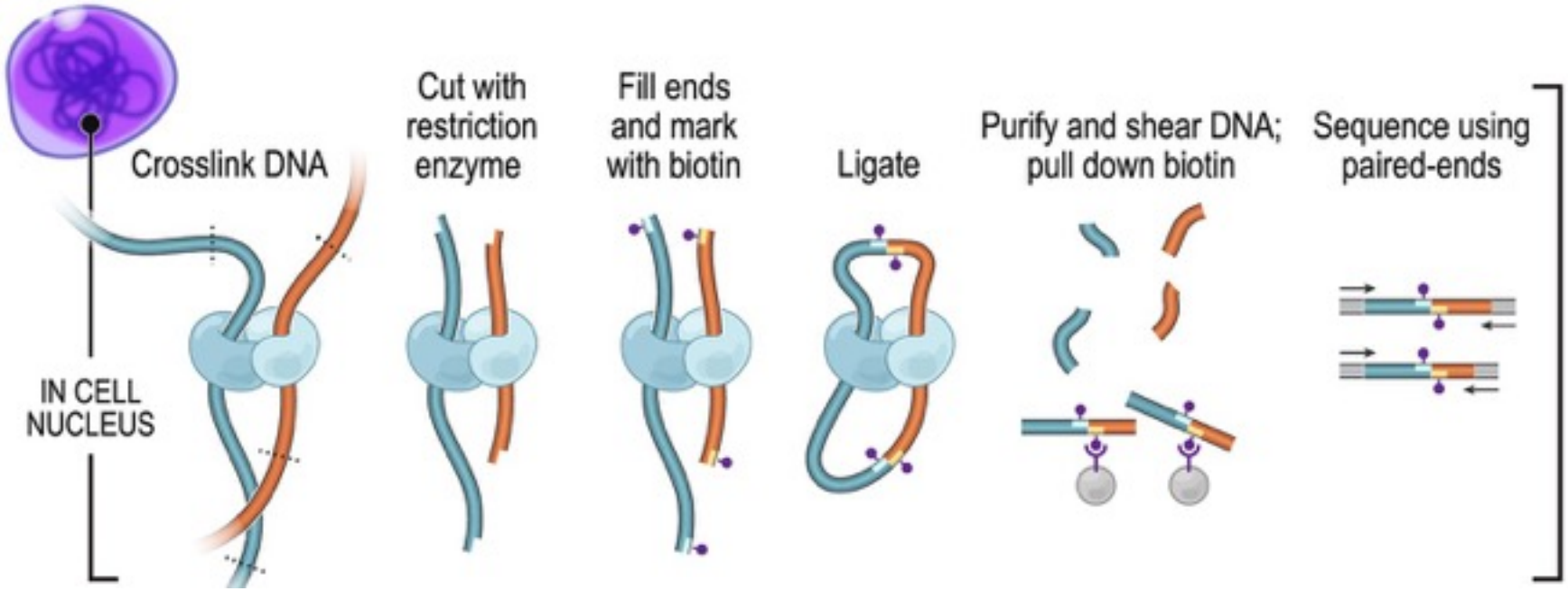
Figure is adapted from Luong, P. Basic Principles of Genetics, Connexions Web site (2009)

Chip-Seq is a method to study protein-DNA interactions



ChIP-sequencing. (2017, February 19). In *Wikipedia, The Free Encyclopedia*.

Hi-C is a method to study 3D genome architecture



Suhas Rao et al, Nature Review Mol. Bio. (2009)

Article

Cell

A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping

Suhas S.P. Rao,[1,2,3,4,10] Miriam H. Huntley,[1,2,3,4,5,10] Neva C. Durand,[1,2,3,4] Elena K. Stamenova,[1,2,3,4] Ivan D. Bochkov,[1,2,3] James T. Robinson,[1,4] Adrian L. Sanborn,[1,2,3,6] Ido Machol,[1,2,3] Arina D. Omer,[1,2,3] Eric S. Lander,[4,7,8,*] and Erez Lieberman Aiden[1,2,3,4,9,*]

[1]The Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030, USA
[2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
[3]Department of Computer Science, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005, USA
[4]Broad Institute of MIT and Harvard, Cambridge, MA 02139, USA
[5]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA
[6]Department of Computer Science, Stanford University, Stanford, CA 94305, USA
[7]Department of Biology, Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA
[8]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[9]Center for Theoretical Biological Physics, Rice University, Houston, TX 77030, USA
[10]Co-first author

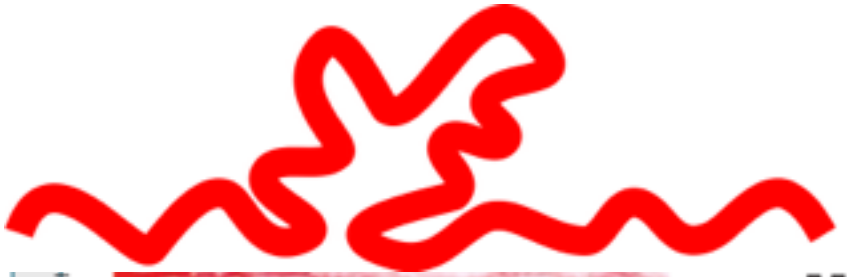
Presented by : Brad Krajina

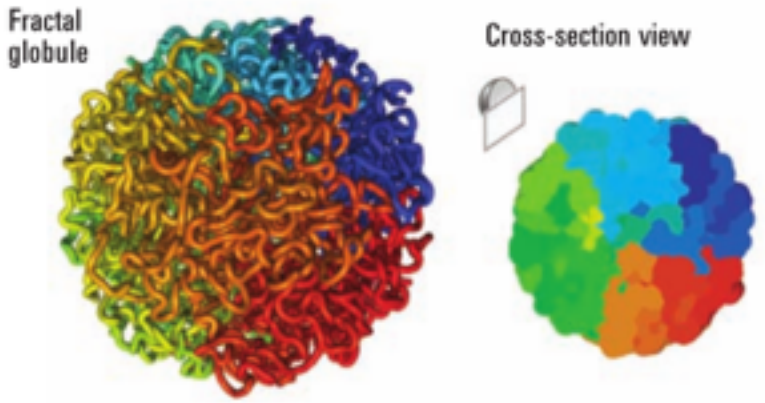
HiC reveals organization of human chromsomes

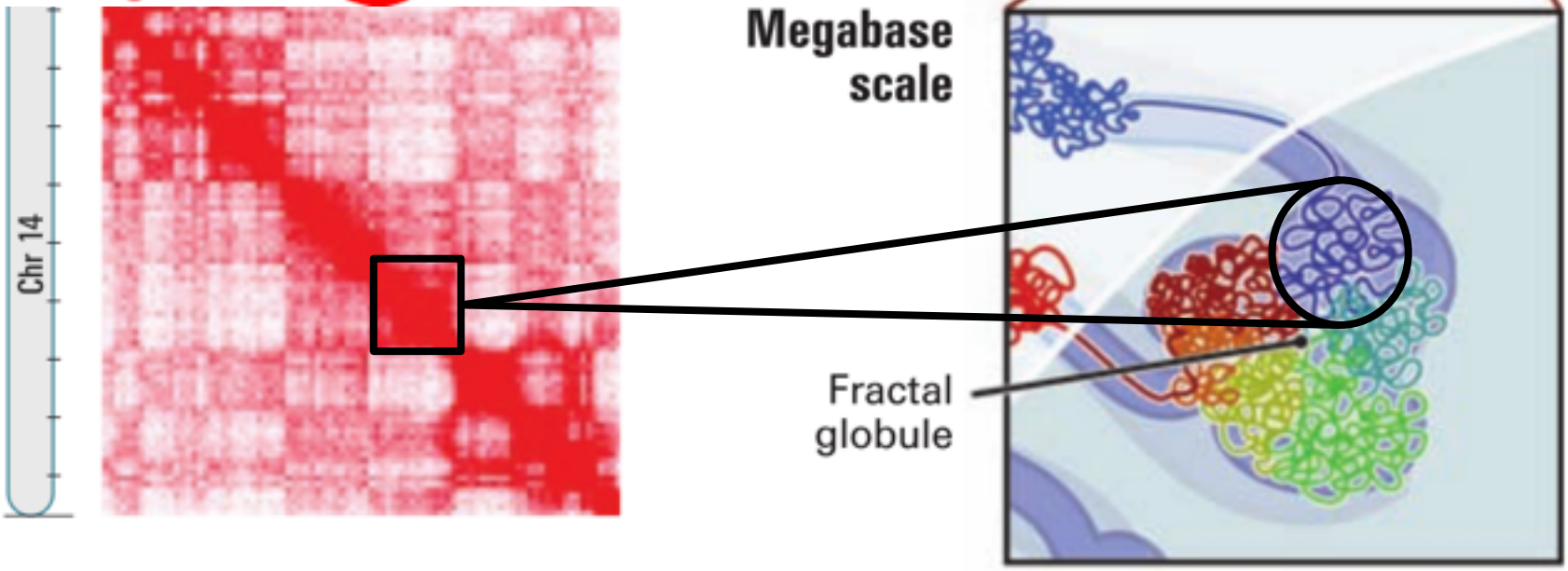
Contact probability maps from HiC reveal regions of the chromosome that preferentially interact

"Unfolded" genome

"Folded" genome with contact domains

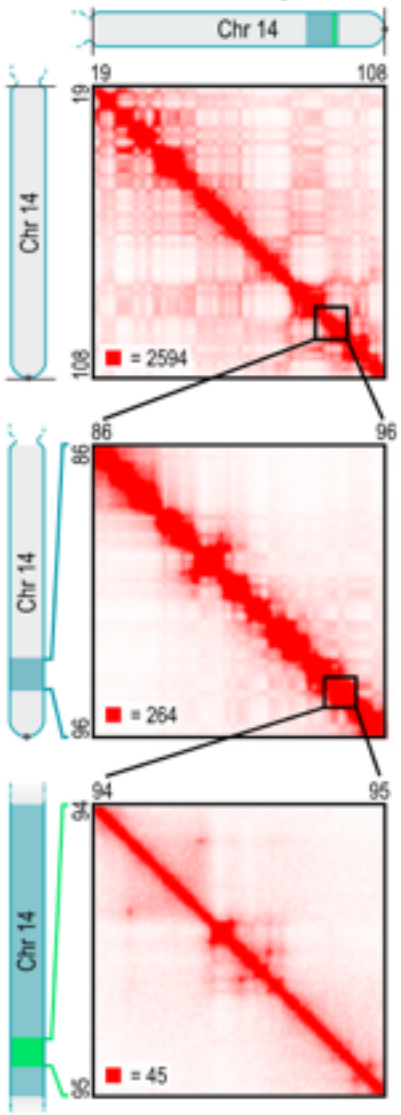




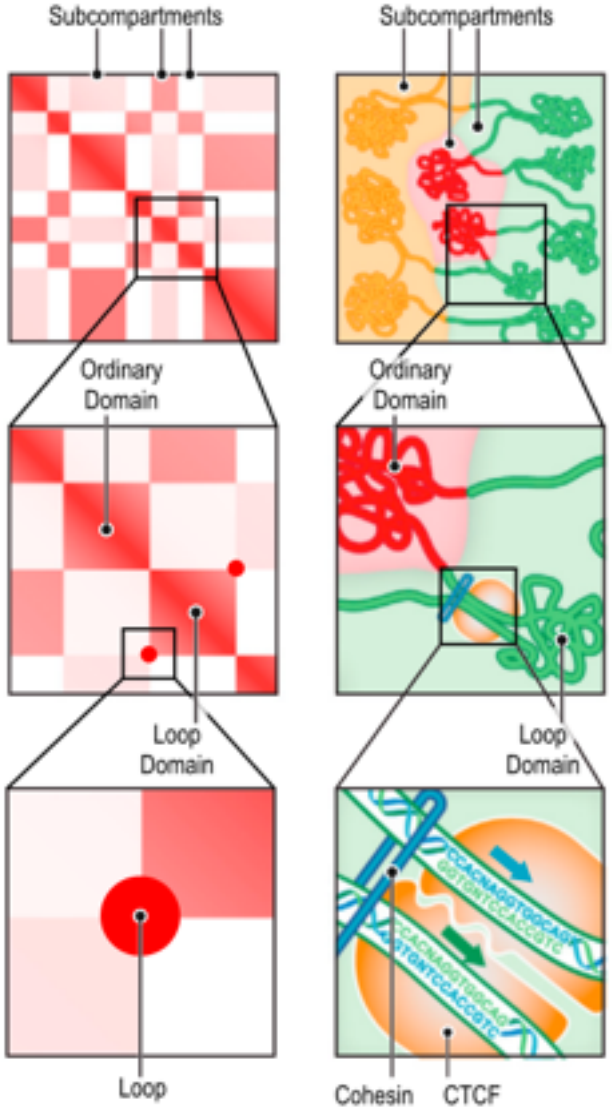


In situ Hi-C reveals hierarchical organization and function of the human genome

In situ Hi-C enables kilobase-scale resolution of contact maps



Kilobase-scale resolution elucidates hierarchical genome organization



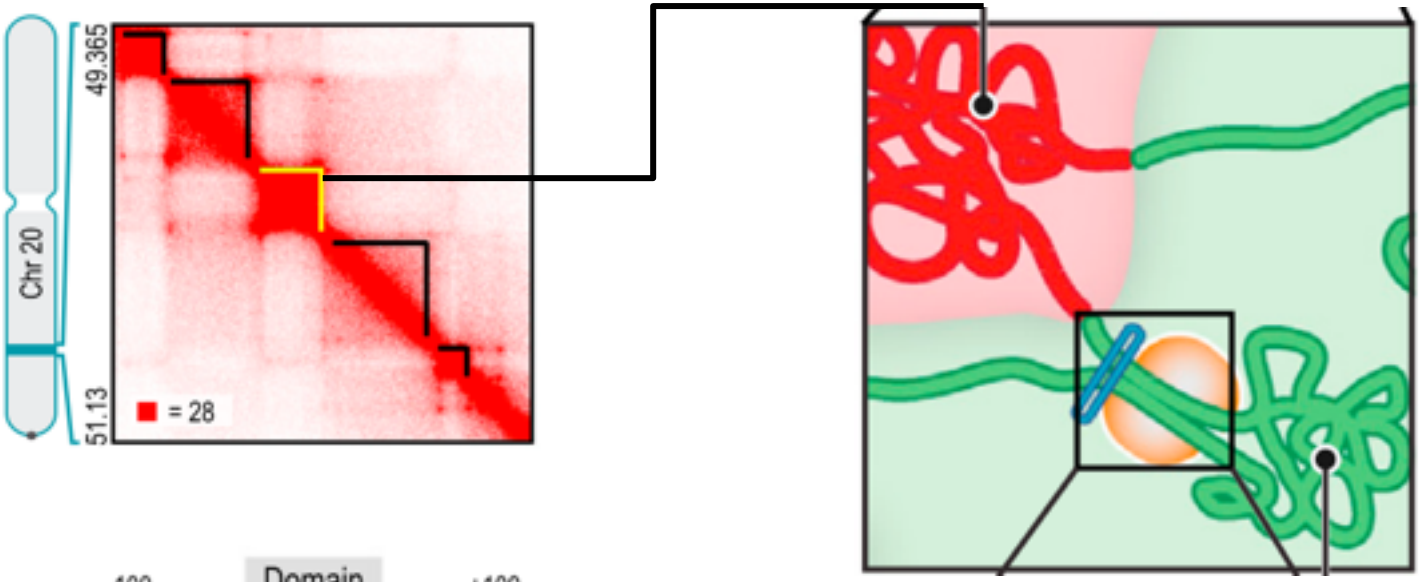
Sub-compartments organize distinct Epigenetic patterns

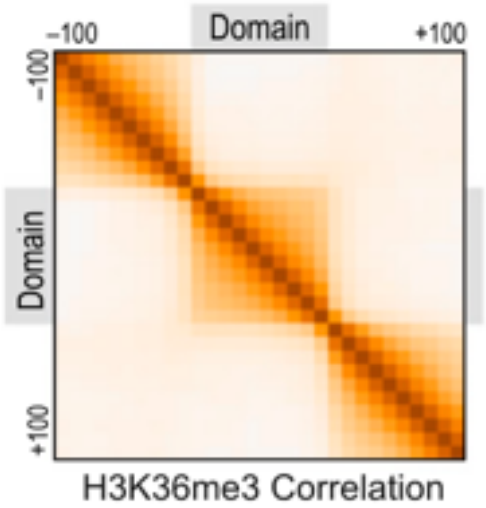
Megabase domains organize preferential regions of self-contact

Protein-mediated loops demarcate domain boundaries and regulate gene activation

Contact domains segregate the chromosome into contact domains with distinct epigenetic profiles

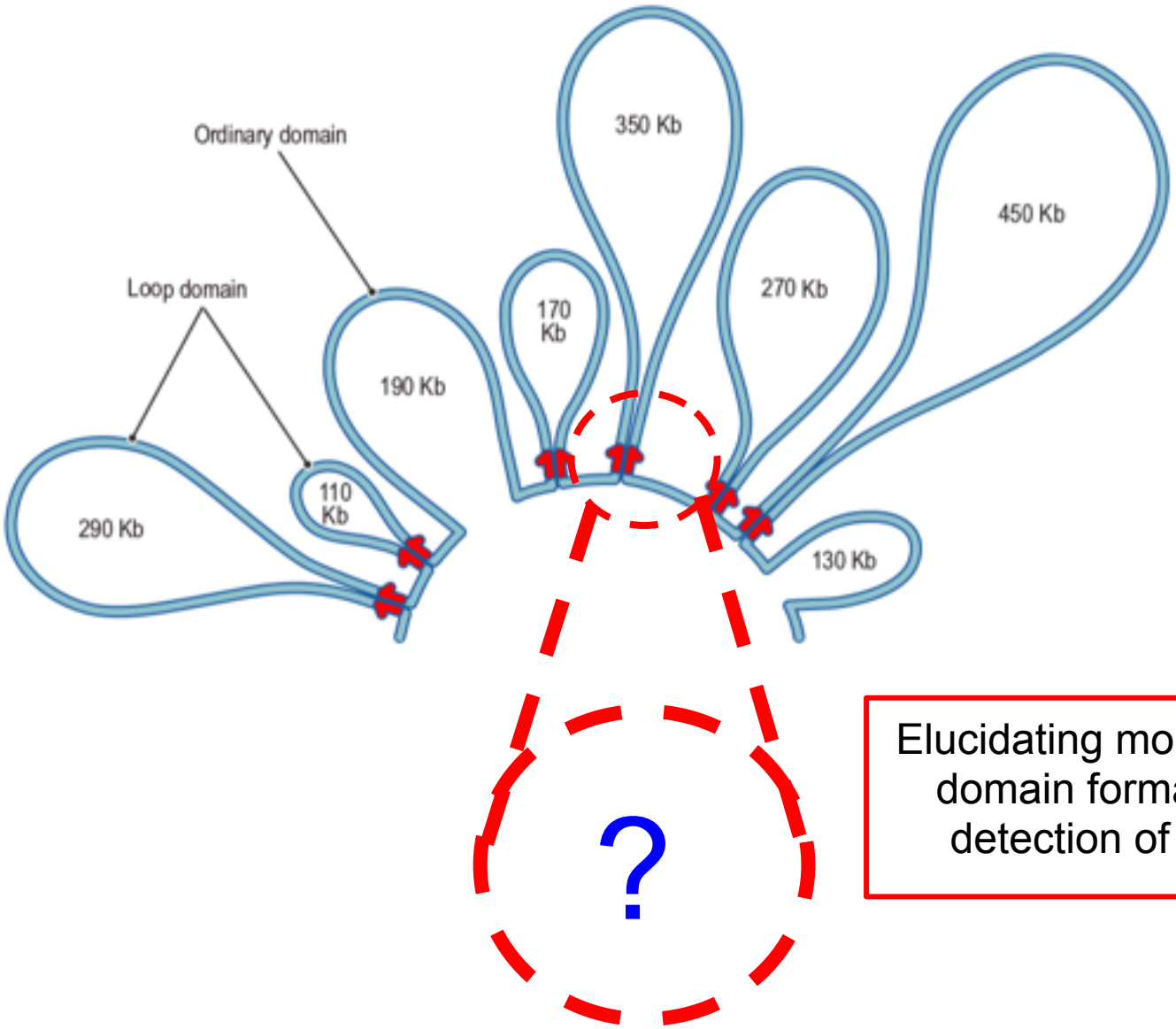
Genome contact maps reveal domains with preferential self-contacts





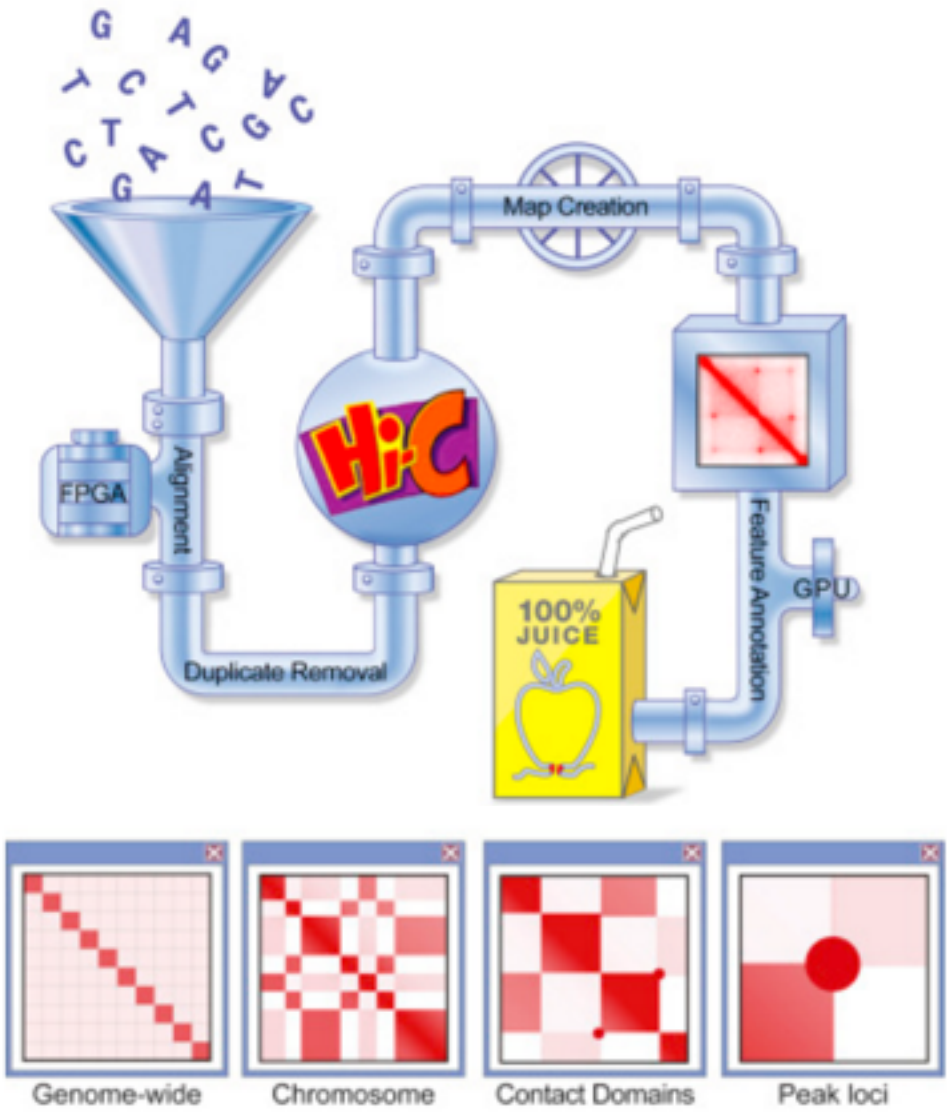
Contact domains exhibit correlations in histone modifications

What are the “molecular ties” that define domains?



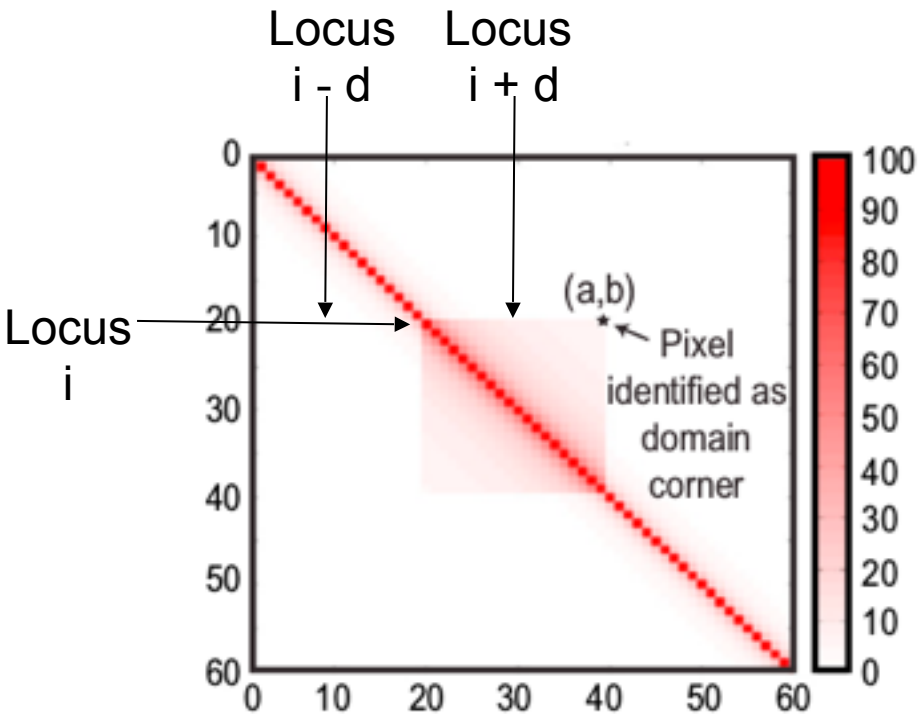
Elucidating molecular mechanisms of domain formation requires robust detection of domain boundaries

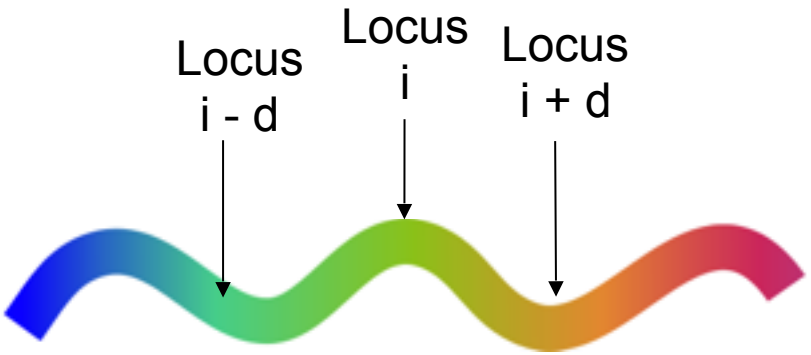
“Juicer” and “Juicebox” provide an automated pipeline for analyzing HiC sequencing data

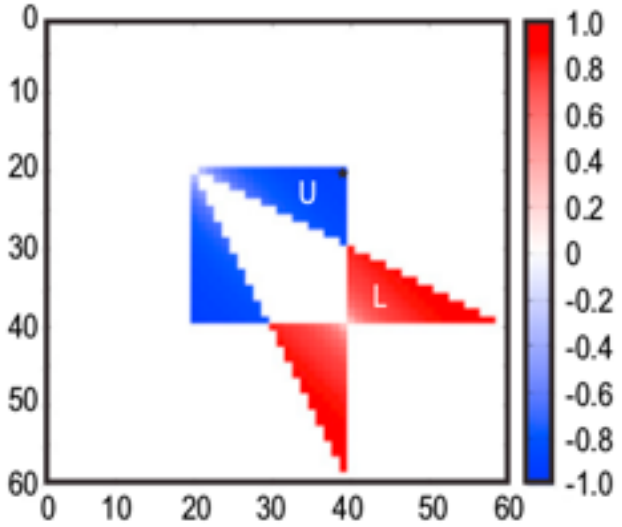


An “arrowhead” matrix algorithm enables more robust detection of contact domain boundaries

The arrowhead matrix algorithm identifies relative enrichment or depletion of contacts with loci an equal distance (d) upstream and downstream of a reference locus



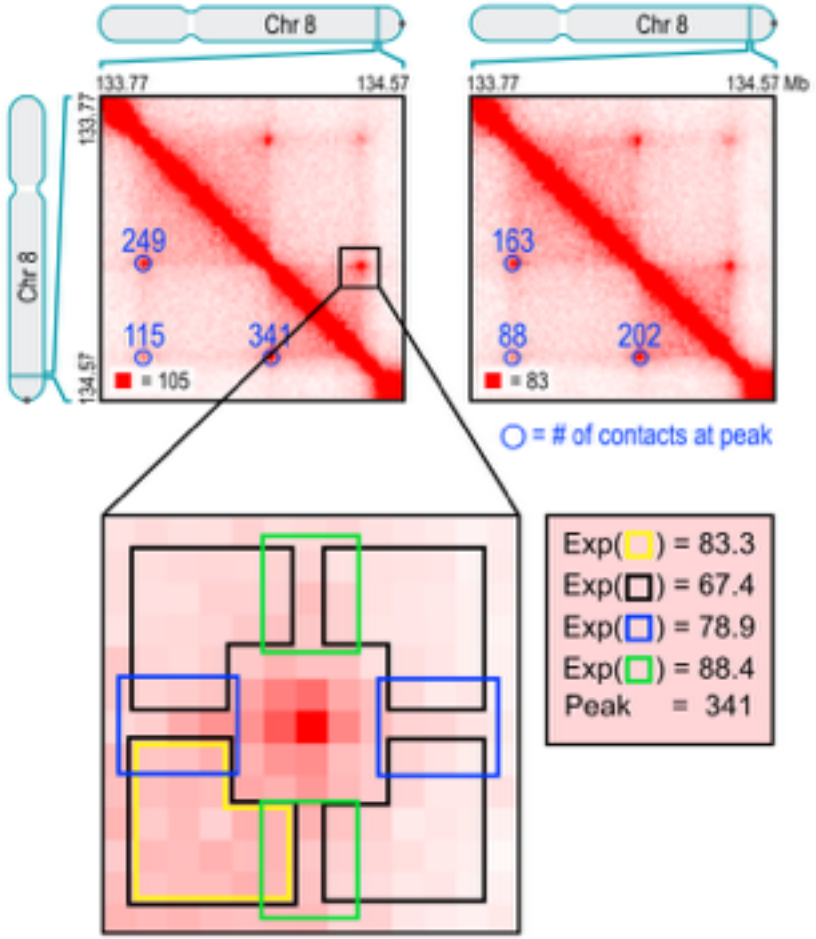




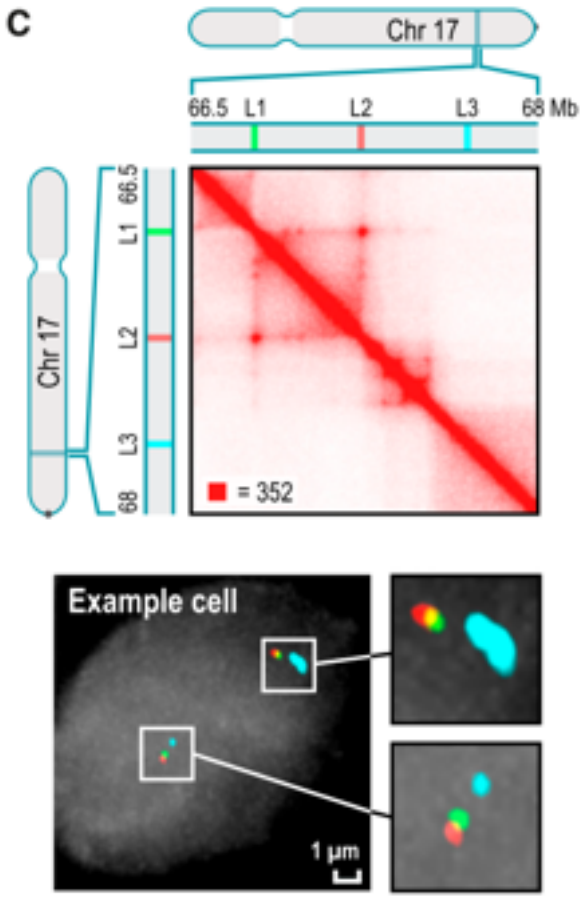
The arrow head matrix is **negative** if locus i + d is **enriched** in contacts with locus i compared to i - d, **positive** if it is **deficient**, and **zero** if it is **neither**

“HICCUP” local background analysis reveals peaks in contact maps that may demarcate chromatin loops

Local background analysis identifies regions within contact domains that exhibit unusually high contact probabilities

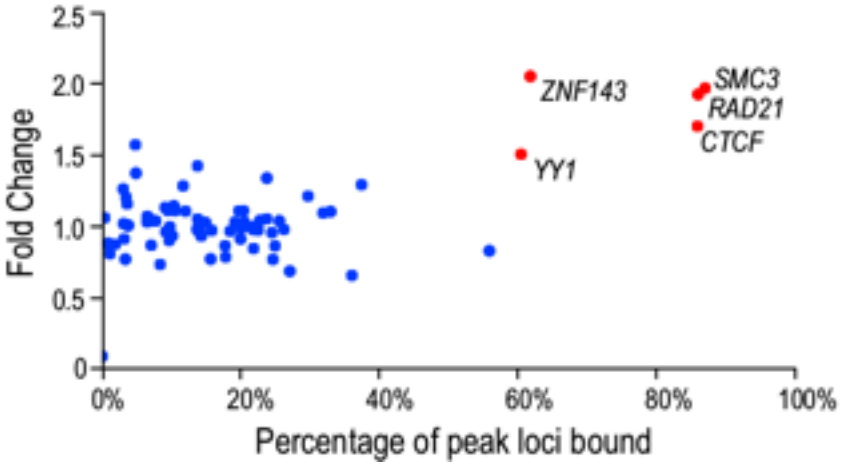


Fluorescence *in situ* hybridization imaging recapitulates contact of loci identified by peak analysis

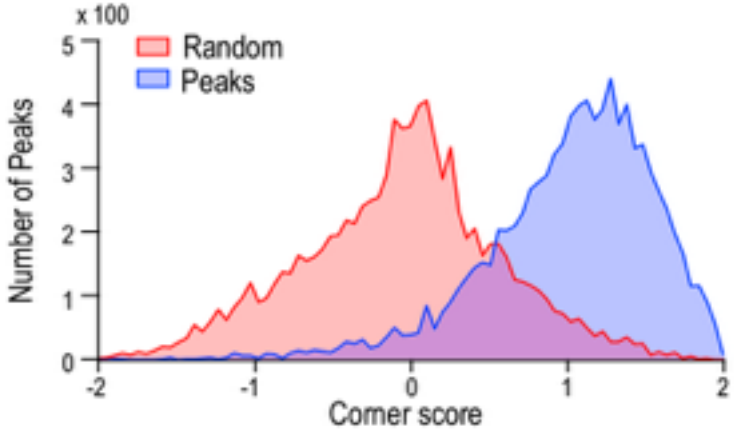


Protein-mediated loops frequently demarcate contact domain boundaries

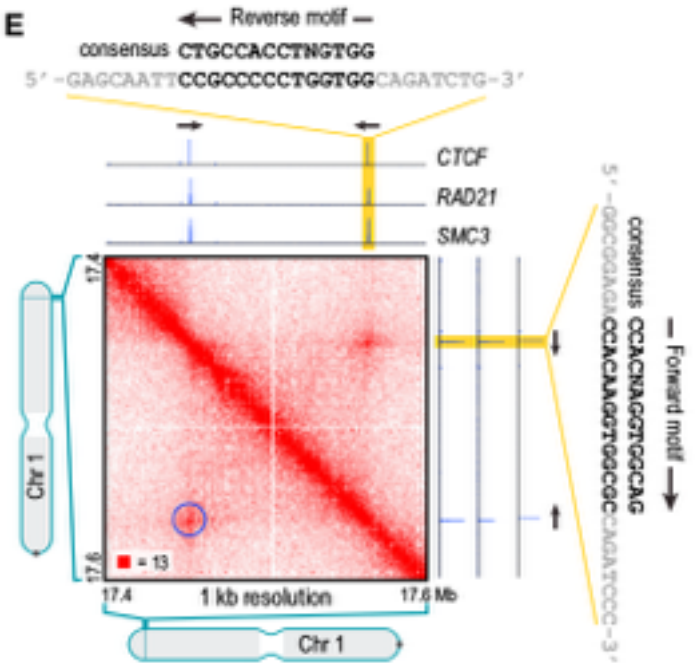
The majority of loops recruit SMC3,RAD21, and CTCF DNA binding proteins

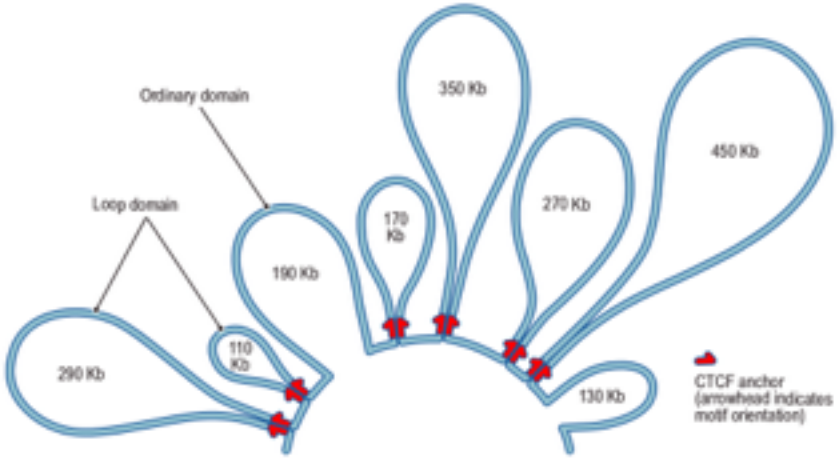


Loops frequently demarcate domain boundaries ("corners")



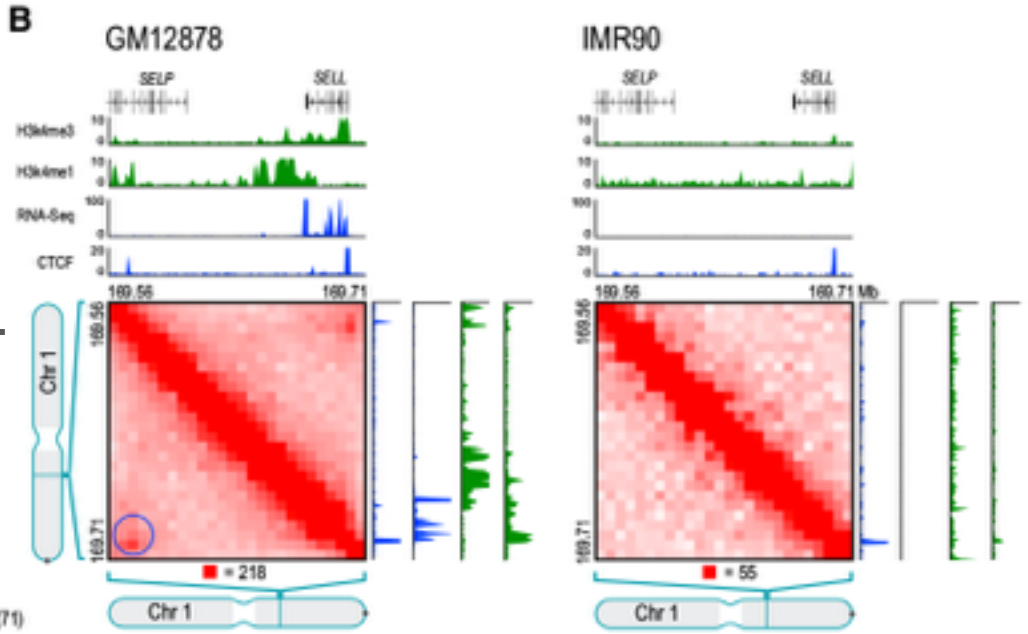
The majority of loops bind CTCF protein pairs in a convergent orientation



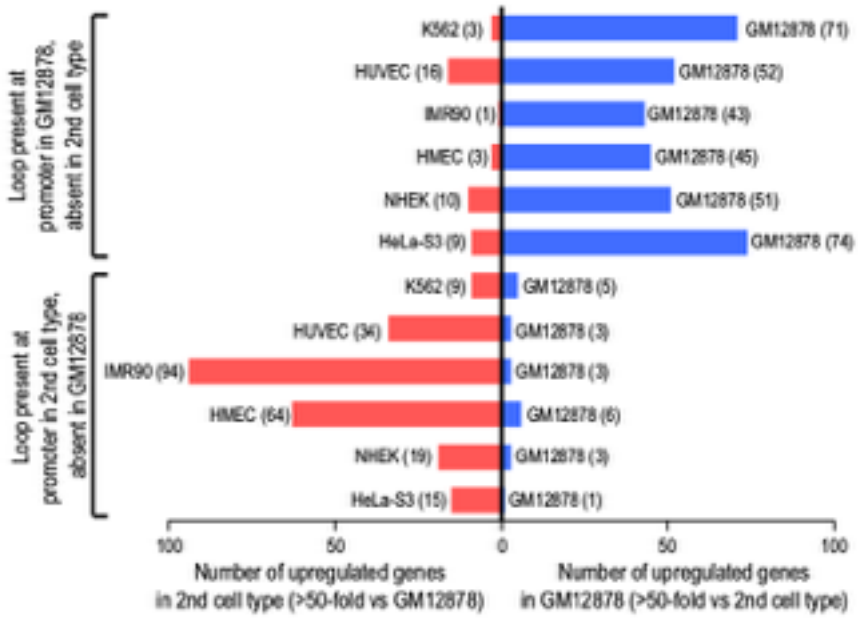


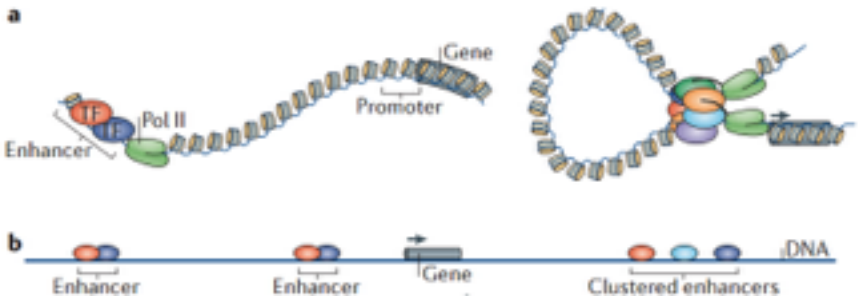
Peak analysis reveals chromatin loops that mediate enhancer-promoter contact and gene activation

In GM12878, SEL gene is on, and a loop forms between the promoter and a distal enhancer.In IMR90, no loop forms and the gene is silenced



Loops present in one cell type and not another correlate with gene activation of phenotype markers



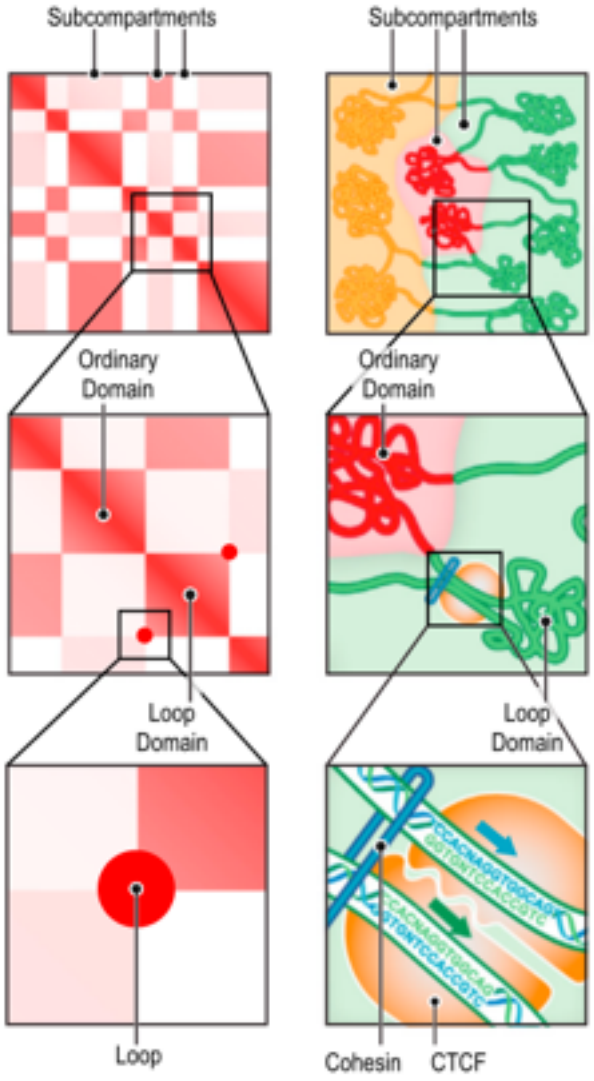


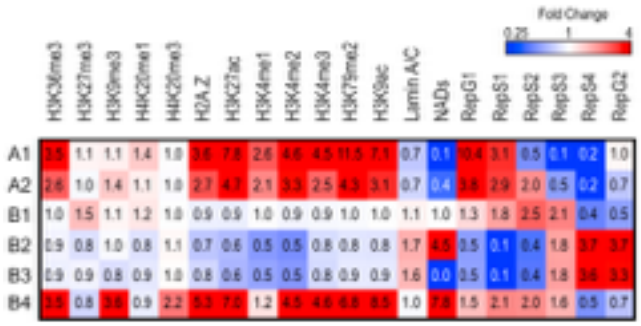
Summary: *in situ* HiC reveals hierarchical folding in the human genome

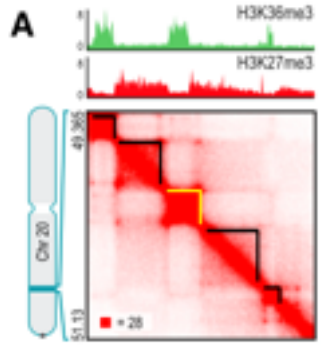
Long-range contact patterns organize the genome into sub-compartments with distinct histo modification patterns

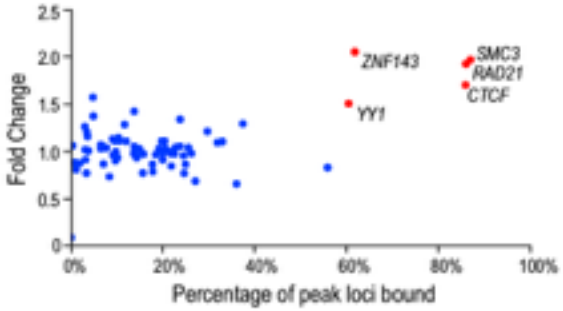
Sub-compartments contain contact domains with internally correlated epigenetic marks

Contact domains are organized by protein mediated loops that frequently demarcate domain boundaries and regulate gene activation



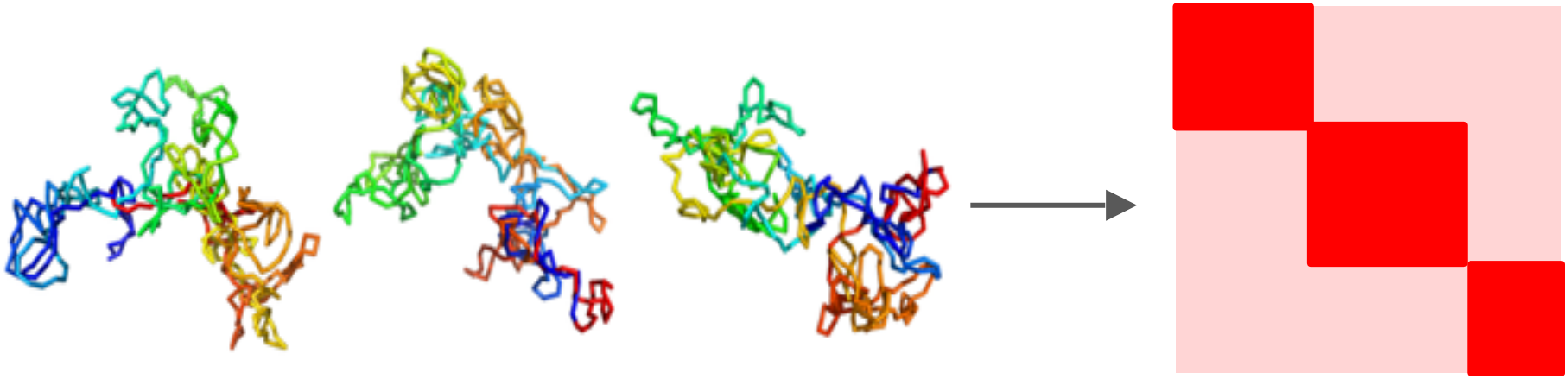




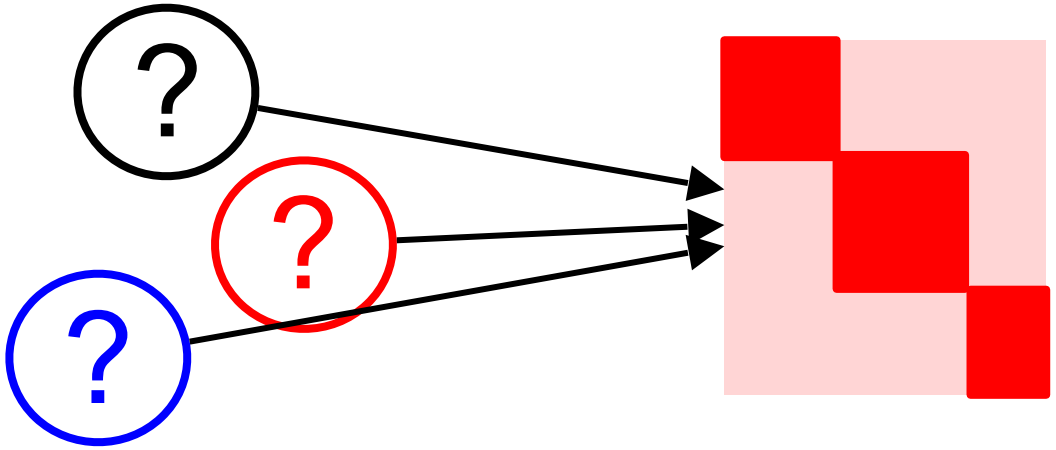


Critiques

Contact maps from cell populations represent averages over **conformational ensembles**, which may conceal genomic architecture within individual cells

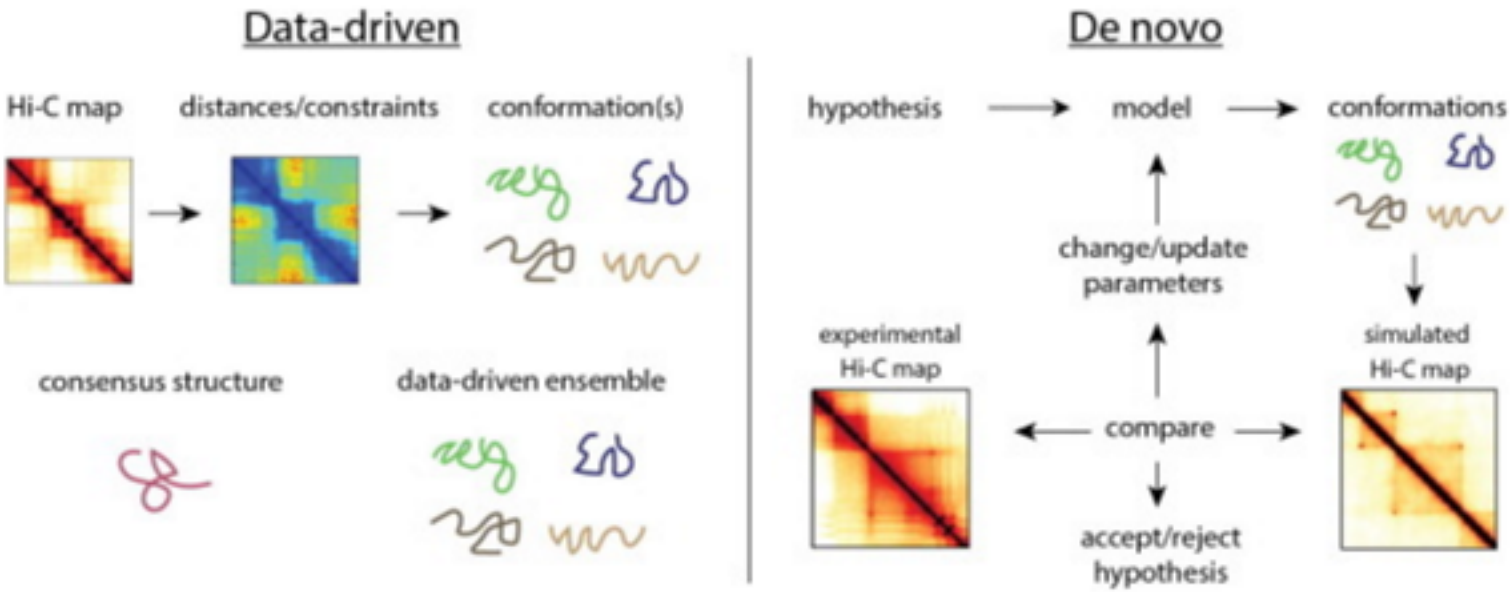


Solving genomic organization from contact maps is highly under-determined: many organizations can produce similar contact probabilities



Outlook on chromsome conformation analysis: a union between biology, data science, and soft matter physics

Polymer physics modeling combined with HiC data may provide new insights into the emergent physics that governs chromosome organization



Imakaev Maxim V.,Fudenberg Geoffrey and Mirny Leonid A. (2015), Modeling chromosomes: Beyond pretty pictures, FEBS Letters, 589

nature biotechnology

Genome architectures revealed by tethered chromosome conformation capture and population-based modeling

Reza Kalhor[1,2], Harianto Tjong[1], Nimanthi Jayathilaka[1,2], Frank Alber[1] & Lin Chen[1,3,4]

A quick overview

Problem: Low signal to noise ratios in Hi-C compromise ability to detect interactions between chromosomes

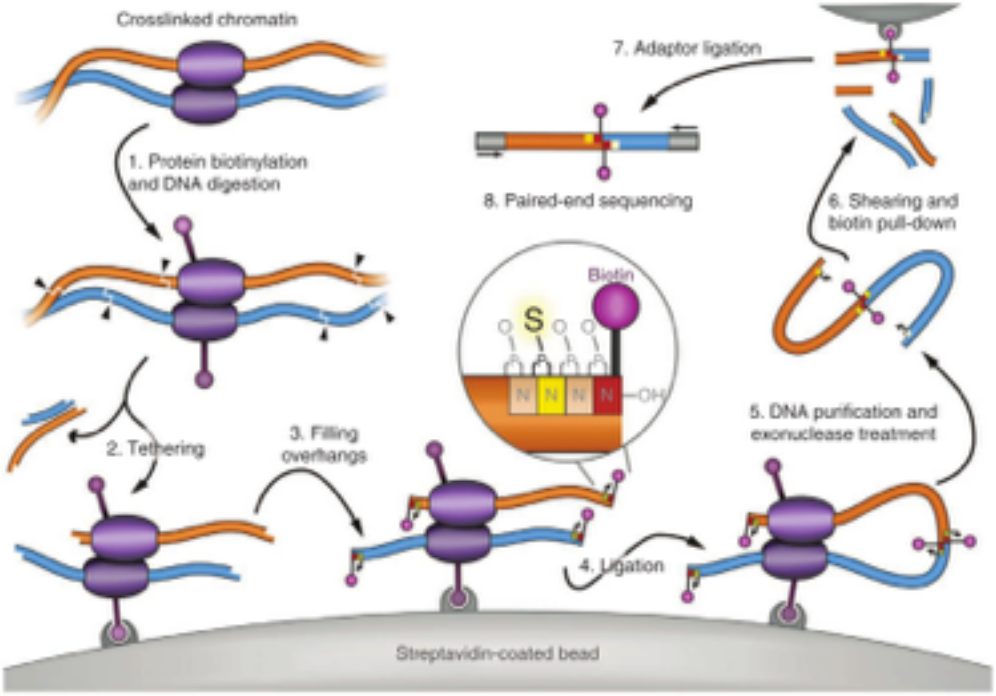
Solution: Develop a new conformational capture experiment with reduced signal to noise

Result: Can use low-noise data to model genome architectures of cell population

Tethered conformational capture (TCC)

Immobilization in TCC improves signal to noise ratio

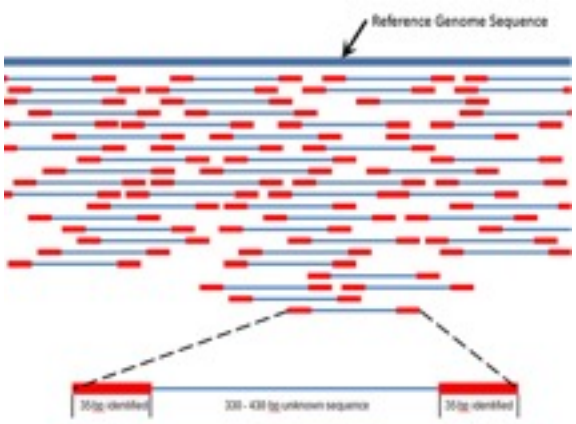
- Can wash away DNA that is not crosslinked
- Immobilization prevents unwanted ligations



Reza Kalhor *et al*, Nature Biotechnology (2011).

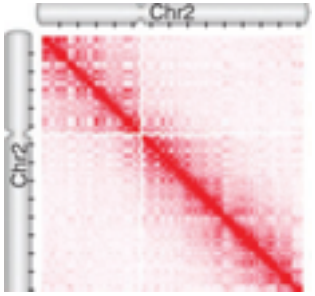
Contact frequency maps can be created from Hi-C reads

Align reads to human genome



Ligation Matrix

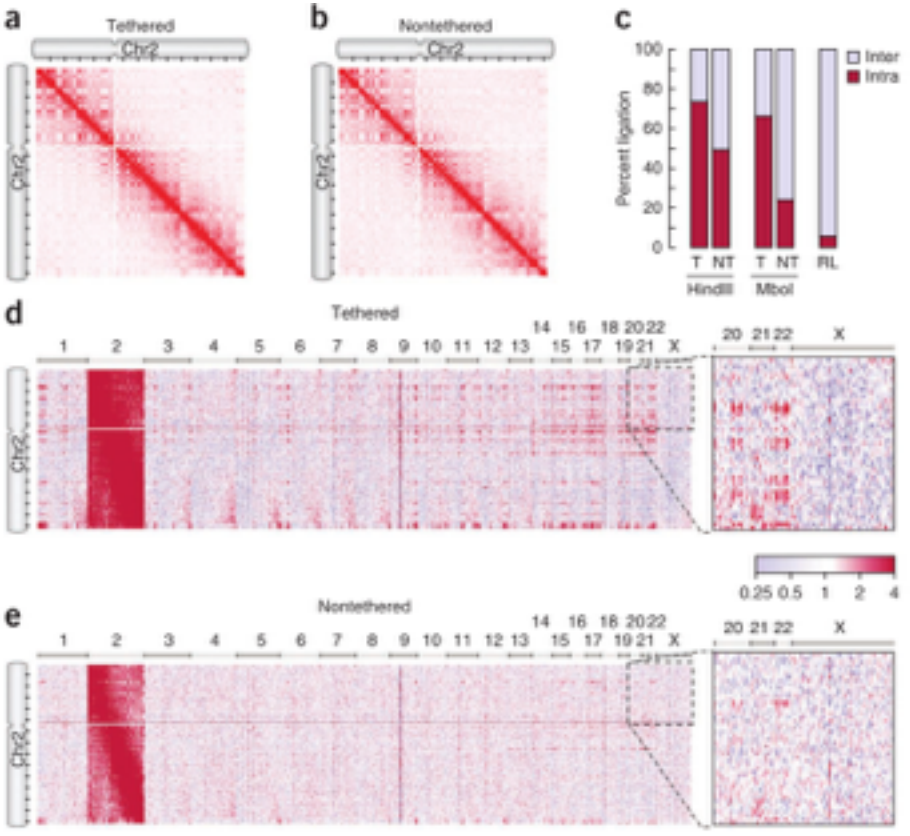
Contact frequency map



A contact frequency map represents the frequency of interactions between regions of chromosomes

Improved signal to noise ratio in tethered libraries

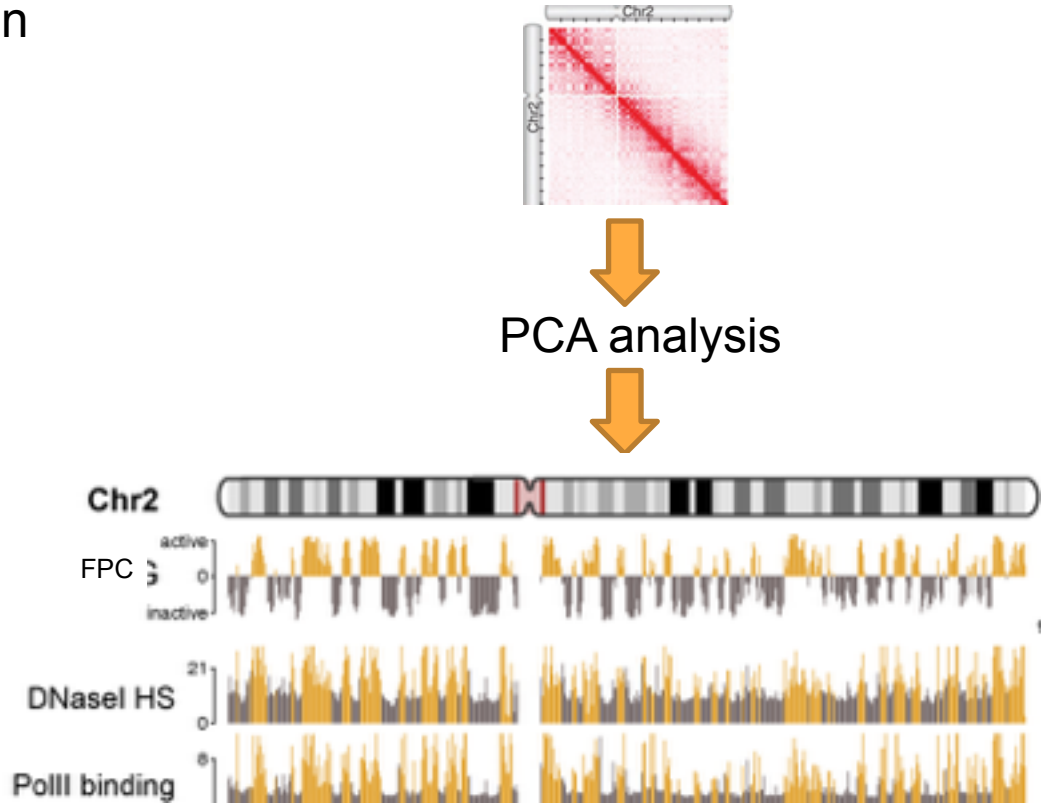
- Good correlation between TCC and Hi-C (**a,b**)
- Less interchromosomal contacts in TCC (**c**)
- Can also compute genome wide enrichment maps (**d,e**)



Reza Kalhor *et al*, Nature Biotechnology (2011).

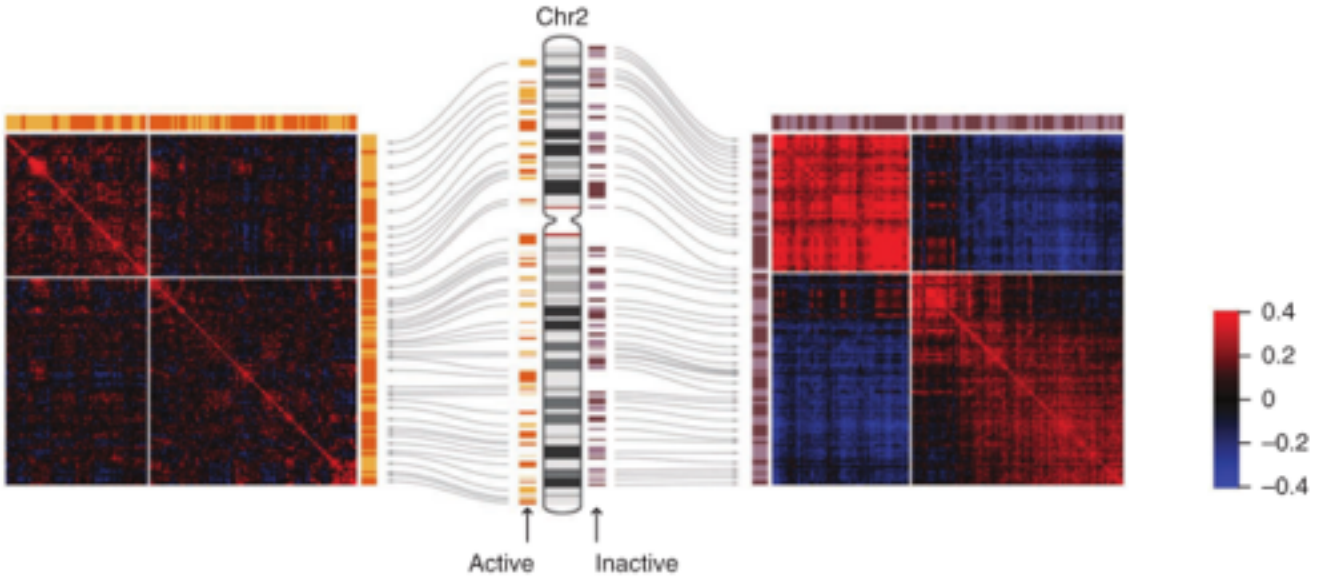
Principal component analysis can be used to determine active vs inactive regions

- First principal component (FPC) can classify chromosome into active or inactive regions
- Active regions have high gene expression, high DNase sensitivity
- Inactive class have low gene expression, low DNase sensitivity
- Confirmed experimentally



Reza Kalhor *et al*, Nature Biotechnology (2011).

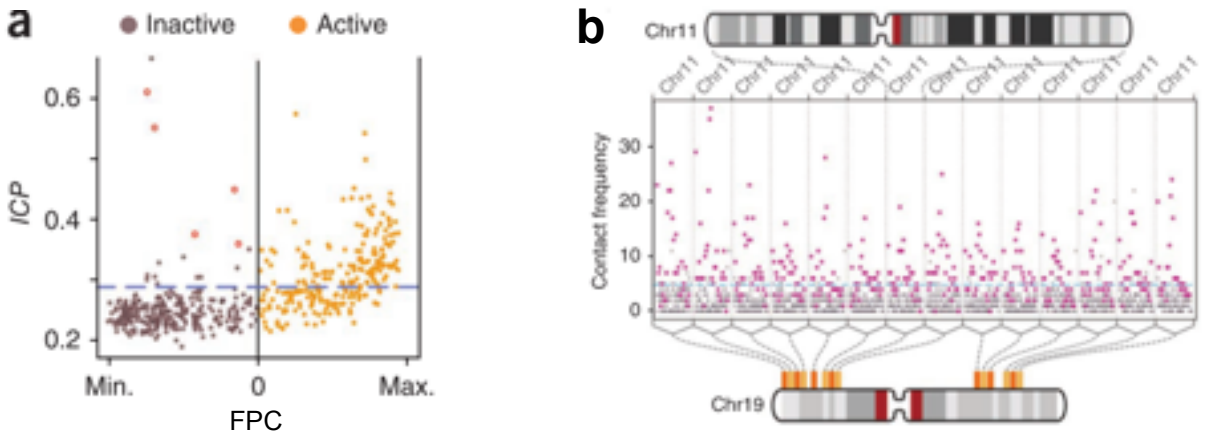
Intrachomosamal interactions can be separated into active-active and inactive-inactive maps



Results suggest that in larger chromosomes inactive regions from opposing chromosomes are largely inaccessible to each other

Reza Kalhor *et al*, Nature Biotechnology (2011).

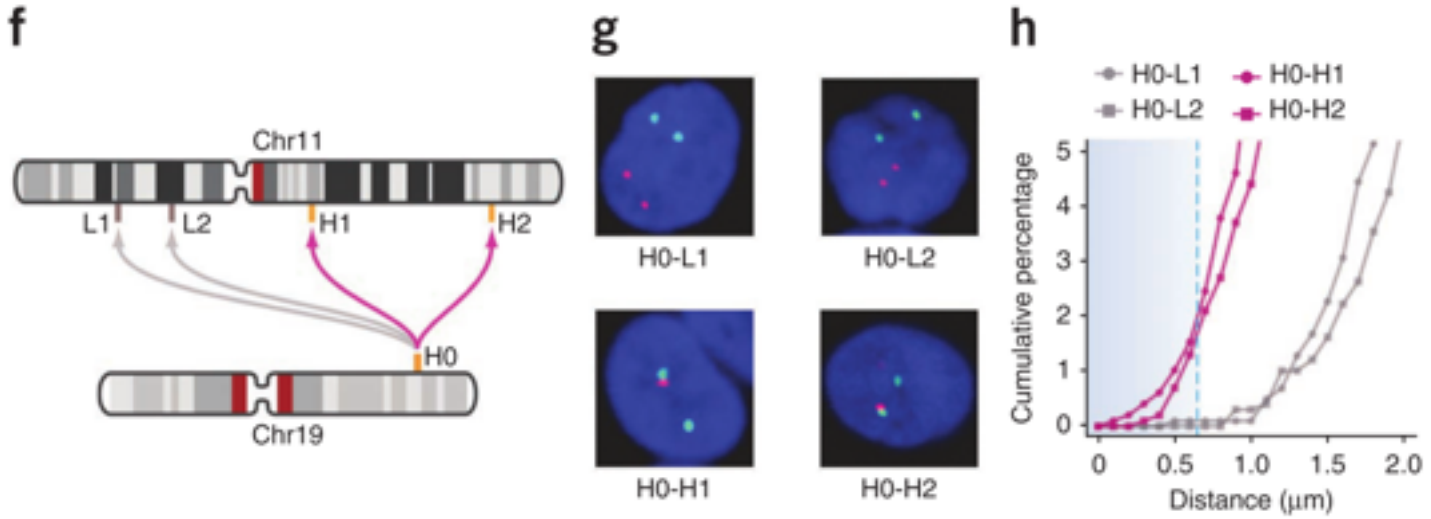
Interchomosomal contacts are mediated by the active class and interact indiscriminately



- Interchomosomal contact probablity (ICP) represents propensity of a region to form contacts with other chromosomes
- Low contact frequencies indicate that formation of interchromosomal contacts largely governed by spatial accessibility (**b**)

Reza Kalhor *et al*, Nature Biotechnology (2011).

3-D fluorescence *in situ* hybridization agrees with TCC interchromosomal contacts



Active region of chromosome 19 interacts more with active regions on chromosome 11 than with inactive regions on chromosome 11

Reza Kalhor *et al*, Nature Biotechnology (2011).

Can 3D genome structures be computed from the interchromosomal contact data?

Wide range of frequencies indicate some contacts may only be present in small fraction of cells -> must generate population of structures

To find these structures, they formed an optimization problem

1. Represent structure of chromosome at appropriate resolution
2. Create a scoring function that uses data
3. A method for optimizing the scoring function

Modeling the 3D organization of the genome

Represent chromosome

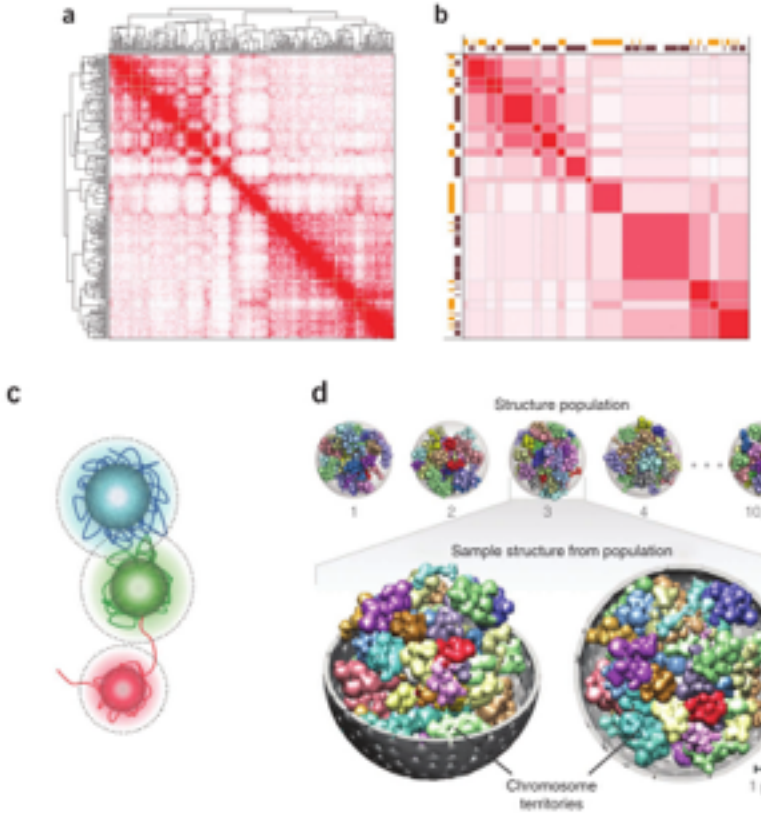
- Divided all chromosomes into 428 blocks (**a,b**)
- Each block represented as a sphere with two radii (hard and soft) (**c**)

Scoring function

- Nuclear volume restraints
- Excluded volume restraints
- Contact restraints

Optimization

- Start from 10,000 random positions and iterate until score=0
- Utilized Integrated Modeling Platform to solve optimization problem

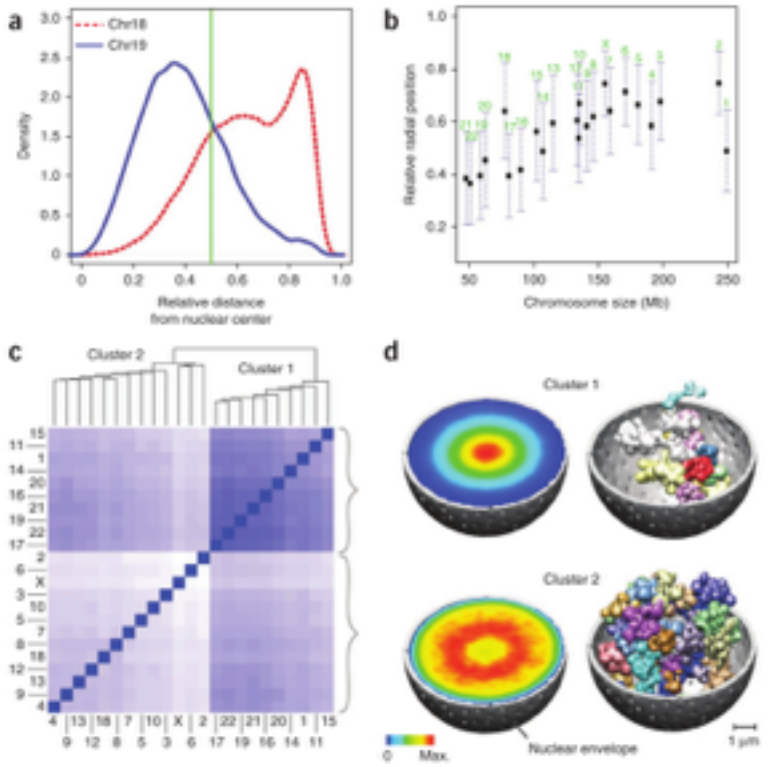


$$S(X_1,\ldots,X_M)=\sum_{m=1}^{M}\sum_{i=1}^{2N}u_{im}^{nuc}+\sum_{m=1}^{M}\sum_{i=1}^{(2N-1)}\sum_{j>i}^{2N}u_{ijm}^{exc}+\sum_{m=1}^{M}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}w_{ijm}u_{ijm}^{con}=0$$

Reza Kalhor *et al*, Nature Biotechnology (2011).

Analysis of computed genome structures

- Large degree of structural variation, contacts only present in small fraction of cells
- On average, only 21% of contacts are shared between any two structures in population
- However radial positions clearly defined and agree with experimental data



Reza Kalhor *et al*, Nature Biotechnology (2011).

A critique of the paper

Strengths

- Good job correcting artifacts through computation
- Thorough documentation of computation (80 page supplemental)

Weaknesses

- More analysis of sources of heterogeneity
- No comparison of genome architecture with Hi-C data
- Weak analysis of results. How will we use this?
- Has not replaced Hi-C as standard for measuring genomic contacts

Future work

- Examine differences between different cell types
- Incorporate experimental data
- Single cell