
3000788 Intro to Comp Molec Biol

Lecture 19: Chromatin conformation capture

Fall 2025



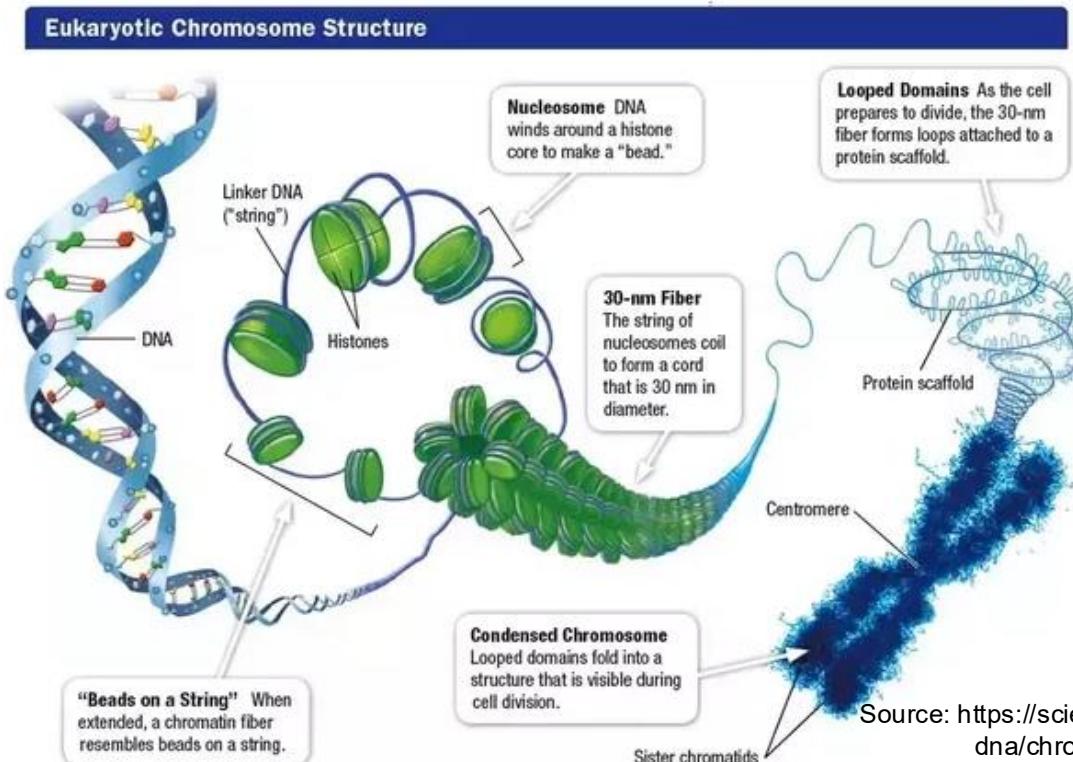
Sira Sriswasdi, PhD

- Research Affairs
- Center of Excellence in Computational Molecular Biology (CMB)
- Center for Artificial Intelligence in Medicine (CU-AIM)

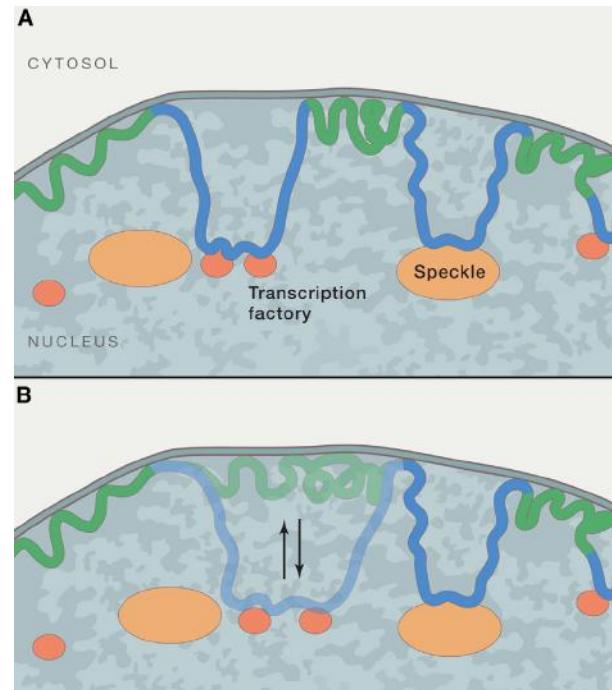
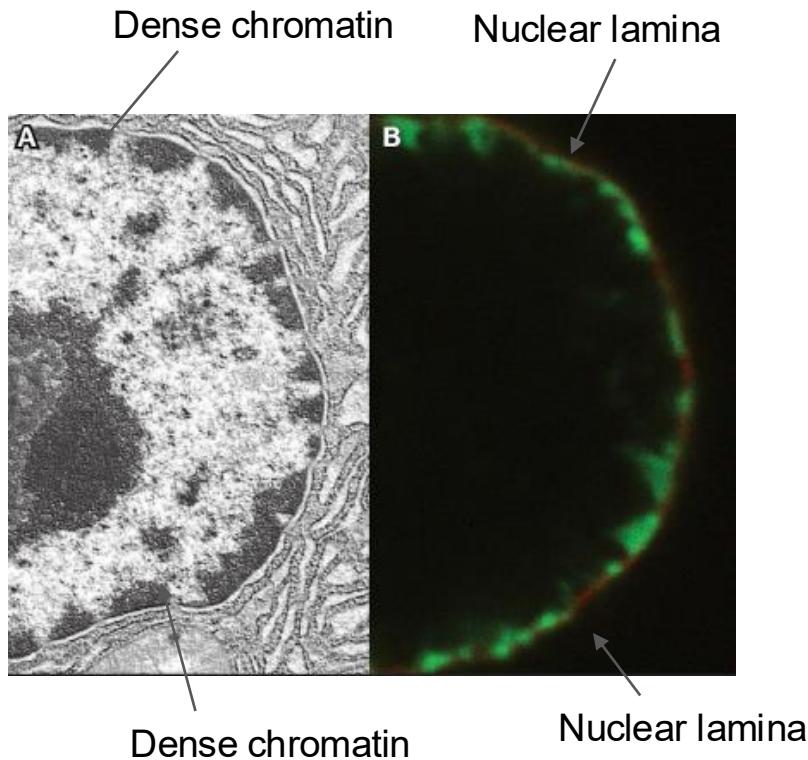
Today's agenda

- Functional roles of chromatin
- Assays for studying chromatin fold
 - 3D structure
 - Proteins complexes involved
- Structural models of chromatin

Chromatin folding for packaging

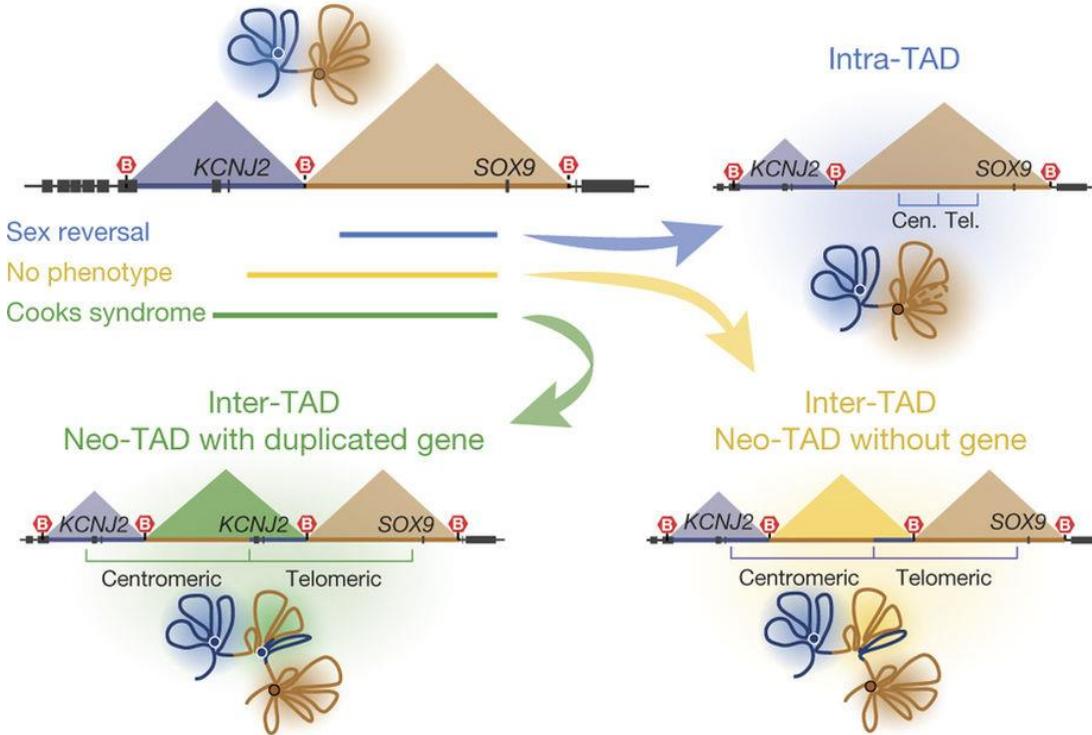


Heterochromatin associates with nuclear lamina



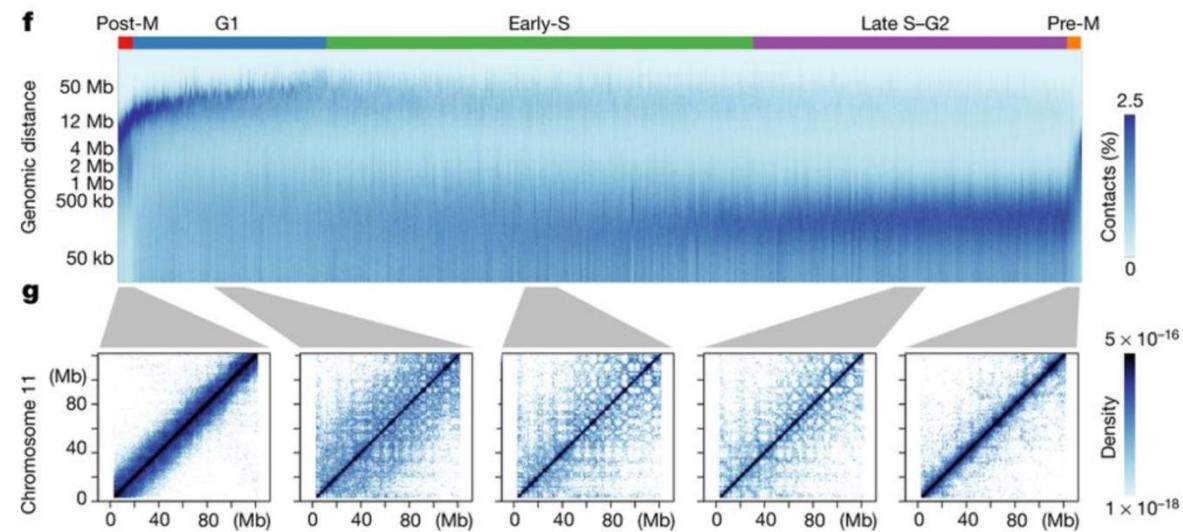
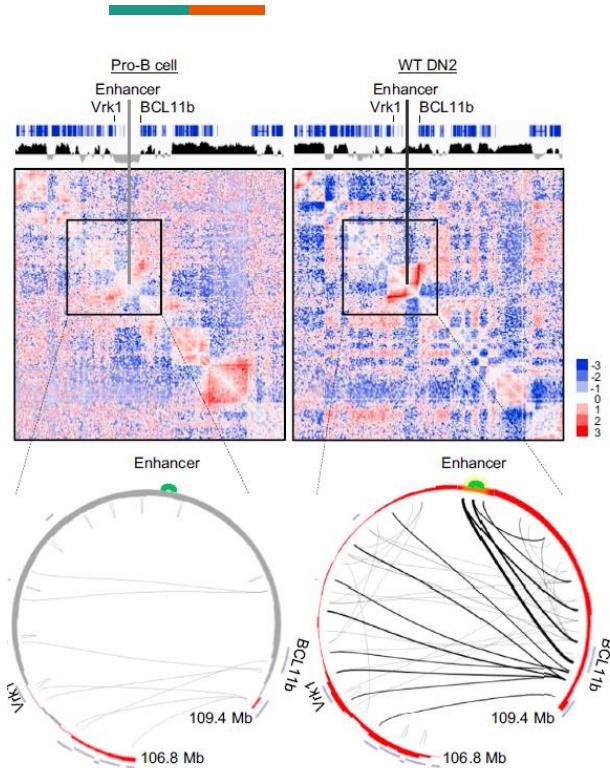
Stensel and Belmont. Cell (2017)

Disease due to chromatin organization



- Cooks syndrome
- Caused by a tandem duplication that **forms a new local structure**
- **TAD** = topologically associating domain
 - Local folding of chromatin

Restructuring of chromatin during development



Nagano et al. Nature (2017)

Isoda et al. Cell (2017)

Mutational hotspots on chromatin



Article | Open Access | Published: 20 October 2016

Chromatin accessibility contributes to simultaneous mutations of cancer genes

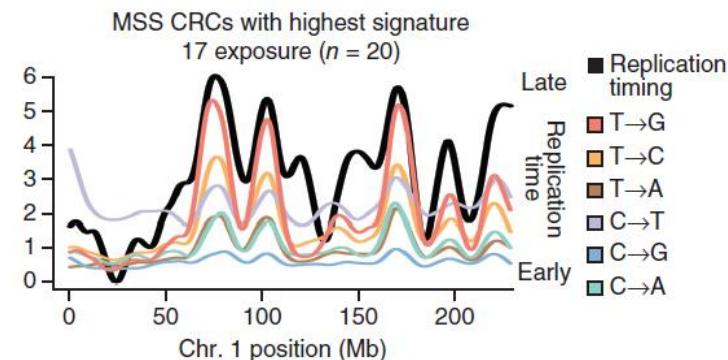
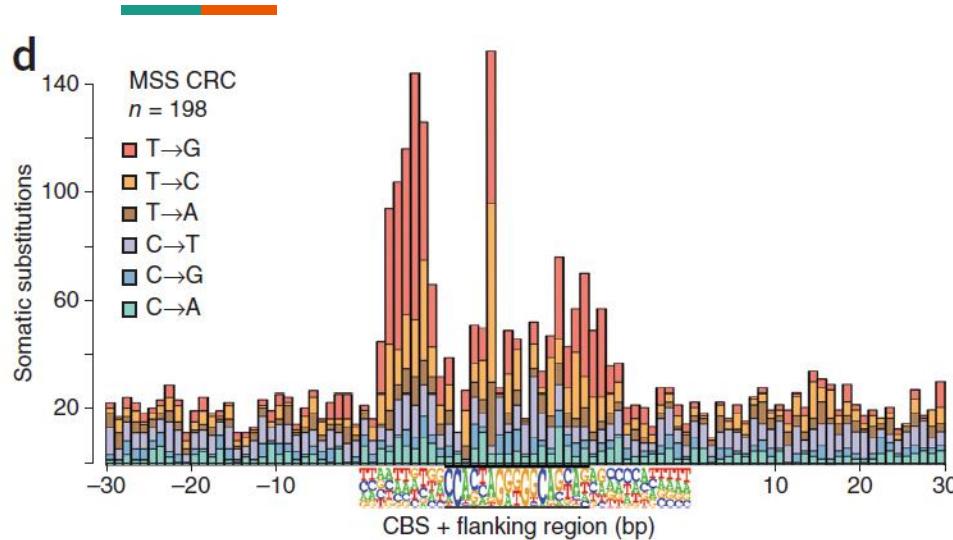
Yi Shi, Xian-Bin Su, Kun-Yan He, Bing-Hao Wu, Bo-Yu Zhang & Ze-Guang Han 

A genome-wide scan for correlated mutations detects macromolecular and chromatin interactions in *Arabidopsis thaliana* 

Laura Perlaza-Jiménez, Dirk Walther 

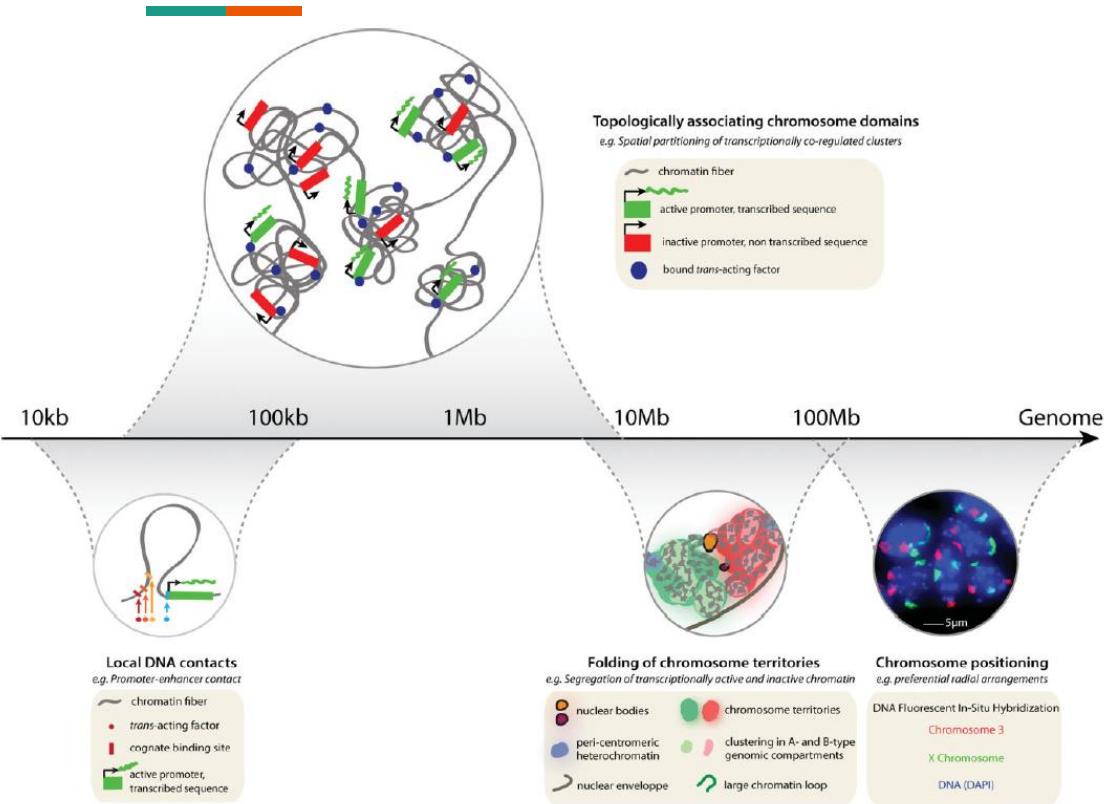
Nucleic Acids Research, Volume 46, Issue 16, 19 September 2018, Pages 8114–8132,

Cancer subtype with destabilized chromatin



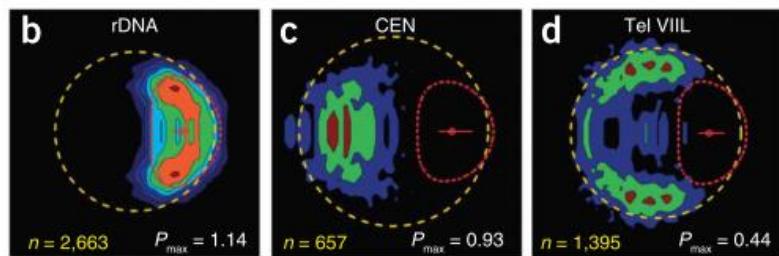
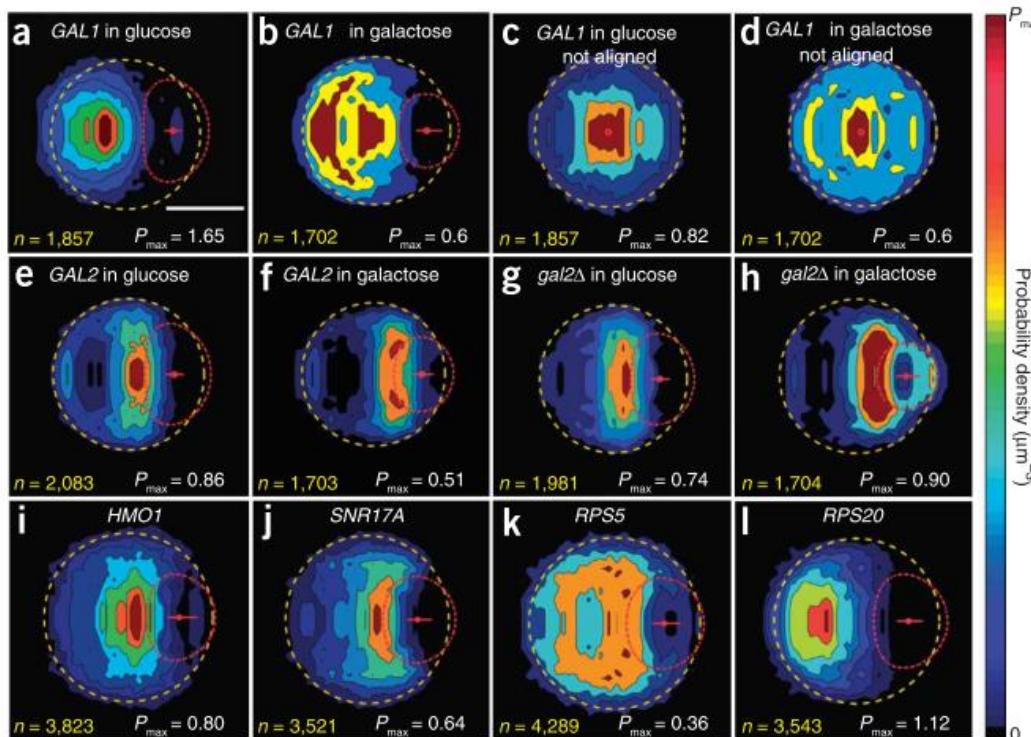
- CBS = CTCF binding site on chromosome to regulate the folding
- Mutations on CBS disrupt chromatin folding
 - Widespread dis-regulations of gene expression

Hierarchical organization of chromatin



- Organized chromatin is easy to split during cell division
- Topologically associating domain (TAD) = local chromatin folds
- Enhancer looping

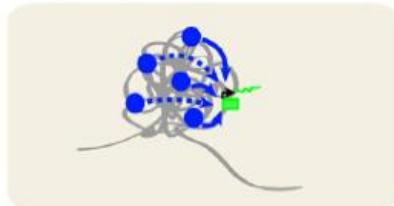
Chromain and gene territories



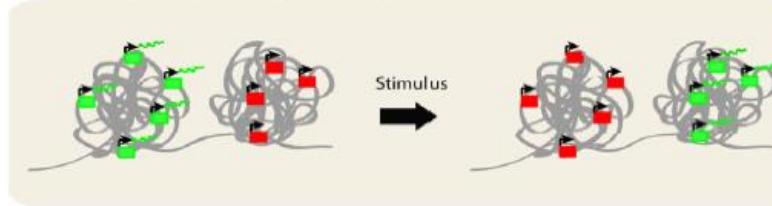
- Chromatin features, such as rDNA, centromere, and telomere, and some genes **occupy specific regions in the nucleus**

Biological implications of gene territories

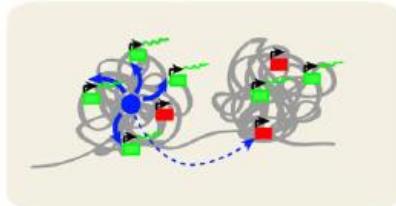
A) Regulatory landscape effect



C) Partitioning of oppositely regulated neighborhoods



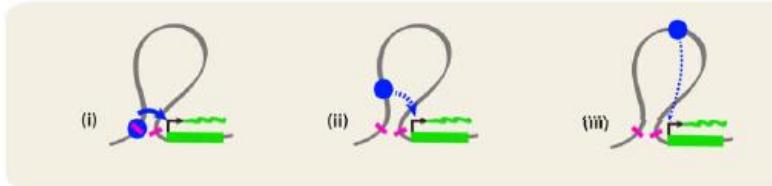
B) Enhancer sharing and allocation



D) Ripple effects of transcriptional activation



E) Architectural and Enhancer elements within TADs



- ~ chromatin fiber
- trans-acting factor
- architectural element
- efficient enhancer-promoter communication
- ↑ inefficient enhancer-promoter communication
- [green box] active promoter, transcribed sequence
- [red box] inactive promoter, non transcribed sequence
- [orange box] spuriously activated promoter, lowly transcribed sequence

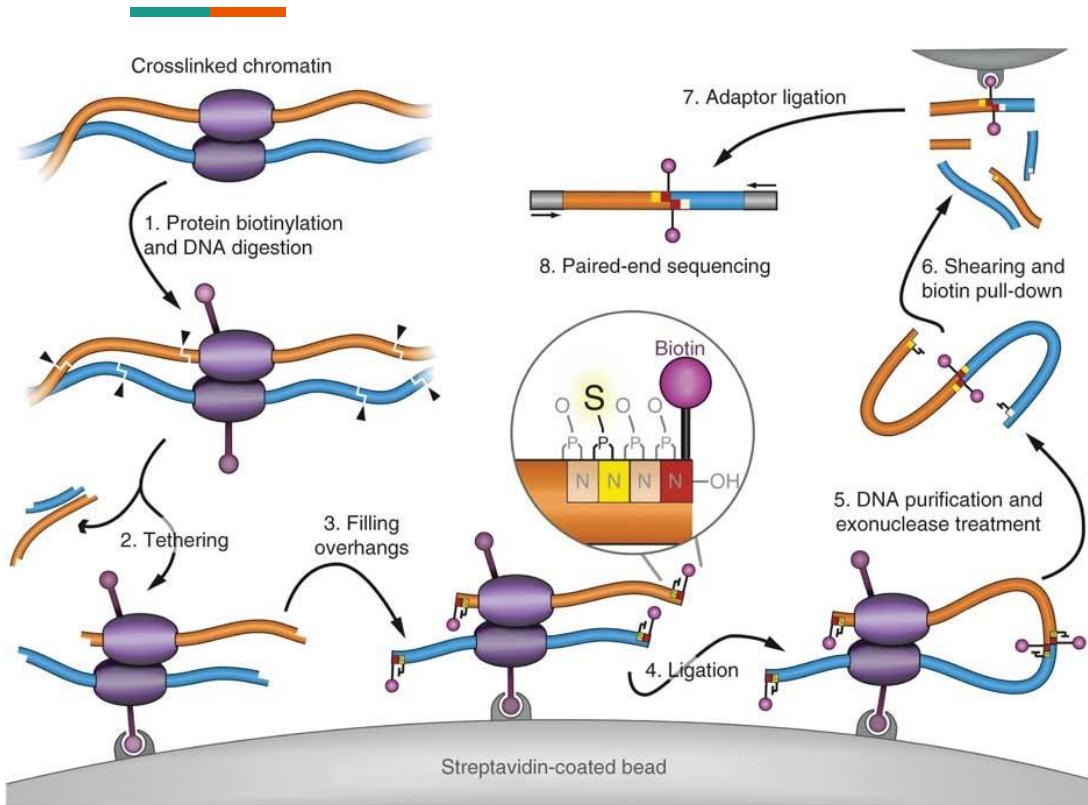
- Localization of co-regulated genes
- Sharing of transcription factors and enhancer

Nora et al. Bioessay (2013)



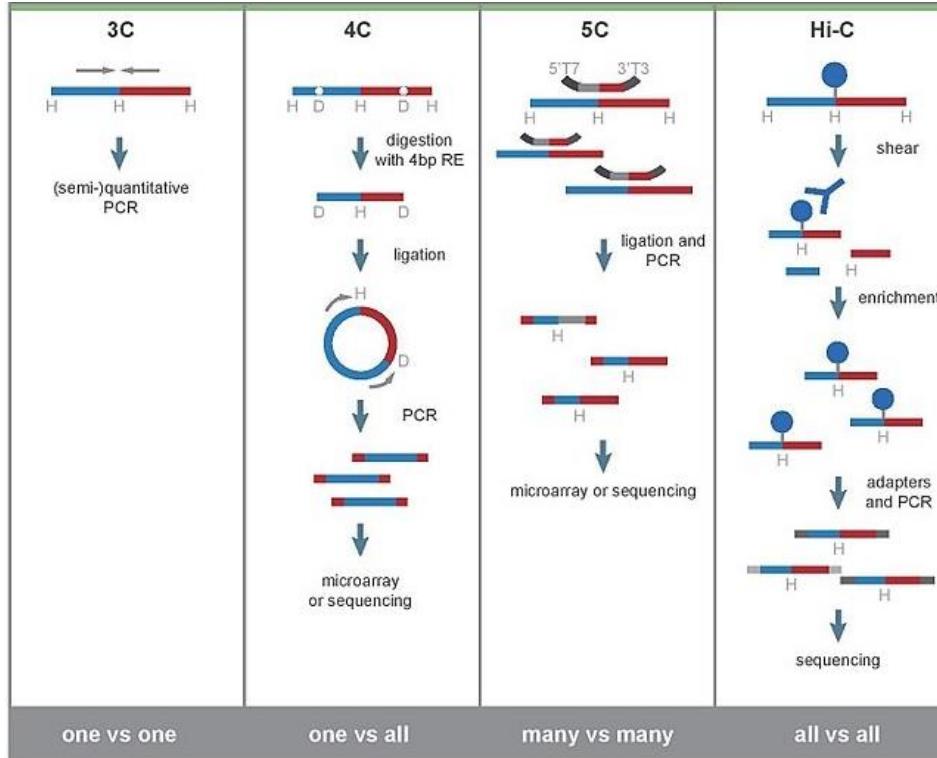
Assays for chromatin structure

Chromatin conformation capture



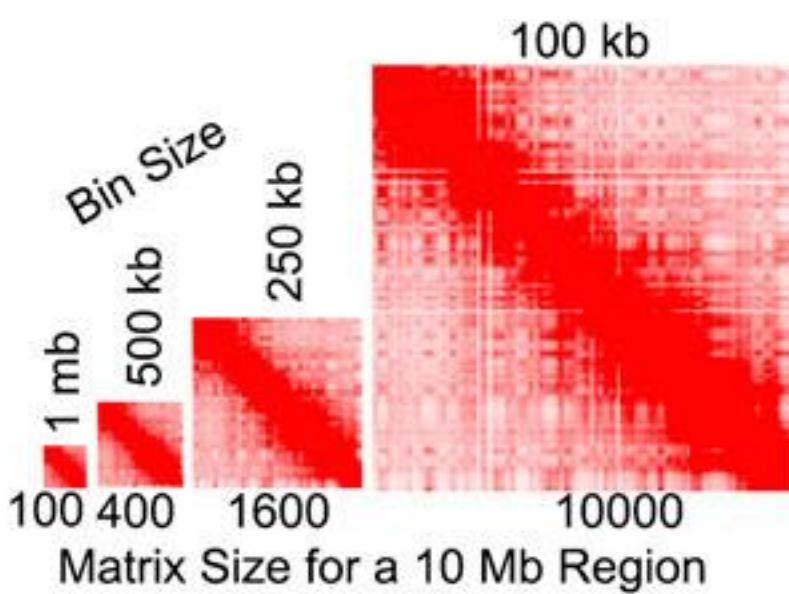
- Induce cross-link between proximal chromatins
- Pull down by biotin
- Generate hybrid fragments
- Identify by high-throughput sequencing

Chromatin conformation capture techniques



- Probe all 3D proximity regardless of mechanisms
- **Hi-C** = high-throughput and broad genome coverage
- **Targeted**
 - 3C = specific interaction
 - 4C = all interactions involving a specific locus
 - 5C = multi-loci vs multi-loci

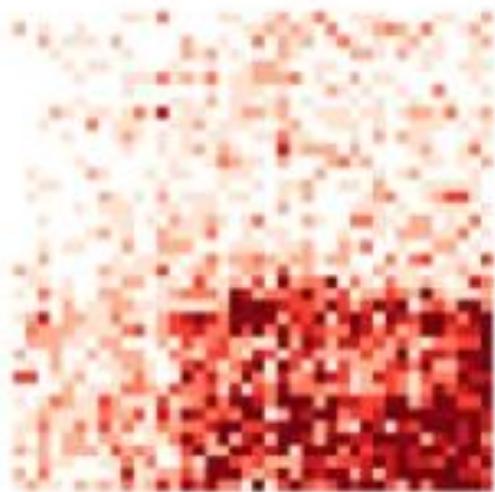
Chromatin structure resolution



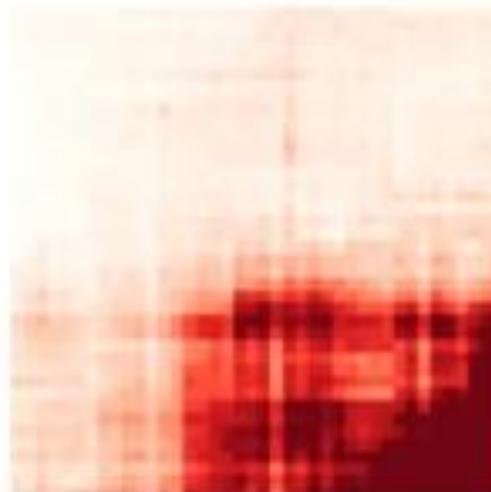
- Resolution = size of 1 bin of genomic region
 - Typically 1kb, 10kb, ..., up to 1Mb
- High resolution → small bin size → more details for downstream structural analysis
- High resolution requires high sequencing depth

Hi-C data requires high sequencing depths

Insufficient depth



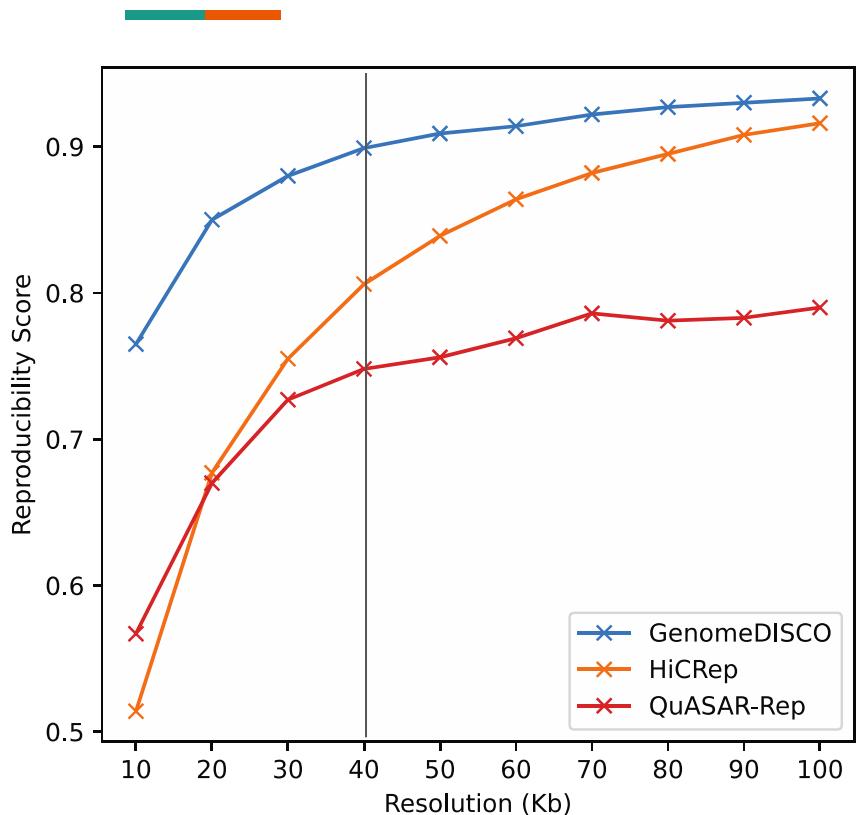
High sequencing depth



Zhang et al., Nat Comm (2018)

- Hi-C data = filling an $N \times N$ matrix with read counts
- Required sequencing depth scales with **genome size squared**
- Low sequencing depth generates spotty Hi-C profiles

Finding appropriate Hi-C resolution for your data



- Similar idea to **rarefaction** for metagenomics data
- Split read data into two halves
- Vary resolution (genomic bin size)
- Identify the highest resolution (smallest bin size) where the two halves produce the same result

QC tools for Hi-C data

Source: github.com/kundajelab/3DChromatin_ReplicateQC

HiCRep

Transformation: 2D mean filter

Comparison: weighted sum of correlation coefficients stratified by distance

Reproducibility score:

$$\sum_d w_d \cdot \rho_d$$

↓ weight ↓ correlation

GenomeDISCO

Transformation: smoothing using graph diffusion

Comparison: difference in smoothed contact maps

Reproducibility score:

$$\frac{\sum_i \sum_j |rw(A)_{ij} - rw(B)_{ij}|}{\# \text{ genomic bins}}$$

HiC-Spector

Transformation: eigen-decomposition of Laplacian

eigenvector 1
eigenvector 2
eigenvector 3
eigenvector 4
eigenvector 5

eigenvector r ...

Comparison: weighted difference of eigenvectors

$$S_d(A, B) = \sum_{i=0}^{r-1} \|v_i^A - v_i^B\|$$

Reproducibility score:

$$\left(1 - \frac{1}{r} \frac{S_d}{l}\right) \quad l = \sqrt{2}$$

QuASAR-Rep

Transformation: correlation matrix of distance-based contact enrichment

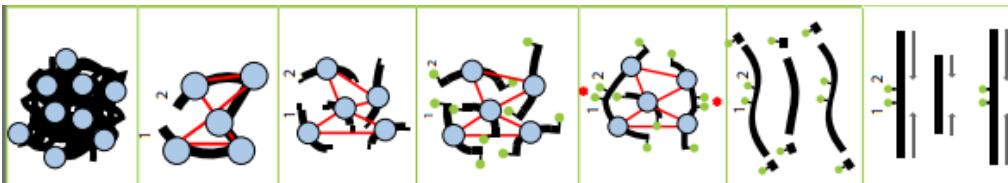
Comparison: compute correlation of values in the 2 transformed matrices

Reproducibility score:

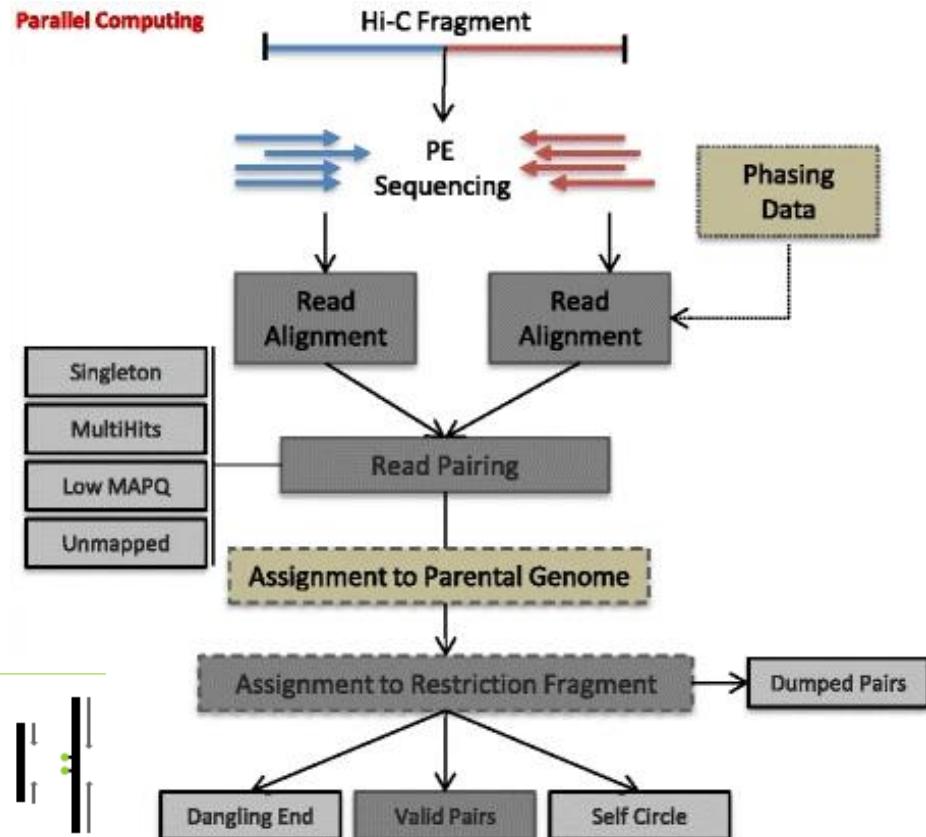
Pearson correlation ($\text{quasar}(A), \text{quasar}(B)$)

Hi-C analysis steps

- Mostly identical to generic DNA sequencing workflow
- Filtering for **valid reads** that came from the Hi-C protocols
 - Orientation and distance of paired reads relative to restriction enzyme site

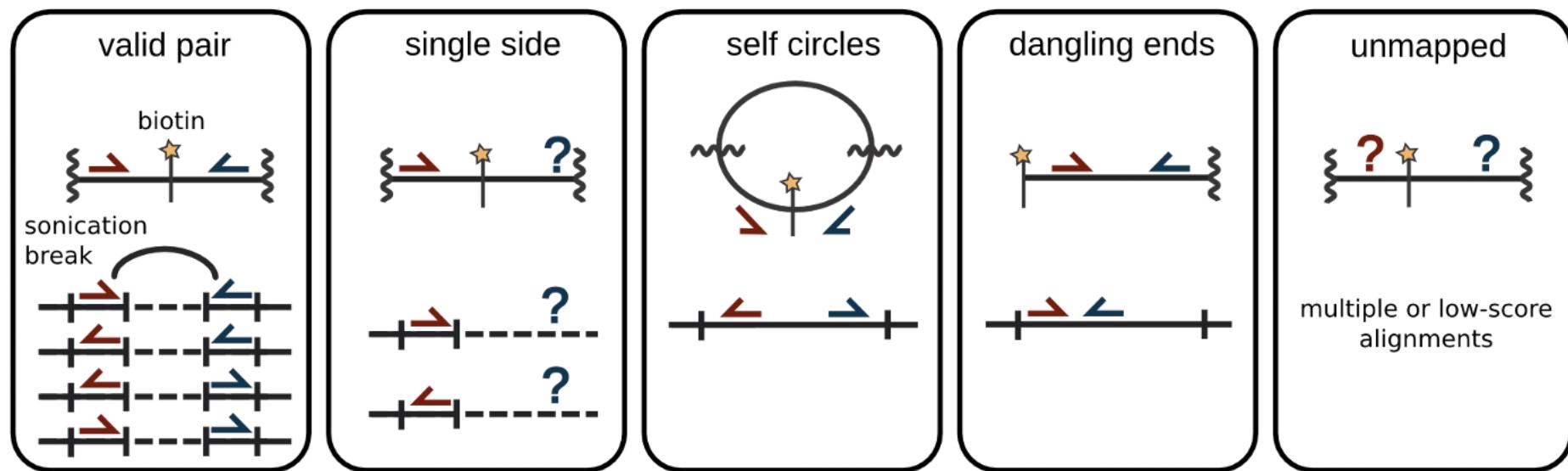


Source: Dovetail Genomics Hi-C pamphlet



Servant et al. Genome Biology (2015)

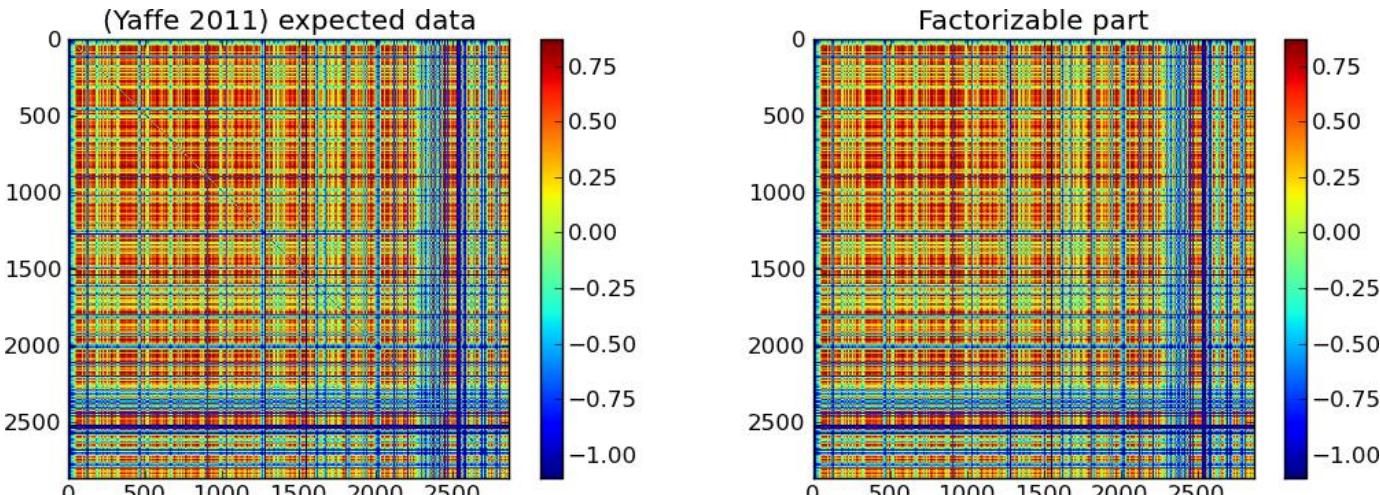
Valid fragments vs unwanted products



Imakaev et al. Nature Methods (2012)

- Look for read pairs whose orientation point away from restriction sites

Normalizing read count bias

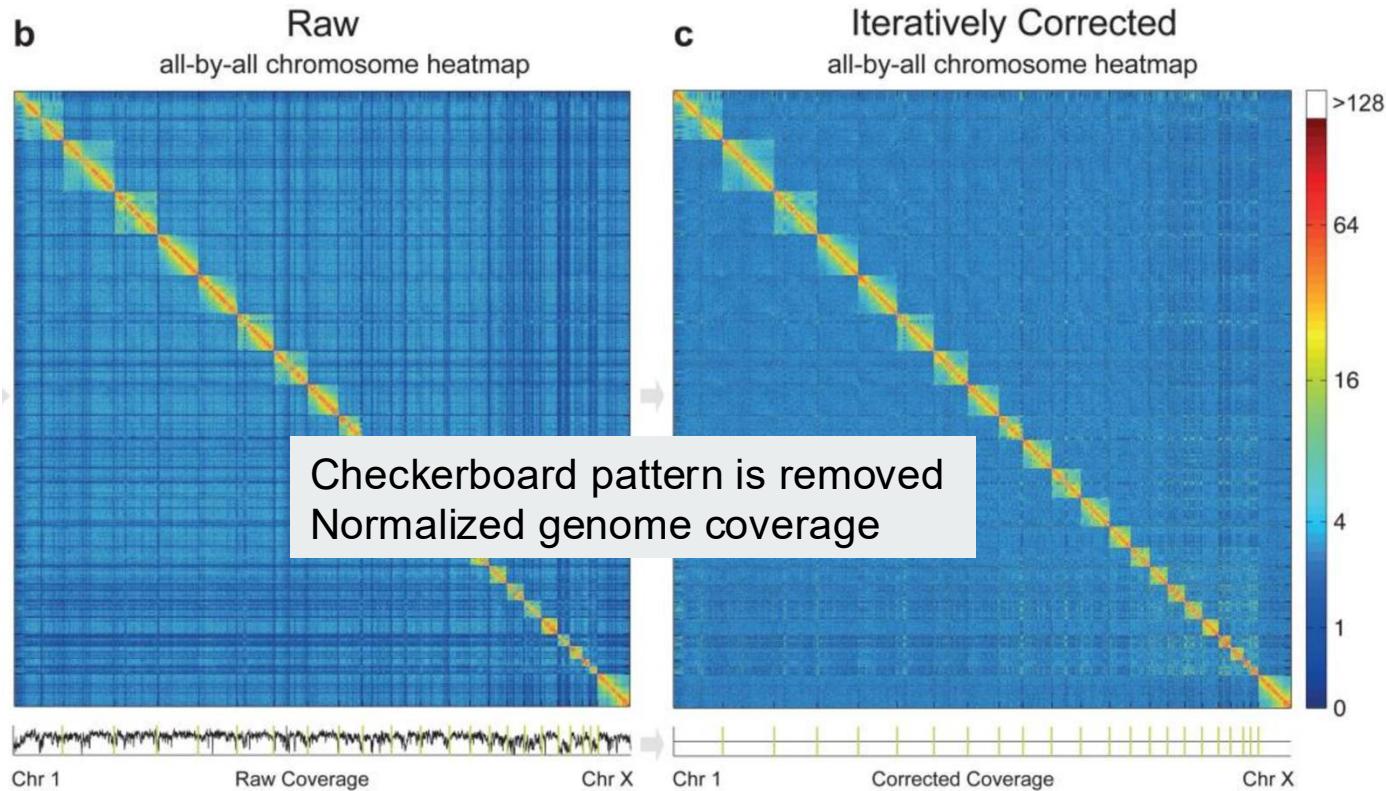


Imakaev et al. Nature Methods (2012)

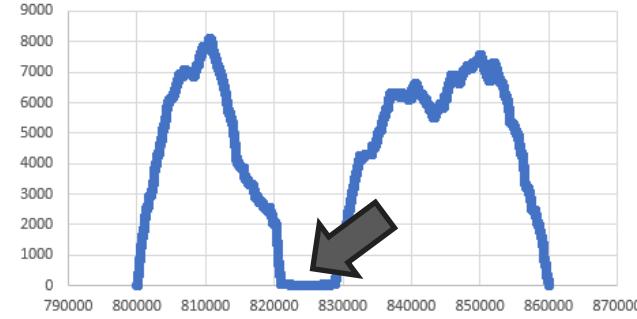
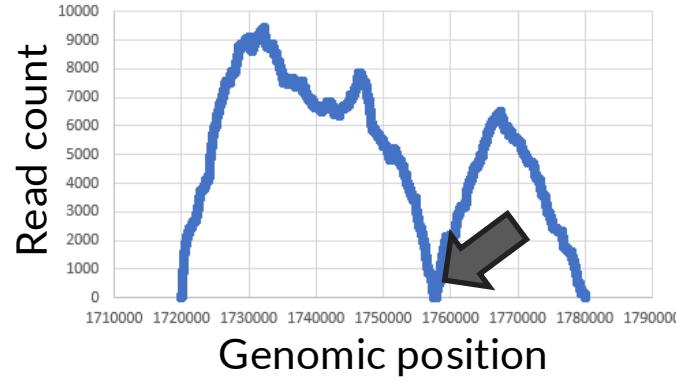
- Total bias for interaction between loci A and B = **bias at A** \times **bias at B**
 - Probability of getting DNA fragment from a genomic region by chance)

Iterative correction and eigenvector decomposition (ICED)

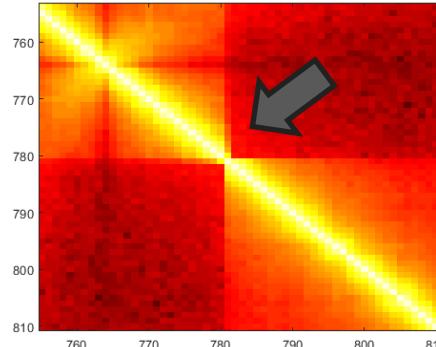
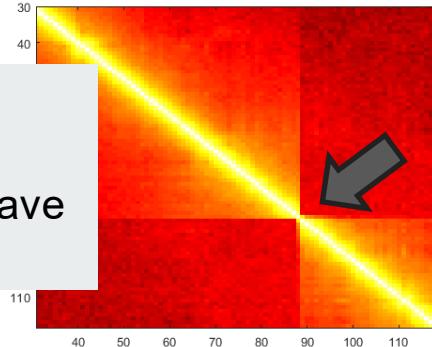
Imakaev et al. Nature Methods (2012)



Hi-C identifies misassembled genomic region



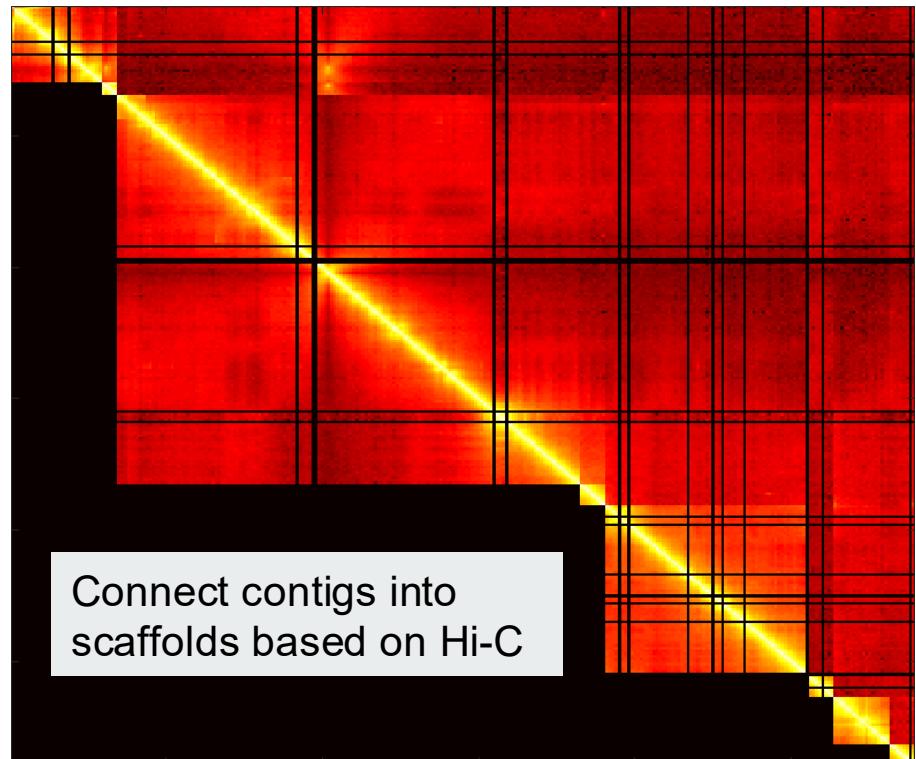
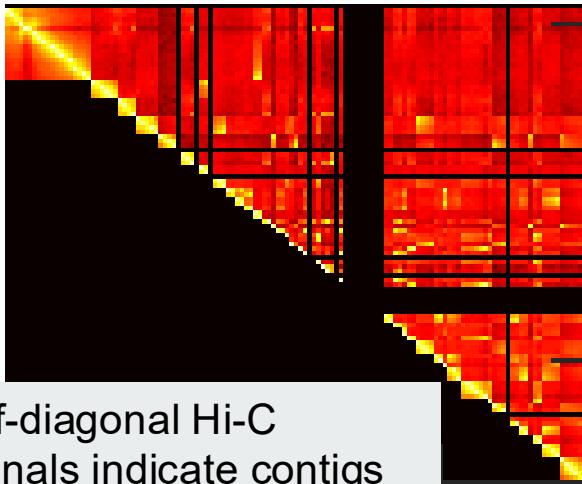
Mis-assembly can occur when two genomic regions have similar sequences



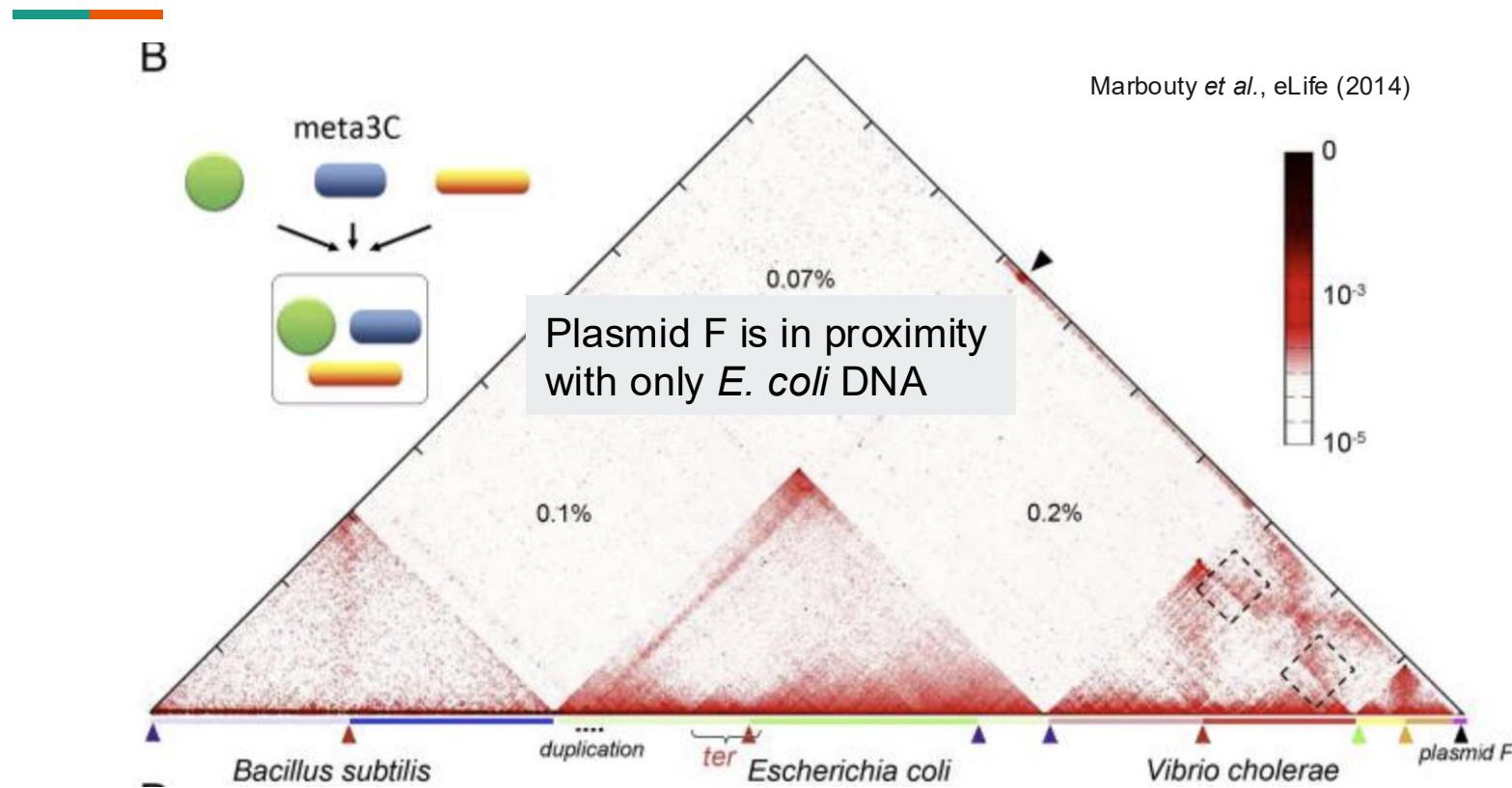
Hi-C guides genome assembly



Many small contigs



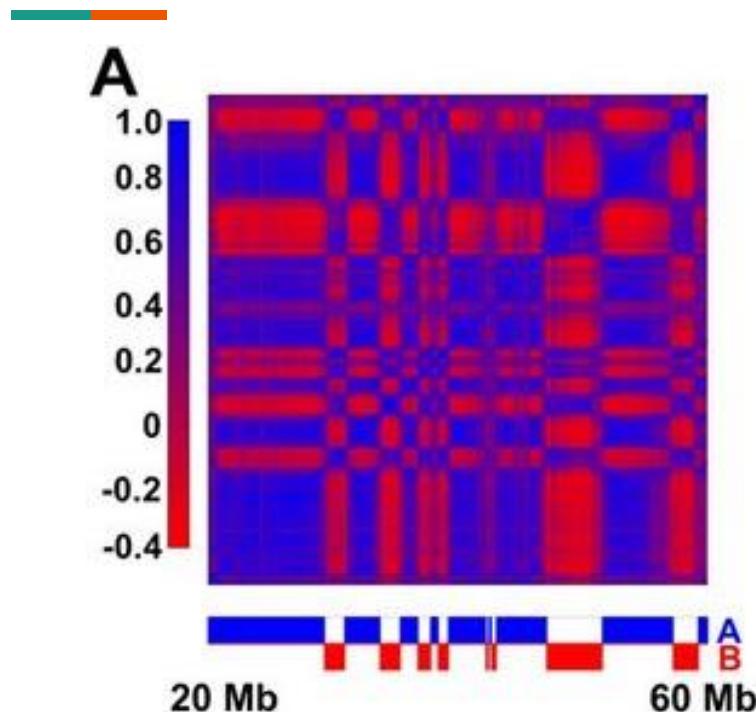
Hi-C guides metagenomic host assignment



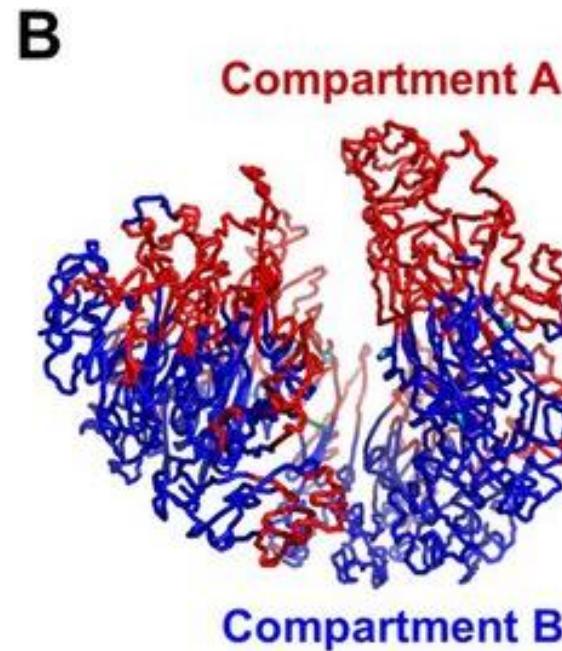


A/B compartment and topologically associating domain (TAD)

A/B compartment

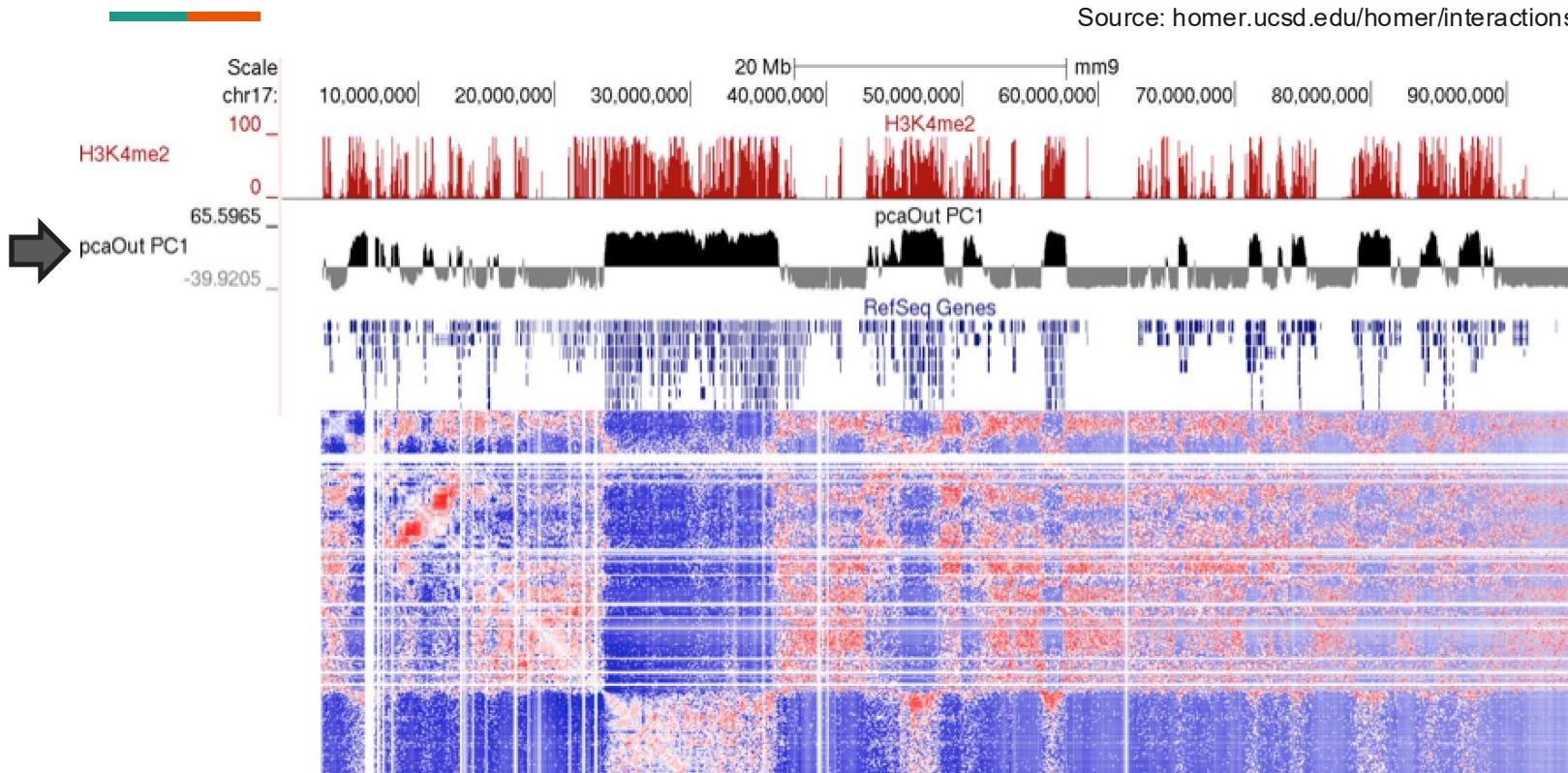


Xie et al. Scientific Reports (2017)



- Caused by localization of euchromatin and heterochromatin

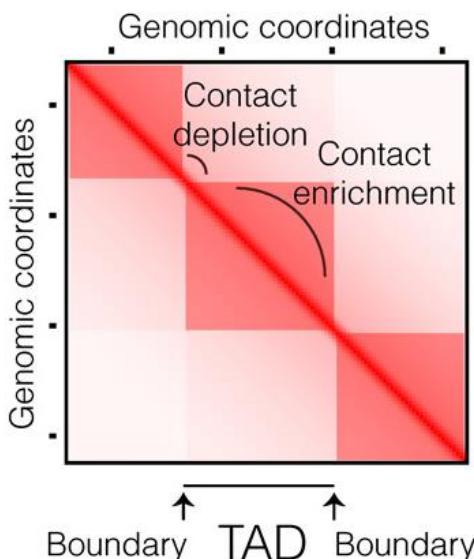
A/B compartment identification with PCA



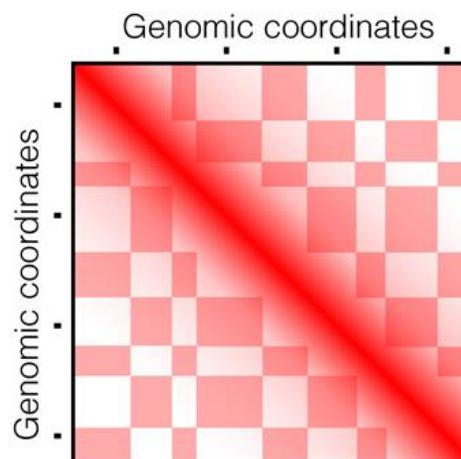
Hierarchical organization of chromatin



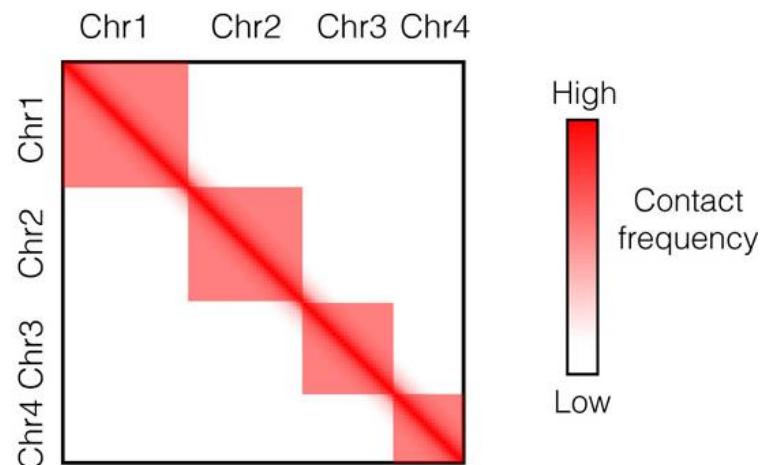
TADs



Compartments



Chromosome territories

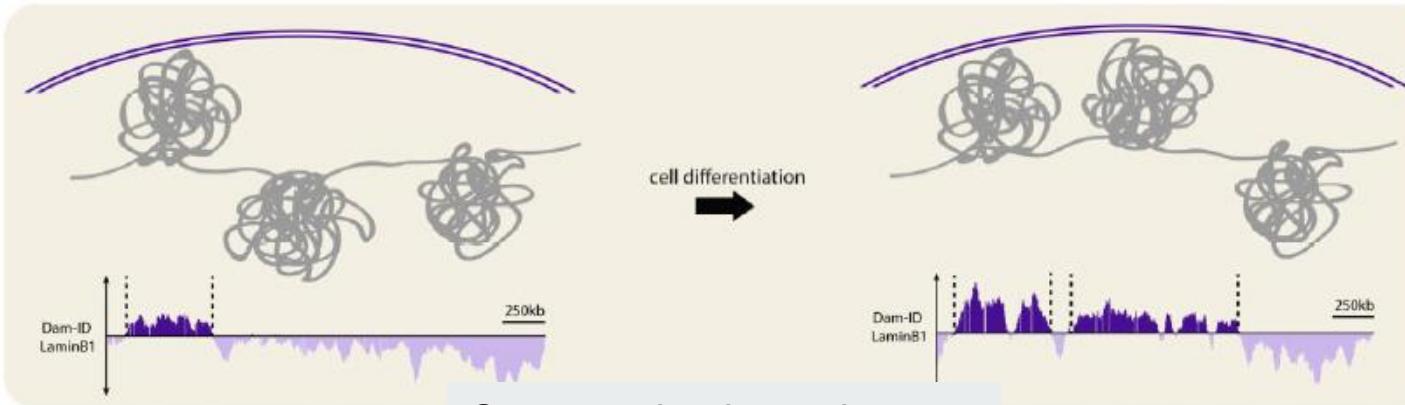


Szabo et al., Sciences (2019)

- TAD size ranges from 100kb to 10Mb

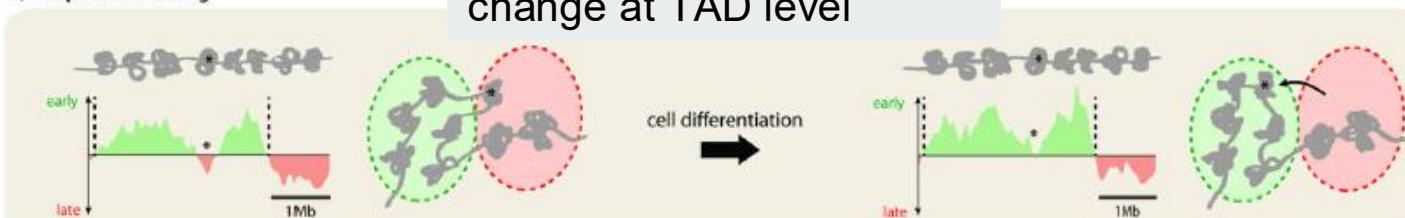
TAD as a functional unit of chromatin structure

D) Lamina association



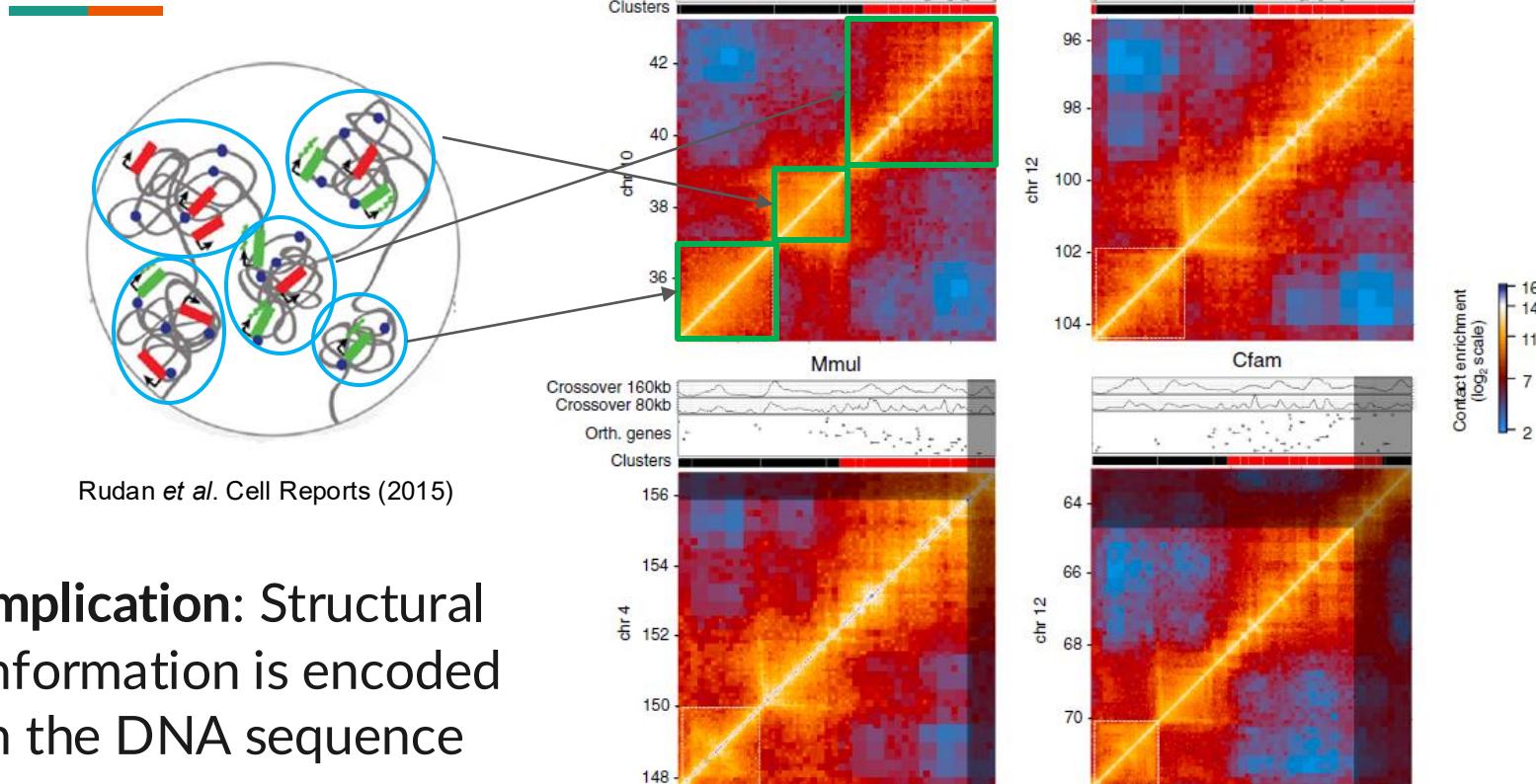
Concerted epigenetic state
change at TAD level

E) Replication Timing



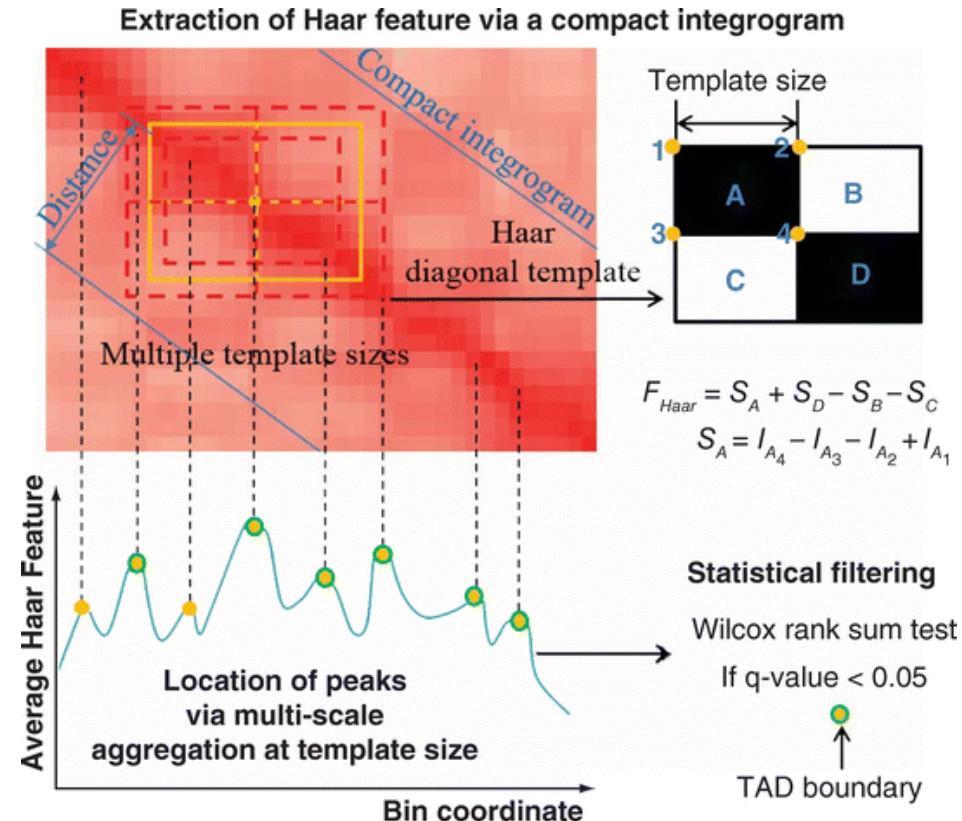
Source: Nora et al. Bioessay (2013)

Evolutionary conservation of TAD

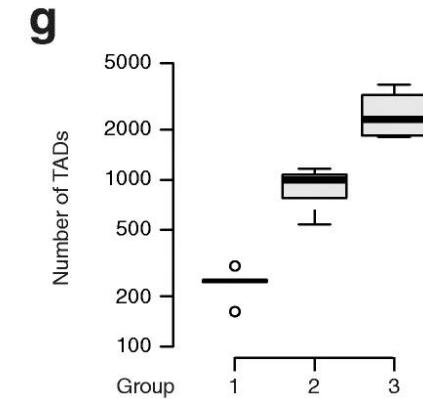
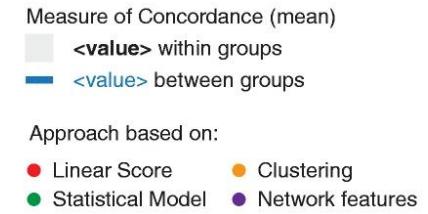
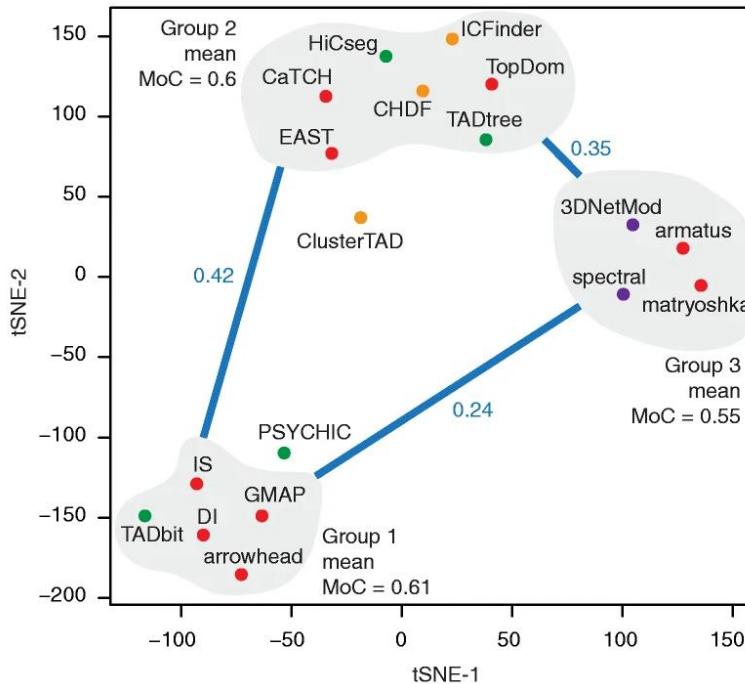
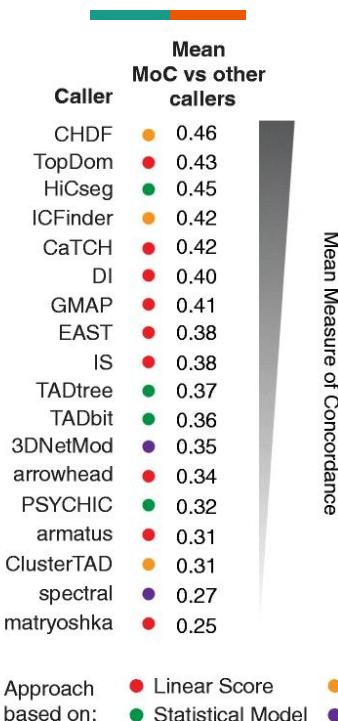


TAD identification from Hi-C data

- TAD boundary = **high interaction difference** between primary diagonal (A+D) and secondary diagonal (B+C) is maximized
- **Peak detection:** Mann-Whitney U test compared to scores from random positions



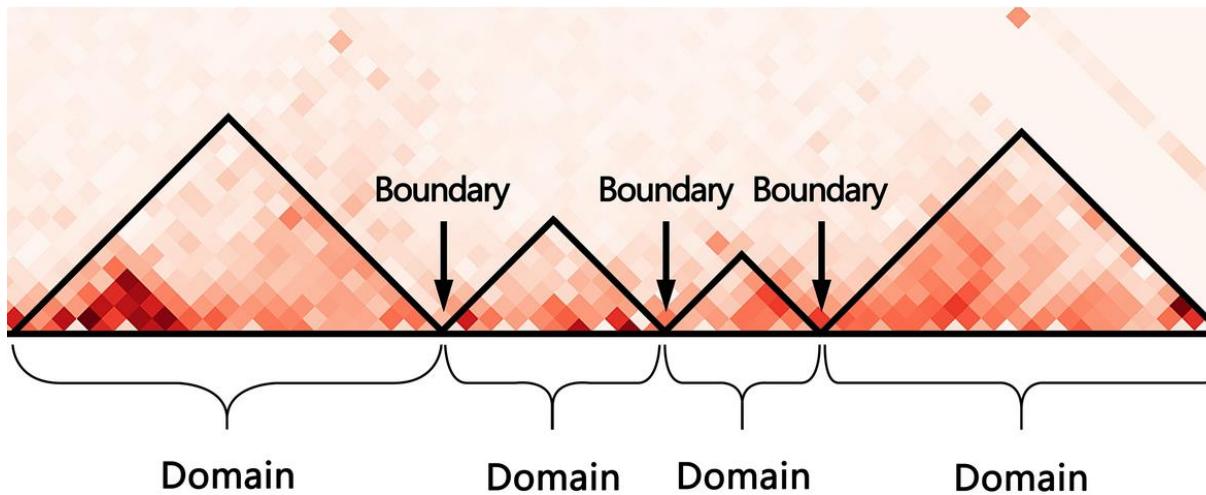
Variety of TAD algorithms



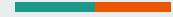
- Many TADs are not clearly delineated, nested inside other TADs

Variety of TAD algorithms

Long, C. et al., Interdisciplinary Sciences: Computational Life Sciences 13:638-651 (2021)

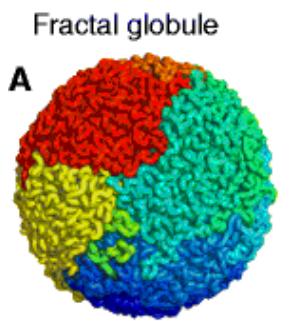


- TAD can exhibit hierarchical structure
- Not all algorithms allow nested or overlapping structure
- Caused by averaging TAD across multiple cells

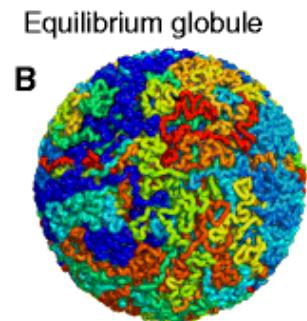


Structure models for chromatin

Early physics-inspired coarse models

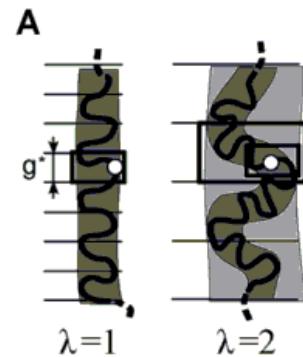


Fractal globule

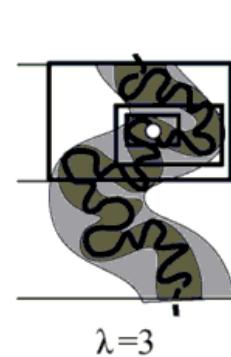


Equilibrium globule

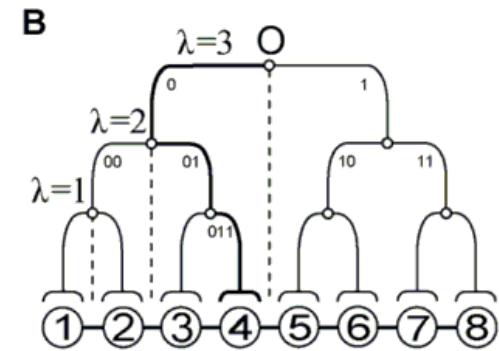
A



B



Mirny. Chromosome Research (2011)

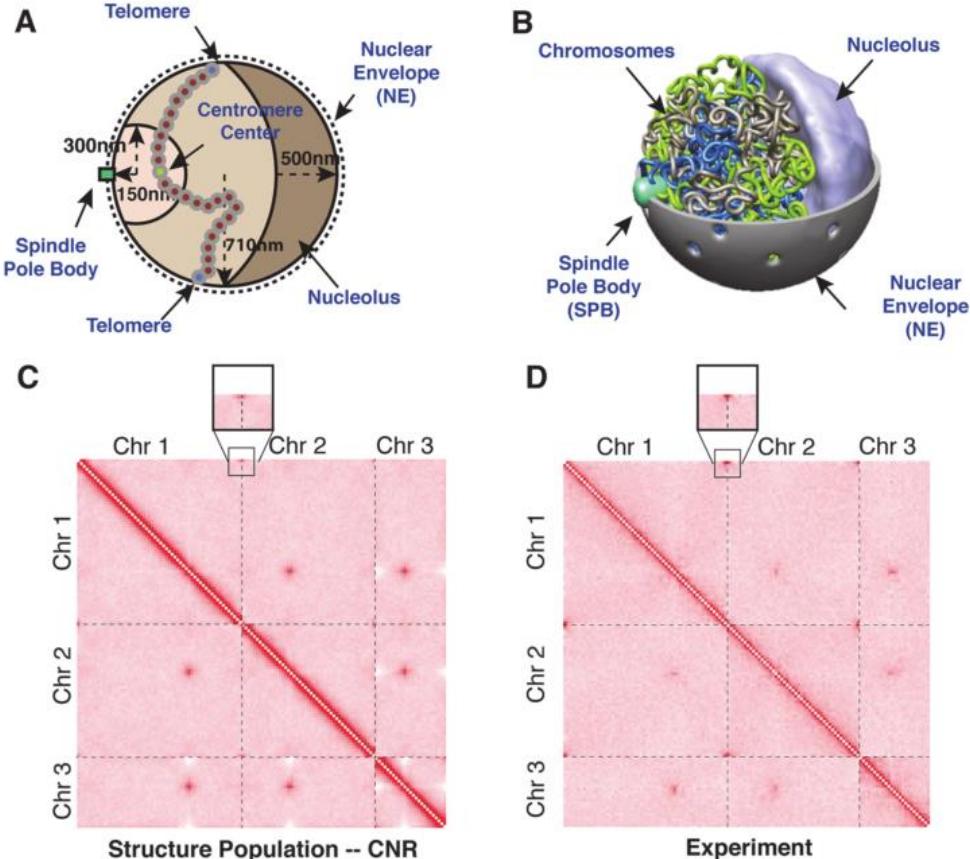


Nazarov et al. Soft Matter (2014)

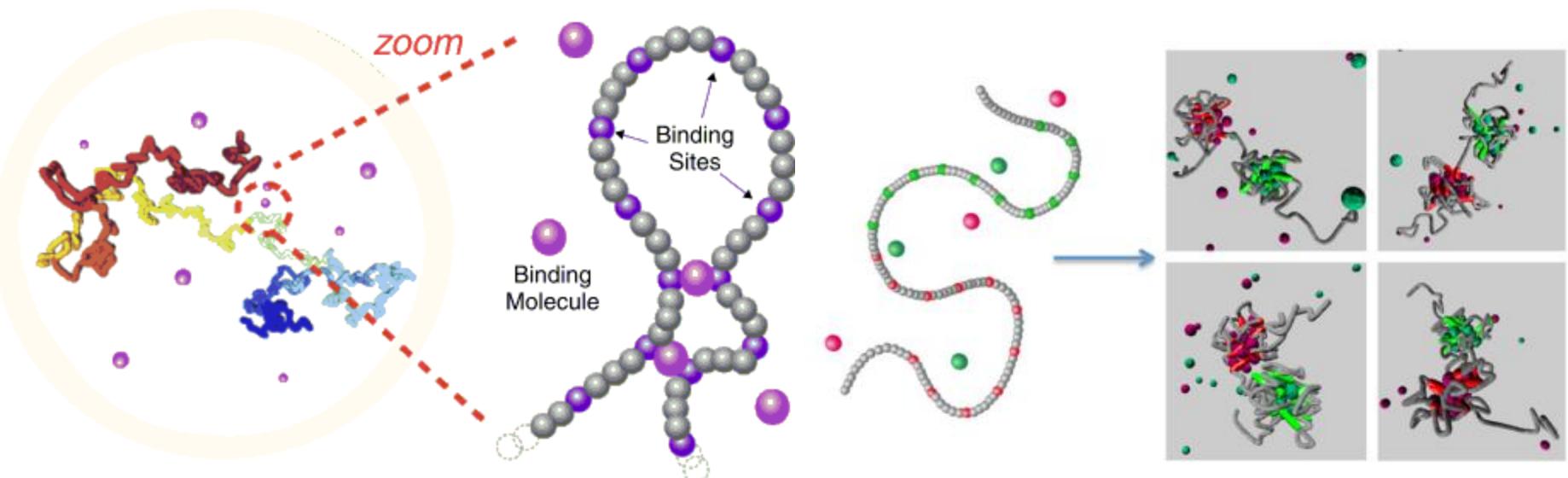
- Average properties of chromatin as a very long polymers
- Assume simplistic, regular folds
- Can already explain chromosome territory

Polymer simulation

- Control the relative positions of centromeres, telomeres, and rDNA
- Let the rest of chromosomes behave like flexible polymer chains
- Constrained by nucleus size



Strings and binders switch model

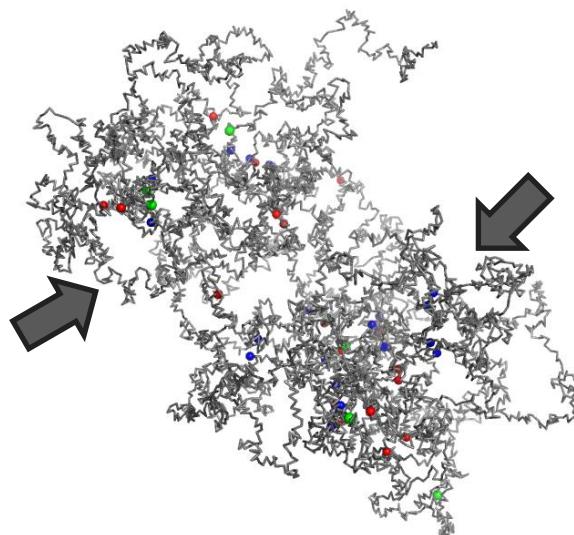


Nicodemi and Pombo. Curr Opin Cell Biol (2014)

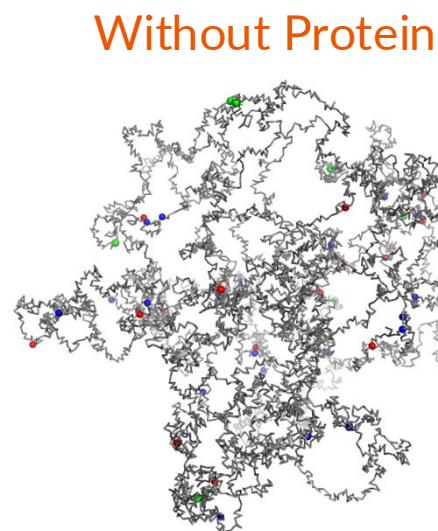
Barbieri, M. et al. PNAS (2012).

- Assumption: Protein molecules bind to multiple genomic loci at once

Strings and binders simulation

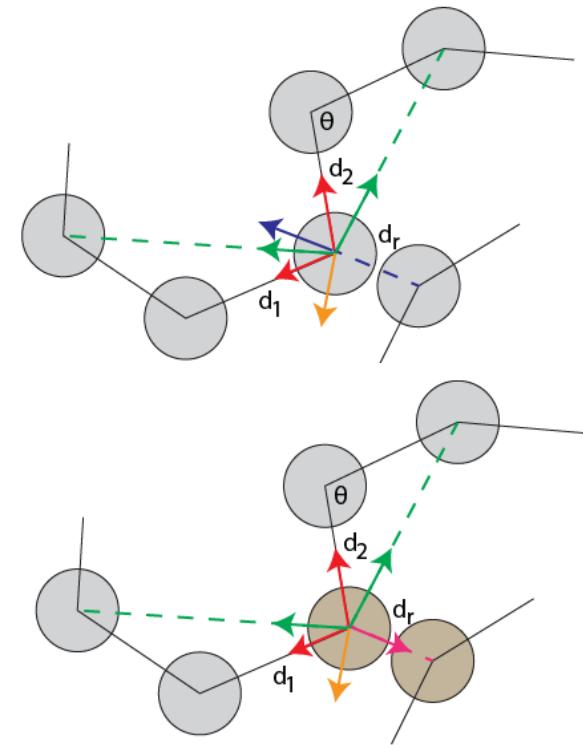


With Protein



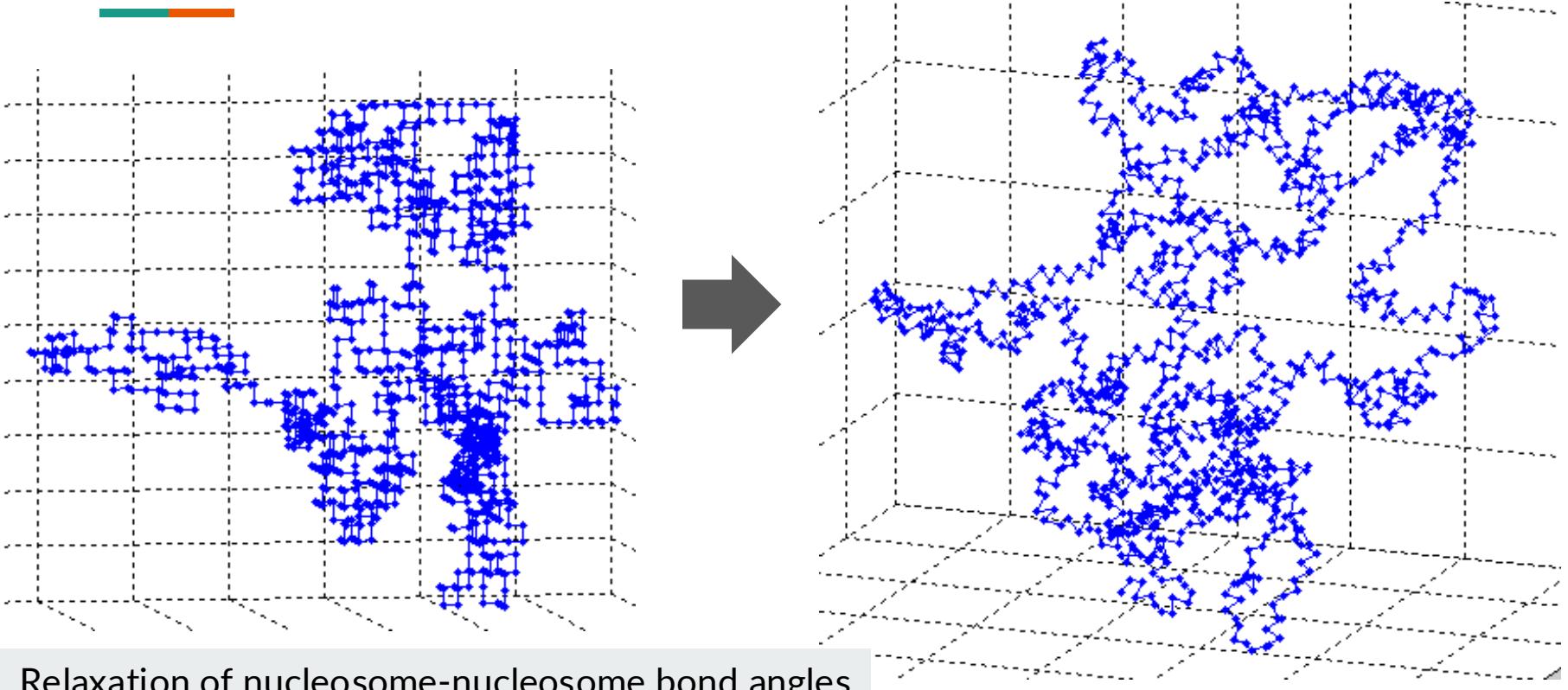
Without Protein

1 Mb represented by a 7000-unit polymer chain
Each unit = 150 bp ~ 1 nucleosome

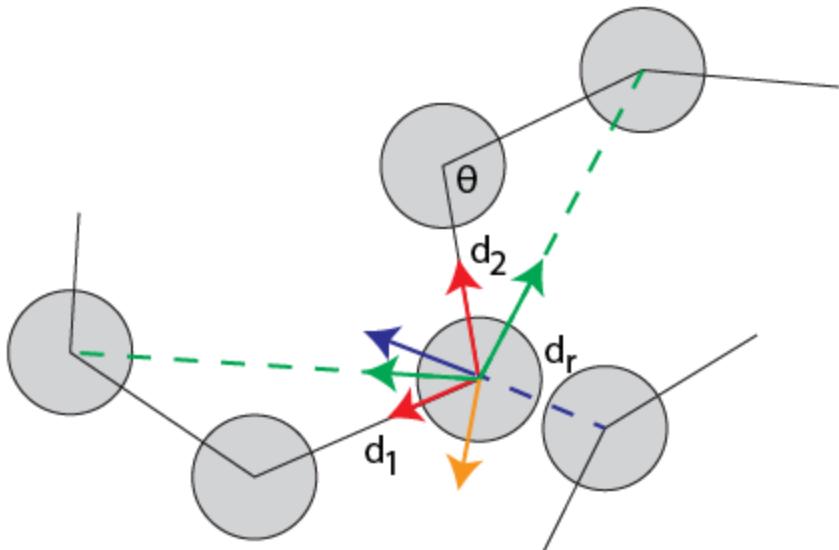


- Regular unit
- Protein binding site

Initialization with self-avoiding random walk



Physical forces involved in chromatin folding

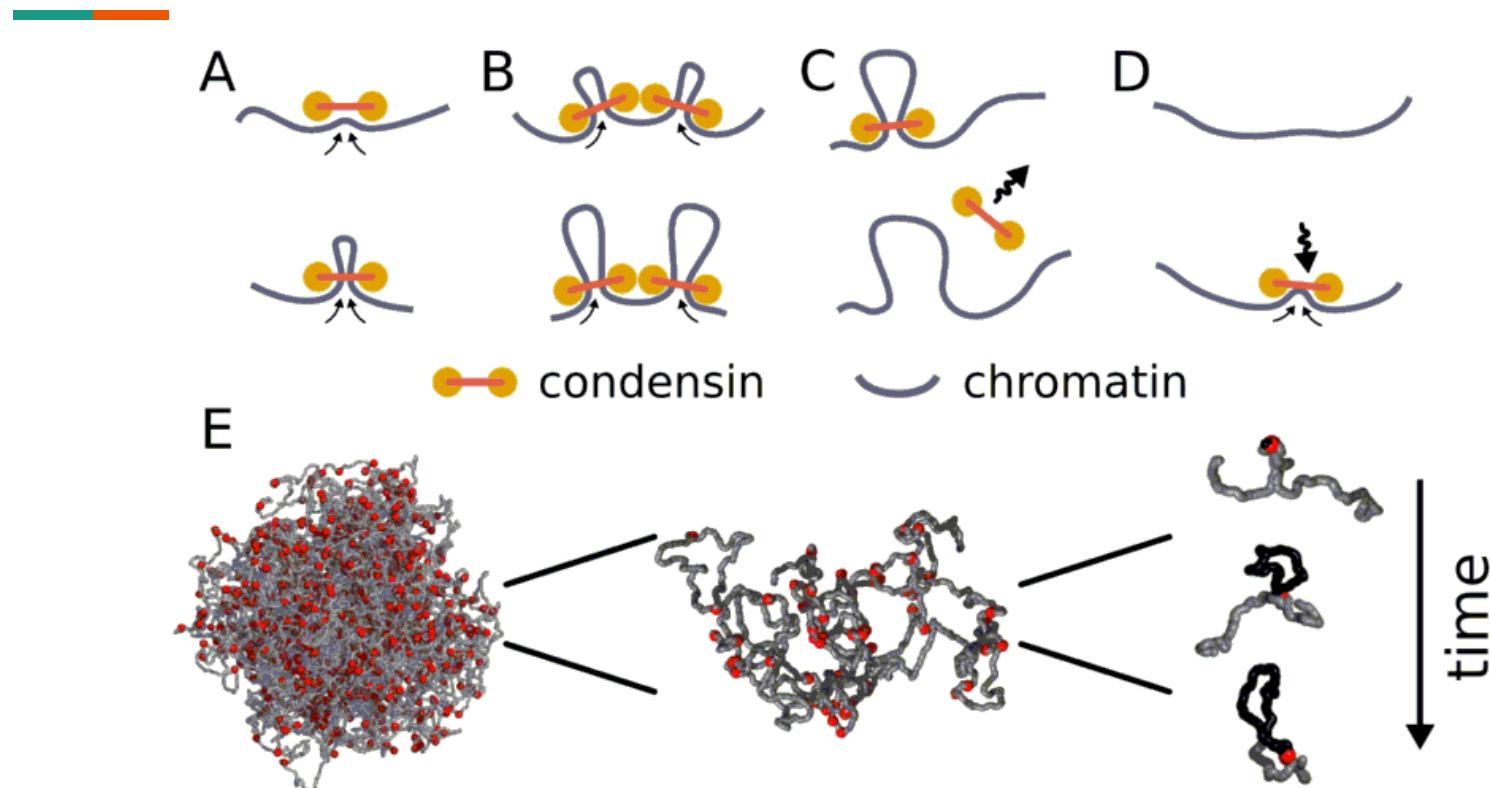


- **Entropic** (diffusional) = constant force with random direction
- **Tension** (bead-bead) = spring force between adjacent beads
- **Repulsion** = dispersion force between close-by beads
- **Attraction** = artificial force to control the distribution of θ



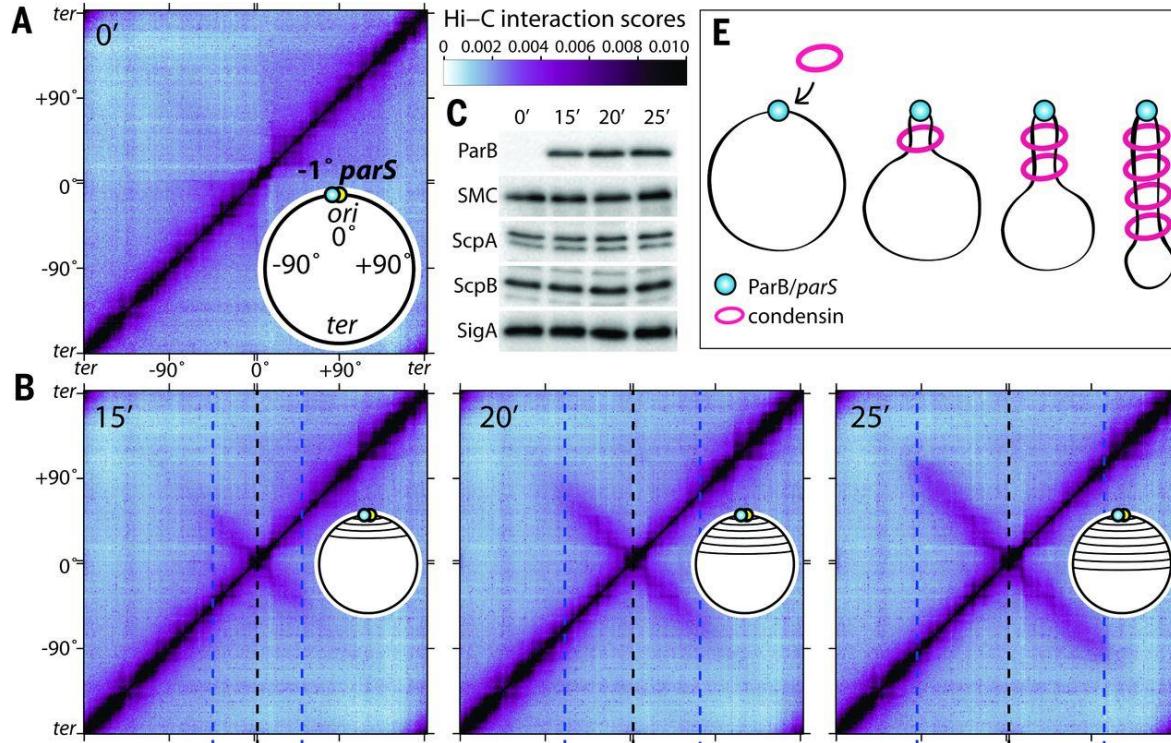
Loop extrusion model

Loop extrusion by proteins



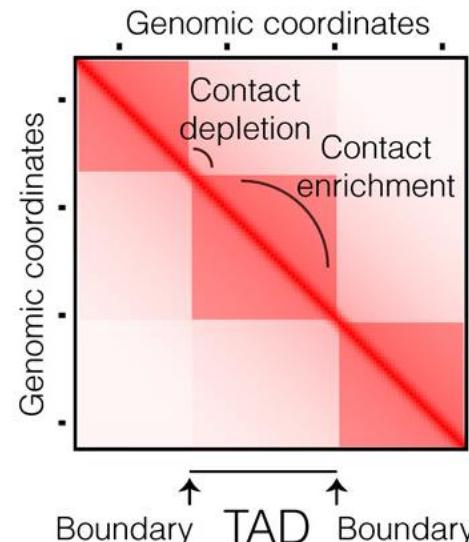
Live loop extrusion in a bacteria

Wang et al. Science 355:524-527 (2017)

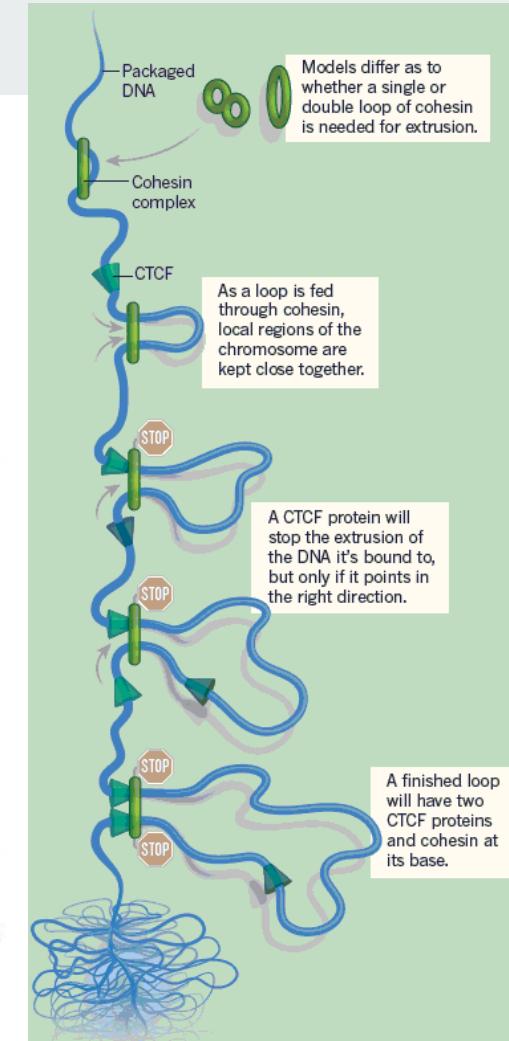


Loop extrusion and CTCF

- Proteins like cohesin and condensin extrude loops
- How to stop at TAD boundary?
- CTCF proteins can stop extrusion process in a specific direction
- Not every species has CTCF

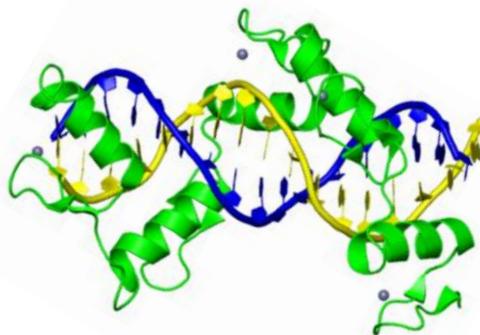


Dolgin. Nature (2017)

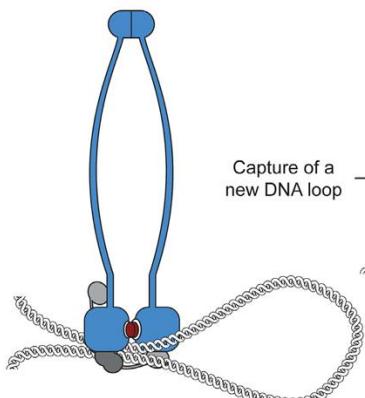


Ingredients of TAD formation

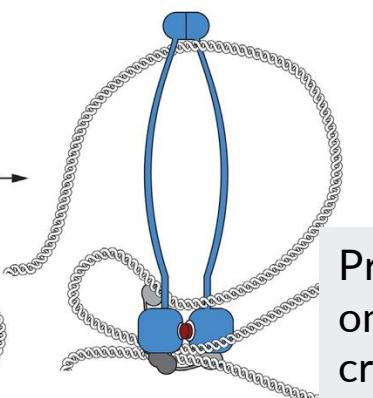
CTCF in complex
with DNA



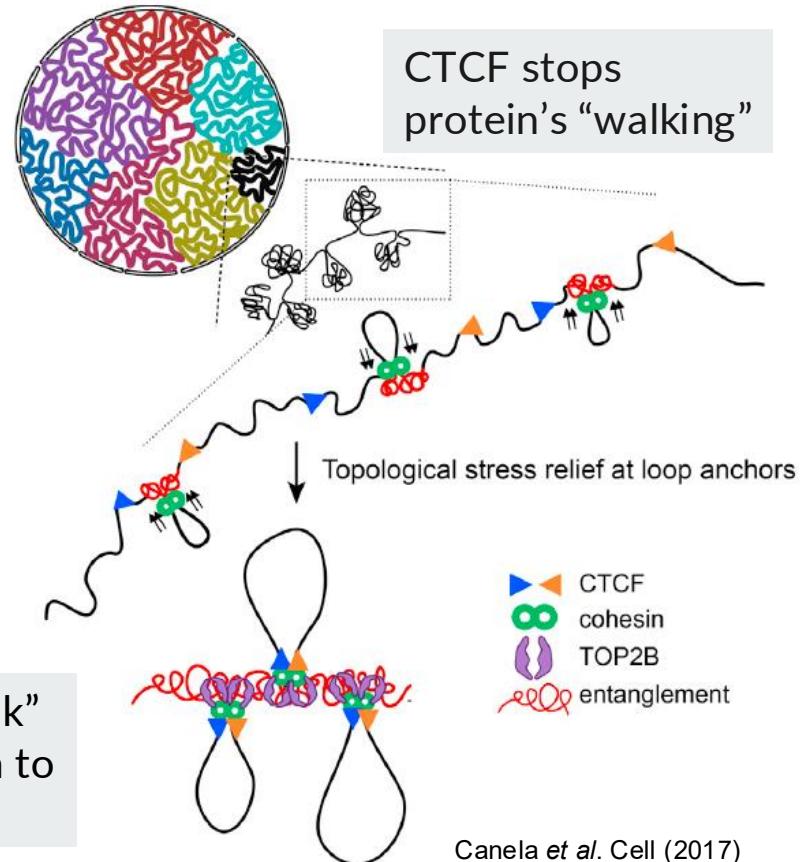
Cohesin



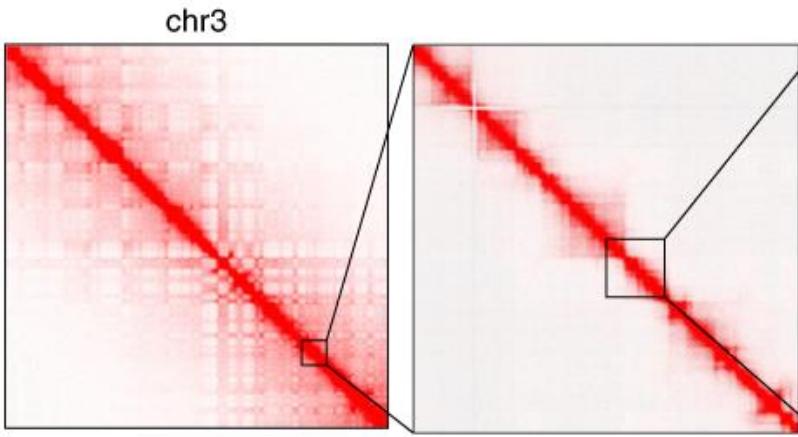
Condensin



Proteins “walk”
on chromatin to
create loop

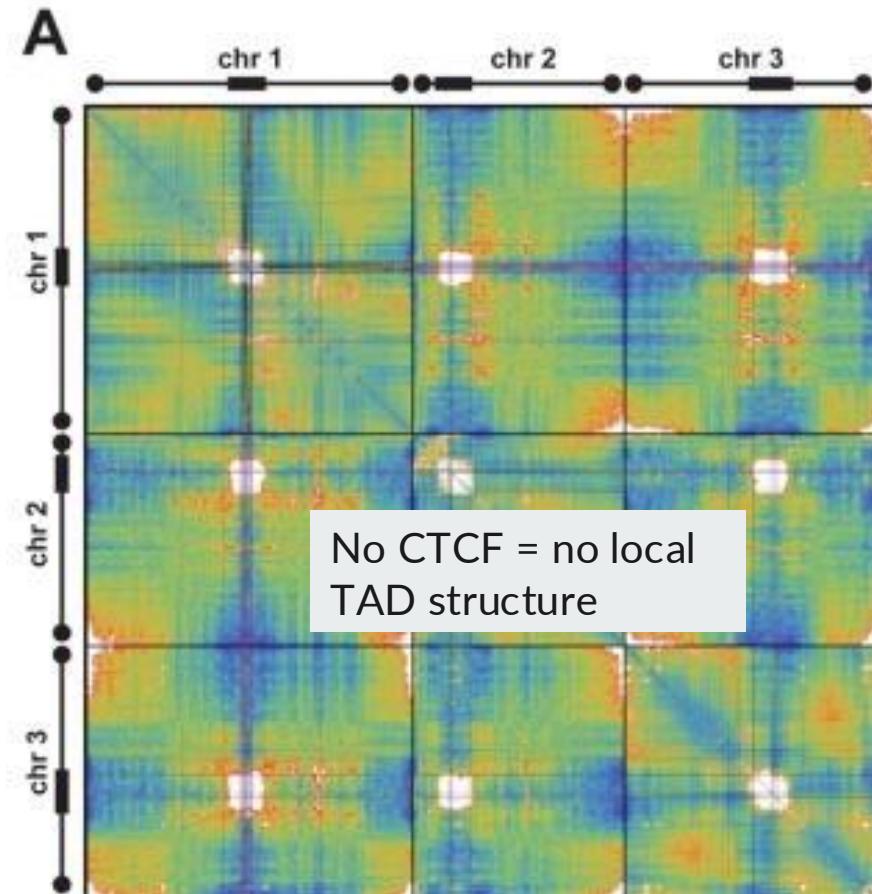


Human vs arabidopsis



Li et al. Nat Methods 16:991-993 (2019)

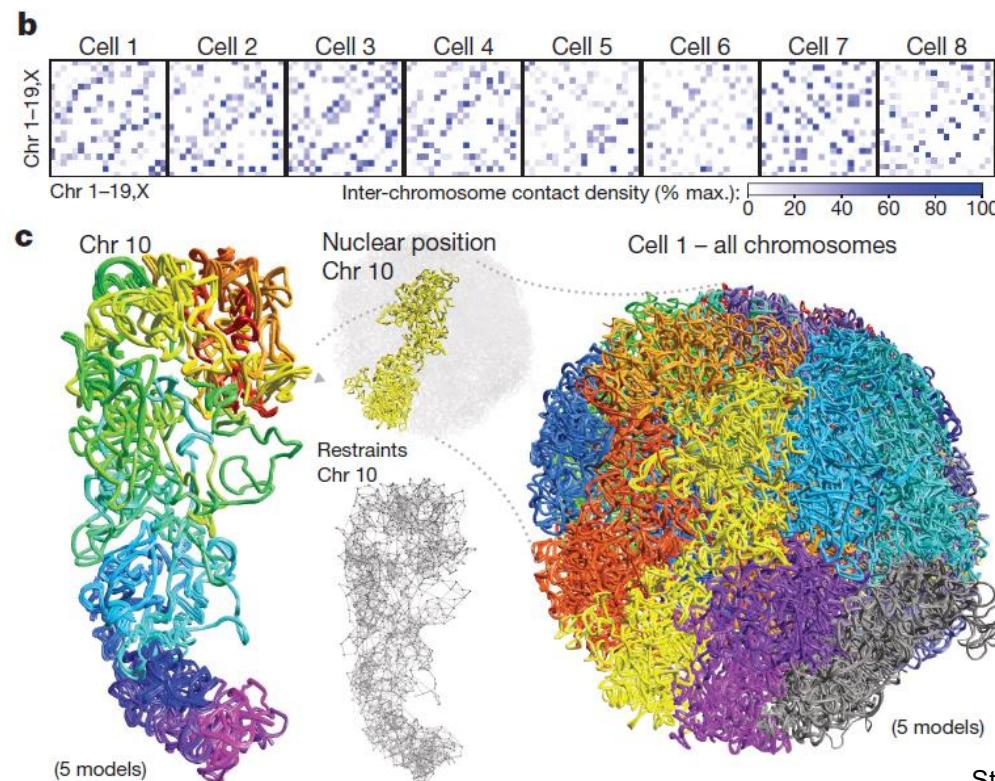
- CTCF is responsible for organized TADs in human genome



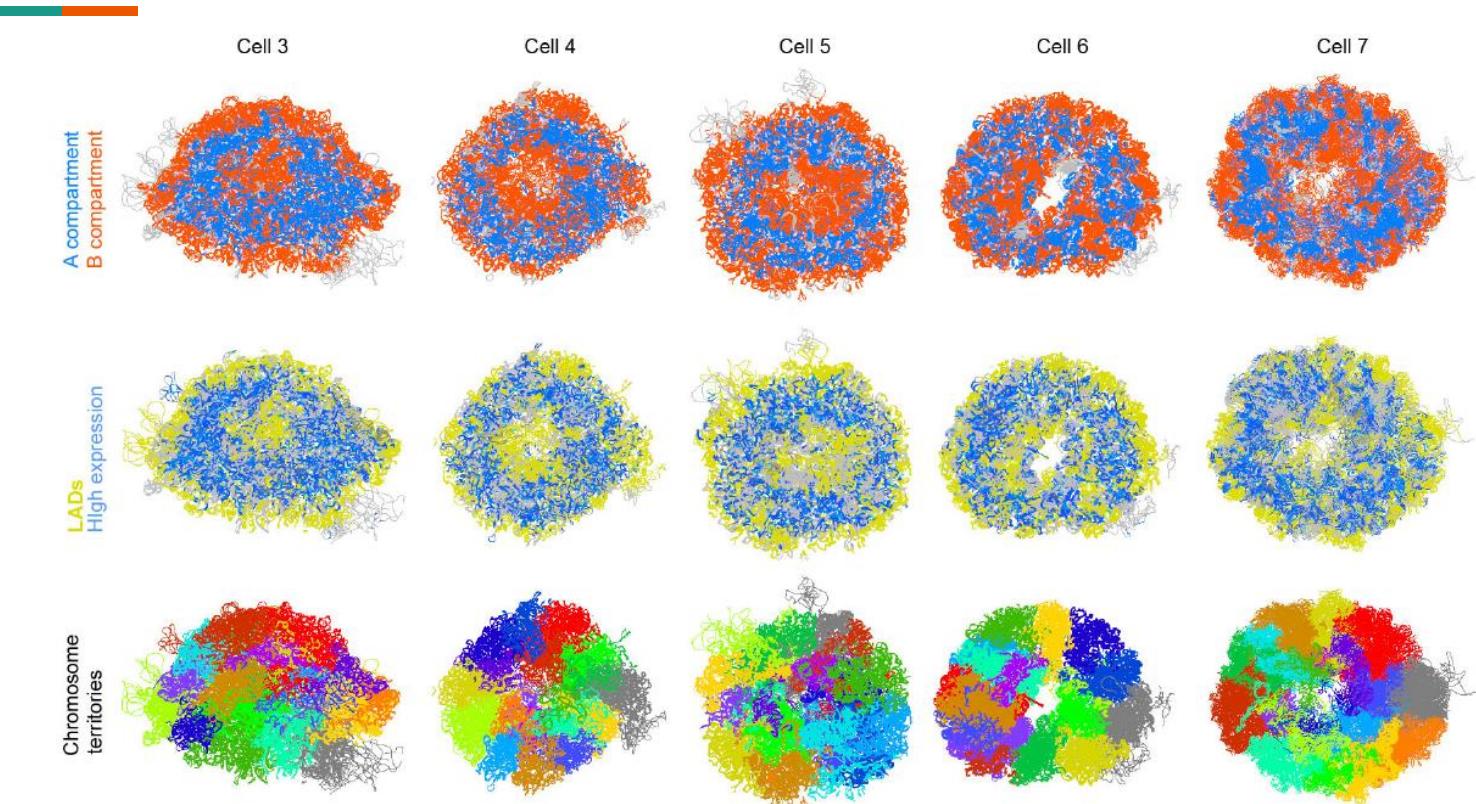


Cell-cell variation of chromatin structure

Single-cell Hi-C of mouse cells

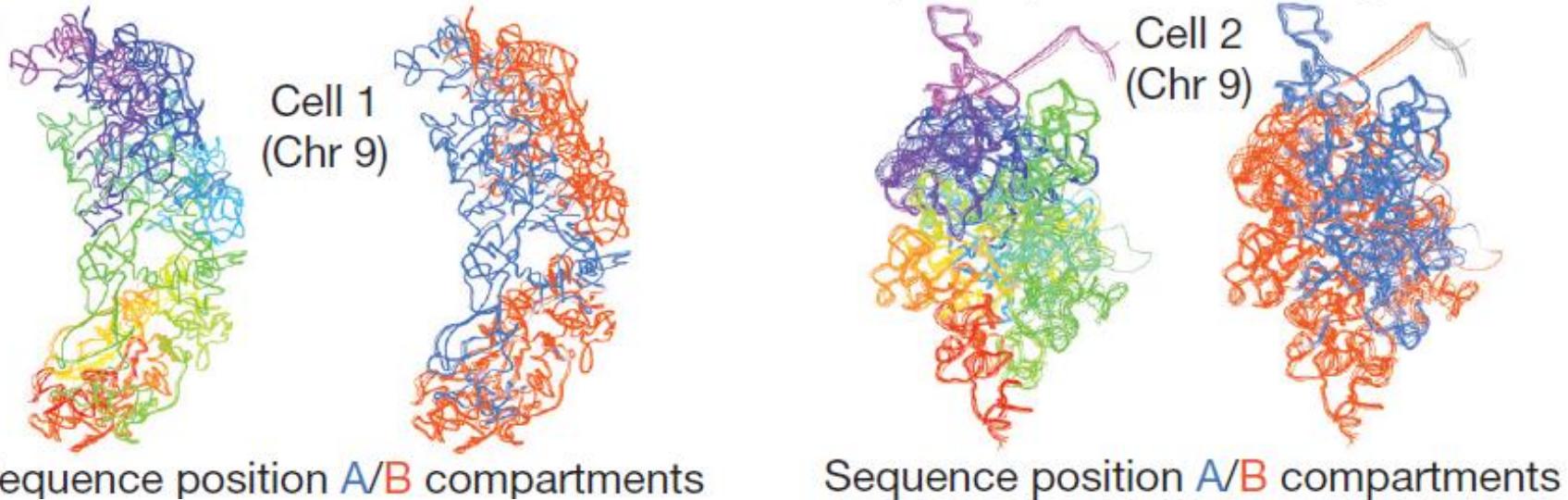


Some conservation across cells



Cell-to-cell variations

d



- 3D structure may not be the same across cells, but the chromatin state (A/B compartment) and TAD formation are somewhat similar

Summary

- Functional roles of chromatin
 - DNA packaging
 - Transcriptional regulation
 - Dynamics during development and cell division
- Chromatin conformation capture assays for studying chromatin structure
- Structural models for chromatin
 - Loop extrusion model
- Cell-to-cell structure variation with conserved states

Any question?

- See you next time