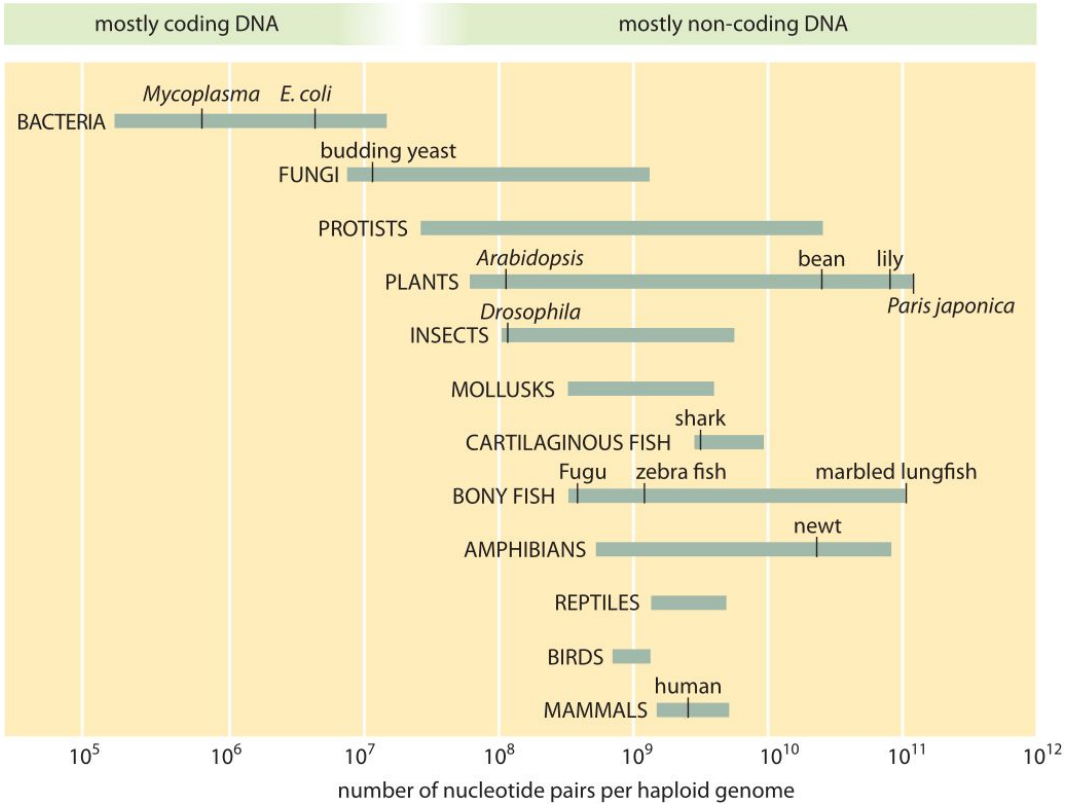
Hi-C method. Chromatin structure.

December 18th 2023.

Anna Litvin, Skoltech, MSU.
litvinannaema@gmail.com
Credits to Ilya Pletnev, Skoltech

Amount of DNA vary in organisms



E.coli – 5.0 Mb in one cell;

Humans — 6.27 Gb in one diploid cell;

http://book.bionumbers.org/how-big-are-genomes/

Why to have chromatin in eukaryotes

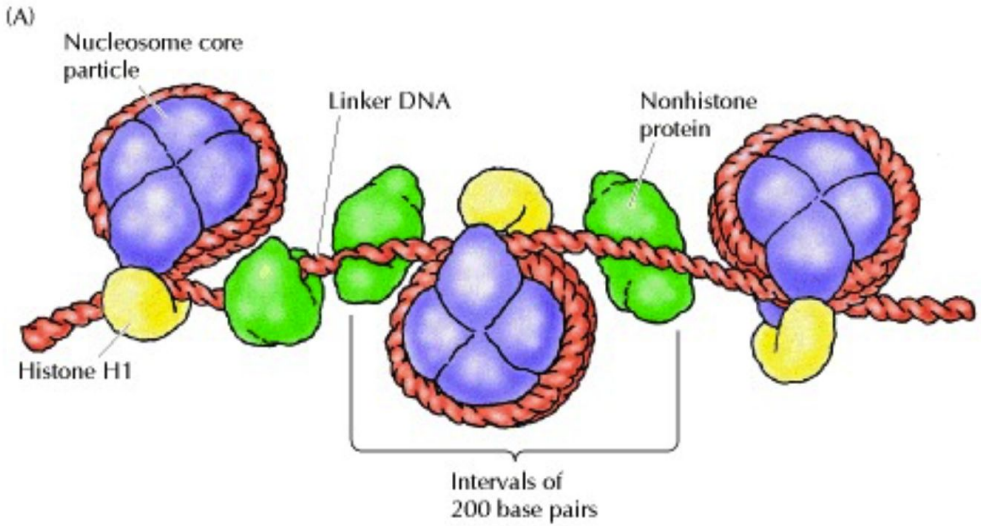
Human cell: 6.27 Gb of DNA = 2 meters of DNA
Human nucleus: 10 micrometers (μm, 10^{-6} m) in diameter

~~How to fit structurally?~~
How to function? How to regulate?

- transcription factors access
- proper replication
- silencers and enhancers access

Pal K, Forcato M, Ferrari F. Hi-C analysis: from data generation to integration. *Biophys Rev*. 2019;11(1):67-78. doi:10.1007/s12551-018-0489-1

Lowest level — nucleosomes



Nucleus mass = 1 part DNA + 1 part histones + 1 part other proteins

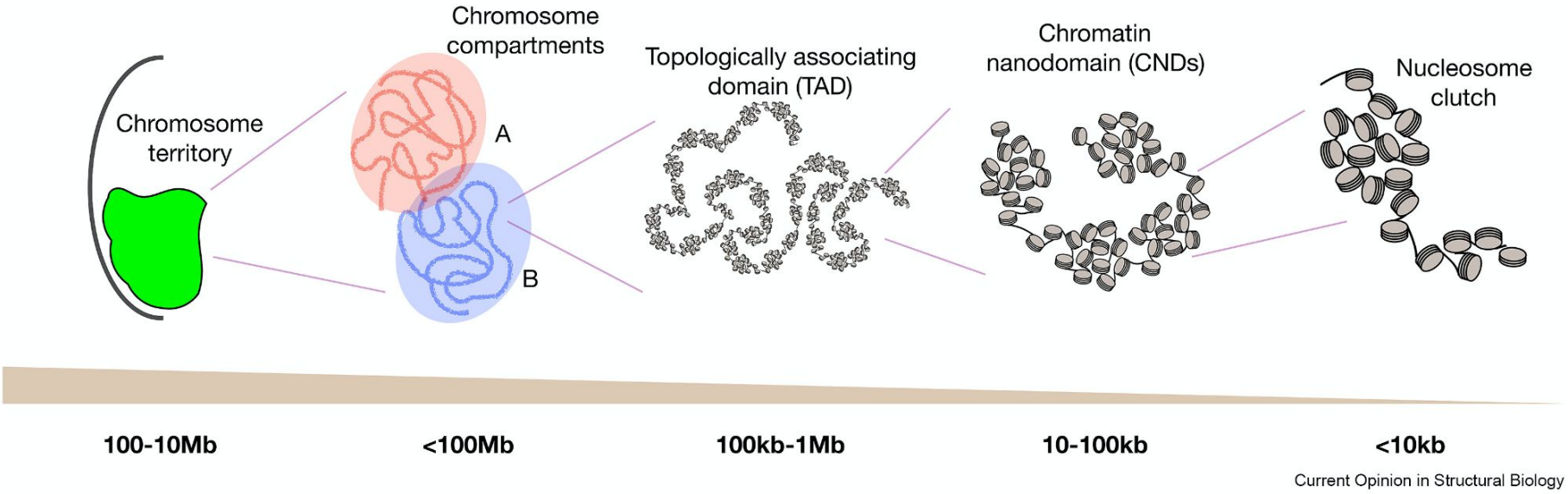
Histones are found in all eukaryotes and even archaea.

→ 10 nm thread of beads

→ 30 nm supercoling??

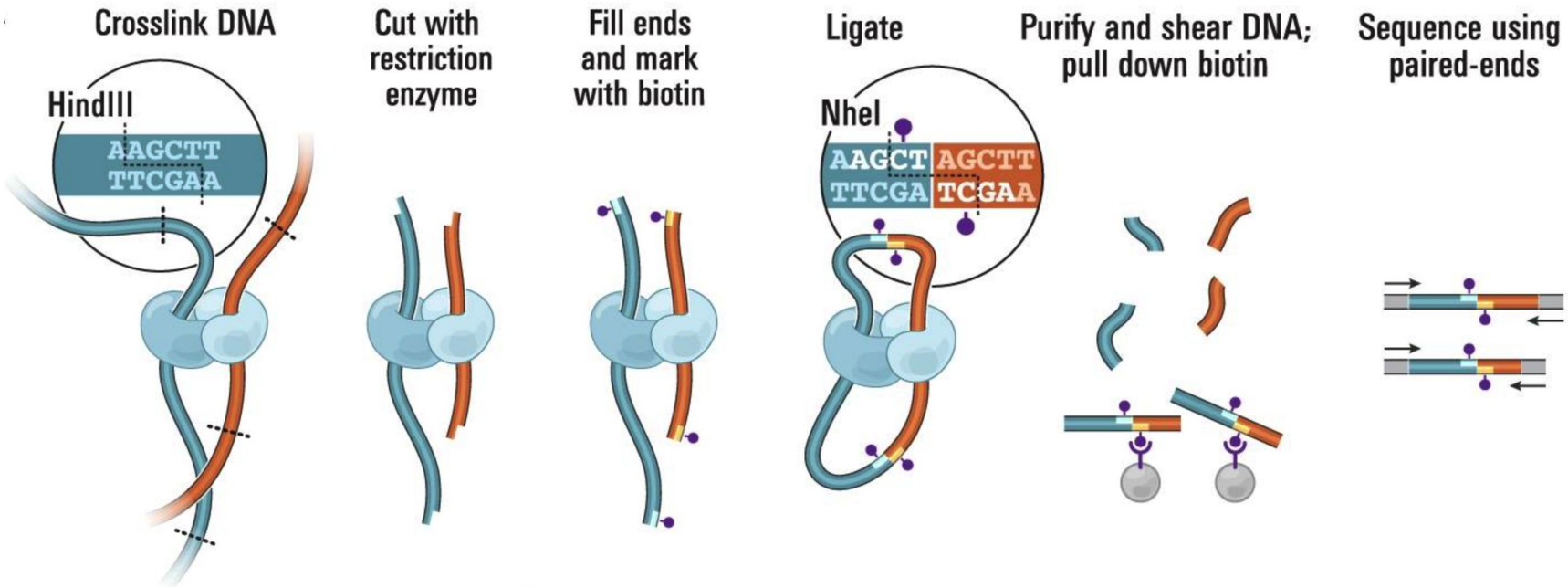
Cooper GM. The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates; 2000. Chromosomes and Chromatin. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9863/

Chromatin in human nucleus



Di Stefano M, Cavalli G. Integrative studies of 3D genome organization and chromatin structure. *Curr Opin Struct Biol*. 2022;77:102493. doi:10.1016/j.sbi.2022.102493

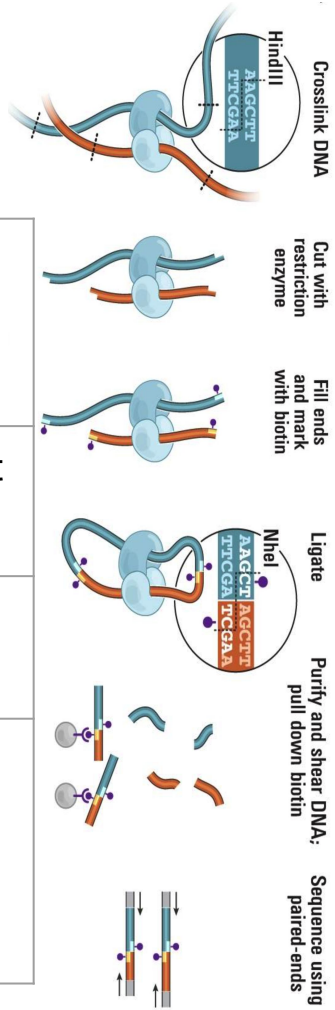
Hi-C method



Lieberman-Aiden E, van Berkum NL, Williams L, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289-293. doi:10.1126/science.1181369

Hi-C method steps

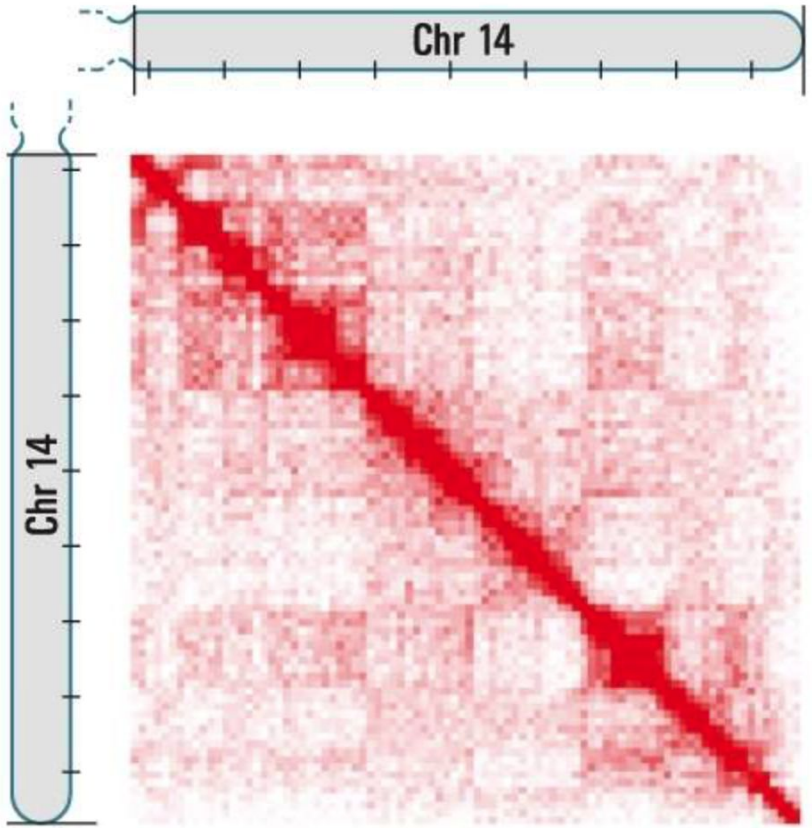
Crosslink chromatin material using formaldehyde	Fix the chromatin state that we want to explore
Fragment, dilute and ligate	Create chimera DNA pieces from DNA fragments that were close to each other in space
Shear DNA and pull the biotin	Isolate fragments with ligated DNA — target pieces (magic with restrictases)
Sequence DNA pieces and understand where they come from	If they come from different genome regions probably they were close due to functional reasons, or not? (here we need statistics)



Original paper

“**We created a Hi-C library** from a karyotypically normal human lymphoblastoid cell line (GM06990) and sequenced it on two lanes of an Illumina Genome Analyzer, **generating 8.4 million read pairs** that could be uniquely aligned to the human genome reference sequence; of these, **6.7 million corresponded to long-range contacts between segments greater than >20 Kb apart**.”

One hit — one chimera DNA piece consisting of two DNA fragments mapped to the genome.



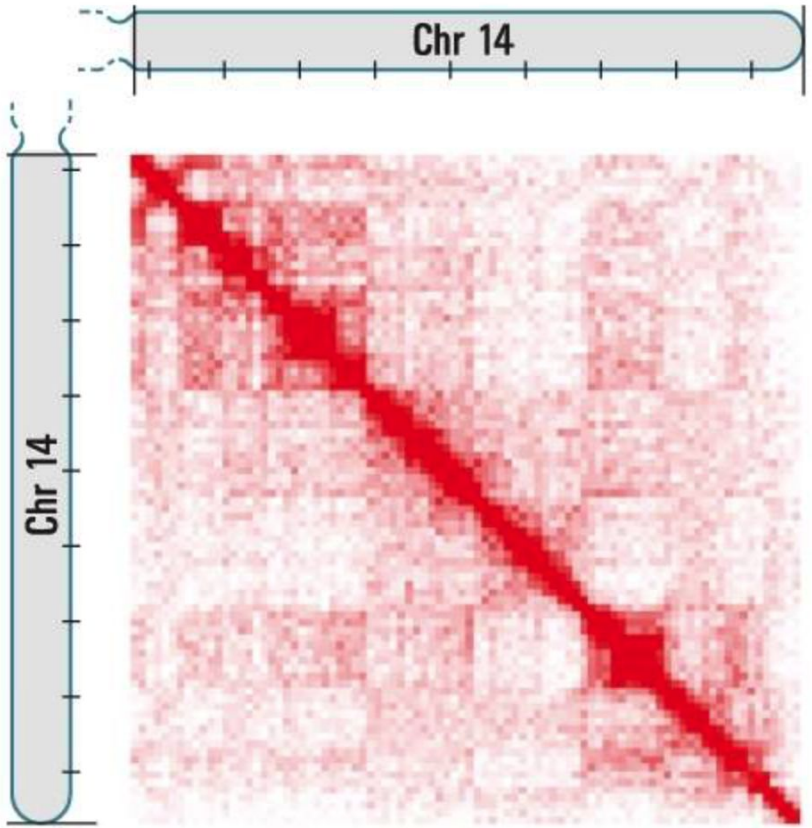
Lieberman-Aiden et al, Science, 2009.

Heatmap for contacts

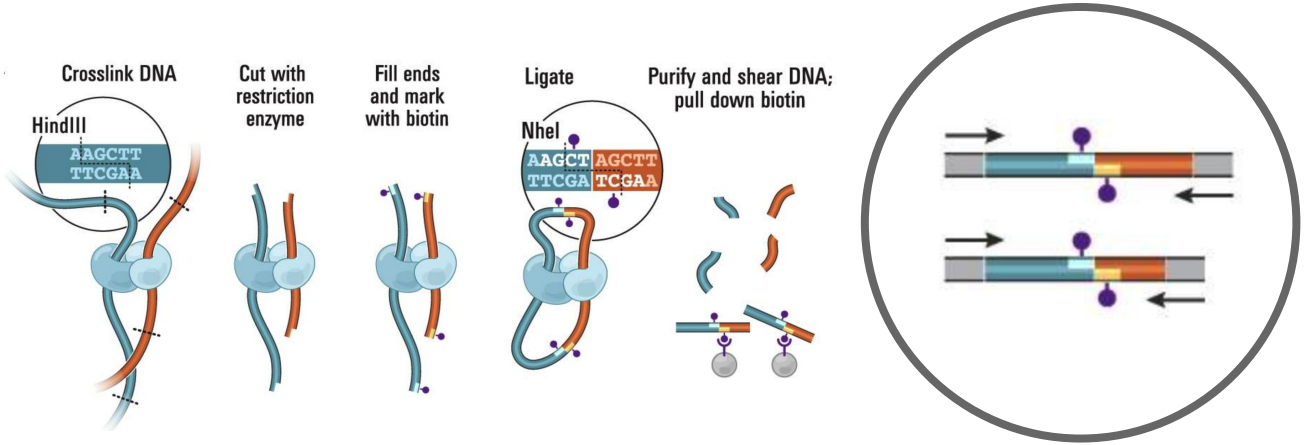
Both x and y axes correspond to the same genome fragment (chromosome 14 in our case).

The sequence is binned into smaller pieces (bins, here 1 Mb).

The color of each square shows the number of contacts between corresponding bins = small sequence fragments.

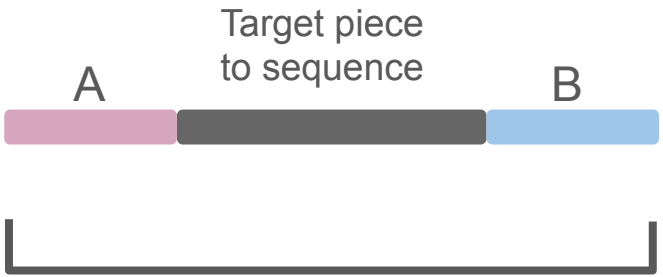


How to get a heatmap from reads?

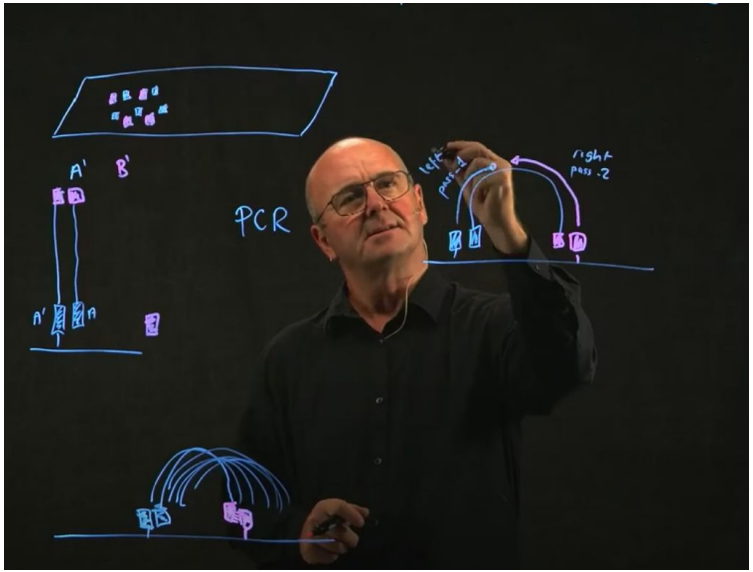


1. Reads are paired. We have only sequenced the ends of the pieces.
2. Each read pair is mapped to the genome — which regions interact.
3. Determine resolutions we want to achieve → select bin size → count interactions in each bin.

Pair end reads

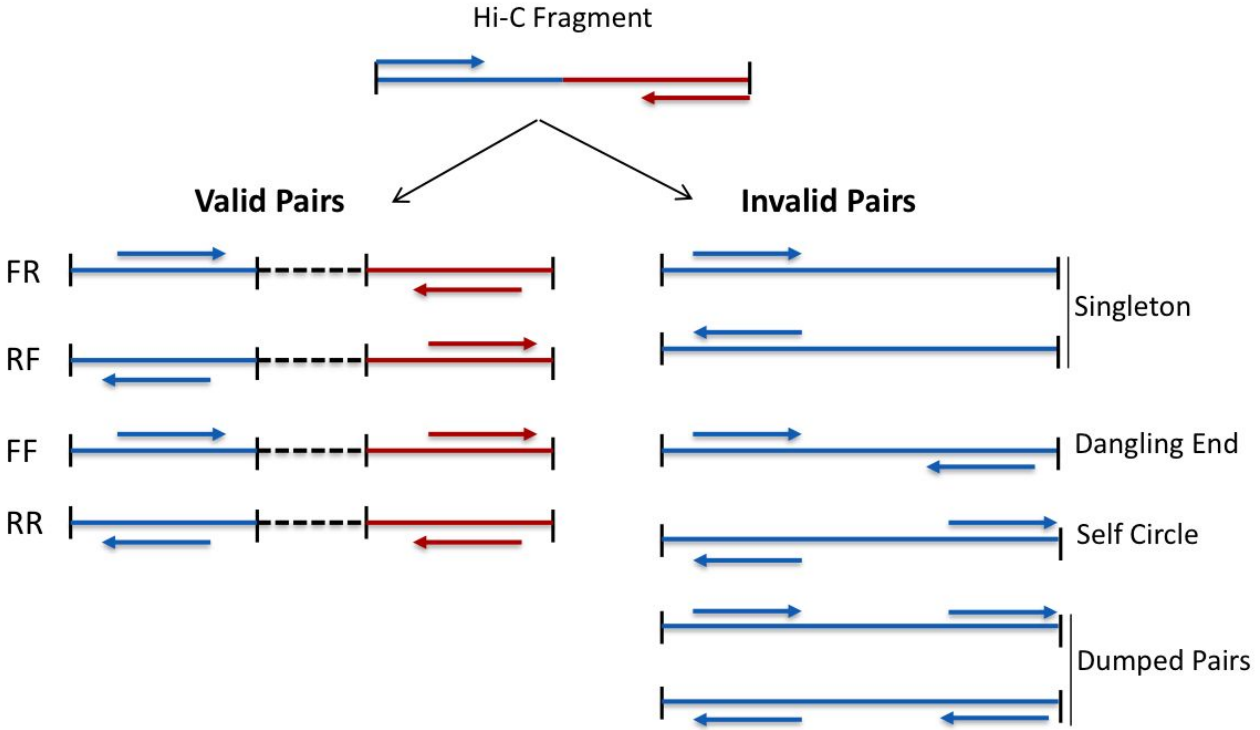


A and B — special different adapters



https://www.youtube.com/watch?v=WneZp3fSJIk

Reads in Hi-C

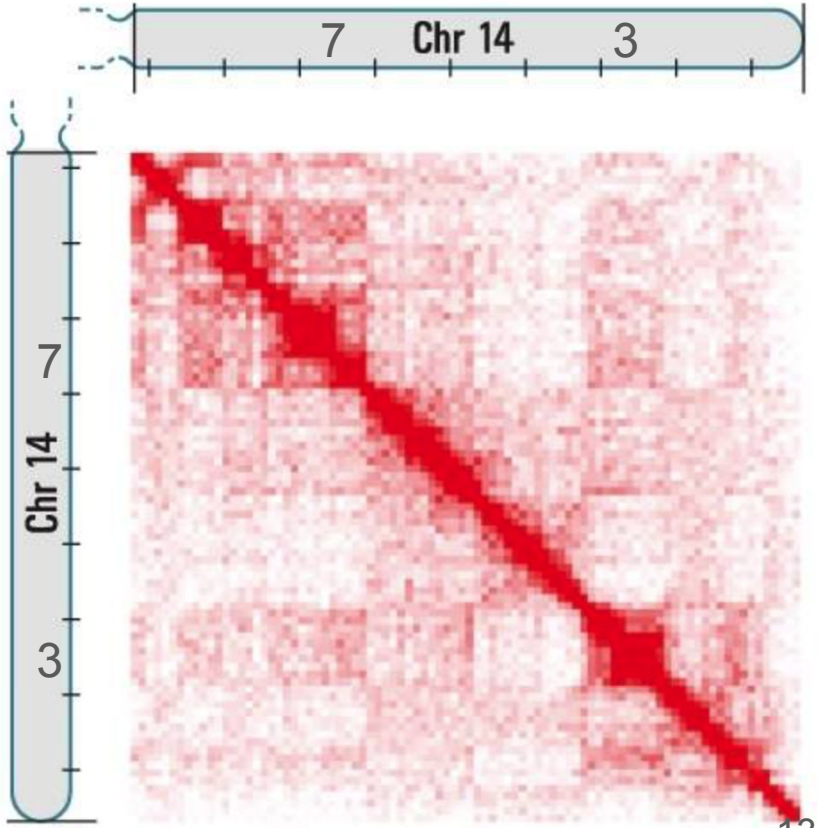


https://mathforgenomics.github.io/servant.pdf

Data analysis. Significance.

Why is diagonal so bright?

What do we expect from the experiment?



Chromosome territories

FISH — **f**luorescent **i**n **s**itu **h**ybridization.

Or in situ hybridization with other dyes.

(F) Simultaneous delineation of all chromosomes in a human fibroblast nucleus (left) and a prometaphase rosette (right) by multi-color FISH.

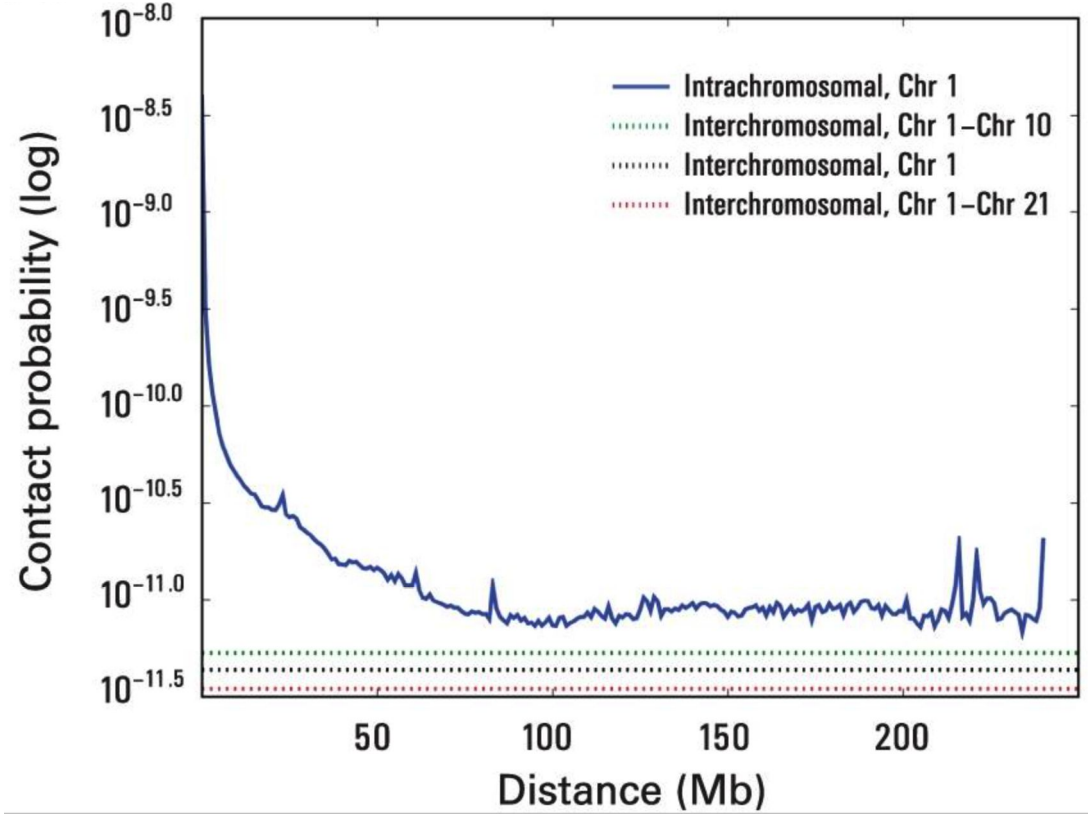
Cremer T, Cremer M. Chromosome territories. Cold Spring Harb Perspect Biol. 2010;2(3):a003889. doi:10.1101/cshperspect.a003889

Chromosome territories with Hi-C

Existence of chromosome territories — probability of intrachromosome (inside) contact is lower than probability of interchromosome (between) contact.

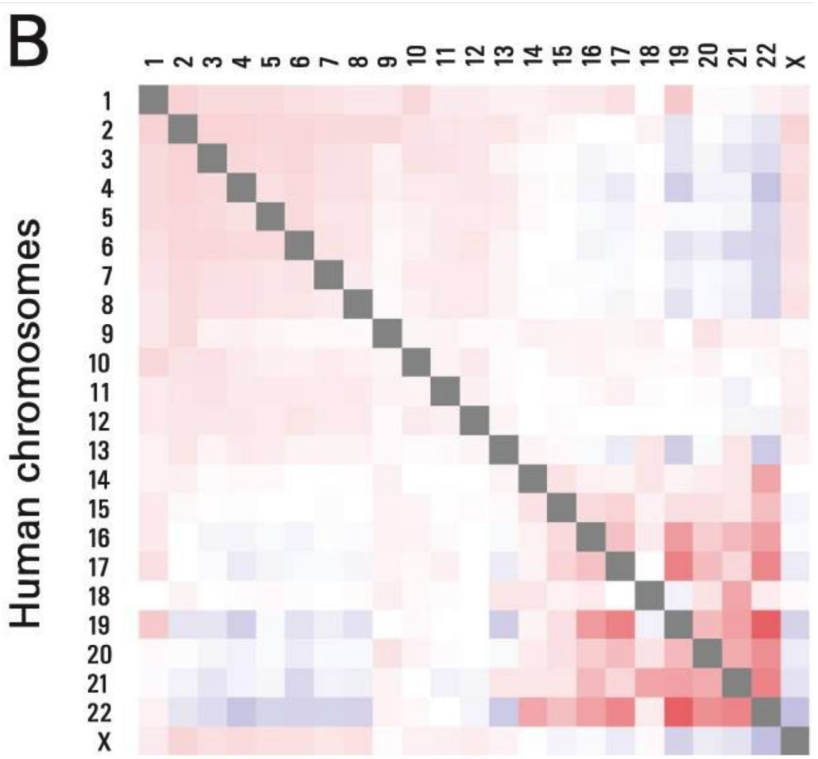
H0: all chromosomes are intertwined into one big lump and interactions between them are equally like.

H1: H0 is not true.



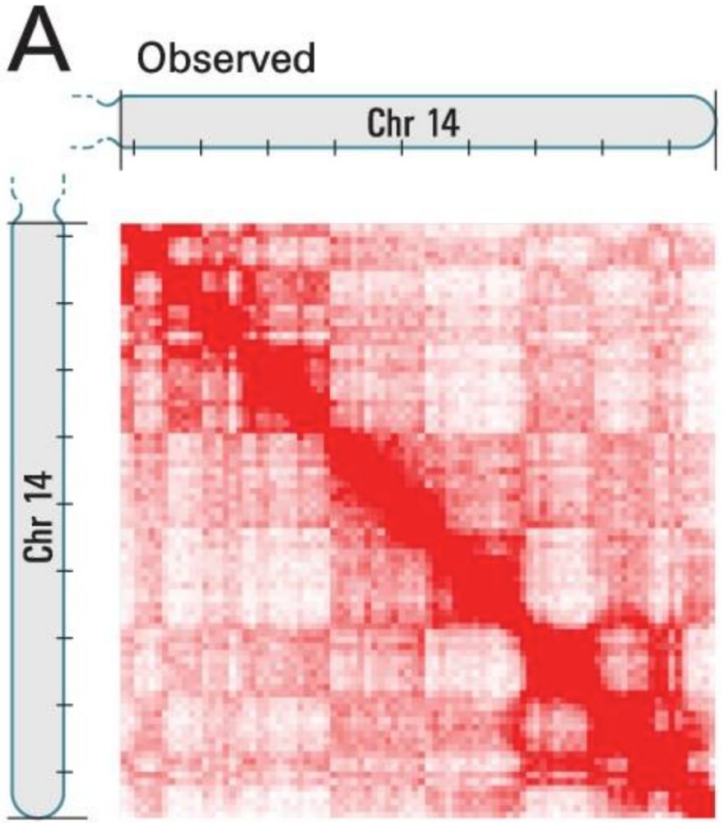
Chromosome territories with Hi-C

(B) Observed/expected number of interchromosomal contacts between all pairs of chromosomes. Red indicates enrichment, and blue indicates depletion (up to twofold).



Inside one chromosome

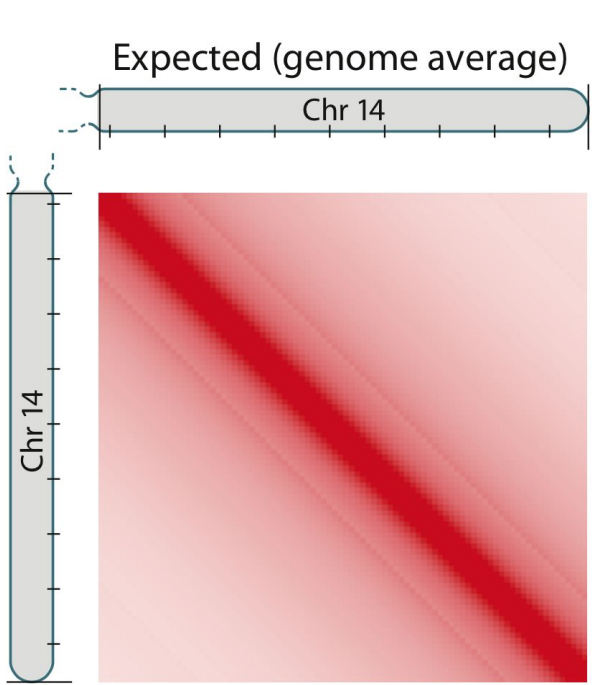
(A) Map of chromosome 14 at a resolution of 1Mb (1 tick mark = 10Mb) exhibits substructure in the form of an intense diagonal and a constellation of large blocks (range: 0-200 reads).

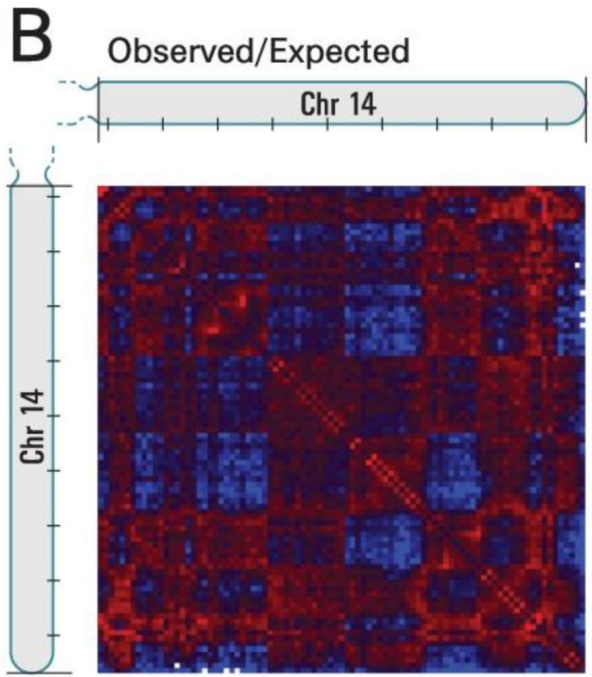


Lieberman-Aiden et al, Science, 2009.

Expected number of contacts

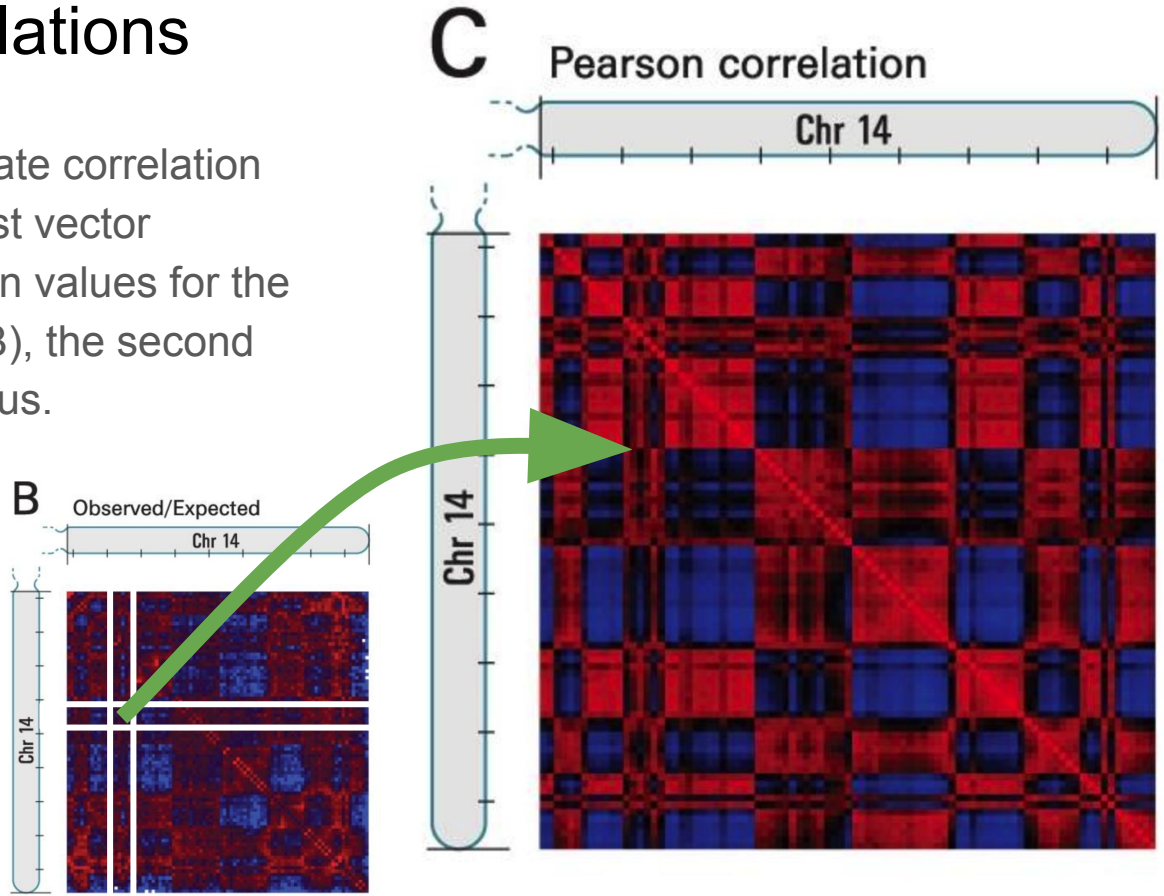
(B) The observed/ expected matrix shows loci with either more (red) or less (blue) interactions than would be expected given their genomic distance (range: 0.2 – 5).





Picture with correlations

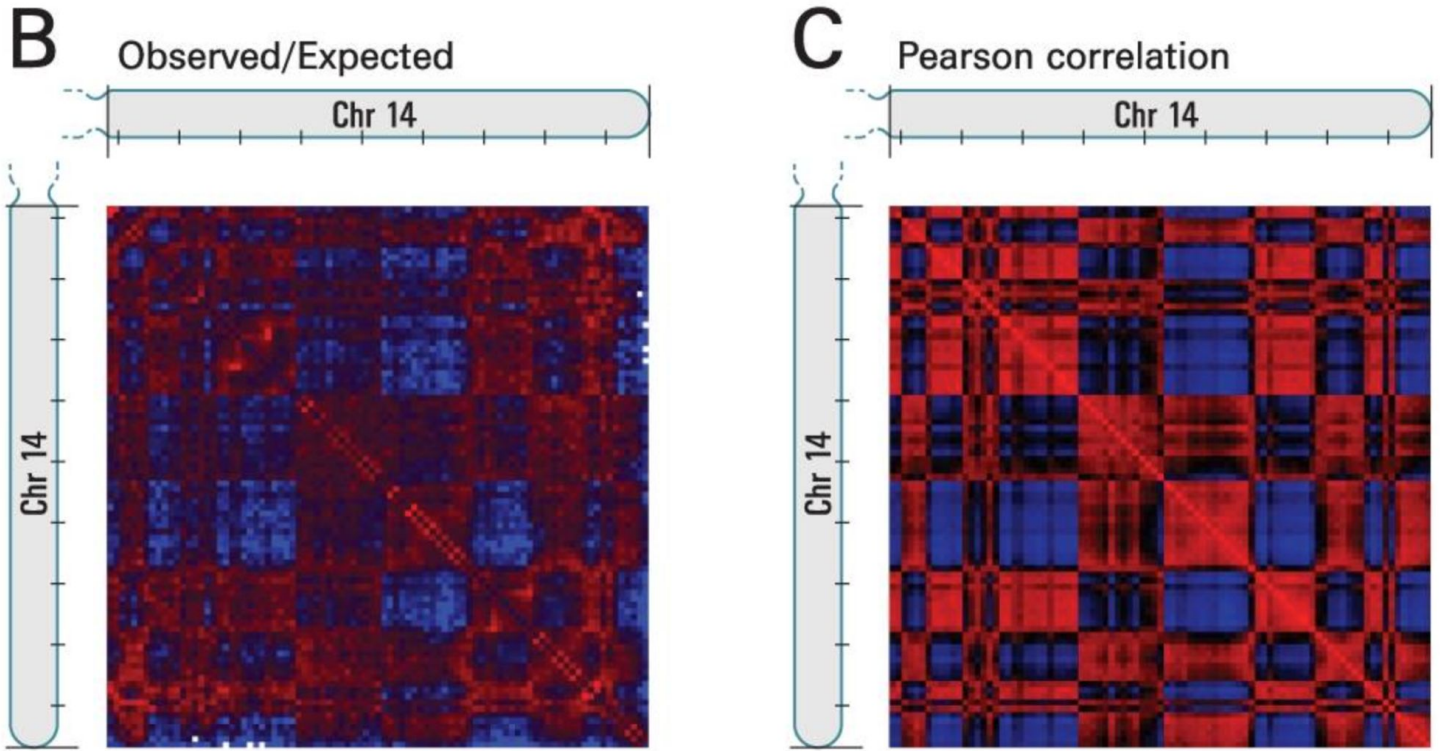
For each loci pair we calculate correlation between teo vectors: the first vector corresponds to all interaction values for the first locus (from the matrix B), the second vector — for the second locus.



Why to correlate?

This process dramatically sharpened the plaid pattern;

71% of the resulting matrix entries represent statistically significant correlations ($p \leq 0.05$).

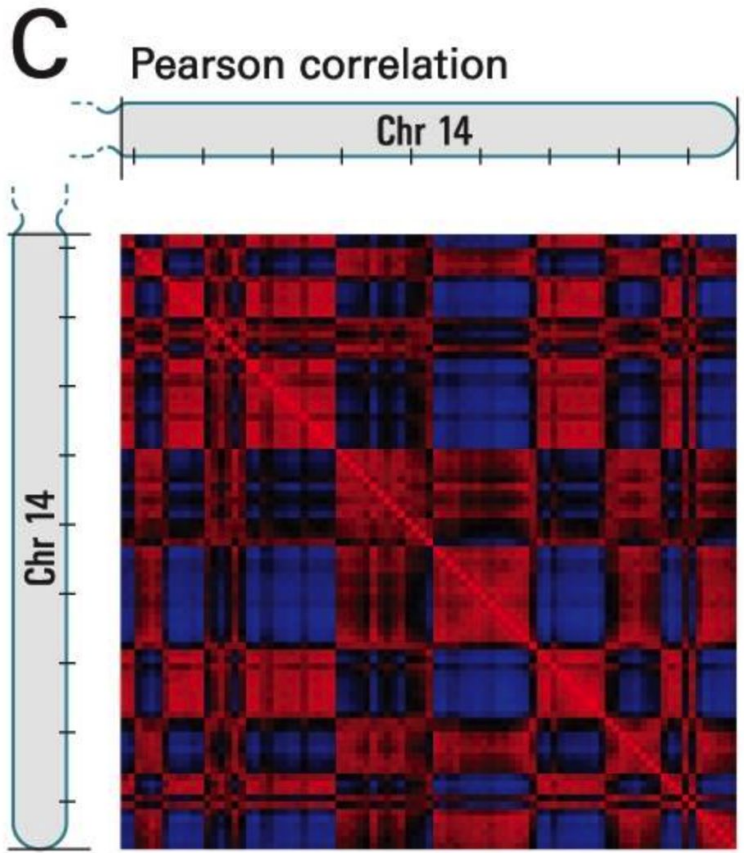


A and B components

Two sets of loci that correlate with each other.

First major component from PCA (principal component analysis) defines A/B compartments (positive/negative).

Who is who — correlate with GC/openess/chip-seq ….

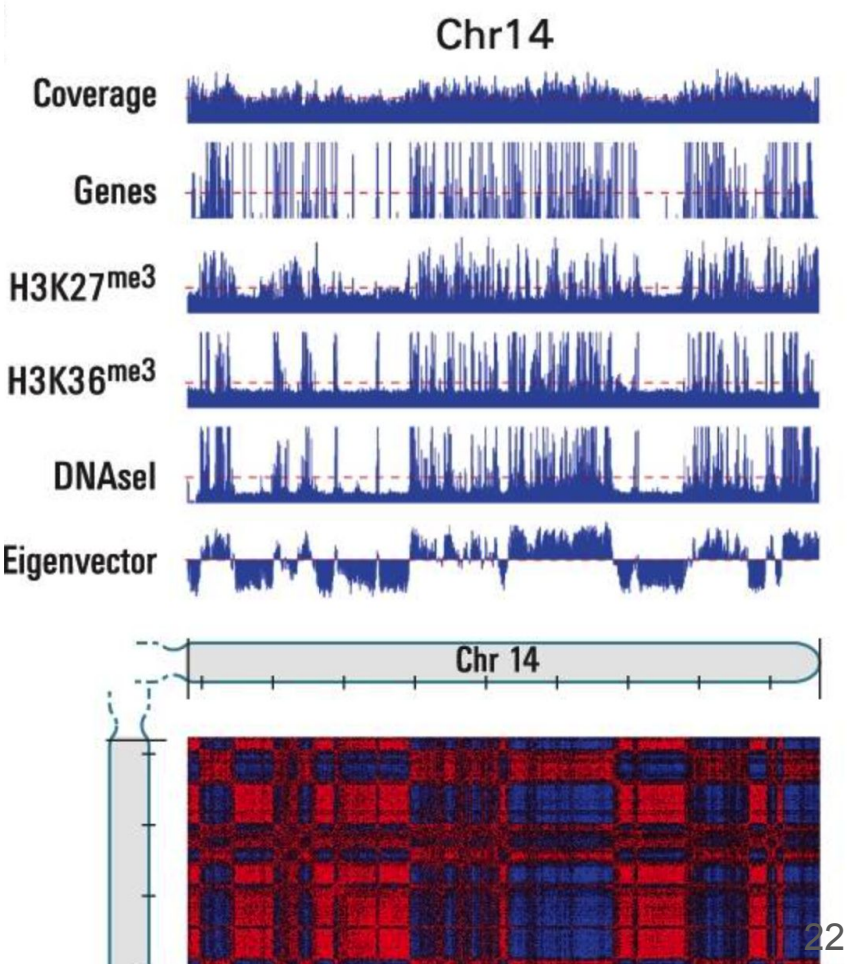


A/B properties

A: presence of genes, higher expression and accessible chromatin, activating (H3K36me3) ~~and repressive (H3K27me3) chromatin marks~~. Opened chromatin — euchromatin.

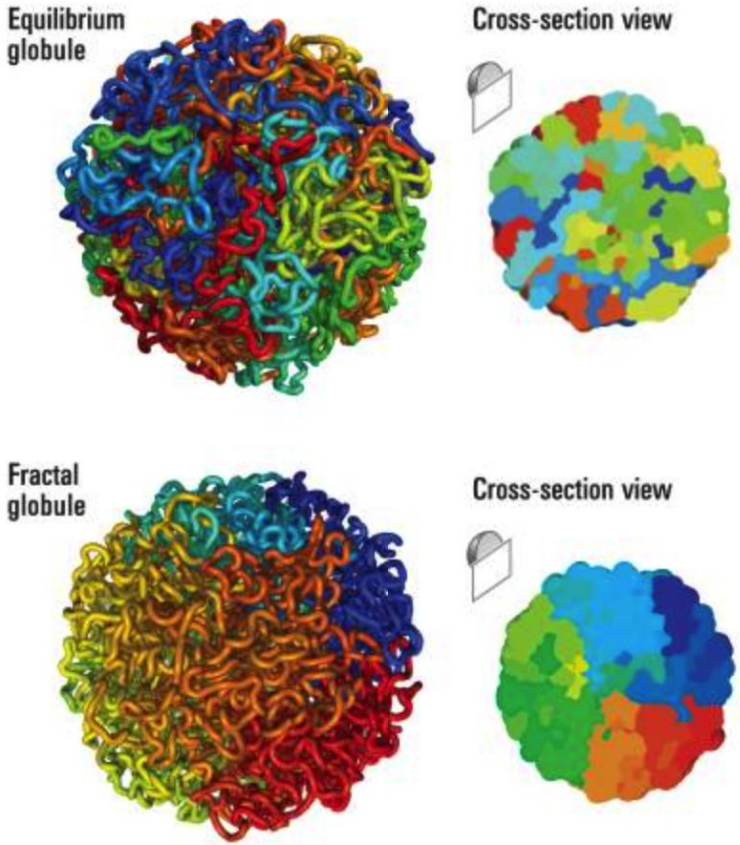
B: Closed chromatin — heterochromatin.

Initially eu- and heterochromatin were seen in microscopy.

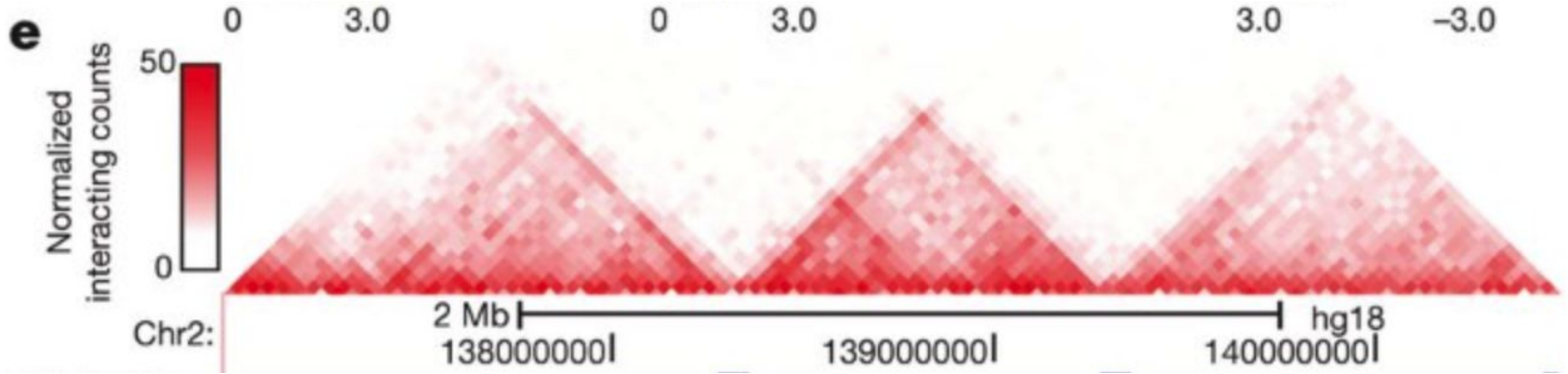


Polymer physics in chromatin

The ‘equilibrium globule’ and ‘fractal globule’ models make very different predictions concerning the scaling of contact probability with genomic distance s. The equilibrium globule model predicts that contact probability will scale as s−3/2, which we do not observe in our data. We analytically derived the contact probability for a fractal globule and found that it decays as s−1 (SOM); this corresponds closely with the prominent scaling we observed (−1.08).



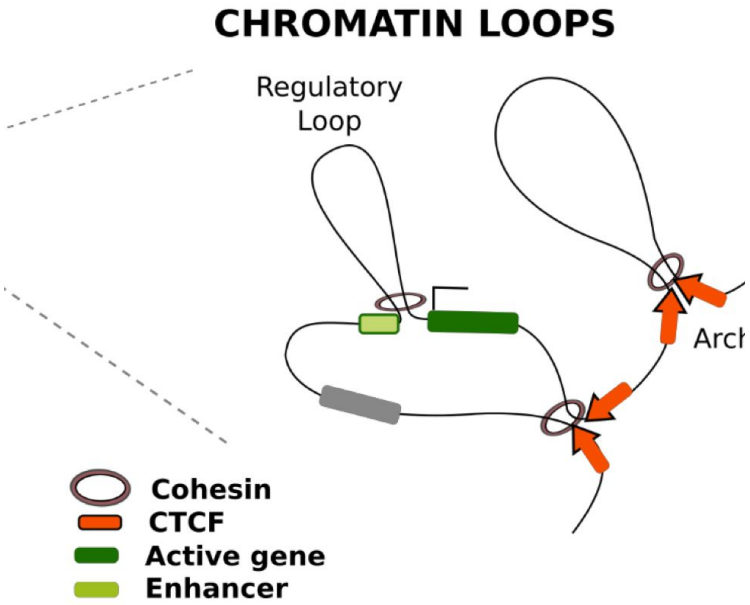
TADs. Topologically associated domains.

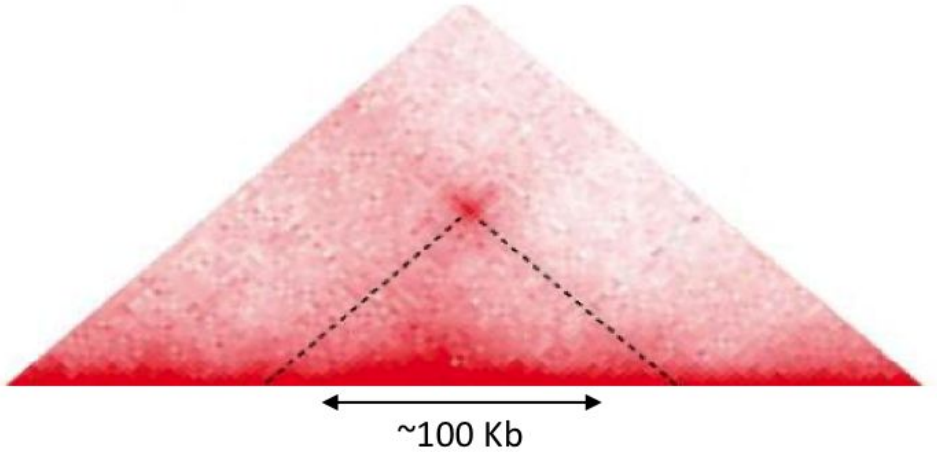


In summary, we show that the mammalian chromosomes are segmented into megabase-sized topological domains. We have identified multiple factors that are associated with the boundary regions separating topological domains, including the insulator binding factor CTCF, housekeeping genes and SINE elements.

Dixon, J., Selvaraj, S., Yue, F. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380 (2012). https://doi.org/10.1038/nature11082

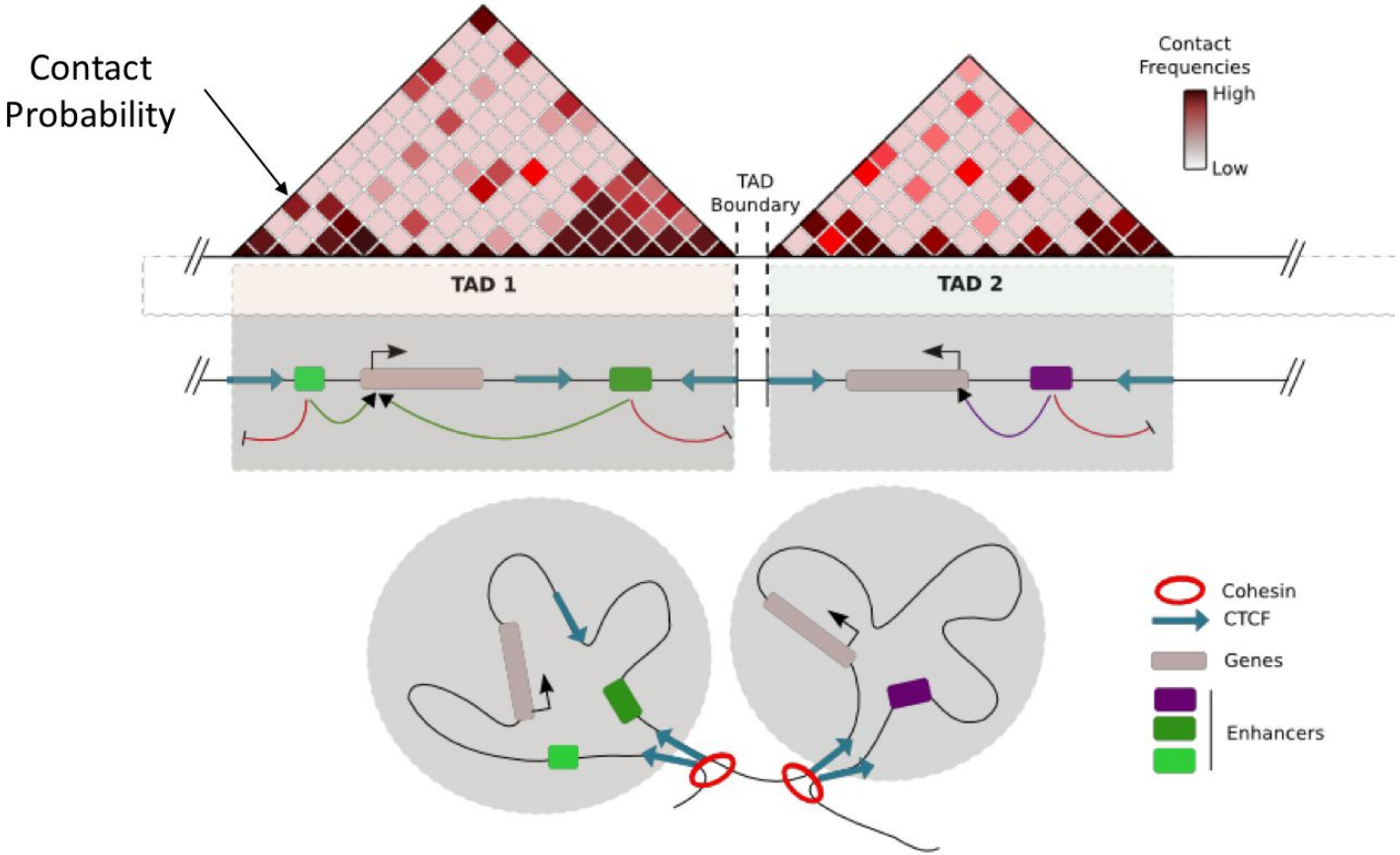
Chromatin loops





https://mathforgenomics.github.io/servant.pdf

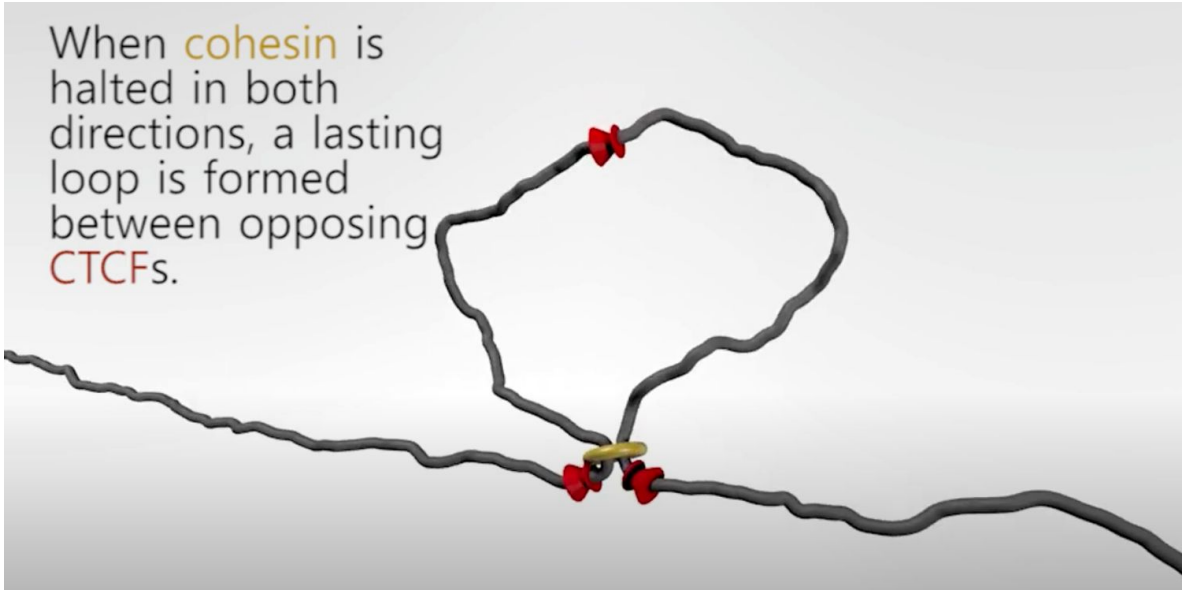
TADs



https://mathforgenomics.github.io/servant.pdf

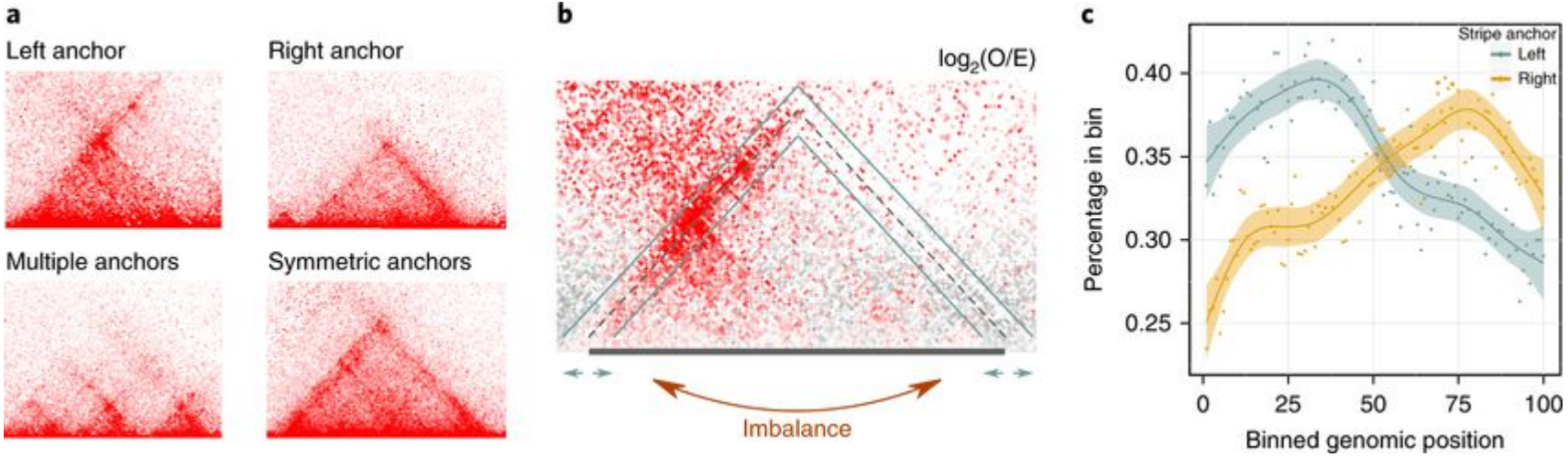
Loop extrusion

TAD is an ensemble of dynamic loops forming in different cells in the sample.



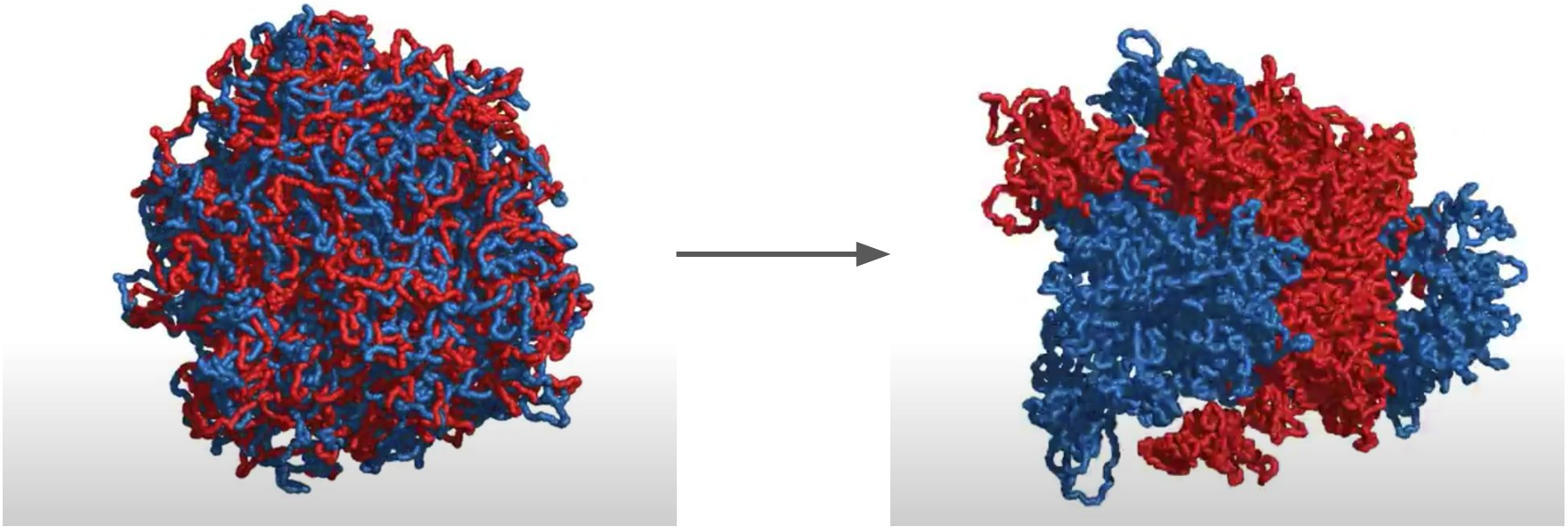
https://mirnylab.mit.edu/projects/emerging-evidence-for-loop-extrusion/

https://www.youtube.com/watch?v=8FW6gOx5lPI&ab_channel=MirnyLab



https://www.researchgate.net/figure/Genome-wide-identification-and-analysis-of-asymmetric-stripes-in-limb-tissue-a-Examples_fig4_331029336

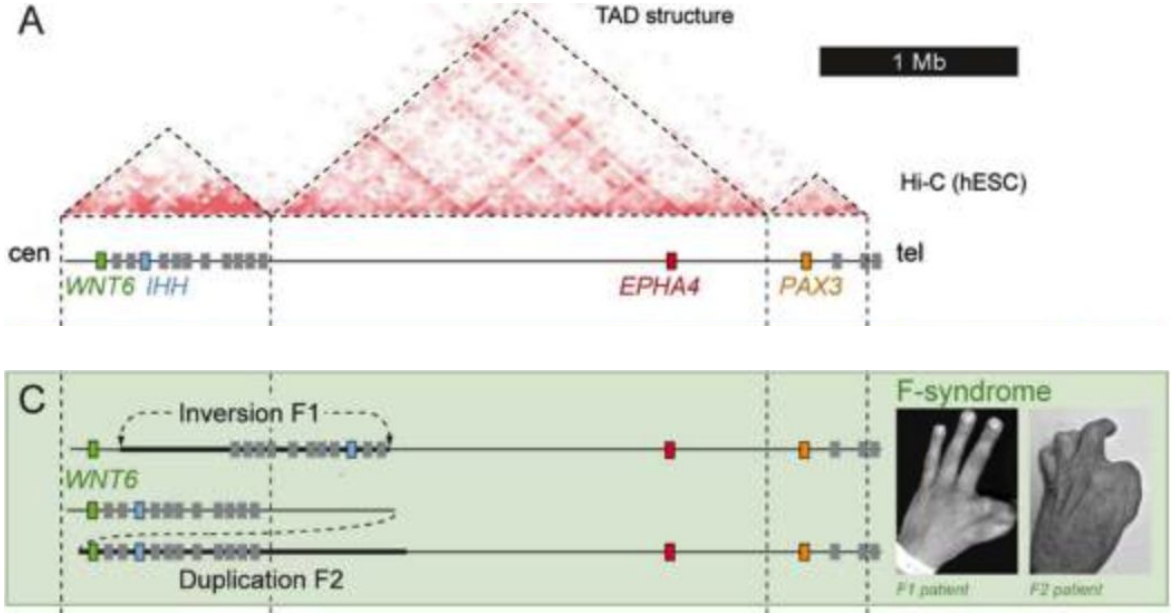
Loop extrusion in mitosis helps to separate chromatids



https://www.youtube.com/watch?v=stZR5s9n32s

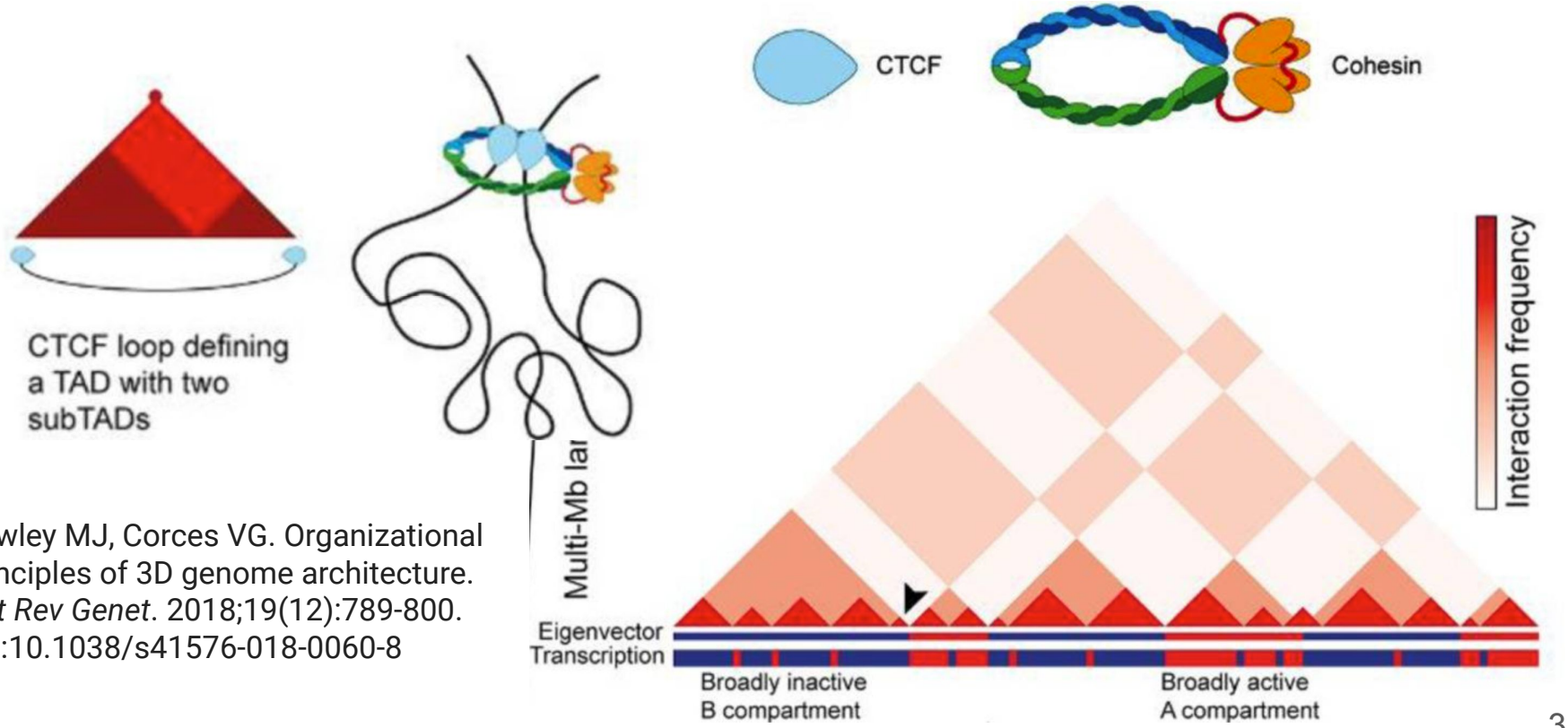
~~cohesin~~ → condensin

TADs influence promoter-enhancer interactions



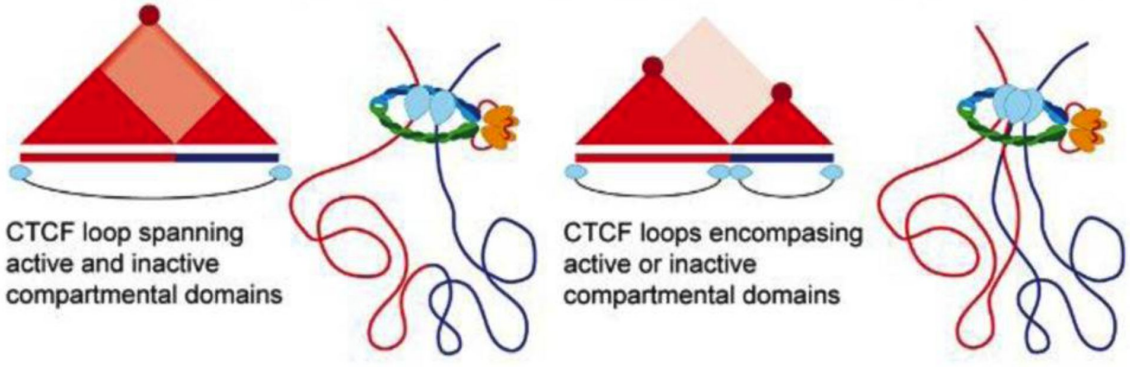
Lupiáñez DG, Kraft K, Heinrich V, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*. 2015;161(5):1012-1025. doi:10.1016/j.cell.2015.04.004

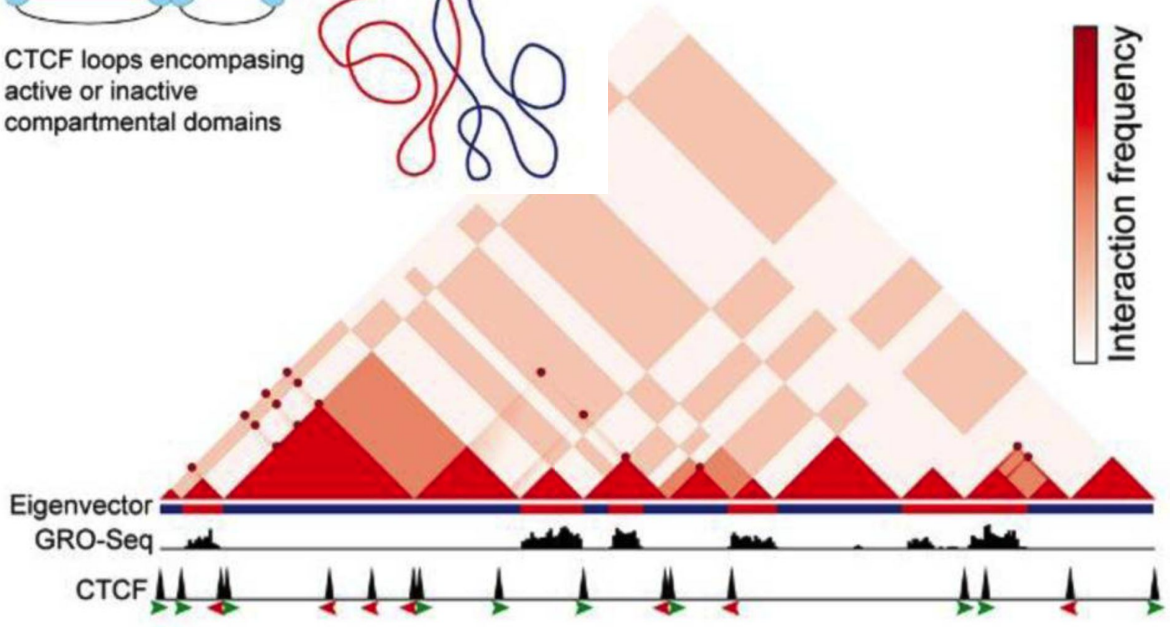
Hierarchical chromatin model



Rowley MJ, Corces VG. Organizational principles of 3D genome architecture. *Nat Rev Genet*. 2018;19(12):789-800. doi:10.1038/s41576-018-0060-8

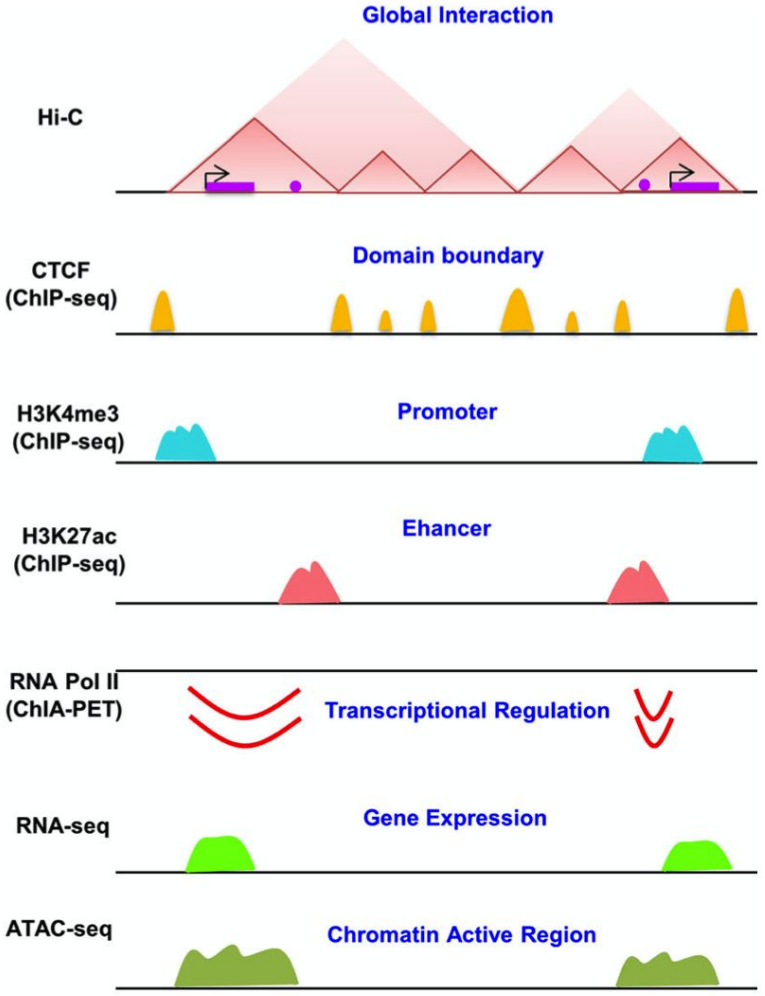
Not so hierarchical?





Multiomics

Add layers of information



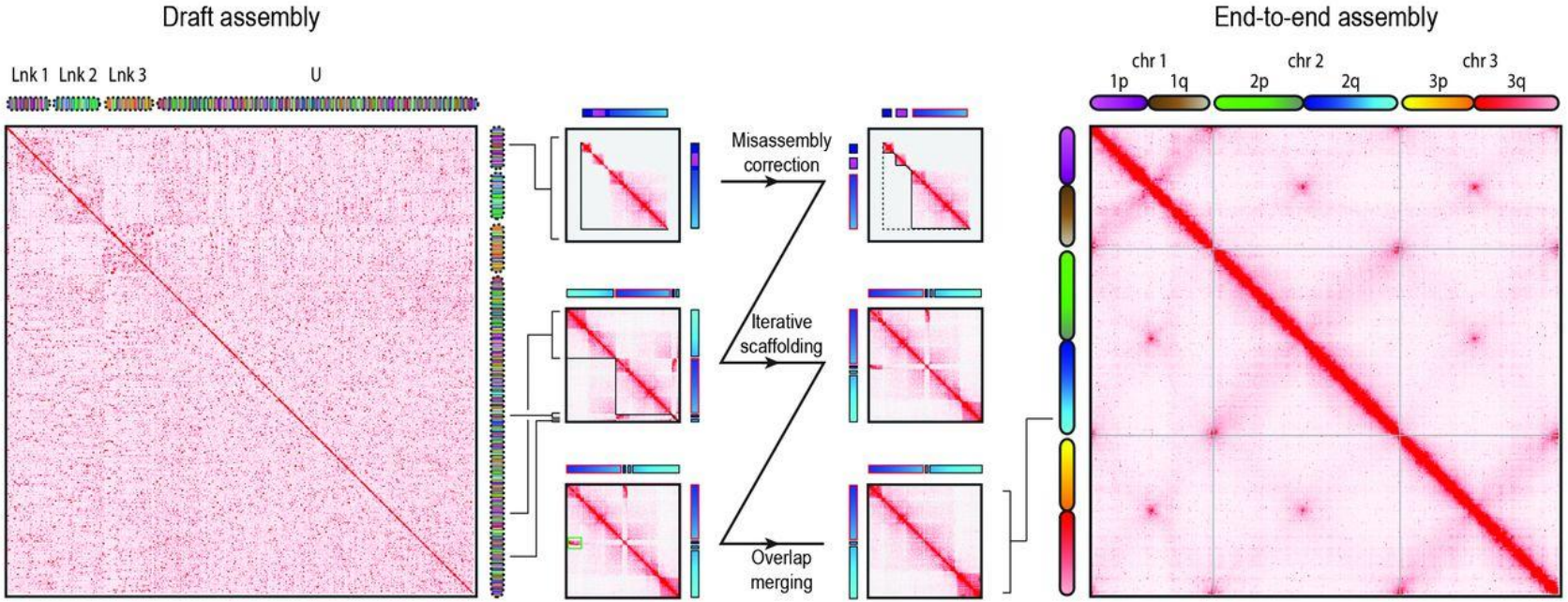
Kong S, Zhang Y. Deciphering Hi-C: from 3D genome to function. *Cell Biol Toxicol*. 2019;35(1):15-32. doi:10.1007/s10565-018-09456-2

What to do with the data?

Initial paper: bulk (many cells are averaged) Hi-C for a human cell line — revealed some organisation levels.

- Differences between tissues in human body or cell lines?
- Differences in one/many tissues during development and aging?
- Differences in pathological cases? (cancer…)
- Different species?
- Single-cell Hi-C — how widely chromatin structure vary between cells in one tissue or cell line?

Hi-C in genome assembly

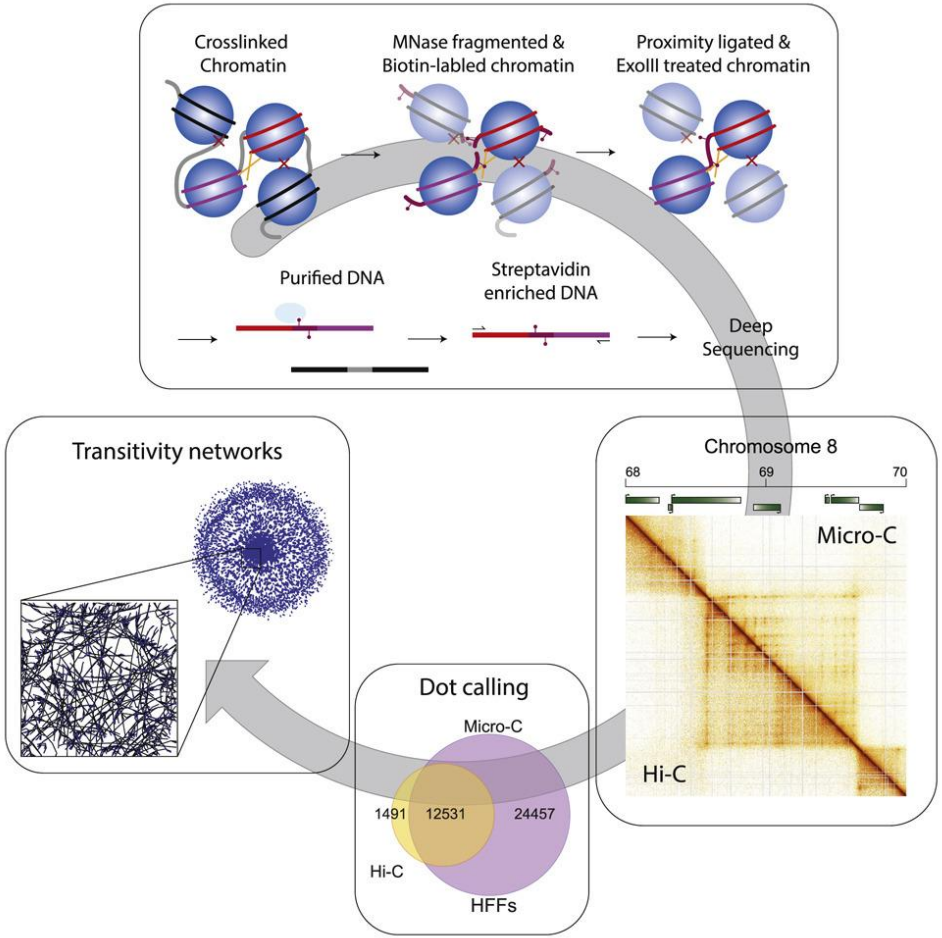


Aedes aegypti mosquito

Olga Dudchenko et al. ,De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds.Science356,92-95(2017).DOI:10.1126/science.aal3327

Micro-C

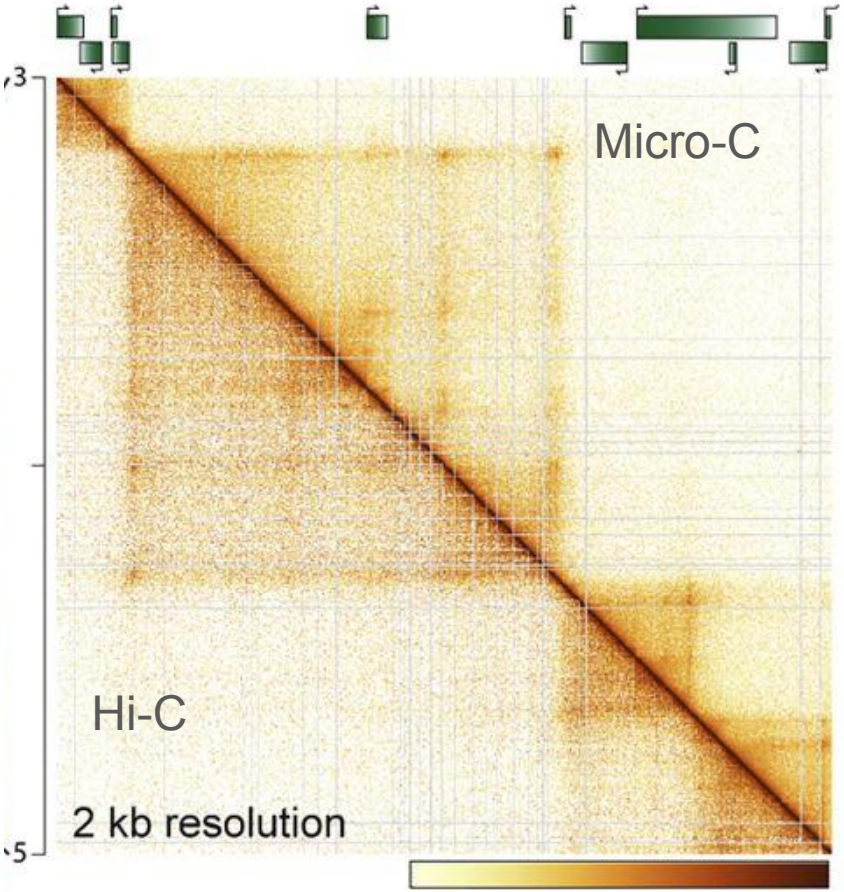
Compared to Hi-C, Micro-C exhibits an order of magnitude greater dynamic range, allowing the identification of ~20,000 additional loops in each cell type.



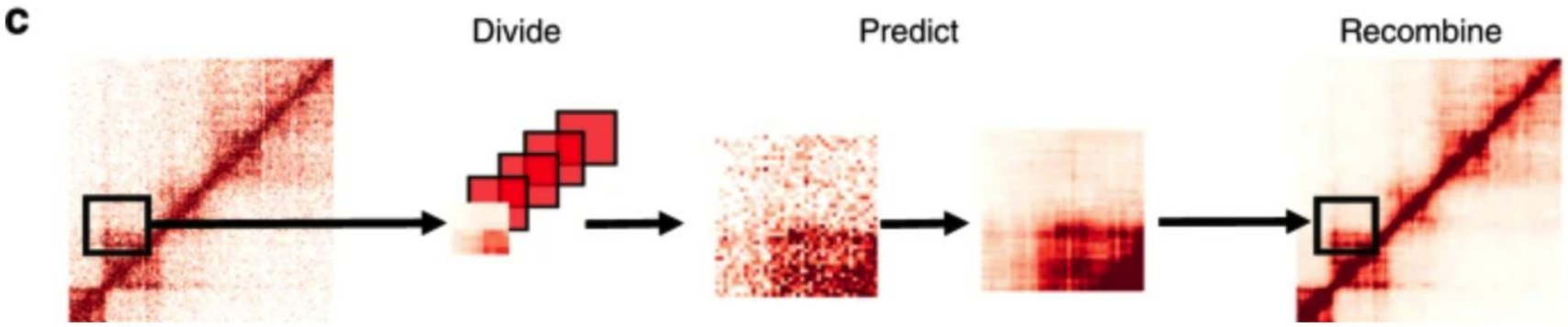
Krietenstein N, Abraham S, Venev SV, et al. Ultrastructural Details of Mammalian Chromosome Architecture. *Mol Cell*. 2020;78(3):554-565.e7. doi:10.1016/j.molcel.2020.03.003

Micro-C

Overall, Micro-C and Hi-C maps reveal the same major classes of patterns that illuminate various features of chromosome folding: at lower resolutions (e.g., ~25 kb bins) the coarse checkerboard pattern reflects the compartmentalization of active and inactive chromatin, while zooming to higher resolutions (e.g., <10 kb bins) reveals finer compartmental segmentation, TADs, and off-diagonal interaction peaks thought to result from a high frequency of CTCF-anchored long-range looping interactions.



Find factors that determine TADs



What sequence/chromatin structure features are important?

Zhang, Y., An, L., Xu, J. *et al.* Enhancing Hi-C data resolution with deep convolutional neural network HiCPlus. *Nat Commun* 9, 750 (2018). https://doi.org/10.1038/s41467-018-03113-2