# Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L | week 12

Pierre-Luc Germain

ETH Zürich

# Plan for today

- Debriefing on the assignment

- **Theory:**

  Chromatin conformation & related technologies

- **Practice:**
  Using long-range interactions to annotate distal sites

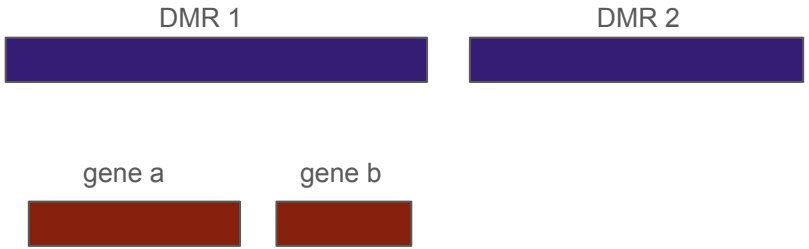
# Debriefing on the assignment

The first part was:

Plot a heatmap of the methylation levels of the **genes** in top 5 DMR regions.

# Debriefing: Differentially methylated ranges (DMRs)

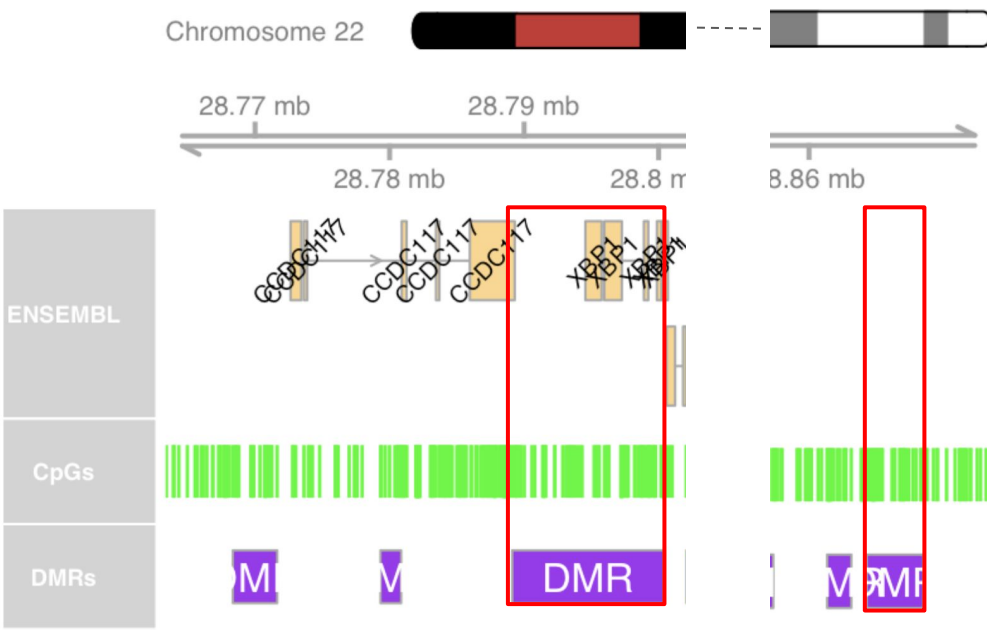
-de-novo detection of DMRs with DMRcate




One DMR can contain several genes but also none or only partially…

# debriefing: Differentially methylated ranges (DMRs)

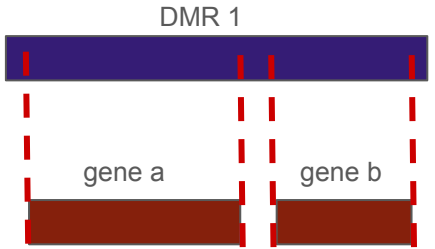
-de-novo detection of DMRs with DMRcate




One DMR can contain several genes but also none.

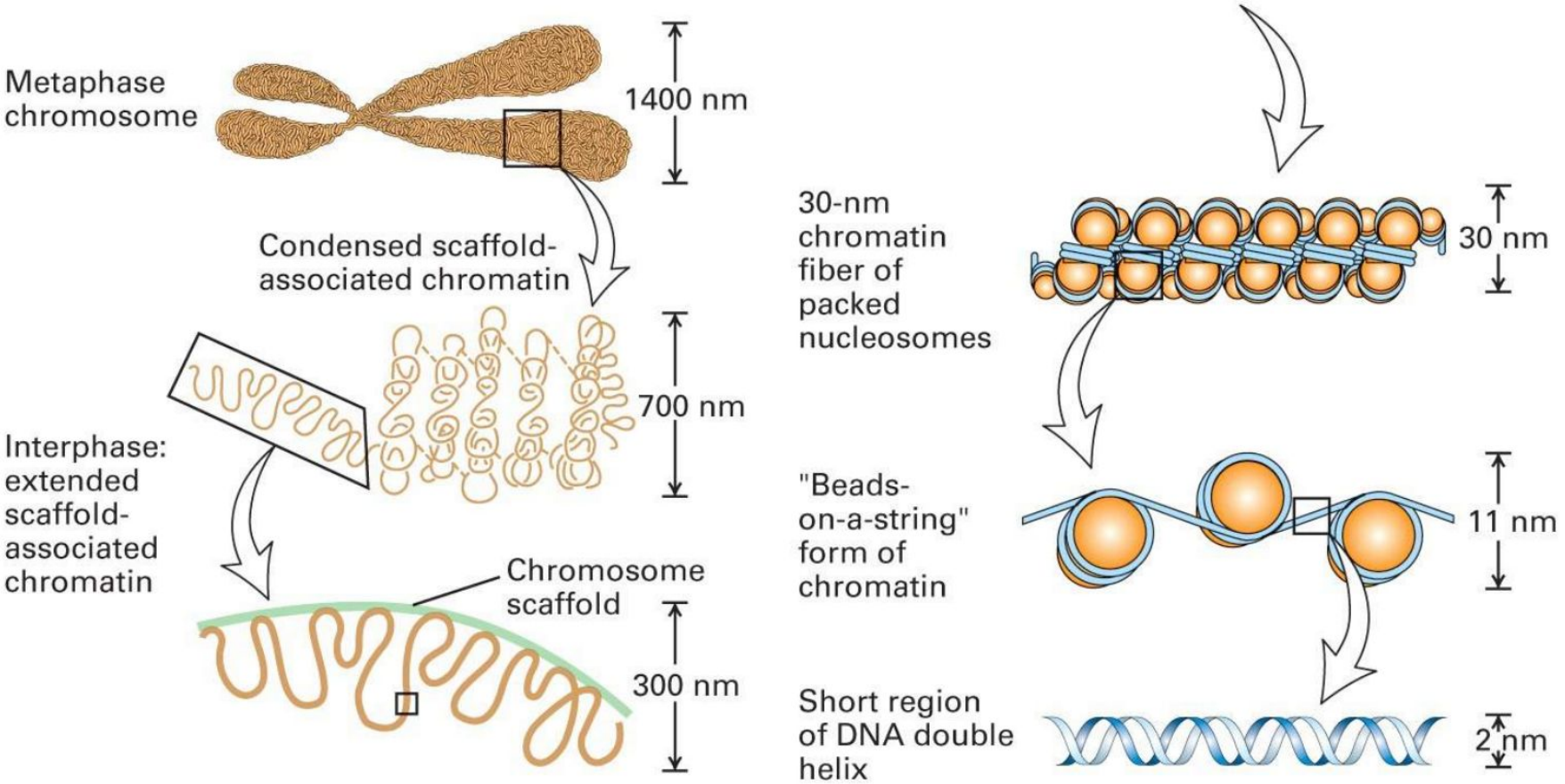
# debriefing: Differentially methylated ranges (DMRs)

-retrieving gene coordinates within the top 10 DMRs:



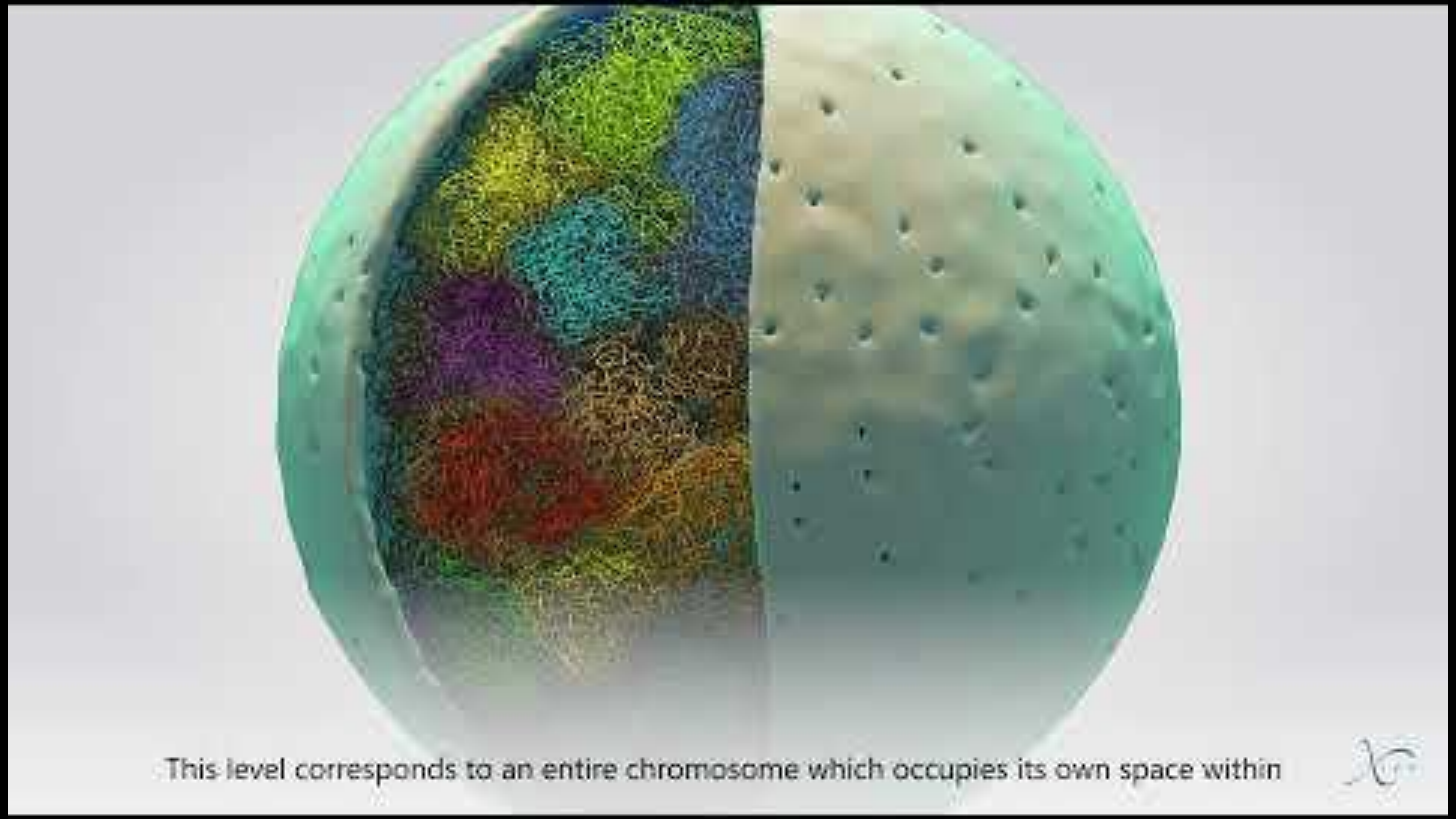

```
chr22 <-  GRanges(seqnames=Rle(c("22")),
                  ranges = IRanges(1, end=50818468))
genesChr22 <- genes(ensdb, columns=c("gene_seq_start", "gene_seq_end", "gene_name"),
                    filter=GRangesFilter(chr22))
seqlevelsStyle(genesChr22) <- "UCSC"
```

```
# Get the genes within Differentially methylated regions
topIdx <- order(dmrRangesGenes$min_smoothed_fdr)[1:10]
genesDmr <- unlist(tstrsplit(dmrRangesGenes[topIdx]$overlapping.genes, split=", "))
genesDmr <- genesDmr[!is.na(genesDmr)]
dmrGenes <- genesChr22[genesChr22$gene_name %in% genesDmr]
dmrGenes
```



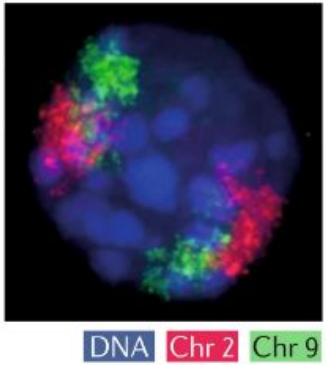

Lodish et al. *Molecular Biology of the Cell* (5th ed.). W.H. Freeman & Co., 2003.

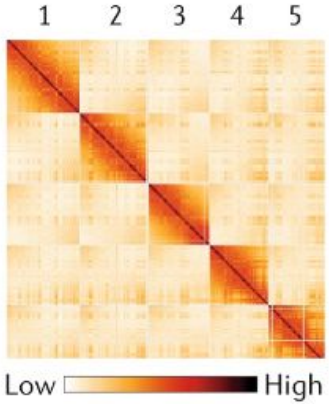
# 3D organization of the genome

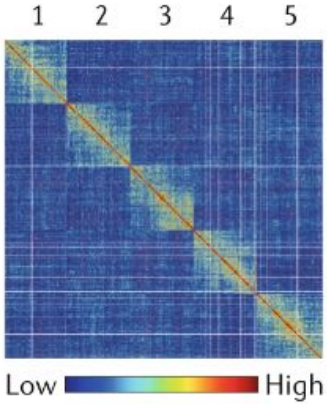


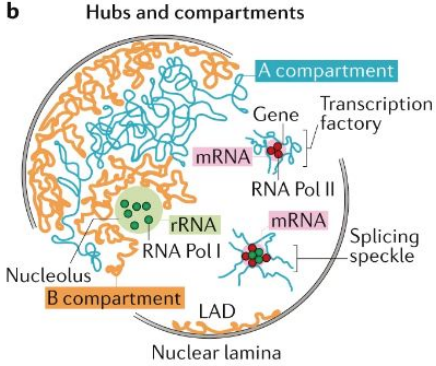

https://youtu.be/Pl44JjA--2k

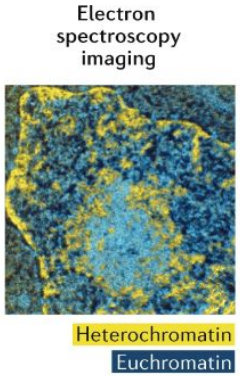
# The nucleus is organized into chromosome territories

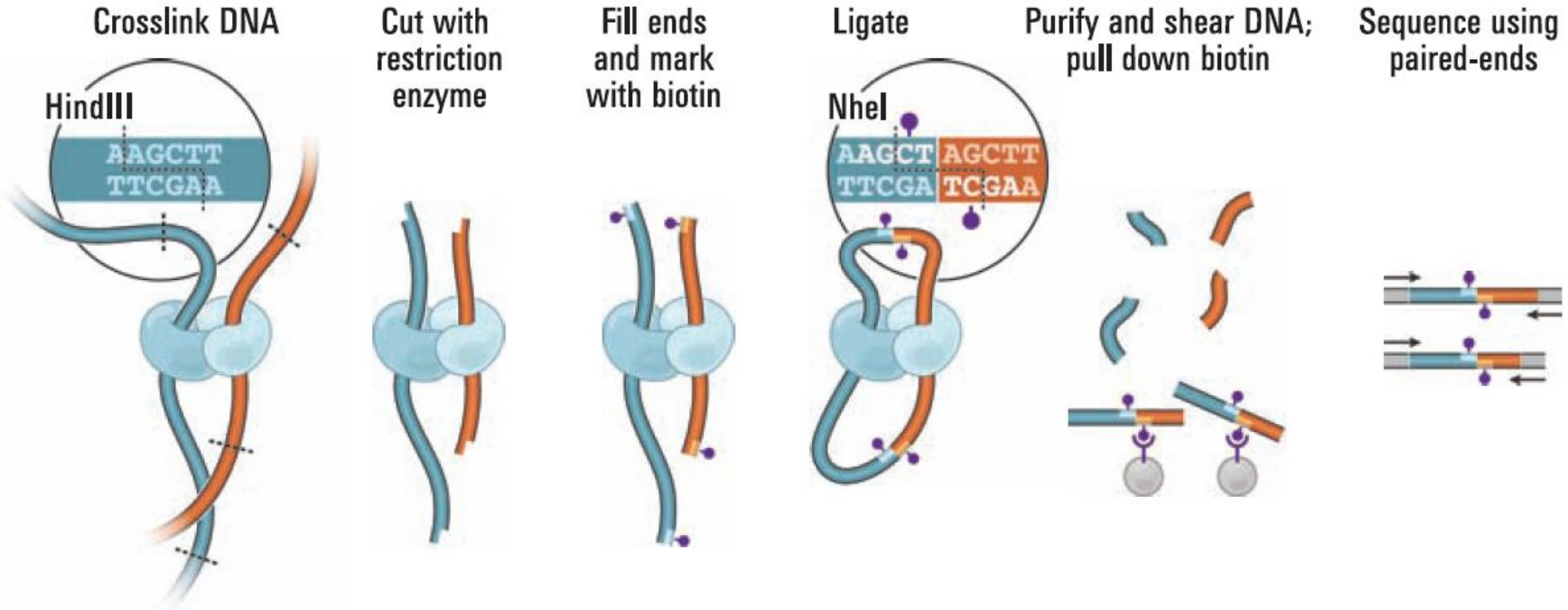




Chromosomes are split into active (A) and inactive (B) compartments

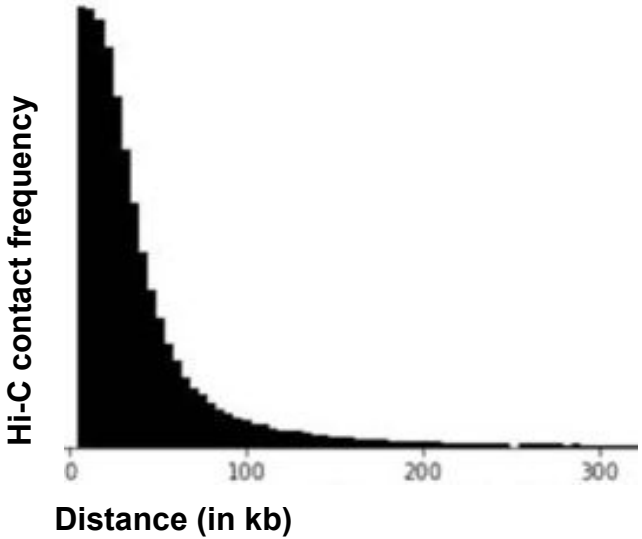
(Adapted from Kempfer and Pombo 2020)

# Sequencing-based conformation capture




(Lieberman-Aiden et al. 2009)

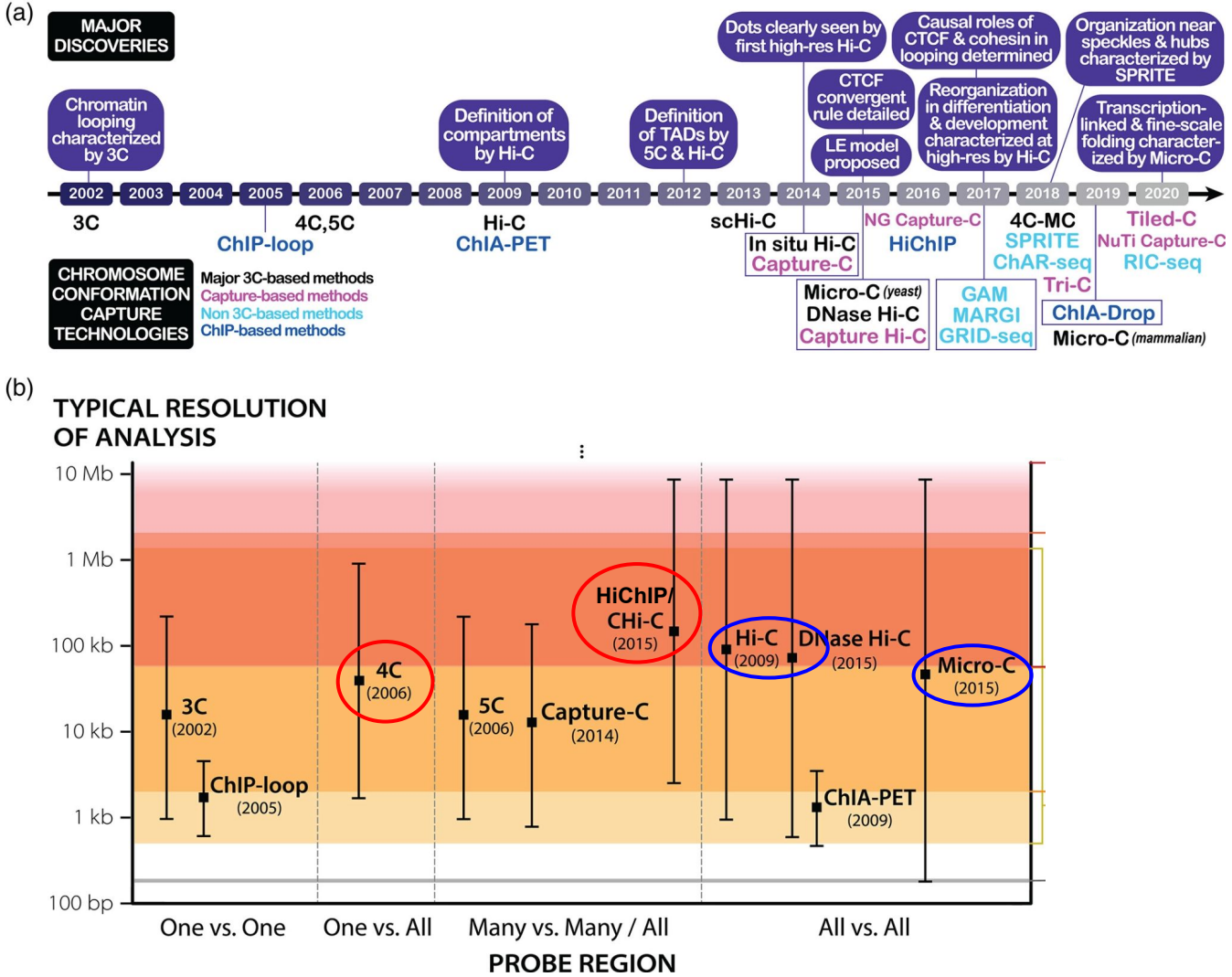
# Sequencing-based conformation capture




Under the null hypothesis of random 3D organization, we expect to have more contacts from DNA regions that are closer to each other.

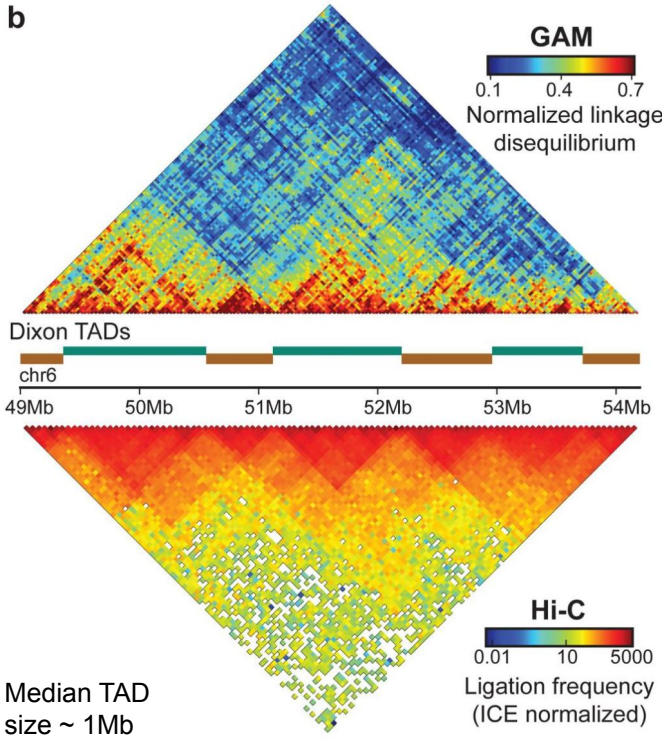
This is why, when identifying interactions, methods look for an excess over this null hypothesis

# Sequencing-based conformation capture



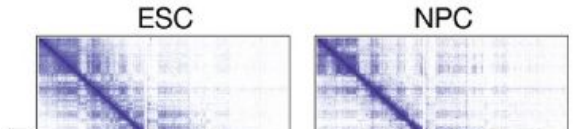

(Adapted from Goel and Hansen 2020)

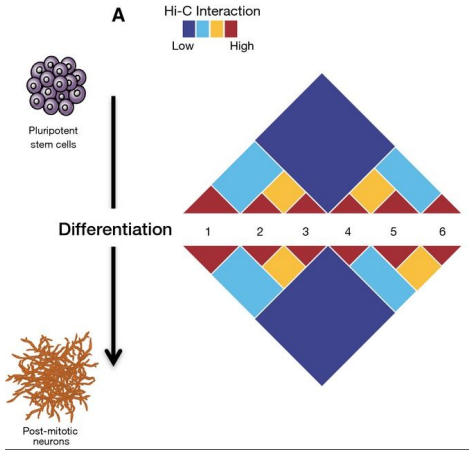
# Chromosomes are organized into topologically associated domains (TADs)




Median TAD size ~ 1Mb

(Beagrie et al., 2017)

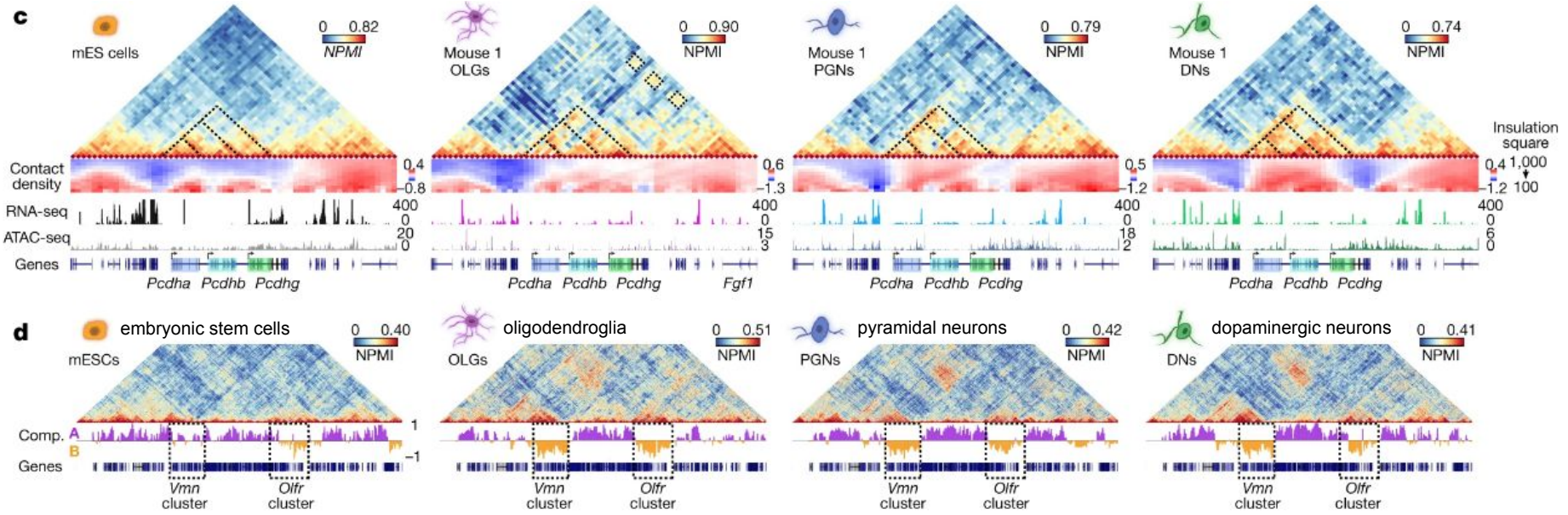


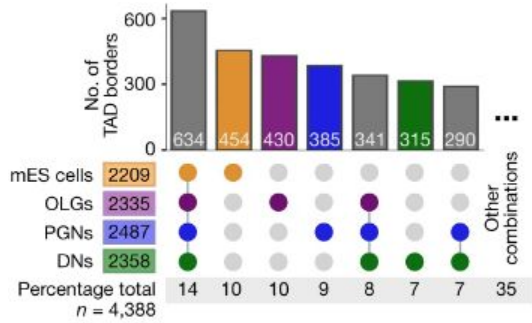

(Fraser et al., 2015)

...that are rather stable across cell types

# TADs rearrangement across cell types



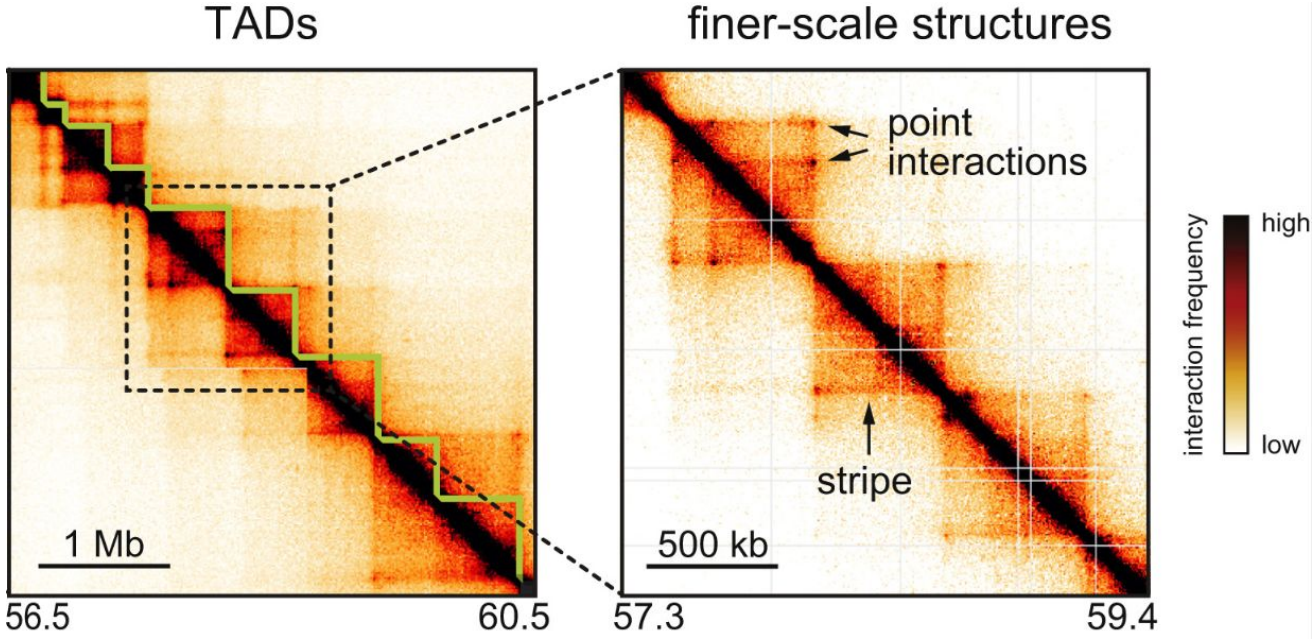

Comparison of TADs borders:




(Winick-Ng et al., *Nature* 2021)

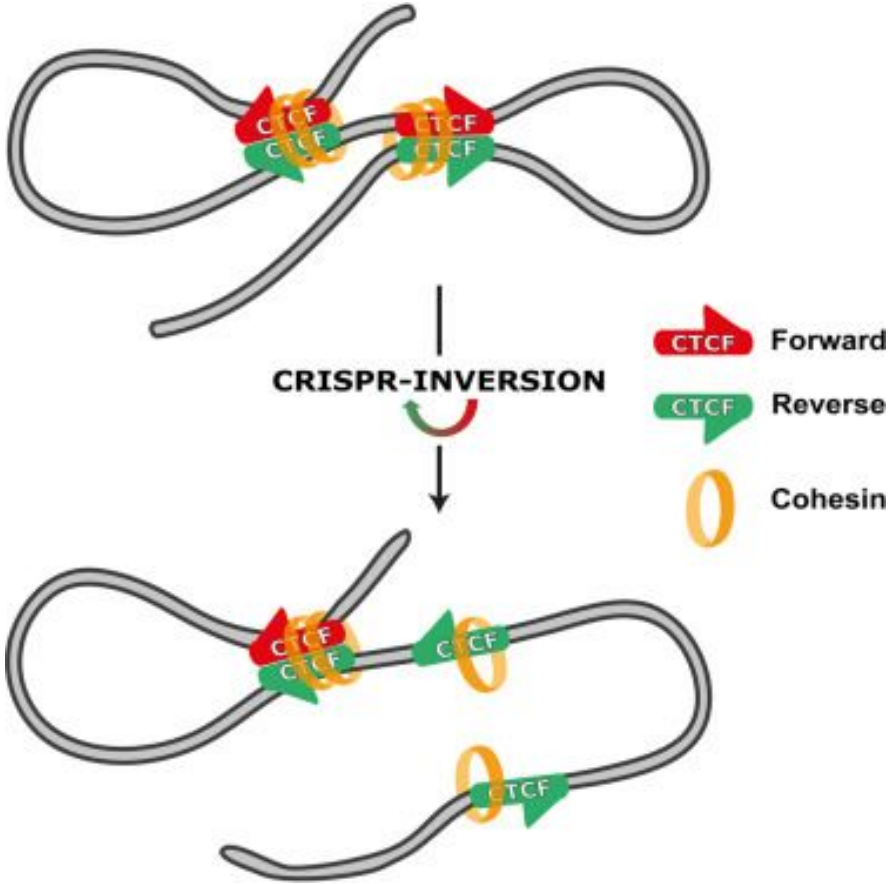
TAD boundaries are defined by very stable point interactions…

…which represent CTCF binding sites




(Adapted from McCord, Kaplan and Giorgetti, Mol Cell 2020)

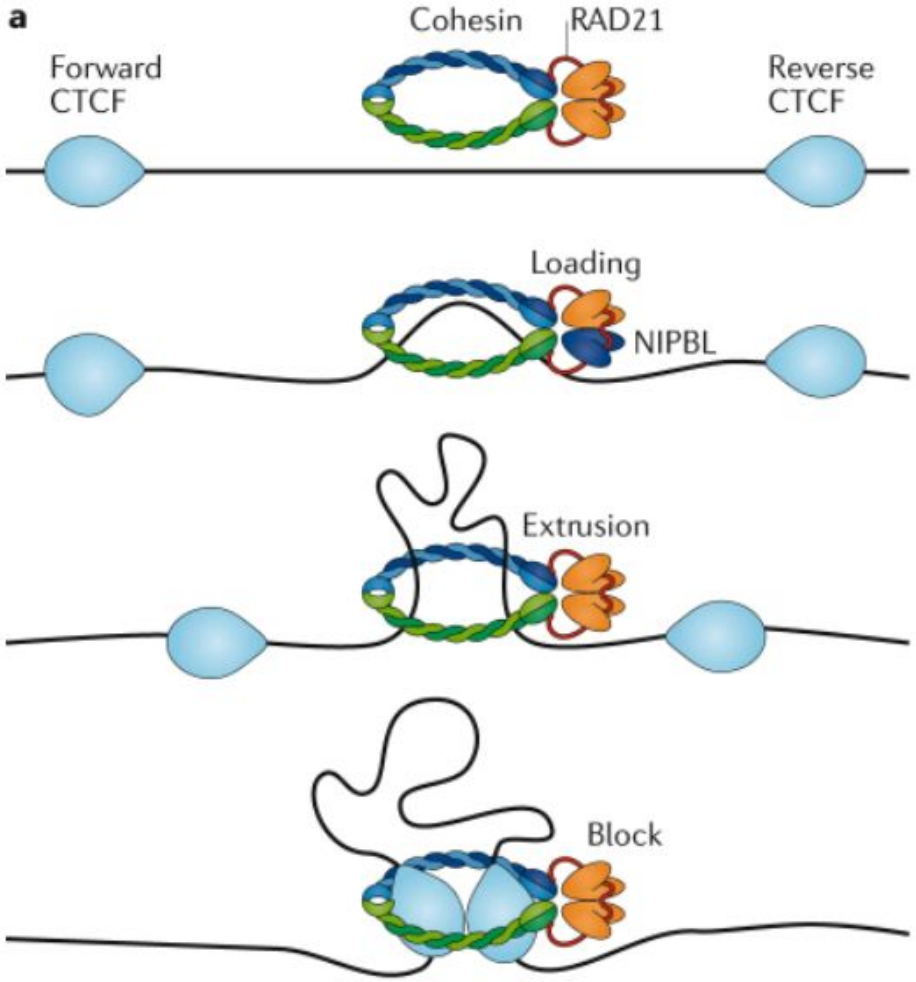
CTCF forms **convergent** dimers at loops



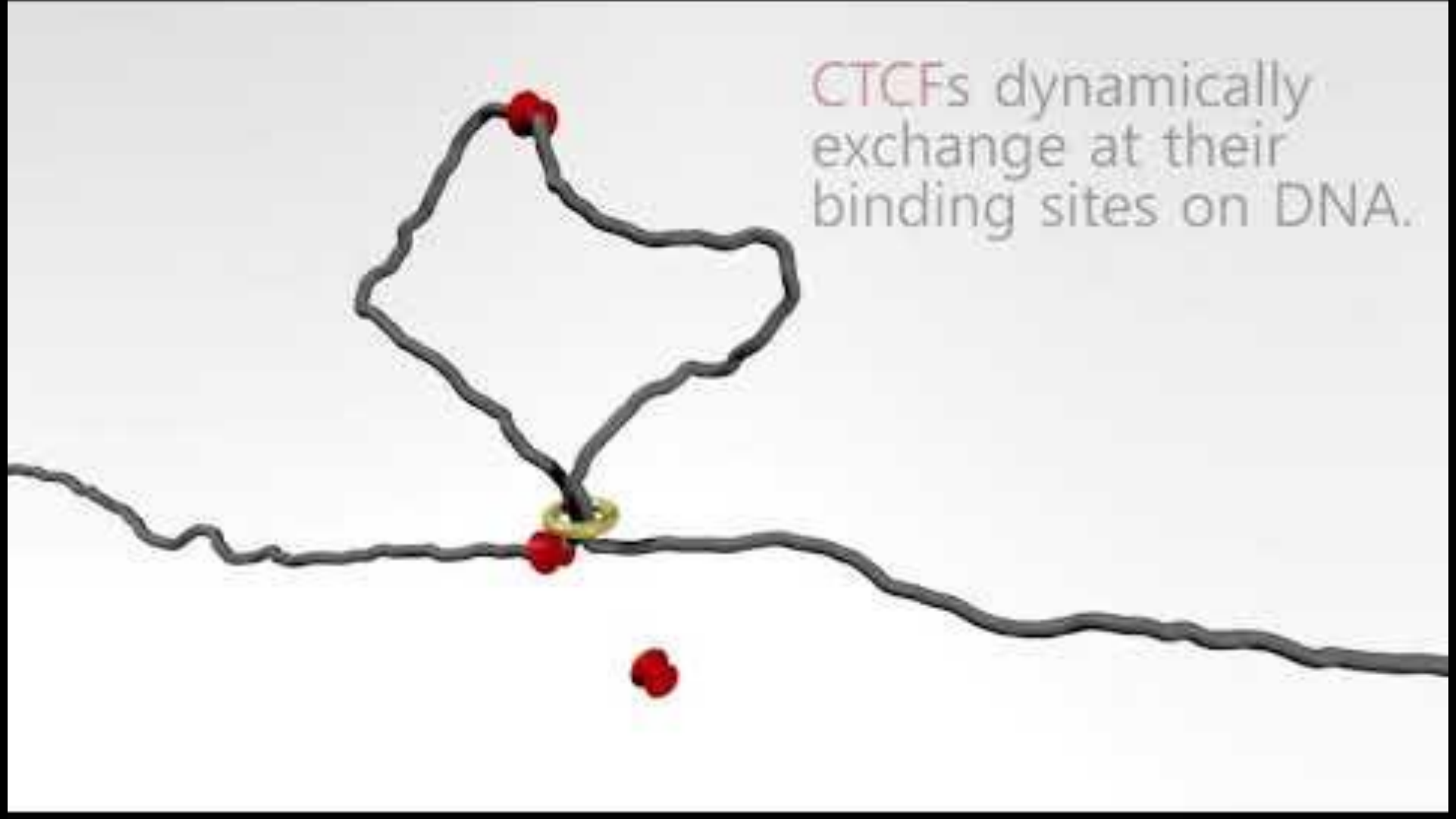

(Adapted from de Wit et al., Mol Cell 2015)

# The loop extrusion model

CTCF dimers form loops
by blocking Cohesin



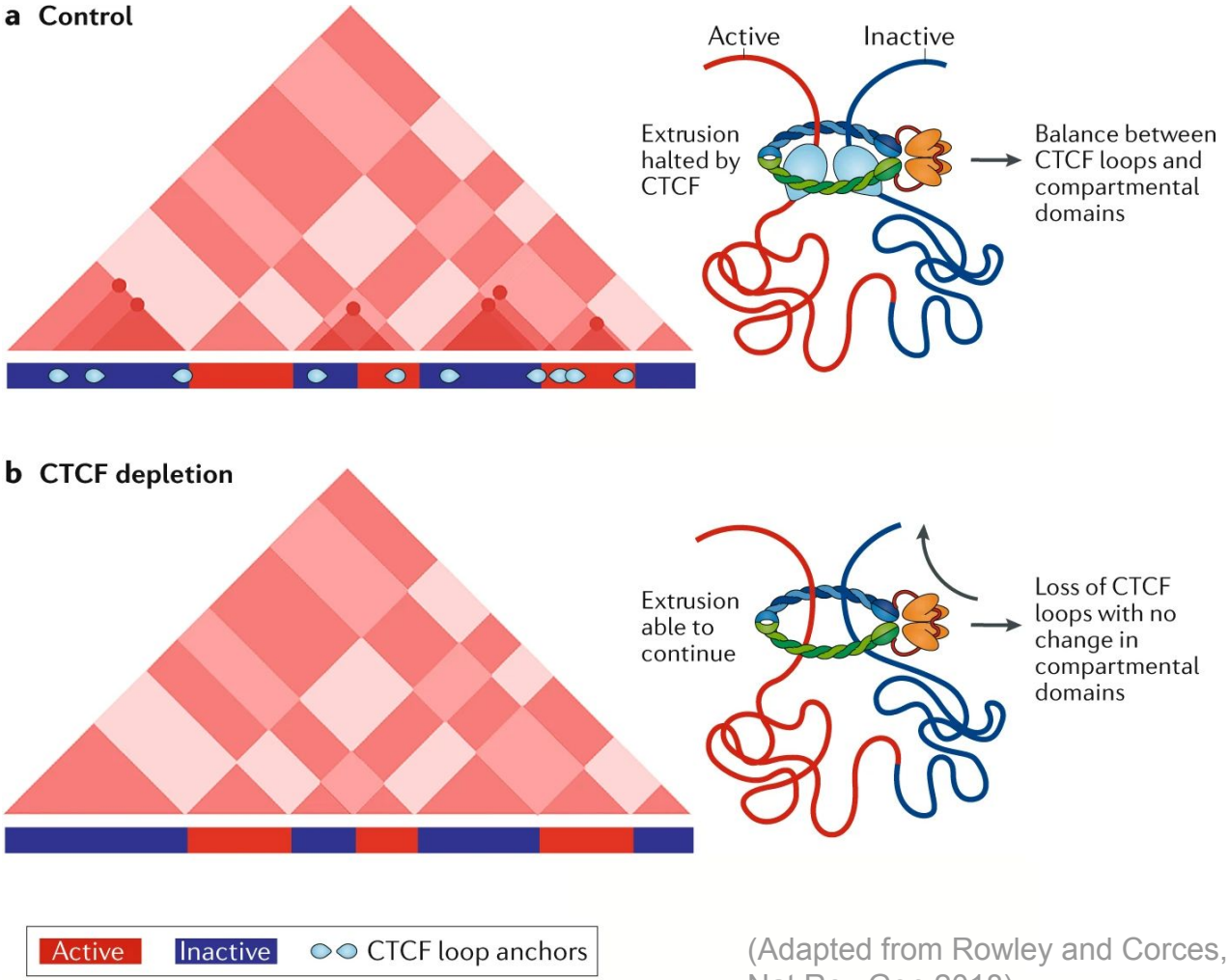

(Adapted from
Rowley and Corces, Nat Rev Gen 2018)


CTCFs dynamically
exchange at their
binding sites on DNA.

CTCF depletion disrupts especially TAD-internal structures

presumed to include promoter-enhancer loops

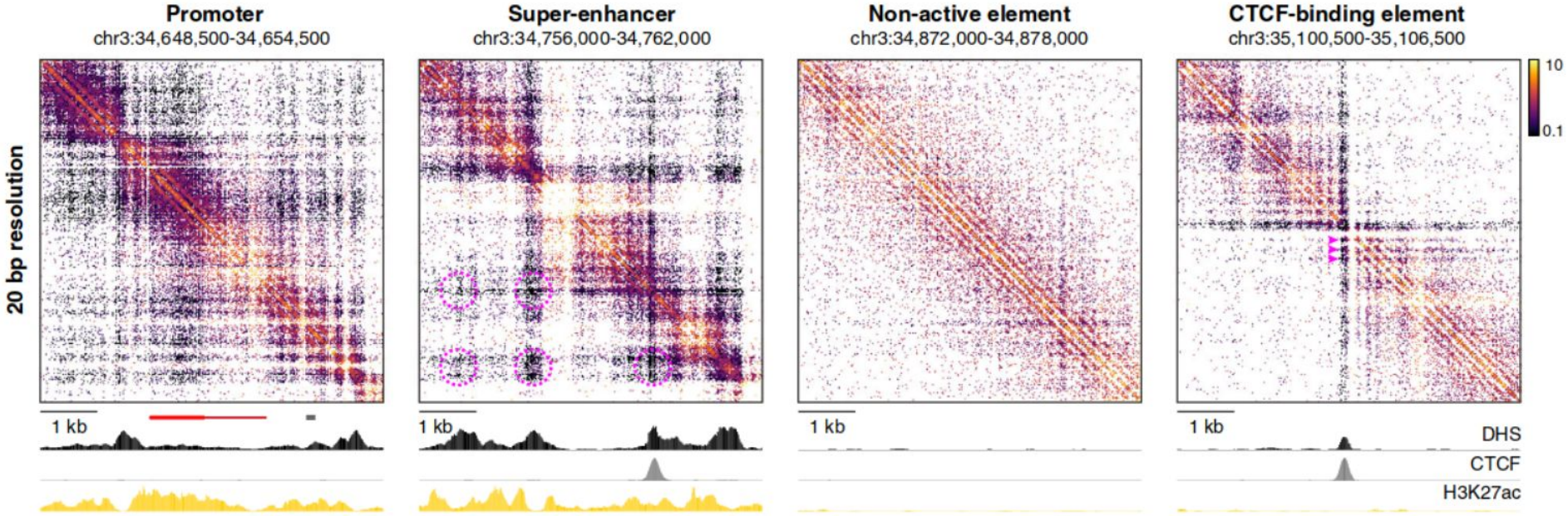
however…




(Adapted from Rowley and Corces, Nat Rev Gen 2018)

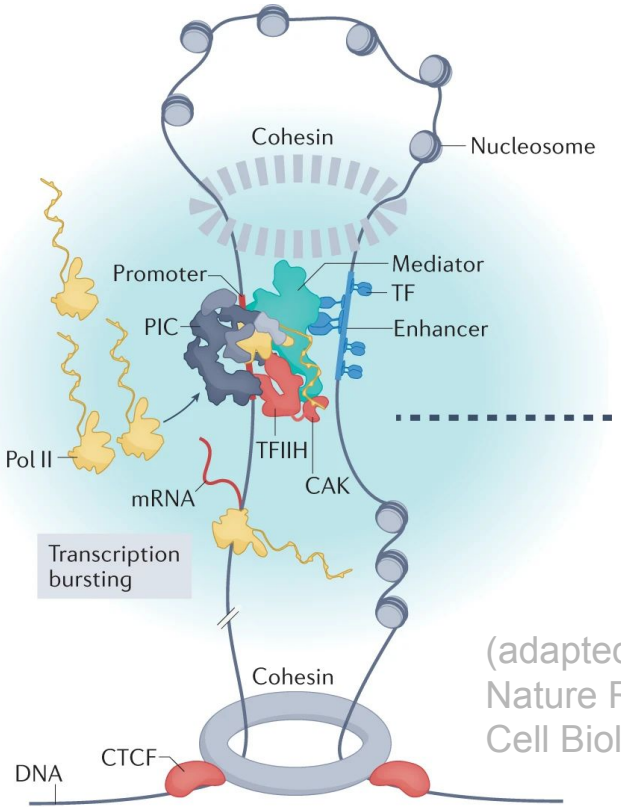
# When going at very high resolutions, similar contact patterns appear, that are however not associated with CTCF

The structures however appear correlated with DHS, suggesting that at this fine-grained level the it is determined by (most likely combinations of) other factors




(20bp-resolution contact maps, adapted from Aljahani et al., *Nat Comm* 2022)

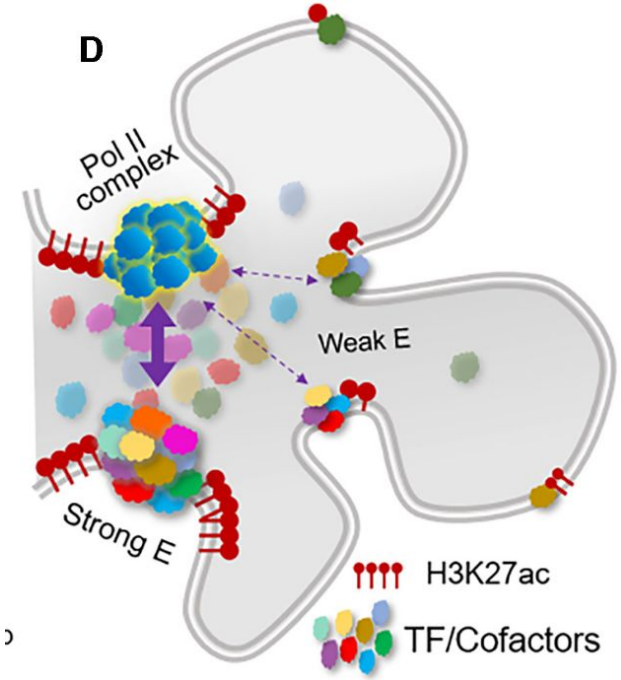
According to the prevailing model, enhancers physically interact with target TSS through combinations of proteins (in particular *Mediator*)




(adapted from Richter et al., Nature Reviews Molecular Cell Biology 2022)

In both views, chromatin loops create local neighborhoods in which these interactions can take place

An alternative view rests on diffusion between regulatory elements

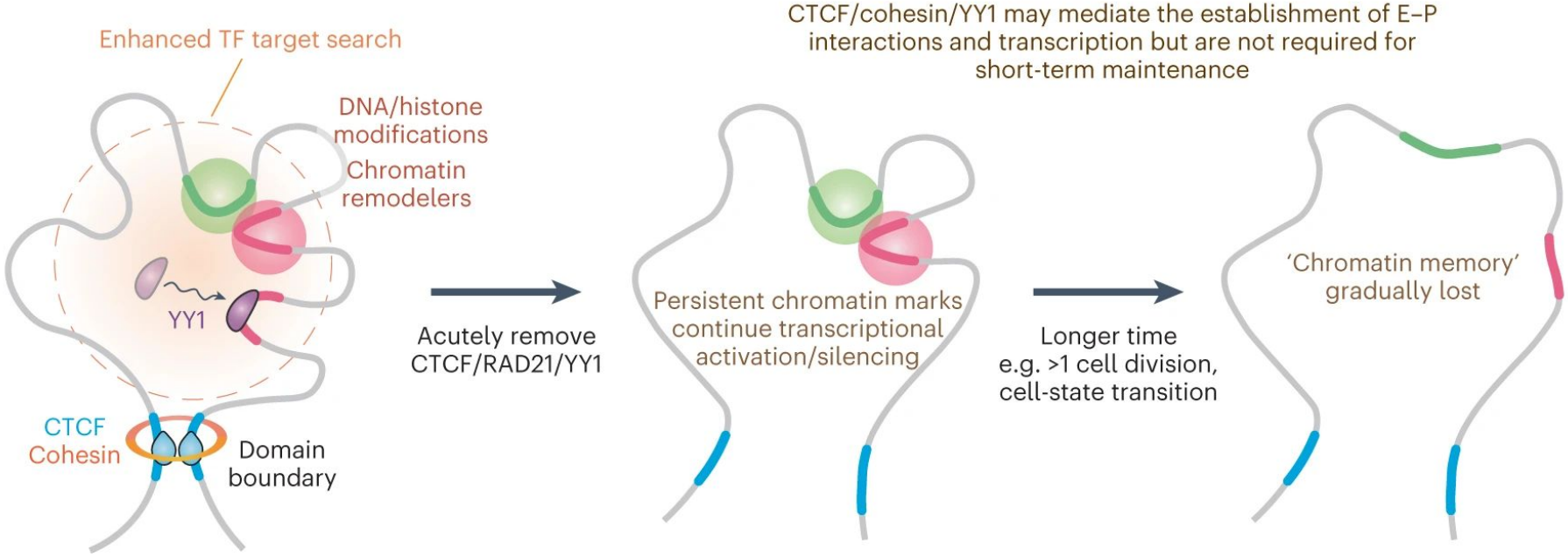



(adapted from Zhu et al., Nucleic Acids Research 2021; see also Galouzis and Furlong, Current Opinion in Cell Biology 2022)

# Enhancer–promoter interactions and transcription are largely maintained upon acute loss of CTCF, cohesin, WAPL or YY1

Tsung-Han S. Hsieh, Claudia Cattoglio, Elena Slobodyanyuk, Anders S. Hansen, Xavier Darzacq ✉ & Robert Tjian ✉

*Nature Genetics* **54**, 1919–1932 (2022) | Cite this article



# Proximity is not sufficient for enhancer-mediated activation

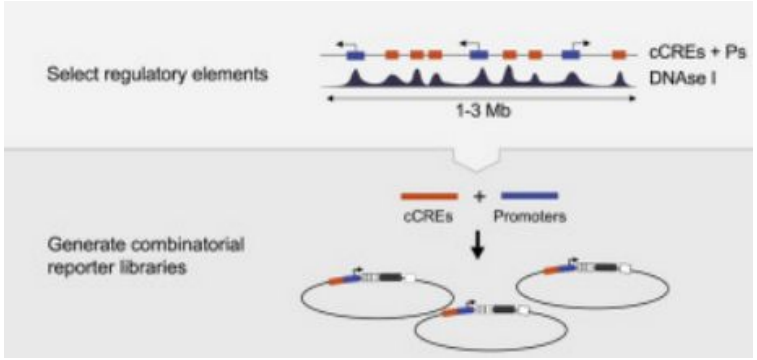
Molecular Cell

Volume 82, Issue 13, 7 July 2022, Pages 2519-2531.e6

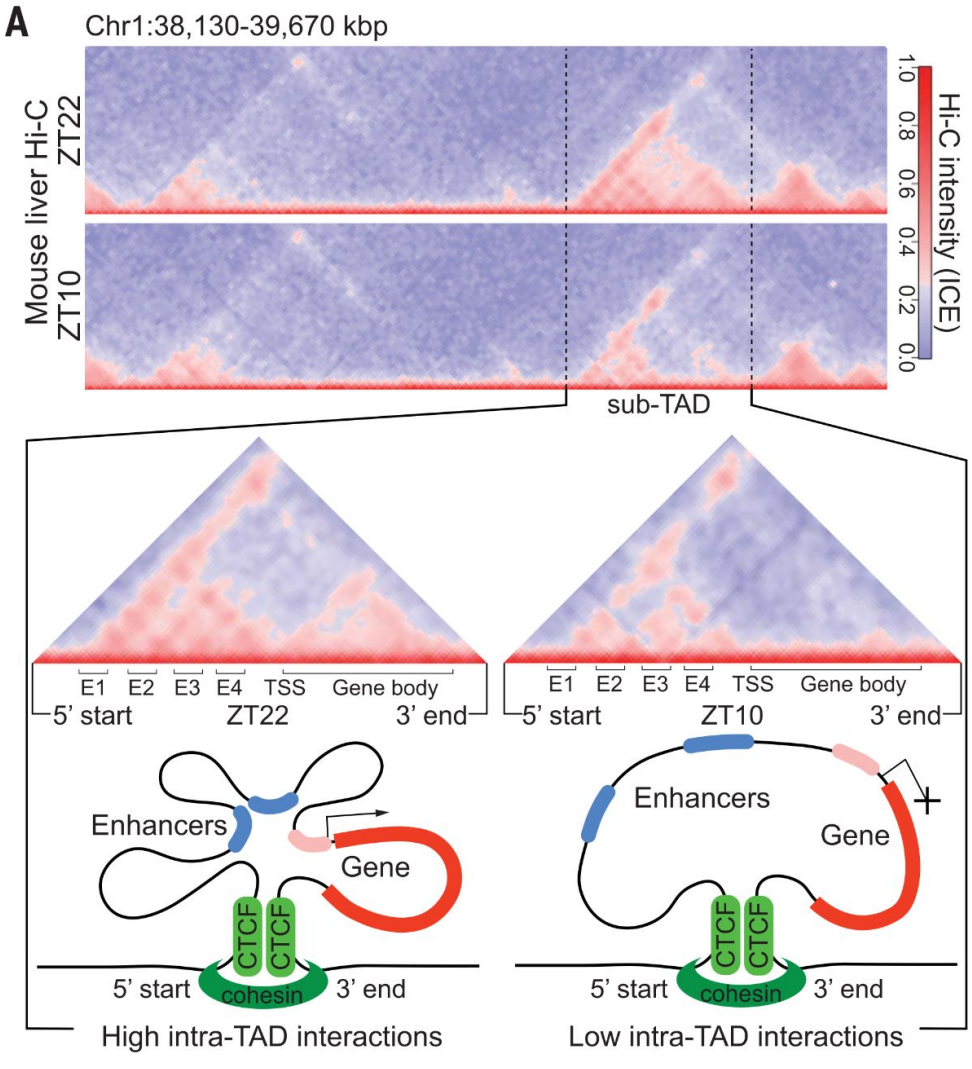
Article

## Systematic analysis of intrinsic enhancer-promoter compatibility in the mouse genome

Miguel Martinez-Ara [1], Federico Comoglio [1 2], Joris van Arensbergen [1 3], Bas van Steensel [1 4]




- Even when enhancers are made to come in proximity to a promoter, **not all enhancers can regulate all promoters**
- E-P compatibility (i.e. the specificity of the interactions) appears to be dictated by TF interactions between the two domains




# Changes in chromatin contact occur within stable TADs/subTADs

Here an example driven by the circadian rhythm (mouse liver, ZT22=5am ZT10=5pm)

(Kim et al., Science 2018)

# A few take-home messages

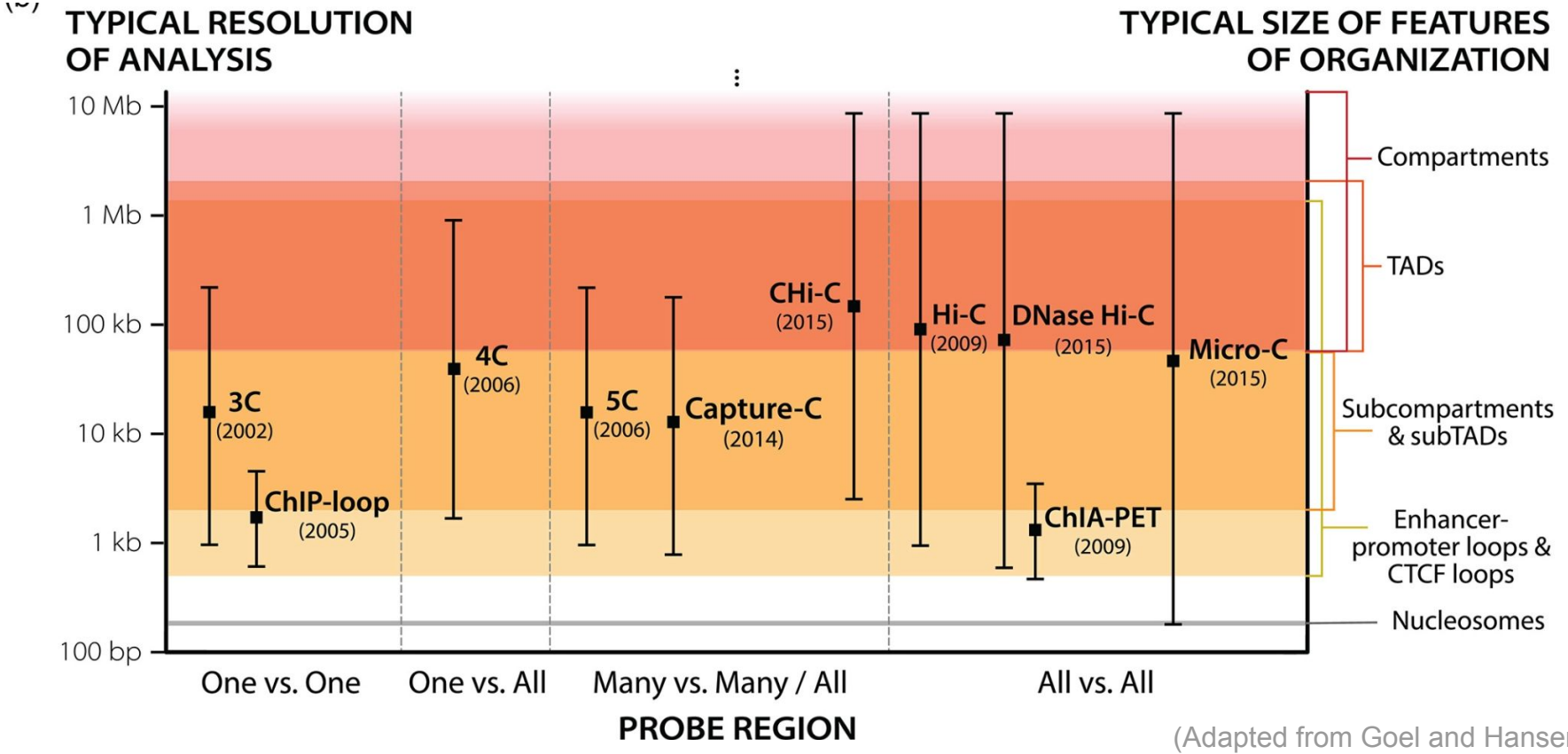
- the nucleus is organized into chromosome territories,
  and chromosomes are organized into active/inactive compartments

- chromosomes are further organized into TADs that are
  - largely stable across cellular states and
  - are delimited by CTCF dimers (loops)

- chromatin loops bring regions of chromatin in proximity,
  - enabling the formation of enhancer-promoter interactions,
  - but are not required for the short-term *maintenance* of these interactions

- Functional enhancer-promoter interactions appear determined by TF interactions, made possible by loops

# Sequencing-based conformation capture




(Adapted from Goel and Hansen 2020)

(See also Hun Lee, Wu and Rhie, Epigenetics & Chromatin 2022 for a review)

# Data analysis - some references

**Method & analysis refs**

- Hi-C:
  - Juicer
- HiChIP:
  - Dovetail doc
  - FitHiChIP

**Main output & formats**

- Contact maps
- TADs/ compartments:
  - bedpe (?)
- Interactions:
  - bedpe / pairix

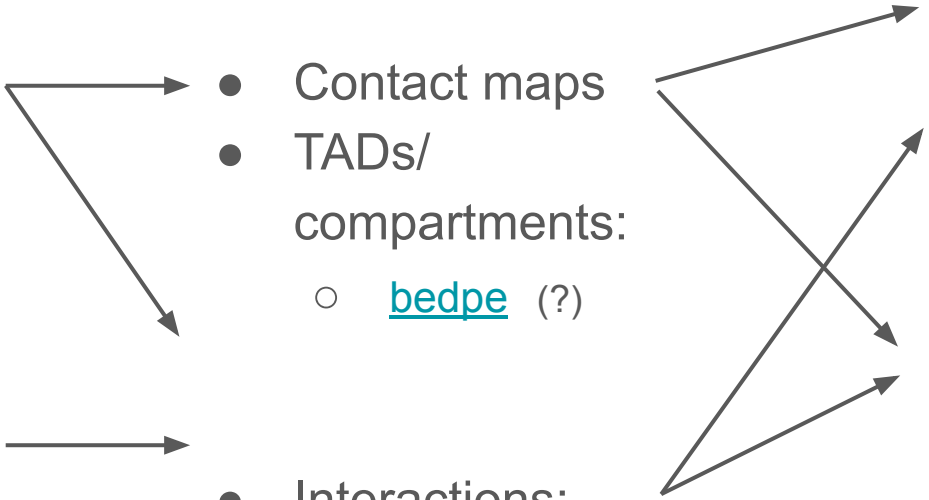
**Visualization in R**

HiCBricks

plotgardener

**Data structure in R**

interactionSet



# Practical

# Enhancer-promoter interactions: what to do with distal regulatory elements?

- Enhancers within <1kb of a TSS nearly always regulate that TSS (Sanyal et al., Nature 2012; Morris et al., Science 2023),
- Beyond 2.5kb, more than half of the interactions are not with the closest gene (Sanyal et al., Nature 2012).

# Enhancer-promoter interactions: what to do with distal regulatory elements?

- Several types of information can be used to link such distal regulatory regulatory elements to target genes:
  - **distance**
  - chromatin **conformation** data (e.g. Fulco et al., Nat Gen 2019; Salviato et al., NAR 2021) → requires very high-depth data for the cell-type of interest
  - expression Quantitative-Trait-Loci (**eQTLs**) (e.g. Fishilevich et al., Database 2017) → depends on well-represented genetic variation profiled in the right cell types
  - statistical **association** (e.g. correlation) between signals (e.g. accessibility or histone modification) at the distal enhancer and gene expression (e.g. Hait et al., NAR 2022)
    → requires a lot of data with sufficient variation → single-cell multi-omic data!

# Enhancer-promoter interactions: what to do with distal regulatory elements?

- For the assignment, you'll work with something easy to get and to use, namely :

  Generic (i.e. non-celltype-specific) enhancer-target predictions based on both association and conformation,

  produced by Salviato et al., NAR 2021

# Assignment

- Choose a transcription factor (e.g. p300), and obtain peaks from ENCODE (ChIP-seq in a human context!)
- Isolate the peaks that are:
  - Between 2.5kb and 10kb from a TSS
  - More than 10kb from a TSS
- For each set of peaks:
  - Subset to those peaks that have a predicted distal target(s) using Salviato et al. (2021)
    - You can download a GRanges of those interactions at https://ethz-ins.org/content/hg38.SalviatoDistalEnhancerTargets.GR.rds
  - Find the nearest TSS for each peak
  - In what proportion of the cases is the predicted target the closest gene?
- Hints:
  - you can use the `annotateRegions` function, as we did in week 4, to get the gene nearest to each peak
  - beware not to count, when calculating proportions, peaks that don't have interactions with any TSS!
- Expected for of the answer:
  - "Of the genes that are between 2.5 and 10kb from the nearest TSS, XX % form an interaction with that nearest gene. Of the genes that are more than 10kb away from the nearest TSS, XX % form an interaction with that nearest gene."