Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L - 2022 | week 11

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: strings and objects as arguments

```
> promoterRegions <- readRDS("data_practical/promoterRegions.rds")
> getMeth(bsseqEx, regions="promoterRegions", type="smooth", what="perRegion")


> getMeth(bsseqEx, regions=promoterRegions, type="smooth", what="perRegion")


> ?getMeth
```

Usage

```
getMeth(BSseq, regions = NULL, type = c("smooth", "raw"),
  what = c("perBase", "perRegion"), confint = FALSE, alpha = 0.95,
  withDimnames = TRUE)
```

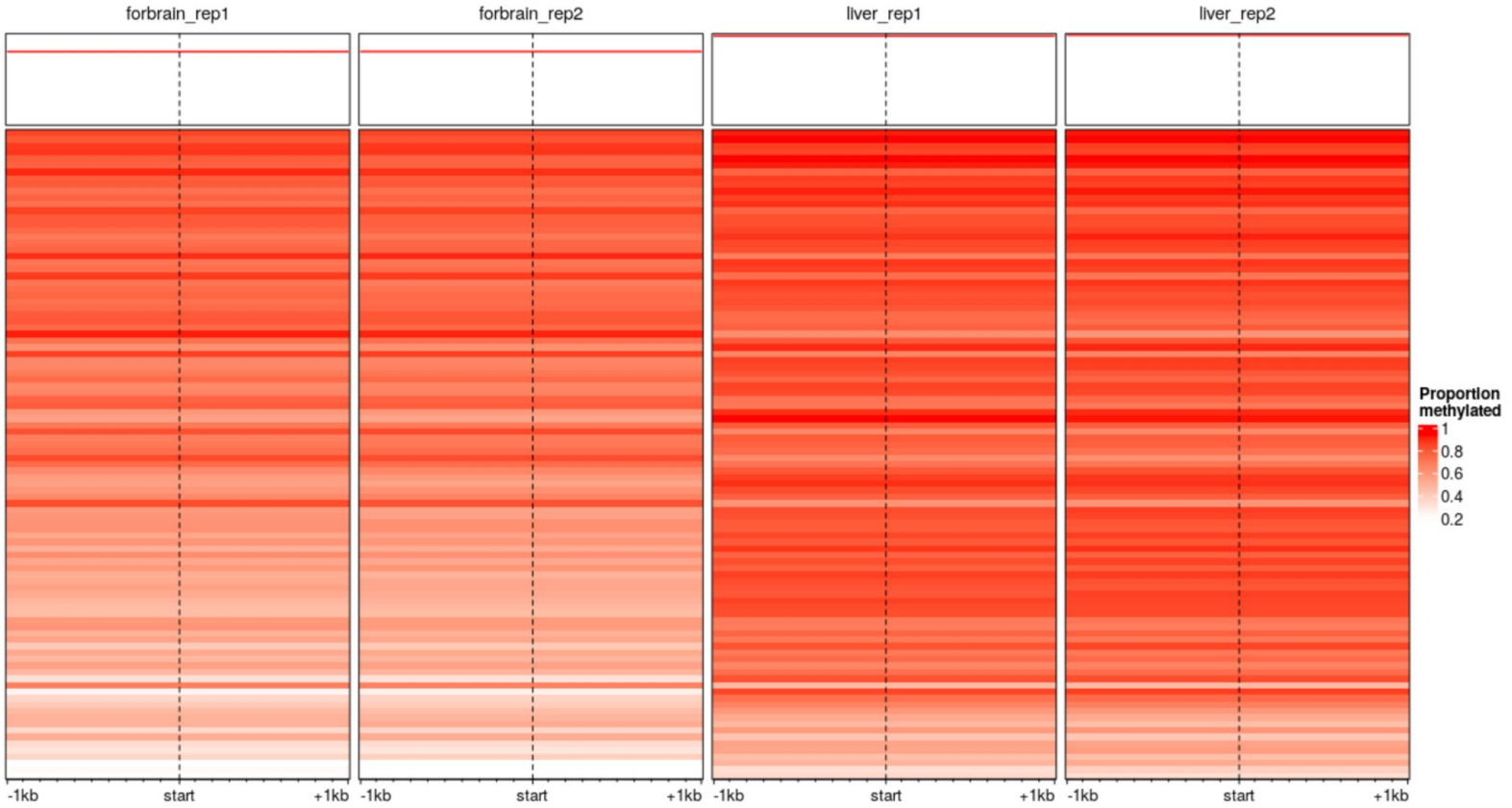
Arguments

BSseq	An object of class BSseq.
regions	An optional data.frame or GenomicRanges object specifying a number of genomic regions.
type	This returns either smoothed or raw estimates of the methylation level.
what	The type of return object, see details.

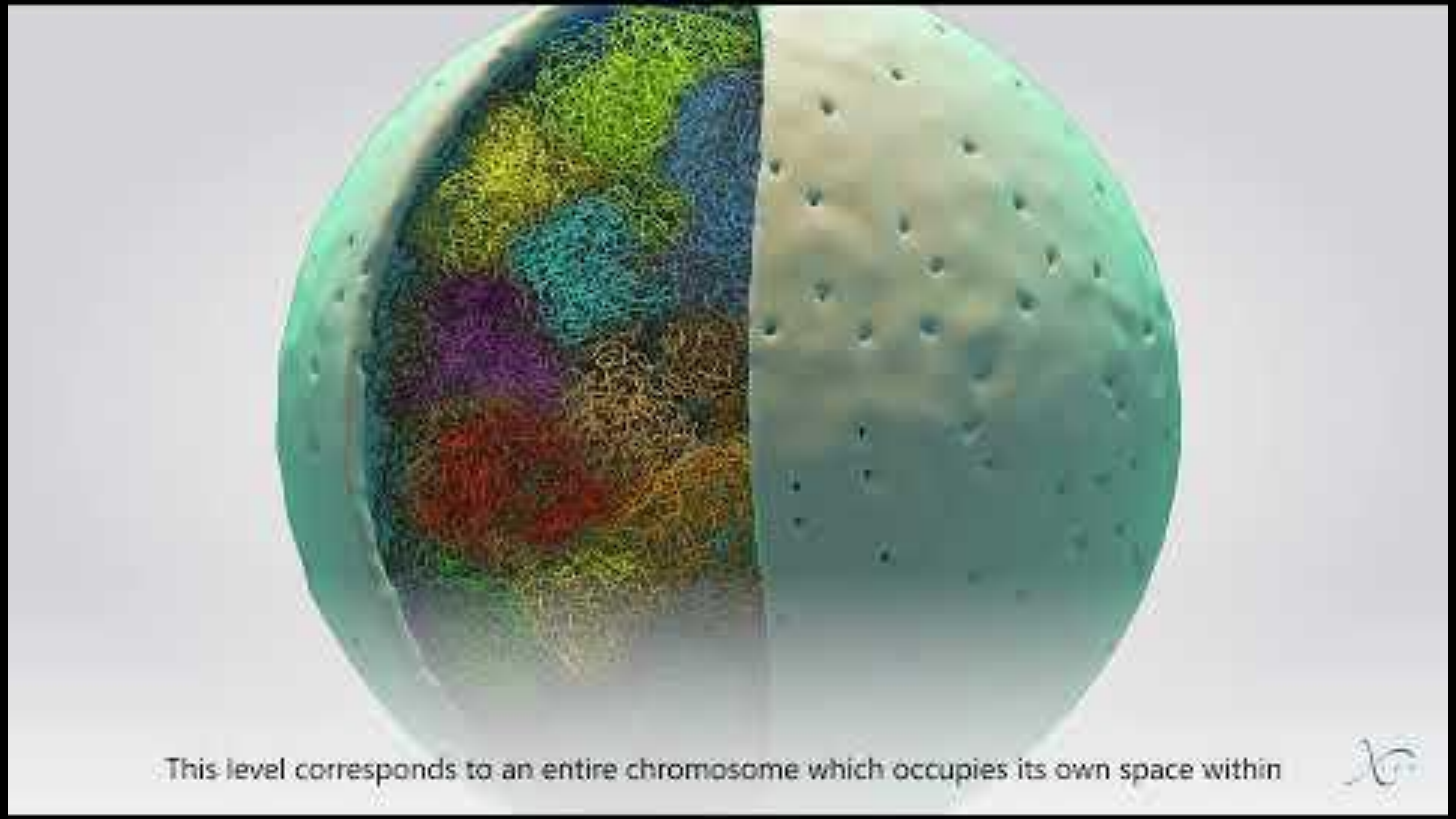
Debriefing on the assignment: getting the signals

```
# For each sample, we extract the methylation scores of CpGs in our regions of interest:
metRaw <- bsseq::getMeth(bsseqOb, regions=dmrRanges, type="raw", what="perRegion")
metRawFbRep1 <- GRanges(seqnames=seqnames(dmrRanges),
                        ranges=ranges(dmrRanges),
                        strand=strand(dmrRanges),
                        score=metRaw[,"E13_5_rep1.bed"])

# we then give the list of objects to signal2Matrix:
tracks <- list("forbrain_rep1"=metRawFbRep1,
               "forbrain_rep2"=metRawFbRep2,
               "liver_rep1"=metRawLiverRep1,
               "liver_rep2"=metRawLiverRep2)

signal2Matrix(tracks, dmrRanges, extend=0, w=20, type="scale")

# then we plot…
```

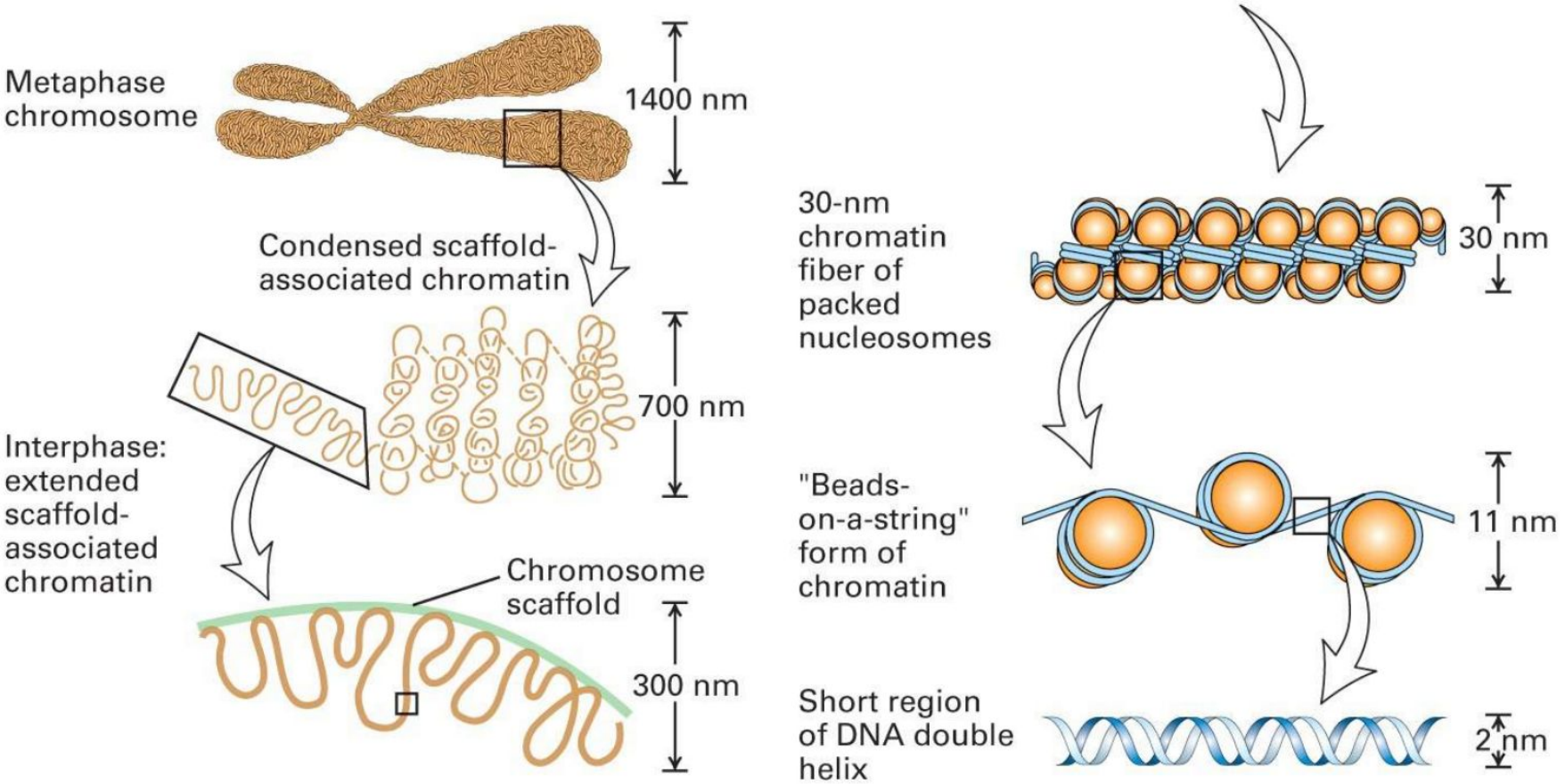
Debriefing on the assignment: result



3D organization of the genome

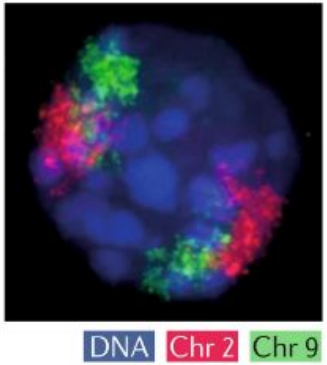


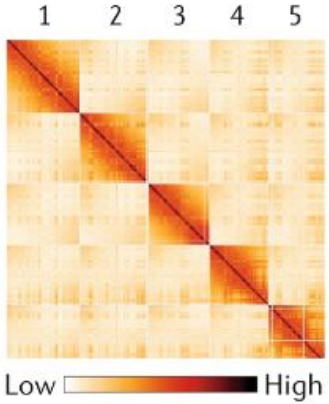
https://youtu.be/Pl44JjA--2k

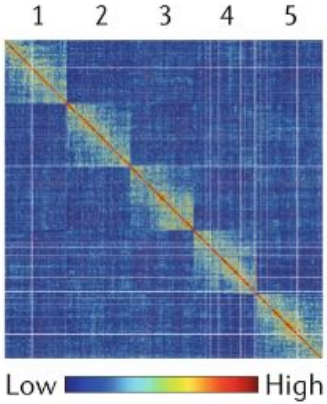


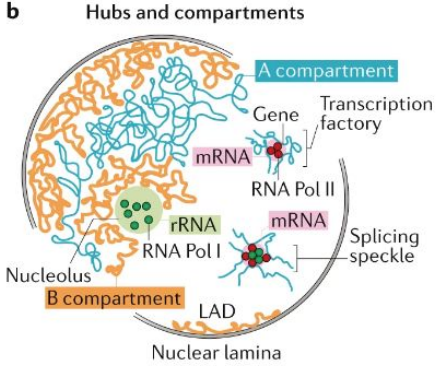
Lodish et al. *Molecular Biology of the Cell* (5th ed.). W.H. Freeman & Co., 2003.

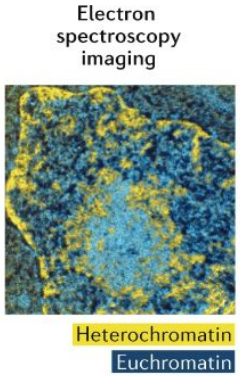
The nucleus is organized into chromosome territories







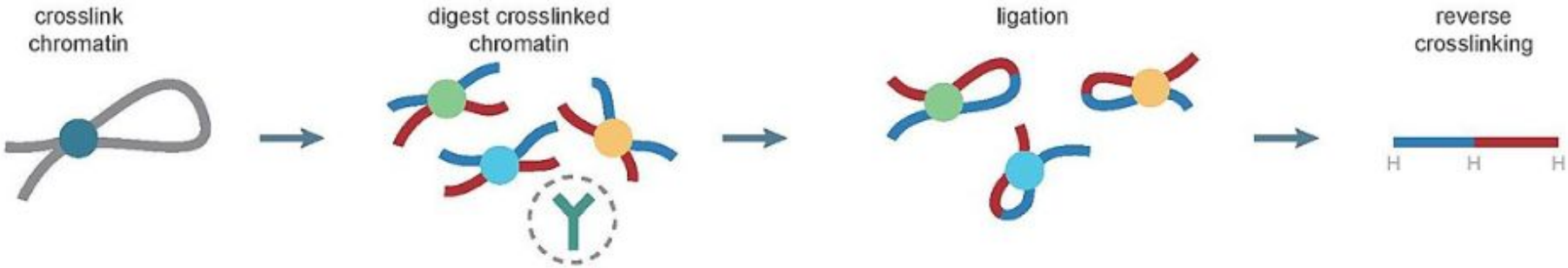




Chromosomes are split into active (A) and inactive (B) compartments

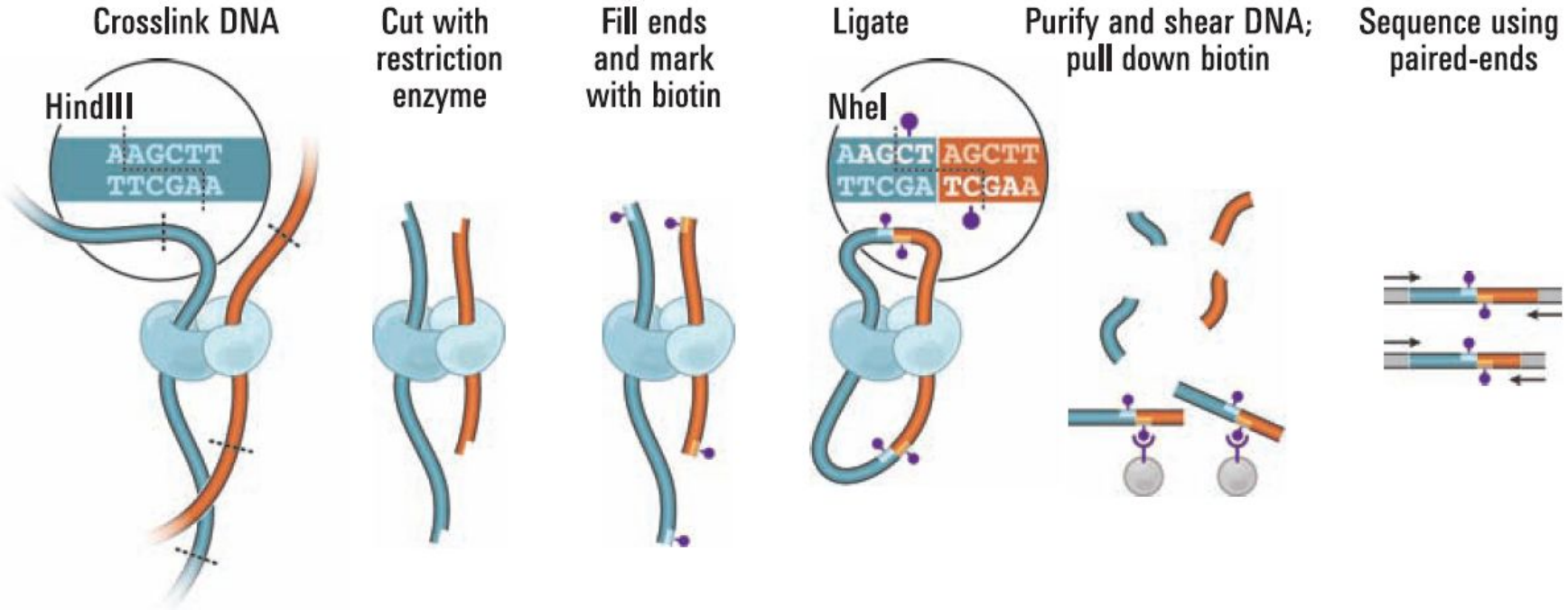
(Adapted from Kempfer and Pombo 2020)

Sequencing-based conformation capture



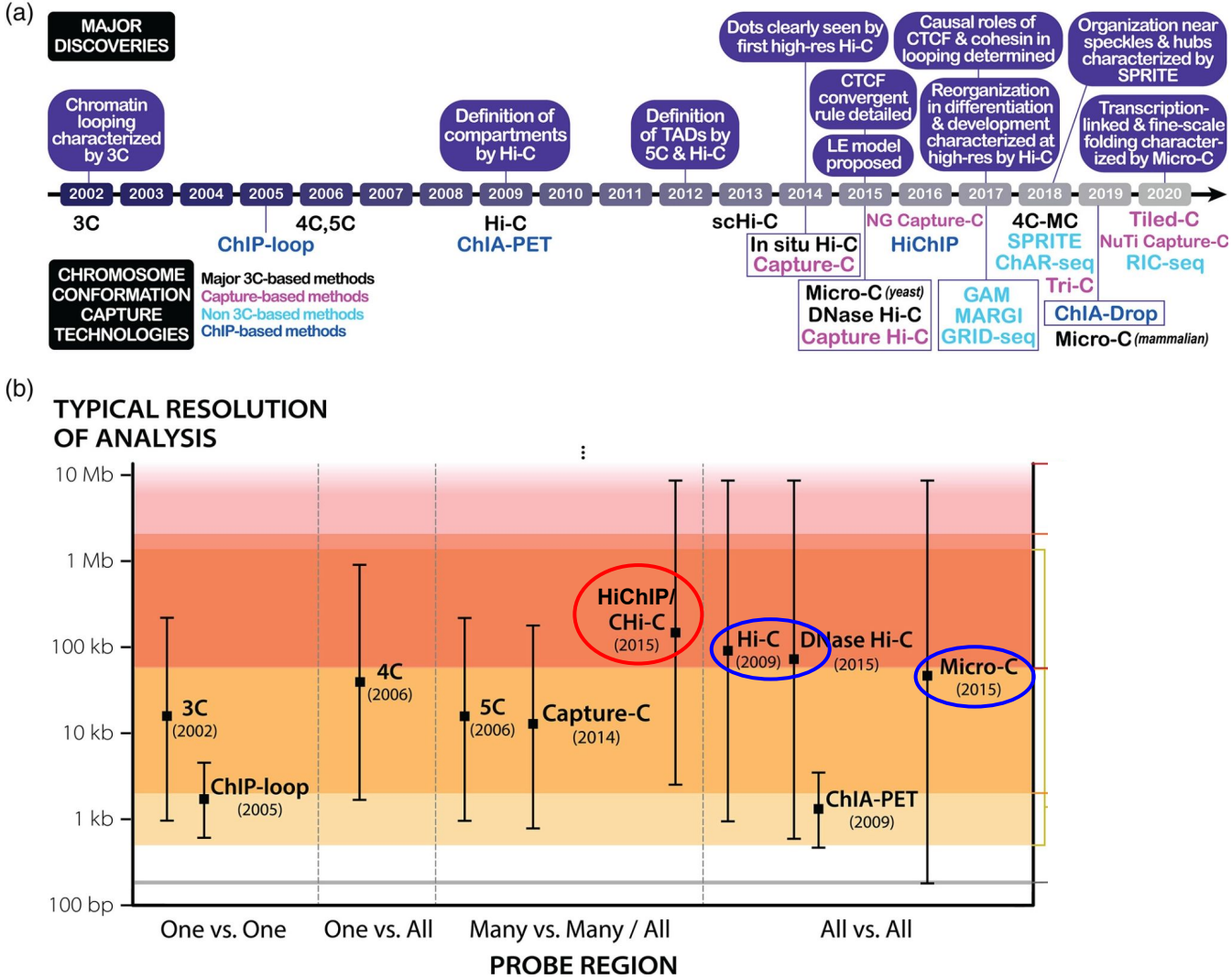
(Adapted from Li et al., 2014)

Sequencing-based conformation capture



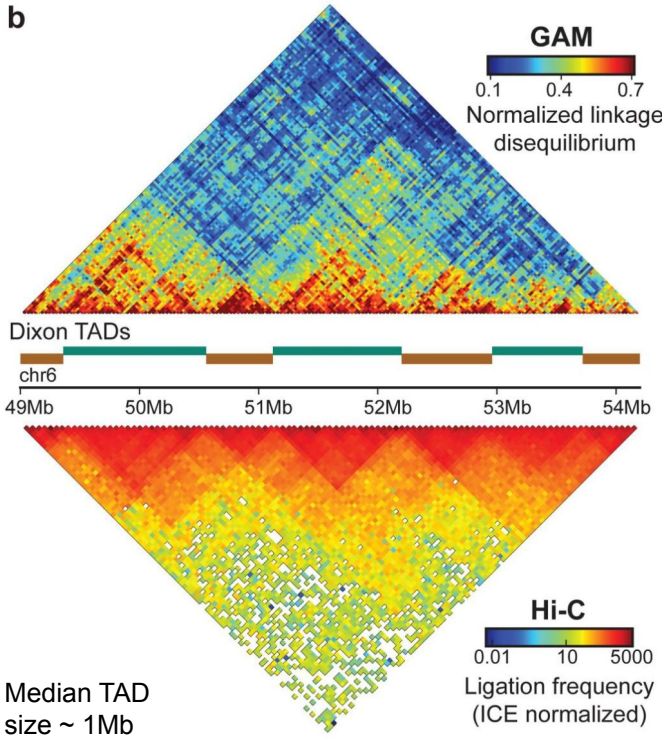
(Lieberman-Aiden et al. 2009)

Sequencing-based conformation capture



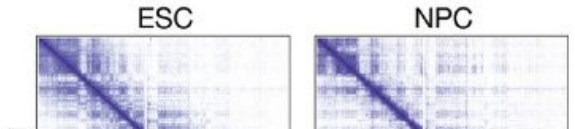
(Adapted from Goel and Hansen 2020)

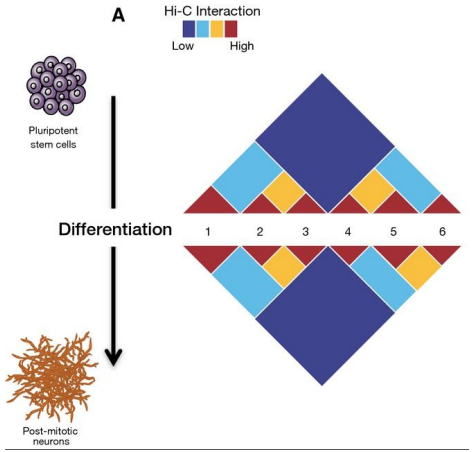
Chromosomes are organized into topologically associated domains (TADs)



Median TAD size ~ 1Mb

(Beagrie et al., 2017)

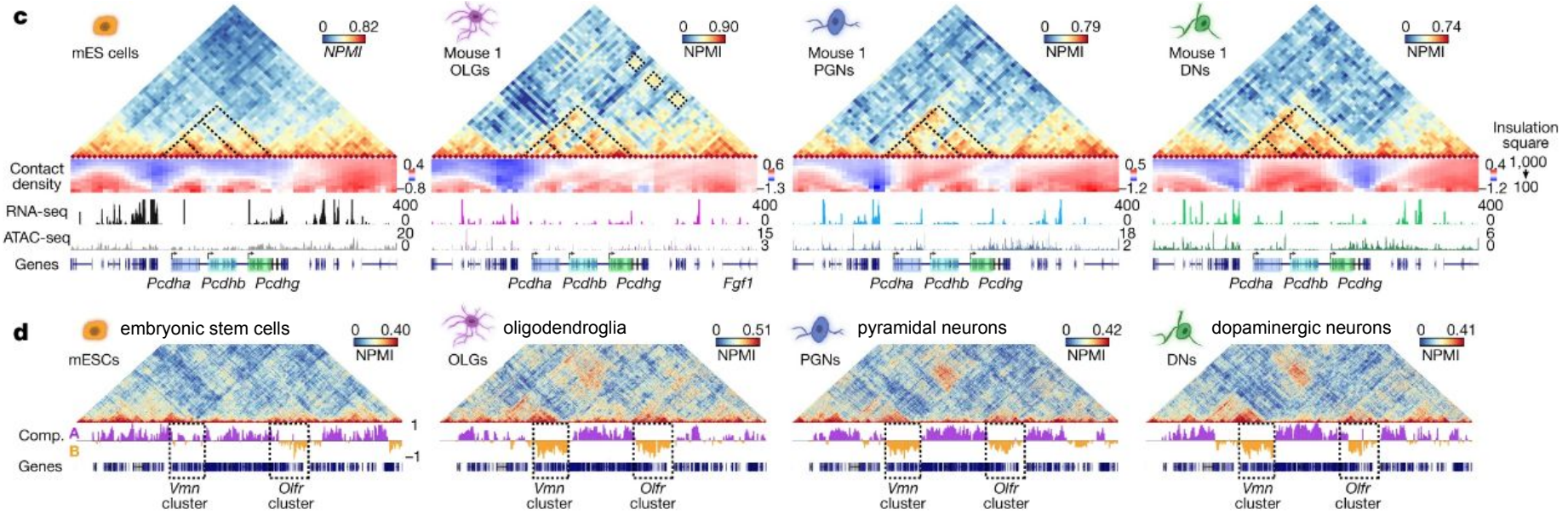




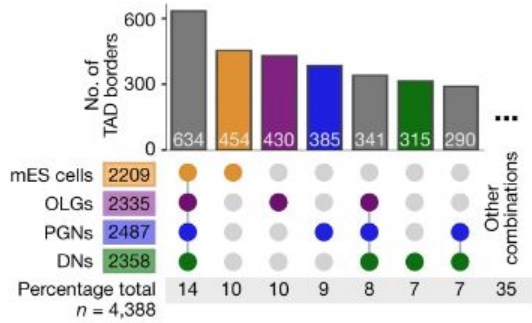
(Fraser et al., 2015)

...that are rather stable across cell types

TADs rearrangement across cell types



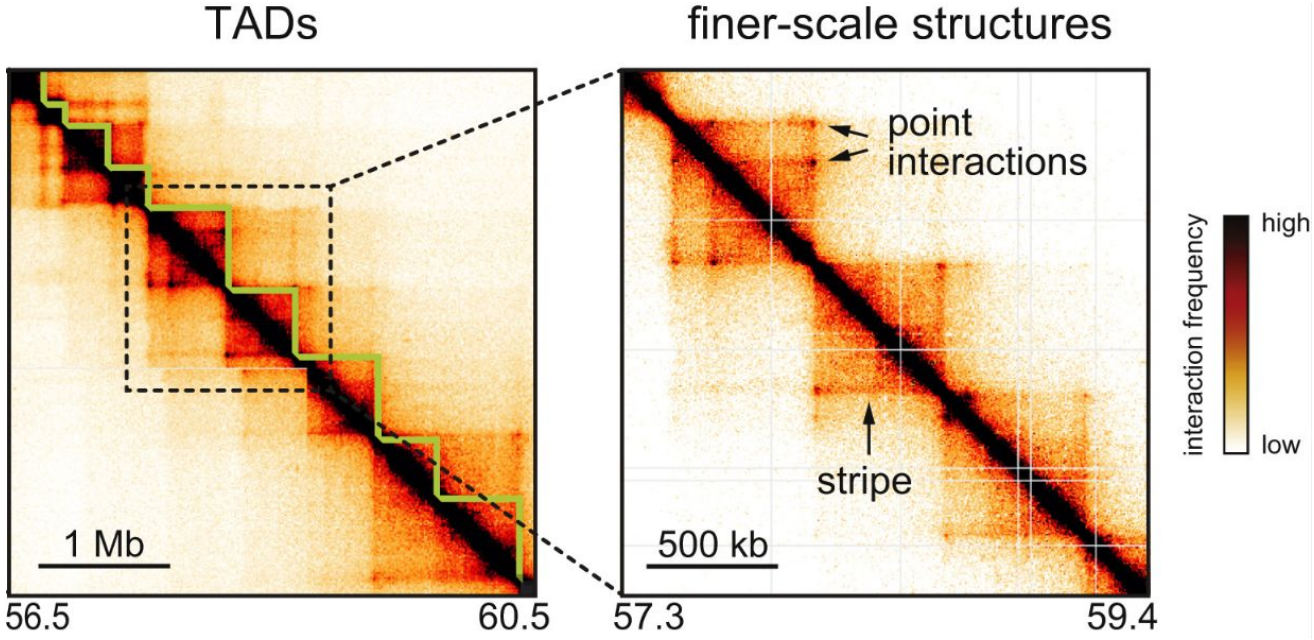
Comparison of TADs borders:



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

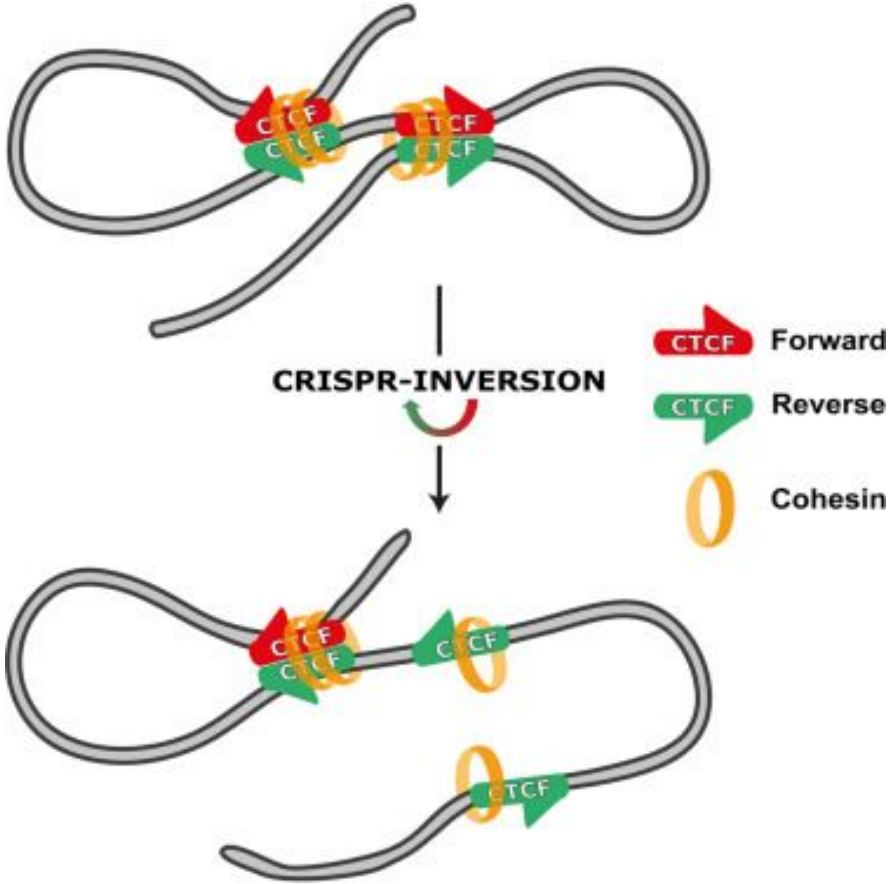
TAD boundaries are defined by very stable point interactions…

…which represent CTCF binding sites



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

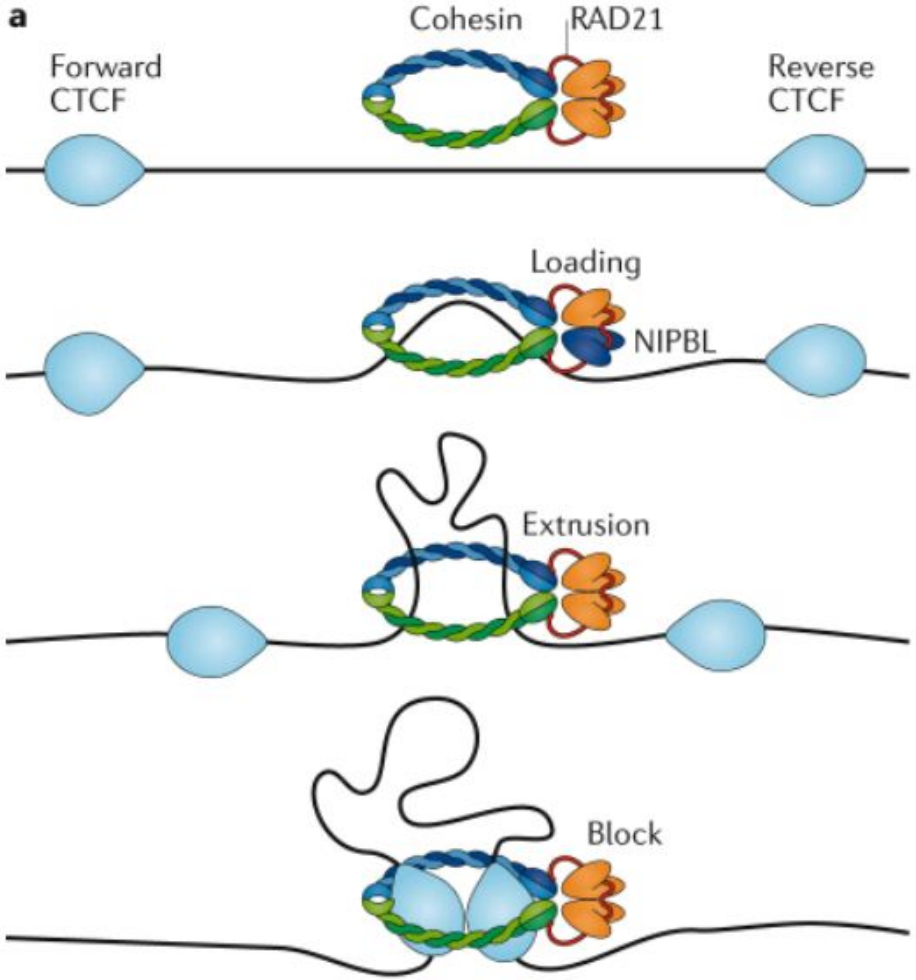
CTCF forms **convergent** dimers at loops



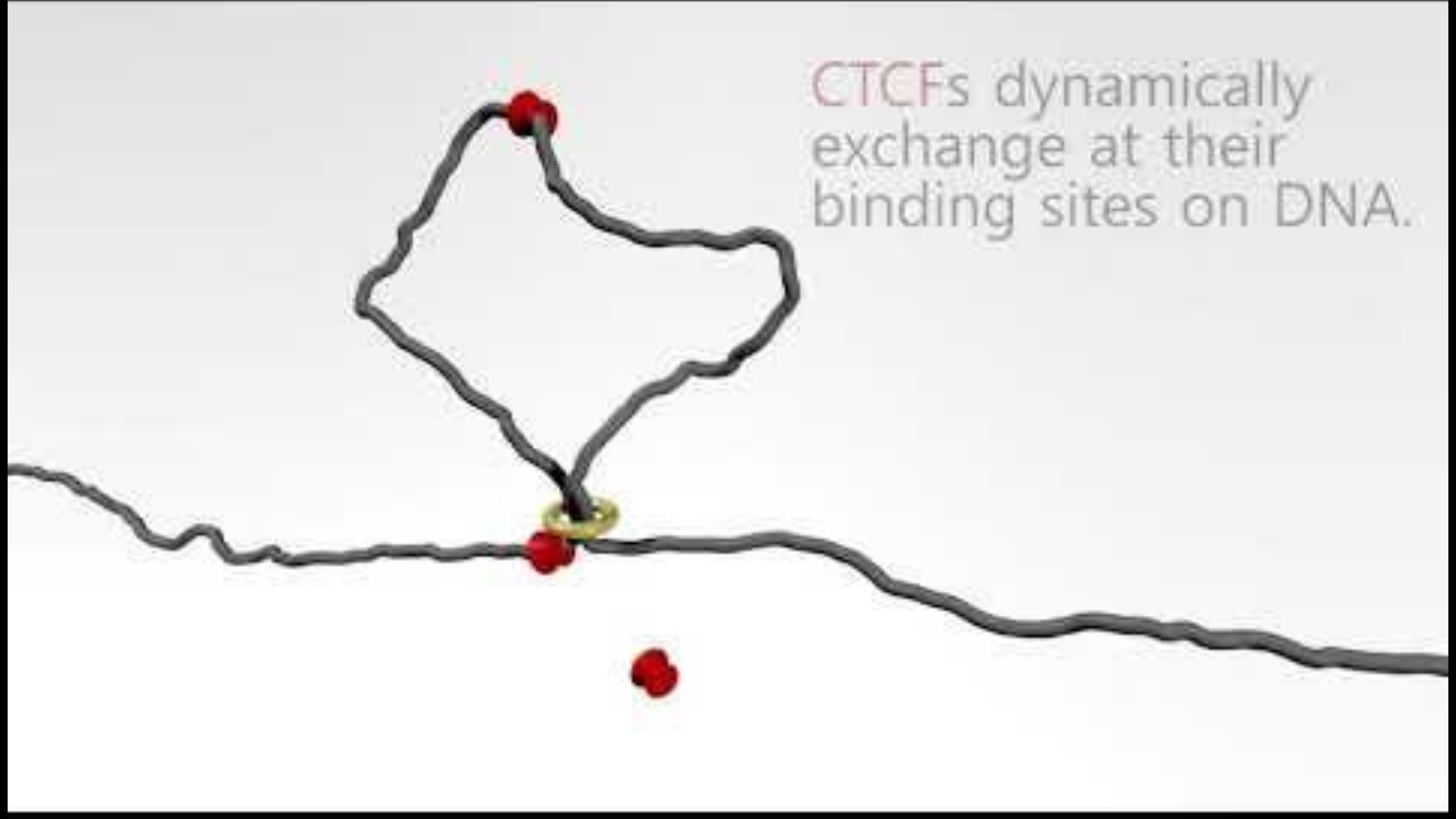
(Adapted from de Wit et al., Mol Cell 2015)

The loop extrusion model

CTCF dimers form loops
by blocking Cohesin

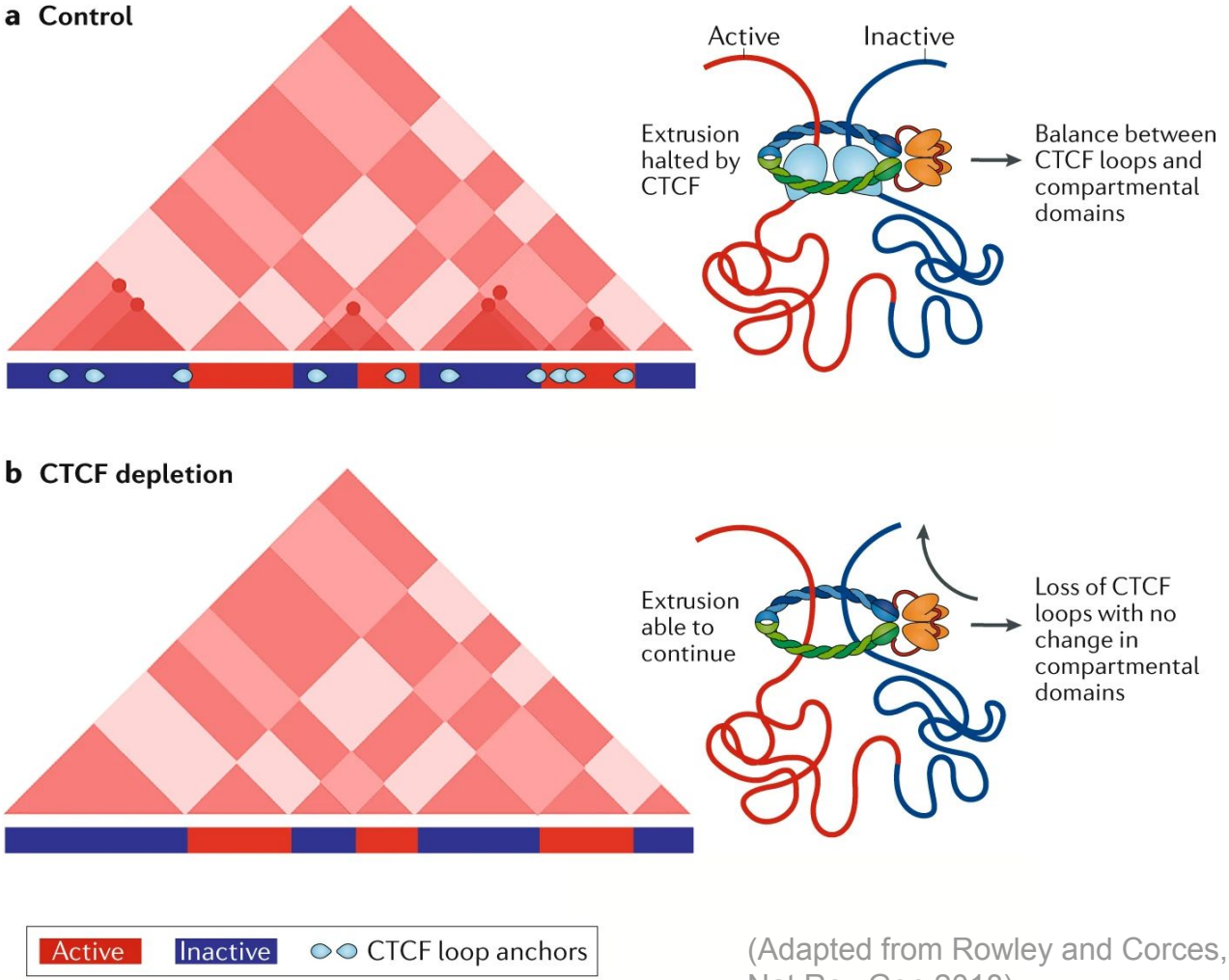


(Adapted from
Rowley and Corces, Nat Rev Gen 2018)

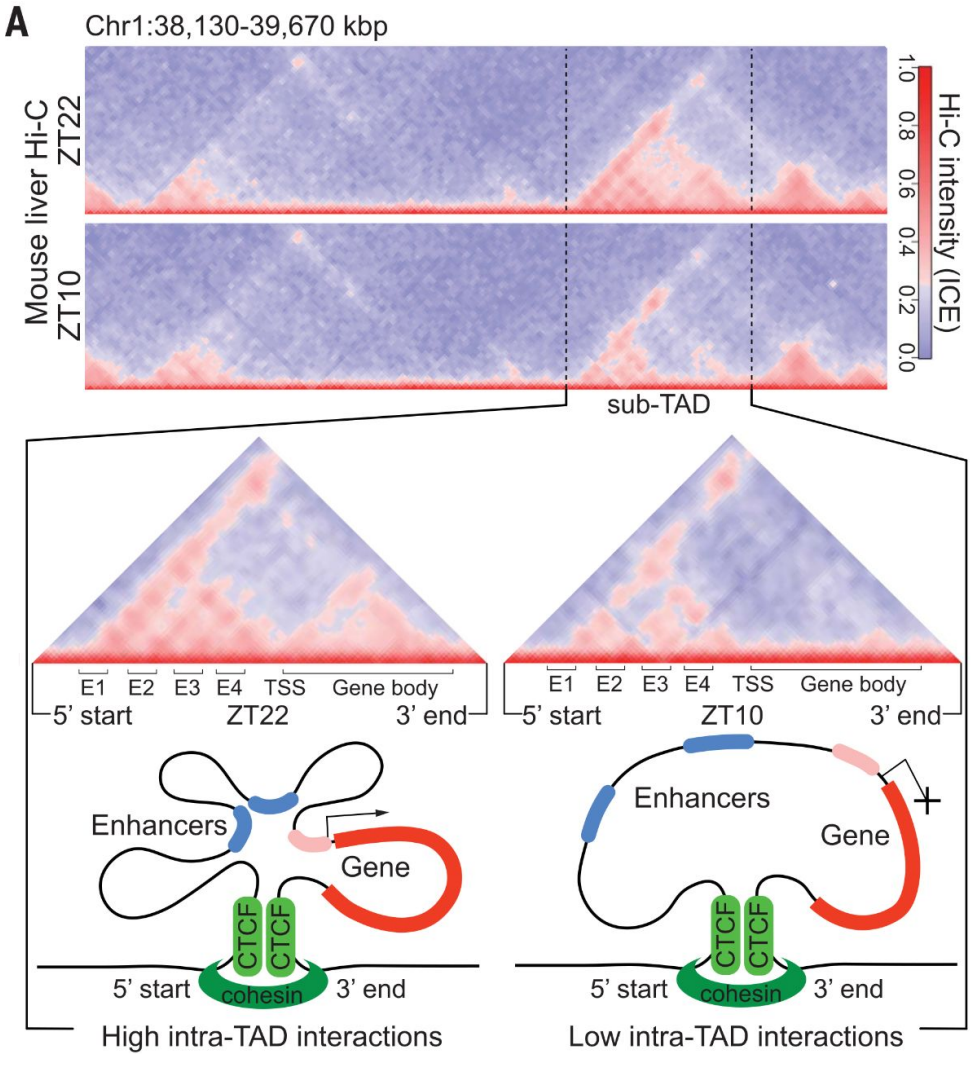

CTCFs dynamically
exchange at their
binding sites on DNA.

CTCF depletion disrupts especially TAD-internal structures

e.g. including promoter-enhancer loops



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

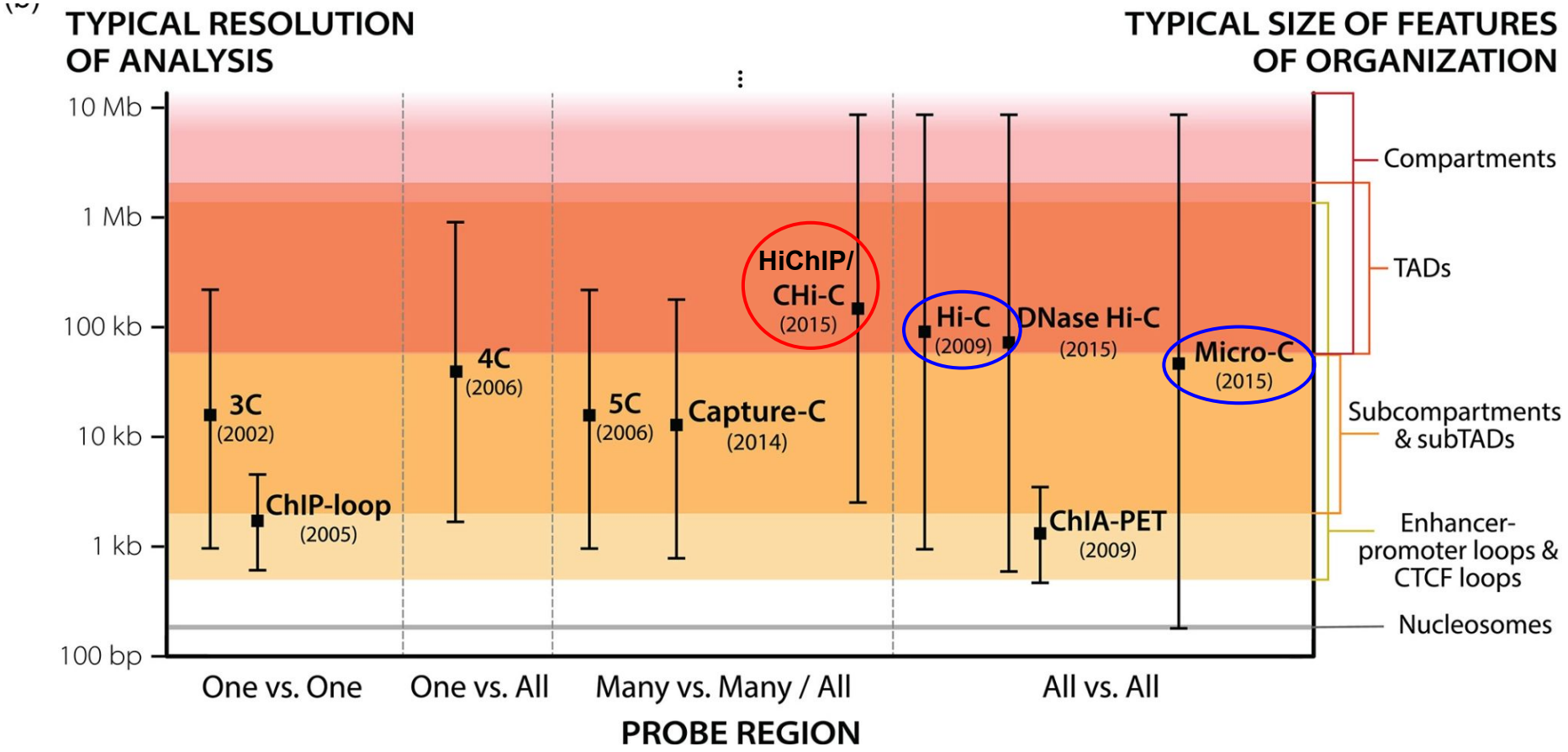


Changes in chromatin contact occur within stable TADs/subTADs

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

(Kim et al., Science 2018)

Sequencing-based conformation capture



(Adapted from Goel and Hansen 2020)

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

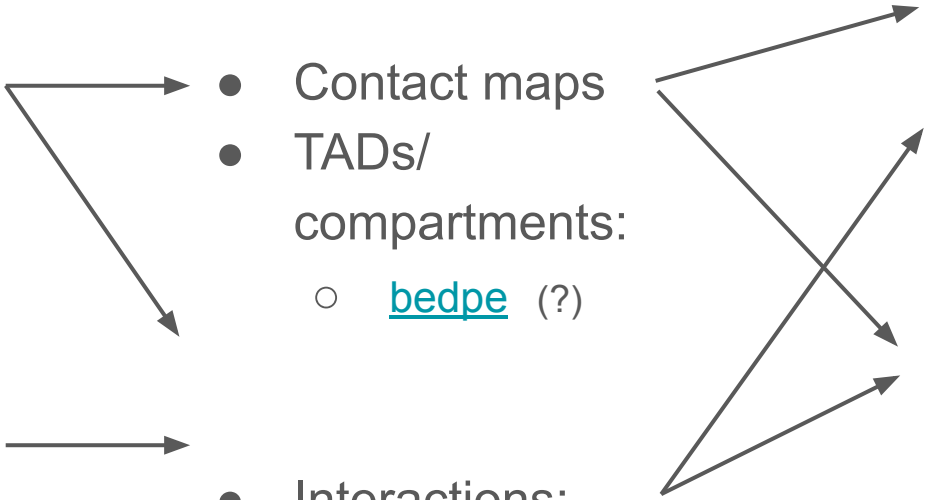
Visualization in R

HiCBricks

plotgardener

Data structure in R

interactionSet



Assignment

- Choose a transcription factor (e.g. p300), and obtain peaks from ENCODE
  - make sure it's the same cell line, i.e. A549, or that Hi-C interactions are available for your cellular 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, identify those that are in contact with a TSS using long-range interactions from ENCODE
- For each set, what proportion of the interactions are with the nearest gene?
- Hint 1: you can use the `annotateRegions` function, as we did in week 4, to get the gene nearest to each peak
- Hint 2: 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."