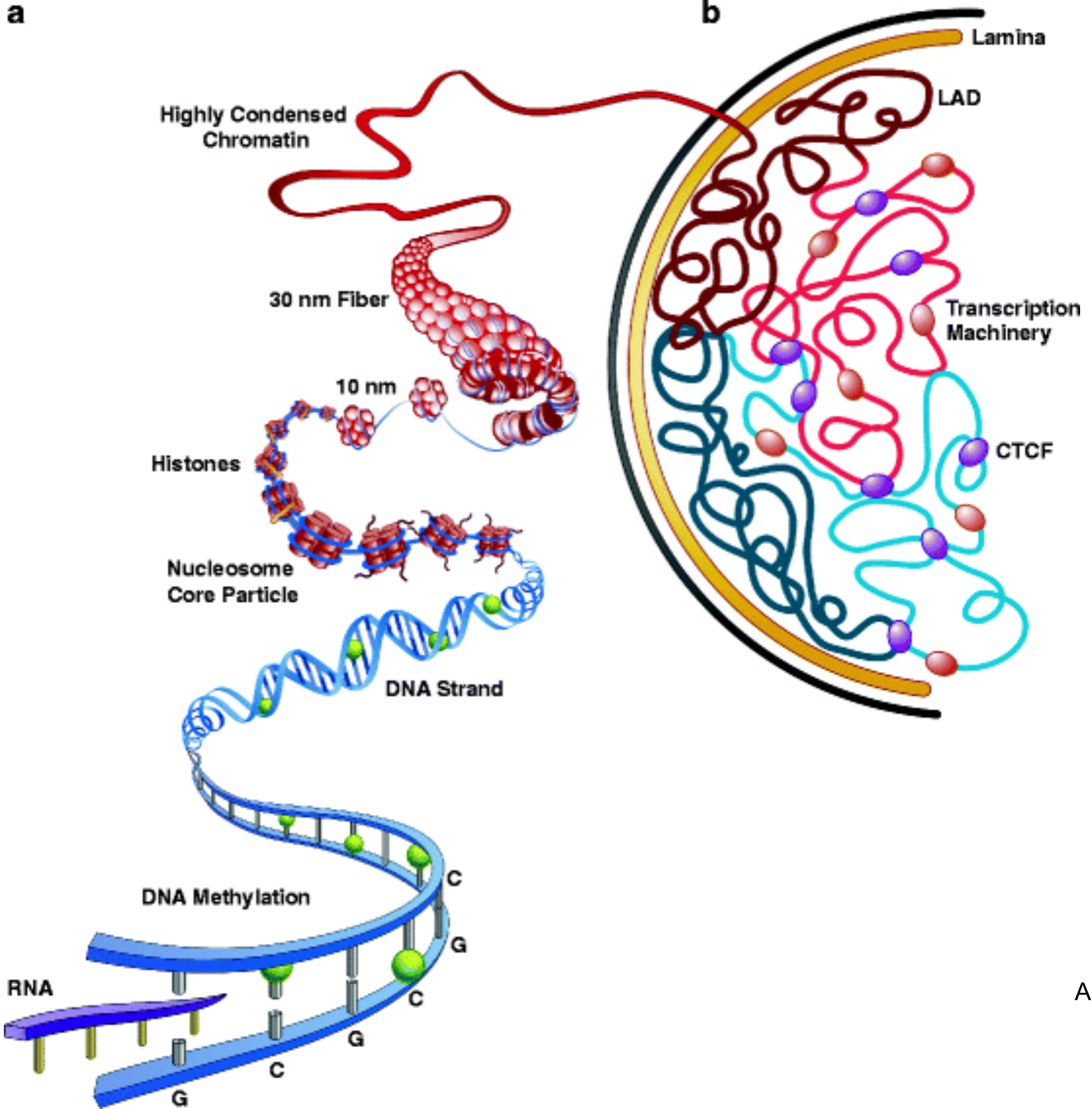
Scaling and Complexity of Genome Structure

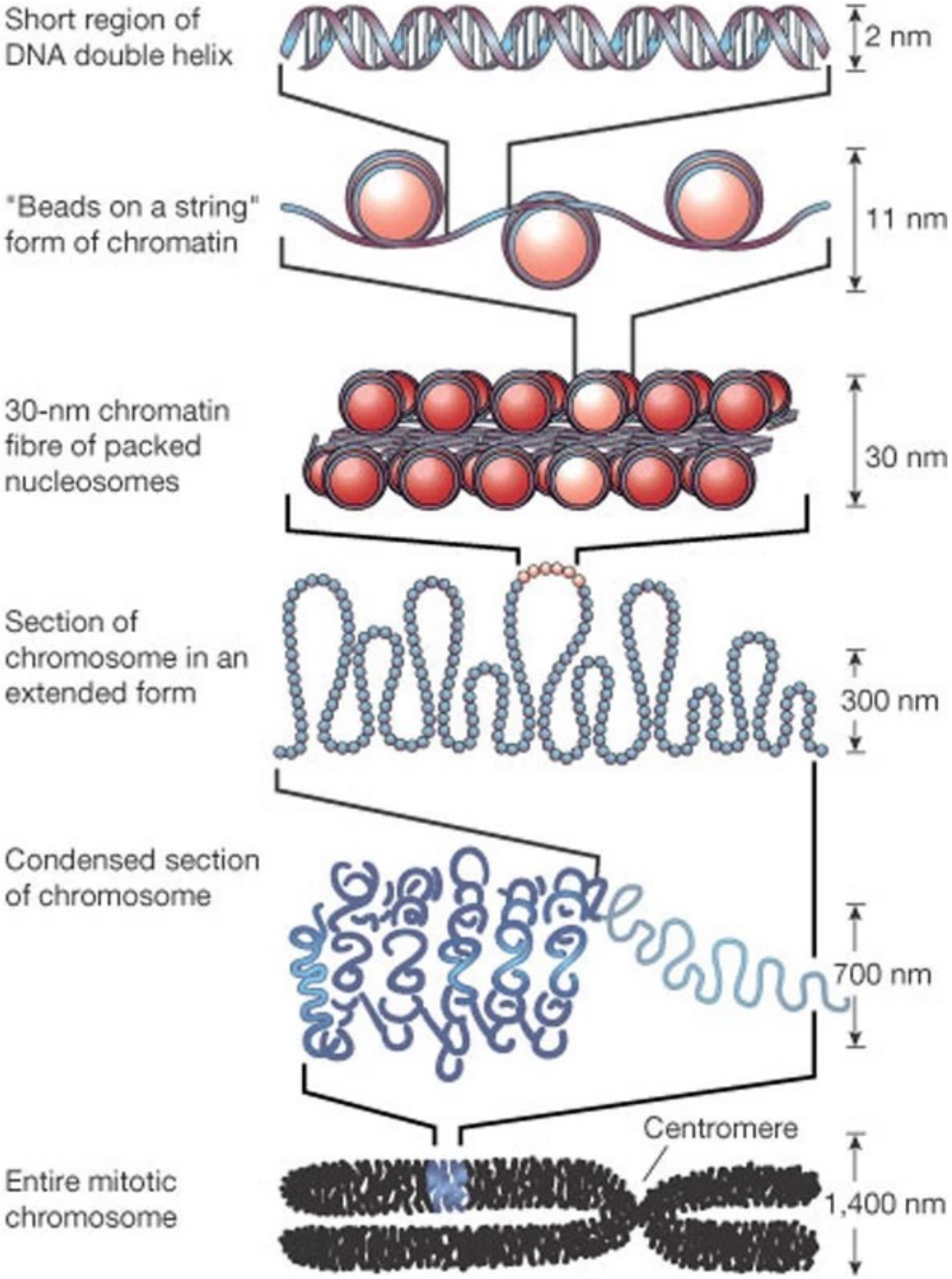
November 10, 2025

Outline

- Broad histone domains
- ChIP-seq signal detection: SICER and RECOGNICER
- Power law and “scale free”
- Hi-C
- Fractals

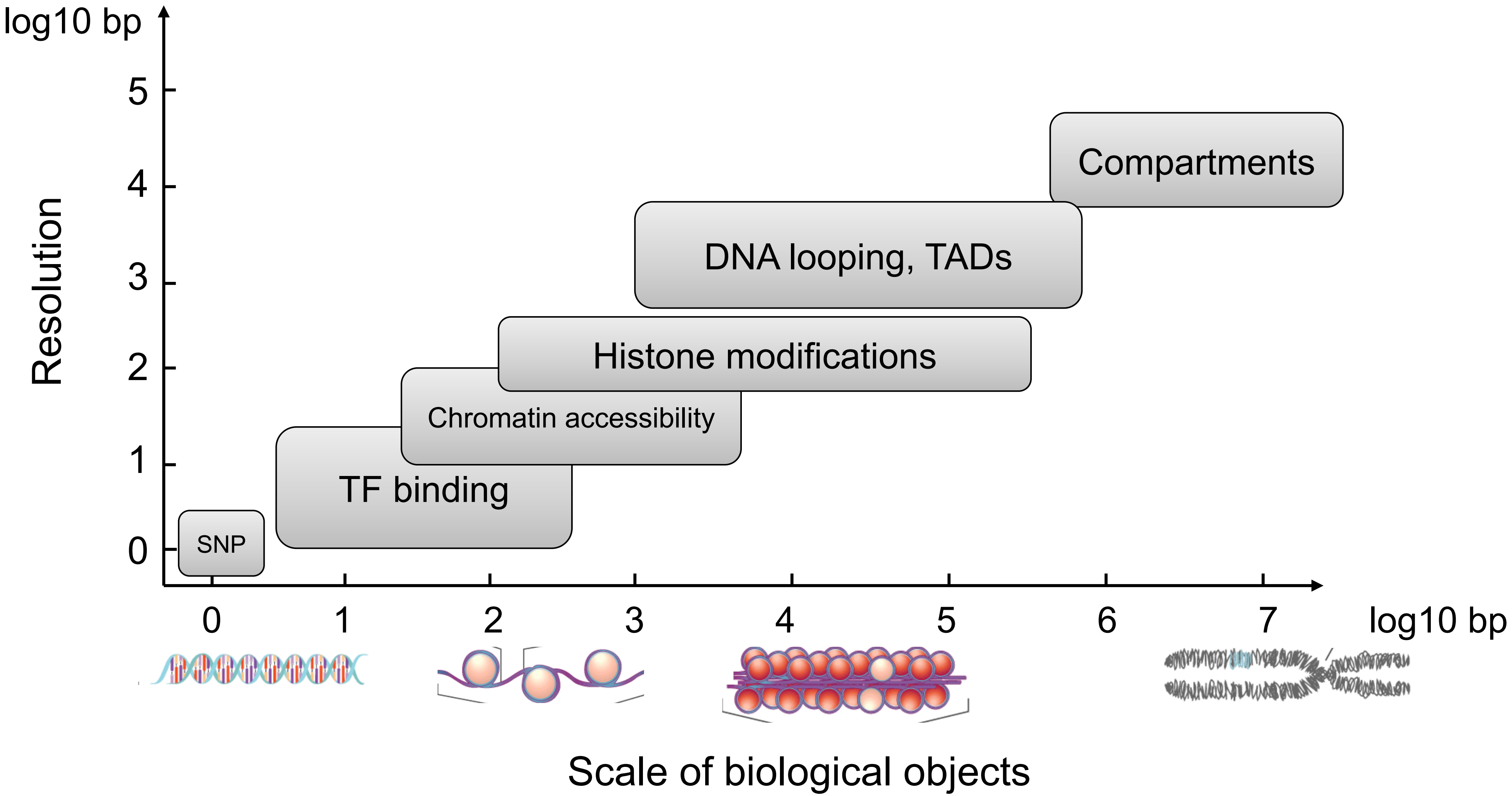


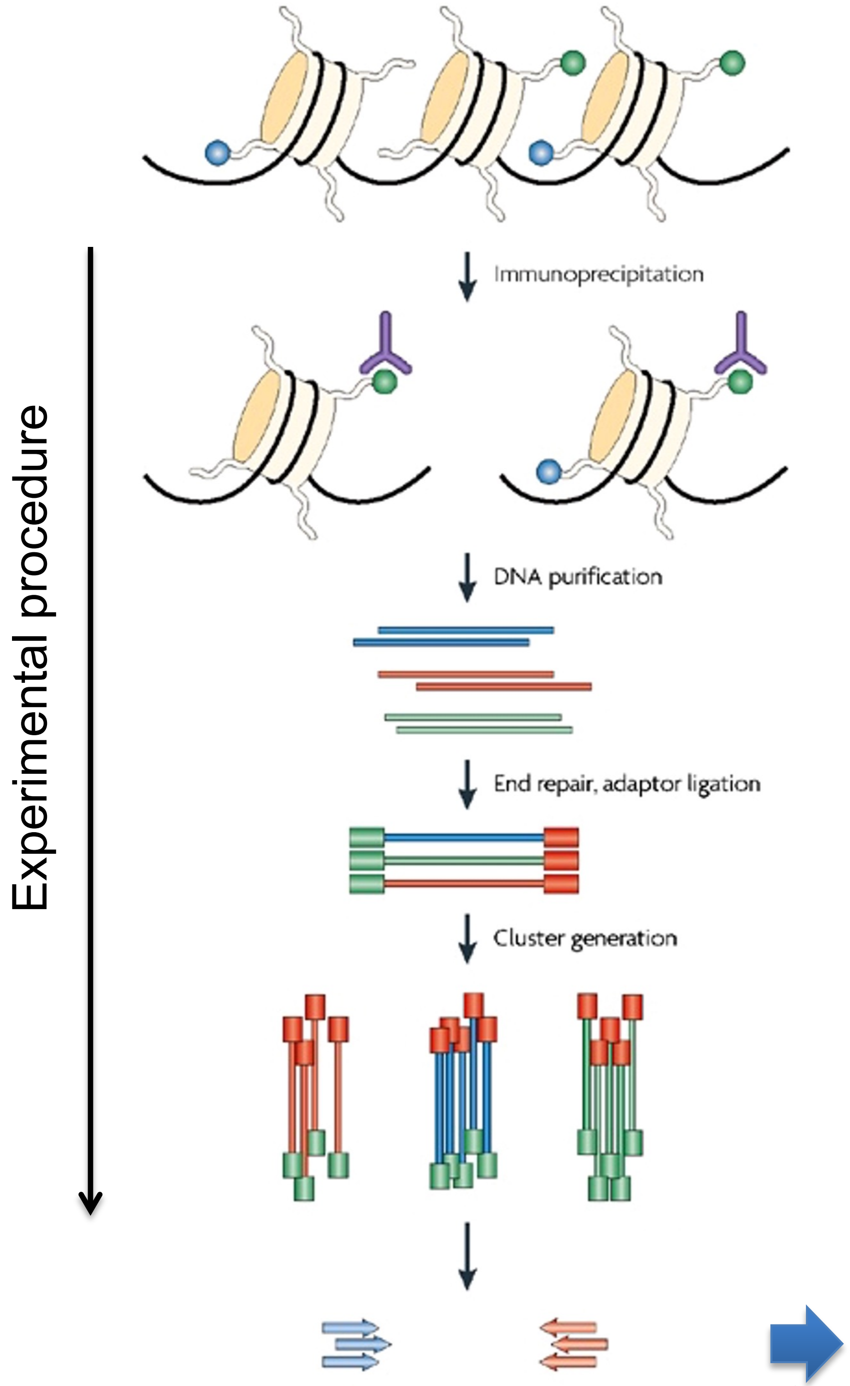
Adomas & Wade 2013



Felsenfeld & Groudine *Nature* 2003

Functional genomic/epigenomic elements range across scales





Biology

↑ *downstream analysis,*
data integration

Peaks (bed)

↑ *peak calling*

Pile-up for visualization
(bedGraph, wig, bigwig)

Non-redundant reads (sam/bam/bed)

↑ *redundancy assessment*

Mapped reads (sam/bam/bed)

↑ *alignment* (**bowtie2/BWA**)

Raw sequence reads (fastq)

macs/
SICER

Computational analysis

Transcription Factors vs. Histone Marks

	DNA-binding proteins (Transcription factors)	Histone Marks (Histone modifications, histone variants, chromatin regulators*)
Cell type specificity	Both factor and profile	Profile
Signal width ("peak size")	Narrow	Narrow or broad
Chromatin accessibility	High	High or low
DNA sequence motif	Yes	No
Resolution	Up to 1-10bp	Nucleosome (200bp)
Peak calling tool	MACS	SICER

Peak calling: Scale matters

- **Sharp peaks (<1kb)**

  transcription factor binding, DNase/ATAC-seq

  MACS (Zhang et al, *Genome Bio* 2008)

  dynamic background

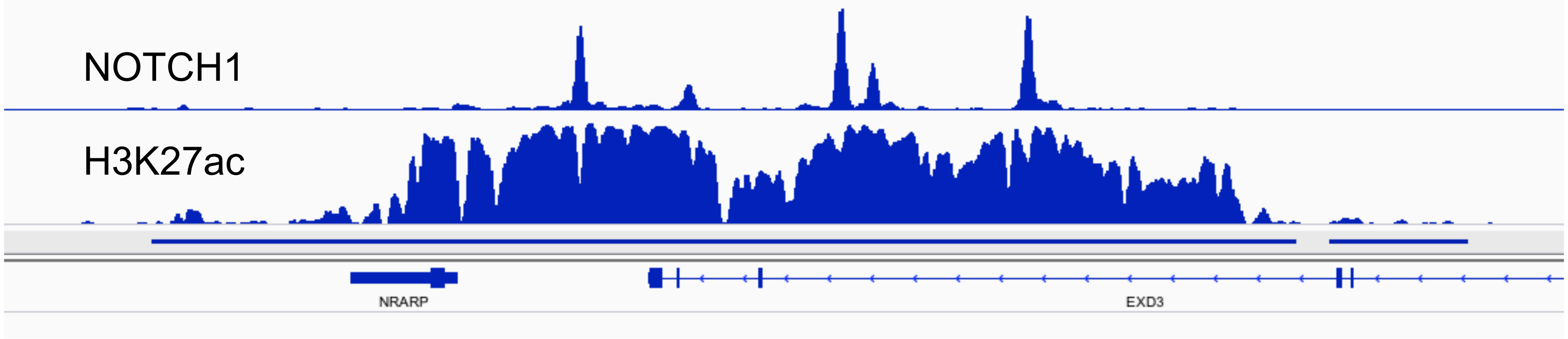
  Poisson model

- **Broad peaks (>1kb)**

  Histone modifications,
  “super-enhancers”
  Diffuse signal

  SICER (Zang et al, *Bioinformatics* 2009)

  Spatial clustering of localized weak signal and integrative Poisson model

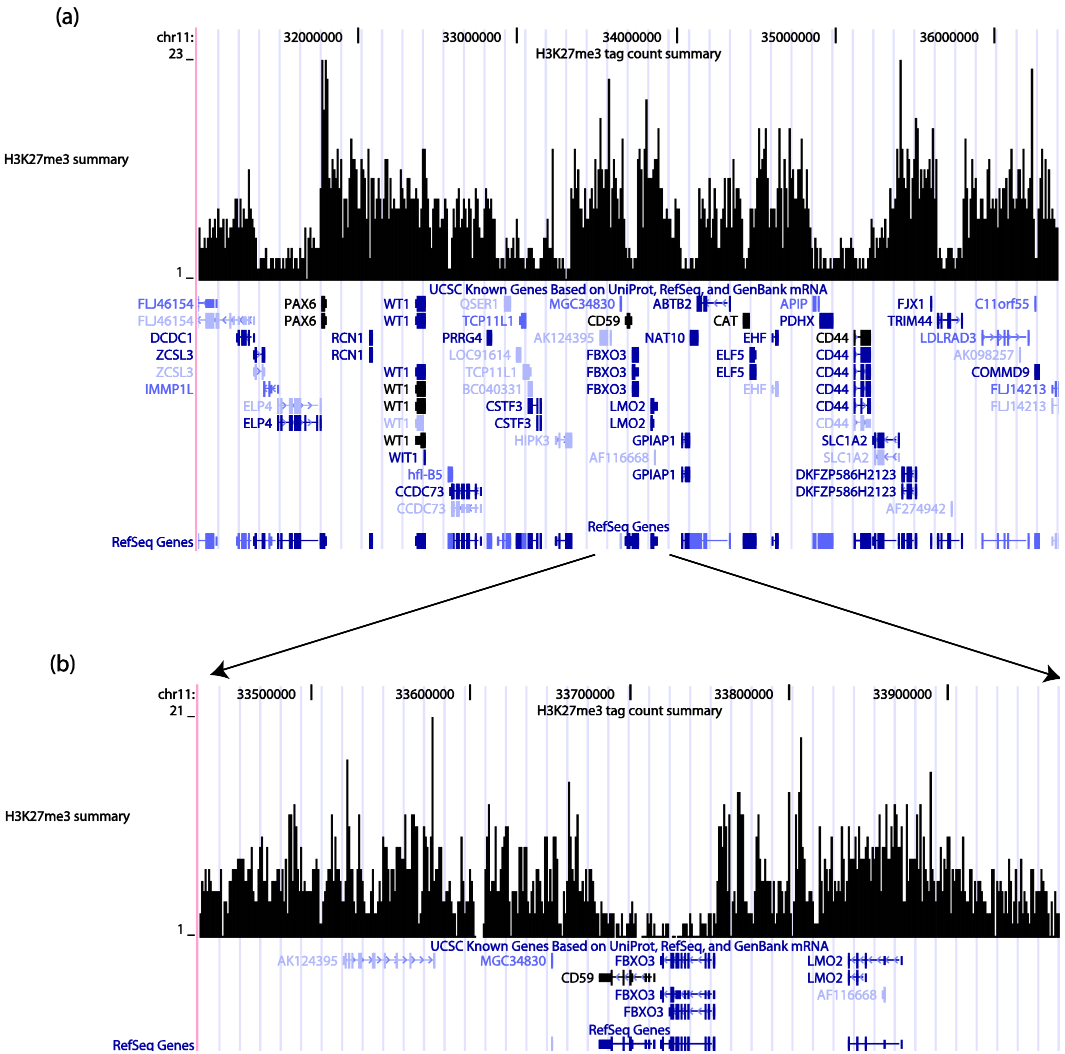


Wang, Zang et al. 2014

Histone modification patterns are intrinsically diffuse

- Noisy
- Hard to see “peaks”
- Enriched regions are spread out
- Lack saturation

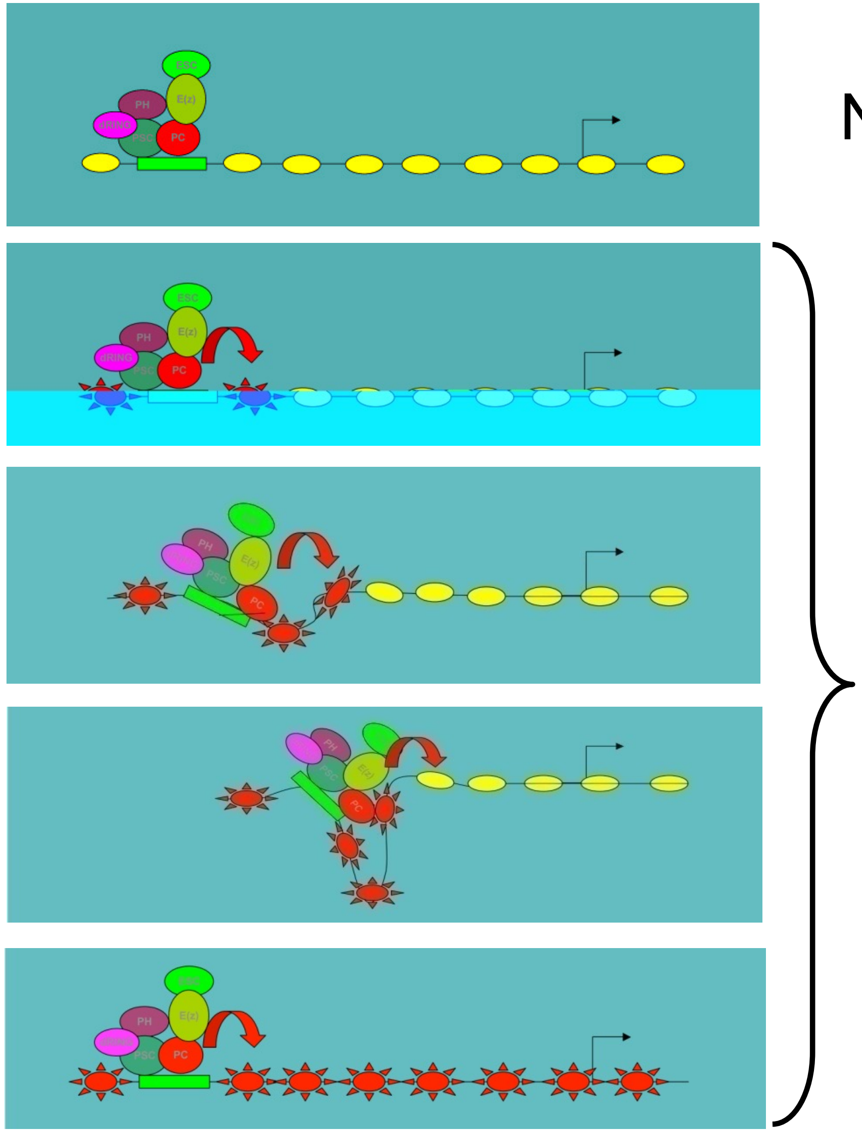
- *Why?*



Histone modification tends to spread out

Domain formation model for repressive marks

- HP1
  H3K9me3
- PRC1/PRC2
  H3K27me3



Nucleation

Propagation

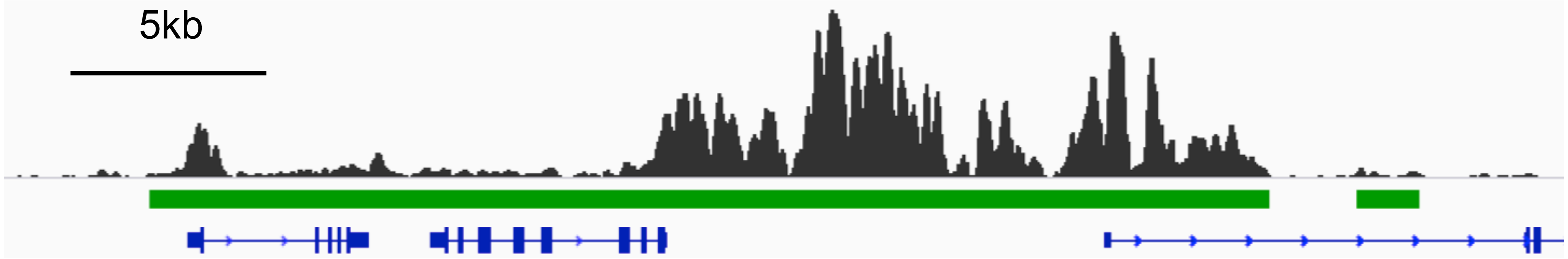
Functional annotation of common histone marks

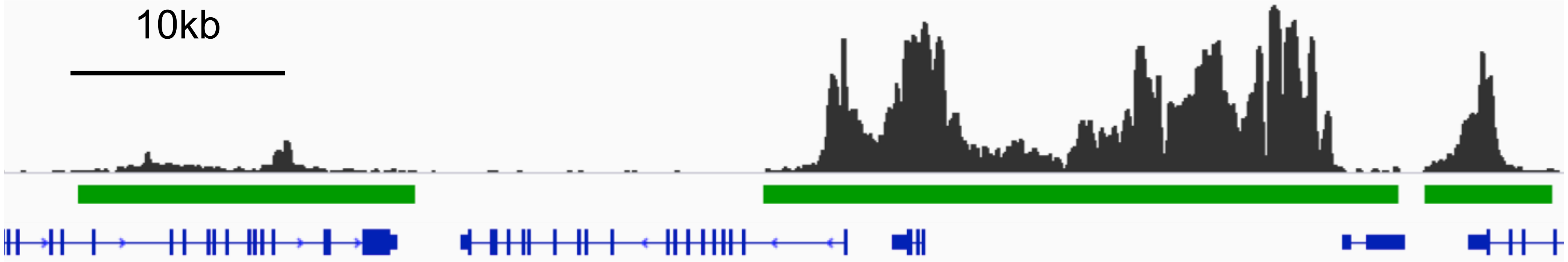
Functional Annotation	Histone Marks
Promoters	H3K4me3
Bivalent/Poised Promoter	H3K4me3/H3K27me3
Transcribed Gene Body	H3K36me3
Enhancer (both active and poised)	H3K4me1
Active Enhancer	H3K4me1/H3K27ac
Polycomb Repressed Regions	H3K27me3
Heterochromatin	H3K9me3

Modified from Rivera & Ren *Cell* 2013

SICER

- **S**patial-clustering **I**dentification of **C**hIP-seq **E**nriched **R**egions





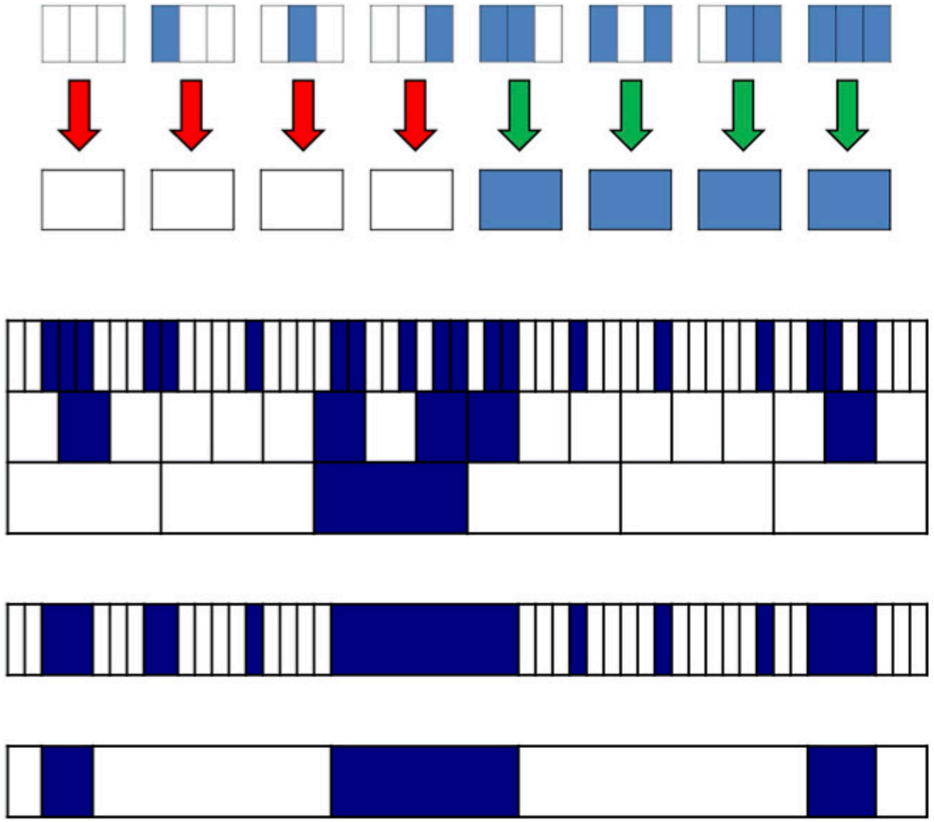
Zang et al. *Bioinformatics* 2009

Other approaches for chromatin domains

- ChromHMM: Hidden Markov Models (Ernst & Kellis)
- Recognicer: Coarse-graining (Zang, et al. 2020)

RECOGNICER: calling ultra-broad ChIP-seq peaks using coarse-graining

- **Re**cursive **Co**arse-**G**rai**n**ing approach for **I**dentification of **C**hIP-seq **E**nriched **R**egions



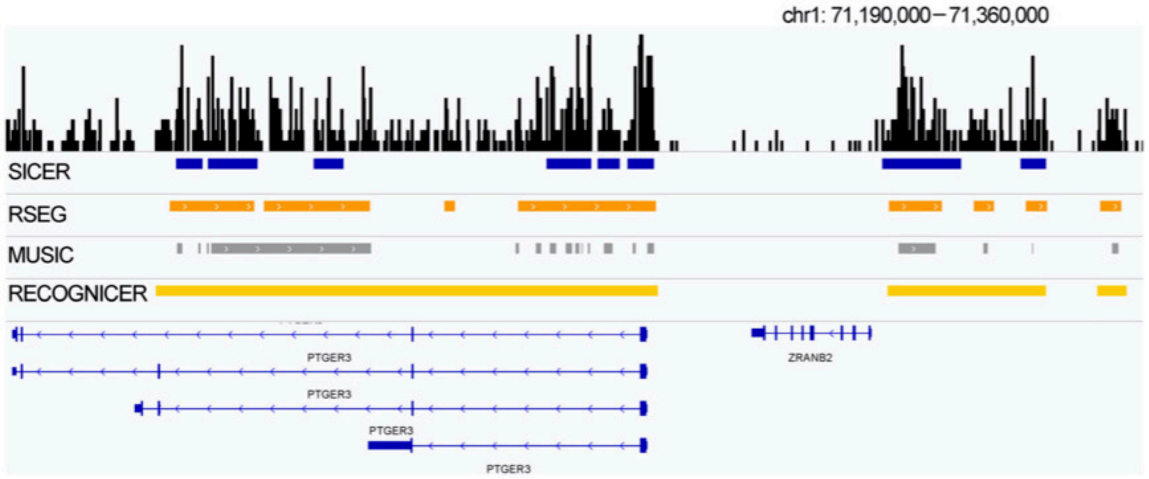
- Block transformation under a majority rule
- Approach:
  - Recursive block transformation
  - Trace back to identify candidate enriched regions
  - Significance determination
  - Scale-free

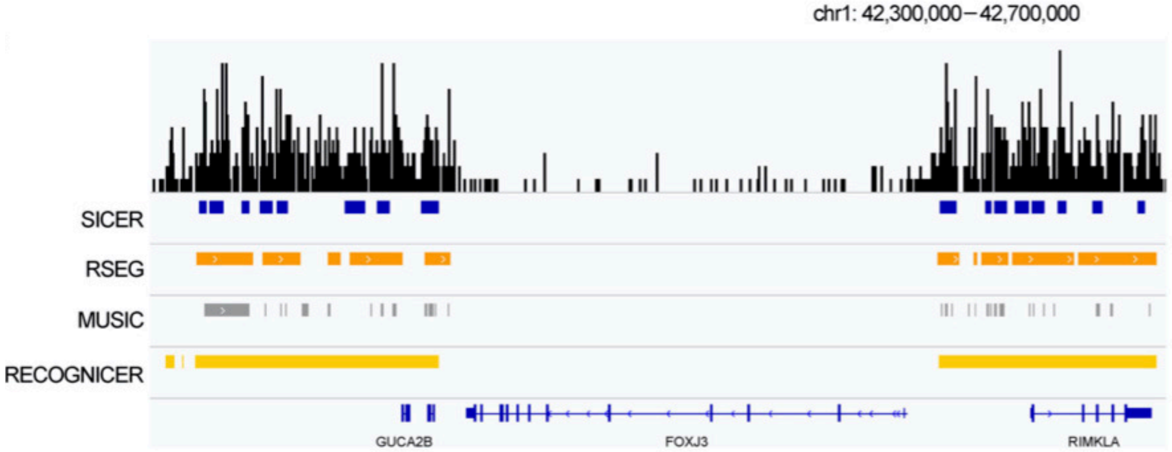
https://zanglab.github.io/SICER2/

Zang et al. *Quantitative Biology* 2020

RECOGNICER: calling ultra-broad peaks from ChIP-seq data

- **Re**cursive **Co**arse-**G**rai**n**ing approach for **I**dentification of **C**hIP-seq **E**nriched **R**egions



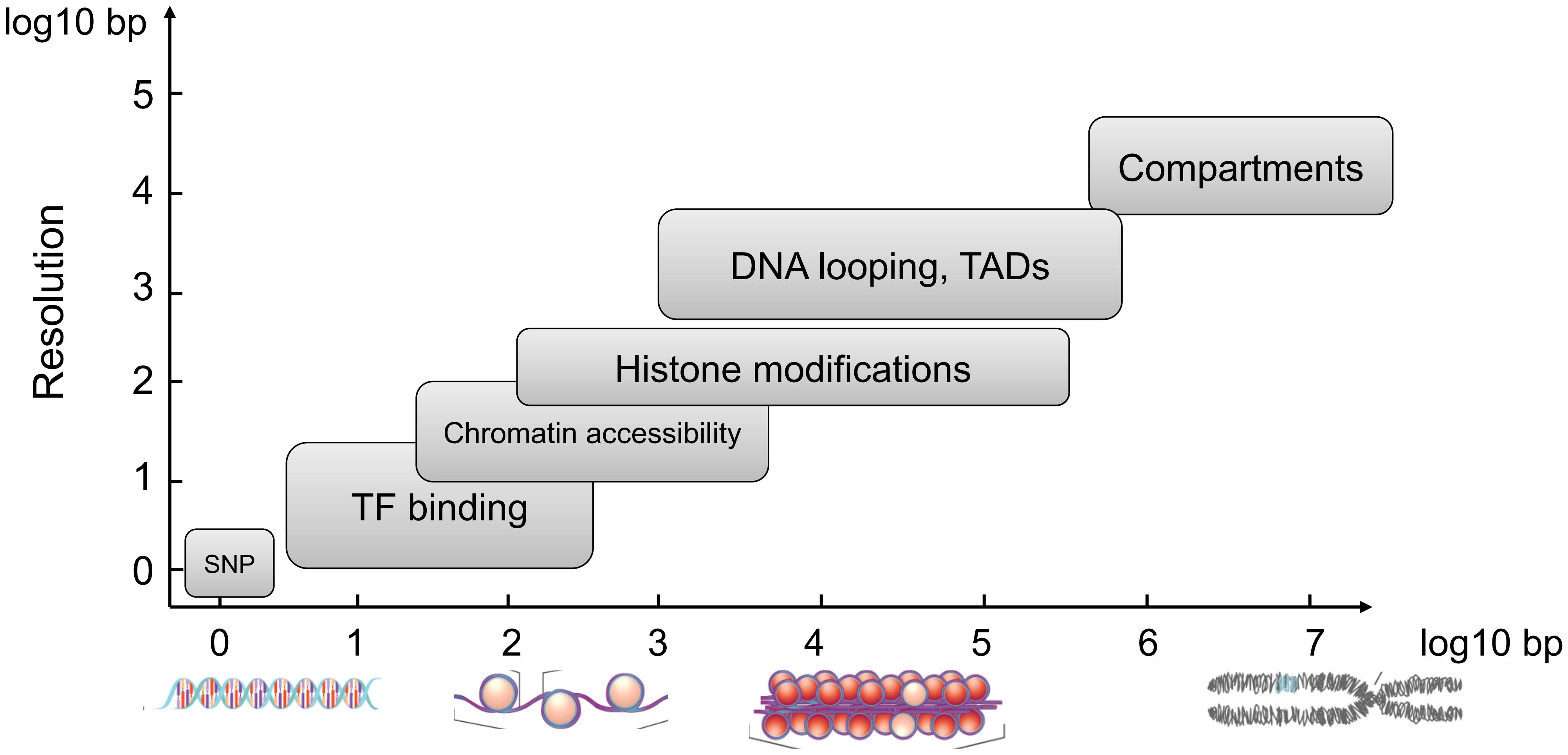


RSEG: Song & Smith 2011
MUSIC: Harmanci et al 2014

https://zanglab.github.io/SICER2/

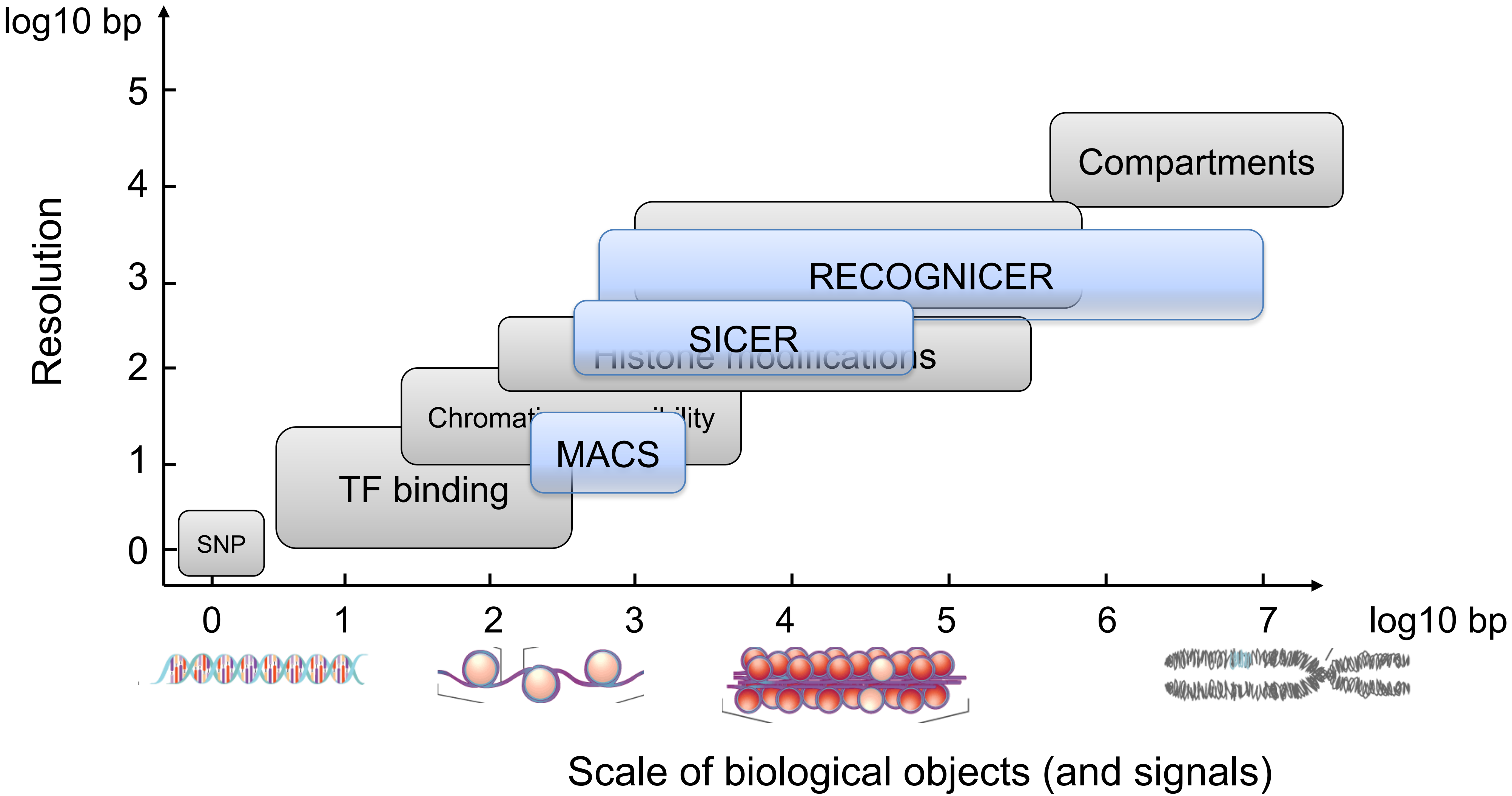
Zang et al. *Quantitative Biology* 2020

Functional genomic/epigenomic elements range across scales

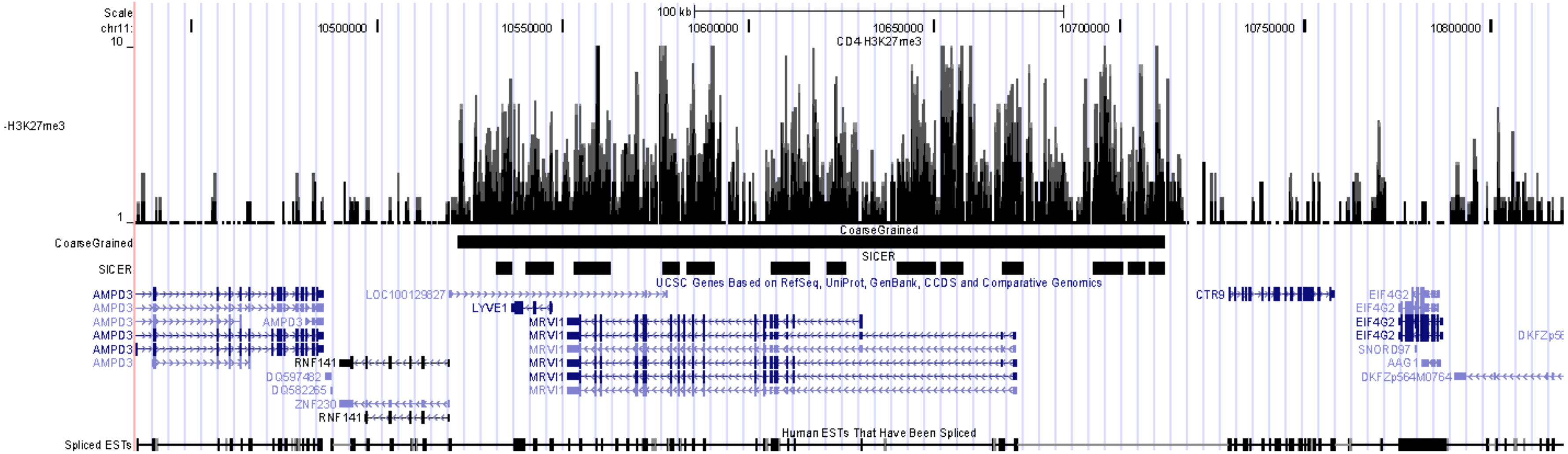


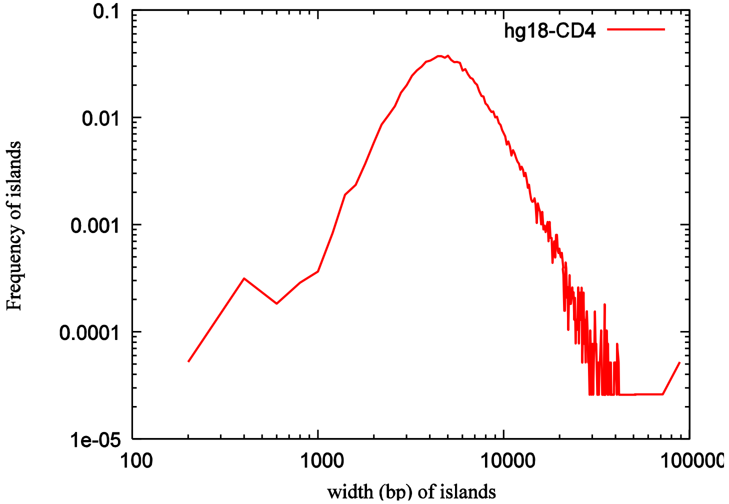
Scale of biological objects

Functional genomic/epigenomic elements range across scales



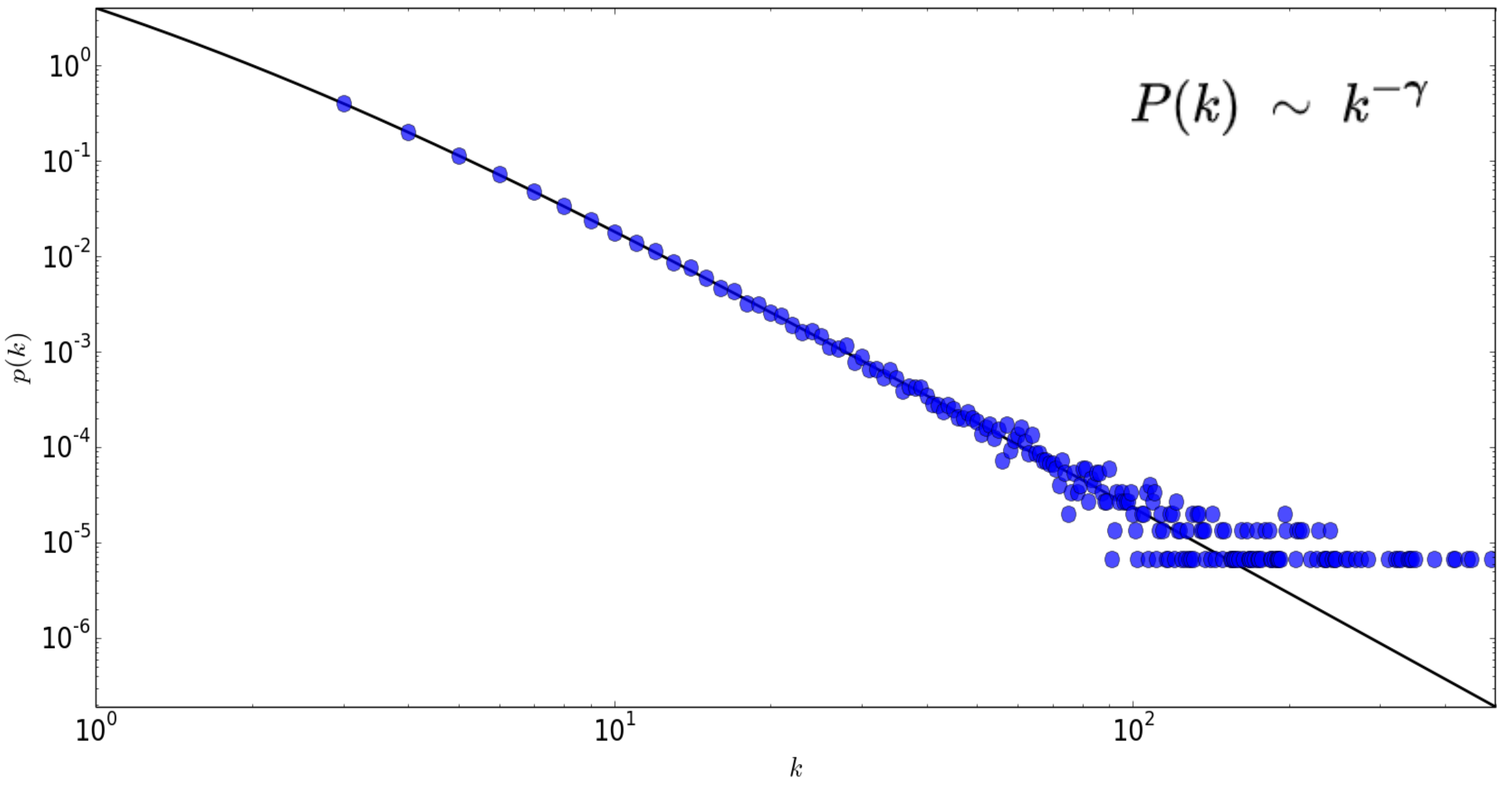
Scale-free property of chromatin domains

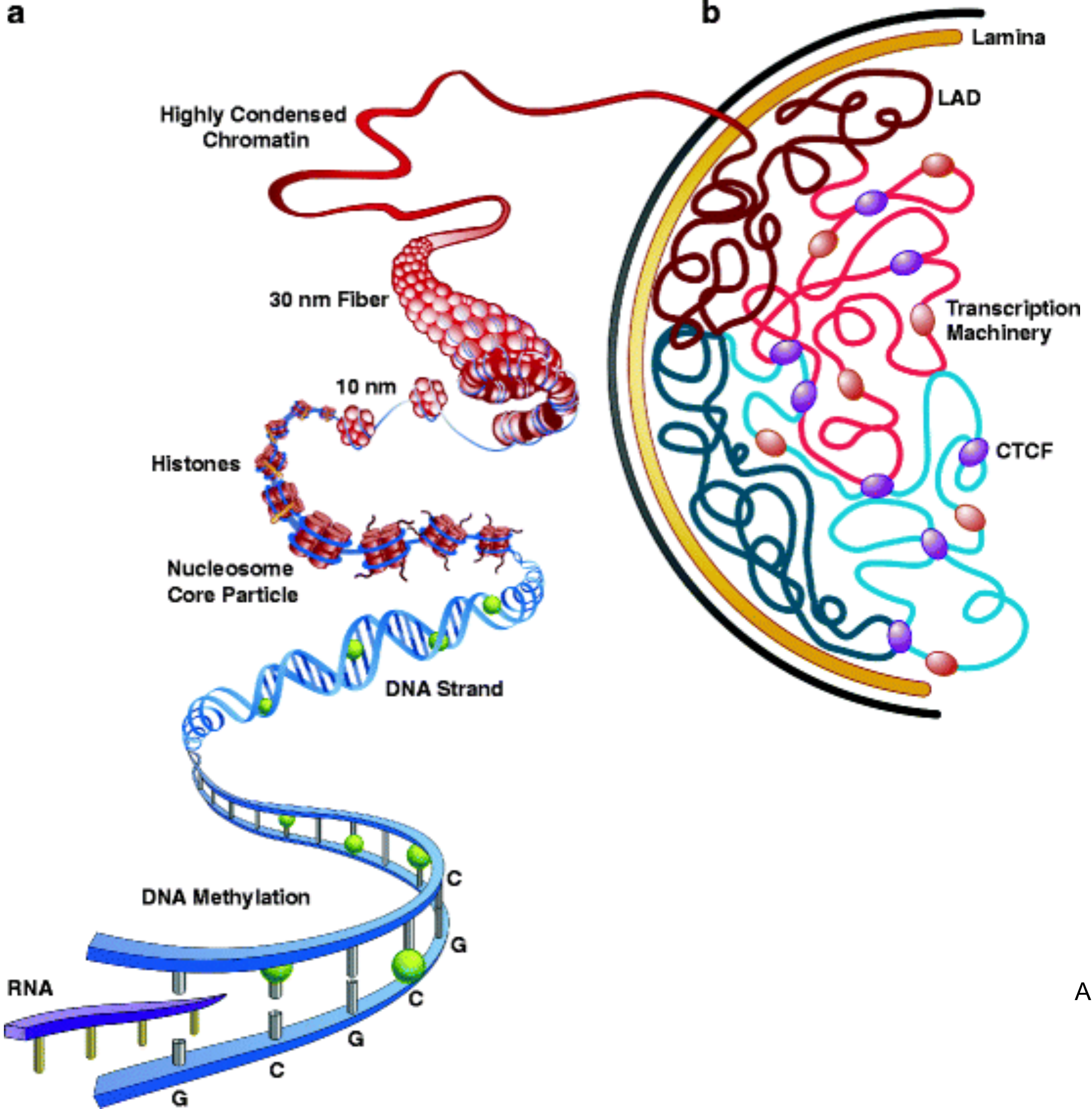




Scale-free: Power-law distribution

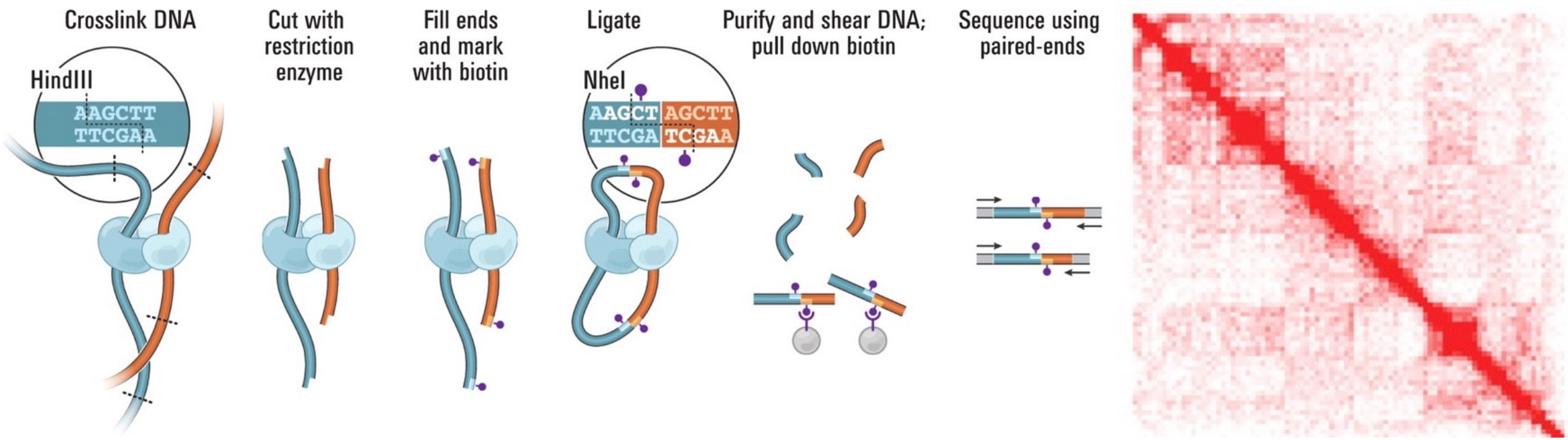
- Scale-free network
  - Fraction of nodes in the network having k connections follows power-law





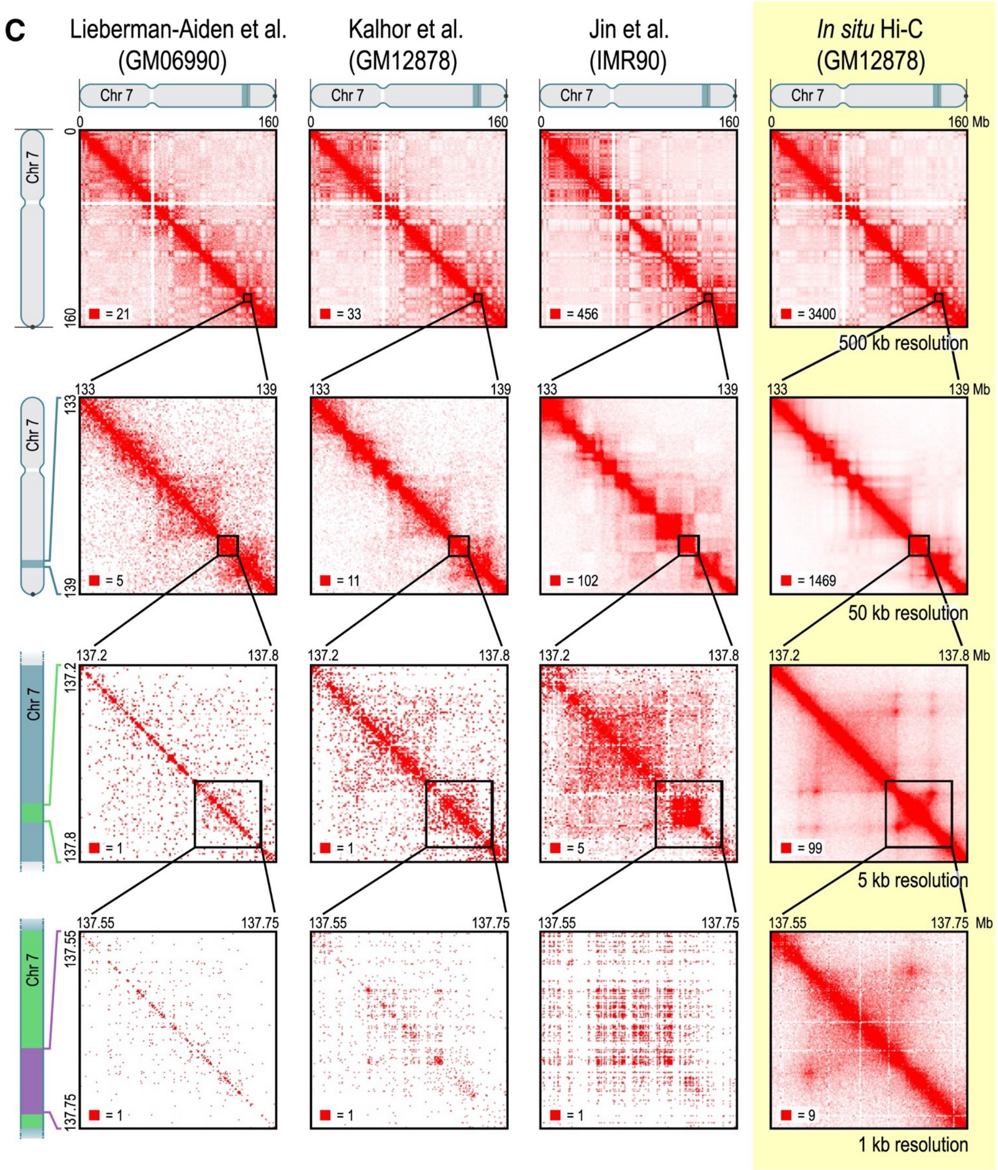
Adomas & Wade 2013

Hi-C



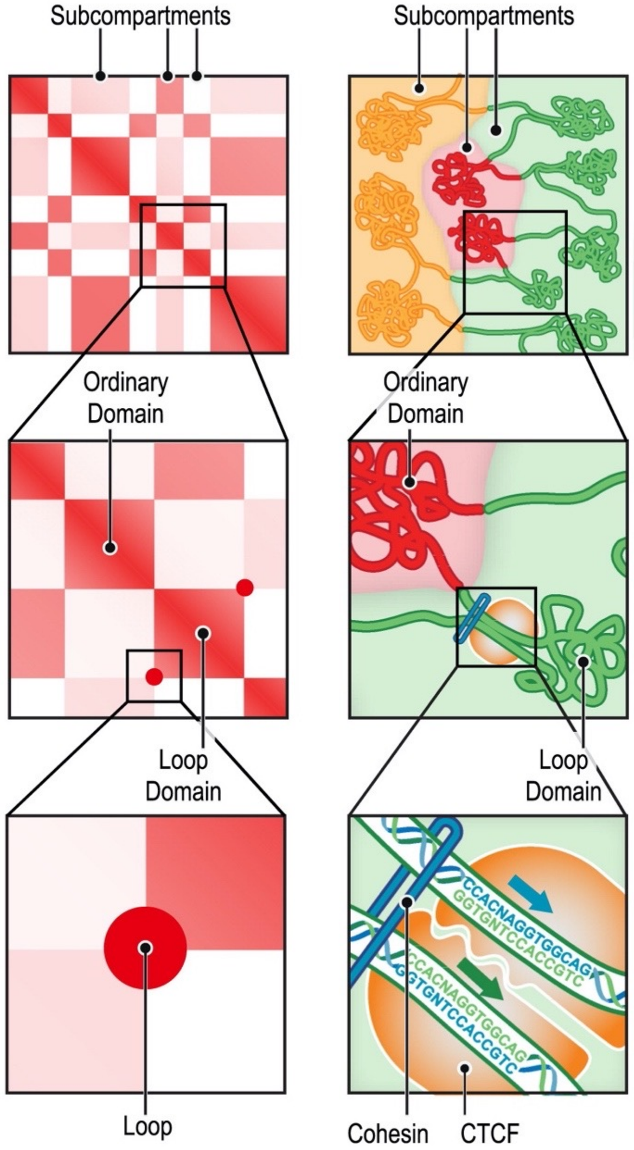
Lieberman-Aiden et al. Science 2009

Hi-C contact heatmap for 3D genome interactions



Rao *et al. Cell* 2014

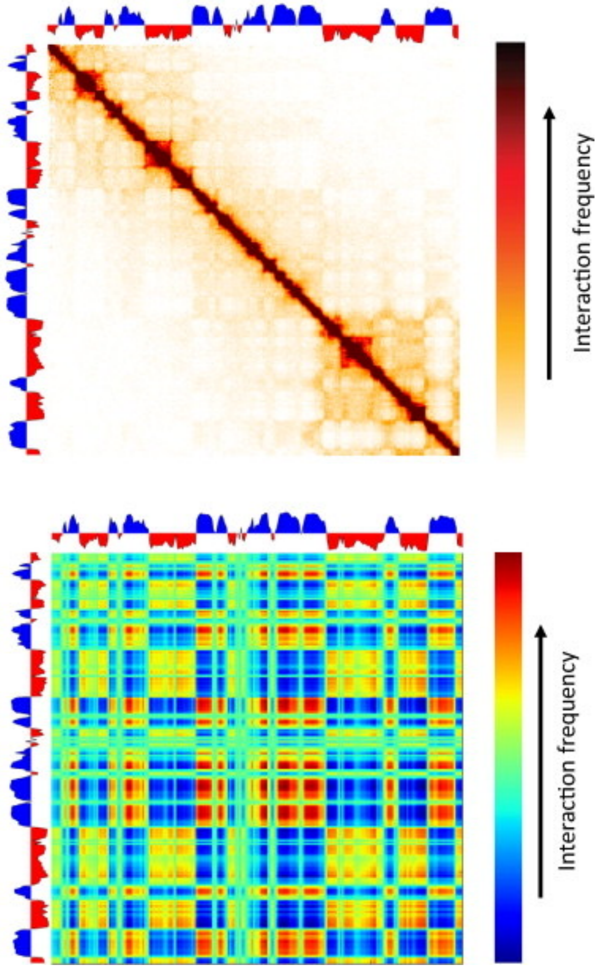
Hi-C contact heatmap for 3D genome interactions



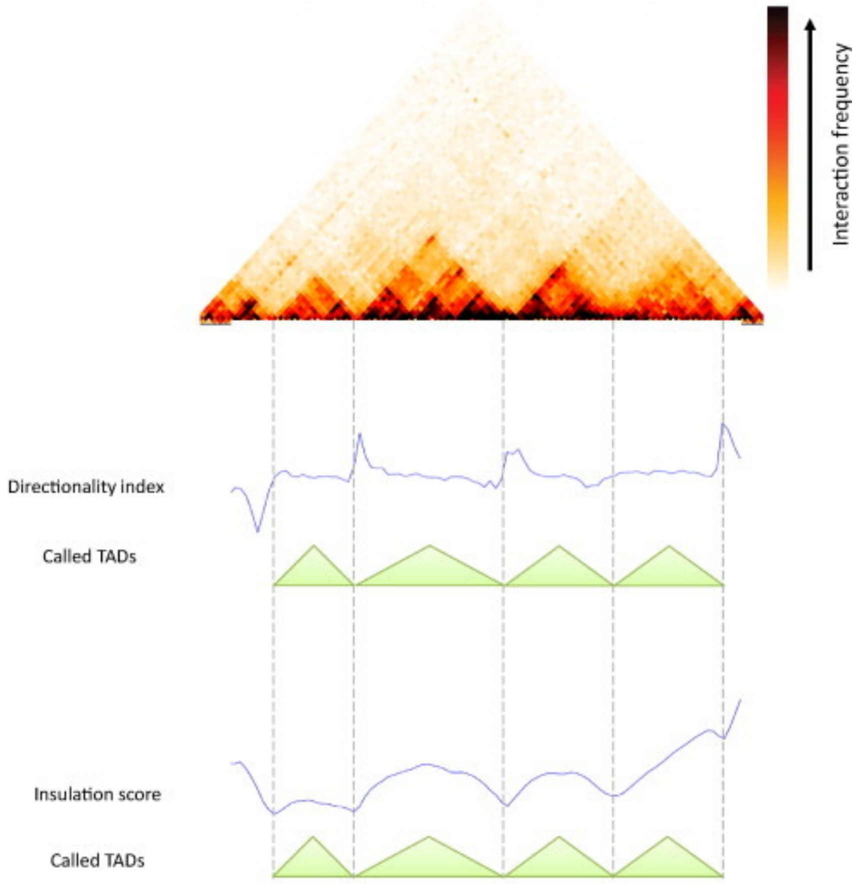
Rao *et al.* *Cell* 2014

Hi-C analysis

- Chromatin compartments

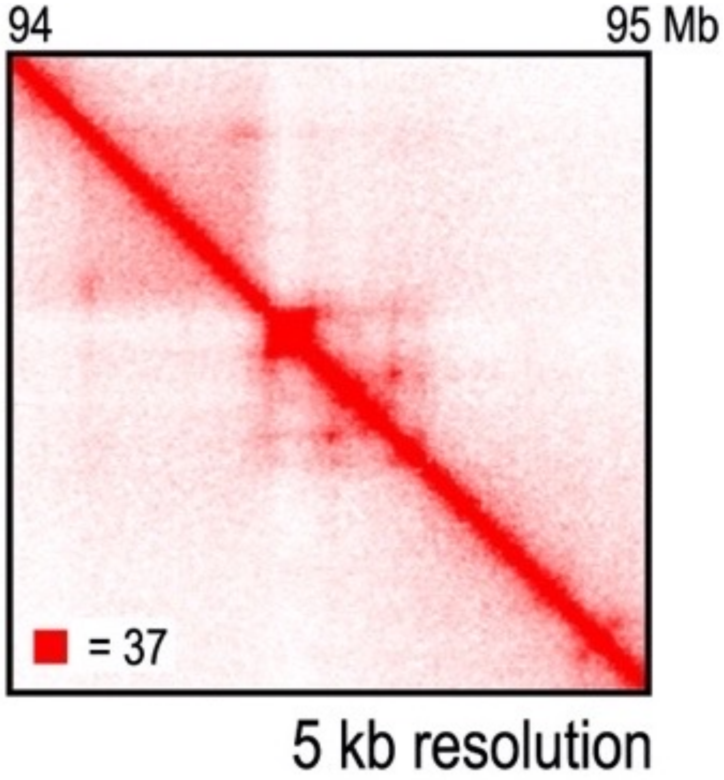


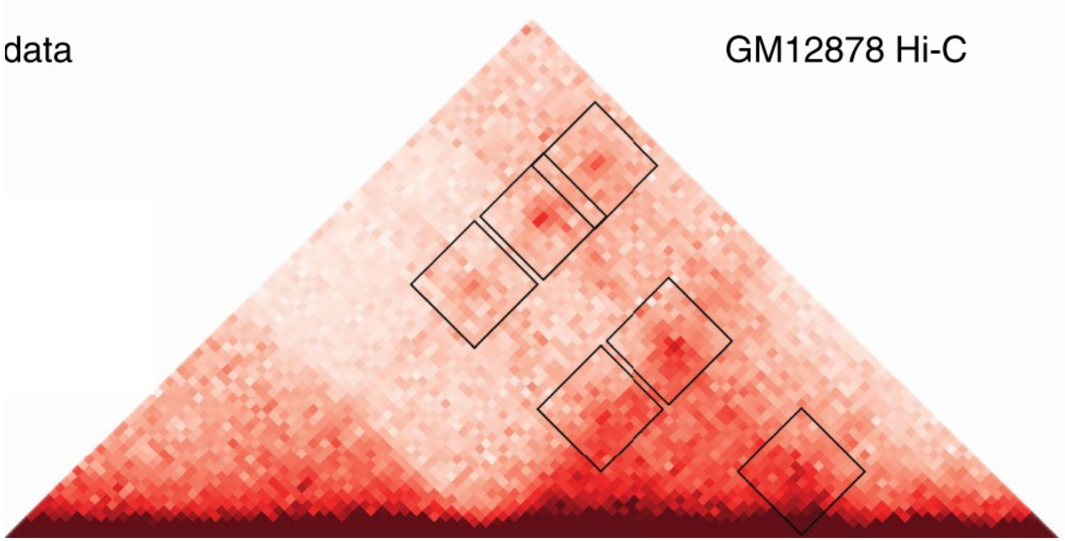
- Topologically Associating Domains (TADs)

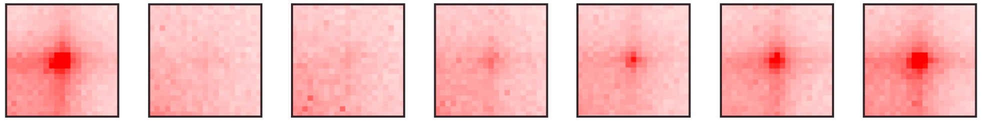


Essential elements in a Hi-C contact heatmap

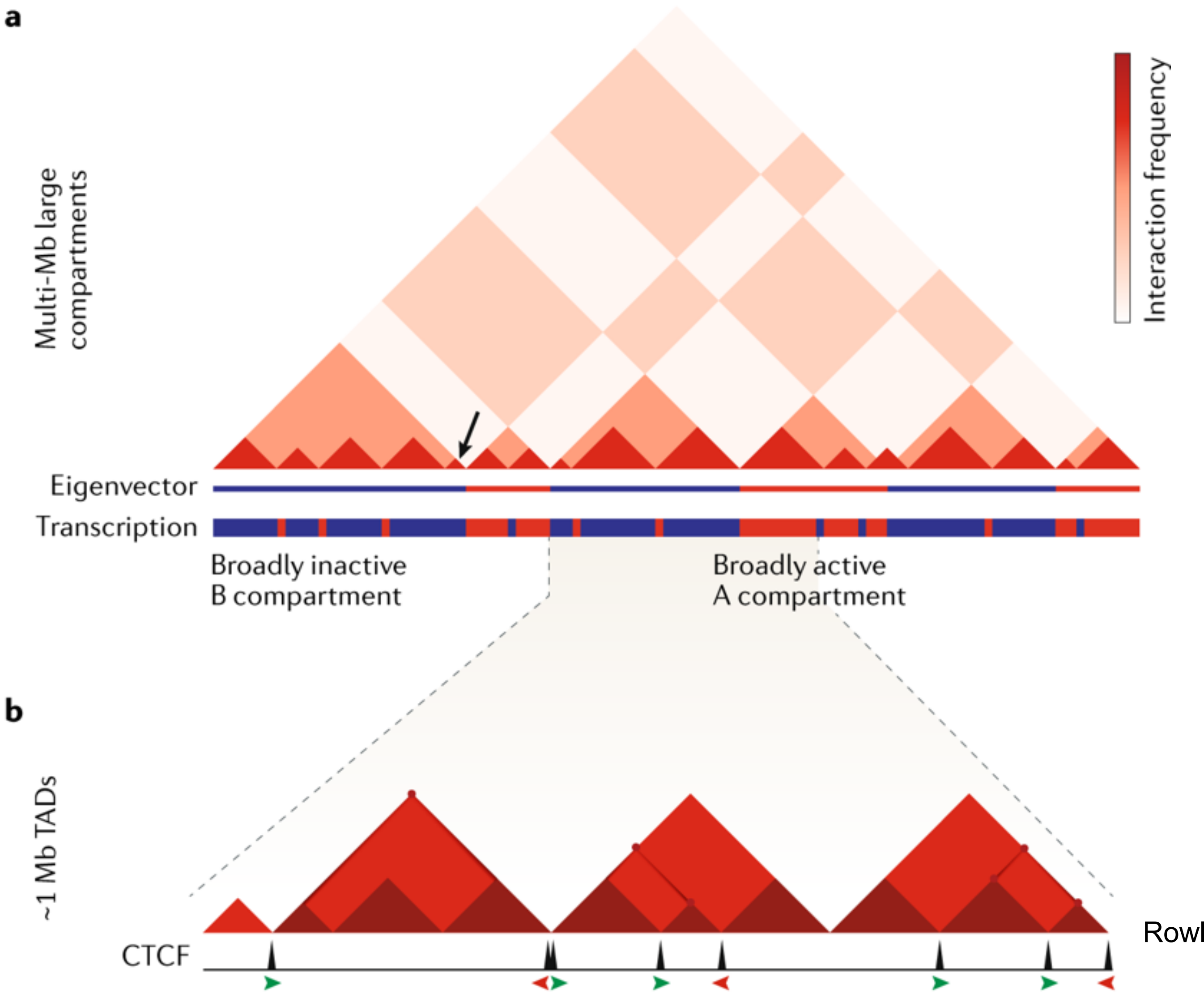
- Scale, scale, scale
- Resolution
- Normalization
- Blocks, stripes, loops (2d peaks)





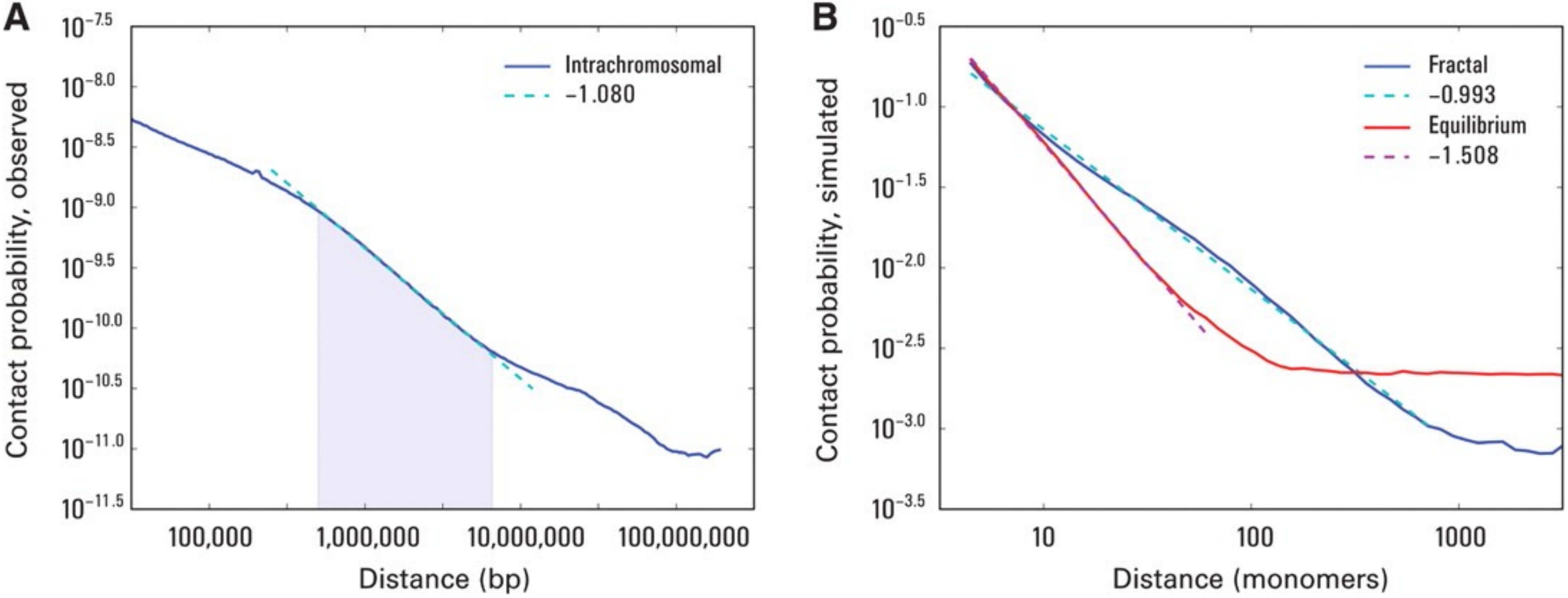


Scale-free property of Hi-C maps



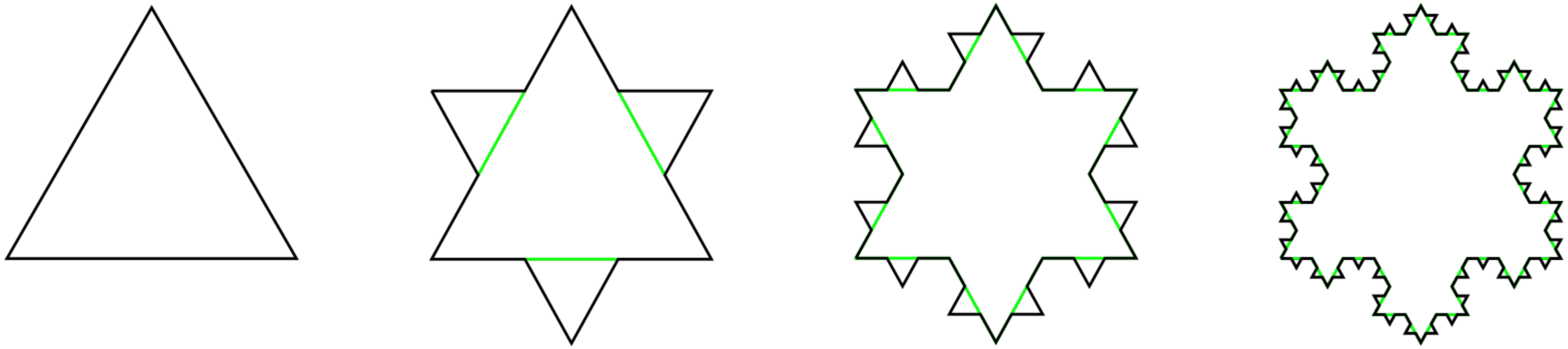
Rowley & Corces, *Nat Rev Genet* (2018)

Hi-C: Power-law property of contact probability distribution

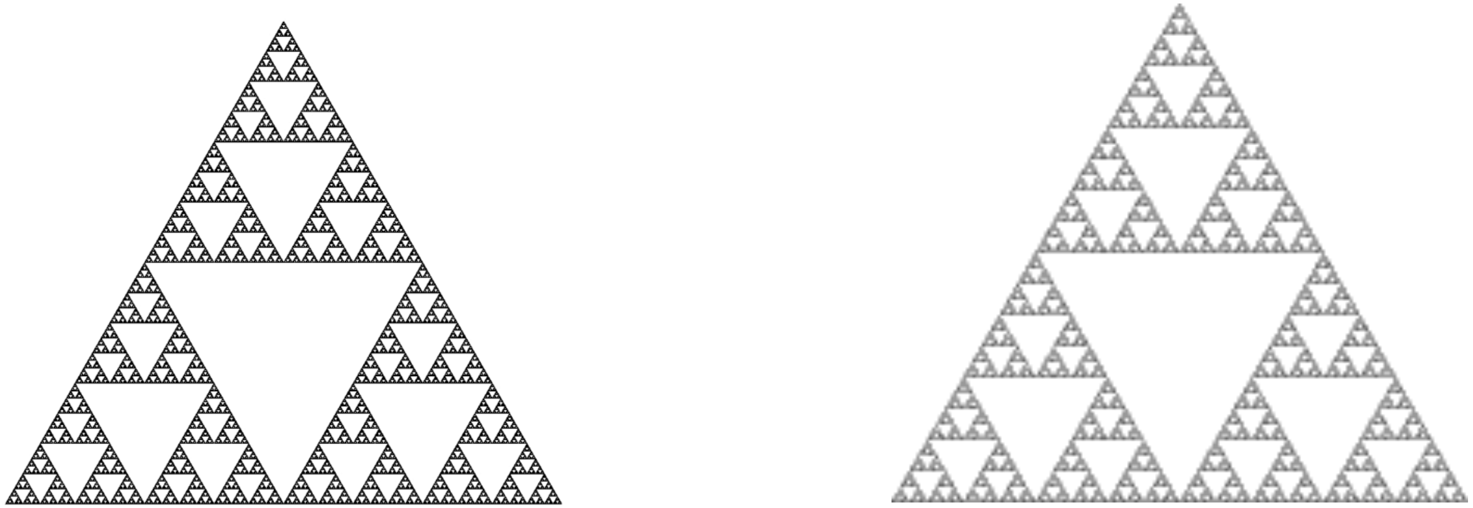


Lieberman-Aiden et al. Science 2009

Fractal Structures



Koch Snowflake



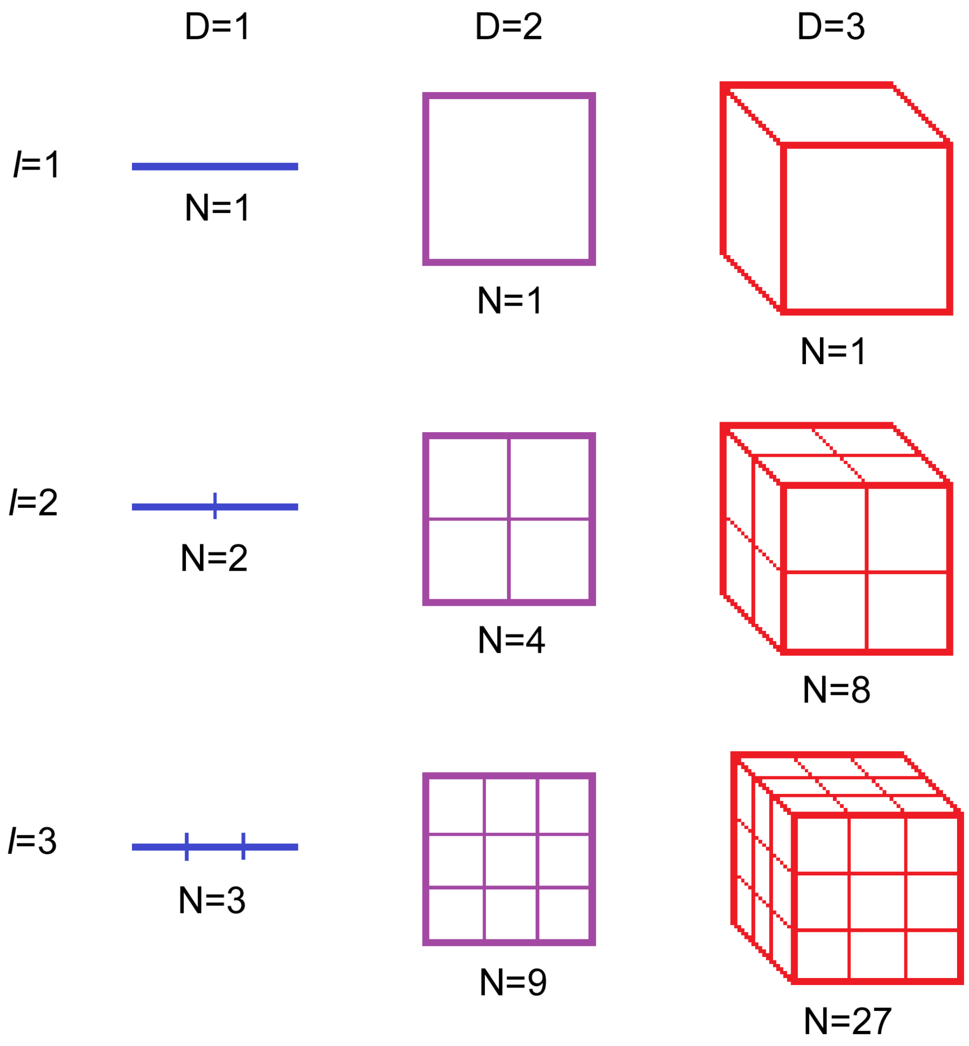
Sierpinski Gasket

Fractal Dimension

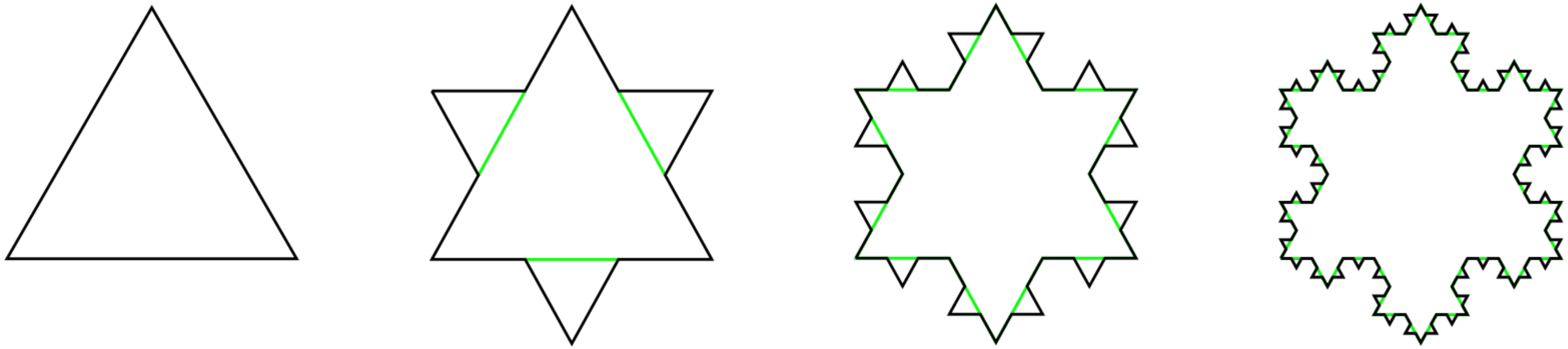
$$N = \varepsilon^{-D}$$

How many (*N*) unit elements can fill in the space of scale-up by *l*

$$D = -\log_{\varepsilon} N = -\frac{\log N}{\log \varepsilon}$$



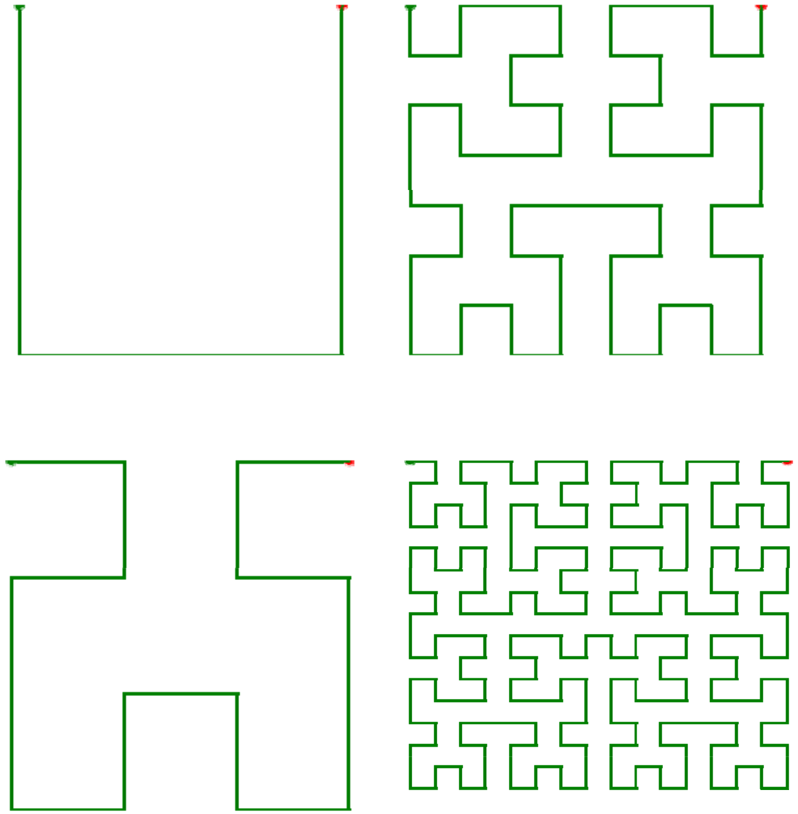
Fractal Dimension: Koch Snowflake

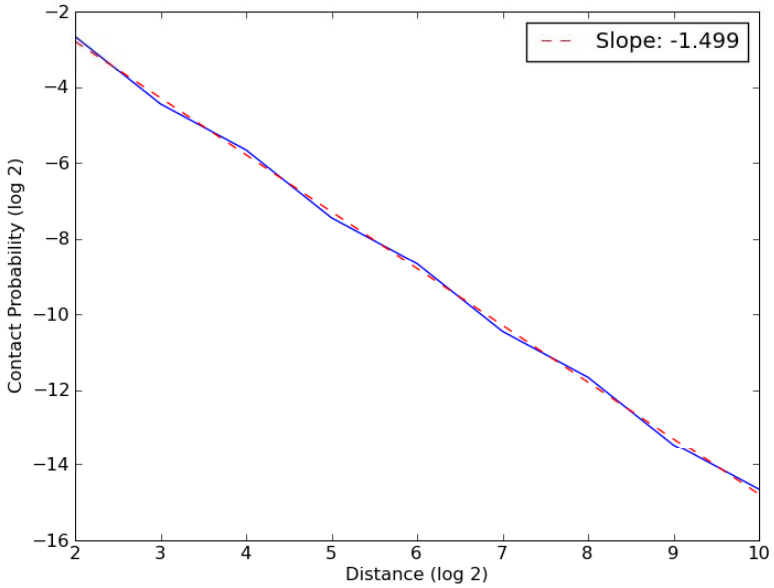


$$D = -\log_{\varepsilon} N = -\frac{\log N}{\log \varepsilon}$$

log4/log3 ~ 1.26

Hilbert Curve



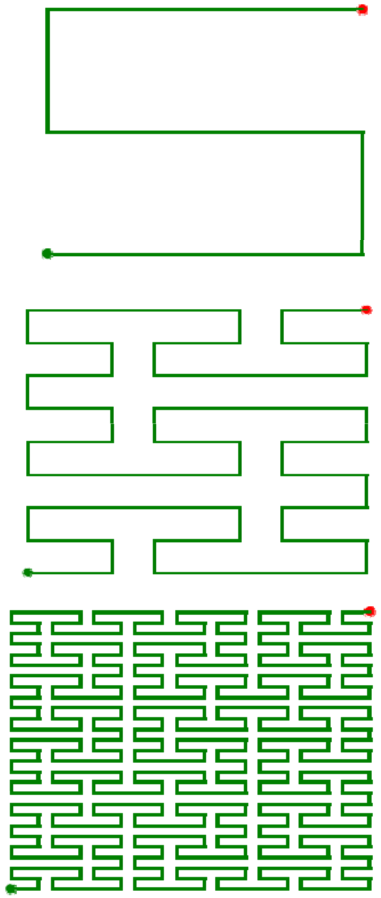


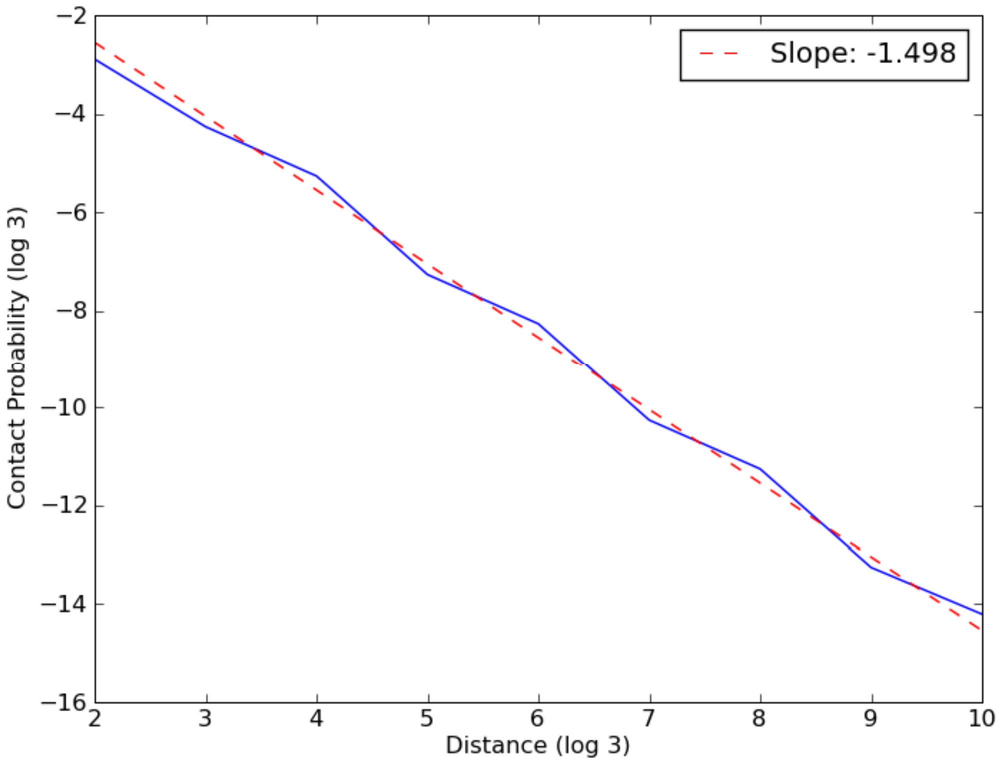
$$D = -\log_{\varepsilon} N = -\frac{\log N}{\log \varepsilon}$$

log5/log3 ~1.465

Lieberman-Aiden et al. Science 2009

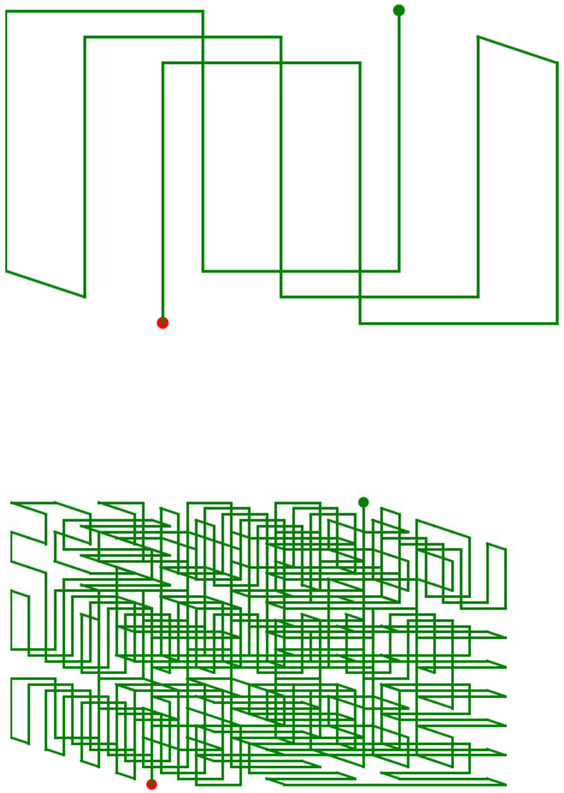
Peano Curve

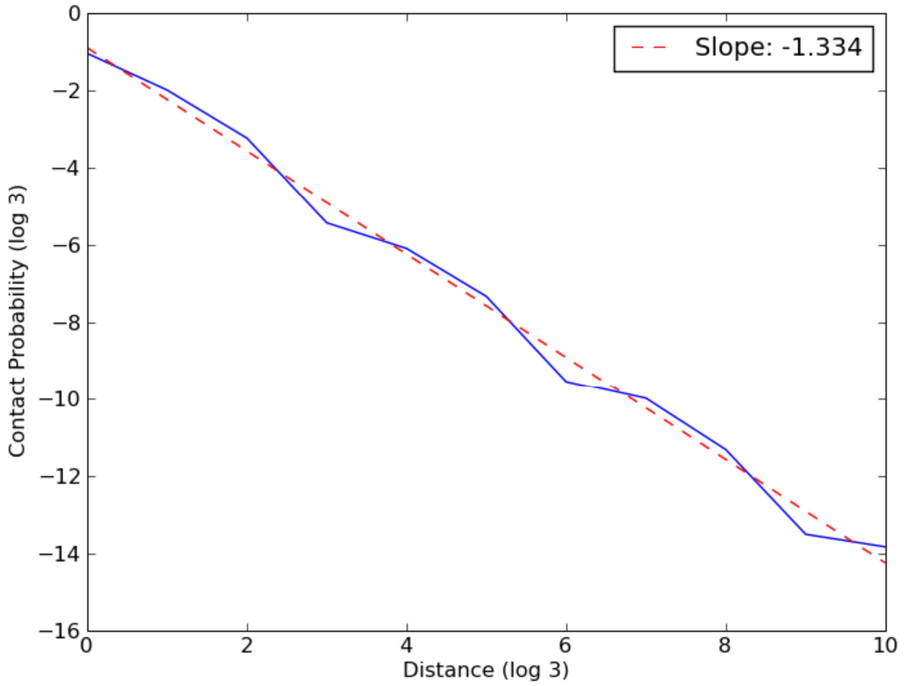




Lieberman-Aiden et al. Science 2009

3D Peano Curve



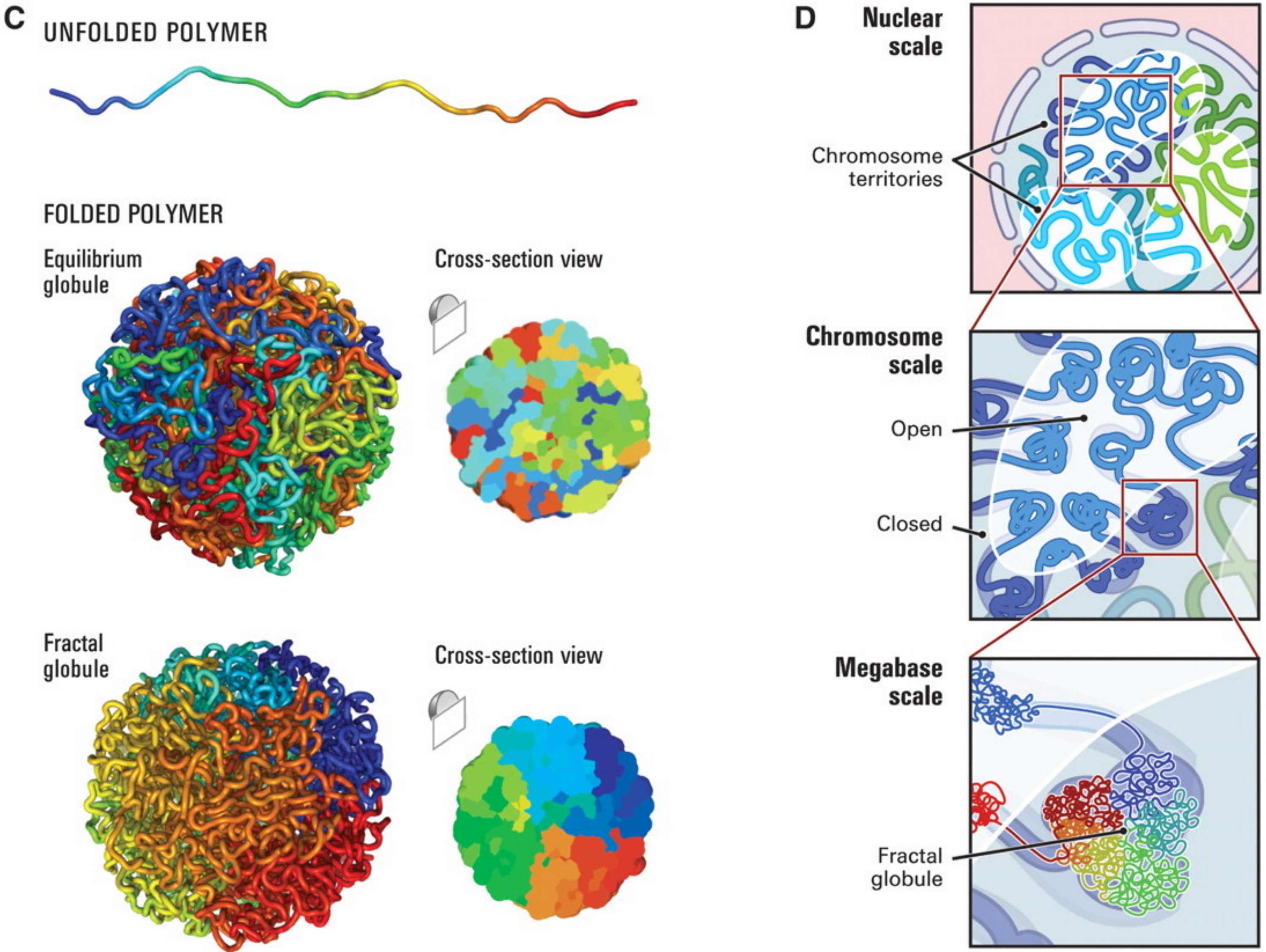


$P_{contact}(x) = kx^{\alpha}$, where α is given by

$\alpha_{smooth} = -(1+\frac{1}{d})$ and $\alpha_{interdigitated} = -1$.

Lieberman-Aiden et al. Science 2009

Fractal Structure of Genome Organization



Lieberman-Aiden et al. Science 2009

Summary

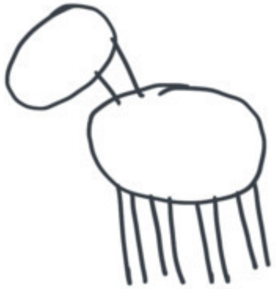
- Genomic elements range across scales
- Effective/meaningful computational analysis should consider scales
- Genome structure is complex
- Fractal and scale-free properties

HOW TO:

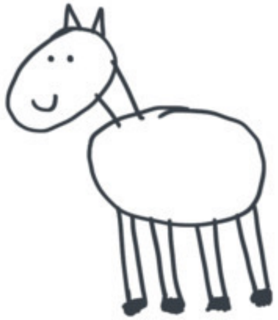
DRAW A HORSE

BY VAN OKTOP

① DRAW 2 CIRCLES

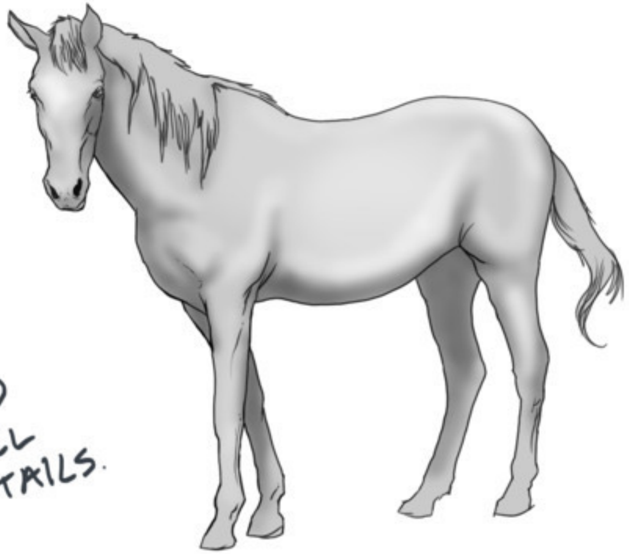


② DRAW THE LEGS



③ DRAW THE FACE

④ DRAW THE HAIR



⑤ ADD SMALL DETAILS.

Scales of histone mark islands and chromatin domains

- Narrow: a few nucleosomes, 0.5kb ~ 5kb
  - H3K4me3, H2A.Z, etc.

- Broad: 5kb~100kb
  - Gene loci, chromatin domains, super-enhancers
  - H3K4me1, H3K27ac, H3K36me3, H3K27me3, etc.

- Very broad: >100kb
  - Large chromatin domains, chromatin compartments
  - H3K9me3, H3K27me3