Introduction to Epigenetics

BMI/CS 776

www.biostat.wisc.edu/bmi776/

Spring 2021

Daifeng Wang

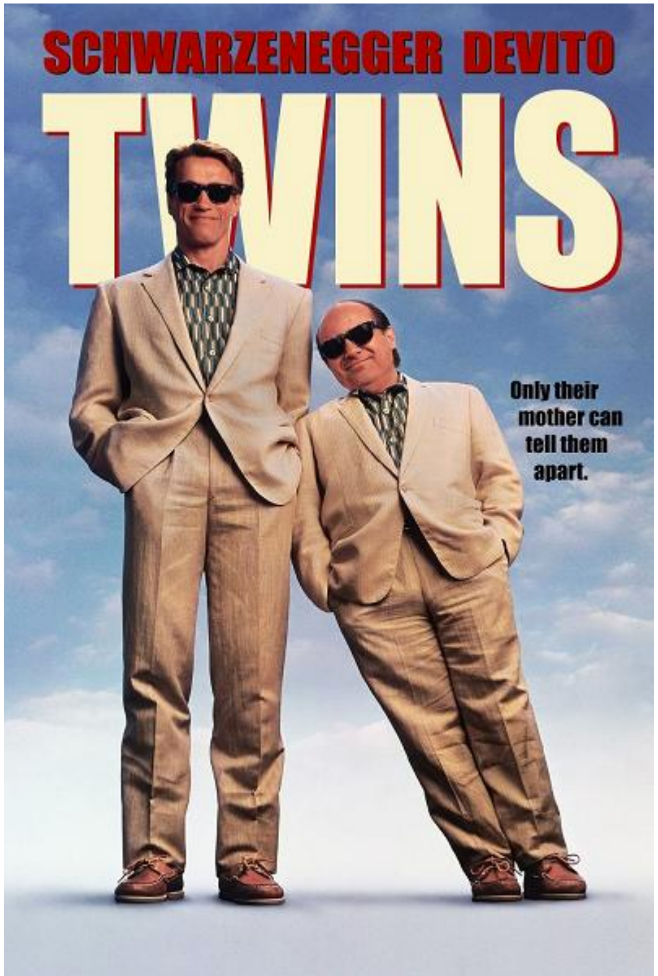
daifeng.wang@wisc.edu

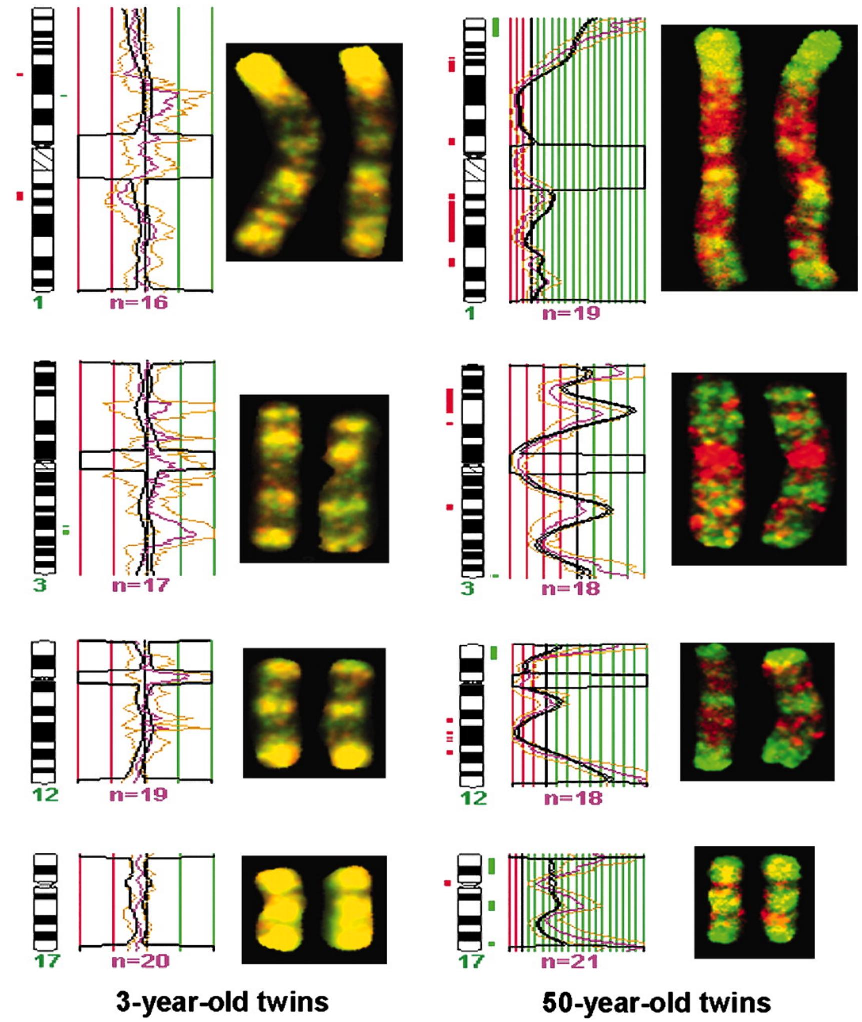
Goals for lecture

Key concepts

- Importance of epigenetic data for understanding transcriptional regulation
- Use of epigenetic data for predicting transcription factor binding sites

Identical DNAs but identical fates?



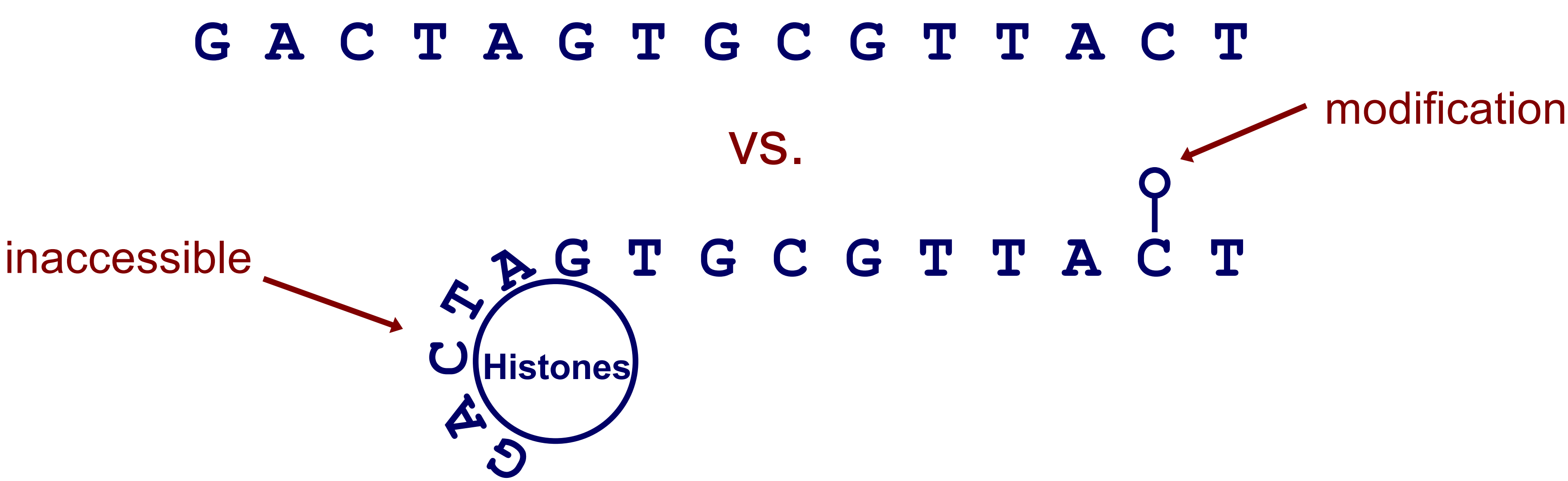


Chromosomes

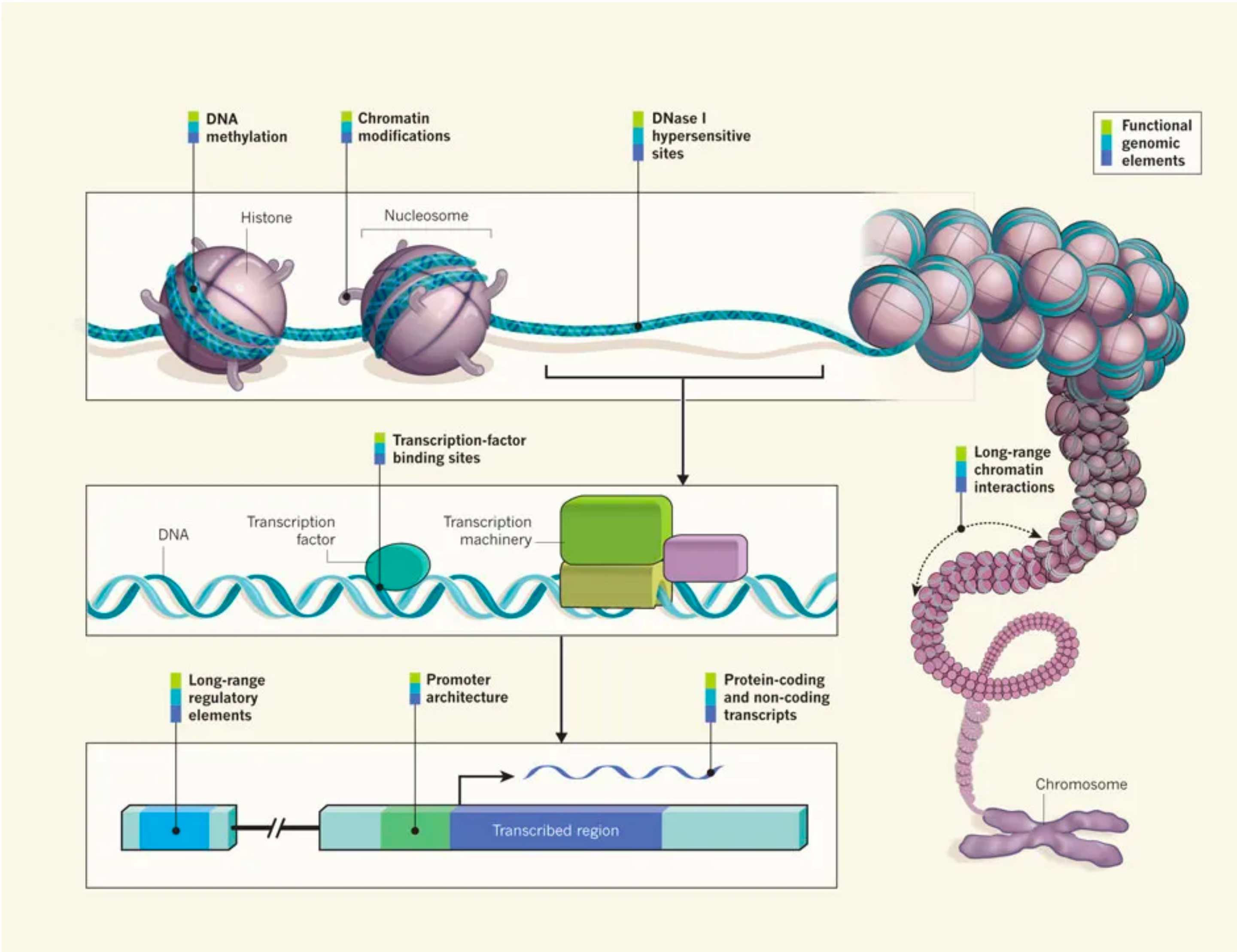
PNAS July 26, 2005 102 (30) 10604-10609; https://doi.org/10.1073/pnas.0500398102

Defining epigenetics

- Formally: attributes that are "in addition to" genetic sequence or sequence modifications
  - "*Epigenetic code*" (vs. genetic code)
- Informally: experiments that reveal the context of DNA sequence
  - DNA has multiple states and modifications



Chromatin packages DNA around Histones



Nature **volume 489**, pages52–54(2012)

Importance of epigenetics

Better understand

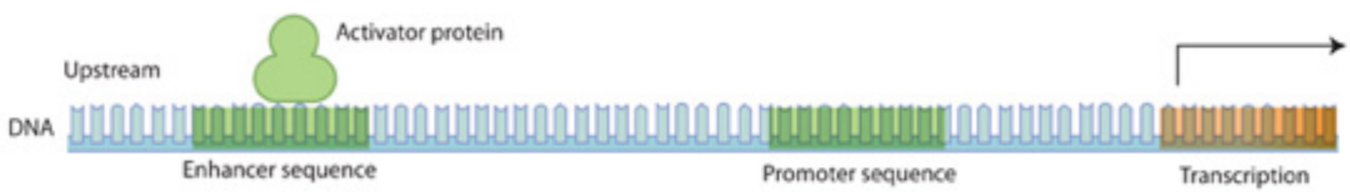
- DNA binding and transcriptional regulation
- Differences between cell and tissue types
- Development and other important processes

PWMs are not enough

- Genome-wide motif scanning is imprecise
- Transcription factors (TFs) bind < 5% of their motif matches
- Same motif matches in all cells and conditions

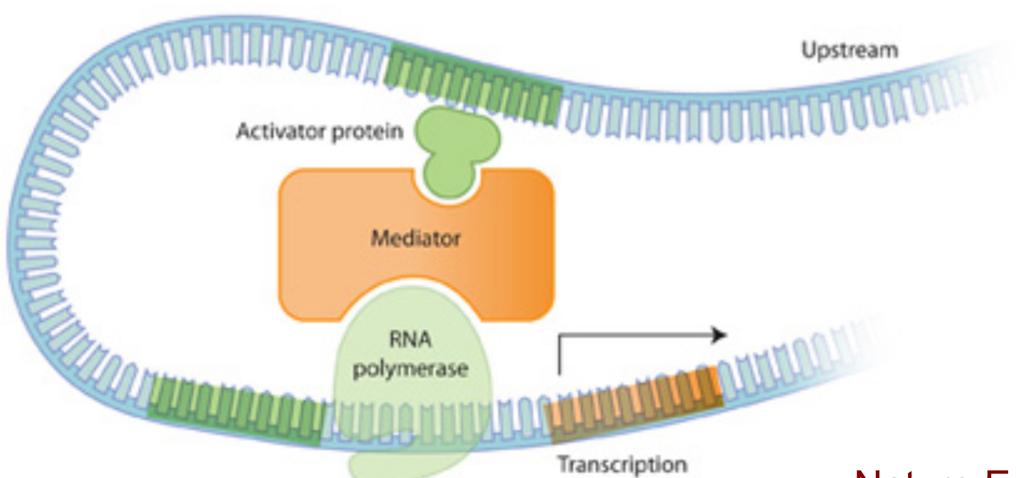
PWMs are not enough

- DNA looping can bring distant binding sites close to transcription start sites
- Which genes does an enhancer regulate?



Promoter: DNA binding site for TFs, close to gene transcription start site

Enhancer: DNA binding site for TFs, can be far from affected gene

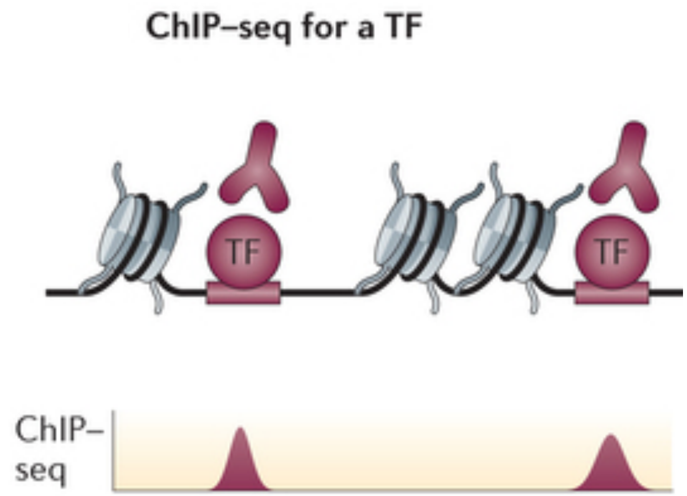


Nature Education 2010

Mapping regulatory elements genome-wide

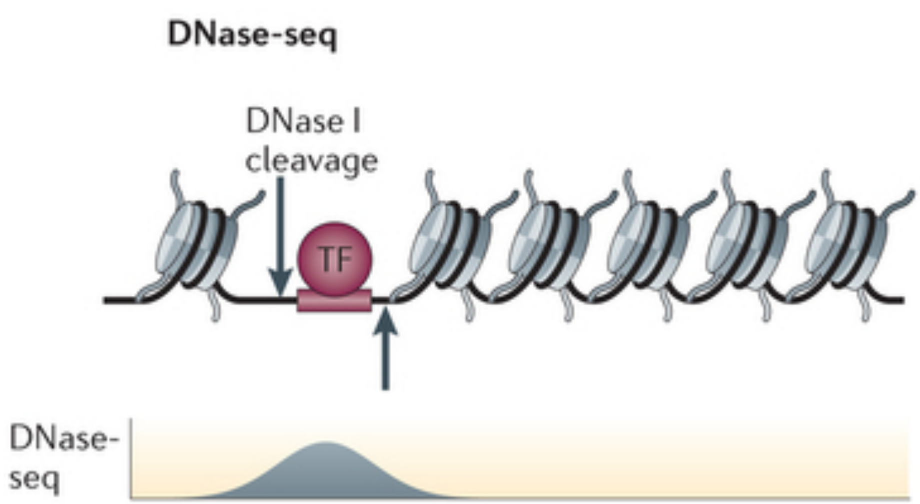
- Can do much better than motif scanning with additional data

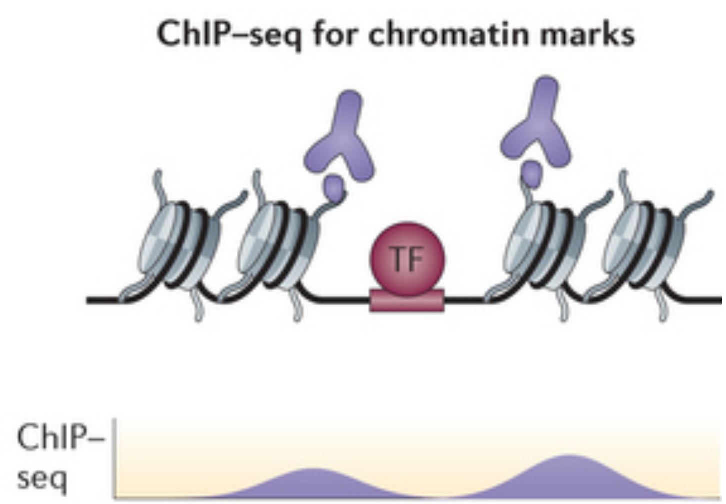
- ChIP-seq measures binding sites for one TF at a time



Shlyueva *Nature Reviews Genetics* 2014

- Epigenetic data suggests where *some* TF binds

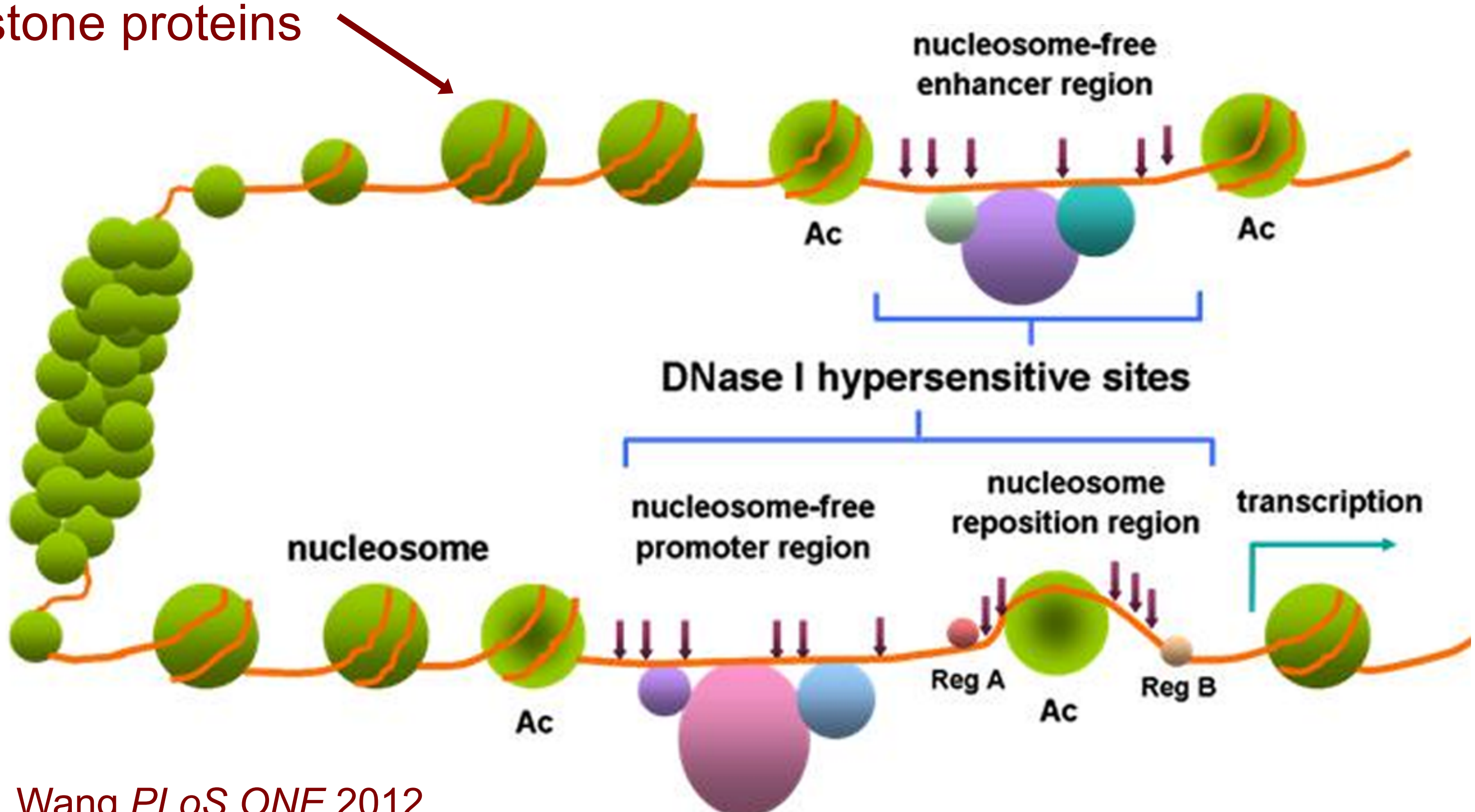




DNase I hypersensitivity

- Regulatory proteins bind accessible DNA
- DNase I enzyme cuts open chromatin regions that are not protected by nucleosomes

Nucleosome: DNA wrapped around histone proteins

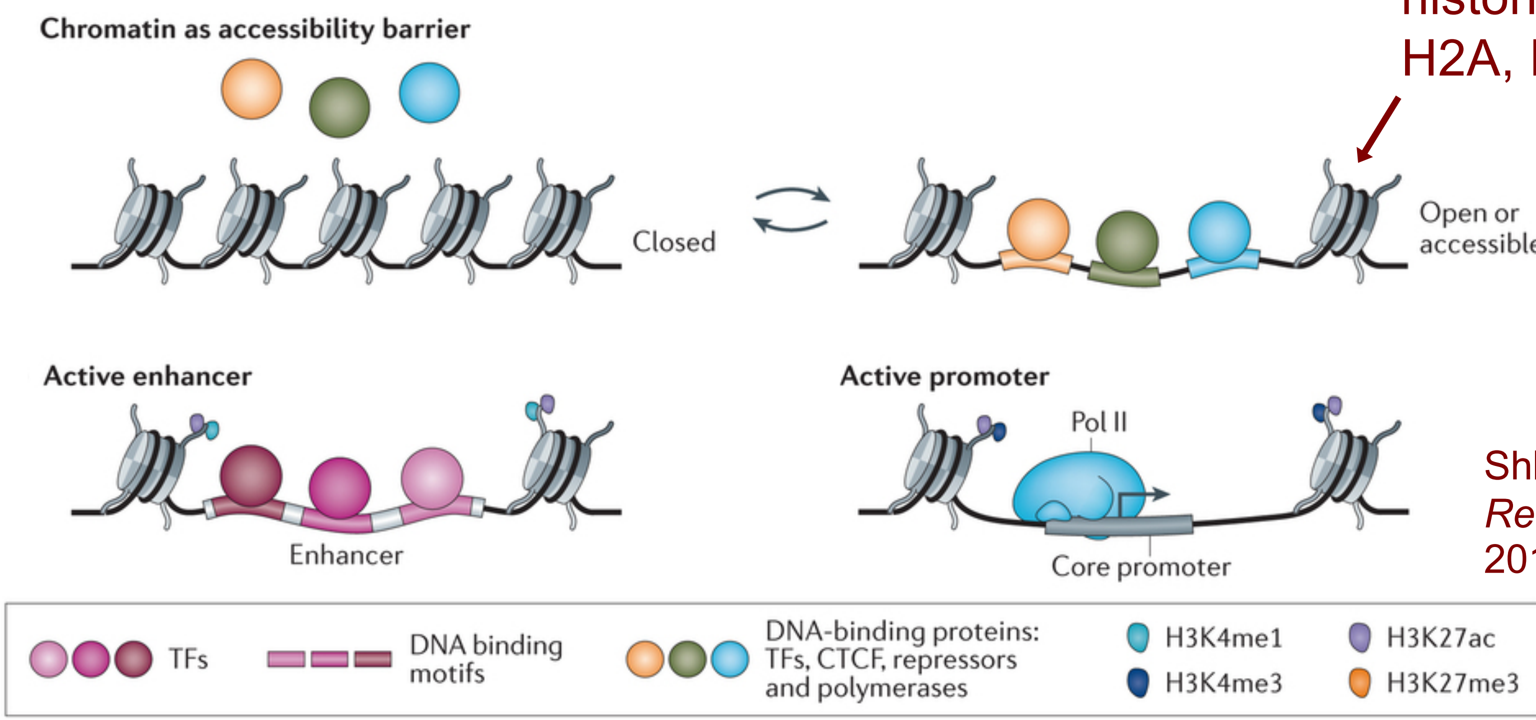


Wang *PLoS ONE* 2012

Histone modifications

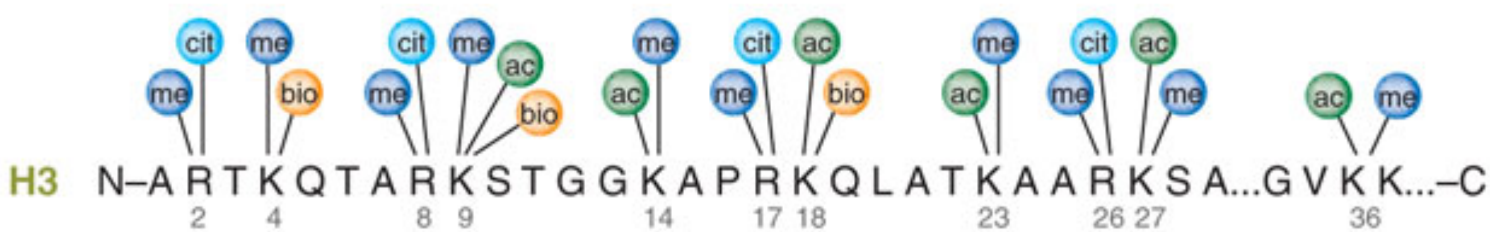
- Mark particular regulatory configurations

Two copies of histone proteins H2A, H2B, H3, H4



Shlyueva *Nature Reviews Genetics* 2014

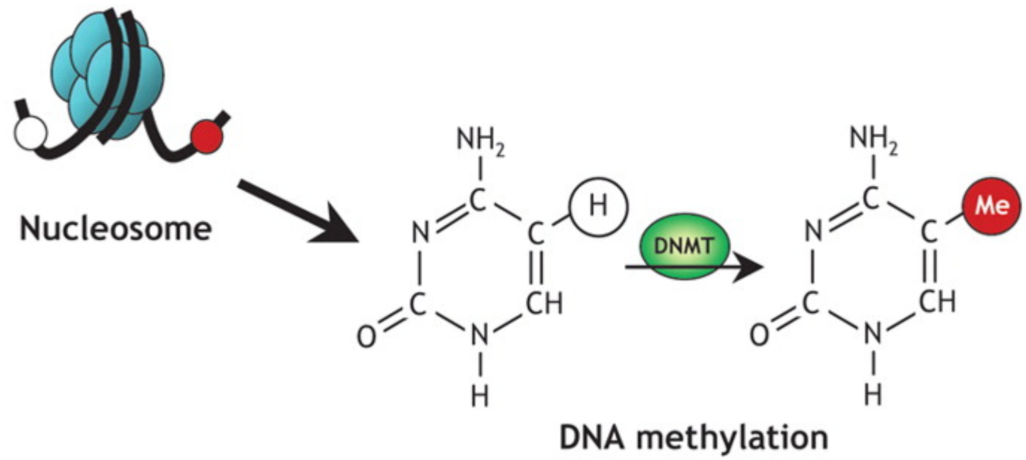
- H3 (protein) K27 (amino acid) ac (modification)

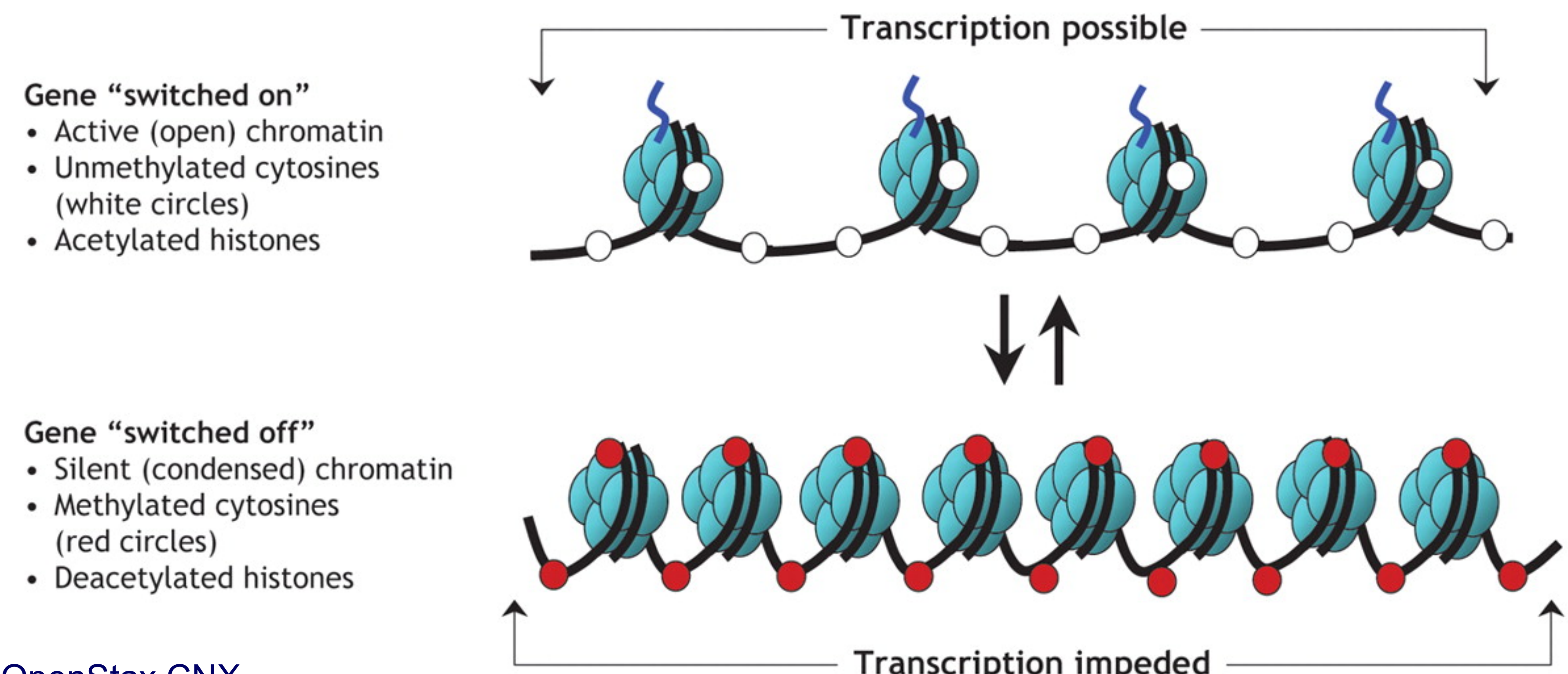


Latham *Nature Structural & Molecular Biology* 2007; Katie Ris-Vicari

DNA methylation

- Reversible DNA modification
- Represses gene expression

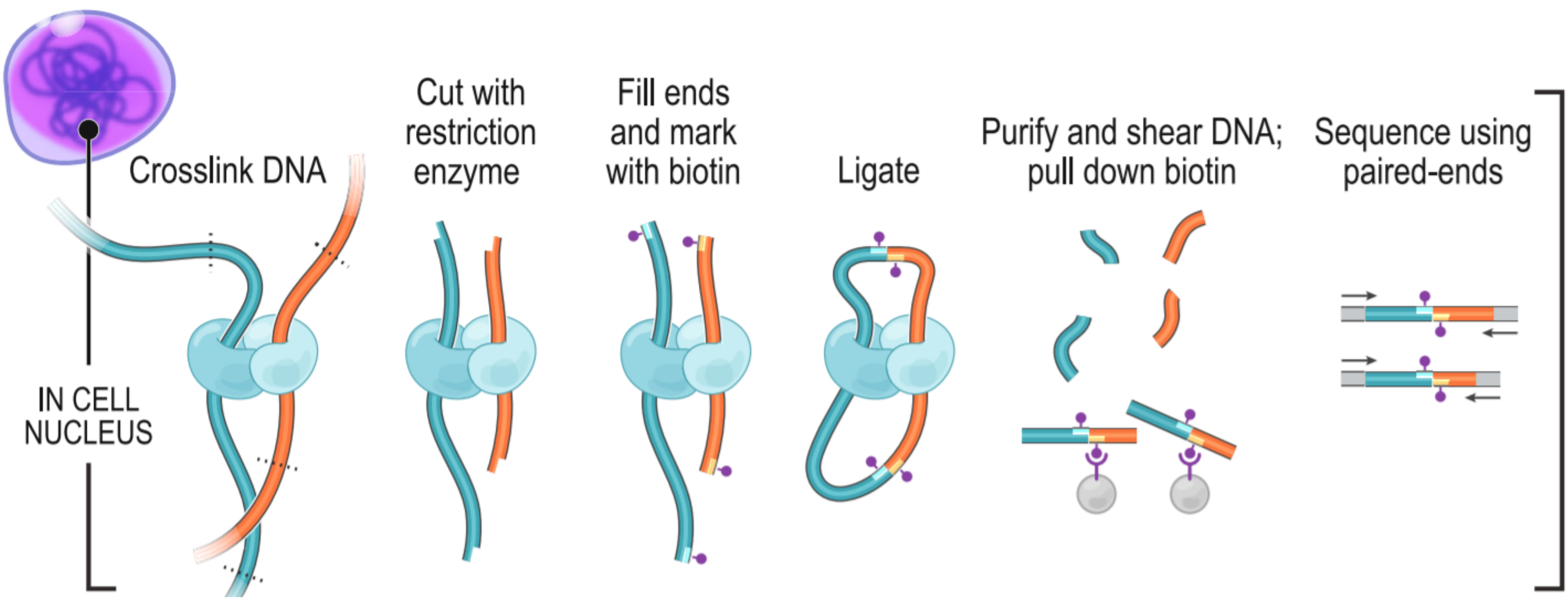




OpenStax CNX

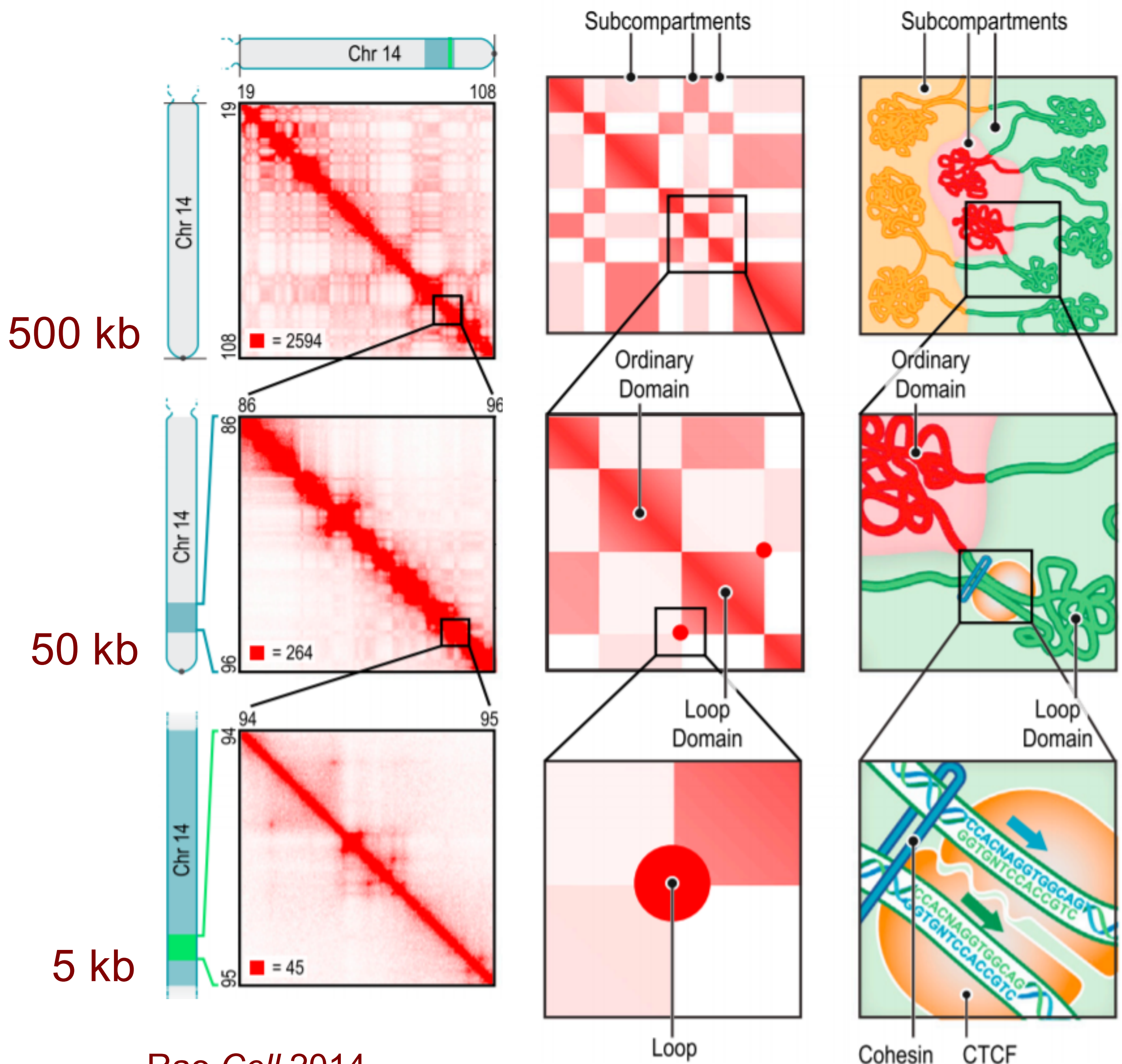
3D organization of chromatin

- Algorithms to predict long range enhancer-promoter interactions
- Or measure with chromosome conformation capture (3C, Hi-C, etc.)



Rao *Cell* 2014

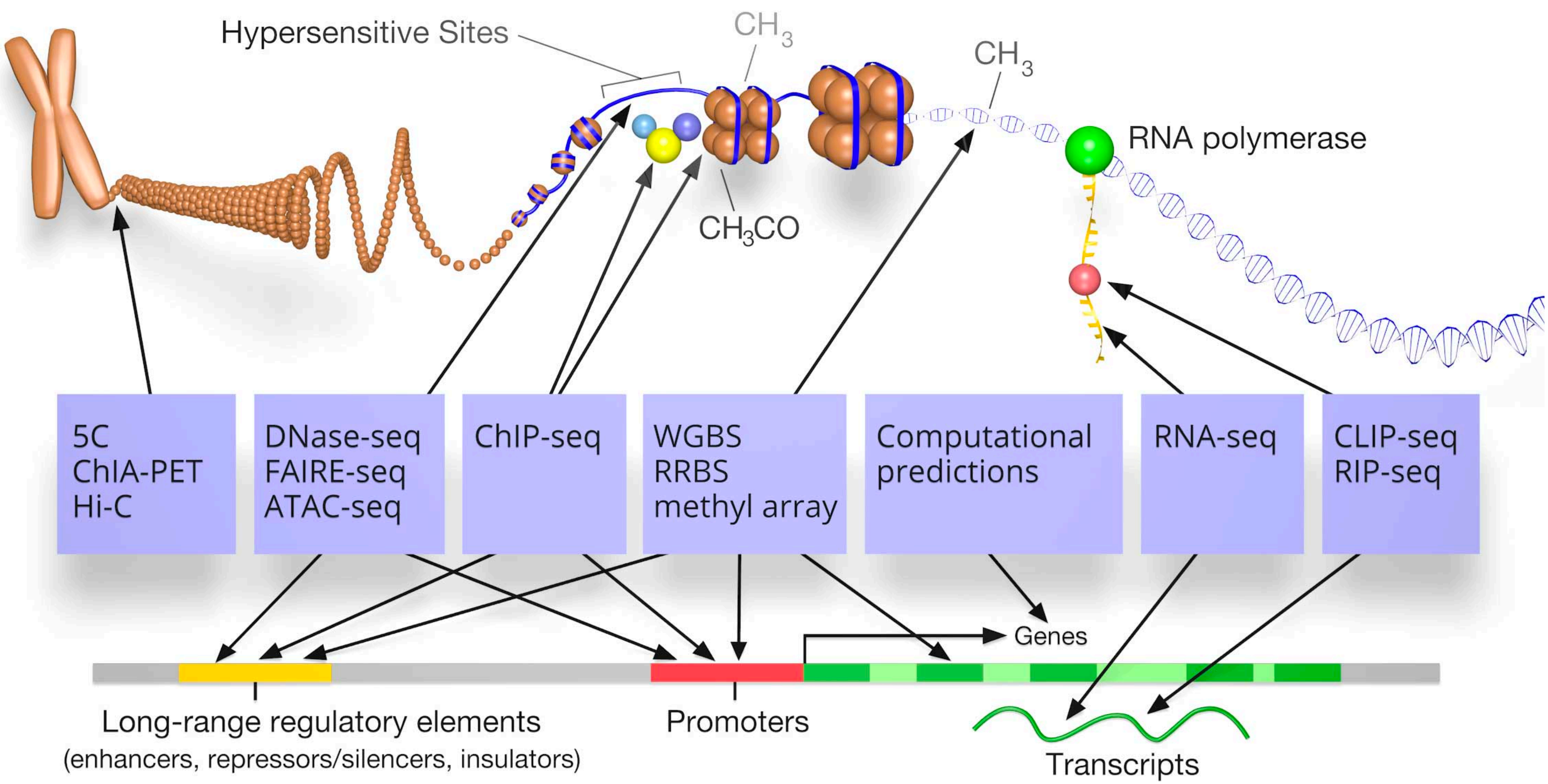
3D organization of chromatin



Rao *Cell* 2014

- Hi-C produces 2D chromatin contact maps
- Learn domains, enhancer-promoter interactions

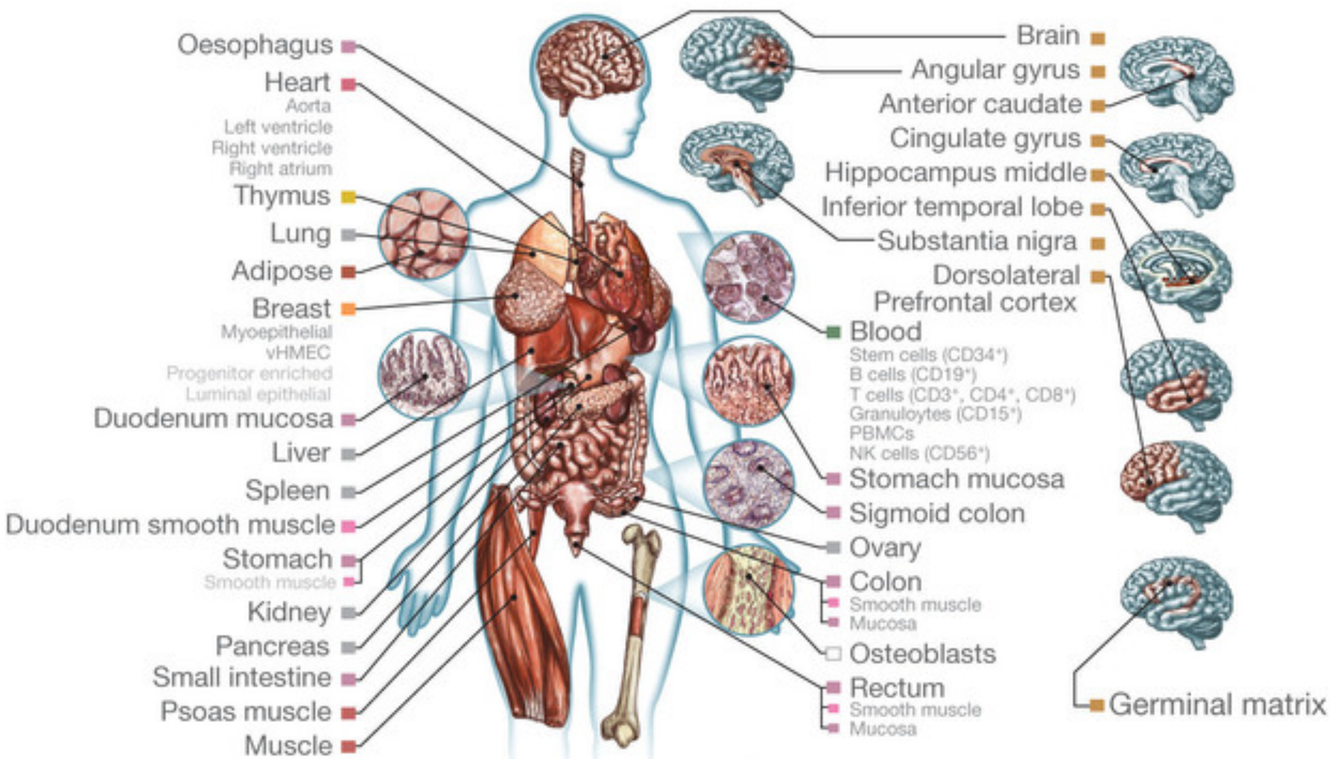
Next Generation Sequencing (NGS) for epigenomics



Based on an image by Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)

Large-scale epigenetic maps

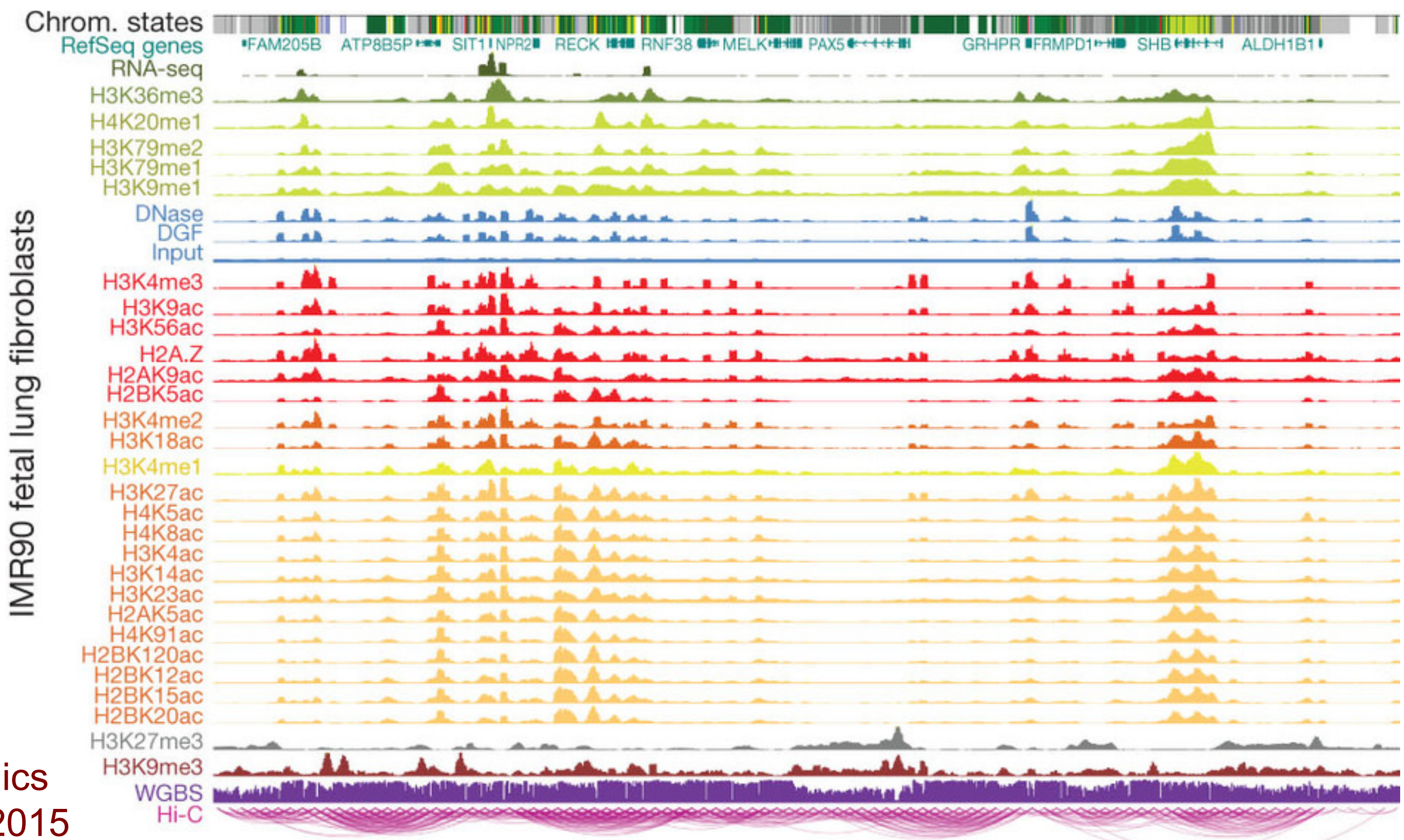
- Epigenomes are condition-specific
- Roadmap Epigenomics Consortium and ENCODE surveyed over 100 types of cells and tissues



Roadmap Epigenomics Consortium *Nature* 2015

Genome annotation

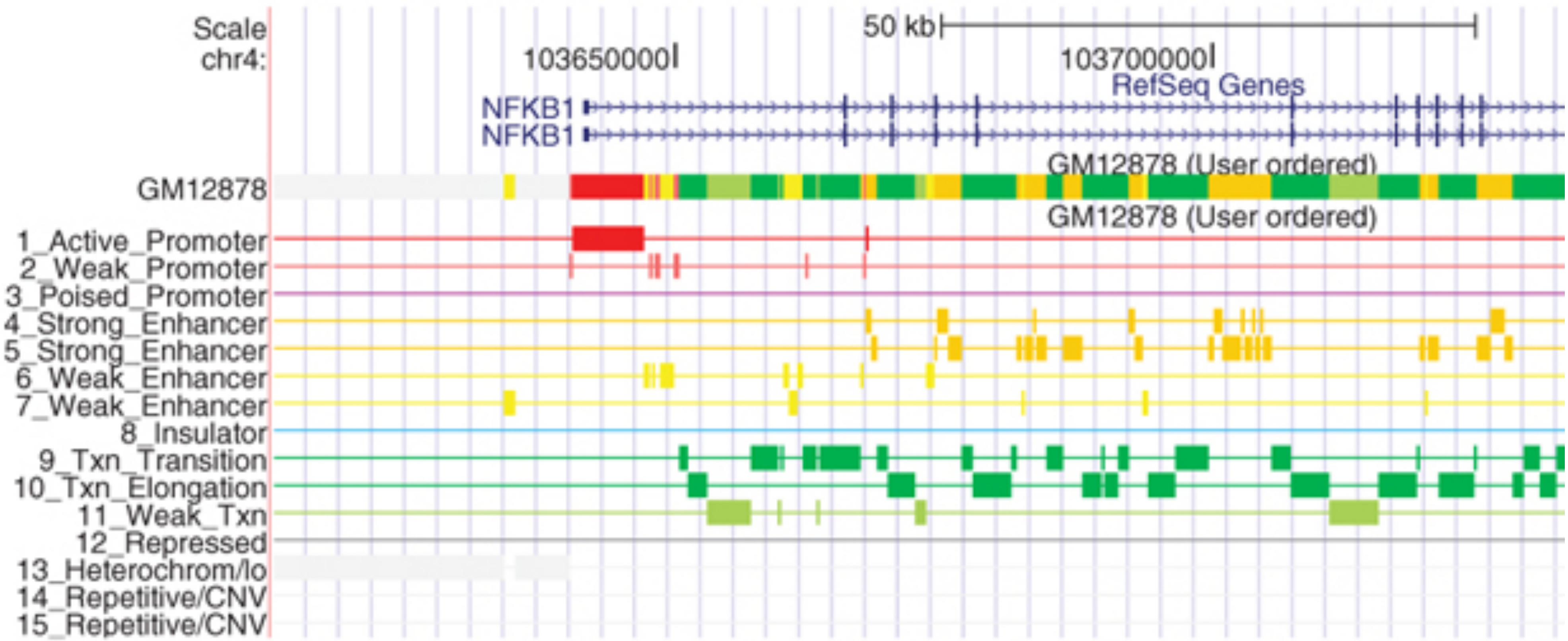
- Combinations of epigenetic signals can predict functional state
  - ChromHMM: Hidden Markov Model
  - Segway: Dynamic Bayesian network



Roadmap Epigenomics Consortium *Nature* 2015

Genome annotation

- States are more interpretable than raw data

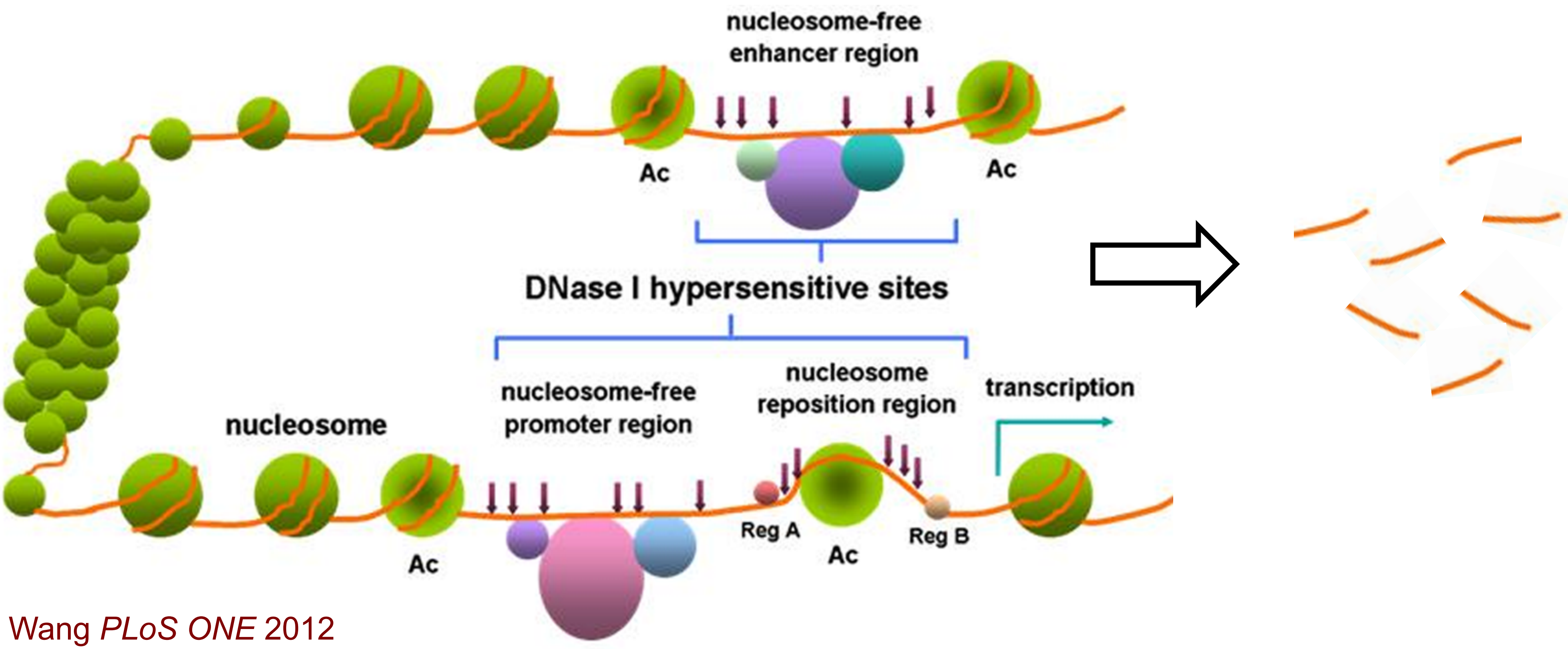


Ernst and Kellis *Nature Methods* 2012

Predicting TF binding with DNase-Seq

DNase I hypersensitive sites

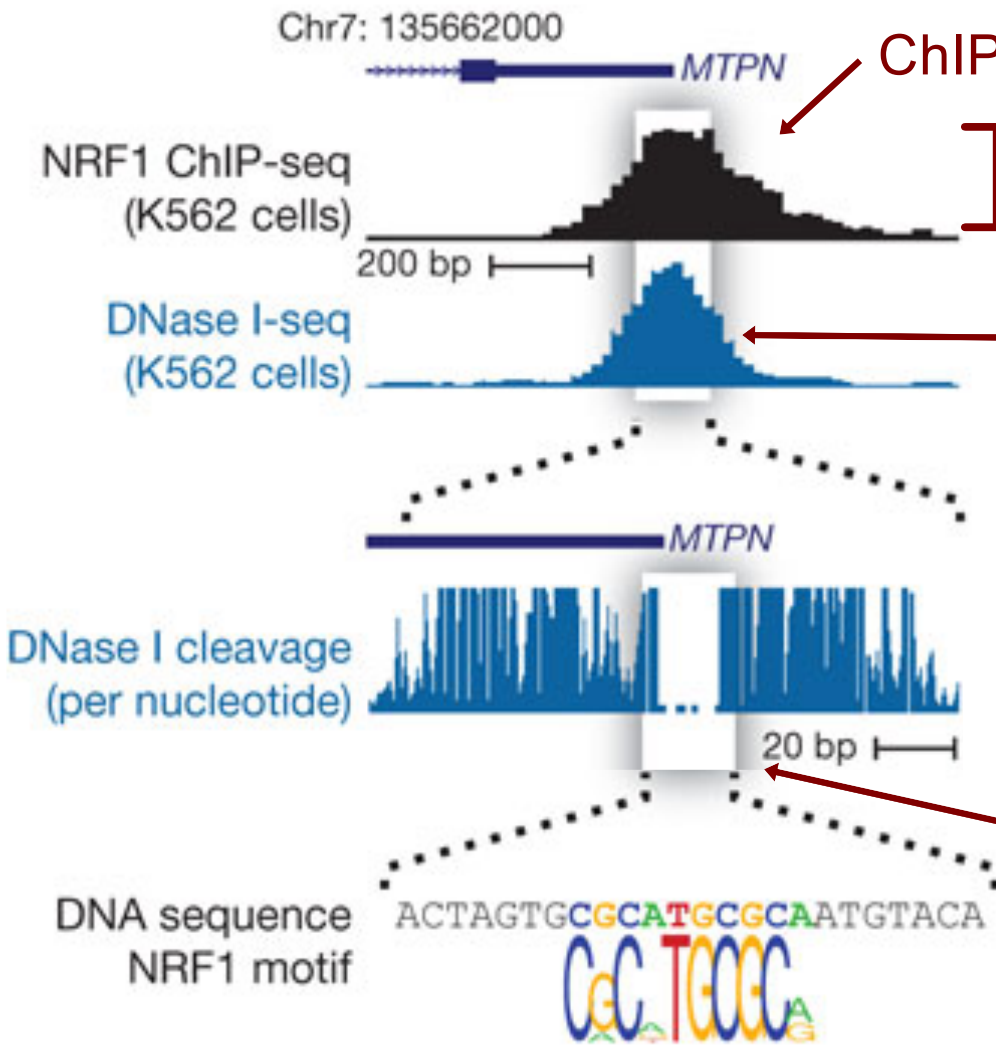
- Arrows indicate DNase I cleavage sites
- Obtain short reads that we map to the genome



Wang *PLoS ONE* 2012

DNase I footprints

- Distribution of mapped reads is informative of open chromatin and specific TF binding sites



ChIP-Seq peak

Read depth at each position

Nucleosome free "open" chromatin

Zoom in

TF binding prevents DNase cleavage leaving Dnase I "footprint", only consider 5′ end

Neph *Nature* 2012

DNase I footprints to TF binding predictions

- DNase footprints suggest that ***some*** TF binds that location
- We want to know ***which*** TF binds that location
- Two ideas:
  - Search for DNase footprint patterns, then match TF motifs
  - Search for motif matches in genome, then model proximal DNase-Seq reads

We'll consider this approach