



Gene regulation and epigenetics

Carl Herrmann

Institut für Pharmazie und Molekulare Biotechnologie (IPMB)

WiSe 2025/2026

Our primary interests

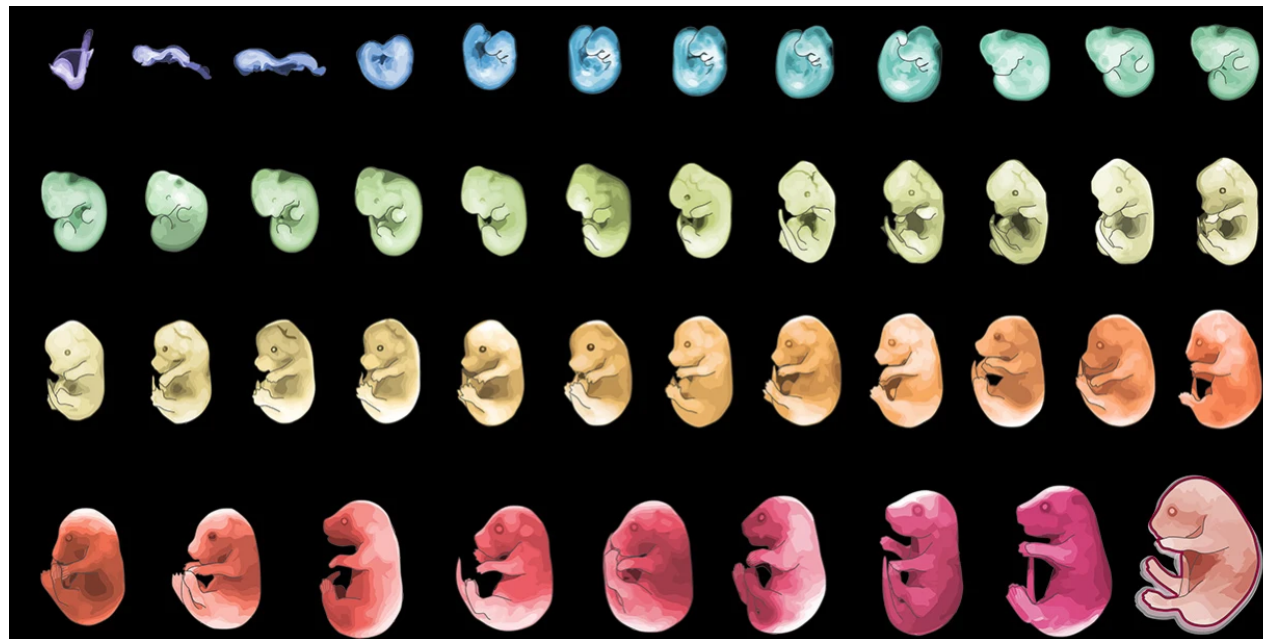
- Understanding the mechanisms of **regulatory genomics** in development and disease (especially cancer)
 - neuroblastoma / glioblastoma
 - epigenomics
 - role of transcription factors
- **Methods development** for integration of omics datasets (especially single-cell)
 - molecular signature extraction
 - single-cell multi-omics
- Integration of clinical and omics data using ML approaches



- Carlos Ramirez (Postdoc)
- Robin Droit (Postdoc)
- Daria Doncevic (Postdoc)
- Albert Li (PhD)
- Thien Duc Bui (PhD)
- Jean Radig (PhD)

Topic of this lecture

- How does the **diversity of tissues** and cell types develop from the **same DNA sequence**?
- How are the developmental programs controlled in **time** (across developmental phases) and **space** (between tissues and cell types)?



[Qiu et al., Nature]

Content of the lecture

- | | |
|--------------|--|
| 23.10 | <ul style="list-style-type: none">1. General aspects of transcriptional regulation2. Introduction to regulatory genomics3. Available data types |
| 30.10 | <ul style="list-style-type: none">4. Bioinformatics methods in regulatory genomics5. Improving regulatory predictions with integrative models6. Hands-on: looking at data! |



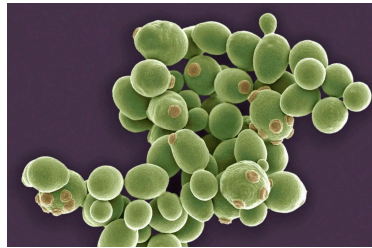
1. Introduction to gene regulation

Complexity of an organism

Bacteria



Yeast
(unicellular eukaryote)



C. elegans
(multicellular metazoan)



Mouse
(multicellular mammal)

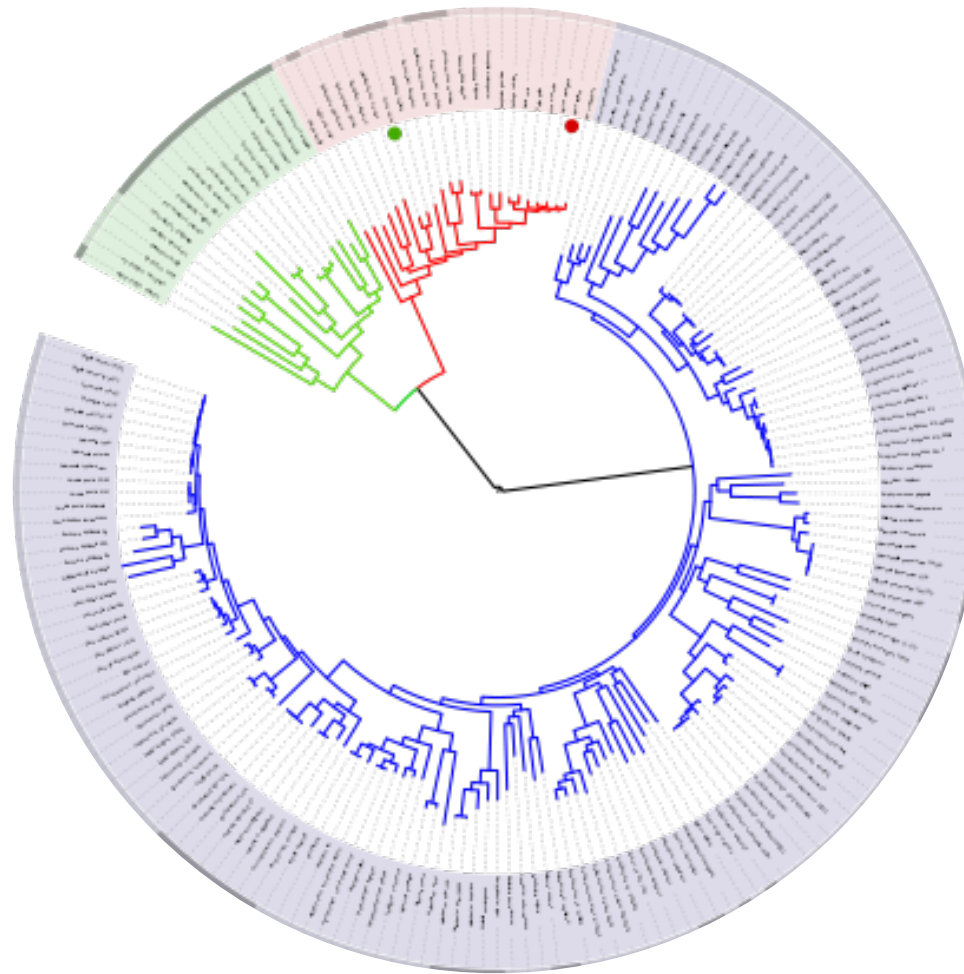


complexity

- More celltypes
- More tissues
- More physiological processes
- More genes ?
- Longer genome ?
- ?

Genome size

Genome size

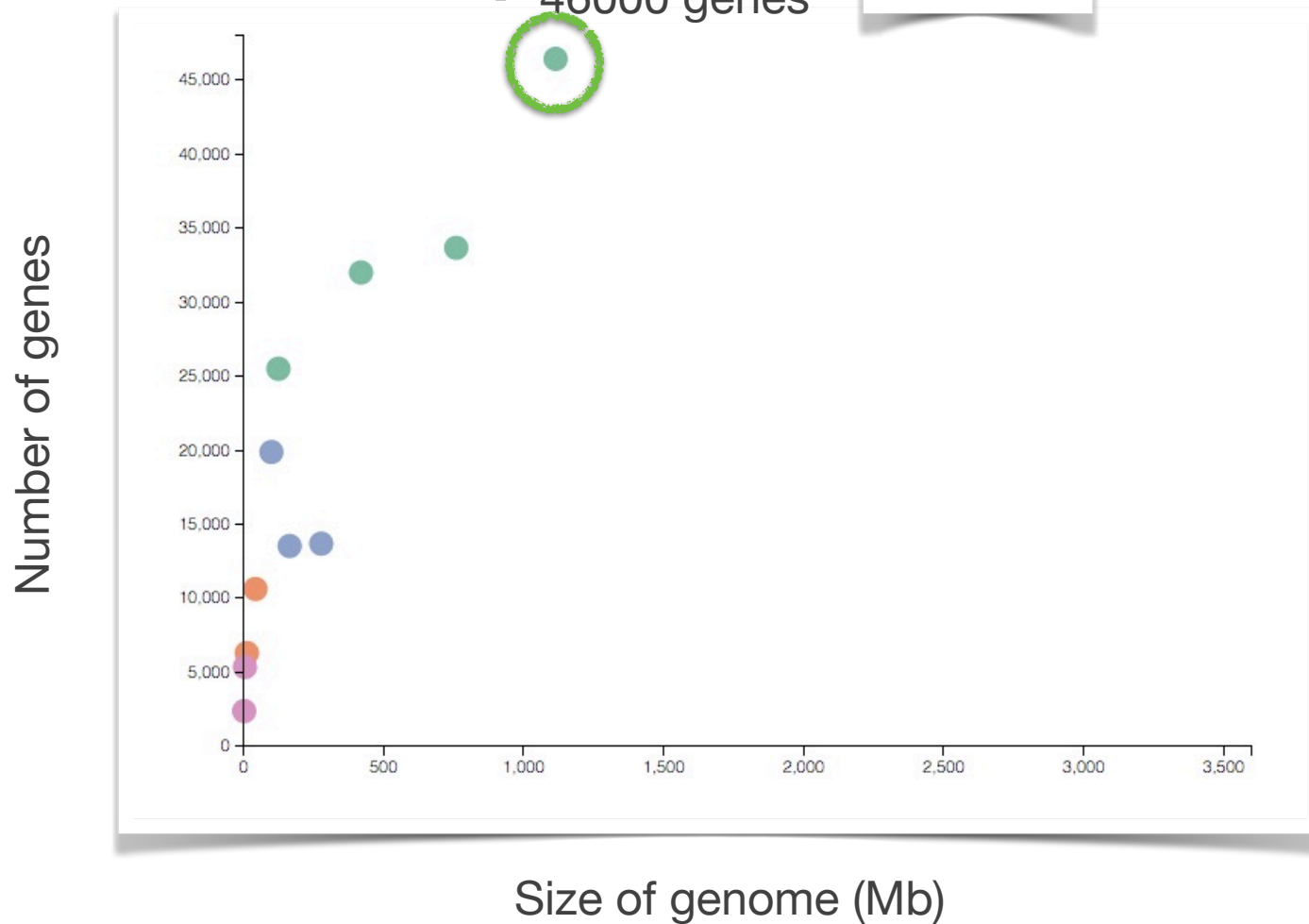


Eukaryotes
Archaea
Bacteria

[interactive Tree of Life]

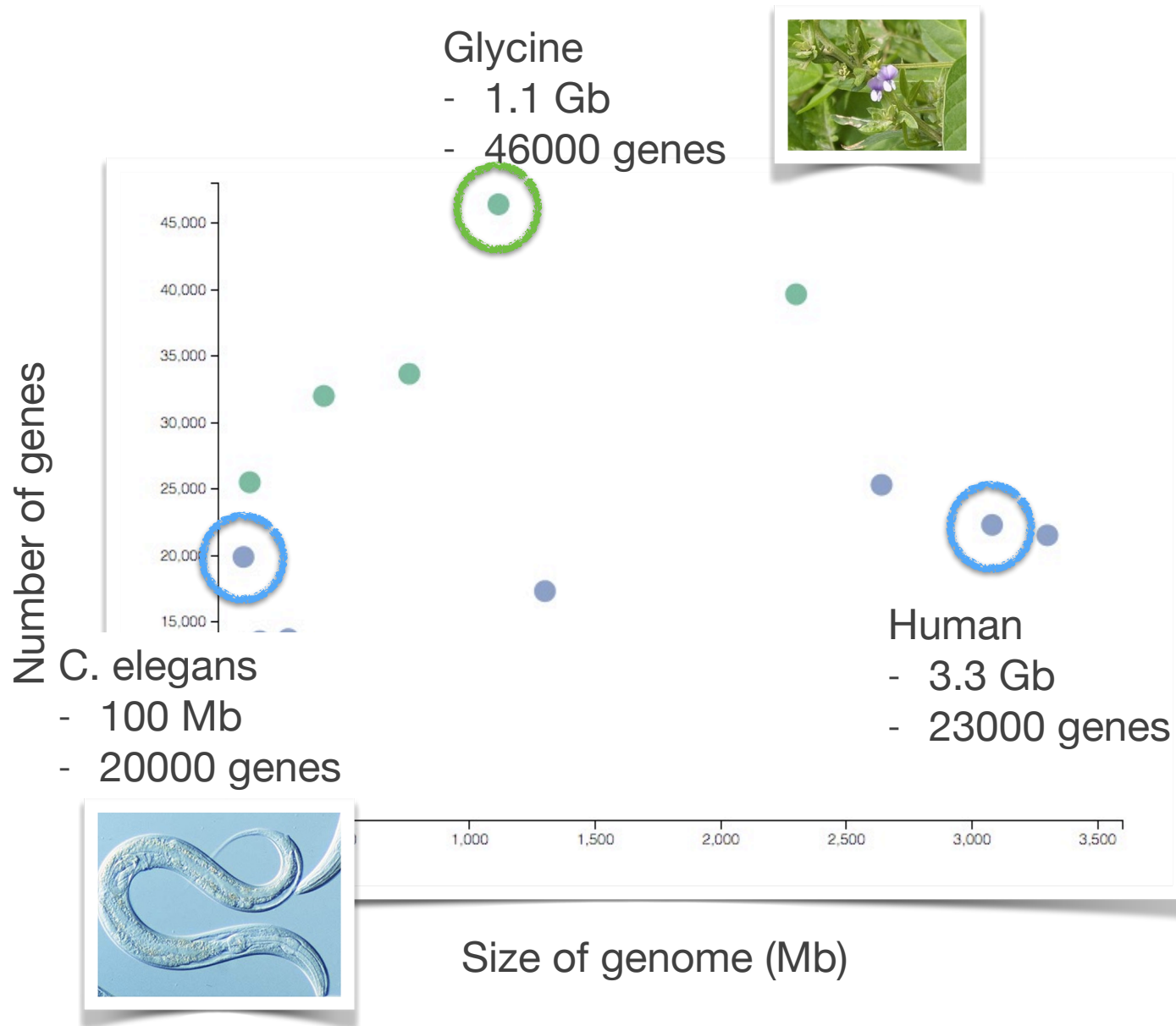
Number of genes

Glycine
- 1.1 Gb
- 46000 genes



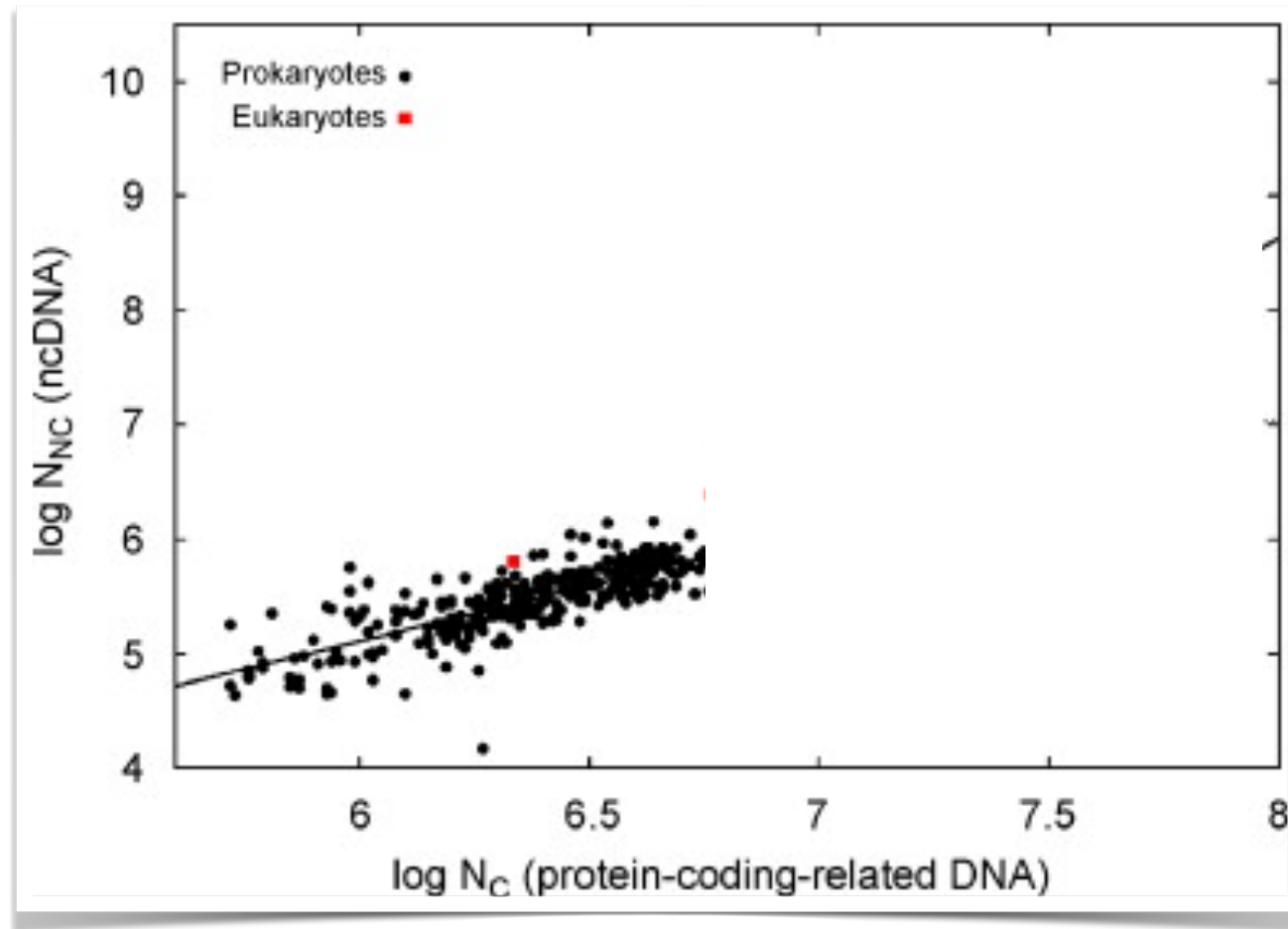
[https://bionumbers.hms.harvard.edu/files/Genomic census for a variety of selected organisms.pdf](https://bionumbers.hms.harvard.edu/files/Genomic%20census%20for%20a%20variety%20of%20selected%20organisms.pdf)

More genes = more complex ?



[<https://gf.neocities.org/gs/genes.html>]

Bigger non-coding genome = higher complexity



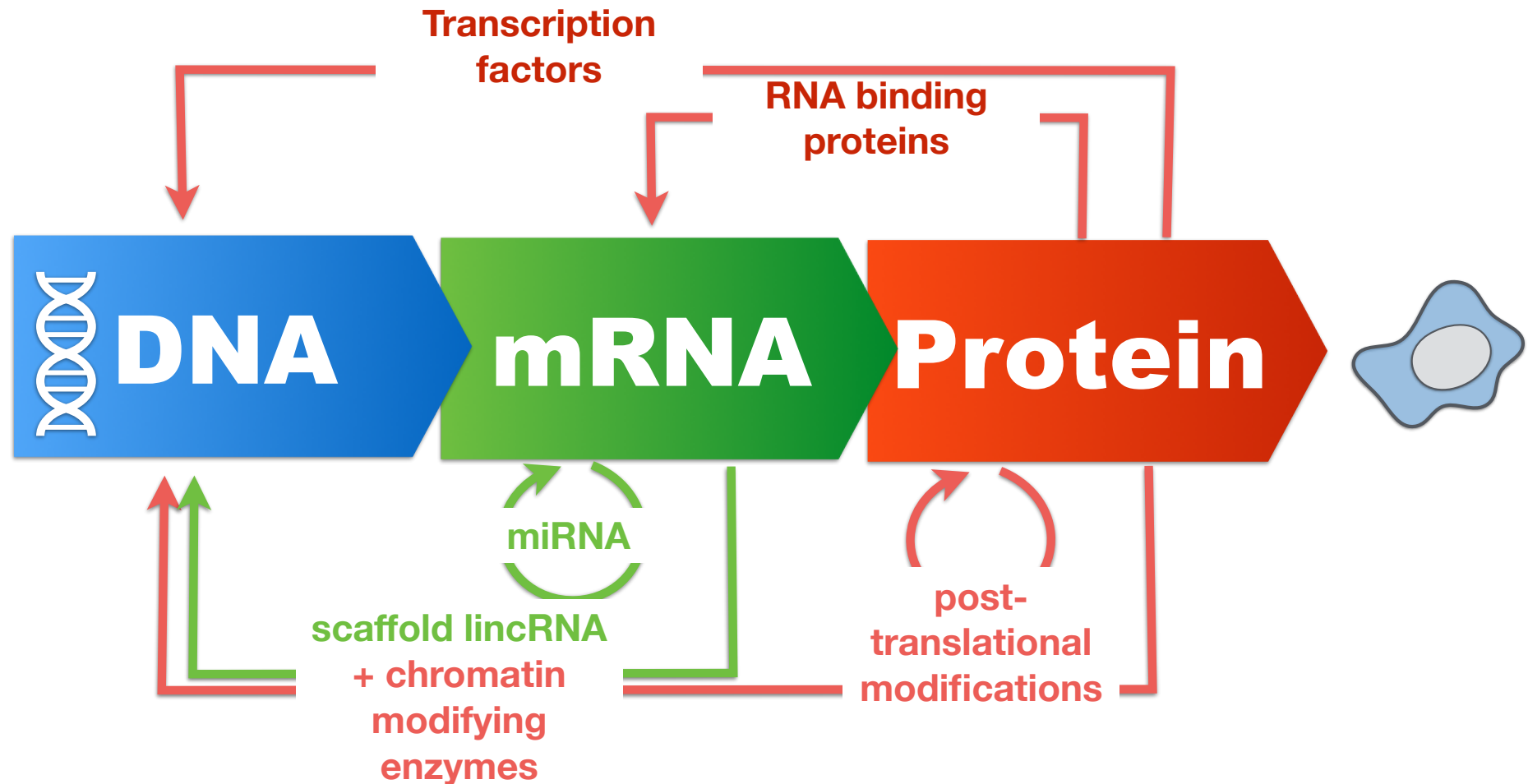
Proportion of **non-coding DNA** correlates with organismal complexity

[Ahnert, Fink, Zinovyev, 2008]



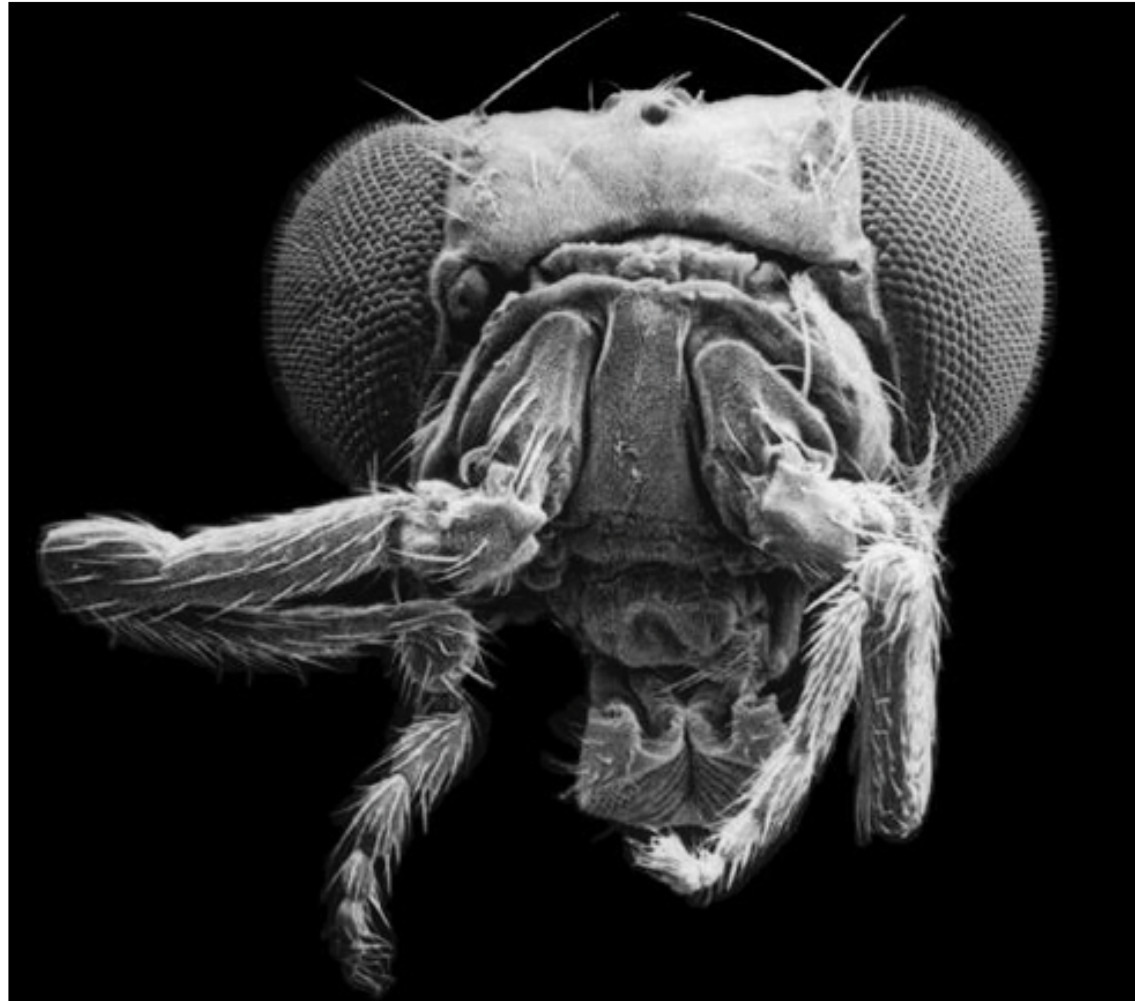
2. Transcriptional regulation

The Dogma in the genomics area



How to ensure that the right genes/proteins are expressed at the right time in the right cells?

Spatial control of gene expression

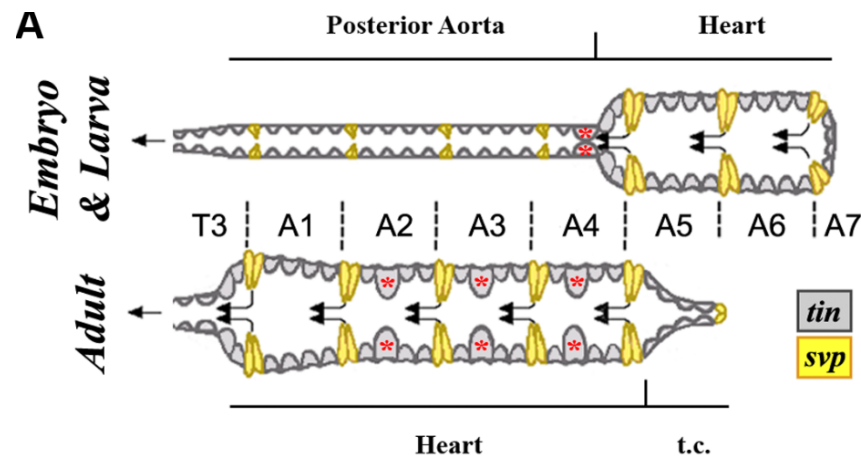


[F.R. Turner]

***Antp* expression in the *Drosophila* head turns antennae into legs...**

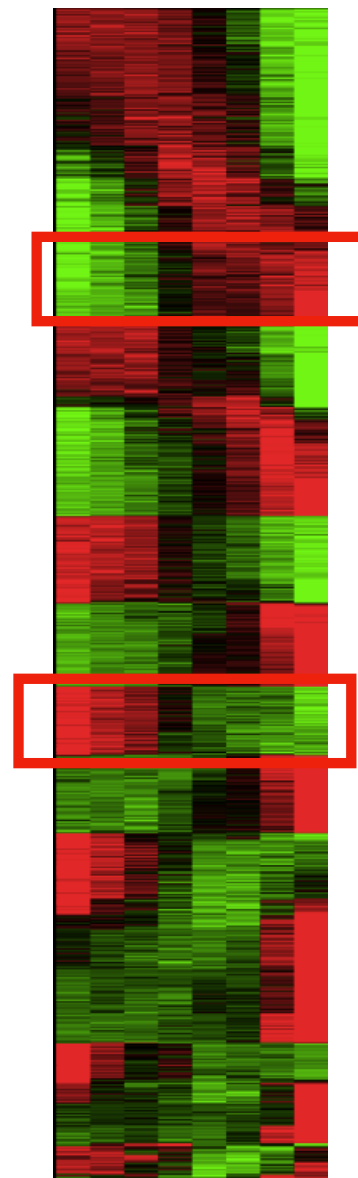
Temporal control of gene expression

- Time-course experiment on developing **heart chamber** in *Drosophila*
- Gene expression profiles show **coordinated dynamical regulation**
- How is this tight expression patterns ensured?

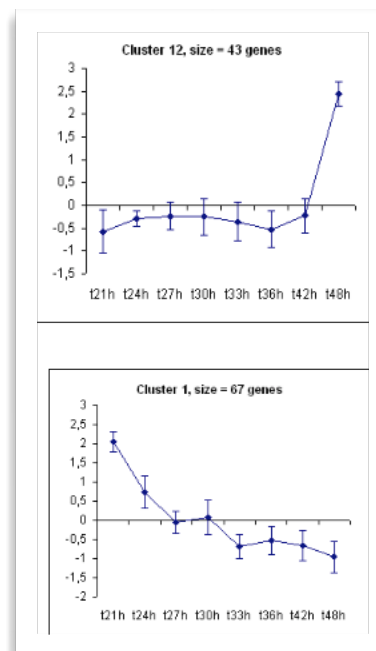


time (h)

21 24 27 30 33 36 42 48

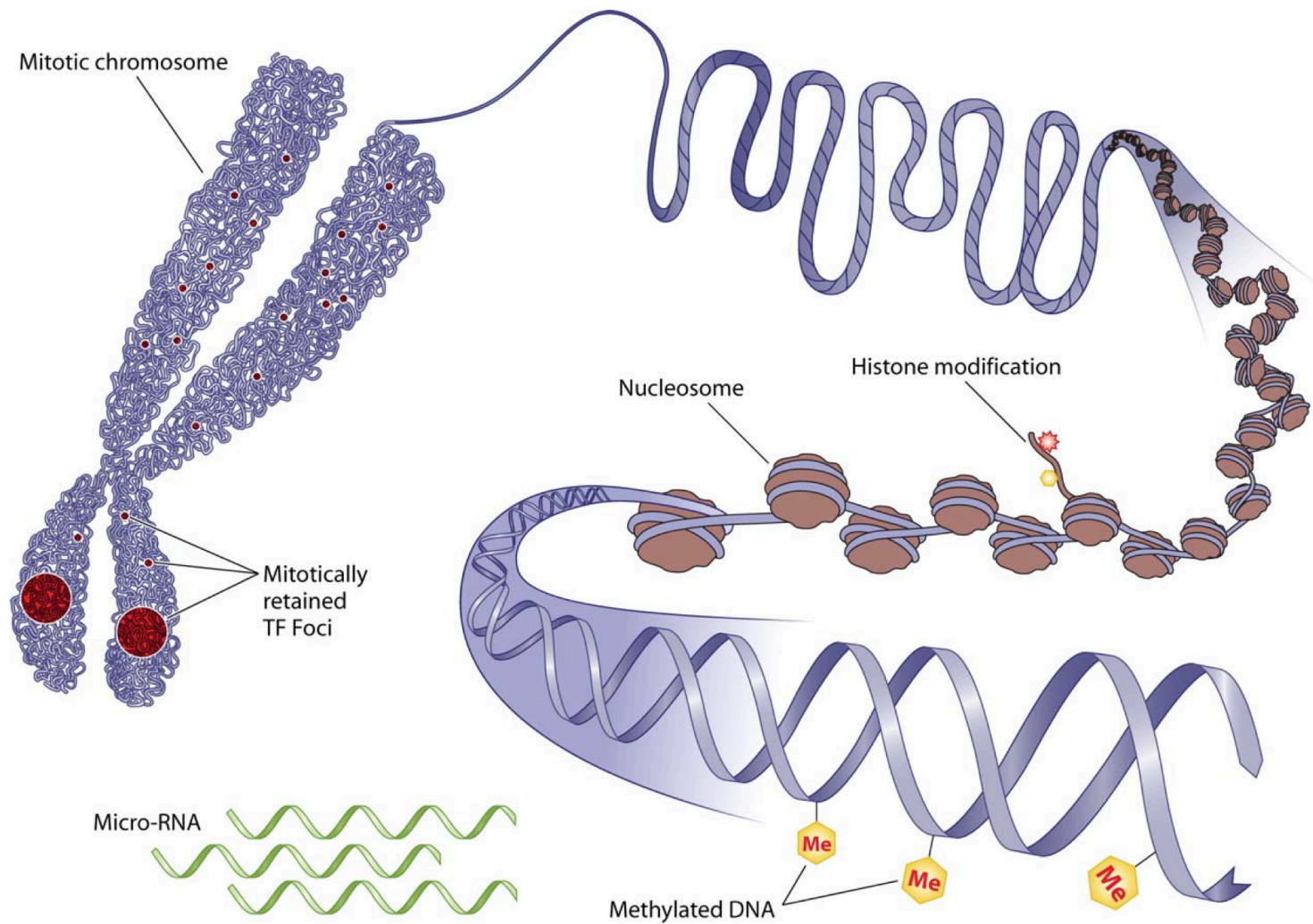


red = upregulation
green = downregulation



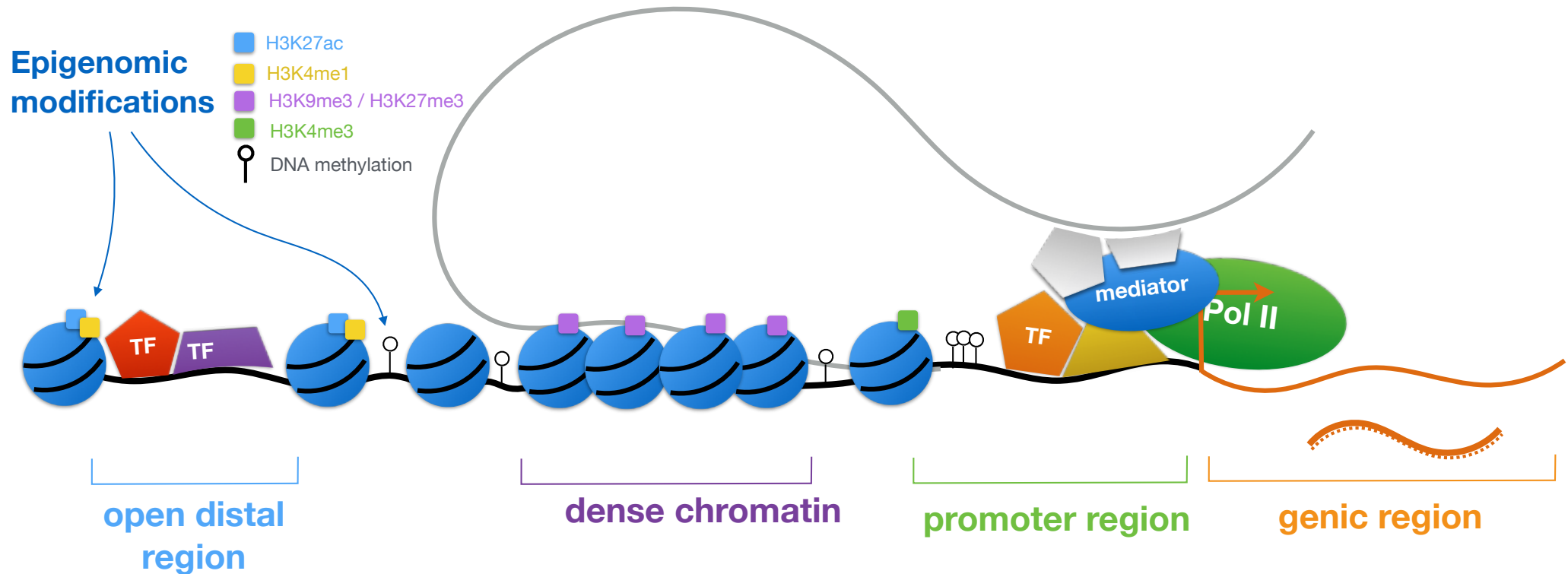
[Zeitouni et al, PLOS Genetics (2007)]

Cellular control room



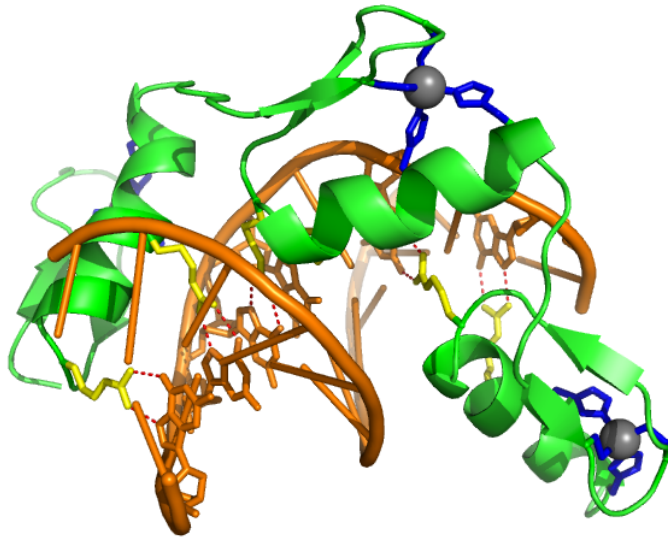
[Zaidi et al., Mol.Cell.Bio 2010]

Transcriptional regulation



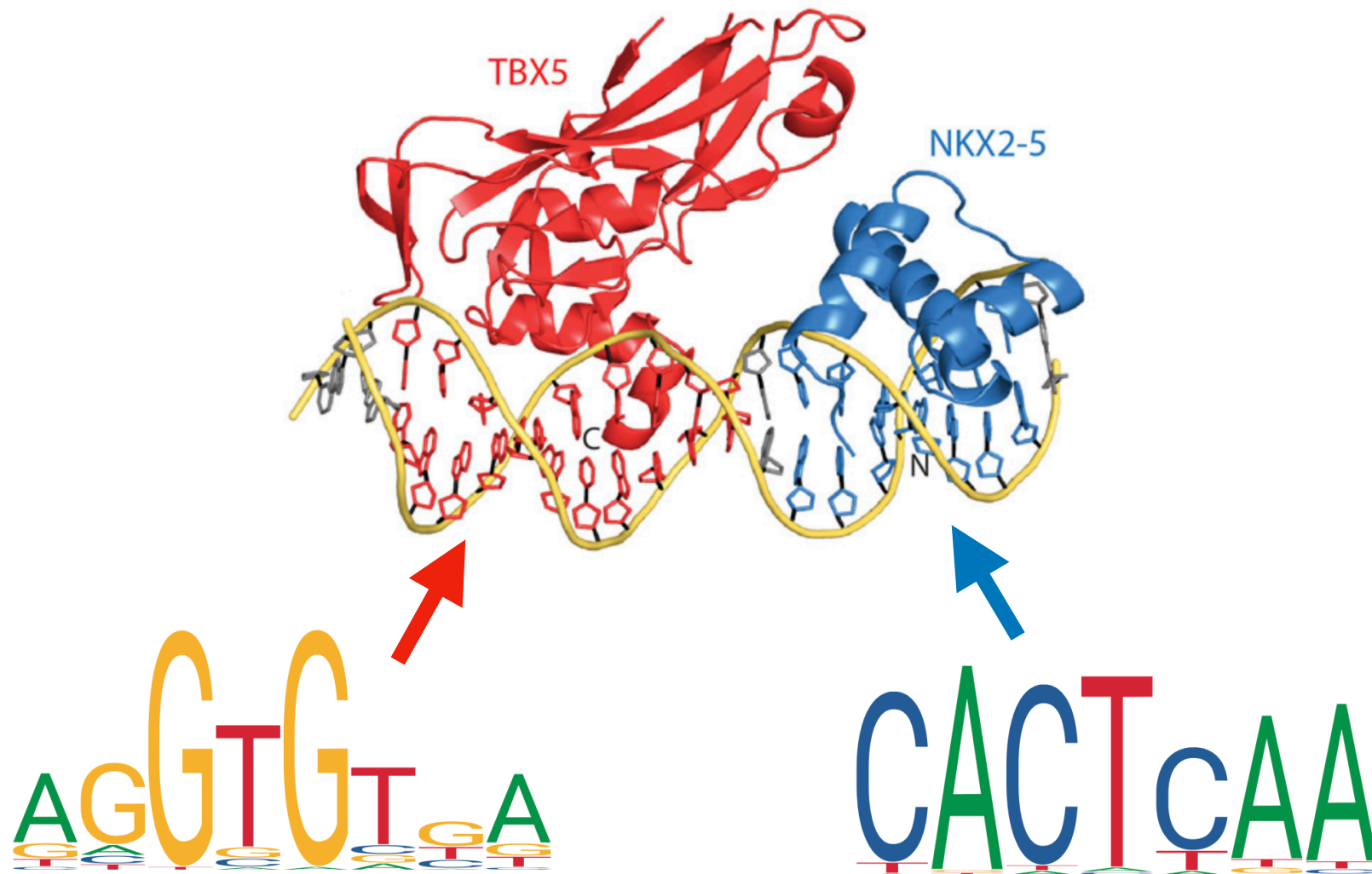
combinatorial interplay of multiple components

Transcription factors



- Transcription factors (TFs) are proteins which can bind the DNA
- TFs contain a **DNA binding domain** (DBD) and a **transcriptional activator** (TA)
- Transcription factors are grouped into structural families

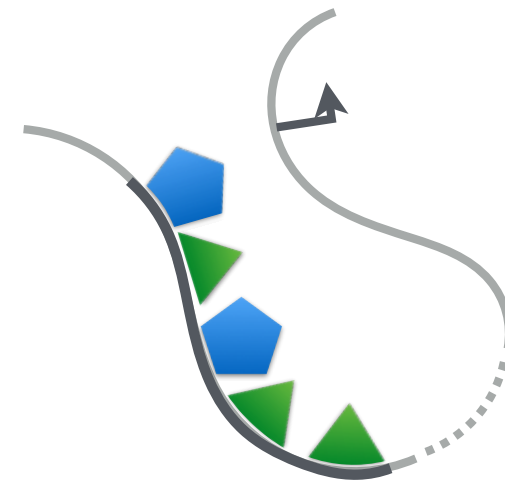
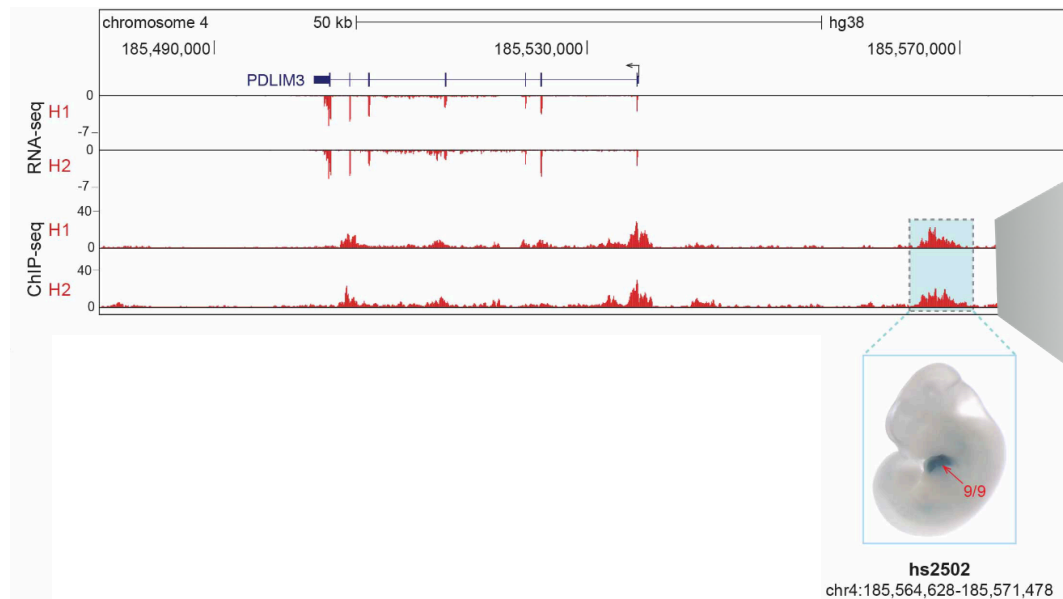
Transcription factors



Transcription factors

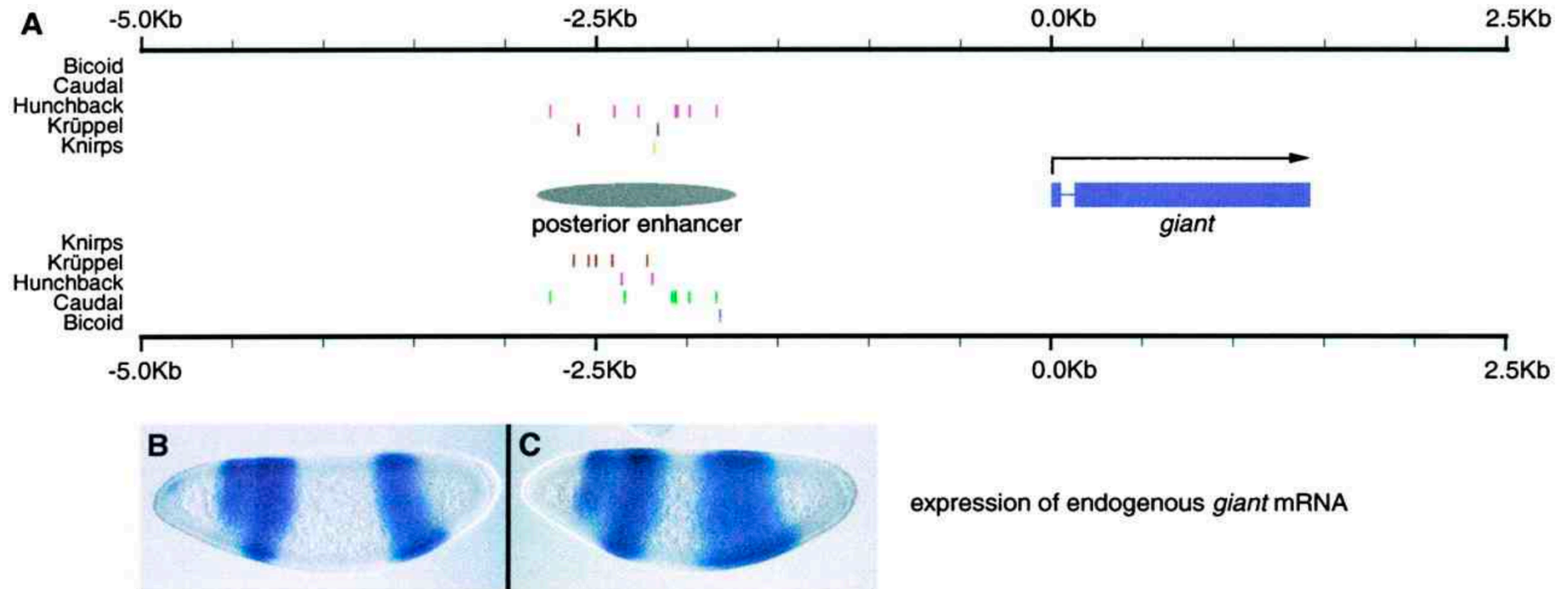
- **Enhancers** are regulatory elements which can be located far from the target genes
- **Multiple binding sites** for different transcription factors
- typical length: few kb → several hundred kb ("superenhancers")
- Organisational principles ("grammar") remains unclear

Identification of heart enhancers

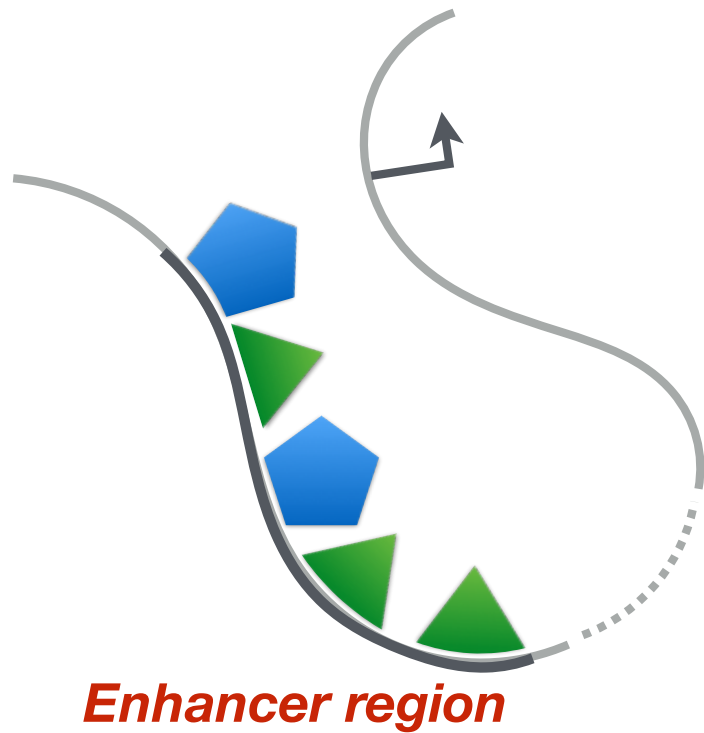


Enhancer region

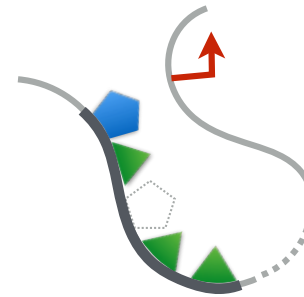
Enhancers in development



Transcriptional regulation

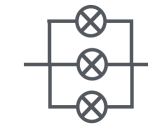


enhanceosome model



off

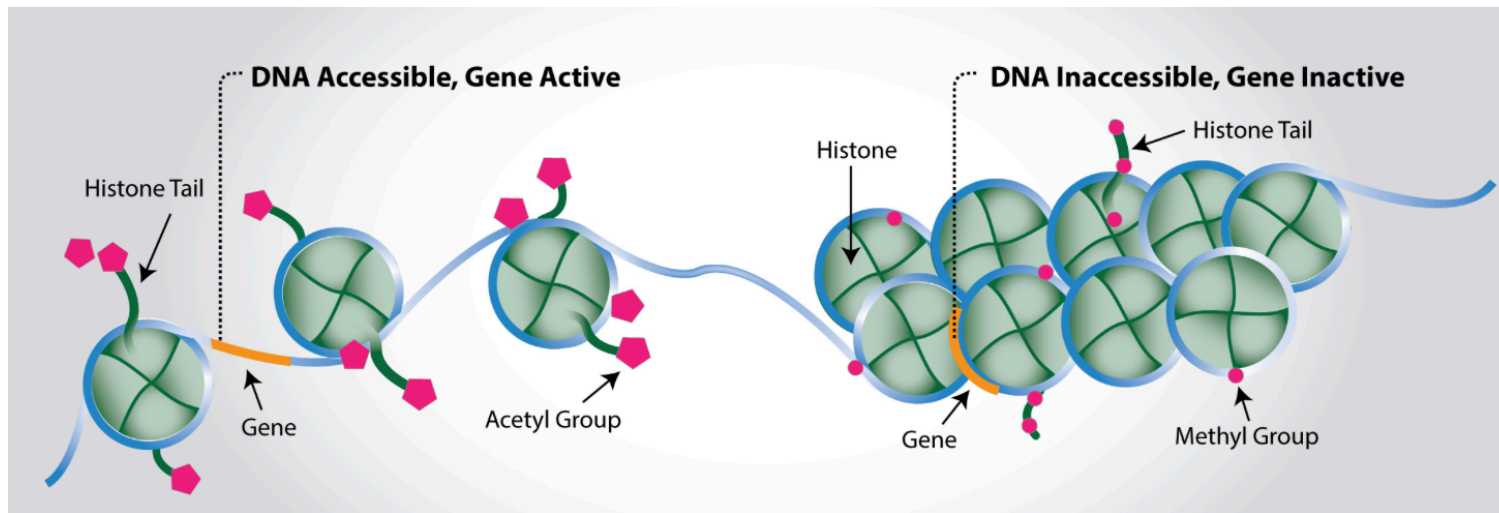
billboard model



on

[Arnosti, Kulkarni, 2005]

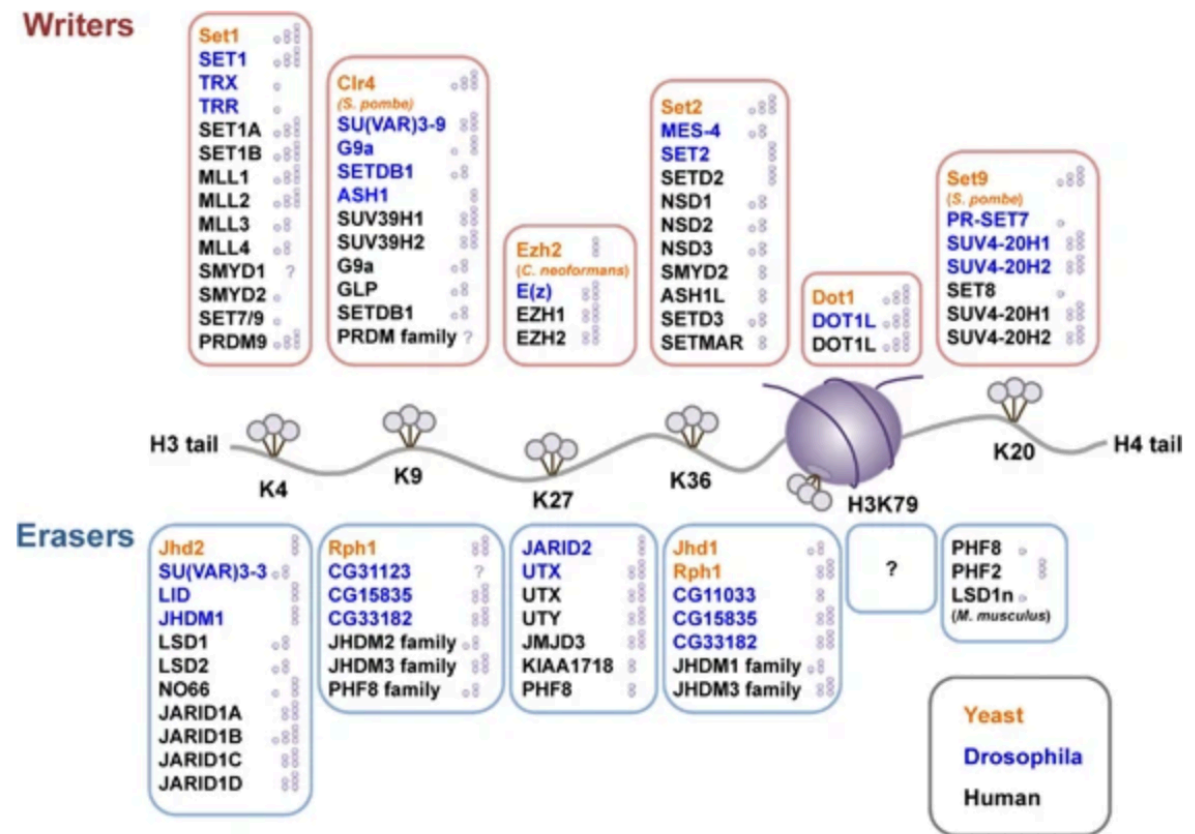
Epigenetic: histone modifications



- **Nucleosomes** are octamers composed of eight different histone proteins (2 x H2A; 2 x H2B; 2 x H3; 2 x H4)
- histones are subject to **post-translational modifications** at their N-terminal tail (acetylation, methylation, ...)
- **modify the physical properties of the DNA-nucleosome interactions**
- **Acetylation:** allows nucleosomes to slide away (→ activating)
- **Methylation** : activating/repressive, depends on the residue

Epigenetic: histone modifications

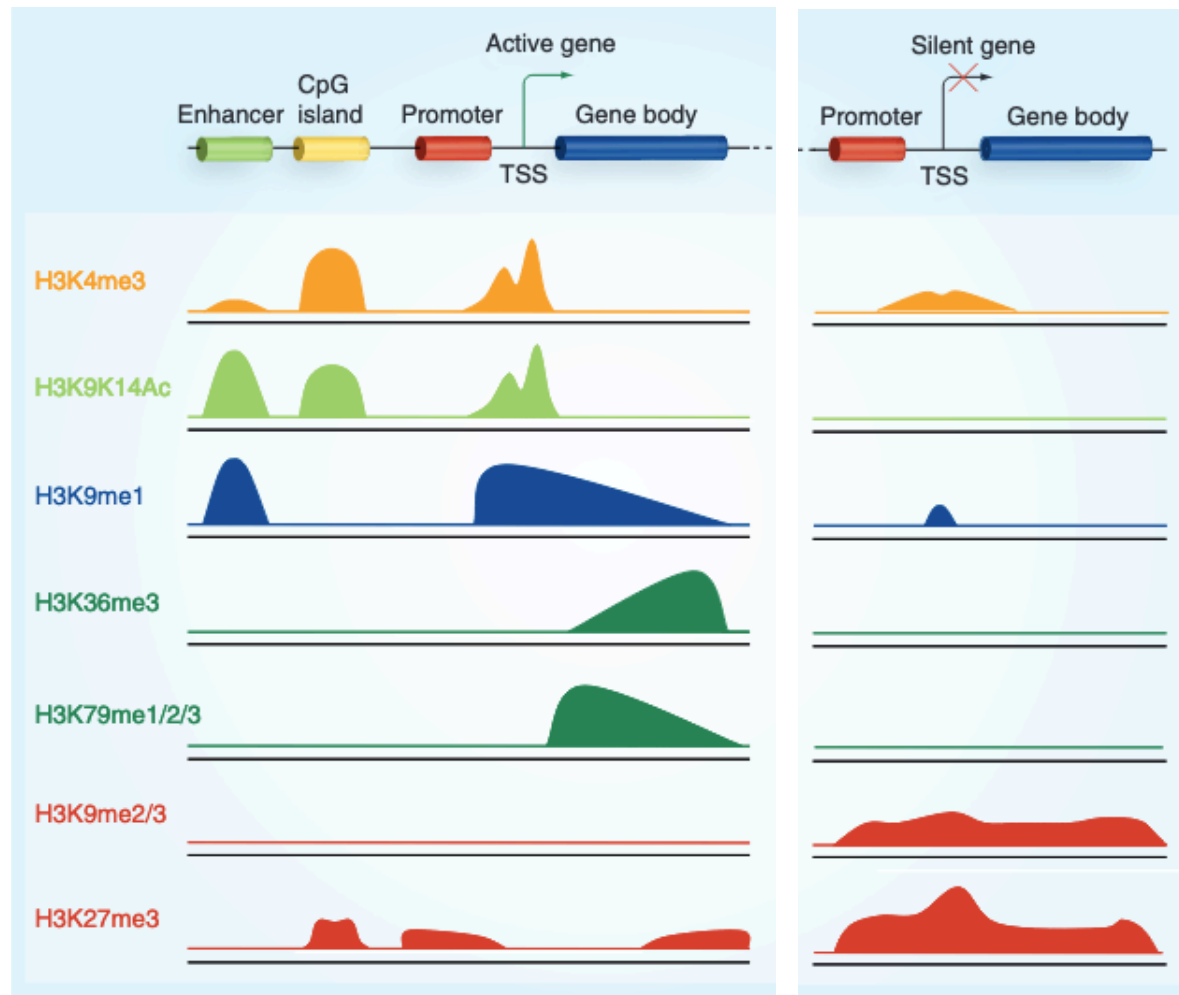
- Many epigenetic enzymes are involved in histone modifications
 - writers:** set the epigenetic modifications (methylation, acetylation, ...)
 - erasers:** unset these modifications
 - readers:** detect and bind to residues that have specific modifications
- Mutations** in these enzymes are often encountered in cancer



[Hyun et al., EMM 2017]

Epigenetic: histone modifications

histone modifications are a good proxy of gene expression and presence of regulatory elements



H3K27ac = acetylation of lysine 27 on histone 3

[Lin, Shannon, Hardy, 2010]

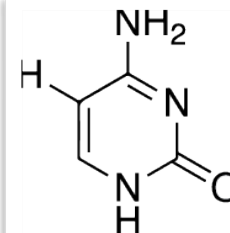
Epigenetics: DNA methylation

- DNA methylation occurs mainly on **cytosines** in **CpG** dinucleotides in the human genome (28 million in human genome!)
- Methylation is set/unset by specific **enzymes**:
 - DNA methyltransferases (DNMT) on C5
 - demethylated by other enzymes (e.g. TET enzyme)
- In the human genome, **3-6% of CpGs are methylated**
- **CpG islands** (high density of CpGs) are found at promoters and are usually unmethylated

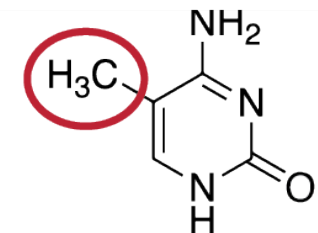
...**A T C * G G A T C T A G**...



...**A T C G G A T C T A G**...

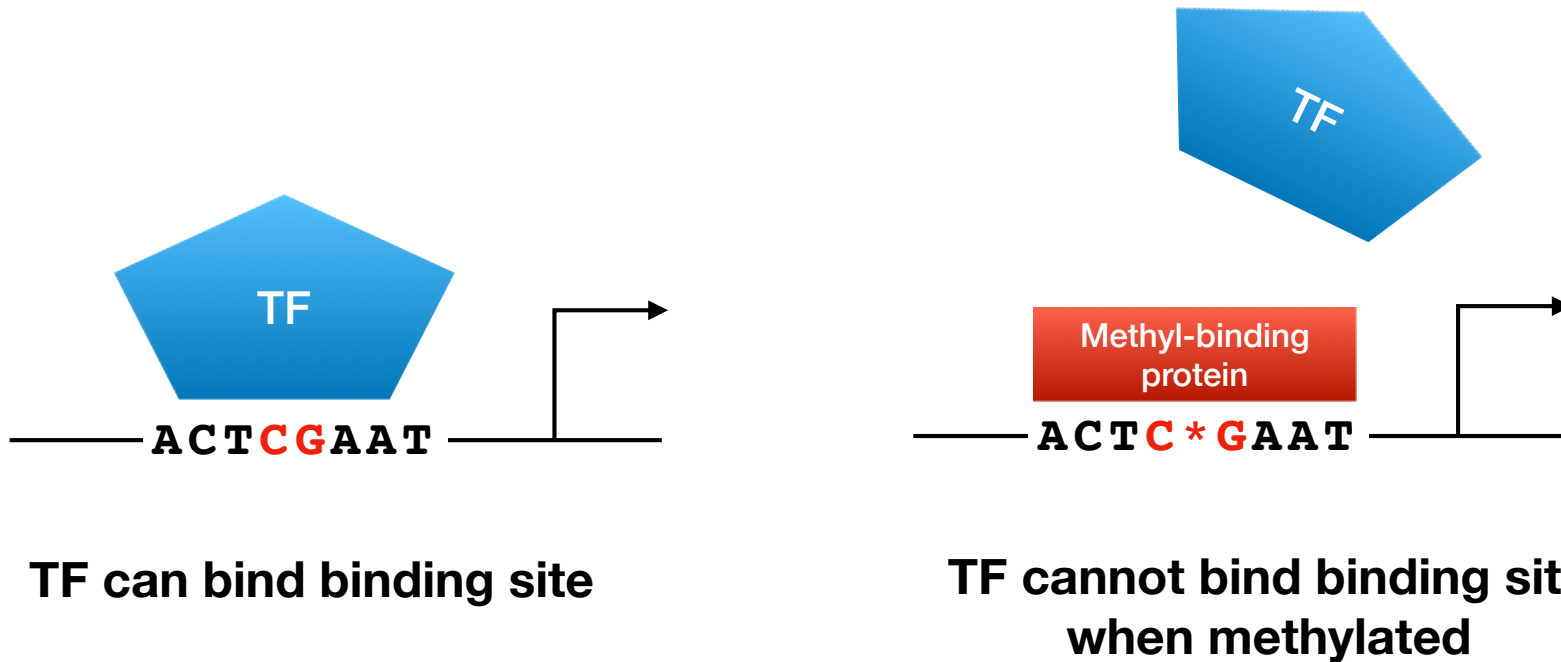


Cytosine



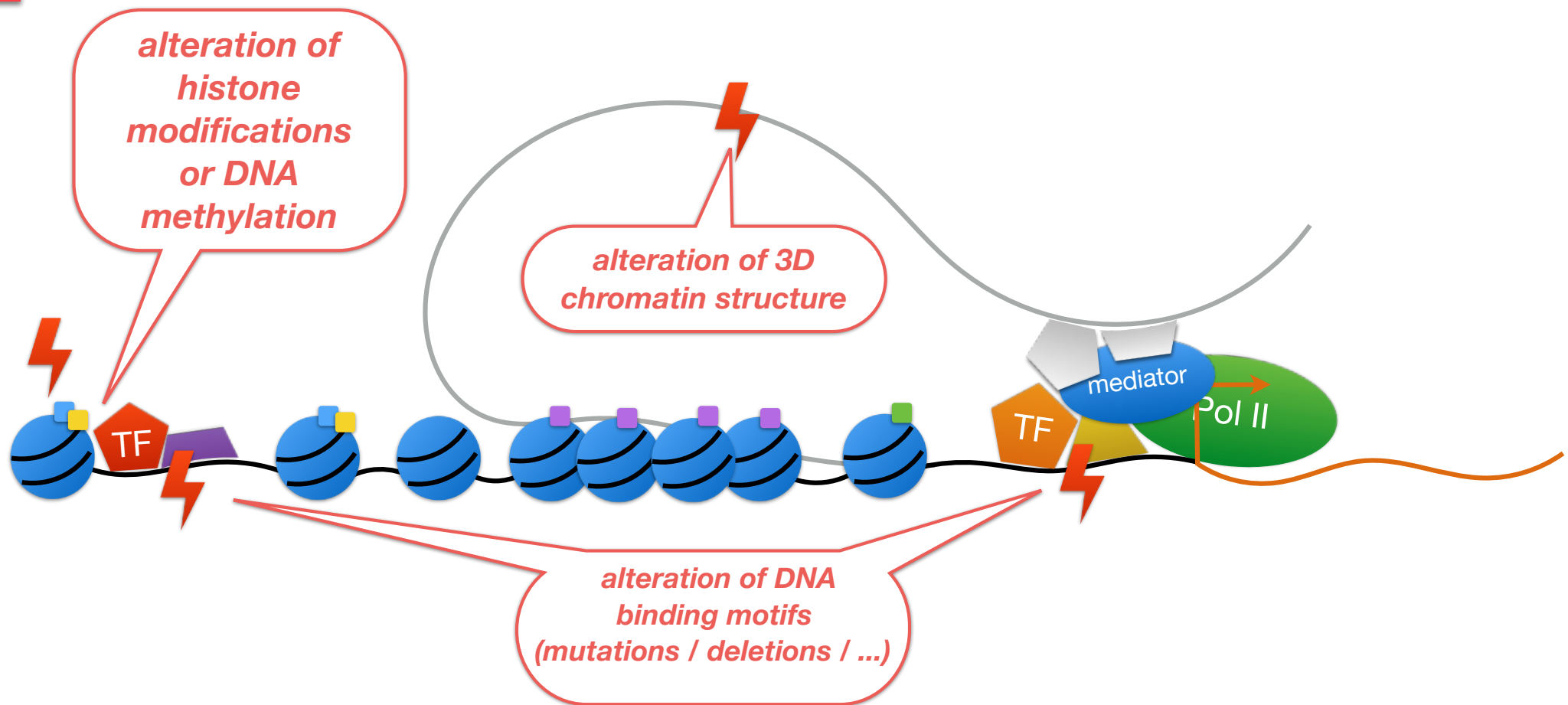
methylated Cytosine

Epigenetics: DNA methylation



- CpG methylation of gene promoters is usually associated with **gene repression**
- Methylation **prevents the binding** of most transcription factors
- Methylated cytosines can recruit methyl-binding proteins (prevent TFs from binding)

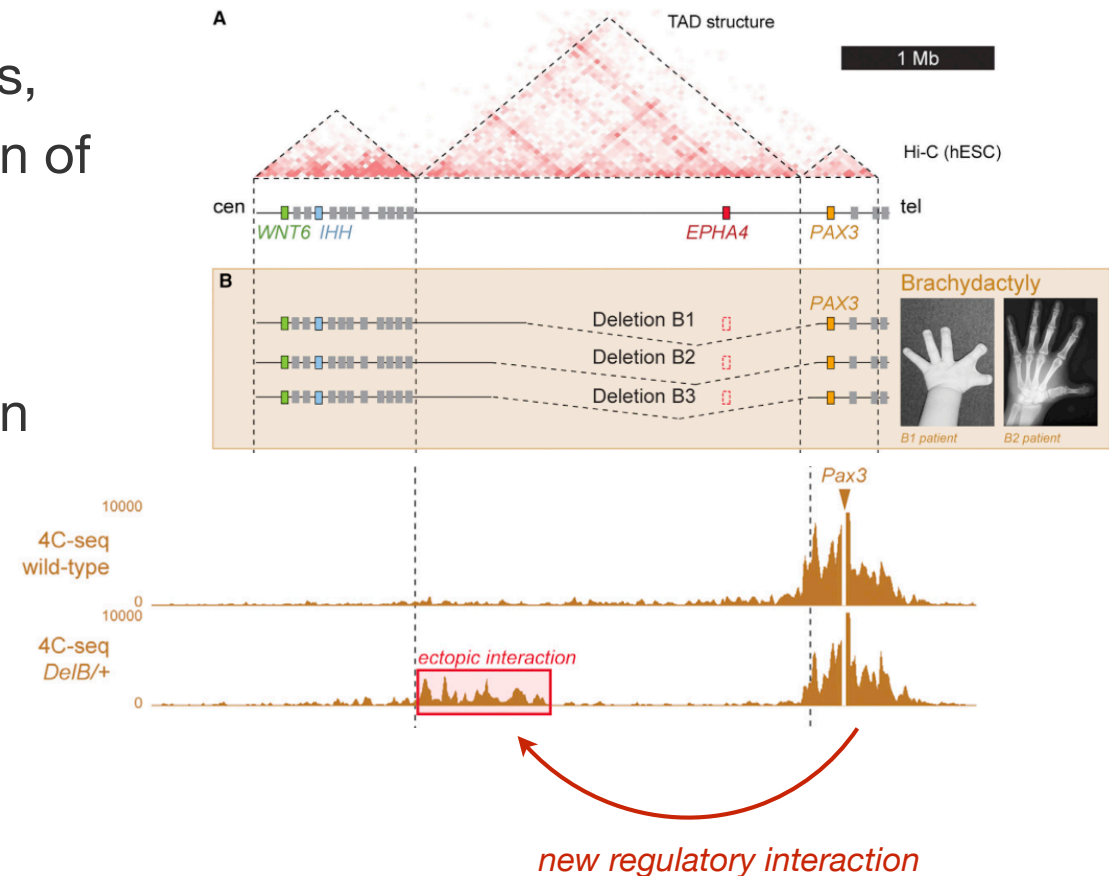
Transcriptional **d**eregulation



complex interplay of multiple components
multiple sources of potential deregulation

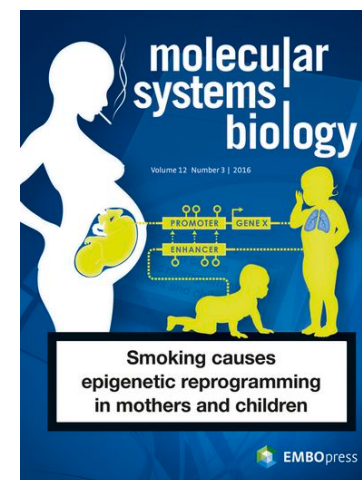
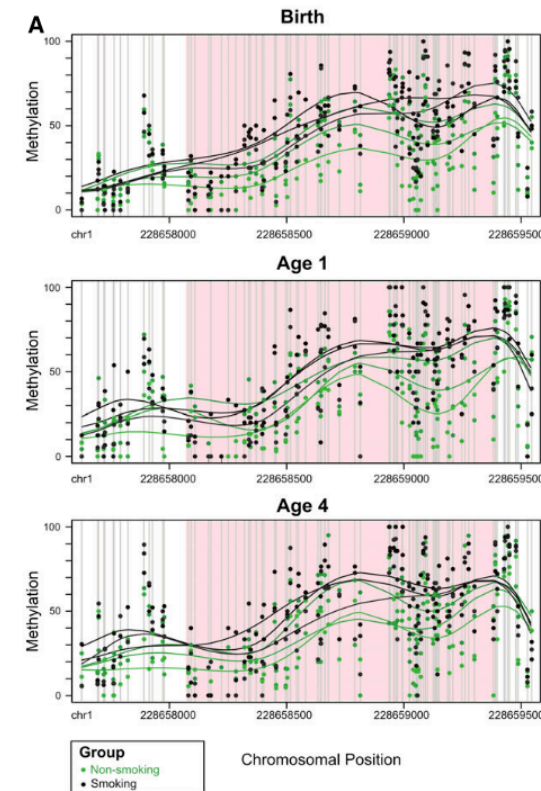
Conformational deregulation

- Chromatin conformation defines **domains**, separated by **insulators**
- Genomic alterations (deletions, inversions...) lead to disruption of 3D conformation
→ ectopic gene activation
- "Enhancer hijacking" has been described in cancer



Epigenetic deregulation

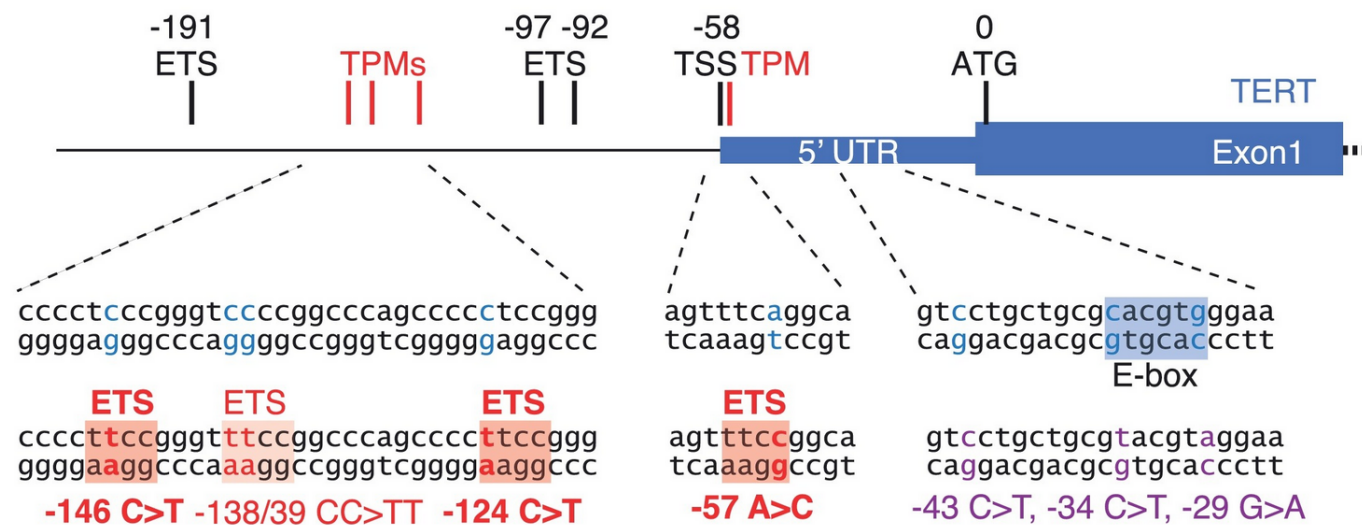
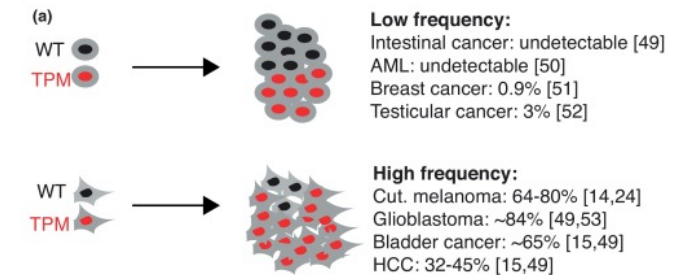
- **Epigenetic marks** (e.g. histone marks or DNA methylation) can encode external environmental cues
- Maternal smoking affects DNA methylation in children at regulatory sites (**differential methylated regions DMR**)
- These regions control **developmental genes** involved e.g. in lung development
→ higher susceptibility to lung diseases



[Bauer et al., MSB (2016)]

Sequence deregulation

- **TERT** promoter mutation
- Mutation in the promoter of the telomerase gene TERT is found in many tumor types
- leads to the creation of a new binding site and overexpression of the gene



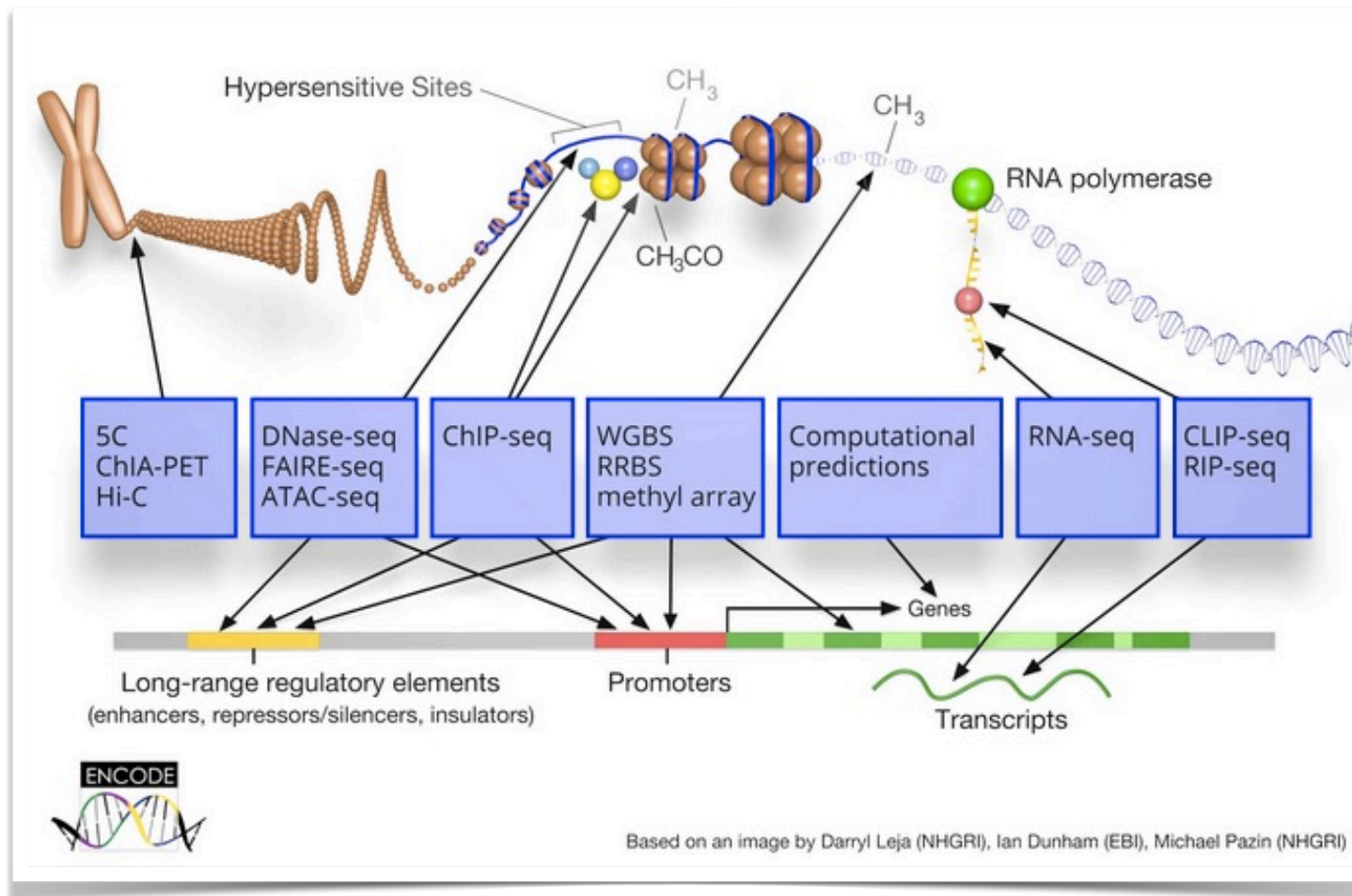
[Lorbeer, Hockemeyer, 2020]



3. Which data is available?

Exploring the genome's activity

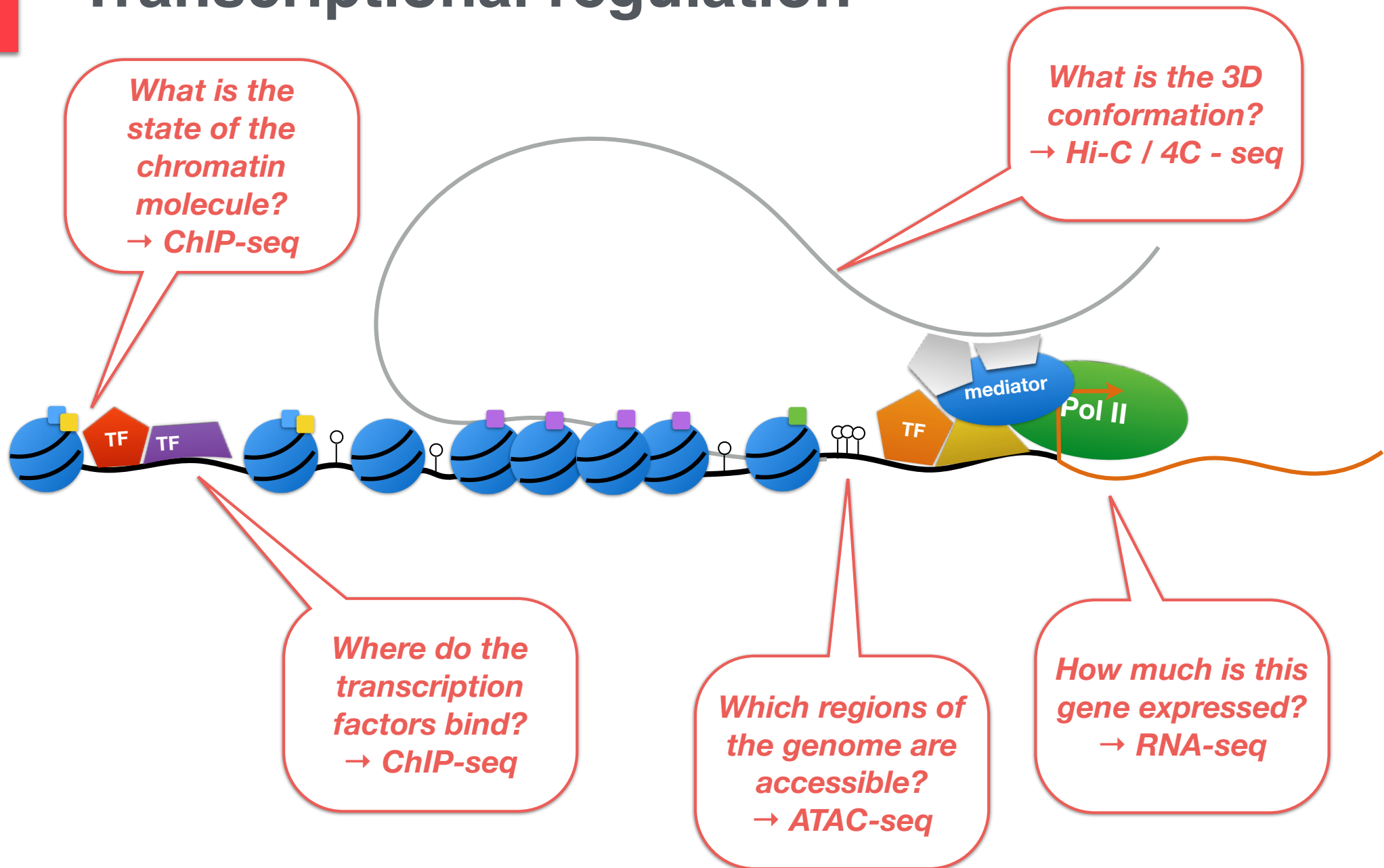
- Large scale consortia (ENCODE, Roadmap, ...) have systematically explored the **activity** of the genome using experimental assays



<https://www.encodeproject.org/matrix/?type=Experiment>

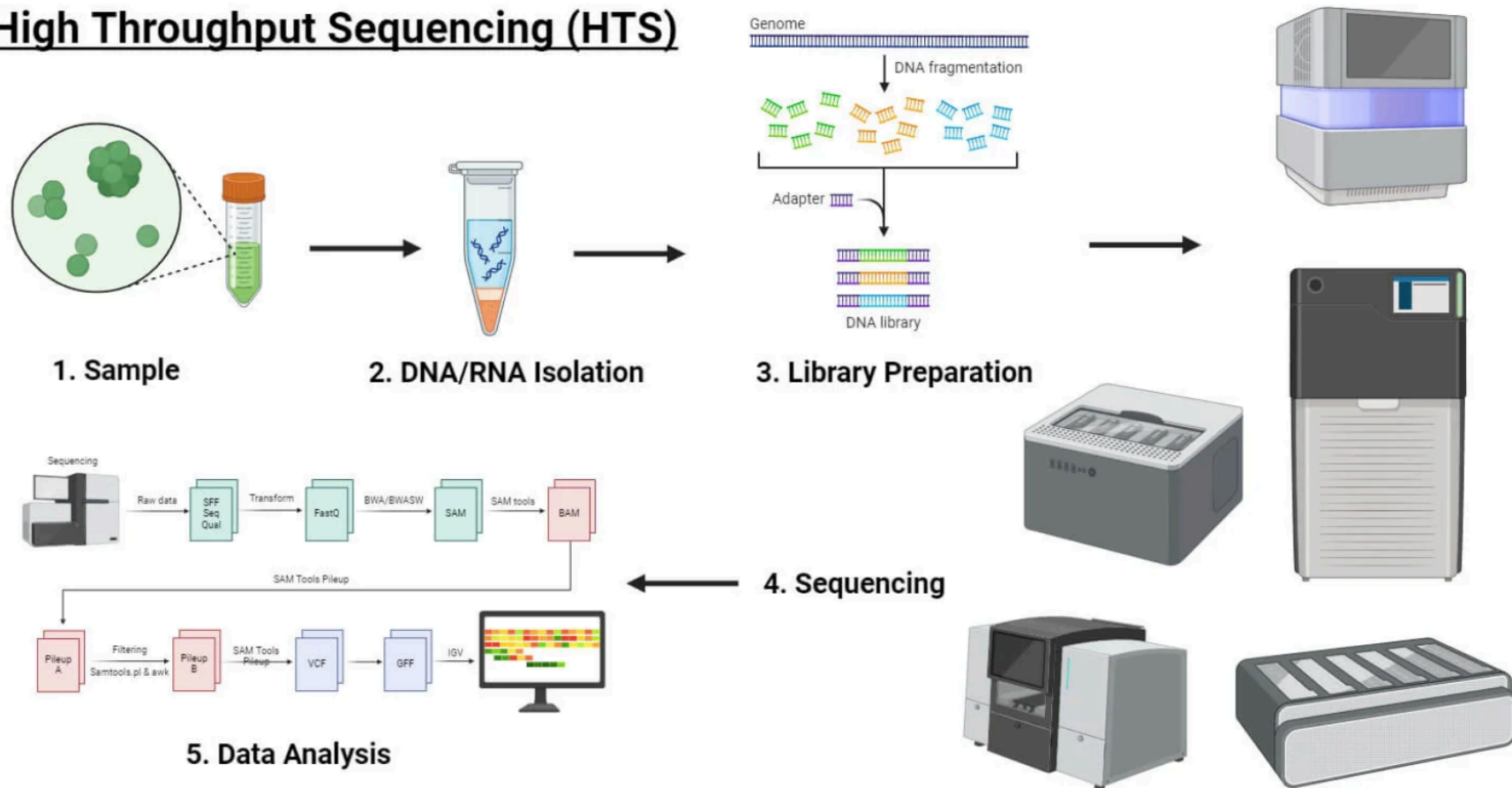
<https://www.encodeproject.org/>

Transcriptional regulation



DNA / RNA sequencing

High Throughput Sequencing (HTS)



<https://microbenotes.com/high-throughput-sequencing-hts/>

DNA/RNA sequencing



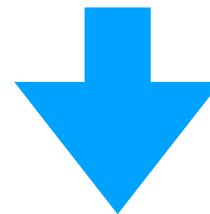
```
@SRR540192.1580 IL34_5480:5:1:6425:1038/1
CATCTTGGCCTCTGTGCAGCATTCTTTCTCCATGGT
+
IIIIIIHHIIIIIIHHIIIIIIIIIIIIHHIIHHID
@SRR540192.1752 IL34_5480:5:1:7005:1052/1
GCTCCCAGAAACCCAGGGCCACTGGCAGCTTCAGGGA
+
GGGGGGGBG@GGGGB@>D<GGGF@<?<?9??; (? :2 (
@SRR540192.1788 IL34_5480:5:1:10167:1053/1
ATGGGCTTCCTCCGGCTTTTCAGCCACCTGCGCCCTGC
+
GG@G>G@E3<B=B;B<E>EDEAAAB:B.:=>A?;A8D
@SRR540192.2271 IL34_5480:5:1:5889:1093/1
TGATCATCTGGCTGATGCGGTGACTGCCACCCTTGAG
+
IIGIIIIIIIIIIIIIDIIIGIIIIHGHHHIIIIHHID
```

identifier of the machine/run

DNA/RNA sequence

Sequencing quality per base

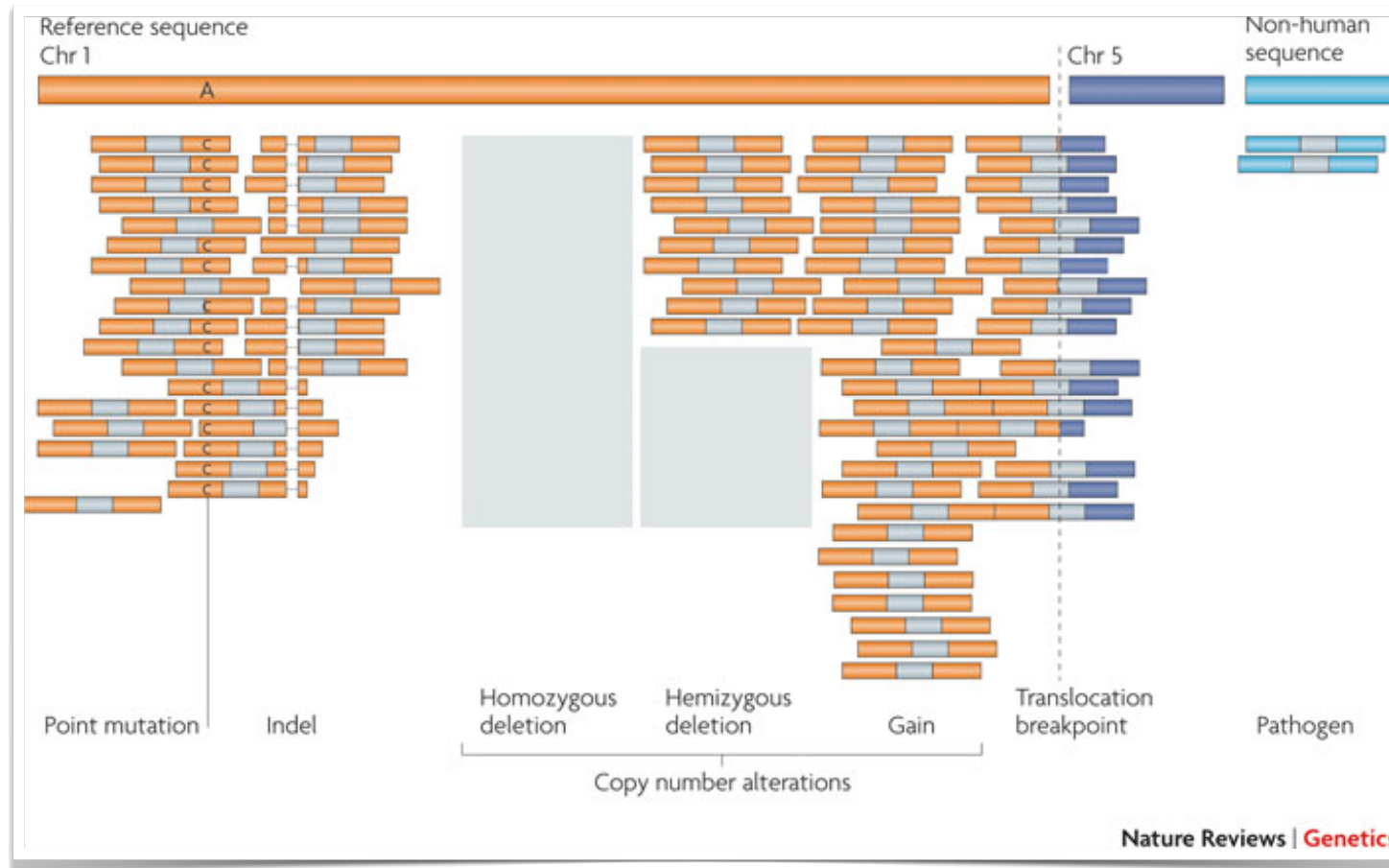
= fastq format



8 Terabase in 2 days

Book
with 3 Billion
pages

Application to cancer genomics



Do we observe mutations?
 Do we observe deletions/insertions?
 Do we observe translocations?

Transcriptomics: RNA-seq

```

TCTAGAACTGTGGGAAAAATAAATTTCTGTTGTTTAAAGCCACCCAATTTG
TGGTAATTTGTTATGGCAGCCCCAAAAAATATTGTAATAAGGGTAGTT
CAATAAATAGTGGTACAGGCCAGGTGTGGTGGCTCACACCTGTAATCTCA
GCACCTTTGGGAGGCCAGGTGGTGGGACTGCTTGAGGCAGGAGTTTGAGA
CCAGCCTGCCCAATGTGGTGAACCTCATCTCTACTAAAAATACAAAAAT
TAGCTGGGCATGGTGGCAGCTGCCTGTAGTCCCAGCTACTCTGGGAGCTG
AGGCAGGAGAATCGCTTGAACCCAGGTGGAGGTTGCAGTGAGCCAAGATT
GTGCCACTGTGCTCCAGCCTGGACGACAGTGAGACTCTGTCTCAATAAAA
TAAATAAATAAATAGTGGTACATACAGTGACTAAAGATGGAAGTAAAAA
CCTCCCATCTATTCCCTTTTTTTCGTTTTTTTTTGGAGACGAGTCTTT
GCCCTGTGTGCTCAGGCTGGAGTCCAGTGGTGAATCTCGGCTCACCTCTG
CTTCTGGGTTCCGGAGATTCTCCTGCCTCAGACTCCAAGTAGTTGGAAT
TACAAGCCCAACACACGCCAGGCTGATTTTTGTATTTTACTAGAGACGG
GGGTTTCACTATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCGGGTGA
TCCGCTCACCTTGGCCTCCAGAGTGTGGGATTACAGGCATGAGCCACC
GTGCCAGCCTATTCTCTTTATGTTTTGGTTTCTGAACATGTGAAAAAT
AATTTATTTTATGAAAAATAAGTACAATTTAAGGGCTGGGTGTGGTGGCT
CATGCCTGTAAATCCTAGCACTTTGGGAGGCGGAGGCAGGAGTATTGCTTA
AGCTCAGGGGTTTGATACAGCTGGGCAACATCGTGAACCCCATCTCT
ACAAAACTAGCCAGGCATGGTGCATGCATCTGTAGTCTCAGCTGCTCA
GGAGGCTGAACAGGTAGGATTGCTTGAGCCTGAGAGGCTGGGGTTGCAAT
GAGTGTGATCGTGCCACTGTACTCCAGCTTGGGGACAGAGTGAGACCTT
GTCTCAAAAAATAAATAAGTAAATATGGCTCACGCCGTGAATCCAGCAG
TTTGTGAAGCTGAGCCAGGCGGATCACCTAAGGTGAGGAGTTCAAGACCA
GCCTGGCCCAACATGGTGAACCCCATCTCTACTAAAAATACAAAAATTA
CTGTATGTGGTGGCTGTGCTGTAGTCCAGCTATTGAGGATGCTGAGG
CAAGAGAATTGCTTCAACTTGGGAGGTGGAGGTTGCAGTGAGCCAAGATG
GCACCATGTACTCCAGCCTGGGCAACAAGAGCGAAACTCCGTCTTGAAA
ATAAATAAATGAATAAATAAATAAAGTAAATGTGATTCTGAGCCTTGT
CTATATAGCACTAAAAGATTAACTCAAAACACATGTTGTAAGAAAACTGG
AAGAGGAAGAGAGTCCCGATCAAGTAGCATAGGAAGAAGAGTTTTATTT
TTATGAATATAAGAAAAACAGCTGGGCAACATGGTGAACCCCATCTCT
ACAAAAAAGTACAAAAGTACTTTTTTTCAGACGTGGTGGCACATACCTGT
AGTCCCAGCTACTCCAGAGGCTAAGGTGGGAGGATCTCTTGAGCCTGGGA
GAGGAGGCTGCAGTGAGCTGTAATTGTGCCACTGCCTCCAGCCTGGGC
AACACAGCAAGACCTTTCTCAAAAAAAAAAAAAAAAAAAAAAAGAG
AAAAAAGATAAGCAAGCATACCCTAGAAGGCTGGGTACTTAGCATACA
TTCTGTTGAATCAAGATGCATGAGAGTCTTAATGATAGTTTACTCATACT
GTACTTAATATGCAAAATTAGAATAATCAAGTAAGAACAATGATTAGATG
CCATTTTTAAATCTAAAATGAATGCCATGTGGGATGAGAGTGCTGGCC
TTGAACTCAGAAAACTCGGTTCAAGTCTGGGTTTTTCCCATTTCTAGGT
TGTGATCACTTCCCTCCCTGAGTCTTAATTAATAATCCAGTTTAGGCC
GGGCGTGGTGGCTCACGCCGTGAATACCAGCACTTTGGAGGGCCGAGGCG
GGTGGATCATGAGGTGAGGAGATCGAAACCATCCTGGCTAACACAGTGAA
ACCTGTCTCTACTAAAAATACAAAAATTAGCCGGCGTGGTGGTGAG
CACTCCCAGTAGTCCAGCTACTCGGGAGGCTGAGGCAAGAGAATGGTG
TGAACCCGGGAGGCGGAGCTTGCACTCAGCCGAGATCACGCCACTGTA
CCAGCCTGGGCGACAGAGTGAGACTCCGTCTCAAAAAATAAATAAATAA
AATATCCAGTTTATAAATAGGGATAACAATATCTGTCTGGCCTGATAACA
AGGGCTTGTAAATGCAAAATGCAATAAGGAGAAAGCAATTCATGGACTTT
GACTCTCAGTACCAGGAAAAGTCTTGCGTGAAGGAAGAACATAGAGG
AAGTCATTAGTGATAAGGACCACAGGTGCAGTTGCAATTCCTGCCCTTT
TGCCAAAAGAAAAAGGAGAGACCTGTGCCAAAATAAGCAGCTTCTAACTCA
AGTCATACCAAGTCAGTCCCTTGCACTGCACATTGCTGTGCTGGC
ACTGCTGCTGGTCTTGGCAGTAATAGGTTGGGTATCCTAATCTACTGCCG
GAACCTTTCAGTGGTTTCCCTCTTCAAGTTTATCTAGTTAAATGTTGCGGATA
ATCCTTCCCTCGGCCCTCCCAACAAGTGAAGAACTCCATCTCAAA
ATGGCTTCAACCAATGCTGAGGTGATTATCTCGCAGAAGAAGAAATCCAG
AGAAAGTGTGGCTTCGTGTTAACAAATACATGGACTGGAAGATGATGTCT

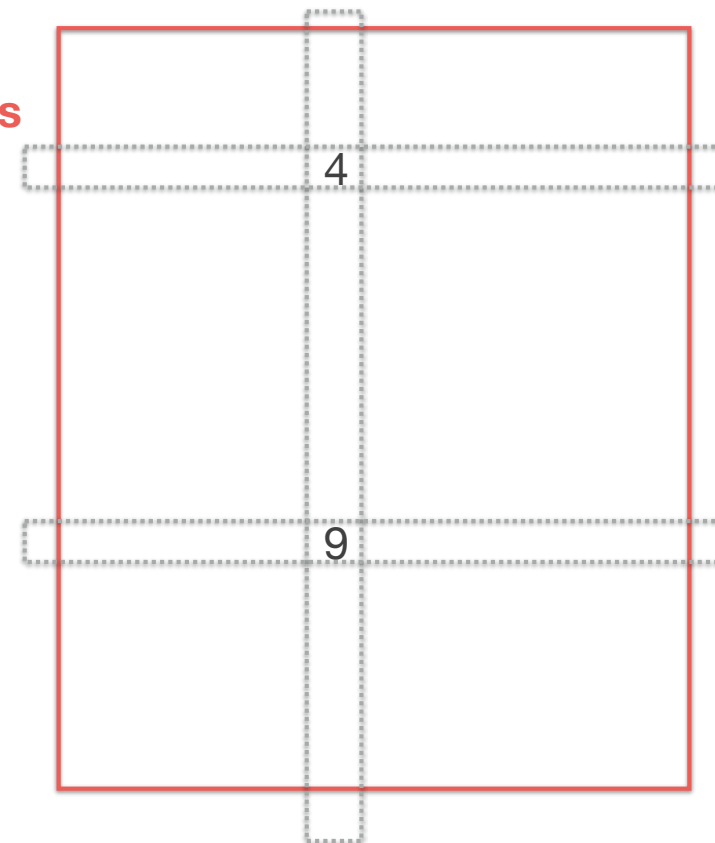
```



mRNA molecules



Patient 1432

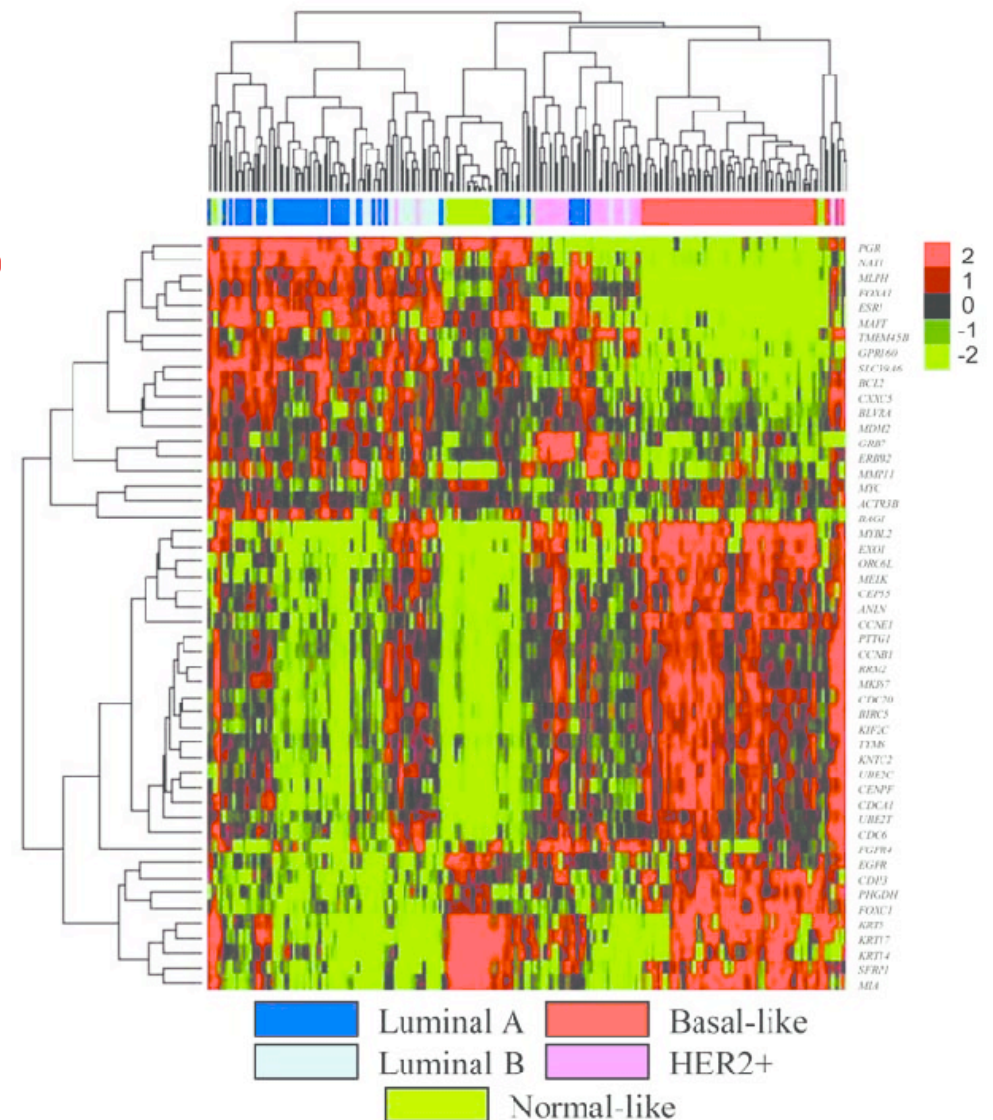


Transcriptomics: RNA-seq

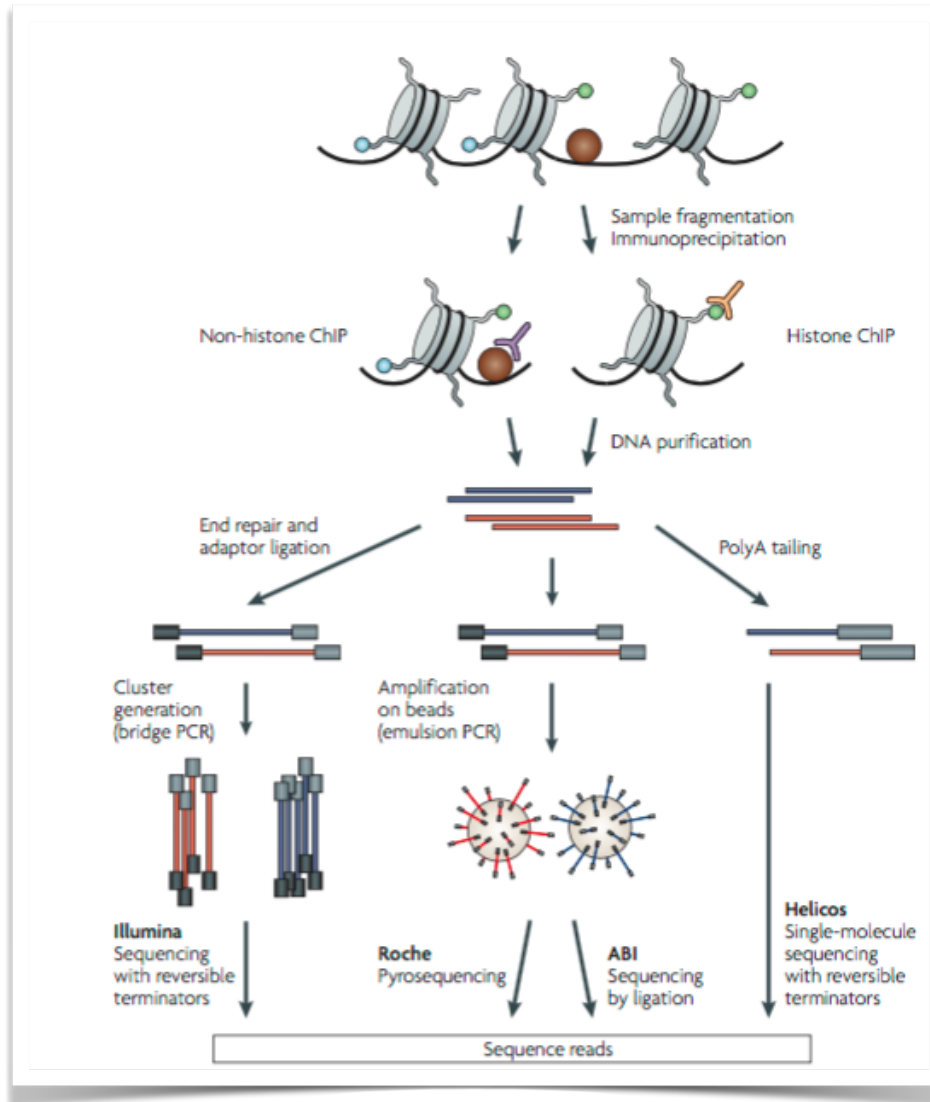
TCTAGAACTGTGGGAAAAATAAATTTCTGTTGTTTAAAGCCACCCAATTTG
 TGGTAATTTGTTATGGCAGCCCCAAAAAATATTGTAATAAGGTTAGTT
 CAATAAATAGTGGTACAGGCCAGGTGTGGTGGCTCACACCTGTAATCTCA
 GCACCTTTGGGAGGCCGAGGTGGTGGAGTGCCTTGAGGCAGGAGTTGAGA
 CCAGCCTGCCCAATGTGGTGAACCTCATCTCTACTAAAAATACAAAAAT
 TAGCTGGGCATGGTGGCAGCTGCCTGTAGTCCCAGCTACTCTGGGAGCTG
 AGGCAGGAGAATCGCTTGAACCCAGGTGGAGGTTGCAGTGAGCCAAGATT
 GTGCCACTGTGCTCCAGCCTGGACGACAGTGAGACTCTGTCTCAAAATAAA
 TAAATAAATAATAGTGGTACATACAGTGACTAAAGATGGAAGTAAAA
 CCTCCCATCTATTCCCTTTTTTTCGTTTTTTTTTGGAGACGAGTCTTT
 GCCCTGTTGCTCAGGCTGGAGTCCAGTGGTGAATCTCGGCTCACCTCTG
 CTTCTGGGTTCCGGAGATTCTCCTGCCTCAGACTCCAAGTAGTTGAAT
 TACAAGCCCCACACACGCCAGGCTGATTTTTGTATTTTACTAGAGACGG
 GGGTTTCACTATGTTGGCCAGGCTGGTCTCAAACTCTGACCTCGGGTGA
 TCCGCTCACCTTGGCCTCCAGAGTGTGGGATTACAGGCATGAGCCACC
 GTGCCAGCCTATTCTCTTTATGTTTTGGTTTCTGAATATGTGAAAAAT
 AATTTATTTTATGAAAAATAAGTACAATTTAAGGGCTGGGTGTGGTGGCT
 CATGCCTGTAATCCTAGCACCTTTGGAGGCGGAGGCAGGAGTATTGCTTA
 AGCTCAGGGGTTTGATACAGCTGGGCAACATCGTGAACCCCATCTCT
 AAAAAAATAGCCAGGCTGGTGGCATGCATCTGTAGTCTCAGCTGCTCA
 GGAGGCTGAACAGGTAGGATTGCTTGAGCCTGAGAGGCTGGGGTTGCAAT
 GAGTGTGATCGTGCCACTGACTCCAGCTTGGGGACAGAGTGAGACCTT
 GTCTCAAAAAATAAATAAGTAAATATGGCTCACGCCCTGTAATCCCAGCAG
 TTTGTGAAGCTGAGCCAGGCGGATCACCTAAGGTCAGGAGTTCAAGACCA
 GCCTGGCCCAACATGGTGAACCCCATCTCTACTAAAAATACAAAAATTAG
 CCTGATGTGGTGGCTGTGCTGTAGTCCCAGCTATTGAGGATGCTGAGG
 CAAGAGAATTGCTTCAACTTGGGAGGTGGAGGTTGCAGTGAGCCAAGATG
 GCACCATTTGACTCCAGCCTGGGCAACAGAGCGAAACTCCGTCTTGAAA
 ATAATAAATGAATAAATAAATAAAGTAAATGTGATTCTGAGCCTTGTG
 CTATATAGCACTAAAAAGATTAACTCAAGGCACATGTTGTAAGAAAACTGG
 AAGAGGAAGAGAGTCCCGATCAAGTAGCATAGGAAGAAGAGTTTTTATTT
 TTATGAATATAAGAAAAACAGCCTGGGCAACATGGTGAACCCCATCTCT
 ACAAAAAAGTACAAAAAGTACTTTTTTTCAGACGTGGTGGCACATACCTGT
 AGTCCCAGCTACTCCAGAGGCTAAGGTGGGAGGATCTCTTGAGCCTGGGA
 GAGGAGGCTGCAGTGAGCTGTAATTGTGCCACTGCACTCCAGCCTGGGC
 AACACAGCAAGACCTTTCTCAAAAAAAGAGAGAGAGAGAGAGAGAGAG
 AAAATAAGATAAGCAAGCATACCCTAGAAGGCTGGGTACTTAGCATACA
 TTCTGTTGAATCAAGATGCATGAGAGTCTTAATGATAGTTTACTCATACT
 GTACTTAATATGCAAAATTAGAATAATCAAGTAAGAACAATGATTAGATG
 CCATTTTTAAATCTAAAATGAATGCCATGTGGGGATGAGAGTCTGGCC
 TTGAACCTCAGAAAACTCGGTTCAAGTCTGGGTTTTTCCCATTTCTAGGT
 TGTGATCACTTCCCTCCCTGAGTCTTAATTAATAATCCAGTTTAGGCC
 GGGCGTGGTGGCTCACGCCGTGAATACCAGCACTTTGGAGGGCGAGGGCG
 GGTGGATCATGAGGTGAGGAGATCGAAACCATCTGGCTAACACAGTGAA
 ACCCTGTCTCTACTAAAAATACAAAAATTAGCCGGGCGTGGTGGTGAG
 CACTCCCAGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATGGTG
 TGAACCCGGGAGGCGGAGCTTGCAAGTCAGCCGAGATCACGCCACTGTACT
 CCAGCCTGGGCGCAGAGTGAGACTCCGTCTCAAAAAATAAATAAATAA
 AATATCCAGTTTATAAATAGGGATAACAATATCTGTCTGGCCTGATAACA
 AGGGCTTGTAGTACAAAATGCAATAAGGAGAAAGCAATTCATGGACTTT
 GACTCTCAGTACCAGGAAAAGTCTTGCGTGAAGGAAGAACATAGAGG
 AAGTCATTAGTGATAAGGACCACAGGTGCAGTTGCAATTCCTGCCCTTTT
 TGGCAAGAAAAAGGAGAGACCTGTGCCAAAAAAGCAGCTTCTAACTCA
 AGTCATACCAAGTCAGTCCCTTGCACTGCACATTGCTGTGCTGGC
 ACTGCTCGGCTTGGCAGTAATAGTTGGGTATCCTAATCTACTGCCCG
 GAACCTTTCAGTGGTTTTCTCTTTCAGTTTATCTAGTTAAATGTTGGGATA
 ATCCTTCCCCTCGGCCCCCTCCCCACAAGTGAAGAACTCCATCTCAAA
 ATGGCTTACCAATGCTGAGGTGATTATCTCGCAGAAGAAGAAATCCAG
 AGAAAGTGTGGCTTCGTGTTAACAAATACATGGACTGGAAGATGATGTCT



mRNA molecule



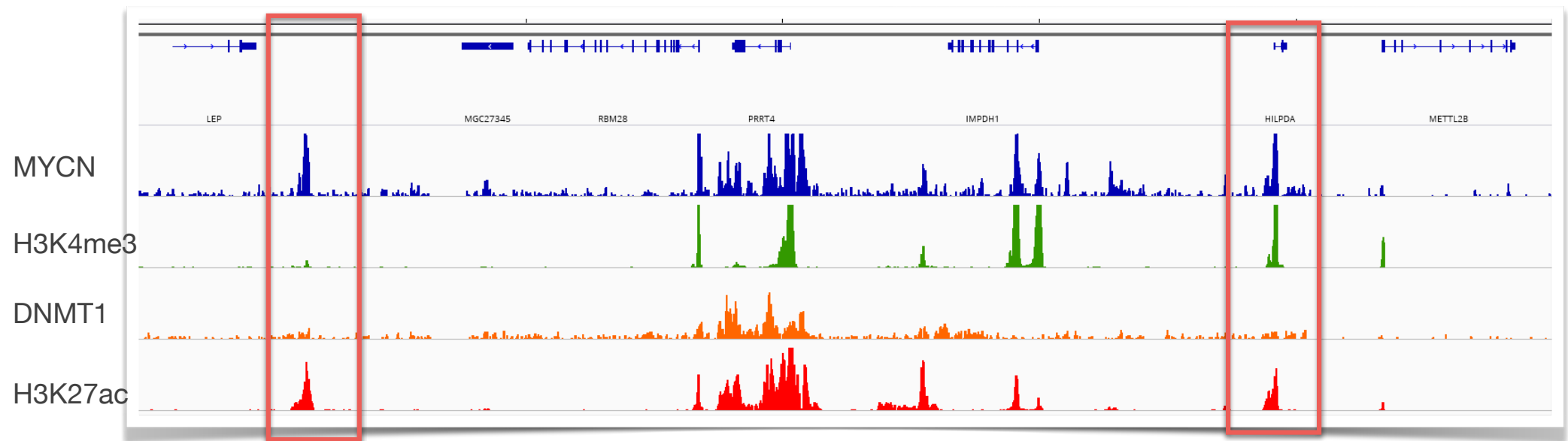
Chromatin Immunoprecipitations



[Park, Nat.Rev. 2009]

- Goal: identify DNA sequences on which **transcription factor** bind or **histones** are modified
- Chromatin immunoprecipitation (ChIP): use **specific antibodies** to pull out DNA fragments
- High-throughput sequencing
- Data analysis: where are the enriched regions?

Example of ChIP-seq signal



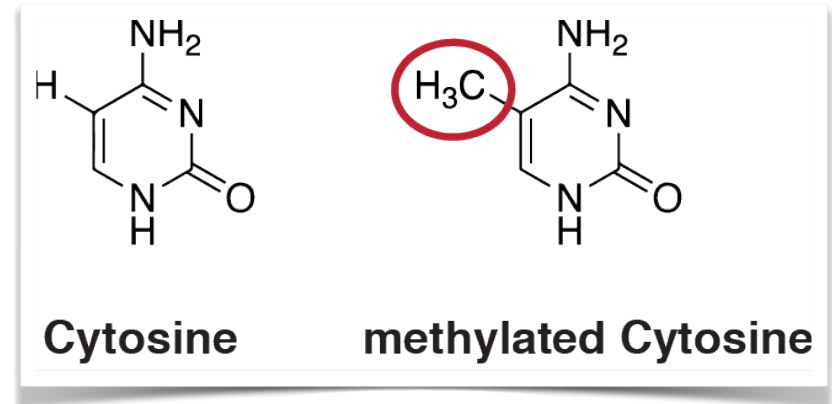
**distal
MYCN peak**
(away from gene
promoters)

MYCN peak
and H3K4me3
enrichment
at promoter

Epigenetics: DNA methylation

- DNA methylation is revealed by using **bisulfite conversion** (HSO_3^-):

- ◉ unmethylated cytosines are converted
 $\text{C} \rightarrow \text{U} \rightarrow \text{T}$
- ◉ methylated cytosines are protected
 $\text{mC} \rightarrow \text{mC}$



- unmethylated CpG are identified by the presence of a **mismatch**
TpG
- 2 approaches:
 - ◉ array based: hybridization to CpG probes on array
 - ◉ sequencing: whole genome bisulfite-sequencing

Measuring DNA methylation

- **Array based methods**
- CpG containing probes on array
 - 27K probes
 - 450K probes
 - 800K (EPIC)
- all probes contain a methylated (C) and unmethylated (T) version
- Cheap but sparse
- **Sequencing base methods**
(whole-genome bisulfite sequencing WGBS)
 - unmethylated C → T
 - methylated C → C
- Shearing, conversion and sequencing (Illumina X-10)
- Information about the 28 million CpGs

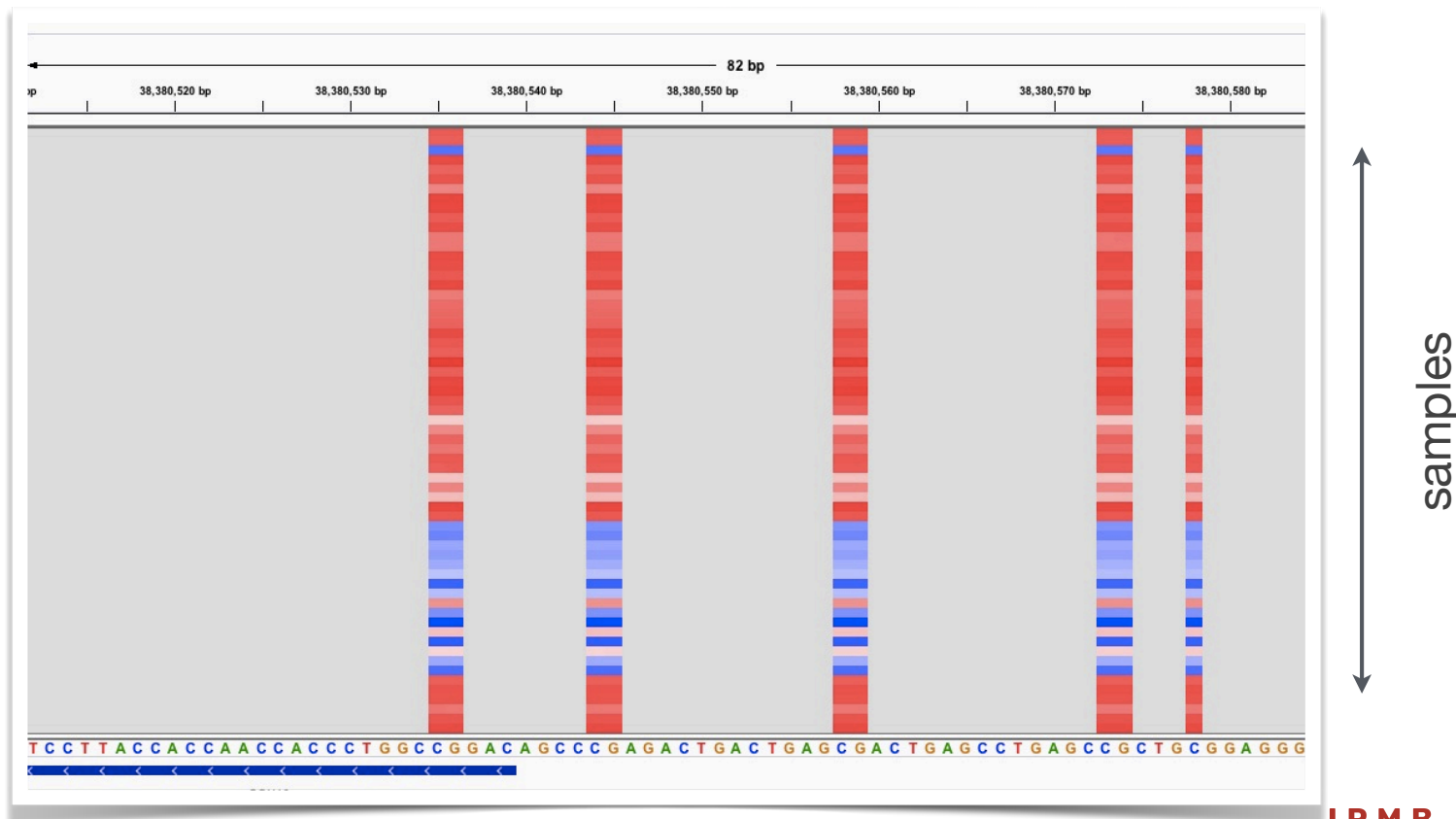


```

--ATGTTTCGTAGATTGTACTGTTGAACGTTATGTTTAATAGATGCGTTGCGAAT--
ATGTTCTAGATTGTA TGTGAA GTTATGTTAA
GTTCTAGATTGTA TTGAA GTTATGTTAATA
TCTAGATTGTA TGT GAA GTTATGTTAATAG
CTAGATTGTA TGT AA GTTATGTTAATAGAT
AGATTGTA TGTGAA GTTATGTTAATAGATG
GATTGTA TGTGAA CG AATAGATCGTTGCGA
ATTGTA TGTGAA CGT TAGATCGTTGCGAAT
TGTATGTTGAACGTTAT
ATTGTTGAACGTTATGTT
  
```

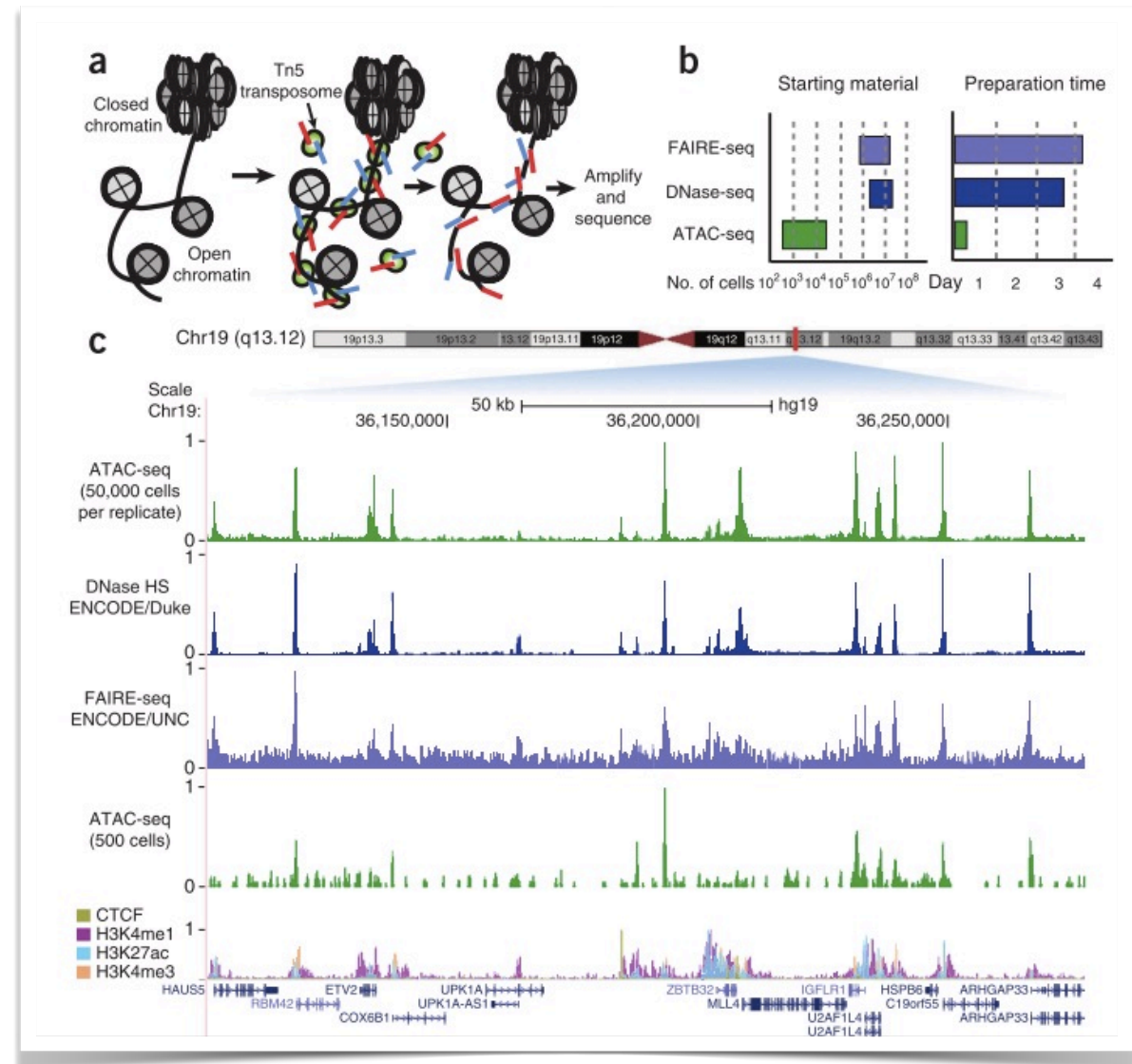
Example DNA methylation

- Whole genome bisulfite sequencing provide information about all CpGs in the genome
- Vertical bars = CpG positions; red = high methylation (100%); blue = no methylation (~10%)



Chromatin accessibility

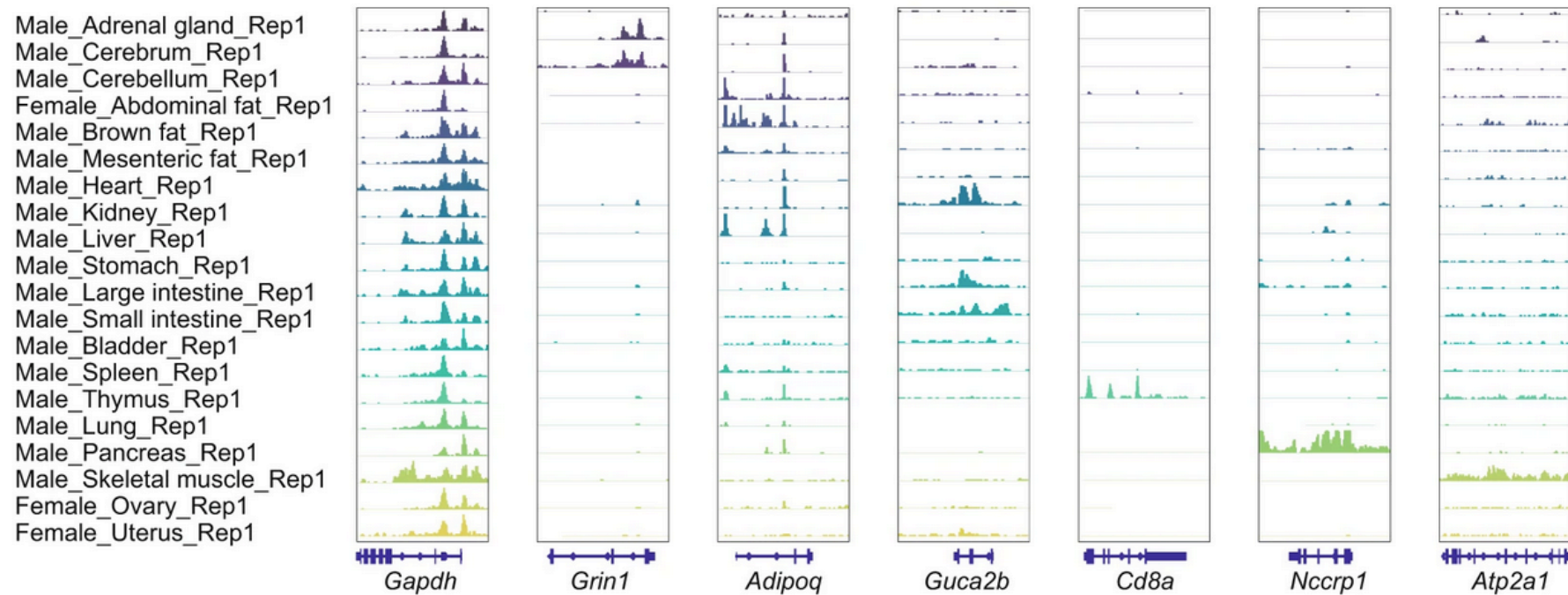
- **ATAC-seq:** using Tn5 transposase prepared with sequencing primers
- requires a small number of input material (~10,000 cells)
- easily adapter to single-cell sequencing
- identification of open chromatin regions (peaks)



[Greenleaf (2013)]

Accessibility atlas

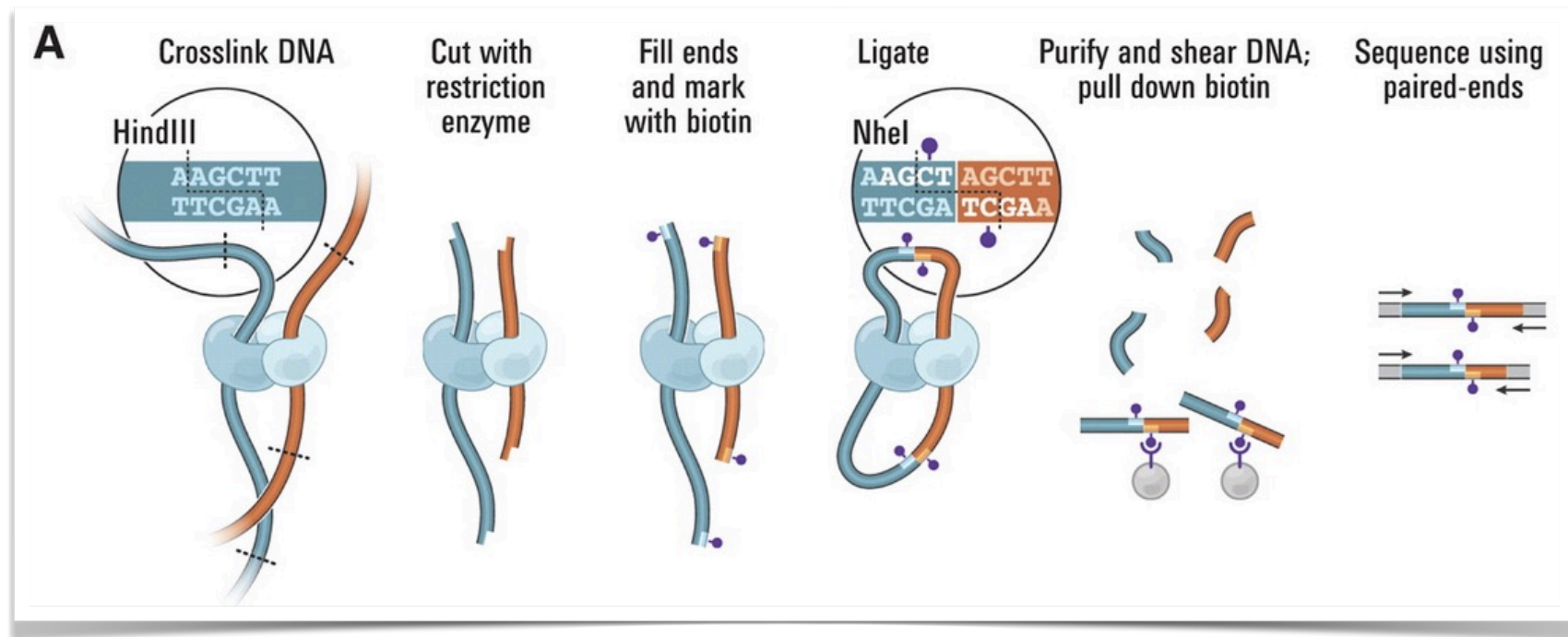
- Patterns of chromatin accessibility are **cell-type specific**



[Liu et al., Scientific Data (2019)]

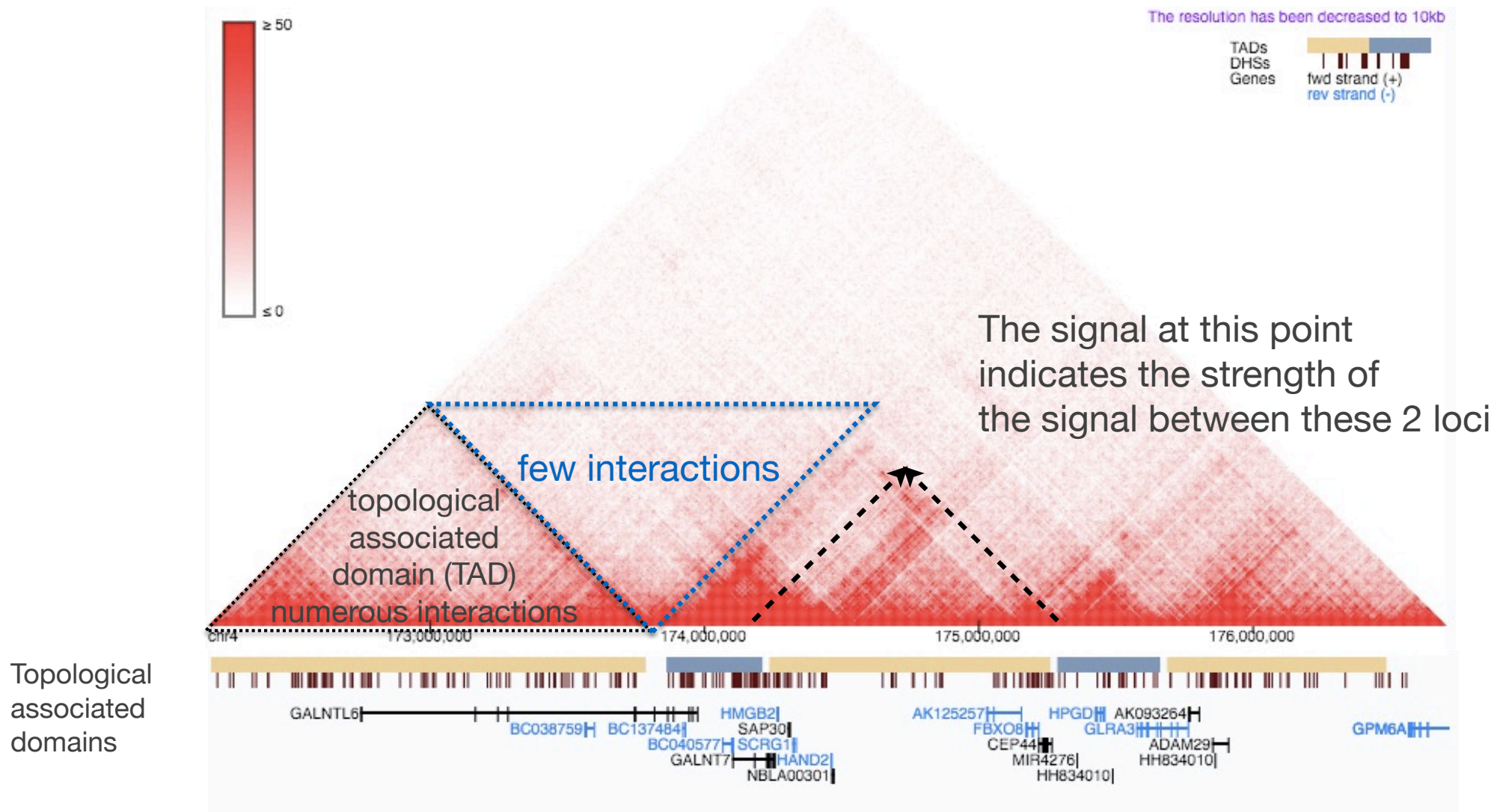
Mapping chromatin interactions

- DNA looping allows interactions between distal DNA loci
- Identification of interacting regions through “**chromatin conformation capture**” methods (3C / 4C / Hi-C)



[Liebermann-Aiden, 2009]

Hi-C and topological domains



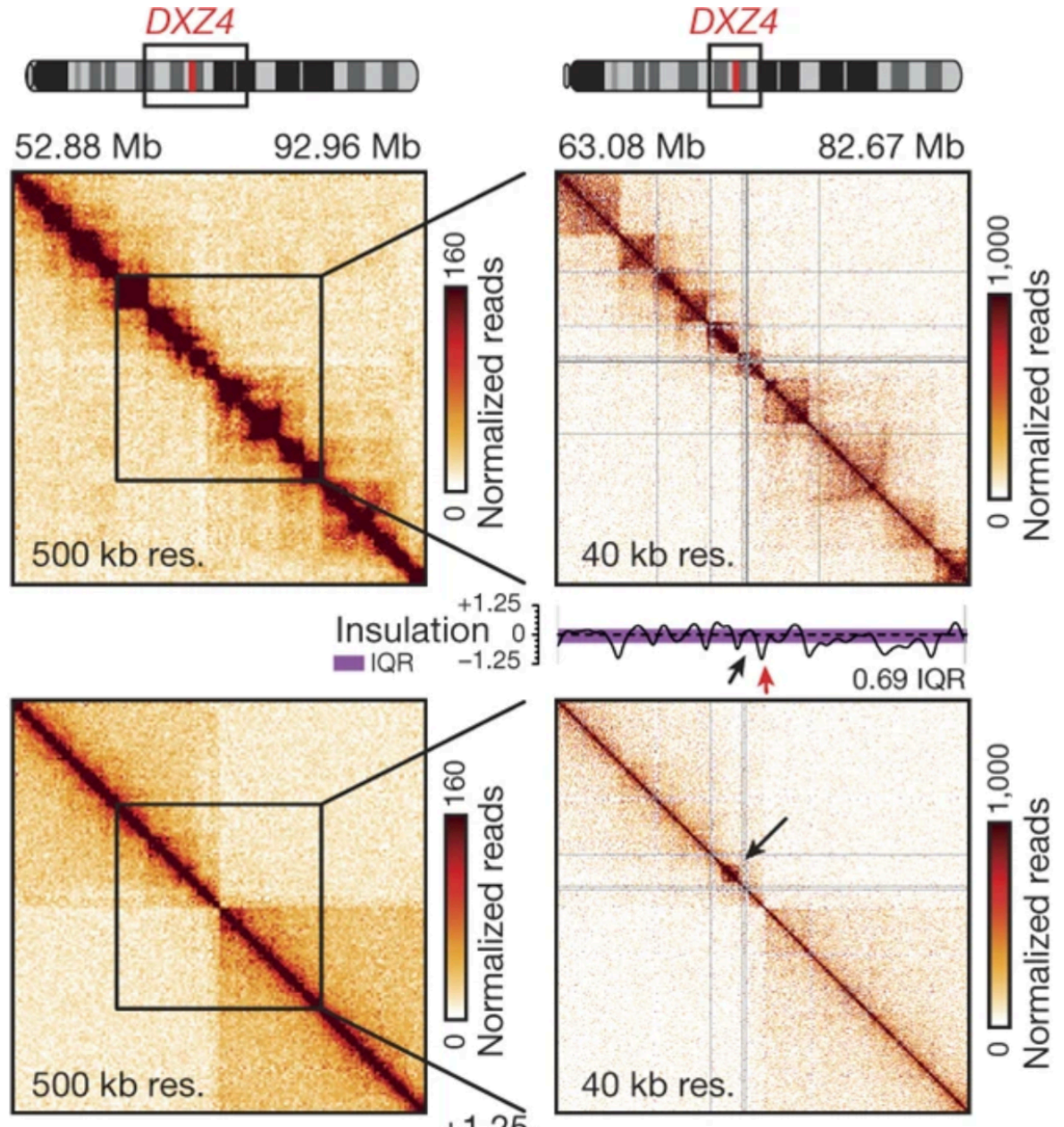
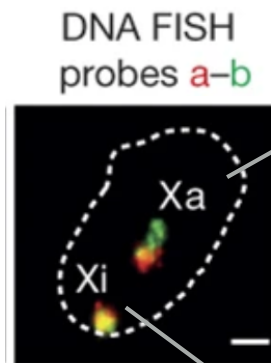
[<http://promoter.bx.psu.edu/hi-c/view.php>]

[Dixon (2012,2015)]

Chromatin organization and cell state

Allele specific Hi-C
in neural progenitor cells

Active/inactive X allele



[Georgetti et al., Nature 2016]

Chromatin organization and differentiation

- Changes in chromatin conformation occur during cell differentiation (e.g. neural development)

