SevenBridges

Epigenetic control

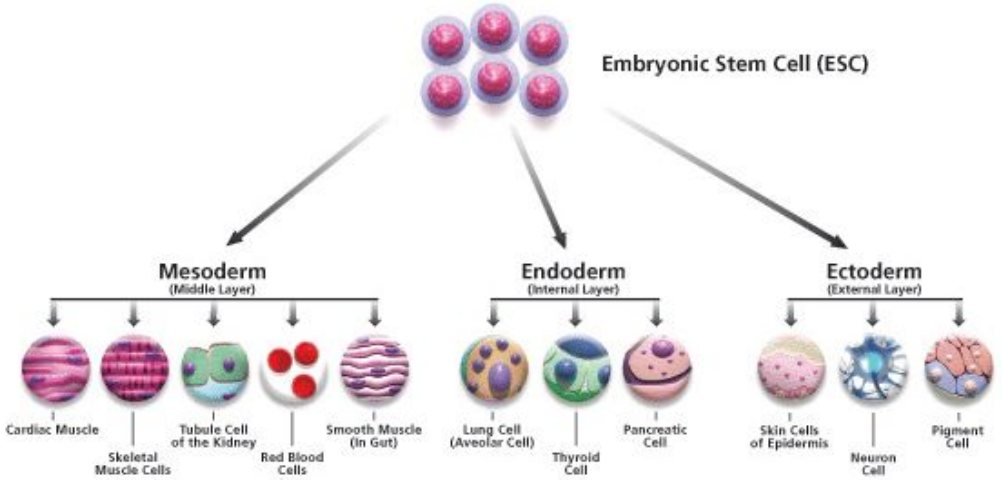
May 2023

Boris Majić, presenter

boris.majic@sbgenomics.com

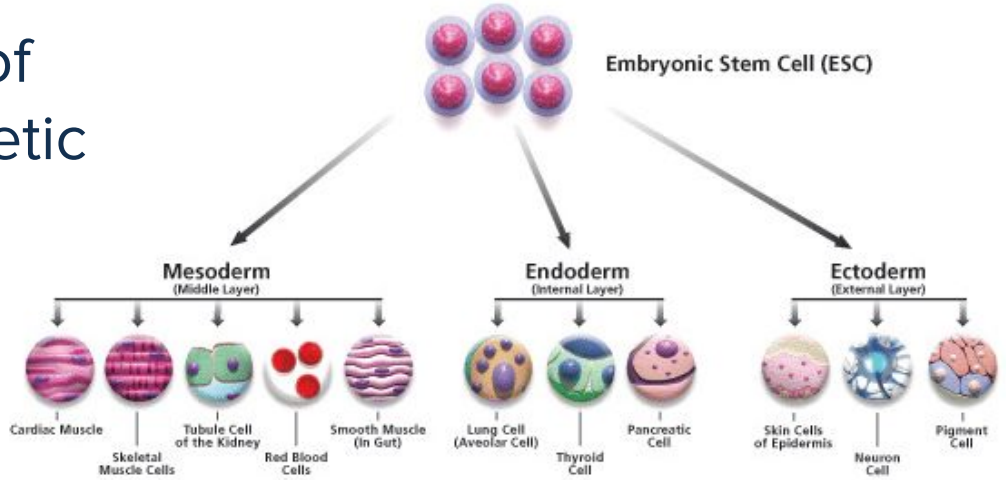
Epigenetic control

- Same genotype - different phenotypes
- How?



Epigenetic control

- **Epigenetics** - extra layer of information on top of genetic information
- Each cell type is defined by genes that are **expressed** in the cell

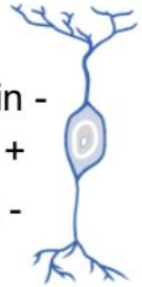


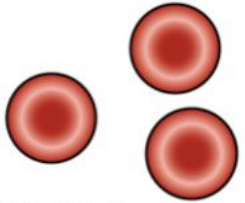
Gene expression

DNA → RNA → protein

Single cell fertilised egg
1 cell type

Neuron
Haemoglobin -
Dopamine +
Myoglobin -





Red blood cells
Haemoglobin +
Dopamine -
Myoglobin -

Muscle cells
Haemoglobin -
Dopamine -
Myoglobin +

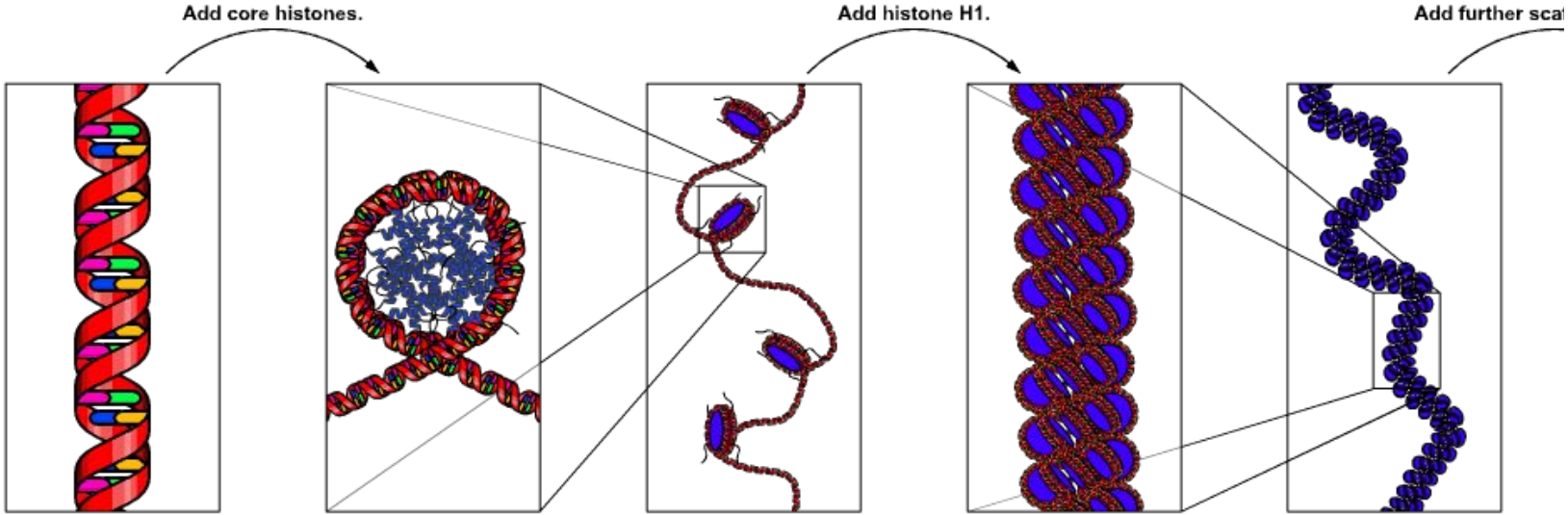
Each cell type only expresses a restricted subset of genes.

Epigenetic modifications

- Demarcate the start and end of genes
- Provide structure to the chromosome
- Alter how we read each and every gene
  - genes being expressed (active)
  - genes being not expressed (silent)

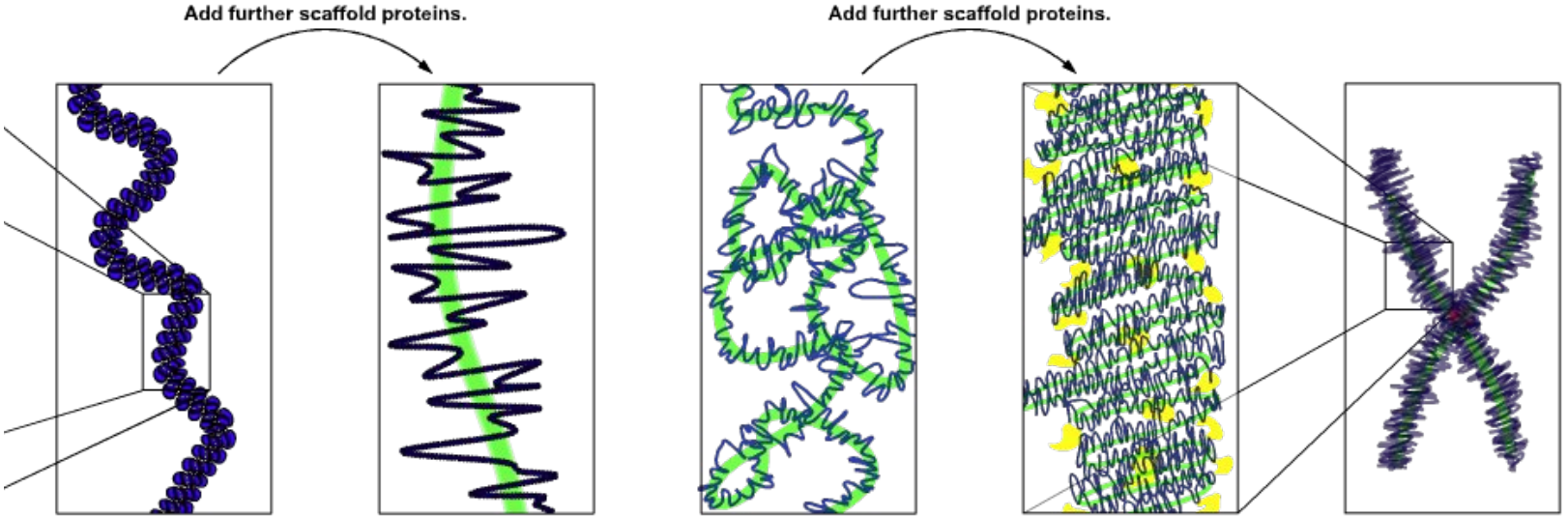
DNA structure refresher

- DNA is compacted into chromatin
- DNA is wrapped around histone proteins



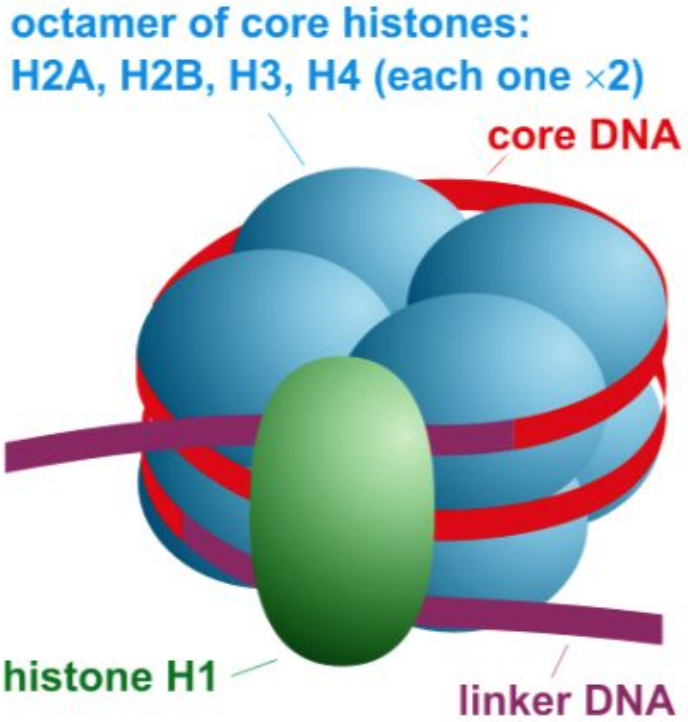
DNA structure refresher

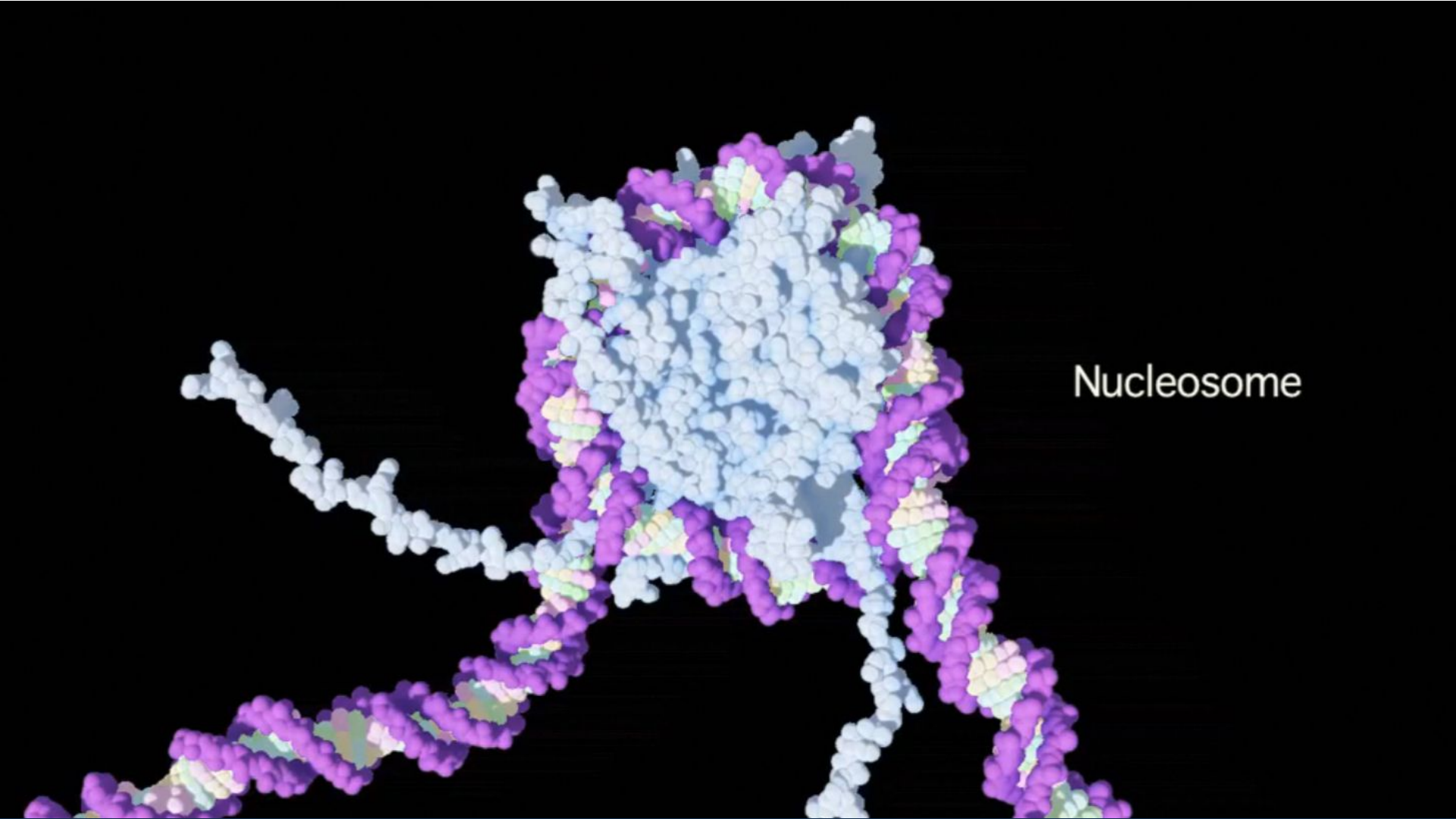
- DNA is compacted into chromatin
- DNA is wrapped around histone proteins



Nucleosome

- 146bp of DNA wrapped around a histone octamer
- Positively charged histones bind to negatively charged DNA




Nucleosome

Heterochromatin versus euchromatin

- Heterochromatin - closed chromatin
  - **Facultative** - can differ by cell type or time
    - tissue specific genes
    - parts of X chromosome
  - **Constitutive** - same in all cell types - structural role
    - Centromeres
    - Telomeres
    - Parts of sex chromosomes
- Euchromatin - open chromatin
  - DNA accessible for transcription

Specific epigenetic modifications

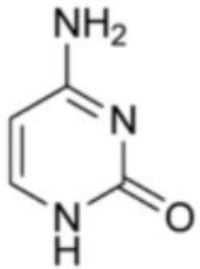
- **DNA methylation**
- **Histone modifications**
- Chromatin remodelling
- Histone variation
- Noncoding RNAs

DNA methylation

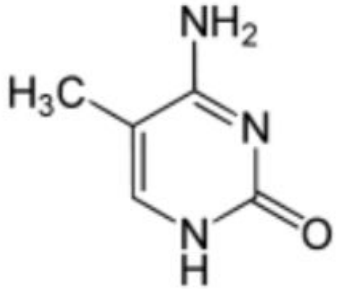
Changes to DNA amino acids

DNA methylation

- Methylation of cytosine nucleotide
- Almost exclusively on CpG dinucleotides → --------CG---------



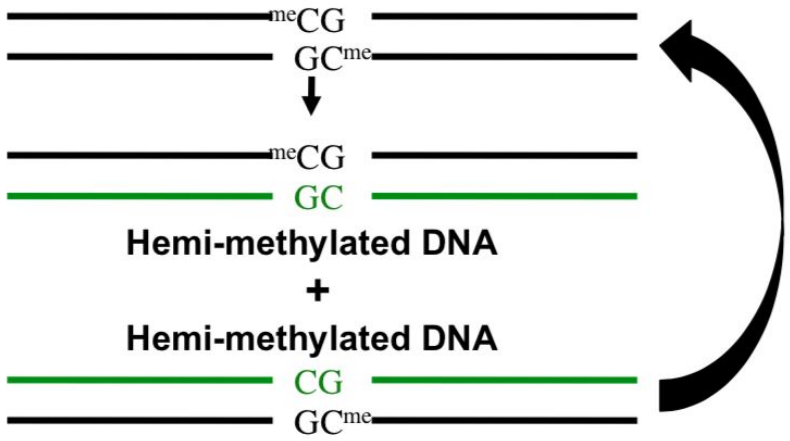
Cytosine



5- methyl cytosine

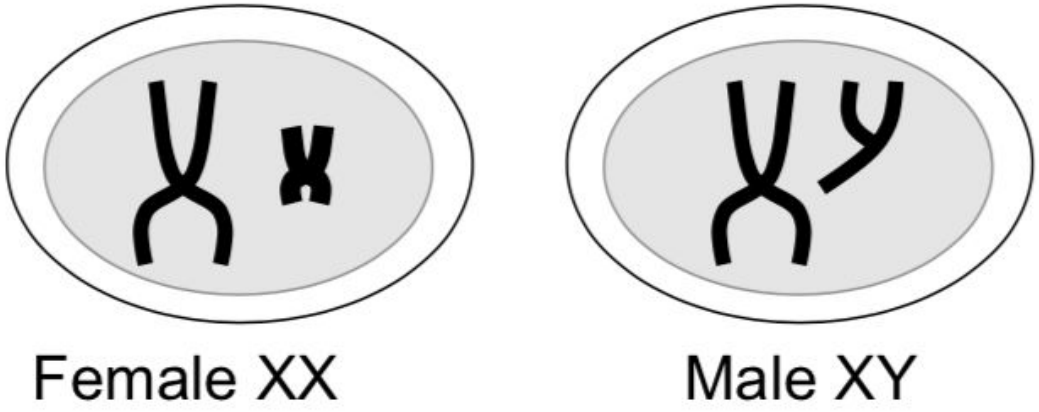
CpG site

- CpG - just a G following a C in sequence
- **CpG islands**
  - Why are they rare?
- CpG islands at gene promoters
  - Methylated ➜ gene expression silenced



X inactivation

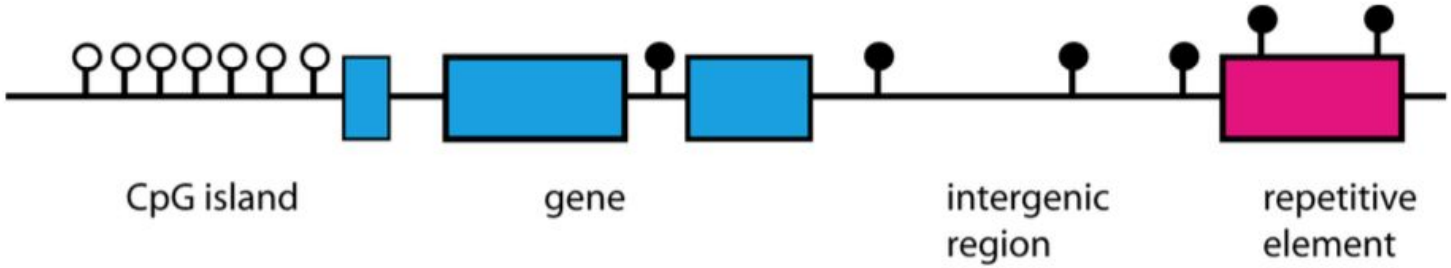
- Epigenetic dosage compensation mechanism in mammals -males and females have the same dose of genes on the X chromosome
- Inactive X chromosome shows DNA methylation of CpG islands



X inactivation

DNA methylation - where does it occur?

- CpG islands – usually unmethylated
- Intergenic regions – usually methylated
- Repetitive elements – usually methylated

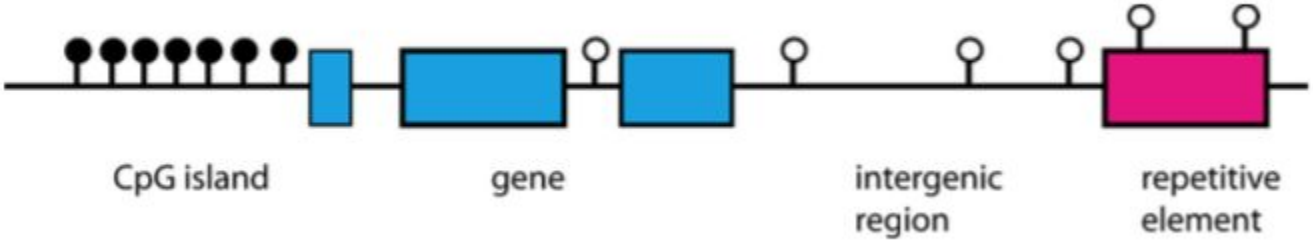


DNA methylation - function

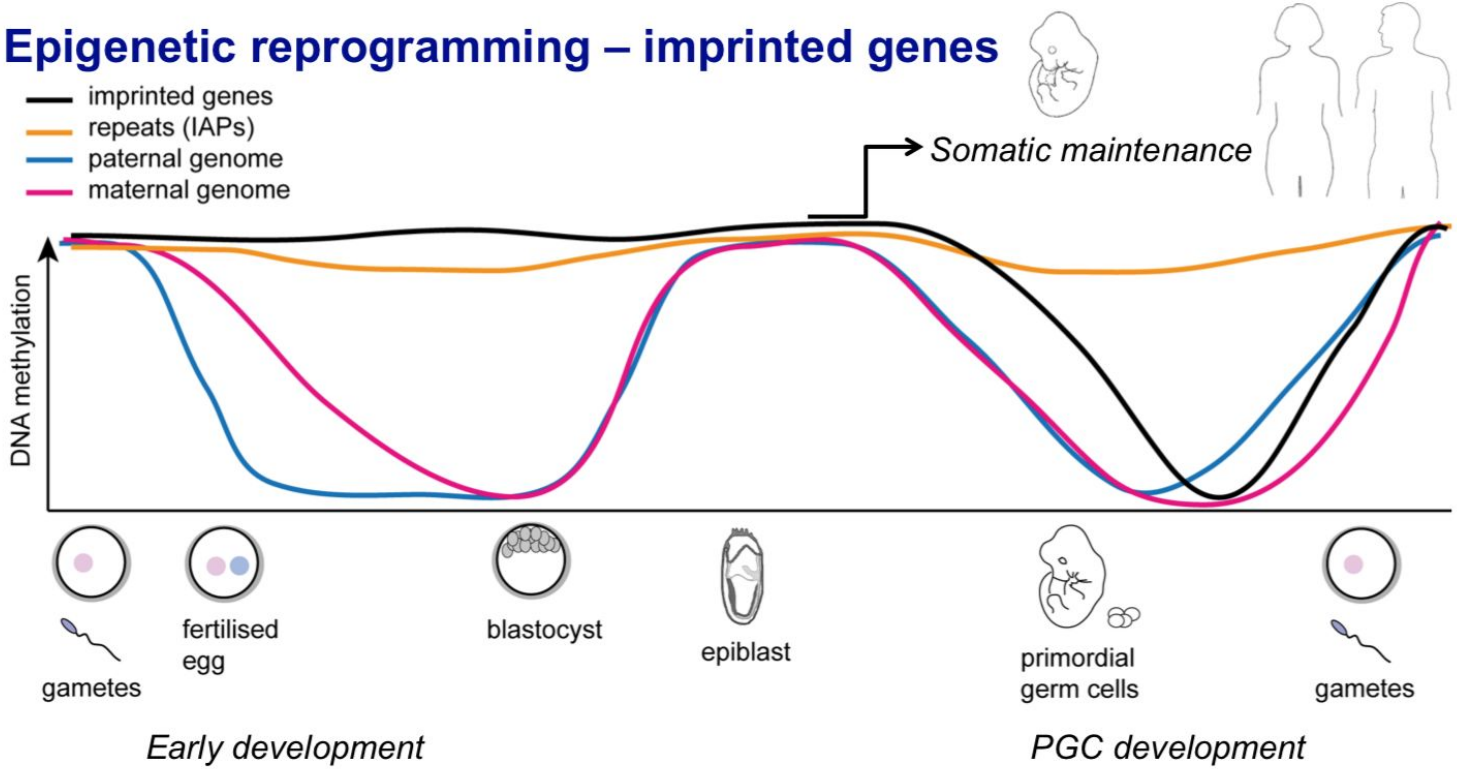
- Intergenic regions - Maintain genomic stability
- Repetitive regions
  - Silence repeats to prevent transposition
  - Mutate transposable elements
- Oncogenes
  - Silence expression

DNA methylation in cancer

- **Hypermethylation** of CpG islands
  - Disables tumor suppressors
- **Hypomethylation** of genome
  - Leads to genomic instability



DNA methylation through aging

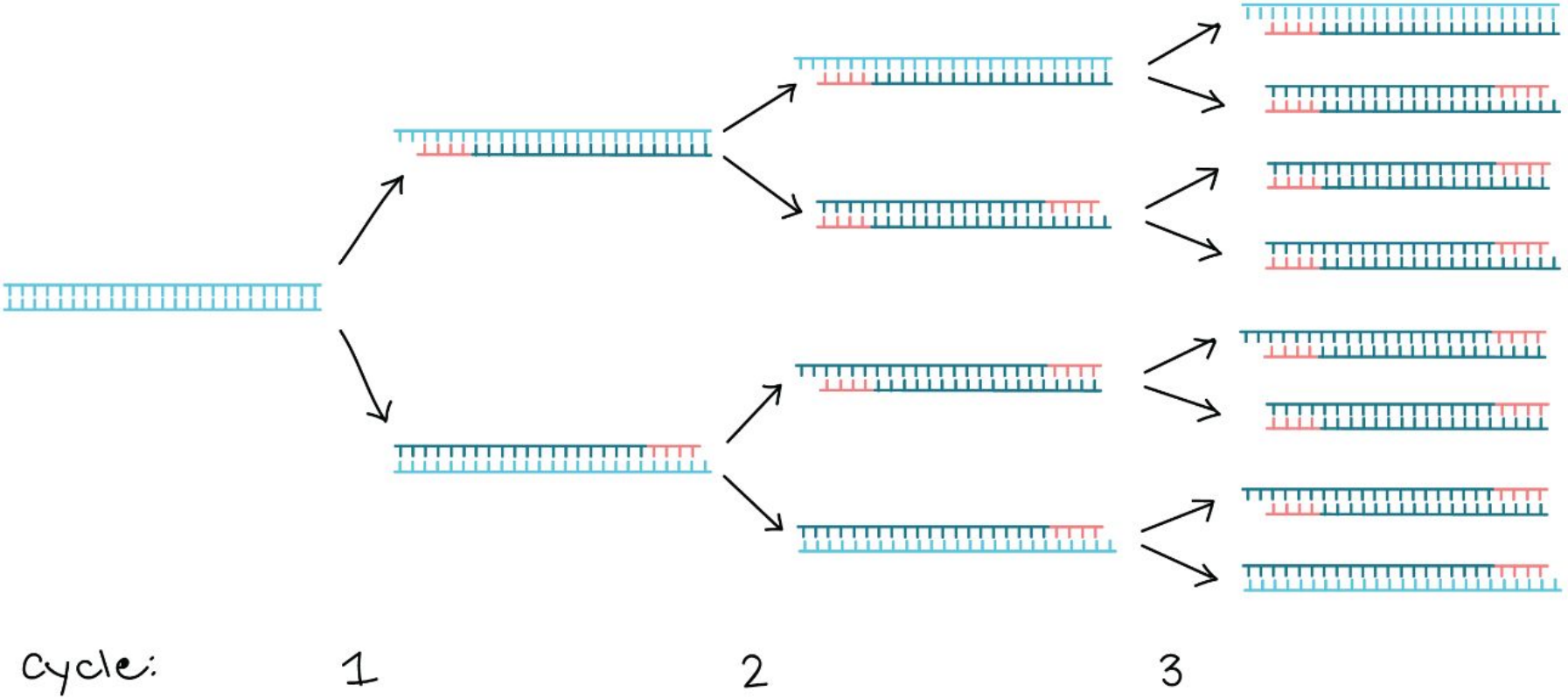


Bisulfite sequencing

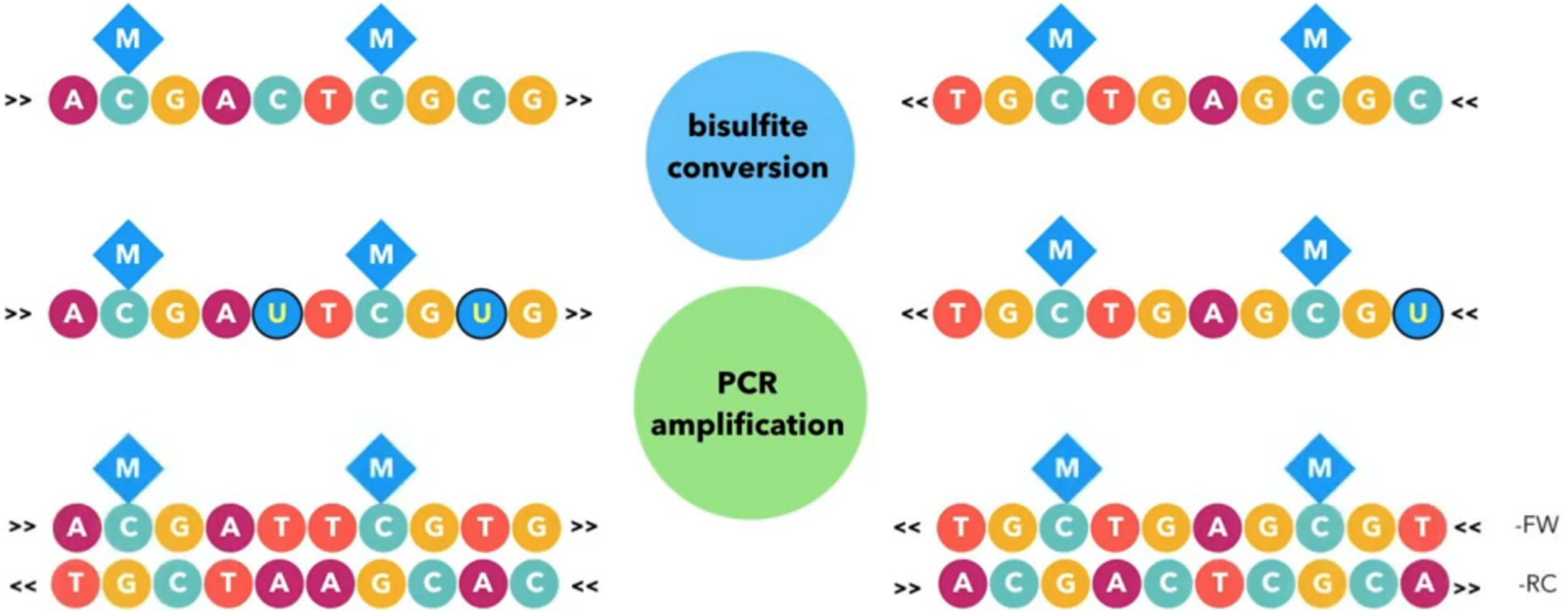
- Bisulfite modification converts non-methylated cytosines to uracils
- PCR amplification results in replacement of uracils with thymines

PCR

Recap



Bisulfite sequencing



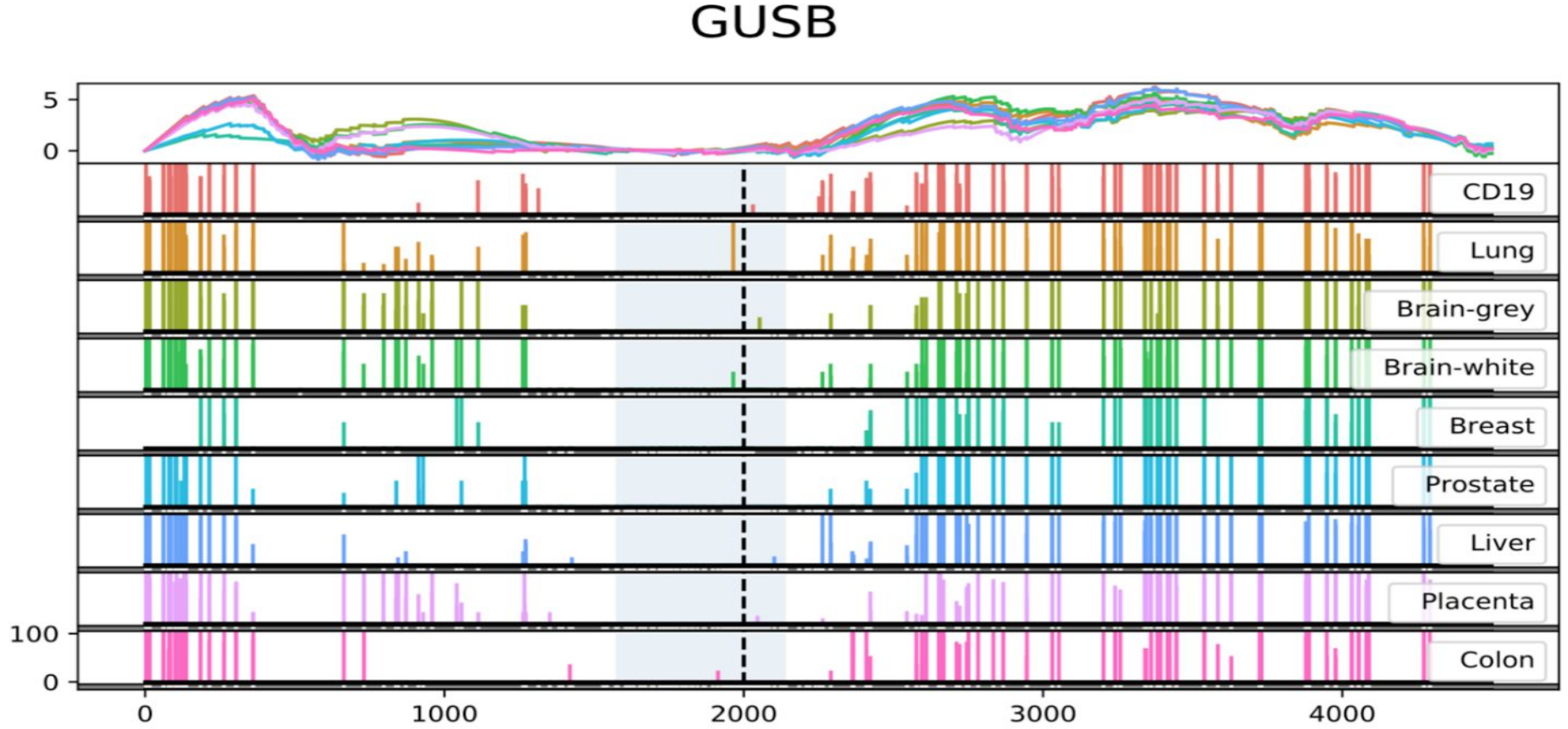
Bisulfite sequencing - alignment

- Similar to regular alignment
- Differences?

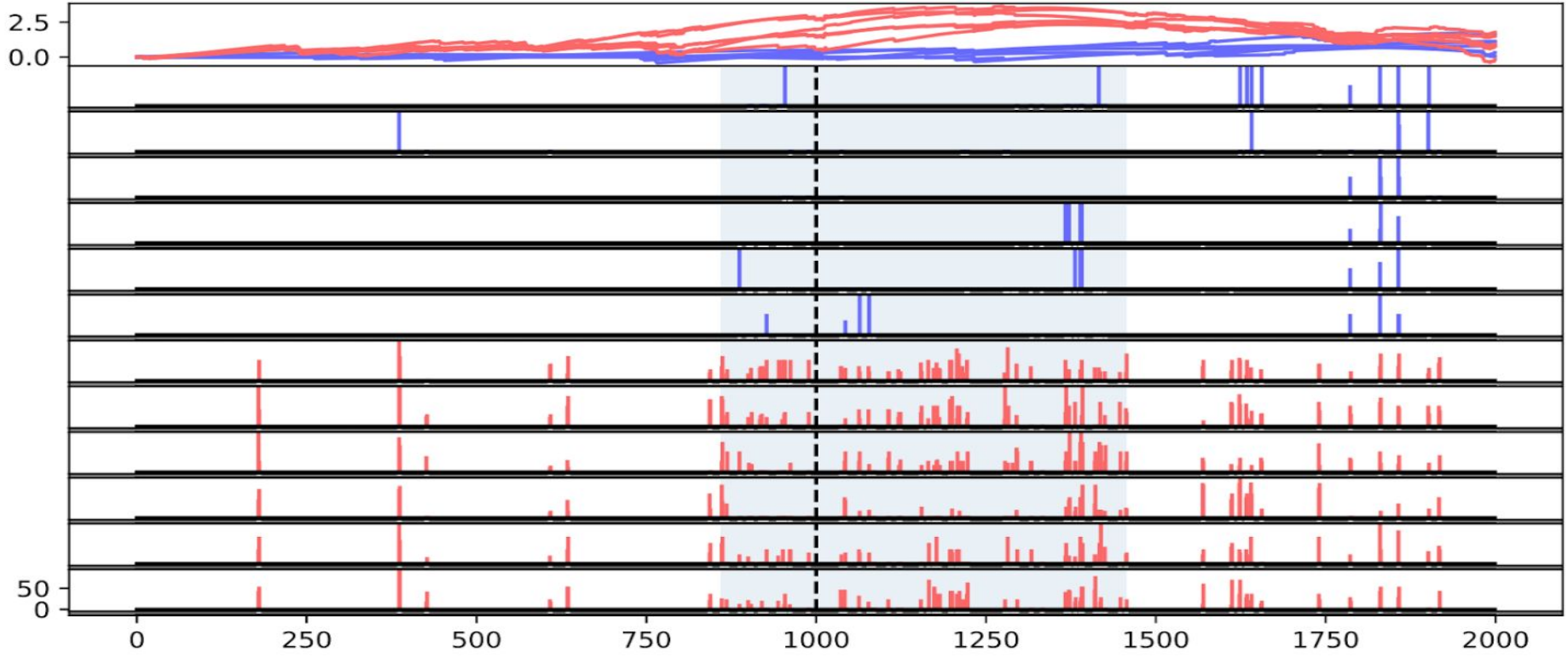
Bisulfite sequencing - alignment

- Dependant on library preparation protocol
  - Case when single strand is used
  - Case when both strands are used

Differential methylation - between cell types

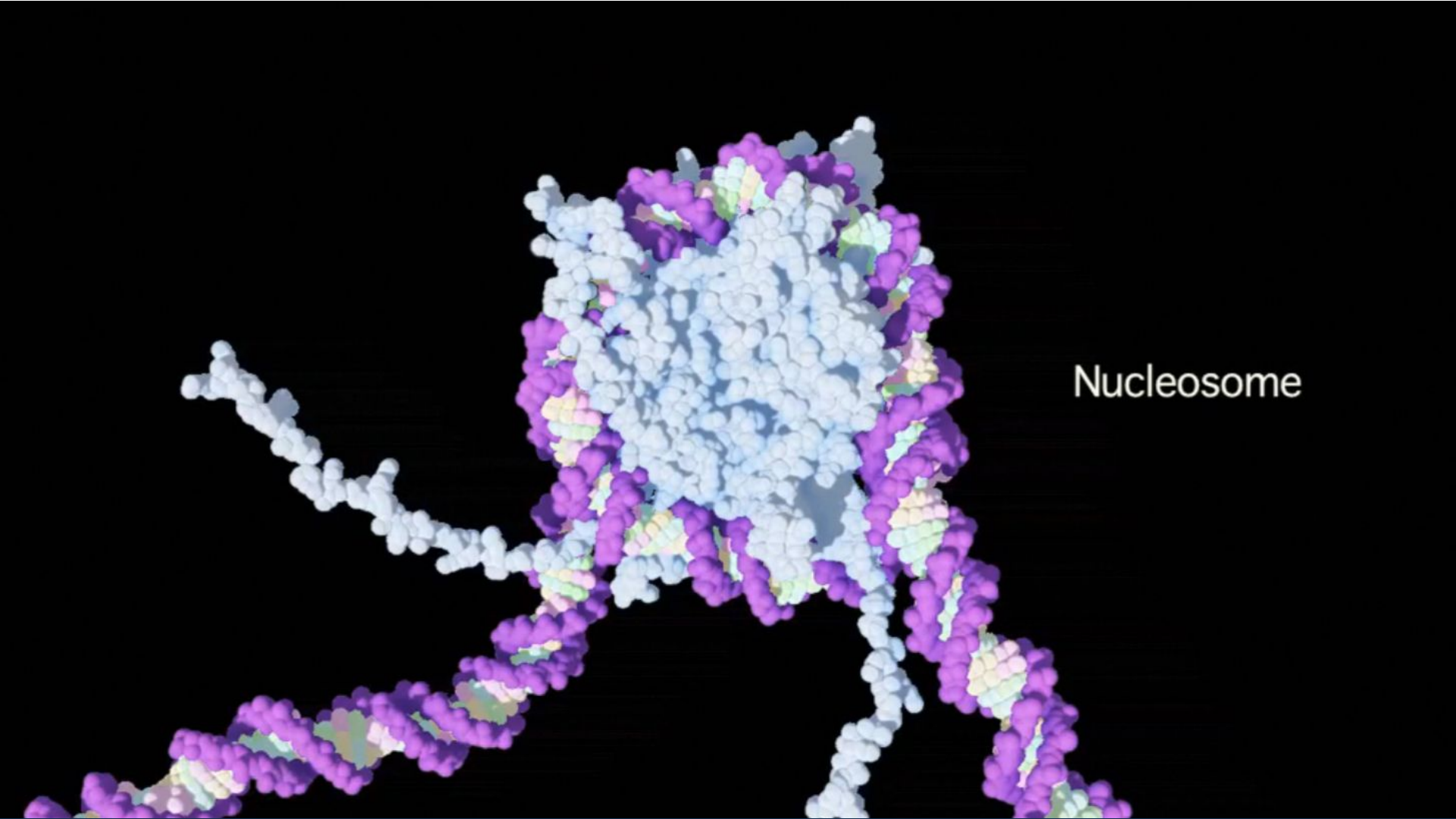


Differential methylation - in tumor



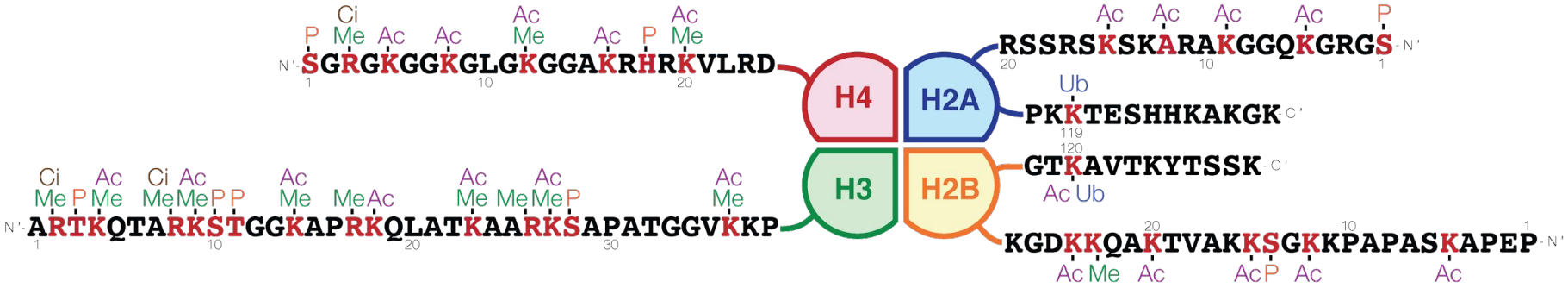
Histone modification

Modifications to structure holding DNA


Nucleosome

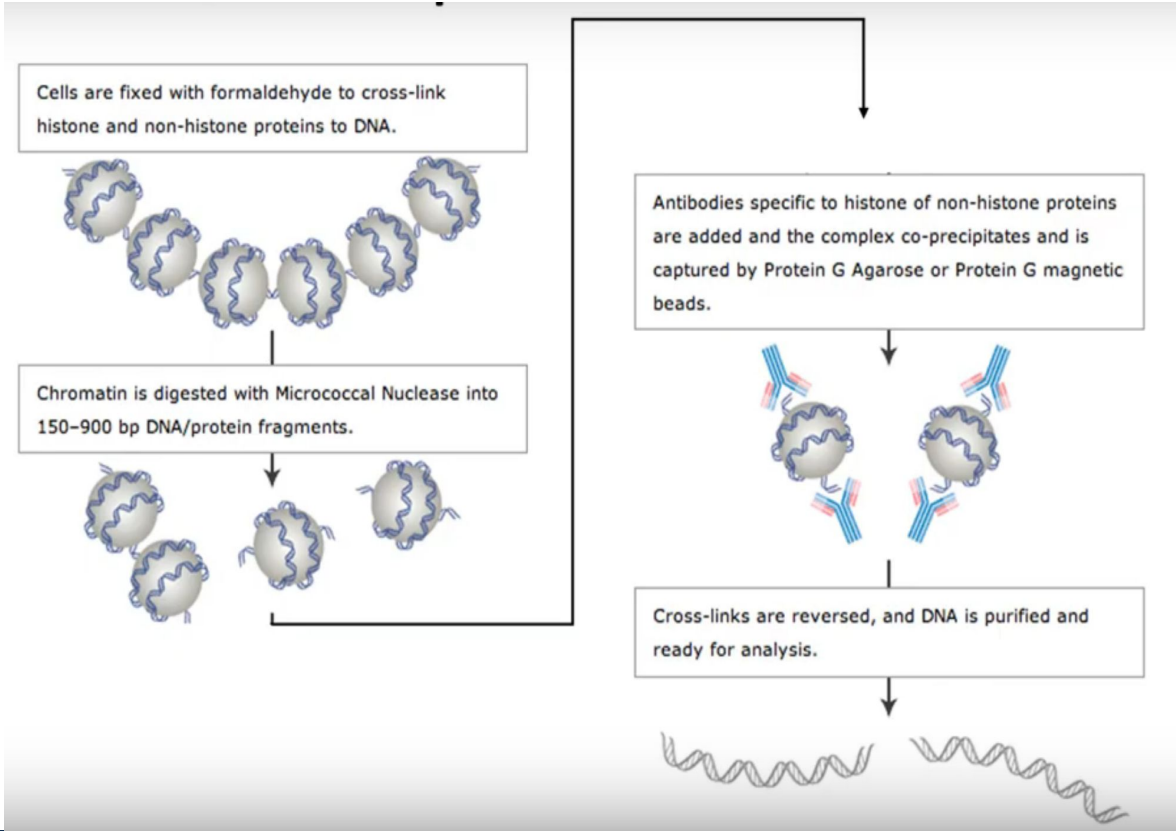
Histone modifications

- Chemical modifications of histone tails
- More than 50 different modifications

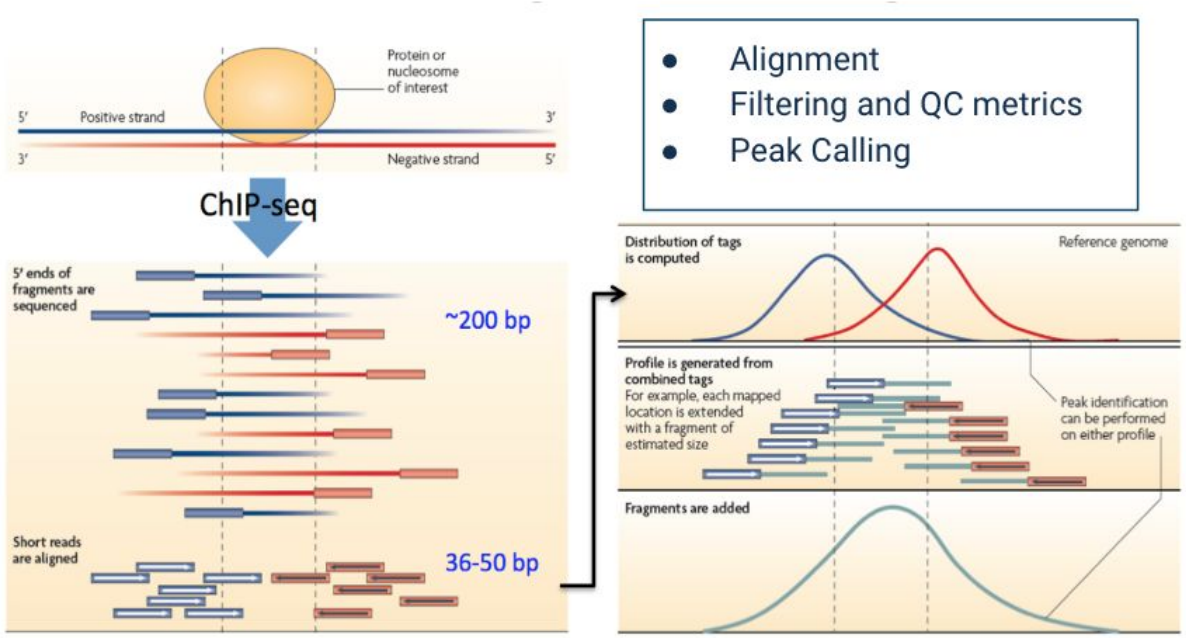


ChIP-seq

- Chromatin ImmunoPrecipitation



ChIP-seq



- Alignment
- Filtering and QC metrics
- Peak Calling

Other epigenetic modifications

- Chromatin remodeling - dependent on ATP
- Histone variants
- Noncoding RNAs
  - piRNAs
  - siRNAs
  - lncRNAs
- RNA modifications

Practical excercise

There are 3 files: Reference fasta file and two fasta files of bisulfite sequenced reads.

Task:

- Calculate GC content in the reference fasta
    - How many GC dinucleotides are present in this region?
    - Is it uniformly distributed or are there CpG islands? If there are CpG islands, how many islands are there?
- Which of given fasta files does represent a tumor sample and which one does represent a normal sample, if we know that this is a region of a tumor suppressor gene?

Literature

- Epigenetics review
- Bisulfite sequencing
- Bismark - bisulfite-seq aligner
- Differential methylation
- Chip-seq

Differential methylation exercise

```
import re
from Bio import SeqIO

reference = SeqIO.parse(open("/sbgenomics/project-files/example_human_reference.fasta"),'fasta')
fasta_1 = SeqIO.parse(open("/sbgenomics/project-files/example_1.fasta"),'fasta')
fasta_2 = SeqIO.parse(open("/sbgenomics/project-files/example_2"),'fasta')
ref_chromosome = list(reference)[0]
f1_chromosome = list(fasta_1)[0]
f2_chromosome = list(fasta_2)[0]

ref_sequence = str(ref_chromosome.seq)
f1_sequence = str(f1_chromosome.seq)
f2_sequence = str(f2_chromosome.seq)
```