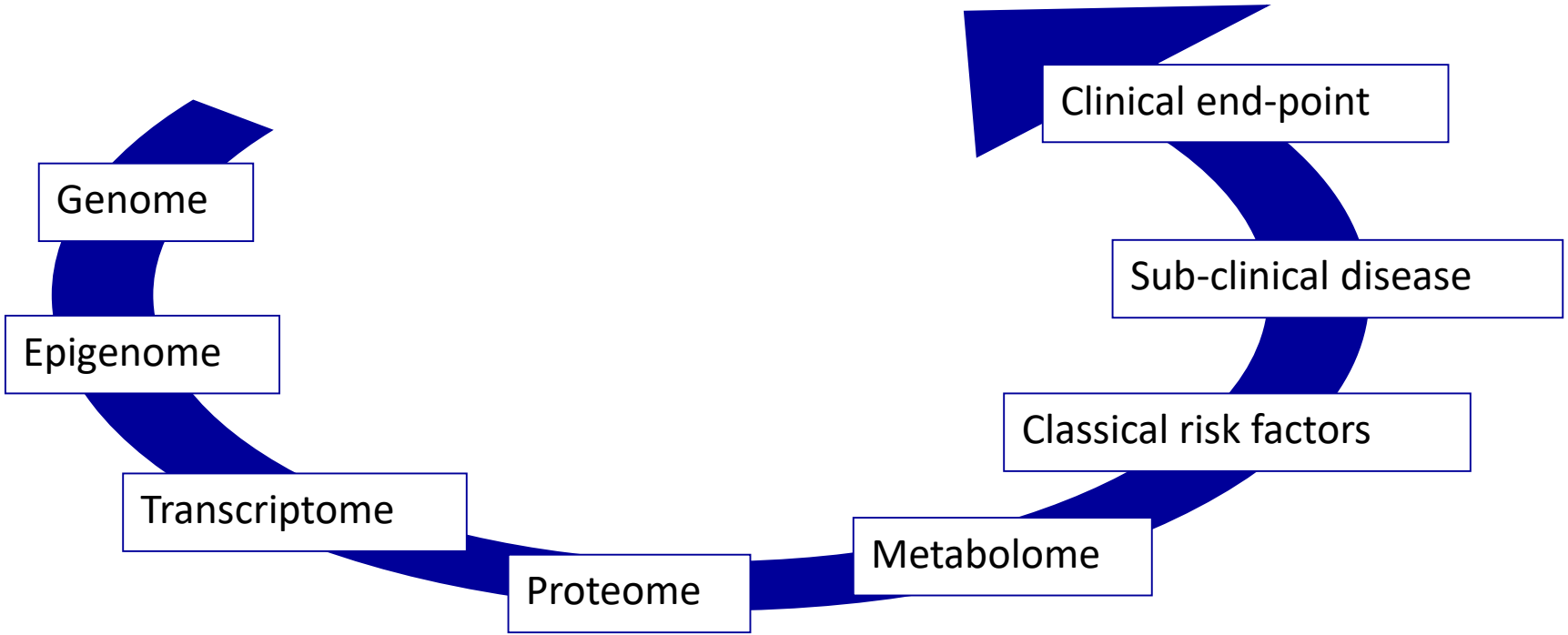
An introduction to the epigenome

Bas Heijmans and Roderick Slieker
Biomedical Data Sciences
Leiden University Medical Center
The Netherlands
bas.heijmans@lumc.nl

FOS course Molecular Data Science – 26 October 2020


Genome
Epigenome
Transcriptome
Proteome
Metabolome
Classical risk factors
Sub-clinical disease
Clinical end-point

On offer

- The molecular basis of epigenetics
- The role of epigenetics in fundamental biology
- Epigenetics as integrator of environmental signals
- Epigenetics in disease mechanisms

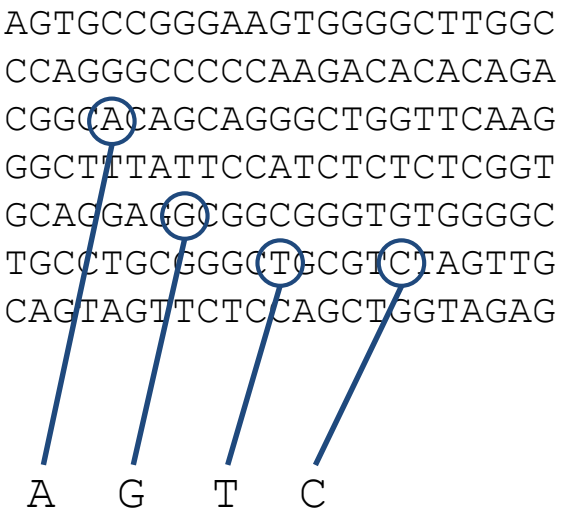
20 hours after fertilization

Day 3, 8 cells

The genetic code is not enough

AGTGCCGGGAAGTGGGGCTTGGC
CCAGGGCCCCCAAGACACACAGA
CGGCACAGCAGGGCTGGTTCAAG
GGCTTTATTCCATCTCTCTCGGT
GCAGGAGGCGGCGGGTGTGGGGC
TGCCTGCGGGCTGCGTCTAGTTG
CAGTAGTTCTCCAGCTGGTAGAG

The genetic code is not enough



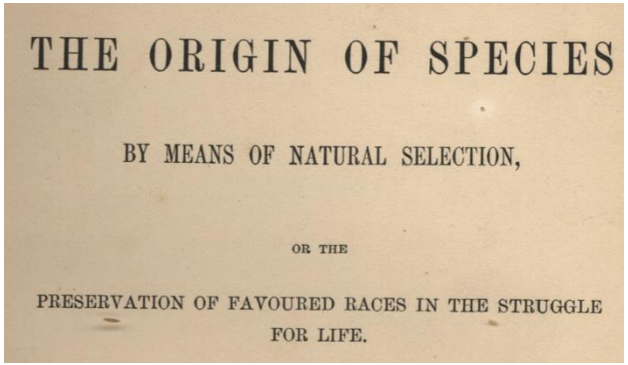
The genetic code is not enough

AGTGCCGGGAAGTGGGGCTTGGC
CCAGGGCCCCCAAGACACACAGA
CGGCACAGCAGGGCTGGTTCAAG
GGCTTTATTCCATCTCTCTCGGT
GCAGGAGGCGGCGGGTGTGGGGC
TGCCTGCGGGCTGCGTCTAGTTG
CAGTAGTTCTCCAGCTGGTAGAG

The genetic code is not enough

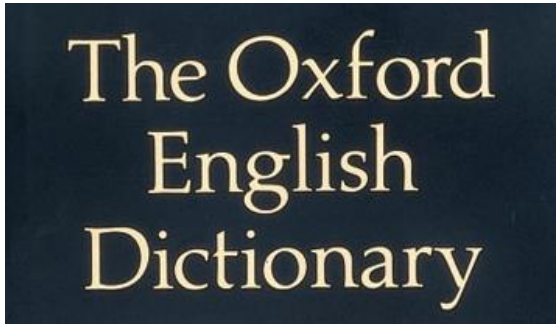
AGTGCCGGGAAGTGGGGCTTGGC
CCAGGGCCCCCAAGACACACAGA
CGGCACAGCAGGGCTGGTTCAAG
GGCTTTATTCCATCTCTCTCGGT
GCAGGAGGCGGCGGGTGTGGGGC
TGCCTGCGGGCTGCGTCTAGTTG
CAGTAGTTCTCCAGCTGGTAGAG

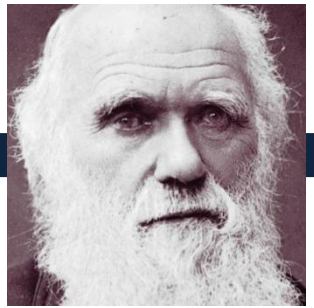
The Oxford English Dictionary

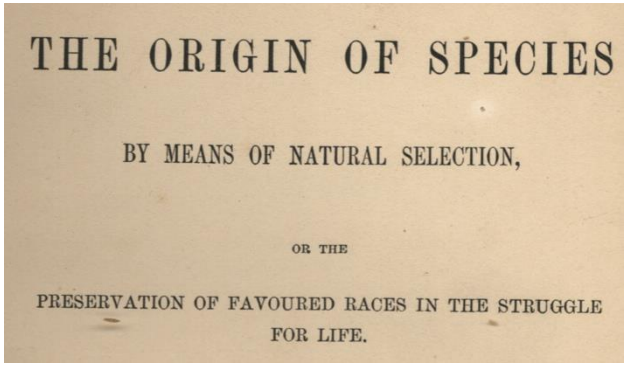


The genetic code is not enough

AGTGCCGGGAAGTGGGGCTTGGC
CCAGGGCCCCCAAGACACACAGA
CGGCACAGCAGGGCTGGTTCAAG
GGCTTTATTCCATCTCTCTCGGT
GCAGGAGGCGGCGGGTGTGGGGC
TGCCTGCGGGCTGCGTCTAGTTG
CAGTAGTTCTCCAGCTGGTAGAG

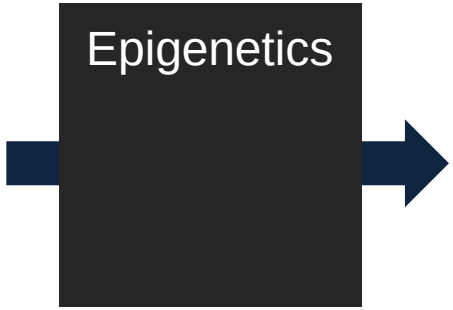






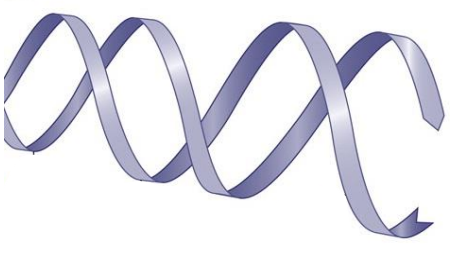
The genetic code is not enough

AGTGCCGGGAAGTGGGGCTTGGC
CCAGGGCCCCCAAGACACACAGA
CGGCACAGCAGGGCTGGTTCAAG
GGCTTTATTCCATCTCTCTCGGT
GCAGGAGGCGGCGGGTGTGGGGC
TGCCTGCGGGCTGCGTCTAGTTG
CAGTAGTTCTCCAGCTGGTAGAG

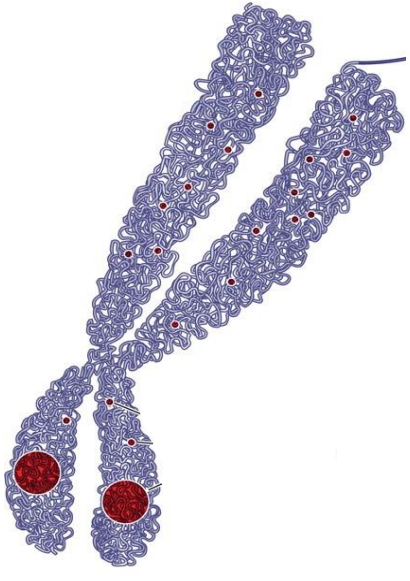


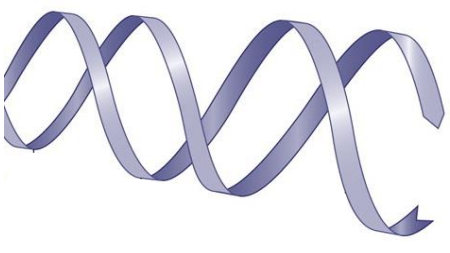
- Instructing the DNA where, when and how much to express a gene.
- Epigenetics provides variation & memory.

Molecular dimmers on the DNA

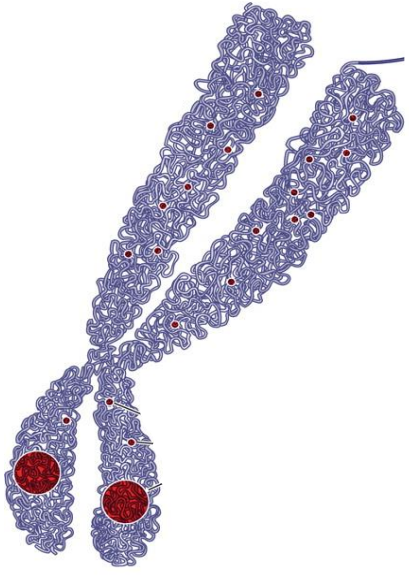


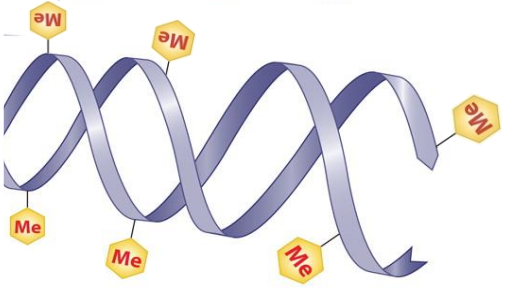
Molecular dimmers on the DNA



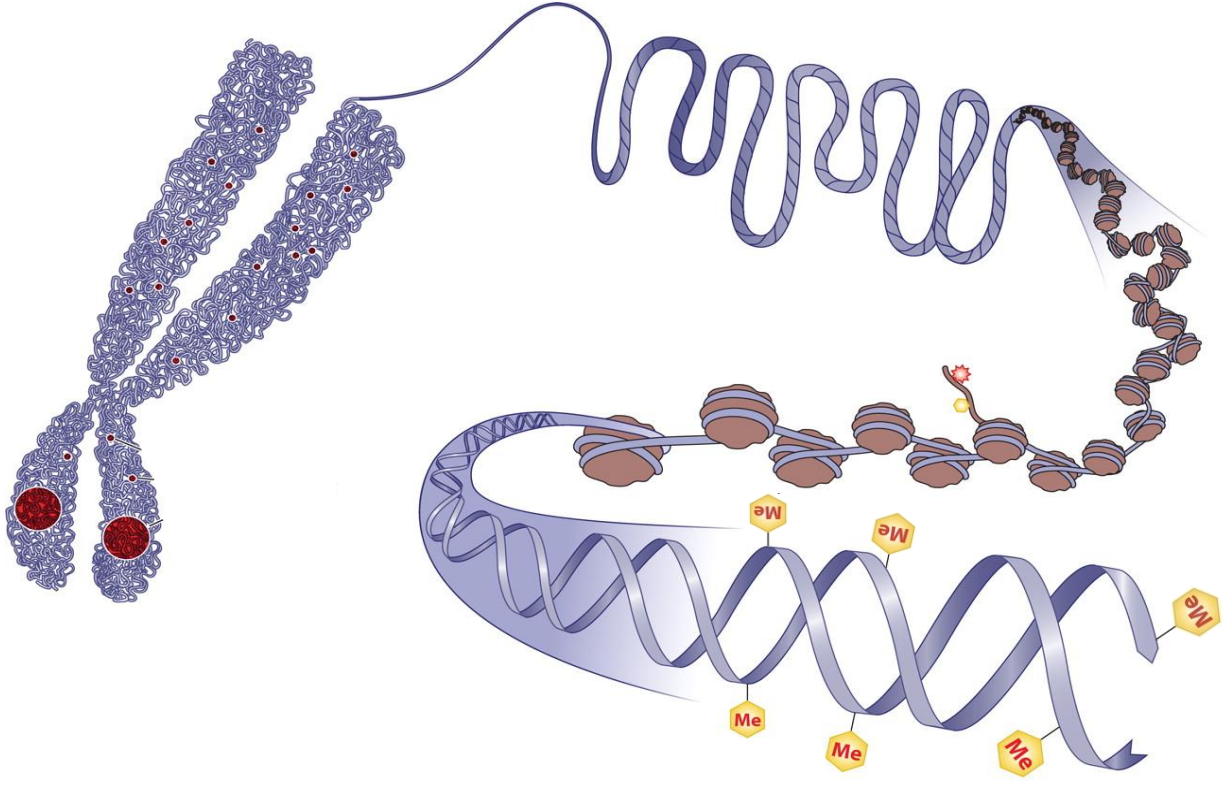


Molecular dimmers on the DNA



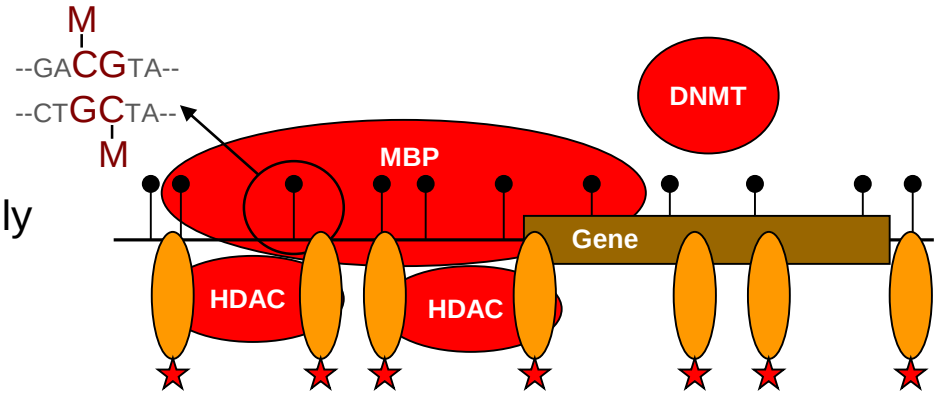


Molecular dimmers on the DNA

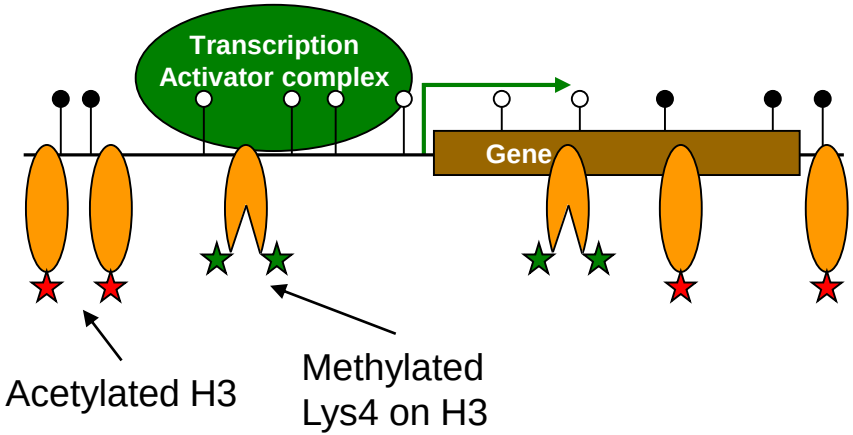


The simplified (and outdated) text book view

Transcriptionally **silent** DNA



Transcriptionally **active** DNA



Disclaimer: The order of events (cause and consequence) is not known

Epigenetic mechanisms

1. Control gene expression
   (by changes in the accessibility of DNA and recruiting regulatory factors like transcription factors and chromatin modifiers).
2. Stable, long-term, but in principle reversible.
3. Transmitted during cell division, particularly mitosis.

Epigenome: the whole of epigenetics marks in a cell.

Epigenome Projects

IHEC

International Human Epigenome Consortium

Home | Areas of focus | Standard Operating Procedures | Tools / Useful Information | Policies and Guidelines | IHEC Structure | Outreach and Training | Feedback | Intranet

Links

The EPIGENOME Network of Excellence

The focal point for the European

Overview

What is the International Human Epigenome Consortium?

Consortium goals

Primary Goals

1. Coordinate the production of reference maps of

There is one human (reference) genome. How many epigenomes?

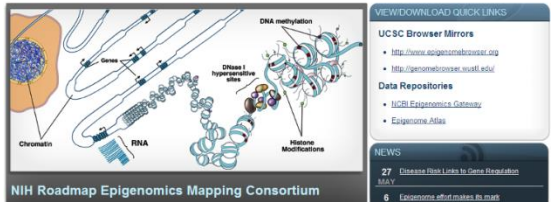
Centre of Excellence funded by The Danish National Research Foundation.

ROADMAP epigenomics

regulating gene expression and in controlling specific cellular functions. Epigenetic processes go beyond DNA-stored information and are essential for packaging and interpreting the genome, are fundamental to normal

years. To reach this goal, the consortium will use robust and validated technologies to generate:

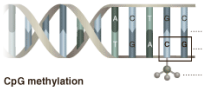
- very high resolution maps of informative histone modifications



LUMC

Mapping the Epigenome

DNA contains the genetic blueprint for all human cells, but the reading and execution of the blueprint inside each cell is controlled in part by chemical markers attached to the DNA. Scientists have begun to map some of these epigenetic markers, including CpG methylation.



CpG methylation
DNA is a code written with four letters: **A**, **T**, **C** and **G**, each standing for one nucleotide.

In CpG methylation, a small marker called a methyl group attaches to the DNA at a CpG site, where a **C** and a **G** nucelotide sit next to each other.

Reading the chart
The outer ring represents 35 million base pairs in Chromosome 22. Orange marks highlight areas of the chromosome that were tested for CpG methylation in a pilot study by the Human Epigenome Project.

Measuring CpG methylation
Bar charts indicate the average amount of CpG methylation found within the tested areas. Each chart covers 100,000 base pairs. Some charts have been shifted, shown with connecting lines.

AMOUNT OF METHYLATION
- 0 to 20%
- 20 to 80%
- 80 to 100% of CpG sites

1 million base pairs

MUSCLE
LIVER
HEART
WHITE BLOOD CELL
SKIN
PLACENTA
SPERM

Chromosome 22
Of the 23 pairs of chromosomes in the human genome, 22 is the second smallest, containing only about 2 percent of DNA in the genome.

Gray and white bands on the circular chart correspond to these bands on the chromosome.

Genes
Some of the known genes from Chromosome 22 that fall within the tested areas are shown outside the chart. CpG methylation is one of several epigenetic factors that is thought to influence how genes are transcribed or silenced within cells.

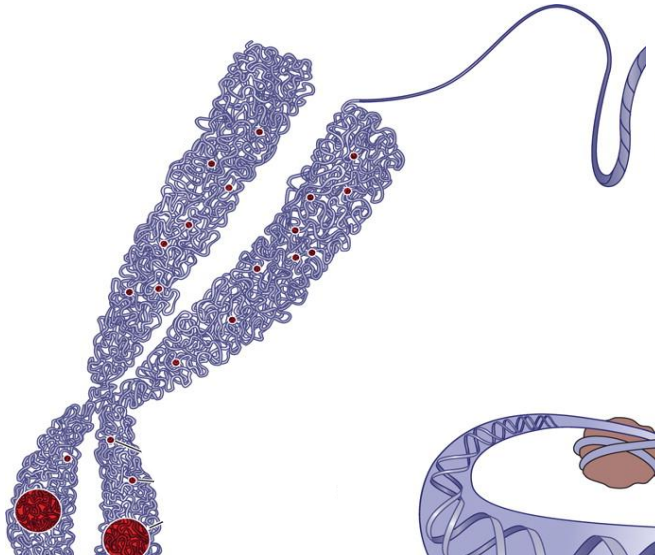
Variation among tissues
Each concentric ring of bar charts represents a different tissue, from muscle cells to sperm cells. Methylation levels that are significantly above or below the average level across all of the tissues are highlighted, indicating possible cell-specific differences.

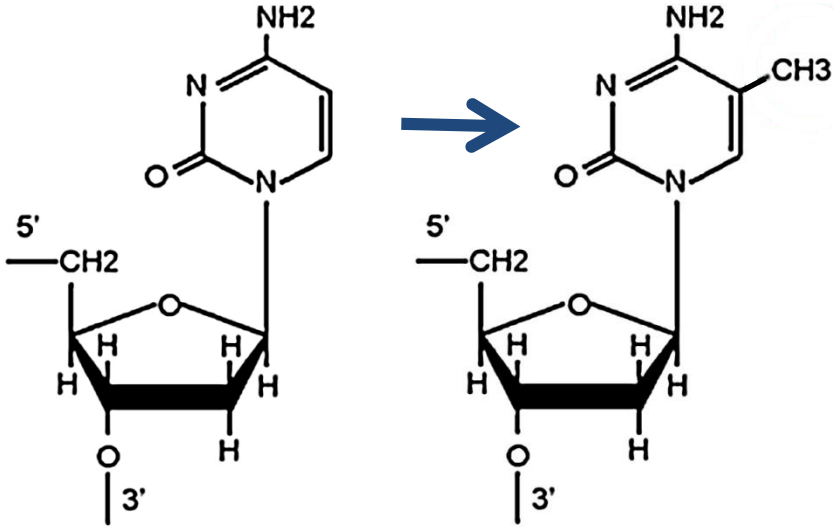
- 20% or more above average
- 20% or more below average

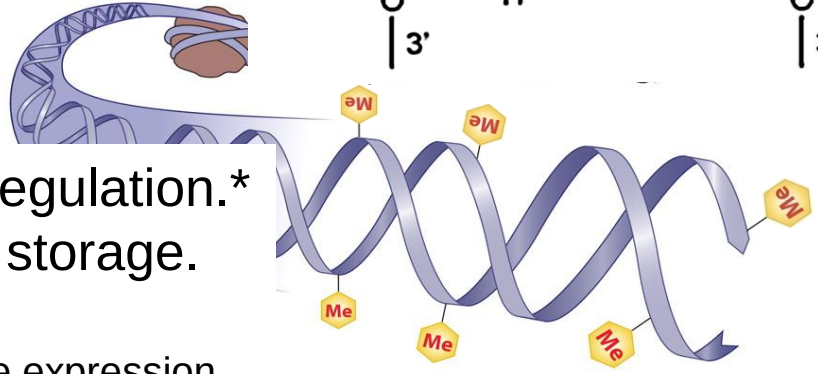
Sources: Human Epigenome Project; Nature Genetics; Dr. Florian Eckhardt; Dr. Stephan Beck

JONATHAN CORUM/THE NEW YORK TIMES AND MARTIN KRZYWINSKI/CANADA'S MICHAEL SMITH GENOME SCIENCES CENTER

Focus on DNA methylation



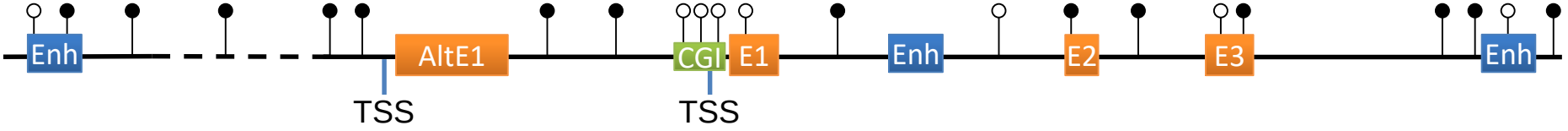




- Signals altered genomic regulation.*
- Standard biomaterial and storage.
- High-throughput profiling.

* Controls, stabilizes or reflects gene expression.

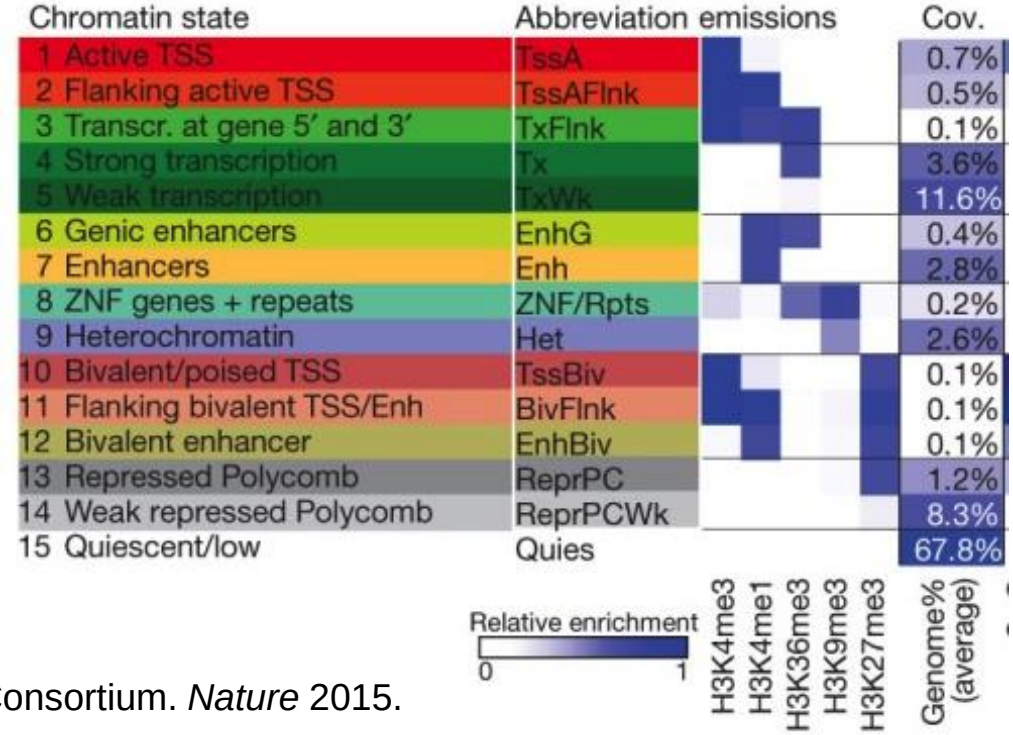
Significance methylation depends on genomic context



→ Genomic annotation is essential when interpreting DNA methylation data

Reference epigenomes inform on biological function

Chromatin states (or segmentation)



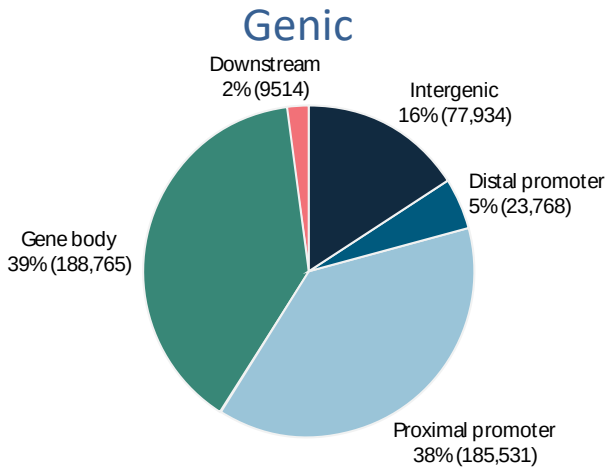
Roadmap Epigenomics Consortium. *Nature* 2015.

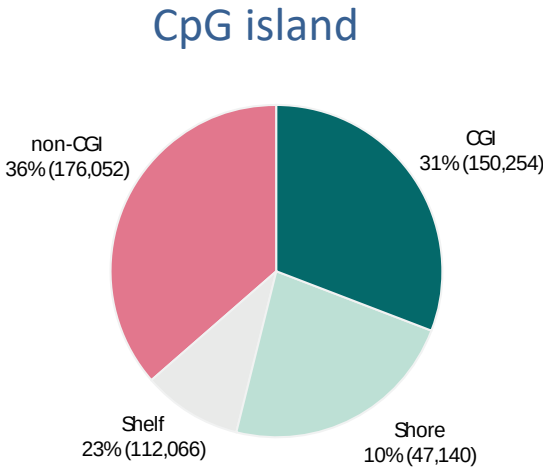
Practical

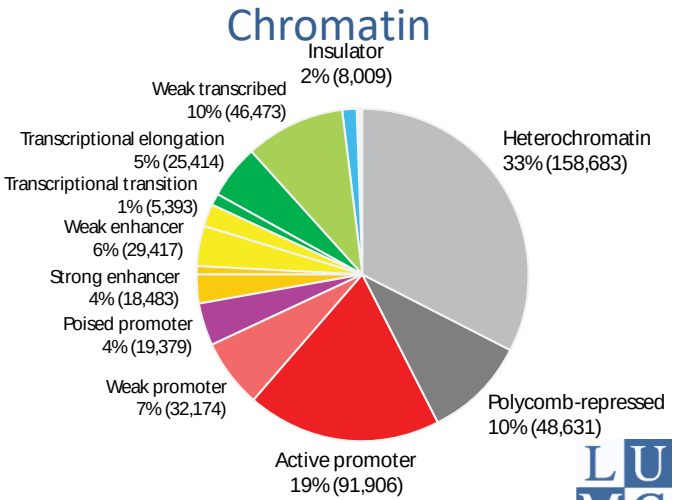
- Characterizing DNA methylation differences between tissues.
- Genome-wide methylation of close to 500 thousand CpG sites.

Illumina 450k DNA methylation array

- Informative: ~483,000 CpGs (<2%) but many annotations
- Affordable: ~200 euro per sample
- Fast: >1000 samples a week
- Data: relatively easy to manage (amount and complexity)







Principle methylation array

M? M? M? M? x 485,000 (out of 28M)

...CCTGGTACGTCTAGCCGTAATTAGCTCGATTCCGGT...

Bisulfite conversion

unmethylated

methylated

...UUTGGTUGTUTAG...

...UUTGGTCGTUTAG...

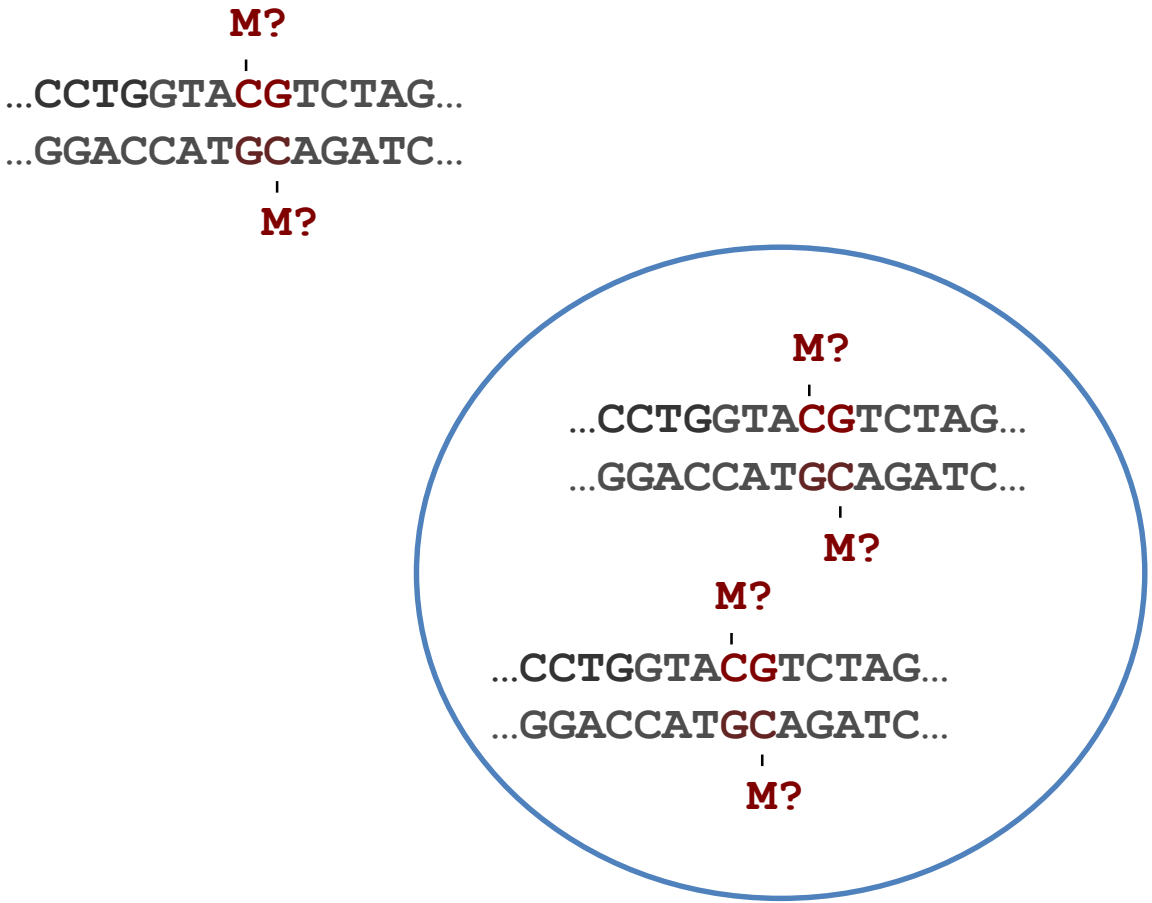
unmeth signal

meth signal

β-value = meth / (meth + unmeth)

Possible DNA methylation levels of one CpG site

- A DNA molecule?
- A cell?


M?
...CCTGGTACGTCTAG...
...GGACCATGCAGATC...
M?
M?
...CCTGGTACGTCTAG...
...GGACCATGCAGATC...
M?
M?
...CCTGGTACGTCTAG...
...GGACCATGCAGATC...
M?

Possible DNA methylation levels of one CpG site

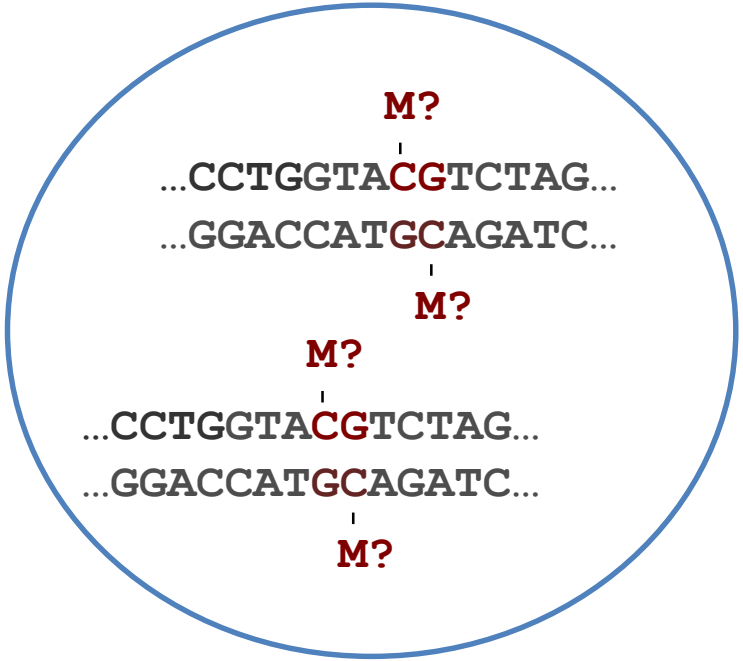
- A DNA molecule?
- A cell?
- A blood sample?
- A muscle biopsy?

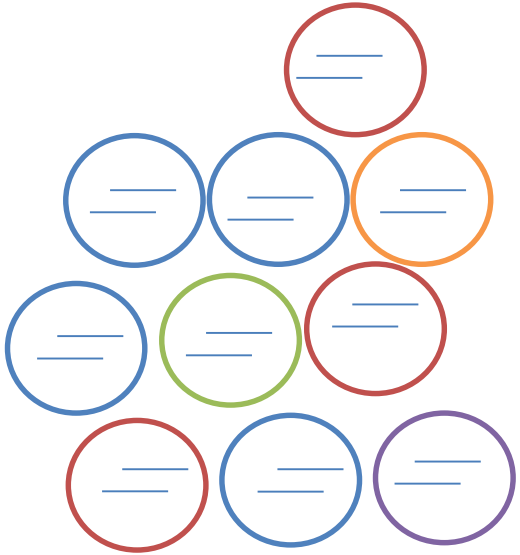
M?

...CCTGGTACGTCTAG...

...GGACCATGCAGATC...

M?





4*1 + 3*0 + 1*0.5 + 1*1 + 1*1


Amazing
DISCOVERIES