An introduction to the epigenome
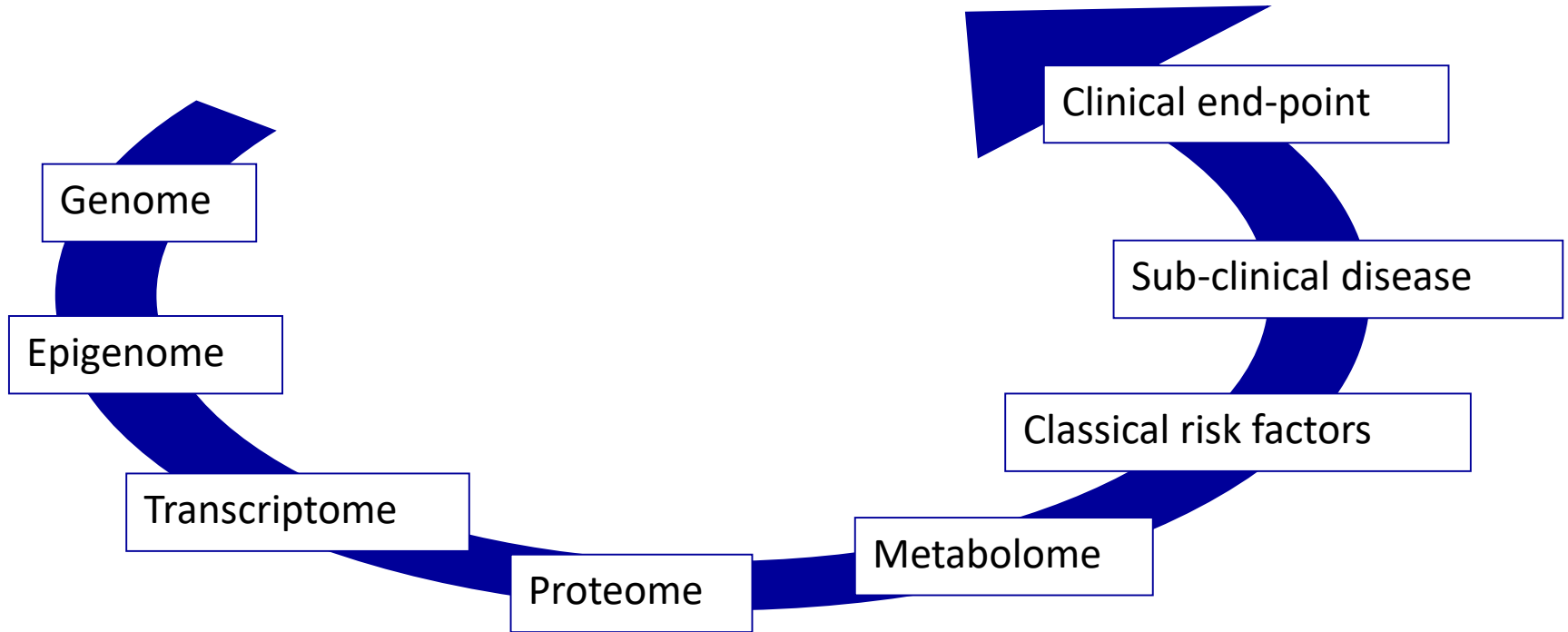
Bas Heijmans
Biomedical Data Sciences
Leiden University Medical Center
The Netherlands
bas.heijmans@lumc.nl

FOS course Molecular Data Science – 28 October 2019

LU
MC

Genome

Epigenome

Transcriptome

Proteome

Metabolome

Classical risk factors

Sub-clinical disease

Clinical end-point

LU
MC

On offer

- The molecular basis of epigenetics

- The role of epigenetics in fundamental biology

- Epigenetics as integrator of environmental signals

- Epigenetics in disease mechanisms

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

A   G   T   C

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

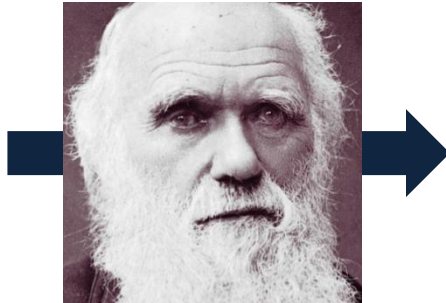The Oxford English Dictionary

THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

OR THE

PRESERVATION OF FAVOURED RACES IN THE STRUGGLE
FOR LIFE.

The genetic code is not enough

AGTGCCGGGAAGTGGGGCTTGGC
CCAGGGCCCCCAAGACACACAGA
CGGCACAGCAGGGCTGGTTCAAG
GGCTTTATTCCATCTCTCTCGGT
GCAGGAGGCGGCGGGTGTGGGGC
TGCCTGCGGGCTGCGTCTAGTTG
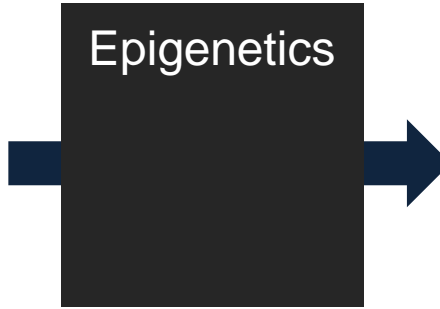CAGTAGTTCTCCAGCTGGTAGAG

The Oxford English Dictionary

THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

OR THE

PRESERVATION OF FAVOURED RACES IN THE STRUGGLE
FOR LIFE.

The genetic code is not enough

```
AGTGCCGGGAAGTGGGGCTTGGC
CCAGGGCCCCCAAGACACACAGA
CGGCACAGCAGGGCTGGTTCAAG
GGCTTTATTCCATCTCTCTCGGT
GCAGGAGGCGGCGGGTGTGGGGC
TGCCTGCGGGCTGCGTCTAGTTG
CAGTAGTTCTCCAGCTGGTAGAG
```
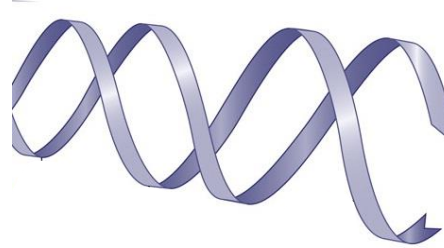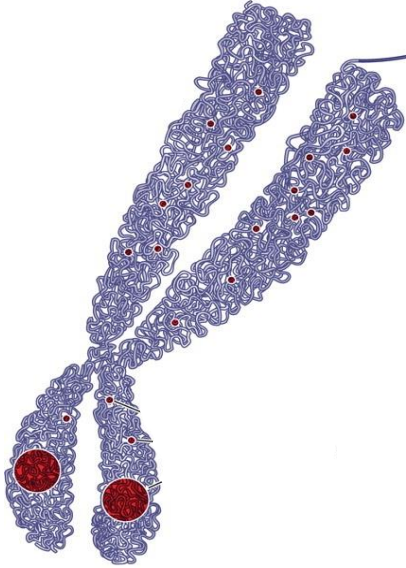
Epigenetics

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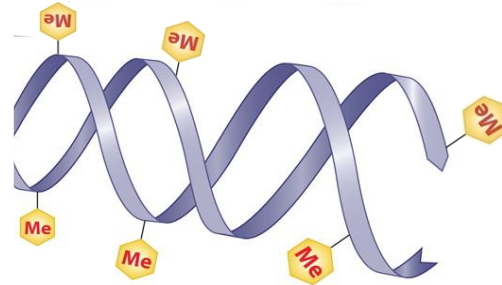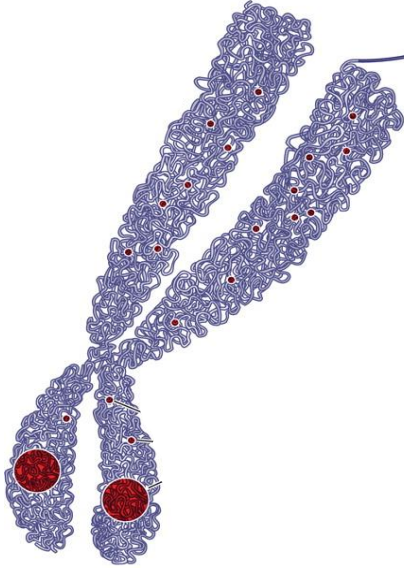- Epigenetics provides variation & memory.
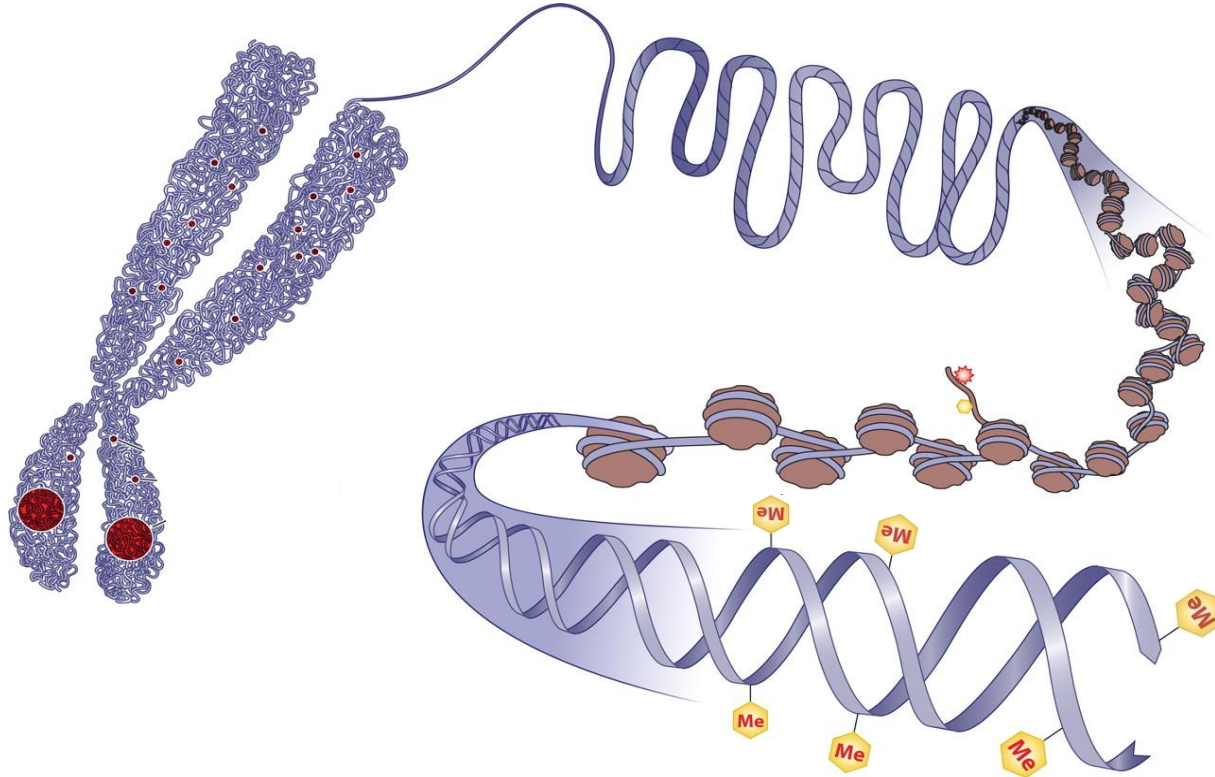
LU
MC

Molecular dimmers on the DNA
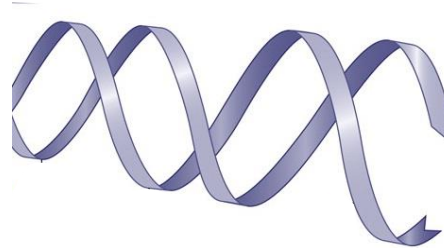
Molecular dimmers on the DNA

Molecular dimmers on the DNA

Molecular dimmers on the DNA

The simplified (and outdated) text book view

Acetylated H3

Methylated Lys4 on H3
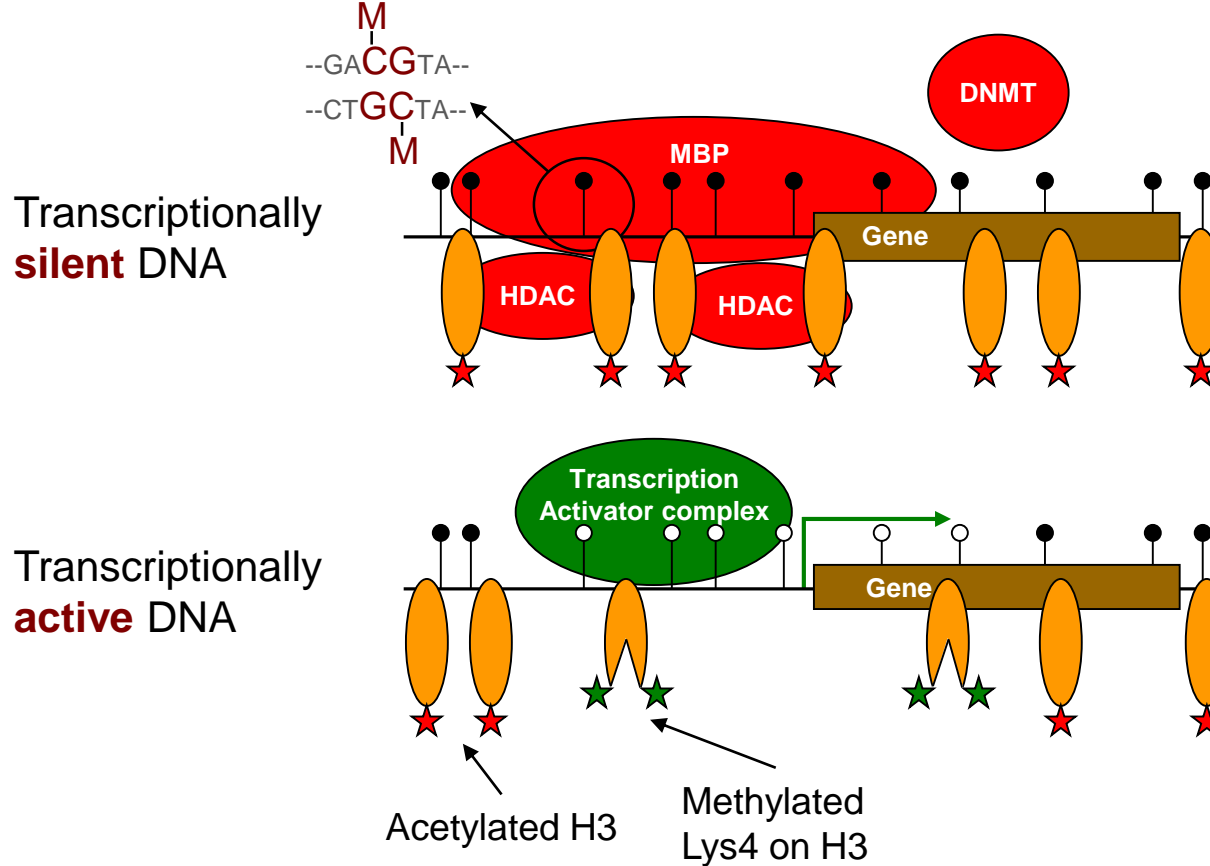
Disclaimer: The order of events (cause and consequence) is not known

20 hours after fertilization

Day 3, 8 cells

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

International Human Epigenome Consortium

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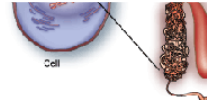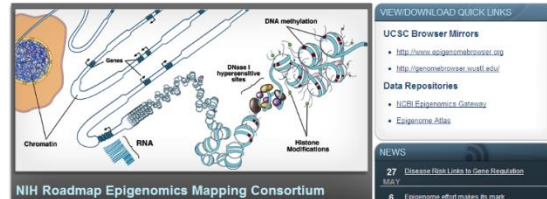
Mapping the Epigenome

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.
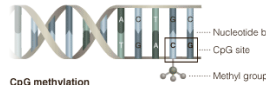
CpG methylation

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** nucleotide sit next to each other.

Nucleotide base pair
CpG site
Methyl group

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.

1 million base pairs

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

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.

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

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

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

Focus on DNA methylation
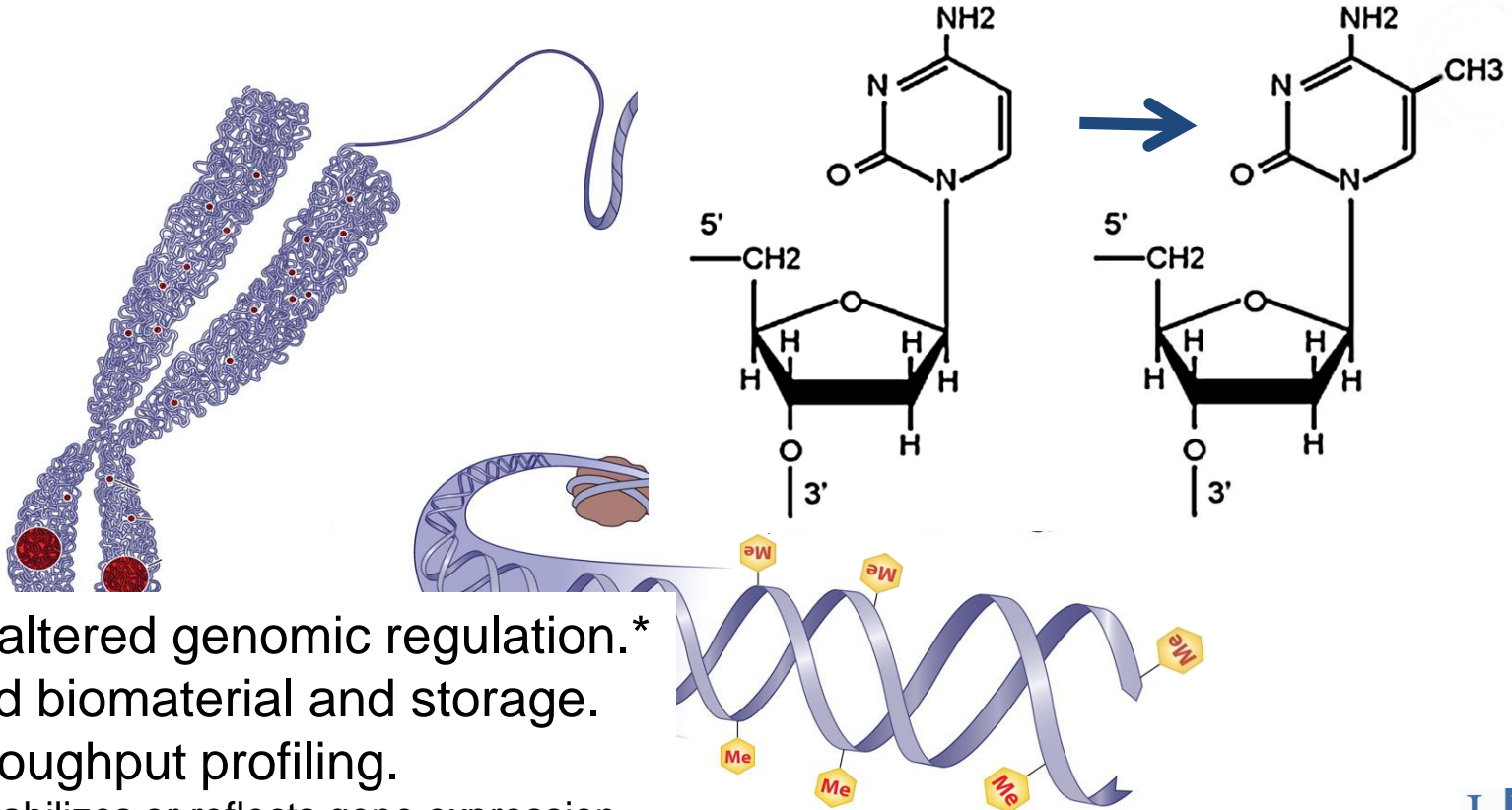
- Signals altered genomic regulation.*
- Standard biomaterial and storage.
- High-throughput profiling.

* 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)

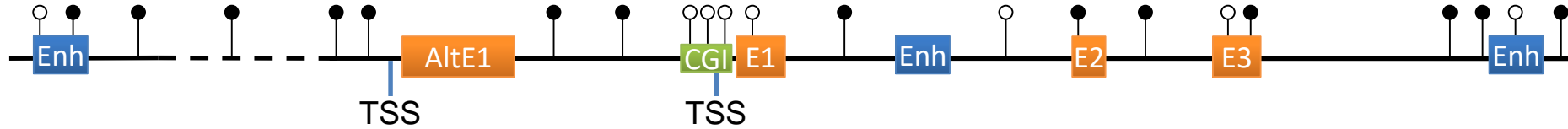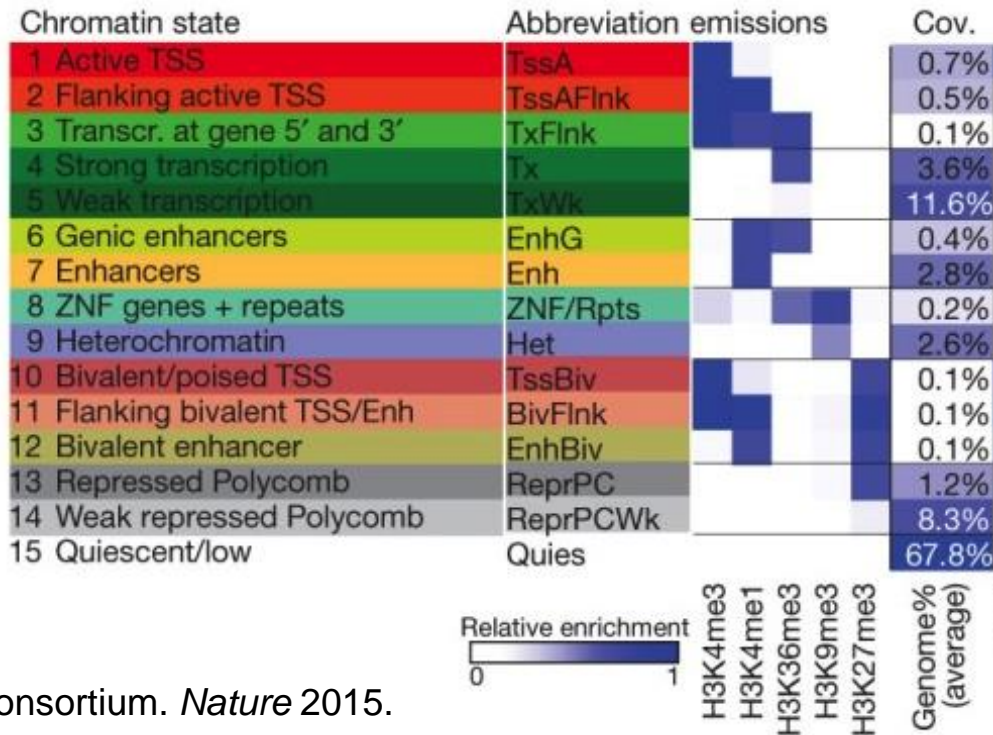| Chromatin state	Abbreviation	emissions					Cov.
1 Active TSS	TssA						0.7%
2 Flanking active TSS	TssAFlnk						0.5%
3 Transcr. at gene 5' and 3'	TxFlnk						0.1%
4 Strong transcription	Tx						3.6%
5 Weak transcription	TxWk						11.6%
6 Genic enhancers	EnhG						0.4%
7 Enhancers	Enh						2.8%
8 ZNF genes + repeats	ZNF/Rpts						0.2%
9 Heterochromatin	Het						2.6%
10 Bivalent/poised TSS	TssBiv						0.1%
11 Flanking bivalent TSS/Enh	BivFlnk						0.1%
12 Bivalent enhancer	EnhBiv						0.1%
13 Repressed Polycomb	ReprPC						1.2%
14 Weak repressed Polycomb	ReprPCWk						8.3%
15 Quiescent/low	Quies						67.8%
		H3K4me3	H3K4me1	H3K36me3	H3K9me3	H3K27me3	Genome% (average)

Relative enrichment 0 — 1

Roadmap Epigenomics Consortium. *Nature* 2015.
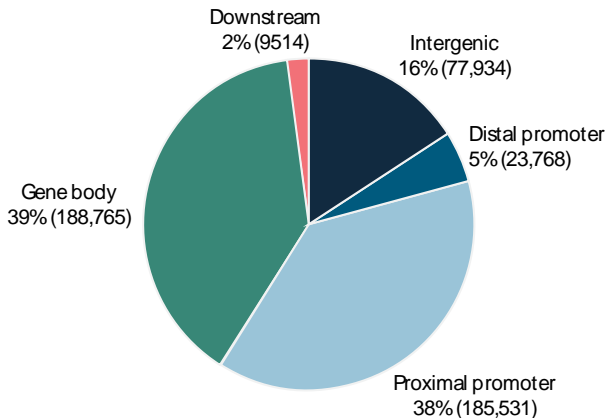
LU MC

Practical

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

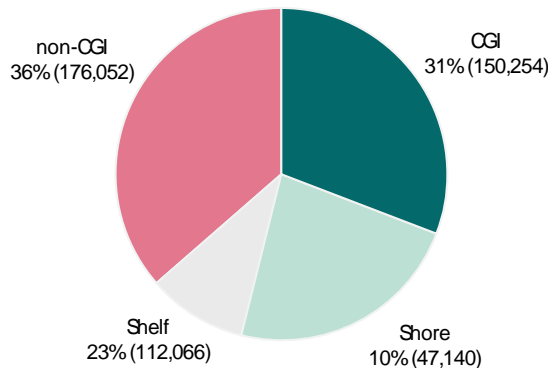# Illumina 450k DNA methylation array

- Informative: ~483,000 CpGs (<2%) but many annotations

- Affordable: ~200 euro per sample

- Fast: >1000 samples a week
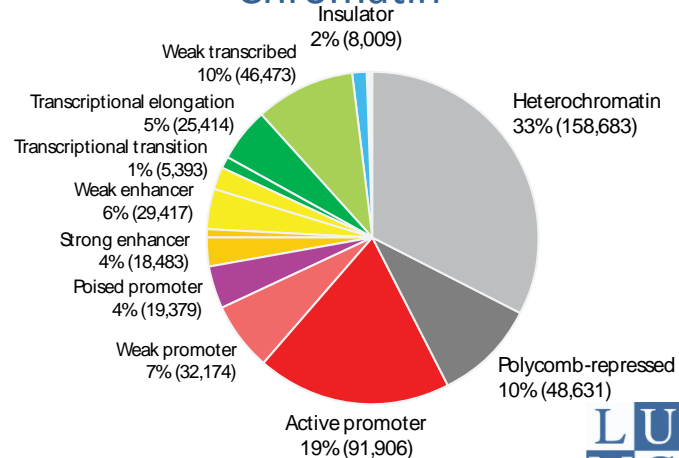
- Data: reasy to manage (amount and complexity)

Genic

- Downstream 2% (9514)
- Intergenic 16% (77,934)
- Distal promoter 5% (23,768)
- Gene body 39% (188,765)
- Proximal promoter 38% (185,531)

CpG island

- non-CGI 36% (176,052)
- CGI 31% (150,254)
- Shelf 23% (112,066)
- Shore 10% (47,140)

Chromatin

- Insulator 2% (8,009)
- Weak transcribed 10% (46,473)
- Transcriptional elongation 5% (25,414)
- Transcriptional transition 1% (5,393)
- Weak enhancer 6% (29,417)
- Strong enhancer 4% (18,483)
- Poised promoter 4% (19,379)
- Weak promoter 7% (32,174)
- Active promoter 19% (91,906)
- Heterochromatin 33% (158,683)
- Polycomb-repressed 10% (48,631)

LUMC

Principle methylation array

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

...CCTGGTA**CG**TCTAGC**CG**TAATTAGCT**CG**ATTC**CG**GT...

Bisulfite conversion

unmethylated   methylated

...UUTGGT**UG**TUTAG...   ...UUTGGT**CG**TUTAG...

unmeth signal   meth signal

β-value = meth / (meth + unmeth)

Possible DNA methylation levels of one CpG site

M?

...CCTGGTA**CG**TCTAG...

...GGACCAT**GC**AGATC...

M?

- A DNA molecule?
- A cell?

Possible DNA methylation levels of one CpG site

```
              M?
              |
...CCTGGTACGTCTAG...
...GGACCATGCAGATC...
              |
              M?
```

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

LUMC

M?

...CCTGGTA**CG**TCTAG...
...GGACCATG**C**AGATC...

M?

M?

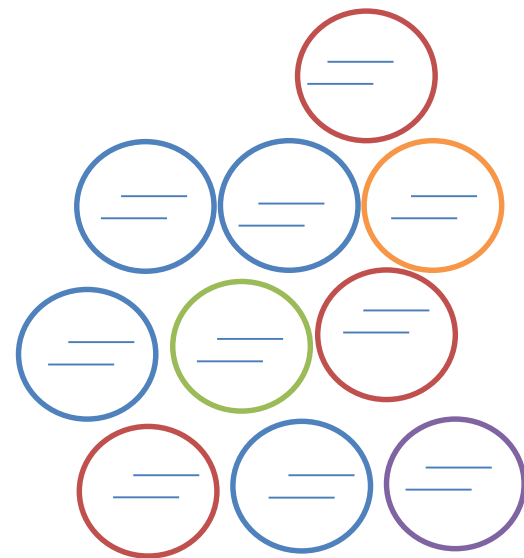...CCTGGTA**CG**TCTAG...
...GGACCATG**C**AGATC...

M?

M?

...CCTGGTA**CG**TCTAG...
...GGACCATG**C**AGATC...

M?

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