Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L - 2022 | week 10

Pierre-Luc Germain / Emanuel Sonder

ETH Zürich

Today: DNA Methylation

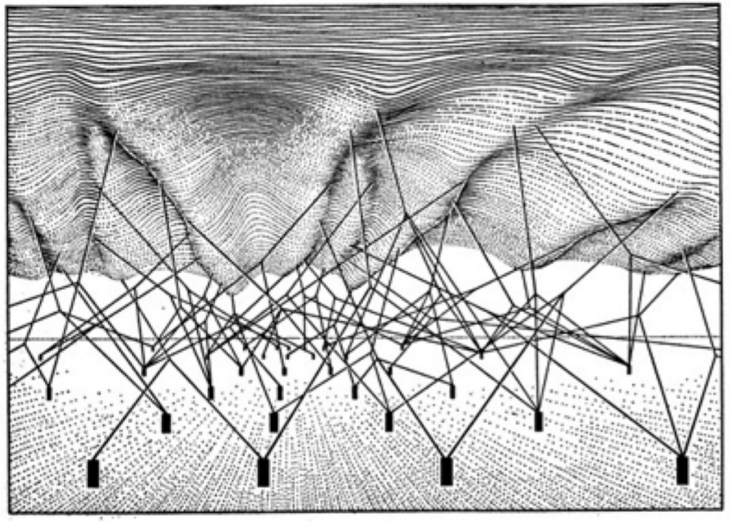


FIGURE 5

The complex system of interactions underlying the epigenetic landscape. The pegs in the ground represent genes; the strings leading from them the chemical tendencies which the genes produce. The modelling of the epigenetic landscape, which slopes down from above one's head towards the distance, is controlled by the pull of these numerous guy-ropes which are ultimately anchored to the genes.

Waddington, 1957

so far we looked at:

- histone marks (chIP-seq)
  - H3K27 acetylation
  - H3K27 methylation
  - H3K9 methylation
  - etc.
- chromatin accessibility (ATAC-seq)

Today: DNA Methylation

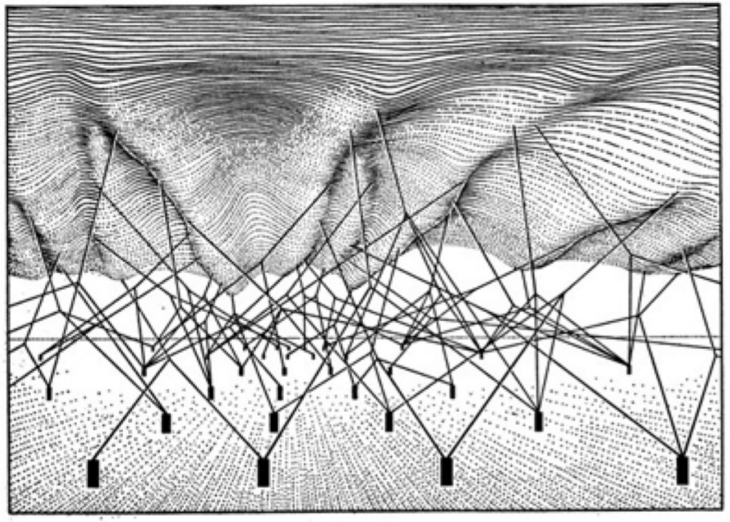


FIGURE 5

The complex system of interactions underlying the epigenetic landscape. The pegs in the ground represent genes; the strings leading from them the chemical tendencies which the genes produce. The modelling of the epigenetic landscape, which slopes down from above one's head towards the distance, is controlled by the pull of these numerous guy-ropes which are ultimately anchored to the genes.

Waddington, 1957

so far we looked at:

- histone marks (chIP-seq)
  - H3K27 acetylation
  - H3K27 methylation
  - H3K9 methylation
  - etc.
- chromatin accessibility (ATAC-seq)

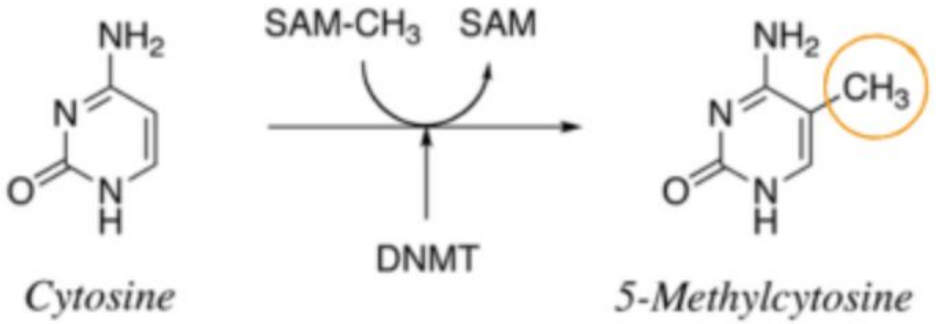
=> Missing piece: DNA methylation

Plan for Today

- Biology of DNA Methylation:
  - Contexts
  - Activation vs repression
- Profiling Techniques
- Practical
  - visualization
  - bsseq objects
  - enrichment

DNA Methylation context

CpG Dinucleotides:

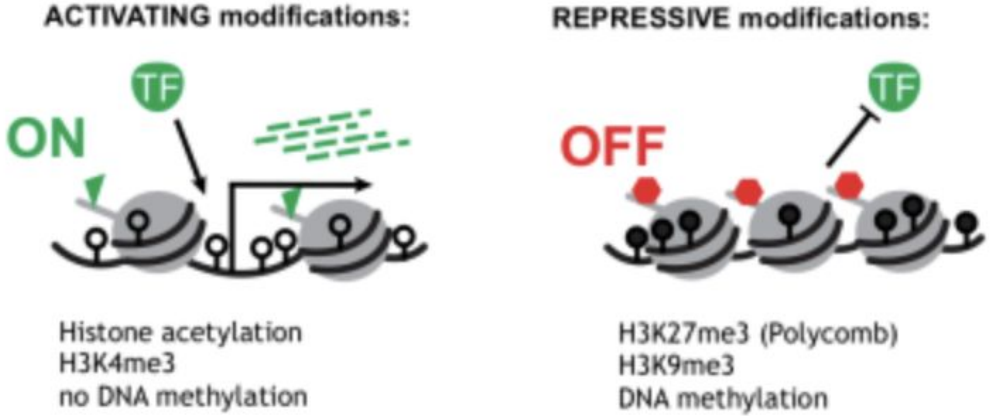


Scheme from Izaskun Mallona/Tuncay Baubec

also in other contexts such as CHG, CHH…

DNA Methylation: Activation vs Repression

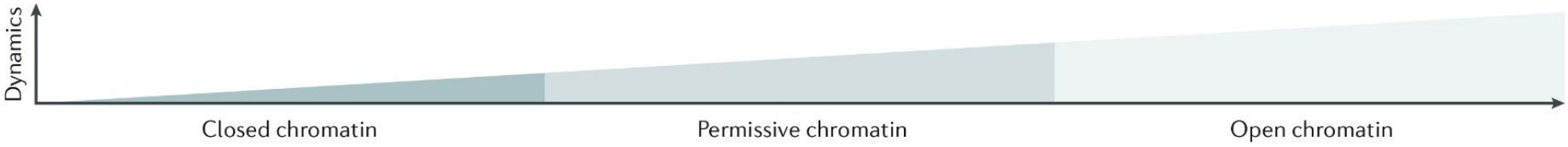
Repressive role of DNA methylation in transcription



Scheme from Izaskun Mallona/Tuncay Baubec

Correlation between DNA methylation and gene silencing

There are degrees of accessibility (activation/repression)



Heterochromatin (stable repression)

Very low accessibility

HP1 binding

H3K9 methylation

?

Poised/silenced

Low accessibility

H3K27 methylation

Active elements

Very high accessibility

RNApol2 binding

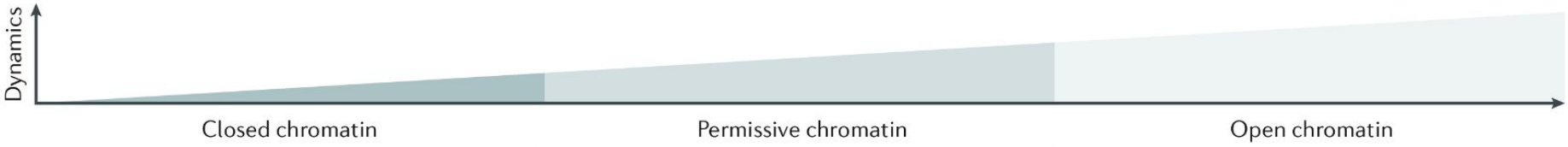
H3K4 methylation

H3K27 acetylation

Fate transitions (e.g. differentiation)

Transient responses

There are degrees of accessibility (activation/repression)



Heterochromatin (stable repression)

Very low accessibility

HP1 binding

DNA methylation

H3K9 methylation

?

Poised/silenced

Low accessibility

H3K27 methylation

Active elements

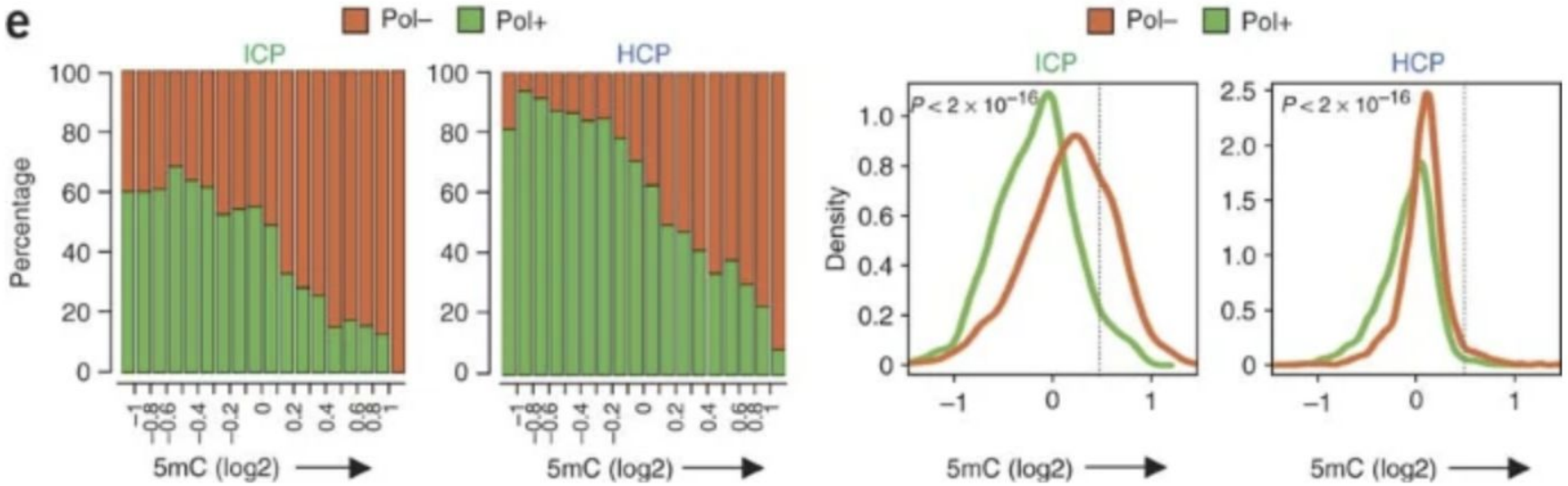
Very high accessibility

RNApol2 binding
H3K4 methylation
H3K27 acetylation

Fate transitions
(e.g. differentiation)

Transient responses

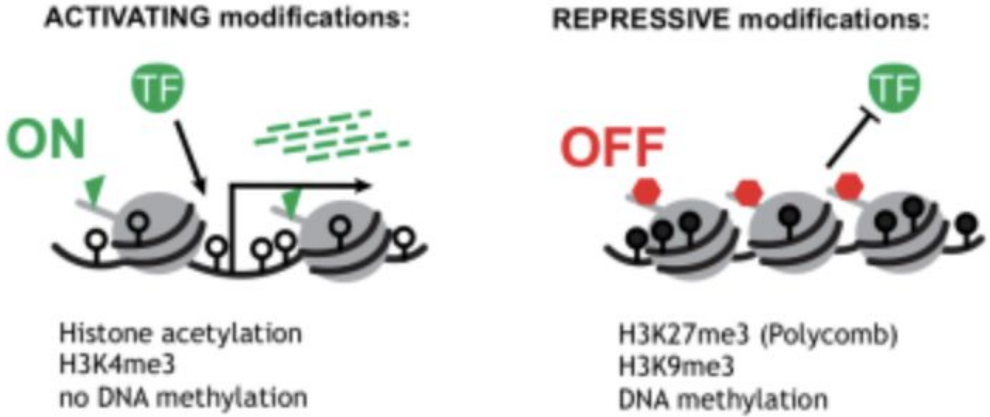
at promoters



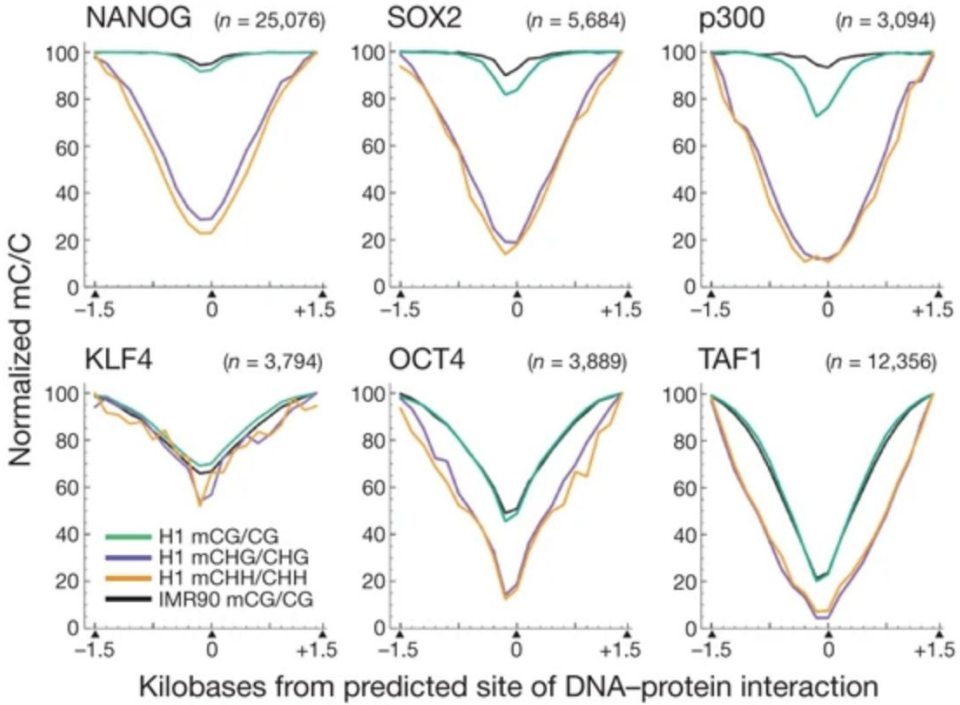
Weber et al., 2007

at promoters

“However, how this leads to transcription inhibition is still not entirely resolved, as the methyl mark per se does not seem to confer silencing. Regions of accessible chromatin are frequently lowly methylated or unmethylated, indicating that binding of transcription factors and DNA methylation are mutually exclusive”. (Greenberg et al. 2019)

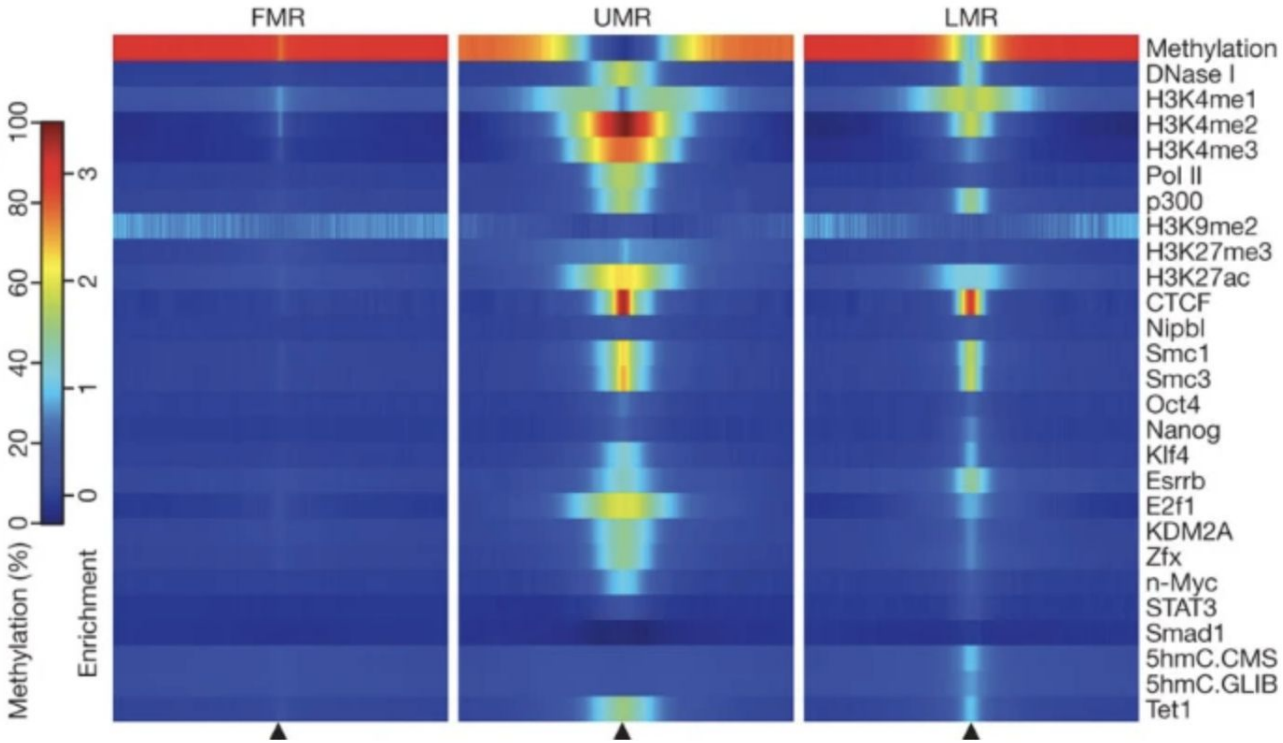


TF Binding & Methylation



Lister et al., 2009

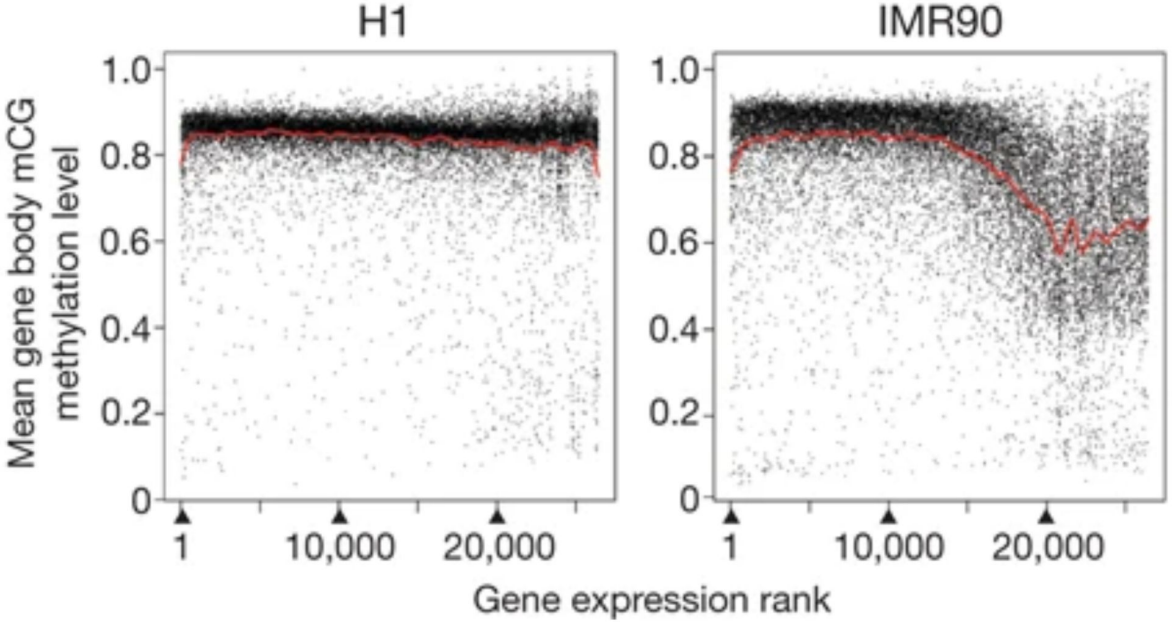
TF Binding & Methylation



Stadler et al., 2011

at gene bodies

“The enrichment of DNA methylation in gene bodies presents a paradox: on the one hand, gene-body methylation is highly conserved across eukaryotes — more than it is conserved at transposable elements, for example — indicating it has an important function. On the other hand, DNA methylation is mutagenic, so why is it so prominent in coding sequences” (Greenberg et al. 2019)

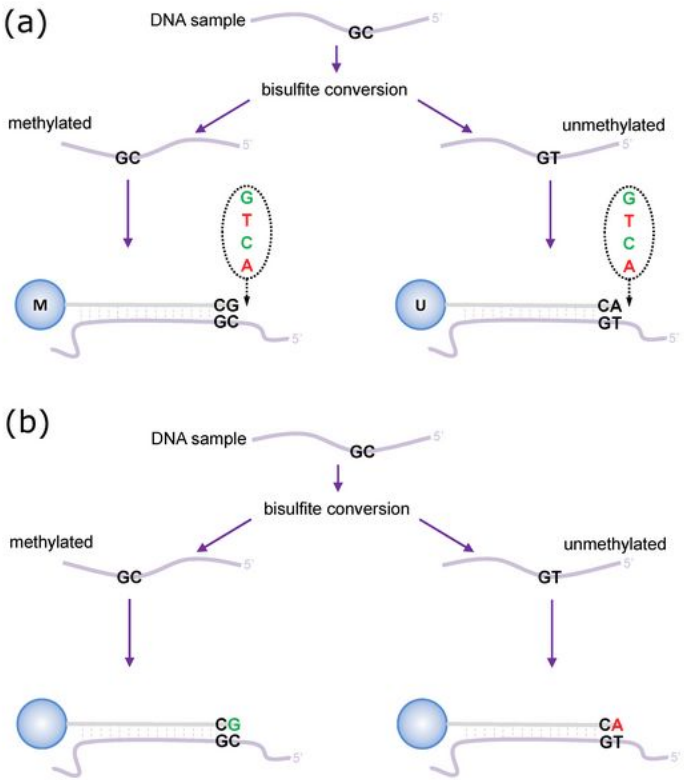


Lister et al., 2009

Profiling techniques

- DNA Methylation Arrays:
  - Illumina 450K/850K
- Bisulfite sequencing:
  - RRBS
  - WGBS
  - scWGBS

Methylation Arrays



- fluorescence signal
- cheap(er)
- again Illumina:
  - Infinium 450K & 850K probing wide ranges of the genome
  - two types of assays: Infinium I (a), Infinium II (b)
  - divergence between the β-values retrieved from the type I and type II Infinium assays. Specifically, the β-values obtained from Infinium II probes were less accurate and reproducible than those obtained from Infinium I probes. (Dedeurwaerder et al., 2011)
- Analysis:
  - A cross-package Bioconductor workflow for analysing methylation array data.

Maksimovic et al., 2016

Beta & M-Values

- Beta-value, ratio of the methylated probe intensity and the overall intensity:

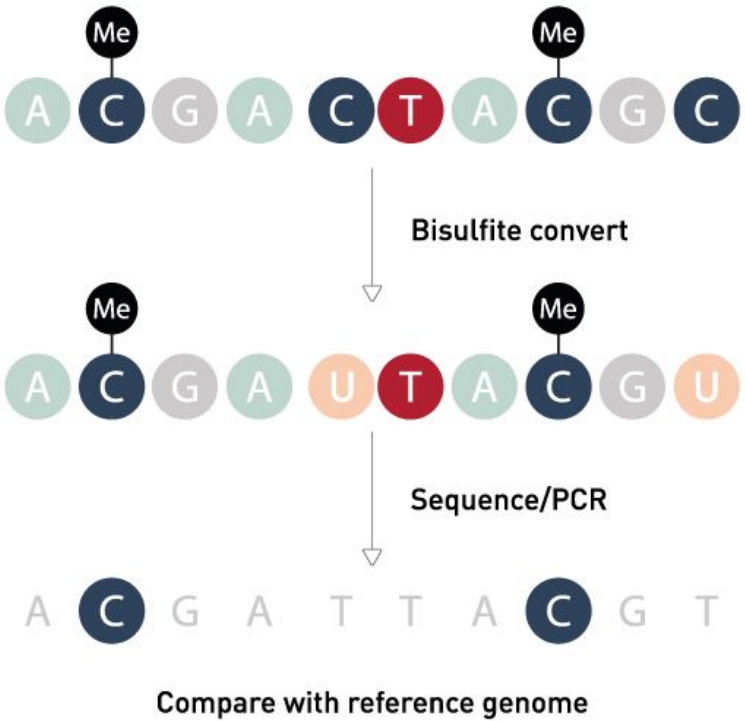
$$Beta_i = \frac{\max(y_{i,methy}, 0)}{\max(y_{i,unmethy}, 0) + \max(y_{i,methy}, 0) + \alpha}$$

- M-value, log2 ratio of the intensities of methylated probe versus unmethylated probe:

$$M_i = \log_2 \left(\frac{\max(y_{i,methy}, 0) + \alpha}{\max(y_{i,unmethy}, 0) + \alpha} \right)$$

- The Beta-value has a more intuitive biological interpretation, but the M-value is more statistically valid for the differential analysis of methylation levels. (Du et al., 2010)

Bisulfite Sequencing



https://www.diagenode.com/en/categories/bisulfite-conversion

Exists in different flavors:

- RRBS
- WGBS
- scWGBS

Bisulfite Sequencing - Aligning

Bismark

CpG-report file:

	chr	pos	strand	met	unmet	context	tri-context
1:	5	11823	+	0	0	CG	CGA
2:	5	11824	-	0	0	CG	CGG
3:	5	11850	+	0	0	CG	CGA
4:	5	11851	-	0	0	CG	CGG
5:	5	11856	+	0	0	CG	CGA

In R: bsseq package:

`read.bismark()`

Bisulfite Sequencing - Aligning

Bismark

CpG-report file:

	chr	pos	strand	met	unmet	c	
1:	5	11823	+	0	0		
2:	5	11824	-	0	0	CG	CGG
3:	5	11850	+	0	0	CG	CGA
4:	5	11851	-	0	0	CG	CGG
5:	5	11856	+	0	0	CG	CGA

met	unmet
0	0
20	20
1	2
10	20

In R: bsseq package:

`read.bismark()`

DNA-Methylation Analysis tasks:

- Smoothing:
  - bsseq
- Differential Methylation testing:
  - DMRcate (defines DM regions itself)
  - edgeR (allows for testing user-defined regions, e.g. genes of interest)
  - Many more (Methylkit, bsseq etc.)
- Calling partially methylated domains (PMDs)
  - MethylSeekR
- Quantifying DNA methylation heterogeneity
  - Quantitative comparison of within-sample heterogeneity scores for DNA methylation (Scherer et al., 2020)

hw

1. Plot the DM Regions called in the practical (dmr.rds) using the provided bsseq object (bsseqEx.rds), a) without smoothing b) with smoothing (bsseq)

   Hint getMeth(..., regions=...)

2. Use the rGREAT enrichment analysis from last week with the DM Regions. Phrase what is your hypothesis for the analysis.

References

- Weber, M., Hellmann, I., Stadler, M. et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nat Genet 39, 457–466 (2007). https://doi.org/10.1038/ng1990
- Lister, R., Pelizzola, M., Dowen, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences . *Nature* **462,** 315–322 (2009). https://doi.org/10.1038/nature08514
- Stadler, M., Murr, R., Burger, L. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480,** 490–495 (2011). https://doi.org/10.1038/nature10716
- Maksimovic J, Phipson B, Oshlack A. A cross-package Bioconductor workflow for analysing methylation array data. *F1000Res*. 2016;5:1281. Published 2016 Jun 8. doi:10.12688/f1000research.8839.3
- Dedeurwaerder S, Defrance M, Calonne E, Denis H, Sotiriou C, Fuks F. Evaluation of the Infinium Methylation 450K technology. Epigenomics. 2011 Dec;3(6):771-84. doi: 10.2217/epi.11.105. PMID: 22126295.
- Du, P., Zhang, X., Huang, CC. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11,** 587 (2010). https://doi.org/10.1186/1471-2105-11-587
- Greenberg, M.V.C., Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat Rev Mol Cell Biol* **20,** 590–607 (2019). https://doi.org/10.1038/s41580-019-0159-6