Bioinformatic approaches to regulatory genomics and epigenomics

376-1347-00L | week 11

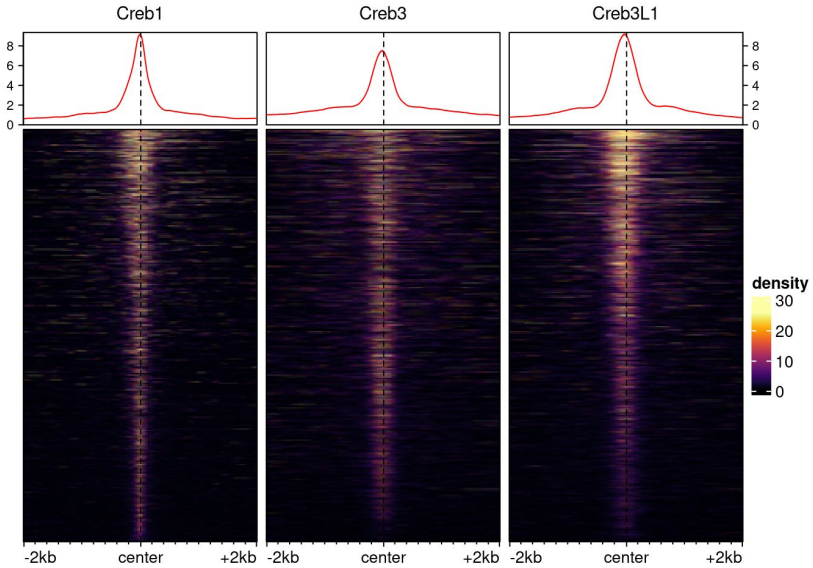
Emanuel Sonder / Pierre-Luc Germain

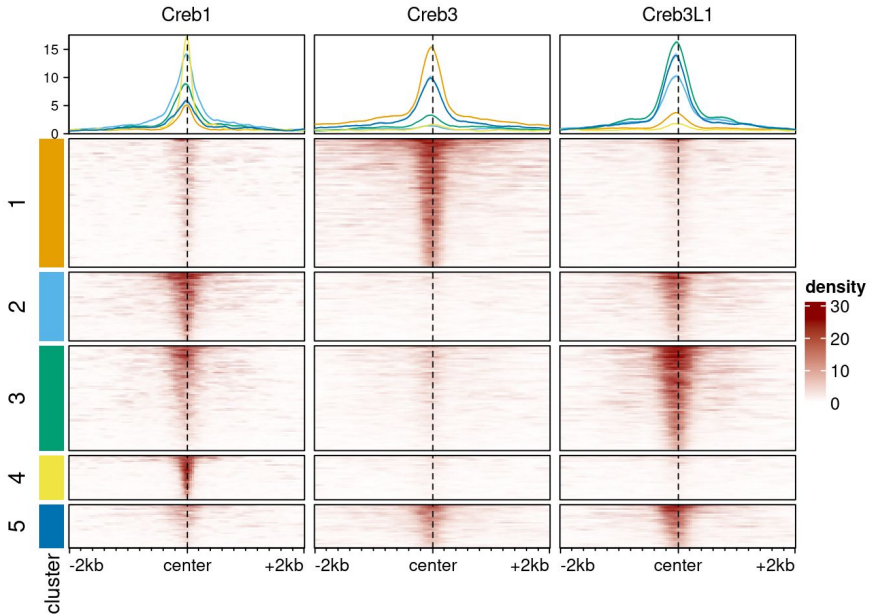
ETH Zürich

Plan for Today

- Debriefing on the assignment
- Some biology behind DNA Methylation
- Practical 1: visualization of DNA methylation
- Profiling Techniques
- Practical 2: statistical analysis

Debriefing on the assignment





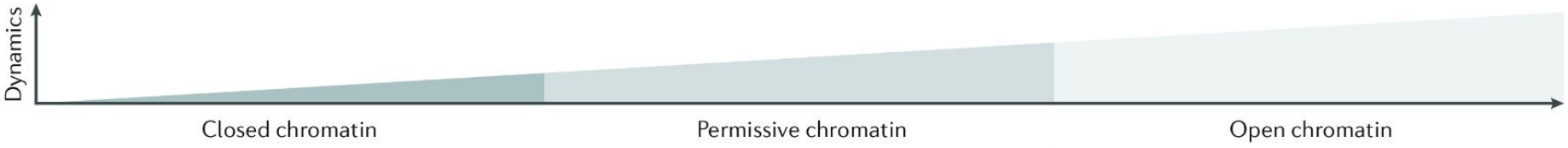
Debriefing on the assignment

Pay attention to the background used in enrichment analysis, and *describe the results accordingly*, e.g.:

If you used the union of all clusters as a background, then you can say that cluster X is enriched in certains terms *in comparison to the binding sites of all CREB family members*.

This tells us how the CREB family members differ from each other, but not necessarily what might be consistent across the family members.

Degrees of activation/repression



Heterochromatin (stable repression)

Very low accessibility

DNA methylation

HP1 binding

H3K9 methylation

Poised/silenced

Low accessibility

H3K27 methylation

Active elements

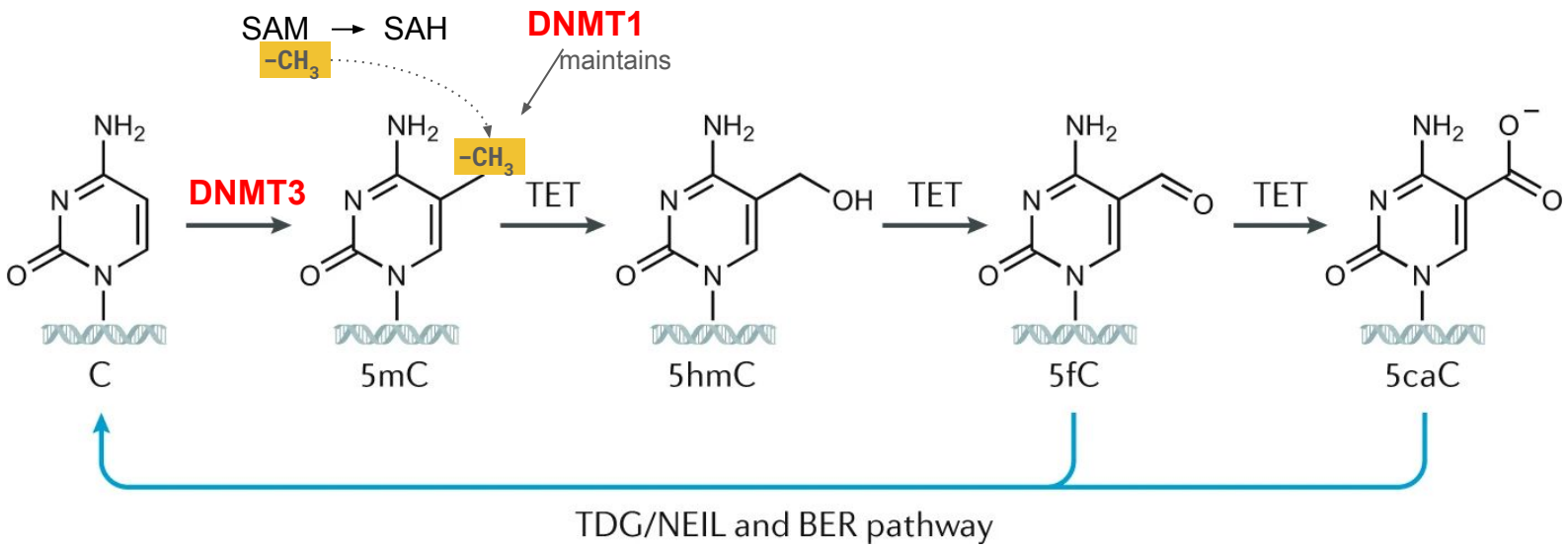
Very high accessibility

RNApol2 binding
H3K4 methylation
H3K27 acetylation

Fate transitions
(e.g. differentiation)

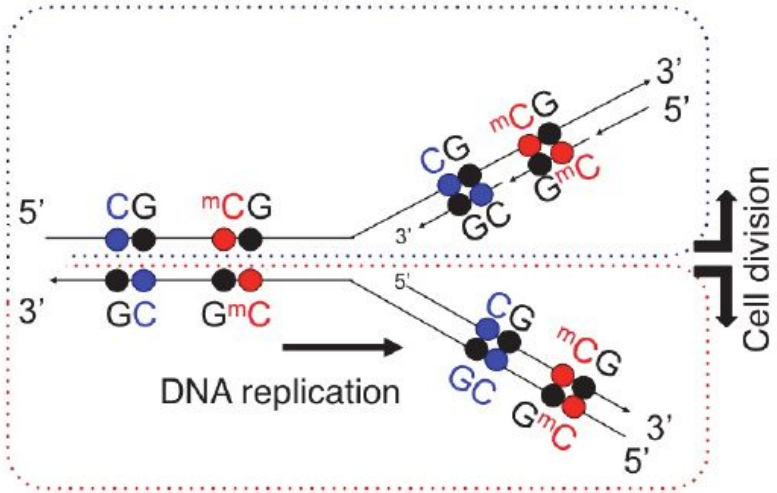
Transient responses

What is DNA methylation?



(Modified from Parry, Rulands and Reik, *Nat Rev Gen* 2021)

Methylation context - why all this talk about CpGs ?



(Adapted from Kubota et al., 2011)

At C-G pairs, DNMT1 ensures that the methylation is copied to the newly synthesized DNA strands

While methylation does occur at non-CpGs, it normally gets diluted out during replication

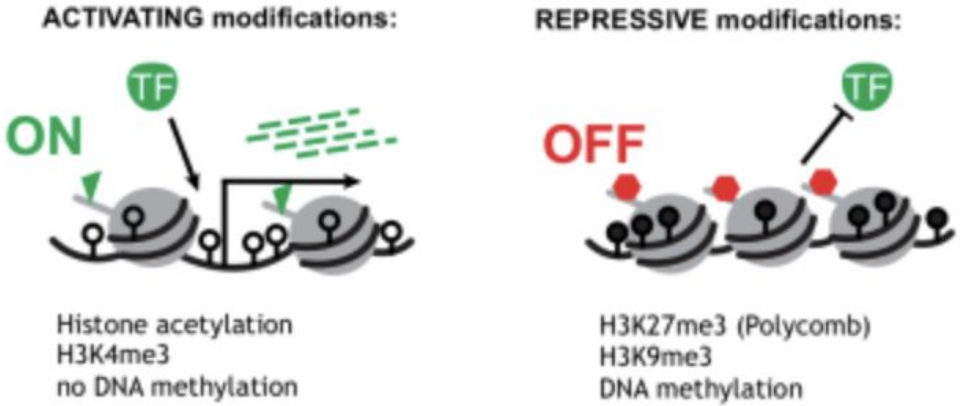
Hence (in mammals) most relevant DNA methylation happens at CpGs.

<u>Some exceptions:</u>

- In embryonic stem cells
- In post-mitotic neurons
- In cancer

DNA Methylation: Activation vs Repression

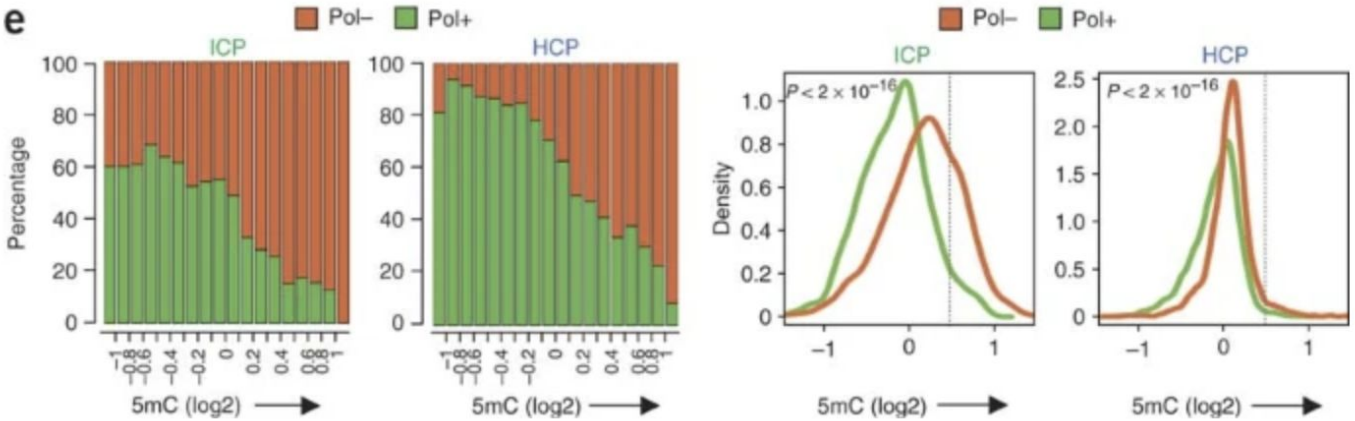
Repressive role of DNA methylation in transcription



Scheme from Izaskun Mallona/Tuncay Baubec

Correlation between DNA methylation and gene silencing

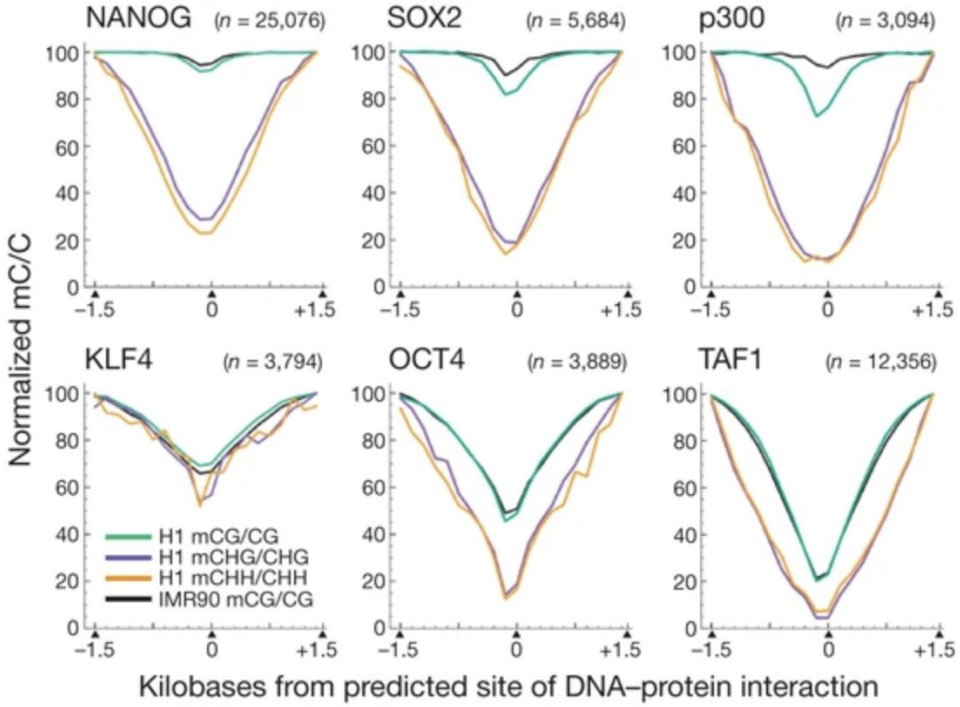
At promoters, DNA methylation has long been associated to repression



(adapted from Weber et al., 2007)

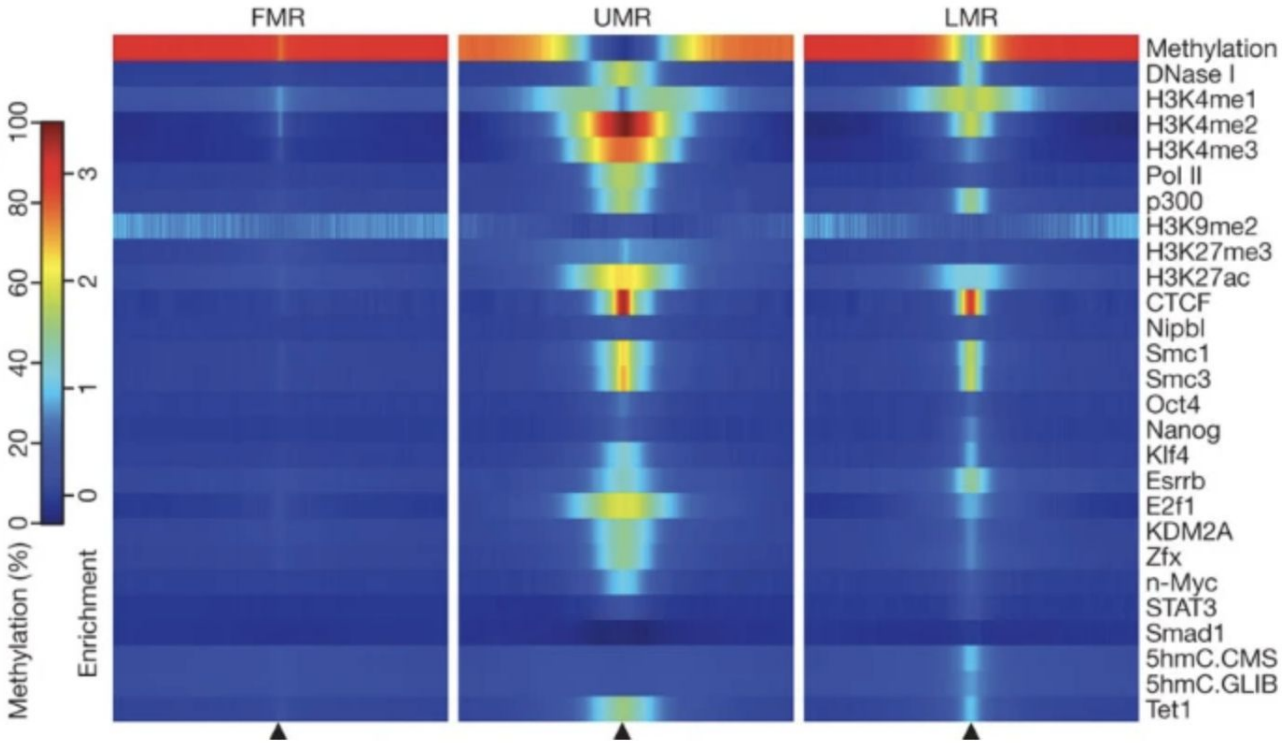
We now know that this repression is mostly happening because the methylation inhibits the binding of transcription factors (TFs) (Kaluscha et al., Nat Gen 2022)

TF Binding & Methylation



(adapted from Lister et al., 2009)

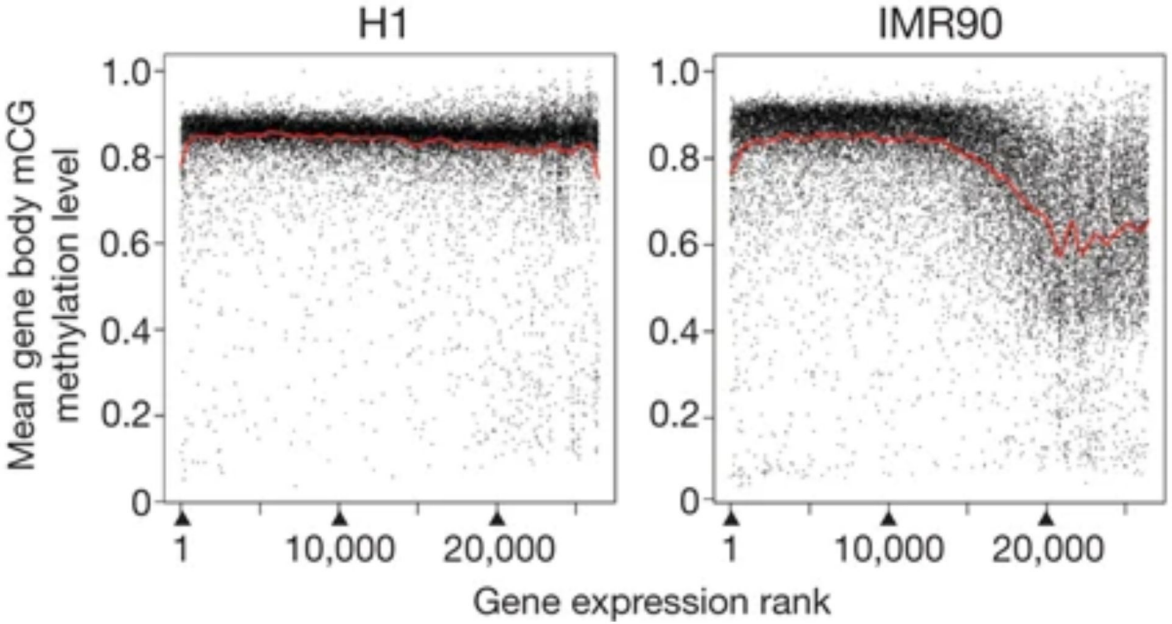
TF Binding & Methylation



(Adapted from Stadler et al., 2011)

DNA methylation of gene bodies

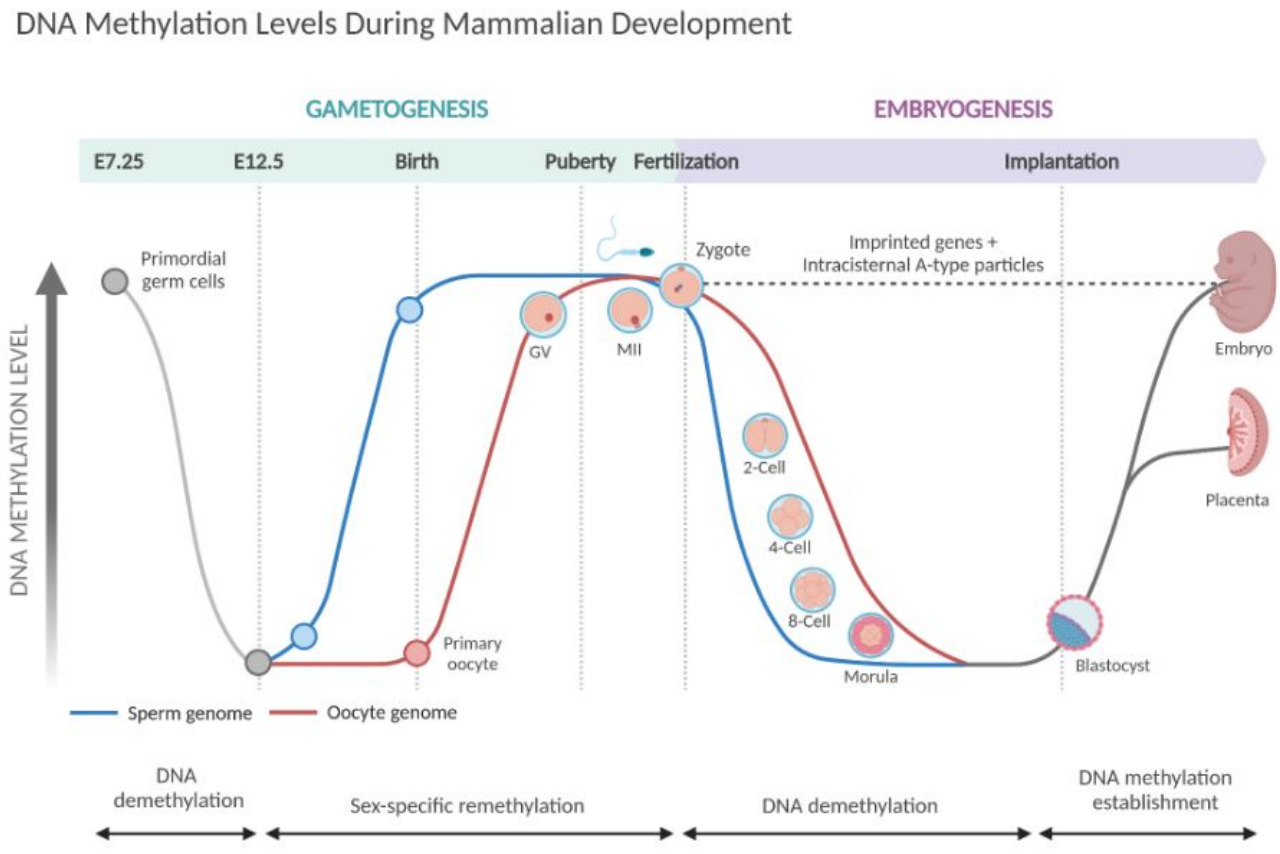
"The enrichment of DNA methylation in gene bodies presents a paradox: on the one hand, gene-body methylation is highly conserved across eukaryotes — more than it is conserved at transposable elements, for example — indicating it has an important function. On the other hand, DNA methylation is mutagenic,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



Differentiated cell types typically show a positive correlation between gene expression and gene body methylation

Lister et al., 2009

Germline DNA methylation reprogramming



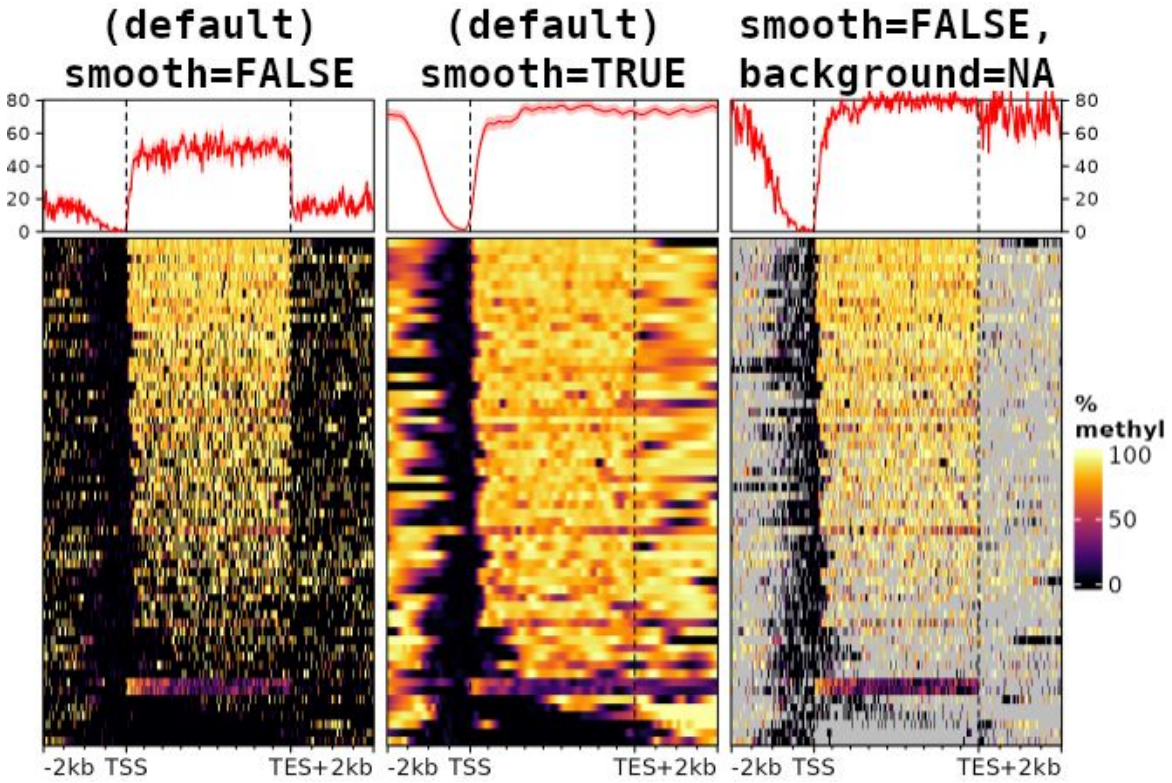
(BioRender)

Practical - Part 1

Visualization

Visualization of DNAme data in epiwraps

DNAme differs from other signals in that there can be an absence of signal not simply because the modification is absent, but because there aren't any C's to be methylated…



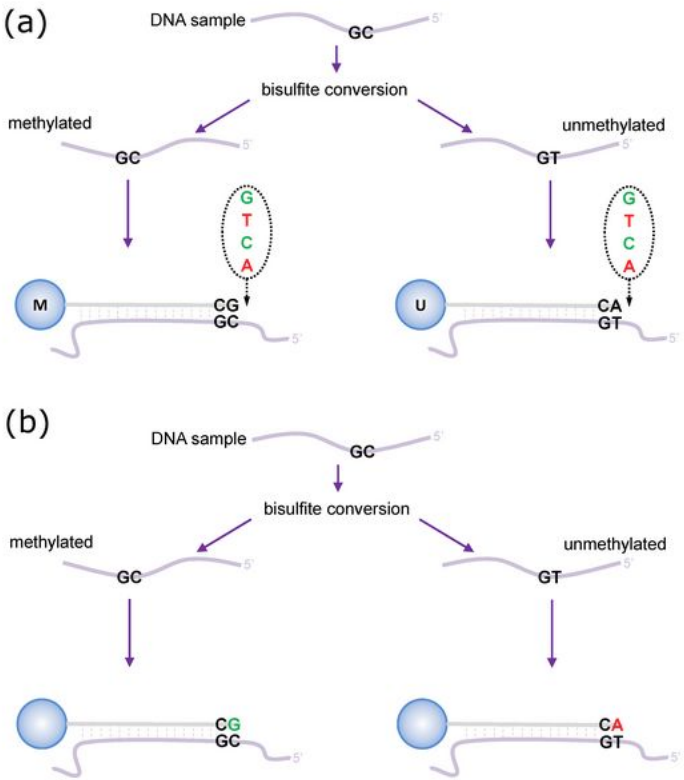
Gene bodies in a cancer cell line

See the [epiwraps documentation]

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

- 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:

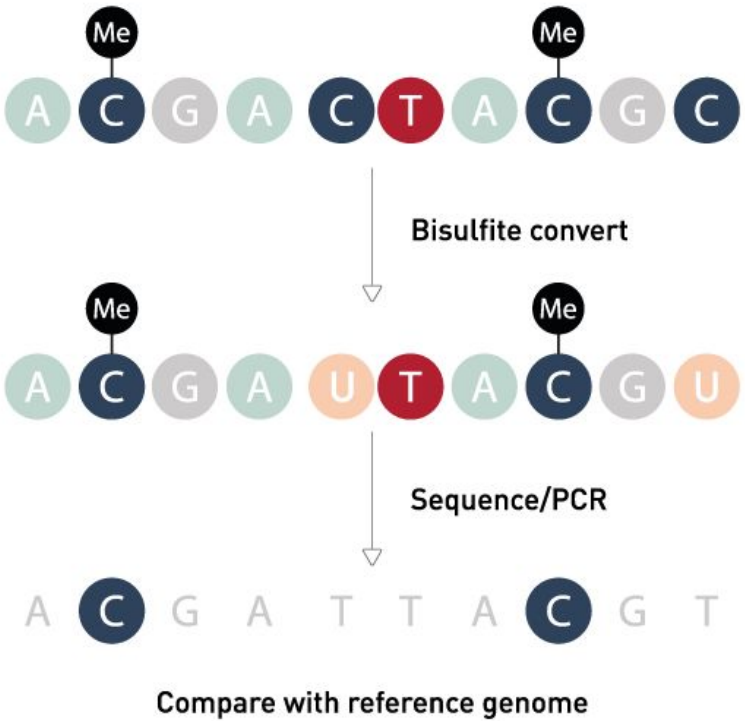
$$Beta_i = \frac{\max(y_{i,methy}, 0)}{\max(y_{i,unmethy}, 0) + \max(y_{i,methy}, 0) + \alpha}$$

- 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	→ 50%
10	20	→ 50%

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)

Practical - Part 2

bsseq objects + Differential Methylation Testing

Assignment

Use the set of differentially-methylated regions (DMRs) we between ESC and astrocytes that we established during the practical. You can download them here

- Plot a heatmap of the methylation levels of the genes in top 20 DMRs located on chr1. You may either do this as we did for promoters during the practical (i.e. using `getMeth()`), or using `signal2Matrix()` and `plotEnrichedHeatmaps()`.

- Run an enrichment analysis (rGREAT) of the significant DMRs from the practical against the entire genome. Describe what your enrichment analysis is testing exactly & describe the results

Save your assignment in a R markdown named assignment.Rmd, render it, and push both the Rmd and html file to your github repository

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
- Kaluscha, S., Domcke, S., Wirbelauer, C. *et al.* Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat Genet* 54, 1895−1906 (2022). https://doi.org/10.1038/s41588-022-01241-6