CS-E5875 High-Throughput Bioinformatics
DNA methylation analysis

Harri Lähdesmäki
(Acknowledgement for slides: Essi Laajala)

Department of Computer Science
Aalto University

November 14, 2023

Contents

- DNA methylation
- Bisulfite sequencing (BS-seq) protocol
- Alignment and quantification of BS-seq data
- Statistical analysis of BS-seq data
- Other details

DNA methylation: what?

- Epigenetic changes are reversible modifications on DNA, or “on top of DNA”, which do not change the DNA sequence itself
- DNA methylation is an epigenetic modification where methyl group is added to the 5 position of a cytosine in DNA
- Methyl group is added enzymatically by DNA methyl transferases (DNMT)
- By far the most extensively studied epigenetic modification on DNA

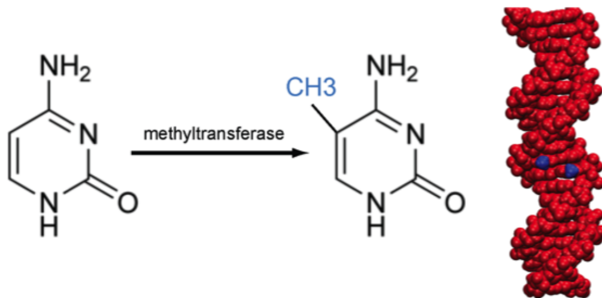


Figure from http://www.ks.uiuc.edu/Research/methylation/

DNA methylation: where?

- In mammaling genomes, DNA methylation primarily occurs in the context of CpG dinucleotides
- Non-CpG methylation found e.g. in stem cells and brain
- CpGs occur with a smaller frequency than expected
  - Human genome GC content is 42%
  - CpGs are expected to occur 4.41% of the time
  - The frequency of CpG dinucleotides is 1%
  - Methylated CpGs are prone to spontaneous deamination to thymines

Figure from (Schubeler, 2009)

DNA methylation: how?

- Two general classes of enzymatic methylation activities
  - De novo methylation (mainly) by DNMT3
  - Maintenance methylation during cell division (mainly) by DNMT1

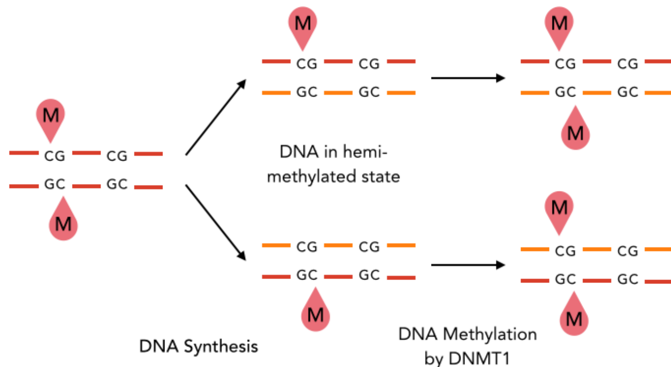


Figure from `http://2014.igem.org/Team:Heidelberg/Project/PCR_2.0`

DNA methylation in gene regulation

- CpG islands (C+G dense $\gtrsim$500 long regions) are present in the 5' regulatory regions of many genes (5' = "beginning")
- Hypermethylation (=overmethylation) of CpG islands near gene promoters contributes to transcriptional silencing by
  - Affecting binding of transcription factors (DNA binding protein that regulate gene transcription)
  - Binding proteins with methyl-CpG-binding domains (MBDs), and recruiting e.g. histone deacetylases and other chromatin remodellers

DNA methylation in gene regulation

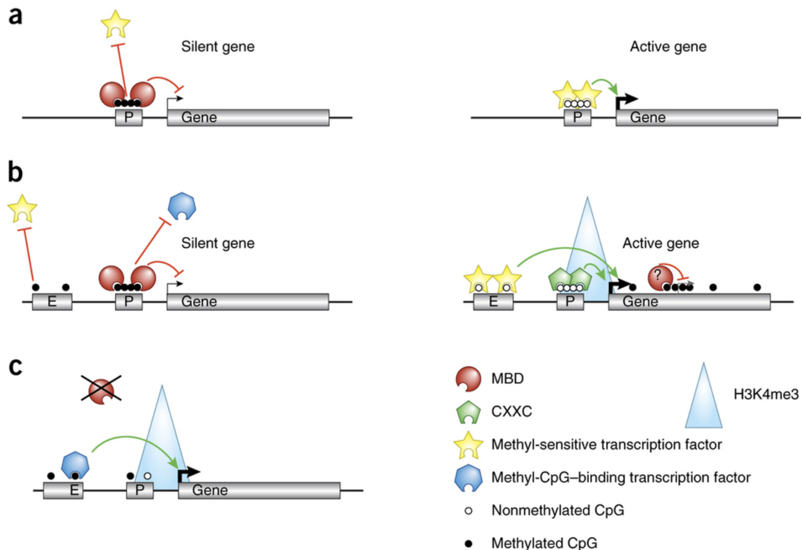


Figure from (Spruijt & Vermeulen, 2014)

DNA methylation in health and disease

- DNA methylation is necessary for normal development, e.g. in
  - Cell differentiation
  - Protection of DNA from transposable elements
  - Genomic imprinting
  - X chromosome inactivation
- Environmental and lifestyle factors may impact DNA methylation, even in-utero
- DNA methylation patterns have been associated with several diseases

Heritability of DNA methylation patterns

- DNA methylation patterns are not directly inheritable
  - Paternal and maternal DNA methylation patterns are almost completely erased during first cell divisions after fertilization in mammals (epigenetic reprogramming)
- However, genomic variants can influence DNA methylation
  - So-called methylation quantitative trait loci (mQTL) effects
- DNA methylation patterns correlate between family members to some extent
  - For example in genome-wide human leukocyte DNA methylation profiles, correlation coefficients between 0.24 and 0.30 have been reported between first-degree relatives, compared to 0.03 between unrelated individuals (Tremblay et al. 2016)
  - These seem to be explained by both genetic inheritance and environmental factors

DNA demethylation

- Until recently, it was believed that methylated DNA can be unmethylated only by dilution during cell differentiation/DNA replication, i.e., incomplete DNA maintenance methylation
- Recently, TET family proteins were shown to be dioxygenases that converted 5mC to 5hmC, 5fC and 5caC, which can be further converted back to unmethylated C
- TETs thus contribute to active demethylation, but 5hmC, 5fC and 5caC can also have multiple functions

DNA demethylation

Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1

Mamta Tahiliani,[1] Kian Peng Koh,[1] Yinghua Shen,[2] William A. Pastor,[1] Hozefa Bandukwala,[1] Yevgeny Brudno,[2] Suneet Agarwal,[3] Lakshminarayan M. Iyer,[4] David R. Liu,[2]* L. Aravind,[4]* Anjana Rao[1]*

15 MAY 2009 VOL 324 SCIENCE www.sciencemag.org

a

Unknown decarboxylase?

TDG and BER

DNMT enzymes?

NH_2 / N / O / N / R — Cytosine

DNMT

NH_2 / N / O / N / R — 5mC

TET protein

NH_2 / OH / N / O / N / R — 5hmC

TET protein

NH_2 / O / N / O / N / R — 5fC

TET protein

NH_2 / O / OH / N / O / N / R — 5caC

AID or APOBEC?

O / OH / HN / O / N / R — 5hmU

TDG or SMUG1 and BER

Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine

Shinsuke Ito,[1,2]* Li Shen,[1,2]* Qing Dai,[3] Susan C. Wu,[1,2] Leonard B. Collins,[4] James A. Swenberg,[2,4] Chuan He,[3] Yi Zhang[1,2]†

2 SEPTEMBER 2011 VOL 333 SCIENCE www.sciencemag.org

BER := base excision repair
TDG := thymine DNA glycosylase
AID := activation-induced deaminase
APOBEC := apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like

Contents

- DNA methylation
- Bisulfite sequencing (BS-seq) protocol
- Bisulfite sequencing data analysis
- Alignment and quantification of BS-seq data
- Statistical analysis of BS-seq data
- Other details

Bisulfite sequencing (BS-seq) protocol

- Bisulfite treatment of genomic DNA converts unmethylated cytosines to uracils which are read as thymine during sequencing
- Methylated (and hydroxymethylated) cytosines are resistant to the conversion and are read as cytosine

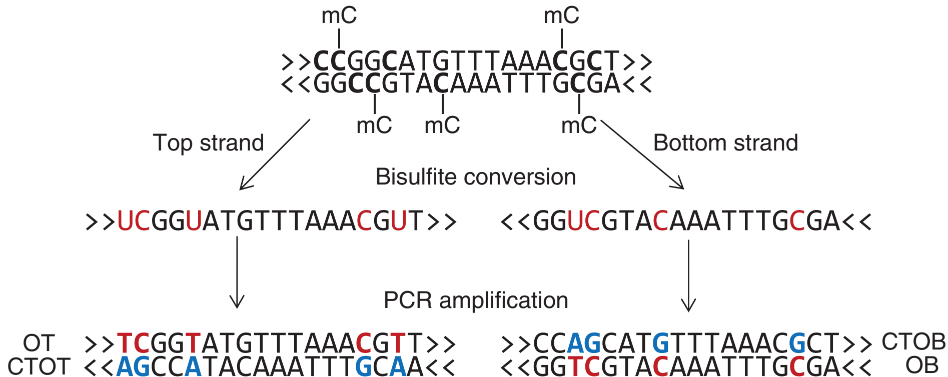
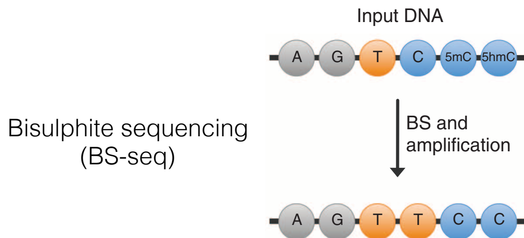


Figure from (Krueger et al, 2012)

Bisulfite sequencing (BS-seq) protocol

- Bisulfite treatment of genomic DNA converts unmethylated cytosines to uracils which are read as thymine during sequencing
- Methylated (and hydroxymethylated) cytosines are resistant to the conversion and are read as cytosine



Cytosine modification

	C	5mC	5hmC
BS-seq	T	C	C

Figure from (Booth et al, 2012)

Reduced representation BS-seq (RRBS-seq)

- BS-seq provides an accurate map of methylation state at single nucleotide resolution
- Whole genome analysis is expensive since only about 1% of the human genome contains CpGs
- Reduced representation BS-seq (RRBS-seq) uses restriction enzymes prior to bisulfite sequencing focus on CpG regions
  - MspI digests genomic DNA in a methylation-insensitive manner
  - MspI targets 5'CCGG3' sequences and cleaves the phosphodiester bonds upstream of CpG dinucleotide.
  - → Each fragment will have a CpG at each end
- RRBS-seq will cover majority of promoters and GC rich regions

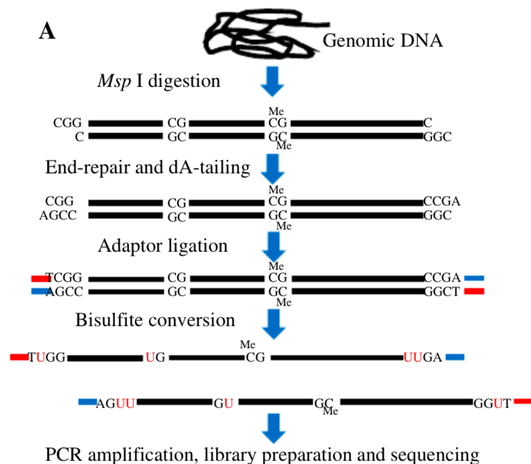


Figure from (Lianga et al, 2014)

Contents

- DNA methylation
- Bisulfite sequencing (BS-seq) protocol
- Alignment and quantification of BS-seq data
- Statistical analysis of BS-seq data
- Other details

RRBS data analysis workflow example

1. RRBS data preprocessing

- TrimGalore (trimming of low-quality bases, end repair biases, adapter sequences, and overlaps between paired-end reads), fastQC before and after TrimGalore
- Bismark-alignment
- Bismark methylation extractor to extract numbers of methylated and unmethylated reads at each CpG site in each sample
- Excluding of samples that have low bisulfite conversion efficiency (estimated as proportion of Cs converted to Ts in the completely unmethylated lambda phage genome)
- Trimming of M-biases

2. SNP detection (e.g. bis-SNP or bs-SNPer)

- SNPs removed from each sample

3. Coverage filtering

- CpG sites with extremely high coverages within each sample removed to avoid PCR duplicates
- Low-coverage CpG sites, e.g. with coverage < 10 reads in more than 50 % of the samples removed

4. Differential methylation analysis

- E.g. beta-binomial models are suitable for read count data:

$$m \sim \mathrm{Bin}(p,n)$$
$$p \sim \mathrm{Beta}(\alpha, \beta)$$

5. Interpretation

- Annotation to genomic parts, nearest genes, regulatory regions
- Known mQTL and eQTM effects

RRBS data analysis workflow example

1. RRBS data preprocessing
- TrimGalore (trimming of low-quality bases, end repair biases, adapter sequences, and overlaps between paired-end reads), fastQC before and after TrimGalore
- Bismark-alignment
- Bismark methylation extractor to extract numbers of methylated and unmethylated reads at each CpG site in each sample
- Excluding of samples that have low bisulfite conversion efficiency (estimated as proportion of Cs converted to Ts in the completely unmethylated lambda phage genome)
- Trimming of M-biases

2. SNP detection (e.g. bis-SNP or bs-SNPer)
- SNPs removed from each sample

3. Coverage filtering
- CpG sites with extremely high coverages within each sample removed to avoid PCR duplicates
- Low-coverage CpG sites, e.g. with coverage < 10 reads in more than 50 % of the samples removed

4. Differential methylation analysis
- E.g. beta-binomial models are suitable for read count data:

$$m \sim \text{Bin}(p,n)$$
$$p \sim \text{Beta}(\alpha, \beta)$$

5. Interpretation
- Annotation to genomic parts, nearest genes, regulatory regions
- Known mQTL and eQTM effects

Aligning BS-seq reads

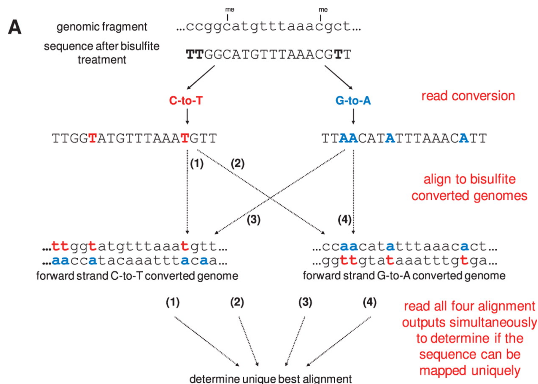
Bisulfite treatment introduces mutations into genomic DNA in a methylation dependent manner

- Alignment of BS-seq reads is more challenging
- Standard alignment methods cannot be used directly

Bismark tool uses the following approach to map BS-seq reads

- Reads from a BS-seq experiment are converted into a C-to-T version and a G-to-A version
- The same conversion for the genome
- Bowtie alignment in the genome that has reduced complexity
- A unique best alignment is determined from four parallel alignment processes (see next page)

Bismark tool



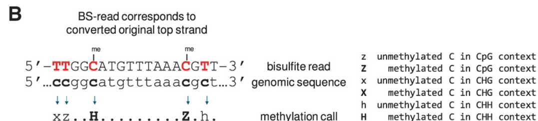


Figure from (Krueger & Andrews, 2011)

Quantifying BS-seq data

- Bismark outputs, among others, one line per read containing useful information
  - Mapping position, alignment strand, the bisulfite read sequence, its equivalent genomic sequence and a methylation call string
- Bismark automatically extracts the methylation information at individual cytosine positions
  - For different sequence contexts (CpG, CHG, CHH; where H can be either A, T or C)
  - Strand-specific or strands merged
- That is, for each cytosine Bismark outputs
  - n_i the number of reads covering the cytosine in sample i
  - m_i the number of methylated readouts (i.e., "C") for the cytosine in sample i
- One way to quantify methylation proportion is

$$\hat{p}_i = \frac{m_i}{n_i} = \frac{\text{the number of C reads overlapping the cytosine}}{\text{the number of C or T reads overlapping the cytosine}}$$

Contents

- DNA methylation
- Bisulfite sequencing (BS-seq) protocol
- Alignment and quantification of BS-seq data
- Statistical analysis of BS-seq data
- Other details

Beta-binomial model

- At the end, one is typically interested in testing a hypothesis, e.g. is there a statistically significant difference in methylation levels between group A and group B
- Some early methods applied e.g. the t-test on the estimated methylation fractions $\hat{p}_i$ (or their logit transformations)
- We will look at RadMeth tool (Dolzhenko and Smith, 2014)
- RadMeth uses the beta-binomial regression model, where beta-binomial is a compound distribution obtained from the binomial by assuming that its probability of success parameter follows a beta distribution

Beta-binomial model

$i = 1, \ldots, s$, where s is the number of biological samples
For each cytosine in the genome we have the following model

- n_i: the number of reads that contain "C" or "T" readout at the cytosine in sample i
- m_i: the number of reads that contain "C" readout (i.e. methylated) at the cytosine in sample i ($0 \leq m_i \leq n_i$)
- If we knew the underlying methylation proportion p_i, then: $M_i \sim \mathrm{Binom}(p_i, n_i)$

Beta-binomial model

$i = 1, \ldots, s$, where s is the number of biological samples
For each cytosine in the genome we have the following model

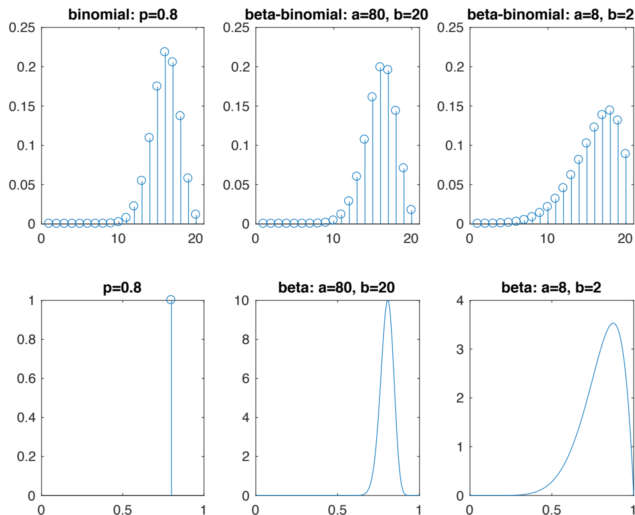
- n_i: the number of reads that contain "C" or "T" readout at the cytosine in sample i
- m_i: the number of reads that contain "C" readout (i.e. methylated) at the cytosine in sample i ($0 \leq m_i \leq n_i$)
- If we knew the underlying methylation proportion p_i, then: $M_i \sim \mathrm{Binom}(p_i, n_i)$
- p_i: the unknown methylation level of the cytosine in sample i
- Instead of assuming a fixed (unknown) methylation level, assume p_i has a compounding distribution $p_i \sim \mathrm{Beta}(\alpha, \beta),\ \alpha \geq 0, \beta \geq 0$
- The probability of observing methylation level $M_i = m_i$ for a coverage n_i follows so called beta-binomial model $\mathrm{BetaBinomial}(n_i, \alpha, \beta)$

$$\begin{aligned} P(M_i = m_i | n_i, \alpha, \beta) &= \int_0^1 \mathrm{Binom}(m_i | p_i, n_i) \mathrm{Beta}(p_i | \alpha, \beta) \mathrm{d} p_i \\ &= \binom{n_i}{m_i} \frac{\mathrm{B}(m_i + \alpha, n_i - m_i + \beta)}{\mathrm{B}(\alpha, \beta)}, \end{aligned}$$

where B is the beta function

Beta-binomial model

- An illustration of binomial / beta / beta-binomial densities



Binomial and beta-binomial densities

Beta-binomial model

- Mean and variance of the beta-binomial random variable are

$$\mu = \frac{n_i\alpha}{\alpha+\beta} \quad \text{and} \quad \sigma^2 = \frac{n_i\alpha\beta(\alpha+\beta+n_i)}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Beta-binomial model

- Mean and variance of the beta-binomial random variable are

$$\mu = \frac{n_i \alpha}{\alpha + \beta} \quad \text{and} \quad \sigma^2 = \frac{n_i \alpha \beta (\alpha + \beta + n_i)}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$$

- Reparameterization
  - $\pi = \frac{\alpha}{\alpha+\beta}$ is the the average methylation level of a set of replicate samples
  - $\gamma = \frac{1}{\alpha+\beta+1}$ is the common dispersion parameter

Beta-binomial model

- Mean and variance of the beta-binomial random variable are

$$\mu = \frac{n_i \alpha}{\alpha + \beta} \quad \text{and} \quad \sigma^2 = \frac{n_i \alpha \beta (\alpha + \beta + n_i)}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$$

- Reparameterization
  - $\pi = \frac{\alpha}{\alpha+\beta}$ is the the average methylation level of a set of replicate samples
  - $\gamma = \frac{1}{\alpha+\beta+1}$ is the common dispersion parameter
- This allows us to write the original model $\mathrm{BetaBinomial}(n_i, \alpha, \beta)$ equivalently as

$$M_i \sim \mathrm{BetaBinomial}(n_i, \pi, \gamma)$$

  where the mean and the variance are now defined as
  - $\mathrm{E}(M_i) = n_i \pi$
  - $\mathrm{Var}(M_i) = n_i \pi (1 - \pi)(1 + (n_i - 1)\gamma)$
- That is, given π and γ, solve for α and β, and evaluate pdf using $\mathrm{BetaBinomial}(n_i, \alpha, \beta)$
- Recall that the variance of the binomial distribution is $n_i \pi (1 - \pi)$ which is smaller than $\mathrm{Var}(M_i)$ for $n_i \geq 2$

GLM with beta-binomial likelihood

- In most of the real world applications, methylation levels can be confounded by one or more factors (e.g. age and smoking)
- The generalized linear model (GLM) generalizes the linear regression to allow for response variables that have likelihood models other than a normal distribution

GLM with beta-binomial likelihood

- For each sample i (and for each cytosine), the mean methylation level π_i depends on covariates $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^T$

$$g(\pi_i) = \sum_{j=1}^{p} x_{ij}\beta_j = \mathbf{x}_i^T \boldsymbol{\beta}$$

where $\boldsymbol{\beta}$ is a $p \times 1$ parameter vector and

$$\begin{aligned} g(\pi) &= \mathrm{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) \\ \pi_i &= \mathrm{logit}^{-1}(\mathbf{x}_i^T \boldsymbol{\beta}) = \mathrm{logistic}(\mathbf{x}_i^T \boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\exp(\mathbf{x}_i^T \boldsymbol{\beta}) + 1} \end{aligned}$$

- $\mathrm{logit}(\cdot) :]0,1[\to \mathbb{R}$, thus $\mathrm{logit}(\cdot)^{-1} : \mathbb{R} \to]0,1[$

Model fitting and inference

- The beta-binomial regression is fit separately for each CpG site
- The parameters $\boldsymbol{\beta}$ and γ are estimated using maximum likelihood
  - E.g. iteratively reweighted least squares algorithm using a Newton-Raphson method
- Test the differential methylation w.r.t. a covariate x_j with parameter β_j:
  - Null hypothesis: $\beta_j = 0$
  - Estimate the full model, and the reduced model without the covariate x_j
  - Compare the models using log-likelihood ratio test

$$D = -2 \ln \left(\frac{\text{likelihood of the reduced model}}{\text{likelihood of the full model}} \right)$$

- p-value from chi-square test with $d_{full} - d_{reduced}$ degrees of freedom, where d_{full} denotes the number of free parameters in the full model
- Remember multiple testing!
- Neighbouring CpGs that are detected to be statistical significant can be combined to form differentially methylated regions (DMRs)

RadMeth application

- Neuron and non-neuron RRBS-seq samples from mouse frontal cortex: $x_{i1} \in \{0, 1\}$
- 6 samples: $s = 6$
- Two additional factors: age ($x_{i2} \in \mathbb{R}_+$), sex ($x_{i3} \in \{0, 1\}$)
- 72 000 differentially methylated regions (DMRs) between neuron and non-neuron samples that contain at least 10 CpGs
- DMRs with minimum methylation difference above 0.55
  - 1708 lowly methylated (active) regions in neurons
  - These regions are associated with (located close to) 1089 genes
  - GO enrichment analysis by DAVID found a strong association of these genes with various aspects of neuronal development and function

RadMeth application

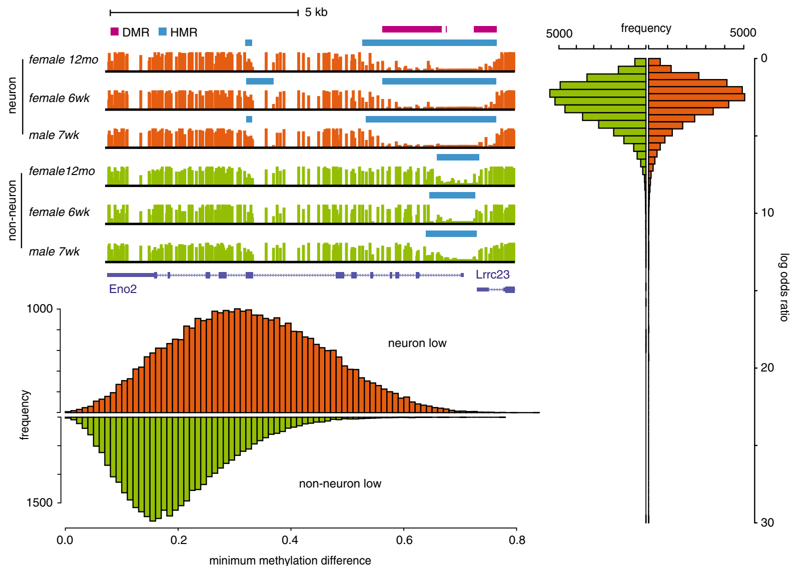


Figure from (Dolzhenko and Andrew, 2014)

Contents

- DNA methylation
- Bisulfite sequencing (BS-seq) protocol
- Alignment and quantification of BS-seq data
- Statistical analysis of BS-seq data
- Other details

RRBS data analysis workflow example

1. RRBS data preprocessing
- TrimGalore (trimming of low-quality bases, end repair biases, adapter sequences, and overlaps between paired-end reads), fastQC before and after TrimGalore
- Bismark-alignment
- Bismark methylation extractor to extract numbers of methylated and unmethylated reads at each CpG site in each sample
- Excluding of samples that have low bisulfite conversion efficiency (estimated as proportion of Cs converted to Ts in the completely unmethylated lambda phage genome)
- Trimming of M-biases

2. SNP detection (e.g. bis-SNP or bs-SNPer)
- SNPs removed from each sample

3. Coverage filtering
- CpG sites with extremely high coverages within each sample removed to avoid PCR duplicates
- Low-coverage CpG sites, e.g. with coverage < 10 reads in more than 50 % of the samples removed

4. Differential methylation analysis
- E.g. beta-binomial models are suitable for read count data:

$$m \sim \mathrm{Bin}(p,n)$$
$$p \sim \mathrm{Beta}(\alpha, \beta)$$

5. Interpretation
- Annotation to genomic parts, nearest genes, regulatory regions
- Known mQTL and eQTM effects

1. Preprocessing

MspI digest

5′ CGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNC −3′
3′ − CNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGC −5′

end repair + A-tailing

5′ − CGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCGA −3′
3′ − AGCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGC −5′

adapter ligation (X = adapter sequence)

5′ −XXXXXXXTCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCGAXXXXXXXXXX−3′ OT
3′ −XXXXXXXAGCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGCTXXXXXXXXXX−5′ OB

bisulfite treatment

PCR amplification of the OT strand

5′ −XXXXXXXTCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCGAXXXXXXXXXX−3′ OT
3′ −XXXXXXXAGCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGCTXXXXXXXXXX−5′ CTOT

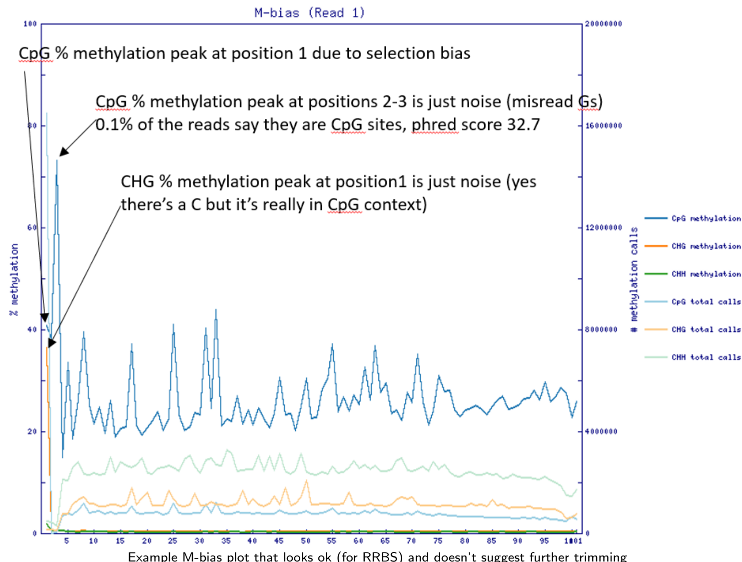
PCR amplification of the OB strand

5′ −XXXXXXXTCGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNCCGAXXXXXXXXXX−3′ CTOB
3′ −XXXXXXXAGCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGCTXXXXXXXXXX−5′ OB

Figure from (Krueger 2013).

- TrimGalore: trimming of
  - Low-quality bases
  - End repair biases
  - Adapter sequences
  - overlaps between paired-end reads

1. Preprocessing (steps after alignment and methylation extractor)



Example M-bias plot that looks ok (for RRBS) and doesn't suggest further trimming

- Excluding of samples that have low bisulfite conversion efficiency (estimated as proportion of Cs converted to Ts in the completely unmethylated lambda phage genome)
- Trimming of M-biases

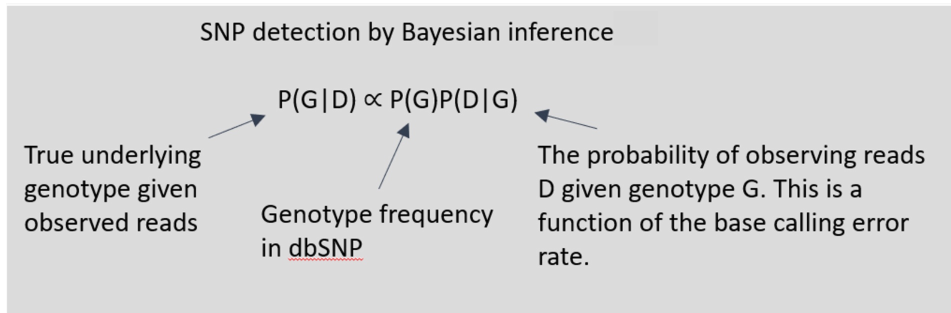
2. SNP detection

- Possible effects of SNPs (Liu et al. 2012):
  - C to T SNPs at methylation sites might be misinterpreted as partially (50%) or completely unmethylated Cs

2. SNP detection

- Possible effects of SNPs (Liu et al. 2012):
  - C to T SNPs at methylation sites might be misinterpreted as partially (50%) or completely unmethylated Cs
- SNP detection: Most bisulfite sequencing protocols only convert unmethylated C, while G on the opposing strand remains unchanged

$\rightarrow$ Detect C to T mutations using the reads from the opposite strand



3. Coverage filtering

- CpG sites with extremely high coverages within each sample removed to avoid PCR duplicates in RRBS (although some number of overlaps are expected)
- For example, low-coverage CpG sites, i.e. with coverage < 10 reads in more than 50 % of the samples removed

RRBS data analysis workflow example

1. RRBS data preprocessing

- TrimGalore (trimming of low-quality bases, end repair biases, adapter sequences, and overlaps between paired-end reads), fastQC before and after TrimGalore
- Bismark-alignment
- Bismark methylation extractor to extract numbers of methylated and unmethylated reads at each CpG site in each sample
- Excluding of samples that have low bisulfite conversion efficiency (estimated as proportion of Cs converted to Ts in the completely unmethylated lambda phage genome)
- Trimming of M-biases

2. SNP detection (e.g. bis-SNP or bs-SNPer)

- SNPs removed from each sample

3. Coverage filtering

- CpG sites with extremely high coverages within each sample removed to avoid PCR duplicates
- Low-coverage CpG sites, e.g. with coverage < 10 reads in more than 50 % of the samples removed

4. Differential methylation analysis

- E.g. beta-binomial models are suitable for read count data:

$$m \sim \text{Bin}(p,n)$$
$$p \sim \text{Beta}(\alpha, \beta)$$

5. Interpretation

- Annotation to genomic parts, nearest genes, regulatory regions
- Known mQTL and eQTM effects

5. Interpretation after the differential methylation analysis

Could the differentially methylated sites/regions have an impact on the expression of some gene?

- Nearest genes
  - E.g. some GO terms enriched in the nearest genes?
- Genomic parts: intron/exon/promoter/intergenic
  - R package genomation

5. Interpretation after the differential methylation analysis

Could the differentially methylated sites/regions have an impact on the expression of some gene?

- Nearest genes
  - E.g. some GO terms enriched in the nearest genes?
- Genomic parts: intron/exon/promoter/intergenic
  - R package genomation
- Enhancer regions
  - GeneHancer database

5. Interpretation after the differential methylation analysis

Could the differentially methylated sites/regions have an impact on the expression of some gene?

- Nearest genes
  - E.g. some GO terms enriched in the nearest genes?
- Genomic parts: intron/exon/promoter/intergenic
  - R package genomation
- Enhancer regions
  - GeneHancer database
- Transcription factor binding sites
  - TRANSFAC (a manually curated database of transcription factor binding sites)

5. Interpretation after the differential methylation analysis

Could the differentially methylated sites/regions have an impact on the expression of some gene?

- Nearest genes
  - E.g. some GO terms enriched in the nearest genes?
- Genomic parts: intron/exon/promoter/intergenic
  - R package genomation
- Enhancer regions
  - GeneHancer database
- Transcription factor binding sites
  - TRANSFAC (a manually curated database of transcription factor binding sites)
- mQTL and eQTM
  - For human whole blood e.g. GoDMC mQTLdb (mqtldb.godmc.org.uk) or BIOS QTL browser (genenetwork.nl)

mQTL (methylation quantitative trait loci) and eQTM (expression quantitative trait methylation)

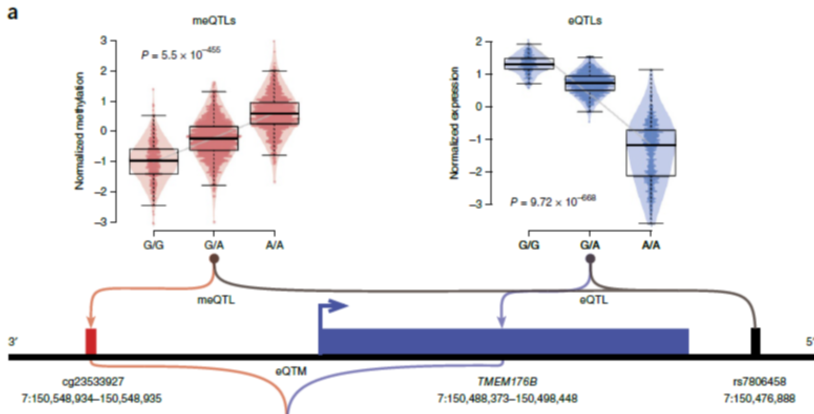


Figure from (Bonder et al. 2017)

References

- Bonder et al., Disease variants alter transcription factor levels and methylation of their binding sites, Nature Genetics, Vol. 49, pp 131â138, 2017
- Michael J. Booth et al., Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution, Science, 336(6063):934-937, 2012
- Jeremy J Day & J David Sweatt, DNA methylation and memory formation, Nature Neuroscience 13:1319-1323, 2010
- Egor Dolzhenko and Andrew D Smith, Using beta-binomial regression for high-precision differential methylation analysis in multifactor whole-genome bisulfite sequencing experiments, BMC Bioinformatics, 15:215, 2014
- Eckhardt F et al., DNA methylation profiling of human chromosomes 6, 20 and 22, Nature Genetics,38(12):1378-85, 2006.
- Felix Krueger and Simon R. Andrews, Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications, Bioinformatics, 27(11):1571-1572, 2011.
- Felix Krueger et al., DNA methylome analysis using short bisulfite sequencing data, Nature Methods 9, 145-151, 2012
- Felix Krueger, Reduced Representation Bisulfite-Seq - A Brief Guide to RRBS, Babraham Bioinformatics, 2013
- Jialong Lianga et al., Single-Cell Sequencing Technologies: Current and Future, Journal of Genetics and Genomics, 41(10):513-528, 2014
- Yaping Liu et al., Bis-SNP: Combined DNA methylation and SNP calling for Bisulfite-seq data, Genome Biology Vol. 13, No. R61, 2012
- Jovana Maksimovic et al., A cross-package Bioconductor workflow for analysing methylation array data, F1000Res, Jun 8;5:1281, 2016
- Alexander Meissner, et al., Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis, Nucleic Acids Res., 33(18):5868-77, 2005.
- Christoph Plass, et al., Mutations in regulators of the epigenome and their connections to global chromatin patterns in cancer, Nature Reviews Genetics 14, 765-780, 2013
- Dirk Schubeler, Epigenomics: Methylation matters, Nature 462:296-297, 2009
- Cornelia G Spruijt & Michiel Vermeulen, DNA methylation: old dog, new tricks?, Nature Structural & Molecular Biology 21, 949-954, 2014
- Benedicte L Tremblay et al., Familial resemblances in blood leukocyte DNA methylation levels, Epigenetics Vol. 11, No. 11, 831â838, 2016