ipDMR: identification of differentially methylated regions with interval P-values

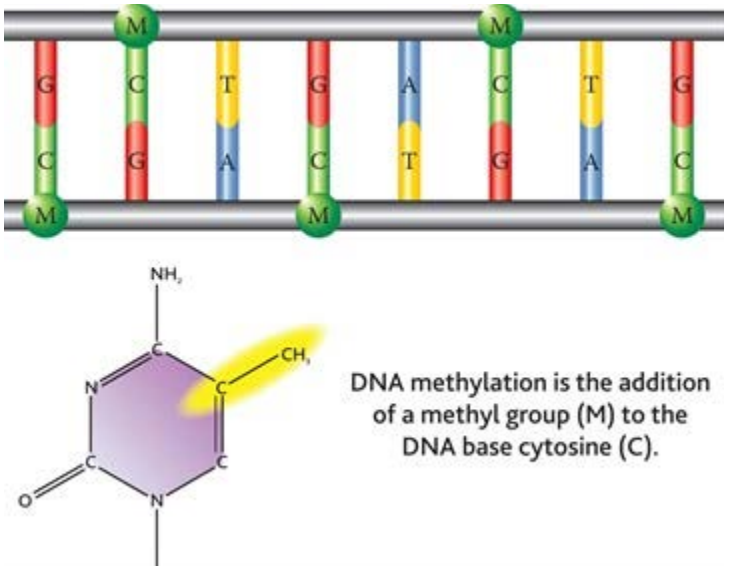
Liang Niu
University of Cincinnati

Overview

- Introduction
- Method
- Results
- Discussion
- Demonstration

Introduction: DNA methylation

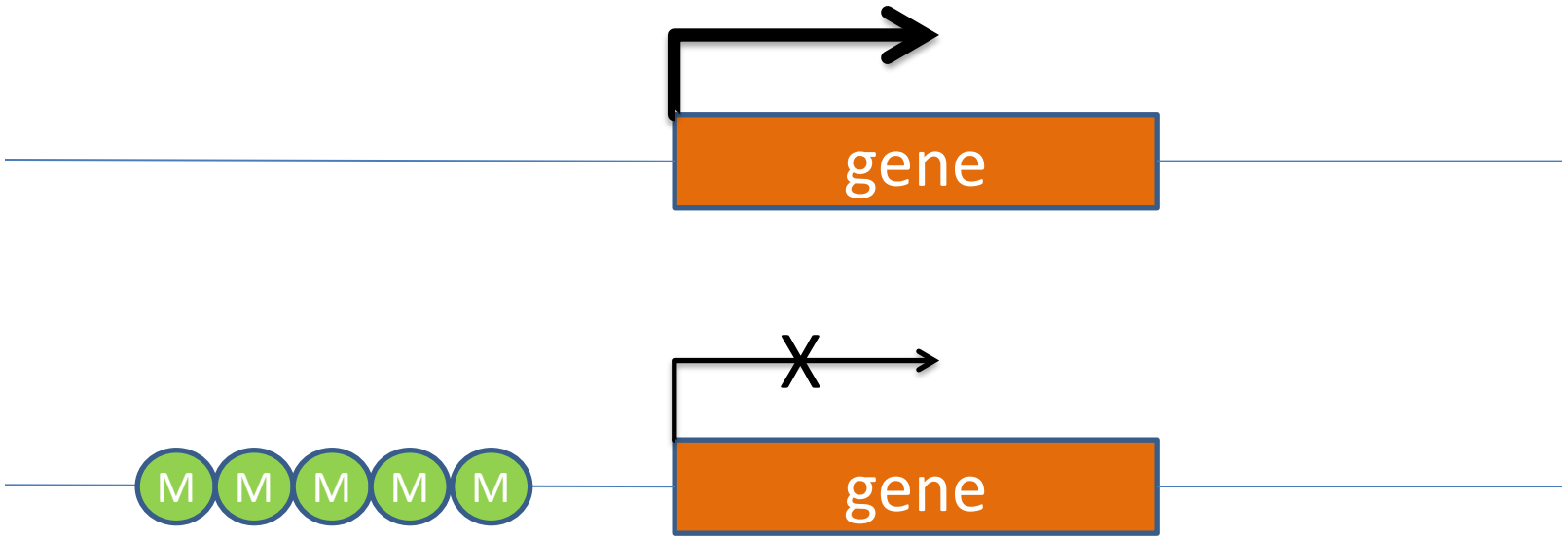
- DNA methylation: a modification of DNA
- In human body cells, DNA methylation typically occurs at CpG dinucleotide context (CpG site)



From http://education.cambridge.org

Introduction: DNA methylation

- When located in a gene promoter, DNA methylation typically represses gene transcription



Introduction: DMR

- Methylation Level (at a DNA locus): the proportion of cells with methylated sites among many cells.
- Differentially Methylated Site (DMS): a DNA locus with different methylation levels between two groups.
- Differentially Methylated Region (DMR): a DNA region with multiple differentially methylated sites.
  - DMRs may have stronger associations with disease than DMS.

Method: ipDMR workflow

- Given p-values for individual CpG sites, ipDMR
  1. calculates a P-value for each small interval, i.e. the interval bordered by two adjacent CpG within a user-specified value (default: 1000 bp);
  2. performs the Benjamini–Hochberg (BH) procedure on the interval P-values to select those significant intervals at a user-specified false discovery rate (FDR) threshold;
  3. joins all nearby significant intervals and CpGs if the gap (the number of bps between two intervals/CpGs) is less than the user-specified value (default: 1000 bp);
  4. recalculates P-values for each combined region using the original P-values for all CpGs in that region;
  5. performs another BH procedure on these region P-values to obtain the FDR-adjusted P-values.

Results: ipDMR vs comb-p (simulation study)

Method	Seed	Bin size	TP (SD)	FD (SD)
ipDMR	0.01	310	0.44 (0.11)	0.21 (0.13)
	0.01	50	0.44 (0.12)	0.19 (0.12)
	0.1	310	0.61 (0.11)	0.59 (0.10)
	0.1	50	0.61 (0.11)	0.57 (0.10)
Comb-p	0.01	310	0.40 (0.11)	0.43 (0.14)
	0.01	50	0.36 (0.11)	0.31 (0.14)
	0.1	310	0.50 (0.11)	0.69 (0.07)
	0.1	50	0.45 (0.10)	0.57 (0.11)

Discussion

- Nazer, Naghme, et al. "A novel approach toward optimal workflow selection for DNA methylation biomarker discovery." BMC bioinformatics 25.1 (2024): 1-19.
  - “Here seven techniques which are amongst the most popular methods were selected for benchmarking. These include BumpHunter, ProbeLasso, seqlm, DMRcate, COHCAP, Comb-p, and ipDMR” (page 8)
  - “The method ipDMR consistently outperformed other DMR finding methods, according to the comparison analyses” (page 14)
  - A validation analysis “further demonstrated that our proposed guidelines are useful for achieving reliable results.” (page 16)

Demonstration: ENmix package

- Bioconductor package ENmix includes ipDMR
- See ENmix user's guide: https://www.bioconductor.org/packages/release/bioc/vignettes/ENmix/inst/doc/ENmix.html

```
library(ENmix)
dat=simubed()
head(dat)
ipdmr(data=dat,seed=0.1)
```