EpiMix: an integrative tool for epigenomic subtyping using DNA methylation

Yuanning Zheng¹, John Jun¹, Kevin Brennan¹ and Olivier Gevaert¹

¹Stanford Center for Biomedical Informatics Research (BMIR), Department of Medicine & Department of Biomedical Data Science, Stanford University

Abstract

Emerging evidence has revealed the regulatory roles of DNA methylation (DNAm) on protein-coding genes and non-coding RNAs, and recent technologies enable genome-wide quantification of DNAm in large human cohorts. This creates the need to use a model-based computational approach to resolve the epigenetic heterogeneity in large human cohorts and to pinpoint the individuals carrying differential methylation profiles. Here we developed EpiMix, a comprehensive tool for population-level analysis of DNAm. EpiMix allows us to detect abnormal DNAm that were presented in only small subsets of a patient cohort and to identify DNAm-associated disease subtypes. Furthermore, we applied this model-based approach to identify abnormal DNAm at functionally diverse genomic elements, including cis-regulatory elements within protein-coding genes, distal enhancers, and genes encoding microRNAs and IncRNAs. In two separate studies, we showed that EpiMix discovered novel epigenetic mechanisms underlying childhood food allergy and survival-associated, methylation-driven non-coding RNAs in non-small cell lung cancer. EpiMix is available as an R package and a web-based tool: https://epimix.stanford.edu

Motivation

Epigenetics and human disease

Epigenetics
- Cell proliferation
- Cell differentiation
- Cardiovascular disease
- Cell metabolism
- Immunological disease

DNA methylation regulates gene expression

Motif DNA methylation transferases

DNA methylation

Cytosine

SAS

BAM

DNA methylation

Methyl group

CpG

CpG

Target gene

Promoter

Epigenetic biomarkers and personalized medicine

Use case 1: CD4+ T cells from children with food allergy

Workflow

i. Downloading

ii. Preprocessing

iii. Methylation modeling

iv. Functional analysis

v. User interface

Customized dataset

DNA methylation (beta values)

TCGA

Gene expression (ENCODE, TCGA, etc.)

Sample annotation

CpG filtering

Sample filtering

Missing value imputation

Batch effect correction

Enhancer mode

Intronic mode

ncRNA mode

Methylation

Expression

Cpg island

Motif

Enhancer

Intronic

ncRNA

Web application

R package

DNA methylation regulates gene expression

Use case 2: human lung cancer

Use case 3: CD4+ T cells from children with food allergy

Use case 4: human lung cancer

Conclusion

EpiMix is a comprehensive tool for genome-wide analysis of DNA methylation. It can be used to identify methylation-associated disease subtypes and improve patient classification. The application of EpiMix has the potential to discover novel epigenetic biomarkers and therapeutic targets for personalized medicine.