EpiMix: an integrative tool for epigenomic subtyping using DNA methylation

Yuanning Zheng[1], John Jun[1], Kevin Brennan[1] and Olivier Gevaert[1]

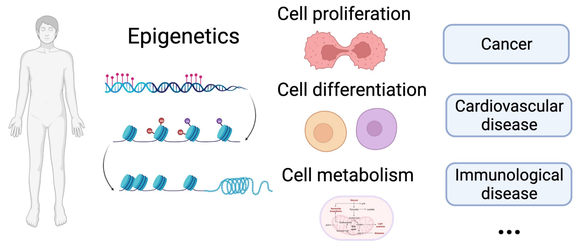
[1]Stanford Center for Biomedical Informatics Research (BMIR), Department of Medicine & Department of Biomedical Data Science, Stanford University

Abstract

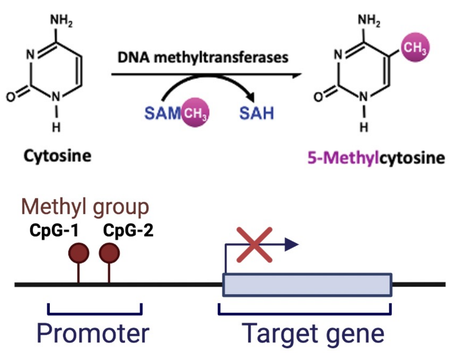
Emerging evidence has revealed the regulatory roles of DNA methylation (DNAme) on protein-coding genes and non-coding RNAs, and recent technologies enable genome-wide quantification of DNAme in large human cohorts. This creates the need to use a model-based computational approach to resolve the epigenetic 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 DNAme. EpiMix allows us to detect abnormal DNAme that were presented in only small subsets of a patient cohort and to identify DNAme-associated disease subtypes. Furthermore, we applied this model-based approach to identify abnormal DNAme at functionally diverse genomic elements, including *cis*-regulatory elements within protein-coding genes, distal enhancers, and genes encoding microRNAs and lncRNAs. 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

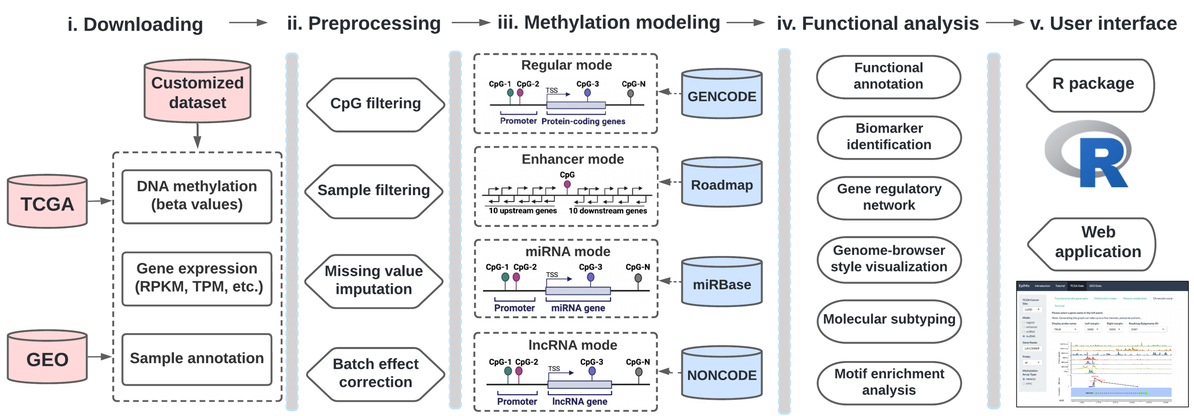
Epigenetics and human disease



DNA methylation regulates gene expression

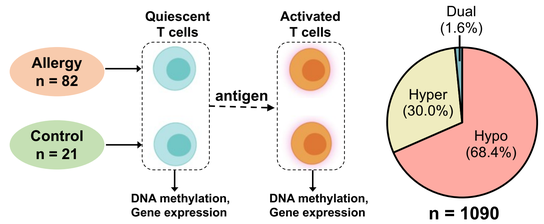


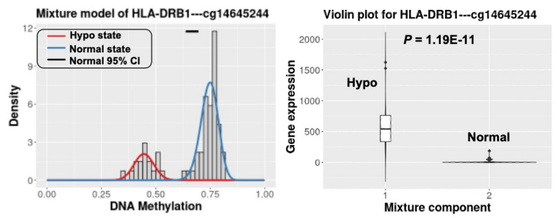
Workflow

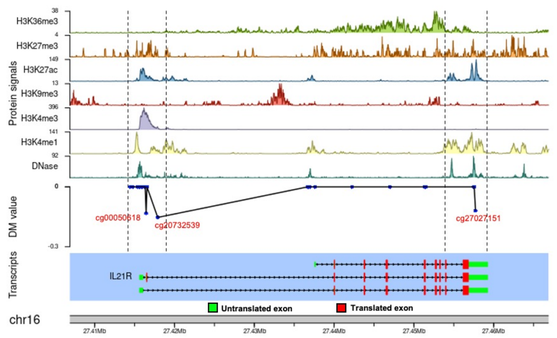


Use case 1: CD4+ T cells from children with food allergy

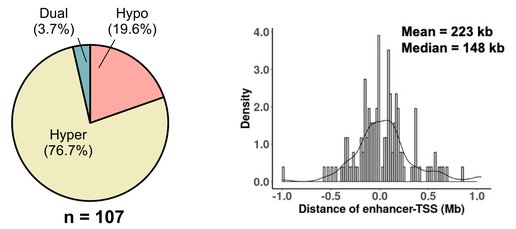
1. Protein-coding genes

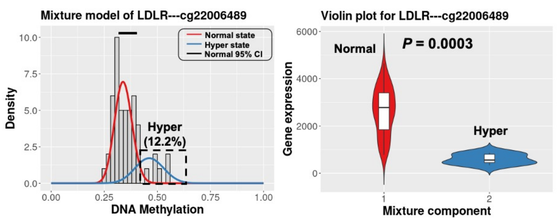




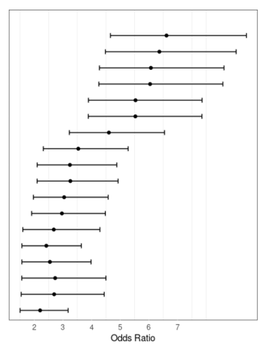


2. Distal enhancers



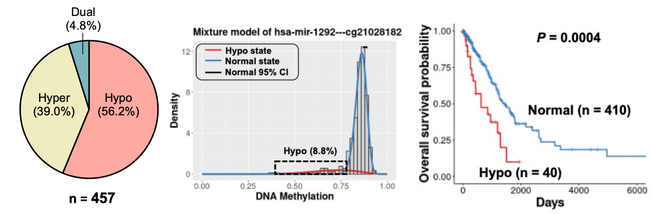


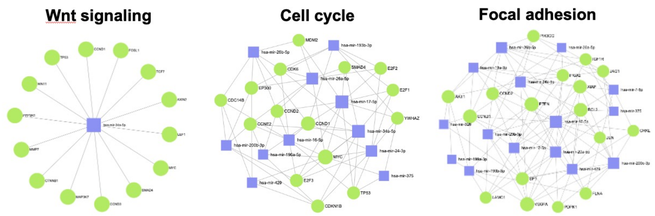
Motif	Odds ratio (95% CI)	# probes (% of paired)
JUN	6.62 (4.66-9.41)	70 (51.09%)
FOSL2	6.37 (4.48-9.05)	69 (50.36%)
FOSL1	6.07 (4.27-8.63)	68 (49.64%)
FOSB	6.04 (4.25-8.58)	67 (48.91%)
JUND	5.53 (3.89-7.86)	65 (47.45%)
FOS	5.53 (3.88-7.85)	64 (46.72%)
JUNB	4.6 (3.22-6.54)	59 (43.07%)
BATF	3.54 (2.31-5.28)	33 (24.09%)
BATF3	3.24 (2.1-4.88)	31 (22.63%)
TF65	3.26 (2.1-4.93)	30 (21.9%)
NFAC1	3.04 (1.97-4.58)	31 (22.63%)
RELB	2.96 (1.91-4.48)	30 (21.9%)
NR1H2	2.68 (1.6-4.29)	21 (15.33%)
IRF7	2.42 (1.57-3.64)	31 (22.63%)
BATF	2.54 (1.56-3.98)	24 (17.52%)
PDX1	2.73 (1.56-4.5)	18 (13.14%)
EMX2	2.69 (1.54-4.44)	18 (13.14%)
IRF2	2.2 (1.5-3.18)	45 (32.85%)



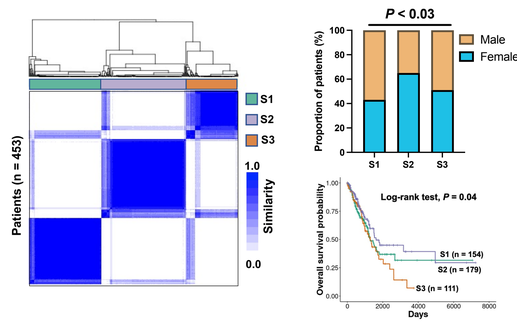
Use case 2: human lung cancer

3. miRNAs

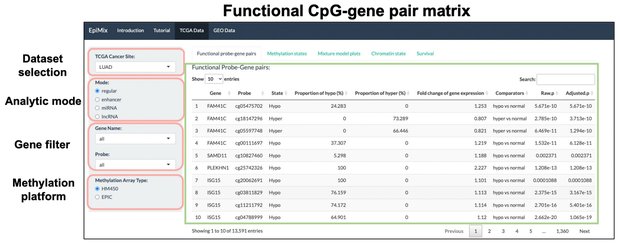




4. lncRNAs



EpiMix web application: https://epimix.stanford.edu



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.