Submitters Name: Youshu Cheng Submitted Email: youshu.cheng@yale.edu
PI Name: Ke Xu PI email: ke.xu@yale.edu

A hierarchical Bayesian interaction model to estimate cell-type-specific methylation quantitative trait loci incorporating priors from cell-sorted bisulfite sequencing data

Youshu Cheng[1,2], Biao Cai[1], Hongyu Li[1], Xinyu Zhang[2,3], Gypsyamber D'Souza[4], Sadeep Shrestha[5], Andrew Edmonds[6], Jacquelyn Meyers[7], Margaret Fischl[8], Seble Kassaye[9], Kathryn Anastos[10], Mardge Cohen[11], Bradley E Aouizerat[12,13], Ke Xu[2,3] & Hongyu Zhao[1,2]

[1]Department of Biostatistics, Yale School of Public Health;
[2]VA Connecticut Healthcare System;
[3]Department of Psychiatry, Yale School of Medicine;
[4]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health;
[5]Department of Epidemiology, University of Alabama at Birmingham;
[6]The University of North Carolina at Chapel Hill;
[7]SUNY Downstate Health Sciences University School of Medicine, Department of Psychiatry;
[8]Department of Medicine, University of Miami School of Medicine;
Division of Infectious Diseases and Tropical Medicine, Georgetown University;
[10]Department of Medicine, Albert Einstein College of Medicine;
[11]Hektoen Institute for Medical Research;
[12]Bluestone Center for Clinical Research, College of Dentistry;
[3]Department of Oral and Maxillofacial Surgery, College of Dentistry

Methylation Quantitative Trait Loci (meQTLs) are chromosomal regions that harbor genetic variants affecting DNA methylation levels. The identification of meQTLs can be accomplished through quantifying the effects of single nucleotide polymorphisms (SNPs) on DNA methylation levels, and these inferred meQTLs can shed light on the complex interplay between the genome and methylome. However, most meQTL studies to date utilize bulk methylation datasets composed of different cell types that may have distinct methylation patterns in each cell type. Current technological challenges hinder the comprehensive collection of large-scale, cell-type-specific (CTS) methylation data, which limits our understanding of CTS methylation regulation. To address this challenge, we propose a hierarchical Bayesian interaction model (HBI) to infer CTS meQTLs from bulk methylation data.

Our HBI method integrates bulk methylations data from a large number of samples and CTS methylation data from a small number of samples to estimate CTS meQTLs. Through simulations, we show that HBI improves the estimation (accuracy and power) of CTS genetic effects on DNA methylation. To systematically characterize genome-wide SNP-methylation level associations in multiple cell types, we apply HBI to bulk methylation data measured in peripheral blood mononuclear cells (PBMC) from a cohort of 431 individuals together with flow-sorted cell-derived methylation sequencing (MC-seq) data measured in isolated white blood cells (CD4+ T-cells, CD8+ T-cells, CD16+ monocytes) from 47 participants. Further, we demonstrate that HBI can identify CTS meQTLs and improve the functional annotation of SNPs.

HBI can incorporate strong and robust signals from MC-seq data to improve the estimation of CTS meQTLs. Applying HBI to link the methylome and genome data helps to identify biologically relevant cell types for complex traits.