Measurement Errors in Array-Based DNA Methylation Analysis

Weihua Guan

University of Minnesota
E-mail: wguan@umn.edu

Abstract: Genome-wide DNA methylation measures are now routinely used to investigate their association with various outcomes of interest. Inherent to the array-based DNA methylation measures, e.g., Illumina HumanMethylation450 (HM450) or Infinium MethylationEPIC chips, is the associated technical variation of non-biological interest. One example of such variation is the bead-to-bead variation due to the use of multiple beads on HM450 or EPIC arrays. These technical variations can be viewed as measurement errors, and are commonly ignored in downstream association analysis. We have proposed a novel statistical framework to take into account of these measurement errors. Specifically, we used a mixed effects model to quantify the measurement error, and developed an expectation–maximization (EM) algorithm to estimate the model parameters. We applied our proposed method to the Atherosclerosis Risk in Communities (ARIC) methylation data (n = 2,843; HM450 array) in an epigenome-wide association study of smoking status, after accounting for the bead-to-bead variation. We identified 14 additional CpG sites associated with smoking, at sites with high technical variation (intraclass correlation coefficient < 0.4). We expect that our new method can improve statistical power of association tests and accuracy of parameter estimates in future epigenome-wide association studies (EWASs) using methylation arrays.