

Inference of Specific Haplotype Effects in Case-Control Studies Using Unphased Genotype Data

Michael P. Epstein, Ph.D.
Emory University

Abstract:

Haplotype-based association methods are powerful and popular procedures for identifying genes that influence complex disease. For a case-control study design, a variety of statistical methods exist that detect haplotype-disease association by comparing haplotype frequencies among sampled cases and controls. Given unrelated samples often consist of unphased genotype data (resulting in haplotype ambiguity), many such statistical methods account for missing haplotype information by using the Expectation-Maximization (EM) algorithm for inference. While existing haplotype-based association methods are important, the majority fail to determine how specific haplotypes influence disease. Inference of specific haplotype effects is valuable—particularly for identifying functional variants of a candidate gene. Therefore, I develop a retrospective likelihood for estimating and testing the effects of specific SNP-based haplotypes on disease in a case-control study assuming unphased genotype data. This proposed method has a flexible structure that allows modeling of main and interaction effects of specific haplotypes on disease. For statistical inference, I apply an Expectation-Conditional-Maximization (ECM) algorithm to account for the ambiguous haplotype information in the genotype data. Using simulation studies, I assess both the power and accuracy of this proposed method for haplotype inference.