Integrating prior genomic information into genome-wide association studies

Thomas DC
Department of Preventive Medicine, Keck School of Medicine, University of Southern California2, Los Angeles, CA 90089-9011, USA
dthomas@usc.edu

Genome-wide association studies have the potential to detect the effects of common low-penetrance genetic variants in complex diseases. Most such studies are based on some form of multistage sampling design. This generally involves an initial scan of hundreds of thousands of single nucleotide polymorphism markers on a sample of cases and controls, followed by testing of a subset of the most promising markers on an independent sample. Additional markers may also be included at this second stage to better characterize the full spectrum of genetic variation in the targeted regions. I will review a number of challenges in the design of such studies, focusing in particular on the use of genomic annotation data to help in the selection of regions to prioritize for the second stage of this process. This is based on a hierarchical regression model incorporating genomic covariates in both the probability that each association is a true positive and the magnitude of the association given that it is a true positive. I will also briefly discuss the use of family-based designs, testing of gene-gene and gene-environment interactions, ethnic heterogeneity, and other practical issues.