DIMENSION REDUCTION APPROACHES IN GENOME-WIDE ASSOCIATION STUDIES

Kent R. Bailey

Division of Biostatistics, Mayo Clinic, Rochester, Minnesota

Genome-wide association studies (GWAS) may currently involve on the order of 10^6 potential predictor variables (Single Nucleotide polymorphisms—SNP's), and a much smaller number of independent phenotypic responses. Methods for reducing the dimensionality of the space of predictors are therefore strongly indicated. Generic methods such as Principal Components Analysis, though not designed for discrete data, are a plausible approach that can be used successfully in the absence of prior information relating genotype to phenotype. In this study, we explore, instead, 2 simple methods which make use of external prior information, based either on prior unstructured analyses on other datasets, or based on a known or postulated relationship such as a race-phenotype association. In the former case, we use a prior gene-centric haplotype analysis in a training dataset to inform a test (validation) analysis on a new dataset. In the latter case, we use a well-known racial difference in response to classes of hypertension medication to inform and reduce the dimension of a search for genetic basis for differential inter-individual pharmacologic responses. Power considerations will be discussed based on simulation of polygenic and single-gene pharmacogenomic effects within actual pharmacogenomic studies (GERA1 and GERA2) involving biracial samples exposed to 2 classes of hypertension medications. In addition, the actual results of applying these methods will be shown for illustration.