ATELIER « MÉTHODES STATISTIQUES COMPUTATIONNELLES EN GÉNOMIQUE ET EN BIOLOGIE SYSTÉMIQUE »

18–22 AVRIL 2011

WORKSHOP ON "COMPUTATIONAL STATISTICAL METHODS FOR GENOMICS AND SYSTEMS BIOLOGY"

APRIL 18–22, 2011

A unified framework for testing multiple phenotypes for association with genetic variants

Matthew Stephens *

stephens@galton.uchicago.edu

In many ongoing genome-wide association studies, multiple related phenotypes are available for testing for association with genetic variants. In most cases, however, these related phenotypes are analysed independently from one another. For example, several studies have measured multiple lipid-related phenotypes, such as LDL-cholestrol, HDL-cholestrol, and Triglycerides, but in most cases the primary analysis has been a simple univariate scan for each phenotype. This type of univariate analysis fails to make full use of potentially rich phenotypic data.

While this observation is in some sense obvious, much less obvious is the right way to go about examining associations with multiple phenotypes. Common existing approaches include the use of methods such as MANOVA, canonical correlations, or Principal Components Analysis, to identify linear combinations of outcome that are associated with genetic variants. However, if such methods give a significant result, these associations are not always easy to interpret. Indeed the usual approach to explaining observed multivariate associations is to revert to univariate tests, which seems far from ideal.

In this work we outline an approach to dealing with multiple phenotypes based on Bayesian model averaging. The method attempts to identify which subset of phenotypes is associated with a given genotype. In this way it incorporates the null model (no phenotypes associated with genotype) ; the simple univariate alternative (only one phenotype associated with genotype) and the general alternative (all phenotypes associated with genotype) into a single unified framework. In particular our approach both tests for and explains multivariate associations within a single model, avoiding the need to resort to univariate tests when explaining and interpreting significant multivariate findings. We illustrate the approach on examples, and show how, when combined with multiple phenotype data, the method can improve both power and interpretation of association analyses.

*Department of Statistics, University of Chicago, Eckhart Hall Room 126, 5734 S. University Avenue, Chicago, IL 60637, USA.