Abstract

Univariate methods have frequently been used to discover Quantitative Trait Loci for gene expression measurements, often with much success. However, correlations caused by Linkage Disequilibrium as well as chance correlations which are functions of the large number of markers typically used in such studies, mean that causative regions can often cause multiple signals. Traditional investigations into the number of QTL for a given phenotype, such as visual inspection of likelihood plots, are not feasible when considering thousands of phenotypes. Stepwise methods have been suggested to counter this, but these are known to produce unstable models and there are difficulties in deriving significance estimates. The Lasso is a shrinkage method which has often been employed to discover true signals when the number of variables exceeds the number of observations. We propose a test statistic based on the threshold at which variables enter the Lasso model, prove analytic properties of this statistic which demonstrate parallels with univariate methods, and demonstrate its utility in proposing candidate QTL. We show that this method controls for LD structure, and the estimates of significance produced have superior properties when compared to those derived by stepwise methods. We test our method using a simulation strategy inspired by a previous eQTL mapping experiment. These simulations find that the ratio of true discoveries to false positives is superior for this method compared to univariate and stepwise approaches. Finally we apply the derived method to data from this experiment to investigate the nature of genetic regulation in this population.