A flexible mixture model to describe multifactorial genetic risk factors of complex disease, with an application for gene-gene and gene-environment interaction discovery.

Joanna L. Davies[1] and Chris Holmes[1]

[1] Department of Statistics, University of Oxford, UK

The same disease can be caused by different combinations of environmental, epigenetic and genetic factors. Different combinations of factors may cause different identifiable subtypes of disease but often differences in etiology cannot be distinguished by the clinical or molecular phenotype. Hence, many analyses of genetic SNP data from complex disease data sets assume that genetic variants have the same effect on disease risk across all individuals. In reality, the effect of a single genetic locus may differ across individuals according to their exposure to other environmental factors and/or the presence of other genetic variants. Models which assume variants at each locus have the same effect on disease risk for all individuals estimate the average effect a locus has on disease risk. Thus if a locus has a very strong effect in a relatively small proportion of cases and little or no effect in remaining cases, the estimated average effect will be small and may go undetected using such an approach.

We introduce a flexible mixture model to test for single marker association with a quantitative trait which allows the marker to affect the trait for some individuals but not others. Such a model accounts for the possibility that the marker is associated with a gene which is involved in gene-gene and/or gene-environmental interactions without direct knowledge of the nature of the interactions. For example if there exists a pairwise interaction between locus A and B, only individuals with a certain genotype at locus B may be affected by locus A. The model can be tailored to test for additive, dominant and recessive genetic effects on a continuous phenotype, although we show that in many circumstances the additive model has good power to detect dominant and recessive effects. Further more, the model can be extended to incorporate additional sources of information such as the relatedness of individuals in the study.

We explore an application of the model to a two stage search strategy for the detection of pairwise gene-gene interactions analogous to that presented by [1]. Exhaustive search strategies for the detection of pairwise interactions face a huge multiple testing correction. We reduce this penalty by testing only a subset of pairwise interactions chosen according to whether marginal effects are detected using our mixture model. We then fit (*a*) the full interaction model to all pairs of markers from those with a significant marginal effect and (*b*) the full interaction model for pairs of loci with at least one marker which has significant marginal effect. We compare power to detect gene-gene interactions using the Bonferroni adjustment for multiple testing with the strategy described by [1] and the exhaustive pairwise search strategy. Our results show that our strategy has greater power than [1] and under some circumstances has greater power to detect interactions than an exhaustive search strategy.

References

[1] Evans DM, Marchini J, Morris AP, Cardon LR, (2006) Two-stage two-locus models in genome-wide association. PLoS Genet 2(9): e157