A shrinkage regression approach to tackle the HLA region

C.Vignal (1,2), A.Bansal (2), C. Hoggart (1), D.Balding (1)
(1)Imperial College, UK, (2)GlaxoSmithKline, UK

Many autoimmune diseases have been associated with the HLA region, but the presence of linkage disequilibrium (LD) has meant that finding causal elements has been difficult. Multivariate association analyses can perform better than univariate methods, however, there can be problems when the number of variables exceeds the number of observations or in the presence of correlated predictors.

We adopt a Bayesian-inspired shrinkage regression approach for multilocus analysis of correlated data in which each regression coefficient is assigned a prior distribution that strongly favors zero values. We consider two shrinkage priors, the Laplace or double exponential distribution, and the normal-exponential-gamma distribution. Parameter inference is based on the posterior mode and terms with non-zero posterior modes indicate marker-disease associations.

We applied this approach to a case-control association study on rheumatoid arthritis (RA) using SNPs spanning the HLA region, together with genotypes from the multiallelic HLA-DRB1 locus. The latter is a known RA risk factor that was included in all our models without shrinkage. After controlling for type-I error, we found fewer positive SNP associations than in single-point tests, suggesting that LD might be better-handled. These results were supported by a simulation study. We selected a set of SNPs in various degrees of LD with HLA-DRB1. For each marker, case-control labels were randomised within the HLA-DRB1 allelic classes to simulate causal SNPs, while maintaining LD with HLA-DRB1. Our results showed that the shrinkage approach provides a substantial benefit, both in terms of maintaining statistical power to detect multiple causal variants and in the reduction of false positive associations.