

Family-based and genetically stratified association analysis

P.C. Sham

Genome Research Centre, the University of Hong Kong

Abstract

A variety of methods for family-based association analysis have been proposed in the last 2 decades. The transmission / disequilibrium test (TDT), which considers “trio” families with an affected child and both parents, has been particularly popular. This has been generalized in various ways to tests based on sibling and subsequently to general pedigrees (the pedigree disequilibrium test, PDT). The main advantage of these methods is that they are robust to false positive association arising from hidden population stratification. Yet, although family-based association studies may be more powerful than standard case-control designs under some extreme conditions, they are generally less efficient in most circumstances. We have recently suggested the inclusion of parental phenotype information into the family-based association analysis, in order to improve statistical power. Another possible approach to improve power is to select particular families that have affected members, although we have found that doing so often results in reduction rather than gain of power.

Whole-genome SNP data offer the possibility of classifying apparently unrelated individuals into genetically homogeneous subgroups, thereby achieving the main advantage of family-based association methods without the need to ascertain families. Methods for deriving homogeneous genetic subgroups can be based on the similarities measures between subjects such as average identity (by state) for all genotyped SNPs followed by hierarchical clustering, or a test of whether two individuals are from the same population, called the pair-wise population concordance (PPC) test. The latter test is based on the ratio of SNPs that are either homozygous but for different alleles in the two individuals (i.e. AA, BB) or heterozygous in both individuals (i.e. AB, AB). Clusters of individuals who can be considered to come from the same population are then subjected to stratified association analysis. Family data can also be considered in a stratified analysis framework, which allows the combination of case-control and family data in a single analysis. These procedures have been implemented in the freely available PLINK software (<http://pngu.mgh.harvard.edu/~purcell/plink/>).