

Homework 4

Style 1: literature project

This involves choosing a paper from the literature that extends or provides additional background on the material of Chapter 6 on family-based genetic association studies using unrelated individuals.

Topics may include: multiple testing, power, the added value of using families in GWAs, incorporating parental phenotype information, genotype imputation while respecting family-relatedness.

Summarize the paper of your choice (not one that has been discussed in class), its objectives, results, and show how it fits into the broader picture.

A good reference of papers can be retrieved from the URL:

<http://www.nslj-genetics.org/ld/>

Style 2: computing project

Use existing software / applets / simulation code, or write your own simulation code on issues that are relevant for population-based association studies. These may include:

- Computing the power of a family-based genetic association study:
 - E.g.,
 - Which factors affect the power (most)?
 - Based on your simulation study, can you make recommendations for good common practice in designing and carrying out a genetic association study?
 - Do these power calculation tools adjust for multiple testing?
 - Added value of the software PBAT?
- Assess the effect of introducing 5%, 10%, 15%, 20%, 25%, 30% missing *parental* genotypes on the power of your study or on study results.
 - E.g.,
 - What is the best way to analyze data with missing genotypes?
 - Impute? Ignore? ...
 - What do you observe?
 - Which percentages give rise to unreliable results?

Useful URLs :

<http://www.dorak.info/hla/stat.html>

<http://cran.r-project.org/web/views/Genetics.html>

Style 3: data analysis project

Introduction

In this homework it is the idea to use the simulated data we introduced before, in conjunction with the R “GenABEL” package, to answer relevant questions in a family-based genetic association context. All of the questions below can be answered by consulting the course notes (Chapter 5 and 6), the GenABEL manual and its tutorial (cfr. Supporting documents to April 2 class).

Specific questions on family-based genetic association analysis

- Q1. How many samples have a call rate less than 98? What is the lowest call rate still acceptable?
- Q2. How many SNPs deviate from HWE? Which is the criterion you would use to decide whether HWE is rejected? Can you motivate this?
- Q3. How many rare SNPs are present in the simulated data? Which ones have MAF < 1%? Why is this information valuable?
- Q4. Run a Quality Control check on SNPs/samples with default values. Are these default values appropriate for these data? Are there SNPs/samples that do not pass the Quality Control?
- Q5. Perform an automated descriptives run on the genotype data, as well as an outlier check on the trait
- Q6. Summarize the distribution of the kinship coefficients. Do you observe a correlation between genomic and pedigree kinships?
- Q7. Analyze the family data using the plain GC method, with and without empirical significance assessment
- Q8. Analyze the family data using FASTA testing (using heritability from a polygenic model). Is it possible to assess “genomewide” significance? If so, how would you implement this?
- Q9. Analyze the family data using GRAMMAS testing, with and without genomewide significance assessment
- Q10. How do Q7-Q9 compare?

Write a small report, including some explanations about how you obtained the answers

Due date: 23 April 2010