

Chapter 7: Bioinformatics

Introduction

We simulated data on a number of families, including parents and/or a number of children. The genotype data file *pedigree.ped* includes relevant data, with column content following “LINKAGE” format (“LINKAGE” is a genetics software package). A second data file, *pheno.dat*, includes data on a continuous phenotype. These data can be linked to the pedigree data using the unique identifiers for family members.

In this homework it is the idea to use the simulated data in conjunction with the “FBAT” software, to answer relevant questions in a family-based genetic association context. All of the questions below can be answered by consulting the course notes, the “FBAT manual” and its tutorial (cfr, “FBAT Tour”), all of which can be downloaded from the URL <http://www.biostat.harvard.edu/~fbat/fbat.htm>.

General task

Make sure the data are in the correct format for usage with FBAT (consult the FBAT manual), in particular the header line and the number of columns should be correct...

Specific questions on family-based genetic association analysis

Q1. How many rare SNPs are present in the simulated data? Which ones have MAF < 1%? Why is this information valuable?

Q2. Analyze the family data using the FBAT statistic on the continuous trait from the phenotype file, and consider the default options. What do you observe? Can you detect a significant marker?

Q3. Are the results in Q2 corrected for multiple testing? If not, how can you achieve overall significance?

Q4. Compute permutation based p-values for the settings of Q2. Vary the number of permutations. What can you tell about the precision of the “estimated” p-values?

Q5. Now re-analyze the same family data, exploring different settings in combination:

- Modes
- Models
- Options `-o` and `-e`

Compare your results with those obtained from Q2.

Your report with answers will serve as the written part of your exam for this chapter. Include a separate Appendix (also having your name on it), with the log file of your fbat analysis.

Due date: 5 May 2010