Chintan Mehta
April 2, 2016
chintan.mehta@yale.edu

Instructions

Place all files (code and data) and run all analysis in same directory. There are three R code files

- Load data with: ***load-PING.R***
- Perform analysis with: ***analysis-PING.R***
- Functions: ***functions.R***

Data files provided by PING:

- Genotype: ***PING_660_final_filtered*** in plink format (bed, bim, and fam files)
- Phenotype: csv file from PING data portal containing neuroimaging, cognitive, and family history data. Rename the file ***PING_FULL.csv*** prior to running analysis.

Steps

1. Obtain genotype PCs and pairwise inheritance-by-descent (IBD) with pLINK 1.90. In Windows terminal, run following commands in Dos terminal. Be sure output is in the working directory.

   plink.exe --bfile PING_660_final_filtered --out PING_PCA --pca 20

   plink.exe --bfile PING_660_final_filtered --out PING_IBD –genome

2. Run ***load-PING.R*** file to load data and create working sample
   a. Load genotype, phenotype, PCA, and IBD files.
   b. Create working data frame containing variables used in analysis.
   c. Set criteria for including subjects in sample (can be adjusted prior to running code).
      i. European genetic ancestry: determined by thresholds for first two PCS (using extreme first percentile of those with self-reported white ancestry)
      ii. Remove siblings (set to IBD > 0.20)
      iii. Minimum age (set to eight)
      iv. No missing cognitive or MRI data
   d. Socioeconomic status variables are important for cognitive responses. Missing values of household income and parental education are imputed using the Amelia package in R. Imputation is performed with household income, parental education, and parents' occupation measures.
   e. Two data frames are used throughout in procedure: ***Z11*** & ***D11***

3. Perform analysis with ***analysis-PING.R***
    a. Remove covariates from cognitive exam scores and MRI measurements prior to performing analysis. Covariates can be easily adjusted.
    b. Define learning performance score (used as response in GWAS)
        i. Average of IBAM, reading, vocabulary, and list sorting exams.
        ii. Determined by sign of loadings on second PC of eight cognitive exam scores.
    c. Define neuroimaging risk score
        i. Perform joint variance test to compare ratio of variances between diagnosed and undiagnosed subjects along neuroimaging variables.
        ii. Evaluate single JVAR test statistic by combining comparisons for thickness, area, and volume measurements of respective ROI
        iii. Significance determined by permutation tests
        iv. Using thickness measurements of six most significant ROIs, identify neuroimaging outliers in diagnosed subgroup.
        v. Evaluate neuroimaging (NSLP) risk score for each subject

4. Perform GWAS
    a. Significance from two tests of genomic effect are reported:
        i. Chi-squared (one dof) in typical main effect model for marginal associations
        ii. Chi-squared (two dof) in gene-risk interaction model (Interaction between SNP & NSLP)
    b. Performed in parallel.
        i. Time ~ 14 minutes on 10 cores
        ii. Memory ~ 300 MB (times # cores)

5. Output
    a. JVAR test: jvar-results.csv
    b. GWAS: gwas-results.csv