Evaluation of epistasis detection methods for quantitative phenotypes

Supplemental File S3

**EpiSNP:**

Used EPISNP1.exe with input formatted as specified in documentation.

Each dataset was represented using a \_\_\_.txt and \_\_\_chr1.dat files.

The former file contained ind\_ID, father\_ID, mother\_ID, sex, and phenotype columns.

The latter file contained fam\_ID, sex, ind\_ID, and a column for every single SNP.

For each comparison only 50 most significant hits were reported in output file.

**Matrix Epistasis:**

Each dataset was represented by a \_\_\_\_.csv and \_\_\_\_\_.pheno file.

Former contained the genotype listed for each SNP. Each column represented a single SNP. Each row represented a single individual.

Latter contained phenotypes listed for each individual in a single column.

Following R code was used to execute tool:

library("MatrixEpistasis")

compute\_matrix\_epistasis <- function(fname) {

fname1 = paste(fname, sep = '', '.csv')

fname2 = paste(fname, sep = '', '.pheno')

fname3 = paste(fname, sep = '', '\_MatrixEpistasis.txt')

data = read.csv(fname1)

snpA = as.matrix(data)

snpB = as.matrix(data)

trait = read.csv(fname2, header = FALSE)

MatrixEpistasis\_main(snpA, snpB, trait=trait, pvalThreshold=5e-8, outputFileName=fname3)

}

**MIDESP (1.2):**

Used Plink formatted file as input.

Used command as specified on MIDESP github repository to execute the tool.

Used all default settings (k = 30, fdr = 0.005, etc.)

-noapc flag was used, since datasets contained 1000 SNPs, and tool developers recommended not using apc with < 5000 SNPs

**PLINK BOOST (1.90b34.2):**

Sample Command:

module load plink2/1.90b3.42

plink --file Pure\_Dominant\_TwoPairs\_BaselineAlpha10\_InteractionAlpha16\_Chr1\_CEU\_SNP1000\_IND1000\_MAF005\_02\_converted --out Pure\_Dominant\_TwoPairs\_BaselineAlpha10\_InteractionAlpha16\_Chr1\_CEU\_SNP1000\_IND1000\_MAF005\_02\_converted

plink --bfile Pure\_Dominant\_TwoPairs\_BaselineAlpha10\_InteractionAlpha16\_Chr1\_CEU\_SNP1000\_IND1000\_MAF005\_02\_converted --allow-no-sex --fast-epistasis boost --epi1 5e-08 --out Pure\_Dominant\_TwoPairs\_BaselineAlpha10\_InteractionAlpha16\_Chr1\_CEU\_SNP1000\_IND1000\_MAF005\_02\_converted\_epistasis

**PLINK Epistasis (1.90b34.2):**

Sample Command:

module load plink2/1.90b3.42

plink --file Pure\_Dominant\_TwoPairs\_BaselineAlpha10\_InteractionAlpha16\_Chr1\_CEU\_SNP1000\_IND1000\_MAF005\_02 --out Pure\_Dominant\_TwoPairs\_BaselineAlpha10\_InteractionAlpha16\_Chr1\_CEU\_SNP1000\_IND1000\_MAF005\_02

plink --bfile Pure\_Dominant\_TwoPairs\_BaselineAlpha10\_InteractionAlpha16\_Chr1\_CEU\_SNP1000\_IND1000\_MAF005\_02 --allow-no-sex --epistasis --epi1 5e-08 --out Pure\_Dominant\_TwoPairs\_BaselineAlpha10\_InteractionAlpha16\_Chr1\_CEU\_SNP1000\_IND1000\_MAF005\_02\_epistasis

**QMDR (3.0.2):**

Formatted data into format specified by QMDR documentation.

Used graphical interface to perform all analyses.

Loaded prepared .csv file using Load Datafile button.

Then ran analysis using Run Analysis button.

Used all default settings, except that the attribute count range was changed from 1:3 to 2:2 since we were only interested in pairwise interactions.

**MDR:**

Used QMDR tool with discretized (binary phenotype) version of our datasets.

Identical procedure to the one for QMDR.