UNIVERSITY OF QUEENSLAND

Epistasis follow-up

*** summary ***

28th Nov 2014

1 Analysis summary

Original discovery analysis identified 501 interactions comprising of 781 unique SNPs and 238 genes (probes). These were 26 cis-cis, 462 cis-trans and 13 trans trans. The majority of our discovery interactions were composed of one SNP that was significantly associated with the gene expression level in the discovery data set, and one SNP that had no previous association (439 out of 501, Methods). Only nine interactions were between SNPs that both had known main effects, whereas 64 were between SNPs that had no known main effects.

2 Methods

The following analyses have been conducted;

1. Determining the empirical p-values for each of the 501 interactions

The initial analysis used F-tests and some simulation work to determine the expected Type 1 error rate in the 1st stage of the discovery process. The 1st stage was followed by a 2nd stage where the interaction model was fitted. Subsequent simulations and theoretical calculations have suggested that the Type 1 error rate of the 2nd stage is not correct when there is a large main effect and / or in the presence of LD.

a. Identify which of the two snps in the original epistasis pair has the largest additive effect.

b. Treat the largest additive SNP as a fixed SNP and perform a genome-wide analysis using the 8df and 4df epistasis model.

c. This generates $\approx 500,000$ interaction p-values. Apply the same snp-pair filtering as used in the manuscript. Namely, LD ($r^2 < 0.1$), nclass = 9, and minclass > 5. Any SNP with +/- 5MB of original epistasis SNP pairs were also removed.

d. Use the (filtered) interaction p-values to determine the empirical distribution of null p-values.

2. Identify the largest additive eQTL for the probe (irrespective of effect size). Regress out the effect of the additive loci and use the adjusted phenotype for 8df and 4df model analysis. Results reported are the 8df and 4df of original pair and the empirical p-values from genome-wide analysis fitting each of the two snps as fixed. As before these are determined by filtering out the snps on the same chromosome as the original fixed snp and also within +/- 5MB of the second snp.

2. Prediction

Of the 501 snp pairs, 484 have both snps in the EGCUT data. Most of the Inchianti snps need to be imputed, but we expect most to pass filtering. For pairs without and Inchiamti snp I propose using the snp with the largest additive effect in the egcut data.

For each pair;

a. Predict the phenotype in egcut data using a predictor with effects estimated from

4df model (estimated in BSGS)
8df model (estimated in BSGS)
1df model (estimated in BSGS using the Inchianti snp)

3 Results

4 Methods