

Assignment 3. eQTL analysis

Please note: each student should solve the assignment independently, although general discussion between students is allowed. Copying code and solutions is forbidden.

In this assignment, you will investigate the genetic basis of immune DCs' transcriptional responses to pathogenic-like stimulations in BXD recombinant inbred lines.

Choose *one* gene response dataset from the moodle website. In your input matrix, rows are annotated with gene identifier; columns are BXD strain names.

Apply the following preprocessing steps:

- (1) Use Genotyping file from assignment 2; use only representative genomic loci by filtering those loci that have exactly the same information as neighboring loci across the BXD strains under study.
- (2) Gene expression data should be averaged across different individuals of the same strain before applying the association test.

Next, run an association test on all genes using each of your selected representative SNPs in the genotyping file. For each gene, select those eQTLs that are best associated with the expression of the gene (genes with weak associations for all SNPs should be excluded).

Explain the way you chose the final collection of eQTLs and discuss their statistical significance (due to multiple testing, most results may be insignificant. In this part don't use permutation, but multiple testing correction).

Report your results:

- 1) How many different significant eQTLs? How many of them are *cis*-acting and how many are *trans*-acting? For this assignment, we consider eQTLs to be *cis*-acting when the distance between the SNP and the gene is 2Mbp or less.
- 2) Plot the number of genes associated with each eQTL across the genome.
- 3) Compare the distribution of association P-values scans between *cis*-associated genes and *trans*-associated genes.
- 4) Create a scatter plot visualization (as seen in lecture) and indicate *trans*-acting hotspots and clarity of *cis*-acting eQTLs within it.

The final report should include the required visualizations and a detailed discussion of your data processing steps and results.

Submission Guidelines:

- 1) Submit a zip file named hw3_<your_id>, for example, if your id is 123456789 the submission file should be hw3_123456789.zip
- 2) Provide your code and report.
- 3) The report needs to be submitted as a PDF file.
- 4) Add the results of your code to the report, don't make us run it.
- 5) We advise you to submit a clean and organized code. In case of wrong answers, it will assist us in finding the cause and reduce points deduction.

Submission deadline:

Option 1: 5/7/2023 - in order to get grade and feedback before the final project submission

Option 2: with final project submission, estimated to be during September.