

In this assignment, you are to analyze gene expression data from Affymetrix microarrays. You are to preprocess the data, conduct QA/QC, generate a short list of “interesting” genes and annotate the list of interesting genes as further specified below.

You are to find your own data set for this assignment. We encourage you to use data from your own research, other researchers, or students. Affymetrix microarray data also can be obtained from the Gene Expression Omnibus at [www.ncbi.nlm.nih.gov/geo/](http://www.ncbi.nlm.nih.gov/geo/).

### 1. Pre-Process the Data

- a) Conduct all steps necessary to pre-process your microarray data, including conducting background adjustment, normalization, summarization and quality assessment/quality control.
- b) Generate appropriate plots and summary statistics to justify your choice of methods and demonstrate the effectiveness of the pre-processing.

### 2. Identify “Interesting” Genes

Identify genes that are differentially expressed (in one group or between two or more groups) with gene filtering and multiple testing procedures.

- a) First, use simple selection criteria or tests to reduce the number genes to consider
- b) Second, apply multiple testing procedures to develop your short list of interesting genes. Specify your parameters of interest, null and alternative hypotheses, the test statistic and null distribution to be used to test the hypotheses, the Type I error rate you will control and the procedure for controlling your selected Type I error rate.
- c) Summarize the results of the filtering and testing procedures and compile a short list of interesting genes.

### 3. Annotate the “Interesting” Genes

- a) Find the names of your interesting genes and identify their GenBank accession numbers.
- b) Generate an HTML report of your analyses. The report should include links to databases such as PubMed, GenBank, LocusLink for each your interesting genes, as well as other information that would help other researchers use your results (e.g., such as the chromosome number and location on the chromosome of each interesting gene).
- c) Use Bioconductor to find publications about your interesting genes. Are any pairs of your genes cited in the same articles?

### SUBMITTAL

Write up your analysis as a paper for submittal to a scientific publication, including Introduction, Methods, Results, and Discussion sections. Limit the main text including figures to 5 to 8 pages. Additional supporting materials may be included as an appendix. Please include your R code as an appendix.