

BioSys PhD - Statistical Methods for Microarray Data

EXERCISES

Consider a dataset consisting of 10 slides (5 - DIS; 5 - NOR; DIS correspond to odd slides) with 23064 genes each. The aim is to identify marker genes and proteins showing robust differential expression in human Disease-vs-Normal cells. Frequently, `.txt` files with microarray data are given already normalized. That is the case of our data, which is already preprocessed and available at `Disease.txt`.

1. Read and visualize the data.
2. Apply `limma` package to the data:
 - (a) Create the appropriate design matrix, as well as a linear model fitting for each gene;
create a contrast matrix;
 - (b) Compute the empirical Bayes statistics for differential expression;
 - (c) Obtain a summary table for the top genes;
 - (d) How many genes are chosen genes for $B > 0$?
 - (e) Highlight these genes in the volcano plot.
3. Apply `RankProd` package to the data:
 - (a) Create the vector containing the class labels of the samples;
 - (b) Compute the RP statistics for differential expression;
 - (c) Obtain a summary table for the top genes;
 - (d) For a fixed FDR, represent graphically the RP values;
 - (e) How many genes are chosen genes for that FDR?