Problem #2: Apolipoprotein AI and Liver Expression in Mice.

From Canvas, download microarray data (**module2-hackathon2-arraydata.txt**) and gene names (**module2-hackathon2-genenames.txt**) in the **module2-hackathon2** directory. This is a gene expression study in liver with eight mice in the control group (C57Bl/6 strain) and eight mice with the apolipoprotein AI (apo AI) gene knocked out. The data have already been pre-processed and log transformed. The gene names have been saved in a separate file since there are duplicate names (keep any eye on that). In all parts below, list the genes by name.

We are interested in testing whether there are genes that are differentially expressed between the control and knock-out group for each gene. Obtain the p-values from a two sided t-test for differential expression using permutations (B=12870 possibilities), where the labels of the two mouse groups are shuffled. How many genes are significant at the 0.01 level? Print out a table with the top 10 genes, their t-statistics, permutation p-value and fold change.

HINTS:

* You many need to use the gtools packing (using install.packages).
* Use blank.lines.skip = FALSE when reading in the gene names. Some genes have no annotation, if this option is TRUE, then those genes will be skipped.
* apply() may be handy to perform operations on each gene.
* To get all permutations use the “combinations” function in gtools. This can take up to 3 hours. Try a few permutations first to see if it is working. Do not use parallel computing methods in R.