Functional analysis of differentially expressed genes

Juan R Gonzalez

BRGE - Bioinformatics Research Group in Epidemiology
Center for Research in Environmental Epidemiology (CREAL)
http://www.creal.cat/jrgonzalez/software.htm

For the exercise we are going to use the packages `tweeDEseq` and `tweeDEseqCountData` and the `tweeDE` object `resPT`, obtained as a result of the exercise in the previous session (file 'resPT.Rdata'), and resulting from analyzing differential expression with `tweeDEseq` on a subset of genes for RNA-seq data published by Pickrell *et al.* (*Nature*, 464:768-772, 2010) consisting of female and male samples derived from lymphoblastoid cell lines of unrelated Nigerian individuals.

1. Load the mentioned packages and the `resPT` object and select a subset of DE genes with minimum fold-change of 1.5 at a maximum FDR of 5 %.

2. Create a subset of the genes with sex-specific expression formed by those that occur in the `resPT` object (use the code seen in previous exercise).

3. Assess by a one-tailed Fisher's exact test whether genes with sex-specific expression are enriched in our list of DE genes.

4. Do the same with the housekeeping genes (the list of those genes are in the object `hkGenes` that can be obtained by executing `data(hkGenes)`) Do you get the same result? Why?

5. Using the package `GOstats` and the human annotation package `org.Hs.eg.db`, perform a GO enrichment analysis on the subset of DE genes with the conditional hypergeometric test. Human annotations are anchored at Entrez Gene IDs and, for this reason, you will have to translate first the Ensembl Gene IDs obtained from `resPT` into Entrez IDs by using the appropriate mapping in `org.Hs.eg.db`. Do the enriched GO terms seem to be related to sex-specific expression?