

RNA-seq analysis - Final Exercise

Juan R Gonzalez

*BRGE - Bioinformatics Research Group in Epidemiology
Barcelona Institute for Global Health, ISGlobal (former CREAL)
<http://www.creal.cat/jrgonzalez/software.htm>*

File `gbm.Rdata` is an R file containing data about an RNA-seq experiment from The Cancer Genome Atlas project (cancergenome.nih.gov). The file includes a `SummarizedExperiment` object called `rse`. This object corresponds to a RNA-seq count data of 174 samples diagnosed with glioblastoma as well as information about different phenotypes. In 2011, TCGA scientists described four types distinct subtypes of this tumor (Classical, Mesenchymal, Neural and Proneural) by using clinical and gene expression data obtained from microarray data. We are interested in detecting those genes that are associated with each of the 4 subtypes. The variable 'Cluster' in the `rse` object (remember that phenotypic data can be retrieved by using `colData` function) contains information about each subtype (Classical can be considered as the reference category). NOTE: There are individuals having other categories in the 'Cluster' variable than Classical, Mesenchymal, Neural and Proneural. Those individuals having missing values or different subtypes than these four types must be removed from the analysis.

Table of raw counts can be retrieved using `assays` function (use data in the 'raw_counts' element of the list). Perform all the required steps to determine those genes that are differentially expressed for each subtype and investigate which are the biological pathways (e.g. gene sets) that are associated with those genes.

NOTE: See <http://cancergenome.nih.gov/researchhighlights/researchbriefs/foursubtypes> for further information about the clusters and the related paper. This may help when interpreting the results.

TO DELIVER: Write a report describing the analyses you have carried out, interpret the results and include some discussion making reference to the tables and figures you think are necessary to be showed in the report. Write the R code in the Appendix. Save it in a pdf file and upload this file to the required link. The report can also be created by using knitr or Sweave. In that case the Appendix is not necessary since the chunks will be showed (e.g. use `echo=TRUE` in the whole document).