## Differential expression analyses with DESeq2 and snakemake

Your job for the last laboratory is to re-organize the code for DEG analyses we looked at in a snakemake pipeline.

The input of the pipeline will be:

1. a matrix with read counts from a bulk RNAseq experiments (/home/stud0/aibh\_integromics/05\_deseq/data/counts\_HXO.tsv.gz)
2. the lengths of the involved genes (/home/stud0/aibh\_integromics/05\_deseq/data/gene\_lengths.txt)
3. a sample sheet where rownames are sample ids (corresponding to columns in the count matrix) and there are other columns of possible covariates for the DESeq2 analyses/groups on which we want to compute DEG (/home/stud0/aibh\_integromics/05\_deseq/data/samplesheet.tsv)

The required outputs are:

1. a PCA plot of all the samples colored according to all the columns in the sample sheet
2. a volcano plot and a tsv file with the differential analyses results on any possible combinations of the samples annotation found in the sample sheet, where the user will need to be able to define these combinations depending on the file name he asks snakemake to produce
3. a plot showing Hallmarks GSEA enrichment results alongside a tsv with the same information, the GSEA analysis should be doable for all the possible DEG results of point 2 (you can use /home/stud0/aibh\_integromics/05\_deseq/data/hallm.rds as input for this rule)
4. a normalized matrix of robustly corrected FPKM for all the samples

You can use the same thresholds on n\_reads/n\_samples we adopted for our previous lab!