## Exam III Outline – Gene Expression Analysis

1. Different ways of coding explanatory variables in linear models
2. Relationship between two-sample t-test and a linear model with coded variables using *treatment contrasts* (x = 0 and x = 1).
3. RNA-seq biology (read counts)
4. Processing of RNA-seq data: FPKM/RPKM, TMM, and TMM
5. Retrieving data using the *UCSCXenaTools* R package
6. Understanding the format of RNA-seq data and accessing clinical (phenotype) variables
7. Constructing side-by-side boxplots comparing the gene expression across groups
8. Identifying differentially expressed probes/genes using the *limma* package and understanding the false discovery rate (FDR)
9. Generating heatmaps
10. Converting a probe name to the corresponding gene symbol and vice versa using the probe map data provided by UCSC Xena
11. Classification using *k*-nearest neighbors *(knn)*, including leave-one-out cross-validation, optimization, and making predictions in a test dataset.
12. Perform a gene set enrichment analysis using DAVID, based on a list of genes