Problem set 7

Online databases and tools

Here is a small RNA-seq dataset of 151 cancer patients downloaded from The Cancer Genome Atlas and the group labels. We will analyze this dataset using online tool instead of on R program locally.

Q1: Perform differential expression analysis using the DESeq2 module on the GenePattern platform (registration needed but free to use). Show a screen capture of how you set up the parameters for the analysis on the website.

Q2: Show the fold changes and p-values for the top 5 up-regulated genes in Group2 (sorted by p-values).

Q3: Show the fold changes and p-values for the top 5 down-regulated genes in Group2 (sorted by p-values).

Q4: What is the gene ID system used in this dataset?

Q5: What are the gene symbols for the top 5 up-regulated genes in Group 2?

Q6: What are the enriched functions among significantly up-regulated genes in Group 2?

For the following questions, you need to find an appropriate online platform to solve them. Explain the tool you used along with your answer for each question.

Q7: What is the binding motif for SMAD2 transcription factor?

Q8: Which miRNAs can bind to the 3'UTR of LIN28B genes to regulate its expression?

Q9: Which proteins can bind to ACTB?

Q10: If you want a list of genes involved in **interferon gamma response**, from which online databases can you get them?

This paper used microarray to identify gene markers for keratinocyte colony types: holoclone, meroclone, and paraclone. The author has uploaded their data on Gene Expression Omnibus.

Q11: What is the GEO ID for the data generated by this paper?

Q12: In addition to microarray data, does this paper provide any other data? What are they?

Q13: Re-analyze the microarray data using GEO2R function on the GEO website. What are the top 3 up-regulated genes in holoclones compared to other colony types?