

## Homework 1: RNA-sequencing Analysis (20 points total)

*If applicable, please remember to show your work for any calculations or code!*

1. Describe the main steps of a full differential expression RNA-seq analysis, up to and including performing differential gene analysis. For each of the main steps, give an example of the tool you would use (if relevant) and most importantly the purpose of what is done computationally. (10 points)
2. What kind of normalization method should you perform to your count matrix if you choose to use DESeq2 for your differential gene expression analysis? (1 point)
3. Say you are looking at two genes in an experiment, *Gene A* and *Gene B*. If the  $\log_2$  fold change of *Gene A* to *Gene B* is a value of 3, what is the absolute difference in gene expression between *Gene A* and *Gene B* in terms of **A**? (2 points)
4. Given the gene expression table below with read counts for Rep1, Rep2, and Rep3, which two samples are the **most similar** to one another? (*Hint*: Calculate TPMs and use a scale factor of 10 instead of 1 million) (7 points)

Gene name	Rep 1	Rep 2	Rep 3
A (2kb)	10	12	30
B (4kb)	20	25	60
C (1kb)	5	8	15
D (10kb)	0	0	1