

CMM 262 2022

Homework 2: RNA-seq and single cell RNA-seq (Priya Pantham and Robert Morey)

Instructions

Answer [the following questions](#) in your own words and upload a PDF of your answers to Gradescope. Make sure to write your name and PID at the start of your answers. This assignment is due **2/10/22 at 9:00AM**.

Part 1: RNA-seq

- For the following main steps of a full differential expression RNA-seq analysis, explain the purpose of what is done computationally and give an example of a tool you would use (if relevant). (4 points)
 - QC the FASTQ files
 - Align the FASTQ files
 - Sort and index the aligned reads
 - QC the bam file
 - Count reads for every gene
 - Normalize the counts
 - Perform differential expression analysis
 - Visualize the results
- When using DESeq2 to perform your differential expression analysis do you need to normalize your count matrix? If not, why not? (1 point)
- Say you are looking at two genes in an experiment, *Gene A* and *Gene B*. If the log2 fold change of *Gene A* to *Gene B* is 3 ($\log_2(\frac{\text{Gene A}}{\text{Gene B}}) = 3$), what is the difference in gene expression between *Gene A* and *Gene B* in terms of A? (1 point)
- 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) (3 points)

Gene name	Rep 1	Rep 2	Rep 3
A (2kb)	15	3	30
B (4kb)	10	22	60
C (1kb)	5	10	15
D (10kb)	0	0	1

Part 2: Single cell RNA-seq

5. Describe the experimental differences between droplet based methods and physical separation methods. Start by explaining each. Then discuss when it might be better to use one or the other. (3 point)
6. Describe what a doublet is and how it is generated. Why do you not want doublets in your dataset when you perform analysis? (2 points)
7. When analyzing scRNA-seq data, how do you know which cell a read comes from? (1 point)
8. When preprocessing scRNA-seq data how do you control for technical and biological covariates? Which three metrics do you perform QC on and what does each control for? (3 points)
9. Come up with an experiment in which you'd prefer to use scRNA-seq rather than RNA-seq. Justify why you believe this assay would be preferable for this experiment. (2 points)