

RNA-seq Homework

BMSC8203, Spring 2020

Instructor: Kelsey Keith

Please send me your answers in either a word document or a PDF with your LAST NAME in the file name and NO SPACES. If you need any clarification on the assignment, please email me at kelsey.h.keith@gmail.com.

Multiple Choice

Please indicate your choice by changing the color of the answer or making it bold.

1. What is the MINIMUM number of samples you need per condition for RNA-seq?
 - a. 1
 - b. 2
 - c. 3
 - d. 6
2. When would you need to use ribosomal RNA depletion to make an RNA-seq library instead of using poly-A capture?
 - a. When you want to look at only mature messenger RNAs with a poly-A tail.
 - b. When you want to study non-coding RNAs that don't have a poly-A tail.
 - c. You don't need to get rid of ribosomal RNA before making an RNA-seq library.
 - d. When you want to study alternative splicing of RNA transcripts.
3. Which of the following statements about sequencing depth is correct?
 - a. Genes that don't have sufficient coverage should be excluded from analysis.
 - b. Sequencing depth is not an important consideration in NGS analysis.
 - c. You cannot ask sequencing facilities to resequence your RNA-seq library if you don't have enough depth.
 - d. As sequencing depth decreases, confidence in your data increases.
4. What do the results of differential expression analysis tell you?
 - a. Whether the sequencing depth was sufficient for the analysis.
 - b. Which genes stayed the same, were differentially expressed, between experimental conditions.
 - c. Whether your experiment changed transcript splicing.
 - d. Which genes increased or decreased, were differentially expressed, between experimental conditions.
5. What is the correct interpretation of an MA plot?
 - a. As a gene moves farther away from the horizontal line running through 0 on the y-axis, the greater the change in gene expression between conditions.
 - b. The closer a gene is to the horizontal line running through 0 on the y-axis, the greater the change in gene expression between conditions.
 - c. Only genes that have low mean expression can be differentially expressed.
 - d. The MA plot shows the statistical significance of all genes tested.

6. What is the correct interpretation of a volcano plot?
 - a. As you move farther away from 0 on the x-axis, the difference in expression between conditions decreases and as you move up the y-axis the statistical significance of the difference increases, so the points at the top middle of the plot are the most significantly differentially expressed genes.
 - b. As you move farther away from 0 on the x-axis, the difference in expression between conditions decreases and as you move up the y-axis the statistical significance of the difference decreases, so the points at the center bottom of the plot are the most significantly differentially expressed genes.
 - c. As you move farther away from 0 on the x-axis, the difference in expression between conditions increases and as you move up the y-axis the statistical significance of the difference increases, so the points at the top right and left of the plot are the most significantly differentially expressed genes.
 - d. As you move farther away from 0 on the x-axis, the difference in expression between conditions increases and as you move up the y-axis the statistical significance of the difference decreases, so the points at the bottom right and left of the plot are the most significantly differentially expressed genes.
7. Why do pathway analysis?
 - a. Because papers will not be accepted for publication without pathway analysis.
 - b. To aggregate data at a higher level to figure out what biological processes are changing as a result of the experiment.
 - c. To figure out which genes are differentially expressed.
 - d. If there aren't any genes differentially expressed, pathway analysis is another way of looking at the data.
8. Which of the following can you **NOT** analyze using RNA-seq data?
 - a. alternative splicing
 - b. mutations
 - c. changes in gene expression
 - d. methylation
9. What is an advantage of single cell RNA-seq over traditional, bulk RNA-seq?
 - a. You get better coverage over the genes in single cell than you do in bulk.
 - b. Single cell RNA-seq is cheaper than bulk RNA-seq.
 - c. Single cell RNA-seq is cheaper than bulk RNA-seq.
 - d. You can measure different cell types simultaneously in single cell RNA-seq; in bulk RNA-seq you get an average expression from all cell types sequenced.
10. What is a disadvantage of single cell RNA-seq over traditional, bulk RNA-seq?
 - a. Single cell allows you measure more diversity in gene expression than bulk RNA-seq.
 - b. Single cell RNA-seq is sparse; most genes are not measured in most cells.
 - c. You can dissociate and sequence whole tissues in bulk RNA-seq, but you can only sequence cell lines in single cell RNA-seq.
 - d. You can dissociate and sequence whole tissues in single cell RNA-seq and get expression for individual cells.

Short Answer

In this brief short answer section, we'll use the TCC-GUI web app, <https://infinityloop.shinyapps.io/TCC-GUI/>, to examine how changing the minimum read threshold changes the results of the differential expression analysis. For this section you should type a number or a brief sentence next to the question.

To start, go to the Exploratory Analysis tab and under Select Sample Data, select and import the hypoData example dataset, just as we did in class. After importing the data, click the Assign Group button, and move on to the TCC Computation section so you can adjust the minimum read threshold.

1. Set the Filtering Threshold for Low Count Genes slide bar to the following levels, leave all the other options at their default settings, and click Run TCC Computation. For each answer, you should give a single number from the MA Plot section.
 - a. Set the threshold to 0.
 - b. Set the threshold to 10.
 - c. Set the threshold to 30.
2. Why does increasing the minimum read count threshold increase the number of differentially expressed genes?