

# Bioinformatics Homework

BMSC8503, Spring 2020

Instructor: Kelsey Keith

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

For this homework assignment, we'll use the TCC-GUI web app, <https://infinityloop.shinyapps.io/TCC-GUI/>, to examine how changing parameters changes the results of the differential expression analysis. Things you should look for in the web app have been emphasized by writing them in Courier New.

## Questions:

1. 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 tab so you can adjust the minimum read threshold. 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. You should report a single number from the MA Plot tab as your answer.
  - 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? (1-2 sentences)
3. Now go back to the TCC Computation tab. Leaving all other options at their defaults, this time, we'll change the Normalization Method. For the following questions, set the Normalization Method to the given value, click Run TCC Computation, and report a single number from the MA Plot section as your answer.
  - a. Set the Normalization Method to TMM
  - b. Set the Normalization Method to DESeq
4. Now go back to the TCC Computation tab. Leaving all other options at their defaults, this time, we'll change the DEG Identification Method. For the following questions, set the DEG Identification Method to the given value, click Run TCC Computation, and report a single number from the MA Plot section as your answer.
  - a. Set the DEG Identification Method to edgeR
  - b. Set the DEG Identification Method to DESeq
  - c. Set the DEG Identification Method to DESeq2

5. Which of the three parameters, minimum read threshold, normalization method, or identification method, have the largest effect on the number of genes?
6. Why is it important to report exactly what method you used for normalization and differential gene expression identification? (1-2 sentences)