Bioinformatic Exercise: Technical evaluation part 2

Laleh Varghaei

June 20, 2024

1 Question 1:

The written code is working and it sends a warning email if some origins produce more than 10% failed samples. We can also use Cron job scheduler which is a command line tool in order to run the python script once a week.

2 Question 2:

2.1 a.

The workflow should start with the preprocessing and checking the quality of the data. This can be done by FastQC. Then the reads should be aligned to a reference genome using Bowtie2 or STAR. The quantification of gene expression can be done using HTSeq-count, which produce a count matrix where rows represent genes and columns represent samples. For differential expression analysis, DESeq2 can be used to model the comparison between cancer samples and matched adjacent normal samples.

2.2 b.

The first limitation that comes to my mind is a very small sample size. Larger sample sizes are necessary to have sufficient statistical power to detect significant changes. The other limitation can be insufficient sequencing depth that can affect the sensitivity of detecting lowly expressed genes.

2.3 c.

When analysing the heatmap, it is crucial to examine the clustering patterns of samples. Analysing the heatmap A, we should see if healthy samples and patient samples group into distinct clusters. If they do, this means their gene expression profiles are significantly different. Clear clustering shows that there are noticeable differences between the two groups. We should also check for downregulated and upregulated genes and also outliers. In the heatmap of DE analyses (B), the focus should be on genes that show significant differential expression. These genes are likely to be biologically relevant and may play important roles in cancer development and progression.

2.4   Question 3

The solution for me is planning and prioritizing. I would create a clear plan for completing the first three projects and allocating a portion of my time each day to address the reviewer comments of the fourth user. I would also build in some buffer time to ensure that if I do not finish a task on schedule I have time to complete it. For example, I would dedicate two days to each of the three projects, while allocating a portion of my time each day to the fourth user's needs. On the last day of the week I would review all tasks to ensure everything is completed. If there are any remaining tasks, I would use this buffer day to address any leftovers.

2.5   Question 4

First, I would ensure that I fully understand the goals and requirements of the project by asking detailed questions during the meeting. After that, I would conduct a literature review to familiarize myself with the methodology and best practices for this project. I would also consult with colleagues who have done similar analyses and ask for any advice on the methodology that I should apply. Additionally, I would maintain regular communication with the group, providing updates on my progress and any challenges that I may encounter.

2.6   Question 5

I would assure them that I will investigate the problem. Then, I would start by carefully reviewing the entire analysis and calculations to identify any errors. I would analyze the steps and reasons behind the error to understand its cause. After identifying and correcting the error, I would ensure that the corrected analysis produces accurate results. I would then report the new results, explaining how the previous miscalculation affected the overall process and results and detailing the differences in the corrected analysis. Finally, I would apologize for my mistake, emphasize my commitment to carefully reviewing the steps and errors that led to this issue and express my intent to learn from it.