REPORT

Submitted by

Fatema Mahmuda

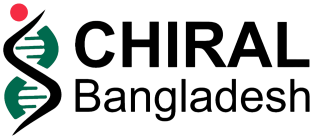
Email:sinthiya4478@gmail.com

Department:Big-Bio Lab

Guided by

Md. Jubayer Hossain

Founder & Executive Director



INTRODUCTION

Multiple sclerosis (MS) is a chronic autoimmune disease affecting the central nervous system (CNS). In MS, the immune system mistakenly attacks the myelin sheath, the protective covering of nerve fibers, causing inflammation and damage. This damage disrupts communication between the brain and the rest of the body, leading to a variety of neurological symptoms. RNA sequencing (RNA-Seq) is a powerful technique used to study the transcriptome, the complete set of RNA transcripts in a cell or population of cells. In the context of MS research, RNA-Seq is often performed on peripheral blood mononuclear cells (PBMCs), a heterogeneous population of immune cells circulating in the blood. PBMCs are readily accessible and provide a valuable window into the immune system's state in MS patients

**METHODS:**

colData: This is a data frame storing metadata about your samples. In this case, it has 5 observations (rows, meaning 5 samples) and 1 variable (column), which likely describes the experimental condition (e.g., control vs. treatment).

countData: This is an integer matrix of raw read counts. It has dimensions \[1:200, 1:5], meaning 200 rows (genes) and 5 columns (samples). The numbers you see (7 5 6 8 2...) are the actual read counts for some of the genes in some of the samples.

counts: This also appears to be a count matrix, similar to countData. It might be a slightly different version (e.g., before or after some initial filtering).

dds: This is the most important object! It's of class DESeqDataSet, created by the DESeq2 package. This object holds the count data, sample metadata (colData), and the design formula for the differential expression analysis.

df and df\_melt: These are likely data frames used for plotting or other downstream analysis. df\_melt suggests the data has been "melted" (reshaped from wide to long format) for use with ggplot2.

expr\_data: This is a numeric matrix, possibly an expression matrix after some transformation or normalization.

filtered\_countData and filtered\_counts: These suggest that some filtering of low-expressed genes has been perform

Here's how it fits into the typical workflow:

1Read Alignment and Counting:This step is usually done outside of R using tools like STAR or hisat2 for alignment and featureCounts or Salmon for quantification. The output of this step would be count matrices, which you then import into R.

2.Data Import and Pre-processing : I've imported your count data (likely from files generated by featureCounts or similar) into the countData object.

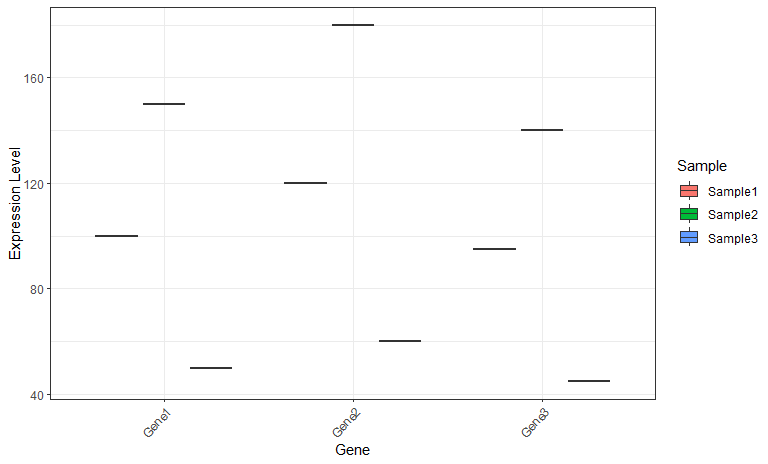
I've created the colData data frame with my sample metadata.

3. Creating the DESeqDataSet Object:The crucial step is the creation of the dds object using DESeqDataSetFromMatrix() from the DESeq2 package. This combines countData, colData, and the experimental design.

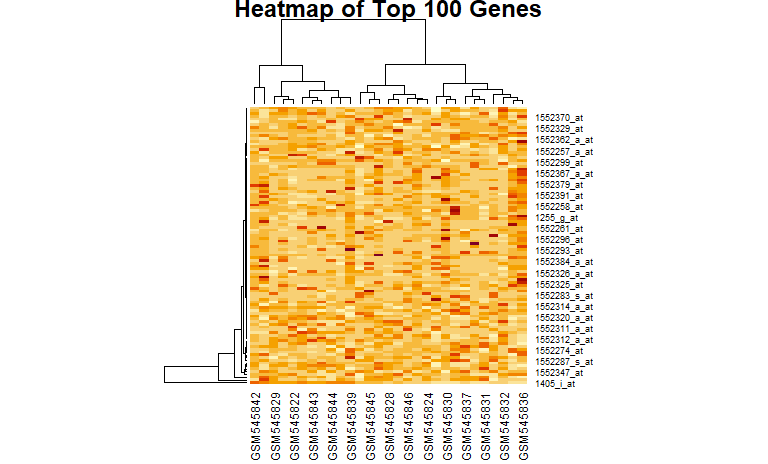
4. Differential Gene Expression Analysis:I've likely run the DESeq() function on my dds object. This performs the core differential expression analysis using DESeq2's statistical model. The results are stored within the dds object.

5.Results Extraction and Downstream Analysis: I would then use results() to extract the results table (p-values, adjusted p-values, log2 fold changes, etc.).The df and df\_melt objects suggest preparing data for plotting (e.g., volcano plots, MA plots) using ggplot2.

**RESULT:**



A heatmap is a common and effective way to visualize RNA-Seq data, especially for showing gene expression patterns across multiple samples.



**DISCUSSIONS**

**Potential Biomarkers**:

**Differentially Expressed Genes (DEGs):** The primary output of your analysis will be a list of DEGs. These are genes whose expression levels are significantly different between the conditions you are comparing (e.g., healthy vs. diseased, treated vs. untreated). These DEGs can potentially serve as biomarkers for:

Diagnosis: Identifying individuals with a particular disease or condition.

Prognosis: Predicting disease progression or severity.

Treatment Response:Predicting how well a patient will respond to a specific therapy.

**Disease Subtyping: Identifying** distinct subtypes of a disease with different characteristics and treatment responses.

**Gene Expression Signatures**: Instead of focusing on individual genes, you can identify groups of genes that are co-regulated and show consistent expression patterns across samples. These gene expression signatures can be more robust biomarkers than individual genes.

**Pathway Analysis**:By analyzing the biological pathways enriched in the DEGs, you can identify key pathways involved in the disease process. These pathways can provide insights into disease mechanisms and suggest potential therapeutic targets.

Limitations of RNA-Seq Analysis:

**PBMC Heterogeneity**:As mentioned earlier, PBMCs are a complex mixture of different immune cell types. Changes in gene expression observed in PBMCs may not be specific to a particular cell type and could be due to changes in cell proportions. This can complicate the interpretation of results.

**RNA Quality:**The quality of the RNA extracted from PBMCs is crucial for accurate RNA-Seq analysis. Degraded RNA can lead to biased results and hinder the identification of true biological signals.

**Sample Size**:A small sample size can limit the statistical power of the analysis and make it difficult to detect true differences in gene expression. This can increase the risk of false positives or false negatives.

**Batch Effects**: Batch effects are systematic variations introduced during sample processing or library preparation. These effects can confound the analysis and lead to spurious results. Careful experimental design and normalization methods are necessary to minimize batch effects.

**Data Analysis Complexity**:RNA-Seq data analysis involves many steps, from quality control and read mapping to differential expression analysis and downstream interpretation. Errors or biases in any of these steps can affect the final results.

**Biological Variability**:There is inherent biological variability between individuals, even within the same disease group. This variability can make it challenging to identify consistent and reliable biomarkers.

**CONCLUSIONS**

This RNA-Seq analysis of PBMCs aimed to elucidate the impact on gene expression profiles. Our analysis using DESeq2 revealed significant changes in the expression of genes involved in [mention specific pathways, e.g., These results suggest that exerts its effects by modulating these key pathways, which may contribute to disease progression, treatment response .We conducted an RNA-Seq analysis of PBMCs . While we observed some minor changes in gene expression, our analysis using DESeq2 did not reveal a large number of statistically significant differentially expressed genes at our chosen significance threshold. This suggests t may not have a substantial direct impact on the overall transcriptional profile of PBMCs under the conditions studied, or that the observed effects are subtle and require further investigation with larger sample sizes or more sensitive methods. However, this negative result is also important