# Title: RNA-Seq Analysis of Peripheral Blood Mononuclear Cells in Multiple Sclerosis Patients and healthy control

## Introduction

Multiple sclerosis (MS) is a chronic autoimmune disease of the central nervous system (CNS) characterized by inflammation, demyelination, and neurodegeneration. It is a highly heterogeneous disease with an unpredictable clinical course, varying from mild to severe disability. Although the precise etiology of MS remains elusive, it is widely accepted that a combination of genetic predisposition, environmental factors, and immune dysregulation contribute to its pathogenesis. Immune cells, particularly T cells, B cells, and monocytes, play a central role in driving the inflammatory processes in MS, making the study of immune profiles a critical area of investigation.

Peripheral blood mononuclear cells (PBMCs) represent a diverse population of immune cells, including lymphocytes and monocytes, which can provide valuable insights into systemic immune dysregulation in MS. PBMCs are accessible, minimally invasive to obtain, and serve as a proxy for understanding the immune mechanisms underlying MS. The study of PBMCs has yielded important findings regarding cytokine dysregulation, immune cell phenotypes, and transcriptomic changes associated with MS. However, given the complexity of immune interactions, there is a need for more comprehensive approaches to capture the full spectrum of molecular alterations in PBMCs from MS patients.

RNA sequencing (RNA-seq) has emerged as a powerful tool for transcriptomic profiling, offering high sensitivity and resolution in quantifying gene expression. Unlike traditional gene expression methods, such as microarrays, RNA-seq allows for the detection of low-abundance transcripts, alternative splicing events, and novel transcripts. RNA-seq has been instrumental in identifying molecular signatures and pathways associated with various diseases, including MS. By analyzing the transcriptome of PBMCs, RNA-seq provides an opportunity to uncover key dysregulated genes and pathways involved in MS pathogenesis.

Several studies have used RNA-seq to investigate transcriptomic changes in PBMCs from MS patients, revealing critical insights into disease mechanisms. For instance, altered expression of genes involved in immune signaling, mitochondrial function, and cellular stress responses have been reported. Additionally, transcriptomic analyses have shed light on the role of specific immune cell subsets, such as CD4+ and CD8+ T cells, in MS progression. Furthermore, these studies have contributed to the identification of potential biomarkers for MS diagnosis, prognosis, and treatment response. Despite these advances, challenges remain in interpreting RNA-seq data, particularly in the context of the immune system's complexity and inter-patient variability.

In MS, disease heterogeneity poses a significant challenge to the identification of universal molecular signatures. Factors such as disease subtype (e.g., relapsing-remitting, secondary progressive, and primary progressive MS), treatment status, and comorbidities can influence gene expression profiles in PBMCs. Additionally, the dynamic nature of the immune response, characterized by fluctuations in immune cell composition and activation states, further complicates transcriptomic analyses. Addressing these challenges requires careful experimental design, robust analytical methods, and validation in independent cohorts.

The application of RNA-seq to PBMCs in MS holds promise for advancing our understanding of disease biology and identifying novel therapeutic targets. By integrating RNA-seq data with clinical, genetic, and epigenetic information, researchers can develop a more comprehensive view of the molecular networks driving MS. Furthermore, longitudinal studies that track transcriptomic changes over time and in response to therapy can provide valuable insights into disease progression and treatment efficacy.

In this study, we performed RNA-seq analysis of PBMCs from MS patients to investigate transcriptomic alterations associated with disease pathogenesis. We aimed to identify differentially expressed genes (DEGs), enriched pathways, and potential biomarkers that could inform the development of targeted therapies. Our study included MS patients with different disease subtypes and treatment statuses, enabling us to explore the impact of these variables on gene expression. By leveraging advanced bioinformatics approaches, we sought to address the challenges associated with transcriptomic analysis in a heterogeneous disease like MS.

## Materials and Methods

### Data Collection

The RNA-Seq dataset of peripheral blood mononuclear cells (PBMCs) from multiple sclerosis (MS) patients was obtained from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) repository. The dataset corresponds to the accession ID **GSE21942**. The dataset includes raw counts of RNA-Seq data from PBMCs of MS patients and healthy controls.

### Data Preprocessing

Data preprocessing was conducted in R Studio (version ≥4.0.0). The following steps outline the process:

1. **Downloading the Dataset**: The GEOquery package was used to retrieve the dataset.

```
library(GEOquery)
#import dataset
gset <- getGEO("GSE21942", GSEMatrix = TRUE, AnnotGPL=TRUE)

#extract expression matrix
expression_matrix_data<- exprs(gset[[1]])

#extract meta data
metadata<- pData(gset[[1]])
```

2. **Normalization**: Normalization of raw counts was done using the log2 transformation to correct for library size differences.

```
ex <- exprs(gset)
qx <- as.numeric(quantile(ex, c(0., 0.25, 0.5, 0.75, 0.99, 1.0), na.rm=T))
LogC <- (qx[5] > 100) ||
  (qx[6]-qx[1] > 50 && qx[2] > 0)
if (LogC) { ex[which(ex <= 0)] <- NaN
exprs(gset) <- log2(ex) }
```

3. **Differential Gene Expression Analysis:** Import gene tables and identify top 10 over expressed and under expressed significant genes and export it in csv files.

```
#import gene table
gene_table <- read_table("data/GSE21942.top.table.tsv")
#convert data structure
gene_table$ID <- as.factor(gene_table$ID)
gene_table$Genesymbol <- as.factor(gene_table$Genesymbol)
gene_table$Genetitle <- as.factor(gene_table$Genetitle)
head(gene_table)

#identify 10 over expressed significant genes
over_expressed_genes <- gene_table |>
  filter(adjPVal <= 0.05 & logFC > 1 ) |>
  arrange(desc(logFC)) |>
  head(10)
write.xlsx(over_expressed_genes, "outputs/over_expressed_top10_genes.xlsx")
# identify 10 under expressed significant genes
under_expressed_genes <- gene_table |>
  filter(adjPVal <= 0.05 & logFC < -1  ) |>
  arrange(desc(logFC)) |>
  head(10)
write.xlsx(under_expressed_genes, "outputs/under_expressed_top10_genes.xlsx")
```

**#join significant gene**
significant_gene <- bind_rows(over_expressed_genes, under_expressed_genes)

**#export into csv**
write.csv(significant_gene, "outputs/top20_significant_genes.csv")

1. **Experimental Design**: Metadata containing sample information (e.g., patient/control status) was used to construct a design matrix.
2. design <- model.matrix(~0 + metadata$group)
3. colnames(design) <- levels(metadata$group)

## Tools and Resources

- **R Packages**: GEOquery, limma, clusterProfiler, EnhancedVolcano, pheatmap.

## Result:

### 1. 20 significant genes

Table 1:Top 20 significant genes

| ID | adjPVal | PValue | t | B | logFC | Genesymbol | Genetitle |
|---|---|---|---|---|---|---|---|
| AFFX-r2-Bs-dap-M_at | 0.0000485 | 7.42E-07 | 6.31 | 5.92 | 4.47 | | |
| AFFX-r2-Bs-dap-5_at | 0.000117 | 0.00000256 | 5.85 | 4.71 | 4.17 | | |
| AFFX-r2-Bs-dap-3_at | 0.0000342 | 4.38E-07 | 6.51 | 6.43 | 4.16 | | |
| AFFX-DapX-3_at | 0.0000258 | 2.94E-07 | 6.66 | 6.81 | 4.15 | | |
| AFFX-DapX-M_at | 0.0000485 | 7.41E-07 | 6.31 | 5.92 | 4.1 | | |
| AFFX-r2-Bs-thr-3_s_at | 0.000014 | 1.17E-07 | 7.01 | 7.71 | 3.99 | | |
| 201123_s_at | 0.000198 | 0.00000527 | 5.59 | 4.01 | 3.3 | EIF5A | eukaryotic |
| AFFX-DapX-5_at | 0.000136 | 0.00000315 | 5.78 | 4.51 | 3.26 | | |
| AFFX-ThrX-3_at | 0.000014 | 1.17E-07 | 7.01 | 7.71 | 3.14 | | |
| 228697_at | 8.05E-08 | 5.44E-12 | 11.3 | 17.2 | 3.1 | HINT3 | histidine |
| 244515_at | 0.00000113 | 2.56E-09 | -8.52 | 11.4 | -1.01 | | |
| 222737_s_at | 0.0000832 | 0.00000158 | -6.03 | 5.18 | -1.01 | BRD7 | bromodomain |
| 236533_at | 0.00251 | 0.000158 | -4.35 | 0.727 | -1.01 | ASAP1 | ArfGAP |
| 236368_at | 0.0027 | 0.000174 | -4.32 | 0.634 | -1.01 | KIAA0368 | KIAA0368 |
| 240636_at | 0.00846 | 0.000777 | -3.76 | -0.798 | -1.01 | | |
| 225173_at | 0.00933 | 0.000882 | -3.71 | -0.919 | -1.01 | ARHGAP18 | Rho |
| 230170_at | 4.42E-07 | 4.61E-10 | -9.24 | 13 | -1.02 | OSM | oncostatin |
| 222891_s_at | 0.0000076 | 4.85E-08 | -7.34 | 8.56 | -1.02 | BCL11A | B-cell |
| 1561923_a_at | 0.0000453 | 6.73E-07 | -6.35 | 6.01 | -1.02 | SF3B6 | splicing |
| 228818_at | 0.000869 | 0.0000382 | -4.87 | 2.09 | -1.02 | | |

### 2. Table of 10 overexpressed gene

Table 2: Top10 over expressed genes

| ID | adjPVal | PValue | t | B | logFC | Genesymbol | Genetitle |
|---|---|---|---|---|---|---|---|
| AFFX-r2-Bs-dap-M_at | 0.0000485 | 7.42E-07 | 6.31 | 5.92 | 4.47 | | |
| AFFX-r2-Bs-dap-5_at | 0.000117 | 0.00000256 | 5.85 | 4.71 | 4.17 | | |
| AFFX-r2-Bs-dap-3_at | 0.0000342 | 4.38E-07 | 6.51 | 6.43 | 4.16 | | |
| AFFX-DapX-3_at | 0.0000258 | 2.94E-07 | 6.66 | 6.81 | 4.15 | | |
| AFFX-DapX-M_at | 0.0000485 | 7.41E-07 | 6.31 | 5.92 | 4.1 | | |
| AFFX-r2-Bs-thr-3_s_at | 0.000014 | 1.17E-07 | 7.01 | 7.71 | 3.99 | | |
| 201123_s_at | 0.000198 | 0.00000527 | 5.59 | 4.01 | 3.3 | EIF5A | eukaryotic |
| AFFX-DapX-5_at | 0.000136 | 0.00000315 | 5.78 | 4.51 | 3.26 | | |
| AFFX-ThrX-3_at | 0.000014 | 1.17E-07 | 7.01 | 7.71 | 3.14 | | |
| 228697_at | 8.05E-08 | 5.44E-12 | 11.3 | 17.2 | 3.1 | HINT3 | histidine |

### 3. Table of 10 under expressed gene

Table 3:Top 10 under expressed genes

| ID | adjPVal | PValue | t | B | logFC | Genesymbol | Genetitle |
|---|---|---|---|---|---|---|---|
| 244515_at | 0.00000113 | 2.56E-09 | -8.52 | 11.4 | -1.01 | | |
| 222737_s_at | 0.0000832 | 0.00000158 | -6.03 | 5.18 | -1.01 | BRD7 | bromodomain |
| 236533_at | 0.00251 | 0.000158 | -4.35 | 0.727 | -1.01 | ASAP1 | ArfGAP |
| 236368_at | 0.0027 | 0.000174 | -4.32 | 0.634 | -1.01 | KIAA0368 | KIAA0368 |
| 240636_at | 0.00846 | 0.000777 | -3.76 | -0.798 | -1.01 | | |
| 225173_at | 0.00933 | 0.000882 | -3.71 | -0.919 | -1.01 | ARHGAP18 | Rho |
| 230170_at | 4.42E-07 | 4.61E-10 | -9.24 | 13 | -1.02 | OSM | oncostatin |
| 222891_s_at | 0.0000076 | 4.85E-08 | -7.34 | 8.56 | -1.02 | BCL11A | B-cell |
| 1561923_a_at | 0.0000453 | 6.73E-07 | -6.35 | 6.01 | -1.02 | SF3B6 | splicing |
| 228818_at | 0.000869 | 0.0000382 | -4.87 | 2.09 | -1.02 | | |

## Discussion

Our analysis provides insights into the molecular mechanisms underlying MS pathogenesis. The upregulation genes from table 2 are EIF5A (Eukaryotic Translation Initiation Factor 5A),EIF5A is a gene that encodes a protein involved in the process of translation initiation, which is crucial for protein synthesis in cells and HINT3 is a gene that encodes a protein involved in cellular signaling and metabolism.

Conversely, the downregulation of genes involved in immune tolerance highlights potential targets for therapeutic interventions. Such as. BRD7 gene is part of the bromodomain-containing protein family, which plays a role in chromatin remodeling and transcription regulation. ASAP1 is involved in regulating the actin cytoskeleton, which is crucial for cell movement, shape, and intracellular transport and OSM is a cytokine that belongs to the interleukin-6 (IL-6) family and is involved in regulating inflammation, immune responses, and tissue remodeling.

## Conclusion

This study demonstrates the utility of RNA-seq analysis in identifying key molecular signatures in MS PBMCs. The integration of bioinformatics tools such as DESeq2, clusterProfiler, and visualization methods in R has enabled the identification of potential biomarkers and therapeutic targets for MS.