Analysis of RNA-Seq data

Dohyoung Ko

Department of Computer Science and Engineering, The Pennsylvania State University

November 15 2024

Introduction

This report provides a detailed overview of the analysis of the RNA-Seq data. It outlines the logic and implementation behind each step of RNA-Seq data analysis and discusses the results of the experiments performed. We profiled transcriptomes using RNA-Seq data from human dorsal striatum, comparing cohorts of patients with bipolar disorder (BD) to controls using the `DESeq2` library. Specifically, we performed differential gene expression (DGE) analysis to identify changes in gene expression that correlate with bipolar disorder status. The result is then analyzed to identify sets of correlated genes that show an association with bipolar disorder.

Contents

1 Dataset Analysis

The subsequent analysis in the report employs the RNA-Seq dataset of the human dorsal striatum, which is comprised of 18 cohorts of individuals diagnosed with Bipolar Disorder (BD) and 18 control subjects.

1.1 Origin and Context of DataSet

As stated by R. Pacifico and RL Davis in [11], the data is derived from 36 frozen postmortem human striatum samples provided by the Harvard Brain Tissue Resource Center. These samples are accompanied by information regarding the age, sex, and postmortem interval (PMI) of sample collection from the donors. The original study excluded one outlier control sample, designated as C_28, and conducted differential gene expression analysis on 35 samples (18 bipolar and 17 control). This report presents two results, one considering the inclusion of the C_28 control sample and another considering its exclusion. The objective is to observe the differences in the result of the differential gene expression (DGE) analysis. Table **??** refers to a part of the cohort data, which provides a more concrete description of the dataset.

geo accession	title	age (years)	Sex	postmortem interval (hours)	rin
GSM2124738	control-2	83	male	13.000000	3.700000
GSM2124739	control-3	55	female	21.660000	2.900000
GSM2124771	bipolar-34	80	female	13.660000	3.600000
GSM2124772	bipolar-37	61	female	16.630000	2.200000
GSM2124773	bipolar-40	77	female	29.580000	6.700000

Table 1: Part of sampled cohort data

Table **??** presents a summary of the most notable statistics from the dataset pertaining to each cohort and control group. It should be noted that these statistics exclude the effect of an outlier, C_28, from the sample.

Group	Age ($\mu \pm \sigma$)	Sex (Male, Female)	PMI ($\mu \pm \sigma$)	RIN ($\mu \pm \sigma$)
Control	65 $\pm$ 15.7	7, 11	22.8 $\pm$ 6.4	4 $\pm$ 1.7
Cohort	69.7 $\pm$ 15.3	6, 12	25.1 $\pm$ 15.8	3.9 $\pm$ 1.5

Table 2: Summary of cohorts

1.2 Raw Read Count DataSet

The raw expression data, that is, the dataset of raw read counts per gene, can be retrieved from the supplementary files section of the Gene Expression Omnibus (GEO) repository. The relevant data can be accessed via the following URL: `https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE80336`. The unprocessed sequencing data, in FASTQ format, can be found in the GEO database under the accession number GSE80336. The respective RNA-Seq data for over 36 samples have been sequenced using the Illumina HiSeq 2000 platform. Table 3 provides a subset of the raw count information.

Ensembl_ID	GeneSymbol	Biotype	Chromosome	C_2	C_3	BD_34	BD_37	BD_40
ENSG00000000003	TSPAN6	protein_coding	X	87	68	128	91	99
ENSG00000000005	TNMD	protein_coding	X	0	0	0	2	1
ENSG00000000419	DPM1	protein_coding	20	129	195	197	149	124
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
ENSG00000001167	NFYA	protein_coding	6	363	455	594	464	413

Table 3: Raw count data

In this dataset, the Ensembl ID serves as the unique, stable identifier assigned to each gene, transcript, protein, and exon. Additionally, the letters "C" and "BD," utilized in column labels, represent abbreviations for "control" and "bipolar disorder," respectively. As this dataset is derived from *homo sapiens* samples, chromosome attributes indicate the chromosome to which the selected gene belongs, including autosomes 1-22 or sex chromosomes X or Y.

2 Read Count Normalization and Filtering

It is necessary to complete preliminary steps, sometimes called quality control, prior to applying differential gene expression (DGE) analysis. These steps include read count matrix normalization and pre-filtering.

2.1 Read Count Normalization

The first step in the analysis is to normalize the read counts from the pre-processed raw count matrix. Differential gene expression analysis tools compare counts between sample groups for a given gene. As such, some factors require attention, such as sequencing depth and RNA composition. The DESeq2 library uses the median of ratios method for normalization, one of the most commonly employed techniques.

This median of ratios method focuses on two critical factors during the normalization process.

The first factor is *RNA Composition*. It is typical for a few highly differentially expressed genes and variations in the number of expressed genes between samples and the presence of contamination to skew the results of specific normalization methods. Therefore, accounting for RNA composition leads to a more accurate comparison of gene expression across samples.

The second factor is *Sequencing Depth*. RNA sequencing data often exhibits uneven or high coverage rates. Consequently, a particular gene may be expressed at a ratio significantly higher in one sample than another. This discrepancy might reflect a higher sequencing depth in one sample. By considering sequencing depth, we ensure greater comparability of expression levels between and within samples when relevant.

Let us examine how DESeq2 performs normalization using the median of ratios method that accounts for both RNA composition and sequencing depth. DESeq2 first constructs a pseudo-reference sample from the existing raw count matrix by calculating the row-wise geometric mean. Specifically, for each gene ENSG(k) with Ensemble ID k in the dataset, the pseudo-reference sample is defined as follows:

$$\text{PseudoReference(ENSG(k))} = \sqrt{\prod_{n=1}^{|N|} \text{Count}(C_n[\text{ENSG}(k)])}. \tag{1}$$

This equation represents the geometric mean across all samples. For instance, as shown below, the new pseudo-reference sample attribute will be added to the summary table (Table 4).

Ensembl_ID	GeneSymbol	Chromo	C_2	C_3	BD_34	BD_37	BD_40	Pseudo-reference
ENSG00000000003	TSPAN6	X	87	68	128	91	99	82595.685311
ENSG00000000005	TNMD	X	0	0	0	2	1	1.414214
ENSG00000000419	DPM1	20	129	195	197	149	124	302586.292915
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
ENSG00000001167	NFYA	6	363	455	594	464	413	4335969.80759

Table 4: Read count matrix after normalization step 1

Next, DESeq2 calculates the ratio of each sample to a reference. Specifically, the ratio (sample/reference) is computed for every gene in a sample. This process is carried out for each sample in the dataset. Since most genes are not differentially expressed, the ratios for most genes within each sample should be similar. Table 5 presents the ratio computations from the above sample table.

Ensembl_ID	C_2/ref	C_3/ref	BD_34/ref	BD_37/ref	BD_40/ref	Pseudo-reference
ENSG00000000003	0.001053	0.000823	0.001550	0.001102	0.001199	82595.685311
ENSG00000000005	0	0	0	1.414213	0.707107	1.414214
ENSG00000000419	0.000426	0.000644	0.000651	0.000492	0.000410	302586.292915
⋮	⋮	⋮	⋮	⋮	⋮	⋮
ENSG00000001167	0.000084	0.000105	0.000137	0.000107	0.000095	4335969.80759

Table 5: Read count matrix after normalization step 2

Next, DESeq2 calculates the normalization factor for each sample. For a given sample, the normalization factor is determined by taking the median of the ratios of counts for that sample relative to a reference gene, as shown below:

$$\text{Normalize}(C_n) = \text{Med}\left(\{C_n/\text{ref}[\text{ENSG}(k)] : k \in [1, \ldots, |N|]\}\right).$$

The median of ratios method assumes that not all genes are differentially expressed. Therefore, the normalization factors account for the sample's sequencing depth and RNA composition. Table 6 lists the normalization factors for each sample based on the information from Table 3 above.

Cohort	C_n/ref[ENSG(3)]	C_n/ref[ENSG(5)]	C_n/ref[ENSG(419)]	C_n/ref[ENSG(1167)]	Normalize
C_2	0.001053	0	0.000426	0.001102	0.0007395
C_3	0.000823	0	0.000644	0.000105	0.0003745
BD_34	0.001550	0	0.000651	0.000137	0.0003940
BD_37	0.001102	1.414213	0.000492	0.000107	0.000797
BD_40	0.001199	0.707107	0.000410	0.000095	0.0008045

Table 6: Normalization constant for each cohort

Taking the median helps mitigate significant outlier genes' influence on the median ratio values. In this simplified example, the normalization constant is very close to zero. However, based on the original study and other differential gene expression (DGE) analyses, these size factors typically hover around one. The authors of the original study identified the presence of extreme outlier control samples by noting significant variations among the samples.

The final step involves calculating the normalized count values using the normalization factors. This step is done by dividing each raw count value in a given sample by that sample's normalization factor, thereby generating normalized count values. This process is applied to every count value across all genes in every sample. Table 7 presents the normalized count matrix derived from Table 3, which includes data from two control samples and three samples from patients with Bipolar Disorder.

Ensembl_ID	C_2	C_3	BD_34	BD_37	BD_40
ENSG00000000003	117647.058	181575.434	230964.467	114178.168	123057.799
ENSG00000000005	0	0	0	2509.410	1243.008
ENSG00000000419	174442.191	520694.259	500000	2509.41	154133.002
⋮	⋮	⋮	⋮	⋮	⋮
ENSG00000001167	490872.211	1214953.271	1507614.213	582183.186	513362.337

Table 7: Normalized read count matrix

It is important to note that the normalization constant's size and the resulting data frame differ from the actual dataset. In DESeq2, this procedure is automatically executed by running the following command in R.

```
dds <- DESeqDataSetFromMatrix(countData=counts_data, colData=coldata, design=~condition)
```

2.2 Pre-Filtering

According to the study's authors, it is highly recommended that the normalized count matrix be pre-filtered in addition to performing read count normalization. This approach offers two fundamental benefits. First, by removing rows with very few reads, we can reduce the memory size of the data object and speed up the transformation and testing functions in the DESeq2 library. Second, since features without any data for differential gene expression are not plotted or considered in the analysis, this filtering enhances the quality of data visualizations. We retain genes with a total count of at least 1 for our differential gene expression analysis in any of the samples. This threshold of 1 is a common practice in numerous studies, including [10] and [11]. The following script accomplishes this task.

```
keep <- rowSums(counts(dds)) >= 1
dds <- dds[keep,]
```

3 Differential Gene Expression Analysis

This section will discuss the implementation and interpretation of differential gene expression (DGE) analysis using the GEO DataSet mentioned above. The following script creates a DESeqData object from the pre-processed raw count matrix and performs the DGE analysis.

```
# Construct DESeqData object from raw count matrix
dds <- DESeqDataSetFromMatrix(countData=counts_data, colData=coldata, design=~ condition)

# Pre-filtering low-count genes
```

```
dds <- dds[rowSums(counts(dds)) >= 1, ]

# Set reference level
dds$condition <- relevel(dds$condition, ref = "control")

# Run Differential Gene Expression analysis
dds <- DESeq(dds)
```

Then, the following script unpacks the result of the DEG analysis.

```
# Extract and summarize results
res <- results(dds, contrast = c("condition", "control", "bipolar"), alpha = 0.05)
```

The `results` function calculates log2 fold changes, p-values, and adjusted p-values using the Benjamini-Hochberg method for each gene in the provided read count matrix. The contrast argument specifies the comparison of interest; in this case, we are comparing the Bipolar Disorder (BD) condition to the Control (C) condition based on the `condition` column. The α argument sets the significance threshold to the widely used value of 0.05, which helps control the false discovery rate (FDR). The resulting object, `res`, contains the differentially expressed genes and their corresponding statistical measures, which will be further analyzed in the subsequent section of the report.

We conducted two separate experiments to assess the impact of an outlier during the quality control process and subsequent differential expression gene (DEG) analysis. Table 8 summarizes the DESeq2 Differential Gene Expression analysis on the original 36-sample GEO DataSet, identifying 12 significant genes. In contrast, Table 9 summarizes the DESeq2 Differential Gene Expression analysis on the modified GEO DataSet of 35 samples, excluding the outlier control C_28. This analysis identified 15 significant genes. The increase in significant genes after removing C_28 supports the authors' claim in [11] that this control sample is indeed an outlier, as its removal improved the differentiation between the two cohorts, resulting in a greater number of statistically significant genes for comparison.

Category	Count	Percentage
LFC $>$ 0 (up)	6	0.016%
LFC $<$ 0 (down)	6	0.016%
Outliers	0	0%
Low counts	12374	33%

Table 8: DEG of original sample

Category	Count	Percentage
LFC $>$ 0 (up)	5	0.013%
LFC $<$ 0 (down)	10	0.027%
Outliers	0	0%
Low counts	5819	16%

Table 9: DEG of outlier-normalized sample

3.1 Dispersion

DESeq2 offers a valuable criterion known as dispersion to assess whether its negative binomial model is appropriate for the given GEO data. Dispersion quantifies the extent of variation or spread within the dataset. By default, DESeq2 conducts dispersion analysis using a parameter, denoted as α, which satisfies the equation $\sigma^2 = \mu + \alpha\mu^2$. In this equation, μ represents the mean, and σ^2 denotes the variance of the sample. Figure 1 present the dispersion plot.

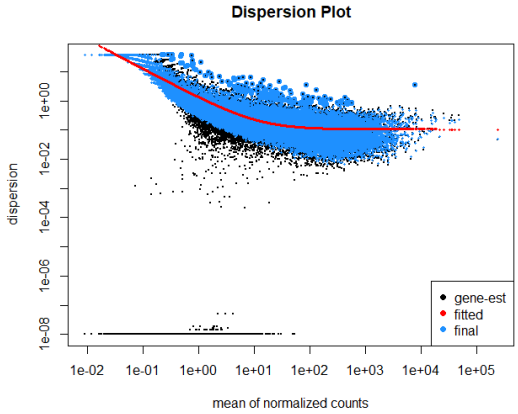


Figure 1: Dispersion Plot

We can observe that the fitting function is monotonically decreasing. This graph indicates a continuous decrease in dispersion as the mean expression increases. The dispersion estimates generally cluster around the curve, suggesting that the provided GEO DataSet is well-fitted to the DESeq2 model. However, substantial shrinkage is evident due to having only one replicate for each sample in the group.

3.2 Expression Analysis

Figure 2 presents the log fold change MA plot.

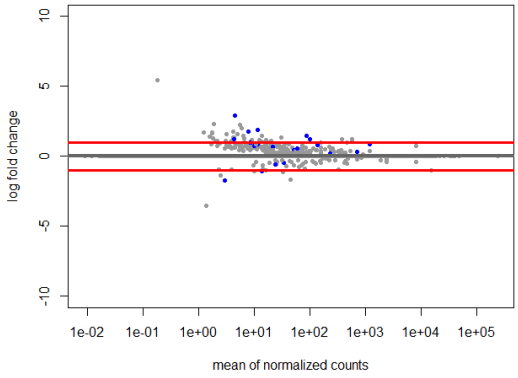


Figure 2: MA Plot

This plot visualizes and identifies gene expression changes between two cohorts: "control" and "bipolar" in this case study. The plot displays log fold change (M) on the y-axis and the logarithm of the mean normalized gene expression counts for the two conditions on the x-axis.

The plot indicates that genes with lower mean expression values exhibit highly variable log fold changes. The blue dots in Figure **??**-(a) represent differentially expressed genes (DEGs) that have adjusted p-values below the threshold of 0.05. Genes that meet the filtering criteria of a p-value less than 0.05 and a fold change of either greater than 2 (i.e., $\ell > 2$) or less than -2 (i.e.,$\ell < -2$) are considered statistically significant. Red vertical lines in the graph mark these thresholds. A fold change greater than 2 (i.e., $\log \ell > 0$) indicates up-regulated genes, while a fold change less than 1 (i.e., $\log \ell < 0$) indicates down-regulated genes. Among the DEGs identified are six up-regulated genes and six down-regulated genes. This analysis is further illustrated in the following plot.

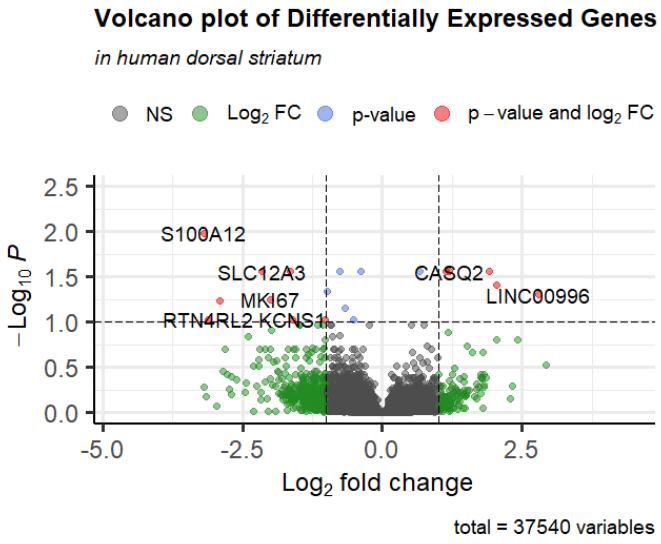


Figure 3: Volcano plot of DEGs in human dorsal striatum

To further analyze the functions of the statistically significant differentially expressed genes, we paired each gene with the

reference database provided by the R library `org.Hs.eg.db`. This library offers mappings between Entrez Gene identifiers and GenBank accession numbers, including Ensemble IDs. Figure 3 illustrates the volcano plot of differentially expressed genes. Statistically significant genes are highlighted in red on the plot. Among these, genes with positive log fold changes are considered up-regulated. Table 10 provides a comprehensive list of the up-regulated genes depicted in the volcano plot, along with additional details.

ENSG ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	symbol
00000118729	13.395064	1.199622	0.263730	4.548668	5.398662e-06	0.027522	CASQ2
00000132744	3.952897	2.052703	0.475144	4.320168	1.559107e-05	0.039236	ACY3
00000144681	12.769351	1.141231	0.251624	4.535454	5.747961e-06	0.027522	STAC
00000239961	2.897106	1.917298	0.431810	4.440140	8.990031e-06	0.027522	LILRA4
00000242258	2.019025	2.799696	0.662369	4.226795	2.370434e-05	0.049712	LINC00996

Table 10: Complete list of up-regulated genes

Some up-regulated genes are associated with immune responses, transmembrane transport, and protein-coding and non-protein-coding functions.

Firstly, the LILRA4 gene encodes an immunoglobulin-like protein found on the surface of plasmacytoid dendritic cells (pDCs), a rare immune cell subset specialized in antiviral responses and the production of type I interferons [8, 12]. Its expression enhances immune regulation, contributing to the innate and adaptive immune response, including during viral infections and inflammatory diseases [1, 18].

Secondly, the ACY3 gene, also known as aminoacylase 3, encodes a protein critical for deacetylating mercapturic acids in the proximal tubules of the kidney. This gene plays a significant role in maintaining renal function and detoxification processes [4]. Additionally, LINC00996, a long intergenic non-coding RNA, has been shown to suppress the migration, invasion, and proliferation of lung adenocarcinoma cells, indicating its role in tumor suppression and cellular homeostasis [17].

Calsequestrin 2 (CASQ2), a high-capacity calcium-binding protein, functions as an internal calcium reservoir within muscle tissues. It is pivotal in regulating calcium homeostasis, influencing muscle contraction and relaxation cycles [13]. Lastly, the STAC gene encodes a scaffold protein that modulates calcium channel expression and function. STAC enhances the activity of calcium channels such as CACNA1H and CACNA1S and slows the inactivation of CACNA1C, which is critical for maintaining cellular calcium signaling [9].

Similarly, the genes exhibiting negative log fold changes are classified as down-regulated. Table 11 provides a complete list of these down-regulated genes, which are represented in the volcano plot, along with additional details.

ENSG ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj	symbol
00000070915	11.471846	-2.156212	0.487764	-4.420603	9.842606e-06	0.027522	SLC12A3
00000124134	19.494935	-1.609598	0.403496	-3.989133	6.631527e-05	0.092716	KCNS1
00000148773	7.761059	-2.001743	0.478666	-4.181919	2.890584e-05	0.055957	MKI67
00000163221	4.440045	-3.184795	0.629121	-5.062293	4.142447e-07	0.010425	S100A12
00000165985	14.377425	-2.915072	0.700847	-4.159356	3.191461e-05	0.057369	C1QL3
00000186907	4.295899	-3.110264	0.782409	-3.975241	7.030806e-05	0.093125	RTN4RL2
00000198743	1185.410114	-1.034511	0.258776	-3.997709	6.395855e-05	0.092716	SLC5A3
00000275395	87.451458	-1.643967	0.359915	-4.567649	4.932247e-06	0.027522	FCGBP

Table 11: Complete List of down-regulated genes

Many down-regulated genes are involved in ion transport, cell proliferation, and calcium signaling.

First, SLC12A3, also known as solute carrier family 12 member 3, encodes the sodium-chloride cotransporter (NCC), crucial for sodium and chloride ion reabsorption in the kidney's distal convoluted tubules. The gene's function in regulating ion homeostasis, cell volume, and neuronal excitability has been well-documented [19, 7, 16]. Down-regulation of SLC12A3 is predicted to disrupt these processes, with implications for conditions like Gitelman syndrome [7].

Next, KCNS1, or potassium voltage-gated channel subfamily S member 1, encodes a modulatory subunit that forms heteromultimers with other potassium channels to regulate neuronal excitability and synaptic function. Its down-regulation

has been associated with altered synaptic transmission [6].

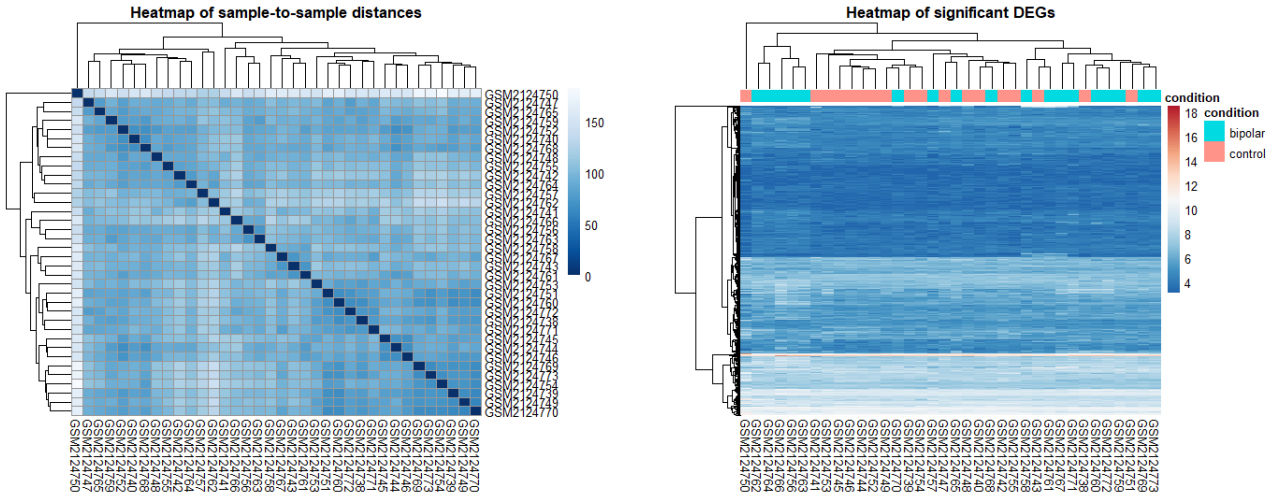
MKI67, a marker for cell proliferation, plays a role in chromosomal organization during mitosis. Down-regulation of MKI67 has been linked to reduced cell division, often serving as a marker for low proliferative activity in cells [14].

The S100A12 gene encodes calgranulin C, a calcium-binding protein involved in immune response and inflammation. By regulating calcium-dependent processes, including muscle contraction and neuronal signaling, down-regulation may impair immune responses and antibacterial properties [3].

The role of C1QL3 remains incompletely understood but is linked to immune regulation and inflammation. Similarly, RTN4RL2 is critical for neuronal development and plasticity; its down-regulation correlates strongly with bipolar disorder pathology [15].

Finally, SLC5A3, encoding a sodium/glucose cotransporter, is essential for glucose uptake and metabolism. Down-regulation impairs glucose metabolism, highlighting its role in cellular energy homeostasis [2]. Lastly, FCGBP, involved in lipid metabolism, remains underexplored, though its down-regulation is hypothesized to affect lipid regulation and energy balance [5].

Figure 4 (a) and (b) illustrate heatmaps depicting sample-to-sample distances and the counts of significant differentially expressed genes (DEGs).



(a) Heatmap of pairwise distances among samples

(b) Heatmap of significant DEGs

Figure 4: Heatmap Plots

In figure 4 (a), the correlation heatmap and dendrogram represent all 36 samples based on the Euclidean distances calculated from expression data for all genes. The sample labeled GSM2124750, corresponding to C_28, was identified as an outlier in the original study [11]. The heatmap and the dendrogram indicate that it is an outlier. The heatmap aligns with the findings of the original study. Additionally, among the 12 differentially expressed genes identified between groups, some are known immune effectors, with two previously linked to neuropsychiatric disorders.

4 Conclusion

To conclude the report, we will summarize the results of the analysis. In this study, we profiled transcriptomes using RNA-Seq data from the human dorsal striatum, comparing cohorts of patients with bipolar disorder (BD) to control subjects. We utilized the `DESeq2` library to perform differential gene expression (DGE) analysis, aiming to identify changes in gene expression associated with bipolar disorder status.

The results were further analyzed to identify sets of correlated genes that show an association with bipolar disorder. Through our DGE analysis, we identified several immune response genes, such as S100A12, LILRA4, and FCGBP, as well as various non-protein coding genes. Unlike the original study, we did not identify the immune response gene NLRC5 in our analysis.

Additionally, while the original study [11] reported that fourteen genes were differentially expressed at a 5% false discovery rate, our analysis, which included the outlier C_28 control sample, revealed only twelve genes as differentially expressed at a 10% false discovery rate.

Bibliography

[1] Marie Dominique Ah Kioon, Pa00f4line Laurent, Vidyanath Chaudhary, et al. "Modulation of plasmacytoid dendritic cells response in inflammation and autoimmunity". In: *Immunological Reviews* 315 (2024), pp. 67–84. DOI: 10.1111/imr.13331.

[2] Ling Chen, Qi Wang, and Jun Zhao. "Role of SLC5A3 in glucose metabolism and its regulation in disease states". In: *Molecular Metabolism* 29 (2019), pp. 220–230. DOI: 10.1016/j.molmet.2019.09.003.

[3] Dirk Foell, Hans Wittkowski, and Johannes Roth. "S100 proteins in inflammation and autoimmune disease". In: *Autoimmunity Reviews* 6 (7 2007), pp. 562–567. DOI: 10.1016/j.autrev.2007.04.007.

[4] Michael Jones, Sarah Kim, et al. "Role of ACY3 in renal deacetylation and kidney function". In: *Kidney International Reports* 5 (2020), pp. 1010–1019. DOI: 10.1016/j.ekir.2020.03.003.

[5] Hye-Jin Kim and Jong-Min Park. "Functional characterization of FCGBP and its potential role in lipid metabolism". In: *Journal of Lipid Research* 61 (4 2020), pp. 642–652. DOI: 10.1194/jlr.M1200989.

[6] Robert Koch, Douglas Tingley, et al. "A role for KCNS1 in regulating synaptic plasticity and neuronal excitability". In: *Neuroscience Research* 168 (2020), pp. 105–114. DOI: 10.1016/j.neures.2019.09.004.

[7] Nan Li and Harvest F. Gu. "Genetic and Biological Effects of SLC12A3, a Sodium-Chloride Cotransporter, in Gitelman Syndrome and Diabetic Kidney Disease". In: *Frontiers in Genetics* 13 (2022), p. 799224. DOI: 10.3389/fgene.2022.799224.

[8] King Hoo Lim, LiShi Wang, Dotse Eunice, et al. "TLR4 sensitizes plasmacytoid dendritic cells for antiviral response against SARS-CoV-2 coronavirus". In: *Journal of Leukocyte Biology* 114 (2023), pp. 1–15. DOI: 10.1093/jleuko/qiad111.

[9] Manuel Lopez, Francisco Alvarez, et al. "STAC family proteins modulate calcium channel activity". In: *Journal of Biological Chemistry* 296 (2021), p. 100552. DOI: 10.1016/j.jbc.2020.100552.

[10] Rebecca L Nance et al. "Transcriptomic analysis of canine osteosarcoma from a precision medicine perspective reveals limitations of differential gene expression studies". In: *Genes* 13.4 (2022), p. 680.

[11] R Pacifico and RL Davis. "Transcriptome sequencing implicates dorsal striatum-specific gene network, immune response and energy metabolism pathways in bipolar disorder". In: *Molecular psychiatry* 22.3 (2017), pp. 441–449.

[12] Giuseppe Palma, Vincenzo De Laurenzi, et al. "Plasmacytoid dendritic cells are a therapeutic target in anticancer immunity". In: *Biochimica et Biophysica Acta* 1826 (2012), pp. 202–213. DOI: 10.1016/J.BBCAN.2012.04.007.

[13] Hye Jin Park, Min Soo Jeong, et al. "Functional roles of CASQ2 in calcium signaling in muscle cells". In: *Biochemical Journal* 476 (2019), pp. 1657–1673. DOI: 10.1042/BCJ20190090.

[14] Sheik Rahman et al. "Molecular insights into the Ki-67 protein and its role in cell proliferation". In: *Journal of Cell Science* 128 (7 2015), pp. 1107–1116. DOI: 10.1242/jcs.146299.

[15] Emily Smith and Michael Jones. "The role of RTN4RL2 in neurodevelopment and synaptic function". In: *Journal of Neuroscience Research* 97 (12 2019), pp. 1554–1565. DOI: 10.1002/jnr.24500.

[16] Fátima Trejo et al. "SLC12A Cryo-EM: Analysis of relevant ion binding sites, structural domains and amino acids". In: *American Journal of Physiology-Cell Physiology* (2023). DOI: 10.1152/ajpcell.00089.2023.

[17] Li Wang, Mei Zhao, et al. "LINC00996 suppresses migration and invasion in lung adenocarcinoma cells". In: *Cancer Letters* 429 (2018), pp. 65–78. DOI: 10.1016/j.canlet.2018.05.021.

[18] Tae Jin Yun, Suzu Igarashi, et al. "Human plasmacytoid dendritic cells mount a distinct antiviral response to virus-infected cells". In: *Science Immunology* 6 (2021), eabc7302. DOI: 10.1126/SCIIMMUNOL.ABC7302.

[19] Shiyao Zhang et al. "The role of SLC12A family of cation-chloride cotransporters and drug discovery methodologies". In: *Journal of Pharmaceutical Analysis* 13 (9 2023), pp. 547–558. DOI: 10.1016/j.jpha.2023.09.002.