

Analysis of ABCB4 Gene Expression

August 29, 2024

1 Introduction

In this study, we conduct a Large-Scale Multi-Omic Analysis to investigate the relationship between ABCB4 gene expression and COVID-19 severity [1], utilizing data from the GSE157103 dataset. The ABCB4 gene, crucial for phospholipid transport across cell membranes, has been implicated in various liver diseases due to its role in genetic defects

2 Methods

The analysis was carried out using R version 4.1.1 [4], with essential packages including `tidyverse` for data manipulation and visualization [5], `pheatmap` for generating heatmaps, and `knitr` for dynamic report generation [6]. Data preprocessing involved extracting relevant columns to focus on gene expression profiles, with particular emphasis on the ABCB4 gene for detailed analysis. `kable` was used to construct a summary table to assist in generalizing analysis [7]. Heatmaps were created using `pheatmap` to visually represent gene expression patterns [3]. To cluster the data, we employed hierarchical clustering using the Euclidean distance metric, which measures the straight-line distance between points in multidimensional space. This approach facilitated grouping based on similarity in gene expression patterns, enabling a clear understanding of the data structure [2]. The combination of these tools and methods provided a comprehensive framework for effective data analysis and visualization.

Table 1: Summary Statistics Stratified by Sex showing the standard deviation or IQR for continuous variables and the n percent for categorical variables

Variable	Female	Male	Unknown
Age, mean (SD)	59.88 (18.22)	62.64 (14.65)	83.00 (NA)
Ferritin (ng/ml), median [IQR]	318.0 [547.0]	755.0 [849.0]	NA [NA]
D-dimer (mg/l FEU), median [IQR]	1.37 [5.86]	2.21 [10.55]	NA [NA]
Disease Status, n (%)	51 (74.5%)	74 (83.8%)	1 (100%)
Mechanical Ventilation, n (%)	51 (68.6%)	74 (52.7%)	1 (100%)

3 Results

3.1 Histogram of ABCB4 Gene Expression



Figure 1: Histogram showing the distribution of ABCB4 gene expression.

3.2 Scatter Plot of ABCB4 Gene Expression and Age

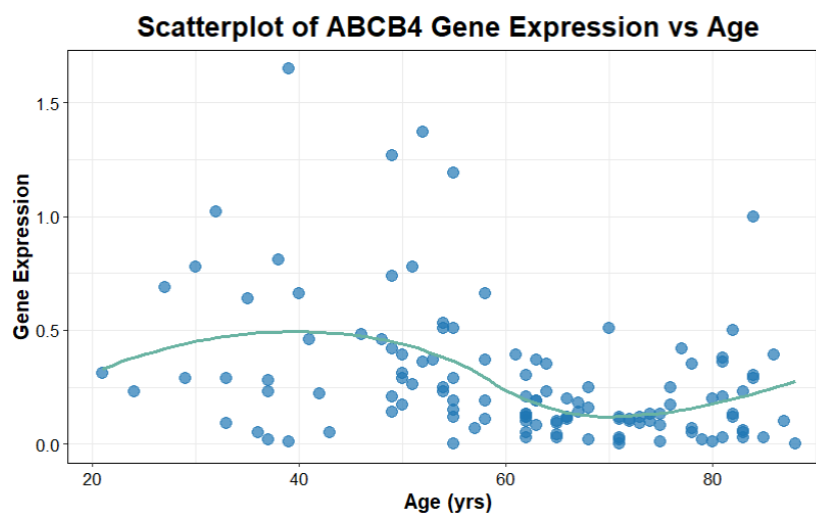


Figure 2: Scatter plot showing the relationship between ABCB4 gene expression and age.

3.3 Boxplot of ABCB4 Gene Expression Stratified by Disease Status and Sex

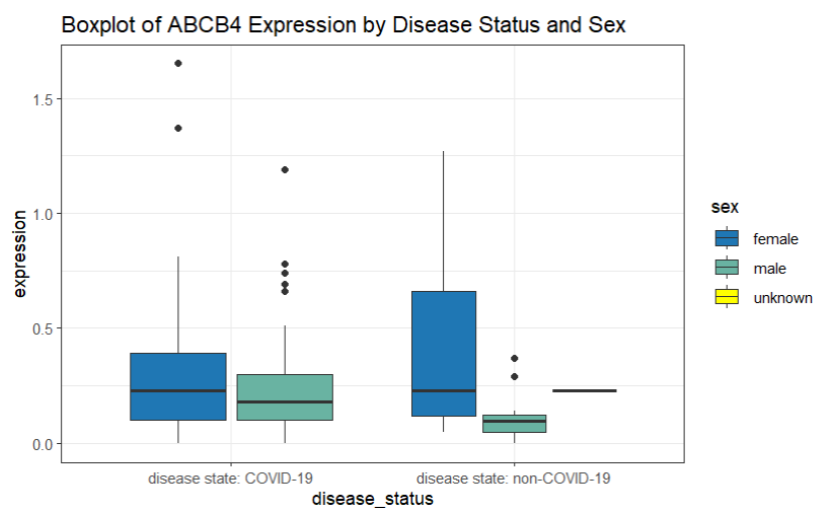


Figure 3: Boxplot showing ABCB4 gene expression stratified by disease status and sex.

3.4 Heatmap of Gene Expression

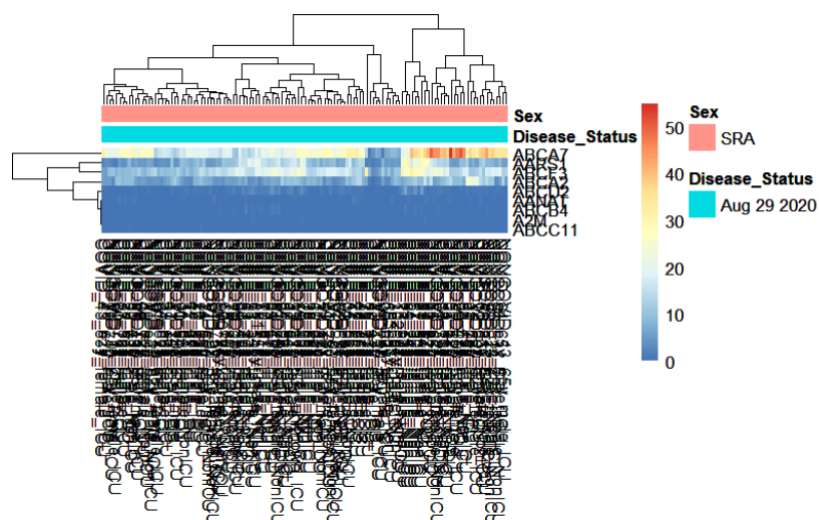


Figure 4: Heatmap showing the expression levels of selected genes.

3.5 Lollipop Plot

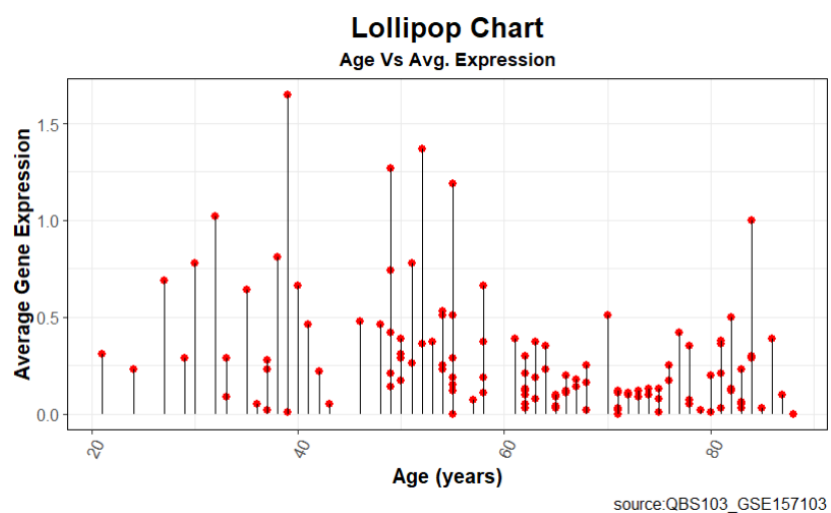


Figure 5: Lollipop Plot showing the relationship between Age and Gene Expression

4 Discussion

The first table (table 1) showcases the relationship between the variables age, ferritin, D-dimer levels, disease status, and mechanical ventilation stratified by sex. Interestingly, there was a noticeable difference between ferritin levels between males and females, but more analysis is necessary to make further conclusions. The specific analysis of the ABCB4 gene expression provided insights into its variability across participants. The histogram (Figure 1) plot showed a relatively normal distribution, with a greater amount of gene expression being correlated with a decrease in the count, while the scatter plot (Figure 2) suggested no strong correlation between ABCB4 expression and age. The box-plot (Figure 3) indicated some variability in expression based on disease status and sex, with females who did not have covid-19 having much more variability in the gene expression, while men were fairly standard. Interestingly, males and females were very similar in gene expression when they did have covid-19. Further, the heatmap (Figure 4) highlighted overall gene expression patterns across different samples showcasing that certain genes, such as ABCA7 had much higher expression than a gene such as ABCC11. Lastly, the Lollipop Plot (Figure 5) showcases the relationship between Age and Gene expression, which allowed us to compare gene expression to age, showcasing the wide variability.

5 References

References

- [1] Gse157103 dataset, 2024. Accessed: 2024-08-29.
- [2] GeeksforGeeks. Draw heatmap with clusters using pheatmap in r, 2024.
- [3] R Kolde. pheatmap: Pretty heatmaps, 2019. R package version 1.0.12.
- [4] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024.
- [5] Hadley Wickham. tidyverse: Easily install and load the tidyverse, 2019. R package version 1.3.0.
- [6] Yihui Xie. knitr: A general-purpose package for dynamic report generation in r, 2023. R package version 1.43.
- [7] H Zhu. kableextra: Construct complex table with kable() in r, 2023. R package version 1.3.4.