QBS 103: Final Project Submission

Angelique Cortez

August 25 2024

Contents

1 Introduction **1**

2 Methods **1**

3 Results **2**
3.1 Table of Summary Statistics 2
3.2 Histogram of ABI1 Gene Expression 2
3.3 Scatter Plot of Gene + Continuous Covariate 3
3.4 Boxplot of Gene Stratified by Two Categorical Covariates 3
3.5 Heatmap of ABI1 Gene 4
3.6 PCA plot 4

4 Conclusions **5**

1 Introduction

The Abelson interactor 1 (*ABI1* gene is shown to be an important regulator of actin filament regulator during cytoskeltal reorganization. Additionally, in a study condicted by Steinstel et al.,*ABI1* has been indicated as a downtream target of inflammation. As we are interested in the severity of Covid-19 and inflammation, we looked at *ABI1* expression and examined covariates such as age, sex, ICU status, Ferritin(ng/mL), Crp(ng/mL), and mechanical ventilation.

2 Methods

Utilizing the publicly available dataset (GSE157103) from Overmyer et al., I formatted the gene expression and matrix files for downstream data analysis. To understand the distribution of the data, I created a histogram using ggplot2. Additionally, I used ggplot2 to observe if there was a correlation between ABI1 gene expression and the covariate age and added a correlation line with geom smooth. A boxplot stratified by sex and mechanical ventilation covariates was made via ggplot2. Pheatmap was used to visualize ABI1 and 19 other genes,

displaying stratification by mechanical ventilation and ICU status. Principal Component Analysis (PCA) of patient data colored by ICU Status was plotted using ggplot2.

3 Results

3.1 Table of Summary Statistics

Categorical Variables			**Age**		**Ferritin (ng/mL)**		**Crp (ng/mL)**	
icu_status	sex	n	mean_age	sd_age	mean_ferritin	sd_ferritin	mean_crp	sd_crp
no	female	2700	54.92	18.15	536.48	983.14	79.22	68.76
no	male	3200	61.81	16.62	901.12	1116.34	138.75	106.34
yes	female	2400	64.04	16.00	713.36	1100.11	151.33	112.64
yes	male	4100	62.63	12.27	1067.30	928.44	148.58	99.82

Table 1: Summary Statistics of Covariates: Age, Ferritin(ng/mL), Crp(ng/mL) by ICU status and Sex.

3.2 Histogram of ABI1 Gene Expression

As shown in Figure 1, the histogram illustrates a normal distribution of ABI1 gene expression in this dataset.

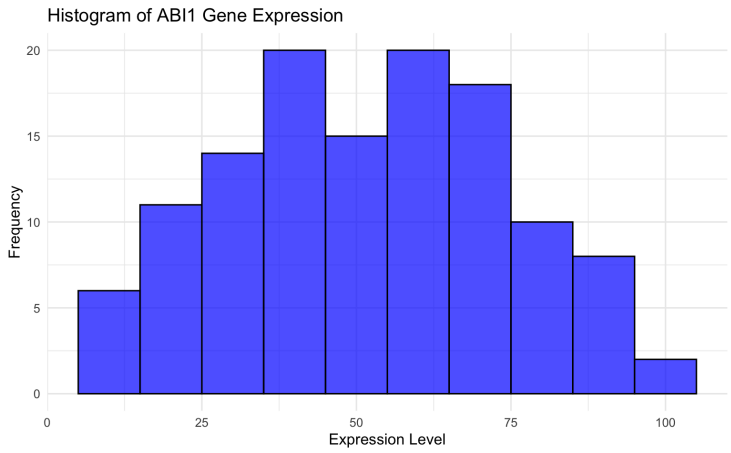

Figure 1: Frequency of ABI1 Gene Expression

3.3 Scatter Plot of Gene + Continuous Covariate

As shown in Figure 2, the scatterplot illustrates the non significant negative correlation between ABI1 gene expression and age. Furthermore, there is an even distribution of sex with no clear stratification.

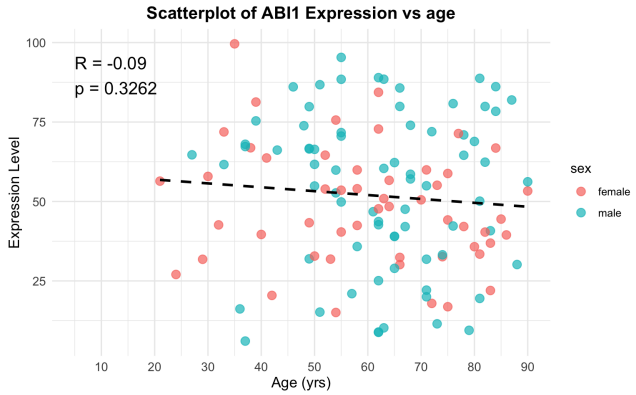

Figure 2: Correlation Analysis of ABI1 Gene Expression and Age (yrs)

3.4 Boxplot of Gene Stratified by Two Categorical Covariates

As shown in Figure 3, the boxplot illustrates the frequency distribution of ABI1 gene expression by sex and mechanical ventilation. Those who are male have an overall higher median expression level in both mechanical ventilation groups compared to females. Those that required mechanical ventilation male and female, had higher median expression of ABI1 gene.

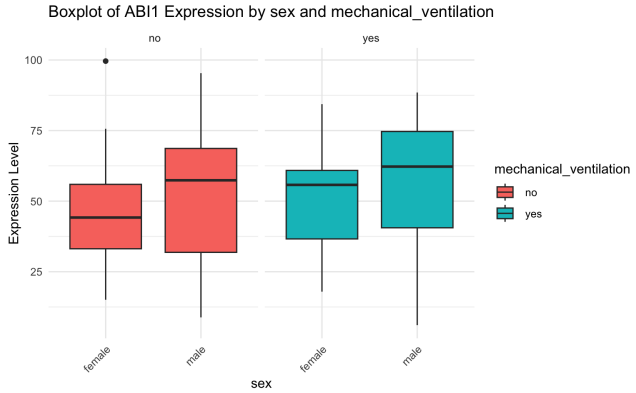

Figure 3: ABI1 Gene Expression by Mechanical Ventilation

3.5 Heatmap of ABI1 Gene

As shown in Figure 4, the heatmap illustrates the frequency distribution of genes including ABI1 gene expression by ICU Status and Mechanical ventilation Status. It appears that for the first 25 participants in the heat-map who were admitted to the ICU and had mechanical ventilation had an increased expression of the ABI1 gene.

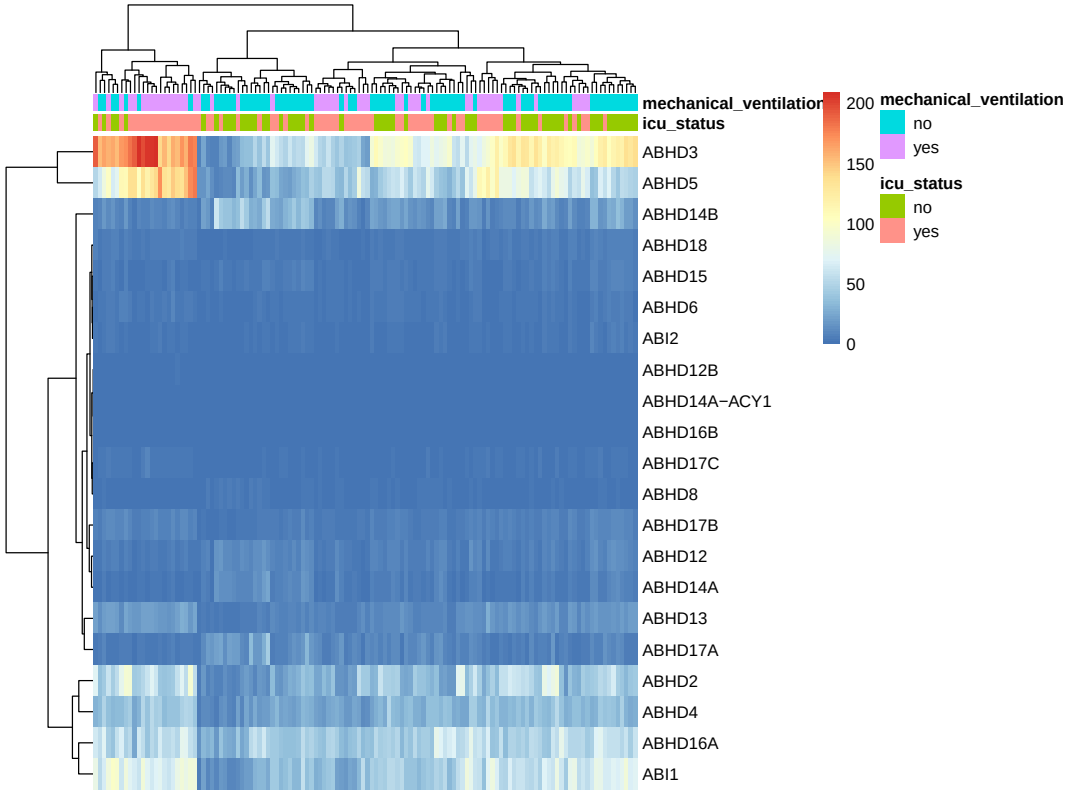

Figure 4: Gene Expression by Mechanical Ventilation and ICU Status

3.6 PCA plot

As shown in Figure 5, the PCA plot illustrates the variance of patient data colored by ICU Status. It appears that there is a mild stratification between those who had been admitted to the ICU and thos who did not, demonstrating that the variance may be driven by ICU Status.

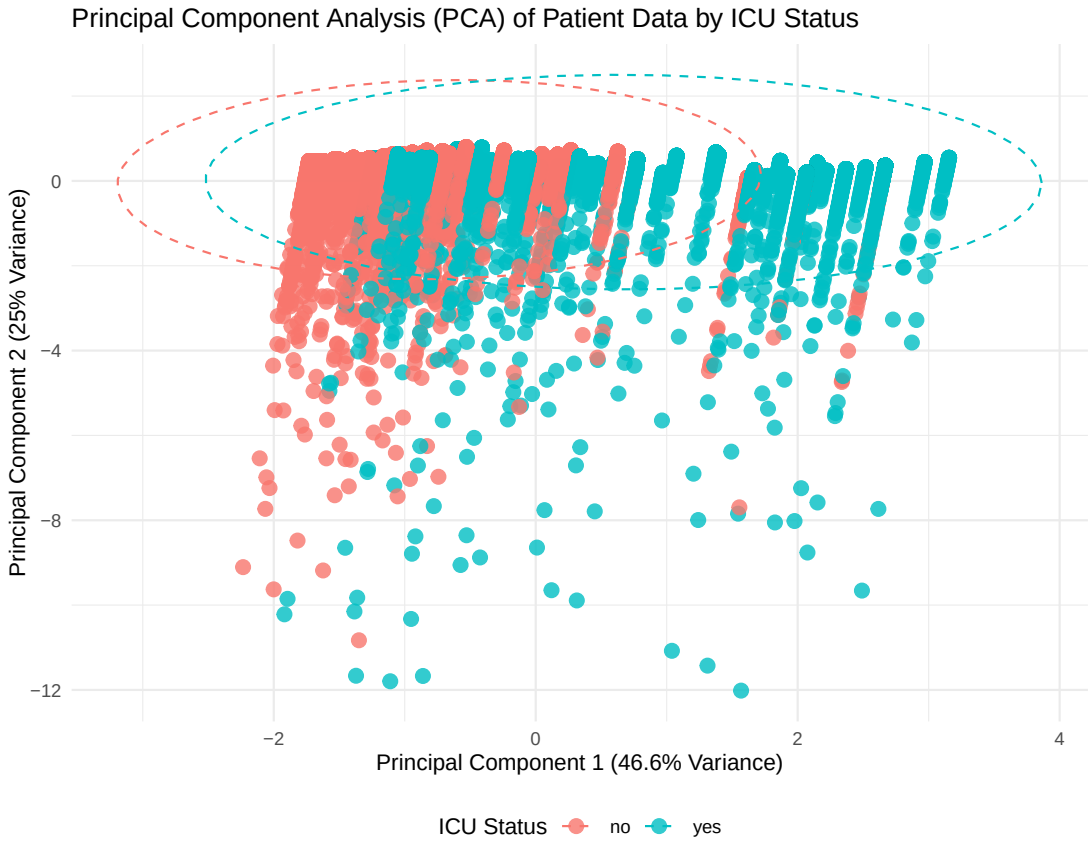

Figure 5: Principal Component Analysis (PCA) of Patient Data by ICU Status

4 Conclusions

Figure 1, shows a histogram of the ABI1 gene shows a normal distribution. There is a non significant negative correlation between gene expression and the covariate age in Figure 2, additionally there is not defined grouping based on sex further demonstrating ABI1 gene expression does not associate with sex. ABI1 gene expression in Figure 3 does appear to be overall increased in males vs females. Additionally, we see that expression level increased for both sexes when the patient required mechanical ventilation. We further see an increase in ABI1 expression associated with mechanical ventilation and ICU status (Figure 5). Overall we see an association of between ABI1 gene expression and mechanical ventilation and ICU Status but not sex.

References

[1] Kassambara, A. (2023). *ggpubr: 'ggplot2' Based Publication Ready Plots*. R package version 0.6.0, https://CRAN.R-project.org/package=ggpubr.

[2] Kolde, R. (2019). *pheatmap: Pretty Heatmaps*. R package version 1.0.12, https://CRAN.R-project.org/package=pheatmap.

[3] Müller, K., Wickham, H. (2023). *tibble: Simple Data Frames*. R package version 3.2.1, https://CRAN.R-project.org/package=tibble.

[4] Overmyer, K. A., Shishkova, E., Miller, I. J., Balnis, J., Bernstein, M. N., Peters-Clarke, T. M., Meyer, J. G., Quan, Q., Muehlbauer, L. K., Trujillo, E. A., He, Y., Chopra, A., Chieng, H. C., Tiwari, A., Judson, M. A., Paulson, B., Brademan, D. R., Zhu, Y., Serrano, L. R., Linke, V., ..., Jaitovich, A. (2021). Large-Scale Multi-omic Analysis of COVID-19 Severity. *Cell Systems*, 12(1), 23–40.e7. https://doi.org/10.1016/j.cels.2020.10.003.

[5] Steinestel, K., Brüderlein, S., Steinestel, J., Märkl, B., Schwerer, M. J., Arndt, A., Kraft, K., Pröpper, C., Möller, P. (2012). Expression of Abelson interactor 1 (Abi1) correlates with inflammation, KRAS mutation and adenomatous change during colonic carcinogenesis. *PloS One*, 7(7), e40671. https://doi.org/10.1371/journal.pone.0040671.

[6] Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

[7] Wickham, H., François, R., Henry, L., Müller, K. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.3, https://CRAN.R-project.org/package=dplyr.

[8] Wickham, H., Girlich, M. (2023). *tidyr: Tidy Messy Data*. R package version 1.3.0, https://CRAN.R-project.org/package=tidyr.

[9] Zhu, H. (2023). *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. R package version 1.3.4, https://CRAN.R-project.org/package=kableExtra.