

Final Project

Jennifer Faith

August 2025

Contents

1	Introduction	1
2	Methods	1
3	Results	2

1 Introduction

For this project, we analyzed two datasets that explored gene expression across patients from a hospital. The dataset provided various health measures, including the Charlson score, ferritin levels, and CRP, for 126 patients, as well as measures of their gene expression across 100 genes. The datasets also provided information on whether the individuals were positive for COVID-19.

Throughout this analysis, the focus was placed on gene ABCF1 because it is believed to have a strong role in regulating immune response by shifting the body from attack mode back to a more relaxed mode at the end of an illness [5]. The gene has been shown across other studies to play an important role in recovery from illnesses like sepsis [1]. Our objective was to analyze a specific gene and investigate whether any patterns exist between that gene expression and health outcomes.

2 Methods

The data sources analyzed were provided by faculty and emerged from a study previously conducted. Throughout this project, the datasets were analyzed using R version 4.5.1. Many additional R libraries were utilized for various graphs, including bar charts, scatter plots, etc. These notable packages were tidyr [4], tidyverse [3], knitr[6], and kableExtra [7]. To create a clustered heatmap, the R package pheatmap [2] was used, and the Euclidean clustering algorithm was applied. This allows for the organization and sorting of data points to find the clearest pattern in the data.

3 Results

As shown in the summary table 1, across all of the variables, there were a few notable trends. For starters, the mean age was on the older side, at 62.8 for healthy, non-COVID-19 positive patients, and the patients positive with COVID-19 were also in their 60s, with a mean age of 60.8. The reported gender of the patients was fairly balanced, but the study data showed a higher number of males testing positive for COVID. Finally, a higher percentage of patients without COVID-19 were in the ICU compared with patients who had COVID-19.

Table 1: Summary statistics for covariates

Table 2: Summary Table		
Variable	Healthy (n=26)	Diseased (n=99)
Gene ABCF1 Expression mean (sd)	12.4 (5.5)	14.9 (7.2)
Age mean (sd)	62.8 (15.6)	60.8 (16.1)
Hospital Free Days mean (sd)	32.3 (14.8)	21.9 (16.6)
Sex n (%)		
Female	13 (50)	38 (38.4)
Male	12 (46.2)	61 (61.6)
Unknown	1 (3.8)	0 (0)
ICU Status (%)		
No	10 (38.5)	49 (49.5)
Yes	16 (61.5)	50 (50.5)

While analyzing a histogram of gene expression vs. count of participants as seen in Figure 1, we observe a slight trend indicating that patients typically have lower gene expression of ABCF1 overall. Only a few patients had a gene expression of over 20 with the majority clustered around 5-15.

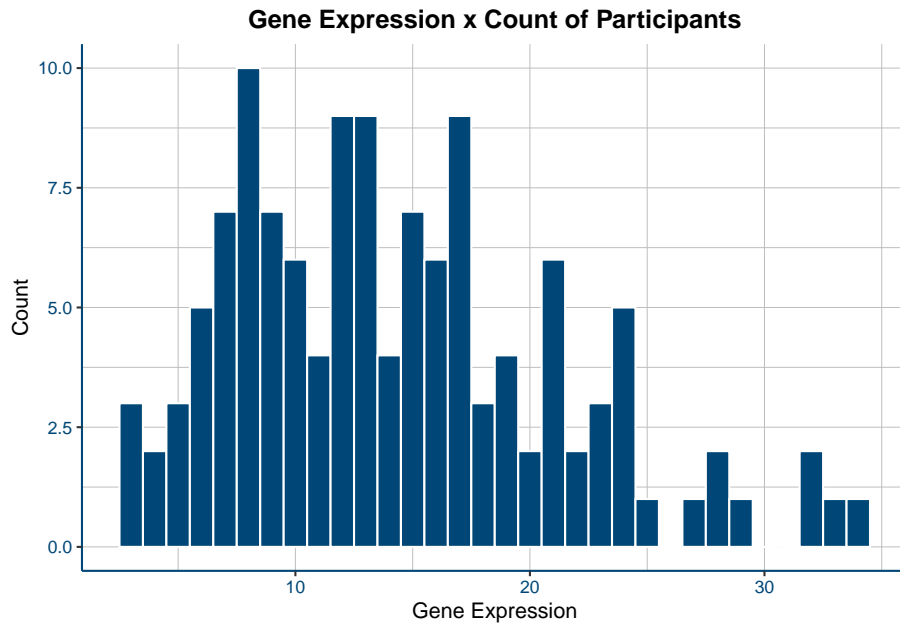


Figure 1: Histogram of Gene Expression

The analysis of ABCF1 gene expression compared with hospital-free days also showed an interesting trend when plotted in a scatterplot as seen in Figure 2. Hospital-free days are the number of days in a 45-day period that an individual is not hospitalized. This means those with higher hospital-free days spent less time in the hospital. Those with more hospital-free days showed overall to have a slightly higher gene expression than the other groups. This is expected based on the research in the Introduction, which outlines ABCF1's role in immune system recovery. We would expect those healing from an illness to have a higher expression of ABCF1.

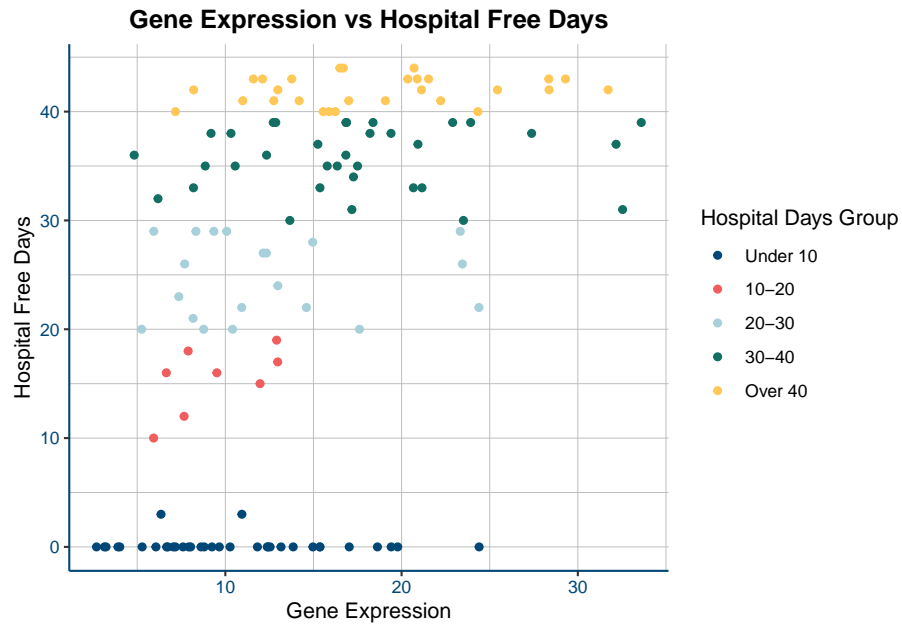


Figure 2: Gene Expression vs Hospital Free Days

The analysis of ABCF1 gene expression compared with hospital-free days through a heatmap in Figure 3 shows a similar trend as the scatterplot, but helps highlight more clearly that the highest rates of gene expression for ABCF1 occurred in patients who spent less time in the hospital.



Figure 3: Gene Expression vs Hospital Free Days

We then compared COVID-19 status, gene expression, and sex through a boxplot seen in Figure 4. This analysis indicated that the ABCF1 gene expression occurred at a higher rate in those who tested positive with COVID-19. Females overall tend to have a slightly higher rate of ABCF1 gene expression as well.

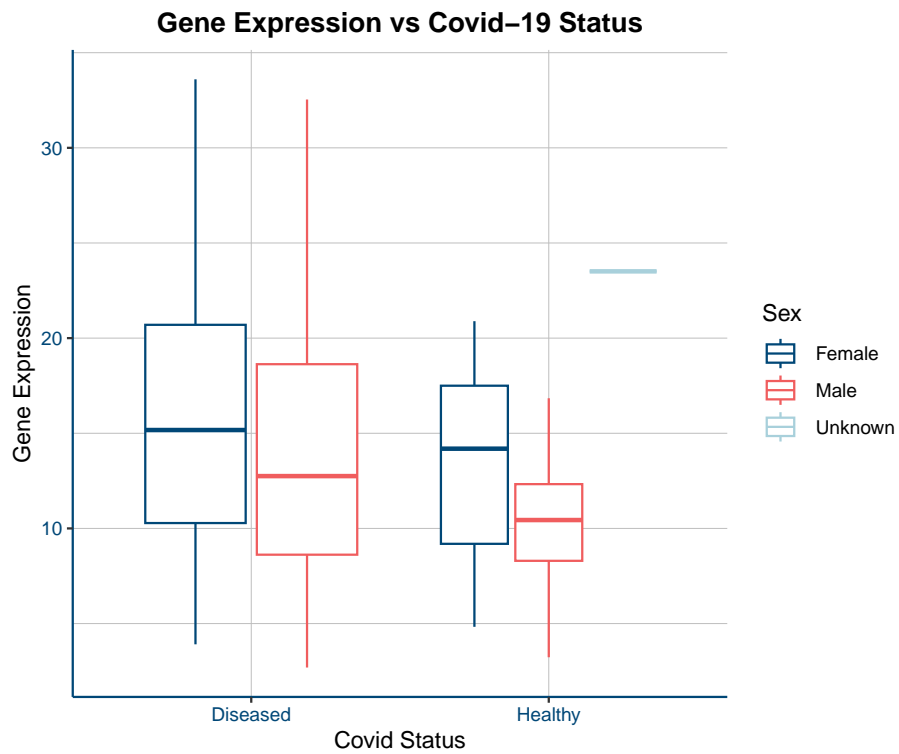


Figure 4: Gene Expression vs Covid-19 Status

A heatmap shown in Figure 5 provided a visualization on how ABCF1's expression would compare to other gene expressions. The plot indicates that while other genes also had some levels of expression, ABCF1 appears to have some of the strongest expression and tended to be seen more in those with COVID-19.

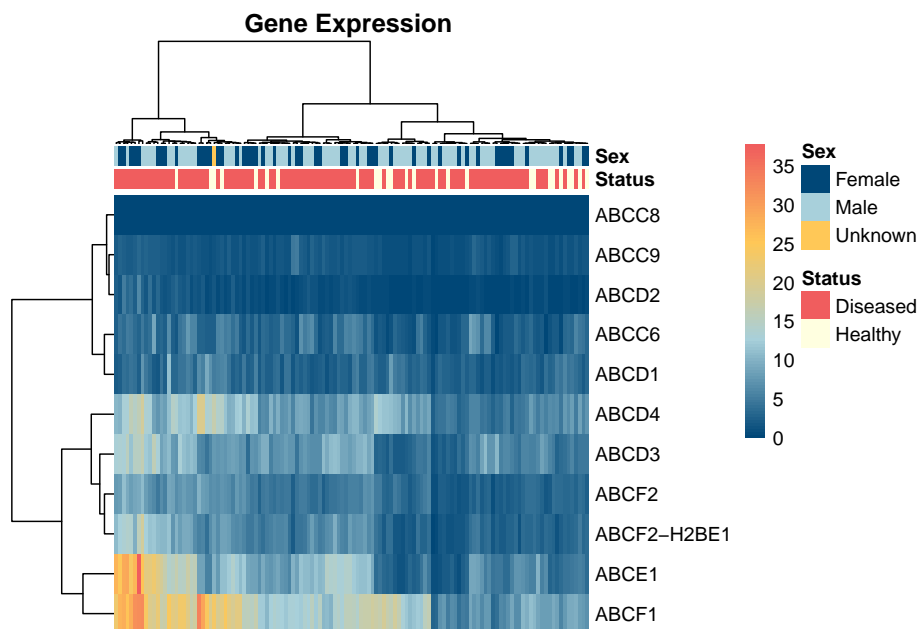


Figure 5: Gene Expression Summary

References

- [1] Hitesh Arora, Sara Morgan Wilcox, Laura Alexandra Johnson, Lonna Munro, Brett Alexander Eyford, Cheryl Gurine Pfeifer, Ian Welch, and Wilfred Arthur Jefferies. The ATP-binding cassette gene ABCF1 functions as an e2 ubiquitin-conjugating enzyme controlling macrophage polarization to dampen lethal septic shock. 50(2):418–431.e6.
- [2] Raivo Kolde. *pheatmap: Pretty Heatmaps*, 2025. R package version 1.0.13.
- [3] Hadley Wickham, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, Max Kuhn, Thomas Lin Pedersen, Evan Miller, Stephan Milton Bache, Kirill Müller, Jeroen Ooms, David Robinson, Dana Paige Seidel, Vitalie Spinu, Kohske Takahashi, Davis Vaughan, Claus Wilke, Kara Woo, and Hiroaki Yutani. Welcome to the tidyverse. *Journal of Open Source Software*, 4(43):1686, 2019.
- [4] Hadley Wickham, Davis Vaughan, and Maximilian Girlich. *tidyr: Tidy Messy Data*, 2024. R package version 1.3.1.
- [5] Sara M. Wilcox, Hitesh Arora, Lonna Munro, Jian Xin, Franz Fenninger, Laura A. Johnson, Cheryl G. Pfeifer, Kyung Bok Choi, Juan Hou, Pamela A.

Hoodless, and Wilfred A. Jefferies. The role of the innate immune response regulatory gene ABCF1 in mammalian embryogenesis and development. 12(5):e0175918. Publisher: Public Library of Science.

- [6] Yihui Xie. knitr. *Journal of Open Source Software*, 2025.
- [7] Hao Zhu. *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*, 2024. R package version 1.4.0.