

# Gene Expression During the Covid-19 Pandemic

Seamus Stein

August 22, 2023

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	<i>APOM</i> Background . . . . .	2
<b>2</b>	<b>Subject Demographics and Clinical Characteristics</b>	<b>2</b>
<b>3</b>	<b>Histogram of <i>APOM</i> Expression</b>	<b>3</b>
<b>4</b>	<b>Scatter plot of <i>APOM</i> Expression by Age</b>	<b>4</b>
<b>5</b>	<b>Boxplot of <i>APOM</i> Expression by ICU Status and Sex</b>	<b>4</b>
<b>6</b>	<b>Heatmap of Gene Expression</b>	<b>5</b>
<b>7</b>	<b>Raincloud Plot</b>	<b>6</b>
<b>8</b>	<b>Acknowledgements</b>	<b>7</b>

# 1 Introduction

The purpose of this study is to examine gene expression data for various genes and their relationship to other covariates during the COVID-19 pandemic [1]. The gene expression data was linked to its metadata for analysis. In this study the covariates of interest are the age and sex of the patients as well as their ICU status. In this project the relationships will be examined by visualizing the data through the use of the ggplot2 package in R. In section 1.1 the biological function of *APOM* gene which was selected for this study will be described.

## 1.1 *APOM* Background

The gene that is of interest is the *APOM* gene. The *APOM* gene encodes a protein that is an apolipoprotein. The encoded protein is secreted through the plasma membrane but remains membrane-bound. It functions in the transportation of lipids. [2]

In the following sections data will be presented on the *LCN2* and *APOA2* gene. The functions of these genes as well as the 10 genes that will be presented in the heatmap will not be discussed in this study.

# 2 Subject Demographics and Clinical Characteristics

Table 1 presents the patient’s demographics and clinical characteristics including summary statistics of the genes and covariate variables of interest in the study.

Variable	n = 74
<b>ICU Status</b> n (%)	
Not Admitted	25 (33.78)
Admitted	49 (66.22)
<b>Sex</b> n (%)	
Female	25 (33.78)
Male	49 (66.22)
<b>Mechanical Ventilation</b> n (%)	
Not on Mechanical Ventilation	36 (48.65)
On Mechanical Ventilation	38 (51.35)
<b>Age</b> Mean (sd)	63.31 (14.58)
<b>Ferritin (ng/l)</b> Mean (sd)	834.26 (935.89)
<b>Lactate (mmol/l)</b> Mean (sd)	0.95 (1.23)
<b>APOM</b> Median [IQR]	3.29 [2.37, 4.53]
<b>LCN2</b> Median [IQR]	136.43 [81.2, 586.76]
<b>APOA2</b> Median [IQR]	0 [0, 0.2]

Table 1: Subject Demographics and Clinical Characteristics

### 3 Histogram of *APOM* Expression

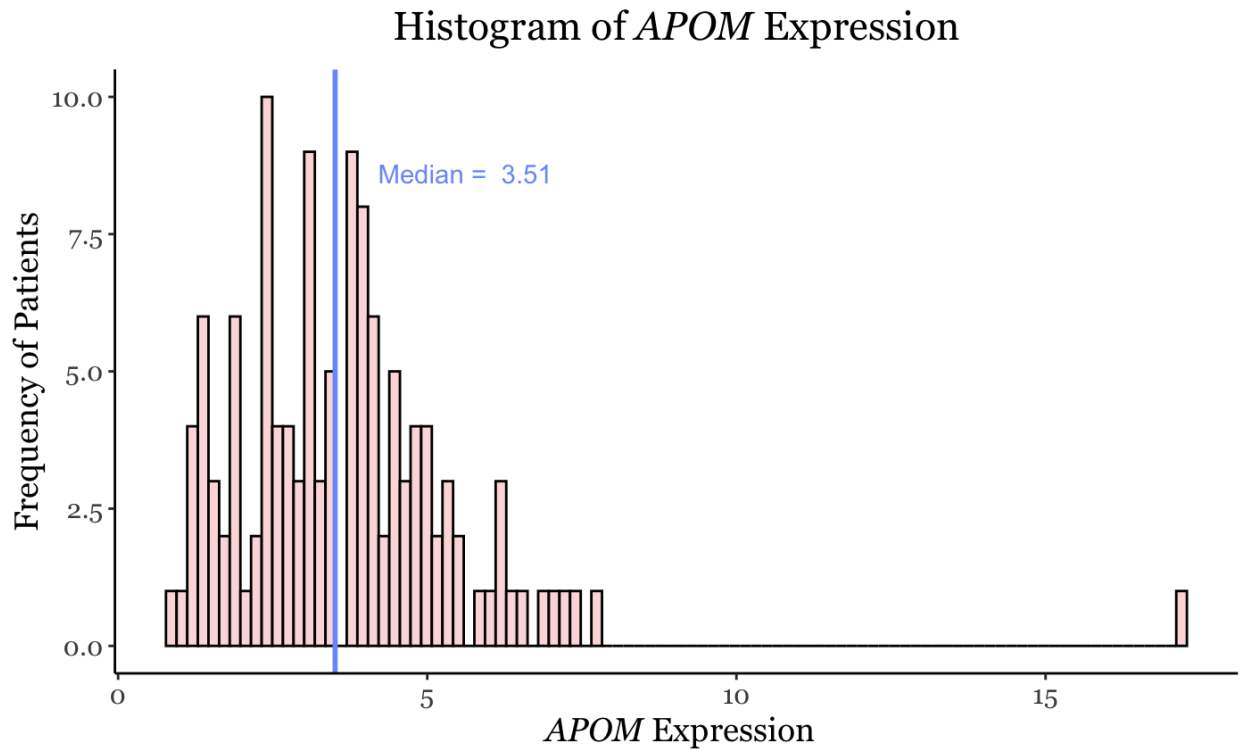


Figure 1: Histogram of *APOM* Expression

When examining Figure 1 we notice that the the median *APOM* expression level is 3.51. This is illustrated by the vertical line in the graph. As seen, the observed distribution of *APOM* expression in the patients of the study is skewed to the right as there is a tail produced by the presence of the outlier on the right hand side of the graph.

## 4 Scatter plot of *APOM* Expression by Age

Next, we will examine the relationship between age, the continuous variable of interest, and *APOM* expression.

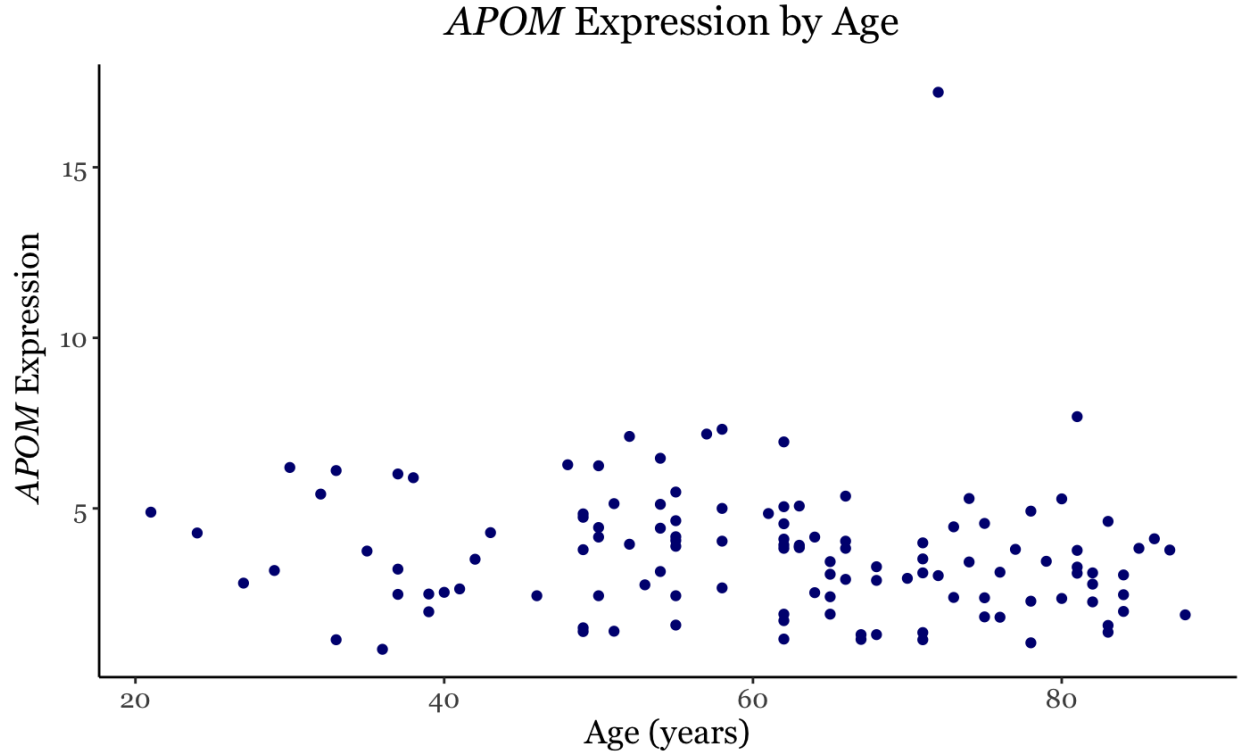


Figure 2: Scatter plot of *APOM* Expression and Age

Figure 2 indicates that there is no relationship between age and *APOM* expression. As age increases the expression of the *APOM* gene does not increase or decrease.

## 5 Boxplot of *APOM* Expression by ICU Status and Sex

The third graph that will be examined is a boxplot. In this graph the *APOM*'s gene expression will be examined by two categorical variables, sex and ICU status, respectively. In the analysis, there was a patient with an unknown in the sex variable. Since this could be interpreted as a missing value it was dropped from analysis.

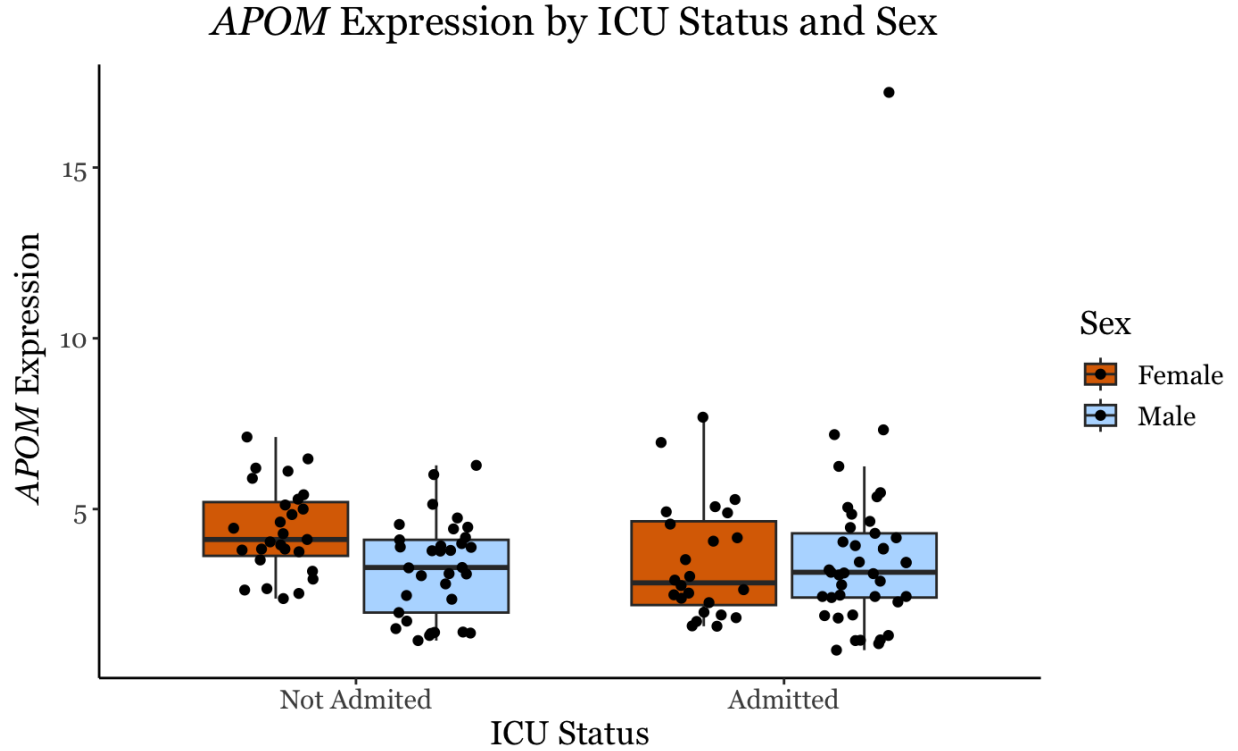


Figure 3: Boxplot of *APOM* Expression by ICU Status and Sex

When examining Figure 3, it appears that among those who do are not admitted to the ICU the median of *APOM* expression is greater among females compared to males. However, among those admitted to the ICU, the median level of expression is greater among males compared to females. The interquartile range for females is noticeably narrower for those not admitted to the ICU compared to those admitted to the ICU. Finally, there is an outlier, a male who was admitted to the ICU has an unusually high level of *APOM* expression.

## 6 Heatmap of Gene Expression

In the following heatmap the expression 10 genes will be visualized. The genes that are considered include *APOM*, *LCN2*, *APOA1*, *APOA2*, *MPO*, *BPI*, *DEFA1*, *DEFA4*, *PRTN3* and *CTSG*.

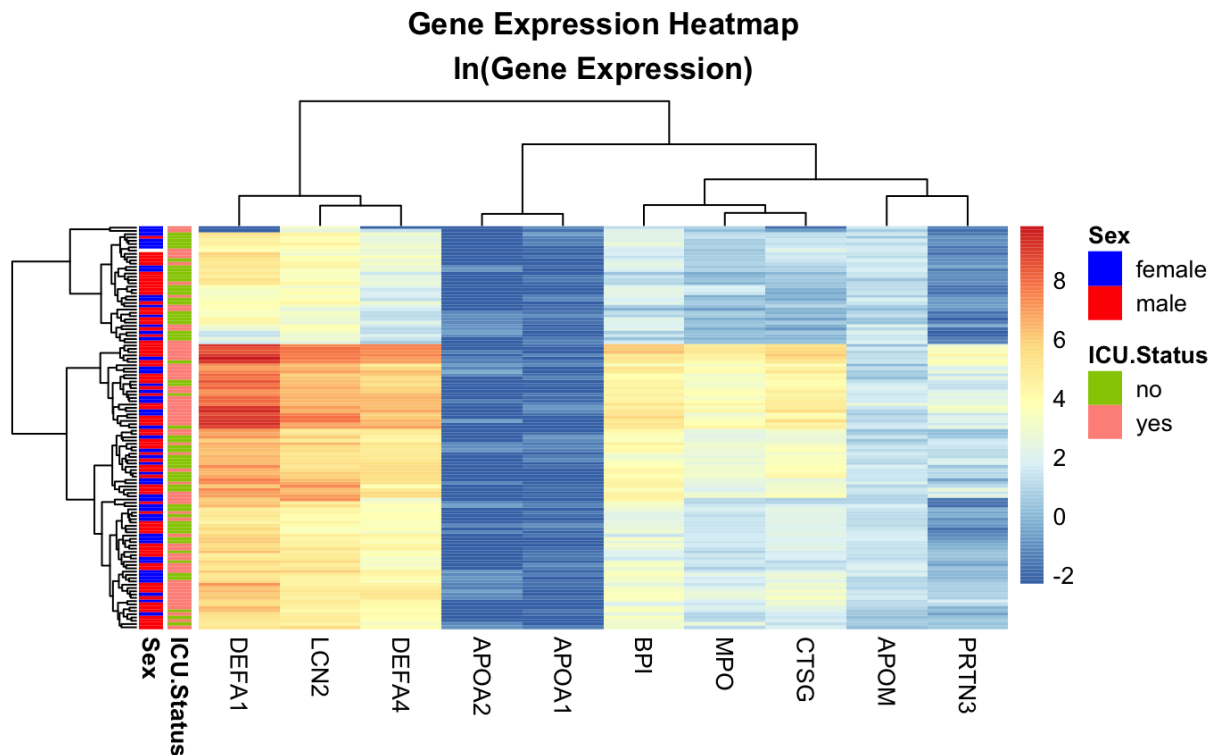


Figure 4: Heatmap of Log Transformed Gene Expression using the Euclidean Algorithm

In Figure 4 there are tracking bars for the two categorical variables of interest, sex and ICU status. The euclidean distance clustering algorithm was performed to create clusters of similar patient's with gene expression. The gene expression was log transformed by means of the natural log. When examining Figure 4 many of the genes have low to medium gene expression in the clusters. The *DEFA1* gene appears to have the greatest level of expression as it has the darkest color according to the tracking bar. When examining the summary statistics for the *DEFA1* gene its maximum expression is 18863.9 or approximately 9.845 on the natural log scale. *DEFA1*'s minimum expression is 0. This range is visualized by the heatmap tracking bar. The next greatest range of expression is found in the *LCN2* gene (the minimum and maximum gene expression is 10.66 and 2928.23 respectively) followed by the *DEFA4* gene (minimum expression is 0.0 and maximum expression is 1981.42). Compared to other genes selected, these genes had the greatest range of expression.

## 7 Raincloud Plot

The final type of plot that will be examined is a raincloud plot. Figure 5 displays a density of all the patients at particular levels of *APOM* expression given by the x-axis. The expression levels are stratified by the

patient's ICU status. Underneath each of the stratified density plots lies a horizontal bar plot. Created with a jitter, one can see for each strata where most of the patients lie.

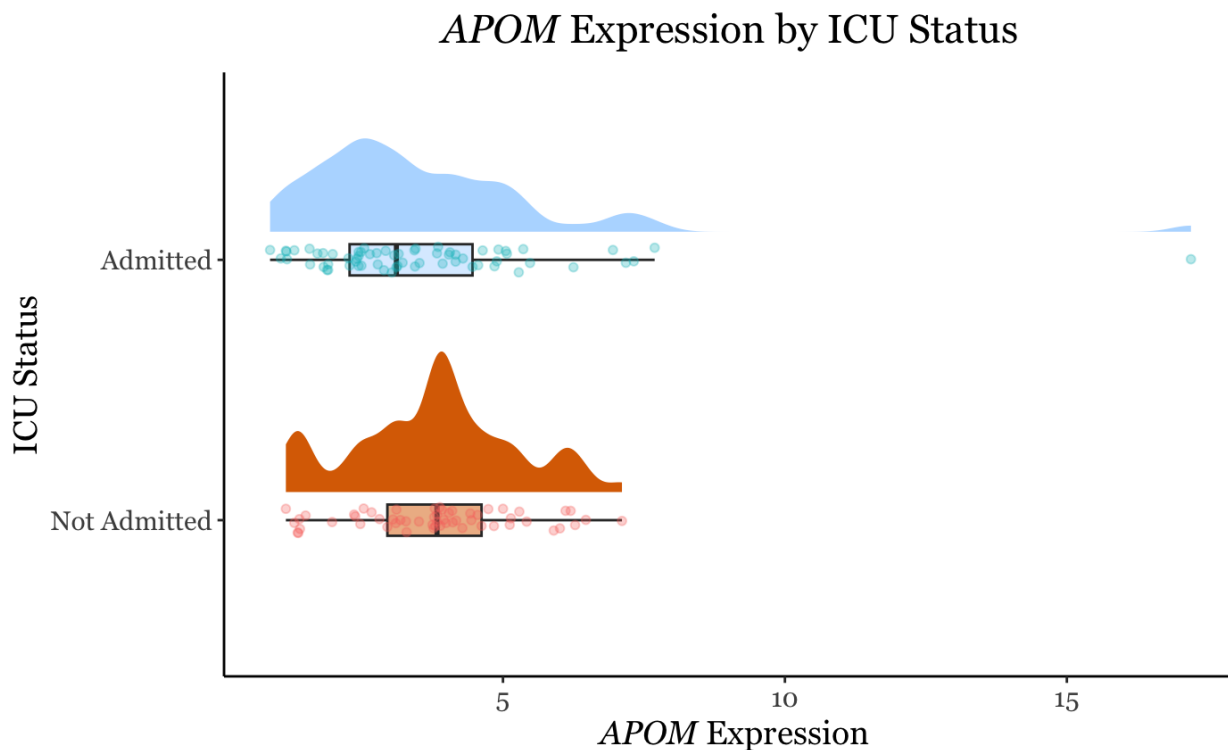


Figure 5: Raincloud Plot: *APOM* Expression by ICU Status

When examining Figure 5, for patients admitted to the ICU there is a greater density below the median resulting in a higher peak. This is in contrast to patients not admitted to the ICU where there is a greater amount of patients surrounding the median expression level for that strata. When comparing patients admitted to the ICU and those not admitted, those not admitted to the ICU appear to have a higher *APOM* expression. It is important to note that this differs from 3 in that we are not stratifying by the patient's sex.

Raincloud plots could be useful and further analysis could be done to compare the distribution of patients and density of expression in different genes of interest. One could also be interested in a different categorical variable and could expand the study to examine and compare and contrast expression levels, and resulting densities, of different genes by multiple categorical variables.

## 8 Acknowledgements

I would like to thank Graham who helped me create better bins on my histogram. I also would like to thank Elizabeth as we worked together on parts of our respective projects together. Other sources that I used

when coding the project can be found in the references.[3][4][5][6][7][8]

## References

- [1] 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, Drake, L. A, Adam, A. P, Schwartz, B. S, Singer, H. A, Swanson, S, Mosher, D. F, Stewart, R, Coon, J. J, & Jaitovich, A. (2021) Large-scale multi-omic analysis of covid-19 severity. *Cell Systems* **12**.
- [2] PubChem. (year?) APOM - apolipoprotein m (human).
- [3] baptiste. (year?) Answer to "using italic() with a variable in ggplot2 title expression".
- [4] mcjudd. (year?) Using italic() with a variable in ggplot2 title expression.
- [5] (year?) Add mean & median to histogram (4 examples) | base r & ggplot2.
- [6] (year?) Graphics in r with ggplot2.
- [7] (year?) how to display multiple row annotations on pheatmap.
- [8] (year?) [solved]-combining paste() and expression() functions in plot labels-r.