Word Count: 1486 Student Number: 190260662

BMS3025 Data Analysis Report of Blanco-Melo et al 2020

Introduction:

Coronaviridae is a family of single-stranded, positive-sense RNA viruses capable of infecting various vertebrate hosts[1]. They are primarily transmitted through respiratory droplets and close contact[2,3]. Generally, these viruses cause respiratory symptoms such as the common cold in humans[4]. However, there have been multiple outbreaks of far more pathogenic variants in recent decades.

The COVID-19 pandemic, which first began in Wuhan, China, in 2019, has been linked to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). As of January 2023, over 100 million people have been infected by the virus, resulting in over 2 million deaths worldwide[5]. The symptoms exhibited due to SARS-CoV-2 infection are typical of those seen in coronaviruses, including a fever and cough[6]. However, more severe cases, particularly in the elderly population, have been shown to cause acute lung injury and acute respiratory distress syndrome, contributing to a high mortality rate compared to previous variants[7].

Respiratory viruses are common and are not limited to coronaviruses. Influenza A Virus (IAV), also known as the flu, is a highly contagious respiratory virus known to cause yearly outbreaks. It is characterised by symptoms such as a fever, cough, and body aches which can be severe in some individuals, particularly the elderly or those with underlying health conditions[8]. Respiratory syncytial virus (RSV) is another common respiratory virus associated with respiratory symptoms, particularly in young children. These can include fever, cough, and difficulty breathing[9].

Despite similarities in the pathology of these respiratory viruses, the infection rate, and the mortality of SARS-CoV-2 are alarming. Therefore, research into the exact mechanisms of the virus has been vital in controlling the pandemic.

Blanco-Melo et al. (2020) compared the transcriptional responses of SARS-Cov-2 with other common respiratory viruses. They found that SARS-CoV-2 differed in the ability of the host to generate a sufficient Interferon I and III response to the virus. This suggests that

individuals with comorbidities, who may have a weaker immune response, may be at a higher risk of experiencing more severe cases of COVID-19 due to the virus's ability to replicate more efficiently. This report aims to explore a subset of the data generated by Blanco-Melo et al. further. Bioinformatic analysis was performed using RStudio, comparing the host response to SARS-CoV-2 with other respiratory viruses.

Methods:

The DEseq2 package within RStudio was used to further analyse the dataset from Blanco-Melo et al. 2020 due to its robustness and accuracy compared to other methods[10]. Firstly, the data was used to create a DEseqDataSet which included only the results from Covid, RSV, and their respective mock conditions (see Figure 1).

```
# The Blanco-Melo et al 2020 dataset is first converted to a data frame to ensure compatability with DEseq2.

dat <- as.data.frame(dat)

# The column names are then modified.

names(dat)[names(dat) == "X"] <- "Gene"

# A matrix of count data and a sample data frame are created. The matrix is subsetted to only include the
Covid, RSV, and their respective mock conditions.

dat3 <- as.matrix(dat[,c(14:17,22:27)])
rownames(dat3) <- dat$Gene
sample_table2 <- data.frame(sample_id=colnames(dat3),
                            condition=c('Mock_RSV', 'Mock_RSV', 'RSV', 'RSV',
                                        'Mock_Covid', 'Mock_Covid', 'Mock_Covid', 'Covid', 'Covid', 'Covid'))

# The DESeqDataSetFromMatrix function is used to create a DESeqDataSet object.The DESeq function is applied to
the DESeqDataSet object.

dds3 <- DESeqDataSetFromMatrix(dat3, colData = sample_table, design=~condition)
dds4 <- DESeq(dds3)
```

Figure 1. Creating a DEseqDataSet for just RSV and SARS-CoV-2

From here, RStudio's plot function was used to generate a volcano plot of RSV versus its mock conditions (see Figure 2).

```
# Objects were created to filter the gene expression data from 'dds4' into results from RSV, COVID and their
mock condidtions.

rsv_results <- results(dds4, contrast = c('condition', 'RSV', 'Mock_RSV'))
covid_results <- results(dds4, contrast = c('condition', 'Covid', 'Mock_Covid'))

# Vectors were  created to differentiate the genes expression via colour. Those genes significantly
upregulated/downregulated compared to their mock conditions would be highlighted.

cols <- ifelse(covid_results$padj <= 0.05 & abs(covid_results$log2FoldChange) >= 1, "red", "black")
cols2 <- ifelse(rsv_results$padj <= 0.05 & abs(rsv_results$log2FoldChange) >= 1, "yellow", "black")

# The volcano plot was  generated with cut-off lines included to indicate significance.

plot(rsv_results$log2FoldChange, -log10(rsv_results$padj),
     col=cols2,
     bg=cols,
     pch = 21,
     xlab= "Log2 Fold Change",
     ylab= "-Log10 Adjusted P Value",
     abline(v = c(-1,1), h = -log10 (0.05), col= "blue", lty=2, lwd=3),
     main = "RSV vs RSV-Mock")
```

Figure 2. Generating RSV Data and its respective Volcano Plot. Log2 fold change was plotted on the x axis and -Log10 adjusted P value on the Y. 'cols' and 'cols2' used to indicate significance.

Results:

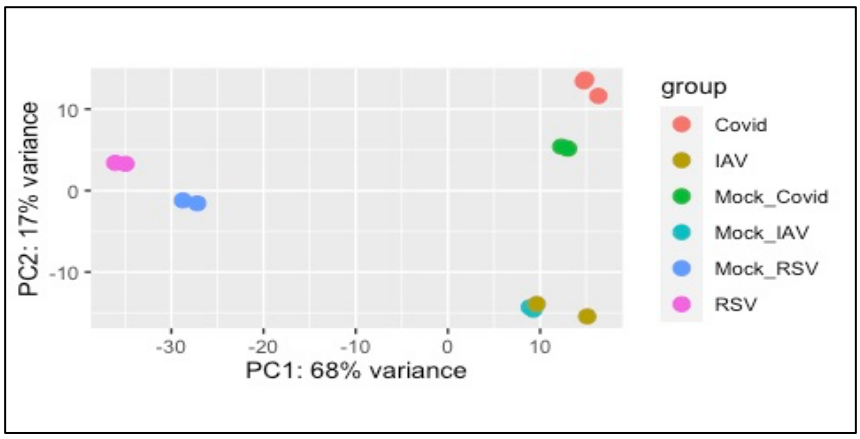

Figure 3. A Principal Component Analysis Plot of the A549 data from RSV, IAV and SARS-CoV-2 infections

A principal component analysis (PCA) **(Figure 3)** was conducted on the data from Blanco-Melo et al. to visualise the relationships between samples based on their gene expression profiles. The first two principal components, which captured 68% and 17% of the data's variance, were plotted on the x- and y-axes. The samples were colored based on the virus species.

The plot shows that the SARS-CoV-2 samples are clustered separately from the RSV samples, with influenza A virus (IAV) samples located closer to SARS-CoV-2 on the plot. These results suggest a significant difference in the gene expression profiles of SARS-CoV-2 and RSV, whereas there is only a slight variance in the profiles of SARS-CoV-2 and IAV. This indicates that SARS-CoV-2 and IAV may not be as genetically distinct from each other as SARS-CoV-2 and RSV.

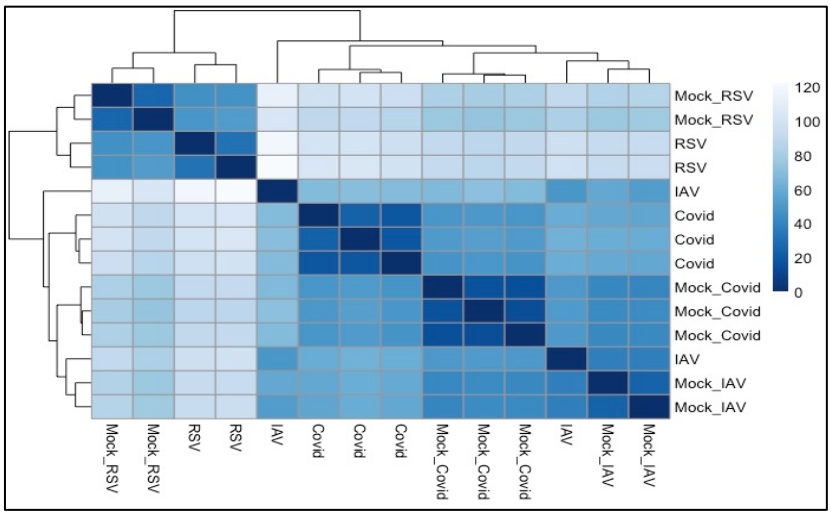

Figure 4. pheatmap of a sample distance matrix of the PCA plot data

The pheatmap package was then used to generate a heatmap of a sample distance matrix of the gene expression data **(Figure 4)**. The heatmap depicts the expression levels of the different viruses. Results showed that RSV and SARS-CoV-2 had the most distinct expression patterns, while IAV exhibited a more similar pattern to SARS-CoV-2. Again, these findings suggest that there may be some overlap in the gene expression patterns of IAV and SARS-CoV-2 but that RSV exhibits a unique expression profile.

Given the goal of specifically comparing the gene expression of SARS-CoV-2 to viruses that are more different from it, it was appropriate to exclude IAV from further analysis and to focus solely on SARS-CoV-2 and RSV.

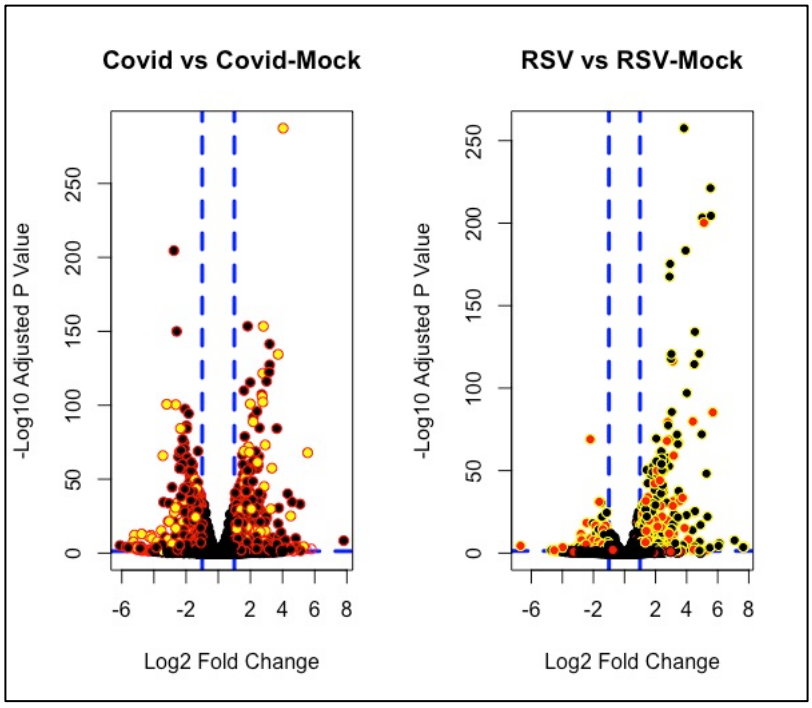

Figure 5. A volcano plot merging the RSV/SARS infection response.

Two volcano plots were generated to visualise the results of the differential expression analysis of COVID and RSV compared to their respective mock conditions **(Figure 5)**. The plot displayed the log2 fold change on the x-axis and the -log10 adjusted p-value on the y-axis. In the Covid plot, genes that were significantly up/downregulated compared to their mock conditions are indicated by a yellow background for RSV and a red outline for COVID. In the RSV plot red background indicates significance from the mock in COVID and the yellow outline in RSV. Black points indicate no significance.

Upon examination, it was observed that some genes were differentially expressed in both COVID and RSV compared to their mock conditions. Specifically, these genes showed significant upregulation or downregulation in both conditions, indicating shared regulation of gene expression between COVID and RSV. On the other hand, some genes were upregulated or downregulated in COVID but not in RSV, while others showed differential expression in RSV but not in COVID. These findings suggest that there is both overlap and unique regulation of gene expression in COVID and RSV compared to their mock conditions.

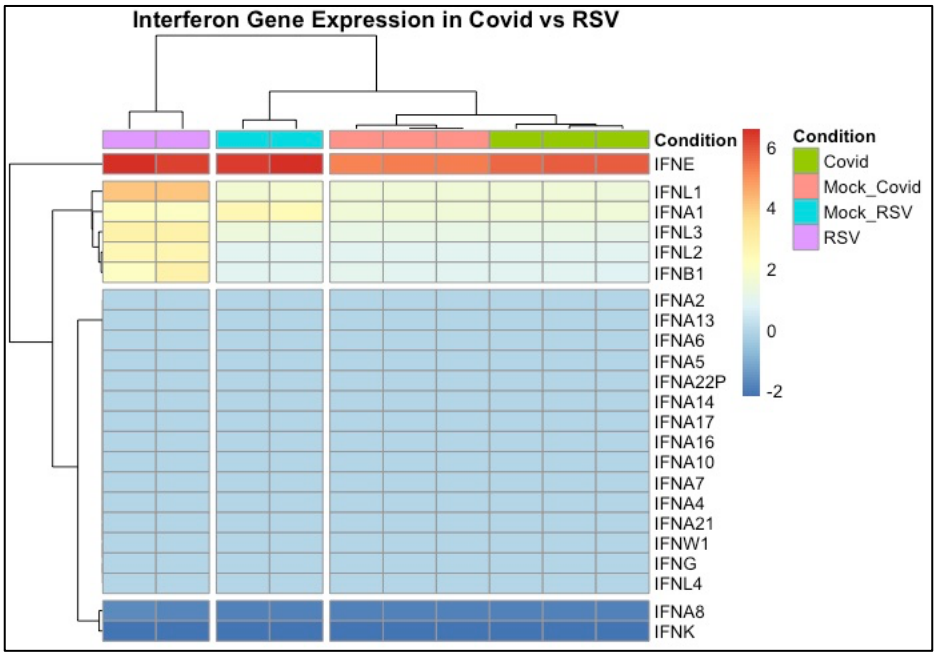

Figure 6. A heatmap of Interferon gene expression comparing RSV and SARS-CoV-2 infection.

The pheatmap package was then used to investigate the interferon gene expression patterns in COVID and RSV infections. A vector of the interferon genes from the Blanco-Melo et al. dataset was created, which allowed us to create the following heatmap **(Figure 6).** The heatmap revealed differences in the expression of several interferon genes between the two viruses. IFNE, IFNL1, IFNA1, IFNL3, IFNL2, and IFNB1 were found to be more highly upregulated in RSV compared to COVID, while IFNA8 and IFNK were downregulated to a similar level in both viruses. The remaining interferon genes showed similar expression levels between each virus and their respective mocks. These findings highlight key differences in SARS-CoV-2's ability to affect the host interferon response compared to RSV. These mechanisms may be contributing to the often-worse symptoms associated with COVID-19.

Discussion:

The results of our bioinformatic analysis demonstrate both similarities and differences in the interferon gene response to RSV versus SARS-CoV-2 infections. Our findings suggest that the differential expression of interferon genes may play a role in the severity of symptoms observed in these viral infections. Specifically, the upregulation of certain interferon genes in RSV may contribute to the generally milder symptoms observed in RSV infections, as interferons have antiviral properties and play a vital role in the immune response to viral infections.

On the other hand, the downregulation of IFNA8 and IFNK in both COVID and RSV, and the similar expression of the remaining genes, may partially explain the similarities in the symptoms displayed by these two viruses. However, it is important to note that multiple other factors likely influence the pathology of COVID and RSV. Further research is necessary to understand the complex immune responses to these viral infections fully.

The original SARS-CoV virus has been shown to use various mechanisms to avoid a type I interferon response in the host. SARS-CoV nucleocapsid protein has been shown to interfere with the function of interferon regulatory factor 3 (IRF3)[11]. Other mechanisms include non-structural protein 3 (Nsp3) blocking the phosphorylation of IRF3 as well as further inhibition via accessory proteins such as ORF3b and 6[12,13]. Recent studies sequencing the SARS-CoV-2 genome have shown 82% similarity between the two nucleotides and homology between proteins[14]. This suggests that SARS-CoV-2's inhibitory effects on type I and III IFN responses are due to similar mechanisms to those of SARS-CoV.

In contrast to SARS-CoV-2, RSV appears to have a less pronounced effect on the Interferon-I/III pathway. Interestingly, RSV has been shown to have NS1 and NS2 proteins, which suppress the induction of type 3 interferons in vitro[15]. Selvaggi et al. suggest that RSV may cause the activation of interferon subtypes which could indirectly regulate antiviral mechanisms[16]. Nevertheless, further research into exactly how RSV suppresses the interferon response will be key in developing our understanding of the role of interferon suppression in respiratory disease.

In future studies, it would be interesting to compare the interferon gene response to other more novel COVID-19 variants to gain a more comprehensive understanding of the commonalities and differences in immune evasion strategies among these viruses. It would also be valuable to continue investigating the specific molecular mechanisms by which SARS-CoV-2 and RSV affect the interferon pathway and examine the potential consequences of this inhibition for host defense and disease severity.

Finally, studying the interferon gene response in patients with COVID-19 or those dealing with long-covid side effects may provide further insight into the potential utility of interferon-based therapies for this disease. Overall, this research has provided important insights into the complex interplay between viruses and the host immune system, which may be identified as potential targets for future antiviral therapies.

References:

1 Cui J, Li F, Shi Z-L. Origin and evolution of pathogenic coronaviruses. Nature reviews microbiology. 2019;17:181-192.

2 Chan JF-W, Yuan S, Kok K-H, To KK-W, Chu H, Yang J, Xing F, Liu J, Yip CC-Y, Poon RW-S. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The lancet. 2020;395:514-523.

3 Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KS, Lau EH, Wong JY. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. New England journal of medicine. 2020.

4 Weiss SR, Navas-Martin S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. Microbiology and molecular biology reviews. 2005;69:635-664.

5 World Health Organization. Coronavirus disease (COVID-19) pandemic. . [cited 2023 January]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

6 Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. The lancet. 2020;395:507-513.

7 Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, Niemeyer D, Jones TC, Vollmar P, Rothe C. Virological assessment of hospitalized patients with COVID-2019. Nature. 2020;581:465-469.

8 Dawood FS, Iuliano AD, Reed C, Meltzer MI, Shay DK, Cheng P-Y, Bandaranayake D, Breiman RF, Brooks WA, Buchy P. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. The Lancet infectious diseases. 2012;12:687-695.

9 Hall CB, Weinberg GA, Iwane MK, Blumkin AK, Edwards KM, Staat MA, Auinger P, Griffin MR, Poehling KA, Erdman D. The burden of respiratory syncytial virus infection in young children. New England Journal of Medicine. 2009;360:588-598.

10 Love MI, Anders S, Huber W. Analyzing RNA-seq data with DESeq2. Bioconductor. 2017;2:1-63.

11 Kopecky-Bromberg SA, Martínez-Sobrido L, Frieman M, Baric RA, Palese P. Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. Journal of virology. 2007;81:548-557.

12 Devaraj SG, Wang N, Chen Z, Chen Z, Tseng M, Barretto N, Lin R, Peters CJ, Tseng C-TK, Baker SC. Regulation of IRF-3-dependent innate immunity by the papain-like

protease domain of the severe acute respiratory syndrome coronavirus. Journal of Biological Chemistry. 2007;282:32208-32221.

13 Lu X, Pan Ja, Tao J, Guo D. SARS-CoV nucleocapsid protein antagonizes IFN-β response by targeting initial step of IFN-β induction pathway, and its C-terminal region is critical for the antagonism. Virus genes. 2011;42:37-45.

14 Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. Cell host & microbe. 2020;27:325-328.

15 Spann KM, Tran K-C, Chi B, Rabin RL, Collins PL. Suppression of the induction of alpha, beta, and gamma interferons by the NS1 and NS2 proteins of human respiratory syncytial virus in human epithelial cells and macrophages. Journal of virology. 2004;78:4363-4369.

16 Selvaggi C, Pierangeli A, Fabiani M, Spano L, Nicolai A, Papoff P, Moretti C, Midulla F, Antonelli G, Scagnolari C. Interferon lambda 1–3 expression in infants hospitalized for RSV or HRV associated bronchiolitis. Journal of Infection. 2014;68:467-477.