Pathology of SARS-CoV-2 in Calu3 Cells

## Abstract

The COVID-19 pandemic has led to millions of deaths since the start of the pandemic. Understanding the SARS-CoV-2 virus that causes COVID-19 can help prevent further spread of the disease and make better treatments for COVID-19. Viruses use the cells in our body to replicate themselves. To use the cells in the body, viruses have to employ techniques to stop the immune system and use the host's cellular machinery. Many of the deaths caused by COVID-19 disease are a result of respiratory failure after the SARS-CoV-2 virus infects the lungs. By comparing RNA sequence data of lung cells infected with SARS-CoV-2 and uninfected lung cells, we can better understand the pathology of the SARS-CoV-2 in the lungs. Information about the pathology may include molecules to signal the immune system, viral RNA that changes cellular metabolism or viral RNA that affects protein coding in the cell. Infected and uninfected cells were sequenced at 3, 5, and 8 hours after infection. The cells used were from the Calu3 immortal cell line of bronchial epithelial cells. After comparing the RNA expression data of infected lung cells with uninfected lung cells, we found many of the differentially expressed genes were used to signal the immune system or relate to cell metabolism. Our results suggest that cells infected with covid 19 show many differences in gene expression. Understanding these differences can help researchers find better treatments for COVID-19 and better prepare us for future coronaviruses.

## Introduction

The COVID-19 pandemic has gripped the world since its rapid spread began in early 2020. The outbreak has been caused by an RNA virus, SARS-CoV-2. Although the Sars-CoV-2 virus is new, there have been historical outbreaks of very similar viruses in the same coronavirus family. The SARS (severe acute respiratory virus) outbreak in 2003, started in China and spread to other countries. The 2003 outbreak is the first new contagious disease in the 21st century. Similar to COVID-19 and other respiratory illnesses, SARS can spread through airborne saliva droplets. SARS can also spread via surfaces with viral droplets. (3, 13) Another historic outbreak was caused by the Middle Eastern Respiratory Virus (MERS). Similar to COVID-19 and SARS, MERS is also caused by a coronavirus. First reports of the disease were in the Arabian peninsula. The symptoms of MERS are similar to that of COVID-19 and other infections: fever, cough, and shortness of breath. The MERS virus is more fatal than SARS-CoV-2, killing 3 to 4 out of every 10 patients. (4, 12) Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the cause of the ongoing coronavirus disease 19 (COVID-19) pandemic.

To understand the difference in RNA expression between infected and uninfected cells, the data we used here were Calu3 cells infected with SARS-CoV-2 as well as uninfected ones. Calu3 cells were infected with the SARS-CoV-2 virus. Most of the calu3 cells were successfully infected, and cells were sampled at 3, 5, and 8 hours post infection. (1)

The question we seeked to answer is to explore the differences in RNA expression between infected and uninfected cells. We approached this by performing differential expression analysis in this research. We used PCA plot to check results, biological processes ontology to see the effects outside cells, and dot plot to see many of the relevant genes found in the analysis. We also performed unsupervised analysis. We performed clustering methods with varying numbers of genes and generated heatmaps with annotation sidebar. We performed chi-squared tests for each pairwise comparison of clustering results from each method as well.

Many of the genes between the two groups were significantly expressed. After analyzing the infected and uninfected cells, we found three main differences between infected cells and uninfected cells: a decrease in RNA expression in infected cells, changes to cellular metabolism, and an immune response in the infected cells. Other studies have found that COVID-19 causes similar changes in infected cells, including the decrease in host expression and an immune response. Understanding the genetic changes that SARS-CoV-2 causes to infected lung cells will be an important part of developing new medications for the disease. It will also become an essential part of understanding the pathology of COVID-19.

## Methods

We performed differential expression analysis using DESeq2, to find genes that were significantly differentially expressed between the two groups and normalize the data. For the analysis we used the counts for each gene in each of the samples, and labels for what group that sample belonged to. After running the differential analysis, we used those results to make a PCA plot of the samples and a Volcano plot. We used the function plotPCA() to generate the PCA plot. The input parameters were vsd and intgroup = c(“condition”). To generate the volcano plot, we used EnhancedVolcano() function and the input parameters were deseq\_df, lab = deseq\_df$Gene, x = "log2FoldChange", y = "padj", and pCutoff = 0.01. Using the normalized gene counts, we used clusterProfiler with the biological processes ontology to understand the biological processes that the differentially expressed genes are related to and see groupings of those genes.

We also performed unsupervised analysis. This analysis used two different clustering algorithms, K-Means and consensus clustering. We used the K-Means algorithm to identify the number of centroids and then allocated every data point to the nearest cluster. With K-Means, we clustered the data into two groups by specifying k =2. The parameters we used in function heatmap() are as.matrix(subset), name = subset, and column\_km = 2. We did the rerun of the clustering method using 10, 100, 1000, and 10000 genes to observe the BSS/TSS ratios. We used the consensus clustering algorithm to assess the stability of the clusters. We decided how many iterations we wished to run and chose a set of K values (1-6) to test. Therefore, the input parameters for consensus clustering method were 1, 2, 3, 4, 5, and 6. We also generated one heatmap with the consensus clustering algorithm by specifying k = 5. We performed chi-squared tests for each pairwise comparison of clustering results from each method as well. We generated three tables to record the input parameters of these clusters and methods. The input parameters for the starting group for both cluster 1 (infected) and cluster 2 (uninfected) clusters in consensus cluster and K-Means were 4. The input parameters for consensus cluster in cluster 1 was 6 and for consensus cluster in cluster was 2. The input parameters for K-Means in cluster 1 was 5 and for K-Means in cluster 2 was 3.

All code and analysis are available at our github repository: <https://github.com/michaelmez39/CIS4930CovidInLungs>

## Results

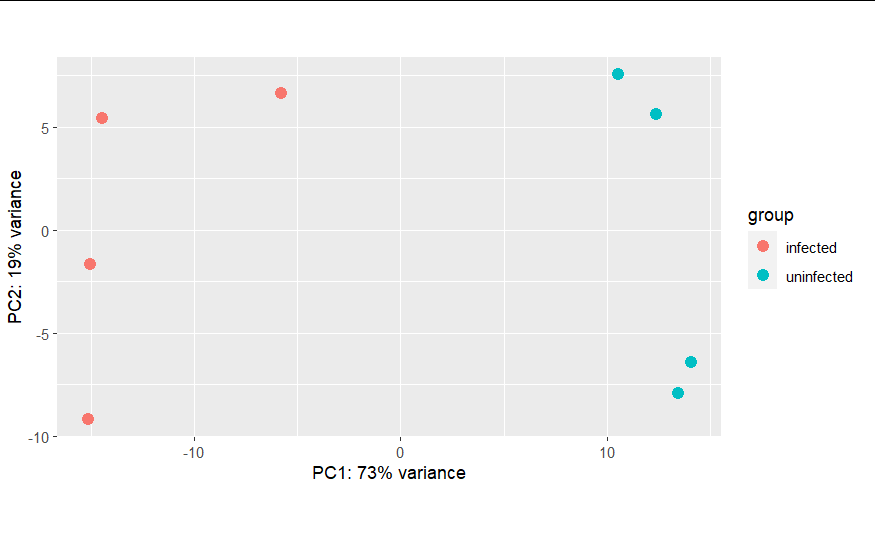


Figure 1: PCA Plot of infected and uninfected groups

In Figure 1, after running differential expression analysis, we can see that the uninfected and infected groups are clearly in separate groups, especially along PC1. This suggests that there is a strong difference between the two groups.

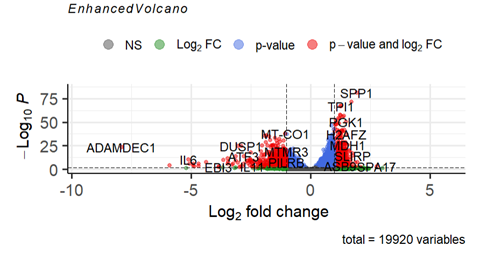


Figure 2: Volcano Plot of differentially expressed genes

Using the volcano plot, we can examine some of the differentially expressed genes between the two samples. On the left are the genes more strongly expressed by the infected group and on the right are the genes more strongly expressed by the uninfected group. Some of the interesting genes expressed by the infected cells are Interleukin 6 and14 both cytokines involved in the immune response. ADAMDEC1 is a signaling molecule for dendritic cells, often involved in the viral immune response. On the uninfected side we see SPP1, involved in type 1 immune response (5). Also we see TPI3 and MDH1 both involved in cellular metabolism (6).

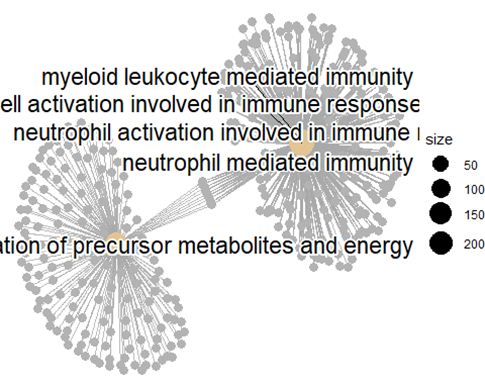


Figure 3: Cnet Plot from BP Gene Ontology

Looking at the results from the gene ontology, we can see two clusters around immune response and mediation of precursor metabolites and energy.

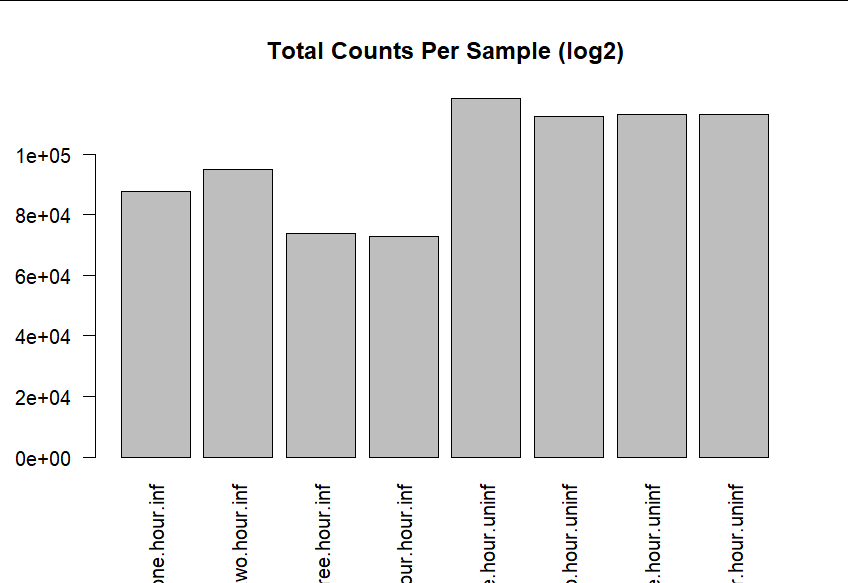


Figure 4: Total RNA expression of each sample (excluding viral RNA)

Looking at the total RNA expression counts, we see that the infected cells have a much lower amount of RNA expression as compared to the uninfected cells. Since our data shows only host RNA sequences, a decrease in the total RNA sequences that we see indicates a decrease in host RNA, not necessarily viral RNA.

|  |  |
| --- | --- |

Figure 5: Heatmaps of genes separated into clusters (left: kmean, right: consensus cluster)

Lastly, we can examine the results of clustering by looking at the heatmap and dendrograms of

the clusters. For consensus cluster results, we found K=5 to be a good number of groups to minimize the CDF. We can see that generally, consensus cluster has the groups split between infected and uninfected. Looking back at Figure 4, we see that many of these cluster groups are related to the amount of RNA expression in the sample. Although this does illustrate the changes in RNA expression between the two groups, it may not show much more than that.

Comparing the RNA expression data between infected and uninfected cells, we can see several patterns emerge between the infected and uninfected cells. Looking at both the gene ontology and differential analysis we find that infected cells show increases in genes involved in the immune response and metabolic processes. From the results of clustering, we can see that healthy cells however show much more RNA expression overall. These results highlight some important differences between cells infected with Covid-19 and uninfected cells.

## 

|  | Starting Group | Consensus Cluster |
| --- | --- | --- |
| Cluster 1 / Infected | 4 | 6 |
| Cluster 2 / Not Infected | 4 | 2 |

𝚾2 = 0.266

P-value = 0.6956

|  | Starting Group | KMeans |
| --- | --- | --- |
| Cluster 1 / Infected | 4 | 5 |
| Cluster 2 / Not Infected | 4 | 3 |

𝚾2 = 0

P-value = 1

|  | Consensus Cluster | KMeans |
| --- | --- | --- |
| Cluster 1 / Infected | 6 | 5 |
| Cluster 2 / Not Infected | 2 | 3 |

𝚾2 = 0.266

P-value = 0.6956

Figure 6: Statistical Significance of Clustering

One of the main weaknesses of our results was that the clustering did not provide much useful analysis or significant information for several different reasons. Having a low number of samples means that it is very difficult to show significance in the clusters that formed, as seen in Figure 6, the statistics don’t work well with a small number of samples. Another problem with the clustering is that the changes in RNA expression from the virus seem to have a large effect on the clustering that covers any other signals in the data. One last consideration of our results is that we found a difference in cellular metabolism between the two groups, however because the cells we used are derived from a cancer cell line their metabolism may not function exactly like regular lung cells.

After looking at the results of our analysis, there are three main differences we found between infected and uninfected cells: altered cellular metabolism, immune signaling molecules, and decrease in host rna expression. From the gene enrichment analysis, we can see that infected cells have an altered cellular metabolism and that they begin to make signalling molecules related to viral infection, like interleukin-6. The gene ontology analysis gives similar results to manually looking at the significantly expressed genes, showing clusters around cellular metabolism and immune response. Other researchers have found similar patterns, with many looking at the role cytokines like interleukin play in Covid-19 progression (8, 11).

The decrease in gene expression found in infected cells has been found in previous studies 1, 2. It is believed that COVID-19 reduces host expression and increases the expression of virus RNA. By increasing viral RNA expression, viruses can replicate faster. We discarded expression data not found in both infected and uninfected groups, so we cannot see SARS-CoV-2 RNA sequences in our data. Since our data shows only host RNA sequences, a decrease in the total RNA sequences that we see only indicates a decrease in host RNA. In our data we do clearly see a reduction in the amount of RNA data, which can be seen by comparing the infected and uninfected heatmaps of the count

Our results show many possible future areas of research. One interesting area of research would be to further examine the immune response of Covid19 and to do a comparison to similar viruses like MERS. Other researchers have looked at the effects of Sars-Cov-2 on mitochondria and some believe that a therapeutic medicine could target the changes to mitochondria to help reduce the symptoms of covid19. (14)

## Conclusion

By examining the difference in RNA expression between uninfected and infected cells, we found many important parts of Covid-19 pathology, like some of the signalling molecules it triggers in lung cells related to the immune response and changes it makes to cellular metabolism. This helped us understand our goal to identify differences in RNA expression between infected and uninfected lung cells. One of the most clear results of our analysis is that cells infected with Covid-19 have reduced RNA expression compared to uninfected cells. To improve our research in the future, using a larger number of samples would help to find more significant differences in infected and uninfected cells. Some of the main goals of similar research could be to find therapies that are effective in treating Covid-19, understanding comorbidities like diabetes and understanding how Covid-19 spreads between people.

## 

## References

1 Article our data is from:

[SARS-CoV-2 uses a multipronged strategy to impede host protein synthesis](https://www.nature.com/articles/s41586-021-03610-3)

2 Covid two prong host suppression:

[A two-pronged strategy to suppress host protein synthesis by SARS coronavirus Nsp1 protein](https://www.nature.com/articles/nsmb.1680)

3 Historic SARS Virus 2003 Outbreak:

[The aetiology, origins, and diagnosis of severe acute respiratory syndrome](https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(04)01172-7/fulltext)

4 Middle East Outbreak:

[Middle East Respiratory Syndrome Coronavirus nsp1 Inhibits Host Gene Expression by Selectively Targeting mRNAs Transcribed in the Nucleus while Sparing mRNAs of Cytoplasmic Origin | Journal of Virology](https://journals.asm.org/doi/10.1128/JVI.01352-15)

5 SPP1 about:

[SPP1 - Osteopontin precursor - Homo sapiens (Human) - SPP1 gene & protein](https://www.uniprot.org/uniprot/P10451)

6 TPI1 about:

[TPI1 - Triosephosphate isomerase - Homo sapiens (Human) - TPI1 gene & protein](https://www.uniprot.org/uniprot/P60174)

7 Sars-CoV-2 cellular metabolism:

[SARS-CoV-2 infection rewires host cell metabolism and is potentially susceptible to mTORC1 inhibition](https://www.nature.com/articles/s41467-021-22166-4)

8 SARS-CoV-2 immune response:

[JCI Insight - COVID-19 and RA share an SPP1 myeloid pathway that drives PD-L1+ neutrophils and CD14+ monocytes](https://insight.jci.org/articles/view/147413)

9 Viruses and Metabolism:

[Viral activation of cellular metabolism - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S0042682215000847)

10 Krebs Cycle:

[The Krebs Cycle — Harnessing Chemical Energy for Cellular Respiration](https://www.sigmaaldrich.com/US/en/technical-documents/technical-article/research-and-disease-areas/metabolism-research/citric-acid-cycle)

11 More on Immune System and Coronavirus 2: [Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients](https://www.science.org/doi/10.1126/science.abc6027)

12 Introduction about previous COVID outbreak in middle east:

<https://www.cdc.gov/coronavirus/mers/index.html>

13 Introduction about previous COVID outbreak in China:

<https://www.who.int/health-topics/severe-acute-respiratory-syndrome#tab=tab_1>

14 Covid 19 and cellular metabolism

[Virus-Host Cell Interactions and the Viral Life Cycle: Basic Science to Therapeutics: Mitochondrial metabolic manipulation by SARS-CoV-2 in peripheral blood mononuclear cells of patients with COVID-19](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7816428/)