## Comparing Genomic Expression Between Adults and Children Infected with Covid-19

### Abstract

Bioinformatics seeks to solve large-scale biological problems using computational methods. With the recent emergence of the SARS-Cov-2 infection, bioinformatic methods can be used to find differences between adult infections and child infections. Utilizing high throughput sequencing expression data shows the responses induced by children and adults over time. These responses can then be used to find how gene expression differs between adults infected with the SARS-Cov-2 infection and children infected with SARS-Cov-2 infection. We were not able to find statistically significant differences between the gene expressions of infected adults and children. We utilized clustering methods to look for gene expressions that may have formed distinct clusters and we used chi squared test of independence to evaluate independence between variables. We have seen from these results that the expression differences between adults and children are not significant and they did not show any dependence either. These results show that the gene expression between adults and children that are infected with SARS-Cov-2 do not show a significant difference. This study can be further expanded upon and improved in many ways. The results we found may have been limited by the dataset we selected. Different data sets and different methods may show results that we were unable to prove. Improvements may allow this study to prove useful in developing better medication and vaccines.

### Introduction

The data we used was an RNA high throughput sequencing dataset that had the gene signatures of Covid-19 throughout the duration of the disease in two different age groups, children and adults. We seeked to answer whether there was a difference in the way gene signatures of Covid-19 were expressed between children and adults.Our hypothesis was that the signatures would be expressed differently between the two, as it is often said that children have much milder symptoms of Covid-19 than their parents.The results of our project showed that there was no significant difference in the way gene signatures were expressed between children and adults infected by Covid-19. Our results seem reasonable considering the Covid-19 vaccine and treatments are the same between adults and children. Furthermore, viruses and infections generally have the same effect between children and adults, so it makes sense that the gene signatures expressed make sense. The broader impact of our results is a better understanding of the way Covid-19 is expressed in children and adults. It helps us understand how the virus works and this knowledge can be used in developing vaccines and fighting the virus. Furthermore, it may help in giving insight into the potential different or similar treatment options between children and adults affected with Covid-19.

### Methods

There were various methods used during this project. These method types include expression analysis, differential analysis, enrichment analysis, unsupervised learning, and statistical analysis. To perform expression analysis, two techniques were used: differential and enrichment analysis. Differential analysis was done using the DESeq2 function to find differentially expressed genes. These genes were then represented visually with a volcano plot. Enrichment analysis was used to perform gene ontology using the gProfiler2 method.

We then utilized unsupervised learning techniques such as clustering to find any correlations among the genes. We then performed hierarchical clustering using hclust and PAM clustering using pam. A heatmap was created to represent the results of the clustering. Finally we performed statistical analysis to determine correlation. We used Pearson’s chi squared test of independence to find any associations among the genes. These methods allowed us to determine if our hypothesis was met.

### Results

We were able to find an answer to our original question regarding whether different ages (children vs adults) resulted in varying gene signatures of Covid-19. We found that there appeared to be no significant differences between the expressions of gene signatures of Covid-19 between adults and children.The differential expression analysis identifies the genes with significantly varying expression. The Chi squared test of independence greatly helped us reach an answer. The p values from our chi squared tests were quite high and not statistically significant with an alpha of 0.05. This meant we could not find evidence that there was a relationship or a dependence between the age group and the gene signature expressions of Covid-19. Overall, even though our hypothesis was false, this project was still successful as we were able to reach conclusions and gain insight. A bioethical issue involving our project was the privacy of individuals involved in the dataset. We were able to avoid this because the dataset had any identifiable information of individuals hidden. A weakness in our project is the reliance on one main statistical test, the chi squared test of independence.. An improvement to this project would be possible by applying different techniques when finding significant genes and expanding our use of clustering to find significant results. Future work on this project might involve comparing how gene expression differs between adults and children who have received the Covid-19 mRNA vaccine.

### Conclusion:

In our search to find out whether there was a difference in genes expressed by Covid-19 between children and adults, we found that our findings did not support our hypothesis. Our hypothesis was that the signatures would be expressed differently between the two. We came up with this hypothesis as it is often said that children have much milder symptoms of Covid-19 than their parents, so we believed that this may have been due to fewer genes being expressed in children than in adults. However, our findings showed that there was no significant difference in the way that the gene signatures were being expressed between the two groups. From assignment 3, where we did predictive modeling, the chi squared test of independence showed that the age group and the expression of the gene signatures were independent as our p-value was not sufficiently low enough to reject the null hypothesis that there is a relationship or that they are not independent. Our clusters were not distinct enough to show that there were clearly defined groups for the genes of children and adults. In assignment 2, the heat map also did not show that there were many significantly differentiated expressed genes. Although out PCA plot in assignment 2 did appear to show some sort of trend in that there were two groups, one adult group and one children group, this was not backed up by any of our further exploration and predictive models. Ultimately, we found that age did not have a true impact on how the gene signatures of Covid-19 were expressed.

### References:

Ackerman, Margaret E, et al. “A Robust, High-Throughput Assay to Determine the Phagocytic Activity of Clinical Antibody Samples.” *Journal of Immunological Methods*, U.S. National Library of Medicine, 7 Mar. 2011, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3050993/.

Aleebrahim-Dehkordi, Elahe, et al. “Human Coronaviruses SARS-COV, MERS-COV, and SARS-COV-2 in Children.” *Journal of Pediatric Nursing*, Elsevier Inc., 2021, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7580518/.

Anderson, Elizabeth M, et al. “Seasonal Human Coronavirus Antibodies Are Boosted upon SARS-COV-2 Infection but Not Associated with Protection.” *Cell*, Elsevier Inc., 1 Apr. 2021, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7871851/.

Bunyavanich, Supinda, et al. “Nasal Gene Expression of Angiotensin-Converting Enzyme 2 in Children and Adults.” *JAMA*, American Medical Association, 16 June 2020, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7240631/.

“Geo Accession Viewer.” *National Center for Biotechnology Information*, U.S. National Library of Medicine, https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE166190.

“GSM5065639: Cluster5 Child1 Positive interval5; Homo Sapiens; RNA-Seq.” *National Center for Biotechnology Information*, U.S. National Library of Medicine, https://www.ncbi.nlm.nih.gov/sra?term=SRP304757.

Meštrović, Dr. Tomislav. “Study Shows Children and Adults Have a Comparable Expression of SARS-COV-2 Entry-Related Genes.” *News*, 27 May 2020, https://www.news-medical.net/news/20200527/Study-shows-children-and-adults-have-a-comparable-expression-of-SARS-CoV-2-entry-related-genes.aspx.

Pierce CA;Sy S;Galen B;Goldstein DY;Orner E;Keller MJ;Herold KC;Herold BC; “Natural Mucosal Barriers and Covid-19 in Children.” *JCI Insight*, U.S. National Library of Medicine, https://pubmed.ncbi.nlm.nih.gov/33822777/.

“Understanding Covid-19 in Children and Adults (ID 699562).” *National Center for Biotechnology Information*, U.S. National Library of Medicine, https://www.ncbi.nlm.nih.gov/bioproject/PRJNA699562.

Vono, Maria, et al. “Robust Innate Responses to SARS-COV-2 in Children Resolve Faster than in Adults without Compromising Adaptive Immunity.” *Cell Reports*, The Authors., 5 Oct. 2021, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8440231/.

Github Repository:

* <https://github.com/ammarsyed/BioinformaticsProject>