Original article

Comprehensive RNA-Seq analysis of molecular signatures between COVID-19 and Idiopathic Pulmonary Fibrosis

Campbell L

Submitted: July 23, 2023, Revised: version 1, September 13, 2023
Accepted: September 16, 2023

Abstract

The aetiology of respiratory illnesses such as COVID-19 and Idiopathic Pulmonary Fibrosis (IPF) is known to be influenced by intricate genetic factors. An in-depth investigation of gene expression patterns was conducted using RNA-Seq analysis in RStudio. The objective was to elucidate the shared molecular pathways between COVID-19, IPF, and healthy control samples. Differentially expressed genes were identified through the implementation of advanced statistical techniques, including DESeq2, Limma Voom, and Limma Trend. The production of a high-quality gene count dataset required extensive preprocessing and normalization measures. The procedure was carried out by performing the necessary steps. Enriched pathways and potential biomarkers were unveiled through the use of functional annotation and visualization tools. The study findings provide novel insights into the genetic associations between COVID-19 and IPF, which have significant implications for identifying therapeutic targets and developing personalized treatment strategies.

Keywords

COVID19, Idiopathic Pulmonary Fibrosis, RStudio, RNA sequencing, Differentially Expressed Genes, Biomarkers, Gene expression, Signaling pathway, Genetic association

Luca Campbell, Lovejoy High School, 2350 Estates Pkwy, Lucas, TX 75002, USA.
lucaacampbell@gmail.com

Introduction

SARS-CoV-2-associated coronavirus disease 2019 (COVID-19) primarily affects pulmonary parenchyma. According to recent studies, SARS-CoV-2 utilizes the ACE2 cell membrane receptor to invade host cells with the assistance of transmembrane protease serine 2. The pulmonary vasculature is a possible target for SARS-CoV-2, which can lead to infection (1,2). Furthermore, chest CT scans of patients with COVID-19 indicated notable vascular thickening in comparison to those with non-COVID-19 viral pneumonia (3). Similarly, idiopathic pulmonary fibrosis (IPF) is a devastating disease characterized by the progressive scarring and stiffening of lung tissue (4). This chronic interstitial lung disease leads to irreversible damage and a decline in lung function over time, ultimately resulting in respiratory failure and death (5). Currently, the IPF affects more than 3 million individuals globally, which places a significant strain on healthcare systems (6). Both IPF and COVID-19 share some common risk factors and molecular characteristics. Furthermore, studies have shown that patients with IPF are more susceptible to severe complications of COVID-19 (7).

The latest developments in single-cell RNA sequencing technology have made it possible to thoroughly analyze the diversity of cells within complex tissues. Single-cell RNA sequencing allows for the profiling of individual cells' transcriptomes. This technique can identify unique cell populations and their gene expression patterns, providing valuable information about the development of various diseases (8).

To date, there has not been a direct comparison between COVID-19 vs. pulmonary fibrosis. Our objective was to identify the molecular pathways and cellular mechanisms that could contribute to the development and progression of pulmonary fibrosis and COVID-19. Additionally, we aimed to compare the gene expression profiles of COVID-19 and pulmonary fibrosis to identify any unique features. Our research results could offer valuable understanding regarding the development of COVID-19 and IPF and could aid in the creation of specific diagnostic markers for these severe complications.

Methods

snRNA-seq dataset

The snRNA-seq data were retrieved from publicly available databases such as GEO, and ArrayExpress. The snRNA-seq datasets contained (i) Gene IDs (2) Cell IDs, and (3) matrix of counts. The datasets were selected based on their relevance to the present research, which focused on pulmonary vasculature in patients with lethal COVID-19 and idiopathic pulmonary fibrosis. The datasets obtained contained snRNA-seq data from pulmonary vasculature cells in both diseases. Once the relevant data was identified, it was downloaded and preprocessed to make sure high-quality reads were included in downstream analysis. The snRNA-seq data was then subjected to post-genome alignment which included parsing, filtering, and normalization of the input data files.

snRNA-seq data analysis

Quality control

After the retrieval and primary screening, the dataset was further processed in R (Version 4.0.1). After downloading and decompressing data in R, the data was converted into quality metrics and poor-quality cells were excluded. Minimal filtering was performed to exclude cells with less than 100 UMIs. The cells with UMI count greater than 500 indicated deeper sequencing and were included.

Datasets normalization and scaling

The sequencing data were then normalized using the function known as NormalizeData, which was included in the Seurat package (v3.1). Following the completion of the normalization, the top 2000 genes with extremely variable expression levels were determined through the application of the FindVariableFeatures function. Next, the data were scaled through the application of the ScaleData function.

Principal Component Analysis (PCA)

The data obtained were subjected to principal component analysis (PCA) using the RunPCA function to summarize it. After conducting the PCA analysis, the FindNeighbour and FindCluster functions were used to cluster the datasets and create a shared nearest-neighbour (SNN) graph. The clustered data was then visualized using uniform manifold approximation and projection. This was performed by running a runUMAP function.

Markers identification

The FindMarkers function was used to calculate the marker genes for each of the obtained clusters. The subset () function was utilized to annotate clusters and divide them into major cellular lineages based on the expression levels of canonical markers. Individual sub-clustering for epithelial, stromal, immune, and endothelial subsets was performed.

Results

A collection of gene counts for COVID-19 and IPF (Idiopathic Pulmonary Fibrosis) samples, totaling 10 samples, was obtained. The dataset included 4 COVID-19 samples, 4 IPF samples, and 2 healthy (control) samples. Our next goal was to normalize the RNA-Seq data using the typical data processing methodology. We used a desktop computer with a dual-core CPU and 8 GB of RAM running the Windows 8.1 operating system to complete the normalization work. Before utilizing iDEP95 (9), an interactive online tool to find differentially expressed genes, we used the R programming language to normalize the gene counts. We then used three distinct methods—DESeq2, Limma Trend, and Limma Voom—to find the differentially expressed genes. We used the KNIME analytics data mining tool to create machine learning models after identifying the differentially expressed genes. Lastly, we chose the experimental parameters for the prediction models.

Read count and normalization

The number of reads that encompass a certain feature, such as a gene, is referred to as the

read count for that feature. The substitution of a gene's read counts is what is done to quantify the gene's expression. Read counts must be connected to the reference genome and added together before differential expression analysis can begin. This is done so that interpreted genes may be formed. The per library raw read count is shown as a bar graph in Figure 1. The total count method was utilized to normalize read counts, which enabled comparisons to be made across separate experiments. This experiment made use of the DESeq2 method, which is included in R. This algorithm first converts the read counts data into a stable dispersion. The DESeq2 package's primary goals are to normalize, visualize, and perform statistical analysis on differentially expressed genes that make use of high-dimensional read counts. The normalized read count is shown in Figure 2.

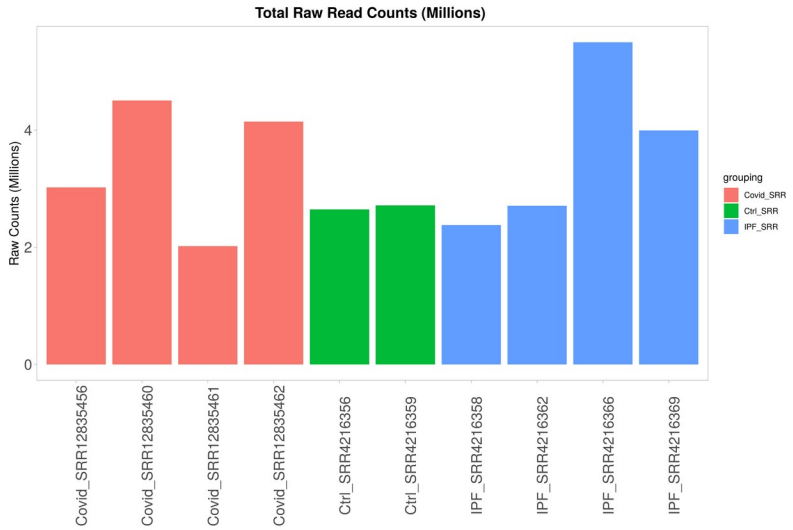

Figure 1. Aggregate read count for selected libraries. Read counts are shown in millions and depict raw data.

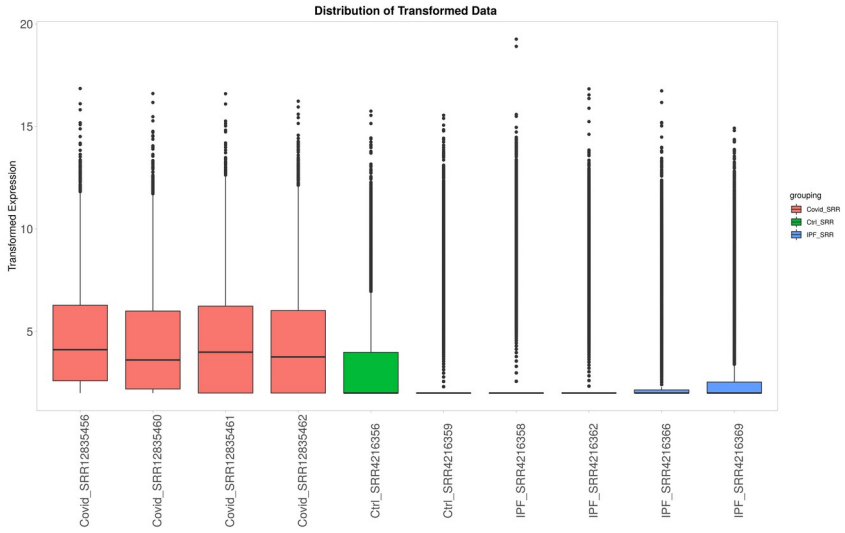

Figure 2. Normalized Distribution of Transformed Expression Data. DESeq2 package in RStudio was applied to normalize the raw data by using the total count method.

Differentially expressed genes analysis

We performed high-throughput sequencing analysis on the GSE86618 dataset to find highly important genes in the RNA-Seq expression of endothelial cells. Our goal was to examine the gene expression patterns of healthy controls and patients with COVID-19 and IPF in the dataset we used. Using the R programming language, the raw gene counts from the dataset were pre-processed and standardized. Differentially expressed genes (DEGs) were discovered using iDEP95, a web-based application, and three distinct approaches: DESeq2, Limma Trend, and Limma Voom. Significant regulators were identified using a combination of up-and down-regulated genes and statistical methods including fold change and adjusted p-values. The highest number of DEGs were found to be influenced in COVID-19 vs IPF where 8608 genes were up-regulated, while 768 genes were down-regulated (Figure 3). Through Venn diagrams, further insight into the number of DEGs influenced mutually in each group was also studied. Two DEGs were up-regulated and twenty-eight were down-regulated in COVID-19, IPF, and control. Among COVID-19 vs control and IPF vs control, 116 DEGs were up-regulated and 592 were down-regulated (Figure 4 & 5).

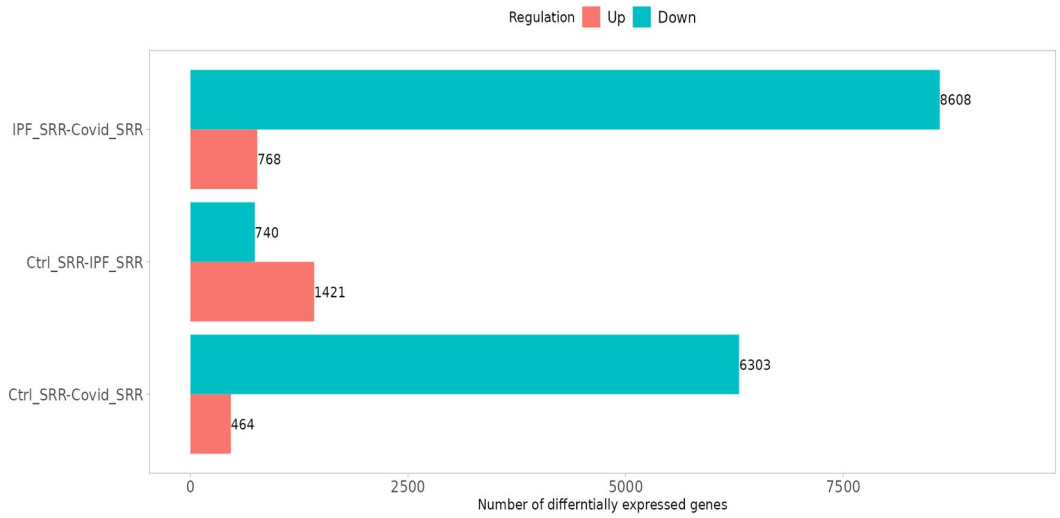

Figure 3. Number of differentially expressed genes in COVID-19, IPF, and control. The comparison is performed in three pairs: COVID-19 vs Control, IPF vs Control, and COVID-19 vs IPF.

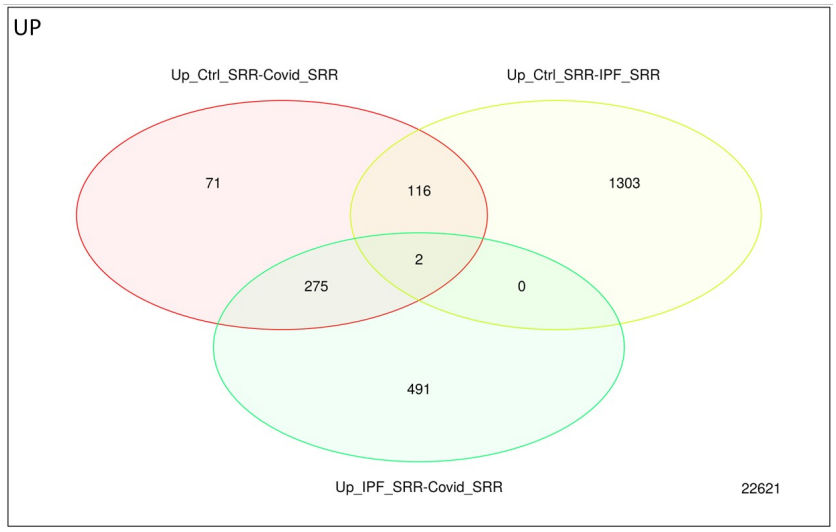

Figure 4. Venn diagram of up-regulated DEGs in IPF, COVID-19, and control.

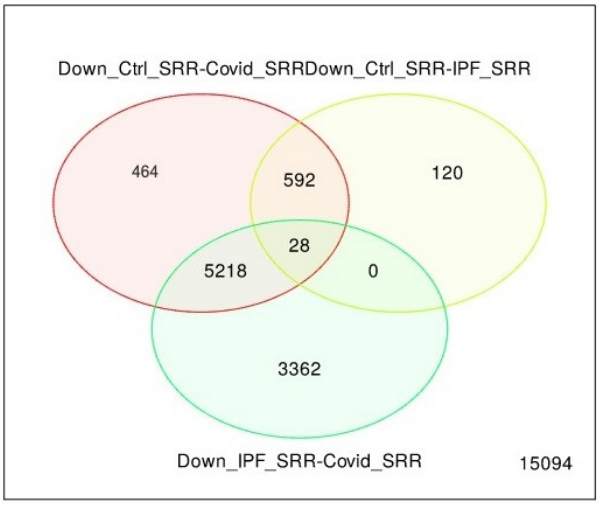

Figure 5. Venn diagram of up-regulated DEGs in IPF, COVID-19, and control.

When analyzing differential expressions, the fold change is a conventionally used statistical metric. It calculates the ratio of means between diseased and healthy groups, expressed as a log2 fold change (log2FC), to quantify the amount of gene expression alterations between the two respiratory illnesses. Up-regulation is indicated by a log2FC value > +1.0, whereas down-regulation occurs at a log2FC value of 1.0. In addition, the false discovery rate (FDR) was decreased by calculating adjusted p-values (padj) to account for multiple testing. An adjusted p-value cutoff of 0.05 was chosen that indicated 5 per cent of the tests could be false positives. The MA plots and Violin plots for up-regulated and down-regulated DEGs in COVID-19 and IPFs in comparison to control are shown in Figures 6 through 11. These diagrams show the average gene expression in COVID-19, IPF, and control groups plotted against the log2 fold change.

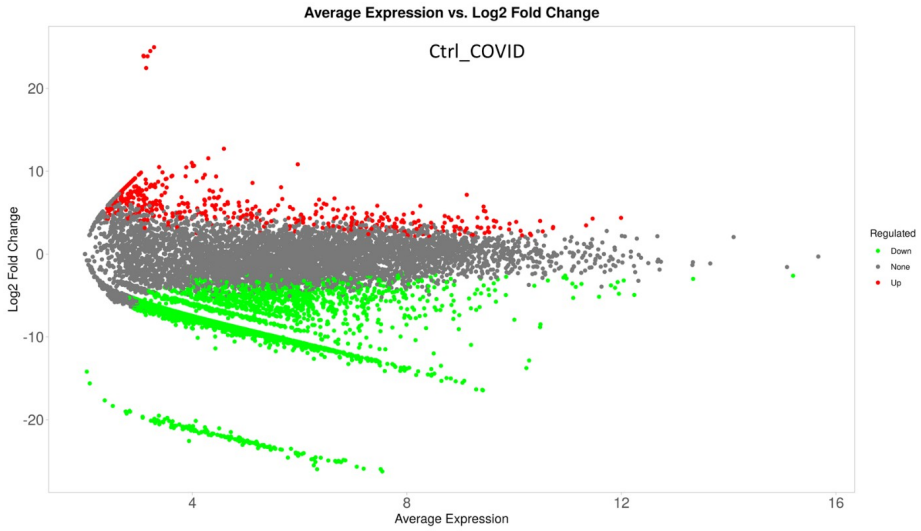

Figure 6. MA Plot of differentially expressed genes in COVID-19 patients in comparison to control on log2FC ≥ 2 using DESeq2. Green dots indicate expression upregulation, red dots indicate expression down-regulation, and unaffected genes are shown with grey.

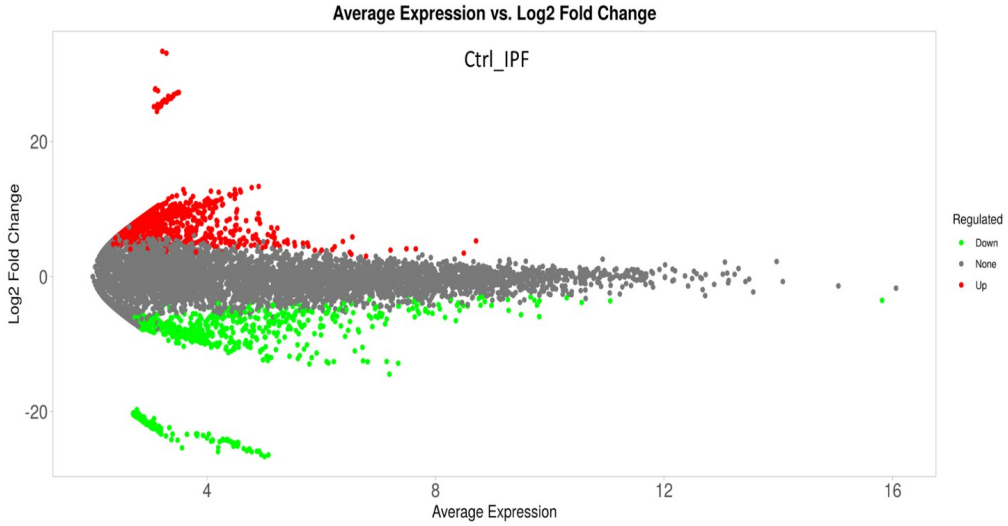

Figure 7. MA Plot of differentially expressed genes in IPF patients in comparison to control on log2FC ≥ 2 using DESeq2. Green dots indicate expression upregulation, red dots indicate expression down-regulation, and unaffected genes are shown with grey.

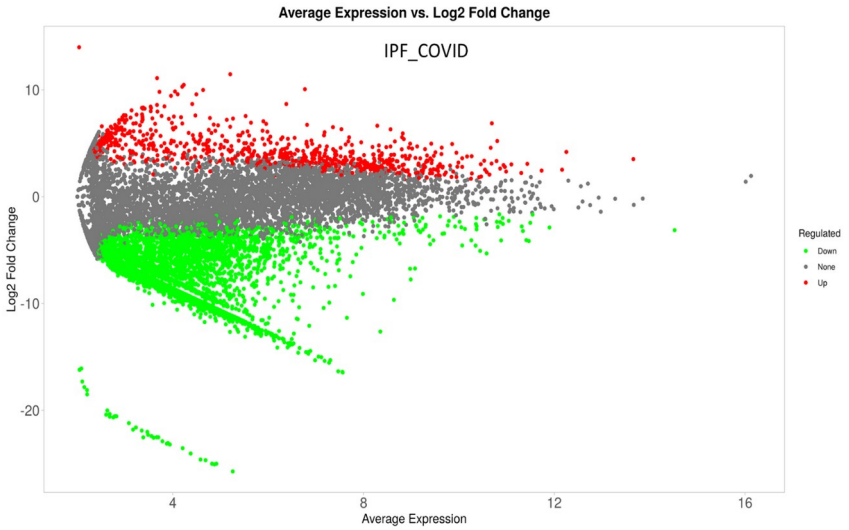

Figure 8. MA Plot of differentially expressed genes in IPF patients in comparison to COVID-19 on log2FC ≥ 2 using DESeq2. Green dots indicate expression upregulation, red dots indicate expression down-regulation, and unaffected genes are shown with grey.

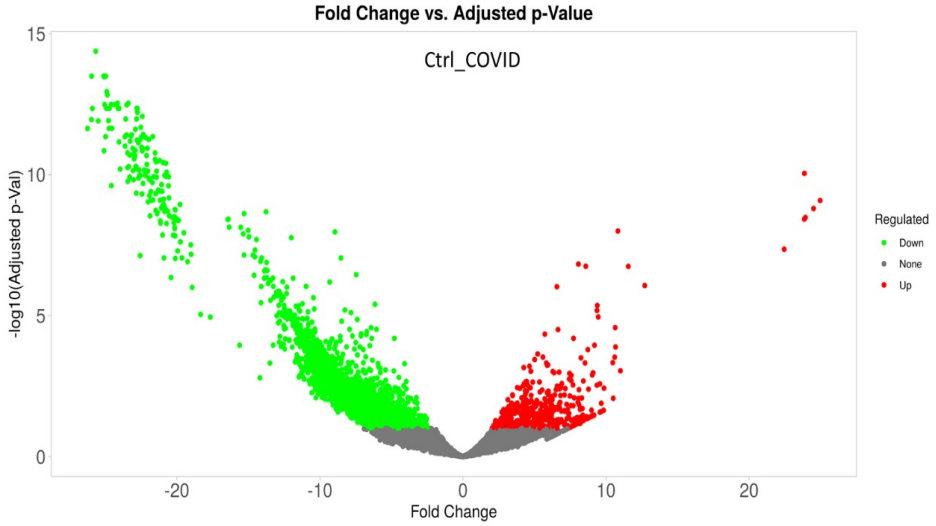

Figure 9. Volcano Plot of differentially expressed genes in COVID-19 patients in comparison to control on log2FC ≥ 2 using DESeq2. Green dots indicate expression upregulation, red dots indicate expression down-regulation, and unaffected genes are shown with grey.

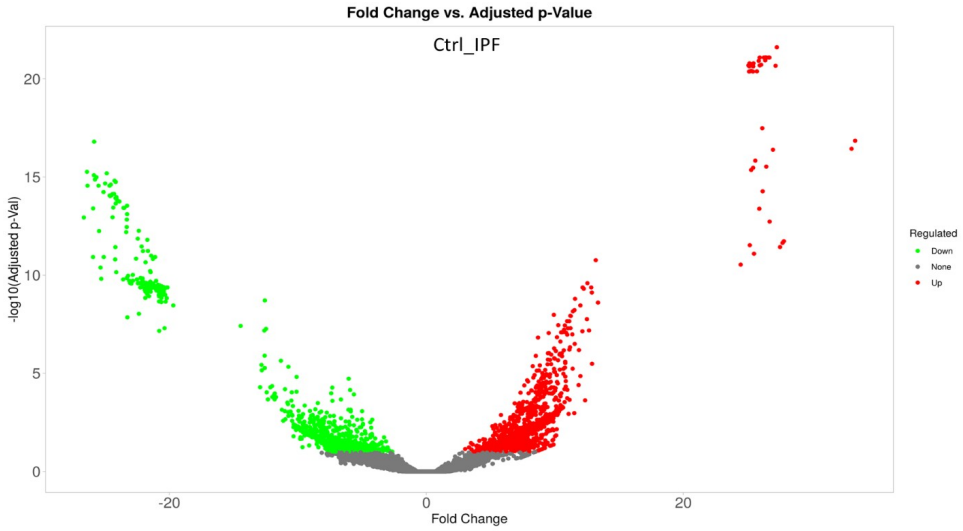

Figure 10. Volcano Plot of differentially expressed genes in IPF patients in comparison to control on log2FC ≥ 2 using DESeq2. Green dots indicate expression upregulation, red dots indicate expression down-regulation, and unaffected genes are shown with grey.

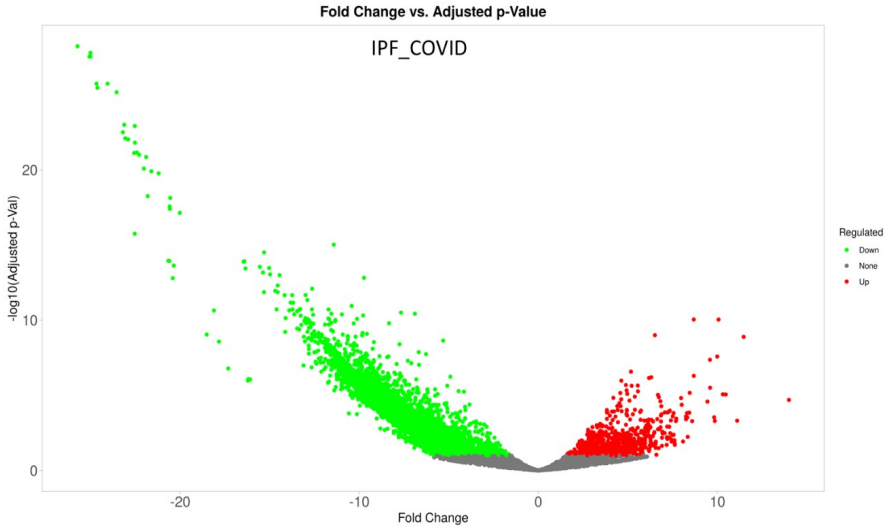

Figure 11. Volcano Plot of differentially expressed genes in COVID-19 patients in comparison to IPF on log2FC ≥ 2 using DESeq2. Green dots indicate expression upregulation, red dots indicate expression down-regulation, and unaffected genes are shown with grey.

Set of gene markers identified

Twenty genes were identified, based on P-value and adjusted P-value, whose expression was found to be differentially expressed in all three groups. These genes include CFB, CF1, CFTR, COL12A1, COL15A1, COL18A1, COL3A1, COMP, CRABP2, CTHRC1, DCLK1, FHL1, LTBP1, MMP1, MUC5B, RTEL1, SFTPA2, SFTPC, SULF1, and VCAN. The expressions of COL3A1, COL14A1, COL15A1, COMP, CRABP2, CTHRC1, DCKL1, MUC5B, and RTEL1 were high in COVID19 patients in comparison to control and IPF. These genes do not have much expression difference in IPF and control group. The expression of SFPA2, and SFTPC was low in IPF patients relative to control and COVID-19 patients. The expression of CFB was down-regulated in COVID-19 patients in comparison to IPF but was up-regulated in COVID-19 relative to healthy individuals. The expression of the gene markers is shown in Figures 12.

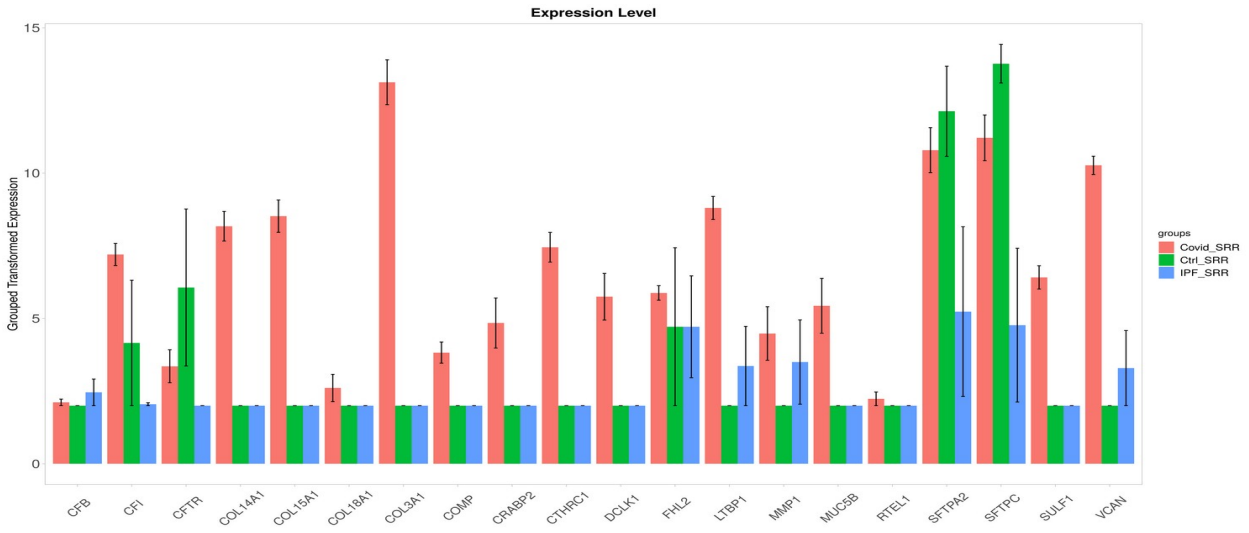

Figure 12. Comparison of various Genes reported to be linked with IPF and COVID-19

Gene set enrichment analysis

The gene set enrichment analysis was performed to understand the overall association between DEGs, and pathways involved. A hierarchical cluster was constructed and the association between DEGs was determined and involved pathways were identified (Figure 13). The top 15 pathways for hierarchical clustering were then identified and sorted for further assessment (Figure 14 and Table 1). Most of the pathways identified were stimulated in response to external chemicals or substances. Immune pathways, pathways activated in response to stress, and pathways linked with host-pathogen interactions were also identified. More than 600 genes were involved in response to the organic substances' pathway with a ≥ 1.3-fold enrichment.

These pathways were then further investigated for their potential association with each other and were subjected to network analysis. Network analysis revealed that genes associated with cell adhesion regulation were not associated with any network. However, genes activated in response to external stimuli, stress, biotics, chemical compounds, or endogenous substances were found to have an association with each other. Their network is presented in Figure 15.

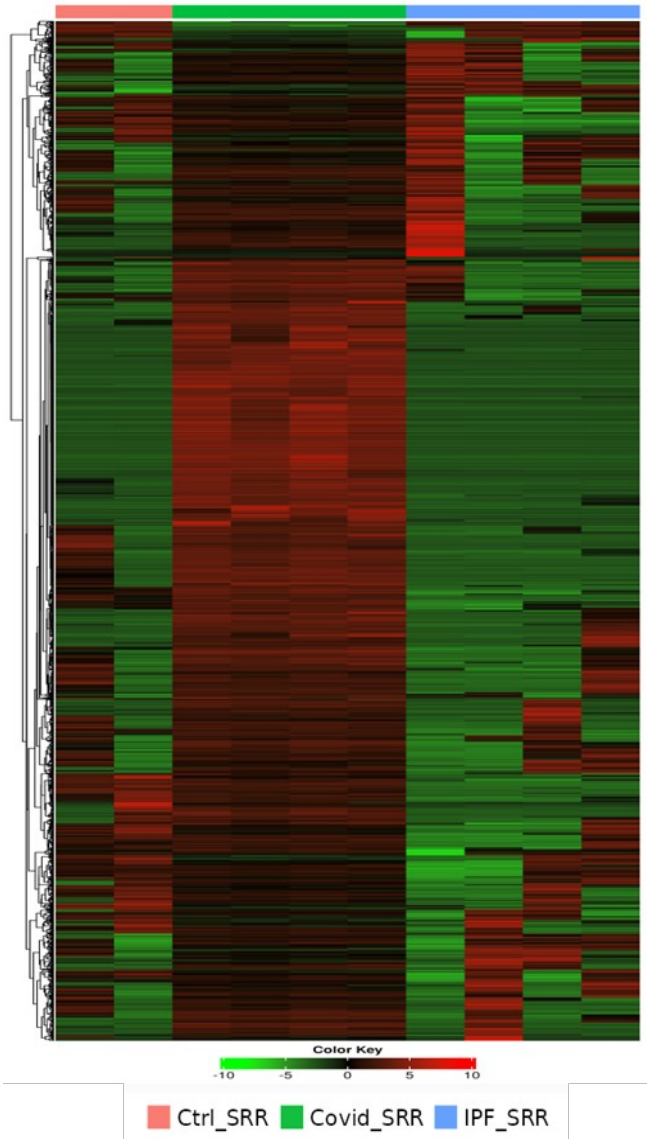

Figure 13. Hierarchical cluster for DEGs found in COVID-19, IPF, mad control groups

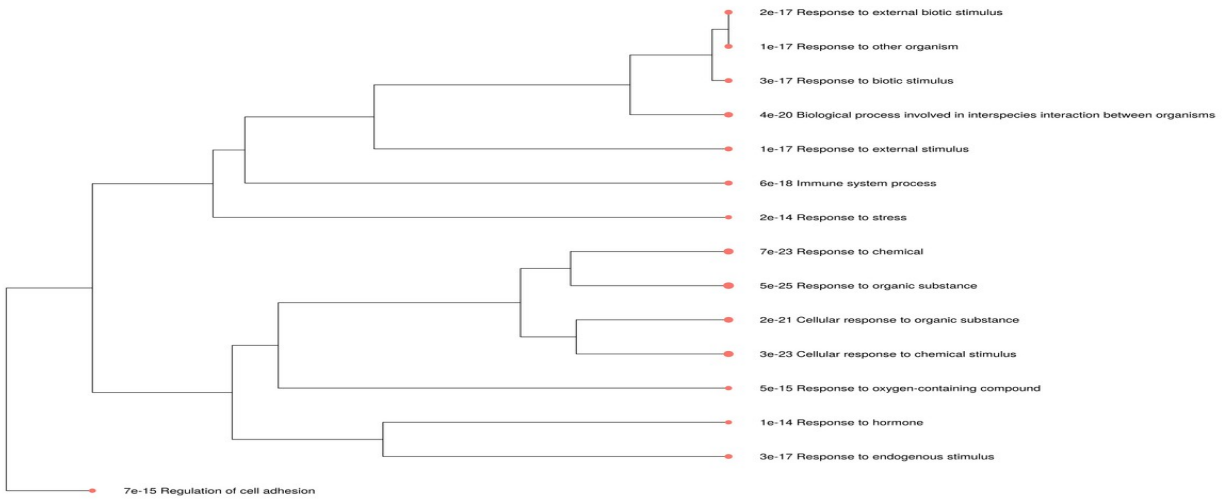

Figure 14. Top 15 Pathways of Hierarchal Clustering. Red circles at nodes depict FDRs. Smaller the FDR; the more significant the association is.

Table 1. Top 15 pathways in hierarchical clustering

P-adjusted	Fold Change	Pathway	cytoscape
5.47e-25	1,4	Response to organic substance	COL15A1, COL18A1, SULF1
2.89e-23	1,4	Cellular response to chemical stimulus	CF1, DCLK1, RTEL1
7.21e-23	1,3	Response to chemical	CFTR
2.48e-21	1,4	Cellular response to organic substance	COL15A1, COL18A1, CTHRC1, SULF1
6.17e-18	1,4	Immune system process	CFB, SFTPA2, SFTPC
1.09e-17	1,4	Response to external stimulus	CFB, COL18A1, CTHRC1, SULF1
1.67e-17	1,6	Response to external biotic stimulus	MUC5B
3.22e-17	1,5	Response to endogenous stimulus	COL15A1, MUC5B
7.36e-15	1,7	Regulation of cell adhesion	VCAN, COL12A1, COL3A1, COMP, LTBP1, MMP1
1.19e-14	1,7	Response to hormone	CFTR, CRABP2
1.51e-14	1,3	Response to stress	FHL1, RTEL1

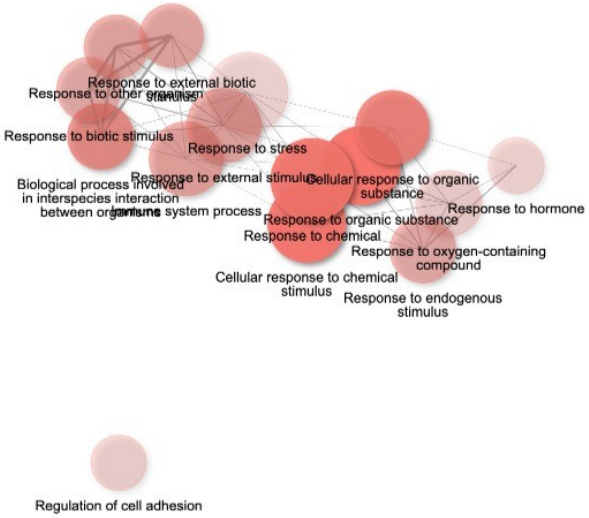

Figure 15. A network analysis of the top 15 pathways determined through hierarchical clustering.

Principal Component Analysis

Principal component analysis was performed to further understand the diversity of the data. The analysis revealed that DEGs in COVID-19 were grouped while the DEGs of IPF and control were placed in one quarter (Figure 16). This suggested that specific COVID-19 biomarkers could be used to distinguish it from pulmonary fibrosis. The Pearson linkage analysis further validated the potential of COVID-19 endothelial cell markers to distinguish COVID-19 from IPF (Figure 17).

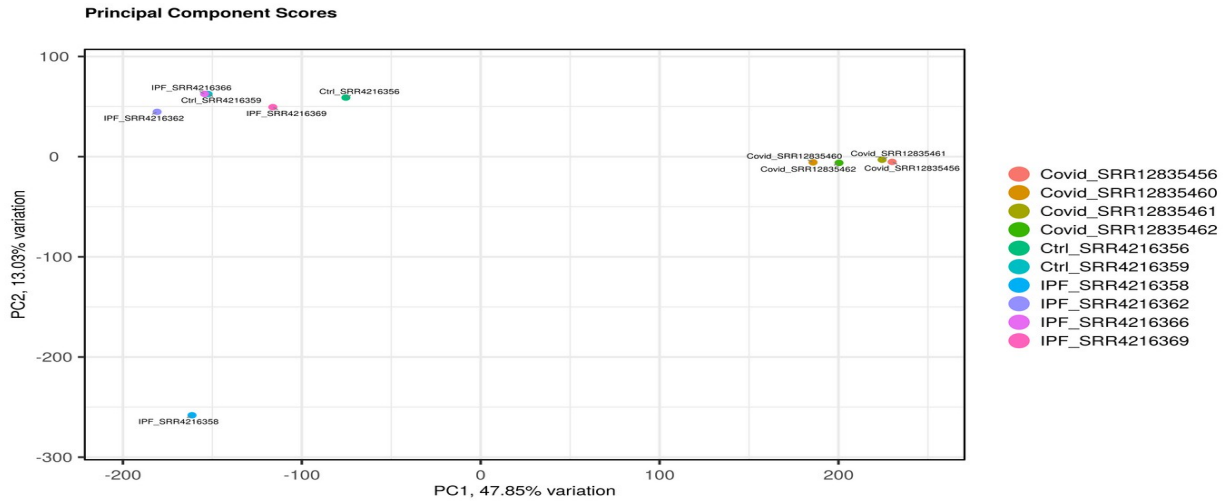

Figure 16. Principal component analysis of COVID, IPF, and control DEGs

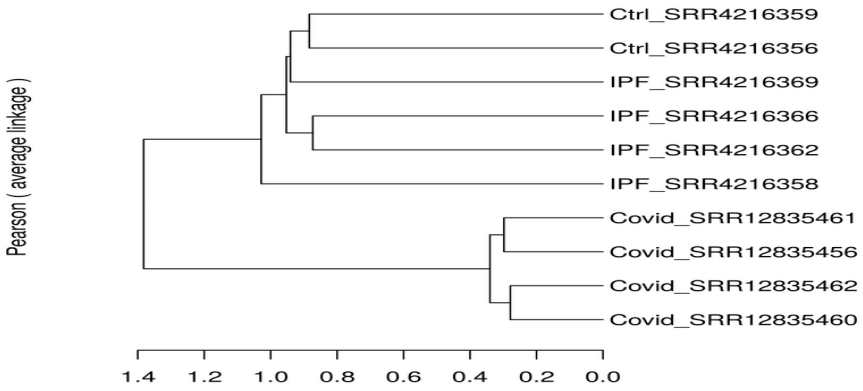

Figure 17. Pearson linkage depicting the association between study groups.

Discussion

Respiratory disorders, particularly COVID-19 and Idiopathic Pulmonary Fibrosis (IPF), are characterized by convoluted genetic components that play a role in their development. The SARS-CoV-2 virus is the causative agent of COVID-19, which mostly affects the respiratory system and may, in rare instances, result in severe lung injury (10,11). Idiopathic pulmonary fibrosis (IPF) is a chronic and irreversible lung disease that is characterized by the abnormal production of scar tissue inside the lungs (12). Recent research has demonstrated that the disorders share similar molecular pathways and gene expression patterns. In both COVID-19 and IPF, the expression of genes that are involved in immune response, fibrosis, inflammation, and tissue remodeling are dysregulated (13). The elucidation of genetic correlations between these diseases has substantial repercussions for the process of finding possible therapeutic targets and developing individualized treatment protocols.

Using RNA-Seq analysis carried out in RStudio, the objective of the current study was to carry out a comprehensive analysis of the gene expression profiles that were identified in specimens obtained from COVID-19 patients, individuals with Idiopathic Pulmonary Fibrosis (IPF), and healthy control participants. This was accomplished by conducting a comprehensive examination of the gene expression profiles. The RNA-Seq method is a cutting-edge sequencing technique that facilitates the measurement and description of the entire transcriptome (14,15). The expression levels of different genes can be understood through this tool which can help in identifying key molecular components that play role in diseases such as cancer or COVID-19 (15,16).

Normalization techniques such as Trimmed Mean of M-values (TMM) and Fragments Per Kilobase of Transcript per Million Mapped Reads (FPKM) were utilized to account for variations in sequencing depth and gene length. This allowed for accurate comparisons between the COVID-19 and IPF sample that were included in the study (17). The normalization of the data excluded noise and improved the accuracy of the results.

Advanced statistical techniques, including DESeq2, Limma Trend, and Limma Voom, available in Rstudio (18-20), were employed to identify genes that displayed significant differential expression between the COVID-19 and IPF groups, as well as the control group. To ensure the production of consistent and dependable outcomes, the methodologies utilized considered a range of variables. This encompassed various factors such as sample size, biological variability, and experimental design, among others. Our aim was to identify genes exhibiting distinct expression levels across the groups, in order to gain deeper insights into the molecular alterations associated with COVID-19 and IPF. To do this, we compared the expression levels of genes across all the diverse categories. A large number of DEGs were detected in the COVID-

19 vs control group. The expression fold change analysis also showed expression difference of DEGs between COVID-19 and IPF group. Allen et al. also performed the study to identify genetic overlaps in the diseases and found several genes that were expressed in both diseases such as MUC5B, ATP11A and DPP9 (13).

The research utilized data visualization techniques such as MA plots and volcano plots to facilitate data interpretation. The purpose of the graphical representations that have been provided in this article is to serve as a method of graphically depicting the degree of change that has occurred in the levels of gene expression, as well as the statistical significance of such changes. The use of this method makes it possible to identify genes whose expression levels have been significantly changed. We identified a set of genes whose expression was significantly up-regulated in only COVID-19 patients (COL3A1, COL14A1, COL15A1, COMP, CRABP2, CTHRC1, DCKL1, MUC5B, and RTEL1). We also identified genes SFPA2, and SFTPC whose expression was solely up-regulated in IPF patients. Future investigation may identify these genes as potential biomarkers or therapeutic targets.

Our goal in this study was to examine COVID-19 and IPF as a pair in their whole due to the complexity of their connection. Both of these diseases have a major influence on the respiratory system, and there is some evidence to suggest that their underlying pathogenic processes may be similar (21). There are currently few therapy options for IPF, a degenerative lung disease. Our mission is to investigate DEGs associated with IPF to discover novel treatment targets for this condition. This objective not only shows harmony with the larger scientific effort to create individualized treatment plans, but it also suggests the possibility of repurposing drugs for the treatment of IPF. This is because one of the goals of modern medicine is to develop individualized treatment plans for each patient. According to this paradigm, existing treatments for COVID-19 might be adapted into effective means of managing IPF (22). This study adds significantly to the growing body of scholarly literature that seeks to decipher the intricate genetic links between different diseases.

A detailed investigation of the gene expression patterns connected with COVID-19, IPF, and control samples was made possible by the usage of RNA-Seq analysis in RStudio. Because of how we approached the problem, we were able to draw certain conclusions about the gene expression patterns observed in these conditions. The results of this study not only enhance our understanding of the molecular processes at work in these diseases, but also have the potential to influence the development of individualized therapeutic approaches and diagnostic techniques. This research not only adds a fresh perspective to the existing body of knowledge, but it also breaks new ground by shedding light on potential therapy targets that are common to several disorders. The

discipline of personalized medicine has made significant progress because of this development. Future research may use these findings to validate and expand upon our previous findings, ultimately leading to the development of more effective methods for treating and managing patients with COVID-19 and IPF.

Conclusion

Genes that are differentially or commonly expressed between COVID-19, Idiopathic Pulmonary Fibrosis and control were identified, using datasets from the National Library of Medicine that utilized single cell RNA sequencing. The results obtained may be used to identify biomarkers specific to COVID-19 and IPF, to enable accurate diagnosis. The results may also drive the development of individualized therapeutic approaches for COVID-19 and IPF.

Acknowledgement

Thank you to the donors whose data was accessed and used in this study.

References

1. L. Chen, R. Tang, H. Chen, Y. Huang, S. Yi, Q. Bai, Y. Zhu, C. Hong, C. Liu, N. Zhang, T. Wang. Pulmonary Vasculature: A Target for COVID-19. *Am J Respir Crit Care Med*. **203**, 260-261 (2021). https://doi.org/10.1164/rccm.202007-2858LE

2. Z. Varga, A. J. Flammer, P. Steiger, M. Haberecker, R. Andermatt, A. S. Zinkernagel, M. R. Mehra, R. A. Schuepbach, F. Ruschitzka, H. Moch. Endothelial cell infection and endotheliitis in COVID-19. *Lancet*. **395**, 1417-1418 (2020). https://doi.org/10.1016/S0140-6736(20)30937-5

3. H. X. Bai, B. Hsieh, Z. Xiong, K. Halsey, J. W. Choi, T. M. L. Tran, I. Pan, L. B. Shi, D. C. Wang, J. Mei, X. L. Jiang, Q. H. Zeng, T. K. Egglin, P. F. Hu, S. Agarwal, F. F. Xie, S. Li, T. Healey, M. K. Atalay, W. H. Liao. Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT. *Radiology*. **296**, E46-e54 (2020). https://doi.org/10.1148/radiol.2020200823

4. A. V. Gaikwad, W. Lu, S. Dey, P. Bhattarai, C. Chia, J. Larby, G. Haug, S. Myers, J. Jaffar, G. Westall, G. K. Singhera, T.-L. Hackett, J. Markos, M. S. Eapen, S. S. Sohal. Vascular remodelling in idiopathic pulmonary fibrosis patients and its detrimental effect on lung physiology: potential role of endothelial-to-mesenchymal transition. **8**, 00571-2021 (2022). https://doi.org/10.1183/23120541.00571-2021

5. M. S. Eapen, A. V. Gaikwad, I. E. Thompson, W. Lu, S. Myers, P. Sharma, S. S. J. J. o. T. D. Sohal. The effectiveness of immunosuppressive cyclosporin in attenuating the progression of interstitial lung diseases. **11**, S1139 (2019). https://doi.org/10.1183/13993003.01139-2019

6. M. K. J. A. J. M. C. Glassberg. Overview of idiopathic pulmonary fibrosis, evidence-based guidelines, and recent developments in the treatment landscape. **25**, S195-S203 (2019). https://doi.org/10.1080/17476348.2019.1672397

7. S. F. Naqvi, D. A. Lakhani, A. H. Sohail, J. Maurer, S. Sofka, A. Sarwari, Y. B. Hadi. Patients with idiopathic pulmonary fibrosis have poor clinical outcomes with COVID-19 disease: a propensity matched multicentre research network analysis. **8**, e000969 (2021). https://doi.org/10.1136/bmjresp-2021-000969

8. S. S. J. N. R. N. Potter. Single-cell RNA sequencing for the study of development, physiology and disease. **14**, 479-492 (2018). https://doi.org/10.1038/s41576-018-0023-0

9. S. X. Ge, E. W. Son, R. Yao. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinform*. **19**, 1-24 (2018). https://doi.org/10.1186/s12859-018-2486-6

10. L. Yang, S. Liu, J. Liu, Z. Zhang, X. Wan, B. Huang, Y. Chen, Y. Zhang. COVID-19: immunopathogenesis and Immunotherapeutics. *Signal transduction and targeted therapy*. **5**, 128 (2020). https://doi.org/10.1038/s41392-020-00243-2

11. K. Yuki, M. Fujiogi, S. Koutsogiannaki. COVID-19 pathophysiology: A review. *Clin Immunol*. **215**, 108427 (2020). https://doi.org/10.1016/j.clim.2020.108427

12. B. Ballester, J. Milara, J. Cortijo. Idiopathic pulmonary fibrosis and lung cancer: mechanisms and molecular targets. *International journal of molecular sciences*. **20**, 593 (2019). https://doi.org/10.3390/ijms20030593

13. R. J. Allen, B. Guillen-Guio, E. Croot, L. M. Kraven, S. Moss, I. Stewart, R. G. Jenkins, L. V. Wain. Genetic overlap between idiopathic pulmonary fibrosis and COVID-19. *Eur Respir J*. **60**, (2022). https://doi.org/10.1183/13993003.03144-2021

14. M. Z. Hasan, S. Islam, K. Matsumoto, T. Kawai. Meta-analysis of single-cell RNA-seq data reveals phenotypic switching of immune cells in severe COVID-19 patients. *Comput Biol Med*. **137**, 104792 (2021). https://doi.org/10.1016/j.compbiomed.2021.104792

15. N. Iqbal, P. Kumar. Integrated COVID-19 Predictor: Differential expression analysis to reveal potential biomarkers and prediction of coronavirus using RNA-Seq profile data. *Comput Biol Med*. **147**, 105684 (2022). https://doi.org/10.1016/j.compbiomed.2022.105684

16. J. Wang, D. C. Dean, F. J. Hornicek, H. Shi, Z. Duan. RNA sequencing (RNA-Seq) and its application in ovarian cancer. *Gynecologic oncology*. **152**, 194-201 (2019). https://doi.org/10.1016/j.ygyno.2018.11.017

17. J. Simoneau, S. Dumontier, R. Gosselin, M. S. Scott. Current RNA-seq methodology reporting limits reproducibility. *Brief Bioinformatics*. **22**, 140-145 (2021). https://doi.org/10.1093/bib/bbz158

18. C. R. Yendrek, E. A. Ainsworth, J. Thimmapuram. The bench scientist's guide to statistical analysis of RNA-Seq data. *BMC research notes*. **5**, 1-10 (2012). https://doi.org/10.1186/1756-0500-5-506

19. S. H. Yip, P. Wang, J.-P. A. Kocher, P. C. Sham, J. Wang. Linnorm: improved statistical analysis for single cell RNA-seq expression data. *Nucleic Acids Res*. **45**, e179-e179 (2017). https://doi.org/10.1093/nar/gkx828

20. D. Fanidis, P. Moulos. Integrative, normalization-insusceptible statistical analysis of RNA-Seq data, with improved differential expression and unbiased downstream functional analysis. *Brief Bioinformatics*. **22**, bbaa156 (2021). https://doi.org/10.1093/bib/bbaa156

21. J. Fadista, L. M. Kraven, J. Karjalainen, S. J. Andrews, F. Geller, J. K. Baillie, L. V. Wain, R. G. Jenkins, B. Feenstra. Shared genetic etiology between idiopathic pulmonary fibrosis and COVID-19 severity. *EBioMedicine*. **65**, (2021). https://doi.org/10.1016/j.ebiom.2021.103282

22. R. Kircheis, M. Schuster, O. Planz. COVID-19: mechanistic model of the African paradox supports the central role of the NF-κB pathway. *Viruses*. **13**, 1887 (2021). https://doi.org/10.3390/v13101887