



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



# Identification of novel regulatory pathways across normal human bronchial epithelial cell lines (NHBEs) and peripheral blood mononuclear cell lines (PBMCs) in COVID-19 patients using transcriptome analysis

Lavanya C, Aajnaa Upadhyaya, Arpita Guha Neogi, Vidya Niranjana \*

Department of Biotechnology, R. V. College of Engineering, Bengaluru, India

## ARTICLE INFO

### Keywords:

Gene regulatory networks  
Neurodegenerative disorders  
Differential gene expression  
Pathway analysis  
Drug repurposing

## ABSTRACT

The SARS-CoV-2 is one of the most infectious and deadly coronaviruses, which has gripped the world, causing the COVID-19 pandemic. Despite the numerous studies being conducted on this virus, many uncertainties are with the disease. This is exacerbated by the speedy mutations acquired by the viral strain, which enables the disease to present itself differently in different people, introducing new factors of uncertainty. This study aims at the identification of regulatory pathways across two cell lines, namely, the peripheral blood mononuclear cell line (PBMC) and the normal human bronchial epithelial (NHBE) cell line. Both the above-mentioned cell lines were considered because they support viral replication. Furthermore, the NHBE cell line captures vital changes in the lungs, which are the main organs affected by the COVID-19 patients, and the PBMC cell line is closely linked to the body's immune system. RNA-Seq analysis, differential gene expression and gene set enrichment analysis for pathway identification were followed. Pathway analysis throws light upon the various systems affected in the body due to the COVID-19. Gene regulatory networks associated with the significant pathways were also designed. These networks aid in identifying various gene targets, along with their interactions. Studying the functionality of the pathways and the gene interactions associated with them, aided by long COVID studies, will provide immense clarity about the current COVID-19 scenario. In the long term, this will help in the design of therapeutic approaches against the SARS-CoV-2 and can also contribute to drug repurposing studies. Ultimately, this study identifies and analyses the relationship of various undiscovered or lesser explored pathways in the human body to the SARS-CoV-2 and establish a clearer picture of the association to help streamline further studies and approaches.

## 1. Introduction

Coronaviruses, known for their spiked structure, belong to the *Coronaviridae* family, which infects the upper gastrointestinal and respiratory tract in mammals [1]. SARS-CoV-2 outbreaks in 2002 and 2003, as well as MERS-CoV, which has been causing an epidemic in some parts of Africa since 2012, have aided researchers in their understanding of coronaviruses. COVID-19, which started in late December 2019, was declared a pandemic by the World Health Organization on March 11, 2020 [2]. Although there is no effective cure for the infection, which affects general health, vaccines have shown promising results. Among the unavailability of treatment regimens, an overwhelming amount of research has been carried out on drug discovery. These include, but are not limited to, understanding the role of repurposed drugs like

Hydroxychloroquine, Favipiravir, remdesivir [3] along with carbon fullerene and nanotubes as potential binding agents against the protein targets [4]. Apart from synthetic compounds, Phyto-actives are under intense investigation for its anti-viral activity [5]. The potential targets also inculcated signalling pathways like NF- $\kappa$ B (inflammatory and apoptosis related) pathways, which are targeted during drug discovery [6]. Researchers are still trying to understand the virus and its potential mutations to find a permanent solution to this global concern. However, the key to understanding the secrets of this virus lies in its genetic material. Coronaviruses have the largest genome of approximately 26–32 kB among the RNA viruses. SARS-CoV-2 has a positive-sense single-stranded RNA as its genetic material [7]. Transcriptome studies of other types of coronaviruses have revealed much information about the kind of infection and the pathways that affect the body, and they have also exposed novel approaches. However, it is uncertain if SARS-CoV-2 has

\* Corresponding author. Department of Biotechnology, RV College of Engineering, RV Vidyaniketan Post, Mysore Road, Bengaluru, 560059, India.  
E-mail address: [vidya.n@rvce.edu.in](mailto:vidya.n@rvce.edu.in) (V. Niranjana).

ABBREVIATIONS	
SARS CoV-2	Severe Acute Respiratory Syndrome Coronavirus-2
MERS CoV	Middle East Respiratory Syndrome Coronavirus
COVID-19	Coronavirus Disease 2019
PBMC	Peripheral Blood Mononuclear Cells
NHBE	- Normal Human Bronchial Epithelial Cells
ACE2	Angiotensin Converting Enzyme 2
NCBI	National Center for Biotechnology Information
ENA	European Nucleotide Archive
GRCh	Genome Reference Consortium Human genome
GFF	General Feature Format
GTF	General Transfer Format
FNA	- FASTA Nucleic Acid
SRA	Sequence Read Archive
HISAT	Hierarchical Indexing for Spliced Alignment of Transcripts
SAM	Sequence Alignment Map
BAM	Binary Alignment Map
DGE	- Differential Gene Expression
KEGG	Kyoto Encyclopedia of Genes and Genomes
KOBAS	KEGG Orthology-Based Annotation System
GSEA	Gene Set Enrichment Analysis
ENCODE	ENCyclopedia Of DNA Elements
GO	Gene Ontology
NDD	Neurodegenerative disease

the same transcriptome properties as SARS-CoV – 1 and whether the prior techniques would be effective against this virus. As a result, transcriptomics is being used in this study to unveil previously undisclosed information regarding COVID-19 and SARS-CoV – 2.

Transcriptomics is an NGS-based approach that helps analyse the region of the actively transcribed genome. RNA-Seq involves the sequencing of the entire RNA population without any annotation to the genome. More reads are generated if the sequence is transcriptionally active, which plays a significant role in the analysis [8]. RNA-Seq analysis yields valuable information about the differentially expressed genes, which helps in discovering the novel pathways associated with these genes.

To characterize the genes involved and analyse the affected pathways, transcriptomics of cell lines derived from lungs and blood were considered for this study, and were later subjected to gene enrichment analysis. Normal human bronchial epithelial cell lines were considered for the study, as the virus directly affects the lungs of the infected person. Peripheral blood mononuclear cells (PBMCs) were considered for the following reasons: (i) viral replication in these cell lines, (ii) dynamic changes in the disease during the infection could be detected in these cell lines [9], and (iii) these cell lines are directly related to the immune system.

This study draws attention to the pathways at a more intricate level to obtain valuable information about the virus and COVID-19 using two different cell lines. Since not much work has been done concerning these two cell lines, this study can provide potential insights into the virus and

viral infection.

## 2. Materials and methodology

Fig. 1 represents the overall methodology that is followed for this study.

### 2.1. Dataset collection

From the NHBE and PBMC cell lines, three healthy and patient samples were obtained. Run Selector is a tool available through the Sequence Read Archive (SRA), which fine-tunes web-based search results based on the two dozen fields to filter SRA data in Run Selector. In this study, the Run Selector was narrowed down to search transcriptome data, and the RNA-Seq assay type was selected. Clicking on the hyperlinks inside the data table will lead to the NCBI page for the SRA Study or BioProject. The healthy and patient sample IDs from both the cell lines were collected.

The or file transfer protocol (FTP) links to fetch these samples were taken from European Nucleotide Archive (ENA). The human reference genome data were available from the most recent version, GRCh38, released in 2013, also called hg38. GRCh38 V.32 was used by downloading all the GTF and GFF files from Gencode.

NHBE samples were taken from an open-access project. In the mentioned project, biological replicates were created using uninfected human lung biopsies derived from one female aged 60 and one male

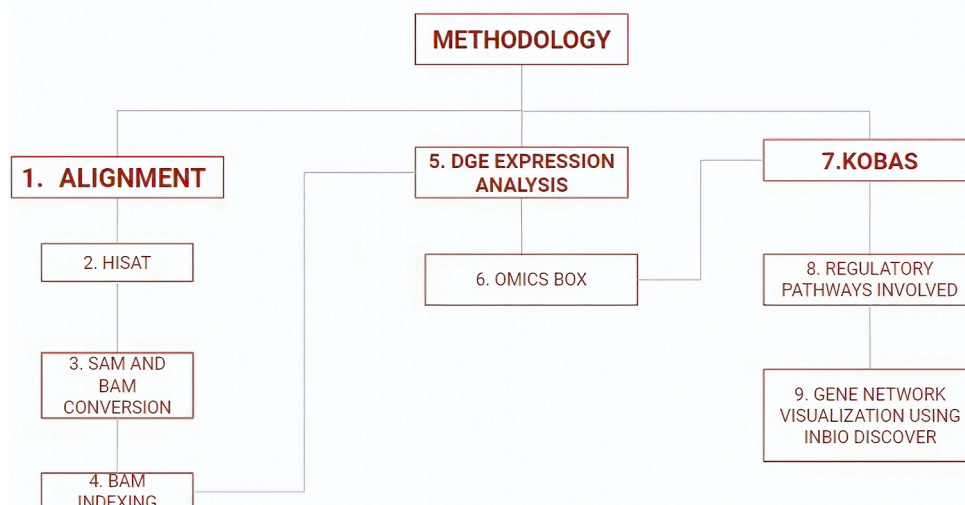


Fig. 1. Pipeline methodology flowchart.

aged 72. Technical replicates were processed by deriving lung samples from a single deceased male, a COVID-19 patient aged 74. The biosamples were accessed from the BioProject ID PRJNA615032 (<https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA615032>) [10,11].

The samples for PBMCs were collected from a study that consisted of 6 convalescent patients, 10 RTP patients, and 10 healthy controls for the analysis of the immunological characteristics of PBMCs. Transcriptome sequencing performed a comprehensive characterization of the transcriptional changes in the three groups. The biosamples were accessed from the BioProject ID SRP304889 (<https://www.ncbi.nlm.nih.gov/sra/?term=SRP304889>).

## 2.2. Indexing

Indexing the sequences help differentiate between multiple sequences by using various techniques and algorithms. These techniques also help reduce the computer memory acquired by the human genome, thus enabling a faster and efficient indexing of the sequence. This characteristic allows the aligner to narrow down the query sequence of the genome, which in turn saves time and memory [8]. In this study, the human reference genome was indexed before the sample reads were mapped and aligned for better results. Once the human reference was indexed, the NHBE and PBMC sample sets were mapped and aligned with it.

## 2.3. Mapping and alignment

After the human reference genome is indexed, the sample sequences must be mapped or aligned to the human reference genome or a de novo assembly must be conducted. This is one of the primary steps in transcriptome analysis. Mapping is done to locate the origins of the reads in the human reference genome [8] and to locate and identify the distances between the genes in the chromosome [12]. Therefore, the three sample sets of both cell lines in this work, including the healthy and patient conditions, were mapped and aligned to the human reference genome collected using the HISAT2 tool.

The HISAT2 tool is a mapping and alignment tool that yields higher accuracy compared to other alignment tools. It uses several fast algorithms like the Burrows–Wheeler Aligner and Bowtie [12].

The mapping results from the HISAT2 tool are stored in the Sequence Alignment Map (SAM) format (output file has a .sam extension). SAM is a tab-delimited format that is readable, slow to parse and easily examinable. However, it must be converted to Binary Alignment Map (BAM) format to reduce the size of the samples and parse the sequences faster [8]. BAM is the compressed, binary version of SAM. BAM format compresses up to 128 Mb. The SAM/BAM format contains a header and an alignment section with 11 mandatory fields, with each line starting with the symbol '@' [8].

Samtools were used to convert SAM into BAM. These tools were downloaded and installed using anaconda. Samtools looks for matches and mismatches at each genome coordinate of the reads and removes duplicates while compressing the files and changing the file format from the .sam to the .bam format [8].

The bam files obtained are indexed to obtain a companion file, called an index file or the bam index file. This file has the same name, suffixed with .bai. This allows programs to skip through some areas of the sequences and jump directly to specific parts of the bam file. The bai file is valid only when there is a corresponding bam file. For each alignment file in this study, the .bam.bai was kept in the same directory as the bam files of the sample.

## 2.4. Differential gene expression

RNA-Seq provides count data for quantitative readouts, statistical analysis and visualization with the aid of tools such as DESeq or R/Bioconductor packages [13–15]. In transcriptome analysis, different

conditions are compared using differential gene expression (DGE) analysis [16]. Here, the samples were subjected to pairwise DGE that used the count table generated from RNA-Seq for analysis using a tool called OmicsBox.

OmicsBox is a bioinformatics software solution that has various modules to run with different types of NGS-based approaches [17]. However, the transcriptomics module was used in this study, and under this module, the DGE analysis tool was selected. The results obtained from the alignment in the form of BAM files and the human reference genome in the gff format were directly uploaded to OmicsBox. The genes between the experimental and contrast conditions were run over time with well-known and versatile statistical packages like NOISeq, edgeR and maSig pro [18–21].

Differential expression analysis identifies which genes are expressed under specified conditions and the extent to which they are expressed. These genes offer biological insight into the processes affected by the conditions of interest. The count data used for differential expression analysis represent the number of sequence reads originating from a particular gene. A higher number of counts imply more reads associated with a particular gene and a strong assumption that there is a higher expression level of that gene in the sample. In this work, the count data were first normalized to account for differences between the library sizes and RNA composition between samples. Following this, the normalized counts were used to make some plots for QC at the gene and sample levels. Finally, differential gene expression analysis was performed. The gene count of counts per million (CPM) was the filter, which compared replicates between the same sample group.

## 2.5. Gene set enrichment and pathway analysis

The differentially expressed genes obtained after DGE from Omics-Box were subjected to gene set enrichment (GSE). This analysis is a popular agenda to interpret and analyse the.

Information obtained from the DGE was used for pathway summary. This approach is flexible, robust, dramatically reduces background noise and can be used for highly heterogeneous datasets. It also helps in detecting pathway activity changes based on the differentially expressed genes, and the employment of GSE methods in this pipeline has helped model the associated pathways [22].

Kobas 3.0 tool was used for pathway analysis, and EnrichR tool was used to optimize the pathways obtained. Visualization and networking of pathways were done using InBio Map™. KOBAS version 3.0, named KOBAS intelligent version (KOBAS-i), is a web server and software that annotates an input set of genes or proteins by mapping to genes with known pathways in the KEGG PATHWAY database. It uses a hypergeometric test to identify significant pathways. Moreover, it uses five different pathway databases, namely KEGG PATHWAY, PID, BioCyc, Reactome and Panther, and five human disease databases, including OMIM, KEGG DISEASE, FunDO, GAD and NHGRI GWAS Catalog. KOBAS-i accepts different types of inputs, like gene IDs, symbols, FASTA sequence, or tabular BLAST output [23,24].

The two programs in KOBAS 3.0 are 'annotate' and 'identify'. The annotate program identifies the coding regions and locations of the genes. The enrichment module provides information about the pathways and GO terms that are statistically significantly associated with the input gene list or expression [23]. Two different enrichment analyses are available: gene-list enrichment and exp-data enrichment. For this study, the KEGG pathway database was chosen as the filter to map the input genes to the KEGG pathway database. The upregulated and down-regulated gene lists were uploaded separately after selecting the Gene-set enrichment option. The two lists were uploaded separately due to the input limit of 3000 gene symbols set by KOBAS-i. This tool outputs statistically significant pathways associated with the statistically significant genes obtained after gene set enrichment analysis in the form of bar and bubble plots. KOBAS 3.0 is freely available at: <http://kobas.cbi.pku.edu.cn/>.

The Enrichr tool was used to optimize the pathways. Enrichr currently contains an extensive and diverse collection of 102 gene set libraries available for analysis [25]. Features such as submission of fuzzy sets, uploading of BED (Browser Extensible Data) files and visualization in the form of clustergrams are incorporated. The clustergram generates correlation plots of the most significant pathways. The upregulated and downregulated cell lines from both the cell lines were uploaded to Enrichr to output a bar plot representing the most significant pathways obtained from different pathway databases. Enrichr is freely available at <http://amp.pharm.mssm.edu/Enrichr> [26–28].

For networking, interpretation and visualising the pathways obtained from KOBAS 3.0 and Enrichr, another tool called InBio Map™ was used. With this tool, relevant annotations from the enrichment analysis results were selected. Custom network visualizations were created using annotations, and gene expression datasets about this study were given as input. Networks were filtered based on topology or proteins. InBio Map™ is freely available at <https://inbio-discover.com/> [29].

Further details and the codes used are available in the published protocol [30].

### 3. Results and discussion

#### 3.1. Results obtained from omics box

##### 3.1.1. Count table generation

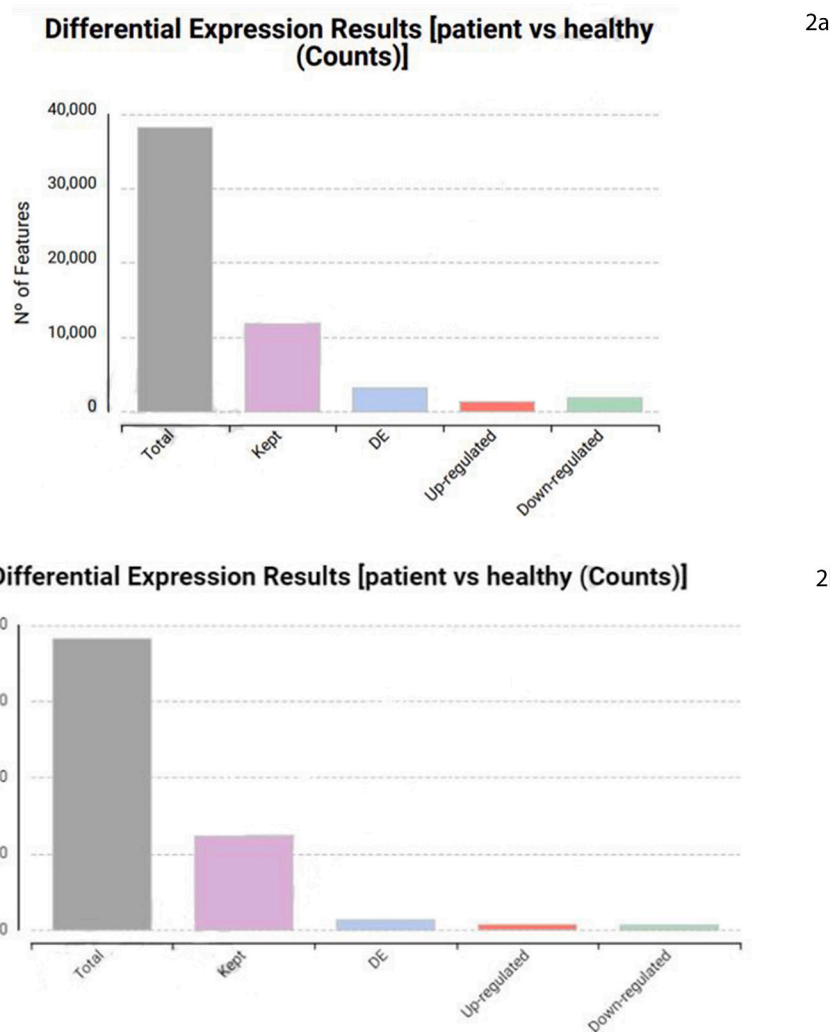
In the count tables, 38,199 genes are acquired from both cell lines. The gene expression of RNA-seq studies is approximated using the count table. The number of reads that overlap a certain feature, such as a gene, is referred to as a count. [Supplementary Tables S2A–S2B](#) provides count tables for both cell lines.

In Supplementary figures, S1A–S1D, bar charts depicting counts per category and box plots depicting count distributions are presented.

##### 3.1.2. Differential gene expression

Tables containing differentially expressed genes obtained from the two cell lines were generated and are provided in [Supplementary Tables S2C and S2D](#). These tables show whether a gene is upregulated ( $FDR \leq 0.05$ ,  $\log_{2}FC \geq 1$ ) or downregulated ( $FDR \leq 0.05$ ,  $\log_{2}FC$

$< -1$ ). Genes that have not passed the filtering step are not shown here. The  $\log_{2}FC$  values describe the gene expression changes between different experimental conditions (healthy and patient). The  $\log_{2}$  CPM tag averages the  $\log_{2}$ -counts-per-millions. The False discovery rate (FDR) is generally estimated by the Benjamini–Hochberg method. Typically, the p-values indicate how significant the results are. A p-value



**Fig. 2.** Fig. 2A: DGE results from the NHBE cell line. Fig. 2B: DGE results from the PBMC cell line. The first rectangle represents the total number of features (genes) obtained from the count table. The next block represents the number of genes that were obtained after applying the normalisation filter. The third block represents the genes that did not show any expression pattern. The last two blocks show the number of upregulated and downregulated genes, respectively.

less than 0.05 (typically  $\leq 0.05$ ) is statistically significant. FDR is an adjusted p-value to trim false-positive results. Since there are hundreds and thousands of genes, false positives may enter by chance. Therefore, FDR offers more confidence than just a p-value. FDR values less than 0.05 are considered real. The likelihood ratio (LR) statistic for the generalized linear model (LR test) evaluates the integrity of two competing statistical models based on the ratio of their commonness or likelihood. The results from the differential gene expression analysis of both cell lines are summarized in the bar plots shown in Fig. 2. The heatmap showing the regulation of the top 50 differentially expressed genes is represented in Fig. 3.

The volcano plots shown in Fig. 4 represent the regulation of genes based on their FDR and logFC values. The FDR values are less than 0.05 for both upregulated and downregulated genes, and the logFC values are higher than 1 for the upregulated genes and lower than -1 for the downregulated genes. Genes that do not show any expression pattern are denoted by black. In Figs. 3 and 4, the upregulated genes are denoted by red, and the downregulated genes are denoted by green. The MA and MDS plots representing the differentially expressed genes of both cell lines are provided in the Supplementary Figs. S1E-S1H.

### 3.2. Results from pathway analysis

#### 3.2.1. KOBAS 3.0 pathway analysis

In KOBAS 3.0, the gene lists from both samples were uploaded separately, and the above-mentioned plots containing the significant pathways were obtained. This tool allows the user to visualize the output in different ways, as follows.

Enriched pathways visualized in cirFunMap: Each node represents an enriched pathway, and the node colour portrays different clusters. The node size corresponds to six levels of enriched p-values. Furthermore, the node size ranges from small to large: [0.05,1], [0.01,0.05], [0.001,0.01], [0.0001,0.001], [1e-10,0.0001] and [0,1e-10].

Enriched pathways visualized in barplot: Each row represents an enriched pathway, and the length of the bar represents the enrich ratio, which is given by the ratio of the number of input genes to the background gene number. Here, the genes present in the whole human genome are considered the background genes; however, the users can input their own set of background genes. The colour of the bars

corresponds to the different clusters to which they belong. If there are more than five pathways for each cluster, the top five pathways with the highest enrich ratio will be displayed. Hence, only the most significant pathways are displayed in the plot. Fig. 5 represents the bar plots derived from the upregulated and downregulated genes of the NHBE cell line. Fig. 6 represents the bar plots derived from the upregulated and downregulated genes of the PBMC cell line.

Bubble plots visualising the enriched pathways were also obtained for both cell lines and are provided in the Supplementary Fig. S1I-S1L.

#### 3.2.2. Enrichr

The gene lists from the two cell lines were uploaded to Enrichr, which then generated significant pathways from different pathway databases. All the pathways generated by Enrichr and the significant genes associated with them are provided in the supplementary tables. The corresponding p-values and corrected p-values are also mentioned. A corrected p-value provides more confidence in the results obtained.

The bar chart represents the top 10 enriched pathways and their p-values. The blue coloured bars indicate that the pathways have significant p-values ( $< 0.05$ ). An asterisk (\*) next to the p-value indicates that the pathway also has a significant adjusted p-value ( $< 0.05$ ).

One of the most significant pathways represented in Fig. 7A is the coronavirus disease pathway. The figure implies that most of the significant genes correlate with this disease, which also implies that the other associated pathways must be closely observed to draw conclusive results about the coronavirus infection.

Fig. 7B shows the bar plot obtained from the gene set enrichment of the PBMC cell line, which implies that the significant genes obtained from the upregulated and downregulated gene lists of the PBMC cell line correlate with the immunological pathways, which is one of the most heavily affected systems of the human body in COVID-19.

## 4. Discussions

Based on the results obtained, all the significant pathways can be broadly classified into glucose metabolism pathways, immunological pathways, pathways of neurodegeneration, cellular physiology, and signalling pathways. Gene networks for some categories mentioned above were drawn by loading the significant genes from the different

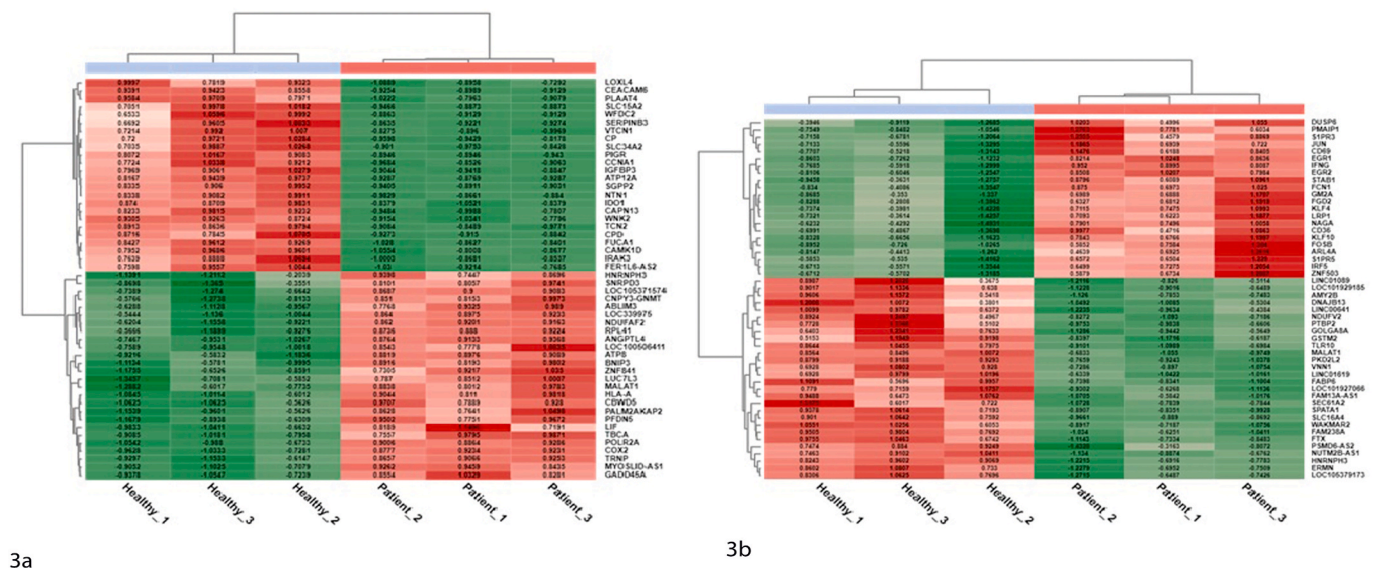
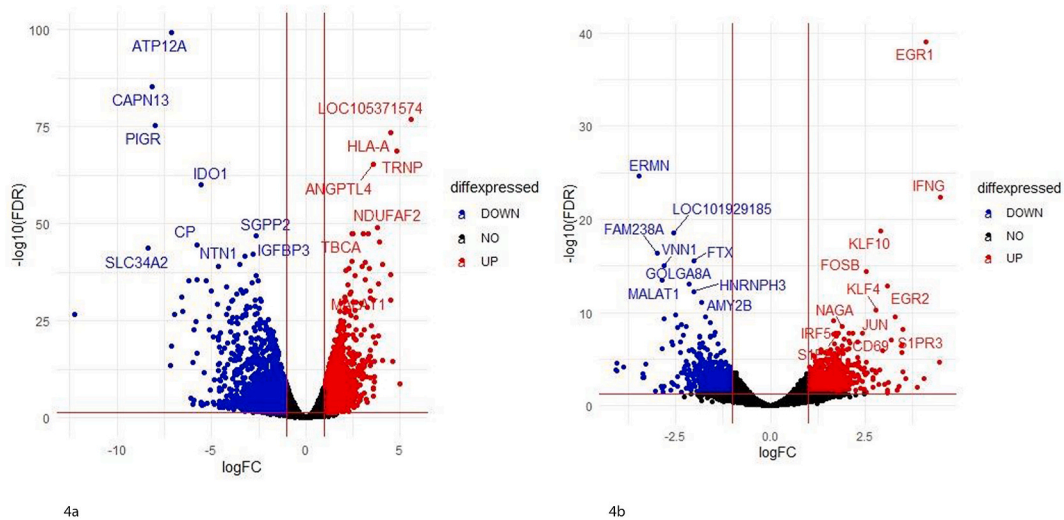
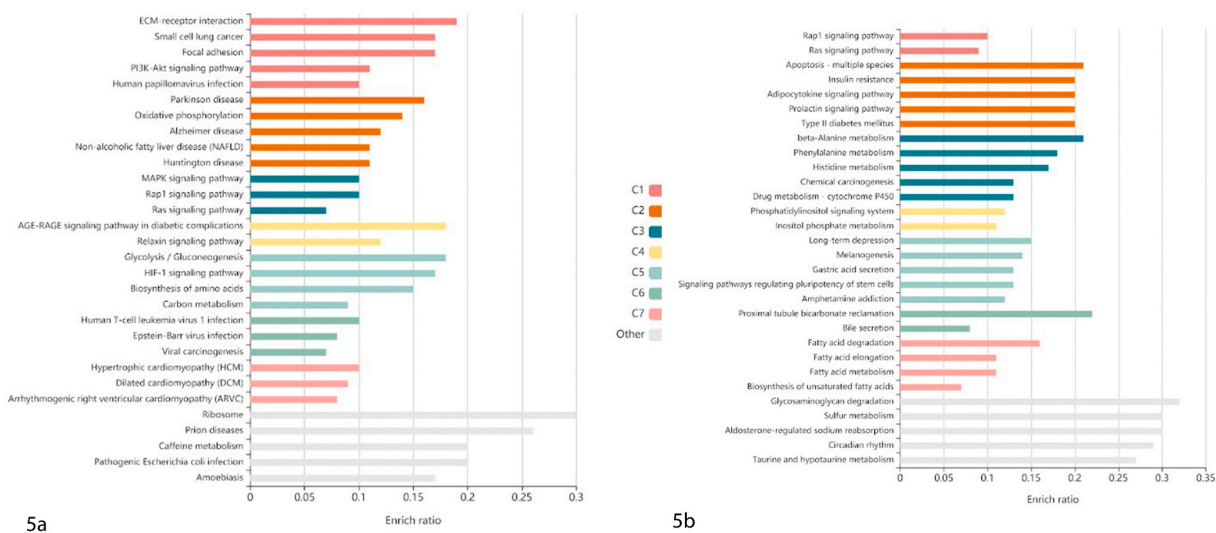


Fig. 3. Fig. 3A represents the heatmap for the differentially expressed genes of the NHBE cell line. The first half of the genes (till FER1L6-CS2) are upregulated in the healthy samples and downregulated in the patient samples. The second half of the genes (from HNRNP3) are downregulated in the healthy samples and upregulated in the patient samples. Fig. 3B represents the heatmap for the differentially expressed genes from the PBMC cell line. The first half of the genes (till ZNF503) are upregulated in the healthy samples and downregulated in the patient samples. The second half of the genes (from LINC01089) are downregulated in the healthy samples and upregulated in the patient samples.



**Fig. 4.** Fig. 4A represents the volcano plot visualising the differentially expressed genes from the NHBE cell line and Fig. 4B represents the volcano plot visualising the differentially expressed genes from the PBMC cell line.



**Fig. 5.** Fig. 5A shows the bar plots visualising the significant pathways from the upregulated genes of the NHBE cell line. Fig. 5B shows the bar plots visualising the significant pathways from the downregulated genes of the NHBE cell line.

significant pathways obtained, to InBio Discover™.

#### 4.1. Association of glucose metabolism pathways with COVID-19

According to the American Diabetes Association, people with diabetes are more likely to get COVID-19. They are also more prone to severe complications after contracting this disease [31]. Diabetes is one of the major comorbidities correlated with severe COVID-19. The replication rate of the causative virus increases with the increase in glucose levels in the host. Such an increase in the replication rate is followed by a cytokine storm and ACE-2 upregulation [32].

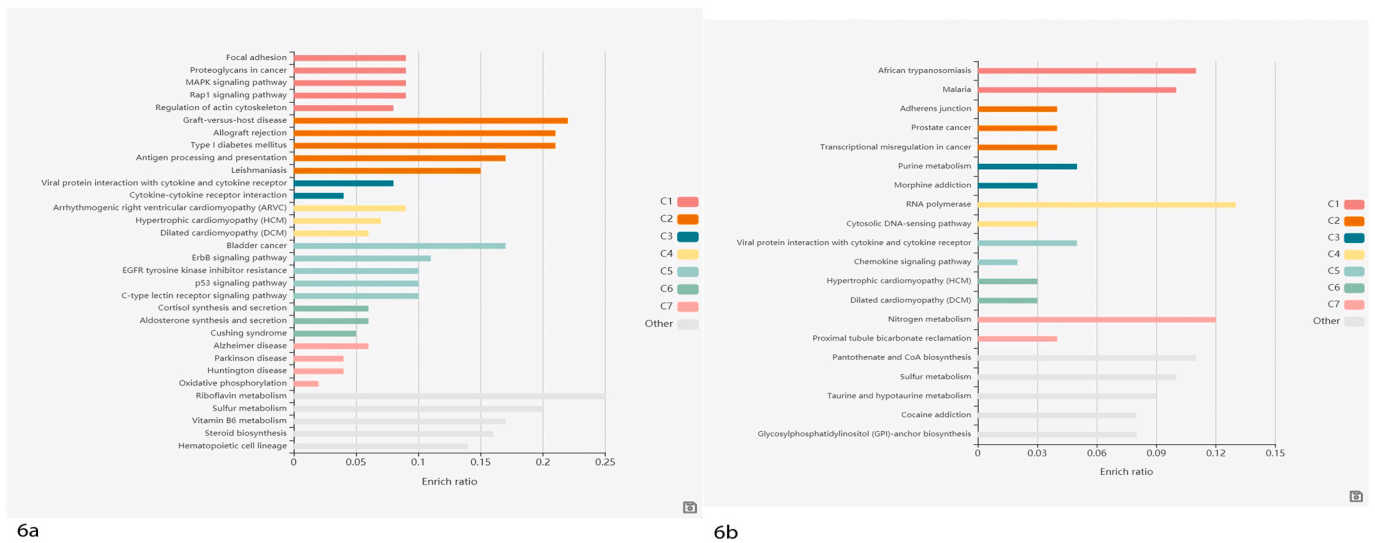
Several studies have attempted to model the important role of the pathways associated with glucose metabolism in increased SARS-CoV-2 replication, which in turn elicits an immune response in the host.

The results showed that in diabetes complications, the pathways classified under the glucose metabolism pathways are glycolysis/gluconeogenesis, insulin resistance, insulin signalling, type 1 diabetes mellitus, and AGE-RAGE signalling pathways. The insulin resistance and signalling pathways were downregulated, whereas the other pathways

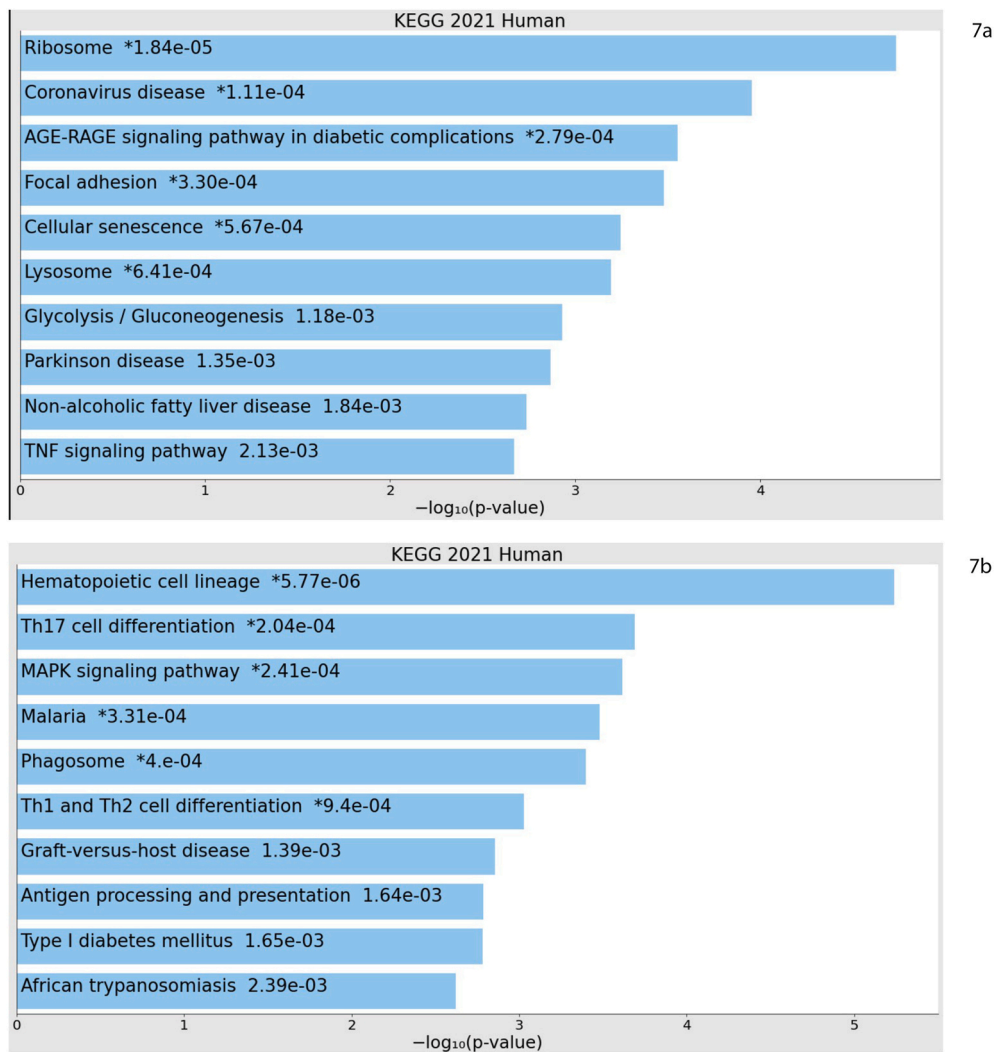
were upregulated. Targeting those pathways related to glucose metabolism will eventually shed light on the changes in all systems in a body under viral attack. This will help design new anti-viral approaches that target the pathogen. A gene network consisting of the significant genes belonging to the above-mentioned pathways was constructed to examine the interaction between the families of genes, as shown in Fig. 8.

#### 4.2. Association of neurodegenerative pathways with COVID-19

The hospital admission data indicated that the mortality rate is higher in COVID-19 patients who have dementia and other neurodegenerative diseases, than in patients who do not suffer from any neurodegenerative comorbidities. To date, no particular cause for the above phenomena has been noted. However, SARS-CoV-2 infection possibly damages the nasal epithelium, which is connected to the central nervous system. It was further observed that, due above, neurodegenerative diseases may occur in the future [33]. During the 15th International Virtual Conference on Alzheimer’s Disease (AD) and Parkinson’s



**Fig. 6.** Fig. 6A shows the bar plots visualising the significant pathways from the upregulated genes of the PBMC cell line. Fig. 6B shows the bar plots visualising the significant pathways from the downregulated genes of the PBMC cell line.



**Fig. 7.** The above bar plots show the most significant pathways obtained from the KEGG database for the NHBE and PBMC cell lines.



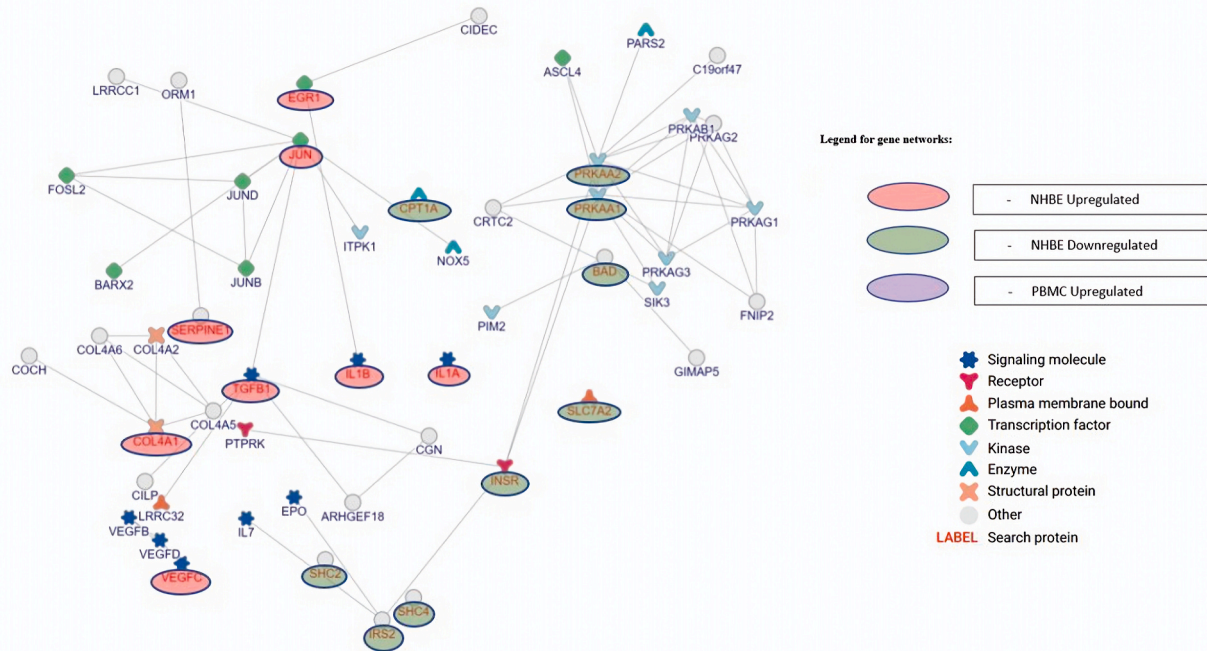


Fig. 8. represents the gene network of the glucose metabolism pathways derived from the significantly regulated genes of the NHBE cell line.

Disease (PD) 2021, the clinical and social implications of COVID-19 on neurodegenerative disorders were evaluated. A group of physicians and researchers are currently conducting clinical trials to establish conclusive results regarding the same [34].

Our study has shown neurodegenerative disorders, such as prion disease, Parkinson's disease (PD), Huntington's disease (HD), Alzheimer's disease (AD) and amyotrophic lateral sclerosis (ALS), as some of the highly significant pathways associated with COVID-19. All the above pathways were observed to be upregulated in the NHBE cell line.

The significant genes associated with the abovementioned pathways were ND6, NOTCH1, NDUFB10, UQCRB, NDUFA12, ATP6, ATP12, ITPR1, UQCR10, UBE2L4, TUBA1B, TUBB6, LRP1, and UCHL1 (S2E).

There were scattered families of involved genes belonging to the above-mentioned significant pathways. Gene association and interaction studies must be conducted to obtain a consolidated gene network. The finding above suggests that COVID-19 is significantly related to neurodegenerative disorders, which must be further studied to obtain conclusive results.

#### 4.3. Association of immunological pathways with COVID-19

The immune system is one of the most heavily affected systems in COVID-19. Many studies have shown that a host immune response is elicited by the SARS-CoV-2. SARS-CoV-2 infection has been observed to trigger both the adaptive and innate immune systems. Furthermore, B and T cells are essential for generating an immunogenic response against the virus. Moreover, COVID-19 is associated with cytokine storms. The genes associated with cytokine storms have also been identified [35].

Most of the upregulated immunological pathways were observed in the PBMC cell line. The pathways in this category include the antigen processing and presentation pathway, the phagosomal pathway, the hematopoietic cell lineage pathway, the influenza A pathway, the allograft rejection pathway and the lysosomal pathway. Lysosomes create an acidic environment that helps in combating the SARS-CoV-2 [36]. All the above-mentioned pathways are significant.

The CD74 gene is associated with pulmonary inflammation in COVID-19. The HLA family of genes regulates the immune system to

combat pathogens. The regulation of the HLA family of genes depends on many factors, such as disease severity and ethnicity of the test and control groups [37]. A gene network was obtained from the significant genes of the pathways belonging to this category, as shown in Fig. 9. Knowledge about the genes responsible for a host's immune response helps in monitoring the spread of the infection and can aid in further mutation studies.

#### 4.4. Association of cellular signalling and physiological pathways with COVID-19

Pathogens utilize the regulatory and signalling pathways of a host organism to provide for themselves, and this phenomenon exerts a profound impact on signalling pathways during a viral infection. Grimes et al. (2020) observed that the MAPK (Mitogen-activated protein kinase) signalling pathway is upregulated, serving as a pro-inflammatory response against pathogenic attacks [38]. The AMPK (adenosine monophosphate-activated protein kinase) pathway controls the autophagy response against viral infections. Downregulation of this pathway may result in a cytokine storm, which is an essential host immune response against COVID-19. The pathways in this category are the focal adhesion, ECM-receptor interaction, cellular senescence, TNF signalling, AMPK signalling and MAPK signalling pathways. All of these pathways are involved in cell proliferation, differentiation and, in the maintenance of cell integrity. In the NHBE cell line, some of the most significant pathways associated with the upregulated genes are the focal adhesion, ECM-receptor interaction and cellular senescence pathways, whereas the AMPK and TNF (Tumour Necrosis Factor) signalling pathways are associated with the downregulated genes. The presence of upregulated genes in the PBMC cell line indicates that the MAPK signalling pathway is one of the most significant pathways. Studying the regulation of cellular signalling pathways associated with COVID-19 will help us identify the different receptors and cell targets in order to narrow down and verify the various therapeutic approaches for COVID-19. A gene network for the above pathways is shown in Fig. 10.

Furthermore, a genetic network developed from the most significant genes associated with COVID-19 is presented in Fig. 11. These genes

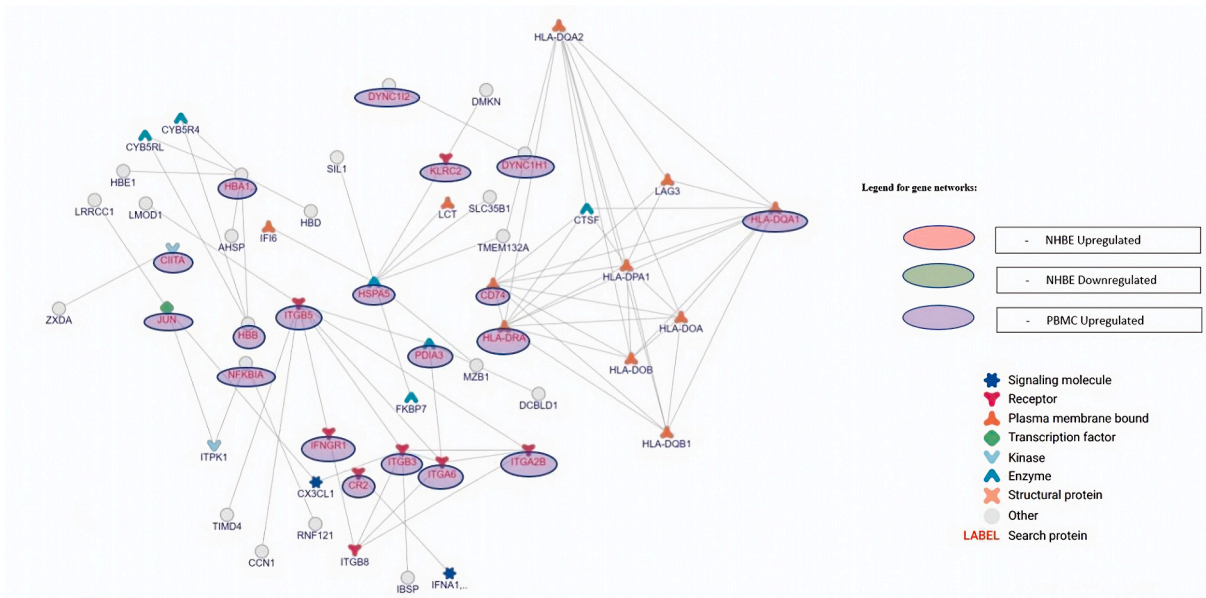


Fig. 9. represents the gene network for the immunological pathways derived from the significantly regulated genes of the PBMC cell line.

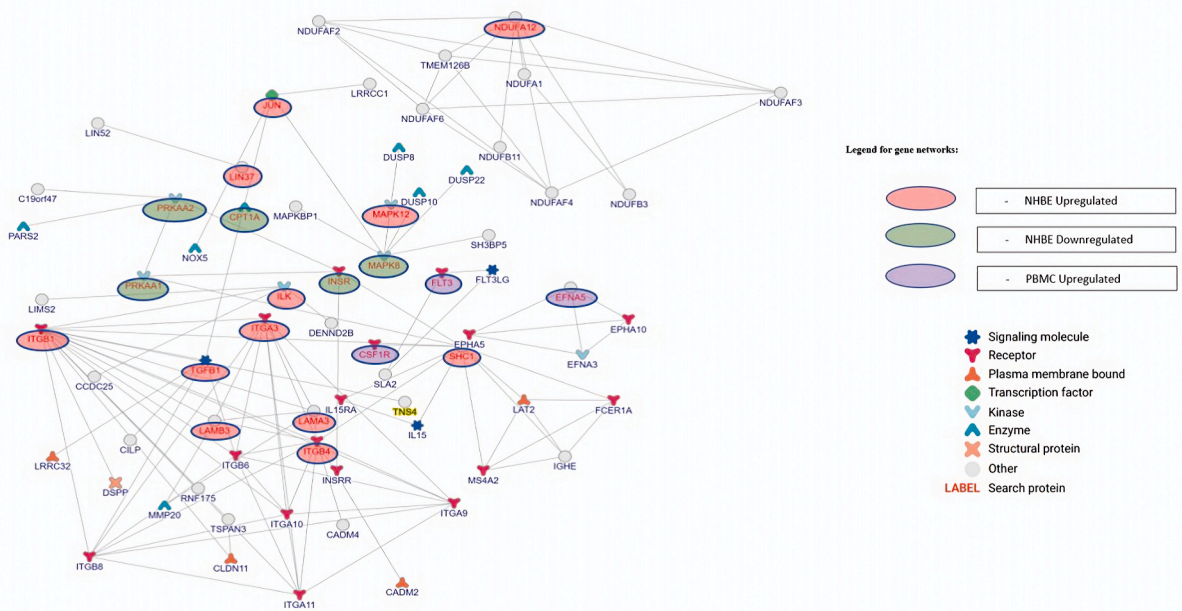


Fig. 10. visualises the gene network of the cellular signalling and physiology pathways derived from the significantly regulated genes of both the cell lines.

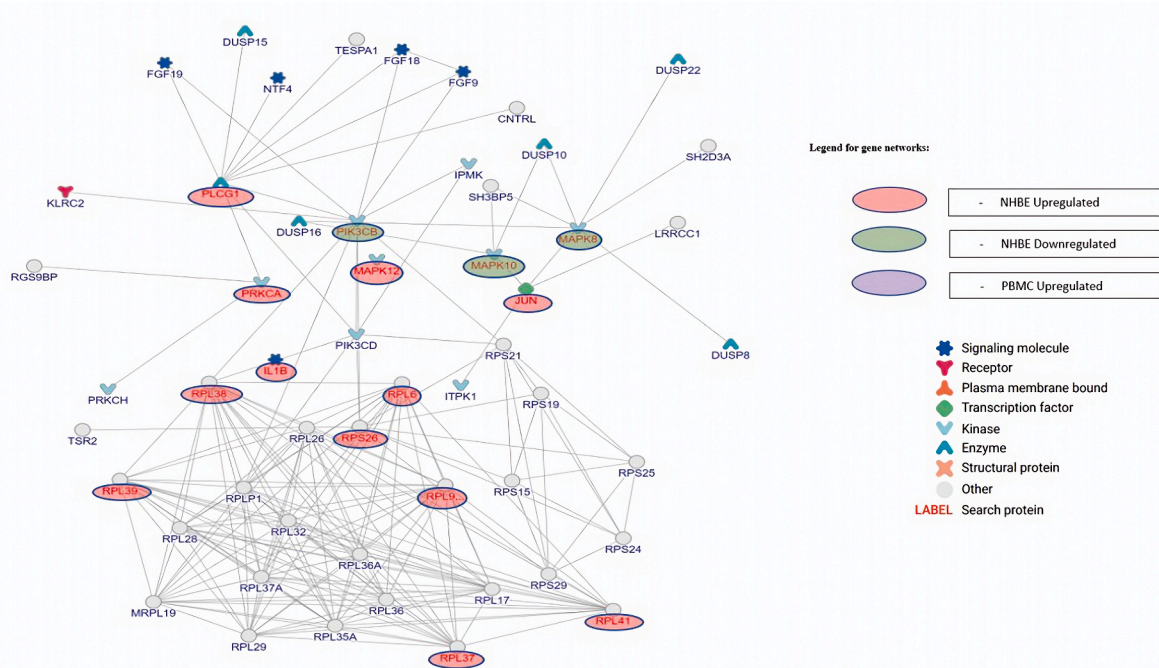
proved to be potential therapeutic targets against viral infection. Further gene functionality studies must be conducted to obtain conclusive results regarding the same.

### 5. Future work and conclusion

The coronavirus epidemic continues to spread over the world on an unpredictable trajectory. However, with the knowledge of pathways involved in COVID-19, the predictability of the disease increases and the control taken against this disease is also more assertive. Gene Regulatory pathways have become a wide focus of interest recently. The pathways obtained can determine the long-term effects of the coronavirus infection. Further, it can also help evaluate various therapeutic approaches based on the different gene targets found in the above-mentioned pathways. Understanding changes in glucose metabolism pathways

can help assess the contributing factors of long covid effects on diabetic patients. Although evidence from further studies is required, the link between neurodegenerative diseases and COVID-19 can ensure better care facilities for patients with neurodegenerative disorders during the pandemic. Early detection of patients with characteristic host immune response markers can help identify the progression of the disease in such patients and can guarantee appropriate measures accordingly. Comprehensive assessment of cellular physiology and signalling pathways paves the way toward novel therapeutic approaches and drug repurposing studies. Other important cell signalling targets of bronchial epithelial cells of COVID-19 patients must be identified in order to verify the current treatment strategies and propose potential strategies against the same.

Gene networks provide information regarding the interaction of various genes and can also throw light on potential protein interactions.



**Fig. 11.** shows the gene network derived from the significant genes associated with the COVID-19 disease.

The need of the hour is to perform long COVID studies to determine the exact impact of COVID-19 on individuals and societies. The SARS-CoV-2 mutates at an exponential rate and each variant presents different symptoms and attacks different pathways in the body. Under such circumstances, expanding research on the scope of the disease by going back to the basics will generate new and valuable information about the virus. This knowledge will help us combat the disease in a better and more appropriate manner.

### Funding

The research work conducted has no source of funding.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this study.

### Acknowledgement

We would like to sincerely thank the staff and non-staff members of the Department of Biotechnology at R V College of Engineering, Bangalore, India, for equipping us with the skills to conduct this project. We extend our thanks to the Center of Excellence, Computational genomics, R. V. College of Engineering, for the guidance to explore various topics and their constant support in this venture. We would like to thank Department of Computer science and engineering, R V College of Engineering for providing access to NVIDIA P1000 GPU to perform the analysis.

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.imu.2022.100979>.

### References

- [1] Ali I, Alharbi O. COVID-19: disease, management, treatment, and social impact. The Science of the total environment; 2020, 138861. <https://doi.org/10.1016/j.scitotenv.2020.138861>. 728.
- [2] Rabi FA, Al Zoubi MS, Kasasbeh GA, Salameh DM, Al-Nasser AD. SARS-CoV-2 and coronavirus disease 2019: what we know so far. Pathogens 2020;9(3):231. <https://doi.org/10.3390/pathogens9030231>.
- [3] Skariyachan S, Gopal D, Chakrabarti S, Kempanna P, Uttarkar A, Muddebihalkar AG, Niranjan V. Structural and molecular basis of the interaction mechanism of selected drugs towards multiple targets of SARS-CoV-2 by molecular docking and dynamic simulation studies- deciphering the scope of repurposed drugs. Comput Biol Med 2020;126:104054. <https://doi.org/10.1016/j.combiomed.2020.104054>.
- [4] Skariyachan Sinosh, Gopal Dharshini, Deshpande Dhriti, Joshi Anusha, Uttarkar Akshay, Niranjan Vidya. Carbon fullerene and nanotube are probable binders to multiple targets of SARS-CoV-2: insights from computational modeling and molecular dynamic simulation studies. Infect Genet Evol 2021:105155. <https://doi.org/10.1016/j.meegid.2021.105155>.
- [5] Skariyachan S, Gopal D, Muddebihalkar AG, Uttarkar A, Niranjan V. Structural insights on the interaction potential of natural leads against major protein targets of SARS-CoV-2: molecular modelling, docking and dynamic simulation studies. Comput Biol Med 2021;132:104325. <https://doi.org/10.1016/j.combiomed.2021.104325>.
- [6] Ahmad Shaban, Bhanu Piyush, Kumar Jitendra, Pathak Ravi Kant, Mallick Dharmendra, Uttarkar Akshay, Niranjan Vidya, Mishra Vachaspati. Molecular dynamics simulation and docking studies reveal NF- $\kappa$ B as a promising therapeutic drug target for COVID-19, Research Square. 2021. <https://doi.org/10.21203/rs.3.rs-469785/v2>.
- [7] Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. The architecture of SARS-CoV-2 transcriptome. Cell 2020;181(4):914–21. <https://doi.org/10.1016/j.cell.2020.04.011>. e10.
- [8] Wang X. Next-generation sequencing data analysis. first ed. CRC Press; 2016. <https://doi.org/10.1201/b19532>.
- [9] Li L, Wo J, Shao J, Zhu H, Wu N, Li M, Yao H, Hu M, Dennin RH. SARS-coronavirus replicates in mononuclear cells of peripheral blood (PBMCs) from SARS patients. J Clin Virol: Off. Pub. Pan. Am. Soc. Clin. Virol. 2003;28(3):239–44. [https://doi.org/10.1016/s1386-6532\(03\)00195-1](https://doi.org/10.1016/s1386-6532(03)00195-1).
- [10] Daamen AR, Bachali P, Owen KA, Kingsmore KM, Hubbard EL, Labonte AC, Robl R, Shrotri S, Grammer AC, Lipsky P. EComprehensive transcriptomic analysis of COVID-19 blood, lung, and airway. Sci Rep 2021;11(1):7052. <https://doi.org/10.1038/s41598-021-86002-x>.
- [11] Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Møller R, Jordan TX, Oishi K, Panis M, Sachs D, Wang TT, Schwartz RE, Lim JK, Albrecht RA, tenOever BR. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell 2020;181(5):1036–45. <https://doi.org/10.1016/j.cell.2020.04.026>. e9.
- [12] Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 2016;11(9):1650–67. <https://doi.org/10.1038/nprot.2016.095>.

- [13] Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol* 2010;11(10):R106. <https://doi.org/10.1186/gb-2010-11-10-r106>.
- [14] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15(12):550. <https://doi.org/10.1186/s13059-014-0550-8>.
- [15] Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* 2015;4:1521. <https://doi.org/10.12688/f1000research.7563.2>.
- [16] OmicsBox - BioBam | bioinformatics made easy - <https://www.biobam.com/omicsbox/>.
- [17] Robinson MD, McCarthy DJ, Smyth GK edgeR. A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26(1):139–40. <https://doi.org/10.1093/bioinformatics/btp616>.
- [18] Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;31(2):166–9. <https://doi.org/10.1093/bioinformatics/btu638>.
- [19] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9(4):357–9. <https://doi.org/10.1038/nmeth.1923>.
- [20] Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf* 2011;12:323. <https://doi.org/10.1186/1471-2105-12-323>.
- [21] Hänzelmann S, Castelo R, Guinney JGSA. Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf* 2013;14:7. <https://doi.org/10.1186/1471-2105-14-7>.
- [22] Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, Wu Y, Zhao L, Liu J, Guo J, Fang S, Cao W, Yi L, Zhao Y, Kong L, Kobas-i. Intelligent prioritisation and exploratory visualisation of biological functions for gene enrichment analysis. *Nucleic Acids Res* 2021;49:W317–25. <https://doi.org/10.1093/nar/gkab447>.
- [23] Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, Kong L, Gao G, Li C-Y, Wei L. Kobas 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res* 2011;39:W316–22. <https://doi.org/10.1093/nar/gkr483>.
- [24] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;44:W90–7. <https://doi.org/10.1093/nar/gkw377>.
- [25] Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinf* 2013;14:128. <https://doi.org/10.1186/1471-2105-14-128>.
- [26] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;44(W1):W90–7. <https://doi.org/10.1093/nar/gkw377>.
- [27] Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, Lachmann A, Wojciechowicz ML, Kropiwnicki E, Jagodnik KM, Jeon M, Ma'ayan A. Gene set knowledge discovery with Enrichr. *Curr. Protoc.* 2021;1:e90. <https://doi.org/10.1002/cpz1.90>.
- [28] Li T, Wernersson R, Hansen RB, Horn H, Mercer J, Slodkowitz G, Workman CT, Rigina O, Rapacki K, Starfeldt HH, Brunak S, Jensen TS, Lage K. A scored human protein-protein interaction network to catalyse genomic interpretation. *Nat Methods* 2017;14:61–4. <https://doi.org/10.1038/nmeth.4083>.
- [29] Lavanya C, Vidya Niranjana Dr, Upadhyaya Aajnaa, Guha Neogi Arpita. Identification of regulatory pathways across different cell lines (NHBE and PBMC) in COVID-19 patients using transcriptome analysis. <https://dx.doi.org/10.17504/protocols.io.bx6aprae>; 2021.
- [30] Diabetes and covid-19, American Diabetes Association, <https://www.diabetes.org/coronavirus-covid-19>.
- [31] Codo AC, Davanzo GG, Monteiro LB, de Souza GF, Muraro SP, Virgilio-da-Silva JV, Prodonoff JS, Carregari VC, de Biagi Junior C, Crunfli F, Jimenez Restrepo JL, Vendramini PH, Reis-de-Oliveira G, Bispo Dos Santos K, Toledo-Teixeira DA, Parise PL, Martini MC, Marques RE, Carmo HR, Borin A, Moraes-Vieira PM. Elevated glucose levels favor SARS-CoV-2 infection and monocyte response through a HIF-1 $\alpha$ /Glycolysis-Dependent Axis. *Cell Metabol* 2020;32(3):437–46. <https://doi.org/10.1016/j.cmet.2020.07.007>. e5.
- [32] Mahalaxmi I, Kaavya J, Mohana Devi S, Balachandrar V. COVID-19 and olfactory dysfunction: a possible associative approach towards neurodegenerative diseases. *J Cell Physiol* 2021;236(2):763–70. <https://doi.org/10.1002/jcp.29937>.
- [33] AD/PD 2021: impact of Covid-19 on neurodegenerative diseases. 2021. Pharmaceutical Technology, <https://www.pharmaceutical-technology.com/comment/ad-pd-2021-impact-covid-19-neurodegenerative-diseases/>. accessed 1.8.22.
- [34] Catanzaro M, Fagiani F, Racchi M, Corsini E, Govoni S, Lanni C. Immune response in COVID-19: addressing a pharmacological challenge by targeting pathways triggered by SARS-CoV-2. *Signal Transduct Targeted Ther* 2020;5:84. <https://doi.org/10.1038/s41392-020-0191-1>.
- [35] Coronaviruses hijack lysosomes to exit cells. 2020. National Institutes of Health (NIH). URL, <https://www.nih.gov/news-events/nih-research-matters/coronaviruses-hijack-lysosomes-exit-cells>. accessed 1.8.22.
- [36] Tavasolian F, Rashidi M, Hatam GR, Jeddi M, Hosseini AZ, Mosawi SH, Abdollahi E, Inman RD. HLA, immune response, and susceptibility to COVID-19. *Front Immunol* 2021;11:3581. <https://doi.org/10.3389/fimmu.2020.601886>.
- [37] Grimes, J.M., Grimes, K.V., p38 MAPK inhibition: a promising therapeutic approach for COVID-19. *J Mol Cell Cardiol* 2020;144, 63–65. <https://doi.org/10.1016/j.yjmcc.2020.05.007>.
- [38] Bhutta MS, Gallo ES, Borenstein R. Multifaceted role of AMPK in viral infections. *Cells* 2021;10:1118. <https://doi.org/10.3390/cells10051118>.