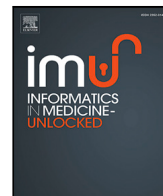




Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

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.



Systems Biology and Bioinformatics approach to Identify blood based signatures molecules and drug targets of patient with COVID-19

Md. Imran Hasan^{a,1}, Md Habibur Rahman^{a,1}, M. Babul Islam^b,
Md Zahidul Islam^c, Md Arju Hossain^d, Mohammad Ali Moni^{e,f,*}

^a Department of Computer Science and Engineering, Islamic University, Kushtia 7003, Bangladesh

^b Department of Electrical and Electronic Engineering, University of Rajshahi, 6205, Bangladesh

^c Department of Information and Communication Technology, Islamic University, Kushtia 7003, Bangladesh

^d Department of Biotechnology and Genetic Engineering, Mawlana Bhashani Science and Technology University, Tangail 1092, Bangladesh

^e Department of Computer Science and Engineering, Pabna University of Science and Technology, Pabna 6600, Bangladesh

^f Artificial Intelligence & Digital Health Data Science, School of Health and Rehabilitation Sciences, Faculty of Health and Behavioural Sciences, The University of Queensland, St Lucia, QLD 4072, Australia

ARTICLE INFO

Keywords:

COVID-19
Blood common gene
Differentially expressed genes
Biomarker signatures
Gene ontology
PBMCs cell

ABSTRACT

Severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) infection results in the development of a highly contagious respiratory ailment known as new coronavirus disease (COVID-19). Despite the fact that the prevalence of COVID-19 continues to rise, it is still unclear how people become infected with SARS-CoV-2 and how patients with COVID-19 become so unwell. Detecting biomarkers for COVID-19 using peripheral blood mononuclear cells (PBMCs) may aid in drug development and treatment. This research aimed to find blood cell transcripts that represent levels of gene expression associated with COVID-19 progression. Through the development of a bioinformatics pipeline, two RNA-Seq transcriptomic datasets and one microarray dataset were studied and discovered 102 significant differentially expressed genes (DEGs) that were shared by three datasets derived from PBMCs. To identify the roles of these DEGs, we discovered disease-gene association networks and signaling pathways, as well as we performed gene ontology (GO) studies and identified hub protein. Identified significant gene ontology and molecular pathways improved our understanding of the pathophysiology of COVID-19, and our identified blood-based hub proteins TPX2, DLGAP5, NCAPG, CCNB1, KIF11, HJURP, AURKB, BUB1B, TTK, and TOP2A could be used for the development of therapeutic intervention. In COVID-19 subjects, we discovered effective putative connections between pathological processes in the transcripts blood cells, suggesting that blood cells could be used to diagnose and monitor the disease's initiation and progression as well as developing drug therapeutics.

1. Introduction

SARS coronavirus 2 (SARS-CoV-2) is a new encapsulated RNA beta-coronavirus. The current coronavirus disease pandemic 2019 (COVID-19) is a worldwide public health emergency announced by the World Health Organization (WHO) and caused by the highly contagious severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Later, etiological investigations and deep sequencing revealed that a novel extremely infectious coronavirus, formally named SARS-CoV-2, was the causal pathogen of the outbreak [1–3]. At the end of June 2021, over 182 million cases of SARS-CoV-2 infection have been reported worldwide. After its discovery in December 2019, COVID-19 has spread throughout the world [4]. The widespread behavior of the virus has greatly influenced

the death rate, proving it to be the most internecine global epidemic of the twenty-first century. According to recent findings, SARS-CoV-2 infections do not impact all persons in the same manner, and that some people are more susceptible than others, with symptoms ranging from mild to severe, including a dry cough and sore throat, fever, and loss of taste and smell, to more severe symptoms such as trouble breathing or shortness of breath, necessitating the utilization of intensive care hospital facilities and, in some cases, death [5].

Some COVID-19 individuals, on the other hand, become ill with severe pneumonia symptoms and consequences, such as acute respiratory distress syndrome (ARDS), pulmonary oedema, acute kidney

* Corresponding author.

E-mail address: moni@pust.ac.bd (M.A. Moni).

¹ These authors contributed equally and hold joint first authorship of this work.

damage, or multiple organ failure, relatively fast after exposure [6–8]. COVID-19's spread has had a significant effect on the number of blood donors, blood supplies, and blood safety. SARS-CoV-2, a new coronavirus that causes the respiratory disease COVID-19, has recently spread rapidly. Unlike previous pandemic coronaviruses SARS and MERS, which demonstrated symptom co-occurrence with infectiousness, asymptomatic human to human transmission remain a difficult feature of the viral containment campaign [9,10]. An outbreak of pneumonia with an unclear cause emerged in early December 2019 [11]. Fever, dizziness, and cough were among the clinical symptoms, which were close to those of viral pneumonia.

The presence of genetic and transcript levels in patient tissues will aid in the detection of potential new biomarkers for identifying the remarkably similar phenotypes seen in COVID-19 patients [12,13]. Actually, PBMCs are a diverse group of immune cells that act as a first line defense against infections and disease-causing microorganisms. Jiang et al. reported that a significant decrease in CD3+ T cells, CD4+ T cells, CD8+ T cells and, natural killer cells in PBMCs indicates the severity of COVID-19 patients compared to moderate patients [14]. Despite the fact that the pathogenesis of COVID-19 is likely multifactorial, the use of molecular methods to improve diagnosis and evaluation of the disease has not yet produced conclusive results, prompting a renewed focus on the search for early COVID-19 biomarkers in PBMCs. According to Changfu Yao [15], single-cell RNA sequencing (scRNA-seq) of PBMCs enables in-depth analysis of transcriptional alterations in immune cells of COVID-19 patients. The transcriptome data from PBMCs infected with COVID-19 and CKD were analyzed in this study that make more reliable on finding common DEGs between them [16].

The discovery of these molecular blood biomarkers could have a significant effect on COVID-19 diagnosis, care, and treatment.

There have been a number of studies that have identified biomarkers that could be utilized in risk stratification models to predict severe and catastrophic outcomes COVID-19. It was discovered that biomarkers related to heart and muscle injuries, as well as lower enzymes, were considerably higher in COVID-19 patients who were severely ill or died [17]. IL-6 and IFN- γ were found to be significantly elevated only in late stages of severe infection, implying cytokine storms are the result of severe COVID-19 infection rather than the cause [18]. Yao et al. demonstrated that D-dimer levels in PBMCs correlate with disease severity, making it a reliable prognostic marker for in-hospital mortality in COVID-19 cases [19]. The SARS outbreak showed that low platelet counts were also related to greater illness severity; they occurred in up to 55% of patients, and they were connected with an increased risk of mortality [20].

More research into transcriptome analysis of coding and non-coding elements, as well as convalescent blood samples from both severe and mild COVID-19 cases, should help to identify molecular risk factors that can be used to predict highly susceptible individuals for severe COVID-19 infection, allowing for early intervention and customized treatment options.

Several gene expression profiling studies in COVID-19 have been conducted to characterize the association of the diseases. Because the functional relationship between gene products was not taken into consideration, these findings were limited to the transcript level [15,16]. Integrative studies within the network medicine framework are important to understand the molecular mechanisms behind diseases and to recognize crucial biomolecules since biological molecules interact with each other to carry out roles in biological processes in cells and tissues. An integrative strategy was employed to identify molecular biomarker profiles for COVID-19 that are expressed under the same genetic regulation in peripheral blood cells. This was accomplished through the use of transcriptome analysis.

In this study, we used publicly available PBMCs datasets for computational and transcriptome analyses. This research focused on biomarker signatures at the transcriptional and translational (hub proteins and transcription factors (TFs)) levels, to better understand COVID-19

pathogenic pathways. Each dataset's GO enrichment is done using a Fisher exact test and Kolmogorov–Smirnov test based on gene counts and gene ranks, respectively. DEGs has used genomics data to demonstrate the essential GO terms of genetic interrelationship. The research is significant because it is the largest comparative and transcriptomic study ever conducted on SARS-CoV-2 infection responses in human blood PBMC cells. The significance of the potential biomarkers that we have been able to identify in terms of appropriate immune responses has been demonstrated. Based on the transcriptomic analysis of SARS-CoV-2, the following analyses attempt to identify cell informative pathways. The genomic analysis was first used to identify genomic differences in the effect of SARS-CoV-2 on Homosapiens. This genomic level study will allow researchers to focus on SARS-CoV-2 and the major risk factors in the future.

1.1. Materials and methods

There are seven major phases in this work's overall approach. The dataset for blood (PBMC) cells is obtained in the first step. This step aims to make sure that the samples are taken from COVID-19. The differentially expressed genes (DEGs) from each of the selected datasets were determined in the second step of our analytical approach. In step three, we look for common DEGs in the COVID-19 blood PBMCs cell datasets. The next step was to conduct gene set enrichment analysis to determine the biological significance of the DEGs discovered. We concentrated on revealing protein–protein interaction networks in step five. In step six, we found gene regulatory network (GRN) interactions. The final stage of our investigation uncovered drug–molecule interactions. We looked at three SARS-CoV-2 contaminated datasets in this paper. The proposed workflow shows in Fig. 1. Each tissue group's preprocessing steps were completed independently.

We have considered the following points in selecting datasets for this study:

- (1) We removed duplicate samples that were included in multiple datasets from our study.
- (2) Blood (PBMC) cells are related in a lot of COVID-19 databases. However, we only look at total samples that are either case or control. Besides, tests from disease datasets that are unrelated to the regulation have been excluded.
- (3) Several datasets have been labeled with specifics related to a specific diagnosis, with a particular emphasis on biological interactions, but the results do not apply to the diagnosis and are therefore inappropriate.
- (4) Only human data was used and non-human datasets were discarded.
- (5) We count the number of Differentially Expressed Genes (DEGs) with an absolute log fold change value greater than or equal to 1 and a p -value less than 0.05.

1.2. Details about datasets

We looked at two RNA-Seq transcriptomic datasets and one microarray dataset related to COVID-19. One came from a study of PBMCs from SARS-CoV-2 patients at the Beijing Institute of Genomics Genome Sequence Archive in BIG Data Center (<https://bigd.big.ac.cn/>), P.R. China, with the data accession number CRA002390. Others, two datasets GSE152418 and GSE164805, were assembled from the Gene Expression Omnibus (GEO) database [21], where GSE152418 is an RNA-Seq dataset, and GSE164805 is a microarray dataset.

We discovered that the datasets we choose are appropriate for our study when compared to other available datasets. We filtered the datasets to select those with the least bias and noise for this work. Case and control samples were included in the datasets. Then, to approximate normality and minimize the effect of outliers, we

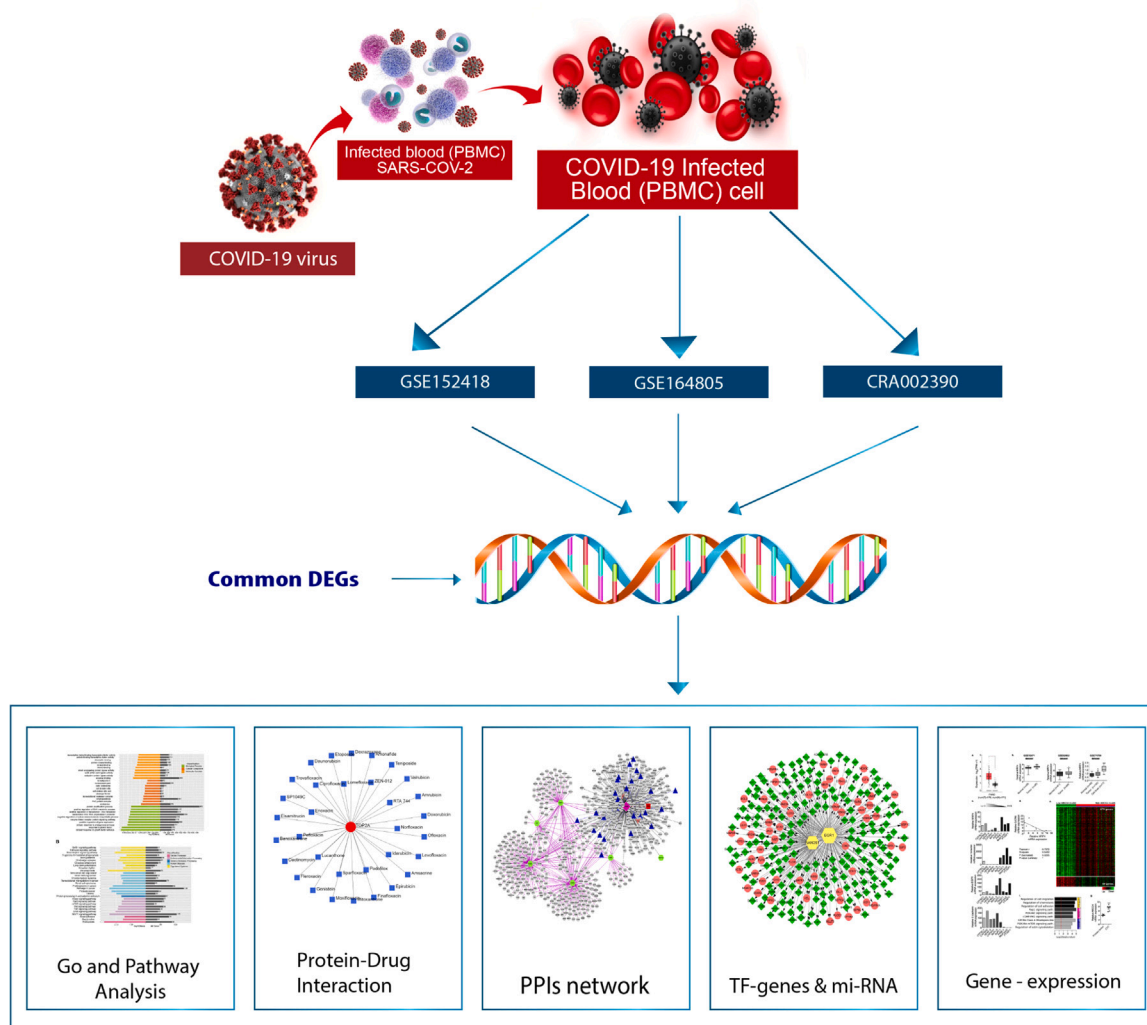


Fig. 1. The full workflow in this study. Gene expression datasets from COVID-19 matched control comparison studies of blood tissue were obtained from the Gene Expression (GEO) repository. These datasets were analyzed to identify common differentially expressed genes (DEGs) among blood tissue. The significantly enriched pathways and Gene Ontology (GO) terms were identified through enrichment analysis. Protein-protein interaction network was analyzed to identify hub proteins. Transcription Factor (TF)-target gene interactions and RNA-seq target gene interactions were also studied to identify regulatory biomolecules.

applied a logarithmic transformation to all the blood cell datasets. GREIN [22] was used to evaluate the RNA-seq data (GSE152418). Many more articles were used GREIN [22] to evaluate their RNA-seq dataset [23–25]. We used the GEO2R online tools to classify DEGs from the PBMCs microarray GSE164805 dataset using linear models. The p -values were modified using the Benjamini–Hochberg (BH) procedure [26–28]. The DEGs that differed among the three datasets were taken into consideration for future investigation.

1.3. DEGs of COVID-19 immune responses in common with blood datasets

In human tissues affected by COVID-19, gene expression analysis based on microarray and RNA-seq datasets can be a sensitive technique for studying global gene expression and identifying plausible molecular pathways that are activated, and this can be done with high sensitivity [29]. The transcriptome profile of diseased tissue was compared to the transcriptome profile of control (non-diseased) tissues to generate all of these microarrays and RNA-seq-based datasets. We can use this information to find biomarker genes linked to COVID-19 progression. In complex disease prognostic studies, achieving this can be difficult, but it can lead to a method for making more accurate prognosis [30].

1.4. Gene ontology and molecular pathways are discovered by gene set enrichment analysis

Gene Set Enrichment Analysis (GSEA) is a method for interpreting gene expression data and functionally enriched GO terms on various conditions or disease states. We identified the cells signaling pathways involving the DEGs found in blood PBMCs cell and then determined which other genes may be involved in those pathways. Due to hard-thresholding, a biological system may produce too few or too many genes as statistically significant, which can vary from one study to the next for a given set of genes. The GSEA method examines data from gene sets that are based on prior biological knowledge, such as gene pathways and gene expression profiles [31,32]. Gene set enrichment analysis is a computational and statistical method for determining whether a set of genes has statistical significance under various biological conditions [33].

GO resources include structural and computational information for gene product-based functions [34,35]. For gene product annotation, GO is divided into three sections: molecular function, biological process, and cellular component [36]. Signaling pathway analysis and gene ontology analysis are also part of gene collection enrichment analysis.

The biological significance of the established DEGs is determined through signaling pathways and ontology analysis. Enrichr was used to

find signaling pathways and ontology concepts. The Enrichr (<https://amp.pharm.mssm.edu/Enrichr/>) platform was used to collect GO terms for current analysis. Enrichr is a web-based program that contains massive gene sets made up of 102 libraries and runs genome-based experiments [37].

In experimental biology labs, fundamental interactions within complicated biological systems have frequently been organized into computable pathway representations [38]. In the context of precision medicine, databases may contain diverse representations of the same biological pathway [38]. Also, pathways are frequently described at various levels of detail, with a variety of data kinds and lack of clearly defined boundaries [39]. The Kyoto Encyclopedia of Genes and Genomes (KEGG) [40], Reactome [41], Wiki Pathways [42], and BioCarta databases are used for cell informative pathway research. The Enrichr framework is also used to apply the database performance. We determined other genes may play a role in the cell signaling pathways involving the DEGs found in the COVID-19 blood datasets and then which other genes may play a role in those pathways. In this enrichment study, we incorporated all of the DEGs discovered in COVID-19 blood PBMC cells.

1.5. Protein–protein interaction network analysis

Using all common DEGs among COVID-19 blood PBMCs, we created a protein–protein interaction (PPI) network using STRING (<https://string-db.org/>) [43]. A standard pathway starts with an extracellular signaling molecule that stimulates a specific receptor, triggering a series of protein–protein or protein–small molecule interactions. The study of protein interactions, which is regarded as the primary phase in drug discovery and systems biology, yields significant knowledge about the functions of proteins [44,45]. The advanced research of PPIs networks determines the number of complex biological processes [46].

1.6. The establishment of a topological algorithm on the PPIs network, as well as the identification of hub genes

Hub genes are genes that are strongly interconnected in a large-scale complex PPIs network [47]. The PPI network is made up of genes, edges, and their connections, with hub genes being the most entangled nodes. The maximal clique centrality (MCC) topological algorithm determines the hub genes for the current study. The MCC algorithm is applied to the PPIs network through cytoHubba, a Cytoscape software plugin (<http://apps.cytoscape.org/apps/cytohubba>). CytoHubba is a Cytoscape plugin that consists of 11 topological algorithms for ranking nodes in a particular network [48].

1.7. Analysis of TF–miRNA co-regulatory network

Transcription factors (TFs) bind to specific genes and regulate the rate of transcription of genetic information; thus, they are critical for molecular insights. The network repository was used to produce the TF–miRNA co-regulatory network review [49]. We used the NetworkAnalyst (<https://www.networkanalyst.ca/>) [50] platform to search the JASPAR database [51] for topologically credible TFs connected to our mutual DEGs. JASPAR is a freely accessible database of TF profiles from different species belonging to six taxonomic groups [52]. The co-regulatory network, which regulates DEGs at the transcriptional levels, is used to identify the miRNAs and TFs. The network was visualized using the NetworkAnalyst web-based framework. As the need for gene expression-based datasets grow, NetworkAnalyst has been used as a leading bioinformatics method for system-level data understanding [53,54]. We used the Network repository's TF-RNAseq coregulatory interactions to find regulatory TFs that control DEGs of interest at the transcriptional and post-transcriptional levels. The DEGs shared by COVID-19 blood PBMC cells were used. In this study, DEGs

derived from COVID-19 patients' peripheral blood cells were considered. The number of connections a node has with other nodes in the network determines its degree. We consider 9 degrees in the TF–miRNA network. Nodes with a higher degree are thought to be effective network hubs [55,56]. Furthermore, the node sizes are important. Nodes indicating genes that have strong connections with other differentially expressed genes appear to be greater in size when compared to the other nodes in the network [57]. We also discovered a network of interactions between transcription factors (TFs) and DEGs.

1.8. The analysis of suggested drugs identifies protein–drug interactions

In this study, the common DEGs revealed in the interaction of blood PBMC cells in COVID-19 were employed. We discovered protein–drug interactions that may affect these genes. The DEGs discovered in peripheral blood cells have been combined. To find potential drugs for the treatment of COVID-19, NetworkAnalyst was used to search the DrugBank database for protein–drug interactions [58]. A set of protein–drug interactions was chosen based on statistical significance thresholds for drug–protein interactions and the potential role of the targeted protein in COVID-19 pathogenesis, and simulations were run to analyze the binding affinities of identified drugs with their target protein.

2. Results

To examine the genetic and transcriptomic interactions of the COVID-19 with blood cells, we created a systematic and quantitative research process. Much of the data was gathered from publicly accessible sources.

2.1. Differential gene expression analysis identifies significant blood cells in COVID-19

A sensitive tool for investigating global gene expression and finding probable molecular pathways that are activated in human tissues impacted by COVID-19 is gene expression analysis utilizing microarray and RNA-seq datasets [59]. To better understand the transcriptional effects of COVID-19 on blood PBMC cells, we used a strict cut-off threshold of $\log_{2}(\text{foldchange}) \geq 1$ and a p -value < 0.05 to find genes that were differentially expressed compared to control patients. We compared the upregulated and downregulated genes with the significant upregulated and downregulated genes of blood cells in COVID-19 before moving on to other studies. We had 2621 DEGs after processing CRA002390 with 1654 up-regulated genes and 967 down-regulated genes. We also ran a comparison study to see the shared DEGs among the COVID-19 blood PBMCs cell datasets. A total of 2514 DEGs were discovered as a result of the analysis of GSE152418, with 2170 up-regulated genes and 344 down-regulated genes. On the other hand, after processing GSE164805, a total of 12,809 DEGs were discovered, with 6,705 up-regulated genes and 6,104 down-regulated genes.

We can see the Ven diagram for upregulated common gene in Fig. 2(C) and downregulated common gene in Fig. 2(D).

Each dataset's DEGs have been established, and a number of overlapping DEGs have been discovered. We can get the 87 upregulated and 15 downregulated common genes after comparing GSE152418, GSE164805, and CRA002390 their upregulated and downregulated genes. We also created heat maps to display the relationship among the overlapping DEGs. The heat map Fig. 2(A) represents the interaction between genes from the perspective of p -value, while the heat map in Fig. 2(B) shows the relationship between genes in terms of log fold change values [29,37]. Changing the significance level of differentially expressed gene products and the fold change cut-offs can reveal different results that imply different signaling pathways and functions involved. Several statistical test like p -value have been widely used for large sampling sizes (15,000 genes) which can influence the amount

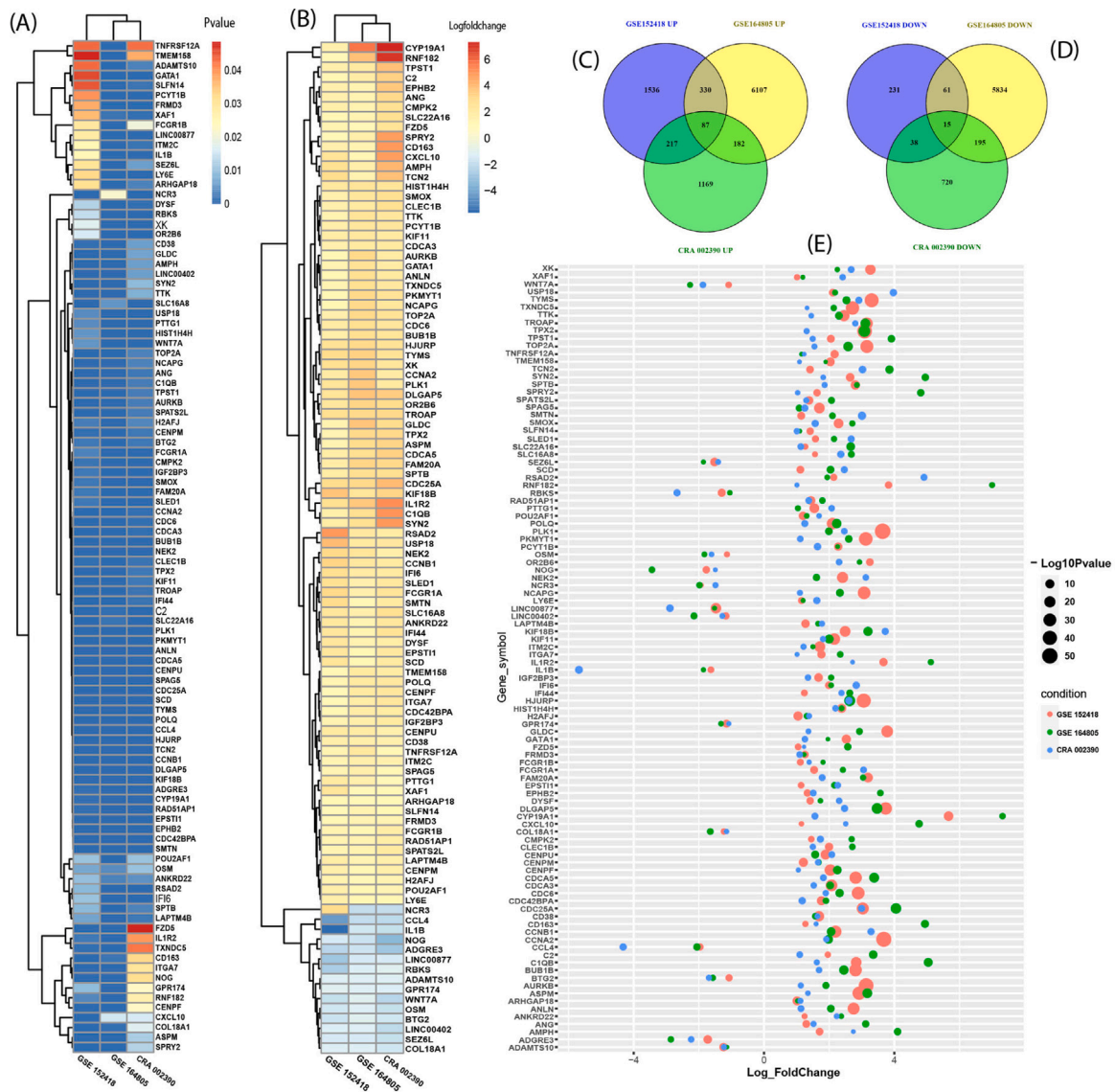


Fig. 2. Genes that are shared between COVID-19 infected patient blood (PBMC) cells are discovered by comparison of analyses. (A) In COVID-19, a heat map of the log fold change for the genes that are shared by all blood cells is depicted. (B) The heat map depicts the p-values for the genes that are shared by all of the blood cells in COVID-19. (C) The number of upregulated common important genes of COVID-19 in relation to the blood (PBMC) cells is represented by a Venn diagram. (D) The number of downregulated common important genes in COVID-19 as compared to blood (PBMC) cells is depicted in a Venn diagram. (E) The combined log-fold changes and p-values for the common genes shared by all blood (PBMC) cells in the experiment are depicted in the bubble plot.

of false positive rate and may indicate little about the biology. On the other hand, fold change allows for a more biologically meaningful assessment but still has difficulties recognizing what is significant to the organism [60]. Therefore, applying both criteria to generate heat map may assist but not singly solve the microarray and high throughput data analysis problem. For this reason, we applied together *p*-value and logFC value to generate heat map for connecting biological and statistical test and minimize background noises of DEGs. We created We also show that the striking essence of the unique transcriptional signature induced in blood PBMCs is visualized using bubble plots shows in Fig. 2(E).

2.2. Pathway enrichment analysis of DEGs reveals shared biological function between blood PBMCs cells in COVID-19

The pathway enrichment test measures the importance of a group of genes/proteins/molecules' overlap with an annotated group of genes/proteins/molecules known as *a priori* for their specific biological role, namely pathways, to determine their functional relevance.

Following the discovery of specific DEGs associated with SARS-CoV-2 infection profiles in blood PBMCs, a variety of databases (KEGG, Reactome, Wiki, BioCarta, and GO) were used to classify GO words and cell informative pathways. We discovered the cell signaling pathways in COVID-19 that include DEGs found in blood cells and then looked for other genes that could be involved in those pathways. We incorporated all of the DEGs found in blood and lung cells, as well as COVID-19 immune response cells, in this enrichment study. Using three ontology GO analysis databases, including GO Biological Process, GO Cellular Component, and GO Molecular Function, we identified the top 20 GO pathway from each database analysis of the common DEGs between blood cells in COVID-19 [61]. As shown in Fig. 4(A, B, C) we combined the ontology GO analysis from these databases and plotted the most important pathways based on the *p*-value. Pathways with a higher logarithmic *p*-value have a higher level of enrichment. Mitotic sister chromatid segregation (GO: 0000070), spindle assembly checkpoint (GO: 0071173), spindle pole (GO: 0000922) and kinase binding (GO: 0019900) were identified as the most significant GO pathway in our research.

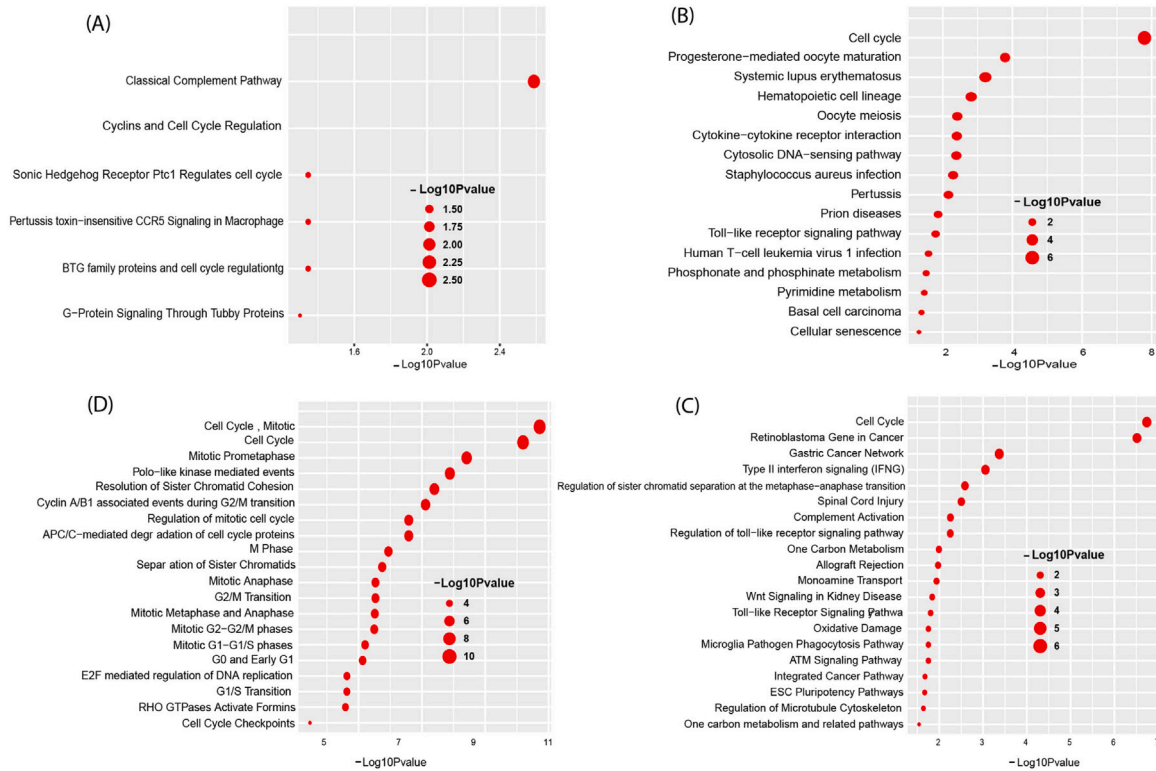


Fig. 3. Here the pathway analysis of blood PBMCs cell on COVID-19. We consider the p -value for each term. (A) indicates the biocarta, (B) indicates the KEGG, (C) and (D) indicates wiki & reactome Pathway analyses.

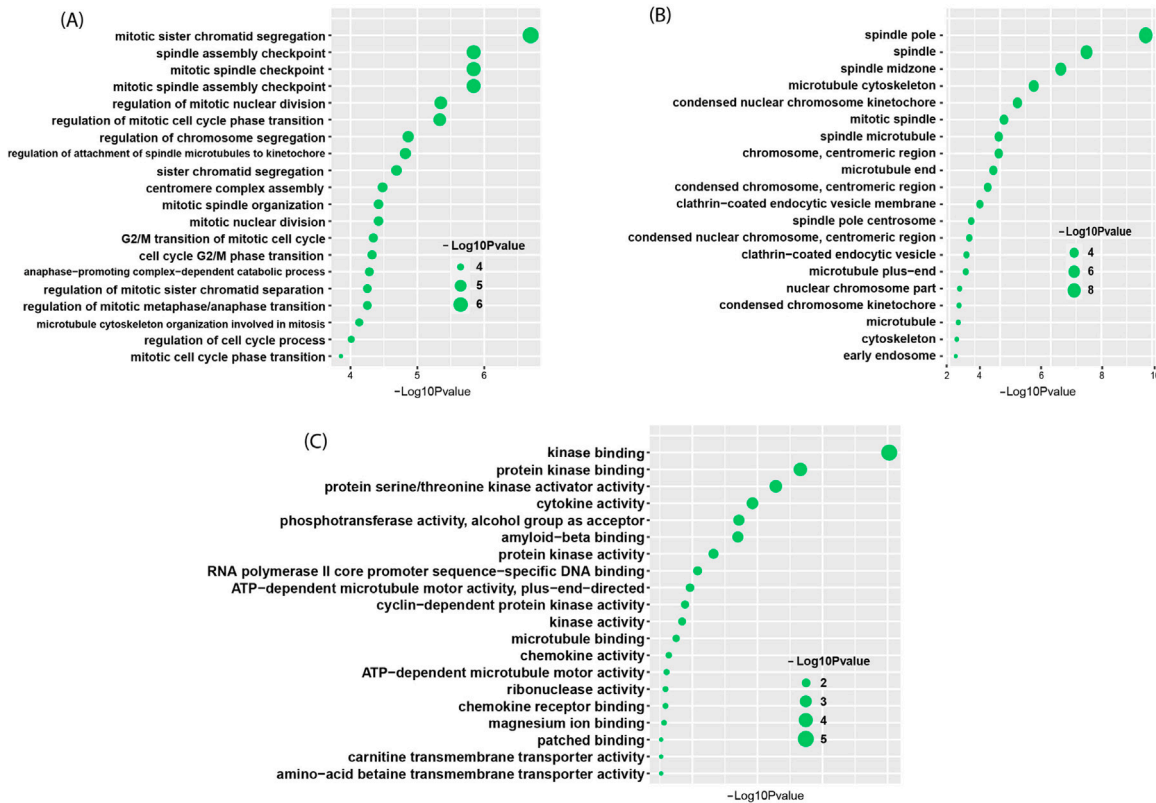


Fig. 4. Ontology analysis of blood PBMCs cell on COVID-19. We get top 20 term on depend of p -value. (A) GO Molecular Function, (B) GO Cellular Component, (C) GO Biological Process.

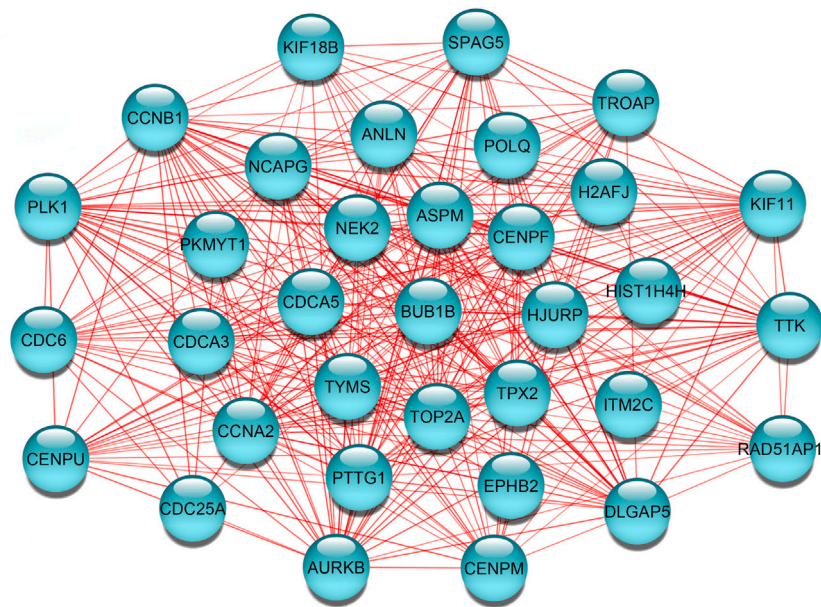


Fig. 5. Investigation of protein–protein interactions (PPI) in the vicinity of the proteins represented by the overlapping differentially expressed genes were carried out using a comprehensive PPI database referred to as STRING, which was searched using NetworkAnalyst, and the results were analyzed. The proteins are represented by the nodes, while the interactions between the proteins are represented by the edges.

In COVID-19 commonly enriched pathways against DEGs, we hypothesized that such a pathway enrichment test would reveal mutual biological functions of blood cells. An over-representation statistical test for the DEGs was performed with p -values < 0.05 to obtain significantly enriched pathways using KEGG pathways as known pathway annotation. We also conducted Reactome and Wiki Pathways analyses with the top significant genes between blood and lung cells in COVID-19. The most significant signaling pathway were BTG family proteins, cell cycle regulation, Cytosolic DNAsensing route, Tolllike receptor signaling pathway, and Type II interferon signaling (IFNG) can all be included for pathway analysis. Details were Shown in Fig. 3(A, B, C, D) respectively Biocarta, KEGG, WiKi and Reactome pathway analyses.

2.3. Protein–protein interaction network analysis

The PPI network was created using the STRING [43] web-based visualization resource, and the network is shown in Fig. 5. The figure represents the signature genes' participation and interaction in the PPI network. From the standpoint of PPI, we may also observe the relationship among the cell's genes. The network has 68 nodes and 502 edges, according to our findings. The proteins are represented by the nodes, and the interactions between the proteins are represented by the edges.

2.4. Hub proteins were identified from protein–protein interaction analysis

To predict typical DEG interactions and adhesion pathways, we examined the PPI network from STRING and visualized it in Cytoscape [62]. In a PPI network, the most interconnected nodes are known as hub genes. We identified the top 10 DEGs as the most influential genes based on the PPI network analysis in Cytoscape using the Cytohubba plugin. TPX2, DLGAP5, NCAPG, CCNB1, KIF11, HJURP, AURKB, BUB1B, TTK, and TOP2A are the hub genes. Hub proteins are thought to be drug targets. These hub genes may be used as biomarkers, which could contribute to new therapeutic approaches for the diseases being studied. As a result of the PPI research, the hub proteins were discovered. Using the degree steps, a protein–protein interaction network was built using DEGs to expose the central protein, the so-called hub proteins. These are potential biomarkers that could contribute to the

discovery of new COVID-19 therapeutic targets. Using a comprehensive PPI database called STRING, queried via NetworkAnalyst, protein–protein interactions (PPI) were performed around the proteins encoded by the overlapped differentially expressed genes [63,64]. The proteins are represented by the nodes, and the interactions between the proteins are represented by the edges. The hub gene network was created using the STRING web-based visualization resource, and it is depicted in Fig. 6. The figure depicts the signature genes' participation and interaction in the PPI network. From the standpoint of PPI, we may also observe the relationship among the cells.

2.5. Analysis of TF–miRNA co-regulatory network

A network-based method was used to unravel the regulatory TFs and miRNAs of the hub protein or DEGs, and TFs–miRNA linkages networks were studied to uncover transcriptional and post-transcriptional regulatory fingerprints of common DEGs. Fig. 7 depicts the interactions between TFs and miRNA. When it comes to a degree it all comes down to the number of connections the node has with other nodes in the network. We consider 9 degrees in the TF–miRNA network. Nodes with a higher degree are thought to be effective network hubs. Furthermore, the node sizes are important. If we look at all of the nodes in the network, the ones representing genes that have strong relationships with other differentially expressed genes appear to be larger than the others. In the TFs–miRNA network, the green color represents the TF (i.e. TFAP2A, TFAP2C, E2F1, E2F2, E2F3, E2F4, EGR1, NFYA, MYC, JUN, SP1, HNF4A, CTCF, TF53, USF1, MAX, MXI1, 23601) and the blue color represents the miRNA (i.e. hsa-let-7i, has-let-7e, hsa-let-7a, hsa-let-7b, hsa-let-7g, hsa-miR-98).

2.6. Candidate drugs are identified and evaluated

The goal was to identify possible medication candidates that could affect COVID-19 while simultaneously exploring the protein–drug interaction. The study of protein–drug interactions is crucial to comprehend the features required by sensitive receptors [65]. As a result of the protein–drug interaction research, it was discovered that the medication had an interaction with a hub protein. In Fig. 8 shows the association of two therapeutic compounds, Glycine and Pyridoxal phosphate, with the hub proteins of the GLDC, is demonstrated.

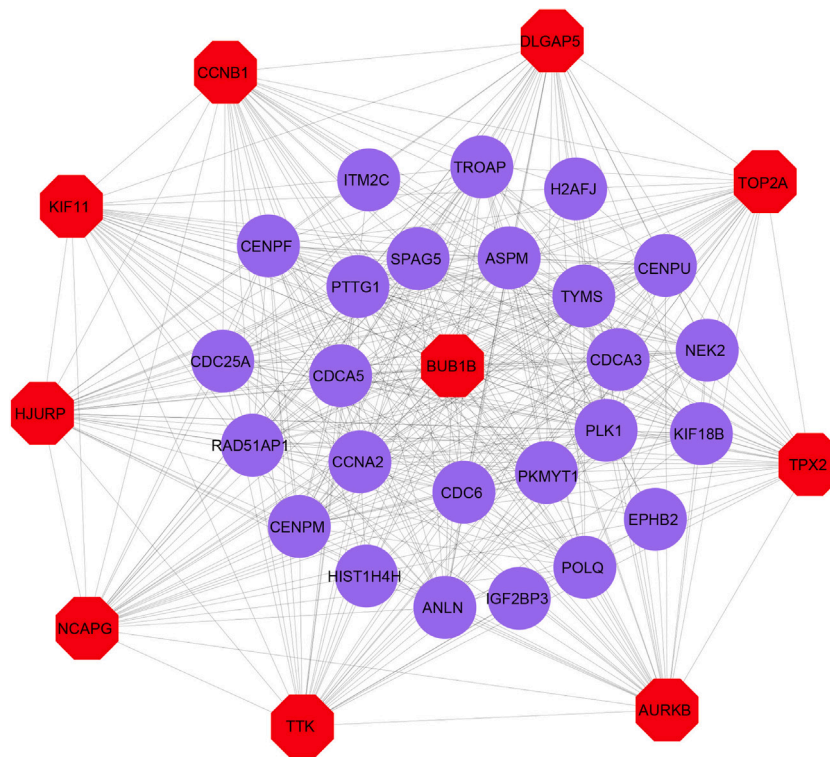


Fig. 6. The Cytohubba plugin in Cytoscape was used to determine the hub genes in the PPI network by analyzing the PPI network. To obtain the hub genes, the Cytohubba plugin was used in conjunction with the most recent MCC method. Here, the orange nodes indicate the highlighted top 10 hub genes and their interactions with other molecules..

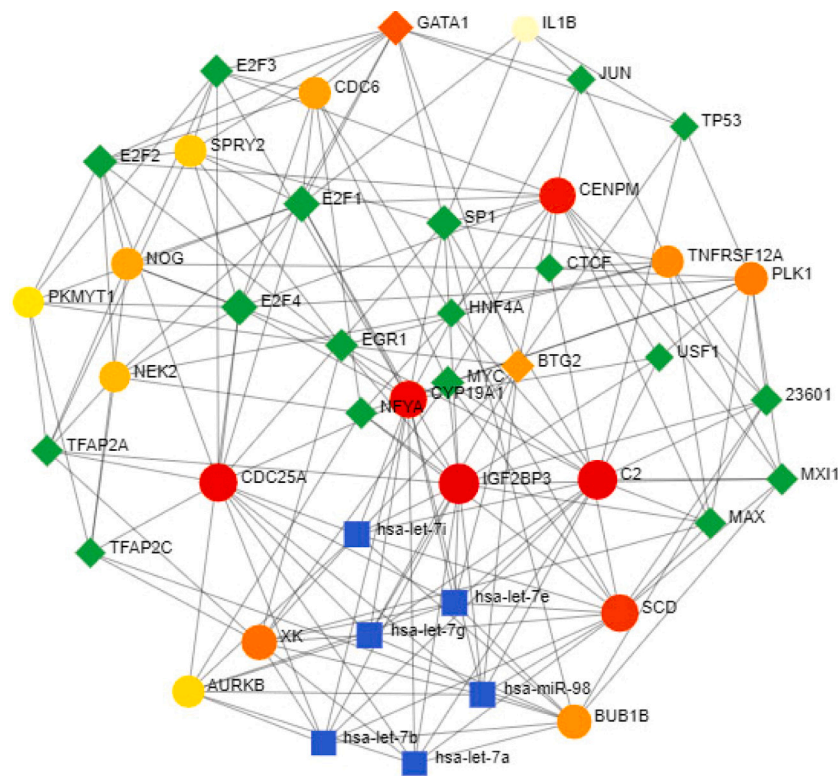


Fig. 7. The TF-miRNA regulatory interaction network is a network of interrelated regulatory interactions. There are TF genes shown with green and miRNAs indicated with blue in the network, and they are interacting..



Fig. 8. Protein drug interaction for drug analysis. It is possible to see the interaction between a hub protein and its drugs. The red circle GLDC indicates the hub protein and the other two yellow square boxes indicate Glycine and Pyridoxal Phosphate are possible drugs..

3. Discussion

COVID-19 has been shown to affect a variety of body systems, though it is debatable if it specifically affects the blood cells, therefore influences human behavior. Furthermore, previous analysis indicates that a portion of gene expression in PBMCs is associated with gene expression of COVID-19. Changfu Yao claims that [15], single-cell RNA sequencing (scRNA-seq) of PBMCs permits in-depth study of transcriptional changes in COVID-19 patients' immune cells. Another study looked at the transcriptome data from COVID-19 infected PBMC and CKD patients to see if there were any similar DEGs between them [16]. The lack of COVID-19 biomarkers in the peripheral blood has prompted attempts to find much-needed methods for the early detection of this debilitating disease.

Gene expression-based biological information can be discovered using large-scale microarray datasets. By contributing to the rapidly growing genome sequencing field, high-throughput sequencing has had a significant impact on biomedical research. SARS-CoV has already been subjected to high-throughput sequencing-based analysis, which has yielded impressive gene expression results [66–68].

To find possible biomarker candidates, we studied two gene expression datasets from the COVID-19 patients' peripheral blood. The discovery of peripheral biomarkers can also shed light on the molecular mechanisms of COVID-19 and allow for treatment monitoring. Transcriptomics analysis (via RNA-seq and microarray) is widely used to find candidate biomarkers for a variety of diseases. The three transcriptomic datasets of PBMCs blood tissues shared their DEGs, according to our study. We can get the 87 upregulated and 15 downregulated common genes after comparing GSE152418, GSE164805, and CRA002390 their upregulated and downregulated genes. Gene over-representation analysis and gene ontology (GO) analysis were performed on mutually dysregulated DEGs among blood cells, referred to as core DEGs in COVID-19. The enriched pathways by the established DEGs were then identified using pathway enrichment analysis, which included mitotic sister chromatid segregation (GO: 0000070), spindle pole (GO: 0000922) and kinase binding (GO: 0019900). In recent two studies, Chen et al. identified mitotic sister chromatid segregation and spindle pole GO pathway through integrative analysis in Hepatitis B virus-associated hepatocellular carcinoma [69] and mitotic sister chromatid segregation was identified in glioblastoma multiforme diseases [70].

Toll like receptor signaling pathway and Type II interferon signaling were the most potential pathway identified in our study. The SARS-CoV-2 spike protein S1 subunit activates TLR4 signaling to induce pro-inflammatory responses [71] and to increase ACE2 cell surface expression protein which facilitating the viral entry into the host cell [72]. The activated TLR4 causes the host's lung to react aggressively, resulting in a cytokine storm, building up secretions and impeding oxygenation of the blood, and attacking the body with the immune system, which leads to numerous organ failure [73].

TLR4 signaling in macrophages may therefore be a viable target in COVID-19 patients for the control of excessive inflammation. Khanmohammadi and Rezaei (2021) suggested that, in the early phases of the condition, TLRs might be a viable target to manage infection and production of SARS-CoV-2 vaccine [74].

Several studies conducted on human and animal models have revealed that interferon type 1 and 3 signaling are associated with

SARS-CoV-2 infection, and also suggest that dysregulated interferon signaling is a frequent molecular mechanism that develop the COVID-19 infection [75–77]. It is surprisingly said that type 2 interferon was identified in our study also would be used as a therapeutic target of SARS-COV-2 infection which was not previously identified in SARS-COV-2 development confirmed by literature analysis.

We also discovered dysregulated central hub proteins including TPX2, DLGAP5, NCAPG, CCNB1, KIF11, HJURP, AURKB, BUB1B, TTK, and TOP2A that govern many cellular processes using protein–protein interaction networks [78–80]. These hub proteins are thought to be important players in the disease's mechanisms. Among them, the Aurora Kinase B (AURKB) protein is effective for Hydrodynamic Analysis and Protein Interactions of the Chromosomal Passenger Complex [77]. Kim and Shin (2021) demonstrated that SARS-CoV-2 has been identified as DEGs in Caco-2 cells of the Aurora Kinase B (AURKB) and the Aurora Kinase A (AURKA) proteins [77]. The Aurora A Kinase (AURKA) is activated by TPX2 and contributes to cell cycle progression regulation. The overexpression of TPX2 improved the proliferative, invasive and migrating abilities of prostate cancer cells [81]. A recent study also suggesting that TPX2 genes may be useful targets for both the diagnosis and prognosis of hepatocellular carcinoma (HCC) patients who have been infected with HVB (hepatitis B virus) [82]. Auwul et al. found PLK1, AURKB, AURKA, CDK1, CDC20, KIF11, CCNB1, KIF2C, DTL and CDC6 hub genes which were serve as a potential biomarker of PBMCs in COVID-19 datasets that support our findings [83]. Y. Song et al. identified several hub genes including DLGAP5, TOP2A, AURKA, AURKB, and CCNA2 from lung adenocarcinoma cell. In clinical samples, qRT-PCR confirmed the presence of these hub genes which could serve as a therapeutic target for molecular cancer therapy [84].

Regulatory biomolecules are being investigated more and more as possible biomarkers for serious illnesses like neurodegenerative diseases. Multiple differentially expressed genes were identified during infection, suggesting that they could be used as disease biomarkers for SARS-CoV-2 viral infections. Proteins like these can play a role in the development and progression of COVID-19. DEGs, as well as TFs and miRNAs, were discovered to have a substantial effect on gene expression at the transcriptional and post-transcriptional stages. These DEGs was next analyzed in further depth to determine whether any regulatory factors, such as transcription factors (TFs), were present that could influence DEG levels in COVID-19 affected tissues.

Eventually, the potential medicine was discovered utilizing the signature gene reversal technique according to Auwul et al. and Fagone et al. studies [85,86]. Among them, Glycine is an important non-essential amino acid, which is to be investigated as a positive mitigator in COVID-19 patients for cell damage and proinflammatory storms [87]. The medication Pyridoxal supplementation may potentially alleviate the symptoms COVID-19 by reducing both the immunological suppression causing viral spread as well as the pathological hypersecretion of inflammatory cytokines [88–90]. We recommended sending these candidate medications for prospective use in COVID-19 therapy for biological and clinical testing.

For COVID-19 diagnostic development, PBMC gene expression analysis can play a putative role. Our findings indicate that the evolution of emerging diseases can be observed and analyzed using bioinformatics techniques, as it allows for the development of a better understanding of different cells. Understanding comorbidity associations is gaining popularity among scientists because it can reveal new information about disease causes as well as potential therapeutic strategic goals. This study highlights the significance of using advanced bioinformatics and system biology to uncover possible disease interactions and drug development opportunities. The research also focused on the examination of the gene expression in PBMC in order to learn about the possible use of the discovered hubs proteins for COVID-19 diagnostics, then we chose to classify the bioinformatics method which has been frequently utilized.

Although, some limitations of the study should be acknowledged because the results of this study rely on only bioinformatics analysis where biomedical analysis in the wet lab is mandatory for better confirmation. For this reason, possible caution should be considered during the interpretation of data analysis. In addition, PBMCs gene expression in COVID-19 was reversed in the transcriptomic analysis to identify potential drugs, but lung tissues are the primary organs affected by SARS-CoV-2. Besides, the transcriptomic results were obtained in our study for a small number of samples infected with SARS-CoV-2 while a greater number of samples will result in a substantial number of concordant genes. Despite the importance of the current systems biology study of COVID-19 gene expression to identify putative biomarkers, we propose wet-lab experimentation to confirm these candidates through *in vivo* analysis to develop them as new biomarkers in COVID-19 disease progression.

4. Conclusion

In this research, we looked at the transcriptomics of blood PBMCs cells to find DEGs that were shared among the three datasets. These common DEGs were added to the investigation of protein–protein interactions in the context of pathway analysis, transcription factors, and miRNA. RNA-seq and microarray data from blood cells and found 102 DEGs. Toll like receptor, type II interferon signaling pathway, mitotic sister chromatid segregation, spindle pole and kinase binding were most significant molecular and gene ontology pathway involved in COVID-19 pathogenesis. The 10 hub genes including TPX2, DLGAP5, NCAPG, CCNB1, KIF11, HJURP, AURKB, BUB1B, TTK, and TOP2A were identified from the PPI networks of these 102 DEGs. Several TFs (TFAP2A, E2F1, NFYA, MYC, JUN, SP1, TF53, USF1, and MAX) and the miRNA (let-7i, let-7e, let-7a, let-7b, and miR-98) were identified as putative transcriptional and post-transcriptional regulators of the DEGs. We have found two potential drugs Glycine and Pyridoxal which target the biomarkers that we discovered for COVID-19 pathogenesis. Several TFs and miRNAs were identified as putative transcriptional and post-transcriptional regulators of the DEGs that we identified. These results add to our knowledge of COVID-19's relationship with these blood response genes and demonstrate how the infection could be investigated for other diseases. Although, the results of this study rely on only bioinformatics analysis where biomedical analysis in the wet lab is mandatory for better confirmation. Besides, the samples of SARS-CoV-2 were collected at different times and the sample number is lower. We now recommended a more rigorous validation of this identifying biomarkers through in wet lab experiments for a therapeutic intervention to identify in COVID-19 pathogenesis.

Abbreviations

- (SARS-CoV-2) - Severe acute respiratory syndrome coronavirus-2.
- (PBMCs) - Peripheral Blood Mononuclear Cells.
- (DEGs) - Differentially Expressed Genes.
- (WHO) - World Health Organization.
- (scRNA-seq) - single-cell RNA sequencing.
- (CKD) - Chronic kidney disease.
- (TFs) - Transcription factors.
- (GRN) - Gene Regulatory Network.
- (GEO) - Gene Expression Omnibus.
- (GREIN) - GEO RNA-seq experiments interactive navigator.
- (GSEA) - Gene Set Enrichment Analysis.
- (KEGG) - The Kyoto Encyclopedia of Genes and Genomes.
- (BH) - Benjamini–Hochberg.
- (PPI) - Protein–Protein Interactions.
- (MCC) - maximal clique centrality.
- (HCC) - hepatocellular carcinoma.

CRedit authorship contribution statement

Md. Imran Hasan: Conception and design, or analysis and interpretation of the data, Writing – original draft, Writing – review & editing. **Md Habibur Rahman:** Conception and design, or analysis and interpretation of the data, Writing – original draft, Writing – review & editing. **M. Babul Islam:** Writing – original draft, Writing – review & editing. **Md Zahidul Islam:** Writing – original draft, Writing – review & editing. **Md Arju Hossain:** Writing – original draft, Writing – review & editing. **Mohammad Ali Moni:** Writing – original draft, Writing – review & editing.

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 paper.

Acknowledgment

All authors approved the version of the manuscript to be published.

References

- [1] Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, Si H-R, Zhu Y, Li B, Huang C-L, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;579(7798):270–3.
- [2] Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, Hu Y, Tao Z-W, Tian J-H, Pei Y-Y, et al. A new coronavirus associated with human respiratory disease in China. *Nature* 2020;579(7798):265–9.
- [3] Rahman MH, Rana HK, Peng S, Kibria MG, Islam MZ, Mahmud SH, Moni MA. Bioinformatics and system biology approaches to identify pathophysiological impact of COVID-19 to the progression and severity of neurological diseases. *Comput Biol Med* 2021;104859.
- [4] WHO, WHO Coronavirus (COVID-19) Dashboard, <https://covid19.who.int/>.
- [5] Mattiuzzi C, Lippi G. Which lessons shall we learn from the 2019 novel coronavirus outbreak? *Ann Transl Med* 2020;8(3).
- [6] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395(10223):497–506.
- [7] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020;395(10223):507–13.
- [8] Guan W-j, Ni Z-y, Hu Y, Liang W-h, Ou C-q, He J-x, Liu L, Shan H, Lei C-l, Hui D-S, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020;382(18):1708–20.
- [9] Nashiry A, Sarmin Sumi S, Islam S, Quinn JM, Moni MA. Bioinformatics and system biology approach to identify the influences of COVID-19 on cardiovascular and hypertensive comorbidities. *Brief Bioinform* 2021;22(2):1387–401.
- [10] Birse CE, Lagier RJ, FitzHugh W, Pass HI, Rom WN, Edell ES, Bungum AO, Maldonado F, Jett JR, Mesri M, et al. Blood-based lung cancer biomarkers identified through proteomic discovery in cancer tissues, cell lines and conditioned medium. *Clin Proteom* 2015;12(1):1–12.
- [11] Yang X, Yu Y, Xu J, Shu H, Liu H, Wu Y, Zhang L, Yu Z, Fang M, Yu T, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med* 2020;8(5):475–81.
- [12] Geng R-X, Li N, Xu Y, Liu J-h, Yuan F-e, Sun Q, Liu B-H, Chen Q-X. Identification of core biomarkers associated with outcome in glioma: evidence from bioinformatics analysis. *Disease Markers* 2018;2018.
- [13] Bost P, Giladi A, Liu Y, Bendjelal Y, Xu G, David E, Blecher-Gonen R, Cohen M, Medaglia C, Li H, et al. Host-viral infection maps reveal signatures of severe COVID-19 patients. *Cell* 2020;181(7):1475–88.
- [14] Jiang M, Guo Y, Luo Q, Huang Z, Zhao R, Liu S, Le A, Li J, Wan L. T-cell subset counts in peripheral blood can be used as discriminatory biomarkers for diagnosis and severity prediction of coronavirus disease 2019. *J Infect Dis* 2020;222(2):198–202.
- [15] Yao C, Bora SA, Chen P, Goodridge HS, Gharib SA. Sample processing and single cell RNA-sequencing of peripheral blood immune cells from COVID-19 patients. *STAR Protocols* 2021;2(2):100582.
- [16] Rabiul Auwal M, Zhang C, Rezanur Rahman M, Shahjaman M, Alyami SA, Ali Moni M. Network-based transcriptomic analysis identifies the genetic effect of COVID-19 to chronic kidney disease patients: A bioinformatics approach. 2021.

- [17] Henry BM, De Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. *Clin Chem Lab Med* 2020;58(7):1021–8.
- [18] Zhao Y, Qin L, Zhang P, Li K, Liang L, Sun J, Xu B, Dai Y, Li X, Zhang C, et al. Longitudinal COVID-19 profiling associates IL-1RA and IL-10 with disease severity and RANTES with mild disease. *JCI Insight* 2020;5(13).
- [19] Yao Y, Cao J, Wang Q, Shi Q, Liu K, Luo Z, Chen X, Chen S, Yu K, Huang Z, et al. D-dimer as a biomarker for disease severity and mortality in COVID-19 patients: a case control study. *J Intensive Care* 2020;8(1):1–11.
- [20] Lippi G, Plebani M, Henry BM. Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: a meta-analysis. *Clin Chim Acta* 2020;506:145–8.
- [21] Clough E, Barrett T. The gene expression omnibus database. In: *Statistical genomics*. Springer; 2016, p. 93–110.
- [22] Al Mahi N, Najafabadi MF, Pilarczyk M, Kouril M, Medvedovic M. GREIN: An interactive web platform for re-analyzing GEO RNA-seq data. *Sci Rep* 2019;9(1):1–9.
- [23] Mahi N. Connectivity analysis of single-cell RNA-seq derived transcriptional signatures. [Ph.D. thesis], University of Cincinnati; 2020.
- [24] Bernstein MN, Ni Z, Collins M, Burkard ME, Kendzioriski C, Stewart R. CHARTS: a web application for characterizing and comparing tumor subpopulations in publicly available single-cell RNA-seq data sets. *BMC Bioinformatics* 2021;22(1):1–9.
- [25] Gavriush GE, Chistyakov DV, Sergeeva MG. ARGEOS: A new bioinformatic tool for detailed systematics search in GEO and ArrayExpress. *Biology* 2021;10(10):1026.
- [26] Nienhold R, Ciani Y, Koelzer VH, Tzankov A, Haslbauer JD, Menter T, Schwab N, Henkel M, Frank A, Zsikla V, et al. Two distinct immunopathological profiles in autopsy lungs of COVID-19. *Nature Commun* 2020;11(1):1–13.
- [27] Shaath H, Toor S, Nair VS, Elkord E, Alajez NM. Transcriptomic analyses revealed systemic alterations in gene expression in circulation and tumor microenvironment of colorectal cancer patients. *Cancers* 2019;11(12):1994.
- [28] Rahman MH, Peng S, Hu X, Chen C, Uddin S, Quinn JM, Moni MA. Bioinformatics methodologies to identify interactions between type 2 diabetes and neurological comorbidities. *IEEE Access* 2019;7:183948–70.
- [29] Rahman MH, Rana HK, Peng S, Hu X, Chen C, Quinn JM, Moni MA. Bioinformatics and machine learning methodologies to identify the effects of central nervous system disorders on glioblastoma progression. *Brief Bioinform* 2021;bbaa365.
- [30] Liang J, Lv X, Lu C, Ye X, Chen X, Fu J, Luo C, Zhao Y. Prognostic factors of patients with Gliomas—an analysis on 335 patients with Glioblastoma and other forms of Gliomas. *BMC Cancer* 2020;20(1):1–7.
- [31] Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci* 2005;102(43):15545–50.
- [32] Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* 2010;26(7):976–8.
- [33] Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for gene set enrichment analysis. *Bioinformatics* 2007;23(23):3251–3.
- [34] Consortium GO. The gene ontology resource: 20 years and still going strong. *Nucleic Acids Res* 2019;47(D1):D330–8.
- [35] Podder NK, Rana HK, Azam MS, Rana MS, Akhtar MR, Rahman MR, Rahman MH, Moni MA. A system biological approach to investigate the genetic profiling and comorbidities of type 2 diabetes. *Gene Reports* 2020;21:100830.
- [36] Doms A, Schroeder M. GoPubMed: exploring PubMed with the gene ontology. *Nucleic Acids Res* 2005;33(suppl_2):W783–6.
- [37] Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;44(W1):W90–7.
- [38] Mubeen S, Hoyt CT, Gemünd A, Hofmann-Apitius M, Fröhlich H, Domingo-Fernández D. The impact of pathway database choice on statistical enrichment analysis and predictive modeling. *Front Genet* 2019;10:1203.
- [39] Domingo-Fernández D, et al. CompPath: An ecosystem for exploring, analyzing, and curating mappings across pathway databases. 2018, bioRxiv 353235.
- [40] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 2000;28(1):27–30.
- [41] Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, Haw R, Jassal B, Korninger F, May B, et al. The reactome pathway knowledgebase. *Nucleic Acids Res* 2018;46(D1):D649–55.
- [42] Slenter DN, Kutmon M, Hanspers K, Riutta A, Windsor J, Nunes N, Mélius J, Cirillo E, Coort SL, Digles D, et al. WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res* 2018;46(D1):D661–7.
- [43] Mering Cv, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res* 2003;31(1):258–61.
- [44] Šikić M, Tomić S, Vlahoviček K. Prediction of protein–protein interaction sites in sequences and 3D structures by random forests. *PLoS Comput Biol* 2009;5(1):e1000278.
- [45] Mahmud SH, Al-Mustanjid M, Akter F, Rahman MS, Ahmed K, Rahman MH, Chen W, Moni MA. Bioinformatics and system biology approach to identify the influences of SARS-CoV-2 infections to idiopathic pulmonary fibrosis and chronic obstructive pulmonary disease patients. *Brief Bioinform* 2021.
- [46] Pagel P, Kovac S, Oesterheld M, Brauner B, Dunger-Kaltenbach I, Frishman G, Montrone C, Mark P, Stümpflen V, Mewes H-W, et al. The MIPS mammalian protein–protein interaction database. *Bioinformatics* 2005;21(6):832–4.
- [47] Hsing M, Byler KG, Cherkasov A. The use of Gene Ontology terms for predicting highly-connected hub nodes in protein–protein interaction networks. *BMC Syst Biol* 2008;2(1):1–14.
- [48] Chin C-H, Chen S-H, Wu H-H, Ho C-W, Ko M-T, Lin C-Y. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 2014;8(4):1–7.
- [49] Liu Z-P, Wu C, Miao H, Wu H. RegNetwork: an integrated database of transcriptional and post-transcriptional regulatory networks in human and mouse. *Database* 2015;2015.
- [50] Xia J, Gill EE, Hancock RE. NetworkAnalyst for statistical, visual and network-based meta-analysis of gene expression data. *Nat Protoc* 2015;10(6):823–44.
- [51] Fornes O, Castro-Mondragon JA, Khan A, Van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranašić D, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2020;48(D1):D87–92.
- [52] Khan A, Fornes O, Stigliani A, Gheorghe M, Castro-Mondragon JA, Van Der Lee R, Bessy A, Cheneby J, Kulkarni SR, Tan G, et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* 2018;46(D1):D260–6.
- [53] Zhou G, Soufan O, Ewald J, Hancock RE, Basu N, Xia J. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res* 2019;47(W1):W234–41.
- [54] Al-Mustanjid M, Mahmud SH, Royel MRI, Rahman MH, Islam T, Rahman MR, Moni MA. Detection of molecular signatures and pathways shared in inflammatory bowel disease and colorectal cancer: A bioinformatics and systems biology approach. *Genomics* 2020;112(5):3416–26.
- [55] Oany AR, Mia M, Pervin T, Junaid M, Hosen SZ, Moni MA. Design of novel viral attachment inhibitors of the spike glycoprotein (S) of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) through virtual screening and dynamics. *Int J Antimicrob Ag* 2020;56(6):106177.
- [56] Thomas JG, Olson JM, Tapscott SJ, Zhao LP. An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Res* 2001;11(7):1227–36.
- [57] Gaiteri C, Sibille E. Differentially expressed genes in major depression reside on the periphery of resilient gene coexpression networks. *Front Neurosci* 2011;5:95.
- [58] Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018;46(D1):D1074–82.
- [59] Rahman MH, Peng S, Hu X, Chen C, Rahman MR, Uddin S, Quinn JM, Moni MA. A network-based bioinformatics approach to identify molecular biomarkers for type 2 diabetes that are linked to the progression of neurological diseases. *Int J Environ Res Public Health* 2020;17(3):1035.
- [60] Dalman MR, Deeter A, Nimishakavi G, Duan Z-H. Fold change and p-value cutoffs significantly alter microarray interpretations. In: *BMC Bioinformatics*. 13, (2):BioMed Central; 2012, p. 1–4.
- [61] Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One* 2010;5(11):e13984.
- [62] Smoot ME, Ono K, Ruscheinski J, Wang P-L, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011;27(3):431–2.
- [63] Ekman D, Light S, Björklund AsK, Elofsson A. What properties characterize the hub proteins of the protein–protein interaction network of *saccharomyces cerevisiae*? *Genome Biol* 2006;7(6):1–13.
- [64] Tsai C-J, Ma B, Nussinov R. Protein–protein interaction networks: how can a hub protein bind so many different partners? *Trends Biochem Sci* 2009;34(12):594–600.
- [65] Mahmud SH, Chen W, Liu Y, Awal MA, Ahmed K, Rahman MH, Moni MA. PredTIs: prediction of drug–target interactions based on multiple feature information using gradient boosting framework with data balancing and feature selection techniques. *Brief Bioinform* 2021.
- [66] Wichert S, Fokianos K, Strimmer K. Identifying periodically expressed transcripts in microarray time series data. *Bioinformatics* 2004;20(1):5–20.
- [67] Soon WW, Hariharan M, Snyder MP. High-throughput sequencing for biology and medicine. *Mol Syst Biol* 2013;9(1):640.
- [68] Irigoyen N, Firth AE, Jones JD, Chung BY-W, Siddell SG, Brierley I. High-resolution analysis of coronavirus gene expression by RNA sequencing and ribosome profiling. *PLoS Pathogens* 2016;12(2):e1005473.
- [69] Chen Z, Chen J, Huang X, Wu Y, Huang K, Xu W, Xie L, Zhang X, Liu H. Identification of potential key genes for hepatitis B virus-associated hepatocellular carcinoma by bioinformatics analysis. *J Comput Biol* 2019;26(5):485–94.
- [70] Yang J, Wang L, Xu Z, Wu L, Liu B, Wang J, Tian D, Xiong X, Chen Q. Integrated analysis to evaluate the prognostic value of signature mRNAs in glioblastoma multiforme. *Front Genet* 2020;11:253.

- [71] Aboudounya MM, Heads RJ. COVID-19 and toll-like receptor 4 (TLR4): SARS-CoV-2 may bind and activate TLR4 to increase ACE2 expression, facilitating entry and causing hyperinflammation. *Mediators Inflamm* 2021;2021.
- [72] Edokpayi JN, Odiyo JO, Popoola OE, Msagati TA. Evaluation of contaminants removal by waste stabilization ponds: A case study of Siloam WSPs in Vhembe district, South Africa. *Heliyon* 2021;7(2):e06207.
- [73] Brandão SCS, Ramos JdOX, Dompieri LT, Godoi ETAM, Figueiredo JL, Sarinho ESC, Chelvanambi S, Aikawa M. Is toll-like receptor 4 involved in the severity of COVID-19 pathology in patients with cardiometabolic comorbidities? *Cytokine Growth Factor Rev* 2020.
- [74] Khanmohammadi S, Rezaei N. Role of toll-like receptors in the pathogenesis of COVID-19. *J Med Virol* 2021;93(5):2735–9.
- [75] Blanco-Melo D, Nilsson-Payant BE, Liu W-C, Møller R, Panis M, Sachs D, Albrecht RA, et al. SARS-CoV-2 launches a unique transcriptional signature from in vitro, ex vivo, and in vivo systems. *BioRxiv* 2020.
- [76] Casadevall A, Pirofski L-a. In fatal COVID-19, the immune response can control the virus but kill the patient. *Proc Natl Acad Sci* 2020;117(48):30009–11.
- [77] Kim Y-M, Shin E-C. Type I and III interferon responses in SARS-CoV-2 infection. *Experiment Mol Med* 2021;53(5):750–60.
- [78] Huang H, Zheng J, Shen N, Wang G, Zhou G, Fang Y, Lin J, Zhao J. Identification of pathways and genes associated with synovitis in osteoarthritis using bioinformatics analyses. *Sci Rep* 2018;8(1):1–9.
- [79] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43(7). e47.
- [80] Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* 2015;43(D1):D447–52.
- [81] Zou J, Huang R-Y, Jiang F-N, Chen D-X, Wang C, Han Z-D, Liang Y-X, Zhong W-D. Overexpression of TPX2 is associated with progression and prognosis of prostate cancer. *Oncology Lett* 2018;16(3):2823–32.
- [82] Ji Y, Yin Y, Zhang W. Integrated bioinformatic analysis identifies networks and promising biomarkers for hepatitis B virus-related hepatocellular carcinoma. *Int J Genomics* 2020;2020.
- [83] Auwal MR, Rahman MR, Gov E, Shahjaman M, Moni MA. Bioinformatics and machine learning approach identifies potential drug targets and pathways in COVID-19. *Brief Bioinform* 2021.
- [84] Song Y, Tang W, Li H. Identification of KIF4A and its effect on the progression of lung adenocarcinoma based on the bioinformatics analysis. *Biosci Rep* 2021;41(1). BSR20203973.
- [85] Rahman MH, Sarkar B, Islam MS, Abdullah MI. Discovering biomarkers and pathways shared by Alzheimer's disease and parkinson's disease to identify novel therapeutic targets. *Int J Eng Res Technol* 2020;6.
- [86] Fagone P, Ciurleo R, Lombardo SD, Iacobello C, Palermo CI, Shoenfeld Y, Bendtzen K, Bramanti P, Nicoletti F. Transcriptional landscape of SARS-CoV-2 infection dismantles pathogenic pathways activated by the virus, proposes unique sex-specific differences and predicts tailored therapeutic strategies. *Autoimmun Rev* 2020;19(7):102571.
- [87] Li C-Y. Can glycine mitigate COVID-19 associated tissue damage and cytokine storm? *Radiat Res* 2020;194(3):199–201.
- [88] Shakoor H, Feehan J, Mikkelsen K, Al Dhaheri AS, Ali HI, Platat C, Ismail LC, Stojanovska L, Apostolopoulos V. Be well: A potential role for vitamin B in COVID-19. *Maturitas* 2021;144:108–11.
- [89] Rahman MH, Peng S, Chen C, Moni MA, et al. Genetic effect of type 2 diabetes to the progression of neurological diseases. *BioRxiv* 2018;480400.
- [90] Desbarats J. Pyridoxal 5'-phosphate to mitigate immune dysregulation and coagulopathy in COVID-19. 2020.