Received: 1 July 2021 | Revised: 14 July 2021 | Accepted: 18 July 2021

DOI: 10.1049/syb2.12034

IET Systems Biology

The Institution of Engineering and Technology WILEY

ORIGINAL RESEARCH PAPER

# Comprehensive investigation of RNA-sequencing dataset reveals the hub genes and molecular mechanisms of coronavirus disease 2019 acute respiratory distress syndrome

**Wangsheng Deng**[1] | **Jiaxing Zeng**[1] | **Shunyu Lu**[2] | **Chaoqian Li**[3]

[1]Emergency Department, The Fourth Affiliated Hospital of Guangxi Medical University, Liuzhou, China

[2]Department of Pharmacy, Affiliated Tumor Hospital of Guangxi Medical University, Nanning, China

[3]Emergency Department, The First Affiliated Hospital of Guangxi Medical University, Nanning, China

**Correspondence**

Wangsheng Deng, Emergency Department, The Fourth Affiliated Hospital of Guangxi Medical University, Liu Shi Road 1, Liuzhou, 545005, Guangxi Zhuang Autonomous Region, China.
Email: dengwslzgryy@163.com
Chaoqian Li, Emergency Department, The First Affiliated Hospital of Guangxi Medical University, Shuang Yong Road 6, Nanning, 530021, Guangxi Zhuang Autonomous Region, China.
Email: Leechaoqian@163.com

**Abstract**

The goal of this study is to reveal the hub genes and molecular mechanisms of the coronavirus disease 2019 (COVID-19) acute respiratory distress syndrome (ARDS) based on the genome-wide RNA sequencing dataset. The RNA sequencing dataset of COVID-19 ARDS was obtained from GSE163426. A total of 270 differentially expressed genes (DEGs) were identified between COVID-19 ARDS and control group patients. Functional enrichment analysis of DEGs suggests that these DEGs may be involved in the following biological processes: response to cytokine, G-protein coupled receptor activity, ionotropic glutamate receptor signalling pathway and G-protein coupled receptor signalling pathway. By using the weighted correlation network analysis approach to analyse these DEGs, 10 hub DEGs that may play an important role in COVID-19 ARDS were identified. A total of 67 potential COVID-19 ARDS targetted drugs were identified by a complement map analysis. Immune cell infiltration analysis revealed that the levels of T cells CD4 naive, T cells follicular helper, macrophages M1 and eosinophils in COVID-19 ARDS patients were significantly different from those in control group patients. In conclusion, this study identified 10 COVID-19 ARDS-related hub DEGs and numerous potential molecular mechanisms through a comprehensive analysis of the RNA sequencing dataset and also revealed the difference in immune cell infiltration of COVID-19 ARDS.

**KEYWORDS**

acute respiratory distress syndrome, coronavirus disease 2019, GSE163426, hub genes, molecular mechanism

## 1 | INTRODUCTION

Coronavirus disease 2019 (COVID-19), named by the World Health Organization, refers to pneumonia caused by the 2019 novel coronavirus infection. The International Committee on Classification of Viruses also named this new type of coronavirus 'SARS-CoV-2' (Severe Acute Respiratory Syndrome Coronavirus 2) [1]. At present, there are more than 100 million patients diagnosed with COVID-19 worldwide, and more than 3 million patients have died. There are still 30 million patients with COVID-19. The clinical manifestations of COVID-19-infected pneumonia patients include fever, fatigue, and dry cough, which are its main manifestations [2]. Upper respiratory symptoms such as nasal congestion and runny nose are rare, and hypoxia is present [3]. About half of the patients have difficulty breathing more than a week later. In severe cases, it leads to acute respiratory distress syndrome (ARDS), septic shock, metabolic acidosis, and coagulation dysfunction [4, 5]. ARDS is also the main cause of death for COVID-19 patients. However, it is still unclear which people are more likely to develop ARDS, and which are the related factors in patients who develop ARDS that lead to death. Therefore, we urgently need a better understanding of COVID-19 ARDS.

Wangsheng Deng and Jiaxing Zeng contributed equally to this work.

ARDS can originate on the gaseous side or the vascular side of the alveoli. ARDS can occur in patients with severe COVID-19 and causes unique lung damage [6, 7]. COVID-19 is a systemic disease that mainly damages the vascular endothelium, and this vascular damage may require treatment that is different from the conventional approach to ARDS management [7]. Immune activation may also be an important reason for patients to develop ARDS [8–10]. Therefore, it is urgent to understand the difference between COVID-19 ARDS and conventional ARDS so that we can develop more targetted treatment strategies. In addition to clinical manifestations, we also need to further understand its potential molecular mechanisms and potential hub genes. The goal of our study is to reveal the hub genes and molecular mechanisms of COVID-19 ARDS based on the genome-wide RNA sequencing dataset as well as the difference of immune cell infiltration in COVID-19 ARDS.

## 2 | MATERIALS AND METHODS

### 2.1 | Data processing

The raw count expression profiling by the high-throughput sequencing dataset of GSE163426 was downloaded from the Gene Expression Omnibus (GEO) database (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163426). The GSE163426 dataset contains a high-throughput sequencing dataset for 52 human tracheal aspirates samples, including 15 samples from COVID-19 ARDS patients. The other 37 cases were derived from subjects with ARDS due to other causes and mechanically ventilated controls without evidence of pulmonary disease, which we identified as control groups in the current study. The original count data is normalised by edgeR, and differentially expressed genes (DEGs) are screened [11, 12]. The DEGs' standard is defined as $|\log_2 \text{fold change (FC)}| > 1$, $P < 0.05$ and a false discover rate (FDR) $< 0.05$ [13, 14]. All the datasets in the present study were downloaded from the GEO public database and do not involve any animal or human study. Therefore, no additional ethics committee approval is required.

### 2.2 | Function enrichment and gene interaction analysis

In order to further understand the differences in molecular mechanisms between COVID-19 ARDS patients and control patients, we used the database for Annotation, Visualisation and Integrated Discovery (DAVID) v6.8 (https://david.ncifcrf.gov/home.jsp) to enrich the functions of these DEGs [15, 16]. We also used gene set enrichment analysis (GSEA: http://software.broadinstitute.org/gsea/index.jsp) to further explore the mechanisms between COVID-19 ARDS patients and control group patients [17, 18]. GSEA results were set as follows: $|\text{Normalised Enrichment Score (NES)}| > 1$, $P < 0.05$ and FDR $< 0.25$. In order to further understand the interaction relationship between these genes, a gene-gene interaction analysis was conducted on these DEGs by using GeneMANIA

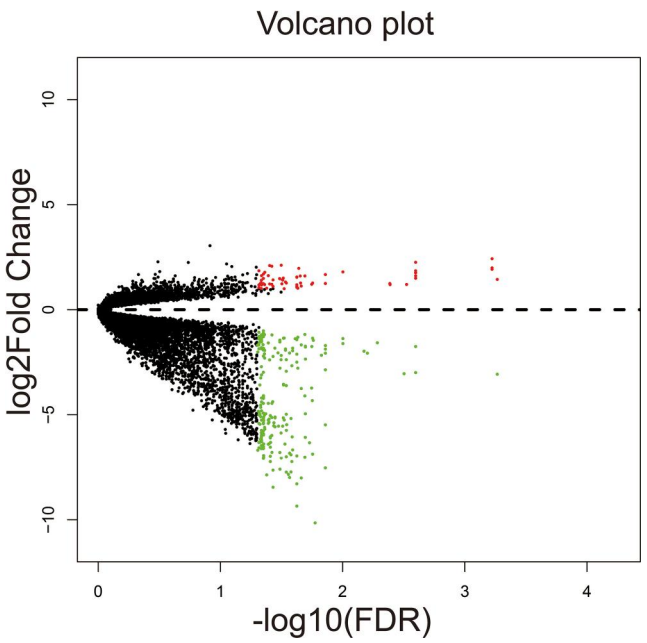


**FIGURE 1** Volcano plot of differentially expressed gene between COVID-19 acute respiratory distress syndrome and control group patients

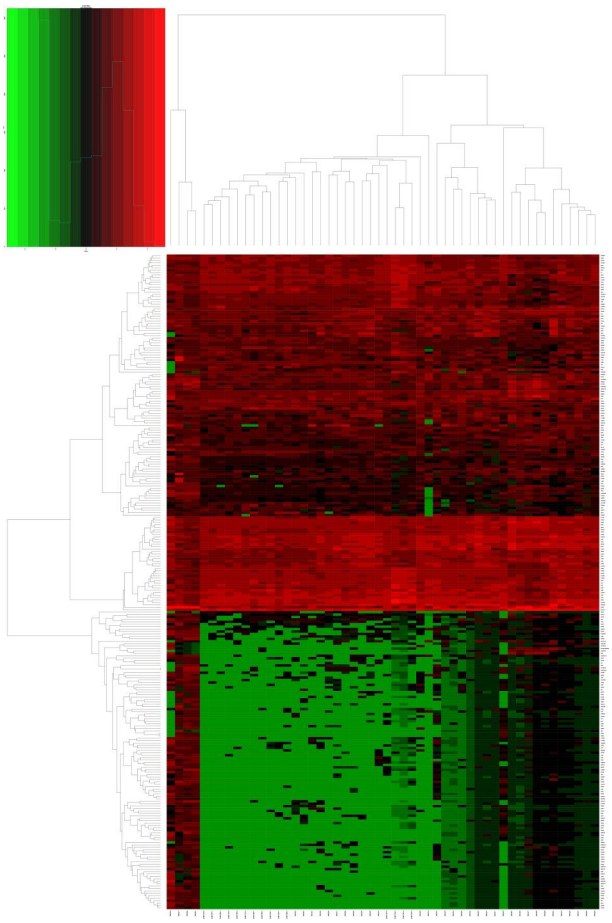

**FIGURE 2** Heat map of differentially expressed gene between COVID-19 acute respiratory distress syndrome and control group patients

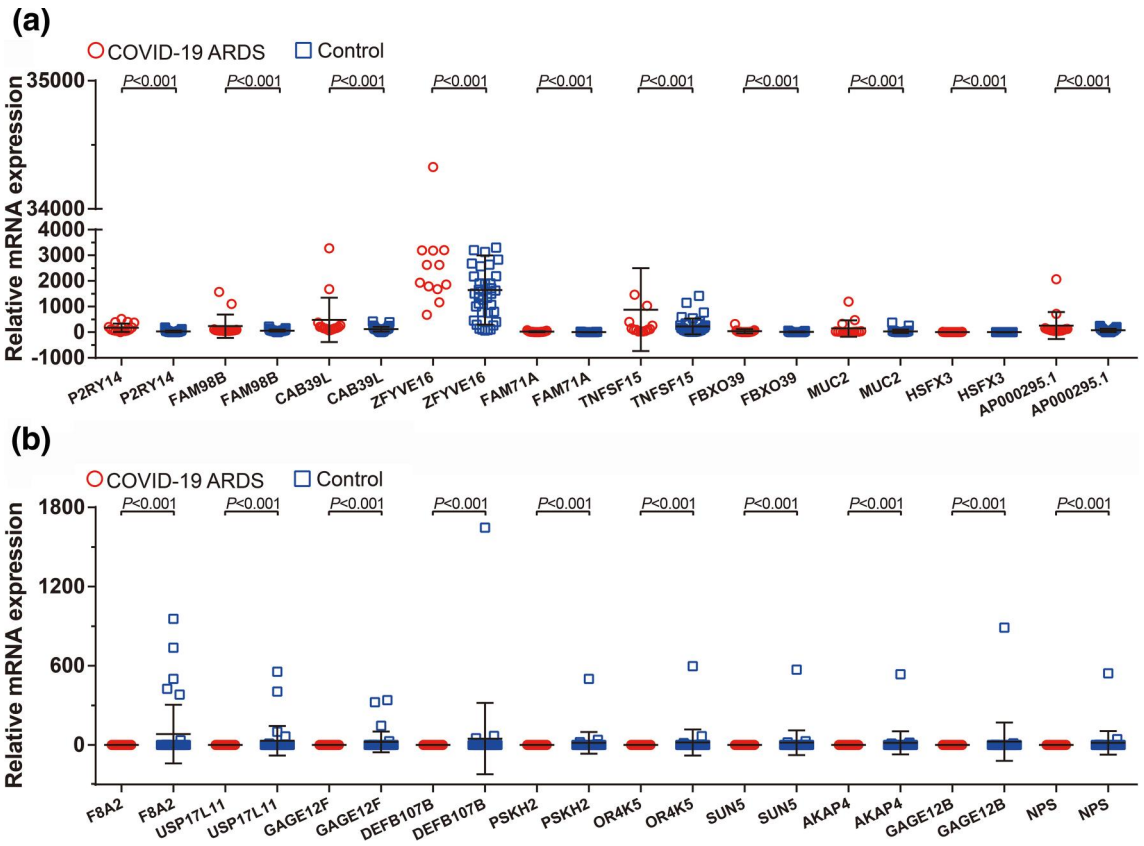


**FIGURE 3** Scatter plot of the top 10 up-regulated differentially expressed genes (DEGs) and down-regulated DEGs: (a) scatter plot of up-regulated DEGs; (b) scatter plot of down-regulated DEGs

(http://genemania.org/) [19, 20] and STRING (https://string-db.org) online tools [21–25]. Subsequently, we further verified the gene-gene interaction relationship between these DEGs using the RNA sequencing dataset of COVID-19 ARDS patients, and we used the weighted gene co-expression network analysis (WGCNA) approach to verify the co-expression interaction relationship between these DEGs [26]. Hub genes in COVID-19 ARDS were determined according to the WGCNA network by using the degree. The concept of degree in the WCGAN network is defined as the connectivity degree of each gene, which is the sum of the edge attributes of the genes connected to it. We then used these DEGs to predict potential targetted drugs based on the Complement Map (CMAP, https://portals.broadinstitute.org/cmap/) database [27, 28]. Small molecule drugs with a negative mean connective score and $P < 0.05$ were considered statistically significant and had inhibitory effects. The chemical structures of these small molecule compounds were obtained from the PubChem (https://pubchem.ncbi.nlm.nih.gov) database [29–33].

## 2.3 | Immune cell infiltration analysis

In order to further understand the relationship between COVID-19 ARDS and immune cells, we used the established computational approach (CIBERSORT) package to analyse and compare the infiltration of immune cells between COVID-19 ARDS and control group patients, so as to understand the immunological characteristics of COVID-19 ARDS patients [34, 35]. A violin plot and heat maps were drawn using the ggplot2 package.

## 2.4 | Statistical analysis

A scatter plot of the mRNA expression was obtained using an independent sample *t* test. The Wilcoxon test was used for comparison of immune cells between COVID-19 ARDS patients and control group patients. All statistical analyses were carried out using SPSS version 22.0 and the R platform that uses version R4.0.2. $P < 0.05$ was considered to be statistically significant.

# 3 | RESULTS

## 3.1 | Screening of DEGs

We used edgeR to screen the differentially expressed genes between COVID-19 ARDS and control group patients. We obtained a total of 270 DEGs, of which 208 DEGs were down-regulated and 62 DEGs were up-regulated in the

COVID-19 ARDS group. In the COVID-19 ARDS group, the most significantly up-regulated DEG was the purinergic receptor P2Y14 (P2RY14, $\log_2$FC = 2.42, $P = 1.05 \times 10^{-7}$, FDR = 0.0006, Table S1, Figure 1), and the most significantly down-regulated DEG was the coagulation factor VIII associated 2 (F8A2, $\log_2$FC = −10.15, $P = 2.98 \times 10^{-5}$, FDR = 0.017, Table S1, Figure 1). Subsequently, the expression level distribution of these DEGs in the two groups of samples is shown in Figure 2 in the form of a heat map. In addition, the top ten up-regulated genes and the top ten down-regulated genes are also shown in the form of a scatter plot in Figure 3a,b and Table 1.

**TABLE 1** The top 10 up-regulated and down-regulated DEGs

| ID | $Log_2FC$ | *P*-value | FDR |
|---|---|---|---|
| F8A2 | −10.15208739 | 2.98E−05 | 0.01676952 |
| DEFB107B | −9.346607155 | 9.31E−05 | 0.023688991 |
| GAGE12B | −8.454216021 | 0.000273381 | 0.037101321 |
| GAGE12F | −8.289891203 | 8.96E−05 | 0.023688991 |
| USP17L11 | −8.011558919 | 6.46E−05 | 0.021766209 |
| OR4K5 | −7.985552841 | 0.000125822 | 0.02717469 |
| NPS | −7.860072646 | 0.000380604 | 0.041830422 |
| SUN5 | −7.837855053 | 0.000130093 | 0.027774149 |
| AKAP4 | −7.729202685 | 0.000141693 | 0.028864229 |
| PSKH2 | −7.719765307 | 0.000117433 | 0.026434761 |
| ZFYVE16 | 1.8512934 | 1.62E−06 | 0.002524835 |
| AP000295.1 | 1.852209732 | 0.000700273 | 0.048340144 |
| FAM98B | 1.924358675 | 1.30E−07 | 0.000597912 |
| TNFSF15 | 1.966295459 | 7.26E−05 | 0.0228549 |
| CAB39L | 1.992787287 | 1.61E−07 | 0.000597912 |
| MUC2 | 2.065428132 | 0.000298573 | 0.037984264 |
| HSFX3 | 2.096915061 | 0.000334409 | 0.039562487 |
| FBXO39 | 2.113460379 | 0.000192157 | 0.031771397 |
| FAM71A | 2.253072759 | 1.63E−06 | 0.002524835 |
| P2RY14 | 2.423663806 | 1.05E−07 | 0.000597912 |

Abbreviations: DEGs, differentially expressed genes; FC, fold change; FDR, false discover rate.

## 3.2 | Function enrichment and gene interaction analysis

Then, we explored the function of these DEGs. Gene Ontology (GO) term analysis suggested that these DEGs are significantly enriched in chemical synaptic transmission, calcium ion binding, proteinaceous extracellular matrix, movement of cell or subcellular component, response to mechanical stimulus, secretory granule, G-protein coupled receptor activity, extracellular-glutamate-gated ion channel activity, response to cytokine, lysosome, platelet alpha granule lumen, ionotropic glutamate receptor signalling pathway and G-protein coupled receptor signalling pathway (Figure 4). The Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis suggested that these genes are significantly enriched in neuroactive ligand-receptor interaction (hsa04080) and nicotine addiction (hsa05033) (Figure 4). In order to further understand the molecular mechanism differences between the two groups of patients, GSEA was also used for the analysis. When we used c5 as the reference gene set (c5.all.v7.4.symbols.gmt), none of the enrichment results reached statistical significance. When we used c2 as the reference gene set (c2.all.v7.4.symbols.gmt, Table S2), only one enrichment result reached statistical significance (Figure 5, Table S3). Through the GSEA analysis, we found that there was a statistically significant difference in

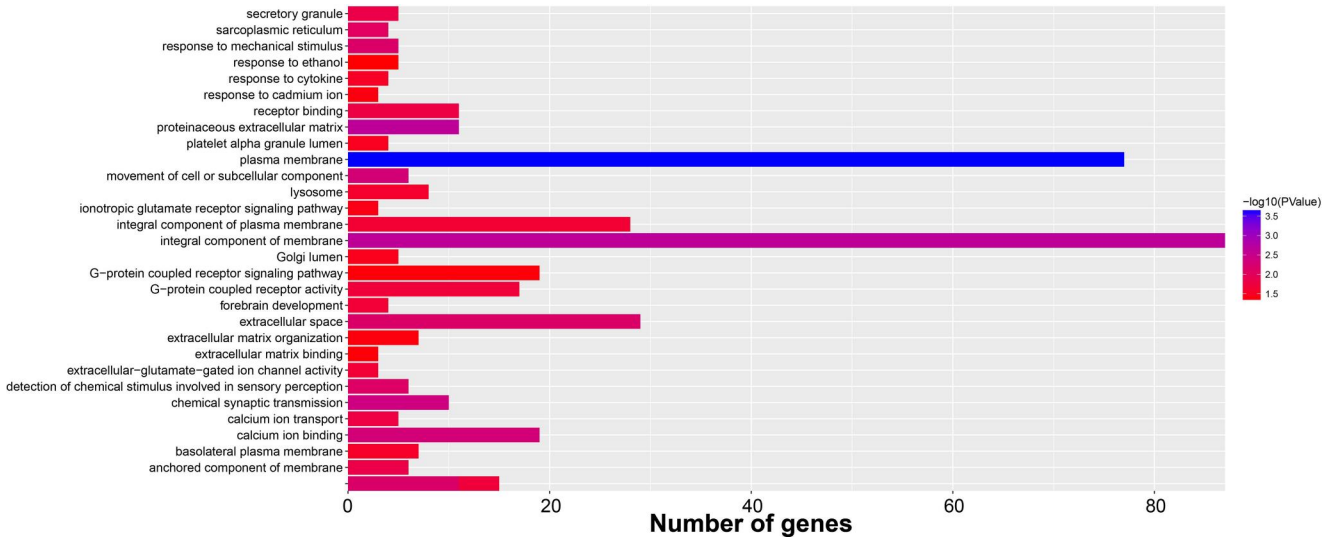


**FIGURE 4** Bar chart of Gene Ontology and Kyoto Encyclopedia of Genes and Genomes enrichment analysis results of differentially expressed genes between COVID-19 acute respiratory distress syndrome and control group patients

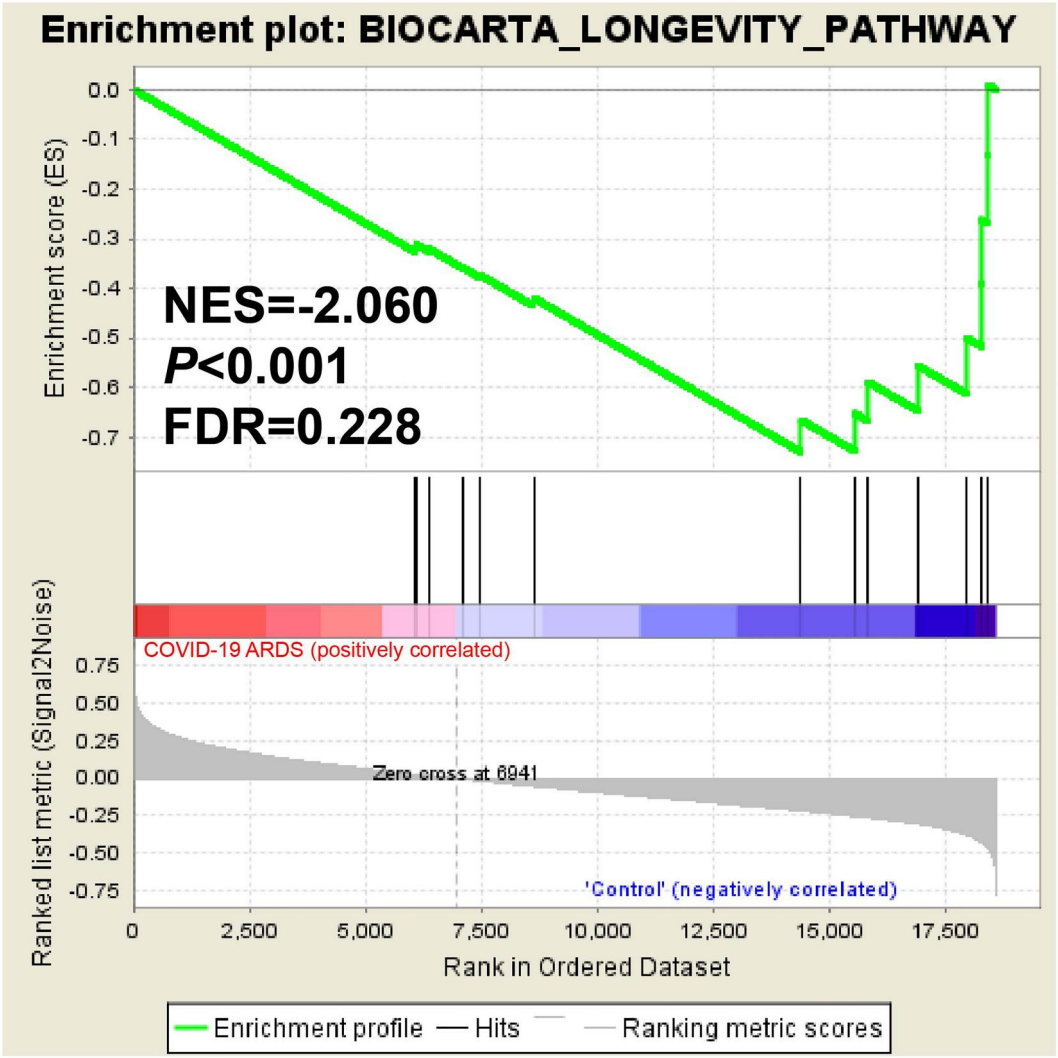


**FIGURE 5** Gene set enrichment analysis results between COVID-19 acute respiratory distress syndrome and control group patients using the c2 reference gene set

longevity pathway (NES = −2.060, $P < 0.001$, FDR = 0.228, Figure 5) between the two groups.

In order to understand the gene-gene interaction relationships among these DEGs, we used the STRING online analysis tool to construct the DEGs gene–gene interaction network (Figure 6). We found that there were complex gene–gene interaction relationships among these DEGs, including co-expression, experimentally determined etc. Subsequently, we also used GeneMANIA online analysis tool to verify the co-expression gene–gene interaction relationship of these DEGs (Figure 7). Based on the above two online gene interaction analysis tools, we found that there was a complex co-expression interaction between these DEGs. In order to further understand the co-expression relationship of these DEGs in COVID-19 ARDS patients, we further explored the co-expression interaction relationship of these DEGs by the WGCNA method. In the WGCNA analysis, according to the data distribution, we choose soft threshold beta as four (Figure 8a,b).

In the WGCNA network modular analysis of these DEGs, we differentiated a cluster of genes defined as turquoise modules (Figure 8c,d). In the WGCNA network, a total of 178 DEGs with 529 edges were included to construct a weighted co-expression gene-gene interaction network (Figure 9). Among them, 47 DEGs belong to the turquoise module, and the remaining genes belong to the grey module, that is, these DEGs do not belong to any module. Then, we performed a functional enrichment of the DEGs of the turquoise module. We did not obtain statistically significant KEGG results. We obtained 1 GO TERM analysis result. The GO term analysis suggested that these turquoise module DEGs are significantly enriched in positive regulation of T cell tolerance induction. Subsequently, hub DEG screening was performed using the WGCNA co-expression interaction network. The importance of DEG in the network is evaluated by degree. Through analysis, we found that the DEGs with the highest degree in the WGCNA network were Erbb2 interacting protein (ERBIN), pericentriolar material 1 (PCM1), zinc finger

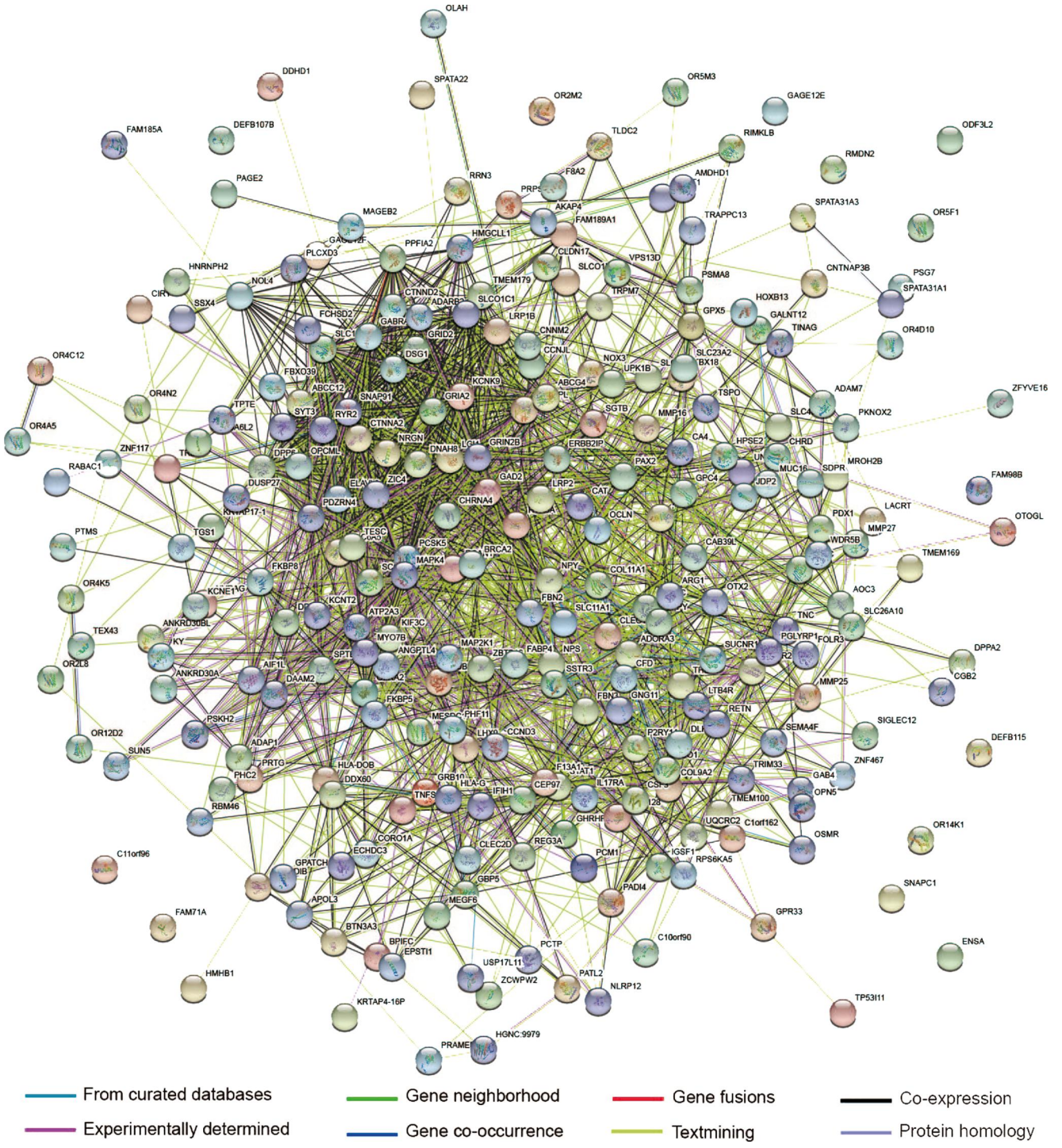


**FIGURE 6** Differentially expressed genes interaction network generated by the STRING online analysis tool

FYVE-Type containing 16 (ZFYVE16) and small nuclear RNA activating complex polypeptide 1 (SNAPC1). Both of them have degrees of 33 in the WGCNA network (Table 2). In the WGCNA network, the second ranking DEGs for the degree are calcium binding protein 39 Like (CAB39L), occludin (OCLN), UV radiation resistance associated (UVRAG), major histocompatibility complex, class I, G (HLA-G), C-type lectin domain family 2 member D (CLEC2D) and zinc finger protein 117 (ZNF117), all of which have 32° (Table 2). Based on the WGCNA network of DEG in patients with COVID-19 ARDS, we identified a total of 10 hub DEGs that are mentioned above and may have molecular characteristics specific to COVID-19 ARDS patients. Among the 10 hub DEGs, ZFYVE16 and CAB39L are the top ten up-regulated DEGs in

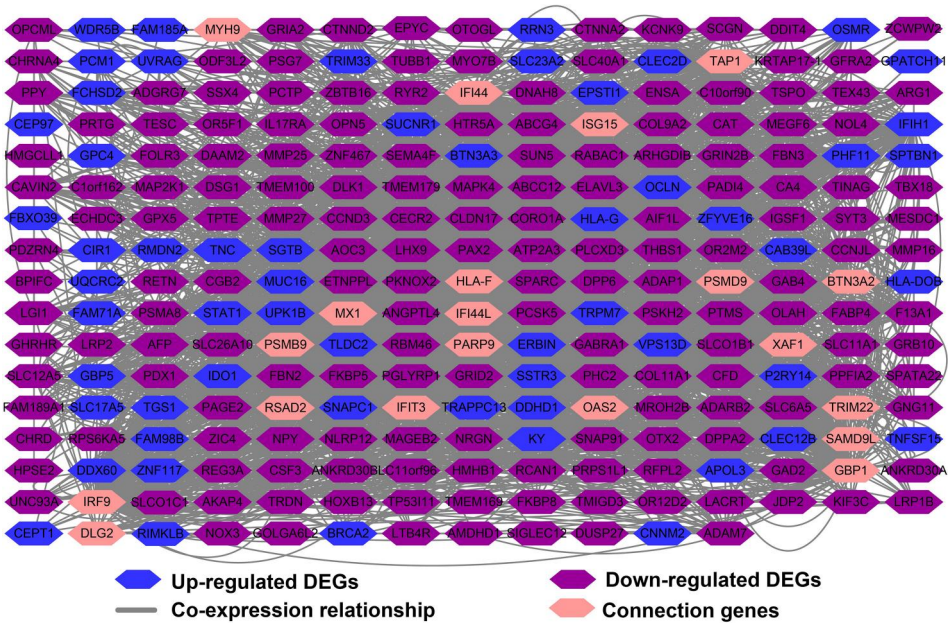


**FIGURE 7** Differentially expressed genes co-expression interaction network generated by the GeneMANIA online analysis tool

fold change. Through the CMAP analysis, we obtained a total of 67 small molecule drugs with a *P*-value <0.05 and a negative mean connective score (Table S4). Among them, there were 16 small molecule drugs with a negative mean connective score less than −0.6, and they were camptothecin, alsterpaullone, GW-8510, doxorubicin, irinotecan, H-7, 0175029-0000, 8-azaguanine, Gliclazide, Meticrane, 1,4-chrysenequinone, 6-azathymine, azacitidine, ellipticine, mitoxantrone and apigenin (Figure 10a–n, Table 3).

## 3.3 | Immune cell infiltration analysis

We all know that COVID-19 is a viral infectious disease, which will inevitably cause changes in the body's immune system after infecting the human body. Therefore, we feel it is necessary to compare the immune cells between COVID-19 ARDS patients and control patients. The CIBERSORT software package can calculate the proportion of 22 immune cells in the human body through the RNA sequencing dataset. The 22 types of human immune cells include B cells naive, B cells memory, plasma cells, T cells CD8, T cells CD4 naive, T cells CD4 memory resting, T cells CD4 memory activated, T cells follicular helper, T cells regulatory (Tregs), T cells gamma delta, NK cells resting, NK cells activated, monocytes, macrophages M0, macrophages M1, macrophages M2, dendritic cells resting, dendritic cells activated, mast cells resting, mast cells activated, eosinophils, and neutrophils. By drawing a histogram of the proportions of 22 immune cells in the two groups of patients, we found that B cells and mast cells accounted for a larger proportion in COVID-19 ARDS patients, followed by macrophages (Figure 11). Then, we showed the activation levels of 22 immune cells through heat maps. We found that the activation levels of mast cells activated and neutrophils were much higher than those of other immune cells in the two groups of patients (Figure 12). In addition, we also compared the differences in immune cells between the two groups of patients. The violin plot revealed that the level of T cells CD4 naive in COVID-19 ARDS patients was significantly lower than that in the control group patients (Figure 13). The levels of T cells follicular helper, macrophages M1 and eosinophils in COVID-19 ARDS patients were significantly higher than those in the control group patients (Figure 13).

# 4 | DISCUSSION

A multicentre study suggests that ARDS may serve as an important clinical subtype for the classification of different treatment strategies and outcomes in patients with COVID-19. They divided COVID-19 patients into four sub-phenotypes. One of the subtypes is characterised by high C-reactive protein, early need for mechanical ventilation, and the highest rate of ARDS, and has different outcomes and treatment strategies [36]. Patients with ARDS associated with COVID-19 often have different clinical characteristics than those with ARDS due to other causes, and case-control studies have shown that ARDS is a major cause of mortality in patients with COVID-19 [37]. Therefore, there is an urgent need to understand the molecular mechanism and core genes of COVID-19-related ARDS, so as to develop targetted drugs for the diagnosis of COVID-19-related ARDS. There is a need to develop new drugs against COVID-19 and reduce the mortality. Through this research,

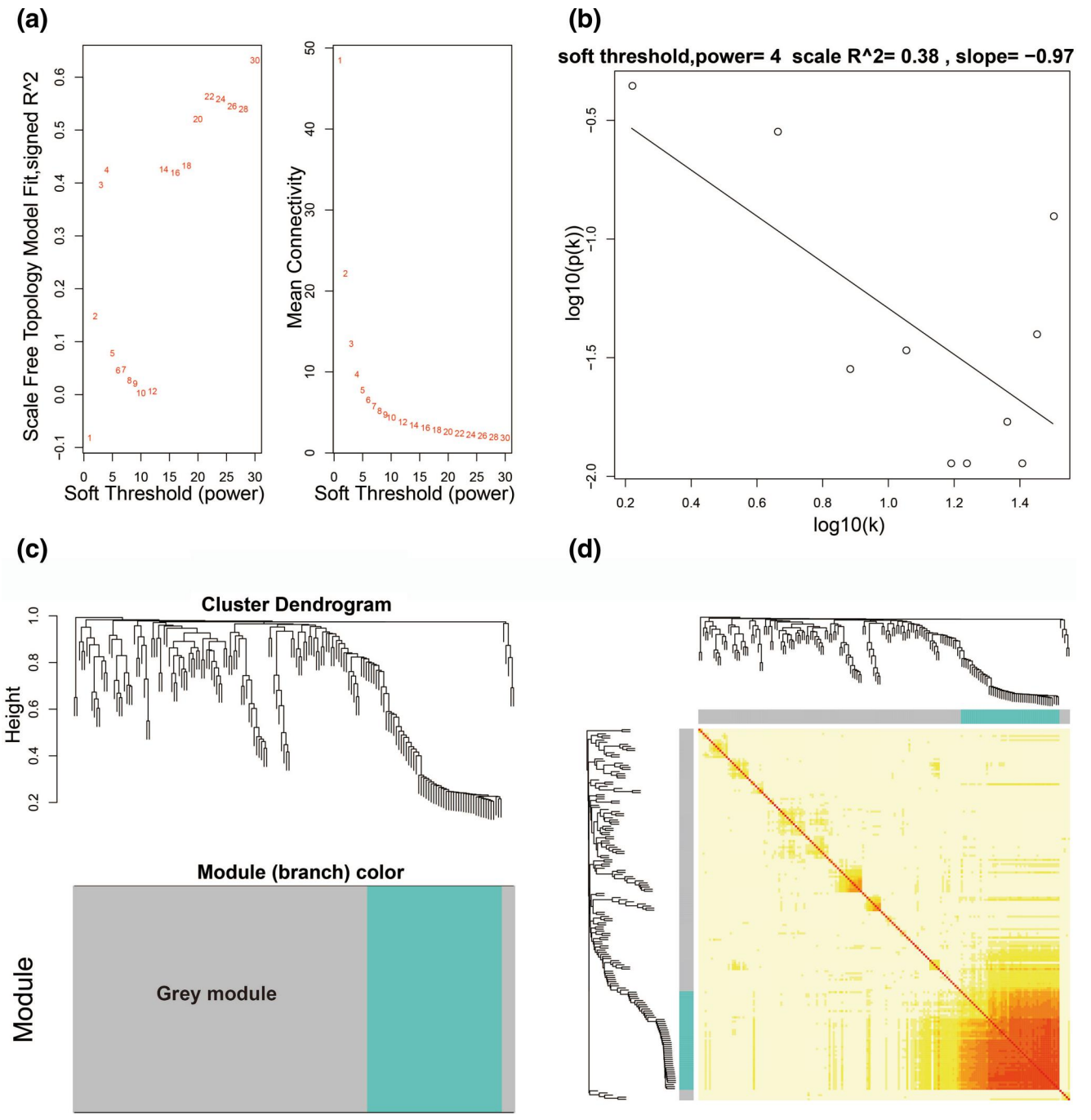


**FIGURE 8** Weighted gene co-expression network analysis analysis results of differentially expressed genes (DEGs) between COVID-19 acute respiratory distress syndrome and control group patients: (a) soft threshold screening plot; (b) scale-free topology plot; (c) clustering dendrograms of DEGs; (d) TOM plot

we have screened the molecular mechanisms of multiple COVID-19-related ARDS hub genes through whole-genome RNA sequencing data and screened a batch of potential targetted small molecule drugs. Previous studies have also used RNA sequencing methods to perform functional enrichment on specimens of patients with COVID-19 ARDS to understand the molecular mechanisms of this disease, in order to develop new therapeutic drugs or strategies. Wu et al. analysed the RNA-sequence dataset of ACE2-expressed-A549 and NHBE cells with SARS-CoV-2 infection and found the genes associated with it by bioinformatics enrichment analysis with immune-related pathways, responses of host cells after an intracellular infection, steroid hormone biosynthesis, receptor signalling, and the complement system, which are closely related to a COVID-19 infection [38]. Amira Mohammed et al. used staphylococcal enterotoxin B to induce ARDS a mouse model through single-cell RNA sequencing and transcriptome

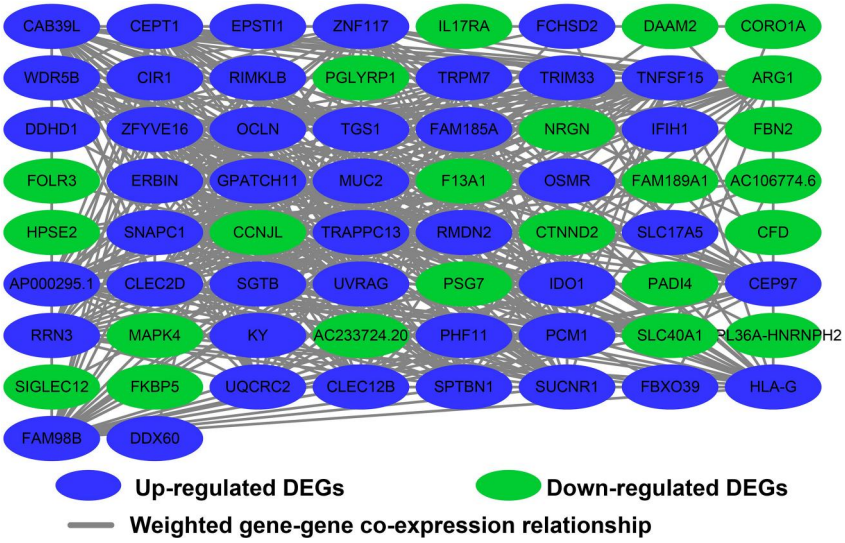


**FIGURE 9** Weighted gene co-expression network analysis co-expression networks of differentially expressed genes between COVID-19 acute respiratory distress syndrome and control group patients

**TABLE 2** Hub DEGs list of COVID-19 ARDS patients in WGCNA network

| ID | $log_2FC$ | *P*-value | FDR | Degree in WGCNA |
|---|---|---|---|---|
| ERBIN | 1.758416 | 1.58E−06 | 0.002525 | 33 |
| PCM1 | 1.495209 | 8.43E−07 | 0.002525 | 33 |
| ZFYVE16 | 1.851293 | 1.62E−06 | 0.002525 | 33 |
| SNAPC1 | 1.410232 | 0.000183 | 0.03095 | 33 |
| CAB39L | 1.992787 | 1.61E−07 | 0.000598 | 32 |
| OCLN | 1.603824 | 6.94E−05 | 0.022209 | 32 |
| UVRAG | 1.105065 | 8.43E−05 | 0.023689 | 32 |
| HLA-G | 1.418098 | 0.000274 | 0.037101 | 32 |
| CLEC2D | 1.178746 | 0.000572 | 0.046022 | 32 |
| ZNF117 | 1.220263 | 0.000654 | 0.047449 | 32 |

Abbreviations: ARDS, acute respiratory distress syndrome; COVID-19, coronavirus disease 2019; DEGs, differentially expressed genes; FC, fold change; FDR, false discover rate; WGCNA, weighted gene co-expression network analysis.

analysis and found that ARDS is involved in the apoptosis of immune cells in the mitochondrial pathway. The gene expression dataset of bronchoalveolar lavage fluid of COVID-19 patients shows that there are some similarities in the cytokine and apoptosis genes in the ARDS sequencing results induced by Staphylococcal enterotoxin B [39]. Single cell RNA sequencing of bronchoalveolar lavage fluid from COVID-19 patients revealed significant amplification of transmembrane 4 L six family member 1 (TM4SF1)+ and keratin 5 (KRT5)+ lung progenitor cells in severe COVID-19 patients, suggesting a synergistic role of lung progenitor cells in the prevention and supplementation of alveolar loss in COVID-19 patients [40].

Among the selected hub genes, we found that HLA-G and OCLN genes were reported to be related to COVID-19 in previous studies. Lung progenitor cells of KLT5 + can differentiate into HOP homeobox (HOPX)+ OCLN+ alveolar barrier cells to supplement alveolar loss after SARS-CoV-2 infection, restore epithelial barrier and effectively prevent infiltration of inflammatory cells [40]. Zhang et al. found that the proportion of HLA-G+ T cells showed a high-low-high trend by detecting the expression level of HLA-G in the peripheral blood of COVID-19 patients in the convalescent stage of severe pneumonia [41]. The researchers also prospected the application of HLA-G in COVID-19, suggesting that the targetted treatment of HLA-G may be effective against COVID-19 [42]. Among the top drugs screened by CMAP, we also found multiple drugs that have been reported to have antiviral effects in previous studies. Previous studies based on RNA sequencing data and bioinformatics analysis found that camptothecin had a potential antiviral effect against SARS-CoV-2 [43, 44]. The potential mechanism by which Camptothetin acts against SARS-CoV-2 may be through the possible blocking of the spiking glycoprotein interaction with the angiotensin-converting enzyme 2 (ACE2) receptor found on host cells [43]. Fadhl M Alakwaa have developed a bioinformatics pipeline for screening COVID-19 targetted therapeutics by using bioinformatics analysis methods [44]. Qazi et al. used Autodock to conduct computer simulations of the antibiotic class of viral proteases and peptidases and found that doxorubicin can be used to inhibit viral proteases that may prevent entry into host cells to control the COVID-19 disease [45]. Bragi Lovetrue built an accurate COVID-19 disease model. The combination therapy of Irinotecan + Etoposide discovered by AI may be an effective method to protect critically ill COVID-19 patients from death caused by COVID-19-specific cytokine storm [46]. Prabhat Pratap Singh Tomar et al. screened two compounds inhibiting E protein based on the library of ion channel blockers—gliclazide and metadine, which may exert an antiviral effect by inhibiting the E protein of SARS-CoV-2 [47]. Kiran Bharat Lokhande et al. based on Selleckchem Inc. Through molecular docking and molecular dynamics simulation, it was found that Mitoxantrone can be used as a potential inhibitor of SARS-CoV-2 MPRO and has a

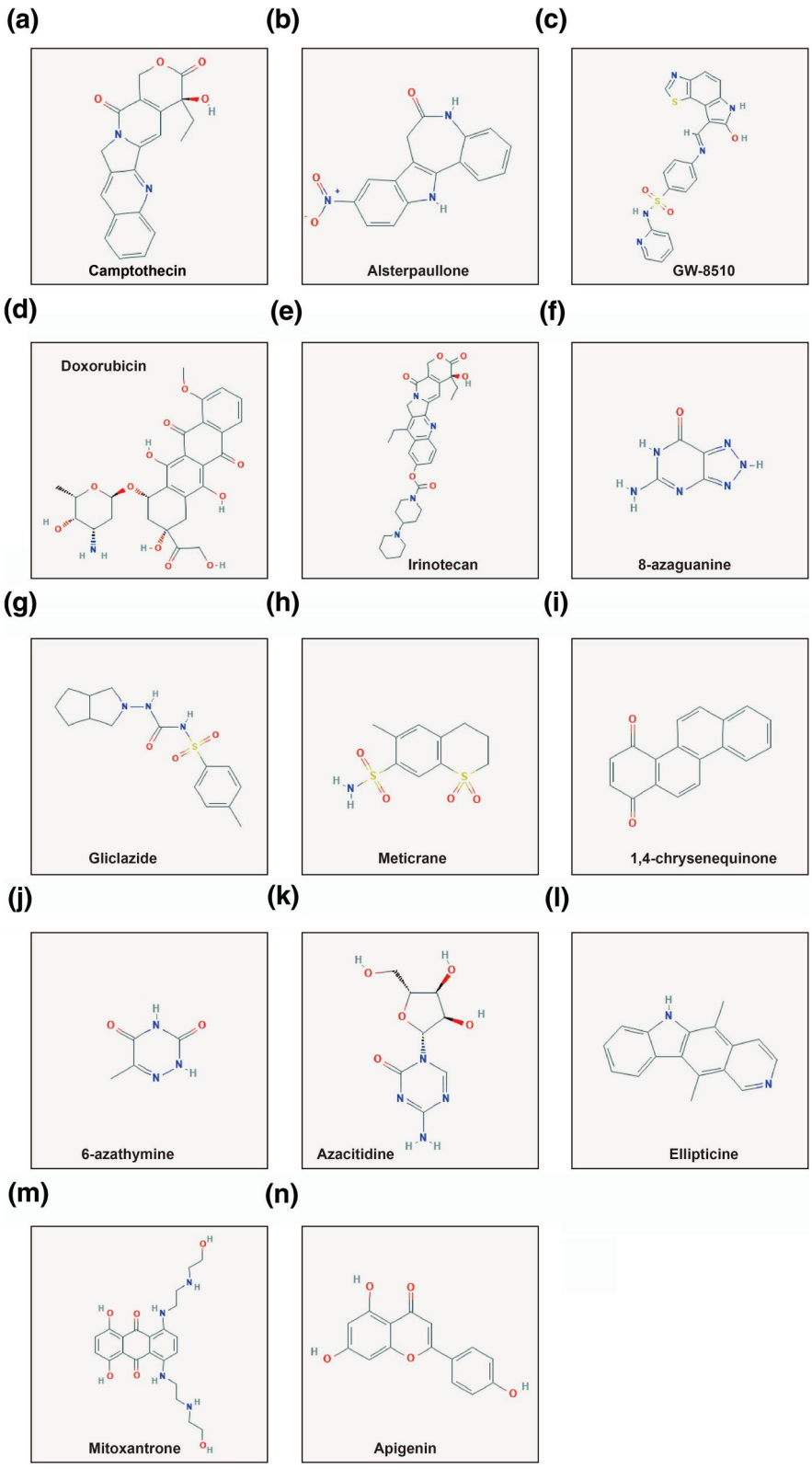


**FIGURE 10** The chemical structure of these small molecule compounds that is identified by CMap analysis with the mean connective score that is less than −0.6: the compound structure of (a) camptothecin; (b) alsterpaullone; (c) GW-8510; (d) doxorubicin; (e) irinotecan; (f) 8-azaguanine; (g) gliclazide; (h) meticrane; (i) 1,4-chrysenequinone; (j) 6-azathymine; (k) azacitidine; (l) ellipticine; (m) mitoxantrone; and (n) apigenin

potential antiviral effect [48]. Based on the bioinformatics analysis of the present study, we screened the hub genes and targetted drugs for COVID-19 ARDS, which had some overlap and consistency with the previous reports. At the same time, we also identified several hub genes and drugs related to COVID-19 ARDS that have not been reported in previous studies. Based on a review of the literature, it was previously reported that the new coronavirus is closely related to the imbalance of immune cells. In this study, by analysing the difference in the degree of infiltration of immune cells, we found

**TABLE 3** CMap analysis results with mean connective score less than −0.6

| CMap name | Mean connective score | *n* | Enrichment | *P*-value | Specificity | Percent non-null |
|---|---|---|---|---|---|---|
| Camptothecin | −0.859 | 3 | −0.998 | <0.001 | <0.05 | 100 |
| Alsterpaullone | −0.811 | 3 | −0.988 | <0.001 | 0.0214 | 100 |
| GW-8510 | −0.797 | 4 | −0.907 | 0.0001 | 0.0992 | 100 |
| Doxorubicin | −0.739 | 3 | −0.859 | 0.00563 | 0.1278 | 100 |
| Irinotecan | −0.7 | 3 | −0.861 | 0.00543 | 0.087 | 100 |
| H-7 | −0.683 | 4 | −0.841 | 0.00117 | 0.1667 | 100 |
| 0175029-0000 | −0.674 | 6 | −0.852 | 0.00002 | 0.0282 | 100 |
| 8-Azaguanine | −0.628 | 4 | −0.878 | 0.00052 | <0.05 | 100 |
| Gliclazide | −0.626 | 4 | −0.801 | 0.00304 | 0.01 | 100 |
| Meticrane | −0.612 | 5 | −0.841 | 0.00028 | <0.05 | 100 |
| 1,4-Chrysenequinone | −0.612 | 2 | −0.863 | 0.03746 | 0.0859 | 100 |
| 6-Azathymine | −0.607 | 4 | −0.761 | 0.00668 | 0.0115 | 100 |
| Azacitidine | −0.607 | 3 | −0.714 | 0.04795 | 0.1827 | 100 |
| Ellipticine | −0.605 | 4 | −0.719 | 0.01261 | 0.0763 | 100 |
| Mitoxantrone | −0.603 | 3 | −0.796 | 0.01729 | 0.1064 | 100 |
| Apigenin | −0.602 | 4 | −0.738 | 0.00937 | 0.0815 | 100 |

Abbreviation: CMap, Complement Map.

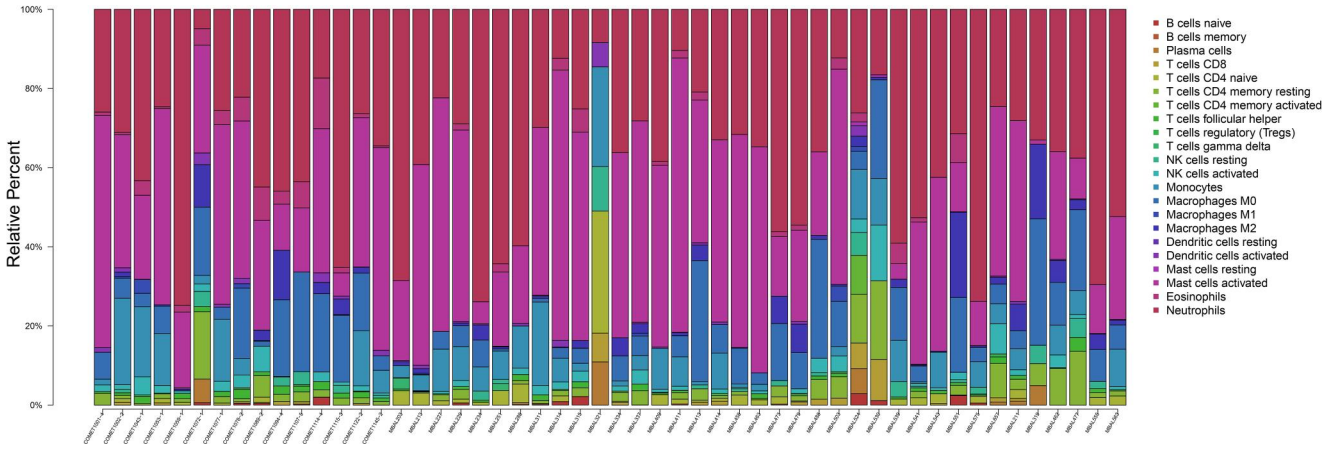


**FIGURE 11** Histogram of all samples based on the ratio of 22 immune cells

that there were significant differences in multiple immune cells between patients of COVID-19 ARDS and control groups.

There are still some limitations of this study that need to be noted. First, this study is a single queue study based on the GEO public database and lacks additional validation cohort. Second, the samples used in this study are tracheal aspirates, which are not as representative of the disease as lung tissue. Third, a large number of potential molecular mechanisms and drugs screened in this study are yet to be further verified by in vivo and in vitro experiments. Despite the above research limitations, this study is the first study to identify the potential molecular mechanism of COVID-19 ARDS and the hub gene through whole-genome RNA sequencing, which can provide a theoretical basis for future COVID-19 ARDS prevention and treatment strategies.

## 5 | CONCLUSIONS

In conclusion, a total of 270 DEGs were screened related to COVID-19 ARDS, and 10 (ERBIN, PCM1, ZFYVE16, SNAPC1, CAB39L, OCLN, UVRAG, HLA-G, CLEC2D and ZNF117) of them were identified as hub genes of COVID-19

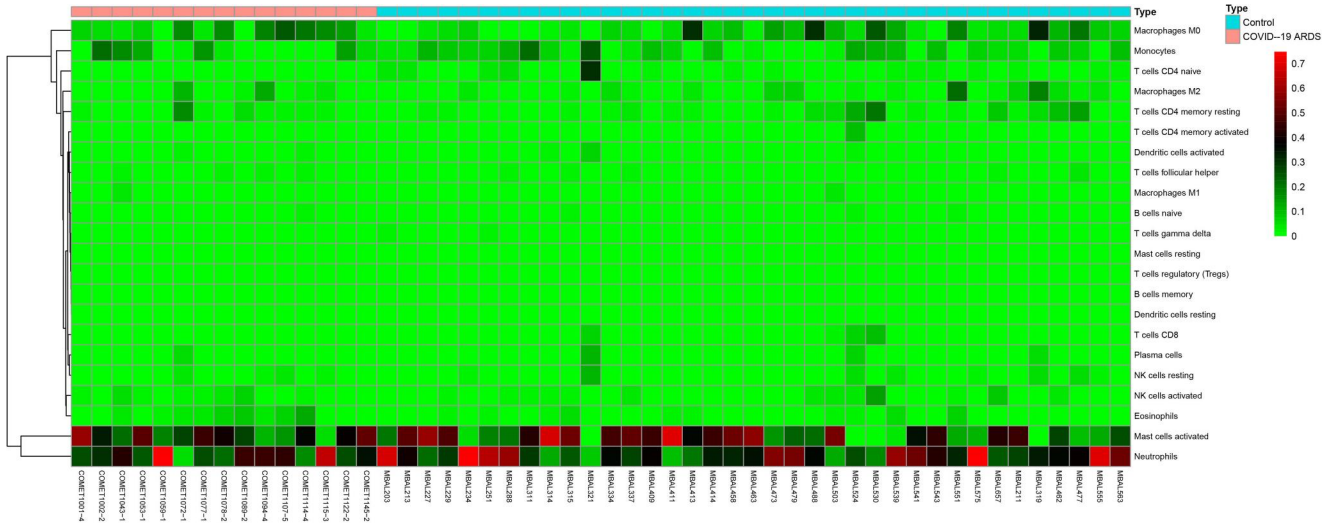


**FIGURE 12** Heat map of all samples based on the level of 22 immune cells

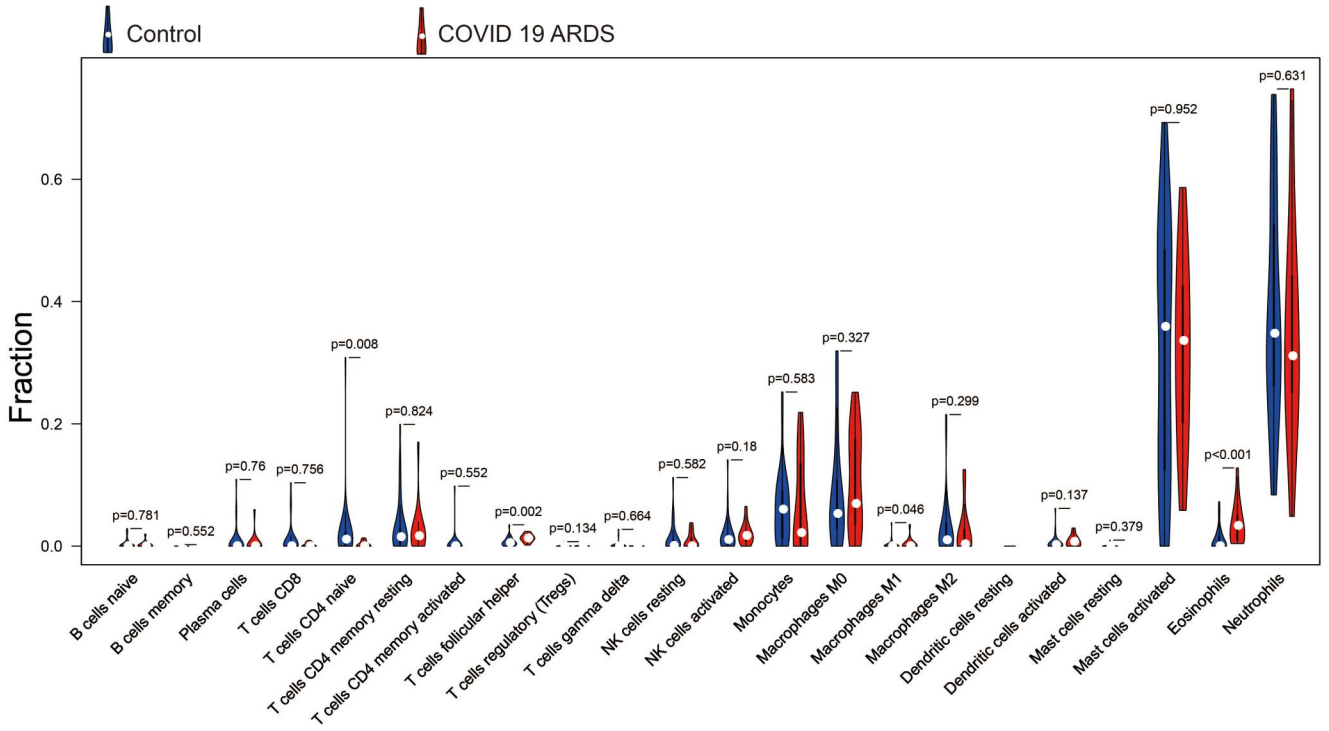


**FIGURE 13** Violin plot of the level of 22 immune cells between COVID-19 acute respiratory distress syndrome and control group patients

ARDS. Response to cytokine, G-protein coupled receptor activity, ionotropic glutamate receptor signalling pathway and G-protein coupled receptor signalling pathway, and longevity pathway may be the potential molecular mechanisms of COVID-19 ARDS. Sixty-seven targetted drugs of COVID-19 ARDS were screened by CMAP analysis. Immune cell infiltration analysis revealed that the levels of T cells CD4 naive, T cells follicular helper, macrophages M1 and eosinophils in COVID-19 ARDS patients were significantly different than those in control group patients. Through our current study, we have identified COVID-19 ARDS-related hub DEGs and molecular mechanisms through a comprehensive analysis of the RNA sequencing dataset and also revealed the difference in immune cell infiltration of COVID-19 ARDS. However, our results still need to be verified in additional cohorts as well as in vivo and in vitro experiments verification.

## ACKNOWLEDGEMENTS

The authors also thank the contributors of GSE163426 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163426) for the open access sharing of the COVID-19 ARDS dataset. In addition, we would like to acknowledge the helpful comments on this article received from our reviewers.

## CONFLICT OF INTERESTS

The authors declare that they have no competing interests.

## DATA AVAILABILITY STATEMENT

The datasets used during the present study are available from the corresponding author upon reasonable request. The raw whole genome expression microarray dataset of COVID-19 ARDS can be obtained from the GSE163426 dataset (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE163426).

## ORCID

*Wangsheng Deng* https://orcid.org/0000-0002-2367-3174

## REFERENCES

1. Asselah, T., et al.: Covid-19: Discovery, diagnostics and drug development. J. Hepatol. 74(1), 168–184 (2021)
2. Machhi, J., et al.: The natural history, pathobiology, and clinical manifestations of SARS-Cov-2 infections. J. Neuroimmune Pharmacol. 15(3), 359–386 (2020)
3. de Souza, T.H., et al.: Clinical manifestations of children with Covid-19: A systematic review. Pediatr. Pulmonol. 55(8), 1892–1899 (2020)
4. Guan, W.J., et al.: Clinical characteristics of coronavirus disease 2019 in China. N. Engl. J. Med. 382(18), 1708–1720 (2020)
5. Wang, T., et al.: Clinical features of coronavirus disease 2019 patients with mechanical ventilation: A nationwide study in China. Crit. Care Med. 48(9), e809–e812 (2020)
6. Kommoss, F.K.F., et al.: The pathology of severe Covid-19-related lung damage. Dtsch. Arztebl. Int. 117(29-30), 500–506 (2020)
7. Li, X., Ma, X.: Acute respiratory failure in Covid-19: is it typical ARDS? Crit. Care. 24(1), 198 (2020)
8. Wu, C., et al.: Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. JAMA Intern. Med. 180(7), 934–943 (2020)
9. Felsenstein, S., et al.: Covid-19: Immunology and treatment options. Clin. Immunol. 215, 108448 (2020)
10. Wilk, A.J., et al.: A single-cell atlas of the peripheral immune response in patients with severe Covid-19. Nat. Med. 26(7), 1070–1076 (2020)
11. Robinson, M.D., McCarthy, D.J., Smyth, G.K.: edgeR: A bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26(1), 139–140 (2010)
12. McCarthy, D.J., Chen, Y., Smyth, G.K.: Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 40(10), 4288–4297 (2012)
13. Reiner, A., Yekutieli, D., Benjamini, Y.: Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics. 19(3), 368–375 (2003)
14. Benjamini, Y., et al.: Controlling the false discovery rate in behavior genetics research. Behav. Brain Res. 125(1-2), 279–284 (2001)
15. Dennis, G., Jr., et al.: DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol. 4(5), P3 (2003)
16. Jiao, X., et al.: DAVID-WS: A stateful web service to facilitate gene/protein list analysis. Bioinformatics. 28(13), 1805–1806 (2012)
17. Liberzon, A., et al.: The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 1(6), 417–425 (2015)
18. Subramanian, A., et al.: Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102(43), 15545–15550 (2005)
19. Mostafavi, S., et al.: A real-time multiple association network integration algorithm for predicting gene function. Genome Biol. 9(Suppl 1), S4 (2008)
20. Warde-Farley, D., et al.: The geneMANIA prediction server: Biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 38, W214–W220 (2010). (Web Server issue)
21. Kuhn, M., et al.: STITCH: Interaction networks of chemicals and proteins. Nucleic Acids Res. 36, D684–688 (2008). (Database issue)
22. Kuhn, M., et al.: STITCH 2: An interaction network database for small molecules and proteins. Nucleic Acids Res. 38, D552–D556 (2010). (Database issue)
23. Jensen, L.J., et al.: STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 37, D412–D416 (2009). (Database issue)
24. Franceschini, A., et al.: STRING V9.1: Protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 41, D808–D815 (2013). (Database issue)
25. Szklarczyk, D., et al.: STRING V11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res. 47(D1), D607–D613 (2019)
26. Langfelder, P., Horvath, S.: Wgcna: An R package for weighted correlation network analysis. BMC Bioinf. 9, 559 (2008)
27. Lamb, J., et al.: The Connectivity Map: Using gene-expression signatures to connect small molecules. Genes, and Disease', Science. 313(5795), 1929–1935 (2006)
28. Lamb, J.: The Connectivity Map: A new tool for biomedical research. Nat. Rev. Cancer. 7(1), 54–60 (2007)
29. Wang, Y., et al.: PubChem Bioassay: 2017 update. Nucleic Acids Res. 45(D1), D955–D963 (2017)
30. Hahnke, V.D., Kim, S., Bolton, E.E.: PubChem chemical structure standardization. J Cheminform. 10(1), 36 (2018)
31. Kim, S., et al.: PubChem 2019 update: Improved access to chemical data. Nucleic Acids Res. 47(D1), D1102–D1109 (2019)
32. Elmassry, M.M., Kim, S., Busby, B.: Predicting drug-metagenome interactions: Variation in the microbial beta-glucuronidase level in the human gut metagenomes. PloS One. 16(1), e0244876 (2021)
33. Kim, S., et al.: Teaching cheminformatics through a collaborative intercollegiate online chemistry course (OLCC). J. Chem. Educ. 98(2), 416–425 (2021)
34. Newman, A.M., et al.: Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods. 12(5), 453–457 (2015)
35. Chen, B., et al.: Profiling tumor infiltrating immune cells with CIBERSORT. Methods Mol. Biol. 1711, 243–259 (2018)
36. Vasquez, C.R., et al.: Identification of distinct clinical subphenotypes in critically ill patients with Covid-19. Chest (2021). https://doi.org/10.1016/j.chest.2021.04.062
37. Pan, F., et al.: Factors associated with death outcome in patients with severe coronavirus disease-19 (Covid-19): A case-control study. Int J Med Sci. 17(9), 1281–1292 (2020)
38. Wu, Y.H., et al.: Severe acute respiratory syndrome coronavirus (SARS-Cov)-2 infection induces dysregulation of immunity: In silico gene expression analysis. Int. J. Med. Sci. 18(5), 1143–1152 (2021)
39. Mohammed, A., et al.: Delta9-tetrahydrocannabinol prevents mortality from acute respiratory distress syndrome through the induction of apoptosis in immune cells, leading to cytokine storm suppression. Int. J. Mol. Sci. 21(17) (2020)
40. Zhao, Z., et al.: Single-cell analysis identified lung progenitor cells in Covid-19 patients. Cell Prolif. 53(12), e12931 (2020)
41. Zhang, S., et al.: Dynamics of peripheral immune cells and their HLA-G and receptor expressions in a patient suffering from critical COVID-19 pneumonia to convalescence. Clin. Transl. Immunol. 9(5), e1128 (2020)

42. Zidi, I.: Puzzling out the COVID-19: therapy targeting HLA-G and HLA-E. Hum. Immunol. 81(12), 697–701 (2020)
43. Mamkulathil Devasia, R., et al.: Enhanced production of camptothecin by immobilized callus of Ophiorrhiza mungos and a bioinformatic insight into its potential antiviral effect against SARS-Cov-2. J. King Saud Univ. Sci. 33(2), 101344 (2021)
44. Alakwaa, F.M.: Repurposing didanosine as a potential treatment for COVID-19 using single-cell RNA sequencing data. mSystems. 5(2) (2020)
45. Sajid Jamal, Q.M., Alharbi, A.H., Ahmad, V.: Identification of doxorubicin as a potential therapeutic against SARS-Cov-2 (COVID-19) protease: A molecular docking and dynamics simulation studies. J. Biomol. Struct. Dyn. 1–15 (2021)
46. Lovetrue, B.: The AI-discovered aetiology of COVID-19 and rationale of the irinotecan+ etoposide combination therapy for critically ill COVID-19 patients. Med. Hypotheses. 144, 110180 (2020)
47. Singh Tomar, P.P., Arkin, I.T.: SARS-Cov-2 E protein is a potential ion channel that can be inhibited by gliclazide and memantine. Biochem. Biophys. Res. Commun. 530(1), 10–14 (2020)
48. Lokhande, K.B., et al.: Molecular docking and simulation studies on SARS-Cov-2 $M^{pro}$ reveals Mitoxantrone, Leucovorin, Birinapant, and Dynasore as potent drugs against COVID-19. J. Biomol. Struct. Dyn. 1–12 (2020)

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Deng, W., et al.: Comprehensive investigation of RNA-sequencing dataset reveals the hub genes and molecular mechanisms of coronavirus disease 2019 acute respiratory distress syndrome. IET Syst. Biol. 1–14 (2021). https://doi.org/10.1049/syb2.12034