

RESEARCH PAPER



## COVID-19 of differing severity: from bulk to single-cell expression data analysis

Linlin Tian<sup>a\*</sup>, Min He<sup>a\*</sup>, Huafeng Fan<sup>a</sup>, Hongying Zhang<sup>a</sup>, Xiaoxiao Dong<sup>a</sup>, Mengkai Qiao<sup>a</sup>, Chenyu Tang<sup>a</sup>, Yan Yu<sup>a</sup>, Tong Chen<sup>b</sup>, and Nan Zhou<sup>a</sup>

<sup>a</sup>Nanjing Municipal Center for Disease Control and Prevention, Nanjing, Jiangsu, P.R. China; <sup>b</sup>Department of Urology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, P.R. China

### ABSTRACT

Coronavirus disease 2019 (COVID-19) is raging worldwide and causes an immense disease burden. Despite this, the biomarkers and targeting drugs of COVID-19 of differing severity remain largely unknown. Based on the GSE164805 dataset, we identified modules most critical for mild COVID-19 (mCOVID-19) and severe COVID-19 (sCOVID-19) through WGCNA, respectively. We subsequently constructed a protein–protein interaction network, and detected 16 hub genes for mCOVID-19 and 10 hub genes for sCOVID-19, followed by the prediction of upstream transcription factors (TFs) and kinases. The enrichment analysis then showed downregulation of TNFA signaling via NFKB for mCOVID-19, as well as downregulation of MYC targets V1 for sCOVID-19. Infiltration degrees of many immune cells, such as macrophages, were also sharply different between mCOVID-19 and sCOVID-19 samples. Predicted protein targeting drugs with the highest scores nearly all belong to naturally derived or synthetic glucocorticoids. For the two single-cell RNA-seq datasets, we explored the expression distribution of hub genes for mCOVID-19/sCOVID-19 in each cell type. The expression levels of PRKCA, MCM5, TYMS, RBBP4, BCL6, FLOT1, KDM6B, and TLR2 were found to be cell-type-specific. Furthermore, the expression levels of 10 hub genes for mCOVID-19 were significantly upregulated in PBMCs between eight healthy controls and eight mCOVID-19 patients at our institution. Collectively, we detected critical modules, pathways, TFs, kinases, immune cells, targeting drugs, hub genes, and their expression distributions in different cell types that may involve the pathogenesis of COVID-19 of differing severity, which may propel earlier diagnosis and more effective treatment of this intractable disease in the future.

### ARTICLE HISTORY

Received 23 February 2023  
Revised 3 April 2023  
Accepted 24 June 2023

### KEYWORDS





COVID-19; bulk; single-cell

## Introduction


Coronavirus disease 2019 (COVID-19) pandemic threatens the worldwide healthcare system and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) serves as the causal virus. Clinical presentations of COVID-19 vary greatly between individuals, ranging from asymptomatic infection to respiratory failure, and even death [1]. Approximately, 20% of the patients with COVID-19 experience exacerbations, and 5% require intensive care [2].

In order to identify novel biomarkers, previous transcriptomic studies of COVID-19 were on the basis of lung epithelial cell [3], peripheral blood mononuclear cells (PBMCs) [4], etc. To date, almost all previous transcriptomic studies of COVID-19 have simply divided included subjects into COVID-19 patients and healthy people, and

have not further subdivided COVID-19 patients [4,5]. In doing so, the difference in gene expression profiles of COVID-19 patients of differing severity remains largely unknown. Given that the variation in therapeutic measures for COVID-19 patients ranges widely from at-home isolation to intensive care, it is therefore necessary to find novel efficient biomarkers for early patient stratification. The exploration of the molecular mechanisms behind COVID-19 of differing severity is also particularly important for therapeutic tailoring. Gratifyingly, a previous attempt was made to find inflammatory and immune alterations of COVID-19 patients of differing severity based on bulk transcriptome analysis of PBMCs taken from mild COVID-19 (mCOVID-19) and severe COVID-19 (sCOVID-19) patients, as well

**CONTACT** Nan Zhou  [zhounan0408@126.com](mailto:zhounan0408@126.com)  Nanjing Municipal Center for Disease Control and Prevention, Nanjing, Jiangsu, P.R. China; Tong Chen  [doctor\\_chentong@163.com](mailto:doctor_chentong@163.com)  Department of Urology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, P.R. China

\*Linlin Tian and Min He contributed equally to this study.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/15384101.2023.2239620>

as healthy people [6]. In addition, two single-cell RNA-seq (scRNA-seq) transcriptome studies with mCOVID-19 and sCOVID-19 patients were also conducted for in-depth analyses of different cell populations [7,8].

The weighted gene co-expression network analysis (WGCNA) acts as a useful methodology that enables the identification of interactions between genes and unveils complex biological mechanisms behind various physiological or pathophysiological states [9]. WGCNA thus far has been used to identify critical genes and modules in COVID-19 [10–13], but nevertheless rarely applied in COVID-19 of differing severity. Here, we performed WGCNA to obtain critical modules in mild and severe COVID-19 patients. We subsequently detected hub genes from protein–protein interaction (PPI) network. In addition, single sample gene set enrichment analysis (ssGSEA) was utilized for the identification of the abundance of infiltrating immune cells. We also applied the Drug Gene Interaction Database (DGIdb) to find the targeting drugs. Furthermore, upstream potential transcription factors (TFs) and kinases, as well as enriched downstream pathways were also predicted. We also explored the expression distribution of hub genes in each cell type through two scRNA-seq datasets. Furthermore, qPCR was utilized to validate the mRNA expression levels of hub genes in PBMCs between healthy controls and mCOVID-19 patients at our institution. Based on the above-mentioned comprehensive genomic analyses, we sought to determine the potential critical genes, hub modules, TFs, kinases, signaling pathways, infiltrating immune cells, and drug targets that may involve the pathogenesis of mild and severe COVID-19.

## Materials and methods

### Data sources

We achieved gene-expression profiling from the Gene Expression Omnibus (GEO) database. Inclusion criteria of this study for bulk or single-cell dataset selection were as follows: (1) transcriptome data should be available for COVID-19 samples of differing severity; (2) gene expression data should contain at least 2000 genes. Correspondingly,

the GSE164805 dataset, containing COVID-19 expression profile from PBMCs was the unique eligible bulk microarray dataset. Specifically, the GSE164805 dataset was analyzed using GPL26963 platform and contained 5 samples from healthy controls, 5 samples from mCOVID-19 patients, and 5 samples from sCOVID-19 patients. Patient demographic characteristics of the GSE164805 dataset, including age and gender, are shown in Table S1, and a comparison of these two demographic characteristics among groups is presented in Table S2. Here, there was no significant difference among groups with regard to age and gender. Also, three other datasets containing PBMC samples in the context of COVID-19 (GSE152418, GSE206263, and GSE157103) were included for further exploration. The GSE152418 dataset, containing 17 samples from healthy controls, 4 samples from mCOVID-19 patients, and 8 samples from sCOVID-19 patients, was analyzed using GPL24676 platform [4]. In addition, there were 7 samples from healthy controls, 5 samples from mCOVID-19 patients, and 4 samples from moderate COVID-19 patients in the GSE206263 dataset, which was analyzed through the GPL24676 platform [14]. As for the GSE157103 dataset, 50 samples from ICU COVID-19 patients and 50 samples from non-ICU COVID-19 patients were included [15].

In addition to the above-mentioned bulk datasets, two scRNA-seq datasets, namely Schulte-Schrepping dataset and the GSE216020 dataset, were also included in this study. Schulte-Schrepping dataset contained two independent cohorts of COVID-19 cases [8], namely the Berlin cohort (cohort 1) [12] and the Bonn cohort (cohort 2). Specifically, cohort 1 consisted of 49 COVID-19 samples (8 mild and 10 severe patients at different time points) and 22 control samples, while cohort 2 consisted of 50 COVID-19 samples (8 mild and 9 severe patients at different time points) and 14 control samples (13 controls at different time points). Besides, the GSE216020 dataset included 24 samples and was analyzed through GPL24676 platform. In order to guarantee the background concordance of different groups, we selected those samples taken at day 0. Correspondingly, 5 samples (GSM6656095, GSM6656096, GSM6656097, GSM6656098, and GSM6656099) from healthy controls, 4 samples (GSM6656081, GSM6656085, GSM6656101, and GSM6656104)

from mCOVID-19 patients, and 6 samples (GSM6656082, GSM6656084, GSM6656088, GSM6656090, GSM6656091, and GSM6656100) from sCOVID-19 patients were applied for further analysis.

### **Data processing of bulk RNA-seq**

The expression matrix and sample information of the bulk RNA-seq datasets were downloaded from the GEO website. FeatureCounts function from the *subread* R package was utilized to summarize counts per gene. Then, we imported, organized, filtered, and normalized the data using the *edgeR* package. To be specific, we filtered out the low-abundance genes using the filterByExpr function, and we normalized the data using the TMM method. As the datasets were downloaded from different studies and different platforms, we applied the ComBat function in the *sva* R package [16] in order to perform batch effect correction. Subsequently, for the assessment of sample similarity and putative batch effect, principal component analysis (PCA) was conducted before and after batch effect correction. The log 2-fold change (logFC) and false discovery rate-adjusted *P* value (FDR) of hub genes were calculated between healthy controls and COVID-19 PBMC samples through the *limma* R package [17]. A moderate *t* test was performed, and Benjamini–Hochberg method was applied for the adjustment of the *P* value.

### **WGCNA analysis**

We constructed signed weighted gene co-expression network using WGCNA R package [9]. For the detection of numerous soft power threshold  $\beta$  over  $R^2$ , pickSoftThreshold function was utilized. The power of  $\beta$  was set at 14 for ensuring the scale-free network (Figure S1C, S1D). We transformed Pearson correlation matrix into the adjacency matrix and subsequently further transformed it into the topological overlap matrix (TOM) in order to compute the inconsistency. We incorporated very consistent modules and the height cutoff was set at 0.25. According to Pearson association, we assessed the module–trait relationship. For

specific gene, we calculated module membership (MM) on the basis of correlations between module eigengene and the expression levels across samples. Gene significance (GS) refers to the correlation of gene expression level with the phenotype. For the critical module, we calculated GS and MM of every gene within this module. The specific gene within one module was considered critical when both  $MM > 0.8$  and  $GS > 0.2$  were met.

### **PPI network establishment**

We utilized Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) to explore PPI network [18]. Subsequently, we applied Cytoscape software to better visualize the PPI network [19]. Cytoscape plugin CytoHubba [20] was adopted for the identification of key genes in the network. Lastly, we intersected the top 40 genes in every method within the CytoHubba plugin, and the intersected genes were treated as hub genes.

### **Upstream regulatory network prediction**

In order to determine TFs and kinases that may regulate COVID-19 related genes, we applied eXpression2Kinases (X2K) (<https://amp.pharm.mssm.edu/X2K>) for the identification and ranking of potential TFs, intermediate proteins, and kinases that were likely responsible for the dysregulation of gene expression [21]. We obtained the top 10 TFs and top 10 kinases on the basis of hypergeometric *P* value, followed by the construction and visualization of the regulatory network.

### **Drug target establishment**

The drug–gene interactions were analyzed using DGIdb (<https://dgidb.org/>) [22]. Drugs and compounds were predicted to target critical genes. Here, we only included approved drugs that had activation or inhibitory interaction with the critical gene. After downloading, we input the interaction network into the Cytoscape to better visualize the network.

### Gene set enrichment analysis (GSEA) and gene set variation analysis (GSVA)

GSEA was conducted through the *clusterProfiler* R package [23] and the Benjamini–Hochberg (BH) method was applied for the adjustment of *P* value. False discovery rate < 0.1 and *P* value < 0.05 were set as the threshold for the confirmation of significant enrichment. GSVA is a nonparametric unsupervised method to display differential enrichment pathways [24]. We used GSVA R package to explore hallmark pathways of immune-related genes. Based on the Molecular Signature Database (<http://software.broadinstitute.org/gsea/msigdb/index.jsp>), we obtained gene set “h.all.v7.1.symbols.gmt”, which was used as the reference.

### Immune cells infiltration analysis

We implemented ssGSEA through GSVA R package for the assessment of enrichment scores of PBMCs samples [24]. Here, we applied gene expression to PBMCs samples and meta-genes within 28 types of immune cells. Immune infiltration levels of each immune cell enrichment were indicated by the normalized enrichment score. Subsequently, the correlation of each hub gene with each immune cell was also computed.

### ScRNA-seq data analysis

The cell unique molecular identifier (UMI) matrix was converted to a Seurat object through the package Seurat v3 [25]. The criteria for the filtration of the matrix were a minimum number of cells equal to 3 and a minimum number of features equal to 200. We merged the three caput samples and detected single-cell transcriptomes after filtration. We detected neighbors through seven dimensions. As for the unsupervised clustering, a suitable resolution parameter was selected. PCA dimensions were reduced via Uniform Manifold Approximation and Projection (UMAP). Besides, feature plots were generated through the package Seurat v3.

### Random forest (RF) model construction

In order to explore the discriminatory ability of hub gene expression between non-ICU COVID-19 and ICU COVID-19 patients, we constructed an RF model using the *randomForest* R package [26]. For the identification of a better RF model, we endowed *mtry* and *ntree* with 3 and 500. Genes with an importance value > 2 were considered ICU COVID-19-specific genes. We randomly split the data into 70% training and 30% testing partitions. In addition, we plotted the area under the receiver operating characteristic (ROC) curve (AUC) for assessment of the RF model.

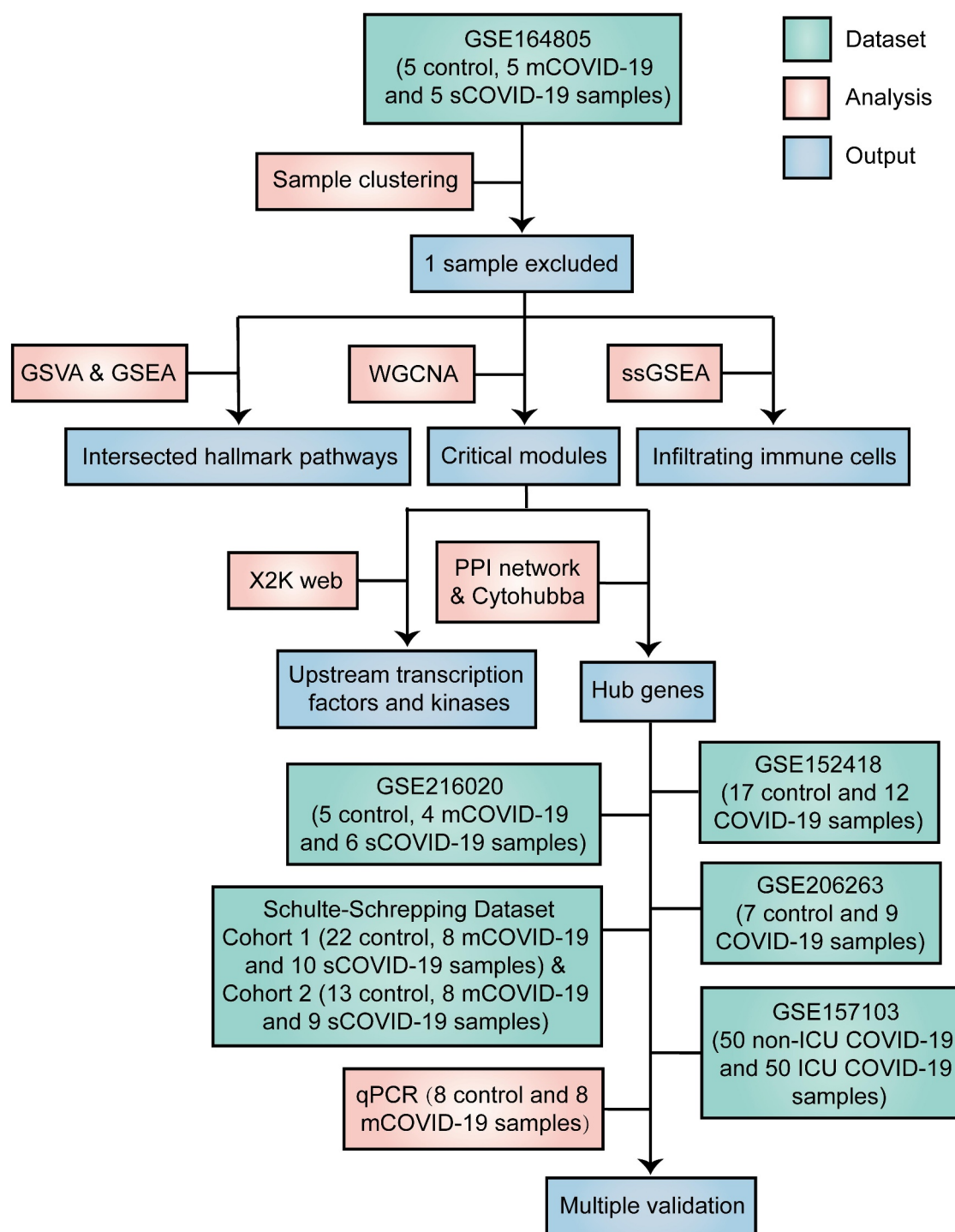
### RNA extraction and quantitative polymerase chain reaction (qPCR)

We extracted PBMCs in blood samples from eight healthy controls and eight mCOVID-19 patients at our institution. Then, we extracted total mRNAs from the PBMCs through RNAPrep pure blood kit (Tiangen Biotech, Beijing, China). Subsequently, a reverse transcription kit (Servicebio, Wuhan, China) was used to synthesize cDNA. We conducted qPCR using SYBR Green Master Mix (Servicebio, Wuhan, China) on a QuantStudio 5 Real-Time instrument (Applied Biosystems). In accordance with the comparative  $C_t$  method ( $\Delta\Delta C_t$  method), we conducted the normalization of relative mRNA level to GAPDH mRNA level. The synthesization of primers was performed by Tsingke (Nanjing, China). The primer sequence of each hub gene is listed in Table S6.

## Results

### Hub modules establishment

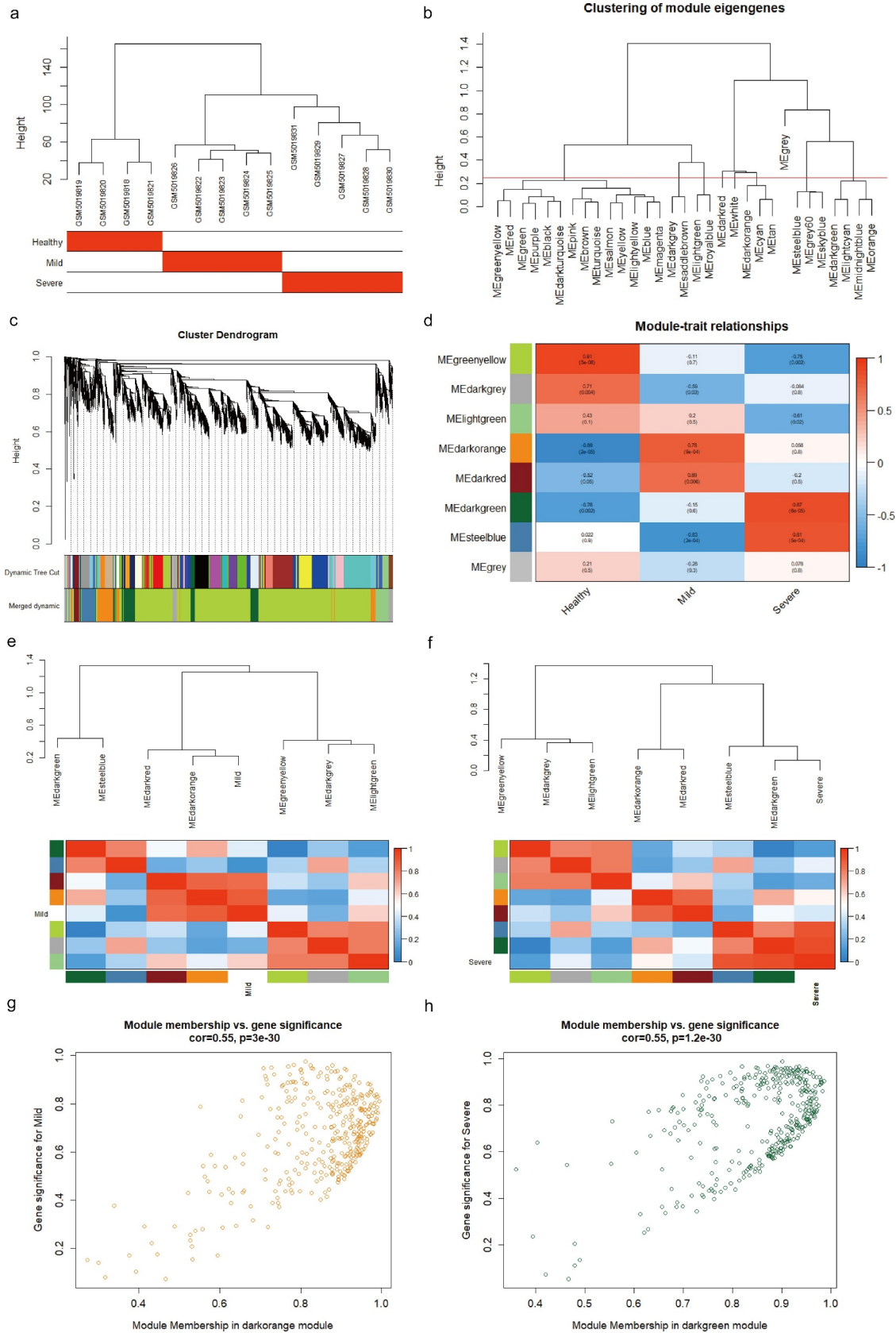
A detailed flowchart depicting the design of this study is provided in Figure 1. Initially, we clustered all 15 samples. One sample from healthy controls was considered the outlier and therefore excluded (GSM5019817; Figure S1A). For the identification of COVID-19-related modules, the top 25% variant genes across 14 samples (from 4414 gene expression matrix) were used for WGCNA (Figure 2a). Besides,  $\beta = 14$  and  $R^2 = 0.7$  were considered optimal for ensuring the scale-free network (Figure S1C, S1D). In the process of incorporating



**Figure 1.** The flowchart of this study design. GSEA, gene set enrichment analysis; GSVA, gene set variation analysis; Mcovid-19, mild COVID-19; Scovid-19, severe COVID-19; WGCNA, weighted gene co-expression network analysis.

similar modules, clustering height cutoff was set to 0.25, and 8 modules were achieved (Figure 2b,c). The darkorange module exhibited the most relevant association with mCOVID-19 ( $R = 0.78$ ,  $P = 9e-04$ , Figure 2d), whereas the darkgreen module indicated the

most relevant association with sCOVID-19 ( $R = 0.87$ ,  $P = 6e-05$ , Figure 2d). The interactions among these co-expressed modules were shown through the module eigengene dendrogram. Notably, the darkorange module was in close proximity to mCOVID-19, and exhibited

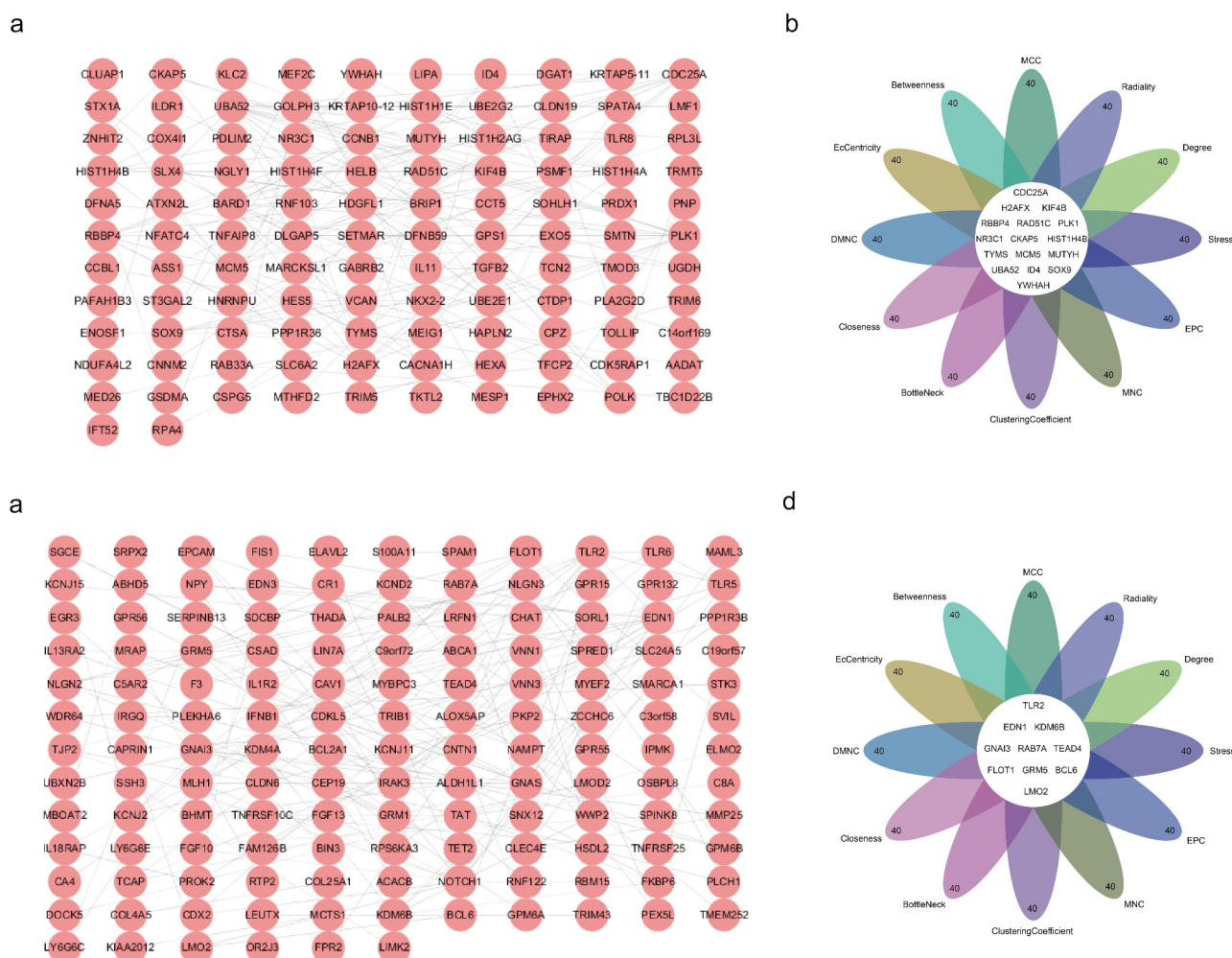


**Figure 2.** Construction of co-expression modules using WGCNA. (a) Clustering dendrogram of the samples. (b) the cluster dendrogram of module eigengenes. (c) the cluster dendrogram and module color. (d) Module-trait relationships between different modules and traits. (e) the correlations among Mcovid-19 modules. (f) the correlations among Scovid-19 modules. (g) Scatter plot showing the correlation between GS and MM according to genes within Mcovid-19 darkorange module. (h) Scatter plot showing the correlation between GS and MM according to genes within Scovid-19 darkgreen module. WGCNA, weighted gene co-expression network analysis. Mcovid-19, mild COVID-19. Scovid-19, severe COVID-19.

a highly relevant association with mCOVID-19 (Figure 2e). Similar results were found in darkgreen module for sCOVID-19 (Figure 2f). Intramodular results indicated that genes in the darkorange module were positively correlated with mCOVID-19, with  $p < 3e-30$  and correlation = 0.55, while genes in the darkgreen module showed negative correlation with sCOVID-19, with  $P < 1.2e-30$  and correlation = 0.55 (Figure 2g,h). Given that  $\text{cor. geneModuleMembership} > 0.8$  and  $\text{cor. geneTraitSignificance} > 0.2$  were set as the cut-off values, we obtained 201 critical genes in the darkorange module for mCOVID-19 and 249 critical genes in the darkgreen module for sCOVID-19.

### PPI network establishment

According to 201 critical genes for mCOVID-19 and 249 critical genes for sCOVID-19, we established PPI networks for mCOVID-19 and sCOVID-19, respectively (Figure 3a,c). We applied all 12 approaches in the CytoHubba plugin, and intersected the top 40 genes of every approach (Figure 3b,d). In doing so, 16 intersected genes, including CDC25A, H2AFX, KIF4B, RBBP4, RAD51C, PLK1, NR3C1, CKAP5, HIST1H4B, TYMS, MCM5, MUTYH, UBA52, ID4, SOX9, and YWHAH were obtained for mCOVID-19 (Figure 3b; Table S3), while 10 hub genes, including TLR2, EDN1, KDM6B, GNAI3, RAB7A, TEAD4, FLOT1, GRM5, BCL6, and LMO2, were achieved for sCOVID-19 (Figure 3d; Table S4).



**Figure 3.** PPI network and hub genes detection. (a) PPI network according to genes from Mcovid-19 darkorange module. (b) Hub genes within Scovid-19 darkgreen module. (c) PPI network based on genes from Scovid-19 darkgreen module. (d) Hub genes within Scovid-19 darkgreen module. Degree, node connect degree; DMNC, density of maximum neighborhood component; EPC, edge percolated component; MCC, maximal clique centrality; MNC, maximum neighborhood component; Closeness, node connect closeness; PPI, protein–protein interaction.

### Upstream regulatory network

For the understanding of upstream regulatory molecules, we utilized X2K web based on the gene expression within the most critical module for mCOVID-19 or sCOVID-19. Among mCOVID-19 related TFs, PML and SIN3A exhibited the most associations with intermediate proteins and kinases (Figure 4a). Furthermore, MAPK3, CDK4, CK2ALPHA, MAPK14, AKT1, GSK3B were the top kinases with the most associations for mCOVID-19 (Figure 4a). STAT3, RUNX1, NFE2L2, and EGR1 were found to be the most significantly correlated TFs that regulated the sCOVID-19 related genes (Figure 4b). In addition, DNAPK, CK2ALPHA, CSNK2A1, PRKCA, MAPK3, ERK1, and ABL1 were the kinases that showed the most associations for sCOVID-19 (Figure 4b).

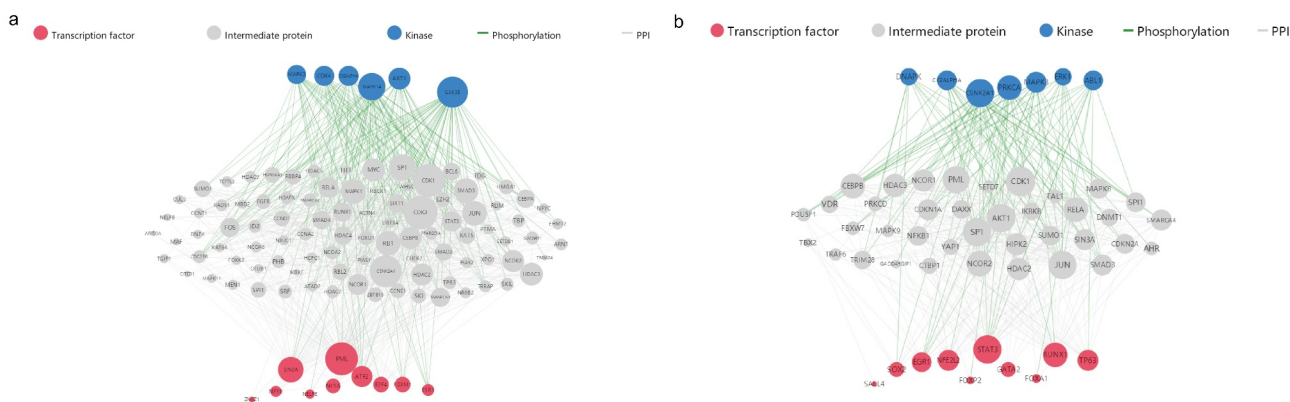
### Functional enrichment

For the exploration of putative hallmark pathways correlated with mCOVID-19 and sCOVID-19, gene expression was applied for exploring the association between each hallmark pathway expression and every sample using GSVA (Figure S2A, S2C). In addition, the dysregulated hallmark pathways in mCOVID-19 and sCOVID-19 are visualized through lollipop plot (Figure S2B, S2D). The significantly enriched hallmark pathways for mCOVID-19 included apoptosis, DNA repair, interferon gamma response, oxidative phosphorylation, interferon alpha response, and

TNFA signaling via NFKB, which are shown through gseaplot (Figure S3A). Additionally, GSEA analysis demonstrated DNA repair, inflammatory response, MYC targets V1, oxidative phosphorylation, and TNFA signaling via NFKB as primarily associated hallmark pathways for sCOVID-19 (Figure S3C). The most significantly enriched hallmark pathways for mCOVID-19 and sCOVID-19 are plotted through cnetplot (Figure S3B, S3D). Then, through the intersection of significantly associated hallmark pathways using GSEA and GSVA, we obtained one upregulated and three downregulated pathways for mCOVID-19 and sCOVID-19, respectively (Figure 5a,b). According to this, we found that inflammatory signaling, including interferon alpha response and inflammatory response were upregulated in mCOVID-19 and sCOVID-19 patients (Figure 5a,b). On the other hand, a pronounced downregulation of DNA repair and oxidative phosphorylation was observed in both mCOVID-19 and sCOVID-19 patients (Figure 5a,b). Furthermore, downregulation of TNFA signaling via NFKB was detected for mCOVID-19 (Figure 5a), while downregulation of MYC targets V1 was observed for sCOVID-19 (Figure 5b).

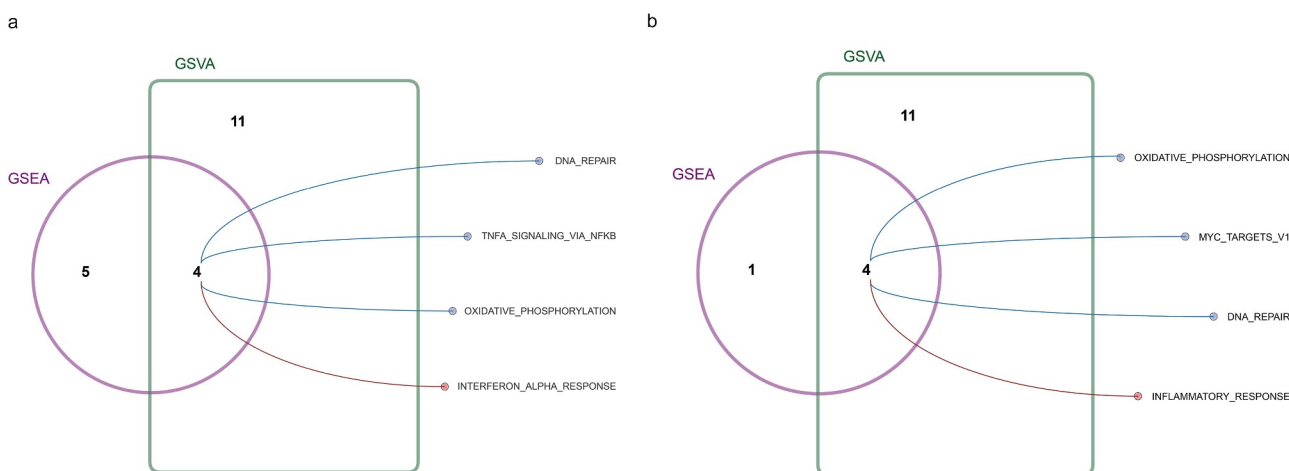
### Immune cells abundance

For the exploration of immune fluctuations in the process of COVID-19 progression, infiltrating immune cells of all included samples were evaluated. Through the heatmap, the respective relationships of infiltrating immune cells between



**Figure 4.** Predicted upstream regulatory network through the X2K Web. (a) Upstream regulatory network based on genes from Mcovid-19 darkorange module. (b) Upstream regulatory network based on genes from Scovid-19 darkgreen module.





**Figure 5.** Intersected hallmark pathways for COVID-19. The intersection of significantly enriched hallmark pathways in (a) Mcovid-19 and (b) Scovid-19.

healthy and mCOVID-19 samples were described (Figure 6a). Subsequently, the correlation between infiltrated levels of immune cells for healthy and mCOVID-19 samples were also shown (Figure 6b). Notably, a significant decrease of various immune cells was found for mCOVID-19, including activated CD8 T cell, CD56<sup>bright</sup> natural killer cell, central memory CD4 T cell, effector memory CD8 T cell and type 1 T helper cell. In contrast, a significant increase was observed for immature B cell, monocyte, neutrophil, plasmacytoid dendritic cell, regulatory T cell, and T follicular helper cell (Figure 6c).

Analogously, the relationships of infiltrating immune cells between healthy and sCOVID-19 samples were mapped (Figure 7a), and the correlations of immune cells in healthy and sCOVID-19 samples were shown (Figure 7b). A significant decrease in certain immune cells (activated CD8 T cell, CD56<sup>bright</sup> NK cell, CD56<sup>dim</sup> NK cell, etc.) was observed for sCOVID-19 samples. Conversely, an increase in neutrophil, eosinophil, immature dendritic cell, macrophage, monocyte, plasmacytoid dendritic cell, and type 17 T helper cell were detected in sCOVID-19 samples (Figure 7c).

### Interaction network of drugs and hub genes

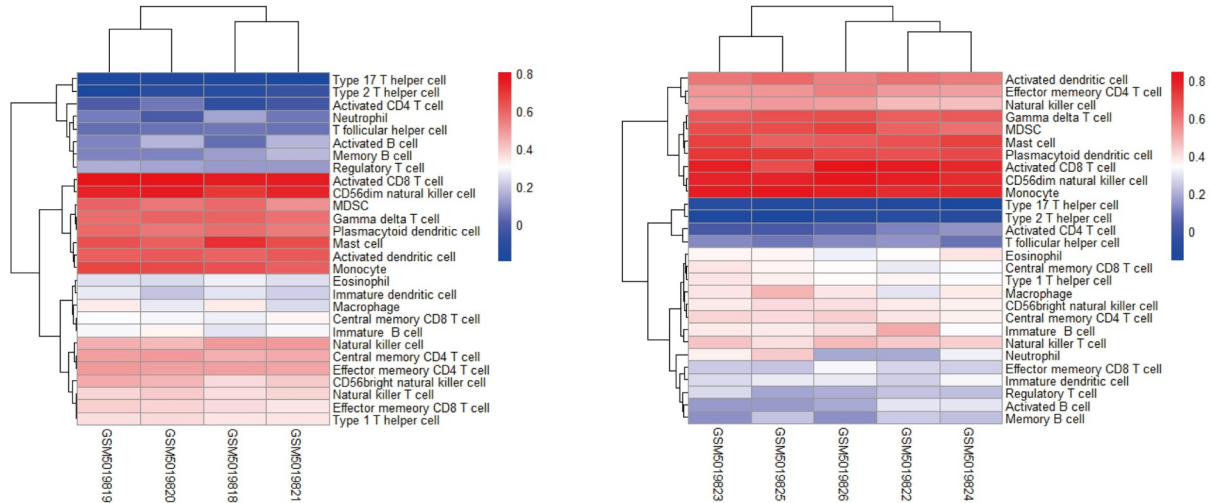
In order to predict the drugs that target the hub genes, 16 hub genes for mCOVID-19 and 10 hub genes for sCOVID-19 obtained from

respective PPI networks were applied for the prediction of potential drugs. Only the approved drugs (inhibitors or activators) with obvious pharmacological effects were taken into consideration. NR3C1 and TYMS were found to be potentially targeted by drugs for mCOVID-19, and a total of 64 FDA-approved drugs were predicted for this disease status (Figure 8). Cortisone acetate and fluorometholone served as the two drugs with the highest query scores as well as interaction scores. As for sCOVID-19, only GRM5 was predicted to have potential interaction with FDA-approved drug (Figure 8a). The interactions between drugs and NR3C1, TYMS, and GRM5 are depicted in Table S5. Furthermore, a significant correlation between NR3C1, TYMS, and GRM5 and immune cells is also depicted (Figure 8b).

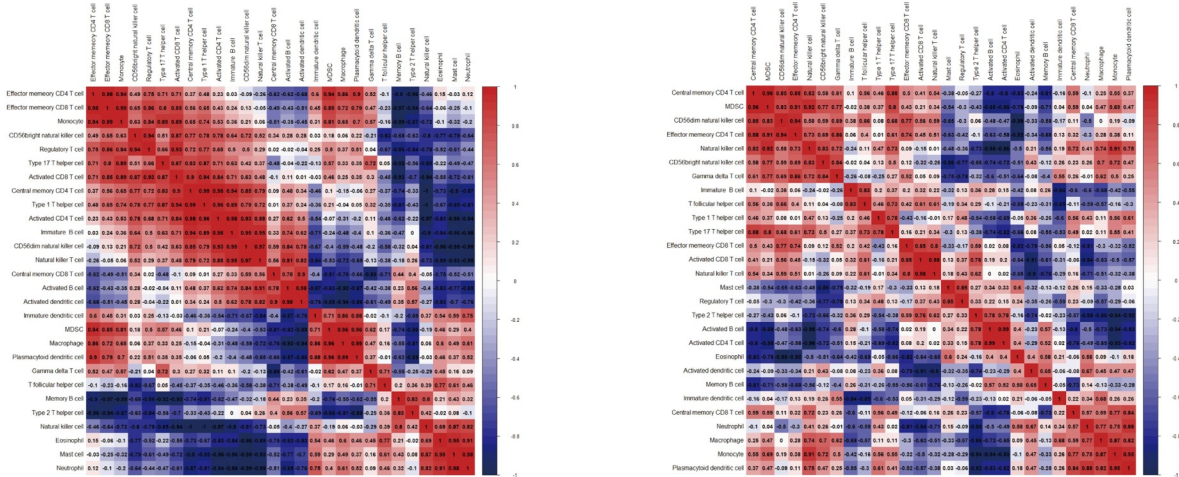
### Expression distribution of hub genes in each cell type for Mcovid-19 or Scovid-19

In order to unveil the expression distribution of hub genes in different cell types, scRNA-seq analysis of the Schulte-Schrepping dataset and GSE216020 dataset was conducted. With regard to cohort 1 in the Schulte-Schrepping dataset, we found 17 clusters of cells, which were annotated as classical monocytes, HLA-DR<sup>hi</sup> CD83<sup>hi</sup> monocytes, CD163 monocytes (sample ID1\_d7), HLA-DR<sup>lo</sup> S100A<sup>hi</sup> monocytes, non-classical monocytes, neutrophils, immature neutrophils, mDCs, pDCs,

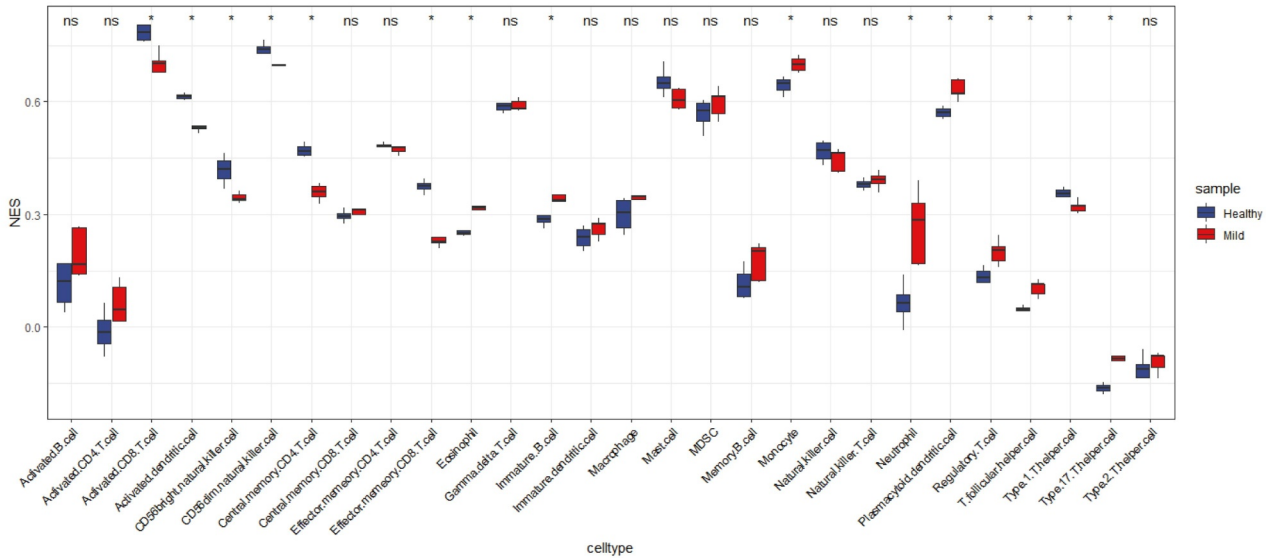
a



b

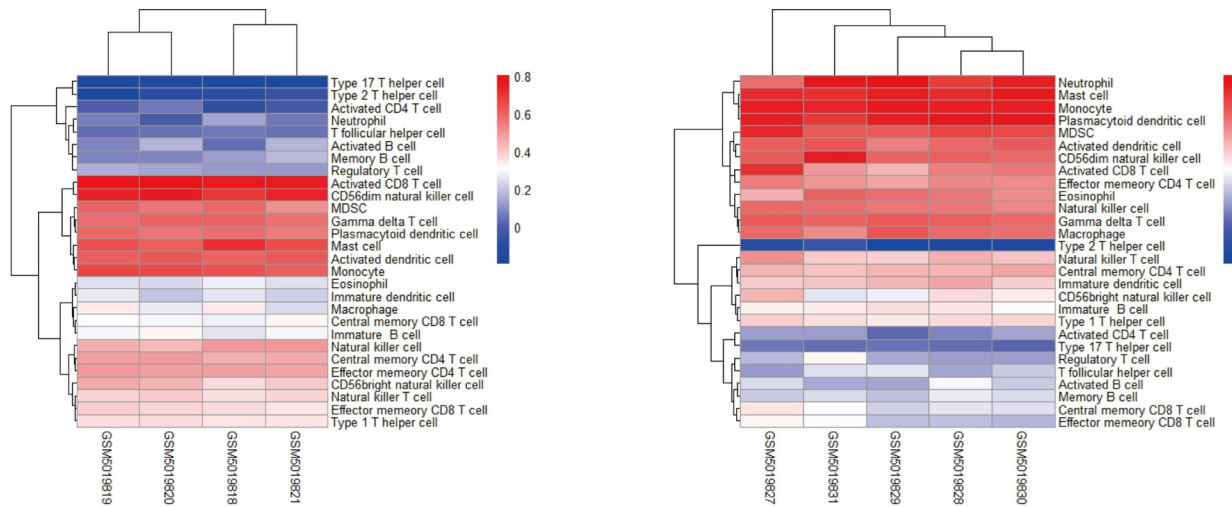


c

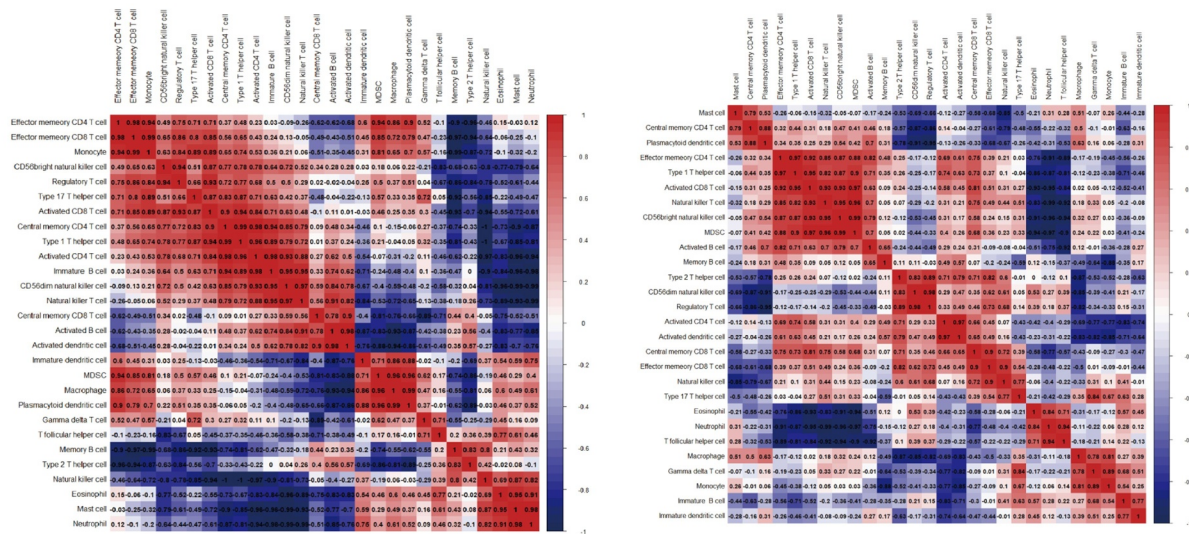


**Figure 6.** Infiltrating immune cells differences between Mcovid-19 and control samples. (a) the heatmap summarizing the correlation of immune cells with Mcovid-19 (right) and healthy (left) samples. (b) Heatmap of correlation showing the association between immune cells based on Mcovid-19 (right) and healthy (left) samples. (c) Bar graph nested by violin plot showing infiltrating immune cells difference between Mcovid-19 and healthy samples.

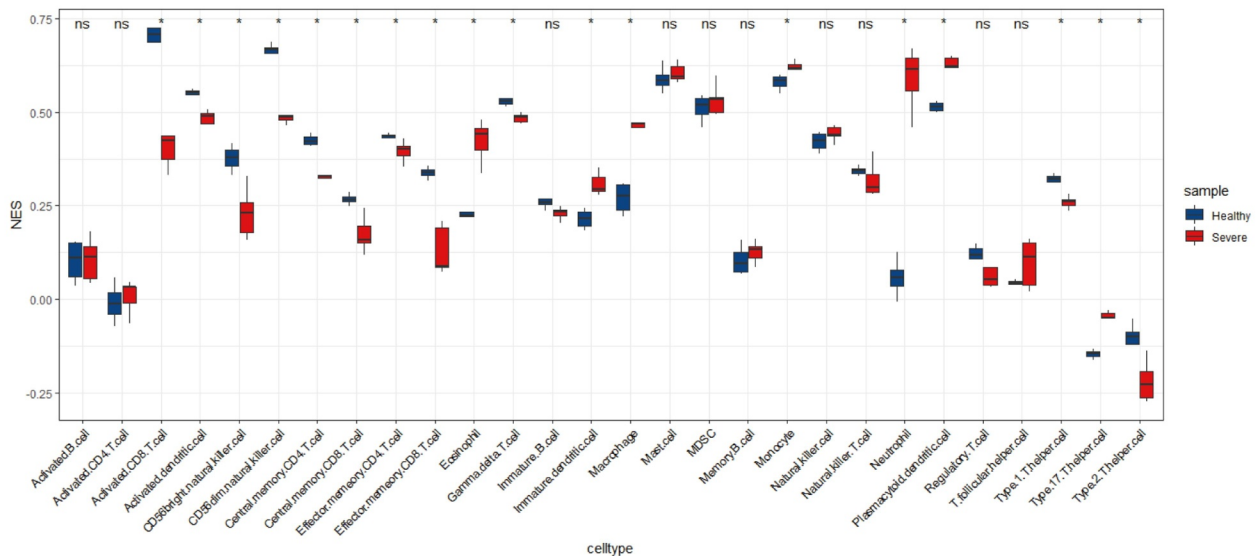
a



b

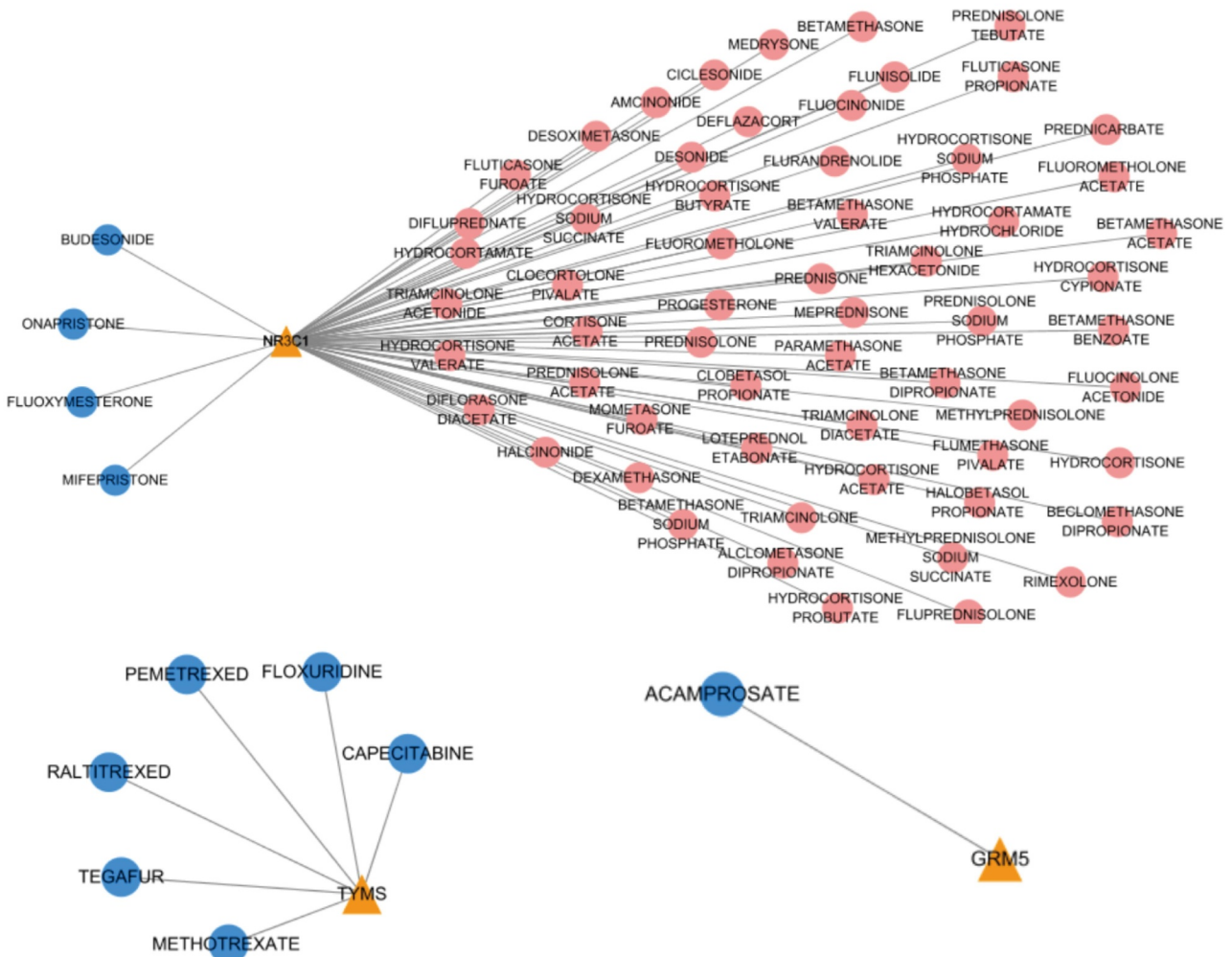


c

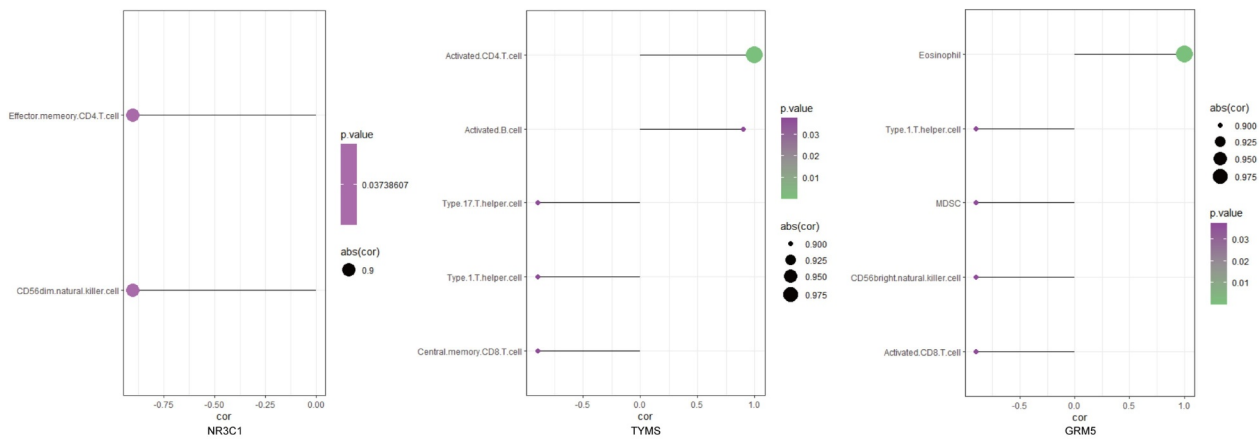


**Figure 7.** Infiltrating immune cells differences between Scovid-19 and control samples. (a) the heatmap summarizing the correlation of immune cells with Scovid-19 (right) and healthy (left) samples. (b) Heatmap of correlation showing the association between immune cells based on Scovid-19 (right) and healthy (left) samples. (c) Bar graph nested by violin plot showing infiltrating immune cells difference between Scovid-19 and healthy samples.

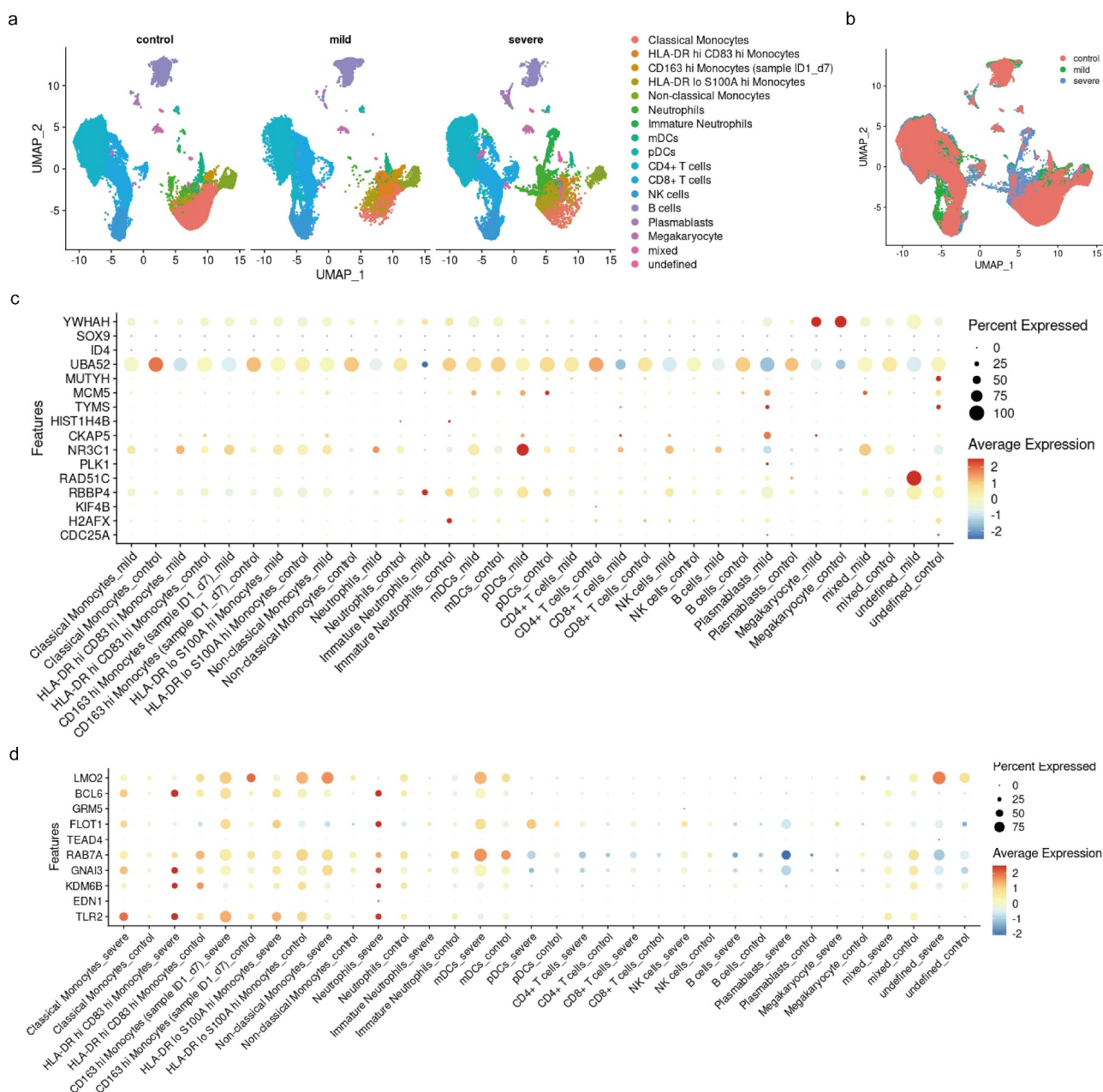
a



b



**Figure 8.** Drugs and immune cells related to NR3C1, TYMS and GRM5. (a) the potential drugs targeting NR3C1, TYMS and GRM5. Genes are shown as orange triangle. The pink circle represents the approved drug with activation effects, and the blue circle represents the inhibitor. (b) the bubble plots showing the significant associations between immune cells and NR3C1, TYMS and GRM5. Adjusted *P* value and number of genes are represented by color and size of each point for each immune cell.



**Figure 9.** Expression distribution of hub genes in each cell type based on cohort 1 from the Schulte-Schrepping dataset. (a) UMAP dimension reduction plot of control, Mcovid-19 and Scovid-19 samples. (b) UMAP plot of cells from all samples. Cells are colored for different groups. Feature dot plot showing mRNA expression levels of hub genes for (c) Mcovid-19 or (d) Scovid-19 in each cell type. UMAP, uniform manifold approximation and projection.

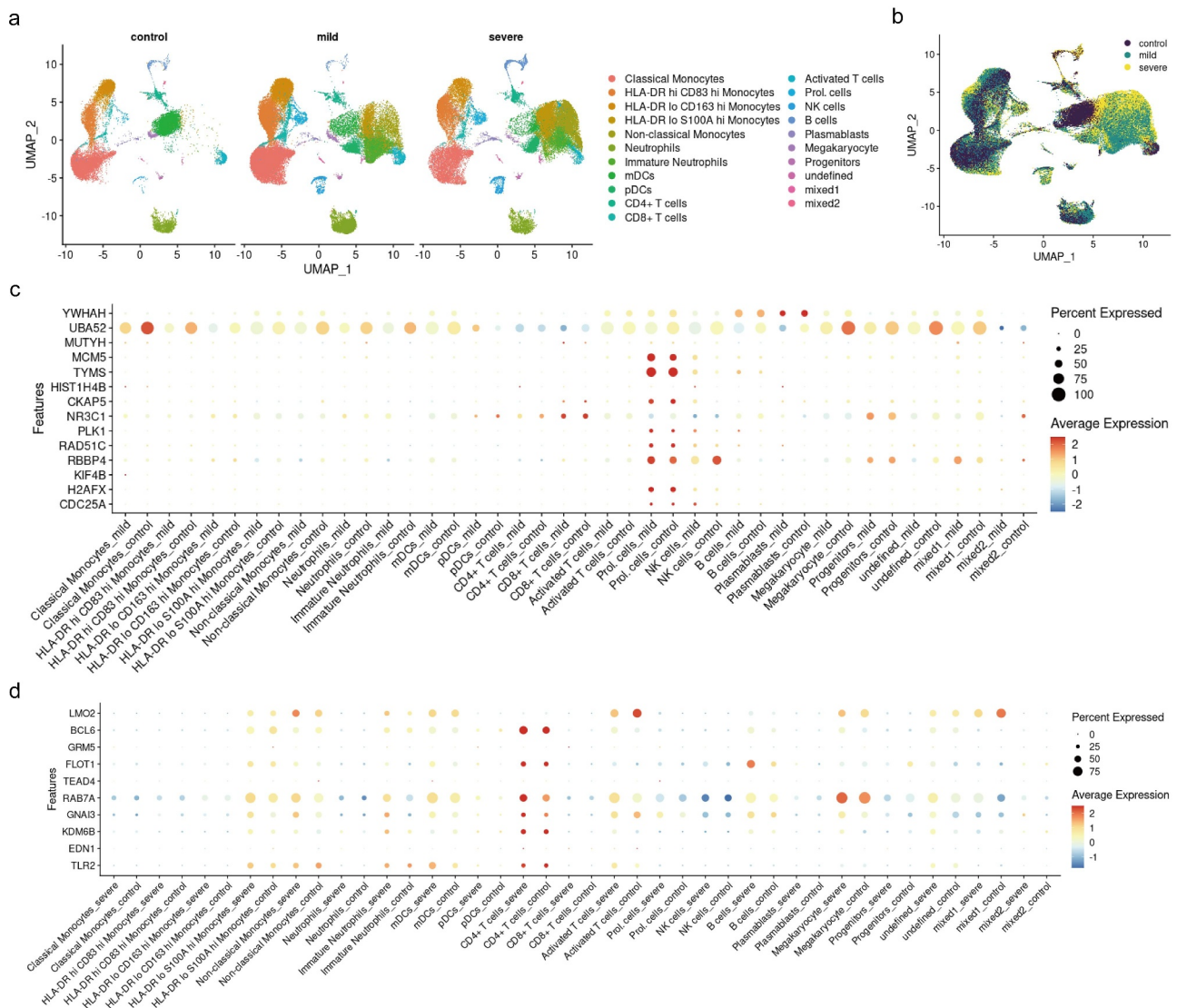
CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, NK cells, B cells, plasmablasts, megakaryocyte, mixed and undefined cells (Figure 9a). UMAP analysis of these cells showed that the heterogeneity in different groups was evident (Figure 9b). UBA52, one hub gene for mCOVID-19, was found to be detectably expressed in a high percent of cells across different cell types. The expression of NR3C1 was pDCs-specific in the mCOVID-19 group (Figure 9c).

Besides, BCL6, GNAI3, KDM6B and TLR2 yielded elevated expression in HLA-DR<sup>hi</sup> CD83<sup>hi</sup> monocytes and neutrophils in the sCOVID-19 group (Figure 9d). In cohort 2 of the Schulte-Schrepping dataset, 21 clusters of cells, including classical monocytes, HLA-DR<sup>hi</sup> CD83<sup>hi</sup> monocytes, HLA-DR<sup>lo</sup> CD163<sup>hi</sup> monocytes, HLA-DR<sup>lo</sup> S100A<sup>hi</sup> monocytes, non-classical monocytes, neutrophils, immature neutrophils, mDC, pDC, CD4<sup>+</sup>

T cells, CD8<sup>+</sup> T cells, activated T cells, prol. Cells, NK cells, B cells, plasmablasts, megakaryocytes, progenitors, undefined, mixed 1, and mixed 2 cells, were identified and annotated (Figure 10a). The heterogeneity of cells between healthy controls and COVID-19 patients was obvious (Figure 10a,b). Moreover, the expression levels of MCM5, TYMS, and RBBP4 were prol. cell-specific in both control and mCOVID-19 groups (Figure 10c), while the expression of BCL6, FLOT1, KDM6B, and TLR2 was CD4<sup>+</sup> cell-specific in the control and sCOVID-19 groups. Similar to the cohort 1, the expression of UBA52

and RAB7A in cohort 2 was detectable in a high percent of cells across the majority of cell types (Figure 10c,d).

As for the GSE216020 dataset, we identified up to 9 clusters of cells, which were subsequently annotated as naive CD4<sup>+</sup> T cell, CD8<sup>+</sup> T cell, CD14<sup>+</sup> monocyte, platelet, NK cell, B cell, plasma cell, FCGR3A<sup>+</sup> monocyte, and CD4<sup>+</sup> memory T cell (Figure S4A,B,C). Besides, the expression distribution of hub genes for mCOVID-19 (Figure S4D) and sCOVID-19 (Figure S4E) was shown. Notably, it was found that the expression of PRKCA was mainly distributed in Naive CD4<sup>+</sup>



**Figure 10.** Expression distribution of hub genes in each cell type based on cohort 2 from the Schulte-Schreppling dataset. (a) UMAP dimension reduction plot of control, Mcovid-19 and Scovid-19 samples. (b) UMAP plot of cells from all samples. Cells are colored for different groups. Feature dot plot showing mRNA expression levels of hub genes for (c) Mcovid-19 or (d) Scovid-19 in each cell type. UMAP, uniform manifold approximation and projection.

T cell (Figure S4E,F). In addition, PRKCA expression was significantly decreased in sCOVID-19 patients in comparison with healthy controls.

### **Discriminatory ability of hub genes expression for COVID-19 (severity)**

To further investigate the discriminatory ability of hub gene expression for COVID-19 severity more broadly, we applied machine learning modeling to rank the association between the hub genes in the context of COVID-19 severity. As the GSE157103 dataset included samples from both ICU COVID-19 patients and non-ICU COVID-19 patients, the expression levels of all hub genes for mCOVID-19 or sCOVID-19 were combined to distinguish ICU and non-ICU patients (Figure S5). We constructed the RF model with the least residuals. A total of 19 hub genes were ranked according to the gene importance based on RF model, which indicated that FLOT1, MUTYH, and RAB7A possessed a high priority within the model (Figure S5A). The 500 trees were chosen as the variables of the present model (Figure S5B). Subsequently, the accuracy of the RF model was determined by the ROC curve, which yielded an excellent AUC value (Figure S5C).

For the assessment of broadly discriminative ability of hub gene expression between healthy controls and COVID-19 patients in other datasets, we combined the PBMC samples from two datasets, namely GSE152418 and GSE206263. The expression levels of samples from these two datasets before and after normalization were visualized (Figure S6A, S6B). In addition, we conducted PCA of these samples before and after batch effect correction (Figure S6C, S6D). The heatmap indicated an ideally discriminative ability of the hub gene expression between healthy controls and COVID-19 patients for the GSE152418 dataset, but not for the GSE206263 dataset (Figure S6E). We provided detailed results of the comparison of hub genes expression between healthy controls and COVID-19 patients from the combined GSE152418 and GSE206263 datasets (Table S7). Specifically, the expression levels of 24 hub genes were detectable. Among them, logFC of eight hub genes (PLK1, TYMS, NR3C1, ID4, KDM6B, TEAD4, GRM5, and BCL6) was over 1, while logFC of another

eight hub genes (CDC25A, MUTYH, RBBP4, EDN1, TLR2, SOX9, RAB7A, and YWHAH) was between 0.5 and 1.

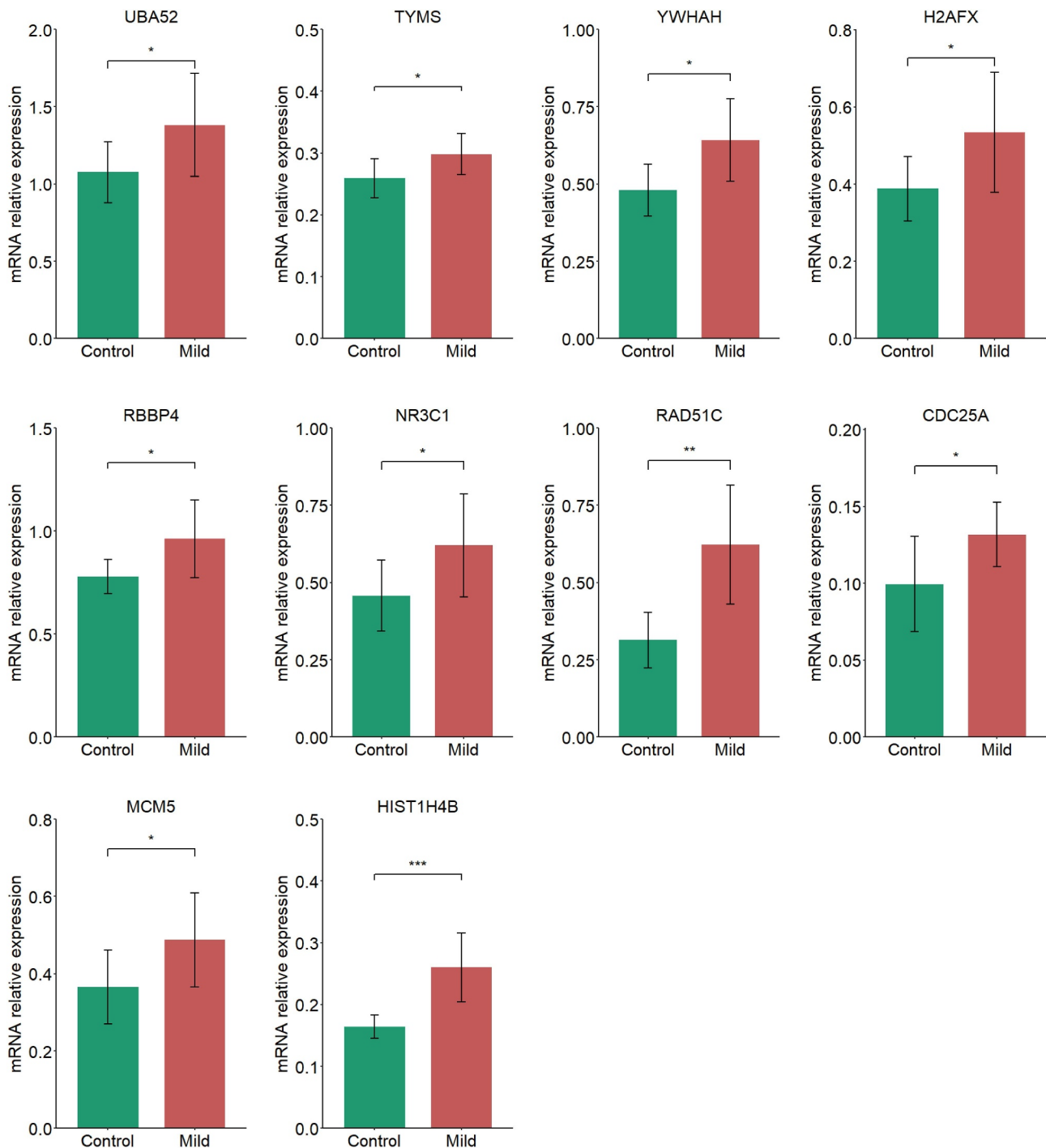
### **Validation of Mcovid-19 hub genes expression through qPCR**

As for the PBMC samples from eight healthy controls and eight mCOVID-19 patients at our institution, the mRNA expression levels of 16 mCOVID-19 hub genes were detected through qPCR. The mRNA expression values of 10 hub genes (CDC25A, H2AFX, RBBP4, RAD51C, NR3C1, HIST1H4B, TYMS, MCM5, UBA52, and YWHAH) were significantly elevated for mCOVID-19 patients when compared with healthy controls (Figure 11). On the other hand, the mRNA expression levels of CKAP5, MUTYH, KIF4B, and SOX9 were statistically insignificant between the two groups (Figure S7), while PLK1 and ID4 were undetectable.

### **Discussion**

COVID-19 has rapidly spread worldwide and caused a humanitarian disaster with social and economic consequences [27]. Due to the constant mutation of SARS-CoV-2, the efficacy of vaccines developed for COVID-19 tends to decrease [28]. The development of potential precision theranostics strategy requires a deeper understanding of the molecular mechanisms of COVID-19 of differing severity.

The current study identified many hub genes as potential molecular blood signatures for mCOVID-19 and sCOVID-19, respectively. As for mCOVID-19, hub genes included CDC25A, H2AFX, KIF4B, etc. Among these hub genes, PLK1 is located on the centrosome during mitosis and plays an essential role in cell division cycle. The overexpression of PLK1 has been reported in human tumors, such as digestive system tumor [29]. Previous study reported that NR3C1 was upregulated in mCOVID-19 group, but was downregulated in the sCOVID-19 group [30], which was similar to our results. SOX9 belongs to X-linked genes encoding proteins that recognize the sequence CCTTGAG along with other DNA-binding proteins [31]. Interestingly, X-linked



**Figure 11.** The comparison of Mcovid-19 hub genes expression in PBMCs between eight healthy controls and eight Mcovid-19 patients at our institution. \* $P$  value < 0.05; \*\* $P$  value < 0.01; \*\*\* $P$  value < 0.001. PBMCs, peripheral blood mononuclear cells.

genes such as FOXP3 and TLR7/8 and Y-linked genes such as SRY and SOX9 could be the reason for sex difference in COVID-19 lethality [32]. Besides, the ACE2<sup>+</sup>SOX9<sup>+</sup> cells are readily infected by SARS-CoV-2 pseudovirus [33].

In the case of sCOVID-19, hub genes included TLR2, EDN1, KDM6B, etc. TLR2, one of the Toll-

like receptor family, involves pathogen recognition and innate immunity activation. Besides, TLR2 plays important roles in sensing various pathogens, including bacteria, viruses, fungi, and parasites [33]. In addition, TLR2 appears to participate in early-phase NF- $\kappa$ B activation, and prophylactic administration of TLR2/6 agonist attenuates the



transmission of SARS-CoV-2 and protects subjects from COVID-19 [34]. RAB7A, a small GTPase, modulates vesicular transport and membrane trafficking [35]. RAB7A loss decreases viral entry through sequestering ACE2 receptor and altering endosomal trafficking [36]. BCL6 inhibits activatory and proliferative capacities of monocyte and macrophage, and the transcriptive capability of chemokines and IL-6 [37,38]. In addition, the upregulation of BCL6 was found in SARS-CoV, pneumonia, and severe acute respiratory syndrome [39]. PRKCA is a family member of the protein kinase C (PKC) and can be activated by calcium and diacylglycerol. Each PKC family member possesses specific cellular functions attributable to the specific expression distribution. Interestingly, PRKCA was found to be a naive CD4<sup>+</sup> T cell-specific gene and downregulated in sCOVID-19 patients (Figure S4E,F). It was reported that PRKCA is activated in respiratory disease originated from virus and/or bacteria [40]. PRKCA upregulates the outward transition of the ribonucleoprotein complex from the nucleus in affected cells [41]. The multiplication of SARS-Cov-2 can be accelerated through the activation of RKCA/mTORC2 and PI3K/PDK1 signalings [42]. T cell plays critical roles in the clearance of virus, and naive CD4<sup>+</sup> T cell serves as an important member of T cell that can be transformed into differentiated effector cells upon encountering foreign antigens [43]. On the other hand, the role of PRKCA in naive CD4<sup>+</sup> T cell and the mechanism of its downregulation in sCOVID-19 patients remains to be clarified in future studies.

The upstream regulatory network that controls the most critical module for mCOVID-19 or sCOVID-19 was also assessed in this study. PML, SIN3A, and ATF2 were the most critical TFs associated with mCOVID-19. PML is a nuclear protein that forms nuclear bodies and exhibits antiviral activity against both DNA and RNA viruses. SIN3A, a transcription regulator, regulates histone deacetylase activity and inhibits many tumor-related factors [44]. ATF2 is a member of AP-1 TF family, and can be activated through phosphorylating Thr69 and Thr71 once encountering extracellular stress. As for kinases, we identified CDKs and MAPKs most significantly enriched for

mCOVID-19. MAPKs are serine/threonine protein kinases and share common features with CDKs. MAPK14 represses cell cycle progression through downregulating cyclins and upregulating CDK inhibitors [45]. In the case of sCOVID-19, STAT3, RUNX1, and TP63 were identified as the most correlated TFs. STAT3 plays a role in inducing many proinflammatory and immune response genes [46]. The activation STAT3 induces COVID-19 to occur through upregulating cytokine storm [47]. Furthermore, RUNX1 is implicated in the angiogenesis and fibrosis, which are essential components of cell responses to SARS-CoV-2 infection.

Our results from the enrichment and immune cells demonstrated hyperproliferation of peripheral blood cells and activation of immune cells in COVID-19 patients. Concerning the infiltrating immune cells in the blood, upregulated abundance of neutrophils suggested important roles of neutrophils in COVID-19 pathogenesis. Conversely, decreased abundances of CD8<sup>+</sup> T cells were found for both mCOVID-19 and sCOVID-19 (Figures 6c and 7c). Previous study reported that CD8<sup>+</sup> T cells could express functional exhaustion molecules, such as TIM-3 [48]. In our current study, an upregulated abundance of macrophage was observed for sCOVID-19 but not mCOVID-19 (Figures 6c and 7c). In other words, it seems that macrophage-derived inflammation elevated to a large extent only when COVID-19 progresses to a severe stage. Future studies with more comprehensive study designs such as the addition of research into macrophage polarization are warranted to unveil this phenomenon.

The mainstay of current pharmacologic treatment is mainly composed of antiviral drugs, anti-SARS-CoV-2-neutralizing monoclonal antibodies, and immunomodulator agents [49]. Overall, antiviral drugs and monoclonal antibodies are suitable for patients at an early stage, and immunomodulator agents or combination therapy are appropriate for patients at an advanced stage [50]. Bioinformatic approaches provide the possibility of identifying putative drugs that may be helpful for future treatment of COVID-19. Herein, NR3C1, TYMS, and GRM5 were predicted to achieve the drug target. Among these three critical genes, NR3C1 had the largest number of protein

targeting drugs. NR3C1 may bind glucocorticoid response products within the promoter of glucocorticoid responsive gene for the activation of its transcription [51]. TYMS catalyzes the synthesis of thymidylate or dTMP from dUMP [52] and maintains the dTMP pool, a necessary element for DNA synthesis. GRM5 can activate phospholipase C and participate in multiple biological processes [53]. Among the protein targeting drugs, we found that drugs with the highest scores nearly all belong to naturally derived or synthetic glucocorticoids such as cortisone acetate, fluorometholone, and difluprednate.

Glucocorticoids are common ingredients of drugs for inflammation-related disorders due to anti-inflammatory and immunosuppressive properties. The pathophysiological features of sCOVID-19 are characterized by a hyperinflammatory state and elevated concentrations of inflammatory cytokines in bodily fluids. Recently, a large and open-label randomized controlled trial assessed the effects of dexamethasone in patients hospitalized with COVID-19, and found that dexamethasone users who received respiratory support had a significantly lower 28-day mortality [54]. In clinical management, dexamethasone has been routinely selected for hospitalized patients who require respiratory support. On the other hand, for patients with mild-to-moderate acute respiratory distress syndrome caused by COVID-19, it was reported that no clinical benefit was observed for high-dose administration of glucocorticoids [55]. In addition, moderate-quality evidence arising from a previous meta-analysis suggested that in patients with mCOVID-19, glucocorticoids was correlated with over three-fold higher risk of mortality [56]. In contrast, most of the glucocorticoids in this study were predicted based on NR3C1 and TYMS, which served as two hub genes originated from a comparison between mCOVID-19 and control samples. Several potential explanations may account for this discrepancy. The beneficial effect of glucocorticoids in COVID-19 patients relies on appropriate selections of the timing, dose, duration, and pharmacotyping of drugs. Although glucocorticoids are not contraindicated in the presence of mCOVID-19, high dose might be more harmful than helpful for such patients. Meanwhile, there may be the potential for mCOVID-19 patients to achieve therapeutic efficacy on the basis of short-

term low-dose administration of glucocorticoids. In the era of precision and personalized medicine, one critical issue is how to choose an appropriate amount of drugs based on the specific medical condition and the stratification of heterogeneous patients. Despite these promising findings, there are several limitations of this study. Firstly, the experimental validation of hub gene expression is only conducted between healthy controls and mCOVID-19 patients at our institution. Secondly, it should be noted that the small sample size in the control and case groups during network analysis is the major limitation to this exploratory study, which hinders the possibility of comparing mCOVID-19 and sCOVID-19 cases for the prediction of potential drugs. Future studies with large sample sizes are needed to confirm our results and to investigate severity-specific treatment targets for mCOVID-19 and sCOVID-19 cases.

## Conclusions

In conclusion, we identified critical modules, signaling pathways, TFs, kinases, immune cells, targeting drugs, hub genes, and their expression distributions in different cell types that may involve the pathogenesis of COVID-19 of differing severity, and fuel future advances in early diagnosis, patient stratification, and precision theranostics strategy of this notorious disease.

## Acknowledgements

We would like to thank Yali Zhou from Department of Microbiology, Faculty of Basic Medical Sciences, Guilin Medical University for the proofreading of our manuscript.

## Disclosure statement

No potential conflict of interest was reported by the author(s). All coauthors demonstrated no conflict of interest in connection with the present study.

## Funding

This study was supported by Jiangsu Funding Program for Excellent Postdoctoral Talent (2022ZB730) and Nanjing Important Science & Technology Specific Projects (2021-11005).

## Authors' contributions

Study design: N.Z and J.D.; Administrative support: X.X.D., M.K.Q., and C.Y.T.; Data analysis and visualization: L.L.T., M.H., H.F.F., H.Y.Z., and Y.Y. Manuscript writing: L.L.T., T.C. and N.Z.

## Ethical Approval and Consent to participate

The medical research ethics committee of Microbiology Laboratory, Nanjing Municipal Center for Disease Control and Prevention identified and approved study design of the current study.

## Availability of supporting data

The dataset used for the current study can be achieved from the GEO website.

## Authors' information

L.L.T., M.H., H.F.F., H.Y.Z., X.X.D., M.K.Q., C.Y.T., Y.Y., T.C., and N.Z. are from Nanjing Municipal Center for Disease Control and Prevention. In addition, T.C. is from Department of Urology, The First Affiliated Hospital of Nanjing Medical University.

## ORCID

Tong Chen  <http://orcid.org/0000-0001-6786-0310>

## References

- [1] Henry BM, Vikse J. Clinical characteristics of Covid-19 in China. *N Engl J Med.* 2020;382(19):1860–1861.
- [2] Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 cases from the chinese center for disease control and prevention. *JAMA.* 2020;323(13):1239–1242. doi: 10.1001/jama.2020.2648
- [3] Islam T, Rahman MR, Aydin B, et al. Integrative transcriptomics analysis of lung epithelial cells and identification of repurposable drug candidates for COVID-19. *Eur J Pharmacol.* 2020;887:173594. doi: 10.1016/j.ejphar.2020.173594
- [4] Arunachalam PS, Wimmers F, Mok CKP, et al. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science.* 2020;369(6508):1210–1220. doi: 10.1126/science.abc6261
- [5] Blanco-Melo D, Nilsson-Payant BE, Liu W-C, et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell.* 2020;181(5):1036–1045.e9. 2020 Cell. doi: 10.1016/j.cell.2020.04.026
- [6] Zhang Q, Meng Y, Wang K, et al. Inflammation and antiviral immune response associated with severe progression of COVID-19. *Front Immunol.* 2021;12:631226. doi: 10.3389/fimmu.2021.631226
- [7] Xu J, He B, Carver K, et al. Heterogeneity of neutrophils and inflammatory responses in patients with COVID-19 and healthy controls. *Front Immunol.* 2022;13:970287. doi: 10.3389/fimmu.2022.970287
- [8] Schulte-Schrepping J, Reusch N, Paclik D, et al. Severe COVID-19 is Marked by a Dysregulated Myeloid Cell Compartment. *Cell.* 2020;182(6):1419–1440.e23. doi: 10.1016/j.cell.2020.08.001
- [9] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf.* 2008;9(1):559. doi: 10.1186/1471-2105-9-559
- [10] Hasankhani A, Bahrami A, Sheybani N, et al. Differential co-expression network analysis reveals key hub-high traffic genes as potential therapeutic targets for COVID-19 Pandemic. *Front Immunol.* 2021;12:789317. doi: 10.3389/fimmu.2021.789317
- [11] Karami H, Derakhshani A, Ghasemigol M, et al. Weighted Gene Co-expression network analysis combined with machine learning validation to identify key modules and hub genes associated with SARS-CoV-2 Infection. *J Clin Med.* 2021;10(16):3567. doi: 10.3390/jcm10163567
- [12] Hernández Cordero AI, Li X, Yang CX, et al. Gene expression network analysis provides potential targets against SARS-CoV-2. *Sci Rep.* 2020;10(1):21863. doi: 10.1038/s41598-020-78818-w
- [13] Alarabi AB, Mohsen A, Mizuguchi K, et al. Co-expression analysis to identify key modules and hub genes associated with COVID-19 in platelets. *BMC Med Genomics.* 2022;15(1):83. doi: 10.1186/s12920-022-01222-y
- [14] Giroux NS, Ding S, McClain MT, et al. Differential chromatin accessibility in peripheral blood mononuclear cells underlies COVID-19 disease severity prior to seroconversion. *Sci Rep.* 2022;12(1):11714. doi: 10.1038/s41598-022-15668-8
- [15] Overmyer KA, Shishkova E, Miller IJ, et al. Large-scale multi-omic analysis of COVID-19 Severity. *Cell Syst.* 2021;12(1):23–40.e7. doi: 10.1016/j.cels.2020.10.003
- [16] Leek JT, Johnson WE, Parker HS, et al. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics.* 2012;28(6):882–883. doi: 10.1093/bioinformatics/bts034
- [17] Ritchie ME, Phipson B, Wu D, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015;43(7):e47. doi: 10.1093/nar/gkv007
- [18] Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43(Database issue):D447–52. doi: 10.1093/nar/gku1003
- [19] Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of

- biomolecular interaction networks. *Genome Res.* 2003;13(11):2498–2504. doi: [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303)
- [20] Chin CH, Chen S-H, Wu H-H, et al. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol.* 2014;8(Suppl 4):S11. doi: [10.1186/1752-0509-8-S4-S11](https://doi.org/10.1186/1752-0509-8-S4-S11)
- [21] Chen EY, Xu H, Gordonov S, et al. Expression2kinases: mRNA profiling linked to multiple upstream regulatory layers. *Bioinformatics.* 2012;28(1):105–111. doi: [10.1093/bioinformatics/btr625](https://doi.org/10.1093/bioinformatics/btr625)
- [22] Cotto KC, Wagner AH, Feng Y-Y, et al. Dgidb 3.0: a redesign and expansion of the drug–gene interaction database. *Nucleic Acids Res.* 2018;46(D1):D1068–d1073. doi: [10.1093/nar/gkx1143](https://doi.org/10.1093/nar/gkx1143)
- [23] Yu G, Wang L-G, Han Y, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS.* 2012;16(5):284–287. doi: [10.1089/omi.2011.0118](https://doi.org/10.1089/omi.2011.0118)
- [24] Hänzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf.* 2013;14(1):7. doi: [10.1186/1471-2105-14-7](https://doi.org/10.1186/1471-2105-14-7)
- [25] Butler A, Hoffman P, Smibert P, et al. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.* 2018;36(5):411–420. doi: [10.1038/nbt.4096](https://doi.org/10.1038/nbt.4096)
- [26] Wang H, Zhou L. Random survival forest with space extensions for censored data. *Artif Intell Med.* 2017;79:52–61. doi: [10.1016/j.artmed.2017.06.005](https://doi.org/10.1016/j.artmed.2017.06.005)
- [27] Pullano G, Di Domenico L, Sabbatini CE, et al. Underdetection of cases of COVID-19 in France threatens epidemic control. *Nature.* 2021;590(7844):134–139. doi: [10.1038/s41586-020-03095-6](https://doi.org/10.1038/s41586-020-03095-6)
- [28] Wu R, Wang L, Kuo H-CD, et al. An Update on Current Therapeutic Drugs Treating COVID-19. *Curr Pharmacol Rep.* 2020;6(3):56–70. doi: [10.1007/s40495-020-00216-7](https://doi.org/10.1007/s40495-020-00216-7)
- [29] Takahashi T, Sano B, Nagata T, et al. Polo-like kinase 1 (PLK1) is overexpressed in primary colorectal cancers. *Cancer Sci.* 2003;94(2):148–152. doi: [10.1111/j.1349-7006.2003.tb01411.x](https://doi.org/10.1111/j.1349-7006.2003.tb01411.x)
- [30] Park JH, Lee HK. Re-analysis of single cell transcriptome reveals that the NR3C1-CXCL8-neutrophil axis determines the severity of COVID-19. *Front Immunol.* 2020;11:2145. doi: [10.3389/fimmu.2020.02145](https://doi.org/10.3389/fimmu.2020.02145)
- [31] Bowles J, Schepers G, Koopman P. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Dev Biol.* 2000;227(2):239–255. doi: [10.1006/dbio.2000.9883](https://doi.org/10.1006/dbio.2000.9883)
- [32] Ghosh S, Klein RS. Sex drives dimorphic immune responses to viral infections. *J Immunol.* 2017;198(5):1782–1790. doi: [10.4049/jimmunol.1601166](https://doi.org/10.4049/jimmunol.1601166)
- [33] Zhang Z, Guo L, Lu X, et al. Clinical analysis and pluripotent stem cells-based model reveal possible impacts of ACE2 and lung progenitor cells on infants vulnerable to COVID-19. *Theranostics.* 2021;11(5):2170–2181. doi: [10.7150/thno.53136](https://doi.org/10.7150/thno.53136)
- [34] Dowling JK, Mansell A. Toll-like receptors: the swiss army knife of immunity and vaccine development. *Clin Transl Immunology.* 2016;5(5):e85. doi: [10.1038/cti.2016.22](https://doi.org/10.1038/cti.2016.22)
- [35] Guerra F, Bucci C. Multiple Roles of the Small GTPase Rab7. *Cells.* 2016;5(3). doi: [10.3390/cells5030034](https://doi.org/10.3390/cells5030034)
- [36] Daniloski Z, Jordan TX, Wessels H-H, et al. Identification Of required host factors for SARS-CoV-2 infection in human cells. *Cell.* 2021;184(1):92–105. e16. doi: [10.1016/j.cell.2020.10.030](https://doi.org/10.1016/j.cell.2020.10.030)
- [37] Yu RY, Wang X, Pixley FJ, et al. BCL-6 negatively regulates macrophage proliferation by suppressing autocrine IL-6 production. *Blood.* 2005;105(4):1777–1784. doi: [10.1182/blood-2004-08-3171](https://doi.org/10.1182/blood-2004-08-3171)
- [38] Toney LM, Cattoretti G, Graf JA, et al. BCL-6 regulates chemokine gene transcription in macrophages. *Nat Immunol.* 2000;1(3):214–220. doi: [10.1038/79749](https://doi.org/10.1038/79749)
- [39] Nain Z, Rana HK, Liò P, et al. Pathogenetic profiling of COVID-19 and SARS-like viruses. *Brief Bioinform.* 2021;22(2):1175–1196. doi: [10.1093/bib/bbaa173](https://doi.org/10.1093/bib/bbaa173)
- [40] Maharaj NP, Wies E, Stoll A, et al. Conventional protein kinase C- $\alpha$  (PKC- $\alpha$ ) and PKC- $\beta$  negatively regulate RIG-I antiviral signal transduction. *J Virol.* 2012;86(3):1358–1371. doi: [10.1128/JVI.06543-11](https://doi.org/10.1128/JVI.06543-11)
- [41] Marjuki H, Alam MI, Ehrhardt C, et al. Membrane accumulation of influenza A virus hemagglutinin triggers nuclear export of the viral genome via protein kinase c-mediated activation of ERK signaling. *J Biol Chem.* 2006;281(24):16707–16715. doi: [10.1074/jbc.M510233200](https://doi.org/10.1074/jbc.M510233200)
- [42] Partovian C, Ju R, Zhuang ZW, et al. Syndecan-4 Regulates Subcellular Localization of mTOR Complex2 and Akt Activation in a PKC $\alpha$ -dependent manner in endothelial cells. *Mol Cell.* 2008;32(1):140–149. doi: [10.1016/j.molcel.2008.09.010](https://doi.org/10.1016/j.molcel.2008.09.010)
- [43] Bousso P, Wahn V, Douagi I, et al. Diversity, functionality, and stability of the T cell repertoire derived in vivo from a single human T cell precursor. *Proc Natl Acad Sci U S A.* 2000;97(1):274–278. doi: [10.1073/pnas.97.1.274](https://doi.org/10.1073/pnas.97.1.274)
- [44] Das TK, Sangodkar J, Negre N, et al. Sin3a acts through a multi-gene module to regulate invasion in *Drosophila* and human tumors. *Oncogene.* 2013;32(26):3184–3197. doi: [10.1038/onc.2012.326](https://doi.org/10.1038/onc.2012.326)
- [45] Cargnello M, Roux PP. Activation and function of the MAPKs and their substrates, the MAPK-activated protein kinases. *Microbiol Mol Biol Rev.* 2011;75(1):50–83. doi: [10.1128/MMBR.00031-10](https://doi.org/10.1128/MMBR.00031-10)
- [46] Zhong Z, Wen Z, Darnell JE Jr. Stat3: a STAT family member activated by tyrosine phosphorylation in response to epidermal growth factor and interleukin-6. *Science.* 1994;264(5155):95–98. doi: [10.1126/science.8140422](https://doi.org/10.1126/science.8140422)
- [47] Jafarzadeh A, Nemati M, Jafarzadeh S. Contribution of STAT3 to the pathogenesis of COVID-19. *Microb Pathog.* 2021;154:104836. doi: [10.1016/j.micpath.2021.104836](https://doi.org/10.1016/j.micpath.2021.104836)

- [48] Zhang W, Zhao Y, Zhang F, et al. The use of anti-inflammatory drugs in the treatment of people with severe coronavirus disease 2019 (COVID-19): The Perspectives of clinical immunologists from China. *Clin Immunol.* 2020;214:108393. doi: [10.1016/j.clim.2020.108393](https://doi.org/10.1016/j.clim.2020.108393)
- [49] Coopersmith CM, Antonelli M, Bauer SR, et al. The Surviving Sepsis Campaign: Research Priorities for Coronavirus Disease 2019 in Critical Illness. *Crit Care Med.* 2021;49(4):598–622. doi: [10.1097/CCM.0000000000004895](https://doi.org/10.1097/CCM.0000000000004895)
- [50] Gandhi RT, Lynch JB, Del Rio C, et al. Mild or Moderate Covid-19. *N Engl J Med.* 2020;383(18):1757–1766. doi: [10.1056/NEJMcp2009249](https://doi.org/10.1056/NEJMcp2009249)
- [51] Lu NZ, Cidlowski JA. Translational regulatory mechanisms generate N-terminal glucocorticoid receptor isoforms with unique transcriptional target genes. *Mol Cell.* 2005;18(3):331–342. doi: [10.1016/j.molcel.2005.03.025](https://doi.org/10.1016/j.molcel.2005.03.025)
- [52] Yu MC, Yuan JM, Lu SC. Alcohol, cofactors and the genetics of hepatocellular carcinoma. *J Gastroenterol Hepatol.* 2008;1(Suppl 1):S92–7. doi: [10.1111/j.1440-1746.2007.05293.x](https://doi.org/10.1111/j.1440-1746.2007.05293.x)
- [53] Luo WY, Xing SQ, Zhu P, et al. PDZ Scaffold Protein CAL Couples with Metabotropic Glutamate Receptor 5 to Protect Against Cell Apoptosis and is a Potential Target in the Treatment of Parkinson's Disease. *Neurotherapeutics.* 2019;16(3):761–783. doi: [10.1007/s13311-019-00730-7](https://doi.org/10.1007/s13311-019-00730-7)
- [54] Horby P, Lim WS, Emberson JR, et al. Dexamethasone in Hospitalized Patients with Covid-19. *N Engl J Med.* 2021;384(8):693–704. doi: [10.1056/NEJMoa2021436](https://doi.org/10.1056/NEJMoa2021436)
- [55] Jamaati H, Hashemian SM, Farzanegan B, et al. No clinical benefit of high dose corticosteroid administration in patients with COVID-19: A preliminary report of a randomized clinical trial. *Eur J Pharmacol.* 2021;897:173947. doi: [10.1016/j.ejphar.2021.173947](https://doi.org/10.1016/j.ejphar.2021.173947)
- [56] Lu S, Zhou Q, Huang L, et al. Effectiveness and safety of glucocorticoids to treat COVID-19: a rapid review and meta-analysis. *Ann Transl Med.* 2020;8(10):627. doi: [10.21037/atm-20-3307](https://doi.org/10.21037/atm-20-3307)