



Tracing the cell-type-specific modules of immune responses during COVID-19 progression using scDisProcema



Anyao Li ^{a,b,1}, Jihong Yang ^{a,c,d,1}, Jingyang Qian ^{a,b}, Xin Shao ^{a,b}, Jie Liao ^{a,b}, Xiaoyan Lu ^{a,b,c}, Xiaohui Fan ^{a,b,c,*}

^a College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China

^b Innovation Center of Yangtze River Delta, Zhejiang University, Jiaxing 314102, China

^c Innovation Center in Zhejiang University, State Key Laboratory of Component-Based Chinese Medicine, Hangzhou 310058, China

^d Zhang Boli Intelligent Health Innovation Lab, Hangzhou 311100, China

ARTICLE INFO

Article history:

Received 14 March 2022

Received in revised form 30 June 2022

Accepted 30 June 2022

Available online 2 July 2022

Keywords:

COVID-19

scRNA-seq

WGCNA

Disease progression

ABSTRACT

COVID-19 has caused severe threats to lives and damage to property worldwide. The immunopathology of the disease is of particular concern. Currently, researchers have used gene co-expression networks (GCNs) to deepen the study of molecular mechanisms of immune responses to COVID-19. However, most efforts have not fully explored dynamic changes of cell-type-specific molecular networks in the disease process. This study proposes a GCN construction pipeline named single-cell Disease Progression cellular module analysis (scDisProcema), which can trace dynamic changes of immune system response during disease progression using single-cell data. Here, scDisProcema considers changes in cell fate and expression patterns during disease development, identifying gene modules responsible for different immune cells. The hub genes are screened for each module by the specific expression level and the intercellular connectivity of modules. Based on functional items enriched by each gene module, we elucidate the biological processes of different cells involved in disease development and explain the molecular mechanisms underlying the process of cell depletion or proliferation caused by disease. Compared with traditional WGCNA methods, scDisProcema can make more convenient use of the heterogeneity information provided by scRNA-seq data and has great potential in exploring molecular changes during disease progression and organ development.

© 2022 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Coronavirus disease 2019 (COVID-19), caused by Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2), has been spreading progressively since the end of 2019. As a result, by February 2022, the cumulative number of confirmed cases and the death toll had exceeded 430 million and 50 million, respectively, posing a significant threat to the lives and health of people worldwide. Immune disorders caused by COVID-19 are highly associated with poor outcomes [1], and therefore, immunological issues related to the disease have received widespread attention. In general, during the development and recovery of the disease, the proportion of immune cells and the expression level of inflammatory factors would have different dynamic changing trends or

degrees [2,3]. Therefore, deepening the understanding of these dynamic changes is of great significance to identify factors that contribute to the deterioration of COVID-19 and its sequelae during patients' recovery process.

Single-cell RNA sequencing (scRNA-seq) technology [4] has made great progress, significantly improving the understanding of cell heterogeneity and cell-cell communication [5] masked by bulk sequencing. At present, this technique has been applied in transcriptome analysis of SARS-CoV-2 infected cells [6–8]. An integrated scRNA-seq dataset [9], including 1.46 million single cells, characterized the immune landscape of COVID-19 patients. The authors collected samples of patients at different progression or convalescent stages. Based upon the severity, five states were defined, namely mild/moderate progression (MP), severe/critical progression (SP), mild/moderate convalescence (MC), severe/critical convalescence (SC), and healthy control (H) state. Such a comprehensive data set provides the fundamental for further analysis.

With the popularization of scRNA-seq technology, a variety of network analysis algorithms and tools have been developed,

* Corresponding author at: College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China.

E-mail address: fanxh@zju.edu.cn (X. Fan).

¹ These authors contribute equally to this work.

among which the weighted gene co-expression network analysis (WGCNA) algorithm [10] is widely used for the detection of key genes related to phenotypic traits. WGCNA identifies co-expressed gene modules, in which the most closely related genes are identified as hub genes, which are usually functionally important and represent major roles in diseases, facilitating the identification of diagnostic biomarkers and potential drug targets. WGCNA has also been used in COVID-19 research to predict important genes, or gene modules [11,12]. However, current studies lack exploring the dynamic change, which refers to the change of gene-expression patterns underlying the cell proportion fluctuation during the disease development. Therefore, we proposed the scDisProcema pipeline to capture the dynamic change in this study.

First, we inferred different immune cell types' gene co-expression network modules during disease progression. Then, we identified the key modules underlying the dynamic changes of immune regulation in disease processes by correlating module expression patterns with cell proportion change. Furthermore, hub gene and functional enrichment analysis of the key module elucidated the underlying biological processes. The above steps traced the dynamic proportion changes of immune cells during the disease development and recovery process and revealed the gene expression patterns corresponding to these cellular changes. The pipeline is downstream of quality control and cell clustering, which are performed by using the R package Seurat [13], and the Seurat object can be directly utilized by scDisProcema. An overview of the present work can be seen in Fig. 1A.

2. Materials and methods

2.1. Cell preprocessing, clustering and annotation

Data filtering, normalization, dimensionality, and clustering were performed by R package Seurat (4.0.4). Cells with less than 200 detected genes or higher than 5000 detected genes and higher than 25% mitochondrial gene ratio were removed. The raw read counts were normalized with the NormalizeData function. The top 2000 highly variable genes (HVGs) were identified by the FindVariableFeatures function, then scaled using the ScaleData function. The RunPCA function was used to perform principal component analysis, and then the RunHarmony function of R package harmony [14] (0.1.0) was used to reduce the batch effect. The top 30 principal components (PCs) were used to perform t-distributed stochastic neighbor embedding (tSNE) to embed the dataset into two dimensions. Then FindNeighbors and FindClusters functions were adopted to construct the nearest neighbor graph and cluster cells. Recognized marker genes were used to annotate the cell type for each cluster. We annotated nine cell types, namely B, CD4 + T, CD8 + T, monocyte-derived dendritic cell (mDC), megakaryocyte (Mega), monocyte or macrophage (Mono/Macro), natural killer cell (NK), plasmacytoid dendritic cell (pDC), $\gamma\delta$ T cell ($\gamma\delta$ T). Function FindAllMarkers was used to identify differentially expressed genes (DEGs) of each cell cluster, and we set parameters “only.pos” as TRUE, “logfc.threshold” as 0.5 and “min.pct” as 0.5. Details of parameters can be seen in Appendix A.

2.2. The scDisProcema workflow

The workflow of the scDisProcema can be divided into three steps: 1) state-gene matrix as input to WGCNA for gene co-expression network module searching; 2) module dynamic analysis for obtaining the degree of dynamic change and relevance to cells over the course of disease; 3) key module identification for different cell types.

2.2.1. Gene co-expression network module searching

Based on feature selection and cell clustering results, we then used R package WGCNA (version. 1.70–3) to infer the gene co-expression network. The average expression values of each HVGs of different cell types in various states were calculated as input data. First, the genes with median absolute deviation greater than 0.01 were retained and then the goodSamplesGenes function was used to screen out samples and genes with zero-variance. To construct a scale-free network, the pickSoftThreshold function was used to select the best soft threshold and set the filtering criteria to be 0.9. The network was constructed by the blockwiseModules function, and we set the “maxBlockSize” to be the total number of genes, “minModuleSize” as 20, “deepSplit” as 4, “mergeCutHeight” as 0.1, and numericLabels as TRUE to name modules with colors. The module eigengene (ME) was calculated to represent the expression level of the module. Intramodular connectivity (KWithin) was calculated for each module. Details of parameters can be seen in Appendix A.

2.2.2. Module dynamic analysis

With cell type annotation, we got n cell types ($C_i, i = 1, \dots, n$) and m gene modules ($M_j, j = 1, \dots, m$). For C_i and M_j , we first calculated the cell-type-specific correlation coefficient score ($S_{(CC)}$) by the following formula:

$$S_{(CC)ij} = \frac{\text{cov}(\text{Pro}_i, \text{ME}_j)}{\sigma_{\text{Pro}_i} \sigma_{\text{ME}_j}} \quad (1)$$

$S_{(CC)ij}$ represents the Pearson correlation between ME and cell proportion changes during disease progression. Pro_i and ME_j are vectors containing cell proportion values of C_i and module eigengene values of M_j in C_i across all states, respectively.

Then we calculated the activity fluctuation score ($S_{(AF)}$) as follows to measure dynamic change degree of M_j across all states:

$$S_{(AF)j} = \sqrt{\frac{\sum_{s=1}^S ((\text{ME}_j)_s) - \bar{\text{ME}}_j}{S-1}} \quad (2)$$

S represents the number of states (here is 5). The $(\text{ME}_j)_s$ means the ME value of M_j in the state s ($s = 1, \dots, 5$).

2.2.3. Key module identification

The product of (1) and (2) was calculated and then scaled using z-score normalization, namely $Z_{\text{score}}(S_{(CC)} \times S_{(AF)})$. The absolute value of the above-scaled score was used as module significance score ($S_{(MS)}$). Thus, we obtained a matrix of n cell types and m gene modules, the elements of which are $S_{(MS)ij}$. A large $S_{(MS)}$ value indicates a significant dynamic change of module expression across different states and a strong correlation with the cell ratio variation. The module with the largest $S_{(MS)}$ value was identified as its key module for each cell type.

2.3. Hub genes selection

Hub genes are genes with high correlation in the module. Here, genes that meet the following two requirements are defined as hub genes: 1) specifically expressed in corresponding cell types; 2) owned high connectivity in the network module. Thus, markers with the top five intercellular connectivity were chosen. If no member in the key module belongs to the marker gene set, we used another strategy to select hub genes. That is, the log2 fold change (log2FC) of a gene's mean expression level in the corresponding type against other types was used to measure the specificity, and the genes with top five KWithin values and absolute log2FC larger than 0.5 were set as the hub genes.

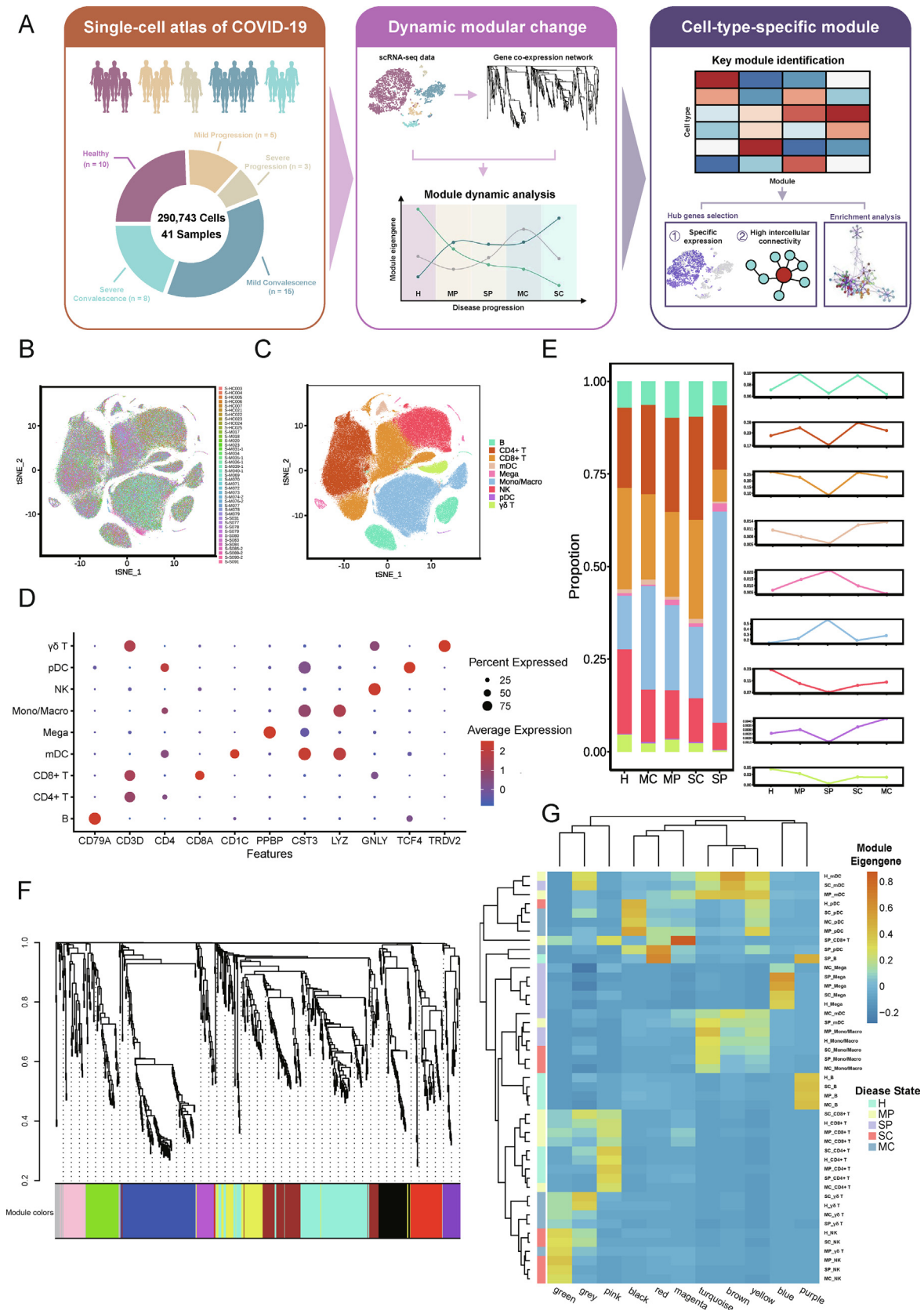


Fig. 1. The overall framework and WGCNA results. (A) The overall framework. (B-C) The t-SNE projection of single-cell data by samples (B) and by cell types(C). (D) The dot plot of marker genes of nine major cell types. (E) The bar plot and the line plots show the variation trend of each cell type during disease progression and recovery. (F) WGCNA hierarchical clustering results. (G) The heatmap shows module eigengene values across different states in nine cell types. tSNE for t-Distributed Stochastic Neighbor Embedding; WGCNA for weighted gene co-expression network analysis.

2.4. Network visualization

The network was visualized using the Cytoscape (version. 3.9.0) software [15]. We calculated the log₂ expression change of genes in the disease state compared with healthy state for each cell type as log₂FC. The edge represents the association between two genes in the visualized network, and the nodes represent genes. We visualized the size of nodes and their color according to the values of KWithin and log₂FC, respectively. The color of nodes in the control network of healthy state was grey, and upregulated and downregulated genes were mapped with red and blue, respectively. When visualizing the PPI network built by Metascape [16], we show the label of each node using gene name and moderately adjust the tightness of the network.

3. Results

3.1. Sample selection and clustering analysis of COVID-19 samples under different states

The dynamic change of immune status of SARS-CoV-2 infection has been widely concerned. The landscape in the *Cell* paper mentioned above [9] gave a relatively complete picture of this dynamic due to the comprehensive samples and high-quality data, improving our knowledge of immunology in disease outcomes. Thus, we extracted 41 samples covering all states for single-cell level profiling (Fig. 1. A-B) and gene network analysis. Furthermore, to reduce the batch effects and other possible influencing factors, such as effects of different experimental operations and comorbidities on the expression profile, we selected the fresh peripheral blood samples from patients without other comorbidities. Detailed information can be seen in Table 1 and Supplementary Table *1*.

A total of 290,743 cells were divided into 27 clusters by tSNE clustering analysis, covering nine cell types in the immune system (Fig. 1C), including B cells (CD79A), CD4 + T cells (CD3D & CD4), CD8 + T cells (CD3D & CD8A), monocyte-derived DCs (CD1C), megakaryocytes (PPBP), monocytes or macrophages (CST4 & LYZ), natural killer cells (GNLY), plasmacytoid DCs (TCF4) and $\gamma\delta$ T cells (CD3D & TRDV2) (Fig. 1D), and their proportions varied by states (Fig. 1E). All the above cell types were annotated based upon the cell type information in reference [9]. Generally, the lymphocytes, such as T cells, B cells, and pDCs, had the lowest proportion in SP state and rose gradually in the convalescence period. However, the cell ratio of myeloid cells except mDCs showed the opposite dynamic trend.

The cell proportion changes can partially reflect the effect of virus infection on immune cells at different stages [2,3], which may be closely related to the occurrence of inflammatory reaction and secretion of cytokines.

3.2. Gene co-expression module construction of COVID-19 processes

Our network is built based on the WGCNA tool. The input data was the cell states-HVGs expression matrix. After quality control, 705 genes were selected and divided into 11 gene modules (Fig. 1F). The heatmap (Fig. 1G) displayed the ME values across cell types, visualizing the changes of module expression during the progression and recovery. In addition, the hierarchical clustering of modules and cells was carried out, respectively. Cell types at different states were clustered together, suggesting that the co-expression modules identified by WGCNA were cell-type-specific.

Table 1
COVID-19 sample information.

State	# of samples	# of cells
Healthy (H)	10	58,058
Mild progression (MP)	5	44,201
Severe progression (SP)	3	28,110
Mild convalescence (MC)	15	120,744
Severe convalescence (SC)	8	39,630
Total	41	290,743

3.3. Key module identification of COVID-19 processes

We identified the module responsible for the patients' immune changes during COVID-19 processes by correlating cell proportion change with module expression pattern change. For each cell type, we considered the degree of dynamic change in module expression and the correlation with the change in cell proportions. As a result, the module with the highest module significance score, namely $S_{(MS)}$, was identified as the key module of the cell type. The complete process can be seen in **subsection 2.2 and Fig. 2**, and the results are shown in the heatmap (Fig. 3A). As a result, the red module was identified as the key module for B cell and pDC, magenta module for CD8 + T cells, blue module for megakaryocytes, brown module for monocytes or macrophages, and grey module for CD4 + T, mDC, NK, and $\gamma\delta$ T cells. The corresponding relationship between modules and cell types, as well as the hub genes and main functions of each module, are discussed below, which are all shown in Table 2.

There are some reliable cell type-specific marker datasets, for example, PanglaoDB Augmented (2021) [17], CellMarker Augmented (2021) [18], and CellMatch (<https://github.com/ZJUFan-Lab/scCATCH/tree/master/data>, which is incorporated into the tool scCATCH [19]). Enrichr [20] is a comprehensive interactive enrichment tool that can use these datasets to perform robust type annotation of gene modules. Therefore, we used Enrichr to evaluate whether the enrichment results in the key modules could match the corresponding cell types. The results were shown in Fig. 3B-F. In the enriched items of the red module (Fig. 3B), the second and fifth terms were B cells and plasmacytoid dendritic cells, respectively. Macrophages and monocytes were ranked as the brown module's top three and four enriched items (Fig. 3E), respectively. Additionally, different cell types could share the same key modules, indicating a functional connection between these cell types.

In general, the matched enriched terms of identified modules indicate the rationality of identified key modules.

3.4. Dynamic modular change of key modules in COVID-19 processes

We used the Cytoscape to depict the above key modules. To reflect the degree of module dynamic change, we calculated the expression change under certain disease states compared with the healthy state (see details in **subsection 2.4**). The networks in Fig. 4 visualized the co-expression pattern of each module across five states. Different cell types could share the same module, and the expression trend could also be similar (Supplementary Fig. *1*). Genes in red, magenta, and blue modules were inclined to upregulate as the disease worsened and downregulated during recovery (Fig. 4A-D). In contrast, the brown and grey modules showed more complex fluctuation (Fig. 4E-I). These modules were generally upregulated in mild patients but remarkably downregulated in severe ones, indicating the difference between patients with mild and severe syndromes.

The dynamic change trend of gene modules could be the same as that of cell proportion, e.g., megakaryocytes and mDCs. In con-

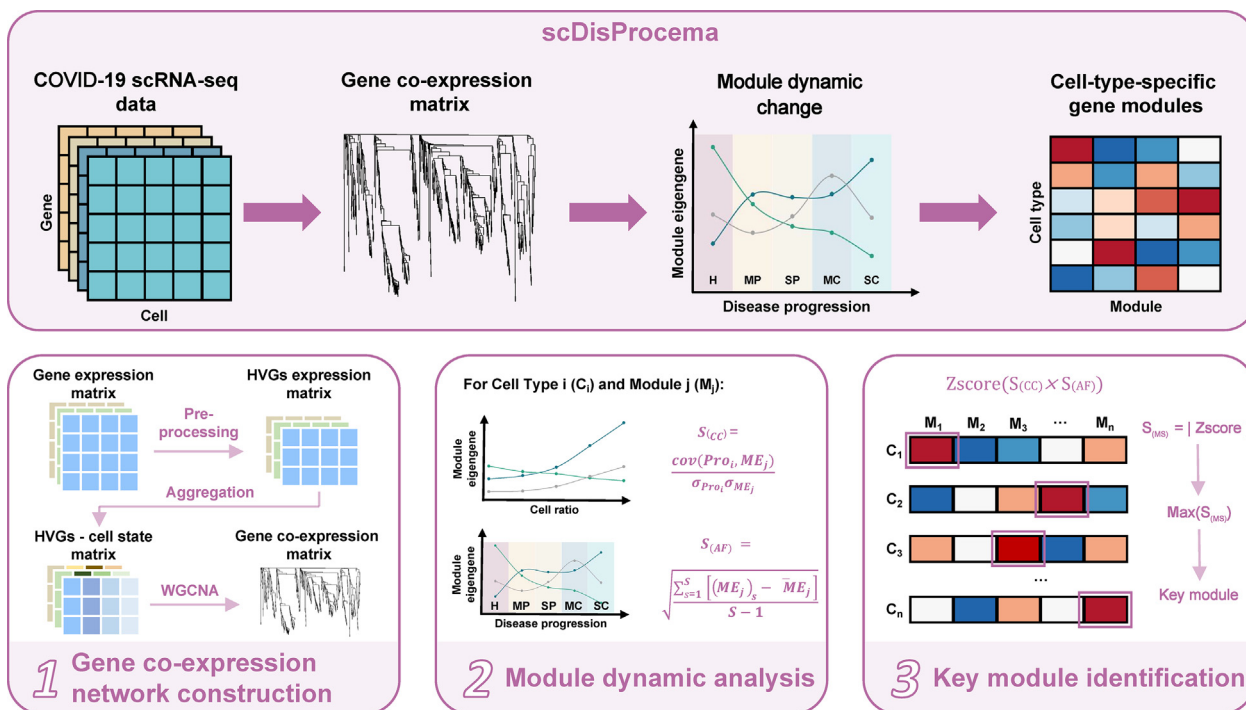


Fig. 2. The workflow of scDisProcema. The workflow consists of three steps, namely gene co-expression network module construction, module dynamic analysis, and key module identification.

trast, some are opposite, e.g., B cells, pDCs, CD8 + T cells, suggesting that biological processes involved in these gene modules may be responsible for either cell proliferation or depletion, which can be further verified. Moreover, the dynamic change patterns of different cell types in the same module were correlated, suggesting potential interactions among different cell types (Supplementary Fig. *1*).

3.5. Hub genes in the COVID-19

Hub genes are generally the genes that can represent the characteristics of the networks, which are often defined according to genes' intermodular connectivity [21–23]. We selected intersected genes between module hubs and differentially expressed marker genes considering both cell-specificity and connectivity (see **subsection 2.3** for more details). All hub genes are shown in Table 2. We examined the expression of these hub genes and found that most of them were upregulated explicitly in their corresponding cell types, confirming the rationality of the defined hub genes (Supplementary Fig. *2*). However, no hub gene in the magenta module overlapped with the corresponding cell type marker genes, suggesting that such a module might not directly determine the cell type. In contrast, hub genes in the magenta module might regulate CD8 + T cell function through cell cycle and DNA damage processes (see Table 2 and below Fig. 5B for details).

3.6. Module function and protein–protein interaction

To explore the biological processes involved in gene modules, we utilized the Metascape software, a comprehensive functional enrichment tool to conduct functional enrichment (Fig. 5A–E; Table 2) and protein–protein interaction (PPI) network topology (Fig. 5F–J) analysis. Particularly, connected network components were identified using the “Molecular Complex Detection” (MCODE) algorithm [24], and the top three function terms of the components were shown.

In the present study, we showed that functions of red module genes were directly or indirectly related to ER stress (Fig. 5A). Endoplasmic Reticulum Stress (ER stress) is a protective cellular response caused by the accumulation of misfolded proteins in response to stimuli such as viral infection [25]. GRP94 (human gene HSP90B1) is a key marker of this intracellular process [26]. ER stress is also associated with apoptosis, autophagy, inflammatory response, angiogenesis, and other biological processes. MCODEs identified from PPI networks were also associated with ER protein folding and signal recognition [27] (Fig. 5F). In addition to HSP90B1, other two hubs of pDC, SEC11C and SPCS3, which both encode signal peptidase complex subunits, are also involved in endoplasmic reticulum signal transduction [28]. In addition, a large number of genes in the red module, including the unique hubs of B cells, namely IGHM and ISG20, are related to the activation of B cells, which might be one of the pro-inflammatory responses of ER stress. Invasion of SARS-CoV-2 could activate ER stress-related pathways, leading to cellular inflammatory response and cell apoptosis, consistent with the upregulation of the red module in the severe stage.

The magenta module was mainly involved in the cell cycle, apoptosis, damage repair of genetic material (Fig. 5B). We found that the genes enriched in cell division and DNA replication were also involved in the DNA damage and repair process, suggesting cells were damaged or apoptotic, consistent with the depletion of CD8 + T cells in the SP state [29]. Most of the involved genes were histone genes and MCM family genes, which have been shown to play an important role in regulating the stability of DNA structure [30,31]. There was a strong interaction between these two kinds of genes, and both were detected as independent MCODEs in the PPI network (Fig. 5G).

The key module of megakaryocytes is module blue, and most genes of which were involved in coagulation function (Fig. 5C). Besides, these genes are also implicated in human cytomegalovirus (HCMV) infection and cell interactions, which are critical for platelet adhesion [32]. A large number of interacting histone genes were

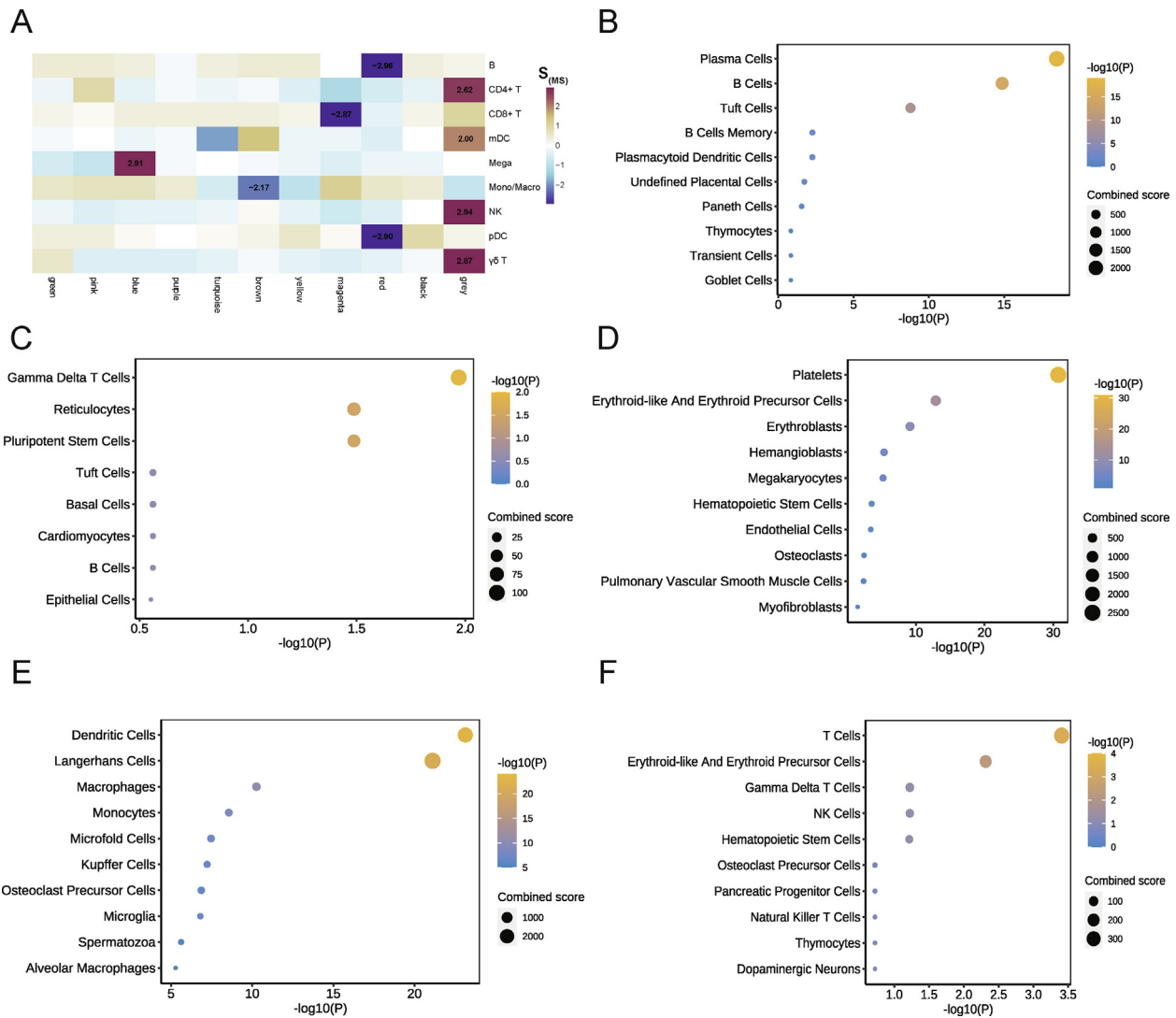


Fig. 3. The results of key module identification. (A) Modules with maximum or minimum scaled values are marked in the heatmap and identified as key modules. Module red is the key module for B cell and pDCs, magenta is for CD8 + T cells, blue is for megakaryocytes, brown is for monocytes/macrophages, grey is for CD4 + T, mDCs, NK, and γδ T cells. (B-F) The cell type enrichment results inferred by enrichR platform. Enrichment results for module red (B), module magenta (C), module blue (D), module brown (E), and module grey (F). The color represents the $-\log(p\text{-value})$ and the dot size represents the combined score inferred by enrichR. pDC for plasmacytoid dendritic cells; NK for natural killer; mDC for monocyte-derived dendritic cells. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Hub genes of key modules.

Module	Module IX (Red)		Module VII (Magenta)		Module III (Blue)	Module VI (Brown)	Module XI (Grey)			
Cell Type	B cell	pDC	CD8 + T		Mega	Mono/Macro	CD4 + T	mDC	NK	γδ T
Hub Genes	IGHM	HSP90B1	HIST1H1B		RGS18	CST3	HBA1	TRGV4	HBB	DUSP2
	ISG20	SEC11C	DHFR		MMD	ANXA5	TRGV4	HSPA1A	HSPA1B	TNFAIP3
	HSP90B1	SPCS3	PCNA		RUFY1	ANXA2	HBA2	TRGV5	HBA2	CXCR4
		MYDGF	SMC2		CA2	PPT1	TRGV5	HSPA1B	HSPA1A	
Main Function	ER stress, activation of B cells		Cell cycle, apoptosis, damage repair		PLA2G12A	CTSH	HSPA1B	CXCR4	TRGV5	
					Coagulation, HCMV infection	Immunity	Oxygen transport and oxidative stress			

enriched in items related to HCMV infection (Fig. 5H). This is possibly because epigenetic modifications such as histone acetylation are closely related to HCMV infection and reactivation [33], and such modifications also occur during SARS-CoV-2 infection [34]. Megakaryocytes, on the other hand, could release histone-rich platelets to promote clotting [35,36], which may be one of the mech-

anisms for thrombosis in COVID-19 patients [37]. Furthermore, the development of abnormal coagulation (DIC) and thrombosis is highly correlated with the poor prognosis of COVID-19 [38,39]. Here, the dynamic expression of blue module was consistent with the change of the proportion of megakaryocytes, which might reflect the degree of clotting in the patient.

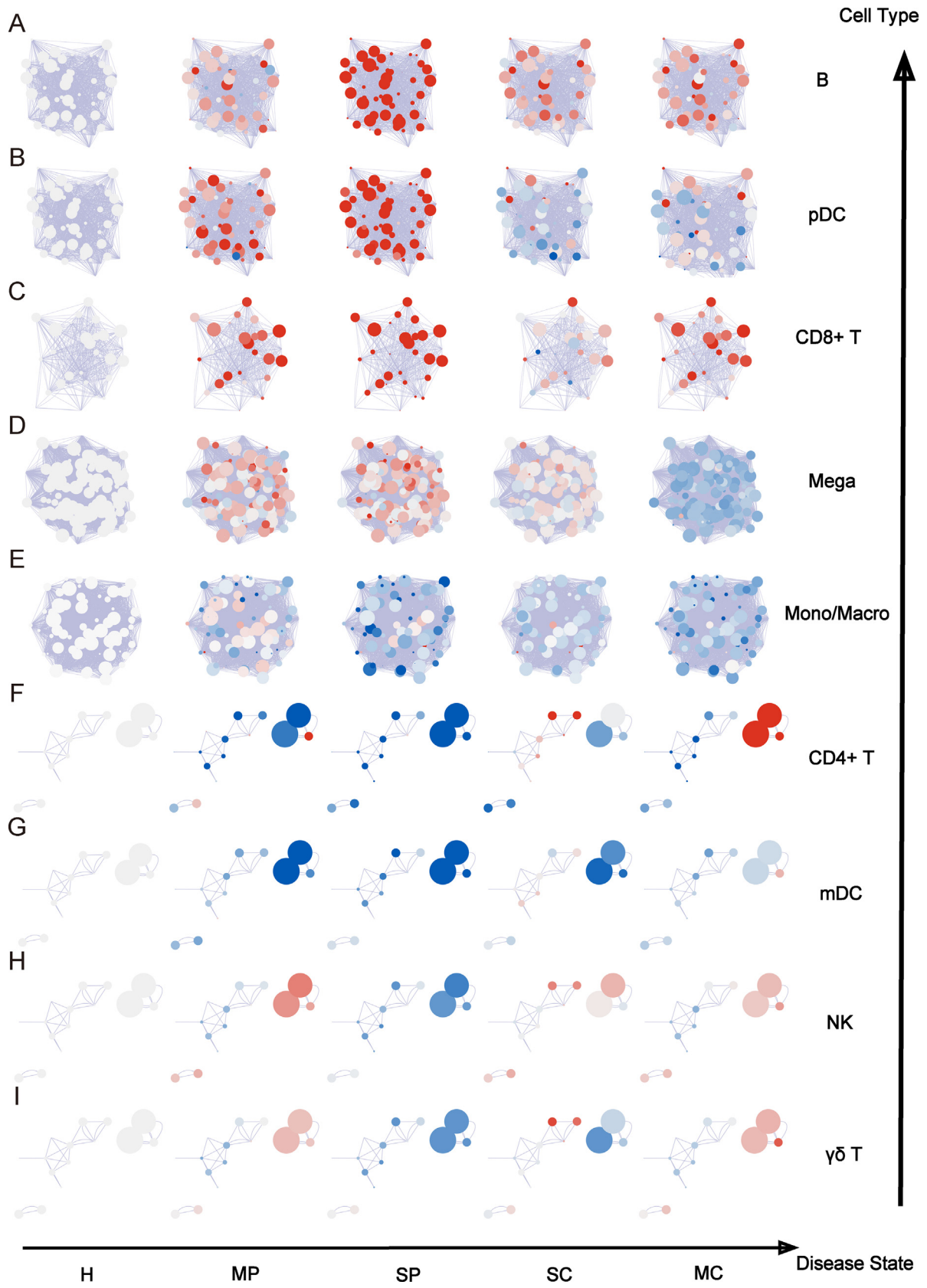


Fig. 4. The results of module dynamic analysis. (A-I) represents the network modules of B, pDC, CD8 + T, megakaryocyte, monocyte/macrophage, CD4 + T, mDC, NK, and $\gamma\delta$ T cells, respectively. Edges represent the association, and the nodes represent the genes. The node color indicates whether the gene expression level is higher (red) or lower (blue) than the control group. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

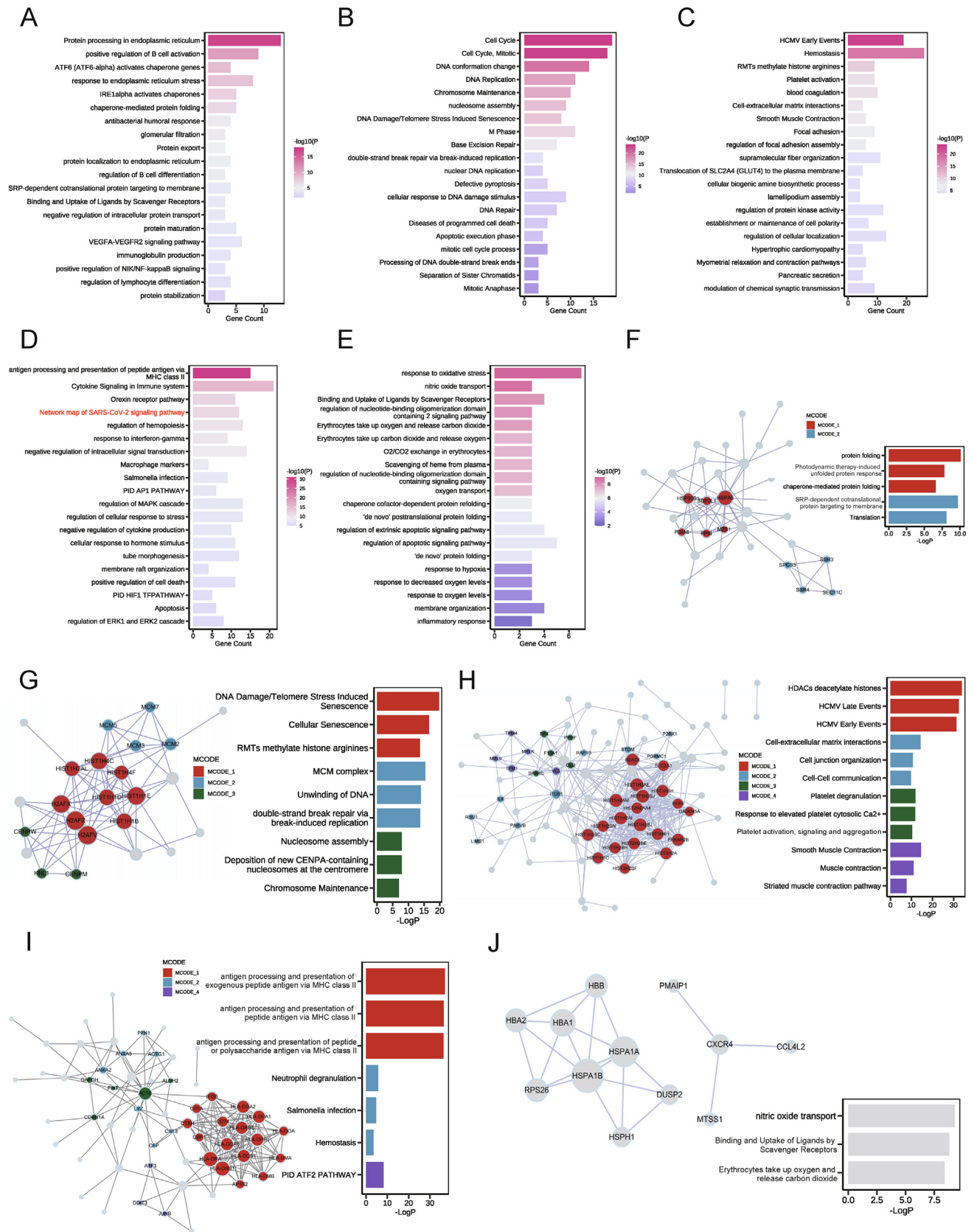


Fig. 5. Enrichment items and PPI network of each key module inferred by Metascape. (A-E) The enriched biological process terms by module red, magenta, blue, brown, and grey, respectively. The color represents the $-\log(p\text{-value})$ and the bar size represents the number of genes involved. (F-J) The inferred PPI network (left side). The order of modules was the same as those in (A-E). Each MCODE's top three enrichment items are shown on the right side. MCODE for identified "Molecular Complex Detection". (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The function of the brown module was closely related to immunity, which was consistent with the well-known function of corresponding Mono/Macro cells (Fig. 5D). In particular, we obtained a term named “Network Map of SARS-CoV-2 Signaling Pathway”, suggesting the important role of the brown module in regulating immune response in COVID-19. Densely correlated HLA genes indicated a higher antigen presentation function of MHC class II molecules (Fig. 5I). Other MCODEs were also related to immunity. SARS-CoV-2 induces cytokine storm, impels interferon response, and inhibits the function of MHC class I and II molecules [40]. The alteration of MHC II presentation might further contribute to immune escape [41]. This was consistent with the downregulation of the brown module in SP. In addition, the module participated in the process of apoptosis, which might explain the proliferation and decrease of monocytes and macrophages.

The grey module was closely related to oxygen transport and oxidative stress (Fig. 5E). Acute respiratory distress syndrome (ARDS) is one of the results of severe COVID-19 [42], which may be related to the downregulation of oxygen transport-related genes in the severe stage. In addition, oxidative stress plays an important role in the inflammatory response of innate immune cells [43], which explains the critical effect of this module on CD4 + T, mDC, NK, and $\gamma\delta$ T cells. Hub genes of the grey module, such as HBA1, CXCR4, and HSPA1B, were located at the center position of the PPI network and participated in oxidative stress regulation (Fig. 5J).

In general, we provided potential molecular mechanisms under dynamic cell heterogeneity of COVID-19 using scDisProcema.

4. Discussion

The COVID-19 pandemic poses serious health threats and economic costs to people worldwide. This disease leads to severe immune response and inflammatory factor storms, affecting the development and outcome of the disease. scRNA-seq has been utilized to study COVID-19 immunopathology because of its ability to delineate more detailed cell expression profiles than the traditional bulk RNA-seq. In this study, scRNA-seq data of COVID-19 patients under different states were collected to explore dynamic changes of immune regulation, providing a new perspective for the treatment and recovery of COVID-19.

The difference in gene regulation among different biological conditions, such as samples with different treatment doses or under different disease stages, is worth studying. As a traditional and uncomplicated tool for constructing a gene co-expression network, WGCNA has been widely used in studying immune regulation in COVID-19 patients. However, among these works [12,44–46], only the most common features of WGCNA were used. By using bulk RNA-seq data, the DEGs-samples matrix was extracted and used to identify gene modules for function and pathway enrichment [45,46]. The calculated ME values were used to compare differences between groups. Then key modules were selected by measuring the correlation between phenotypic traits and modules, and hub genes could be considered as important targets. Cells will be analyzed instead of samples when using single-cell data [44] to study COVID-19, and pipelines that apply WGCNA to scRNA-seq or other single-cell omics data have also been proposed [47,48]. In these pipelines, pseudocells or metacells that combine information of multiple cells were used instead of samples as input to WGCNA, which could moderately eliminate the problem of excessive dimensionality. However, there is no clear clinical meaning for pseudocells, so it is not reliable to explain underlying biological processes based upon WGCNA module analysis. Particularly, it is important to identify genes associated with dynamic changes during disease development or embryogenesis

[49], which is a difficult task for these methods. To address these challenges, we proposed scDisProcema, which, on the one hand, utilized the information obtained from upstream analysis of single-cell data. On the other hand, key modules that changed dynamically for COVID-19 across different states were identified.

Here, we conduct a follow-up exploration based on the analysis results of the Seurat package. Using the average expression of all major cell types in different states as input for scDisProcema, we obtained 11 gene modules. Considering the dynamics of gene expression and the correlation with changes in cell proportion, we identified the key modules that best reflected the specific biological changes in each cell type during disease development, that is, the red module for B cells and pDCs, magenta for CD8 + T cells, blue for megakaryocytes, brown for monocytes or macrophages, grey for CD4 + T, mDCs, NK and $\gamma\delta$ T cells. In the screening of hub genes, different from the previous methods that only considered connectivity degree, we also considered the specific expression of genes in cell types. Therefore, we screened hub genes in corresponding key modules for each cell type, which might play an important role in the dynamic process of disease development.

Then we performed functional enrichment analysis and PPI network analysis for each module. The red module is associated with ER stress, which induces inflammation response, apoptosis, and angiogenesis [50]. The magenta module contains many histone genes, which are closely linked to each other, and associated with DNA damage repair, confirming that DNA damage levels are elevated in COVID-19 patients [51]. The blue module is closely related to the coagulation function. The significantly enhanced expression level of this module in severe patients is consistent with the occurrence of thrombosis or DIC in severe patients [52]. Brown module was responsible for the immune responses of various cytokines, which might play an important role in the progression and prognosis of patients [53,54]. The grey module plays a vital role in oxygen transport and oxidative stress, the latter is an important part of the pathogenesis of COVID-19 [55] and interacts with the patients' inflammatory response [56]. In addition, through PPI network analysis, we found that most of the identified hub genes were involved in MCODEs related to the cell invasion process induced by viruses, indicating the significant role of these genes in responding to virus infection. Moreover, key gene modules showed different expression levels in healthy and patients under different disease stages, and regulated corresponding immune cells to undergo different biological processes, depicting the dynamic changes of the COVID-19 immune scene from multiple perspectives.

Through scDisProcema, we used the information of single cells to capture the dynamic change of samples at different time points or disease stages. With this approach, more applications can be explored. For example, to capture the dynamic change during aging, development, and regeneration processes. We can also explore the modular difference among patients with different clinical outcomes. In general, scDisProcema provides a practical way for investigating dynamic complex biological systems. However, in the current state, it has not yet been possible for scDisProcema to characterize heterogeneity information in disease processes at a finer granularity, such as the behavior of individual cells and inter-cellular variation—which was considered of great importance by some researchers [57]. We believe this could be a good direction for our future study.

CRediT authorship contribution statement

Anyao Li: Software, Data curation, Formal analysis, Writing – original draft. **Jihong Yang:** Conceptualization, Methodology, Writing – review & editing. **Jingyang Qian:** Visualization. **Xin Shao:** Writing – review & editing. **Jie Liao:** Writing – review & editing.

Xiaoyan Lu: Writing – review & editing. **Xiaohui Fan:** Conceptualization, Supervision, Writing – review & editing.

Declaration of Competing Interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China (81973701) and Innovation Team and Talents Cultivation Program of National Administration of Traditional Chinese Medicine (ZYXCXTD-D-202002).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.csbj.2022.06.066>.

References

- Liao M, Liu Y, Yuan J, Wen Y, Xu G, Zhao J, et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med* 2020;26(6):842–4.
- Huang W, Li M, Luo G, Wu X, Su B, Zhao L, et al. The inflammatory factors associated with disease severity to predict COVID-19 progression. *J Immunol* 2021;206(7):1597–608.
- Wen W, Su W, Tang H, Le W, Zhang X, Zheng Y, et al. Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. *Cell Discov* 2020;6:31.
- Chen G, Ning B, Shi T. Single-cell RNA-Seq technologies and related computational data analysis. *Front Genet* 2019;10:317.
- Shao X, Lu X, Liao J, Chen H, Fan X. New avenues for systematically inferring cell-cell communication: through single-cell transcriptomics data. *Protein Cell* 2020;11(12):866–80.
- Tang X, Uhl S, Zhang T, Xue D, Li B, Vandana JJ, et al. SARS-CoV-2 infection induces beta cell transdifferentiation. *Cell Metab* 2021;33(8):1577–1591 e1577.
- Xu G, Qi F, Li H, Yang Q, Wang H, Wang X, et al. The differential immune responses to COVID-19 in peripheral and lung revealed by single-cell RNA sequencing. *Cell Discov* 2020;6:73.
- Zheng Y, Liu X, Le W, Xie L, Li H, Wen W, et al. A human circulating immune cell landscape in aging and COVID-19. *Protein Cell* 2020;11(10):740–70.
- Ren X, Wen W, Fan X, Hou W, Su B, Cai P, et al. COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell* 2021;184(23):5838.
- Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinf* 2008;9:559.
- See P, Lum J, Chen J, Ginhoux F. A Single-Cell Sequencing Guide for Immunologists. *Front Immunol* 2018;9:2425.
- Verstockt B, Verstockt S, Abdu Rahiman S, Ke BJ, Arnauts K, Cleynen I, et al. Intestinal receptor of SARS-CoV-2 in inflamed IBD tissue seems downregulated by HNF4A in ileum and upregulated by interferon regulating factors in colon. *J Crohns Colitis* 2021;15(3):485–98.
- Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;36(5):411–20.
- Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* 2019;16(12):1289–96.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13(11):2498–504.
- Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;10(1):1523.
- Franzen O, Gan LM, Björkregren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* 2019, 2019.
- Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, et al. Cell Marker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res* 2019;47(D1):D721–8.
- Shao X, Liao J, Lu X, Xue R, Ai N, Fan X. scCATCH: automatic annotation on cell types of clusters from single-cell RNA sequencing data. *iScience* 2020;23(3):100882.
- Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;44(W1):W90–7.
- Dong J, Horvath S. Understanding network concepts in modules. *BMC Syst Biol* 2007;1:24.
- Tian Z, He W, Tang J, Liao X, Yang Q, Wu Y, et al. Identification of important modules and biomarkers in breast cancer based on WGCNA. *Oncotargets Ther* 2020;13:6805–17.
- Wan Q, Tang J, Han Y, Wang D. Co-expression modules construction by WGCNA and identify potential prognostic markers of uveal melanoma. *Exp Eye Res* 2018;166:13–20.
- Bader GD, Hogue CW. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinf* 2003;4:2.
- Rashid HO, Yadav RK, Kim HR, Chae HJ. ER stress: autophagy induction, inhibition and selection. *Autophagy* 2015;11(11):1956–77.
- Marzec M, Eletto D, Argon Y. GRP94: an HSP90-like protein specialized for protein folding and quality control in the endoplasmic reticulum. *Biochim Biophys Acta* 2012;1823(3):774–87.
- Berndt U, Oellerer S, Zhang Y, Johnson AE, Rospert S. A signal-anchor sequence stimulates signal recognition particle binding to ribosomes from inside the exit tunnel. *Proc Natl Acad Sci U S A* 2009;106(5):1398–403.
- Shelness GS, Lin L, Nicchitta CV. Membrane topology and biogenesis of eukaryotic signal peptidase. *J Biol Chem* 1993;268(7):5201–8.
- Laing AG, Lorenc A, Barrio I DMD, Das A, Fish M, Monin L, Munoz-Ruiz M, McKenzie DR, Hayday TS, Francos-Quijorna I, et al. A dynamic COVID-19 immune signature includes associations with poor prognosis. *Nat Med* 2020;26(10):1623–35.
- Bailis JM, Forsburg SL. MCM proteins: DNA damage, mutagenesis and repair. *Curr Opin Genet Dev* 2004;14(1):17–21.
- Hauer MH, Gasser SM. Chromatin and nucleosome dynamics in DNA damage and repair. *Genes Dev* 2017;31(22):2204–21.
- Dib PRB, Quirino-Teixeira AC, Merij LB, Pinheiro MBM, Rozini SV, Andrade FB, et al. Innate immune receptors in platelets and platelet-leukocyte interactions. *J Leukoc Biol* 2020;108(4):1157–82.
- Reeves MB. Cell signaling and cytomegalovirus reactivation: what do Src family kinases have to do with it? *Biochem Soc Trans* 2020;48(2):667–75.
- Atlante S, Mongelli A, Barbi V, Martelli F, Farsetti A, Gaetano C. The epigenetic implication in coronavirus infection and therapy. *Clin Epigenetics* 2020;12(1):156.
- Frydman GH, Tessier SN, Wong KHK, Vanderburg CR, Fox JG, Toner M, et al. Megakaryocytes contain extranuclear histones and may be a source of platelet-associated histones during sepsis. *Sci Rep* 2020;10(1):4621.
- Lam FW, Cruz MA, Leung HC, Parikh KS, Smith CW, Rumbaut RE. Histone induced platelet aggregation is inhibited by normal albumin. *Thromb Res* 2013;132(1):69–76.
- Becker RC. COVID-19-associated vasculitis and vasculopathy. *J Thromb Thrombolysis* 2020;50(3):499–511.
- Maquet J, Lafaurie M, Sommet A, Moulis G. Thrombocytopenia is independently associated with poor outcome in patients hospitalized for COVID-19. *Br J Haematol* 2020;190(5):e276–9.
- Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *J Thromb Haemost* 2020;18(4):844–7.
- Taefeshokr N, Taefeshokr S, Hemmat N, Heit B. Covid-19: perspectives on innate immune evasion. *Front Immunol* 2020;11:580641.
- Obermair FJ, Renoux F, Heer S, Lee CH, Cereghetti N, Loi M, et al. High-resolution profiling of MHC II peptide presentation capacity reveals SARS-CoV-2 CD4 T cell targets and mechanisms of immune escape. *Sci Adv* 2022;8(17):eab15394.
- Asselah T, Durantel D, Pasmant E, Lau G, Schinazi RF. COVID-19: Discovery, diagnostics and drug development. *J Hepatol* 2021;74(1):168–84.
- Chernyak BV, Popova EN, Prikhodko AS, Grebenchikov OA, Zinovkina LA, Zinovkin RA. COVID-19 and oxidative stress. *Biochemistry (Mosc)* 2020;85(12):1543–53.
- Ma D, Liu S, Hu L, He Q, Shi W, Yan D, et al. Single-cell RNA sequencing identify SDCBP in ACE2-positive bronchial epithelial cells negatively correlates with COVID-19 severity. *J Cell Mol Med* 2021;25(14):7001–12.
- Mo S, Dai L, Wang Y, Song B, Yang Z, Gu W. Comprehensive analysis of the systemic transcriptomic alternations and inflammatory response during the occurrence and progress of COVID-19. *Oxid Med Cell Longev* 2021;2021:9998697.
- Zhang J, Lin D, Li K, Ding X, Li L, Liu Y, et al. Transcriptome analysis of peripheral blood mononuclear cells reveals distinct immune response in asymptomatic and re-detectable positive COVID-19 patients. *Front Immunol* 2021;12:716075.
- Feregino C, Tschopp P. Assessing evolutionary and developmental transcriptome dynamics in homologous cell types. *Dev Dyn* 2021.
- Morabito S, Miyoshi E, Michael N, Shahin S, Martini AC, Head E, et al. Single-nucleus chromatin accessibility and transcriptomic characterization of Alzheimer's disease. *Nat Genet* 2021;53(8):1143–55.
- Shao L, Xue R, Lu X, Liao J, Shao X, Fan X. Identify differential genes and cell subclusters from time-series scRNA-seq data using scTITANS. *Comput Struct Biotechnol J* 2021;19:4132–41.
- Banerjee A, Czinn SJ, Reiter RJ, Blanchard TG. Crosstalk between endoplasmic reticulum stress and anti-viral activities: a novel therapeutic target for COVID-19. *Life Sci* 2020;255:117842.

- [51] Lorente L, Martin MM, Gonzalez-Rivero AF, Perez-Cejas A, Caceres JJ, Perez A, et al. DNA and RNA oxidative damage and mortality of patients with COVID-19. *Am J Med Sci* 2021;361(5):585–90.
- [52] Mei H, Luo L, Hu Y. Thrombocytopenia and thrombosis in hospitalized patients with COVID-19. *J Hematol Oncol* 2020;13(1):161.
- [53] Lowery SA, Sariol A, Perlman S. Innate immune and inflammatory responses to SARS-CoV-2: Implications for COVID-19. *Cell Host Microbe* 2021;29(7):1052–62.
- [54] Ramasamy S, Subbian S. Critical determinants of cytokine storm and type I interferon response in COVID-19 pathogenesis. *Clin Microbiol Rev* 2021;34(3).
- [55] Wieczfinska J, Kleniewska P, Pawliczak R. Oxidative stress-related mechanisms in SARS-CoV-2 infections. *Oxid Med Cell Longev* 2022;2022:5589089.
- [56] Vollbracht C, Kraft K. Oxidative stress and hyper-inflammation as major drivers of severe COVID-19 and long COVID: implications for the benefit of high-dose intravenous vitamin C. *Front Pharmacol* 2022;13:899198.
- [57] Ferrari Gianlupi J, Mapper T, Segó TJ, Sluka JP, Quinney SK, Craig M, et al. Multiscale model of antiviral timing, potency, and heterogeneity effects on an epithelial tissue patch infected by SARS-CoV-2. *Viruses* 2022;14(3).