

# Delimiting COVID-19 immunological features using single-cell RNA sequencing

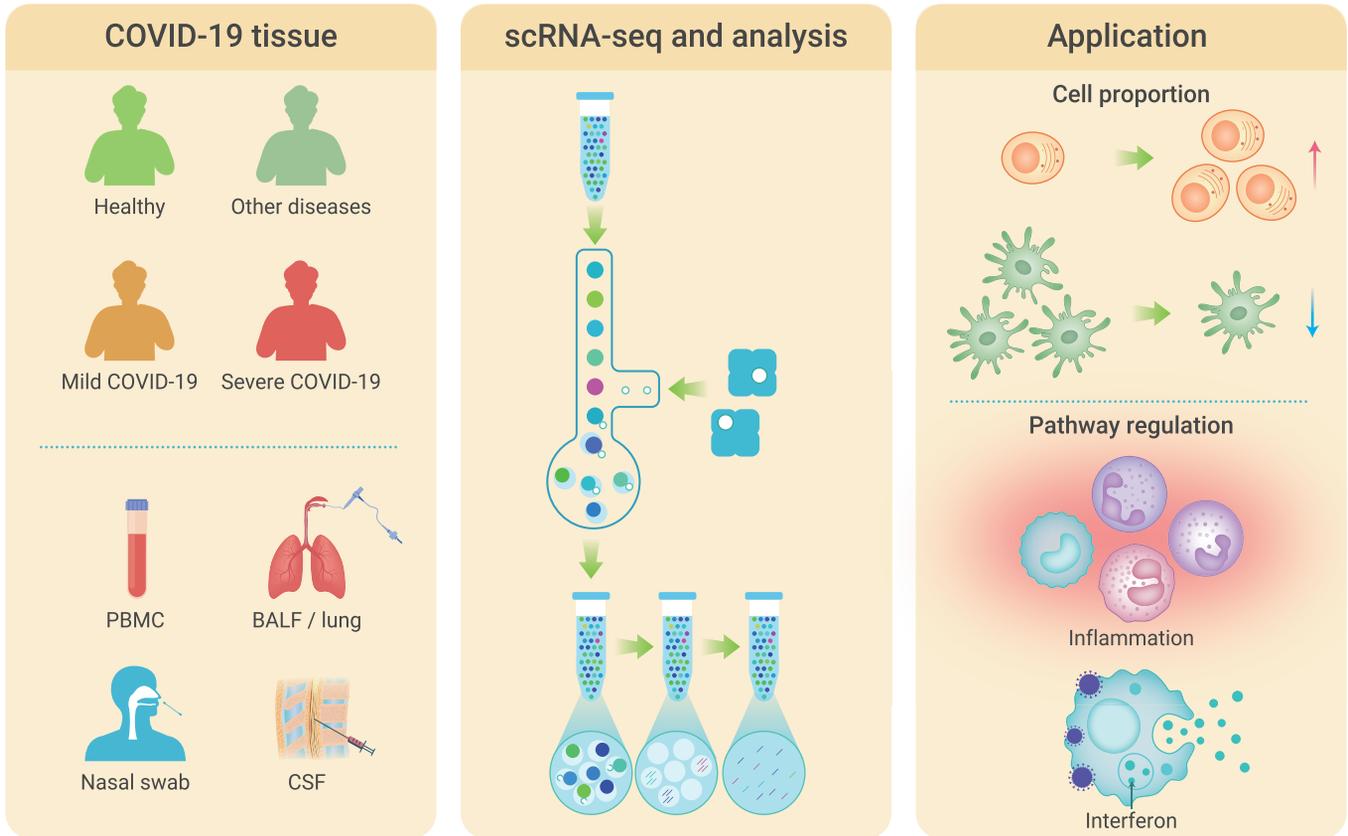
Wendao Liu,<sup>1,2</sup> Johnathan Jia,<sup>1,2</sup> Yulin Dai,<sup>2</sup> Wenhao Chen,<sup>3,4</sup> Guangsheng Pei,<sup>2</sup> Qiheng Yan,<sup>2</sup> and Zhongming Zhao<sup>1,2,5,6,\*</sup>

\*Correspondence: [zhongming.zhao@uth.tmc.edu](mailto:zhongming.zhao@uth.tmc.edu)

Received: April 11, 2022; Accepted: July 16, 2022; Published Online: July 21, 2022; <https://doi.org/10.1016/j.xinn.2022.100289>

© 2022 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## GRAPHICAL ABSTRACT



## PUBLIC SUMMARY

- Single-cell RNA sequencing is a powerful biotechnology in COVID-19 research
- It reveals multiple features of SARS-CoV-2 pathogenesis and host immune response
- Single-cell immune profiling characterizes the adaptive immune response in COVID-19
- Analysis of B cell receptors accelerates the neutralizing antibody identification



# Delineating COVID-19 immunological features using single-cell RNA sequencing

Wendao Liu,<sup>1,2</sup> Johnathan Jia,<sup>1,2</sup> Yulin Dai,<sup>2</sup> Wenhao Chen,<sup>3,4</sup> Guangsheng Pei,<sup>2</sup> Qiheng Yan,<sup>2</sup> and Zhongming Zhao<sup>1,2,5,6,\*</sup>

<sup>1</sup>The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX 77030, USA

<sup>2</sup>Center for Precision Health, School of Biomedical Informatics, the University of Texas Health Science Center at Houston, Houston, TX 77030, USA

<sup>3</sup>Immunobiology and Transplant Science Center, Department of Surgery, Houston Methodist Research Institute and Institute for Academic Medicine, Houston Methodist Hospital, Houston, TX 77030, USA

<sup>4</sup>Department of Surgery, Weill Cornell Medicine, Cornell University, New York, NY 10065, USA

<sup>5</sup>Human Genetics Center, School of Public Health, the University of Texas Health Science Center at Houston, Houston, TX 77030, USA

<sup>6</sup>Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN 37203, USA

\*Correspondence: [zhongming.zhao@uth.tmc.edu](mailto:zhongming.zhao@uth.tmc.edu)

Received: April 11, 2022; Accepted: July 16, 2022; Published Online: July 21, 2022; <https://doi.org/10.1016/j.xinn.2022.100289>

© 2022 The Author(s). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Citation: Liu W., Jia J., Dai Y., et al., (2022). Delineating COVID-19 immunological features using single-cell RNA sequencing. *The Innovation* **3**(5), 100289.

Understanding the molecular mechanisms of coronavirus disease 2019 (COVID-19) pathogenesis and immune response is vital for developing therapies. Single-cell RNA sequencing has been applied to delineate the cellular heterogeneity of the host response toward COVID-19 in multiple tissues and organs. Here, we review the applications and findings from over 80 original COVID-19 single-cell RNA sequencing studies as well as many secondary analysis studies. We describe that single-cell RNA sequencing reveals multiple features of COVID-19 patients with different severity, including cell populations with proportional alteration, COVID-19-induced genes and pathways, severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) infection in single cells, and adaptation of immune repertoire. We also collect published single-cell RNA sequencing datasets from original studies. Finally, we discuss the limitations in current studies and perspectives for future advance.

## INTRODUCTION

As of June 11<sup>th</sup>, 2022, the estimated coronavirus disease 2019 (COVID-19) global caseload and mortality are over 535 million cases and over 6.3 million deaths respectively (<https://ourworldindata.org/grapher/cumulative-deaths-and-cases-covid-19>). The severity of this global emergency has provided fuel for COVID-19 research. This is reflected in the sharp increase in COVID-19-related publications, with over 100,000 articles estimated in 2020 alone (<https://www.nature.com/articles/d41586-020-03564-y>), many of these published as preprints.

Single-cell RNA sequencing (scRNA-seq) has become one of the most powerful tools to understand the dynamics of gene expression and genomics both within the cell and in the cellular environment. First developed in 2009 to sequence a mouse blastomere, subsequent developments have made this high-resolution methodology readily available and widely applied for dissecting heterogeneity of human tissues and underlying diseases. The clinical symptoms of severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) infection severity varies dramatically, ranging from asymptomatic, mild, severe, critical, to even death. Furthermore, viral infection of host cells causes dramatic changes in the immune response.<sup>1–4</sup> scRNA-seq offers a high-resolution view into the cell and cellular environment. Thus, it has been an important tool for studying the molecular mechanisms of COVID-19, including the dynamic cellular changes in response to viral infection.

Accordingly, we have seen over 200 scRNA-seq publications for COVID-19 since the spring 2020. Here, we systematically summarize our search of these studies, the related datasets, as well as their discoveries. We first present a summary of the literature and their classifications. In addition, we describe the published datasets and their features by categories. We then discuss the standards of and definitions of COVID-19 infection severity in these studies. Next, the findings from these studies regarding immune cell subpopulations and differential gene expression are compared and discussed. In addition, the pathways of interest and their functions in COVID-19 infection severity are reviewed. Furthermore, we highlight some important findings in cellular communication, cell trajectory inference, and other applications such as using novel variable, diversity and joining (VDJ) sequencing to identify COVID-19-specific T and B cell responses. Finally, we review the applications of these datasets, their limitations, and

potential improvements to be made in future. We further discuss the gaps in knowledge in the field and its current direction.

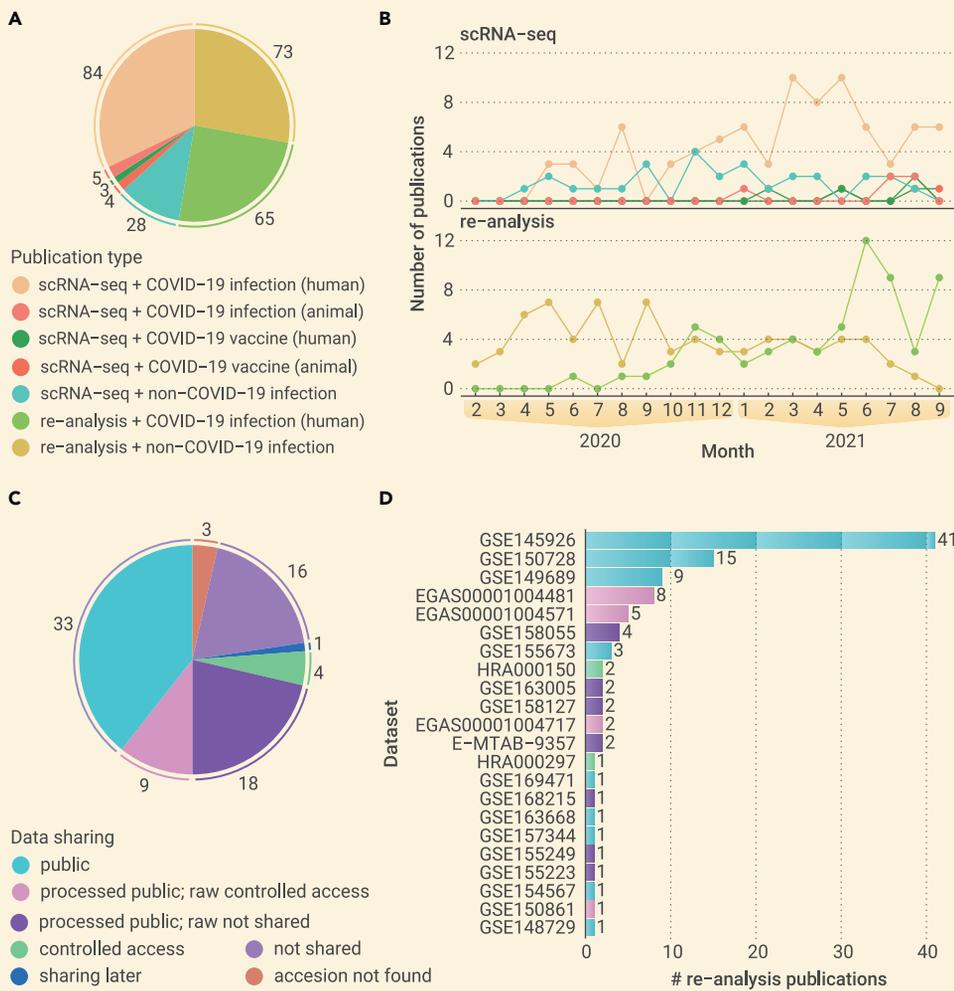
## RAPID ADOPTION OF SCRNA-SEQ IN COVID-19 RESEARCH

Since early spring 2020, over 80 original studies using scRNA-seq in COVID-19 research and over 60 bioinformatic re-analysis studies of those original datasets have been published. We retrieved a collection of articles about this topic from PubMed with query “(single-cell) AND (sequencing OR seq) AND (COVID OR SARS-CoV-2)” on October 4<sup>th</sup>, 2021. After removing duplicates, replacing published preprint articles, and filtering out irrelevant articles, we obtained a total of 262 articles (Table S1). We classified articles according to experimental methods, species, and whether subjects were infected with COVID-19. Among these articles, nearly half of the studies performed scRNA-seq on infected or uninfected subjects, while another half performed bioinformatic analysis on published datasets but did not perform original sequencing (Figure 1A).

Next, we analyzed the trend of COVID-19 scRNA-seq studies by looking into the online publication date for these articles. For the studies performing scRNA-seq on infected subjects, there is a gradual increase in the number of article since the first publication by Wen et al. in May 2020 (Figure 1B).<sup>5</sup> There are also 28 studies performing scRNA-seq on uninfected subjects and analyzing potential implications of SARS-CoV-2 infection (labeled “non-COVID-19 infection” in Figure 1A). They quantified the expression of *ACE2* and *TMPRSS2* in certain organs, tissues, or cell types to predict their susceptibility to SARS-CoV-2 infection.<sup>6</sup> scRNA-seq studies using vaccinated human or animal subjects to study vaccine efficacy and vaccine-induced immune responses started to appear in 2021. Since June 2021, there has been a trend of more articles focusing on analyzing COVID-19 scRNA-seq datasets than non-COVID-19 datasets. This is likely due to the accumulating number of COVID-19 scRNA-seq studies that shared data in the past year. With the ongoing global pandemic, we expect that more scRNA-seq and bioinformatic analysis studies will appear soon.

Data sharing is a critical issue in biomedical research. As the datasets generated in COVID-19 scRNA-seq studies can be re-analyzed by others, we evaluate data sharing of these studies (Figure 1C). About two-thirds of the studies share both raw (FASTQ) and processed data, or at least share the processed data. However, there are also studies that did not share their generated scRNA-seq datasets on any public databases or websites. In addition, several studies claim they uploaded their datasets into specific data repository, but there are issues preventing access, including invalid accession numbers and long-term delayed sharing. This issue is unexpected, as most journals now require data sharing for peer review. To evaluate the effect of data sharing, we counted the frequency of published COVID-19 scRNA-seq datasets that were re-analyzed in bioinformatic analysis articles. The top three datasets<sup>7–9</sup> are all publicly accessible datasets deposited to the Gene Expression Omnibus (GEO) database, and most of the re-analyzed datasets have publicly accessible processed data (Figure 1D). It is apparent that data sharing promoted follow-up studies in COVID-19 research. We conclude that data sharing is necessary and strongly encouraged for ongoing and future COVID-19 studies.

To facilitate further COVID-19 research, we curated a collection of COVID-19 scRNA-seq datasets (Table S2). We made notes of the sampling tissues/organs, number of recruited subjects, sequencing protocols, and accessibility of



**Figure 1. The summary of COVID-19 scRNA-seq publications as of October 1<sup>st</sup>, 2021** (A) Number of published articles by study categories. "scRNA-seq" refers to the original studies by generating scRNA-seq data. "re-analysis" indicates the re-analysis studies of the published scRNA-seq datasets. "COVID-19 infection" refers to those studies including COVID-19-infected human or animal subjects. "Non-COVID-19 infection" refers to those studies using non-COVID-19-infected human or animal subjects to predict the susceptibility of cells to SARS-CoV-2. (B) Distribution of the number of published articles by month in 2020 and 2021. (C) Status of scRNA-seq data sharing of the original studies with COVID-19 patients. Raw data refer to the FASTQ files. Processed data refer to the count matrices or R/Python objects. (D) The number of publications (re-analysis) per each COVID-19 scRNA-seq dataset.

provided by the National Health Commission of China (NHCC).<sup>16</sup> The NHCC standard defines four groups: mild, moderate, severe, and critical (Table 1). Our review indicates that the NHCC's standard is used the most among those COVID-19 scRNA-seq studies, followed by the WHO's standard. Interestingly, some researchers used multiple datasets where both standards of severity were used. In addition to the two most popular standards from the WHO and NHCC, other criteria, such as intensive care unit (ICU) admission and mechanical ventilation, a modified Murray score, or the National Early Warning Score (NEWS), were adopted. The modified Murray score is a four-point scale based on five clinical criteria. The NEWS is used in the UK to determine the degree of illness and need for critical care (<https://www.mdcalc.com/national-early-warning-score-news>).

Our comparison shows no consensus agreement on which standard to use in a study. For

example, Ren et al. categorized the patients based on the WHO clinical guidelines to develop a large single-cell transcriptomic atlas.<sup>10</sup> They found that megakaryocytes and monocytes might contribute to the cytokine storms observed in severe infections. On the other hand, Ziegler et al. used the WHO's nine-point scale to classify scRNA-seq data from nasopharyngeal swabs. They reported that, despite a similar viral load, the epithelial cells expressed antiviral genes in mild infections, while the antiviral responses in nasal epithelia were impaired in severe infections.<sup>17</sup>

### COVID-19 INFECTION SEVERITY

these datasets (Figure 2). Most datasets are generated using peripheral blood mononuclear cells (PBMCs) from subjects. Other common tissue or organ samples include bronchoalveolar lavage fluid (BALF), nasopharyngeal swab, cerebrospinal fluid (CSF), and lung. Most studies recruit up to 20 COVID-19 patients and 20 healthy subjects as controls, but there are a few studies with over 50 COVID-19 patients.<sup>10–13</sup> For library preparation and sequencing protocols, most studies use the 10x Genomics platform. The 5' gene expression library, including VDJ library (total 38), is preferred compared with the 3' library (total 28). Most datasets share processed data like count matrices or R/Python objects, and 36 out of the 65 datasets share raw sequencing data.

One of the major goals of COVID-19 scRNA-seq research has been uncovering the genetic and cellular factors that drive COVID-19 disease severity. However, how do we measure COVID-19 infection severity? The spectrum of symptoms seen in COVID-19 infections is broad. While some patients do not display any signs of infection (asymptomatic), there are others that suffer critical or fatal infections. Infection severity can be classified into different groups, such as mild, moderate, and critical or severe infection based on symptoms. In this section, we present some of the standards used for classifying infection severity in COVID-19 scRNA-seq research.

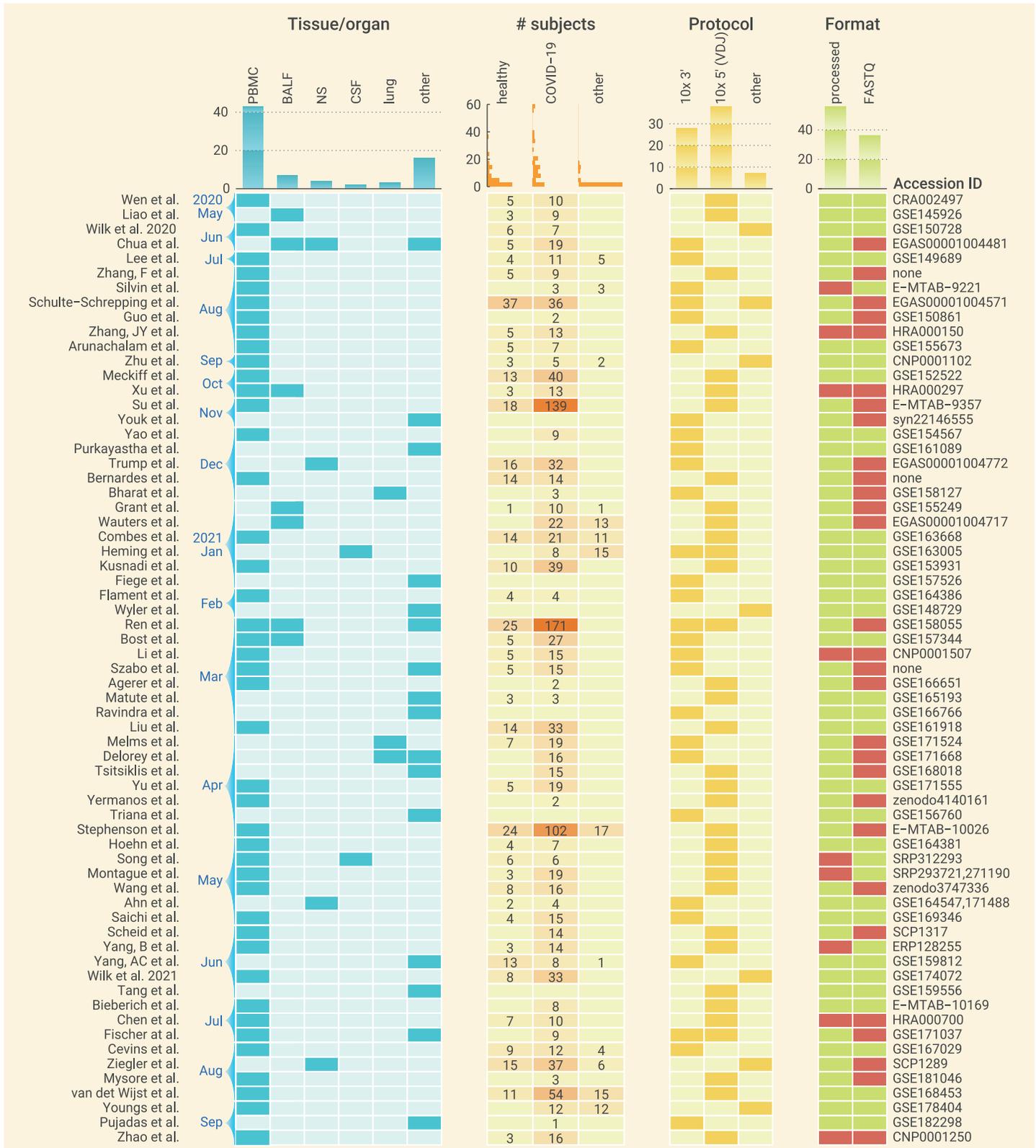
The World Health Organization (WHO) developed a COVID-19 infection severity classification scale for clinical management<sup>14</sup> and a nine-point WHO ordinal scale (WOS) for diagnosing COVID-19 infection severity in clinical trials.<sup>15</sup> Under the clinical guidelines, there are three distinct groups: non-severe (also called moderate), severe, and critical COVID-19. The nine-point scale categorizes the patient state as uninfected, ambulatory, mild, severe, and dead based on a score from 0 to 8, with 0 being uninfected and 8 being dead. The score is assigned according to the description of the patient (Table 1). Another infection severity standard is

provided by the National Health Commission of China (NHCC).<sup>16</sup> The NHCC standard defines four groups: mild, moderate, severe, and critical (Table 1). Our review indicates that the NHCC's standard is used the most among those COVID-19 scRNA-seq studies, followed by the WHO's standard. Interestingly, some researchers used multiple datasets where both standards of severity were used. In addition to the two most popular standards from the WHO and NHCC, other criteria, such as intensive care unit (ICU) admission and mechanical ventilation, a modified Murray score, or the National Early Warning Score (NEWS), were adopted. The modified Murray score is a four-point scale based on five clinical criteria. The NEWS is used in the UK to determine the degree of illness and need for critical care (<https://www.mdcalc.com/national-early-warning-score-news>).

Our comparison shows no consensus agreement on which standard to use in a study. For example, Ren et al. categorized the patients based on the WHO clinical guidelines to develop a large single-cell transcriptomic atlas.<sup>10</sup> They found that megakaryocytes and monocytes might contribute to the cytokine storms observed in severe infections. On the other hand, Ziegler et al. used the WHO's nine-point scale to classify scRNA-seq data from nasopharyngeal swabs. They reported that, despite a similar viral load, the epithelial cells expressed antiviral genes in mild infections, while the antiviral responses in nasal epithelia were impaired in severe infections.<sup>17</sup>

In May 2020, Liao et al. published the first COVID-19 lung tissue scRNA-seq study in Nature Medicine.<sup>7</sup> They developed a single-cell transcriptomic atlas directly from BALF tissue.<sup>7</sup> They used the NHCC infection severity classification. Liu et al. linked immune response variation to disease severity over time by performing single-cell analysis of PBMCs.<sup>18</sup> They found that severe patients exhibited low levels of cellular inflammation early in their hospitalization, but, at 17 to 23 days after symptom onset, the inflammatory responses were significantly elevated. Hasan et al. classified infections by defining ICU admission and mechanical ventilation treatment as severe infections.<sup>19</sup> Lam et al. stratified patients by measuring lung injury with a modified Murray score that assigns a score between 0 and 4 based on five criteria.<sup>20</sup> Lee et al. used NEWS standard to stratify the disease status and found that type I interferons (IFNs) exacerbate inflammation in severe infections.<sup>9</sup>

Since the classification systems from the WHO and NHCC have similar guidelines, we expect the results to be comparable between similar infection severities. However, the extent of the difference in the results remains unclear and warrants some future investigation. We believe a universal standard would be beneficial as standardization is important for comparison between cohorts. However, we also acknowledge that various factors such as country of origin for data collection may determine which standard is used.



**Figure 2. A detailed view of 65 COVID-19 scRNA-seq datasets** On the top, it shows several categories: tissue/organ, number of subjects, sequencing protocol, accessibility of data format (raw or processed). On the left, it shows the datasets by the first author and publication month/year. Top bar plots show the numbers of datasets belonging to each category. For number of subjects, a blank cell indicates zero subject in this category. The top histogram shows the distribution of subjects in each category among datasets. Datasets with zero subjects indicate that they are generated from cell culture. For data format columns, green indicates the dataset being publicly accessible, while red indicates not (including dataset not shared, controlled access, and invalid accession ID). See [Table S2](#) for details. BALF, bronchoalveolar lavage fluid; CSF, cerebrospinal fluid; NS, nasopharyngeal swab; PBMC, peripheral blood mononuclear cell.

### ALTERED CELL PROPORTIONS IN COVID-19

scRNA-seq enables the identification of cell populations in samples and comparison of cell proportions in different conditions. When comparing cell proportions from samples in different conditions, a higher proportion usually implies

stronger local cell proliferation, transition from other cell types, or recruitment of cells from adjacent tissues like blood. On the contrary, a lower proportion usually implies greater cell death, transition to other cell types, or migration outward. Many COVID-19 scRNA-seq studies analyzed cell proportion alteration

**Table 1.** Summary of infection severity definition and number of COVID-19 studies in each definition

Definition <sup>a</sup>	WHO	WOS	NHCC	Other
Mild	Absence of any signs of severe or critical COVID-19	Score 3: hospitalized, no oxygen therapy Score 4: oxygen by mask or nasal prongs	Mild symptoms with no sign of pneumonia on imaging	Other criteria
Moderate			Fever and respiratory symptoms with radiological findings of pneumonia	
Severe	Any of: (1) oxygen saturation <90% on room air (2) signs of severe respiratory distress in adults; i.e., respiratory rate >30 breaths per minute (3) presence of danger signs in children such as cyanosis	Score 5: non-invasive ventilation or high-flow oxygen Score 6: intubation and mechanical ventilation Score 7: ventilation + additional organ support: pressors, RRT, ECMO	Any of: (1) respiratory distress (defined by as >30 breaths per minute) (2) oxygen saturation <93% at rest (3) PaO <sub>2</sub> /FiO <sub>2</sub> < 300 mmHg (4) over 50% progression of lung lesions within 24–48 h (diagnosed by lung imaging)	
Critical	Any of: (1) ARDS (2) sepsis (3) septic shock (4) requiring any ventilation support such as mechanical ventilation (any kind) and/or vasopressor therapy		Any of: (1) respiratory failure and invasive mechanical ventilation (2) shock (3) multi-organ dysfunction requiring ICU admission and monitoring	
No. of studies	3	12	17 <sup>b</sup>	14

ARDS, acute respiratory distress syndrome; ECMO, extracorporeal membrane oxygenation; NHCC, National Health Commission of China; ICU, intensive care unit; PaO<sub>2</sub>/FiO<sub>2</sub>, ratio of arterial oxygen partial pressure to fraction of inspired air; RRT, renal replacement therapy; WHO World Health Organization; WOS, WHO ordinal scales.

<sup>a</sup>Including the studies using the datasets generated by other studies.

<sup>b</sup>One study that used both WHO and NHCC standards is not included here.

in COVID-19 patients of different severities compared with healthy controls or controls with other diseases (Figure 3; Table S3).

Most studies used PBMC samples.<sup>8–11,21–33</sup> In general, these studies reported that neutrophils, plasma cells, and classical monocytes had significantly increased proportions in the blood of COVID-19 patients. Higher proportions of these cells were also found in severe patients than in moderate patients.<sup>7–9,22,28,29,31,32</sup> Cell populations with a decreased proportion in the blood of COVID-19 patients included the overall T cells, natural killer (NK) cells, dendritic cells (DCs), and non-classical monocytes. Severe patients had even lower proportions of these cells than moderate patients.<sup>8–11,21–25,27,29–32</sup> Analysis of bone marrow mononuclear cells (BMMCs) showed a significant increase of immature myeloid progenitors and a dramatic decrease of lymphoid progenitors in severe patients, suggesting that dysregulated hematopoiesis contributes to the proportional changes of immune cell subsets in COVID-19 patients.<sup>30,34</sup> Of note, each immune cell subset underwent dynamic cell state transitions in response to viral infection. For example, Xu et al. found that the proportions of the activated CD4<sup>+</sup> T cell subsets (e.g., T helper [Th] 1, Th2, and Th17-like) among the T cell compartment were increased in the severe patients even though overall T cells were decreased in the PBMCs from severe COVID-19 patients. How this reflects disease severity needs further investigation, and each immune cell subset in COVID-19 patients needs more detailed analysis to reflect a spectrum of dynamic cell states.

In BALF, epithelial cells were decreased in both moderate and severe COVID-19 patients.<sup>4</sup> As lung epithelial cells are primarily the targets of SARS-CoV-2 attack, the loss of epithelial cells may reflect virally induced cell death.<sup>2</sup> Several research groups revealed that, compared with the healthy controls, the numbers of CD8<sup>+</sup> T cells and NK cells were increased in the moderate and severe patients.<sup>1,4,7</sup> The severe patients had higher proportions of neutrophils and monocytes, and lower proportions of alveolar macrophages, than both the controls and moderate patients, but there were no significant differences between the controls and moderate patients regarding the proportions of those cell subsets. In contrast, the proportion of plasmacytoid DCs (pDCs) in BALF was comparable between healthy controls and severe patients, and it was only increased in moderate patients.<sup>1,7</sup>

In a scRNA-seq study of nasopharyngeal swab samples, Chua et al. found that both the moderate and severe COVID-19 patients had higher proportions of proliferative secretory cells, monocyte-derived DCs, NK T cells, NK cells, and neutrophils compared with the healthy controls.<sup>35</sup> Particularly, an increase in the proportion of neutrophils was observed in association with disease severity (none in the healthy controls versus 61% of immune cells in the severe patients). In addition, COVID-19 patients had lower proportions of basal cells, secretory cells, squamous cells, resident macrophages, monocyte-derived macrophages, and cyto-

toxic T cells. Remarkably, the proportion of non-resident macrophages was decreased in the moderate patients but increased in the severe patients who experienced a severe extravasation event.<sup>35</sup>

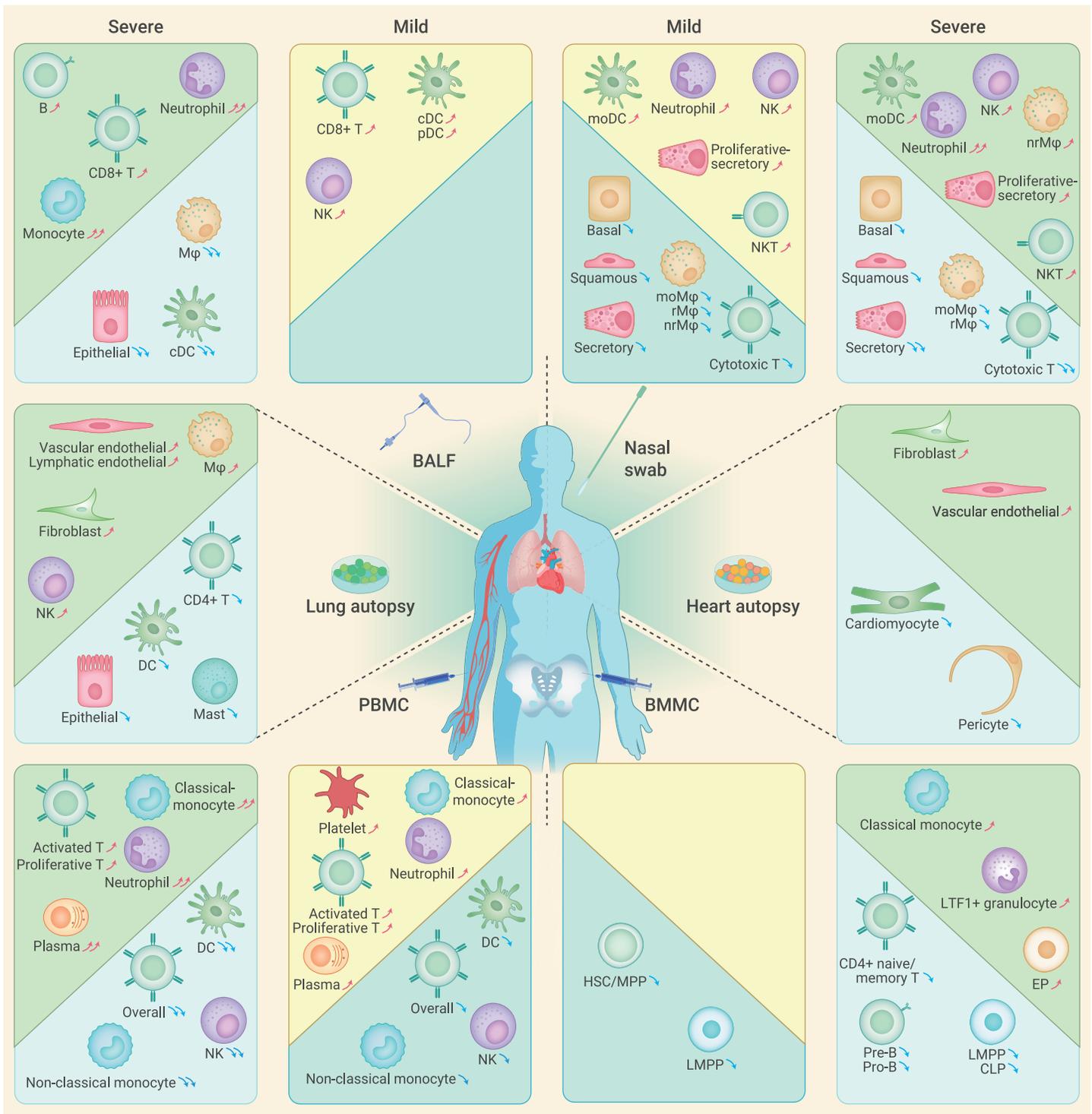
Cell proportion alteration in the lung and heart from autopsy was also examined for the patients who died of COVID-19. In the lung of COVID-19 patients, macrophages, NK cells, fibroblasts, and endothelial cells were increased, while epithelial cells were decreased.<sup>2,36</sup> This is consistent with the results from BALF studies. When studying the heart tissue of COVID-19 patients, fibroblasts and vascular endothelial cells were significantly increased, while cardiomyocytes and pericytes were decreased, suggesting stress-induced apoptosis of these cells.<sup>36</sup>

In summary, scRNA-seq has demonstrated great sensitivity and accuracy in quantifying cell proportions and facilitates the identification of cells with altered proportions. The significantly altered cell populations from most studies are generally reliable. However, it is worth noting that different studies usually have highly variable clustering results, and they use different methods to determine significant changes in cell proportions. This potentially leads to some inconsistent results among different studies. It remains unclear whether the results would show more consistency if a universal pipeline was applied to all samples. Such an analysis using the same pipeline is warranted in future.

### COVID-19-INDUCED GENES AND PATHWAYS

One unique feature of scRNA-seq is its capacity to analyze the transcriptomic changes of different cell types in response to SARS-CoV-2 infection. Differential expression analysis on scRNA-seq data is a common approach to find upregulated and downregulated genes between COVID-19 patients and healthy controls, as well as between the patients of different severity. Those differentially expressed genes (DEGs) are usually followed by gene set enrichment analysis to identify the activated and deactivated pathways that are most associated with the immune responses against SARS-CoV-2 infection (Figure 4).

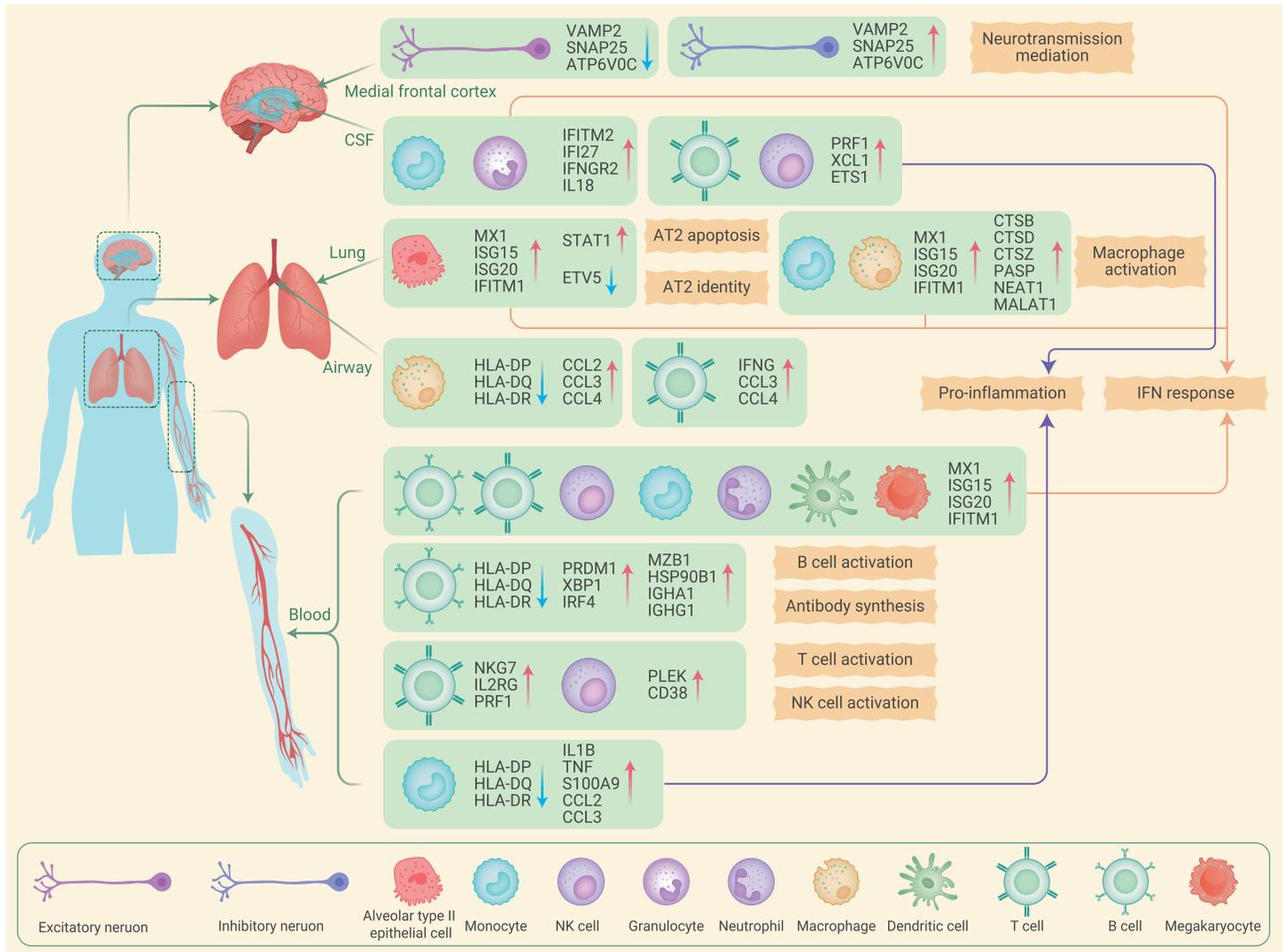
In PBMCs from COVID-19 patients, IFN-stimulated genes (ISGs) were the most commonly found upregulated genes in most cell types, including B cells/plasma cells,<sup>11,21,31</sup> T cells,<sup>8,11,31</sup> NK cells,<sup>25,31,37</sup> monocytes,<sup>9,13,31,37,38</sup> DCs,<sup>31,38</sup> neutrophils,<sup>22</sup> and megakaryocytes.<sup>21</sup> This gene expression change is expected because the type I IFN system is triggered by SARS-CoV-2, and ISGs have antiviral functions. ISGs usually had highest expression in moderate patients and lower expression in severe patients,<sup>11,13,21,22,38</sup> indicating the decreased immunity of severe patients. Similarly, human leukocyte antigen (HLA)-II genes also showed downregulation in B cells and myeloid cells from COVID-19 patients compared with healthy controls,<sup>1,8,12,38</sup> suggesting dysregulation of immune cell crosstalk. Moreover, genes in some inflammation-mediating pathways, like interleukin (IL)-1 and tumor necrosis factor (TNF) pathways,



**Figure 3. Major cell populations with significantly altered proportions in COVID-19** Red and blue arrows indicate cell populations with significantly altered proportions in six different tissues from moderate and severe COVID-19 patients. Double arrows indicate that the proportion of the cell populations in severe patients is significantly higher or lower than both healthy controls and moderate patients. BALF, bronchoalveolar lavage fluid; BMMC, bone marrow mononuclear cell; cDC, conventional dendritic cell; CLP, common lymphoid progenitor cell; EP, erythrocyte progenitor cell; HSC/MPP, hematopoietic stem cell/multipotent progenitor cell; LMPP, lymphoid-primed multipotent progenitor; moDC, monocyte-derived dendritic cell; moMφ, monocyte-derived macrophage; nrMφ, non-resident macrophage; PBMC, peripheral blood mononuclear cell; pDC, plasmacytoid dendritic cell; rMφ, resident macrophage.

as well as genes encoding pro-inflammatory molecules like S100A8, S100A9, and S100A12, were found upregulated in several cell types,<sup>9–11,25,31,38</sup> suggesting hyperinflammation in COVID-19. The burst of alarmins S100A8/A9 is considered to contribute to COVID-19 pathogenesis by driving the formation of aberrant classical monocytes and immature neutrophils, and to trigger the cytokine storm that characterizes severe COVID-19. These aberrant monocytes and neutrophils have an immunosuppressive phenotype, suggesting the defective monocyte activation and emergency myelopoiesis.<sup>27,28</sup> In addition to the

genes involved in common pathways, each cell type may have its specific DEGs. For example, B cells were characterized with upregulated activation-associated genes and protein transport-associated genes,<sup>5,31</sup> showing their function in antibody synthesis. Effector T cells and NK cells upregulated genes involved in activation, cytokine production, and lymphocyte cytotoxicity.<sup>8,11,25,31</sup> As major inducers and regulators of inflammation, monocytes in blood and macrophages in BALF upregulated genes encoding many inflammatory cytokines like CCL2, CCL3, and CCL4.<sup>1,7,31,38</sup>



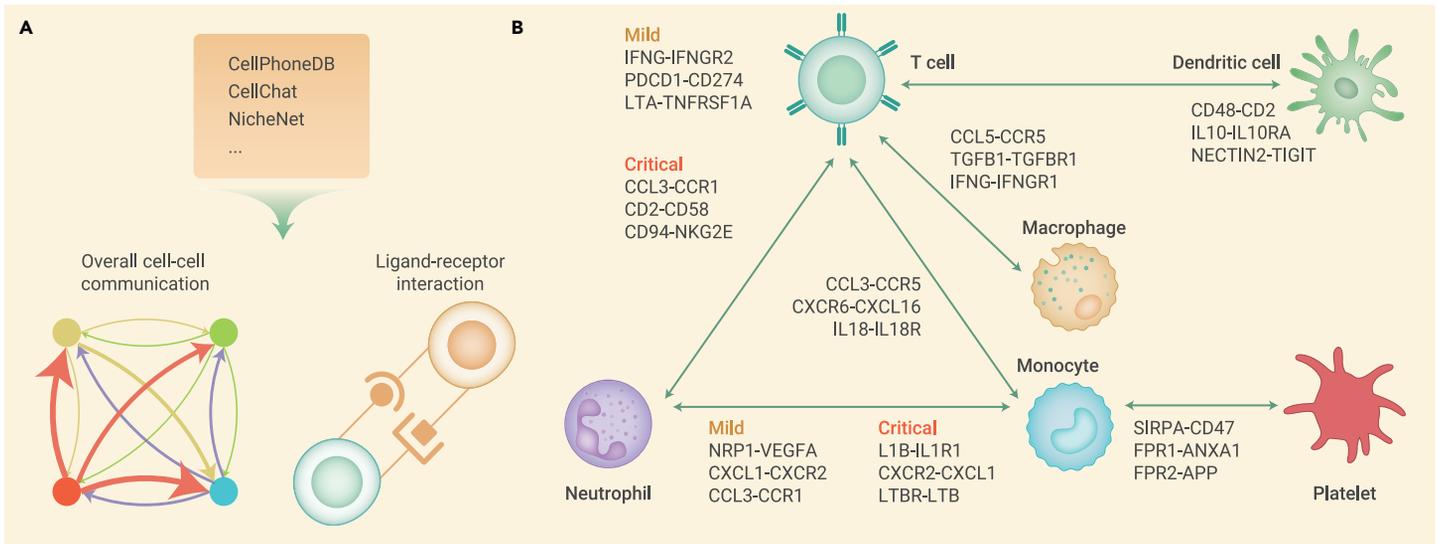
**Figure 4. scRNA-seq reveals major induced genes and pathways in different cell types from brain, lung, and blood of COVID-19 patients** Genes involved in interferon (IFN) response and inflammation were most commonly induced in multiple cell types from different tissues in COVID-19 patients. Each cell type also has induced genes and pathways associated with its biological functions. AT2, alveolar type II epithelial cell; CSF, cerebrospinal fluid.

Samples from other organs of COVID-19 patients displayed similar immune response-associated differential expression patterns, and showed altered gene expression in non-immune cells. In the lungs of COVID-19 patients, higher expression of ISGs was found in both immune cells and alveolar epithelial cells.<sup>2,36</sup> Furthermore, alveolar type II epithelial cells (AT2) upregulated genes for programmed cell death like *STAT1* and downregulated a gene required for maintaining AT2 identity, *ETV5*.<sup>2,36</sup> These results suggest impaired alveolar epithelial regeneration activity conferred by AT2. Monocyte-derived macrophages upregulated genes associated with activation and aberrant activation, contributing to the dysregulated inflammation.<sup>36</sup> In the brain, Yang et al. revealed that neurotransmission-mediating synaptic genes, such as *VAMP2*, *SNAP25*, and *ATP6V0C*, were downregulated in excitatory neurons but upregulated in proximal inhibitory neurons, suggesting the dysfunction of upper-layer cortical circuitry in COVID-19 patients.<sup>2</sup> Heming et al. revealed that, in CSF of COVID-19 patients, many immune cells also increased the expression of ISGs such as *IFITM2*, *IFI27*, *IFNGR2*, and *IL18* in monocytes and granulocytes and *PRF1*, *XCL1*, and *ETS1* in T and NK cells.<sup>39</sup>

In addition to comparing COVID-19 patients with healthy controls, comparison of same cell types between different tissues is an intriguing aspect. For instance, T cell activation assessed in the blood circulation of COVID-19 patients does not properly reflect T cell responses in the respiratory tract. Szabo et al. compared T cells and myeloid cells between airways and blood from COVID-19 patients.<sup>29</sup> They found that blood T cells expressed high levels of genes associated with quiescence and lymphoid homing, like *TCF7*, *LEF1*, and *SELL*. In contrast, airway T cells exhibited an activated and pro-inflammatory state and expressed high

levels of genes encoding cytokines and chemokines, like *IFNG*, *CCL2*, and *CCL4*. Airway myeloid cells also had higher expression of multiple chemokines for recruiting monocytes and macrophages, lymphocytes, neutrophils, and complement components. These results suggest that immune cells of the same type in different tissues could exert different functions during infection.

Instead of using differential expression analysis to reveal overall upregulated and downregulated genes in certain cell types, many studies performed subpopulation analysis with the identification of marker genes. This approach could help find genes with altered expression in a subset of cells from the tissue of COVID-19 patients. However, due to the variations in sampling strategy, sequencing platform and depth, and data analysis in different studies, the obtained cell subclusters were found to vary among studies and they usually showed low consistency. A large cohort of the recruited subjects is highly recommended to obtain general results shared by most COVID-19 patients, although it is often practically difficult. For example, using PBMC and BALF samples from 171 COVID-19 patients and 25 healthy controls, Ren et al. obtained 64 cell clusters from over a million cells and analyzed their association with severity.<sup>10</sup> They found some severity-associated clusters, such as *XBP1*<sup>+</sup> plasma cells, *MKI67*<sup>+</sup> plasma cells, *LILRA4*<sup>+</sup> DCs, *CCL3*<sup>+</sup>*CD14*<sup>+</sup> monocytes, *CST7*<sup>+</sup> neutrophils, *ANXA2*<sup>+</sup>*CD4*<sup>+</sup> T cells, and *IL2RB*<sup>+</sup>*CD8*<sup>+</sup> T cells. As a larger sample size is usually required to generate more reliable and general cell clusters, it is good practice to integrate previously published datasets with new sequencing data for clustering.<sup>4,18,21</sup> Previously published datasets could also be used for cross-validation of results. Factors such as batch effect, sequencing platform, and sample and clinical information need to be considered in such integrative studies.



**Figure 5. scRNA-seq reveals vital cell-cell communication in COVID-19** (A) Bioinformatic methods enable measuring the intensity of overall cell-cell communication and ligand-receptor interaction among cell populations from scRNA-seq data. (B) Vital cell-cell communication among immune cell populations highlighted in COVID-19 scRNA-seq studies. Some upregulated ligand-receptor interactions are labeled between cell populations.

In summary, both differential expression analysis and subpopulation analysis provide reliable insights into gene expression regulation during SARS-CoV-2 infection. DEGs involved in some vital antiviral pathways are shared by many immune cell types, while cell-type-specific, tissue-specific, or subcluster-specific DEGs reflect their distinct functional changes. SARS-CoV-2 infection generally induces robust innate and adaptive immune responses, as shown by the dramatic upregulation of cell activation-associated genes in a variety of immune cell types as well as the upregulation of genes encoding chemokines and inflammatory cytokines. These immune responses protect most virus-infected individuals from progression to life-threatening disease. Analysis of single-cell atlas further uncovers many pathogenic mechanisms underlying severe COVID-19, such as defects in type I IFN response and ISGs expression, aberrant activation of myeloid cells, and impaired alveolar epithelial regeneration. To deeply understand the multi-organ impact of COVID-19, future studies should make efforts to generate scRNA-seq data from less acquirable tissues than blood, and integrate new data with published data to characterize tissue-specific immune response and pathogenesis.

### CELL-CELL COMMUNICATION IN COVID-19

The immune response against SARS-CoV-2 infection and within the infected cells requires the cooperation of many cell types. The recent development of bioinformatic methods for scRNA-seq data enables prediction of cell-cell communication through a ligand-receptor interaction network. So far, some scRNA-seq studies performed cell-cell communication analysis from COVID-19 patients. The most commonly used method to infer cell-cell communication among these studies is CellPhoneDB,<sup>40</sup> while others used methods including CellChat,<sup>41</sup> ICELLNET,<sup>42</sup> NicheNet,<sup>43</sup> and CSomap.<sup>44</sup>

Most studies analyzed altered cell-cell communication among certain immune cell types in blood or BALF from COVID-19 patients (Figure 5A). The increased interaction of T cell-monocyte/macrophage interaction was underlined in several studies,<sup>34,35,45–48</sup> which constitutes an important step in immune cell activation. Other highlighted immune cell interactions include monocytes/macrophages with neutrophils,<sup>4</sup> monocytes with platelets,<sup>11</sup> and DCs with T cells (Figure 5B).<sup>38,48</sup> Using other sampled tissues from COVID-19 patients, communication among non-immune cell types was also predicted. For example, Yang et al. found increased intercellular communication from the choroid plexus epithelium to the cortex across inflammatory pathways,<sup>49</sup> and Melms et al. found enriched transforming growth factor beta (TGFβ) signaling across all major lung cell types, which promotes fibrosis.<sup>36</sup> Some studies recruited patients with different severity. Taking severity into consideration, the authors provided insights into progressive alteration of the cell-cell communication network. When comparing patients with severe/critical condition with mild/moderate condition, several studies reported that communication between T cells and antigen-pre-

sentation cells decreased,<sup>4,38,47</sup> suggesting dysfunction of the immune system at late infection stages.

Apart from overall communication among cell types, predicted ligand-receptor pairs could also reveal some features of immune response (Figure 5). Enhanced inflammatory signals were found in both immune and non-immune cells from blood and brain through two chemokine subclasses (CCL and CXCL) and interleukin families,<sup>2,4,45,46</sup> which might contribute to hyperinflammation in COVID-19. The strong SIRPA-CD47 interaction was predicted to mediate platelet activation by monocytes, leading to coagulation abnormalities.<sup>11</sup> On the other hand, decreased signaling through HLA-II-related genes was found in moderate and severe patients.<sup>38</sup> These results revealed potentially activated and deactivated signaling pathways during infection.

In summary, cell-cell communication analysis can track down the altered interactions between different cell types and the cell signaling through ligand-receptor pairs in COVID-19. However, it should be emphasized that these interactions are putative based on previous knowledge, and further validation is required before drawing definite conclusions. In addition, most studies do not analyze the effect of ligand-receptor interactions on target gene expression. We suggest that future study should pay more attention to the functional effects of cell-cell communication networks in COVID-19 using methods like NicheNet, which enables the identification of target genes affected by each ligand and involved signaling mediators.

### CELL TRAJECTORY IN COVID-19

Studying the trajectory of a disease at the cellular level is critically important for our understanding of the disease and development of precision medicine approaches. scRNA-seq provides unique opportunities for studying cellular dynamic processes by using trajectory inference (TI) analysis, also called pseudo-time analysis.<sup>50</sup> In disease studies, TI can reveal abnormal cell differentiation and gene expression alteration induced by disease progression. The detailed features and findings of trajectory analysis in COVID-19 scRNA-seq studies are summarized in Table S4.

When examining hematopoietic stem cells (HSCs) and multipotent progenitor cells (MPPs), Wang et al. identified attenuated HSCs/MPPs to megakaryocyte-erythrocyte progenitor cell (MEP) differentiation potential in COVID-19 patients. Indeed, more HSCs/MPPs from COVID-19 patients differentiated into granulocyte-monocyte progenitor cells (GMPs) rather than lymphoid progenitors.<sup>30</sup> This could potentially explain increased proportions of myeloid cells and decreased proportions of lymphocytes in both bone marrow and blood (Figure 3). Trajectory analysis also revealed uncommon paths of differentiation. Wilk et al. observed a possible lymphocyte-to-granulocyte transdifferentiation process from plasmablasts to developing neutrophils in patients with acute respiratory distress syndrome

(ARDS).<sup>8</sup> In addition, many lymphoid and myeloid cell lineages, such as DCs,<sup>38</sup> monocytes,<sup>9</sup> CD4<sup>+</sup> T, CD8<sup>+</sup> T cells, and B cells,<sup>27,51</sup> displayed continuous trajectories along COVID-19 severity, suggesting gradual changes of transcriptional profiles in response to COVID-19 infection. For example, Wauters et al. found that resident memory CD8<sup>+</sup> T cells and exhausted CD8<sup>+</sup> T cells from BALF exhibited good effector functions in mild patients but reduced effector functions in critical patients. Myeloid cells were characterized with pro-phagocytic and antigen-presentation-facilitating functions in mild patients, whereas they had disease-worsening characteristics in critical patients.<sup>4</sup> Some cell subpopulations with significant alteration in severe COVID-19 are usually located in the end of the trajectories, like differentiated neutrophils with an ISG signature and young reticulated platelets.<sup>22</sup> Another special feature was that more diverse trajectories of T cell and macrophages were observed in BALF from severe COVID-19 compared with the mild COVID-19 and control groups.<sup>52</sup> When studying the cell trajectory of recovered COVID-19 patients, Luo et al. revealed that the transcriptional profiles of CD8<sup>+</sup> T cells from discharged patients shifted from antiviral immune response to metabolism adaptation.<sup>53</sup> An integrated analysis of PBMC and BALF data by Xu et al. revealed a differentiation trajectory from circulating monocytes to macrophages in respiratory epithelium, suggesting the recruitment of peripheral monocytes into inflammatory tissues (Figure 6A).<sup>33</sup>

Critical disease severity and viral spread are often linked to the interactions between epithelial cells and immune cells. In addition to the classical path where basal cells differentiate into ciliated cells through intermediate club cells,<sup>54</sup> Chua et al. identified a subcluster of basal cells differentiating to ciliated cells, called basal-diff, bypassing the intermediate secretory cell stage.<sup>35</sup> In addition, trajectory analysis can annotate the cells with severe SARS-CoV-2 infection. For example, to determine an unknown "SARS-CoV-2<sup>hi</sup>" cell population, Ahn et al. applied trajectory analysis and identified them as ciliated cells based on their position in the trajectory.<sup>55</sup> Furthermore, TI showed the club and goblet cells could also be major precursors of differentiated multiciliated cells.<sup>55</sup> Last, Ziegler et al. inferred the trajectory of ciliated cell subtypes and showed two decreased terminally differentiated ciliated cell subclusters in severe patients, suggesting that COVID-19 preferentially affected terminally differentiated subsets. Conversely, they found an increase in the proportion of IFN-responsive ciliated cells in mild/moderate patients (Figure 6B).<sup>17</sup>

With the access to other organ samples from autopsy and organoid models, trajectory analysis indicates that the differentiation of multiple cell types is indirectly disturbed in COVID-19 patients, mostly due to the hyperinflammation across the body. In the brain of COVID-19 patients, a microglia cluster associated with COVID-19 was found to derive from a homeostatic cluster, showing the response of microglia to an inflamed central nervous system environment.<sup>49</sup> In the lung of COVID-19 patients, a trajectory of putative tuft cells was identified among airway epithelial cells. Melms et al. found that the number of tuft cells increased 3-fold in COVID-19 patients. As they are involved in airway inflammation as well as myeloid cell recruitment, these putative tuft cells might contribute to the pathophysiology of COVID-19.<sup>36</sup> In the pancreatic tissues of COVID-19 patients, Tang et al. revealed beta cell to alpha cell transdifferentiation was correlated with SARS-CoV-2 gene expression. They also identified the eIF2 pathway regulating this process.<sup>56</sup> By profiling SARS-CoV-2-infected colon and ileum organoids, Triana et al. reported that IFN-mediated signaling was correlated with the trajectory in non-infected cells beside infected cells (i.e., bystander cells).<sup>57</sup>

In summary, abnormal cell differentiation and transition were identified in multiple tissues and organs from COVID-19 patients by trajectory analysis. Immune cell populations showed skewed differentiation from HSCs/MPPs to myeloid cells and transition of transcriptional profiles along disease severity. This infection-induced altered myelopoiesis, also called emergency myelopoiesis, is a hallmark of severe COVID-19. For non-immune cells like epithelial cells, their trajectories usually corresponded to SARS-CoV-2 infection and abnormal paths of differentiation. Even though the abnormal differentiation and transdifferentiation have been discovered in many diseases, the interpretation of such cell trajectories should be carefully examined because the findings of abnormal transdifferentiation may not be supported in other datasets.<sup>28</sup> Therefore, validation using other datasets as well as other bioinformatic or experimental methods is strongly recommended.

## SINGLE-CELL ANALYSIS OF IMMUNE REPERTOIRE IN COVID-19

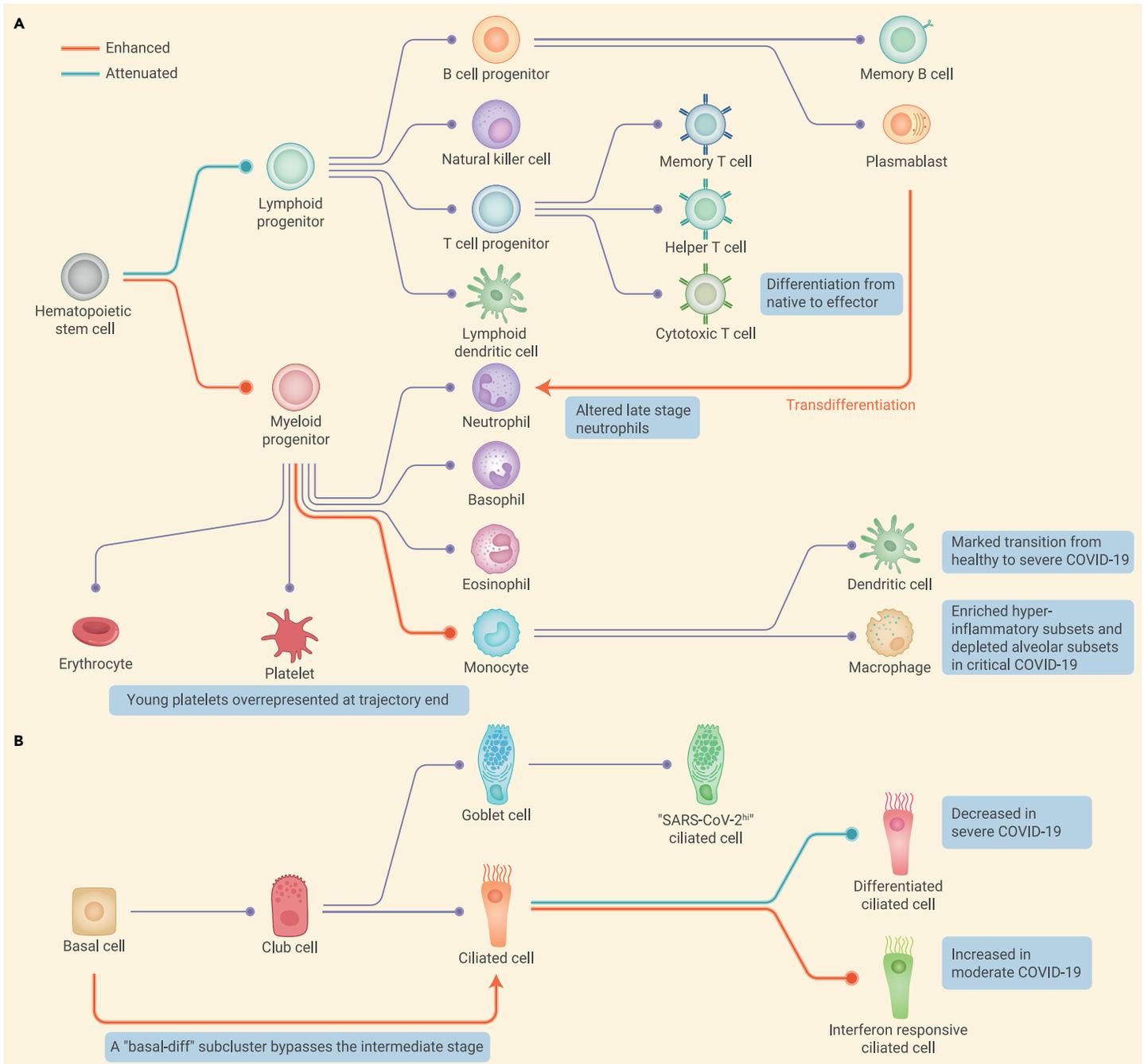
T cells and B cells mediate antigen-specific adaptive immunity by recognizing and eliminating antigens originating from infection and disease. Antigen specificity refers to the ability of a particular T cell or B cell to recognize the antigen epitope in a specific manner. Expression of a unique T cell receptor (TCR) or B cell receptor (BCR; membrane-bound immunoglobulin [Ig]) on the surface of each T cell or B cell is central to specific antigen recognition, respectively. The vast diversity of TCRs and BCRs expressed on an enormous number of T cells and B cells allows for the recognition of a broad spectrum of antigens. The immune repertoire refers to the collection of TCRs and BCRs (or T cell and B cell clonotypes) present in an individual.

The vast diversity of TCRs and BCRs is mainly generated in developing T cells and B cells by a somatic recombination process called V(D)J recombination. Through V(D)J recombination, the variable region of TCR or BCR genes is assembled from component V, D, and J gene segments, resulting in a potential diversity of over 10<sup>13</sup> unique TCR and BCR sequences (or T cell and B cell clonotypes), respectively. Through targeted enrichment, transcripts containing VDJ regions are sequenced in single T cells or B cells using a similar protocol to scRNA-seq, named single-cell VDJ sequencing (scVDJ-seq). As scVDJ-seq enables studying adaptive immune response against foreign antigens, many studies have used it to analyze the alteration of TCRs and BCRs after COVID-19 infection.

T cell clone diversity and expansion in COVID-19 patients are primary interests in many COVID-19 scVDJ-seq studies (Figure 7A). Most studies profiled T or B cells in PBMCs using scVDJ-seq. Several studies obtained the same results that TCR diversity decreased significantly in severe and critical patients compared with that in mild/moderate patients and healthy people.<sup>10,58–60</sup> Slightly different results were obtained in bone marrow. Wang et al. found lower diversity in CD4<sup>+</sup> T cells but higher diversity in CD8<sup>+</sup> T cells from BMDCs of severe patients.<sup>30</sup> For convalescent patients recovered from mild COVID-19, their TCR diversity was high at the recovering stage but quickly decreased after that.<sup>53,59</sup> As for T cell clonal expansion, many studies found higher expansion in mild/moderate patients compared with healthy patients,<sup>7,12,31,58</sup> but there are contradictory results for severe patients. The results in several studies reported some inconsistent observations. For example, Liu et al. and Stephenson et al. showed that there were increased clonal expansion levels in CD8<sup>+</sup> T cells in severe patients,<sup>11,18</sup> but Liao et al. and Xu et al. showed that there were decreased clonal expansion levels in CD8<sup>+</sup> T cells.<sup>7,27</sup> Meckiff et al. and Xu et al. reported increased clonal expansion levels in CD4<sup>+</sup> T cells in severe patients,<sup>26,27</sup> but Liu et al. showed decreased clonal expansion levels CD4<sup>+</sup> T cells.<sup>18</sup> In addition, T cell clonal expansion in convalescent patients was found either decreased<sup>5,53</sup> or increased<sup>31</sup> compared with healthy people. Discordance was also found in TCR gene preponderance from multiple studies,<sup>30,31,60–62</sup> suggesting that many TCR clonotypes could confer recognition capability. These contradictory results may be due to different severity standards, infection stages, individual specificity, or technical variation among studies. Therefore, these results require further investigation and validation in future studies.

Integrated analysis of single-cell VDJ data and transcriptomic data could reveal the heterogeneity of TCR alteration in cell subpopulations (Figure 7B). For example, Meckiff et al. found the greatest CD4<sup>+</sup> cytotoxic T cell clonal expansion in hospitalized patients.<sup>26</sup> Stephenson et al. revealed that the proportion of expanded effector CD8<sup>+</sup> T cells increased, but that of expanded effector memory CD8<sup>+</sup> T cells decreased in severe patients.<sup>11</sup> Wang et al. reported that expanded T cells in PBMCs mainly comprised CD4<sup>+</sup>GZMH<sup>high</sup> T cells, CD8<sup>+</sup>GZMH<sup>high</sup> T cells, and CD8<sup>+</sup>GZMK<sup>high</sup> T cells.<sup>34</sup> Wen et al. found that naive and central memory T cells were slightly expanded, while effector memory T cells, terminal effector CD8<sup>+</sup> T cells, and proliferating T cells were highly expanded in patients at the early recovery stage.<sup>5</sup> These results highlight vital T cell subpopulations with the most clonal expansion during COVID-19 infection. Currently, it is difficult to compare the results from different studies because they usually have different classification and annotation of subpopulations. To address this issue, follow-up studies performing integration on multiple datasets are required. As we discussed earlier, data sharing is important for the research community.

Apart from its application in T cells, scVDJ-seq has been used to study BCR repertoire in COVID-19 patients. In contrast to TCRs, there are more consistent results for BCRs. COVID-19 patients generally had lower BCR diversity and higher BCR expansion,<sup>5,11,21,31,59,60,63</sup> but BCR expansion decreased during recovery (Figure 7A).<sup>58</sup> Skewed usage of IGHV3 gene



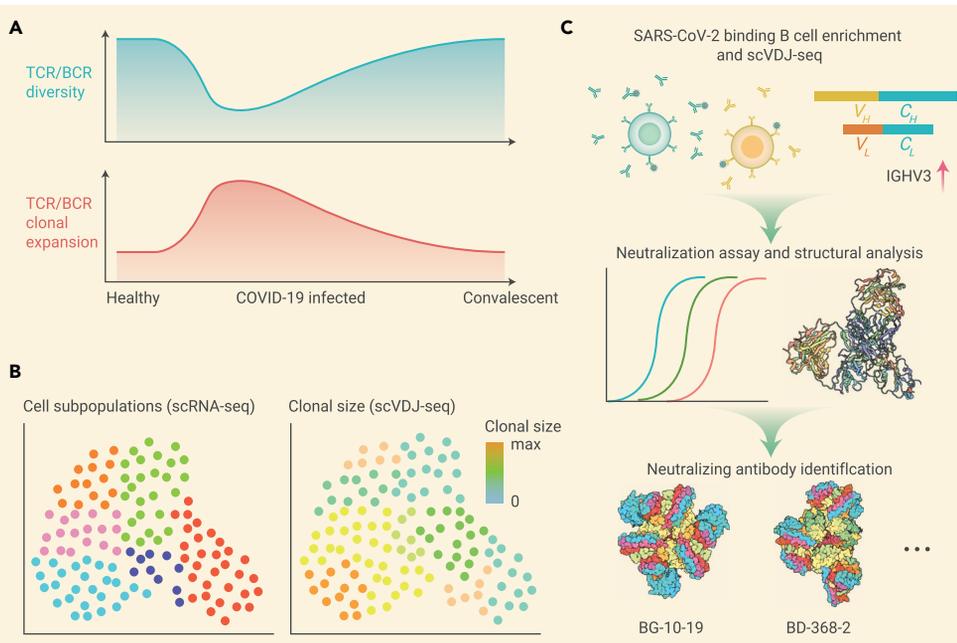
**Figure 6. Differentiation trajectory of two major cell lineages in the analysis of scRNA-seq data from COVID-19 patients** (A) Differentiation trajectory of hematopoietic stem cells. (B) Differentiation trajectory of epithelial cells. Lineages highlighted in red or blue indicate an enhanced or attenuated differentiation pattern respectively in COVID-19 patients. The red arrows indicate potential differentiation trajectory particular to COVID-19 patients, bypassing conventional differentiation.

family was revealed by most studies,<sup>5,10,21,31,51,60</sup> like *IGHV3-23* and *IGHV3-30*, but the usage of other gene families like *IGHV1* and *IGHV4*, was also reported.<sup>58,63–65</sup> In addition, scVDJ-seq could be used to identify somatic hypermutations within the complementarity-determining regions (CDRs) in BCRs of COVID-19 patients.<sup>11,30,51,60</sup> Considering the effect of factors like age and sex, more BCR mutations were found in female than male patients,<sup>11</sup> and young patients than old patients.<sup>51</sup> Similarly, higher TCR diversity was found in female and young patients.<sup>10</sup> These results collectively suggest clinal features of COVID-19 patients have impacts on their adaptive immune response.

An emerging application of scVDJ-seq is to accelerate the identification of neutralizing antibodies for COVID-19 treatment (Figure 7C).<sup>48,66–68</sup> Prior to the invention of scVDJ-seq, RT-PCR was used to obtain antibody sequences from a limited number of individual B cells. The scVDJ-seq technique has outperformed RT-PCR and can obtain tens of thousands of antibody sequences in one run. Cao

et al. have recently used scVDJ-seq to analyze antigen-binding B cells from convalescent COVID-19 patients. They identified 14 potent neutralizing antibodies from 8,558 antigen-binding IgG1<sup>+</sup> clonotypes.<sup>66</sup> Scheid et al. investigated SARS-CoV-2 spike-specific B cell responses in COVID-19 patients who had recovered from the disease. By combining analyses of scVDJ-seq, scRNA-seq, and monoclonal antibody (mAb) structures, they identified a mAb clone, BG10-19, that could neutralize different SARS-CoV-2 variants by locking the virus spike trimer in a closed conformation.<sup>68</sup> These studies demonstrate the efficiency of scVDJ-seq for identifying SARS-CoV-2-neutralizing antibodies.

In summary, scVDJ-seq characterizes the change of immune repertoire in response to SARS-CoV-2 infection, and helps identify SARS-CoV-2-neutralizing antibodies. Integrated analysis of scRNA-seq and scVDJ-seq can help identify the highly expanded cell subpopulations. Regarding the BCR, scVDJ-seq can dramatically accelerate the identification of neutralizing antibodies, which currently represents the most significant therapeutic application.



**Figure 7. scVDJ-seq characterizes immune repertoire in COVID-19** (A) A diagram showing the trend of the diversity and clonal expansion of TCR and BCR in COVID-19. Measured from scVDJ-seq data, the diversity of TCR and BCR generally decreases in the condition of COVID-19 infection, and the clonal expansion generally increases in COVID-19 infection. The curves may not apply to all severity and cell subpopulations. (B) Integrated analysis of transcriptomic and VDJ data could identify cell subpopulations with high clonal expansion. Cell subpopulations are identified using scRNA-seq data, and the clonal sizes of corresponding cells are estimated using scVDJ-seq data. (C) scVDJ-seq helps find potent neutralizing antibodies against SARS-CoV-2. Other methods like neutralization assay and cryo-EM structural analysis are usually needed for the final identification and validation of neutralizing antibodies.

cytoplasm rather than in the nucleus of host cells, little or no detection of viral reads in several organs is more likely a consequence of technical limitation rather than the low rate of SARS-CoV-2 infection in these cells.<sup>2</sup>

In addition to scRNA-seq, other molecular approaches, such as qPCR,<sup>66</sup> digital spatial profiling,<sup>72</sup> and bulk RNA sequencing, have

been applied to detect the expression of SARS-CoV-2 genes. For example, by using bulk RNA sequencing to quantify viral gene expression in an infected human lung epithelial cell line, Wyler et al. found an increased relative amount of *ORF7a* at a late time post infection.<sup>73</sup> However, these approaches lack single-cell resolution. Future COVID-19 scRNA-seq studies could address this issue by including the analysis of SARS-CoV-2 gene expression.

In summary, appropriate scRNA-seq library preparation and analysis methods will enable us to detect SARS-CoV-2 in different cell populations. We strongly recommend 5' library preparation protocols for more accurate quantification of SARS-CoV-2 gene expression at single-cell resolution.

#### SCRNA-SEQ FOR MULTI-OMICS ANALYSIS

The scRNA-seq technology has been used with other genomic evidence (such as common and rare variants) to unveil COVID-19 disease etiology. Large-scale genome-wide association studies (GWASs) have revealed that many host genetic factors confer intrinsic susceptibility to COVID-19 severity.<sup>74,75</sup> Dai et al. conducted a systematic fine-mapping of the COVID-19 severity GWAS at the 3p21.31 locus using tissue and cell-type expression quantitative trait loci (eQTL) data.<sup>76</sup> They further validated that the protective gene *CXCR6* had lower expression in severe patients than in moderate patients from BALF scRNA-seq data, indicating *CXCR6* might play an important role in first-line defense of human lung tissue-resident memory T cells. Qi et al. developed a network-based approach that is implemented in a user-friendly software Network Calculator to measure the proximity of the risk genes of GWAS severe COVID-19 symptoms and virus-host interactome genes.<sup>77</sup> They identified that these genes tend to have more interactions, providing some insights for the genetic basis of host susceptibility. They further identified these genes were highly enriched in the macrophages, T cells, and epithelial cells of severe patients compared with mild patients. To understand the genetic basis of severe COVID-19 in non-elderly adults, Zhang et al. performed whole-exome sequencing and identified common and rare disease-associated genetic variants within over 1,000 risk genes.<sup>78</sup> By integrating single-cell multi-omics profilings of human lung data, they identified particularly enriched risk genes in NK cells. The Mendelian randomization indicates the proportion of NK cells have a causal relationship with critical illness of COVID-19. Last, the 5' scRNA-seq technology could accurately capture the transcriptome of both host and virus. Liu et al. explicitly identified and quantified the cells with virus transcripts in COVID-19 scRNA-seq data.<sup>70</sup> They further identified that the viral *ORF10* transcript was differentially expressed between COVID-19 severe patients and moderate patients, suggesting an association between viral transcripts and COVID-19 severity.

scRNA-seq technology has many applications along with other measurements, such as ligand-receptor analysis, cell surface protein prediction, and cell

#### QUANTIFICATION OF SARS-COV-2 GENE EXPRESSION

Less attention has been paid to the expression of SARS-CoV-2 genes compared with host genes in most COVID-19 scRNA-seq studies. Only a few studies analyzed the expression of SARS-CoV-2 genes in individual cells. This was usually achieved by using a merged human-SARS-CoV-2 genome with merged annotation as reference in read alignment. The aligned reads were then used to quantify individual viral gene expression or measure overall coverage on the viral genome.<sup>2,56</sup> Some studies also considered strand information involving viral replication.<sup>3,17</sup> Bost et al. developed Viral-Track,<sup>1</sup> a computational method to detect viral RNA in scRNA-seq, and applied it to SARS-CoV-2 expression data. Several recent studies also used Viral-Track to detect viral reads<sup>2,30,32</sup> and then quantify viral gene expression.<sup>4</sup> Among non-immune cells from BALF, there was a strong enrichment of viral reads in ciliated and epithelial progenitor cells,<sup>1,17</sup> which express both *ACE2* and *TMPRSS2*. Among various immune cell types, strong enrichment of viral reads was found in macrophages,<sup>1,3,4</sup> suggesting that macrophages could be infected by SARS-CoV-2 or phagocytose viral particles.<sup>1,3</sup> Infection was also observed in other cell types, like neutrophils, monocytes, DCs, and T cells, which do not necessarily express *ACE2* and *TMPRSS2*.<sup>2-4,66</sup> Moreover, Wauters et al. reported the specificity of viral gene expression in different cell types. Viral spike protein coding gene *S* was almost exclusively detected in epithelial cells, while nucleocapsid protein coding gene *N* was detected more frequently in myeloid and lymphoid cells, especially in neutrophils and macrophages, than in epithelial cells.<sup>4</sup> Differential expression analysis of viral genes in the infected cells of patients with different severity revealed the correlation between viral gene *ORF10* expression level and severity.<sup>69,70</sup> The expression of viral genes was also examined in other organs. Tang et al. found high viral gene expression in acinar cells, alpha cells, beta cells, ductal cells, and fibroblasts from the infected islet.<sup>56</sup> Delorey et al. revealed strong enrichment of viral reads in myeloid cells and slight enrichment in endothelial cells from the lung.<sup>2</sup> Interestingly, a recent study by Yang et al. did not detect viral reads in the brain, possibly due to the brain barriers.<sup>49</sup> This preliminary study may need further investigation.

Due to the limitation of single-cell library preparation methods, some protocols capturing the 3' end of transcripts, like 10x Genomics 3' Gene Expression, are inappropriate to quantify SARS-CoV-2 gene expression. The alignment results from these methods will have extremely high coverage on the 3' end of SARS-CoV-2 genomes and low coverage elsewhere,<sup>32,69</sup> leading to inaccurate quantification of viral genes on the 3' end of the SARS-CoV-2 genome except for *ORF1ab*. This is because subgenomic RNAs transcribed from those genes share the same 3' end sequence and poly(A) tail.<sup>71</sup> The other inappropriate protocol for the quantification of SARS-CoV-2 gene expression is single-nucleus RNA sequencing (snRNA-seq), which only profiles transcripts in the nucleus by isolating nuclei from single cells. Since the transcription of SARS-CoV-2 primarily occurs in the

surface imaging technology, to uncover the cellular microenvironment and inter-cellular signaling alteration in COVID-19 patients. Ramaswamy et al. profiled samples from COVID-19-induced multisystem inflammatory syndrome in children (MIS-C), adult COVID-19, and healthy controls using scRNA-seq and serum proteomics.<sup>79</sup> By analyzing the antigen receptor repertoire, they reported that MIS-C patients had increased S100A alarmin expression and decreased antigen-presentation signatures in myeloid cells, which is consistent with the serum proteome signature in MIS-C patients with myeloid inflammatory responses. Cheng et al. developed a multilayer network method (scMLnet) to characterize the inter-/intra-cellular signaling network constructed from ligand-receptor interactions, receptor-transcription factor (TF) pathways, and TF-target interactions.<sup>80</sup> By using COVID-19 scRNA-seq data, they successfully predicted the elevation of the cell surface protein ACE2 by extracellular cytokines. He et al. used COVID-19 scRNA-seq data and known ligand-receptor interactions to train a risk model that could predict inflammatory damage score and counts of T cells using serum cytokine markers.<sup>81</sup> Lee et al. identified that *SLC2A3* in macrophage groups could be a possible surface protein to distinguish severe COVID-19 patients by using COVID-19 scRNA-seq data.<sup>82</sup> This study provided a pipeline to use scRNA-seq data to identify candidate surface and druggable targets with molecular imaging technology.

In summary, scRNA-seq has been widely integrated with omics data and cell surface protein activity prediction in COVID-19 research, which greatly advances our knowledge of COVID-19 at the molecular level. Furthermore, there are many other studies using scRNA-seq from COVID-19 patients to conduct therapeutic drug development or validation.<sup>81,83,84</sup> Many studies also explored the post-COVID-19 symptoms by cytokine storm on diverse organs by using scRNA-seq and other omics data.<sup>2,85</sup> Mutual corroboration of results in transcriptomics and other omics substantially improves the reliability of new findings, which has been more recognized in the current COVID-19 research.<sup>11,12,21,86</sup>

## CONCLUDING REMARKS AND PERSPECTIVES

scRNA-seq has demonstrated strength and potential in COVID-19 research. In this article, we review its broad application and major findings from over 80 original studies as well as many secondary analysis studies. Here, we discuss the limitations of current studies and propose some perspectives for future advance.

Defining severity is the foundation for COVID-19 studies, and many COVID-19 scRNA-seq studies included patients with different severity. Up to now, there are several standards for severity definition. It is unknown whether and how using different severity definitions would generate inconsistent results. For scRNA-seq studies, a clear and universal severity definition could guarantee replicable analyses and results, and therefore the most used classification is suggested for future studies.

Up to now, the sampled tissue is limited to PBMC in most COVID-19 scRNA-seq studies, as blood is easy to collect from subjects, while other tissue, such as lung, is very difficult to obtain. Despite the difficulty in acquiring samples, more efforts are needed to study other tissues and organs with SARS-CoV-2 infection. These studies should cover topics that cannot be explained by the alteration of PBMCs, such as the immune response of tissue-resident immune cells, the recruitment of circulating immune cells into tissues, non-immune cell driving pathological symptoms like lung fibrosis and neural disorders, as well as the susceptibility and vulnerability of different cell types against SARS-CoV-2.

In addition to using samples from COVID-19 patients for research, developing animal models for COVID-19 and subjecting them to scRNA-seq could facilitate preclinical analysis of vaccines and therapeutic agents. Although there are many studies using infected animals for COVID-19 research,<sup>87,88</sup> only a few studies performed scRNA-seq on COVID-19-infected animals like monkeys,<sup>89</sup> hamsters,<sup>90</sup> macaques,<sup>91</sup> ferrets,<sup>92</sup> and K18-hACE2 transgenic mice.<sup>93</sup> Several studies demonstrated the antiviral immune responses of animal models after receiving newly developed vaccines with scRNA-seq,<sup>94–97</sup> revealing induced gene modules of different immune cell populations. As it is much easier to obtain samples from various organs in animal models than in human patients and at consecutive time points after COVID-19 infection, vaccination, or therapy, performing temporal-spatial analysis at the single-cell resolution in animal models for COVID-19 could be a promising future direction. Among those animal studies, the findings from non-human primates will be more useful in general, but they are also more expensive and under stronger regulation.

Another important issue is the selection of scRNA-seq protocols. In terms of COVID-19 studies, 5'-end library preparation is prioritized over 3' end. First, it enables more read coverage of SARS-CoV-2 genes and more accurate quantification. Subsequent bioinformatic analysis could use either a merged human-SARS-CoV-2 genome as reference or Viral-Track to detect and quantify viral gene expression in single cells. Second, 5'-end library preparation can be integrated with scVDJ-seq to profile both transcriptome and immune repertoire. Therefore, we recommend 5'-end scRNA-seq for future studies concerning SARS-CoV-2 gene expression and TCR/BCR profiling. In addition, snRNA-seq is not recommended because it is unlikely to capture viral transcripts in the cytoplasm.

During bioinformatic analysis of scRNA-seq data, the clustering and annotation of cell subpopulations usually vary in studies, even if samples are collected from the same tissues. This has a strong impact on downstream analyses like identifying cell proportion alteration and differential expression analysis. Authors could obtain different results with distinct cluster annotations, which makes it hard to compare and evaluate the results of subpopulation analysis from different studies. Although it is impossible to create a comprehensive atlas as a reference for every study, we suggest integrating multiple datasets for analysis as a reasonable compromise. This could also help study whether the viral strains prevalent in different locations and time points induce similar immune responses. Therefore, data sharing and reproducibility are important for follow-up studies during the ongoing pandemic.

While scRNA-seq has improved our understanding of the infection and immune response of COVID-19, integrating it with other technology to generate multi-omics data would provide valuable insights from other aspects. There are many promising technologies and methods to study COVID-19 in combination with scRNA-seq, such as GWAS for identifying risk genes or alleles for COVID-19,<sup>78</sup> single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq) for profiling chromatin accessibility,<sup>45,86,98</sup> cellular indexing of transcriptomes and epitopes by sequencing (CITE-seq) for profiling surface proteome,<sup>11</sup> and spatial transcriptomics for acquiring spatial information.<sup>2</sup> The multi-omics data would facilitate characterizing an integrated framework of COVID-19 pathogenesis and immune response.

Finally, thoroughness is a limitation of this review. First, the collection of articles from PubMed using our query does not include every paper in this field. Second, as COVID-19 is an urgently needed area of research, new studies are consistently published while writing this review. To address this concern, a database tracking and curating single-cell COVID-19 studies and datasets would facilitate future research.

## REFERENCES

- Bost, P., Giladi, A., Liu, Y., et al. (2020). Host-viral infection maps reveal signatures of severe COVID-19 patients. *Cell* **181**, 1475–1488.e12.
- Delorey, T.M., Ziegler, C.G.K., Heimberg, G., et al. (2021). COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature* **595**, 107–113.
- Grant, R.A., Morales-Nebreda, L., Markov, N.S., et al.; NU SCRIPT Study Investigators (2021). Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia. *Nature* **590**, 635–641.
- Wauters, E., Van Mol, P., Garg, A.D., et al. (2021). Discriminating mild from critical COVID-19 by innate and adaptive immune single-cell profiling of bronchoalveolar lavages. *Cell Res.* **31**, 272–290.
- Wen, W., Su, W., Tang, H., et al. (2020). Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. *Cell Discov.* **6**, 41.
- Hoffmann, M., Kleine-Weber, H., Schroeder, S., et al. (2020). SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e8.
- Liao, M., Liu, Y., Yuan, J., et al. (2020). Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* **26**, 842–844.
- Wilk, A.J., Rustagi, A., Zhao, N.Q., et al. (2020). A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nat. Med.* **26**, 1070–1076.
- Lee, J.S., Park, S., Jeong, H.W., et al. (2020). Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Sci. Immunol.* **5**, eabd1554.
- Ren, X., Wen, W., Fan, X., et al. (2021). COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell* **184**, 5838–1913e19.
- Stephenson, E., Reynolds, G., Botting, R.A., et al. (2021). Single-cell multi-omics analysis of the immune response in COVID-19. *Nat. Med.* **27**, 904–916.
- Su, Y., Chen, D., Yuan, D., et al. (2020). Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19. *Cell* **183**, 1479–1495.e20.
- van der Wijst, M.G.P., Vazquez, S.E., Hartoularos, G.C., et al.; UCSF COMET consortium (2021). Type I interferon autoantibodies are associated with systemic immune alterations in patients with COVID-19. *Sci. Transl. Med.* **13**, eabh2624.

14. WHO (2021). Therapeutics and COVID-19 (WHO). <https://apps.who.int/iris/bitstream/handle/10665/342368/WHO-2019-nCoV-therapeutics-2021.2-eng.pdf>.
15. WHO (2020). R&D Blueprint: Novel Coronavirus: COVID-19 Therapeutic Trial Synopsis (WHO). <https://www.who.int/docs/default-source/blue-print/covid-19-therapeutic-trial-synopsis.pdf>.
16. NHCC (2020). Diagnosis and treatment protocol for novel coronavirus pneumonia (Trial Version 7). *Chin. Med. J. (Engl.)* **133**, 1087–1095.
17. Ziegler, C.G.K., Miao, V.N., Owings, A.H., et al. (2021). Impaired local intrinsic immunity to SARS-CoV-2 infection in severe COVID-19. *Cell* **184**, 4713–4733.e22.
18. Liu, C., Martins, A.J., Lau, W.W., et al. (2021). Time-resolved systems immunology reveals a late juncture linked to fatal COVID-19. *Cell* **184**, 1836–1857.e22.
19. Hasan, M.Z., Islam, S., Matsumoto, K., and Kawai, T. (2021). Meta-analysis of single-cell RNA-seq data reveals phenotypic switching of immune cells in severe COVID-19 patients. *Comput. Biol. Med.* **137**, 104792.
20. Lam, M.T.Y., Duttke, S.H., Odish, M.F., et al. (2021). Profiling transcription initiation in peripheral leukocytes reveals severity-associated cis-regulatory elements in critical COVID-19. Preprint at bioRxiv. <https://doi.org/10.1101/2021.08.24.457187>.
21. Bernades, J.P., Mishra, N., Tran, F., et al.; HCA Lung Biological Network; Deutsche COVID-19 Omics Initiative DeCOI (2020). Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19. *Immunity* **53**, 1296–1314.e9.
22. Combes, A.J., Courau, T., Kuhn, N.F., et al. (2021). Global absence and targeting of protective immune states in severe COVID-19. *Nature* **591**, 124–130.
23. de Cevins, C., Luka, M., Smith, N., et al. (2021). A monocyte/dendritic cell molecular signature of SARS-CoV-2-related multisystem inflammatory syndrome in children with severe myocarditis. *Med* **2**, 1072–1092.e7.
24. Flament, H., Rouland, M., Beaudoin, L., et al. (2021). Outcome of SARS-CoV-2 infection is linked to MAIT cell activation and cytotoxicity. *Nat. Immunol.* **22**, 322–335.
25. Krämer, B., Knoll, R., Bonaguro, L., et al. (2021). Early IFN- $\alpha$  signatures and persistent dysfunction are distinguishing features of NK cells in severe COVID-19. *Immunity* **54**, 2650–2669.e14.
26. Meckiff, B.J., Ramírez-Suástegui, C., Fajardo, V., et al. (2020). Imbalance of regulatory and cytotoxic SARS-CoV-2-reactive CD4(+) T cells in COVID-19. *Cell* **183**, 1340–1353.e16.
27. Schulte-Schrepping, J., Reusch, N., Paclik, D., et al.; Deutsche COVID-19 OMICS Initiative DeCOI (2020). Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell* **182**, 1419–1440.e23.
28. Silvino, A., Chapuis, N., Dunsmore, G., et al. (2020). Elevated calprotectin and abnormal myeloid cell subsets discriminate severe from mild COVID-19. *Cell* **182**, 1401–1418.e18.
29. Szabo, P.A., Dogra, P., Gray, J.I., et al. (2021). Longitudinal profiling of respiratory and systemic immune responses reveals myeloid cell-driven lung inflammation in severe COVID-19. *Immunity* **54**, 797–814.e6.
30. Yang, B., Fan, J., Huang, J., et al. (2021). Clinical and molecular characteristics of COVID-19 patients with persistent SARS-CoV-2 infection. *Nat. Commun.* **12**, 3501.
31. Zhu, L., Yang, P., Zhao, Y., et al. (2020). Single-cell sequencing of peripheral mononuclear cells reveals distinct immune response landscapes of COVID-19 and influenza patients. *Immunity* **53**, 685–696.e3.
32. Bost, P., De Sanctis, F., Canè, S., et al. (2021). Deciphering the state of immune silence in fatal COVID-19 patients. *Nat. Commun.* **12**, 1428.
33. Xu, G., Qi, F., Li, H., et al. (2020). The differential immune responses to COVID-19 in peripheral and lung revealed by single-cell RNA sequencing. *Cell Discov.* **6**, 73.
34. Wang, X., Wen, Y., Xie, X., et al. (2021). Dysregulated hematopoiesis in bone marrow marks severe COVID-19. *Cell Discov.* **7**, 60.
35. Chua, R.L., Lukassen, S., Trump, S., et al. (2020). COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* **38**, 970–979.
36. Melms, J.C., Biermann, J., Huang, H., et al. (2021). A molecular single-cell lung atlas of lethal COVID-19. *Nature* **595**, 114–119.
37. Zhang, J.Y., Wang, X.M., Xing, X., et al. (2020). Single-cell landscape of immunological responses in patients with COVID-19. *Nat. Immunol.* **21**, 1107–1118.
38. Saichi, M., Ladjemi, M.Z., Korniotis, S., et al. (2021). Single-cell RNA sequencing of blood antigen-presenting cells in severe COVID-19 reveals multi-process defects in antiviral immunity. *Nat. Cell Biol.* **23**, 538–551.
39. Heming, M., Li, X., Räuber, S., et al. (2021). Neurological manifestations of COVID-19 feature T cell exhaustion and dedifferentiated monocytes in cerebrospinal fluid. *Immunity* **54**, 164–175.e6.
40. Efremova, M., Vento-Tormo, M., Teichmann, S.A., and Vento-Tormo, R. (2020). CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. *Nat. Protoc.* **15**, 1484–1506.
41. Jin, S., Guerrero-Juarez, C.F., Zhang, L., et al. (2021). Inference and analysis of cell-cell communication using CellChat. *Nat. Commun.* **12**, 1088.
42. Noël, F., Massenet-Regad, L., Carmi-Levy, I., et al. (2021). Dissection of intercellular communication using the transcriptome-based framework ICELLNET. *Nat. Commun.* **12**, 1089.
43. Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162.
44. Ren, X., Zhong, G., Zhang, Q., et al. (2020). Reconstruction of cell spatial organization from single-cell RNA sequencing data based on ligand-receptor mediated self-assembly. *Cell Res.* **30**, 763–778.
45. Zheng, Y., Liu, X., Le, W., et al. (2020). A human circulating immune cell landscape in aging and COVID-19. *Protein Cell* **11**, 740–770.
46. Fischer, D.S., Ansari, M., Wagner, K.I., et al. (2021). Single-cell RNA sequencing reveals ex vivo signatures of SARS-CoV-2-reactive T cells through 'reverse phenotyping'. *Nat. Commun.* **12**, 4515.
47. Xie, X., Cheng, X., Wang, G., et al. (2021). Single-cell transcriptomes of peripheral blood cells indicate and elucidate severity of COVID-19. *Sci. China Life Sci.* **64**, 1634–1644.
48. Song, E., Bartley, C.M., Chow, R.D., et al. (2021). Divergent and self-reactive immune responses in the CNS of COVID-19 patients with neurological symptoms. *Cell Rep. Med.* **2**, 100288.
49. Yang, A.C., Kern, F., Losada, P.M., et al. (2021). Dysregulation of brain and choroid plexus cell types in severe COVID-19. *Nature* **595**, 565–571.
50. Saelens, W., Cannoodt, R., Todorov, H., and Saeys, Y. (2019). A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* **37**, 547–554.
51. Bieberich, F., Vazquez-Lombardi, R., Yermanos, A., et al. (2021). A single-cell atlas of lymphocyte adaptive immune repertoires and transcriptomes reveals age-related differences in convalescent COVID-19 patients. *Front. Immunol.* **12**, 701085.
52. Jeong, H.H., Jia, J., Dai, Y., et al. (2021). Investigating cellular trajectories in the severity of COVID-19 and their transcriptional programs using machine learning approaches. *Genes* **12**, 635.
53. Luo, L., Liang, W., Pang, J., et al. (2021). Dynamics of TCR repertoire and T cell function in COVID-19 convalescent individuals. *Cell Discov.* **7**, 89.
54. Pardo-Saganta, A., Law, B.M., Tata, P.R., et al. (2015). Injury induces direct lineage segregation of functionally distinct airway basal stem/progenitor cell subpopulations. *Cell Stem Cell* **16**, 184–197.
55. Ahn, J.H., Kim, J., Hong, S.P., et al. (2021). Nasal ciliated cells are primary targets for SARS-CoV-2 replication in the early stage of COVID-19. *J. Clin. Invest.* **131**, e148517.
56. Tang, X., Uhl, S., Zhang, T., et al. (2021). SARS-CoV-2 infection induces beta cell transdifferentiation. *Cell Metab.* **33**, 1577–1591.e7.
57. Triana, S., Metz-Zumaran, C., Ramirez, C., et al. (2021). Single-cell analyses reveal SARS-CoV-2 interference with intrinsic immune response in the human gut. *Mol. Syst. Biol.* **17**, e10232.
58. Zhao, X.N., You, Y., Cui, X.M., et al. (2021). Single-cell immune profiling reveals distinct immune response in asymptomatic COVID-19 patients. *Signal Transduct. Target. Ther.* **6**, 342.
59. Huang, L., Shi, Y., Gong, B., et al. (2021). Dynamic blood single-cell immune responses in patients with COVID-19. *Signal Transduct. Target. Ther.* **6**, 110.
60. Zhao, Z., Zhao, Y., Zhou, Y., et al. (2020). Single-cell analysis identified lung progenitor cells in COVID-19 patients. *Cell Prolif.* **53**, e12931.
61. Agerer, B., Kobischke, M., Gudipati, V., et al. (2021). SARS-CoV-2 mutations in MHC-I restricted epitopes evade CD8(+) T cell responses. *Sci. Immunol.* **6**, eabg6461.
62. Shiakolas, A.R., Kramer, K.J., Wrapp, D., et al. (2021). Cross-reactive coronavirus antibodies with diverse epitope specificities and Fc effector functions. *Cell Rep. Med.* **2**, 100313.
63. Jin, X., Zhou, W., Luo, M., et al. (2021). Global characterization of B cell receptor repertoire in COVID-19 patients by single-cell V(D)J sequencing. *Brief. Bioinform.* **22**, bbab192.
64. He, B., Liu, S., Wang, Y., et al. (2021). Rapid isolation and immune profiling of SARS-CoV-2 specific memory B cell in convalescent COVID-19 patients via LIBRA-seq. *Signal Transduct. Target. Ther.* **6**, 195.
65. Montague, Z., Lv, H., Otwinowski, J., et al. (2021). Dynamics of B cell repertoires and emergence of cross-reactive responses in patients with different severities of COVID-19. *Cell Rep.* **35**, 109173.
66. Cao, Y., Su, B., Guo, X., et al. (2020). Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* **182**, 73–84.e16.
67. Dugan, H.L., Stamper, C.T., Li, L., et al. (2021). Profiling B cell immunodominance after SARS-CoV-2 infection reveals antibody evolution to non-neutralizing viral targets. *Immunity* **54**, 1290–1303.e7.
68. Scheid, J.F., Barnes, C.O., Eraslan, B., et al. (2021). B cell genomics behind cross-neutralization of SARS-CoV-2 variants and SARS-CoV. *Cell* **184**, 3205–3221.e24.
69. Ravindra, N.G., Alfajaro, M.M., Gasque, V., et al. (2021). Single-cell longitudinal analysis of SARS-CoV-2 infection in human airway epithelium identifies target cells, alterations in gene expression, and cell state changes. *PLoS Biol.* **19**, e3001143.
70. Liu, T., Jia, P., Fang, B., and Zhao, Z. (2020). Differential expression of viral transcripts from single-cell RNA sequencing of moderate and severe COVID-19 patients and its implications for case severity. *Front. Microbiol.* **11**, 603509.
71. Kim, D., Lee, J.Y., Yang, J.S., et al. (2020). The architecture of SARS-CoV-2 transcriptome. *Cell* **181**, 914–921.e10.
72. Pujadas, E., Beaumont, M., Shah, H., et al. (2021). Molecular profiling of coronavirus disease 2019 (COVID-19) autopsies uncovers novel disease mechanisms. *Am. J. Pathol.* **191**, 2064–2071.
73. Wyler, E., Mösbauer, K., Franke, V., et al. (2021). Transcriptomic profiling of SARS-CoV-2 infected human cell lines identifies HSP90 as target for COVID-19 therapy. *iScience* **24**, 102151.
74. Severe Covid-19 GWAS Group, Ellinghaus, D., Degenhardt, F., Buti, M., et al. (2020). Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* **383**, 1522–1534.
75. Pairo-Castineira, E., Clohisey, S., Klaric, L., et al. (2021). Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92–98.
76. Dai, Y., Wang, J., Jeong, H.H., et al. (2021). Association of CXCR6 with COVID-19 severity: delineating the host genetic factors in transcriptomic regulation. *Hum. Genet.* **140**, 1313–1328.

77. Qi, H.X., Shen, Q.D., Zhao, H.Y., et al. (2021). Network-based analysis revealed significant interactions between risk genes of severe COVID-19 and host genes interacted with SARS-CoV-2 proteins. *Brief. Bioinform.* **23**, bbab372.
78. Zhang, S., Cooper-Knock, J., Weimer, A.K., et al. (2021). Common and rare variant analyses combined with single-cell multiomics reveal cell-type-specific molecular mechanisms of COVID-19 severity. Preprint at medRxiv. <https://doi.org/10.1101/2021.06.15.21258703>.
79. Ramaswamy, A., Brodsky, N.N., Sumida, T.S., et al. (2021). Immune dysregulation and autoreactivity correlate with disease severity in SARS-CoV-2-associated multisystem inflammatory syndrome in children. *Immunity* **54**, 1083–1095.e7.
80. Cheng, J., Zhang, J., Wu, Z., and Sun, X. (2021). Inferring microenvironmental regulation of gene expression from single-cell RNA sequencing data using scMLnet with an application to COVID-19. *Brief. Bioinform.* **22**, 988–1005.
81. Pal, L.R., Cheng, K., Nair, N.U., et al. (2021). Synthetic lethality-based prediction of anti-SARS-CoV-2 targets. Preprint at bioRxiv. <https://doi.org/10.1101/2021.09.14.460408>.
82. Lee, H., Park, J., Im, H.J., et al. (2021). Discovery of potential imaging and therapeutic targets for severe inflammation in COVID-19 patients. *Sci. Rep.* **11**, 14151.
83. Li, Y., Duche, A., Sayer, M.R., et al. (2021). SARS-CoV-2 early infection signature identified potential key infection mechanisms and drug targets. *BMC Genom.* **22**, 125.
84. Wang, Z., Guo, K., Gao, P., et al. (2021). Repurposable drugs for SARS-CoV-2 and influenza sepsis with scRNA-seq data targeting post-transcription modifications. *Precis. Clin. Med.* **4**, 215–230.
85. Zhou, Y., Xu, J., Hou, Y., et al. (2021). Network medicine links SARS-CoV-2/COVID-19 infection to brain microvascular injury and neuroinflammation in dementia-like cognitive impairment. *Alzheimer's Res. Ther.* **13**, 110.
86. Wilk, A.J., Lee, M.J., Wei, B., et al. (2021). Multi-omic profiling reveals widespread dysregulation of innate immunity and hematopoiesis in COVID-19. *J. Exp. Med.* **218**, e20210582.
87. Muñoz-Fontela, C., Dowling, W.E., Funnell, S.G.P., et al. (2020). Animal models for COVID-19. *Nature* **586**, 509–515.
88. Shou, S., Liu, M., Yang, Y., et al. (2021). Animal models for COVID-19: hamsters, mouse, ferret, mink, tree shrew, and non-human primates. *Front. Microbiol.* **12**, 626553.
89. Speranza, E., Williamson, B.N., Feldmann, F., et al. (2021). Single-cell RNA sequencing reveals SARS-CoV-2 infection dynamics in lungs of African green monkeys. *Sci. Transl. Med.* **13**, eabe8146.
90. Yang, L., Nilsson-Payant, B.E., Han, Y., et al. (2021). Cardiomyocytes recruit monocytes upon SARS-CoV-2 infection by secreting CCL2. *Stem Cell Rep.* **16**, 2274–2288.
91. Singh, D., Aladyeva, E., Das, S., et al. (2021). Myeloid cell interferon responses correlate with clearance of SARS-CoV-2. *Res. Sq.* <https://doi.org/10.21203/rs.3.rs-664507/v1>.
92. Lee, J.S., Koh, J.Y., Yi, K., et al. (2021). Single-cell transcriptome of bronchoalveolar lavage fluid reveals sequential change of macrophages during SARS-CoV-2 infection in ferrets. *Nat. Commun.* **12**, 4567.
93. Qin, Z., Liu, F., Blair, R., et al. (2021). Endothelial cell infection and dysfunction, immune activation in severe COVID-19. *Theranostics* **11**, 8076–8091.
94. Brunk, F., Moritz, A., Nelde, A., et al. (2021). SARS-CoV-2-reactive T-cell receptors isolated from convalescent COVID-19 patients confer potent T-cell effector function. *Eur. J. Immunol.* **51**, 2651–2664.
95. Meyer, M., Wang, Y., Edwards, D., et al. (2021). Attenuated activation of pulmonary immune cells in mRNA-1273-vaccinated hamsters after SARS-CoV-2 infection. *J. Clin. Invest.* **131**, e148036.
96. Ragan, I.K., Hartson, L.M., Dutt, T.S., et al. (2021). A whole virion vaccine for COVID-19 produced via a novel inactivation method and preliminary demonstration of efficacy in an animal challenge model. *Vaccines* **9**, 340.
97. Routhu, N.K., Cheedarla, N., Gangadhara, S., et al. (2021). A modified vaccinia Ankara vector-based vaccine protects macaques from SARS-CoV-2 infection, immune pathology, and dysfunction in the lungs. *Immunity* **54**, 542–556.e9.
98. Li, S., Wu, B., Ling, Y., et al. (2021). Epigenetic landscapes of single-cell chromatin accessibility and transcriptomic immune profiles of T cells in COVID-19 patients. *Front. Immunol.* **12**, 625881.

## ACKNOWLEDGMENTS

The authors thank all the members in the Bioinformatics and Systems Medicine Laboratory (BSML) for valuable discussion and help, especially Dr. Hyun-Hwan Jeong who helped create Table S4. We thank the three reviewers whose comments helped improve the manuscript. We also thank those investigators who generated the scRNA-seq data for COVID-19 studies. We are sorry for those related studies that we were not able to include in this review article due to our limited search function and space. The diagrams were created with Biorender.com. Z.Z. was partially supported by National Institutes of Health grants (R01LM012806, R01DE030122, and R01DE029818) and Cancer Prevention and Research Institute of Texas (CPRIT RP180734 and RP210045). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. Funding for open access charge: CPRIT (RP180734).

## AUTHOR CONTRIBUTIONS

Z.Z., W.L., and Y.D. conceived the project. W.L., J.J., and Q.Y. collected and analyzed the data. W.L., J.J., Y.D., G.P., W.C., and Z.Z. wrote the manuscript. Z.Z. supervised the project. All authors read and approved the final manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xinn.2022.100289>.

## LEAD CONTACT WEBSITE

<https://sbmi.uth.edu/faculty-and-staff/Zhongming-Zhao.htm>.