



## ARTICLE OPEN

# Adaptive immune responses to SARS-CoV-2 infection in severe versus mild individuals

Fan Zhang<sup>1</sup>, Rui Gan<sup>1</sup>, Ziqi Zhen<sup>1</sup>, Xiaoli Hu<sup>2</sup>, Xiang Li<sup>1</sup>, Fengxia Zhou<sup>1</sup>, Ying Liu<sup>3</sup>, Chuangeng Chen<sup>1</sup>, Shuangyu Xie<sup>1</sup>, Bailing Zhang<sup>1</sup>, Xiaoke Wu<sup>4,5</sup> and Zhiwei Huang<sup>1</sup>

The global Coronavirus disease 2019 (COVID-19) pandemic caused by SARS-CoV-2 has affected more than eight million people. There is an urgent need to investigate how the adaptive immunity is established in COVID-19 patients. In this study, we profiled adaptive immune cells of PBMCs from recovered COVID-19 patients with varying disease severity using single-cell RNA and TCR/BCR V(D)J sequencing. The sequencing data revealed SARS-CoV-2-specific shuffling of adaptive immune repertoires and COVID-19-induced remodeling of peripheral lymphocytes. Characterization of variations in the peripheral T and B cells from the COVID-19 patients revealed a positive correlation of humoral immune response and T-cell immune memory with disease severity. Sequencing and functional data revealed SARS-CoV-2-specific T-cell immune memory in the convalescent COVID-19 patients. Furthermore, we also identified novel antigens that are responsive in the convalescent patients. Altogether, our study reveals adaptive immune repertoires underlying pathogenesis and recovery in severe versus mild COVID-19 patients, providing valuable information for potential vaccine and therapeutic development against SARS-CoV-2 infection.

*Signal Transduction and Targeted Therapy* (2020)5:156

; <https://doi.org/10.1038/s41392-020-00263-y>

## INTRODUCTION

Coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has raised a global health emergency. Worldwide studies have contributed to the characterization, diagnosis, and treatment of the disease.<sup>1–4</sup> However, the pathogenesis of SARS-CoV-2 infection in humans remains unclear. Previous studies on severe acute respiratory syndrome (SARS),<sup>5</sup> Middle East respiratory syndrome,<sup>6</sup> and influenza<sup>7</sup> demonstrated that immune changes, especially those in peripheral blood lymphocyte subsets, play a critical role in defense against coronavirus infections. Consistently, several studies of COVID-19 patients showed that both humoral and cellular immunity are involved in the pathogenesis of COVID-19.<sup>8–10</sup> Although most COVID-19 patients presented mild-to-moderate symptoms, some infected individuals did develop severe or critical outcomes. However, the immunological features associated with the disease severity remains largely unknown. In addition, earlier studies on the recovery of SARS patients have shown that complete restoration of peripheral lymphocyte may require a longer period.<sup>11</sup> Thus, studies of the immune system of convalescent COVID-19 patients will facilitate understanding of their recovery state and establish the relationship between adaptive immune responses and disease severity if it exists.

Here, we collected peripheral blood CD3<sup>+</sup> T cells and CD3<sup>+</sup>CD19<sup>+</sup>CD20<sup>+</sup>CD27<sup>+</sup> antigen-experienced B cells (AEBCs) from five severe/critical and four mild-to-moderate convalescent COVID-19 patients. These cells were analyzed with single-cell TCR

sequencing (scTCR-seq), single-cell BCR sequencing (scBCR-seq), and single-cell RNA sequencing (scRNA-seq). By using these techniques, we found SARS-CoV-2-specific shuffling of adaptive immune repertoires and COVID-19-induced remodeling of peripheral lymphocytes. Characterization of variations in cell composition and functional status of the peripheral T and B cells of the recovered COVID-19 patients showed a more robust humoral immune response and T-cell immune memory in the severe patients (SPs). Data from scRNA-seq and scTCR-seq revealed concomitant responses of three major clusters of memory T cells in the adaptive immune system of the COVID-19 patients. Importantly, functional assays indicate that peptides derived from the M protein of SARS-CoV-2 are active in inducing T-cell response in most of the COVID-19 patients. Altogether, our data not only unveil the adaptive immunological features of convalescent COVID-19 patients but also provide information on the development vaccine and therapeutic agents against SARS-CoV-2 infection.

## RESULTS

### Study design

To investigate whether immunological memory is established in the SARS-CoV-2-infected individuals, peripheral blood mononuclear cells (PBMCs) were isolated from a cohort of nine convalescent patients of COVID-19 and three age-matched healthy controls (HCs). These cells were further sorted into CD19<sup>+</sup>CD20<sup>+</sup>CD27<sup>+</sup> AEBCs and CD3<sup>+</sup> T-cell populations via flow

<sup>1</sup>HIT Center for Life Sciences, School of Life Science and Technology, Harbin Institute of Technology, Harbin 150080, China; <sup>2</sup>Department of Infectious Diseases, Heilongjiang Provincial Hospital, Harbin Institute of Technology, Harbin 150030, China; <sup>3</sup>Harbin Blood Center, Harbin 150056, China; <sup>4</sup>Centre for Reproductive Medicine, Heilongjiang Provincial Hospital, Harbin Institute of Technology, Harbin 150030, China and <sup>5</sup>Department of Obstetrics and Gynecology, First Affiliated Hospital, Heilongjiang University of Chinese Medicine, Harbin 150040, China

Correspondence: Zhiwei Huang ([huangzhiwei@hit.edu.cn](mailto:huangzhiwei@hit.edu.cn))

These authors contributed equally: Fan Zhang, Rui Gan, Ziqi Zhen, Xiaoli Hu

Received: 24 June 2020 Revised: 17 July 2020 Accepted: 22 July 2020

Published online: 14 August 2020

cytometry. Listed in Supplementary Table 1 are the demographic and clinical characteristics of the nine COVID-19 donors including four mild-to-moderate cases (M1–M4), four severe cases (S1, S2, S3, and S5), and one critical case (S4). Diagnosis of SARS-CoV-2 was based on clinical symptoms, exposure history, and chest radiography, and SARS-CoV-2 infection was confirmed by using commercial quantitative PCR with reverse transcription assay on throat swab samples from the respiratory tract. The classification of disease severity for COVID-19 is based on the Chinese Clinical Guidance for COVID-19 Pneumonia Diagnosis and Treatment. The blood samples of these patients were collected about 4 weeks post discharge. Besides the three enrolled healthy donors (H1–H3), two more HCs were acquired from the Single Cell Immune Profiling Datasets of 10x genomics and labeled as H4 and H5. To understand how SARS-CoV-2 infection perturbs B-cell and T-cell immune repertoires, we applied scTCR-seq and scBCR-seq to sequence the V(D)J region of CD3+ T cells and AEBCs from 12 donors. After filtering, a total of 50,075 T cells (mean: 3070 cells for five SPs/critical patients, 3762 cells for four mild/moderate patients (MPs), 3936 for five HCs), and 29,899 AEBCs (mean: 2454 for SPs, 2652 for MPs, and 1404 for HCs) were obtained. Totally, 39,775 TCR and 21,339 BCR clones were acquired from these cells (Supplementary Table 2). Nonproductive (or nonfunctional) sequences were discarded in the following analysis. Furthermore, to further assess the cellular population composition and functional status of the peripheral T cells and AEBCs during the convalescing phase, we also conducted scRNA-seq on the T and AEBC cells from all the 12 enrolled donors. In total, 83,816 cells were recovered including 27,216 cells from SPs (mean: 5443 cells), 28,942 cells from MPs (mean: 7236 cells), and 27,658 cells from HCs (mean: 5532 cells) (Fig. 1a). By integrative analysis of single-cell transcriptome and immune profiling, we aim to reveal the adaptive immunological features related to the disease severity in the convalescent COVID-19 patients (Fig. 1a).

#### T- and B-cell immune repertoires in COVID-19 patients

A direct consequence of SARS-CoV-2 infection is TCR or BCR clonal expansion and expanded memory B-cell clones have been used to screen neutralizing antibodies against SARS-CoV-2.<sup>12</sup> We first investigated the clonal expansion by quantifying the ratio of expanded clones, clonal abundance distributions<sup>13</sup>, and clonal diversity<sup>14</sup> in the whole TCR or BCR repertoires. We observed comparable T-cell clonal expansion in the SPs and MPs (Fig. 1b), whereas overall increased TCR clonal abundance (Fig. 1c) and reduced diversity (Fig. 1d) were observed in the SPs. Notably, more expanded BCR clones (Fig. 1e), increased clonal abundance (Fig. 1f), and reduced clonal diversity (Fig. 1g) were present in the SPs compared with the MPs and the HCs. In contrast, no significant BCR clonal expansion was observed in most of the MPs (except for M2 (56 years), the oldest individual in the mild group (mean: 46 years)) (Fig. 1e). Taken together, these results show the persistence of robust T-cell responses in most of the convalescent COVID-19 patients, and stronger B-cell immune responses to SARS-CoV-2 infection in the SPs than in the MPs.

To further characterize BCR clonal expansion in these patients, we applied BASELINE<sup>15</sup> to the entire Ig sequences to estimate the selection strength in the CDR and FWR regions. The analysis revealed a clear positive selection in the CDR region of the COVID-19 patients but not in the HCs (Fig. 1h), suggesting that the patients' Ig repertoires were modified to produce high-affinity memory and plasma cells against SARS-CoV-2 infection. An increased frequency of replacement (non-synonymous) mutations occurs in the BCR CDR region during the process of antigen-driven BCR affinity maturation.<sup>16</sup> Indeed, we observed higher mutation frequencies of the top expanded clones in the SPs and MPs compared with the HCs (Fig. 1i). Consistently, higher mutation frequencies were also found in the more expanded clones than in the less expanded clones (Supplementary Fig. 1). IgG and IgA isotypes were overrepresented

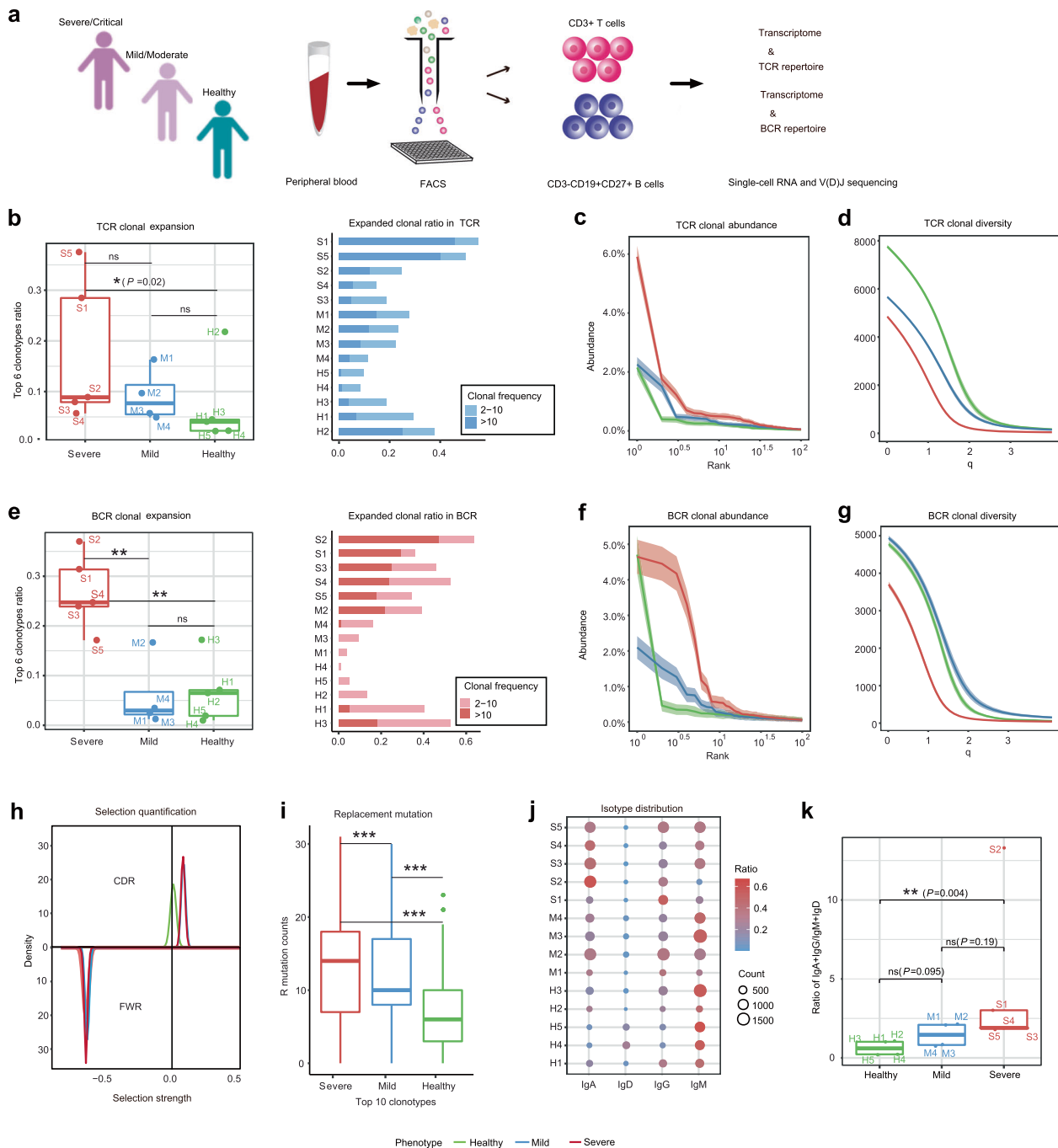
in the convalescent patients, especially in the severe cases in a comparison with the HCs (Fig. 1j). The ratio of (IgA + IgG) to (IgD + IgM) increased with disease severity in these COVID-19 patients (Fig. 1k), suggesting that SARS-CoV-2 induced a more intensive antibody response in the SPs than in the MPs.

Next, to reveal the unique gene patterns and preferences for BCR or TCR in the COVID-19 patients, we compared the usage of V (D)J genes among the participants. We first evaluated the overall immune repertoire differences among them. Principle component analysis (PCA) was used to explore the inter-individual variations based on the usage properties of IGHV genes of BCRs and TRAV or TRBV genes of TCRs. PCA plot showed that clonotype composition of the healthy and the two distinct severity groups of COVID-19 patients are apart from each other (Fig. 2a), suggesting shuffling of the V gene usage preference in the adaptive immune repertoires of the patients. Importantly, the IGHV-gene repertoire was remarkably biased in the SPs (Fig. 2b, c), with IGHV3 family genes including *IGHV3-23* (16.8%), *IGHV3-7* (10.2%), *IGHV3-48* (4.5%), *IGHV3-21* (3.3%), *IGHV3-11* (2.24%), and *IGHV3-15* (1.64%) accounting for ~40% of the whole expanded cohort. Similarly, enhanced usage of the IGHV3 family genes was also observed in human antibodies against other viruses such as cytomegalovirus (CMV),<sup>17</sup> influenza virus,<sup>18</sup> and Ebola virus.<sup>19</sup> Interestingly, *IGHV4-34* B-cell clones, rarely present in IgG memory B cells from healthy individuals,<sup>20</sup> were highly represented in one of the severe cases (S2) (Fig. 2c). Moreover, the top two pairing VJ segments *IGHV3-23-IGHJ4* and *IGHV5-51-IGHJ4* of BCR clones in the SPs appeared SARS-CoV-2 specific<sup>10</sup> (Fig. 2c). When comparing severe group with the mild group, *IGHV3-23*, *IGHV3-48*, *IGHV1-2*, and *IGHV4-34* were dominated in severe group (Supplementary Fig. 2). We similarly discerned gene usage preference of *TRAV14/DV4*, *TRAV20*, *TRAV23/DV6*, *TRAV25*, *TRBV13*, *TRBV14*, and *TRBV28* in the SPs and MPs (Fig. 2a–c), and some of them (*TRAV14/DV4*, *TRAV20*, and *TRBV13*) were also observed to expand in response to a range of viruses including influenza virus,<sup>21</sup> CMV,<sup>22</sup> and simian immunodeficiency virus (SIV).<sup>23</sup>

To identify convergent antibodies for COVID-19, we pooled the BCR data from the 14 individuals together and carried out clonal grouping using Change-O toolkit,<sup>24</sup> based on common genes of IGHV and IGHJ and nucleotide similarity of CDR3 sequences. Public antibody sequences present in more than a single donor were identified and extracted for multiple alignment analysis of their CDR3 regions (Supplementary Fig. 3). The data from the analysis revealed a repertoire of public clusters (0.786, 0.6, and 0.92% of total IgA, IgG, and IgM clusters) in the nine COVID-19 patients but not in the five HCs (0.156, 0, and 0% of total IgA, IgG, and IgM clusters) (Supplementary Fig. 4a), presumably due to the infection of SARS-CoV-2.<sup>25</sup> In total, we identified 19 convergent IgG and 25 IgA antibodies shared by the COVID-19 patients (Supplementary Table 3, the human antibodies sequences will be provided upon request), though their SARS-CoV-2 neutralizing activity warrants future investigations. During preparation of the manuscript, a recently published paper reported<sup>26</sup> convergent antibodies of *IGHV3-30/IGKV1-39* from two COVID-19 convalescent donors. Next, we used GLIPH<sup>27</sup> to analyze TCRβ sequences and grouped them according to the CDR3 sequence similarity. Likewise, we found more public TCR clusters in the SPs than in the MPs or HCs (1.8, 0.62, and 0.66% of total TCR clusters in SPs, MPs, and HCs, respectively) (Supplementary Fig. 4b). Taken together, these results support the notion that severe and mild COVID-19 patients experience distinct humoral and cell-mediated adaptive immune responses.

Characterization of variations in cell composition and functional status of the peripheral T and B cells in recovered COVID-19 patients

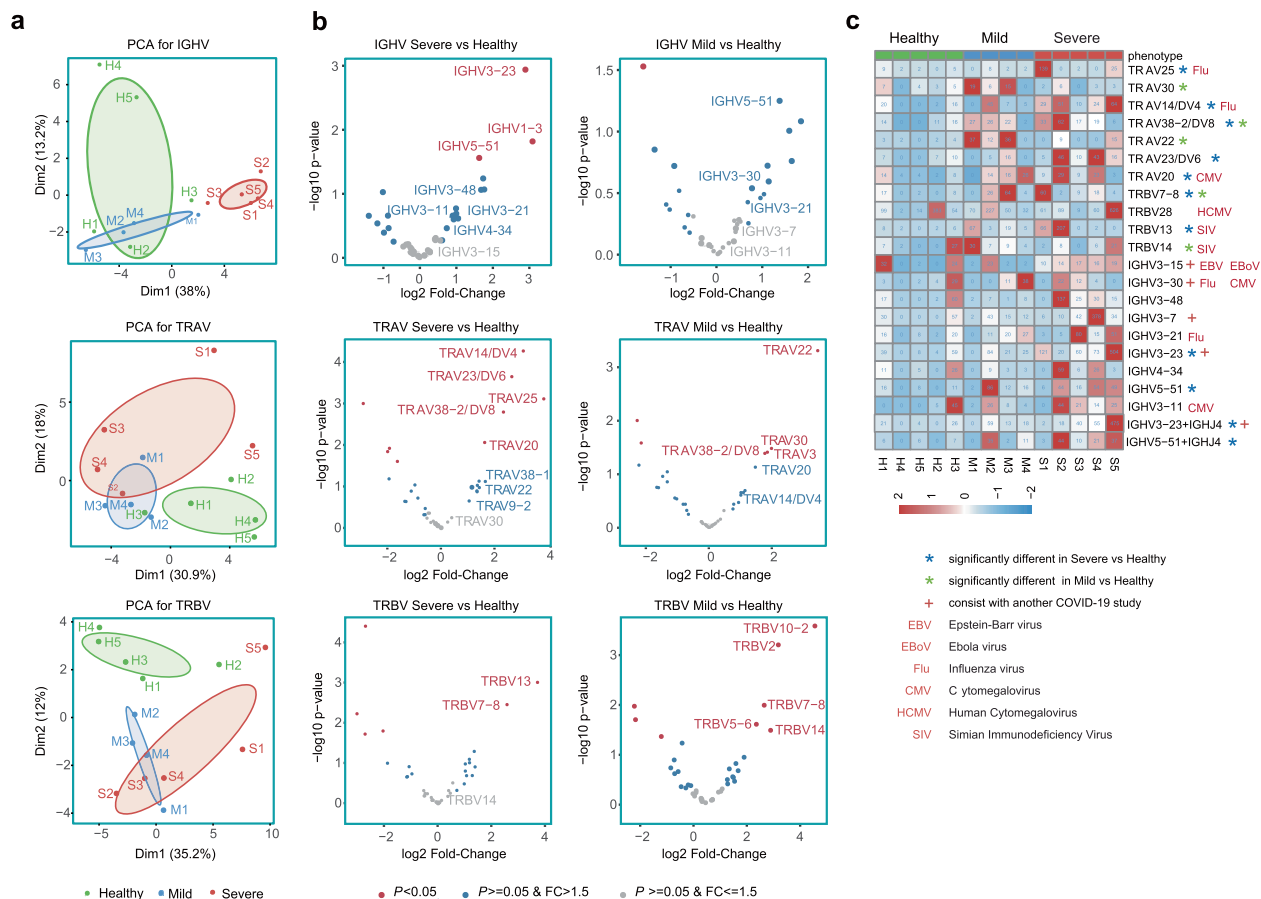
To characterize the adaptive immune system of the convalescent COVID-19 patients and understand their recovery state, we performed scRNA-seq analysis on CD3+ T cells and AEBCs from



**Fig. 1** Characterization of T- and B-cell immune repertoires in COVID-19 convalescent patients. **a** Schematic representation of overall study design. Blood samples from recovered patients with severe/critical infection or mild/moderate infection, and healthy controls were collected and sorted to obtain CD3+ T cells and CD19+CD20+CD27+ antigen-experienced B cells (AEBCs). Sorted cells underwent high-throughput single-cell RNA and TCR/BCR V(D)J sequencing for further analyses. Characterization and comparison of TCR clonal expansion among severe/critical patients (SPs, red), mild/moderate patients (MPs, blue), and healthy controls (HCs, green), by quantifying the ratio of expanded clones (**b**), clonal abundance (**c**), and clonal diversity (**d**). Characterization and comparison of BCR clonal expansion among SPs, MPs, and HCs, by quantifying the ratio of expanded clones (**e**), clonal abundance (**f**), and clonal diversity (**g**). **h** Comparison of selection strength on complementary-determining region (CDR, top) and framework region (FWR, bottom) of the BCRs among SPs, MPs, and HCs by density plot. **i** Comparison of replacement mutation frequency on IGHV gene segments among SPs, MPs, and HCs by boxplot. **j** Comparison of isotype distributions among SPs, MPs, and HCs. **k** Comparison of the ratio of (IgA + IgG) to (IgD + IgM) in SPs, MPs, and HCs by boxplot.  $***P < 0.001$ ;  $**P < 0.01$ ;  $*P < 0.05$ ; n.s. not significant using Wilcoxon rank sum test for pairwise comparisons

the SPs (S1–S5), MPs (M1–M4), and HCs (H1–H5) using Cell Ranger count pipeline. After quality control, a total of 83,817 cells were obtained for downstream analysis. Using a Louvain clustering algorithm<sup>28</sup> and automated reference-based annotation tools (Scibet<sup>29</sup> and Singler<sup>30</sup>) combined with expression of canonical

genes, we identified ten distinct clusters representing different T-cell subsets and two distinct clusters representing different B-cell subsets (Supplementary Table 4 and Supplementary Fig. 5). Then t-distributed stochastic neighbor embedding (t-SNE) was performed to visualize the cells in 2D space (Fig. 3a). Mucosal



**Fig. 2** Comparison of V gene usage of IGH, TRA, and TRB among SPs, MPs, and HCs. **a** Principal component analysis (PCA) for IGHV-gene usage and TRAV or TRBV-gene usage in SPs (S1–S5), MPs (M1–M4), and HCs (H1–H5). **b** Volcano plot representing differences of IGHV-gene usage in BCR repertoires or TRAV/TRBV-gene usage of TCR repertoires. Positive fold-change values denote more frequent IGHV genes or TRAV/TRBV genes in COVID-19 patients. Genes with  $P$  value  $< 0.05$  are displayed in red. Genes with a  $P$  value  $> 0.05$  but fold-change (FC) value  $> 1.5$  are displayed in blue. **c** Heatmap of comprehensive gene segments enriched in SPs or MPs compared with in HCs. Colors denote frequencies of each V gene segment used in each sample. The V genes overrepresented in SPs or MPs ( $P$  value  $< 0.05$ ) are indicated using a blue or a green star, respectively. Other known biasedly used gene segments related to virus-specific antibodies are marked using relevant virus name (EBV, EBoV, Flu, CMV, HCMV, and SIV). Genes consistent with another COVID-19 recovered patient study<sup>10</sup> are denoted with a cross

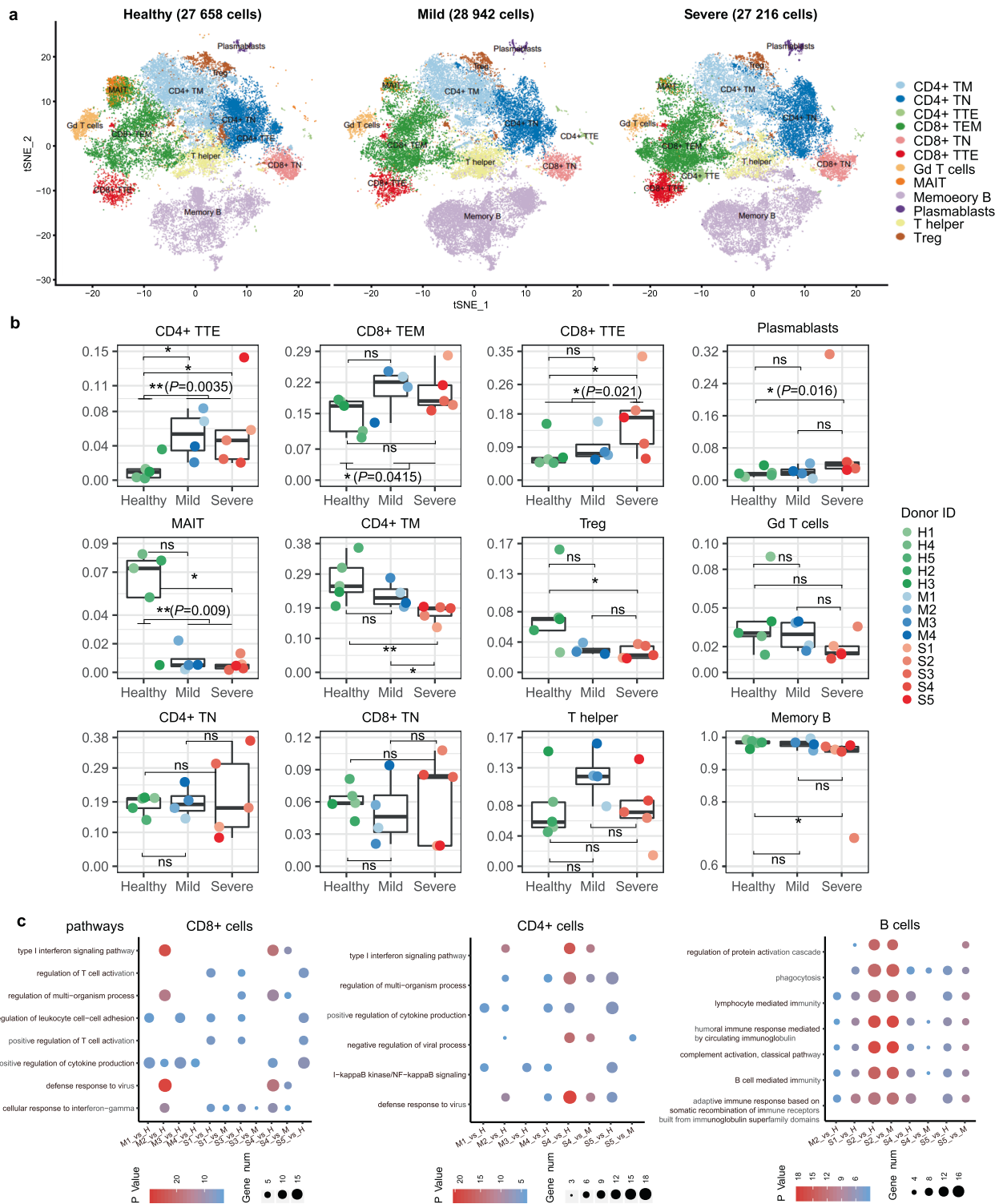
associated invariant T (MAIT) cells were characterized by the invariant alpha chain *TRAV1-2* in conjunction with *TRAJ33*, *TRAJ20*, or *TRAJ12*.  $V\gamma9/V\delta2^+$  T cells (the major Gd T-cell fraction in the peripheral blood) and cytotoxic T cells were identified in the T-cell subsets based on expression of the marker genes *TRDV2*, *TRGV9*, and *TRDC* and cytotoxic effector molecules of *GZMA*, *GZMB*, *PRF1*, and *NKG7*, respectively (Supplementary Fig. 5).

To reveal COVID-19-driving changes in the compositions of the T- or B-cell subsets, we quantified the proportions of all cell types in each individual from the SP, MP, and HC categories (Fig. 3b). Overall, we observed no significant difference in the CD4/CD8 naive, T helper, and Gd T-cell subsets among the SPs, MPs, and HCs. In contrast, both SPs and MPs exhibited elevated percentages of CD8<sup>+</sup> effector memory (TEM) cells (mean:  $\sim 20.1\%$ ) compared with HCs (mean:  $\sim 14.6\%$ ,  $P = 0.0415$ , Wilcoxon rank sum test), suggesting that the proliferation of these cells correlates with COVID-19. Notably, the subset of CD8<sup>+</sup> TEM cells re-expressing *CD45RA* (named CD8<sup>+</sup> terminal effector (TTE) cells) were significantly higher in the SPs (mean:  $\sim 17\%$ ) than those in the MPs and HCs (mean:  $\sim 8\%$ ,  $P = 0.021$ , Wilcoxon rank sum test). These CD8<sup>+</sup> TTE cells are mostly cytotoxic T cells expressing cytotoxic effector genes, such as *GZMA*, *GZMB*, *PRF1*, and *NKG7*. Indeed, we found that the COVID-19 patients possessed a small CD4<sup>+</sup> TTE cell population (mean:  $\sim 5.6\%$ ), which was barely detected in the healthy individuals (mean:  $\sim 1.3\%$ ,  $P = 0.0035$ ,

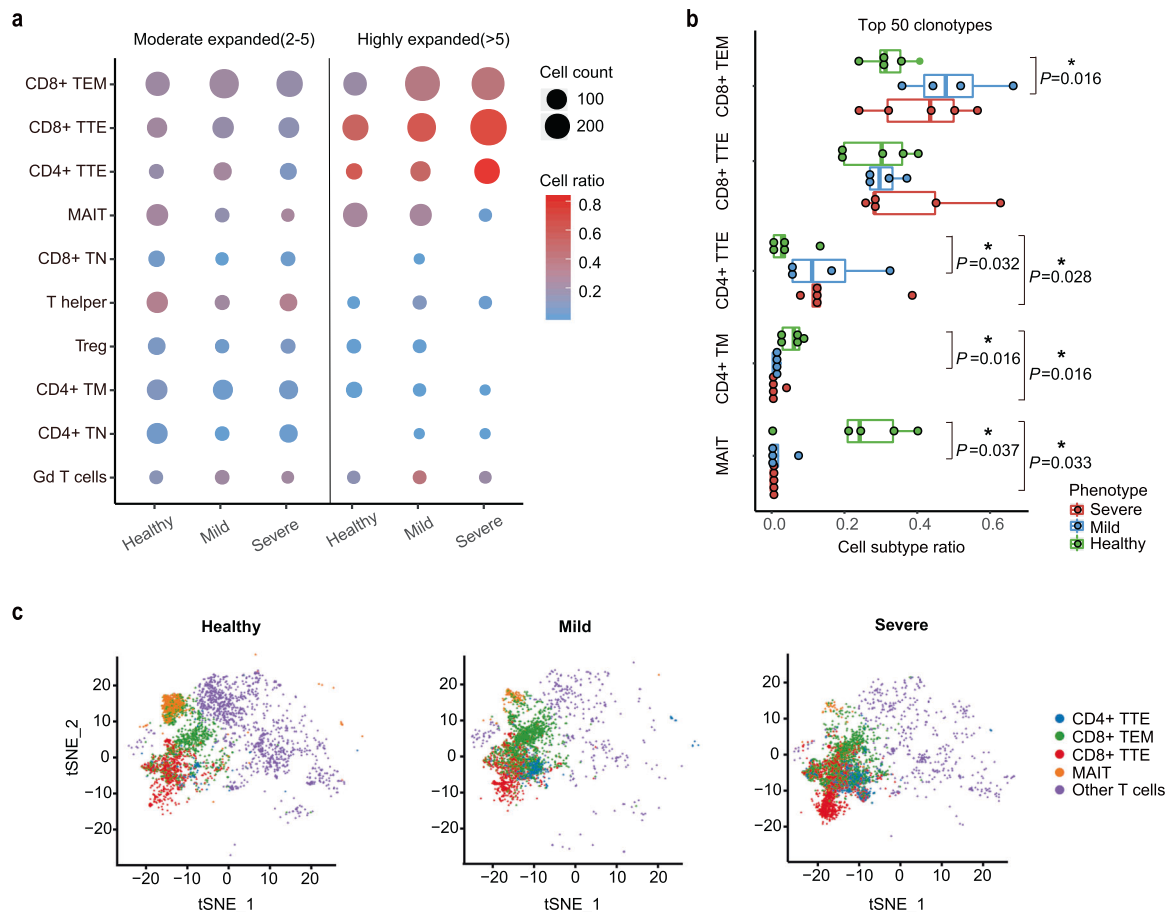
Wilcoxon rank sum test). Notably, the MAIT cells were substantially diminished in both SPs and MPs (patient versus health,  $P = 0.009$ , Wilcoxon rank sum test) compared with the HCs. Although the significance of this is still unclear, this result indicates that the immune system of COVID-19 patients has not been fully restored at early recovery stage, consistent with the study of recovery of SARS patients. Of note, depletion of MAIT cells was also observed in other viral diseases, such as HIV and HCV.<sup>31</sup> In contrast, SARS-CoV-2-induced depletion of  $V\gamma9/V\delta2^+$  T cells<sup>8</sup> was recovered in the convalescent patients (Fig. 3b). In addition, plasmablasts were enriched in the SPs ( $P = 0.016$ ), further supporting the conclusion that a more robust humoral immune response is stimulated in severe cases compared with the mild cases (Fig. 3b).<sup>4,32</sup>

To further characterize the functional status of T cells and AEBs in the convalescent SPs and MPs, we conducted differentially expressed gene (DEG) analysis, gene pathway enrichment analysis, and gene set enrichment analysis across the CD4<sup>+</sup>, CD8<sup>+</sup>, and memory B-cell subsets. Considering the highly heterogeneous immune responses among the COVID-19 patients,<sup>4</sup> we detected differentially expressed immune-related and inflammation-related pathways in every patient compared with HCs in CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, and AEBs, respectively, and further investigated the upregulated canonical pathways shared by multiple COVID-19 patients (Fig. 3c). As a result, we found that the signaling pathways of “cytokine production” and “leukocyte cell-cell adhesion” are the





**Fig. 3** Characterization of variations in cell composition and functional status of the peripheral T and B cells of the recovered COVID-19 patients. **a** Two-dimensional t-SNE visualization of 12 major cell types identified from healthy controls (HCs, left), mild patients (MPs, middle), and severe patients (SPs, right) including four for CD8+ T cells, five for CD4+ T cells, one for  $\gamma\delta$  T cells, and two for B cells. Each dot corresponds to one single cell, and colored according to cell types. The number of total cells showed in each t-SNE map was labeled on the top. **b** Boxplot representing proportions of each cell type in each sample. Each dot corresponds to one sample, and colored by group-specific color scheme (severe, red gradients; mild, blue gradients; healthy, green gradients). \*\* $P < 0.01$ ; \* $P < 0.05$ ; n.s. not significant by Wilcoxon rank sum test. **c** Dot plot depicting upregulated canonical pathways in CD4+ T cells, CD8+ T cells, and AEBs shared by multiple COVID-19 patients. Differentially expressed genes and their enriched pathways were calculated as each COVID-19 patient relative to all healthy controls (S# versus H or M# versus H), and each severe patient relative to all mild cases (S# versus M)

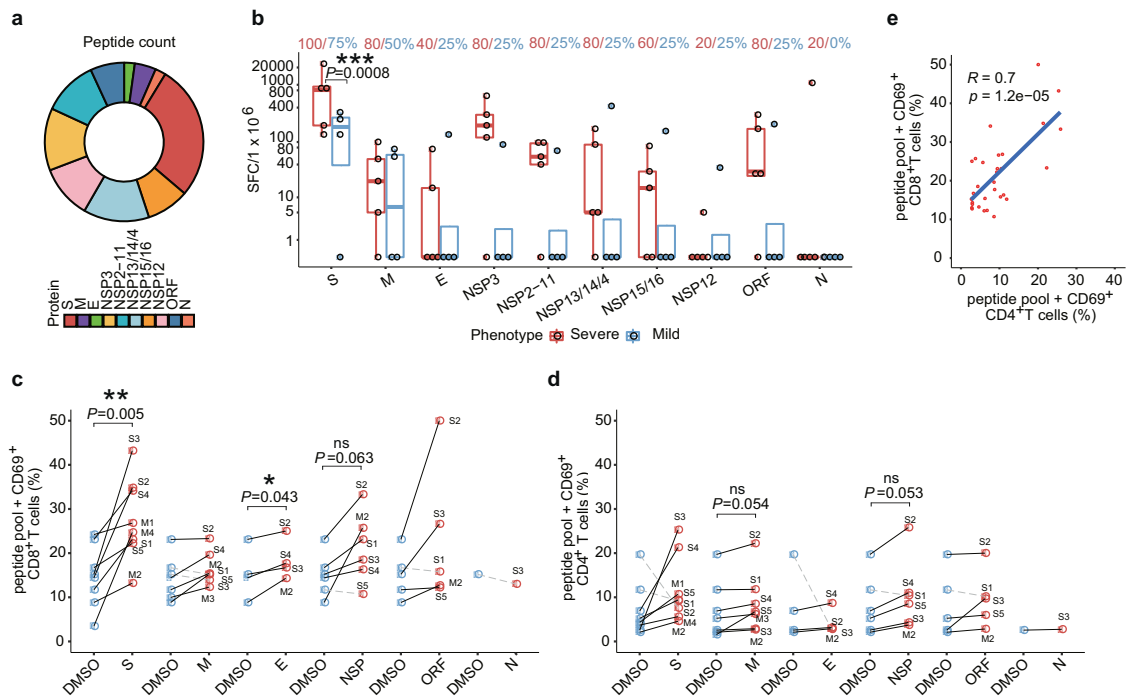


**Fig. 4** Comparison of TCR clonal expansion in T-cell subsets among severe patients (SPs), mild patients (MPs) and healthy controls (HCs). **a** Dot plot depicting clonal expansion status of distinct T-cell subsets in HCs, MPs, and SPs. Dot size denotes cell counts in each T-cell subset belonging to more expanded (frequency > 5) or less expanded ( $2 \leq \text{frequency} \leq 5$ ) clonotypes. Red-blue color range denotes the ratio of expanded cells occupied in each T-cell subset. **b** Boxplot depicting the top 50 TCR clonotypes in each donor consists of various T-cell subsets of different proportions. \* $P < 0.05$ ; n.s. not significant using Wilcoxon rank sum test for pairwise comparisons. **c** t-SNE projection of clonally expanded T cells (blue, CD4+ TTE; green, CD8+ TEM; red, CD8+ TTE; orange, MAIT; violet, other T cells) in HCs, MPs, and SPs

most commonly upregulated in either the CD4+ or the CD8+ T cell sets of the COVID-19 patients (Fig. 3c). The T-cell activation pathway was activated in the CD8+ T cells in three severe cases, whereas the pathways involving antiviral immunity, such as “type I interferon signaling” and “defense response to virus” were highly upregulated in the T cells of S4 and M2. Compared with the HCs and the MPs, the SPs clearly expressed more genes in the pathways of “B-cell-mediated immunity,” “complement activation,” “humoral immune response mediated by circulating immunoglobulin,” and “phagocytosis,” indicating a more extensive humoral immune response in the severe cases. This conclusion is further supported by analyses of BCR clonal expansion (Fig. 1e) and plasmablasts enrichment (Fig. 3b). The upregulation of the complement pathway in the SPs suggests a systemic pro-inflammatory response induced by SARS-CoV-2 infection as documented before.<sup>33</sup> Severe COVID-19 patients are normally associated with increased levels of interleukin (IL)-6, IL-10, and tumor necrosis factor- $\alpha$ .<sup>34,35</sup> However, substantial expression of pro-inflammatory cytokines such as IL-6, IL-1 $\beta$ , IL-2, IL-10, IL1A, IL-8, and CXCL-10 was not detected in most of the profiled T or B cells from all COVID-19 samples (Supplementary Fig. 6), suggesting that (IL)-6, IL-10, and probably other cytokines returned back to normal levels after recovery of these COVID-19 patients. Notably, the mitogen-activated protein kinase (MAPK) pathway (i.e., *FOS*, *JUN*, *JUNB*, and *DUSP1*) was greatly suppressed in all recovered patients compared with that in the HCs. This is in

full agreement with a previous study suggesting that inhibition of the MAPK signaling pathway is a recovery sign of COVID-19 patient (Supplementary Fig. 7).<sup>36</sup>

**Generation of specific T-cell subsets in response to SARS-CoV-2**  
To identify SARS-CoV-2-specific T-cell subsets in the convalescent COVID-19 patients, we combined scRNA-seq and scTCR-seq to assess clonal expansion of different T-cell subsets. Full-length TCRs with both alpha and beta chains were obtained from 15,818 T cells of the 10 clusters in COVID-19 patients, including 11,805 harboring unique TCRs and 4013 (25.4%) harboring repeated TCRs, indicative of clonal expansion of T cells in COVID-19 patients. Compared with other T-cell subsets, CD8+ TEM, CD8+ TTE, and CD4+ TTE were substantially clonally expanded in the COVID-19 patients, with more expanded clonotypes containing more CD8+ TEM and CD8+ TTE cells or 75.52% CD8+ TTE cells in the SPs, whereas a lower copy number of these clones existed in the HCs (Fig. 4b). These clonally expanded T cells may represent SARS-CoV-2-specific CD8+ or CD4+ T cells. Interestingly, the clonally expanded T cells showed an aggregative distribution on the t-SNE map (Fig. 4c), indicating transcriptional homogeneity in the COVID-19 patients and further supporting the idea that they are SARS-CoV-2 specific. Altogether, our data indicate concomitant response of CD8+ and CD4+ T cells in the adaptive immune



**Fig. 5** Assessment of SARS-CoV-2 specific T-cell immunity in COVID-19 convalescent patients by IFN- $\gamma$  ELISpot and FACS analysis. **a** Composition and fraction of peptides used in ten SARS-CoV-2 antigen pools. **b** Boxplot comparing the number of spot-forming cells (y axis) from ELISpot assay between stimulated severe and mild samples for each antigen group (x axis). Each donor was tested in duplicates. For each donor, the average number of spots from stimulated samples subtract that of the unstimulated samples was shown as a dot on the plot. Negative values were set to zero. Red and blue dots indicate SPs and MPs, respectively. \*\*\* $P < 0.001$  by unpaired Student's  $t$  test. SFC spot-forming cells. The blue or red text on top indicates the percentage of stimulated mild or severe donors shown positive in ELISpot, respectively. Antigen-specific CD69<sup>+</sup>CD8<sup>+</sup> T cell (**c**) and CD4<sup>+</sup> T cell (**d**) reactivity in ELISpot-positive COVID-19 cases between the negative control (DMSO, blue dots) and antigen pool stimulations (red dots). \*\* $P < 0.01$ ; \* $P < 0.05$  by paired  $t$ -test. **e** Pearson's correlation between SARS-CoV-2-specific CD4<sup>+</sup> T cells (x axis) and SARS-CoV-2-specific CD8<sup>+</sup> T cells (y axis)

system of the COVID-19 patients, and the clonally expanded CD8<sup>+</sup> effector cells in the patients may ultimately develop into long-lived memory T cells. These data are further supported by a recent study showing a robust adaptive immune response in mild COVID-19 patients.<sup>3</sup> Interestingly, MAIT clones were not expanded in the COVID-19 patients but were preferentially amplified in the healthy donors, though the significance of this remains unclear.

To experimentally validate the SARS-CoV-2 specific T-cell immunity in the COVID-19 convalescent patients, we synthesized 276 potential T-cell epitope peptides from all the 29 proteins of SARS-CoV-2,<sup>37–39</sup> with 210 corresponding to HLA class I CD8<sup>+</sup> T-cell epitopes and 66 to HLA class II CD4<sup>+</sup> T-cell epitopes (Fig. 5a). These peptides were divided into ten groups (as described in Supplementary Table 5) to stimulate PBMCs followed by interferon gamma (IFN- $\gamma$ ) ELISpot analysis. IFN- $\gamma$ -secreting T cells from each donor were detected and counted. A donor was considered positive to a peptide group when there was a more than twofold increase in the numbers of IFN- $\gamma$ -secreting T cells in the peptide-treated sample compared with the unstimulated control. At this cutoff point, we found that all recovered patients demonstrated specific memory T-cell responses against at least one group of the SARS-CoV-2 peptides (Fig. 5b). IFN- $\gamma$ -secreting T cells in response to the S and the M peptides were detected in all SPs and most MPs (75 and 50%). Interestingly, the S peptides stimulated a markedly higher percentage of IFN- $\gamma$ -secreting cells in both SPs and MPs than the other types of viral peptide. Each of the SPs demonstrated T-cell responses (mean: 12-fold increase) against at least five viral proteins, whereas most of the MPs (75%) only responded to the S protein (mean: 3.46-fold). Taken together, the ELISpot data indicate that the memory T cells specific for the S and other proteins were generated and sustained in the COVID-19 convalescent patients, which is further supported by a recent

study.<sup>40</sup> Importantly, we found that the SPs exhibited stronger T-cell immune responses than the MPs, suggesting that SPs may retain a more robust T-cell memory against SARS-CoV-2 than MPs.

To further determine the subpopulations of cytokine-producing memory T cells, totally 31 ELISpot-positive PBMCs from recovered donors were stained with antibodies against CD4/CD69 and CD8/CD69 for FACS analysis. SARS-CoV-2 Spike-specific CD8<sup>+</sup>/CD69<sup>+</sup> T cells were detected in all tested cases ( $P = 0.005$  versus DMSO control, paired  $t$ -test, Fig. 5c). CD8<sup>+</sup>/CD69<sup>+</sup> T cells were detected in 71, 75, 83, and 80% tested cases after stimulation with M, E, NSP, and ORFs (ORF3, ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10), respectively. These results show that the convalescent COVID-19 patients generated a substantial CD8<sup>+</sup> memory T-cell response against SARS-CoV-2 antigens. CD4<sup>+</sup> T-cell activation in response to S, M, E, NSP, and other ORFs was also detected in most tested cases ( $\geq 75\%$ ) (Fig. 5d), which is well correlated with CD8<sup>+</sup> T responses ( $R = 0.7$ ,  $P = 1.2 \times 10^{-5}$ , Pearson's correlation test, Fig. 5e).

## DISCUSSION

Differential adaptive immune responses between severe and mild COVID-19 patients remain largely unknown. This study aims to investigate the dynamics of antibody levels and T-cell responses against SARS-CoV-2 according to disease severity. In this respect, we observed multiple differences between the SPs and the MPs including TCR and BCR clonal expansion and diversity (Fig. 1), isotype distribution of antibody sequences (Fig. 1), V(D)J gene segments usage preference (Fig. 2), and dysregulation of peripheral blood lymphocyte subsets (Fig. 3). Higher levels of BCR clonal expansion (Fig. 1) and B-cell activation (Fig. 3) are

present in the SP group, indicating a more robust humoral immune response happened in severe infection. Taken together, these results denote that SPs and MPs may experience different cellular and humoral immune responses, which likely are related to different degrees of disease severity.

The upregulated immune pathways indicate that the immune systems of the convalescent patients still remain active, although they are no longer symptomatic. This agrees well with the notion that a longer period of time may require for complete restoration from the SARS-CoV-2 infection. A detailed analysis of peripheral immune cell compositions and functional status of peripheral lymphocyte is essential to comprehensively evaluate patient recovery stage.

Early studies have demonstrated an important role of type I and III IFN signaling pathways in SARS-CoV-2 or SARS-CoV infection,<sup>4,41,42</sup> and proposed prophylactic and therapeutic potential of these interferons in COVID-19 patients. Our data show that T-cell IFN response could only be detected in one critical infection case (S4) and one mild case (M2) (Fig. 3c and Supplementary Fig. 6). This may reflect the heterogeneous response of IFN induction in COVID-19 patients<sup>4</sup> and even their heterogeneous pathogenesis. However, activation of IFN-mediated pathways of T cells during SARS-CoV-2 infection and its relationship with clinical outcomes and antiviral immunity remains to be elucidated. Interestingly, M2 is the only individual in mild group that shares some common features with severe COVID-19 patients in single-cell immune profiling and transcriptome, such as highly clonal expansion (Fig. 1b, e) and upregulated pathways (Fig. 3c). M2 is also the only enrolled COVID-19 patient with hypertension and diabetes. This may suggest that some chronic basic diseases may associate with COVID-19<sup>43</sup> via affecting adaptive immune responses to SARS-CoV-2. The heterogeneity in immune responses of COVID-19 patients was also observed in the youngest (44 years) patient S2. The IGHV4-34 B-cell clones, rarely present in IgG memory B cells from healthy individuals,<sup>20</sup> were highly expressed in this patients. Interestingly, a TCR clone (frequency = 38) from the CD8<sup>+</sup> effector T cells of S2 was predicted to target an antigen peptide (LLLGIGILV) from an interferon-stimulated gene and human bone marrow stromal cell antigen 2 (BST2) (Supplementary Fig. 9). This suggests the possibility that autoimmunity induced by SARS-CoV-2 infection may occur to S2, thus contributing to the severe symptom of the patient.

To meet the urgent need for vaccine development, it is imperative to understand the magnitude and composition of human CD4<sup>+</sup> and CD8<sup>+</sup> T-cell responses to SARS-CoV-2. Our data of scRNA-seq and scTCR-seq revealed generation of specific T-cell subsets induced by the infectious virus. In strong support of this conclusion, many peptides derived from SARS-CoV-2 induced specific memory T-cell responses in the recovered patients. Interestingly, in addition to peptides from the S protein, which acts as the major target in most vaccine development efforts,<sup>40</sup> peptides derived from the M protein are also reactive in T-cell response in most of the COVID-19 patients. Our data showed that the reactive major viral proteins are potential targets for future vaccine development. Our data further revealed concomitant response of three major clusters of memory T cells (CD8<sup>+</sup> TEM, CD8<sup>+</sup> TTE, and CD4<sup>+</sup> TTE) in the adaptive immune system of the COVID-19 patients. This appears to be independent of the severity of the patients and is consistent with the data that robust adaptive immune responses are activated in mild COVID-19 patients.<sup>3</sup>

## MATERIALS AND METHODS

### Ethics statement

This study was conducted according to the principles expressed in the Declaration of Helsinki. Ethical approval was obtained from the Research Ethics Committee of Heilongjiang Provincial Hospital

(2020-112). All participants provided written informed consent for sample collection and subsequent analyses.

### Patients and clinical sample collection

All nine patients with COVID-19 in this study were enrolled from the Heilongjiang provincial Hospital from March to April, 2020. The clinical classifications was defined as mild to moderate, severe, and critical, according to the "Diagnosis and Treatment Protocol of COVID-19 (the 7th Tentative Version)" by the National Health Commission of China issued on 3 March 2020 (<http://www.nhc.gov.cn/zyygj/s7653p/202003/46c9294a7dfe4cef80dc7f5912eb1989.shtml>).

Briefly, moderate cases have fever, respiratory symptoms, and pneumonia evidenced by computed tomography (CT) imaging. This study enrolled two patients with moderate infection with bilateral pneumonia, labeled M1 and M2, two patients with mild infection, labeled M3 and M4. Patients with severe pneumonia were diagnosed on the basis of one of the following criteria: (1) respiratory distress with respiratory rate  $\geq 30$  times  $\text{min}^{-1}$ ; (2) fingertip oxygen saturation  $\leq 93\%$  at room air; (3) ratio of partial pressure of arterial oxygen to fraction of inspired oxygen ( $\text{PaO}_2/\text{FiO}_2$ )  $\leq 300$  mmHg (1 mmHg = 0.133 kPa); and (4) obvious progression of lesions in 24–48 h shown by pulmonary CT imaging  $> 50\%$ . Two patients with severe pneumonia were enrolled, labeled as S1, S2, S3, and S5. These four patients showed  $\text{PaO}_2/\text{FiO}_2 \leq 300$  mmHg. The patients with critical pneumonia (S4) were diagnosed on the basis of one of the following criteria: (1) respiratory failure and an artificial airway required for invasive mechanical ventilation; (2) shock; and (3) combined failure of other organs that required intensive care unit monitoring. S4 is the patient with critical pneumonia received invasive mechanical ventilation.

### Isolation of PBMCs

PBMCs were isolated from anticoagulant blood using Ficoll under the biosafety level 2 facility. To isolate PBMCs, blood was diluted with the same volume of PBS, then gently put on the equal volume of Ficoll in a 15 ml centrifuge tube and centrifuged for 35 min at  $700 \times g$  without brake. Discard the first layer that contains serum and platelet, carefully transfer the buffy coat to a new centrifuge tube, wash with 10 ml PBS, and centrifuge at  $350 \times g$ , room temperature, 5 min. The supernatant was discarded and the pellet was washed again. After that, recollect these PBMCs, count the cell number, and store in the liquid nitrogen.

### Isolation of T and B cells

Approximately 20 ml of human peripheral blood was obtained and placed on ice, and returned to room temperature before use. PBMCs isolation was processed within 2 h in a BSL-2 laboratory. PBMCs were isolated from whole blood samples by Ficoll density gradient centrifugation. Then, cells were counted in 0.4% Trypan blue,  $1 \times 10^7$  PBMCs was prepared for flow cytometry and cell sorting, and residual cells were centrifuged and resuspended at a concentration of  $2 \times 10^6/\text{ml}$  for further use.

### Flow cytometry and cell sorting

A FACS Aria cell sorter (BD Biosciences) was used to isolate T cells (CD3<sup>+</sup>), and CD3<sup>+</sup>CD19<sup>+</sup>CD20<sup>+</sup>CD27<sup>+</sup> AEBCs (comprised of mostly memory B cells with a small number of peripheral plasmablasts). Surface marker staining and cell sorting was performed by pelleting and resuspending cells in FACS buffer (2% FBS in PBS) with antibodies at indicated concentrations for 20 min at room temperature in the dark. Cells were washed once in the FACS buffer before resuspension. PBMCs were filtered with a 100- $\mu\text{m}$  nylon cell strainer to remove clumps and debris before flow cytometry.



### IFN- $\gamma$ ELISpot

IFN- $\gamma$ -secreting T cells were detected by Human IFN- $\gamma$  ELISpot pro 3 kits (MABTECH AB, Sweden) according to the manufacturer's protocol. Briefly, frozen PBMCs in cryotubes were placed in a 37 °C water bath till completely thawed, then transferred to a new 15-ml centrifuge tubes containing 10 ml complete RPMI medium (90% RPMI-1640, 10% heat-inactivated FBS, 1% penicillin–streptomycin), and centrifuged at 1000 rpm, room temperature for 5 min. The cell pellet was resuspended in 2–4 ml complete RPMI and balanced at 37 °C with 5% CO<sub>2</sub> for 6 h. Cell concentrations were adjusted to  $\sim 2 \times 10^6$ /ml and plated at 100 K per well in duplicate and incubated for 36 h with 10  $\mu$ g/ml of peptide pools (the concentration of each peptide is 10  $\mu$ g/ml). Spots were then counted using an Elispot Reader System (AT-Spot2100, atyx). The number of spots was converted into the number of spots-forming cells (SFCs) per million cells and the mean of duplicate wells was plotted.

We collected 276 peptides identified as potential T-cell epitopes for SARS-CoV-2 by at least two previous studies.<sup>37–39</sup> This pool of 9–15-mer peptides was derived from 29 proteins spanning the whole proteome of SARS-CoV-2, including 76 epitopes from spike (S), 6 from nucleocapsid (N), 12 from membrane (M), 6 from envelope (E), 157 from Orf1ab (Nsp1–Nsp16), and 20 from 7 other ORFs (ORF3, ORF3a, ORF6, ORF7a, ORF7b, ORF8, and ORF10). We divided these 276 peptides into ten groups, with peptides from the same protein in the same group. Details of the groups are shown in Supplementary Table 5.

### Cells activation and surface marker staining

Cells were thawed in complete RPMI 1640 culture media and then left overnight at 37 °C, 5% CO<sub>2</sub>. 2e5 PBMCs were plated in a 96-well plate and stimulated with peptides that shown positive in ELISPOT. After 24 h culture, PBMCs from each well were collected and washed with FACS buffer and surface-stained for CD4 (FITC, clone OKT4, Biolegend), CD8 (Brilliant Violet 510™, clone SK1, Biolegend) and CD69 (PE/Cyanine7, clone FN50, Biolegend) according to the manufacturer's instructions. All data were collected using a flow cytometer (LSRFortessa; BD Biosciences) and analyzed with FlowJo software (Tree Star, Ashland, OR, USA).

### Publicly available HC data

The V(D)J repertoires of T and B cells integrated with 5' gene expression data from PBMCs of two additional HCs were acquired from the Single Cell Immune Profiling Datasets of 10x genomics, with one donor generated using a 5' V2 chemistry kit (labeled as HC2), and the other using a 5' V3 chemistry kit (labeled as HC3) both on Chromium Single Cell Controller (10x genomics). Data could be accessed via <https://support.10xgenomics.com/single-cell-vdj/datasets/3.0.0/> for H4 and via <https://support.10xgenomics.com/single-cell-vdj/datasets/2.2.0/> for H5.

### scRNA-seq and preprocessing

**Single-cell 5' RNA sequencing.** The library of single-cell 5' RNA was established following the protocol from 10x genomics Chromium Single Cell Immune Profiling Solution. In brief, the RNA of the captured cells was released by lysis, barcoded via reverse transcription process, and amplified until acquiring sufficient cDNA for constructing 5' gene expression libraries. The Illumina Novaseq 6000 with PE150 was applied to sequence the 5' transcripts.

**Sample demultiplexing.** FASTQ files within read2 files were demultiplexed, and the cell barcodes and single-cell 5' unique molecular identifiers within read1 files were achieved by Cell Ranger Single-Cell Software Suite (v3.1.0). Each sample was analyzed using Cell Ranger count pipeline with default parameters and GRCh38 as the reference human genome.

### Single-cell transcriptome data analysis

**Cell aggregation and quality control.** Seurat (v3.1.4)<sup>44</sup> was applied to aggregate all samples, and perform quality control on raw gene expression matrix obtained from Cell Ranger pipeline. The filtering criteria for each cell were: (1) the number of genes expressed ranges from 200 to 4000, and (2) the percentage of the mitochondrial genes is less than 25%. Cells that did not satisfy the above criteria were discarded. Gene expressions were normalized by log normalization with factor 10,000, and the top 2000 highly variable genes were selected for PCA.

**Batch-effect correction.** Though many genes show significant differences in expression, these differences may come from experimental batch effect. Therefore, Harmony<sup>45</sup> was utilized for batch-effect correction based on the top 50 principal components of PCA.

**Cluster identification.** After removing the batch effect, dimension reduction was performed using t-SNE algorithm, and cell clusters were identified at resolution 3.0 by FindClusters function of Seurat (v3.1.4) utilizing a shared nearest neighbor graph constructed by FindNeighbors function. Finally, 59 clusters were obtained.

**Cell type annotation.** The corresponding cell types of the clusters were annotated using SingleR<sup>30</sup> and Scibet<sup>29</sup> with the reference datasets of human immune cells. The annotation was corrected manually in accordance with known gene markers.

### Differential gene expression analysis

Differential gene expression analysis was performed using Seurat v.3 (FindMarkers function) with default parameters. We identified differentially expressed genes of each COVID-19 sample by comparing CD4+, CD8+, or B cells of each COVID-19 sample to those cells of all HCs. Genes with adjusted  $P < 0.05$  and fold change  $> 1.5$  were considered as significantly upregulated in patients and were used for further GO enrichment analysis.

### GO and pathway enrichment analysis

GO and pathway enrichment analysis were performed using clusterProfiler<sup>46</sup> ( $P$ valueCutoff = 0.01 and  $q$ valueCutoff = 0.05). For each comparison, the top ten enriched pathways containing at least five upregulated genes were kept. Significantly upregulated pathways shared by at least two patients were selected to display in bubble charts.

### Single-cell BCR/TCR V(D)J sequencing and analysis

Chromium single cell V(D)J enriched libraries were quantified, normalized, and sequenced according to the User Guide for Chromium Single Cell V(D)J Reagent kits. The V(D)J enriched libraries were pooled and sequenced with an Illumina NovaSeq 6000 S2 Reagent Kit (300 cycles) (Illumina). BCR/TCR sequences for each single B/T cell were assembled by Cell Ranger vdj pipeline (v3.1.0), with CDR3 sequences and rearranged full-length BCR/TCR V(D)J segments as well as clonotype frequency obtained. Only productive BCR/TCR sequences were kept for subsequent analysis. The assembled FASTA sequences of IGH chains were collected and assigned germline sequences using Change-O<sup>24</sup> to perform clonal grouping and mutation analysis. V(D)J segment usage counts were calculated for all cohorts and PCA analysis was performed to compare the V(D)J segment usage among SPs, MPs, and HCs.

### BCR/TCR clonal grouping

All BCR/TCR clonotypes were identified by Cell Ranger vdj pipeline (v3.1.0). The BCR/TCR clonotypes from SPs, MPs, and HCs were pooled together before carrying out the clonal grouping step. TCR clusters were predicted by GLIPH<sup>27</sup> to bind the same MHC-restricted peptide antigen. Public TCR clusters were detected for patients and HCs, respectively. BCR clusters were identified using

Change-O based on hierarchical clustering approach. First, IGH sequences with the same V and J segment calls and CDR3s of the same length were grouped. Second, IGH sequences within each group were clustered into BCR clusters based on the hamming distances of CDR3 nucleotide sequences. Public BCR clusters of each isotype (IgM, IgA, and IgG) were detected for patients and HCs, separately.

#### Clonal abundance and diversity

Clonal abundance distributions were inferred using resampling strategies to correct for variations in sequencing depth.<sup>13</sup> Clonal diversity distributions were calculated by plotting diversity index<sup>14</sup> over a range of diversity orders ( $q$ ) to generate a smooth curve. All the corresponding analyses were performed using R package alakazam.<sup>47</sup>

Clonal abundance was defined by the number of BCR or TCR sequences in each cluster that gathered similar clones according to a sequence-based distance measure defined by nucleotide similarity of the junction regions.<sup>24</sup> The complete clonal abundance distribution were inferred using a novel estimator of the complete clones-rank abundance distribution by separately adjusting the sample relative abundances for the set of clones detected in different donor groups and estimating the relative abundances for the set of clones undetected in the group but inferred to be present in the assemblage. The colored bands indicate the middle 95% percentiles of the clone distribution with 200 bootstrap estimates in each donor group.

The Hill diversity index ( $^qD$ ), which measures diversity in a population, was calculated for the set of BCR IGH or TCR clones to determine whether the diversity in the repertoires differed between the donor groups.<sup>47</sup> For each donor group, the repertoire was subsampled to the number of sequences in the smallest sample and the Hill diversity index was calculated independently in 200 equally spaced  $q$  values between 0 and 4. Because  $q$  varies from 0 to infinity, the diversity ( $^qD$ ) depends less on rare species and more on common ones as  $q$  increases, thus encompassing a range of definitions that can be visualized as a single curve. For  $q = 0$ , the diversity is defined as the total number of clones. As  $q$  approaches infinity, the diversity is given by one over the frequency of the largest clone. At a given value of  $q$  ( $x$  axis), lower values of  $^qD$  ( $y$  axis) indicate lower diversity. The colored bands indicate the middle 95% percentiles of the sampled distribution; thus, when three lines are separated and fall outside of the colored band, the difference between the donor groups is significant.

#### Mutation and selection analysis

Replacement (R) mutations within subregions of the V segment were identified by comparing the input BCR sequences with the germline sequences using Change-O. In addition, the quantification of selection pressure was performed for SPs, MPs, and HCs using BASELINE<sup>15</sup> based on an underlying SHM targeting model. The probability density function (PDF) for selection strength in CDR and FWR regions was calculated using a Bayesian estimation model.

#### Statistics

DESEQ2<sup>48</sup> was used to find differential segment usages in BCR and TCR. The  $P$  value and fold change were plotted as volcano map using ggplot2. Wilcoxon signed ranks test and  $t$ -test was used pairwise comparisons.

#### DATA AVAILABILITY

The datasets used for the current study are available from the corresponding author upon reasonable request.

#### ACKNOWLEDGEMENTS

This research was funded by the National Natural Science Foundation of China grant no. 31825008 and 31422014 to Z.H. and grant no. 61872117 to F.Z.

#### AUTHOR CONTRIBUTIONS

F.Z., Z.Z., X.L., S.X., B.Z., Y.L., and X.W. collected clinical samples from COVID-19-infected patients and healthy donors. F.Z., R.G., and F.Z. performed bioinformatics analysis. Z.Z. and X.L. performed the experiments. F.Z. and Z.H. wrote the manuscript. Z.H. conceived and supervised the project. All authors contributed to the discussions.

#### ADDITIONAL INFORMATION

The online version of this article (<https://doi.org/10.1038/s41392-020-00263-y>) contains supplementary material, which is available to authorized users.

**Competing interests:** The authors declare no competing interests.

#### REFERENCES

- Long, Q. X. et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat. Med.* <https://doi.org/10.1038/s41591-020-0897-1> (2020).
- Yu, F., Du, L., Ojcius, D. M., Pan, C. & Jiang, S. Measures for diagnosing and treating infections by a novel coronavirus responsible for a pneumonia outbreak originating in Wuhan, China. *Microbes Infect.* **22**, 74–79 (2020).
- Liao, M. et al. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* <https://doi.org/10.1038/s41591-020-0901-9> (2020).
- Wilk, A. J. et al. A single-cell atlas of the peripheral immune response to severe COVID-19. <https://doi.org/10.1101/2020.04.17.20069930> (2020).
- Li, T. et al. Significant changes of peripheral T lymphocyte subsets in patients with severe acute respiratory syndrome. *J. Infect. Dis.* **189**, 648–651 (2004).
- Ko, J. H. et al. Predictive factors for pneumonia development and progression to respiratory failure in MERS-CoV infected patients. *J. Infect.* **73**, 468–475 (2016).
- Mohn, K. G. et al. Immune responses in acute and convalescent patients with mild, moderate and severe disease during the 2009 influenza pandemic in Norway. *PLoS ONE* **10**, e0143281 (2015).
- Wolfe, R. et al. Virological assessment of hospitalized patients with COVID-2019. *Nature* <https://doi.org/10.1038/s41586-020-2196-x> (2020).
- Ni, L. et al. Detection of SARS-CoV-2-specific humoral and cellular immunity in COVID-19 convalescent individuals. *Immunity*. <https://doi.org/10.1016/j.immuni.2020.04.023> (2020).
- Wen, W. et al. Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. *Cell Discov.* **6**, 31 (2020).
- Li, T. et al. Long-term persistence of robust antibody and cytotoxic T cell responses in recovered patients infected with SARS coronavirus. *PLoS ONE* **1**, e24 (2006).
- Cao, Y. et al. Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells. *Cell* **182**, 73–84. e16 (2020).
- Chao, A., Hsieh, T. C., Chazdon, R. L., Colwell, R. K. & Gotelli, N. J. Unveiling the species-rank abundance distribution by generalizing the good-turing sample coverage theory. *Ecology* **96**, 1189–1201 (2015).
- Hill, M. O. Diversity and evenness: a unifying notation and its consequences. *Ecology* **54**, 2 (1973).
- Yaari, G., Uduman, M. & Kleinstein, S. H. Quantifying selection in high-throughput immunoglobulin sequencing data sets. *Nucleic Acids Res.* **40**, e134 (2012).
- Hershberg, U., Uduman, M., Shlomchik, M. J. & Kleinstein, S. H. Improved methods for detecting selection by mutation analysis of Ig V region sequences. *Int. Immunol.* **20**, 683–694 (2008).
- Bryson, S. et al. Structures of preferred human IgV genes-based protective antibodies identify how conserved residues contact diverse antigens and assign source of specificity to CDR3 loop variation. *J. Immunol.* **196**, 4723–4730 (2016).
- Fu, Y. et al. A broadly neutralizing anti-influenza antibody reveals ongoing capacity of haemagglutinin-specific memory B cells to evolve. *Nat. Commun.* **7**, 12780 (2016).
- Ehrhardt, S. A. et al. Polyclonal and convergent antibody response to Ebola virus vaccine rVSV-ZEBOV. *Nat. Med.* **25**, 1589–1600 (2019).
- Tsay, G. J. & Zouali, M. The interplay between innate-like B cells and other cell types in autoimmunity. *Front. Immunol.* **9**, 1064 (2018).
- Pizzolla, A. et al. Influenza-specific lung-resident memory T cells are proliferative and polyfunctional and maintain diverse TCR profiles. *J. Clin. Invest.* **128**, 721–733 (2018).

22. Attaf, M. et al. Major TCR repertoire perturbation by immunodominant HLA-B (\*) 44:03-restricted CMV-specific T cells. *Front. Immunol.* **9**, 2593 (2018).
23. Price, D. A. et al. T cell receptor recognition motifs govern immune escape patterns in acute SIV infection. *Immunity* **21**, 793–803 (2004).
24. Gupta, N. T. et al. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* **31**, 3356–3358 (2015).
25. Galson, J. D. et al. In-depth assessment of within-individual and inter-individual variation in the B cell receptor repertoire. *Front. Immunol.* **6**, 531 (2015).
26. Robbiani, D. F. et al. Convergent antibody responses to SARS-CoV-2 infection in convalescent individuals. <https://doi.org/10.1101/2020.05.13.092619> (2020).
27. Glanville, J. et al. Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**, 94–98. <https://doi.org/10.1038/nature22976> (2017).
28. Waltman, L. & van Eck, N. J. A smart local moving algorithm for large-scale modularity-based community detection. *Eur. Phys. J. B* **86**, 471 (2013).
29. Li, C. et al. SciBet as a portable and fast single cell type identifier. *Nat. Commun.* **11**, 1818 (2020).
30. Aran, D. et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
31. Ussher, J. E., Willberg, C. B. & Klenerman, P. MAIT cells and viruses. *Immunol. Cell Biol.* **96**, 630–641 (2018).
32. Zhao, J. et al. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. *Clin. Infect. Dis.* <https://doi.org/10.1093/cid/ciaa344> (2020).
33. Gralinski, L. E. et al. Complement activation contributes to severe acute respiratory syndrome coronavirus pathogenesis. *mBio* **9**. <https://doi.org/10.1128/mBio.01753-18> (2018).
34. Chen, G. et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. *J. Clin. Investig.* **130**, 2620–2629 (2020).
35. Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
36. Huang, L. et al. Blood single cell immune profiling reveals the interferon-MAPK pathway mediated adaptive immune response for COVID-19. <https://doi.org/10.1101/2020.03.15.20033472> (2020).
37. Prachar, M. et al. COVID-19 vaccine candidates: prediction and validation of 174 SARS-CoV-2 epitopes. <https://doi.org/10.1101/2020.03.20.000794> (2020).
38. Campbell, K. M., Steiner, G., Wells, D. K., Ribas, A. & Kalbasi, A. Prediction of SARS-CoV-2 epitopes across 9360 HLA class I alleles. <https://doi.org/10.1101/2020.03.30.016931> (2020).
39. Fast, E. & Chen, B. Potential T-cell and B-cell epitopes of 2019-nCoV. <https://doi.org/10.1101/2020.02.19.955484> (2020).
40. Grifoni, A. et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell*. <https://doi.org/10.1016/j.cell.2020.05.015> (2020).
41. Hadjadj, J. et al. Impaired type I interferon activity and exacerbated inflammatory responses in severe Covid-19 patients. <https://doi.org/10.1101/2020.04.19.20068015> (2020).
42. Channappanavar, R. et al. Dysregulated type I interferon and inflammatory monocyte-macrophage responses cause lethal pneumonia in SARS-CoV-infected mice. *Cell Host Microbe* **19**, 181–193 (2016).
43. Zhang, J. et al. Association of hypertension with the severity and fatality of SARS-CoV-2 infection: a meta-analysis. *Epidemiol. Infect.* **148**, e106 (2020).
44. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
45. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
46. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284–287 (2012).
47. Stern, J. N. et al. B cells populating the multiple sclerosis brain mature in the draining cervical lymph nodes. *Sci. Transl. Med.* **6**, 248ra107 (2014).
48. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020