



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



## Profiling of the immune repertoire in COVID-19 patients with mild, severe, convalescent, or retesting-positive status

Yonggang Zhou<sup>a,b,c,1</sup>, Jinhe Zhang<sup>b,c,1</sup>, Dongyao Wang<sup>a,b,c,1</sup>, Dong Wang<sup>b,c</sup>, Wuxiang Guan<sup>d</sup>, Jingkun Qin<sup>b,c</sup>, Xiuxiu Xu<sup>a,b,c</sup>, Jingwen Fang<sup>e</sup>, Binqing Fu<sup>a,b,c</sup>, Xiaohu Zheng<sup>b,c</sup>, Dongsheng Wang<sup>a</sup>, Hong Zhao<sup>f</sup>, Xianxiang Chen<sup>g</sup>, Zhigang Tian<sup>b,c</sup>, Xiaoling Xu<sup>a</sup>, Guiqiang Wang<sup>f,\*\*</sup>, Haiming Wei<sup>a,b,c,2,\*</sup>

<sup>a</sup> Respiratory and Critical Care Medicine, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, 230001, China

<sup>b</sup> Hefei National Laboratory for Physical Sciences at Microscale, The CAS Key Laboratory of Innate Immunity and Chronic Disease, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, 230027, China

<sup>c</sup> Institute of Immunology, University of Science and Technology of China, Hefei, 230027, China

<sup>d</sup> Center for Emerging Infectious Diseases, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, Hubei, 430071, China

<sup>e</sup> HanGene Biotech, Xiaoshan Innovation Polis, Hangzhou, Zhejiang, 311200, China

<sup>f</sup> Department of Infectious Diseases, Peking University First Hospital, Beijing, 100034, China

<sup>g</sup> Department of Tuberculosis, Wuhan Pulmonary Hospital, Wuhan, 430030, China

### ARTICLE INFO

#### Keywords:

COVID-19 predictor  
Immune repertoire  
Dual-injury mechanism  
BCR  
TCR

### ABSTRACT

Forty-seven samples of peripheral blood mononuclear cells from four groups of coronavirus disease (COVID)-19 patients (mild, severe, convalescent, retesting-positive) and healthy controls were applied to profile the immune repertoire of COVID-19 patients in acute infection or convalescence by transcriptome sequencing and immune-receptor repertoire (IRR) sequencing. Transcriptome analyses showed that genes within principal component group 1 (PC1) were associated with infection and disease severity whereas genes within PC2 were associated with recovery from COVID-19. A “dual-injury mechanism” of COVID-19 severity was related to an increased number of proinflammatory pathways and activated hypercoagulable pathways. A machine-learning model based on the genes associated with inflammatory and hypercoagulable pathways had the potential to be employed to monitor COVID-19 severity. Signature analyses of B-cell receptors (BCRs) and T-cell receptors (TCRs) revealed the dominant selection of longer V–J pairs (e.g., IGHV3-9–IGHJ6 and IGHV3-23–IGHJ6) and continuous tyrosine motifs in BCRs and lower diversity of TCRs. These findings provide potential predictors for COVID-19 outcomes, and new potential targets for COVID-19 treatment.

### 1. Introduction

Major infectious diseases pose huge threats to human health and great challenges to the security of global public health, and can cause more premature deaths than any other disease [1,2]. In the last two decades, infectious diseases caused by members of the coronavirus family have raged three times in the population: severe acute respiratory

syndrome coronavirus (SARS-CoV) [3], middle East respiratory syndrome coronavirus (MERS-CoV) [4] and SARS-CoV-2 [5].

At present, the outbreak caused by SARS-CoV-2 infection has been upgraded to a global pandemic. Worldwide, the number of known cases has exceeded 83 million, and led to >1.8 million deaths, according to data released by the World Health Organization (WHO) on January 3, 2021. This situation has placed extreme constraints on public-health

\* Corresponding author. Respiratory and Critical Care Medicine, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, 230001, China.

\*\* Corresponding author.

E-mail addresses: [john131212@sina.com](mailto:john131212@sina.com) (G. Wang), [ustcwhm@ustc.edu.cn](mailto:ustcwhm@ustc.edu.cn) (H. Wei).

<sup>1</sup> These authors contributed equally.

<sup>2</sup> Lead Contact.

resources worldwide [6–8]. The pandemic has not been brought under control effectively worldwide.

The acute infectious pneumonia caused by SARS-CoV-2 was named “coronavirus disease 2019” (COVID-19) by the WHO on February 11, 2020 [9]. The initial symptoms of COVID-19 are mostly mild (fever, dry cough, fatigue) and some patients develop severe symptoms (e.g., dyspnea) gradually [8]. Although most people with COVID-19 have mild disease, compared with the 2009 influenza pandemic, the proportion of COVID-19 patients requiring hospitalization is higher, and COVID-19 patients have about a fivefold increased risk of admission to the intensive care unit (ICU) [7].

SARS-CoV-2 targets cells in the nasal cavity and respiratory tract. It focuses on cells carrying angiotensin-converting enzyme 2 (ACE2) receptors and transmembrane serine protease 2 to infect the host [8,10]. Pulmonary cells infected with SARS-CoV-2 release inflammatory signals that activate the body’s antiviral immune response. However, SARS-CoV-2 infection causes an excessive non-effective response and severe inflammation in some COVID-19 patients by pathogenic T helper (Th)1 cells and inflammatory cluster of differentiation (CD)14<sup>+</sup>CD16<sup>+</sup> monocytes [11]. The level of proinflammatory cytokines such as granulocyte-macrophage colony-stimulating factor (GM-CSF), interleukin (IL)-1 and IL-6 released by these cells in the serum of patients with severe COVID-19 is increased, which results in recruitment of more activated immune cells into the lung to form “cytokine release syndrome” (CRS) [8,11–14]. Monoclonal-antibody drugs that target these proinflammatory cytokines, such as tocilizumab (*anti*-IL6 receptor) or mavrilimumab (*anti*-GM-CSF receptor- $\alpha$ ) can alleviate CRS and improve survival chances in patients with severe COVID-19 [11,15,16]. Nevertheless, some patients miss the period when efficacious medication can be administered due to the rapid progress of COVID-19. Finding and identifying features to improve the potential ability to predict the COVID-19 outcome after SARS-CoV-2 infection is expected to optimize the therapeutic effect of the targeted drugs. This strategy could help to reduce the incidence and mortality of severe COVID-19.

According to preliminary calculations based on the latest data released by the WHO, the overall prevalence of mortality in COVID-19 patients was ~2% (1.8/83). The mortality prevalence for COVID-19 varies markedly by age, and older patients ( $\geq 85$  years) are more at risk than children (aged 5–17 years) in the USA [8]. Most hospitalized patients can recover from COVID-19 after symptomatic treatment, thereby reaching the standard for hospital discharge (negative SARS-CoV-2 nucleic acid test (NAT)). However, unlike SARS and MERS, some COVID-19 patients discharged from hospital re-test positive for SARS-CoV-2 [17]: these patients are known as “retesting-positive” (RTP) patients. Upon re-admission to hospital, these patients show no significant clinical symptoms or disease progression, with normal or improved computed-tomography images and normal levels of proinflammatory cytokines [18]. In a clinical follow-up study from Wuhan (China), the prevalence of RTP patients was ~3%, which increased the risk of COVID-19 re-spread [19]. Therefore, looking for relevant potential predictors of COVID-19 outcomes involving RTP patients in the convalescent period is of great importance for the development of targeted treatment and early preventive measures.

Compared with SARS and MERS, COVID-19 caused by SARS-CoV-2 infection is relatively mild. However, SARS-CoV-2 is more contagious and has a longer and more complicated course, including an acute-infection period with obvious symptoms and a chronic convalescence period that may occur repeatedly. Looking for markers of patients with the severity or the RTP status is important opportunities for observing COVID-19 outcomes.

We wondered if we could identify markers that have the potential to predict COVID-19 outcomes. We employed comprehensive immune analyses using high-throughput sequencing of the transcriptome and the *repertoire* of immune receptors from the peripheral blood mononuclear cells (PBMCs) of patients. In this way, we wished to provide new insights into the pathogenesis of severe COVID-19 and the cause of RTP.

## 2. Materials and methods

### 2.1. Study design

The study protocol was approved by the Ethics Committee of the First Affiliated Hospital of the University of Science & Technology of China (2020-XG(H)-005) and Peking University First Hospital (2020-Research-112) for Emerging Infectious Diseases. The study was conducted in accordance with the International Conference on Harmonization Guidelines for Good Clinical Practice and the Declaration of Helsinki 1964 and its later amendments.

We profiled the transcriptome and the *repertoire* of immune receptors of PBMCs of COVID-19 patients, and to find potential clues to analyze COVID-19 outcomes. We addressed this objective by: (i) collecting PBMCs from 47 samples involved in five courses of disease; (ii) undertaking sequencing of the transcriptome and *repertoire* of immune receptors; (iii) summarizing the differences in the transcriptome or immune-receptor *repertoire* of different courses. Patients were divided into groups based on the clinical diagnosis.

### 2.2. Human samples

Forty-seven PBMC samples were from nine HCs, 16 non-ICU patients, six ICU patients, six CPs with negative viral NAT, and 10 CPs with positive viral NAT (RTP). The detailed patient information is shown in [Supplementary Table 1](#). The samples of HCs, non-ICU patients and ICU patients were collected from the First Affiliated Hospital of the University of Science and Technology of China (Hefei, China). All non-ICU patients and ICU patients were hospitalized patients diagnosed with SARS-CoV-2 infection. Patients in the ICU were admitted mainly because SpO<sub>2</sub> was lower ( $\leq 93\%$  while breathing room air), and because high-flow nasal intubation or higher-level oxygen support were required to correct hypoxemia. PBMC samples of CPs and RTP patients were collected in the Wuhan Pulmonary Hospital (Wuhan, China). CPs and RTP patients were convalescent patients with COVID-19 who had reached negative viral NAT and discharged after treatment. In the follow-up after hospital discharge, CP’s viral NAT remained negative, whereas that of RTP patients was positive, and they were re-admitted to hospital. Neither CPs nor RTP patients had obvious symptoms.

### 2.3. Transcriptome sequencing

Forty-seven PBMC samples from the five groups were subjected to transcriptome sequencing by the BGISEG platform (Beijing Genomics Institution (BGI), Beijing, China). Briefly, 500  $\mu$ L of whole blood from each sample underwent centrifugation at 300 $\times$ g for 5 min at room temperature. Red blood cell lysis buffer was used to remove red blood cells. Then 1 mL of TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) was added to lyse the remaining cells and stored at  $-20$  °C. Total RNA was extracted and purified by phenol–chloroform extraction and treated with DEPC-treated water. RNA libraries were prepared for sequencing using the MGISEQ-2000 R S High Throughput (Rapid) Sequencing Reagent Kit according to standard BGI protocols. Raw data of RNA-Seq were deposited in the genome sequence archive (GSA) database of national genomics data center (accession number, CRA003083).

### 2.4. Mapping and counting mRNA transcripts

Cutadapt 2.10 (<https://cutadapt.readthedocs.io/en/stable/index.html>) was used for adapter trimming. Clean reads were aligned to the hg38 reference genome using hisat2. Then, they were annotated and counted with gene annotations gencode v30 using feature counts. All samples were merged to produce a digital counts matrix and normalized to TPM with customized R script. The gene-expression matrix was filtered to retain only genes with an average TPM  $>0.5$ ; this strategy resulted in 20,717 expressed genes.

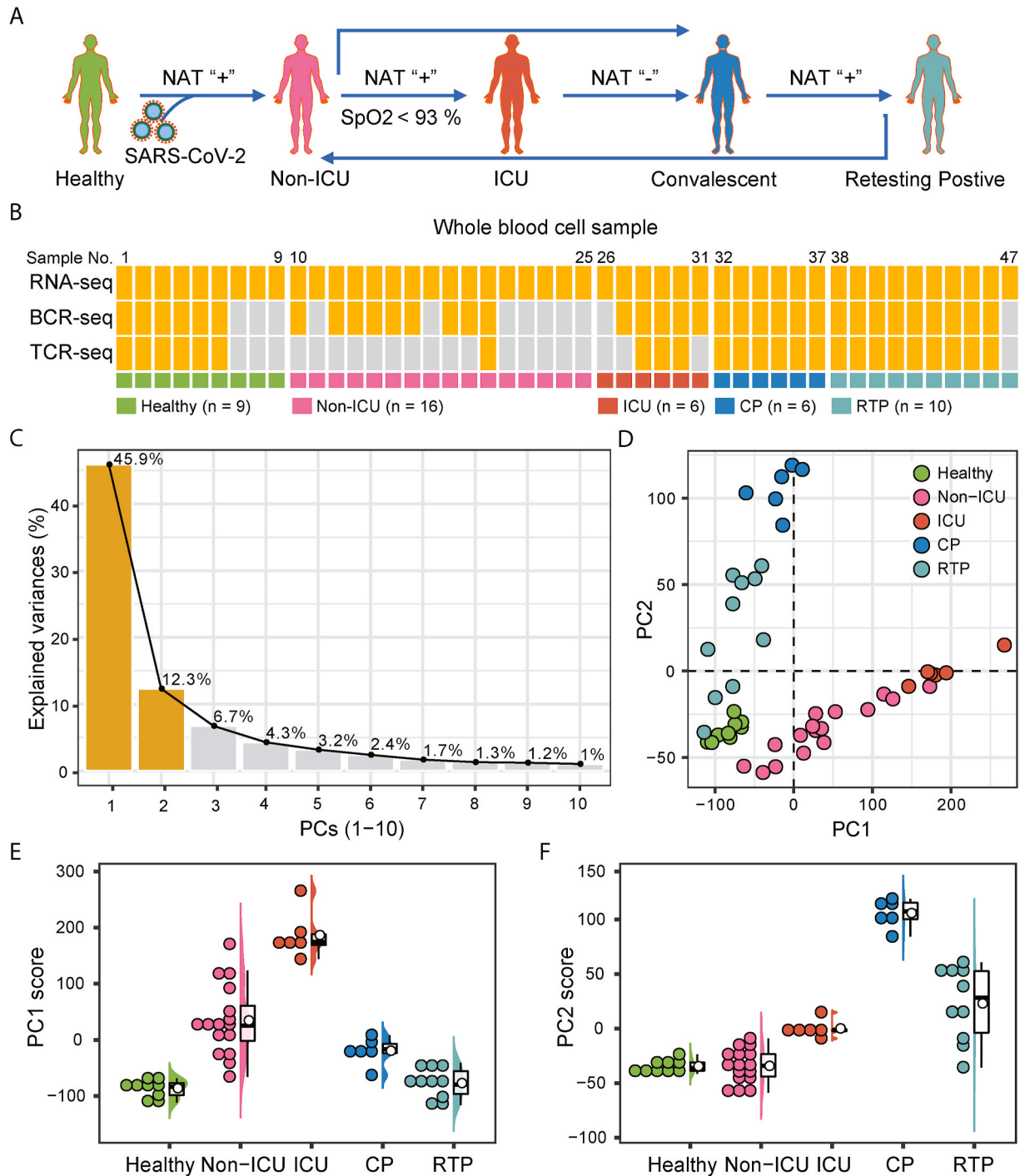
2.5. PCA

To uncover the biological knowledge from our time-course design, we undertook on a logarithmic TPM matrix to identify disease-related factors. We selected the top-100 genes with the highest contribution to each component to enriched GO terms using cluster profiler. List of

genes contained in PC1 and PC2 showed in the [Supplementary Table 2](#).

2.6. GSVA of pathways

To identify the gene-expression signatures and pathway activation, GSVA (a non-parametric and unsupervised algorithm) was employed.



**Fig. 1.** PC1 and PC2 distinguish between infected and convalescent patients with COVID-19. (A) A model diagram depicting the COVID-19 course at full-scale. NAT "+/-" denotes that the nucleic-acid test result of SARS-CoV-2 is positive ("+") or negative ("-"). (B) Overview of the experimental design and number of samples for RNA-seq, BCR-seq and TCR-seq. Healthy control (n = 9): healthy people not infected with SARS-CoV-2. Non-ICU patients (n = 16): patients with mild COVID-19 admitted to the general ward. ICU patients (n = 6): patients with severe COVID-19 admitted to the intensive care unit (ICU). CP patients (n = 6): convalescent patients with a negative viral nucleic-acid re-test result. RTP patients (n = 10): convalescing patients with a positive viral nucleic-acid re-test result. (C) The proportion of PC1-10 in the PCA of transcriptome data from the 47 patients in the five groups. (D) PCA of transcriptome data from the five groups based on PC1 and PC2. (E and F) Correlation analyses between PC1(E) or PC2(F) genes and patients in the five groups.

Gene-set collections in MsigDB, including H (hallmarker), C2 (canonical pathways), C5 (gene ontology) and C7 (immunologic signatures) were selected. Gene sets that were changed significantly ( $P$  adjusted  $<0.1\%$ ) compared with those of HCs and which correlated significantly ( $P$  adjusted  $<0.1\%$ ) with PC1 and PC2 were utilized.

## 2.7. Immune infiltration

To infer the composition of immune cells in PBMCs, the TPM matrix was processed by xCell algorithm v1.1.0 with default settings except the cell\_types.use parameter (basophils, eosinophils, megakaryocytes, monocytes, neutrophils, NK cells, plasma cells, and all B- and T-cell subtypes were selected). The signature genes involved in xCell showed in the [Supplementary Table 3](#).

## 2.8. MLM

To establish a model for evaluating COVID-19 severity based on the dual-injury mechanism, we created the learning set based on the 632 inflammation- and vascular injury mechanism-related genes from non-ICU and ICU groups using the leave-one-out method. With glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) as an internal reference, the relative gene expression was calculated. Using the  $P$  value as the screening condition, a logistic regression algorithm was used to model, and the final 16 featured genes and data models were determined through 100 permutations. The transcriptome data of 10 COVID-19 patients who were RTP and the dynamic data of three cases during hospitalization in the Array Express database (E-MTAB-8871) [20] were used as the test set. All the in house developed codes/scripts were uploaded to Github website ([https://github.com/USTC-IOI/RNA\\_IR\\_analysis\\_of\\_SARS-nCov](https://github.com/USTC-IOI/RNA_IR_analysis_of_SARS-nCov)).

## 2.9. Sequencing of the repertoire of immune receptors

PBMC samples of some patients in the five groups used for the sequencing of the immune-receptor repertoire are shown in [Fig. 1B](#). BCR or TCR libraries were prepared for sequencing using standard protocols based on the Illumina Adaptor Primer and Multiplex PCR kit (Qiagen, Stanford, VA, USA). A bioanalyzer (2100; Agilent Technologies, Santa Clara, CA, USA) was used to ascertain the quality of the libraries. The BCR region of the immune-receptor repertoire was sequenced using an Illumina platform (PE150). The TCR region of the immune-receptor repertoire was sequenced using another Illumina platform (PE101). Raw data of the repertoire of immune receptors were deposited in the GSA database of national genomics data center (accession number, CRA003098).

## 2.10. Mapping of the immune-receptor repertoire

SOAPnuke was used for data filtering. Software that compares immune-receptor repertoire (MiXCR) was used for sequence comparison. The comparison library is the software's own gene library. Sequences that were identical were classified as the same clone type (clonotype). Reads with lower-quality values were compared twice to reduce errors introduced by PCR and sequencing. The final result file was a tabbed text file, and each line contained information such as the amount, frequency, CDR3 sequence, and V/D/J gene of a clone type.

## 2.11. CDR3 cluster analysis

Because the CDR3 of BCR or TCR differs greatly between groups, most of them are expressed within a single group or even within a single sample, and it is difficult to perform statistical analysis between groups, so cluster analysis (component analysis) is performed. Group by CDR3 length, hamming distance and the number of different amino acids between different CDR3 groups were calculated. Using a hamming

distance as a condition, perform network discovery clustering (netsig), and construct components. For each component, draw the amino acid ratio diagram (seqlogo diagram) and network diagram (network) at each position, and count the relevant V and J regions.

## 2.12. Structural simulation of antibody binding to the spike protein

The homology of antibody structure is relatively high. Hence, homology-modeling method was used to simulate the structure of the heavy chain of an antibody. We used swissmodel ([swissmodel.expasy.org](http://swissmodel.expasy.org)) for homology modeling. Enter the heavy chain sequence we obtained through BCR sequencing to this website, it would give the antibody with the highest homology to the existing structure, and build the target protein model. Then, we used Zdock (<http://zdock.umassmed.edu/>) for rigid protein docking. Here, we used the modeled protein structure to dock with the spike protein of SARS-CoV-2 and ACE2 complex and, finally, obtained docking scores.

## 2.13. Chemiluminescence assay

Collect the serum of CP and RTP patients for SARS-CoV-2 IgG and IgM detection. Perform chemiluminescence detection of SARS-CoV-2 IgG and IgM according to the kit instructions (YHLO, Cat. No.: C86095 M and Cat. No.: C86095G).

## 2.14. Statistical analyses

Statistical significance between two groups was determined using unpaired two-tailed  $t$ -tests, and between multiple groups by one-way analysis of variance. Data are the mean  $\pm$  SEM.

## 3. Results

### 3.1. Differences in principal component analysis (PCA) of COVID-19 patients with different courses

COVID-19 patients may go through a relatively complicated course from being infected with SARS-CoV-2 to being "cured" ([Fig. 1A](#)): mild, severe, convalescence, or RTP. Viral NAT is a core indicator for evaluating infection, and blood oxygen saturation ( $SpO_2$ )  $<93\%$  is a key indicator for evaluating severe cases.

We wished to profile the immune characteristics of COVID-19 patients with different courses to identify potential predictors of the outcome and guide treatment. Hence, we collected PBMCs from 47 patients for high-throughput sequencing: nine healthy controls (HCs), 16 non-ICU patients, six ICU patients, six convalescing patients (CPs) and 10 RTP patients ([Fig. 1B](#)). PCA was used to reduce dimensionality to analyze the transcriptome data of 47 samples in five groups.

We found that principal component (PC)1 (45.9%) and PC2 (12.3%) were the main components, and that the proportion of other PCs was  $<10\%$  ([Fig. 1C](#)). Using PC1 and PC2 to profile the characteristics of the transcriptome, we found that PC1 and PC2 could distinguish patients in the five groups. Non-ICU patients and ICU patients were distributed in the PC1 area, whereas CPs and RTP patients were distributed in the PC2 area ([Fig. 1D](#)). Furthermore, the PC score was used to assess the correlation between the patients in the five groups and PC1 or PC2. The PC1 score increased in HCs, non-ICU patients and ICU patients, and rose to a peak in ICU patients, whereas CPs and RTP patients showed a low correlation with PC1 ([Fig. 1E](#)). These data showed that PC1 was a gene set related to SARS-CoV-2 infection and COVID-19 severity. The PC2 score was higher in CPs and RTP patients than that in the other three groups, and highest in CPs. These observations showed that PC2 was a gene set related to convalescence from COVID-19.

However, we did not detect the nucleic-acid sequence of SARS-CoV-2 in the transcriptome data of all samples ([Supplementary Fig. 1](#)). The difference in gene-expression characteristics between HCs and the other

four groups was not caused directly by SARS-CoV-2 infection, but likely was the response of blood cells after systemic infection (especially immune cells). Therefore, based on transcriptome data, we used xCell algorithm v1.1.0 to analyze the changes in the proportion of various immune cells in the sample. Using this method, our analyses showed that compared with HCs, the proportion of CD4<sup>+</sup> T cells and CD8<sup>+</sup> T cells decreased in non-ICU and (especially) ICU patients, whereas the proportion of B cells and plasma cells increased in non-ICU and ICU patients (Supplementary Fig. 2). This decrease in the proportion of T cells is consistent with the flow cytometry data we have reported [11]. Also, the increase in the proportion of B cells may be related to upregulation of expression of B cell-characteristic genes (e.g., *PRDM1*, *XBPI1*, *IRF4*) as reported in a single-cell sequencing study by Zhu and colleagues [21]. These results suggested that PC1 and PC2 may be gene sets related to changes in the immune response of COVID-19 patients upon acute infection or long-term convalescence.

### 3.2. A “dual-injury” mechanism based on inflammatory and hypercoagulable pathways is present in severe COVID-19

A decrease in the number of T cells (especially CD4<sup>+</sup> T cells) is an obvious feature of COVID-19 [8,11,22]. Based on the xCell algorithm, contrary to the decrease in the proportion of CD4<sup>+</sup> T cells in acute infection with SARS-CoV-2, the proportion of polarized Th1 and Th2 cells increased (Supplementary Fig. 2A), and these cells participated in CRS development. In contrast, the proportion of T regulatory cells (which negatively regulate the inflammatory response) was reduced in non-ICU patients and ICU patients, and increased obviously in recovered patients and RTP patients (Supplementary Fig. 2A). The proportion of CD8<sup>+</sup> T cells (the main force of the antiviral immune response) showed a slight increase in non-ICU patients, but a visible decrease in ICU patients (Supplementary Fig. 2B).

In addition to relying on marker molecules to analyze changes in types of immune cells, gene set variation analysis (GSVA) was undertaken to assess the enrichment of signaling pathways in the patients in the five groups to enhance understanding of functional activation of immune cells (Supplementary Fig. 3). Focusing on the gene set in PC1 related to acute infection, the humoral immune response, complement activation, the chemotaxis activity of immune cells, and inflammatory response-related signaling pathways were enriched sufficiently in non-ICU and ICU groups compared with those of other groups (Fig. 2A and Supplementary Fig. 3). The humoral immune response is mediated not only by antimicrobial peptides but also by immunoglobulin (Ig)-producing B cells, and activation of B cells occurs more often in infection (Supplementary Fig. 3). Contrary to the immune-activation signal, expression of some classic immunosuppressive pathways was down-regulated in non-ICU and ICU groups and correlated negatively with PC1, such as the suppressor of cytokine signaling 3 (SOCS3), forkhead box P3 (FOXP3), IL-10, and transforming growth factor (TGF)- $\beta$  (Fig. 2B and Supplementary Fig. 3). This lack of negative regulatory signals leads to an immune imbalance and accelerates CRS development. However, antiviral pathways, such as natural killer (NK) cells, CD8<sup>+</sup> T cells and type-I interferon receptor signaling, were negatively related to PC1, especially in the ICU group. These data suggested that the low antiviral immune response of patients in the ICU group was also an important factor in disease deterioration (Fig. 2C and Supplementary Fig. 3). The proportion of NK cells and CD8<sup>+</sup> T cells decreased in the ICU group compared with that in the non-ICU group (Supplementary Figs. 2B and D). The proportion of neutrophils that may be associated with excessive inflammation increased in the ICU group (Supplementary Fig. 2E). To evaluate further the pathways involved in PC1, we discovered that expression of proinflammatory pathways increased in acute infection, and was most enriched in the ICU group (Fig. 2D (left), 2E). Conversely, expression of anti-inflammatory pathways increased in recovered patients and RTP patients (Fig. 2D (middle), 2E). Combining positive and negative factors, we found that from patients with mild disease to severe

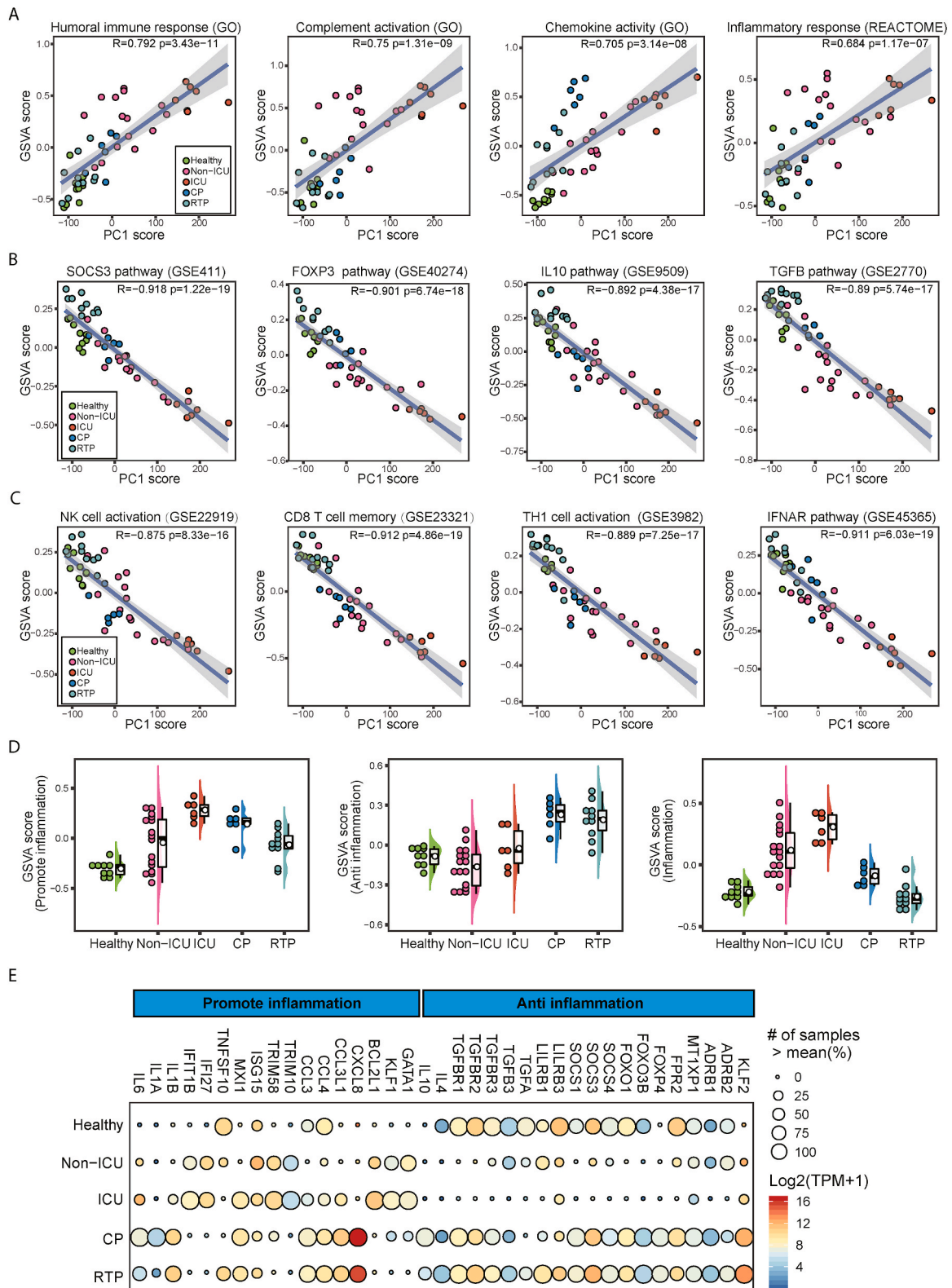
disease, the immune imbalance developed to CRS, but the immune balance was restored in the convalescent period (Fig. 2D, right).

In addition to inflammation, decreased SpO<sub>2</sub> due to dyspnea is a typical feature of severe COVID-19 [8]. Differential analyses of the top-100 genes in PC1 showed increased expression of the hemoglobin-1 gene in the ICU group (Supplementary Figs. 4A and B). Upon functional-enrichment analyses of these genes using the Gene Ontology database, we found that these genes were not only involved in the immune response but also related to the hypoxic response of erythrocytes (Supplementary Fig. 4C). GSVA of these genes showed that expression of blood oxygen-related pathways (e.g., gas transport, hemoglobin complex, erythrocyte maturation) was enriched significantly upon infection, especially in the ICU group (Fig. 3A). Over-enrichment of oxygen transport-related pathways corresponds to the clinical features of hypoxia in severe COVID-19. The hypoxic environment *in vivo* changes oxidative metabolism and accelerates the production of peroxides, and enrichment of peroxide metabolic pathways was also observed in non-ICU and ICU groups (Fig. 3A and Supplementary Fig. 5). Excessive accumulation of peroxides not only aggravates the inflammatory response, it also carries the risk of damaging vascular endothelial cells (together with proinflammatory cytokines such as IL-1, IL-6 and TNF- $\alpha$ ). In addition, fibrinolysis, collagen-trimer, platelet and blood-microparticle pathways related to blood coagulation were highly enriched in patients upon infection and even in some convalescent patients (Fig. 3A). Hypercoagulation caused an increase in blood pressure and vascular permeability, which put greater pressure on endothelial cells (Fig. 3A). Further focusing on the functional changes of vascular endothelial cells, pathways that negatively regulate the proliferation and repair of endothelial cells were enriched upon infection, suggesting that endothelial cells in COVID-19 patients are dysfunctional (Fig. 3A and B). Importantly, expression of “don’t eat me” signals was reduced in the ICU group, such as CD47/SIRPA, CD24/SIGLEC10 and SLAMF7 signals, which can aggravate the damage of activated macrophages to endothelial cells (Fig. 3C). Due to the wide distribution of blood vessels in the body, endothelial dysfunction leads to a further increase in permeability, and release of more proinflammatory cytokines and peroxides in tissues further aggravates damage to blood vessels and tissues, thereby forming a dual-injury mechanism of systemic multi-organ injury in COVID-19 patients. This mechanism echoes the histopathological results of exudative inflammation and hypercoagulation presented in the autopsy of COVID-19 victims [23,24].

### 3.3. A machine-learning model (MLM) based on the dual-injury mechanism to screen potential predictors of COVID-19 severity

Once severe multi-organ injuries occur in COVID-19 patients, the cure rate is low, which seriously threatens the lives of patients [7,8]. Unfortunately, COVID-19 often accelerates suddenly. Hence, clarifying the COVID-19 outcome as early as possible after infection is crucial for timely targeted treatment.

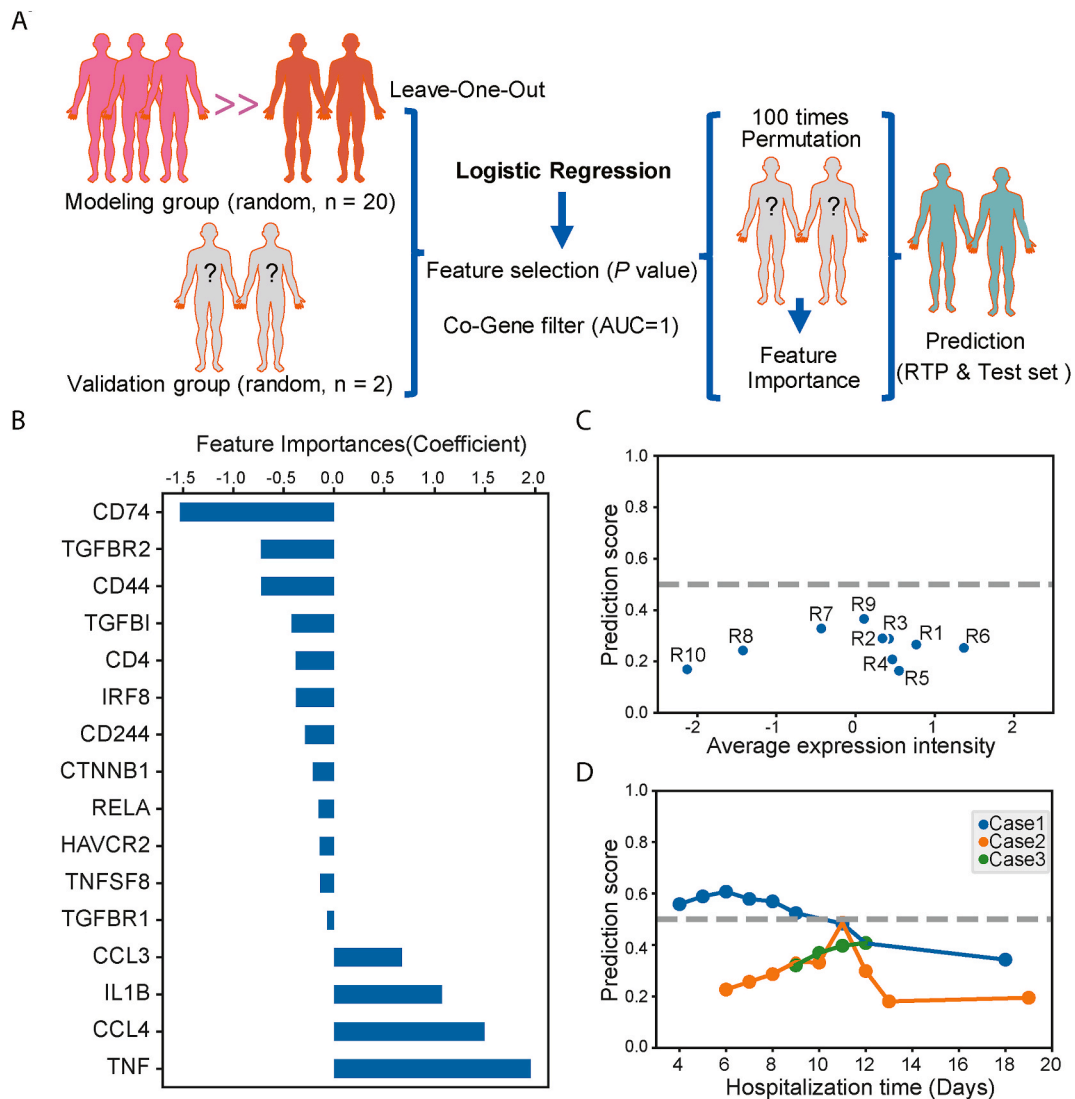
To screen potential predictors of COVID-19 progression in the early infection period, we built a MLM based on the genes associated with the dual-injury mechanism from the PC1 gene set. Based on the tag per kilobase million values of related genes from 16 non-ICU samples and six ICU samples, the leave-one-out method was employed to establish a modeling group and validation group (Fig. 4A). The size of each modeling group was 20 samples and that of the validation group was two samples. Using the *P*-value as the screening criterion, a mathematical model was established through logistic regression, and 100 permutations were performed to screen and filter genes. Finally, a feature set containing 16 genes was obtained (Fig. 4A and B). Among them, *TGFB1*, *TGFB1*, *TGFB2* and *CD74* (as immune-negative signals) and *CD44*, *IRF8* and *CD244* (as antiviral-activity signals) could inhibit COVID-19 severity, whereas *IL-1*, *TNF*, *CCL3* and *CCL4* (as proinflammatory cytokines) promoted COVID-19 severity (Fig. 4B). These genes also had an important role in defining the dual-injury mechanism



**Fig. 2.** An imbalanced immune response to inflammation is related to triggering of COVID-19 severity. (A, B, C) Correlation analyses between PC1 and GSVAs enrichment score of proinflammatory pathways (A), anti-inflammatory pathways (B) and antiviral immune-response pathway (C). (D) Dynamics of the degree of GSVAs enrichment of proinflammatory-related pathways (left), anti-inflammatory-related pathways (middle) and a combination of the two (right) in the five groups of healthy controls, non-ICU, ICU, CP and RTP. (E) Bubble chart showing the dynamic difference in expression of featured genes in the proinflammatory and anti-inflammatory pathways enriched by GSVAs in the five groups. “# of samples > mean (%)” show the proportion of samples whose expression is greater than the mean. TPM, transcripts per kilobase million.







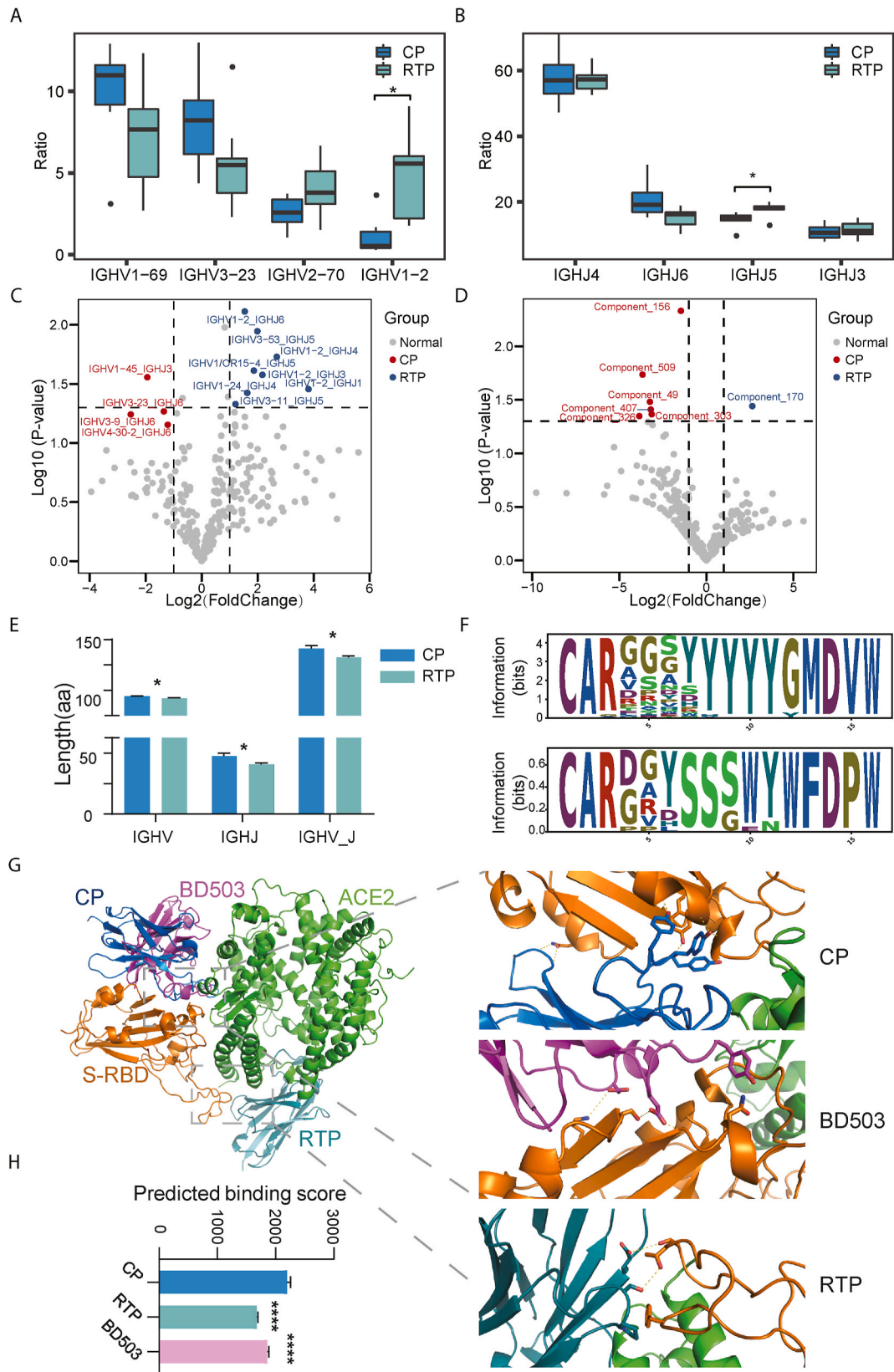
**Fig. 4.** An AI-based gene model that has the potential to estimate COVID-19 severity.

(A) Study design of a machine-learning model for screening genes associated with COVID-19 severity based on transcriptome data from non-ICU and ICU patients with COVID-19 and a logistic regression algorithm. (B) The top-16 featured genes of COVID-19 severity calculated by the machine-learning model. (C) Performance of the machine-learning model in a test cohort of 10 RTP patients with COVID-19. (D) Performance of the machine-learning model in a test cohort of dynamic transcriptomic data of PBMCs from three COVID-19 cases in the array express database (E-MTAB-8871).

In Case 1 (a 66-year-old man with a history of fever and cough and a chest radiograph showing typical COVID-19 pathology), SpO<sub>2</sub> decreased and supplemental oxygen was given 3 days after hospital admission [20]. Our MLM suggested that the score of Case 1 would increase 4–6 days after hospital admission, and would then decrease to <0.5 on day 10 (Fig. 4D). This calculation corresponded to the clinical manifestations of Case 1: SpO<sub>2</sub> was lowest on 3–5 days, and then restored, and oxygen supplementation was stopped on day 12 [20]. Cases 2 and 3 had the mild clinical manifestations of COVID-19, with no records of decreased SpO<sub>2</sub> or the requirement for supplemental oxygen [20]. Our MLM showed that the scores of Cases 2 and 3 would be < 0.5 in the 2 weeks when test data were available, and they had mild COVID-19 (Fig. 4D). These test results showed that 16 genes as potential predictors of COVID-19 severity identified through MLM screening based on genes related to the dual-injury mechanism have the potential to be used for analyzing the condition of a patient with COVID-19.

#### 3.4. Diversity of the immune-receptor repertoire in convalescent patients with COVID-19 is associated with outcomes

After treatment, the SARS-CoV-2 load in the body is reduced greatly, viral NAT is negative, and the patient enters a long convalescence period with slow repair of tissue damage [8,25]. Some convalescent patients with a weak immune response after SARS-CoV-2 infection are positive for SARS-CoV-2 NAT, which harbors risks to the treatment and prevention of COVID-19 [17,18]. The PC2 gene set showed the potential to distinguish between patients who had infection and those convalescing (Fig. 1D). Hence, we focused on whether the PC2 gene set could distinguish between CPs and RTP patients. The expression of the top-100 genes of PC2 was enriched in the CP group and RTP group compared with that in the other groups (Supplementary Figs. 6A and B). These genes were involved mainly in the regulation of immune balance and repair of various tissue injuries (Supplementary Fig. 6C). Furthermore, we used GSVA to assess differences in enrichment of multiple pathways related to immune recognition, immune activation, immune regulation, and vascular repair in PC2 (Supplementary Fig. 7). Compared with other



**Fig. 5.** Select shorter BCRs are related to RTP in convalescent patients.

(A and B) Differences in IGHV (A) and IGHJ (B) selection between the CP group and RTP group. (C) Differential dominance in IGHV-IGHJ pairs between the CP group and RTP group. (D) Difference in statistical parameters of IGHV-IGHJ pairs between the CP group and RTP group. (E) Differential dominance in components between the CP group and RTP group. (F) The CDRH3 motif was the most significant difference in the RTP group and CP group. (G and H) Structural simulation (G) and predictive binding force statistics (H) of the simulated neutralizing antibody competing with the S protein and ACE2 protein complex.

groups, the signaling pathways involving activation of toll-like receptors (TLR) signaling, T cells and NK cells, inhibitory immune regulation, lymphocyte migration and angiogenesis were enriched in CP and RTP groups, and the degree of enrichment of the two groups was similar (Supplementary Fig. 7). These data suggest that the genes involved in the enrichment pathway in PC2 had limited ability to distinguish between the CP group and RTP group. Although the pathways of B-cell receptor (BCR) and T-cell receptor (TCR) activation had similar enrichment levels in the CP group and RTP group, there may have been key differences in the diversity of BCRs and TCRs between these two groups to distinguish RTP.

To find clues from differences in BCRs or TCRs, we undertook sequencing analyses of the *repertoire* of immune receptors in PBMC samples from some patients in the five groups (Fig. 1B). BCR sequencing showed that the diversity of the immune-receptor *repertoire* decreased upon infection, and recovered in the convalescent period (Supplementary Fig. 8A). Expression of the top 5% of BCRs was more enriched in the recovery period, especially in the ICU group (Supplementary Fig. 8B). The Shannon-Wiener index of BCR also suggested that their decrease upon infection and return in the convalescent period may be related to the specific activation and expansion of B cells after infection (Supplementary Fig. 8C). Analyses of BCR rearrangement were done between the five groups to determine the dominant selection in regions V and J (Supplementary Figs. 8D and E). Focusing on convalescence, the CP group had greater levels of immunoglobulin G heavy chain variable (IGHV)1-69 and IGHV3-23, and less IGHV1-2, compared with that in the RTP group (Fig. 5A, left). In the J region, the CP group had more IGHJ6 instead of IGHJ5 compared with that in the RTP group (Fig. 5A, right). V-J rearrangements in the CP group were more concentrated, whereas in the RTP group there were more V-J rearrangements and greater diversity (Fig. 5B). Research on the human immunodeficiency virus (HIV) and influenza B virus has shown that neutralizing antibodies with more amino acids to form a longer CDR-H3 loops have a better chance of breaking through the blockade of viral-antigen glycosylation to recognize viruses [26,27]. Hence, we compared the amino-acid number of V-J-dominant combinations from the CP group and RTP group. The amino-acid number of the dominant selection in the V region and J region in the CP group was more than that in the RTP group, so the V-J-dominant combinations were also longer in the CP group (Fig. 5C). These findings suggested that the neutralizing antibodies produced by B cells in the RTP group may not have been able to break through glycosylation of the spike protein effectively to recognize SARS-CoV-2 due to the shorter CDR-H3 loops.

CDR3 is the key to antibody affinity. However, CDR3 is extremely diverse, which results in very large differences between samples. To analyze the difference of the CDR3 sequence of the BCR between CP and RTP groups, we grouped them according to the amino-acid number of the CDR3 sequence and then undertook component analyses. The sequence of component-156 was the most significant in the CP group, and component-170 was the most significant in the RTP group (Fig. 5D). Further analyses of the sequences of components revealed obvious differences in the central regions of component-156 and component-170 (Fig. 5E). The core region of component-156 contained five consecutive tyrosine residues, whereas the core region of component-170 contained more serine residues (Fig. 5E). Therefore, we hypothesized that the antibodies produced by the most dominant B-cell clones in the CP group and RTP group may recognize different antigens or have different affinities. To verify this hypothesis further, we combined the most dominant component and V-J regions of the CP group and RTP group, and simulated the antibody heavy-chain structure based on the V-CDR3-J sequence. The neutralizing antibody BD503 has been shown to block binding of the receptor binding domain (RBD) of SARS-CoV-2 to the target cell receptor (ACE2) [28]. We undertook dynamic simulation analyses of the RBD of the spike protein and structure of the ACE2 complex derived from the database and simulated the antibody structure. The most dominant antibody in the CP group had similar binding

positions with that of BD503 in the spike protein, and had strong binding potential (Fig. 5F and G). These observations imply that the antibodies produced by B cells in the CP group were better at blocking SARS-CoV-2 invasion than those in the RTP group. In addition, although patients in the RTP group seemed to have more antibodies, they failed to eliminate SARS-CoV-2 effectively (Supplementary Figs. 8F and G), further suggesting that the antibodies produced by the B cells in the RTP group may not be able to recognize and neutralize SARS-CoV-2.

T cells also play an important part in clearing SARS-CoV-2 infection [29,30]. TCR sequencing of samples from five groups was undertaken to study changes in the diversity of T-cell clones after SARS-CoV-2 infection. Similar to BCRs, the diversity of TCRs also decreased significantly in non-ICU and ICU groups, and recovered in convalescence (Supplementary Figs. 9A, B, C). GSVA of transcriptome data suggested that T cells had more obvious activation of the antiviral response in convalescence than that in infection (Supplementary Fig. 3 and Supplementary Fig. 7). Focusing on convalescent patients, we found that the TCR diversity of the RTP group was greater than that of the CP group (Supplementary Figs. 9A, B, C). Next, to further understand the rearrangement of TCRs, we separately analyzed the dominant selection of V and J regions. In the V region, TRBV28 was selected more often in the CP group compared with that in the RTP group (Supplementary Fig. 9D). In the J region, TRBJ2-7 was selected more often in the CP group, whereas TRBJ1-5 and TRBJ1-4 were selected more often in the RTP group (Supplementary Fig. 9E). Further evaluation of the diversity of the pairs of the dominant selected V and J regions revealed that, compared with that in HCs, V-J combinations were more concentrated and less diverse in the convalescent period (Fig. 6A). These observations further suggested that T-cell clones may have expanded and participated in elimination of residual SARS-CoV-2 in convalescence. However, compared with the CP group, the V-J combinations in the RTP group were more discrete, with greater diversity, which also indicated that the T-cell clones in the RTP group may not have been amplified effectively (Fig. 6B). In addition, component analyses were carried out to identify the dominant CDR3 sequences in each group. Due to the highly diverse nature of CDR3, although we undertook cluster analyses, only a few consensus sequences were identified, of which eight CDR3 sequences shared in all five groups and three CDR3 sequences shared in four groups of patients may have been related to SARS-CoV-2 infection (Fig. 6C). The eight TCR sequences was consistent with the activation time of T cells, which increased in infection and decreased in convalescence, and was higher in the CP group than that in the RTP group (Fig. 6D). These findings suggested that a reduction in the number of these TCRs may be related to the weak clearance of virus by T cells in the RTP group, which must be confirmed in future studies.

#### 4. Discussion

SARS-CoV-2 has infected >83 million people and wrought havoc upon healthcare systems worldwide. We collected PBMC samples from five groups involved in the complete course of COVID-19 for high-throughput sequencing: HCs, mild (non-ICU), severe (ICU), CPs, and RTP. Based on the sequencing data of the transcriptome and *repertoire* of immune receptors from the PBMCs of COVID-19 patients, we tried to find clues to identify potential predictors related to the COVID-19 outcomes of patients upon infection and in the convalescent period (Fig. 7).

Inflammation is a major cause of COVID-19 severity [8,11]. Transcriptome data showed that the increased inflammation shown in patients after SARS-CoV-2 infection was related to immune imbalance. In ICU patients, the imbalance was more prominent, the antiviral immune response of NK cells and CD8 T cells was low, activation of humoral immunity and the complement system further aggravated the inflammatory response, and decline of some classic immune-negative signals enhanced the deterioration of inflammation (Fig. 2). Peripheral lymphocytes from COVID-19 patients over-express exhaustion-related makers [11,31,32]. The serum level of IgA is positively correlated with

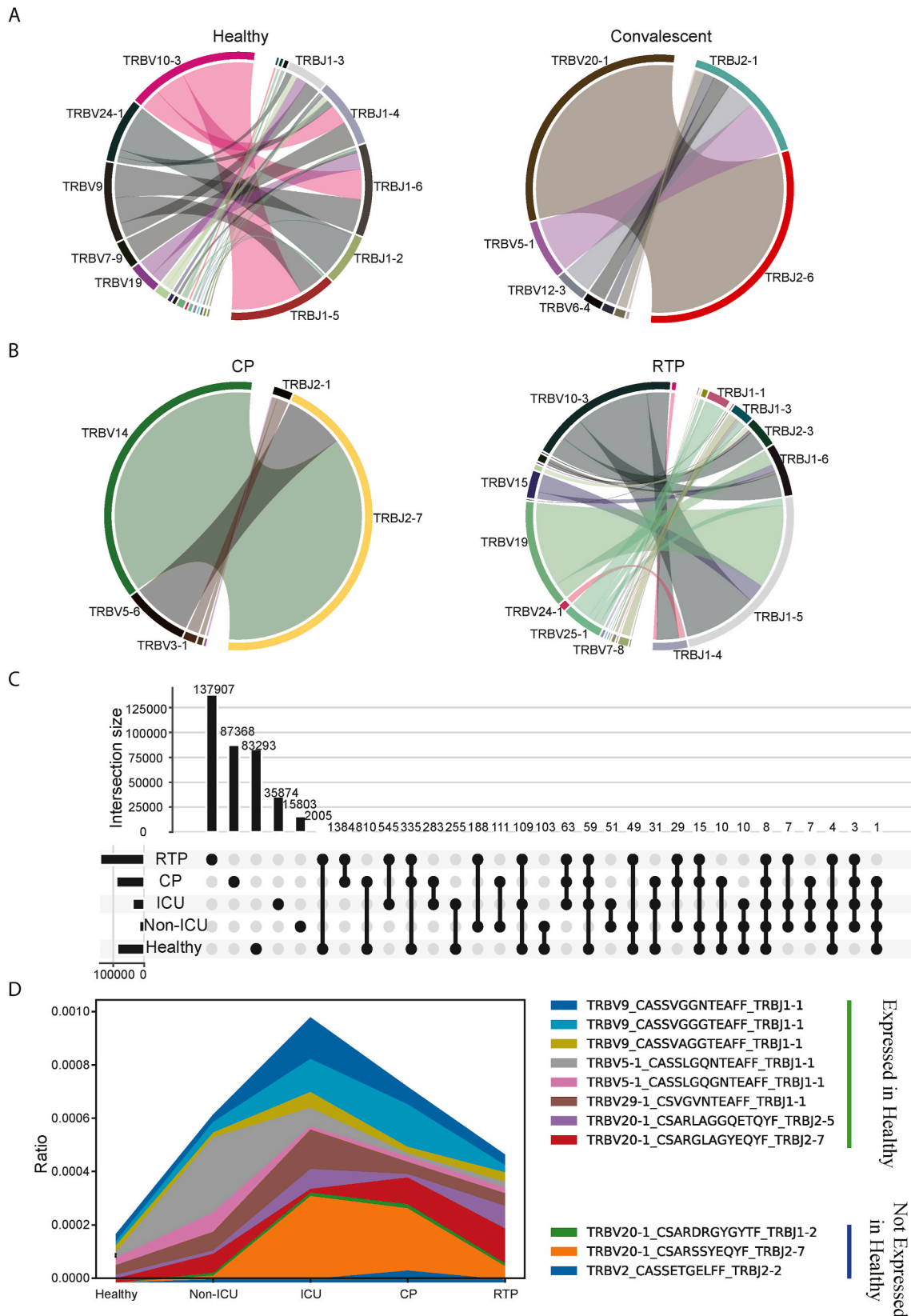


Fig. 6. Scattered TCR clone type is related to RTP in convalescent patients.

(A) Differential dominance in TRBV-TRBJ pairs between healthy controls and convalescent patients. (B) Differential dominance in TRBV-TRBJ pairs between the CP group and RTP group. (C) Upset chart of common components among the five groups. (D) The potential dynamic changes in the SARS-CoV-2-specific TCR sequence in the five groups of COVID-19 patients.

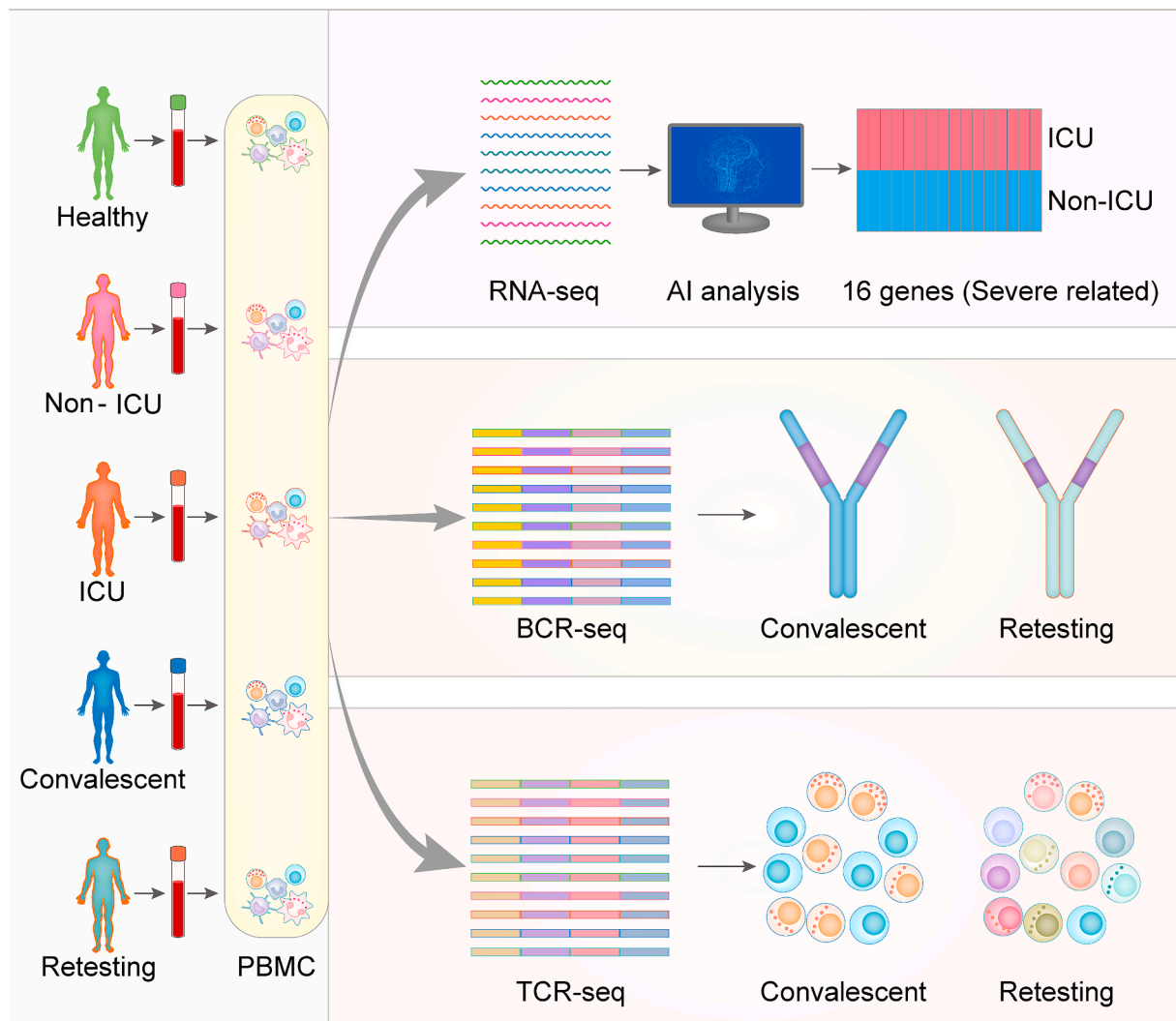


Fig. 7. Featured markers associated with immune response in the COVID-19 severity and the retesting-positive status.

COVID-19 severity [33]. Those observations support our analyses of the role of inflammation in COVID-19. In addition, COVID-19 patients often have platelet-fibrin thrombi in small arteries [23]. Excessive activation of platelets may be related to hypercoagulation [34]. We found that the proportion of megakaryocytes (main source of platelets) was correlated with COVID-19 severity (Supplementary Fig. 2F). However, this seems to be only part of the cause of hypercoagulation. On the basis of inflammation, we showed that the increased peroxide accumulation and hypercoagulable pathways associated with hypoxia and inflammation caused further dysfunction of vascular endothelial cells (Fig. 3). Clinically, endothelial dysfunction is associated with hypercoagulation [35]. Endothelial dysfunction was closely related to an increase in vascular permeability in patients with severe COVID-19 (Fig. 3). This increase in vascular permeability would allow more inflammatory factors and inflammatory cells to enter tissues, thereby aggravating tissue injury and even causing multi-organ injury. Our results suggest that loss of the don't-eat-me signal may also be an important mechanism of endothelial dysfunction in COVID-19 patients (Fig. 3). The classic don't-eat-me signaling molecule [36], CD24Fc, has been used to test alleviation of the pathological injury caused by COVID-19 [37]. We propose a dual-injury mechanism based on inflammatory and hypercoagulable pathways in the mechanism of COVID-19 severity.

Gradually, COVID-19 is being recognized as a systemic disease [8, 38]. In addition to an excessive inflammatory response and increased vascular-injury signals, transcriptome analyses suggested that the sense

of smell and taste may also be disturbed in patients after SARS-CoV-2 infection (Supplementary Fig. 5). Also, anosmia and dysgeusia are being considered as new typical symptoms of COVID-19 [39]. We also found that the bone-mineralization pathway was downregulated and a cartilage-development pathway was upregulated in convalescent patients, suggesting that osteoporosis is a potential sequela of COVID-19 (Supplementary Fig. 7).

The complex systemic damage and increase in mortality caused by COVID-19 have highlighted the importance of understanding COVID-19 outcomes. Among them, appraisal of COVID-19 severity is one of the important factors for judging outcomes. Some artificial intelligence (AI)-based models have been developed for predicting COVID-19 severity according to the characteristics of serum proteomics or clinical symptoms [40,41]. However, disease-related changes in gene expression occur earlier than changes in protein levels and clinical manifestations. Therefore, an assessment model based on gene-expression changes has the advantage of being more sensitive than a model based on changes in protein levels and clinical manifestations. Such a model can effectively supplement the deficiencies of existing prediction systems and can be used for primary screening. Based on the 16 genes serving as potential predictors involved in the dual-injury mechanism, we built a MLM with the potential to analyze COVID-19 severity (Fig. 4D). Although expression of these genes was derived from high-throughput sequencing, the calculation of relative expression with *GAPDH* as an internal reference was introduced in the modeling,

which made our MLM model compatible with the data derived from quantitative polymerase chain reactions. This MLM will reduce the cost of detection, and increase the convenience of detection operations, which is conducive to the popularization of AI-prediction models based on specific genes.

Another important opportunity for understanding COVID-19 outcomes is in convalescence. Identifying RTP patients earlier will help issuing of “immune passports” to prevent the re-spread of SARS-CoV-2. RTP patients have similar clinical manifestations and immune activation to that of COVID-19 patients in the convalescent period: obvious severe symptoms are absent [17,18]. Therefore, we did not analyze the differences in functional pathways, but focused on the changes in the diversity of BCRs and TCRs in RTP patients. Compared with RTP patients, we found that CPs had a longer V region and J region (e.g., V<sub>H</sub>3–9 and J<sub>H</sub>6) to achieve longer V–J pairing (Fig. 5B). SARS-CoV-2 is a highly glycosylated spherical particle whose degree of glycosylation is much higher than that of influenza viruses and the HIV, so many neutralizing antibodies cannot block SARS-CoV-2 from invading cells [42,43]. A neutralizing antibody with a longer CDR-H3 loops has been shown to bind effectively to the HIV [26], which suggests that an antibody with a longer V–J amino-acid sequence in the CP group may also have a better ability to bind to SARS-CoV-2. Also, the central region of CDRH3 of BCRs in CP patients had tyrosine residues instead of serine residues in the RTP group (Fig. 5E). These BCR differences not only helped to explain why the antibodies in the RTP group did not neutralize SARS-CoV-2 (Supplementary Figs. 8F and G), but also helped understanding of which type of antibodies may be effective. This information will provide new ideas for the screening of neutralizing antibodies and evaluation of vaccine-protection effects.

#### Author contributions

Yonggang Zhou: conceptualized the study, analyzed and interpreted the data, and wrote the manuscript. Jinhe Zhang: carried out the experiments and assisted with interpretation of transcriptomic data. Dongyao Wang: carried out the experiments and assisted with interpretation of transcriptomic data. Dong Wang: carried out the experiments and assisted with interpretation of transcriptomic data. Wuxiang Guan: undertook the experiments and collected whole-blood samples and patient information. Jingkun Qin: assisted with data interpretation of immune-receptor repertoire. Xiuxiu Xu: assisted with data interpretation of immune-receptor repertoire. Jingwen Fang: assisted with data interpretation of immune-receptor repertoire. Binqing Fu: undertook the experiments and collected whole-blood samples and patient information. Xiaohu Zheng: undertook the experiments and collected whole-blood samples and patient information. Dongsheng Wang: undertook the experiments and collected whole-blood samples and patient information. Hong Zhao: undertook the experiments and collected whole-blood samples and patient information. Xianxiang Chen: undertook the experiments and collected whole-blood samples and patient information. Zhigang Tian: provided strategic planning and interpreted some data. Xiaoling Xu: provided strategic planning and interpreted some data. Guiqiang Wang: conceptualized the study, analyzed and interpreted the data, and wrote the manuscript. Haiming Wei: conceptualized the study, analyzed and interpreted the data, and wrote the manuscript.

#### Declaration of competing interest

The authors declare no competing interests.

#### Acknowledgments

This work was supported by China National Center for Biotechnology Development (2020YFC0843800 to H.W. and 2020YFC0846800 to G.W.) and Jack Ma Foundation (ZB910000001 to H.W.).

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jaut.2021.102596>.

#### Data and materials availability

All data and codes to understand and assess the conclusions of this research are available in the main text and supplementary material. Information on resources and reagents are available from Haiming Wei under a material agreement with University of Science and Technology of China.

(A) A model diagram depicting the COVID-19 course at full-scale. NAT “+/-” denotes that the nucleic-acid test result of SARS-CoV-2 is positive (“+”) or negative (“-“). (B) Overview of the experimental design and number of samples for RNA-seq, BCR-seq and TCR-seq. Healthy control (n = 9): healthy people not infected with SARS-CoV-2. Non-ICU patients (n = 16): patients with mild COVID-19 admitted to the general ward. ICU patients (n = 6): patients with severe COVID-19 admitted to the intensive care unit (ICU). CP patients (n = 6): convalescent patients with a negative viral nucleic-acid re-test result. RTP patients (n = 10): convalescing patients with a positive viral nucleic-acid re-test result. (C) The proportion of PC1-10 in the PCA of transcriptome data from the 47 patients in the five groups. (D) PCA of transcriptome data from the five groups based on PC1 and PC2. (E and F) Correlation analyses between PC1(E) or PC2(F) genes and patients in the five groups.

(A, B, C) Correlation analyses between PC1 and GSVA enrichment score of proinflammatory pathways (A), anti-inflammatory pathways (B) and antiviral immune-response pathway (C). (D) Dynamics of the degree of GSVA enrichment of proinflammatory-related pathways (left), anti-inflammatory-related pathways (middle) and a combination of the two (right) in the five groups of healthy controls, non-ICU, ICU, CP and RTP. (E) Bubble chart showing the dynamic difference in expression of featured genes in the proinflammatory and anti-inflammatory pathways enriched by GSVA in the five groups. “# of samples > mean (%)” show the proportion of samples whose expression is greater than the mean. TPM, transcripts per kilobase million.

(A) Heatmap of GSVA-enrichment degree of oxygen saturation-, peroxide-, hypercoagulable-, and vascular injury-related pathways in each sample of the five groups. (B) Dynamics of the GSVA-enrichment degree of promoted vascular injury-related pathways (left), anti-vascular injury-related pathways (middle) and a combination of the two (right) in the five groups of healthy controls, non-ICU, ICU, CP and RTP. (C) Bubble chart showing the dynamic difference in expression of featured genes in the promoted vascular injury- and anti-vascular injury-pathways enriched by GSVA in the five groups. “# of samples > mean (%)” show the proportion of samples whose expression is greater than the mean. TPM, transcripts per kilobase million.

(A) Study design of a machine-learning model for screening genes associated with COVID-19 severity based on transcriptome data from non-ICU and ICU patients with COVID-19 and a logistic regression algorithm. (B) The top-16 featured genes of COVID-19 severity calculated by the machine-learning model. (C) Performance of the machine-learning model in a test cohort of 10 RTP patients with COVID-19. (D) Performance of the machine-learning model in a test cohort of dynamic transcriptomic data of PBMCs from three COVID-19 cases in the array express database (E-MTAB-8871).

(A and B) Differences in IGHV (A) and IGHJ (B) selection between the CP group and RTP group. (C) Differential dominance in IGHV-IGHJ pairs between the CP group and RTP group. (D) Difference in statistical parameters of IGHV-IGHJ pairs between the CP group and RTP group. (E) Differential dominance in components between the CP group and RTP group. (F) The CDRH3 motif was the most significant difference in the RTP group and CP group. (G and H) Structural simulation (G) and predictive binding force statistics (H) of the simulated neutralizing antibody competing with the S protein and ACE2 protein complex.

(A) Differential dominance in TRBV-TRBJ pairs between healthy controls and convalescent patients. (B) Differential dominance in TRBV-TRBJ pairs between the CP group and RTP group. (C) Upset chart of common components among the five groups. (D) The potential dynamic changes in the SARS-CoV-2-specific TCR sequence in the five groups of COVID-19 patients.

## References

- [1] W. H. Organization, World Health Statistics, Cause-specific Mortality and Morbidity, World Health Organization, 2009, pp. 47–57.
- [2] B.S. Graham, N.J. Sullivan, Emerging viral diseases from a vaccinology perspective: preparing for the next pandemic, *Nat. Immunol.* 19 (2018) 20–28.
- [3] C. Drosten, et al., Identification of a novel coronavirus in patients with severe acute respiratory syndrome, *N. Engl. J. Med.* 348 (2003) 1967–1976.
- [4] A.M. Zaki, S. van Boheemen, T.M. Bestebroer, A.D. Osterhaus, R.A. Fouchier, Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia, *N. Engl. J. Med.* 367 (2012) 1814–1820.
- [5] C. Wang, P.W. Horby, F.G. Hayden, G.F. Gao, A novel coronavirus outbreak of global health concern, *Lancet* 395 (2020) 470–473.
- [6] WHO, Coronavirus disease (COVID-19), Situation report– 161. [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200629-covid-19-sitre-p-161.pdf?sfvrsn=74fde64e\\_2](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200629-covid-19-sitre-p-161.pdf?sfvrsn=74fde64e_2). (Accessed 29 June 2020).
- [7] E. Petersen, et al., Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics, *Lancet Infect. Dis.* 9 (2020) e238–e244, [https://doi.org/10.1016/S1473-3099\(20\)30484-9](https://doi.org/10.1016/S1473-3099(20)30484-9).
- [8] W.J. Wiersinga, A. Rhodes, A.C. Cheng, S.J. Peacock, H.C. Prescott, Pathophysiology, transmission, diagnosis, and treatment of Coronavirus disease 2019 (COVID-19): a review, *JAMA* 8 (2020) 782–793, <https://doi.org/10.1001/jama.2020.12839>.
- [9] P. Yang, X. Wang, COVID-19: a new challenge for human beings, *Cell. Mol. Immunol.* 17 (2020) 555–557.
- [10] Y.J. Hou, et al., SARS-CoV-2 reverse genetics reveals a variable infection gradient in the respiratory tract, *Cell* 2 (2020) 429–446.e14, <https://doi.org/10.1016/j.cell.2020.05.042>.
- [11] Y. Zhou, et al., Pathogenic T-cells and inflammatory monocytes incite inflammatory storms in severe COVID-19 patients, *Natl. Sci. Rev.* 7 (2020) 998–1002.
- [12] C. Huang, et al., Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, *Lancet* 395 (2020) 497–506.
- [13] P. Mehta, et al., COVID-19: consider cytokine storm syndromes and immunosuppression, *Lancet* 395 (2020) 1033–1034.
- [14] F.M. Lang, K.M. Lee, J.R. Tejjaro, B. Becher, J.A. Hamilton, GM-CSF-based treatments in COVID-19: reconciling opposing therapeutic approaches, *Nat. Rev. Immunol.* 20 (2020) 507–514.
- [15] X. Xu, et al., Effective treatment of severe COVID-19 patients with tocilizumab, *Proc. Natl. Acad. Sci. U. S. A.* 117 (2020) 10970–10975.
- [16] G. De Luca, et al., GM-CSF blockade with mavrilimumab in severe COVID-19 pneumonia and systemic hyperinflammation: a single-centre, prospective cohort study, *Lancet Rheumatol.* 8 (2020) e465–e473, [https://doi.org/10.1016/S2665-9913\(20\)30170-3](https://doi.org/10.1016/S2665-9913(20)30170-3).
- [17] L. Lan, et al., Positive RT-PCR test results in patients recovered from COVID-19, *Jama* 323 (2020) 1502–1503.
- [18] J. An, et al., Clinical characteristics of the recovered COVID-19 patients with re-detectable positive RNA test, *Ann Transl Med.* 17 (2020) 1084–1503, <https://doi.org/10.21037/atm-20-5602>.
- [19] Q. Mei, et al., Assessment of patients who tested positive for COVID-19 after recovery, *Lancet Infect. Dis.* 9 (2020) 1004–1005, [https://doi.org/10.1016/S1473-3099\(20\)30433-3](https://doi.org/10.1016/S1473-3099(20)30433-3).
- [20] E.Z. Ong, et al., A dynamic immune response shapes COVID-19 progression, *Cell Host Microbe* 27 (2020) 879–882.
- [21] L. Zhu, et al., Single-cell sequencing of peripheral mononuclear cells reveals distinct immune response landscapes of COVID-19 and influenza patients, *Immunity* 53 (2020) 685–696.
- [22] L. Zhu, et al., Single-cell sequencing of peripheral blood mononuclear cells reveals distinct immune response landscapes of COVID-19 and influenza patients, *Immunity* 3 (2020) 685–696.e3, <https://doi.org/10.1016/j.immuni.2020.07.009>.
- [23] X.-W. Bian, The COVID-19 Pathology Team. Autopsy of COVID-19 victims in China, *Natl. Sci. Rev.* 7 (2020) 1414–1418, <https://doi.org/10.1093/nsr/nwaa123>.
- [24] L. Carsana, et al., Pulmonary post-mortem findings in a series of COVID-19 cases from northern Italy: a two-centre descriptive study, *Lancet Infect. Dis.* 10 (2020) 1135–1140, [https://doi.org/10.1016/S1473-3099\(20\)30434-5](https://doi.org/10.1016/S1473-3099(20)30434-5).
- [25] Y. Wang, et al., Organizing pneumonia of COVID-19: time-dependent evolution and outcome in CT findings, *medRxiv* (2020), <https://doi.org/10.1101/2020.05.22.20109934>.
- [26] J.H. Lee, et al., A broadly neutralizing antibody targets the dynamic HIV envelope trimer apex via a long, rigidified, and anionic  $\beta$ -hairpin structure, *Immunity* 46 (2017) 690–702.
- [27] A. Madsen, et al., Human antibodies targeting influenza B virus neuraminidase active site are broadly protective, *Immunity* 53 (2020) 852–863.
- [28] Y. Cao, et al., Potent neutralizing antibodies against SARS-CoV-2 identified by high-throughput single-cell sequencing of convalescent patients' B cells, *Cell* 1 (2020) 73–84.e16, <https://doi.org/10.1016/j.cell.2020.05.025>.
- [29] Z. Chen, E. John Wherry, T cell responses in patients with COVID-19, *Nat. Rev. Immunol.* 9 (2020) 529–536, <https://doi.org/10.1038/s41577-020-0402-6>.
- [30] M. Leslie, T cells found in coronavirus patients 'bode well' for long-term immunity, *Science* 368 (2020) 809–810.
- [31] M. Zheng, et al., Functional exhaustion of antiviral lymphocytes in COVID-19 patients, *Cell. Mol. Immunol.* 17 (2020) 533–535.
- [32] H.Y. Zheng, et al., Elevated exhaustion levels and reduced functional diversity of T cells in peripheral blood may predict severe progression in COVID-19 patients, *Cell. Mol. Immunol.* 17 (2020) 541–543.
- [33] H. Ma, et al., Serum IgA, IgM, and IgG responses in COVID-19, *Cell. Mol. Immunol.* 17 (2020) 773–775.
- [34] B.K. Manne, et al., Platelet gene expression and function in patients with COVID-19, *Blood* 11 (2020) 1317–1329, <https://doi.org/10.1182/blood.2020007214>.
- [35] G. Goshua, et al., Endotheliopathy in COVID-19-associated coagulopathy: evidence from a single-centre, cross-sectional study, *Lancet Haematol* 7 (2020) e575–e582.
- [36] A.A. Barkal, et al., CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy, *Nature* 572 (2019) 392–396.
- [37] R.R. Tian, et al., CD24Fc protects against viral pneumonia in simian immunodeficiency virus-infected Chinese rhesus monkeys, *Cell. Mol. Immunol.* 17 (2020) 887–888.
- [38] A. Gupta, et al., Extrapulmonary manifestations of COVID-19, *Nat. Med.* 26 (2020) 1017–1032.
- [39] C. Menni, C.H. Sudre, C.J. Steves, S. Ourselin, T.D. Spector, Quantifying additional COVID-19 symptoms will save lives, *Lancet* 395 (2020) e107–e108.
- [40] B. Shen, et al., Proteomic and metabolomic characterization of COVID-19 patient sera, *Cell* 182 (2020) 59–72.
- [41] W. Liang, et al., Early triage of critically ill COVID-19 patients using deep learning, *Nat. Commun.* 11 (2020) 3543.
- [42] Y. Watanabe, J.D. Allen, D. Wrapp, J.S. McLellan, M. Crispin, Site-specific glycan analysis of the SARS-CoV-2 spike, *Science* 369 (2020) 330–333.
- [43] C. Yi, et al., Key residues of the receptor binding motif in the spike protein of SARS-CoV-2 that interact with ACE2 and neutralizing antibodies, *Cell. Mol. Immunol.* 17 (2020) 621–630.