Computational and Structural Biotechnology Journal 21 (2023) 1362–1371

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

CoV2-TCR: A web server for screening TCR CDR3 from TCR immune repertoire of COVID-19 patients and their recognized SARS-CoV-2 epitopes

Xingxing Jian [a,*,1], Yu Zhang [b,c,1], Jingjing Zhao [b,d,1], Zhuoming Zhao [b,d], Manman Lu [b], Lu Xie [a,b,**]

[a] Bioinformatics Center & National Clinical Research Centre for Geriatric Disorders & Department of Geriatrics, Xiangya Hospital, Central South University, Changsha, Hunan, China
[b] Shanghai-MOST Key Laboratory of Health and Disease Genomics (Chinese National Human Genome Center at Shanghai), Institute of Genome and Bioinformatics, Shanghai Institute for Biomedical and Pharmaceutical Technologies, Shanghai, China
[c] School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai, China
[d] College of Food Science and Technology, Shanghai Ocean University, Shanghai, China

ARTICLE INFO

Article history:
Received 25 September 2022
Received in revised form 8 January 2023
Accepted 26 January 2023
Available online 27 January 2023

Keywords:
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
Coronavirus disease 2019 (COVID-19)
T cell receptor (TCR)
Complementary determining regions 3 (CDR3)
Antigen epitope
Immunogenic peptide
CoV2-TCR

ABSTRACT

Although multiple vaccines have been developed and widely administered, several severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants have been reported to evade immune responses and spread diffusely. Here, 108 RNA-seq files from coronavirus disease 2019 (COVID-19) patients and healthy donors (HD) were downloaded to extract their TCR immune repertoire by MiXCR. Those extracted TCR repertoire were compared and it was found that disease progression was related negatively with diversity and positively with clonality. Specifically, greater proportions of high-abundance clonotypes were observed in active and severe COVID-19 samples, probably resulting from strong stimulation of SARS-CoV-2 epitopes and a continued immune response in host. To investigate the specific recognition between TCR CDR3 and SARS-CoV-2 epitopes, we constructed an accurate classifier CoV2-TCR with an AUC of 0.967 in an independent dataset, which outperformed several similar tools. Based on this model, we observed a huge range in the number of those TCR CDR3 recognizing those different peptides, including 28 MHC-I epitopes from SARS-CoV-2 and 22 immunogenic peptides from SARS-CoV-2 variants. Interestingly, their proportions of high-abundance, low-abundance and rare clonotypes were close for each peptide. To expand the potential application of this model, we established the webserver, CoV2-TCR, in which users can obtain those recognizing CDR3 sequences from the TCR repertoire of COVID-19 patients based on the 9-mer peptides containing mutation site(s) on the four main proteins of SARS-CoV-2 variants. Overall, this study provides preliminary screening for candidate antigen epitopes and the TCR CDR3 that recognizes them, and should be helpful for vaccine design on SARS-CoV-2 variants.

© 2023 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Since late 2019, the emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) rapidly evolved into a pandemic disease, officially named coronavirus disease 2019 (COVID-19) by the World Health Organization. To date, it has been broken out in more than 200 countries or regions worldwide, resulting in more than 534.5 million confirmed cases and more than 6.3 million deaths (Accessed on 16 June 2022, https://covid19.who.int). Critically, although multiple vaccines have been developed and widely

* Corresponding author.
** Corresponding author at: Bioinformatics Center & National Clinical Research Centre for Geriatric Disorders & Department of Geriatrics, Xiangya Hospital, Central South University, Changsha, Hunan, China.
E-mail addresses: jianxingxing@foxmail.com (X. Jian), xielu@sibpt.com (L. Xie).
[1] These authors equally contributed to this study.

https://doi.org/10.1016/j.csbj.2023.01.038
2001-0370/© 2023 The Authors. Published by Elsevier B.V. on behalf of Research Network of Computational and Structural Biotechnology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

administered, several SARS-CoV-2 variants derived from mutations in the major histocompatibility complex class I (MHC-I)-restricted epitopes have been reported to evade immune response [1]. Wang et al. investigated the susceptibility to immune evasion of the SARS-CoV-2-specific antibody in COVID-19 convalescents for the 2 years following discharge and showed that Omicron variants largely escaped any preexisting immunity in recovered individuals [2]. In addition, developing new and effective vaccine was recommended as a strategy to limit transmission of Omicron and emerging variants [3].

As with any viral infection, pathogen-derived antigens peptides (8–11-mer) are presented together with MHC-I (HLA-I in human) molecules on the surface of infected cells [4]. Then, the exhibited peptide-MHC complex (pMHC) is specifically recognized by CD8 + T cells *via* T cell receptors (TCRs), leading to the eradication of infected cells, where naive T cells are induced for clonal expansion of various subsets with identical TCRs [5]. In addition, TCRs on circulating T cells are mostly heterodimers with alpha (α) and beta (β) subunits, and the hypervariable TCR repertoire mainly results from the highly diverse complementary determining regions (CDRs) shaped by VDJ genes recombination localized in the TCR α and β chains [6]. Notably, the third CDR (CDR3) in the TCR β chain is responsible for the recognition of the antigen peptide in pMHC. Thus, those diverse CDR3 in the TCRβ repertoire of COVID-19 patients play a key role in the clearance of SARS-CoV-2.

Reportedly, SARS-CoV-2 infection activates the host's adaptive immune system against virus, including both antibody immunity produced by B cells and T cell immunity delivered by $CD4^+$ T cells as the helper T cells with $CD8^+$ T cells as the cytotoxic cells [7–10]. Therein, neutralizing antibodies can prevent the binding and entry of virus. As a supplement to antibodies, $CD8^+$ T cells prevent effectively the establishment of infection by clearing the infected cells after exposure [11], especially tissue-resident memory T cells [12]. Grau-Expósito et al [13]. showed the response of peripheral and lung resident memory T cell against SARS-CoV-2 during acute infection. Moreover, the clinical data from severe COVID-19 patients showed that the $CD8^+$ T cells were significantly higher in those surviving patients [14]. A higher proportion of SARS-CoV-2 specific $CD8^+$ T cells were observed in mild cases and these $CD8^+$ T cells have extensive and strong memory after the recovery period of COVID-19 [15]. Terminally differentiated $CD8^+$ $GZMB^+$ effector cells were clonally expanded both during and after the infection, whereas $CD8^+$ $GZMK^+$ lymphocytes were further expanded after infection and represented bona fide memory precursor effector cells [16]. Thus, the anti-SARS-CoV-2 T-cell-mediated immune response could be crucial for immune memory against SARS-CoV-2 infection.

In addition, there is evidence that the SARS-CoV-2 specific T cells were found in both COVID-19 patients and unexposed healthy donors [7,17,18]. This suggests that memory T cells recognizing the SARS-CoV-2 variants may have already existed in human beings. These existing memory T cells may result from previous cross-reactivity of T cells primed by seasonal coronaviruses, such as SARS or influenza [17]. Therefore, we hypothesized, in TCR repertoire of COVID-19 patients, there existed TCR CDR3 sequences that can recognize specific epitopes from any SARS-CoV-2 variants.

Previously, we have constructed a prediction tool of immunogenic neoantigens, namely DeepCNN-Ineo, aimed at screening those tumor neoantigens bound by HLA molecular and recognized by TCR sequences [19]. In this study, we obtained 108 RNA-seq files of peripheral blood, including 69 samples from COVID-19 patients and 39 samples from HD, to extract their TCRβ immune repertoire, which was found to correlate with the disease progression of COVID-19 samples. Furthermore, an accurate model configured in an easy-to-use web server, called CoV2-TCR, was constructed and implemented to investigate those TCR CDR3 recognizing epitopes from SARS-CoV-2 or its variants. As a result, those epitopes recognized by most TCR CDR3 sequences may be selected as candidate antigen peptides for vaccine design, while the number of these recognizing CDR3 sequences may reflect, to some extent, the protective coverage of the vaccine.

2. Materials and methods

2.1. Samples and data collection

In this study, 108 RNA-seq files of peripheral blood samples were downloaded from European Nucleotide Archive database with the search ID PRJEB38339, including 69 COVID-19 samples from 37 SARS-CoV-2 infected patients and 39 samples from 19 healthy donors (HD). Referring to a previous study on PRJEB38339 [20], the 69 COVID-19 samples can be divided into active and recovered groups according to the disease progression and further assigned into asymptomatic, mild, moderate, severe, and fatal groups according to the severity observed in patients. Of note, as described in the data source publication [20], the severity categories 'asymptomatic' and 'mild' are almost entirely included within the disease progression category 'recovered', while the 'moderate', 'severe' and 'fatal' are almost entirely included within the 'active' disease group.

2.2. Extraction and analysis of TCR immune repertoire

Based on the RNA-seq files (fastq) obtained from all samples, MiXCR software (copyright 2018, MiLaboratory, LLC Revision 225a025b) was used to extract TCR immune repertoire of each sample [21], in which those columns include cloneID, cloneCount, cloneFraction, allVHitsWithScore, allDHitsWithScorec, allJHitsWithScorec, nSeqCDR3, aaSeqCDR3 and *etc.* Since the CDR3 in the TCRβ chain is responsible for the recognition of the antigen peptide in pMHC, only TCRβ CDR3 (TRB) was utilized. That is, the unique TCRβ CDR3 sequences were treated as clonotypes in this study. On the R (version 4.1.2) platform, R package immunarch (version 0.6.7) were applied for analysis of TCR immune repertoire, including diversity, clonotype distribution and tracking.

2.3. Construction of deep learning convolution neural network model

We constructed a deep learning convolution neural network model to explore the specific recognition pairs between TCR CDR3 sequences and SARS-CoV-2 epitopes (Fig. S1). The pairs between SARS-CoV-2 antigen epitopes and TCR CDR3 amino-acid sequences were first encoded into a feature matrix by using one-hot encoding. Here, we focused on those HLA-I antigens epitopes and TCR CDR3 sequences with the length of (8–11)-mer and (~25)-mer, respectively. Thus, the paired epitope-CDR3 sequences were encoded and zero-padded as a matrix with dimension of 36×20.

The epitope-CDR3 pairs were encoded to obtain the feature matrix, and two 1D convolutional layers were applied. Specifically, one layer with 32 filters and another with 64, was followed by a batch normalization layer with momentum of 0.99. Next, a max pooling layer with pooling kernel size of 2 was applied, aiming to reduce the feature dimension and avoid overfitting. There were three fully connected layers with 128, 64 and 32 variables following the flattened layer. Each of the fully connected layers had the rectified linear unit (ReLU) activation function. Subsequently, before the last output layer, a dropout layer with probabilities of 0.3 was added for randomly discarding variables in the hidden layer. Eventually, an output layer with two variables and the Softmax activation function was adopted to represent the final statistical classification score, which was used to judge the specific recognition of peptides by TCR CDR3. Here, we regarded the predictive score of more than 0.95 to be valid recognition.

An adagrad optimizer with a learning rate of 0.005 was used to automatically decay learning rate by using the number of iterations and accumulated gradients. MSE was chosen as the objective function. The neural network was trained for 100 epochs and the batch size was set to 20.

We constructed and trained this model using TensorFlow 1.14.0 and Python (3.7.6) packages Numpy (version 1.19.5), Pandas (version 0.25.3), Biopython (version 1.76), scikit-learn (version 0.24.2), and scipy (version 1.4.1).

2.4. The datasets for model construction and validation

We first searched those recognized pairs between SARS-CoV-2 epitopes and TCR CDR3 sequences in IEDB [22] and VDJdb [23]. And, 11,388 and 381 recognized pairs were downloaded from IEDB and VDJdb, respectively. These recognized pairs have been verified experimentally, and were therefore regarded as positive pairs for model construction and validation.

In addition, in IEDB and TCRdb [24], we also obtained 22,805 TCR CDR3 sequences derived from seven types of tumors, *i.e.* breast cancer, colorectal cancer, gastric cancer, glioblastoma, lung cancer, melanoma, ovarian cancer. In IEDB, we also downloaded 50,639 recognized pairs between EBV epitopes and TCR CDR3 sequences, and 408 recognized pairs between HIV epitopes and TCR CDR3 sequences. Additionally, 1313 TCR CDR3 sequences derived from healthy people were obtained as well. We hypothesized that these TCR CDR3 sequences completely unrelated to SARS-CoV-2 epitopes should not specifically recognize SARS-CoV-2 epitopes. In order to construct negative pairs, these TCR CDR3 non-induced from SARS-CoV-2 epitopes were randomly paired with those SARS-CoV-2 epitopes from the positive set above.

We first merged the positive pairs downloaded from IEDB and negative pairs derived from tumors and EBV, and then removed duplicated pairs and those pairs containing unusual amino acids (*e.g.* B, J, O, U, X, Z). Further, we only retained those pairs involving (8–11)-mer epitopes and (~25)-mer CDR3. In total, 50,949 pairs were obtained for model construction, including 11,056 positive pairs and 39,893 negative pairs. Subsequently, with the ratio of 8:2, these pairs were randomly divided into training dataset (40,759 pairs) and test dataset (10,190 pairs).

Similarly, we also merged an independent validation dataset for model validation, including those positive pairs downloaded from TCRdb and those negative pairs derived from HIV and healthy people. After carrying out the same filtration and selection as above, our validation dataset totally contained 880 pairs, including 334 positive pairs and 546 negative pairs.

2.5. Screening of immunogenic peptides in SARS-CoV-2 variants

In 2019nCoVR (https://ngdc.cncb.ac.cn/ncov/knowledge/compare) [25–28], we obtained those amino-acid mutations exhibiting a frequency of greater than 0.7 in the four main proteins of the seven SARS-CoV-2 variants, *i.e.* alpha, beta, delta, gamma, lambda, mu, omicron, respectively. Then, based on those mutation sites and the reference amino-acid sequences of the four main proteins in SARS-CoV-2, the (8–11)-mer peptides containing mutation site(s) were extracted for subsequent immunogenicity prediction. Meanwhile, netMHCpan 4.1 EL [29] and IEDB Analysis Resource (Class I Immunogenicity) [22] were respectively applied, in which the 12 HLA types with high proportion in the world were used, namely HLA-A* 01:01, HLA-A* 02:01, HLA-A* 03:01, HLA-A* 11:01, HLA-A* 23:01, HLA-A* 24:02, HLA-B* 07:02, HLA-B* 08:01, HLA-B* 35:01, HLA-B* 40:01, HLA-B* 44:02, HLA-B* 44:03. Those peptides with %rank of less than 0.5 and IEDB immunogenicity score of greater than 0.15 were considered as strong immunogenicity. And, we screened 22 predicted immunogenic peptides (Table 2).

2.6. Establishment of online web tool

Based on the model CoV2-TCR constructed above, we established a user-friendly web server. This tool required users to select a specific protein from SARS-CoV-2, *i.e.* spike glycoprotein (S), membrane protein (M), envelop protein (E), or nucleocapsid protein (N), and its specified mutation amino-acid site. After submitting, users would be provided with those TCR CDR3 sequences that can recognize the related mutant peptides.

To implement an interaction interface and improve user experience, the front-end web framework was built by Vue.js and Element-UI, and the back-end was realized by Flask.

3. Results

3.1. Comparison of diversity, richness and clonality of the TCR immune repertoire

In this study, 108 RNA-seq files of peripheral blood, including 69 COVID-19 samples and 39 HD samples, were individually used to extract TCR immune repertoire. Subsequently, for each sample, the Chao1 index and Shannon index were separately evaluated to measure the richness and diversity of TCR immune repertoire. Here, the general richness and diversity in COVID-19 samples were observed with no significant difference compared to that in HC (Fig. 1a, d). However, when dividing those COVID-19 samples into active and recovered groups according to the disease progression, the richness and diversity of those active samples were significantly lower, while those recovered patients were significantly higher values (Fig. 1b, e). Further, according to the disease severity observed in the COVID-19 patients, a significant decrease in richness and diversity was observed in mild, moderate, and severe groups respectively (Fig. 1c, f). Meanwhile, we observed that those mild samples presented higher richness and diversity than those of HD, and no significant difference between those of severe and fatal samples. Thus, we considered that the richness and diversity of TCR immune repertoire in COVID-19 patients could be used to negatively indicate the disease progression and severity degree. Noteworthy, as described in the Method, the severity categories 'asymptomatic' and 'mild' are almost entirely included within the disease progression category 'recovered', while the 'moderate', 'severe' and 'fatal' are almost entirely included within the 'active' disease group.

In addition, the clonality from each sample was separately estimated as well. As expected, there was a significant increase in clonality in COVID-19 samples due to the SARS-CoV-2 infection (Fig. 1g). In particular, those active patients showed a remarkably higher clonality compared to that in HC, while those recovered patients presented a significantly lower clonality (Fig. 1h). Consistently, we observed that a significantly increased clonality was detected in those patients with mild, moderate, and severe COVID-19 samples respectively (Fig. 1i), suggesting that the clonality of TCR immune repertoire could be used to positively indicate the disease progression and severity degree. Taken together, SARS-CoV-2 infection resulted in clonal expansion in TCR immune repertoire of COVID-19 patients, which is positively correlated with the disease severity of the patients.

3.2. Investigation of clonotypes in patients with COVID-19 infection

The immune system is central to maintaining the body in a healthy state by detecting and evicting those constantly exposed threats, including various pathogens and cancer cells. In general, the recognition of antigens by T cells activates their proliferation and/or phenotypic changes, leading to alteration of TCR repertoire [30,31]. Here, a total of 209,428 and 398,033 unique clonotypes was respectively detected in HD and COVID-19 samples, of which 17,728

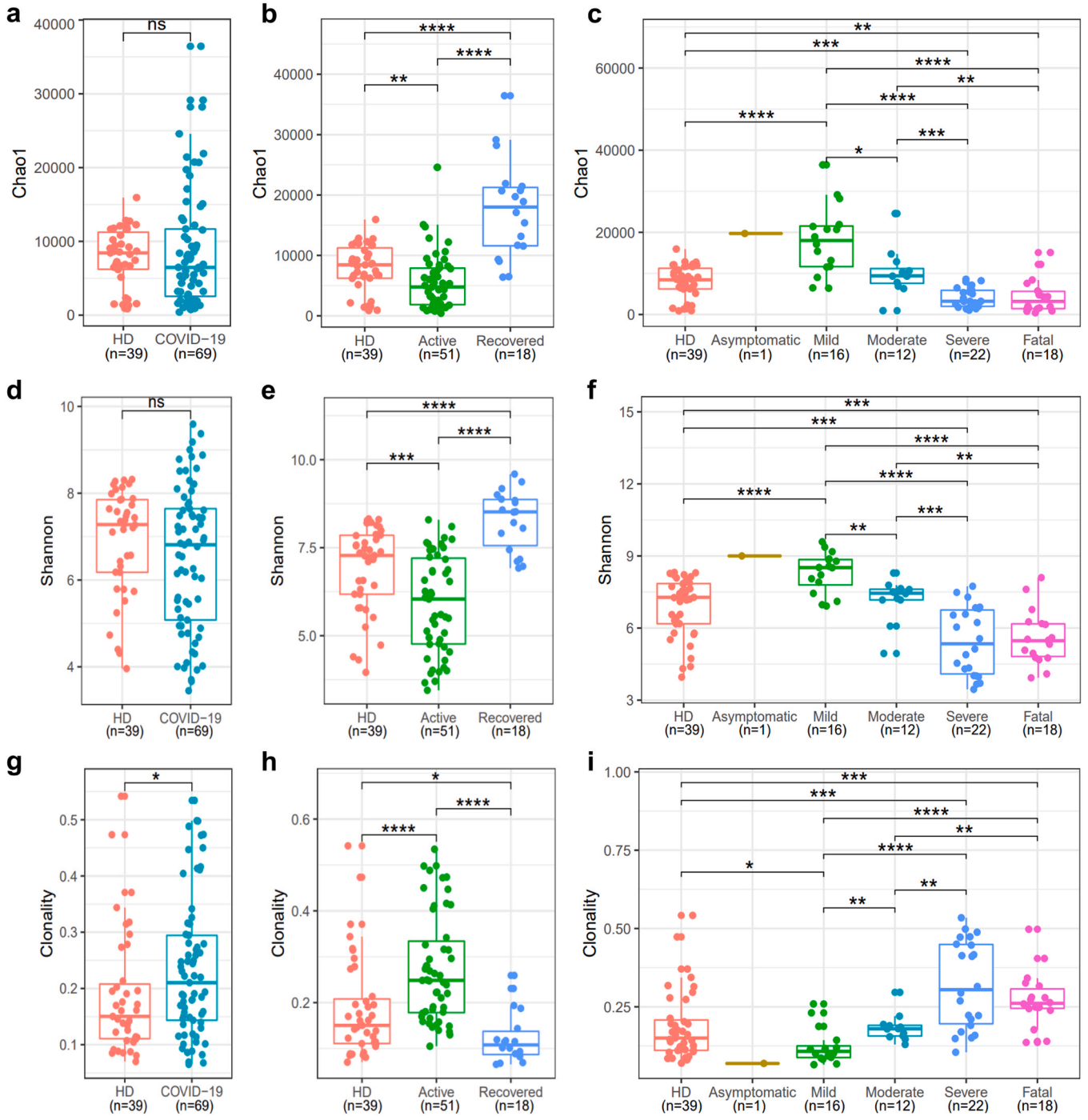

Fig. 1. Comparison of TCR immune repertoire. a, d, g) Comparison of Chao1 (a), Shannon (d), and Clonality (g) between HD and COVID-19 samples. b, e, h) Comparison of Chao1 (b), Shannon (e), and Clonality (h) between HD, active COVID-19 samples, and recovered COVID-19 samples. c, f, i) Comparison of Chao1 (c), Shannon (f), and Clonality (i) between HD and COVID-19 samples in different severity degree.

clonotypes were shared (Fig. 2a). Comparatively, those unshared clonotypes from COVID-19 samples (380,305) were hypothesized to be driven and produced by SARS-CoV-2 epitopes. Therefore, when assigning those COVID-19 samples into active and recovered groups according to disease progression, we found that fewer unshared clonotypes were obtained from those active samples than those from recovered samples (Fig. 2b). Similarly, fewer unshared clonotypes were obtained from those fatal/severe samples than from those mild/moderate samples (Fig. 2c). These observations were consistent with the low diversity in active samples and in severe/fatal samples described above, resulting from the strong stimulus of SARS-CoV-2 epitopes.

Next, to investigate those biased clonotypes derived from SARS-CoV-2 infection in COVID-19 samples, the distribution between the number of clonotypes and the number of cells per clonotype was further analyzed. As shown in Fig. 2d, the number of clonotypes decreased gradually with the increase of the number of cells characterized by a certain clonotype in the immune repertoire. That is to say, as a result of continued stimulation by SARS-CoV-2 antigens, more characteristic clonotypes were observed to appear in COVID-19

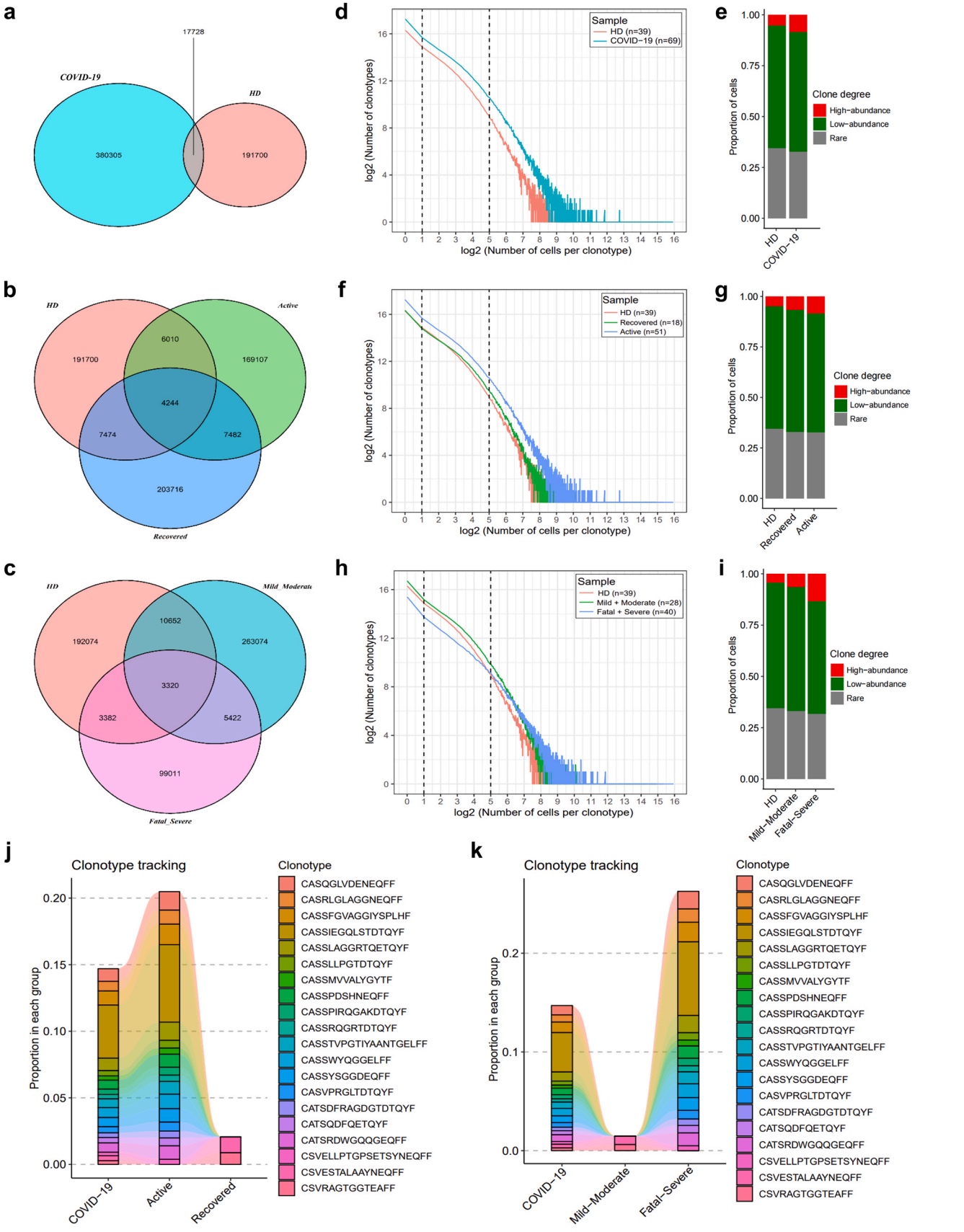

Fig. 2. Investigation of TCR immune repertoire. a, d, e) Venn (a), clonotypes distribution (d), and clone degree (e) of TCR repertoire between HD and COVID-19 samples. b, f, g) Venn (b), clonotypes distribution (f), and clone degree (g) of TCR repertoire between HD, active COVID-19 samples, and recovered COVID-19 samples. c, h, i) Venn (c), clonotypes distribution (h), and clone degree (i) of TCR repertoire between HD and COVID-19 samples in different severity degree. (j) Clone tracking of top 20 high-abundance clonotypes in COVID-19 patients with the different disease progression. (k) Clone tracking of top 20 high-abundance clonotypes in COVID-19 patients with the different severity degree.

samples. To compare them, we characterized those clonotypes with the number of T cells (*i.e.* abundance) less than 2^1, less than 2^5, and greater than 2^5 as rare, low-abundance, and high-abundance, respectively. Those high-abundance clonotypes were considered to result from the strong stimulus from SARS-CoV-2 antigens. Therein, high-abundance clonotypes accounted for 4.42% of the total in HD samples, while a higher proportion (8.75%) was detected in COVID-19 samples (Fig. 2e). Consistently, the higher proportion of high-abundance clonotypes was found in active patients than that in recovered patients (Fig. 2f, g), and the higher proportion of high-abundance T cells was found in fatal/severe samples than that in mild/moderate samples (Fig. 2h, i). These observations echoed the fewer clonotypes in active samples and in severe/fatal samples, also resulting from host's immune response to SARS-CoV-2 epitopes.

In addition, we further focused on those 380,305 unique clonotypes unshared with 39 HD samples, in which the number of high-abundance, low-abundance, and rare clonotypes was 37,562, 236,182, and 106,561, respectively. Then, the top 20 high-abundance clonotypes were singled out for subsequent clonotype tracking. As shown in Fig. 2j, k, we tracked their proportions at each group, illustrating their changes in immune repertoire of the COVID-19 samples at different infected periods. Consistent with the observation above, the sum proportion of recovered samples and mild/moderate samples were both lower. Notably, some sequences presented a high proportion in active samples and fatal/severe samples, while showed little proportion in recovered samples and mild/moderate samples, *e.g* CASSIEGQLSTDTQYF, CASSFGVAGGIYSPLHF, CASRLGLAGGNEQFF, and CASQGLVDENEQFF. In addition, we also found that the proportion of CSVESTALAAYNEQFF and CSVRAGTGGTEAFF presented little in active samples and fatal/severe samples but presented a high proportion in recovered samples and mild/moderate samples. That indicated that the high-abundance clonotypes may be associated with the severity degree and outcomes in different patients.

3.3. Construction of a classifier to identify TCR CDR3 with their recognizing SARS-CoV-2 epitopes

Those experimentally-verified recognition pairs between SARS-CoV-2 epitopes and TCR CDR3 sequences were downloaded from IEDB as positive pairs [22]. To obtain negative pairs, those TCR CDR3 sequences derived from seven tumors and EVB were collected from IEDB and TCRdb [24], and randomly paired with those SARS-CoV-2 epitopes of the positive pairs. Subsequently, the positive pairs and the negative pairs were merged and filtered, a total of 50,949 pairs were obtained, and then randomly sampled them as training dataset and test dataset according to the ratio of 8:2 for model construction.

Based on the training dataset and test dataset above, a deep learning convolution neural network model, named CoV2-TCR, was constructed (Fig. S1). Those epitope-CDR3 pairs were first encoded by using one-hot and zero-padded into a feature matrix, and then two 1D convolutional layers were applied for batch normalization. Next, a max pooling layer with pooling kernel size of 2 was followed to reduce the feature dimension and avoid overfitting. Three fully connected layers with an ReLU activation function followed the flattened layer, and then a dropout layer with probabilities of 0.3 was added for randomly discarding variables. Eventually, an output layer with two variables and the Softmax activation function was adopted to represent the final statistical classification score. Here, the predictive score of greater than 0.95 was treated as threshold of valid recognition between TCR CDR3 and SARS-CoV-2 epitopes. Consequently, we observed that the stable accuracy both converged in training dataset and test dataset, and the AUC reached 0.963 in the test dataset (Fig. 3a).

Furthermore, to verify the performance of the model CoV2-TCR, an independent validation dataset was prepared and applied. We downloaded those TCRβ CDR3 and their recognizing SARS-CoV-2 epitopes from VDJdb [23]. After excluding those pairs duplicated with positive pairs downloaded from IEDB, we treated the rest as positive pairs of the validation dataset. To produce negative pairs of the validation dataset, those TCR CDR3 derived from healthy people or induced by HIV were randomly paired with SARS-CoV-2 epitopes. Then, the positive pairs and negative pairs were merged and considered as an independent validation dataset. Here, the AUC of our model CoV2-TCR achieved at 0.967 in this independent validation dataset (Fig. 3b), indicating that model CoV2-TCR was an excellent classifier to probe the specific recognition between TCR CDR3 and SARS-CoV-2 epitopes.

Further, we compared the CoV2-TCR with other similar models, *i.e.* imRex [32] and ERGO [33]. By contrast, the imRex represents the interaction between epitopes and TCR CDR3 by combining amino-acid physicochemical properties and BLOSUM matrices, which are then used as inputs to construct convolutional neural network model. The ERGO was trained with two models, *i.e.* ERGO-AE and ERGO-LSTM, by using autoencoder (AE) and long short-term memory (LSTM), respectively. Based on the independent validation dataset, we found that the CoV2-TCR outperformed imRex, ERGO-LSTM, and ERGO-AE with a higher AUC (Fig. 3c), indicating that our model had better performance on predicting TCR CDR3 and their recognizing epitopes from SARS-CoV-2.

3.4. Identification of TCR CDR3 binding to SARS-CoV-2 antigen epitopes based on model CoV2-TCR

To classify those TCR CDR3 induced by SARS-CoV-2 epitopes from those clonotypes from COVID-19 samples, the model CoV2-TCR was applied for subsequent analysis. In a prior study [34], according to the amino-acid sequences of the four main proteins from SARS-COV-2, *i.e.* spike glycoprotein (S), membrane protein (M), envelop protein (E), and nucleocapsid protein (N), we screened 28 predicted MHC-I antigen peptides with the length of (8–11)-mer by performing a series of bioinformatics tools. Subsequently, based on those 380,305 unique clonotypes from COVID-19 patients but not from HD, including 37,562 High-abundance, 236,182 Low-abundance, 106,561 Rare clonotypes, the model CoV2-TCR was used to identify those TCR CDR3 that can specifically recognize these 28 predicted epitopes (Table S1). As shown in Table 1, we individually enumerated the 28 antigens and the number of TCR CDR3 sequences recognized, from 11 to 244,792, indicating a huge difference in T cell immune recognition to epitopes. Further, based on the clonal degree as described above, those TCR CDR3 sequences recognized to each epitope were separately divided into three groups (*i.e.* High-abundance, Low-abundance, Rare clonotypes). In those TCR CDR3 recognizing each epitope, we counted respectively their proportions in the total High-abundance, Low-abundance, and Rare clonotypes, and found that more TCR CDR3 did belong to Low-abundance clonotypes, but the proportions of High-abundance, Low-abundance, Rare clonotypes were close.

Besides, based on the four main proteins in SARS-CoV-2 and those amino-acid mutation site(s) of the seven SARS-CoV-2 variants (*i.e.* alpha, beta, delta, gamma, lambda, mu, omicron), we screened 22 predicted immunogenic peptides with %rank of less than 0.5 and IEDB immunogenicity score of greater than 0.15 (Table 2). Similarly, the CoV2-TCR model was applied to identify those TCR CDR3 that can recognize specifically these selected peptides (Table S2). Consistently, indeed, these peptides all can bind to some TCR CDR3 in those unshared clonotypes from COVID-19 patients, although the number of TCR CDR3 still presented a huge difference, ranging from 1 to 80,772. Also, for those TCR CDR3 recognizing each peptide, the proportions of High-abundance, Low-abundance, Rare clonotypes were still close (Table 2). That denoted some immunogenic peptides induced adaptive immune response, leading to uniform T-cell

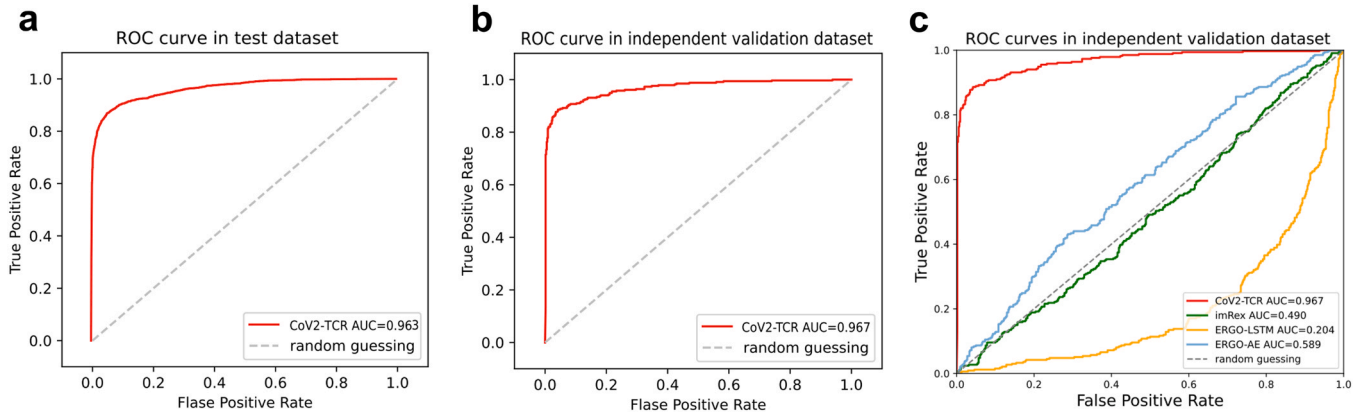

Fig. 3. Performance of model CoV2-TCR specifically recognized to SARS-CoV2 epitopes. (a, b) The ROC curve of model CoV2-TCR in test dataset (a) and in independent validation dataset (b), respectively. (c) The ROC curves of model CoV2-TCR, imRex, ERGO-LSTM, and ERGO-AE in independent validation dataset.

expansion of High-abundance, Low-abundance, Rare clonotypes. In addition, those recognizing TCR CDR3 were compared with those downloaded from IEDB and VDJdb. We found that 16 peptides, of which TCR CDR3 recognizing sequences did overlap with those from the two databases (*i.e.* IEDB, VDJdb), although the peptides from SARS-CoV-2 variants may be completely different from those from positive pairs (Table S3).

3.5. A web-based tool to identify TCR CDR3 recognizing SARS-CoV-2 peptides

Given the results of the investigation above, the model CoV2-TCR was implemented in a user-friendly web server (http://www.biostatistics.online/CoV2-TCR/#/). In addition, the four main proteins in SARS-CoV-2, *i.e.* spike protein (S), membrane protein (M), envelope protein (E), and nucleocapsid protein (N), have been integrated as well. As we have known, based on the amino-acid sequences of these proteins, there were four probable mutation types, including single amino acid variation (SAAV), insertion and deletion of amino acids (InDelAA), insertion of amino acids (InAA), deletion of amino acids (DelAA) [35].

In the process of operating the webserver, before submitting a task, users need to select a mutated protein and its mutation type, and input mutation site(s) and amino acid change. Then, nine peptides containing mutation site(s) (9-mer) will be respectively extracted for subsequent prediction by CoV2-TCR. Those pairs between peptide and TCR CDR3 with predictive score of greater than 0.95 will be recorded in an Excel file and sent to the user as an email. The output file will consisted of five columns, including SARS-CoV-2 Epitope, TCR CDR3, CDR3 clone degree, Predictive score, and Interaction.

During the infection process of SARS-CoV-2, the spike protein plays a crucial role in interacting with the human angiotensin-converting enzyme 2 (hACE2) receptor [36]. Singh et al [37]. identified the highly flexible region in the receptor binding domain of SARS-CoV-2, starting from residue 475–485 of spike protein. Particularly,

Table 1
Those TCR CDR3 recognizing the 28 predicted MHC-I peptides from SARS-CoV-2.

	Protein	Position	Peptide	Length	TCR CDR3 [score > 0.95]	High-abundance (%)	Low-abundance (%)	Rare (%)	TCR CDR3 [top 1]
1	Envelop	E_{19-26}	LFLAFVVF	8	11	0 (0)	6 (0.003)	5 (0.005)	CASTNQTSMYLCASSLASGNYEQYF
2	Envelop	E_{20-27}	FLAFVVFL	8	222	18 (0.05)	129 (0.06)	75 (0.07)	CASSSQTVIPRARANYGYTF
3	Envelop	E_{29-37}	VTLAILTAL	9	265	75 (0.20)	102 (0.04)	88 (0.08)	CASSIEAWLNIQYFGAGTRLGELFF
4	Envelop	E_{29-38}	VTLAILTALR	10	7021	1239 (3.30)	3629 (1.54)	2153 (2.02)	CSVLSSNTGELFF
5	Envelop	E_{30-37}	TLAILTAL	8	276	49 (0.13)	145 (0.06)	82 (0.08)	CASSIEAWLNIQYFGAGTRLGELFF
6	Membrane	$M_{134-143}$	LESELVIGAV	10	1590	130 (0.35)	968 (0.41)	492 (0.46)	CASSQEPPELAGWAGELFF
7	Membrane	$M_{134-144}$	LESELVIGAVI	11	3932	356 (0.95)	2314 (0.98)	1262 (1.19)	CASSQEPPELAGWAGELFF
8	Membrane	$M_{135-144}$	ESELVIGAVI	10	20,387	2113 (5.63)	12,561 (5.32)	5713 (5.36)	CASSIEAWLNIQYFGAGTRLGELFF
9	Membrane	$M_{136-146}$	SELVIGAVILR	11	64,488	5696 (15.17)	40,879 (17.31)	17,913 (16.82)	CASSPQPNYRVWDTDTQYF
10	Membrane	M_{19-28}	QWNLVIGFLF	10	25,960	2608 (6.94)	16,204 (6.86)	7148 (6.71)	CASTNQTSMYLCASSLASGNYEQYF
11	Membrane	M_{4-14}	SNGTITVEELK	11	975	114 (0.30)	573 (0.24)	288 (0.27)	CSVEFRPPYGYTFGSGTSSYNEQFF
12	Membrane	M_{73-82}	INWITGGIAI	10	51,034	5201 (13.85)	31,568 (13.37)	14,265 (13.39)	CSVEVRWEGALYEQYF
13	Membrane	M_{76-84}	ITGGIAIAM	9	1877	144 (0.38)	1224 (0.52)	509 (0.48)	CASSSQTVIPRARANYGYTF
14	Nucleocapsid	N_{10-17}	RNAPRITF	8	994	65 (0.17)	611 (0.26)	318 (0.30)	CASSSQTVIPRARANYGYTF
15	Nucleocapsid	$N_{111-121}$	YYLGTGPEAGL	11	204,512	19,865 (52.89)	131,190 (55.55)	53,457 (50.19)	CSLMRGALYNEQFF
16	Nucleocapsid	$N_{360-369}$	YKTFPPTEPK	10	40,796	4066 (10.83)	24,637 (10.43)	12,093 (11.35)	CSLMRGALYNEQFF
17	Nucleocapsid	N_{9-17}	QRNAPRITF	9	467	39 (0.10)	265 (0.11)	163 (0.15)	CASSSQTVIPRARANYGYTF
18	Spike	$S_{1215-1224}$	YIWLGFIAGL	10	17,270	1817 (4.84)	10,795 (4.57)	4658 (4.37)	CASSSQTVIPRARANYGYTF
19	Spike	$S_{1220-1230}$	FIAGLIAIVMV	11	244,792	24,980 (66.51)	154,075 (65.24)	65,737 (61.71)	CSVVPSSPYNEQFF
20	Spike	$S_{232-241}$	GINITRFQTL	10	2078	277 (0.74)	1179 (0.50)	622 (0.58)	CASSSQTVIPRARANYGYTF
21	Spike	S_{55-65}	FLPFFSNVTWF	11	740	46 (0.12)	460 (0.20)	234 (0.22)	CASSSQTVIPRARANYGYTF
22	Spike	S_{62-70}	VTWFHAIHV	9	261	36 (0.10)	133 (0.06)	92 (0.09)	CSVEFRPPYGYTFGSGTSSYNEQFF
23	Spike	$S_{627-636}$	DQLTPTWRVY	10	9771	1023 (2.72)	5780 (2.45)	2968 (2.79)	CAWSPSWGDSRRNTEAFF
24	Spike	$S_{628-636}$	QLTPTWRVY	9	382	56 (0.15)	197 (0.08)	129 (0.12)	CSLMRGALYNEQFF
25	Spike	$S_{825-835}$	KVTLADAGFIK	11	10,070	1114 (2.97)	5900 (2.50)	3056 (2.87)	CSVEDLGQGLEINTEAFF
26	Spike	$S_{826-835}$	VTLADAGFIK	10	3344	407 (1.08)	2060 (0.87)	877 (0.82)	CASTNQTSMYLCASSLASGNYEQYF
27	Spike	$S_{827-835}$	TLADAGFIK	9	1449	270 (0.72)	778 (0.33)	401 (0.38)	CASTNQTSMYLCASSLASGNYEQYF
28	Spike	$S_{876-884}$	ALLAGTITS	9	2893	442 (1.18)	1507 (0.64)	944 (0.89)	CSVSLLNNWELTEAFF

Table 2
Those TCR CDR3 recognizing the 22 predicted immunogenic peptides from seven SARS-CoV-2 variants.

	Ancestry	Gene	Variant	Position	peptide	Rank% [0–0.5]	Immunogen [score > 0.15]	TCR CDR3 [score> 0.95]	High-abundance (%)	Low-abundance (%)	Rare (%)	TCR CDR3 [top1]
1	Alpha	S	VY143–144 V	136–144	CNDPFLGVY	0.27	0.152	72	6 (0.02)	43 (0.02)	23 (0.02)	CASSSQTVIPRARANYGYTF
2	Alpha	S	D1118H	1111–1121	EPQIITDNTF	0.483	0.340	20,492	2110 (5.62)	12,708 (5.38)	5674 (5.33)	CSLMRGALYNEQFF
3	Alpha	S	VY143–144 V	135–144	FCNDPFLGVY	0.26	0.166	25,331	2568 (6.84)	15,355 (6.50)	7408 (6.96)	CSLMRGALYNEQFF
4	Alpha	S	T716I	714–722	IPINFTISV	0.32	0.203	96	9 (0.02)	63 (0.03)	24 (0.02)	CSVEFRPPYGYTFGSGTSSYNEQFF
5	Beta	S	TLLA240–243 T	233–241	INITRFQTL	0.05	0.168	59	11 (0.03)	20 (0.008)	28 (0.03)	CASSSQTVIPRARANYGYTF
6	Beta/Gamma/Omicron	S	K417N	414–423	QTGNIADYNY	0.26	0.186	573	62 (0.17)	322 (0.14)	189 (0.18)	CASSIEAWLNIQYFGAGTRLGELFF
7	Delta	S	G142D	136–145	CNDPFLDVYY	0.444	0.153	230	24 (0.06)	128 (0.05)	78 (0.07)	CASSSQTVIPRARANYGYTF
8	Delta	S	G142D	135–144	FCNDPFLDVY	0.249	0.156	46,978	5338 (14.21)	28,156 (11.92)	13,484 (12.66)	CSVRRGNTGELFF
9	Gamma	S	K417T	415–424	TGTIADYNYK	0.421	0.195	3376	414 (1.10)	2067 (0.88)	895 (0.84)	CASTNQTSMYLCASSLASGNYEQYF
10	Mu	S	V143VT	136–145	CNDPFLGVTY	0.452	0.187	1399	141 (0.38)	833 (0.35)	425 (0.40)	CASSSQTVIPRARANYGYTF
11	Mu	S	V143VT	136–146	CNDPFLGVTYY	0.434	0.202	1398	183 (0.49)	766 (0.32)	449 (0.42)	CSVEFRPPYGYTFGSGTSSYNEQFF
12	Mu/Omicron	S	T95I	89–97	GVYFASIEK	0.014	0.170	764	62 (0.17)	492 (0.21)	210 (0.20)	CASTNQTSMYLCASSLASGNYEQYF
13	Mu	S	V143VT	137–146	NDPFLGVTYY	0.161	0.206	95	11 (0.03)	56 (0.02)	28 (0.03)	CASSIEAWLNIQYFGAGTRLGELFF
14	Omicron	S	G339D	339–347	DEVFNATRF	0.206	0.225	1	0 (0)	0 (0)	1 (0.001)	CASSIEAWLNIQYFGAGTRLGELFF
15	Omicron	M	Q19E	18–26	EEWNLVIGF	0.041	0.226	6481	537 (1.43)	3968 (1.68)	1976 (1.86)	CASSYLGPSNTGELFF
16	Omicron	M	Q19E	18–28	EEWNLVIGFLF	0.358	0.349	80,772	7241 (19.28)	50,959 (21.58)	22,572 (21.19)	CASSQDAGPDSLFDTRELFF
17	Omicron	E	T9I	4–12	FVSEEIGTL	0.356	0.321	1043	101 (0.27)	661 (0.28)	281 (0.26)	CASTNQTSMYLCASSLASGNYEQYF
18	Omicron	M	Q19E	15–23	KLLEEWNLV	0.046	0.391	516	60 (0.16)	292 (0.12)	164 (0.15)	CASSLQWGDSVSTEAFF
19	Omicron	M	A63T	56–64	LIWPVTLIC	0.453	0.151	14,949	1574 (4.19)	8989 (3.81)	4386 (4.12)	CASSSQTVIPRARANYGYTF
20	Omicron	S	S375F	370–378	NSASFFTFK	0.07	0.172	464	36 (0.10)	260 (0.11)	168 (0.16)	CASSIEAWLNIQYFGAGTRLGELFF
21	Omicron	E	T9I	6–14	SEEIGTLIV	0.433	0.304	1326	183 (0.49)	671 (0.28)	472 (0.44)	CASSSQTVIPRARANYGYTF
22	Omicron	S	A67V	62–70	VTWFHVIHV	0.473	0.391	2946	339 (0.90)	1692 (0.72)	915 (0.86)	CSVEFRPPYGYTFGSGTSSYNEQFF

the two amino-acid exchanges, *i.e.* S477G and S477N, were reported to strengthen the binding of the SARS-COV-2 spike with the hACE2 receptor [37]. Subsequently, here as a case, the two mutations were separately inputted into the CoV2-TCR, and those TCR CDR3 recognizing those peptides containing mutation site was identified (Tables S4 and S5). The peptides TEIYQAG(G)T and TEIYQAG(N)T were recognized respectively by the most TCR CDR3, suggesting their potential as candidate vaccine peptides.

4. Discussion

In this study, on the basis of trancriptomic files from peripheral blood of COVID-19 patients and HD subjects, the TCR immune repertoires were investigated and analyzed. Meanwhile, we found the severity degree of the COVID-19 patients was negatively correlated with the diversity of TCR repertoires, and was positively correlated to their clonality. Furthermore, when we assigned those clonotypes into High-abundance, Low-abundance, Rare groups, the proportion of high-abundance clonotypes was found to be associated with the severity degree of COVID-19 patients. Indeed, the more cells infected with SARS-CoV-2 in host, the more severe the disease became [38]. Sustained stimulation by antigen epitopes from SARS-CoV-2 would cause biased cloning expansion, leading to a bias T cell immune repertoire with low diversity.

Furthermore, a clear difference was discovered when the proportions of the top 20 TCR CDR3 were explored in COVID-19 patients. Interestingly, we observed that two sequences, *i.e.* CSVESTALAAYNEQFF and CSVRAGTGGTEAFF presented little in active samples and fatal/severe samples, but high proportion in recovered samples and mild/moderate samples. That indicated the severity degree of disease may be related to the amplified TCR CDR3 sequences. In addition, previous reports showed that those unexposed-unvaccinated individuals carried a significant fraction of circulating CD8 + T cells reactive to epitopes from SARS-CoV-2. These CD8 + T cells may belong to memory and naive T cells derived from other coronavirus infections [17,18,39]. The public shared T cells repertoire in COVID-19 patients may contain an unexpected fraction of TCR clonotypes that can react to different epitopes from SARS-CoV-2 variants. Thus, we believed that, among the TCR CDR3 in COVID-19 patients, there existed some TCR CDR3 that can recognize those mutated epitopes from any SARS-CoV-2 variants.

To investigate the TCR repertoire of COVID-19 patients and their recognized epitopes from SARS-CoV-2 or its variants, a precise model CoV2-TCR was constructed and outperformed the similar models, *i.e.* imRex [32], ERGO-LSTM [33], and ERGO-AE [33]. In addition, by comparison with the TCRmatch [40] and TCRdist [41] constructed by the similarity-based or distance-based approaches, the CoV2-TCR was established by convolution neural network, whose advantage lies in its capacity of feature extraction. NetTCR-2.0 was trained to predict interactions between TCRs and their cognate HLA-A* 02:01-restricted peptides [42]. Due to the small number of training peptides, the tool can only be applied to the limited peptides included in the training data. The CoV2-TCR we proposed in this study provided a generic strategy that can specifically recognize MHC-I epitopes from SARS-CoV-2 or its variants. We believe that our method could be extended to some similar protein-binding situations, helping to guide the selection of candidate vaccine polypeptides and immunotherapeutic TCR-T.

When the CoV2-TCR was used to explore those TCR CDR3 recognizing those predicted epitopes from SARS-CoV-2 or its variants, the number of TCR CDR3 was found to exhibit a huge variability. This is why immunogenic epitopes need to be screened for use as peptide vaccines. Indeed, the CoV2-TCR could provide preliminary screening, in which those epitopes from SARS-CoV-2 variants recognized by most TCR CDR3 may be selected as candidate antigen peptides for vaccine design. In addition, in those TCR CDR3 recognizing each

epitope, we observed the close proportions of High-abundance, Low-abundance, Rare clonotypes in the total population, suggesting that those immunogenic peptides may cause uniform immune response in host.

In conclusion, this study investigated the TCR repertoire of SARS-CoV-2 infected patients, and those TCR CDR3 existed in COVID-19 patients were found to recognize epitopes from mutation(s) of SARS-CoV-2 variants. Notably, we established an easy-to-use web server for screening TCR CDR3 sequences that can recognize a specific SARS-CoV-2 mutated peptide, which would be helpful to identify candidate antigen peptides responding to SARS-CoV-2 variants.

Funding

This work was supported by the National Natural Science Foundation of China (31900102, 31870829), Shanghai Municipal Health Commission Collaborative Innovation Cluster Project (2019CXJQ02), and Natural Science Foundation of Hunan Province (2021JJ40995).

CRediT authorship contribution statement

Xingxing Jian: Conceptualization, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Yu Zhang**: Investigation, Software, Methodology. **Jingjing Zhao**: Conceptualization, Investigation, Methodology. **Zhuoming Zhao**: Investigation, Discussion. **Manman Lu**: Investigation, Discussion. **Lu Xie**: Conceptualization, Investigation, Supervision, Writing – review & editing.

Declaration of Competing Interest

All authors declare no competing interests.

Acknowledgments

We appreciate the resources provided by the High Performance Computing Center of Central South University.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2023.01.038.

References

[1] Agerer B, Koblischke M, Gudipati V, Montano-Gutierrez LF, Smyth M, Popa A, et al. SARS-CoV-2 mutations in MHC-I-restricted epitopes evade CD8(+) T cell responses. Sci Immunol 2021;6(57):eabg6461.
[2] Wang M, Zhou B, Fan Q, Zhou X, Liao X, Lin J, et al. Omicron variants escape the persistent SARS-CoV-2-specific antibody response in 2-year COVID-19 convalescents regardless of vaccination. Emerg Microbes Infect 2023;12(1):2151381.
[3] Dhama K, Nainu F, Frediansyah A, Yatoo MI, Mohapatra RK, Chakraborty S, et al. Global emerging Omicron variant of SARS-CoV-2: Impacts, challenges and strategies. J Infect Public Health 2023;16(1):4–14.
[4] Tang-Huau TL, Gueguen P, Goudot C, Durand M, Segura EJNC. Human in vivo-generated monocyte-derived dendritic cells and macrophages cross-present antigens through a vacuolar pathway. Nat Commun 2018 2018;9(1):2570.
[5] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395(10223):497–506.
[6] Dash P, Fiore-Gartland AJ, Hertz T, Wang GC, Sharma S, Souquette A, et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. Nature 2017;547(7661):89–93.
[7] Le Bert N, Tan AT, Kunasegaran K, Tham CYL, Hafezi M, Chia A, Chng MHY, Lin M, Tan N, Linster M. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. Nature 2020;584(7821):457–62.
[8] Mathew D, Giles JR, Baxter AE, Oldridge DA, Greenplate AR, Wu JE, et al. Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. Science 2020;369(6508):eabc8511.
[9] Zhang X, Tan Y, Ling Y, Lu G, Liu F, Yi Z, et al. Viral and host factors related to the clinical outcome of COVID-19. Nature 2020;583(7816):437–40.
[10] McMahan K, Yu J, Mercado NB, Loos C, Tostanoski LH, Chandrashekar A, et al. Correlates of protection against SARS-CoV-2 in rhesus macaques. Nature 2021;590(7847):630–4.
[11] Hansen SG, Ford JC, Lewis MS, Ventura AB, Hughes CM, Coyne-Johnson L, et al. Profound early control of highly pathogenic SIV by an effector memory T-cell vaccine. Nature 2011;473(7348):523–7.
[12] Masopust D, Soerens AG. Tissue-resident T cells and other resident leukocytes. Annu Rev Immunol 2019;37:521–46.
[13] Grau-Exposito J, Sanchez-Gaona N, Massana N, Suppi M, Astorga-Gamaza A, Perea D, et al. Peripheral and lung resident memory T cell responses against SARS-CoV-2. Nat Commun 2021;12(1):3010.
[14] Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. JAMA Intern Med 2020;180(7):934–43.
[15] Peng Y, Mentzer AJ, Liu G, Yao X, Yin Z, Dong D, et al. Broad and strong memory CD4(+) and CD8(+) T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. Nat Immunol 2020;21(11):1336–45.
[16] Notarbartolo S, Ranzani V, Bandera A, Gruarin P, Bevilacqua V, Putignano AR, et al. Integrated longitudinal immunophenotypic, transcriptional and repertoire analyses delineate immune responses in COVID-19 patients. Sci Immunol 2021;6(62):eabg5021.
[17] Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Moderbacher CR, et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. Cell 2020;181(7):1489–501.
[18] Braun J, Loyal L, Frentsch M, Wendisch D, Georg P, Kurth F, et al. SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. Nature 2020;587(7833):270–4.
[19] Lu M, Xu L, Jian X, Tan X, Zhao J, Liu Z, et al. dbPepNeo2.0: A database for human tumor neoantigen peptides from mass spectrometry and TCR recognition. Front Immunol 2022;13:855976.
[20] Schultheiss C, Paschold L, Simnica D, Mohme M, Willscher E, Wenserski L, et al. Next-generation sequencing of T and B cell receptor repertoires from COVID-19 patients showed signatures associated with severity of disease. Immunity 2020;53(2):442–55.
[21] Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods 2015;12(5):380–1.
[22] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. Nucleic Acids Res 2019;47(D1):D339–43.
[23] Shugay M, Bagaev DV, Zvyagin IV, Vroomans RM, Crawford JC, Dolton G, et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. Nucleic Acids Res 2018;46(D1):D419–27.
[24] Shomuradova AS, Vagida MS, Sheetikov SA, Zornikova KV, Kiryukhin D, Titov A, et al. SARS-CoV-2 epitopes are recognized by a public and diverse repertoire of human T cell receptors. Immunity 2020;53(6):1245–57. e5.
[25] Gong Z, Zhu JW, Li CP, Jiang S, Ma LN, Tang BX, et al. An online coronavirus analysis platform from the National Genomics Data Center. Zool Res 2020;41(6):705–8.
[26] Song S, Ma L, Zou D, Tian D, Li C, Zhu J, et al. The global landscape of SARS-CoV-2 genomes, variants, and haplotypes in 2019nCoVR. Genom Prote Bioinf 2020;18(6):749–59.
[27] Yu D, Yang X, Tang B, Pan YH, Yang J, Duan G, et al. Coronavirus GenBrowser for monitoring the transmission and evolution of SARS-CoV-2. Brief Bioinform 2022;23(2):bbab583.
[28] Zhao WM, Song SH, Chen ML, Zou D, Ma LN, Ma YK, et al. The 2019 novel coronavirus resource. Yi Chuan 2020;42(2):212–21.
[29] Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res 2020;48(W1):W449–54.
[30] Kumar BV, Connors TJ, Farber DL. Human T cell development, localization, and function throughout life. Immunity 2018;48(2):202–13.
[31] Katayama Y, Yokota R, Akiyama T, Kobayashi TJ. Machine learning approaches to TCR repertoire analysis. Front Immunol 2022;13:858057.
[32] Moris P, De Pauw J, Postovskaya A, Gielis S, De Neuter N, Bittremieux W, et al. Current challenges for unseen-epitope TCR interaction prediction and a new perspective derived from image classification. Brief Bioinform 2021;22(4):bbaa318.
[33] Springer I, Besser H, Tickotsky-Moskovitz N, Dvorkin S, Louzoun Y. Prediction of specific TCR-peptide binding from large dictionaries of TCR-peptide pairs. Front Immunol 2020;11:1803.
[34] Zhao J, Tan X, Wang G, Ouyang J, Jian X, Xie L. Bioinformatics prediction of candidate epitopes for peptide vaccines based on pathogen antigenic protein sequences. Life Sci Res 2021;25(04):363–71.
[35] Jian X, Xu L, Zhao J, Wang Y, Zhou W, Xie L. NAIRscore as a biomarker for the quality of immune response to neoantigens is related with an increased overall survival in multiple myeloma. Mol Ther Nucleic Acids 2022;29:285–95.
[36] Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 2020;181(2):281–92. e6.
[37] Singh A, Steinkellner G, Kochl K, Gruber K, Gruber CC. Serine 477 plays a crucial role in the interaction of the SARS-CoV-2 spike protein with the human receptor ACE2. Sci Rep 2021;11(1):4320.
[38] Luo L, Liang W, Pang J, Xu G, Chen Y, Guo X, et al. Dynamics of TCR repertoire and T cell function in COVID-19 convalescent individuals. Cell Disco 2021;7(1):89.

[39] Quiros-Fernandez I, Poorebrahim M, Fakhr E, Cid-Arregui A. Immunogenic T cell epitopes of SARS-CoV-2 are recognized by circulating memory and naive CD8 T cells of unexposed individuals. EBioMedicine 2021;72:103610.
[40] Chronister WD, Crinklaw A, Mahajan S, Vita R, Kosaloglu-Yalcin Z, Yan Z, et al. TCRMatch: Predicting T-Cell receptor specificity based on sequence similarity to previously characterized receptors. Front Immunol 2021;12:640725.
[41] Olson BJ, Schattgen SA, Thomas PG, Bradley P, Matsen Iv, FA. Comparing T cell receptor repertoires using optimal transport. PLoS Comput Biol 2022;18(12):e1010681.
[42] Montemurro A, Schuster V, Povlsen HR, Bentzen AK, Jurtz V, Chronister WD, et al. NetTCR-2.0 enables accurate prediction of TCR-peptide binding by using paired TCR alpha and beta sequence data. Commun Biol 2021;4(1):1060.