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.

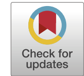
International Journal of Infectious Diseases 113 (2021) 308–317

Contents lists available at ScienceDirect

International Journal of Infectious Diseases

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

T-cell receptor repertoires as potential diagnostic markers for patients with COVID-19



Xianliang Hou[a,1], Guangyu Wang[c,1], Wentao Fan[b,1], Xiaoyan Chen[d], Chune Mo[a], Yongsi Wang[b], Weiwei Gong[a], Xuyan Wen[e], Hui Chen[b], Dan He[b], Lijun Mo[f], Shaofeng Jiang[g], Minglin Ou[a], Haonan Guo[h,**], Hongbo Liu[i,*]

[a] *Central Laboratory, Guangxi Health Commission Key Laboratory of Glucose and Lipid Metabolism Disorders, the Second Affiliated Hospital of Guilin Medical University, Guilin, 541199, China*
[b] *Guangzhou Huayin Health Medical Group Co., Ltd, Guangzhou, China*
[c] *College of Laboratory Medicine, Guilin Medical University, Guilin, 541199, China*
[d] *Department of State Owned Assets Management, Affiliated Hospital of Guilin Medical University, Guilin, 541001, China*
[e] *Department of Pathology, Affiliated Hospital of Guilin Medical University, Guilin, 541001, China*
[f] *Clinical Laboratory, the Second Affiliated Hospital of Guilin Medical University, Guilin, 541199, China*
[g] *Guangxi Key Laboratory of Tumor Immunology and Microenvironmental Regulation, Guilin Medical University, Guilin, Guangxi, 541199, China*
[h] *Department of Clinical Laboratory, The Affiliated Hospital of Guilin Medical University, Guilin, Guangxi, 541001, China*
[i] *Department of Laboratory Medicine, the Second Affiliated Hospital of Guilin Medical University, Guilin, 541199, China*

ARTICLE INFO

Article history:
Received 20 June 2021
Revised 25 September 2021
Accepted 15 October 2021

KEYWORDS:
Coronavirus disease 2019
T-cell receptor
SARS-CoV-2
Adaptive immunity

ABSTRACT

Objective: Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is an ongoing global health emergency. T-cell receptors (TCRs) are crucial mediators of antiviral adaptive immunity. This study sought to comprehensively characterize the TCR repertoire changes in patients with COVID-19.

Methods: A large sample size multi-center randomized controlled trial was implemented to study the features of the TCR repertoire and identify COVID-19 disease-related TCR sequences.

Results: It was found that some T-cell receptor beta chain (TCRβ) features differed markedly between COVID-19 patients and healthy controls, including decreased repertoire diversity, longer complementarity-determining region 3 (CDR3) length, skewed utilization of the TCRβ variable gene/joining gene (TRBV/J), and a high degree of TCRβ sharing in COVID-19 patients. Moreover, this analysis showed that TCR repertoire diversity declines with aging, which may be a cause of the higher infection and mortality rates in elderly patients. Importantly, a set of TCRβ clones that can distinguish COVID-19 patients from healthy controls with high accuracy was identified. Notably, this diagnostic model demonstrates 100% specificity and 82.68% sensitivity at 0–3 days post diagnosis.

Conclusions: This study lays the foundation for immunodiagnosis and the development of medicines and vaccines for COVID-19 patients.

© 2021 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has spread rapidly worldwide (Lai et al., 2020). As of June 20, 2021, SARS-CoV-2 had affected more than 179 060 045 people globally, caus-

Abbreviations: COVID-19, coronavirus disease 2019; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; CDR3, complementarity-determining region 3; TCR, T-cell receptor; TCRβ, T-cell receptor beta chain; TRBV, TCR beta chain variable gene; TRBJ, TCR beta chain joining gene; HC, Healthy control; DLS, Discovery Life Science; BWNW, Bloodworks Northwest; HUniv12Oct, Hospital Univesitario 12 de Octubre.

* Corresponding authors: Hongbo Liu, Department of Laboratory Medicine, the Second Affiliated Hospital of Guilin Medical University, Guilin, 541199, China.
** Haonan Guo, Department of Clinical Laboratory, The Affiliated Hospital of Guilin Medical University, Guilin, Guangxi 541001, China.

E-mail addresses: guohaonan@glmc.edu.cn (H. Guo), hbliu@glmc.edu.cn (H. Liu).

[1] Xianliang Hou, Wentao Fan, and Guangyu Wang contributed equally to this work.

https://doi.org/10.1016/j.ijid.2021.10.033
1201-9712/© 2021 The Author(s). Published by Elsevier Ltd on behalf of International Society for Infectious Diseases. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

ing over 3.87 million deaths. In the USA, as many as 34 401 766 individuals had tested positive for COVID-19, and the death toll had reached 617 091 people (Baidu, 2021). The symptoms of COVID-19 include a dry cough, fever, diarrhea, fatigue, pneumonia, and conjunctivitis. Some patients develop acute respiratory distress syndrome (ARDS), severe pneumonia, or multiple organ failure (Ahn et al., 2020; Guo et al., 2020; Lai et al., 2020). Following its global spread, the World Health Organization declared the current outbreak of coronavirus a public health emergency of global concern. Although clinicians and scientists worldwide have made great efforts to explore antiviral drugs and produce vaccines (Ahn et al., 2020; Guo et al., 2020; Lai et al., 2020), there is still no highly effective clinical treatment and specific medicine for COVID-19. Thus, there is an urgent need to better understand the host immune response in SARS-CoV-2 infection to better devise diagnostic and prognostic biomarkers and design effective therapeutic interventions for the disease.

T-cells are the central mediators of antiviral adaptive immunity and play critical roles in clearing SARS-CoV-2 infections, directly influencing patient clinical outcomes (Ahn et al., 2020; Guo et al., 2020; Lai et al., 2020). T-cell antigen recognition requires T-cell receptors (TCRs), which are expressed on the T-cell surface. The antigen specificity of each TCR is primarily determined by the hypervariable complementarity-determining region 3 (CDR3) of the receptor chain, which originates from the recombination of the V (variable), D (diversity), and J (joining) gene segments and the deletion and insertion of nucleotides at the V(D)J junctions (Hou et al., 2019; Hou et al., 2019). These processes generate a diverse TCR repertoire capable of recognizing an extensive and unpredictable range of antigens and providing potent antiviral immunity (Hou et al., 2019; Hou et al., 2019). The composition and the diversity of the TCR repertoire changes in response to cancer, aging, chronic and acute infection, and many other internal and external forces, making the TCR repertoire highly dynamic. Specific recognition of antigens results in clonal expansion of antigen-specific T-cells, leading to skewing of the TCR repertoire (TCR bias) to favor antigen-specific T-cells (Hou et al., 2016; Wen et al., 2020). Evidence of TCR bias in antiviral immunity has been demonstrated for many viral infections, including those caused by Epstein–Barr virus (EBV) (Gil et al., 2020), severe acute respiratory syndrome coronavirus (SARS-CoV) (Gutierrez et al., 2020), and the Middle East respiratory syndrome coronavirus (MERS-CoV) (Channappanavar et al., 2014). With the development of technology, immunoSEQ Technology now allows ultra-deep sequencing of the T-cell receptor beta chain (TCRβ) CDR3 region, revealing the composition and characterization of T-cell populations (Channappanavar et al., 2014). Relevant to this investigation, Zhang and co-workers found COVID-19-induced remodeling of peripheral lymphocytes and SARS-CoV-2-specific shuffling of adaptive immune repertoires. They also indicated that peptides derived from the M protein of SARS-CoV-2 are active in inducing T-cell responses in most COVID-19 patients (Zhang et al., 2020).

In the present study, a large-scale and multi-center comprehensive immunological analysis was performed to decode the adaptive immune response directed against SARS-CoV-2 using high-throughput immune sequencing. The study comprehensively addressed the correlations between TCR diversity and immune responses against viral antigens, and explored to what extent SARS-CoV-2 influences the TCR repertoire, including TCR diversity, CDR3 frequency distribution, CDR3 length distribution, V/J usage, V–J pairing, and overlap indices. In particular, it was sought to identify the TCRs specific for SARS-CoV-2 viral antigens in COVID-19 patients. Through this study, it is hoped that a better understanding of the adaptive immune response to SARS-CoV-2 infection will be gained, which will provide a theoretical basis for the development of effective drugs or vaccines against SARS-CoV-2.

2. Methods

2.1. Biological materials

The TCR sequences for all study subjects were obtained from the ImmuneRACE study (Nolan et al., 2020). As described previously, the ImmuneRACE study is a prospective, single group, multi-cohort exploratory study of participants with COVID-19 (Nolan et al., 2020). The study includes the T-cell repertoire data of 593 individuals from three global collaborators. Participants aged 8 to 89 years and residing in 24 different geographical areas across the USA were consented and enrolled via a virtual study design. All of the samples were collected from patients who were actively suffering from or had recovered from COVID-19. In the COVID-19-BWNW group, whole blood samples were collected at Bloodworks Northwest (Seattle, WA, USA). In the COVID-19-DLS group, whole blood samples were collected at Discovery Life Sciences (Huntsville, AL, USA). In the COVID-19-HUniv12Oct group, whole blood samples were collected at the Hospital Universitario 12 de Octubre (Madrid, Spain). In addition, TCR sequences for healthy controls (HCs) were obtained from the Adaptive Biotechnologies immuneACCESS site (https://doi.org/10.21417/B7SG6T). A total of 43 healthy subjects were included in the study. Peripheral blood samples were collected from healthy donors who tested negative for anti-hepatitis B surface antigen (anti-HBsAg) antibodies and anti-HIV antibodies and exhibited no clinical or laboratory signs of other infectious diseases or immunological disorders. Among these 43 healthy donors, 23 were female and 20 were male, and they had a mean age of 46.16 ± 15.56 years, ranging from 30 to 61 years.

2.2. Genomic DNA extraction and high-throughput sequencing and analysis

Whole blood samples were taken from each volunteer and collected in K2EDTA tubes. Samples were stored at the institution and sent to Adaptive Biotechnologies as frozen whole blood, isolated peripheral blood mononuclear cells (PBMCs), and DNA extracted from either sample type for TCRβ analysis via immunoSEQ. Immunosequencing of the TCRβ CDR3 regions was performed using the immunoSEQ assay as described previously (Robins et al., 2012; Carlson et al., 2013). In brief, a bias-controlled multiplex-PCR system was designed to amplify the extracted genomic DNA. Subsequently, high-throughput sequencing was performed. Raw data processing and analysis were performed with the immunoSEQ Analyzer software (http://www.adaptivebiotech.com/immunoseq). Demultiplexed reads were then further processed to reduce amplification and sequencing bias. TCRβ V, D, and J gene definitions were provided by the IMGT database (www.imgt.org). As any given CDR3 sequence can be produced in multiple ways, the probability distribution of hidden recombination events cannot be inferred directly from the observed sequences (Elhanati et al., 2018; Pogorelyy et al., 2018). Therefore, existing methods were used to calculate the probability of TCR nucleotide sequences from a given recombination model (Murugan et al., 2012). The probabilistic model predicts the generation probability of any specific CDR3 sequence by the primitive recombination process, which is sufficient to allow annotation of the V(N)D(N)J genes constituting each unique CDR3 and to obtain the corresponding AA sequence. Moreover, Batch correction was performed to eliminate the batch effect between different datasets. In addition, multiple TCR data statistics were performed, including CDR3 frequency distribution, CDR3 length distri-

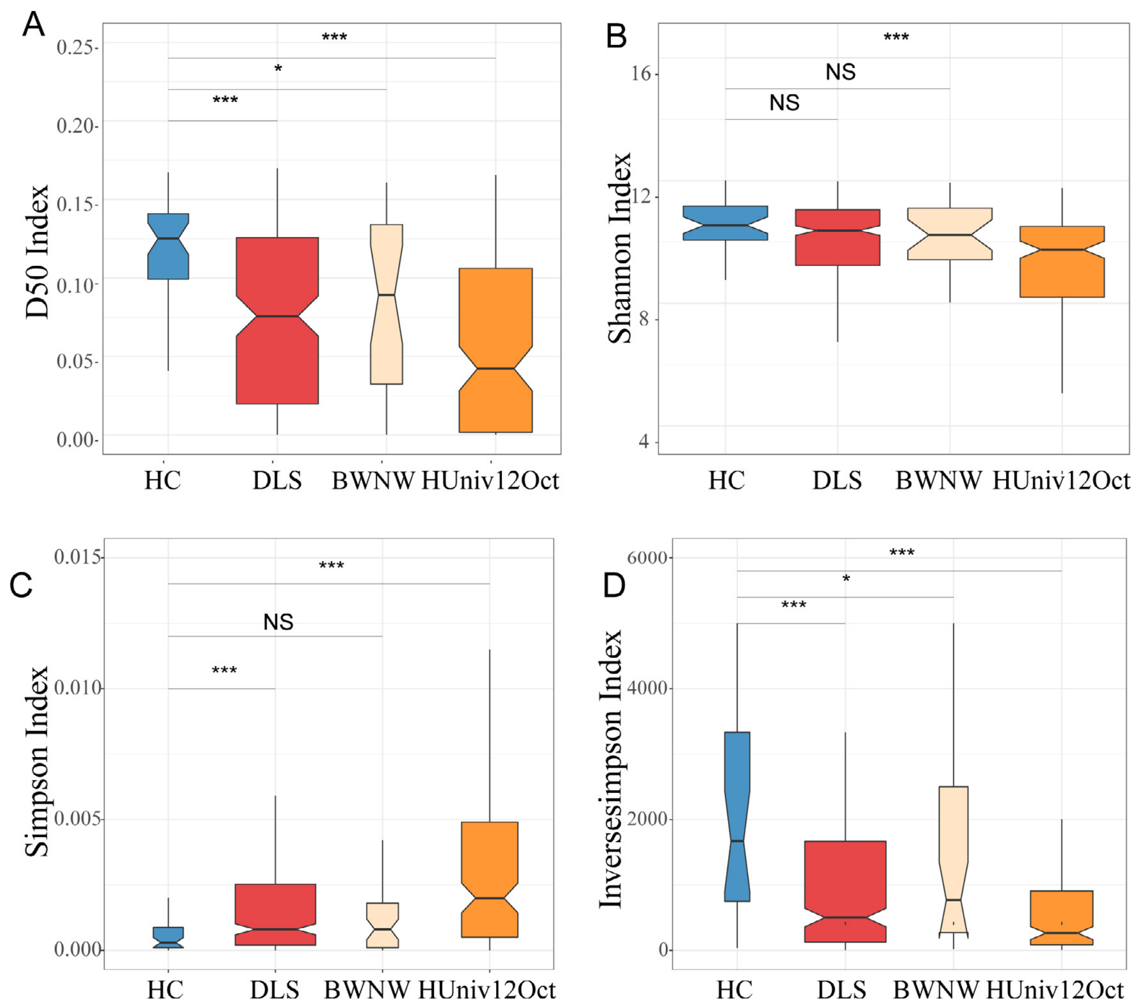


Figure 1. Diversity indices of the TCRβ repertoire in COVID-19 patients and healthy controls. The diversity measured by (A) D50 index, (B) Shannon index, (C) Simpson index, and (D) and inverse Simpson index. Data are presented as the mean ± standard deviation values; comparisons were made using the *t*-test. $^{*}P < 0.05$, $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$ (two-tailed).

bution, V/J usage, V–J pairing, and the length distribution of Vdels, Jdels, D5dels, D3dels, n1ins, and n2ins. All of these analyses were assessed based on earlier published work (Gomez-Tourino et al., 2017; Hou et al., 2019; Huang et al., 2019). TCR repertoire diversity was calculated based on the D50 index, Shannon index, Simpson index, and inverse Simpson index (Stewart et al., 1997; Venturi et al., 2007). Sharing among TCR repertoires was quantified by calculating the overlap coefficient (overlap $(X, Y) = |X$ and $Y|/\min(|X|, |Y|)$ for amino acid sequences (species = nucleotide sequence). Moreover, to further identify the COVID-19-associated clones, we searched for the clones that were highly abundant in the COVID-19 group but rare in the HC group, using methods described previously (Huang et al., 2019).

2.3. Definition of COVID-19 disease-associated clones

Disease-associated clones were defined as those TCRβ presenting in at least four COVID-19 patients and fewer than three healthy individuals. The disease-related clones were obtained through screening.

2.4. Classification of the COVID-19 patients according to the disease-associated clones

The random forest model in R package was used, based on two features: the proportion of unique disease-associated clones present in the sample (the number of disease-associated clones divided by the total number of clones in the sample) and the proportion of total disease-associated clones present in the sample (the sum frequency of disease-associated clones in the sample). Exhausted leave-one-out cross-validation was used to assess the identifier's performance during model training. More specifically, given that there were N samples in each group, $4/5N$ samples were used as training data and processed using the above classification model. The remaining $1/5N$ samples were used as testing data to perform the classification. The cross-validations were repeated $5N$ times until every sample had been used as testing data five times (Huang et al., 2019). Receiver operating characteristic (ROC) curves and area under the ROC curve (AUC) values were calculated using the predicted probability of a given TCRβ chain belonging to the COVID-19 population. Accuracy was calculated as the percentage of correct predictions divided by the total number of predictions made. Subsequently, a new independent cohort of COVID donors ($n = 538$), HCs ($n = 36$), and rheumatoid arthritis ($n = 65$) patients was used to prove the potential use of these disease-associated TCRβ clones as biomarkers of SARS-CoV-2 infection.

2.5. Statistical analysis

The assessment of statistical significance was performed using the Mann–Whitney *U*-test or unpaired *t*-test where appropriate. The statistical analyses were performed using IBM SPSS Statistics

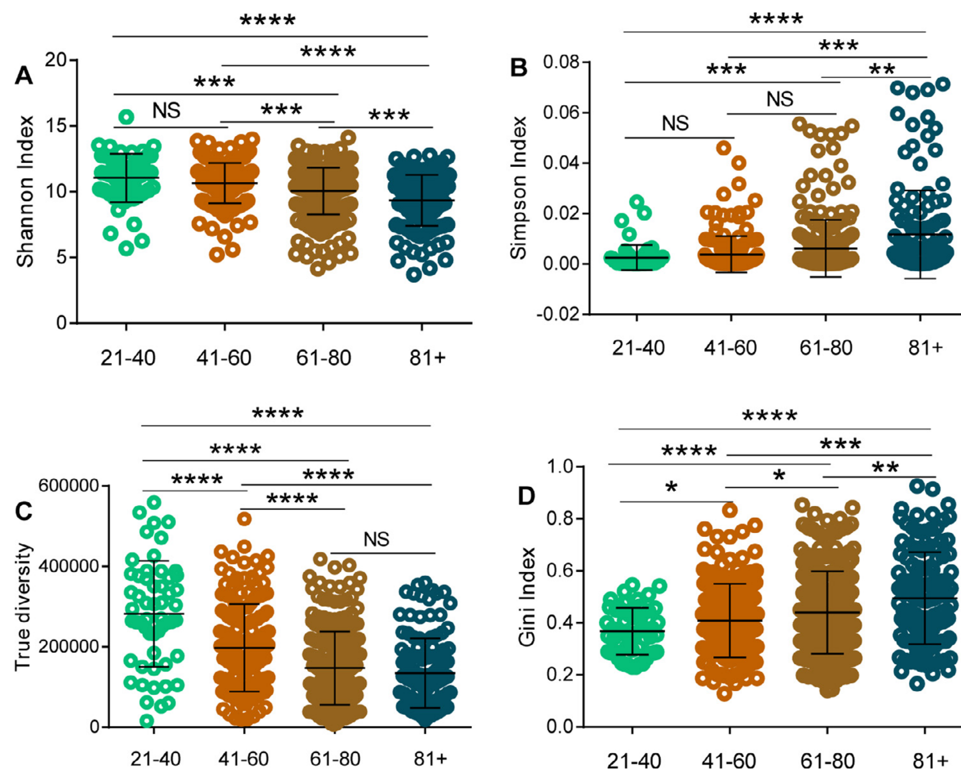


Figure 2. TCR repertoire diversity declines with aging in COVID-19 patients. The COVID-19 patients were assigned to four groups according to their age: 21–40, 41–60, 61–80, 80+. The diversity measured by (A) Shannon index, (B) Simpson index, (C) true diversity index, and (D) Gini index. Data are presented as the mean ± standard deviation values; comparisons were made using the *t*-test. *$P < 0.05$, **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$ (two-tailed).

version 20, and a two-tailed *P*-value less than 0.05 was considered significant.

3. Results

3.1. Decreased TCRβ repertoire diversity in COVID-19 patients

The Simpson index and inverse Simpson index were used to investigate the TCRβ diversity of the PBMCs of COVID-19 patients from Bloodworks Northwest (BWNW), Discovery Life Sciences (DLS), and Hospital Universitario 12 de Octubre (HUniv12Oct). Regarding the Simpson index and Gini index, the smaller the index value, the greater the CDR3 diversity. With regard to the D50, Shannon, inverse Simpson, and true diversity indices, the greater the index value, the greater the CDR3 diversity. As shown in Figure 1, the highest diversity was in the HC group, and a lower TCR diversity was observed in COVID-19 patients than in the controls, especially in the DLS group and HUniv12Oct group. However, the differences were not always statistically significant ($P < 0.05$), especially for the comparison between the BWNW group and HC group. The COVID-19 patients were then divided into four groups according to their age (21–40, 41–60, 61–80, 80+), and it was found that TCR repertoire diversity declines with aging (Figure 2).

3.2. TRBV/J gene usage is strongly skewed in patients with COVID-19

In the context of similar HLA molecules, antigen-driven stimulation results in oligoclonal expansion of T-cells via common V–D–J segments. To study the preference genes and unique changes of TCR in COVID-19 patients, the usage frequencies of TRBV and TRBJ genes in COVID-19 patients were compared with those in the HCs (Figure 3A–F). The results revealed that TRBV3, TRBV9, TRBV11, TRBV12, TRBV15, TRBV21, TRBV24, and TRBV27 showed higher usage, while TRBV1, TRBV5, TRBV6, and TRBV19 showed significantly lower usage in COVID-19 patients when compared with HCs. Of note, it was observed that the specific skewed usages of TRBV3, TRBV9, TRBV11, TRBV12, TRBV15, and TRBV24 were found in COVID-19 patients from all three global collaborators (Figure 3A–C). The preferred TRBJs were TRBJ1-2 and TRBJ1-6, whereas the low usage TRBJs were TRBJ1-4 and TRBJ2-6, results that were consistent in all three global collaborators (Figure 3D–F). Moreover, the top three pairing frequencies in COVID-19 patients were TRBV7–TRBJ2-7, TRBV5–TRBJ2-7, and TRBV6–TRBJ2-3 (Figure 3G–I; **Supplementary Material** Figure S1).

3.3. TCRβ reveals a longer CDR3 length in COVID-19 patients

TCR CDR3 loops can vary in both sequence and length, and this gives them the ability to recognize a large and diverse array of antigens. It was found that CDR3 length distributions differed between COVID-19 patients and HCs, showing a shift towards longer clonotypes in COVID-19 patient cells versus HC cells, regardless of the group – BWNW, DLS, or HUniv12Oct (Figure 4A–C). Long TCRβ CDR3 lengths might arise from abnormal insertions and/or deletions at the V–D–J recombinant junctions. Therefore, to understand the molecular mechanism of these features, recombination events (indels) were analyzed in each of the six rearrangement positions (Vdels, Jdels, D5dels, D3dels, n1ins, and n2ins). The results showed that the length distribution of the six recombination events (indels) clearly differed between COVID-19 patients and HCs (Figure 4D–I; **Supplementary Material** Figures S2 and S3).

3.4. Higher degree of TCRβ sharing in COVID-19 patients

As with any other viral infection, the ability of the individual's repertoire to effectively clear SARS-CoV-2 infection is mostly

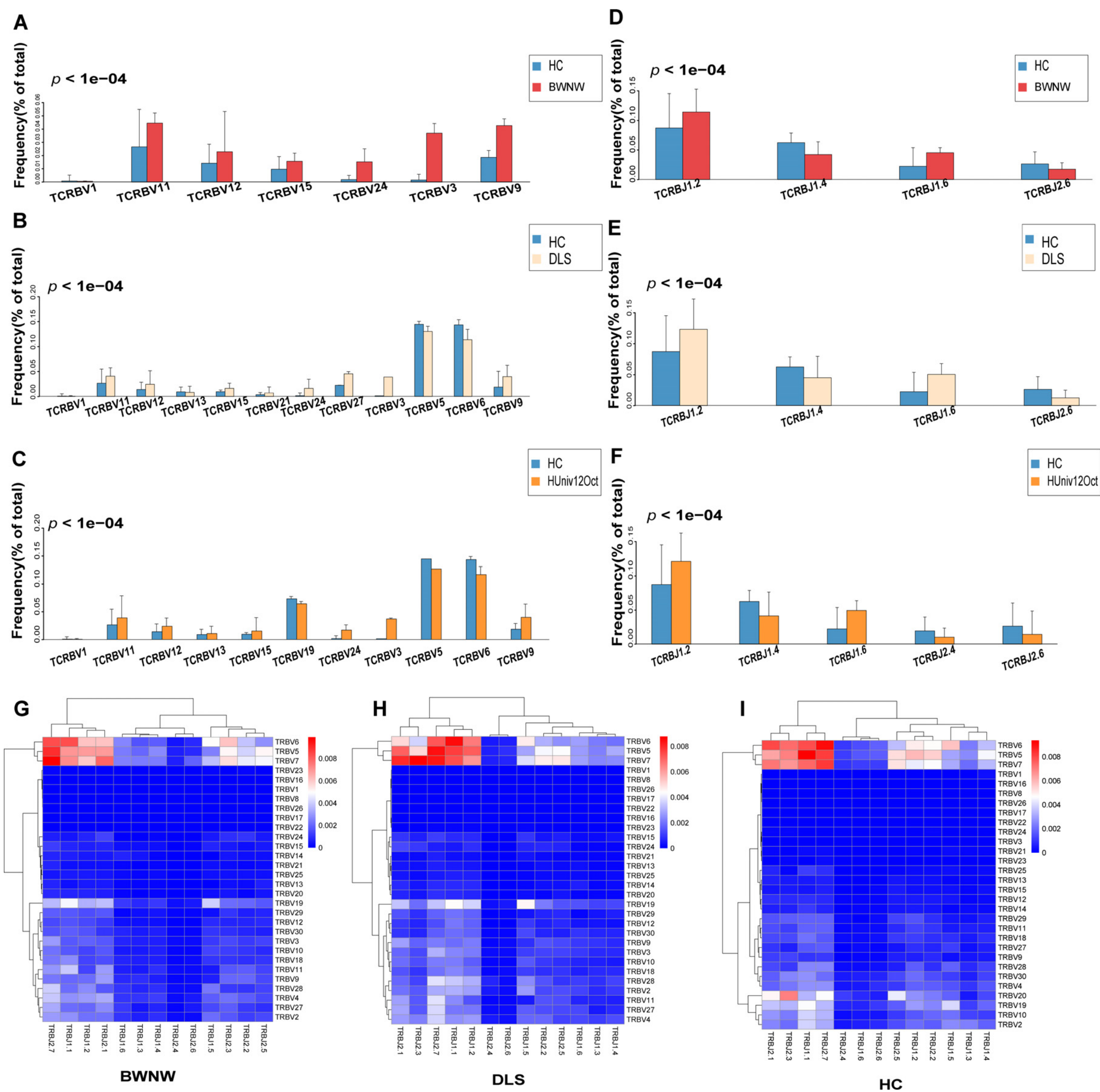


Figure 3. The skewed use of TRBV/J segments in patients with COVID-19. Significant differences in TRBV segment usage in COVID-19 patients in the (A) BWNW, (B) DLS, and (C) HUniv12Oct groups, compared with the normal control group. Differential usage of the TRBJ gene segment in the (D) BWNW, (E) DLS, and (F) HUniv12Oct groups, compared with the normal control group. Significant differences were analyzed statistically by unpaired *t*-test and FDR correction. All of the TRBVs and TRBJs that were found to differ significantly ($P < 1 \times 10^{-4}$) between COVID-19 patients and healthy controls are presented. Data are presented as the mean ± standard deviation values. Heat maps of TRBV gene usage in conjunction with TRBJ usage showing preferred TRBV–TRBJ pairs in patients with COVID-19 in the (G) BWNW and (H) DLS groups, and (I) healthy controls.

dependent on its composition. To further assess how commonly TCR sequences were shared among the different individuals in each group, overlap indices were calculated for TCRβ amino acid repertoires. The results showed a high degree of sharing of TCRβ amino acid sequences among COVID-19 patients (Figure 5A–C). It was found that a large number of the TCRβ amino acid sequences within an individual were shared with at least one of the other donors in each group (259 669, HC; 724 559, BWNW; 1 653 741, HUniv12Oct; 2 838 484, DLS). Thus, the extent of TCRβ sharing between larger groups of individuals should be potentially much greater. Moreover, it was found that 365, 58, 1297, and 140 TCRβ amino acid sequences were shared by 50% of the individuals in the HC, DLS, BWNW, and HUniv12Oct groups, respectively. In addition, the levels of overlap between any two samples in HCs and COVID-19 patients were calculated and displayed using heat maps (Figure 5D, E).

3.5. Disease-associated TCRβ clones can distinguish COVID-19 patients from healthy controls

A particular interest of this study was the appearance of SARS-CoV-2-associated T-cells in the peripheral blood and their potentiality to serve as biomarkers for COVID-19 diagnosis. COVID-19-associated clones are usually abundantly and widely represented in COVID-19 patients but rarely appear in HCs. As supporting evidence, 15 TCRβ clones that were widely represented in the DLS, BWNW, and HUniv12Oct groups but were rare in the HC group were identified, and it was found that the unique (irrespective of each clone's abundance) and total (including the abundance of each clone) disease-associated TCRβ clone frequencies in PBMCs were significantly higher in COVID-19 patients than in HCs (Figure 6A–F). Overall, these disease-associated TCRβ clones separated COVID-19 patients from HCs with an accuracy of 96.76%,

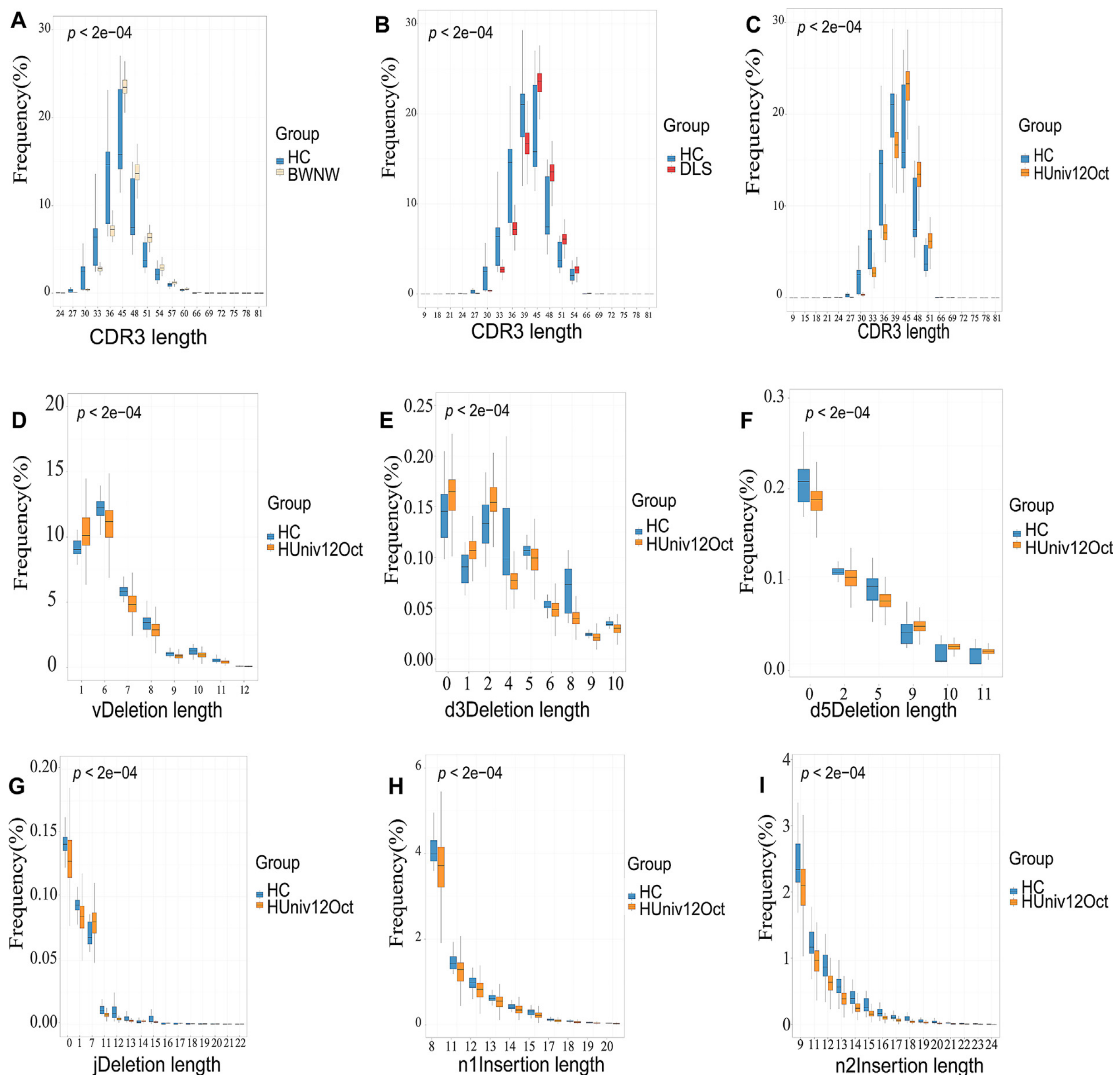


Figure 4. TCRβ CDR3s are longer and show abnormal InDel (insertion–deletion) length in rearrangement positions among COVID-19 patients. In comparison with the healthy controls, a significant increase in CDR3 length was observed in COVID-19 patients in the (A) BWNW group, (B) DLS group, and (C) HUniv12Oct group, represented by higher frequencies of long TCRβ CDR3s (and lower frequencies of short ones) in patients with COVID-19. The insertion (or deletion) lengths clearly differed between the COVID-19 patients in the HUniv12Oct group and healthy controls at the six rearrangement sites: (D) Vdel, (E) D3del, (F) D5del, (G) Jdel, (H) n1ins, (I) n2ins. All of the CDR3 lengths, insertion (or deletion) lengths that differed significantly ($p < 2 \times 10^{-4}$) between COVID-19 patients and healthy controls are presented. Data are presented as the mean ± standard deviation values; comparisons were made using an unpaired *t*-test.

97.67%, and 93.98% in the DLS group, BWNW group, and HUniv12Oct group, respectively (Figure 6G–I), supporting the possible use of these disease-associated TCRβ clones as biomarkers for COVID-19.

Considering diagnostic strategies, when would be the best time to collect samples? To answer this question, the HUniv12Oct cohort with the COVID-19-associated TCRβ clones was used to evaluate whether these diagnostic strategies were more or less accurate depending on the number of days since diagnosis. Overall, the diagnostic model was highly sensitive and specific for each time period: 0–3 days, 4–9 days, 10–25 days, 26–40 days, and 41+ days post diagnosis (**Supplementary Material** Table S1, Figure S4). The classifier demonstrated 100% specificity and 82.68% sensitivity at 0–3 days post diagnosis and 94.66% specificity and 81.34% sensitivity at 4–9 days post diagnosis, further increasing to 98.46% specificity and 92.3% sensitivity at 10–25 days post diagnosis. Notably, there was some reduced signal at 26–40 days post diagnosis (89.22% specificity and 93.84% sensitivity); this subsequently increased further to 95.38% specificity and 89.22% sensitivity at 41+ days post diagnosis. Therefore, the specificity was highest soon after diagnosis (0–3 days post diagnosis) and lowest at 26–40 days post diagnosis.

In addition, it was found that there was a high degree of overlap of each group's 15 disease-associated TCRβ amino acid sequences between the BWNW group, DLS group, and HUniv12Oct group (Figure 7A). Among them, five clones were shared by the three groups. After removing the repeat sequence, there were 25 unique COVID-19-associated clones among the three global collaborators. It is worth noting that two prominent amino acid motifs were identified in these disease-associated clones: L/E-G-

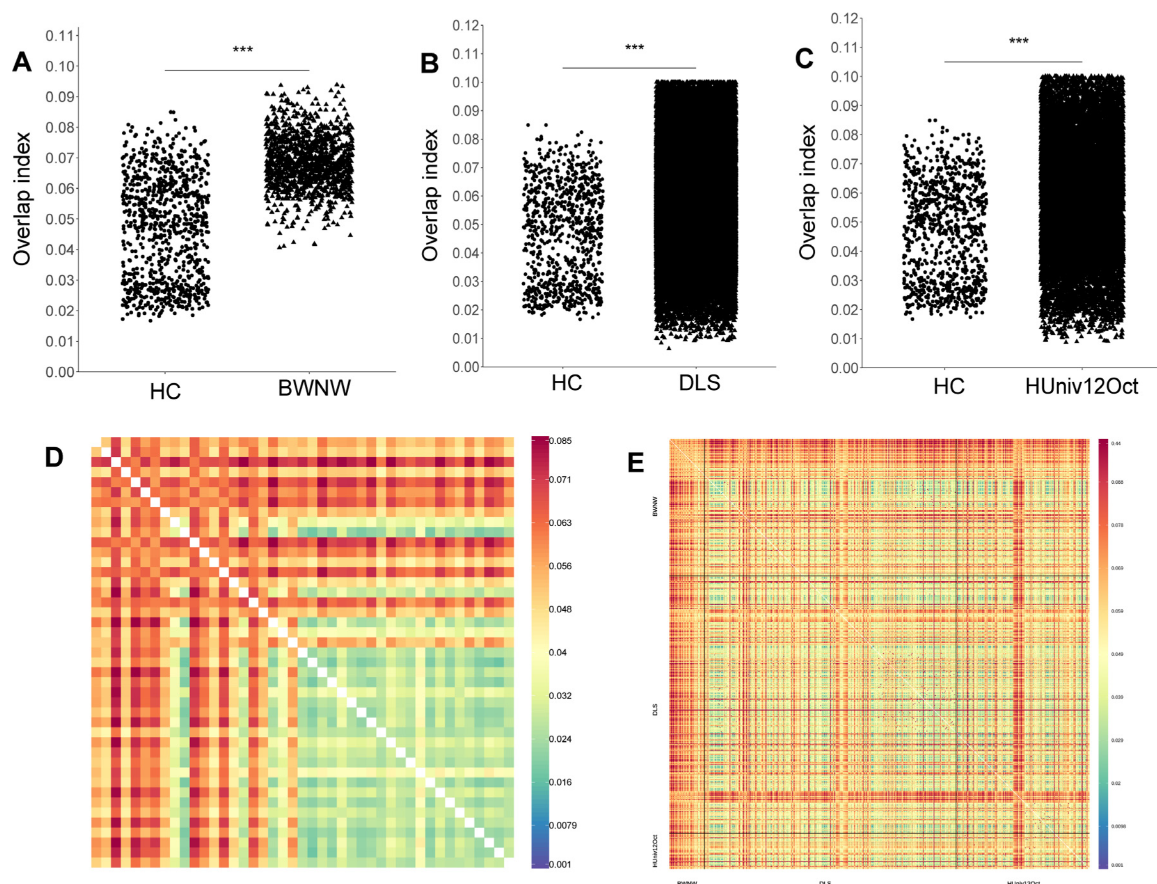


Figure 5. TCRβ CDR3s are highly shared across COVID-19 patients. The overlap indices of TCRβ CDR3 amino acid sequences were found to be significantly larger in COVID-19 patients in the (A) BWNW group, (B) DLS group, and (C) HUniv12Oct group, compared with the healthy controls. The level of overlap between any two samples in (D) healthy controls, and (E) COVID-19 patients. Data were compared using an unpaired *t*-test; $^{***}P < 0.001$ (two-tailed).

S-N and R/P-G-G/Q (Figure 7B). Besides, the phylogenetic tree showed that these 25 disease-associated clones are related to each other (Figure 7C). Remarkably, the five clones (CASSPWTGQETQYF, CASSLNRAGNTIYF, CASSPGGRGNQPQHF, CASSARLAGGTDTQYF, CASSVGRGSYNEQFF) shared by the three global collaborators belonged to four major branches in the phylogenetic tree.

Subsequently, a new independent cohort of COVID donors ($n = 538$), HCs ($n = 36$), and rheumatoid arthritis ($n = 65$) patients was used to prove the potential use of these disease-associated TCRβ clones as biomarkers of SARS-CoV-2 infection. The test results showed that 11 disease-associated TCRβ clones could separate COVID-19 patients from HCs and rheumatoid arthritis patients with an accuracy of 83.75% (**Supplementary Material** Figure S5); these disease-associated TCRβ clones were CASSRGGSSGNTIYF, CASSLQGASEKLFF, CASSFRSSYNSPLHF, CASSLNRAGNTIYF, CASSIRGQPQHF, CASSLLVNTGELFF, CASSVGRGSYNEQFF, CASSPGGRGNQPQHF, CASSPGQEYGYTF, CASSPGITDTQYF, CASSTGVGNTIYF.

4. Discussion

As of June 20, 2021, COVID-19, caused by SARS-CoV-2, had affected over 179 060 045 people, killing more than 3.87 million. As with any virus, the innate and adaptive immune system plays a critical role in clearing SARS-CoV-2 infection (Nussing et al., 2018; Walker 2019), while a failed immune response might result in viral spreading, cytokine storm, and a high mortality rate (Nussing et al., 2018; Walker 2019). In the present study, it was sought to comprehensively characterize the composition and diversity of the TCR repertoire in PBMCs of COVID-19 patients using immunoSEQ Technology. This study comprised a large sample size multi-center randomized controlled trial, and aimed at clustering T-cells relevant for immunity against SARS-CoV-2 (Nolan et al., 2020). The results showed that the TCRβ diversity was clearly lower in COVID-19 patients when compared to HCs, and that TCR repertoire diversity decreases with increasing age. This confirms the results of Britanova et al., who found that TCRβ diversity of naïve T-cells was 60–120 million for individuals in the first two decades of life, decreasing to 8–57 million in individuals over 70 years old (Britanova et al., 2016). This may be a contributing factor in the lower mortality rate found in young and middle-aged COVID-19 patients when compared to elderly patients with COVID-19 (Gutierrez et al., 2020).

The distribution of CDR3 sequence lengths is another key feature that provides an integrative view of repertoire composition. Biases in CDR3 length are often observed in epitope-specific T-cell repertoires (Pickman et al., 2013; Liu et al., 2017; Huang et al., 2019). Indeed, it was found in the present study that there was a shift towards longer clonotypes in COVID-19 patients when compared to HCs. These results appear to confirm previous findings that virus-specific TCRβ clonotypes show increased TCRβ CDR3 length when compared to autoantigen-specific clonotypes (Gomez-Tourino et al., 2017). Moreover, our experimental results showed that the length distribution of the six recombination events (indels) differed obviously between the COVID-19 patients and HCs. Different rearrangements may lead to variable CDR3 lengths. TCR CDR3 loops can vary in both length and sequence, allowing diverse antigens to be recognized.

Furthermore, the analysis results showed that the usage frequency of the TRBV/TRBJ segments differed noticeably between COVID-19 patients and HCs. The skewed use of the TRBV/J segments may be associated with the immune dysfunction and the

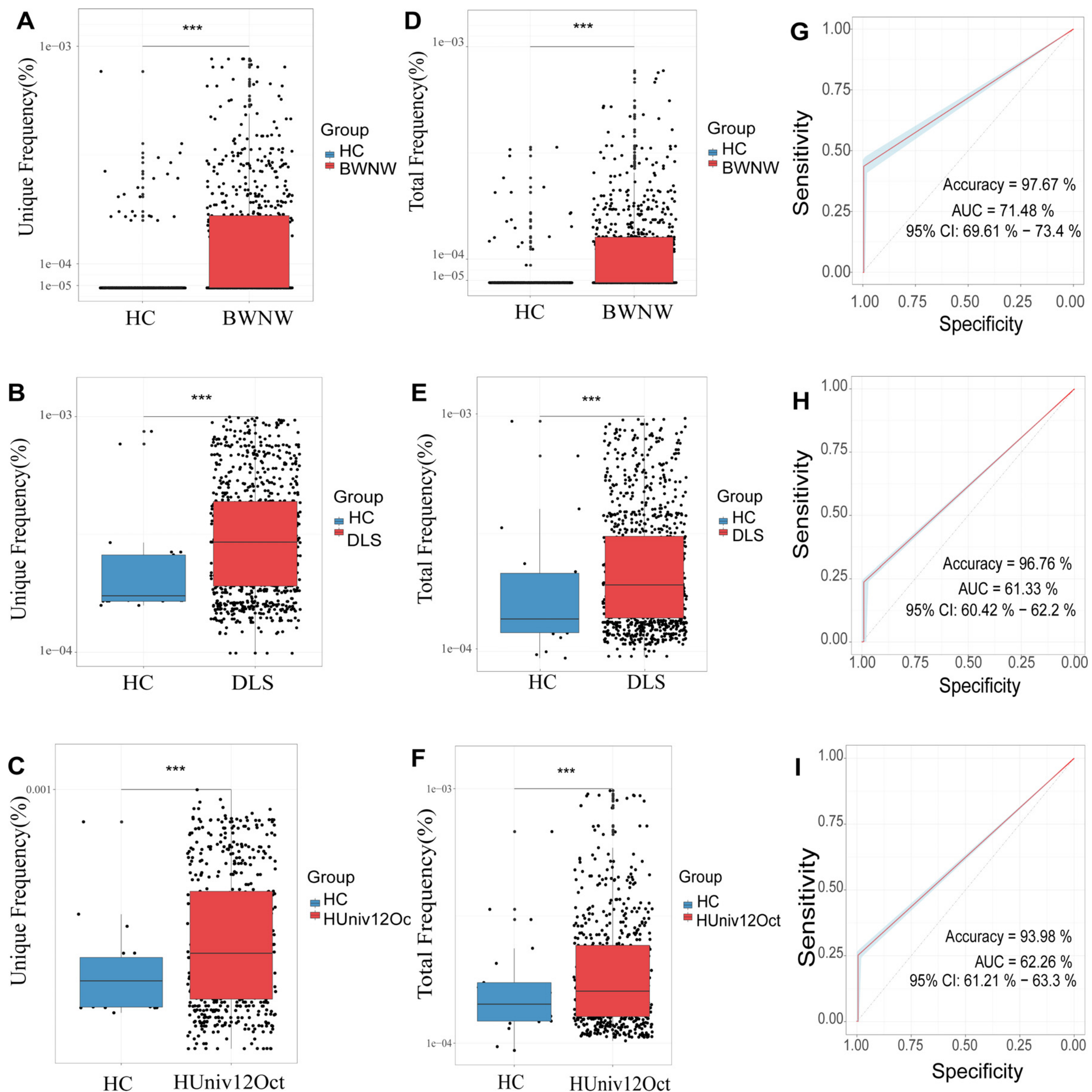


Figure 6. Disease-associated TCRβ clones can distinguish patients with COVID-19 from healthy controls. The unique frequency (irrespective of each clonotype frequency) and total frequency (including the abundance of each clonotype) of disease-associated TCRβ clones were found to be significantly higher in the peripheral blood mononuclear cells of COVID-19 patients in the (A, D) BWNW group, (B, E) DLS group, and (C, F) HUniv12Oct group than in the healthy controls. ROC curves of the classification of healthy controls and COVID-19 patients in the (G) BWNW group, (H) DLS group, and (I) HUniv12Oct group according to disease-associated TCRβ clones. Data are presented as the mean ± standard deviation values; comparisons were made using the unpaired *t*-test. $^{***}P < 0.001$.

pathogenesis of the disease. In the case of disease, stimulation by SARS-CoV-2 antigens can lead to the targeted rearrangement and excessive abnormal cloning of one or a few TRBV subfamilies, and the cloning of other T-cells may be suppressed by the dominant T-cell clones, which may result in impaired immune function and decreased ability to clear the virus (Li et al., 2011; Sui et al., 2015). Relevant to this investigation, Wen et al. applied single-cell RNA sequencing to characterize the changes in PBMCs from 10 COVID-19 patients. They identified an over-representation of the IGHV3 family in COVID-19 patients compared to HCs, especially IGHV3-21, IGHV3-7, IGHV3-30, IGHV3-15, and IGHV3-23 (Wen et al., 2020). The skewed use of the TRBV/IGHV genes offers a framework for the rational design of SARS-CoV-2 vaccines.

In addition, we found that the degree of overlap of the TCR repertoire was significantly higher in COVID-19 patients compared with HCs. TCRβ clonotypes that are shared between individuals are likely raised against common antigens, and are thought to play an essential role in the efficacy of pathogen-specific responses and the control of infection (Price et al., 2009; Venturi et al., 2011; Zhao et al., 2016). Similarly, a recent study demonstrated the presence of influenza-specific public T-cells in both infected patients and HCs (Price et al., 2009; Venturi et al., 2011; Zhao et al., 2016). Thus, if linked to certain infections, such TCRs could become invaluable tools for immunodiagnosis of human disease and vaccine development. Moreover, we identified a set of SARS-CoV-2-associated TCRβ that can distinguish patients with COVID-19 from HCs with an accuracy rate of more than 93%. Overall, this study demonstrates the potential of disease-associated TCRβ clones as alternative biomarkers for the screening and diagnosis of COVID-19. However, due to the limited number of samples in this analysis, validation of the performance of these biomarkers and the evaluation of the accuracy and reliability of this method are required through the performance of more studies in other laboratories. In this study, the HUniv12Oct cohort with the COVID-19-

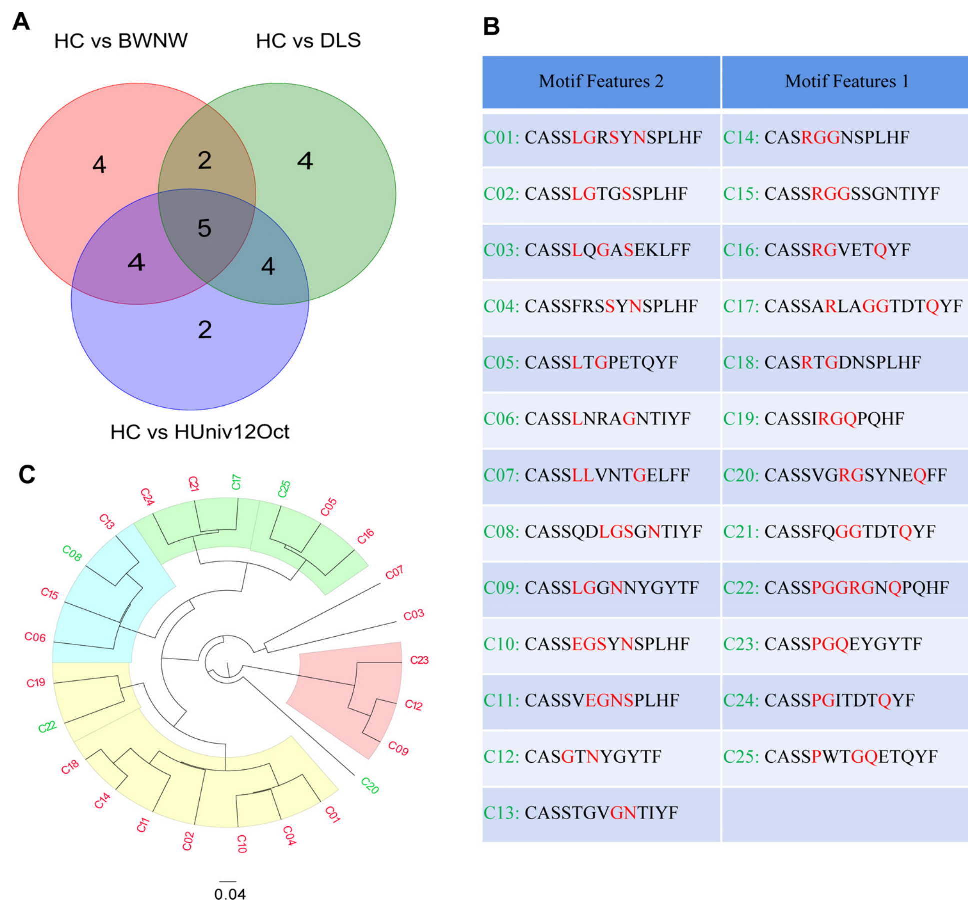


Figure 7. Repertoire characteristics of COVID-19-associated clones. (A) The degree of overlap of each group's 15 disease-associated TCRβ amino acid sequences between the BWNW group, DLS group, and HUniv12Oct group. The number of amino acid sequences unique to each group is shown in the non-overlapping sections, and the number of amino acid sequences common to two and to three groups is indicated in the relevant overlapping areas. (B) Two motif features are observed in the 25 non-repetition COVID-19-associated clones. (C) The evolutionary tree of the 25 COVID-19-associated clones. The five COVID-19-associated clones that are common to the three groups are marked in green.

associated TCRβ clones was used to evaluate whether these diagnostic strategies are more or less accurate depending on the number of days post diagnosis. It was found that the diagnostic model was highly sensitive and specific in each time period, with around 95–100% specificity over the three periods of 0–3 days, 4–9 days, and 10–25 days post diagnosis. Relevant to this investigation, Snyder et al. (Snyder et al., 2020) also trained a classifier to diagnose SARS-CoV-2 infection based solely on TCR sequencing from blood samples, and at 99.8% specificity they observed high early sensitivity soon after diagnosis (day 3–7 = 85.1%; day 8–14 = 94.8%), as well as lasting sensitivity after recovery (day 29+/convalescent = 95.4%).

It is worth noting that the potential role of cross-reactive immunity may affect the diagnostic effect. Spike (S) proteins of SARS-CoV-2 share about 97% and 76% amino acid identity with coronavirus RaTG13 and SARS-CoV, respectively (Ou et al., 2020). Many previous studies have highlighted significant serological cross-reactivity between SARS-CoV-2 and other coronaviruses (MERS-CoV, SARS-CoV, HCoV-OC43, HCoV-HKU1, HCoV-229E, RaTG13, and HCoV-NL63) (Mateus et al., 2020; Hicks et al., 2021; Liu et al., 2021) and dengue virus (Lustig et al., 2020), among others. Cross-reactivity between SARS-CoV-2 and other viruses may interfere with accurate clinical diagnosis and lead to false-positive dengue serology among COVID-19 patients. However, another study involving a similar experimental design produced contradictory results: Ou et al. demonstrated limited cross-neutralization between convalescent sera from severe acute respiratory syndrome (SARS) patients and COVID-19 patients (Ou et al., 2020). In short, further research is needed to verify the accuracy of diagnosis based on TCRβ clones.

In summary, this study presents a comprehensive overview of the TCRβ CDR3 repertoire in COVID-19 patients, including the reduced TCRβ diversity, increased CDR3 length, skewed usage of TRBV/J, and high degree of TCRβ sharing. The most important discovery is that using the defined COVID-19-associated TCRβ clones, COVID-19 patients could be distinguished from healthy individuals. These findings demonstrate that disease-associated TCRβ clonotypes could work as potential biomarkers to help diagnose COVID-19, at least in COVID-19 screening.

Declarations

Funding: This work was Supported by the China Postdoctoral Science Foundation (2021M691239), the Guangxi Natural Science Foundation (2019GXNSFBA245032), the Guangxi Science and Technology Plan Project (Gui Ke AD20238021), the Guangxi Natural Science Foundation (2020GXNSFDA297027), the open funds of the Guangxi Key Laboratory of Tumor Immunology and Microenvironmental Regulation (2020KF010), the Guilin Science Research and Technology Development Project (20190218-5-5), and the Research Capability Improvement Project for Young and Middle-aged teachers in Guilin Medical University (2018glcy09).

Ethical approval: This ImmuneRACE study was approved by Western Institutional Review Board (WIRB reference number 1-

1281891- 1, Protocol ADAP-006). All participants were consented for sample collection and metadata use via electronic informed consent processes.

Conflict of interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Data availability

All sequencing data have been deposited and made public at the Adaptive Biotechnologies immuneACCESS site (https://clients.adaptivebiotech.com/pub/ covid-2020). All other data are available from the authors upon reasonable request.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijid.2021.10.033.

References

Ahn DG, Shin HJ, Kim MH, Lee S, Kim HS, Myoung J, et al. Current Status of Epidemiology, Diagnosis, Therapeutics, and Vaccines for Novel Coronavirus Disease 2019 (COVID-19). J Microbiol Biotechnol 2020;30:313–24. doi:10.4014/jmb.2003.03011.

Baidu, Novel coronavirus pneumonia, Real time big data report of epidemic situation https://voice.baidu.com/act/newpneumonia/newpneumonia (accessed 15 August 2021).

Britanova OV, Shugay M, Merzlyak EM, Staroverov DB, Putintseva EV, Turchaninova MA, et al. Dynamics of Individual T Cell Repertoires: From Cord Blood to Centenarians. J Immunol 2016;196:5005–13. doi:10.4049/jimmunol.1600005.

Carlson CS, Emerson RO, Sherwood AM, Desmarais C, Chung MW, Parsons JM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. Nat Commun 2013;4:2680. doi:10.1038/ncomms3680.

Channappanavar R, Zhao J, Perlman S. T cell-mediated immune response to respiratory coronaviruses. Immunol Res 2014;59:118–28. doi:10.1007/s12026-014-8534-z.

Elhanati Y, Sethna Z, Callan CJ, Mora T, Walczak AM. Predicting the spectrum of TCR repertoire sharing with a data-driven model of recombination. Immunol Rev 2018;284:167–79. doi:10.1111/imr.12665.

Gil A, Kamga L, Chirravuri-Venkata R, Aslan N, Clark F, Ghersi D, et al. Epstein-Barr Virus Epitope-Major Histocompatibility Complex Interaction Combined with Convergent Recombination Drives Selection of Diverse T Cell Receptor alpha and beta Repertoires. mBio 2020;11. doi:10.1128/mBio.00250-20.

Gomez-Tourino I, Kamra Y, Baptista R, Lorenc A, Peakman M. T cell receptor beta-chains display abnormal shortening and repertoire sharing in type 1 diabetes. Nat Commun 2017;8:1792. doi:10.1038/s41467-017-01925-2.

Guo YR, Cao QD, Hong ZS, Tan YY, Chen SD, Jin HJ, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak - an update on the status. Mil Med Res 2020;7:11. doi:10.1186/s40779-020-00240-0.

Gutierrez L, Beckford J, Alachkar H. Deciphering the TCR Repertoire to Solve the COVID-19 Mystery. Trends Pharmacol Sci 2020;41:518–30. doi:10.1016/j.tips.2020.06.001.

Hicks J, Klumpp-Thomas C, Kalish H, Shunmugavel A, Mehalko J, Denson JP, et al. Serologic cross-reactivity of SARS-CoV-2 with endemic and seasonal Betacoronaviruses. J Clin Immunol 2021;41:906–13. doi:10.1101/2020.06.22.20137695.

Hou X, Yang Y, Chen J, Jia H, Zeng P, Lv L, et al. TCRbeta repertoire of memory T cell reveals potential role for Escherichia coli in the pathogenesis of primary biliary cholangitis. Liver Int 2019;39:956–66. doi:10.1111/liv.14066.

Hou XL, Wang L, Ding YL, Xie Q, Diao HY. Current status and recent advances of next generation sequencing techniques in immunological repertoire. Genes Immun 2016;17:153–64. doi:10.1038/gene.2016.9.

Huang C, Li X, Wu J, Zhang W, Sun S, Lin L, et al. The landscape and diagnostic potential of T and B cell repertoire in Immunoglobulin A Nephropathy. J Autoimmun 2019;97:100–7. doi:10.1016/j.jaut.2018.10.018.

Lai CC, Shih TP, Ko WC, Tang HJ, Hsueh PR. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): The epidemic and the challenges. Int J Antimicrob Agents 2020;55. doi:10.1016/j.ijantimicag.2020.105924.

Li Y, Geng S, Du X, Chen S, Yang L, Wu X, et al. Restricted TRBV repertoire in CD4+ and CD8+ T-cell subsets from CML patients. Hematology 2011;16:43–9. doi:10.1179/102453311X12902908411634.

Liu K, Pan X, Li L, Yu F, Zheng A, Du P, et al. Binding and molecular basis of the bat coronavirus RaTG13 virus to ACE2 in humans and other species. Cell 2021;184:3438–51 e10. doi:10.1016/j.cell.2021.05.031.

Liu S, Hou XL, Sui WG, Lu QJ, Hu YL, Dai Y. Direct measurement of B-cell receptor repertoire's composition and variation in systemic lupus erythematosus. Genes Immun 2017;18:22–7. doi:10.1038/gene.2016.45.

Lustig Y, Keler S, Kolodny R, Ben-Tal N, Atias-Varon D, Shlush E, et al. Potential antigenic cross-reactivity between SARS-CoV-2 and Dengue viruses. Clin Infect Dis 2020:ciaa1207. doi:10.1093/cid/ciaa1207.

Mateus J, Grifoni A, Tarke A, Sidney J, Ramirez SI, Dan JM, et al. Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. Science 2020;370:89–94. doi:10.1126/science.abd3871.

Murugan A, Mora T, Walczak AM, Callan CJ. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. Proc Natl Acad Sci U S A 2012;109:16161–6. doi:10.1073/pnas.1212755109.

Nolan S, Vignali M, Klinger M, Dines JN, Kaplan IM, Svejnoha E, et al. A large-scale database of T-cell receptor beta (TCRbeta) sequences and binding associations from natural and synthetic exposure to SARS-CoV-2. Res Sq 2020 rs.3.rs-51964. doi:10.21203/rs.3.rs-51964/v1.

Nussing S, Sant S, Koutsakos M, Subbarao K, Nguyen T, Kedzierska K. Innate and adaptive T cells in influenza disease. Front Med 2018;12:34–47. doi:10.1007/s11684-017-0606-8.

Ou X, Liu Y, Lei X, Li P, Mi D, Ren L, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nat Commun 2020;11:1620. doi:10.1038/s41467-020-15562-9.

Pickman Y, Dunn-Walters D, Mehr R. BCR CDR3 length distributions differ between blood and spleen and between old and young patients, and TCR distributions can be used to detect myelodysplastic syndrome. Phys Biol 2013;10. doi:10.1088/1478-3975/10/5/056001.

Pogorelyy MV, Minervina AA, Chudakov DM, Mamedov IZ, Lebedev YB, Mora T, et al. Method for identification of condition-associated public antigen receptor sequences. Elife 2018;7:e33050. doi:10.7554/eLife.33050.

Price DA, Asher TE, Wilson NA, Nason MC, Brenchley JM, Metzler IS, et al. Public clonotype usage identifies protective Gag-specific CD8+ T cell responses in SIV infection. J Exp Med 2009;206:923–36. doi:10.1084/jem.20081127.

Robins H, Desmarais C, Matthis J, Livingston R, Andriesen J, Reijonen H, et al. Ultra-sensitive detection of rare T cell clones. J Immunol Methods 2012;375:14–19. doi:10.1016/j.jim.2011.09.001.

Snyder TM, Gittelman RM, Klinger M, May DH, Osborne EJ, Taniguchi R, et al. Magnitude and Dynamics of the T-Cell Response to SARS-CoV-2 Infection at Both Individual and Population Levels. medRxiv [Preprint]. 2020: 2020. 07. 31. 20165647. doi: 10.1101/2020.07.31.20165647.

Stewart JJ, Lee CY, Ibrahim S, Watts P, Shlomchik M, Weigert M, et al. A Shannon entropy analysis of immunoglobulin and T cell receptor. Mol Immunol 1997;34:1067–82. doi:10.1016/s0161-5890(97)00130-2.

Sui W, Hou X, Zou G, Che W, Yang M, Zheng C, et al. Composition and variation analysis of the TCR beta-chain CDR3 repertoire in systemic lupus erythematosus using high-throughput sequencing. Mol Immunol 2015;67:455–64. doi:10.1016/j.molimm.2015.07.012.

Venturi V, Kedzierska K, Turner SJ, Doherty PC, Davenport MP. Methods for comparing the diversity of samples of the T cell receptor repertoire. J Immunol Methods 2007;321:182–95. doi:10.1016/j.jim.2007.01.019.

Venturi V, Quigley MF, Greenaway HY, Ng PC, Ende ZS, McIntosh T, et al. A mechanism for TCR sharing between T cell subsets and individuals revealed by pyrosequencing. J Immunol 2011;186:4285–94. doi:10.4049/jimmunol.1003898.

Walker CM. Adaptive Immune Responses in Hepatitis A Virus and Hepatitis E Virus Infections. Cold Spring Harb Perspect Med 2019;9. doi:10.1101/cshperspect.a033472.

Wen W, Su W, Tang H, Le W, Zhang X, Zheng Y, et al. Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. Cell Discov 2020;6:31. doi:10.1038/s41421-020-0168-9.

Zhao Y, Nguyen P, Ma J, Wu T, Jones LL, Pei D, et al. Preferential Use of Public TCR during Autoimmune Encephalomyelitis. J Immunol 2016;196:4905–14. doi:10.4049/jimmunol.1501029.

Zhang F, Gan R, Zhen Z, Hu X, Li X, Zhou F, et al. Adaptive immune responses to SARS-CoV-2 infection in severe versus mild individuals. Signal Transduct Target Ther 2020;5:156. doi:10.1038/s41392-020-00263-y.