Identification, crystallization and epitope determination of public TCR shared and expanded in COVID-19 patients

Xiuyuan Lu¹, Yuki Hosono^{1,2,3}, Shigenari Ishizuka^{1,2}, Masamichi Nagae^{1,2}, Eri Ishikawa^{1,2}, Daisuke Motooka⁴, Yuki Ozaki⁴, Nicolas Sax⁵, Ryo Shinnakasu⁶, Takeshi Inoue⁶, Taishi Onodera⁷, Takayuki Matsumura⁷, Masaharu Shinkai⁸, Takashi Sato⁸, Shota Nakamura⁴, Shunsuke Mori^{9,10}, Teru Kanda¹¹, Emi E. Nakayama¹², Tatsuo Shioda¹², Tomohiro Kurosaki^{6,13}, Hisashi Arase^{9,10}, Kazuo Yamashita⁵, Yoshimasa Takahashi⁷, Sho Yamasaki^{1,2,14,15,*}

¹ Laboratory of Molecular Immunology, Immunology Frontier Research Center, Osaka University, Suita, 565-0871, Japan

² Department of Molecular Immunology, Research Institute for Microbial Diseases, Osaka University, Suita, 565-0871, Japan

³ Department of Respiratory Medicine and Clinical Immunology, Graduate School of Medicine, Osaka University, Suita, 565-0871, Japan.

⁴ Department of Infection Metagenomics, Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University, Suita, 565-0871, Japan

⁵ KOTAI Biotechnologies, Inc, Suita, 565-0871, Japan

⁶ Laboratory of Lymphocyte Differentiation, Immunology Frontier Research Center, Osaka University, Suita, 565-0871, Japan

⁷ Department of Immunology, National Institute of Infectious Diseases, Tokyo, 162-8640, Japan

⁸ Tokyo Shinagawa Hospital, Tokyo, 140-8522, Japan

⁹ Department of Immunochemistry, Research Institute for Microbial Diseases, Osaka University, Suita, 565-0871, Japan

¹⁰ Laboratory of Immunochemistry, Immunology Frontier Research Center, Osaka University, Suita, 565-0871, Japan

¹¹ Division of Microbiology, Faculty of Medicine, Tohoku Medical and Pharmaceutical University, Sendai, 983-8536, Japan

¹² Department of Viral Infections, Research Institute for Microbial Diseases, Osaka

University, Suita, 565-0871, Japan

¹³ Laboratory of Lymphocyte Differentiation, RIKEN Center for Integrative Medical Sciences, Yokohama, 230-0045, Japan

¹⁴ Division of Molecular Immunology, Medical Mycology Research Center, Chiba University, Chiba, 260-8673, Japan

¹⁵ Division of Molecular Design, Research Center for Systems Immunology, Medical Institute of Bioregulation, Kyushu University, Fukuoka, 812-8582, Japan

*Corresponding to: yamasaki@biken.osaka-u.ac.jp

Summary

T cells play pivotal roles in protective immunity against SARS-CoV-2 infection. Follicular helper T (Tfh) cells mediate the production of antigen-specific antibodies; however, T cell receptor (TCR) clonotypes used by SARS-CoV-2-specific Tfh cells have not been well characterized. Here, we first identified and crystallized public TCR of Tfh clonotypes that are shared and expanded in unhospitalized COVID-19-recovered patients. These clonotypes preferentially recognized SARS-CoV-2 spike (S) protein epitopes which are conserved among emerging SARS-CoV-2 variants. These clonotypes did not react with S proteins derived from common cold human coronaviruses, but cross-reacted with symbiotic bacteria, which might confer the publicity. Among SARS-CoV-2 S epitopes, S₈₆₄₋₈₈₂, presented by frequent HLA-DR alleles, could activate multiple public Tfh clonotypes in COVID-19-recovered patients. Furthermore, S₈₆₄₋₈₈₂-loaded HLA tetramer preferentially bound to CD4⁺ T cells expressing CXCR5. In this study, we identified and crystallized public TCR for SARS-CoV-2 that may contribute to the prevention of COVID-19 aggravation.

Keywords

Public TCR, Tfh, epitope, SARS-CoV-2, spike, commensal bacteria, cross-reactive antigen

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused a worldwide pandemic with emerging mutations and continuous development of robust vaccine is imperative. Neutralizing antibodies are a major component of a protective immunity against SARS-CoV-2 (Tay et al., 2020), and thus information about virus-specific CD4⁺ T cells that help B cells to produce antigen-specific antibodies is needed. Among CD4⁺ T cells, follicular helper T (Tfh) cells mediate germinal center B cells maturation and production of high affinity antibodies (Crotty, 2019). Some deceased COVID-19 patients exhibited incomplete Tfh cell development (Kaneko et al., 2020), suggesting that efficient Tfh induction is a key to protective immunity against SARS-CoV-2 (Gong et al., 2020; Zhang et al., 2020). Furthermore, the characterization of SARS-CoV-2-specific Tfh cells bearing common TCR shared between different individuals will help us to understand and predict host responses during infection. However, such public Tfh clonotypes specific for SARS-CoV-2 have not yet been reported.

Given that V(D)J recombination is a random process, presence of shared TCR between individuals is generally thought to be a rare event (Soto et al., 2020; Venturi et al., 2008). Common antigens are one of the reasons that maintain such public clonotypes in the periphery (Li et al., 2012; Venturi et al., 2008). Recently, some T cells recognizing common cold human coronaviruses (HCoVs) are reported to cross-react with SARS-CoV-2 (Braun et al., 2020; Mateus et al., 2020; Woldemeskel et al., 2020), whereas SARS-CoV-2-specific T cells that do not react with HCoVs are frequently

detected (Bacher et al., 2020; Mateus et al., 2020). Currently, other cross-reactive antigens recognized by SARS-CoV-2-specific T cells have not been well examined.

In this study, using single cell-based paired TCR analysis together with a rapid TCR reconstitution and epitope determination platform, we identified 1) public TCRs of Tfh cells specific for SARS-CoV-2 expanded in convalescent patients, 2) their epitope peptides and restricting HLA alleles, and 3) their cross-reactive proteins from symbiotic bacteria.

Results

Identification of SARS-CoV-2 spike protein-specific Tfh clones in convalescent COVID-19 patients

We first analyzed SARS-CoV-2-specific T cell subsets and their clonotypes using a single cell-based RNA sequencing platform. Peripheral blood mononuclear cells (PBMCs) isolated from healthy donors and convalescent COVID-19 patients (Table S1 and S2) were stimulated with antigens derived from SARS-CoV-2, including inactivated virus, recombinant S protein, overlapped peptide pools derived from S protein (S peptide pool) or membrane (M) and nucleocapsid (N) proteins (MN peptide pool). Activation marker-positive T cells were sorted and analyzed for their T cell receptor (TCR) sequences together with RNA expression by single cell TCR-seq and RNA-seq analyses (Figure S1). UMAP embedding and clustering allowed identification of a cluster (#8) consisting of CD4⁺ cells expressing Tfh-related genes, such as *CD200*, *PDCD1*, *ICOS*, *CXCL13*, *CD40LG* and *CXCR5*, suggesting that cluster #8 includes circulating Tfh cells (Crotty, 2019; Schmitt et al., 2014) (Figure 1A and 1B). Within this Tfh cluster, we could identify 1735 TCRs (Table S3). Furthermore, in the S-reactive Tfh cluster, we detected clones bearing TCRαβ pairs shared between different patients, Ts-017 and Ts-018, who had antibodies against receptor binding domain (RBD) of S protein and neutralizing antibodies (Table S1). Considering the cell number of single cell TCR-seq, these are likely to be frequent public TCRs and we designated them as TCR-017 and TCR-018, respectively. TCR-017 and -018 had the same Vα, Jα, Vβ and Jβ usages and their complementarity-determining region 3

(CDR3) sequences were identical except for one amino acid in CDR3 β (Figure 1C and 1D). TCR-018 was detected in patient Ts-018 as three different barcoded clones within S protein/peptides-reactive Tfh cluster (Table S3).

Tfh clones derived from different patients recognized an identical S epitope restricted by the same HLAs

To examine the antigen specificity of TCR-017 and -018, the respective TCR α and β chains were reconstituted in a TCR-deficient T cell hybridoma bearing an NFAT-GFP reporter (Figure 1E). TCR-transfectants were stimulated with SARS-CoV-2 antigens in the presence of transformed B cells derived from the same patients as autologous antigen-presenting cells (APCs). Cells expressing TCR-017 and -018 responded to recombinant S protein and S peptide pool, but not to MN peptide pool (Figure 2A), which is consistent with the initial antigen specificity revealed in single cell analysis (Table S3). Among two halves of pooled S peptides, #1 (S₁₋₆₄₃) and #2 (S₆₃₃₋₁₂₇₃), only #2 pool reacted with both TCRs (Figure 2B). However, another SARS-CoV-2 S peptide pool (S₃₀₄₋₃₃₈, 421-475, 492-519, 683-707, 741-770, 785-802, 885-1273) did not activate either clone (Figure S2A and Figure 2B). The above results suggest that the antigen epitope(s) of TCR-017 and -018 is located in regions 633-682, 708-740, 771-784 or 803-884 of S protein (Figure S2A).

To further reduce the number of epitope candidates, we next determined the HLA alleles that restrict TCR-017 and -018. Heterologous APCs from patient Ts-017 could activate TCR-018, and *vice versa*, demonstrating that these APCs are

interchangeable (Figure S2B). Thus, the shared HLA class II alleles, DRA-DRB1*15:01, DQA1*01:02-DQB1*06:02 or DPA1*02:02-DPB1*05:01 (shown in red in Table S2), are likely to restrict this recognition. Since DRB1*15:01 and DQB1*06:02 are genetically linked and would not be segregated from each other (Begovich et al., 1992), these HLA alleles were separately transfected into HEK293T cells to examine their ability to present S peptides to TCR-017 and -018. Only cells transfected with DRA and DRB1*15:01 could activate both clones in the presence of S peptide pool #2 (Figure 2C), suggesting that the reactivity is determined by DRB1*15:01, as DRA is monomorphic. Two epitopes, S₈₂₈₋₈₄₆ and S₈₆₄₋₈₈₂, were predicted as strong binders to DRA-DRB1*15:01 using NetMHC server software (Reynisson et al., 2020) within the candidate regions described above (Figure S2A). Among them, S₈₆₄₋₈₈₂ was identified as the epitope for both clones (Figure 2D). Judging from their dose-response curves, relative affinities of TCR-017 and -018 to this peptide appeared comparable (Figure 2E). Furthermore, antigenic activities of serial overlapping peptides indicated that S₈₇₀₋₈₇₈ is a minimum epitope recognized by these TCRs (Figure 2F). Taken together, both TCR-017 and -018 recognized the same epitope on the same HLA with similar affinities, indicating that TCR-017 and -018 are considered as one clonotype, clonotype-017/018, shared by different donors.

Among other HLAs predicted to strongly bind S₈₆₄₋₈₈₂ (Reynisson et al., 2020), we found that DRB1*15:02/*15:03/*15:04, but not DRB1*16:02 and DQA1*01:01-DQB1*05:01, were restricting HLA that present this epitope to activate clonotype-017/018 (Figure S2C and S2D).

Public clonotypes significantly expanded in recovered-COVID-19 patients preferentially recognize conserved S epitopes

Consistent with the high frequency of DRB1*15:02 (10.6%) and *15:01 (7.7%) in the Japanese population (<http://hla.or.jp>), bulk TCR sequencing of pre-pandemic Japanese donors revealed that TCR α and β chains of clonotype-017/018 were broadly shared by 14.8% and 33.3%, respectively, of the population (Figure 2G). As DRB1*15 are not minor alleles worldwide (Gonzalez-Galarza et al., 2020), we analyzed TCR databases of healthy and convalescent COVID-19 cohorts containing donors of various ethnicities (Emerson et al., 2017; Nolan et al., 2020) and found that 17% and 24%, respectively, possessed TCR β -017/018 (Figure 2H). These results suggest that clonotype-017/018 is a public clonotype in multiple ethnicities.

We therefore examined whether any of 1735 paired TCRs detected in our sample pool are found in these cohorts. 855 out of 1735 were revealed to be public Tfh TCRs (Table S3), indicating that TCR-017/018 is not an exceptional clonotype. Among them, we identified 10 Tfh clonotypes which were significantly expanded in recovered COVID-19 patients compared with the healthy cohort (Figure 3A, column J, asterisks). Of note, a half of such clonotypes were indeed detected in multiple patients in our sample pool by single cell TCR sequencing (Figure 3A, column A, #), and TCR-017/018 was the fifth most significantly expanded clonotype (Figure 3A, column A and J).

To efficiently determine the epitopes of these clonotypes, we established a

rapid epitope-determination platform (Figure S3A) and validated it using TCR-017/018 (Figure 3B, clonotype 5). By using this platform, we further determined the epitopes for other COVID-19-expanded Tfh clonotypes (Figure 3B). Restricting HLAs of these epitopes were determined using overlapped HLA alleles in the combination of transformed B cells from different individuals, and individual HLA transfectants (Table S2 and Figure S3B-S3F). These results indicate that the public Tfh clonotypes which significantly expanded in unhospitalized, recovered patients preferentially recognized S epitopes (Figure 3A and 3B). All identified epitopes were located in two major hotspots within the conserved trimer-forming interface of S protein with low mutation rates (Figure 3C)(Elbe and Buckland-Merrett, 2017; Singer et al., 2020). In addition, the donors of these clonotypes had anti-RBD and neutralizing antibodies (Table S1). These results suggest that public Tfh clonotypes specific for conserved S epitopes have expanded during SARS-CoV-2 infection and presumably contributed to the recovery from mild symptoms (Emerson et al., 2017; Nolan et al., 2020).

Crystal structure revealed sequence flexibility of CDR3 that extends the publicity of SARS-CoV-2-specific clonotypes

We demonstrated that TCR β -017/018 is highly public and expanded in the recovered patients (Figure 3A). Since the nonidentical amino acid, Q94 and P94 of TCR β -017 and -018, respectively, have distinct characteristics, we suspected that any amino acid at this position allows epitope recognition. Indeed, any of 17 substitutions, except for D94, retained the reactivity to S₈₆₄₋₈₈₂ epitope on DRA-DRB1*15:01 (Figure

4A). Indeed, crystallographic analysis of TCR-017 complex demonstrated that the residue at position 94 in CDR3 β is located away from CDR3 α (Figure 4B and 4C) and thus unlikely to contact with peptide-MHC. In addition, C90, A91, S92 and S93 were more distant from CDR3 α (Figure 4C), suggesting that this TCR β may mainly utilize a C-terminal half of CDR3 β for recognition and allow diverse TCR sequence variants or even other TRBVs.

This promiscuity within CDR3 β sequence should expand the pool of this public clonotype, which gives us a more precise expansion rate in recovered patients. Indeed, the expansion of the extended clonotype in the COVID-19-recovered cohort became much more significant ($p = 1.6 \times 10^{-6}$) (Figure 4D) than the value shown in Figure 3A ($p = 0.003$). Furthermore, its occurrence became more than half in all three cohorts (Figure 4E), suggesting that the extended clonotype 5 is broadly shared among a large population.

Public Tfh clonotypes did not cross-react with HCoV but with symbiotic bacteria

We next asked how these clonotypes, particularly clonotype 5, acquired publicity. Cross-reactivity with the common cold coronaviruses is one of the possible mechanisms (Braun et al., 2020; Grifoni et al., 2020; Mateus et al., 2020). However, none of the clonotypes identified in this study including TCR-017/018 reacted with S proteins derived from common cold human coronavirus (HCoV) strains (Figure 3B and Figure S3A, S4A and S4B). Indeed, these HCoVs do not have identical sequences to SARS-CoV-2 epitopes determined in Figure 3B (Figure 5A). Furthermore, peptides in

the corresponding regions derived from these coronaviruses including severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome-related coronavirus (MERS-CoV) were not recognized by TCR-017/018 (Figure S4C and S4D). Thus, these clonotypes are unlikely to acquire publicity by cross-reacting with reported human coronaviruses present before the outbreak of COVID-19. This raises a hypothesis that some other common antigen(s) may confer publicity on these clonotypes. In line with this, BLAST homology search of S₈₇₀₋₈₇₈ core epitope revealed that oral commensal bacteria, *Selenomonas noxia*, express a protein, multidrug and toxic compound extrusion (MATE) family efflux transporter, that includes the same epitope with S₈₇₀₋₈₇₉, IAQYTSALLA (Figure 5A). Indeed, the synthetic peptides (MATE₂₄₁₋₂₆₀ and ₂₄₂₋₂₅₆) were recognized by both TCR-017 and -018 on DRB1*15:01 (Figure 5B). In addition, if any amino acids within this S epitope are replaceable, further cross-reactive proteins might be found. To identify such acceptable substitutions, we performed Alanine/Glycine scanning of S₈₇₀₋₈₇₈ peptides and found that seven single substitutions, except for positions Q872 and T874, retained antigenicity of the epitope (Figure 5C). Among these mutated epitopes, the sequence of L877A (IAQYTSAAL) was found in a protein expressed by various bacteria, histidine ammonia-lyase (HAL, EC 4.3.1.3)(Table S4 and Figure 5A).

Other 9 public Tfh clonotypes significantly expanded in patients were also unresponsive to HCoV (Figure 3B). Instead, homology searches for the epitopes of these clonotypes (Figure 3B) revealed that S₇₅₃₋₇₅₉, an epitope of clonotype 2, was contained in proteins of various organisms including symbiotic bacteria, such as

Bacteroidales bacteria and *Klebsiella pneumoniae* (Atarashi et al., 2017; Sefik et al., 2015) (Table S5 and Figure 5A). Indeed, these bacterial peptides, such as TonB-dependent receptor (TBDR)₁₆₉₋₁₈₃ and yersiniabactin ABC transporter ATP-binding/permease protein (YbtP)₂₇₀₋₂₈₄, respectively, could activate T cells reconstituted with clonotype 2 TCR (Figure 5D). Furthermore, *E. coli* transformed with the expression vectors for these proteins could activate public clonotypes in the presence of HLA-matched dendritic cells (Figure 5E), confirming that cross-reactive antigenic epitopes are processed and presented by APCs. These observations suggest that continuous priming via symbiotic bacteria might explain the high frequency of the public clonotypes worldwide (Figure 4E). Further cohort studies will clarify the etiological relationships between these bacteria and resistance against SARS-CoV-2 infection.

S₈₆₄₋₈₈₂ could activate multiple public Tfh clonotypes

Among 1735 TCRs derived from Tfh cluster (Table S3), we identified another clonotype from patient Ts-017 that recognizes the exact same epitope, S₈₆₄₋₈₈₂, restricted by DRB1*15:01, despite it had distinct V/J usages and CDR3 sequences/lengths from TCR-017/018 (Figure S5). This clonotype was also public as it was detected in cohorts (Table S3, Seurat barcode, DB2_CAAGAAAGTGCGGTAA). Thus, this epitope, S₈₆₄₋₈₈₂, could activate multiple public Tfh clonotypes even in the same patient.

To further examine the ability of this epitope to activate multiple Tfh clonotypes, PBMCs from additional convalescent/healthy donors possessing

DRB1*15:01/*15:02 were stimulated with S₈₆₄₋₈₈₂ *in vitro*. After epitope-induced expansion for 10 days, CD4⁺ T cells were subjected to single cell-TCR- and RNA-seq to determine the antigen-responsive clonotypes (Figure 6A and Table S6). As a result, we newly identified 104 public Tfh clonotypes that responded to S₈₆₄₋₈₈₂ after filtering by two criteria; 1) expanded during stimulation based on single cell sequencing (> 0.1% of total CD4⁺ T cells) and 2) detected in cohort databases (Table S6, column K). Indeed, bulk TCR sequencing demonstrated that both α and β chains of all 104 clonotypes were expanded after S₈₆₄₋₈₈₂ stimulation (Table S6, column I). For most of these clonotypes, Tfh characteristics were observed even after *in vitro* stimulation (Table S6, column J). These results indicate that S₈₆₄₋₈₈₂ epitope is capable of expanding multiple public Tfh clones in DRB1*15:01/02 donors.

To examine the presence of these S₈₆₄₋₈₈₂-reactive clones in the periphery of donors before *in vitro* expansion, we assembled all expanded TCR $\alpha\beta$ pairs with bulk TCR sequences of unstimulated peripheral T cells (Figure 6A, day 0). Approximately 20% of S₈₆₄₋₈₈₂-reactive TCR α and β chains detected at day 10 (Table S6) were present in the ‘day 0’ bulk TCR sequences from all recovered patients possessing anti-S protein antibodies (Table S1), whereas none of such S₈₆₄₋₈₈₂-reactive TCRs were detectable in healthy control before stimulation (Figure 6B). This data provides an implication that some S₈₆₄₋₈₈₂-reactive Tfh clones have been expanded in recovered patients during SARS-CoV-2 infection and maintained as a memory pool after infection.

To directly detect S₈₆₄₋₈₈₂-reacting T cells in PBMCs, we generated DRA-DRB1*15:01 tetramer loaded with this epitope, which was confirmed to bind to

recombinant soluble TCR-017 $\alpha\beta$ *in vitro* (Figure S6A and S6B). In healthy donors possessing DRB1*15:01, tetramer-positive CD4⁺ T cells (approximately 0.005% of T cells) expressed higher CXCR5 than tetramer-negative CD4⁺ cells (Figure 6C). After *in vitro* activation by S₈₆₄₋₈₈₂ peptide, the percentages of tetramer-positive CD4⁺ T cells were markedly increased and CXCR5⁺ population was further accumulated (Figure 6C), suggesting that S₈₆₄₋₈₈₂ peptide is capable of generating activated CD4⁺ T cells with Tfh profiles.

Discussion

Recently, public TCR for SARS-CoV-2 has been extensively studied (Shomuradova et al., 2020). In this study, we report the first identification and structural determination of SARS-CoV-2-specific public TCR of CD4⁺ T cells. Furthermore, we identified multiple public Tfh clonotypes that were significantly expanded in COVID-19-recovered patients by preferentially recognizing conserved S epitopes.

These observations suggest that the individuals possessing such public Tfh clonotypes could efficiently respond to S protein and produce antibodies against the cognate antigen, which may contribute to the prevention of disease aggravation and subsequent recovery (Table S1)(Emerson et al., 2017; Nolan et al., 2020). As these public clonotypes were widely detected in multiple ethnicities, the precise monitoring of the frequencies of public Tfh clonotypes using tetramer or clonotypic mAb could provide a novel option for prognosis prediction. Furthermore, such common S epitopes are candidates for next-generation vaccine that promote Tfh responses against SARS-CoV-2 and prevent aggravation.

The current study still has several limitations. 1) We analyzed only PBMCs. Although circulating Tfh cells in the periphery can reflect Tfh cells in germinal center (Crotty, 2019; Heit et al., 2017; Hill et al., 2019; Locci et al., 2013; Schmitt et al., 2014), the analysis of tissue resident Tfh cells of biopsy samples is ideal to fully characterize Tfh clonotype. 2) Although the present study suggests that these Tfh clonotypes contribute to amelioration of COVID-19 in patients recovered from mild symptoms, additional studies on patients with severe symptoms are helpful for further aspects. 3) The restricting HLA of the public clonotypes expanded in the patients were DRA-DRB1*15:01 (8.8%), DRA-DRB1*15:02 (2.5%), DRA-DRB1*15:03 (2.0%), DRA-DRB1*15:04 (0.02%), DRA-DRB4*01:03 (33.9% in local)(González-Quezada et al., 2019), DPA1*01:03 (60.7%)-DPB1*04:02 (16.5%), and DPA1*02:02 (24.7%)-DPB1*05:01 (14.8%), which cover a large part of, but not all, population worldwide (Gonzalez-Galarza et al., 2020).

Notwithstanding, the current study proposes that combination of single cell paired TCR sequencing and public database is a potent tool to identify public clonotypes which actually responded to SARS-CoV-2 in patients. The rapid reconstitution and determination platform enables us to identify epitopes and responsible HLA simultaneously. As the identified T cell epitopes are conserved among emerging SARS-CoV-2 mutants (Elbe and Buckland-Merrett, 2017; Singer et al., 2020) (Figure 3C), these peptides could become candidates for a robust vaccine against mutants, such as B.1.1.7 (UK), B.1.351 (South Africa), P.1 (Brazil) and presumably other future variants (Faria et al., 2021; Rambaut et al., 2020; Tegally et al., 2020).

Furthermore, the identification of cross-reacting symbiotic bacteria in this study would provide novel prognosis prediction and therapeutic options. To extend a peptide pool for the vaccine development, identification of additional public clonotypes and their epitopes is necessary and now under investigation.

Acknowledgements

We thank T. Ito, K. Toyonaga, Y. Adachi, S. Moriyama, Y. Harima and D. Azizah for experimental support; C. Schutt, W. Ise, J. B. Wing, D. Okuzaki, D. Standley, S. Teraguchi, H. Nakagami, S. Futami and T. Kobayashi for discussion. This research was supported by AMED (19fk0108161h0001 (HA), JP20fk0108265, JP20nf0101623, JP20gm0910010, JP20ak0101070 and JP20fk0108075 (SY)) and JSPS KAKENHI JP18H05279 (HA).

Author Contributions

S.Y. conceptualized research; X.L., Y.H., S.I., E.I., M.N. and Y.O. did investigation; R.S., T.I., T.O., T.M., M.S., T.S., S.M., T.K., E.N., T.S., T.K., H.A. and Y.T. provided resources; D.M., N.S., S.N. M.N. and K.Y. did data curation; S.Y. supervised the research; X.L., Y.H. and S.Y. wrote the manuscript.

Declaration of Interests

Authors do not have any conflict of interest.

Figure Legends

Figure 1. Shared TCR expressed by SARS-CoV-2-reactive Tfh cells from convalescent patients.

(A) UMAP projection of T cells in single-cell analysis of PBMCs identified a cluster representing Tfh cells (cluster #8). Each dot corresponds to a single cell and was colored according to the cluster.

(B) The expressions of canonical Tfh cell markers, CD200, PDCD1, ICOS, CXCL13, CD40LG and CXCR5 were shown as heatmaps in the UMAP plot.

(C) Usages of variable (V) and Joint (J) gene segments and CDR3 sequences of α and β chains of TCR-017 and TCR-018. For TCR-018, three different barcoded clones having the same nucleotide sequences detected in Ts-018 were shown.

(D) Alignment of CDR3 α and CDR3 β nucleotide and amino acid sequences of TCR-017 and TCR-018. Nucleotide sequences in white boxes indicate N-nucleotides.

(E) Expression of paired TCRs on reporter cells indicated by surface CD3 expression. Closed, anti-mouse CD3 mAb; open, isotype control antibody.

See also Figure S1 and Table S1-S3.

Figure 2. TCR-017 and -018 recognize the same core epitope, S₈₇₀₋₈₇₈, presented on DRB1*15:01.

(A) Cells expressing TCR-017 or TCR-018 were stimulated with 10 μ g/ml recombinant S protein, 1 μ g/ml S protein peptide pool (S pep pool) or 1 μ g/ml M and N protein peptide pool (MN pep pool) in the presence of autologous APCs and analyzed for GFP

and CD69 expressions.

(B-D) Cells expressing TCR-017 or TCR-018 were stimulated with 0.3 µg/ml S peptide pool #1, #2 or Selected S pool (S₃₀₄₋₃₃₈, 421-475, 492-519, 683-707, 741-770, 785-802, 885-1273) (B), 1 µg/ml S peptide pool #2 in the presence of HEK293T cells expressing indicated HLAs (C), or 1 µg/ml single peptides S₈₂₈₋₈₄₆ or S₈₆₄₋₈₈₂ (D).

(E) Sensitivity of TCR-017 and TCR-018 to the epitope. TCR transfectants were stimulated with indicated amount of S₈₆₄₋₈₈₂ and APCs from Ts-018. Activities were shown as percentages to the maximum responses induced by plate-coated anti-CD3 mAb. Data are shown in mean ± SD of triplicates.

(F) Core epitope recognized by TCR-017 and TCR-018 (shown in red). Cells were stimulated with 1 µg/ml of serial overlapped 15-mer peptides covering S₈₆₁₋₈₈₇ region. Data are shown in mean ± SD of triplicates.

(G and H) Occurrence of TCRα-017/018 (CDR3: CVVNRGSSYKLIF) and TCRβ-017/018 (CDR3: CASS[Q or P]TYEQYF) in a pre-pandemic healthy cohort from Japan (n = 27) (G), and in cohorts of healthy donors (n = 786) and convalescent COVID-19 patients (n = 1,413) (H). Numbers of individuals possessing the TCRs among cohorts were indicated.

All data are representatives of at least two independent experiments.

See also Figure S2 and Table S3.

Figure 3. Determination of epitopes of public Tfh clonotypes expanded in COVID-19 patients within conserved S regions with low mutation rates.

(A) Characterization of public clonotypes. Clonotypes detected in multiple donors from our sample pool are shown as # (column A). Clonotype occurrences were calculated based on a cohort of pre-pandemic healthy donors ($n = 27$) and public database including healthy donors ($n = 786$) and convalescent COVID-19 patients ($n = 1,413$) datasets (column K and I). For clonotypes that contain different TCR α or TCR β sequences, occurrences are calculated using pooled sequences (column K and I). Clonotypes were listed following the order of significance of TCR β expansion. *, Clonotypes expanded significantly in recovered patients ($p < 0.05$) (column J).

(B) TCR pair of each clonotype listed in (A) was reconstituted in reporter cells and tested as described in Figure S3A in the presence of shared HLA-expressing HEK293T cells or autologous APCs. Expressions of paired TCRs on reporter cells indicated by surface CD3 expression were shown in histograms. Closed, TCR-pairs-reconstituted reporter cells; open, parental cells. GFP reporter activities are shown as heatmaps. Restricting HLAs were determined as described in Figure S3B-F, Table S2 and Methods. Clonotypes 1 and 6 did not respond to S peptide pools. TCR $\alpha\beta$ of clonotype 8 was not expressed on the cell surface.

(C) Locations of epitopes described in (B) are highlighted in red in 3D structure of SARS-CoV-2 S protein (PDB code: 6XR8). Frequencies of mutation (mutation rate) defined relative to Wuhan-Hu-1 were calculated based on the data from CoV-GLUE that enabled by data from GISAID.

See also Figure S3 and Table S3.

Figure 4. Structural basis of peptide recognition by public clonotype 5 (017/018).

(A) TCR β -017, -018 or mutants at position 94 were co-expressed with TCR α -017/018 and human CD4 in reporter cells. Cells were kept unstimulated or stimulated with 1 μ g/ml of peptide S₈₆₄₋₈₈₂ or anti-CD3 antibody in the presence of APCs from Ts-018. Percentages of GFP⁺ cells in the CD3⁺ populations are shown. Data are shown in mean $\pm$ SD of duplicates.

(B and C) Crystal structure of TCR $\alpha\beta$ heterodimer of clonotype 5 (TCR-017). Whole view (B) and close-up view of CDR3 domains (C) are shown. Q94, C90, A91, S92 and S93 within CDR3 β are indicated.

(D) Frequencies of functional antigen-reactive TCR β chains harboring acceptable 17 substitutions at position 94 except for D94 of TCR β -017/018 determined in (A) in the healthy donors and convalescent COVID-19 patients. Horizontal line, median; bottom/top, first/third quartile; whiskers, 1.5 times the interquartile range.

(E) Occurrences of extended antigen-reactive TCR β chains harboring acceptable 17 substitutions at position 94 (shown as X) except for D94 of TCR β -017/018 determined in (A) in the individuals in a pre-pandemic Japanese healthy cohort, a healthy cohort and a convalescent COVID-19 cohort described in Figure 2G and 2H. Numbers of individuals possessing this TCR β among cohorts are indicated.

See also Table S7.

Figure 5. Cross-reactivity of SARS-CoV-2-specific public Tfh clonotypes.

(A) Alignment of cross-reacting epitopes. Similar amino acids to SARS-CoV-2 epitopes

found in HCoV and microbes are shown in black. Details of the microbes and full list of homologous search are shown in Table S4 and S5.

(B) Cells expressing clonotype 5 (TCR-017 or TCR-018) were stimulated with 1 µg/ml of peptides derived from MATE family efflux transporter of *Selenomonas noxia*.

(C) Cells expressing clonotype 5 (TCR-017 or TCR-018) were stimulated with 1 µg/ml of S₈₆₇₋₈₈₁ peptides containing indicated mutations in the presence of APCs from Ts-018.

Data are shown in mean ± SD of triplicates.

(D) Cells expressing TCR from clonotype 2 were stimulated with 1 µg/ml of peptides derived from *Klebsiella pneumoniae* yersiniabactin ABC transporter ATP-binding/permease protein (YbtP) or *Bacteroidales* bacterium TonB-dependent receptor (TBDR). Accession numbers of TBDRs from different *Bacteroidales* bacteria are shown.

(E) Cells expressing TCRs from clonotype 5 (TCR-017) or clonotype 2 were stimulated with heat-killed *E. coli* transformed with the expression vectors for indicated microbial proteins in the presence of monocyte-derived dendritic cells expressing DRB1*15:01.

See also Figure S4 and Table S4 and S5.

Figure 6. Characterization of S₈₆₄₋₈₈₂-reactive T cells.

(A) Workflow of the analysis for the S₈₆₄₋₈₈₂-induced T cell expansion. PBMCs from convalescent COVID-19 patients (red) / healthy (grey) donors possessing DRB1*15:01 or *15:02 were analyzed by bulk TCR-seq, or stimulated with 1 µg/ml of S₈₆₄₋₈₈₂ peptide for 10 days before CD4⁺ T cells were sorted and analyzed by single cell TCR-

and RNA-seq as well as bulk TCR-seq.

(B) Numbers of TCR α and β of expanded clonotypes (proportion > 0.1% in single cell TCR-seq) detected in bulk TCR-seq described in Table S6 are shown.

(C) PBMCs from healthy donors possessing DRB1*15:01 were stained with S₈₆₇₋₈₈₁-loaded DRA-DRB1*15:01 tetramer along with anti-CD4 and anti-CXCR5 antibodies before or after S₈₆₄₋₈₈₂ stimulation followed by flowcytometric analysis. Percentages of CXCR5⁺ cells within tetramer-positive (Tet⁺) and -negative (Tet⁻) CD4⁺ cells are shown.

All data are representative of at least two independent experiments.

See also Figure S5 and S6 and Table S6.

Methods

Human subjects

The Institutional Review Boards of Osaka University (approval number 898-4) and National Institute of Infectious Diseases (approval number 1137) approved blood draw protocols for convalescent COVID-19 and healthy individuals. The Institutional Review Board of nonprofit organization MINS (approval number 190210) approved the analyses of blood samples from healthy individuals in KOTAI Biotechnologies. The research was performed in accordance with all relevant guidelines and regulations. Written informed consent was obtained from all participants or designated healthcare surrogates if participants were unable to provide informed consent. Study enrollment criteria included subjects over 20 years old, regardless of disease severity and genders (Table S1). Prior to enrollment in this study, all COVID-19 donors were confirmed to be positive for SARS-CoV-2 by PCR using nasopharyngeal swab specimens. Blood from convalescent COVID-19 donors and healthy donors was obtained at Tokyo Shinagawa Hospital and Osaka University. Samples were de-identified prior to analyses. For PBMC preparation, whole blood was collected in heparin-coated tube and centrifuged to separate the cellular fraction and plasma, followed by density-gradient sedimentation. For neutralizing antibody assay, plasma samples from patients were heat-inactivated for 30 min at 56°C. Plasma diluted at 1:5 followed by two-fold serial dilutions was incubated with equal volume of solution containing 100 TCID₅₀ of SARS-CoV-2 for 1 hour at 37°C and added to VeroE6/TMPRSS2 cells. After incubation for 5 days, the highest plasma dilution that protected 100% of cells from

cytopathic effect was taken as the neutralization titer. Disease severity was defined as mild, moderate I, moderate II or severe in accordance with the Japanese COVID-19 clinical practice guideline ver. 2.2 (<https://www.mhlw.go.jp/content/000650160.pdf>).

SARS-CoV-2

For the preparation of inactivated SARS-CoV-2, SARS-CoV-2 (KNG19-020) was kindly supplied by Dr. Tomohiko Takasaki (Kanagawa Prefectural Institute of Public Health). The virus was propagated in VeroE6/TMPRSS2 cells (JCRB1819) and purified by sucrose gradient centrifugation (Dent and Neuman, 2015). Concentrated virus was then inactivated by ultraviolet light (0.6 J/cm^2).

Antigens

Inactivated virus was used at 2.9×10^6 viral particles/ml (approximately $1 \mu\text{g S protein/ml}$) for stimulation. For the preparation of recombinant SARS-CoV-2 S protein, the ectodomain of S protein was cloned into a mammalian expression vector pCMV, with a foldon sequence followed by $9 \times \text{His-tag}$ and Strep-tag at the C-terminus. Polybasic cleavage site (RRAR) of S protein was replaced by a single alanine, and K986P and V987P mutations were introduced to stabilize the conformation as previously described (Amanat et al., 2020). After transient transfection into Expi293F cells (Gibco), secreted protein was purified from the culture supernatant 4 days post transfection using TALON metal affinity resin (Clontech) and Amicon Ultra 10K filter device (Millipore). For the expression of S proteins of SARS-CoV-2 and HCoV-OC43,

pME18S expression vectors for each protein were transfected into HEK293T cells. For the expression of bacterial proteins in *E. coli*, cDNA for MATE family efflux transporter, TonB-dependent receptor and TonB-dependent receptor plug domain-containing protein were cloned into pCold expression vector and transformed into *E. coli* BL21 competent cells Champion21 (SMBIO Technology). The protein expressions were induced by the addition of 0.5 or 1.0 mM isopropyl- β -D-thiogalactopyranoside (IPTG) at 37°C. Bacterial pellets were washed and resuspended with PBS and incubated at 95°C for 5 min for inactivation. The expressions of the proteins were confirmed by SDS-PAGE and CBB staining. PepMix SARS-CoV-2 (Spike Glycoprotein)(contains pool #1 and #2) was purchased from JPT Peptide Technologies. PepTivator SARS-CoV-2 for S, M and N protein were purchased from Miltenyi Biotec. Domains of SARS-CoV-2 S protein showed in Figure S2A was previously described (Wrapp et al., 2020). All of other peptides were from GenScript. 3D structure of SARS-CoV-2 S protein (PDB code: 6XR8) was depicted with program PyMOL ver. 2.0 (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.).

In vitro stimulation of PBMCs

Cryopreserved PBMCs were thawed and washed with RPMI 1640 medium (Sigma) supplemented with 5% human AB serum (GeminiBio), Penicillin (Sigma), streptomycin (MP Biomedicals) and 2-mercaptoethanol (Nacalai Tesque). 5×10^5 PBMCs were stimulated in the same medium with inactivated SARS-CoV-2, 1 or 10 μ g/ml of

recombinant S protein, 1 $\mu\text{g/ml}$ of S peptide pool or 1 $\mu\text{g/ml}$ of M + N (MN) peptide pool for 20 hours, followed by staining with anti-human CD3, CD69, CD137, CD154 and TotalSeq-C Hashtags antibodies. CD69⁺, CD137⁺ or CD154⁺ cells within CD3⁺-gated population were sorted by cell sorter SH800S (SONY) and used for single cell TCR and RNA sequencing analyses. For epitope-specific clonal expansion, $1-5 \times 10^5$ PBMCs were stimulated with 1 $\mu\text{g/ml}$ of S₈₆₄₋₈₈₂ for 10 days and recombinant human IL-2 (1 ng/ml, Peprotech) was added at day 4 and day 7. CD4⁺ T cells were sorted and analyzed by single cell TCR and RNA sequencing.

Single cell-based transcriptome and TCR repertoire analysis

Single cell capturing and library preparation were performed using following reagents; Chromium Next GEM Single Cell 5' Library & Gel Bead Kit v1.1, 16rxns, PN-1000165; Chromium Next GEM Chip G Single Cell Kit, 48rxns, PN-1000120; Chromium Single Cell V(D)J Enrichment Kit, Human T Cell, 96 rxns, PN-1000005; Single Index Kit T Set A, 96 rxns, PN-1000213; Chromium Single Cell 5' Feature Barcode Library Kit, 16 rxns, PN-1000080; Single Index Kit N Set A, 96 rxns, PN-1000212. Single cell suspension containing approximately 2×10^4 cells were loaded into Chromium microfluidic chips to generate single-cell gel bead-in-emulsion using Chromium controller (10x Genomics) according to the manufacturer's instructions. RNA from the barcoded cells for each sample was subsequently reverse-transcribed inside gel bead-in-emulsion using Veriti Thermal Cycler (Thermo Fisher Scientific), and all subsequent steps to generate single-cell libraries were performed according to

the manufacturer's protocol, with 14 cycles used for cDNA amplification. Then ~50 ng of cDNA was used for gene expression library amplification by 14 cycles in parallel with cDNA enrichment and library construction for TCR libraries. Fragment size of the libraries was confirmed with Agilent 2100 Bioanalyzer (Agilent). Libraries were sequenced on Illumina NovaSeq 6000 as paired-end mode (read1: 28bp; read2: 91bp). The raw reads were processed by Cell Ranger 3.1.0 (10x Genomics). Gene expression-based clustering was performed using the Seurat R package (v3.1)(Hafemeister and Satija, 2019). Briefly, cells with a mitochondrial content above 10%, cells with less than 200 or more than 4,000 genes detected were considered outliers (dying cells, empty droplets and doublets, respectively) and filtered out. The Seurat SCTransform function was used for normalization, and data was integrated without performing batch-effect correction as all samples were processed simultaneously. Hashtag oligo demultiplexing was performed on CLR-normalized hashtag UMI counts, and clonotypes were matched to the gene expression data through their droplet barcodes, using Python scripts. Only cells assigned a single hashtag and a beta-chain clonotype were retained for downstream analyses.

Bulk TCR sequencing and analysis

$1-3 \times 10^5$ PBMCs were lysed in QIAzol (Qiagen). Full-length cDNA was then synthesized using the SMARTer technology (Takara Bio), and the variable region of TCR α and β genes were amplified using TRAC/TRBC-specific primers. After sequencing of the variable region amplicons, each pair of reads was assigned a

clonotype (defined as TR(A/B)V and TR(A/B)J gene, and the complementarity-determining region 3) using the MiXCR software (Bolotin et al., 2015). For each α/β clonotype, expansion was defined as the fraction of reads for that clone divided by the total number of reads for the α/β chain, respectively, and fractions were converted to log10 for plotting and statistical analyses.

HLA class II typing

For HLA class II typing, genomic DNA samples were isolated from PBMCs using QIA DNA Mini Kit (Qiagen). AllType FASTplex NGS 11 Loci Flex Kit (ONE LAMBDA) was used to prepare DNA sequence libraries (DPA1, DPB1, DQA1, DQB1, DRB1 and DRB3/4/5) according to the manufacturer's protocol. Sequencing was performed on Illumina MiSeq (Illumina). TypeStream Visual ver. 2.0 (ONE LAMBDA) was used to analyze the DNA sequences. For DRB1*15:01/02 typing, genomic DNA was amplified with a DRB1*15/16-specific forward primer (5'-CGT TTC CTG TGG CAG CCT AAG AGG-3') and a DRB1*15-specific reverse primer (5'-CCG CGC CTG CTC CAG GAT-3') followed by DNA sequencing.

Antigen-presenting cells

Recombinant Epstein-Barr virus (EBV) was produced as previously reported (Kanda et al., 2015). Viral stocks were obtained by concentrating virus-containing culture supernatants by ultracentrifugation at 32,000 rpm for 1 hour. 3×10^5 PBMCs were incubated with an aliquot of the viral stock for 1 hour at 37°C. The infected cells were

cultured with RPMI 1640 medium supplemented with 20% fetal bovine serum (FBS, CAPRICORN SCIENTIFIC GmbH) containing 0.1 µg/ml cyclosporine A (CsA, Cayman Chemical). EBV-immortalized B lymphoblastoid cell lines were obtained after 3 weeks culture and used as APCs. To generate APCs expressing specific HLA, HEK293T cells were transfected with plasmids encoding HLA class II alleles as previously described (Jiang et al., 2013). EBV experiments were approved by the Minister of Education, Culture, Sports, Science and Technology (approval number 539) and the Institutional Review Boards of Osaka University (approval number 04658). To generate monocyte-derived dendritic cells, CD14⁺ cells were isolated from PBMCs using CD14 MicroBeads, human (Miltenyi Biotec), and cultured in RPMI 1640 medium supplemented with 10% FBS, 0.1 mM Non-Essential Amino Acid Solution (Gibco), 1mM Sodium Pyruvate (Gibco), 10 ng/ml of human GM-CSF (Peprotech) and IL-4 (Peprotech) for 6 days.

TCR reconstitution and stimulation

TCRα and β chain cDNA sequences were synthesized and cloned into retroviral vectors pMX-IRES-rat CD2 and pMX-IRES-human CD8, respectively. Two vectors containing paired TCRα and β chains were transfected together into Phoenix packaging cells using PEI MAX (Polysciences). Supernatant containing retroviruses was used for infection into mouse T cell hybridoma with an NFAT-GFP reporter gene (Matsumoto et al., 2021) to reconstitute TCRαβ pairs. TCRβ mutants were constructed by site-directed mutagenesis. For antigen stimulation, TCR-reconstituted cells were co-cultured with

stimulants in the presence of immortalized autologous B cells unless indicated otherwise. After 20 hours, T cell activation was assessed by GFP or CD69 expression.

Rapid epitope-determination platform

For the preparation of pooled peptide matrix, 15-mer peptides with 11 amino acids overlap that cover the full length of S protein of SARS-CoV-2 were individually synthesized (GenScript). Peptides were dissolved in DMSO at 12 mg each peptide/ml and 8-12 peptides were mixed to create 75 different semi-pools so that the responsible epitopes can be determined from the reactivities of horizontal and vertical pools. Pooled peptides were diluted 10 times with water, followed by adding 1 μ l of the solution into each corresponding well. S peptide pools and S₈₆₄₋₈₈₂ peptide were also loaded as controls. S peptide pools of common cold HCoV-229E were loaded to assess cross-reactivities. For reporter cell assay, 100 μ l of media containing T cells and APCs were added to the wells so that T cells are stimulated with each peptide at 1 μ g/ml. After 20 hours, GFP reporter induction was assessed in T cell-gated population using AttuneNxT flow cytometer (Thermo Fisher Scientific).

Restricting HLA determination

For the clonotypes detected in more than one donor from our sample pool, reporter cells were stimulated in the presence of HEK293T cells expressing individual pair of alleles that shared by the donors. Otherwise, reporter cells were stimulated in the presence of transformed B cells from various donors that have partial overlapped HLAs with the

original donor of the clonotype.

Tetramer staining

Freshly-isolated or *in vitro*-expanded PBMCs from DRB1*15:01 donors were stained with S₈₆₇₋₈₈₁ peptide-loaded DRA-DRB1*15:01 tetramer. For expansion, PBMCs from DRB1*15:01-possessing donors were stimulated with 1 µg/ml of peptide S₈₆₄₋₈₈₂ for 9 days. Cells were treated with 100 nM protease kinase inhibitor (dasatinib, Axon Medchem) at 37°C for 30 min and stained with tetramer conjugated with PE (Medical & Biological Laboratories) at 4°C for 1 hour. Cells were then stained with biotin-conjugated anti-PE antibody together with antibodies for surface markers for flow cytometric analysis using AttuneNxT (Thermo Fisher Scientific) and FlowJo 10.5.2 (BD Biosciences).

Database analyses

Public repertoire datasets of pre-pandemic healthy donors (n = 786, Adaptive ImmuneACCESS)(Emerson et al., 2017) or unhospitalized COVID-19 convalescent donors who had mild symptoms (n = 1,413, Adaptive ImmuneRACE)(Nolan et al., 2020) of various ethnicities from US, Italy, and Spain were downloaded from the Adaptive Biotechnologies website and V gene names were renamed to match the IMGT nomenclature. Finally, clonotypes and their expansion were defined in the same way as in-house bulk TCR sequencing. TCR occurrence in a dataset was defined as the fraction of patients whose repertoire contained the given TCR, regardless of its expansion.

Expression and purification of soluble TCR $\alpha\beta$ heterodimer

Expression constructs encoding the extracellular domains of TCR α -017 (Q1-D207 in mature protein) and TCR β -017 (G1-G241) subunits were incorporated into pCold vector including 6 $\times$ His-tag and a tobacco etch virus protease cleavage site. For crystallization, the point mutations (T159C in α and S166C/C184A in β) were introduced to form artificial disulfide bond as described previously (Boulter et al., 2003). The plasmids were transformed into *E. coli* BL21 competent cells Champion21 and Rosetta2 (DE3) (Novagen), respectively. The protein expression was induced by the addition of 0.5 mM IPTG at 18°C. Cells were suspended with 500 mM NaCl-containing Tris-HCl buffer (pH 8.0) and disrupted with sonication. The inclusion bodies including target proteins were collected by centrifugation. The inclusion body was then solubilized by 50 mM Tris-HCl buffer (pH 8.0) containing 6 M guanidine HCl, 10 mM EDTA and 2 mM DTT at room temperature. The equal amount of solubilized TCR α and TCR β (35 mg each) was mixed and rapidly diluted with 1 L of 100 mM Tris-HCl buffer (pH 8.0) containing 5 M urea, 0.4 M L-arginine, 5 mM reduced glutathione and 0.5 mM oxidized glutathione at 4°C. The diluted solution was further dialyzed against 10 mM Tris-HCl buffer (pH 8.0) at 4°C for two days. The dialyzed solution was applied onto 5 ml Ni-NTA agarose (Fuji-film Wako) and His-tagged TCR $\alpha\beta$ were eluted with elution buffer (50 mM Tris-HCl (pH 8.0), 300 mM NaCl and 250 mM imidazole). After removal of His-tag by TEV protease, the eluted protein was concentrated and further applied to Superdex75 (GE Healthcare) equilibrated with 20 mM Tris-HCl (pH 8.0)

buffer containing 100 mM NaCl. The TCR $\alpha\beta$ heterodimer fractions were concentrated up to 1.2 mg/ml by Amicon Ultra (molecular weight cut off: 10 kDa). The purity of the proteins was assessed by SDS-PAGE and CBB staining. The numbering of amino acids in CDR3 sequences of TCRs was based on mature protein.

Biolayer interferometry analysis

Octet RED96 System (fortéBio, Pall Life Science) was used to detect the binding kinetics of TCR $\alpha\beta$ heterodimer to peptide-loaded HLA molecule (Medical & Biological Laboratories). All assays were carried out at 25°C on a black 96-well plate with agitation set at 1,000 rpm and 200 μ l/well for all solutions. Super Streptavidin biosensors (fortéBio, Pall Life Science) were dipped in PBS prior to use and reacted with 5.8 μ g/ml biotinylated peptide-loaded HLA molecule. The association of TCR $\alpha\beta$ heterodimer at concentrations of 0.1, 0.2 and 0.4 μ M was measured for 900 sec, and the dissociation in PBS was measured for 900 sec. The obtained data was analyzed and processed using Octet data analysis 11.1 software.

Crystallization, data collection, and structure determination of TCR-017 ectodomain

All crystallization trials were performed by sitting drop vapor diffusion method. Initial crystallization conditions were screened using Index (Hampton Research), SG1 Screen, and SG2 Screen (Molecular Dimensions). The best diffracted crystal was obtained under the condition of 0.1 M HEPES-Na (pH 7.0) and 1.1 M sodium malonate at 20°C. Prior to X-ray diffraction experiments, crystals were soaked in the reservoir containing

20% ethylene glycol and flash-cooled in liquid nitrogen. X-ray diffraction data sets were collected at the beamline BL-1A in the Photon Factory (Tsukuba, Japan). Diffraction data were integrated with program XDS (Kabsch, 2010) and scaled with program SCALA (Evans, 2006). The phases of datasets were determined by molecular replacement method using program MOLREP (Vagin and Teplyakov, 2010) with the coordinate of 1E6 TCR (PDB code: 5C0B)(Cole et al., 2016). After initial phase determination, the model buildings were manually performed using program COOT (Emsley et al., 2010). Refinement was performed using program REFMAC5 (Vagin et al., 2004) at the initial step and Phenix.refine for final model (Afonine et al., 2012). The stereochemical quality of the final model was assessed by program MolProbity (Williams et al., 2018). Data collection and refinement statistics were summarized in Table S7. Structural factors and the atomic coordinates of TCR $\alpha\beta$ -017 ectodomain have been deposited in Protein Data Bank under the accession code of 7EA6. All figures of 3D structure were depicted with program PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.).

Antibodies

Anti-human CD3 (HIT3a), anti-human CD8 (RPA-T8), anti-human CD69 (FN50), anti-human CD137 (4B4-1), anti-human CD154 (24-31), anti-human CD4 (OKT4), anti-human CXCR5 (J252D4), TotalSeq-C Hashtags (LNH-94; 2M2), anti-mouse CD3 (17A2), anti-mouse CD69 (H1.2F3), anti-rat CD2 (OX-34) and biotin anti-phycoerythrin (PE)(PE001) antibodies were purchased from BioLegend. Rat IgG2b

κ Isotype Control (eB149/10H5) was purchased from eBioscience.

Statistical analysis

Statistical analysis was performed with *t*-test using SciPy and p-values were indicated in Figure 3A and 4D.

Data availability

Recombinant SARS-CoV-2 S protein, TCR-reconstituted cells, soluble TCRαβ heterodimers under a standard material transfer agreement, and single cell-based transcriptome data are available from the corresponding author on request as supplies permit. The datasets supporting the current study have not been deposited in a public repository because permission was not acquired from the blood donors involved in this study but are available from the corresponding author on request. Other data needed to support the conclusion of this manuscript are included in the main text and Supplemental information.

Supplemental Figure Legends

Figure S1. Workflow for single cell-based analyses of SARS-CoV-2-responsive T cells, related to Figure 1.

PBMCs from healthy donors and convalescent COVID-19 patients were stimulated with SARS-CoV-2-derived antigens as indicated for 20 hours followed by staining with eight different hashtag antibodies (#1-8), anti-CD3 and indicated activation marker antibodies. Hashtagged cells from two patients were pooled and analyzed with cell sorter. Antigen-responded ($CD69^+$ or $CD137^+$ or $CD154^+$) $CD3^+$ T cells were sorted and immediately subjected to single-cell TCR and RNA sequencing. Each collection day after diagnosis is shown after the sample ID.

Figure S2. Epitope peptide and restricting HLAs of clonotype-017/018, related to Figure 2.

(A) Relative positions that are covered by each S peptide pool are shown. NTD, N-terminal domain; RBD, receptor-binding domain; FP, fusion peptide; TM, transmembrane domain. Strong binding peptides ($\%RANK \leq 2$) presented by DRA-DRB1*15:01 in region 633-682, 708-740, 771-784 and 803-884 (pink regions) were predicted by NetMHC 4.0 Server. 19-mer peptides (red bars) that cover the strong binding peptides were synthesized and used for further epitope identification.

(B) Cells expressing TCR-017 or -018 were stimulated with 1 μ g/ml of S peptide pool #2 in the presence of heterologous APCs from exchanged donors (Ts-018 and Ts-017, respectively).

(C and D) Cells expressing TCR-017 or -018 were stimulated with 1 μ g/ml of peptide

S₈₆₄₋₈₈₂ in the presence of HEK293T cells expressing indicated HLAs (C) and immortalized B cells possessing indicated HLAs (D).

Figure S3. Identification of the epitopes of public clonotypes, related to Figure 3.

(A) A scheme of rapid epitope-determination platform for identification of the epitope of T cell clonotypes. 15-mer peptides with 11 amino acids overlap that cover the full length of S protein of SARS-CoV-2 were synthesized and pooled as described in Materials and Methods. Each peptide is present in a unique combination of two different pools, thus the epitope can be revealed by the common peptide(s) in the pools that activated the clonotype. Peptide pools of S, M and N proteins of SARS-CoV-2 or HCoV S proteins were also added in the matrix as controls. S864 indicates S₈₆₄₋₈₈₂ peptide.

(B) Identification of DRB1*15:01 as a restricting HLA of S₇₅₃₋₇₅₉ for clonotype 2. Reporter cells expressing clonotype 2 were activated by S₇₄₉₋₇₆₃ in the presence of HEK293T cells transfected with DPA1*02:01-DRB1*15:01. Data are shown in mean ± SD of triplicates.

(C) Exclusion of DPA1*02:02-DPB1*05:01 as a restricting HLA of S₃₅₃₋₃₆₇ for clonotype 3. Reporter cells expressing clonotype 3 were not activated by S₃₅₃-including peptide pools, such as A5, E8 and E9 (S pool #1) in Figure S3A, in the presence of HEK293T cells transfected with DPA1*02:02-DPB1*05:01.

(D) Identification of DPA1*02:02-DPB1*05:01 as a restricting HLA of S₅₅₇₋₅₆₇ for clonotype 7. Reporter cells expressing clonotype 7 were activated by S₅₅₃- or S₅₅₇-including peptide pools, such as B7, B8, F4 and E9 (S pool #1) in Figure S3A, in

the presence of HEK293T cells transfected with DPA1*02:02-DPB1*05:01.

(E) Identification of DQA1*01:01-DQB1*05:01 as a restricting HLA of S₈₃₃₋₈₄₃ for clonotype 4 and 9. Reporter cells expressing clonotype 4 and 9 were activated by S₈₂₈₋₈₄₆ in the presence of HEK293T cells transfected with DQA1*01:01-DQB1*05:01. Data are shown in mean $\pm$ SD of triplicates.

(F) Identification of DPA1*02:02-DPB1*05:01 as a restricting HLA of S₉₅₇₋₉₆₇ for clonotype 10. Reporter cells expressing clonotype 10 were activated by S953- or S957-including peptide pools, such as C11, C12, G4 and E10 (S pool #2) in Figure 3A, in the presence of HEK293T cells transfected with DPA1*02:02-DPB1*05:01.

Figure S4. No cross-reactivity of clonotype-017/018 to other human coronaviruses, related to Figure 5.

(A) Cells expressing TCR-017 or -018 were co-cultured with HEK293T cells expressing indicated HLAs and S protein (S) derived from indicated coronavirus.

(B) Cells expressing TCR-017 or -018 were stimulated with 1 μ g/ml of peptide pool derived from HCoV-OC43 S protein (HCoV-OC43 S pep pool).

(C) Sequences of peptides from human coronaviruses corresponding to S₈₆₇₋₈₈₁ of SARS-CoV-2. Amino acids that are same as those in SARS-CoV-2 in the core epitope are shown in black.

(D) Cells expressing TCR-017 or -018 were stimulated with 1 μ g/ml of peptides shown in (C).

Figure S5. Another public Tfh clonotype that recognizes the same epitope S₈₇₀₋₈₇₈ as clonotype-017/018, related to Figure 6.

(A) V/J usage and CDR3 sequences of the newly-identified clonotype that recognize S₈₇₀₋₈₇₈ peptide.

(B-D) Reporter cells expressing TCR $\alpha\beta$ in (A) was stimulated with 1 μ g/ml of S peptide pool #1, #2 or selected S pool (S₃₀₄₋₃₃₈, 421-475, 492-519, 683-707, 741-770, 785-802, 885-1273) (B), 1 μ g/ml of S peptide pool #2 in the presence of HEK293T cells expressing indicated HLAs (C), or 1 μ g/ml of peptides S₈₂₈₋₈₄₆ or S₈₆₄₋₈₈₂ (D).

(E) Cells were stimulated with pooled peptide matrix described in Figure S3A. Wells containing S865 or S869 (C1, C2 and G3), S pool #2 (E10) and S₈₆₄₋₈₈₂ (F11) were positive.

(F) 1 μ g/ml of serial overlapped 15-mer peptides covering S₈₆₁₋₈₈₇ region were tested for the reactivity to this clonotype to determine minimum epitope (shown in red). Data are shown in mean $\pm$ SD of triplicates.

Figure S6. S₈₆₇₋₈₈₁-loaded HLA-DR interacted with soluble TCR-017 heterodimer, related to Figure 6.

(A) The purity of soluble TCR-017 $\alpha\beta$ heterodimer (indicated by arrow head) was assessed by SDS-PAGE in reduced or non-reduced conditions.

(B) Interaction between S₈₆₇₋₈₈₁-loaded HLA DRA-DRB1*15:01 monomer and soluble TCR-017 by biolayer interferometry. Indicated concentrations of TCR $\alpha\beta$ heterodimer were applied to 5.8 μ g/ml immobilized recombinant HLA proteins.

Table S1. Donor characteristics, related to Figure 1.

Table S2. HLA class II types of involved blood donors, related to Figure 1.

Table S3. Tfh clonotypes identified in single cell analysis, related to Figure 1-3.

Table S4. Proteins possessing epitopes of TCR-017/018 (IAQY TSAAL), related to Figure 5.

Table S5. Proteins possessing epitopes of clonotype 2 (LLQYGSF), related to Figure 5.

Table S6. S₈₆₄₋₈₈₂-reacted public Tfh clonotypes, related to Figure 6.

Clonotypes excluding those with Treg characteristics were assembled with in-house bulk TCR-seq data as well as those in cohort databases. Clonotypes that satisfied following two criteria are shown: 1) expanded during stimulation based on single cell sequencing ($> 0.1\%$ of total CD4⁺ T cells) and 2) detected in cohort databases.

Table S7. Data collection and refinement statistics of the crystallographic analysis of TCRαβ-017, related to Figure 4.

References

- Afonine, P. V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 68, 352–367.
- Amanat, F., Strohmeier, S., Rathnasinghe, R., Schotsaert, M., Coughlan, L., García-Sastre, A., and Krammer, F. (2020). Introduction of two prolines and removal of the polybasic cleavage site leads to optimal efficacy of a recombinant spike based SARS-CoV-2 vaccine in the mouse model. *BioRxiv*, 10.1101/2020.09.16.300970.
- Atarashi, K., Suda, W., Luo, C., Kawaguchi, T., Motoo, I., Narushima, S., Kiguchi, Y., Yasuma, K., Watanabe, E., Tanoue, T., et al. (2017). Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation. *Science* 358, 359–365.
- Bacher, P., Rosati, E., Esser, D., Martini, G.R., Saggau, C., Schiminsky, E., Dargvainiene, J., Schröder, I., Wieters, I., Khodamoradi, Y., et al. (2020). Low-Avidity CD4⁺ T Cell Responses to SARS-CoV-2 in Unexposed Individuals and Humans with Severe COVID-19. *Immunity* 53, 1258-1271.e5.
- Begovich, A.B., McClure, G.R., Suraj, V.C., Helmuth, R.C., Fildes, N., Bugawan, T.L., Erlich, H.A., and Klitz, W. (1992). Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. *J. Immunol.* 148, 249–258.
- Bolotin, D.A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I.Z., Putintseva, E. V., and Chudakov, D.M. (2015). MiXCR: Software for comprehensive adaptive immunity profiling. *Nat. Methods* 12, 380–381.
- Boulter, J.M., Glick, M., Todorov, P.T., Baston, E., Sami, M., Rizkallah, P., and Jakobsen, B.K. (2003). Stable, soluble T-cell receptor molecules for crystallization and therapeutics. *Protein Eng.* 16, 707–711.
- Braun, J., Loyal, L., Frentsch, M., Wendisch, D., Georg, P., Kurth, F., Hippenstiel, S., Dingeldey, M., Kruse, B., Fauchere, F., et al. (2020). SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature* 587, 270–274.
- Cole, D.K., Bulek, A.M., Dolton, G., Schauenberg, A.J., Szomolay, B., Rittase, W.,

- Trimby, A., Jothikumar, P., Fuller, A., Skowera, A., et al. (2016). Hotspot autoimmune T cell receptor binding underlies pathogen and insulin peptide cross-reactivity. *J. Clin. Invest.* *126*, 2191–2204.
- Crotty, S. (2019). T Follicular Helper Cell Biology: A Decade of Discovery and Diseases. *Immunity* *50*, 1132–1148.
- Dent, S., and Neuman, B.W. (2015). Purification of coronavirus virions for Cryo-EM and proteomic analysis. *Methods Mol. Biol.* *1282*, 99–108.
- Elbe, S., and Buckland-Merrett, G. (2017). Data, disease and diplomacy: GISAID’s innovative contribution to global health. *Glob. Challenges* *1*, 33–46.
- Emerson, R.O., DeWitt, W.S., Vignali, M., Gravley, J., Hu, J.K., Osborne, E.J., Desmarais, C., Klinger, M., Carlson, C.S., Hansen, J.A., et al. (2017). Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat. Genet.* *49*, 659–665.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* *66*, 486–501.
- Evans, P. (2006). Scaling and assessment of data quality. In *Acta Crystallographica Section D: Biological Crystallography*, (Acta Crystallogr D Biol Crystallogr), pp. 72–82.
- Faria, N.R., Claro, I.M., Candido, D., Franco, L.A.M., Andrade, P.S., Thais, M., Silva, C.A.M., Sales, F.C., Erika, R., Aguiar, R.S., et al. (2021). Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. *Virological.Org*, 586.
- Gong, F., Dai, Y., Zheng, T., Cheng, L., Zhao, D., Wang, H., Liu, M., Pei, H., Jin, T., Yu, D., et al. (2020). Peripheral CD4⁺ T cell subsets and antibody response in COVID-19 convalescent individuals. *J. Clin. Invest.* *130*, 6588–6599.
- Gonzalez-Galarza, F.F., McCabe, A., Santos, E.J.M. Dos, Jones, J., Takeshita, L., Ortega-Rivera, N.D., Cid-Pavon, G.M.D., Ramsbottom, K., Ghattaoraya, G., Alfievic, A., et al. (2020). Allele frequency net database (AFND) 2020 update: Gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res.* *48*, D783–D788.

González-Quezada, B.A., Creary, L.E., Munguia-Saldaña, A.J., Flores-Aguilar, H., Fernández-Viña, M.A., and Gorodezky, C. (2019). Exploring the ancestry and admixture of Mexican Oaxaca Mestizos from Southeast Mexico using next-generation sequencing of 11 HLA loci. *Hum. Immunol.* *80*, 157–162.

Grifoni, A., Weiskopf, D., Ramirez, S.I., Mateus, J., Dan, J.M., Moderbacher, C.R., Rawlings, S.A., Sutherland, A., Premkumar, L., Jadi, R.S., et al. (2020). Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell* *181*, 1489-1501.e15.

Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* *20*, 1–15.

Heit, A., Schmitz, F., Gerdt, S., Flach, B., Moore, M.S., Perkins, J.A., Robins, H.S., Aderem, A., Spearman, P., Tomaras, G.D., et al. (2017). Vaccination establishes clonal relatives of germinal center T cells in the blood of humans. *J. Exp. Med.* *214*, 2139–2152.

Hill, D.L., Pierson, W., Bolland, D.J., Mkindi, C., Carr, E.J., Wang, J., Houard, S., Wingett, S.W., Audran, R., Wallin, E.F., et al. (2019). The adjuvant GLA-SE promotes human Tfh cell expansion and emergence of public TCRβ clonotypes. *J. Exp. Med.* *216*, 1857–1873.

Jiang, Y., Arase, N., Kohyama, M., Hirayasu, K., Suenaga, T., Jin, H., Matsumoto, M., Shida, K., Lanier, L.L., Saito, T., et al. (2013). Transport of misfolded endoplasmic reticulum proteins to the cell surface by MHC class II molecules. *Int. Immunol.* *25*, 235–246.

Kabsch, W. (2010). XDS. *Acta Crystallogr. D. Biol. Crystallogr.* *66*, 125–132.

Kanda, T., Miyata, M., Kano, M., Kondo, S., Yoshizaki, T., and Iizasa, H. (2015). Clustered MicroRNAs of the Epstein-Barr Virus Cooperatively Downregulate an Epithelial Cell-Specific Metastasis Suppressor. *J. Virol.* *89*, 2684–2697.

Kaneko, N., Kuo, H.H., Boucay, J., Farmer, J.R., Allard-Chamard, H., Mahajan, V.S., Piechocka-Trocha, A., Lefteri, K., Osborn, M., Bals, J., et al. (2020). Loss of Bcl-6-Expressing T Follicular Helper Cells and Germinal Centers in COVID-19. *Cell*

183, 143-157.e13.

Li, H., Ye, C., Ji, G., and Han, J. (2012). Determinants of public T cell responses. *Cell Res.* 22, 33–42.

Locci, M., Havenar-Daughton, C., Landais, E., Wu, J., Kroenke, M.A., Arlehamn, C.L., Su, L.F., Cubas, R., Davis, M.M., Sette, A., et al. (2013). Human circulating PD-1+CXCR3-CXCR5+ memory Tfh cells are highly functional and correlate with broadly neutralizing HIV antibody responses. *Immunity* 39, 758–769.

Mateus, J., Grifoni, A., Tarke, A., Sidney, J., Ramirez, S.I., Dan, J.M., Burger, Z.C., Rawlings, S.A., Smith, D.M., Phillips, E., et al. (2020). Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science* 370, 89–94.

Matsumoto, Y., Kishida, K., Matsumoto, M., Matsuoka, S., Kohyama, M., Suenaga, T., and Arase, H. (2021). A TCR-like antibody against a proinsulin-containing fusion peptide ameliorates type 1 diabetes in NOD mice. *Biochem. Biophys. Res. Commun.* 534, 680–686.

Nolan, S., Vignali, M., Klinger, M., Dines, J.N., Kaplan, I.M., Craft, T., Boland, K., Mazza, M., and Dobbs, K. (2020). A large-scale database of T-cell receptor beta (TCR β) sequences and binding associations from natural and synthetic exposure to SARS-CoV-2. *Res. Square*, 10.21203/RS.3.RS-51964/V1.

Rambaut, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., Connor, T., Peacock, T., Robertson, D.L., and Volz, E.M. (2020). Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *Virological.Org*, 563.

Reynisson, B., Barra, C., Kaabinejadian, S., Hildebrand, W.H., Peters, B., Peters, B., Nielsen, M., and Nielsen, M. (2020). Improved Prediction of MHC II Antigen Presentation through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data. *J. Proteome Res.* 19, 2304–2315.

Schmitt, N., Bentebibel, S.E., and Ueno, H. (2014). Phenotype and functions of memory Tfh cells in human blood. *Trends Immunol.* 35, 436–442.

Sefik, E., Geva-Zatorsky, N., Oh, S., Konnikova, L., Zemmour, D., McGuire, A.M., Burzyn, D., Ortiz-Lopez, A., Lobera, M., Yang, J., et al. (2015). Individual intestinal

symbionts induce a distinct population of ROR γ ⁺ regulatory T cells. *Science* 349, 993–997.

Shomuradova, A.S., Vagida, M.S., Sheetikov, S.A., Zornikova, K. V., Kiryukhin, D., Titov, A., Peshkova, I.O., Khmelevskaya, A., Dianov, D. V., Malasheva, M., et al. (2020). SARS-CoV-2 epitopes are recognized by a public and diverse repertoire of human T cell receptors. *Immunity* 53, 1245-1257.e5.

Singer, J.B., Gifford, R.J., Cotten, M., and Robertson, D.L. (2020). CoV-GLUE: A Web Application for Tracking SARS-CoV-2 Genomic Variation. Preprints, 10.20944/preprints202006.0225.v1.

Soto, C., Bombardi, R.G., Kozhevnikov, M., Sinkovits, R.S., Chen, E.C., Branchizio, A., Kose, N., Day, S.B., Pilkinton, M., Gujral, M., et al. (2020). High Frequency of Shared Clonotypes in Human T Cell Receptor Repertoires. *Cell Rep.* 32, 107882.

Tay, M.Z., Poh, C.M., Rénia, L., MacAry, P.A., and Ng, L.F.P. (2020). The trinity of COVID-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.* 20, 363–374.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., et al. (2020). Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *MedRxiv*, 10.1101/2020.12.21.20248640.

Vagin, A., and Teplyakov, A. (2010). Molecular replacement with MOLREP. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 66, 22–25.

Vagin, A.A., Steiner, R.A., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G.N. (2004). REFMAC5 dictionary: Organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 60, 2184–2195.

Venturi, V., Price, D.A., Douek, D.C., and Davenport, M.P. (2008). The molecular basis for public T-cell responses? *Nat. Rev. Immunol.* 8, 231–238.

Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: More and

better reference data for improved all-atom structure validation. *Protein Sci.* 27, 293–315.

Woldemeskel, B.A., Kwaa, A.K., Garliss, C.C., Laeyendecker, O., Ray, S.C., and Blankson, J.N. (2020). Healthy donor T cell responses to common cold coronaviruses and SARS-CoV-2. *J. Clin. Invest.* 130, 6631–6638.

Wrapp, D., Wang, N., Corbett, K.S., Goldsmith, J.A., Hsieh, C.L., Abiona, O., Graham, B.S., and McLellan, J.S. (2020). Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* 367, 1260–1263.

Zhang, J., Wu, Q., Liu, Z., Wang, Q., Wu, J., Hu, Y., Bai, T., Xie, T., Huang, M., Wu, T., et al. (2020). Spike-specific circulating T follicular helper cell and cross-neutralizing antibody responses in COVID-19-convalescent individuals. *Nat. Microbiol.* 6.

Figure 1

Figure 3

A

Clonotype	Donor	TRBV	CDR3 β	TRBJ	TRAV	CDR3 α	TRAJ	β occurrence (%)		β expansion p-value	$\alpha\beta$ occurrence (%)	
								Healthy	COVID		Healthy (pre-pandemic) α	β
1 #	Ts-018	7-3	CASSLAGGSTDTQYF	2-3	8-2	CVVSRITGGGKNTLF	10	14.5	21.9	2.4e-8*	14.8	25.9
	Ts-009	7-3	CASSQAGGSTDTQYF	2-3	2	CAALNAGNMLTF	39					
2	Ts-009	6-1	CASSEGASNPQHF	1-5	12-1	CVVNGSSASKIIF	3	3.1	13.8	5.0e-7*	11.1	3.7
3 #	Ts-018	20-1	CSARAGVQETQYF	2-5	12-3	CAMTYSGGYQKVTF	13	6.0	15.1	0.0001*	22.2	18.5
	Ts-025	20-1	CSARAGVQETQYF	2-5	12-3	CAMTPSGGYQKVTF	13					
4	Ts-025	20-1	CSARDWREETQYF	2-5	8-4	CAVSDQGRDDKIIF	30	2.0	16.6	0.0003*	3.7	0.0
5 # (TCR-017/018)	Ts-017	11-2	CASSQTYEQYF	2-7	12-1	CVVNRGSSYKLIF	12	17.1	24.1	0.003*	14.8	33.3
	Ts-018	11-2	CASSPTYEQYF	2-7	12-1	CVVNRGSSYKLIF	12					
6	Ts-002	20-1	CSAKTSGRGETQYF	2-5	13-2	CAENSGGSNYKLTF	53	1.5	8.8	0.006*	96.3	0.0
7 #	Ts-010	7-9	CASSSREDTQYF	2-3	23	CAATLYSGAGSYQLTF	28	1.7	3.1	0.01*	7.4	3.7
	Ts-013	7-9	CASSSREDTQYF	2-3	23	CAATSYSGAGSYQLTF	28					
8	Ts-013	29-1	CSVYRGLDEKLFF	1-4	23	CAASPPSYGGSQGNLIF	42	0.3	0.1	0.02*	0.0	0.0
9	Ts-025	20-1	CSARDWRGETQYF	2-5	9-2	CALSGGNRDDKIIF	30	1.9	10.3	0.02*	7.4	3.7
10 #	Ts-013	11-2	CASSHGTGGRYEQYF	2-7	13-1	CAAQYALNF	41	1.5	2.6	0.03*	0.0	3.7
	Ts-025	11-2	CASSLGTSGRYEQYF	2-7	13-1	CAAGYALNF	41					

B

C

Figure 4

Figure 5

A

Organism	Clonotype 2 (S ₇₅₃₋₇₅₉)	3 (S ₃₅₃₋₃₆₇)	4, 9 (S ₈₃₃₋₈₄₃)	5 (S ₈₆₉₋₈₇₉)	7 (S ₅₅₇₋₅₆₇)	10 (S ₉₅₇₋₉₆₇)
SARS-CoV-2	LLQYGSF	WNRKRISNCVADYSV	FIK-QYGDCLGD	MIAQYTSALLA	KKFL-PFQQFGR	QALNTLVKQLS
HCoV-NL63	LKQYISA	YYQHTDINF-----	TVDVDYKSCITKG	RMAMYTGSLLIG	QSLAGG---ITY	SALNHLTSQLR
HCoV-OC43	LVEYGSF	WERKTFSNCFNMSS	FVE-AYNNCTGG	QISGYTAAATS	ATYYNSWQNLLY	EALNLLQQLS
HCoV-HKU1	LSEYGTIF	WERKIFSNCFNLST	FVE-AYNNCTGG	QISGYTAAATV	AVYYNSWQNLLY	QALNLLQQLF
HCoV-229E	LKQYISA	YHKHTFIVLY-----	TVDADYKCKITKG	RMAMYTGSLLIG	DTFLNG---ITY	NSLNHLTSQLR
<i>Selenomonas</i>	-	-	-	IAQYTSALLA	-	-
<i>Pseudomonas</i>	-	-	-	MIAQYTSALL	-	-
<i>Klebsiella</i>	LLQYGSF	-	-	-	-	-
<i>Bacteroidales</i>	LLQYGSF	-	-	-	-	-

B

C

D

E

Figure 6

A

B

Donor ID	Ts-003		Ts-005		Ts-038		Ts-022		NS#63		
<i>In vitro</i> stimulation (days)	0	10	0	10	0	10	0	10	0	10	
Responded TCR (#)	α	2	19	3	21	10	33	4	25	0	6
	β	2	19	4	21	10	33	4	25	0	6

C

Figure S1

Figure S2

Figure S3

Figure S4

Figure S6

Table S1. Donor characteristics, related to Figure 1.

Donor ID	Age (years)	Gender	Collection (days)	RBD IgG (μg/ml)	NAb Titer	Peak Disease Severity ^a
Ts-001	73	Male	30	28.5	80	Moderate I
			93	31.6	80	
Ts-002	36	Male	50	4.2	10	Severe
			106	0.8	10	
Ts-004	56	Female	57	12.5	80	Moderate II
			106	5.9	40	
Ts-009	49	Male	51	0.7	5	Severe
			107	0.2	2.5	
Ts-010	55	Male	64	1.5	10	Moderate II
			120	0.2	5	
Ts-013	22	Female	61	7.4	20	Moderate I
			110	3.9	10	
Ts-025	72	Female	41	10.7	20	Moderate II
			83	6.1	20	
Ts-016	28	Male	71	1.1	10	Moderate I
Ts-017	85	Female	132	3.4	10	Severe
Ts-018	71	Female	64	3.1	20	Moderate II
Ts-020	50	Male	42	11.5	20	Severe
Oo-001	31	Male	N/A	N/A	N/A	N/A
HIWA	43	Male	N/A	N/A	N/A	N/A
Ts-003	39	Female	53	2.5	10	Moderate I
Ts-005	53	Male	61	2.8	20	Moderate I
Ts-038	49	Male	96	0.8	5	Moderate I
Ts-022	63	Male	86	1.4	20	Severe
NS#63	42	Male	N/A	N/A	N/A	N/A

RBD, receptor binding domain; NAb, neutralizing antibody; N/A, not applicable.

^aJapanese COVID-19 clinical practice guideline ver. 2.2

Table S2. HLA class II types of involved blood donors, related to Figure 1.

Donor ID	DPA1	DPB1	DQA1	DQB1	DRB1	DRB3/4/5
Ts-001	02:02/02:02	05:01/05:01	03:01/03:02	03:02 ^b /03:03	04:03/12:01 ^c	4*01:04/3*01:01
Ts-002	01:03/02:02	02:01/03:01	05:03/05:08	03:01 ^d /03:01 ^d	12:01 ^c /14:03	3*01:01/3*01:01
Ts-004	01:03/02:02	02:01/05:01	01:02/03:02	03:03/06:04	09:01/13:02	4*01:03/3*03:01
Ts-009	02:02/02:02	02:01/02:02	01:02/03:03	04:01/06:02	04:05/15:01	4*01:03/5*01:01
Ts-010	02:01/02:02	05:01/09:01	01:03/01:03	06:01/06:01	08:03/15:02 ^e	— ^a /5*01:02
Ts-013	01:03/02:02	04:02/05:01	01:02/03:03	04:02/05:02 ^f	04:04/16:02	4*01:03/5*02:02
Ts-025	01:03/02:02	04:02/05:01	01:01/03:01	03:02 ^b /05:01	01:01/04:03	— ^a /4*01:03
Ts-016	01:03/02:02	02:01/05:01	01:03/05:08	03:01 ^d /06:01	12:01 ^c /15:02 ^e	3*01:01/5*01:02
Ts-017	02:02/02:02	02:01/05:01	01:02/01:02	06:02/06:04	13:02/15:01	3*03:01/5*01:01
Ts-018	02:02/02:02	03:01/05:01	01:02/03:02	03:03/06:02	09:01/15:01	4*01:03/5*01:01
Ts-020	02:02/02:02	05:01/05:01	01:02/01:03	06:01/06:02	08:03/15:01	5*01:01/— ^a
Oo-001	01:03/02:02	02:01/04:02	01:01/03:01	03:02 ^b /05:01	01:01/04:03	4*01:03/— ^a
HIWA	01:03/02:02	02:01/05:01	01:05/03:03	03:01 ^d /05:01	04:01/10:01	4*01:02/— ^a
Ts-003	02:02/02:02	05:01/05:01	01:02/05:05	03:01 ^g /05:02 ^f	11:01/15:02 ^e	3*02:02/5*01:01
Ts-005	02:02/02:02	02:01/05:01	01:02/05:07	03:01 ^d /06:02	14:03/15:01	3*01:01/5*01:01
Ts-038	01:03/02:01	02:01 ^h /09:01 ⁱ	01:02/01:03	06:01/06:04	13:02/15:02 ^e	3*03:01/5*01:02
Ts-022	01:03/02:02	02:01/05:01	01:03/01:04	05:03/06:01	14:54/15:02 ^e	3*02:02/5*01:02
NS#63	02:01/02:01	09:01/09:01	01:02/01:03	06:01/06:02	15:01/15:02 ^e	5*01:01/5*01:02

^aA dash (—) indicates no allele expressed.

^{b–i}We cannot exclude ^bDQB1 03:289/03:416, ^cDRB1 12:10, ^dDQB1 03:276N, ^eDRB1 15:140/15:149,

^fDQB1 05:241, ^gDQB1 03:297/03:419, ^hDPB1 352:01 and ⁱDPB1 899:01.

Table S4. Proteins possessing epitopes of TCR-017/018 (IAQYTS~~A~~AL), related to Figure 5.

Cross-reactive protein [Organism]	Identity (%)	Accession
histidine ammonia-lyase [<i>Haloarcula salaria</i>]	100	WP_188853020.1
histidine ammonia-lyase [<i>Haloarcula argentinensis</i>]	100	WP_005537844.1
histidine ammonia-lyase [<i>Halorubellus</i> sp. JP-L1]	100	WP_166041822.1
histidine ammonia-lyase [<i>Deinococcus wulumuqiensis</i>]	100	WP_114672982.1
histidine ammonia-lyase [<i>Deinococcus fonticola</i>]	100	WP_135229009.1
histidine ammonia-lyase [<i>Deinococcus radiodurans</i>]	100	WP_010889407.1
histidine ammonia-lyase [<i>Deinococcus wulumuqiensis</i>]	100	WP_017871478.1
histidine ammonia-lyase [<i>Deinococcus</i> sp. K2S05-167]	100	WP_119766766.1
TPA: histidine ammonia-lyase [<i>Ignavibacteriales</i> bacterium]	100	HCY74963.1
histidine ammonia-lyase [<i>Leptolyngbya valderiana</i>]	100	WP_063716510.1
histidine ammonia-lyase [<i>Anaerolineae</i> bacterium]	100	NPV66375.1
histidine ammonia-lyase [<i>Candidatus</i> Parcubacteria bacterium]	100	PLX26207.1
histidine ammonia-lyase [<i>Candidatus</i> Parcubacteria bacterium]	100	RJQ35361.1
Histidine ammonia-lyase [<i>Synergistetes</i> bacterium ADurb.BinA166]	100	OPZ34295.1
histidine ammonia-lyase [<i>Candidatus</i> Fermentibacter daniensis]	100	NLI03160.1
hypothetical protein AO395_09060 [<i>Candidatus</i> Fermentibacter daniensis]	100	KZD18851.1
histidine ammonia-lyase [<i>Aminivibrio pyruvatiphilus</i>]	100	TDY64917.1
aromatic amino acid lyase [<i>Sciscionella marina</i>]	100	WP_020501521.1
histidine ammonia-lyase [<i>Synergistaceae</i> bacterium]	100	NLM72193.1
aromatic amino acid lyase [<i>Sciscionella</i> sp. SE31]	100	WP_031466349.1
histidine ammonia-lyase [<i>Pseudomonas</i> sp. CMR12a]	100	WP_124346689.1
histidine ammonia-lyase [<i>Gammaproteobacteria</i> bacterium]	100	PCI68862.1
histidine ammonia-lyase [<i>Kangiella</i> sp.]	100	PHS17023.1
histidine ammonia-lyase [<i>Candidatus</i> Lokiarchaeota archaeon]	100	NHJ25594.1
histidine ammonia-lyase [<i>Candidatus</i> Lokiarchaeota archaeon]	100	MBD3194455.1
histidine ammonia-lyase [<i>Deinococcus proteolyticus</i>]	100	WP_013622983.1
histidine ammonia-lyase [<i>Deinococcus piscis</i>]	100	WP_189641912.1
histidine ammonia-lyase [<i>Synergistales</i> bacterium]	100	NCC56342.1
histidine ammonia-lyase [<i>Aminivibrio pyruvatiphilus</i>]	100	WP_133955739.1
hypothetical protein AO394_00950 [<i>Candidatus</i> Fermentibacter daniensis]	100	KZD16260.1
histidine ammonia-lyase [<i>Synergistaceae</i> bacterium]	100	NLD96450.1
histidine ammonia-lyase [<i>Synergistales</i> bacterium]	100	NCB17450.1
histidine ammonia-lyase [<i>Deltaproteobacteria</i> bacterium HGW-Deltaproteobacteria-20]	100	PKN47942.1
Histidine ammonia-lyase [candidate division Hyd24-12 bacterium ADurb.Bin004]	100	OQC66456.1
aromatic amino acid lyase [<i>Synergistaceae</i> bacterium]	100	NLK18261.1
aromatic amino acid lyase [<i>Phycisphaerales</i> bacterium]	100	NRA57503.1
Histidine ammonia-lyase [<i>Synergistetes</i> bacterium ADurb.BinA166]	100	OPZ38667.1

Table S5. Proteins possessing epitopes of clonotype 2 (LLQYGSF), related to Figure 5.

Cross-reactive protein [Organism]	Identity (%)	Accession
Ectopic P granules protein 5 [<i>Dissophora ornata</i>]	100	KAF8936033.1
autotransporter-associated beta strand repeat-containing protein [<i>Akkermansiaceae</i> bacterium]	100	MBC8095538.1
hypothetical protein [<i>Cuniculiplasma divulgatum</i>]	100	WP_077076102.1
NAD-dependent histone deacetylase SIR2 [<i>Talaromyces islandicus</i>]	100	CRG88369.1
excinuclease ABC subunit UvrA [<i>Paenibacillus</i> sp. B01]	100	WP_153783582.1
excinuclease ABC subunit UvrA [<i>Paenibacillus pasadenensis</i>]	100	WP_028597746.1
TonB-dependent receptor [<i>Bacteroidales</i> bacterium]	100	NLC85804.1
TonB-dependent receptor [<i>Bacteroidales</i> bacterium]	100	NLY24744.1
TonB-dependent receptor [<i>Proteiniphilum</i> sp.]	100	MBF6628795.1
TonB-dependent receptor [<i>Proteiniphilum</i> sp.]	100	MBK5194537.1
hypothetical protein [<i>Panacibacter ginsenosidivorans</i>]	100	WP_147189841.1
TonB-dependent receptor [<i>Gramella forsetii</i>]	100	WP_011708978.1
yersiniabactin-iron ABC transporter [<i>Klebsiella pneumoniae</i> subsp. pneumoniae Kp13]	100	AHE43795.1
TPA: hypothetical protein [<i>Porphyromonadaceae</i> bacterium]	100	HBT85488.1
yersiniabactin ABC transporter ATP-binding/permease protein YbtP [<i>Klebsiella pneumoniae</i>]	100	WP_070181785.1
yersiniabactin ABC transporter ATP-binding/permease protein YbtP [<i>Klebsiella pneumoniae</i>]	100	WP_025368101.1
AAA family ATPase [<i>Gemmatimonadetes</i> bacterium]	100	NNE43540.1
hypothetical protein AA313_de0209348 [<i>Arthrobotrys entomopaga</i>]	100	KAF3919251.1
zinc ribbon domain-containing protein [<i>Fictibacillus macauensis</i>]	100	WP_007203242.1
helix-turn-helix domain-containing protein [<i>Flavobacterium</i> sp.]	100	MBL0013419.1
sugar porter family MFS transporter [<i>Dyadobacter jiangsuensis</i>]	100	WP_106596419.1
sugar porter family MFS transporter [<i>Dyadobacter</i> sp. SG02]	100	WP_090384436.1
sugar porter family MFS transporter [<i>Dyadobacter</i> sp. 3J3]	100	WP_159475098.1
sugar porter family MFS transporter [<i>Dyadobacter arcticus</i>]	100	WP_167273011.1
sugar porter family MFS transporter [<i>Dyadobacter fermentans</i>]	100	WP_015811577.1
cell division protein FtsA [<i>Gammaproteobacteria</i> bacterium]	100	MAQ10824.1
TonB-dependent receptor plug domain-containing protein [<i>Bacteroidales</i> bacterium]	100	NLY25865.1
MFS transporter [<i>Candidatus Desulfosporosinus infrequens</i>]	100	WP_106800250.1
M20/M25/M40 family metallo-hydrolase [<i>Proteobacteria</i> bacterium]	100	NCG32350.1
MBL fold metallo-hydrolase [<i>Verrucomicrobia</i> bacterium]	100	NBR86662.1
hypothetical protein B9S33_12740 [<i>Pedospaera</i> sp. Tous-C6FEB]	100	PAW83959.1
MBL fold metallo-hydrolase [<i>Verrucomicrobia</i> bacterium]	100	MBI5802291.1
FAD-dependent oxidoreductase [<i>Bacillus horikoshii</i>]	100	WP_187442366.1
FAD-dependent oxidoreductase [<i>Bacillus</i> sp. m3-13]	100	WP_029326587.1
pyridine nucleotide-disulfide oxidoreductase [<i>Bacillus horikoshii</i>]	100	TYS73947.1
TPA: heat-inducible transcription repressor HrcA [<i>Tissierellia</i> bacterium]	100	HHT20653.1
TPA: hypothetical protein [<i>Ruminococcus</i> sp.]	100	HBI85508.1

Table S6. S864-882-reacted public Tfh clonotypes, related to Figure 6. Clonotypes excluding those with Treg characteristics were assembled with in-house bulk TCR-seq data as well as those in cohort databases. Clonotypes that satisfied following two criteria are shown: 1) expanded during stimulation based on single cell sequencing (> 0.1% of total CD4⁺ T cells) and 2) detected in cohort databases.

Donor	TRAV	CDR3 α	TRAJ	TRBV	CDR3 β	TRBJ	Frequency of $\alpha\beta$ TCR (d10)		Expansion (d0 < d10)		RNA-seq			CITE-seq	Occurrence (%)			
							%	#	α	β	CD200	PDCD1	ICOS	PD-1	Healthy β	COVID β	Healthy (pre-2019)	
																	α	β
Ts-003	20	CAVQLPGTGNQFYF	49	10-3	CAISRLQETQYF	2-5	3.3	252	+	+					0.1	0.1	0.0	0.0
	26-1	CIVRVRYNFKFYF	21	11-2	CASSLLSETQYF	2-5	1.0	78	+	+					0.8	2.5	3.7	7.4
	4	CLVGGDTGFQKLVF	8	11-2	CASSLVSGGNTAEAF	1-1	5.4	421	+	+					0.0	0.3	14.8	0.0
	21	CAVIAGGADGLTF	45	4-2	CASSQDPGQSNQPHF	1-5	0.6	46	+	+					0.9	1.3	0.0	0.0
	8-4	CAVSDRGGTASKLTF	44	5-6	CASSQLPMTGELFF	2-2	2.4	186	+	+					0.0	0.1	11.1	0.0
	26-1	CIVRVAGAAGNKLTF	17	7-2	CASSPDPVGTEAFF	1-1	0.9	70	+	+					0.0	0.1	18.5	0.0
	8-2	CVVILQAGTALIF	15	4-1	CASSYGTGKGEKLF	1-4	0.5	37	+	+					0.0	0.1	0.0	0.0
	12-1	CVVNKGSNYQLIW	33	14	CASSQAYTQYF	2-3	1.4	109	+	+					1.4	2.1	14.8	3.7
	26-1	CIVRVAGGAGDGLTF	45	5-1	CASSLNVLMTAEAF	1-1	0.5	41	+	+					0.1	0.0	7.4	0.0
	26-2	CILRAQGSSEKLVF	57	9	CASSVGGQGINSPLHF	1-6	0.6	44	+	+					0.0	0.6	7.4	0.0
	13-2	CAEILNNAGNMLTF	39	2	CASSEQGETQYF	2-5	0.3	26	+	+					8.9	8.4	25.9	11.1
	35	CAGQARLMF	31	19	CASSVDSMNTAEAF	1-1	0.1	11	+	+					1.1	0.6	0.0	0.0
	19	CALSRGGGADGLTF	45	7-9	CASSFLRGGRTDTQYF	2-3	0.2	12	+	+					0.1	0.1	7.4	0.0
	8-3	CAVGVSGNTGKLIF	37	14	CASSQAGNTIYF	1-3	0.6	45	+	+					8.4	9.1	7.4	0.0
	1-2	CAVRGEASGSARQLTF	22	5-1	CASSPGPTSYEQYF	2-7	0.2	17	+	+					1.5	1.2	0.0	0.0
	10	CVVSPGGYNKLIF	4	3-1	CASSQVRAYEQYF	2-7	0.2	12	+	+					2.9	2.5	7.4	3.7
	8-4	CVPNYGGSGNKLIF	42	7-8	CASSLIVGGYGYTF	1-2	0.1	7	+	+					0.0	0.1	0.0	0.0
	9-2	CALSDTGTASKLTF	44	3-1	CASSQEQTQAF	1-1	0.1	10	+	+					0.0	0.1	44.4	0.0
	23	CAASMGATGNQFYF	49	3-1	CASSLLTAGYTF	1-2	0.2	14	+	+					0.1	0.1	0.0	0.0
Ts-005	4	CLVGGDTGFQKLVF	8	11-2	CASSLVSGGANTEAFF	1-1	4.4	347	+	+					0.1	0.5	14.8	0.0
	26-1	CIATLYNFKFYF	21	11-2	CASSLSLAGELF	2-2	3.3	260	+	+					0.0	0.8	0.0	0.0
	35	CAGQNVNAGNMLTF	39	29-1	CSVGRDFYGYTF	1-2	1.3	100	+	+					0.4	0.1	0.0	0.0
	9-2	CAPLNSGGSNYKLTF	53	7-3	CASSLYNEQFF	2-1	1.8	139	+	+					5.2	4.7	3.7	0.0
	13-1	CALHTGGFKTIF	9	20-1	CSARDPTGRNTEAFF	1-1	1.7	132	+	+					2.4	1.6	11.1	3.7
	21	CAVHGNDMRF	43	29-1	CSVEQGMNTEAFF	1-1	0.9	68	+	+					8.4	7.2	3.7	3.7
	30	CGTATYNTDKLIF	34	11-2	CASSFGTGELFF	2-2	1.0	77	+	+					6.2	6.4	0.0	3.7
	27	CAGVYTNAGKSTF	27	20-1	CSARDSGGISSETQYF	2-5	0.5	37	+	+					0.3	0.0	0.0	0.0
	12-1	CVVNKGSNYQLIW	33	14	CASSQGFVAF	1-1	1.1	85	+	+					0.3	1.1	14.8	0.0
	8-3	CAVVIPLSGSYIPTF	6	18	CASSRGGSYEQYF	2-7	0.1	8	+	+					4.2	4.9	0.0	3.7
	6	CVATTSGNQFYF	49	5-4	CASSLRGRSTGELFF	2-2	0.5	37	+	+					0.3	0.3	0.0	0.0
	8-4	CAVTLTGANNLFF	36	11-2	CASSLISDTQYF	2-3	0.2	12	+	+					2.7	2.3	7.4	7.4
	38-2	CAYRVAGATSYGKLTFF	52	5-1	CASSFSQGNTEAFF	1-1	0.3	24	+	+					0.0	0.1	0.0	0.0
	8-6	CAVNPIAGGTSYGKLTFF	52	20-1	CSARAGQGNTAEAF	1-1	0.1	9	+	+					6.9	7.2	0.0	3.7
	6	CALVTNQAGTALIF	15	7-3	CASSLSGGQPQHF	1-5	0.2	13	+	+					2.7	4.5	0.0	0.0
	12-1	CVVNPNARLMF	31	5-1	CASSLLTGSTDTQYF	2-3	0.4	34	+	+					2.7	4.1	11.1	0.0
	26-1	CIVRPANQAGTALIF	15	3-1	CASSQVAGEQFF	2-1	0.3	22	+	+					5.5	10.4	3.7	11.1
	9-2	CALSVTNTPLVF	29	30	CAWVTGVSNTIYF	1-3	0.5	39	+	+					0.0	0.1	0.0	0.0
	23	CAASETGGSNYKLTF	53	5-4	CASSPGLARYEQYF	2-7	0.4	35	+	+					0.9	0.6	0.0	0.0
	17	CAAYTTGGSEKLVF	57	13	CASSLRSTGIPYEQYF	2-7	0.3	21	+	+					0.0	0.1	0.0	0.0
Ts-038	26-1	CIVRVGAQKLVF	54	14	CASSQDLFWGYTF	1-2	0.5	37	+	+					0.0	0.3	44.4	0.0
	13-1	CAASSNNFNKYF	21	20-1	CSARPGVDTQYF	2-3	6.8	376	+	+					1.1	2.1	7.4	0.0
	14	CAMRGVGGATNKLIF	32	28	CASSLGTDYNEQFF	2-1	0.7	39	+	+					2.5	2.1	0.0	3.7
	29	CAVTNFGNEKLTFF	48	5-4	CASSLDLKETQYF	2-5	2.4	133	+	+					0.6	1.1	0.0	0.0
	26-1	CIVRGANQAGTALIF	15	3-1	CASSQVSAEQFF	2-1	0.8	44	+	+					1.1	0.8	11.1	3.7
	17	CATDAFGNEKLTFF	48	7-3	CASSRTGGSGGYTF	1-2	0.9	48	+	+					0.0	0.1	18.5	0.0
	29	CAARHDTGGFKTIF	9	25-1	CASSAGHEQFF	2-1	1.2	64	+	+					0.1	0.1	0.0	3.7
	12-1	CVVNPSSTNGKLTFF	37	14	CASSHDSNYGYTF	1-2	2.1	115	+	+					0.8	1.6	3.7	0.0
	12-1	CVVNPNARLMF	31	5-1	CASSSLTGSTDTQYF	2-3	0.8	42	+	+					1.8	2.9	11.1	0.0
	13-1	CAALPNFNNEKLTFF	48	5-1	CASSPSGGGETQYF	2-5	0.3	17	+	+					4.6	5.9	0.0	22.2
	25	CAGRRLTGGGNKLTF	10	5-1	CASSTRPSGGYQETQYF	2-5	0.9	50	+	+					0.0	0.1	0.0	0.0
	12-3	CAMSVFIFVIRGGSEKLVF	57	20-1	CSAPPGTGGEQYF	2-7	0.1	6	+	+					0.4	0.1	0.0	0.0
	8-3	CAVGALNNNDMRF	43	25-1	CASSDRQGAGGELFF	2-2	0.1	7	+	+					0.3	0.2	7.4	0.0

(Continued on next page)

(continued)

Donor	TRAV	CDR3 α	TRAJ	TRBV	CDR3 β	TRBJ	Frequency of $\alpha\beta$ TCR (d10) (%)	Expansion (d0 < d10) α β	RNA-seq			CITE-seq PD-1	Occurrence (%)			
													Healthy β	COVID β	Healthy (pre-2019)	
															α	β
Ts-038	8-1	CAVNAVNRDDKIIF	30	2	CASSEGGTTLYEQYF	2-7	0.4 23	+ +					0.0	0.1	11.1	0.0
	29	CAASPAQGGSEKLVF	57	7-9	CASSLDWYEQYF	2-7	0.1 6	+ +					2.3	2.1	0.0	7.4
	21	CAVNSGGGYNKLIF	4	7-2	CASSLGLAREQYF	2-7	0.2 11	+ +					2.7	3.9	0.0	11.1
	21	CAVRHRAGNTPLVF	29	19	CASSRTPGGINYGTYF	1-2	0.4 21	+ +					0.0	0.1	0.0	0.0
	8-3	CAVVFTSGTYKYIF	40	4-3	CASSQGLAGGETQYF	2-5	0.1 6	+ +					6.5	4.5	7.4	14.8
	19	CALSEARGFKTIF	9	7-3	CASSLKSQETQYF	2-5	0.2 11	+ +					1.0	0.2	0.0	3.7
	17	CATEGELGDDKIIF	30	2	CASSRGPSETQYF	2-5	0.1 7	+ +					0.1	0.1	0.0	0.0
	26-1	CIAQSNNGYALNF	41	29-1	CSVAGGGGEKLVF	1-4	0.1 6	+ +					1.0	0.7	0.0	0.0
	8-3	CAVGGQGFQNVLHC	35	15	CATSRLAGEQFF	2-1	0.1 7	+ +					0.4	0.6	0.0	0.0
	8-3	CAVVDNMNRDDKIIF	30	6-5	CASRNDNYGYTF	1-2	0.1 6	+ +					0.3	0.1	0.0	0.0
	13-1	CAAVSVGSARQLTF	22	5-1	CASSLGGGYNQEFF	2-1	0.1 6	+ +					15.5	13.8	0.0	37.0
	12-1	CVVNMFNARLMP	31	19	CASSILAGAYEQYF	2-7	0.3 17	+ +					8.8	8.8	0.0	7.4
	14	CAMREKDNFNKPYF	21	5-1	CASSLGTGINTAEPF	1-1	0.2 12	+ +					2.2	1.6	0.0	0.0
	38-1	CAATGNGQFYF	49	3-1	CASSQGLAGLTDQYF	2-3	0.2 9	+ +					1.8	2.4	0.0	3.7
	9-2	CALFTRGSNYKLTF	53	20-1	CSARGLYEQYF	2-7	0.1 6	+ +					12.7	11.6	0.0	14.8
	8-3	CAVGNPFGNEKLTF	48	2	CASREDRVGANVLTF	2-6	0.1 6	+ +					0.0	0.1	18.5	0.0
	13-1	CAASKAGGGNKLTF	10	6-1	CASDGNNGYTF	1-2	0.1 6	+ +					0.9	0.1	51.9	0.0
	29	CAASGRNTGNQFYF	49	5-1	CASSLGVGRYEQYF	2-7	0.1 6	+ +					1.3	0.5	7.4	0.0
	35	CAGLFGFRTSGSRLTF	58	20-1	CSASRLAGSEQYF	2-7	0.1 7	+ +					0.6	0.6	0.0	0.0
	9-2	CALSGSGNTPLVF	29	20-1	CSARDRANTGELFF	2-2	0.2 10	+ +					10.3	12.5	18.5	14.8
	9-2	CALNTNAGKSTF	27	10-2	CASSENNGRYEQYF	2-7	0.1 8	+ +					1.8	2.1	70.4	0.0
Ts-022	12-1	CVVSGGSNYKLTF	53	11-2	CASSLSPEQYF	2-7	4.2 231	+ +					0.9	1.6	11.1	0.0
	27	CAGEWNTPLVF	29	3-1	CASSQTQYNEQFF	2-1	1.2 65	+ +					0.5	0.6	3.7	0.0
	27	CAGAGGNTPLVF	29	3-1	CASSQTQYSYEQYF	2-7	6.0 334	+ +					0.1	0.1	3.7	0.0
	12-2	CVGNFNKPYF	21	4-2	CASSQVKQSNEKLFF	1-4	8.2 457	+ +					0.3	0.0	0.0	0.0
	8-3	CAVGAPDNYQLIW	33	5-1	CASSRTVSGELFF	2-2	2.3 127	+ +					0.1	0.1	11.1	0.0
	3	CAVSIEDTGRRALT	5	5-1	CASSLSVSSEQYF	2-7	1.6 87	+ +					1.0	0.7	0.0	0.0
	25	CAGAGNNRKLW	38	3-1	CASSRSGAGTDQYF	2-3	1.0 55	+ +					0.0	2.8	11.1	0.0
	13-2	CAESLGSWKLQF	24	20-1	CSARARDRVAEAF	1-1	0.6 36	+ +					0.0	0.1	0.0	0.0
	12-1	CAVRSNTGKLIF	37	11-2	CASSYGPLEQFF	2-1	1.0 56	+ +					0.1	0.0	0.0	0.0
	21	CAVYTGTASKLTF	44	7-9	CASGKGGTGELFF	2-2	0.6 35	+ +					0.0	0.1	40.7	0.0
	14	CAMREGTDSYKLIF	12	20-1	CSASLARDQPQHF	1-5	0.1 7	+ +					0.4	0.2	3.7	0.0
	12-3	CAMRAESSGSARQLTF	22	5-1	CASSRTVTGELFF	2-2	0.6 35	+ +					3.1	2.0	0.0	7.4
	12-3	CAMRSSSGGSNYKLTF	53	9	CASSVVSAYEQYF	2-7	0.3 19	+ +					1.3	0.7	0.0	0.0
	26-1	CIPETGANNLFF	36	7-3	CASSLYAGNYGYTF	1-2	0.3 14	+ +					0.3	0.3	0.0	0.0
	25	CAGRFEAGSARQLTF	22	5-1	CASSFSLTGELFF	2-2	0.2 9	+ +					1.3	1.1	0.0	14.8
	1-1	CAPGGSYIPTF	6	5-1	CASSLSLTGEQYF	2-7	0.2 10	+ +					0.5	0.2	0.0	0.0
	12-1	CAVNGGNGQFYF	49	5-1	CASSLAQPGETQYF	2-5	0.2 9	+ +					1.5	0.7	3.7	3.7
	17	CATDRVCVRDRDDKIIF	30	5-1	CASSPGPTGEQFF	2-1	0.2 9	+ +					0.0	0.3	0.0	0.0
	5	CAESTGGYQKVTF	13	11-3	CASSRLTGGGDTQYF	2-3	0.2 11	+ +					0.1	0.0	40.7	0.0
	12-3	CAMRPISGGSARQLTF	22	5-1	CASSLSITGEAEPF	1-1	0.3 15	+ +					0.1	0.0	0.0	0.0
	1-1	CAVRSISSGSARQLTF	22	5-1	CASSLTVTSYEQYF	2-7	0.3 15	+ +					1.0	1.6	7.4	7.4
	22	CAVARDVAGSARQLTF	22	5-1	CASSPHVNTAEPF	1-1	0.2 10	+ +					0.3	0.2	0.0	0.0
	1-2	CVLMDSSYKLIF	12	6-1	CASSEAPGFSYNEQFF	2-1	0.1 6	+ +					0.0	0.1	0.0	0.0
	17	CATEATGANSKLTF	56	7-3	CASSPIGSTDTQYF	2-3	0.2 10	+ +					0.6	0.3	0.0	0.0
	38-1	CAFTSGGSNYKLTF	53	6-5	CASSYGDSSDTQYF	2-3	0.1 8	+ +					0.1	0.1	3.7	0.0
NS#63	13-1	CAASIYTDKLIF	34	20-1	CSALGAGEQFF	2-1	0.9 43	+ +					0.0	0.2	7.4	0.0
	35	CAGLLSGSARQLTF	22	6-5	CASSPGLAGGYEQFF	2-1	0.2 10	+ +					0.4	0.4	7.4	0.0
	39	CAVDPRETGANSKLTF	56	19	CASSISHEKLFF	1-4	0.2 10	+ +					0.1	0.5	0.0	0.0
	4	CLVAKPKAAGNKLTF	17	27	CASSPAGVNEKLFF	1-4	0.1 7	+ +					0.6	0.7	0.0	0.0
	9-2	CALSDRRRAAGGADGLTF	45	15	CATSRELLNTGELFF	2-2	0.1 6	+ +					0.4	0.3	0.0	0.0
	29	CAASTKTSYDKVIF	50	12-3	CASSLKGSPLHF	1-6	0.1 7	+ +					1.5	3.7	7.4	0.0

min max n = 786 1413 27

Table S7. Data collection and refinement statistics of the crystallographic analysis of TCR $\alpha\beta$ -017, related to Figure 4.

	TCR $\alpha\beta$ -017
Data collection statistics	
Space group	$C222_1$
Cell constants (Å)	$a=81.1, b=81.2, c=315.1$
Resolution (Å) ^a	46.4-2.18 (2.30-2.18)
R_{sym} (%) ^a	10.0 (88.0)
Completeness (%) ^a	99.8 (99.6)
$\langle I/\sigma \rangle$ ^a	13.7 (2.3)
Multiplicity ^a	6.8 (7.1)
Refinement statistics	
Resolution (Å)	42.4-2.18
R (%)	22.2
R_{free} (%)	27.6
Root mean square deviations from ideal values	
Bond length (Å)	0.008
Bond angle (°)	1.038
Ramachandran plot	
Favored (%)	93.8
Outlier (%)	0.9
Rotamer outlier (%)	0.4

^aValues in the parenthesis are in the highest resolution shells.