SARS-CoV-2 epitopes are recognized by a public and diverse repertoire of human T-cell receptors

Alina S. Shomuradova[1], Murad S. Vagida[1], Savely A. Sheetikov[1], Ksenia V. Zornikova[1], Dmitriy Kiryukhin[1], Aleksei Titov[1], Iuliia O. Peshkova[1], Alexandra Khmelevskaya[1], Dmitry V. Dianov[1], Maria Malasheva[1], Anton Shmelev[1], Yana Serdyuk[1], Dmitry V. Bagaev[2], Anastasia Pivnyuk[3], Dmitrii S. Shcherbinin[4,5], Alexandra V. Maleeva[1], Naina T. Shakirova[1], Artem Pilunov[1], Dmitry B. Malko[1], Ekaterina G. Khamaganova[1], Bella Biderman[1], Alexander Ivanov[6], Mikhail Shugay[3,4,5*] and Grigory A. Efimov[1**]

1. National Research Center for Hematology, Moscow, Russia
2. Eindhoven University of Technology, Eindhoven, Netherlands
3. Center of Life Sciences, Skolkovo Institute of Science and Technology
4. Pirogov Russian Medical State University, Moscow, Russia
5. Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia
6. Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

*e-mail: mikhail.shugay@gmail.com twitter: @antigenomics
**e-mail: efimov.g@blood.ru twitter: @grigory_efimov

Abstract

Understanding the determinants of adaptive immune response to SARS-CoV-2 is critical for fighting the ongoing COVID-19 pandemic. Here we assayed both antibody and T-cell reactivity to SARS-CoV-2 antigens in COVID-19 convalescent patients and healthy donors sampled before and during the pandemic. Our results show that while anti-SARS-CoV-2 antibodies can distinguish convalescent patients from healthy donors, the magnitude of T-cell response was more pronounced in healthy donors sampled during COVID-19 pandemic than in donors sampled before the outbreak. This hints at the possibility that some individuals have encountered the virus but were protected by T-cell cross-reactivity observed. A public and diverse T-cell response was observed for two A*02-restricted SARS-CoV-2 epitopes, revealing a set of T-cell receptor motifs displaying germline-encoded features. Bulk CD4+ and CD8+ T-cell response to SARS-CoV-2 glycoprotein S is characterized by multiple groups of homologous T-cell receptor sequences some of which are shared across multiple donors, indicating the existence of immunodominant epitopes. Overall, our findings indicate that T cells form an efficient response to SARS-CoV-2 and alongside the antibodies can serve as a useful biomarker for surveying SARS-CoV-2 exposure and immunity. We hope that data, including the set of specific T-cell receptors identified in this study can serve as a basis for future developments of SARS-CoV-2 vaccinations and monitoring.

Keywords

*SARS-CoV-2, COVID-19, T-cell response, T-cell epitopes, T-cell receptor, TCR sequencing, clonal structure, MHC, HLA-A*02:01*

Introduction

SARS-CoV-2 is currently causing the global pandemic of COVID-19 (*1-5*). Elucidating the mechanisms of the adaptive immune response to SARS-CoV-2 is crucial for the prediction of vaccine efficacy and assessment of the possibility of reinfection.

It is commonly assumed that the antibody response is required for viral clearance (*6*). Multiple serological tests for detection of the SARS-CoV-2-specific antibodies are being developed (*7-9*) and massive efforts are undertaken to estimate the number of seropositive individuals in the population. Moreover, monoclonal antibodies (*10-12*) and plasma of the convalescent patients were proposed for COVID-19 therapy (*13, 14*). Administration of monoclonal neutralizing antibodies against glycoprotein S (S-protein) of SARS-CoV-2 protected experimental animals from the high dose of SARS-CoV-2 (*15*). However, around 30% of convalescent patients have no or very low titers of SARS-CoV-2 neutralizing antibodies (*16*) suggesting the involvement of the other immune mechanisms in the viral clearance. There is strong evidence of an important role of T-cellular immunity in the clearance of the respiratory viruses, such as SARS-CoV which caused atypical pneumonia outbreak in 2003. Memory T-cell responses to the SARS-CoV epitopes were detectable in 50% of convalescent patients 12 months post-infection (*17*). Moreover, CD8+ cells specific to the immunodominant epitope of S-protein of the

mouse-adapted SARS-CoV strain protected aged mice from otherwise lethal infection (*18*). In another study the adoptive transfer of SARS-CoV-specific CD8+ or CD4+ cells or immunization with peptide-pulsed dendritic vaccine reduced virus titers in the lungs and enhanced survival of mice, thus proving that T cells are sufficient for virus clearance in the absence of antibodies or activation of the innate immune system (*19*). The other studies suggest a leading role of CD4+ cells in the SARS-CoV clearance. Depletion of the CD4+ T cells at the time of infection delayed the virus clearance whereas depletion of CD8+ T cells had no such effect (*20*). Furthermore, CD4+ cell response correlated with the positive outcome of SARS-CoV infection (*19*).

In humans, the severe SARS-CoV infection was characterized by the delayed development of the adaptive immune response and prolonged virus clearance (*21*). Decreased numbers of T cells strongly correlated with disease severity (*22*). T cells not only contributed to the resolution of the disease but also formed a long-lasting memory response. The CD8+ and CD4+ cells isolated from the SARS-CoV convalescent patients 4 years after recovery demonstrated IFNγ secretion in response to stimulation by peptide pools derived from S- N- and M-proteins of SARS-CoV (*23*). SARS-CoV antigen-specific T cells persisted in the convalescent patients for up to 11 years post-infection. The S-protein was shown to be the most immunogenic among all the SARS-CoV antigens (*17, 23*). Moreover, the particular CD4+ and CD8+ T-cell epitopes of SARS-CoV S-protein were identified (*24*). Two CD8+ epitopes were presented in HLA-A*02:01 and elicited a specific response in the SARS-CoV convalescent patients but not in healthy donors (*24*). Additionally, several CD8+ T-cell epitopes were identified and characterized in the M (*25, 26*) and N (*26*) proteins of SARS-CoV.
At present, growing evidence confirms the importance of T-cell response to the SARS-CoV-2 in disease control. High CD8+ T-cell counts in the lungs correlated with better control of SARS-CoV-2 progression (*27*).

The presence of the T-follicular helper cells and CD8+ T cells with activated phenotype in the blood at the time of virus clearance suggests their active involvement in the immune response in the recovered patient (*28*). On the contrary, the exhausted phenotype of CD8+ T cells in the peripheral blood may serve as an indicator of poor disease prognosis (*29*).

Most probably, the immune memory is capable of protecting against the SARS-CoV-2. The studies on primates demonstrated that the repeated virus challenge failed to provoke reinfection once the virus was eliminated (*30*).

The T cells of convalescent patients are responsive to stimulation by the peptide pools covering the SARS-CoV-2 proteins. In particular, S-protein of SARS-CoV-2 was a strong inducer of Th1-type response in the CD4+ cells (*31*).

Interestingly the CD4+ cells reactive to the SARS-CoV-2 antigens were found not only in the COVID-19 convalescent patients but also in healthy donors that most likely could be explained by the T-cell cross-reactivity (*32*). Further studies are necessary to clarify whether this preexisting T-cell response is protective.

Other studies hinted on the possible role of patient HLA-genotype in the reaction to the virus. According to the bioinformatic predictions, some HLA alleles present more of the SARS-CoV-2 epitopes than the others, possibly affecting the severity of the COVID-19 (*33*). Some known immunogenic T-cell epitopes of SARS-CoV are conserved in SARS-CoV-2 (*34*), suggesting that they might also play a role in the immune response to the SARS-CoV-2. However, no experimental data about the targets of T-cell reactivity is currently available.

Thus, we analyzed the immune response to SARS-COV-2 in the COVID-19 convalescent patients (CP) aiming to describe the underlying structure, clonality, and epitope-specificity of the T-cellular immune response to the SARS-CoV-2 S-protein.

Results

SARS-CoV-2-specific T cells but not antibodies are present in healthy donors sampled during COVID-19 pandemic

To study the adaptive immune response to SARS-CoV-2 we recruited 31 COVID-19 convalescent patients (CP). According to the classification by the US National Institutes of Health (*35*) the patients were categorized as having asymptomatic (n = 2), mild (n = 19) and moderate/severe (n = 10) clinical types of disease. None of the patients required either treatment in the intensive care unit, oxygen supplementation, or invasive ventilation support. The cohort was gender-balanced (15 males, 16 females) and donor age was from 17 to 59 years with a median of 35 years. Peripheral blood was collected at days 17 - 49 (median 33) after the onset of the disease or positive PCR-test (Fig.1a). The control group of healthy donors was formed by recruiting 14 healthy volunteers during the COVID-19 pandemic - HD (CoV) with no symptoms and a negative PCR test. Additionally, we used 10 samples of peripheral blood mononuclear cells (PBMC) from biobanked healthy hematopoietic stem cell donors - HD (BB), which were cryopreserved no later than September 2019, and 10 serum samples of healthy blood donors cryopreserved no later than 2017 - HD (S). The presence of SARS-CoV-2 specific antibodies was tested in the serum samples. Cell samples were used to study T-cellular response to the peptide pools of M-, N-, S- proteins of SARS CoV-2, and to the recombinant surface glycoprotein S (S-protein) as well as to determine the repertoire of T-cell receptors (TCRs) of S-protein specific cells (Fig.1b).

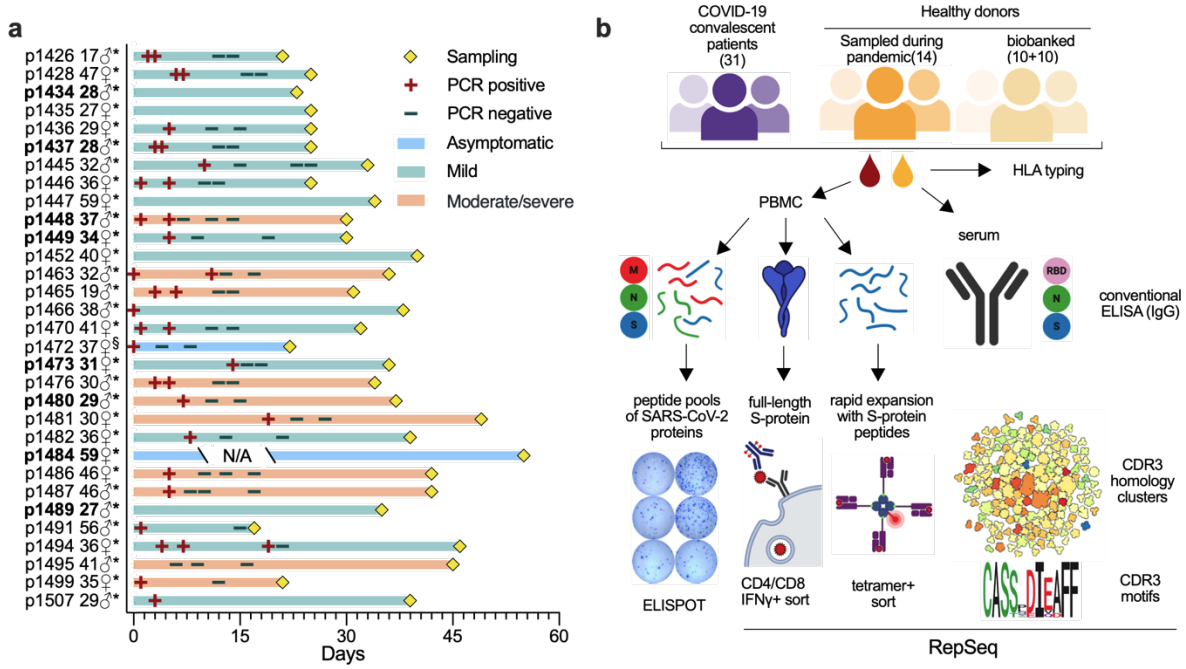


Fig.1 Experimental pipeline and COVID-19 patient data. a, Clinical data of COVID-19 CP. Age and gender are indicated on the left of the Y-axis. The time point of sampling, results of PCR tests and severity of disease are provided on the swimmer's plot. * - days are calculated since the onset of symptoms, § - days are calculated since the positive PCR test. N/A - information is not available. Donors for which the TCR repertoire was analyzed are shown in bold. **b**, The cohorts of COVID-19 convalescent patients - CP (n=31), healthy donors sampled during COVID-19 pandemic - HD(CoV) (n=14) and obtained from the biobank - HD(BB) (n=10) and serum bank - HD(S) (n=10) were included in study. Peripheral blood was collected and used for in vitro assays with several SARS-CoV-2 antigens. Antigen-specific IFNγ production by T cells in CP to S-protein and peptide pools of SARS-CoV-2 (M, N, and S) was measured by ELISPOT. The S-protein-directed T-cell response was assessed by IFNγ secretion assay after stimulation of peripheral blood mononuclear cells (PBMC) with recombinant S-protein, FACS-sorted IFNγ-secreting CD4+ and CD8+ T-cells were used for sequencing their T-cell receptor (TCR) beta repertoire. Simultaneously, rapid in vitro expansions with predicted HLA-A*02:01-presented S-protein derived peptides were set up, tetramer-positive cells were FACS-sorted and TCR alpha and beta repertoire sequencing was performed. Bioinformatic analysis was used to reveal shared CDR3 motifs in antigen and epitope-specific populations. Antibody-mediated immune response in CP and HD was measured by conventional ELISA with recombinant N-, S-protein, and RBD domain of S-protein. Illustration was created with BioRender.

Analysis of the humoral immune response to SARS-CoV-2 demonstrated that the IgG antibodies of the majority of COVID-19 -CPs were specific to one or more viral antigens. IgGs from HD (CoV) and HD (BB) groups showed no reactivity to S-protein of SARS-CoV-2 or its receptor-binding domain (RBD) (Fig.2a,b). The presence of antibodies specific to the nucleoprotein (N-protein) of SARS-CoV-2 in HD could be explained by either their cross-reactivity or by recognition of some bacterial products co-purified with the antigen (Fig.2a and Extended Data Fig.1a).

Despite the variability of the antibody response, response to all three antigens distinguished CP from HD (Fig.2a,b and Extended Data Fig.1a,b) using RBD resulted in the lowest background. Levels of IgG antibodies specific to different antigens positively correlated in CP (Fig.2e and Extended Data Fig.1c-e) with the strongest correlation between RBD and S-protein. Only two patients (p1472 and p1473) did not demonstrate IgG response to any of the tested viral antigens.

T-cell response to the SARS-CoV-2 S-protein was highly variable across the donors with some CP lacking detectable virus-reactive T cells (Fig.2c,d). We have not observed any clear association between the magnitude of T-cell response and the HLA genotype of the donor (Extended Data Table1), time after the disease onset, or patient age (Extended Data Fig.2). We observed only mild correlation between the magnitude of T-cell and humoral response in our cohort whereas the magnitude of CD8+ and CD4+ responses were interdependent (Fig.2e).

All tested HD(CoV) lacked antibodies against any of SARS-CoV-2 antigens, which probably excludes the possibility of past asymptomatic infection. Surprisingly some of them had frequencies of S-protein specific T-cells comparable with those in the CP group (Fig.2c). Apart from the significant difference in T-cell response between CP and HD(BB), we also observed the significant increase of both CD4+ and CD8+ S-protein specific T-cells in HD(CoV) as compared with HD(BB) group (Fig.2d). This might indicate that some HD(CoV) were exposed to the virus but did not develop the disease.

S-protein specific T cells in CP exhibited conventional phenotype distribution typical to CD4+ and CD8+ cells. S-protein reactive CD4+ lymphocytes were represented predominantly by central memory phenotype (CD45RO+, CD197+), followed by the effector memory phenotype (CD45RO+, CD197-). Antigen-specific CD8+ cells had mostly effector memory phenotype, followed by the terminal effector cells (CD45RO-, CD197-) (Fig.2f). Nevertheless, we observed a significant increase of activated (CD38+, HLA-DR+) CD4+ in CP group compared to HD(Cov) (Fig.2g). The level of PD-1 expression by CD4+ but not CD8+ cells was significantly higher in the IFNγ-secreting population (Fig.2h).

We also measured T-cell immune response to the recombinant SARS-CoV-2 S-protein using ELISPOT and to the peptide pools covering SARS-CoV-2 S, M, and N-proteins. Some patients responded to recombinant S-protein while they demonstrated no response to S-protein derived peptide pools (Extended Data Fig.3). This might be explained by the incomplete coverage of the protein sequence (refer to the discussion for details). Activation of T cells upon full-length S-protein stimulation was equally effective in both CD4+ and CD8+ lymphocytes, regardless of a better potency for presentation HLA class II epitopes in such conditions (Extended Data Fig.3a,b) The response to the proteins was significantly different (M- vs N-protein, $p=0.0352$), as well as varied significantly ($p<0.0001$) across the individuals (Extended Data Fig.3c,d).

All CP exhibited either CD8+ or CD4+ T-cell reactivity to at least one of the proteins of SARS CoV-2 (Fig.2c, Extended Data Fig.3).

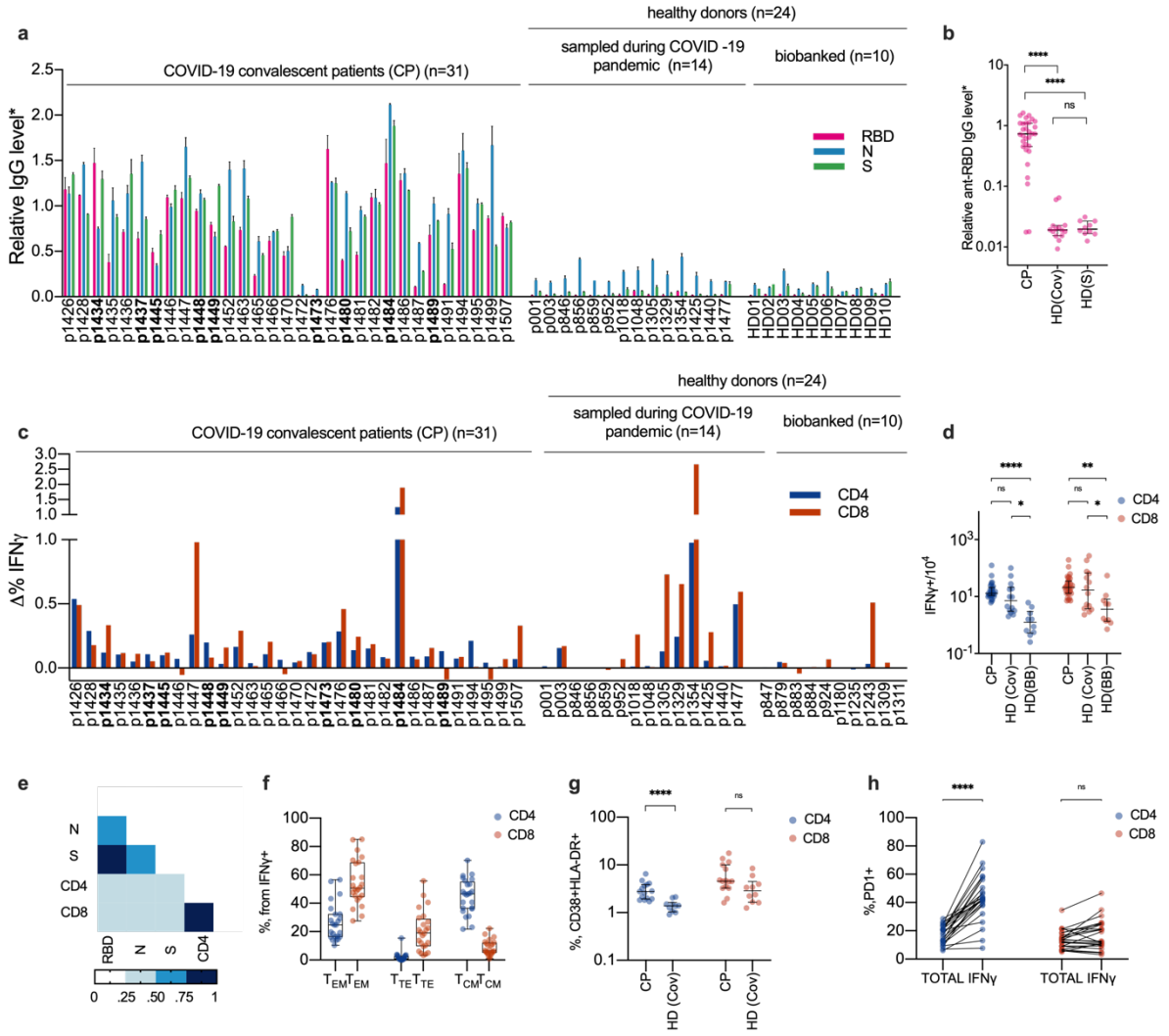


Fig.2 Healthy donors sampled during the COVID-19 pandemic have increased numbers of SARS-CoV-2-specific T cells but not antibodies. **a-b**, Relative levels of anti-RBD, anti-N and anti-S IgG were measured by conventional ELISA in COVID-19 CP (n=31), HD (CoV) (n=14) and HD (S) (n=10). Plotted data are mean of two independent measurements ± SD (**a**) and medians with bars representing interquartile range **(b)**. (**c-d**), T-cell response to S-protein was measured by frequencies of IFNγ-producing CD4+ and CD8+ cells in COVID-19 CP (n=31), HD (CoV) (n=14) and HD (BB) (n=10). Data are delta between frequencies of activated T-cells in stimulated and unstimulated samples **(c)** and medians of the frequencies with bars represent the interquartile range **(d)**. **e**, Correlation between relative levels of antibodies and T-cell response in CP group (n=31). Heat-map of intervals of spearman's coefficients of correlation is shown. **f**, Phenotype of IFNγ-secreting CD4+ and CD8+ cells in CP (n=24). TEM - T effector memory (CD4+5RO+, CD197-), TTE - T terminal memory (CD4+5RO-, CD197-) and TCM - T central memory (CD4+5RO+, CD197+). Box represents interquartile range with the median line, whiskers represent min and max. (n=24) **g**, Percentage of activated (CD38+ HLA-DR+) CD4+ and CD8+ cells in CP (n=15) and HD (CoV) (n=10) groups. Data are median with interquartile range. **h**, Comparison of frequencies between PD1+ cells in CD4+ and CD8+ not-naive T cells (TOTAL) and not-naive activated subpopulation (IFNγ) in CP group (n=24). For group comparison, we used the Kruskal–Wallis test and Dunn's multiple comparison test (**b**,**c**) and Mann-Whitney test (**g**, **h**). p-value *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001. CP - convalescent patients, HD (CoV) - healthy donors sampled during COVID-19 pandemic, HD (BB), and HD (S) - biobanked healthy donor cell or serum samples, respectively.

Immune response to two HLA-A*02:01-restricted SARS-CoV-2 S-protein epitopes allowed to discriminate convalescent patients from healthy donors

The most common MHC I allele in the CP cohort was HLA-A*02:01 (Extended Data Table1) present in 14 of the 31 patients. We selected 13 potential S-protein epitopes predicted to be presented by HLA-A*02:01, some of them shared 100% sequence homology with SARS-CoV and were previously shown to be immunogenic (Table1). Total magnitude of S-protein directed CD8+ response was in some patients less than 0.1% from the total CD8+ population, so we decided to perform rapid

in vitro antigen-specific expansion according to the protocol which was previously proposed to expand antigen-specific cells from the memory subpopulation (*36*). Epitope-specific cells were detected by flow cytometry using MHC-tetramers (Fig.3a,b). Strikingly, 12 of 14 CPs with HLA-A*02:01 allele reacted to a single epitope of S-protein$_{269\text{-}277}$ YLQPRTFLL (YLQ), while only 1 HD(Cov) had the cells recognizing this epitope. The response to the second epitope S-protein$_{1000\text{-}1008}$ RLQSLQTYV (RLQ) was also, though to a somewhat lesser extent, pointing to the past COVID-19 (Fig.3a). Share of the positive wells in expansion for both epitopes could be used to confirm previous COVID-19 (Fig.3c). Almost complete absence of the response to these epitopes in HD confirmed that the detected antigen-specific cells were derived from the memory subpopulation. The remaining 11 of the tested epitopes yielded only sporadic T-cell response (Fig.3a).

To describe the structure and clonality of SARS-CoV-2 directed T-cell immune response we performed analysis of T-cell receptor (TCR) repertoires of FACS-sorted IFNγ-secreting CD8+/CD4+ cells and MHC-tetramer-positive populations as well as the total fraction of PBMC by high throughput sequencing using the Illumina platform.

Antigen and epitope-specific TCRs were defined as being strongly and significantly enriched in the respective fraction (Fig.3d).

Clonality of IFNγ-secreting cells was higher in CD4+ T cells (Fig.3e) with the number of IFNγ-secreting CD4+ clones ranging between 19 and 545 (median=115). The CD8+ IFNγ-secreting cell population had a median of 18 (2 to 74) clones (Fig.3e). Sequences of YLQ- and RLQ-specific clonotypes were deposited to the VDJdb database (vdjdb.cdr3.net). Clonotypes significantly enriched in the IFNγ-secreting CD4+ and CD8+ populations are listed in Extended Data Table 2.

We observed only a negligible intersection between MHC-tetramer-positive and IFNγ-secreting populations, resulting from a minimal presence of RLQ- and YLQ-specific clones in the peripheral blood. The YLQ-specific cells were significantly more diverse than the RLQ-specific clones with a median of 37 and 8 clones per individual, respectively (Fig.3f). T-cell clones specific for both antigens were either undetectable or observed with a very low frequency in the total TCR repertoire of the peripheral blood (Fig.3g).

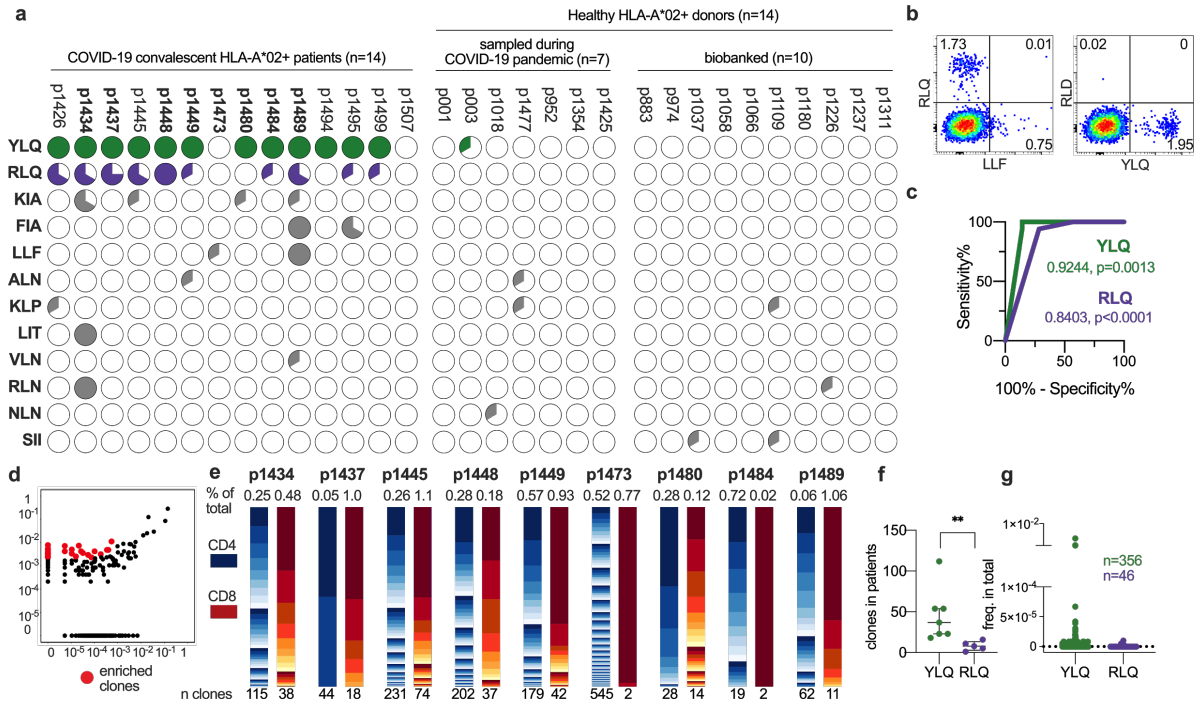


Fig.3 Epitope specificity of CD8+ T-cell response to SARS-CoV-2 S-protein is significantly different between HLA-A*02+ COVID-19 convalescent patients and healthy donors. (**a-b**), PBMC of HLA-A*02:01 positive CP (n=14), HD(Cov) (n=7) and HD(BB) (n=10) were stimulated with a mix of 13 predicted peptides and expanded for 8-12 days followed by MHC-tetramer staining. **a**, Pie charts represent fractions of wells containing tetramer-

positive cells after expansion. **b**, Representative MHC-tetramer staining at day 9 (patient p1445, well 3). **c**, ROC curve of the CP/HD classifier based on the presence of YLQ- or RLQ- epitope-specific cells after expansion. **(d-e)**, The IFNγ-secreting CD4+ and CD8+ cells were FACS-sorted after stimulation with the S-protein and their TCR beta repertoire was sequenced. **d**, Representative enrichment plot is shown (patient p1448, IFNγ-secreting CD8+ vs total PBMC). Red dots represent strongly (>10X) and significantly ($p<10^{-8}$, exact Fisher test) enriched clones, assumed to be antigen-specific. **e**, Clonal structure of the CD4+ (blue) and CD8+ (orange) antigen-specific T-cell population. Numbers below the bars indicate the total number of antigen-specific clones, numbers above - their combined share in the total T-cell repertoire. **f**, MHC-tetramer-positive clones after rapid in vitro expansion were FACS sorted and their TCR repertoires were sequenced. Numbers of YLQ- and RLQ-specific clones in each patient are plotted (p=0.0013 by Kolmogorov-Smirnov unpaired t-test). **g**,YLQ-and RLQ- specific T cell clones occupy only a negligible fraction of the total T cells repertoire. Frequencies of each antigen-specific T cell clone in the PBMC are plotted.

TCRs specific to two SARS-CoV-2 epitopes display prominent sequence motifs shared across individuals

Statistical analysis of V(D)J rearrangement properties of RLQ- and YLQ-specific TCRs revealed biases in the CDR3 length distribution (Fig.4a) and Variable gene usage (Fig.4b) for both TCR alpha (TRA) and beta (TRB) chains. CDR3 regions of these TCRs, with the exception of RLQ-specific TRB, appear to be substantially shorter than those observed in a control PBMC repertoire, a feature previously shown to be associated with “public” TCRs that have higher V(D)J rearrangement probability and incidence rate across individuals (*37*). TRB CDR3 length difference can also explain why YLQ-specific T-cells are more frequent than RLQ-specific ones.

We observed some notable differences in the frequency of certain Variable genes: for example, TRAV12-1 and TRBV7-9 were used by 71% and 16% YLQ-specific TCRs, TRAV13-2 and TRBV6-5 were used by 15% and 25% RLQ-specific TCRs, compared to just 3-4% gene usage in control TCRs. Strong biases observed for gene usage might suggest the importance of germline-encoded features in TCR recognition of RLQ and YLQ epitopes and is reminiscent of previously reported TRAV12 bias for the Yellow fever virus LLW epitope (*38, 39*)

We performed TCR sequence similarity analysis to extract the set of motifs that governs the recognition of RLQ and YLQ epitopes (see Materials and Methods). Our analysis revealed a set of three distinct CDR3 alpha and three distinct CDR3 beta motifs, each containing more than 10 highly similar sequences coming from YLQ-specific T-cells (Fig.4c). Interestingly, all of the motifs were encountered in most of the donors surveyed (Fig.4d), suggesting the public nature of the response and little difference between motifs in terms of publicity. Moreover, some YLQ-specific TCRs were shared between multiple individuals and others exhibited a high degree of global similarity (Extended Data Fig.4). No significant correlation in frequency was observed between pairs of TRA and TRB motifs, suggesting that the pairing between motifs of different chains may be entirely random, in line with recent observations (*40*). Position-weight matrices of the motifs demonstrate a set of highly dissimilar consensuses (Fig.4e) suggesting that while RLQ- and YLQ-specific TCR repertoires are highly public, they are also diverse.

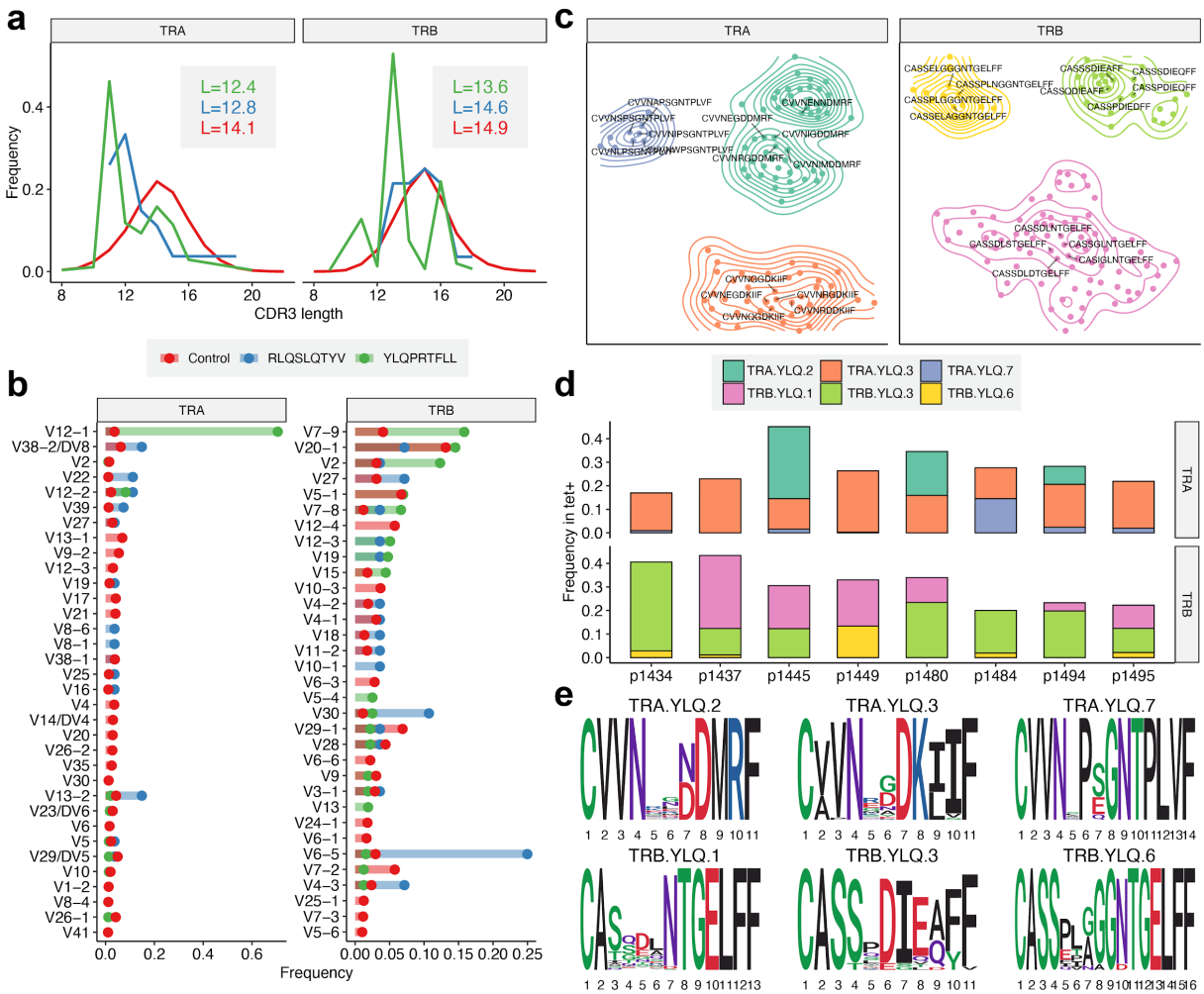


Fig.4 YLQ-specific clones display prominent CDR3 motifs that are shared across individuals. **a,** Length distribution of CDR3 amino acid sequences of tetramer-positive T cell repertoires and control (background) dataset. Line color indicates the dataset: blue-TCRs specific to epitopes RLQ, green-TCRs specific to epitopes YLQ, red- control PBMC TCR repertoire, L-mean CDR3 lengths The insert shows mean CDR3 lengths. **b,** A histogram of Variable gene usage across datasets. Only Variable genes with frequency not less than 1% in any of the datasets are shown. **c,** CDR3 sequence similarity maps of TCR motifs discovered in TCR alpha (TRA) and beta (TRB) chain tetramer-positive T cell repertoires. Sets of highly homologous CDR3 sequences with at least ten members are shown, contour plots show connected components of corresponding sequence similarity graph, labels highlight five - representative CDR3 sequences for each cluster. **d,** Distribution of TCRs corresponding to each CDR3 motif across donor samples. The color of each bar corresponds to specified CDR3 motif. Height of each colored bar corresponds to the total fraction of T-cells having a given CDR3 in each donor. **e,** Position-weight matrices (PWMs) for CDR3 sequences of TRA and TRB chain motifs found in tetramer-positive T-cell repertoires.

TCR repertoire structure analysis of CD4+ and CD8+ T-cell responses to SARS-CoV-2 S-protein reveals public TCR motifs

We analyzed the repertoire sequencing data for bulk T-cell responses to the full-length S-protein of SARS-CoV-2 using TCRNET algorithm (see Materials and Methods) that was successfully employed for similar tasks previously (*41*). This algorithm detects groups of homologous TCR sequences that are unlikely to arise due to convergent V(D)J recombination and thus are a hallmark of an antigen-specific response. We scored each TCR sequence in a pooled dataset of IFNγ-producing CD4+ and CD8+ T cells by quantifying the number of similar TCR sequences observed in the same dataset and in the control PBMC dataset (Fig.5a). This analysis revealed 732 and 517 unique TRB clonotypes that are part of homologous TCR clusters for CD4+ and CD8+ subsets, respectively (Extended Data Table 3). Of them, 433 of CD4+ and 195 of CD8+ TRB clonotypes overlapped with those determined above based on the increased clonotype abundance in stimulated fraction

compared to control. Only two CDR3 amino acid sequences of tetramer-positive T-cells were matched to these clusters when allowing a single amino acid substitution, both coming from YLQ-specific T-cells.

Plotting the CDR3 similarity graph revealed many large (in terms of the number of members) homologous clusters of various CDR3 lengths, with CD4+ subset displaying higher cluster density than CD8+ (Fig.5b). The total number of clusters was higher for CD4+ subset than CD8+ (208 vs 174 clusters), the average CD4+ cluster size was larger than CD8+ in terms of number of unique CDR3 amino acid sequences (5.8 vs 4.0), yet the average frequency of a given cluster in corresponding dataset in terms of number of cells was larger for CD8+ than CD4+ (0.22% vs 0.12%), altogether highlighting higher diversity of CD4+ response in line with previous reports (*42*).

We then mapped the set of detected TCR clusters back to original donor samples (Fig.5c) finding that all donors feature some level of homologous TCR response, yet there are some prominent inter-donor differences: for example, donors p1448 and p1149 displayed many CD4+ clusters, while donor p1480 displayed almost exclusively CD8+ clusters, and donor p1484 displayed few clusters some of which are nevertheless of high frequency in terms of the number of cells. Analysis of cluster sharing across donors (Fig.5d) revealed that most clusters are private to donors that is expected as donors have multiple unmatched HLAs and cells are stimulated with whole protein, however, there were 9 CD4+ and 8 CD8+ clusters shared between stimulated cells of multiple donors (Fig.5e). One of the “public” clusters (C2 in Fig.5e) contains sequences matching those found in YLQ-specific T-cells and was further explored by mapping it to unstimulated PBMC samples of all donors (Fig.5f). Exploring both stimulated IFNγ-producing CD8+ cells and PBMCs revealed a set of TCR sequences in multiple donors matching those of cluster C2, all sharing the same highly conserved CDR3 motif “CASS[YD]SDTGELFF”.

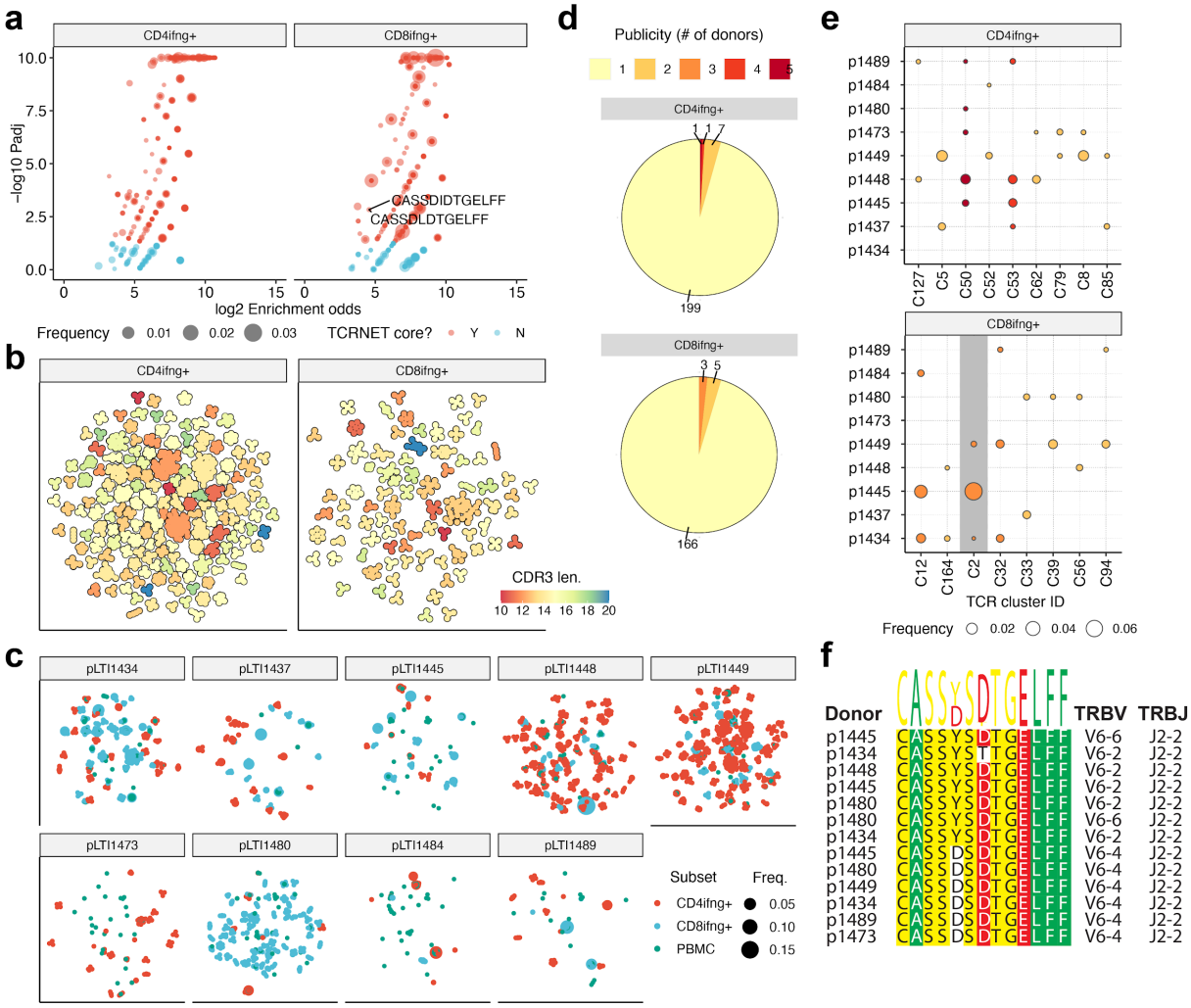


Fig.5 IFNγ-producing T-cells from stimulated T-cell pools feature public TCR motifs. **a**, T-cell neighborhood enrichment (TCRNET) analysis results showing TCRs related to expanded T-cell families. The plot quantifies the number of homologous TCRs expanded upon stimulation compared to control dataset. Each point represents a TCR sequence and point size represents total frequency in a pooled sample. X axis corresponds to the logarithm of the observed to expected number of homologous TCRs (neighbours) for a given sequence, Y axis is the logarithm of adjusted P-value for TCRNET test, capped at 10^{-10}. Red points specify selected TCRs (enrichment odds greater than 4x, Padj < 0.05). Labels indicate two shared TCRs found in tetramer positive T-cell dataset that are YLQ-specific. Only TCRs with Padj < 1 are shown in order to avoid overplotting. **b**, CDR3 similarity map showing the structure of motifs discovered for CD4+ and CD8+ T-cells. Points representing individual CDR3 sequences are placed according to the layout of a TCR similarity graph, built based on CDR3 sequences identified by TCRNET. Black outline shows groups of homologous CDR3 sequences, color indicates the CDR3 region length for a given cluster. Only clusters with at least 3 members are shown. **c**, Distribution of TCRs coming from TCRNET-identified clusters across donors. Each point represents a unique TCRs, point sizes are scaled according to TCR frequency in a given donor and subset, points are colored according to T-cell subset. **d**, Publicity of CDR3 motifs discovered in IFNγ -producing CD8+ and CD4+ T-cell datasets. Labels indicate the number of distinct CDR3 motifs found in a given number of donors. **e**, Distribution and frequency of TCRs associated with public CDR3 motifs in IFNγ -producing CD8+ and CD4+ T cell populations of different donors. Point size and color corresponds to a given cluster frequency in a given donor and publicity respectively (see **d**). Cluster C2 that contains TCRs found in tetramer positive T-cells (see **a**) is highlighted with grey background. Only clusters present in at least two donors are shown. **f**, Multiple sequence alignment of the CDR3 region, Variable and Joining genes of distinct TCR variants corresponding to cluster C2 found in PBMC (unstimulated) samples from different donors.

Table1 Set of HLA-A*02:01-restricted peptides of S-protein used in this study.

N	Start position	End position	Length	AA sequence	HLA restriction	Cleavage probability	Binding Score	Binding Rank	SARS-CoV-1	Link
1	269	277	9	YLQPRTFLL	A*02:01	0.977383	0.973032	0.015	6/9	(*43*)
2	417	425	9	KIADYNYKL	A*02:01	0.966616	0.908998	0.0583	8/9	-
3	424	433	10	KLPDDFTGCV	A*02:01	0.896237	0.591582	0.3676	8/10	(*34*)
4	691	699	9	SIIAYTMSL	A*02:01	0.951629	799,513	0,11	8/9	(*44*)
5	821	829	9	LLFNKVTLA	A*02:01	0.859019	0.785743	0.1599	9/9	(*45*)
6	958	966	9	ALNTLVKQL	A*02:01	0.943025	0.450592	0.6159	9/9	(*34*)
7	976	984	9	VLNDILSRL	A*02:01	0.97061	0.950712	0.0356	9/9	(*44*)
8	983	991	9	RLDKVEAEV	A*02:01	0.969105	0.860941	0.0962	9/9	-
9	996	1004	9	LITGRLQSL	A*02:01	0.891839	0.957880	1.98	9/9	(*24*)
10	1000	1008	9	RLQSLQTYV	A*02:01	0.748401	0.743108	0.199	9/9	(*46*)
11	1185	1193	9	RLNEVAKNL	A*02:01	0.9655	0.618928	0.3295	9/9	(*46*)
12	1192	1200	9	NLNESLIDL	A*02:01	0.945978	0.697243	0.2411	9/9	(*44*)
13	1220	1228	9	FIAGLIAIV	A*02:01	0.179154	0.82072	0.1288	9/9	(*24*)

Discussion

Here we analyzed T cell and humoral immune response to SARS-COV-2 in 31 donors recently recovered from COVID-19 and a control cohort of healthy donors sampled both before the onset of COVID-19 pandemic and during it. Two patients (p1472 and p1473) had no detectable antibody levels in the serum to any of the tested SARS-CoV-2 antigens and no T-cellular response to any of the peptide pools, albeit they had T-cells reactive to the recombinant S-protein. One possible explanation is the delayed seroconversion (blood was sampled at 22 and 36 days since the onset of symptoms or positive PCR test) though it contradicts the recent study reporting that antibodies develop in 100% patients sampled at days 17-19 (*47*). Other possibility is the false positive PCR test. In our study detection of anti-RBD IgG yielded more reliable results than other antigens we tested. Also we observed a moderate correlation of anti-RBD antibody levels. It is in accordance with previously published data (*16, 48*). It was previously demonstrated that titers of anti-SARS-CoV-2 antibodies positively correlate with patient age (*16*). In our cohort we did not observe that (Extended Data Fig.2e-h).

The accumulating data indicate that some healthy COVID-19-naive donors have T cells specific to SARS-CoV-2 antigens and S-protein in particular (*32, 34, 49*). Our findings are in line with that with a caveat that we observed the significant increase of SARS-CoV-2 S-protein reactive T-cells in healthy donors sampled during the COVID-19 pandemic (Fig.2c,d). Combined with the complete lack of SARS-CoV-2 specific antibodies in that group this suggests that some of the donors may have had contact with SARS-CoV-2 before blood sampling. Due to cross-reactivity induced by other coronaviruses their T cells might have protected them from developing the full scale infection. This is illustrated by the case of p1477 who cohabited with COVID-19 patient but was negative in multiple PCR tests, did not have any COVID-19 typical or flu-like symptoms and had no detectable antibodies to any of the SARS-CoV-2 antigens. This hypothesis, although, needs to be validated on a larger cohort of donors.

As others have shown before, we observed that significantly more CD4+ cells in convalescent donors express HLA-DR and CD38 (*32*). We have also shown the same tendency for CD8+ T cells, though the difference was less significant (Fig.2g). As was shown before (*31*) the majority of SARS-

CoV-2 specific CD4+ belonged to the T_{CM} subpopulation while CD8+ were predominantly of T_{TE} and T_{EM} phenotype.

In this work we used the IFNγ-secretion upon antigen stimulation as a criterion for antigen-specific cells. Of course, there are other ways of T-cell reactivity, this is in particular true for CD4+ T cells. Thus, our approach might potentially miss some relevant T cells. But previous studies show that CD4+ T cells react to stimulation by SARS-CoV-2 antigens (S-protein in particular) by Th1-type response (*31*). In another study IFNγ was also the predominant cytokine produced by memory T cells after stimulation with SARS-CoV-2 peptides (*23*).

The results we obtained for T-cells stimulated by the full-length S protein reveal a highly specific response in terms of TCR repertoire structure of both CD8+ and CD4+ T cells. IFNγ-producing T-cell repertoires featured multiple groups of homologous TCR sequences that are in a good agreement with TCR variants found to be enriched based on corresponding clonal abundances. Thus, we were able to identify hundreds of TCR motifs, some of which were shared across multiple donors. Interestingly, only one of the detected motifs was found in tetramer positive CD8+ T-cell fraction, suggesting that immediate response may be targeting a wealth of distinct S-protein epitopes, some of which are presented by HLA alleles other than A*02. This single motif, however, displayed extensive sharing between donors and was also found in an unstimulated PBMC fraction.

In this work we tested 13 epitopes for 11 of which presentation by HLA-A*02 was previously confirmed and immunogenicity demonstrated in SARS-CoV convalescent patients (*50*) and healthy donors (*44*). Despite antigen-specific expansion being performed for all peptides to detect even the small T-cell clones, we observed consistent epitope-specific response to only 2 of 13 peptides (Fig.3a). In particular we did not see a strong response to RLN peptide despite its previously reported immunodominance in SARS-CoV-2 convalescent patients (*46*). KLP-specific T cells were detected only in 1 convalescent patient, contrary to study with SARS-CoV-1 (*34*), which showed this epitope to be highly immunogenic in convalescent patients but not healthy donors. KLP peptide of SARS-CoV-2 differs from SARS-CoV-1 by single amino acid substitution (M-T), whereas RLN peptides are 100% homologs. This lack of the immune response to homologous epitopes in SARS-CoV-2 convalescents may be attributed to the existence of more immunodominant epitopes, that elicit T cell expansion, which crowds out RLN- and KLP-specific T cells. It was previously shown that T-cell immune response in healthy donors is focused predominantly on epitopes in C-terminal part of SARC-CoV-2 S-protein which may be explained by higher homology of this region with common cold coronaviruses (Braun). In our study only one epitope (YLQ) was derived from the N-terminal part of the S-protein, but indeed we saw almost no response to this epitope by healthy donors.

We described two HLA-A*02:01 epitopes of SARS-CoV-2 S-protein YLQ and RLQ with immune response to them detected in 12 of 14 and 10 of 14 of convalescent donors, respectively. T cells specific to both epitopes occupied only a negligible fraction of total TCR repertoire, explaining almost no intersection between tetramer-positive and IFNγ-secreting repertoires. It is possible that YLQ and RLQ-specific clones are localized in the peripheral tissues and only a limited number of cells are present in the circulation. It was previously shown that clonal CD8+ expansions in SARS-CoV-2 are tissue resident (*27*).

YLQ epitope was suggested before as potentially immunogenic in SARS-CoV-2 (*43, 51*). It was also predicted to bind to HLA-A*01:01 and HLA-C*07:02, besides HLA-A*02:01 (*52*). In this study we showed that it is not only highly immunogenic but also is recognized by T cells sharing the same TRA V-segment (TRAV12-1), suggesting the important role of the TCR alpha chain in recognition of this epitope. Of note this epitope an a substantial share of N-terminal part of S-protein was not present in the utilized peptide pool, possibly explaining the discrepancy between reactivity to the recombinant S-protein and peptide pools as seen on Fig.3 of the Extended Data.

Our analysis of T cells specific to RLQ and YLQ epitopes obtained using tetramer-based enrichment revealed a set of highly conserved TCR sequences shared across multiple donors. These sequences feature highly restricted Variable segment usage and relatively short CDR3 length, suggesting that RLQ and YLQ are targeted by public TCRs that rely on germline-encoded motifs to recognize them. The YLQ epitope is recognized by several unrelated motifs that are shared across several donors, suggesting that the response is both public and diverse.

Alongside (*53*) this study provides a first glimpse into the structure of T-cell response to SARS-CoV-2. Further studies on the specificity of the SARS-CoV-2 targeted response and deconvolution of SARS-CoV-2 epitopes would provide crucial information for vaccine design and disease diagnosis.

Acknowledgments

Authors would like to acknowledge their gratitude to all donors who volunteered for our study; to the nurses who performed the venipuncture: Yulia Fadeeva, Anastasya Nisanova, Valentyna Mirponova, and Lubov Piskunova; to our colleagues: Anastasia Minervina, Mikhail Pogorelyy, Vasily Lazarev, Maria Lagarkova, Ivan Zvyagin, Igor Fabrichny, Anastassia Semikhina, Alexey Panov, Ekaterina Avilova, Artem Demidenko, Alexander Veretennikov, Fedor Rozov, Elena Osipova, Nikolay Mugue, Igor Fabrichny, Yuri Lebedin and Ekaterina Morozova for their kind help with experiments and reagents and the most valuable advice. Mikhail Shugay was supported by RFBR grant No 19-34-70011. Alexander Ivanov was supported by the Ministry of Science and Higher Education of the Russian Federation [Agreement No. 075-15-2019-1660].

Material and methods

Donor selection. 31 COVID-19 convalescent patients from Moscow, Russia were recruited voluntarily. COVID-19 was confirmed either by positive SARS-CoV-2 RT-PCR test or retrospectively by the detection of anti-RBD antibodies. All donors signed the informed consent approved by the National Research Center for Hematology ethical committee before the enrollment. According to the classification by the US National Institutes of Health, the severity of disease was defined on patient's case history - asymptomatic (lack of symptoms), mild severity (fever, cough, muscle pain, but without respiratory difficulty or abnormal chest imaging) and moderate/severe (lower respiratory disease at CT scan or clinical assessment, a saturation of oxygen (SaO2) >93% on room air, but lung infiltrates less than 50%). In this study were also included 7 volunteers, sampled during COVID-19 pandemic but without known contact with COVID-19 patients (except p1477, who was cohabiting with a COVID-19 patient). Additionally, 10 healthy hematopoietic stem cell donor samples and 10 healthy donor serum samples were obtained from the blood bank with the approval of the local ethical committee. Cell samples were cryopreserved no later than September 2019, serum samples no later than 2017.

HLA genotyping. For most donors HLA genotyping was performed by next-generation sequencing using ALLType kit (One Lambda, Canoga Park, California). This kit uses a single multiplex polymerase chain reaction to amplify the full HLA-A/B/C gene sequences and from the exon 2 to the 3'UTR of the HLA-DRB1/3/4/5/DQB1 genes. Prepared libraries were run on an Illumina MiSeq sequencer using standard flow cell with 2 x 150 paired-end sequencing. Reads were analyzed using the HLA TypeStream Visual Software (TSV) (One Lambda), version 2.0.0.27232 and the IPD-IMGT/HLA database 3.39.0.0. Other donors were HLA genotyped by Sanger sequencing was performed for loci HLA-A,B,C, DRB1 and DQB1 using Protrans S4 and Protrans S3 reagents respectively. The PCR product for sequencing was prepared by BigDye Terminator v1.1. Capillary electrophoresis was performed on Genetic Analyser Nanophore05. One donor HLA genotyping was determined based on exome sequencing data.

SARS-CoV-2 S-protein peptides. Putative HLA-A*02:01 epitopes of viral S-protein were included in the analysis if they meet the following criteria: 1) weak binders (0.5<rank<2) or strong binders (rank<0.5) according to NetMHCpan 4.0 2) full or partial homologs of existing SARS-CoV S-protein epitopes (identity >60%) (1) Detailed information about selected peptides is listed in Table 1. Predicted proteasomal cleavage score of the C-terminal amino acid was estimated using NetChop 3.1 (*54*). HLA-A02:01 binding affinity score and rank were estimated by NetMHCpan 4.0 (*55*). SARS-CoV identity was measured as the count of identical positions in the alignment of amino acid sequences of SARS-CoV and SARS-CoV-2 (MN908947.3) S-protein performed by QIAGEN CLC Genomic Workbench software. Peptides were synthesized by using a solid-phase synthesis method and purified by high-performance liquid chromatography (HPLC) (greater than 95% purity). All peptides were dissolved in DMSO, except cysteine-containing peptides which were dissolved in MES buffer, pH 6.5 / isopropanol mixture (1:1 vol.).

PBMC isolation. Venous blood from healthy donors and recovered COVID-19 patients was collected into EDTA tubes and subjected to Ficoll (Paneco) density gradient centrifugation (400g, 30 min). Isolated PBMC were washed and used for multiple assays.

Flow cytometry. A surface staining and phenotype analysis of the PBMC were performed with CD3-AF700 (OKT3; Sony), CD4-FITC (RPA-T4; Sony), CD8-PerCP/Cy5.5 (RPA-T8; Sony), CCR7 (CD197)-

PE/Dazzle594 (G043H7; Sony), PD1 (CD279)-BV421 (EH12.2H7; Sony), CD27-BV711 (0323; Sony), CD28-BV785 (CD28.2; Sony) and CD45RO-PE/Cy7 (UCHL1; Sony). Cells were analyzed on the FACS Aria III cell sorter (BD Biosciences). FlowJo Software (version 10.6.1., Tree Star, Ashland, OR) was used for analyzing data.

Tetramer staining. Antigen-specific cells were detected by staining, with CD3-AF700 (OKT3; Sony), CD8-FITC (Sony), 7AAD (Sony) along with 7 combinations of two different peptide-tetramer complexes conjugated with Streptavidin-Allophycocyanin and Streptavidin-R-Phycoerythrin (Thermo Scientific). FACS Canto II cell analyzer and Aria III cell sorter (both BD Biosciences) were used. Data were analyzed using FlowJo Software.

IFNγ secretion assay. Measurement of IFNγ secretion in CD4+ and CD8+ T cells was performed using IFN-γ Secretion Assay-Detection Kit (APC) (Miltenyi Biotec) according to the manufacturer's protocol. Briefly, fresh PBMC isolated from biobanked healthy donors (BB), sampled during COVID-19 pandemic and COVID-19 convalescent patients (CP) were resuspended in RPMI 1640 culture medium (Gibco) supplemented with 5% normal human A/B serum (obtained from pooled inactivated human AB Rh- male serum) and 1mM sodium pyruvate (Gibco) and plated at density 1-10x10^6 cells/ml. Cells were treated for 16h with 10µg/mL of SARS-CoV-2 S-protein, followed by incubation for 5 min at 4°C with IFNγ Catchmatrix Reagent (Miltenyi Biotec). Cells were then transferred into a warm medium (37°C) for 45 min to re-initiate secretion of IFNγ, washed and stained with surface and phenotype markers together with IFNγ Detection Antibody-APC (Miltenyi Biotec) for 10 min at 4°C. CD4+IFNγ+ and CD8+IFNγ+ populations were sorted directly to TRIzol™ Reagent (Thermo Fisher Scientific) using FACS Aria III cell sorter (BD Biosciences). Data were analyzed using FlowJo Software (version 10.6.1., Tree Star, Ashland, OR). For group comparison we used Kruskal–Wallis test and Dunn's multiple comparison test and Mann-Whitney test.

Immunomagnetic isolation of CD4+ and CD8+ T-cells. CD4+ and CD8+ T-cells were isolated using human CD4+ MicroBeads (Miltenyi Biotec) and human CD8+ MicroBeads (Miltenyi Biotec), respectively, according to the manufacturer's protocol. Briefly, the PBMC isolated from COVID-19 convalescent donors (CD), were incubated for 15 min at 4°C with lyophilized CD8+ MicroBeads, washed and loaded onto MS MACS Column (Miltenyi Biotec), which were placed in MACS Separator. After columns were removed and magnetically labeled CD8+ cells were eluted, the flow-through fraction was collected and used for isolation of CD4+ cells by lyophilized CD4+ MicroBeads. Isolated CD8+, CD4+ T-cells, and unlabeled cells (source of APC) were counted and used for IFNγ ELISPOT assay.

IFNγ ELISPOT assay. Measurement of antigen-specific IFNγ production by T cells was performed using ImmunoSpot Human IFN-γ Single-Color ELISPOT kit (CTL, HIFNγ P-2M/5) with 96-well precoated with human IFNγ capture antibody nitrocellulose plate. CD8+ and CD4+ T cells isolated with the use of immunomagnetic beads were plated at density 10^5 cells/well in duplicate. Unlabeled cells obtained after the selection of CD8+ and CD4+ T cells were used as APC at density 2x10^5/well. SARS-CoV-2 S-protein was added at a final concentration of 10µg/mL in Serum-free Testing Medium (CTL) containing 1mM GlutaMAX (Gibco) at a final volume 200µL/well. Total PBMC were seeded at a concentration 5x105/well in duplicates and pulsed with M-, N- or S peptide pools (Miltenyi Biotec) at a final concentration 1µM. Plates were incubated for 18h at 37°C in 9% CO_2. After, plates were washed two times with PBS and then two times with PBS, containing 0.05% Tween-20, followed by incubation with biotinylated anti-human IFN-γ Detection antibody for 2h at room temperature (RT). Wells were washed three times with 0.05% Tween-20/PBS and Streptavidin-AP were added for 30 min at RT. After a few washings, the colorimetric reaction was started by adding substrate components for 15 min at RT. The reaction was stopped by gently rinsing the plate with tap water. Spots numbers were counted by CTL ImmunoSpot® Analyzer using ImmunoSpot® Software. For group comparison the data were log(2)-transformed, the normality was assessed by Shapiro-Wilk test and two-way ANOVA with Tukey's multiple comparisons test was performed.

Expression and purification of recombinant proteins. The recombinant S-protein-His6 of SARS-CoV-2 was encoded by the plasmid kindly provided by prof. Florian Krammer (*7*) was expressed in the Expi293 Expression System (ThermoFisher Scientific) for five days. After harvesting the medium was centrifuged at 10000g, the supernatant was concentrated 10 times and diafiltered into buffer A (10 mM phosphate buffer, 2,7 mM KCl, 500 mM NaCl, pH 8.0) using ÄKTA™ flux tangential flow filtration system (Cytiva, filter cartridge UFP-10-C-4X2MA (cat. # 56-4102-11)). The concentrate was mixed with Ni-NTA agarose resin (Qiagen) previously equilibrated with buffer A and incubated 2h at 22°C with agitation.

The resin mix was packed into a column and washed with 10 volumes of buffer A with 30 mM imidazole and eluted with buffer A with 200 mM imidazole. The eluate was dialyzed against 100 volumes of PBS (10 mM phosphate buffer, 2,7 mM KCl, 137 mM NaCl, pH 7,5) using Slide-A-Lyzer™ Dialysis cassettes (20K MWCO, Thermo Fisher Scientific).
Biotinylated MHC class I/UV-cleavable peptide complexes for UV-mediated ligand exchange were produced as described (*56, 57*). Briefly, heavy (HLA-A*02:01 with biotinylation tag) and light (beta-2 microglobulin) chains were expressed in *Escherichia coli* strain BL21(DE3) pLysS in the form of inclusion bodies. Proteins were dissolved in the denaturation buffer (50 mM Tris-HCl, pH 8.0, and 8 M urea). In vitro folding was set up in folding buffer (100 mM Tris-HCl, 400 mM arginine , 5 mM reduced glutathione, 0.5 mM oxidized glutathione, 2 mM EDTA, protease inhibitors, 1 mM PMSF, pH = 8.0). UV-cleavable peptide (KILGFVFJV for HLA-A*02:01 and AARGJTLAM for HLA-B*07:02, Thermo Scientific custom peptide synthesis), light and heavy chains were mixed in the folding buffer at the 30 : 2 : 3 molar ratio. Correctly folded complexes were purified on Superdex 75 pg column (Cytiva) using tris-buffered saline (20 mM Tris-HCl, 150 mM NaCl, pH 8.0) as mobile phase. Complexes were biotinylated by in-house made biotin ligase and purified on Superdex 75 pg columns.
Concentrations were determined using specific molar absorption coefficient *A*1cm0.1%280 = 1.03, 2.36, and 1.68 for S-protein-His6, HLA-A, and hB2M, respectively (calculated in SnapGene® Viewer based on amino acid sequence).
The recombinant SARS-CoV-2 N-protein was a generous gift from Vasily Lazarev.

ELISA
Clinically approved ELISA kit developed by the National Research Center for hematology was used for the detection of anti-RBD IgG according to the manufacturers' instructions. For detection antibodies to N-protein and full-length S-protein, we used in-house ELISA assay according to the protocol adapted from Kramer et al. (*7*). In brief, 96 well plates (Thermo Scientific, cat. 9502227) were coated with 50 µL per well of 1.4 µg/mL solution of N-protein or 0.4 µg/mL solution of S-protein. The proteins were diluted in a coating buffer (bicarbonate/carbonate 100 mM, pH 9.6). 14h later the plates were washed 3 times with 250 µL of PBS with 0.1% Tween 20 (TPBS) and blocked with 200 ul of 3% non-fat dry milk (Thermo) in PBS for 1.5h. Then the plates were washed thrice and 100 ul of serum samples diluted (1:100) in 1% non-fat dry milk prepared in TPBS were added in duplicates and incubated for 2h. Next, the wells were washed 3 times and were incubated for one more hour with 100 µL of anti-human IgG monoclonal HRP-conjugated antibodies (supplied with RBD ELISA Kit). Finally, the plates were washed 3 times and 100 µL of 3',5,5'-Tetramethylbenzidine (TMB) substrate were added to each well. 10 min after 50 ul of 1 M H3PO4 was added as a stop-solution and optical density (OD) was measured at 450 nm with a reference of 650 nm.
To compare the samples within one plate, as well as normalize the values across the different plates and compare them, we performed the serial dilutions of p1484 serum (from 1:200 to 1:51200) in each plate. A sigmoid four-parameter logistic (4PL) fitting curves model was used to fit the calibration curve based on the serial dilution. For each plate, an EC50 value (the half OD between the top and bottom segment of a curve) was used as a coefficient of normalization. The mean of two OD values for each sample was divided by the coefficient of normalization for the given plate. For group comparison Kruskal–Wallis test and Dunn's multiple comparison test was used.

Antigen-specific T-cell expansions. PBMC of HLA-A*02:01 positive donors were used for rapid in vitro expansion. Briefly, $3x10^6$ cells were incubated for 8-12 days in RPMI 1640 culture medium supplemented with 10% normal human A/B serum, 1mM sodium pyruvate, IL-7 (25ng/mL), IL-15 (40ng/mL), and IL-2 (50ng/mL) at final volume 2ml/well. Half of the medium was replaced on day 3, 5, and 7. A mix of HLA-A*02:01-restricted peptides (see Table 1) of S-protein in DMSO or MES buffer (Sigma-Aldrich) (final concentration of each in medium 10ng/mL) was added at day 0.

TCR repertoire sequencing
TCR libraries were processed as described previously (*58*). cDNA synthesis reaction for alpha and beta chains of T cell receptor was carried out with primer to C-terminal region and SMART-Mk, providing 5' template switch effect and containing sample barcode for contamination control and unique molecular identifier
TCR repertoire sequencing data was analyzed using MIXCR software with default settings. Tetramer-positive TCR sequencing data was formatted and deposited to the VDJdb database (vdjdb.cdr3.net).

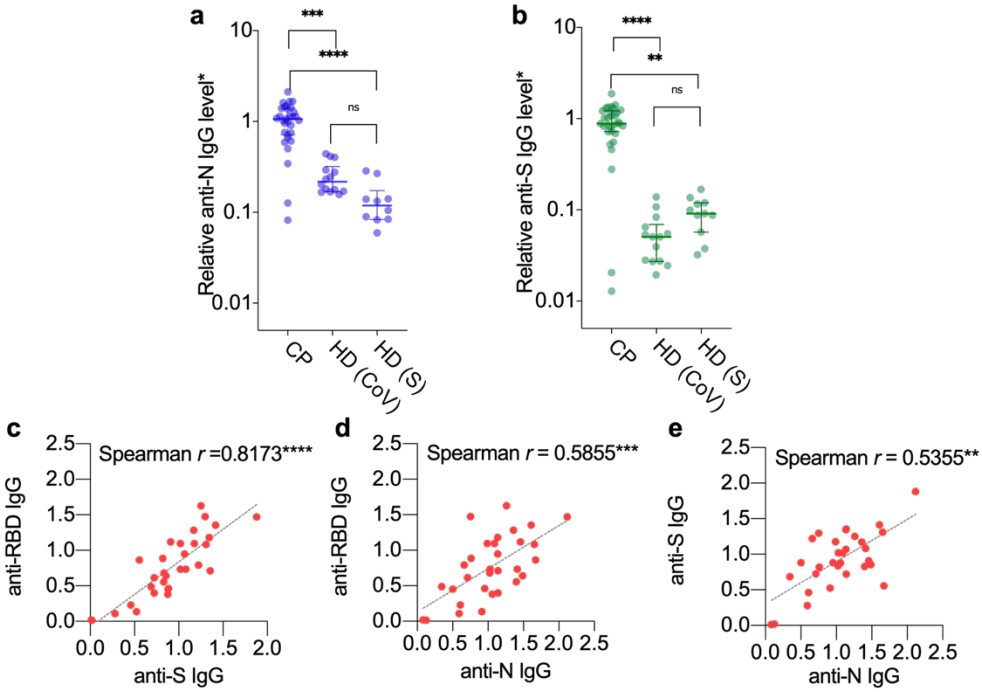
TCR repertoire motif discovery and motif analysis. Motif discovery for TCR repertoires corresponding to tetramer-positive TCRs specific to SARS-CoV-2 epitopes was performed as described

previously (*59*). Briefly, the TCR similarity network was constructed allowing a single amino acid substitution in CDR3 sequence (Hamming distance of 1), the number of similar sequences ("neighbors") for each CDR3 was counted and compared to the expected number of neighbors predicted using a reference dataset containing ~10^7 amino acid sequences of either TCR alpha and beta chains. CDR3 sequences having more neighbors than would be expected at random were considered to be representative ("core") sequences, TCR motifs were defined as connected components of TCR similarity network containing at least one core sequence.

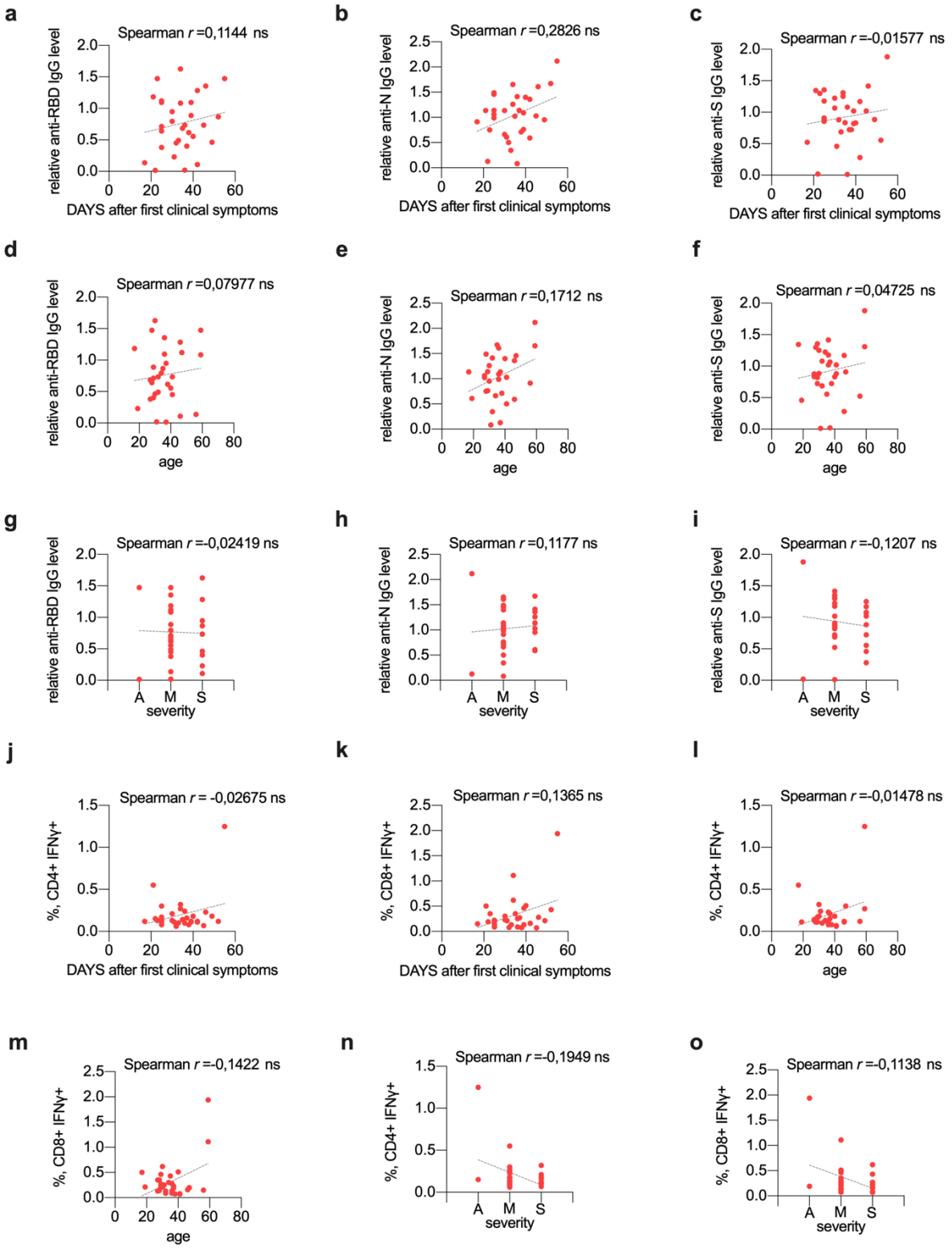
A similar analysis was performed to detect TCR motifs in pooled IFNγ+ fractions of stimulated CD4+ and CD8+ T cells, except for using a control made from pooled PBMC repertoire of corresponding donors and counting neighbors based on CDR3 nucleotide sequences as described in (*37*).

R markdown notebooks used for data analysis are available at https://github.com/antigenomics/covid19-tcr-analysis.

Supplementary information

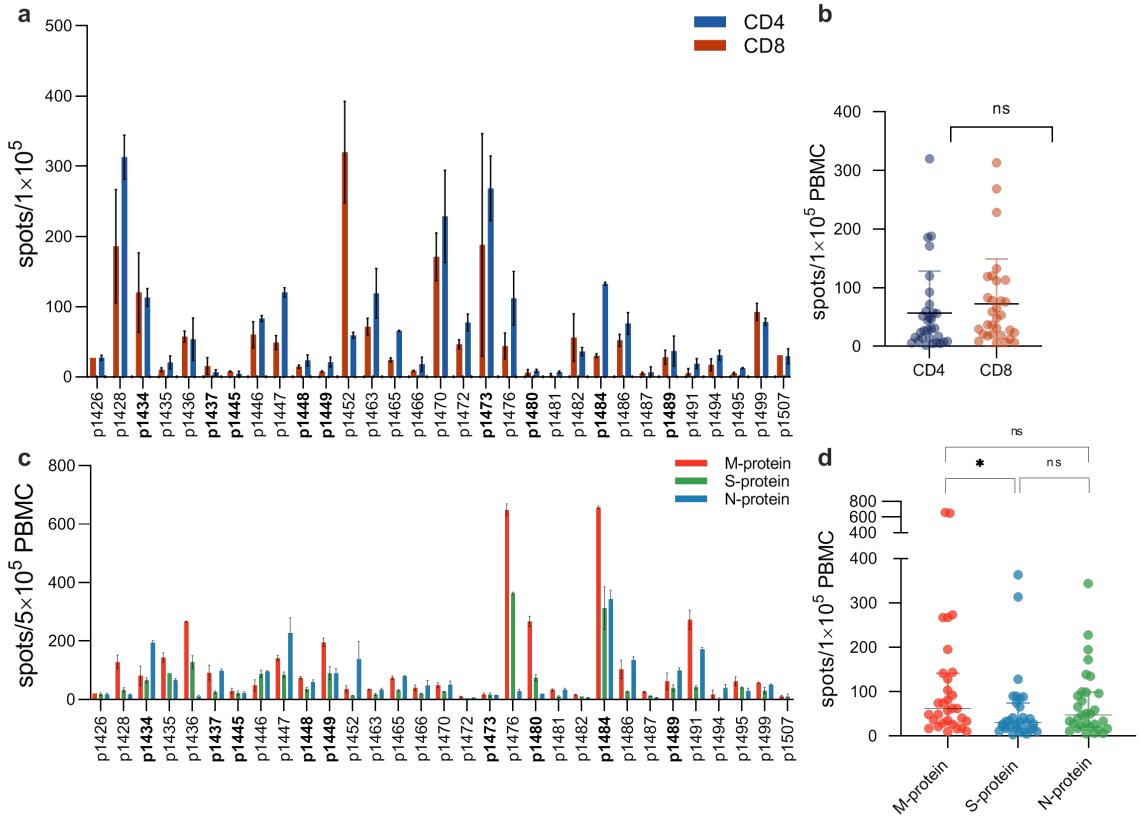


Extended Data Fig.1 Antibody response to the different SARS-CoV-2 antigens. For detection antibodies to N-protein (**a**) and full-length S-protein (**b**) in-house ELISA assay was used. OD_{650} was subtracted from OD_{450} for each well. Mean OD for each serum sample was divided to normalising coefficient (EC50 of the calibration curve) in order to compare the samples across different plates.
The Spearman correlation and linear regression between anti-RBD and anti-S IgG (**c**), anti-RBD and anti-N IgG (**d**), as well as anti-S and anti-N IgG (**e**) is shown. For group comparison Kruskal–Wallis test and Dunn's multiple comparison test was used. RBD - receptor binding domains CP - convalescent patients HD (CoV) - healthy donors sampled during COVID-19 pandemic, and HD (S) - biobanked healthy donors of serum

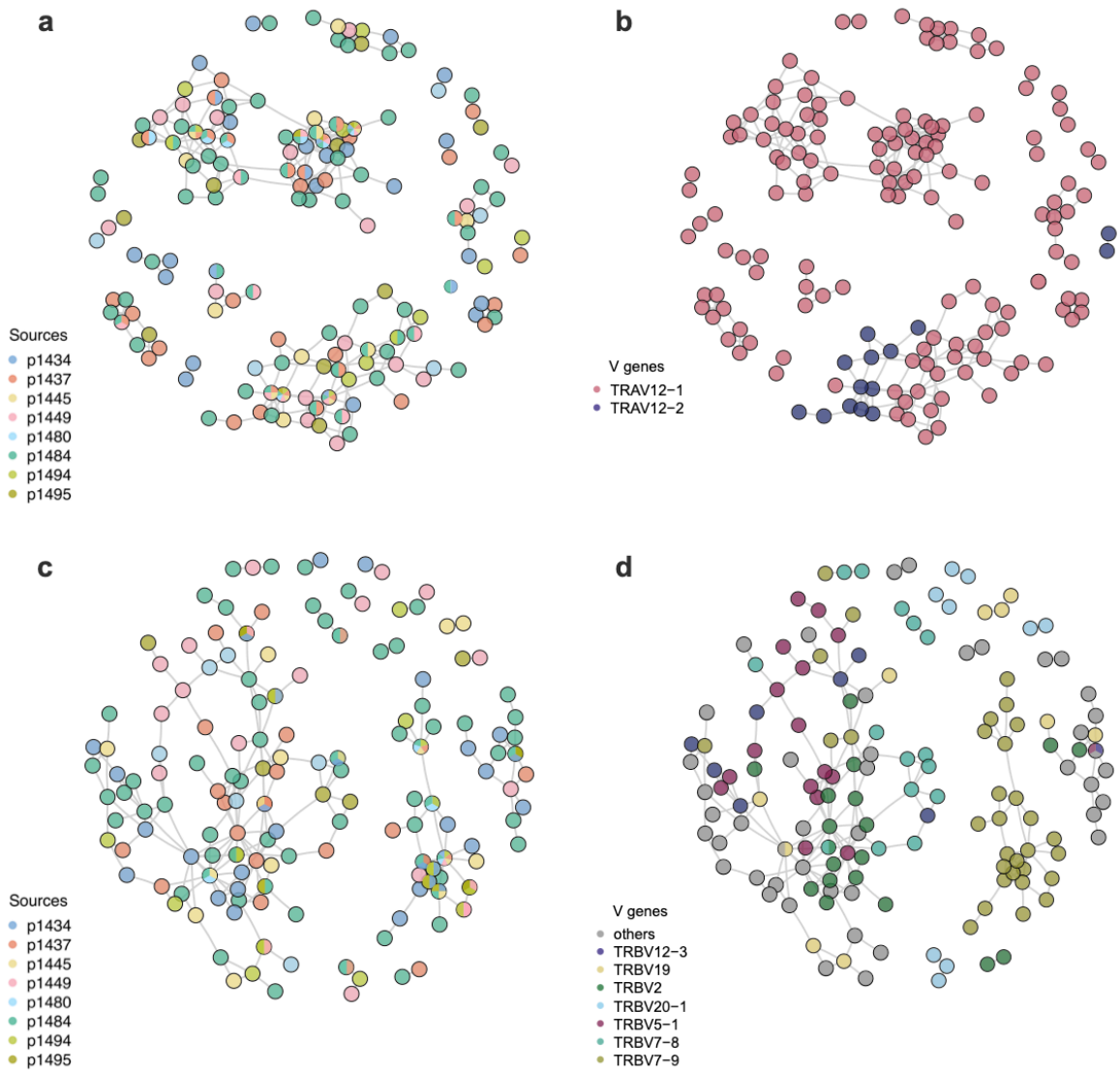


Extended Data Fig.2 Correlation between clinical data and cellular and antibody response in convalescent donors. **(a-c)** Correlation between antibody response and time after the onset of the disease or positive PCR-test. Spearman's coefficient of correlation between time and relative levels of anti-RBD **(a)**, anti-N **(b)**, anti-S **(c)** IgGs and linear regression was plotted (n=31). **(d-f)** Correlation between antibody response and age of convalescent patient. Spearman's

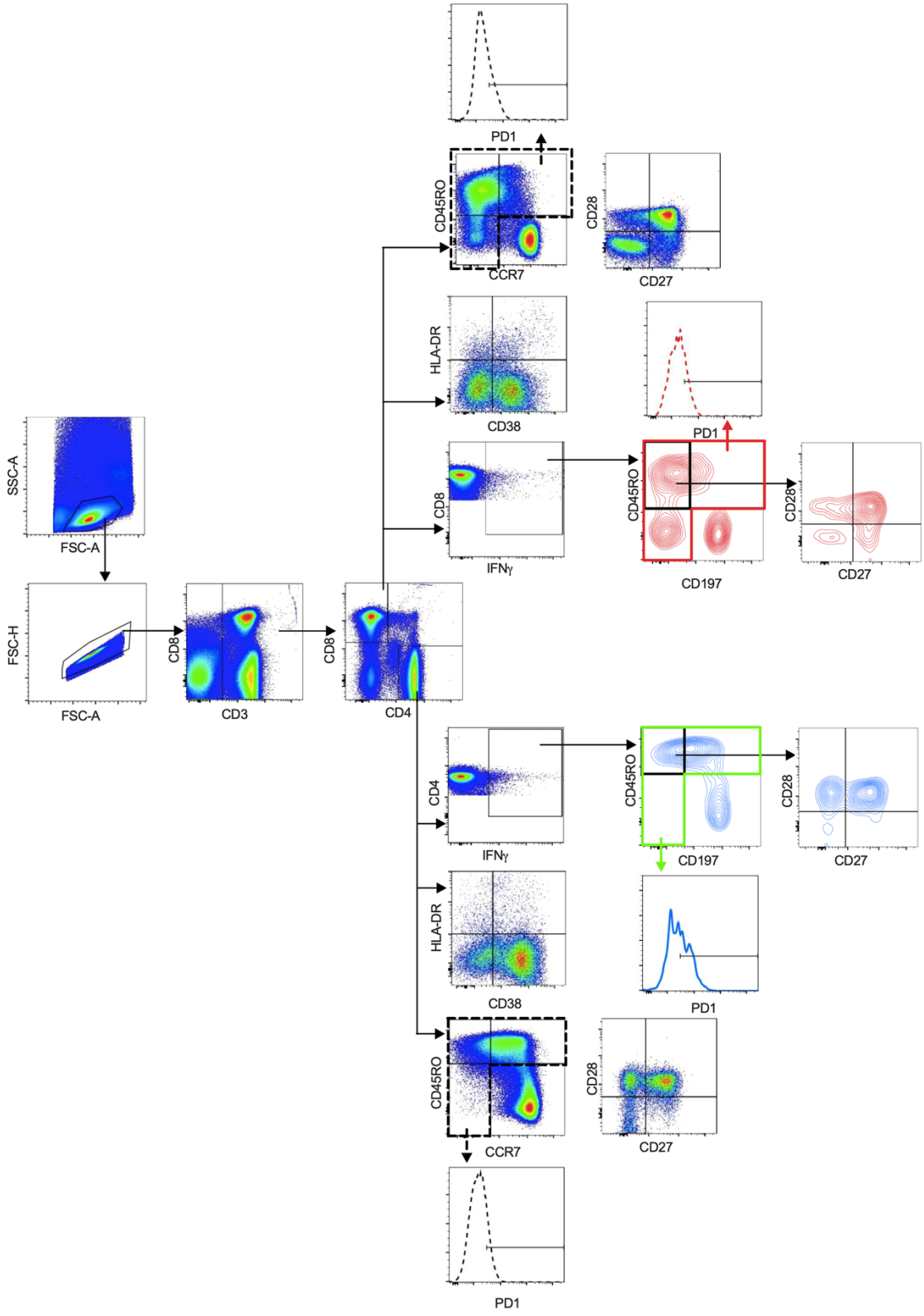
coefficient of correlation between age and relative levels of anti-RBD **(d)**, anti-N **(e)**, anti-S **(f)** IgGs and linear regression was plotted (n=31). **(g-i)** Correlation between antibody response and severity of disease of convalescent patient. Spearman's coefficient of correlation between asymptomatic (A), mild (M) and moderate/severe (S) groups of CP and relative levels of anti-RBD **(g)** anti-N **(h)**, anti-S **(i)** IgGs, linear regression was plotted (n=31). **(j-k)** Correlation between T cell response and time after the onset of the disease or positive PCR-test. Spearman's coefficient of correlation between time and frequencies of IFNγ producing CD4 **(j)** or CD8 **(k)** and time, linear regression was plotted (n=31). **(l-m)** Correlation between T cell response and age. Spearman's coefficient of correlation between age and frequencies of IFNγ producing CD4 **(l)** or CD8 **(m)** and time, linear regression was plotted (n=31). **(n-o)** Correlation between antibody response and severity of disease of convalescent patients. Spearman's coefficient of correlation between asymptomatic (A), mild (M) and moderate/severe (S) groups of CP and frequencies of IFNγ producing CD4 **(n)** or CD8 **(o)** and time, linear regression was plotted (n=31). Clinical significance p-value *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.



Extended Data Fig.3 Variability of T-cell response to SARS-CoV-2 antigens. Magnetically separated CD8+ and CD4+ cells and total PBMC, isolated from COVID-19 convalescent donors (n=31) were stimulated with recombinant glycoprotein S or with pools of peptides (M, N and S) for 18 hours, respectively. INFγ response was assessed by ELISPOT. (**a-b**) Number of antigen-specific CD8+ and CD4+ T-cells. (**c-d**) Number of antigen-specific T-cells. Spots were quantified by automated digital image analysis in duplicate wells. Donors used for TCR repertoire analysis are shown in bold. For group comparison the data were log(2)-transformed, the normality was assessed by Shapiro-Wilk test and two-way ANOVA with Tukey's multiple comparisons test was performed.*p<0.05



Extended Data Fig.4 Global similarity of YLQ-specific CDR3 amino acid sequences. Graph shows CDR3 amino acid sequences of MHC-tetramer-positive clones with Hamming distance between sequences 1 or 0. (**a-b**) TCRα CDR3 amino acid sequences. (**c-d**) TCRβ CDR3 amino acid sequences. (**a, c**) Colors correspond to different CP. (**b, d**) Color corresponds to different V-genes. Each dot represents 1 CDR3 amino acid sequence.



Extended data Fig.5 Flow cytometry gating strategy

Extended Data Table 1 HLA genotyping (Donors for which TCR sequencing was performed are shown in bold; * - donors HLA-typed by flow cytometry)

	HLA I class			HLA II class	
Patient ID	HLA-A*	HLA-B*	HLA-C*	HLA-DR*	HLA-DQB1*
			COVID-19 convalescent donors		
p1426	02:01:01:01; 24:02:01:01	38:01:01:01; 40:01:02	03:04:01:01; 12:03:01:01	DRB1*14:01P/14:216;16:01:01 DRB3*02:02:01:06 DRB5*02:02:01	05:02P/05:241; 05:03:01
p1428	01:01:01:01; 24:02:01:01	08:01:01:01; 38:01:01:01	07:01:01; 12:03:01:01	DRB1*03:01:01:01;14:01P/14:216 DRB3*01:01:02:05;02:02:01:06	02:01:01:01/ 02:163N;05:03:01
p1434	02:01:01:01; 24:02:01:02L	15:01:01:01; 39:01:01:05	03:04:01:01; 12:03:01:01	DRB1*03:01:01:03;04:01:01:01 DRB3*02:02:01:10 DRB4*01:03:01	02:01:01:01/ 02:163N;03:02P
p1435	24:02:01:01; 30:01:01:01	13:02:01:01; 35:02:01:02	04:01:01:06; 06:02:01:01	DRB1*07:01:01:01;11:04:01 DRB3*02:02:01:02 DRB4*01:03:01:04	02:01:01G/ 02:02:06:02/ 02:163N;03:01
p1436	11:01:01:01; 23:01:01:01	44:03:01:19; 55:01:01:01	03:03:01:01; 04:01:01	DRB1*07:01:01:01;12:01P DRB3*02:02:01:09 DRB4*01:03:01:02N	03:01; 03:03:02:01
p1437	01:01:01:01; 02:01:01:01	35:01:01:05; 35:02:01:02	04:01:01:06; 06:02:01:01	DRB1*11:01:01:01;11:04:01 DRB3*02:02:01:04;02:02:01:04	03:01; 03:01P
p1445	02:01:01:01; 03:01:01:01	44:02:01:01; 44:27:01:01	05:01:01:02; 07:04:01	DRB1*12:01;/16:01:01 DRB3*02:02:01:09 DRB5*02:02:01	03:01; 05:02P/05:241
p1446	01:01:01:01; 23:01:01:01	08:01:01:01; 44:03:01:19	04:01:01; 07:01:01	DRB1*03:01:01:01;07:01:01:01 DRB3*01:01:02 DRB4*01:01:01:01	02:01:01:01/ 02:163N; 02:01:01G/ 02:02:06:02/ 02:163N
p1447	31:01:02:01; 31:01:02:01	40:01:02; 55:02:01:03	01:02:01:01; 03:04:01:01	DRB1*04:04:01;15:01:01:03 DRB4*01:03:01 DRB5*01:01:01	03:02P; 06:02:01
p1448	02:01:01:01; 26:01:01:01	27:05:02:01; 39:01:01:05	07:02:01:03; 12:03:01:01	DRB1*11:04:01;12:01P DRB3*02:02:01;02:02:01:09	03:01P; 03:01
p1449	02:01:01:01; 02:01:01:01	44:02:01:01; 50:01:01:02	04:01:01; 05:01:01:02	DRB1* *04:05:01;12:01P DRB3*02:02:01:09 DRB4*01:03:01	03:01; 03:02P
p1452	03:01:01:01; 29:01:01:01	35:01:01:05; 44:02:01:01	04:01:01; 16:04:01:01	DRB1*11:04:01;15:01:01:03 DRB3*02:02:01:04 DRB5*01:01:01:01	03:01P; 06:02:01
p1463	01:01:01:01; 32:01:01:01	27:05:02:05; 37:01:01:01	01:02:01:01; 06:02:01:01	DRB1*08:01:01;10:01:01:03	04:02:01:04; 05:01
p1465	25:01:01:01; 32:01:01:01	27:07:01; 39:01:01:05	12:03:01:01; 15:02:01:01	DRB1*01:01:01;11:04:01 DRB3*02:02:01;02:02:01	03:01P; 05:01
p1466	11:01:01:01; 32:01:01:01	08:01:01:02; 52:01:01:02	07:02:01:01; 12:02:02:01	DRB1*03:01:01:03;15:02P DRB3*02:02:01:10 DRB5*01:02	02:01:01:01/ 02:163N;06:01

p1470	03:01:01:01; 23:01:01:01	35:01:01:05; 44:03:01:19	04:01:01; 04:09N	DRB1*01:01:01;07:01:01:01 DRB4*01:01:01:01;01:01:01:01	02:01:01G/ 02:02:06:02/ 02:163N;05:01
p1472	03:01:01:01; 25:01:01:01	13:02:01:01; 18:01:01	06:02:01:01; 12:03:01:01	DRB1*07:01:01:01;15:01:01:03 DRB4*01:03:01:04 DRB5*01:01:01	02:01:01G/ 02:02:06:02/ 02:163N; 06:02:01
p1473	02:01:01:01; 30:01:01:01	13:02:01:01; 44:02:01:01	05:01:01:02; 06:02:01:01	DRB1*07:01:01:01;08:01:01 DRB4*01:03:01:04;01:03:01:04	02:01:01G/ 02:02:06:02/ 02:163N; 04:02:01:04
p1476	03:02:01; 23:01:01:01	08:01:01; 44:03:01:19	04:01:01; 07:02:01:01	DRB1*03:01:01:03;16:01:01 DRB3*02:02:01:10 DRB5*02:02:01	02:01:01:01/ 02:163N; 05:02P/05:241
p1480	02:01:01:01; 03:01:01:03	40:01:02; 58:01:01:03	03:02:02:05; 03:04:01:01	DRB1*01:01:01;13:02:01:03 DRB3*03:01:01:03;03:01:01:03	05:01; 06:09:01:01
p1481	01:01:01:01; 32:01:01:01	35:01:01:05; 44:02:01:01	04:01:01/ 04:01:01Q; 05:01:01:02	DRB1*04:01:01:01;11:01:01 DRB3*02:02:01:02 DRB4*01:03:01	03:01:01G; 03:01P
p1482	01:01:01:01; 68:12:01	15:01:01:01; 18:03:01:01	06:02:01:01; 07:01:01	DRB1*08:01:01;11:04:01 DRB3*02:02:01:04;02:02:01:04	03:01P; 04:02:01:04
p1484	02:01:01G; 31:01:02G	27:05:02G; 39:01:01G	01:02:01G; 12:03:01G	DRB1*01:01:01G;11:01:01G	03:01:01G; 05:01:01G
p1486	01:01:01:01; 03:01:01:01	35:01:01:05; 52:01:01:02	04:01:01; 12:02:02:01	DRB1*01:01:01;11:04:01 DRB3*02:02:01:02;02:02:01:02	03:01; 05:01
p1487	26:01:01:01; 68:02:01:01	14:02:01:01; 38:01:01:01	08:02:01:01; 12:03:01:01	DRB1*04:02:01;11:04:01 DRB3*02:02:01:02 DRB4*01:03:01:10	03:01:01/03:01:41; 03:02P
p1489	02:01:01:01; 02:01:01:01	07:02:01:01; 27:02:01:04	02:02:02:01; 07:02:01:03	DRB1*11:01:01:01;13:01:01:02 DRB3*02:02:01:02;02:02:01:02	03:01P; 06:03:01
p1491	03:01:01:01; 24:02:01:01	35:01:01:05; 35:03:01	04:01:01; 04:01:01	DRB1*01:01:01;12:01P DRB3*02:02:01:09;02:02:01:09	03:01; 05:01
p1494	02:01:01G; 25:01:01G	15:01:01G; 51:01:01G	04:01:01G; 14:02:01G	DRB1*04:01:01;08:01:01G	03:02:01G; 04:02:01G
p1495	02:01:01G; 23:01:01G	27:05:02G; 44:03:01G	02:02:02G; 04:01:01G	DRB1*07:01:01G;16:01:01	02:02:01G; 05:02:01G
p1499	02:01:01G; 03:01:01G	07:17; 51:01:01G	02:02:02G; 07:02:01G	DRB1*01:01:01G;15:01:01G	05:01:01G; 06:02:01G
p1507	02:01:01G; 03:01:01G	13:02:01G; 15:01:01G	03:03:01G; 06:02:01G	DRB1*07:01:01G;13:01:01G	02:02:01G; 06:03:01G
			Healthy donors sampled during COVID-19 pandemic		
p001	02:01:01:01; 24:02:01:01	13:02:01; 39:01:01:03	06:02:01:01; 07:02:01:01	DRB1*07:01:01;04:04:01	02:02:01; 03:02:01
p003	02; 02	27; 35	01; 04	DRB1*08;15	04; 06
p846	01:01; 01:01	08:01; 08:01	07:01; 07:01	DRB1*03:01;03:01	02:01; 02:01

p856	11:01; 11:01	35:03; 39:01	12:03; 12:03	DRB1*04:08:01;16:01:01	03:04:01; 05:02:01
p952	02; 03	49; 51	04; 07	DRB1*04:07;13:02	03:01; 06:04
p1018	02:01; 01:01	35:03; 27:05	01:02; 04:01	N/A	N/A
p1048	01:01; 03:01	41:02; 52:01	07:01; 17:01	DRB1*13:03;15:02	03:01; 06:01
p1305	24:02; 33:01	14:02; 40:02	03:04; 08:02	DRB1*01:01;01:02	05:01; 05:01
p1329	03:01; 03:01	07:02; 13:02	06:02; 07:02	DRB1*07:01;13:01	02:02; 06:03
p1354	2*	N/A	N/A	N/A	N/A
p1425	2*	N/A	N/A	N/A	N/A
p1440	N/A	N/A	N/A	N/A	N/A
p1477	02:01:01:01 02:17:02:01	51:01:01:07 15:01:01:01	03:03:01:01 15:02:01:01	DRB1*04:01:01:01; 13:01:01 DRB3*01:01:02:04 DRB4*01:03:01	03:02P 06:03:01
			Biobanked healthy donors		
p847	03; 03	07; 07	07:02; 07:02	DRB1*07;16	03; 05
p879	11; 24	35; 51	04:01; 15:02	DRB1*03;11	02; 03
p883	01; 02	08; 38	07; 12	DRB1*03:01;15:01	02; 06
p884	03; 03	35; 40	02; 04	DRB1*01;11	03; 05
p924	02; 68	38; 40(61)	02; 12:03	DRB1*11 DRB2*15	03; 06
p1180	01; 02	27; 56:01	01; 02	DRB1*04:01;11:01	03; 03
p1235	01:01:01G; 02:01:01G	44:02:01G; 57:01:01G	05:01:01G; 06:02:01G	DRB1*04:01:01G;07:01:01G	03:01:01G; 03:03:02G
p1243	03; 26	03; 35	12; N/A	DRB1*04;11	03; 01
p1309	03:01; 33:01	51; 58	03:02; 14	DRB1*03:01;11:04	02:01; 03:01
p1311	01:01; 02	57; 44	06; 05	DRB1*11:03;07	03; 03

Extended Data Table 2 Enriched IFNγ+ clones

Extended Data Table 3 Homologous TCR clusters

References

1. F. Wu *et al.*, A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269 (2020).
2. N. Zhu *et al.*, A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* **382**, 727-733 (2020).
3. A. L. Phelan, R. Katz, L. O. Gostin, The Novel Coronavirus Originating in Wuhan, China: Challenges for Global Health Governance. *JAMA*, (2020).
4. P. J. Lillie *et al.*, Novel coronavirus disease (Covid-19): The first two patients in the UK with person to person transmission. *J Infect* **80**, 578-606 (2020).
5. P. Zhou *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
6. A. T. Huang *et al.*, A systematic review of antibody mediated immunity to coronaviruses: antibody kinetics, correlates of protection, and association of antibody responses with severity of disease. *medRxiv*, 2020.2004.2014.20065771 (2020).
7. F. Amanat *et al.*, A serological assay to detect SARS-CoV-2 seroconversion in humans. *Nat Med*, (2020).
8. L. Guo *et al.*, Profiling Early Humoral Response to Diagnose Novel Coronavirus Disease (COVID-19). *Clin Infect Dis*, (2020).
9. F. Krammer, V. Simon, Serology assays to manage COVID-19. *Science*, (2020).
10. B. Shanmugaraj, K. Siriwattananon, K. Wangkanont, W. Phoolcharoen, Perspectives on monoclonal antibody therapy as potential therapeutic intervention for Coronavirus disease-19 (COVID-19). *Asian Pac J Allergy Immunol* **38**, 10-18 (2020).
11. C. Wang *et al.*, A human monoclonal antibody blocking SARS-CoV-2 infection. *Nat Commun* **11**, 2251 (2020).
12. D. Pinto *et al.*, Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature*, (2020).
13. V. W. Wong, D. Dai, A. K. Wu, J. J. Sung, Treatment of severe acute respiratory syndrome with convalescent plasma. *Hong Kong Med J* **9**, 199-201 (2003).
14. L. Chen, J. Xiong, L. Bao, Y. Shi, Convalescent plasma as a potential therapy for COVID-19. *Lancet Infect Dis* **20**, 398-400 (2020).
15. T. F. Rogers *et al.*, Rapid isolation of potent SARS-CoV-2 neutralizing antibodies and protection in a small animal model. *bioRxiv*, 2020.2005.2011.088674 (2020).
16. F. Wu *et al.*, Neutralizing antibody responses to SARS-CoV-2 in a COVID-19 recovered patient cohort and their implications. *medRxiv*, 2020.2003.2030.20047365 (2020).
17. C. K. Li *et al.*, T cell responses to whole SARS coronavirus in humans. *J Immunol* **181**, 5490-5500 (2008).
18. R. Channappanavar, C. Fett, J. Zhao, D. K. Meyerholz, S. Perlman, Virus-specific memory CD8 T cells provide substantial protection from lethal severe acute respiratory syndrome coronavirus infection. *J Virol* **88**, 11034-11044 (2014).
19. J. Zhao, S. Perlman, T cell responses are required for protection from clinical disease and for virus clearance in severe acute respiratory syndrome coronavirus-infected mice. *J Virol* **84**, 9318-9325 (2010).
20. J. Chen *et al.*, Cellular immune responses to severe acute respiratory syndrome coronavirus (SARS-CoV) infection in senescent BALB/c mice: CD4+ T cells are important in control of SARS-CoV infection. *J Virol* **84**, 1289-1301 (2010).
21. M. J. Cameron, J. F. Bermejo-Martin, A. Danesh, M. P. Muller, D. J. Kelvin, Human immunopathogenesis of severe acute respiratory syndrome (SARS). *Virus Res* **133**, 13-19 (2008).
22. T. Li *et al.*, Significant Changes of Peripheral T Lymphocyte Subsets in Patients with Severe Acute Respiratory Syndrome. *The Journal of Infectious Diseases* **189**, 648-651 (2004).

23. Y.-Y. Fan *et al.*, Characterization of SARS-CoV-specific memory T cells from recovered individuals 4 years after infection. *Archives of Virology* **154**, 1093-1099 (2009).
24. Y. D. Wang *et al.*, T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus spike protein elicit a specific T-cell immune response in patients who recover from SARS. *J Virol* **78**, 5612-5618 (2004).
25. L. Yang *et al.*, Persistent memory CD4+ and CD8+ T-cell responses in recovered severe acute respiratory syndrome (SARS) patients to SARS coronavirus M antigen. *The Journal of general virology* **88**, 2740-2748 (2007).
26. L.-T. Yang *et al.*, Long-lived effector/central memory T-cell responses to severe acute respiratory syndrome coronavirus (SARS-CoV) S antigen in recovered SARS patients. *Clinical immunology (Orlando, Fla.)* **120**, 171-178 (2006).
27. M. Liao *et al.*, The landscape of lung bronchoalveolar immune cells in COVID-19 revealed by single-cell RNA sequencing. *medRxiv*, 2020.2002.2023.20026690 (2020).
28. I. Thevarajan *et al.*, Breadth of concomitant immune responses prior to patient recovery: a case report of non-severe COVID-19. *Nature Medicine* **26**, 453-455 (2020).
29. H.-Y. Zheng *et al.*, Elevated exhaustion levels and reduced functional diversity of T cells in peripheral blood may predict severe progression in COVID-19 patients. *Cellular & Molecular Immunology* **17**, 541-543 (2020).
30. L. Bao *et al.*, Reinfection could not occur in SARS-CoV-2 infected rhesus macaques. *bioRxiv*, 2020.2003.2013.990226 (2020).
31. D. Weiskopf *et al.*, Phenotype of SARS-CoV-2-specific T-cells in COVID-19 patients with acute respiratory distress syndrome. *medRxiv*, 2020.2004.2011.20062349 (2020).
32. J. Braun *et al.*, Presence of SARS-CoV-2 reactive T cells in COVID-19 patients and healthy donors. *medRxiv*, 2020.2004.2017.20061440 (2020).
33. A. Nguyen *et al.*, Human leukocyte antigen susceptibility map for SARS-CoV-2. *Journal of Virology*, JVI.00510-00520 (2020).
34. A. Grifoni *et al.*, A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* **27**, 671-680.e672 (2020).
35. https://covid19treatmentguidelines.nih.gov/overview/management-of-covid-19/.
36. L. Danilova *et al.*, The Mutation-Associated Neoantigen Functional Expansion of Specific T Cells (MANAFEST) Assay: A Sensitive Platform for Monitoring Antitumor Immunity. *Cancer Immunol Res* **6**, 888-899 (2018).
37. M. V. Pogorelyy *et al.*, Exploring the pre-immune landscape of antigen-specific T cells. *Genome Medicine* **10**, 68 (2018).
38. A. Bovay *et al.*, T cell receptor alpha variable 12-2 bias in the immunodominant response to Yellow fever virus. *European Journal of Immunology* **48**, 258-272 (2018).
39. A. A. Minervina *et al.*, Primary and secondary anti-viral response captured by the dynamics and phenotype of individual T cell clones. *Elife* **9**, (2020).
40. D. S. Shcherbinin, V. A. Belousov, M. Shugay, Comprehensive analysis of structural and sequencing data reveals almost unconstrained chain pairing in TCRαβ complex. *PLoS Comput Biol* **16**, e1007714 (2020).
41. P.-G. Ritvo *et al.*, High-resolution repertoire analysis reveals a major bystander activation of Tfh and Tfr cells. *Proceedings of the National Academy of Sciences* **115**, 9604-9609 (2018).
42. Q. Qi *et al.*, Diversity and clonal selection in the human T-cell repertoire. *Proceedings of the National Academy of Sciences* **111**, 13139-13144 (2014).
43. V. Baruah, S. Bose, Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. *Journal of Medical Virology* **92**, 495-500 (2020).

44. Y. Lv, Z. Ruan, L. Wang, B. Ni, Y. Wu, Identification of a novel conserved HLA-A*0201-restricted epitope from the spike protein of SARS-CoV. *BMC Immunology* **10**, 61 (2009).
45. J. Ishizuka *et al.*, Quantitating T Cell Cross-Reactivity for Unrelated Peptide Antigens. *The Journal of Immunology* **183**, 4337 (2009).
46. B. Wang *et al.*, Identification of an HLA-A*0201–restricted CD8+ T-cell epitope SSp-1 of SARS-CoV spike protein. *Blood* **104**, 200-206 (2004).
47. Q.-X. Long *et al.*, Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nature Medicine*, (2020).
48. L. Ni *et al.*, Detection of SARS-CoV-2-Specific Humoral and Cellular Immunity in COVID-19 Convalescent Individuals. *Immunity*, (2020).
49. C. Dong *et al.*, Characterization of anti-viral immunity in recovered individuals infected by SARS-CoV-2. *medRxiv*, 2020.2003.2017.20036640 (2020).
50. M. Zhou *et al.*, Screening and Identification of Severe Acute Respiratory Syndrome-Associated Coronavirus-Specific CTL Epitopes. *The Journal of Immunology* **177**, 2138-2145 (2006).
51. R.-L. Jp *et al.* (Research Square, 2020).
52. C. a. K. Lee, H, In silico identification of vaccine targets for 2019-nCoV [version 2; peer review: 3 approved]. *F1000Research* **9**, (2020).
53. A. A. Minervina *et al.*, Longitudinal high-throughput TCR repertoire profiling reveals the dynamics of T cell memory formation after mild COVID-19 infection. *bioRxiv*, 2020.2005.2018.100545 (2020).
54. M. Nielsen, C. Lundegaard, O. Lund, C. Keşmir, The role of the proteasome in generating cytotoxic T-cell epitopes: insights obtained from improved predictions of proteasomal cleavage. *Immunogenetics* **57**, 33-41 (2005).
55. B. Reynisson, B. Alvarez, S. Paul, B. Peters, M. Nielsen, NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res*, (2020).
56. B. Rodenko *et al.*, Generation of peptide–MHC class I complexes through UV-mediated ligand exchange. *Nature Protocols* **1**, 1120-1132 (2006).
57. A. H. Bakker *et al.*, Conditional MHC class I ligands and peptide exchange technology for the human MHC gene products HLA-A1, -A3, -A11, and -B7. *Proceedings of the National Academy of Sciences* **105**, 3825-3830 (2008).
58. I. V. Zvyagin *et al.*, Tracking T-cell immune reconstitution after TCRαβ/CD19-depleted hematopoietic cells transplantation in children. *Leukemia* **31**, 1145-1153 (2017).
59. D. V. Bagaev *et al.*, VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res* **48**, D1057-D1062 (2020).