Check for updates

SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition

Annika Nelde [1,2,3,21], Tatjana Bilich [1,2,3,21], Jonas S. Heitmann[1,3,21], Yacine Maringer[1,2,3], Helmut R. Salih[1,3,4], Malte Roerden [2,3,5], Maren Lübke[2], Jens Bauer [1,2], Jonas Rieth[1,2], Marcel Wacker[1,2], Andreas Peter[6], Sebastian Hörber [6], Bjoern Traenkle[7], Philipp D. Kaiser[7], Ulrich Rothbauer [7,8], Matthias Becker [7], Daniel Junker[7], Gérard Krause [9,10,11], Monika Strengert[9,10], Nicole Schneiderhan-Marra [7], Markus F. Templin [7], Thomas O. Joos[7], Daniel J. Kowalewski [12], Vlatka Stos-Zweifel[12], Michael Fehr[2], Armin Rabsteyn[3,13], Valbona Mirakaj[14], Julia Karbach[15], Elke Jäger[15], Michael Graf[16], Lena-Christin Gruber[1], David Rachfalski[1], Beate Preuß[5], Ilona Hagelstein[1,3], Melanie Märklin [1,3], Tamam Bakchoul [17], Cécile Gouttefangeas[2,3,4], Oliver Kohlbacher [16,18,19,20], Reinhild Klein[5], Stefan Stevanović[2,4], Hans-Georg Rammensee[2,3,4] and Juliane S. Walz [1,2,3] ✉

T cell immunity is central for the control of viral infections. To characterize T cell immunity, but also for the development of vaccines, identification of exact viral T cell epitopes is fundamental. Here we identify and characterize multiple dominant and subdominant SARS-CoV-2 HLA class I and HLA-DR peptides as potential T cell epitopes in COVID-19 convalescent and unexposed individuals. SARS-CoV-2-specific peptides enabled detection of post-infectious T cell immunity, even in seronegative convalescent individuals. Cross-reactive SARS-CoV-2 peptides revealed pre-existing T cell responses in 81% of unexposed individuals and validated similarity with common cold coronaviruses, providing a functional basis for heterologous immunity in SARS-CoV-2 infection. Diversity of SARS-CoV-2 T cell responses was associated with mild symptoms of COVID-19, providing evidence that immunity requires recognition of multiple epitopes. Together, the proposed SARS-CoV-2 T cell epitopes enable identification of heterologous and post-infectious T cell immunity and facilitate development of diagnostic, preventive and therapeutic measures for COVID-19.

T cells control viral infections and provide immunological memory that enables long-lasting protection[1–3]. Whereas CD4+ helper T cells orchestrate the immune response and enable B cells to produce antibodies, CD8+ cytotoxic T cells eliminate virus-infected cells. For both, recognition of viral antigens in the form of short peptides presented on HLAs is fundamental. In consequence, characterization of such viral T cell epitopes[4–6] is crucial for the understanding of immune defense mechanisms, but also a prerequisite for the development of vaccines and immunotherapies[3,7–9].

The SARS-CoV-2 coronavirus causes COVID-19, which has become a worldwide pandemic with dramatic socioeconomic consequences[10,11]. Available treatment options are limited, and despite intensive efforts a vaccine is so far not available. Knowledge obtained from the two other zoonotic coronaviruses SARS-CoV-1 and MERS-CoV indicates that coronavirus-specific T cell immunity is an important determinant for recovery and long-term protection[12–15]. This T cell-mediated immune response is even more important as studies on humoral immunity to SARS-CoV-1 provided

[1]Clinical Collaboration Unit Translational Immunology, German Cancer Consortium (DKTK), Department of Internal Medicine, University Hospital Tübingen, Tübingen, Germany. [2]Institute for Cell Biology, Department of Immunology, University of Tübingen, Tübingen, Germany. [3]Cluster of Excellence iFIT (EXC2180) 'Image-Guided and Functionally Instructed Tumor Therapies', University of Tübingen, Tübingen, Germany. [4]German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Partner Site Tübingen, Tübingen, Germany. [5]Department of Hematology, Oncology, Clinical Immunology and Rheumatology, University Hospital Tübingen, Tübingen, Germany. [6]Institute for Clinical Chemistry and Pathobiochemistry, Department for Diagnostic Laboratory Medicine, University Hospital Tübingen, Tübingen, Germany. [7]NMI, Natural and Medical Sciences Institute at the University of Tübingen, Reutlingen, Germany. [8]Pharmaceutical Biotechnology, University of Tübingen, Tübingen, Germany. [9]Department of Epidemiology, Helmholtz Centre for Infection Research, Braunschweig, Germany. [10]TWINCORE GmbH, Centre for Experimental and Clinical Infection Research, a joint venture of the Hannover Medical School and the Helmholtz Centre for Infection Research, Hannover, Germany. [11]German Center for Infection Research, Braunschweig, Germany. [12]Immatics Biotechnologies GmbH, Tübingen, Germany. [13]Department of General Pediatrics, Oncology/Hematology, University Children's Hospital Tübingen, Tübingen, Germany. [14]Department of Anesthesia and Intensive Care Medicine, University Hospital Tübingen, Tübingen, Germany. [15]Department of Oncology and Hematology, Krankenhaus Nordwest, Frankfurt, Germany. [16]Applied Bioinformatics, Center for Bioinformatics and Department of Computer Science, University of Tübingen, Tübingen, Germany. [17]Institute for Clinical and Experimental Transfusion Medicine, University Hospital Tübingen, Tübingen, Germany. [18]Institute for Bioinformatics and Medical Informatics, University of Tübingen, Tübingen, Germany. [19]Biomolecular Interactions, Max-Planck-Institute for Developmental Biology, Tübingen, Germany. [20]Institute for Translational Bioinformatics, University Hospital Tübingen, Tübingen, Germany. [21]These authors contributed equally: Annika Nelde, Tatjana Bilich, Jonas S. Heitmann. ✉e-mail: juliane.walz@med.uni-tuebingen.de

evidence that antibody responses are short-lived and can even cause or aggravate virus-associated lung pathology[16,17]. With regard to SARS-CoV-2, very recent studies[18–20] described CD4+ and CD8+ T cell responses to viral peptide megapools in donors that had recovered from COVID-19 and individuals not exposed to SARS-CoV-2, the latter being indicative of potential T cell cross-reactivity[21,22]. The exact viral epitopes that mediate these T cell responses against SARS-CoV-2, however, were not identified and characterized in detail in these studies, but are prerequisite (1) to delineate the role of post-infectious and heterologous T cell immunity in COVID-19, (2) for establishing diagnostic tools to identify SARS-CoV-2 immunity and, most importantly, (3) to define target structures for the development of SARS-CoV-2-specific vaccines and immunotherapies. In this study, we define SARS-CoV-2-specific and cross-reactive CD4+ and CD8+ T cell epitopes in a large collection of SARS-CoV-2 convalescent as well as nonexposed individuals and their relevance for immunity and the course of COVID-19 disease.

Results

Identification of SARS-CoV-2-derived peptides. A new prediction and selection workflow, based on the integration of the algorithms SYFPEITHI and NetMHCpan, identified 1,739 and 1,591 auspicious SARS-CoV-2-derived HLA class I- and HLA-DR-binding peptides across all ten viral open-reading frames (ORFs) (Fig. 1a and Extended Data Fig. 1a,b). Predictions were performed for the ten and six most common HLA class I (HLA-A*01:01, -A*02:01, -A*03:01, -A*11:01, -A*24:02, -B*07:02, -B*08:01, -B*15:01, -B*40:01 and -C*07:02) and HLA-DR (HLA-DRB1*01:01, -DRB1*03:01, -DRB1*04:01, -DRB1*07:01, -DRB1*11:01 and -DRB1*15:01) allotypes covering 91.7% and 70.6% of the world population with at least one allotype, respectively[23,24] (Extended Data Figs. 1c and 2a). To identify broadly applicable SARS-CoV-2-derived T cell epitopes, we selected 100 SARS-CoV-2-derived HLA class I-binding peptides comprising ten peptides per HLA class I allotype across all ten viral ORFs for immunogenicity screening (range 3–20 peptides per ORF, mean 10; Fig. 1b,c, Extended Data Fig. 1d–m and Supplementary Table 1). In addition, 20 SARS-CoV-2-derived promiscuous HLA-DR-binding peptides across all ORFs from peptide clusters of various HLA-DR allotype restrictions representing 99 different peptide-allotype combinations were included (Fig. 1d,e, Extended Data Fig. 2b–k and Supplementary Tables 2 and 3). Of these HLA-DR-binding peptides, 14 of 20 (70%) contained embedded SARS-CoV-2-derived HLA class I-binding peptides for 7 of 10 HLA class I allotypes. The complete panel of 120 SARS-CoV-2-derived peptides comprised 10% of the total SARS-CoV-2 proteome (57% and 12% of nucleocapsid and spike protein, respectively; Extended Data Fig. 2l) and showed an equally distributed origin of structural ORF proteins (61 of 120 (51%)) encompassing spike, envelope, membrane and nucleocapsid proteins as well as nonstructural or accessory ORFs (59 of 120 (49%)). The broad HLA class I and HLA-DR allotype restriction of the selected SARS-CoV-2-derived peptides covering ten common HLA class I and six common HLA-DR allotypes allowed for a total coverage of at least one HLA allotype in 97.6% of the individuals of the world population (Fig. 1f). Recurrent mutations of SARS-CoV-2 (refs. [25,26]) affected only a minority of selected SARS-CoV-2-derived peptides with 14 of 120 (12%) sequences (1.7% at anchor position), including reported mutation sites (Supplementary Fig. 1 and Supplementary Tables 4 and 5). Taken together, we predicted for the most common HLA allotypes SARS-CoV-2-derived peptides across all ten viral ORFs and selected 100 HLA class I- and 20 HLA-DR-restricted epitope candidates for further immunological characterization.

Characterization of SARS-CoV-2-derived T cell epitopes. Interferon (IFN)-γ ELISPOT screening of in vitro amplified T cells from patients convalescing from SARS-CoV-2 (SARS group 1,

n = 116, Table 1 and Supplementary Table 6) and donors never exposed to SARS-CoV-2 (PRE group A, *n* = 104, samples collected before SARS-CoV-2 pandemic; Table 1 and Supplementary Table 7) validated 29 of 100 (29%) SARS-CoV-2-derived HLA class I-binding peptides (3 of 10 HLA-A*01; 2 of 10 HLA-A*02; 3 of 10 HLA-A*03; 2 of 10 HLA-A*11; 5 of 10 HLA-A*24; 2 of 10 HLA-B*07; 4 of 10 HLA-B*08; 0 of 10 HLA-B*15; 5 of 10 HLA-B*40; and 3 of 10 HLA-C*07) and 20 of 20 (100%) HLA-DR-binding peptides as naturally occurring T cell epitopes (Figs. 2a,b, 3a,b, Tables 2 and 3, Supplementary Figs. 2 and 3 and Supplementary Table 8). Additional flow-cytometry-based analyses for selected SARS-CoV-2-derived T cell epitopes revealed that T cell responses directed against HLA class I-binding peptides were mainly driven by CD8+ T cells and that HLA-DR-binding peptides were recognized by CD4+ T cells, notably in one single donor also by CD8+ T cells (Fig. 3c,d and Supplementary Table 9). Amplified CD4+ T cells often showed multifunctionality (expressing IFN-γ, tumor necrosis factor (TNF) and CD107a), whereas CD8+ T cells mainly produced only IFN-γ upon peptide stimulation (Fig. 3e and Supplementary Table 9). Twelve of 29 (41%) and 11 of 20 (55%) SARS-CoV-2-derived CD8+ and CD4+ T cell epitopes were dominant epitopes (recognized by ≥50% of SARS donors) with recognition frequencies up to 83% (A01_P01) and 95% (DR_P16), respectively (Fig. 2a,b and Tables 2 and 3). T cell responses showed high inter-individual as well as inter-peptide intensity variation (in terms of spot counts per 5 × 10^5 cells). Overall, the intensity of HLA-DR-specific T cell responses in the SARS group was more pronounced compared to those directed against HLA class I T cell epitopes (Fig. 4a,b). All SARS-CoV-2-derived HLA-DR-binding peptides were found to be immunogenic, independently of the source ORF. SARS-CoV-2-derived HLA class I T cell epitopes showed an equally distributed origin from structural (13 of 29 (45%)) and nonstructural or accessory (16 of 29 (55%)) ORFs (Table 2). However, ORF-specific differences regarding the proportion of validated HLA class I T cell epitopes were observed, revealing the highest frequencies for ORF9 (50%, nucleocapsid protein), ORF1 (45%) and ORF3 (38%; Fig. 4c). The highest recognition in SARS donors was observed for HLA class I T cell epitopes derived from ORF2 (55%, spike protein), ORF5 (52%, membrane protein) and ORF3 (45%), as well as for HLA-DR T cell epitopes derived from ORF5 (95%, membrane protein), ORF8 (68%) and ORF4 (55%, envelope protein; Fig. 4d). In summary, we identified and characterized multiple dominant and subdominant SARS-CoV-2-derived HLA class I and HLA-DR T cell epitopes in patients convalescing from COVID-19.

Cross-reactive T cell responses in unexposed individuals. Upon screening the PRE group A, cross-reactive T cell responses to 9 of 29 (31%) of the validated HLA class I and to 14 of 20 (70%) HLA-DR T cell epitopes were detected. Recognition frequencies (donors with T cell responses normalized to all tested donors) of single SARS-CoV-2 HLA class I and HLA-DR T cell epitopes in the PRE group A were lower compared to that of SARS group 1 (up to 27% for B08_P05 and 44% for DR_P01; Fig. 2a,b, Tables 2 and 3 and Supplementary Fig. 4). Recognition frequencies of HLA class I and HLA-DR T cell epitopes in individual donors differed profoundly between the PRE and the SARS group within the different ORFs. ORF1-derived HLA class I (9%) and ORF8-derived HLA-DR (25%) T cell epitopes showed the highest recognition frequencies in the PRE group, whereas none of the T cell epitopes from ORF5 (membrane protein) and ORF10 that were frequently recognized in SARS donors were detected by T cells in PRE donors (Fig. 4d). Donor-specific recognition rates (recognized peptides/ tested peptides) of HLA class I and HLA-DR SARS-CoV-2 T cell epitopes were significantly lower in the PRE group (HLA class I, mean 26 ± 9; HLA-DR, mean 10 ± 5) than in the SARS group (HLA class I, mean 52 ± 23; HLA-DR, mean 52 ± 23; Fig. 5a). Alignments
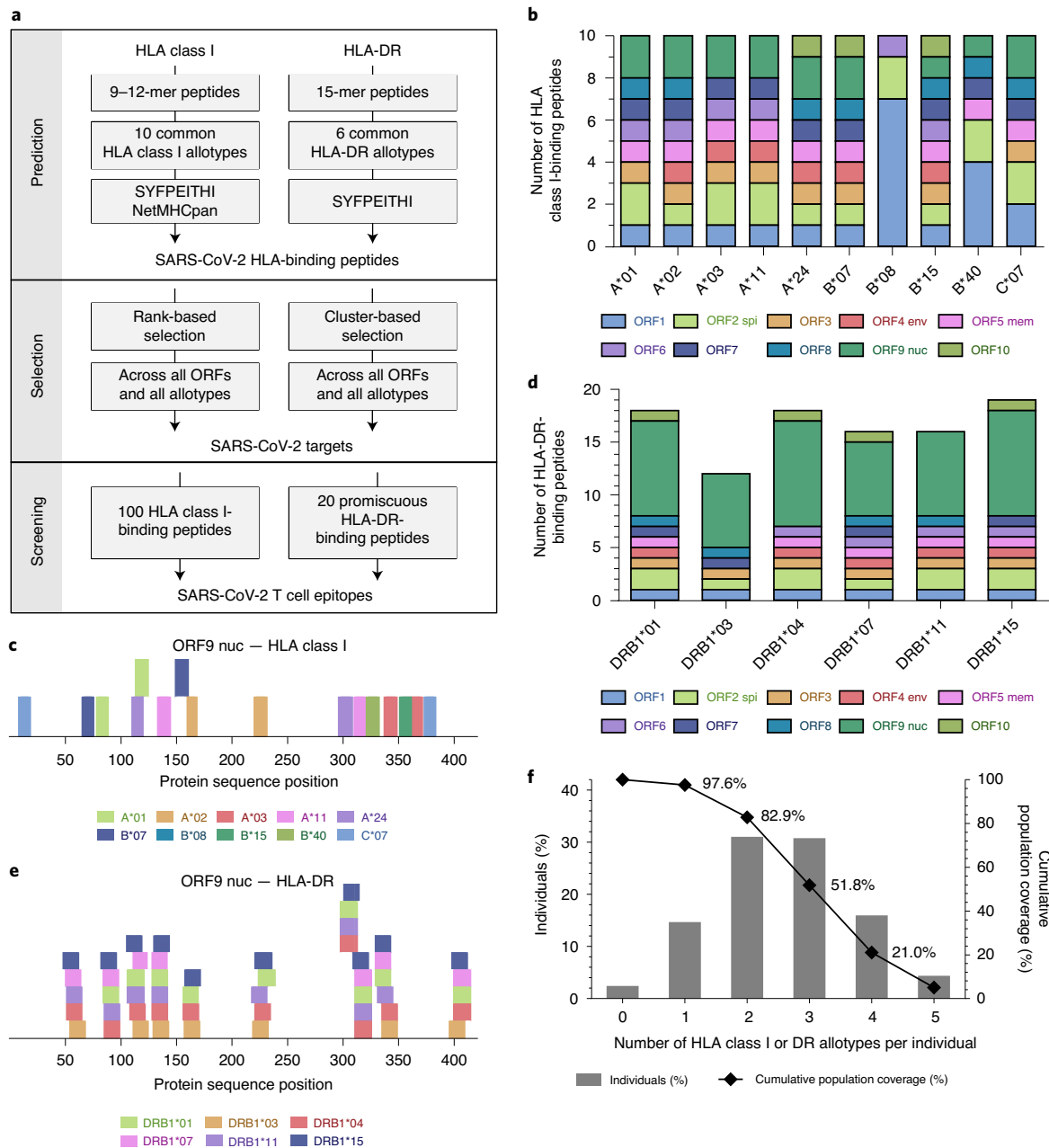
Fig. 1 | Identification and selection of SARS-CoV-2-derived HLA class I- and HLA-DR-binding peptides. a, Schematic overview of our prediction and selection approach and workflow to identify and finally select 120 broadly applicable SARS-CoV-2 HLA class I- and HLA-DR-binding peptides for further screening and validation as T cell epitopes. **b**, Selected HLA class I-binding peptides for the ten most common HLA class I allotypes. Each color represents a distinct ORF. spi, spike protein; env, envelope protein; mem, membrane protein; nuc, nucleocapsid protein. **c**, HLA class I peptide distribution within the ORF9 nucleocapsid protein (for ORF1–ORF8 and ORF10, refer to Extended Data Fig. 1e–m). Each color represents a distinct HLA class I allotype. **d**, Selected HLA-DR-binding peptides for the six most common HLA-DR allotypes. Each color represents a distinct ORF. **e**, HLA-DR peptide cluster distribution within the ORF9 nucleocapsid protein (for ORF1–ORF8 and ORF10, refer to Extended Data Fig. 2c–k). Each color represents a distinct HLA-DR allotype. **f**, Population coverage achieved with the selection of ten common HLA class I and six common HLA-DR allotypes for SARS-CoV-2 T cell epitope screening as compared to the world population. The percentage of individuals within the world population carrying up to five HLA class I or HLA-DR allotypes (x axis) are indicated as gray bars on the left y axis. The cumulative percentage of population coverage is depicted as black dots on the right y axis.

of the SARS-CoV-2 T cell epitopes recognized by unexposed individuals revealed similarities to the four seasonal human common cold coronaviruses (HCoV-OC43, HCoV-229E, HCoV-NL63, HCoV-HKU1) with regard to amino acid sequences, physiochemical and/or HLA-binding properties for 14 of 20 (70%) of the epitopes, thereby providing clear evidence for SARS-CoV-2

T cell cross-reactivity (Fig. 5b, Supplementary Tables 10 and 11 and Supplementary Data 1). Together, cross-reactive T cell responses to SARS-CoV-2 HLA class I and HLA-DR T cell epitopes were identified in unexposed individuals. These cross-reactive peptides showed similarity to common cold coronaviruses, providing functional basis for heterologous immunity in SARS-CoV-2 infection.

Table 1 | Donor characteristics

	SARS collection (*n* = 180)		PRE collection (*n* = 185)	
	group 1	group 2	group A	group B
Number of donors	116	86	104	94
Age (years)				
Range	18–75	18–75	21–70	21–68
Median	44	44	45	50
NA	0	0	5	1
Sex (*n* (%))				
Female	55 (47)	44 (51)	23 (23)	21 (22)
Male	61 (53)	42 (49)	77 (77)	73 (78)
NA	–	–	4	–
Sample collection date	Apr–May 2020	Apr–May 2020	Jun 2007–Nov 2019	May 2017–May 2019
SARS-CoV-2 PCR positivity (*n* (%))	116 (100)	86 (100)	NA	NA
Antibody response (*n* (%))				
positive	96 (84)	71 (83)		
negative	18 (16)	15 (17)	NA	NA
NA	2	0		
Interval positive test to sample collection (days)				
Range	19–52	30–59	NA	NA
Mean	37.7	43.5		
Awareness of symptoms (*n* (%))				
No	10 (9)	6 (7)		
Mild	26 (22)	13 (15)	NA	NA
Moderate	52 (45)	47 (55)		
Severe	28 (24)	20 (23)		
Febrile illness (≥38.0 °C)				
Yes	62 (53)	50 (58)	NA	NA
No	54 (47)	36 (42)		
SC (*n* (%))				
Low SC	50 (43)	36 (42)	NA	NA
High SC	66 (57)	50 (58)		
Hospitalized patients (*n* (%))	3 (3)	2 (2)	NA	NA
Past medical history (*n* (%))				
Arterial hypertension	12 (10)	12 (14)		
Pre-existing autoimmune disease	9 (8)	7 (8)		
Intake of immunosuppressive drugs	5 (4)	6 (7)		
Diabetes mellitus	1 (1)	1 (1)		
Malignant disease	4 (3)	3 (4)	NA	NA
Liver disease	1 (1)	0 (0)		
Lung disease	3 (3)	4 (5)		
Smoking	11 (10)	7 (8)		

Summary of donor characteristics of the SARS groups 1 and 2 as well as of the PRE groups A and B. Out of the SARS and PRE collections two groups were built for (1) T cell epitope screening (group 1 and A) and (2) standardized immunity evaluation (group 2 and B). Donors were assigned to groups according to time of sample acquisition and available sample cell number. Antibody response indicates EUROIMMUNE test results. Awareness of symptoms indicates patient-subjective disease severity. SC was determined by combining objective (fever ≥38.0 °C) and subjective disease symptoms. NA, not available.

T cell responses in convalescent and unexposed individuals.
Epitope screening in SARS and PRE donors enabled the identification of SARS-CoV-2-specific T cell epitopes recognized exclusively in convalescent patients after SARS-CoV-2 infection and of cross-reactive T cell epitopes recognized by both, convalescent patients and SARS-CoV-2 unexposed individuals. To allow for standardized evaluation and determination of T cell response

frequencies to SARS-CoV-2, we designed broadly applicable HLA class I and HLA-DR SARS-CoV-2-specific and cross-reactive T cell epitope compositions (ECs) (Fig. 5c and Extended Data Fig. 6). These ECs were utilized for IFN-γ ELISPOT assays after 12-d in vitro pre-stimulation in groups of convalescent patients (SARS group 2, *n* = 86; Table 1 and Supplementary Table 6) and unexposed donors (PRE group B, *n* = 94; Table 1 and Supplementary Table 7).
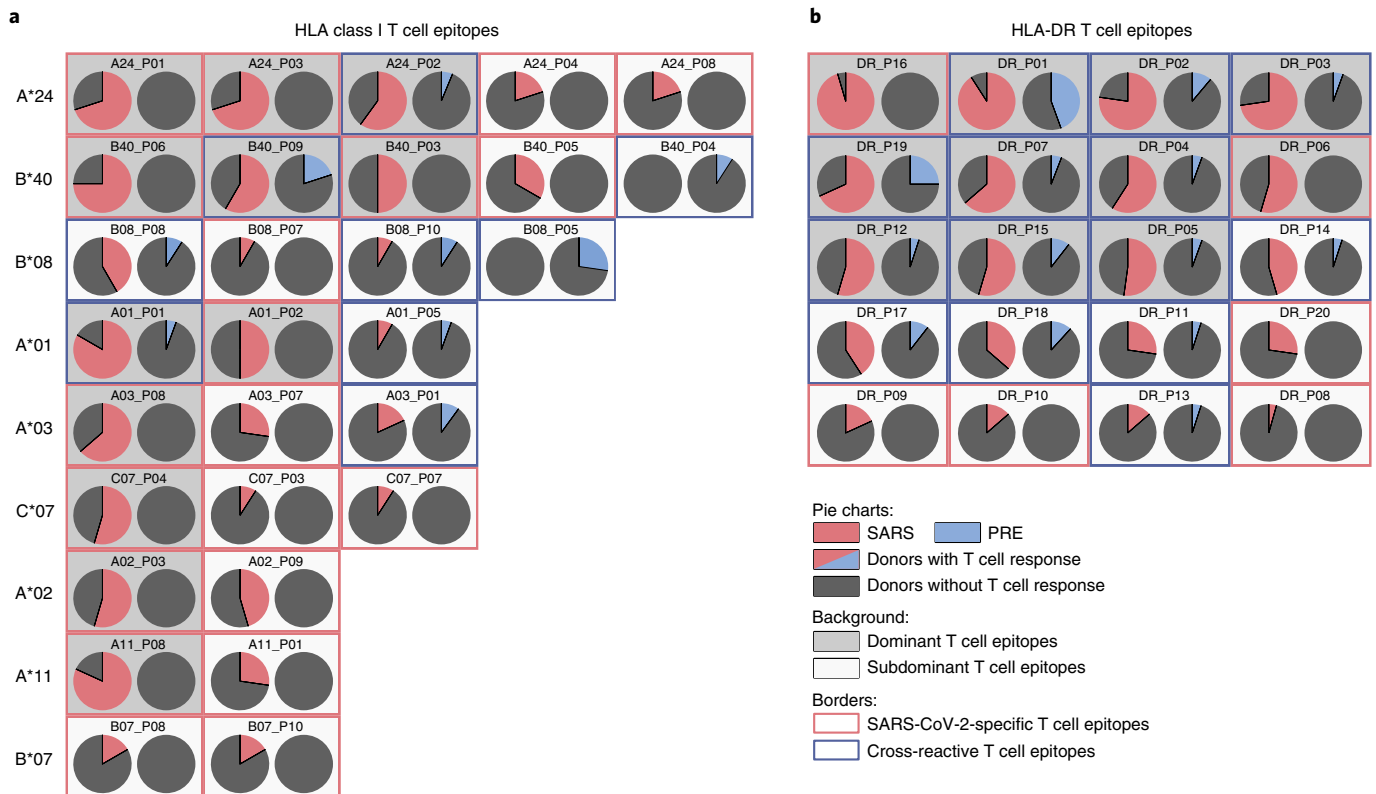
Fig. 2 | Validation of SARS-CoV-2-derived HLA class I and HLA-DR T cell epitopes. a,b, Recognition frequency- and allotype-sorted pie charts of SARS-CoV-2-derived HLA class I (**a**) and HLA-DR (**b**) T cell epitopes. Recognition frequency of T cell epitopes (donors with T cell responses/tested donors) in groups of HLA class I-matched convalescent donors of SARS-CoV-2 infection (SARS group 1, total $n = 116$, left pie chart, red) and donors never exposed to SARS-CoV-2 (PRE group A, total $n = 104$, right pie chart, blue) were assessed by ELISPOT assays after 12-d in vitro pre-stimulation. Dominant (immune responses in $\geq 50\%$ of SARS donors) and subdominant T cell epitopes are marked with dark gray and light gray backgrounds, respectively. SARS-CoV-2-specific T cell epitopes with responses detected exclusively in the SARS group are marked with a red frame, cross-reactive epitopes with immune responses detected in the PRE group are marked with a blue frame.

Of the SARS donors, 100% showed T cell responses to cross-reactive and/or specific ECs (HLA class I 86%, HLA-DR 100%; Fig. 5d,e), whereas 81% of PRE donors showed HLA class I (16%) and/or HLA-DR (77%) T cell responses to cross-reactive ECs (Fig. 5d). In line with the findings obtained with the screening group (SARS group 1), the intensity (in terms of spot counts per 5×10^5 cells) of HLA class I T cell responses was significantly lower compared to HLA-DR T cell responses, both for specific (median calculated spot count HLA class I 379, HLA-DR 760) and cross-reactive ECs (median calculated spot count HLA class I 86, HLA-DR 846; Fig. 5f,g). In line with the differences in recognition rates observed between SARS group 1 and PRE group A, the intensity of T cell responses to cross-reactive ECs was significantly lower in the PRE group (median calculated spot count HLA class I 14, HLA-DR 346) compared to the SARS group (Fig. 5g).

In addition, we evaluated SARS-CoV-2 T cell responses to our ECs ex vivo without 12-d pre-stimulation. Whereas the low-frequent pre-existing SARS-CoV-2 T cells detecting the cross-reactive ECs could not be delineated without pre-stimulation in PRE donors (0 of 42), ex vivo T cell responses to SARS-CoV-2 cross-reactive and/or specific ECs were observed in 96% (45 of 47) of SARS donors (58% HLA class I, 96% HLA-DR; Extended Data Fig. 3a,b). Intensity of T cell responses (in terms of spot counts per 5×10^5 cells) were lower in ex vivo analyses, showing a significant expansion of SARS-CoV-2-specific T cells upon pre-stimulation (Extended Data Fig. 3c–f). In addition to our convalescent SARS collection, including mainly

donors with a mild course of COVID-19, we further evaluated SARS-CoV-2 T cell immunity in a group of hospitalized SARS donors ($n = 21$; Extended Data Fig. 3g). In 81% (17 of 21) of the severely ill patients, T cell responses targeting our specific (71%) or cross-reactive (76%) ECs could be detected ex vivo (Extended Data Fig. 3h). Compared to the ex vivo analyzed donors of SARS group 2, recognition frequencies in the hospitalized group differed most in cross-reactive EC HLA-DR (94% nonhospitalized versus 71% hospitalized). Taken together, SARS-CoV-2 T cell epitopes enabled detection of post-infectious T cell immunity in 100% of individuals convalescing from COVID-19 and revealed pre-existing T cell responses in 81% of unexposed individuals.

Relationship of SARS-CoV-2 T cell and antibody responses. Anti-SARS-CoV-2 IgG responses in SARS donors were analyzed in two independent assays. The anti-SARS-CoV-2 S1 IgG ELISA assay directed against the S1 domain of the viral spike protein, including the immunologically relevant receptor binding domain, revealed 149 of 178 (84%), 7 of 178 (4%) and 22 of 178 (12%) donors with positive, borderline and no anti-S1 response, respectively (Fig. 6a). Of the borderline/nonresponders, 18 of 29 (62%) were also negative in a second, independent anti-nucleocapsid immunoassay (Fig. 6b). However, SARS-CoV-2-specific CD8+ and/or CD4+ T cell responses after a 12-d in vitro pre-stimulation were detected in 10 of 18 (56%) of the 'antibody double-negative' donors (Fig. 6c). The intensity of SARS-CoV-2-specific and cross-reactive HLA-DR T cell responses
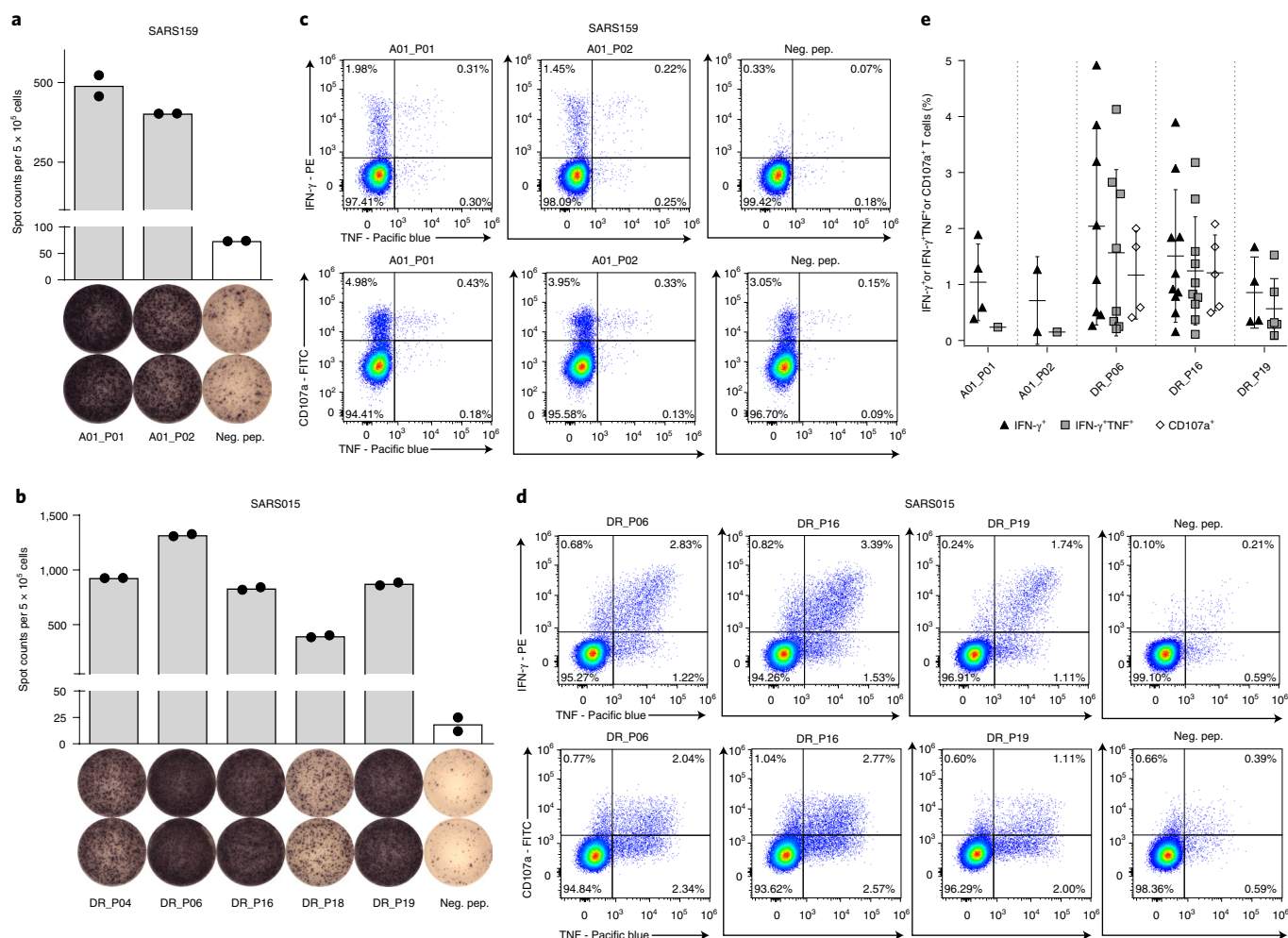
Fig. 3 | Immunological characterization of SARS-CoV-2-derived HLA class I and HLA-DR T cell epitopes. a–d, IFN-γ ELISPOT assay (**a,b**) and flow cytometry-based characterization (**c,d**) of peptide-specific T cells from convalescent SARS donors after 12-d in vitro pre-stimulation with SARS-CoV-2-derived HLA class I- (**a,c**) and HLA-DR-binding (**b,d**) peptides. Flow cytometry data of indicated cytokines and surface markers are shown for CD8⁺ (**c**) and CD4⁺ (**d**) T cells. T cell responses were considered positive when mean spot counts in ELISPOT assays or detected frequency in intracellular cytokine staining was at least threefold higher than the negative control. ELISPOT data are presented as a scatter dot plot with mean. Neg. pep., negative control using an irrelevant HLA-matched peptide. **e**, Percentage of IFN-γ⁺, IFN-γ⁺TNF⁺ or CD107a⁺CD8⁺ (for HLA class I peptides) or CD4⁺ T cells (for HLA-DR peptides) across multiple donors. Indicated percentage depicts the frequency in the sample stimulated with the test peptide minus the frequency of the negative control stimulated with an irrelevant control peptide. Each data point represents one single donor analyzed within one single experiment. Horizontal lines indicate mean with s.d. (error bar). The gating strategy applied for the evaluation of flow-cytometry-acquired data presented in this figure is provided in Supplementary Fig. 5.

correlated with antibody titers (Fig. 6d,e), whereas no correlation was observed with HLA class I T cell responses (Extended Data Fig. 4a,b). No correlation between antibody titers directed against the nucleocapsid of human common cold coronaviruses (HCoV-229E, HCoV-NL63 and HCoV-OC43), as determined by bead-based serological multiplex assays and the intensity of cross-reactive CD4⁺ and CD8⁺ T cell responses in the SARS group, was detected (Extended Data Fig. 4c–h). In conclusion, SARS-CoV-2-specific peptides enable the detection of post-infectious T cell responses, even in seronegative convalescents.

Association of antibody and T cell responses with COVID-19. Finally, the association of anti-SARS-CoV-2 antibody and T cell responses after a 12-d in vitro pre-stimulation with disease severity as assessed by a combinatorial symptom score (SC) of objective (fever ≥38.0 °C) and patient-subjective disease symptoms was determined (Table 1). Alike in critically ill patients[27], independently of age: high-antibody ratios were significantly associated with disease

severity in our collection of convalescent SARS donors (n = 180, group 1 and 2), who in general were in good health and had not been hospitalized (Fig. 6f and Extended Data Fig. 5a). Neither the intensity of SARS-CoV-2-specific nor of cross-reactive T cell responses to HLA class I or HLA-DR ECs correlated with demographics (sex, age or body mass index; Supplementary Tables 12 and 13) or disease severity (Fig. 6g). Rather, diversity of T cell responses in terms of recognition rate of SARS-CoV-2 T cell epitopes (number of recognized epitopes normalized to the total number of tested epitopes in the respective donor) was decreased in patients with more severe COVID-19 symptoms (Fig. 6h and Extended Data Fig. 5b), providing evidence that development of protective immunity requires recognition of multiple SARS-CoV-2 epitopes.

Discussion
This study reports the characterization of multiple broadly applicable SARS-CoV-2-specific and cross-reactive T cell epitopes of various HLA allotype restrictions across all viral ORFs identified in two

Table 2 | Immunogenic SARS-CoV-2-derived HLA class I T cell epitopes

Peptide ID	Sequence	Protein	Protein class	HLA restriction	SARS group 1 (positive/ tested (%))	PRE group A (positive/ tested (%))
A01_P01	TTDPSFLGRY	ORF1	nonstructural	A*01	**10/12 (83%)**	1/18 (6%)
A01_P02	LTDEMIAQY	ORF2 spi	structural	A*01	**6/12 (50%)**	0/18 (0%)
A01_P05	RTFKVSIWNLDY	ORF6	accessory	A*01	1/12 (8%)	1/18 (6%)
A02_P03	ALSKGVHFV	ORF3	accessory	A*02	**6/11 (55%)**	0/8 (0%)
A02_P09	LLLLDRLNQL	ORF9 nuc	structural	A*02	5/11 (45%)	0/8 (0%)
A03_P01	KLFAAETLK	ORF1	nonstructural	A*03	2/11 (18%)	1/10 (10%)
A03_P07	QLRARSVSPK	ORF7	accessory	A*03	3/11 (27%)	0/10 (0%)
A03_P08	KTFPPTEPKK	ORF9 nuc	structural	A*03	**7/11 (64%)**	0/10 (0%)
A11_P01	ASMPTTIAK	ORF1	nonstructural	A*11	3/11 (27%)	0/9 (0%)
A11_P08	ATEGALNTPK	ORF9 nuc	structural	A*11	**9/11 (82%)**	0/9 (0%)
A24_P01	VYIGDPAQL	ORF1	nonstructural	A*24	**7/10 (70%)**	0/17 (0%)
A24_P02	QYIKWPWYI	ORF2 spi	structural	A*24	**6/10 (60%)**	1/16 (6%)
A24_P03	VYFLQSINF	ORF3	accessory	A*24	**7/10 (70%)**	0/17 (0%)
A24_P04	FYVYSRVKNL	ORF4 env	structural	A*24	2/10 (20%)	0/17 (0%)
A24_P08	DYKHWPQIAQF	ORF9 nuc	structural	A*24	2/10 (20%)	0/16 (0%)
B07_P08	FPRGQGVPI	ORF9 nuc	structural	B*07	2/12 (17%)	0/9 (0%)
B07_P10	NPANNAAIVL	ORF9 nuc	structural	B*07	2/12 (17%)	0/9 (0%)
B08_P05	TPKYKFVRI	ORF1	nonstructural	B*08	0/12 (0%)	3/11 (27%)
B08_P07	FVKHKHAFL	ORF1	nonstructural	B*08	1/12 (8%)	0/11 (0%)
B08_P08	DLKGKYVQI	ORF1	nonstructural	B*08	5/12 (42%)	1/11 (9%)
B08_P10	EAFEKMVSL	ORF1	nonstructural	B*08	1/12 (8%)	1/11 (9%)
B40_P03	SELVIGAVIL	ORF5 mem	structural	B*40	**6/12 (50%)**	0/11 (0%)
B40_P04	YEGNSPFHPL	ORF7	accessory	B*40	0/12 (0%)	1/11 (9%)
B40_P05	LEYHDVRVVL	ORF8	accessory	B*40	4/12 (33%)	0/11 (0%)
B40_P06	MEVTPSGTWL	ORF9 nuc	structural	B*40	**9/12 (75%)**	0/11 (0%)
B40_P09	IEYPIIGDEL	ORF1	nonstructural	B*40	**7/12 (58%)**	2/10 (20%)
C07_P03	YYQLYSTQL	ORF3	accessory	C*07	1/11 (9%)	0/9 (0%)
C07_P04	NRFLYIIKL	ORF5 mem	structural	C*07	**6/11 (55%)**	0/9 (0%)
C07_P07	QRNAPRITF	ORF9 nuc	structural	C*07	1/11 (9%)	0/9 (0%)

Summary of immunogenic SARS-CoV-2-derived HLA class I T cell epitopes as defined by IFN-γ ELISPOT assays with detected recognition frequencies in the SARS and PRE groups. Dominant T cell epitopes (immune responses in ≥50% of SARS donors) are marked in bold. ID, identification number.

large collections of donors recovered from SARS-CoV-2 infection as well as unexposed individuals. Our findings aid SARS-CoV-2 research with regard to the understanding of SARS-CoV-2 post-infectious and heterologous T cell responses, but also regarding the development of prophylactic and therapeutic measures.

To allow for the detection of even very small SARS-CoV-2 epitope-recognizing T cell populations especially in unexposed donors, where SARS-CoV-2 cross-reactive T cells were below the detection limit in ex vivo analyses, epitope definition was based on a 12-d pre-stimulation protocol before a routine 18–24-h ELISPOT assay. The requirement of this pre-stimulation protocol is further supported by a recent work characterizing human cytomegalovirus-derived T cell epitopes, showing a loss, even of dominant human cytomegalovirus-derived T cell epitopes when analyzing T cell responses ex vivo without previous amplification[5]. However, as in vitro culture might distort cytokine production or proportions of specific T cell subsets, further studies have to evaluate the physiological cytokine profile and phenotype of SARS-CoV-2-specific T cells in more extensive ex vivo studies. Further validation of the proposed T cell epitopes requires confirmation of MHC binding of the respective peptides, which could be achieved

by refolding experiments to build monomers (MHC–peptide complexes) followed by tetramer staining of T cells or by cytotoxicity experiments utilizing, for example, SARS-CoV-2-infected cell lines.

At present, determination of immunity to SARS-CoV-2 relies on the detection of SARS-CoV-2 antibody responses. However, despite the high sensitivity reported for several assays there is still a substantial percentage of patients with negative or borderline antibody responses and thus unclear immunity status after SARS-CoV-2 infection[28]. Our SARS-CoV-2-specific T cell epitopes, which are not recognized by T cells of unexposed donors, allowed for detection of specific T cell responses even in donors without antibody responses, thereby providing evidence for T cell immunity upon infection. In additional analyses of T cell immunity in hospitalized donors, we could prove SARS-CoV-2 T cell responses also in severely ill patients with COVID-19.

In line with previous data on acute and chronic viral infection[29,30], our data indicate an important role of SARS-CoV-2 CD4+ T cell responses in the natural course of infection, with the identification of multiple dominant HLA-DR T cell epitopes that elicit more frequent and intense immune response in SARS donors compared to the HLA class I T cell epitopes. This guides selection of T cell

Table 3 | Immunogenic SARS-CoV-2-derived HLA-DR T cell epitopes

Peptide ID	Sequence	Protein	Protein class	SARS group 1 positive/tested (%)	PRE group A positive/tested (%)
DR_P01	KDGIIWVATEGALNT	ORF9 nuc	structural	**20/22 (91%)**	8/18 (44%)
DR_P02	GTWLTYTGAIKLDDK	ORF9 nuc	structural	**17/22 (77%)**	2/18 (11%)
DR_P03	RWYFYYLGTGPEAGL	ORF9 nuc	structural	**16/22 (73%)**	1/18 (6%)
DR_P04	ASWFTALTQHGKEDL	ORF9 nuc	structural	**13/22 (59%)**	1/18 (6%)
DR_P05	ASAFFGMSRIGMEVT	ORF9 nuc	structural	**12/23 (52%)**	1/18 (6%)
DR_P06	IGYYRRATRRIRGGD	ORF9 nuc	structural	**12/22 (55%)**	0/17 (0%)
DR_P07	LLLLDRLNQLESKMS	ORF9 nuc	structural	**14/22 (64%)**	1/17 (6%)
DR_P08	AADLDDFSKQLQQSM	ORF9 nuc	structural	1/23 (4%)	0/17 (0%)
DR_P09	AIVLQLPQGTTLPKG	ORF9 nuc	structural	4/22 (18%)	0/17 (0%)
DR_P10	YKHWPQIAQFAPSAS	ORF9 nuc	structural	3/22 (14%)	0/16 (0%)
DR_P11	LDDFVEIIKSQDLSV	ORF1	nonstructural	6/22 (27%)	1/20 (5%)
DR_P12	ITRFQTLLALHRSYL	ORF2 spi	structural	**12/22 (55%)**	1/20 (5%)
DR_P13	FNGLTVLPPLLTDEM	ORF2 spi	structural	3/22 (14%)	1/20 (5%)
DR_P14	FMRIFTIGTVTLKQG	ORF3	accessory	10/22 (45%)	1/20 (5%)
DR_P15	FYVYSRVKNLNSSRV	ORF4 env	structural	**12/22 (55%)**	2/19 (11%)
DR_P16	LSYYKLGASQRVAGD	ORF5 mem	structural	**21/22 (95%)**	0/19 (0%)
DR_P17	IWNLDYIINLIIKNL	ORF6	accessory	9/22 (41%)	2/19 (11%)
DR_P18	QEEVQELYSPIFLIV	ORF7	accessory	8/22 (36%)	2/17 (12%)
DR_P19	SKWYIRVGARKSAPL	ORF8	accessory	**15/22 (68%)**	4/16 (25%)
DR_P20	INVFAFPFTIYSLLL	ORF10	accessory	6/22 (27%)	0/15 (0%)

Summary of immunogenic SARS-CoV-2-derived HLA-DR T cell epitopes as defined by IFN-γ ELISPOT assays with detected recognition frequencies in the SARS and PRE group. Dominant T cell epitopes (immune responses in ≥50% of SARS donors) are marked in bold.

epitopes for vaccine design, also in light of the CD4[+] T cell–dependent stimulation of a protective antibody responses.

Cross-reactivity of T cells for different virus species or even among different pathogens is a well-known phenomenon[31,32] postulated to enable heterologous immunity to a pathogen after exposure to a nonidentical pathogen[21,22,33]. This heterologous immunity facilitated by cross-reactive T cell responses can mediate either beneficial or adverse effects[34,35] such as in Epstein–Barr virus infection, where influenza immunity and the cross-reactive T cell antigen receptor repertoire can lead to protective immunity to Epstein–Barr virus infection[36] or to severe symptoms of infectious mononucleosis[37]. Using predicted or random SARS-CoV-2-derived peptide pools, very recent studies reported pre-existing SARS-CoV-2-directed T cell responses in small groups of unexposed as well as individuals who are seronegative for SARS-CoV-2, thereby suggesting cross-reactivity between human common cold coronaviruses and SARS-CoV-2 (refs. [18–20]). In our study we identified and characterized the exact T cell epitopes that govern SARS-CoV-2 cross-reactivity and proved similarity to human common cold coronaviruses regarding individual peptide sequences, physiochemical and HLA-binding properties[38,39]. Notably, we detected SARS-CoV-2 cross-reactive T cells in 81% of unexposed individuals after a 12-d pre-stimulation. Furthermore, evidence was provided for a lower recognition frequency of cross-reactive HLA-DR EC in hospitalized patients compared to donors with mild COVID-19 course, which might suggest a lack of pre-existing SARS-CoV-2 T cells in severely ill patients. To determine whether expandable, cross-reactive T cells indeed mediate beneficial heterologous immunity and whether this explains the relatively small proportion of severely ill or, even in general, infected patients during this pandemic[40,41], a dedicated study using for example a matched case control or retrospective cohort design applying our cross-reactive SARS-CoV-2 T cell epitopes would be required. Moreover, it has to be emphasized that

the approach of sequence alignments using National Center for Biotechnology Information (NCBI) BLAST[42,43] mainly allows for the detection of cross-reactive epitopes with high sequence similarity, while cross-reactive epitopes with similarities in physiochemical properties within other ORFs of human common cold coronaviruses as well as in other human viruses such as influenza[44] might not be identified.

Our observation that intensity of T cell responses and recognition rate of T cell epitopes was significantly higher in convalescent patients compared to unexposed individuals suggests that not only expansion, but also a spread of SARS-CoV-2 T cell response diversity occurs upon active infection.

The pathophysiological involvement of the immune response in the course of COVID-19 is a matter of intense debate. We showed a correlation of high antibody titers with enhanced COVID-19 symptoms in our cohort of nonhospitalized patients. This finding is in line with recent data describing a correlation of high antibody titers with disease severity in hospitalized patients[27]. Our data together with a recently published study[20] provide evidence that, on the other hand, the intensity of T cell responses does not correlate with disease severity. This finding is of high relevance for the design of vaccines, as it provides evidence that disease-aggravating effects might not hamper the development of prophylactic and therapeutic vaccination approaches aiming to induce SARS-CoV-2-specific T cell responses. In contrast to the intensity of the T cell response, we showed that recognition rates of SARS-CoV-2 T cell epitopes by individual donors were lower in individuals with more severe COVID-19 symptoms. This observation, together with our data on increased T cell epitope recognition rates after SARS-CoV-2 infection compared to pre-existing T cell responses in unexposed individuals and reports from other active or chronic viral infections associating diversity of T cell response with antiviral defense[45–47], provides evidence that natural development and vaccine-based induction of immunity to
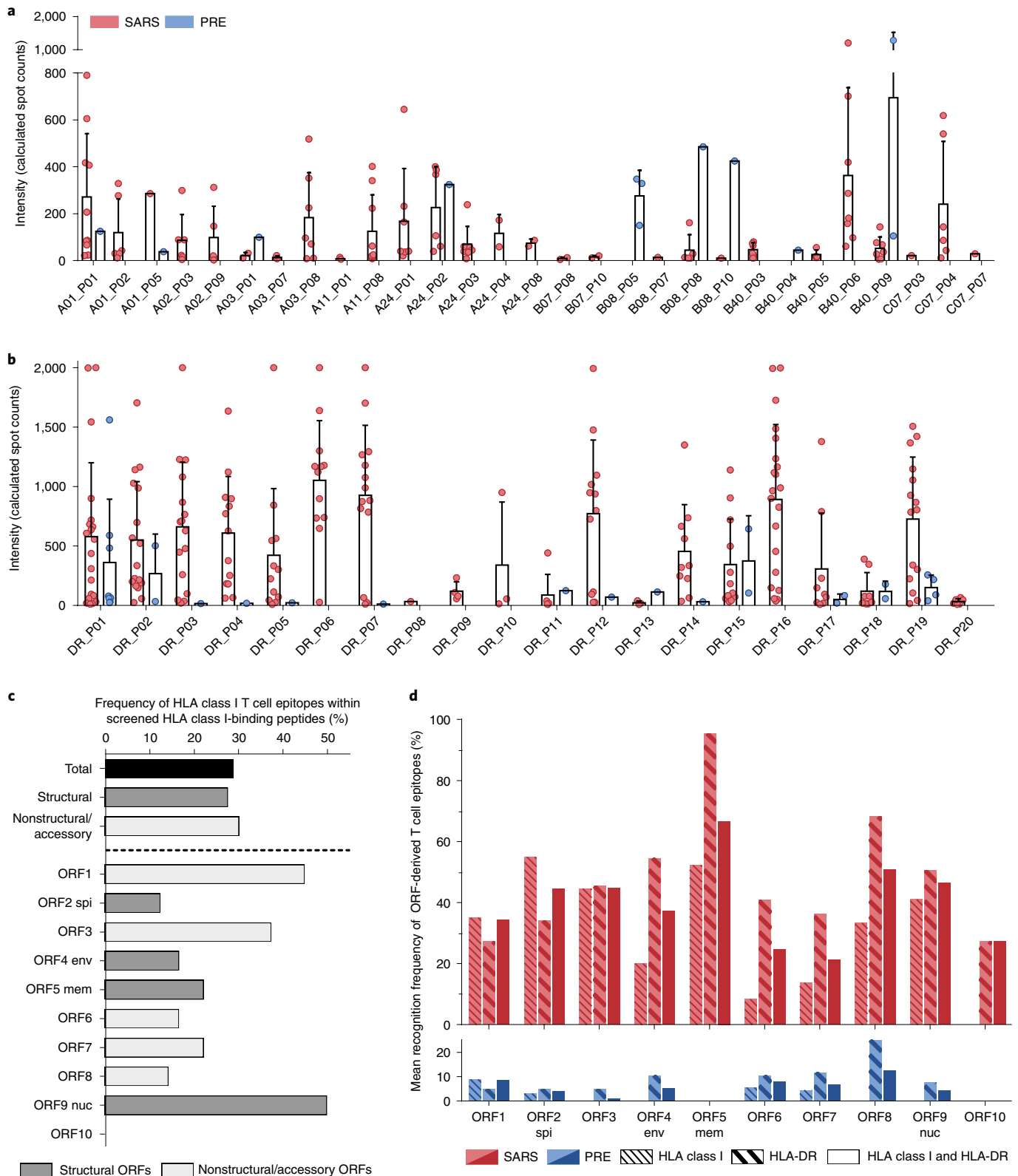
Fig. 4 | Intensity of T cell responses against SARS-CoV-2 HLA class I and HLA-DR T cell epitopes and immunogenicity of different SARS-CoV-2 ORFs.
a,b, Intensity of T cell responses in terms of calculated spot counts in IFN-γ ELISPOT assays after 12-d pre-stimulation against the respective SARS-CoV-2 HLA class I (**a**) and HLA-DR (**b**) T cell epitopes using peripheral blood mononuclear cells (PBMCs) from convalescent SARS-CoV-2-infected donors (SARS) as well as unexposed donors (PRE). Dots represent data from individual donors. Bars represent mean with s.d. (error bar). **c**, Frequency of validated HLA class I T cell epitopes for structural (dark gray) and nonstructural/accessory (light gray) ORFs. **d**, Mean recognition frequency of HLA class I and HLA-DR T cell epitopes by SARS (red) and PRE donors (blue) within the different ORFs.

Fig. 5 | Detection and characterization of T cell responses to SARS-CoV-2-derived HLA class I and HLA-DR T cell epitopes in unexposed individuals.
a, Recognition rate of HLA class I and HLA-DR SARS-CoV-2 T cell epitopes (recognized peptides/tested peptides) in samples of donors from the SARS group 1 (n = 116) and PRE group A (n = 104), respectively (data shown for donors with T cell responses, mean with s.d. (error bars), two-sided Mann–Whitney U-test). **b**, Representative sequence and physiochemical property alignments of the cross-reactive SARS-CoV-2 T cell epitope A24_P02 with the four seasonal human common cold coronaviruses (HCoV-OC43, HCoV-229E, HCoV-NL63, HCoV-HKU1; for other cross-reactive peptides refer to Supplementary Tables 10 and 11 and Supplementary Data 1). Physiochemical properties were calculated by the PepCalc software. Column directions (up versus down) indicate hydrophilicity according to the Hopp–Woods scale. **c**, Schematic overview of the definition of SARS-CoV-2-specific and cross-reactive ECs for standardized evaluation of SARS-CoV-2 T cell responses in a group of convalescent individuals from SARS-CoV-2 infection (SARS group 2, n = 86) and a group of unexposed individuals (PRE group B, n = 94). **d,e**, Recognition frequency (donors with T cell responses/tested donors) of cross-reactive (**d**) and SARS-CoV-2-specific (**e**) ECs by T cells in the SARS group 2 and PRE group B. **f,g**, Calculated spot counts for SARS-CoV-2-specific (HLA class I, n = 68; HLA-DR, n = 78) (**f**) and cross-reactive ECs (**g**) in the SARS group 2 (HLA class I, n = 51; HLA-DR, n = 86) and PRE group B (HLA class I, n = 15; HLA-DR, n = 73) (boxes represent median and 25th to 75th percentiles, whiskers are minimum to maximum, two-sided Mann–Whitney U-test).

SARS-CoV-2 requires recognition of multiple SARS-CoV-2 epitopes. Confirmation of this observation in a larger SARS cohort, including more hospitalized patients is warranted and requires single epitope-based methods to determine T cell epitope recognition rates, as enabled by our SARS-CoV-2 T cell epitopes. Moreover, our data underline the high importance of the identified T cell epitopes for further studies of SARS-CoV-2 immunity, but also for the development of preventive and therapeutic COVID-19 measures. Using the SARS-CoV-2 T cell epitopes we are currently preparing two clinical studies (EudraCT 2020-002502-75; EudraCT 2020-002519-23) to evaluate a multi-peptide vaccine for induction of broad T cell immunity to SARS-CoV-2 to combat COVID-19.
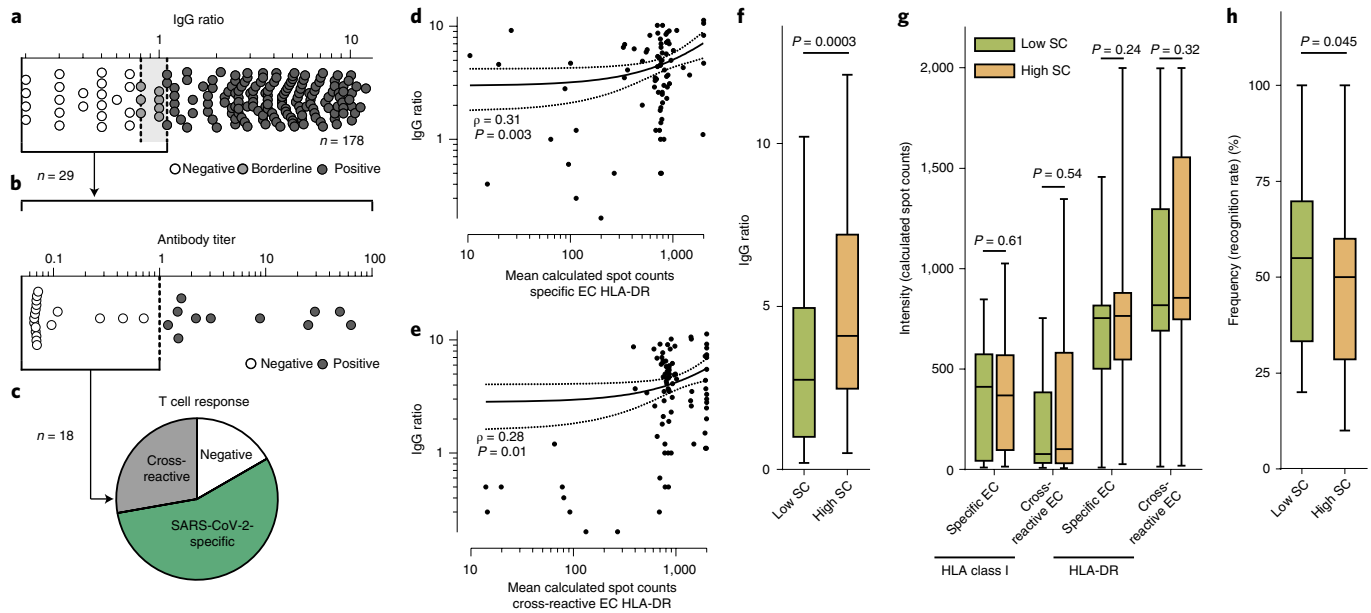
Fig. 6 | SARS-CoV-2-directed antibody and T cell responses in the course of COVID-19. a,b, SARS-CoV-2 serum IgG S1 ratio (EUROIMMUN) in SARS donors ($n=178$) (**a**) and anti-nucleocapsid antibody titers (Elecsys immunoassay) of SARS donors with borderline/negative responses in EUROIMMUN assay ($n=29$) (**b**). Donors with negative and borderline responses are marked in white and gray, respectively. **c,** The pie chart displays T cell responses (positive, $n=15$; negative, $n=3$) to SARS-CoV-2-specific ($n=10$) and cross-reactive ($n=5$) T cell epitopes in donors without antibody responses ($n=18$, assessed in two independent assays). **d,e,** Correlation analysis of IgG ratios (EUROIMMUN) to SARS-CoV-2 with spot counts assessed by ELISPOT assays for HLA-DR SARS-CoV-2-specific ($n=78$) (**d**) and cross-reactive ($n=86$) (**e**) ECs in SARS group 2 (dotted lines, 95% confidence level, Spearman's rho (ρ) and P value). **f,g,** IgG antibody response (EUROIMMUN) to SARS-CoV-2 ($n=178$) (**f**) and T cell response to SARS-CoV-2-specific (HLA class I, $n=68$; HLA-DR, $n=78$) (**g**) and cross-reactive ECs (HLA class I, $n=51$; HLA-DR, $n=86$), respectively, in SARS donors with low and high SC (combining objective (fever ≥38 °C) and subjective disease symptoms) in the course of COVID-19. **h,** Recognition rate of T cell epitopes (recognized peptides/tested peptides) in SARS donors (group 1) with low and high SC in the course of COVID-19 ($n=84$). Boxes represent median and 25th to 75th percentiles, whiskers are minimum to maximum, two-sided Mann–Whitney U-test (**f,g**); boxes represent median and 25th to 75th percentiles, whiskers are minimum to maximum, one-sided Student's t-test (**h**).

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41590-020-00808-x.

References

1. Seder, R. A., Darrah, P. A. & Roederer, M. T-cell quality in memory and protection: implications for vaccine design. *Nat. Rev. Immunol.* **8**, 247–258 (2008).
2. Swain, S. L., McKinstry, K. K. & Strutt, T. M. Expanding roles for CD4+ T cells in immunity to viruses. *Nat. Rev. Immunol.* **12**, 136–148 (2012).
3. Rosendahl Huber, S., van Beek, J., de Jonge, J., Luytjes, W. & van Baarle, D. T cell responses to viral infections—opportunities for peptide vaccination. *Front. Immunol.* **5**, 171 (2014).
4. Khan, N. et al. T cell recognition patterns of immunodominant cytomegalovirus antigens in primary and persistent infection. *J. Immunol.* **178**, 4455–4465 (2007).
5. Lübke, M. et al. Identification of HCMV-derived T cell epitopes in seropositive individuals through viral deletion models. *J. Exp. Med.* **217**, e20191164 (2020).
6. Falk, K. et al. Analysis of a naturally occurring HLA class I-restricted viral epitope. *Immunology* **82**, 337–342 (1994).
7. Einsele, H. et al. Infusion of cytomegalovirus (CMV)-specific T cells for the treatment of CMV infection not responding to antiviral chemotherapy. *Blood* **99**, 3916–3922 (2002).
8. Tan, A. C. et al. The design and proof of concept for a CD8+ T cell-based vaccine inducing cross-subtype protection against influenza A virus. *Immunol. Cell Biol.* **91**, 96–104 (2013).
9. Mudd, P. A. et al. Vaccine-induced CD8+ T cells control AIDS virus replication. *Nature* **491**, 129–133 (2012).
10. Mo, P. et al. Clinical characteristics of refractory COVID-19 pneumonia in Wuhan, China. *Clin. Infect. Dis.* https://doi.org/10.1093/cid/ciaa270 (2020).
11. Khan, S. et al. Emergence of a novel coronavirus, severe acute respiratory syndrome coronavirus 2: biology and therapeutic options. *J. Clin. Microbiol.* https://doi.org/10.1128/JCM.00187-20 (2020); erratum https://doi.org/10.1128/JCM.01297-20 (2020).
12. Zhao, J. et al. Airway memory CD4+ T cells mediate protective immunity against emerging respiratory coronaviruses. *Immunity* **44**, 1379–1391 (2016).
13. Zhao, J., Zhao, J. & Perlman, S. 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).
14. Channappanavar, R., Fett, C., Zhao, J., Meyerholz, D. K. & Perlman, S. Virus-specific memory CD8+ T cells provide substantial protection from lethal severe acute respiratory syndrome coronavirus infection. *J. Virol.* **88**, 11034–11044 (2014).
15. Ng, O. W. et al. Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. *Vaccine* **34**, 2008–2014 (2016).
16. Liu, L. et al. Anti-spike IgG causes severe acute lung injury by skewing macrophage responses during acute SARS-CoV infection. *JCI Insight* https://doi.org/10.1172/jci.insight.123158 (2019).
17. Tang, F. et al. Lack of peripheral memory B cell responses in recovered patients with severe acute respiratory syndrome: a six-year follow-up study. *J. Immunol.* **186**, 7264–7268 (2011).
18. Grifoni, A. et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181**, 1489–1501 (2018).
19. Braun, J. et al. SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19 *Nature* https://doi.org/10.1038/s41586-020-2598-9 (2020).
20. Le Bert, N. et al. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **584**, 457–462 (2020).

21. Petrova, G., Ferrante, A. & Gorski, J. Cross-reactivity of T cells and its role in the immune system. *Crit. Rev. Immunol.* **32**, 349–372 (2012).

22. Su, L. F., Kidd, B. A., Han, A., Kotzin, J. J. & Davis, M. M. Virus-specific CD4+ memory-phenotype T cells are abundant in unexposed adults. *Immunity* **38**, 373–383 (2013).

23. Bui, H. H. et al. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinf.* **7**, 153 (2006).

24. Vita, R. et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* **43**, D405–D412 (2014).

25. Wang, C. et al. The establishment of reference sequence for SARS-CoV-2 and variation analysis. *J. Med. Virol.* **92**, 667–674 (2020).

26. Phan, T. Genetic diversity and evolution of SARS-CoV-2. *Infect. Genet. Evol.* **81**, 104260 (2020).

27. Long, Q. X. et al. Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat. Med.* **26**, 845–848 (2020).

28. Kruttgen, A. et al. Comparison of four new commercial serologic assays for determination of SARS-CoV-2 IgG. *J. Clin. Virol.* **128**, 104394 (2020).

29. Wilkinson, T. M. et al. Preexisting influenza-specific CD4+ T cells correlate with disease protection against influenza challenge in humans. *Nat. Med.* **18**, 274–280 (2012).

30. Soghoian, D. Z. et al. HIV-specific cytolytic CD4+ T cell responses during acute HIV infection predict disease outcome. *Sci. Transl. Med.* **4**, 123ra125 (2012).

31. Vali, B. et al. Characterization of cross-reactive CD8+ T-cell recognition of HLA-A2-restricted HIV-Gag (SLYNTVATL) and HCV-NS5b (ALYDVVSKL) epitopes in individuals infected with human immunodeficiency and hepatitis C viruses. *J. Virol.* **85**, 254–263 (2011).

32. Acierno, P. M. et al. Cross-reactivity between HLA-A2-restricted FLU-M1:58-66 and HIV p17 GAG:77-85 epitopes in HIV-infected and uninfected individuals. *J. Transl. Med.* **1**, 3 (2003).

33. Friberg, H. et al. Memory CD8+ T cells from naturally acquired primary dengue virus infection are highly cross-reactive. *Immunol. Cell Biol.* **89**, 122–129 (2011).

34. Welsh, R. M., Che, J. W., Brehm, M. A. & Selin, L. K. Heterologous immunity between viruses. *Immunol. Rev.* **235**, 244–266 (2010).

35. Benn, C. S., Netea, M. G., Selin, L. K. & Aaby, P. A small jab—a big effect: nonspecific immunomodulation by vaccines. *Trends Immunol.* **34**, 431–439 (2013).

36. Watkin, L. B. et al. Unique influenza A cross-reactive memory CD8+ T-cell receptor repertoire has a potential to protect against EBV seroconversion. *J. Allergy Clin. Immunol.* **140**, 1206–1210 (2017).

37. Aslan, N. et al. Severity of acute infectious mononucleosis correlates with cross-reactive influenza CD8+ T-cell receptor repertoires. *MBio* https://doi.org/10.1128/mBio.01841-17 (2017).

38. Yin, Y. & Mariuzza, R. A. The multiple mechanisms of T cell receptor cross-reactivity. *Immunity* **31**, 849–851 (2009).

39. Borbulevych, O. Y. et al. T cell receptor cross-reactivity directed by antigen-dependent tuning of peptide-MHC molecular flexibility. *Immunity* **31**, 885–896 (2009).

40. Robert Koch Institute. *Coronavirus Disease 2019 (COVID-19) Daily Situation Report of the Robert Koch Institute 29/04/2020* https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Situationsberichte/2020-04-29-en.pdf?__blob=publicationFile (2020).

41. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).

42. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

43. Johnson, M. et al. NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36**, W5–W9 (2008).

44. Sidhom, J.-W. & Baras, A. S. Analysis of SARS-CoV-2 specific T-cell receptors in ImmuneCode reveals cross-reactivity to immunodominant Influenza M1 epitope. Preprint at *bioRxiv* https://doi.org/10.1101/2020.06.20.160499 (2020).

45. Messaoudi, I., Guevara Patino, J. A., Dyall, R., LeMaoult, J. & Nikolich-Zugich, J. Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense. *Science* **298**, 1797–1800 (2002).

46. Tan, A. C., La Gruta, N. L., Zeng, W. & Jackson, D. C. Precursor frequency and competition dictate the HLA-A2-restricted CD8+ T cell responses to influenza A infection and vaccination in HLA-A2.1 transgenic mice. *J. Immunol.* **187**, 1895–1902 (2011).

47. Kiepiela, P. et al. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. *Nat. Med.* **13**, 46–53 (2007).

Methods

Patients and blood samples. Blood and serum samples as well as questionnaire-based assessment of donor characteristics and disease symptoms from convalescent volunteers after SARS-CoV-2 infection were collected at the University Hospital Tübingen and the Cancer Research Department Rhein-Main (Hospital Nordwest) from April to July 2020 (SARS collection, $n = 180$). The collection of unexposed individuals (PRE collection, $n = 185$) includes samples of healthy blood donors (blood donations for research purpose from the Department of Transfusion Medicine, University Hospital Tübingen) that were never exposed to SARS-CoV-2, as the PBMCs of these donors were isolated and asserted (Department of Immunology, Tübingen) before the SARS-CoV-2 pandemic (June 2007 to November 2019). Informed consent was obtained in accordance with the Declaration of Helsinki protocol. The study was approved by and performed according to the guidelines of the local ethics committees (179/2020/BO2, MC 288/2015). Out of the SARS ($n = 180$) and PRE ($n = 185$) collections, two groups were built for (1) single-peptide-based T cell epitope screening (SARS group 1 and PRE group A) and (2) standardized immunity evaluation of ECs using IFN-γ ELISPOT assays after in vitro expansion as well as directly ex vivo (SARS group 2 and PRE group B). Donors were assigned to groups according to time of sample acquisition and available sample cell number. Some donors were analyzed in both groups (1 and 2 or A and B for SARS or PRE, respectively). In addition, samples from hospitalized severely ill SARS donors were collected for ex vivo T cell immunity evaluation. SARS-CoV-2 infection was confirmed by PCR test after nasopharyngeal swab. SARS donor recruitment was performed by online and paper-based calls. Sample collection for SARS donor of group 1 and 2 was performed approximately 3–8 weeks after the end of symptoms and/or negative virus smear. Sample collection of hospitalized SARS donors was performed 5–112 d after positive SARS-CoV-2 PCR. PBMCs were isolated by density gradient centrifugation and stored at −80 °C until further use. Serum was separated by centrifugation for 10 min and supernatant was stored at −80 °C. HLA typing was carried out by Immatics Biotechnology GmbH and the Department of Hematology and Oncology at the University Hospital Tübingen. SC was determined by combining objective (fever ≥38.0 °C) and subjective disease symptoms (no/mild/moderate versus severe, reported by questionnaire) of individual donors. Donors with severe disease symptoms and/or fever were classified as 'high SC' and all others as 'low SC'. Detailed SARS and PRE donor characteristics as well as information on allocation of the donors to the experimental groups are provided in Table 1, Supplementary Tables 6 and 7 and Extended Data Fig. 3g.

Data retrieval. The complete highly conserved and representative annotated proteome sequence of SARS-CoV-2 isolate Wuhan-Hu-1 containing ten different ORFs was retrieved from the NCBI database with the accession number MN908947 (ref. [48]). The amino acid sequence is identical to the reference sequence (EPI_ISL_412026) defined by Wang et al. conducting multiple sequence alignments and phylogenetic analyses of 95 full-length genomic sequences[25].

Prediction of SARS-CoV-2-derived HLA class I-binding peptides. The protein sequences of all ten ORFs were split into 9–12 amino acid-long peptides covering the complete proteome of the virus. The prediction algorithms NetMHCpan 4.0 (ref. [49–51]) and SYFPEITHI 1.0 (ref. [52]) were used to predict the binding of peptides to HLA-A*01:01, -A*02:01, -A*03:01, -A*11:01, -A*24:02, -B*07:02, -B*08:01, -B*15:01, -B*40:01 and -C*07:02. Only peptides predicted as HLA-binding peptides by both algorithms (SYFPEITHI score ≥60%, NetMHCpan rank ≤2) for the respective allotype were further examined. Peptides containing cysteines were excluded to avoid dimerization in a potential subsequent vaccine production process. Peptides derived from the ORF1 polyprotein spanning the cleavage sites of the comprised different protein chains were excluded. An averaged rank combining NetMHCpan- and SYFPEITHI-derived prediction scores was calculated and peptides were ranked for each allotype and ORF separately. Through rank-based selection one peptide for each ORF and each allotype, respectively was selected. For peptides with equal averaged ranks, peptides with higher SYFPEITHI scores were nominated. For some HLA allotypes not every ORF gave rise to an appropriate HLA-binding peptide. To receive ten peptides per HLA allotype and ORF, remaining slots were filled with additional peptides from the ORF9 nucleocapsid protein, the ORF2 spike protein and ORF1.

Prediction of SARS-CoV-2-derived HLA-DR-binding peptides. For HLA-DR predictions all ten ORFs were split into peptides of 15 amino acids, resulting in a total of 9,561 peptides. The prediction algorithm SYFPEITHI 1.0 was used to predict the binding to HLA-DRB1*01:01, -DRB1*03:01, -DRB1*04:01, -DRB1*07:01, -DRB1*11:01 and -DRB1*15:01. The 5% (2% for ORF1) top-scoring peptides of each ORF (based on the total length of each ORF) and each HLA-DR allotype were selected. Position-based sorting of peptides within each ORF revealed peptide clusters of promiscuous peptides binding to several HLA-DR allotypes. Through cluster-based selection, peptide clusters of promiscuous peptides with a common core sequence of nine amino acids were selected. Thereby, ten and two clusters were selected for the ORF9 nucleocapsid and the ORF2 spike protein as well as one cluster for each of the remaining ORFs. Of each selected cluster one representative peptide was selected for immunogenicity analysis excluding cysteine-containing peptides.

Sequence and physiochemical property alignments to human common cold coronaviruses. Potential cross-reactive epitopes of SARS-CoV-2-derived peptides from the four seasonal human common cold coronaviruses (HCoV-OC43, HCoV-229E, HCoV-NL63 and HCoV-HKU1) were identified by sequence alignments of the SARS-CoV-2-derived peptide sequences with the sequences of the common cold coronaviruses using NCBI BLAST[42,43]. The HLA binding of the common cold coronavirus-derived peptides to the HLA allele of the corresponding SARS-CoV-2 peptide were predicted by the algorithms NetMHCpan 4.0 (refs. [49–51]) and SYFPEITHI 1.0 (refs. [52]). Physiochemical property alignments of the SARS-CoV-2-derived peptide sequences with the human common cold coronaviruses were performed by PepCalc (https://pepcalc.com/).

IFN-γ ELISPOT assay following 12-d in vitro stimulation or ex vivo without pre-stimulation. Synthetic peptides were provided by EMC Microcollections and INTAVIS Bioanalytical Instruments. For the 12-d in vitro stimulation, PBMCs were pulsed with HLA class I or HLA-DR peptide pools (1 µg ml⁻¹ per peptide for class I or 5 µg ml⁻¹ for HLA-DR) and cultured for 12 d adding 20 U ml⁻¹ interleukin-2 (Novartis) on days 3, 5 and 7. Peptide-stimulated (expanded/in vitro pre-stimulated) or freshly thawed (ex vivo) PBMCs were analyzed by enzyme-linked immunospot (ELISPOT) assay in duplicates (if not mentioned otherwise). A total of $2–8 \times 10^5$ cells per well were incubated with 1 µg ml⁻¹ (class I) or 2.5 µg ml⁻¹ (HLA-DR) single peptides in 96-well plates coated with anti-IFN-γ (clone 1-D1K, 2 µg ml⁻¹, MabTech). PHA (Sigma-Aldrich) served as positive control, irrelevant HLA-matched control peptides as negative control (negative control peptides are listed in Supplementary Table 14). After 22–24 h incubation, spots were revealed with anti-IFN-γ biotinylated detection antibody (clone 7-B6-1, 0.3 µg ml⁻¹, MabTech), ExtrAvidin-alkaline phosphatase (1:1,000 dilution, Sigma-Aldrich) and BCIP/NBT (5-bromo-4-chloro-3-indolyl-phosphate/nitro-blue tetrazolium chloride, Sigma-Aldrich). Spots were counted using an ImmunoSpot S5 analyzer (CTL) and T cell responses were considered positive when mean spot count was at least threefold higher than the mean spot count of the negative control. The intensity of T cell responses is depicted as calculated spot counts, which were calculated as the mean spot count of duplicates normalized to 5×10^5 cells minus the normalized mean spot count of the respective negative control. In contrast, the recognition frequency of T cell responses within a donor group indicates the relative number of donors that can recognize the respective peptides or ECs (positive donors/tested donors) (Figs. 2a,b, 4d and 5d,e). The frequency (recognition rate) for single donors represents the number of recognized SARS-CoV-2-derived peptides (positive peptides/tested peptides) (Figs. 5a and 6h). For HLA-C*07-restricted peptides, screening in PRE donors was performed using samples of HLA-B*07+ samples due to unavailable HLA-C typing and the known linkage disequilibrium of HLA-B*07 and -C*07 (refs. [53,54]).

Intracellular cytokine and cell surface marker staining. Peptide-specific T cells were further characterized by intracellular cytokine and cell surface marker staining. PBMCs were incubated with 10 µg ml⁻¹ of peptide, 10 µg ml⁻¹ brefeldin A (Sigma-Aldrich) and a 1:500 dilution of GolgiStop (BD) for 12–16 h. Staining was performed using Cytofix/Cytoperm solution (BD), APC/Cy7 anti-human CD4 (1:100 dilution, BioLegend), PE/Cy7 anti-human CD8 (1:400 dilution, Beckman Coulter), Pacific blue anti-human TNF (1:120 dilution, BioLegend), FITC anti-human CD107a (1:100 dilution, BioLegend) and PE anti-human IFN-γ monoclonal antibodies (1:200 dilution, BioLegend). PMA (5 µg ml⁻¹) and ionomycin (1 µM, Sigma-Aldrich) served as positive control. Viable cells were determined using Aqua live/dead (1:400 dilution, Invitrogen). All samples were analyzed on a FACS Canto II cytometer (BD) and evaluated using FlowJo software v.10.0.8 (BD). The gating strategy applied for the evaluation of flow cytometry-acquired data is provided in Supplementary Fig. 5.

SARS-CoV-2 IgG ELISA. The 96-well SARS-CoV-2 IgG ELISA assay (EUROIMMUN, 2606A_A_DE_C03, as constituted on 22 April 2020) was performed on an automated BEP 2000 Advance system (Siemens Healthcare Diagnostics) according to the manufacturer's instructions. The ELISA assay detects anti-SARS-CoV-2 IgG directed against the S1 domain of the viral spike protein and relies on an assay-specific calibrator to report a ratio of specimen absorbance to calibrator absorbance. The final interpretation of positivity is determined by ratio above a threshold value given by the manufacturer: positive (ratio ≥1.1), borderline (ratio 0.8–1.0) or negative (ratio <0.8). Quality control was performed following the manufacturer's instructions on each day of testing.

Elecsys anti-SARS-CoV-2 immunoassay. The Elecsys anti-SARS-CoV-2 assay is an electrogenerated chemiluminescence immunoassay (Roche Diagnostics) and was used according to manufacturer's instructions (v.1.0, as constituted in May 2020). It is intended for the detection of high-affinity antibodies (including IgG) directed against the nucleocapsid protein of SARS-CoV-2 in human serum. Readout was performed on a Cobas e411 analyzer. Negative results were defined by a cutoff index of <1.0. Quality control was performed following the manufacturer's instructions on each day of testing.

Generation of expression constructs for the production of viral antigens. The complementary DNAs encoding the nucleocapsid proteins of HCoV-OC43,

HCoV-NL63 and HCoV-229E (NCBI gene bank accession numbers YP_009555245.1; YP_003771.1; NP_073556.1) were produced with an N-terminal hexahistidine (His₆)-tag by gene synthesis (Thermo Fisher Scientific) and cloned using standard techniques into NdeI/HindIII sites of the bacterial expression vector pRSET2b (Thermo Fisher Scientific).

Protein expression and purification. To express the viral nucleocapsid proteins the respective expression constructs were transformed in *Escherichia coli* BL21(DE3) cells. Protein expression was induced in 1 l TB medium at an optical density (OD_{600}) of 2.5–3 by addition of 0.2 mM isopropyl-β-D-thiogalactopyranoside for 16 h at 20 °C. Cells were collected by centrifugation (10 min, 6,000g) and pellets were suspended in binding buffer (1× PBS, 0.5 M NaCl, 50 mM imidazole, 2 mM PMSF, 2 mM MgCl₂, 150 μg ml⁻¹ lysozyme (Merck) and 625 μg ml⁻¹ DNase I (Applichem)). Cell suspensions were sonified for 15 min (Bandelin Sonopuls HD70, power MS72/D, cycle 50%) on ice, incubated for 1 h at 4 °C in a rotary shaker and sonified again. After centrifugation (30 min at 20,000g) urea was added to a final concentration of 6 M to the soluble protein extract. The extract was filtered through a 0.45-μm filter and loaded on a pre-equilibrated 1-ml HisTrap^FF column (GE Healthcare). The bound His-tagged nucleocapsid proteins were eluted by a linear gradient (30 ml) ranging from 50 to 500 mM imidazole in elution buffer (1× PBS, pH 7.4, 0.5 M NaCl, 6 M Urea). Elution fractions (0.5 ml) containing the His-tagged nucleocapsid proteins were pooled and dialyzed (D-Tube Dialyzer Mega, Novagen) into PBS. All purified proteins were analyzed via standard SDS–PAGE, followed by staining with InstantBlue (Expedeon) and immunoblotting using an anti-His (1:1,000 dilution, QIAGEN) in combination with a donkey anti-mouse labeled with AlexaFluor647 (1:2,000 dilution, Invitrogen) on a Typhoon Trio (GE Healthcare, excitation 633 nM, emission filter settings 670 nM BP 30) to confirm protein integrity.

Preparation of beads for serological multiplex assay. Antigens were covalently immobilized on spectrally distinct populations of carboxylated paramagnetic beads (MagPlex Microspheres, Luminex Corporation) using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide/sulfo-N-hydroxysuccinimide chemistry. For immobilization, a magnetic particle processor (KingFisher 96, Thermo Fisher Scientific) was used. Bead stocks were vortexed thoroughly and sonicated for 15 s. A 96-deep-well plate and tip comb was blocked with 1.1 ml 0.5% (v/v) Triton X-100 for 10 min. Afterwards, 83 μl of 0.065% (v/v) Triton X-100 and 1 ml bead stock were added to each well. Finally, each well contained 0.005% (v/v) Triton X-100 and 12.5×10⁷ beads of one single bead population. The beads were washed twice with 500 μl activation buffer (100 mM Na₂HPO₄, pH 6.2, 0.005% (v/v) Triton X-100) and beads were activated for 20 min in 300 μl activation mix containing 5 mg ml⁻¹ 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide and 5 mg ml⁻¹ sulfo-N-hydroxysuccinimide in activation buffer. Following activation, the beads were washed twice with 500 μl coupling buffer (500 mM MES, pH 5.0 + 0.005% (v/v) Triton X-100). Antigens were diluted to 39 μg ml⁻¹ in coupling buffer and incubated with activated beads for 2 h at 21 °C to immobilize antigens on the surface. Antigen-coupled beads were washed twice with 800 μl wash buffer (1× PBS + 0.005% (v/v) Triton X-100) and finally, were resuspended in 1 ml storage buffer (1× PBS + 1% (w/v) BSA + 0.05% (v/v) ProClin). The beads were stored at 4 °C until further use.

Bead-based serological multiplex assay. To detect human IgG directed against nucleocapsid proteins from three different coronavirus species (HCoV-229E, HCoV-NL63 and HCoV-OC43), a bead-based multiplex assay was performed. All antigens were immobilized on different bead populations as described above. The individual bead populations were combined in a bead mix. A total of 25 μl of diluted serum sample were added to 25 μl of the bead mix resulting in a final sample dilution of 1:400 and incubated for 2 h at 21 °C. Unbound antibodies were removed by washing the beads three times with 100 μl wash buffer (1× PBS + 0.05% (v/v) Tween20) per well using a microplate washer (Biotek 405TS, Biotek Instruments). Bound antibodies were detected by incubating the beads with PE-labeled goat-anti-human IgG detection antibodies (Jackson Dianova) at a final concentration of 5 μg ml⁻¹ for 45 min at 21 °C. Measurements were performed using a Luminex FlexMap 3D instrument using Luminex xPONENT Software v.4.3 (sample size, 80 μl; 100 events; gate, 7,500–15,000; reporter gain, standard PMT). Data analysis was performed on mean fluorescence intensity.

Software and statistical analysis. The population coverage of HLA allotypes was calculated by the IEDB population coverage tool (www.iedb.org). Flow cytometric data were analyzed using FlowJo v.10.0.8 (BD). Data are displayed as mean with s.d., box plots as median with 25th or 75th quantiles and min/max whiskers. Continuous data were tested for distribution and individual groups were tested by use of an unpaired Student's *t*-test, Mann–Whitney *U*-test or Kruskal–Wallis test and corrected for multiple comparison as indicated. Spearman's rho (ρ) was calculated for correlation between continuous data. A logistic regression model was used to calculate odds ratios and 95% confidence intervals. Factors before the outcome and measured continuous variables were included in the model. Missing data were included in tables and in descriptive analysis. Graphs were plotted using GraphPad Prism v.8.4.0. Statistical analyses were conducted using GraphPad Prism v.8.4.0 and JMP Pro (SAS Institute, v.14.2) software. *P* values <0.05 were considered statistically significant.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
Data relating to the findings of this study are available from the corresponding author upon request. Source data are provided with this paper.

References
48. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269 (2020).
49. Hoof, I. et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 61, 1 (2009).
50. Nielsen, M. & Andreatta, M. NetMHCpan-3.0: improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med.* 8, 33 (2016).
51. Jurtz, V. et al. NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* 199, 3360–3368 (2017).
52. Rammensee, H., Bachmann, J., Emmerich, N. P., Bachor, O. A. & Stevanovic, S. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics* 50, 213–219 (1999).
53. Schlott, F. et al. Characterization and clinical enrichment of HLA-C*07:02-restricted cytomegalovirus-specific CD8⁺ T cells. *PLoS ONE* 13, e0193554 (2018).
54. Schmidt, A. H. et al. Estimation of high-resolution HLA-A, -B, -C, -DRB1 allele and haplotype frequencies based on 8,862 German stem cell donors and implications for strategic donor registry planning. *Hum. Immunol.* 70, 895–902 (2009).

Author contributions
A.N., H.-G.R., S.S., C.G. and J.S.W. designed the study; A.N., S.S. and J.S.W. performed in silico prediction and selection of candidate peptides; T. Bilich, Y.M., M.L., A.N., J.B., J.R., M.W., M.F., I.H. and M.M. conducted in vitro T cell experiments. B.P., R.K., D.J.K. and V.S.-Z. conducted HLA allotype analysis. B.T., P.D.K. and U.R. generated expression plasmids and purified proteins for the multiplex serological Luminex assay, which was developed and conducted by M.B., D.J., G.K., M.S., N.S.-M., M.F.T. and T.O.J. SARS-CoV-2 IgG was detected by S.H. and A.P. J.S.H., M.R., A.R., V.M., J.K., E.J., T. Bakchoul, L.-C.G., D.R., H.R.S. and J.S.W. conducted patient data and sample collection as well as medical evaluation and analysis. A.N., T. Bilich, J.S.H., M.G., O.K. and J.S.W. analyzed data and performed statistical analyses. A.N., T. Bilich, J.S.H., H.R.S. and J.S.W. drafted the manuscript. H.-G.R., S.S. and J.S.W. supervised the study.

a

ORF	ORF length (aa)	# 9-12mers	# HLA class I peptides
ORF1	7,096	28,346	1,327
ORF2 spi	1,273	5,054	192
ORF3	275	1,062	52
ORF4 env	75	262	17
ORF5 mem	222	850	43
ORF6	61	206	16
ORF7	121	446	20
ORF8	121	446	13
ORF9 nuc	419	1,638	52
ORF10	38	114	7
total	9,701	38,424	1,739

b

ORF	A*01	A*02	A*03	A*11	A*24	B*07	B*08	B*15	B*40	C*07
ORF1	106	285	248	240	183	102	26	173	30	143
ORF2 spi	13	32	32	44	24	17	3	31	4	19
ORF3	3	17	7	9	9	6	0	5	0	4
ORF4 env	0	7	2	2	2	1	0	4	0	0
ORF5 mem	3	6	6	10	9	2	0	4	1	10
ORF6	1	7	5	2	0	2	1	2	0	0
ORF7	1	6	2	2	3	2	0	3	1	2
ORF8	1	4	0	0	2	1	0	2	1	3
ORF9 nuc	5	5	14	10	4	9	0	7	1	2
ORF10	0	0	0	0	4	2	0	1	0	0
total	133	369	316	319	240	142	30	232	38	183

c

d

e ORF1

f ORF2 spi

g ORF3

h ORF4 env

i ORF5 mem

j ORF6

k ORF7

l ORF8

m ORF10

Legend: A*01, A*02, A*03, A*11, A*24, B*07, B*08, B*15, B*40, C*07

Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Prediction of SARS-CoV-2-derived HLA class I-binding peptides. a, Overview of amino acid lengths of SARS-CoV-2 ORFs, total number of 9–12 amino acid-long peptides and number of predicted HLA class I-binding peptides. aa, amino acid; spi, spike protein; env, envelope protein; mem, membrane protein; nuc, nucleocapsid protein. **b**, Number of predicted HLA class I-binding peptides for each HLA class I allotype. **c**, HLA class I allotype population coverage achieved with the selection of HLA class I allotype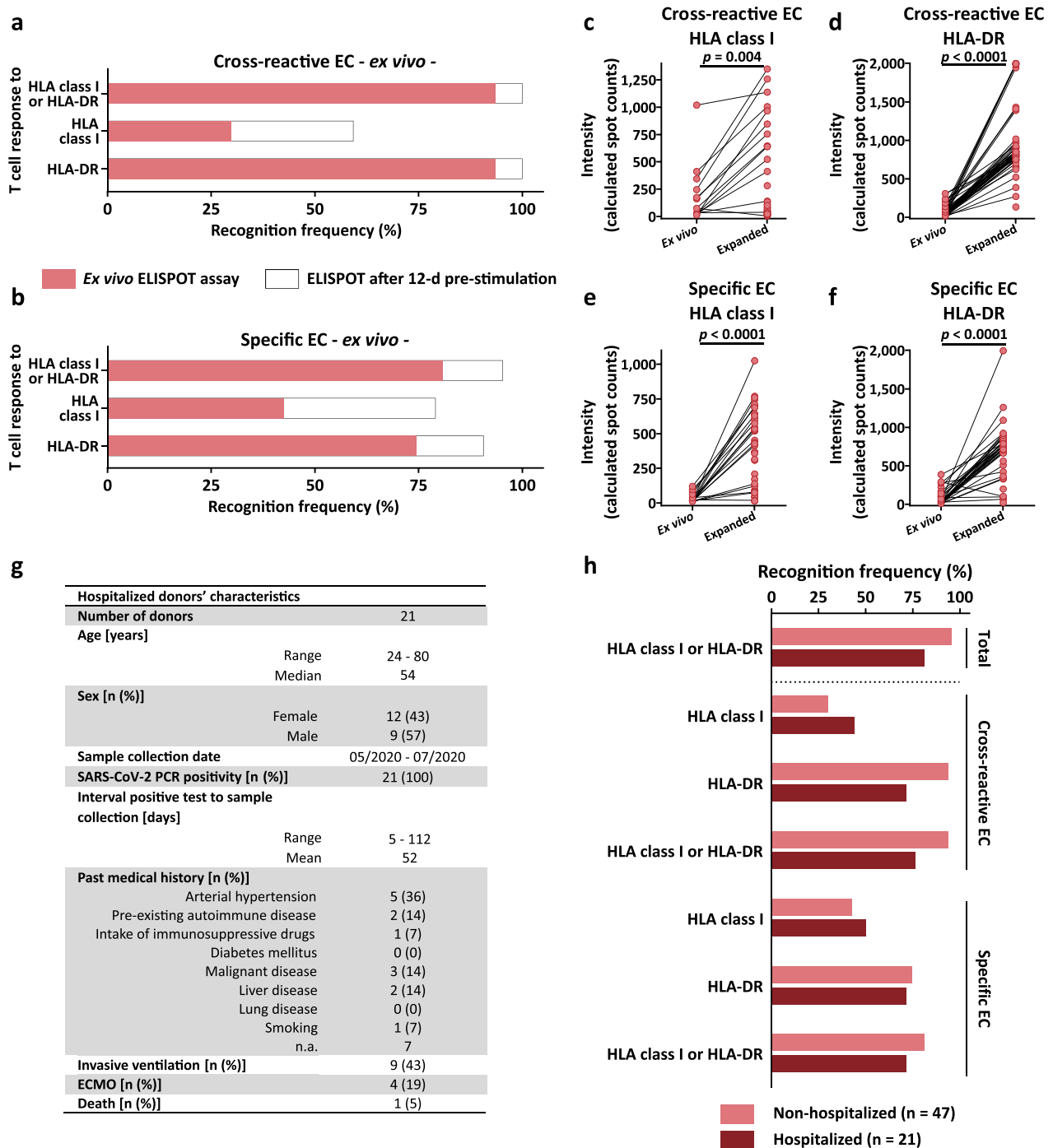s compared to the world population. The frequencies of individuals within the world population carrying up to five HLA allotypes (*x* axis) are indicated as gray bars on the left *y* axis. The cumulative percentage of population coverage is depicted as black dots on the right *y* axis. **d**, Distribution of different HLA class I-restricted peptides within SARS-CoV-2 ORFs. Each color represents a distinct HLA class I allotype. **e-m**, HLA class I-binding peptide distribution within the different SARS-CoV-2 ORFs. Chains (gray) and domains (red) of ORF1 and ORF2 are indicated. Each color represents a distinct HLA class I allotype. nsp, nonstructural protein; p, proteinase; poly., polymerase; heli., helicase; G-tra., Guanine-N7 methyltransferase; endo., endoribonuclease; O-tra., 2'-O-methyltransferase; sig. pep., signal peptide; RBD, receptor binding domain.

Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Prediction of SARS-CoV-2-derived HLA-DR-binding peptides and ORF coverage with predicted HLA class I- and HLA-DR-binding peptides. a, HLA-DR allotype population coverage achieved with the selection of HLA-DR allotypes compared to the world population. The frequencies of individuals within the world population carrying up to two HLA-DR allotypes (x axis) are indicated as gray bars on the left y axis. The cumulative percentage of population coverage is depicted as black dots on the right y-axis. **b**, Number of predicted HLA-DR-binding peptides for each HLA-DR allotype covered with the selected peptide clusters. Each color represents a distinct HLA-DR allotype. spi, spike protein; env, envelope protein; mem, membrane protein; nuc, nucleocapsid protein. **c-k**, Distribution of peptide clusters selected for immunogenicity screening within the different SARS-CoV-2 ORFs. Chains (gray) and domains (red) of ORF1 and ORF2 are indicated. Each color represents a distinct HLA-DR allotype. nsp, nonstructural protein; p, proteinase; poly., polymerase; heli., helicase; G-tra., Guanine-N7 methyltransferase; endo., endoribonuclease; O-tra., 2'-O-methyltransferase; sig. pep., signal peptide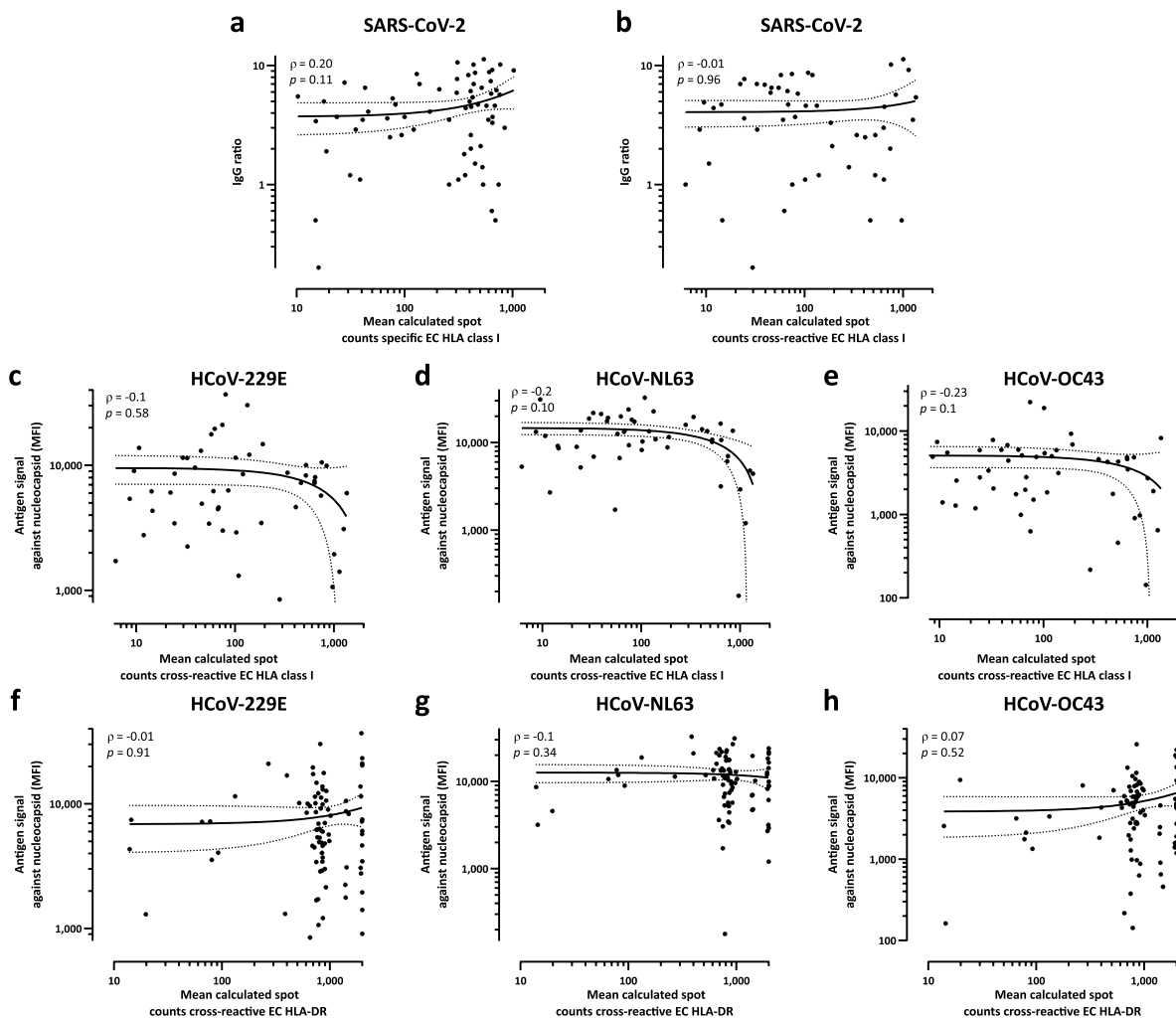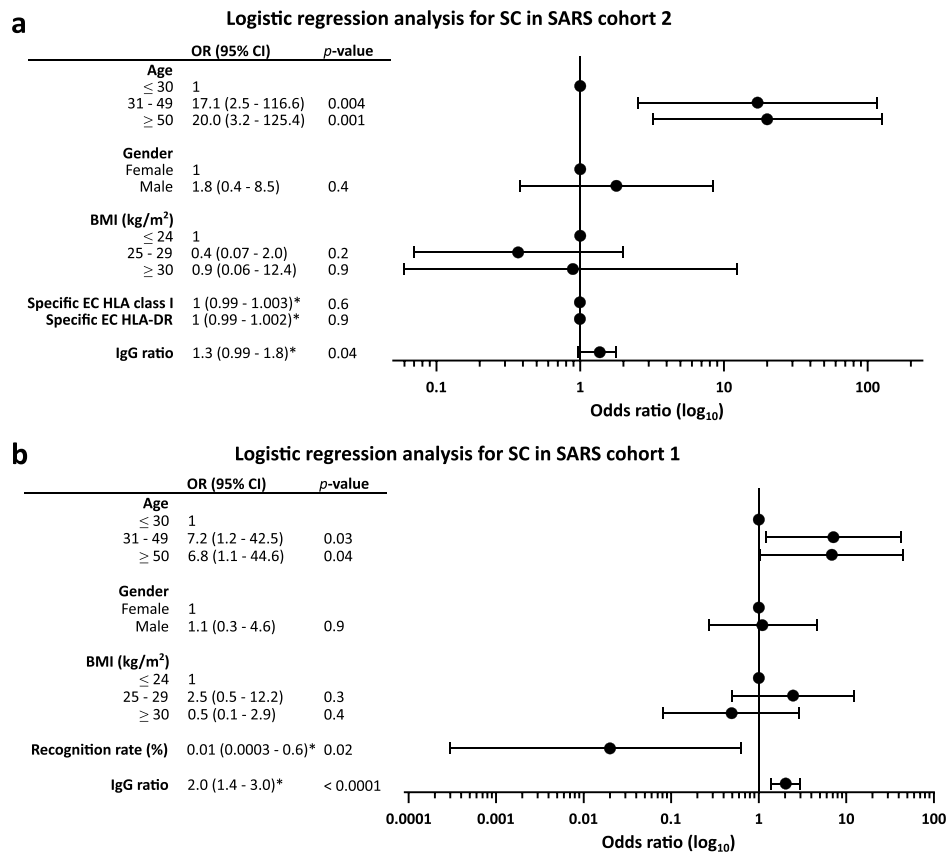; RBD, receptor binding domain. **l**, Protein coverage of total SARS-CoV-2 proteome and individual SARS-CoV-2 ORFs with selected HLA class I- and HLA-DR-binding peptides (left y axis). Striped bars indicate the proportion of individual ORF protein lengths within the total SARS-CoV-2 proteome (right y axis). The protein coverage of HLA-DR-derived peptides for ORF1 (marked with #) amounts to 0.3%.

Extended Data Fig. 3 | Ex vivo T cell responses to SARS-CoV-2-derived HLA class I and HLA-DR T cell epitope compositions in SARS-CoV-2 convalescent and hospitalized COVID-19 patients. a,b, Ex vivo recognition frequency (donors with T cell responses/tested donors) of (**a**) cross-reactive and (**b**) SARS-CoV-2-specific ECs by nonexpanded T cells of SARS donors ($n = 47$) of group 2 analyzed by ex vivo IFN-γ ELISPOT assay. Red bars depict the recognition frequency in ex vivo analyses whereas the white bars indicate the frequency after a 12-d pre-stimulation. **c–f,** Intensity of T cell responses in terms of calculated spot counts against the (**c,d**) cross-reactive and (**e,f**) SARS-CoV-2-specific (**c,e**) HLA class I and (**d,f**) HLA-DR ECs, respectively directly ex vivo and after a 12-d expansion in samples of SARS convalescent donors ($n = 47$). Each spot represents a single donor, paired samples are connected by continuous lines, two-sided Wilcoxon test. Only positive donors are depicted. T cell responses were considered positive when mean spot counts were at least threefold higher than the negative control. **g,** Characteristics of hospitalized donors ($n = 21$) analyzed in ex vivo IFN-γ ELISPOT assay. ECMO, extracorporeal membrane oxygenation; *n*, number; n.a., not available. **h,** Ex vivo recognition frequency (donors with T cell responses/tested donors) of SARS donors of group 2 compared to hospitalized donors analyzed by ex vivo IFN-γ ELISPOT assay.

Extended Data Fig. 4 | Correlation of antibody response with measured T cell intensity. a,b, Correlation analysis of IgG serum ELISA ratios (EUROIMMUN) to SARS-CoV-2 and calculated spot counts assessed in IFN-γ ELISPOT assays after a 12-d in vitro pre-stimulation for HLA class I-restricted (**a**) SARS-CoV-2-specific ($n = 68$) and (**b**) cross-reactive ($n = 51$) ECs in SARS group 2 (dotted lines: 95% confidence level, Spearman's rho (ρ) and P value). **c–h**, Correlation analysis of IgG (MFI signals) to nucleocapsid protein of three common cold coronaviruses (HCoV-OC43, HCoV-229E, HCoV-NL63) and calculated spot counts assessed in IFN-γ ELISPOT assays after a 12-d in vitro pre-stimulation for (**c-e**) HLA class I ($n = 51$) and (f-h) HLA-DR ($n = 86$) cross-reactive ECs in SARS group 2 (dotted lines: 95% confidence level, Spearman's rho (ρ) and P value). MFI, mean fluorescence intensity.

a

Logistic regression analysis for SC in SARS cohort 2

	OR (95% CI)	*p*-value
Age		
≤ 30	1	
31 - 49	17.1 (2.5 - 116.6)	0.004
≥ 50	20.0 (3.2 - 125.4)	0.001
Gender		
Female	1	
Male	1.8 (0.4 - 8.5)	0.4
BMI (kg/m²)		
≤ 24	1	
25 - 29	0.4 (0.07 - 2.0)	0.2
≥ 30	0.9 (0.06 - 12.4)	0.9
Specific EC HLA class I	1 (0.99 - 1.003)*	0.6
Specific EC HLA-DR	1 (0.99 - 1.002)*	0.9
IgG ratio	1.3 (0.99 - 1.8)*	0.04

Odds ratio (log₁₀): 0.1, 1, 10, 100

b

Logistic regression analysis for SC in SARS cohort 1

	OR (95% CI)	*p*-value
Age		
≤ 30	1	
31 - 49	7.2 (1.2 - 42.5)	0.03
≥ 50	6.8 (1.1 - 44.6)	0.04
Gender		
Female	1	
Male	1.1 (0.3 - 4.6)	0.9
BMI (kg/m²)		
≤ 24	1	
25 - 29	2.5 (0.5 - 12.2)	0.3
≥ 30	0.5 (0.1 - 2.9)	0.4
Recognition rate (%)	0.01 (0.0003 - 0.6)*	0.02
IgG ratio	2.0 (1.4 - 3.0)*	< 0.0001

Odds ratio (log₁₀): 0.0001, 0.001, 0.01, 0.1, 1, 10, 100

Extended Data Fig. 5 | Predictors of symptom severity in patient collection. a,b, Odds ratios (ORs) for age, sex, body mass index (BMI) groups, IgG antibody responses (EUROIMMUN), and (**a**) intensity of T cell responses to HLA class I and HLA-DR SARS-CoV-2-specific ECs or (**b**) recognition rate based on an adjusted model of predictors for low versus high symptom score (SC) in SARS donors of group (**a**) 2 and (**b**) 1, respectively. * adjusted OR per unit increase in continuous variable.

HLA class I T cell epitope compositions

	SARS-CoV-2-specific EC				Cross-reactive EC		
Peptide ID	Sequence	ORF	HLA restriction	Peptide ID	Sequence	ORF	HLA restriction
A01_P02	LTDEMIAQY	ORF2 spi	A*01	A01_P01	TTDPSFLGRY	ORF1	A*01
A02_P03	ALSKGVHFV	ORF3	A*02	A01_P05	RTFKVSIWNLDY	ORF6	A*01
A02_P09	LLLLDRLNQL	ORF9 nuc	A*02	A03_P01	KLFAAETLK	ORF1	A*03
A03_P07	QLRARSVSPK	ORF7	A*03	A24_P02	QYIKWPWYI	ORF2 spi	A*24
A03_P08	KTFPPTEPKK	ORF9 nuc	A*03	B08_P05	TPKYKFVRI	ORF1	B*08
A11_P01	ASMPTTIAK	ORF1	A*11	B08_P08	DLKGKYVQI	ORF1	B*08
A11_P08	ATEGALNTPK	ORF9 nuc	A*11	B08_P10	EAFEKMVSL	ORF1	B*08
A24_P01	VYIGDPAQL	ORF1	A*24	B40_P04	YEGNSPFHPL	ORF7	B*40
A24_P03	VYFLQSINF	ORF3	A*24	B40_P09	IEYPIIGDEL	ORF1	B*40
B07_P08	FPRGQGVPI	ORF9 nuc	B*07				
B07_P10	NPANNAAIVL	ORF9 nuc	B*07				
B08_P07	FVKHKHAFL	ORF1	B*08				
B40_P03	SELVIGAVIL	ORF5 mem	B*40				
B40_P06	MEVTPSGTWL	ORF9 nuc	B*40				
C07_P03	YYQLYSTQL	ORF3	C*07				
C07_P04	NRFLYIIKL	ORF5 mem	C*07				

HLA-DR T cell epitope compositions

	SARS-CoV-2-specific EC				Cross-reactive EC		
Peptide ID	Sequence	ORF	HLA restriction	Peptide ID	Sequence	ORF	HLA restriction
DR_P06	IGYYRRATRRIRGGD	ORF9 nuc	DR	DR_P01	KDGIIWVATEGALNT	ORF9 nuc	DR
DR_P09	AIVLQLPQGTTLPKG	ORF9 nuc	DR	DR_P02	GTWLTYTGAIKLDDK	ORF9 nuc	DR
DR_P10	YKHWPQIAQFAPSAS	ORF9 nuc	DR	DR_P03	RWYFYYLGTGPEAGL	ORF9 nuc	DR
DR_P16	LSYYKLGASQRVAGD	ORF5 mem	DR	DR_P04	ASWFTALTQHGKEDL	ORF9 nuc	DR
DR_P20	INVFAFPFTIYSLLL	ORF10	DR	DR_P05	ASAFFGMSRIGMEVT	ORF9 nuc	DR
				DR_P07	LLLLDRLNQLESKMS	ORF9 nuc	DR
				DR_P15	FYVYSRVKNLNSSRV	ORF4 env	DR
				DR_P17	IWNLDYIINLIIKNL	ORF6	DR
				DR_P18	QEEVQELYSPIFLIV	ORF7	DR
				DR_P19	SKWYIRVGARKSAPL	ORF8	DR

This table depicts the selection of HLA class I and HLA-DR peptides included in the SARS CoV 2-specific and cross-reactive HLA class I and HLA-DR T cell epitope composition (EC) for standardized evaluation in SARS group 2 and PRE group B. EC, epitope composition; ID, identification number; spi, spike protein; env, envelope protein; mem, membrane protein; nuc, nucleocapsid protein.

Extended Data Fig. 6 | SARS-CoV-2-specific and cross-reactive HLA class I and HLA-DR T cell epitope compositions.

nature research

Corresponding author(s): Juliane Walz

Last updated by author(s): Sep 7, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	SYFPEITHI 1.0, NetMHCpan 4.0, PepCalc (release date 2015), Luminex FlexMap 3D instrument, Luminex xPONENT Software version 4.3, Typhoon Trio (GE Healthcare), BEP 2000 Advance® system (Siemens Healthcare Diagnostics GmbH), Cobas e411 analyzer, ImmunoSpot S5 analyzer, FACS Canto II cytometer, paper-based questionary for SARS-CoV-2 convalescent donor data.
Data analysis	GraphPad Prism 8.4.0, JMP® Pro (SAS Institute Inc., version 14.2) software, FlowJo 10.0.8 (BD).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data relating to the findings of this study are available from the corresponding author upon request. Source data are provided with this paper.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	This was a discovery project for identifying new T cell epitopes in an unpublished German population. No sample size calculation was performed. Samples sizes were based on maximal available sample sets where detailed clinical and serological data were also available. A total of n = 180 convalescent volunteer samples after SARS-CoV-2 infection (SARS collection) and n = 185 unexposed volunteer samples (PRE collection) asserted prior to SARS-CoV-2 pandemic were included. For all analyses similar numbers of SARS and PRE donors were used to allow for comparable results of post-infectious and pre-existing T cell responses. Out of the SARS and PRE collections two groups were built for (i) T cell epitope screening (SARS group 1 (n = 116), PRE group A (n = 104)) and (ii) standardized immunity evaluation (SARS group 2 (n = 86), PRE group B (n = 94)). In addition, hospitalized, severely ill COVID-19 patients (n = 21) were collected at different time points during or after infection. As SARS-CoV-2 T cell responses turned out to be very frequent in the analyzed SARS (up to 100%) and PRE (up to 81%) collection the number of recruited individuals is sufficient and allowed for the identification of multiple dominant and subdominant SARS-CoV-2 T cell epitopes.
Data exclusions	All samples with available data (as shown in Table 1 and Supplementary Tables 6,7) were included in the respective analyses.
Replication	Predicted SARS-CoV-2 peptides were tested in at least n = 19 different donors to calculate recognition frequencies of T cell epitopes. Attempts at replications were successfully conducted. Elispot were assessed as duplicates and only reliable replicates were analyzed. We analyzed several donors using expanded T cells after 12-days pre-stimulation as well as T cells without pre-stimulation. Both assays showed consistent results. Anti-SARS-CoV-2 antibodies were tested by two independent assays (EUROIMMUNE and Roche).
Randomization	No randomization was performed. Out of the SARS (n = 180) and PRE (n = 185) collections two groups were built for (i) T cell epitope screening (SARS group 1 (n = 116), PRE group A (n = 104)) and (ii) standardized immunity evaluation (SARS group 2 (n = 86), PRE group B (n = 94)). Donors were assigned to groups according to time of sample acquisition and available sample cell number. Some donors were analyzed in both groups (1 and 2 or A and B for SARS or PRE, respectively). Covariates such as age and gender were equally distributed between SARS group 1 and 2. In addition, samples from hospitalized severely ill SARS donors (n = 21) were collected for evaluations of T cell response in ex vivo assays.
Blinding	Blinding was not appropriate for this study of SARS-CoV-2 T cell responses in COVID-19 convalescent patients and unexposed individuals, with no associated therapeutic intervention.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	APC/Cy7 anti-human CD4, BioLegend, Cat# 300518, RRID:AB_314086 PE/Cy7 anti-human CD8, Beckman Coulter, Cat# 737661, RRID:AB_1575980 Pacific Blue anti-human TNF-alpha, BioLegend, Cat# 502920, RRID:AB_528965 FITC anti-human CD107a, BioLegend, Cat# 328606, RRID:AB_1186036 PE anti-human IFNγ antibody, BioLegend, Cat# 506507, RRID:AB_315440 anti-IFNγ antibody, MabTech, Cat# 3420-3-250, RRID:AB_907283 anti-IFNγ biotinylated detection antibody, MabTech, Cat# 3420-6-250, RRID:AB_907273

Penta-His antibody, Qiagen, Cat# 34660, RRID:AB_2619735
AlexaFluor647 donkey-anti-mouse antibody, Invitrogen, Cat# A-31571, RRID:AB_162542
PE-labeled goat-anti-human IgG detection antibody, Jackson Dianova, Cat# 109-116-098, RRID:AB_2337678

| Validation | All antibodies were purchased from the above stated companies. Antibodies are well described and published elsewhere. Informations can be sought from the manufactures website under catalogue number. |

Human research participants

Policy information about studies involving human research participants

Population characteristics	Donors with proven SARS-CoV-2 infection and SARS-CoV-2 unexposed individuals. Summary statistics of donors are provided in Table 1, Supplementary Table 6 and 7, and Extended Data Figure 3g.
Recruitment	Recruitment of convalescent SARS-CoV-2 donors was performed by online and paper-based calls. Samples from hospitalized COVID-19 patients were collected during or after infection. All SARS-CoV-2 donors were tested positive for SARS-CoV-2 by PCR. Using this recruitment strategy mild and moderate cases were more abundant than severe cases. The likelihood for recruitment of older patients with COVID-19 is lower with our approach. The potential bias, such as age of the SARS donors are unlikely to impact the results, as no correlation was observed between the T cell response and age of the donors. Furthermore, as our recruitment strategy mainly recruited SARS donors with mild symptoms of COVID-19, we conducted an additional analysis of T cell reponses to our SARS-CoV-2 T cell epitopes in a group of hospitalized SARS donors.
Ethics oversight	Written informed consent was obtained from all subjects. Study was approved by the ethic committee of the University Hospital Tuebingen (study code: 179/2020/BO2) and by the ethic committee of the state medical association of Hessen (MC 288/2015).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	PBMCs were incubated with 10 µg/mL of peptide, 10 µg/mL Brefeldin A (Sigma-Aldrich), and a 1:500 dilution of GolgiStop (BD) for 12 - 16 h. Staining was performed using Cytofix/Cytoperm solution (BD), APC/Cy7 anti-human CD4 (BioLegend), PE/Cy7 anti-human CD8 (Beckman Coulter), Pacific Blue anti-human TNF, FITC anti-human CD107a, and PE anti-human IFNγ monoclonal antibodies (BioLegend). PMA (5 µg/ml) and ionomycin (1 µM, Sigma-Aldrich) served as positive control. Viable cells were determined using Aqua live/dead (Invitrogen).
Instrument	FACS Canto II cytometer (BD)
Software	FlowJo software version 10.0.8 (BD)
Cell population abundance	Cells have not been enriched or sorted prior to in vitro stimulation. Culturing of peripheral blood mononuclear cells for 12 days in the presence of IL-2 results in T cell enrichment.
Gating strategy	Lymphocytes were selected in the FSC/SSC gate. Single cell were selected. Viable cells were determined using Aqua live/dead (Invitrogen). Cells expressing either CD4 or CD8 were analyzed separately for production of indicated cytokines and expression of degranulation marker.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.