SARS-CoV-2 genome-wide T cell epitope mapping reveals immunodominance and substantial CD8+ T cell activation in COVID-19 patients

Sunil Kumar Saini1, Ditte Stampe Hersby2*, Tripti Tamhane3, Helle Rus Povlsen4, Susana Patricia Amaya Hernandez1, Morten Nielsen3, Anne Ortved Gang2, Sine Reker Hadrup1*

T cells are important for effective viral clearance, elimination of virus-infected cells, and long-term disease protection. To examine the full spectrum of CD8+ T cell immunity in COVID-19, we experimentally evaluated 3141 major histocompatibility complex (MHC) class I–binding peptides covering the complete SARS-CoV-2 genome. Using DNA-barcoded peptide-MHC complex multimers combined with a T cell phenotype panel, we report a comprehensive list of 122 immunogenic and a subset of immunodominant SARS-CoV-2 T cell epitopes. Substantial CD8+ T cell recognition was observed in patients with COVID-19, with up to 27% of all CD8+ lymphocytes interacting with SARS-CoV-2–derived epitopes. Most immunogenic regions were derived from open reading frame 1 (ORF1) and ORF3, with ORF1 containing most of the immunodominant epitopes. CD8+ T cell recognition of lower affinity was also observed in healthy donors toward the SARS-CoV-2–derived epitopes. This preexisting T cell recognition signature was partially overlapping with the epitope landscape observed in patients with COVID-19 and may drive the further expansion of T cell responses to SARS-CoV-2 infection. The phenotype of the SARS-CoV-2–specific CD8+ T cells revealed a strong T cell activation in patients with COVID-19, whereas minimal T cell activation was seen in healthy individuals. We found that patients with severe disease displayed significantly larger SARS-CoV-2–specific T cell populations compared with patients with mild diseases, and these T cells displayed a robust activation profile. These results further our understanding of T cell immunity to SARS-CoV-2 infection and hypothesize that strong antigen-specific T cell responses are associated with different disease outcomes.

INTRODUCTION

The COVID-19 (coronavirus disease 2019) pandemic caused by the highly infectious SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) has challenged public health at an unprecedented scale, causing the death of more than 2 million people worldwide so far (1). T cells perform essential functions in the control and elimination of viral infections; CD8+ T cells are critical for efficient clearance of virus-infected cells, whereas CD4+ T cells are important for supporting both the CD8+ T cell response and B cell–mediated production of specific antibodies. Characteristics from the ongoing pandemic suggest that T cell recognition will be critical to mediate long-term protection against SARS-CoV-2 (2), because the antibody-mediated response seems to decline in a follow-up evaluation of convalescent patients, although it is not yet understood how this affects the risk of reinfection and what antibody levels are required for disease protection (3–5). Furthermore, studies of the closely related SARS-CoV infection show persistent memory CD8+ T cell responses even after 11 years in SARS recovered patients without B cell responses (6, 7), emphasizing the potential role of CD8+ memory T cells in long-term protection from coronaviruses.

Several recent studies have reported robust T cell immunity in SARS-CoV-2–infected patients (8–10), and unexposed healthy individuals also showed functional T cell reactivity restricted to SARS-CoV-2 (9, 11–15). The observed T cell cross-reactivity is hypothesized to derive from routine exposure to common cold coronaviruses [human coronavirus (HCoV)] (HCoV-OC43, HCoV-HKU1, HCoV-NL63, and HCoV-229E) that widely circulate, with 90% of the human population being seropositive for these viruses (16, 17) and substantial sequence homology to the SARS-CoV-2 genome (18, 19). However, the influence of such preexisting immunity to the T cell recognition associated with COVID-19 disease is poorly understood.

SARS-CoV-2 infection can result in mild to severe disease (including death), but a large number of asymptomatic infections are also described (20–22). The presence of preexisting T cell immunity, represented by cross-reactive T cells, could have strong implications for how individuals respond to SARS-CoV-2 infection. However, their biological role upon encounter with SARS-CoV-2 infection remains unclear, and their contribution to disease protection needs to be determined. Furthermore, in severe clinical disease, cytokine release syndrome is reported and might, in some cases, be dampened by immunosuppressive medication or anti–interleukin-6 (IL-6) antibody therapy (23, 24). Such clinical characteristics point to a potential uncontrolled immune response with the involvement of strong T cell activation.

CD8+ T cells are activated by a specific interaction between the T cell receptor (TCR) and the peptide antigen presented by major histocompatibility complex class I (MHC-I) molecules on the surface of virus-infected cells. Although SARS-CoV-2–specific immunity has been reported both in the context of COVID-19 and preexisting T cells, the full spectrum of exact antigens (minimal peptide epitope) within the viral genome, associated with this immunity and their immunodominance in SARS-CoV-2–infected patients, is not fully described. Using our large-scale T cell detection technology based on DNA-barcoded peptide-MHC (pMHC) multimers (25), we have mapped T cell recognition throughout the complete SARS-CoV-2
RESULTS

SARS-CoV-2–specific CD8+ T cells recognize a broad range of epitopes

To reveal the full spectrum of T cell immunity in COVID-19 disease, we used a complete SARS-CoV-2 genome sequence (26) to identify immunogenic minimal epitopes recognized by CD8+ T cells. Using NetMHCpan 4.1 (27), we selected 2204 potential human leukocyte antigen (HLA)–binding peptides (9 to 11 amino acids) for experimental evaluation. These peptides were predicted to bind one or more of 10 prevalent MHC-I molecules, including HLA-A (A01:01, A02:01, A03:01, and A24:02), HLA-B (B07:02, B08:01, and B15:01), and HLA-C (C06:02, C07:01, and C07:02) loci, leading to a total of 3141 pMHC specificities for experimental evaluation (Fig. 1A and table S1). Epitope predictions are covering the full viral genome, with open reading frame 1 (ORF1) being the largest gene region and hence including the highest number of predicted peptides (Fig. 1B). T cell reactivity toward these peptides was analyzed for 18 patients with COVID-19. In this cohort, 11 patients had severe disease requiring hospital care, and 7 patients had mild disease not requiring hospitalization. Blood samples were collected during the active phase of the infection, as close as possible after the first positive SARS-CoV-2 test (Fig. S1A) (25). Briefly, each pMHC complex is multim- erized on a PE (phycoerythrin)–labeled or APC (allophycocyanin)– labeled dextran backbone and tagged with a unique DNA barcode. DNA-barcode pMHC multimers are then pooled to generate an HLA-matching patient-tailored pMHC multimer panel, which is in- cubated with patient-derived PBMCs (peripheral blood mononuclear cells), and multimers bound to CD8+ T cells are sorted and sequenced to identify T cell recognition toward the probed pMHC complexes. For comparative evaluation, we also included 39 T cell epitopes from common viruses: cytomegalovirus (CMV), Epstein-Barr virus (EBV), and influenza (flu) virus (CEF) (Fig. 1C and table S3).

We found broad and strong SARS-CoV-2–specific CD8+ T cell responses in patients with COVID-19, contributing up to 27% of the total CD8+ T cells (Fig. 1D). A substantial selection of T cells specific to individual immunogenic epitopes measuring up to 14% of the total T cells was detected in several patients (Fig. 1D, fig. S2, and table S4). In total, we identified T cell responses to 142 pMHC complexes corres- ponding to 122 unique SARS-CoV-2 T cell epitopes across the 10 analyzed HLAAs (Fig. 1E) dominated by peptides with high-affinity binding to their corresponding HLA molecule (fig. S1B). We also de- tected 25 T cell responses to CEF-derived peptides across the 18 pa- tients with COVID-19 (Fig. 1E and table S5). For the SARS-CoV-2–derived peptides, HLA-A01:01, HLA-A02:01, and HLA-B15:01 presentation dominated in terms of the total number of identified epitopes as well as the “immunogenicity score” (i.e., the number of T cell re- sponses normalized to the number of probing pMHC multimers and the number of patients analyzed) (Fig. 1F). HLA-A03:01– and C07:01-specific peptides showed the least T cell reactivity (three epitopes each) de- spite being analyzed in nine and six patients, respectively (Fig. 1E). Most of the immunogenic epitopes were mapped to the ORF1 pro-tein, followed by S and ORF3 proteins (Fig. 1, G and H, and table S4). Given the size difference of the viral proteins, the immuno- genicity score was used to evaluate their relative contribution to T cell recognition. On the basis of such evaluation, we observe that peptides derived from ORF3 displayed the highest relative immuno- genicity (in terms of T cell recognition), followed by ORF1 protein (Fig. 1H). The overall frequency of SARS-CoV-2–reactive T cells (the sum of estimated frequencies for all SARS-CoV-2–specific T cells) in individual patients with COVID-19 showed a broad range of T cell involvement and variation in terms of T cell recognition to individual SARS-CoV-2 proteins (Fig. 1I).

In summary, we report SARS-CoV-2–specific CD8+ T cell im- munity toward several epitopes and a substantially high presence of SARS-CoV-2–specific T cells in several patients with COVID-19. The ORF1 protein not only contributes the most to T cell recog- nition of SARS-CoV-2 but is also by far the largest group of proteins. When protein size is considered, ORF3 and ORF1 are the viral re- gions most frequently recognized by CD8+ T cells.

Strong immunodominance of SARS-CoV-2–derived peptides in patients with COVID-19

Of the 122 epitopes recognized by T cells in the patient cohort, 5 were determined as “immunodominant” based on the presence of T cell recognition in >50% of the tested samples with the given HLA molecule and T cell detection identified in at least two or more patients (Fig. 2A). Unexpectedly, in our patient cohort, none of the immuno- dominant epitopes were derived from the S protein, despite this being the second largest protein (Fig. 2B). Among the immuno- dominant epitopes, a very robust HLA-associated immunodominance was observed for two of the epitopes: HLA-A01:01–TTDPSFLGRY– specific (and its variant peptides TTDPSFLGRYM and HLA-A01:01- TTDPSFLGRY), with specific T cells detected in all five analyzed patients (estimated frequency reaching up to 25% of total CD8+ T cells), and HLA-B07:02–SPRWYFYYL, with specific T cells ob- served in four of the five patients evaluated (estimated frequency up to 10%) (Fig. 2A and table S4). To validate the T cell responses iden- tified for the two most immunodominant epitopes (TTDPSFLGRY and SPRWYFYYL), we determined the presence of these T cells using conventional fluorophore-labeled pMHC tetramers in seven patients with COVID-19. For both immunodominant epitopes, the frequency of T cells determined by the individually labeled pMHC tetramers correlated to the frequencies determined based on the DNA barcode–labeled MHC multimer reagents (at a range from 0.01 to 11% of the total CD8+ T cells) (Fig. 2, C and D). Next, we evaluated the cytokine secretion capacity of the SARS-CoV-2–specific T cells by stimulating PBMCs (same time point as used for T cell identification) with respective epitopes. SARS-CoV-2–peptide–induced secretion of interferon-γ (IFN-γ) and tumor necrosis factor–α (TNF-α) was detected in all seven patients, confirming functional activation of T cells raised against dominant and nondominant epi- topes (Fig. 2E and table S6).
Representative COVID-19 patient (AP0301)

DNA barcodes associated with each of the identified on the basis of the enrichment of recognition to individual epitopes was identified on the basis of their HLA restriction across the HLA-matching donors (% normalized).

Bar plots show the distribution of SARS-CoV-2–related to their HLA-restriction (3142 peptide-HLA pairs) across the viral genome. Total pMHC specificities analyzed for each protein are shown in parentheses next to the respective SARS-CoV-2–derived peptides identified in the 18 analyzed patients with COVID-19. In parentheses, number of peptides tested for each HLA pool for each HLA (top row) and the number of all T cell recognition to SARS-CoV-2–specific T cell responses separated based on the peptide-HLA combination per patient and is colored according to their origin of protein, similar to that shown in (A). The black dots show CD8+ T cells reactive to one of the CEF peptides (here, CMV pp65; YSEHPFTFSQY–HLA-A01:01). All peptides with no significant enrichments are shown as gray dots. (E) Summary of all T cell recognition to SARS-CoV-2–derived peptides identified in the 18 analyzed patients with COVID-19. In parentheses, number of peptides tested for each HLA (top row) and the number of patients analyzed for each HLA pool (bottom row). Each dot represents one peptide-HLA combination per patient and is colored according to their origin of protein, similar to that shown in (A). The black dots show CD8+ T cells reactive to the CEF peptides in all analyzed patients. (F) Bar plots summarize the number of HLA-specific SARS-CoV-2 epitopes identified and the HLA-restricted immunogenicity (% immunogenic peptides) in the analyzed patient cohort. Immunogenicity represents the fraction of T cell–recognized peptides out of the total number of peptides analyzed for a given HLA restriction across the HLA-matching donors (% normalized). (G) Similar to (E), a summary of SARS-CoV-2–specific T cell responses separated based on the protein of origin. (H) Bar plots show the number of epitopes derived from each of the SARS-CoV-2 protein and their immunogenicity score (% immunogenic peptides). (I) Estimated frequencies (% of total CD8+ T cells) as the sum of all SARS-CoV-2 epitope–reactive T cells identified in individual patients with COVID-19. Bars are colored according to the protein origin of the recognized epitopes.
Low-avidity recognition toward SARS-CoV-2–derived peptides in healthy individuals

To examine the potential for preexisting SARS-CoV-2–reactive T cells, we next analyzed healthy individuals for T cell recognition against all 3141 SARS-CoV-2–derived peptides. We selected two healthy donor cohorts: The first cohort included SARS-CoV-2–unexposed healthy individuals (HD-1; n = 18 individuals, PBMCs collected before the COVID-19 pandemic), and the second cohort included...
health care staff at high risk of SARS-CoV-2 exposure but who did not test positive (HD-2; \( n = 20 \) individuals, PBMCs collected during COVID-19 pandemic). CD8+ T cells from SARS-CoV-2–unexposed healthy individuals showed broad-scale T cell recognition toward SARS-CoV-2–derived peptides across the whole viral genome (Fig. 3A, fig. S3, and table S7). Cumulatively, 214 SARS-CoV-2–derived peptides were recognized by T cells in 16 of the 18 analyzed samples. The high-risk COVID-19 healthy cohort showed similar T cell recognition toward 178 SARS-CoV-2–derived epitopes (Fig. 3B and table S7) in 15 of the 20 donors. T cell recognition in healthy donors was directed

---

Fig. 3. Broad reactivity toward SARS-CoV-2–derived peptides in healthy individuals. (A) CD8+ T cell recognition to individual SARS-CoV-2–derived peptides (table S7) and CEF peptides (table S5) in the pre–COVID-19 healthy donor cohort (\( n = 18 \) individuals) identified based on the enrichment of DNA barcodes associated with each of the tested peptide specificities (LogFC > 2 and \( P < 0.001 \), Barracoda). Significant SARS-CoV-2–specific T cell recognition of individual peptide sequences is colored and segregated based on their protein of origin. The black dots show CD8+ T cells reactive to the CEF peptides in all analyzed donors. (B) T cell recognition in the high-exposure risk healthy donor cohort (tables S5 and S7) (\( n = 20 \) individuals). (C) Staining index of CD8+ T cells binding SARS-CoV-2–specific pMHC multimers in the three evaluated cohorts. One-way ANOVA (Kruskal-Wallis test) **** \( P < 0.0001 \) (patient versus HD-1 < 0.0001 and patient versus HD-2 < 0.0001); \( n = 18 \) (patient), \( n = 18 \) (HD-1), and \( n = 20 \) (HD-2). (D) Flow cytometry dot plots showing in vitro expanded T cells from healthy donors recognizing SARS-CoV-2–derived epitopes, detected by combinatorial tetramer staining. T cell binding to each pMHC specificity is detected using pMHC tetramers prepared in a two-color combination (blue dots), with gray dots showing tetramer-negative T cells, and the number on the plots shows the frequency (%) of tetramer+ of the CD8+ T cells. Gating strategy used for the flow cytometry analysis is shown in fig. S11A. (E) Venn diagram illustrating the overlap of T cell recognition toward SARS-CoV-2–derived peptides in the COVID-19 patient and healthy donor cohorts.
equally toward ORF1 and S proteins, whereas ORF3-derived peptides were recognized less in the healthy donor cohort compared with the COVID-19 patient cohort (fig. S3B). The immunodominant T cell epitopes from ORF1 identified in the patient cohort were not among the most prevalent responses in the healthy donors (fig. S3C).

Despite such broad T cell recognition in both healthy donor cohorts, the presence of SARS-CoV-2–recognizing T cells seems to be of low frequency with limited separation of the CD8+ T cells binding to the pool of DNA-barcoded pMHC multimers (fig. S4A) and measured by a significantly lower staining index of the pMHC multimer binding in healthy donors compared with patients (fig. 3C). Consequently, a direct estimate of the frequency of the SARS-CoV-2–reactive T cell populations in the individual healthy donors was not feasible. The low frequency and limited separation of these T cells were confirmed by independent analysis using conventional pMHC tetramers for several individual epitopes in healthy donor PBMCs (fig. S4B). Together, these data suggest a lower TCR avidity to the probed pMHC in healthy individuals compared with patients with COVID-19, which could represent potential cross-reactivity to the probed pMHC in healthy individuals compared with patients (fig. 3C).

Consequently, a direct estimate of the frequency of the SARS-CoV-2–reactive T cell populations in the individual healthy donors was not feasible. The low frequency and limited separation of these T cells were confirmed by independent analysis using conventional pMHC tetramers for several individual epitopes in healthy donor PBMCs (fig. S4B). Together, these data suggest a lower TCR avidity to the probed pMHC in healthy individuals compared with patients with COVID-19, which could represent potential cross-reactivity to the probed pMHC in healthy individuals compared with patients (fig. 3C).

To further validate the presence of low-frequency T cells in healthy donors, we expanded T cells in vitro from several COVID-19–unexposed healthy donors and measured T cell binding using conventional pMHC tetramers. On the basis of the in vitro peptide-driven expansion, pMHC tetramer binding T cell populations were verified in multiple donors, recognizing SARS-CoV-2–derived peptides, including immunodominant epitopes across four HLAAs (A01:01-TTDPFSLRGY, A02:01-LLLLDRNLQL, A02:01-KLDCVMYA, A24:02-FYAYLRKHFN, and B07:02-SPRWYFYYL) (Fig. 3D). Although these T cell responses were of low frequency, a functional cytokine response (measured by IFN-γ and TNF-α production) was observed in in vitro expanded T cell cultures when restimulated with individual peptide epitopes or epitope pools (fig. S5). Forty-one of the COVID-19 immunogenic peptides, including the immunodominant peptides, identified in the patient cohort were also recognized by T cells of healthy donors; this includes the two most frequently observed epitopes of SARS-CoV-2: HLA-A01:01-TTDPFSLRGY and HLA-B07:02-SPRWYFYYL (Fig. 3E and table S7). Together, we show a full spectrum of T cell recognition toward SARS-CoV-2–derived peptides in healthy donors; this is detected at low frequency and shows characteristics of low-avidity interaction based on the staining index of the pMHC multimer interaction.

Enhanced activation profile of SARS-CoV-2–specific CD8+ T cells associated with COVID-19 disease severity

For phenotypic characterization of SARS-CoV-2–specific CD8+ T cells, we combined pMHC multimer analysis with a 13-parameter antibody panel (table S8) and evaluated the phenotype of the SARS-CoV-2–reactive T cell populations in patients with COVID-19 and healthy donors. This furthermore allowed us to evaluate whether the multimer-specific T cell profile of the high-risk COVID-19 healthy cohort (HD-2) has any distinct features compared with the unexposed cohort (HD-1), despite both cohorts containing presumably unexposed individuals. Dimensional reduction using Uniform Manifold Approximation and Projection (UMAP) showed distinct clustering of SARS-CoV-2 multimer-reactive T cells of the COVID-19 patient cohort compared with the two healthy donor cohorts with higher expression of activation markers CD38, CD69, CD39, HLA-DR, and CD57 and reduced expression of CD8 and CD27 (fig. S6). Compared with both healthy donor cohorts, we observed that more SARS-CoV-2–reactive T cells from patients with COVID-19 expressed the activation markers CD38, CD39, CD69, and HLA-DR and showed a late-differentiated effector memory (EM) profile of reduced CD27 (Fig. 4A). We did not observe activation of SARS-CoV-2–specific multimer+ T cells in the high-risk COVID-19 healthy cohort, except for nonsignificant trends for reduced CD27 and increased CD57 expression (Fig. 4A). SARS-CoV-2–reactive T cells in patients and healthy donor cohorts showed a similar distribution of memory subsets (determined by CCR7 and CD45RA expression); however, higher expression of T cell activation markers (fig. S7) was observed in EM and TEMRA (terminally differentiated EM) subsets in patients. Furthermore, the highly activated and differentiated T cell phenotype in patients with COVID-19 was distinct to SARS-CoV-2–specific T cells and not observed for CEF-specific T cells detected in the same cohort (fig. 4B). We also observed no difference in CEF-specific multimer+ T cells between the three cohorts in a similar analysis (fig. S8A). In addition, we compared the expression of T cell activation markers in combination with the inflammatory response marker CD38 on multimer+ CD8+ T cells across the three cohorts, which showed significantly enhanced expression of activation molecules (CD39, CD69, and HLA-DR) and PD-1 inhibitory receptor on CD38+ T cells only in the patient cohort (fig. S8, B and C).

We next evaluated the association of SARS-CoV-2–specific CD8+ T cell presence in the patient cohort related to their requirement for hospital care. No overall difference in the total number of recognized SARS-CoV-2–derived epitopes was observed between severely diseased patients requiring hospitalization (n = 11 individuals) and patients with mild symptoms not requiring hospital care (outpatient; n = 7 individuals) (Fig. 4C). For phenotype characterization, 23 additional patient samples (total, n = 41 patients; hospitalized, n = 21; outpatients, n = 20) were analyzed using a patient HLA-matching pMHC multimer library combined with the 13-parameter antibody panel, similar to the initial 18 patients but without resolving individual epitope specificities. On the basis of this extended cohort, a significantly higher frequency of SARS-CoV-2–specific CD8+ T cells was observed in the hospitalized patients compared with outpatient samples (Fig. 4D). Furthermore, a significant increase in the fraction of such cells expressing CD38, CD39, HLA-DR, and PD-1 was observed in the hospitalized patients (Fig. 4E). By measuring the coexpression of immune activation markers—CD38 together with CD39, PD-1, and HLA-DR—a strong elevation in T cells expressing these combinations of activation markers was observed among the hospitalized patients (Fig. 4F). Together, the increased frequency and activation signature suggest a role for SARS-CoV-2–specific CD8+ T cells in severe COVID-19 disease.

We also examined the phenotype of CD8+ T cells specific to the two most immunodominant epitopes TTDPFSLRGY and SPRWYFYYL with respect to disease severity (in eight patients; four hospitalized and four outpatients) using conventional pMHC tetramer–based evaluation of individual T cell specificities. Hospitalized patients displayed increased PD-1 expression compared with the same T cell populations in the outpatients (fig. S9A). Furthermore, a higher frequency of T cells reactive to these two SARS-CoV-2 immunodominant epitopes was observed in the hospitalized patients, but the functional evaluation upon peptide stimulation revealed that only a subtraction
Fig. 4. Enhanced activation profile of SARS-CoV-2–specific T cells correlates with COVID-19 disease severity. (A) Box plots comparing percentages of SARS-CoV-2 pMHC multimer binding CD8+ T cells expressing the indicated phenotype surface markers in the COVID-19 patient and the two healthy donor cohorts (n = 18 individuals for each cohort). Each dot represents one sample. Frequencies were quantified from flow cytometry data processed using the gating strategy applied in fig. S11. P values for one-way ANOVA (Kruskal-Wallis test): CD38 < 0.0001 (HD-1 versus patient < 0.0001 and HD-2 versus patient < 0.0001), CD39 < 0.0001 (HD-1 versus patient = 0.006 and HD-2 versus patient < 0.0001), CD69 < 0.0001 (HD-1 versus patient < 0.0001 and HD-2 versus patient < 0.0001), HLA-DR = 0.002 (HD-1 versus patient = 0.02 and HD-2 versus patient < 0.004), and CD27 = 0.03 (HD-1 versus patient = 0.03). (B) Box plots comparing the percentage of SARS-CoV-2 pMHC multimer+ (n = 18 patients) and CEF pMHC multimer+ (n = 14 patients) CD8+ T cells expressing the indicated surface markers in the COVID-19 patient cohort. Each dot represents one sample. P values for hypothesis (Mann-Whitney) test: P = 0.0002 (CD38), P < 0.0001 (CD39), P = 0.0001 (CD69), P = 0.009 (HLA-DR), and P = 0.04 (CD27). (C) Number of SARS-CoV-2 epitopes recognized by T cells in outpatient (n = 7) and hospitalized (n = 11) patient samples. (D) Box plots show frequencies of SARS-CoV-2 pMHC multimer+ CD8+ T cells in outpatient (n = 20) and hospitalized patients (n = 21). P value (Mann-Whitney test) of ≤0.0001. (E) Box plots showing the percentage of SARS-CoV-2 pMHC multimer+ CD8+ T cells expressing the indicated surface markers in outpatients (n = 20) and hospitalized patients (n = 21). Each dot represents one sample. P values for hypothesis (Mann-Whitney) test: P = 0.0001 (CD39), P = 0.0001 (PD-1), and P = 0.027 (HLA-DR). (F) Comparison of the frequency of SARS-CoV-2 pMHC multimer+ CD8+ T cells expressing activation markers (CD39 and HLA-DR) and PD-1 in combination with CD38 [as shown in the representative plots (fig. S8), in hospitalized and outpatient samples]. P values for hypothesis (Mann-Whitney) testing: P = 0.04 (CD38 + CD39), P = 0.005 (CD38 + HLA-DR), and P < 0.0001 (CD38 + PD-1). (G) Comparison of tetramer binding (conventional single-color tetramers) and functional (cytokine-secreting) T cells recognizing the two immunodominant epitopes in 10 patients, grouped according to COVID-19 disease severity. *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001, and ****P ≤ 0.0001.
of these high-frequency T cells were responsive to antigen exposure (Fig. 4G and fig. S9B). These data, together with increased PD-1 expression, suggest a functional impairment or selective inhibition of these high-frequency T cell populations, as observed by a recent study (5).

A fraction of SARS-CoV-2 epitopes share sequence homology with widely circulating common cold coronaviruses
Preexisting T cell immunity, in the context of SARS-CoV-2–reactive T cells in unexposed healthy individuals, has been reported by several studies (13–15, 17, 19), and it has been hypothesized that this is due to the shared sequence homology between the SARS-CoV-2 genome and other common cold coronaviruses (HCoV-OC43, HCoV-HKU1, HCoV-NL63, and HCoV-229E). Having evaluated the full spectrum of minimal epitopes for T cell recognition, we sought to evaluate the sequence homology at the peptide level and its association with the SARS-CoV-2 T cell reactivity that we observed in healthy donors. First, we searched for immunogenic hotspots across the full SARS-CoV-2 proteome by comparing the number of identified epitopes (in the patient cohort) with the total number of predicted peptides in any given region of the proteins. In general, the epitopes were spread over the full length of the protein sequences while clustering in minor groups throughout all regions of the viral proteome (Fig. 5A). Regions indicated by an asterisk demonstrate significant enrichment of T cell recognition relative to the number of MHC-I–binding peptides in a given region. Both the C- and N-terminal regions of the ORF1 seem to hold fewer T cell epitopes compared with the rest of this protein. When similarly mapping the T cell recognition of SARS-CoV-2–derived peptides observed in healthy donors, we detected a comparable spread of T cell recognition in the healthy donor cohort. Most T cell epitope clusters in the patient cohort coincide with T cell recognition in the healthy donor cohort. The few regions that distinguish the T cell recognition observed in healthy donors from that observed in patients include the C- and N-terminal regions of ORF1, parts of the N, and, in general, a higher level of T cell recognition to S. In these regions, T cell recognition in healthy donors exceeded the observation from patients with COVID-19 (Fig. 5A). When evaluating the prevalence of T cell recognition for the epitopes identified in >25% of the patient (Fig. 2A and table S9) or the healthy donor cohort (fig. S3C and table S9), we observed that most of these T cell responses frequently observed in patients with COVID-19 are also detected in healthy donors, whereas a large fraction of epitopes dominating in healthy individuals were not detected in our patient cohort (Fig. 5B). However, several SARS-CoV-2 reactivities that were identified only in the healthy donors in our study were shown to be present in patients with COVID-19 analyzed by other studies (table S9), which strongly points to a substantial degree of cross-recognition to SARS-CoV-2 from preexisting T cell populations and that such populations might drive the further expansion of T cell responses to SARS-CoV-2 infection.

To further elucidate the potential origin of such a cross-reactive T cell population in the healthy donor cohort, we next evaluated the sequence homology of SARS-CoV-2 MHC-I–binding peptides with the four common cold coronaviruses: HCoV-HKU1, HCoV-NL63, HCoV-OC43, and HCoV-229E. With a variation limit of up to two amino acids in each peptide sequence, 15% of the total predicted peptides showed sequence similarity with one or more HCoV peptide sequence (Fig. 5C, gray pie). Among the T cell–recognized peptides, in both the patient and healthy donor cohorts, this fraction was comparable with 19 and 16%, respectively, of T cell–recognized peptides sharing sequence homology with one or more HCoV (Fig. 5C). As an alternative approach, the similarities were calculated by kernel method for amino acid sequences using BLOSUM62, indicating comparable sequence similarity of the peptides recognized by T cells and those not recognized in reference to HCoV. However, peptides with the lowest similarity to HCoV were not recognized by T cells in the patient cohort (fig. S10).

Because T cell cross-recognition can often be driven by a few key interaction points, predominantly in the “core” of the peptide sequence (i.e., positions 3 to 8) (28, 29), we restricted the sequence similarity to the core of the peptide that would be most likely to interact with the TCR (30). On the basis of the protein core only, up to 74% of all the identified epitopes showed sequence homology to HCoV (one or more) (Fig. 5C), suggesting these common cold viruses as a potential source of the observed low-avidity interactions in healthy donors. Furthermore, when evaluating peptides frequently recognized by T cells in both patients with COVID-19 and healthy individuals, we find evidence of substantial homology, as exemplified with the peptide sequences listed in Fig. 5D. However, similar sequence homology is observed for the peptide sequences that are recognized only in the patient cohort (Fig. 5D). Thus, at present, our data point to substantial T cell cross-recognition being involved in shaping the T cell response to SARS-CoV-2 in patients with COVID-19; however, we find no specific enrichment of T cell recognition to peptide sequences with large sequence homology compared with the total peptide library being evaluated, and the responses identified exclusively in the patient samples are not more specific to SARS-CoV-2 compared with those recognized in both cohorts. ORF1 displayed the highest T cell recognition immunogenicity and also the highest sequence identity to HCoV (40%, as opposed to 22 to 34% for all other SARS-CoV-2 proteins, calculated using direct sequence alignment). Future studies seeking to fully understand the role and origin of the underlying T cell cross-recognition will likely require an in-depth evaluation of pre- and postinfection samples.

DISCUSSION
Several studies using overlapping peptide pools spanning different regions of SARS-CoV-2 viral proteins have shown a broad range of T cell activation in convalescent COVID-19 patients (8, 9, 11, 14, 15, 31–35). Our work now provides a detailed characterization of minimal epitopes derived from the complete SARS-CoV-2 genome for their CD8 T cell immunogenicity, immunodominance, and functional and phenotypical characteristics in patients with COVID-19 and healthy donors. We identified CD8 T cell responses to 122 epitopes in 18 patients with COVID-19 after screening for T cell recognition based on 3141 peptides derived from the full SARS-CoV-2 genome and selected based on their predicted HLA-binding capacity. Of these, a few immunodominant T cell epitopes were recognized in most of the patients. Both dominant and subdominant T cell epitopes were cross-recognized by low-level preexisting T cell populations in SARS-CoV-2–unexposed healthy individuals. We have observed that the SARS-CoV-2 dominant epitopes mount very strong T cell responses, with up to 27% of all CD8 T lymphocytes recognizing a single epitope (two overlapping peptides with the same peptide core).

Initial analysis of SARS-CoV-2–unexposed individuals revealed substantial presence of CD4+ and CD8+ T cells cross-reactive to SARS-CoV-2 peptides (11, 13, 15, 17, 19, 36, 37). Longitudinal analysis
Fig. 5. A fraction of SARS-CoV-2 epitopes share sequence homology with widely circulating common cold coronaviruses. (A) SARS-CoV-2 T cell immunogenicity map across the viral proteome comparing the distribution of identified SARS-CoV-2 epitopes (patient cohort, orange line; n = 16 patients) with the total analyzed peptides (gray line). The height of a peak indicates the number of ligands (right y axis) analyzed in a particular region and the number of identified epitopes (left y axis). The bottom panel similarly maps epitopes and ligands from healthy donors (green line, n = 31 individuals). Positions significantly enriched (P < 0.05) with epitopes compared with the number of tested ligands are marked with an asterisk. (B) T cell epitopes selected on the basis of their immunodominant characteristics either in the patient (orange) or healthy donor (green) cohort or represented in both (red) are evaluated for their T cell recognition prevalence in both cohorts. (C) Sequence similarity of SARS-CoV-2 peptides with the other four common cold coronaviruses (HCoV) HCoV-HKU1, HCoV-NL63, and HCoV-229E. The gray pie chart indicates the sequence similarity of the total predicted peptides from SARS-CoV-2 with any one (+1), two (+2), three (+3), or all four (+4) HCoV peptides with a variation limit of up to two amino acids within the full-length peptide. The colored pie chart shows a similar analysis for epitopes detected in the patient (n = 16) or healthy donor cohort (combined analysis of HD-1 and HD-2, n = 31) for full-length peptide and peptide core. (D) Examples of sequence homology for shared (between patient and healthy donors) and patient-specific T cell epitopes with one or more HCoV peptide sequence. Nonmatching amino acids are shown in gray.

of cross-reactive and induced CD8+ T cells before and after SARS-CoV-2 infection has been followed in individual cases (37), but the role of preexisting T cells in overall immune response and disease outcome is not yet established. Using a genome-wide screen of expanded T cells, a recent study reported cross-reactivity to SARS-CoV epitopes in patients with COVID-19 but not to other commonly circulating coronaviruses (38). Our ex vivo evaluation of all 3141 SARS-CoV-2–derived minimal epitopes in two healthy cohorts (COVID-19–unexposed and high risk) shows extensive but low-frequency and low-avidity interaction with CD8+ T cells. Preexisting immunity based on cross-reactive T cells can influence how the immune system reacts upon viral exposure, including through faster expansion of preexisting memory cells upon initial exposure to viral infection. A similar outcome and benefit of preexisting T cell immunity have been shown in the case of the flu pandemic virus H1N1 (39, 40). However, active stimulation of cross-reactive T cells could also lead to exhaustion of rapidly expanded T cells, similar to the higher PD-1 expression and reduced cytokine secretion of the SARS-CoV-2 immunodominant T cells observed by us and others (5, 41, 42). In addition, hyperactivation of preexisting T cells could contribute to short- and long-term disease...
severity via inflammation and autoimmunity, because increased production of IFN-γ by CD4+ and CD8+ T cells has been observed in patients with severe COVID-19 (43). Furthermore, it has been reported (44) that SARS-CoV-2 infection can be a triggering factor for autoimmune reactions and severe pneumonia with sepsis leading to acute respiratory distress syndrome, bone marrow infection with pancytopenia, and organ-specific autoimmunity (45–47). Preexisting T cell immunity can influence vaccination outcomes, because they may induce a faster but possibly selective immune response. The ORF1 protein regions are highly conserved within coronaviruses (48) and show the highest HCoV identity among SARS-CoV-2 proteins, and most of the immunodominant epitopes that we have identified belong to the ORF1 region. Thus, a detailed evaluation of these T cell epitopes could be of value in vaccine design.

Most vaccine development efforts are currently focusing on mounting antibody responses to the spike protein, with limited focus on T cell immunity. This is due to the receptor binding domain being the main target for neutralizing antibodies produced by B cells (49). However, several studies have pointed out relatively low antibody titers in COVID-19 recovered patients (3, 50–52). In conditions where antibody titers cannot sufficiently protect against infections, T cell immunity may sustain the antibody responses and provide a direct source of T cells for clearing virus–infected cells. For the involvement of T cell immunity in vaccine development, our data suggest that the inclusion of other virus proteins, such as ORF1 or ORF3, might be highly relevant. For now, the role of antibody- and T cell–mediated immune response after natural infection or after vaccination is not yet resolved and requires extensive longitudinal analysis comparing antibody and T cell kinetics to determine a synergistic or specific effect in long-term disease protection.

T cell recognition of SARS-CoV-2–derived peptides in both patients with COVID-19 and healthy donors has prompted us to understand the role of T cell cross-reactivity in controlling infections. In recent years, technology improvements in TCR characterization have allowed us to interrogate the TCR-pMHC interaction from a structural approach while obtaining experimental information related to the peptide amino acids that are crucial to T cell recognition (53–58). Such efforts have taught us that T cell cross-recognition is very difficult to predict, without knowing the precise interaction required for the given TCR, because even T cell epitopes with as low as 40% sequence homology can be recognized by a given TCR (30). Therefore, the underlining source of T cell cross-reactivity might arise from a larger set of epitopes within the HCoV viruses, including sequences with larger variation than those evaluated here (i.e., maximum of two amino acid variants per peptide sequence/peptide core).

Although T cell recognition itself was largely overlapping in identity between patients and healthy donors, the magnitude of T cell responses and particularly the phenotype of SARS-CoV-2–specific T cells were substantially different. We detected a strong activation profile of SARS-CoV-2–specific T cells only in patients with COVID-19, and this strong “activation signature” (high expression of CD38, CD39, PD-1, and HLA-DR) was further enhanced in patients requiring hospitalization. Such highly activated T cell responses should facilitate viral clearance, and hence, our data further support the notion that some severely affected patients might suffer from hyperactivation of their T cell compartment as a consequence of their primary viral infection, which may even be cleared. Additional signs of functional impairment were observed, and cytokine secretion upon antigen stimulation was incomplete for the high-frequency populations of SARS-CoV-2–specific T cells.

A limitation of the current study relates to the lack of information related to the precise date of infection. This may differ by up to 1 week, because symptoms and hence diagnosis can be delayed. Consequently, differences in T cell mobilization and/or activation may be observed as a function of time, which is not controlled in the present study. However, a measurement of symptoms before the first positive SARS-CoV-2 test indicates that samples were collected at about the same time relative to symptom onset in the two groups of patients, except for three patients from the intensive care unit included later after infection. In addition, although our T cell screening strategy allows for high-throughput epitope mapping, determination of individual responses can only be estimated following the barcode deconvolution strategy, in relation to the pool of pMHC multimer T cells upon sorting. For the healthy donor population, the separation was insufficient to precisely determine the frequency of this T cell population, whereas for the patient cohort, both measurements demonstrated strong correlation with measurements of the individual responses using conventional pMHC tetramers.

Together, COVID-19 disease drives substantial T cell activation, with T cell recognition of a large number of SARS-CoV-2–derived peptides. There is also considerable T cell recognition of such peptides in healthy donors, arguing for cross-recognition, potentially from T cells raised against other coronaviruses. The activation profile clearly distinguishes patients from healthy individuals. Patients who required hospitalization for COVID-19 demonstrated a significantly higher frequency of SARS-CoV-2–specific T cells and a more activated phenotype compared with patients with milder disease. The data presented here support a role for T cell recognition in COVID-19 and hypothesize that such T cells are associated with COVID-19 disease severity. Preexisting T cell immunity likely influences the immune response to SARS-CoV-2, which could be leveraged to fight novel infections.

MATERIALS AND METHODS

Study design

This study aimed to identify a full repertoire of CD8+ T cell–mediated immune response to SARS-CoV-2 infection. For a comprehensive evaluation, we determined potential T cell epitopes within the complete SARS-CoV-2 genome and analyzed the resulting 3141 peptides for their T cell recognition, immunodominance, breadth of the T cell response, functional and phenotype of reactive T cells, and contribution in COVID-19 disease severity. We used a DNA barcode–based MHC multimer T cell detection technology in combination with a 13-parameter flow cytometry phenotyping panel for T cell identification in PBMCs in a cohort of 18 patients with COVID-19 (composed of severe and mild disease) and compared with T cell recognition in two healthy donor cohorts (18 COVID-19–unexposed individuals and 20 high-risk health care staff). To understand the association of SARS-CoV-2–specific T cells in disease severity, we included an additional 23 patients for T cell phenotype analysis.

Clinical samples

Approval for the study design and sample collection was obtained from the Committee on Health Research Ethics in the Capital Region of Denmark. All included patients and health care employees gave their informed written consent for inclusion. PBMC samples
from 18 SARS-CoV-2–infected patients were used in this study. Blood samples were collected as close as possible to the first COVID-19–positive test. The patient cohort included samples from individuals with severe symptoms who required hospital care (hospitalized; n = 11) and patients with mild symptoms not requiring hospital care (outpatient; n = 7). For hospitalized patients, we collected full data from the medical record regarding disease course, age, gender, travel history, performance status, symptoms, comorbidity, medications, laboratory findings, diagnostic imaging, treatment, need of oxygen, need for intensive care, and an overall estimate of disease severity (table S2). For outpatients, we used a questionnaire to collect data on comorbidity, travel history, medications, and performance status.

SARS-CoV-2 infection was diagnosed by one of four platforms as follows: BGI (BGI COVID-19 RT-PCR kit), Panther Fusion (Hologic), Roche Flow (Roche MagNA Pure 96 and Roche LightCycler 480 II real-time PCR), and Qiiaflow (QIAsymphony or RotorGene, Qiagen). In the last three platforms, LightMix Modular SARS-CoV (COVID-19) E-gene (#53-0776-96) has been used. The diversity of platforms used was due to supply issues. All platforms were validated using validation kits and panels from the Statens Serum Institute, Denmark. Most patients had more than one positive test for COVID-19. Swabs, sputum, and tracheal secretion were used depending on the setting.

Patients were attempted for inclusion soon after diagnosis. The samples were collected within 2 weeks of COVID-19 diagnosis (except for three patients who were at intensive care after diagnosis). The average number of days with symptoms before sample collection matches closely in the two patient cohorts (10.85 days for the hospitalized group and 10.45 days for the outpatient group) (table S2); however, it was not possible to determine the exact date of infection.

For the pre–COVID-19 healthy donor cohort (n = 18), we used samples collected before October 2019 and obtained from the central blood bank, Rigshospitalet, Copenhagen, in an anonymized form. In addition, we included 20 health care employees from Herlev Hospital during the COVID-19 pandemic, who were at high risk of SARS-CoV-2 infection but not detected to be positive, as a cohort to follow immune responses in a potentially exposed population. PBMCs from all three cohorts were isolated immediately after sampling using Ficoll-Paque PLUS (GE Healthcare) density gradient centrifugation and were cryopreserved thereafter at a density of 2 × 10⁶ to 20 × 10⁶ cells/ml.

SARS-CoV-2 peptide selection
Potential HLA class I–binding peptides were predicted from the complete set of 8- to 11-mer peptides contained within the Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1 (GenBank ID: MN908947.3) to a set of 10 prevalent and functionally diverse HLA molecules (HLA-A01:01, HLA-A02:01, HLA-A03:01, HLA-A24:02, HLA-B07:02, HLA-B08:01, HLA-B15:01, HLA-C06:02, HLA-C07:01, and HLA-C07:02) using a preliminary version of NetMHCpan 4.1 (www.cbs.dtu.dk/services/NetMHCpan/index_v0.php) (PMID: 32406916). For peptides predicted from ORF1 protein, a percenttile rank binding threshold of 0.5% was used, and for peptides derived from all other proteins, a threshold of 1% was used. Together, 2203 peptides were selected, binding to one or more HLA molecules, generating 3141 peptide–HLA pairs for experimental evaluation (table S1). All peptides were custom-synthesized by Pepscan Presto BV, Lelystad, The Netherlands. Peptide synthesis was done at a 2-μmol scale with ultraviolet (UV) and mass spectrometry quality control analysis for 5% random peptides with an estimated purity of 70 to 92% by the supplier.

MHC-I monomer production
All 10 MHC-I monomer types were produced using methods previously described (59). Briefly, MHC-I heavy chain and human β2-microglobulin (β2m) were expressed in Escherichia coli using pET series expression plasmids. Soluble denatured proteins of the heavy chain and β2m were harvested using inclusion body preparation. The folding of these molecules was initiated in the presence of UV-labile HLA-specific peptide ligands (60). HLA-A02:01 and A24:02 molecules were folded and purified empty, as described previously (61). Folded MHC-I molecules were biotinylated using the BirA biotin-protein ligase standard reaction kit (Avidity LLC, Aurora, CO), and MHC-I monomers were purified using size exclusion chromatography (HPLC, Waters Corporation, USA). All MHC-I folded monomers were quality-controlled for their concentration, UV degradation, and biotinylation efficiency and stored at −80°C until further use.

DNA-barcoded multimer library preparation
The DNA-barcoded multimer library was prepared using the method developed by Bentzen et al. (25). Unique barcodes were generated by combining different A and B oligos, with each barcode representing a 5’ biotinylated unique DNA sequence. These barcodes were attached to PE or APC and streptavidin-conjugated dextran (Fina Biosolutions, Rockville, MD, USA) by incubating them at 4°C for 30 min to generate a DNA barcode dextran library of 1325 unique barcode specificities. SARS-CoV-2 pMHC libraries were generated by incubating 200 μM peptide of each peptide with 100 μg/ml of the respective MHC molecules for 1 hour using UV-mediated peptide exchange (HLA-A01:01, A03:01, B07:02, B08:01, B15:01, C06:02, C07:01, and C07:02) or direct binding to empty MHC molecules (HLA-A02:01 and A24:02). HLA-specific DNA-barcoded multimer libraries were then generated by incubating pMHC monomers to their corresponding DNA barcode–labeled dextrans at 4°C for 30 min, thus providing a DNA barcode–labeled pMHC multimer specifically to probe the respective T cell population. A similar process was followed to generate DNA-barcoded pMHC multimers for CEF epitopes (HLA-A and HLA-B) using APC- and streptavidin-conjugated dextran attached with unique barcodes.

T cell staining with DNA-barcoded pMHC multimers and phenotype panel
All COVID-19 patient and healthy donor samples were HLA-genotyped for HLA-A, HLA-B, and HLA-C loci (next-generation sequencing; IMGM Laboratories GmbH, Germany) (table S10). Patient and healthy donor HLA-matching SARS-CoV-2 and CEF pMHC multimer libraries were pooled [as described previously (25)] and incubated with 5 × 10⁶ to 10 × 10⁶ PBMCs [thawed and washed twice in RPMI and 10% fetal calf serum (FCS) and washed once in barcode cytometry buffer] for 15 min at 37°C at a final volume of 60 μl. Cells were then mixed with 40 μl of phenotype panel containing surface marker antibodies (table S8) and a dead cell marker (LIVE/DEAD Fixable Near-IR; Invitrogen, L10119) (final dilution 1/1000) and incubated at 4°C for 30 min. Cells were washed twice with barcode cytometry buffer and fixed in 1% paraformaldehyde.

Cells fixed after staining with pMHC multimers were acquired on a FACSAria flow cytometer instrument (AriaFusion, Becton
Dickinson) and gated by the FACSDiva acquisition program (Becton Dickinson), and all the PE-positive (SARS-CoV-2 multimer binding) and APC-positive (CEM multimer binding) cells of CD8⁺ gate were sorted into presaturated tubes (2% bovine serum albumin and 100 μl of barcode cytometry buffer) (fig. S11A). Sorted cells belonging to each sample were then subjected to polymerase chain reaction (PCR) amplification of its associated DNA barcode(s). Cells were centrifuged for 10 min at 5000g, and the supernatant was discarded with minimal residual volume. The remaining pellet was used as the PCR template for each of the sorted samples and amplified using the Taq PCR Master Mix Kit (Qiagen, 201443) and the sample-specific forward primer (serving as sample identifier) A-key36. PCR-amplified DNA barcodes were purified using the QIAquick PCR Purification kit (Qiagen, 28104) and sequenced at PrimBio (USA) using the Ion Torrent PGM 314 or 316 chip (Life Technologies).

DNA barcode sequence analysis and identification of pMHC specificities
To process the sequencing data and automatically identify the barcode sequences, we designed a specific software package, "Barracoda" (https://services.healthtech.dtu.dk/service.php?Barracoda-1.8). This software tool identifies the barcodes used in a given experiment, assigns sample ID and pMHC specificity to each barcode, and calculates the total number of reads and clonally reduced reads for each pMHC-associated DNA barcode. Furthermore, it includes statistical processing of the data. Details are given in (25). The analysis of barcode enrichment was based on methods designed for the analysis of RNA sequencing data and was implemented in the R package edgeR. Fold changes in read counts mapped to a given sample relative to mean read counts mapped to replicate baseline samples were estimated using normalization factors determined by the trimmed mean of M values. P values were calculated by comparing each experiment individually to the mean baseline sample reads using a negative binomial distribution with a fixed dispersion parameter set to 0.1 (25). False discovery rates (FDRs) were estimated using the Benjamini-Hochberg method. Specific barcodes with FDR < 0.1% were defined as significant, determining T cell recognition in the given sample. At least 1 per 1000 reads associated with a given DNA barcode relative to the total number of DNA barcode reads in that given sample was set as the threshold to avoid false-positive detection of T cell populations due to the low number of reads in the baseline samples. T cell frequency associated with each significantly enriched barcode was measured on the basis of the percentage read count of the associated barcode out of the total percentage multimer⁺ CD8⁺ T cell population in patient samples. In healthy donors, T cell recognition was identified on the basis of barcode enrichment analysis, the same as in patient samples; however, a frequency estimate of the corresponding T cell populations was not determined for significant responses identified in healthy donors because of insufficient separation of multimer⁺ cells. To exclude potential pMHC elements binding to T cells in a nonspecific fashion, non–HLA-matching healthy donor material was included as a negative control. Any T cell recognition determined in these samples was subtracted from the full dataset.

T cell functional analysis
For functional evaluation of T cells from PBMCs of patients with COVID-19 or PBMCs expanded from healthy donors, 1 × 10⁶ to 2 × 10⁶ cells were incubated with 1 μM SARS-CoV-2 minimal epitope or pool of up to 10 epitopes (1 μM per peptide) for 9 hours at 37°C in the presence of protein transport inhibitor (final dilution 1/1000; GolgiPlug; BD Biosciences, 555029). Functional activation of T cells was measured using intracellular cytokines IFN-γ (final dilution 1/20; BD Biosciences, 341117) and TNF-α (final dilution 1/20; BioLegend, 502930). Cells incubated with Leukocyte Activation Cocktail (final dilution 1/500; BD Biosciences, 550583) were used as a positive control, and HLA-specific irrelevant peptides were used as negative controls. Surface marker antibodies CD3-FITC (final dilution 1/20; BD Biosciences, 345764), CD4-APC (final dilution 1/20; BD Biosciences, 345764), CD8-PerCP (final dilution 1/20; BD Biosciences, 345764), and CD4-PE (final dilution 1/20; BD Biosciences, 345764) were used as negative controls. Surface marker antibodies CD3-FITC (final dilution 1/20; BD Biosciences, 345764), CD4-APC (final dilution 1/20; BD Biosciences, 345764), CD8-PerCP (final dilution 1/20; BD Biosciences, 345764), and CD4-PE (final dilution 1/20; BD Biosciences, 345764) were used as negative controls.
Flow cytometry analysis

For phenotype analysis, all samples were analyzed using FlowJo data analysis software (FlowJo LLC). Frequencies of specific cell populations were calculated according to the gating strategy shown in fig. S11B. For combinatorial tetramer staining, T cell binding to specific pMHC tetramers was identified using the gating plan described in the original study (63). For UMAP analysis (64), FCS (Flow Cytometry Standard) files of samples from the patient and healthy cohorts were concatenated (160,000 total cells), downsampled (FlowJo plugin), and visualized using UMAP (version 2.2, FlowJo plugin) analysis based on the following selected markers: CD3, CD4, CD8, CD38, CD39, CD69, CD137, HLA-DR, PD-1, CCR7, CD45RA, CD27, and CD57.

Sequence homology analyses

To evaluate the homology between SARS-CoV-2 and HCoV, both epitopes (peptides recognized by T cells) and ligands (peptide not recognized by T cells) were mapped to their respective source protein from the SARS-CoV-2 proteome. Enrichment analysis of the epitopes in the given region of the proteins was based on testing whether the count of observed epitopes exceeded what we expected from the number of ligands tested at each position. Epitopes were considered successes, and the count of ligands was regarded as the number of trials in a binomial test. The probability of success was derived from the average ratio of epitope to ligand per protein across each protein. The test was “one-sided” with a significance level at 0.05.

The similarity of SARS-CoV-2 ligands and epitopes from both patient and healthy donor cohorts to a set of human common cold corona viruses (HCoV-HKU1, HCoV-229E, HCoV-NL63, and HCoV-OC43) was tested using two methods. The first approach used a kernel method for amino acid sequences using BLOSUM62 (65). The second approach was a simple string search allowing up to two mismatches. The similarity of SARS-CoV-2 ligands and epitopes from both SARS-CoV-2 in SARS-recovered patients.

Data processing and statistics

T cell recognition was determined on the basis of the DNA-barcoded pMHC multimer analysis and evaluated through Barraconda (see above). The data were plotted using Python 3.7.4. For all plots, peptide sequences with no significant enrichments are shown as gray dots, and all peptides with a negative enrichment are set to LogFc equal zero (Figs. 1, D, E, and G, and 3 and A, B, and Fig. S2). Box plots for data quantification and visualization were generated, and their related statistical analyses were performed using GraphPad Prism (GraphPad Software Inc.) (Figs. 3C and 4, A to F, and figs. S1, A and B, S7B, S8, B and C, and S9A) or R studio (fig. S10). For unpaired comparisons, Mann-Whitney test was applied, and to compare more than two groups, one-way analysis of variance (ANOVA) (Kruskal-Wallis) test was performed using GraphPad Prism. All P values are indicated in the figure legends. Flow cytometry data were analyzed using FlowJo (version 10). Immunogenicity scores (Fig. 1, F and H, and fig. S3) were calculated (as percentage) by dividing the total identified T cell reactivity associated with an HLA or protein with the total number of specificities analyzed in a given cohort (number of peptides multiplied by the number of patient with a given HLA). Staining index (Fig. 3C) was calculated as follows: [mean fluorescence intensity (MFI) of multimer+ cells − MFI of multimer− cells]/(2 × SD of multimer− cells). MFI of multimer+ and multimer− CD8+ T cells and the SD of the multimer CD8+ T cells are from FlowJo analysis for patient and healthy donor samples.

SUPPLEMENTARY MATERIALS

immunology.sciencemag.org/cgi/content/full/6/58/eabf7550/DC1

Tables S1 to S11

Fig. S1 to S11

View/request a protocol for this paper from Bio-protocol.

REFERENCES AND NOTES


Acknowledgments: We thank all patients and healthy donors for participating in and contributing the analyzed samples; B. Rotbel, A. G. Burlak, A. F. Laye, and P. T. Petersen for excellent technical support; the clinical research unit for patient inclusion—K. F. Kokholm, M. L. Sieg, and J. Kock; the Department of Microbiology, Herlev Hospital, L. Nielsen, and all the collaborators for active participation to this work. Funding: This work is supported by the Independent Research Fund Denmark (grant no. 0214-00053B, 2020) to S.R.H. Author contributions: S.K.S. conceived the idea, designed and performed experiments, analyzed data, prepared figures, and wrote the manuscript. D.S.H. designed the research, performed experiments, and analyzed data. H.R.P. analyzed data, performed bioinformatic analysis, prepared figures, and wrote the manuscript. S.P.A.H. performed the experiments. M.N. supervised and performed the research and wrote the manuscript. A.O.G. conceived the idea; supervised the clinical study, patient participation, and data and sample collection; and wrote the manuscript. S.R.H. conceived the idea, designed the research, supervised the research, performed experiments, and analyzed data. H.R.P. analyzed data, performed bioinformatic analysis, prepared figures, and wrote the manuscript. S.P.A.H. performed the experiments. M.N. supervised and performed the research and wrote the manuscript. A.O.G. conceived the idea; supervised the clinical study, patient participation, and data and sample collection; and wrote the manuscript. S.R.H. conceived the idea, designed the research, supervised the research, performed experiments, and analyzed data. H.R.P. analyzed data, performed bioinformatic analysis, prepared figures, and wrote the manuscript. S.P.A.H. performed the experiments. M.N. supervised and performed the research and wrote the manuscript. A.O.G. conceived the idea; supervised the clinical study, patient participation, and data and sample collection; and wrote the manuscript.

Data and materials availability: All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials. The raw sequencing data can be accessed from the corresponding author upon reasonable request. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party: obtain authorization from the rights holder before using such material.
Mapping SARS-CoV-2 T cell recognition

Cellular immunity mediated by cytotoxic CD8+ T cells contributes to protection against viral infection, but the full spectrum of SARS-CoV-2 T cell recognition and role of preexisting T cell immunity remain incompletely understood. Saini et al. used DNA-barcoded peptide–MHC-I multimers to scan the SARS-CoV-2 genome for CD8+ T cell recognition in patients with COVID-19. Across 10 analyzed HLA molecules, 122 unique SARS-CoV-2 CD8+ T cell epitopes were detected, including 5 immunodominant epitopes primarily concentrated within ORF1. Healthy donors displayed broad T cell recognition of lower affinity and shared epitopes could be partially attributed to homology with seasonal human coronaviruses. The frequency and activation of SARS-CoV-2–specific CD8+ T cells were increased during severe compared with mild disease, highlighting differences in T cell responses associated with disease progression.