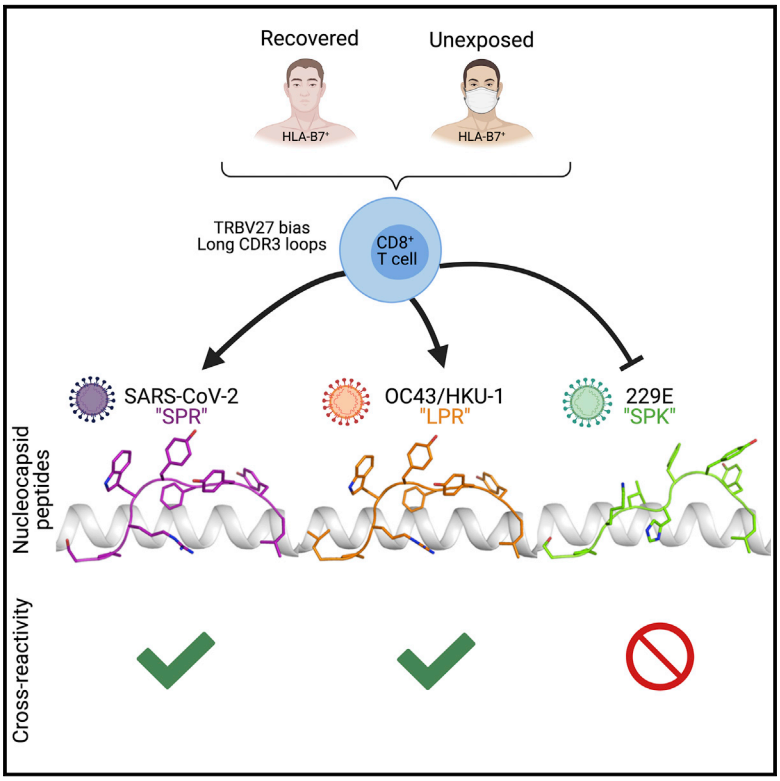
Article

Immunity

CD8+ T cells specific for an immunodominant SARS-CoV-2 nucleocapsid epitope cross-react with selective seasonal coronaviruses

Graphical abstract



Highlights

- ICS assay detects SARS-CoV-2-specific T cells in unexposed and recovered donors
- Epitope mapping identifies one dominant HLA-B7-restricted epitope
- Identical TCR clonotypes recognize SARS-CoV-2 and betacoronaviruses epitopes
- Crystal structures reveal distinct peptide conformations between alpha- and betacoronaviruses

Authors

Katie E. Lineburg, Emma J. Grant, Srividhya Swaminathan, ..., Kirsty R. Short, Corey Smith, Stephanie Gras

Correspondence

corey.smith@qimrberghofer.edu.au (C.S.), s.gras@latrobe.edu.au (S.G.)

In brief

The impact of seasonal coronaviruses on immune responses to SARS-CoV-2 is an active area of research. Lineburg et al. identify CD8+ T cells specific for a conserved and immunodominant SARS-CoV-2 epitope in HLA-B7+ individuals. Furthermore, SARS-CoV-2 epitope-specific CD8+ T cells display cross-reactivity to beta- but not alphacoronaviruses because of distinct peptide-HLA conformations.

Lineburg et al., 2021, Immunity *54*, 1055–1065
May 11, 2021
https://doi.org/10.1016/j.immuni.2021.04.006

CellPress

Immunity

CellPress

Article

CD8+ T cells specific for an immunodominant SARS-CoV-2 nucleocapsid epitope cross-react with selective seasonal coronaviruses

Katie E. Lineburg,[1,10] Emma J. Grant,[2,3,10] Srividhya Swaminathan,[1,4,10] Demetra S.M. Chatzileontiadou,[2,3,10] Christopher Szeto,[2,3] Hannah Sloane,[2,3] Archana Panikkar,[1] Jyothy Raju,[1] Pauline Crooks,[1] Sweera Rehan,[1] Andrea T. Nguyen,[2,3] Lea Lekieffre,[1] Michelle A. Neller,[1] Zhen Wei Marcus Tong,[5] Dhilshan Jayasinghe,[2,3] Keng Yih Chew,[5] Christian A. Lobos,[2,3] Hanim Halim,[2] Jacqueline M. Burrows,[1] Alan Riboldi-Tunnicliffe,[6] Weisan Chen,[3] Lloyd D'Orsogna,[7,8] Rajiv Khanna,[1] Kirsty R. Short,[5,9] Corey Smith,[1,4,11,*] and Stephanie Gras[2,3,11,12,*]

[1]QIMR Berghofer Centre for Immunotherapy and Vaccine Development and Translational and Human Immunology Laboratory, Department of Immunology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia
[2]Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia
[3]Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, VIC 3086, Australia
[4]Faculty of Medicine, The University of Queensland, Brisbane, QLD 4072, Australia
[5]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, QLD 4072, Australia
[6]Australian Synchrotron, ANSTO, Clayton, VIC 3168, Australia
[7]Department of Clinical Immunology, PathWest Laboratory Medicine, Fiona Stanley Hospital, Murdoch, WA 6150, Australia
[8]School of Medicine, University of Western Australia, Nedlands, WA 6009, Australia
[9]Australian Infectious Diseases Research Centre, The University of Queensland, Brisbane, QLD 4072, Australia
[10]These authors contributed equally
[11]These authors contributed equally
[12]Lead contact
*Correspondence: corey.smith@qimrberghofer.edu.au (C.S.), s.gras@latrobe.edu.au (S.G.)
https://doi.org/10.1016/j.immuni.2021.04.006

SUMMARY

Efforts are being made worldwide to understand the immune response to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus responsible for the coronavirus disease 2019 (COVID-19) pandemic, including the impact of T cell immunity and cross-recognition with seasonal coronaviruses. Screening of SARS-CoV-2 peptide pools revealed that the nucleocapsid (N) protein induced an immunodominant response in HLA-B7+ COVID-19-recovered individuals that was also detectable in unexposed donors. A single N-encoded epitope that was highly conserved across circulating coronaviruses drove this immunodominant response. *In vitro* peptide stimulation and crystal structure analyses revealed T cell-mediated cross-reactivity toward circulating OC43 and HKU-1 betacoronaviruses but not 229E or NL63 alphacoronaviruses because of different peptide conformations. T cell receptor (TCR) sequencing indicated that cross-reactivity was driven by private TCR repertoires with a bias for TRBV27 and a long CDR3β loop. Our findings demonstrate the basis of selective T cell cross-reactivity for an immunodominant SARS-CoV-2 epitope and its homologs from seasonal coronaviruses, suggesting long-lasting protective immunity.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is an emerging virus responsible for the ongoing coronavirus disease 2019 (COVID-19) pandemic. Over 97 million individuals have been infected and over 2 million individuals have succumbed to infection (Dong et al., 2020). Although some vaccines against SARS-CoV-2 are already being administered, and others are in development, many questions remain regarding the immune response toward this virus. Cytotoxic CD8+ T cells are key players in the immune response to viral infections because they participate directly in viral clearance. Among the 26 viral proteins of SARS-CoV-2, some surface proteins, such as the spike protein (S), are more variable, whereas others are internal and more conserved, such as the nucleocapsid protein (N). The sequence conservation of non-surface proteins makes them ideal vaccine targets for activating cytotoxic CD8+ T cells. CD8+ T cells recognize small peptides (typically 8–10 amino acids) together with human leukocyte antigen (HLA) molecule with varying affinity. Although HLA-A2, the most prevalent HLA molecule (~40% frequency worldwide; Ellis et al., 2000), can present SARS-CoV-2 N-derived peptides (Szeto et al., 2021), they are only weakly immunogenic (Habel et al., 2020). However, it remains unclear whether this is a characteristic

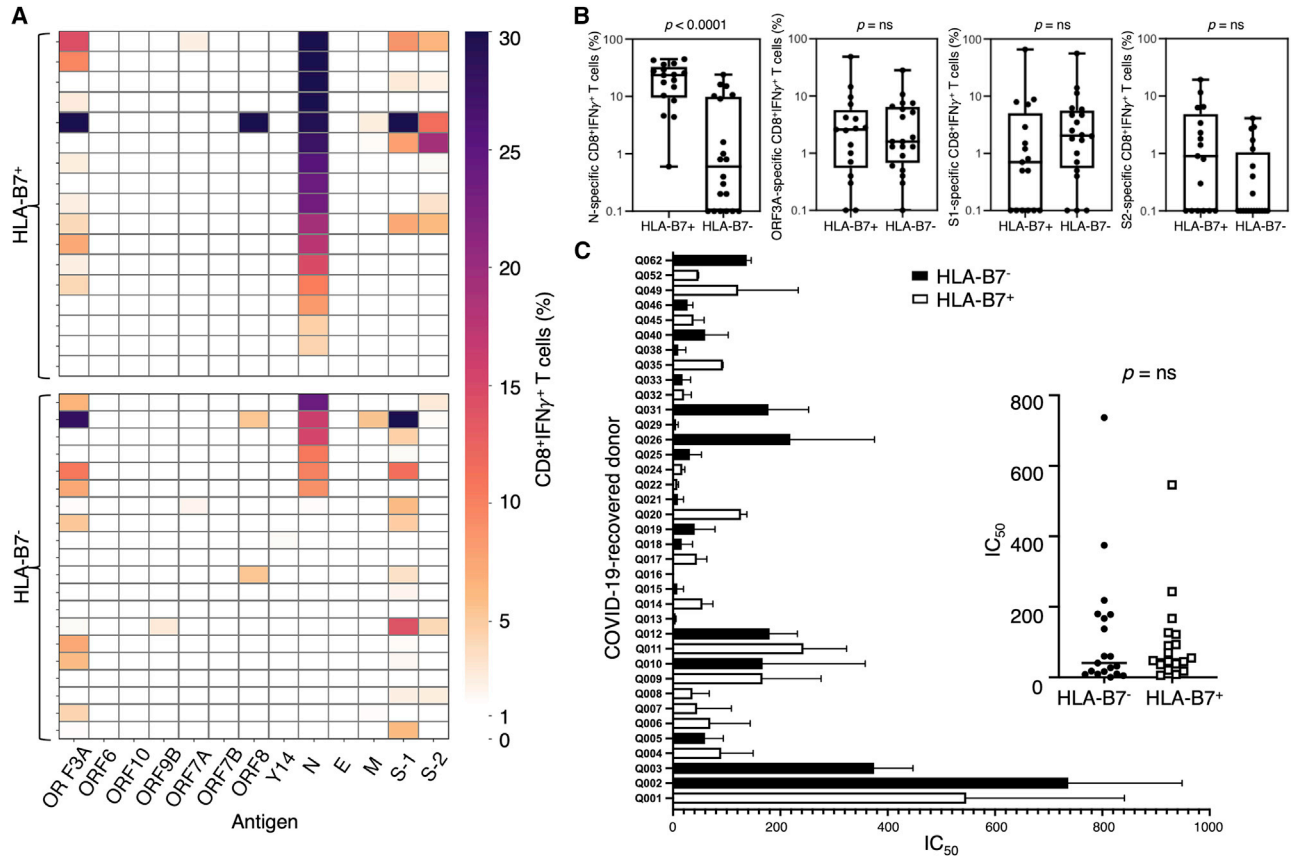


Figure 1. Immunodominant response to N in HLA-B7$^+$ COVID-19-recovered donors

PBMCs from COVID-19-recovered donors were stimulated with overlapping peptides from SARS-CoV-2 antigens, cultured for 2 weeks in the presence of IL-2, and then assessed for IFN-γ production following recall with the cognate pool.

(A) Heatmap representing the frequency of CD8$^+$ IFN-γ-producing cells responding to each antigen from a total of 37 COVID-19-recovered donors (n = 17 HLA-B7$^+$ and n = 20 HLA-B7$^-$ donors).

(B) Comparative analysis of T cell responses in HLA-B7$^+$ and HLA-B7$^-$ donors in response to different antigens. Data are represented as box-and-whisker plots displaying the median, including minimum to maximum range. p values were calculated using a Mann-Whitney test.

(C) Serum samples were serially diluted and tested for neutralizing activity against SARS-CoV-2. Neutralization curves were generated, and the respective inhibitory concentration at 50% (IC_{50}) was analyzed using a "log (inhibitor) versus normalized response" equation (GraphPad). IC_{50} values of study subjects were graphed individually (left) and grouped according to their HLA-B7 expression profile (right).

See also Figures S1 and S6 and Tables S1 and S2.

specific to the selected N-derived peptides because HLA-A2$^+$ individuals also demonstrate a strong CD8$^+$ T cell response to an S-derived epitope (Shomuradova et al., 2020). Therefore, it is imperative to identify and characterize novel immunogenic CD8$^+$ T cell epitopes against SARS-CoV-2.

Another important aspect of understanding the immune response to this new coronavirus is to understand the properties of protective pre-existing immunity at a population level (Karlsson et al., 2020). It has been proposed that pre-existing immune memory may be generated by previous infection with seasonal coronaviruses (OC43, 229E, NL63, and HKU-1) (Mateus et al., 2020; Ng et al., 2020). However, the identities of the T cell epitopes that give rise to pre-existing immunity are only starting to emerge (Bacher et al., 2020; Braun et al., 2020; Karlsson et al., 2020; Le Bert et al., 2020; Mateus et al., 2020; Nelde et al., 2021; Schulien et al., 2021; Sette and Crotty, 2020; Shomuradova et al., 2020). The SARS-CoV-2 N protein sequence is 90.3% identical to the SARS-CoV-1 N protein but shares only 29% identity with the N proteins of OC43 and HKU-1 and 23% with the NL63 and 229E virus strains. The majority of this sequence homology occurs in the N-terminal domain of the protein, which also contains the immunogenic $N_{105–113}$ peptide (SPRWYFYYL, hereafter referred to as SPR) (Ferretti et al., 2020; Kared et al., 2020; Peng et al., 2020; Schulien et al., 2021; Sekine et al., 2020; Snyder et al., 2020). The SPR peptide sequence is identical in SARS-CoV-2 and SARS-CoV-1; it differs by only one residue in OC43 and HKU-1 (LPRWYFYYL, hereafter referred to as LPR), three residues in 229E (SPKLHFYYL, hereafter referred to as SPK), and four in NL63 viruses (PPKVHFYYL, hereafter referred to as PPK).

Here we demonstrated that HLA-B7$^+$ individuals who had recently recovered from COVID-19 exhibited a dominant CD8$^+$ T cell response that was highly specific to the SPR epitope of the SARS-CoV-2 N protein. In addition, we identified SPR-specific CD8$^+$ T cells in multiple unexposed HLA-B7$^+$ individuals that showed cross-reactivity with the homologous LPR peptide

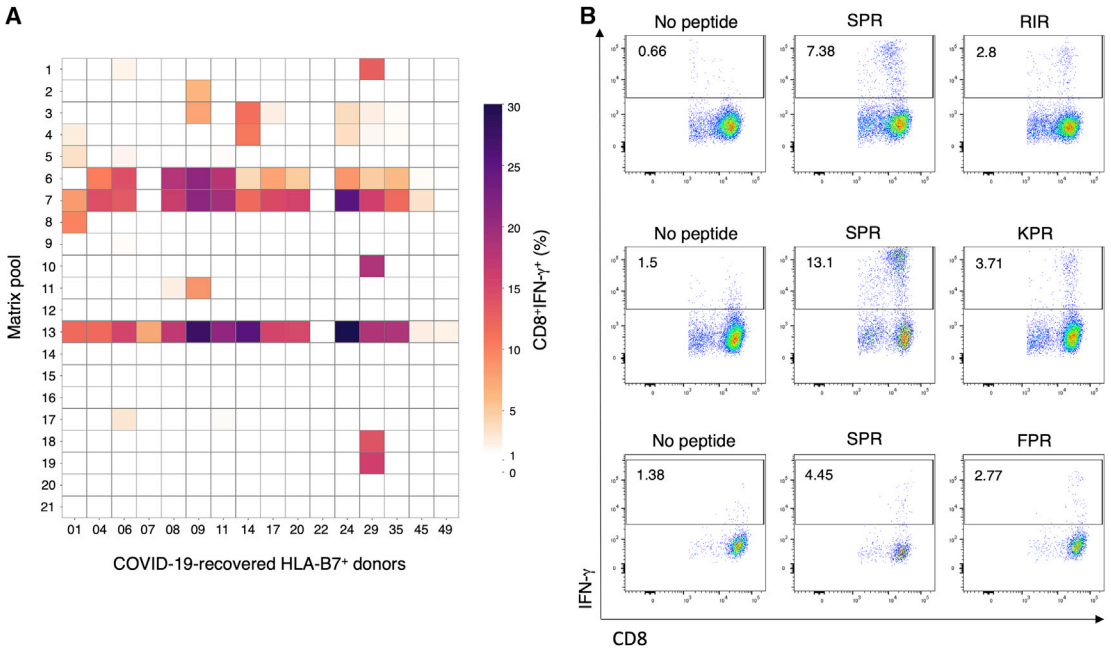


Figure 2. Peptide matrix analysis of T cell responses to N in COVID-19-recovered donors

(A) T cell cultures from affected individuals showing reactivity to the nucleocapsid (N) were stimulated with matrix pools covering the entire N protein and assessed for cytokine production by ICS. Shown is a heatmap representing the frequency of $CD8^+$ IFN-γ-producing cells responding to each matrix pool from a total of 16 $HLA\text{-}B7^+$ COVID-19-recovered donors (the Q052 donor did not show an N protein response and was excluded).

(B) Representative flow plots displaying COVID-19-recovered donors' $CD8^+$ IFN-γ responses to the SPR, RIR, KPR, and FPR peptides or the no-peptide control. See also Figure S6 and Table S3.

from the OC43 and HKU-1 strains. However, this same cross-recognition did not extend to the similarly homologous SPK and PPK peptides from the 229E and NL63 viruses, respectively. Using the structural landscape of HLA-B7 bound to SPR and its variants, we delineate the molecular basis for selective T cell cross-reactivity toward specific seasonal coronaviruses.

RESULTS

SARS-CoV-2 N is highly immunogenic in $HLA\text{-}B7^+$ COVID-19-recovered individuals

Although many $CD8^+$ T cell studies have focused on different proteins or a limited number of selected peptides (Habel et al., 2020; Kared et al., 2020; Sekine et al., 2020; Shomuradova et al., 2020; Szeto et al., 2021), the specific SARS-CoV-2 proteins that consistently elicit the strongest $CD8^+$ T cell responses remain ill defined (Ferretti et al., 2020; Grifoni et al., 2020; Nelde et al., 2021; Peng et al., 2020). To assess this, we used peripheral blood mononuclear cells (PBMCs) derived from COVID-19-recovered individuals (n = 37) to screen 13 overlapping peptide pools spanning 12 SARS-CoV-2 proteins (Figure 1; Tables S1 and S2). Although all 37 COVID-19-recovered donors displayed immunogenicity to SARS-CoV-2 peptide stimulation, 70% of donors or more demonstrated a $CD8^+$interferon γ $(IFN\text{-}\gamma)^+$ response of 5% or higher when stimulated with ORF3A, N, S-1, or S-2 peptide pools (26 of 37). The most consistent $CD8^+$ T cell responses were generated against the N protein-derived peptide pool in conjunction with expression of HLA-B7 (Figure 1A). The majority (76%) of these $HLA\text{-}B7^+$ individuals demonstrated a dominant $CD8^+$ T cell response with an average $IFN\text{-}\gamma^+$ frequency of 10% or higher (13 of 17). Conversely, only 25% of $HLA\text{-}B7^-$ individuals displayed comparable responses (5 of 20). Significant differences in $CD8^+$ T cell responses between $HLA\text{-}B7^+$ and $HLA\text{-}B7^-$ individuals were only observed in response to the N protein (Figure 1B; $p < 0.0001$).

To assess whether $CD8^+$ T cell reactivity to the N protein was associated with the presence of elevated antibody titers, we examined neutralizing anti-S immunoglobulin G (IgG) titers in COVID-19-recovered individuals and did not observe any significant differences between $HLA\text{-}B7^+$ and $HLA\text{-}B7^-$ donors (n = 37) (Figure 1C). Similarly, the strong $CD8^+$ T cell response was comparable between males (n = 14) and females (n = 23) and similar within donors categorized as younger (n = 18) or older (n = 19) than 50 years. The distribution of disease severity reported was comparable between $HLA\text{-}B7^+$ and $HLA\text{-}B7^-$ cohorts (Table S2).

To define the immunogenic peptide(s) responsible for this $HLA\text{-}B7^+$ restricted N-specific $CD8^+$ T cell response, we used a matrix of peptide pools to identify which specific peptide elicited the strongest IFN-γ response. Through this process, we identified an immunogenic hotspot (Figure 2) corresponding to the same N protein region that contains the known immunogenic HLA-B7-restricted SPR ($N_{105–113}$) epitope (Table S3; Ferretti et al., 2020; Kared et al., 2020; Schulien et al., 2021; Sekine et al., 2020; Snyder et al., 2020). Eighty percent of $HLA\text{-}B7^+$ COVID-19-recovered donors showed a strong $CD8^+$ T cell response to this peptide (14 of 17), consistent with other reports (Ferretti et al., 2020; Kared et al., 2020; Schulien et al., 2021; Sekine et al., 2020; Snyder et al., 2020). Three additional

Table 1. Coronaviruses peptide homologs to SARS-CoV-2 SPR

Virus	Peptide Sequence	Peptide Name	Peptide Sequence Identity (%)[a]	N Protein Sequence Identity (%)[a]	Tm (°C)
SARS-CoV-2	SPRWYFYYL	SPR	–	–	62.8 ± 0.7
SARS-CoV-1	SPRWYFYYL	SPR	100	90.3	62.8 ± 0.7
OC43	*L*PRWYFYYL	LPR	88.9	29.2	59.8 ± 0.7
HKU-1	*L*PRWYFYYL	LPR	88.9	29.6	59.8 ± 0.7
229E	SP*KLH*FYYL	SPK	66.7	23.2	55.7 ± 0.8
NL63	*PPKVH*FYYL	PPK	55.5	24.9	48.2 ± 1.2

Mutations from the SARS-CoV-2 peptide are denoted in italic, and anchor residues are underlined. Tm is determined at 50% of its normalized fluorescence intensity and indicative of the temperature required to unfold 50% of the protein. The Tm value is represented as the mean ± SEM of n = 2.
[a]Sequence identity with SARS-CoV-2-derived SPR peptide.

subdominant HLA-B7-restricted epitopes from the N protein were identified: RIRGGDGKM (RIR; $N_{93–101}$), FPRGQGVPI (FPR; $N_{66–74}$), and KPRQKRTAT (KPR; $N_{257–265}$) (Figure 2B; Table S3). Although FPR and KPR are known to be immunogenic in COVID-19-recovered individuals (Schulien et al., 2021), RIR has been predicted but not yet described as an immunogenic SARS-CoV-2 $CD8^+$ T cell epitope (Sekine et al., 2020).

These data show that the SARS-CoV-2 N protein contains multiple immunogenic epitopes that induce a strong $CD8^+$ T cell response in COVID-19-recovered HLA-$B7^+$ individuals and that the dominant $CD8^+$ T cell response is directed against the SPR peptide.

$CD8^+$ T cells exhibit cross-reactivity between SARS-CoV-2- and OC43/HKU-1-derived N peptides

Given that this dominant SPR-specific $CD8^+$ T cell response has been observed in HLA-$B7^+$ individuals (Ferretti et al., 2020; Kared et al., 2020; Schulien et al., 2021; Sekine et al., 2020; Snyder et al., 2020) and that pre-existing immunity toward this peptide has been suggested for unexposed individuals (Ferretti et al., 2020; Nelde et al., 2021; Schulien et al., 2021), we next sought to identify the basis of the T cell cross-reactivity that elicits the cross-recognition observed against SARS-CoV-2. Sequence alignment comparing SARS-CoV-2 with other coronaviruses demonstrated that, although the entire N protein sequence homology is below 30%, the region containing the SPR peptide is more conserved, displaying 55%–89% sequence homology (Table 1). The homologous peptides from SARS-CoV-1, OC43, HKU-1, 229E, and NL63 share a proline at position 2 (P2-P) and a leucine at position 9 (P9-L). These residues are critical anchors for HLA-B7 binding (Sette and Sidney, 1999) and suggest that SPR peptide homologs from these seasonal viral strains might bind to the HLA-B7 molecule. Alignment of 26,158 N protein sequences from SARS-CoV-2 revealed a 100% conservation of the SPR peptide within circulating isolates, including the variants isolated in the United Kingdom (B.1.1.7), South Africa (B.1.351), and Brazil (P.1) (Table S4). In addition, circulating isolates of the homologous peptides (LPR, SPK, and PPK) are entirely conserved within their respective viruses (Table S4), demonstrating their global conservation.

Considering the high degree of conservation between SPR and these homologous peptides (Table 1), we investigated whether SARS-CoV-2-specific $CD8^+$ T cells were able to cross-recognize any of the other conserved homologous peptides. We first expanded SPR-specific T cells from PBMCs isolated from COVID-19-recovered individuals (Figure 3). Following re-stimulation of the SPR-specific $CD8^+$ T cells with the SPR peptide, a high level of specificity for the SPR peptide was detected (Figure 3A), with an average $CD8^+IFN\text{-}\gamma^+$ response of 23.13%. The same $CD8^+$ T cells were able to cross-react when restimulated with the LPR peptide (Figure 3A), resulting in a similar $CD8^+IFN\text{-}\gamma^+$ T cell response of 18.86%. Conversely, we observed little to no reactivity when the SPR-specific $CD8^+$ T cells were restimulated with the SPK or PPK peptide (Figure 3A; Figure S1A).

Because it was evident that $CD8^+$ T cells were able to cross-recognize the variant peptides from selected seasonal coronaviruses, we next assessed the level of cross-reactive $CD8^+$ T cells in SARS-CoV-2-unexposed individuals. We expanded peptide-specific $CD8^+$ T cells against each of the four peptide variants using PBMCs from unexposed donors (n = 13 for SPR, LPR, and SPK lines and n = 10 for PPK lines). After 10–14 days of expansion, we re-stimulated with each of the individual peptide variants and assessed the ability of $CD8^+$ T cells to produce IFN-γ using an intracellular cytokine staining (ICS) assay (Figures 3B and 3C; Figure S1). Peptide-specific $CD8^+$ T cells generated against SPR, LPR, and SPK peptides were most specific for their cognate peptide (Figure 3C; Figure S1); however, T cells generated against the PPK peptide demonstrated limited specificity (Figure 3C; Figure S1E). Peptide-specific $CD8^+$ T cells generated in response to SPR or LPR peptide stimulation displayed cross-reactivity in the majority of unexposed donors without a significant difference between the two peptides (Figure 3C). Conversely, those generated in response to SPK or PPK stimulation displayed little to no cross-reactivity. Functional avidity analysis confirmed comparable recognition of SPR and LPR peptides by T cells generated from unexposed individuals; it also indicated that SPR-specific T cells generated by COVID-19-recovered donors display significantly higher avidity (Figure 3D; Figure S2A), similar to that reported in $CD4^+$ T cells (Bacher et al., 2020). A higher magnitude of SPR-specific $CD8^+$ T cells was observed from COVID-19-recovered donors (median, 16%) than from unexposed donors (median, 0.4%) (Figure S2B).

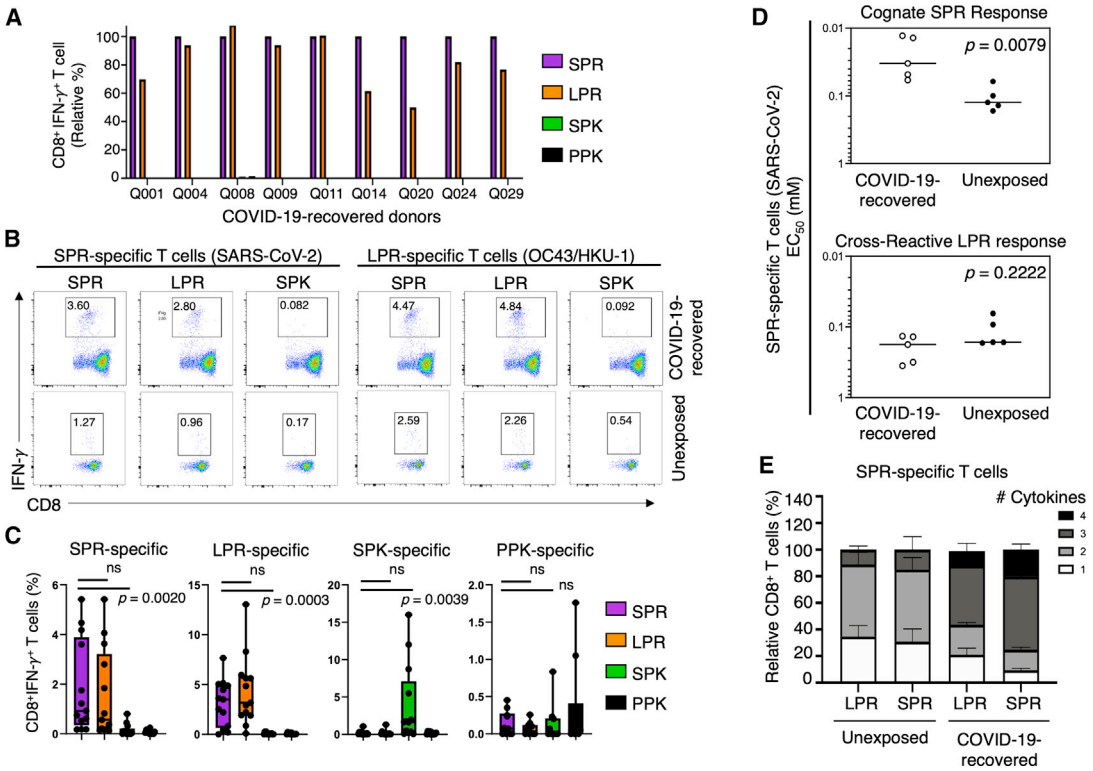


Figure 3. CD8⁺ T cell responses toward N peptides derived from seasonal and pandemic coronaviruses

PBMCs from unexposed and COVID-19-recovered donors were stimulated with SPR (SARS-CoV-2), LPR (OC43/HKU-1 betacoronavirus), SPK (229E alphacoronavirus), or PPK (NL63 alphacoronavirus) peptides, cultured for 2 weeks in the presence of IL-2, and then assessed for IFN-γ production following recall with their cognate peptide or each of the homologous peptides listed above.

(A) $CD8^+IFN\text{-}\gamma^+$ responses of SPR-specific T cells expanded from the PBMCs of COVID-19-recovered donors following recall stimulation with the SPR, LPR, SPK, or PPK peptide in an ICS.

(B) Representative flow plots of $CD8^+IFN\text{-}\gamma^+$ responses in SPR-specific and LPR-specific T cells expanded from COVID-19-recovered and unexposed donors' PBMCs, displaying cross-recognition between SPR and LPR peptides following cognate peptide stimulation.

(C) Graph displaying the frequency of $CD8^+IFN\text{-}\gamma^+$ T cell responses in SPR, LPR, SPK, and PPK-specific T cells following recall stimulation on day 14 of culture. Each peptide-specific T cell line was re-stimulated individually with its cognate peptide or one of the homologous peptides (SPR, LPR, SPK, or PPK), and their $IFN\text{-}\gamma^+$ response was measured by ICS. Data are represented as median, displaying minimum to maximum range. p values were calculated using a Mann-Whitney test.

(D) Graphs displaying the avidity at the effective concentration used to induce half maximal response (EC_{50}) of SPR-specific T cells in response to cognate SPR peptide and cross-presentation of the LPR peptide, with the median response indicated by a line. Statistical analysis was performed using a Mann-Whitney test, comparing COVID-19-recovered (n = 5) and unexposed (n = 5) donors.

(E) Graph displaying the polyfunctionality of $CD8^+$ SPR-specific T cells from unexposed and COVID-19-recovered donors following re-stimulation with SPR or LPR peptide. Data are represented as relative frequency (percent) of total $CD8^+$ T cells.

See also Figures S1 and S2 and Tables S1 and S2.

To assess the polyfunctionality of activated $CD8^+$ T cells in response to peptide recall, we measured the production of IFN-γ, tumor necrosis factor (TNF), and interleukin-2 (IL-2) as well as CD107a expression. SPR-specific $CD8^+$ T cells from COVID-19-recovered individuals exhibited broader polyfunctionality than those from unexposed donors, with the majority expressing three or four functions (IFN-γ, TNF, CD107a ± IL-2) (Figure 3E; Figure S3A). In comparison, responses of unexposed individuals, irrespective of the peptide used, were dominated by $CD8^+$ T cells generating one or two cytokines that included IFN-γ, TNF, and/or CD107a but not IL-2 (Figure 3E; Figure S3).

These data demonstrate that exposed and unexposed individuals can generate SPR-specific $CD8^+$ T cells that efficiently cross-recognize the homologous LPR peptide from circulating betacoronaviruses.

SPR-specific $CD8^+$ T cells share biased TRBV27 gene use and long CDR3β in unexposed and COVID-19-recovered individuals

To further characterize these cross-reactive $CD8^+$ T cells, we next assessed their TCR repertoires. SPR-specific $CD8^+$ T cells from COVID-19-recovered donors were stained with the HLA-B7-SPR multimer, and single cells were isolated by fluorescence-activated cell sorting (FACS) (Figure 4A). Additionally, SPR- and LPR-specific T cells from unexposed individuals were stained individually with the HLA-B7-SPR or HLA-B7-LPR multimer, and single cells were sorted (Figure S2C). The TCR repertoire was determined using multiplex RT-PCR (Grant et al., 2018; Wang et al., 2012). A total of 767 T cells were sequenced from COVID-19-recovered (n = 5) and unexposed (n = 7) individuals. This analysis revealed that clonotypic

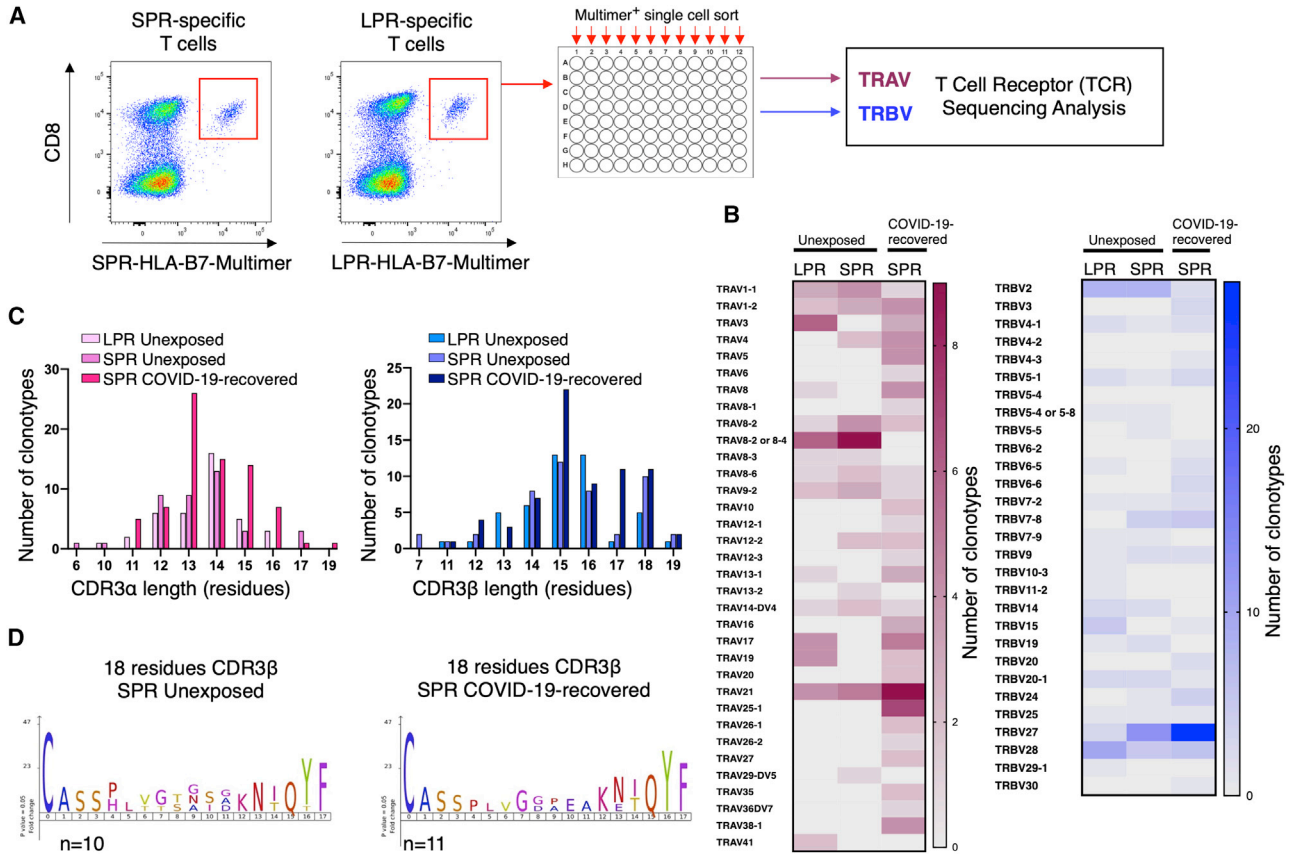


Figure 4. Diverse TCR repertoires are utilized for recognition of SPR and LPR peptides in unexposed and COVID-19-recovered individuals

PBMCs from COVID-19-recovered or unexposed individuals were stimulated with the SPK or LPR peptides and cultured for 10–14 days in the presence of IL-2. $CD8^+$ T cell lines were stained with the SPK or LPR multimer. $Multimer^+$ cells were single cell sorted, and the TCR repertoire was determined by multiplex PCR.

(A) Schematic displaying representative multimer staining for single-cell isolation and TCR sequencing of the SPR and LPR-specific T cells in $HLA\text{-}B7^+$ unexposed donors.

(B) A heatmap displaying preferred TRAV (left panel) and TRBV (right panel) use of SPR and LPR-specific TCRs in unexposed and COVID-recovered individuals.

(C) Summary of CDR3α (left panel) and CDR3β (right panel) lengths for distinct peptide-specific TCR clonotypes.

(D) Motif of the 18-amino-acid-long CDR3β loops from unexposed (left panel) and COVID-19-recovered (right panel) donors.

See also Figures S2 and S4 and Table S5.

diversity underpins the response to SPR and LPR peptides, with a total of 209 clonotypes isolated and 80 unique αβ TCR pairs (Table S5). The TCR repertoires for each individual were private, with no shared (public) TCRs identified between individuals. This suggests that the number of distinct clonotypes capable of responding to the SPR epitope likely underpins the consistent immunodominant response observed in convalescent participants. Cross-reactive $CD8^+$ T cell clones recognizing the SPR and LPR peptides were observed in the majority of unexposed individuals (n = 6 of 7). However, a unique clonotypic profile was evident in all donors' PBMCs following SPR and LPR stimulation. This suggests that subtle differences in peptide recognition by different TCRs not only influence clonal expansion but also affect the efficiency and avidity of T cell cross-recognition toward homologous peptides (Figure S2A). Comparison of the SPR-specific T cell αβTCR sequences identified a strong T cell receptor beta variable region (TRBV) use bias, with 30% and 40% of these clonotypes expressing the TRBV27 gene in unexposed and COVID-19-recovered individuals, respectively, whereas LPR-specific T cells displayed a bias for TRBV28 (22%; Figure 4B). In comparison, α chain sequencing revealed limited shared use between COVID-19-recovered and unexposed individuals (Figure 4B). A use bias for TRAV8-2 in SPR- and LPR-specific clonotypes was seen in unexposed individuals, and shared TRAV21 use (~10%–12%) in unexposed and COVID-19-recovered donors (Figure 4B; Table S5). Although we observed a typical distribution in CDR3α length, a preference for long CDR3β was seen in all populations, with the majority of identified sequences greater than 15 amino acids in length (Figure 4C). Although analysis of the CDR3 loop sequences did not reveal a shared motif in the α or β chains, reflecting the private nature of the TCR repertoires (Figures S4), a shared ^{108}PxxGx[P/G/A]x^{114} motif (where x is any of V/T/S/L/G) was observed in the 18-residue-long CDR3β loops for the SPR-specific TCRs from COVID-19-recovered and unexposed donors (Figure 4D).

Although not definitive, the presence of this shared motif in clonotypes from exposed and unexposed donors suggests

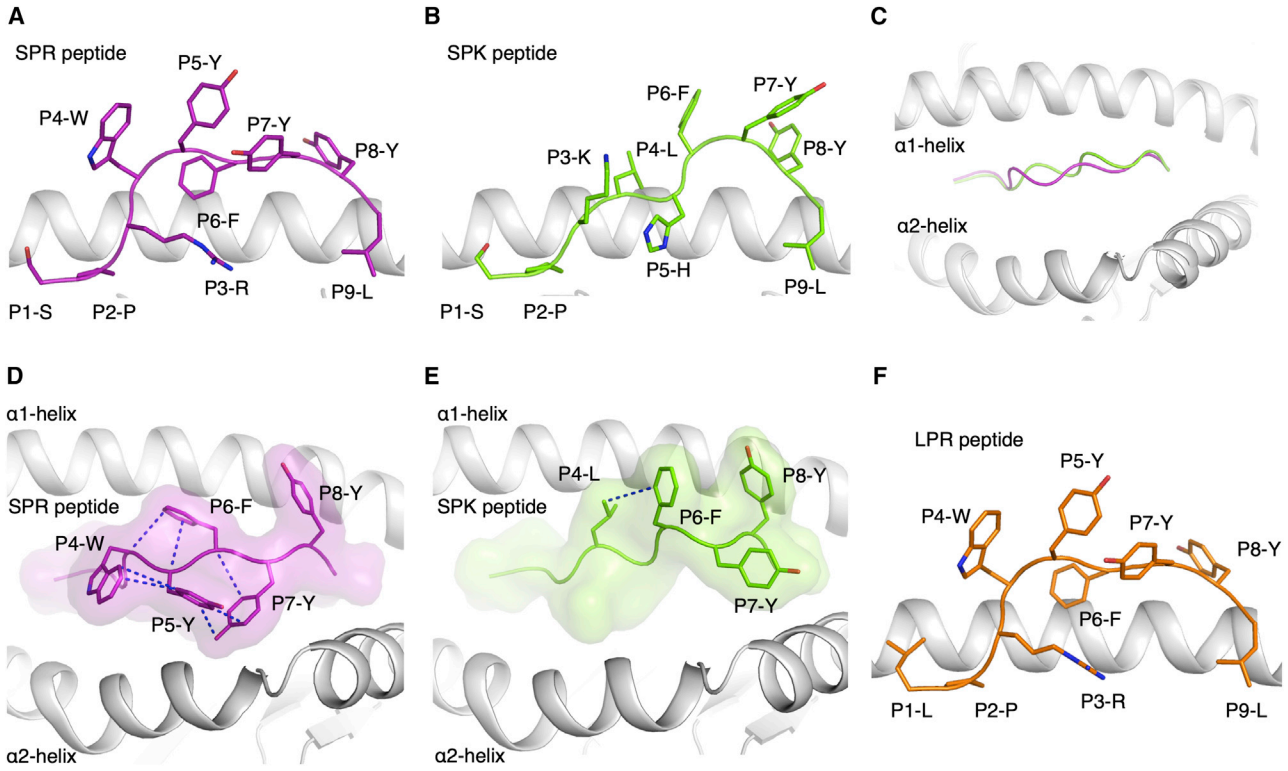


Figure 5. A structural basis for selective T cell cross-reactivity

(A and B) Crystal structures of the (A) SPR (purple stick) and (B) SPK (green stick) peptides presented by the HLA-B7 molecule (white cartoon), represented in the same orientation.

(C) Superposition of the HLA-B7-SPR (purple) and HLA-B7-SPK (green) structures represented as a cartoon from a top-down view of the antigen binding cleft.

(D) Top view of the HLA-B7-SPR structure, with a surface representation of the SPR peptide in purple with transparency and the solvent-exposed residues as sticks. The blue dashed lines represent intra-peptide interactions between the bulky aromatic and solvent-exposed residues of the SPR peptide.

(E) Top view of the HLA-B7-SPK complex in the same orientation as in (D). The SPK peptide is represented as green sticks for the surface-exposed residues and as a green surface. The blue dashed lines represent the intra-peptide interaction.

(F) Model of the structure of the HLA-B7 (white cartoon) molecule presenting the LPR peptide (orange stick), based on modeling from the HLA-B7-SPR complex. See also Figure S5 and Table S6.

that the pre-existing immunity in HLA-B7$^+$ individuals favored clonal expansion.

Distinct epitope presentation drives selective T cell cross-reactivity in HLA-B7$^+$ individuals

To determine the molecular basis of this selective CD8$^+$ T cell cross-reactivity, we assessed the ability of the SPR and homologous peptides to form a stable complex with HLA-B7. Each peptide refolded efficiently with the HLA-B7 molecule, with the exception of the PPK peptide (poor yield). Next, the thermostability of each peptide-HLA (pHLA) complex was assessed (Figure S5A). The most stable complexes identified were HLA-B7-SPR and HLA-B7-LPR (melting temperature, Tm $\geq$ 60°C), followed by HLA-B7-SPK (Tm $\approx$ 56°C), whereas HLA-B7-PPK was less stable than the others (Tm $\approx$ 48°C) (Table 1). Considering that the SPR and SPK peptides demonstrated similar thermostability in complex with HLA-B7 but that CD8$^+$ T cells were not able to cross-recognize both peptides, we next determined the crystal structures of these peptides bound to the HLA-B7 molecule (Figure 5).

The crystal structures of HLA-B7-SPR and HLA-B7-SPK were solved at 2.88 Å and 1.97 Å, respectively (Table S6). The electron density was well defined for both peptides (Figure S5), showing a stable conformation of the peptides in the cleft of HLA-B7 (Figure 5). Comparison of both structures revealed that different conformations were adopted by SPR and SPK peptides. The arginine (R) at position (P) three (P3-R) in the SPR peptide acted as a secondary anchor residue (Figure 5A), whereas for SPK, the lysine at P3 was pushed out of the cleft by the histidine at P5 that acted as a secondary anchor in the cleft (Figure 5B). The sequence differences at P3 and P5 of the peptides altered the SPK conformation compared with the SPR conformation (Figure 5C) despite the shared ^{6}FYYL^9 motif. Four of the five aromatic residues of the SPR peptide formed a network of interaction, stabilizing each other on the surface of the HLA-B7 cleft (Figure 5D). This formed a compact and large binding surface for potential interactions with the TCR. The large binding surface might explain the expansion of T cells with a long CDR3β loop (>16 residues) observed in 34% and 37% of the clonotypes in COVID-19-recovered and unexposed donors, respectively (Figure 4C), which would be able to contact a greater surface area of the peptide because of their length. In comparison, the conformation adopted by the SPK peptide only exposed three aromatic residues located at the C-terminal part of the peptide (^{6}FYY^8)

(Figure 5E). The LPR peptide only differs from SPR at P1, and this single substitution is unlikely to change the peptide conformation, given that HLA-B7 accommodates a wide range of residues at P1, based on previously reported structures (Brennan et al., 2012; Chan et al., 2018; Du et al., 2016; Rowntree et al., 2018, 2020). Therefore, because of the high sequence identity, the LPR peptide is likely to adopt a conformation similar to the SPR peptide (Figure 5F), forming the basis of the $CD8^+$ T cell cross-reactivity observed in $HLA\text{-}B7^+$ individuals. Despite sharing 67% sequence identity, the different structures adopted by the SPR (SARS-CoV-2) and SPK (229E) peptides provide a basis for a low, if any, level of $CD8^+$ T cell cross-reactivity between these peptides, highlighting the fine specificity of $CD8^+$ T cells.

DISCUSSION

T cell mediated cross-reactivity in response to seasonal coronaviruses has the potential to limit the development and severity of COVID-19. Here we demonstrated that a robust $CD8^+$ T cell response in $HLA\text{-}B7^+$ individuals is primarily directed toward a single immunodominant epitope, SPR, encoded by the N protein of SARS-CoV-2, recognized by more than 80% of tested COVID-19-recovered donors. We also identified SPR-specific $CD8^+$ T cells in more than 90% of tested unexposed $HLA\text{-}B7^+$ individuals who had not been exposed to SARS-CoV-2. Although comparative analysis of circulating seasonal coronaviruses revealed a high level of conservation in the homologous peptide sequences, including conservation of the critical proline residue at P2 of the epitope, detailed analysis revealed that T cell responses in unexposed volunteers were driven by cross-reactive $CD8^+$ T cells specific for the LPR peptide from the OC43 and HKU-1 seasonal coronaviruses. It is possible that these cross-reactive memory $CD8^+$ T cells facilitate the strong, polyfunctional, high avidity $CD8^+$ T cell response toward the SPR peptide observed in T cells expanded from COVID-19-recovered individuals. In contrast, despite sharing the C-terminal $^6FYY^8$ motif and the HLA-B7-favored anchor residues (P2-P and P9-L) with the PPK and SPK peptides from the 229E and NL63 viruses, respectively, there was no $CD8^+$ T cell cross-reactivity between these and the SPR or LPR peptides. Considering that an average 4.1% of the global population is $HLA\text{-}B7^+$ and it is the sixth most frequent HLA-B molecule worldwide (frequency 0%–20%, depending on ethnicity' Gonzalez-Galarza et al., 2020; Solberg et al., 2008), these results may have profound implications for T cell-mediated protection against COVID-19. However, they also demonstrated the fine specificity of T cells for their cognate antigen, which may prevent cross-recognition of less conserved epitopes.

Studies have provided tantalizing evidence that individuals not exposed to SARS-CoV-2 may harbor T cells capable of cross-recognizing epitopes presented by HLA class I and class II molecules (Bacher et al., 2020; Braun et al., 2020; Karlsson et al., 2020; Le Bert et al., 2020; Nelde et al., 2021; Schulien et al., 2021; Shomuradova et al., 2020). However, several factors, including lack of sequence conservation between several of the reported cross-reactive epitopes, the non-physiological quantities of antigen used in detection assays, and the limited number of unexposed donors assessed, may contribute to overestimation of the prevalence of pre-existing T cell immunity when it comes to defining truly cross-reactive T cells that could protect against COVID-19. It is therefore critical to undertake a detailed analysis of the molecular characteristics that promote or limit T cell cross-reactive immunity (Karlsson et al., 2020; Sette and Crotty, 2020). It is well established from studies of other human viral infections, including HIV, cytomegalovirus (CMV), and influenza, that even a single amino acid change, particularly in a TCR contact region, can have a profound effect on cross-reactivity (Geldmacher et al., 2009; Goulder et al., 1997; Gras et al., 2009, 2010; Iglesias et al., 2011; Kløverpris et al., 2015; Smith et al., 2014). Here we revealed different conformations of the SPR and SPK peptides in complex with HLA-B7, providing a basis for the observed low or lack of T cell cross-reactivity. This shows that, despite the sequence similarity and conserved C-terminal motif between the peptides, T cell cross-reactivity will not occur toward all seasonal coronaviruses because of fine TCR specificity. Although we were unable to solve the structure of the LPR epitope in complex with HLA-B7, our modeling revealed significant structural homology with minimal changes induced by the S-to-L substitution at P1. Similarly, SPR- and LPR-specific T cells efficiently cross-recognized both peptides, displaying similar functional avidity and polyfunctionality. TCR sequence analysis of SPR- and LPR-specific T cells revealed that, although some biased gene use was observed (TRAV8-2, TRBV2, TRBV27, and TRBV28), the TCR repertoires were entirely private with no public TCR shared between individuals. However, we observed key TCRs able to cross-react toward the SPR and LPR peptides in unexposed individuals. This likely contributes to the strong response observed from COVID-19-recovered donors' expanded T cells and may suggest a level of pre-existing immunity. Although the TCR repertoire was private, we observed biased TRBV27 gene use with long CDR3β loops expressed preferentially for SPR-specific TCRs in unexposed and COVID-19-recovered individuals. Despite sequence diversity in the TCR repertoire, the bias for long length of CDR3 loops might reflect a common docking mechanism between the clonotypes and facilitate recognition of the numerous large aromatic residues of the SPR peptide.

It is clear that, although pre-existing immunity toward SARS-CoV-2 exists and may provide an advantage to certain individuals, this same advantage might be limited to a few epitopes specific to certain HLA molecules and dependent on previous encounters with certain coronaviruses. In addition, the presence of pre-existing immunity suggests that memory T cells could be activated upon SARS-CoV-2 infection, important for long-lasting protection that will be critical for vaccination (Cañete and Vinuesa, 2020; Dan et al., 2020; Jarjour et al., 2021). Those cross-reactive T cells could also be used as biomarkers to help predict the severity of disease or the efficacy of a vaccine. The discovery of defined epitopes that elicit strong $CD8^+$ T cell responses, particularly those that are pre-existing, is important because these epitopes could be exploited to boost or prime the immune response in a vaccination strategy and therefore could provide significant protection of the global population.

Limitations of study

Our study established that one single epitope dominates the response toward the SARS-CoV-2 N protein in $HLA\text{-}B7^+$

individuals who recovered from COVID-19. Although HLA-B7 is the most common HLA-B molecule in populations of European decent, it is the sixth most frequent HLA-B molecule worldwide and is expressed at a frequency of 4% globally. Therefore, although generalization at the global population level is not possible because of our focus on a single allele, our study does provide a potential mechanism for the immunodominant response observed in HLA-B7$^+$ individuals via cross-recognition by seasonal coronavirus T cells. Given the conservation between SARS-CoV-2 and seasonal coronaviruses, especially with beta-coronaviruses, it is likely that pre-existing and cross-reactive T cell responses occur for other peptides restricted to diverse HLAs. However, our study also demonstrates how even a few changes in amino acid sequence can affect T cell recognition, suggesting that the presence of immunodominant cross-reactive CD8$^+$ T cells could be restricted to specific pHLA complexes. More detailed and precise research in this direction is critical to determine the actual presence of cross-reactive and pre-existing T cells between SARS-CoV-2 variants and seasonal coronaviruses and their potential role in protection against COVID-19.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Study participants
- METHOD DETAILS
  - Peripheral blood mononuclear cells
  - Generation of peptide-specific CD8$^+$ T cell lines
  - Intracellular cytokine assay
  - Multimer staining
  - Single-cell multiplex PCR
  - Protein **expression, refold and purification**
  - Differential scanning fluorimetry
  - Crystallization and structural determination
  - Model building
  - SARS-CoV-2 microneutralization assay
  - Sequence alignment of SPR and homologs from coronaviruses isolates
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.immuni.2021.04.006.

ACKNOWLEDGMENTS

The authors would like to thank Queensland Health Forensic & Scientific Services, Queensland Department of Health who provided the SARS-CoV-2 isolate QLD02; Monash facilities (Flow Core, Macromolecular Crystallization Facility, Imaging Facility); VTIS, PathWest Laboratory Medicine and CareDx for HLA Typing; and the MX team for assistance at the Australian Synchrotron. The authors would also like to thank all participants who took part in the study. This work was supported by generous donations from the QIMR Berghofer COVID-19 appeal and financial contributions from Monash University, the Australian Nuclear Science and Technology Organisation (ANSTO; AINSE ECR grants), the Australian Research Council (ARC), the National Health and Medical Research Council (NHMRC, GNT1132519), and the Medical Research Future Fund (MRFF, APP2005654). S.S. and H.S. are supported by an Australian Government Research Training Program scholarship. E.J.G. was supported by an NHMRC C.J. Martin fellowship (1110429) and is supported by an Australian Research Council DECRA (DE210101479). K.R.S. is supported by an Australian Research Council DECRA (DE180100512). S.G. is supported by an NHMRC SRF (1159272).

AUTHOR CONTRIBUTIONS

Conceptualization, K.E.L., E.J.G., S.S., D.S.M.C., C. Smith, and S.G.; formal analysis, K.E.L., E.J.G., S.S., D.S.M.C., H.S., C. Szeto., C. Smith, and S.G.; funding acquisition, K.R.S., C. Smith, and S.G.; investigation and methodology, K.E.L., E.J.G., S.S., D.S.M.C., C. Szeto, H.S., A.P., J.R., P.C., S.R., A.T.N., L.L., M.A.N., Z.W.M.T., D.J., K.Y.C., C.A.L., H.H., J.M.B., K.R.S., and S.G.; project administration, C. Smith and S.G.; resources, W.C., L.D., C. Smith, and S.G.; supervision, K.E.L., E.J.G., S.S., D.S.M.C., C. Smith, and S.G.; validation, K.E.L., E.J.G., S.S., D.S.M.C., C. Smith, and S.G.; visualization, K.E.L., E.J.G., S.S., D.S.M.C., H.S., C. Smith, and S.G. ; writing – original draft, K.E.L., E.J.G., S.S., D.S.M.C., H.S., C. Smith, and S.G. ; writing – review & editing, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure that the study questionnaires were prepared in an inclusive way. One or more of the authors of this paper self-identifies as living with a disability. The author list of this paper includes contributors from the location where the research was conducted who participated in the data collection, design, analysis, and/or interpretation of the work.

Received: January 22, 2021
Revised: February 26, 2021
Accepted: April 9, 2021
Published: April 13, 2021

REFERENCES

Aragão, D., Aishima, J., Cherukuvada, H., Clarken, R., Clift, M., Cowieson, N.P., Ericsson, D.J., Gee, C.L., Macedo, S., Mudie, N., et al. (2018). MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. J. Synchrotron Radiat. *25*, 885–891.

Bacher, P., Rosati, E., Esser, D., Martini, G.R., Saggau, C., Schiminsky, E., Dargvainiene, J., Schröder, I., Wieters, I., Khodamoradi, Y., et al. (2020). Low-Avidity CD4$^+$ T Cell Responses to SARS-CoV-2 in Unexposed Individuals and Humans with Severe COVID-19. Immunity *53*, 1258–1271.e5.

Braun, J., Loyal, L., Frentsch, M., Wendisch, D., Georg, P., Kurth, F., Hippenstiel, S., Dingeldey, M., Kruse, B., Fauchere, F., et al. (2020). SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. Nature *587*, 270–274.

Brennan, R.M., Petersen, J., Neller, M.A., Miles, J.J., Burrows, J.M., Smith, C., McCluskey, J., Khanna, R., Rossjohn, J., and Burrows, S.R. (2012). The impact of a large and frequent deletion in the human TCR β locus on antiviral immunity. J. Immunol. *188*, 2742–2748.

Bricogne, G.B.E., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O.S., Vonrhein, C., and Womack, T.O. (2011). BUSTER version 2.10 (Global Phasing Ltd.).

Brochet, X., Lefranc, M.P., and Giudicelli, V. (2008). IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. Nucleic Acids Res. *36*, W503-8.

Cañete, P.F., and Vinuesa, C.G. (2020). COVID-19 Makes B Cells Forget, but T Cells Remember. Cell *183*, 13–15.

Chan, K.F., Gully, B.S., Gras, S., Beringer, D.X., Kjer-Nielsen, L., Cebon, J., McCluskey, J., Chen, W., and Rossjohn, J. (2018). Divergent T-cell receptor recognition modes of a HLA-I restricted extended tumour-associated peptide. Nat. Commun. *9*, 1026.

Collaborative Computational Project, Number 4 (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D Biol. Crystallogr. *50*, 760–763.

Dan, J.M., Mateus, J., Kato, K., Hastie, K.M., Faliti, C.E., Ramirez, S.I., Frazier, A., Yu, E.D., Grifoni, A., Rawlings, S.A., et al. (2020). Immunological memory to SARS-CoV-2 assessed for greater than six months after infection. bioRxiv. https://doi.org/10.1101/2020.11.15.383323.

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

Du, V.Y., Bansal, A., Carlson, J., Salazar-Gonzalez, J.F., Salazar, M.G., Ladell, K., Gras, S., Josephs, T.M., Heath, S.L., Price, D.A., et al. (2016). HIV-1-Specific CD8 T Cells Exhibit Limited Cross-Reactivity during Acute Infection. J. Immunol. *196*, 3276–3286.

Ellis, J.M., Henson, V., Slack, R., Ng, J., Hartzman, R.J., and Katovich Hurley, C. (2000). Frequencies of HLA-A2 alleles in five U.S. population groups. Predominance Of A*02011 and identification of HLA-A*0231. Hum. Immunol. *61*, 334–340.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486–501.

Ferretti, A.P., Kula, T., Wang, Y., Nguyen, D.M.V., Weinheimer, A., Dunlap, G.S., Xu, Q., Nabilsi, N., Perullo, C.R., Cristofaro, A.W., et al. (2020). Unbiased Screens Show CD8$^+$ T Cells of COVID-19 Patients Recognize Shared Epitopes in SARS-CoV-2 that Largely Reside outside the Spike Protein. Immunity *53*, 1095–1107.e3.

Geldmacher, C., Metzler, I.S., Tovanabutra, S., Asher, T.E., Gostick, E., Ambrozak, D.R., Petrovas, C., Schuetz, A., Ngwenyama, N., Kijak, G., et al. (2009). Minor viral and host genetic polymorphisms can dramatically impact the biologic outcome of an epitope-specific CD8 T-cell response. Blood *114*, 1553–1562.

Giudicelli, V., Brochet, X., and Lefranc, M.P. (2011). IMGT/V-QUEST: IMGT standardized analysis of the immunoglobulin (IG) and T cell receptor (TR) nucleotide sequences. Cold Spring Harb. Protoc. *2011*, 695–715.

Gonzalez-Galarza, F.F., McCabe, A., Santos, E.J.M.D., Jones, J., Takeshita, L., Ortega-Rivera, N.D., Cid-Pavon, G.M.D., Ramsbottom, K., Ghattaoraya, G., Alfirevic, A., et al. (2020). Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Res. *48* (D1), D783–D788.

Goulder, P.J., Phillips, R.E., Colbert, R.A., McAdam, S., Ogg, G., Nowak, M.A., Giangrande, P., Luzzi, G., Morgan, B., Edwards, A., et al. (1997). Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nat. Med. *3*, 212–217.

Grant, E.J., Josephs, T.M., Loh, L., Clemens, E.B., Sant, S., Bharadwaj, M., Chen, W., Rossjohn, J., Gras, S., and Kedzierska, K. (2018). Broad CD8$^+$ T cell cross-recognition of distinct influenza A strains in humans. Nat. Commun. *9*, 5427.

Gras, S., Saulquin, X., Reiser, J.B., Debeaupuis, E., Echasserieau, K., Kissenpfennig, A., Legoux, F., Chouquet, A., Le Gorrec, M., Machillot, P., et al. (2009). Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope. J. Immunol. *183*, 430–437.

Gras, S., Kedzierski, L., Valkenburg, S.A., Laurie, K., Liu, Y.C., Denholm, J.T., Richards, M.J., Rimmelzwaan, G.F., Kelso, A., Doherty, P.C., et al. (2010). Cross-reactive CD8+ T-cell immunity between the pandemic H1N1-2009 and H1N1-1918 influenza A viruses. Proc. Natl. Acad. Sci. USA *107*, 12599–12604.

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

Habel, J.R., Nguyen, T.H.O., van de Sandt, C.E., Juno, J.A., Chaurasia, P., Wragg, K., Koutsakos, M., Hensen, L., Jia, X., Chua, B., et al. (2020). Suboptimal SARS-CoV-2-specific CD8$^+$ T cell response associated with the prominent HLA-A*02:01 phenotype. Proc. Natl. Acad. Sci. USA *117*, 24384–24391.

Iglesias, M.C., Almeida, J.R., Fastenackels, S., van Bockel, D.J., Hashimoto, M., Venturi, V., Gostick, E., Urrutia, A., Wooldridge, L., Clement, M., et al. (2011). Escape from highly effective public CD8+ T-cell clonotypes by HIV. Blood *118*, 2138–2149.

Jarjour, N.N., Masopust, D., and Jameson, S.C. (2021). T Cell Memory: Understanding COVID-19. Immunity *54*, 14–18.

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

Kared, H., Redd, A.D., Bloch, E.M., Bonny, T.S., Sumatoh, H., Kairi, F., Carbajo, D., Abel, B., Newell, E.W., Bettinotti, M.P., et al. (2020). CD8+ T cell responses in convalescent COVID-19 individuals target epitopes from the entire SARS-CoV-2 proteome and show kinetics of early differentiation. bioRxiv. https://doi.org/10.1101/2020.10.08.330688.

Karlsson, A.C., Humbert, M., and Buggert, M. (2020). The known unknowns of T cell immunity to COVID-19. Sci. Immunol. *5*, eabe8063.

Kløverpris, H.N., Cole, D.K., Fuller, A., Carlson, J., Beck, K., Schauenburg, A.J., Rizkallah, P.J., Buus, S., Sewell, A.K., and Goulder, P. (2015). A molecular switch in immunodominant HIV-1-specific CD8 T-cell epitopes shapes differential HLA-restricted escape. Retrovirology *12*, 20.

Le Bert, N., Tan, A.T., Kunasegaran, K., Tham, C.Y.L., Hafezi, M., Chia, A., Chng, M.H.Y., Lin, M., Tan, N., Linster, M., et al. (2020). SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. Nature *584*, 457–462.

Lineburg, K.E., Srihari, S., Altaf, M., Swaminathan, S., Panikkar, A., Raju, J., Crooks, P., Ambalathingal, G.R., Martins, J.P., Matthews, K.K., et al. (2020). Rapid detection of SARS-CoV-2-specific memory T-cell immunity in recovered COVID-19 cases. Clin. Transl. Immunology *9*, e1219.

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

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Cryst. *40*, 658–674.

Nelde, A., Bilich, T., Heitmann, J.S., Maringer, Y., Salih, H.R., Roerden, M., Lübke, M., Bauer, J., Rieth, J., Wacker, M., et al. (2021). SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition. Nat. Immunol. *22*, 74–85.

Ng, K.W., Faulkner, N., Cornish, G.H., Rosa, A., Harvey, R., Hussain, S., Ulferts, R., Earl, C., Wrobel, A.G., Benton, D.J., et al. (2020). Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. Science *370*, 1339–1343.

Peng, Y., Mentzer, A.J., Liu, G., Yao, X., Yin, Z., Dong, D., Dejnirattisai, W., Rostron, T., Supasa, P., Liu, C., et al.; Oxford Immunology Network Covid-19 Response T cell Consortium; ISARIC4C Investigators (2020). Broad and strong memory CD4$^+$ and CD8$^+$ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. Nat. Immunol. *21*, 1336–1345.

Rowntree, L.C., Nguyen, T.H.O., Halim, H., Purcell, A.W., Rossjohn, J., Gras, S., Kotsimbos, T.C., and Mifsud, N.A. (2018). Inability To Detect Cross-Reactive Memory T Cells Challenges the Frequency of Heterologous Immunity among Common Viruses. J. Immunol. *200*, 3993–4003.

Rowntree, L.C., Nguyen, T.H.O., Farenc, C., Halim, H., Hensen, L., Rossjohn, J., Kotsimbos, T.C., Purcell, A.W., Kedzierska, K., Gras, S., and Mifsud, N.A.

(2020). A Shared TCR Bias toward an Immunogenic EBV Epitope Dominates in HLA-B*07:02-Expressing Individuals. J. Immunol. *205*, 1524–1534.

Schulien, I., Kemming, J., Oberhardt, V., Wild, K., Seidel, L.M., Killmer, S., Sagar, Daul, F., Salvat Lago, M., Decker, A., et al. (2021). Characterization of pre-existing and induced SARS-CoV-2-specific $CD8^+$ T cells. Nat. Med. *27*, 78–85.

Sekine, T., Perez-Potti, A., Rivera-Ballesteros, O., Strålin, K., Gorin, J.B., Olsson, A., Llewellyn-Lacey, S., Kamal, H., Bogdanovic, G., Muschiol, S., et al.; Karolinska COVID-19 Study Group (2020). Robust T Cell Immunity in Convalescent Individuals with Asymptomatic or Mild COVID-19. Cell *183*, 158–168.e14.

Sette, A., and Crotty, S. (2020). Pre-existing immunity to SARS-CoV-2: the knowns and unknowns. Nat. Rev. Immunol. *20*, 457–458.

Sette, A., and Sidney, J. (1999). Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics *50*, 201–212.

Shomuradova, A.S., Vagida, M.S., Sheetikov, S.A., Zornikova, K.V., Kiryukhin, D., Titov, A., Peshkova, I.O., Khmelevskaya, A., Dianov, D.V., Malasheva, M., et al. (2020). SARS-CoV-2 Epitopes Are Recognized by a Public and Diverse Repertoire of Human T Cell Receptors. Immunity *53*, 1245–1257.e5.

Smith, C., Gras, S., Brennan, R.M., Bird, N.L., Valkenburg, S.A., Twist, K.A., Burrows, J.M., Miles, J.J., Chambers, D., Bell, S., et al. (2014). Molecular imprint of exposure to naturally occurring genetic variants of human cytomegalovirus on the T cell repertoire. Sci. Rep. *4*, 3993.

Snyder, T.M., Gittelman, R.M., Klinger, M., May, D.H., Osborne, E.J., Taniguchi, R., Zahid, H.J., Kaplan, I.M., Dines, J.N., Noakes, M.N., et al. (2020). Magnitude and Dynamics of the T-Cell Response to SARS-CoV-2 Infection at Both Individual and Population Levels. medRxiv. https://doi.org/10.1101/2020.07.31.20165647.

Solberg, O.D., Mack, S.J., Lancaster, A.K., Single, R.M., Tsai, Y., Sanchez-Mazas, A., and Thomson, G. (2008). Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. Hum. Immunol. *69*, 443–464.

Szeto, C., Chatzileontiadou, D.S.M., Nguyen, A.T., Sloane, H., Lobos, C.A., Jayasinghe, D., Halim, H., Smith, C., Riboldi-Tunnicliffe, A., Grant, E.J., and Gras, S. (2021). The presentation of SARS-CoV-2 peptides by the common HLA-A*02:01 molecule. iScience *24*, 102096.

Wang, G.C., Dash, P., McCullers, J.A., Doherty, P.C., and Thomas, P.G. (2012). T cell receptor αβ diversity inversely correlates with pathogen-specific antibody levels in human cytomegalovirus infection. Sci. Transl. Med. *4*, 128ra42.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD107a-FITC	BD Biosciences/eBioscience	Cat# 555800; RRID:AB_396134
CD107a-AF488	Thermofisher	Cat# 53-1079-42; RRID:AB_2016657
CD8-PerCP-Cy5.5	BD Biosciences/eBioscience	Cat# 565310; RRID:AB_2687497
CD4-PE-Cy7	BD Biosciences	Cat# 560649; RRID:AB_1727475
CD4-Pacific Blue	BD Biosciences	Cat# 558116; AB_397037
CD4-BUV395	BD Biosciences	Cat# 563550; AB_2738273
Live/Dead Fixable Near-IR Dead Cell Stain	Life Technologies	Cat# L34975
IFN-γ-AF700	BD Biosciences	Cat# 557995; RRID:AB_396977
IFN-γ-PE	BD Biosciences	Cat# 554701; RRID:AB_395518
IFN-γ-V421	BD Biosciences	Cat# 562988; 2737934
CD14-APCH7	BD Biosciences	Cat# 560180 RRID:AB_1645464
CD19-APCH7	BD Biosciences	Cat# 560727; RRID:AB_1727437
SARS-CoV/SARS-CoV-2 Nucleocapsid Antibody, Rabbit mAb	Sino Biological	40143-R001; RRID:AB_2827974
Goat anti-Rabbit IgG (H+L) Cross-adsorbed Secondary Ab, HRP	Thermo Fischer	A16104; RRID:AB_2534776
Bacterial and virus strains		
E. coli BL21	ATCC	N/A
SARS-CoV-2 virus	Qld Health	hCoV-19/Australia/QLD02/2020
Biological samples		
Donor-derived PBMCs	QIMR, QLD, Australia	N/A
Human Sera samples (heat-treated)	Mater Hospital, QLD, Australia	N/A
Donor-derived PBMCs	Monash University, VIC, Australia	N/A
Donor-derived PBMCs	Blood bank, VIC, Australia	N/A
Chemicals, peptides, and recombinant proteins		
RPMI-1640	Thermofisher	Cat# 21870092
MEM nonessential amino acid solution	Thermofisher	Cat# 11140050
HEPES	Life Technologies	Cat# 15630080
l-glutamine	GIBCO	Cat# 25030149
penicillin/streptomycin	Life Technologies	Cat# 15140122
FCS	Scientifix	Cat# SFBS-NZ
IL-2	BD Biosciences	Cat# 559334
GolgiPlug Protein Trnsp Inhb	BD Biosciences	Cat# 555029
Cytofix/Cytoperm W/Golgi Stop Kit	BD Biosciences	Cat# 554715
Paraformaldehyde (BD cytofix)	BD Biosciences	Cat# 554655
Triton X-100	Astral	Cat# 786-514
ExoSAP	Life Technologies	Cat# 78201.1.ML
Urea	Thermofisher	Cat# AJA817-5KG
L-Arginine	Sigma	Cat# A5131-1KG
Tris-HCl	Thermofisher	Cat# FSBBP152-5
EDTA	BDH	Cat# BDH9232-500G
glutathione reduced	Goldbio	Cat# G-155-500
glutathione oxidised	Goldbio	Cat# G-060-25
SYPRO orange	Thermofisher	Cat# S6650

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
NaCl	Thermofisher	Cat# CHE700/NACL-25KG
NH4SO4	Astral	Cat# BIOADB0060-500 g
PEG3350	Sigma	Cat# 88276-1KG-F
Potassium Iodide	Sigma	Cat # 221945-500G
Ethylene glycol	Sigma	Cat# 293237-1L
Minimum Essential Media	GIBCO	11095080
FCS	GIBCO	16000044
PBS	GIBCO	003002
Triton X-100	Sigma Aldrich	T8787
Tween-20	Sigma	003005
2,2′-Azino-bis(3-ethylbenzothiazoline-6-sulfonic acid) diammonium salt	Sigma Aldrich	A1888
SARS-CoV-2 antigen overlapping peptide pools POOLS CoV1-5 contain: (AP31, NS6, ORF10, ORF9B, NS7A, NS7B, NS8, Y14, NCAP, VEMP, VME1, Spike Glycoprotein)	JPT	PM-WCPV-AP3A-1 PM-WCPV-NS6-1 PM-WCPV-ORF10-1 PM-WCPV-ORF9B-1 PM-WCPV-NS7A-1 PM-WCPV-NS7B-1 PM-WCPV-NS8-1 PM-WCPV-Y14-1 PM-WCPV-NCAP-1 PM-WCPV-VEMP-1 PM-WCPV-WME1-1 PM-WCPV-S-1
SPR, SPK, PPR, LPR peptides	Genscript	N/A
Critical commercial assays		
VILO cDNA synthesis kit	Life Technologies	Cat# 11754250
Deposited data		
Crystal structure of HLA-B7-SPR	This paper	7LGD (PDB code)
Crystal structure of HLA-B7-SPK	This paper	7LGT (PDB code)
TCR sequences deposited in Mendeley Data	This paper	https://doi.org/10.17632/p34mzy8jfx.1
Experimental models: cell lines		
Vero cells	ATCC	ATCC® CRL-1586
Oligonucleotides		
Primer for TCR alpha and beta chain sequencing	Sigma	N/A
Recombinant DNA		
Human beta 2 microglobulin	genscript	N/A
HLA-B*07:02 soluble fraction (1-275 residues)	genscript	N/A
Software and algorithms		
FlowJo software (TreeStar)	FlowJo, LLC	N/A
FACSDiva software	BD Biosciences	N/A
FinchTV	Geospiza	N/A
IMGT software	IMGT	N/A
GraphPad Prism 8 (version 8.4.2)	Graphpad	N/A
XDS	XDS	N/A
PHASER	CCP4	N/A
CCP4 suite	CCP4	N/A
COOT	Coot	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BUSTER	BUSTER	N/A
Pymol	Schrödinger	N/A
Nonlinear regression analysis, log (inhibitor) versus normalized response	Graphpad	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and materials should be directed to the Lead Contact, Prof. Stephanie Gras (S.Gras@latrobe.edu.au)

Materials availability

Materials are available upon reasonable request.

Data and code availability

The final crystal structure models for the peptide-HLA-B*07:02 complexes have been deposited to the Protein Data Bank (PDB) under the following accession codes: 7LGD for HLA-B7-SPR and 7LGT for HLA-B7-SPK.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study participants

This study was performed according to the principles of the Declaration of Helsinki. Ethics approval to undertake the research was obtained from the QIMR Berghofer Medical Research Institute Human Research Ethics Committee and Monash University Human Research Ethics Committee. COVID-19-recovered donors were over the age of 18, had been clinically diagnosed by PCR with SARS-CoV-2 infection, and had subsequently been released from isolation following resolution of symptomatic infection. A total of 37 participants were recruited in May and June 2020 from the south-east region of Queensland, Australia. The majority of participants were returned overseas travelers. Participants ranged in age from 20 to 75, 14 were male and 23 were female, and were a median of 62 (46 – 124) days post-initial diagnosis. Blood samples were collected from all participants to isolate peripheral blood mononuclear cells (PBMCs) to assess SARS-CoV-2 immunity. Healthy donors over the age of 18, with no known COVID-19 infection or exposure, were recruited. These donors are referred to as unexposed throughout the manuscript. A total of 17 unexposed donors were recruited, ranging in age from 19 to 56 (average of 33 years old), 9 were male, 7 were female and 1 was undetermined. Informed consent was obtained from all participants. The HLA typing was performed by AlloSeq Tx17 (CareDx Pty Ltd, Fremantle, Australia), or Australian Red Cross Victorian Transplant and Immunogenetics Service (Melbourne, Australia), or PathWest Laboratory Medicine, Fiona Stanley Hospital using AllType NGS high resolution typing on the IonTorrent NGS platform, and these details are provided in Table S1.

METHOD DETAILS

Peripheral blood mononuclear cells

PBMCs were separated from whole blood or buffy coats using density gradient centrifugation. PBMCs were used fresh or were cryogenically stored until use.

Generation of peptide-specific CD8$^+$ T cell lines

CD8$^+$ T cell lines were generated as previously described (Grant et al., 2018; Lineburg et al., 2020). Briefly, PBMCs were incubated with 10 μM SARS-CoV-2 overlapping peptide pools, or 10 μM of individual peptides and cultured for 10-14 days in RPMI-1640 supplemented with 2 mM MEM nonessential amino acid solution (Sigma), 100 mM HEPES (Sigma), 2 mM L-glutamine (Sigma), penicillin/streptomycin (Life Technologies), 50 mM 2-ME (Sigma) and 10% heat-inactivated (FCS; Scientifix). Cultures were supplemented with 10IU IL-2 2-3 times weekly. CD8$^+$ T cell lines were used freshly harvested, or were cryogenically stored for subsequent analysis.

Intracellular cytokine assay

CD8$^+$ T cell lines were stimulated with cognate SARS-CoV-2 antigen overlapping peptide pools, or 10 μM individual peptides and were incubated for 4-5 hours in the presence of GolgiPlug (BD Biosciences), GolgiStop (BD Biosciences) and anti-CD107a-FITC or -AF488 (BD Biosciences/eBioscience). Following stimulation, cells were surface stained for 30 mins with anti-CD8-PerCP-Cy5.5 (eBioscience/BD Biosciences), anti-CD4-PE-Cy7 or -Pacific Blue or -BUV395 (all BD Biosciences) and Live/Dead Fixable Near-IR Dead Cell Stain (Life Technologies). Cells were fixed and permeabilised using BD Cytofix/Cytoperm solution (BD

Biosciences) and then intracellularly stained with anti-IFN-γ-AF700 or -PE or -V421 (all from BD Biosciences) as well as anti-TNF-PE-Cy7, and IL2-PE (all BD Biosciences) for a further 30 minutes. Cells were acquired on a BD LSRFortessa with FACSDiva software. Post-acquisition analysis was performed using FlowJo software (TreeStar). Cytokine detection levels identified in the no-peptide control condition were subtracted from the corresponding test conditions in all summary graphs to account for non-specific, spontaneous cytokine production.

Multimer staining

$CD8^+$ T cell lines were multimer stained for 1 hr at room temperature. Cells were washed and surface stained with anti-CD8-PerCP-Cy5.5 (BD Biosciences), anti-CD4-BUV395 (BD Biosciences) anti-CD14-APCH7, anti-CD19-APCH7 and Live/Dead Fixable Near-IR Dead Cell Stain (Life Technologies). Cells were either fixed with 1% paraformaldehyde and acquired on the BD LSR Fortessa, or were directly single-cell sorted into PCR plates (Eppendorf) using a BD Aria Fusion. Plates were centrifuged, and stored at −80°C until use.

Single-cell multiplex PCR

Single-cell multiplex PCR was carried out as previously described (Grant et al., 2018; Wang et al., 2012). Briefly, cDNA was generated using the VILO cDNA synthesis kit (Invitrogen) at 1/20 of the manufacturer's recommendation with 0.1% Triton X. Nested PCR comprising 40 α- and 27 β-chains was subsequently undertaken. PCR products were purified using ExoSAP (GE Healthcare) and were sequenced at AGRF (Melbourne, Australia) or The QIMR Berghofer Sequencing Facility (Brisbane, Australia). Sequences were analyzed using FinchTV (Geospiza) and IMGT software (Brochet et al., 2008; Giudicelli et al., 2011). CDR3 sequences shown are all productive (no stop codons).

Protein expression, refold and purification

DNA plasmids encoding HLA-B7 α chain and β-2-microglobulin were transformed separately into a BL21 strain of *E. coli*. Recombinant proteins were expressed individually, where inclusion bodies were extracted and purified from the transformed *E. coli* cells. Soluble pHLA complexes were produced by refolding 30 mg of HLA-B7 α chain with 10 mg of β-2-microglobulin and 5 mg of peptide (Genscript) into a buffer of 3M Urea, 0.5 M L-Arginine, 0.1 M Tris-HCl pH 8.0, 2.5 mM EDTA pH 8.0, 5 mM glutathione (reduced), 1.25 mM glutathione (oxidised). The refold mixture was dialysed into 10 mM Tris-HCl pH 8.0 and soluble pHLA was purified using anion exchange chromatography using a HiTrapQ column (GE Healthcare).

Differential scanning fluorimetry

Differential Scanning fluorimetry was performed in a QIAGEN RG6 real-time PCR machine, with pHLA samples heated from 30 to 95°C at a rate of 0.5°C/min using a default excitation and emission channel set to yellow (excitation of ~530 nm and detection at ~557 nm). The experiment was set up using two concentrations of pHLA (5 μM and 10 μM), each in duplicate. Each sample was dialysed in 10mM Tris-HCl pH 8.0, 150mM NaCl and contained a final concentration of 10X SYPRO Orange Dye. Fluorescence intensity data was normalized and plotted using GraphPad Prism 8 (version 8.4.2). The T_m value is determined by the temperature when 50% of maximum fluorescence intensity is reached, and summarized in Table 1.

Crystallization and structural determination

Crystals of pHLA complexes were grown via sitting-drop, vapor diffusion method at 20°C with a protein: reservoir drop ratio of 1:1, at a concentration of 7 mg/mL in 10 mM Tris-HCl pH 8.0, 150 mM NaCl. Crystals of HLA-B7 in complex with SARS-CoV-2 $N_{105-113}$ (SPRWYFYYL) were grown in 2M ammonium sulfate, 0.1M HEPES pH 7.5; or with 229E $N_{105-113}$ (SPKLHFYYL) were grown in 18% PEG3350, 0.2 M KI. These crystals were soaked in a cryoprotectant containing mother liquor and 20% EG or 30% PEG3350 (w/v) and then flash-frozen in liquid nitrogen. The data were collected on the MX2 beamline at the Australian Synchrotron, part of ANSTO, Australia (Aragão et al., 2018). The data were processed using XDS (Kabsch, 2010) and the structures were determined by molecular replacement using the PHASER program (McCoy et al., 2007) from the CCP4 suite (Collaborative Computational Project, Number 4, 1994) with a model of HLA-B7 without the peptide (derived from PDB ID: 5WMN; Rowntree et al., 2018). Manual model building was conducted using COOT (Emsley et al., 2010) followed by refinement with BUSTER (Bricogne et al., 2011). The final model has been validated using the wwPDB OneDep System with the accession number of 7LGD for HLA-B7-SPR and 7LGT for HLA-B7-SPK structures. The final refinement statistics are summarized in Table S6. All molecular graphics representations were created using PyMOL.

Model building

Model building of the structure of HLA-B7-LPR complex was performed using the crystal structure of HLA-B7-SPR as a starting model. The SPR peptide P1-Ser residue was mutated into a P1-Leu residue using the crystallographic software, COOT (Emsley et al., 2010), where the side chain rotamer was selected based on the least steric clashes, as evaluated using MolProbity.

SARS-CoV-2 microneutralization assay

Vero cells were cultured in 96 well plates. Convalescent serum harvested from COVID-19-recovered patients was heat-treated at 56°C for 1 hour. The sera was then serial diluted with minimum essential medium (MEM) (GIBCO) supplemented with 2% FCS. In physical containment 3 settings, the diluted sera were incubated with the SARS-CoV-2 (QLD/02; MOI 1) for 1 hour at room

temperature. The serum-virus mixture was transferred to the cultured vero cells and further incubated for 1 hour at room temperature for infection. The inoculum was then removed and replaced with MEM (GIBCO) supplemented with 2% FCS (GIBCO). The cells were incubated at 37°C for 72 hours. The cells were fixed with 10% formaldehyde for 24 hours.

Cells were permeabilised with PBS containing 0.1% Trition-X-100 (Sigma) for 15mins at room temperature. The plates were blocked using PBS (Sigma) supplemented with 0.1% Tween-20 (Sigma) and 3% skim milk for 1 hour at room temperature. The cells were primarily stained with rabbit anti-NP (1:3000)(Sino Biological) for 1 hour at room temperature. The plates were washed with PBS (Sigma) supplemented with 0.1% Tween-20 (Sigma). The cells were stained using goat anti-rabbit immunoglobulin horse-radish protein (HRP) (Thermo Fischer) (1:3000) for 1 hour at room temperature. The plates were washed with PBS (Sigma) supplemented with 0.1% Tween-20 (Sigma). The plates were incubated with 2,2′-azino-bis(3-ethylbenzothiazoline-6-sulfonic acid (ABTS) (Thermo Fischer) substrate for color development and incubated for 20 minutes. Measurements were done at 405 nm.

Sequence alignment of SPR and homologs from coronaviruses isolates

Complete full length protein sequences for the SARS-CoV-2 (taxid ID 2697049, 419 amino acids), SARS-CoV-1 (taxid ID 694069, 422 amino acids), OC43 (taxid ID 31631, 448 amino acids), HKU-1 (taxid ID 290028, 441 amino acids), 229E (taxid ID 11137, 389 amino acids) and NL63 (taxid ID 277944, 377 amino acids) coronaviruses were obtained from the NCBI virus database http://www.ncbi.nlm.nih.gov/labs/virus on the 10th of November 2020. Sequences were aligned using https://www.fludb.org/brc/home.spg?decorator=influenza. All sequences were used except a single SARS-CoV-2 North American sequence that was unable to be aligned, and a single SARS-CoV-2 Europe sequence with an unknown amino acid (denoted as X) within the peptide sequence.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

GraphPad Prism 8.2.1 (San Diego, CA) was used to perform statistical analysis. Statistical comparisons between participant groups (unexposed and COVID-19-recovered) were made using unpaired Mann-Whitney *U* Wilcoxon rank-sum tests. $p < 0.05$ was considered statistically significant.

Immunity, Volume *54*

Supplemental information

$CD8^+$ T cells specific for an immunodominant SARS-CoV-2 nucleocapsid epitope cross-react with selective seasonal coronaviruses

Katie E. Lineburg, Emma J. Grant, Srividhya Swaminathan, Demetra S.M. Chatzileontiadou, Christopher Szeto, Hannah Sloane, Archana Panikkar, Jyothy Raju, Pauline Crooks, Sweera Rehan, Andrea T. Nguyen, Lea Lekieffre, Michelle A. Neller, Zhen Wei Marcus Tong, Dhilshan Jayasinghe, Keng Yih Chew, Christian A. Lobos, Hanim Halim, Jacqueline M. Burrows, Alan Riboldi-Tunnicliffe, Weisan Chen, Lloyd D'Orsogna, Rajiv Khanna, Kirsty R. Short, Corey Smith, and Stephanie Gras

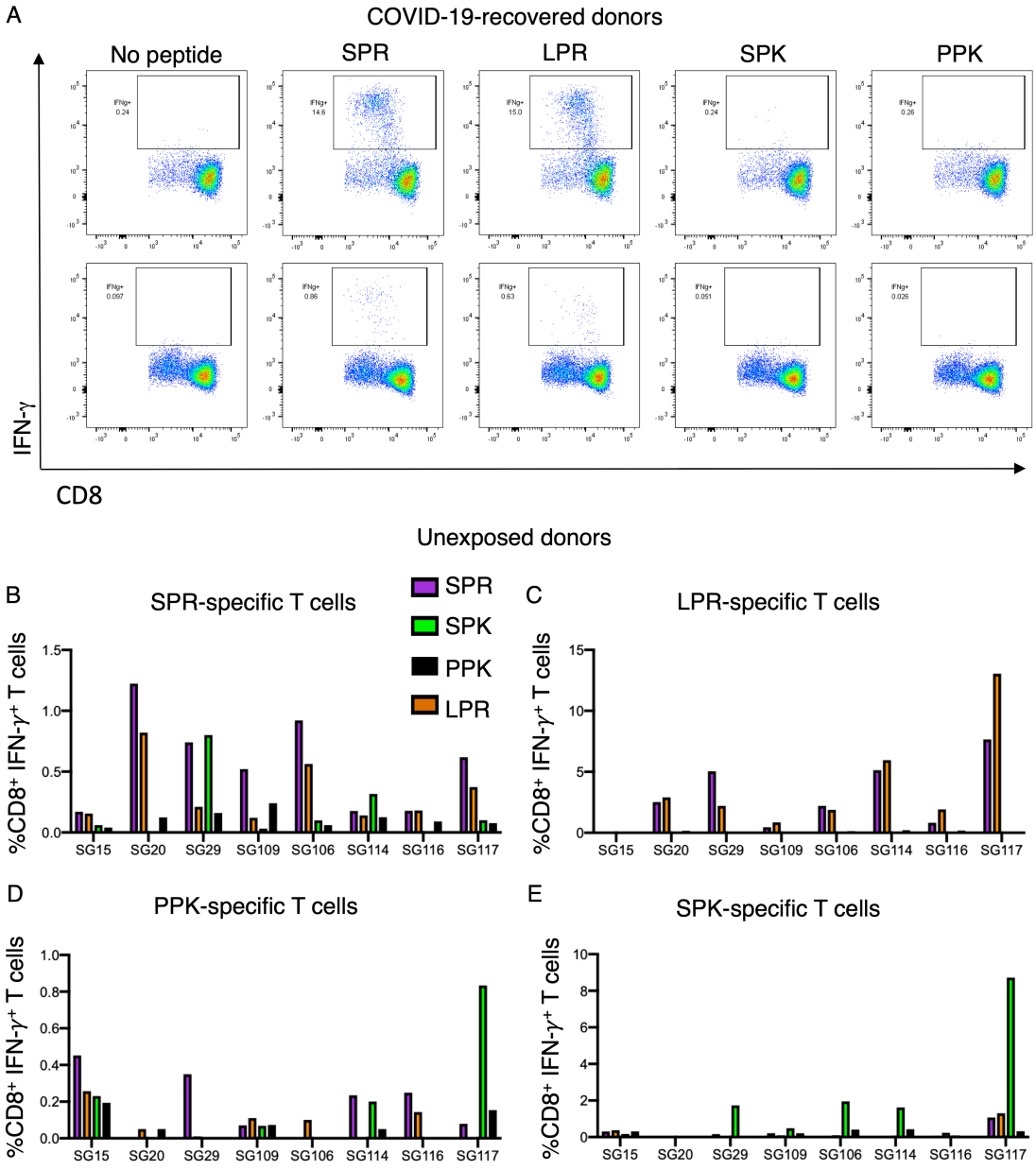


Figure S1. Cross-reactivity of SPR-specific CD8$^+$ T cells in COVID-19-recovered donors and from unexposed donors towards coronavirus peptides, Related to Figure 3.

(A) Following 14 days in culture, SPR-specific T cells were re-stimulated with five separate peptide conditions: no peptide control, cognate SPR, LPR, SPK or PPK and IFN-γ production was measured using an ICS assay. Representative flow cytometry displaying IFN-γ production in CD8$^+$ T cells generated from two separate COVID-19-recovered donors (top and bottom panels). (B-E) Peptide-specific CD8$^+$ T cells were expanded from unexposed donors (n=8) by stimulation with the (B) SPR, (C) LPR, (D) SPK or (E) PPK peptides and then re-stimulated separately with each peptide. The frequency of CD8$^+$ IFN-γ producing T cells was determined using an ICS assay. Summary of CD8$^+$ IFN-γ responses towards each of the variant peptides with the no peptide control subtracted.

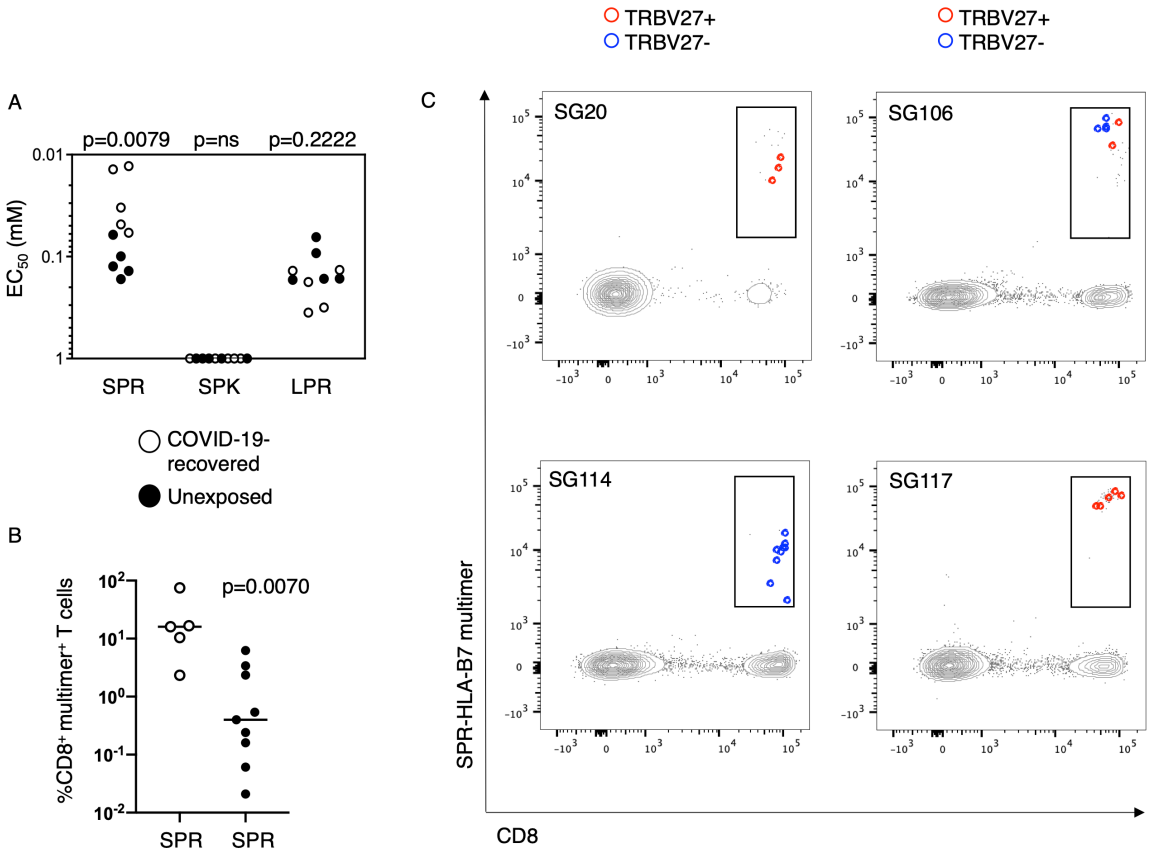


Figure S2. Functional avidity and magnitude of epitope-specific CD8$^+$ T cells from COVID-19-recovered and unexposed donors, and index sorting of SPR-specific T cells from unexposed donors, Related to Figure 3.

(A-B) Peptide-specific CD8$^+$ T cells were expanded from PBMCs derived from unexposed and COVID-19-recovered individuals. (A) Graph comparing the avidity (EC_{50}) of SPR-specific T cells derived from both COVID-19-recovered (white) and unexposed (black) individuals in response to cognate SPR peptide and homologous SPK and LPR peptides measured by IFN-γ production in an ICS assay. (B) Graph comparing the frequency of CD8$^+$ multimer$^+$ T cells between COVID-19 recovered (white) and unexposed (black) donors. The groups were compared using the Mann-Whitney test with $p < 0.05$ for significance, ns: non-significant. (C) SPR-specific CD8$^+$ T cells from unexposed donors were stained with SPR-HLA-B7 multimer and multimer$^+$CD8$^+$ cells were single-cell index sorted. A few clonotypes utilising TRBV27$^+$ (red) or TRBV27$^-$ (blue) overlaid on the total CD3$^+$/CD8$^+$ population are represented to show the wide range of functional avidities of the clones independently of the TRBV27 gene expression.

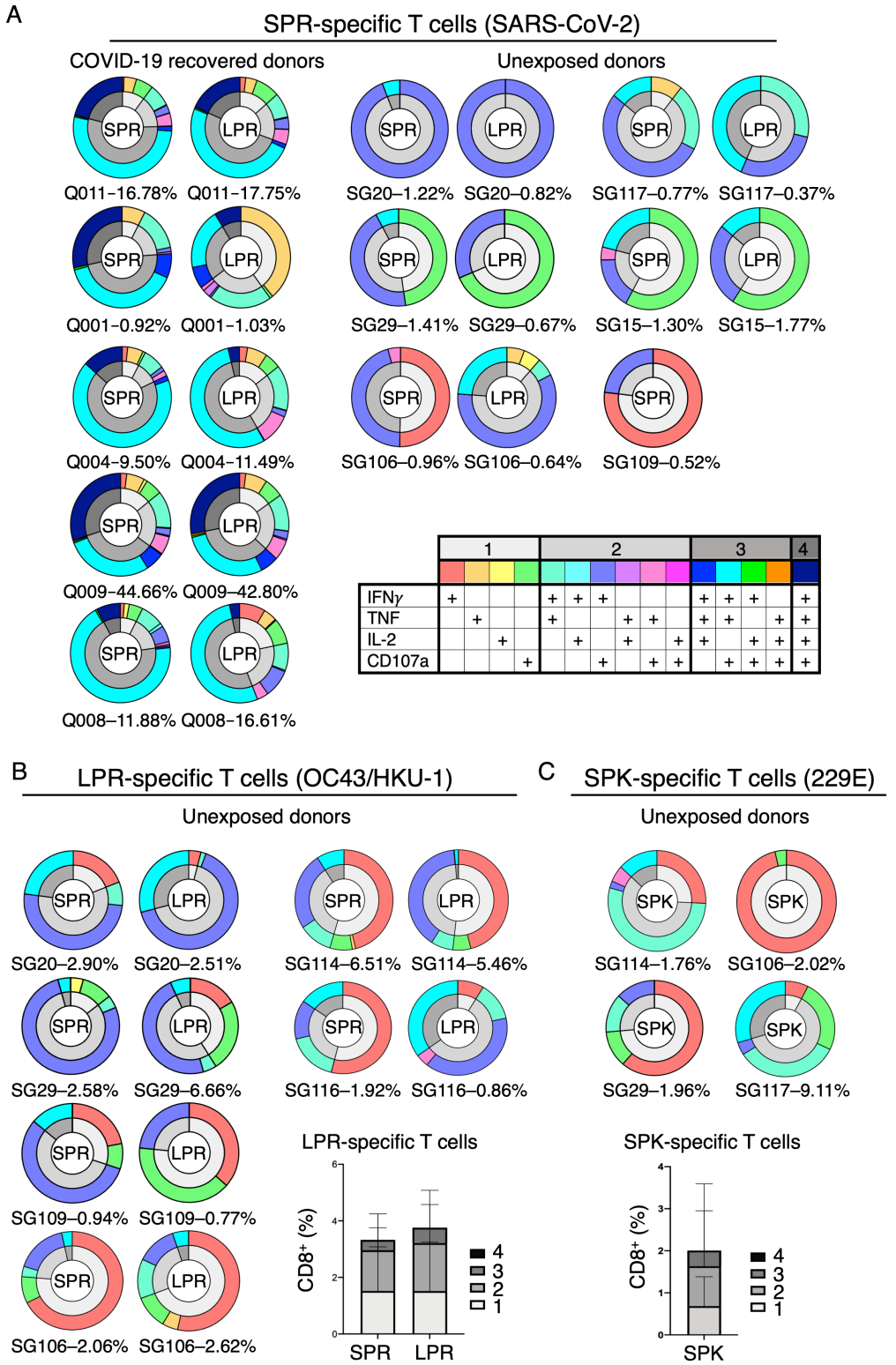


Figure S3. Polyfunctionality of SPR-specific CD8⁺ T cells from COVID-19-recovered and unexposed donors and LPR- and SPK-specific T cells from unexposed donors, Related to Figure 3.

(A) SPR-specific T cell lines were established from COVID-19-recovered (n=5) and unexposed donors (n=6), and re-stimulated individually with the SPR and LPR peptides in an ICS assay. (B) LPR- and (C) SPK-specific T cells were established and re-stimulated individually with the SPR, LPR, and SPK peptides in an ICS assay. The frequency of CD8⁺ T cells with different cytokine production profiles (IFN-γ, TNF,

IL-2 and CD107a) was determined, along with the number of different cytokines produced, minus the no peptide control. The outer ring of the double ring pie shows the cytokine profile in different colours, and the inner ring represents the number of cytokines produced with black for 4 and white for one.

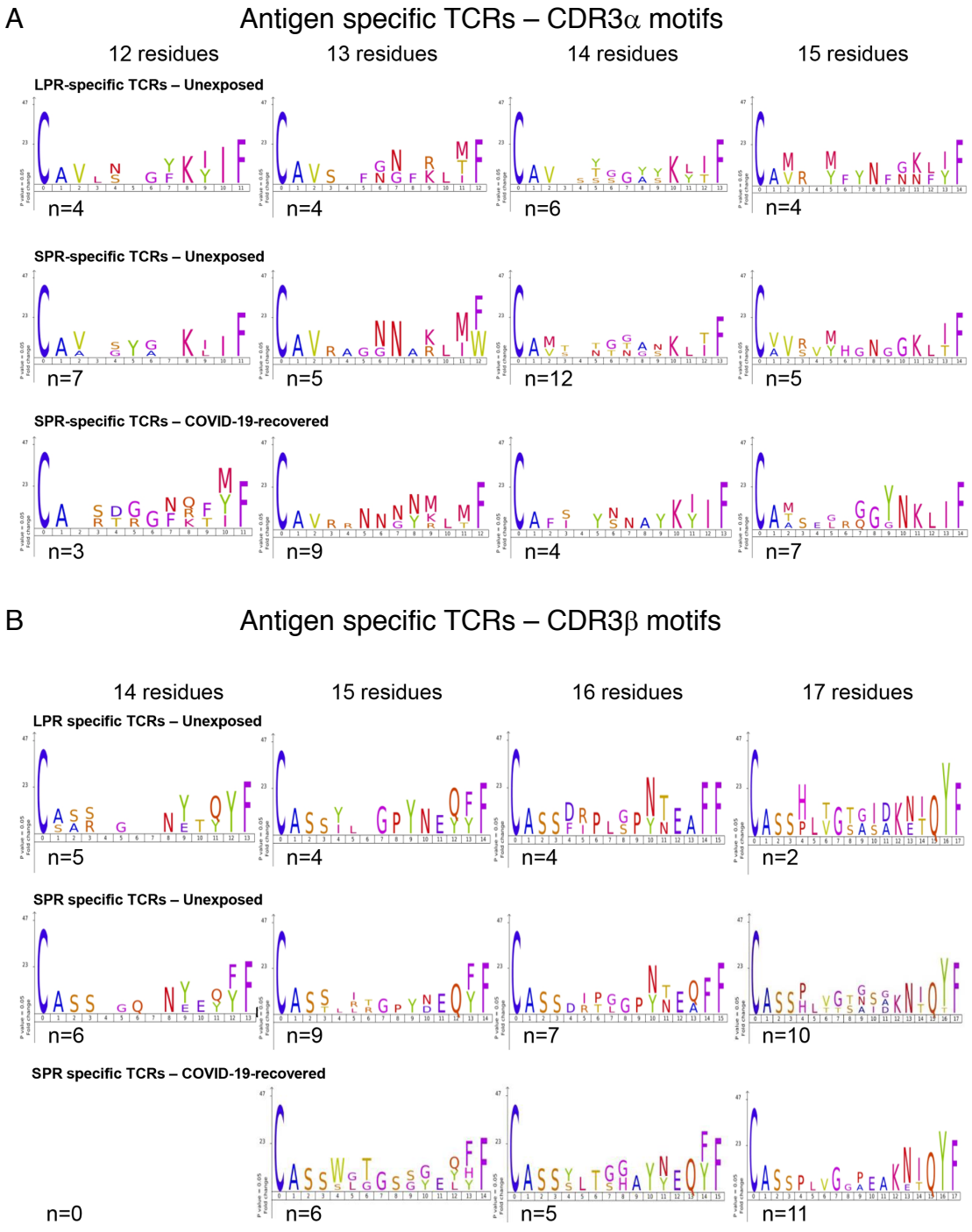


Figure S4. CDR3α and CDR3β analysis of the LPR- and SPR-specific CD8$^+$ T cells, Related to Figure 4.

PBMCs from COVID-19-recovered or unexposed individuals were stimulated with the SPK or LPR peptides and cultured for 10-14 days in the presence of IL2. CD8$^+$ T cell lines were stained with SPK or LPR multimer and multimer$^+$ cells were single-cell sorted and the TCR repertoire was determined by multiplex PCR. Analysis of the motifs of CDR3α and CDR3β sequences from distinct SPR and LPR-specific CD8$^+$ T cell clonotypes in unexposed and COVID-19-recovered individuals is shown.

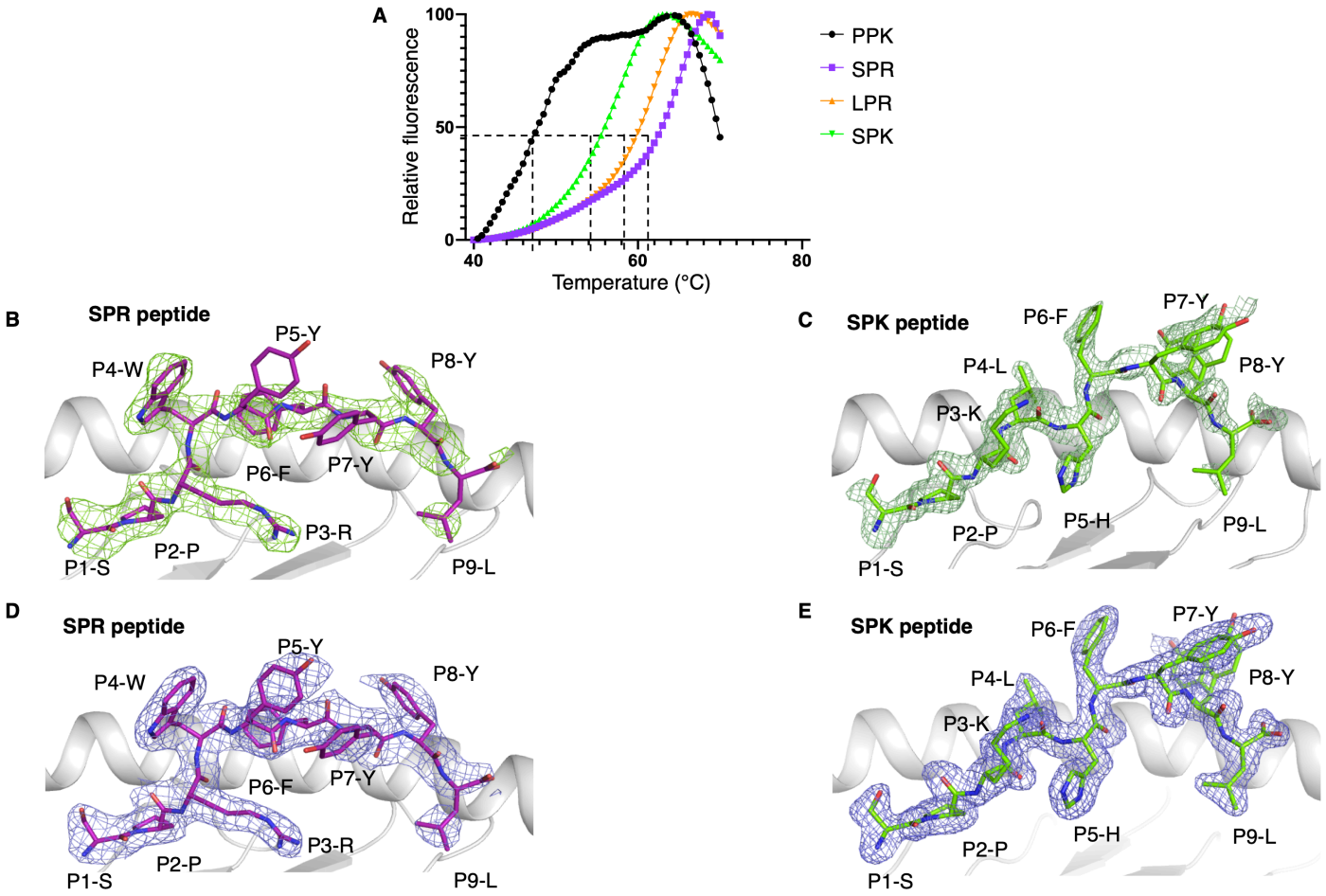


Figure S5. Thermal stability of HLA-B7 presenting coronaviruses peptides and electron density maps of the SPR and SPK peptides presented by HLA-B7, Related to Figure 5.

(**A**) Thermal stability profile of HLA-B7 in complex with coronavirus peptides derived from $N_{105\text{-}113}$ of viral strains SARS-CoV-2 (purple), or its homologous from NL63 (black), 229E (green) and OC43 (orange). Normalised fluorescence intensity follows a sigmoidal curve with an increase in temperature, as SYPRO Orange dye binds to exposed hydrophobic regions of pHLA-B7 during its thermal denaturation. The thermal midpoint temperature is determined at 50% of its normalised fluorescence intensity, allowing for the comparison of stability between pHLA-B7 complexes (summarised in **Table S4**). (**B-E**) HLA-B7 is represented as a white cartoon, the SPR peptide as purple sticks and the SPK peptide as green sticks. The **B** and **C** panels show the 2mFo-Fc electron density map, before building the peptide, contoured at 3σ and coloured in green, while the **D** and **E** panels show the Fo-Fc electron density maps, with the peptides build in and refinement cycles, contoured at 1σ and coloured in blue.

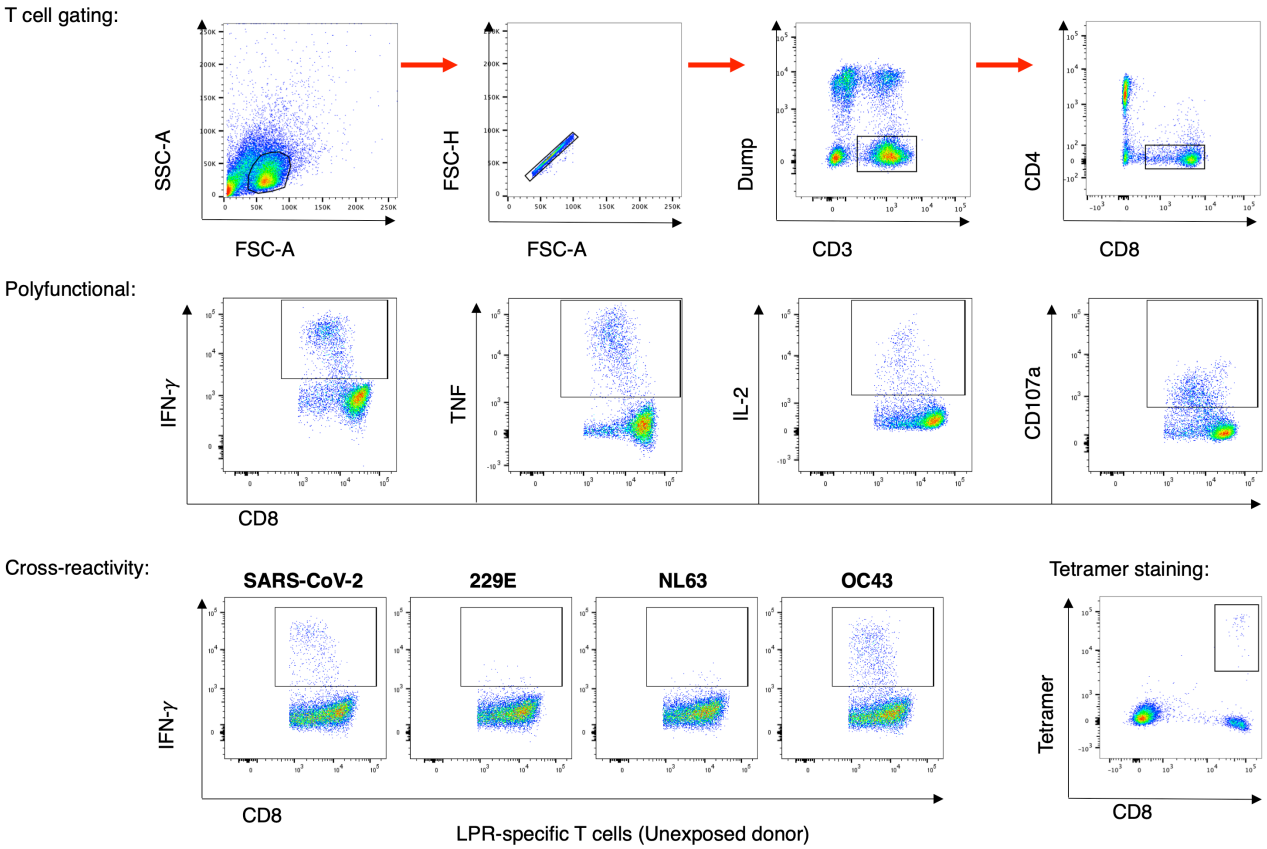


Figure S6. Gating strategy for our study, Related to Figures 1, 2, 3, and 4.
Peptide-specific $CD8^+$ T cells were derived by stimulating PBMCs from unexposed or COVID-19-recovered donors with various peptides as indicated throughout. $CD8^+$ T cell responses were subsequently assessed by either multimer staining or function in an ICS assay. Representative gating for the identification of $CD8^+$ T cells in all assays (top panel). Representative gating for the assessment of polyfunctionality (middle panel) or cross-reactivity (bottom panel, left) in an ICS assay. Representative flow cytometry plot of the identification of $multimer^+$ populations (bottom panel, right).

Table S1. HLA typing of COVID-19-recovered and unexposed donors, Related to Figures 1, 2, 3, and 4.

Donor ID	HLA-A	HLA-B	Age	Sex	Day post CoV PCR test
SG15	A*11:01, A*30:01	B*07:06, B*44:03	27	M	NA
SG20	A*32:01	B*07:02, B*44:02	32	F	NA
SG29	A*02:01, A*03:01	B*07:02, B*35:01	24	M	NA
SG105	A*02:01, A*29:02	B*07:02, B*44:03	N/A	N/A	NA
SG3	A*02:01, A*24:02	B*07:02, B*14:02	32	M	NA
SG34	A*02:01, A*24:02	B*07:02, B*08:01	24	F	NA
SG106	A*02:01, A*26:01	B*07:02, B*55:01	23	F	NA
SG109	A*02:01, A*26:01	B*07:02, B*44:02	19	F	NA
SG114	A*01:01, A*31:01	B*07:02, B*08:01	39	M	NA
SG116	A*03:01, A*29:02	B*07:02, B*44:03	32	F	NA
SG117	A*02:01, A*03:01	B*07:02, B*44:02	38	M	NA
GR001	A*01:01, A*03:01	B*07:02, B*08:01	42	M	NA
GR016	A*02:1	B*07:02	27	M	NA
GR036	A*02:01, A*03:01	B*07:02, B*44:02	40	F	NA
GR067	A*03:01	B*07:02	31	F	NA
GR082	A*01:01, A*02:01	B*07:02, B*44:02	56	M	NA
GR086	A*03:01	B*07:02	45	M	NA
Q001	A*03:01	B*07:02	20	M	58
Q002	A*02:01, A*32:01	B*15:01, B*44:02	53	F	49
Q003	A*01:01, A*25:01	B*08:01, B*35:01	52	M	53
Q004	A*02:01, A*03:01	B*07:02, B*44:02	30	F	51
Q005	A*02:01, A*11:01	B*15:01, B*35:01	33	M	58
Q006	A*02:01, A*24:02	B*07:02, B*18:01	48	M	48
Q007	A*02:01, A*31:01	B*07:02, B*40:01	20	F	53
Q008	A*02:01, A*03:01	B*07:02, B*44:02	48	F	58
Q009	A*03:01, A*26:01	B*07:02, B*27:05	63	F	62
Q010	A*02:01	B*15:01, B*44:03	63	M	49

Q011	A*02:01	B*07:02, B*44:02	59	F	67
Q012	A*02:01, A*03:01	B*44:02, B*51:01	71	F	53
Q013	A*24:02, A*25:01	B*08:01	51	F	54
Q014	A*03:01, A*29:02	B*07:02, B*44:03	61	F	54
Q015	A*02:01	B*35:03	30	F	46
Q016	A*03:03, A*24:03	B*35:01, B*38:01	25	F	71
Q017	A*03:01, A*26:01	B*07:02	58	F	102
Q018	A*02:05, A*30:01	B*13:02, B*49:01	63	M	102
Q019	A*01:01, A*24:02	B*08:01, B*27:05	60	F	59
Q020	A*24:02, A*31:01	B*07:02, B*35:02	58	M	58
Q021	A*01:01, A*24:02	B*08:01, B*35:02	26	F	56
Q022	A*01:01, A*31:01	B*07:02, B*08:01	21	F	57
Q024	A*01:01, A*03:01	B*07:02, B*08:01	56	F	83
Q025	A*01:01, A*02:01	B*08:01, B*40:01	67	F	67
Q026	A*01:01, A*30:02	B*40:01, B*49:01	39	F	61
Q029	A*02:01, A*31:02	B*07:02, B*40:01	31	F	69
Q031	A*02:01, A*29:02	B*15:01, B*44:03	61	M	85
Q032	A*02:01	B*27:05, B*44:02	25	F	70
Q033	A*02:01, A*24:02	B*15:01, B*18:04	25	M	76
Q035	A*02:01, A*03:01	B*07:02, B*08:01	62	F	76
Q038	A*01:01, A*02:01	B*08:01, B*44:02	41	F	88
Q040	A*02:01, A*32:01	B*44:02, B*44:03	74	F	82
Q045	A*01:01, A*03:01	B*07:02, B*14:02	21	F	67
Q046	A*02:01, A*24:02	B*15:01, B*40:01	44	M	80
Q049	A*01:01, A*03:01	B*07:02, B*56:01	51	M	116
Q052	A*01:01, A*24:02	B*07:02, B*44:03	67	M	75
Q062	A*02:01	B*08:01, B*18:01	22	F	124

All donors with ID SG or GR are unexposed donors, all donors with ID Q0 are COVID-19-recovered donors, M: male, F: female.

Table S2. COVID-19 symptoms and classification, Related to Figures 1, 2, 3, and 4.

HLA-B7^{+} donor ID	**Symptom classification**	**Symptoms included**
Q001	Other	FE|CO|SO|HE|LO|SK|GA
Q004	Mild	HE|LO|OT
Q006	Severe	CO|SH|SO|HE|LO|JO|GA
Q007	NA	NA
Q008	Moderate	CO|SH|SO|HE|SK|GA
Q009	Mild	CO
Q011	Severe	SO|HE|LO|OT
Q014	Mild	FE|CO|HE|LO|GA
Q017	Mild	SO|HE|OT
Q020	Moderate	FE|CO|SH|HE|LO|JO|OT
Q022	Mild	CO|SO|LO|JO|OT
Q029	NA	NA
Q035	Moderate	FE|CO|SH|SO|OT
Q045	NA	NA
Q049	Mild	SO|HE
Q052	Mild	OT
HLA-B7^{-} donor ID	**Symptom classification**	**Symptoms included**
Q002	Mild	CO|LO|OT
Q003	Mild	CO|HE|JO
Q005	Severe	CO|SH|SO|HE|JO|OT
Q010	Mild	FE|CO|LO|GA
Q012	NA	NA
Q013	Mild	LO|OT
Q015	NA	NA
Q016	Mild	HE|OT
Q018	Mild	CO|SO|OT
Q019	Mild	SO|HE|LO|JO|OT
Q021	NA	NA
Q025	NA	NA

Q026	Mild	FE|CO|HE|LO|GA|OT
Q031	Mild	FE|CO
Q032	Moderate	FE|CO|SH|HE|LO|GA|OT
Q033	NA	NA
Q038	Moderate	FE|SH|HE|LO|GA
Q040	Severe	FE|HE|LO|JO|GA
Q046	NA	NA
Q062	Mild	LO|OT

FE: fever; CO: cough; SO: Sore throat; SH: Short of breath; HE: headache; LO: Loss of taste or smell; JO: Joint pain; GA: Gastrointestinal symptoms; OT: other; SK: Skin rash.

Table S3. HLA-B7-resticted epitopes defined in SARS-CoV-2 nucleocapsid, Related to Figures 1 and 2.

Peptide origin	Peptide name	Peptide sequence	Number of Responders*
N_{66-74}	FPR	FPRGQGVPI	1 / 17
N_{93-101}	RIR	RIRGGDGKM	4 / 17
$N_{105-113}$	SPR	SPRWYFYYL	14 / 17
$N_{257-265}$	KPR	KPRQKRTAT	1 / 17

*Responders are COVID-19-recovered HLA-B7^{+} individuals.

Table S4. Sequence variation in coronaviruses isolates of the SPR peptide and homologues, Related to Figure 3.

Virus	Sequence	Oceania	Asia	Europe	Africa	North America	South America	Total number of sequences
SARS-CoV-2	SPRWYFYYL	100% (8,693)	100% (1,330)	100% (438)	100% (281)	100% (15,273)	100% (143)	26,158
SARS-CoV-1	SPRWYFYYL	NA	100% (2)	NA	NA	100% (11)	NA	13
OC43	LPRWYFYYL	NA	100% (61)	100% (3)	100% (3)	100% (69)	NA	136
HKU-1	LPRWYFYYL	NA	100% (3)	100% (1)	NA	100% (12)	NA	16
229E	SPKLHFYYL	NA	NA	100% (3)	NA	100% (23)	NA	26
NL63	PPKVHFYYL	NA	100% (8)	NA	100% (7)	100% (43)	NA	58

Complete full length sequences were obtained from the NCBI virus database http://www.ncbi.nlm.nih,gov/labs/virus and were aligned using https://www.fludb.org/brc/home.spg?decorator=influenza. The frequency of peptide conservation is shown in green, with the number of sequences aligned from each geographic region shown in parenthesis. Mutations from the SARS-CoV-2 peptide are denoted in blue, and anchor residues are underlined. NA refers to no sequences being available for that geographic region.

Table S6. Data collection and refinement statistics, Related to Figure 5.

Data Collection Statistics	**HLA-B7-SPR**	**HLA-B7-SPK**
Space group	**$P2_12_12_1$**	**P1**
Cell Dimensions (a,b,c) (Å)	64.45, 107.18, 174.19	57.14, 62.85, 63.01 α=77.15°, β=77.00°, γ=77.86°
Resolution (Å)	46.65 – 2.88 (3.04 – 2.88)	44.53 – 1.97 (2.02 – 1.97)
Total number of observations	191654 (26453)	204851 (14311)
Nb of unique observation	28141 (4027)	56880 (3904)
Multiplicity	6.8 (6.6)	3.6 (3.7)
Data completeness (%)	100 (100)	98.1 (96.8)
I/σ_I	8.0 (1.6)	5.5 (1.7)
R_{pim}[a] (%)	8.0 (52.9)	7.4 (43.8)
$CC_{1/2}$ (%)	89.8 (57.8)	99.0 (55.2)
Refinement Statistics		
R_{factor}[b] (%)	20.6	18.6
R_{free}[b] (%)	26.9	23.0
rmsd from ideality		
Bond lengths (Å)	0.009	0.008
Bond angles (°)	1.038	0.955
Ramachandran plot (%)		
Favoured	96.3	97.9
Allowed	3.6	1.7
Disallowed	0.1	0.4
PBD code	**7LGD**	**7LGT**

[a]$R_{p.i.m} = \Sigma_{hkl} [1/(N-1)]^{1/2} \Sigma_i | I_{hkl,\,i} - \langle I_{hkl} \rangle | / \Sigma_{hkl} \langle I_{hkl} \rangle$. [b]$R_{factor} = \Sigma_{hkl} \; | \, | F_o | - | F_c | \, | / \Sigma_{hkl} | F_o |$ for all data except ≈ 5% which were used for R_{free} calculation. Values in parentheses are for the highest resolution-shell.