Genome-Wide B Cell, CD4⁺, and CD8⁺ T Cell Epitopes That Are Highly Conserved between Human and Animal Coronavirus, Identified from SARS-CoV-2 as Targets for Preemptive Pan-Coronavirus Vaccines


*J Immunol* 2021; 206:2566-2582; Prepublished online 28 April 2021; doi: 10.4049/jimmunol.2001438
http://www.jimmunol.org/content/206/11/2566

Supplementary Material
http://www.jimmunol.org/content/suppl/2021/04/28/jimmunol.2001438.DCSupplemental

References
This article cites 89 articles, 21 of which you can access for free at:
http://www.jimmunol.org/content/206/11/2566.full#ref-list-1

**Why The JI?** Submit online.

- **Rapid Reviews!** 30 days* from submission to initial decision
- **No Triage!** Every submission reviewed by practicing scientists
- **Fast Publication!** 4 weeks from acceptance to publication

*average

Subscription
Information about subscribing to *The Journal of Immunology* is online at:
http://jimmunol.org/subscription

Permissions
Submit copyright permission requests at:
http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts
Receive free email-alerts when new articles cite this article. Sign up at:
http://jimmunol.org/alerts

*The Journal of Immunology* is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852
Copyright © 2021 by The American Association of Immunologists, Inc. All rights reserved.
Print ISSN: 0022-1767 Online ISSN: 1550-6606.
Genome-Wide B Cell, CD4⁺, and CD8⁺ T Cell Epitopes That Are Highly Conserved between Human and Animal Coronaviruses, Identified from SARS-CoV-2 as Targets for Preemptive Pan-Coronavirus Vaccines

Swayam Prakash,⁎,1 Ruchi Srivastava,⁎,1 Pierre-Gregoire Coulon,⁎,1 Nisha R. Dhanushkodi,⁎,1 Aziz A. Chentoufi,⁎ Delia F. Tifrea, † Robert A. Edwards, ‡ Cesar J. Figueroa,§ Sebastian D. Schubl,‡ Lanny Hsieh,§ Michael J. Buchmeier,¶ Mohammed Bouziane,‖ Anthony B. Nesburn,* Baruch D. Kuppermann,* and Lbachir BenMohamed⁎,†,##

Over the last two decades, there have been three deadly human outbreaks of coronaviruses (CoVs) caused by SARS-CoV, MERS-CoV, and SARS-CoV-2, which has caused the current COVID-19 global pandemic. All three deadly CoVs originated from bats and transmitted to humans via various intermediate animal reservoirs. It remains highly possible that other global COVID pandemics will emerge in the coming years caused by yet another spillover of a bat-derived SARS-like coronavirus (SL-CoV) into humans. Determining the Ag and the human B cells, CD4⁺ and CD8⁺ T cell epitope landscapes that are conserved among human and animal coronaviruses should inform in the development of future pan-coronavirus vaccines. In the current study, using several immunoinformatics and sequence alignment approaches, we identified several human B cell and CD4⁺ and CD8⁺ T cell epitopes that are highly conserved in 1) greater than 81,000 SARS-CoV-2 genome sequences identified in 190 countries on six continents; 2) six circulating CoVs that caused previous human outbreaks of the common cold; 3) nine SL-CoVs isolated from bats; 4) nine SL-CoVs isolated from pangolins; 5) three SL-CoVs isolated from civet cats; and 6) four MERS strains isolated from camels. Furthermore, the identified epitopes: 1) recalled B cells and CD4⁺ and CD8⁺ T cells from both COVID-19 patients and healthy individuals who were never exposed to SARS-CoV-2, and 2) induced strong B cell and T cell responses in humanized HLA-DR1/HLA-A*02:01 double-transgenic mice. The findings pave the way to develop a preemptive multipletapepang-coronavirus vaccine to protect against past, current, and future outbreaks. The Journal of Immunology, 2021, 206: 2566–2582.

The first two deadly coronaviruses, the MERS-CoV and the SARS-CoV, originated from bats, their natural hosts and reservoirs, and were transmitted to humans from intermediate animals, namely camels and civet cats, respectively (10, 13–16). The third deadly SARS-CoV-2 appears to be 96% identical to a bat SARS-like coronavirus (SL-CoV) strain termed Bat-CoV-RaTG13 and transmitted to humans from a yet-to-be determined intermediate animal (17, 18). Although human-to-human spread of the common cold coronaviruses occurs frequently, only rarely do animal-to-human coronavirus transmissions occur (19). However, the highly contagious SARS-CoV-2 successfully produces both animal-to-human spread and human-to-human transmission (20, 21). The first known human-to-human transmission of SARS-CoV-2, which causes coronavirus...
disease 2019 (COVID-19), was reported in late January 2020, prompting the World Health Organization and United States authorities to declare a global public health emergency (22).

All human coronaviruses are associated with respiratory illnesses, ranging from mild common colds to more severe lower respiratory tract symptoms (23). Within 2–14 d after SARS-CoV-2 exposure, newly infected individuals may develop fever, fatigue, myalgia, and respiratory symptoms, including cough and shortness of breath (24). Although 40–45% of newly infected individuals remained asymptomatic, 55–60% individuals are symptomatic, ranging from mild/45% to critically ill patients, especially the elderly and those with comorbidities: they develop severe pulmonary inflammatory disease and may need a rapid medical intervention to prevent acute respiratory distress syndrome and death (24, 26). The SARS-CoV-2 infection induces antiviral CD4+ T cells, helping the production of neutralizing/blocking Abs and the formation of effector IFN-γ-producing CD4+ T cells and cytotoxic CD8+ T cells, all arms of immunity critical in reducing viral load in the majority of asymptomatic and converse patients (27–33). Although SARS-CoV-2-specific IgG/IgM Abs and CD4+ and CD8+ T cells are critical to reducing viral infection in a majority of asymptomatic and converse patients (32), an excessive proinflammatory cytokine storm appears to lead to acute respiratory distress syndrome and death in many symptomatic individuals (34–40). Thus, it is crucial to determine the B cell and T cell epitope specificities and the repertoire, phenotype, and function of B cells and CD4+ and CD8+ T cells that are associated with natural resistance seen in asymptomatic patients (41–43). The information in this study should guide in the development of pan-coronavirus vaccines.

In the current study, we identified several human B and CD4+ and CD8+ T cell epitopes that are highly conserved among six strains of coronaviruses previously reported to infect humans and over 81,000 genome sequences of SARS-CoV-2 that currently circulate in 190 countries on six continents. Moreover, as immune targets for preemptive pan-coronavirus vaccines, we identified the epitopes that are common among the above human coronaviruses and 25 animal strains isolated from bats, pangolins, civet cats, and camels. We demonstrated the antigenicity of these epitopes in both SARS-CoV-2 patients and unexposed healthy individuals and their immunogenicity in humanized HLA-DR1/HLA-A*02:01 doubly–transgenic (Tg) mice. Our findings pave the way for incorporating these highly conserved B cell and T cell epitopes in future preemptive multiepitope pan-coronavirus vaccines that would be expected to not only protect against COVID-19 but also against subsequent global outbreaks.

Materials and Methods

Human study population

Sixty-three COVID-19 patients and ten unexposed healthy individuals who had never been exposed to SARS-CoV-2 or COVID-19 patients were enrolled in this study (Table I). Seventy-eight percent were non-White (African, Asian, Hispanic, and others), and 22% were White. Forty-four percent were females and 56% were males, with an age range of 26–95 y (median, 62 y).

Detailed clinical and demographic characteristics of the COVID-19 patients and the unexposed healthy individuals with respect to age, gender, HLA-A*02:01 and HLA-DRB1 distribution, COVID-19 disease severity, comorbidity, and biochemical parameters are presented in Table I. None of the symptomatic patients were on antiviral or anti-inflammatory drug treatments at the time of blood sample collections. The COVID-19 patients (n = 63) were divided into four groups depending on the severity of the symptoms: group 1 comprised of SARS-CoV-2–infected patients who never developed any symptoms or any viral diseases (i.e., asymptomatic patients; n = 11), group 2 had mild symptoms (i.e., in-patient only; n = 32), group 3 had moderate symptoms (i.e., intensive care unit [ICU] admission; n = 11), and group 4 had severe symptoms (i.e., ICU admission ± intubation or death; n = 9). As expected, compared with the asymptomatic group, all of the three symptomatic groups (i.e., mild, moderate, and severe) had higher percentages of comorbidities, including diabetes (22–64%), hypertension (64–78%), cardiovascular disease (11–18%), and obesity (9–50%) (Table I). The final group, group 5, was comprised of unexposed healthy individuals (controls) with no history of COVID-19 or contact with COVID-19 patients (n = 10) collected prior to 2019. All subjects were enrolled at the University of California Irvine under Institutional Review Board–approved protocols (no. 2020-5779). A written informed consent was received from all participants prior to inclusion in this study.

Sequence comparison among SARS-CoV-2 and previous coronavirus strains

We retrieved 81,963 human SARS-CoV-2 genome sequences from GISAID database, representing countries from North America, South America, Central America, Europe, Asia, Oceania, and Africa (Fig. 1). Furthermore, the full-length sequences of SARS-CoV-2 strains (SARS-CoV-2-Wuhan-Hu-1 (MN908947.3), SARS-CoV-Urbani (AY278741.1), HKU1–genotype B (AY884001), CoV-OC43 (KP293903), CoV-NL63 (NC_005831), CoV-229E (KY983587), and MERS (NC_019843) found in the human host were obtained from the National Center for Biotechnology Information (NCBI) GenBank. SARS-CoV-2 genome sequences from bat (RATG13 [MK584930.1], ZC211691 [MK746933.2], ZC211692 [MK746934.1], and ZC211693 [MK746937]), and pangolin (GX-P2V [MT072864.1], GX-P5E [MT040336.1], GX-P1E [MT040334.1], GX-P4L [MT040333.1], GX-P3B [MT072865.1], MP789 [MT121216.1], and Guangdong-P2S [EPISL_410544]) were obtained from NCBI (www.ncbi.nlm.nih.gov/nucleotide) and GISAID (www.gisaid.org). More so, the SARS-CoV strains from bat (WIV16 [KT448582.1], WIV1 [KF367457.1], YNL3_31C [KP868808.1], Rs672 [KF588686.1], and recombinant strain [FJ211859.1]), camel (KT368891.1, MN514967.1, KF917527.1, and NC_028752.1), and civet (Civet007, A022, and B039) were also retrieved from the NCBI GenBank. The sequences were aligned using ClustalW algorithm in MEGAX.

Sequence conservation analysis of SARS-CoV-2

The SARS-CoV-2-Wuhan-Hu-1 (MN908947.3) protein sequence was compared with SARS-CoV- and MERS-CoV–specific protein sequences obtained from human, bat, pangolin, civet, and camel. The Sequence Variation Analysis was performed on the consensus-aligned protein sequences from each virus strain. This Sequence Homology Analysis identified consensus protein sequences from the SARS-CoV-2 and MERS-CoV and predicted the Epitope Sequence Analysis.

SARS-CoV-2 CD8 and CD4 T cell epitope prediction

Epitope prediction was carried out using the 12 proteins predicted for the reference SARS-CoV-2 isolate Wuhan-Hu-1. The corresponding SARS-CoV-2 protein accession identification numbers obtained from NCBI (www.ncbi.nlm.nih.gov/protein) are as follows: YP_009724389.1 (open reading frame [ORF]1ab), YP_009724395.1 (ORF1a), YP_009724390.1 (spike glycoprotein [S]), YP_009724391.1 (envelope protein [E]), YP_009724393.1 (membrane glycoprotein), YP_009724394.1 (ORF6), YP_009724395.1 (ORF7a), YP_009724396.1 (ORF8), YP_009724397.2 (nucleocapsid phosphoprotein), and YP_009725255.1 (ORF10). The tools used for CD8+ T cell–based epitope prediction were SYFPEITHI, MHC class I (MHC-I) binding predictions, and class I immunogenicity. Of these, the latter two were hosted on the Immune Epitope Database (IEDB) platform. For the prediction of CD4+ T cell epitopes, we used multiple databases and algorithms, namely SYFPEITHI, MHC class II (MHC-II) binding predictions, TepiTool, and TEPITOPEpan. For the CD8+ T cell epitope prediction, we selected the five most frequent HLA-A class I alleles (HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*11:01, and HLA-A*23:01) with large coverage of the world population, regardless of race and ethnicity (Supplemental Fig. 1A, 1C), using a phenotypic frequency cutoff of ≥6%. Similarly, for CD4 T cell epitope prediction, we selected HLA-DRB1*04:01, HLA-DRB1*11:01, HLA-DRB1*15:01, HLA-DRB1*03:01, and HLA-DRB1*04:01 alleles with large population coverage (Supplemental Fig. 1B, 1D). Subsequently, using NetMHC, we analyzed the SARS-CoV-2 protein sequence against all the aforementioned MHC-I and MHC-II alleles. Epitopes with 9-mer length for MHC-I and 15-mer length for MHC-II were predicted. Subsequently, the peptides were analyzed for binding stability to the respective HLA allotype. Our stringent epitope selection criteria were based on picking the top 1% epitopes focused on prediction percentile scores.
Table I. Clinical and demographic features of symptomatic and asymptomatic COVID-19 patients and unexposed healthy individuals enrolled in the study

<table>
<thead>
<tr>
<th>Patient Characteristics</th>
<th>Severe Symptoms</th>
<th>Moderate Symptoms</th>
<th>Mild Symptoms</th>
<th>Asymptomatic</th>
<th>Healthy Individuals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 9)</td>
<td>(n = 11)</td>
<td>(n = 32)</td>
<td>(n = 11)</td>
<td>(n = 10)</td>
</tr>
<tr>
<td>Demographic features</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>62 (26–95)</td>
<td>56 (24–91)</td>
<td>62 (24–87)</td>
<td>54 (22–78)</td>
<td>51 (25–67)</td>
</tr>
<tr>
<td>Gender (male/female)</td>
<td>5/4 (56%/44%)</td>
<td>9/2 (82%/18%)</td>
<td>21/11 (66%/34%)</td>
<td>2/9 (18%/82%)</td>
<td>6/4 (60%/40%)</td>
</tr>
<tr>
<td>Race (White/Non-White [%])</td>
<td>2/7 (22%/78%)</td>
<td>1/10 (9%/91%)</td>
<td>4/28 (12%/88%)</td>
<td>5/6 (45%/55%)</td>
<td>5/5 (50%/50%)</td>
</tr>
<tr>
<td>HLA phenotype</td>
<td>HLA-A*0201 (+ve)</td>
<td>4/9 (44%)</td>
<td>5/11 (45%)</td>
<td>14/32 (44%)</td>
<td>4/11 (36%)</td>
</tr>
<tr>
<td></td>
<td>HLA-DRB1 (+ve)</td>
<td>9/9 (100%)</td>
<td>11/11 (100%)</td>
<td>32/32 (100%)</td>
<td>11/11 (100%)</td>
</tr>
<tr>
<td>Clinical parameters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>25.2 (20.3–57.9)</td>
<td>26.5 (20.9–33.5)</td>
<td>30.3 (21.1–46.5)</td>
<td>29.3 (17.6–60.8)</td>
<td></td>
</tr>
<tr>
<td>Temperature/fever/chills</td>
<td>98.4 (97.9–99.9)</td>
<td>100.3 (97.7–102.8)</td>
<td>99.1 (97.8–102.8)</td>
<td>98.7 (97.7–102.5)</td>
<td>—</td>
</tr>
<tr>
<td>Shortness of breath</td>
<td>4 (44%)</td>
<td>6 (55%)</td>
<td>16 (50%)</td>
<td>1 (9%)</td>
<td>—</td>
</tr>
<tr>
<td>Fatigue/maligia</td>
<td>0 (0%)</td>
<td>4 (36%)</td>
<td>15 (47%)</td>
<td>1 (9%)</td>
<td>—</td>
</tr>
<tr>
<td>Headache</td>
<td>4 (44%)</td>
<td>6 (54%)</td>
<td>16 (50%)</td>
<td>1 (9%)</td>
<td>—</td>
</tr>
<tr>
<td>Ventilator support</td>
<td>9 (100%)</td>
<td>11 (100%)</td>
<td>2 (6%)</td>
<td>0 (0%)</td>
<td>—</td>
</tr>
<tr>
<td>WBC</td>
<td>10.9 (7.4–14.8)</td>
<td>8 (6–29.8)</td>
<td>7.1 (3.9–18.9)</td>
<td>30.6 (4.9–60.8)</td>
<td>—</td>
</tr>
<tr>
<td>RBC</td>
<td>4.07 (2.97–5.92)</td>
<td>4.04 (2.68–4.59)</td>
<td>4.4 (2.69–5.41)</td>
<td>7.1 (4.17–16.2)</td>
<td>—</td>
</tr>
<tr>
<td>Hemoglobin (g/l)</td>
<td>11.1 (8.3–16.2)</td>
<td>12.1 (8.4–13.8)</td>
<td>13.1 (8.1–16.9)</td>
<td>4.4 (4.01–12.9)</td>
<td>—</td>
</tr>
<tr>
<td>Comorbidities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes</td>
<td>2 (22%)</td>
<td>7 (64%)</td>
<td>18 (56%)</td>
<td>5 (46%)</td>
<td>—</td>
</tr>
<tr>
<td>Hypertension</td>
<td>7 (78%)</td>
<td>7 (64%)</td>
<td>22 (69%)</td>
<td>4 (36%)</td>
<td>—</td>
</tr>
<tr>
<td>Cardiovascular disease</td>
<td>1 (11%)</td>
<td>2 (18%)</td>
<td>4 (13%)</td>
<td>1 (9%)</td>
<td>—</td>
</tr>
<tr>
<td>CAD</td>
<td>0 (0%)</td>
<td>1 (9%)</td>
<td>2 (6%)</td>
<td>0 (0%)</td>
<td>—</td>
</tr>
<tr>
<td>Asthma/COPD</td>
<td>1 (11%)</td>
<td>3 (27%)</td>
<td>4 (13%)</td>
<td>0 (0%)</td>
<td>—</td>
</tr>
<tr>
<td>Obesity</td>
<td>3 (33%)</td>
<td>1 (9%)</td>
<td>16 (50%)</td>
<td>5 (46%)</td>
<td>—</td>
</tr>
<tr>
<td>Cancer</td>
<td>1 (11%)</td>
<td>0 (0%)</td>
<td>6 (19%)</td>
<td>1 (9%)</td>
<td>—</td>
</tr>
</tbody>
</table>

Patients were scored on a scale of 1 to 4 and then classified into three groups of symptomatic patients (severe symptoms [i.e., ICU admission] and death), moderate symptoms [i.e., ICU admission], mild symptoms [i.e., in-patient only], and asymptomatic patients [i.e., infected patients but with no symptoms]). Unexposed healthy individuals with no history of COVID-19 or contact with COVID-19 patients. Median values are shown along with range. Dashes (—) show the absence of parameters.

BMI, body mass index; CAD, coronary artery disease; COPD, chronic obstructive pulmonary disease; ESRD, end-stage renal disease.

SARS-CoV-2 B cell epitope prediction

Linear B cell epitope predictions were carried out on the S, the primary target of B cell immune responses for SARS-CoV-2. We used the BepiPred 2.0 algorithm embedded in the B cell prediction analysis tool hosted on IEDB platform. For each protein, the epitope probability score for each amino acid was calculated and the probability of exposure were retrieved. Potential B cell epitopes were predicted using a cutoff of 0.55 (corresponding to a specificity >0.81 and sensitivity <0.3) and considering sequences having more than 5-aa residues. This screening process resulted in 28 B cell peptides (Supplemental Table III). From this pool, we selected 10 B cell epitopes with 19-62 aa lengths. Three B cell epitopes were observed to possess receptor-binding domains and sensitivity regression. SInter shows the similarity of amino acids of the B cell peptides and the template-target similarity measured by the protein structure similarity score. Higher SInter score represents a more significant binding affinity among the ACE2 molecule and B cell peptides. Subsequently, molecular docking models were built based on distance restraints for protein–peptide pairs using GalxyPepDock. Based on the optimized energy scores, docking models were ranked.

While performing the protein–peptide docking analysis for CD8+ T cell epitope peptides, we used the x-ray crystal structure of HLA-A*0201 in complex 4UQ3 available on the Protein Data Bank and for CD4 peptides x-ray crystallographic structure HLA-DM–HLA-DRB1 complex 4GBX.

Epitope conservancy analysis

The epitope conservancy analysis tool was used to compute the degree of the conservancy of CD8+ T cell, CD4+ T cell, and B cell epitopes within a given protein sequence of SARS-CoV-2 set at 100% identity level. The fraction of protein sequences that contain the regions similar to epitopes were evaluated on the degree of similarity or correspondence among two sequences. The CD8+ T cell and CD4+ T cell epitopes were screened against all the 12 structural and nonstructural proteins of SARS-CoV-2, namely YP_009724390.1 (S), YP_009724391.1 (ORF3a), YP_009724392.1 (E), YP_009724393.1 (membrane glycoprotein). The conservancy of CD8+ T cell epitopes was screened using the x-ray crystal structure of HLA-A*0201 in complex 4UQ3 available on the Protein Data Bank and for CD4 peptides x-ray crystallographic structure HLA-DM–HLA-DRB1 complex 4GBX.
Population coverage–based T cell epitope selection

For a robust epitope screening, we evaluated the conservancy of CD8+ T cell, CD4+ T cell, and B cell epitopes within human-SARS-CoV-2 genome sequences representing North America, South America, Africa, Europe, Asia, and Australia. As of August 27, 2020, the Nextstrain database recorded 81,963 human-SARS-CoV-2 genome sequences, and the number of genome sequences continues growing daily. In the present analysis, 81,963 human-SARS-CoV-2 genome sequences were extrapolated from the GISAID and NCBI GenBank databases. We therefore considered all the 81,963 SARS-CoV-2 genome sequences representing six continents for subsequent conservancy analysis. We set a threshold for a candidate CD8+ T cell, CD4+ T cell, and B cell epitope if the epitope showed 100% sequence conservancy in ≥95 human-SARS-CoV-2 genome sequences. Furthermore, population coverage calculation was carried out using the Population Coverage software hosted on IEDB platform (45). Population coverage calculation was performed to evaluate the distribution of screened CD8+ and CD4+ T cell epitopes in world population at large in combination with HLA-I (HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*11:01, and HLA-A*23:01) and HLA-II (HLA-DRB1*01:01, HLA-DRB1*04:01, HLA-DRB1*11:01, HLA-DRB1*15:01, HLA-DRB1*03:01, and HLA-DRB1*04:01) alleles.

Peptide synthesis

Potential peptide epitopes (9-mer long for CD8+ T cell epitopes and 15-mer long for CD4+ T cell epitopes) identified from 12 human-SARS-CoV-2 proteins, namely ORF1ab, ORF1a, S, ORF3a, E membrane glycoprotein, ORF6, ORF7a, ORF7b, ORF8, nucleocapsid phosphoprotein, and ORF10, were synthesized using solid-phase peptide synthesis and standard 9-fluorenylmethoxycarbonyl technology (21st Century Biochemicals, Marlborough, MA). The purity of peptides was over 90%, as determined by reversed-phase HPLC (Vydac C18) and mass spectroscopy (Voyager MALDI-TOF System). Stock solutions were made at 1 mg/ml in 10% DMSO in PBS.
method of synthesis was used for B cell peptide epitopes from the spike protein of SARS-CoV-2.

Cell lines
T2 (174 × CEM.T2) mutant hybrid cell line derived from the T lymphoblast cell line CEM was obtained from the American Type Culture Collection (www.atcc.org). The T2 cell line was maintained in IMDM (American Type Culture Collection, Manassas, VA) supplemented with 10% heat-inactivated FCS and 100 U of penicillin/ml, 100 U of streptomycin/ml (Sigma-Aldrich, St. Louis, MO). T2 cells lack the functional TAP heterodimer and failed to express normal amounts of HLA-A*02:01 on the cell surface. HLA-A*02:01 surface expression is stabilized following the binding of exogenous peptides to these MHC-I molecules.

Stabilization of HLA-A*02:01 on class I HLA-transfected B × T hybrid cell lines
To determine whether synthetic peptides could stabilize HLA-A*02:01 molecule expression on the T2 cell surface, peptide-inducing HLA-A*02:01 upregulation on T2 cells was examined according to a previously described protocol (46, 47). T2 cells (3 × 10^6 per well) were incubated with different concentrations (30, 10, and 3 μM) of 91 individual CD8+ T cell–specific peptides in 48-well plates for 18 h at 26 °C. Cells were then incubated at 37 °C for 3 h in the presence of 0.7 μg/ml BD GolgiStop to block cell surface expression of newly synthesized HLA-A*02:01 molecules and human β-2 microglobulin (1 μg/ml). The cells were subsequently washed with FACS buffer (1% BSA and 0.1% sodium azide in PBS) and stained with anti-HLA-A2–specific mAb (clone BB7.2) (BD Pharmingen, San Diego, CA) at 4 °C for 30 min. After incubation, the cells were washed with FACS buffer, fixed with 2% paraformaldehyde in PBS, and analyzed by flow cytometry using a Fortessa (Becton Dickinson) flow cytometer equipped with a BD High Throughput Sampler for rapid analysis of samples prepared in plate format. The acquired data were analyzed with FlowJo software (BD Biosciences, San Jose, CA), and expression was measured by mean fluorescence intensity (MFI). Percentage of MFI increase was calculated as follows: percentage MFI increase = (MFI with the given peptide – MFI without peptide)/MFI without peptide × 100. Each experiment was performed three times, and means ± SD values were calculated.

HLA-A*02:01 and HLA-DR1 double-transgenic mice
A colony of HLA class I and class II double-Tg mice was maintained at the University of California Irvine (48) vivarium and treated in accordance with the Association for Assessment and Accreditation of Laboratory Animal Care according to Institutional Animal Care and Use Committee–approved animal protocols (ns. 2020-19-111) and National Institutes of Health guidelines. The HLA-A1/HLA-B8-Tg mice retain their endogenous mouse H2 locus and express human HLA-A*02:01 and HLA-DRB1*01 under the control of its normal promoter (49, 50). Prior to this study, the expression of HLA-A*02:01 and DR1 molecules on the PBMCs of each HLA-Tg mouse was confirmed by FACS.

Immunization of mice
Groups of age-matched HLA-transgenic mice/B6 mice (n = 3) were immunized s.c. on days 0 and 14, with a mixture of four SARS-CoV-2–derived human CD4+ T/CD8+ T/B cell peptide epitopes delivered in aluminum hydroxide (alum) and CpG1668 adjuvants. As a negative control, mice received adjuvants alone (mock immunized).

Splenocytes
Spleens were harvested from mice in 2 wk after second immunization. Spleens were placed in 10 ml of cold PBS with 10% FBS and 2× antibiotic–antimycotic (Life Technologies, Carlsbad, CA). Spleens were minced finely and sequentially passed through a 100-μm screen and a 70-μm screen (BD Biosciences). Cells were then pelleted via centrifugation at 400 × g for 10 min at 4 °C. RBCs were lysed using a lysis buffer (ammonium chloride) and washed again. Isolated splenocytes were diluted to 1 × 10^6 viable cells/ml in RPMI media with 10% (v/v) FBS and 2× antibiotic–antimycotic. Viability was determined by trypan blue staining.

Flow cytometry analysis
PBMCs/splenocytes were analyzed by flow cytometry. The following Abs were used: CD8, CD4, CD45, CD107a/b, CD44, CD69, TNF-α, and IFN-γ. For surface staining, mAbs against various cell markers were added to a total of 1 × 10^6 cells in PBS containing 1% FBS and 0.1% sodium azide (FACS buffer) and left for 45 min at 4 °C. At the end of the incubation period, the cells were washed twice with FACS buffer. A total of 100,000 events were acquired by LSR II (Becton Dickinson, Mountain View, CA), followed by analysis using FlowJo software (Tree Star, Ashland, OR).

ELISpot assay
All reagents used were filtered through a 0.22-μm filter. Wells of 96-well MultiscreenHTS plates (MilliporeSigma, Billerica, MA) were prewet with 30% ethanol and then coated with 100 μl primary anti–IFN-γ Ab solution (10 μg/ml of 1-D1K coating Ab from Mabtech in PBS [pH 7.4], V-E4) overnight at 4 °C. After washing, nonspecific binding was blocked with 200 μl of RPMI media with 10% (v/v) FBS for 2 h at room temperature. Following the blockade, 0.5 × 10^5 cells from patients PBMCs (or from mouse splenocytes) in 100 μl of RPMI were mixed with 10 μg individual peptides (with DMSO for no stimulation or with individual peptide at a final concentration of 10 μg/ml). After incubation in humidified 5% CO2 at 37 °C for 72 h (samples from COVID-19 patients) or 5 d (for healthy donor samples to recall their T cell memory), cells were removed by washing (using PBS and RPMI 02% solution), and 100 μl of biotinylated secondary anti–IFN-γ Ab (clone 7-B6-1; Mabtech) in blocking buffer (PBS with 0.5% FBS) was added to each well. Following a 2-h incubation and washing, HRP-conjugated streptavidin was diluted 1:1000, and wells were incubated with 100 μl for 1 h at room temperature. Following washing, wells were incubated for 1 h at room temperature with 100 μl of tetramethylbenzidine detection reagent and spots counted with an automated EliSpot Reader System (ImmunoSpot reader; Cellular Technology, Shaker Heights, OH).

ELISA-based assay to access the efficacy of RBD region toward inducing specificAbs against B cell epitopes in HLA-A2–treated mice
The efficacy of our B cell peptide epitopes toward inducing specific Abs was measured in the HLADR1/A*02:01-immunized mice by ELISA. ELISA plates (catalog M5785; Sigma-Aldrich) were first coated overnight at 4 °C with 10 μg/ml of each B cell peptide epitope. Subsequently, plates were washed five times with PBS–Tween 0.01% before starting the blocking by adding PBS with 1% BSA for 3 h at room temperature, followed by a second wash. Sera of C57BL/6 mice immunized either with pool B cell peptides/alum/CpG or adjuvant alone (control) were added into the wells at varying dilutions (1/5, 1/25, 1/125, and 1/625 or PBS only in triplicates). Plates were incubated at 4 °C overnight with the sera, then washed with PBS–Tween 0.01% before to add anti-mouse IgG Ab (1/500 dilution; Mabtech). After the last washing, streptavidin–HRP (1/1000 dilution; Mabtech) was added for 30 min at room temperature. Finally, we added 100 μl of filtered tetramethylbenzidine substrate for 15 min and blocked the reaction with H2SO4 before the readout (OD measurement was done at 450 nm on the Bio-Rad iMark microplate reader). The same procedure was followed to measure the titers of Abs specific against our 15 screened B cell epitopes in the sera of COVID-19 patients (n = 40) and healthy donors (n = 10), using anti-human IgG Ab as the secondary Ab (1/500 dilution; Mabtech).

Constructing the phylogenetic tree
Phylogenetic analyses were conducted in MEGAX. The evolutionary history was performed, and phylogenetic tree was constructed using the maximum likelihood method and Tamura-Nei model. The maximum likelihood method assumes that each locus evolves independently by pure genetic drift. The tree with the highest log likelihood was selected. Initial tree(s) for the heuristic search were obtained by applying neighbor-joining and BioNJ algorithms to a matrix of pairwise genetic distances estimated using the Tamura-Nei model and then selecting the topology with superior log likelihood value. This analysis involved available nucleotide sequences of SARS-CoV-2 from human (Homo sapiens), bat (Rhinolophus affinis and R. malayanus), and pangolin (Manis javanica). In addition, genome sequences from previous outbreaks of SARS-CoV in human, bat, civet, and camel were taken into consideration while performing the evolutionary analyses.

Data and code availability
The human-specific SARS-CoV-2 complete genome sequences were retrieved from the GISAID database, whereas the SARS-CoV-2 sequences for pangolin (M. javanica), and bat (R. affinis and R. malayanus) were retrieved from NCBI. Genome sequences of previous strains of SARS-CoV for human, bat, civet, and camel were retrieved from the NCBI GenBank.

Statistical analyses
Data for each differentially expressed markers among blockade-treated and mock-treated groups of HLAs-Tg mice were compared by ANOVA and Student t test using GraphPad Prism version 6 (GraphPad Software, La Jolla, CA). Statistical differences observed in the measured CD8+ and CD4+ T cells and Ab responses between healthy donors and COVID-19 patients
were calculated using ANOVA and multiple t test comparison procedures in GraphPad Prism. Data are expressed as the mean ± SD. Results were considered statistically significant at \( p \leq 0.05 \).

**Results**

**Evolutionary convergence of human SARS-CoV-2 into bat- and pangolin-derived SL-CoVs**

Understanding the animal origins of SARS-CoV-2 is critical for the development of a preemptive pan-coronavirus vaccine to protect from future human outbreaks and deter future zoonosis.

We first screened for the evolutionary relationship among human SARS-CoV-2 and SARS-CoV/MERS-CoV strains from previous outbreaks (i.e., Urbani, MERS-CoV, OC43, NL63, 229E, and HKU1-genotype B) along with 25 SL-CoV genome sequences obtained from different animal species: bats (R. affinis and R. malayanus), civet cats (Paguma larvata), and pangolins (M. javanica) and MERS-CoVs from camels (Camelus dromedarius and C. bactrianus) (Fig. 1). These sequence alignments revealed similarity of the original human-SARS-CoV-2 strain found in Wuhan, China, to four bat SL-CoV strains: hCoV-19-bat-Yunnan-RmYN02, bat-CoV-19-XZC21, and hCoV-19-bat-Yunnan-RaTG13 obtained from the Yunnan and Zhejiang provinces of China (Fig. 1A). With further genetic distance analysis, we discovered the least evolutionary divergence between SARS-CoV-2 isolate Wuhan-Hu-1 and the above-mentioned three SL-CoV isolates from bats, namely 1) Bat-CoV-RaTG13 (0.1), 2), bat-CoV-19-XZC21 (0.1), and 3) Bat-CoV-YN02 (0.2) (Fig. 1B, 1C). Moreover, the phylogenetic analysis performed among the whole genome sequences of a total of 81,963 SARS-CoV-2 strains for which sequences have been reported in circulation in 190 countries suggest an evolutionary convergence of bat and pangolin SL-CoVs into the human SARS-CoV-2 strains (Fig. 1D, 1E). Furthermore, through a complete genome tree derived from the 81,963 SARS-CoV-2 genome sequences submitted from Asian, African, North American, South American, European, and Oceanian regions, we confirmed that the least evolutionary divergence for SARS-CoV-2 strains is in SL-CoVs isolated from bats and pangolins (Fig. 1D–F).

Altogether, the phylogenetic analysis and genetic distance suggest that the highly contagious and deadly human-SARS-CoV-2 strain originated from bats, most likely from either the Bat-CoV-19-ZXC21 (MG772934.1) or Bat-CoV-RaTG13 (MN996532.2) strains that spilled over into humans after further mutations and/or recombination. These mutations and/or recombination(s) possibly contributed to the rapid global expansion of the highly contagious and deadly SARS-CoV-2 (51, 52).

**Genome-wide identification of SARS-CoV-2 CD8\(^{+}\) T cell epitopes that are highly conserved between human and bat/pangolin coronaviruses**

We first predicted potential CD8\(^{+}\) T cell epitopes from the entire genome sequence of the first SARS-CoV-2-Wuhan-Hu-1 strain (NCBI GenBank accession number MN908947.3) (53–59). For this, we used multiple databases and algorithms, including the SYFPEITHI, MHC-I processing predictions, MHC-I binding predictions, MHC-I immunogenicity, and IEDB (56, 60). We focused on epitopes restricted to the five most frequent HLA class I alleles with large coverage in worldwide human populations, regardless of race and ethnicity (i.e., HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*11:01, and HLA-A*23:01) (61–63) (Supplemental Fig. 1A, 1C).

Using the aforementioned criteria, we originally identified a total of 9,660 potential CD8\(^{+}\) T cell epitopes derived from 12 structural proteins (S, membrane glycoprotein, and nucleocapsid phosphoprotein) and ORFs of SARS-CoV-2-Wuhan-Hu-1 strain (MN908947.3) (Supplemental Table I). Subsequently, this large pool of epitopes was narrowed down to 91 epitopes that are highly conserved among 1) over 81,000 SARS-CoV-2 strains (that currently circulate in 190 countries on six continents); 2) the four major common cold coronaviruses that caused previous outbreaks (i.e., hCoV-OC43 [KF923903], hCoV-229E [KY983587], hCoV-HKU1 genotype B [AY884401], and hCoV-NL63 [NC_005831]); and 3) the SL-CoVs that are isolated from bats, civet cats, pangolins, and camels (Fig. 2A). Although the highest degree of similarity (expressed as percentage of resemblance) was identified among 81,963 SARS-CoV-2 strains, 6 strains of previous human SARS-CoV-2s and 18 animal SL-CoVs strains isolated from bats and pangolins, only a small percentage of similarity was found between the SARS-CoV-2 and MERS-CoV strains.
Supplemental Fig. 2). However, a significantly lower degree of similarity was recorded among the SARS-CoV-2 and the SL-CoVs strains isolated from civet cats’ and camels’ CoVs (Supplemental Fig. 2).

We further identified 27 SARS-CoV-2 human CD8+ T cell epitopes out of the 91 epitopes that bound with high affinity with HLA-A*02:01 molecules using in vitro peptide–HLA binding assay (Fig. 2A). Four epitopes were found to be very high affinity binders (Fig. 2B). The 27 epitopes with high binding affinity were later confirmed in silico using molecular docking models across five major HLA-A*01:01, HLA-A*02:01, HLA-A*03:01, HLA-A*11:01, and HLA-A*23:01 haplotypes (Supplemental Fig. 3) (64). The highest binding affinity to HLA-A*02:01 molecules with the highest S_inter scores (blue squares) were recorded for ORF1ab_6749–6757, S_2–10, S_858–966, S_1220–1228, E_26–34, ORF8_83–91, ORF10_3–11, and ORF10_5–13, whereas minimum S_inter score was observed for ORF1ab_1732–1740, S_691–699, and membrane protein (M)939–947. Other CD8+ T cell epitopes like ORF1ab_1675–1683, ORF1ab_2363–2371, ORF1ab_3013–3021, and ORF7b_26–34 were also found with intermediate S_inter scores (Supplemental Fig. 3A, 3B). Although the identified highly conserved CD8+ T cell epitopes were distributed within 8 of the 12 structural and nonstructural ORFs (i.e., ORF1ab, S, E, M, ORF6, ORF7b, ORF8, and ORF10), the highest numbers of epitopes were localized in the replicase polyprotein 1ab/1a (ORF1ab) (nine epitopes) followed by the spike glycoprotein (five epitopes) (Supplemental Figs. 2, 8).

Altogether, our findings identified 27 highly conserved potential human CD8+ T cell epitopes from the sequence of SARS-CoV-2 that are highly conserved among 81,963 SARS-CoV-2 strains, the four major common cold coronaviruses (i.e., hCoV-OC43, hCoV-229E, hCoV-HKU1 genotype B, and hCoV-NL63), newly found highly transmissible variants (Supplemental Fig. 9), and several SL-CoV strains that are isolated from bats and pangolins. These results suggest that both the structural and the nonstructural proteins

![FIGURE 3.](https://www.jimmunol.org/)

CD8+ T cells specific to highly conserved SARS-CoV-2 epitopes detected in COVID-19 patients and unexposed healthy individuals. (A) Experimental design: PBMCs from HLA-A*02:01 positive COVID-19 patients (n = 30) (B) and controls unexposed healthy individuals (n = 10) (C) were isolated and stimulated overnight with 10 μM of each of the 27 SARS-CoV-2–derived CD8+ T cell epitopes. The number of IFN-γ-producing cells were quantified using ELISpot assay (B, C, and D). Dotted lines represent threshold to evaluate the relative magnitude of the response: mean SFCs between 25 and 50 correspond to a medium/intermediate response, whereas a strong response is defined for mean SFCs >50. PBMCs from HLA-A*02:01–positive COVID-19 patients (E) were further stimulated for an additional 5 h in the presence of mAbs specific to CD107α and CD107β and GolgiPlug and GolgiStop. Tetramers specific to Spike peptides, CD107α/β and CD69, and TNF-α expression were then measured by FACS. Representative FACS plot showing the frequencies of tetramer+CD8+ T cells, CD107α/β+CD8+ T cells, CD69+CD8+ T cells, and TNF-α+CD8+ T cells following priming with a group of four Spike CD8+ T cell epitope peptides. Average frequencies of tetramer+CD8+ T cells, CD107α/β+CD8+ T cells, CD69+CD8+ T cells, and TNF-α+CD8+ T cells. *p < 0.05.
are immunodominant Ags that are targeted by human CD8\(^+\) T cells from both COVID-19 patients and common cold coronavirus–infected healthy individuals.

**In silico screening of potential promiscuous SARS-CoV-2 CD4\(^+\) T cell epitopes that are highly conserved between human and bat/pangolin coronaviruses**

We subsequently identified a total of 9,594 potential HLA-DR–restricted CD4\(^+\) T cell epitopes from the whole genome sequence of SARS-CoV-2-Wuhan-Hu-1 strain (MN908947.3) using multiple databases and algorithms including the SYFPEITHI, MHC-II binding predictions, TepTool, and TEPITOPEpan (Supplemental Table II). These potential promiscuous CD4\(^+\) T cell epitopes were screened in silico against the five most frequent HLA-DR alleles with large coverage in the human population, regardless of race or ethnicity: HLA-DRB1\(*\)01:01, HLA-DRB1\(*\)11:01, HLA-DRB1\(*\)15:01, HLA-DRB1\(*\)03:01, and HLA-DRB1\(*\)04:01 (Supplemental Fig. 1B, 1D). The number of potential CD4\(^+\) T cell epitopes was later narrowed down to 16 epitopes based on 1) the epitope sequences that are highly conserved among 81,963 SARS-CoV-2 strains, the four major common cold and 25 SL-CoV strains isolated from bats, civet cats, pangolins, and camels (Supplemental Fig. 4); and 2) their high binding affinity to HLA-DR molecules using in silico molecular docking models (Supplemental Fig. 5). The sequences of most of the 16 CD4\(^+\) T cell epitopes are 100% conserved and common among 81,963 SARS-CoV-2 strains currently circulating in six continents (Supplemental Fig. 4). A high degree of sequence similarities was also identified in the sequences of most 16 CD4\(^+\) T cell epitopes from SARS-CoV-2 strains and the SL-CoV strains isolated from bats and pangolins. In contrast, a lower sequence similarity was identified among CD4\(^+\) T cell epitopes from SARS-CoV-2 strains and the SL-CoV strains isolated from civet.

**FIGURE 4.** CD4\(^+\) T cells specific to highly conserved SARS-CoV-2 epitopes detected in COVID-19 patients and unexposed healthy individuals. (A) Experimental design: PBMCs from HLA-DRB1–positive COVID-19 patients (\(n=30\)) (B) and controls unexposed healthy individuals (\(n=10\)) (C) were isolated and stimulated for 48 h. with 10 \(\mu\)M of each of the 16 SARS-CoV-2-derived CD4\(^+\) T cell epitope peptides. The number of IFN-\(\gamma\)-producing cells were quantified using ELISpot assay (B, C, and D). Dotted lines represent a threshold to evaluate the relative magnitude of the response: mean SFCs between 25 and 50 correspond to a medium/intermediate response, whereas a strong response is defined for mean SFCs >50. PBMCs from HLA-DRB1–positive COVID-19 patients (E) were further stimulated for an additional 5 h in the presence of mAbs specific to CD107\(a/b\), CD69, and TNF-\(\alpha\) expression were then measured by FACS. Representative FACS plot showing the frequencies of tetramer\(^+\)CD4\(^+\) T cells, CD107\(^ab\)CD4\(^+\) T cells, CD69\(^+\)CD4\(^+\) T cells, and TNF-\(\alpha\)-CD4\(^+\) T cells following priming with a group of two Spike CD4\(^+\) T cell epitope peptides. Average frequencies are shown for tetramer\(^+\)CD4\(^+\) T cells, CD107\(^ab\)CD4\(^+\) T cells, CD69\(^+\)CD4\(^+\) T cells, and TNF-\(\alpha\)-CD4\(^+\) T cells. \(^*\)\(p<0.05\).
cats, followed by MERS-like CoV strains isolated from camels (Supplemental Figs. 4, 8).

The 16 highly conserved CD4+ T cell epitopes are distributed within 9 out of the 12 structural and nonstructural ORFs (i.e., ORF1ab, S, E, M, ORF6, ORF7a, ORF7b, ORF8, and nuclease protein [N]). The highest numbers of epitopes were localized in the replicase polyprotein ORF1ab/1a (five epitopes) followed by ORF7a (three epitopes) (Supplemental Figs. 4, 8). Unlike the human CD8+ T cell epitopes, the human CD4+ T cell epitopes are found to be expressed in each of the structural S, E, M, and N proteins. Two epitopes are from the E, one epitope from the M, one epitope from the N protein, and one epitope from the spike protein. The remaining CD4+ T cell epitopes are distributed among the ORF6, ORF7a, ORF7b, and ORF8 proteins (Supplemental Figs. 4, 8).

Altogether, these results identified 16 potential CD4+ T cell epitopes from the whole sequence of SARS-CoV-2 that cross-react and have high sequence similarity among 81,963 SARS-CoV-2 strains, the main four major common cold coronaviruses, and the S-CoV strains isolated from bats and pangolins. Similar to CD8+ T cell epitopes, the replicase polyprotein ORF1ab appeared to be the most immunodominant Ag, with a high number of conserved epitopes that may possibly be targeted by human CD4+ T cells.

Cross-reactive human and animal coronavirus-derived epitopes, spanning the whole virus proteome, are targeted by memory CD4+ and CD8+ T cells from SARS-CoV-2 patients and unexposed healthy individuals

Next, we assessed whether the potential SARS-CoV-2 CD4+ and CD8+ T cell epitopes that are highly conserved between human and animal coronaviruses would recall memory CD8+ T cells from COVID-19 patients as well as from healthy individuals who have never been exposed to SARS-CoV-2 or to COVID-19 patients (i.e., from healthy individuals’ blood samples that were collected from 2014 to 2018; Figs. 3A, 4A). Detailed clinical and demographic characteristics of the COVID-19 patients and the unexposed healthy individuals enrolled in the current study, with respect to age, gender, HLA-A*02:01 and HLA-DRB1 distribution, COVID-19 disease severity, comorbidity, and biochemical parameters, are described in Table 1 and in Materials and Methods.

Blood-derived PBMCs from COVID-19 patients (black, Fig. 3B) and healthy individuals (white, Fig. 3C) were analyzed by ELISPOT for frequencies in SARS-CoV-2 epitope–specific, IFN-γ–producing CD8+ T cells. As shown in Fig. 3B and 3D, significant numbers of SARS-CoV-2 epitope–specific memory CD8+ T cells producing IFN-γ were detected in PBMCs of COVID-19 patients. Out of the 27 highly conserved cross-reactive SARS-CoV-2 CD8+ T cell epitopes (Supplemental Fig. 2) selected for their binding affinity with HLA-A*02:01 molecules (Fig. 2B), strong T cell responses (mean spot-forming T cells [SFCs] > 50 per 0.5 × 10⁶ PBMCs fixed as threshold) were detected in COVID-19 patients against 10 epitopes derived from 1) structural proteins like Spike (i.e., S958-968, S974-984, S1000-1008, and S1220-1228) or the Es (i.e., E26-34a); and 2) nonstructural proteins (i.e., ORF1ab9175-1003, ORF1ab2110-2218, ORF1ab2689-2677, ORF691-111, and ORF1093-111) (Fig. 3B, 3D). In addition, 12 other SARS-CoV-2 CD8+ T cell epitopes from structural of nonstructural SARS-CoV-2 proteins induced an intermediate response (with a mean SFCs between 25 and 50 per 0.5 × 10⁶ PBMCs) in COVID-19 patients: ORF1ab864-921, ORF1ab1013-1021, ORF1ab1833-1911, ORF1ab3732-3740, ORF1ab4283-4291, ORF1ab4629-4647, S2-10, S691-699, E206-208, M2-60, M3-90-97, and ORF1051-13.

Moreover, among the 27 SARS-CoV-2 epitopes, seven epitopes recalled a strong memory CD8+ T cell response (mean SFCs > 50) from unexposed healthy individuals (i.e., ORF1ab1675-1683, ORF1ab732-7340, ORF1ab6749-6757, S974-984, S1000-1008, and S1220-1228), and five epitopes recalled a memory CD8+ T cell response that was intermediate (ORF1ab4149-4167, S2-10, E26-34a, ORF1031-11, and ORF1043-13) (Fig. 3C, 3D). However, the unexposed healthy individuals exhibited a different pattern of CD8+ T cell immunodominance as compared with COVID-19 patients. We then compared the epitopes specificity and function of memory CD8+ T cells in HLA-A*02:01–positive COVID-19 patients and healthy individuals using flow cytometry (Fig. 3E). For a better comparison, a similar FACS gating strategy was applied to PBMC-derived T cells from both COVID-19 and healthy donors (data not shown). Our COVID-19 patients appeared to have a higher frequency of CD8+ T cells compared with healthy donors (Fig. 3E). Tetramer staining showed that many of SARS-CoV-2 epitope–specific CD8+ T cells are multifunctional, producing IFN-γ and TNF-α and expressing CD69 and CD107αβ markers of activation and cytotoxicity in COVID-19 patients (Fig. 3E).

Similar to SARS-CoV-2 memory CD8+ T cells, memory CD4+ T cells specific to several highly conserved SARS-CoV-2 epitopes were detected in both COVID-19–recovered patients and unexposed healthy individuals (Fig. 4B–D). Out of the 16 highly conserved cross-reactive SARS-CoV-2 CD4+ T cell epitopes (Supplemental Fig. 4), strong T cell responses (mean SFCs > 50 per 0.5 × 10⁶ PBMCs fixed as threshold) were detected in COVID-19 patients against two epitopes, one derived from the structural protein M (M176-190) and one from the nonstructural protein ORF1a (ORF1a1350-1365) (Fig. 4B, 4D). Moreover, six additional SARS-CoV-2 CD8+ T cell epitopes from nonstructural SARS-CoV-2 proteins (i.e., ORF1a1350-1365, ORF668-6102, ORF1a6420-6434, ORF62-126, ORF7a3-17, and ORF8b1-15) and two more epitopes from structural proteins (i.e., S1-13 and N388-403) induced an intermediate CD4+ T cell response (mean SFCs between 25 and 50 per 0.5 × 10⁶ PBMCs) in COVID-19 patients (Fig. 4B, 4D).

Besides, among the 16 SARS-CoV-2 epitopes, two epitopes recalled a strong memory CD4+ T cell response (mean SFCs > 50) from unexposed healthy individuals with no history of COVID-19 (i.e., ORF1a1350-1365 and ORF612-26) (Fig. 4C, 4D). Furthermore, five additional epitopes recalled an intermediate CD4+ T cell response in these unexposed healthy individuals (i.e., ORF1a1801-1815, S1-13, M176-190, ORF8b1-15, and N388-403). Unlike for CD8+ T cell responses, the unexposed healthy individuals exhibited a similar pattern of CD4+ T cell immunodominance as compared with COVID-19 patients, with few differences in the magnitude of the responses only. Multifunctional SARS-CoV-2 epitope–specific CD4+ T cells expressing CD69, CD107αβ, and TNF-α were detected using specific tetramers in PBMCs of HLA-DR1–positive COVID-19 patients and healthy individuals (Fig. 4E) with a trend showing higher percentage of these cells in COVID-19 patients, although not significantly higher.

The immunogenicity of the identified SARS-CoV-2 human epitopes was assessed in humanized HLA-DR1/HLA-A*02:01 double-transgenic mice (Fig. 5A, 6A). A mixture of peptides incorporating CD4+ T cell or CD8+ T cell epitopes was delivered with CpG and alum, as shown in Figs. 5A and 6A and detailed in the Materials and Methods. As a negative control, mice received adjuvant alone. The induced SARS-CoV-2 epitope–specific CD4+ T cell responses were determined in the spleen using multiple immunological assays, including IFN-γ ELISPOT, FACS surface markers of activation, markers of cytotoxic degranulation, and intracellular cytokine staining. The gating strategy used for mice is shown in Figs. 5B and 6B. Two weeks after the second immunization with the mixture of CD8+ T cell peptides, 10 out of 27 highly conserved SARS-CoV-2 human CD8+ T cell epitope peptides were immunogenic in humanized HLA-DR1/HLA-A*02:01 double-transgenic mice (Fig. 5C, 5D). The
remaining 17 CD8 T cell epitopes presented moderate/low immunogenicity levels in HLA-DR1/HLA-A*02:01 double-transgenic mice. The immunogenic epitopes were derived from both structural Spike protein (S2, 10, S958–966, S1000–1009, and S1220–1228) and E (E20–28) and from nonstructural proteins (i.e., ORF1ab1334–1372, ORF1ab2363–2371, ORF1ab4178–4270, ORF1ab4578–5478, ORF8752–81, and ORF1064–643). Moreover, 7 out of 16 SARS-CoV-2 peptides induced significant CD4 T cell responses in humanized HLA-DR1/HLA-A*02:01 double-transgenic mice (Fig. 6C, 6D). The immunogenic epitopes were derived from both structural Spike protein (S1–13) and M (M176–190) and from nonstructural proteins (ORF1a1350–1365, ORF1b8019–9033, ORF1a6420–6434, ORF612–26, ORF7b54–52, and ORF861–75). The remaining 9 CD4 T cell epitopes presented moderate/low level of immunogenicity in HLA-DR1/HLA-A*02:01 double-transgenic mice.

Altogether, these results indicate that preexisting memory CD4 T and CD8 T cells specific to both structural and nonstructural protein Ags and epitopes are present in COVID-19 patients and unexposed healthy individuals. Although SARS-CoV-2–specific CD4 and CD8 T cells in COVID-19 patients and healthy donors target epitopes from the whole virus proteome, most T cell epitopes are concentrated in the nonstructural proteins, with ORF1a/b being the most targeted Ags. These memory T cells recognized highly conserved SARS-CoV-2–2 epitopes that cross-react with the human and animal coronaviruses. It is likely that infection with a common cold coronavirus and/or human exposition with animal- and pet-related coronaviruses induced long-lasting memory CD4 and CD8 T cells specific to the structural and nonstructural SARS-CoV-2 epitopes in healthy unexposed individuals. Heterologous immunity and heterologous immunopathology orchestrated by these cross-reactive, epitope-specific memory CD4 and CD8 T cells following previous multiple exposures to common cold coronaviruses may have shaped protection versus susceptibility to SARS-CoV-2 infection and disease with a yet-to-be determined mechanism(s).
Identification of B cell epitopes from SARS-CoV-2 Spike protein that are highly conserved between human and animal coronaviruses that are antigenic in humans and immunogenic in humanized HLA-transgenic mice

We next predicted potential linear B cell (Ab) epitopes on Spike protein sequence of the first SARS-CoV-2-Wuhan-Hu-1 strain (NCBI GenBank accession number MN908947.3) using BepiPred 2.0, with a cutoff of 0.55 (corresponding to a specificity >0.81 and sensitivity <0.30) and considering sequences having more than 5-aa residues (65). This stringent screening process initially resulted in the identification of 28 linear B cell epitopes (Supplemental Table III). From this pool of 28 potential epitopes, we later selected 15 B cell epitopes (19–62 aa in length) based on 1) their sequences being highly conserved between SARS-CoV-2, the main four major common cold coronaviruses (CoV-OC43 [KF923903], CoV-229E [KY983587], CoV-HKU1 [AY884001], and CoV-NL63 [NC_005831]) (66), and the SL-CoVs that are isolated from bats, civet cats, pangolins, and camels; and 2) the probability of exposure each linear epitope to the surface of infected target cells (Supplemental Fig. 6). The Spike epitope sequences highlighted in blue indicate a high degree of homology among the currently circulating 81,963 SARSCoV-2 strains and at least a 50% conservancy among two or more human SARS-CoV strains from previous outbreaks and the SL-CoV strains isolated from bats, civet cats, pangolins, and camels (Supplemental Fig. 6). Two of the fifteen B cell epitopes, namely S 369–393 and S440–501, overlap with the Spike’s RBD regions that bind to the ACE2 receptor (designated as RBD-1 and RBD-2 in Supplemental Fig. 7A). Higher S_inter scores were observed for RBD-derived epitopes S369–393 and S471–501 when molecular docking was performed against the ACE2 receptor (Supplemental Fig. 7B). Upon screening for the glycosylation regions, we observed B cell epitopes S13–37, S59–81, S329–363, S601–640, and S1133–1172 with Asparagines predicted to be N-glycosylated. In contrast, B cell epitopes
S516, S536, S598, and S819 were observed to be the O-glycosylated. The remaining B cell epitopes S287–317, S304–322, S369–393, S404–501, S440–501, S672–690, and S688–909 were found to possess no glycosylation.

We later determined the ability of each of the 15 B cell epitopes selected from the Spike protein, which showed a high conservancy between human and animal coronaviruses to induce SARS-CoV-2 epitope-specific, Ab-producing plasma B cells and IgG Abs in B6 mice (Fig. 7). Synthetic peptides corresponding to each linear B cell epitope were produced. Because four epitopes were too long to synthesize (e.g., 62 aa), they were divided into two or three short fragments, resulting in a total of 22 B cell epitope peptides (Supplemental Table III). As illustrated in Fig. 7A, groups of five B6 mice each received two s.c. injections with mixtures of three to four B cell epitope peptides mixed with CpG and alum adjuvants. Alum/CpG adjuvants alone were used as negative controls (mock immunized). B and C The frequencies of IgG-producing CD3−CD138−B220+ plasma B cells were determined in the spleen of immunized mice by flow cytometry. B cells were then gated by the expression of CD3− and B220+ cells and CD138 expression on plasma B cells determined. Representative FACS plot (left panels) and average frequencies (right panel) of plasma B cells detected in spleen of immunized mice. The percentages of plasma CD138+ B220+B cells is indicated on the top left of each dot plot. D SARS-CoV-2–derived B cell epitope–specific IgG responses were quantified in immune serum, 14 d after second immunization (i.e., day 28) by ELISpot (number of IgG+ spots). Representative ELISpot images (left panels) and average frequencies (right panel) of anti-peptide specific IgG-producing B cell spots (1 × 105 splenocytes per well) following 4 d in vitro B cell polyclonal stimulation with mouse Poly-S (ImmunoSpot). The top/left of each ELISpot image shows the number of IgG-producing B cells per half a million cells. ELISA plates were coated with each individual immunizing peptide. The B cell epitope–specific IgG concentrations (micrograms per milliliter) measured by ELISA in E levels of IgG detected in peptide-immunized B6 mice after subtraction of the background measured from mock-vaccinated mice. The dashed horizontal line indicates the limit of detection and in F level of IgG specific to each of the 22 Spike peptides detected SARS-CoV-2–infected patients (n = 40) after subtraction of the background measured from healthy nonexposed individuals, as shown in G (n = 10). Black bars and gray bars show high and medium immunogenic B cell peptides, respectively. The dashed line indicates the limit of detection.
the frequencies of plasma B cells in the spleen is shown in Fig. 7B. Out of the 22 Spike B cell epitopes, seven epitopes (S13–37, S287–317, S524–558, S544–578, S565–598, S601–628, and S614–640) induced high frequencies of CD138⁺B220⁺ plasma B cells in the spleen of B6 mice (Fig. 7C). The IgG ELISpot assay confirmed that 7 out of the 22 Spike B cell epitopes induced significant numbers of IgG-producing B cells in the spleen (Fig. 7D). Moreover, significant amounts of IgG were detected in the serum of the immunized B6 mice. These IgG Abs were specific to 6 out of the 22 Spike B cell peptide epitopes (S13–37, S59–81, S287–317, S565–598, S601–628, and S614–640) (Fig. 7E). As expected, nonimmunized animals or those that received adjuvant alone did not develop detectable IgG responses. Of these six highly immunogenic B cell peptides, five peptides (S13–37, S59–81, S287–317, S601–628, and S614–640) were highly antigenic, as they were recognized by serum IgG from COVID-19 patients, confirming the presence of at least one native linear B cell epitope in each peptide (Fig. F). In summary, we identified five highly conserved immunogenic and antigenic human B cell target epitopes from the Spike SARS-CoV-2 virus that recall IgG Abs from COVID-19 patients (Fig. G). This study further discovered five highly conserved B cell epitopes from SARS-CoV-2: S13–37, S287–317, S318–363, S614–640, and S1133–1160, which are recognized by IgG Abs from healthy individuals who were never exposed to COVID-19, suggesting B cell epitopes cross-reactivity to other human coronaviruses (Fig. 7G).

Discussion

Although the current COVID-19 pandemic will likely be overcome through the implementation of physical distancing, barriers together with a mass vaccination, it is indispensable that a safe and effective preemptive vaccine be developed and in place ready to protect against another inevitable COVID pandemic that will emerge in the years to come.

Toward this goal of developing a multi-epitope preemptive pan-coronavirus human vaccine, we identified several cross-reactive human B and T cell epitopes of SARS-CoV-2 that are highly conserved among the human SARS-CoVs and animal SL-CoVs. Although antiviral SARS-CoV-2–specific Abs and CD4⁺ and CD8⁺ T cell responses appear crucial in protecting asymptomatic patients and convalescent patients, very little information exists with regards to the repertoire of targeted SARS-CoV-2 B and T cell epitopes that are common within a substantial group of human and animal coronaviruses (56, 67, 68). In agreement with our results, 4 out of the 27 CD8⁺ T cell epitopes reported in this study have been recently reported to be cross-reactive between SARS-CoV and SARS-CoV-2: ORF1ab2363–2371, ORF1ab3013–3021, S558–565, and S1220–1228.

FIGURE 8. Illustrations of the SARS-CoV/SARS-CoV-2 genome-wide location of the highly conserved, antigenic, and immunogenic CD4⁺ T cell, CD8⁺ T cell, and B cell epitopes. (A) Enveloped, spherical, ~120 nm in diameter SARS-CoV/SARS-CoV-2 genome encodes four structural proteins: spike, E, M, and N, highlighted in blue, green, gray, and black, respectively. (B) The SARS-CoV/SARS-CoV-2 genome encodes two large nonstructural genes, ORF1a (green) and ORF1b (gray), encoding 16 nonstructural proteins (NSP1–NSP16). The genome encodes at least six accessory proteins (shades of light gray) that are unique to SARS-CoV/SARS-CoV-2 in terms of number, genomic organization, sequence, and function. The common SARS-CoV, SARS-CoV-2, and SL-CoVs–derived human B (blue), CD4⁺ (green), and CD8⁺ (black) T cell epitopes are shown. Structural and nonstructural ORFs used in this study were from SARS-CoV-2-Wuhan-Hu-1 strain (NCBI accession number MN908947.3). The amino acid sequence of the SARS-CoV-2-Wuhan-Hu-1 structural and nonstructural proteins was screened for human B and CD4⁺ and CD8⁺ T cell epitopes using different computational algorithms as previously described in Materials and Methods. Shown are genome-wide identified SARS-CoV-2 human B cell epitopes (in blue), CD4⁺ T cell epitopes (in green), and CD8⁺ T cell epitopes (in black) that are highly conserved between human and animal coronaviruses.
(56, 68). Similarly, B cell epitope S\textsubscript{287-317} has been reported as cross-reactive between SARS-CoV and SARS-CoV-2 (56, 68). However, to the best of our knowledge, none of the 16 CD\textsuperscript{4}\ T cell epitopes identified in this study have been reported previously.

The highly conserved human B and T cell epitopes reported in this study have massive implications for the development of a universal preemptive pan-coronavirus vaccine to induce (or to boost) neutralizing Abs, CD\textsuperscript{4}\ Ths (Th1), and antiviral CD\textsuperscript{8}\ cytotoxic T cells (21, 69, 70). Some of our identified epitopes are similar to those recently reported by Grifoni et al. (56, 68), whereas other epitopes have never been reported. Moreover, in agreement with recent reports (56, 68, 71), our study revealed a high degree of similarity among SARS-CoV-2, SARS-CoV, and bat-SL-CoV epitope sequences, but not the MERS-CoV epitope sequences. In the current study, we have identified B cell epitopes S\textsubscript{13-37}, S\textsubscript{99-81}, S\textsubscript{329-363}, S\textsubscript{601-640}, and S\textsubscript{1133-1172} with N-glycosylated regions and S\textsubscript{116-536}, S\textsubscript{224-598}, and S\textsubscript{802-819} with O-glycosylated regions. Extensive glycosylation has been observed in CoV’s S, representing the most extensive known class I viral fusion proteins. SARS spike glycoprotein is known to encode 69 N-linked glycan sequons per trimeric spike, with SARS-CoV-2 containing 66 sites (72). These modifications may mask immunogenic B cell epitopes from the host humoral immune system by occluding them with host-derived glycans. However, some studies underscore the importance of glycosylation in the lack of immunogenicity and viral immune evasion. Watanabe et al. (72) have reported that extensive N-linked glycan modifications of SARS and MERS-CoV S do not constitute an effective shield compared with glycan shields of certain other viruses, which is reflected by the overall structure, density, and oligomannose abundance across the corresponding trimeric glycoproteins (66).

Whether the identified epitopes will contribute to protection is beyond the scope of the current study. Unlike most coronavirus subunit vaccines (73–75), our multiepitope coronavirus vaccine (e.g., the pan-coronavirus candidate I illustrated in Supplemental Fig. 8) incorporates multiple human asymptomatic B and CD\textsuperscript{4}\/CD\textsuperscript{8} T cell epitopes that are selected carefully from the whole genome of SARS-CoV-2 for being recognized by Abs and CD\textsuperscript{4}/CD\textsuperscript{8} T cells from asymptomatic and convalescent patients that are naturally protected from COVID-19. The present study employed a combinatorial approach for designing an all-in-one multiepitope pan-coronavirus vaccine candidate (Supplemental Fig. 8) by applying highly conserved genome-wide human B and T cell epitopes from 12 genome-derived antigenic proteins of SARS-CoV-2. The present study focused on HLA-A*02:01–restricted epitopes represented by more than 50% of the human population. However, epitopes restricted to other HLA-A, HLA-B, and HLA-C haplotypes, including the forecasted population coverage of the chosen T cell epitope ensemble (combined HLA class I), are expected to cover 99.8% of the global population, regardless of race and ethnicity. In addition, for a wider vaccine coverage (i.e., close to 99%), our multiepitope pan-coronavirus vaccine platform would be easily adapted to include CD\textsuperscript{8} T cell epitopes for other HLA supertypes that are distributed in the various human populations. The polymorphic HLA molecules can be clustered into a handful of HLA-A supertypes that bind largely overlapping peptide repertoires (76). Moreover, such a multiepitope vaccine would be easily adapted to exclude undesirable epitopes that are restricted to HLA-B*44 and HLA-C*01 alleles, which appear to correlate with SARS-CoV-2 virus spreading across certain countries (77), and HLA-B*35 allele, which appear to be associated with severe pneumonia developed by SARS-CoV-2 in young patients (78).

We do not exclude that the highly conserved B cell and CD\textsuperscript{4} and CD\textsuperscript{8} T cell epitopes identified from bat’s coronavirus variants will mutate following recombination that often occurs for zoonotic events before an animal SL-CoV spills over into humans. In this context, our preemptive multiepitope pan-coronavirus vaccine is highly adaptable to newly mutated coronavirus strains. If a coronavirus epitope mutates, that single epitope can be easily adjusted and replaced in the multiepitope vaccine (79). In lieu of this, we have screened all of our candidate epitopes against SARS-CoV-2 variants, which have been evidenced with increased transmissibility, including B.1.1.7 variant emerging from the U.K. (variant 20I/501Y.V1), B.1.351 variant emerging from South Africa (variant 20H/501Y.V2), B.1.1.28 variant emerging from Brazil (P.1 variant 20J/501Y.V3), and CAL.20C variant observed in California (Supplemental Fig. 9). We found 100% conservancy for 15 out of 16 CD\textsuperscript{4} T cell epitopes against B.1.1.7 (U.K.), B.1.351 (South Africa), and B.1.1.28 (Brazil) variants. One mutation in S131 region that was found in the CAL.20C variant is observed in one of our CD\textsuperscript{4} T cell epitope S\textsubscript{1-13} (MVFVLVLLPLVSS), whereas the remaining 15 CD\textsuperscript{4} T cell epitopes showed 100% conservancy CAL.20C variant. Notably, all our 27 CD\textsuperscript{8} T cell epitopes showed conservancy against highly transmissible South African, Brazilian, and Californian (CAL.20C) variants. However, one region specific to the nonsynonymous mutation S982A from the U.K. variant was observed in our CD\textsuperscript{8} T cell epitope S\textsubscript{796-798} (VLNDILSRL). The remaining 26 CD\textsuperscript{8} T cell epitopes showed 100% conservancy with the B.1.1.7 (U.K.) variant. In comparison, two of the screened B cell epitopes (S\textsubscript{99-81} and S\textsubscript{601-640}) belonged to regions specific to South African variant B.1.351 (20H/501Y.V2). More so, two B cell epitopes (S\textsubscript{644-646} and S\textsubscript{540-501}) represent the regions specific to South African variant B.1.351 (20H/501Y.V2), Brazilian variant B.1.1.28 (P.1 variant 20J/501Y.V3), and Californian variant (CAL.20C). Two B cell epitopes, S\textsubscript{524-599} and S\textsubscript{772-750}, represent regions specific to nonsynonymous spike protein–specific mutations A570D and P681H found in the B.1.17 strain from the U.K. This emphasizes that our preemptive multiepitope pan-coronavirus vaccine strategy could be easily adapted to any variant as well as to any new zoonotic bat SL-CoVs that may spill over into humans in the future. This high adaptability is expected to speed up the implementation of a future preemptive multiepitope vaccine before a local outbreak spreads and transforms into a global pandemic.

It is inevitable that future COVID-like outbreaks, caused by yet another spillover of a bat SL-CoV, could lead to other COVID-like pandemics with global health, social, and economic disasters in the years to come. However, because it is almost impossible to predict which viral strain might cause the next coronavirus pandemic, it is urgent to develop a pan-coronavirus vaccine that targets a wide range of human and animal coronavirus strains. Unlike conventional monovalent vaccines made from epitopes selected from a single virus strain, a preemptive multiepitope pan-coronavirus vaccine (Supplemental Fig. 8) that includes several highly conserved human B and CD\textsuperscript{4} and CD\textsuperscript{8} T cell epitopes identified from the entire genome sequences of human SARS-CoVs that cross-react and are shared with bat and pangolin SL-CoVs (18, 80–83). The current ongoing collaborative research efforts should not only focus on developing a vaccine for COVID-19 but should also be oriented toward developing preemptive pan-coronavirus vaccines. Such a proactive vaccine strategy would help fight and contain future local outbreaks and epicenters of highly contagious and deadly zoonotic coronaviruses globally before becoming the next deadly pandemic worldwide (21, 70). Moreover, because it is impossible to predict the time and location of the next deadly global pandemic, it is essential to have ready, at least preclinically, several pan-coronavirus vaccine candidates that would be quickly implemented in a clinical trial against a substantial group of coronaviruses before an outbreak spreads into a global pandemic(s).
In the current study, CD4+ and CD8+ T cells specific to highly conserved SARS-CoV-2 epitopes were detected in healthy adults, recruited between 2014 and 2018, who have never been exposed to the SARS-CoV-2 virus. These findings suggest cross-reactive T cells between current SARS-CoV-2 and previous circulating common cold coronaviruses, as confirmed by recent reports (68, 70, 71). However, because it is unknown whether the healthy adults used in this study were indeed exposed to any common cold coronaviruses, such an assertion may not be conclusive. Among the many circulating common cold coronaviruses known to infect humans, four serotypes that cause severe respiratory infections are highly seasonal: CoV-OC43, CoV-229E, CoV-HKU1, and CoV-NL63 (66), and appear to have a similar transmission potential to influenza A (H3N2). The seasonality of these common cold coronaviruses is predictable, as their outbreaks often emerged in December, peaked in January/February, and began to decrease in March of every year (66).

The human SARS-CoV-2 CD4+ and CD8+ T cell epitopes identified in this study are highly conserved between 81,963 strains of SARS-CoV-2 and CoV-OC43, CoV-229E, CoV-HKU1, and CoV-NL63 (Supplemental Figs. 2, 4). Whether these apparent cross-reactive CD4+ and CD8+ T cells play a protective or a harmful role or an entirely negligible role in SARS-CoV-2 infection and disease remains to be determined (70, 71). Nevertheless, because common cold coronavirus infections are frequent in children, it will be interesting to determine whether children who appeared more resistant to COVID-19 compared with adults will have robust antiviral immunity. T cell responses to some of the common SARS-CoV-2 epitopes identified in this study. Stronger CD4+ and CD8+ T cell responses to common coronavirus epitopes in children would shed some light on the unique situation currently seen in COVID-19 in which immune children tend to be more resistant to SARS-CoV-2 infection and disease, as compared with more susceptible adults (84, 85). Such a result would also imply that a pan-coronavirus vaccine incorporating these cross-reactive highly conserved SARS-CoV-2 human CD4+ and CD8+ T cell epitopes would boost protective T cell immunity that would have been previously induced by a common cold coronavirus. This would protect not only from seasonal circulating common cold coronaviruses but also from SARS-CoV-2 infection and disease.

Even though the highly conserved coronavirus human CD4+ and CD8+ T cell epitopes identified in this report can be enlightening for a pan-coronavirus vaccine, humans are not immunologically naive, and they often have memory CD4+ and CD8+ T cell populations that can cross-react with and respond to other infectious agents, a phenomenon termed heterologous immunity (86). Therefore, we cannot exclude that some SARS-CoV-2-specific CD4+ and CD8+ T cell epitopes identified in this study are cross-reactive with other viral pathogen-derived epitopes, such as epitopes from circulating seasonal influenza or common cold coronaviruses (87, 88). This may explain, in part, the high proportion of asymptomatic infections with SARS-CoV-2 in the current pandemic. The latter is supported by a recent elegant study that detected SARS-CoV-2-reactive CD8+ and CD4+ T cells in healthy individuals that were never exposed to SARS-CoV-2 (68). SARS-CoV-2–specific, but cross-reactive, CD4+ and CD8+ T cells can become activated and modulate the immune responses and clinical outcome of subsequent heterologous SARS-CoV-2 infections. Therefore, T cell cross-reactivity may be crucial in protective heterologous immunity instead of damaging heterologous immunopathology, as has been reported in other systems (89). To confirm SARS-CoV-2–heterologous CD4+ and CD8+ T cell epitopes that may potentially cross-react with other pathogenic (noncoronaviruses) epitopes, we are currently comparing the CD4+ and CD8+ T cell response to those highly conserved SARS-CoV-2 epitopes identified using human CD4+ and CD8+ T cell responses to those of pathogen-free SARS-CoV-2–infected transgenic mice.

In conclusion, we report, in this study, to our knowledge, several previously unknown human universal B and CD4+ and CD8+ T cell target epitopes identified from the whole SARS-CoV-2 genome. These epitopes are highly conserved and common between SARS-CoV-2 Wuhan strain and 1): seven circulating common cold human coronaviruses that caused previous human SARS and MERS outbreaks (70); 2) 81,963 strains of human SARS-CoV-2 that now circulate in six continents; 3) several bat-derived SL-CoV strains (13, 14); and 4) several SL-CoV strains isolated from pangolins (90). The findings from this report pave the way for the development of preemptive multiprotein pan-coronavirus vaccine candidates that would target not only the current human COVID-19 outbreak but also possible future coronavirus outbreaks that might come from a bat-derived SL-CoV strain that would spill over again into humans.

Acknowledgments

The authors thank Dr. Dale Long from the National Institutes of Health Tetramer Facility (Emory University, Atlanta, GA) for providing the tetrarmers used in this study. We thank University of California Irvine Center for Clinical Research and Institute for Clinical and Translational Science for providing human blood samples used in this study. A special thanks to Dr. Delia F. Tifrea for continuous efforts and dedication in providing COVID-19 samples that are crucial for this clinical research. We also thank Dr. Steven A. Goldstein, Dr. Michael J. Stamos, Dr. Suzanne B. Sandmeyer, Jim Mazzo, Dr. Daniela Bota, Dr. Beverly L. Alger, Dr. Dan Forthal, Dr. Tasheen Muzaffar, Dr. Ilhem Messaoudi, Anju Subba, Janice Briggs, Marge Brannon, Beverley Alberola, Jessica Sheldon, Rosie Magallon, and Andria Pontello for contributing directly or indirectly to this COVID-19 vaccine project.

Disclosures

The authors have no financial conflicts of interest. The University of California Irvine has filed a patent application on the results reported in this manuscript.

References


