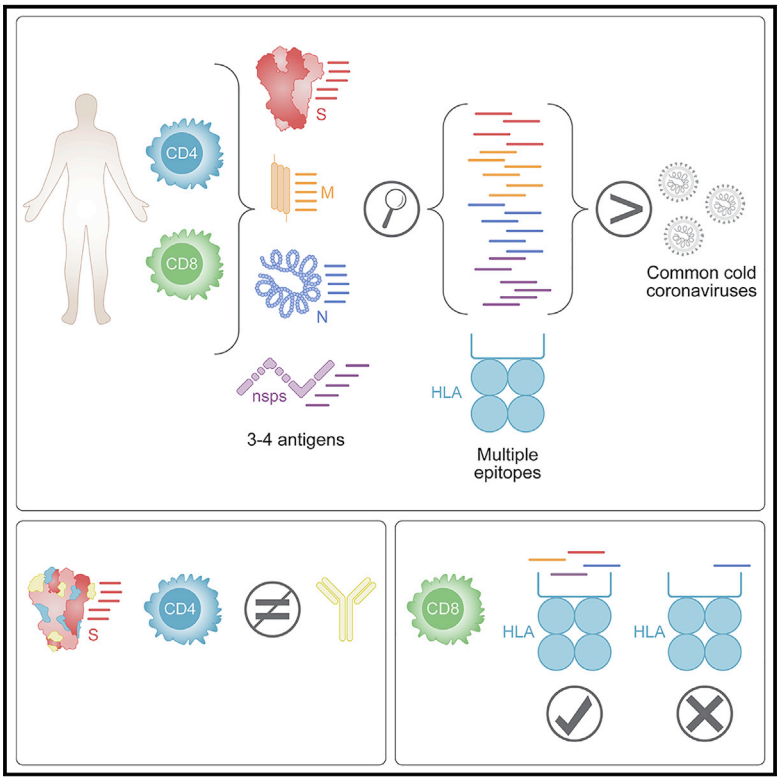
Cell Reports
Medicine

Article

Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases

Graphical Abstract



Highlights

- T cell responses recognize at least 30–40 epitopes in each donor
- Immunodominance is correlated with HLA binding
- Immunodominant regions for $CD4^+$ T cells have minimal overlap with antibody epitopes
- $CD8^+$ T cell responses depend on the repertoire of HLA class I alleles

Authors

Alison Tarke, John Sidney, Conner K. Kidd, ..., Daniela Weiskopf, Alba Grifoni, Alessandro Sette

Correspondence

agrifoni@lji.org (A.G.), alex@lji.org (A.S.)

In Brief

Tarke et al. show a broad T cell repertoire, suggesting that viral escape of T cell immunity is unlikely. CD4 immunodominant regions correlate with HLA binding and not with high common cold coronavirus homology. RBD is poorly recognized by CD4s. Epitope pools can be used to optimize detection of T cell responses.

Tarke et al., 2021, Cell Reports Medicine 2, 100204
February 16, 2021
https://doi.org/10.1016/j.xcrm.2021.100204

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases

Alison Tarke,[1,2] John Sidney,[1] Conner K. Kidd,[1] Jennifer M. Dan,[1,3] Sydney I. Ramirez,[1,3] Esther Dawen Yu,[1] Jose Mateus,[1] Ricardo da Silva Antunes,[1] Erin Moore,[1] Paul Rubiro,[1] Nils Methot,[1] Elizabeth Phillips,[4] Simon Mallal,[4] April Frazier,[1] Stephen A. Rawlings,[3] Jason A. Greenbaum,[1] Bjoern Peters,[1,3] Davey M. Smith,[3] Shane Crotty,[1,3] Daniela Weiskopf,[1] Alba Grifoni,[1,5,*] and Alessandro Sette[1,3,5,6,*]

[1]Center for Infectious Disease and Vaccine Research, La Jolla Institute for Immunology (LJI), La Jolla, CA 92037, USA
[2]Department of Internal Medicine and Center of Excellence for Biomedical Research (CEBR), University of Genoa, Genoa 16132, Italy
[3]Division of Infectious Diseases and Global Public Health, Department of Medicine, University of California, San Diego (UCSD), La Jolla, CA 92037, USA
[4]Institute for Immunology and Infectious Diseases, Murdoch University, Perth, WA, Australia
[5]These authors contributed equally
[6]Lead contact
*Correspondence: agrifoni@lji.org (A.G.), alex@lji.org (A.S.)
https://doi.org/10.1016/j.xcrm.2021.100204

SUMMARY

T cells are involved in control of SARS-CoV-2 infection. To establish the patterns of immunodominance of different SARS-CoV-2 antigens and precisely measure virus-specific CD4$^+$ and CD8$^+$ T cells, we study epitope-specific T cell responses of 99 convalescent coronavirus disease 2019 (COVID-19) cases. The SARS-CoV-2 proteome is probed using 1,925 peptides spanning the entire genome, ensuring an unbiased coverage of human leukocyte antigen (HLA) alleles for class II responses. For HLA class I, we study an additional 5,600 predicted binding epitopes for 28 prominent HLA class I alleles, accounting for wide global coverage. We identify several hundred HLA-restricted SARS-CoV-2-derived epitopes. Distinct patterns of immunodominance are observed, which differ for CD4$^+$ T cells, CD8$^+$ T cells, and antibodies. The class I and class II epitopes are combined into epitope megapools to facilitate identification and quantification of SARS-CoV-2-specific CD4$^+$ and CD8$^+$ T cells.

INTRODUCTION

The severity of the associated coronavirus disease 2019 (COVID-19) ranges from asymptomatic or mild self-limiting disease to severe pneumonia and acute respiratory distress syndrome (World Health Organization [WHO]; https://www.who.int/publications/i/item/clinical-management-of-covid-19). We and others have started to delineate the role of SARS-CoV-2-specific T cell immunity in COVID-19 clinical outcomes.[1–8] A growing body of evidence points to a key role for SARS-CoV-2-specific T cell responses in COVID-19 disease resolution and modulation of disease severity.[2,6,9] Milder cases of acute COVID-19 were associated with coordinated antibody, CD4$^+$ and CD8$^+$ T cell responses, whereas severe cases correlated with a lack of coordination of cellular and antibody responses and delayed kinetics of adaptive responses.[2,6]

To date, most studies have utilized pools of predicted or overlapping peptides spanning the sequence of different SARS-CoV-2 antigens,[1,2,5–10] but the exact T cell epitopes and immunodominant antigen regions have not been comprehensively determined. Several studies have mapped different epitopes or the corresponding T cell receptors (TCRs), providing important insights into the frequency and phenotype of epitope-specific CD8$^+$ and CD4$^+$ T cells in COVID-19 using *ex vivo* studies,[4,10–12] but have been biased in their approach due to sampling only a limited number of cells,[7,11,13] using human leukocyte antigen (HLA) predictions focused on a limited number of allelic variants not representative of the majority of the human population,[11,13] or detecting responses mediated by only a few cytokines, potentially largely underestimating total responses.[4,13] Other important studies, although providing critical knowledge about T cell recognition per se, utilize *in vitro* re-stimulation protocols.[13,14]

Defining a comprehensive set of epitope specificities is important for several reasons. First, it allows us to determine whether, within different SARS-CoV-2 antigens, certain regions are immunodominant. This will be important for vaccine design so as to ensure that vaccine constructs include not only regions targeted by neutralizing antibodies, such as the receptor binding domain (RBD) in the spike (S) region, but also include regions capable of delivering sufficient T cell help and are suitable targets of CD4$^+$ T cell activity. Second, a comprehensive set of epitopes helps define the breadth of responses in terms of the average number of different CD4$^+$ and CD8$^+$ T cell SARS-CoV-2 epitopes generally recognized by each individual. This is key because some reports have described a T cell repertoire focused on few viral

epitopes,[11] which would be concerning for potential viral escape from immune recognition via accumulated mutations that can occur during replication or through viral reassortment. Third, a comprehensive survey of epitopes restricted by a set of different HLAs representative of the diversity present in the general population is important to ensure that results obtained are generally applicable across different ethnicities and racial groups and also to lay the foundations to examine the potential associations of certain HLAs with COVID-19 severity. Finally, the definition of the epitopes recognized in SARS-CoV-2 infection is relevant in the context of the debate on the potential influence of SARS-CoV-2 cross-reactivity with endemic "common cold" coronaviruses (CCC).[3,4] Several studies have defined the repertoire of SARS-CoV-2 epitopes recognized in unexposed individuals,[3,14,15] but the correspondence between that repertoire and the epitope repertoire elicited by SARS-CoV-2 infection has not been evaluated.

In this study, we report a comprehensive map of epitopes recognized by $CD4^+$ and $CD8^+$ T cell responses across the entire SARS-CoV-2 viral proteome. Importantly, these epitopes have been characterized in the context of a broad set of HLA alleles using a direct *ex vivo*, cytokine-independent approach.

RESULTS

Characteristics of the study participants

To broadly define the pattern of immunodominance and epitope recognition associated with SARS-CoV-2 infection, we studied peripheral blood mononuclear cell (PBMC) samples from 99 adult convalescent COVID-19 donors. Their age ranged from 19 to 91 years (median 41), with a gender ratio of about 2M:3F (male 41%; female 59%). Ethnic breakdown was reflective of the demographics of the local enrolled population. Samples were obtained 3 to 184 days post-symptom onset (median 67 days). Peak COVID-19 disease severity was representative of the distribution observed in the general population to date (mild 91%; moderate 2%; severe and critical 7%; Table S1).

SARS-CoV-2 infection was determined by PCR-based testing during the acute phase of infection, if available (79% of the cases), and/or verified by plasma SARS-CoV-2 S protein RBD immunoglobulin G (IgG) ELISA[16] using plasma from convalescent phase blood draws. All donors were seropositive at the time of blood donation, with the exception of two mildly symptomatic donors with positive PCR results from the acute phase of illness but seronegative results at time of blood donation (at 55 and 148 days post-symptom onset [PSO], respectively).

All donors were HLA typed at both class I and class II loci (Table S2). The HLA class I and II alleles frequently observed in the enrolled cohort were largely reflective of what is found in the worldwide population, as reported by the Allele Frequency Net Database[17] and as retrieved from the Immune Epitope Database's (IEDB) (http://www.iedb.org) population coverage tool (Figure S1).[18,19] Of the 20 different HLA class I alleles with phenotypic frequencies >5% in our cohort, 15 (75%) are also present in the most common and representative class I alleles in the worldwide population (Figures S1A and S1B).[20] Likewise, of the 34 different HLA class II alleles with phenotypic frequencies >5% in our cohort, 26 (76%) are also present in the worldwide population with frequencies >5%. These alleles correspond to 16 of the 27 (59%) alleles included in a reference panel of the most common and representative class II alleles in the general population (Figures S1D–S1F).[21] In conclusion, our cohort is largely representative of the HLA allelic variants commonly expressed worldwide.

Pattern of antigen immunodominance in $CD4^+$ and $CD8^+$ T cell responses to SARS-CoV-2 antigens

To study adaptive immune responses in COVID-19 convalescent donors, we previously utilized TCR-dependent activation induced marker (AIM) assays to quantify SARS-CoV-2-specific $CD4^+$ and $CD8^+$ T cells utilizing the combination of markers $OX40^+CD137^+$ and $CD69^+CD137^+$ for $CD4^+$ and $CD8^+$ T cells, respectively.[1,6,15] To define the global pattern of immunodominance in the study cohort, we tested PBMCs from each donor with sets of overlapping peptides spanning the various SARS-CoV-2 proteins, as previously described (Figures 1A and 1B).[1] These data also defined the specific viral antigens recognized by each donor and therefore highlight the specific antigens/donor pairs suitable for further epitope identification studies, as shown in Figures 1C and 1E.

For each SARS-CoV-2 protein antigen (Table S3), we recorded the % of donors in which a positive response was detected and the total response counts (positive cells/million detected in the AIM assay). This information was used to tabulate the percentage of the total response ascribed to each protein and calculate the cumulative coverage provided by the most immunodominant proteins.

For $CD4^+$ T cell responses, 9 viral proteins (non-structural protein [nsp] 3, nsp4, nsp12, nsp13, S, open reading frame 3a [ORF3a], membrane [M], ORF8, and nucleocapsid [N]) accounted for 83% of the total response. In the context of $CD8^+$ T cell responses, 8 viral proteins (nsp3, nsp4, nsp6, nsp12, S, ORF3a, M, and N) accounted for 81% of the total response. These results confirmed the pattern previously observed with a more limited (n = 20) number of COVID-19 patients[1] and highlight a broad pattern of immunodominance, where 8 to 9 antigens are required to cover 80% of the response.

We further evaluated the number of antigens recognized in each of the individual donors analyzed. To this end, we focused on antigens associated with a sizeable response, arbitrarily defined herein as those antigens individually accounting for at least 10% of the total response. We found that, per donor, an average of 3.2 and 2.7 proteins were recognized by 10% or more of the total $CD4^+$ and $CD8^+$ SARS-CoV-2-specific T cells, respectively (Figures 1D and 1F).

Functional consequences of SARS-CoV-2-specific $CD4^+$ T cell responses directed against different antigens

We next investigated whether the recognition of different SARS-CoV-2 antigens by $CD4^+$ T cells correlated with functional antibody and/or $CD8^+$ T cell responses. Consistent with the wide range of blood collection time points (day PSO) and peak disease severity in the COVID-19 donor cohort, we observed a wide range of RBD IgG responses (Figure 2A). Combined $CD4^+$ T cell responses did not significantly correlate with the antibody response to RBD ($R = 0.1285$; $p = 0.2051$; Figure 2B).

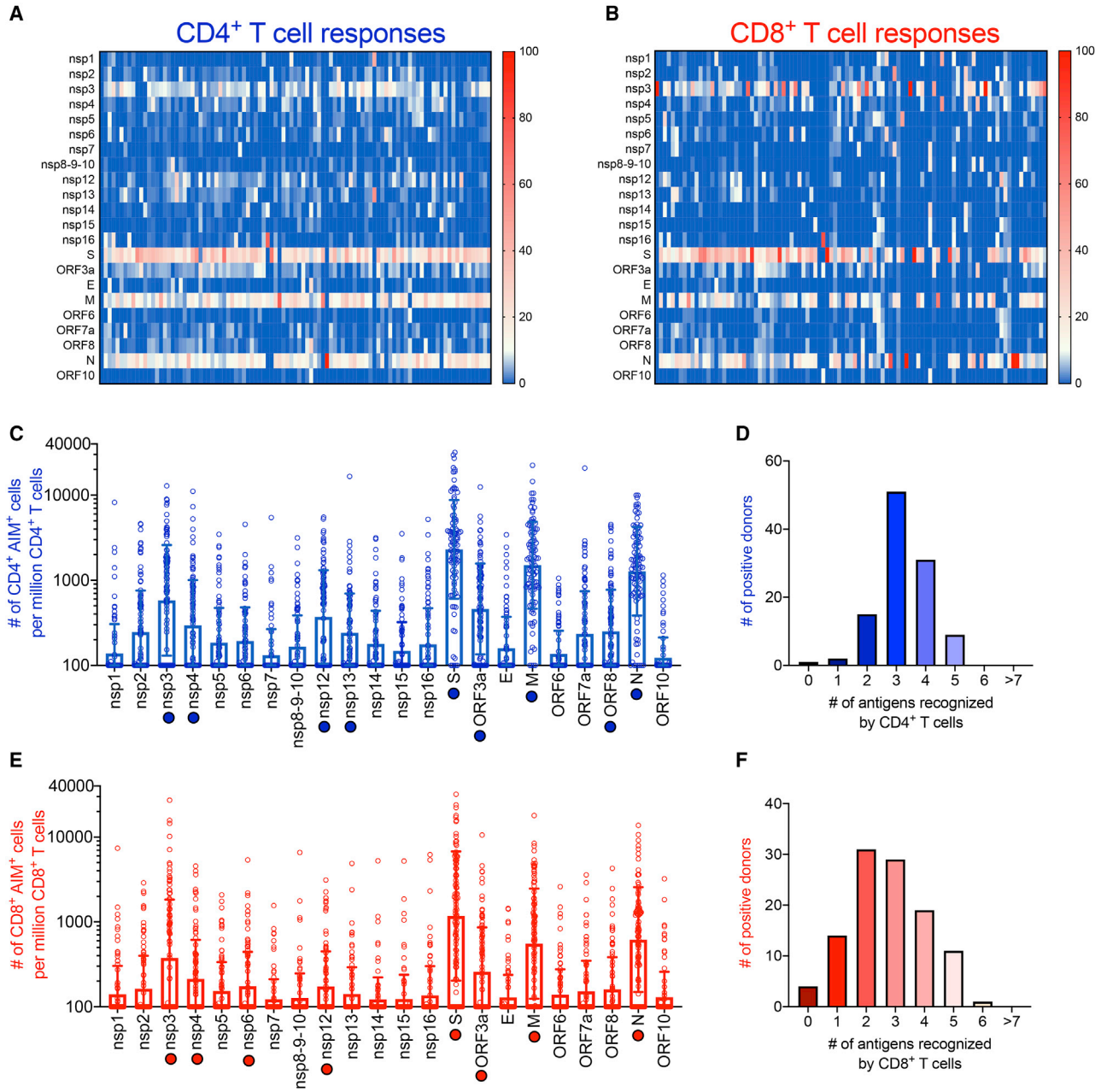


Figure 1. SARS-CoV-2-specific T cell reactivity per protein

PBMCs from convalescent COVID-19 donors (n = 99) were analyzed for reactivity against SARS-CoV-2 (A–F). Heatmaps of T cell reactivity across the entire SARS-CoV-2 proteome and as a function of the donor tested are shown for $CD4^+$ (A) and $CD8^+$ (B) T cells. The x axis shows individual donors' responses to the indicated SARS-CoV-2 protein. Immunodominance at the ORF/antigen level and breath of T cell responses are shown for $CD4^+$ (C) and $CD8^+$ (E) T cells. Data are shown as geometric mean ± geometric SD. The numbers of donors recognizing one or more antigens with a response >10%, normalized per donor to account for the differences in magnitude based on days PSO, are shown for $CD4^+$ (D) and $CD8^+$ (F) T cells. Empty blue and red circles represent $CD4^+$ and $CD8^+$ T cell reactivity per protein, respectively. Filled blue and red circles highlight the immunodominant antigens recognized by $CD4^+$ and $CD8^+$ T cells, respectively.

Breaking the correlation down for individual antigens showed that two correlations had $p < 0.05$, namely spike ($R = 0.2223$; $p = 0.0270$) and M protein ($R = 0.2117$; $p = 0.0354$), but these would not be significant when performing a multiple hypothesis comparison taking all other antigens into account (Figures 2C–2E). In contrast, the correlation between $CD4^+$ and $CD8^+$ T cell responses was highly significant in aggregate ($R = 0.6756$; $p = 1.70 \times 10^{-14}$; Figure 2F) and was significant for each of the individual antigen comparisons (Figures 2G–2I). The same was observed when the correlations of the matched protein-specific $CD4^+$ and $CD8^+$ T cell responses were considered (Figures S2A–S2C).

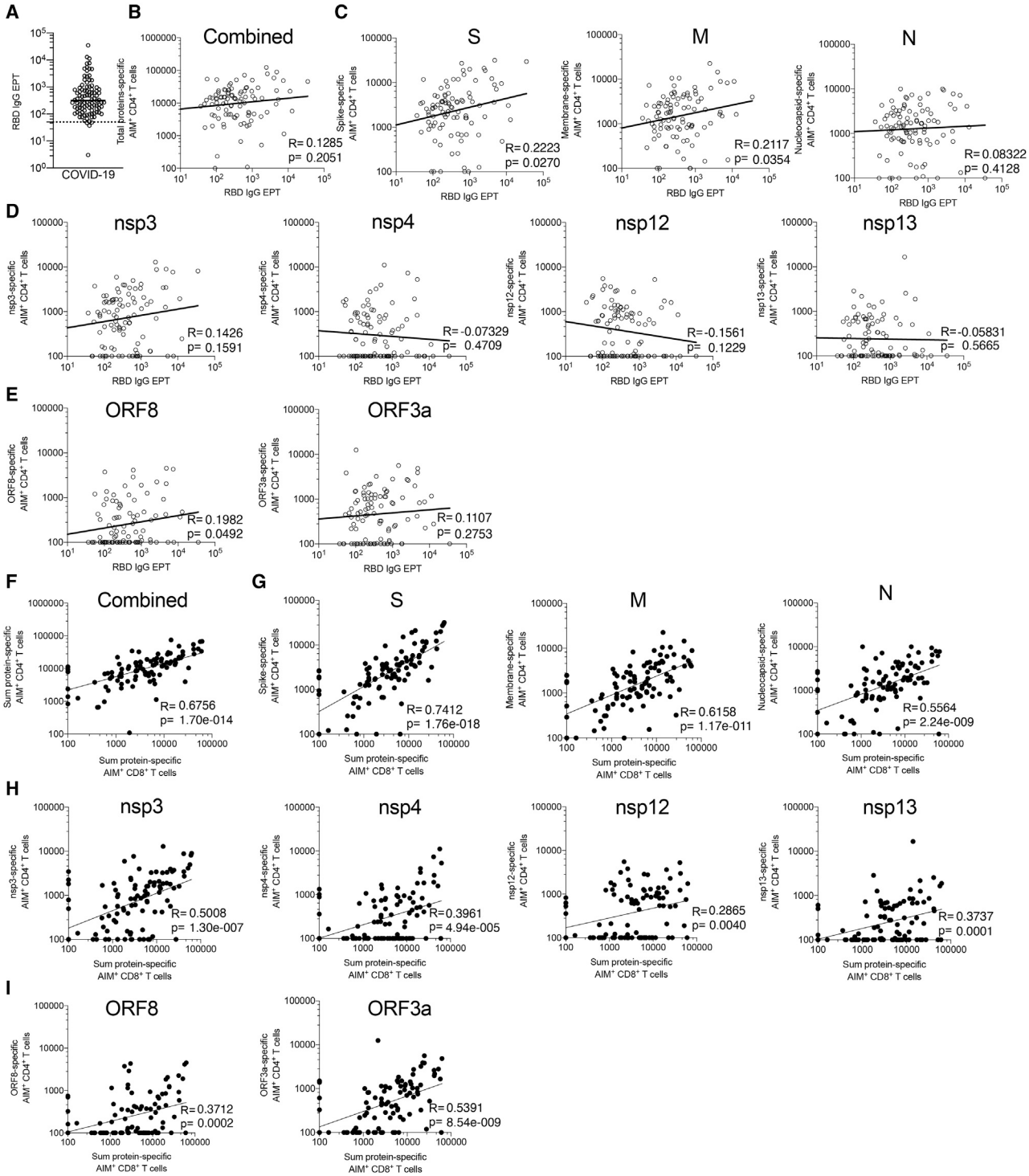


Figure 2. SARS-CoV-2-specific CD4$^+$ T cell reactivities and their correlations with antibody production and CD8$^+$ T cell reactivity

(A) RBD IgG serology is shown for all the convalescent COVID-19 donors (n = 99) of this cohort.

(B–E) Serology data of (A) are correlated with CD4$^+$ T cell reactivities specific against all combined proteins (B), structural proteins S, M, and N (C), non-structural proteins nsp3, nsp4, nsp12, and nsp13 (D), and ORF8 and ORF3a (E).

(F–I) The total CD8$^+$ T cell reactivity is correlated with the total CD4$^+$ T cell reactivity (F) and the CD4$^+$ T cell reactivity against structural proteins S, M, and N (G), non-structural proteins nsp3, nsp4, nsp12, and nsp13 (H), and ORF8 and ORF3a (I).

Empty and filled circles represent correlation between CD4$^+$ T cell reactivity and serology or CD8$^+$ T cell reactivity, respectively. All analyses were performed using Spearman correlation, and the p values shown were not corrected for multiple hypothesis testing.

These data overall suggest that the $CD4^+$ T cell response against all dominant antigens is potentially relevant in terms of providing helper function for $CD8^+$ T-cell-specific responses. However, it cannot be excluded that what we observed is due only to the $CD4^+$ T cell help or to other extrinsic and intrinsic properties of antigen presentation and responsiveness within an individual. This might reflect that T cell responses correlate with gene expression. S, N, and M may be immunodominant because of the very high gene expression for each.[22] In this context, it is perhaps surprising that a strong $CD4^+$ and $CD8^+$ T cell response was elicited by nsp3, which is not known to be expressed at high levels.[22]

SARS-CoV-2 peptides and epitope screening strategy

The analysis of the SARS-CoV-2 proteome summarized above identified the major viral antigens accounting for 80% or more of the total $CD4^+$ and $CD8^+$ T cell response. These antigens were then introduced into the epitope screening pipeline (Figure S3A). Because class II epitope prediction is not as robust as class I prediction,[23] and because of the high degree of overlap in binding capacity of different HLA class II alleles, to determine $CD4^+$ T cell reactivity in more detail, we followed a comprehensive and unbiased approach based on the use of complete sets of overlapping peptides spanning each antigen and composition of antigen-specific peptide pools. Positivity was defined as net AIM^+ counts (background subtracted by the average of triplicate negative controls) >100 and a stimulation index (SI) > 2, as previously described.[24] Positive peptide pools were deconvoluted to identify the specific 15-mer peptide(s) recognized. For large proteins, such as S, an intermediate "mesopool" step was used to optimize use of reagents.

In parallel, we synthesized panels of predicted HLA class I binders for the 28 most common allelic variants (Table S4), as described in the STAR methods section. The top two hundred predicted peptides were synthesized for each allele, leading to 5,600 predicted HLA binders in total. To identify $CD8^+$ T cell epitopes, we tested individual peptides derived from the specific antigen(s) recognized by $CD8^+$ T cells of individual donors and that were predicted to bind the HLA class I alleles expressed by the respective donor (Figures 1B and 1E). To quantify the population coverage provided by the HLA class I alleles selected for study, we tabulated the fraction of the donor cohort studied where allele matches were identified for 0, 1, 2, 3, or 4 of the respective HLA A and B alleles expressed by the donor. We found that 98% of the participants in our cohort were covered by at least one allele, 91% by 2 or more, and 74% were covered by 3 or more of the alleles in our panel (Figure S1C). As shown in Table S3, focusing on the 8 most dominant SARS-CoV-2 antigens for the purpose of epitope identification allowed mapping of 80% or more of the response, although screening only 35%–40% of the total peptides.

To broadly identify T cell epitopes recognized in a cytokine-independent manner, we used the AIM assay mentioned above.[1,25] Examples of gating strategies, pool deconvolution, and epitope identification for both $CD4^+$ and $CD8^+$ T cell responses are shown in Figure S3B. AIM^+ cell counts were calculated per million $CD4^+$ or $CD8^+$ T cells, respectively.

Summary of $CD4^+$ T cell epitope identification results

To identify specific $CD4^+$ T cell epitopes, we deconvoluted peptide pools corresponding to antigens previously identified as positive for $CD4^+$ T cell activity in each specific donor (Figure 1A). In instances where not all positive pools could be deconvoluted due to limited cell availability, peptide pools were selected for screening to ensure that each of the 9 major antigens was tested in at least 10 donors. Overall, we were able to test each peptide for these antigens in a median of 13 donors (range 10–17). Each donor was previously determined to be positive for $CD4^+$ T cell responses to that specific antigen.

Taken together, a total of 280 SARS-CoV-2 $CD4^+$ T cell epitopes were identified, including 3 nsp16 (this protein was not included in the top proteins studied) epitopes identified in parallel experiments in 2 donors (Table S5). We found that each donor responded to an average of 3.2 viral antigens (Figure 1D), and 5.9 $CD4^+$ T cell epitopes were recognized per antigen for the top 80% most immunodominant antigens (data not shown). For each epitope/responding donor combination, potential HLA restrictions were also inferred based on the predicted HLA binding capacity of the epitope for the HLA alleles present in the respective responding donor (listed in Table S2), as previously described (Voic et al., 2020).[15,26] Table S6 provides the spectrum of distributions of the magnitudes of T cell responses to all peptides tested at the level of the individual donors.

HLA binding capacity of dominant epitopes

A total of 109 of the 280 epitopes were recognized by 2 or more donors, accounting for 71% of the total response. The 49 most dominant epitopes, recognized in 3 or more donors, accounted for 45% of the total response (Figure 3A).

Because dominant epitopes are associated with promiscuous HLA class II binding,[27,28] defined as the capacity to bind multiple HLA allelic variants, we investigated the role of HLA binding in determining immunodominant SARS-CoV-2 epitopes. Specifically, we measured the *in vitro* binding capacity of the 49 most dominant epitopes (positive in 3 or more donors, as mentioned above) for a panel of 15 of the most common DR alleles using individual peptides and purified HLA class II molecules.[29] The results are provided in Table S7. It was noted that, in general, a good correlation was observed between predicted and measured binding ($R = 0.6604$; $p = 2.97 \times 10^{-93}$; Figure S4A). Based on these results, we further characterized those 49 most dominant epitopes using predicted binding for additional HLA class II alleles, including a panel of the 12 most common HLA-DQ and DP allelic variants, and all HLA class II variants (DR, DQ, and DP) expressed in the cohort.

Overall, the 49 most dominant epitopes showed significantly higher binding promiscuity (number of alleles bound at the 1,000 nM or better threshold)[30,31] for the panel of common HLA class II than a control group of 49 non-epitopes derived from the same proteins (average number of HLA predicted to be bind ± SD epitopes = 10.8 ± 6.5; non-epitopes = 5.7 ± 6; $p = 0.0001$ by Mann-Whitney; Figures S4B and S4C). The same conclusion was reached when the full set of HLA

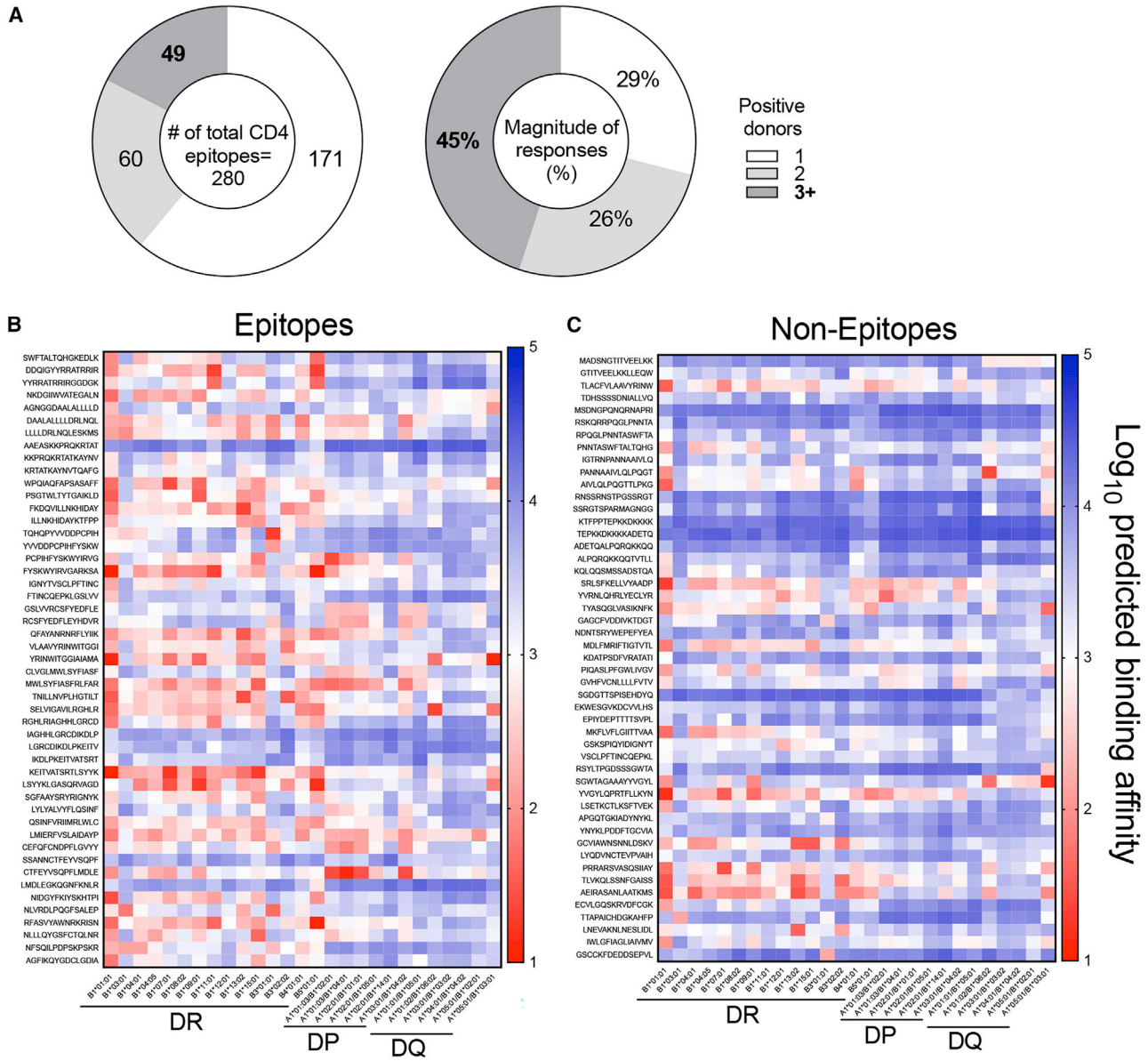


Figure 3. Heat maps of HLA predicted binding patterns in the 27 most frequent HLA class II alleles

(A) SARS-CoV-2 CD4$^+$ T cell epitopes as a function of the number of responding donors (n = 44 convalescent COVID-19 donors) recognized and strength of responses.

(B and C) Predicted binding patterns for the top 49 most immunodominant SARS-CoV-2 CD4$^+$ T cell epitopes (B) are compared with a set of matched non-epitopes (C). Predicted half maximal inhibitory concentration (IC_{50}) was calculated using NetMHCIIpan and converted to $\log_{10}$ scale. Lower values indicate stronger predicted binding affinity and are highlighted at the red end of the spectrum. Predicted values with an IC_{50} < 1 000 nM ($\log_{10}$ scale < 3) are considered positive binders.

alleles present in the cohort was considered using the same criteria (average ± SD epitopes = 24.3 ± 15.2; non-epitopes = 13.2 ± 14.1; p = 0.0003 by Mann-Whitney; Figures S4D and S4E).

Heatmaps of the 49 epitopes and non-epitopes considering the panel of common HLA DR, DP, and DQ are shown in Figures 3B and 3C. These results confirm that broad HLA binding capacity is a key feature of dominant epitopes. It further indicates that, because of their broad binding capacity, these epitopes are likely to be recognized in different geographical settings and different ethnicities.

Similarity of SARS-CoV-2 CD4$^+$ T cell epitopes to CCC sequences

Several studies have reported significant pre-existing immune memory to SARS-CoV-2 peptides in unexposed donors.[1,3,4,15] This reactivity was shown to be associated, at least in some instances, with memory T cells specific for human CCCs cross-

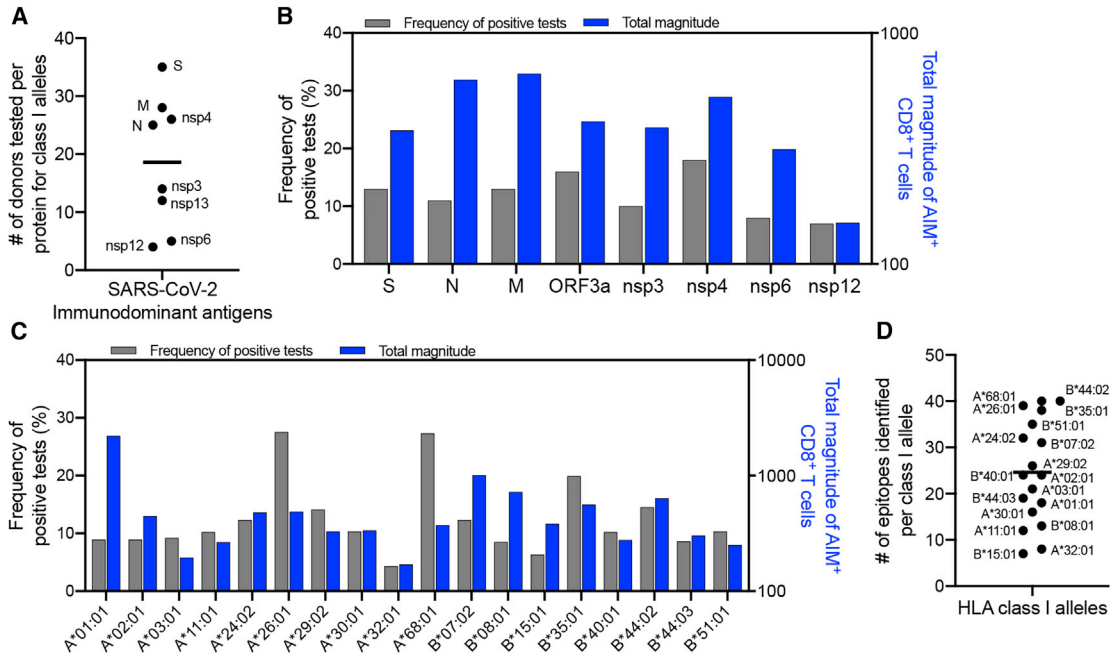


Figure 4. Distribution of SARS-CoV-2 CD8+ T cell responses by antigen and class I allele

(A) The number of donors tested with their HLA-matched class I peptides for each of the 8 dominant proteins for $CD8^+$ (n = 40 convalescent COVID-19 donors with a range of 4 to 35 donors tested per protein).

(B and C) The distribution of allele-specific $CD8^+$ responses for the 18 class I alleles that were tested in 3 or more donors is shown as function of protein composition (B) or the HLA class I alleles tested (C). Blue bars represent the total magnitude of AIM^+ $CD8^+$ T cells divided by the number of positive donors. Gray bars represent the frequency of positive tests.

(D) The total number of epitopes identified for each class I allele is shown in panel.

reactively recognizing SARS-CoV-2 sequences.[3,15] In particular, it was shown that the SARS-CoV-2 epitopes recognized in unexposed donors had significantly higher homology to CCC than SARS-CoV-2 sequences not recognized in unexposed donors. Here, using the exact same methodology,[15] we performed the converse analysis, namely an analysis of the homology between the $CD4^+$ T cell epitopes experimentally identified in COVID-19 donors (Figure S5) and sequences of peptides derived from the four widely circulating human CCCs (NL63, OC43, HKU1, and 229E). No significant differences were observed based on percent sequence identity between epitopes recognized from the COVID-19 cohorts and non-epitope controls in structural proteins S, M, and N and accessory proteins encoded by ORF3a and ORF8 or non-structural proteins (Figure S5A).

Indeed, in our previous studies,[1,15] we noted that the pattern of antigen recognition in exposed and unexposed donors was significantly different. Here, having defined the actual epitopes recognized in COVID-19, we compared them to the epitopes previously identified in unexposed donors. The present study re-identified 50% of the epitopes in our COVID-19 cohort but in addition identified 227 $CD4^+$ T cell epitopes specific for SARS-CoV-2 infection (Figure S5C). Thus, more than 80% (227/280) of the epitopes identified herein were not previously seen in the unexposed cohort. These results are consistent with the notion that, although a cross-reactive repertoire is present in unexposed donors, SARS-CoV-2 infection elicits a vast repertoire of additional T cell specificities.

Summary of $CD8^+$ T cell epitope identification results

Following the approach described above, a total of 523 SARS-CoV-2 $CD8^+$ T cell epitopes were identified (Table S8). These epitopes are associated with 26 different HLA restrictions, based on predicted HLA binding capacity matched to the HLA alleles of the responding donor. A complete list of the synthesized class I peptides and the corresponding magnitude of T cell responses for individual donors can be found in Table S6. For eight HLAs, only 1 to 2 donors expressing the matching HLA could be tested. Predicted binders for the remaining 18 HLAs were tested in a median of 5 donors (range 3–9). The 8 most immunodominant proteins were screened in an average of 19 donors (range 4–35; Figure 4A). Of the 523 $CD8^+$ T cell epitopes identified, 61 were recognized in 2 or 3 different donor-allele combinations, meaning that there were 454 unique peptides recognized. Of these, 101 (22%) were recognized by 2 or more donors, accounting for 49% of the total response. We found that each donor recognized an average of 2.7 antigens (Figure 1F) and responded to an average of 1.6 $CD8^+$ T cell epitopes per antigen per HLA allele (data not shown). Considering 4 HLA A and B alleles in each donor, we expect at least 17 epitopes per donor for class I ($2.7 \times 1.6 \times 4 = 17.3$).

Figure 4 shows the frequency of positive epitopes (identified epitopes/peptides screened), and the average magnitude of epitope responses (total magnitude of response normalized by the number of positive donors), as a function of protein (Figure 4B) or HLA class I allele (Figure 4C) analyzed. Each HLA

was associated with an average of 25 epitopes (range 7–40; median 24; Figure 4D). Interestingly, as also previously detected in other systems,[32,33] there was a wide variation as a function of HLA allele. Some alleles, such as A*03:01 and A*32:01, were associated with responses that were both infrequent and weak; in other cases (e.g., A*01:01), responses were infrequent but when observed were of high magnitude. Finally, and conversely, other alleles were associated with relatively frequent but low-magnitude responses (e.g., A*68:01). This effect was previously linked to differences in the size of peptide repertoires associated with different HLA motifs.[20]

In terms of antigen specificity of $CD8^+$ T cell responses, relatively similar epitope-specific response frequencies were observed for the various antigens, with the exception of nsp12, which was associated with responses of low frequency and magnitude (Figure 4B). These results should be interpreted with the caveat in mind that the donors screened were pre-selected on the basis of association with positive responses to that particular antigen; thus, these data do not directly address protein immunodominance, which is instead addressed in Table S3. These data instead point to the relative frequency and magnitude of responses at the level of individual epitopes associated with a given antigen, which were found to be overall similar.

To address the potential relationship between $CD8^+$ T cell epitope recognition and CCC homology, as performed above in the case of $CD4^+$ T cell epitopes, we analyzed the homology of the $CD8^+$ T cell epitopes to CCCs (NL63, OC43, HKU1, and 229E), as compared to the homolog to the same CCC viruses detected in the case of peptides that tested negative in all donors tested, regardless of the HLA restriction (Figure S5). Similar to what was observed in the context of $CD4^+$ T cell responses, the $CD8^+$ T cell epitopes recognized in convalescent COVID-19 donors were not associated with higher sequence identity to CCC as compared to non-epitopes when structural, accessory, or non-structural proteins (Figure S5B) were considered.

Distribution of $CD4^+$ and $CD8^+$ T cell epitopes within dominant SARS-CoV-2 antigens

We next analyzed the distribution of $CD4^+$ and $CD8^+$ T cell epitopes within the dominant SARS-CoV-2 S, N, and M antigens (Figure 5). For each antigen, we show the frequency (red line) and magnitude (black line) of $CD4^+$ T cell responses along the antigen sequence, considering regions with response frequency above 20% as immunodominant. Based on the results presented above, we also plotted HLA class II binding promiscuity (defined as the number of HLA allelic variants expressed in the donor cohort predicted to be bound by a given peptide) and the degree of homology of each 15-mer peptide for aligned CCC antigen sequences. The bottom panel represents the distribution of $CD8^+$ T cell epitopes (black) and non-epitopes (red) along the antigen sequence.

Responses to S peptides with a frequency of 20% or higher were focused on discrete regions of the protein involving the N-terminal domain (NTD), the C-terminal (CT) 686–816 region, and the neighboring fusion protein (FP) region; only a few responses were focused on the RBD. These immunodominant regions are boxed in red in Figure 5A. We expected HLA-binding capacity to be associated with T cell immunodominant regions and indeed found a significant positive correlation with the frequency of responses ($R = 0.2231$; $p = 0.0003$ by Spearman correlation; Figure S5D). No significant correlation ($R = -0.03144$; $p = 0.6187$ by Spearman correlation; Figure S5E) was found with sequence homology to CCC (calculated as maximum sequence homology to the four main CCC species). As indicated in the 3D rendering of the S crystal structure (PDB: 6XR8), these immunodominant regions were mostly located in the surface-exposed portions of the S monomer and were not particularly influenced by the glycosylation pattern (shown in Figure 5A as stars in the linear structure description and based on experimental identification by Cai and co-authors[34]). The glycosylation patterns are also shown in the 3D rendering of the corresponding crystal structure, based on curation done by the authors of the same manuscript, and shown as gray dots (Figure 6A). We further explored the correlation between $CD4^+$ T cell immunodominance and location of proteolytic cleavage sites, utilizing the major histocompatibility complex class II (MHCII)-NP algorithm.[35] The results did not reveal any significant correlation between the predicted cleavage sites and immunodominant regions (Spearman correlation has $R = -0.08426$ and $p = 0.1816$; Figure S5F). This is consistent with previous results that indicated that predicted cleavage sites do not significantly improve epitope predictions.[35] Finally, $CD8^+$ T cell reactivity did not reveal any particular immunodominant region in S, with epitopes and non-epitopes roughly equally distributed along the sequence (Figure 5A).

In the same way, we compared responses observed within the N and M proteins as a function of structural protein composition, HLA promiscuity, and CCC homology (Figures 5B, 5C, 6B, and 6C). For the N protein (Figure 5B), the majority of the response was focused on the NTD and CTD regions, with lower contributions from the linker region (all outlined in red boxes); segments in the middle and toward the ends of the protein were devoid of any reactivity. The correlation between immunodominance and HLA binding promiscuity was even stronger than observed for S ($R = 0.4725$; $p = 7.41 \times 10^{-6}$; Figure S5G). Similar to what was observed for the S protein, no significant correlation between the frequency of positive responses was observed with CCC similarity ($R = 0.1660$; $p = 0.1362$; Figure S5H) or predicted cleavage sites ($R = -0.009245$; $p = 0.9343$; Figure S5I). The immunodominance of N-specific $CD8^+$ T cell responses mirrors the one observed for the $CD4^+$ T cell counterpart, highlighting that, in general, the N-terminal and C-terminal domains are the major immunodominant regions of N recognized by both T cell types.

$CD4^+$ T cell immunogenic regions were distributed across the entire span of the M protein (Figure 5C), including the transmembrane region (Figure 6C). No significant correlation was observed when investigating HLA binding promiscuity ($R = 0.2374$; $p = 0.1253$; Figure S5J), CCC similarity ($R = 0.07648$; $p = 0.6259$; Figure S5K), or predicted cleavage sites ($R = 0.08421$; $p = 0.5913$; Figure S5L). The lack of correlation between M epitopes and HLA binding is consistent with the interpretation that M is a prominent antigen because it is highly expressed, not because it contains high-quality epitopes. No particular immunodominance patterns were observed for the M protein with respect to $CD8^+$ epitopes.

Finally, when we investigated the location of immunodominant T cell regions relative to the main sites identified for

CellPress
OPEN ACCESS

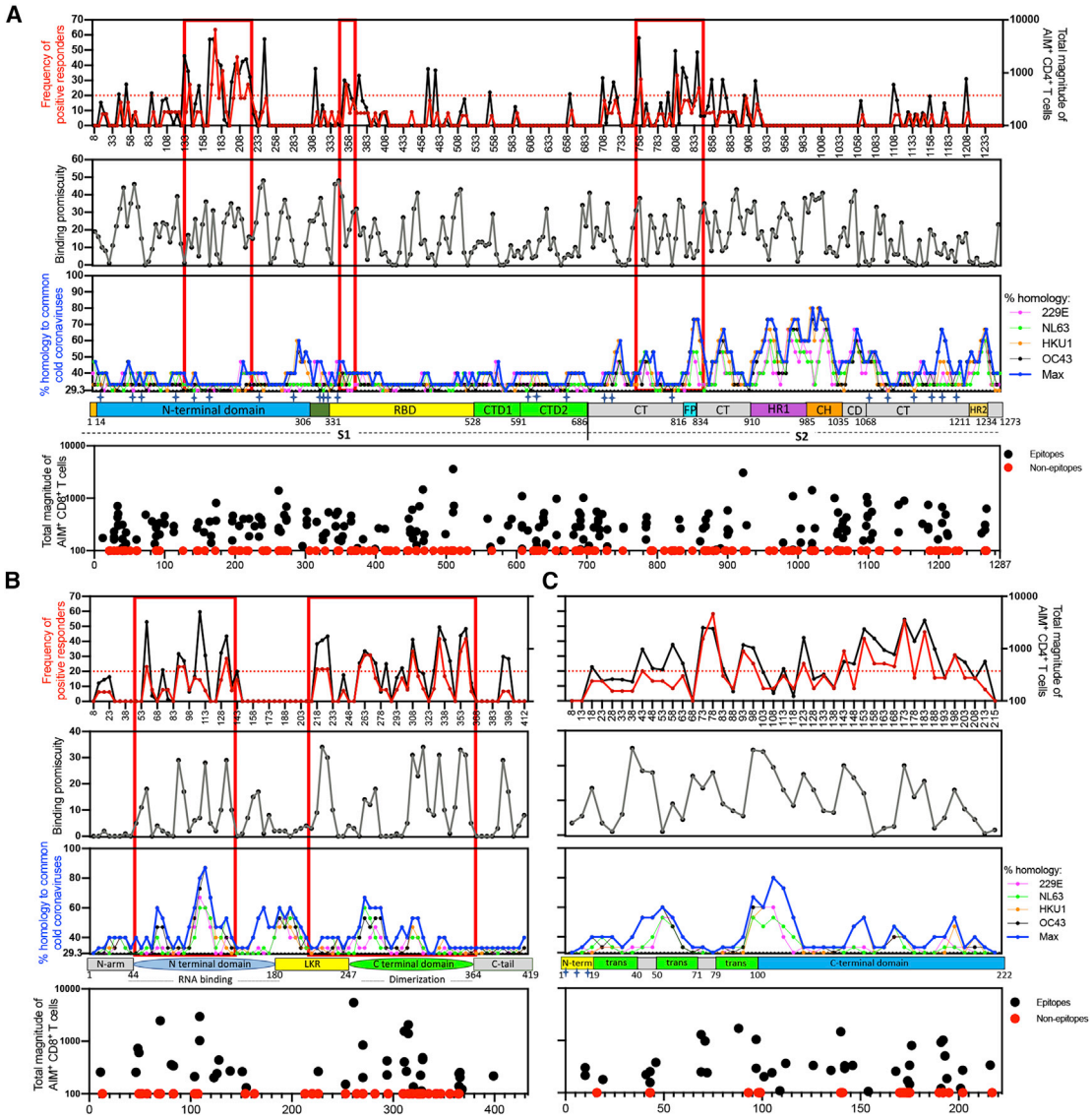


Figure 5. Immunodominant regions for CD4+ T cells S, N, and M proteins.

(A) S, (B) N, and (C) M proteins as a function of the frequency of positive response (red) and total magnitude (black) in the topmost panel. The dotted red line indicates the cutoff of 20% frequency of positivity used to define the immunodominant regions boxed in red and also shown in red in Figure 6. The x axis labels in this topmost panel indicate the middle position of the peptide. Binding promiscuity was calculated based on NetMHCIIpan predicted IC_{50} for the alleles present in the cohort of donors tested and is shown in gray on the upper middle panel. The lower middle panel shows the % homology of SARS-CoV-2 to the four most frequent CCC (229E in pink, NL63 in green, HKU1 in orange, and OC43 in black) and the max value (blue). The linear structure of each protein is drawn below the graph of homology. The magnitude of $CD8^+$ responses to class I predicted epitopes is shown in the bottom panel, where black dots represent epitopes and red dots represent non-epitopes, each centered on the middle position of the peptide. PBMCs from convalescent COVID-19 donors (n = 44) were tested for reactivity to the peptides indicated in the topmost and bottommost panels (A)–(C).

antibody reactivity,[36] the $CD4^+$ T cell immunodominant regions identified in S and N showed minimal overlap with immunodominant linear regions targeted by antibody responses (Figure 6). The $CD4^+$ T cell epitope recognition patterns of ORF3a, ORF8, nsp3, nsp4, nsp12, and nsp13 are shown in Figure S6. The ORF8 protein was similar to M in that epitopes throughout both of these small proteins were recognized. ORF3a had clear regions of response clustered in the middle and at the C terminus. Nsp3, which was the 4th most immunodominant antigen, was associated with a rather striking immunodominant region centered around residue 1,643. Other non-structural proteins were less immunodominant overall but had discreet regions targeted by $CD4^+$ T cell responses (i.e., residue 5,253 for nsp12).

Reactivity of megapools based on the experimentally identified epitopes

The experiments described above identified a total of 280 $CD4^+$ and 454 $CD8^+$ T cell epitopes. These epitopes were arranged

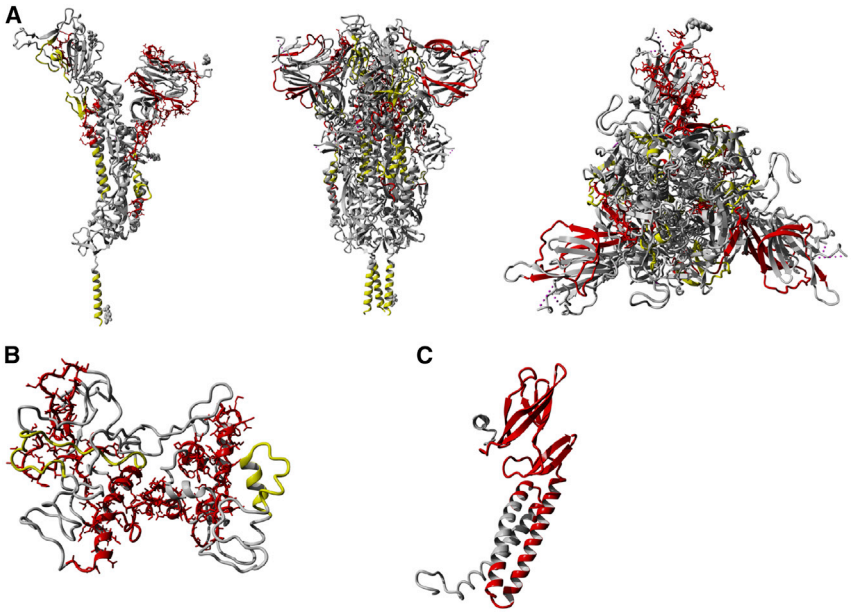


Figure 6. Immunodominant regions for CD4+ T cells and B cells in relation to the 3D rendering of S, N, and M proteins

3D rendering of S (A), N (B), and M (C) proteins. The drawings show in gray the 3D structures, in red the $CD4^+$ T cell immunodominant regions for each protein with frequency of positive responses >20% (also shown in red in Figure 5), and in yellow the B cell immunodominant regions for each protein based on the work of Shrock et al.[36] Glycosylation sites for S are shown as gray dots and are based on information embedded in the original crystal structure shown to map the immunodominant regions (PDB: 6XR8).

(A) The S protein is shown as monomer on the left and trimer in the middle and on the right (side and top views).

(B) N protein 3D rendering was based on a model generated using Phyre2. Additional details about the N model are available in the STAR methods section.

(C) The M protein is shown as a monomer according to a model previously described by Heo et al.[37]

All the 3D renderings have been performed using the free version of YASARA.

into two epitope megapools (MPs), CD4-E and CD8-E, respectively (where the E denotes "experimentally defined"). These MPs were tested in a new cohort of 31 COVID-19 convalescent donors (none of these donors were utilized in the epitope identification experiments) and 25 unexposed controls (Table S1). MP reactivity was assessed for all donors using AIM and interferon γ (IFNγ) FluoroSpot assays.

To put the results in context, we also tested peptides contained in the CD4-R and CD4-S and CD8-A and CD8-B MPs previously utilized to measure SARS-CoV-2 $CD4^+$ and $CD8^+$ T cell responses, respectively.[1,2,6,15] These MPs are based on either overlapping peptides spanning the entire S sequence (CD4-S) or predicted peptides (all other proteins). Although these pools contain a larger total number of peptides (474 for CD4-R + CD4-S and 628 for the CD8-A + CD8-B) than the corresponding experimentally defined sets, we expected that the experimentally defined peptide sets would be able to recapitulate the reactivity observed with the previously utilized MPs. As a further context, we also tested the T cell epitope compositions (ECs) class I and EC class II pools of experimentally defined $CD8^+$ and $CD4^+$ epitopes described by Nelde et al.,[14] encompassing 29 and 20 epitopes each, which prior to this study represented the most comprehensive set of experimentally defined epitopes.

As might be expected, the results showed that the AIM assay was more sensitive than the FluoroSpot assay (Figure 7). On the other hand, as a tradeoff for the lower signal, the FluoroSpot assay showed higher specificity in the responses detected, with fewer unexposed individuals showing any reactivity compared to the AIM assay. For $CD4^+$ T cell responses as detected in the AIM assay (Figure 7A), the CD4-E MP recapitulated the reactivity observed with the MPs of larger numbers of predicted peptides (CD4-R+S) and showed significantly higher reactivity ($p = 4.30 \times 10^{-6}$ by Mann-Whitney) as compared to the EC class II pool. A similar picture was observed when the FluoroSpot assay was utilized (Figure 7B), with a significantly higher reactivity of the CD4-E MP compared to the CD4-R+S ($p = 0.0208$ by Mann-Whitney) and to the EC class II pool ($p = 1.39 \times 10^{-7}$ by Mann-Whitney). In both AIM and FluoroSpot assays, the CD4-E MP showed the highest capacity to discriminate between COVID-19 convalescent and unexposed donors ($p = 3.19 \times 10^{-10}$ and $p = 1.56 \times 10^{-9}$, respectively, by Mann-Whitney).

A similar picture was noted in the case of $CD8^+$ T cell reactivity (Figures 7C and 7D), where the CD8-E MP recapitulated the reactivity observed with the MPs of larger numbers of predicted peptides (CD8-A+B), with a strong trend ($p = 0.0551$ by Mann-Whitney) toward more reactivity than the EC class II pool. In the case of the FluoroSpot assay, we noted equivalent reactivity for the CD8-E and CD8-A+B MPs and significantly higher reactivity ($p = 0.0219$ by Mann-Whitney) than the EC class II pool (Figure 7D). In both assays, the CD8-E MP showed highest capacity to discriminate between COVID-19 convalescent and unexposed subjects ($p = 1.47 \times 10^{-8}$ and $p = 1.48 \times 10^{-8}$, respectively, by Mann-Whitney). To test how well the different T cell responses measured separate individuals that have been exposed to SARS-Cov-2 versus those that do not, we performed receiver operating characteristic curve (ROC) analyses (Figures 7E–7H), which allow us to directly compare the classification success based on true- and false-positive rates. The CD4-E and CD8-E response data were associated with the best performance.

Considering that a potential practical limitation in the characterization of SARS-CoV-2 responses is the number of cells available for study, in selected COVID-19 donors, we titrated the number of PBMCs/well to determine whether a response could be measured with lower cell numbers. As expected, as the cell input was decreased, the magnitude of responses decreased correspondingly. Although marginal responses were seen with

CellPress
OPEN ACCESS

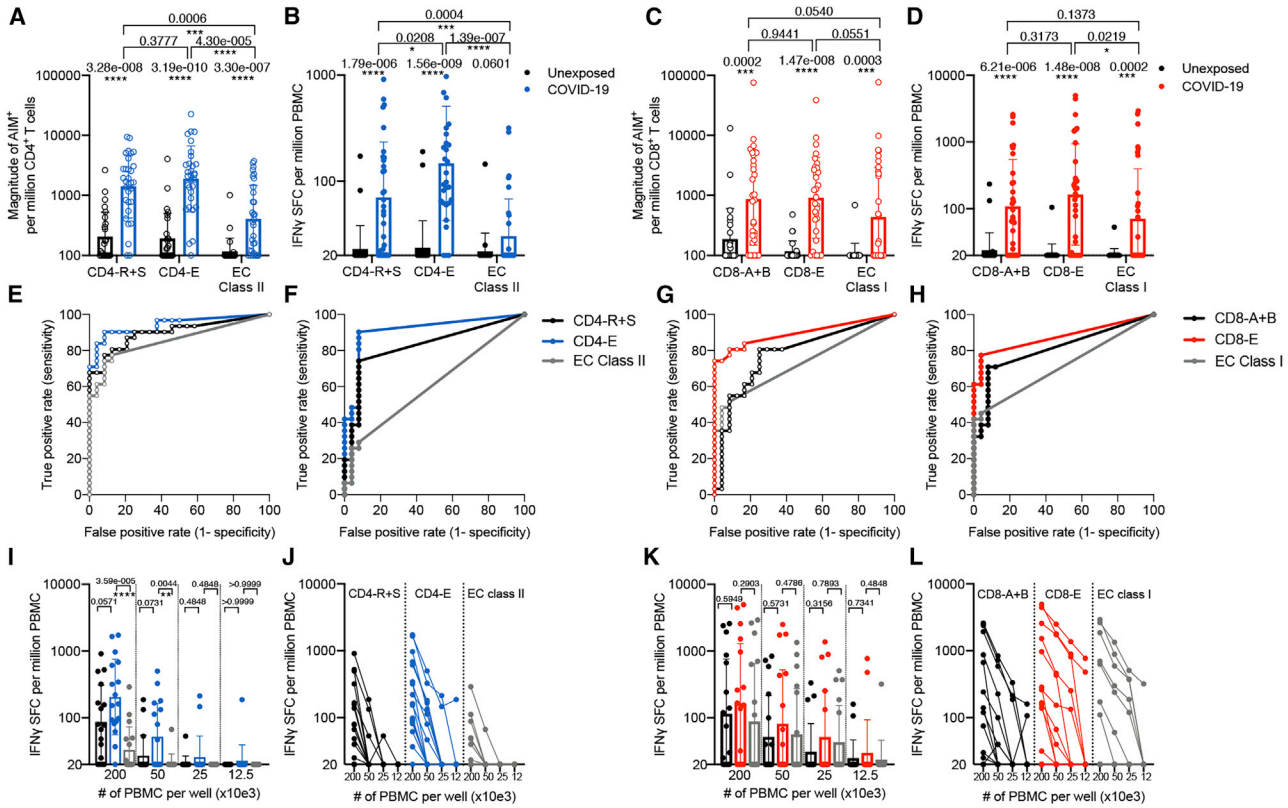


Figure 7. T cell responses to SARS-CoV-2 megapools as measured in AIM (empty circles) and FluoroSpot (filled in circles) assays

(A–D) Twenty-five unexposed and 31 convalescent COVID-19 donors were tested in the AIM assays (A and C), and all donors were also tested in the FluoroSpot assays (B and D).

(A and B) $CD4^+$ T cell responses to CD4-R+S (previously described), CD4-E (280 class II epitopes identified in this study), and EC class II[14] megapools were measured via AIM (A) and FluoroSpot (B). Bars represent geometric mean ± geometric SD, and p values were calculated by Mann-Whitney.

(C and D) $CD8^+$ T cell responses to CD8-A+B (previously described), CD8-E (454 class I epitopes identified in this study), and EC class I[14] megapools were measured via AIM (C) and FluoroSpot (D). Bars represent geometric mean ± geometric SD, and p values were calculated by Mann-Whitney.

(E–H) ROC analysis for $CD4^+$ and $CD8^+$ T cell response data in FluoroSpot (F–H) and AIM (E–G) assays.

(I–L) Additionally, we further tested 17 of these COVID-19 convalescent donors in FluoroSpot with a titration of 200, 50, 25, and 12.5 × 10^3 cells per well with the indicated CD4-MPs (I and J) and CD8-MPs (K and L). (I and K) Bars represent geometric mean ± geometric SD, and p values were calculated by Mann-Whitney.

25,000 cells/well and below, a sizeable response was still detectable with 50,000 cells/well, with 8 out of 17 donors responding for the CD4-E MP (as compared to 16 out of 17 in the case of 200,000 cell level). Similarly, in the case of the CD8-E MP, 8 out of 17 donors responded with 50,000 cells/well (as compared to 11 out of 17 in the case of 200,000 cell level). The frequency and magnitude of responses of CD4-E were higher compared to the EC class II ($p = 3.59 \times 10^{-5}$ and $p = 0.0044$ by Mann-Whitney; Figures 7I and 7J). The CD8-E MP was also associated with a higher magnitude of response than the EC class I pool (Figures 7K and 7L). In conclusion, these results underline the biological relevance of the more comprehensive CD4-E and CD8-E MPs.

DISCUSSION

This study presents a comprehensive analysis of the patterns of epitope recognition associated with SARS-CoV-2 infection in a cohort of approximately 100 different convalescent donors spanning a range of peak COVID-19 disease severity representative of the observed distribution in the San Diego area. SARS-CoV-2 was probed using 1,925 different overlapping peptides spanning the entire viral proteome, ensuring an unbiased coverage of the different HLA class II alleles expressed in the donor cohort. For HLA class I, we used an alternative approach, selecting 5,600 predicted binders for 28 prominent HLA class I alleles, representing 61% of the HLA A and B allelic variants in the worldwide population, and affording an overall 98.8% HLA class I coverage at the phenotypic level.

The biological relevance of the epitope characterization studies summarized here is underlined by the use of the *ex vivo* AIM assay that does not require *in vitro* stimulation, which potentially skews the results by eliciting responses from naive cells. The AIM assay is also more agnostic for different types of $CD4^+$ T cells, as it measures all activated cells, regardless of T cell subset or any particular pattern of cytokine secretion.

We are not aware of any study that describes the repertoire of $CD4^+$ and $CD8^+$ T cell epitopes recognized in SARS-CoV-2 infection with a comparable level of granularity or breadth. Although

several previous reports have described SARS-CoV-2 epitopes, and accordingly represent very useful advances, these studies either utilized *in vitro* expansion,[14] were limited in the number of proteins analyzed,[4] characterized responses in fewer than 10 HLA types,[10,11,14] or focused on TCR repertoire after *in vitro* expansion of small numbers of cells.[12] Comparing our results with those obtained in those previous studies, we note that, of the 20 HLA class II peptides identified by Nelde and co-authors,[14] 14 were contained within proteins we mapped here in detail, and we independently re-identified 12 (86%) of them (identical or largely overlapping sequences). Of 137 class I peptides reported thus far,[10,11,14] 98 were contained within the viral proteins we mapped in detail, and we independently re-identified 68 (69%) of them (identical or largely overlapping sequences).

Importantly, because SARS-CoV-2 antigen-specific T cell responses were evaluated in a systematic and unbiased fashion, quantitative estimates of the size of the repertoire of T cell epitope specificities recognized in each donor can be derived. Determining the breadth of responses is of relevance, because previous studies[11,12] have suggested narrow SARS-CoV-2-specific T cell repertoires in COVID-19 patients; notably, a limited repertoire could favor viral mutation, a particular concern with this RNA virus. Based on our results, we expect that each donor would be able to recognize about 19 $CD4^+$ T cell epitopes, on average. Likewise, for $CD8^+$ T cells, we expect at least 17 epitopes per donor to be recognized. Overall, T cell responses in SARS-CoV-2 are estimated to recognize even more epitopes per donor than seen in the context of other RNA viruses, such as dengue,[38,39] where 11.6 and 7 $CD4^+$ and $CD8^+$ T cell epitopes, respectively, were recognized on average. This analysis should allay concerns over the potential for SARS-CoV-2 to escape T cell recognition by mutation of a few key viral epitopes.

We defined the patterns of immunodominance across the various antigens encoded in the SARS-CoV-2 genome recognized in COVID-19 donors. Consistent with earlier reports from our group[1] and others,[10] we see clear patterns of immunodominance, with a limited number of antigens accounting for about 80% of the total response. In general, the same antigens are dominant for both $CD4^+$ and $CD8^+$ responses, with some differences in relative ranking, such as in the case of nsp3, which is relatively more dominant for $CD8^+$ than $CD4^+$ T cell responses. Immunodominance at the protein level correlated with protein abundance/gene, as previously noted for $CD4^+$ T cell responses,[22] although we note that the accessory proteins and nsps also account for a significant fraction of the response despite their predicted lower abundance in infected cells.

Because of their role in instructing both antibody and $CD8^+$ T cell responses, we correlated $CD4^+$ T cell activity on a per donor and per antigen level with antibody and $CD8^+$ T cell adaptive responses. This enabled establishing which antigens have functional relevance in terms of eliciting $CD4^+$ T cell responses correlated with antibody and $CD8^+$ T cell responses. At the level of antibody responses, S and M were correlated with RBD antibody titers, highlighting their capacity to support antibody responses, presumably by a deterministic linkage (viral antigen bridge) and cognate interactions.[40] Surprisingly, N-specific $CD4^+$ T cell responses did not correlate with S RBD antibody titers, suggesting unexpected complexity of the N-specific $CD4^+$ T cell response. By contrast with these selective effects, $CD4^+$ T cell activity against any of the antigens correlated with the total $CD8^+$ T cell activity, suggesting that the role of $CD4^+$ T cell responses driven by the different proteins is determinant in its helper function for either RBD-specific antibody production or $CD8^+$ T cell responses. This was particularly true in both contexts when looking specifically at the S and M proteins, which are also the strongest and most frequently recognized antigens for both $CD4^+$ and $CD8^+$ T cells.

After examining relative immunodominance at the level of the different SARS-CoV-2 antigens, we probed for variables that may influence which specific peptides are recognized within a given antigen/ORF. Previously, we have shown that SARS-CoV-2 sequences recognized in unexposed individuals were associated with a higher degree of similarity to sequences encoded in the genome of various CCCs. Here, repeating the same analysis with the SARS-CoV-2 epitopes recognized in COVID-19 donors, we found no significant correlation. We further show that although a large fraction of the epitopes previously identified in unexposed donors are re-identified in COVID-19 donors, about 80% of the epitopes are not previously seen in unexposed, suggesting that the SARS-CoV-2-specific T cell repertoire of COVID-19 cases is overlapping but substantially different from the SARS-CoV-2-cross-reactive memory T cell repertoire of unexposed donors. This is consistent with our previous observation of a different pattern of reactivity[15] and consistent with reports from other groups.[4,14]

HLA binding capacity was a major determinant of immunogenicity for $CD4^+$ T cells (the influence of HLA binding was not evaluated for $CD8^+$ T cell, because the tested epitope candidates were picked based of their predicted HLA binding capacity). As found in several previous large-scale, pathogen-derived epitope identification studies, immunodominant epitopes were also found to be promiscuous HLA class II binders.[27,41] Binding to multiple HLA allelic variants is an important mechanism to amplify the potential immunogenicity of peptide epitopes and specific regions within an antigen. It is possible that the dominance of particular regions might further correlate with processing. However, at this juncture, HLA class II processing algorithms do not effectively predict epitope recognition.[35,42,43]

Further analysis projected the $CD4^+$ T cell dominant regions on known or predicted SARS-CoV-2 protein structures. This established that the dominant epitope regions are different for B and T cells. This is of relevance for vaccine development, as inclusion of antigen sub-regions selected on the basis of dominance for antibody reactivity might result in an immunogen devoid of sufficient $CD4^+$ T cell activity. In this context, it is important to note that the RBD region had very few $CD4^+$ T cell epitopes recognized in COVID-19 donors, but inclusion of regions neighboring the RBD N and C termini would be expected to provide sufficient $CD4^+$ T cell help.

In contrast to the clear demarcation of dominant regions for antibody and $CD4^+$ T cell responses, $CD8^+$ T cell epitopes were uniformly dispersed throughout the various antigens, consistent with previous in-depth analyses revealing little positional effect in $CD8^+$ T cell epitope distribution.[44] In the case of $CD8^+$ T cell responses, our data highlight HLA-allele-specific differences in the frequency and magnitude of responses. This

effect was noted before in the case of dengue virus[32] and related to potential HLA-linked protective versus susceptibility effects. The current study is not powered to test these potential effects, leaving it to future studies to examine this possibility. Regardless, our study provides a roadmap for inclusion of specific regions or discrete epitopes to allow for $CD8^+$ T cell epitope representation across a variety of different HLAs.

Finally, the functional relevance of our study was highlighted by the generation of improved epitope MPs for measuring T cell responses to SARS-CoV-2; these experimentally defined pools are associated with increased activity and lower complexity when compared to our previous MPs based on overlapping and predicted peptides. We plan to make these epitope pools available to the scientific community at large and expect that they will facilitate further investigation of the role of T cell immunity in SARS-CoV-2 infection and COVID-19.

In conclusion, we identify several hundred different HLA class I and class II restricted SARS-CoV-2-derived epitopes. We anticipate that these results will be of significant value in terms of basic investigation of SARS-CoV-2 immune responses and in the development of both multimeric staining reagents and T-cell-based diagnostics. In addition, the results shed light on the mechanisms of immunodominance of SARS-CoV-2, which have implications for understanding host-virus interactions, as well as for vaccine design.

Limitations of study

To maximize cell usage, our analysis was focused on the most dominantly recognized proteins. Screening for less commonly recognized proteins would require a larger cohort to enable identification of a sufficient number of donors responding to each protein. However, such expanded studies would be expected to yield additional epitopes.

The limited number of donors studied also did not allow investigation of responses directed against relatively rare HLA alleles, and HLA restrictions were not experimentally verified. The predictions utilized for HLA class I included the top 200 candidates for each allele. Utilizing more generous prediction thresholds is likely to allow for identification of additional epitopes. The limited number of donors also did not allow for the evaluation of potential differences in terms of ethnic background, disease severity, age, and gender. Future investigations will include validation of the epitope pools as potential diagnostic tools, establish a robust, user-friendly T cell assay, and investigate differences in T cell reactivity as a function of ethnicity, disease severity, age, and gender.

STAR★METHODS

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

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.xcrm.2021.100204.

ACKNOWLEDGMENTS

This study has been funded by the NIH NIAID (award AI42742 to S.C. and A.S., contract no. 75N9301900065 to A.S. and D.W., contract no. 75N93019C00001 to A.S. and B.P., NIH grant U01 CA260541-01 to D.W., K08 award AI135078 to J.M.D., and AI036214 to D.M.S.). Additional support has been provided by UCSD T32s (AI007036and AI007384 to S.A.R. and S.I.R.) and the Jonathan and Mary Tu Foundation (D.M.S.). A.T. was supported by a PhD student fellowship through the Clinical and Experimental Immunology Course at the University of Genoa, Italy. We thank Gina Levi and the LJI clinical core for assistance in sample coordination and blood processing and Erica Ollmann Saphire, Michael Norris, and Sara Landeras-Bueno for useful discussions and input on 3D modeling.

AUTHOR CONTRIBUTIONS

Conceptualization, A.T., A.G., S.C., and A.S.; data curation and bioinformatic analysis, J.A.G. and B.P.; formal analysis, A.T., J.S., C.K.K., A.G., D.W., J.M.D., J.M., and E.D.W.; funding acquisition, S.C., A.S., D.W., S.I.R., S.A.R., and J.M.D.; investigation, A.T., E.D.W., C.K.K., N.M., J.M.D., J.M., J.S., E.M., P.R., D.W., A.S., and A.G.; project administration, A.F.; resources, S.I.R., S.A.R., S.M., E.P., D.M.S., S.C., and A.S.; supervision, B.P., J.S., S.C., D.W., R.d.S.A., A.S., and A.G.; writing, A.T., S.C., A.S., and A.G.

DECLARATION OF INTERESTS

A.S. is a consultant for Gritstone, Flow Pharma, Merck, Epitogenesis, Gilead, and Avalia. S.C. is a consultant for Avalia. All other authors declare no competing interests. LJI has filed for patent protection for various aspects of vaccine design and identification of specific epitopes.

Received: November 18, 2020
Revised: December 17, 2020
Accepted: January 20, 2021
Published: February 16, 2021

REFERENCES

1. 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.
2. Rydyznski Moderbacher, C., Ramirez, S.I., Dan, J.M., Grifoni, A., Hastie, K.M., Weiskopf, D., Belanger, S., Abbott, R.K., Kim, C., Choi, J., et al. (2020). Antigen-specific adaptive immunity to SARS-CoV-2 in acute COVID-19 and associations with age and disease severity. Cell *183*, 996–1012.e19.

3. 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.

4. 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.

5. Altmann, D.M., and Boyton, R.J. (2020). SARS-CoV-2 T cell immunity: specificity, function, durability, and role in protection. Sci. Immunol. *5*, eabd6160.

6. Weiskopf, D., Schmitz, K.S., Raadsen, M.P., Grifoni, A., Okba, N.M.A., Endeman, H., van den Akker, J.P.C., Molenkamp, R., Koopmans, M.P.G., van Gorp, E.C.M., et al. (2020). Phenotype and kinetics of SARS-CoV-2-specific T cells in COVID-19 patients with acute respiratory distress syndrome. Sci. Immunol. *5*, eabd2071.

7. 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.

8. Meckiff, B.J., Ramírez-Suástegui, C., Fajardo, V., Chee, S.J., Kusnadi, A., Simon, H., Eschweiler, S., Grifoni, A., Pelosi, E., Weiskopf, D., et al. (2020). Imbalance of regulatory and cytotoxic SARS-CoV-2-reactive CD4$^+$ T cells in COVID-19. Cell *183*, 1340–1353.e16.

9. Schub, D., Klemis, V., Schneitler, S., Mihm, J., Lepper, P.M., Wilkens, H., Bals, R., Eichler, H., Gärtner, B.C., Becker, S.L., et al. (2020). High levels of SARS-CoV-2-specific T cells with restricted functionality in severe courses of COVID-19. JCI Insight *5*, 142167.

10. 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.

11. 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.

12. 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.T., 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.

13. Keller, M.D., Harris, K.M., Jensen-Wachspress, M.A., Kankate, V., Lang, H., Lazarski, C.A., Durkee-Shock, J.R., Lee, P.-H., Chaudhry, K., Webber, K., et al. (2020). SARS-CoV-2 specific T-cells are rapidly expanded for therapeutic use and target conserved regions of membrane protein. Blood, Published online October 26, 2020. https://doi.org/10.1182/blood.2020008488.

14. Nelde, A., Bilich, T., Heitmann, J.S., Maringer, Y., Salih, H.R., Roerden, M., Lübke, M., Bauer, J., Rieth, J., Wacker, M., et al. (2020). SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition. Nat. Immunol. *22*, 74–85.

15. 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.

16. Stadlbauer, D., Amanat, F., Chromikova, V., Jiang, K., Strohmeier, S., Arunkumar, G.A., Tan, J., Bhavsar, D., Capuano, C., Kirkpatrick, E., et al. (2020). SARS-CoV-2 seroconversion in humans: a detailed protocol for a serological assay, antigen production, and test setup. Curr. Protoc. Microbiol. *57*, e100.

17. 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.

18. Dhanda, S.K., Mahajan, S., Paul, S., Yan, Z., Kim, H., Jespersen, M.C., Jurtz, V., Andreatta, M., Greenbaum, J.A., Marcatili, P., et al. (2019). IEDB-AR: immune epitope database-analysis resource in 2019. Nucleic Acids Res. *47* (*W1*), W502–W506.

19. Bui, H.H., Sidney, J., Dinh, K., Southwood, S., Newman, M.J., and Sette, A. (2006). Predicting population coverage of T-cell epitope-based diagnostics and vaccines. BMC Bioinformatics *7*, 153.

20. Paul, S., Weiskopf, D., Angelo, M.A., Sidney, J., Peters, B., and Sette, A. (2013). HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity. J. Immunol. *191*, 5831–5839.

21. Greenbaum, J., Sidney, J., Chung, J., Brander, C., Peters, B., and Sette, A. (2011). Functional classification of class II human leukocyte antigen (HLA) molecules reveals seven different supertypes and a surprising degree of repertoire sharing across supertypes. Immunogenetics *63*, 325–335.

22. Xie, X., Muruato, A., Lokugamage, K.G., Narayanan, K., Zhang, X., Zou, J., Liu, J., Schindewolf, C., Bopp, N.E., Aguilar, P.V., et al. (2020). An infectious cDNA clone of SARS-CoV-2. Cell Host Microbe *27*, 841–848.e3.

23. Peters, B., Nielsen, M., and Sette, A. (2020). T cell epitope predictions. Annu. Rev. Immunol. *38*, 123–145.

24. da Silva Antunes, R., Quiambao, L.G., Sutherland, A., Soldevila, F., Dhanda, S.K., Armstrong, S.K., Brickman, T.J., Merkel, T., Peters, B., and Sette, A. (2020). Development and validation of a *Bordetella pertussis* whole-genome screening strategy. J. Immunol. Res. *2020*, 8202067.

25. Reiss, S., Baxter, A.E., Cirelli, K.M., Dan, J.M., Morou, A., Daigneault, A., Brassard, N., Silvestri, G., Routy, J.P., Havenar-Daughton, C., et al. (2017). Comparative analysis of activation induced marker (AIM) assays for sensitive identification of antigen-specific CD4 T cells. PLoS ONE *12*, e0186998.

26. Voic, H., de Vries, R.D., Sidney, J., Rubiro, P., Moore, E., Phillips, E., et al. (2020). Identification and Characterization of CD4 $^+$ T Cell Epitopes after Shingrix Vaccination. J Virol. *94*, e01641-20.

27. Oseroff, C., Sidney, J., Kotturi, M.F., Kolla, R., Alam, R., Broide, D.H., Wasserman, S.I., Weiskopf, D., McKinney, D.M., Chung, J.L., et al. (2010). Molecular determinants of T cell epitope recognition to the common Timothy grass allergen. J. Immunol. *185*, 943–955.

28. Lindestam Arlehamn, C.S., Gerasimova, A., Mele, F., Henderson, R., Swann, J., Greenbaum, J.A., Kim, Y., Sidney, J., James, E.A., Taplitz, R., et al. (2013). Memory T cells in latent Mycobacterium tuberculosis infection are directed against three antigenic islands and largely contained in a CXCR3+CCR6+ Th1 subset. PLoS Pathog. *9*, e1003130.

29. Sidney, J., Southwood, S., Moore, C., Oseroff, C., Pinilla, C., Grey, H.M., and Sette, A. (2013). Measurement of MHC/peptide interactions by gel filtration or monoclonal antibody capture. Curr. Protoc. Immunol. *Chapter 18*, Unit 18.3.

30. Southwood, S., Sidney, J., Kondo, A., del Guercio, M.F., Appella, E., Hoffman, S., Kubo, R.T., Chesnut, R.W., Grey, H.M., and Sette, A. (1998). Several common HLA-DR types share largely overlapping peptide binding repertoires. J. Immunol. *160*, 3363–3373.

31. Paul, S., Grifoni, A., Peters, B., and Sette, A. (2020). Major histocompatibility complex binding, eluted ligands, and immunogenicity: benchmark testing and predictions. Front. Immunol. *10*, 3151.

32. Weiskopf, D., Angelo, M.A., de Azeredo, E.L., Sidney, J., Greenbaum, J.A., Fernando, A.N., Broadwater, A., Kolla, R.V., De Silva, A.D., de Silva, A.M., et al. (2013). Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. Proc. Natl. Acad. Sci. USA *110*, E2046–E2053.

33. 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.

34. Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S.M., Walsh, R.M., Jr., Rawson, S., Rits-Volloch, S., and Chen, B. (2020). Distinct conformational states of SARS-CoV-2 spike protein. Science *369*, 1586–1592.

35. Paul, S., Karosiene, E., Dhanda, S.K., Jurtz, V., Edwards, L., Nielsen, M., Sette, A., and Peters, B. (2018). Determination of a predictive cleavage motif for eluted major histocompatibility complex class II ligands. Front. Immunol. *9*, 1795.

36. Shrock, E., Fujimura, E., Kula, T., Timms, R.T., Lee, I.H., Leng, Y., Robinson, M.L., Sie, B.M., Li, M.Z., Chen, Y., et al.; MGH COVID-19 Collection & Processing Team (2020). Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. Science *370*, eabd4250.

37. Heo, L., and Feig, M. (2020). Modeling of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Proteins by Machine Learning and Physics-Based Refinement. bioRxiv. https://doi.org/10.1101/2020.03.25.008904.

38. Weiskopf, D., Cerpas, C., Angelo, M.A., Bangs, D.J., Sidney, J., Paul, S., Peters, B., Sanches, F.P., Silvera, C.G., Costa, P.R., et al. (2015). Human CD8+ T-cell responses against the 4 dengue virus serotypes are associated with distinct patterns of protein targets. J. Infect. Dis. *212*, 1743–1751.

39. Grifoni, A., Angelo, M.A., Lopez, B., O'Rourke, P.H., Sidney, J., Cerpas, C., Balmaseda, A., Silveira, C.G.T., Maestri, A., Costa, P.R., et al. (2017). Global assessment of dengue virus-specific $CD4^+$ T cell responses in dengue-endemic areas. Front. Immunol. *8*, 1309.

40. Sette, A., Moutaftsi, M., Moyron-Quiroz, J., McCausland, M.M., Davies, D.H., Johnston, R.J., Peters, B., Rafii-El-Idrissi Benhnia, M., Hoffmann, J., Su, H.P., et al. (2008). Selective CD4+ T cell help for antibody responses to a large viral pathogen: deterministic linkage of specificities. Immunity *28*, 847–858.

41. Lindestam Arlehamn, C.S., McKinney, D.M., Carpenter, C., Paul, S., Rozot, V., Makgotlho, E., Gregg, Y., van Rooyen, M., Ernst, J.D., Hatherill, M., et al. (2016). A quantitative analysis of complexity of human pathogen-specific CD4 T cell responses in healthy M. tuberculosis infected South Africans. PLoS Pathog. *12*, e1005760.

42. Barra, C., Alvarez, B., Paul, S., Sette, A., Peters, B., Andreatta, M., Buus, S., and Nielsen, M. (2018). Footprints of antigen processing boost MHC class II natural ligand predictions. Genome Med. *10*, 84.

43. Cassotta, A., Paparoditis, P., Geiger, R., Mettu, R.R., Landry, S.J., Donati, A., Benevento, M., Foglierini, M., Lewis, D.J.M., Lanzavecchia, A., et al. (2020). Deciphering and predicting CD4+ T cell immunodominance of influenza virus hemagglutinin. J. Exp. Med. *217*, e20200206.

44. Kim, Y., Yewdell, J.W., Sette, A., and Peters, B. (2013). Positional bias of MHC class I restricted T-cell epitopes in viral antigens is likely due to a bias in conservation. PLoS Comput. Biol. *9*, e1002884.

45. Vita, R., Mahajan, S., Overton, J.A., Dhanda, S.K., Martini, S., Cantrell, J.R., et al. (2019). The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. *47*, D339–D343.

46. Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., and Nielsen, M. (2017). NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. J. Immunol. *199*, 3360–3368.

47. Paul, S., Lindestam Arlehamn, C.S., Scriba, T.J., Dillon, M.B.C., Oseroff, C., Hinz, D., et al. (2015). Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. J Immunol Methods. *422*, 28–34.

48. Paul, S., Sidney, J., Sette, A., and Peters, B. (2016). TepiTool: A Pipeline for Computational Prediction of T Cell Epitope Candidates. Curr Protoc Immunol. *114*, 18.19.1–18.19.24.

49. Carrasco Pro, S., Sidney, J., Paul, S., Lindestam Arlehamn, C., Weiskopf, D., Peters, B., and Sette, A. (2015). Automatic generation of validated specific epitope sets. J. Immunol. Res. *2015*, 763461.

50. Dan, J.M., Lindestam Arlehamn, C.S., Weiskopf, D., da Silva Antunes, R., Havenar-Daughton, C., Reiss, S.M., Brigger, M., Bothwell, M., Sette, A., and Crotty, S. (2016). A cytokine-independent approach to identify antigen-specific human germinal center T follicular helper cells and rare antigen-specific CD4+ T cells in blood. J. Immunol. *197*, 983–993.

51. Dhanda, S.K., Vita, R., Ha, B., Grifoni, A., Peters, B., and Sette, A. (2018). ImmunomeBrowser: a tool to aggregate and visualize complex and heterogeneous epitopes in reference proteins. Bioinformatics *34*, 3931–3933.

52. Karosiene, E., Rasmussen, M., Blicher, T., Lund, O., Buus, S., and Nielsen, M. (2013). NetMHCIIpan-3.0, a common pan-specific MHC class II prediction method including all three human MHC class II isotypes, HLA-DR, HLA-DP and HLA-DQ. Immunogenetics *65*, 711–724.

53. Zeng, W., Liu, G., Ma, H., Zhao, D., Yang, Y., Liu, M., Mohammed, A., Zhao, C., Yang, Y., Xie, J., et al. (2020). Biochemical characterization of SARS-CoV-2 nucleocapsid protein. Biochem. Biophys. Res. Commun. *527*, 618–623.

54. Kelley, L.A., and Sternberg, M.J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. *4*, 363–371.

55. Land, H., and Humble, M.S. (2018). YASARA: a tool to obtain structural guidance in biocatalytic investigations. Methods Mol. Biol. *1685*, 43–67.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
M5E2 (V500) [anti-CD14]	Becton Dickinson	561391 (RRID:AB_10611856)
HIB19 (V500) [anti-CD19]	Becton Dickinson	561121 (RRID:AB_10562391)
RPA-T4 (BV605) [anti-CD4]	Becton Dickinson	562989 (RRID:AB_2737935)
RPA-T8 (BV650) [anti-CD8]	BioLegend	301042 (RRID:AB_2563505)
FN50 (PE) [anti-CD69]	Becton Dickinson	555531 (RRID:AB_2737680)
Ber-ACT35 (PE-Cy7) [anti-OX40]	Biolegend	350012 (RRID:AB_10901161)
4B4-1 (APC) [anti-CD137]	BioLegend	309810 (RRID:AB_830672)
OKT3 (AF700) [anti-CD3]	Biolegend	317340 (RRID:AB_2563408)
Biological samples		
Convalescent donor blood samples	UC San Diego Health	https://health.ucsd.edu/
Convalescent donor blood samples	Sanguine Biosciences	https://www.sanguinebio.com
Convalescent donor blood samples	StemExpress	https://www.stemexpress.com
Convalescent donor blood samples	BioIVT	https://bioivt.com/
Chemicals, peptides, and recombinant proteins		
Synthetic peptides	Synthetic Biomolecules (aka A&A)	http://www.syntheticbiomolecules.com
SARS-CoV-2 Receptor Binding Domain (RBD) protein	Stadlbauer et al.[16]	N/A
Deposited data		
SARS-CoV-2 spike glycoprotein 3D-structure	Cai et al.[34]	PDB: 6XR8
Wuhan-Hu-1 RNA isolate	NCBI nuccore database	GenBank:MN908947
ORF10 protein	NCBI protein database	NCBI: YP_009725255.1
Nucleocapsid phosphoprotein	NCBI protein database	NCBI: YP_009724397.2
ORF8 protein	NCBI protein database	NCBI: YP_009724396.1
ORF7a protein	NCBI protein database	NCBI: YP_009724395.1
ORF6 protein	NCBI protein database	NCBI: YP_009724394.1
membrane glycoprotein	NCBI protein database	NCBI: YP_009724393.1
envelope protein	NCBI protein database	NCBI: YP_009724392.1
ORF3a protein	NCBI protein database	NCBI: YP_009724391.1
surface glycoprotein	NCBI protein database	NCBI: YP_009724390.1
orf1ab polyprotein	NCBI protein database	NCBI: YP_009724389.1
Software and algorithms		
IEDB	Vita et al.[45]	https://www.iedb.org
IEDB-AR (analysis resource)	Dhanda et al.[18]	http://tools.iedb.org/main/
NetMHCpan EL 4.0	Jurtz et al.[46]	http://tools.iedb.org/mhci/
Tepitool	Paul et al.;[47] Paul et al.[48]	http://tools.iedb.org/tepitool/
MHCII NP algorithm	Paul et al.[35]	http://tools.iedb.org/mhciinp/
FlowJo 10	FlowJo, LLC	https://www.flowjo.com
GraphPad Prism 8.4	GraphPad	https://www.graphpad.com:443/
YASARA	YASARA	http://www.yasara.org/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Dr. Alessandro Sette (alex@lji.org).

Materials availability

Epitope pools used in this study will be made available to the scientific community upon request, and following execution of a material transfer agreement, by contacting Dr. Alessandro Sette (alex@lji.org).

Data and code availability

The published article includes all data generated or analyzed during this study, and summarized in the accompanying tables, figures and supplemental materials.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human Subjects

Convalescent COVID-19 Donors utilized for epitope identification

Blood donations from the 99 convalescent donors included in this study's cohort were collected through either the UC San Diego Health Clinic under IRB approved protocols (200236X), or under IRB approval (VD-214) at the La Jolla Institute. Donations obtained through the CROs Sanguine, BioIVT and Stem Express were collected under the same IRB approval (VD-214) at the La Jolla Institute. Details of this cohort can be found in Table S1. All donors were over the age of 18 years and no exclusions were made due to disease severity, race, ethnicity, or gender. All donors were able to provide informed consent, or had a legal guardian or representative able to do so. Study exclusion criteria included lack of willingness or ability to provide informed consent, or lack of an appropriate legal guardian to provide informed consent.

Disease severity was defined as mild, moderate, severe or critical as previously described (Grifoni 2020).[1] In brief, this classification of disease severity is based on a modified version of the WHO interim guidance, ''Clinical management of severe acute respiratory infection when COVID-19 is suspected'' (WHO Reference Number: WHO/2019-nCoV/clinical/2020.4). At the time of enrollment in the study, 80% of donors had been confirmed positive by swab test viral PCR during the acute phase of infection. Plasma samples from all donors were later tested by IgG ELISA for SARS-CoV-2 S protein RBD to verify previous infection (Table S1; Figure 2A).

Healthy Unexposed donors utilized for CD4-E and CD8-E megapool validation

Samples from healthy adult donors were obtained from the San Diego Blood Bank (SDBB). According to the criteria set up by the SDBB if a subject was eligible to donate blood, they were considered eligible for our study. All the donors were tested for SARS-CoV-2 RBD IgG serology and were found negative and therefore considered unexposed. An overview of the characteristics of these donors is provided in Table S1.

Convalescent COVID-19 donors utilized for CD4-E and CD8-E megapool validation

The 31 convalescent donors tested in the megapool AIM and FluoroSpot assays (Figure 7) were collected from the same clinics using the same protocols as described above for the donors utilized for epitope identification. Similarly, no donors enrolled were under the age of 18 and none were excluded due to disease severity, race, ethnicity, or gender. All donors, or legal guardians, gave informed consent. Specific characteristics of these donors can be found in Table S3, including the summary of ELISA testing for SARS-CoV-2 S protein RBD.

METHOD DETAILS

Peptide Pools

Preparation of 15-mers and subsequent megapools and mesopools

To identify SARS-CoV-2-specific T cell epitopes, 15-mer peptides overlapping by 10 amino acids and spanning entire SARS-CoV-2 proteins were synthesized. All peptides were synthesized as crude material (A&A, San Diego, CA) and individually resuspended in dimethyl sulfoxide (DMSO) at a concentration of 10 mg/mL. Aliquots of these peptides were pooled by antigen of provenance into megapools (MP) (as described in Table S3 and sequentially lyophilized as previously reported[49]. Another portion of the 15-mer peptides were pooled into smaller mesopools of ten peptides each. All pools were resuspended at 1 mg/mL in DMSO.

Class I peptide preparation

Class I predicted peptides were designed using the protein sequences derived from the SARS-CoV-2 reference strain (GenBank: MN908947). Predictions were performed as previously reported using NetMHC pan EL 4.0 algorithm[46] for 28 HLA A and B alleles that were selected based on frequency in our cohort and also representative of the worldwide population (Figures S1A and S1B). The top 200 predicted peptides were selected for each allele. In total 5,600 class I peptides were synthesized and resuspended in DMSO at 10 mg/mL.

PBMC isolation and HLA typing

Whole blood was collected from all donors in either Acid Citrate Dextrose (ACD) tubes or heparin coated blood bags. Whole blood was then centrifuged at room temperature for 15 minutes at 1850 rpm to separate the cellular fraction and plasma. The plasma was then carefully removed from the cell pellet and stored at −20C. Peripheral blood mononuclear cells (PBMC) were isolated by density-gradient sedimentation using Ficoll-Paque (Lymphoprep, Nycomed Pharma) as previously described[32]. Isolated PBMC were cryopreserved in cell recovery media containing 10% DMSO (GIBCO), supplemented with 90% heat-inactivated fetal bovine serum, depending on the processing laboratory, (FBS; Hyclone Laboratories, Logan UT) and stored in liquid nitrogen until used in the assays.

Each sample was HLA typed by Murdoch University in Western Australia, an ASHI-Accredited laboratory.[26] Typing was performed for the class I HLA A and B loci and class II DRBI, DQB1, and DPB1 loci.

SARS-CoV-2 RBD ELISA

The SARS-CoV-2 RBD ELISA has been described in detail elsewhere.[1] All convalescent COVID-19 donors had their serology determined by ELISA. Briefly, 96-well half-area plates (ThermoFisher 3690) were coated with 1 ug/mL SARS-CoV-2 Spike (S) Receptor Binding Domain (RBD) and incubated at 4°C overnight. On the following day plates were blocked at room temperature for 2 hours with 3% milk in phosphate buffered saline (PBS) containing 0.05% Tween-20. Then, heat-inactivated plasma was added to the plates for another 90-minute incubation at room temperature followed by incubation with conjugated secondary antibody, detection, and subsequent data analysis by reading the plates on Spectramax Plate Reader at 450 nm using SoftMax Pro. Limit of detection (LOD) was defined as 1:3. Limit of sensitivity (LOS) for SARS-CoV-2 infected individuals was established based on uninfected subjects, using plasma from normal healthy donors not exposed to SARS-CoV-2.

Flow Cytometry

Activation induced cell marker (AIM) assay

The AIM assay was performed as previously described[25,50]. Cryopreserved PBMCs were thawed by diluting the cells in 10 mL complete RPMI 1640 with 5% human AB serum (Gemini Bioproducts) in the presence of benzonase [20 μl/10ml]. Cells were cultured for 20 to 24 hours in the presence of SARS-CoV-2 specific MPs [1 μg/ml], mesopools [1 μg/ml], 15-mers [10 μg/ml], or class I predicted peptides [10 μg/ml] in 96-wells U bottom plates with $1x10^6$ PBMC per well. As a negative control, an equimolar amount of DMSO was used to stimulate the cells as a negative control in triplicate wells, and phytohemagglutinin (PHA, Roche, 1 μg/ml) was included as the positive control. The cells were stained with CD3 AF700 (4:100; Life Technologies Cat# 56-0038-42), CD4 BV605 (4:100; BD Biosciences Cat# 562658), CD8 BV650 (2:100; Biolegend Cat# 301042), and Live/Dead Aqua (1:1000; eBioscience Cat# 65-0866-14). Activation was measured by the following markers: CD137 APC (4:100; Biolegend Cat# 309810), OX40 PE-Cy7 (2:100; Biolegend Cat#350012), and CD69 PE (10:100; BD Biosciences Cat# 555531). All samples were acquired on either a ZE5 cell analyzer (Bio-rad laboratories) or an Aurora flow cytometry system (Cytek), and analyzed with FlowJo software (Tree Star).

HLA binding assays

The binding of selected SARS-CoV-2 15-mer epitopes to HLA class II MHC molecules was measured as previously described (Sidney 2013, Voic 2020).[29] In brief, the binding is quantified by each peptide's capacity to inhibit the binding of a radiolabeled peptide probe to purified MHC in classical competition assays. The probe was incubated with purified MHC, a mixture of protease inhibitors, and different concentrations of unlabeled inhibitor peptide at room temperature or 37°C for 2 days. MHC molecules were subsequently captured on HLA-DR-specific monoclonal antibody (L243) coated Lumitrac 600 plates (Greiner Bio-one, Frickenhausen, Germany) and radioactivity was measured using the TopCount microscintillation counter (Packard Instrument Co., Meriden, CT). Each peptide was tested at 6 concentrations to cover a 100,000-fold dose range, and an unlabeled version of the radiolabeled probe was included in each experiment as a positive control for inhibition. To analyze the results, we calculated the concentration of peptide at which the binding was inhibited by 50% (IC_{50} nM). For these values to approximate true Kd values, the following conditions were met: 1) the concentration of radiolabelled probe is less than the concentration of MHC, and 2) the measured IC_{50} is greater than or equal to the concentration of MHC.

FluoroSpot

PBMCs derived from 25 unexposed donors were stimulated in triplicate at a single density of $200x10^3$ cells/well (one donor was tested at $50x10^3$ due to limitation in cell numbers). PBMCs from a cohort of 31 convalescent COVID-19 donors were stimulated in triplicates of $200x10^3$ cells/well, with the exception of 5 donors tested at $50\text{-}100x10^3$ cells/well due to cell limitations (Figures 7B, 7D, 7F, and 7H). Seventeen of these convalescent donors were further titrated at 200, 50, 25, and $12.5x10^3$ cells/well (Figures 7I–7L). The cells were stimulated with the different MPs analyzed (1μg/mL), PHA (10μg/mL), and DMSO (0.1%) in 96-well plates previously coated with anti-cytokine antibodies for IFNγ, (mAbs 1-D1K; Mabtech, Stockholm, Sweden) at a concentration of 10μg/mL. After 20 hours of incubation at 37°C, 5% CO_2, cells were discarded and FluoroSpot plates were washed and further incubated for 2 hours with cytokine antibodies (mAbs 7-B6-1-BAM; Mabtech, Stockholm, Sweden). Subsequently, plates were washed again with PBS/0.05% Tween20 and incubated for 1 hour with fluorophore-conjugated antibodies (Anti-BAM-490). Computer-assisted image analysis was performed by counting fluorescent spots using an AID iSPOT FluoroSpot reader (AIS-diagnostika, Germany). Each megapool was considered positive compared to the background based on the following criteria: 20 or more spot forming cells (SFC) per 10^6 PBMC after background subtraction for each cytokine analyzed, a stimulation index (S.I.) greater than 2, and statistically different from the background ($p < 0.05$) in either a Poisson or t test.

QUANTIFICATION AND STATISTICAL ANALYSIS

FlowJo 10 and GraphPad Prism 8.4 were used to perform data and statistical analyses, unless otherwise stated. Statistical details of the experiments are provided in the respective figure legends. Data plotted in linear scale are expressed as mean + standard deviation (SD). Data plotted in logarithmic scales are expressed as median + 95% confidence interval (CI) or geometric mean + geometric SD. Statistical analyses were performed using Spearman correlation and Mann-Whitney or Kolmogorov-Smirnov tests for unpaired comparisons. Details pertaining to significance are also noted in the respective figure legends.

AIM assay analysis

In analyzing data from the AIM assays, the counts of AIM^+ $CD4^+$ and $CD8^+$ T cells were normalized based on the counts of $CD4^+$ and $CD8^+$ T cells in each well to be equivalent to $1x10^6$ total $CD8^+$ or $CD4^+$ T cells. The background was removed from the data by subtracting the single or the average of the counts of AIM^+ cells plated as single or triplicate wells stimulated with DMSO. We included the triplicate wells stimulated with DMSO in the mesopools and epitope identification steps to take into account the variability of the weaker signals observed in those two respect to the original MP reactivity[24]. The Stimulation Index (SI) was calculated by dividing the count of AIM^+ cells after SARS-CoV-2 stimulation with the ones in the negative control. A positive response had an SI greater than 2 and a minimum of 100 AIM^+ cells after background subtraction. The gates for AIM^+ cells were drawn relative to the negative and positive controls for each donor. A representative example of the gating strategy is depicted in Figure S3B.

HLA class I nested epitopes

For some alleles and proteins, multiple nested class I predicted peptides were tested in the AIM assay. In cases where a specific donor responded to multiple nested epitopes corresponding to the same allele and protein, the epitope with the highest magnitude of response was classified as the optimal epitope. If multiple nested epitopes had the same response (within a range of 50 AIM^+ cells), the epitope with the shortest length was selected. Nested epitopes corresponding to different donors or different alleles were conserved as separate epitopes.

CCC homology analysis

SARS-CoV-2-derived 15-mer peptides were analyzed for their identity with the common cold coronaviruses (CCC) 229E, NL63, HKU1, and OC43, as previously described[15]. In brief, every SARS-CoV-2 15-mer peptide tested for immunogenicity was compared against every position in the corresponding protein sequences of common coronaviruses obtained from GenBank. The region that best matched the respective SARS-CoV-2 peptide was used to calculate percent sequence identity for each of the four CCC viruses individually, as well as the maximum across all four (Figure S5A). The same methodology was also used to calculate sequence identity for SARS-CoV-2 class I peptides (Figure S5B). Using the same set of common coronavirus reference sequences, an alternative analysis was performed by mapping each SARS-CoV-2 peptide with the S, M and N protein sequences corresponding to the four common coronavirus using Immunobrowser tool[51]. The values resulted from this specific analysis are plotted in Figure 5.

T cell epitope restriction predictions

Putative HLA class II restrictions for individual 15-mer $CD4^+$ T cell epitopes were inferred using the IEDB's TepiTool resource (Paul 2016). All $CD4^+$ T cell prediction analyses were performed applying the NetMHCIIpan algorithm[52]. Prediction analyses were performed to either infer HLA restriction based on the HLA typing of the cohort (Tables S2 and S5) or to assess potential binding promiscuity of experimentally defined epitopes, considering the 27 most frequent class II alleles in the worldwide population[21]. In both types of prediction analyses, a 20th percentile threshold was applied (Table S3), as previously described[15].

Assigning regions within the linear structure

Simple diagrams were created to describe the linear structures of S, N, and M proteins (Figure 4). The different regions of the S protein were defined based on the works of Cai et al., 2020.[34] The structure of the N protein was divided into 3 main regions, the N- and C-terminal domains, and the linker region in between[53]. For the M protein, the regions of the structure were extracted from UniProt (UniProtKB - P59596 (VME1_SARS).

3D-rendering and model design

Three different approaches have been used to map T and B cell immunodominant regions on the 3D-structures for SARS-CoV-2 S, M and N proteins. The S protein model was based on the crystal structure described in Cai et al., 2020[34] (PDB: 6XR8) and using the glycosylation sites annotated in the submitted PDB. The M protein model has been previously described by Heo et al., 2020. The model for the N protein was run on four different homology prediction servers (SWISS-MODEL, RaptorX, iTasser and Phyre2). In order to have a complete N sequence, Phyre2 server was subsequently selected using the intensive mode[54]. The resulting model showed a variable level of confidence with higher percentages (> 90%) in the C-Terminal domain (CTD) and N-terminal domain (NTD) regions and low confidence percentages (> 10%) in the linker domain. The N model was superimposable with both the crystal structures for the CTD (PDB: 6WZO) and NTD (PDB: 6M3M). The current N model has the only purpose of visualization for mapping immunodominant regions. All the mapping analyses have been performed using the free version of YASARA[55].