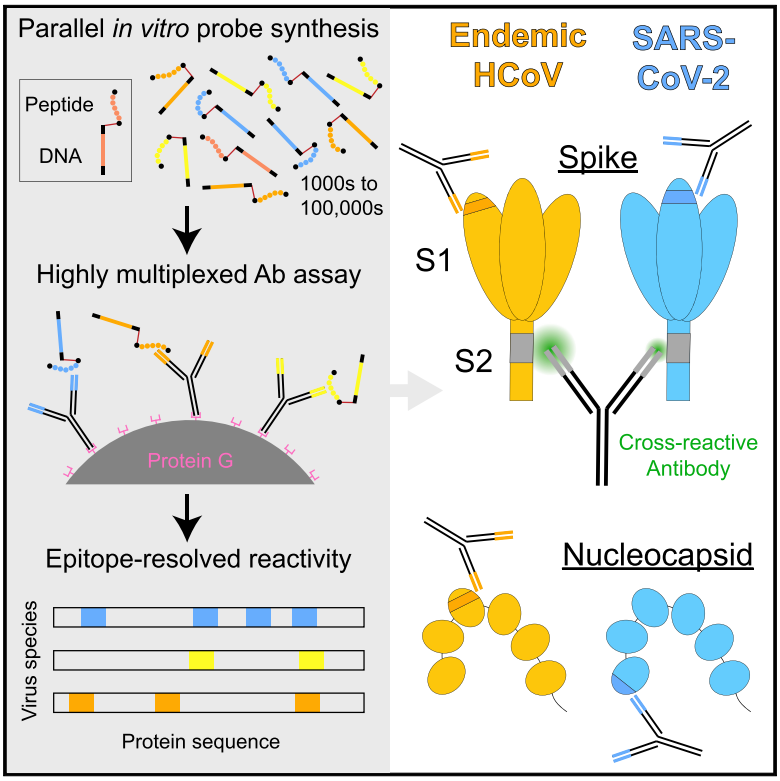
Cell Reports
Medicine

Article

Epitope-resolved profiling of the SARS-CoV-2 antibody response identifies cross-reactivity with endemic human coronaviruses

Graphical Abstract



Highlights

- PepSeq enables fully *in vitro*, highly multiplexed peptide-based antibody assays
- Epitope mapping shows preexisting antibody reactivity to SARS-CoV-2 antigens
- Antibodies cross-recognize endemic and pandemic antigens in the Spike S2 subunit
- Cross-reactive antibodies raised by SARS-CoV-2 preferentially bind endemic homologs

Authors

Jason T. Ladner, Sierra N. Henson, Annalee S. Boyle, ..., Mark S. Chee, Sergey A. Shiryaev, John A. Altin

Correspondence

jason.ladner@nau.edu (J.T.L.), jaltin@tgen.org (J.A.A.)

In Brief

Ladner et al. use a fully *in vitro*, highly multiplexed approach to finely map antibody epitopes across the Spike and Nucleocapsid proteins of all human-infecting coronaviruses. This demonstrates that SARS-CoV-2 elicits antibodies that cross-recognize endemic coronavirus antigens at conserved regions of Spike and that these antibodies preferentially bind endemic homologs.

Ladner et al., 2021, Cell Reports Medicine 2, 100189
January 19, 2021 © 2020 The Author(s).
https://doi.org/10.1016/j.xcrm.2020.100189

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Article

Epitope-resolved profiling of the SARS-CoV-2 antibody response identifies cross-reactivity with endemic human coronaviruses

Jason T. Ladner,[1,*] Sierra N. Henson,[2] Annalee S. Boyle,[2] Anna L. Engelbrektson,[2] Zane W. Fink,[1] Fatima Rahee,[2] Jonathan D'ambrozio,[3] Kurt E. Schaecher,[3] Mars Stone,[4] Wenjuan Dong,[5] Sanjeet Dadwal,[6] Jianhua Yu,[5] Michael A. Caligiuri,[5] Piotr Cieplak,[7] Magnar Bjørås,[8] Mona H. Fenstad,[8,10] Svein A. Nordbø,[8,9] Denis E. Kainov,[8] Norihito Muranaka,[11] Mark S. Chee,[11] Sergey A. Shiryaev,[7] and John A. Altin[2,12,*]

[1]The Pathogen and Microbiome Institute, Northern Arizona University, Flagstaff, AZ, USA
[2]The Translational Genomics Research Institute (TGen), Phoenix and Flagstaff, AZ, USA
[3]Walter Reed National Military Medical Center, Bethesda, MD, USA
[4]Vitalant Research Institute and Department of Laboratory Medicine, University of California, San Francisco, San Francisco, CA, USA
[5]Hematologic Malignancies and Stem Cell Transplantation Institute, City of Hope National Medical Center, Duarte, CA, USA
[6]Division of Infectious Diseases, City of Hope National Medical Center, Duarte, CA, USA
[7]Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA
[8]Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway
[9]Department of Medical Microbiology, St. Olavs Hospital, Trondheim, Norway
[10]Department of Immunology and Transfusion Medicine, St. Olavs Hospital, Trondheim, Norway
[11]Encodia, San Diego, CA, USA
[12]Lead contact
*Correspondence: jason.ladner@nau.edu (J.T.L.), jaltin@tgen.org (J.A.A.)
https://doi.org/10.1016/j.xcrm.2020.100189

SUMMARY

The SARS-CoV-2 proteome shares regions of conservation with endemic human coronaviruses (CoVs), but it remains unknown to what extent these may be cross-recognized by the antibody response. Here, we study cross-reactivity using a highly multiplexed peptide assay (PepSeq) to generate an epitope-resolved view of IgG reactivity across all human CoVs in both COVID-19 convalescent and negative donors. PepSeq resolves epitopes across the SARS-CoV-2 Spike and Nucleocapsid proteins that are commonly targeted in convalescent donors, including several sites also recognized in some uninfected controls. By comparing patterns of homologous reactivity between CoVs and using targeted antibody-depletion experiments, we demonstrate that SARS-CoV-2 elicits antibodies that cross-recognize pandemic and endemic CoV antigens at two Spike S2 subunit epitopes. We further show that these cross-reactive antibodies preferentially bind endemic homologs. Our findings highlight sites at which the SARS-CoV-2 response appears to be shaped by previous CoV exposures and which have the potential to raise broadly neutralizing responses.

INTRODUCTION

Severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) is a single-stranded RNA virus in the Coronaviridae family that emerged in late 2019 and has caused morbidity, mortality, and economic disruption on a global scale with few precedents.[1] The Coronaviridae family includes four species/strains that are endemic in the human population—HCoV-229E, HCoV-NL63, HCoV-HKU1, and HCoV-OC43 (Betacoronavirus 1 species)—and are usually associated with mild, self-limiting upper respiratory tract infections, although they can cause severe illness in immunocompromised patients.[2] Two other species, Middle East respiratory syndrome-CoV (MERS-CoV) and SARS-CoV, have recently emerged and cause severe disease in humans. Like the other human-infecting CoVs (HCoVs),[3,4] SARS-CoV-2 infection can elicit a robust antibody response in humans,[5,6] and this response represents a major focus of widespread efforts to develop accurate diagnostics and strategies for passive and active immunization against infection.[7–9] Existing serological assays for SARS-CoV-2 antibody reactivity generally use full-length viral proteins or domains—Spike (S), Nucleocapsid (N), or the receptor-binding domain (RBD) of S—as antigenic baits, followed by enzyme-linked or fluorescent detection.[9] These assays provide a single measure of antibody reactivity, which represents a composite signal across many epitopes, and are able to detect viral exposure with a range of accuracies.[10,11] Neutralization assays using either native or pseudotyped viruses have also been developed.[12] It remains to be seen how these different assays will perform as diagnostics or correlates of the protection conferred by infection or vaccination.

Relative to protein-based analyses of the humoral response, epitope-level assays have the potential to add several layers of

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

information. First, although SARS-CoV-2 proteins are generally distinct from other HCoVs, some regions of strong conservation exist,[1,13] meaning that there is the potential for immune cross-reactivity that can only be resolved at the epitope level. It was recently demonstrated that a large fraction of non-exposed individuals have T cell reactivity to SARS-CoV-2 peptides, indicating cross-reactivity with existing responses, possibly those generated against homologous peptides from endemic HCoVs.[14] In the case of antibody responses, cross-reactivity has been described between the more closely related SARS-CoV and SARS-CoV-2.[15,16] Epitope-resolved analyses therefore have the potential to identify antigens that may discriminate related CoVs, leading to more specific diagnostic assays. High levels of sequence conservation may also indicate functional essentiality; therefore, by highlighting potentially cross-reactive epitopes in conserved regions of the proteome, epitope-level assays can identify antibodies and targets with therapeutic potential, against which viral escape may be more difficult.[17]

A second rationale for generating epitope-resolved views is that antibody recognition of different protein regions can have divergent functional consequences, including neutralization potential. For CoVs, antibodies binding the surface-exposed, receptor-binding S protein exhibit the greatest neutralizing potential,[18,19] but these antibodies can recognize a wide variety of epitopes within the protein, each with the potential for different functional consequences. This likely accounts for the imperfect correlation between the titers of S-binding antibodies and viral neutralization activity across individuals.[20] Due to its interaction with the host entry receptor (the angiotensin-converting enzyme 2 [ACE2]), the RBD of S represents the predominant target of vaccination and monoclonal antibody development strategies, and a growing number of antibodies against this domain have been described.[20–23] However, the RBD is one of the less conserved regions of the CoV proteome, and antibodies against epitopes outside the RBD have also been shown to have neutralizing activity[21,24]; these may act in various ways, including by preventing important protease cleavage events and/or conformational changes required for successful entry into cells. However, antibodies that recognize epitopes within the N protein, which coats the viral genome and is contained within mature viral particles, likely provide little or no neutralization potential, but may be useful signatures for differentiating vaccine responses from those resulting from natural virus infection, a strategy already used for other viruses.[25,26] In addition to the different neutralization potential, it is possible that unfavorable distributions of epitope reactivity can contribute to immunopathology—for example, through antibody-dependent enhancement,[27–29] although this phenomenon remains to be demonstrated for SARS-CoV-2.[30]

Peptide sub-sequences have been used for decades as probes to detect antibodies recognizing linear epitopes within the full-length proteins from which they are derived.[31,32] Although unable to detect antibodies whose binding depends on elements that are discontinuous in the primary sequence, this strategy has the advantage that it enables the parallel design, synthesis, and assay of thousands of epitope-level antigen baits. In its simplest format, peptides can be used individually—for example, in separate wells in an ELISA. A recent study used this approach to identify two linear epitopes in S protein that were targeted by neutralizing antibodies in SARS-CoV-2 convalescent donors.[24] More powerful assays involve sets of peptides that are assayed in multiplex using either spatial addressing, in the case of peptide arrays,[33] or DNA indexing, in the case of phage display libraries.[34] Using the latter approach, the highly multiplexed and epitope-resolved detection of antibodies to viruses has been demonstrated with high sensitivity and specificity.[35]

Here, we present a synthetic biology approach to highly multiplexed peptide-based serological assays (PepSeq) in which libraries of peptide baits, each covalently coupled to a DNA barcode, are synthesized from high-complexity DNA pools using a simple and fully *in vitro* approach. Library synthesis takes advantage of *in vitro* transcription and translation, including an intramolecular coupling mediated by puromycin,[36,37] and the DNA-barcoded peptides can then be used to probe antibodies using a high-throughput sequencing readout. We use this platform to synthesize libraries of overlapping 30-mer peptides covering all HCoV proteomes and assay these against sera from prepandemic and SARS-CoV-2 convalescent donors. Our results demonstrate the accurate detection of SARS-CoV-2 exposure and reveal multiple recurrent antibody epitopes, including two Spike epitopes at which antibody responses cross-react between SARS-CoV-2 and one or more endemic HCoVs. We further demonstrate that these cross-reactive antibodies preferentially bind to endemic HCoV peptides, suggesting that the response to SARS-CoV-2 at these regions is shaped by previous CoV exposure.

RESULTS

A highly multiplexed peptide assay to evaluate CoV antibody responses

To generate a broad and high-resolution view of the antibody response to HCoVs, including SARS-CoV-2, we designed and synthesized 2 separate DNA-barcoded 30-mer peptide libraries (PepSeq) using the method described previously[36] (Figure 1A). Each library began as a pool of DNA oligonucleotide templates, which was modified using bulk enzymatic steps consisting of transcription, ligation of a puromycin-containing adaptor oligo, translation, and reverse transcription. One library was focused on SARS-CoV-2 (SCV2) and contained 2,107 peptides representing the Spike and Nucleocapsid—the 2 most immunogenic CoV proteins—at high redundancy, with an average of 38 peptides covering each amino acid position (Figure 1B). The other library (human virome [HV]) comprised 244,000 peptides designed to cover the full proteomes of all of the viruses known to infect humans, as of the end of 2018. Therefore, HV included peptides from the complete proteomes of 6/7 HCoVs: HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1, SARS-CoV, and MERS-CoV, but not SARS-CoV-2 (Figure 1C). The SCV2 library also included 373 positive control peptides that we have previously shown are commonly recognized across the human population (unpublished data). These controls represent a subset of the HV peptides and were designed from 55 different virus species.

CellPress
OPEN ACCESS

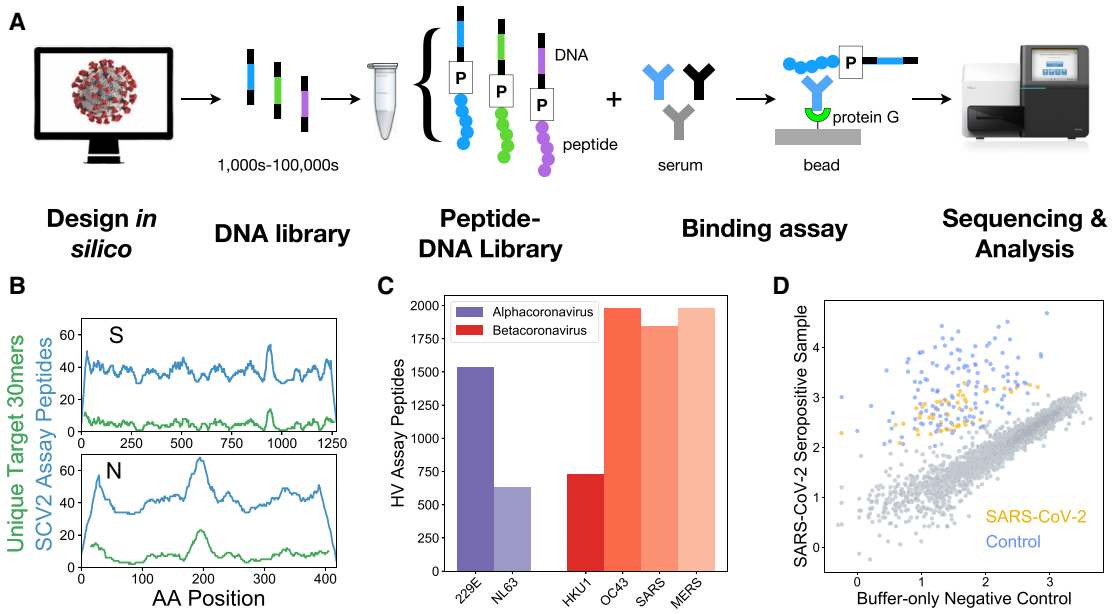


Figure 1. Epitope-resolved CoV serology using a highly multiplexed peptide-based assay (PepSeq)

(A) Platform for customizable highly multiplexed peptide-based serology, comprising the following steps: (1) *in silico* design, (2–3) generation of a library of DNA-barcoded peptides from oligonucleotide templates using bulk *in vitro* reactions (transcription, ligation of a puromycin [P]-containing adaptor, translation, reverse transcription), (4) serum binding assay and protein G capture, and (5) sequencing and analysis of the distribution of binders using their DNA barcodes.

(B) Peptide coverage depth across the SARS-CoV-2 Spike (S) and Nucleocapsid (N) proteins within the SCV2 peptide library. Peptide coverage depth (blue) correlates well with amino acid sequence diversity within the target SARS-CoV-2 sequences (green), calculated as the number of unique 30-mers.

(C) Number of peptides within the HV library that were designed from each of the 6 human CoVs (HCoVs) known before 2019.

(D) Example scatterplot illustrating SCV2 PepSeq assay results for a single serum sample. This plot shows normalized sequence read counts (log10 scale) for each peptide in the SCV2 library. Assay results using antibody-free negative controls are shown on the x axis (average of 8 replicates shown), while the results from a COVID-19 convalescent serum sample are shown on the y axis (average of 2 replicates shown). Gray circles represent unenriched peptides, with a strong correlation between the 2 assays, based on the starting abundance of the different peptides. Colored circles represent SARS-CoV-2 (orange) and non-SARS-CoV-2 control (blue) peptides that have been enriched through interaction with serum antibodies.

In total, we assayed and analyzed 55 coronavirus disease 2019 (COVID-19) convalescent and 69 SARS-CoV-2-negative (both pre- and postpandemic) serum/plasma samples using the SCV2 and/or HV PepSeq libraries (Tables 1 and S1); 96% of the convalescent samples (53/55) and 94% of the negative samples (65/69) were assayed separately with both libraries. For each assay, we incubated our PepSeq probes overnight with serum/plasma (or buffer as a negative control), captured the immunoglobulin G (IgG) on protein G beads, washed away the non-binding library members, eluted binders, and then performed PCR and high-throughput sequencing on the DNA tags to identify the distribution of bound peptides. Each sample was run in duplicate, and we observed strong signal concordance between technical replicates of the same sera, including those run on different days (Figure S1). Comparative analysis of peptide abundance between serum/plasma and buffer-only negative controls revealed a strong correlation in abundance for the majority of peptides, while a subset of peptides showed distinctly higher relative abundance in each serum/plasma sample (Figure 1D). These latter peptides are those that have been enriched by binding to serum IgG. To quantify peptide enrichment, we calculated *Z* scores for each peptide in each sample. For each peptide, relative abundance was normalized to the corresponding value for the buffer-only negative controls, and this normalized value was compared among peptides with similar abundance in the buffer-only controls. Each *Z* score corresponds to the number of standard deviations away from the mean. For a peptide to be considered enriched, we also required a minimum fold-change compared to buffer-only controls and a minimum normalized read count. *Z* score, fold change, and read count thresholds were chosen to minimize false negatives based on the analysis of buffer-only controls.

Characterization of SARS-CoV-2 antibody epitopes

As expected, multiple positive control peptides were found to be enriched in every serum sample that we tested (Figure 2A), and there was no significant difference between convalescent and negative samples in the number of enriched control peptides (t test, $p = 0.47$). In contrast, we detected significantly more SARS-CoV-2 peptides enriched in convalescent samples compared to controls in both target proteins (t test; S: $p = 2.2e{-}7$, 6.2-fold difference; N: $p = 1.9e{-}6$, 15.7-fold difference) (Figure 2A). We observed at least 1 enriched SARS-CoV-2 peptide from 50/55 convalescent samples (91%), with an average of 18 enriched peptides per sample; while enriched SARS-CoV-2 peptides were only observed in 25/68 (37%) negative samples, with an average of 2 enriched peptides per sample. For the convalescent donors, there was no correlation between

Table 1. Summary of samples characterized in this study

Sample type	Sample size	Males/ females/ unreported	Median age, y[a]	Median days from diagnosis
COVID-19 convalescent	55	22/33/0	51	30
Negative control	69	19/14/36	42	–

[a]Median values calculated from a subset of total samples for which information was available. See Table S1 for details.

the number of enriched control and SARS-CoV-2 peptides ($p = 0.94$). Therefore, the absence of SARS-CoV-2 reactivity in some convalescent samples does not appear to be related to sample quality or a generally low concentration of IgG. We also did not observe a significant effect of gender in overall SARS-CoV-2 reactivity in convalescent donors (t test, $p = 0.56$), nor a significant correlation between SARS-CoV-2 reactivity and the number of days between PCR diagnosis and sample collection (Pearson correlation = −0.13, $p = 0.35$) (Figure S2). Notably, however, the five convalescent donors without detectable SARS-CoV-2 peptide enrichment (Table S1) were well below the median age of the full convalescent donor population (22–43 versus 50). In fact, overall, we observed a significant positive correlation between age and the number of enriched SARS-CoV-2 peptides in convalescent donors (Pearson correlation = 0.33, $p = 0.014$), while a weak trend in the opposite direction was observed for the number of enriched control peptides (Pearson correlation = −0.21, $p = 0.13$), indicating that this pattern was not due to an overall higher level of reactivity in samples from older patients.

In total, we identified IgG reactivity (i.e., peptide enrichment) against 229 and 95 SARS-CoV-2 peptides in convalescent and negative control samples, respectively; 70 of these peptides were enriched in both sample types. The peptides enriched in convalescent samples clustered together into 10 putative epitopes within the S protein and 9 putative epitopes within the N protein (Figures 2B and 2C; Table S2). These epitopes were recognized at a range of prevalences across the sampled population. The 6 most widely recognized epitopes—S positions 560–572, 819–824, and 1,150–1,156 and N positions 166–169, 223–229, and 390–402—were each detected in 13%–49% of the convalescent samples tested (median = 28.2%, n = 55), and all of the convalescent samples with at least 1 enriched SARS-CoV-2 peptide were reactive to at least 1 of these 6 immunodominant regions (Figures 2B and 2C). Notably, we also observed the enrichment of peptides from 4/6 of these immunodominant regions in negative control samples, although at much lower rates (1.5%–20% reactive, median = 2.2%, n = 68). At the other extreme, 9 (47%) of the observed epitope regions were each detected in only a single convalescent donor. Overall, relatively little reactivity was detected to peptides within the RBD, suggesting that these epitopes require protein conformations that are not well represented by linear 30-mer peptides.

To evaluate the potential for the highly recurrent S protein epitopes to be targeted by neutralizing antibodies, we evaluated these within the context of the structure of the protein. The inferred core regions (i.e., sequences present in all enriched peptides from assays of convalescent donors) of each of these epitopes were mapped onto a rendering of the three-dimensional structure of the native S trimer (Figure S3A). All three epitope regions are accessible for antibody binding on the surface of the trimer. The most widely recognized region (1,150–1,156) is located within the "stem helix" just upstream and partially overlapping with the heptad repeat region 2 (HR2). This region is proximal to the transmembrane domain and unresolved in the native structure; however, comparison of pre- and postfusion structures indicated that the HR2 epitope lies within a region that undergoes a dramatic conformational rearrangement during fusion (Figure S3B). The second epitope (819–824) resides near the S2′ cleavage site, spanning the fusion peptide (FP), whose exposure and incorporation into the host membrane are essential steps in virus entry into cells. Based on their proximity to these important functional sites, these epitopes are hereafter referred to as HR2 and FP, respectively. Finally, the 560–572 epitope occurs in the subdomain SD1 region (in the S1 subunit but C-terminal of the RBD).

To explore the diagnostic potential of the six highly recurrent S and N epitopes, we compared the maximum *Z* scores per epitope across the full set of convalescent and negative samples (Figure 2D). Across all six epitopes, we observed an overall shift toward higher *Z* scores in convalescent samples, which suggests the presence of additional antibody reactivity at these epitopes that is below our current enrichment thresholds. In fact, at 5/6 of these epitopes, we observed a significant difference in the mean of the *Z* score distributions between convalescent and negative samples (t test, S560: $p = 0.001$, FP: $p = 0.036$, HR2: $p = 0.008$, N223: $p = 0.043$, N390: $p = 0.008$). To estimate the combined diagnostic performance of these six epitopes, we built logistic regression models using the maximum peptide *Z* score for each of the epitopes as features and the donor status (convalescent versus negative) as the predicted outcome. Cross-validated models each trained on a randomly selected subset of 70% of donors and tested on the remaining 30% gave a mean area under the curve of 0.92 (Figure 2E).

Conserved antibody epitopes across the HCoVs

The observation of reactivity against several SARS-CoV-2 epitopes in a small subset of negative donors (Figure 2) suggests the presence of cross-reactive antibodies raised in response to other antigens. We hypothesized that some of these responses may be explained by exposure to conserved protein regions of the related endemic HCoVs. To test this hypothesis, we compared the SARS-CoV-2 reactivity profiles with patterns of reactivity to 244,000 peptides designed from hundreds of other human-infecting viruses, including all of the endemic HCoVs (HV library). Specifically, we performed comprehensive pairwise correlations of reactivity for the 6 immunodominant SARS-CoV-2 epitopes detected using the SCV2 library (described above) versus each of the 244,000 peptides of the HV library, across all convalescent and negative donors (Figure 3A). For 3 of these epitopes (FP, HR2, and N166), we detected a subset of HV peptides whose reactivity was strongly correlated with the SARS-CoV-2 response. Almost exclusively, these highly correlated responses corresponded to CoV peptides with homology

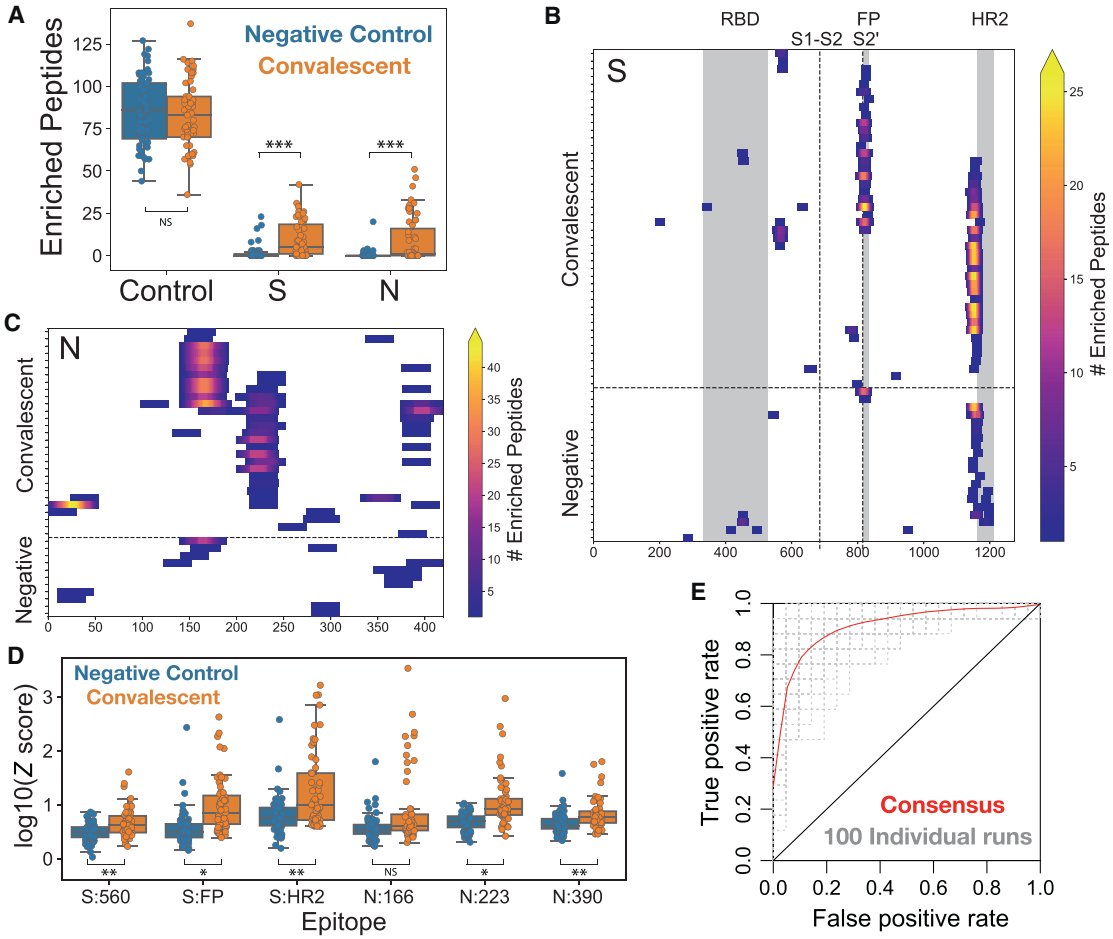


Figure 2. PepSeq identifies recurrent reactivities to SARS-CoV-2 peptides and classifies exposure status

(A) Boxplots showing the number of enriched SCV2 library peptides from assays with negative control (blue, n = 68) and COVID-19 convalescent (orange, n = 55) samples, divided into 3 different categories: non-SARS-CoV-2 control peptides (Control), and SARS-CoV-2 Spike (S) and Nucleocapsid (N) peptides. ***t test with $p < 1e{-}5$, NS, not significant). Individual data points are shown as circles, the limits of the boxes correspond to the 1st and 3rd quartiles, the black line inside each box corresponds to the median, and the whiskers extend to points that lie within 1.5 interquartile ranges of the 1st and 3rd quartiles.

(B and C) Heatmaps showing the locations of enriched SARS-CoV-2 peptides within the S and N proteins, respectively. Each row represents a single serum/plasma sample and each plot includes only samples with at least 1 enriched peptide from the focal protein. Each position is colored according to the number of enriched peptides that overlap that position. The horizontal dashed line separates COVID-19 convalescent samples (top) from negative control samples (bottom). The vertical dashed lines in (B) represent the S1–S2 and S2′ cleavage sites, respectively. The gray boxes indicate selected functional regions: receptor binding domain (RBD), fusion peptide (FP), and heptad repeat 2 (HR2).

(D) Boxplots showing the distribution of *Z* scores across all assayed samples for the 6 most common epitope reactivities observed in (B) and (C). For each sample/epitope combination, the *Z* score of the most enriched, overlapping peptide is presented. Boxplots were drawn as described in (A), with convalescent samples in orange and negative controls in blue. t test: *p < 0.05, **p < 0.01, NS, not significant.

(E) Receiver-operating curves showing sensitivity/specificity across a range of thresholds with which logistic regression models trained on randomly selected subsets of 70% of the donors were able to classify the remaining 30% of donors as either negative control or convalescent, using log-transformed *Z* scores for the 6 epitopes described in (D) as features. The red curve shows the average of 100 individual runs. Each patient sample was assayed in duplicate. Enriched peptides were determined based on consistent signal across replicates and *Z* scores shown as averages across replicates.

to the focal epitope, in each case including peptides from the closely related but non-endemic SARS-CoV. For the N166 epitope in the Nucleocapsid protein, only the homologous SARS-CoV peptides were highly correlated with detected SARS-CoV-2 reactivity, indicating a lack of cross-reactivity with homologous regions in endemic HCoVs. In contrast, at the HR2 epitope, in addition to SARS-CoV peptides, the reactivity profile of SARS-CoV-2 strongly correlated with those of 3 homologous peptides from the more distantly related Betacoronavirus 1 species, which includes HCoV-OC43. This is consistent with the presence of antibodies that cross-react with both SARS and HCoV-OC43 peptides. Finally, we observed evidence for broad cross-reactivity across CoVs at the FP epitope. At this epitope, the SARS-CoV-2 response correlated strongly with 18 homologous CoV peptides, including peptides from all 4 endemic HCoVs, MERS-CoV, SARS-CoV, and Alphacoronavirus 1.

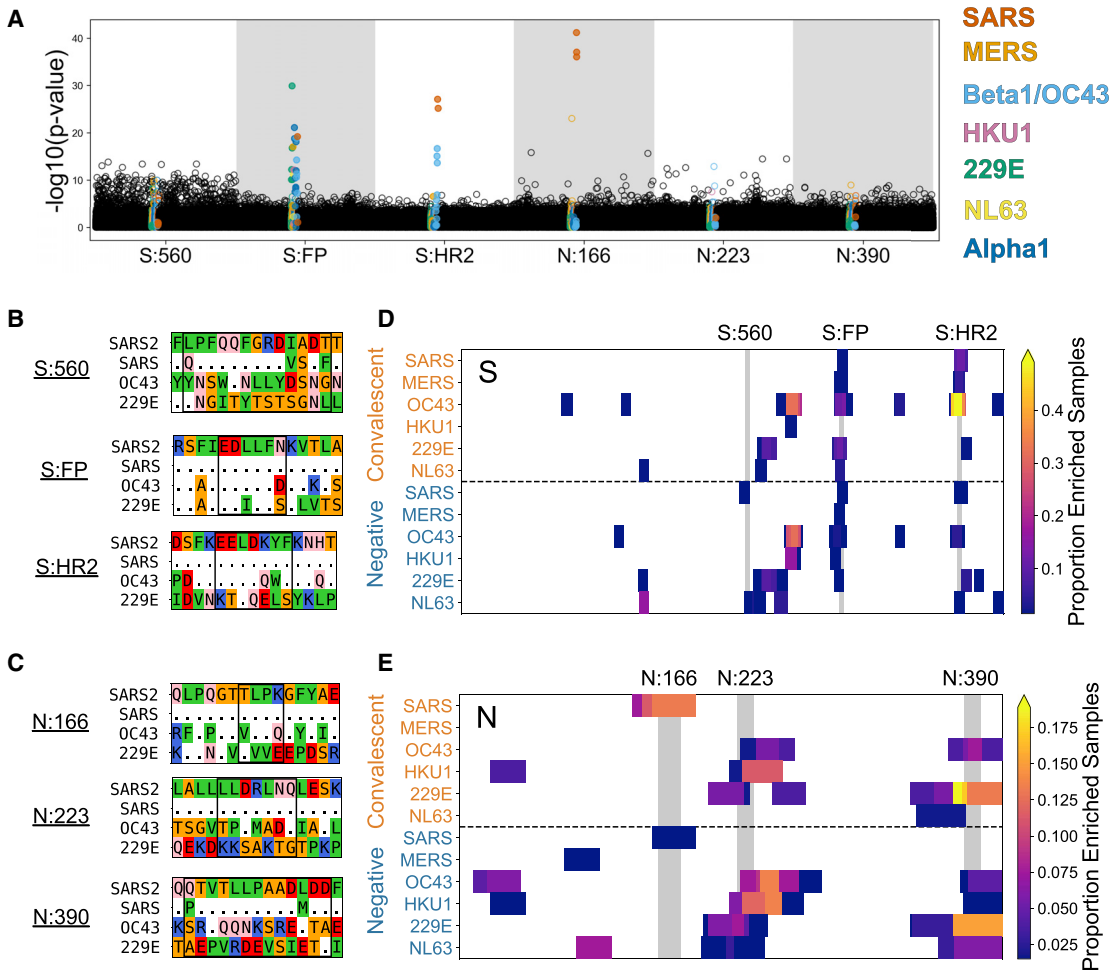


Figure 3. Epitope-level SARS-CoV-2 reactivity strongly correlates with reactivity to endemic CoVs at 2 highly conserved regions of Spike S2

(A) Correlations between reactivity to the 6 recurrent SARS-CoV-2 epitopes identified in Figure 2 and each of the 244,000 human virome-wide peptides in the HV library. Each dot represents the −log10(p value) of a Pearson correlation of log-transformed *Z* scores across all convalescent and control donors for the focal pair. Peptides corresponding to the CoV species in HV are colored as indicated, and filled circles represent peptides that are homologous to the focal SARS-CoV-2 epitope.

(B and C) Multispecies sequence alignments for each of the 6 most commonly observed SARS-CoV-2 epitopes. In each case, 15 amino acids are shown, and the minimally reactive region for each (inferred from the data presented in Figures 2B and 2C) is highlighted with a black box. Residues are colored according to amino acid properties: small non-polar (orange), hydrophobic (green), polar (pink), negatively charged (red), and positively charged (blue). Positions that are identical to the SARS-CoV-2 sequence are indicated with ".". Accessions for sequences in (B) Uniprot: P0DTC2, P59594, P36334, P15423 and (C) Uniprot: P0DTC9, P59595, P33469, P15130.

(D and E) Heatmaps illustrating the relative locations of enriched HV library peptides within the S (D) and N (E) proteins and across all HCoVs. Results have been aggregated across all of the tested samples, and the color at each location indicates the proportion of tested samples with enriched peptides covering that position. Results from convalescent (orange, top) and negative control (blue, bottom) are presented individually and are separated by the dashed line.

Vertical gray bars indicate the locations of the 6 epitopes highlighted in (A), (B), and (C). SARS2, SARS-CoV-2; SARS, SARS-CoV; MERS, MERS-CoV; Beta1, Betacoronavirus 1; OC43, HCoV-OC43; HKU1, HCoV-HKU1; 229E, HCoV-229E; NL63, HCoV-NL63; Alpha1, Alphacoronavirus 1. Each patient sample was assayed in duplicate. Enriched peptides were determined based on consistent signal across replicates.

In further support of antibody cross-reactivity, the observed correlations in reactivity were generally consistent with the patterns of sequence conservation between species at each epitope (Figures 3B and 3C). The FP epitope, for example, is highly conserved across members of both the Alphacoronavirus and Betacoronavirus genera (Figure 3B). This is also the epitope at which we observed correlated reactivity across all of the included CoV species. However, the HR2 epitope is highly conserved between SARS-CoV-2, SARS-CoV, and HCoV-OC43 (all members of Betacoronavirus), but not with HCoV-229E (Alphacoronavirus) (Figure 3B), again consistent with the detected patterns of correlation. Despite modest amino acid-level conservation between SARS-CoV-2 and the endemic HCoVs at N166, we did not observe any evidence of cross-reactivity with endemic species; however, correlation with SARS-CoV peptides at this epitope is consistent with perfect conservation

between the 2 SARS viruses at this region. Finally, the complete lack of sequence conservation between SARS-CoV-2 and the endemic CoVs at the S560, N223, and N390 epitopes is consistent with the absence of any evidence of cross-reactivity. However, the lack of a correlated response against SARS-CoV peptides at these three epitopes was unexpected. Although these regions are still highly conserved across these emergent CoVs, they are the least conserved of the six primary SARS-CoV-2 epitopes, and the divergence between viruses in this region may be sufficient to affect antibody affinity. This absence of SARS-CoV reactivity could also be partly related to the overall reduction in sensitivity that we observed for the HV library compared to the SCV2 library at the FP, HR2, and N166 epitopes (Figure S4). This reduced sensitivity is not surprising because (1) the increased complexity of the HV library (2 orders of magnitude greater than SCV2) resulted in a lower concentration of each peptide within the assay and (2) the HV library contains fewer unique peptides covering each epitope of interest.

These data reveal the existence of 2 distinct classes of linear SARS-CoV-2 epitopes: (1) those that are highly conserved across CoV species and likely elicit cross-reactive antibodies (discussed in more detail below) and (2) those that are not highly conserved and therefore appear to elicit antibodies that uniquely bind to SARS-CoV-2 peptides. Notably, for some of the epitopes contained in this second category, we did observe independent signatures of antibody reactivity to homologous regions of several endemic HCoVs, in both convalescent and negative donors. This is true, for example, for two commonly reactive epitopes in the SARS-CoV-2 Nucleocapsid protein: N223 and N390 (Figure 3E). In fact, within our negative control sample set, we observed enriched peptides, overlapping both of these epitopes, from all four endemic HCoVs, with several of these homologous reactivities observed in >10% of the assayed samples. This finding suggests that despite the lack of sequence conservation, there are general commonalities in the immunogenic features of these protein regions between species. Furthermore, despite the lack of evidence for cross-reactive antibodies that target the S560 and N166 epitopes, we observed rare but significant reactivity to SARS-CoV-2 peptides from these regions in prepandemic serum samples. This finding suggests that non-CoV and potentially non-viral antigens may also have the potential to elicit antibodies that cross-react with common SARS-CoV-2 antibody epitopes.

Interspecies cross-reactivity profiles of antibodies binding Spike S2 epitopes

Our results revealed two immunodominant antibody epitopes that likely elicit antibodies that are reactive across CoV species. Both epitopes are located in the S2 subunit of Spike and both exhibit high sequence conservation across multiple CoV species. Furthermore, at both sites, we observed (1) SARS-CoV-2 reactivity in negative control subjects, (2) strongly correlated reactivity across donors between peptides designed from SARS-CoV-2 and endemic CoVs, and (3) a markedly elevated frequency of reactivity to homologous regions of endemic CoVs in convalescent subjects relative to negative controls (Figures 3D and 3E). These observations highlight an apparently bidirectional cross-reactivity: prior endemic exposures appear to generate reactivity to SARS-CoV-2, and SARS-CoV-2 exposure elicits increased reactivity against the endemic CoVs. To conclusively determine whether the same antibodies are capable of binding to both the pandemic and endemic antigens, we selected three to six convalescent donors who had a high level of reactivity at each epitope and selectively depleted antibodies that bind the SARS-CoV-2 epitope using bead-bound peptides. We then re-assayed these depleted samples with both the SCV2 and HV PepSeq libraries (Figure 4A). As expected, depletion using SARS-CoV-2 HR2 or FP peptides resulted in a marked loss of signal for all of the corresponding SARS-CoV-2 peptides that were enriched in the non-depleted samples (HR2: 2- to 380-fold decrease, median = 13; FP: 8- to 267-fold decrease, median = 42), but no consistent decrease in signal for SARS-CoV-2 peptides outside the depleted epitope. Notably, depletion with the respective SARS-CoV-2 peptides also caused a similar loss of signal for all previously enriched homologous peptides from other endemic and pandemic CoVs: HCoV-OC43 and SARS-CoV peptides for the HR2 epitope and HCoV-229E, HCoV-OC43, and SARS-CoV peptides for the FP epitope. The loss of reactivity to endemic CoV peptides following antibody depletion with SARS-CoV-2 peptides directly demonstrated the existence of cross-reactive antibodies and showed that these cross-reactive antibodies are the primary contributors to the signal we observed for these endemic CoV peptides.

Having confirmed their cross-reactivity, we studied the species-reactivity profiles of these responses by comparing reactivity in the HV library to peptides designed from SARS-CoV and the two most common endemic HCoVs (HCoV-OC43 and HCoV-229E) at both of the Spike S2 epitopes (as well as N166, which served as a control for which we see no evidence of pandemic/endemic cross-reactivity). As the SARS-CoV and SARS-CoV-2 genomes are 100% identical across each of these epitopes, this strategy allowed a stringent comparison between endemic and pandemic reactivities without the potentially confounding effect of differences in sensitivity between the HV and SCV2 libraries. For both the FP and HR2 epitopes, we observed that the maximum reactivity across the endemic homologs was almost universally, and often substantially, higher than the maximum reactivity to the corresponding SARS-CoV/SARS-CoV-2 peptides, across both the convalescent and negative control donors exhibiting evidence of epitope-specific enrichment (FP: 3- to 19-fold difference, median = 5.3; HR2: 0.7- to 2,358-fold difference, median = 8.3) (Figure 4B). However, we observed a very different pattern at the N166 epitope, at which SARS-CoV/SARS-CoV-2 reactivity was universally higher for strongly reactive samples, including both convalescent and negative control donors. Thus, even though SARS-CoV-2 was likely to be the most recent CoV exposure in the majority of the convalescent donors, the fine-level specificity of the responding Spike S2 antibodies appears to be imprinted by previous exposure to endemic CoVs.

To more precisely resolve the specificity of the cross-reactive Spike S2 antibodies between the various endemic CoVs, we selected subsets of convalescent donors with strong reactivity in each of the FP, HR2, and N166 epitopes, and then used the

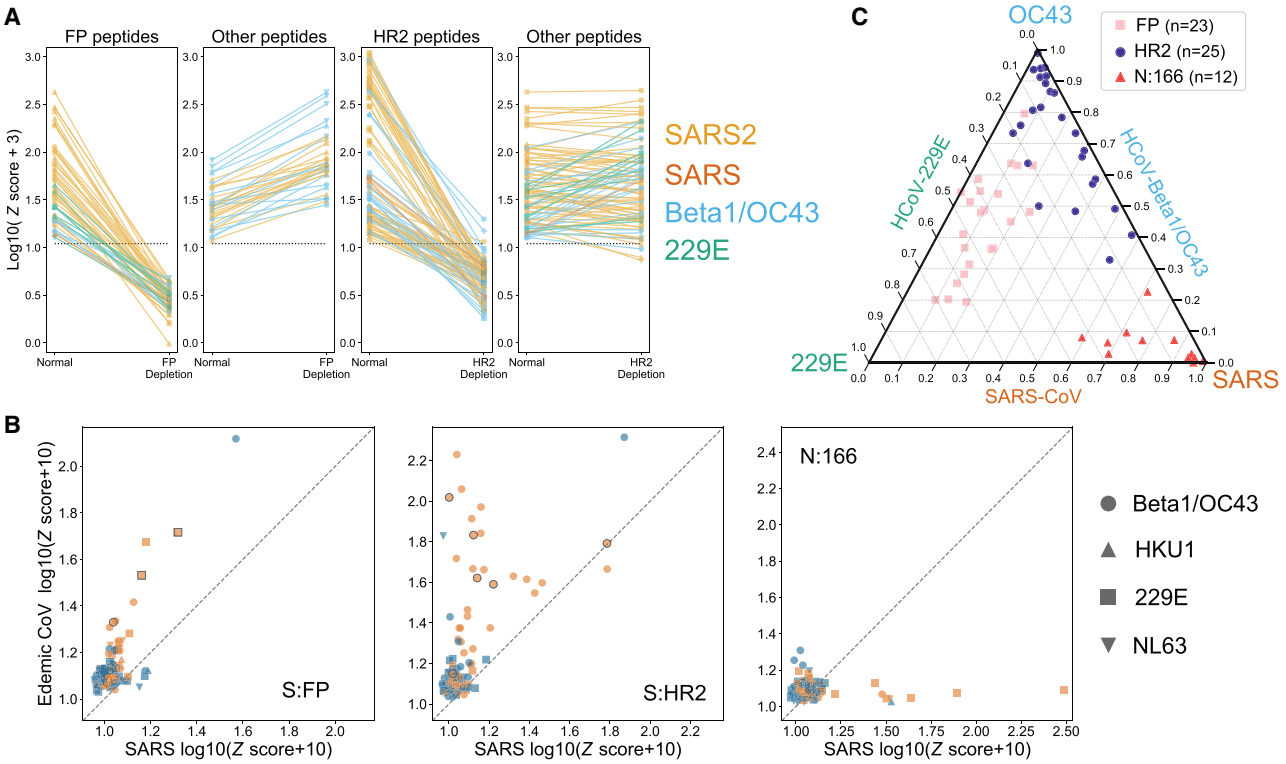


Figure 4. SARS-CoV-2 elicits cross-reactive Spike S2 antibodies that preferentially recognize homologs from the endemic CoVs

(A) Line plots comparing peptide-specific patterns of enrichment before and after targeted depletion of antibodies binding the SARS-CoV-2 FP or HR2 antigens. Each plot compares peptide enrichment (*Z* score) from the same samples before (left) and after (right) antibody depletions. Each line represents a single peptide found to be enriched before antibody depletion, and the color of each line indicates the species from which the peptide was designed. Each plot includes results from 3–6 convalescent donors (peptides from each donor are depicted with different shapes). For each depletion experiment (FP and HR2 targeted), on- and off-target peptides are plotted separately. Dashed horizontal lines represent a *Z* score of 8.

(B) Scatterplots comparing enrichment (*Z* score) between SARS-CoV peptides (x axis) and endemic HCoV peptides (y axis) across 3 epitopes (S:FP, S:HR2, and N:166, respectively) and all samples assayed in duplicate using the HV library (average *Z* score is shown). Convalescent and negative control samples are represented by orange and blue shapes, respectively. The type of shape indicates the endemic CoV species from which the most highly enriched peptide was designed (circle, HCoV-OC43; square, HCoV-229E; upright triangle, HCoV-HKU1; and upside-down triangle, HCoV-NL63). Shapes with black outlines indicate the samples included in the antibody-depletion experiments shown in (A).

(C) Ternary plot showing the relative signal across HV library peptides from 3 HCoVs (SARS-CoV, HCoV-OC43, and HCoV-229E) at 3 commonly reactive epitopes in COVID-19 convalescent patients. Each point represents a single convalescent sample that exhibited at least 1 enriched SCV2 library peptide at the relevant epitope. Position within the triangle was determined by normalizing the maximum peptide *Z* score (averaged across replicates) observed for each of the 3 focal species.

HV data to apportion their relative reactivity across 3 viruses: HCoV-OC43, HCoV-229E, and SARS-CoV (Figure 4C). This was accomplished by normalizing reactivity across the three species using the peptide with the highest *Z* score from each. As expected for an epitope without endemic CoV cross-reactivity, the N166 response showed a SARS-dominant profile, with relatively minor reactivities to HCoV-OC43 and HCoV-229E homologs. In contrast, the magnitude of the FP response was primarily dominated by HCoV-OC43 and HCoV-229E, with the relative loadings varying across samples, and a substantially smaller SARS-CoV component, consistent with the conservation of this epitope across the 3 species, but indicating stronger recognition of the 2 endemic homologs. Finally, the HR2 response showed an HCoV-OC43-dominant pattern with a SARS-CoV component that varied across donors, again indicating a response likely shaped primarily by previous exposure to an endemic virus and reflecting the strong conservation of this region between SARS-CoV/SARS-CoV-2 and HCoV-OC43, but not HCoV-229E (Figure 3B).

DISCUSSION

Like most viruses, SARS-CoV-2 elicits a robust antibody response whose targets are likely to be important determinants of disease outcome and the extent of protection conferred following natural infection or vaccination. In this study, we describe a customizable platform that enables epitope-resolved profiling of the antibody response (PepSeq), and its application to the study of HCoVs, including SARS-CoV-2. Using this system, we identified frequently targeted epitopes in both the S and N proteins, two of which overlap conserved, functional sites in the Spike S2 subunit, and therefore have the potential

to be sites of broadly neutralizing reactivity. By quantifying antibody reactivity to homologous antigens from all HCoVs and using targeted antibody-depletion experiments, we also demonstrate that the antibody response cross-recognizes antigens from SARS-CoV-2 and at least one endemic CoV at both of these conserved S2 epitopes.

By independently testing reactivity across thousands of potential epitopes, we identified several with promise for use in both diagnostics and functional characterization assays. For the two epitopes we detected in the S2 subunit of Spike (each discussed in further detail below) structural considerations and previous characterization of related epitopes[24,38,39] indicate neutralization potential. In these cases, a peptide-based assay may provide a facile means of profiling functional reactivities independently of cell/viral culture, and in a way that complements ACE2:RBD binding inhibition assays that cannot measure S2 reactivity.[40] We also identified a set of 6 recurrent epitopes across the S and N proteins that together exhibit the potential for generating an accurate profile of SARS-CoV-2 exposure. However, our identification of epitopes at which there is cross-reactivity between the pandemic and endemic CoVs also highlights a source of potential false-positive reactivity that may limit SARS-CoV-2 serological assays generally. Future studies are needed to examine the impact of these cross-reactivities on diagnostic assay performance by studying individuals who are known to have recently experienced an endemic HCoV infection. Outside of the present pandemic, assays focused on reactivity to epitopes of the S2 subunit, in particular, may be better suited to the general detection of CoV exposure, rather than as a specific indication of SARS-CoV-2 infection.

Our PepSeq analysis identified a region centered on positions 1,150–1,156 in the Spike S2 domain (EELDKYF, HR2) as the most widely recognized SARS-CoV-2 linear epitope in convalescent donors (Figure 2B). This region is located within the stem helix, directly N-terminal of the HR2 region. While unresolved in the prefusion structure, analysis of postfusion structures of CoV Spike proteins indicate that HR2 undergoes a ~180° reorientation during the formation of the 6-helix bundle in which it comes into close contact with the HR1 region.[41] HR-derived peptides that disrupt the HR1:HR2 interaction have previously been shown to inhibit infection by other CoVs,[42,43] highlighting the strong potential for functional targeting of this region. Moreover, neutralizing monoclonal antibodies raised against related CoVs, including SARS-CoV (which has >95% amino acid-level identity to SARS-CoV-2 at the stem helix of HR2), have been shown to bind a region directly adjacent to the one that we identified in this study[38,39,44] (Figure S3).

The second immunodominant reactivity that we identified in Spike S2 also occurs in a region whose sequence is highly conserved across CoV species. This reactivity is centered on positions 819–824 (EDLLFN, FP), which is adjacent to the S2′ cleavage site and overlaps the FP. Given the proximity of this minimal reactive region to the S2′ cleavage site, this reactivity has the potential to block proteolytic processing and thereby prevent maturation of the S protein. Alternatively, and perhaps additionally, binding of antibody to the FP is expected to prevent its insertion into the host membrane and therefore prevent fusion and cell entry. A recent study, using a lower-throughput peptide-based approach also identified this FP epitope as reactive in two SARS-CoV-2 convalescent donors, and while they did not characterize the mechanism of action, they demonstrated the neutralization potential of antibodies against this epitope using antibody-depletion assays[24] (Figure S3). This study additionally reported an epitope downstream of the Spike RBD to which antibodies also exhibited neutralization potential. We observed reactivity to this same region in 7 of our COVID-19 convalescent donors (N560, minimal reactive sequence: LPFQQFGRDIADT) (Figure S3).

Our results demonstrate that the antibodies recognizing the FP and HR2 SARS-CoV-2 epitopes also cross-react with homologous antigens from common endemic HCoVs, and we observed preexisting reactivity to SARS-CoV-2 peptides to both of these epitopes in several donors without SARS-CoV-2 exposure. This cross-reactivity likely explains some of the immunodominance of these antibodies in response to SARS-CoV-2 infection, as prior infections with endemic CoVs, such as HCoV-OC43 and HCoV-229E, have likely primed responses to these regions. In fact, in the vast majority of samples tested, these cross-reactive antibodies exhibited preferential binding of endemic HCoV peptides, compared to peptides from SARS-CoV/SARS-CoV-2 (described in more detail below). This pattern is consistent with the imprinting of antibody responses on prior antigen exposures,[45,46] and it remains to be seen how such imprinting may affect the neutralization potential of these antibodies against emerging viruses such as SARS-CoV-2.

In addition to the Spike S2 epitopes, we observed some pre-pandemic reactivity to SARS-CoV-2 peptides at the N166 and N390 epitopes. However, in contrast to the S2 epitopes, preexisting reactivity at these epitopes does not appear to be the result of previous CoV exposures. These regions are not well conserved across HCoVs (Figure 3C), and we did not observe any evidence for cross-reactive antibodies that recognize both SARS-CoV-2 and endemic HCoV peptides at these regions (Figure 3A). This suggests that these preexisting reactivities were generated in response to a non-CoV antigen. We did observe several non-CoV HV library peptides that showed some degree of correlation with SARS-CoV-2 reactivity at N166, but none of these peptides exhibited obvious sequence similarity to the SARS-CoV-2 N166 epitope, and we did not find any peptides with responses that were strongly correlated to the SARS-CoV-2 reactivity at N390 (Figure 3A). Therefore, the sources of the antigens that triggered these preexisting reactivities are currently unknown and may not be of viral origin at all. Notably, we did not detect any reactivity within our negative controls to S560 or N223. Therefore, of the six most commonly reactive SARS-CoV-2 epitopes detected here, these may turn out to be the most specific indicators of SARS-CoV-2 exposure.

Despite well-documented serological reactivity in studies using the full-length RBD antigen,[47] we observed very little reactivity to peptides designed from the RBDs of HCoVs, including SARS-CoV-2 (Figures 2B and 3D), and the SARS-CoV-2 reactivity that we did observe in this region was very similar in both convalescent and negative samples. This lack of reactivity in our assay, as well as a similar absence of reactivity in a recent study using a lower-throughput peptide-based approach,[24] suggests that antibodies to the RBD recognize

conformational epitopes and/or depend on posttranslational modifications. Like other peptide-based antibody assays, PepSeq is limited to the detection of epitopes that are well represented by linear peptides and do not require posttranslational modifications. The dependence of RBD epitopes on secondary/tertiary structure is supported by structural analyses of the footprints of neutralizing antibodies bound to Spike RBD, which indicate the involvement of residues that are distal in the linear sequence.[15,48] The identification of epitopes such as these will require lower throughput approaches, including mutagenesis and/or structural studies, or potentially the use of longer peptide fragments.

The observation of markedly increased reactivity to the endemic homologs of the FP and HR2 epitopes in convalescent versus negative donors (Figures 3D and 4B) indicates that infection with SARS-CoV-2 elicits a cross-reactive antibody response at these sites. Conversely, the detection of reactivity to the SARS-CoV-2 homologs of these epitopes in occasional negative donors (Figure 2B) is most likely explained by endemic CoV-generated antibodies that cross-react with SARS-CoV-2. Despite this bi-directionality, we show that the cross-reactive response in SARS-CoV-2 convalescent donors recognizes the endemic homologs substantially more strongly for both epitopes (Figures 4B and 4C), which is most consistent with a model in which preexisting B cell clones previously raised and matured against the endemic CoVs are recruited into the response to SARS-CoV-2. Such preexisting cross-reactive clones would be expected to have a range of intrinsic affinities for the homologous SARS-CoV-2 epitope, and these could be further improved by somatic mutation. However, by analogy with other viruses, there may be a limit to the efficiency with which such preexisting antibodies can be redirected, reflecting their imprinting.[45,46] Under this model, convalescent donors who exhibit detectable reactivity to endemic CoVs but not to SARS-CoV-2 at the HR2 epitope (upper left in Figure 4B, center panel) represent cases in which preexisting antibodies bind only weakly to SARS-CoV-2 (below the threshold of the assay) and have been unable to acquire a high affinity for the new virus. This would suggest that B cell memory against the FP and HR2 epitopes capable of cross-reacting with SARS-CoV-2 is prevalent in the general population, consistent with the near-universal seropositivity reported for endemic CoVs,[49] although often below our limit of detection.

Our findings raise the possibility that the nature of an individual's antibody response to prior endemic CoV infection may affect the course of COVID-19. They also indicate that analysis of S2 reactivity is crucial for a complete assessment of the humoral response to SARS-CoV-2, consistent with the observation that S2-only assays provide an equally strong correlate of neutralization compared to RBD-only assays (J. Nikolich and D. Bhattacharya, personal communication). The HR2 and FP cross-reactivities characterized here represent a possible source of background signal for SARS-CoV-2 serological assays that include the S2 subunit of Spike, which would be absent in those targeting only the RBD, for which sequence conservation is lower across CoV species.[50] However, our findings also indicate that the incorporation of related endemic CoV antigens may improve the sensitivity of SARS-CoV-2 serological analyses, and in particular, that a differential analysis of SARS-CoV-2 and endemic CoV Spike S2 reactivity may provide an important measure of the efficiency with which preexisting cross-reactive responses can be redirected.

The identification of broadly immunogenic epitopes in conserved functional domains of the SARS-CoV-2 Spike S2 subunit, including cross-reactivity with endemic HCoVs, also has implications for the design of therapeutic antibodies and vaccines. SARS-CoV-2 vaccines under development predominantly use two forms of the S antigen—whole protein or the RBD—and in each case are designed primarily to elicit neutralizing antibodies.[51] Relative to RBD-focused vaccines, we hypothesize that vaccines that include the Spike HR2 and FP sites (1) will be able to induce a broader array of neutralizing reactivities, (2) may be more capable of rapidly recruiting preexisting memory B cells that are prevalent in the population, and (3) may be less prone to viral escape due to a lower tolerance for amino acid substitutions. In particular, the identification of HR2 and FP as conserved, functionally important and broadly immunogenic sites capable of eliciting cross-reacting antibodies, makes these regions candidates for the development of broadly neutralizing responses against CoVs. Future work should resolve the functional consequences of these cross-reactive antibody responses, and how an individual's exposure history to endemic CoVs may affect their course of COVID-19.

Limitations of study

Our study is limited by a reliance on 30-mer peptides that contain no posttranslational modifications and are expected to be generally unable to form conformational structures. Therefore, we expect that some CoV-binding antibodies will be missed using our approach. Furthermore, although we observe a number of immunodominant epitopes, our cohort size (n = 55 convalescent donors) may leave some rarer epitopes undetected. In addition, our SARS-CoV-2 PepSeq library focused only on two proteins, S and N, which have generally been shown to be the targets of most anti-CoV antibodies. Nonetheless, this limitation prevents us from drawing conclusions regarding the potential for cross-reactive antibodies that recognize other CoV proteins. Finally, although we speculate on the potential for antibody binding at particular identified Spike epitopes to have neutralizing activity, these hypotheses are based on structural considerations and analogy with prior studies, rather than any direct functional data contained herein.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.xcrm.2020.100189.

ACKNOWLEDGMENTS

This work was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award nos. U24AI152172 and U24AI152172-01S1; the National Institute on Minority Health and Health Disparities of the National Institutes of Health under award no. U54MD012388; and the State of Arizona Technology and Research Initiative Fund (TRIF), administered by the Arizona Board of Regents, through Northern Arizona University. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. Z.W.F. was supported by a Hooper Undergraduate Research Award and a Jean Shuler Research Mini-Grant. We are grateful for assistance from Mike Busch and Phillip Williamson in obtaining samples and Piotr Swiderski in synthesizing a custom oligonucleotide reagent. We are also grateful for the constructive feedback received from Paul Keim, Stephen Daley, Gerard Zurawski, and Erik Settles, which greatly improved the manuscript.

AUTHOR CONTRIBUTIONS

Conceptualization, J.T.L. and J.A.A.; Methodology, J.T.L., N.M., M.S.C., and J.A.A.; Software, J.T.L. and Z.W.F.; Formal Analysis, J.T.L. and J.A.A.; Investigation, S.N.H., A.S.B., A.L.E., and F.R.; Resources, J.D., K.E.S., M.S., W.D., S.D., J.Y., M.A.C., M.B., M.H.F., S.A.N., and D.E.K.; Data Curation, J.T.L., Z.W.F., and S.A.S.; Writing – Original Draft, J.T.L. and J.A.A.; Writing – Review & Editing, S.N.H., A.S.B., A.L.E., Z.W.F., F.R., D.E.K., and S.A.S.; Visualization, J.T.L., S.A.S., and J.A.A.; Supervision, J.T.L. and J.A.A.; Project Administration, J.T.L. and J.A.A.; Funding Acquisition, J.T.L. and J.A.A.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 13, 2020
Revised: November 11, 2020
Accepted: December 17, 2020
Published: January 19, 2021

REFERENCES

1. Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al.; China Novel Coronavirus Investigating and Research Team (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. N. Engl. J. Med. *382*, 727–733.
2. Piñana, J.L., Xhaard, A., Tridello, G., Passweg, J., Kozijn, A., Polverelli, N., Heras, I., Perez, A., Sanz, J., Berghuis, D., et al.; Infectious Diseases Working Party of the European Society for Blood and Marrow Transplantation and Infectious Complications Subcommittee of the Spanish Hematopoietic Stem Cell Transplantation and Cell Therapy Group (GETH) (2020). Seasonal human coronaviruses respiratory tract infection in recipients of allogeneic hematopoietic stem cell transplantation. J. Infect. Dis. https://doi.org/10.1093/infdis/jiaa553.
3. Dijkman, R., Jebbink, M.F., El Idrissi, N.B., Pyrc, K., Müller, M.A., Kuijpers, T.W., Zaaijer, H.L., and van der Hoek, L. (2008). Human coronavirus NL63 and 229E seroconversion in children. J. Clin. Microbiol. *46*, 2368–2373.
4. Callow, K.A., Parry, H.F., Sergeant, M., and Tyrrell, D.A. (1990). The time course of the immune response to experimental coronavirus infection of man. Epidemiol. Infect. *105*, 435–446.
5. Ni, L., Ye, F., Cheng, M.-L., Feng, Y., Deng, Y.-Q., Zhao, H., Wei, P., Ge, J., Gou, M., Li, X., et al. (2020). Detection of SARS-CoV-2-Specific Humoral and Cellular Immunity in COVID-19 Convalescent Individuals. Immunity *52*, 971–977.e3.
6. Liu, A., Li, Y., Peng, J., Huang, Y., and Xu, D. (2020). Antibody responses against SARS-CoV-2 in COVID-19 patients. J. Med. Virol. https://doi.org/10.1002/jmv.26241.
7. Casadevall, A., and Pirofski, L.-A. (2020). The convalescent sera option for containing COVID-19. J. Clin. Invest. *130*, 1545–1548.
8. Le, T.T., Andreadakis, Z., Kumar, A., Gómez Román, R., Tollefsen, S., Saville, M., and Mayhew, S. (2020). The COVID-19 vaccine development landscape. Nat. Rev. Drug Discov. *19*, 305–306.
9. Krammer, F., and Simon, V. (2020). Serology assays to manage COVID-19. Science *368*, 1060–1061.
10. Whitman, J.D., Hiatt, J., Mowery, C.T., Shy, B.R., Yu, R., Yamamoto, T.N., Rathore, U., Goldgof, G.M., Whitty, C., Woo, J.M., et al. (2020). Test performance evaluation of SARS-CoV-2 serological assays. medRxiv. https://doi.org/10.1101/2020.04.25.20074856.
11. Deeks, J.J., Dinnes, J., Takwoingi, Y., Davenport, C., Spijker, R., Taylor-Phillips, S., Adriano, A., Beese, S., Dretzke, J., Ferrante di Ruffano, L., et al.; Cochrane COVID-19 Diagnostic Test Accuracy Group (2020). Antibody tests for identification of current and past infection with SARS-CoV-2. Cochrane Database Syst. Rev. *6*, CD013652.
12. Nie, J., Li, Q., Wu, J., Zhao, C., Hao, H., Liu, H., Zhang, L., Nie, L., Qin, H., Wang, M., et al. (2020). Establishment and validation of a pseudovirus neutralization assay for SARS-CoV-2. Emerg. Microbes Infect. *9*, 680–686.
13. Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., et al. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet *395*, 565–574.
14. 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.
15. Pinto, D., Park, Y.-J., Beltramello, M., Walls, A.C., Tortorici, M.A., Bianchi, S., Jaconi, S., Culap, K., Zatta, F., De Marco, A., et al. (2020). Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. Nature *583*, 290–295.
16. Lv, H., Wu, N.C., Tsang, O.T.-Y., Yuan, M., Perera, R.A.P.M., Leung, W.S., So, R.T.Y., Chan, J.M.C., Yip, G.K., Chik, T.S.H., et al. (2020). Cross-reactive Antibody Response between SARS-CoV-2 and SARS-CoV Infections. Cell Rep. *31*, 107725.
17. Friesen, R.H.E., Lee, P.S., Stoop, E.J.M., Hoffman, R.M.B., Ekiert, D.C., Bhabha, G., Yu, W., Juraszek, J., Koudstaal, W., Jongeneelen, M., et al. (2014). A common solution to group 2 influenza virus neutralization. Proc. Natl. Acad. Sci. USA *111*, 445–450.
18. Du, L., He, Y., Zhou, Y., Liu, S., Zheng, B.-J., and Jiang, S. (2009). The spike protein of SARS-CoV–a target for vaccine and therapeutic development. Nat. Rev. Microbiol. *7*, 226–236.
19. Pillay, T.S. (2020). Gene of the month: the 2019-nCoV/SARS-CoV-2 novel coronavirus spike protein. J. Clin. Pathol. *73*, 366–369.
20. Robbiani, D.F., Gaebler, C., Muecksch, F., Lorenzi, J.C.C., Wang, Z., Cho, A., Agudelo, M., Barnes, C.O., Gazumyan, A., Finkin, S., et al. (2020). Convergent Antibody Responses to SARS-CoV-2 in Convalescent Individuals. Nature *584*, 437–442.
21. Chi, X., Yan, R., Zhang, J., Zhang, G., Zhang, Y., Hao, M., Zhang, Z., Fan, P., Dong, Y., Yang, Y., et al. (2020). A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. Science *369*, 650–655.

22. Hansen, J., Baum, A., Pascal, K.E., Russo, V., Giordano, S., Wloga, E., Fulton, B.O., Yan, Y., Koon, K., Patel, K., et al. (2020). Studies in humanized mice and convalescent humans yield a SARS-CoV-2 antibody cocktail. Science *369*, 1010–1014.

23. Zost, S.J., Gilchuk, P., Case, J.B., Binshtein, E., Chen, R.E., Nkolola, J.P., Schafer, A., Reidy, J.X., Trivette, A., Nargi, R.S., et al. (2020). Potently neutralizing and protective human antibodies against SARS-CoV-2. Nature *584*, 443–449.

24. Poh, C.M., Carissimo, G., Wang, B., Amrun, S.N., Lee, C.Y.-P., Chee, R.S.-L., Fong, S.-W., Yeo, N.K.-W., Lee, W.-H., Torres-Ruesta, A., et al. (2020). Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. Nat. Commun. *11*, 2806.

25. Hoofnagle, J.H., Gerety, R.J., Ni, L.Y., and Barker, L.F. (1974). Antibody to hepatitis B core antigen. A sensitive indicator of hepatitis B virus replication. N. Engl. J. Med. *290*, 1336–1340.

26. Lubroth, J., Grubman, M.J., Burrage, T.G., Newman, J.F., and Brown, F. (1996). Absence of protein 2C from clarified foot-and-mouth disease virus vaccines provides the basis for distinguishing convalescent from vaccinated animals. Vaccine *14*, 419–427.

27. Khurana, S., Loving, C.L., Manischewitz, J., King, L.R., Gauger, P.C., Henningson, J., Vincent, A.L., and Golding, H. (2013). Vaccine-induced anti-HA2 antibodies promote virus fusion and enhance influenza virus respiratory disease. Sci. Transl. Med. *5*, 200ra114.

28. Katzelnick, L.C., Gresh, L., Halloran, M.E., Mercado, J.C., Kuan, G., Gordon, A., Balmaseda, A., and Harris, E. (2017). Antibody-dependent enhancement of severe dengue disease in humans. Science *358*, 929–932.

29. Halstead, S.B., and O'Rourke, E.J. (1977). Antibody-enhanced dengue virus infection in primate leukocytes. Nature *265*, 739–741.

30. Eroshenko, N., Gill, T., Keaveney, M.K., Church, G.M., Trevejo, J.M., and Rajaniemi, H. (2020). Implications of antibody-dependent enhancement of infection for SARS-CoV-2 countermeasures. Nat. Biotechnol. *38*, 789–791.

31. Lucchese, G., Stufano, A., Trost, B., Kusalik, A., and Kanduc, D. (2007). Peptidology: short amino acid modules in cell biology and immunology. Amino Acids *33*, 703–707.

32. Fleri, W., Paul, S., Dhanda, S.K., Mahajan, S., Xu, X., Peters, B., and Sette, A. (2017). The Immune Epitope Database and Analysis Resource in Epitope Discovery and Synthetic Vaccine Design. Front. Immunol. *8*, 278.

33. Price, J.V., Tangsombatvisit, S., Xu, G., Yu, J., Levy, D., Baechler, E.C., Gozani, O., Varma, M., Utz, P.J., and Liu, C.L. (2012). On silico peptide microarrays for high-resolution mapping of antibody epitopes and diverse protein-protein interactions. Nat. Med. *18*, 1434–1440.

34. Larman, H.B., Zhao, Z., Laserson, U., Li, M.Z., Ciccia, A., Gakidis, M.A.M., Church, G.M., Kesari, S., Leproust, E.M., Solimini, N.L., and Elledge, S.J. (2011). Autoantigen discovery with a synthetic human peptidome. Nat. Biotechnol. *29*, 535–541.

35. Xu, G.J., Kula, T., Xu, Q., Li, M.Z., Vernon, S.D., Ndung'u, T., Ruxrungtham, K., Sanchez, J., Brander, C., Chung, R.T., et al. (2015). Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome. Science *348*, aaa0698.

36. Kozlov, I.A., Thomsen, E.R., Munchel, S.E., Villegas, P., Capek, P., Gower, A.J., Pond, S.J.K., Chudin, E., and Chee, M.S. (2012). A highly scalable peptide-based assay system for proteomics. PLoS ONE *7*, e37441.

37. Roberts, R.W., and Szostak, J.W. (1997). RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc. Natl. Acad. Sci. USA *94*, 12297–12302.

38. Lai, S.-C., Chong, P.C.-S., Yeh, C.-T., Liu, L.S.-J., Jan, J.-T., Chi, H.-Y., Liu, H.-W., Chen, A., and Wang, Y.-C. (2005). Characterization of neutralizing monoclonal antibodies recognizing a 15-residues epitope on the spike protein HR2 region of severe acute respiratory syndrome coronavirus (SARS-CoV). J. Biomed. Sci. *12*, 711–727.

39. Keng, C.-T., Zhang, A., Shen, S., Lip, K.-M., Fielding, B.C., Tan, T.H.P., Chou, C.-F., Loh, C.B., Wang, S., Fu, J., et al. (2005). Amino acids 1055 to 1192 in the S2 region of severe acute respiratory syndrome coronavirus S protein induce neutralizing antibodies: implications for the development of vaccines and antiviral agents. J. Virol. *79*, 3289–3296.

40. Tan, C.W., Chia, W.N., Qin, X., Liu, P., Chen, M.I.-C., Tiu, C., Chen, V.C.-W., Hu, Z., Young, B.E., Sia, W.R., et al. (2020). A SARS-CoV-2 surrogate virus neutralization test (sVNT) based on antibody-mediated blockage of ACE2-spike (RBD) protein-protein interaction. Nat. Biotechnol. *38*, 1073–1078.

41. Walls, A.C., Tortorici, M.A., Snijder, J., Xiong, X., Bosch, B.-J., Rey, F.A., and Veesler, D. (2017). Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. Proc. Natl. Acad. Sci. USA *114*, 11157–11162.

42. Xia, S., Yan, L., Xu, W., Agrawal, A.S., Algaissi, A., Tseng, C.K., Wang, Q., Du, L., Tan, W., Wilson, I.A., et al. (2019). A pan-coronavirus fusion inhibitor targeting the HR1 domain of human coronavirus spike. Sci. Adv. *5*, eaav4580.

43. Liu, S., Xiao, G., Chen, Y., He, Y., Niu, J., Escalante, C.R., Xiong, H., Farmar, J., Debnath, A.K., Tien, P., and Jiang, S. (2004). Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. Lancet *363*, 938–947.

44. Routledge, E., Stauber, R., Pfleiderer, M., and Siddell, S.G. (1991). Analysis of murine coronavirus surface glycoprotein functions by using monoclonal antibodies. J. Virol. *65*, 254–262.

45. Monto, A.S., Malosh, R.E., Petrie, J.G., and Martin, E.T. (2017). The Doctrine of Original Antigenic Sin: Separating Good From Evil. J. Infect. Dis. *215*, 1782–1788.

46. Gostic, K.M., Ambrose, M., Worobey, M., and Lloyd-Smith, J.O. (2016). Potent protection against H5N1 and H7N9 influenza via childhood hemagglutinin imprinting. Science *354*, 722–726.

47. Amanat, F., Stadlbauer, D., Strohmeier, S., Nguyen, T.H.O., Chromikova, V., McMahon, M., Jiang, K., Arunkumar, G.A., Jurczyszak, D., Polanco, J., et al. (2020). A serological assay to detect SARS-CoV-2 seroconversion in humans. Nat. Med. *26*, 1033–1036.

48. Yuan, M., Wu, N.C., Zhu, X., Lee, C.D., So, R.T.Y., Lv, H., Mok, C.K.P., and Wilson, I.A. (2020). A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. Science *368*, 630–633.

49. Gorse, G.J., Patel, G.B., Vitale, J.N., and O'Connor, T.Z. (2010). Prevalence of antibodies to four human coronaviruses is lower in nasal secretions than in serum. Clin. Vaccine Immunol. *17*, 1875–1880.

50. Khan, S., Nakajima, R., Jain, A., Romero de Assis, R., Jasinskas, A., Obiero, J.M., Adenaiye, O., Tai, S., Hong, F., Milton, D.K., et al. (2020). Analysis of Serologic Cross-Reactivity Between Common Human Coronaviruses and SARS-CoV-2 Using Coronavirus Antigen Microarray. bioRxiv. https://doi.org/10.1101/2020.03.24.006544.

51. World Health Organization (2020). COVID-19 – landscape of novel coronavirus candidate vaccine development worldwide. https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines.

52. Woolhouse, M.E.J., and Brierley, L. (2018). Epidemiological characteristics of human-infective RNA viruses. Sci. Data *5*, 180017.

53. Jia, N., Liu, H.-B., Ni, X.-B., Bell-Sakyi, L., Zheng, Y.-C., Song, J.-L., Li, J., Jiang, B.-G., Wang, Q., Sun, Y., et al. (2019). Emergence of human infection with Jingmen tick virus in China: a retrospective study. EBioMedicine *43*, 317–324.

54. Shiryaev, S.A., Thomsen, E.R., Cieplak, P., Chudin, E., Cheltsov, A.V., Chee, M.S., Kozlov, I.A., and Strongin, A.Y. (2012). New details of HCV NS3/4A proteinase functionality revealed by a high-throughput cleavage assay. PLoS ONE *7*, e35759.

55. Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veesler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell *181*, 281–292.e6.

CellPress
OPEN ACCESS

56. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. *46* (*W1*), W296–W303.

57. Fink, Z.W., Martinez, V., Altin, J., and Ladner, J.T. (2020). PepSIRF: a flexible and comprehensive tool for the analysis of data from highly-multiplexed DNA-barcoded peptide assays. arXiv, 2007.05050. http://arxiv.org/abs/2007.05050v1.

58. Mina, M.J., Kula, T., Leng, Y., Li, M., de Vries, R.D., Knip, M., Siljander, H., Rewers, M., Choy, D.F., Wilson, M.S., et al. (2019). Measles virus infection diminishes preexisting antibodies that offer protection from other pathogens. Science *366*, 599–606.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
4 NTPs (ATP, CTP, GTP, UTP)	Lucigen	Cat# RN02825
Protease Inhibitor Cocktail	VWR	Cat# G6521
DYNAL MyOne Dynabeads Streptavidin T1	Thermo Fisher Scientific	Cat# 65604D
AcTEV Protease	Thermo Fisher Scientific	Cat# 12575015
RNase Inhibitor, Murine	New England Biolabs	Cat# M0314L
ProtoScript II Reverse Transcriptase	New England Biolabs	Cat# M0368X
RNase H	New England Biolabs	Cat# M02971
Q5 High-Fidelity DNA Polymerase	New England Biolabs	Cat# M0491L
T4 RNA Ligase 2	New England Biolabs	Cat# M0239L
SARS-CoV-2 FP Peptide: PSKRSFIEDLLFNKVTLADA	Sigma-Aldrich	N/A
SARS-CoV-2 HR2 Peptide: LQPELDSFKEELDKYFKNHT	Sigma-Aldrich	N/A
Dynabeads Protein G for Immunoprecipitation	Thermo Fisher Scientific	Cat# 10004D
Critical commercial assays		
Ampliscribe T7-Flash Transcription Kit	Lucigen	Cat# ASF3507
PURExpress *In Vitro* Protein Synthesis Kit	New England Biolabs	Cat# E6800L
Deposited data		
PepSeq read count data	Open Science Framework	https://osf.io/ak2tm/
Oligonucleotides		
Custom 244,000 and 7,500 oligo libraries, CCTATACTTCCAAGGCGCAxxxxxxxxxx xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx xxxxxxxxGGTGACTCTCTGTCTTGGC	Agilent Technologies	G7227A, G7220A
DNA amplification primer (Forward): GCGA ATTAATACGACTCACTATAGGGCTTAAGT ATAAGGAGGAAAAAATATGGGAGAAAAC CTATACTTCCAAGGCGCA	Integrated DNA Technologies	N/A
DNA amplification primer (Reverse): GCTC CTGCTGCATTTCCGTTCAGCAGACGCAG CAGCCAAGACAGAGAGTCACC	Integrated DNA Technologies	N/A
Biotinylated oligo for bead purification: TTTTTCATATTTTTTCCTCCTTATACTT AAGCCC	Integrated DNA Technologies	N/A
Software and algorithms		
Peptide design algorithm	GitHub	https://github.com/LadnerLab/Library-Design
PepSIRF (version 1.3.2)	GitHub	https://github.com/LadnerLab/PepSIRF
Custom Python scripts	GitHub	https://github.com/LadnerLab/PepSIRF/tree/master/extensions
R version 4.0.2	R project	https://www.r-project.org/
PyMol 2.3.2	Schrodinger LLC	https://pymol.org/2/
Other		
NextSeq 500/500 Mid Output Kit v2.5 (150 Cycle)	Illumina	Cat# 20024904

CellPress
OPEN ACCESS

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, John Altin (jaltin@tgen.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The raw peptide count data from this study have been deposited in the Open Science Framework (https://osf.io/ak2tm/), DOI: 10.17605/OSF.IO/AK2TM. All custom code is available via GitHub (https://github.com/LadnerLab).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The age, gender, COVID-19 RT-PCR status, and collection site and date of the serum/plasma donors studied herein are described in Table S1. COVID-19 convalescent serum and plasma samples (n = 55) were collected at sites in California, USA (Vitalant Research Institute and City of Hope National Medical Center) and Norway (St. Olav's Hospital, Trondheim and Oslo University Hospital, Oslo) from patients who had tested positive for SARS-CoV-2 by RT-PCR a median of 30 days prior. Post-pandemic negative control samples (n = 5) were also collected from healthy donors at two of these sites (City of Hope National Medical Center and St. Olav's Hospital). Use of post-pandemic samples was determined to be not human subjects research by TGen's Research Compliance office. Pre-pandemic negative control serum samples (n = 64) were sourced from three locations: Creative Testing Solutions (Phoenix, AZ), Walter Reed National Military Medical Center (Bethesda, MD) and TGen (Flagstaff, AZ). The samples from Creative Testing Solutions were collected during January 2015 from multiple locations in California. Samples from Walter Reed National Military Medical Center were collected during 2019; exact dates were not provided, but the latest collections were during the first week of December. The sample from TGen was collected in November 2018. The use of all pre-pandemic samples was reviewed by the NAU and TGen Research Compliance offices and determined not to be human subjects research.

METHOD DETAILS

PepSeq library design

We designed two different libraries of peptides in order to assess antibody reactivity to SARS-CoV-2 peptides and to peptides from other human-infecting CoVs. For our human virome ('HV') peptide library, we sought to include sequences from all viruses known to infect humans. For viruses with RNA genomes, we obtained a list of 214 virus species from Woolhouse and Brierley.[52] NCBI taxonomy IDs were obtained for each of these species using the "names.dmp" file from the NCBI "new_taxdump" downloaded on November 19, 2018 [note: "Bovine viral diarrhea virus 1" (NCBI:txid11099) was replaced with the corresponding species, "Pestivirus A" (NCBI:txid2170080)]. Taxonomy IDs for human viruses with DNA genomes were obtained using the "host.dmp," "nodes.dmp" and "fullnamelineage.dmp" files from the NCBI "new_taxdump" downloaded on November 26, 2018. In total, we identified 289 taxonomy IDs annotated as virus species with DNA genomes that are known to cause human infections; however, 31 of these were excluded from our design because they clearly belonged to unclassified adenovirus strains, rather than distinct virus species. Finally, we included two taxonomy IDs associated with the Jingmenvirus group, members of which have recently been associated with human infections in China.[53]

On November 19, 2018, we downloaded all viral protein sequences from the UniProt Knowledgebase ("uniprot_sprot_viruses.dat" and "uniprot_trembl_viruses.dat" from https://ftp://ftp.uniprot.org/pub/databases/uniprot/current_release/knowledgebase/taxonomic_divisions/) and extracted the sequences annotated with one of our 474 target species taxonomy IDs. NCBI BLAST was used to identify sequences with non-viral components (i.e., recombinant), specifically those containing common reporter and therapeutic proteins: ubiquitin, luciferase, green fluorescent protein, chloramphenicol acetyltransferase, LacZ, GusA and GusB. These sequences were excluded from the assay design. To identify taxonomically misclassified proteins, we downloaded all of the proteins annotated in the NCBI RefSeq database for our target species, when available (342/474 target species IDs). We then used NCBI BLAST to identify the best matching RefSeq protein for each UniProt protein, and flagged instances when the top hit was "strong" and to a RefSeq protein from a different genus ($\geq$80% nt identity) or species ($\geq$95% nt identity). All of the flagged UniProt proteins were manually investigated, including an additional BLAST to the NCBI nt database, and sequences confirmed to be misclassified were either removed completely or taxonomically relabeled. Finally, we removed all sequences < 30 amino acids in length and collapsed identical sequences to a single representative using a custom python script (https://github.com/LadnerLab/Library-Design/blob/master/scripts/one_hundred_rep.py).

Following our length, identity and taxonomy filters, we were left with 1,300,994 target protein sequences assigned to 443 distinct species-level taxonomy IDs. However, a small subset of viral species contributed the vast majority of protein sequences. For example, 49% of the proteins were from human immunodeficiency virus 1 and 16% were from influenza A virus. To ensure more

even representation of viruses within our assay design, we randomly subsampled the overrepresented species, including no more than 2000 and 4000 protein sequences for viruses with RNA and DNA genomes, respectively. Additional protein sequences were allowed for DNA viruses because they often contain larger genomes and proteomes (i.e., more distinct genes). When down-sampling, priority was given to proteins from the Swiss-Prot database, which have been manually reviewed. Our final down-sampled target set included 148,215 proteins and 88.78 M amino acids.

In order to optimize potential epitope coverage in as few peptides as possible, we utilized a greedy set cover algorithm in which all potential linear epitopes contained within our target sequences were treated as our "elements of interest" and "sets" were defined as the collection of all potential epitopes contained within a potential peptide probe. Each round, a score was calculated for each potential peptide probe, which corresponded to the sum of the frequencies of each contained epitope within the full target set of proteins, and the highest scoring peptide was added to our design. In the event of a tie, a peptide was randomly chosen from the highest scoring subset. All of the potential epitopes contained within the added peptide were then excluded from the calculation of scores in the next round. This procedure was repeated until a targeted proportion of total epitope diversity was contained within the selected peptides. This algorithm was implemented with custom software (https://github.com/LadnerLab/Library-Design). For our design, we focused on optimizing 9-mer (i.e., 9 amino acids long) epitope coverage using 30-mer peptides.

To reduce the runtime and memory requirements of the algorithm, we partitioned our target protein sequences according to taxonomy prior to running our peptide design algorithm. We generated subsets of our target proteins by first dividing according to viral family and finally by genus, if the family-level partition contained > 500,000 unique 9mers. Due to the random nature of peptide selection in the event of a tie, our algorithm is not deterministic. Therefore, we independently ran the design for each partition 5-20 times (depending on the size of the partition), and in each case, we selected the result with the fewest number of chosen peptides.

For a subset of species with low numbers of UniProt sequences per annotated protein, we added unique protein sequences present in GenBank to our list of targets. Additionally, for these species and one other with low overall epitope coverage in our set cover design (severe fever with thrombocytopenia syndrome virus, taxID = 1933190), we redesigned peptides using a sequence-level (i.e., no alignment) sliding window approach (step = 19) in order to optimize epitope coverage. We also included 15 "positive control" peptides, which included epitopes known to be broadly reactive in the human population based on preliminary, unpublished data, and 223 "negative control" peptides designed from an assortment of eukaryotic proteins of exotic species (e.g., coelacanth, coral, great white shark).

In total, this HV design included 244,000 unique 30-mer peptides, and represents approximately 70% of all potential 9-mer epitopes contained within the target protein sequences. Each of these peptides was represented by a single nucleotide encoding. This design does not contain any peptides derived from SARS-CoV-2, but does contain full proteome coverage of the other six CoVs known to infect humans: Human coronavirus 229E (NCBI taxID: 11137), Human coronavirus NL63 (277944), Human coronavirus HKU1 (290028), Betacoronavirus 1 (694003, includes Human coronavirus OC43), Severe acute respiratory syndrome-related coronavirus (694009, SARS), and Middle East respiratory syndrome-related coronavirus (1335626, MERS).

Our second design (SCV2) focused almost entirely on SARS-CoV-2, including high density tiling of peptides across the two most immunogenic SARS-CoV-2 proteins: the spike glycoprotein (S) and the nucleocapsid protein (N). As targets for this design, we utilized 2303 SARS-CoV-2 genome sequences downloaded from GISAID on April 3rd, 2020 (Table S3), along with six locally generated sequences. Using these genomes, we first generated consensus amino acid sequences for the S and N proteins. In our design, we included all of the unique 30-mer peptides contained in these consensus sequences, equivalent to a 1-step sliding window approach.[54] Additionally, we used the same epitope-centric set cover design algorithm used for HV in order to capture amino acid-level polymorphisms present within our full set of target genomes. This aspect of the design ensured that 100% of the unique 16-mer peptides present in the S and N proteins from the 2309 SARS-CoV-2 genomes were represented in our design. In total, this design included 1550 30-mer peptides from the S protein and 557 30-mer peptides from the N protein. Each of these peptides was represented by three different nucleotide encodings. This design also included a set of 373 control peptides. These controls represent a subset of the HV peptides, which we have determined are commonly recognized by IgG antibodies in human sera (unpublished results). Therefore, we expect that some fraction of these controls will be recognized by antibodies in each blood sample tested. Collectively, these peptides were designed from 55 different virus species, including the four endemic human CoVs.

PepSeq library synthesis and assay

Libraries of covalently-coupled peptide:DNA conjugates were prepared from pools of DNA oligonucleotide templates in bulk enzymatic reactions. Pools of ssDNA templates (*Agilent*) were PCR-amplified and the dsDNA products were used as templates for *in vitro* transcription (*Ampliscribe*). The resulting mRNA was ligated to a hairpin oligonucleotide adaptor bearing a puromycin molecule tethered by a PEG spacer and, following buffer exchange, the reaction mix was used as a template in an *in vitro* translation reaction (*PURExpress, NEB*). Constructs bearing mRNA – comprising of (i) mRNA, (ii) mRNA+adaptor, (iii) mRNA+adaptor+peptide – were captured using magnetic beads coated with a biotinylated DNA oligo complementary to a 30 nt sequence in the mRNA constant region. A reverse transcription reaction, primed by the adaptor hairpin, was used to generate cDNA, after which RNase was applied to remove mRNA. Product was buffer-exchanged, quantified by running on a gel against standard DNA oligos of known concentrations, and used without further modifications or purification.

To perform serological assays, 5uL of a 1:10 dilution of serum/plasma in Superblock T20 (*Thermo*) was added to 0.1pmol of PepSeq library for a total volume of 10uL and was incubated at 20°C overnight. The binding reaction was applied to pre-washed protein G-bearing beads (*Thermo*) for 15 minutes, after which beads were washed 10 times with 1x PBST. After the final wash, beads were

resuspended in 30uL of water and heated to 95°C for 5 minutes to elute bound product. Elutions were amplified and indexed using barcoded DNA oligos. Following PCR cleanup, products were pooled, quantified and sequenced on a NextSeq instrument (*Illumina*).

In selected cases, subsets of antibodies were depleted from serum/plasma samples using bead-bound peptides prior to the assay. Chemically synthesized N-terminally biotinylated 20-mer peptides (*Sigma*) with the SARS-CoV-2 sequences (FP: PSKRSFIEDLLFNKVTLADA, or HR2: LQPELDSFKEELDKYFKNHT) were incubated at 3.33 ug/mL in wash buffer (5 mM Tris HCL pH 7.5, 0.5 mM EDTA, 2M NaCl) with 30uL of pre-washed Streptavidin beads (*Thermo*) for 15 minutes and then washed 3 times with wash buffer and re-suspended in superblock. 30uL of serum/plasma was added to 30 uL of buffer aspirated peptide-coated beads and incubated for 15 minutes. Serum was aspirated from the beads re-applied to a fresh peptide-coated bead set and repeated, for a total of 3 depletion cycles, prior to use in the assay.

Visualization of protein structure

To visualize our identified SARS-CoV-2 epitopes within the 3D conformational structure of the S protein, we utilized the cryo-electron microscopy (Cryo-EM) structure available in the RCSB Protein Data Bank (PDB id: 6VYB).[55] To compare epitope positions across CoV species, we built three additional structures using Cryo-EM templates from PDB: 5SZS for HCoV-NL63, 6ACD for SARS-CoV and 6NZK for HCoV-OC43. We performed structural modeling using Swiss-Model software.[56] Structural alignments and image preparation were done with PyMOL (version 2.3.2, Schrodinger, LLC). To build models of the post-fusion state for S2 subunit fragments, we used the Cryo-EM structure for mouse hepatitis virus, determined by Walls et al. (PDB id: 6B3O).[41]

QUANTIFICATION AND STATISTICAL ANALYSIS

We used PepSIRF v1.3.2,[57] along with custom scripts (https://github.com/LadnerLab/PepSIRF/tree/master/extensions), to analyze the PepSeq high-throughput sequencing data. The data analysis included three primary steps: 1) demultiplexing and assignment of reads to peptides, 2) calculation of enrichment *Z* scores individually for each assay and peptide and 3) identification of enriched peptides for each sample based on the consistency of *Z* scores and fold-change across replicates.

Demultiplexing and assignment of reads to peptides was done using the *demux* module of PepSIRF,[57] allowing up to 1 mismatch within each of the index sequences and up to 2 mismatches with the expected DNA tag (90 nt in length). *Z* scores were calculated using a method adapted from Mina et al.[58] This process involved the generation of peptide bins, each of which contained $\geq$ 300 peptides with similar starting abundance in our PepSeq assay. Starting abundance for each peptide was estimated using buffer-only controls. In total, 8-13 independent buffer-only controls were used to generate the bins for this study. The raw read counts from each of these controls were first normalized to reads per million (RPM) using the column sum normalization method in the *norm* module of PepSIRF. This was to ensure that independent assays were weighted evenly, regardless of differences in the depth of sequencing. Bins were then generated using the *bin* PepSIRF module.

Z scores were calculated using the *zscore* PepSIRF module, and each *Z* score corresponds to the number of standard deviations away from the mean, with the mean and standard deviation calculated independently for the peptides from each bin. It is important that the mean and standard deviation reflect the distribution of unenriched peptides within a bin. Therefore, these calculations were based on the 75% and 95% highest density interval of read counts within each bin for the SCV2 and HV libraries, respectively. Prior to *Z* score calculation, RPM counts for each peptide were further normalized by subtracting the average RPM count observed within our buffer-only controls. This second normalization step controlled for variability in peptide starting abundance within a bin. Finally, the *p_enrich* module of PepSIRF was used to determine which peptides had been enriched through our assay. This module identifies peptides that meet or exceed minimum thresholds, in both replicates. For the SCV2 library, we used a minimum *Z* score threshold of 8 along with a minimum RPM fold-change of 4. For the HV library, we required a minimum RPM count of 10, a minimum RPM fold-change of 4 and we used a 2-tier *Z* score threshold, with one replicate needing a *Z* score $\geq$ 10 and both replicates needing a *Z* score $\geq$ 6. All of these thresholds were selected to minimize the number of false positive determinations of peptide enrichment based on the analysis of buffer-only negative controls. For both the SCV2 and HV libraries, we examined a range of thresholds using four buffer-only negative controls (analyzed as 6 pairs of replicates), none of which were considered in the creation of the bins. In both cases, the chosen thresholds resulted in only a single enriched peptide being called in 1/6 of the analyzed pairs.

Minimally reactive regions for each epitope were identified as the linear peptide sequences shared by all enriched peptides across convalescent donors. To compare the relative level of reactivity at each epitope to SARS-CoV, HCoV-OC43/Beta1 and HCoV-229E (Figure 4C), we first identified all of the peptides, designed from each of these species, that overlapped the minimally reactive epitope regions. For each epitope, we then identified the maximum *Z* score for each sample/species pair and then normalized these values by dividing by the sum of the three species-specific *Z* scores, with negative *Z* scores converted to 0 prior to normalization.

Logistic regression was performed using the *glm* function in R using log-transformed *Z* scores for each of the 6 focal epitopes (peptide with maximum *Z* score for each epitope was used) as features to predict convalescent or negative donor status. Cross-validated AUC was calculated by randomly partitioning the data 100 times in 70:30 training:test sets. To quantify correlations between the patterns of reactivity to SARS-CoV-2 epitopes detected by the SCV2 library versus virome-wide peptides in the HV library, we used the *cor.test* function in R to generate all pairwise Pearson product moment comparisons based on log-transformed *Z* scores.