

ReScan, a Multiplex Diagnostic Pipeline, Pans Human Sera for SARS-CoV-2 Antigens

Colin R. Zamecnik, PhD^{1*}, Jayant V. Rajan, MD, PhD^{2*}, Kevin A. Yamauchi³, PhD, Sabrina A. Mann, BS^{3,4}, Gavin M. Sowa, MS⁵, Kelsey C. Zorn, MHS⁴, Bonny D. Alvarenga, BA¹, Mars Stone, PhD^{6,7}, Philip J. Norris, MD^{6,7,8}, Wei Gu, MD, PhD⁷, Charles Y. Chiu, MD, PhD^{7,8}, Joseph L. DeRisi, PhD^{3,4**}, Michael R. Wilson, MD, MAS^{1**,***}

Title Characters with Spaces: 70

Abstract Word Count: 98

Article Word Count: 2,673

Number of references: 38

Number of display items: 5 (4 figures, 1 table)

Supplemental items: 4 (4 tables)

Extended Data: 2 (2 figures)

*Authors contributed equally

**Authors contributed equally

Affiliations:

¹Weill Institute for Neurosciences, Department of Neurology, University of California, San Francisco, San Francisco, CA, USA

²Division of Experimental Medicine, Department of Medicine, Zuckerberg San Francisco General Hospital, University of California, San Francisco, San Francisco, CA, USA

³Chan Zuckerberg Biohub, San Francisco, CA, USA

⁴Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA

⁵UCSF School of Medicine, San Francisco, CA, USA

⁶Vitalant Research Institute, San Francisco, CA, USA

⁷Department of Laboratory Medicine, University of California, San Francisco, San Francisco, USA

⁸Department of Medicine, University of California, San Francisco, San Francisco, USA

***Corresponding Author Information:

Name: Dr. Michael Wilson

Address: UCSF, Department of Neurology, Division of Neuroimmunology and Glial Biology, 675 Nelson Rising Lane,
NS212, Campus Box 3206, San Francisco, CA 94158

Email: michael.wilson@ucsf.edu

Phone: 415-502-7429

Abstract

Serologic assays are needed to determine SARS-CoV-2 seroprevalence, but poor specificity can overestimate exposures. Here, we built a pan-human coronavirus proteome-wide programmable phage display assay (VirScan) to profile coronavirus antigens specifically enriched by 20 COVID-19 patient serum IgG. With ReScan, a new diagnostic development workflow which combines the isolation of phage expressing the most immunogenic peptides with paper-based microarrays manufactured via acoustic liquid handling, we identified 9 candidate antigens from a library of 534 SARS-CoV-2 peptides. These arrays could form the basis of a multiplexed COVID-19 serologic assay with enhanced specificity. ReScan has broad applicability for serologic assay development.

Article (2,000-4,000 words excluding abstract, online methods, refs, figure legends)

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel and deadly betacoronavirus rapidly spreading across the globe¹⁻³. Diagnostic assays are still being developed, with reagents often in short supply. Direct detection of virus in respiratory specimens with SARS-CoV-2 RT-PCR will remain the gold standard for identifying acutely infected patients⁴. While RT-PCR yields clinically actionable information and can be used to estimate incidence, it does not identify past exposure to virus, information necessary for determining population level disease risk and the degree of asymptomatic spread. These data are critical as they inform public policy about returning to normal activity at local and state levels^{5,6}.

Targeted, ELISA-based serologies to detect antibodies to the SARS-CoV-2 whole spike (S) glycoprotein, its receptor binding domain (RBD) or the nucleocapsid (N) protein have been developed and have promising performance characteristics^{4,7-9}. However, enhancing the specificity of SARS-CoV-2 serologic assays is particularly important in the context of low seroprevalence, where even a 1-2% false positive rate could greatly overestimate population level viral exposure¹⁰. Comprehensive and agnostic surveys of the antigenic profile across the entire SARS-CoV-2 proteome have the potential to identify antigens that are less cross-reactive with antibodies elicited by other common human CoV (HuCoV) infections. Multiplexed assays using antigens characterized by such surveys could also provide individualized

portraits of the adaptive immune response to help identify immunophenotypes that correlate with widely varying COVID-19 clinical outcomes^{11,12}. These tools could inform the design and evaluation of urgently needed SARS-CoV-2 vaccines and therapeutic monoclonal antibodies.

We describe here a new programmable phage display HuCoV VirScan¹³ library with overlapping 38 amino acid peptides tiled across the genomes of 9 HuCoVs, including both SARS-CoV-1 and SARS-CoV-2. We screened known COVID-19 patient sera (n=20) against this library and a previously built pan-viral VirScan library¹⁴ using phage-immunoprecipitation sequencing (PhIP-Seq)¹⁵ to identify the most highly enriched viral antigens relative to pre-pandemic controls.

To rapidly progress from broad serological profiling by VirScan to a linear epitope-based serological assay with phage expressing a focused set of highly immunogenic, disease-specific peptides in a microarray format, we used a complementary diagnostic development pipeline, ReScan. With ReScan, we use antibodies from candidate patient sera to physically pan for phage displaying immunogenic antigens. We then isolate and culture individual phage clones, followed by paper-based microarray production via acoustic liquid handling.

We applied ReScan by first immunoprecipitating a focused SARS-CoV-2 T7 phage library containing 534 overlapping 38 amino acid peptides against COVID-19 patient sera to generate a SARS-CoV-2 specific microarray. A larger cohort of positive and uninfected control patient samples was then screened to identify shared and discriminatory antigens with an automated image processing algorithm. These microarrays could serve as the basis for a low-cost, disease-specific, multiplex serologic assay whose antigens are generated by an inexhaustible and rapidly scalable reagent source (Figure 1A). The approach we describe here has broad implications for serologic diagnostic development.

Results:

Patient Cases and Controls

Twenty de-identified subjects who were found to have SARS-CoV-2 infection by clinically validated qRT-PCR by nasopharyngeal swab sampling were enrolled in this study. De-identified, pre-pandemic, adult healthy control sera (n=95) were obtained from the New York Blood Center. Almost all of the COVID-19 subjects had pneumonia (95%), and

the mean time from symptom onset to sample collection was 16.4 days (IQR 12 to 19 days). More than one-third (38%) of patients were intubated.

HuCoV VirScan Library Benchmarking

We designed and built two new VirScan libraries for SARS-CoV-2 and HuCoV, comprised of 534 and 3,670 38mer peptides respectively. Deep sequencing of each phage library was conducted to assess peptide distribution and representation relative to the original library design. For the SARS-CoV-2 and HuCoV libraries, 83.2% and 92% of sequenced phage were perfect matches to the designed oligonucleotides, respectively. Each library contained 99.8% and 100% of the designed peptide sequences at a sequencing depth of 2-2.5 million reads per library. Benchmarking data on the pan-viral VirScan library were previously described.¹⁴

VirScan Detects Coronavirus Antigens in COVID-19 Patient Sera

We identified multiple SARS-CoV-2 peptide sequences that were enriched in each of the COVID-19 subjects relative to 95 pre-pandemic healthy controls by performing PhIP-Seq on both the HuCoV and pan-viral VirScan libraries (Fig 1B-E). Overall, the S and N proteins were the most differentially enriched proteins compared to pre-pandemic control sera in both the HuCoV library and the pan-viral VirScan libraries. In the S protein, both libraries detected two 3' regions (residues 783-839 and 1124-1178). S protein residues 799-836 structurally flank the RBD, while the region where two peptides (spanning residues 1122-1159 and 1141-1178) overlap appear to be on the outer tip of the S protein trimer. Several other antigenic regions in the S protein (residues 188-232, 681-706) were also enriched in addition to N protein residues 209-265 and residues 142-208. This region of the N protein is a key piece of the N terminal RNA binding domain which aids in viral RNA assembly and packaging into the virion¹⁶. We also identified a significant antibody response to the ORF3a accessory protein (residues 171-210) with our HuCoV library.

ReScan Pans COVID-19 Sera to Generate a Focused SARS-CoV-2 Antigen Microarray

To isolate the phage clones displaying candidate antigens from the parent SARS-CoV-2 T7 phage library, we immunostained and picked plaque colonies from phage immunoprecipitated by COVID-19 patient sera. We sequenced the phage stock from each plaque pick and found that a mean sequence purity of 98% (51-100%), and a total of 31

unique peptides were represented across 364 positively staining plaques. Nine of these antigens were 1) identified in more than 4 plaques and 2) had > 80% sequence clonality, with a mean of 98.8% of reads mapping to the top sequence. The nine antigens were restricted to specific regions of the N and S proteins except for a single ORF3a peptide (residues 172-209) (Supplemental Table 1).

We repurposed an acoustic liquid handler (Labcyte Echo) to form 1cm x 2cm 384 spot microarrays of printed phage on nitrocellulose. For analysis, we employed dotblotr [<https://github.com/czbiohub/dotblotr>], a customized microarray image processing software built specifically for ReScan. After staining microarrays with COVID-19 patient (n=20) and a subset of pre-pandemic control (n=37) sera, dotblotr successfully detected 99.84% of all printed dots on the 57 arrays analyzed. No significant difference in spot detection was seen between positive and healthy groups (Supplemental Figure 2).

We found that 19/20 (95%) COVID-19 subjects had a significant number of positively stained dots on at least one of the nine antigens on the array compared to healthy controls ($p < 0.05$, Fisher's exact test) (Fig. 2). One subject (subject 16) was positive on all nine peptides. A single patient (subject 19) did not have any significant hits by ReScan but was subsequently noted to be one of two in the cohort of positive patients on chronic immunosuppression for a solid organ transplant.

Peptides spanning residues 799-836 and 1141-1178 on the S protein were the most broadly reactive antigen targets, called as significant on 12/20 (60%) and 14/20 (80%) of subjects, respectively. Staining on peptides spanning residues 134-171 and 153-190 of the N protein was detected across the 8/12 (66.7%) of the same patients, indicating a likely shared motif in the overlapping region between these consecutive tiled peptides. The single peptide from ORF3a (residues 171-210) was positive on 6/20 patients (30%).

Similarity Analysis of SARS-CoV-2 Antigens Identified by ReScan

Of the four S peptides isolated by ReScan, fragment 30 (residues 552-589) had 0% AA similarity, fragment 43 (residues 799-836) had 15.8% AA similarity, fragment 44 (residues 818-855) had 26.3% AA similarity, and fragment 61 (residues

1141-1178) had 5.3% AA similarity with other HuCoVs (Fig. 3). In contrast, residues 1141-1178 of the S protein were identical between SARS-CoV-2 and SARS-CoV-1, and the first 16 amino acids of this region of the S protein (residues 1141-1156) shared 56% similarity with 9 other human and non-human CoVs from which peptides spanning this region of the S protein were also significantly enriched in our patient samples relative to pre-pandemic healthy sera.

Of the four N peptides isolated, fragments 8 (residues 134-171), fragment 9 (153-190), 12 (210-247) and 20 (362-399) had 5.3%, 2.6%, 0%, and 0% AA similarity with other human CoVs. These fragments had AA similarities of 94.7%, 97.4%, 86.8% and 89.5% with SARS-CoV-1. Multiple sequence alignment of Fragment 9 (residues 153-190), the most conserved of the 4 N protein peptides we identified, demonstrated an N-terminal conserved motif found in peptides from 2 other human and 3 non-human coronavirus species (NXXAIV, 66.6% AA similarity) as well as a C-terminal motif (YAEGRGXSSRXSSRX, 73.7% AA similarity) found only in SARS-CoV-2 and 3 bat CoV species.

ReScan and PhIP-Seq Reveal Concordant Antigens targeted by COVID-19 Subject Sera

Antigens identified by ReScan were a subset of enriched antigens in the PhIP-Seq analysis (Fig 4). 18/20 (90%) patients had at least one peptide fragment identified by ReScan called as significant by PhIP-Seq (threshold $p < 0.01$). The most commonly observed fragments by ReScan mapped to the S protein (residues 1141-1178) and the N protein (residues 153-190 and 210-247), each with 8 concordant hits between VirScan and ReScan.

Discussion

Diagnostic testing is a critical part of an effective public health response to a rapidly spreading pandemic. While deploying these assays is time sensitive, prematurely releasing tests that are not sufficiently sensitive or specific can stymie timely interventions and undermine confidence in government and public health institutions. Rather than *a priori* selecting one or two antigens upon which to base a serologic assay, we used a proteome-wide custom HuCoV VirScan library and COVID-19 patient sera to agnostically identify antigens most enriched relative to pre-pandemic control sera. In addition to using NGS and statistical analysis (PhIP-Seq), we developed ReScan to rapidly select and propagate the most immunogenic peptide-bearing phage clones by immunofluorescence. With this approach, the most critical antigens are identified, and the phage expressing them are simultaneously isolated and cultured, rendering them an immediate,

renewable protein source. Here, ReScan identified 9 antigens derived from 3 SARS-CoV-2 proteins as candidates for more specific, multiplexed SARS-CoV-2 serologic assays.

ReScan has several key advantages over traditional peptide arrays for antigen screening. Diversity of the original library is not restricted by physical space on the array, and the cost of peptide synthesis is decreased by at least two orders of magnitude as the input library is generated via an oligonucleotide pool cloned *en masse* into a phage backbone. Through immunoprecipitation and phage amplification, IgG antibodies from patients with a disease of interest (e.g., COVID-19) select a specific subset of peptides for use in downstream diagnostic assays. This approach is an improvement over standalone phage immunoprecipitations, which can errantly amplify clones due to nonspecific binding of phage to beads and protein-protein interactions outside of antibody-antigen events. These nonspecific binding events can pose significant challenges for the interpretation of PhIP-Seq datasets¹⁷. ReScan circumvents these problems by using plaque lift immunoblotting¹⁸, a well characterized technique that verifies a particular phage colony was generated based on a true antibody-mediated interaction, to select for antigens that also perform well in the context of an immunostaining assay.

Typically, peptides can be challenging to spot or print on microarrays due to their widely variable physical properties and solubility, necessitating bespoke solvent systems and resuspension procedures¹⁹. By hosting these peptides on a stable T7 bacteriophage carrier, treatment of every clone is standardized, with each species undergoing high-throughput, plate-based amplification. This standardization, combined with acoustic liquid handling, permits the generation of protein microarrays through a flexible and rapid printing process with low variability between arrays.

We found that COVID-19 serum IgG antibodies were predominantly directed against epitopes in the S and N proteins. This is not surprising given the rich literature on the sensitivity and specificity of ELISAs using these proteins for SARS-CoV-1²⁰⁻²⁴. While the majority of the focus thus far on serology development for SARS-CoV-2, has been on the S protein^{4,7,25,26}, our data suggest that N protein-based ELISAs should be pursued in parallel⁸. Because of its high degree of glycosylation, peptides from the S protein displayed on phage cannot capture all possible epitope diversity although two of the S protein peptides we detected with VirScan and ReScan (residues 783-839 and residues 1124-1178) do contain N-

linked glycosylated sites in the native S protein²⁷. We identified three critical regions that were concordant between both ReScan and VirScan. All three S protein fragments are on the protein's surface and appear to be solvent accessible based on the 6VSB crystal structure²⁸.

In addition to our identification of the most enriched N protein peptides in the region of the N-terminal RNA binding domain, previous work with SARS-CoV-1 peptides has shown C-terminal dimerization regions to be particularly immunogenic²⁹. This region, spanning peptides 362-399, was also a significant target in both Rescan and HuCoV VirScan methods, though less frequent (seen in 15% and 30% of patients, respectively).

The only additional epitope identified by ReScan outside the N and S proteins was a single region mapping to the ORF3a protein between positions 172-209. ORF3a plays several important roles in viral trafficking, release and fitness³⁰, and this particular region spanning the di-acidic central domain facilitates export from the host cell endoplasmic reticulum to the Golgi apparatus³¹. It was also identified as an antibody target in SARS-CoV-1-infected patients^{32,33}. While only 6/20 patients had significant hits to this region by ReScan, it is a unique accessory protein to the SARS-like family, and therefore may represent an attractive candidate to enhance specificity of SARS-CoV-2 diagnostic assays.

Our study has some limitations. To fully characterize the test performance characteristics of these 9 candidate antigens as the basis for a more specific, high-throughput clinical diagnostic assay, additional patient samples need to be analyzed. However, both the statistically significant enrichment of these antigens above pre-pandemic healthy controls by PhIP-Seq, and the rigorous comparison between staining on HCs and COVID-19 subjects to determine a positive signal on the ReScan microarrays are both supportive of these antigens being highly specific for SARS-CoV-2 exposure. The current IP strategy focuses only on IgG, but future work may center on IgM targets, given their role in the early humoral immune response. Finally, post-translational modifications and tertiary or quaternary conformational epitopes are not represented in the T7 phage display system used here^{15,34}.

The immediate focus of this study was to develop a blueprint for candidate antigens to include in confirmatory SARS-CoV-2 serologic assays and identify a panel of linear antigens with significant discriminatory power. Beyond their use for

diagnostic purposes, the next and critical question is whether a patient's antibody repertoire confers protective immunity. Because our assay characterizes patients' antibodies to multiple SARS-CoV-2 antigens, it enables deeper phenotyping of the adaptive immune response that, in future studies, can be correlated with acute and long-term clinical outcomes.

Acknowledgments

The authors would like to thank Dr. Nicole Repina for graphic design assistance with Figure 1, as well as the entire Chan Zuckerberg Biohub Sequencing Team for sequencing support. The authors would also like to acknowledge Dr. Michael Busch and Vitalant Research Institute for additional serum samples and controls as well as the New York Blood Center for pre-pandemic control sera. We thank Andrew Kung and Ravi Dandekar for their help using Chimera. C.R.Z. and J.V.R. wish to acknowledge Dr. Caleigh Mandel-Brehm for immunoblotting protocols and extensive discussion on PhIP-Seq analysis. We thank Dr. Stephen Hauser, the Weill Institute for Neurosciences, Mr. David Friedberg, Mr. and Mrs. Rachleff, and the Sandler and Bowes Foundations for their support. We thank the patients and their families for participating in this study.

Author Contributions

J.V.R. computationally designed and cloned the VirScan peptide libraries. S.A.M., B.D.A., C.R.Z. performed the VirScan experiments. C.R.Z., G.M.S. and B.D.A. performed ReScan experiments. C.R.Z. and S.A.M. performed library preparation for sequencing of phage libraries. K.A.Y. wrote the code for dotblotr and assisted C.R.Z. with assay development and analysis for ReScan. K.C.Z., W.G., C.Y.C., M.S. and P.J.N. identified patients, performed clinical phenotyping and provided patient samples. J.V.R., C.R.Z., M.R.W. and J.L.D. analyzed VirScan and ReScan data. C.R.Z., J.V.R. J.L.D., and M.R.W. conceived of and wrote the manuscript. All authors discussed the results and contributed critical review to the manuscript.

Competing Interests Statement

The authors have no competing interests to declare.

Disclaimer

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the National Institutes of Health.

References

1. Zhou, P., *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270-273 (2020).
2. Wu, F., *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265-269 (2020).
3. Zhu, N., *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* **382**, 727-733 (2020).
4. Wolfel, R., *et al.* Virological assessment of hospitalized patients with COVID-2019. *Nature* (2020).
5. Petherick, A. Developing antibody tests for SARS-CoV-2. *Lancet* **395**, 1101-1102 (2020).
6. Lipsitch, M., Swerdlow, D.L. & Finelli, L. Defining the Epidemiology of Covid-19 - Studies Needed. *N Engl J Med* **382**, 1194-1196 (2020).
7. Stadlbauer, D., *et al.* SARS-CoV-2 Seroconversion in Humans: A Detailed Protocol for a Serological Assay, Antigen Production, and Test Setup. *Curr Protoc Microbiol* **57**, e100 (2020).
8. Qu, J., *et al.* Profile of IgG and IgM antibodies against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Clin Infect Dis* (2020).
9. Liu, W., *et al.* Evaluation of Nucleocapsid and Spike Protein-based ELISAs for detecting antibodies against SARS-CoV-2. *J Clin Microbiol* (2020).
10. Jackson, B.R. The dangers of false-positive and false-negative test results: false-positive results as a function of pretest probability. *Clin Lab Med* **28**, 305-319, vii (2008).
11. Guan, W.J., *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* (2020).
12. Bhatraju, P.K., *et al.* Covid-19 in Critically Ill Patients in the Seattle Region - Case Series. *N Engl J Med* (2020).
13. Xu, G.J., *et al.* Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome. *Science* **348**, aaa0698 (2015).
14. Schubert, R.D., *et al.* Pan-viral serology implicates enteroviruses in acute flaccid myelitis. *Nat Med* **25**, 1748-1752 (2019).
15. Larman, H.B., *et al.* Autoantigen discovery with a synthetic human peptidome. *Nature biotechnology* **29**, 535-541 (2011).
16. Chang, C.K., Hou, M.H., Chang, C.F., Hsiao, C.D. & Huang, T.H. The SARS coronavirus nucleocapsid protein--forms and functions. *Antiviral Res* **103**, 39-50 (2014).
17. Yuan, T., Mohan D, Laserson U, Ruczinski I, Baer AN, Larman HB. Improved Analysis of Phage ImmunoPrecipitation Sequencing (PhIP-Seq) Data Using a Z-score Algorithm. *bioRxiv* (2018).
18. Tsui, P., Sweet, R.W., Sathe, G. & Rosenberg, M. An efficient phage plaque screen for the random mutational analysis of the interaction of HIV-1 gp120 with human CD4. *The Journal of biological chemistry* **267**, 9361-9367 (1992).
19. Szymczak, L.C., Kuo, H.Y. & Mrksich, M. Peptide Arrays: Development and Application. *Anal Chem* **90**, 266-282 (2018).
20. Meyer, B., Drosten, C. & Muller, M.A. Serological assays for emerging coronaviruses: challenges and pitfalls. *Virus Res* **194**, 175-183 (2014).

21. Chan, P.K., *et al.* Evaluation of a peptide-based enzyme immunoassay for anti-SARS coronavirus IgG antibody. *J Med Virol* **74**, 517-520 (2004).
22. Woo, P.C., *et al.* Differential sensitivities of severe acute respiratory syndrome (SARS) coronavirus spike polypeptide enzyme-linked immunosorbent assay (ELISA) and SARS coronavirus nucleocapsid protein ELISA for serodiagnosis of SARS coronavirus pneumonia. *J Clin Microbiol* **43**, 3054-3058 (2005).
23. Zhao, J., *et al.* Comparison of immunoglobulin G responses to the spike and nucleocapsid proteins of severe acute respiratory syndrome (SARS) coronavirus in patients with SARS. *Clinical and vaccine immunology : CVI* **14**, 839-846 (2007).
24. Zhong, X., *et al.* B-cell responses in patients who have recovered from severe acute respiratory syndrome target a dominant site in the S2 domain of the surface spike glycoprotein. *Journal of virology* **79**, 3401-3408 (2005).
25. Okba, N.M.A., *et al.* Severe Acute Respiratory Syndrome Coronavirus 2-Specific Antibody Responses in Coronavirus Disease 2019 Patients. *Emerg Infect Dis* **26**(2020).
26. Haveri, A., *et al.* Serological and molecular findings during SARS-CoV-2 infection: the first case study in Finland, January to February 2020. *Euro Surveill* **25**(2020).
27. Watanabe, Y.A.J.D., Wrapp D., McLellan J.S., Crispin M. Site-specific glycan analysis of the SARS-CoV-2 spike. *Science* (2020).
28. Wrapp, D., *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263 (2020).
29. He, Q., *et al.* Development of a Western blot assay for detection of antibodies against coronavirus causing severe acute respiratory syndrome. *Clin Diagn Lab Immunol* **11**, 417-422 (2004).
30. de Haan, C.A., Masters, P.S., Shen, X., Weiss, S. & Rottier, P.J. The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host. *Virology* **296**, 177-189 (2002).
31. McBride, R. & Fielding, B.C. The role of severe acute respiratory syndrome (SARS)-coronavirus accessory proteins in virus pathogenesis. *Viruses* **4**, 2902-2923 (2012).
32. Qiu, M., *et al.* Antibody responses to individual proteins of SARS coronavirus and their neutralization activities. *Microbes Infect* **7**, 882-889 (2005).
33. Tan, Y.J., *et al.* Profiles of antibody responses against severe acute respiratory syndrome coronavirus recombinant proteins and their potential use as diagnostic markers. *Clin Diagn Lab Immunol* **11**, 362-371 (2004).
34. Zhu, J., *et al.* Protein interaction discovery using parallel analysis of translated ORFs (PLATO). *Nature biotechnology* **31**, 331-334 (2013).
35. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359 (2012).
36. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614-620 (2014).
37. O'Donovan, B., Mandel-Brehm, C, Vazquez SE, et al. Exploration of Anti-Yo and Anti-Hu paraneoplastic neurological disorders by PhIP-Seq reveals a highly restricted pattern of antibody epitopes. *Biorxiv* (2018).
38. Mandel-Brehm, C., *et al.* Kelch-like Protein 11 Antibodies in Seminoma-Associated Paraneoplastic Encephalitis. *N Engl J Med* **381**, 47-54 (2019).

Table 1. COVID-19 Patient Demographics

	COVID-19 Cases
N	20
Age – mean (IQR), years	61 (50 to 75)
Sex – no. (%)	
Female	4 (20)
Male	16 (80)
Days from symptom onset – mean (IQR)	16..4 (12 to 19.25)
Immunosuppressed – no. (%)	4 (22)
Pneumonia – no. (%)	17 (94)
Intubated – no. (%)	9 (50)
Died – no. (%)	1 (5.6)

Online Methods:

Patient Cohort: We obtained 20 de-identified, acute serum samples from SARS-CoV-2-infected patients in addition to 95 sera from adult, pre-pandemic controls from the New York Blood Bank. Detailed methods for data collection, human subjects review, VirScan, ReScan, and bioinformatics are provided in the Supplemental Appendix.

HuCoV Virscan library design

Refseq sequences were downloaded for SARS-CoV-2 (NC_045512), SARS1 (NC_004718) and 7 other coronaviruses known to infect humans, including beta coronavirus England 1 (NC_038294), HuCoV 229E (NC_002645), HuCoV HKU1 (NC_006577), HuCoV NL63 (NC_005831), HuCoV OC43 (NC_006213), infectious bronchitis virus (NC_001451), and MERS CoV (NC_019843) from the National Center of Biotechnology Information (NCBI). For each of virus, each viral protein

was divided into a sequence of 38mer peptides, with consecutive peptides sharing a 19 amino acid overlap. Peptide amino acid sequences were converted to DNA sequences and a 21 nucleotide 5' linker sequence as well as a nucleic acid sequence encoding a FLAG tag (DYKDDDDK) at the 3' end was added. All final oligonucleotides were 159 nucleotides in length. Oligos were synthesized as a single pool by Twist Biosciences, followed by PCR amplification, restriction enzyme digestion and cloning into a commercially available T7 vector as previously described^{13,14}.

VirScan Bioinformatics

Sequencing reads were aligned to a reference database comprising the full viral peptide library using the Bowtie2 aligner³⁵. All aligned sequences were translated and the frequencies of the aligned sequences that were perfect matches to the input peptide library were determined using an R language script. Counts were determined for each peptide for each individual. Peptide counts were normalized for read depth by dividing each individual peptide count by the total number of aligned peptides and multiplying by 100,000 (reads per hundred thousand (rpK))³⁵⁻³⁸. For both VirScan libraries, the null distribution of each peptide's rpK was modeled using a set of 95 healthy control sera. From these null distributions, p-values were calculated for the observed peptide rpKs, and multiple hypothesis corrected using the Benjamini-Hochberg method. All peptides with a corrected p-value of < 0.01 were considered significantly enriched over the healthy background for that specific peptide. Regions targeted by host antibodies were identified by aligning all significant peptides to concatenated SARS-CoV-2 open reading frames (ORF) and taking the summation of enrichment relative to the healthy background at each position across the SARS-CoV-2 genome. For both HuCoV and pan-viral VirScan libraries, only alignments that spanned at least 38 amino acids were considered.

ReScan protocol

Preparation of Phage Stocks and Microarray Generation

Two rounds of panning and amplification were performed with a SARS-CoV-2 T7 phage library and antibodies from 7 unique COVID-19 subject sera, where each amplified lysate was then subjected to a standard top agar plaque assay. Plaque lift immunoblots were generated by lifting plates with 200-300 plaques using 0.45µm nitrocellulose and stained with the same patient sera to verify antibody-phage interaction. Brightly stained plaques (n=364) alongside negative

controls (n=18) were picked into 2mL deep 96-well plates and propagated twice in BLT5403 *E. Coli*, spun and filtered to use as clonal antigen stocks.

After pooling the phage stocks into a 384 well ECHO-qualified source plate, each corresponding stock was acoustically printed using a Labcyte ECHO acoustic array liquid handler at 25nL per dot to form a 1cm x 2cm microarray with an individual spot size of approximately 1mm in diameter. Arrays were then incubated overnight at 4°C with patient serum and a rabbit anti-T7 phage imaging and antigen control at 1:10,000 and 1:2000 final concentration, respectively. After species-specific secondary staining, blots were imaged on a LICOR Odyssey CLX scanner. Source plate wells were sequenced by NGS to determine clonality and group staining patterns across the array (see Supplemental Methods). Each phage source well was assigned a peptide by its most commonly represented reference peptide in the NGS data.

Image Processing and Hit Calling

Images were processed by a dotblotr, a custom image processing algorithm [<https://github.com/czbiohub/dotblotr>]. To determine which dots were positive, the magnitude of the dots' normalized signal (human IgG signal intensity / anti-T7 tag signal intensity) was compared to the distribution of normalized signal from the negative controls (phage expressing GFAP or Tubulin 1a1). We calculated the mean and standard deviation of the negative control dots on the same strip and called dots 'positive' with a normalized signal at least 3 standard deviations above the mean of the negative control dots.

Since a given peptide was represented by multiple dots on each strip, all dots corresponding to the same peptide were grouped for hit calling. To determine if a given strip, s , was positive for a given peptide, p , the counts of positive and negative dots for peptide p in strip s were compared to the counts of positive and negative dots for peptide p in all healthy control strips using Fisher's exact test. If peptide p was more likely to yield positive dots in strip s than in the healthy controls with $p < 0.05$, we called strip s positive for peptide p .

Dotblotr References:

[dotblotr] <https://github.com/czbiohub/dotblotr>

[numpy] <http://dx.doi.org/10.1109/MCSE.2011.37>

[skimage] <https://doi.org/10.7717/peerj.453>

[scipy] <https://doi.org/10.1038/s41592-019-0686-2>

[pandas] <https://doi.org/10.5281/zenodo.3509134>

[matplotlib] <https://doi.org/10.1109/MCSE.2007.55>

[opencv] <https://github.com/skvark/opencv-python>

Data Availability Statement

VirScan and ReScan data analyzed in the manuscript have been made available at https://github.com/UCSF-Wilson-Lab/sars-cov-2_ReScan_VirScan_complete_analysis

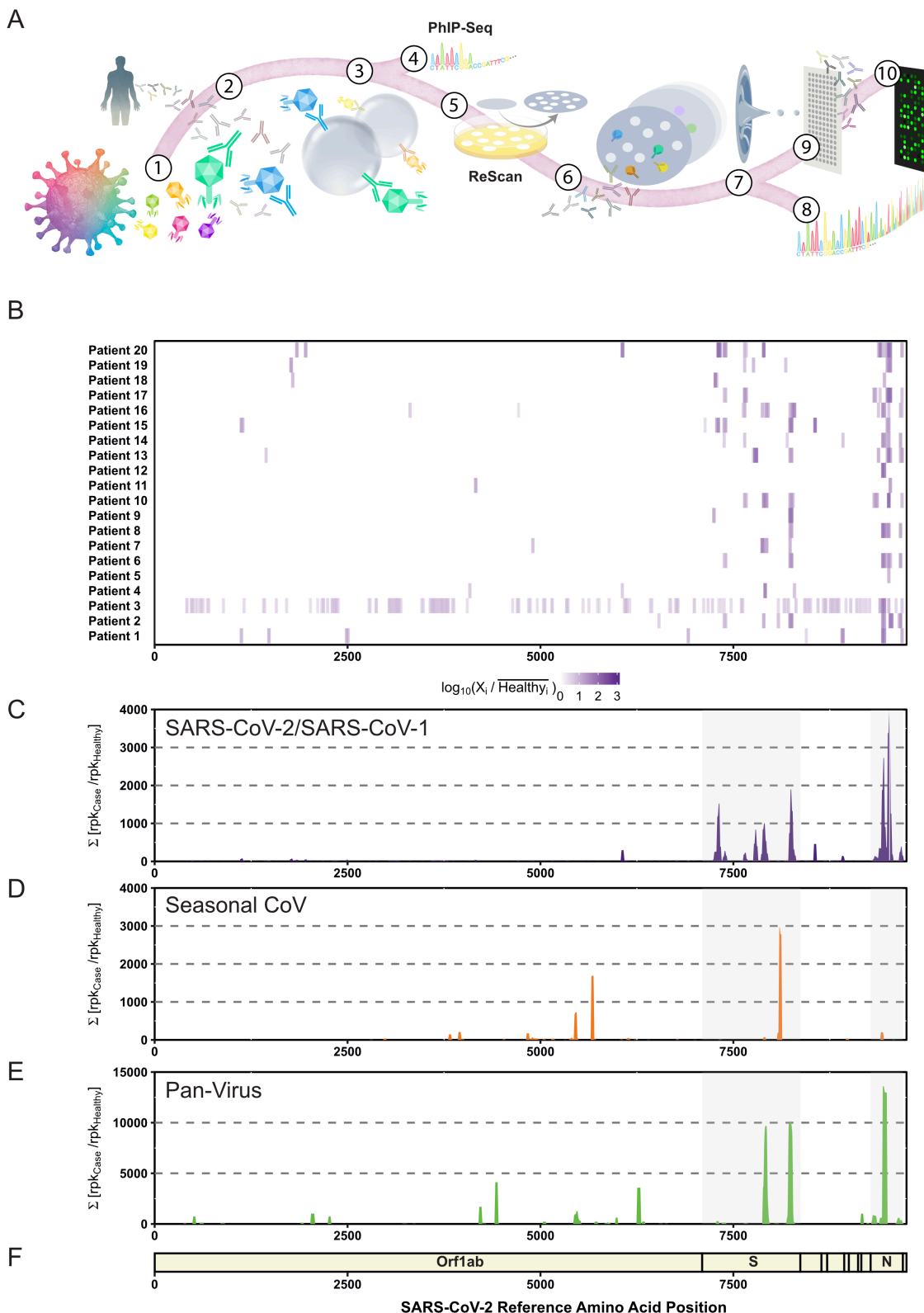


Fig 1. General Workflow for ReScan and Epitope Mapping SARS-CoV-2 using PhIP-Seq. (A) VirScan T7 phage display system with SARS-CoV-2 antigens in overlapping 38aa fragments (1) are incubated with antibodies (2) from patient sera and undergo two rounds of amplification (3). Amplified lysates are either sequenced for phage immunoprecipitation sequencing (PhIP-Seq), or are lifted, stained and isolated for ReScan (4,5). Microarrays of isolated clonal phage populations are printed via acoustic liquid transfer (6), stained with patient sera and analyzed for identity (7, 8) and commonality (9) via dotblot. (B) Heat map displaying results from individual patients and (C) cumulative fold enrichment of significant ($p < 0.01$) SARS-CoV peptides from HuCoV library relative to pre-pandemic controls aligned to SARS-CoV-2 genome. (D) Cumulative fold enrichment of significant other seasonal CoV peptides in the HuCoV library relative to pre-pandemic controls aligned to SARS-CoV-2 genome. (E) Cumulative enrichment of significant pan-viral library peptides relative to pre-pandemic controls aligned to SARS-CoV-2 genome.

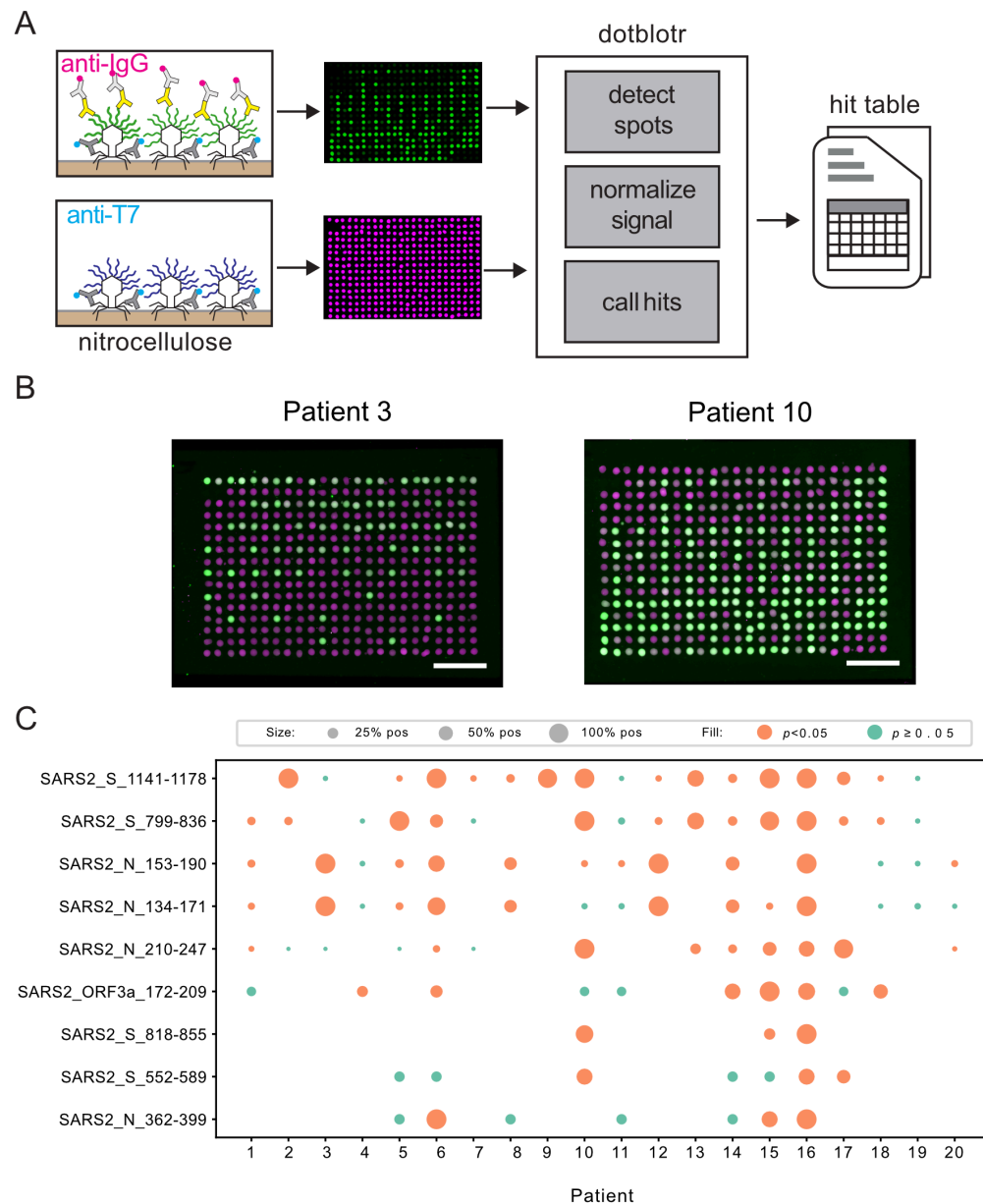


Fig 2. ReScan Analysis using SARS-CoV-2 specific phage display library. (A) Schematic of microarray design. (B) Examples of staining patterns on patients positive for nucleocapsid (Patient 3) and spike peptides (Patient 10), scale bar = 5mm. Anti-T7 tag signal is magenta, anti-IgG signal is green. (C) SARS-CoV-2 peptide analysis from ReScan microarrays; dot size is proportional to the fraction of dots for a given peptide that stained positive by a given patient sample.

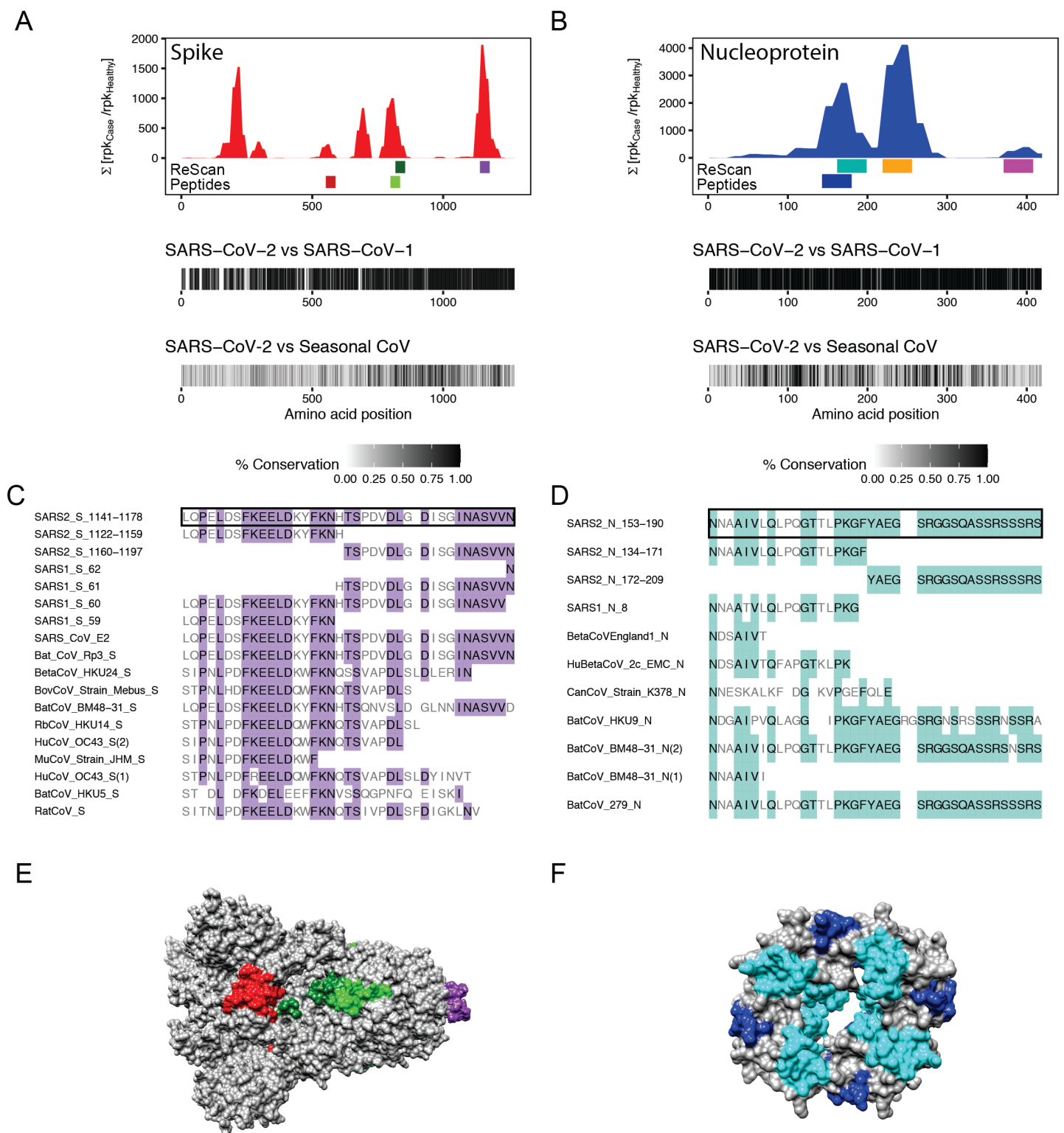


Fig 3. Similarity of Enriched ReScan and VirScan SARS-CoV-2 Peptides to other CoVs. Comparing underlined SARS-CoV-2 spike (A) and nucleocapsid (B) regions found by ReScan with other CoV sequences that were enriched over healthy controls in the HuCoV library (S₅₅₂₋₅₈₉ = red, S₇₉₉₋₈₃₆ = light green, S₈₁₈₋₈₅₅ = dark green, S₁₁₄₁₋₁₁₇₈ = purple; N₁₃₄₋₁₇₁ = blue, N₁₅₃₋₁₉₀ = teal, N₂₁₀₋₂₄₇ = orange, N₃₆₂₋₃₉₉ = pink). Below: homology maps with SARS-CoV-1 and seasonal coronaviruses in the HuCoV libraries. (C,D) Diverse CoV peptide sequences with partial similarity to the regions spanning (1141-1178) and (153-190) of the S and N proteins that were enriched in the both HuCoV and pan-viral libraries. (G,H) Crystal structures with color-coded ReScan peptides overlaid on whole S and N proteins.

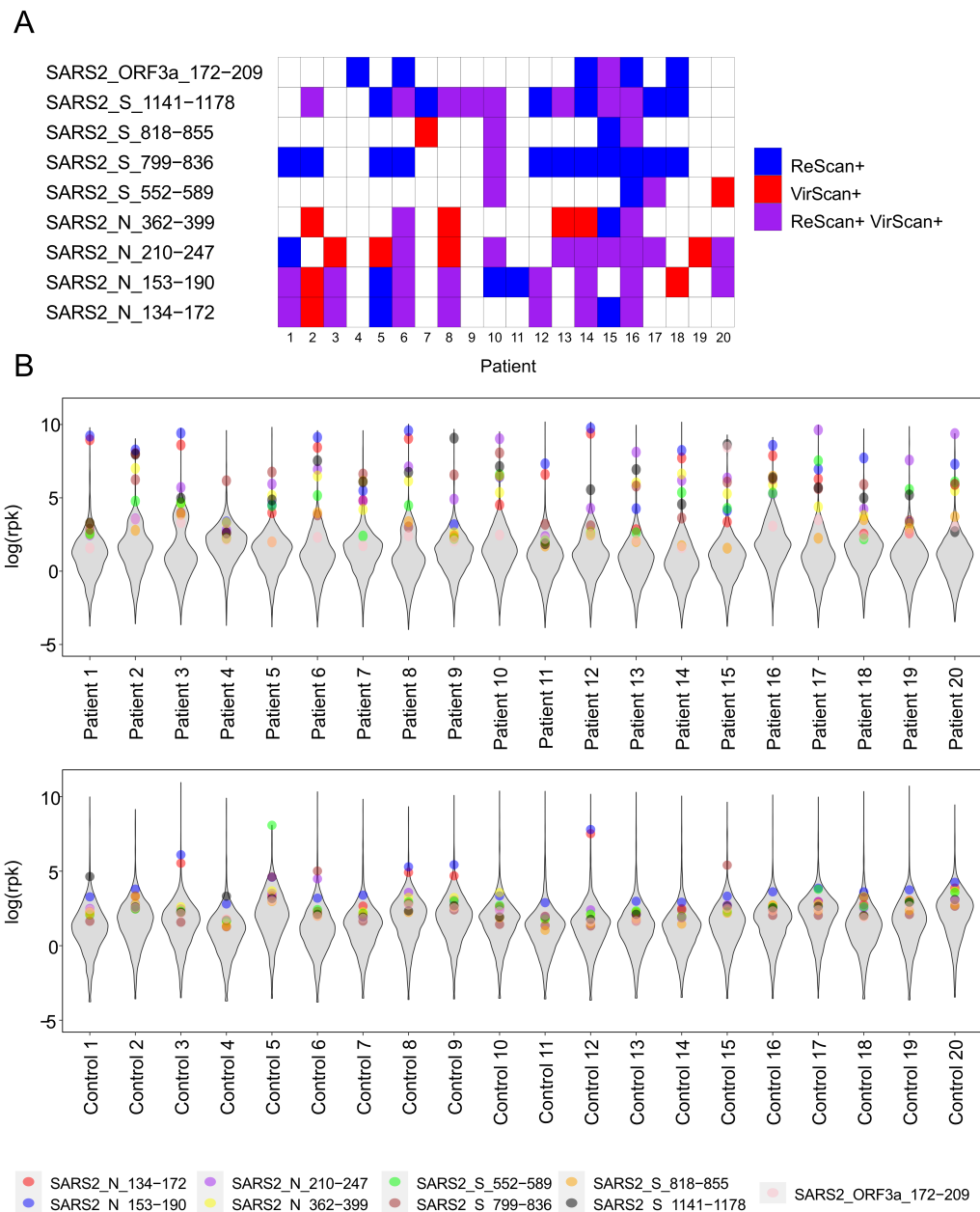
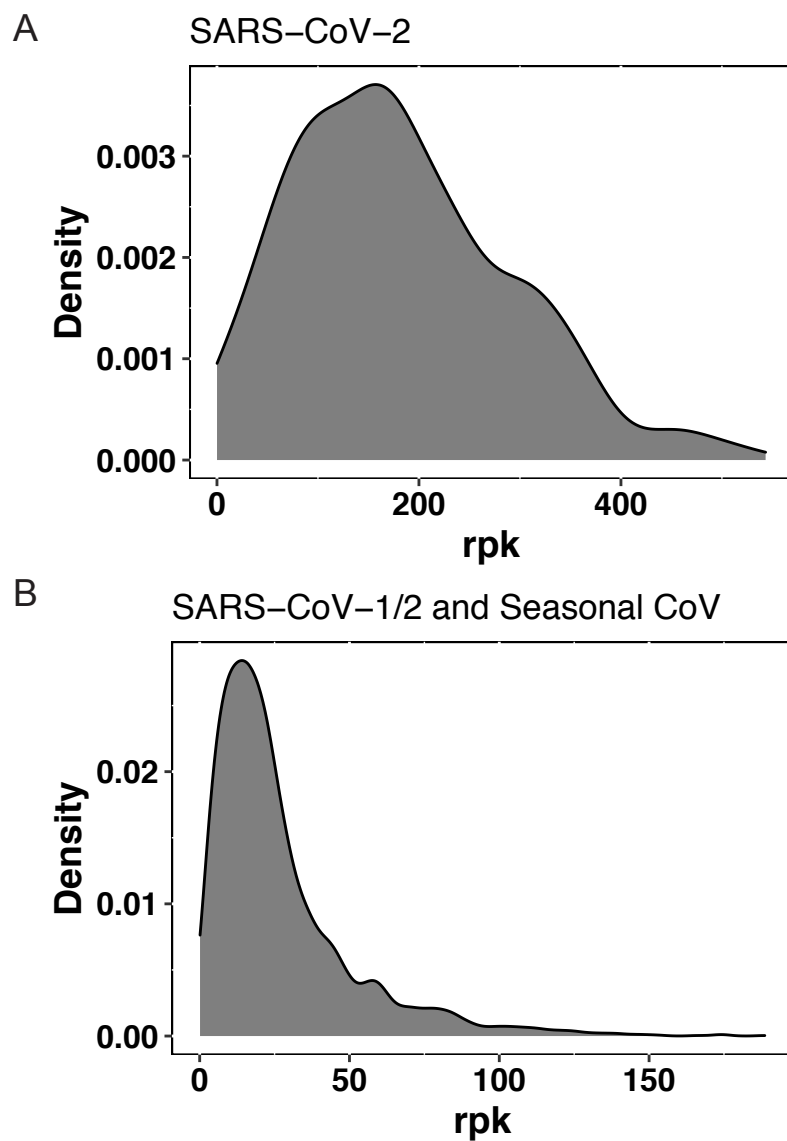
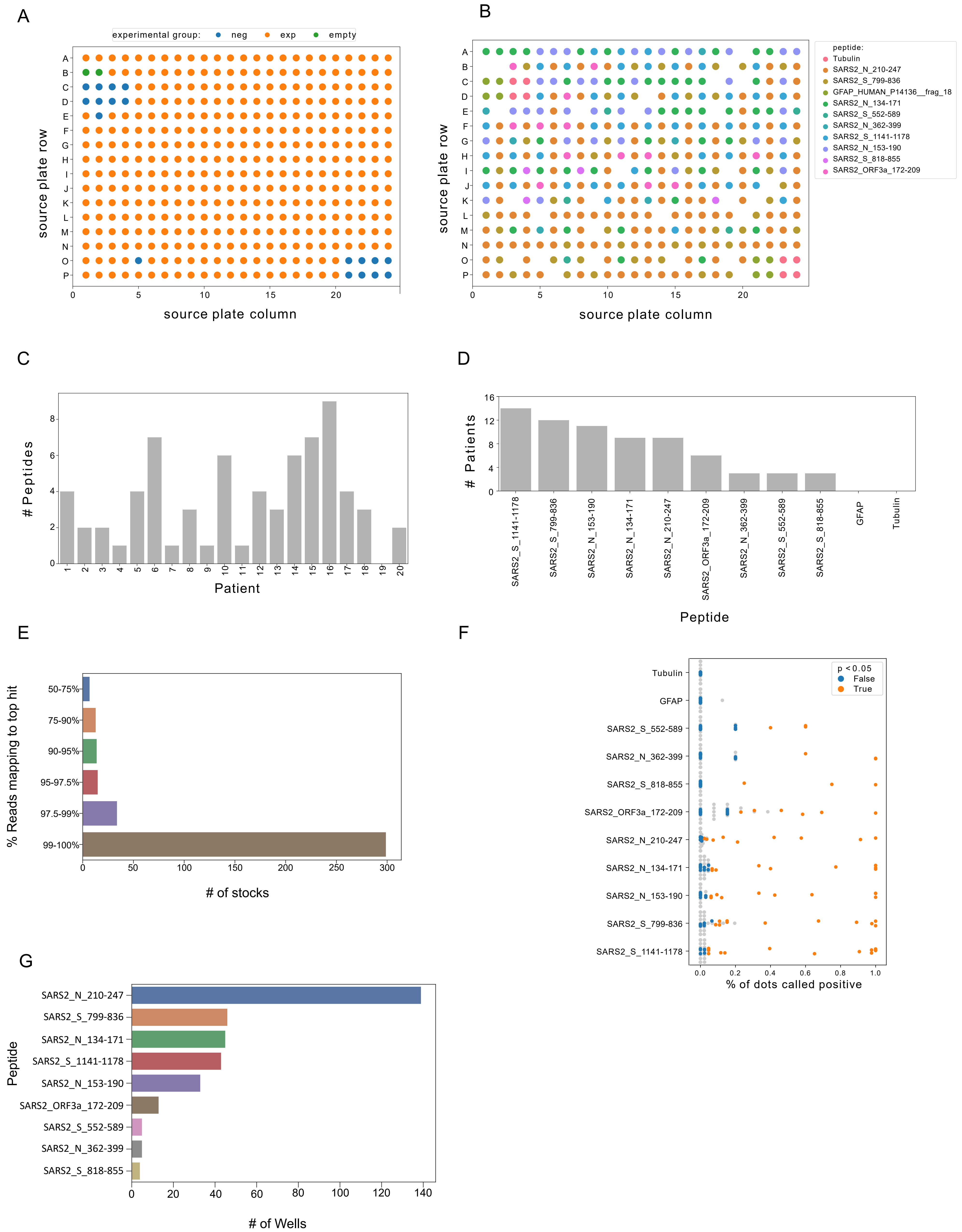


Fig. 4. Concordance between ReScan and VirScan. (A) Heat map of each patient represented across the nine antigenic SARS-CoV-2 peptides isolated by ReScan. ReScan hits were generated from a SARS-CoV-2 specific phage library and VirScan data were derived from the HuCoV library. (B) Violin plots overlaying where each of the nine ReScan peptides map to peptides enrichment distributions from the HuCoV VirScan experiments for each patient. First twenty healthy pre-pandemic controls shown for reference.



Extended Data Figure 1. (A) Distribution of peptides by reads per 100,000 reads (rpk) in the SARS-CoV-2 library used to identify antigens for ReScan. (B) Distribution of peptide rpk in the combined SARS-CoV-2, SARS-CoV-1, and Seasonal CoV library used for VirScan.



Extended Data Figure 2. (A) ReScan assay configuration layout. (B) Map of identity of each clonal phage population on plate. (C) Number of peptide hits for each patient on ReScan and (D) number of times each peptide was called as significant across the COVID-19 cohort (n=20, p<0.05, Fisher's exact test comparing positive signal to healthy distribution). (E) Monoclonality of all wells on ReScan array. (F) Fraction of dots that are positive for each peptide on all strips from COVID-19 positive sera. Strips that are significantly enriched for positive dots over healthy control strips ($p < 0.05$ by Fisher's exact test) are colored orange and strips that were not significantly enriched are colored blue. All healthy control strips are plotted in gray. (G) Number of wells passing quality filters that were mapped to each peptide identity.

SUPPLEMENTAL METHODS

Patient Enrollment and Data Collection

Inclusion criteria for patients with a confirmed diagnosis of COVID-19 were 1) age > 18 years and 2) a positive SARS-CoV-2 RT-PCR from a nasopharyngeal swab performed in a clinical laboratory. Patient 1 was enrolled in a research study at UCSF (IRB number 13-12236). Patient 2 was enrolled in a research study at UCSF (IRB number 10-00476). Patients 3-6 and 9-20 were enrolled in a research study at UCSF (IRB number 18-25287). Patients 7 and 8 were enrolled in a research study at the Blood Systems Research Institute. Because patient samples were received de-identified, the UCSF Institutional Review Board did not require further institutional review. Electronic and paper medical records of patients 2 and 3 from UCSF were reviewed for demographic details, clinical data and laboratory results, and outcome at last follow-up. Patients 3-6, 7, 8, and 9-20 were de-identified with only limited clinical and demographic data available. Their ages were rounded to the nearest mid-decade (e.g., 45 years-old, 55 years-old, etc). The 95 de-identified healthy control sera were obtained from adults who donated blood to the New York Blood Bank before the SARS-CoV-2 pandemic. Sera were transported on dry ice to our laboratory at UCSF and stored immediately upon receipt at -80°C. Upon thawing, aliquots were diluted 1:1 into a glycerol containing storage buffer as described below. There is incomplete information on the handling of all specimens analyzed between the time they were obtained from the patient and ultimately sent to UCSF as these samples were collected by clinical labs. Since the COVID-19 cases and uninfected control samples were all collected from hospital clinical labs it is unlikely there was differential handling procedures between the cases and controls in this study.

Coronavirus library design and cloning

RefSeq sequences for SARS-CoV-2 (NC_045512), SARS-CoV-1 (NC_004718) and 7 other coronaviruses known to infect humans, including beta coronavirus England 1 (NC_038294), HuCoV 229E (NC_002645), HuCoV HKU1 (NC_006577), HuCoV NL63 (NC_005831), HuCoV OC43 (NC_006213), Infectious Bronchitis virus (NC_001451), and MERS CoV (NC_019843) were downloaded from the National Center of Biotechnology Information (NCBI). All open reading frames for each virus were divided into sequences of 38mer peptides, with consecutive peptides sharing a 19 amino acid overlap. All peptide sequences were collapsed on 99% amino acid sequence identity using cd-hit. After collapsing on sequence identity, peptide amino acid sequences were converted to DNA sequences using an R language script, randomizing codon selection. Twenty-one (21) nucleotide 5' linker sequences as well as a nucleic acid sequence encoding a FLAG tag (DYKDDDDK) at the 3' end of each oligonucleotide sequence were added. The following 5' linker sequences were: GCTGTTGCAGTTGTTGGCGCT (SARS-CoV-1 and SARS-CoV-2), GTAGCTGGTGTGTTAGCTGCC (SARS-CoV-1 only), GCAGGAGTAGCTGGTGTGTTGTG (SARS-CoV-2 only), and AGCCATCCGCAGTTCGAGAAA (betacoronavirus England 1, 229E, HKU1, NL63, OC43, Infectious Bronchitis virus, MERS) for a combination SARS-CoV-1/SARS-CoV-2 library, an exclusive SARS1-CoV-1 library, an exclusive SARS-CoV-2 library, and a library spanning other coronaviruses known to cause human infections. The final oligonucleotide sequences were 159 nucleotides in length, output to a FASTA file and sent to Twist Biosciences for synthesis.

A single vial of 10pmoles of lyophilized DNA was received from Twist. The lyophilized oligonucleotides were resuspended in TE to a final concentration of 0.2nM and PCR amplified for

cloning into T7 phage vector arms (Novagen/EMD Millipore, Inc). The following amplification primers and cycling conditions were used:

Forward:

TAGTTAAGCGGAATTCAGCTGTTGCAGTTGTTGGCGCT (SARS-CoV-1/SARS-CoV-2)

TAGTTAAGCGGAATTCAGTAGCTGGTGTGTAGCTGCC (SARS-CoV-1)

TAGTTAAGCGGAATTCAGCAGGAGTAGCTGGTGTGTG (SARS-CoV-2)

TAGTTAAGCGGAATTCAAGCCATCCGCAGTTCGAGAAA (Other Coronaviruses)

Reverse:

ATCCTGAGCTAAGCTTTTACTTATCATCATCATCCTTGTAATCACC

Cycling conditions:

STEP	TEMP	TIME
Initial Denaturation	98°C	30 seconds

20 Cycles	98°C	5 seconds
	70°C	15 seconds
	72°C	10 seconds
Final Extension	72°C	2 minutes
Hold	10°C	-

After amplification, PCR reactions were bead cleaned using Ampure XP beads at 1:1 ratio per the manufacturer's instructions. Bead cleaned PCR products were quantified by using Qubit reagent (Life Technologies, Inc) and 1µg of PCR product was digested with 1µl of EcoRI-HF (NEB) and 1µl of HindIII-HF (NEB) with 1X CutSmart buffer (NEB) in a total volume of 50µl for 1h at 37°C and heat inactivated at 65°C for 10 minutes. Digested DNA was cleaned again using Ampure XP beads and cutting efficiency confirmed by running both the uncut and restriction digested PCR products on a high sensitivity dsDNA Bioanalyzer chip (Applied Biosystems, Inc.).

Cut PCR amplified oligonucleotide pools were ligated into pre-cut T7 vector arms in a 5µl volume at a ratio of 3:1 (0.06 pmol of insert and 0.03 pmol of T7 vector arms) with T4 DNA ligase (NEB) at 16°C for 12h, after which ligations were heat inactivated for 15 minutes at 65°C. A total of 2µl of the ligation reaction was mixed with 12µl of T7 packaging extract, stirring with a the pipet tip and incubating at room temperature for 2h. At the end of two hours, 150µl of cold LB +

Carbenicillin was added to the packaging reaction and it was stored at 4°C. Ten-fold serial dilutions of the raw packaging reaction were done, followed by plaque assays to determine the titer the number of infectious phage in the packaging reaction.

Preparation of phage libraries

Phage libraries were prepared and amplified fresh from packaging reactions. To prepare phage libraries, a 300 mL culture of *E. coli* BLT5403 was incubated at 37°C with shaking until mid-log phase, defined as OD₆₀₀= 0.4. The culture was then inoculated at a multiplicity of infection (MOI) of 0.001 and lysed at 37°C for 2.5 hours. 5M NaCl was added to the lysate on ice for 5 minutes at a 1:10 volume ratio to complete lysis. Lysate was then spun at 4000 g for 15 minutes to remove cellular debris. Phage-containing supernatant was then precipitated with PEG/NaCl (PEG-8000 20%, NaCl 2.5 mM) overnight at 4°C. Precipitated phage were then pelleted for 15 minutes at 3220 g at 4°C and resuspended in storage media (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 6 mM MgCl₂) before 0.22µm filtration. Resulting phage libraries were titered by plaque assay and adjusted to a working concentration of 10¹⁰-10¹¹ pfu/mL before incubation with patient sera.

Immunoprecipitation of phage-bound patient antibodies

96-well, 2mL deep well plates (Genesee Scientific) were incubated with blocking buffer of 3% BSA in 1xTBST overnight to prevent nonspecific binding to plate walls. Blocking buffer was then replaced with 500µL of either library type (HuCoV or pan-viral) and 1µL of human sera in 1:1 storage buffer (0.04% NaN₃, 40% Glycerol, 40mM HEPES to preserve antibody integrity), or 1µg of commercial antibody for a positive immunoprecipitation control. To facilitate antibody

phage binding, the deep well plates with library and sample were incubated overnight at 4°C on a rocker platform in secondary containment. Pierce Protein A/G Beads (ThermoFisher Scientific) were aliquoted 20µL of A and 20µL of G per reaction and were washed 3 times in TNP-40 (140mM NaCl, 10mM Tris-HCL, 0.1% NP-40). After the final wash, beads were resuspended in TNP-40 in half the volume (20uL) and added to the phage-patient antibody mixture and incubated on the rocker at 4°C for 1 hour. Phage bound to antibody bound to bead was then washed five times with RIPA (140mM NaCl, 10mM Tris-HCl, 1% Triton X, 0.1% SDS) using a P1000 multichannel pipet to remove supernatant and add RIPA each time, with each wash incubating the bound A/G beads for 5 minutes with gentle rocking at 4°C. After the fifth wash, immunoprecipitated solution was resuspended in 150µL of LB-Carb and then added to 500µL of log phase *E. coli* for amplification ($OD_{600}=0.4$). After amplification, 5M NaCl was added to lysed *E. coli* to a final concentration of 0.5M NaCl to ensure complete lysis. The lysed solution was spun at 3220 rcf for 15 minutes and the top 500µL was filtered through a 0.2µM PVDF filter plates (Arctic White, #AWFP-F20080) to remove cell debris. Filtered solution was transferred to a new deep well pre-blocked plate and patient sera was added and subjected to another round of immunoprecipitation and amplification. The final lysate was spun, filtered and stored at 4°C for NGS library prep and subsequent ReScan plaque lifts and downstream colony selection.

NGS Library Prep:

All liquid handling steps were performed using the Labcyte Echo 525 and the Integra Via Flow 3520. Immunoprecipitated phage lysate was diluted 1:10 and heated to 70°C for 15 minutes to expose DNA. DNA was then thermocycled in two subsequent reactions, the first of which amplified the unique, programmed peptide insert of the phage genome. The second served to the

add on custom 8-12nt TruSeq sample barcode primers as well as adapter sequences for Illumina flow cell. Libraries were then pooled at equal volumes, followed by cleaning and size selection using Ampure XP magnetic beads at 0.9X ratio according to manufacturer instructions. Eluted libraries were quantified by High Sensitivity DNA Qubit and qualified by High Sensitivity DNA Bioanalyzer after dilution to approximately 1ng/μL. Sequencing was then performed on a NovaSeq S1 (300 cycle kit with 1.3 billion clusters) averaging 5 million reads per sample for HuCoV and pan-viral libraries, and 500,000 reads for SARS-CoV-2 specific libraries.

Reagent	Volume (uL) per reaction	Stock concentration
Nuclease-free water	4.125	--
Phusion HF Buffer (NEB)	2.5	5x
10 mM dNTPs	0.25	10mM
10 uM Forward Primer	0.25	10uM
10 uM Reverse Primer	0.25	10uM
Polymerase	0.125	1U/uL
Phage Lysis	5	Diluted 1:10

STEP	TEMP	TIME
Initial Denaturation	98°C	30 seconds
16 Cycles	98°C	5 seconds
	70°C	20 seconds
	72°C	15 seconds
Final Extension	72°C	2 minutes
Hold	10°C	-

Reagent	Volume (µL) per reaction	Stock concentration
Nuclease-free water	8.125	--
Phusion HF Buffer (NEB)	2.5	5x
10 mM dNTPs	0.25	10mM

Barcode Forward i5/Reverse i7 mixed Primers*	1	5mM
Polymerase	0.125	1U/uL
PCR 1	0.5	1x

STEP	TEMP	TIME
Initial Denaturation	98°C	30 seconds
5 Cycles	98°C	5 seconds
	70°C	20 seconds
	72°C	15 seconds
Final Extension	72°C	2 minutes
Hold	10°C	-

Primers for DNA Amplification:

SARS1/2 Forward_0N	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCTGTTGCAGTTGTTGGC GCT
SARS1/2_Fwd _1N	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNGCTGTTGCAGTTGTTGG CGCT
SARS1/2_Fwd _2N	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNGCTGTTGCAGTTGTTG GCGCT
SARS1/2_Fwd _3N	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNGCTGTTGCAGTTGTT GGCGCT
SARS2_Fwd_0 N	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCAGGAGTAGCTGGTGT TGTG
SARS2_Fwd_1 N	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNGCAGGAGTAGCTGGTG TTGTG
SARS2_Fwd_2 N	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNGCAGGAGTAGCTGGT GTTGTG
SARS2_Fwd_3 N	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNGCAGGAGTAGCTGG TGTTGTG
sCoV_Fwd_0N	ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGCCATCCGCAGTTCGA GAAA
sCov_Fwd_1N	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNAGCCATCCGCAGTTCG AGAAA

sCov_Fwd_2N	ACACTCTTTCCTACACGACGCTCTTCCGATCTNNAGCCATCCGCAGTTC GAGAAA
sCov_Fwd_3N	ACACTCTTTCCTACACGACGCTCTTCCGATCTNNNAGCCATCCGCAGTT CGAGAAA
FLAG_Rev_0N	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTATCATCATCATCCTTG TAGTCACC
FLAG_Rev_1N	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNATCATCATCATCCTT GTAGTCACC
FLAG_Rev_2N	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNATCATCATCATCCT TG TAGTCACC
FLAG_Rev_3N	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNATCATCATCATCC TTGTAGTCACC

Bioinformatic Analysis of PhIP-Seq Data

Sequencing reads were aligned to a reference database of the full viral peptide library using the Bowtie2 aligner, outputting a SAM file. All SAM format files were converted to BAM using the *samtools* view command. For each aligned sequence, the CIGAR string was examined, and all alignments where the CIGAR string did not indicate a perfect match were removed. The filtered set of aligned sequences was then translated and which translated peptides were perfect matches to the input library was determined. Any peptides that were not perfect matches were excluded from further analysis. The final set of peptides was tabulated to generate counts for each peptide

in each individual sample. All of this analysis was automated using an R language script in RStudio.

Peptides counts were normalized for read depth by dividing them by the total number of peptides in the sample and multiplying by 100,000, resulting in a reads/100,000 reads for each peptide.²⁴⁻²⁷ For all VirScan libraries, the null distribution of each peptide's $\log_{10}(\text{rpK})$ was modeled using a set of 95 healthy control sera. All counts were augmented by 1 to avoid zero counts in the healthy control sera samples. Multiple distribution fits were examined for these data, including Poisson, Negative Binomial and Normal distributions. Quantile-quantile plots for each distribution demonstrated that the Normal distribution had the best fit across all peptides, as measured by its median linear correlation coefficient. These null distributions were used to calculate p values for the observed $\log_{10}(\text{rpK})$ of each peptide within a given sample. The calculated p-values were corrected for multiple hypothesis using the Benjamini-Hochberg method. Any peptide with a corrected p value of < 0.01 was considered significantly enriched over the healthy background. For peptides where the number of unique values was less than 10 across the healthy controls, and for which it was thus impossible to estimate a Normal null distribution because of sparsity of data, they were combined into a single distribution and this common distribution was used to calculate p-values. As before, these p-values were corrected for multiple hypothesis testing and a corrected p value of < 0.01 was used as a significance threshold.

To identify regions targeted by host antibodies, all library peptides were aligned to the SARS-CoV-2 reference genome, concatenating all of its open reading frames (ORFs) into a single sequence. This reference sequence was used to generate a blastp database, and all peptides in the

library were aligned to it using NCBI blastp. Using these data, the summation of enrichment relative to the healthy background was calculated at each position across SARS-CoV-2 for all significant peptides for each experimental sample. Finally, the results were summed at each position across all experimental samples and the summed enrichment plotted by position using the R ggplot2 library. Only alignments that were at least 38 amino acids in length (the length of the peptides in the coronavirus VirScan libraries) were included in this analysis.

ReScan Protocol

Plaque Lift Immunoblots

Immunoprecipitations were performed as previously described with a T7 phage display library containing the full-length genome of the SARS2 novel coronavirus with 38 amino acid insert length and 19 amino acid tiling. Standard top-agar plaque assays were performed on freshly filtered second round lysate from a candidate patient IP at 10^6 to 10^7 dilution in LB-Carbenicillin. Plaques were left to grow for 5-6 hours at 37°C or until approximately 2mm in diameter, then placed for 10 min at 4°C to fully set top agar. Pre-cut nitrocellulose (Protran, #10401116) was placed onto the agar plates for 1 minute to adhere phage colonies then carefully lifted and left to dry. Nitrocellulose blots were then washed with TBST for 3 minutes, blocked for 30 minutes at room temperature in TBST with 5% non-fat milk (BioRad, #1706404). The primary buffer consisted of the serum from the patient IP that the plaque assay was cast from diluted to 1:10,000 in blocking buffer, with serum being previously diluted 1:1 in storage buffer. Blots were then incubated overnight with gentle agitation at 4°C. Immunoblots were then washed 3x for 5 minutes with TBST at room temperature, then stained with anti-human IgG antibody (LICOR, #926-32232) at 1:20,000 concentration for 1h and then washed 4x in TBST for 8 minutes each. Immunoblots were then given a final wash in PBS to remove detergent for 2

minutes at room temperature, dried on Whatman 3MM paper, and immediately imaged on a LICOR Odyssey CLX imager.

Plaque Picking and Sub-Library Generation

Plaques were picked with clean 200 μ L pipette tips from corresponding agar plate using immunoblot anti-IgG staining as a guide for which phage colonies represented antibody-specific events. Approximately 40 plaques were each picked from 7 corresponding different patient plaque lifts into 2mL deep well plates into storage media (1xTBS supplemented with 5mM MgCl₂) and pipetted up and down to free phage from tip. 0.5mL of BL5403 E. Coli at an OD₆₀₀=0.4 were then added to each plate and left to lyse for 2.5h on an INFORS shaker at 37°C and 720rpm with hydrophobic breathable seal. 100 μ L of lysis were added to 700 μ L of a second propagation of BL5403 E. Coli at OD₆₀₀=0.4 to improve titer and minimize variation between wells. This was similarly lysed for 1.5h or until visually complete, then 80 μ L of 5M NaCl was added to complete lysis for 5 minutes on ice. Plates were spun at 3200 rcf for 15 minutes at 4°C to pellet debris, then filtered through 0.2 μ m PVDF filter membranes for storage at 4°C. Lysis was used within two weeks.

Acoustic Microarray Printing and Staining

60 μ L of each monoclonal phage stock was added to a unique well on an ECHO qualified source plate (Labcyte, #PPT-0200) and spun at 3200 rcf for 20 min to remove bubbles. Nitrocellulose (0.2 μ m, Whatman) was cut and adhered to a standard Framestar 384 skirted plate (4titude, #0384) as a backing frame and used as a destination plate. Arrays were printed using a Labcyte ECHO 525 instrument on the 384_PP_PLUS_BP setting at a 6,144 well pattern with 2 row

spacing between arrays, with 9 arrays per 384 well plate. 25nL from each source well was printed onto each microarray spot. Arrays were thoroughly dried for 30 minutes, then removed from backing plate, diced and placed into black 7.4cm x 2.4cm western-style boxes at 4°C until staining.

Individual arrays were wetted with 1xTBST for 5 minutes, then blocked with 1X TBST with 5% non-fat milk as previously described for 30 minutes. Primary buffer master mix consisted of cold blocking buffer with 0.02% sodium azide as a preservative, and 1:2000 concentration of anti-T7 Tag (Invitrogen, # PA132386). Each microarray was incubated with patient serum previously diluted 1:1 in storage buffer at 1:5000 dilution for a final dilution of 1:10,000. Arrays were allowed to incubate overnight at 4°C. Blots were then washed 3x for 5 minutes each in 1x TBST and immediately stained in secondary buffer consisting of anti-human IgG 800 and anti-rabbit IgG 680 (LICOR, 926-68071) at 1:20,000 dilution each for 1h under gentle agitation at room temperature. Arrays were then washed 4x in 1X TBST for 8 minutes each, with a final wash in PBS to remove detergent, then dried fully and imaged on a LICOR Odyssey CLX imager.

Miniaturized Next Generation Sequencing of Monoclonal Microarray Library

All liquid transfer steps were done with a Labcyte ECHO 525 acoustic handler. 0.25μL of phage stock was transferred from library source plate into 2.25μL of nuclease free water in a 384 well PCR plate. Phage were lysed as previously described at 70°C for 15 minutes, then the following concentrated master mix was dispensed into each well for a 6.25μL total reaction volume:

Reagents	μL/reaction
5x HF buffer	1.25

10 mM dNTPs	0.125
10mM Forward Primer	0.125
10mM Reverse Primer	0.125
Phusion HF Polymerase	0.0625
Water	2.0625
DNA template	2.5

SARS2 Library Prep Sequencing Primers same as described above.

PCR thermocycler sequence was identical to that used above for library preparation for IP sequencing described above with 13 cycles. 0.5 μ L of product from PCR#1 was used as input for PCR#2, with 2.75 μ L of the following concentrated master mix dispensed into each well:

Reagents	μ L/reaction
5x HF buffer	1.25
10 mM dNTPs	0.125
Phusion HF Polymerase	0.0625
Water	1.3125
DNA template	0.5

8-12 base pair TruSeq barcoding primers were kindly provided by the Chan Zuckerberg BioHub at 5mM concentration in water. These were diluted to 1:5 in nuclease free water and pooled into a 384 well ECHO qualified source plate. 3 μ L of each diluted primer was dispensed into the corresponding well for PCR#2 giving a unique barcode to each reaction after thermocycling protocol identical to what was used above for IP library preparation.

Once complete, the entire 6.25 μ L reaction volume from each well was pooled and purified with Ampure XP bead clean up at 0.9x ratio per manufacturer instructions, then eluted into 600 μ L total volume. Pooled and cleaned DNA was quantified by Qubit and Bioanalyzer, and sequenced on an Illumina Iseq for an average read depth of 4000 reads per well.

Dotblotr image processing

Image processing of ReScan blots

Image processing and visualization of the rescan blot data was done using dotblotr, a python package for analyzing dotblot data [dotblotr]. The processing steps are described below and the source code is available at: <https://github.com/czbiohub/dotblotr>.

First, LI-COR Image Studio was used to crop and rotate the images of the blot such that each image contained a single strip with the rows and columns of the dot array aligned with the rows and columns of pixels in the image. The images were then loaded into numpy arrays [numpy] using skimage [skimage] and the image split into the reference channel (anti-t7 tag antibody) and the probe channel (anti-human IgG antibody).

Each dot was detected using the dot signal in the reference image. Pixels corresponding to a dot were segmented using a threshold selected using otsu's algorithm and connected components labeling. [scipy] The detected spots were filtered based on circularity and size. Each detected dot that passed the quality control filters was assigned a coordinate (letters from rows, numbers for columns) by clustering. Finally, the label image where each dot was given a unique label.

Next, the dot locations were determined from the reference channels to segment the dots in the probe channel. For each dot, the normalized intensity (probe channel / reference channel) was calculated. The calculated data were then output as a results_table in a pandas DataFrame [pandas]. The resulting plots were generated using matplotlib [matplotlib] and openCV [openCV].

Analysis code python notebook, hitcalling.ipynb, is available on the data repository for this manuscript, which takes the results_table using dotblotr (specifically, using the multifile example available on the dotblotr github repo) from both healthy and positive control cohorts.

Array identification with clustering

The row and column corresponding to each dot were determined by clustering the x and y coordinates of the dot centroids, respectively. To determine the row of each dot, only the y coordinate of the centroid of each dot were considered along with the number of rows in the array based on the assay configuration. The y coordinates were then clustered with k-means where k (the number of clusters) was set to the number of rows. Each cluster then corresponds with a row and the identity of the row was determined by the order. The same process was applied for the columns by applying clustering in the same way to the x coordinates of the dot centroids. Thus, the row and column was assigned for each detected dot.

Results table schema

The results_table contains the following columns. Each row contains the results for a single dot.

The list is formatted: **column_name [data type]:** <description>.

- **assay_id [str]:** the name of the assay configuration file
- **strip_id [str]:** the unique identifier for the strip containing the dot
- **dot_name [str]:** the name of the well in the source plate the dot came from
- **source_plate_id [str]:** the name of the source plate the dot came from
- **source_plate_row [str]:** the name of the source plate row the dot came from
- **source_plate_column [str]:** the name of the source plate column the dot came from
- **exp_group [str]:** the name of the experimental group
- **zscore_threshold [float]:** the value of the z score threshold used to call a positive hit
- **top_hit [str]:** name of the most common peptide sequence in the phage source
- **top_hit_pct [str]:** fraction of total reads in the phage source corresponding to the top_hit.
- **row [str]:** row name in the strip the dot is in
- **col [str]:** column name in the strip the dot is in
- **x [float]:** x coordinate of the dot centroid in the image (in pixels)
- **y [float]:** y coordinate of the dot centroid in the image (in pixels)
- **mean_intensity_control [float]:** mean intensity of the dot in the reference channel
- **area [float]:** area of the dot in pixels
- **mean_intensity_probe [float]:** mean intensity of the dot in the probe channel
- **norm_probe_intensity [float]:** mean_intensity_probe / mean_intensity_control
- **positive_threshold [float]:** threshold for norm_probe_intensity value required for the dot to be positive (set by zscore_threshold).
- **pos_hit [bool]:** True if the dot was called positive (based on positive_threshold)

References:

- [dotblotr] <https://github.com/czbiohub/dotblotr>
- [numpy] <http://dx.doi.org/10.1109/MCSE.2011.37>
- [skimage] <https://doi.org/10.7717/peerj.453>
- [scipy] <https://doi.org/10.1038/s41592-019-0686-2>
- [pandas] <https://doi.org/10.5281/zenodo.3509134>
- [matplotlib] <https://doi.org/10.1109/MCSE.2007.55>
- [openCV] <https://github.com/skvarn/opencv-python>

Peptide Name (Figure 2)	Peptide Name (Library Design)	Amino acid sequence
SARS2_N_134-171	SARS2__N__YP_009724397.2__frag_8	ATEGALNTPKDHIGTRNPANNAIAIVLQLPQGTTLPKGF
SARS2_N_153-190	SARS2__N__YP_009724397.2__frag_9	NNAAIVLQLPQGTTLPKGFYAEGSRGGSQASSRSSRS
SARS2_N_210-247	SARS2__N__YP_009724397.2__frag_12	MAGNGGDAALALLLDRLNQLESKMSGKGQQQQGQTVT
SARS2_N_362-399	SARS2__N__YP_009724397.2__frag_20	TFPTEPKDKKKKKADETQALPQRQKKQQTVTLLPAAD
SARS2_S_552-589	SARS2__S__YP_009724390.1__frag_30	LTESNKKFLPFQQFGRDIADTTDAVRDPQTLEILDITP
SARS2_S_799-836	SARS2__S__YP_009724390.1__frag_43	GFNFSQILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQ
SARS2_S_799-836	SARS2__S__YP_009724390.1__frag_44	IEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKF
SARS2_S_819-855	SARS2__S__YP_009724390.1__frag_61	LQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVN
SARS2_ORF3a_172-209	SARS2__ORF3a__YP_009724391.1__frag_10	GDGTTSPISEHDYQIGGYTEKWESGVKDCVVLHSYFTS

Supplementary Table 1. SARS-CoV-2 Peptide Sequences in ReScan Microarray.