

MICROARRAY PROFILING DETECTS LINEAR EPITOPES INDUCED BY SARS-
COV-2 INFECTION AND VACCINATION MAPPING TO
FUNCTIONAL REGIONS

AN HONORS THESIS

SUBMITTED ON THE 6 DAY OF MAY 2022

TO THE DEPARTMENT OF CELL AND MOLECULAR BIOLOGY

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

OF THE HONORS PROGRAM

OF NEWCOMB-TULANE COLLEGE

TULANE UNIVERSITY

FOR THE DEGREE OF

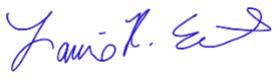
BACHELOR OF SCIENCES

WITH HONORS IN CELL AND MOLECULAR BIOLOGY

BY


LANE PIERSON

APPROVED:



LAURIE EARLS
Co-Director of Thesis



TONY HU
Co-Director of Thesis



BRUCE GIBB
Third Reader

Lane Pierson. MICROARRAY PROFILING DETECTS LINEAR EPITOPES INDUCED BY SARS-COV-2 INFECTION AND VACCINATION MAPPING TO FUNCTIONAL REGIONS

(Professor Laurie Earls, Cell and Molecular Biology and Professor Tony Hu, Biochemistry)

Abstract

The COVID-19 pandemic began in March 2020 and has been ongoing for just over two years now. Rapid identification of epitopes targeted during infection or vaccination can detect therapeutic target sites and evaluate vaccine efficacy and durability. Current methods are low throughput and require challenging mapping studies for epitope identification. Here we employed a peptide microarray to rapidly map SARS-CoV-2 spike and nucleocapsid proteins IgM linear epitopes detected after infection and vaccination. Linear epitope sites detected in non-human primates and patients following SARS-CoV-2 infection revealed extensive overlap and tended to localize to functionally important regions and align with reported neutralizing antibody binding sites. Similar overlap was observed following infection and vaccination, but with group-specific epitope clusters, where specific epitopes mapped to sites known or likely to inhibit protein function. The vaccine-specific epitopes mapped to the central helix and heptad repeat 2, implying differential response to the mRNA-based vaccine spike protein. Mapping linear epitopes to structural regions of known functional importance in this manner may aid in discovery of new targets for antibody therapeutics and the evaluation of vaccine response.

Acknowledgements

I would like to thank the Hu lab, Dr. Tony Hu, Dr. Li Yang, Dr. Christopher Lyon, and every other lab member for all the support and guidance. Without your expertise, editing skills, and ideas, accomplishing this would not have been possible. I am thankful for having the opportunity to join this stellar and innovating lab and being able to gain insights, knowledge, and lifelong friends that will aid me in my future endeavors. Dr. Hu without your constant encouragements, none of the things I have been able to do in the last year and half, receiving grants, an accepted conference abstract, and more, so I am truly thankful and honored to have you as my PI. I would also like to thank my major advisor, Dr. Laurie Earls, along with Dr. Hu, for encouraging me to write an honors thesis and their feedback during the writing process. I would like to thank my third reader, Dr. Bruce Gibb, for being my third reader and making organic chemistry a great class despite the rumors. Lastly, I would like to thank my mom and dad, Veronica and Lawrence Pierson, as well as my friends for the love and support they should along the way.

Table of Contents

<i>Title Page</i>	<i>i</i>
<i>Abstract</i>	<i>ii</i>
<i>Acknowledgements</i>	<i>iii</i>
<i>List of Tables</i>	<i>vi</i>
<i>List of Figures</i>	<i>vi,vii</i>
<i>Introduction</i>	<i>1</i>
<i>Methods</i>	<i>3</i>
Reagents	3
Experimental model and subject details	4
1. Animals	4
2. Human subjects.....	5
2.1. COVID-19 Cohort.....	5
2.2. SARS-CoV-2 mRNA vaccine cohort.....	5
Method details	6
1. Microarray design	6
2. SARS-CoV-2 antibody screening on the proteome microarray	7
3. Immunofluorescence assay	10
4. Structure analysis	11
5. Quantification and statistical analysis.....	11
<i>Results</i>	<i>13</i>
Microarray mapping of the IgM antibody response to SARS-CoV-2 in NHPs	13
Microarray mapping of patient IgM responses to SARS-CoV-2 infection	20
Microarray mapping of patient IgM response to SARS-CoV-2 vaccination	27

Overlapped epitopes in COVID-19 patients and vaccinated individuals.....	34
<i>Discussion:</i>	<i>37</i>
<i>Supplementary Information</i>	<i>44</i>
<i>References</i>	<i>44</i>

List of Tables:

Table 1 List of Reagents.....3-4

Table 2. S protein epitope clusters detected in NHP, COVID-19 and vaccinated participants....29

Table 3. IgM S peptide among the COVID-19 patients, vaccinated and VPI participants.....35-36

List of Figures:

Figure 1. Composition of the proteome microarrays.....7

Figure 2. Proteome Microarray Workflow.....9

Figure 3. Performance of the one-strain SARS-CoV-2 proteome microarray.....9

Figure 4. Performance of the international SARS-CoV-2 microarray.....10

Figure 5. Proteins and peptides identified by proteome microarray in non-human primates....14

Figure 6. Linear peptides mapping of the S protein antibody response by SARS-CoV-2 infected NHPs.....16

Figure 7. Identification and mapping of the antibody response to SARS-CoV-2 N proteins in infected NHPS.....19

Figure 8. Proteins and peptides identified by proteome microarray in COVID-19 patients.....21

Figure 9. Identification and mapping of the antibody response to SARS-CoV-2 S proteins in COVID-19 patients.....23

Figure 10. Identification and mapping of the antibody response to SARS-CoV-2 N proteins in COVID-19 patients.....25

Figure 11. Identification and mapping of the antibody response to SARS-CoV-2 nonstructural proteins in COVID-19 patients.....26

Figure 12. S protein and protein fragment antibody response in SARS-CoV-2 RNA vaccine participants.....28

Figure 13. S peptide mapping of the antibody response in longitudinal SARS-CoV-2 mRNA vaccine participants.....30

Figure 14. S peptide mapping of the antibody response in cross sectional SARS-CoV-2 RNA vaccine participants and VPI participants.....32

Figure 15. Antibody response to non-SARS-CoV-2 respiratory viruses in longitudinal vaccinated participants.....33

Figure 16. Antibody response to non-SARS-CoV-2 respiratory viruses in vaccinated, and VPI participants.....34

Figure 17. Difference in IgM linear peptide epitopes detected in the COVID-19, vaccinated and VPI groups.....37

Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) emerged in 2019, causing a high morbidity and mortality, and became a global pandemic named COVID-19. After two years, COVID-19 remains a global pandemic due to the emergence of new and more infectious strains. New vaccines and therapeutics are urgently needed as new variants, such as Omicron, are more resistant to neutralization by mRNA vaccine generated protection and treatment¹. Development of new vaccines and therapeutics to combat this resistance would benefit from a method to detect linear or tertiary SARS-CoV-2 protein motifs that could function as drug targets.

SARS-CoV-2 is constructed of four structural proteins, the glycospike (S), nucleocapsid (N), envelope (E), and membrane (M), and 16 nonstructural proteins from the ORF1ab, ORF3a, ORF6, ORF7a, ORF8, and ORF10 genomic regions². Studies have shown that the heavily glycosylated trimeric S protein is capable of recognizing the host cell receptor angiotensin converting enzyme 2 (ACE2) and entering the cell after conformational transformation³⁻⁶. To facilitate these functions, the S protein is comprised of a receptor binding domain (RBD), a fusion peptide and heptad repeat 1 and 2 (HR1 and HR2), and a furin cleavage site, which also grants SARS-CoV-2 enhanced transmissibility when compared to SARS-CoV and MERS-CoV⁷. It has two prominent domains, RNA binding domain and dimerization domain, which function to bind to RNA and other N proteins for the encapsulation of the viral RNA⁸⁻¹⁰. The M protein is located in the viral envelope and is essential for viral assembly by interacting with other proteins and defines the shape of the viral envelope^{11,12}. The E protein is located in the viral envelope and helps in virion trafficking and maturation¹³.

Current development studies focus on the dynamics of antibody responses in vaccinated animal models¹⁴⁻¹⁷, COVID-19 patients^{18,19} and vaccinated populations^{20,21}, and then isolating antibodies from these populations to serve as potential blocking or neutralizing antibodies. This approach is low throughput and requires challenging mapping studies for the identification of binding sites which may be nonlinear and therefore complicate epitope identification and the development of vaccination approaches that target them. Other studies have employed the use of immunoinformatic analysis for identification of candidate linear epitopes^{5,22,23}. A few studies have attempted to use antibodies produced by COVID-19 patients to identify linear epitopes^{5,24-26}, due to the simplicity of generating and screening monoclonal antibodies to a specific epitope present on a linear peptide sequence versus a secondary or tertiary protein structure.

With the importance of the S protein and its immunogenic properties it has become one of the main targets of vaccine^{20,21} and antibody therapy designs^{7,27-29}. Many epitope mapping studies have focused on the RBD, due to its importance in cell entry and its showing attenuated virus interaction with ACE2 and cell entry due to targeting antibodies³⁰⁻³³. Other studies have also shown that antibodies to peptides in the NTD of the S protein could elicit a virus neutralization effect^{24,29,30,34}, indicating that additional regions of the S protein may play a role in viral replication and emphasizing the need for more efforts in identifying suppressive linear epitopes in other S protein regions. While the N protein is in the viral particle, it is abundantly released into the blood during infection and is being investigated in potential clinical and therapeutic designs³⁵⁻³⁷. Once in the cell, it has also been shown to have immune suppression capabilities by counteracting RIG-1 and inhibiting avASG formation^{38,39}, thus delaying the innate immune response.

Important questions still remain regarding epitope mapping for vaccine studies that include the degree of overlap between epitopes detected by the immune responses of non-human primates (NHPs) and humans following SARS-CoV-2 infection, and whether the modified spike protein in the Pfizer-BioNTech and Moderna RNA vaccines induces an antibody response similar to that observed following SARS-CoV-2 infection, particularly with regard to the induction of potential blocking and neutralizing antibodies.

In order to solve these issues, we designed an assay that makes use of a proteome microarray to determine the most common linear epitopes recognized by antibodies (IgG and IgM) in SARS-CoV-2 infected NHPs, COVID-19 patients, and individuals vaccinated with an mRNA vaccine. Results from this analysis revealed substantial overlap in epitopes detected in the SARS-CoV-2 infected NHPs and COVID-19 patients and suggest that linear epitope mapping studies in NHPs may be useful in identifying variant- and vaccine-specific epitopes for targeted development of vaccines and monoclonal antibody therapeutics.

Methods

Reagents

The following is the list of reagents.

Table 1. List of Reagents.

Reagent name	Catalog	Source
Goat Polyclonal goat Anti-Monkey IgM Secondary Antibody [DyLight 550]	NBP2-59719R	Novus Biologicals
Goat Polyclonal Goat Anti-Monkey IgG (H+L) Secondary Antibody [DyLight 650]	NB7212C	Novus Biologicals
Alexa Fluor 647-conjugated AffiniPure Goat Anti-Human IgM Fc5 μ Fragment Specific	109-605-043	Jackson ImmunoResearch
Cy3-AffiniPure Donkey Anti-Human IgG (H+L)	709-165-149	Jackson ImmunoResearch
Skim Milk	BD 232100	BD

Phosphate Buffered Saline (PBS)	SH30256.LS	Cytiva HyClone
Tween-20	9005-64-5	Sigma-Aldrich
ACROBiosystems Supplier Diversity Partner Anti-SARS-CoV-2 Antibody IgG Titer Serologic Assay kit (Spike protein S1)	TASK001	ACROBiosystems
ACROBiosystems Supplier Diversity Partner Anti-SARS-CoV-2 Antibody IgM Serologic Assay kit (S1 protein)	TASK012	ACROBiosystems
Peroxidase-AffiniPure Goat Anti-Human IgG (H+L)	109035003	Jackson ImmunoResearch
1-STEP ultra TMB-ELISA	34028	Thermo Scientific

Experimental model and subject details

In this study, we included a SARS-CoV-2 infected NHPs, COVID-19 patient, and SARS-CoV-2 mRNA vaccinated participants. The following is a detailed description of the cohorts themselves.

1. Animals

Plasma samples from a cohort of SARS-CoV-2-infected NHPs (**Table S1**) were analyzed. This cohort was generated at the Tulane National Primate Research Center using an established model of SARS-CoV-2 infection. The Institutional Animal Care and Use Committee of Tulane University reviewed and approved all the procedures for this experiment. The Tulane National Primate Research Center is fully accredited by the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC). All animals are cared for per the NIH Guide to Laboratory Animal Care. The Tulane Institutional Biosafety Committee approved the procedures for sample handling, inactivation, and removal from BSL3 containment. A total of seven male NHPs aged 7 to 11 years were subjected to aerosol inoculation with the SARS-CoV-2 isolate USA-WA1/2020 (CDC). Four African Green Monkeys (AGMs) were exposed to a dose of 1×10^4 TCID₅₀, and three Indian Rhesus Macaques (IRMs) were exposed to a dose of 0.5×10^4

TCID50. The animals were evaluated by twice daily monitoring for 28 days after infection by veterinary staff, and blood samples were collected from all animals 7 days prior to SARS-CoV-2 exposure and at 1, 6-, 13-, 22-, and 28 days post-infection.

2. Human subjects

2.1. COVID-19 Cohort

Thirty-five patients had blood samples collected by Weill Cornell Medicine (**Table S2**), as approved by their institutional review board. Twenty of the participants in this cohort had their blood collected before 2019 and were labelled as pre-COVID-19. The other fifteen participants were hospitalized with COVID-19. Of the fifteen COVID-19 patients (4 women and 11 men; aged 35 to 87 years), eight had blood collected 3-11 days after symptom onset with the remaining seven having blood collected at 20-23 days after symptom onset.

2.2. SARS-CoV-2 mRNA vaccine cohort

Twenty study participants, receiving either the Pfizer-BioNTech or Moderna COVID-19 mRNA vaccine, were enrolled. Five of the twenty participants were enrolled in a longitudinal study to evaluate the dynamic response of linear epitopes produced during the subsequent immune response (**Table S2**). All participants provided written informed consent before study participation approved by the institutional review board of Tulane University (REF#: 2021-040). Fingertip blood samples were collected before vaccination, at 12 days after the first vaccine dose, and at 3, 7, and 14 days, and 1 and 2 months after the second dose. The other fifteen participants were enrolled in a second study to independently confirm the reproducibility of major linear epitopes produced following vaccination (Table S2). All subjects provided written informed consent before study participation approved by the institutional review board of Tulane University (REF#: 2021-040). Additionally, the use of patient specimens was approved by the

Institutional Review Board at the University of Texas Southwestern Medical Center (RED#: STU2021-0001). Lithium heparin plasma specimens were stored at 4°C after collection for up to 72 hours and were then stored at -80°C before analyzing. The participants' blood samples were collected between 7 - 30 days after receipt of their second vaccine dose, and four of these individuals had documented COVID-19 cases prior to their first vaccine dose.

Method details

1. Microarray design

Two SARS-CoV-2 proteomic microarrays were employed, a one strain SARS-CoV-2 proteome array and an international SARS-CoV-2 proteome array. The one-strain SARS-CoV-2 proteome microarray consisted of S, N, and E proteins, as well as spotting peptides that were 15 amino acids in length with overlaps of 5-amino acids, covering the entire SARS-CoV-2 genome (MN908947.3, 966 peptides) (**Table S3**). This array was used to screen NHP cohort. The international SARS-CoV-2 proteome microarray also consisted of S, N, and E proteins, as well as spotting peptides that were 15 amino acids in length with overlaps of 5-amino acids, covering the entire SARS-CoV-2 proteome (MN908947.3). Additionally, it included peptides of 52 other SARS-CoV-2 isolates from different countries, and other potentially cross-reactive proteins and protein fragments from non-SARS-CoV-2 respiratory viruses (**Table S3 and Figure 1**). This array was used to screen the human cohorts.

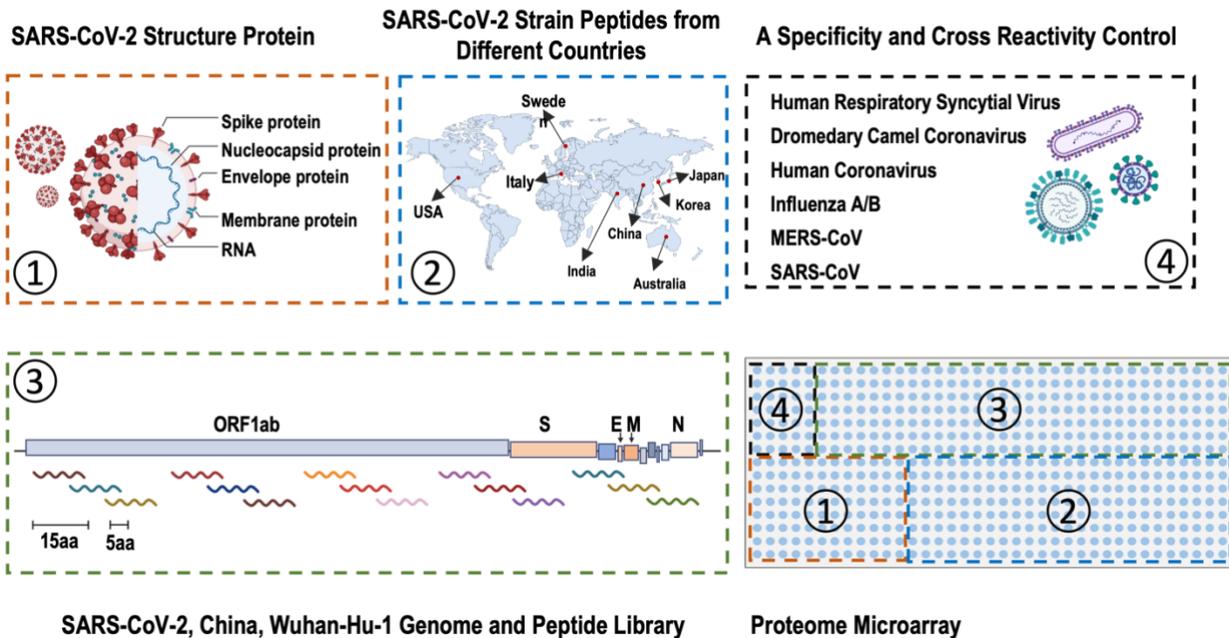


Figure 1. Composition of the proteome microarrays.

Schematic of the design and composition of the one strain microarray (1 and 3) and the international SARS-CoV-2 microarray (1-4). Part 1 describes the SARS-CoV-2 proteins. Part 2 describes where the different variants used to fabricate the array were isolated from. Part 3 describes the peptides on the array. Part 4 list the other non-SARS-CoV-2 viruses related proteins used in the fabrication of the array.

2. SARS-CoV-2 antibody screening on the proteome microarray

All microarray experiments steps took place in a humidified chamber to reduce the chances of artifacts from unequal evaporation. For the one strain microarray chip, after placement into the chamber, the chips were blocked with 400 μ L 5% filtered skim milk at RT for 10 min. Next, 4 μ L of plasma diluted in 400 μ L was added to the chamber and incubated at RT for 30 minutes with gentle rocking. After washing three times with 0.05% PBST for 5 minutes each, the fluorescent secondary antibodies, diluted into 400 μ L 5% skim milk (polyclonal Goat Anti-Monkey IgG (H+L) secondary antibody (Dylight 650) and polyclonal Goat Anti-Monkey IgM secondary antibody (Dylight 550) (working concentration is 4 μ g/mL)), were added into the chamber for 20

minutes and incubated at RT with gentle rocking. Finally, the chips were washed with 0.05% PBST for 5 minutes three times and then with deionized water for 2 minutes two times (**Figure 2**). For the international microarray chip, after placement into the chamber, the chips were blocked with 3000 μ L 5% filtered skim milk at RT for 10 minutes. Next, 30 μ L of plasma diluted in 3000 μ L 5% skim milk was added to the chamber and incubated at RT for 30 minutes with gentle rocking. After washing three times with 0.05% PBST for 5 minutes each, the fluorescent secondary antibodies, diluted into 3000 μ L 5% skim milk (Cy3-conjugated AffiniPure Donkey Anti-Human IgG (H+L) and Alexa Fluor647-conjugated AffiniPure Goat Anti-Human IgM (Fc5 μ fragment specific) (Jackson ImmunoResearch) (working concentration is 4 μ g/mL)), were added into the chamber for 20 minutes and incubated at RT with gentle rocking. Finally, the chips were washed with 0.05% PBST for 5 minutes three times and then with deionized water for 2 minutes two times (**Figure 2**). All chips (one strain and international) were then scanned using an Agilent microarray scanner, and the fluorescence signal intensity of each spot was extracted using GenePix Pro7 software (Molecular Devices). The reproducibility of the antibody detection for both the one strain SARS-CoV-2 proteome microarray and international SARS-CoV-2 proteome microarray was determined by comparing the two spots for one peptide on the same slide and the same peptide spot from two arrays incubated with the same sample (**Figure 3 and Figure 4**).

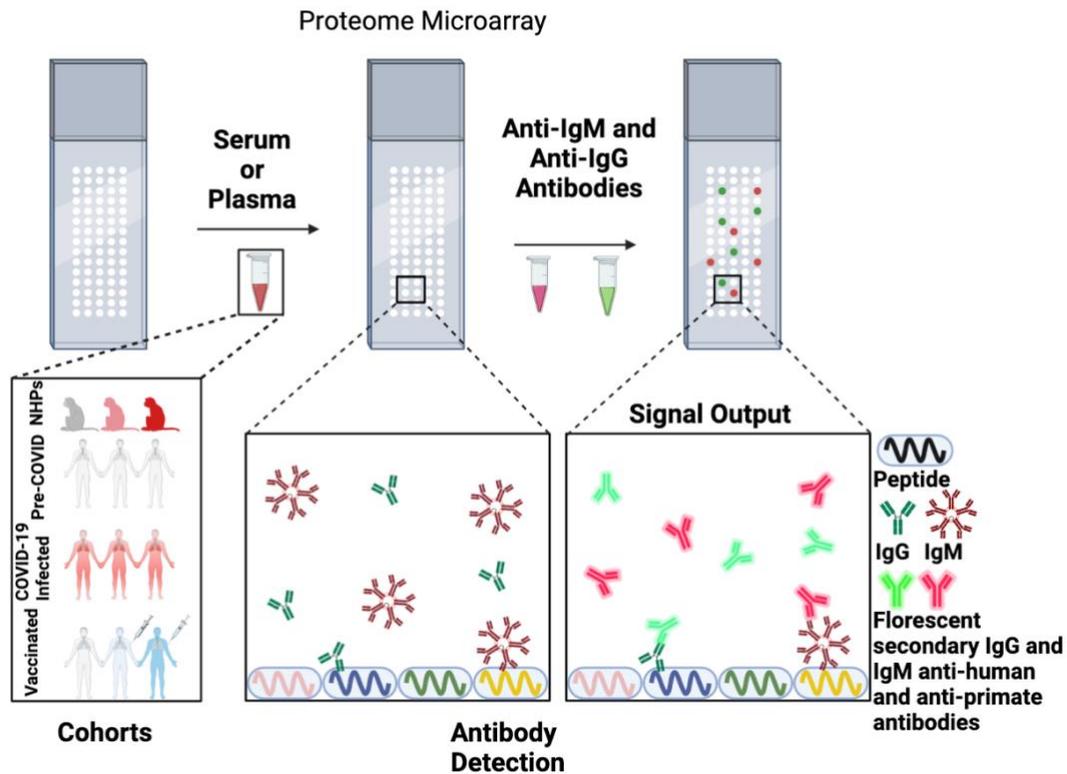


Figure 2. Proteome Microarray Workflow.

Workflow of the SARS-CoV-2 proteome microarray. Plasma from the NHP cohort, and serum and plasma from the human cohorts were incubated onto the array to allow for sample antibody binding to peptides. Next the florescent secondary antibodies were incubated onto the array to allow for their attachment to the sample antibodies.

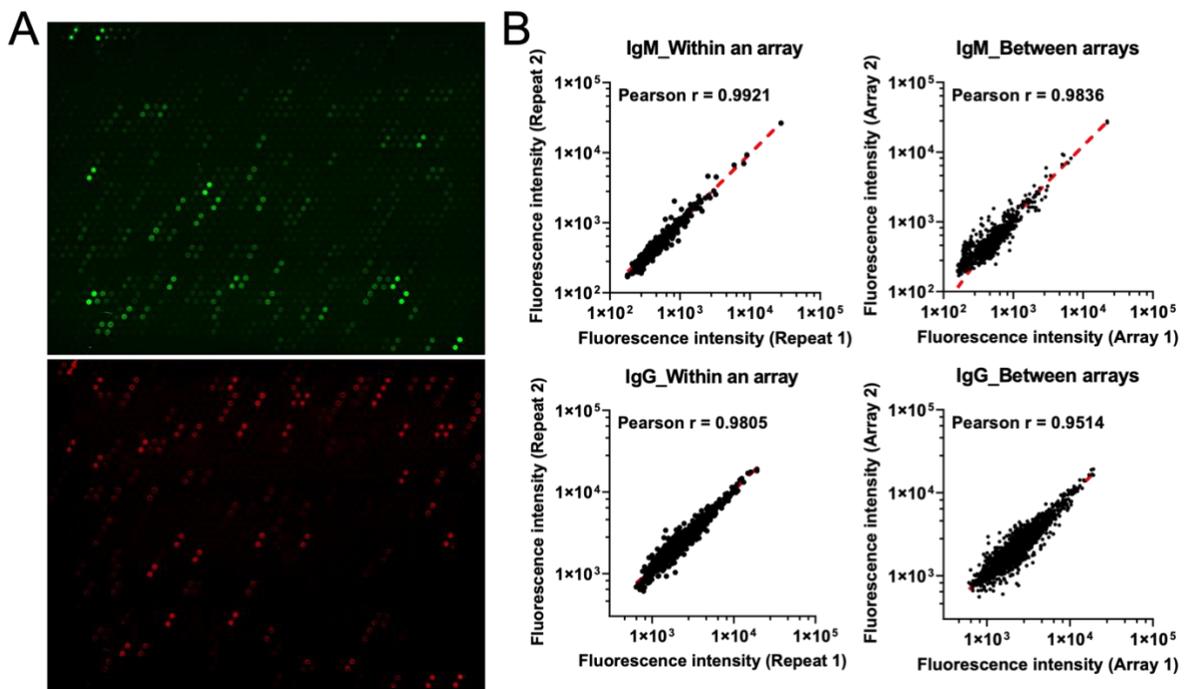


Figure 3. Performance of the one-strain SARS-CoV-2 proteome microarray.

A. The fluorescent images of IgM and IgG show low background noise. Which one is IgG and IgM?
B. Reproducibility of antibody detection using the one-strain SARS-CoV-2 proteome microarray for non-human primates infected by SARS-CoV-2.

4. Structure analysis

Epitope mapping and contact distance evaluation was performed using Chimera X1.2.5 software⁴⁰. Five different 3D models of SARS-CoV-2 proteins, three of the S protein and two of the N protein, were used to map epitopes. The first model, PDB:6VYB, which had one of the RBDs in the “up”, was used to map the epitopes detected in the NHP and COVID-19 cohorts. The second model, PDB:6LZG, modeled the SARS-CoV-2 S protein RBD in its bound state to the ACE2 receptor and was used to determine the interaction between the RBD linear peptide epitope S481-495 and its ACE2 interaction site⁵. The third model, PDB:6VSB, represented the structure of the S protein used in Pfizer and Moderna vaccines, with the two proline substitutions (K986P and V987P) which stabilize the S protein with the RBD in the up position⁴¹. The two N models used for the NHP, and COVID-19 cohorts were the RNA binding domain model (PDB: 6VYO) and a dimerization domain model (PDB: 7C22). The former appears as a tetramer while the later appears as a trimer. The use of two models is due to a full-length N protein not having been isolated and described yet.

Using the model of the S protein RBD bound to the ACE2 receptor, the van der Waals forces between the amino acid residues of the proteins were evaluated using the “Contacts” feature in Chimera X1.2.5. The setting used for this analysis identified pairs of atoms with center-to-center distances ≤ 4.5 angstroms, ignored interactions between atoms four or fewer bonds apart, and included intermodel and intramodel interactions.

5. Quantification and statistical analysis

Microarray signal intensities were normalized using the Z-score method. The raw signal intensity for each peptide was determined by taking the average signal intensity of a spot and then averaging that signal with that peptide’s duplicate spot. For the NHPs, all of the data generated

from a single individual were normalized together. For the human cohorts, normalization occurred for each sample independently.

Identification of peptides detected by IgM and IgG binding was done by applying further statistical test. For the NHPs, peptides were deemed to be epitopes by having p-values < 0.05 from repeated measure ANOVAs with Dunnett's post hoc tests when at least four of the seven NHPs had values greater than the baseline + three times its standard deviation. For the SARS-CoV-2 mRNA longitudinal samples, peptides were deemed to be epitopes by having p-values < 0.05 from repeated measure ANOVAs with Dunnett's post hoc tests and mean values greater than the pre-vaccination sample mean + three times its standard deviation. For the COVID-19 patients and the SARS-CoV-2 mRNA cross-sectional samples, peptides were deemed to be epitopes by having p-values < 0.05 from parametric one-way ANOVA with Dunnett's post hoc test and mean values greater than the mean of the pre-COVID-19 group + three times its standard deviation. ANOVA tests were performed using the “multcomp” R software library and its “mvtnorm”, “survival”, “TH.data”, and “MASS” packages.

Pearson correlation coefficients were analyzed using GraphPad Prism software then used to compare the fluorescence intensity between duplicate spots within an array and the fluorescence intensity of corresponding spots between arrays when these arrays were incubated with the same samples. Non-parametric one-way ANOVAs with Dunn's post-tests performed using GraphPad Prism were used to evaluate the difference between the vaccinated, vaccinated post-infection, and COVID-19 patient groups. Heatmaps indicating antibody responses to SARS-CoV-2 proteins and peptides were constructed using the “pheatmap” package in the R software suite. Graphs were generated with GraphPad Prism and Schematic diagrams were generated using BioRender.

Results

Microarray mapping of the IgM antibody response to SARS-CoV-2 in NHPs

Pre-infection and post-infection plasma from Indian Rhesus macaques (RMs) and African Green monkeys (AGMs) infected with SARS-CoV-2 aerosol exposure to known doses of SARS-CoV-2 isolate strain USA-WA was analyzed by hybridization to a proteomic microarray (**Table S3**).

This allowed for the analysis on the change in SARS-CoV-2-specific antibody responses over time.

Both IgM and IgG antibody response from 1-day post-infection (DPI) plasma revealed weak interactions with SARS-CoV-2 proteins and protein fragments (**Figure 5A**). IgM signals for all proteins began increasing after 6 DPI with three, Protein S RBD, Protein S1+S2ECD, and Protein N, reaching significance at 28 DPI. IgG signals were weaker with a peak at 13 DPI yet stayed significant after 6 DPI. Similarly, IgM also detected more SARS-CoV-2 peptides with stronger signal than IgG (Table S4). Since IgG antibodies derive from the IgM population by isotype switch recombination and since multimeric antibodies, and particularly IgM, play major roles in SARS-CoV-2 neutralization⁴²⁻⁴⁴, all subsequent studies focused on IgM epitopes.

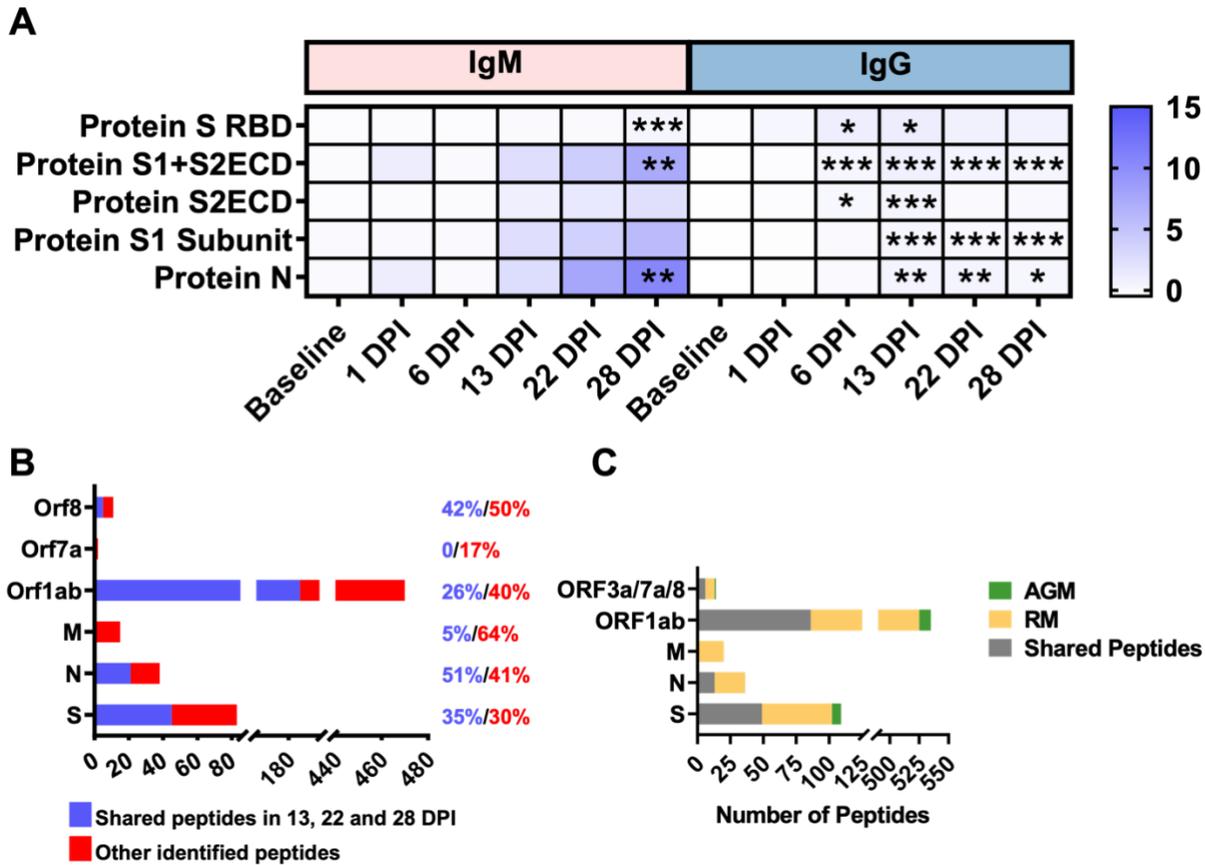


Figure 5. Proteins and peptides identified by proteome microarray in non-human primates.
 A. Heatmap of mean IgM and IgG signals of SARS-CoV-2 proteins and protein fragments.
 B. Number and percentage of unique peptides detected in IgM binding from 13-28 DPI (blue) or at least one of those time points (red).
 C. Relative abundance and overlap among IgM linear epitopes detected in non-human primate models. Bars indicate the number of unique peptides detected in serum from SARS-CoV-2 infected RMs (yellow) and AGMs (green) or shared between the two models (gray).
 * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, by repeated measure ANOVA with Dunnett multiple comparison test. (n = 7)

IgM and IgG binding to peptides mirrored that of the proteins, in which the IgM signals were markedly stronger than IgG's. Along with the difference in signal strengths, there was a pronounced increase in the number of peptides recognized by IgM versus IgG (**Table S4**). With these differences, how IgG antibodies are derived from the IgM population by isotype switch recombination, and multiple studies reporting that multimeric antibodies, particularly IgM, play

major roles in SARS-CoV-2 neutralization⁴²⁻⁴⁴, all peptide analysis will be focused on IgM epitopes.

NHP IgM detected varying amounts of linear peptides from the SARS-CoV-2 coding regions (**Figure 5B**). Binding to peptides derived from Orf3a, Orf6, or Orf10 were not detected. IgM responses were detected for a portion of Orf7a peptides (17%), a moderate amount of Orf1ab, M, and S peptides (65-68%), and a high amount of Orf8 and N peptides (92-93%). Consistent binding of peptides from 13 DPI to 28 DPI was detected for a portion of the detected Orf1ab, Orf8, M, N, and S peptides (26-51%).

Upon species analysis it was found that more peptide epitopes were identified in RMs versus AGMs (**Figure 5C**). Most epitopes recognized by AGMs (5-48%) were a subset of those detected in RMs, with AGMs uniquely detecting 6%, 2% and 7% of the total S, Orf1ab and Orf3a/7a/8 epitopes.

With the S protein being an integral part of cell entry for SARS-CoV-2, S protein peptides detected at successive DPI were mapped out onto a linear S protein schematic to identify the functions of regions exhibiting significant antibody binding. The linear peptide epitopes were found to cluster to the N-terminal, receptor binding, and C-terminal domains of the S1 subunit, and to the fusion protein region of the S2 subunit (**Figure 6A**). Of these peptides, fourteen were detected to be on or adjacent to reported S protein N- or O-glycosylation sites. However, the peptides on the array were not glycosylated, indicating that antibodies generate to the natural S protein are not inhibited by glycosylation modifications and their absence also does not inhibit recognition of the unmodified peptides on the array.

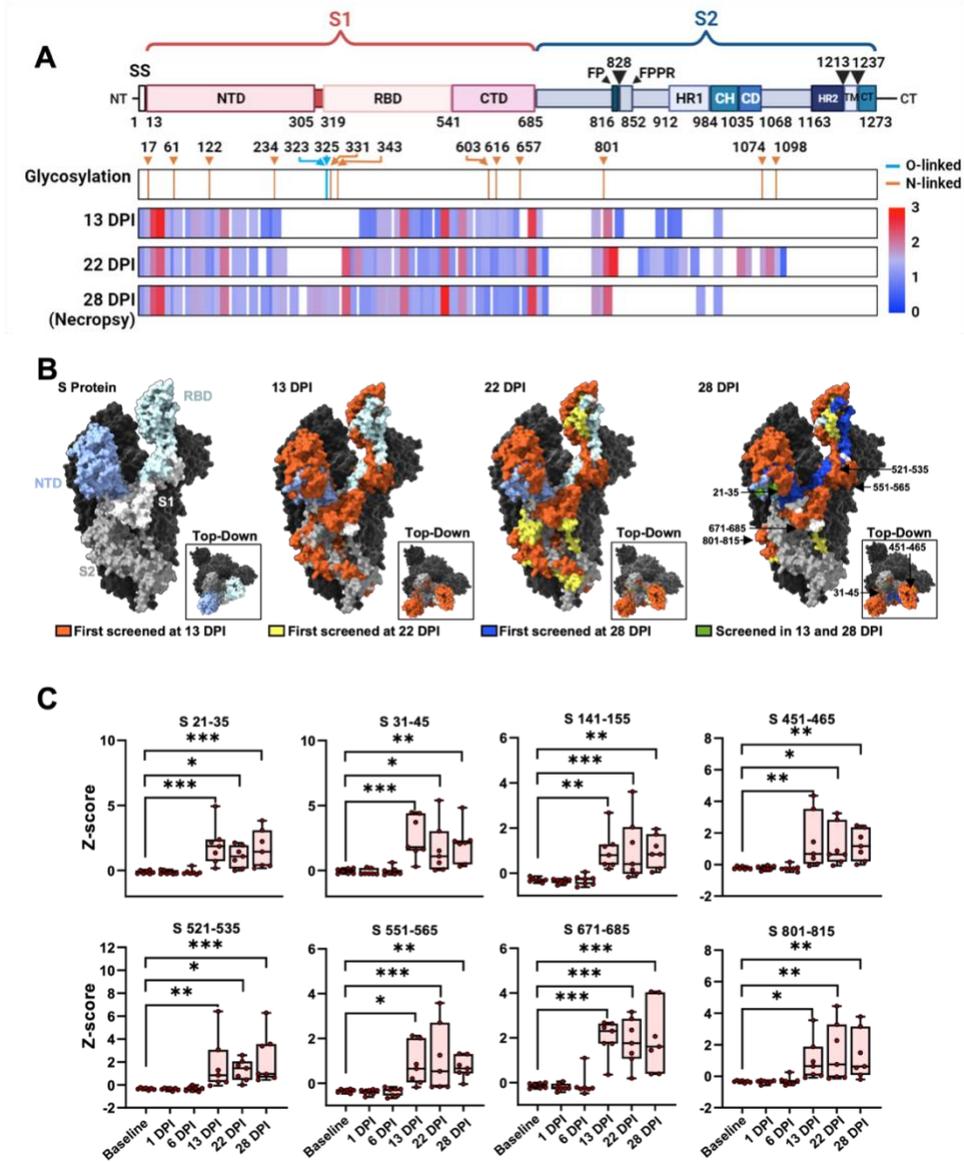


Figure 6. Linear peptides mapping of the S protein antibody response by SARS-CoV-2 infected NHPs.

A. Schematic and heatmap indicating the alignment of detected peptides at the indicated times with S protein features and their mean signal intensities minus baseline signal intensities. O- and N-linked glycosylation positions are numbered and respectively marked by vertical blue and orange lines.

B. 3D structural maps of the SARS-CoV-2 S protein with labeled N-terminal domain (NTD), receptor binding domain (RBD), S1, and S2 peptide sequences and the corresponding peptide sequences bound by IgM at the indicated time points (PDB:6VYB).

C. Reproducibility of IgM signal trends for detected S protein peptides. Graphs indicate each individual value, minimum, lower quartile, median, upper quartile and maximum, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, by repeated measure ANOVA with Dunnett multiple comparison test. (n = 7)

Most peptides detected at 13 DPI were consistently detected afterwards with the tendency for its and its adjacent peptides recognition becoming stronger at increasing DPI (**Figure 6A-B**). This may be due to accumulation and/or affinity maturation if IgM specific sites. Leaner peptide signals detected on the S1 subunit were stronger and more consistent over time than those detected on the S2 subunit which were fleeting, appearing in the 13 and 22 DPI but not in 28 DPI.

Several sites exhibiting strong and persistent IgM signals mapped to regions associated with S protein activity (**Figure 6C**). This included two peptides, S451-465 and S521-535, within the RBD, two others, S671-685 and S801-815, that mapped an area required for activity at the S1/S2 cleavage site and the furin cleavage site adjacent to the S protein fusion peptide, and one, S141-155, in the S1 NTD that is recognized by an antibody, 4A8, demonstrating a high degree of virus neutralization³⁴.

Another protein important in eliciting an immune response is the N protein^{39,45}. The function of the N protein is to encase the RNA genome through the use of three arginine residues and a positively charged pocket in its RNA binding domain³⁶. Although the N protein is inside the viral capsid, it elicited a strong antibody response. Upon further analysis of the N peptides, a high linear peptide coverage (92.7%) of N protein was found. This high degree of coverage came from fleeting signals and persistent signals forming clusters to the NTD to the middle of the RNA binding domain (group one), the end of the RNA binding domain to the middle of the linker (group two), the start of the dimerization domain (group three), and end of the dimerization domain and the CTD (group four) (**Figure 7A**). Of these four groups, groups one and three revealed peptide epitopes with a strong IgM response that increased over the three timepoints that mapped to regions known to have important functions. Structural analysis on the

N protein is hampered by the lack of a full structure isolation^{9,46,47} and therefore models showing only the RNA binding domain and dimerization domain domains in their isolated states^{48,49} were used to map peptides. Almost all peptides found in both the RNA binding domain and the dimerization domain are on exposed surfaces (**Figure 7B**). Linear peptides identified at 13 DPI consistently appeared in later time points while peptides identified at 22 DPI fade away in the RNA binding domain, while for the dimerization domain the epitopes identified at 13 DPI fade with no new peptides being identified after (**Figure 7A-B**).

Several sites exhibited a strong and persistent IgM signal while also mapping to a functional site on the N protein. Sites N 41-55 and N 51-65 both showed strong signals and mapped to the start of the RNA binding domain, while N 91-105 mapped to the middle of the domain (**Figure 7C**). The other five peptides mapped to the CT half of the protein with N 251-265 mapping to the dimerization domain and the other four, N 361-375, N 371-385, N 391-405, and N 401-419, mapping to the CTD.

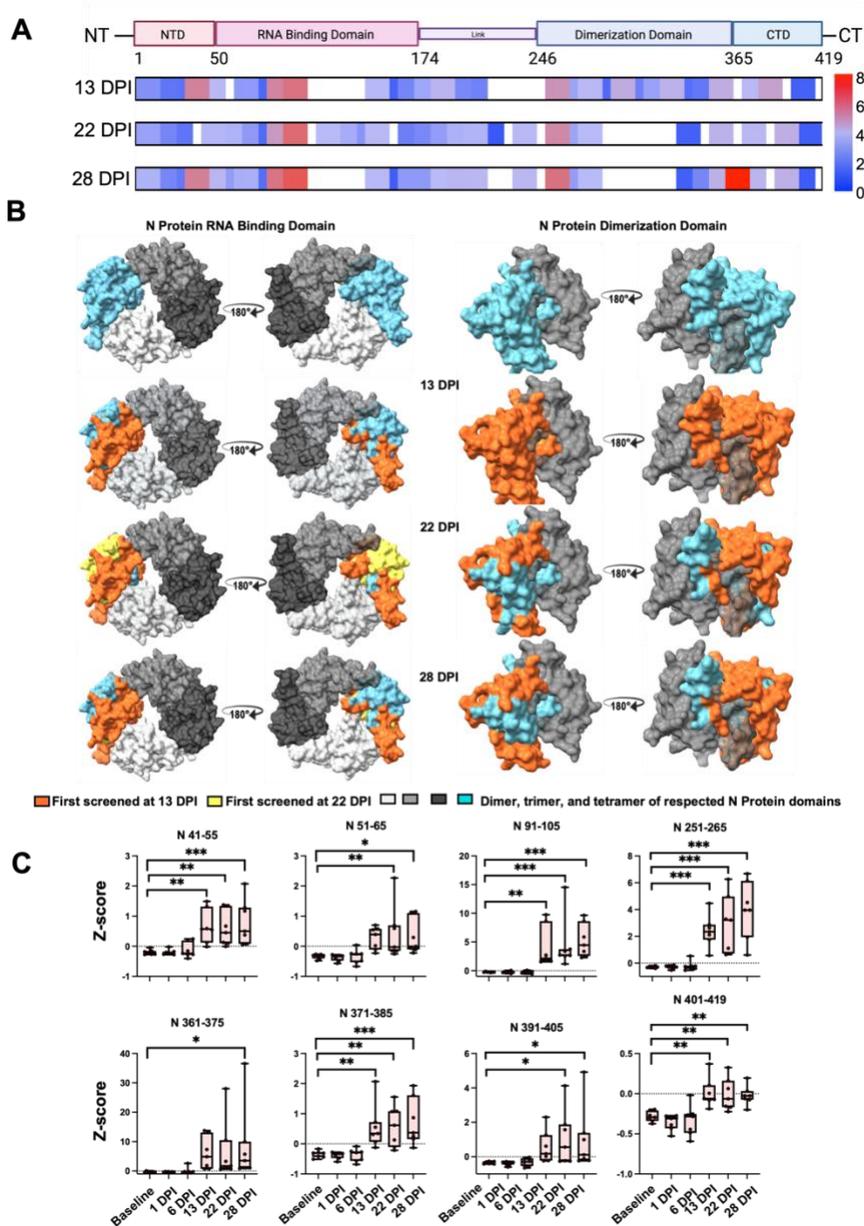


Figure 7. Identification and mapping of the antibody response to SARS-CoV-2 N proteins in infected NHPS.

A. Schematic and heatmap indicated the alignment of detected peptides at the indicated times with N protein features and their mean signal intensities minus baseline signal intensities. .

B. 3D structural maps of the SARS-CoV-2 N Protein RNA binding domain (PDB: 6VYO) and dimerization domain (PDB: 7C22) and the corresponding peptide sequences bound by IgM at the indicated time points.

C. Reproducibility of IgM signal trends for detected N protein peptides. Graphs indicate each individual value, minimum, lower quartile, median, upper quartile and maximum, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, by repeated measure ANOVA with Dunnett multiple comparison test. (n = 7)

Microarray mapping of patient IgM responses to SARS-CoV-2 infection

Serum samples from COVID-19 patients as well as serum from patients before the start of the COVID-19 pandemic were analyzed using the international SARS-CoV-2 proteome microarray which contained proteins and overlapping linear peptides of the SARS-CoV-2 proteome, including peptide variants associated with distinct SARS-CoV-2 isolates, and off-target proteins derived from other human respiratory viruses (**Figure 1 and Table S3**).

IgM and IgG antibody response to SARS-CoV-2 proteins detected in COVID-19 patient serum revealed a stronger IgM than IgG response, similar to the NHPs (**Figure 8A**). This strong response came from samples collected around 20 days after symptom onset and was seen in the full-length N and S proteins, and the S2 ECD, S1, S1 NTD, and S1 CTD protein fragments. The IgG response only showed a weak but significant signal in the full-length S protein and significant but fleeting signals in the full-length N protein and S2 and S1 subunit fragments.

Similar to the NHP results, the COVID-19 patient IgM response varied in its ability to detect linear peptides derived from SARS-CoV-2 coding regions (**Figure 8B and Table S5**). There was no binding detected for Orf6, Orf8 or Orf10 peptides, and only a small percentage of the total Orf1ab, Orf3a, Orf7a, M, and S peptides (1.3 – 8%). A larger percentage of N (23%) peptide sequences were identified. Comparison of the NHPs and COVID-19 cohorts revealed that a substantial overlap, with 92% to 81% for the N and S proteins, respectively, and 64% for Orf1ab (**Figure 9C**). For the detected M proteins, the one detected peptide in the COVID-19 cohort is overlapped with those detected in the NHPs cohort. No overlap was shown for Orf 3a, Orf7a, and Orf8 peptides.

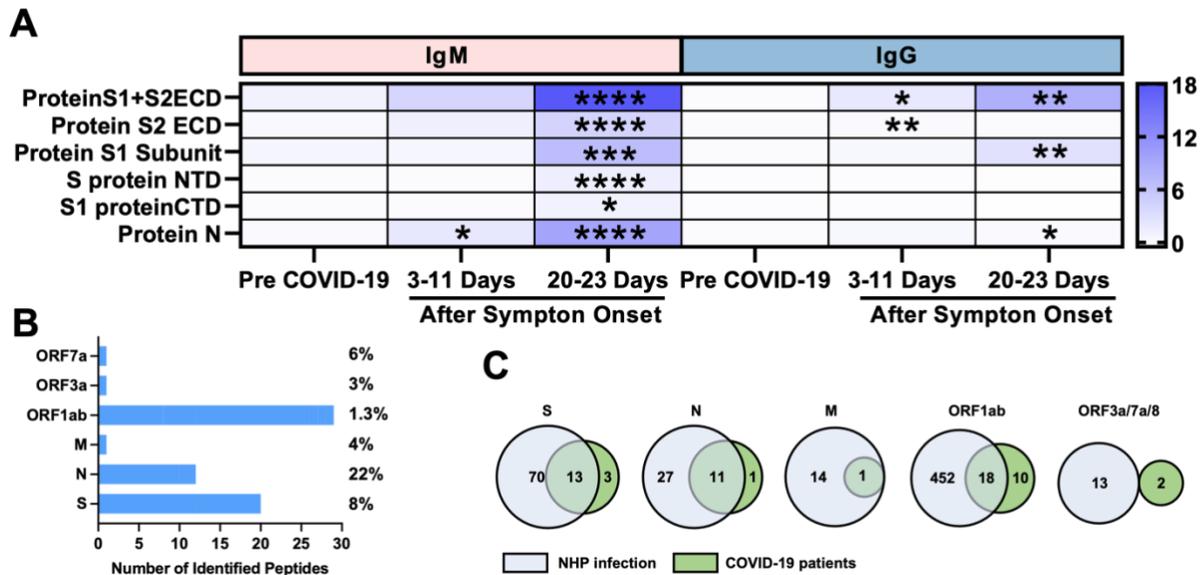


Figure 8. Proteins and peptides identified by proteome microarray in COVID-19 patients.

A. Heatmap of mean IgM and IgG signals of SARS-CoV-2 proteins and protein fragments.

B. Number and percentage of peptides detected by IgM in COVID-19 patients.

C. Overlap of unique linear peptides detected in SARS-CoV-2 infected NHPs and COVID-19 patients.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, by one-way ANOVA with Dunnett post test.

A portion of the linear peptides sites detected had corresponding peptide sequences from two or more SARS-CoV-2 strains represented on the array. Three of the nine detected N peptides had fully overlapped peptide sequences with other two strains, but only one of these variant peptides, N211-225, revealed significant IgM signal (**Table S6**). All 27 detected Orf1ab peptides had overlapping peptide sequences from two or more SARS-CoV-2 strains, although only five revealed complete sequence overlaps with a single substitution, and only one of these, Orf1ab7041-7055, demonstrated significant IgM signal from more than one aligned sequence signal (**Table S6**). This peptide recognized an eleven-residue consensus sequence overlap defined by three variant peptide sequences on the array. Similar to Orf1ab, all 16 detected S protein peptides regions had overlapping peptide sequences from two or more SARS-CoV-2 strains, with all but one revealing full sequence conservation with a single amino acid offset from

the detected peptide (**Table S6**). However, only four sets of these aligned variant peptides, S241-255, S801-815, S811-825, S1251-1265, were also detected on the array, with an additional peptide offset by one amino acid at the S241-255 site not detected, indicating that importance of the terminal amino acid differences among these peptides. Most of the differentially detected peptides, 77%; 10 of 13, at these sites had charged or polar residues at their N- or C-termini (**Table S6**).

Individual data from identified S peptide epitopes showed that a low signal for the samples collected before the COVID-19 outbreak, weak signals for three peptides, S481-495 and S811-825, in the samples collected 3 to 11 days after symptom onset and strong signals for all peptides across most of individual samples collected around 20 days after symptom onset (**Figure 9A**). Surprisingly, a subset of samples drawn before the start of the COVID-19 epidemic revealed IgM binding to the S1+S2 ECD fragment, S1 subunit, and the S2 ECD (**Figure 9A**). This binding activity likely reflected IgM cross-reactivity due to previous exposure to similar linear epitopes on related β coronavirus species responsible for common human respiratory infections (e.g., OC43 and HKU1).

The majority of the S peptide epitopes detected in this analysis localized to the S1 and S2 subunits with similar frequency, with half of the peptides clustering to a 230 amino acid region between the S1 RBD and S2 FP motif (**Figure 9B**). This spread of peptides differed from that found in NHPs, where most epitopes localized to the S1 subunit. Similar to the NHPs, the COVID-19 cohort mapped multiple peptides near or at regions associated with known S protein activities, including its RBD, FP motif, and HR1. One peptides, S481-495, mapped to an RBD region reported to directly interact with the ACE2 receptor^{5,50}.

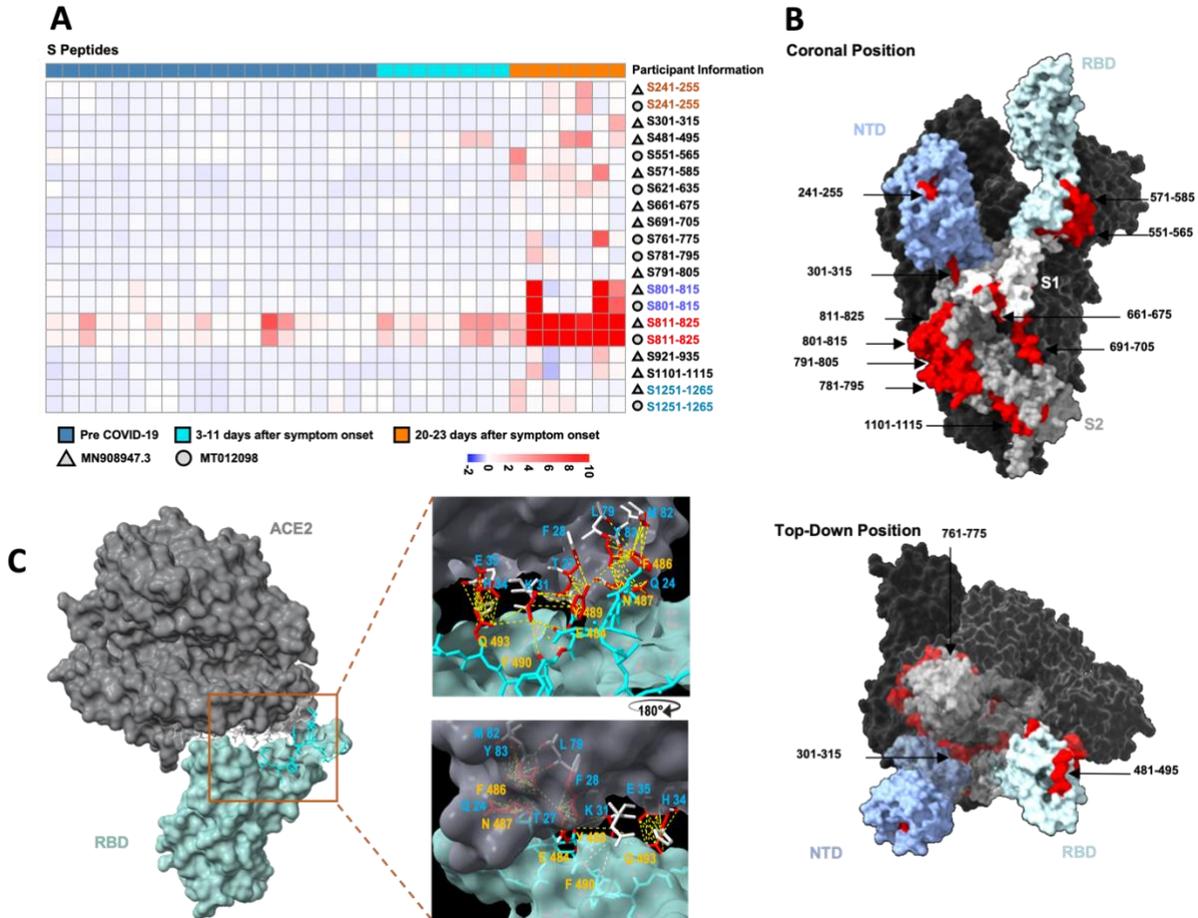


Figure 9. Identification and mapping of the antibody response to SARS-CoV-2 S proteins in COVID-19 patients.

A. Heatmap of IgM signal Z-scores detected for SARS-CoV-2 S protein peptides using serum collected pre-COVID-19 (blue, n = 20) or at 3-11 days (bright blue, n = 8) or at 20-23 days (orange, n = 7) after symptom onset in COVID-19 patients. Colored labels indicate signals detected from overlapping peptide sequences of SARS-CoV-2 strains MT012098 (circles) and MN908947.3 (triangles).

B. 3D structural map indicating IgM binding to linear peptide epitopes on the N-terminal domain (NTD), receptor binding domain (RBD), S1, and S2 peptide sequence regions of the SARS-CoV-2 S protein (PDB:6VYB).

C. 3D structure indicating close interactions (yellow dash lines; ≤ 4.5 angstroms) between predicted contact residues (red) in a detected RBD linear peptide epitope (Blue chain; S481-495) and ACE2 (White chain) (PDB:6LZG).

Furthermore, a structural model of the RBD interaction with ACE2⁵ indicated that six sidechains of the residues in this peptide, E484, F486, N487, Y489, F490, and F493, formed close associations (<4.5 angstroms distant) with sidechains of nine ACE2 residues (**Figure 9B**) This

S1 RBD peptide sequence partially overlaps the binding sites of two known neutralizing antibodies, S2H13²⁹ and F2B-2F6³⁰, that were isolated from COVID-19 patients, suggesting that neutralizing antibodies are routinely generated to this sequence, as it was detected by IgM generated both NHPs and patients after SARS-CoV-2 infection. Two other epitopes that mapped adjacent to the RBD, S551-565 and S571-585, could also hinder RBD interactions with ACE2 via steric hindrance, similar to how another linear epitope S14P5 (S562-579) identified in COVID-19 patients⁵¹. A second IgM binding site at position S661-675 contains four residues, N657, N658, Y660, and E661, reported to contact the furin protease and promote S protein cleavage to induce membrane fusion and SARS-CoV-2 cell entry⁵². Thus, antibodies binding to this interaction site are likely to decrease virulence^{52,53}. Finally, two strongly recognized peptides mapping to residues S801-815 and S811-825 are located near the cleavage sites between S1/S2 and S2^{54,55} located upstream of the fusion peptide involved in cell entry⁵⁴, and antibody recognition of this site may block S protein cleavage or fusion peptide activity.

Besides S peptides, the COVID-19 patient cohort also showed significant IgM interactions with twelve N peptides, of which 11 overlapped with the NHPs. Only one peptide, N161-175, showed strong signals in the samples collected about 3 days after symptom onset, however samples collected around 20 days after symptom onset showed strong signals (**Figure 10A**).

Unlike the NHPs, IgM binding to the N peptides in COVID-19 patients was sporadic (**Figure 10B**). Strong signals mapped to the end of the RNA binding domain, the middle of the linker, the end of the dimerization domain, and the CTD with the peptides being on the surface of the protein (**Figure 10C**).

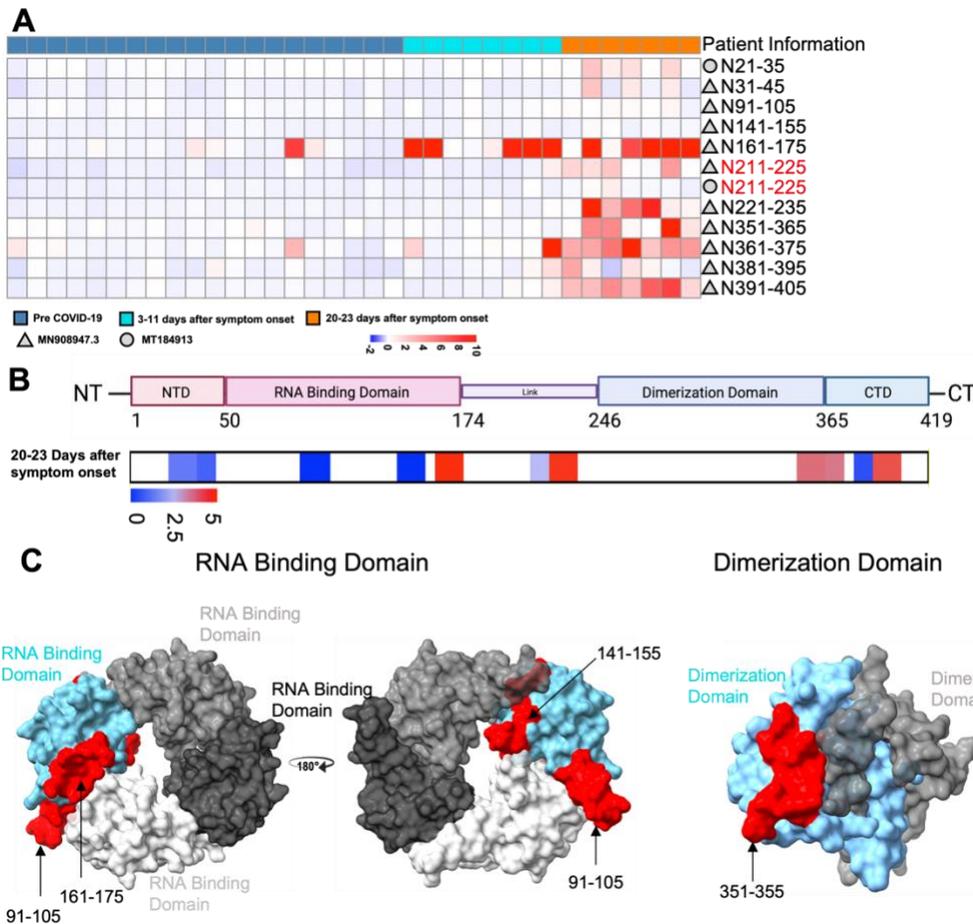


Figure 10. Identification and mapping of the antibody response to SARS-CoV-2 N proteins in COVID-19 patients.

A. Heatmap of IgM signal Z-scores detected for SARS-CoV-2 N peptides using serum collected pre-COVID-19 (blue, n = 20) or at 3-11 days (tan, n = 8) or at 20-23 days (orange, n = 7) after symptom onset in COVID-19 patients. Colored labels indicate signals detected from overlapping peptide sequences of SARS-CoV-2 strains MT184913 (circles) and MN908947.3 (triangles).

B. Schematic and heatmap indicating the alignment of detected peptides at the indicated times with N protein features and their mean signal intensities minus baseline signal intensities.

C. 3D structural maps of the SARS-CoV-2 N protein RNA binding domain (PBD ID: 6VYO) and dimerization domain (PBD ID: 7C22) with the corresponding peptide sequences bound by IgM labelled in red.

All individuals revealed strong binding activity to most proteins encoded by common respiratory viruses, including human respiratory syncytial virus, strains of influenza A and B, multiple common human coronaviruses (HCoV), and the MERS-CoV S protein RBD, regardless of their SARS-CoV-2 infection status (**Figure 11A**). Signals to these viral proteins showed no marked increase, except for the N proteins of MERS-CoV and SARS-CoV which showed a significant

increase (Figure 11B). Signals from COVID-19 patients around 20 days after symptom onset to the S protein of several HCoV strains showed a decrease from the samples collected prior to the COVID-19 pandemic.

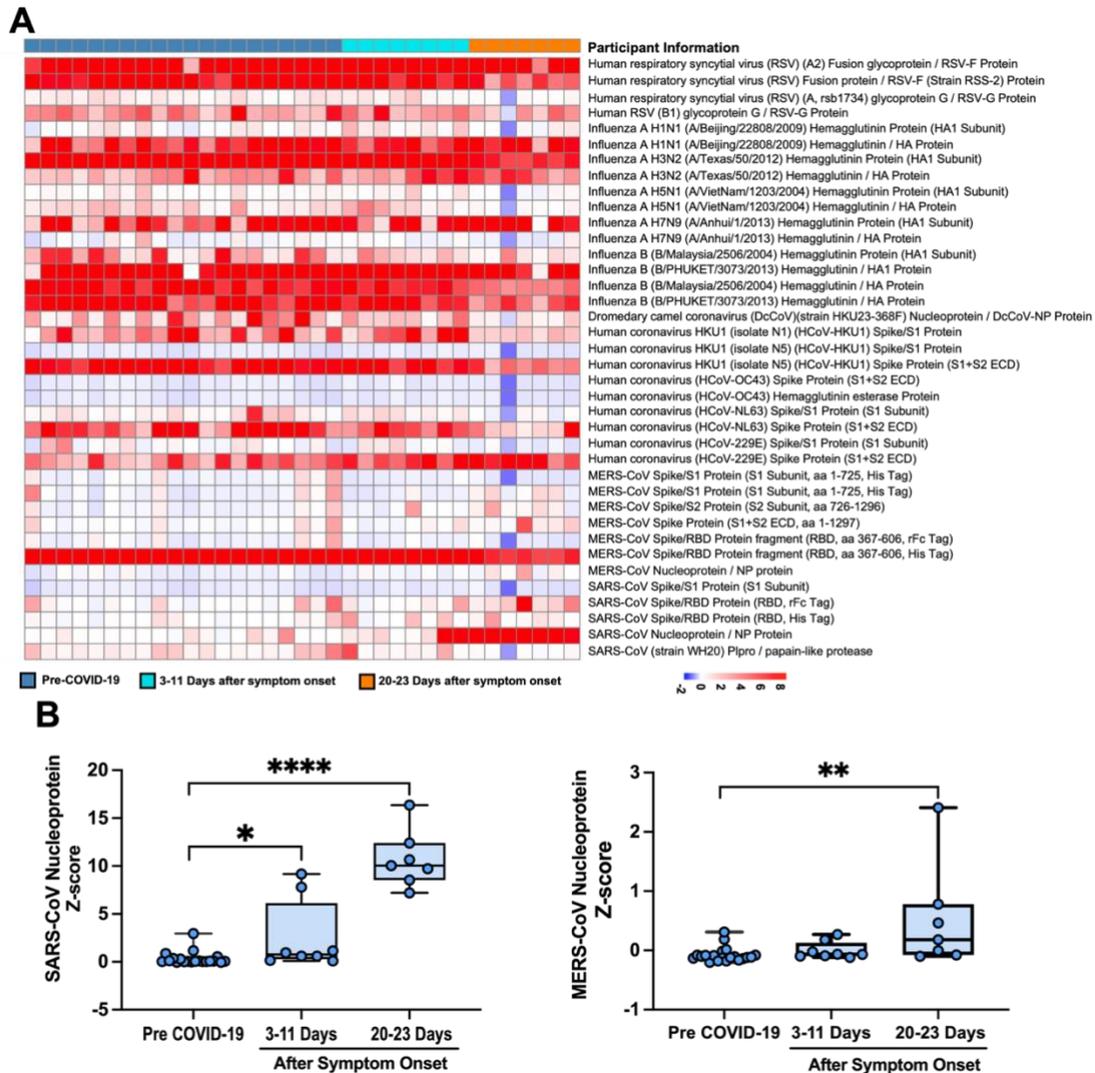


Figure 11. Identification and mapping of the antibody response to non-SARS-CoV-2 respiratory viruses in COVID-19 patients.

A. Heatmap of the IgM signal detected for protein or protein fragments of the indicated respiratory viruses. Pre-COVID-19 (blue, n = 20) or at 3-11 days (tan, n = 8) or at 20-23 days (orange, n = 7) after symptom onset in COVID-19 patients.

B. SARS-CoV and MERS-CoV N protein IgM signal for at the indicated times. Graphs indicate individual, minimum, lower quartile, median, upper quartile and maximum values. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, by non-parametric one-way ANOVA with Dunn's post-test.

Microarray mapping of patient IgM response to SARS-CoV-2 vaccination

SARS-CoV-2 mRNA vaccines contain mRNA that encodes for a modified version of the full-length S protein with two proline substitutions that keep the protein locked in a prefusion conformation to enhance the production of antibodies that recognize this structure and neutralize its fusion activity (Baden et al., 2021; Polack et al., 2020; Wrapp et al., 2020). Due to this it may be likely that S protein specific antibodies produced in response to SARS-CoV-2 infection and vaccination may differ in their S protein sequence coverage and/or affinity for specific sites.

Therefore, the analysis of the antibody responses from a small cohort of longitudinal individuals vaccinated with the Pfizer-BioNTech COVID-19 vaccine and in a cross-sectional participants vaccinated with Pfizer-BioNTech/Moderna COVID-19 vaccine, and the analysis of differential antibody signal in serum drawn before vaccination and at defined intervals after receipt of the first and second vaccine dose was performed.

The longitudinal participants showed a strong IgM response to a S1+S2 ECD fragment and the S1 subunit detected at the 12 days after the first dose and the 3 days after the second dose, respectively, and dramatically increased 14 days after the second dose only to start gradually declining thereafter (**Figure 12**). Weaker IgM responses were detected to the S2 ECD region and the S1 NTD and CTD regions at 12 days and 7 days after the first and second dose, respectively, but there was not an increased response with time following the second dose. IgG binding responses were detected for the S1+S2 ECD and the S1 subunit regions, similar to the results from infected individuals, and binding to these regions peaked at 7 days after the second dose before revealing a rapid decline.

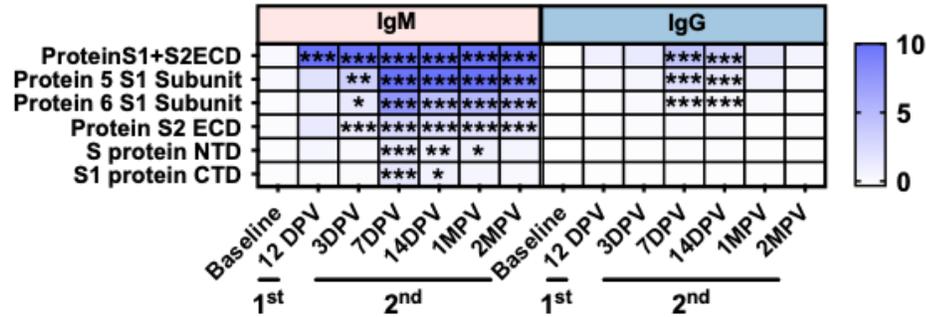


Figure 12. S protein and protein fragment antibody response in SARS-CoV-2 RNA vaccine participants.

Heatmap of the mean IgM and IgG signal in the longitudinal vaccinated participants for the SARS-CoV-2 S proteins and protein fragments at the indicated time points. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, by repeated measure ANOVA with Dunnett multiple comparison test. (n = 5)

The longitudinal participants showed a significant IgM response to sixteen unique peptides (**Figure 13A and Table S7**). Of these sixteen there were five duplicates, S551-565, S621-635, S811-825, S1161-1175, and S1181-S1195, in which the same peptide from two different variants were identified. Three of the five duplicates, S551-565, S811-825, and S1181-S1195, showed a markedly higher response that peaked at 14 days after the 2nd dose, and slightly declined thereafter.

Mapping of these peptides on the S protein structure revealed a different pattern than what was observed in the infected individuals (**Figure 13B**). Between the two cohorts, there was an overlap of six of the 16 unique peptides detected in each group, S241-255, S541-565, S571-585, S621-635, S781-795, and S811-825. Further, peptides detected in these groups formed near-continuous linear clusters specific for each of these group (**Table 2**). Five peptides detected in the infected group localized to two clusters, S661-705, and S761-825, and nine peptides identified in the vaccine group mapped to three clusters, S541-585, S621-645, and S1121-1195. The two clusters detected in the infected group mapped to the S1-S2 junction and contained the S2 region functional domain of the fusion peptide motif, while the clusters detected in the

vaccinated group mapped the adjacent to the RBD domain, within the S1 C-terminal region, and the HR2. Despite this differential pattern, three of the five clusters contained peptides that were detected in both groups, suggesting that these differences represented preferential rather than absolute differences in site recognition, which is supported by the results from the SARS-CoV-2-infected NHPs group, which detected multiple peptides in all but one of these clusters.

Table 2. S protein epitope clusters detected in NHP, COVID-19 and vaccinated participants

Cluster	NHP	COVID-19 patients	Vaccinated participants
	121	--	121
	241	241	221
	301	301	--
	351	--	351
	431	--	431
	481	481	--
Cluster1	541	--	541
	551	551	551
	561	--	561
	--	571	571
Cluster2	621	621	621
	631	--	631
Cluster3	661	661	--
	691	691	--
Cluster4	--	761	--
	781	781	781
	791	791	--
	801	801	--
	--	811	811
	--	921	--
	--	--	1001
	--	1101	--
Cluster5	--	--	1121
	--	--	1161
	--	--	1181
	--	1251	--

Numbers indicate S protein starting positions of detected peptides

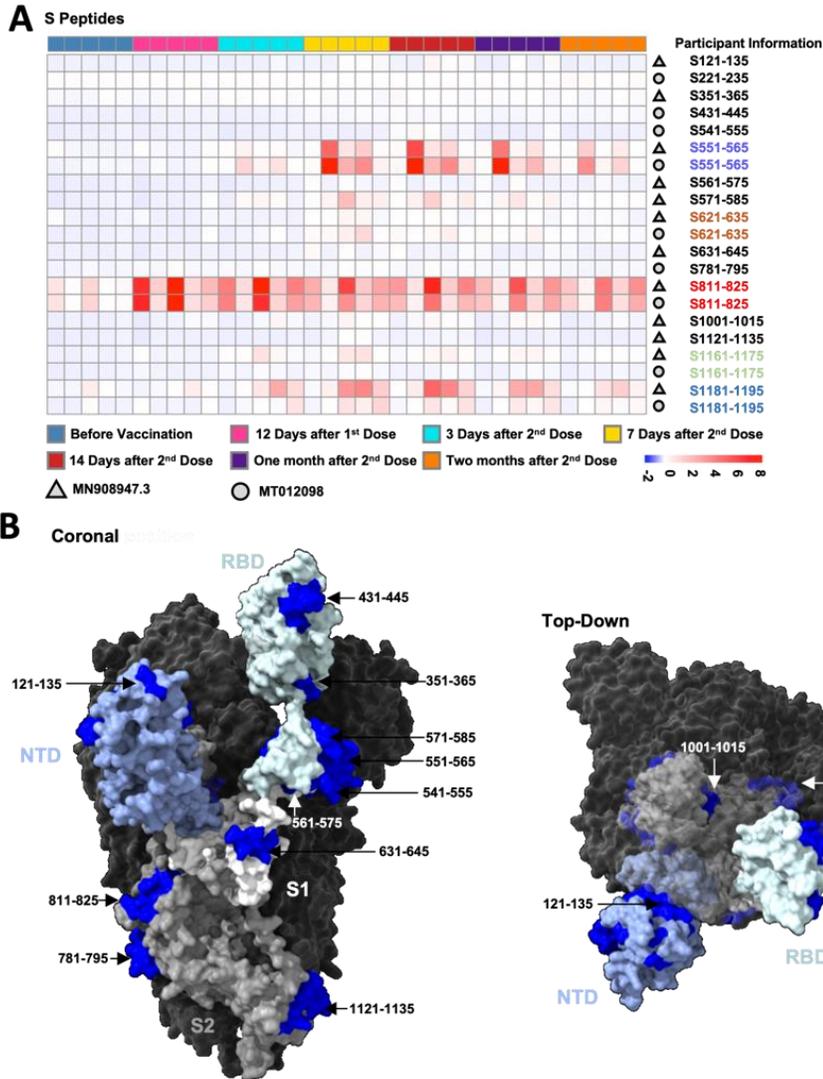


Figure 13. S peptide mapping of the antibody response in longitudinal SARS-CoV-2 mRNA vaccine participants.

A. Heatmap of IgM signal Z-scores for the SARS-CoV-2 S peptides using serum collected at the indicated times before and after vaccination. (n = 5) Colored labels indicate signals detected from overlapping peptide sequences of SARS-CoV-2 strains MT012098 (circles) and MN908947.3 (triangles).

B. 3D structural map (PDB:6VSB) indicating IgM binding to linear peptide epitopes in the N-terminal domain (NTD), receptor binding domain (RBD), S1, and S2 peptide sequence regions of the modified SARS-CoV-2 S protein (K986P and V987P) of the Pfizer and Moderna vaccines.

Analysis of IgM epitopes detected in a second, cross-sectional group of 15 vaccinated participants with blood samples collected from 7 to 30 days after their second vaccine dose, with four being vaccinated after a prior SARS-CoV-2 infection, identified all the peptides detected in the first vaccine cohort as well as one peptide, S661-675, that was originally detected on in the infected cohort. In total, the vaccinated participants identified six peptides, with only a few individuals showing a strong response to these six peptides, and the vaccinated post infection (VPI) individuals showing a significant response to thirty peptides (**Figure 14A and Table S7**). VPI individuals had more peptides with higher Z-scores, including five of the six peptides detected in both the infected and vaccinated cohorts, as well as the S661-675 peptide (**Figure 14B and Table S7**).

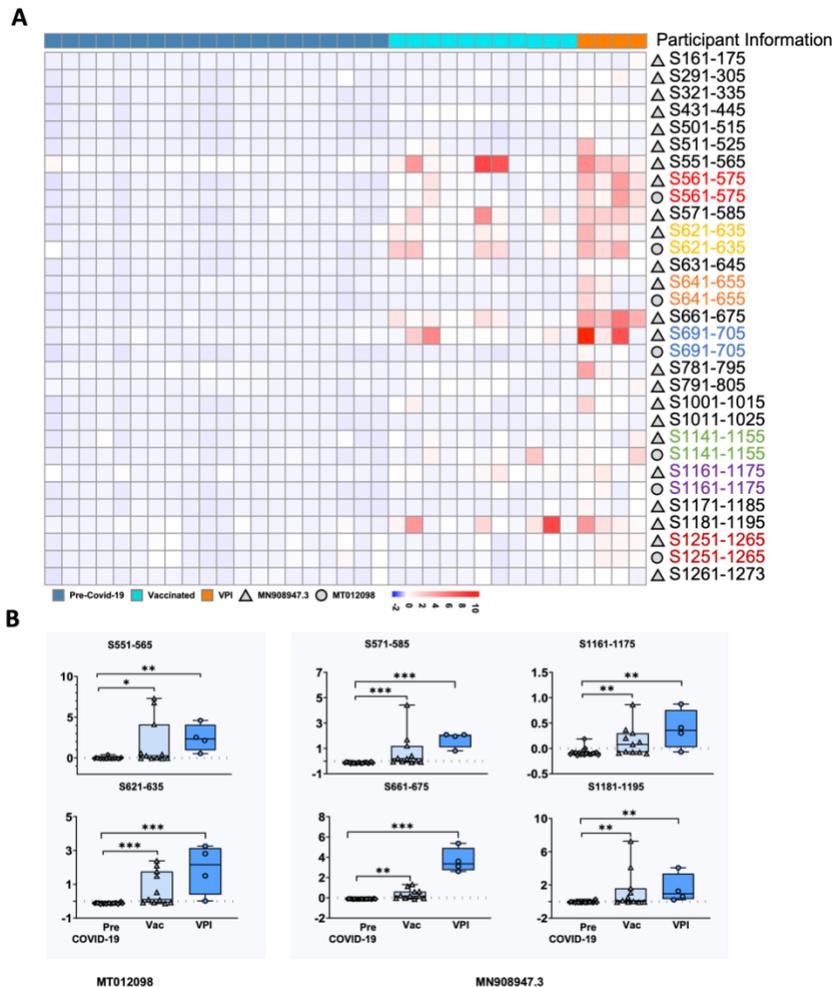


Figure 14. S peptide mapping of the antibody response in cross sectional SARS-CoV-2 RNA vaccine participants and VPI participants.

A. Heatmap of IgM signal Z-scores for the SARS-CoV-2 S peptides using serum collected after vaccination (n=11) and vaccination post infection (VPI) (n=4). Colored labels indicate signals detected from overlapping peptide sequences of SARS-CoV-2 strains MT012098 (circles) and MN908947.3 (triangles).

B. Relative IgM signal Z-scores for the strong linear peptide epitope signals in the vaccinated and VPI participants.

Graphs indicate each individual value, minimum, lower quartile, median, upper quartile and maximum, *p<0.05, **p<0.01, ***p<0.001, by repeated measure ANOVA with Dunnett post test.

Similar to the COVID-19 cohort, most individuals in the longitudinal vaccination cohort revealed strong binding to proteins derived from common respiratory viruses, including binding to several human coronaviruses, which modestly decreased in the post-vaccination group (**Figure 15A**). A moderate increase in binding to MERS-CoV S protein fragments, and a greater increase in binding to the S protein RBD of the SARS-CoV virus was detected, unlike results from the COVID cohort (**Figure 11A and 15A**). These increases were transient and peaked at 7 days after the second vaccine dose, although the IgM response to the MERS-CoV S1+S2 ECD region was not altered by vaccination (**Figure 15B**).

Binding of common respiratory viruses in the cross sectional vaccinated and VPI participants showed similar results to that of the longitudinal group (**Figure 16A**). Significant binding was found in SARS-CoV spike RBD, MERS-CoV spike S1 subunit, and MERS-CoV spike S2 subunit for the VPI participants, but only in the SARS-CoV spike RBD and MERS-CoV spike S2 for the cross-sectional vaccinated participants (**Figure 16B**).

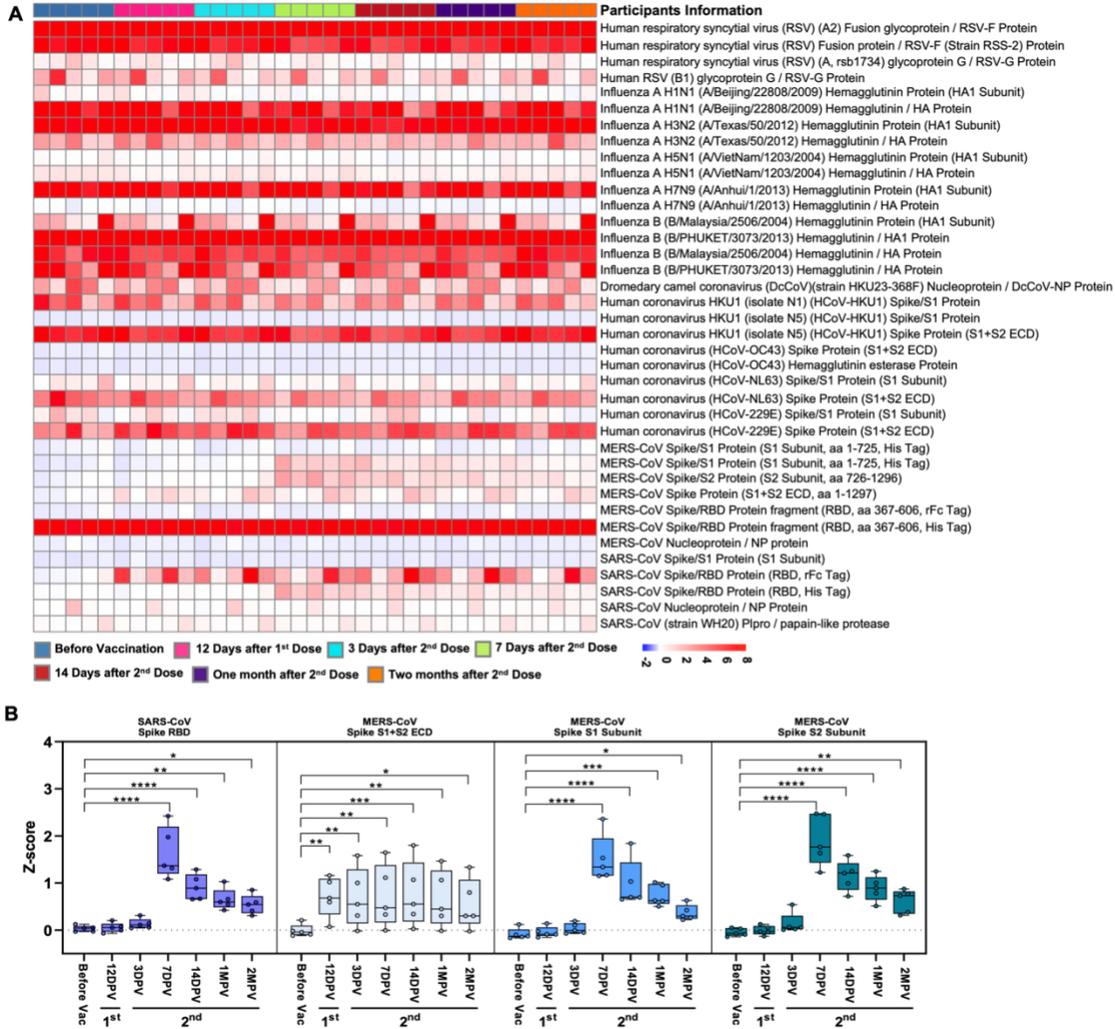


Figure 15. Antibody response to non-SARS-CoV-2 respiratory viruses in longitudinal vaccinated participants.

A. Heatmap of IgM signal Z-scores for non-SARS-CoV-2 respiratory viruses' proteins and protein fragments using serum collected at the indicated times before and after vaccination. (n = 5).

B. Relative IgM signal Z-scores for SARS-CoV and MERS-CoV S protein regions in longitudinal vaccinated participants.

Graphs indicate each individual value, minimum, lower quartile, median, upper quartile and maximum, *p<0.05, **p<0.01, ***p<0.001, by repeated measure ANOVA with Dunnett post test.

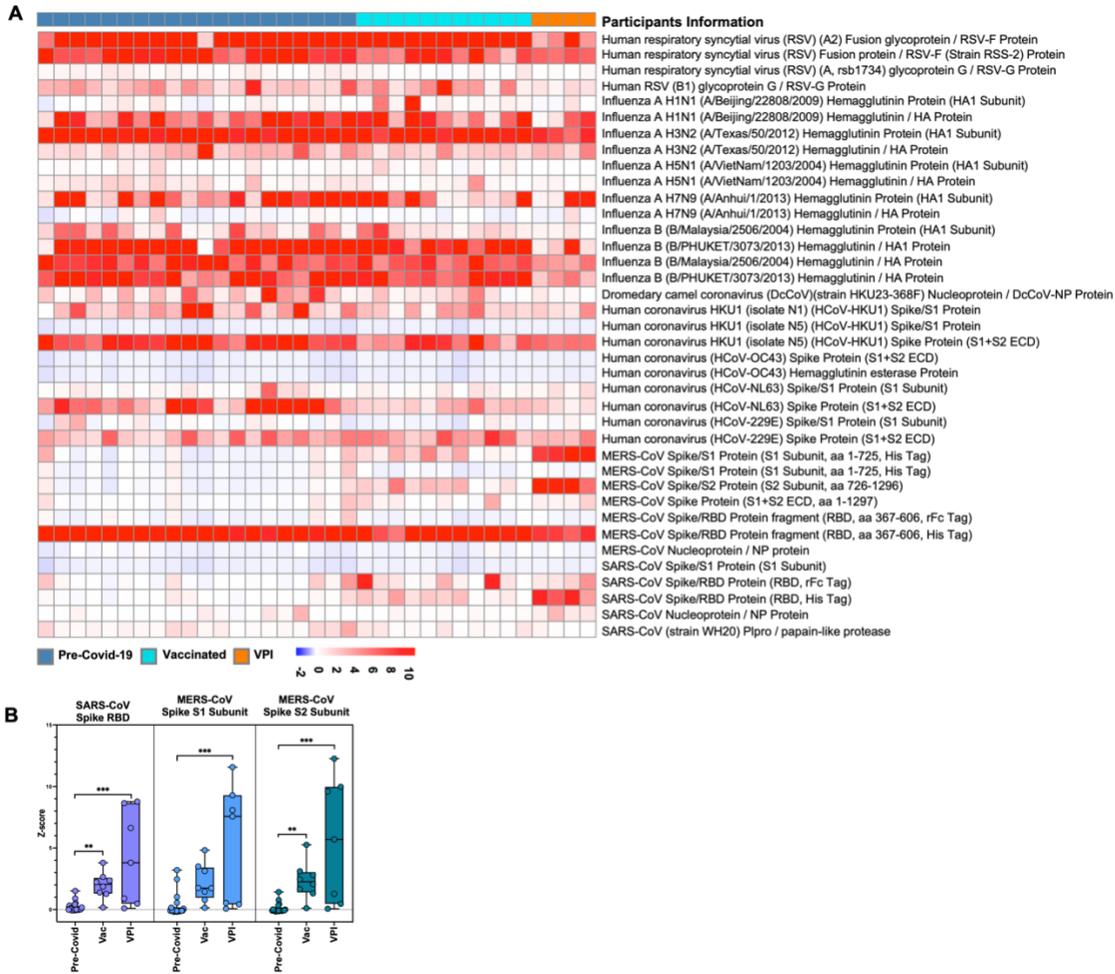


Figure 16. Antibody response to non-SARS-CoV-2 respiratory viruses in vaccinated, and VPI participants.

A. Heatmap of IgM signal Z-scores for non-SARS-CoV-2 respiratory viruses' proteins and protein fragments using serum collected after vaccination and VPI.

B. Relative IgM signal Z-scores for SARS-CoV and MERS-CoV S protein regions in vaccination and vaccinated post infection participants.

Graphs indicate each individual value, minimum, lower quartile, median, upper quartile and maximum, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, by one way ANOVA with Dunnett post test.

Overlapped epitopes in COVID-19 patients and vaccinated individuals

Comparison of peptides detected in COVID-19 patients, vaccination, and VPI showed a substantial overlapping of peptides, although this varied by groups with only five peptides shared among all three (**Figure 17A**). The degree of overlap varied between the infected and vaccinated

(22.2%; 6/27 total), infected and VPI (25.8%; 8/31), and vaccinated and VPI (33.3%; 10/30) individuals, with only five peptides being shared among all three groups. Peptides detected in the vaccinated group tended to cluster to the S1 NTD and RBD, and to the S2 HR1 to HR2 region (**Figure 17B and Table 3**). Peptides in both the COVID-19 cohort and VPI groups clustered to the S2 region outside of the defined functional domains, with those found in vaccinated and VPI groups being detected at multiple functionally regions, including the RBD and HR1 and the central helix (CH) domains, as well as two sites in the C-terminus of the S1 subunit. Five peptides were detected in all three groups, and all but one of these peptides mapped to sites within the S1 subunit C-terminus, with the remaining epitope mapping to the S2 subunit N-terminus. Analysis of the top 12 peptides, defined by Z-score and p-value, found that seven were detected in all the groups at a similar signal intensity, three exclusively detected in the infected group, and two others preferentially in the vaccinated and VPI groups (**Figure 17C**). The three peptides preferentially detected in the infected group also demonstrated a strong signal in the VPI group but failed to significantly differ from the vaccinated group. The infected group's enriched peptides corresponded to the RBD and adjacent to the fusion peptide motif regions, while the vaccinated and VPI groups' enriched epitopes corresponded to the junction between the S1 and S2 subunits and to the HR2 region, which may reflect differential epitope exposure due to conformational differences in the virus- and vaccine-encoded S proteins as result of the two proline substitutions introduced in the mRNA vaccines.

Table 3. IgM S peptide among the COVID-19 patients, vaccinated and VPI participants

Domain	COVID-19 patients	Vaccinated participants	VPI participants
S1 NTD	--	S121-135	--
	--	--	S161-175
	--	S221-235	--
	S241-255	--	--
	--	--	S291-305
	S301-315	--	--

RBD	--	--	S321-335
	--	S351-365	--
	--	S431-445	S431-445
	S481-495	--	--
	--	--	S501-515
	--	--	S511-525
	--	S541-555	--
S1 CTD	S551-565	S551-565	S551-565
	--	S561-575	S561-575
	S571-586	S571-588	S571-587
	S621-635	S621-635	S621-635
	--	S631-645	S631-645
	--	--	S641-655
	S661-675	S661-675	S661-675
S2 NTD	S691-705	--	S691-705
	S761-775	--	--
	S781-795	S781-795	S781-795
	S791-805	--	S791-805
	S801-815	--	--
	S811-825	S811-825	--
HR1	S921-935	--	--
CH	--	S1001-1015	S1001-1015
	--	--	S1011-1025
S2 Internal	S1101-1115	--	--
	--	S1121-1135	--
	--	--	S1141-1155
HR2	--	S1161-1175	S1161-1175
	--	--	S1171-1185
	--	S1181-1195	--
CTD	S1251-1265	--	S1251-1265
	--	--	S1261-1273

Numbers indicate S protein starting positions of detected peptides. Labels indicate the N-terminal, ribosome binding, and C-terminal domains (NTD, RBD, CTD) of the S1 or S2 region and the central helix (CH), heptad repeats 1 and 2 (HR1 and HR2) of the S2 region. VPI: Vaccination post infection.

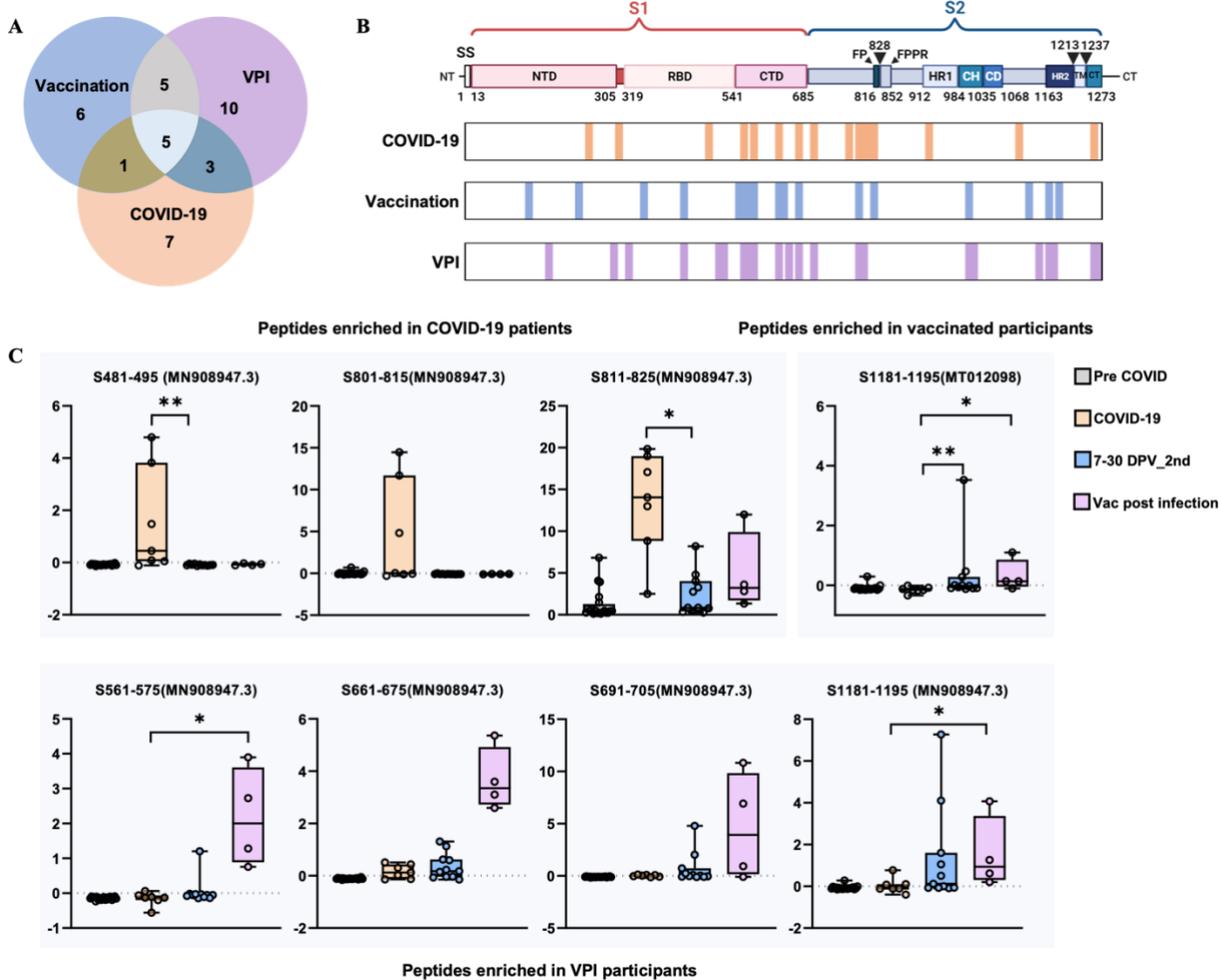


Figure 17. Difference in IgM linear peptide epitopes detected in the COVID-19, vaccinated and VPI groups.

A. Venn diagram of the overlap between IgM linear peptide epitopes detected in SARS-CoV-2 patients (COVID-19), vaccinated individuals (Vac), and individuals vaccinated post-infection (VPI).

B. SARS-CoV-2 S protein schematic indicating sites of IgM linear peptide epitopes specific for and shared among the indicated groups.

C. Relative Z-score differences for IgM linear peptide epitopes preferentially detected in COVID-19 (n = 7), vaccinated (n = 11) and VPI (n = 4) individuals.

Graphs indicate each individual, minimum, lower quartile, median, upper quartile and maximum values; *p<0.05, **p<0.01, ***p<0.001, by non-parametric one-way ANOVA with Dunn's post-test.

Discussion:

An improved understanding of antibody epitopes that block or neutralize critical features of the SARS-CoV-2 S and N protein is needed to guide the design of new vaccine targets and the development of new monoclonal antibody therapeutics, especially with the emergence of mutant

variants. However, most current antibody screening methods are limited in their ability to precisely define identifies epitopes. Employing microarrays with overlapping peptides derived from target proteins for screening of patient serum provide a mean for identification of linear peptide epitopes with great sequence coverage, exact localization, low selection bias and high throughput. Although the proteome microarray lacks the ability to detect conformational epitopes that lack extended linear sequence regions, one study estimates that these antibodies only account for less than 25% of those produced in response to recombinant protein antigens ⁵⁶.

While it is possible that sequence fragmentation among overlapping peptides could lead some epitopes being missed, the reduction in antibody coverage from this effect should be small since 85% of epitopes with linear components recognize five or more contiguous amino acids ⁵⁷, and most recognizing epitopes defined by seven to nine amino acids ⁵⁸. Previously studies that have analyzed linear epitopes selected and screened them using immunoinformatic analysis ⁵⁹⁻⁶³ or identified them by cross-sectional analyses of COVID-19 patients ²⁴⁻²⁶ and primarily focused on IgG linear epitopes. However, multiple studies have reported that multimeric antibodies, and particularly IgM, play major roles in SARS-CoV-2 neutralization ⁴²⁻⁴⁴.

We focused on the IgM response to the S fragments and peptides due to the importance of the S protein in SARS-CoV-2 infection and since IgM epitope responses were detected at greater abundance and with greater intensity than IgG responses following infection or vaccination. Since IgG antibodies come from the IgM population by isotype switch recombination, this difference could result from differential selection and/or from a difference of avidity for same epitopes. Also, immune selection may favor IgG antibodies that recognize non-linear SAR-CoV-2 epitopes or linear epitopes with conformations constrained by protein secondary structure. These conformations are unlikely to be present in the peptides analyzed. The lack of stable

secondary structure negatively affects the association and dissociation rates of antibody-peptide complexes and differentially destabilize IgG versus IgM antibody-peptide complexes due to the more limited ability of monomeric IgG to simultaneously interact with more than one peptide on the array than pentameric IgM.

We found that the IgM linear epitopes for the S protein sequences in the SARS-CoV-2 infected NHPs had a high degree of overlap (81%) with those detected in COVID-19 patients. There were substantially more IgM epitopes detected in the NHP cohort than the COVID-19 cohort, which is partially due to the reduced signal variability in the NHPs compared to the patients, with these epitopes clustering in the S1 region whereas epitopes detected in COVID-19 patients showed similar S1 and S2 region distributions, suggesting the potential for species-specific differences in epitope recognition. Most epitopes strongly detected in both species, however, tended to localize to functionally important regions on the S protein and align with reported neutralizing antibody binding sites.

The S protein IgM linear epitopes detected in serum from vaccinated individuals also revealed a clustered epitope distribution pattern that partially differed from the one detected with serum from COVID-19 patients, although at least one strongly detected epitope in each cluster was shared among these groups and the NHP cohort, which suggests that the differences observed between the groups did not affect the recognition of several epitopes associated with neutralizing activity despite identifying sites that were specifically detected in only in individuals after SARS-CoV-2 infection with or without subsequent vaccination.

Conserved sites that were detected tended to map to be associated with neutralizing activity and map to three major S protein regions. These regions include the S protein RBD, its cleavage

sites, or sequences associated with its cell entry mechanism, including its fusion peptide, and HR1 and HR2.

IgM from the SARS-CoV-2-infected NHP cohort showed almost complete coverage of the array peptides that mapped to the RBD region, which could be from it having fewer glycosylation sites than the rest of the S protein sequence^{64,65}. Multiple antibodies that bind the SARS-CoV-2 RBD have been reported to block viral entry by rendering the RBD unable to bind to ACE2^{50,51,66,67}.

One detected epitope, S481-495, from both SARS-CoV-2-infected NHPs and COVID-19 patients that partially overlapped the binding site of two neutralizing antibodies previously isolated from COVID-19 patients, S2H13²⁹ and F2B-2F6³⁰. Several IgM epitopes were also detected on peptides that mapped to S protein cleavage sites that are recognized by host proteases and associated with virus membrane fusion activity^{68,69}, including polybasic and multi-basic furin cleavage that are partially cleaved during viral production in host cells to promote S protein membrane fusion activity and viral transmission^{7,69-71}. This included a peptide, S661-675, that mapped to a furin recognition site between the S1 and S2 subunits, where antibody binding has been reported to block subunit separation to inhibit conformational changes associated with membrane fusion^{52,53}. Another detected peptide, S801-815, mapped to an S2 subunit site adjacent to the viral fusion peptide motif, where the steric hinderance from a bound antibody could inhibit protease interaction and subsequently inhibit the cleavage event required for fusion peptide insertion into the plasma membrane prior to cell entry. The SARS-CoV fusion peptide motif SFIEDLLFNKV demonstrates strong sequence with that of SARS-CoV-2 and single amino acid substitutions within this sequence can inhibit its membrane fusion activity⁷². A fourth peptide, S811-825, contained this sequence and was strongly bound by serum obtained from SARS-CoV-2 infected and vaccinated individuals. The IgM signal for this peptide was

stronger in infected versus vaccinated individuals, which could be the result of a S protein conformational difference introduced by two proline substitutions in the mRNA vaccines. Further, S811-825 is next to the fusion-peptide proximal region (FPPR), which tightly packs around an internal disulfide bond between cysteines located at S840 and S851⁵⁴. The modification of the S protein sequence in the mRNA vaccines exhibits a looser conformation than the wild-type protein because the K986P mutation results in a net negative charge at the trimer interface⁵⁴, which could affect FPPR conformation and the conformation of the S811-825 epitope. Several other detected peptides, including S1181-1195, were found to map to the two heptad repeat regions (HR1 and HR2) that interact to form six helix fusion cores which bring the viral and cell membranes together to permit membrane fusion^{55,73}. Inhibition of this interaction via steric or conformational effects from the antibody binding is expected to attenuated virus infectivity, similar to the function of the EK1 fusion inhibitor, a peptide inhibitor designed to interact with HR1 to block its ability to bind HR2 to inhibit membrane fusion and cell entry^{74,75}. The efficacy of monoclonal antibody therapy is still being evaluated despite significant in vitro data showing that neutralizing antibodies reduce the infectivity of SARS-CoV-2. However, there is concern that the effectiveness of such therapeutics will be reduced by emerging SARS-CoV-2 variants and strategies that employ multiple neutralizing antibodies targeting distinct epitopes may be required to reduce the effect of mutations on therapeutic efficacy^{76,77}. Given that the RBD exhibits significant mutation, the identification of epitopes that overlap or may alter the conformation of other functional regions will be useful in identifying therapeutic targets for neutralizing antibodies or inhibitors. HR1 and HR2 peptides, and potentially other epitopes identified here that map to or adjacent to sites that regulate the S protein, can serve as target sites for such therapeutics. This includes epitopes shared among infected and vaccinated individuals

that map to the CTD1, S551-565 and S561-575, and to the CTD2, S621-635 and S661-675). In the S protein's 3D conformation both of these regions are located below the RBD and against the S2 and NTD regions ⁷⁸, therefore any antibody binding to them may alter structural interactions among these regions to affect S protein activity. Evidence for this comes from how the CTD1 mediates an interaction between the RBD and FPPR regions for structural rearrangement regulation in such a way that favors membrane fusion activity ⁵⁴, while a CTD2 interaction stabilizes cleaved S1-S2 protein complexes preventing S1 from dissociating ⁷⁹ as well as subsequent S2 structural rearrangements that are involved in membrane fusion ^{80,81}.

For the N protein, we found that the linear epitopes detected on the N protein had high degree of overlap (91.7%) between the SARS-CoV-2 infected NHPs and COVID-19 infected cohorts. The difference in epitopes detected may partially come from the reduced signal variability in the NHPs compared to the SARS-CoV-2 patients. Like the S protein, the N protein had persistent epitopes clustering to the NTD/RNA binding region, the linker region, and the CTD/dimerization domain while the COVID-19 patients showed dispersed epitopes with only a cluster at the CTD/dimerization domain, suggesting the potential for species-specific differences in epitope recognition.

The linear peptide epitopes on the N protein may not be useful in blocking the virus from infecting cells, however they may be useful once the virus is inside the cell. Antibody therapies have been developed for intracellular targets⁸², meaning that the N protein could possibly be targeted as well. The linear peptide epitopes that map to the RNA binding, N91-105, N141-155, and N161-175, and the dimerization domains, N351-365 and N361-375, may hinder the function of the N protein when bound by antibodies. Specifically, the linear peptide epitopes N91-105 and N161-175 were found to have amino acids, K102, D103, L104, S105, G170, P171, and T172,

that were found to interact with RNA⁸³. Another use for these peptides, may be for use as diagnostic markers in diagnostic kits instead of using full length proteins as with SARS-CoV⁸⁴. Our results suggests that the use of a proteomic microarray allows for high-resolution mapping of linear epitopes, their potential functional importance in blocking and neutralization, while also providing valuable information about the adaptive immune response to SARS-CoV-2. However, there are limitations in this study that involve the cohort size, statistical analysis, and antibody binding issues. The size of the cohorts was relatively small, and the array results required normalization, both of which limited the statistical analysis' ability to detect epitopes that had lower affinities with their antibodies or antibodies that had a lower concentration. The COVID-19 patient cohort was greatly affected by the lack of information available regarding exposure dose, infection interval, and virus sequence. To obtain this information a large cohort study performed on at-risk groups exposed to a dominant virus variant or who are evaluated to determine viral sequence information is needed. This type of study is warranted due to the degree of overlap observed between the epitopes detected in infected NHPs and the distinct epitope patterns observed between infected and vaccinated individuals, which suggested a differential rather than an absolute difference that results from reduced epitope coverage. The last limitation is that this analysis focused on IgM epitopes given their greater abundance and signal intensity than IgG epitopes, potentially due to the greater avidity of IgM during competition for same sites. This should not be an issue because IgG epitopes is derived from the IgM pool before subsequent affinity maturation and selection. The limited sample sizes prevented a separate analysis of IgM and IgG in serum samples depleted of the other isotype, but this analysis could be performed in future studies to further confirm the epitopes found here. Finally, further studies are needed to confirm if conserved IgM epitopes detected at functional sites have functional

consequences for S protein activity and thus SARS-CoV-2 virulence when bound to by antibodies.

Overall, our study findings indicate that microarray screening can provide valuable information about the adaptive immune response to SARS-CoV-2 by allowing high resolution mapping of linear peptide epitopes. Our results showed a strong correspondence among linear peptide epitopes NHPs and patients following infection with strongly detected linear peptide epitopes being conserved and mapping to functionally important regions. As well as the different intensities between the infected and vaccinated groups aligned with potential differences in protein conformation among the native and recombinant protein. Taken together, we believe these results indicate the strong potential of employing microarray mapping data with structural and function information to identify candidate targets for therapeutic interventions.

Supplementary Information

[Supplementary Tables](#)

References

- 1 Schmidt, F. *et al.* High genetic barrier to SARS-CoV-2 polyclonal neutralizing antibody escape. *Nature* **600**, 512-516, doi:10.1038/s41586-021-04005-0 (2021).
- 2 Arya, R. *et al.* Structural insights into SARS-CoV-2 proteins. *J Mol Biol* **433**, 166725, doi:10.1016/j.jmb.2020.11.024 (2021).
- 3 Lan, J. *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* **581**, 215-220, doi:10.1038/s41586-020-2180-5 (2020).
- 4 Shajahan, A., Supekar, N. T., Gleinich, A. S. & Azadi, P. Deducing the N- and O-glycosylation profile of the spike protein of novel coronavirus SARS-CoV-2. *Glycobiology* **30**, 981-988, doi:10.1093/glycob/cwaa042 (2020).
- 5 Wang, Q. *et al.* Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. *Cell* **181**, 894-904 e899, doi:10.1016/j.cell.2020.03.045 (2020).
- 6 Watanabe, Y., Allen, J. D., Wrapp, D., McLellan, J. S. & Crispin, M. Site-specific analysis of the SARS-CoV-2 glycan shield. *bioRxiv*, doi:10.1101/2020.03.26.010322 (2020).
- 7 Walls, A. C. *et al.* Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **183**, 1735, doi:10.1016/j.cell.2020.11.032 (2020).

- 8 Gao, T. *et al.* Identification and functional analysis of the SARS-CoV-2 nucleocapsid protein. *BMC Microbiol* **21**, 58, doi:10.1186/s12866-021-02107-3 (2021).
- 9 Kang, S. *et al.* Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm Sin B* **10**, 1228-1238, doi:10.1016/j.apsb.2020.04.009 (2020).
- 10 Zeng, W. *et al.* Biochemical characterization of SARS-CoV-2 nucleocapsid protein. *Biochem Biophys Res Commun* **527**, 618-623, doi:10.1016/j.bbrc.2020.04.136 (2020).
- 11 Gu, W. Y. *et al.* Short hairpin RNAs targeting M and N genes reduce replication of porcine deltacoronavirus in ST cells. *Virus Genes* **55**, 795-801, doi:10.1007/s11262-019-01701-y (2019).
- 12 Neuman, B. W. *et al.* A structural analysis of M protein in coronavirus assembly and morphology. *J Struct Biol* **174**, 11-22, doi:10.1016/j.jsb.2010.11.021 (2011).
- 13 Schoeman, D. & Fielding, B. C. Coronavirus envelope protein: current knowledge. *Virology* **16**, 69, doi:10.1186/s12985-019-1182-0 (2019).
- 14 Corbett, K. S. *et al.* Evaluation of the mRNA-1273 Vaccine against SARS-CoV-2 in Nonhuman Primates. *N Engl J Med* **383**, 1544-1555, doi:10.1056/NEJMoa2024671 (2020).
- 15 Corbett, K. S. *et al.* Evaluation of mRNA-1273 against SARS-CoV-2 B.1.351 Infection in Nonhuman Primates. *bioRxiv*, doi:10.1101/2021.05.21.445189 (2021).
- 16 Klasse, P. J., Nixon, D. F. & Moore, J. P. Immunogenicity of clinically relevant SARS-CoV-2 vaccines in nonhuman primates and humans. *Sci Adv* **7**, doi:10.1126/sciadv.abe8065 (2021).
- 17 Sun, S. H. *et al.* A Mouse Model of SARS-CoV-2 Infection and Pathogenesis. *Cell Host Microbe* **28**, 124-133 e124, doi:10.1016/j.chom.2020.05.020 (2020).
- 18 Long, Q. X. *et al.* Antibody responses to SARS-CoV-2 in patients with COVID-19. *Nat Med* **26**, 845-848, doi:10.1038/s41591-020-0897-1 (2020).
- 19 Suthar, M. S. *et al.* Rapid Generation of Neutralizing Antibody Responses in COVID-19 Patients. *Cell Rep Med* **1**, 100040, doi:10.1016/j.xcrm.2020.100040 (2020).
- 20 Baden, L. R. *et al.* Efficacy and Safety of the mRNA-1273 SARS-CoV-2 Vaccine. *N Engl J Med* **384**, 403-416, doi:10.1056/NEJMoa2035389 (2021).
- 21 Polack, F. P. *et al.* Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. *N Engl J Med* **383**, 2603-2615, doi:10.1056/NEJMoa2034577 (2020).
- 22 Tilocca, B. *et al.* Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. *Microbes Infect* **22**, 188-194, doi:10.1016/j.micinf.2020.04.002 (2020).
- 23 Yarmarkovich, M., Warrington, J. M., Farrel, A. & Maris, J. M. Identification of SARS-CoV-2 Vaccine Epitopes Predicted to Induce Long-Term Population-Scale Immunity. *Cell Rep Med* **1**, 100036, doi:10.1016/j.xcrm.2020.100036 (2020).
- 24 Amrun, S. N. *et al.* Linear B-cell epitopes in the spike and nucleocapsid proteins as markers of SARS-CoV-2 exposure and disease severity. *EBioMedicine* **58**, 102911, doi:10.1016/j.ebiom.2020.102911 (2020).
- 25 Li, Y. *et al.* Linear epitopes of SARS-CoV-2 spike protein elicit neutralizing antibodies in COVID-19 patients. *Cell Mol Immunol* **17**, 1095-1097, doi:10.1038/s41423-020-00523-5 (2020).

- 26 Li, Y. *et al.* Linear epitope landscape of the SARS-CoV-2 Spike protein constructed from 1,051 COVID-19 patients. *Cell Rep* **34**, 108915, doi:10.1016/j.celrep.2021.108915 (2021).
- 27 Chen, X. *et al.* Human monoclonal antibodies block the binding of SARS-CoV-2 spike protein to angiotensin converting enzyme 2 receptor. *Cell Mol Immunol* **17**, 647-649, doi:10.1038/s41423-020-0426-7 (2020).
- 28 Suryadevara, N. *et al.* Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell* **184**, 2316-2331 e2315, doi:10.1016/j.cell.2021.03.029 (2021).
- 29 Piccoli, L. *et al.* Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell* **183**, 1024-1042 e1021, doi:10.1016/j.cell.2020.09.037 (2020).
- 30 Ju, B. *et al.* Human neutralizing antibodies elicited by SARS-CoV-2 infection. *Nature* **584**, 115-119, doi:10.1038/s41586-020-2380-z (2020).
- 31 Robbiani, D. F. *et al.* Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**, 437-442, doi:10.1038/s41586-020-2456-9 (2020).
- 32 Seydoux, E. *et al.* Characterization of neutralizing antibodies from a SARS-CoV-2 infected individual. *bioRxiv*, doi:10.1101/2020.05.12.091298 (2020).
- 33 Wu, Y. *et al.* A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. *Science* **368**, 1274-1278, doi:10.1126/science.abc2241 (2020).
- 34 Chi, X. *et al.* A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **369**, 650-655, doi:10.1126/science.abc6952 (2020).
- 35 Ahlen, G. *et al.* The SARS-CoV-2 N Protein Is a Good Component in a Vaccine. *J Virol* **94**, doi:10.1128/JVI.01279-20 (2020).
- 36 Bai, Z., Cao, Y., Liu, W. & Li, J. The SARS-CoV-2 Nucleocapsid Protein and Its Role in Viral Structure, Biological Functions, and a Potential Target for Drug or Vaccine Mitigation. *Viruses* **13**, doi:10.3390/v13061115 (2021).
- 37 Dutta, N. K., Mazumdar, K. & Gordy, J. T. The Nucleocapsid Protein of SARS-CoV-2: a Target for Vaccine Development. *J Virol* **94**, doi:10.1128/JVI.00647-20 (2020).
- 38 Zheng, Y. *et al.* SARS-CoV-2 NSP5 and N protein counteract the RIG-I signaling pathway by suppressing the formation of stress granules. *Signal Transduct Target Ther* **7**, 22, doi:10.1038/s41392-022-00878-3 (2022).
- 39 Zhao, Y. *et al.* A dual-role of SARS-CoV-2 nucleocapsid protein in regulating innate immune response. *Signal Transduct Target Ther* **6**, 331, doi:10.1038/s41392-021-00742-w (2021).
- 40 Pettersen, E. F. *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-1612, doi:10.1002/jcc.20084 (2004).
- 41 Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263, doi:10.1126/science.abb2507 (2020).
- 42 Gasser, R. *et al.* Major role of IgM in the neutralizing activity of convalescent plasma against SARS-CoV-2. *Cell Rep* **34**, 108790, doi:10.1016/j.celrep.2021.108790 (2021).
- 43 Klingler, J. *et al.* Role of Immunoglobulin M and A Antibodies in the Neutralization of Severe Acute Respiratory Syndrome Coronavirus 2. *J Infect Dis* **223**, 957-970, doi:10.1093/infdis/jiaa784 (2021).

- 44 Wang, Z. *et al.* Enhanced SARS-CoV-2 neutralization by dimeric IgA. *Sci Transl Med* **13**, doi:10.1126/scitranslmed.abf1555 (2021).
- 45 Musico, A. *et al.* SARS-CoV-2 Epitope Mapping on Microarrays Highlights Strong Immune-Response to N Protein Region. *Vaccines (Basel)* **9**, doi:10.3390/vaccines9010035 (2021).
- 46 Cubuk, J. *et al.* The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat Commun* **12**, 1936, doi:10.1038/s41467-021-21953-3 (2021).
- 47 Ye, Q., West, A. M. V., Silletti, S. & Corbett, K. D. Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci* **29**, 1890-1901, doi:10.1002/pro.3909 (2020).
- 48 Chang, C., Michalska, K., Jedrzejczak, R., Maltseva, N., Endres, M., Godzik, A., Kim, Y., Joachimiak, A. (RCSB Protein Data Bank, 2020).
- 49 Zhou, R. J., Zeng, R., Lei, J. (RCSB Protein Data Bank, 2020).
- 50 Casalino, L. *et al.* Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent Sci* **6**, 1722-1734, doi:10.1021/acscentsci.0c01056 (2020).
- 51 Poh, C. M. *et al.* Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat Commun* **11**, 2806, doi:10.1038/s41467-020-16638-2 (2020).
- 52 Vankadari, N. Structure of Furin Protease Binding to SARS-CoV-2 Spike Glycoprotein and Implications for Potential Targets and Virulence. *J Phys Chem Lett* **11**, 6655-6663, doi:10.1021/acs.jpcclett.0c01698 (2020).
- 53 Johnson, B. A. *et al.* Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* **591**, 293-299, doi:10.1038/s41586-021-03237-4 (2021).
- 54 Cai, Y. *et al.* Distinct conformational states of SARS-CoV-2 spike protein. *Science* **369**, 1586-1592, doi:10.1126/science.abd4251 (2020).
- 55 Tang, T., Bidon, M., Jaimes, J. A., Whittaker, G. R. & Daniel, S. Coronavirus membrane fusion mechanism offers a potential target for antiviral development. *Antiviral Res* **178**, 104792, doi:10.1016/j.antiviral.2020.104792 (2020).
- 56 Forsstrom, B. *et al.* Dissecting antibodies with regards to linear and conformational epitopes. *PLoS One* **10**, e0121673, doi:10.1371/journal.pone.0121673 (2015).
- 57 Kringelum, J. V., Nielsen, M., Padkjaer, S. B. & Lund, O. Structural analysis of B-cell epitopes in antibody:protein complexes. *Mol Immunol* **53**, 24-34, doi:10.1016/j.molimm.2012.06.001 (2013).
- 58 Buus, S. *et al.* High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays. *Mol Cell Proteomics* **11**, 1790-1800, doi:10.1074/mcp.M112.020800 (2012).
- 59 Tilocca, B. *et al.* Comparative computational analysis of SARS-CoV-2 nucleocapsid protein epitopes in taxonomically related coronaviruses. *Microbes Infect* **22**, 188-194 (2020).
- 60 Wang, D. L. *et al.* Immunoinformatic Analysis of T- and B-Cell Epitopes for SARS-CoV-2 Vaccine Design. *Vaccines-Basel* **8** (2020).
- 61 Yarmarkovich, M., Warrington, J. M., Farrel, A. & Maris, J. M. Identification of SARS-CoV-2 Vaccine Epitopes Predicted to Induce Long-Term Population-Scale Immunity. *Cell Rep Med* **1**, doi:ARTN 100036 10.1016/j.xcrm.2020.100036 (2020).

- 62 Bhattacharya, M. *et al.* Development of epitope-based peptide vaccine against novel coronavirus 2019 (SARS-CoV-2): Immunoinformatics approach. *J Med Virol* **92**, 618-631, doi:10.1002/jmv.25736 (2020).
- 63 Grifoni, A. *et al.* A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* **27**, 671-+, doi:10.1016/j.chom.2020.03.002 (2020).
- 64 Zhao, X., Chen, H. & Wang, H. Glycans of SARS-CoV-2 Spike Protein in Virus Infection and Antibody Production. *Front Mol Biosci* **8**, 629873, doi:10.3389/fmolb.2021.629873 (2021).
- 65 Barnes, C. O. *et al.* Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* **182**, 828-842 e816, doi:10.1016/j.cell.2020.06.025 (2020).
- 66 Liu, L. *et al.* Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* **584**, 450-456, doi:10.1038/s41586-020-2571-7 (2020).
- 67 Xu, H. *et al.* Structure-based analyses of neutralization antibodies interacting with naturally occurring SARS-CoV-2 RBD variants. *Cell Res*, doi:10.1038/s41422-021-00554-1 (2021).
- 68 Whittaker, G. R. SARS-CoV-2 spike and its adaptable furin cleavage site. *Lancet Microbe*, doi:10.1016/S2666-5247(21)00174-9 (2021).
- 69 Johnson, B. A. *et al.* Furin Cleavage Site Is Key to SARS-CoV-2 Pathogenesis. *bioRxiv*, doi:10.1101/2020.08.26.268854 (2020).
- 70 Xia, X. Domains and Functions of Spike Protein in Sars-Cov-2 in the Context of Vaccine Design. *Viruses* **13**, doi:10.3390/v13010109 (2021).
- 71 Wang, Q. *et al.* A Unique Protease Cleavage Site Predicted in the Spike Protein of the Novel Pneumonia Coronavirus (2019-nCoV) Potentially Related to Viral Transmissibility. *Virol Sin* **35**, 337-339, doi:10.1007/s12250-020-00212-7 (2020).
- 72 Madu, I. G., Roth, S. L., Belouzard, S. & Whittaker, G. R. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. *J Virol* **83**, 7411-7421, doi:10.1128/JVI.00079-09 (2009).
- 73 Harrison, S. C. Viral membrane fusion. *Nat Struct Mol Biol* **15**, 690-698, doi:10.1038/nsmb.1456 (2008).
- 74 Huang, Y., Yang, C., Xu, X. F., Xu, W. & Liu, S. W. Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacol Sin* **41**, 1141-1149, doi:10.1038/s41401-020-0485-4 (2020).
- 75 Wang, X., Xia, S., Zhu, Y., Lu, L. & Jiang, S. Pan-coronavirus fusion inhibitors as the hope for today and tomorrow. *Protein Cell* **12**, 84-88, doi:10.1007/s13238-020-00806-7 (2021).
- 76 Wang, L. S. *et al.* Importance of Neutralizing Monoclonal Antibodies Targeting Multiple Antigenic Sites on the Middle East Respiratory Syndrome Coronavirus Spike Glycoprotein To Avoid Neutralization Escape. *J Virol* **92**, doi:ARTN e02002-17 10.1128/JVI.02002-17 (2018).
- 77 Cao, Y. L. *et al.* Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells. *Cell* **182**, 73-+ (2020).

- 78 Zhang, J., Xiao, T., Cai, Y. & Chen, B. Structure of SARS-CoV-2 spike protein. *Curr Opin Virol* **50**, 173-182, doi:10.1016/j.coviro.2021.08.010 (2021).
- 79 Zhang, J. *et al.* Structural impact on SARS-CoV-2 spike protein by D614G substitution. *Science* **372**, 525-530, doi:10.1126/science.abf2303 (2021).
- 80 Barrett, C. T. *et al.* Effect of clinical isolate or cleavage site mutations in the SARS-CoV-2 spike protein on protein stability, cleavage, and cell-cell fusion. *J Biol Chem* **297**, 100902, doi:10.1016/j.jbc.2021.100902 (2021).
- 81 Letarov, A. V., Babenko, V. V. & Kulikov, E. E. Free SARS-CoV-2 Spike Protein S1 Particles May Play a Role in the Pathogenesis of COVID-19 Infection. *Biochemistry (Mosc)* **86**, 257-261, doi:10.1134/S0006297921030032 (2021).
- 82 Gaston, J. *et al.* Intracellular delivery of therapeutic antibodies into specific cells using antibody-peptide fusions. *Sci Rep* **9**, 18688, doi:10.1038/s41598-019-55091-0 (2019).
- 83 Khan, A. *et al.* Structural insights into the mechanism of RNA recognition by the N-terminal RNA-binding domain of the SARS-CoV-2 nucleocapsid phosphoprotein. *Comput Struct Biotechnol J* **18**, 2174-2184, doi:10.1016/j.csbj.2020.08.006 (2020).
- 84 Shang, B. *et al.* Characterization and application of monoclonal antibodies against N protein of SARS-coronavirus. *Biochem Biophys Res Commun* **336**, 110-117, doi:10.1016/j.bbrc.2005.08.032 (2005).