Check for updates

RESEARCH ARTICLE

Profiling of linear B-cell epitopes against human coronaviruses in pooled sera sampled early in the COVID-19 pandemic

[version 1; peer review: 1 approved with reservations]

Emil Bach[1], Mustafa Ghanizada[1], Nikolaj Kirkby[2], Søren Buus[1], Thomas Østerbye[1]

[1]Department of Immunology and Microbiology, University of Copenhagen, Copenhagen, 2200 N, Denmark
[2]Department of Clinical Microbiology, Copenhagen University Hospital, Copenhagen, 2100 Ø, Denmark

v1 **First published:** 11 Oct 2024, **4**:215
https://doi.org/10.12688/openreseurope.17620.1
Latest published: 11 Oct 2024, **4**:215
https://doi.org/10.12688/openreseurope.17620.1

Abstract

Background

Antibodies play a key role in the immune defence against infectious pathogens. Understanding the underlying process of B cell recognition is not only of fundamental interest; it supports important applications within diagnostics and therapeutics. Whereas conformational B cell epitope recognition is complicated to decipher, linear B cell epitopes offer a straightforward approach that can be reduced to peptide recognition.

Methods

We utilised an overlapping peptide approach encompassing the proteomes of the seven human-infecting coronaviruses. Pooled sera from eight PCR-confirmed COVID-19 convalescents and eight pre-pandemic controls were analysed. 13-mer peptides overlapping by 11 amino acids were synthesised and incubated with the pooled sera. Fluorochrome-labelled anti-IgG and -IgA antibodies were applied to detect antibody-peptide interactions. Strong antibody interactions spanning consecutive peptides were identified as 'high-fidelity regions' (HFRs) and mapped to coronavirus proteomes using a 60% homology threshold for clustering.

Results

Open Peer Review

Approval Status ?

	1
version 1 11 Oct 2024	? view

1. **Erin M Scherer**, Emory University School of Medicine, Atlanta, USA

Any reports and responses or comments on the article can be found at the end of the article.

We found 333 HFRs derived from human coronaviruses. Among these, 98 (29%) mapped to SARS-CoV-2, 144 (44%) to one or more common cold coronaviruses (CCC), and 54 (16%) cross-mapped to both SARS-CoV-2 and CCCs. The remaining 37 (11%) mapped to either SARS-CoV or MERS-CoV. Notably, COVID-19 serum favoured SARS-CoV-2-mapped HFRs, while pre-pandemic serum favoured CCC-mapped HFRs. The primary targets for linear B cell epitopes were the ORF1ab protein (60%), spike protein (21%), and nucleoprotein (15%) in absolute numbers; however, the order was reversed in terms of epitope density.

Conclusion

We identified linear B cell epitopes across coronaviruses, highlighting pan-, alpha-, beta-, or SARS-CoV-2-corona-specific B cell recognition patterns. These insights could aid in understanding past and present coronavirus exposures. Additionally, our results indicate potential cross-reactivity of pre-pandemic anti-CCC antibodies with SARS-CoV-2, possibly influencing COVID-19 outcomes. Lastly, our methodology offers a rapid and thorough approach to high-resolution linear B-cell epitope mapping, which is crucial for future studies of emerging infectious diseases.

Keywords
COVID-19, SARS-CoV2, antibody response, HD peptide array, linear epitopes



This article is included in the Horizon 2020 gateway.



This article is included in the Coronavirus collection.

Corresponding authors: Emil Bach (eb@immunaware.com), Thomas Østerbye (to@immunaware.com)

Author roles: Bach E: Conceptualization, Data Curation, Formal Analysis, Methodology, Visualization, Writing – Original Draft Preparation; **Ghanizada M**: Resources, Writing – Review & Editing; **Kirkby N**: Resources, Writing – Review & Editing; **Buus S**: Formal Analysis, Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **Østerbye T**: Conceptualization, Project Administration, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was supported by the European Commission (Electromed, project id. 862539).
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2024 Bach E *et al*. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Bach E, Ghanizada M, Kirkby N *et al.* **Profiling of linear B-cell epitopes against human coronaviruses in pooled sera sampled early in the COVID-19 pandemic [version 1; peer review: 1 approved with reservations]** Open Research Europe 2024, **4**:215 https://doi.org/10.12688/openreseurope.17620.1

First published: 11 Oct 2024, **4**:215 https://doi.org/10.12688/openreseurope.17620.1

Introduction

As of November 2023, the coronavirus disease 2019 (COVID-19) pandemic has resulted in more than 676 million infections and 6.8 million deaths worldwide (COVID-19 Map - Johns Hopkins Coronavirus Resource Center); a mortality rate of 1%[1]. The causative agent of COVID-19 and its genome was determined in January 2020, identifying a novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[2]. SARS-CoV-2 belongs to the Coronaviridae family, wherein seven virus strains infect humans, all causing respiratory tract infections[3]. These seven human coronaviruses (HCoVs) encompass two genera: Five are in the Betacoronavirus: HCoV-OC43, HCoV-HKU1 (sub-strains N1, N2, and N5), SARS-CoV, SARS-CoV-2, and Middle East respiratory syndrome coronavirus (MERS-CoV). Two are in the Alphacoronavirus genus: HCoV-NL63 and HCoV-229E[4]. OC43, HKU1, NL63, and 229E, are collectively known as the common cold coronavirus (CCC). The CCCs are endemic and responsible for 5–30% of common cold cases[3]. They cause mild respiratory symptoms and are considered non-fatal. In contrast, the epidemic SARS-CoV (emerged in 2002) and MERS-CoV (emerged in 2012) exhibited high mortality rates of approximately 9% and 36%, respectively.

Coronaviruses are enveloped, positive-sense RNA viruses. The SARS-CoV-2 genome contains 15 open reading frames (ORFs) encoding (1) the structural proteins making up the envelope of the virion, the spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins; and (2) the accessory proteins found in the polyproteins ORF1a and ORF1ab together with ORFs 3a, 3b, 6, 7a, 7b, 8, 9b, 9c, and 10[5]. This genomic architecture is similar among the HCoVs, which all share the polyproteins (ORF1a/ab) and structural proteins (S, E, M, and N)[5]. At the sequence level, their nucleotide sequence identities range from 47–80%[6]. Comparing S proteins, SARS-CoV-2 has the highest similarity with SARS-CoV, with a nucleotide sequence identity of 80.4% and protein sequence identity of 77.3%[6]. The HCoV S proteins facilitate their entrance into their host cells via interactions with host cell surface proteins[3,5,7], making them ideal targets for vaccines aimed at inducing neutralising antibodies. Indeed, the S protein is the main target in many SARS-CoV-2 vaccines in use and under development[8,9]. It follows that mutations in the S protein are particularly problematic for the longevity of the immunity induced by these vaccines.

Since its emergence, SARS-CoV-2 has accumulated genomic changes such as mutations, indels (insertions and deletions), and recombinations, resulting in novel variants[10]. Most proteomic mutations have occurred in the ORF1ab protein, followed by the S protein and then the N protein[11–13]. The SARS-CoV-2 Omicron 'complex' (PANGO lineages BA.1, BA.2, BA.3, BA.4, BA.5) developed more than 15 mutations in the S protein receptor-binding domain (RBD) and multiple genomic changes in the amino-terminal domain (NTD)[10]. These mutations have led to antigenic changes in the S protein, allowing variants in the Omicron complex to escape neutralising antibodies in sera from first-generation vaccinees and pre-Omicron convalescents. The dominant variant for the first half of 2023 was the XBB and its sub-lineages, originating from a recombination between two BA.2 sub-lineages[10,14]. The XBB variant exhibited even higher antibody evasion than its parental strains[15]. However, the XBB variant is currently in the process of being outcompeted by EG.5, a new descendant lineage of XBB.19.2[14]. This new variant shows even greater resistance than XBB to neutralisation by sera from vaccinees with or without breakthrough infections with other circulating subvariants compared to the original sub-variants in the Omicron complex[16–19]. Thus, the evasion of humoral immunity appears to be an important driver of SARS-CoV-2 evolution[15].

Antibodies are produced by B cells and the specific regions of an antigen that they recognise are known as B cell epitopes. The identification of virus-specific B cell epitopes is important for rational approaches to prevent, treat, diagnose, and understand virus infections. In general, antibodies against protein antigens can be broadly categorized into those that are specific for conformational (discontinuous) and those that are specific for linear (continuous) B cell epitopes; conformational B cell epitopes are composed of amino acids that are brought into proximity through the folded tertiary structure of a protein antigen, whereas linear B cell epitopes consist of a sequential string of amino acids from the primary structure of a protein antigen.

Vital for their infectivity, viruses use ligand(s) to find and attach to specific target(s) on appropriate host cells (e.g., SARS-CoV-2 uses its spike protein to specifically attach to ACE2 on the target cell). Antibodies that bind to a virus ligand and interfere with the infective virus-host cell interaction are likely to have virus-neutralising activity. For many natural infections, the corresponding neutralising B cell epitope is thought to be conformational. This is highlighted in the recent design of SARS-CoV-2 vaccines. Here, the vaccines have predominantly been targeted against the infective pre-fusion structure of the virus's spike protein. To enhance the vaccine's efficacy, specific mutations have been introduced to stabilise this crucial structure.

Whether a neutralising antibody recognises a conformational or linear B cell epitope has important theoretical and practical considerations. About 90% of B cell epitopes are conformational[20]. Unfortunately, identifying conformational B cell epitopes experimentally tends to involve demanding and low throughput structural studies such as X-ray crystallography of antibody-antigen complexes, NMR or deuterium exchange in the presence or absence of antibody, directed mutagenesis, etc. (reviewed in 21). Machine learning-based predictors of conformational B cell epitopes have been introduced[22,23] and are continuously improving[24–26]. Representing conformational epitopes (e.g. for formulation in vaccines or diagnostics) typically requires recombinant expression in organisms that allow proper folding and post-translational modifications. About 10% of B cell epitopes are linear[20], which typically can be represented by oligopeptides. Comparatively, it is straightforward to identify linear B cell epitopes since large peptide libraries can be generated synthetically (e.g. as peptide microarrays) or

recombinantly (e.g. as phage display libraries). Using overlapping peptide strategies, it is possible to cover entire pathogen proteomes, and using single substitution analysis, it is possible to delineate the length of individual B cell epitopes and define the underlying motifs with several discrete positions where certain amino acids are required, allowed, or disallowed.

We have pioneered a photolithographic peptide microarray technology, which can produce over a hundred thousand peptides per microarray[27]. Using this platform to screen for linear B cell epitopes, we have observed that the typical length of a linear B cell epitope is 5–10 amino acids (range 4–12 amino acids) and that it typically features a compound specificity with a few very dominant positions allowing very few amino acid substitutions, some less dominant position allowing several amino acids and a few "spacer" positions allowing all amino acids[27–29]. Notably, our studies revealed that the specificity of linear B cell epitope recognition demonstrated remarkable conformational sensitivity, enabling precise discrimination among closely related amino acids, such as serine versus threonine, glutamine versus asparagine etc.[27]. Here, we have used this high-density peptide microarray platform to examine and compare the linear B cell epitope repertoire against human coronavirus in pooled sera from eight COVID-19 convalescents sampled around March 2020 (i.e. early in the COVID-19 pandemic) and eight pre-pandemic individuals sampled before October 2019. We report early pandemic and pre-pandemic IgG and IgA responses against linear B cell epitopes from SARS-CoV-2, SARS-CoV, and MERS-CoV, as well as the four human common cold coronaviruses.

Methods

Early COVID-19 pandemic patient and healthy pre-pandemic control serum pools

Pools of anonymised serum from early pandemic, COVID-19-positive convalescent donors or pre-pandemically sampled negative control donors were obtained in March 2020 from the University Hospital of Copenhagen. The COVID-19-positive status of the COVID-19 convalescent donors was confirmed using a commercial IVD-certified RT-PCR test to detect the presence of SARS-CoV-2 RNA in oropharyngeal swaps. Serum from eight COVID-19 convalescents and eight pre-pandemic donors were selected randomly from a larger panel of potential donors, de-identified, and pooled, thus creating anonymous pools of COVID-19 positive and negative donors.

High-density peptide microarrays

As a non-overlapping source of negative control peptides (NP), we randomly generated 3900 random 13 amino acid peptides using the amino acid frequencies of the investigated proteomes. As an overlapping source of negative control peptides (EP), we extracted the proteome of Zaire Ebola virus (strain Mayinga-76, EBOZM), which has not circulated in Denmark. Thus, we did not expect any Danish human serum pool to contain antibodies against EBOZM-derived peptides. As a source of positive control peptides (PP), we extracted the proteome of human cytomegalovirus (HCMV, strain AD169). Due to the strong reactivity and high seroprevalence of HCMV in the human population[30], we expected (and later confirmed) that any human serum pool would contain antibodies against HCMV-derived peptides.

To design experimental libraries of human coronavirus-derived peptides (CP), we extracted the proteomes of the seven human coronaviruses (HCoVs) from the UniProt Knowledgebase[31]: the common cold coronaviruses (CCC), including the alpha coronaviruses 229E and NL63 and the beta coronaviruses OC43 and HKU1(N1 isolate), the epidemic HCoVs MERS-CoV and SARS-CoV, and the novel coronavirus, SARS-CoV-2. All the virus proteomes were represented by overlapping peptide libraries consisting of 13 amino acid long peptides (Table 1) tiled by 2 amino acids (i.e., overlapping by 11 amino acids)

Table 1. Number of peptides from each organism on the peptide microarrays.

Organism (Abbreviation)	Peptide Count*
Zaire ebolavirus (strain Mayinga-76) (EBOZM)	2417
Human cytomegalovirus (strain AD169) (HCMV)	31388
Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)	6330
Human SARS coronavirus (SARS-CoV)	6353
Middle East respiratory syndrome-related coronavirus (isolate United Kingdom/H123990006/2012) (MERS-CoV)	4933
Human coronavirus HKU1 (isolate N1) (HCoV-HKU1(N1))	4998
Human coronavirus OC43 (HCoV-OC43)	5164
Human coronavirus NL63 (HCoV-NL63)	4689
Human coronavirus 229E (HCoV-229E)	4605
Total*	66581

* Total is the union of all peptide sets derived from the viral proteomes.

and filtered for sequence redundancy. A high-density peptide microarray design was generated, distributing the CPs, EPs and PPs in triplicate and the NPs in duplicate randomly across 12 virtual sectors using proprietary software (PepArray, Schafer-N). Peptides were synthesised by Schafer-N (Copenhagen) on amino-functionalized glass microscope slides using a maskless photolithographic synthesis strategy as described elsewhere[27]. After de-protection and washing, the peptide microarrays were used to detect anti-peptide-specific antibodies.

Detecting the peptide-specific binding of IgG and IgA antibodies

Peptide microarrays were washed in PBS supplemented with 0.1% BSA, 0.1% Triton X-100 and incubated with serum pools from convalescent COVID-19 patients or pre-pandemic donors diluted 1:100 in PBS (0.1% BSA, 0.1% Triton X-100) for 2 hours at room temp, washed in PBS (0.1% BSA, 0.1% Triton X-100), and developed by detecting bound IgG or IgA with appropriate fluorochrome-labelled secondary antibodies: 1 µg/ml goat anti-human IgG-Cy3 (Sigma, Cat# C2571) or goat anti-human IgA-Cy5 (Jackson ImmunoResearch, Cat# 109-175-011), respectively. Following post-incubation washes, the microarrays were spin-dried and read in a microarray laser scanner (INNOSCAN 900, Innopsys, France). Finally, the proprietary PepArray software was used to quantify antibody-binding to each peptide at 8-bit resolution.

Identifying peptides binding antibodies with high-fidelity

Staining artefacts were automatically identified by the proprietary software, PepArray (Schafer-N, Copenhagen, Denmark), and the corresponding peptide signals were removed from the final dataset. The remaining data from the high-density peptide microarray were analysed using the R statistical language version 4.3.0[32]. For each serum pool and each Ig isotype, the arithmetic mean signal for each virus-derived peptide and NP was calculated, and the background was subtracted using the median signal of all 3900 random peptides (NPs) to account for non-specific binding (any background corrected value ≤ 0 were recoded to 0, and a value of 1 was added to all signals). The empirical cumulative distribution (Equation 1) of the NP background-corrected signals was used to assess the probability (p-value) that antibody-bindings signals recorded for virus-derived peptides could be fortuitous (i.e., non-specific).

$$\hat{F}(s) = \frac{1}{|NP|}\sum_{i=1}^{|NP|} 1_{[NP_i \geq s]} \qquad \textbf{Equation 1}$$

Where NP indicates the antibody-binding signals of peptides in the NP set, and s indicates an antibody-binding signal of any virus-derived peptide. The virus-derived peptide sequences were mapped to their parent proteins. Each virus-derived peptide sequence was pooled with its neighbouring sequences overlapping with at least 11 positions in the parent protein using the R package, IRanges version 2.36.0[33]. Antibody-binding fidelity for each virus-derived peptide was assessed by combining its p-value with the p-values of its pooled neighbours using Fisher's combined probability test (Equation 2).

$$\chi^2_{2k} \sim -2\sum_{i=1}^{k} \log(p_i) \qquad \textbf{Equation 2}$$

Where k indicates the number of pooled peptides and pi the p-value of peptide i in pool k derived from Equation 1. We reasoned that our overlapping peptide library strategy involving 13-mer peptides overlapping by 11 amino acids would ensure that linear B cell epitopes, largely being less than 10 amino acids long[27], would appear in at least two consecutive overlapping peptides. Moreover, to focus our search on B cell epitopes that yielded signals that were clearly above the background, we imposed a minimum signal requirement of 20% of the maximum signal obtained. Thus, we set a threshold for high-fidelity antibody binding for each virus peptide pair at a Fisher's combined p-value ≤ 0.001 and a standardised and cleaned signal value $\geq 20\%$ of the maximum standardised and background-corrected signal value measured for the serum pool and Ig-type in question. This effectively means that any solitary peptide with a high binding signal will be considered a false positive and assigned to the non-binder group which consists of all remaining virus-derived peptides outside the high-fidelity group. The number of peptides determined to have high-fidelity antibody-binding (HFPs), are summarised for each serum pool and Ig-type (IgG/IgA) in Table 2. Eventually, neighbouring high-fidelity virus-derived peptide pairs were merged into high-fidelity regions (HFR) using the R package IRanges version 2.36.0[33]. The number of HCMV, EBOZM, and HCoV HFRs determined for each serum pool and Ig-type (IgG/IgA) are summarised in Table 3.

Alignment of high-fidelity antibody-binding peptide regions

Unique high-fidelity regions (HFRs) derived from the viruses were uploaded in a single batch to the IEDB Epitope Cluster Analysis web tool (http://tools.iedb.org/cluster/)[34] and clustered against each of the seven HCoV proteomes using a clustering threshold of 60% sequence identity and clustering method 1 ("all connected peptides in a cluster"). The 60% level of homology was empirically selected as the lowest level of homology, which avoided clustering of EBOZM (EP) and HCMV (PP) epitopes with HCoV (CP) epitopes and/or HCoV proteomes (in reality, the 60% sequence identity threshold did inadvertently include a single PP epitope; lower thresholds would have increased these false positive clustering results dramatically). Thus, a clustering threshold of 60% sequence identity would maximise epitope clustering to the HCoV proteomes while avoiding the inclusion of unrelated virus-derived epitopes thereby achieving a reasonable balance between sensitivity and specificity.

The clustered high-fidelity regions were loaded into R[32]. Sequences within each cluster and its subclusters were re-aligned with profile-to-profile alignment using the R package

Table 2. Number of peptides with high-fidelity antibody binding.

Organism	Total	Pre-pandemic IgG		Pandemic IgG		Pre-pandemic IgA		Pandemic IgA	
EBOZM	2417	13 (0.54%)		21 (0.87%)		7 (0.29%)		0 (0.00%)	
HCMV	31388	346 (1.10%)	*	589 (1.88%)	**	38 (0.12%)		11 (0.04%)	
SARS-CoV-2	6330	34 (0.54%)		155 (2.45%)	***	10 (0.16%)		41 (0.65%)	**
SARS-CoV	6353	20 (0.31%)		117 (1.84%)	**	11 (0.17%)		27 (0.42%)	**
MERS-CoV	4933	11 (0.22%)		48 (0.97%)		7 (0.14%)		7 (0.14%)	
HCoV-HKU1(N1)	4998	42 (0.84%)		65 (1.30%)		7 (0.14%)		5 (0.10%)	
HCoV-OC43	5164	23 (0.45%)		47 (0.91%)		9 (0.17%)		5 (0.10%)	
HCoV-NL63	4689	29 (0.62%)		53 (1.13%)		2 (0.04%)		4 (0.09%)	
HCoV-229E	4605	25 (0.54%)		53 (1.15%)		5 (0.11%)		9 (0.20%)	

Levels of statistical significance when compared to EBOZM by 2-sample test for equality of proportions with multiple test correction by the Benjamini-Hochberg procedure: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

HCMV = human cytomegalovirus, EBOZM = Zaire ebolavirus (strain Mayinga-76), HCoV = Human coronavirus.

Table 3. High-fidelity peptides and regions detected in SARS-CoV-2.

		Pre-pandemic IgA			Pre-pandemic IgG			Pandemic IgA			Pandemic IgG		
Protein	Total	HFPs	*Fraction	HFRs	HFPs	*Fraction	HFRs	HFPs	*Fraction	HFRs	HFPs	*Fraction	HFRs
ORF1ab	4884	5	0,10%	4	18	0,37%	11	11	0,23%	5	37	0,76%	18
S	636	1	0,16%	1	11	1,73%	4	11	1,73%	6	36	5,66%	10
ORF3a	133	0	0,00%	0	1	0,75%	1	0	0,00%	0	10	7,52%	4
M	131	0	0,00%	0	0	0,00%	0	0	0,00%	0	13	9,92%	1
N	269	4	1,49%	1	1	0,37%	1	19	7,06%	3	59	21,93%	7
ORF9b	47	0	0,00%	0	3	6,38%	1	0	0,00%	0	0	0,00%	0

* Fraction indicates the fraction of HFPs out of the total number of peptides investigated for each protein.

HFPs = High-fidelity peptides, HFRs = High-fidelity regions

DECIPHER version 2.30.0[35]. Alignment clusters between coronaviridae and HCMV are summarised in Table 4 & Table 5.

Statistics

In Table 2, alongside the counts of high-fidelity peptides (HFPs), the epitope densities for the EBOZM, HCMV, and each of the HCoV-derived peptides are summarised, calculated as the fraction of HFPs determined in the total number of peptides synthesised for each organism. The EBOZM-derived peptides were included to represent a negative control pathogen, where seroreactivity would be unlikely in Danish human serum pools. Thus, the epitope densities observed in EBOZM represented the apparent false discovery rate of the HFPs determined for each serum pool and Ig-type. We applied the 2-sample test for equality of proportions to determine if the epitope densities of the HCMV-derived positive control peptides (PPs) and the HCoV-derived peptides (CPs) were significantly enriched compared to the serum pool and Ig-type FDRs (i.e. EBOZM epitope densities). P-values derived from the two-proportions z-tests were corrected for multiple testing by the Benjamini-Hochberg procedure, and the resulting significance indicators are summarised in Table 2.

To assess whether the pandemic serum pool pathogen-specific IgG epitope densities were significantly enriched compared to the pathogen-specific IgG epitopes observed for the pre-pandemic serum pool, we applied a one-way exact binomial test. In this test, the number of IgG HFPs in the pandemic serum pool was considered the number of successes, the total count of pathogen-specific peptides served as the number of trials, and the IgG epitope density from the pre-pandemic

Table 4. High-fidelity regions detected in human coronaviruses, human cytomegalovirus, and ebolavirus.

Organism	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgG	Pandemic IgA	Organism union*
HCMV	21	96	165	7	271
EBOZM	5	6	11	0	21
SARS-CoV-2	6	18	40	14	72
SARS-CoV	6	10	32	9	53
HCoV-HKU1(N1)	4	17	25	3	47
HCoV-NL63	1	15	24	2	37
HCoV-229E	3	9	21	4	34
HCoV-OC43	5	11	16	4	34
MERS-CoV	4	7	19	4	32
HCoVs union**	26	76	167	34	***280

* Organism total shows the unique high-fidelity region (HFR) sequences detected across all serum pools and Ig-types for the indicated organisms.

** HCoVs total shows the unique HFR sequences in each serum pool and Ig-type across all HCoVs.

*** The number of unique HFR sequences found across all HCoVs, serum pools, and Ig-types.

HCMV = human cytomegalovirus, EBOZM = Zaire ebolavirus (strain Mayinga-76), HCoV = Human coronavirus.

serum pool was used as the hypothetical probability of success. P-values for each organism-wise comparison were corrected for multiple testing by the Benjamini-Hochberg procedure.

Results

This study aimed to identify human coronavirus-specific linear B cell epitopes using serum pools collected before or during the early stages of the COVID-19 pandemic. Overlapping peptides representing the complete proteomes of the seven human coronaviruses (HCoVs) were generated by high-density, in situ synthesised peptide microarrays (Table 1). To analyse the binding of IgG and IgA antibodies from the serum pools to the peptides, we first subtracted background signals to account for non-specific binding. For each serum pool and Ig isotype, the background signal distribution was established using randomly generated peptides as negative control peptides (NP). To identify peptides that qualify as high-fidelity linear B cell epitopes, we exploited that our data came from the analysis of overlapping 13-mer peptide libraries with an overlap of 11-amino acids. Since linear B cell epitopes are often shorter than 10 amino acids long, they would tend to appear in at least two consecutive overlapping peptides, appearing as if "stepping" in and out of an immunogenic region with elevated antibody binding. In the Methods section, we described how we harnessed this information about overlapping peptides by integrating the background signal probabilities of a virus-derived peptide with its immediate neighbors, sharing an 11 amino acid overlap in the parent protein, to calculate a combined background probability within the framework of the parent protein sequence. We applied a statistical threshold of a combined p-value ≤ 0.001 to determine peptides likely harbouring linear B cell epitopes. Additionally, to ensure that only peptides eliciting stronger signals were included, we also applied a threshold for antibody binding signal of $\geq 20\%$ of the maximum signal. Peptides satisfying both criteria were categorised as being high-fidelity binding peptides (HFPs). This approach enabled the quantitative assessment of the numbers of pathogen-specific HFPs found before and early into COVID-19 pandemic (Table 2).

The results of this analysis could be visualised by scatter plots (Figure 1 (IgG) and Figure 2 (IgA). To allow for direct comparison of signal values, signals were scaled to the serum pool and Ig-type maximum values and are plotted as the scaled signals from the pandemic serum pools on the Y-axis against the normalised signals from the pre-pandemic serum pools on the X-axis. Whether a given peptide could be categorised as HFP by both the pre-pandemic and pandemic sera, by the pre-pandemic serum only, or by the pandemic serum only, is indicated by the colouring of the figure symbols (Figure 1 (IgG) and Figure 2 (IgA). To facilitate comparisons of immunogenicity across proteomes of different lengths, we have adjusted for protein length by dividing the number of HFP found in each protein with the total number of overlapping peptides

Table 5. Homology clustering alignments of *coronaviridae* high-fidelity regions.

Align ID	F4 Bar Nr.*	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
1	9	SARS-CoV	P0C6X7	ORF1ab				X	67	79	PYVFIKRSDALST
2	2	SARS-CoV-2	P0DTD1	ORF1ab	X				111	123	VGEIPVAYRKVLL
3	5	HCoV-229E	P0C6X1	ORF1ab				X	161	175	RKPLDYKRQNNLAIE
4	4	MERS-CoV	K9N7C7	ORF1ab				X	177	193	DPKGKYAQNLLKKLIGG
5	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab		X			179	197	DLIISNFKFDDYDFSVEDA
5	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	183	195	----SNFKFDDYDFSVE--
6	7	HCoV-OC43	P0C6X6	ORF1ab	X				187	205	PWVMYLRKRGEKGAYNKDH
7	6	HCoV-NL63	P0C6X5	ORF1ab		X			187	199	LKSGCKLTNAKPP
8	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	205	217	PKGKYSQKAYALL
9	9	SARS-CoV	P0C6X7	ORF1ab		X		X	251	269	HQTPFEIKSAKKFDTFKGE
10	3	SARS-CoV-2	P0DTD1	ORF1ab				X	281	293	--IKTIQPRVEKKKL--
10	3	SARS-CoV	P0C6X7	ORF1ab				X	279	295	SKVKVIQPRVEKKKTEG
11	7	HCoV-OC43	P0C6X6	ORF1ab		X		X	295	313	IVAWHVVRDPRYVMRLQSA
12	6	HCoV-NL63	P0C6X5	ORF1ab				X	359	377	AFLGASVTEDVKFAASTGV
13	9	SARS-CoV	P0C6X7	ORF1ab				X	489	501	DTIKSLDYKSFKT
14	2	SARS-CoV-2	P0DTD1	ORF1ab				X	503	517	VESCGNFKVTKGKAK
14	2	SARS-CoV-2	P0DTD1	ORF1ab		X			505	517	--SCGNFKVTKGKAK
15	5	HCoV-229E	P0C6X1	ORF1ab		X			637	655	DVIDNEIIVKPNISLCVPL
15	5	HCoV-229E	P0C6X1	ORF1ab				X	637	653	DVIDNEIIVKPNISLCV--
16	2	SARS-CoV-2	P0DTD1	ORF1ab			X		691	705	DSIIIGGAKLKALNL
17	7	HCoV-OC43	P0C6X6	ORF1ab		X			723	737	IDGLSCFKIGRRRIC
18	2	SARS-CoV-2	P0DTD1	ORF1ab		X			747	759	LPTEVLTEEVVLK
19	5	HCoV-229E	P0C6X1	ORF1ab				X	753	769	TCAKRFKRWLGILLEAY
20	7	HCoV-OC43	P0C6X6	ORF1ab				X	821	837	VYMAKAGDKYYPVVVDD
21	2	SARS-CoV-2	P0DTD1	ORF1ab				X	831	849	IEVQGYKSVNITFELDERI
22	5	HCoV-229E	P0C6X1	ORF1ab	X				929	941	VDVCEKAIGKKIK

Align ID	F4 Bar Nr.*	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
23	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab		X			945	1101	NDDEDVVTGDNDDEDVVTGDNDDEDVVTGDNDDEDVVTGDND DEDVVTGDNDDEDVVTGDNDDEDVVTGDNDDEDVVTGDNDDE DVVTGDNDDEDVVTGDNDDEDVVTGDNDDEDVVTGDNDDED VVTGDNDDEDVVTGDNNDEEIVTGDNDDQIVV
24	5	HCoV-229E	P0C6X1	ORF1ab	X				1073	1089	YAMQFFKMGRVAKMIER
25	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	1107	1119	DDIESIYDFDTYK
26	6	HCoV-NL63	P0C6X5	ORF1ab				X	1147	1161	LFWIMPYTKLFQKGE
27	4	MERS-CoV	K9N7C7	ORF1ab	X				1153	1171	GGGIAGAINAASKGAVQKE
27	4	MERS-CoV	K9N7C7	ORF1ab			X		1153	1167	GGGIAGAINAASKGA----
28	9	SARS-CoV	P0C6X7	ORF1ab				X	1169	1185	LDNLKPRVEAPKQEEPP
28	9	SARS-CoV	P0C6X7	ORF1ab		X			1169	1183	LDNLKPRVEAPKQEE--
29	4	MERS-CoV	K9N7C7	ORF1ab			X		1201	1219	HVVGPDARAKQDVSLLSKC
30	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab		X			1239	1251	LKCGFSFDLNGLD
31	4	MERS-CoV	K9N7C7	ORF1ab			X		1283	1295	SYDGLRGAIRKAK
32	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab		X			1323	1343	DGKFVTKFSGDKFDFIVGYGM
33	3	SARS-CoV-2	P0DTD1	ORF1ab		X		X	1423	1435	YFYTSKTTVASLI
33	3	SARS-CoV	P0C6X7	ORF1ab				X	1399	1411	FFYTSKEPVASII
34	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	1401	1413	IAVAAGKKFSKET
35	5	HCoV-229E	P0C6X1	ORF1ab		X			1443	1457	QPKIEPKPVSVIKVA
36	2	SARS-CoV-2	P0DTD1	ORF1ab				X	1525	1537	IEFLKRGDKSVYY
37	5	HCoV-229E	P0C6X1	ORF1ab				X	1627	1641	QQVGVIADKDKDLSG
38	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	1701	1713	ILYQYENLSLADI
39	3	SARS-CoV-2	P0DTD1	ORF1ab				X	1765	1777	VEAVMYMGTLSYE-
39	3	SARS-CoV	P0C6X7	ORF1ab				X	1743	1755	-EAVMYMGTLSYDN
40	4	MERS-CoV	K9N7C7	ORF1ab	X				1879	1891	DSSKPVTKKYTYS----
40	4	MERS-CoV	K9N7C7	ORF1ab				X	1881	1895	--SKPVTKKYTYSFLPK
41	4	MERS-CoV	K9N7C7	ORF1ab		X			1933	1945	TNLNKFNRASLRQ
42	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	1945	1961	CLYLKNLTQTFTSMLTN
43	9	SARS-CoV	P0C6X7	ORF1ab	X				1951	1963	HYSASFKKGAKLL

Open Research Europe 2024, 4:215 Last updated: 19 DEC 2024

Align ID	F4 Bar Nr.*	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
44	5	HCoV-229E	P0C6X1	ORF1ab				X	1967	1979	GVVLKRSLKYNLK
45	2	SARS-CoV-2	P0DTD1	ORF1ab				X	1993	2005	HVNNATNKATYKP
46	7	HCoV-OC43	P0C6X6	ORF1ab	X				2037	2051	NIIKLSGVKKPFKVE
47	9	SARS-CoV	P0C6X7	ORF1ab				X	2105	2121	AAINSVPWSKILAYVKP
48	9	SARS-CoV	P0C6X7	ORF1ab	X				2191	2203	GINYVKSPKFSKL
49	6	HCoV-NL63	P0C6X5	ORF1ab		X			2215	2227	YVIIDKVDFVNGF
50	5	HCoV-229E	P0C6X1	ORF1ab				X	2351	2367	NMSLAECKRALGLSISD
51	8	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	2487	2499	--FCYKRNCSVRVKC
51	8	HCoV-OC43	P0C6X6	ORF1ab				X	2397	2411	CLFCYKRNRSLRVKC
52	7	HCoV-OC43	P0C6X6	ORF1ab	X				2585	2597	LSMFDVDKKSLNA
53	6	HCoV-NL63	P0C6X5	ORF1ab				X	2635	2649	NAYYKYDAKNYVRFP
54	2	SARS-CoV-2	P0DTD1	ORF1ab	X				2669	2681	SCNNYMLTYNKVE
55	4	MERS-CoV	K9N7C7	ORF1ab				X	2695	2709	LSDALKRQIRIACRK
56	8	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	2809	2823	VKTGLKLKLTFNKQE
56	8	HCoV-HKU1(N1)	P0C6X2	ORF1ab	X				2809	2821	VKTGLKLKLTFNK--
56	8	HCoV-OC43	P0C6X6	ORF1ab	X				2721	2733	CKTGLKLKLTYNK--
57	2	SARS-CoV-2	P0DTD1	ORF1ab	X				2729	2743	IRSAAKKNNLPFKLT
57	2	SARS-CoV-2	P0DTD1	ORF1ab		X		X	2729	2741	IRSAAKKNNLPFK--
58	6	HCoV-NL63	P0C6X5	ORF1ab	X				2879	2893	EKLKNYAASYNKYKY
58	6	HCoV-NL63	P0C6X5	ORF1ab			X		2879	2891	EKLKNYAASYNKY--
59	4	MERS-CoV	K9N7C7	ORF1ab		X		X	3045	3057	FYYINKVKRAFAD
60	6	HCoV-NL63	P0C6X5	ORF1ab		X			3121	3135	YGNFDDQPSLQVESA
60	6	HCoV-NL63	P0C6X5	ORF1ab				X	3121	3133	YGNFDDQPSLQVE--
61	11	SARS-CoV-2	P0DTD1	ORF1ab			X		3202	3214	-TQYNRYLALYNKY-
61	11	SARS-CoV-2	P0DTD1	ORF1ab	X				3203	3215	--QYNRYLALYNKYK
61	11	SARS-CoV	P0C6X7	ORF1ab			X		3179	3191	-TQYNRYLALYNKY-
61	11	SARS-CoV	P0C6X7	ORF1ab	X				3180	3192	--QYNRYLALYNKYK

Open Research Europe 2024, 4:215 Last updated: 19 DEC 2024

Align ID	F4 Bar Nr.*	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
61	11	HCoV-OC43	P0C6X6	ORF1ab			X		3183	3195	DVAFNRYLSLYNK--
62	3	SARS-CoV-2	P0DTD1	ORF1ab				X	3307	3322	-CTSEDMLNPNYEDLLI
62	3	SARS-CoV	P0C6X7	ORF1ab				X	3283	3299	ICTAEDMLNPNYEDLLI
63	4	MERS-CoV	K9N7C7	ORF1ab				X	3289	3307	VMCPADQLSDPNYDALLIS
64	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	3335	3351	SGIVKMVSPTSKIEPCI
65	7	HCoV-OC43	P0C6X6	ORF1ab				X	3341	3357	NSRTPKYTFGVVKPGET
66	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab	X				3627	3639	VYQQLAGVKLQSK
67	6	HCoV-NL63	P0C6X5	ORF1ab				X	3637	3653	QLIKQLKRAMNIAKSEF
68	5	HCoV-229E	P0C6X1	ORF1ab		X			3637	3651	GMPSFVAYETARQEY
69	4	MERS-CoV	K9N7C7	ORF1ab				X	3651	3663	VANWLAPTNAYMR
70	5	HCoV-229E	P0C6X1	ORF1ab	X				3657	3669	NGSSPQIIKQLKK
71	3	SARS-CoV-2	P0DTD1	ORF1ab				X	3827	3843	GLLPPKNSIDAFKLNIK-
71	3	SARS-CoV	P0C6X7	ORF1ab				X	3805	3821	-LLPPKSSIDAFKLNIKL
72	7	HCoV-OC43	P0C6X6	ORF1ab		X			3833	3845	SQFQSKLTDVKCA
73	6	HCoV-NL63	P0C6X5	ORF1ab		X		X	3839	3851	YAYISNKADLKFV
74	4	MERS-CoV	K9N7C7	ORF1ab				X	3887	3899	ATDPSEAFEKFVS
75	16	SARS-CoV-2	P0DTD1	ORF1ab		X		X	4227	4239	YLYFIKGLNNLNR
75	16	SARS-CoV	P0C6X7	ORF1ab		X		X	4204	4216	YLYFIKGLNNLNR
75	16	HCoV-229E	P0C6X1	ORF1ab		X		X	3907	3919	YLYFVKNLNNLRR
76	3	SARS-CoV-2	P0DTD1	ORF1ab		X			3933	3950	EMLDNRATLQAIASEFSS
76	3	SARS-CoV	P0C6X7	ORF1ab		X			3910	3927	EMLDNRATLQAIASEFSS
77	3	SARS-CoV-2	P0DTD1	ORF1ab		X			3972	3984	DSEVVLKKLKKSL---
77	3	SARS-CoV-2	P0DTD1	ORF1ab				X	3975	3987	---VVLKKLKKSLNVA
77	3	SARS-CoV	P0C6X7	ORF1ab		X			3949	3961	DSEVVLKKLKKSL---
77	3	SARS-CoV	P0C6X7	ORF1ab				X	3952	3964	---VVLKKLKKSLNVA
78	4	MERS-CoV	K9N7C7	ORF1ab				X	3959	3971	TSPQVLKALQKAV
79	7	HCoV-OC43	P0C6X6	ORF1ab		X			4171	4183	LKYTKILKDDGNF

Open Research Europe 2024, 4:215 Last updated: 19 DEC 2024

Align ID	F4 Bar Nr.*	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
80	6	HCoV-NL63	P0C6X5	ORF1ab		X			4173	4185	EQNCDVLKEVLVL
81	6	HCoV-NL63	P0C6X5	ORF1ab		X			4213	4225	ASLGKIVARAMLK
82	8	HCoV-HKU1(N1)	P0C6X2	ORF1ab		X			4339	4355	FSVDPKKTYLDYIQQGG
82	8	HCoV-OC43	P0C6X6	ORF1ab		X			4255	4267	----PKKTYLDFIQQGG
83	10	HCoV-NL63	P0C6X5	ORF1ab		X			4291	4307	ECFVKSDIFGSDFKTFD
83	10	HCoV-NL63	P0C6X5	ORF1ab				X	4293	4307	--FVKSDIFGSDFKTFD
83	10	HCoV-229E	P0C6X1	ORF1ab		X		X	4317	4331	-CFMKSDIFGQDFKTF-
84	14	MERS-CoV	K9N7C7	ORF1ab				X	4791	4803	--GNFNQDFYDFVVS--
84	14	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	4865	4879	-PGNFNQDFYEFVKSK-
84	14	HCoV-OC43	P0C6X6	ORF1ab				X	4777	4789	-PGNFNQDFYDFVL---
84	14	HCoV-NL63	P0C6X5	ORF1ab		X		X	4449	4465	KPGHFNEEFYNFLRLRG
84	14	HCoV-229E	P0C6X1	ORF1ab		X		X	4475	4489	-PGHFNKEFYDFLRSQ-
85	10	HCoV-NL63	P0C6X5	ORF1ab				X	4487	4499	AVKDFDFYRYNKP---
85	10	HCoV-229E	P0C6X1	ORF1ab				X	4513	4527	-IKDFDYYRYNRPTML
86	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	4519	4541	DGNKLDKFFVVKRTNLEVYNKEK
86	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab		X	X		4527	4541	--------FVVKRTNLEVYNKEK
87	12	SARS-CoV-2	P0DTD1	ORF1ab				X	4931	4946	ITQMNLKYAISAKNRA
87	12	SARS-CoV-2	P0DTD1	ORF1ab		X			4933	4945	--QMNLKYAISAKNR-
87	12	SARS-CoV	P0C6X7	ORF1ab				X	4908	4923	ITQMNLKYAISAKNRA
87	12	SARS-CoV	P0C6X7	ORF1ab		X			4910	4922	--QMNLKYAISAKNR-
87	12	MERS-CoV	K9N7C7	ORF1ab				X	4917	4932	MTQMNLKYAISAKNRA
87	12	MERS-CoV	K9N7C7	ORF1ab		X			4919	4931	--QMNLKYAISAKNR-
87	12	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	4993	5007	-TQMNLKYAISAKNRA
87	12	HCoV-HKU1(N1)	P0C6X2	ORF1ab		X			4994	5006	--QMNLKYAISAKNR-
87	12	HCoV-OC43	P0C6X6	ORF1ab				X	4905	4919	-TQMNLKYAISAKNRA
87	12	HCoV-OC43	P0C6X6	ORF1ab		X			4906	4918	--QMNLKYAISAKNR-
87	12	HCoV-NL63	P0C6X5	ORF1ab				X	4580	4592	---LNLKYAISGKERA

Open Research Europe 2024, 4:215 Last updated: 19 DEC 2024

Align ID	F4 Bar Nr.*	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
87	12	HCoV-229E	P0C6X1	ORF1ab				X	4605	4617	---LNLKYAISGKERA
88	8	HCoV-HKU1(N1)	P0C6X2	ORF1ab		X			4723	4735	LVQYDFTDYKLEL
88	8	HCoV-OC43	P0C6X6	ORF1ab		X			4635	4647	LVQYDFTDYKLEL
89	10	HCoV-NL63	P0C6X5	ORF1ab				X	4815	4831	ADISAFKATLYYQNNVF
89	10	HCoV-229E	P0C6X1	ORF1ab				X	4840	4855	ADISAFKATLYYQNGV-
90	17	SARS-CoV-2	P0DTD1	ORF1ab			X		4888	4900	NNLDKSAGFPFNK----
90	17	SARS-CoV	P0C6X7	ORF1ab			X		4865	4877	NNLDKSAGFPFNK----
90	17	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	4953	4965	----KSAGYPFNKFGKA
90	17	HCoV-OC43	P0C6X6	ORF1ab				X	4865	4877	----KSAGYPFNKFGKA
91	3	SARS-CoV-2	P0DTD1	ORF1ab		X			4967	4984	LLKSIAATRGATVVIGTS
91	3	SARS-CoV-2	P0DTD1	ORF1ab				X	4968	4981	-LKSIAATRGATVVI---
91	3	SARS-CoV	P0C6X7	ORF1ab		X			4944	4961	LLKSIAATRGATVVIGTS
91	3	SARS-CoV	P0C6X7	ORF1ab				X	4945	4958	-LKSIAATRGATVVI---
92	6	HCoV-NL63	P0C6X5	ORF1ab		X			5029	5041	DVKKLYLGGLNYY
93	10	HCoV-NL63	P0C6X5	ORF1ab				X	5085	5099	TDVRDYKLANDVKDT-
93	10	HCoV-229E	P0C6X1	ORF1ab				X	5111	5125	-DIRDYKLANDAKESL
94	5	HCoV-229E	P0C6X1	ORF1ab				X	5091	5109	SALGSMDIDVFNKLSTSDW
95	4	MERS-CoV	K9N7C7	ORF1ab		X			5113	5127	TSPDPKFVDKYYAFL
96	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	5171	5187	LSIRNLQKRLYSNVYRT
96	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab		X			5175	5187	----NLQKRLYSNVYRT
97	3	SARS-CoV-2	P0DTD1	ORF1ab				X	5258	5270	AYPLTKHPNQEYA
97	3	SARS-CoV	P0C6X7	ORF1ab				X	5235	5247	AYPLTKHPNQEYA
98	6	HCoV-NL63	P0C6X5	ORF1ab				X	5239	5251	YYQLIGKQKITTI
99	8	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	5553	5567	--ETGKVKPPLNKNYVF
99	8	HCoV-OC43	P0C6X6	ORF1ab				X	5463	5479	SWEIGKVKPPLNKNYVF
100	3	SARS-CoV-2	P0DTD1	ORF1ab			X		5504	5520	YVFTGYRVTKNSKVQIG
100	3	SARS-CoV	P0C6X7	ORF1ab			X		5481	5497	YVFTGYRVTKNSKVQIG
101	19	SARS-CoV-2	P0DTD1	ORF1ab		X			5851	5863	GLPTQTVDSSQGS----

Align ID	F4 Bar Nr.*	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
101	19	SARS-CoV	P0C6X7	ORF1ab		X			5828	5840	GLPTQTVDSSQGS----
101	19	MERS-CoV	K9N7C7	ORF1ab		X			5841	5853	----QTVDSSQGSEYQY
101	19	HCoV-NL63	P0C6X5	ORF1ab		X			5501	5513	---IQTVDSSQGSEYD-
102	3	SARS-CoV-2	P0DTD1	ORF1ab		X			5663	5675	RVECFDKFKVNST
102	3	SARS-CoV	P0C6X7	ORF1ab		X			5640	5652	RVECFDKFKVNST
103	8	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	5901	5917	SQNYVAKRVLGVQTQTV
103	8	HCoV-OC43	P0C6X6	ORF1ab				X	5813	5829	SQNFAAKRVLGLQTQTV
104	6	HCoV-NL63	P0C6X5	ORF1ab		X			5981	5995	VNKHAFHTPAYDKRA
105	15	MERS-CoV	K9N7C7	ORF1ab				X	6451	6465	-EGELPVAVVNDKIFT
105	15	HCoV-NL63	P0C6X5	ORF1ab				X	6103	6117	VDGELPVAVVNDKVF-
106	10	HCoV-NL63	P0C6X5	ORF1ab				X	6139	6155	---AFELFAKRKMGLTPPLS
106	10	HCoV-229E	P0C6X1	ORF1ab				X	6161	6179	TNIAFELFAKRKVGLTPPL-
106	10	HCoV-229E	P0C6X1	ORF1ab			X		6161	6173	TNIAFELFAKRKV-------
107	6	HCoV-NL63	P0C6X5	ORF1ab				X	6259	6271	YIYVRKNGQFQDH
108	3	SARS-CoV-2	P0DTD1	ORF1ab				X	6511	6527	-AKRNIKPVPEVKILNNL-
108	3	SARS-CoV	P0C6X7	ORF1ab				X	6487	6505	WAKRNIKPVPEIKILNNLG
109	3	SARS-CoV-2	P0DTD1	ORF1ab			X	X	6621	6637	GEAVKTQFNYYKKVDGV---
109	3	SARS-CoV	P0C6X7	ORF1ab				X	6599	6617	-ESVKTQFNYFKKVDGIIQQ
109	3	SARS-CoV	P0C6X7	ORF1ab		X			6605	6617	-------FNYFKKVDGIIQQ
110	4	MERS-CoV	K9N7C7	ORF1ab				X	6605	6617	FYLYKKVNNEFID
111	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	6635	6647	EAFRKARNGVFIS
112	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	6665	6681	LNGVIVDKVGELKVEFW
113	1	HCoV-HKU1(N1)	P0C6X2	ORF1ab				X	6785	6803	LHLLIGLFRRKKKSNLLIQ
114	3	SARS-CoV-2	P0DTD1	ORF1ab				X	6933	6951	KTKNVTKENDSKEGFFTYI
114	3	SARS-CoV	P0C6X7	ORF1ab				X	6911	6927	-TKHVTKENDSKEGFFTY-
115	1	HCoV-HKU1(N1)	Q5MQD1	HE		X			186	203	PLCLFKSNFSQYYYNIDT
116	5	HCoV-229E	P15423	S		X			53	73	AFNNWFLLTNTSSVVDGVVRS

Align ID	F4 Bar Nr.*	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
117	1	HCoV-HKU1(N1)	Q5MQD0	S				X	79	93	LKGTTYLSTLWYQKP
118	8	HCoV-HKU1(N1)	Q5MQD0	S		X			97	111	-DFNNGIFSRVKNTKL-----
118	8	HCoV-OC43	P36334	S		X			97	115	SDFINGIFAKVKNTKVIKD--
118	8	HCoV-OC43	P36334	S				X	99	117	--FINGIFAKVKNTKVIKDRV
119	2	SARS-CoV-2	P0DTC2	S			X		139	151	PFLGVYYHKNNKS
120	5	HCoV-229E	P15423	S			X		179	191	TSAFVGALPKTVR
121	8	HCoV-HKU1(N1)	Q5MQD0	S		X			193	205	VSTDWLYFHFYQE--
121	8	HCoV-OC43	P36334	S		X			207	221	VNADYLYFHFYQEGG
122	1	HCoV-HKU1(N1)	Q5MQD0	S				X	257	273	WVTPLSKRQYLLKFDNR
123	3	SARS-CoV-2	P0DTC2	S				X	307	321	TVEKGIYQTSNFRVQ-
123	3	SARS-CoV	P59594	S				X	295	309	-IDKGIYQTSNFRVVP
124	2	SARS-CoV-2	P0DTC2	S			X		347	359	FASVYAWNRKRIS
125	5	HCoV-229E	P15423	S				X	441	459	VTLDKCTKYNIYDVSGVGV
126	2	SARS-CoV-2	P0DTC2	S		X	X		451	467	YLYRLFRKSNLKPFERD
126	2	SARS-CoV-2	P0DTC2	S				X	453	465	--YRLFRKSNLKPFE--
127	6	HCoV-NL63	Q6Q1S2	S		X			453	469	GLQDGFYSANFLDDNVL
127	6	HCoV-NL63	Q6Q1S2	S				X	453	467	GLQDGFYSANFLDDN--
128	4	MERS-CoV	K9N5Q8	S	X	X			497	509	YSYINKCSRFLSD
129	5	HCoV-229E	P15423	S		X			523	541	GFSNVVELPKFFYASNGTY
129	5	HCoV-229E	P15423	S				X	525	541	--SNVVELPKFFYASNGTY
130	2	SARS-CoV-2	P0DTC2	S				X	533	547	LVKNKCVNFNFNGLT
131	2	SARS-CoV-2	P0DTC2	S				X	553	571	TESNKKFLPFQQFGRDIAD
132	2	SARS-CoV-2	P0DTC2	S				X	573	589	TDAVRDPQTLEILDITP
133	2	SARS-CoV-2	P0DTC2	S		X			593	613	GGVSVITPGTNTSNQVAVLYQ
133	2	SARS-CoV-2	P0DTC2	S				X	595	613	--VSVITPGTNTSNQVAVLYQ
134	6	HCoV-NL63	Q6Q1S2	S				X	623	641	VLNNCTKYNIYDYVGTGII
135	12	SARS-CoV-2	P0DTC2	S				X	811	831	KPSKRSFIEDLLFNKVTLADA------

Open Research Europe 2024, 4:215 Last updated: 19 DEC 2024

Align ID	F4 Bar Nr.*	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
135	12	SARS-CoV-2	P0DTC2	S			X		819	831	--------EDLLFNKVTLADA------
135	12	SARS-CoV-2	P0DTC2	S		X			823	837	------------FNKVTLADAGFIKQY
135	12	SARS-CoV	P59594	S				X	793	813	KPTKRSFIEDLLFNKVTLADA------
135	12	SARS-CoV	P59594	S			X		801	813	--------EDLLFNKVTLADA------
135	12	MERS-CoV	K9N5Q8	S				X	883	903	SRSARSAIEDLLFDKVTIADP------
135	12	MERS-CoV	K9N5Q8	S			X		889	901	------AIEDLLFDKVTIA--------
135	12	HCoV-HKU1(N1)	Q5MQD0	S				X	901	920	-SSSRSFFEDLLFDKVKLSDV------
135	12	HCoV-HKU1(N1)	Q5MQD0	S			X		907	920	-------FEDLLFDKVKLSDV------
135	12	HCoV-OC43	P36334	S				X	899	919	KASSRSAIEDLLFDKVKLSDV------
135	12	HCoV-OC43	P36334	S			X		905	919	------AIEDLLFDKVKLSDV------
135	12	HCoV-NL63	Q6Q1S2	S				X	867	885	-IAGRSALEDLLFSKVVTSG-------
135	12	HCoV-NL63	Q6Q1S2	S			X		869	885	---GRSALEDLLFSKVVTSG-------
135	12	HCoV-229E	P15423	S				X	685	705	RVAGRSAIEDILFSKLVTSGL------
135	12	HCoV-229E	P15423	S			X		691	703	------AIEDILFSKLVTS--------
136	2	SARS-CoV-2	P0DTC2	S		X			703	715	NSVAYSNNSIAIP
137	13	HCoV-HKU1(N1)	Q5MQD0	S				X	751	779	SSSSSRRKRRSISASYRFVTFEPFNVSFV
137	13	HCoV-HKU1(N1)	Q5MQD0	S		X			761	779	----------SISASYRFVTFEPFNVSFV
137	13	HCMV	P17151	UL112/ UL113				X	617	629	SSSSSRRSRRRNN----------------
138	3	SARS-CoV-2	P0DTC2	S				X	775	809	DKNTQEVFAQVKQIYKTPPIKDFGGFNFSQILPDP
138	3	SARS-CoV	P59594	S				X	757	773	DRNTREVFAQVKQMYKT-----------------
138	3	SARS-CoV	P59594	S				X	779	791	----------------------FGGFNFSQILPDP
139	7	HCoV-OC43	P36334	S		X		X	759	777	AITTGYRFTNFEPFTVNSV
140	2	SARS-CoV-2	P0DTC2	S			X		765	783	RALTGIAVEQDKNTQEVFA
141	9	SARS-CoV	P59594	S				X	903	915	KQIANQFNKAISQ
142	3	SARS-CoV-2	P0DTC2	S	X				1029	1041	MSECVLGQSKRVD
142	3	SARS-CoV	P59594	S	X				1011	1023	MSECVLGQSKRVD

Open Research Europe 2024, 4:215 Last updated: 19 DEC 2024

Align ID	F4 Bar Nr.*	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
143	1	HCoV-HKU1(N1)	Q5MQD0	S			X		1043	1055	ALNSLLQQLFNKF
144	5	HCoV-229E	P15423	S		X			1049	1065	NYTVPDLVVEQYNQTIL
145	11	SARS-CoV-2	P0DTC2	S				X	1143	1161	PELDSFKEELDKYFKNHTS
145	11	SARS-CoV-2	P0DTC2	S			X		1143	1155	PELDSFKEELDKY------
145	11	SARS-CoV	P59594	S				X	1125	1143	PELDSFKEELDKYFKNHTS
145	11	SARS-CoV	P59594	S			X		1125	1137	PELDSFKEELDKY------
145	11	HCoV-OC43	P36334	S				X	1229	1245	-NLPDFKEELDQWFKNQT-
146	4	MERS-CoV	K9N5Q8	S	X				1167	1179	PVNGYFIKTNNTR
147	4	MERS-CoV	K9N5Q8	S				X	1227	1243	TGIDFQDELDEFFKNVS
148	3	SARS-CoV-2	P0DTC2	S				X	1251	1269	GSCCKFDEDDSEPVLKGVK
148	3	SARS-CoV	P59594	S				X	1233	1251	GSCCKFDEDDSEPVLKGVK
149	6	HCoV-NL63	Q6Q1S2	S		X		X	1341	1356	GSTKLPYYEFEKVHVQ
150	2	SARS-CoV-2	P0DTC3	ORF3a				X	11	25	GTVTLKQGEIKDATP
151	2	SARS-CoV-2	P0DTC3	ORF3a		X		X	63	75	ITLKKRWQLALSK
152	2	SARS-CoV-2	P0DTC3	ORF3a				X	181	195	EHDYQIGGYTEKWES
153	3	SARS-CoV-2	P0DTC3	ORF3a				X	253	273	SGVVNPVMEPIYDEPTTTTSV
153	3	SARS-CoV	P59632	ORF3a				X	253	267	-GVANPAMDPIYDEPT-----
154	5	HCoV-229E	P19739	ORF4a				X	23	37	VSAEVSRQVIQDVKD
155	4	MERS-CoV	K9N643	ORF4b				X	29	45	TKKLRYVKRRFSLLRPE
156	4	MERS-CoV	K9N643	ORF4b				X	55	69	HYVRVTFSDPNMWYL
157	1	HCoV-HKU1(N1)	Q5MQC7	M		X		X	1	15	MNKSFLPQFTSDQAV
158	3	SARS-CoV-2	P0DTC5	M				X	157	205	GRCDIKDLPKEITVATSRTLSYYKLGASQRVAGDSGFAAYSRYRIGNYK
158	3	SARS-CoV	P59596	M				X	156	201	GRCDIKDLPKEITVATSRTLSYYKLGASQRVGTDSGFAAYNRYRIG---
159	4	MERS-CoV	K9N4V7	N				X	13	37	ADNNDITNTNLSRGRGRNPKPRAAP
159	4	MERS-CoV	K9N4V7	N		X			13	31	ADNNDITNTNLSRGRGRNP------

Align ID	F4 Bar Nr.*	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
160	1	HCoV-HKU1(N1)	Q5MQC5	N		X			13	25	ESSRKLLGLTNLS
161	3	SARS-CoV-2	P0DTC9	N				X	19	50	GPSDSTGSNQNGERSGARSKQRRPQGLPNNTA
161	3	SARS-CoV-2	P0DTC9	N			X		27	45	--------NQNGERSGARSKQRRPQGL-----
161	3	SARS-CoV	P59595	N	X		X		27	47	-------NNQNGGRNGARPKQRRPQGLP----
161	3	SARS-CoV	P59595	N				X	35	51	---------------GARPKQRRPQGLPNNTA
162	5	HCoV-229E	P15130	N				X	41	57	RNLVPINKKDKNKLIGY
163	7	HCoV-OC43	Q4VID0	N				X	47	63	LELLSFKRERSLSLQKD
164	2	SARS-CoV-2	P0DTC9	N		X			67	79	PRGQGVPINTNSS
165	8	HCoV-HKU1(N1)	Q5MQC6	N				X	105	119	RRSFKTADGQQKQLL-
165	8	HCoV-OC43	P33469	N				X	107	121	-GSFKTADGNQRQLLP
166	18	SARS-CoV-2	P0DTC9	N				X	179	199	--GSQASSRSSSRSRNSSRNSTP
166	18	SARS-CoV	P59595	N				X	180	197	--GSQASSRSSSRSRGNSRN---
166	18	HCoV-NL63	Q6Q1R8	N				X	145	157	NNSSRASSRSSTR----------
167	7	HCoV-OC43	Q4VID0	N	X				149	161	MRLFRLGFRLARY
168	1	HCoV-HKU1(N1)	Q5MQC5	N	X				151	165	PMFRQGILLLKKLSL
169	3	SARS-CoV-2	P0DTC9	N				X	157	175	IVLQLPQGTTLPKGFYAEG
169	3	SARS-CoV	P59595	N				X	159	176	-VLQLPQGTTLPKGFYAEG
170	6	HCoV-NL63	Q6Q1R8	N				X	161	173	RDSSRSTSRQQSR
171	5	HCoV-229E	P15130	N			X		191	215	KSLGFDKPQEKDKKSAKTGTPKPSR
172	2	SARS-CoV-2	P0DTC9	N				X	201	215	SSRGTSPARMAGNGG
173	6	HCoV-NL63	Q6Q1R8	N		X		X	209	225	GTSTPKKPNKPLSQPRA
174	3	SARS-CoV-2	P0DTC9	N				X	221	261	LLLLDRLNQLESKMSGKGQQQQGQTVTKKSAAEASKKPRQK--
174	3	SARS-CoV-2	P0DTC9	N			X		245	263	----------------------TVTKKSAAEASKKPRQKRT
174	3	SARS-CoV	P59595	N				X	222	262	LLLLDRLNQLESKVSGKGQQQQGQTVTKKSAAEASKKPRQK--

Open Research Europe 2024, 4:215 Last updated: 19 DEC 2024

Align ID	F4 Bar Nr.*	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
174	3	SARS-CoV	P59595	N			X		246	264	----------TVTKKSAAEASKKPRQKRT
175	4	MERS-CoV	K9N4V7	N				X	235	251	PKVITKKDAAAAKNKMR
176	7	HCoV-OC43	P33469	N				X	257	277	HTAKEVRQKILNKPRQKRSPN
176	7	HCoV-OC43	P33469	N			X		261	273	----EVRQKILNKPRQK----
177	2	SARS-CoV-2	P0DTC9	N				X	335	351	GAIKLDDKDPNFKDQVI
178	7	HCoV-OC43	P33469	N			X		335	347	QNLSGNPDEPQKD
179	6	HCoV-NL63	Q6Q1R8	N		X		X	351	367	LNASIPESKPLADDDSA
180	3	SARS-CoV-2	P0DTC9	N				X	357	411	IDAYKTFPPTEPKKDKKKKADETQALPQRQKKQQTVTLLPAADLDD FSKQLQQSM
180	3	SARS-CoV-2	P0DTC9	N			X		358	391	-DAYKTFPPTEPKKDKKKKADETQALPQRQKKQQT---------- --------
180	3	SARS-CoV-2	P0DTC9	N	X				358	374	-DAYKTFPPTEPKKDKKK---------- --------
180	3	SARS-CoV	P59595	N				X	358	391	IDAYKTFPPTEPKKDKKKKTDEAQPLPQRQKKQP---------- --------
180	3	SARS-CoV	P59595	N	X		X		359	375	-DAYKTFPPTEPKKDKKK---------- --------
180	3	SARS-CoV	P59595	N			X		379	393	----------EAQPLPQRQKKQPTV---------- --------
181	4	MERS-CoV	K9N4V7	N				X	359	373	KTFPKKEKKQKAPKE
182	5	HCoV-229E	P15130	N				X	373	385	PVRDEVSIETDII
183	1	HCoV-HKU1(N1)	Q5MQC6	N		X			383	397	DSDSLSSKPQRKRGV--------
183	1	HCoV-HKU1(N1)	Q5MQC6	N	X				389	405	------SKPQRKRGVKQLPEQFD
184	7	HCoV-OC43	P33469	N		X			383	397	QDGMMNMSPKPQRQR
185	1	HCoV-HKU1(N1)	Q5MQC6	N		X			399	411	QLPEQFDSLNLSA
186	2	SARS-CoV-2	P0DTD2	ORF9b		X			53	69	SLNMARKTLNSLEDKAF

* Refers to the bar number the alignment cluster is associatied with in Figure 4.

Open Research Europe 2024, 4:215 Last updated: 19 DEC 2024

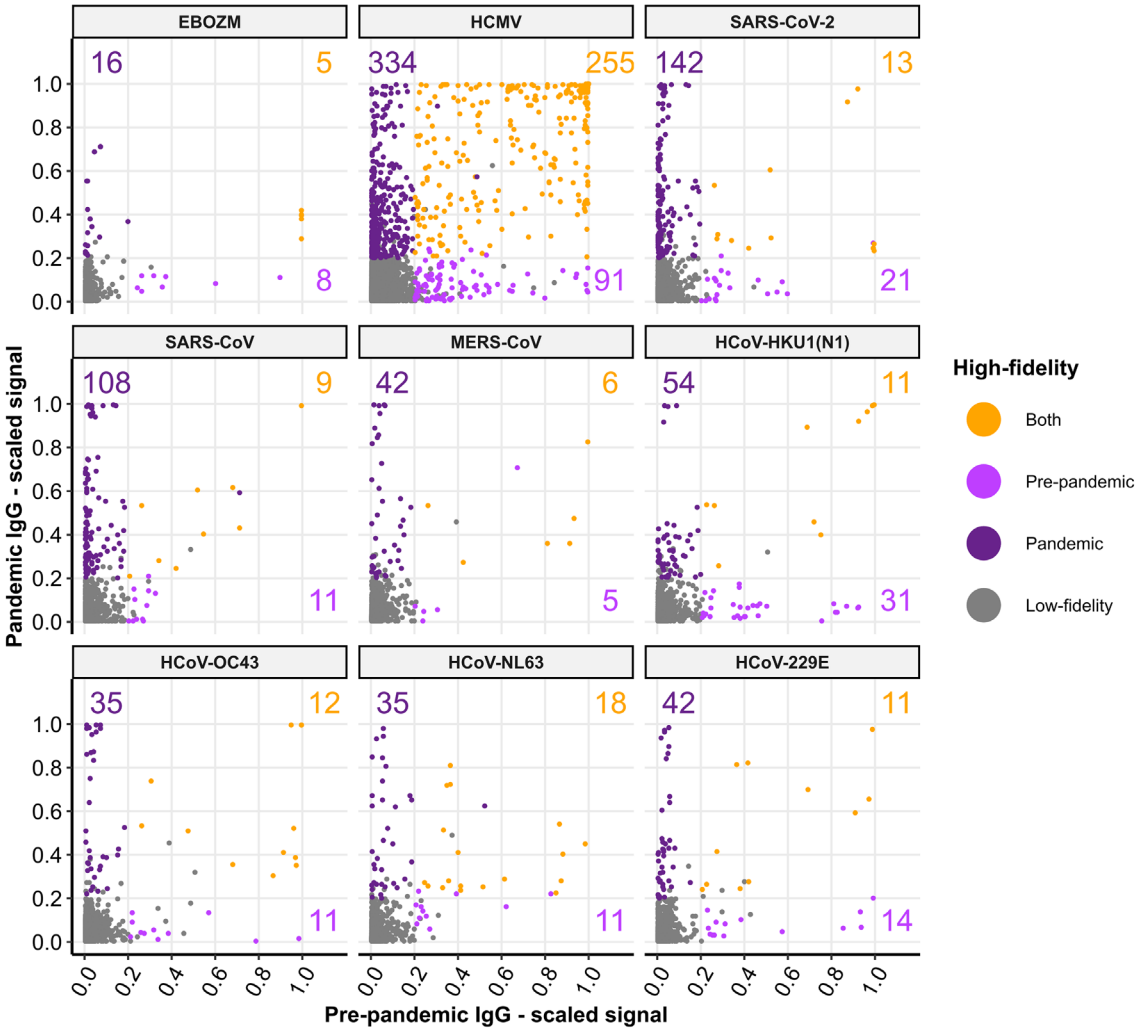

Figure 1. Peptide IgG binding signals on the microarrays. Scatter plot comparing peptide microarray IgG binding signals between the pandemic (y-axis) and pre-pandemic (x-axis) serum pools scaled to the maximum signal measured for each serum pool. Each data point represents one peptide derived from one of the organisms indicated in the plot subtitle. High-fidelity peptides passed our fidelity filters: combined p-value ≤ 0.001 & scaled signal ≥ 20% of the maximum scaled signal measured for the individual serum pool and Ig-type. The legend indicates if a peptide was determined to have high-fidelity IgG binding in the pre-pandemic serum pool (light purple), pandemic serum pool (dark purple), or in both serum pools (orange). Counts indicate the numbers of peptides from each organism determined to have high-fidelity IgG binding distinctly in the pandemic, pre-pandemic, or both serum pools. HCMV = human cytomegalovirus, EBOZM = Zaire ebolavirus (strain Mayinga-76), HCoV = human coronavirus.

representing said proteome (Table 2). Henceforth, we will refer to this length-adjusted measure of epitope recognition as "epitope density", or simply "densities".

As a "non-overlapping" negative control, we used 3900 randomly generated 13-mer peptides (NP). Since these peptides have no overlapping neighbouring peptides, we only applied the "≥ 20% of the maximum signal" criteria. A mere 0.28% of the NP yielded an IgG signal of this magnitude when detected by the pre-pandemic and 0.82% of the NP by the pandemic serum (only 0.10% of the NP were shared). As a more natural "overlapping" negative control, we used 2417 overlapping peptides representing the complete proteome of Zaire ebolavirus (strain Mayinga-76) (EBOZM). Here, 0.54% and 0.87% of the EBOZM-derived peptides (EPs) qualified as HFP when detected by the pre-pandemic and pandemic sera, respectively (only 0.21% of the EP were shared). Using these densities as an approximation, we assume that the IgG HFP false discovery rate (FDR) is 0.54% and 0.87% for the pre-pandemic and pandemic serum pools, respectively (Table 2). By comparison, we found the IgA HFP FDRs to be 0.29% and 0% for the pre-pandemic and pandemic serum pools, respectively (Table 2).

As a positive control (PP), we used 31688 overlapping peptides representing the complete proteome of the HCMV. Given

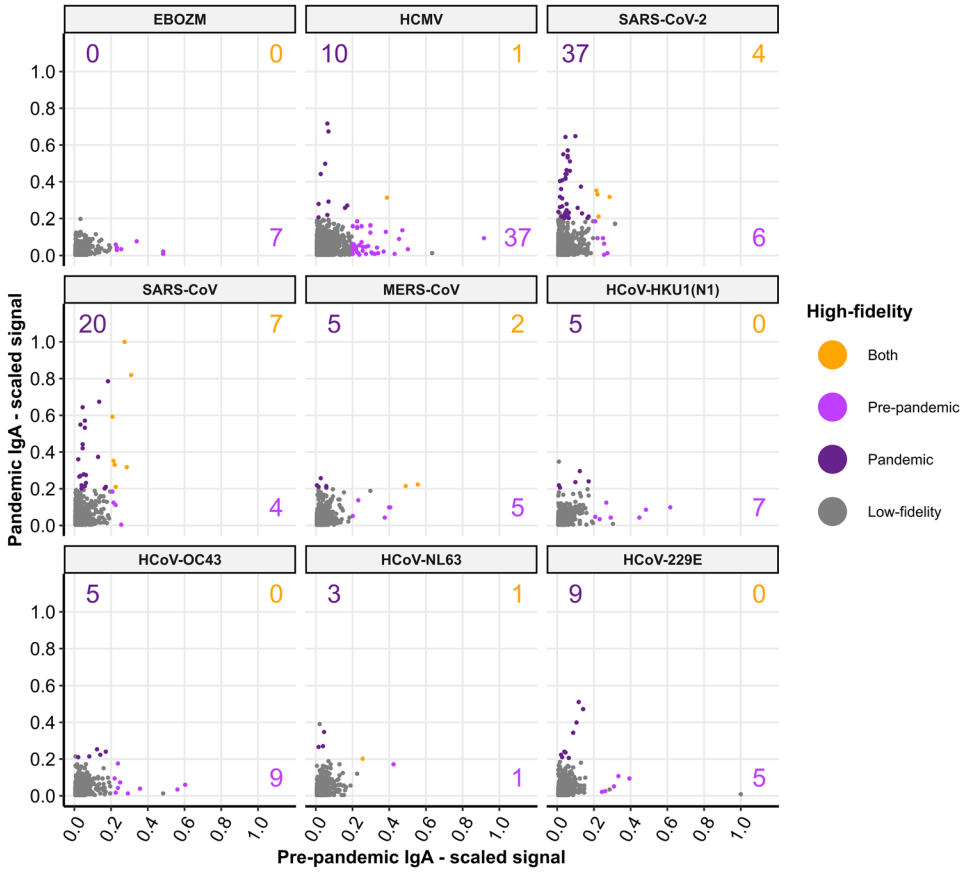

Figure 2. Peptide IgA binding signals on the microarrays. Scatter plot comparing peptide microarray IgA binding signals between the pandemic (y-axis) and pre-pandemic (x-axis) serum pools scaled to the maximum signal measured for each serum pool. Each data point represents one peptide derived from one of the organisms indicated in the plot subtitle. High-fidelity peptides passed our fidelity filters: combined p-value ≤ 0.001 & scaled signal ≥ 20% of the maximum scaled signal measured for the individual serum pool and Ig-type. The legend indicates if a peptide was determined to have high-fidelity IgG binding in the pre-pandemic serum pool (light purple), pandemic serum pool (dark purple), or in both serum pools (orange). Counts indicate the numbers of peptides from each organism determined to have high-fidelity IgA binding distinctly in the pandemic, pre-pandemic, or both serum pools. HCMV = human cytomegalovirus, EBOZM = Zaire ebolavirus (strain Mayinga-76), HCoV = human coronavirus.

HCMV's latent nature and its tendency to reactivate periodically, it has typically triggered a robust immune response. Due to its widespread prevalence in the population, we expected that peptides from HCMV would be recognised by serum pools from any randomly selected group of eight individuals. Our results confirmed this expectation: 1.10% and 1.88% of the peptides in the positive control qualified as HFPs when analysed using pre-pandemic and pandemic sera, respectively, with 0.81% being recognised by both. Comparatively, these densities were significantly higher than the corresponding EP densities ($P < 0.05$ and $P < 0.01$ for the pre-pandemic and pandemic serum pools, respectively (2-sample test for equality of proportions with multiple test correction by the Benjamini-Hochberg procedure. Hence forward: 2StepBH). As expected, the density of shared pre-pandemic/pandemic HCMV epitopes was the highest of any virus tested here.

Early COVID-19 pandemic sera exhibited robust IgG and IgA responses against SARS-CoV and SARS-CoV-2

The strategy described above was used to identify high-fidelity linear B cell epitopes from HCoVs. Compared to the IgG and IgA epitope density of EBOZM, the pre-pandemic serum pool HCoV epitope densities were not significantly enriched (Table 2). Reassuringly, this indicated no prior exposure to any of the epi- or pandemic HCoVs in the pre-pandemic serum pool. In contrast, the pandemic serum exhibited significantly higher densities of IgG and IgA epitopes specific to SARS-CoV-2 and SARS-CoV, at 2.45% and 1.84% for IgG, compared to 0.87% for EBOZM IgG, and at 0.65% and 0.42% for IgA, compared to 0% for EBOZM IgA, ($P < 0.01$, 2StepBH), reflecting the ongoing COVID-19 pandemic. As expected, the SARS-CoV-2 epitope density was the highest of any virus tested in the pandemic serum pool. Finally, neither the pre-pandemic nor the pandemic serum pools had significantly enriched linear B cell epitope densities from any of the common cold coronaviruses (CCCs).

By visual comparison of the response strengths measured as the scaled signals between IgG (Figure 1) and IgA (Figure 2), IgG elicited higher antibody binding signals overall, likely reflecting the serum concentration differences between the two Ig-types. This trend of higher IgG binding translated into

the response breadths measured as the numbers of IgG vs IgA HFPs, which were 6- and 10-fold higher in the pre-pandemic and pandemic serum pools, respectively (Table 2). Interestingly, comparing virus-specific IgG epitope densities between the pandemic and pre-pandemic serum pools (Table 2), we found that the pandemic serum pool identified significantly more IgG HFPs across all the investigated viruses (Exact binomial test with multiple test correction by the Benjamini-Hochberg procedure, p-values < 0.01). The underlying mechanisms contributing to this overall enrichment in pandemic IgG responses are not immediately clear.

Coronavirus high-fidelity peptides primarily map to the nucleoprotein, the spike protein and the ORF1ab

Focusing on the SARS-2-CoV proteome, we mapped the scaled IgG and IgA binding signals of all the SARS-CoV-2 HFPs to their positions in the proteome (Figure 3). To illustrate this comprehensively, the SARS-CoV-2 proteome is depicted as a central line indicating the size and order of the SARS-CoV-2 proteins and the open reading frames (ORFs) in the proteome. The scaled IgG (blue) and IgA (red) signals of each SARS-CoV-2 HFP are shown as bars with a width corresponding to their length, placed corresponding to their position in the proteome and extending from the centre upwards for the pandemic responses or the centre downwards for the pre-pandemic responses. Non-binding peptides are indicated by grey bars. The highest number of SARS-CoV-2-specific linear B cell epitopes were recognised by IgG in the pandemic serum pool (155 HFPs), followed by IgA in the pandemic serum pool (41 HFPs), IgG in the pre-pandemic pool (34 HFPs) and IgA in the pre-pandemic pool (10 HFPs) (Table 2). When stratifying these HFPs across their SARS-CoV-2 parent proteins, the most immunogenic SARS-CoV-2-derived proteins appeared to be the nucleoprotein (59 IG HFPs and 19 IgA HFPs), followed by the ORF1ab (37 IG HFPs and 11 IgA HFPs) and/or the spike protein (36 IG HFPs and 11 IgA HFPs)

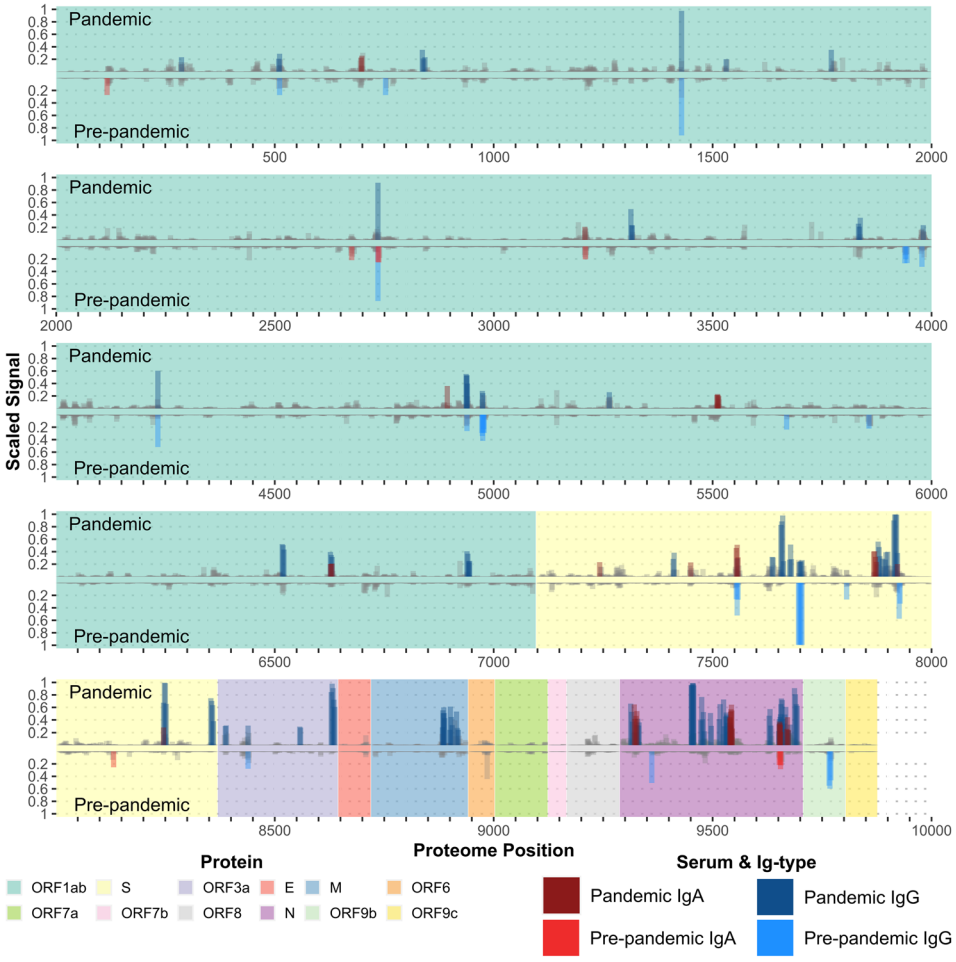

Figure 3. SARS-CoV-2 peptide antibody binding signals mapped to proteome. The y-axis displays standardised IgG and IgA signals on individual microarrays, scaled to their maximum values. Grey bars denote non-binder peptides for both IgG and IgA, while blue/dark blue bars represent high-fidelity IgG-binding peptides (HFPs), and red/dark red bars represent IgA-binding HFPs. Each bar corresponds to the width of a peptide (13 amino acids) and has reduced opacity for overlapping bars. Pandemic serum pool peptide signals are oriented upwards, and pre-pandemic control serum pool peptide signals are oriented downwards. The x-axis shows the SARS-CoV-2 proteome concatenated N- to C-terminal with tick marks at every 50th amino acid position. Each SARS-CoV-2 protein is indicated by the coloured rectangles in the background and is ordered by their start codon position in the reference genome Wuhan-1 (NC_045512.2). HFP = high-fidelity peptide, ORF = Open Reading Frame, S = Spike protein, E = Envelope protein, M = Membrane protein, N = Nucleocapsid protein.

recognised by the pandemic serum pool (Table 3). In contrast, the ORF1ab appeared to be the most frequently recognised SARS-CoV-2 protein by the pre-pandemic serum pool (18 IG HFPs and 5 IgA HFPs), probably due to cross-reactions with CCCs (see below).

Instead of using the absolute numbers of HFP, one could also use the protein-wise linear B cell epitope densities, thereby adjusting for the differences in lengths and numbers of peptides derived from these proteins. The shortest and longest immunogenic SARS-CoV-2 proteins determined here range from ORF9b at 97 amino acids in length, to ORF1ab, which is 7096 amino acids in length, resulting in the synthesis of 47 and 4884 peptide sequences, respectively (Table 3). By this token, the most densely pandemic IgG HFP populated of the five immunogenic proteins remained the nucleoprotein, and the least is the ORF1ab (out of the proteins with responses, Table 3). These findings concerning which coronavirus proteins were the most immunogenic in terms of absolute numbers and densities of epitopes could be extended to the entire set of HCoV (data not shown).

The HFPs recognised by the pre-pandemic serum pool tended to overlap with those recognised by the pandemic serum pool. This was particularly noteworthy for the ORF1ab. One could speculate whether these are caused by pre-existing antibody responses against endemic CCC and/or cross-reactions between SARS-CoV-2 and CCC responses. In contrast, many of the HFPs in the structural proteins of the SARS-CoV-2 virion particle (Spike (S), Membrane (M), and Nucleocapsid (N) proteins) and ORF3a were exclusive to the pandemic serum pool (Figure 3). Several of the HFPs recognised by the pandemic IgA were found in the S and N proteins overlapping the IgG HFPs, suggesting that these proteins are highly immunogenic antigens capable of eliciting diverse antibody responses across multiple Ig-isotypes.

Merging high-fidelity peptides into high-fidelity regions (HFR)

Many of the HFPs from the same serum pools and Ig-types overlapped (Figure 3). The exact footprint(s) of the linear B cell epitopes in the overlapping HFPs cannot be determined from these data; however, we reasoned that the footprints could occur within several overlapping HFPs and that neighbouring and overlapping footprints could merge into high-fidelity regions (HFRs). Such an approach could encompass larger immunogenic structures. Overall, the average length of these HFRs was 18 amino acids, representing 3–4 merged HFPs (range 13 to 157 amino acids, representing 1 to 19 HFPs). The number of unique HFRs detected in each organism, serum pool, and Ig-type grouping are summarised in Table 4. In total, we identified 280 unique HCoV HFRs, 201 HFRs by the pandemic serum pool and 102 by the pre-pandemic pool (53 HFRs were identical between the two serum pools). As expected, and echoing the HFP analysis of HCoV above, the largest contribution, with 72 HFRs, came from SARS-CoV-2, the virus that caused the ongoing pandemic. The second largest contribution, with 53 HFRs, came from the closely related SARS-CoV, likely reflecting cross-reactions from SARS-CoV-2.

Homology clustering of high-fidelity regions suggests antibody cross-reactivity between human coronaviruses

To assess the potential of antibody cross-reactivity against the HCoV-derived HFRs, all the unique HFRs identified from the seven HCoVs, HCMV and EBOZM, were submitted to the Immune Epitope Database (IEDB) Epitope Cluster Analysis web tool (http://tools.iedb.org/cluster/). To avoid false positive clustering, we empirically selected a sequence identity threshold of 60% where only 1 of the 271 HCMV HFRs and none of the 21 Ebola HFR clustered with the HCoV HFRs. The alignments and clustering of all the HCoV HFRs are summarised in Table 5. For completeness, the alignments and clustering of all HCMV and EBOZM HFRs are summarised in Table 6.

Of the 333 (280 unique) HCoV-derived HFRs, 225 HFRs could be grouped into 78 unique clusters of 2–15 HFR members (Table 3). The remaining 108 HFRs were singletons, which did not cluster with any other HCoV HFRs. Each of the 186 clusters and singletons was given a unique Alignment ID number from 1 to 186. Figure 4A features a 4-circle Venn diagram that categorises alignment clusters as either shared or uniquely recognised, contingent upon the 4 combinations of serum pools: pre-pandemic or pandemic and IgG or IgA (detailed in Figure 4B). Notably, 138 out of 186 (74%) of these clusters were exclusive to specific serum and isotype pairings (Figure 4A). Conversely, shared clusters among two, three, and all four combinations were less common, with counts of 41 (22%), 7 (4%), and 0 (0%), respectively (Figure 4A). This relative lack of shared recognition underscores the previously proposed individuality of linear B cell recognition. From a more practical point of view, searching for possible cross-reactions against different HCoV strains becomes more challenging simply because there are fewer examples of shared recognition to investigate.

SARS-CoV, MERS-CoV, and SARS-CoV-2 (before the COVID-19 pandemic) have not circulated in Denmark and should only be recognised by the pre-pandemic serum pool through cross-reactions with common cold coronaviruses (CCCs). Likewise, for the pandemic serum pool, antibody responses targeting SARS-CoV or MERS-CoV should be the result(s) of cross-reactivities with either the CCCs and/or SARS-CoV-2. Using the homology clustering approach (Table 5), for SARS-CoV, we found that 100% of the IgA HFRs and 81% of the IgG HFRs found in the pandemic serum pool could be explained by cross-reactivity to either SARS-CoV-2 and/or one or more of CCCs, consistent with the sequence similarity between SARS-CoV and SARS-CoV-2. However, for MERS-CoV, only 25% and 21% of the IgA and IgG HFRs could be explained by alignment to SARS-CoV-2 and/or the CCCs. This lower degree of alignment is consistent with MERS-CoV sharing the least sequence similarity with the other HCoVs but begs the question of how these antibody responses arose. For the pre-pandemic serum pool, none of the SARS-CoV, MERS-CoV, or SARS-CoV-2-specific IgA HFRs aligned with any of the CCCs, whereas 30%, 29%, and 17% of the SARS-CoV, MERS-CoV, or SARS-CoV-2-specific IgG HFRs could be explained by alignment with one or more of the

Table 6. Homology clustering alignments of *human cytomegalovirus* and *ebolavirus* high-fidelity regions

Align ID	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
5	HCMV	P08318	UL32				X	593	667	--ILTPTPVNPSTAPAPAPTPTFAGTQTPVNGNSPWAPTAPLPGDMNPANWPRERAWALKNPHLAYNPFRMPTTSTA
5	HCMV	P08318	UL32		X			591	621	AAILTPTPVNPSTAPAPAPTPTFAGTQTPVN----------
5	HCMV	P08318	UL32		X			623	645	----------NSPWAPTAPLPGDMNPANWPRER----------
5	HCMV	P08318	UL32		X			653	665	----------HLAYNPFRMPTTS--
6	HCMV	P08318	UL32				X	937	1048	--TTPTYPAVTTVYPPSSTAKSSVSNAPPVASPSILKPGASAALQSRRSTGTAAVGSPVKSTTGMKTVAFDLSSPQKSG TGPQPGSAGMGGAKTPSDAVQNILQKIEKIKNTEE
6	HCMV	P08318	UL32		X			943	987	--------AVTTVYPPSSTAKSSVSNAPPVASPSILKPGASAALQSRRSTGTA---------- ----------
6	HCMV	P08318	UL32		X			991	1023	---------- SPVKSTTGMKTVAFDLSSPQKSGTGPQPGSAGM----------
6	HCMV	P08318	UL32		X			1025	1045	GAKTP---------- ----------SDAVQNILQKIEKIKN---
8	HCMV	P08318	UL32				X	507	583	RVRRDSAWDVRPLTETRGDLFSGDEDSDSSDGYPPNRQDPRFTDTLVDITDTETSAKPPVTTAYKFEQPTLTFGAGV- -
8	HCMV	P08318	UL32		X			519	551	----------LTETRGDLFSGDEDSDSSDGYPPNRQDPRFTDT---------- --
8	HCMV	P08318	UL32		X			559	585	----------ETSAKPPVTTAYKFEQPTLTFGAGVNV
9	HCMV	P08318	UL32				X	675	711	TPRRPSTPRAAVTQTASRDAADEVWALRDQTAESPVE
9	HCMV	P08318	UL32		X			695	711	----------ADEVWALRDQTAESPVE
9	HCMV	P08318	UL32		X			675	689	TPRRPSTPRAAVTQT----------
22	HCMV	P06725	UL83				X	389	435	DEGAAQGDDDVWTSGSDSDEELVTTERKTPRVTGGGAMAGASTSAGR
22	HCMV	P06725	UL83		X			405	419	----------DSDEELVTTERKTPR----------
23	HCMV	P06725	UL83				X	453	473	RGRLKAESTVAPEEDTDEDSD
23	HCMV	P06725	UL83		X			455	469	--RLKAESTVAPEEDTD----
24	HCMV	P08318	UL32		X			885	939	APPSPVKGRGSRVGVPSLKPTLGGKAVVGRPPSVPVSGSAPGRLSGSSRAASTTP
24	HCMV	P08318	UL32				X	885	933	APPSPVKGRGSRVGVPSLKPTLGGKAVVGRPPSVPVSGSAPGRLSGSSR------
25	HCMV	P08318	UL32				X	361	379	RSAAVGYYDEEEKRRERQK
25	HCMV	P08318	UL32		X			363	379	--AAVGYYDEEEKRRERQK
26	HCMV	P09703	US27				X	315	351	CCVKQEIPYQDIDIELQKDIQRRAKHTKRTHYDRKNA
26	HCMV	P09703	US27		X			321	337	------IPYQDIDIELQKDIQRR----------
27	HCMV	P13200	UL99		X			45	75	EGEDDDDGEDDDNEERQQKLRLCGSGCGGND
27	HCMV	P13200	UL99				X	49	75	----DDDGEDDDNEERQQKLRLCGSGCGGND

Align ID	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
28	HCMV	P13200	UL99				X	87	117	DGSKKNAVRSTFREDKAPKPSKQSKKKKKPS
28	HCMV	P13200	UL99		X			87	103	DGSKKNAVRSTFREDKA----------
29	HCMV	P13200	UL99				X	131	190	TDDLDEEDTSIYLSPPPVPPVQVVAKRLPRPDTPRTPRQKKISQRPPTPGTKKPAASLPF
29	HCMV	P13200	UL99		X			131	187	TDDLDEEDTSIYLSPPPVPPVQVVAKRLPRPDTPRTPRQKKISQRPPTPGTKKPAAS---
30	HCMV	P13202	UL123				X	47	73	--------QLSLGDPLFPELAEESLKTFEQVTEDC
30	HCMV	P19893	UL122				X	47	73	--------QLSLGDPLFPELAEESLKTFEQVTEDC
30	HCMV	P13202	UL123		X			39	63	MLRKEVNSQLSLGDPLFPELAEESL----------
30	HCMV	P19893	UL122		X			39	63	MLRKEVNSQLSLGDPLFPELAEESL----------
31	HCMV	P16789	UL98				X	283	297	YKYLRKKEDPFVQNV
31	HCMV	P16789	UL98	X	X			283	295	YKYLRKKEDPFVQ--
32	HCMV	P16790	UL44				X	397	433	GLDRNSGNYFNDAKEESDSEDSVTFEFVPNTKKQKCG
32	HCMV	P16790	UL44		X			399	417	--DRNSGNYFNDAKEESDSED----------
33	HCMV	P16823	UL71				X	331	361	PPPQIPFVGSARELRGVKKKKPTAAALLSSA
33	HCMV	P16823	UL71		X			337	361	------FVGSARELRGVKKKKPTAAALLSSA
34	HCMV	P16823	UL71				X	267	301	VQQQPPVEPVQQQPQGRGSHRRRYKESAPQETLPT
34	HCMV	P16823	UL71		X			283	299	----------RGSHRRRYKESAPQETL--
35	HCMV	P17147	DBP	X				405	421	ARKVLKFQGAPLKDERY
35	HCMV	P17147	DBP			X		405	417	ARKVLKFQGAPLK----
47	HCMV	P06473	gB				X	65	85	HRANETIYNTTLKYGDVVGVN
47	HCMV	P06473	gB		X			67	83	--ANETIYNTTLKYGDVVG--
48	HCMV	P06473	gB				X	807	825	LQAPPSYEESVYNSGRKGP
48	HCMV	P06473	gB		X			811	825	----PSYEESVYNSGRKGP
49	HCMV	P08318	UL32				X	793	831	TVTSTPVQGRLEKQVSGTPSTVPATLLQPQPASSKTTSS
49	HCMV	P08318	UL32		X			793	827	TVTSTPVQGRLEKQVSGTPSTVPATLLQPQPASSK----
50	HCMV	P13200	UL99				X	15	33	TTPGEPLKDALGRQVSLRS
50	HCMV	P13200	UL99		X			17	33	--PGEPLKDALGRQVSLRS
51	HCMV	P69338	UL132				X	221	251	EEDPIRYYVSVYDELTASEMEEPSNSTSWQI
51	HCMV	P69338	UL132		X			221	239	EEDPIRYYVSVYDELTASE----------

Align ID	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
58	HCMV	P08318	UL32		X			433	453	DDDDDEKNIFTPIKKPGTSGK
58	HCMV	P08318	UL32				X	435	453	--DDDEKNIFTPIKKPGTSGK
59	HCMV	P13202	UL123		X			1	21	MESSAKRKMDPDNPDEGPSSK
59	HCMV	P19893	UL122		X			1	21	MESSAKRKMDPDNPDEGPSSK
59	HCMV	P13202	UL123				X	1	19	MESSAKRKMDPDNPDEGPS--
59	HCMV	P19893	UL122				X	1	19	MESSAKRKMDPDNPDEGPS--
60	HCMV	P16732	TRM3		X			83	101	LLFHERLKSALDKLTFRPC
60	HCMV	P16732	TRM3	X				85	99	--FHERLKSALDKLTFR--
61	HCMV	P16753	UL80		X		X	695	708	LNRRIFVAALNKLE
61	HCMV	P16753	UL80	X				695	707	LNRRIFVAALNKL-
62	HCMV	P16761	UL25				X	85	101	------KPSEKKRSSSRRQPQIA
62	HCMV	P16761	UL25			X		79	93	GTGAAKKPSEKKRSS--------
63	HCMV	P16761	UL25				X	107	127	GSPATPKAGKSPKVSRPPSVP
63	HCMV	P16761	UL25		X			111	127	----TPKAGKSPKVSRPPSVP
64	HCMV	P16808	IRL10				X	131	151	HAAKKQNIYERIPYRPSRQKD
64	HCMV	P16808	IRL10		X			131	147	HAAKKQNIYERIPYRPS----
65	HCMV	P16828	UL107		X			69	89	FVVNRLLPRTRSTTKPSCSFA
65	HCMV	P16828	UL107				X	73	85	----RLLPRTRSTTKPS----
66	HCMV	P21602	TRL14				X	163	186	HRIPKKPHYTKLPKYDPDEFWTKA
66	HCMV	P21602	TRL14		X			169	186	------PHYTKLPKYDPDEFWTKA
74	HCMV	P16752	UL79				X	201	215	YFYLIGTGKMRITPD
74	HCMV	P16752	UL79		X			203	215	--YLIGTGKMRITPD
75	HCMV	P16833	UL116		X			139	153	ASDYTPTFDDLENIT
75	HCMV	P16833	UL116				X	141	153	--DYTPTFDDLENIT
80	EBOZM	Q05318	L				X	2117	2129	--IRTAKGRITKLVN
80	EBOZM	Q05318	L	X				2115	2127	HFIRTAKGRITKL--
81	HCMV	P09724	US20		X			323	342	SDGAPLDNGTLTAASTTGKS
81	HCMV	P09724	US20				X	323	337	SDGAPLDNGTLTAAS-----
82	HCMV	P16737	UL106	X				81	99	SLARRHRGPGKRISLRTEI

Open Research Europe 2024, 4:215 Last updated: 19 DEC 2024

Align ID	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
82	HCMV	P16737	UL106				X	85	97	----RHRGPGKRISLRT--
83	HCMV	P16772	UL130				X	33	51	ANQNPSPPWSKLTYSKPHD
83	HCMV	P16772	UL130		X			33	47	ANQNPSPPWSKLTYS----
84	HCMV	P19893	UL122				X	147	163	KKPRKTTRPFKVIIKPP
84	HCMV	P19893	UL122	X	X			149	163	--PRKTTRPFKVIIKPP
90	HCMV	P08318	UL32				X	459	487	GGVSSIFSGLLSSGSQKPTSGPLNIPQQQ
90	HCMV	P08318	UL32		X			459	479	GGVSSIFSGLLSSGSQKPTSG--------
91	HCMV	P16724	TRM1			X		199	217	TIIPNQGRSLNKRLQGLLC
91	HCMV	P16724	TRM1				X	203	217	----NQGRSLNKRLQGLLC
94	EBOZM	Q05318	L		X			1399	1415	PFFQGKRLNIIEDDLIR--
94	EBOZM	Q05318	L				X	1405	1417	------RLNIIEDDLIRLP
95	HCMV	P16784	UL47			X		799	811	AQMSKALKKTASK
95	HCMV	P16791	NEC2	X				197	209	ARMSAALTRTAGK
96	HCMV	P08318	UL32				X	849	871	PSASASVLSPTEDDVVSPATSPL
96	HCMV	P08318	UL32		X			853	867	----ASVLSPTEDDVVSPA----
97	HCMV	P16749	UL69		X			77	95	PCVPARRPRYSKDDDTEGD
97	HCMV	P16749	UL69				X	77	93	PCVPARRPRYSKDDDTE--
98	HCMV	P19893	UL122				X	177	193	EDIKPEPDFTIQYRNKI
98	HCMV	P19893	UL122		X			181	193	----PEPDFTIQYRNKI
105	HCMV	P16760	UL24		X			305	321	YACDVRDDRYIRVGDNL
105	HCMV	P16760	UL24				X	305	317	YACDVRDDRYIRV----
107	HCMV	P09710	HKLF1		X			219	235	PRIRITPISTSPRRKPP
107	HCMV	P09710	HKLF1				X	221	235	--IRITPISTSPRRKPP
108	HCMV	P16766	UL35		X			617	635	MPSTSYGSDVEDPRDDLAE
108	HCMV	P16766	UL35				X	617	633	MPSTSYGSDVEDPRDDL--
109	HCMV	P69338	UL132				X	155	175	SSSPYQRLETRDWDEEEEASA
109	HCMV	P69338	UL132		X			159	171	----YQRLETRDWDEEE----
110	HCMV	P19893	UL122		X			241	259	QMNHPPLPDPLGRPDEDSS
110	HCMV	P19893	UL122				X	243	259	--NHPPLPDPLGRPDEDSS

Align ID	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
111	HCMV	P16790	UL44		X			307	325	GLHDDGPGLDNDLMNEPMG
111	HCMV	P16790	UL44				X	309	325	--HDDGPGLDNDLMNEPMG
113	HCMV	P06473	gB				X	39	57	SHTSRTTSAQTRSVYSQHV
113	HCMV	P06473	gB		X			39	55	SHTSRTTSAQTRSVYSQ--
114	HCMV	P16833	UL116				X	187	205	AVDEEEPELEDDDEYDELW
114	HCMV	P16833	UL116	X				189	201	--DEEEPELEDDDEY----
115	EBOZM	P18272	NCAP	X				105	119	RFEVKKRDGVKRLEE
116	EBOZM	P18272	NCAP	X				163	175	VGEKACLEKVQRQ
117	EBOZM	P18272	NCAP				X	269	281	RTAKVKNEVNSFK
118	EBOZM	Q05323	VP30				X	51	63	PTVFHKKRVEPLT
119	HCMV	P06473	gB		X		X	875	891	QDKGQKPNLLDRLRHRK
120	HCMV	P06726	UL82				X	371	389	TLTPGKPDKIKRPQVQLRA
121	HCMV	P08318	UL32		X		X	749	787	TPEQSTPSRIRKAKLSSPMTTTSTSQKPVLGKRVATPHA
122	HCMV	P09699	US26	X				13	25	VRRTLKGLRKLIL
123	HCMV	P09706	US30				X	295	311	KLQEEKKRQSALVWNVE
124	HCMV	P09731	US7				X	41	53	EFVKRQLAPKPKS
125	HCMV	P16733	gM		X		X	347	372	SHHDSLESRRLREEEDDDDDEDFEDA
126	HCMV	P16733	gM				X	319	331	VRYFRGRGSGSVK
127	HCMV	P16735	UL104				X	247	259	RVFKKVRSERLEA
128	HCMV	P16750	GO				X	29	65	NCKVMSKALYNRPWRGLVLSKIGKYKLDQLKLEILRQ
129	HCMV	P16758	UL20				X	175	193	TVEWYNKSEGDVPEEFMDY
130	HCMV	P16761	UL25				X	63	79	VSSDAPSSFELVRETGG
131	HCMV	P16767	UL36				X	397	409	EGVRQKLRARTGS
132	HCMV	P16785	UL48				X	895	911	GTGASEKRFQALRELVY
133	HCMV	P16785	UL48				X	283	295	KTPEKRRKDLSGS
134	HCMV	P16785	UL48				X	1205	1217	SSVTRVPEKPFKG
135	HCMV	P16789	UL98	X				139	151	RQVELFQLLEKES
136	HCMV	P16790	UL44		X		X	269	293	EEPFQRGDPFDKNYVGNSGKSRGGG
137	HCMV	P16794	NEC1				X	261	279	QRKIDEMDIPNDVSESFER

Align ID	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
138	HCMV	P16798	UL92				X	171	187	VIKVSRKKREDALLKQM
139	HCMV	P16799	CVC1				X	579	591	PETPRVSGRRLPF
140	HCMV	P16823	UL71				X	243	255	FDQRLLEEGDEED
141	HCMV	P16840	US5				X	43	55	AVLKERAFIETIK
142	HCMV	P17147	DBP				X	1049	1061	LIKRSRRENVLFD
143	HCMV	P17151	UL112/UL113				X	227	255	ANGLLRDPRLMNRQKERRPPPSSENDGSP
144	HCMV	P19893	UL122				X	323	337	RKKKSKRISELDNEK
145	HCMV	Q7M6G6	US34A				X	5	17	LLKFRKRRRPVVV
161	HCMV	P06725	UL83	X				163	183	VSGLAWTRQQNQWKEPDVYYT
162	HCMV	P08318	UL32				X	387	401	QREKQQLKAWEERQQ
163	HCMV	P08546	UL54				X	957	973	LMMICKKRYIGKVEGAS
164	HCMV	P16750	GO				X	257	271	QAPVKEQFEKKAKKT
165	HCMV	P16753	UL80		X		X	615	637	TSGGGETPTALLSGGAKVAERAQ
166	HCMV	P16779	UL38		X			179	191	HQLDYEKQVLEAR
167	HCMV	P16785	UL48			X		1021	1033	VQQDLQRVVDAVK
168	HCMV	P16791	NEC2		X			131	149	TKCVVRKSNSASVFNIELI
169	HCMV	P16819	UL62		X			41	55	HVGAQNGMTSERTVE
183	EBOZM	P60170	VSGP				X	261	279	YTSGKRSNTTGKLIWKVNP
183	EBOZM	Q05320	VGP				X	261	279	YTSGKRSNTTGKLIWKVNP
183	EBOZM	Q9YMG2	VSSGP				X	261	279	YTSGKRSNTTGKLIWKVNP
184	EBOZM	Q05128	VP40				X	267	281	LVHKLTGKKVTSKNG
185	EBOZM	Q05318	L	X				545	559	YRNFSFSLKEKELNV
186	EBOZM	Q05320	VGP		X		X	427	447	AGPPKAENTNTSKSTDFLDPA
187	HCMV	P06473	gB	X				403	419	IFNTSYNQTYEKYGNVS
188	HCMV	P06473	gB				X	367	381	SSAKMTATFLSKKQE
189	HCMV	P09722	US22		X			385	401	TRGTDKVPNNSSAWGRT
190	HCMV	P16720	UL6		X			201	215	VSLVASRLTTNSSAF
191	HCMV	P16740	UL120		X			135	155	MVNNTDSNTNTSIIHFVQQDE

Open Research Europe 2024, 4:215 Last updated: 19 DEC 2024

Align ID	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
192	HCMV	P16750	GO				X	67	85	ETTISTKYNVSKQPVKNLT
193	HCMV	P16750	GO				X	389	405	FVENKTASDSNKTTPTS
194	HCMV	P16751	UL78				X	11	29	VTESIMFAIVSFKHMGPFE
195	HCMV	P16782	RIR1		X			61	77	GDKEASAGNTNVSSGVA
196	HCMV	P16801	UL95				X	179	197	SGPEDKKKNVGNKTAGEKN
197	HCMV	P16805	IRL7				X	41	59	LVNQRHKIFTQSFLRTFRN
198	HCMV	P16848	UL31				X	171	183	YNTASNVTEKASY
199	HCMV	P16849	UL33				X	367	383	NQHAKRSVSFNFPSGTW
200	HCMV	Q7M6N6	SCP		X		X	1	19	MSNTAPGPTVANKRDEKHR
220	EBOZM	P60170	VSGP				X	209	225	PSSGYYSTTIRYQATGF
220	EBOZM	Q05320	VGP				X	209	225	PSSGYYSTTIRYQATGF
220	EBOZM	Q9YMG2	VSSGP				X	209	225	PSSGYYSTTIRYQATGF
221	EBOZM	Q05318	L				X	1039	1051	PSGKRLQILGYLE
222	EBOZM	Q05320	VGP		X			323	335	PARTSSDPGTNTT
223	HCMV	P09703	US27				X	57	69	YYRRKKKSPSDTY
224	HCMV	P16719	UL1		X			89	109	YLRNVTHYYTNTSCSPQFMCI
225	HCMV	P16729	MCP				X	831	843	TTSKESTSGVTPE
226	HCMV	P16732	TRM3		X			331	343	TKIIFISSTNTTS
227	HCMV	P16743	UL7				X	75	89	LICFYKRGEGLSTPY
227	HCMV	P16744	UL8				X	75	89	LICFYKRGEGLSTPY
228	HCMV	P16750	GO				X	193	205	PTKVNVDSQTIYF
229	HCMV	P16753	UL80				X	417	429	SYGAPVVGYDQLA
230	HCMV	P16758	UL20		X			63	75	RFTPANTTTNSST
231	HCMV	P16790	UL44		X		X	371	387	KYEQHKITSYLTSKGGS
232	HCMV	P16805	IRL7				X	9	21	NASKRRPSQTQYD
233	HCMV	P16816	UL58				X	9	23	PRWYKRDESGLATSI
234	HCMV	P16824	DUT		X			233	245	VIGPGHTQTVYFD
235	HCMV	P16845	UL22A				X	41	53	GSNTTPSKNVTLS
236	HCMV	P17149	UL70		X			163	177	YYLFEGQYSTISTVE

Align ID	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
237	HCMV	P69338	UL132				X	35	47	STNVPTSTSSRNT
245	EBOZM	P60170	VSGP		X			333	345	TQQLKTTKSWLQK
246	EBOZM	Q05128	VP40		X			219	231	PNKSGKKGNSADL
247	EBOZM	Q05318	L	X				1065	1077	NTETPVLDRLRKI
248	EBOZM	Q05318	L				X	49	63	HIYRLKYDVTVTKFL
249	EBOZM	Q05318	L		X			1475	1489	YLGNTILRTKKLTLD
250	HCMV	P08318	UL32				X	729	747	LGHTTPSSDYNNDVISPPS
251	HCMV	P16727	UL84				X	563	575	DCKLTENTTEKTS
252	HCMV	P16733	gM		X		X	101	117	KISMRKDKGMSLNQSTR
253	HCMV	P16736	HELI				X	13	27	TPKYDDALILNLSSA
254	HCMV	P16766	UL35				X	339	351	YKLLKKIQSVSDA
255	HCMV	P16778	UL37				X	365	377	MCPRFLKRTVKLT
256	HCMV	P16795	gN				X	65	77	TKLSSTTHDPNVM
257	HCMV	P16799	CVC1				X	329	341	LWYVNSFWRSREL
258	HCMV	P16808	IRL10		X			41	53	LTCKCSPNNTSSN
259	HCMV	P16810	IRL12				X	155	167	TNTALISTPWLTN
260	HCMV	P17149	UL70	X				631	645	CRCKGKLGLRIITPF
261	HCMV	P69334	US18				X	153	165	TYLAGNNPTRWKV
262	HCMV	P69338	UL132				X	181	193	KHDPENVIYFRKD
290	HCMV	P13202	UL123				X	31	43	KATTFLQTMLRKE
290	HCMV	P19893	UL122				X	31	43	KATTFLQTMLRKE
292	HCMV	P16784	UL47		X			659	671	LSGVPVTPNVDLL
296	HCMV	P08318	UL32				X	295	309	PSLEFDDDANLLSLG
297	HCMV	P17147	DBP				X	1219	1235	SGGGGGSGLLPAKRSRL
307	EBOZM	Q05318	L				X	475	487	ISDLSIFIKDRAT
308	HCMV	P09715	IRS1	X				591	605	LPEKYIGFYQIRKPP
309	HCMV	P16732	TRM3				X	637	653	TARYSAKRQNRISDDLI
310	HCMV	P16753	UL80		X		X	573	589	IHELKRDLFAARQSSTL
311	HCMV	P16770	UL117	X				33	47	TIVVAPRRGFLRIPP

Align ID	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
312	HCMV	P16772	UL130				X	149	161	AHMVPKQTKLLRF
313	HCMV	P16781	UL43				X	271	287	ALWRLCMKREGSIFSWR
314	HCMV	P69332	US28				X	265	277	FERSLKRALILTE
317	HCMV	P06726	UL82				X	211	225	DALEQLACSDPNTYI
318	HCMV	P09700	US24				X	401	415	PHEGAWERLDPNTLV
319	HCMV	P16724	TRM1				X	21	37	MELECLKYCDPNVLLAE
320	HCMV	P16727	UL84		X		X	393	411	RLRRETVRRPFFSDAPLPY
321	HCMV	P16727	UL84				X	343	359	SVFGRLELDPNESPPDL
322	HCMV	P16761	UL25		X			41	59	STRARSATEDLDRMEAGLS
323	HCMV	P16770	UL117				X	331	345	ATFKFFHQDPNRVLD
324	HCMV	P16778	UL37	X				29	41	RLEDPLPPWLRKK
325	HCMV	P16808	IRL10				X	153	171	SPLIEPTGTDDEEDEDDNV
326	HCMV	P16833	UL116		X			169	193	HNLSIILYEEESQSSVDIAVDEEEP
327	HCMV	P17143	J1L				X	53	71	RRRPERSKTPDKRSAETTQ
328	HCMV	P17150	UL111A		X			109	123	ELHSMRSTLESIYKD
329	HCMV	P17151	UL112/ UL113				X	299	317	SLPLDTSEAVAFLNYSSSS
330	HCMV	P69338	UL132				X	115	129	RFRFADEEAVNLLDD
336	HCMV	P06473	gB				X	627	639	YVDYLFKRMIDLS
337	HCMV	P08560	H301				X	353	368	LRNLAFAWRYRKVRSS
338	HCMV	P16761	UL25				X	211	223	EYLKFVLQDFDVQ
339	HCMV	P16770	UL117				X	409	421	HLPLYNEFTSFRL
340	HCMV	P16792	TRM2	X				95	109	ARLNFVNAGQKYAAF
341	HCMV	P17149	UL70		X			497	509	FYYTRHEVFNERL
342	HCMV	P17150	UL111A		X			157	169	LSELDTLFSRLEE
349	HCMV	P16776	UL5				X	45	59	SCCYDITEAESNKAI
350	HCMV	P16809	IR11				X	169	183	AIISDLKRQWSGLSL
353	HCMV	P09715	IRS1				X	599	615	YQIRKPPWLMEQPPPPS
354	HCMV	P16772	UL130		X			57	73	CPFLYPSPPRSPLQFSG

Open Research Europe 2024, 4:215 Last updated: 19 DEC 2024

Align ID	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
355	HCMV	P16786	UL49	X				251	263	RGLIAALRRYAGK
360	HCMV	P16749	UL69		X			373	385	RQLLFNKLNNLDL
361	HCMV	P16776	UL5				X	149	161	RDIARPLKYRYQR
365	HCMV	P09722	US22	X				319	331	HMFLKCGLLKLRG
366	HCMV	P08318	UL32				X	491	505	AAFSLVSPQVTKASP
367	HCMV	P08318	UL32		X		X	413	427	PARKPSASRRLFGSS
368	HCMV	P13202	UL123		X			463	475	EEEEDGAEEPTAS
369	HCMV	P16727	UL84				X	33	47	SGKCRRQRRALSAPP
370	HCMV	P16753	UL80				X	521	535	SSSDEDLSFPGEAEH
371	HCMV	P16761	UL25				X	25	41	NEDTPADEEAEDSVFTS
372	HCMV	P16767	UL36				X	315	329	HPKKTPDPMIMFDED
373	HCMV	P16782	RIR1				X	53	65	LEEGPLPGGDKEA
374	HCMV	P16789	UL98				X	419	433	SPGDSPPPIPFEDEN
375	HCMV	P16795	gN				X	29	43	SATTSKSSASVSTTK
376	HCMV	P16818	UL61				X	119	131	AQPEPPRGSRRET
377	HCMV	P16823	UL71				X	313	327	HSPDVPREAVMSPTM
380	HCMV	P16749	UL69		X			455	471	RALQVTGTPQFFDQFDT
381	HCMV	P16822	UL66				X	85	103	SGTSCRRPSPGRRDVRRER
383	HCMV	P09700	US24				X	299	311	VYAIQMDDPNHYV
384	HCMV	P09712	US3				X	89	105	HDERGAYFEWNIGGHPV
386	HCMV	P09715	IRS1				X	751	763	RLCKTGPGPPPLP
387	HCMV	P16753	UL80		X		X	477	491	GMDEPPSGWERYDGG
388	HCMV	P17151	UL112/UL113				X	273	287	GGGGKNETEQQSGGA
389	HCMV	P17151	UL112/UL113				X	583	595	PPPSGGRRGRARN
390	HCMV	P17151	UL112/UL113				X	173	185	RDGPSAEEQQRRR
391	HCMV	P69338	UL132				X	257	270	VAMQPVSLRDPEYD
392	HCMV	Q7M6G6	US34A		X			47	59	RYEERLRKNRARR

Align ID	Organism	UniProt ID	Gene	Pre-pandemic IgA	Pre-pandemic IgG	Pandemic IgA	Pandemic IgG	Start	End	Alignment
395	HCMV	P09705	US29		X		X	201	219	TDPSRCDPYQVYFYGLQCP
396	HCMV	P12824	gH				X	669	683	YMHDSDDVLFALDPY
398	HCMV	P06473	gB		X			839	853	YTNEQAYQMLLALAR
399	HCMV	P06725	UL83	X				529	545	GVWQPAAQPKRRRHRQD
400	HCMV	P09715	IRS1				X	823	835	HRPRPPAPPRPKK
401	HCMV	P16725	UL76			X		309	321	VPRDRRGRSFCRP
402	HCMV	P16766	UL35				X	565	577	EQPPRQRRRMVSV
403	HCMV	P16818	UL61		X		X	219	235	GTGVRDPAPRGGRGRAR
404	HCMV	P16838	UL129			X		45	57	GSQPRAAGARRRM
405	HCMV	P13202	UL123		X			67	81	EQVTEDCNENPEKDV
405	HCMV	P19893	UL122		X			67	81	EQVTEDCNENPEKDV
406	HCMV	P09695	TRS1				X	7	23	MSPRPPPLGRGRGAGGP
406	HCMV	P09715	IRS1				X	7	23	MSPRPPPLGRGRGAGGP
407	HCMV	P09699	US26		X			115	127	EDSDDDDFFVYVE
408	HCMV	P09701	US23				X	507	521	SDDDEWEDLGFDLEE
409	HCMV	P09715	IRS1				X	721	737	SEPEDDDEDPTYDELPS
410	HCMV	P16749	UL69		X			17	29	LSERERRARRARR
411	HCMV	P16826	UL101		X			25	37	DGSRRGSRRGLST
412	HCMV	P69332	US28		X		X	11	27	TTEFDYDEDATPCVFTD
414	HCMV	P06726	UL82				X	405	423	FLPQSPGLPPTEEEEEEEE
417	HCMV	P16727	UL84	X				159	171	EPEKKKEKQEKKE

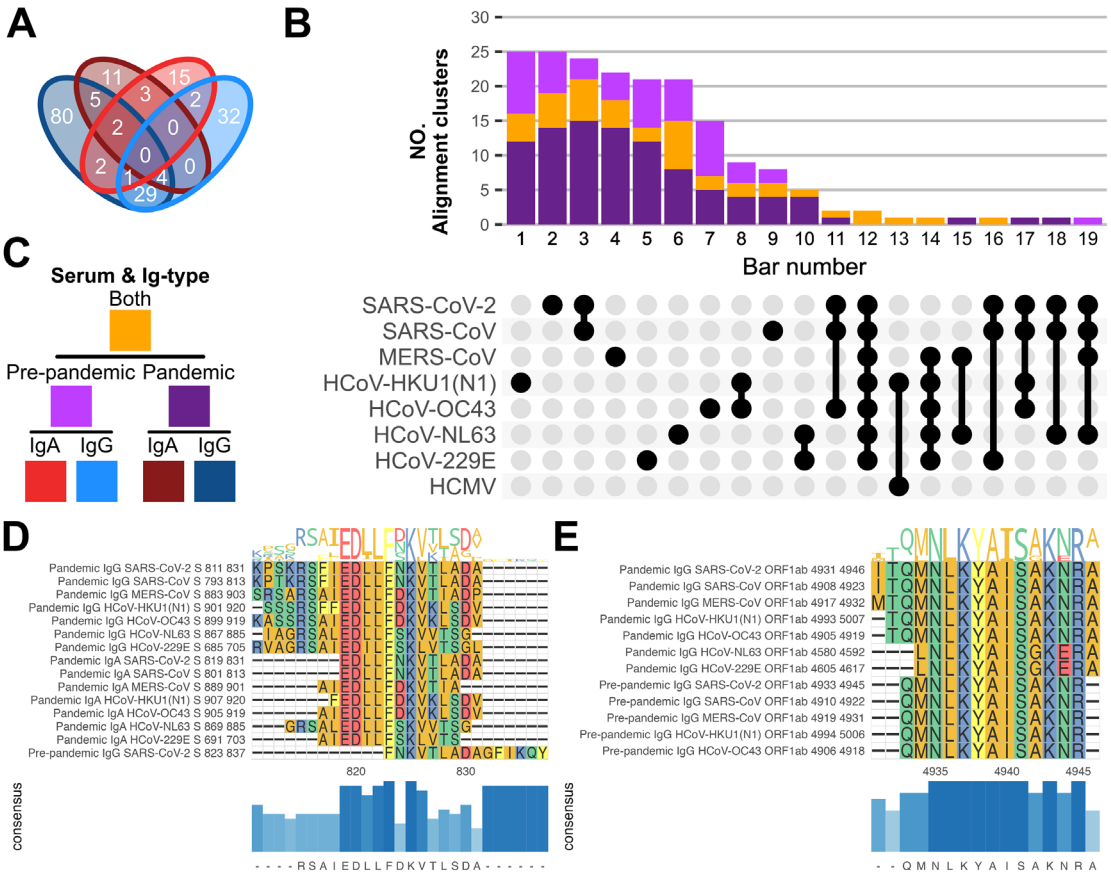

Figure 4. Summary of high-fidelity region alignment clusters. The 333 (280 unique) high-fidelity regions (HFRs) from both serum pools and Ig-types derived from the human coronaviruses (HCoVs) (Table 2) were clustered into 186 alignment clusters using single-linkage clustering with a sequence similarity cut-off at 60% (Table 4). **A**: Venn diagram showing how the 186 alignment clusters are distributed or shared between the serum pool and Ig-types. **B**: Stacked bar plot with counts of how many alignment clusters are specific to the pandemic serum, the pre-pandemic control serum, or if the alignment clusters have HFRs from both sera. The combination matrix below indicates which organisms were aligned. **C**: Is a pedigree chart legend indicating how colours relate to each other in **A** and **B**. **D**: Shows the aligned HFRs with alignment ID 135 (Table 5). The alignment covers positions 811 to 837 in the SARS-CoV-2 Spike (S) protein. **E**: Shows the aligned HFRs with alignment ID 87 (Table 5). The alignment covers positions 4931 to 4947 in the SARS-CoV-2 Open Reading Frame (ORF) 1ab protein. To the left of the aligned HFRs (**DE**) are their serum pools, Ig-types, organisms, parent proteins, and start/end positions. Above, sequence logos represent amino acid frequencies per position. Amino acids are color-coded by side chain chemistry. Below, a bar plot shows each position's most frequent amino acid (or gap). The inferred consensus sequence is at the bottom. HFR = high-fidelity region, HCoV = human coronavirus, HCMV = human cytomegalovirus, NO. = number of, S = Spike protein, ORF = Open Reading Frame.

CCC-derived HFRs, respectively. The presence of unexplained pre-pandemic antibody recognition of SARS-CoV, MERS-CoV, and SARS-CoV-2 derived epitopes suggests that further improvements of the present approach is warranted (see the discussion).

A more detailed analysis of HCoV cross-reactions is summarised in Figure 4C. Alignment clusters were simplified by merging IgG or IgA HFRs, as illustrated in the pedigree tree in Figure 4B. The 19 numbered bars represent the counts of alignment clusters with HFRs from different combinations of HCoVs (Figure 4B). Whether the members of these alignment clusters were recognised by the pre-pandemic or pandemic serum pools or by both is indicated by the segmental colouring of the bars. Alignment cluster bar numbers are cross-referenced in Table 5. By qualitative assessment of the bars (Figure 4B), four main groups were identified: (I) the epi- and pandemic HCoVs of the betacorona genus, SARS-CoV, MERS-CoV, and SARS-CoV-2, represented by bars 2, 3, 4, and 9, which cover the majority of alignment clusters and were predominantly recognised by pandemic serum pool; (II) the CCCs of the alphacorona genus (HCoV-NL63 and HCoV-229E), represented by bars 5, 6, and 10, which were more evenly distributed between the serum pools; (III) the CCCs of the betacorona genus, HCoV-HKU1(N1) and HCoV-OC43, represented by bars 1, 7, and 8, which were also more evenly recognised by the pre-pandemic and pandemic serum; and (IV) a pan-HCoV specific group represented by bar 12, which happens to consist of the two single largest alignment clusters (Table 5, Align ID 87 & 135), one of 15 members from the S protein (Figure 4D) and another of 12 members from the ORF1ab polyprotein (Figure 4E), which were shared between

the pre-pandemic and pandemic serum pools. Finally, eight smaller bars (Figure 4B, bar numbers 11, 13–19) represented alignment clusters with HFRs from various HCoVs across genera.

Figure 4D displays the largest alignment cluster (Align ID 135, Table 5). 14 of 15 HFRs were recognised by the pandemic serum pool (half by IgG, half by IgA); only one HFR was recognised by the pre-pandemic serum pool IgG, and since it was shifted C-terminally compared to all the other members of this alignment group, one could argue that it might not belong to the alignment. Thus, a consensus sequence consisting of 815-RSAIEDLL-822 is likely to be a dominant, pan-HCoV-specific linear B cell epitope footprint elicited by the SARS-CoV-2 infection.

Figure 4E, displays the second largest alignment cluster (Align ID 87, Table 3) located between positions 4931 and 4947 when mapped onto the SARS-CoV-2 ORF1ab protein. Five of the 12 HFRs were recognised by pre-pandemic serum pool IgG and the remaining seven were recognised by pandemic serum pool IgG. This linear B cell epitope cluster exemplifies a pan-HCoV specific, encompassing all seven HCoV strains.

Discussion

We analysed linear B cell epitopes from SARS-CoV-2 and 6 other pandemic or epidemic human coronavirus strains. We obtained pooled sera from COVID-19 convalescents (i.e. SARS-CoV-2 exposed) sampled early in the pandemic (around March 2020) and from pre-pandemic (not SARS-CoV-2 exposed) individuals sampled before October 2019 and tested them for relevant IgG and IgA responses. To detect linear B cell epitopes, high-density peptide microarrays were used to generate overlapping 13 amino acid peptides tiling by 2 amino acids to represent the entire proteomes of all seven HCoVs. These microarrays were incubated with pre-pandemic or pandemic serum pools, and IgG and IgA reactivity against specific peptides were detected. We identified 280 regions of linear B cell epitopes recognised by the pre-pandemic and/or pandemic serum pools. These regions could be homology clustered and aligned to the HCoV proteomes, providing evidence of cross-reactions between the pandemic and epidemic HCoVs. These clusters might be of therapeutic and diagnostic value. Predominantly recognized by the pandemic sera, the largest cluster was pan-HCoV specific and has, by others, been implicated in the neutralising activity of sera from COVID-19 convalescents[36–40]. Recognized by both pandemic and pre-pandemic sera, the second largest cluster was also pan-HCoV-specific, whereas other smaller clusters appeared to be strain or group specific; all of which would be of diagnostic interest.

Here, we have identified a large cohort of SARS-CoV-2 and HCoV-derived linear B cell epitopes and generated data on the IgG and IgA-mediated humoral responses towards these epitopes. At the protein level, we show that the humoral immune system primarily targets the ORF1ab protein, the Spike protein, and the nucleoprotein, with minor targets including ORF3a and ORF9b. Others have obtained similar results studying antibody responses against SARS-CoV-2 proteins after vaccination or natural infection. Currently (November 2023), 8,528 linear B-cell epitopes have been submitted to the Immune Epitope Database (IEDB)[41]. The majority of the methods used for linear antibody epitope discovery in SARS-CoV-2 include enzyme-linked immunosorbent assay (ELISA)[36–38,42–64], peptide arrays[65,66], peptide microarrays[67–81], phage display[82,83–91], or bacterial display[92]. The top three SARS-CoV-2 source proteins of linear B cell epitopes in the IEDB data are the ORF1ab (54%), S (30%), and N (9%) proteins[44].

When trying to understand the interactions between a virus and its host cells that lead to infection, the value of linear B cell epitope information may be questioned. In the case of the SARS-CoV-2 infection, the infective event involves structural rearrangements of the spike protein of the virus (reviewed in 93), and short synthetic peptides are not likely to mimic faithfully structural events involving larger, often discontinuous, viral structures. Nonetheless, linear information might be quite valuable. Poh *et al.* 2020 identified two peptides that successfully adsorbed neutralising activity from the convalescent sera[36]. One of them, S21P2, was contained in the largest cluster found in our study, Alignment ID 135 (SARS-CoV-2 spike protein 811–831, KPSKRSFIEDLLFNKVTLADA, underscoring indicating the rendition of this epitope reported by Poh *et al.* 2020) and the other, S14P5, was identified as a singleton in our study, Alignment ID 131 (SARS-CoV-2 spike protein 553–571, TESNKKFLPFQQFGRDIAD, underscoring indicating the rendition of this epitope reported by Poh *et al.* 2020). Low *et al.* 2022 isolated monoclonal antibodies specific for SARS-CoV-2 spike protein from convalescent donors[40]. Seven of 16 of these recognised a peptide that was also contained in our Alignment ID 135 (SARS-CoV-2 spike protein 811–831, KPSKRSFIEDLLFNKVTLADA; underscoring indicating the rendition of this epitope reported by Low *et al.* 2022). Two of these seven monoclonal antibodies showed pan-HCoV-specific neutralisation in vitro and protection in vivo using animal challenge studies. Linear B cell epitopes might also be of diagnostic value. Shrock *et al.* 2020 used a bacteriophage library approach ("VirScan"[94]) expressing 56-mer peptides tiling by 28 amino acids representing proteomes of all known pathogenic human viruses to examine the serological response against HCoVs and identify linear B cell epitopes[82]. Using this data, they developed a machine learning model that could accurately distinguish between COVID-19 positive and negative individuals.

Apart from the therapeutic and diagnostic implications described above, linear B cell epitope information might also contribute to explaining the natural course of COVID-19. Specifically, the potential cross-reactivity between the pandemic SARS-CoV-2 and endemic CCCs may indicate the presence of pre-existing immunity, offering a possible explanation for the highly variable outcome of the infection. To investigate the possibility of cross-reactions among linear B cell epitopes from different HCoV origins, we conducted a systematic comparison of all identified linear B cell epitopes and asked specifically whether epitope similarities could account for the observed humoral immune responses against SARS-CoV-1

and MERS, despite these viruses not having circulated in Denmark. Since many of the SARS-CoV-2 HFPs in Figure 3 overlapped, we initially merged the overlapping HFPs to reduce the HFPs into high-fidelity regions (HFRs) covering the linear B cell epitope footprints. While the average length of these HFRs was 18 amino acids, representing 3–4 merged HFPs, certain HFRs extended up to 157 amino acids, potentially encompassing multiple linear B cell epitopes simultaneously. We then used the Immune Epitope Database (IEDB) Epitope Cluster Analysis web tool (http://tools.iedb.org/cluster/)[34] to homology cluster HFRs based on a 60% sequence identity threshold. This threshold was selected to avoid false positive clustering of linear B cell epitopes from unrelated organisms; *in casu*, from HCMV. Using this threshold, we identified 78 homology clustered alignments between the HCoVs. Some of the HFRs from SARS-CoV and MERS-CoV occurred in clusters with CCCs and/or SARS-CoV-2, which could explain why they were recognised by the pre-pandemic and/or pandemic sera, respectively. Our data adds to the notion that pre-existing immunity to the endemic CCCs could cross-react with the epi- and pandemic HCoVs. This could lead to partial or complete protection depending on the unknown individual history of CCC exposure; something that for practical purposes is an unpredictable and confounding prognostic factor. Additionally, we observed many homology clusters between SARS-CoV and SARS-CoV-2 (Figure 4B, Bar number 3), which could explain SARS-CoV-1 responses in the pandemic serum pool. The latter potential for antibody cross-reactivity between SARS-CoV and SARS-CoV-2 is well described[58,66,69,70,74,82,83,92,95–99], and some of these antibodies have been shown to lead to cross-neutralization between these two closely related HCoVs[37,39,40,49,54,56,61,64,67,76,78,100–102]. When this is said, homology clustering did not account for all unexpected seroreactivities, as both serum pools retained SARS-CoV- and MERS-CoV-specific antibody responses. Additionally, the pre-pandemic serum exhibited SARS-CoV-2-specific reactivities that did not align with any of the CCCs. The unexplained exclusive HFRs may result from cross-reactive linear B cell epitope footprints below the homology threshold. The alphacorona HCoVs and MERS-CoV overall share less than 30% protein sequence identity in the S protein with the betacorona HCoVs[6]. Alternatively, cross-reactivity to past microorganism exposures not addressed in this study could be at play. For instance, the SARS-CoV-2 S protein S2 subunit (S686-1273) has been reported to harbour cross-reactive epitopes elicited by commensal microorganisms[103]. Finally, homology clustering can only suggest cross-reactivity between the different HCoV HFRs. Alanine scans or complete substitution analyses must confirm cross-reactivity to elucidate the linear B cell epitope footprint positions and amino acid specificities.

An interesting observation was a significant increase in the number of HCMV and Zaire Ebola virus responses in the pandemic vs pre-pandemic serum pools. A priori, one would not expect that an infection with SARS-CoV-2 would increase humoral responses to the unrelated HCMV and Ebola virus. Shrock *et al.* 2020[82] also reported increased seroreactivity against HCMV and herpes simplex virus 1 in non-hospitalized COVID-19 patients compared to hospitalised. We do not know the reason for this increased humoral reactivity in COVID-19 convalescents. Others have suggested that SARS-CoV-2 relaxes peripheral B cell tolerance[104] and/or causes polyclonal "by-stander" B cell activation[105]. Notably, Jiang *et al.* 2023 also reported significantly increased anti-HCMV IgG titers in their COVID-19 patient sera[105].

Finally, we would like to discuss the technological implementation of this technology and possible improvements. A major advantage of addressing linear B cell epitopes as opposed to conformational B cell epitopes lies in the exhaustive nature and relative ease of a linear peptide-driven analysis. Peptides, the embodiment of linear B cell epitopes, can be procured as synthetic high-density peptide microarray libraries featuring hundreds of thousands of addressable peptides, enough to encompass multiple virus proteomes. Which sequences should be used for effort can be determined from DNA sequencing the genomes of interest; for COVID-19, the first complete genome and its proteome derivative became available a few days into the pandemic. Once a peptide has been identified as a linear B cell epitope, the detailed specificity of the (even polyclonal) antibodies recognising this epitope can be virtually atomised using a complete single substitution analysis and, if need be, a multi-substitution analysis. We have demonstrated the power of a complete single substitution analysis (as opposed to a simpler alanine scanning analysis)[27–29] and proposed that an ANOVA analysis can be used to detail this specificity in objective, statistical terms (location and spacing of important residues that positively or negatively affects recognition, and identification of such residues). However, for a 15 (or 13-mer as in this analysis) peptide, this would require that 15 x 20 = 300 peptides (or 260 peptides for a 13-mer) be synthesised for each tiling peptide in an overlapping peptide library. We conducted a complete analysis of linear B cell epitopes in human serum albumin along these lines and used about 180,000 peptides to do this. It would be untenable to do this as a first choice for any proteome – even using high-density peptide microarrays as we did here. To solve this conundrum, we propose using neighbouring peptide signals to down-size this problem. In our past large-scale linear B cell epitope screens, where we characterised linear B cell epitopes using a complete single substitution analysis, we observed that most epitopes were sufficiently short to appear in at least two neighbouring, overlapping peptides[27,28]. Thus, we here introduced a criterium requiring peptides for further analysis to be found in at least one neighbouring peptide. We implemented this criterion by calculating the combined probability of both members of any pair belonging to the background population and required this to be ≤ 0.001. To prepare our data for substitution analysis, we further introduced another criterion that also had to be met, having an antibody binding signal of $\geq 20\%$ of the highest antibody binding signal recorded on the microarray. This ensures that any peptide selected for further study can be examined by substitution analysis, which requires that changes in the signals obtained with analogue peptides could clearly be distinguished from the proband peptide and, ideally, also from the background. With these combined requirements,

we aim to include most actual linear B cell epitopes in our analysis while avoiding those that a priori are unlikely to be such epitopes and/or be subject to substitution analysis. Here, we use this double requirement to condense individual peptide signals into a binary distinction: peptides with high-fidelity antibody binding (HFPs) meeting both requirements or non-binders. We want to add that these simple requirements would support a two-tiered linear B cell epitope discovery process where an initial scan using overlapping 13-mer peptides tiling by two amino acids could address entire virus proteomes and identify putative linear B cell epitopes for subsequent detailed substitution analysis (unfortunately, we did not have the resources to do the latter in this study). Such substitution analyses would identify peptide motifs consisting of a variable number of preferred (and/or disliked) amino acids in an appropriate arrangement and spacing. This information would provide a detailed and high-resolution characterisation of each recognised linear B cell epitope and could be used to improve future searches for potentially cross-reaction epitopes.

Another issue for future improvements is the source of the antibodies used in this analysis. COVID-19 spread rapidly in the early days of the pandemic, reaching thousands of cases by January 2020, tens of thousands by early February, and over a hundred thousand cases by early March 2020[106]. Thus, large numbers of relevant pandemic serums could easily have been obtained from convalescent donors. Likewise, relevant pre-pandemic serum could be sourced from suitable existing biobanks. We suggest that the initial scan always be conducted with pools of relevant serum. We have in a previous study, where individual sera were analysed, observed that the linear B cell repertoire is individualised to the extent that very few shared (or public) linear B cell epitopes exists[107]. Others pursuing similar large-scale, linear B cell epitope discovery using brute-force approaches of testing hundreds of individual sera[82] have demonstrated that very few public epitopes, that are recognised in 30–80% of individuals, exist. In contrast, many private epitopes, recognised in less than 5% of the population, exist[82,108]. Here, we found a mere 30% overlap in HCMV reactivity between the pandemic and pre-pandemic serum pools, which suggests some degree of individualisation of linear B cell responses in our study also (Table 4). Such epitope individualisation would seriously impact epitope discovery if identifying public epitopes is important. We interpret this as another indication of the individualised nature of linear B cell epitope recognition. The serum pools used here were only comprised of eight individuals. We suggest that the largest possible pool be used for the initial scan and any follow-up substitution analysis. Everything else being equal, this should focus the analysis on public epitope, which might be of more general utility. Eventually, when linear B cell epitopes of interest have been identified and characterised, then individual sera can be examined using single peptides (rather than entire peptide microarrays) to define the prevalence of linear B cell epitopes in the population.

Overall, the approach described here allows for a fast response to future pandemics and yields valuable data on linear B cell epitopes. Identifying virus-specific B cell epitopes is important for rational approaches to prevent, treat, diagnose, and understand virus infections.

Conclusion

We provide proof-of-concept for using serum pools on high-density peptide microarrays in linear B-cell epitope discovery from available genetic information. Even with a small pool of relevant pre-pandemic and pandemic sera, our findings correspond with literature regarding antibody reactivity against SARS-CoV-2 and other related human coronaviruses in sera of COVID-19 convalescents and pre-pandemic donors. Homology-clustering of our antibody targets revealed intriguing patterns of potential antibody cross-reactivity between the HCoVs, corroborated in the literature. These findings underscore the complexity of antibody responses and suggested solutions to address cross-reactivity patterns. The risk of future pandemics is substantial given our globalised lifestyle. We propose that the approach taken here, could be a valuable tool in the response to future pandemics allowing the discovery of targets for rational vaccine designs and serological testing.

Data availability

Dryad: Profiling of linear B-cell epitopes against human coronaviruses in pooled sera sampled early in the COVID-19 pandemic, https://doi.org/10.5061/dryad.s1rn8pkg7[109].

Data are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

Acknowledgement

A version of this manuscript was first uploaded to BioXiv: https://doi.org/10.1101/2024.02.29.582263

References

1. Johns Hopkins Center for Systems Science and Engineering (CSSE): **COVID-19 Map - Johns Hopkins Coronavirus Resource Center.** 2023; Accessed 10, 2023.
Reference Source

2. Chen L, Liu W, Zhang Q, *et al.*: **RNA based mNGS approach identifies a novel Human Coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak.** *Emerg Microbes Infect.* 2020; **9**(1): 313–319.
PubMed Abstract | Publisher Full Text | Free Full Text

3. Su S, Wong G, Shi W, *et al.*: **Epidemiology, genetic recombination, and pathogenesis of Coronaviruses.** *Trends Microbiol.* 2016; **24**(6): 490–502.
PubMed Abstract | Publisher Full Text | Free Full Text

4. Gussow AB, Auslander N, Faure G, *et al.*: **Genomic determinants of pathogenicity in SARS-CoV-2 and other Human Coronaviruses.** *Proc Natl Acad Sci U S A.* 2020; **117**(26): 15193–15199.
PubMed Abstract | Publisher Full Text | Free Full Text

5. Kesheh MM, Hosseini P, Soltani S, *et al.*: **An overview on the seven pathogenic Human Coronaviruses.** *Rev Med Virol.* 2022; **32**(2): e2282.
PubMed Abstract | Publisher Full Text

6. Cicaloni V, Costanti F, Pasqui A, *et al.*: **A bioinformatics approach to investigate structural and Non-Structural Proteins in Human Coronaviruses.** *Front Genet.* 2022; **13**: 891418.
PubMed Abstract | Publisher Full Text | Free Full Text

7. Chen B, Tian EK, He B, *et al.*: **Overview of lethal Human Coronaviruses.** *Signal Transduct Target Ther.* 2020; **5**(1): 89.
PubMed Abstract | Publisher Full Text | Free Full Text

8. World Health Organization: **Status of COVID-19 vaccines within WHO EUL/PQ evaluation process.** 2023; Accessed February 11, 2023.
Reference Source

9. World Health Organization: **COVID-19 vaccine tracker and landscape.** 2023; Accessed November 2, 2023.
Reference Source

10. Carabelli AM, Peacock TP, Thorne LG, *et al.*: **SARS-CoV-2 variant biology: immune escape, transmission and fitness.** *Nat Rev Microbiol.* 2023; **21**(3): 162–177.
PubMed Abstract | Publisher Full Text | Free Full Text

11. Yamasoba D, Kimura I, Nasser H, *et al.*: **Virological characteristics of the SARS-CoV-2 Omicron BA.2 spike.** *Cell.* 2022; **185**(12): 2103–2115.e19.
PubMed Abstract | Publisher Full Text | Free Full Text

12. Cao Y, Wang J, Jian F, *et al.*: **Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies.** *Nature.* 2022; **602**(7898): 657–663.
PubMed Abstract | Publisher Full Text | Free Full Text

13. Omotuyi O, Olubiyi O, Nash O, *et al.*: **SARS-CoV-2 Omicron spike glycoprotein receptor binding domain exhibits super-binder ability with ACE2 but not convalescent monoclonal antibody.** *Comput Biol Med.* 2022; **142**: 105226.
PubMed Abstract | Publisher Full Text | Free Full Text

14. Nextstrain team: **Genomic epidemiology of SARS-CoV-2 with subsampling focused globally over the past 6 months.** 2023; Accessed November 3, 2023.
Reference Source

15. Cao Y, Jian F, Wang J, *et al.*: **Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution.** *Nature.* 2023; **614**(7948): 521–529.
PubMed Abstract | Publisher Full Text | Free Full Text

16. Faraone JN, Qu P, Goodarzi N, *et al.*: **Immune evasion and membrane fusion of SARS-CoV-2 XBB subvariants EG.5.1 and XBB.2.3.** *bioRxiv.* 2023; 2023.08.30.555188.
Publisher Full Text

17. Wang Q, Guo Y, Zhang RM, *et al.*: **Antibody neutralisation of emerging SARS-CoV-2 subvariants: EG.5.1 and XBC.1.6.** *Lancet Infect Dis.* 2023; **23**(10): e397–e398.
PubMed Abstract | Publisher Full Text

18. Lasrado N, Collier ARY, Hachmann NP, *et al.*: **Neutralization escape by SARS-CoV-2 Omicron subvariant BA.2.86.** *bioRxiv.* 2023; 2023.09.04.556272.
Publisher Full Text

19. Hu Y, Zou J, Kurhade C, *et al.*: **Less neutralization evasion of SARS-CoV-2 BA.2.86 than XBB sublineages and CH.1.1.** *bioRxiv.* 2023; 2023.09.10.557047.
Publisher Full Text

20. Ferdous S, Kelm S, Baker TS, *et al.*: **B-cell epitopes: discontinuity and conformational analysis.** *Mol Immunol.* 2019; **114**: 643–650.
PubMed Abstract | Publisher Full Text

21. Sharon J, Rynkiewicz MJ, Lu Z, *et al.*: **Discovery of protective B-cell epitopes for development of antimicrobial vaccines and antibody therapeutics.** *Immunology.* 2014; **142**(1): 1–23.
PubMed Abstract | Publisher Full Text | Free Full Text

22. Haste Andersen P, Nielsen M, Lund O: **Prediction of residues in discontinuous B-cell epitopes using protein 3D structures.** *Protein Sci.* 2006; **15**(11): 2558–2567.
PubMed Abstract | Publisher Full Text | Free Full Text

23. Kringelum JV, Lundegaard C, Lund O, *et al.*: **Reliable B cell epitope predictions: impacts of method development and improved benchmarking.** *PLoS Comput Biol.* 2012; **8**(12): e1002829.
PubMed Abstract | Publisher Full Text | Free Full Text

24. Jespersen MC, Mahajan S, Peters B, *et al.*: **Antibody specific B-cell epitope predictions: leveraging information from antibody-antigen protein complexes.** *Front Immunol.* 2019; **10**: 298.
PubMed Abstract | Publisher Full Text | Free Full Text

25. Jespersen MC, Peters B, Nielsen M, *et al.*: **BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes.** *Nucleic Acids Res.* 2017; **45**(W1): W24–W29.
PubMed Abstract | Publisher Full Text | Free Full Text

26. Cia G, Pucci F, Rooman M: **Critical review of conformational B-cell epitope prediction methods.** *Brief Bioinform.* 2023; **24**(1): bbac567.
PubMed Abstract | Publisher Full Text

27. Buus S, Rockberg J, Forsstrom B, *et al.*: **High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays.** *Mol Cell Proteomics.* 2012; **11**(12): 1790–1800.
PubMed Abstract | Publisher Full Text | Free Full Text

28. Hansen LB, Buus S, Schafer-Nielsen C: **Identification and mapping of linear antibody epitopes in human serum albumin using high-density peptide arrays.** *PLoS One.* 2013; **8**(7): e68902.
PubMed Abstract | Publisher Full Text | Free Full Text

29. Hansen CS, Østerbye T, Marcatili P, *et al.*: **ArrayPitope: automated analysis of amino acid substitutions for peptide microarray-based antibody epitope mapping.** *PLoS One.* 2017; **12**(1): e0168453.
PubMed Abstract | Publisher Full Text | Free Full Text

30. Lachmann R, Loenenbach A, Waterboer T, *et al.*: **Cytomegalovirus (CMV) seroprevalence in the adult population of Germany.** *PLoS One.* 2018; **13**(7): e0200267.
PubMed Abstract | Publisher Full Text | Free Full Text

31. UniProt Consortium: **UniProt: the universal protein knowledgebase in 2021.** *Nucleic Acids Res.* 2021; **49**(D1): D480–D489.
PubMed Abstract | Publisher Full Text | Free Full Text

32. R Core Development Team: **A language and environment for statistical computing.** R Foundation for Statistical Computing; 2019.
Reference Source

33. Lawrence M, Huber W, Pagès H, *et al.*: **Software for computing and annotating genomic ranges.** *PLoS Comput Biol.* 2013; **9**(8): e1003118.
PubMed Abstract | Publisher Full Text | Free Full Text

34. Dhanda SK, Vaughan K, Schulten V, *et al.*: **Development of a novel clustering tool for linear peptide sequences.** *Immunology.* 2018; **155**(3): 331–345.
PubMed Abstract | Publisher Full Text | Free Full Text

35. Wright E: **DECIPHER: tools for curating, analyzing, and manipulating biological sequences**. 2021.
Publisher Full Text

36. Poh CM, Carissimo G, Wang B, *et al.*: **Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients.** *Nat Commun.* 2020; **11**(1): 2806.
PubMed Abstract | Publisher Full Text | Free Full Text

37. Vanderheijden N, Stevaert A, Xie J, *et al.*: **Functional analysis of human and feline coronavirus cross-reactive antibodies directed against the SARS-CoV-2 fusion peptide.** *Front Immunol.* 2021; **12**: 790415.
PubMed Abstract | Publisher Full Text | Free Full Text

38. Sun X, Yi C, Zhu Y, *et al.*: **Neutralization mechanism of a human antibody with pan-coronavirus reactivity including SARS-CoV-2.** *Nat Microbiol.* 2022; **7**(7): 1063–1074.
PubMed Abstract | Publisher Full Text

39. Dacon C, Tucker C, Peng L, *et al.*: **Broadly neutralizing antibodies target the coronavirus fusion peptide.** *Science.* 2022; **377**(6607): 728–735.
PubMed Abstract | Publisher Full Text | Free Full Text

40. Low JS, Jerak J, Tortorici MA, *et al.*: **ACE2-binding exposes the SARS-CoV-2 fusion peptide to broadly neutralizing coronavirus antibodies.** *Science.* 2022; **377**(6607): 735–742.
PubMed Abstract | Publisher Full Text | Free Full Text

41. Vita R, Mahajan S, Overton JA, *et al.*: **The Immune Epitope Database (IEDB): 2018 update.** *Nucleic Acids Res.* 2019; **47**(D1): D339–D343.
PubMed Abstract | Publisher Full Text | Free Full Text

42. Amrun SN, Lee CY, Lee B, *et al.*: **Linear B-cell epitopes in the spike and nucleocapsid proteins as markers of SARS-CoV-2 exposure and disease severity.** *EBioMedicine.* 2020; **58**: 102911.
PubMed Abstract | Publisher Full Text | Free Full Text

43. Simula ER, Manca MA, Jasemi S, *et al.*: **HCoV-NL63 and SARS-CoV-2 share recognized epitopes by the humoral response in sera of people collected pre- and during CoV-2 pandemic.** *Microorganisms.* 2020; **8**(12): 1993.
PubMed Abstract | Publisher Full Text | Free Full Text

44. Lu S, Xie XX, Zhao L, *et al.*: **The immunodominant and neutralization linear epitopes for SARS-CoV-2.** *Cell Rep.* 2021; **34**(4): 108666.
PubMed Abstract | Publisher Full Text | Free Full Text

45. Kang S, Yang M, He S, *et al.*: **A SARS-CoV-2 antibody curbs viral nucleocapsid protein-induced complement hyperactivation.** *Nat Commun.* 2021; **12**(1): 2697.
PubMed Abstract | Publisher Full Text | Free Full Text

46. Wang X, Lam JY, Chen L, *et al.*: **Mining of linear B cell epitopes of SARS-CoV-2 ORF8 protein from COVID-19 patients.** *Emerg Microbes Infect.* 2021; **10**(1): 1016–1023.
PubMed Abstract | Publisher Full Text | Free Full Text

47. Prakash S, Srivastava R, Coulon PG, *et al.*: **Genome-Wide B cell, CD4$^+$, and CD8$^+$ T cell epitopes that are highly conserved between human and animal Coronaviruses, identified from SARS-CoV-2 as targets for preemptive pan-Coronavirus vaccines.** *J Immunol.* 2021; **206**(11): 2566–2582.
PubMed Abstract | Publisher Full Text | Free Full Text

48. Zhang Y, Yang Z, Tian S, *et al.*: **A newly identified linear epitope on non-RBD region of SARS-CoV-2 spike protein improves the serological detection rate of COVID-19 patients.** *BMC Microbiol.* 2021; **21**(1): 194.
PubMed Abstract | Publisher Full Text | Free Full Text

49. Pinto D, Sauer MM, Czudnochowski N, *et al.*: **Broad betacoronavirus neutralization by a stem helix-specific human antibody.** *Science.* 2021; **373**(6559): 1109–1116.
PubMed Abstract | Publisher Full Text | Free Full Text

50. Patarroyo ME, Patarroyo MA, Alba MP, *et al.*: **The first chemically-synthesised, highly immunogenic Anti-SARS-CoV-2 peptides in DNA genotyped aotus monkeys for human use.** *Front Immunol.* 2021; **12**: 724060.
PubMed Abstract | Publisher Full Text | Free Full Text

51. Polvere I, Voccola S, Parrella A, *et al.*: **A peptide-based assay discriminates individual antibody response to the COVID-19 Pfizer/BioNTech mRNA vaccine.** *Vaccines (Basel).* 2021; **9**(9): 987.
PubMed Abstract | Publisher Full Text | Free Full Text

52. Polyiam K, Phoolcharoen W, Butkhot N, *et al.*: **Immunodominant linear B cell epitopes in the spike and membrane proteins of SARS-CoV-2 identified by immunoinformatics prediction and immunoassay.** *Sci Rep.* 2021; **11**(1): 20383.
PubMed Abstract | Publisher Full Text | Free Full Text

53. Gao F, Huang J, Li T, *et al.*: **A highly conserved peptide vaccine candidate activates both humoral and cellular immunity against SARS-CoV-2 variant strains.** *Front Immunol.* 2021; **12**: 789905.
PubMed Abstract | Publisher Full Text | Free Full Text

54. Li W, Chen Y, Prevost J, *et al.*: **Structural basis and mode of action for two broadly neutralizing antibodies against SARS-CoV-2 emerging variants of concern.** *Cell Rep.* 2022; **38**(2): 110210.
PubMed Abstract | Publisher Full Text | Free Full Text

55. Maghsood F, Shokri MR, Jeddi-Tehrani M, *et al.*: **Identification of immunodominant epitopes on nucleocapsid and spike proteins of the SARS-CoV-2 in Iranian COVID-19 patients.** *Pathog Dis.* 2022; **80**(1): ftac001.
PubMed Abstract | Publisher Full Text | Free Full Text

56. Zhou P, Yuan M, Song G, *et al.*: **A human antibody reveals a conserved site on beta-coronavirus spike proteins and confers protection against SARS-CoV-2 infection.** *Sci Transl Med.* 2022; **14**(637): eabi9215.
PubMed Abstract | Publisher Full Text | Free Full Text

57. Herrscher C, Eymieux S, Gaborit C, *et al.*: **ELISA-Based analysis reveals an Anti-SARS-CoV-2 protein immune response profile associated with disease severity.** *J Clin Med.* 2022; **11**(2): 405.
PubMed Abstract | Publisher Full Text | Free Full Text

58. Yu J, Qin Z, Liu X, *et al.*: **High-specificity targets in SARS-CoV-2 N protein for serological detection and distinction from SARS-CoV.** *Comput Biol Med.* 2022; **143**: 105272.
PubMed Abstract | Publisher Full Text | Free Full Text

59. Yang H, Cao J, Lin X, *et al.*: **Developing an effective peptide-based vaccine for COVID-19: preliminary studies in mice models.** *Viruses.* 2022; **14**(3): 449.
PubMed Abstract | Publisher Full Text | Free Full Text

60. Tajuelo A, Lopez-Siles M, Mas V, *et al.*: **Cross-Recognition of SARS-CoV-2 B-Cell epitopes with other betacoronavirus nucleoproteins.** *Int J Mol Sci.* 2022; **23**(6): 2977.
PubMed Abstract | Publisher Full Text | Free Full Text

61. Geanes ES, LeMaster C, Fraley ER, *et al.*: **Cross-reactive antibodies elicited to conserved epitopes on SARS-CoV-2 spike protein after infection and vaccination.** *Sci Rep.* 2022; **12**(1): 6496.
PubMed Abstract | Publisher Full Text | Free Full Text

62. Garanina E, Hamza S, Stott-Marshall RJ, *et al.*: **Antibody and T cell immune responses to SARS-CoV-2 peptides in COVID-19 convalescent patients.** *Front Microbiol.* 2022; **13**: 842232.
PubMed Abstract | Publisher Full Text | Free Full Text

63. Pratesi F, Errante F, Pacini L, *et al.*: **A SARS-CoV-2 spike receptor binding motif peptide induces Anti-Spike antibodies in mice andIs recognized by COVID-19 patients.** *Front Immunol.* 2022; **13**: 879946.
PubMed Abstract | Publisher Full Text | Free Full Text

64. Garrido JL, Medina MA, Bravo F, *et al.*: **IgG targeting distinct seasonal coronavirus-conserved SARS-CoV-2 spike subdomains correlates with differential COVID-19 disease outcomes.** *Cell Rep.* 2022; **39**(9): 110904.
PubMed Abstract | Publisher Full Text | Free Full Text

65. Yi Z, Ling Y, Zhang X, *et al.*: **Functional mapping of B-cell linear epitopes of SARS-CoV-2 in COVID-19 convalescent population.** *Emerg Microbes Infect.* 2020; **9**(1): 1988–1996.
PubMed Abstract | Publisher Full Text | Free Full Text

66. Gregory DJ, Vannier A, Duey AH, *et al.*: **Repertoires of SARS-CoV-2 epitopes targeted by antibodies vary according to severity of COVID-19.** *Virulence.* 2022; **13**(1): 890–902.
PubMed Abstract | Publisher Full Text | Free Full Text

67. Ng KW, Faulkner N, Cornish GH, *et al.*: **Preexisting and de novo humoral immunity to SARS-CoV-2 in humans.** *Science.* 2020; **370**(6522): 1339–1343.
PubMed Abstract | Publisher Full Text | Free Full Text

68. Musico A, Frigerio R, Mussida A, *et al.*: **SARS-CoV-2 epitope mapping on microarrays highlights strong immune-response to N protein region.** *Vaccines (Basel).* 2021; **9**(1): 35.
PubMed Abstract | Publisher Full Text | Free Full Text

69. Ladner JT, Henson SN, Boyle AS, *et al.*: **Epitope-resolved profiling of the SARS-CoV-2 antibody response identifies cross-reactivity with endemic human coronaviruses.** *Cell Rep Med.* 2021; **2**(1): 100189.
PubMed Abstract | Publisher Full Text | Free Full Text

70. Holenya P, Lange PJ, Reimer U, *et al.*: **Peptide microarray-based analysis of antibody responses to SARS-CoV-2 identifies unique epitopes with potential for diagnostic test development.** *Eur J Immunol.* 2021; **51**(7): 1839–1849.
PubMed Abstract | Publisher Full Text | Free Full Text

71. Voss C, Esmail S, Liu X, *et al.*: **Epitope-specific antibody responses differentiate COVID-19 outcomes and variants of concern.** *JCI Insight.* 2021; **6**(13): e148855.
PubMed Abstract | Publisher Full Text | Free Full Text

72. Mishra N, Huang X, Joshi S, *et al.*: **Immunoreactive peptide maps of SARS-CoV-2.** *Commun Biol.* 2021; **4**(1): 225.
PubMed Abstract | Publisher Full Text | Free Full Text

73. Schwarz T, Heiss K, Mahendran Y, *et al.*: **SARS-CoV-2 proteome-wide analysis revealed significant epitope signatures in COVID-19 patients.** *Front Immunol.* 2021; **12**: 629185.
PubMed Abstract | Publisher Full Text | Free Full Text

74. Heffron AS, McIlwain SJ, Amjadi MF, *et al.*: **The landscape of antibody binding in SARS-CoV-2 infection.** *PLoS Biol.* 2021; **19**(6): e3001265.
PubMed Abstract | Publisher Full Text | Free Full Text

75. Nitahara Y, Nakagama Y, Kaku N, *et al.*: **High-Resolution linear epitope mapping of the Receptor Binding Domain of SARS-CoV-2 spike protein in COVID-19 mRNA vaccine recipients.** *Microbiol Spectr.* 2021; **9**(3): e0096521.
PubMed Abstract | Publisher Full Text | Free Full Text

76. Camerini D, Randall AZ, Trappl-Kimmons K, *et al.*: **Mapping SARS-CoV-2 antibody epitopes in COVID-19 patients with a multi-coronavirus protein microarray.** *Microbiol Spectr.* 2021; **9**(2): e0141621.
PubMed Abstract | Publisher Full Text | Free Full Text

77. Gattinger P, Niespodziana K, Stiasny K, *et al.*: **Neutralization of SARS-CoV-2 requires antibodies against conformational receptor-binding domain epitopes.** *Allergy.* 2022; **77**(1): 230–242.
PubMed Abstract | Publisher Full Text | Free Full Text

78. Li M, Liu J, Lu R, *et al.*: **Longitudinal immune profiling reveals dominant epitopes mediating long-term humoral immunity in COVID-19-convalescent individuals.** *J Allergy Clin Immunol.* 2022; **149**(4): 1225–1241.
PubMed Abstract | Publisher Full Text | Free Full Text

79. Levy Y, Alcalay R, Zvi A, *et al.*: **Immunodominant linear B-cell epitopes of SARS-CoV-2 spike, identified by sera from K18-hACE2 mice infected with the WT or variant viruses.** *Vaccines (Basel).* 2022; **10**(2): 251.
PubMed Abstract | Publisher Full Text | Free Full Text

80. Chen L, Pang P, Qi H, *et al.*: **Evaluation of spike protein epitopes by assessing the dynamics of humoral immune responses in moderate COVID-19.** *Front Immunol.* 2022; **13**: 770982.
PubMed Abstract | Publisher Full Text | Free Full Text

81. Hotop SK, Reimering S, Shekhar A, *et al.*: **Peptide microarrays coupled to machine learning reveal individual epitopes from human antibody responses with neutralizing capabilities against SARS-CoV-2.** *Emerg Microbes Infect.* 2022; **11**(1): 1037–1048.
PubMed Abstract | Publisher Full Text | Free Full Text

82. Shrock E, Fujimura E, Kula T, *et al.*: **Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity.** *Science.* 2020; **370**(6520): eabd4250.
PubMed Abstract | Publisher Full Text | Free Full Text

83. Stoddard CI, Galloway J, Chu HY, *et al.*: **Epitope profiling reveals binding signatures of SARS-CoV-2 immune response in natural infection and cross-reactivity with endemic human CoVs.** *Cell Rep.* 2021; **35**(8): 109164.
PubMed Abstract | Publisher Full Text | Free Full Text

84. Guo JY, Liu IJ, Lin HT, *et al.*: **Identification of COVID-19 B-cell epitopes with phage-displayed peptide library.** *J Biomed Sci.* 2021; **28**(1): 43.
PubMed Abstract | Publisher Full Text | Free Full Text

85. Tang J, Lee Y, Ravichandran S, *et al.*: **Epitope diversity of SARS-CoV-2 hyperimmune intravenous human immunoglobulins and neutralization of variants of concern.** *iScience.* 2021; **24**(9): 103006.
PubMed Abstract | Publisher Full Text | Free Full Text

86. Li L, Gao M, Li J, *et al.*: **Methods to identify immunogenic peptides in SARS-CoV-2 spike and protective monoclonal antibodies in COVID-19 patients.** *Small Methods.* 2021; **5**(7): 2100058.
PubMed Abstract | Publisher Full Text | Free Full Text

87. Ravichandran S, Tang J, Grubbs G, *et al.*: **SARS-CoV-2 immune repertoire in MIS-C and pediatric COVID-19.** *Nat Immunol.* 2021; **22**(11): 1452–1464.
PubMed Abstract | Publisher Full Text

88. Garrett ME, Galloway JG, Wolf C, *et al.*: **Comprehensive characterization of the antibody responses to SARS-CoV-2 Spike protein finds additional vaccine-induced epitopes beyond those for mild infection.** *eLife.* 2022; **11**: e73490.
PubMed Abstract | Publisher Full Text | Free Full Text

89. Kumar G, Sterrett S, Hall L, *et al.*: **Comprehensive mapping of SARS-CoV-2 peptide epitopes for development of a highly sensitive serological test for total and neutralizing antibodies.** *Protein Eng Des Sel.* 2022; **35**: gzab033.
PubMed Abstract | Publisher Full Text | Free Full Text

90. Willcox AC, Sung K, Garrett ME, *et al.*: **Detailed analysis of antibody responses to SARS-CoV-2 vaccination and infection in macaques.** *PLoS Pathog.* 2022; **18**(4): e1010155.
PubMed Abstract | Publisher Full Text | Free Full Text

91. Ballmann R, Hotop SK, Bertoglio F, *et al.*: **ORFeome phage display reveals a major immunogenic epitope on the S2 subdomain of SARS-CoV-2 spike protein.** *Viruses.* 2022; **14**(6): 1326.
PubMed Abstract | Publisher Full Text | Free Full Text

92. Haynes WA, Kamath K, Bozekowski J, *et al.*: **High-resolution epitope mapping and characterization of SARS-CoV-2 antibodies in large cohorts of subjects with COVID-19.** *Commun Biol.* 2021; **4**(1): 1317.
PubMed Abstract | Publisher Full Text | Free Full Text

93. Le K, Kannappan S, Kim T, *et al.*: **Structural understanding of SARS-CoV-2 virus entry to host cells.** *Front Mol Biosci.* 2023; **10**: 1288686.
PubMed Abstract | Publisher Full Text | Free Full Text

94. Xu GJ, Kula T, Xu Q, *et al.*: **Viral immunology. Comprehensive serological profiling of human populations using a synthetic human virome.** *Science.*

2015; **348**(6239): aaa0698.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

95. Zamecnik CR, Rajan JV, Yamauchi KA, *et al.*: **ReScan, a multiplex diagnostic pipeline, pans human sera for SARS-CoV-2 antigens.** *Cell Rep Med.* 2020; **1**(7): 100123.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

96. Wang H, Wu X, Zhang X, *et al.*: **SARS-CoV-2 proteome microarray for mapping COVID-19 antibody interactions at amino acid resolution.** *ACS Cent Sci.* 2020; **6**(12): 2238–2249.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

97. Jones BE, Brown-Augsburger PL, Corbett KS, *et al.*: **The neutralizing antibody, LY-CoV555, protects against SARS-CoV-2 infection in nonhuman primates.** *Sci Transl Med.* 2021; **13**(593): eabf1906.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

98. Sen SR, Sanders EC, Gabriel KN, *et al.*: **Predicting COVID-19 severity with a specific nucleocapsid antibody plus disease risk factor score.** *mSphere.* 2021; **6**(2): e00203–21.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

99. Woudenberg T, Pelleau S, Anna F, *et al.*: **Humoral immunity to SARS-CoV-2 and seasonal coronaviruses in children and adults in north-eastern France.** *EBioMedicine.* 2021; **70**: 103495.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

100. Lv H, Wu NC, Tsang OTY, *et al.*: **Cross-reactive antibody response between SARS-CoV-2 and SARS-CoV infections.** *Cell Rep.* 2020; **31**(9): 107725.
PubMed Abstract | **Publisher Full Text**

101. Song G, He WT, Callaghan S, *et al.*: **Cross-reactive serum and memory B-cell responses to spike protein in SARS-CoV-2 and endemic coronavirus infection.** *Nat Commun.* 2021; **12**(1): 2938.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

102. Dacon C, Peng L, Lin TH, *et al.*: **Rare, convergent antibodies targeting the stem helix broadly neutralize diverse betacoronaviruses.** *Cell Host Microbe.* 2022; **31**(1): 97–111.e12.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

103. Jia L, Weng S, Wu J, *et al.*: **Preexisting antibodies targeting SARS-CoV-2 S2 cross-react with commensal gut bacteria and impact COVID-19 vaccine induced immunity.** *Gut Microbes.* 2022; **14**(1): 2117503.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

104. Castleman MJ, Stumpf MM, Therrien NR, *et al.*: **SARS-CoV-2 infection relaxes peripheral B cell tolerance.** *J Exp Med.* 2022; **219**(6): e20212553.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

105. Jiang W, Johnson D, Adekunle R, *et al.*: **COVID-19 is associated with bystander polyclonal autoreactive B cell activation as reflected by a broad autoantibody production, but none is linked to disease severity.** *J Med Virol.* 2023; **95**(1): e28134.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

106. Center for Systems Science and Engineering (CSSE): **COVID-19 data repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University.** 2020; Accessed February 20, 2024.
Reference Source

107. Bach E, Fitzgerald SF, Williams-MacDonald SE, *et al.*: **Genome-wide epitope mapping across multiple host species reveals significant diversity in antibody responses to *Coxiella burnetii* vaccination and infection.** *Front Immunol.* 2023; **14**: 1257722.
PubMed Abstract | **Publisher Full Text** | **Free Full Text**

108. Lo KC, Sullivan E, Bannen RM, *et al.*: **Comprehensive profiling of the rheumatoid arthritis antibody repertoire.** *Arthritis Rheumatol.* 2020; **72**(2): 242–250.
PubMed Abstract | **Publisher Full Text**

109. Bach E, Ghanizada M, Kirkby N, *et al.*: **Profiling of linear B-cell epitopes against human coronaviruses in pooled sera sampled early in the COVID-19 pandemic.** [Dataset]. Dryad. 2024.
http://www.doi.org/10.5061/dryad.s1rn8pkg7

Open Peer Review

Current Peer Review Status: ?

Version 1

Reviewer Report 09 December 2024

https://doi.org/10.21956/openreseurope.19039.r45851

© **2024 Scherer E.** This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

?

Erin M Scherer
Emory University School of Medicine, Atlanta, Georgia, USA

This manuscript by Bach et al examine linear antibody epitopes against the seven human coronaviruses using peptide microarrays. They include scrambled peptides and ebolavirus peptides as negative controls and CMV peptides as a positive control. They use 8 pooled convalescent sera from early in the pandemic and 8 pooled pre-pandemic healthy sera to map epitopes. They find:

- the lowest IgG and IgA reactivity to the negative control peptides,
- no significant reactivity detected against the common cold coronavirus peptides,
- the highest pre-pandemic IgG reactivity to CMV peptides,
- the highest convalescent IgG and IgA reactivity to SARS-CoV-2 peptides,
- the next highest convalescent IgG reactivity to CMV peptides,
- and some cross-reactivity to SARS-CoV peptides among convalescent IgG and IgA.

The authors map the regions in the SARS-CoV-2 proteome that these reactivities correspond to. Most epitopes aligned in nucleocapsid (both in terms of number and density (number relative to protein length), but notable reactivities were seen in ORF1ab, spike, ORF3a and membrane protein. They dissect and present which epitopes correspond to pre-, post- or both pre- and post-pandemic IgG/IgA reactivities. They do not assess these antibodies for neutralizing capacity, e.g., by enriching peptide-binding polyclonal IgG from the sera to see if these neutralize. They also do not assess reactivities to SARS-CoV-2 variant peptides. No concerns about execution of experiments or analysis of experimental results. More concerns about approach. Well written for the most part – some concerns with introductory material.

Major comments

- Eight human specimens is very small group size and doesn't even meet the minimum threshold recommended for a pilot study (Ref 1.)
- Pooling serum prevents understanding of individual response variation and reduces chance of identifying public epitopes for diagnostic purposes or really any purpose related to public health (e.g., identification of protective epitopes). It is only through testing of individual sera

and identification of public epitopes among them that one can identify that an epitope is immunodominant across multiple persons, not just a strong response from one person in the pool. Moreover, given that the convalescent pooled sera were only tested at a 1:100 dilution in the current study, the probability of identifying peptide binding from potent antibody responses of individual persons in the pool increases.
- Peptide tables (e.g., Tables 5, 6) really belong in supplementary

Minor comments
- Need citation for: “In contrast, the epidemic SARS-CoV (emerged in 2002) and MERS-CoV (emerged in 2012) exhibited high mortality rates of approximately 9% and 36%, respectively. “
- I think the citation for this statement “Comparing S proteins, SARS-CoV-2 has the highest similarity with SARS-CoV, with a nucleotide sequence identity of 80.4% and protein sequence identity of 77.3%" needs to be checked. 80.04% refers to nucleotide homology of SARS-CoV-2 and SARS-CoV genomes in paper referenced and 76% - as found in other papers - refers to homology of SARS-CoV-2 and SARS-Cov S protein amino acid sequences.
- This statement: “Most proteomic mutations have occurred in the ORF1ab protein, followed by the S protein and then the N protein11–13” are not supported by the citations 11-13, and are not supported by other citations, e.g.,: Gayvert, K., McKay, S., Lim, W.K. *et al.* Evolutionary trajectory of SARS-CoV-2 genome shifts during widespread vaccination and emergence of Omicron variant. *npj Viruses* **1**, 5 (2023). Please revisit.
- The statement: “However, the XBB variant is currently in the process of being outcompeted by EG.5, a new descendant lineage of XBB.19.2.” Is XBB.19.2 correct? XBB.19.2 (or XBB.1.9.2) is not shown as a clade on Nextstrain. XBB.1.9/23D and EG.5.1/23F are shown on Nextstrain. Did you mean one of these? Or did you mean EG.5?
- What isolate of SARS-CoV-2 used for peptides? After the lengthy discussion of different clades in the introduction, this isn’t mentioned. Assuming it is the reference isolate?

References

1. Julious S: Sample size of 12 per group rule of thumb for a pilot study. *Pharmaceutical Statistics*. 2005; **4** (4): 287-291 Publisher Full Text
2. Gayvert K, McKay S, Lim W, Baum A, et al.: Evolutionary trajectory of SARS-CoV-2 genome shifts during widespread vaccination and emergence of Omicron variant. *npj Viruses*. 2023; **1** (1). Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and does the work have academic merit?
Partly

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
I cannot comment. A qualified statistician is required.

Are all the source data underlying the results available to ensure full reproducibility?
No

Are the conclusions drawn adequately supported by the results?
Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Antibody/B cell immunity; SARS-CoV-2, SARS-CoV-2 vaccines

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.