Check for updates

# Deep Sequencing of B Cell Receptor Repertoires From COVID-19 Patients Reveals Strong Convergent Immune Signatures

Jacob D. Galson[1*], Sebastian Schaetzle[1], Rachael J. M. Bashford-Rogers[1,2], Matthew I. J. Raybould[3], Aleksandr Kovaltsuk[3], Gavin J. Kilpatrick[1], Ralph Minter[1], Donna K. Finch[1], Jorge Dias[1], Louisa K. James[4], Gavin Thomas[4], Wing-Yiu Jason Lee[4], Jason Betley[5], Olivia Cavlan[1], Alex Leech[1], Charlotte M. Deane[3], Joan Seoane[6], Carlos Caldas[7], Daniel J. Pennington[4], Paul Pfeffer[4] and Jane Osbourn[1]

[1] Alchemab Therapeutics Ltd, London, United Kingdom, [2] Wellcome Centre for Human Genetics, Oxford, United Kingdom, [3] Oxford Protein Informatics Group, Department of Statistics, University of Oxford, Oxford, United Kingdom, [4] Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom, [5] Illumina, Inc., Illumina Centre, Cambridge, United Kingdom, [6] Translational Research Program, Vall d'Hebron Institute of Oncology, Barcelona, Spain, [7] Cancer Research UK Cambridge Institute and Department of Oncology, Li Ka Shing Centre, University of Cambridge, Cambridge, United Kingdom

Deep sequencing of B cell receptor (BCR) heavy chains from a cohort of 31 COVID-19 patients from the UK reveals a stereotypical naive immune response to SARS-CoV-2 which is consistent across patients. Clonal expansion of the B cell population is also observed and may be the result of memory bystander effects. There was a strong convergent sequence signature across patients, and we identified 1,254 clonotypes convergent between at least four of the COVID-19 patients, but not present in healthy controls or individuals following seasonal influenza vaccination. A subset of the convergent clonotypes were homologous to known SARS and SARS-CoV-2 spike protein neutralizing antibodies. Convergence was also demonstrated across wide geographies by comparison of data sets between patients from UK, USA, and China, further validating the disease association and consistency of the stereotypical immune response even at the sequence level. These convergent clonotypes provide a resource to identify potential therapeutic and prophylactic antibodies and demonstrate the potential of BCR profiling as a tool to help understand patient responses.

Keywords: COVID-19, SARS-CoV-2, B-cell repertoire, BCR, antibody, convergence

## INTRODUCTION

Since the report of the first patients in December 2019 (1, 2), the unprecedented global scale of the COVID-19 pandemic has become apparent. The infectious agent, the SARS-CoV-2 betacoronavirus (3), causes mild symptoms in most cases but can cause severe respiratory diseases such as acute respiratory distress syndrome in some individuals. Risk factors for severe disease include age, male gender, and underlying co-morbidities (4).

Understanding the immune response to SARS-CoV-2 infection is critical to support the development of therapies. Recombinant monoclonal antibodies derived from analyses of B cell receptor (BCR) repertoires in infected patients or the immunization of animals have been shown to be effective against several infectious diseases including Ebola virus (5), rabies (6), and respiratory syncytial virus disease (7). Such therapeutic antibodies have the potential to protect susceptible populations as well as to treat severe established infections.

While many vaccine approaches are underway in response to the SARS-CoV-2 outbreak, many of these compositions include as immunogens either whole, attenuated virus or whole spike (S) protein—a viral membrane glycoprotein which mediates cell uptake by binding to host angiotensin-converting enzyme 2 (ACE2). The antibody response to such vaccines will be polyclonal in nature and will likely include both neutralizing and non-neutralizing antibodies. It is hoped that the neutralizing component will be sufficient to provide long-term SARS-CoV-2 immunity following vaccination, although other potential confounders may exist, such as raising antibodies which mediate antibody-dependent enhancement (ADE) of viral entry (8–10). While ADE is not proven for SARS-CoV-2, prior studies of SARS-CoV-1 in non-human primates showed that, while some S protein antibodies from human SARS-CoV-1 patients were protective, others enhanced the infection via ADE (11). An alternative could be to support passive immunity to SARS-CoV-2, by administering one, or a small cocktail of, well-characterized, neutralizing antibodies.

Patients recovering from COVID-19 have already been screened to identify neutralizing antibodies, following analysis of relatively small numbers (100–500) of antibody sequences (12–14). A more extensive BCR repertoire analysis was performed on six patients in Stanford, USA with signs and symptoms of COVID-19 who also tested positive for SARS-CoV-2 RNA (15). Although no information was provided on the patient outcomes in that study, the analysis demonstrated preferential expression of a subset of immunoglobulin heavy chain (IGH) V gene segments with relatively little somatic hypermutation and showed evidence of convergent antibodies between patients.

To drive a deeper understanding of the nature of humoral immunity to SARS-CoV-2 infection and to identify potential therapeutic antibodies to SARS-CoV-2, we have evaluated the BCR heavy chain repertoire from 31 individuals at various stages of their immune response. We show that there are stereotypic responses to SARS-CoV-2 infection, that infection stimulates both naïve and memory B cell responses, that sequence convergence can be used to identify putative SARS-CoV-2 specific antibodies, and that sequence convergence can be identified between different SARS-CoV-2 studies in different locations and using different sample types.

## METHODS

### Clinical Information Gathering

Peripheral blood was obtained from patients admitted with acute COVID-19 pneumonia to medical wards at Barts Health NHS Trust, London, UK, after informed consent by the direct care team (NHS HRA RES Ethics 19/SC/0361). Venous blood was collected in 20-ml EDTA Vacutainers (BD). Patient demographics and clinical information relevant to their admission were collected by members of the direct care team, including duration of symptoms prior to blood sample collection. Current severity was mapped to the WHO Ordinal Scale of Severity. Whether patients at time of sample collection were clinically Improving, Stable or Deteriorating was subjectively determined by the direct clinical team prior to any sample analysis. This determination was primarily made on the basis of whether requirement for supplemental oxygen was increasing, stable, or decreasing comparing current day to previous 3 days. Lymphocyte counts were determined using a standard clinical cytometer.

### Sample Collection and Initial Processing

Blood samples were processed within 1 h of collection in order to limit RNA degradation. Blood was first centrifuged at 150 x$g$ for 15 min at room temperature to separate plasma. The cell pellet was then resuspended with phosphate-buffered saline (PBS without calcium and magnesium, Sigma) to 20 ml, layered onto 15-ml Ficoll-Paque Plus (GE Healthcare) and then centrifuged at 400 x$g$ for 30 min at room temperature without brake. Mononuclear cells (PBMCs) were extracted from the buffy coat and washed twice with PBS at 300 x$g$ for 8 min. PBMCs were counted with Trypan blue (Sigma) and viability of >96% was observed. PBMCs ($5 \times 10^6$) were immediately resuspended in RLT buffer to stabilize the RNA (Qiagen) and incubated at room temperature for 10 min prior to storage at –80°C. Consecutive donor samples with sufficient RLT samples progressed to RNA preparation and BCR preparation and are included in this manuscript.

Metastatic breast cancer biopsy samples were collected and RNA extracted as part of a previously reported cohort (16).

### RNA Prep and BCR Sequencing

Total RNA from $5 \times 10^6$ PBMCs was isolated using RNeasy kits (Qiagen). First-strand cDNA was generated from total RNA using SuperScript RT IV (Invitrogen) and IgA and IgG isotype specific primers (17) including UMIs at 50°C for 45 min (inactivation at 80°C for 10 min).

The resulting cDNA was used as template for High Fidelity PCR amplification (KAPA, Roche) using a set of 6 FR1-specific forward primers (17) including sample-specific barcode sequences (6 bp) and a reverse primer specific to the RT primer (initial denaturation at 95°C for 3 min, 25 cycles at 98°C for 20 s, 60°C for 30 s, 72°C for 1 min and final extension at 72°C for 7 min). The amount of BCR heavy chain amplicons (~450 bp) was quantified by TapeStation (Beckman Coulter) and gel-purified.

Dual-indexed sequencing adapters (KAPA) were ligated onto 500-ng amplicons per patient using the HyperPrep library construction kit (KAPA) and the adapter-ligated libraries were finally PCR-amplified for 3 cycles (98°C for 15 s, 60°C for 30 s, 72°C for 30s, final extension at 72°C for 1 min). Pools of 10, 9, and 12 libraries were sequenced across three runs on an Illumina MiSeq using 2 × 300 bp chemistry.

## Sequence Processing

The Immcantation framework (docker container v3.0.0) was used for sequence processing (18, 19). Briefly, paired-end reads were joined based on a minimum overlap of 20 nt, and a max error of 0.2, and reads with a mean phred score below 20 were removed. Primer regions, including UMIs and sample barcodes, were then identified within each read, and trimmed. Together, the sample barcode, UMI, and constant region primer were used to assign molecular groupings for each read. Within each grouping, usearch (20) was used to subdivide the grouping, with a cutoff of 80% nucleotide identity, to account for randomly overlapping UMIs. Each of the resulting groupings is assumed to represent reads arising from a single RNA. Reads within each grouping were then aligned, and a consensus sequence determined. Finally, duplicate reads were collapsed into a single processed sequence for analysis. Collapsing duplicate reads ensures that each processed sequence represents a sequence from a single B cell and our analysis is not confounded by expression level.

For each processed sequence, IgBlast (21) was used to determine V, D and J gene segments, and locations of the complementarity determining regions (CDRs) and framework regions (FWRs). Isotype was determined based on comparison to germline constant region sequences. Sequences annotated as unproductive by IgBlast were removed. The number of mutations within each sequence was determined using the shazam R package (19).

## Public BCR Sequence Data Processing

The healthy control BCR sequence dataset used here has been described previously (22). It was prepared using the same primer set that was used in the current study, so comparisons should not suffer from protocol-specific biases. Only samples from participants aged 10 years or older, and from peripheral blood were used. This resulted in samples from 40 different participants, with a mean age of 28 (range: 11–51). Furthermore, only class-switched sequences were considered.

The influenza vaccine BCR sequence dataset used here was a combination of data from two different studies, and amounted to samples from 6 different participants in total (23, 24). All participants were administered a seasonal influenza vaccine, and peripheral blood was taken for BCR sequencing 6–9 days following vaccination. The processed data from these studies was obtained directly from the Observed Antibody Space database, and only the class-switched sequences were considered (25).

The publicly available BCR sequence dataset from six COVID-19 patients in Stanford, USA was also obtained directly from the Observed Antibody Space database. To circumvent the disparity in collapsed dataset sizes between pairs of replicates, we selected the replicate with the highest number of sequences for downstream analysis.

## Clonotyping

BCR sequences were clustered to identify those arising from presumably clonally related B cells; a process termed clonotyping.

Sequences from the 31 COVID-19 patients, the 7 metastatic cancer patients, the 6 influenza vaccine participants, and the 40 healthy controls were clustered together to also identify convergent clusters between samples. Clustering was performed using a previously described algorithm (26). Clustering required identical V and J gene segment usage, identical CDRH3 length, and allowed 1 AA mismatch for every 10 AAs within the CDRH3. Cluster centers were defined as the most common sequence within the cluster. Lineages were reconstructed from clusters using the alakazam R package (27).

## Bronchoalveolar Lavage RNAseq Data Processing

The bronchoalveolar lavage data comes from a previously published study of SARS-CoV-2 infection (28), with data available under the PRJNA605983 BioProject on NCBI. MIXCR v3.0.3 was used, with default settings, to extract reads mapping to antibody genes from the total RNASeq data (29).

## Convergent Clonotype Matching to Public Data

The public COVID-19 BCR sequence dataset (15), the output from the SARS-CoV-2 RNAseq data processing (28), and sequences from CoV-AbDab (30) were scanned for matches to our 1,254 convergent clonotype cluster centers. For comparison to the COVID-19 BCR sequence dataset and the RNAseq dataset, matching was performed using the same threshold that was used for clonotyping – requiring identical V and J genes, the same length CDRH3, and within 1 AA mismatch per 10 CDRH3 AAs to the convergent clonotype cluster center. For matching to the CoV-AbDab, a more lenient threshold was used. Sequences from CoV-AbDab were clustered alongside the representative CDRH3 sequence from each of our 1,254 convergent clones using the cd-hit-2d script from the CD-HIT package (31), at an 80% sequence identity threshold. Clustered CDRH3s were not required to have the same length, however the shorter CDRH3 was required to be at least 90% of the length of the longer CDRH3. Cluster centers containing at least one CoV-AbDab CDRH3 and one convergent clonotype CDRH3 were further investigated.

## Statistical Analysis and Graphing

Statistical analysis and plotting were performed using R (32). Plotting was performed using ggplot2 (33). Specific statistical tests used are detailed in the figure legends. Correlations of IGHV4-34 autoreactive motifs was performed by MANOVA in R.

# RESULTS

## COVID-19 Disease Samples

Blood samples were collected from thirty-one patients admitted to hospital with acute COVID-19 pneumonia. The mean age of patients was 56.2 (SD 17.7) years and 20 (65%) were male (**Supplementary Table 1**). All patients had a clinical history

consistent with COVID-19 and typical radiological changes. Twenty-eight patients had a confirmatory positive PCR test for SARS-CoV-2. The patients experienced an average of 16 days (range 4–55) of symptoms prior to the day on which the blood sample was collected. Twelve of the patients were still requiring hospital care but not oxygen therapy on day of sample collection (WHO Ordinal Scale Score 3), while 13 were hospitalized requiring oxygen by conventional mask or nasal prongs (WHO Ordinal Scale Score 4) and six were hospitalized with severe COVID-19 pneumonia requiring high-flow nasal oxygen or continuous positive airway pressure by mask (WHO Ordinal Scale Score 5). On the day of sample collection, the direct clinical care team considered two patients to be deteriorating, ten improving and the remaining nineteen were clinically stable.

It has previously been observed that lymphopenia is seen in severe COVID-19 patients (34). Our results recapitulate these findings, showing a correlation between lymphocyte count and clinical status (**Supplementary Figure 1**). As COVID-19 lymphopenia is caused by a reduction in T cell rather than B cell counts, we do not expect the different lymphocyte counts from our patients to impact the BCR sequencing results.

## SARS-CoV-2 Infection Results in a Stereotypic B Cell Response

IGHA and IGHG BCR sequencing yielded on average 154,717 unique sequences, and 24,043 clonotypes per sample (**Supplementary Table 1**). To characterize the B cell response in COVID-19, we compared this BCR repertoire data to BCR repertoire data from healthy controls obtained in a separate study (**Supplementary Table 2**), but amplified using the same primer set (22). Comparing IGHV gene segment usage revealed a significantly different IGHV gene usage in COVID-19 patients compared to the healthy controls, most notably with increases in the usage of IGHV2-70 (3.6× IGHA, 3.9× IGHG increase), IGHV3-30 (1.7× IGHA, 1.4× IGHG increase), IGHV5-51 (3.2× IGHA, 3.2× IGHG increase), and IGHV4-34 (1.2× IGHA, 2.2× IGHG increase) in the COVID-19 patients (**Figure 1A**).

We further investigated the IGHV4-34 difference, as IGHV4-34 has been shown to produce self-reactive antibodies (35). The IGHV4-34 germline sequence contains two sequence motifs (an "AVY" hydrophobic patch in the FWR1, and an "NHS" glycosylation sequon in the CDR2) that are not found in other IGHV gene segments and are thought to drive the self-reactivity. We
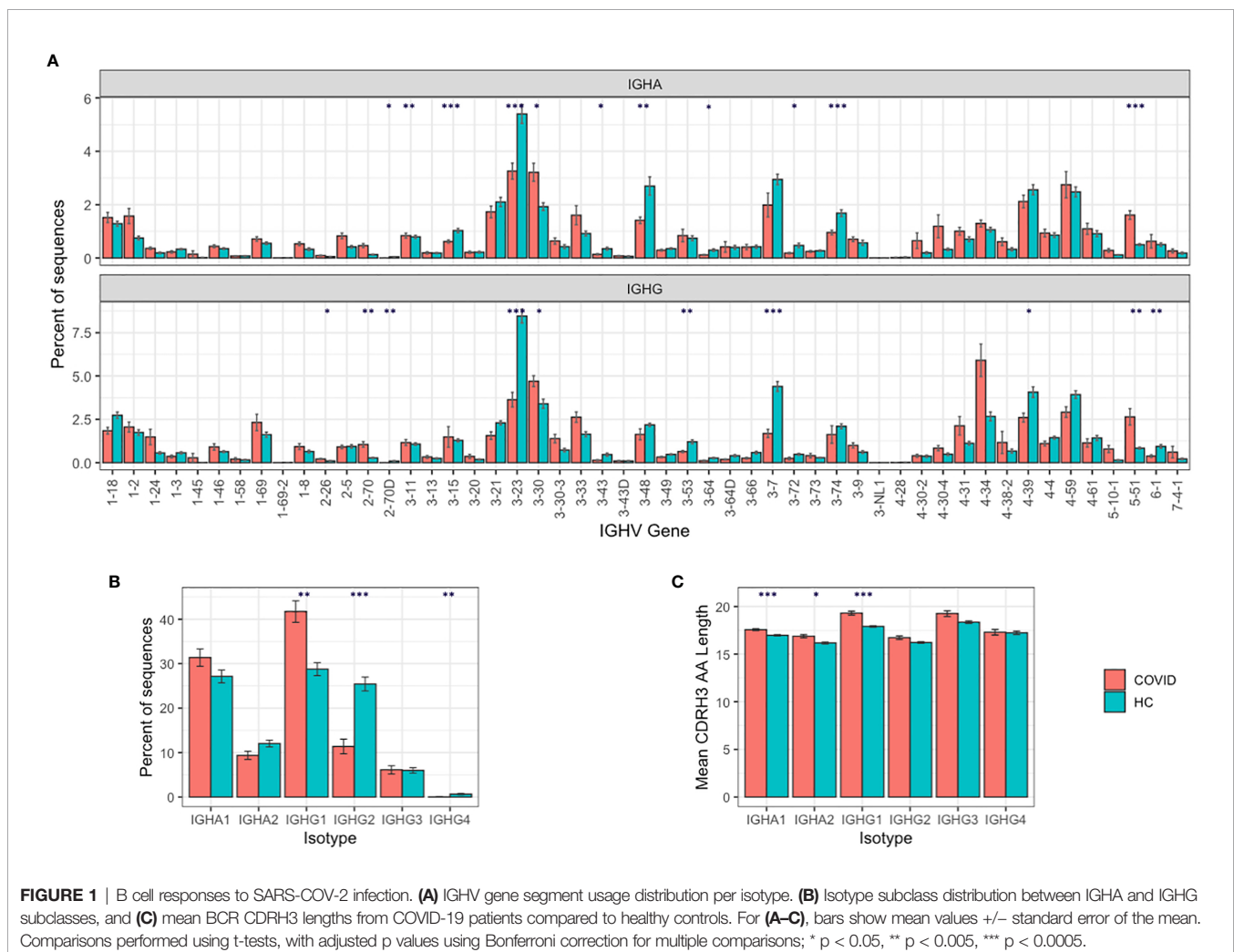


**FIGURE 1** | B cell responses to SARS-COV-2 infection. **(A)** IGHV gene segment usage distribution per isotype. **(B)** Isotype subclass distribution between IGHA and IGHG subclasses, and **(C)** mean BCR CDRH3 lengths from COVID-19 patients compared to healthy controls. For **(A–C)**, bars show mean values +/– standard error of the mean. Comparisons performed using t-tests, with adjusted p values using Bonferroni correction for multiple comparisons; * p < 0.05, ** p < 0.005, *** p < 0.0005.

show that the proportion of sequences containing the autoreactive AVY & NHS sequence motifs is increased in improving COVID-19 patients compared to stable or deteriorating COVID-19 patients, specifically in the IGHG1 isotype subclass (p-value = 0.013; **Supplementary Figure 2**).

Comparing isotype subclasses showed a significant increase in the relative usage of IGHG1 in COVID-19 patients (**Figure 1B**)—this is the first isotype subclass that is switched to upon activation of IGHM (36). There was also an increase in the mean CDRH3 length of the BCRs in the COVID-19 patients, that was most pronounced in the IGHA1, IGHA2, and IGHG1 isotype subclasses (**Figure 1C**).

## SARS-CoV-2 Infection Stimulates Both Naïve and Memory Responses
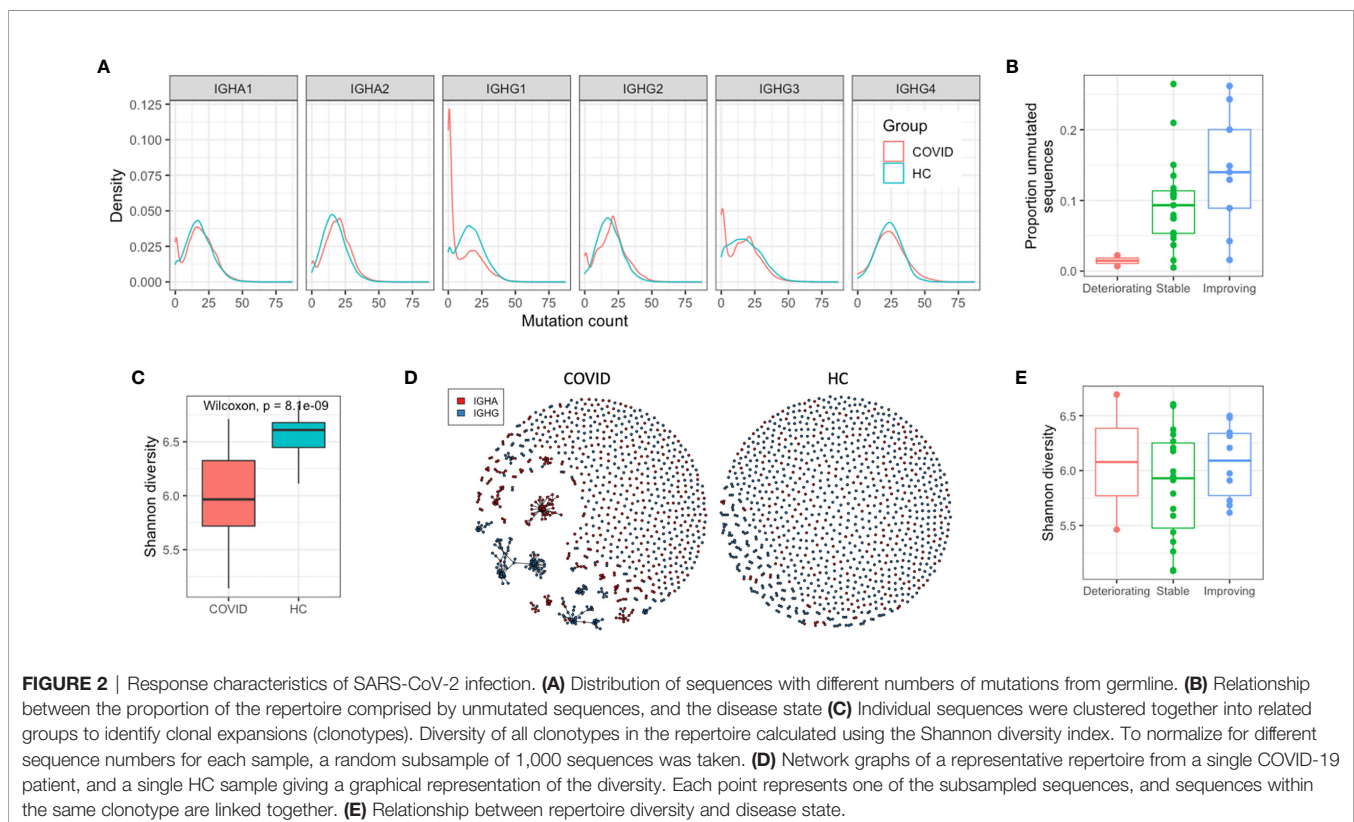
To further investigate the COVID-19-specific B cell response, we analyzed the characteristics of the BCR sequences that are consistent with recent B cell activation—somatic hypermutation and clonal expansion. In healthy controls, for class-switched sequences, there is a clear unimodal distribution of sequences with different numbers of mutations, and a mean mutation count across IGHA and IGHG isotypes of 17.6 (**Figure 2A**). In the COVID-19 samples, the mean mutation count was 14.8, and there was a bimodal distribution with a separate peak of sequences with no mutations. This bimodal distribution was most pronounced in the IGHG1, IGHG3, and IGHA1 isotype subclasses, corresponding to the increased isotype usages. Such a distribution is consistent with an expansion of recently class-switched B cells that have yet to undergo somatic hypermutation. There was considerable variation

between participants in the proportion of unmutated sequences (**Supplementary Figure 3**), which had no significant correlation with the number of days since symptom onset (R = −0.098, p = 0.61), but was lower in the deteriorating compared to improving patients (**Figure 2B**).

To investigate the clonal expansion in these antigen experienced BCR repertoires in more detail, the Shannon diversity index of each repertoire was calculated (while accounting for differences in read depth through subsampling). Lower diversity is indicative of a small number of dominant clonal expansions, while higher diversity is indicative or a larger number of smaller clonal expansions. The class-switched BCR repertoires of the COVID-19 patients were significantly less diverse than the BCR repertoires of the healthy controls due to the presence of large clonal expansions (**Figures 2C, D**), but there was no relationship between disease state and BCR repertoire diversity (**Figure 2E**). Calculating the mean mutation count of the clonal expansions showed that the largest ones were all highly mutated; the mean mutation of the ten largest clonal expansions in each COVID-19 patient repertoire was 18.4. Taken together, these observations suggest the response to COVID-19 consists of both recently activated B cells with low levels of mutation, and historic memory B cells that have high levels of mutation, and form particularly large clonal expansions.

## Sequence Convergence Can Be Used to Identify Putative SARS-CoV-2 Specific Antibodies

Given the skewing of the B cell response in the COVID-19 patients to specific IGHV genes, we next investigated whether



**FIGURE 2** | Response characteristics of SARS-CoV-2 infection. **(A)** Distribution of sequences with different numbers of mutations from germline. **(B)** Relationship between the proportion of the repertoire comprised by unmutated sequences, and the disease state **(C)** Individual sequences were clustered together into related groups to identify clonal expansions (clonotypes). Diversity of all clonotypes in the repertoire calculated using the Shannon diversity index. To normalize for different sequence numbers for each sample, a random subsample of 1,000 sequences was taken. **(D)** Network graphs of a representative repertoire from a single COVID-19 patient, and a single HC sample giving a graphical representation of the diversity. Each point represents one of the subsampled sequences, and sequences within the same clonotype are linked together. **(E)** Relationship between repertoire diversity and disease state.
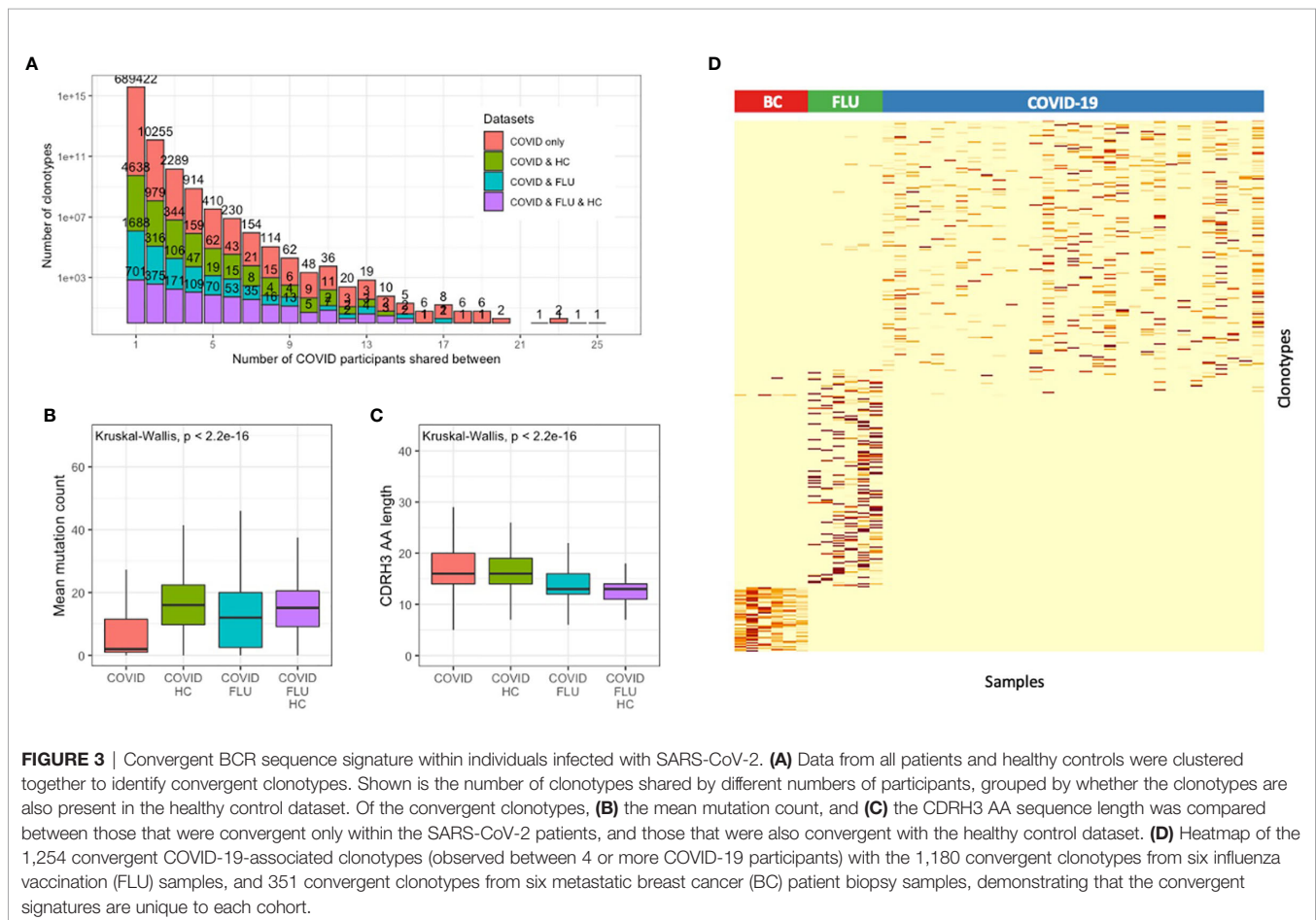
the same similarity was also seen on the BCR sequence level between different participants. Such convergent BCR signatures have been observed in response to other infectious diseases (37) and may be used to identify disease-specific antibody sequences.

Of the 714,107 total clonotypes across all the COVID-19 patients, 17,658 (2.5%) were shared between at least two of the participants (**Figure 3A**). Sample-to-sample contamination within sequencing batches can erroneously be interpreted as convergence, so we used a dual sample barcoding method, and strict laboratory procedures to minimize this. Although an element of contamination can never entirely be ruled out, we found no batch effects between the three sequencing runs we conducted (**Supplementary Figure 4**). As convergence could also occur by chance or be due to an unrelated memory response from commonly encountered pathogens, additional datasets were incorporated into the convergence analysis. The healthy control dataset was used to identify BCR sequences that may be shared by chance or come from commonly encountered pathogens. In addition, a BCR sequence dataset of six individuals who had received an influenza vaccination 6–9 days prior to sampling, was used to further identify BCR sequences that could arise during general respiratory tract infection (23, 24).

Of the 17,658 convergent clonotypes, 2,530 (14.3%) were also present in at least one of the 40 healthy control samples, 1,400 (7.9%) were also present in at least one of the 6 influenza samples, and 869 (4.9%) were also present in at least one of the healthy controls and one of the influenza samples. As expected, of the convergent clonotypes that were also present in the healthy control or influenza samples, the mean mutation count was significantly greater (**Figure 3B**), and the mean CDRH3 length significantly shorter (**Figure 3C**) than the convergent clonotypes that were unique to the COVID-19 patients.

To identify a set of potentially SARS-CoV-2-specific antibody sequences with high confidence, we first selected convergent clonotypes that were shared between at least four of the COVID-19 patients, but not seen in the healthy controls. This revealed 1,337 total clonotypes. For comparison, the same analysis was also conducted on the influenza sample dataset but using a more lenient threshold of sharing between at least two samples in order to identify a similar total number of 1,180 total clonotypes. Finally, as a technical control to determine whether the convergent signature could represent laboratory contamination, the same analysis was performed on an unrelated set of six metastatic breast cancer patient biopsy samples (16), which identified 351



**FIGURE 3** | Convergent BCR sequence signature within individuals infected with SARS-CoV-2. **(A)** Data from all patients and healthy controls were clustered together to identify convergent clonotypes. Shown is the number of clonotypes shared by different numbers of participants, grouped by whether the clonotypes are also present in the healthy control dataset. Of the convergent clonotypes, **(B)** the mean mutation count, and **(C)** the CDRH3 AA sequence length was compared between those that were convergent only within the SARS-CoV-2 patients, and those that were also convergent with the healthy control dataset. **(D)** Heatmap of the 1,254 convergent COVID-19-associated clonotypes (observed between 4 or more COVID-19 participants) with the 1,180 convergent clonotypes from six influenza vaccination (FLU) samples, and 351 convergent clonotypes from six metastatic breast cancer (BC) patient biopsy samples, demonstrating that the convergent signatures are unique to each cohort.

convergent clonotypes. These convergent clonotypes identified from each of the COVID-19, influenza and metastatic cancer cohorts were highly specific to each sample cohort (**Figure 3D**). Removing the small amount of overlap that the 1,337 COVID-19 convergent clonotypes had with the influenza and metastatic cancer samples resulted in a final signature of 1,254 convergent clonotypes.

The 1,254 COVID-19 convergent clonotypes had low mutation levels, with a mean mutation count of 2.3, and only 115 clonotypes with a mean mutation greater than 5. The sequences within the convergent clonotypes were primarily of the IGHG1 (71%) and IGHA1 (15%) subclasses (**Supplementary Figure 5A**). The convergent clonotypes used a diversity of IGHV gene segments, with IGHV3-30, IGHV3-30-3 and IGHV3-33 as the most highly represented (**Supplementary Figure 5B**). This IGHV gene usage distribution differs between that of the total repertoire, and IGHV3-30 is also the most highly used IGHV gene in the CoV-AbDab (30). The proportion of the repertoire comprised by each of these convergent clonotypes was also calculated for each patient sample. This proportion significantly correlated with decreased disease severity for one of the convergent clonotypes (IGHV3-30, CARGPGDSIDYW, IGHJ4) even after Bonferroni correction for multiple testing (p = 0.0052).

## BCR Clonotype Sequence Convergence Signatures are Shared Between Different COVID-19 Studies in Different Locations and From Different Anatomical Sites

To further explore whether the convergent clonotypes observed in our study were indeed disease specific, and to determine whether such convergence was common across studies and geographic regions, we compared these 1,254 convergent clonotypes to public B cell datasets. In addition, to ensure any matches to public datasets were not due to chance, we also selected at random 1,254 clonotypes from the healthy control set and performed the same comparison in parallel.

First, we compared our data to RNAseq data of bronchoalveolar lavage fluid obtained from five of the first infected patients in Wuhan, China (28). These samples were obtained for the purpose of metagenomic analyses to identify the aetiological agent of the novel coronavirus but were re-analyzed to determine whether we could extract any transcripts from BCRs. From the 10,038,758 total reads, we were able to identify 16 unique CDR3 AA sequences (**Supplementary Table 2**). Of these, one had an exact AA match to a CDR3 sequence in our data (CTTDLHDYGDSDYW) and shared the same V gene segment (IGHV3-15), and J gene segment (IGHJ4) usage. The clonotype that the sequence belonged to contained 796 total sequences and was convergent between 10 of our 31 COVID-19 patients. The clonotype was highly diverse, and the sequences had evidence of low mutation from germline, with a mean mutation count over all sequences of 5.3 (**Supplementary Figure 6**). None of the 16 CDR3 sequences from the RNAseq data matched to any of the CDR3 sequences in the 1,254 healthy control clonotypes, or indeed from any of the clonotypes in the entire dataset of 40 healthy control samples, indicating that the

observed matching in the COVID-19 patients is unlikely just due to chance alone.

Next, we compared our 1,254 convergent clonotypes to CoV-AbDab – the Coronavirus Antibody Database (accessed 5th August 2020) (30). At the time of access, this database contained 1,396 redundant and 1,151 non-redundant CDRH3 sequences from published and patented antibodies proven to bind coronaviruses including SARS-CoV-1 and SARS-CoV-2. We found 39 of our 1,254 convergent clonotypes to have high CDRH3 homology to the antibodies in CoV-AbDab (**Table 1** and **Supplementary Table 4**). In contrast, only one of the 1,254 healthy control clonotypes had high CDR3 homology to an antibody in CoV-AbDab. Of the 39 convergent clonotype matches, 10 also shared the same IGHV and IGHJ gene segment usage and would be classed as belonging to the same clonotype (**Table 1**); the healthy control clonotype match did not share the same IGHV or IGHJ gene segment usage.

Finally, we compared our data to a publicly available BCR deep sequencing dataset from six COVID-19 patients from Stanford, USA. Four hundred sixty-three of our 1,254 convergent clonotypes matched (using the same definition we used for clonotyping within our dataset) sequences in this dataset showing the high level of convergence between studies (**Figure 4A** and **Supplementary Table 5**). The average number of clonotype matches to the Stanford COVID-19 patient repertoires was 103, but this varied considerably between patients and timepoints. Two of the six patients were seronegative at the day of sampling (7451 and 7453), and these two patients had the fewest clonotype matches (8 and 31, respectively). Patient 7453 had an additional sample taken two days later (following seroconversion), and at this point had a large increase in the number of clonotype matches to 201. There was one of the 1,254 convergent clonotypes that was found across all six of the Stanford patients, and 2 clonotypes that were found in all five samples from seroconverted patients, but not found in the seronegative patients. In contrast, only 19 of the 1,254 healthy control clonotypes matched to sequences in the Stanford dataset (**Figure 4B**).

## DISCUSSION

We have used deep sequencing of the BCR heavy chain repertoire to evaluate the B cell responses of 31 individuals with COVID-19. In agreement with previous studies, there was a skewing of the repertoire in the response to SARS-CoV-2 infection, with an increased use of certain V genes, an increase in the proportion of antibodies with longer CDRH3s, and an altered isotype subclass distribution (15). The significantly increased usage of IGHA1 observed in the COVID-19 patients is in line with mucosal responses, where the longer hinge in IGHA1 compared to IGHA2 may offer advantages in antigen recognition by allowing higher avidity bivalent interactions with distantly spaced antigens (42).

As anticipated, given the novel nature of the virus, SARS-CoV-2 infection largely stimulated a characteristically naïve-like

**TABLE 1 |** Convergent COVID-19 clonotypes with high CDRH3 homology to antibodies in the CoV-AbDab, and which share the same IGHV and IGHJ gene usage and CDRH3 length.

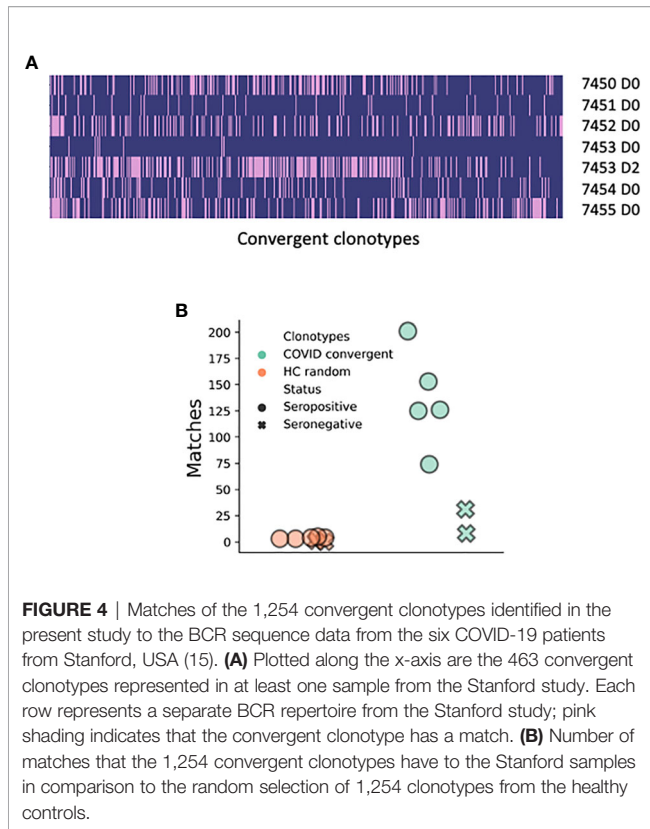| Name | CDRH3 | IGHV | IGHJ | Binds | Neutralizes | Reference |
|---|---|---|---|---|---|---|
| C154 | AKQAGPYCSGGSCYSAPFDY | 3-30 | 4 | CoV-1, CoV-2 | CoV-2 | Robbiani et al. (14) |
| ALC_3983948 | **AKVSGPYCSGGSCYSFYFDY** | **3-30** | **4** | | | |
| COV2-2068 | ARSYDILTGYRDAFDI | 3-53 | 3 | CoV-2 | CoV-2 | Zost et al. (38) |
| ALC_2318471 | **VRNYDILTGYSDAFDI** | **3-53** | **3** | | | |
| COV2-2007 | AKVSATYYYYYYGMDV | 3-30 | 6 | CoV-1, CoV-2 | | Zost et al. (38) |
| CC12.17 | AKSSGSYYYYYYGMDV | 3-30 | 6 | CoV-2 | CoV-2 | Rogers et al. (39) |
| ALC_2318830 | **AKVMTTYYYYYYGMDV** | **3-30** | **6** | | | |
| COVA2-14 | ARVRYYDSSGYYEDY | 1-69 | 4 | CoV-1, CoV-2 | | Brouwer et al. (12) |
| ALC_1780442 | **ARYDYYDSSGYYLDY** | **1-69** | **4** | | | |
| COV2-2270 | AITYYYDSSGYWWDD | 1-69 | 4 | CoV-1, CoV-2 | | Zost et al. (38) |
| ALC_1781971 | **ASTYYYDSSGYWFDY** | **1-69** | **4** | | | |
| COVA2-40 | AGRYCSGGRCGWFDP | 4-4 | 5 | CoV-2 | | Brouwer et al. (12) |
| ALC_1784026 | **ESRYCSGGSCGWFDP** | **4-4** | **5** | | | |
| COV2-2147 | ARSTSGSYYYGMDV | 3-30-3 | 6 | CoV-1, CoV-2 | | Zost et al. (38) |
| CV34 | ARSYGGSYYYGMDV | 3-30-3 | 6 | CoV-2 | | Seydoux et al. (40) |
| ALC_1249094 | **ARGTRGSYYYGMDV** | **3-30-3** | **6** | | | |
| COV2-2006 | ARPQSGGYYAPLDY | 3-30-3 | 4 | CoV-1, CoV-2 | | Zost et al. (38) |
| ALC_1246650 | **ARPYSGSYYAPLDY** | **3-30-3** | **4** | | | |
| S304 | ARGDSSGYYYYFDY | 3-13 | 4 | CoV-1, CoV-2 | CoV-1, CoV-2 | Pinto et al. (41) |
| ALC_1245048 | **ARGYSSGYYYYFDY** | **3-13** | **4** | | | |
| COV2-2027 | AIYGYYYYGLDV | 3-30 | 6 | CoV-2 | | Zost et al. (38) |
| ALC_480504 | **AVYGYYYYGMDV** | **3-30** | **6** | | | |

See **Supplementary Table 4** for convergent clonotypes with high CDRH3 homology to antibodies in the CoV-AbDab, but not sharing the same IGHV and IGHJ gene usage and allowing CDRH3 length mismatch. Each row represents a different group of similar sequences. Name: ID of the sequence. Black text represents the convergent clonotypes from this study, and blue text represents antibodies from CoV-AbDab. CDRH3, CDRH3 amino acid sequence. Green residues represent sequence matches between the convergent clonotype and the CoV-AbDab sequence, and red residues represent mismatches. IGHV: IGHV gene segment of the sequence. Green text represents matches between the convergent clonotype and the CoV-AbDab query sequence, and red text represents mismatches. IGHJ, IGHJ gene segment of the sequence. Green text represents matches between the convergent clonotype and the CoV-AbDab query sequence, and red text represents mismatches. ND, not determined. Binds/Neutralizes: Experimental confirmation of whether the sequence binds SARS-CoV-1 or SARS-CoV-2.

response, rather than a reactivation of pre-existing memory B cells. There was an increased prevalence of unmutated, but class-switched BCR sequences, an increase in the diversity of BCRs present in the class-switched B cell compartment, and an increase in the usage of isotype subclasses that are associated with recent viral immunity. It is interesting that the levels of mutation were so low given that the time point at which we sampled these patients is on average 16 days since symptom onset, and therefore even longer since initial infection. In controlled vaccination studies, it has been seen that 10 days post novel antigen challenge are sufficient to generate high levels of mutation (43). Such expansion of unmutated but class-switched BCR sequences has however also been seen in humans following rotavirus (44) and dengue virus (45) infection. In mice infected with vesicular stomatitis virus, it is seen that there is an unmutated IgG response upon initial infection which confers protection, but secondary challenge results in increased mutation and a change in IGHV gene segment usage (46). These observations are attributed to natural, or autoreactive antibodies, which are produced rapidly upon initial viral infection and recognize repetitive structures such as viral capsids. The increased prevalence of autoreactive IGHV4-34 sequences in improving COVID-19 patients compared to stable or deteriorating COVID-19 patients further

suggests a role for such natural or autoreactive antibodies in resolving infection and lowering the risk of pathology.

In addition to the naïve response, analysis of the largest clonal expansions in the SARS-CoV-2 patients, revealed them to be highly mutated, indicating that they may have arisen from memory recall. Such a secondary response to SARS-CoV-2 has been previously observed (47) and may be due to recall of B cells activated in response to previously circulating human coronaviruses, as recently highlighted (48, 49).

We observed a potential relationship between the observed repertoire characteristics and disease state. Improving patients showed a tendency toward a higher proportion of unmutated sequences compared to deteriorating patients, indicating that a naïve B cell response may be more effective than a memory recall response at clearing infection. Consistent with this hypothesis, Wec et al. found that B cells derived from memory recall generated low affinity, but SARS-CoV-1/2 cross-reactive antibodies, whereas B cells with the lowest mutation counts generated higher affinity SARS-CoV-2 specific antibodies (47). There is a clear need to expand on these findings by using larger sample cohorts and gathering more clinical data to aid understanding of the differences between individuals that respond with mild versus severe disease and have different recovery patterns. Building upon

**FIGURE 4** | Matches of the 1,254 convergent clonotypes identified in the present study to the BCR sequence data from the six COVID-19 patients from Stanford, USA (15). **(A)** Plotted along the x-axis are the 463 convergent clonotypes represented in at least one sample from the Stanford study. Each row represents a separate BCR repertoire from the Stanford study; pink shading indicates that the convergent clonotype has a match. **(B)** Number of matches that the 1,254 convergent clonotypes have to the Stanford samples in comparison to the random selection of 1,254 clonotypes from the healthy controls.

these observations could help to inform the future development of diagnostic assays to monitor and predict the progression of disease in infected patients.

A large number (1,254) of highly convergent clonotypes unique to COVID-19 were identified. Our approach of subtracting the convergent clonotypes also observed in healthy controls, and controls receiving influenza vaccination (22–24), allowed us to identify convergence specific to the disease cohort. The unbiased nature of the BCR repertoire analysis approach means that, while these convergent clonotypes are likely to include many antibodies to the spike protein and other parts of the virus they may also include other protective antibodies, including those to host proteins. Characterization of the heavy chains we have identified, coupled with matched light chains to generate functional antibodies will permit analysis of the binding sites and neutralizing potential of these antibodies. The report that plasma derived from recently recovered donors with high neutralizing antibody titers can improve the outcome of patients with severe disease (50), supports the hypotheses that intervention with a therapeutic antibody has the potential to be an effective treatment. A manufactured monoclonal antibody or combination of antibodies would also provide a simpler, scalable and safer approach than plasma therapy.

Sequence convergence between our 1,254 convergent clonotypes with heavy chains from published and patented SARS-CoV-1 and SARS-CoV-2 antibodies (30) supports several observations. Firstly, it demonstrates that our approach of finding a convergent sequence signature is a useful method for enriching disease-specific antibodies, as we find matches to known SARS-CoV spike-binding antibodies. Secondly, it shows that the clonotypes observed in response to SARS-CoV-2 overlap with those to SARS-CoV-1, presumably explained by the relatively high homology of the two related viruses (3). Indeed, here we show that there is an overrepresentation of clonotypes that correlate with patient clinical symptoms than is expected by chance, and these BCR sequences are associated with the dominant IgA1 and IgG1 responses. Finally, it shows that the convergence extends beyond our UK COVID-19 disease cohort.

Further evidence for convergence extending beyond our disease cohort came from the comparisons of our 1,254 convergent clonotypes to deep sequencing datasets from China (28) and the USA (15). The dataset from the USA is also from BCR sequencing of the peripheral blood of COVID-19 patients, and here we found matches to 463 of our 1,254 clonotypes. The dataset from China was from total RNA sequencing of the bronchoalveolar lavage fluid of SARS-CoV-2 infected patients. Only 16 unique CDRH3 sequences could be identified in this whole dataset, but one of them matched a convergent clonotype in the current study, showing that convergence can be seen both between different locations, and different sample types. We believe that the identification of such high BCR sequence convergence between geographically distinct and independent datasets could be highly significant and validates the disease association of the clonotypes, as well as the overall approach.

In summary, our BCR repertoire analysis provides information on the specific nature of the B cell response to SARS-CoV-2 infection. The information generated has the potential to facilitate the treatment of COVID-19 by supporting diagnostic approaches to predict the progression of disease, informing vaccine development and enabling the development of therapeutic antibody treatments and prophylactics.

## DATA AVAILABILITY STATEMENT

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by NHS HRA RES Ethics 19/SC/0361. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

JO, OC, AL, GT, DP, and PP conceived and designed the study. JG, SS, MR, RB-R, and AK conducted the data analysis with input from JO, RM, JD, GK, DF, LJ, RB-R, CD, OC, W-YL, GT, PP, and DP. GT, PP, and DP recruited COVID-19 participants

and executed the clinical protocols. SS, JD, and W-YL processed the COVID-19 clinical samples. CC and JS recruited the breast cancer participants. SS and JD processed the breast cancer samples. JB oversaw sequencing of the libraries. JG, JO, GK, SS, and RM wrote the manuscript with input from all co-authors. All authors contributed to the article and approved the submitted version.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2020.605170/full#supplementary-material

## REFERENCES

1. Lu H, Stratton CW, Tang YW. Outbreak of pneumonia of unknown etiology in Wuhan, China: The mystery and the miracle. *J Med Virol* (2020) 92:401–2. doi: 10.1002/jmv.25678
2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* (2020) 395:497–506. doi: 10.1016/S0140-6736(20)30183-5
3. Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J Virol* (2020) 94(7):e00127–20. doi: 10.1128/jvi.00127-20
4. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* (2020) 395:507–13. doi: 10.1016/S0140-6736(20)30211-7
5. Mulangu S, Dodd LE, Davey RT, Mbaya OT, Proschan M, Mukadi D, et al. A randomized, controlled trial of Ebola virus disease therapeutics. *N Engl J Med* (2019) 381:2293–303. doi: 10.1056/NEJMoa1910993
6. Gogtay NJ, Munshi R, Narayana DHA, Mahendra BJ, Kshirsagar V, Gunale B, et al. Comparison of a Novel Human Rabies Monoclonal Antibody to Human Rabies Immunoglobulin for Postexposure Prophylaxis: A Phase 2/3, Randomized, Single-Blind, Noninferiority, Controlled Study. *Clin Infect Dis* (2018) 66:387–95. doi: 10.1093/cid/cix791
7. Johnson S, Oliver C, Prince GA, Hemming VG, Pfarr DS, Wang S, et al. Development of a Humanized Monoclonal Antibody (MEDI-493) with Potent In Vitro and In Vivo Activity against Respiratory Syncytial Virus. *J Infect Dis* (1997) 176:1215–24. doi: 10.1086/514115
8. Wang SF, Tseng SP, Yen CH, Yang JY, Tsao CH, Shen CW, et al. Antibody-dependent enhancement of SARS coronavirus infection is mediated by antibodies against spike proteins. *Biochem Biophys Res Commun* (2014) 451:208–14. doi: 10.1016/j.bbrc.2014.07.090
9. Tetro JA. Is COVID-19 receiving ADE from other coronaviruses? *Microbes Infect* (2020) 22:72–3. doi: 10.1016/j.micinf.2020.02.006
10. Sharma A. It is too soon to attribute ADE to COVID-19. *Microbes Infect* (2020) 22:158. doi: 10.1016/j.micinf.2020.03.005
11. Wang Q, Zhang L, Kuwahara K, Li L, Liu Z, Li T, et al. Immunodominant SARS coronavirus epitopes in humans elicited both enhancing and neutralizing effects on infection in non-human primates. *ACS Infect Dis* (2016) 2:361–76. doi: 10.1021/acsinfecdis.6b00006
12. Brouwer PJM, Caniels TG, van der Straten K, Snitselaar JL, Aldon Y, Bangaru S, et al. Potent neutralizing antibodies from COVID-19 patients define multiple targets of vulnerability. *Sci (80- )* (2020) 369:eabc5902. doi: 10.1126/science.abc5902
13. Andreano E, Nicastri E, Paciello I, Pileri P, Manganaro N, Piccini G, et al. Identification of neutralizing human monoclonal antibodies from Italian Covid-19 convalescent patients. *bioRxiv* (2020). doi: 10.1101/2020.05.05.078154
14. Robbiani DF, Gaebler C, Muecksch F, Lorenzi JCC, Wang Z, Cho A, et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* (2020) 584:437–42. doi: 10.1038/s41586-020-2456-9
15. Nielsen SCA, Yang F, Jackson KJL, Hoh RA, Röltgen K, Jean GH, et al. Human B Cell Clonal Expansion and Convergent Antibody Responses to SARS-CoV-2. *Cell Host Microbe* (2020) 28:516–25.e5. doi: 10.1016/j.chom.2020.09.002
16. De Mattos-Arruda L, Sammut SJ, Ross EM, Bashford-Rogers R, Greenstein E, Markus H, et al. The Genomic and Immune Landscapes of Lethal Metastatic Breast Cancer. *Cell Rep* (2019) 27:2690–708.e10. doi: 10.1016/j.celrep.2019.04.098
17. van Dongen JJM, Langerak AW, Brüggemann M, Evans PAS, Hummel M, Lavender FL, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. *Leukemia* (2003) 17:2257–317. doi: 10.1038/sj.leu.2403202
18. Vander Heiden JA, Yaari G, Uduman M, Stern JNH, O'Connor KC, Hafler DA, et al. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* (2014) 30:1930–2. doi: 10.1093/bioinformatics/btu138
19. Gupta NT, Vander Heiden JA, Uduman M, Gadala-Maria D, Yaari G, Kleinstein SH. Change-O: A toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* (2015) 31:3356–8. doi: 10.1093/bioinformatics/btv359
20. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* (2010) 26:2460–1. doi: 10.1093/bioinformatics/btq461
21. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res* (2013) 41:W34–40. doi: 10.1093/nar/gkt382
22. Ghraichy M, Galson JD, Kovaltsuk A, von Niederhäusern V, Schmid JP, Recher M, et al. Maturation of the Human B-Cell Receptor Repertoire with Age. *Front Immunol* (2020) 11:1734. doi: 10.2139/ssrn.3518535
23. Tipton CM, Fucile CF, Darce J, Chida A, Ichikawa T, Gregoretti I, et al. Diversity, cellular origin and autoreactivity of antibody-secreting cell population expansions in acute systemic lupus erythematosus. *Nat Immunol* (2015) 16:755–65. doi: 10.1038/ni.3175
24. Gupta NT, Adams KD, Briggs AW, Timberlake SC, Vigneault F, Kleinstein SH. Hierarchical Clustering Can Identify B Cell Clones with High Confidence in Ig Repertoire Sequencing Data. *J Immunol* (2017) 198:2489–99. doi: 10.4049/jimmunol.1601850
25. Kovaltsuk A, Leem J, Kelm S, Snowden J, Deane CM, Krawczyk K. Observed Antibody Space: A Resource for Data Mining Next-Generation Sequencing of Antibody Repertoires. *J Immunol* (2018) 201:2502–9. doi: 10.4049/jimmunol.1800708

26. Galson JD, Clutterbuck EA, Trück J, Ramasamy MN, Münz M, Fowler A, et al. BCR repertoire sequencing: different patterns of B cell activation after two Meningococcal vaccines. *Immunol Cell Biol* (2015) 93:885–95. doi: 10.1038/icb.2015.57

27. Gupta N, Vander Heiden J, Uduman M, Gadala-Maria Heiden D, Yaari G, Kleinstein S. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics* (2015) 1–3. doi: 10.1093/bioinformatics/btv359

28. Zhou P, Lou YX, XG W, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* (2020) 579:270–3. doi: 10.1038/s41586-020-2012-7

29. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva EV, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* (2015) 12:380–1. doi: 10.1038/nmeth.3364

30. Raybould MIJ, Kovaltsuk A, Marks C, Deane CM. CoV-AbDab: the coronavirus antibody database. *Bioinformatics* (2020) btaa739. doi: 10.1093/bioinformatics/btaa739

31. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* (2012) 28:3150–2. doi: 10.1093/bioinformatics/bts565

32. R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing (2008). Available at: http://www.r-project.org.

33. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. 1st ed. Springer-Verlag New York: Springer (2009). Available at: https://ggplot2.tidyverse.org.

34. Tavakolpour S, Rakhshandehroo T, Wei EX, Rashidian M. Lymphopenia during the COVID-19 infection: What it shows and what can be learned. *Immunol Lett* (2020) 225:31–2. doi: 10.1016/j.imlet.2020.06.013

35. Pascual V, Victor K, Lelsz D, Spellerberg MB, Hamblin TJ, Thompson KM, et al. Nucleotide sequence analysis of the V regions of two IgM cold agglutinins: Evidence that the V(H)4-21 gene segment is responsible for the major cross-reactive idiotype. *J Immunol* (1991) 146:4385–91.

36. Horns F, Vollmers C, Croote D, Mackey SF, Swan GE, Dekker CL, et al. Lineage tracing of human B cells reveals the in vivo landscape of human antibody class switching. *Elife* (2016) 5:1–20. doi: 10.7554/eLife.16578

37. Parameswaran P, Liu Y, Roskin KM, Jackson KKL, Dixit VP, Lee J-Y, et al. Convergent Antibody Signatures in Human Dengue. *Cell Host Microbe* (2013) 13:691–700. doi: 10.1016/j.chom.2013.05.008

38. Zost SJ, Gilchuk P, Chen RE, Case JB, Reidy JX, Trivette A, et al. Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nat Med* (2020) 26:1422–7. doi: 10.1038/s41591-020-0998-x

39. Rogers TF, Zhao F, Huang D, Beutler N, Burns A, He W, et al. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Sci (80- )* (2020) 369(6506):956–63. doi: 10.1126/science.abc7520

40. Seydoux E, Homad LJ, MacCamy AJ, Parks KR, Hurlburt NK, Jennewein MF, et al. Characterization of neutralizing antibodies from a SARS-CoV-2 infected individual. *bioRxiv* (2020). doi: 10.1101/2020.05.12.091298

41. Pinto D, Park YJ, Beltramello M, Walls AC, Tortorici MA, Bianchi S, et al. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* (2020) 583:290–5. doi: 10.1038/s41586-020-2349-y

42. Li Q, Hou S, Schiller ZA, Wallace AL, Kurt Yilmaz N, Toomey JR, et al. IgA MAb blocks SARS-CoV-2 Spike-ACE2 interaction providing mucosal immunity. *bioRxiv* (2020). doi: 10.1101/2020.05.15.096719

43. Wec AZ, Haslwanter D, Abdiche YN, Shehata L, Pedreño-Lopez N, Moyer CL, et al. Longitudinal dynamics of the human B cell response to the yellow fever 17D vaccine. *Proc Natl Acad Sci USA* (2020) 117:6675–85. doi: 10.1073/pnas.1921388117

44. Tian C, Luskin GK, Dischert KM, Higginbotham JN, Shepherd BE, Crowe JE. Immunodominance of the VH1-46 antibody gene segment in the primary repertoire of human rotavirus-specific B cells is reduced in the memory compartment through somatic mutation of nondominant clones. *J Immunol* (2008) 180:3279–88. doi: 10.4049/jimmunol.180.5.3279

45. Godoy-Lozano EE, Téllez-Sosa J, Sánchez-González G, Sámano-Sánchez H, Aguilar-Salgado A, Salinas-Rodríguez A, et al. Lower IgG somatic hypermutation rates during acute dengue virus infection is compatible with a germinal center-independent B cell response. *Genome Med* (2016) 8:23. doi: 10.1186/s13073-016-0276-1

46. Kalinke U, Oxenius A, López-Macías C, Zinkernagel RM, Hengartner H. Virus neutralization by germ-line vs. hypermutated antibodies. *Proc Natl Acad Sci USA* (2000) 97:10126–31. doi: 10.1073/pnas.97.18.10126

47. Wec AZ, Wrapp D, Herbert AS, Maurer D, Haslwanter D, Sakharkar M, et al. Broad sarbecovirus neutralizing antibodies define a key site of vulnerability on the SARS-CoV-2 spike protein. *bioRxiv* (2020). doi: 10.1101/2020.05.15.096511

48. Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Rydyznski Moderbacher R, et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* (2020) 181(7):1489–501. doi: 10.1016/j.cell.2020.05.015

49. Ng K, Faulkner N, Cornish G, Rosa A, Harvey R, Hussain S, et al. Pre-existing and de novo humoral immunity to SARS-CoV-2 in humans. *Sci (80- )* (2020) eabe1107. doi: 10.1101/2020.05.14.095414

50. Duan K, Liu B, Li C, Zhang H, Yu T, Qu J, et al. Effectiveness of convalescent plasma therapy in severe COVID-19 patients. *Proc Natl Acad Sci* (2020) 117:202004168. doi: 10.1073/pnas.2004168117

51. Galson JD, Schaetzle S, Bashford-Rogers RJM, Raybould MIJ, Kovaltsuk A, Kilpatrick GJ, et al. Deep sequencing of B cell receptor repertoires from COVID-19 patients reveals strong convergent immune signatures. *bioRxiv* (2020). doi: 10.1101/2020.05.20.106294. 2020.05.20.106294.

52. Corrie BD, Marthandan N, Zimonja B, Jaglale J, Zhou Y, Barr E, et al. iReceptor: A platform for querying and analyzing antibody/B-cell and T-cell receptor repertoire data across federated repositories. *Immunol Rev* (2018) 284:24–41. doi: 10.1111/imr.12666