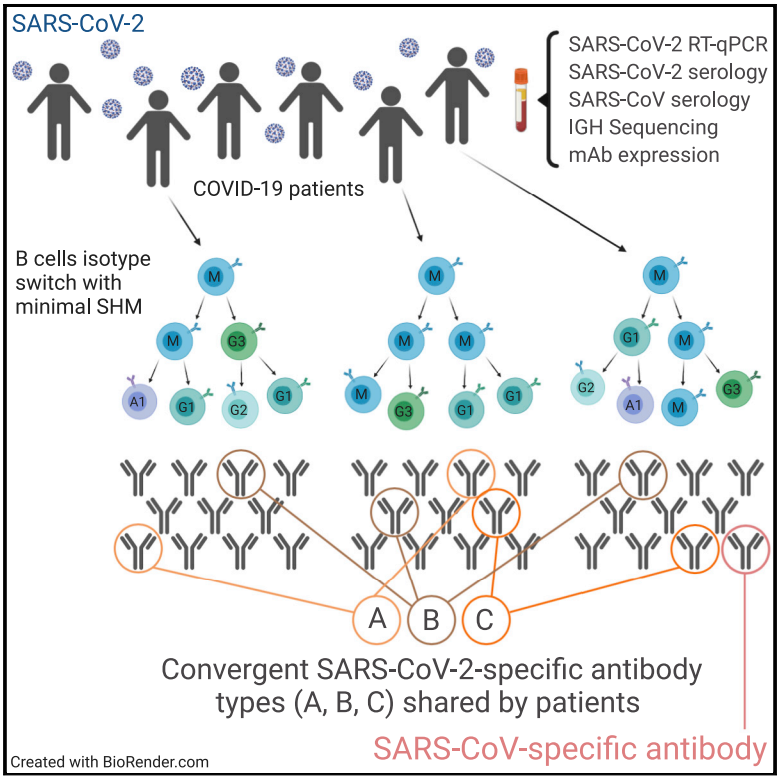
Short Article

Cell Host & Microbe

Human B Cell Clonal Expansion and Convergent Antibody Responses to SARS-CoV-2

Graphical Abstract



Highlights

- Human IGH repertoire sequencing identifies SARS-CoV-2-specific B cell clones
- Convergent virus-specific antibody sequences are shared between COVID-19 patients
- SARS-CoV antibodies are detected by sequence in COVID-19 patients

Authors

Sandra C.A. Nielsen, Fan Yang, Katherine J.L. Jackson, ..., Benjamin A. Pinsky, Catherine A. Blish, Scott D. Boyd

Correspondence

cblish@stanford.edu (C.A.B.), sboyd1@stanford.edu (S.D.B.)

In Brief

B cells produce antibodies and provide protective immunity to viruses. In longitudinal data from COVID-19 patients, Nielsen et al. analyze the development of antibody gene repertoires responding to SARS-CoV-2. COVID-19 patients share subsets of similar antibody sequences that bind SARS-CoV-2 antigens, with rare antibodies that also recognize SARS-CoV.

Nielsen et al., 2020, Cell Host & Microbe *28*, 516–525
October 7, 2020
https://doi.org/10.1016/j.chom.2020.09.002

CellPress

CellPress

Cell Host & Microbe

Short Article

Human B Cell Clonal Expansion and Convergent Antibody Responses to SARS-CoV-2

Sandra C.A. Nielsen,[1,14] Fan Yang,[1,14] Katherine J.L. Jackson,[2,14] Ramona A. Hoh,[1,14] Katharina Röltgen,[1] Grace H. Jean,[3] Bryan A. Stevens,[1] Ji-Yeun Lee,[1] Arjun Rustagi,[4] Angela J. Rogers,[5] Abigail E. Powell,[6] Molly Hunter,[7] Javaria Najeeb,[8] Ana R. Otrelo-Cardoso,[8] Kathryn E. Yost,[9] Bence Daniel,[1] Kari C. Nadeau,[5,10] Howard Y. Chang,[9,11] Ansuman T. Satpathy,[1] Theodore S. Jardetzky,[8,10] Peter S. Kim,[6,12] Taia T. Wang,[4,12,13] Benjamin A. Pinsky,[1] Catherine A. Blish,[4,12,*] and Scott D. Boyd[1,10,15,*]

[1]Department of Pathology, Stanford University, Stanford, CA 94305, USA
[2]Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia
[3]Department of Developmental Biology, Stanford University, Stanford, CA 94305, USA
[4]Department of Medicine, Division of Infectious Diseases and Geographic Medicine, Stanford University, Stanford, CA 94305, USA
[5]Department of Medicine, Division of Pulmonary, Allergy and Critical Care Medicine, Stanford University, Stanford, CA 94305, USA
[6]Stanford ChEM-H and Department of Biochemistry, Stanford University, Stanford, CA 94305, USA
[7]ATUM, Newark, CA, 94560, USA
[8]Department of Structural Biology, Stanford University, Stanford, CA 94305, USA
[9]Center for Personal Dynamic Regulomes, Stanford University, Stanford, CA 94305, USA
[10]Sean N. Parker Center for Allergy and Asthma Research, Stanford University School of Medicine, Stanford, CA, 94305, USA
[11]Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA
[12]Chan Zuckerberg Biohub, San Francisco, CA 94158, USA
[13]Department of Microbiology and Immunology, Stanford University, Stanford, CA 94305, USA
[14]These authors contributed equally
[15]Lead Contact
*Correspondence: cblish@stanford.edu (C.A.B.), sboyd1@stanford.edu (S.D.B.)
https://doi.org/10.1016/j.chom.2020.09.002

SUMMARY

B cells are critical for the production of antibodies and protective immunity to viruses. Here we show that patients infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) who develop coronavirus disease 2019 (COVID-19) display early recruitment of B cells expressing a limited subset of IGHV genes, progressing to a highly polyclonal response of B cells with broader IGHV gene usage and extensive class switching to IgG and IgA subclasses with limited somatic hypermutation in the initial weeks of infection. We identify convergence of antibody sequences across SARS-CoV-2-infected patients, highlighting stereotyped naive responses to this virus. Notably, sequence-based detection in COVID-19 patients of convergent B cell clonotypes previously reported in SARS-CoV infection predicts the presence of SARS-CoV/SARS-CoV-2 cross-reactive antibody titers specific for the receptor-binding domain. These findings offer molecular insights into shared features of human B cell responses to SARS-CoV-2 and SARS-CoV.

INTRODUCTION

The novel human severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiological agent of the coronavirus disease 2019 (COVID-19) pandemic (Huang et al., 2020; Zhu et al., 2020). Prior to the emergence of SARS-CoV-2, six human coronaviruses (hCoVs) were known; four seasonal hCoVs (hCoV-229E, -NL63, -HKU1, and -OC43) (Su et al., 2016) causing usually mild upper respiratory illness, and the two more recently discovered are SARS-CoV (Peiris et al., 2003) and MERS-CoV (Zaki et al., 2012), viruses that arose from spillover events of virus from animals into humans. Humans are expected to be naive to SARS-CoV-2 and display a primary immune response to infection. Humoral immune responses will likely be critical for the development of protective immunity to SARS-CoV-2. Recently, SARS-CoV-2 neutralizing antibodies from convalescent COVID-19 patients have been reported (Cao et al., 2020; Ju et al., 2020; Robbiani et al., 2020; Wu et al., 2020b), offering an important resource of potential protective or therapeutic antibodies. Further analysis of the B cell antigen receptors (BCRs) specific for this virus is needed to define the shared or distinct features of humoral responses elicited in comparison with other viral infections and to assess the extent to which responses to SARS-CoV-2 have breadth extending to other coronaviruses within the subgenus *Sarbecovirus*.

RESULTS

SARS-CoV-2 Infection Causes Global Changes in the Antibody Repertoire

High-throughput DNA sequencing of BCR heavy chain genes defines clonal B cell lineages based on their unique receptor sequences and captures the hallmarks of clonal evolution, such as somatic hypermutation (SHM) and class-switch recombination in the humoral response (Zhou and Kleinstein, 2019). To study SARS-CoV-2-specific humoral responses, we collected a total of 38 peripheral blood specimens from 13 adult patients, sampled at a median of three time points (range 1–5) each, admitted to Stanford Healthcare with COVID-19 confirmed by quantitative reverse transcription PCR (RT-qPCR) testing. Sampling time points were measured as days post-symptom onset (DPSO). All patients showed SARS-CoV-2 receptor-binding domain (RBD)-specific IgA, IgG, and IgM antibodies (Table 1). Immunoglobulin heavy chain (IGH) repertoires were sequenced and compared to those from a healthy human control (HHC) dataset of 114 adults (Nielsen et al., 2019). Data from a HHC individual matched to the COVID-19 cohort by mean sequencing depth and B cell clones is shown in Figure 1 (top panel). In HHC at baseline, IgM and IgD are primarily derived from naive B cells with unmutated IGHV genes, whereas class-switched cells expressing IgA or IgG subclasses show IGHV SHM. In contrast, SARS-CoV-2 seroconverted patients (blue labels in Figure 1) show a highly polyclonal burst of B cell clones expressing IgG, and to a lesser extent IgA, with little to no SHM. Data from a patient prior to and after seroconversion shows an increase in the proportion of class-switched low-SHM clones (bottom panels in Figure 1). Seronegative samples (red labels in Figure 1) had IGH repertoires similar to uninfected HHC, suggesting an earlier stage in the immune response to infection at these time points.

The increased fraction of unmutated (<1% SHM in IGHV gene) clones in the IgG subclasses in seroconverted COVID-19 patients in comparison with HHC was statistically significant (IgG1: p value = 1.884e-08; IgG2: 1.554e-08; IgG3: 3.754e-08; IgG4: 0.00044) (Figures 2A and 2B). SHM frequencies and proportions of unmutated clones in each sample are in Figure S1. Most B cells in COVID-19 patients prior to seroconversion showed IgG SHM comparable to HHC, but the proportion of IgG-expressing (IgG+) B cells with low SHM increased rapidly over time (Figure 2A, lower panels). Notably, prior to seroconversion, B cells expressing some IGHV genes, particularly IGHV3-30-3 and IGHV1-2, showed earlier changes than the rest of the repertoire, with increased IgG class-switched low-SHM clones (Figures 2C and 2D). IGHV3-9 showed a similar trend but was not significant (Figures 2C and 2D).

We previously reported a similar influx of low-mutation clones into the IgG compartment in acute Ebola virus (EBOV) infection (Davis et al., 2019), but there was a prolonged delay, lasting months, in accumulation of SHM in those clones. In contrast, clones detected at two or more time points in COVID-19 patients for whom the longest time courses were available (patients 7453 and 7455; Figure 2B) show that after the initial appearance of low-SHM clone members, the proportion of IgG+ B cells with intermediate SHM frequencies (2%–5%) increases in the first three weeks post-onset of symptoms. Examination of the total clones within each isotype shows the appearance of low-SHM clones post-seroconversion within the first two weeks post-onset of symptoms and subsequent increases in SHM over the following two weeks (Figure 2E). In further contrast to EBOV, COVID-19 primary infection stimulated polyclonal B cell responses with both IgG and, in some patients, IgA subclasses, rather than predominantly IgG (Figure 1). Overall, among IgG+ B cells in COVID-19 patients, the proportion of IgG1+ cells was increased, with decreases in IgG2 and IgG3. Median usage of IgG1 was 1.7-fold greater than that seen in HHC B cells (Table S1).

Comparison of the IGHV genes used by COVID-19 and HHC individuals (Figure S2) revealed COVID-19 skewing of the responding IGH repertoires away from common IGHV genes in HHC, such as IGHV3-7, IGHV3-23, and IGHV5-51, and enrichment of IGHV1-24, IGHV3-9, IGHV3-13, and IGHV3-20 in IgG+ B cells. IGHV3-53 is reported to be enriched in RBD-binding B cells (Cao et al., 2020; Robbiani et al., 2020) but was not significantly overrepresented in the total IGH repertoires in the COVID-19 patients during acute infection. Preferential selection of B cells with particular IGHV is reported in other contexts, such as increased IGHV1-69 in response to some influenza virus antigens (Avnir et al., 2016). Highly utilized IGHV genes in IgG-seroconverted COVID-19 patients show low median IgG1 SHM (range 2.4%–7.5%), compared to higher median IgA1 SHM (range 5.8%–9.5%) (Figure 2C).

In COVID-19 patients, expanded clones detected in two or more replicate IGH sequence libraries showed increased proportions of low-SHM members (Figures S3A and S3B; IgG1, p values = 0.005; IgG2, 0.014, IgG3, 0.036). Expanded clones had longer and more hydrophobic IGH complementarity-determining region-3 (CDR-H3) sequences in class-switched isotypes in comparison with HHC (Figure 2F), consistent with a rapid proliferation of cells recently differentiated from naive B cells (Grimsholm et al., 2020). CDR-H3 charge and aromaticity showed modest differences in COVID-19 patients in comparison with HHC, including more negative charge in non-expanded clones (Figure S3C). Notably, the IGHV gene usage frequencies in expanded clones in comparison with non-expanded clones of COVID-19 patients showed a different pattern than overall IGHV gene usage, with IGHV1-24, IGHV3-13, and IGHV3-20 frequencies increased in the total repertoire but used less often in expanded clones (Figures S2 and S3D), suggesting that the B cells expressing these IGHV genes are highly polyclonal with small clone sizes. Eleven IGHV genes were significantly enriched in expanded clones versus non-expanded clones (Figure S3D), suggesting preferential recruitment and viral epitope binding of B cells by using these germline IGHV.

B Cells Expressing Convergent IGH Are Detected in COVID-19 Patients

Despite the diversity of antigen-driven antibody responses, we and others have previously identified patterns of highly similar, "convergent" antibodies shared by different individuals in response to pathogens such as EBOV (Davis et al., 2019), dengue virus (Parameswaran et al., 2013), and influenza virus (Jackson et al., 2014). Such convergent antibodies make up a small proportion of the total virus-specific B cell response in each individual (Davis et al., 2019). To identify putative SARS-CoV-2-specific antibody signatures, we analyzed clones with shared IGHV, IGHJ, and CDR-H3 region length and clustered

Table 1. Individual COVID-19 Patient Sample Information in DPSO

Patient ID	Age	Sex	Status[a]	DPSO	IgA[b]	IgG[c]	IgM[d]	gDNA Library	cDNA Library
7450	73	F	ICU	9	yes	yes	yes	yes	yes
				22	yes	yes	yes	yes	yes
				25	yes	yes	yes	yes	no
				27	yes	yes	yes	yes	yes
7451	62	M	inpatient	8	no	no	no	yes	yes
7452	61	F	inpatient	18	yes	yes	yes	yes	yes
7453	64	M	ICU	8	no	no	no	yes	yes
				11	no	no	no	yes	yes
				15	yes	yes	yes	yes	yes
				18	yes	yes	yes	yes	yes
				20	yes	yes	yes	yes	no
7454	42	M	ICU	16	yes	yes	yes	yes	yes
7455	36	M	ICU	9	yes	yes	yes	yes	yes
				11	yes	yes	yes	yes	yes
				12	yes	yes	yes	yes	yes
				16	yes	yes	yes	yes	yes
				19	yes	yes	yes	yes	yes
7480	40	F	ICU	11	yes	yes	yes	yes	yes
				14	yes	yes	yes	yes	yes
				17	yes	yes	yes	yes	no
7481	66	M	ICU	32	yes	yes	yes	yes	yes
				35	yes	yes	yes	yes	no
				37	yes	yes	yes	yes	no
7482	88	F	inpatient	5	no	no	no	yes	yes
				8	no	no	no	yes	yes
				10	no	no	yes	yes	yes
				12	no	yes	yes	yes	yes
7483	77	F	inpatient	35	yes	yes	yes	yes	yes
				40	yes	yes	yes	yes	yes
7484	36	M	ICU	11	yes	yes	yes	yes	yes
				14	yes	yes	yes	yes	yes
				17	yes	yes	yes	yes	yes
7485	45	F	ICU	8	no	yes	yes	no	yes
				10	yes	yes	yes	yes	yes
				12	yes	yes	yes	yes	yes
7486	69	F	inpatient	14	no	yes	yes	yes	yes
				17	yes	yes	yes	yes	yes
				19	yes	yes	yes	yes	yes

Seroconversion was determined by ELISA to the SARS-CoV-2 RBD antigen (STAR Methods).
[a]All patients were inpatients; eight patients required ICU care during their hospital stay.
[b]IgA seroconversion.
[c]IgG seroconversion.
[d]IgM seroconversion.

the CDR-H3 sequences at 85% amino acid identity to find clusters spanning two or more COVID-19 patients and absent from the 114 HHC individuals. One thousand two hundred and thirty-six convergent clusters met these criteria and showed SHM frequencies averaging 1.7% (range 0.5%–5.5%). An average of 196 convergent clusters were found per patient, ranging from 69 clusters in patient 7485 to 477 clusters in patient 7455. One thousand one hundred and seventy-one clusters were shared pairwise between two patients, 53 clusters spanned three patients, nine clusters spanned four patients, and three clusters spanned five patients (Figure 3A). To assess the significance of these convergent clones in COVID-19 patients, we undertook repeated resampling of 13 randomly selected HHC samples with the same parameters for 100 iterations. The number of

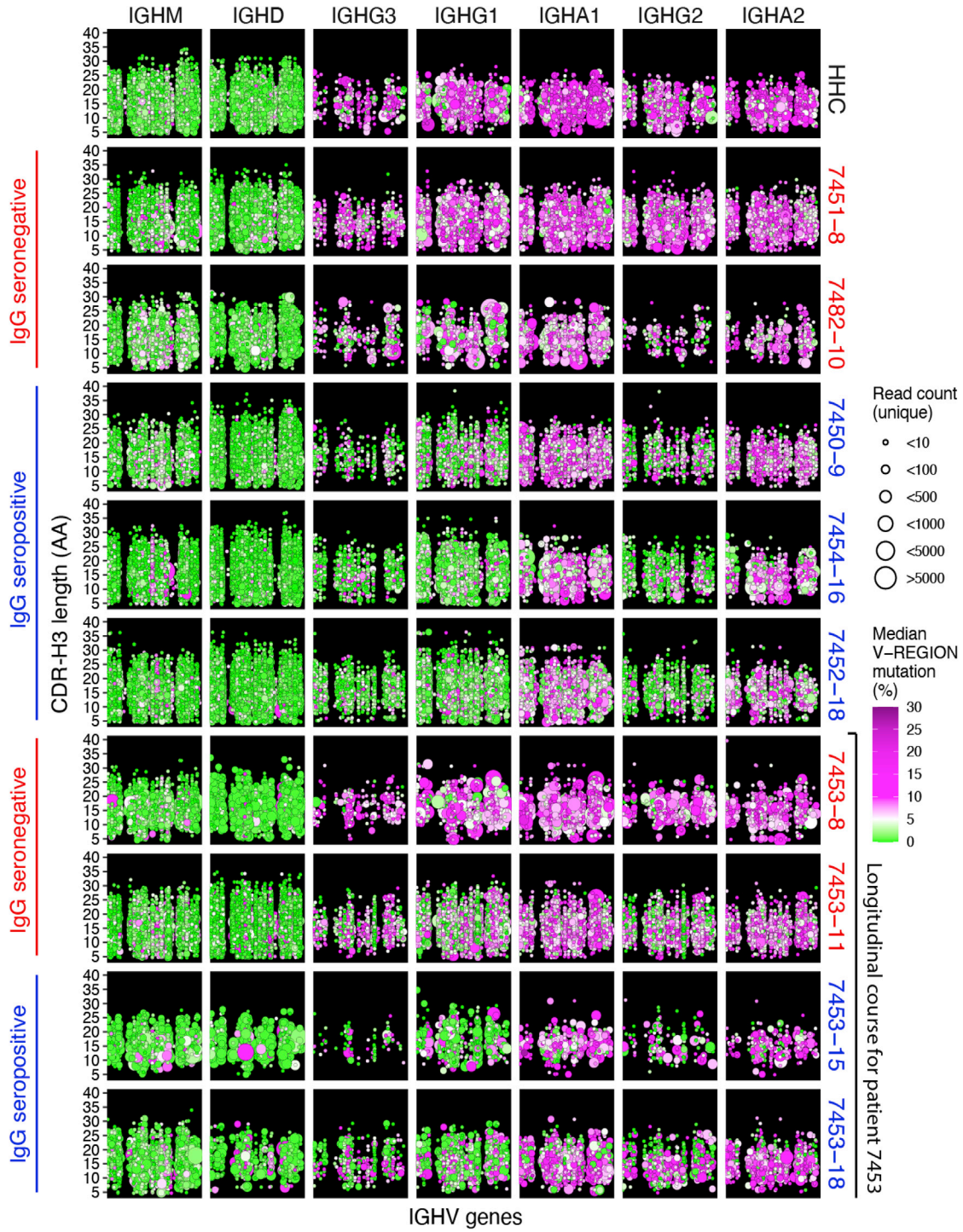
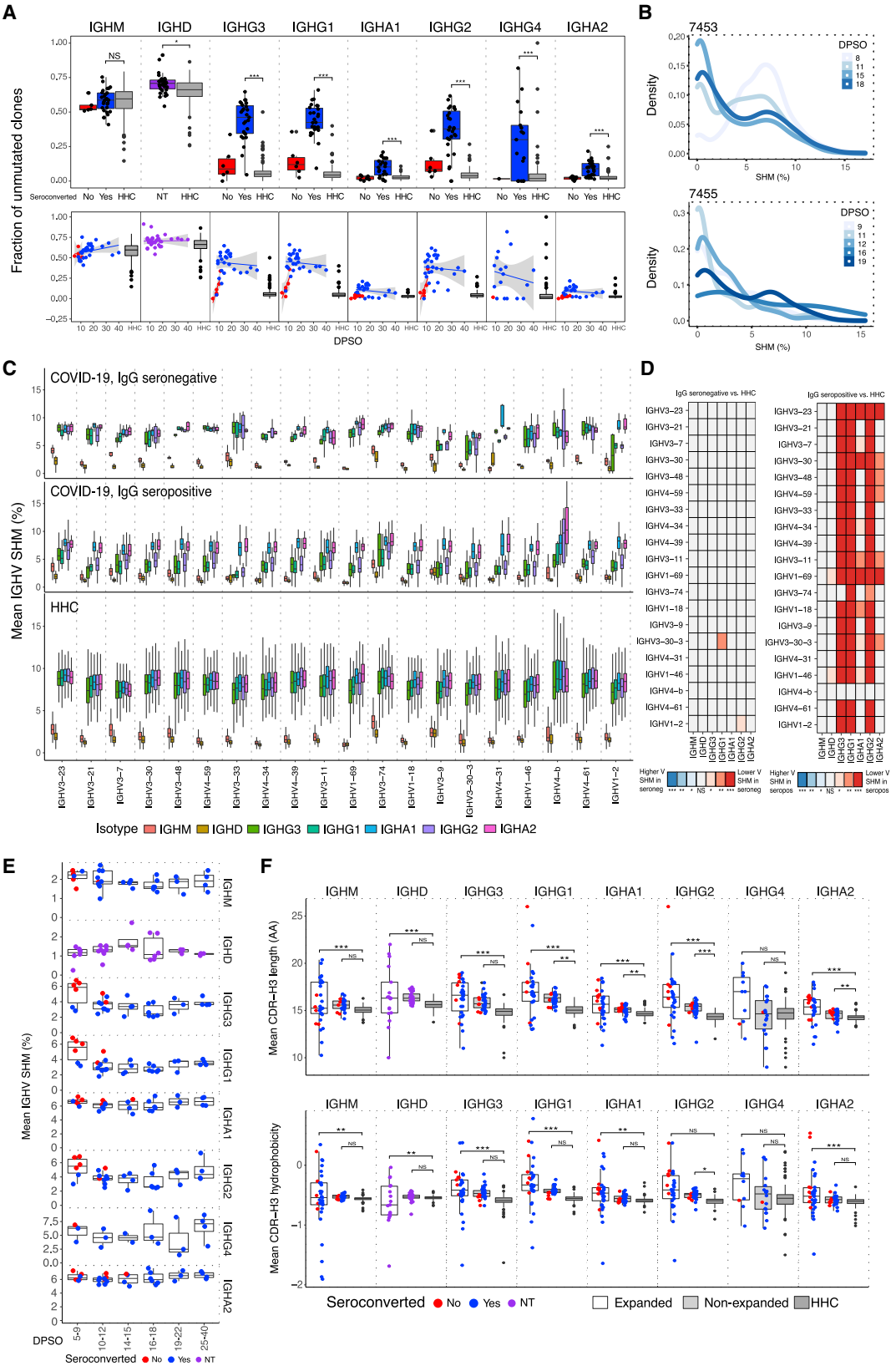


Figure 1. COVID-19 Patient IGH Repertoires Show Early and Extensive Class Switching to IgG and IgA Subclasses without Significant Somatic Mutation

Points indicate B cell clonal lineages. Panel rows are different blood samples (HHC or patient ID), and panel column indicates the clone member's isotype. Within panels, IGHV gene is indicated on the x axis, in the same order and position in each panel, not listed by name due to space constraints, and CDR-H3 length in amino acids (AAs) is indicated on the y axis. Point color indicates median IGHV SHM frequency for each clone, and point size indicates the number of unique reads in the clone. Points are jittered to decrease over-plotting. Patient label colors indicate IgG seroconversion status (red is seronegative, and blue is seropositive) and the number following the patient ID corresponds to days post-symptom onset. The final four rows of panels show the IGH repertoire changes within a single participant (7453) prior to and after seroconversion.



(legend on next page)

convergent clones shared by COVID-19 patients greatly exceeded the mean convergent clone counts from the HHC resampling (Figure 3B), consistent with antigen-driven shared selection of the convergent clones identified in COVID-19 patients. To directly test the antigen specificity of these convergent clones, we expressed human IgG1 monoclonal antibodies (mAbs 2A and 4A; Table S2) from two COVID-19 convergence groups in two patients after paired immunoglobulin heavy and light chain sequencing of single B cells with the 10× Genomics platform. mAb2A had a single nonsynonymous mutation at the beginning of the CDR-H2 of IGHV3-30-3, whereas mAb4A expressed germline IGHV3-15. In enzyme-linked immunosorbent assay (ELISA) testing, both mAbs bound SARS-CoV-2 spike and S1 domain but not the RBD or nucleocapsid (Figure 3C). We further identified 10 sequence clusters shared with independent external COVID-19 patient datasets and reported to be specific for SARS-CoV-2 RBD (Cao et al., 2020; Noy-Porat et al., 2020; Shi et al., 2020). Nine of 13 COVID-19 patients (69%) in our study (7450, 7452, 7454, 7455, 7480, 7483, 7484, 7485, and 7486) had these RBD-specific clonotypes (Figures S4A and S4B), although they make up only a small fraction of each individual's responding BCR repertoire. RBD-specific antibodies are primary candidates for virus-neutralizing, potentially protective antibodies (Ju et al., 2020; Robbiani et al., 2020). With IGH from two COVID-19 patients (7455 and 7480), we expressed four mAbs (CoV.1–CoV.4) from a cluster that is convergent with BD-494 (Yuan et al., 2020), pairing these with the germline-reverted BD-494 light chain and reported CDR-L3. All four mAbs bound SARS-CoV-2 spike protein, S1, and RBD but not nucleocapsid (Figure 3C). BD-494 belongs to a cluster of SARS-CoV-2 RBD-specific antibodies (BD-500, BD-506, and BD-507) using IGHV3-53 and IGHJ6 with CDR-H3 length of 11 amino acid residues (Figure 3D). IGHV3-53 antibodies are reported to contact RBD with several germline-encoded residues in CDR-H1 and CDR-H2, as well as residues in CDR-H3 (Yuan et al., 2020). Contact residues in CDR-H1 and CDR-H2 had the germline IGHV-encoded sequence in CoV.2–CoV.4, whereas CoV.1 had a threonine to isoleucine change in CDR-H1 and an overall IGHV SHM frequency of 4.4% (Figure 3E). CoV.1–CoV.4 each differed from BD-494 by one amino acid residue in CDR-H3.

We hypothesized that in addition to sharing common RBD-binding antibody types, some COVID-19 patients might also demonstrate breadth in their antibody responses and recognize antigens from the distinct but related sarbecovirus, SARS-CoV that was responsible for SARS. Comparison with published SARS patient IGH data (Coughlin et al., 2007; Pinto et al., 2020; Robbiani et al., 2020) revealed convergent SARS-CoV RBD-specific antibodies in two COVID-19 patients, 7453 and 7455 (Figures S4B and S4C). These SARS-CoV-specific convergent IGH sequences were not detected in the 114 HHC samples. To evaluate whether such *in silico* IGH sequence comparisons could predict the serological responses of patients, we tested the plasma samples from COVID-19 patients in SARS-CoV RBD ELISAs and detected cross-reactivity in five of the 13 patients (Figure 3F). Strikingly, the two patients with the highest ELISA OD_{450} values for SARS-CoV RBD were those who had demonstrated convergent IGH sequences specific for SARS-CoV RBD.

DISCUSSION

In these initial months of the COVID-19 pandemic, understanding human antibody responses to SARS-CoV-2 has become a global priority. Our results provide several key findings that could lend some support for vaccine strategies currently under development and suggest that most individuals convalescent from SARS-CoV-2 infection could be, at least for some time, protected against reinfection by commonly elicited RBD-specific antibodies. The IGH repertoires of patients with diagnostically confirmed SARS-CoV-2 infection reveal robust polyclonal responses with early class switching to IgG, and to a lesser extent, IgA isotypes, as well as evidence of accumulating SHM in responding clones within the first month after onset of symptoms, rather than the delayed SHM seen in Ebola patients (Davis et al., 2019). We note that the current COVID-19 study and prior analysis of EBOV infection are among very few published studies of human IGH repertoire longitudinal responses to primary infections; examples from acute dengue virus infection (Appanna et al., 2016; Godoy-Lozano et al., 2016) or H5N6 avian influenza virus (Peng et al., 2019) either had few patients with true primary infection or did not analyze SHM development in responding B cells. Additional follow up of convalescent patients should further clarify the rates of SHM accumulation in SARS-CoV-2-specific BCRs.

Nine of 13 COVID-19 patients (69%) demonstrated convergent IGH specific for the viral RBD, a major target for potentially

Figure 2. IGH Repertoire Signatures of SARS-CoV-2 Infection

(A) Fraction of unmutated (<1% SHM) B cell lineages for each isotype subclass grouped by seroconversion status (top) or plotted by DPSO (bottom). Box colors indicate seronegative (red) or seropositive (blue) for the isotype (IgM serology for IgM; IgG serology for IgG subclasses; and IgA serology for IgA subclasses), with gray for HHC. IgD is indicated in purple as serology was not tested (NT). Points are shown for all COVID-19 samples, and for outliers in the 114 HHC samples.

(B) Distribution of clone SHM percentage plotted as kernel density for clones detected at multiple time points from patients 7453 and 7455. Lines are colored by DPSO. Statistical test: two-sided Wilcoxon-Mann-Whitney.

(C) Mean IGHV SHM percentage for each isotype subclass in IgG seronegative (top), IgG seropositive (middle), or HHC samples (bottom). The 20 most common IGHVs are ordered by frequency in IgM in the patients. Isotypes are plotted by chromosomal ordering. Rare outlier points with extreme values are not shown but were included in all analyses.

(D) Heatmap of IGHV gene SHM for non-seroconverted and seroconverted samples in comparison with HHC. The color scale encodes the significance level and SHM increase (blue) or decrease (red) in COVID-19 in relation with HHC. Statistical test: paired Wilcoxon test with Bonferroni correction for multiple hypothesis testing.

(E) Longitudinal SHM for COVID-19 patients is binned by DPSO and colored by seroconversion status.

(F) Mean CDR-H3 length (top) or hydrophobicity (bottom) for each isotype. COVID-19 patient expanded (white) or non-expanded (light gray) clones, and total clones from HHC (dark gray) are shown. Statistical test: one-way ANOVA with Tukey's HSD test.

Significance cut-offs: NS, not significant (>0.05), *$p \leq 0.05$, **$p < 0.01$, ***$p < 0.001$. See also Figures S1–S3.

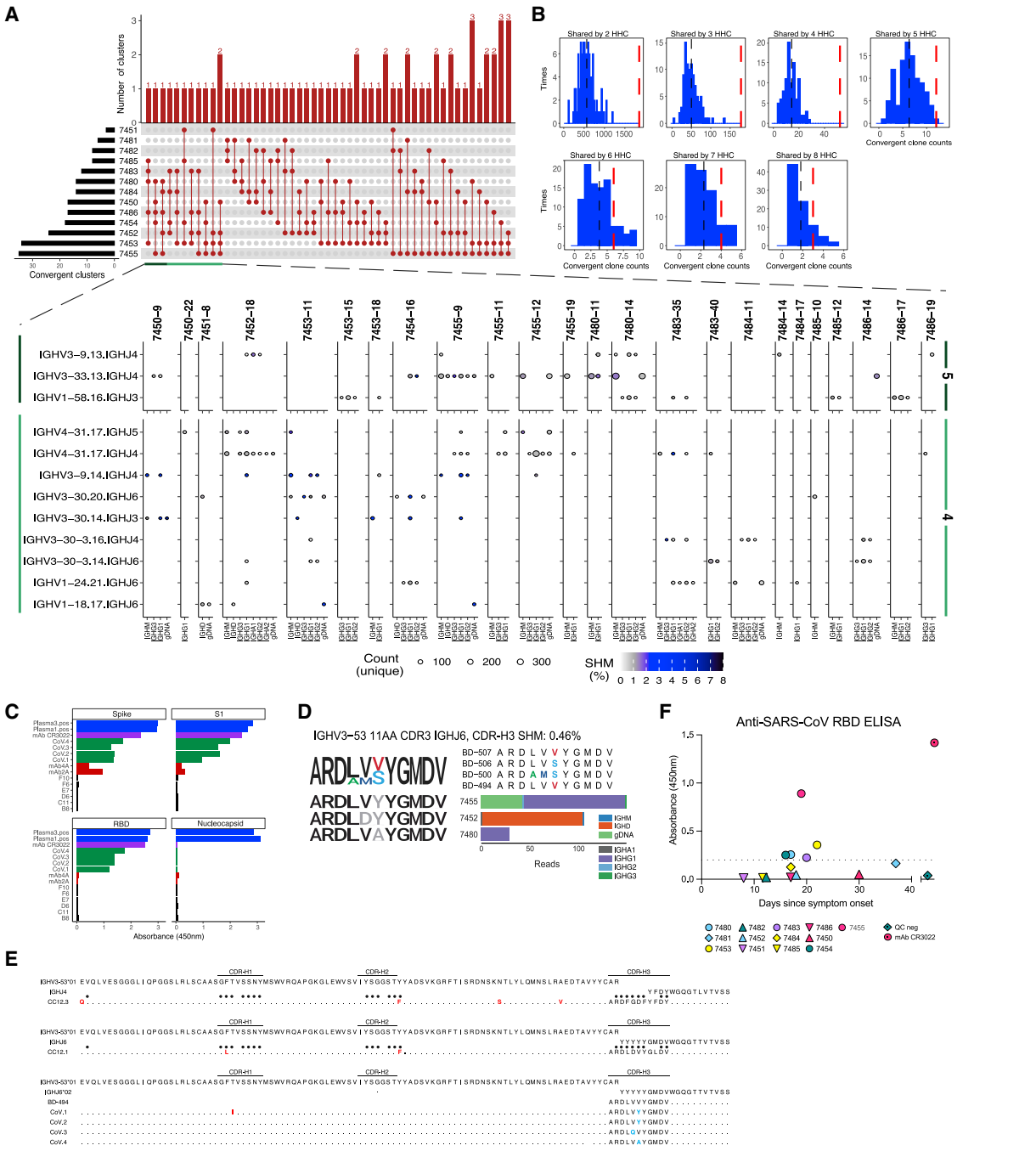


Figure 3. Convergent IGH Sequences among COVID-19 Patients and to Reported Antigen-Specific IGH for SARS-CoV-2 and SARS-CoV

(A) Convergent IGH clusters among patient samples (top panel). The sample distribution is indicated by the lines and dots with the number of clusters sharing that sample distribution indicated by the vertical histogram bars. The total number of convergent clusters identified in each sample is indicated in the histogram to the left of the plot.

(Expanded panel) Lineages belonging to convergent clusters (y axis) shared across four or five patients (light green or dark green brackets, respectively) are plotted by the expressed isotype (x axis). Point color indicates SHM and size shows the number of unique reads.

(B) Distribution of the number of convergent clones shared in two to eight HHC subjects, determined by using 100 resamplings of 13 HHC individuals from the 114 total HHCs. The histogram shows the distribution of convergent clones shared by HHC samples over these trials. The black dashed line is the mean value, and the wide red dashed line is the number of convergent clones shared among the 13 COVID-19 participants.

(C) ELISA results for mAb binding to SARS-CoV-2 spike protein (Spike), S1 domain (S1), RBD, and nucleocapsid. Purified mAbs (mAb2A, mAb4A, B8, C11, D6, E7, F6, F10, mAb CR3022) or HEK293 cell supernatants (CoV.1–CoV.4) were tested. mAbs generated from COVID-19 patient convergent clones (red) or clones convergent with BD-494 (green) were compared to negative control mAbs B8, C11, D6, E7, F6, and F10 (black). The SARS-CoV-2 RBD-binding mAb CR3022

(legend continued on next page)

neutralizing antibodies. SARS-CoV-2 neutralizing serum antibodies are reported to be present in 67%–90% of patients post-infection, depending on the severity of disease, neutralization assay, and threshold for positive results (Robbiani et al., 2020; Suthar et al., 2020; Wu et al., 2020a). It seems reasonable to predict that vaccines based on spike or RBD antigens will also stimulate B cells expressing these common antibody types in a significant fraction of the human population. A subset of patients also had B cell clones expressing convergent IGH to previously described SARS-CoV RBD antibodies; strikingly, the patients with these SARS-CoV-2/SARS-CoV clonotypes also had the highest SARS-CoV RBD binding serum antibody IgG levels. This association suggests that it could become possible to predict the fine specificity of human serological responses from IGH sequence data, as the number of documented antigen-specific clonotypes in public databases increases. This example also highlights the possibility that common modes of human antibody response could enable some breadth of protection or humoral memory against other sarbecoviruses in the future. Longitudinal tracking of IGH repertoires in larger patient cohorts; further investigation into the binding properties, functional activity, and serum antibody levels produced by convergent responding clones in patients; and assessment of clinical outcomes under conditions of exposure to infection will be important next steps toward determining the immunological correlates of protection against SARS-CoV-2 infection.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - COVID-19 Patient Samples
  - Healthy Human Control Samples
  - Cell Lines
- METHOD DETAILS
  - Viral and Serological Testing
  - IGH Library Prep and Sequencing
  - IGH Sequence Analysis
  - Single-Cell Immunoglobulin Analysis
  - IGH Sequence Analysis for mAb Expression
  - mAb Cloning
  - mAb Expression and Purification
  - ELISA Testing of mAbs
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.chom.2020.09.002.

ACKNOWLEDGMENTS

We thank the patients for their contribution to this study. We also thank Hannah K. Frank for initial advice on sequence alignments; Shilpa A. Joshi for editing the paper; Jonasel Roque, Philip Grant, and Aruna Subramanian for assistance with recruiting and consenting subjects; and Aaron J. Wilk, Nancy Q. Zhao, Giovanny J. Martinez-Colon, Julia L. McKechnie, Geoffrey Ivison, Thanmayi Ranganath, Rosemary Vergara, and Laura J. Simpson for processing of samples. We additionally thank the ATUM company for SARS-CoV-2 antigen proteins. This work was supported by NIH/NIAID T32AI007502-23 (A.R.), NIH/NHLBI K23HL125663 (A.J.R.), NIH/NHGRI RM1-HG007735 (H.Y.C.), NIH/NCI K08CA230188 (A.T.S.), Burroughs Wellcome Fund Career Award for Medical Scientists (A.T.S.), Cancer Research Institute Technology Impact Award (A.T.S.), NIH/NIDA DP1DA04608902 (C.A.B.), Burroughs Wellcome Fund Investigators in the Pathogenesis of Infectious Diseases 1016687 (C.A.B.), the Searle Scholars Program (T.T.W.), NIH/NIAID U19AI111825(T.T.W.), NIH/NIAID R01AI139119 (T.T.W.), NIH/NIAID R01AI127877 (S.D.B), NIH/NIAID R01AI130398 (S.D.B.), an endowment to S.D.B. from the Crown Family Foundation, and a gift from an anonymous donor. H.Y.C. is an Investigator of the Howard Hughes Medical Institute. C.A.B. is the Tashia and John Morgridge Faculty Scholar in Pediatric Translational Medicine from the Stanford Maternal Child Health Research Institute.

AUTHOR CONTRIBUTIONS

C.A.B. and S.D.B. conceived the project. C.A.B., K.C.N., A.R., and A.J.R. recruited, enrolled, and consented patients and contributed to clinical sampling and processing and support of clinical sample collection. S.C.A.N., R.A.H., K.R., G.H.J., B.S., J.-Y.L., A.E.P., M.H., J.N., A.R.O.-C., K.E.Y., B.D., and B.A.P. performed the experiments. M.H., H.Y.C., A.T.S., K.C.N., T.S.J., P.S.K., and T.T.W. contributed to serological or 10× assays and analysis. S.C.A.N., F.Y., K.J.L.J., R.A.H., K.R., G.H.J., S.D.B., and B.A.P. contributed to data analysis. S.C.A.N., F.Y., K.J.L.J., R.A.H., K.R., K.E.Y., C.A.B., and S.D.B. wrote the manuscript. All authors edited the manuscript.

DECLARATION OF INTERESTS

A.T.S. is a scientific founder of Immunai and receives research funding from Arsenal Biosciences not related to this study. M.H. is an employee of ATUM. S.D.B. has consulted for Regeneron, Sanofi, and Novartis on topics unrelated to this study. S.D.B., K.R., P.S.K., and A.E.P. have filed provisional patent applications related to serological tests for SARS-CoV-2 antibodies. K.C.N. reports grants from National Institute of Allergy and Infectious Diseases (NIAID), Food Allergy Research & Education (FARE), and End Allergies Together (EAT);

(purple) and pooled SARS-CoV-2 patient plasma (Plasma1.pos and Plasma3.pos) were positive controls. Purified mAbs were tested at 50 μg/mL except for mAb CR3022, which was tested at 0.506 μg/mL.

(D) CDR-H3 AA sequence logos from anti-SARS-CoV-2 convergent IGH. The reported antigen-specific CDR-H3 is shown at the top, and clones from each patient are aligned below (black if matching a conserved residue in the reported CDR-H3, colored for non-conserved, or gray if not matching). Read count per patient, by isotype, is plotted. The SHM frequency is shown after the convergent IGH label.

(E) AA sequence alignment for BD-494-convergent mAbs from (C). SARS-CoV-2-neutralizing mAbs CC12.1 and CC12.3 (Yuan et al., 2020) are shown for reference. Contact residues for CC12.1 and CC12.3 with SARS-CoV-2 RBD are indicated by filled black circles. Residue changes relative to germline IGHV are shown in red. Bases in the CDR-H3 of CoV.1-CoV.4 that differ from the BD-494 CDR-H3 are highlighted in blue. Identical AAs are indicated by small dots.

(F) SARS-CoV RBD IgG ELISA of plasma samples from COVID-19 patients, using the latest sample available for each patient. A pre-pandemic healthy blood donor sample pool was used as a negative quality control (QC), and mAb CR3022 was a positive control for SARS-CoV RBD binding (ter Meulen et al., 2006). The dotted line denotes the cut-off value for seroconversion. Assays were performed in duplicate and mean OD values are shown.

See also Figure S4 and Table S2.

National Heart, Lung, and Blood Institute (NHLBI), and National Institute of Environmental Health Sciences (NIEHS), and is the director of FARE and World Allergy Organization (WAO) Center of Excellence at Stanford; advisor at Cour Pharma; co-founder of Before Brands, Alladapt, Latitude, and IgGenix; National Scientific Committee member at Immune Tolerance Network (ITN) and National Institutes of Health (NIH); a recipient of a Research Sponsorship from Nestle; Consultant and Advisory Board Member at Before Brands, Alladapt, Iggenix, NHLBI, and Probio; Data and Safety Monitoring Board member at NHLBI; and has US patents for basophil testing, multifood immunotherapy and prevention, monoclonal antibodies from plasmablasts, and devices for diagnostics. The remaining authors declare that they have no competing interests.

Received: July 8, 2020
Revised: August 13, 2020
Accepted: August 27, 2020
Published: September 3, 2020

REFERENCES

Appanna, R., Kg, S., Xu, M.H., Toh, Y.X., Velumani, S., Carbajo, D., Lee, C.Y., Zuest, R., Balakrishnan, T., Xu, W., et al. (2016). Plasmablasts During Acute Dengue Infection Represent a Small Subset of a Broader Virus-specific Memory B Cell Pool. EBioMedicine *12*, 178–188.

Avnir, Y., Watson, C.T., Glanville, J., Peterson, E.C., Tallarico, A.S., Bennett, A.S., Qin, K., Fu, Y., Huang, C.Y., Beigel, J.H., et al. (2016). IGHV1-69 polymorphism modulates anti-influenza antibody repertoires, correlates with IGHV utilization shifts and varies by ethnicity. Sci. Rep. *6*, 20842.

Cao, Y., Su, B., Guo, X., Sun, W., Deng, Y., Bao, L., Zhu, Q., Zhang, X., Zheng, Y., Geng, C., et al. (2020). Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-Throughput Single-Cell Sequencing of Convalescent Patients' B Cells. Cell *182*, 73–84.e16.

Corman, V.M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D.K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M.L., et al. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Euro Surveill. *25*, 2000045.

Coughlin, M., Lou, G., Martinez, O., Masterman, S.K., Olsen, O.A., Moksa, A.A., Farzan, M., Babcook, J.S., and Prabhakar, B.S. (2007). Generation and characterization of human monoclonal neutralizing antibodies with distinct binding and sequence features against SARS coronavirus using XenoMouse. Virology *361*, 93–102.

Davis, C.W., Jackson, K.J.L., McElroy, A.K., Halfmann, P., Huang, J., Chennareddy, C., Piper, A.E., Leung, Y., Albariño, C.G., Crozier, I., et al. (2019). Longitudinal Analysis of the Human B Cell Response to Ebola Virus Infection. Cell *177*, 1566–1582.e17, e1517.

Durocher, Y., Perret, S., and Kamen, A. (2002). High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. Nucleic Acids Res. *30*, E9.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics *28*, 3150–3152.

Godoy-Lozano, E.E., Téllez-Sosa, J., Sánchez-González, G., Sámano-Sánchez, H., Aguilar-Salgado, A., Salinas-Rodríguez, A., Cortina-Ceballos, B., Vivanco-Cid, H., Hernández-Flores, K., Pfaff, J.M., et al. (2016). Lower IgG somatic hypermutation rates during acute dengue virus infection is compatible with a germinal center-independent B cell response. Genome Med. *8*, 23.

Grimsholm, O., Piano Mortari, E., Davydov, A.N., Shugay, M., Obraztsova, A.S., Bocci, C., Marasco, E., Marcellini, V., Aranburu, A., Farroni, C., et al. (2020). The Interplay between $CD27^{dull}$ and $CD27^{bright}$ B Cells Ensures the Flexibility, Stability, and Resilience of Human B Cell Memory. Cell Rep. *30*, 2963–2977.e6.

Hogan, C.A., Sahoo, M.K., and Pinsky, B.A. (2020). Sample Pooling as a Strategy to Detect Community Transmission of SARS-CoV-2. JAMA *323*, 1967–1969.

Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet *395*, 497–506.

Jackson, K.J.L., Liu, Y., Roskin, K.M., Glanville, J., Hoh, R.A., Seo, K., Marshall, E.L., Gurley, T.C., Moody, M.A., Haynes, B.F., et al. (2014). Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. Cell Host Microbe *16*, 105–114.

Ju, B., Zhang, Q., Ge, J., Wang, R., Sun, J., Ge, X., Yu, J., Shan, S., Zhou, B., Song, S., et al. (2020). Human neutralizing antibodies elicited by SARS-CoV-2 infection. Nature *584*, 115–119.

Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics *27*, 2957–2963.

Nielsen, S.C.A., Roskin, K.M., Jackson, K.J.L., Joshi, S.A., Nejad, P., Lee, J.Y., Wagar, L.E., Pham, T.D., Hoh, R.A., Nguyen, K.D., et al. (2019). Shaping of infant B cell receptor repertoires by environmental factors and infectious disease. Sci. Transl. Med. *11*, eaat2004.

Noy-Porat, T., Makdasi, E., Alcalay, R., Mechaly, A., Levi, Y., Bercovich-Kinori, A., Zauberman, A., Tamir, H., Yahalom-Ronen, Y., Israeli, M.A., et al. (2020). Tiger team: a panel of human neutralizing mAbs targeting SARS-CoV-2 spike at multiple epitopes. bioRxiv. Online (Bergh.), 20–2020, https://doi.org/10.1101/2020.05.20.106609.

Parameswaran, P., Liu, Y., Roskin, K.M., Jackson, K.K., Dixit, V.P., Lee, J.Y., Artiles, K.L., Zompi, S., Vargas, M.J., Simen, B.B., et al. (2013). Convergent antibody signatures in human dengue. Cell Host Microbe *13*, 691–700.

Peiris, J.S.M., Lai, S.T., Poon, L.L.M., Guan, Y., Yam, L.Y.C., Lim, W., Nicholls, J., Yee, W.K.S., Yan, W.W., Cheung, M.T., et al.; SARS study group (2003). Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet *361*, 1319–1325.

Peng, W., Liu, S., Meng, J., Huang, J., Huang, J., Tang, D., and Dai, Y. (2019). Profiling the TRB and IGH repertoire of patients with H5N6 Avian Influenza Virus Infection by high-throughput sequencing. Sci. Rep. *9*, 7429.

Pinto, D., Park, Y.J., Beltramello, M., Walls, A.C., Tortorici, M.A., Bianchi, S., Jaconi, S., Culap, K., Zatta, F., De Marco, A., et al. (2020). Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. Nature *583*, 290–295.

Robbiani, D.F., Gaebler, C., Muecksch, F., Lorenzi, J.C.C., Wang, Z., Cho, A., Agudelo, M., Barnes, C.O., Gazumyan, A., Finkin, S., et al. (2020). Convergent antibody responses to SARS-CoV-2 in convalescent individuals. Nature *584*, 437–442, https://doi.org/10.1038/s41586-020- 2456-9.

Shi, R., Shan, C., Duan, X., Chen, Z., Liu, P., Song, J., Song, T., Bi, X., Han, C., Wu, L., et al. (2020). A human neutralizing antibody targets the receptor-binding site of SARS-CoV-2. Nature *584*, 120–124.

Stadlbauer, D., Amanat, F., Chromikova, V., Jiang, K., Strohmeier, S., Arunkumar, G.A., Tan, J., Bhavsar, D., Capuano, C., Kirkpatrick, E., et al. (2020). SARS-CoV-2 Seroconversion in Humans: A Detailed Protocol for a Serological Assay, Antigen Production, and Test Setup. Curr. Protoc. Microbiol. *57*, e100.

Su, S., Wong, G., Shi, W., Liu, J., Lai, A.C.K., Zhou, J., Liu, W., Bi, Y., and Gao, G.F. (2016). Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. Trends Microbiol. *24*, 490–502.

Suthar, M.S., Zimmerman, M.G., Kauffman, R.C., Mantus, G., Linderman, S.L., Hudson, W.H., Vanderheiden, A., Nyhoff, L., Davis, C.W., Adekunle, O., et al. (2020). Rapid generation of neutralizing antibody responses in COVID-19 patients. Cell Rep Med *1*, 100040.

Team, R.C. (2017). R: A Language and Environment for Statistical Computing. https://cran.r-project.org/doc/manuals/fullrefman.pdf.

ter Meulen, J., van den Brink, E.N., Poon, L.L.M., Marissen, W.E., Leung, C.S.W., Cox, F., Cheung, C.Y., Bakker, A.Q., Bogaards, J.A., van Deventer, E., et al. (2006). Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. PLoS Med. *3*, e237.

van Dongen, J.J., Langerak, A.W., Brüggemann, M., Evans, P.A., Hummel, M., Lavender, F.L., Delabesse, E., Davi, F., Schuuring, E., García-Sanz, R., et al. (2003). 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 *17*, 2257–2317.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis. (Springer-Verlag New York).

Wu, F., Wang, A., Liu, M., Wang, Q., Chen, J., Xia, S., Ling, Y., Zhang, Y., Xun, J., Lu, L., et al. (2020a). Neutralizing antibody responses to SARS-CoV-2 in a COVID-19 recovered patient cohort and their implications. medRxiv 2020.2003.2030.20047365.

Wu, Y., Wang, F., Shen, C., Peng, W., Li, D., Zhao, C., Li, Z., Li, S., Bi, Y., Yang, Y., et al. (2020b). A noncompeting pair of human neutralizing antibodies block COVID-19 virus binding to its receptor ACE2. Science *368*, 1274–1278.

Ye, J., Ma, N., Madden, T.L., and Ostell, J.M. (2013). IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res. *41*, W34-40.

Yuan, M., Liu, H., Wu, N.C., Lee, C.D., Zhu, X., Zhao, F., Huang, D., Yu, W., Hua, Y., Tien, H., et al. (2020). Structural basis of a shared antibody response to SARS-CoV-2. Science *369*, 1119–1123.

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D., and Fouchier, R.A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N. Engl. J. Med. *367*, 1814–1820.

Zhou, J.Q., and Kleinstein, S.H. (2019). Cutting Edge: Ig H Chains Are Sufficient to Determine Most B Cell Clonal Relationships. J. Immunol. *203*, 1687–1692.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., et al.; China Novel Coronavirus Investigating and Research Team (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. N. Engl. J. Med. *382*, 727–733.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat anti-Human IgM-HRP	Sigma	Cat#A6907; lot#SLCD6915; RRID: AB_258318
Goat anti-Human IgG-HRP	Thermo Fisher	Cat#62-8420; lot#VC298928; RRID: AB_88136
Rabbit anti-Human IgA-HRP	Dako	Cat#P0216
Rabbit anti-Human IgG-HRP	Agilent	Cat#P0214
mAb CR3022	Expressed for this paper Original: (ter Meulen et al., 2006)	N/A
mAb2A	This paper	N/A
mAb4A	This paper	N/A
CoV.1	This paper	N/A
CoV.2	This paper	N/A
CoV.3	This paper	N/A
CoV.4	This paper	N/A
Biological Samples		
COVID-19 patient samples	Stanford Healthcare	N/A
Healthy adult control samples	Stanford Blood Center	N/A
Chemicals, Peptides, and Recombinant Proteins		
SARS-CoV RBD protein	This paper	N/A
SARS-CoV-2 RBD protein	ATUM	Cat#CoV2-RBD-His
SARS-CoV-2 spike ectodomain protein	ATUM	Special request
SARS-CoV-2 spike S1 protein	ATUM	Cat#CoV2-S1-Ctag
SARS-CoV-2 nucleocapsid	ATUM	Special request
Critical Commercial Assays		
AllPrep DNA/RNA kit	QIAGEN	Cat#80204
MiSeq Reagent Kit v3 for isotype library sequencing	Illumina	Cat#MS-102-3003
MiSeq Reagent Kit v2 for gDNA library sequencing	Illumina	Cat#MS-102-2003
MiSeq Reagent Kit v3 for 10x library sequencing	Illumina	Cat#MS-102-3001
10x Single Cell Immune Profiling Solution Kit (v1.1)	10x Genomics	Cat#1000127; Cat#1000167; Cat#1000016; Cat#1000213
HiTrap Protein A HP columns for antibody purification	GE healthcare	Cat#29-0485-76
Deposited Data		
COVID-19 IGH repertoire data (Illumina NGS) 10x Genomics generated reads	This paper	BioProject: PRJNA628125
Healthy adult control IGH repertoire data (Illumina NGS)	Nielsen et al., 2019	BioProject: PRJNA491287
Experimental Models: Cell Lines		
HEK293-EBNA1-6E	National Research Council, Canada	RRID: CVCL_HF20
Expi293F	Thermo Fisher	Cat#A14528; RRID: CVCL_D615
Oligonucleotides		
Primers for isotype/gDNA libraries (PCR1), see Table S3	Integrated DNA Technologies	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Primer for isotype/gDNA libraries (PCR2) P5: 5'-AATGATACGGCGACCACCGAGA TCTACACTCTTTCCCTACACGACGCTCT TCCGATC	Integrated DNA Technologies	N/A
Primer for isotype/gDNA libraries (PCR2) P7: 5'- CAAGCAGAAGACGGCATACGA GATCGGTCTCGGCATTCCTGCTGAAC CGCTC	Integrated DNA Technologies	N/A
Recombinant DNA		
pTT5 vector	National Research Council, Canada (Durocher et al., 2002)	Special request
gBlocks with full-length V(D)J regions	Integrated DNA Technologies	N/A
Software and Algorithms		
IgBLAST	Ye et al., 2013	https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
FLASH	Magoč and Salzberg, 2011	https://ccb.jhu.edu/software/FLASH/
Single-linkage clustering	SciPy	https://docs.scipy.org/doc/scipy/reference/generated/scipy.cluster.hierarchy.linkage.html
CD-HIT	Fu et al., 2012	http://cd-hit.org
Prism8	GraphPad Software	https://www.graphpad.com/scientific-software/prism/
R	Team, 2017	https://www.r-project.org
Cell Ranger for V(D)J immune profiling (v.3.1.0)	10x Genomics	https://support.10xgenomics.com/single-cell-vdj/software/pipelines/latest/what-is-cell-ranger

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to the Lead Contact, Dr. Scott D. Boyd (sboyd1@stanford.edu).

Materials Availability

Unique reagents generated in this study will be provided upon reasonable request.

Data and Code Availability

All data is available in the main text or the extended materials. Code will be provided to readers upon request. The accession number for the 10x Genomics data used for monoclonal antibody expression and the IGH repertoire data reported in this paper is BioProject: PRJNA628125. The IGH repertoire data for the healthy human control (HHC) dataset is in BioProject:PRJNA491287.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

COVID-19 Patient Samples

Patients admitted to Stanford Hospital with signs and symptoms of COVID-19 and confirmed SARS-CoV-2 infection by reverse-transcription quantitative polymerase chain reaction (RT-qPCR) of nasopharyngeal swabs were recruited. Demographic and clinical data for patients are provided in Table 1. Venipuncture blood samples were collected in dipotassium ethylenediaminetetraacetic acid (K_2EDTA)- or sodium heparin-coated vacutainers for peripheral blood mononuclear cell (PBMC) isolation or serology on plasma, respectively. Recruitment of COVID-19 patients, documentation of informed consent, collections of blood samples, and experimental measurements were carried out with Institutional Review Board approval (IRB-55689 and IRB-48973) at Stanford University.

Healthy Human Control Samples

The dataset containing healthy adult control immunoglobulin heavy chain (IGH) repertoires has been described previously (Nielsen et al., 2019). In summary, healthy adults with no signs or symptoms of acute illness or disease were recruited as volunteer blood donors at the Stanford Blood Center. Pathogen diagnostics were performed for cytomegalovirus, human immunodeficiency virus,

hepatitis C and B virus, West Nile virus, human T cell leukemia virus, TPPA (Syphilis), and *Trypanosoma cruzi*. Volunteer age range was 17-87 with median and mean of 52 and 49, respectively.

Cell Lines

HEK293-EBNA1-6E cells were from National Research Council (NRC), Canada; RRID:CVCL_HF20). Expi293F cells were from Thermo Fisher (cat. A14528; RRID:CVCL_D615). Cell lines were cultured using manufacturer's guidelines and used as described in Method Details below.

METHOD DETAILS

Viral and Serological Testing

SARS-CoV-2 infection in patients was confirmed by RT-qPCR testing of nasopharyngeal swab specimens, using the protocols described in (Corman et al., 2020; Hogan et al., 2020). An enzyme-linked immunosorbent assay (ELISA) based on a protocol described in (Stadlbauer et al., 2020) was performed to detect anti-SARS-CoV and anti-SARS-CoV-2 spike receptor-binding domain (RBD) antibodies in plasma samples from COVID-19 patients. Briefly, 96-well high binding plates (Thermo Fisher) were coated with either SARS-CoV or SARS-CoV-2 spike RBD protein (0.1 μg per well) overnight at 4°C. After blocking plates with 3% non-fat milk in phosphate-buffered saline (PBS) containing 0.1% Tween 20 (PBST), plasma samples were incubated at a dilution of 1:100 and bound antibodies were detected with goat anti-human IgM/horseradish peroxidase (HRP) (Sigma: cat. A6907, 1:6,000 dilution), goat anti-human IgG/HRP (Thermo Fisher: cat. 62-8420, 1:6,000 dilution), or rabbit anti-human IgA/HRP (Dako: cat. P0216, 1:5,000 dilution). Assays were developed by addition of 3,3',5,5'-tetramethylbenzidine (TMB) substrate solution. After stopping the reaction with 0.16 M sulfuric acid, the optical density (OD) at 450 nanometers was read using an EMax Plus microplate reader (Molecular Devices). The cut-off value for seroconversion was calculated as OD_{450} = 0.2 for the anti-SARS-CoV IgG assay and as OD_{450} = 0.3 for anti-SARS-CoV-2 IgM, IgG, and IgA assays after analyzing SARS-CoV-2 pre-pandemic negative control samples from healthy blood donors.

IGH Library Prep and Sequencing

The AllPrep DNA/RNA kit (QIAGEN) was used to extract genomic DNA (gDNA) and total RNA from PBMCs. For each blood sample, six independent gDNA library PCRs were set up using up to100 ng template/library. Multiplexed primers to IGHJ and the IGHV FR1 or FR2 framework regions (3 FR1 and 3 FR2 libraries), per the BIOMED-2 design were used (van Dongen et al., 2003) with additional sequence representing the first part of the Illumina linkers (Table S3). In addition, for each sample, total RNA was reverse-transcribed to cDNA using Superscript III RT (Invitrogen) with random hexamer primers (Promega). Total RNA yield varied between patients and between 6 ng-100 ng was used for each of the isotype PCRs using IGHV FR1 primers based on the BIOMED-2 design (van Dongen et al., 2003) and isotype specific primers located in the first exon of the constant region for each isotype category (IgM, IgD, IgE, IgA, IgG). Primers contain additional sequence representing the first part of the Illumina linkers (Table S3). The different isotypes were amplified in separate reaction tubes. Eight-nucleotide barcode sequences were included in the primers to indicate sample (isotype and gDNA libraries) and replicate identity (gDNA libraries). Four randomized bases were included upstream of the barcodes on the IGHJ primer (gDNA libraries) and constant region primer (isotype libraries) for Illumina clustering (Table S3). PCR was carried out with AmpliTaq Gold (Applied Biosystems) following the manufacturer's instructions, and used a program of: 95°C 7 min; 35 cycles of 94°C 30 s, 58°C 45 s, 72°C 60 s; and final extension at 72°C for 10 min. A second round of PCR using QIAGEN's Multiplex PCR Kit was performed to complete the Illumina sequencing adapters at the 5-prime and 3-prime ends of amplicons; cycling conditions were: 95°C 15 min; 12 cycles of 95°C 30 s, 60°C 45 s, 72°C 60 s; and final extension at 72°C for 10 min. Products were subsequently pooled, gel purified (QIAGEN), and quantified with the Qubit fluorometer (Invitrogen). Samples were sequenced on the Illumina MiSeq using 600 cycle kits for isotype libraries (PE300) and 500 cycle kits for gDNA libraries (PE250).

IGH Sequence Analysis

Paired-end reads were merged using Fast Length Adjustment of Short reads software (Magoč and Salzberg, 2011), demultiplexed (100% barcode match), and primer trimmed. The IGHV, IGHD, and IGHJ gene segments and IGHV-IGHD (N1), and IGHD-IGHJ (N2) junctions were identified using the IgBLAST alignment program (Ye et al., 2013). Quality filtering of sequences included keeping only productive reads with a CDR-H3 region, and minimum IGHV-gene alignment score of 200. Sample cDNA or gDNA libraries with poor read coverage were excluded from further analysis (Table 1). For cDNA-templated IGH reads, isotypes and subclasses were called by exact matching to the constant region gene sequence upstream of the primer. Clonal identities within each subject were inferred using single-linkage clustering and the following definition: same IGHV and IGHJ usage (disregarding allele call), equal CDR-H3 length, and minimum 90% CDR-H3 nucleotide identity. A total of 1,259,882 clones (per sample, mean number of clones: 33,154; median number of clones: 18,503) were identified. A total of 24,888,790 IGH sequences amplified from cDNA were analyzed for the COVID-19 subjects (mean: 754,205 per sample; median: 650,812) and 68,831,446 sequences from healthy adult controls (mean: 603,785 per individual; median: 637,269). Each COVID-19 patient had on average 372,304 in-frame gDNA sequences and each adult control had an average of 8,402 in-frame gDNA sequences.

For each clone, the median somatic mutation frequency of reads was calculated. Mean mutation frequencies for all clonal lineages from a sample for each isotype were calculated from the median mutation frequency within each clone, and so represent the mean of

the median values. Clones with < 1% mutation were defined as unmutated and clones with greater or equal to1% mutation were defined as being mutated. Subclass fractions were determined for each subject by dividing the number of clones for a given subclass by the total number of clones for that isotype category. Expanded clones within each sample were defined as clones that were present in two or more of the gDNA replicate libraries. Clonal expansion in the isotype data was inferred from the gDNA data. Analyses were conducted in R (Team, 2017) using base packages for statistical analysis and the ggplot2 package for graphics (Wickham, 2016).

To identify convergent rearranged IGH among patients with SARS-CoV-2 infection, heavy-chain sequences annotated with the same IGHV and IGHJ segment (not considering alleles) and the same CDR-H3 length were clustered based on 85% CDR-H3 amino acid sequence similarity using CD-HIT (Fu et al., 2012). To exclude IGH that are generally shared between humans and to enrich the SARS-CoV-2-specific IGH that are likely shared among the patients, clusters were selected as informative if (1) they contained at least five IGH sequences from each COVID-19 patient and were present in at least two subjects; (2) no IGH sequences from HHC samples (collected prior to the 2019 SARS-CoV-2 outbreak) were identified in the same convergent cluster. The same selection criteria were used to determine the convergent clusters between the COVID-19 samples and previously reported IGH sequences specific to SARS-CoV and SARS-CoV-2. Convergent IGH sequences between the deeply sequenced COVID-19 patients and 10x Genomics single B cell immune profiling on two COVID-19 patients were selected for mAb expression. To assess whether the number of convergent clones identified among COVID-19 patients was significantly greater than the number expected by chance in HHC repertoires, we randomly selected with replacement sets of 13 HHC individuals from the 114 available, determining the number of convergent clones detected in two, three, four or larger numbers of participants, and repeated this process 100 times to generate the distribution of convergent clone counts shown in Figure 3B.

Single-Cell Immunoglobulin Analysis

Single-cell immunoglobulin (Ig) libraries were prepared using the 10x Single Cell Immune Profiling Solution Kit (v1.1 Chemistry), according to the manufacturer's instructions. Briefly, cells were washed once with PBS + 0.04% BSA. Following reverse transcription and cell barcoding in droplets, emulsions were broken and cDNA purified using Dynabeads MyOne SILANE (Thermo Fisher) followed by PCR amplification (98°C for 45 s; 15 cycles of 98°C for 20 s, 67°C for 30 s, 72°C for 1 min; 72°C for 1 min). For targeted Ig library construction, 2 μL of amplified cDNA was used for target enrichment by PCR (Human B cell primer sets 1 and 2: 98°C for 45 s; 6 and 8 cycles of 98°C for 20 s, 67°C for 30 s, 72°C for 1 min; 72°C for 1 min). Following Ig enrichment, up to 50 ng of enriched PCR product was fragmented and end-repaired, size selected with SPRIselect beads (Beckman Coulter), PCR amplified with sample indexing primers (98°C for 45 s; 9 cycles of 98°C for 20 s, 54°C for 30 s, 72°C for 20 s; 72°C for 1 min), and size selected with SPRIselect beads. Targeted single-cell Ig libraries were sequenced on an Illumina MiSeq to a minimum sequencing depth of 5,000 reads/cell using the read lengths 26bp Read1, 8bp i7 Index, 91bp Read2 and reads were aligned to the prebuilt GRCh38/Ensembl/10x reference dataset and consensus Ig annotation was performed using Cell Ranger (10x Genomics, version 3.1.0).

IGH Sequence Analysis for mAb Expression

IGH sequences from single cells with paired productive heavy and light chains were searched against COVID-19 patient bulk IGH repertoires to identify convergent sequences according to the following criteria: utilization of the same IGHV and IGHJ genes; same CDR-H3 lengths; and CDR-H3 amino acid sequences that were within a Hamming distance cutoff of 15% of the length of the CDR-H3. Two native heavy and light chain pairs, designated mAb2A and mAb4A, which were found in convergent clusters characterized by low- to mid-SHM frequencies and included at least one class-switched member, were selected for cloning and expression. Additionally, four IGHs were selected from clones that clustered with SARS-CoV-2 RBD-specific antibody BD-494. Full-length light chain sequence for the BD-494 was not available so to produce mAbs for CoV.1-CoV.4, the four IGH were paired with a germline-reverted light chain that utilizes IGKV1-9, IGKJ3, and the published CDR-L3 sequence for BF-494, 'QQLNSYPFT'.

mAb Cloning

Heavy and light chain sequences were synthesized by Integrated DNA Technologies as gBlocks encoding full-length heavy and light chain V(D)J regions. gBlocks were resuspended at 50 ng/μL and amplified with AmpliTaq Gold (Applied Biosystems) following the manufacturer's instructions, using a program of: 95°C 7 min; 30 cycles of 94°C for 30 s, 55°C for 45 s, 72°C for 60 s; and final extension at 72°C for 10 min. Products were gel purified (QIAGEN) and cloned as in-frame fusions to human IgG1, IgK or IgL constant regions into the pTT5 vector (Durocher et al., 2002) (National Research Council (NRC), Canada) using Gibson Assembly Master Mix (New England Biolabs) for 45 min at 50°C. Assembled constructs were verified by Sanger sequencing.

mAb Expression and Purification

Constructs were transiently transfected in HEK293-EBNA1-6E cells (NRC) or Expi293F cells (Thermo Fisher) at a density of 1.2-1.6 million cells/mL or 3 million cells/mL, respectively, using 25 kDa linear polyethylenimine (PEI) at a 3:1 PEI:DNA ratio in OptiMEM reduced serum medium (GIBCO), with a heavy chain: light chain ratio of 1:1. HEK293-EBNA1-6E cells were maintained in Freestyle 293 Expression Medium (GIBCO) and were supplemented with 0.5% tryptone 24-36 h after transfection. Expi293F cells were maintained in Expi293 Expression Medium (Thermo Fisher). Cell supernatants were harvested after 96 h and filtered through 0.45-μM filters (Millipore). Antibodies were purified via HiTrap Protein A HP columns (GE Healthcare) run at a flow rate of 0.5-1 mL/min on Äkta Start protein purification system (GE Healthcare). Antibodies were eluted using 0.1M glycine pH 2, dialyzed with 3 changes

of PBS pH 7.4 using Slide-A-Lyzer-G2 10K dialysis cassettes (Thermo Fisher), and concentrated using 30,000 kDa molecular weight cutoff polyethersulfone membrane spin columns (Pierce). Final concentrations of purified antibodies were quantified with Nanodrop 2000 (Thermo Fisher).

ELISA Testing of mAbs

ELISA conditions for mAbs were as described for COVID-19 plasma samples with the following modifications. Purified mAbs were tested at 50 μg/mL for intra-COVID-19 convergent antibodies or peanut-specific negative mAb controls, or at 0.506 μg/mL for mAb CR3022; antibody supernatants were diluted 1:1 with PBST plus 1% milk; plates were coated overnight with RBD (0.1 μg per well), S1 (0.5 μg per well), spike protein (0.1 μg per well), or nucleocapsid (0.1 μg per well); and bound mAbs were detected with rabbit anti-human IgG gamma chain-specific/HRP (Agilent: cat. P0214, 1:15,000 dilution).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical tests were performed in R (Team, 2017) using base packages for statistical analysis and the ggplot2 package was used for graphics (Wickham, 2016). Box-whisker plots show median (horizontal line), interquartile range (box), and 1.5 times the interquartile range (whiskers). Linear regression lines were plotted with 95% confidence intervals calculated using the function stat_smooth. In all analyses where statistical significance was tested, significance was defined as: ***p value < 0.001; **p value < 0.01; *p value < 0.05; not significant (NS): p value > 0.05. Differences between the seropositive group and HHC (Figure 2A) was tested using two-sided Wilcoxon–Mann-Whitney (for patients with more than one sample, the mean value of these was used). Patient IGHV gene SHM for seroconverted and non-seroconverted samples were compared to HHC (Figure 2D) using paired Wilcoxon tests with Bonferroni correction for multiple hypothesis testing. Differences between the expanded/non-expanded groups and HHC (Figure 2F) were tested using one-way ANOVA with Tukey's HSD test.

Cell Host & Microbe, Volume *28*

Supplemental Information

Human B Cell Clonal Expansion and Convergent Antibody Responses to SARS-CoV-2

Sandra C.A. Nielsen, Fan Yang, Katherine J.L. Jackson, Ramona A. Hoh, Katharina Röltgen, Grace H. Jean, Bryan A. Stevens, Ji-Yeun Lee, Arjun Rustagi, Angela J. Rogers, Abigail E. Powell, Molly Hunter, Javaria Najeeb, Ana R. Otrelo-Cardoso, Kathryn E. Yost, Bence Daniel, Kari C. Nadeau, Howard Y. Chang, Ansuman T. Satpathy, Theodore S. Jardetzky, Peter S. Kim, Taia T. Wang, Benjamin A. Pinsky, Catherine A. Blish, and Scott D. Boyd

Supplemental Information

Table S1. IgG and IgA subclass proportions, Related to Figure 2. COVID-19 and healthy human control (HHC) median isotype subclass proportions +/- median absolute deviation (MAD) summarized for all samples within each group. p-values were calculated by two-sided Wilcoxon–Mann-Whitney tests.

	COVID-19 (median +/- MAD)	**HHC** (median +/- MAD)	**Two-sided Wilcoxon–Mann-Whitney test**
IGHA1	70.1 +/- 9.3	65 +/- 6.6	p-value = 0.2018
IGHA2	29.9 +/- 9.4	35 +/- 6.6	p-value = 0.2018
IGHG1	69.2 +/- 11.5	41.4 +/- 7.3	p-value = 9.114e-09
IGHG2	16.3 +/- 11.2	40.2 +/- 8.5	p-value = 8.855e-08
IGHG3	8.7 +/- 3.3	14.6 +/- 5.9	p-value = 6.281e-05
IGHG4	0.3 +/- 0.2	1.6 +/- 1.9	p-value = 0.003414

Table S2. Heavy and light chain sequence features for convergent monoclonal antibodies (mAbs) 2A and 4A, Related to Figure 3. The percent identity (ID %) of the IGHV, IGKV or IGLV gene sequence relative to germline is shown.

mAb	IGHV	IGHD	IGHJ	IGHC	IGHV ID (%)	CDR-H3 AA	IGK/LV	IGK/LJ	IGK/LV ID (%)	CDR-L3 AA
mAb2A	IGHV3-30-3*01	IGHD3-22*01	IGHJ3*02	IGHG1	99.7	ARDSGSAFDI	IGLV3-1*01	IGLJ2*01, IGLJ3*01	100	QAWDSSTVV
mAb4A	IGHV3-15*01	IGHD3-16*02	IGHJ4*02	IGHM	100	TTDRHYDYVWGSYRYPDY	IGKV1-39*01	IGKJ5*01	100	QQSYSTPT

Table S3. First round PCR primer sequences for isotype and gDNA libraries for Illumina high-throughput sequencing, Related to STAR METHODS.

Primer name	Sequence (5'-3')
Isotype/gDNA FR1 primers	
IGHV1	GGCATTCCTGCTGAACCGCTCTTCCGATCT[8nt barcode]GGCCTCAGTGAAGGTCTCCTGCAAG
IGHV2	GGCATTCCTGCTGAACCGCTCTTCCGATCT[8nt barcode]GTCTGGTCCTACGCTGGTGAAACCC
IGHV3	GGCATTCCTGCTGAACCGCTCTTCCGATCT[8nt barcode]CTGGGGGGTCCCTGAGACTCTCCTG
IGHV4	GGCATTCCTGCTGAACCGCTCTTCCGATCT[8nt barcode]CTTCGGAGACCCTGTCCCTCACCTG
IGHV5	GGCATTCCTGCTGAACCGCTCTTCCGATCT[8nt barcode]CGGGGAGTCTCTGAAGATCTCCTGT
IGHV6	GGCATTCCTGCTGAACCGCTCTTCCGATCT[8nt barcode]TCGCAGACCCTCTCACTCACCTGTG
gDNA FR2 primers	
IGHV1	GGCATTCCTGCTGAACCGCTCTTCCGATCT[8nt barcode]CTGGGTGCGACAGGCCCCTGGACAA
IGHV2	GGCATTCCTGCTGAACCGCTCTTCCGATCT[8nt barcode]TGGATCCGTCAGCCCCCAGGGAAGG
IGHV3	GGCATTCCTGCTGAACCGCTCTTCCGATCT[8nt barcode]GGTCCGCCAGGCTCCAGGGAA
IGHV4	GGCATTCCTGCTGAACCGCTCTTCCGATCT[8nt barcode]TGGATCCGCCAGCCCCCAGGGAAGG
IGHV5	GGCATTCCTGCTGAACCGCTCTTCCGATCT[8nt barcode]GGGTGCGCCAGATGCCCGGGAAAGG
IGHV6	GGCATTCCTGCTGAACCGCTCTTCCGATCT[8nt barcode]TGGATCAGGCAGTCCCCATCGAGAG
IGHV7	GGCATTCCTGCTGAACCGCTCTTCCGATCT[8nt barcode]TTGGGTGCGACAGGCCCCTGGACAA
Isotype constant region primers	
IgA	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN[8nt barcode]GAAGACCTTGGGGCTGGT
IgD	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN[8nt barcode]CCCTGATATGATGGGGAACA
IgE	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN[8nt barcode]GAAGACGGATGGGCTCTGT
IgG	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN[8nt barcode]TTCGGGGAAGTAGTCCTTGA
IgM	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN[8nt barcode]GGGAATTCTCACAGGAGACG
gDNA IGHJ primer	
IGHJ	ACACTCTTTCCCTACACGACGCTCTTCCGATCTNNNN[8nt barcode]CTTACCTGAGGAGACGGTGACC

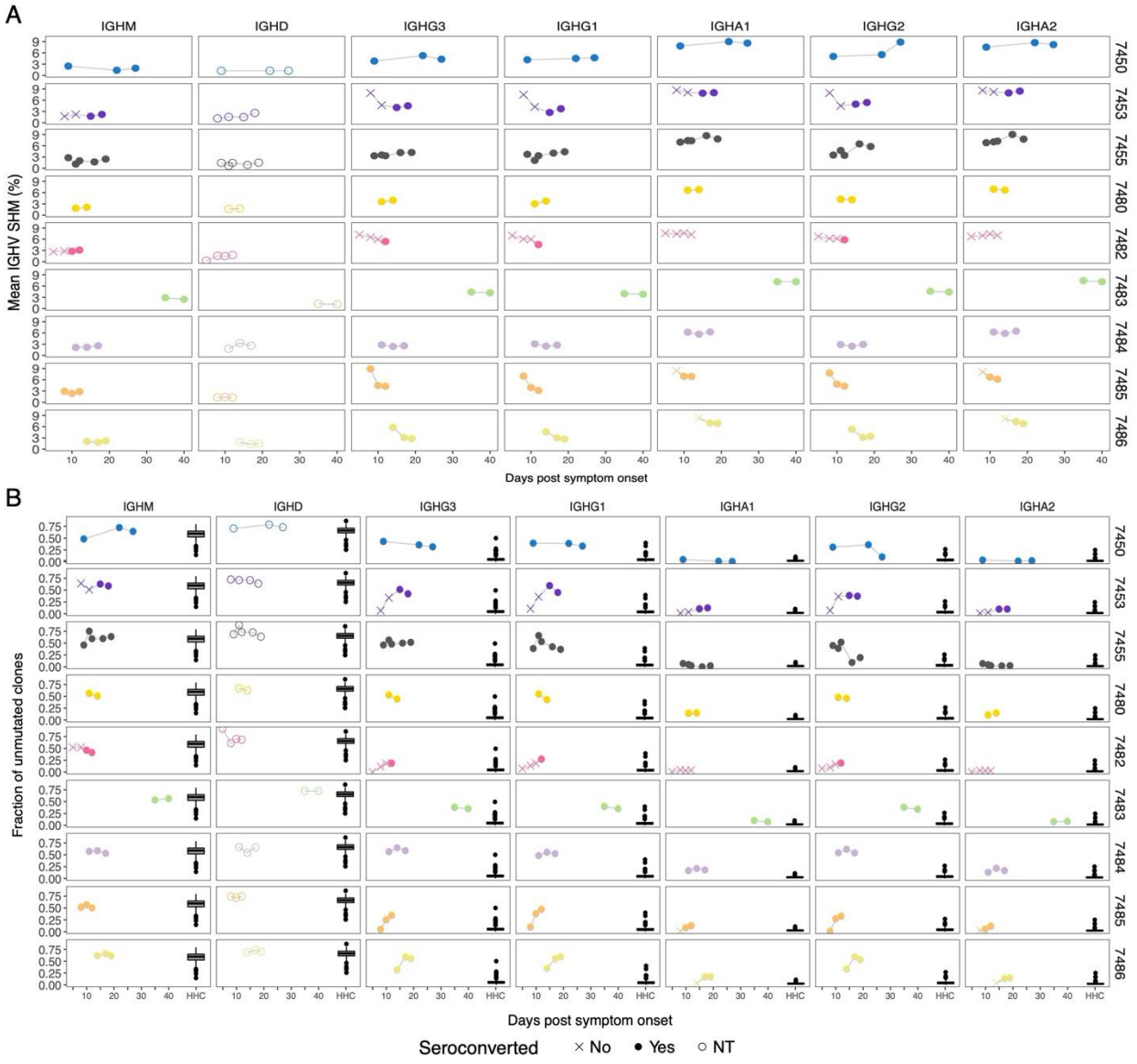


Figure S1. Mean SHM percentage and unmutated clone fractions over time for each patient, Related to Figure 2. (A) The average SHM (%) for the IGHV gene segment of expressed antibodies of the indicated isotype (panel column) for each individual (panel row). SHM for each isotype in each sample was summarized as the median SHM of reads expressed as the indicated isotype within each clone, then taking the mean SHM percentage over all clones. **(B)** The fraction of unmutated (<1% IGHV SHM, y-axis) B cell lineages for each IGH of the indicated isotype (panel column) for each individual (panel row). **(A)** and **(B)** Days post symptom onset on x-axis. Point shapes indicate seronegative (x) or seropositive (filled circle) for the isotype (IgM serology for IgM; IgG serology for IgG subclasses; and IgA serology for IgA subclasses), with gray for HHC. IgD is indicated with an open circle as serology was not tested (NT).

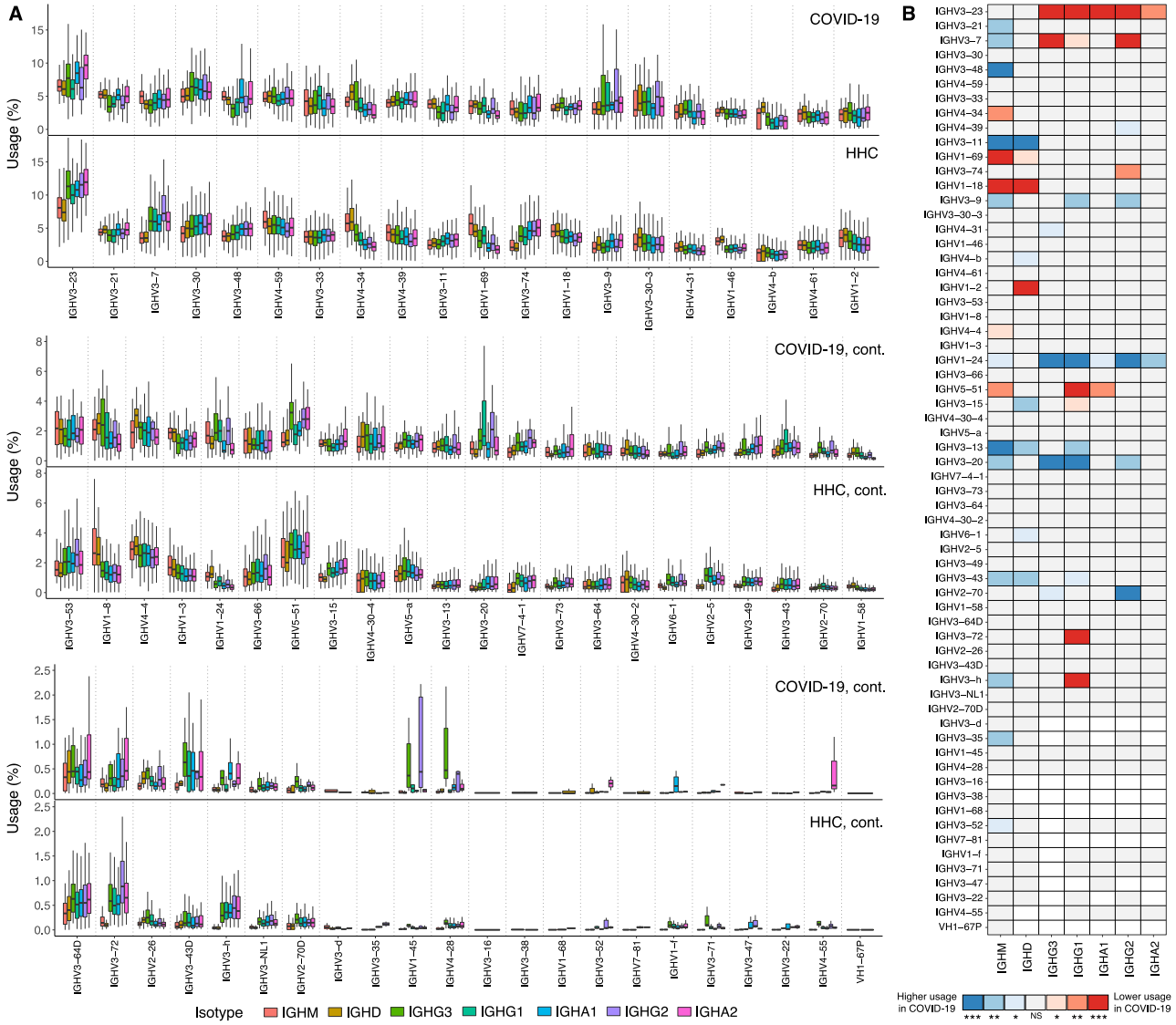


Figure S2. IGHV usage for COVID-19 patients compared to a healthy human control (HHC) cohort, Related to Figure 2. Median frequency of IGHV gene utilization for each isotype subclass observed for the 13 COVID-19 patients compared to a cohort of 114 HHC. **(A)** IGHV usage is shown as the median for clones within each COVID-19 patient and HHC. **(B)** Gene utilization frequencies for each isotype subclass for each IGHV gene were compared between COVID-19 and HHC (paired Wilcoxon tests with Bonferroni correction for multiple hypothesis testing). Where gene usage differed significantly between the cohorts and the gene was utilized at a higher median frequency among the COVID-19 patients the adjusted significance is shown in blue, whereas genes that differed significantly but with lower median usage for the COVID-19 patients relative to the HHC are plotted in red. Instances with insufficient data for test (only 1 or no data points) are in white. The plot y-axes were chosen to show the box-whiskers on a readable scale; rare outlier points with extreme values are not shown but were included in all analyses.

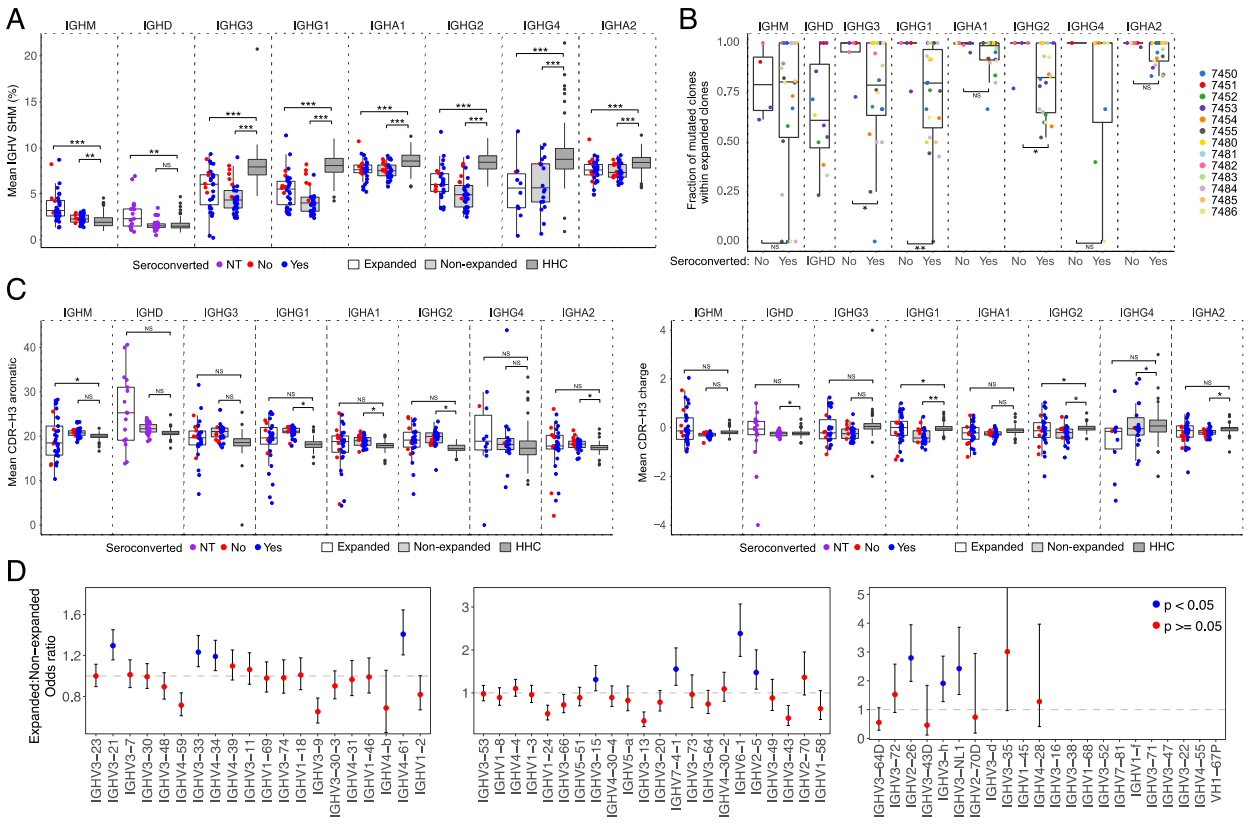


Figure S3. IGHV gene and CDR-H3 features between expanded and non-expanded patient clones and healthy human controls (HHC), Related to Figure 2. (A) Mean IGHV SHM of expanded and non-expanded clones in COVID-19 patients compared to HHC. SHM frequency for each isotype in each sample was summarized as the median SHM of reads expressed as the indicated isotype within each clone, then taking the mean SHM percentage over all clones. Points are jittered on the x-axis to decrease over-plotting of samples with the same value (y-axis). p-values were calculated by two-sided Wilcoxon–Mann-Whitney tests. **(A)** and **(C)** COVID-19 patient samples grouped by expanded clones (white) or non-expanded clones (light gray), and total clones from healthy human (HHC) (dark gray, only outlier points are displayed for this group). Points indicate seronegative (red) or seropositive (blue) for the isotype (IgM serology for IgM; IgG serology for IgG subclasses; and IgA serology for IgA subclasses), with gray for HHC. IgD is indicated in purple as serology was not tested (NT). **(B)** Fraction of mutated expanded clones (y-axis) by seroconversion status (x-axis). Points are colored by participant. p-values were calculated by Fisher's exact tests. **(C)** Mean CDR-H3 charge and aromaticity of expanded clones in COVID-19 patients. Percent aromaticity (left panel) and mean charge (right panel) for CDR-H3 amino acid residues. p-values were calculated by one-way ANOVA with Tukey's HSD test. **(A-C)**, ***p-value < 0.001; **p-value < 0.01; *p-value ≤ 0.05; NS: p-value > 0.05. **(D)** Odds ratio (OR) of IGHV gene usage in expanded clones compared to non-expanded clones in COVID-19 patients. Each dot is the OR value for each IGHV gene, the bars represent confidence interval. The plot y-axis was chosen to show the points on a readable scale; extreme values are not shown but were included in all analyses. Each point was colored by p-value tested by Fisher's exact test, blue: p-value < 0.05, red: p-value >= 0.05.

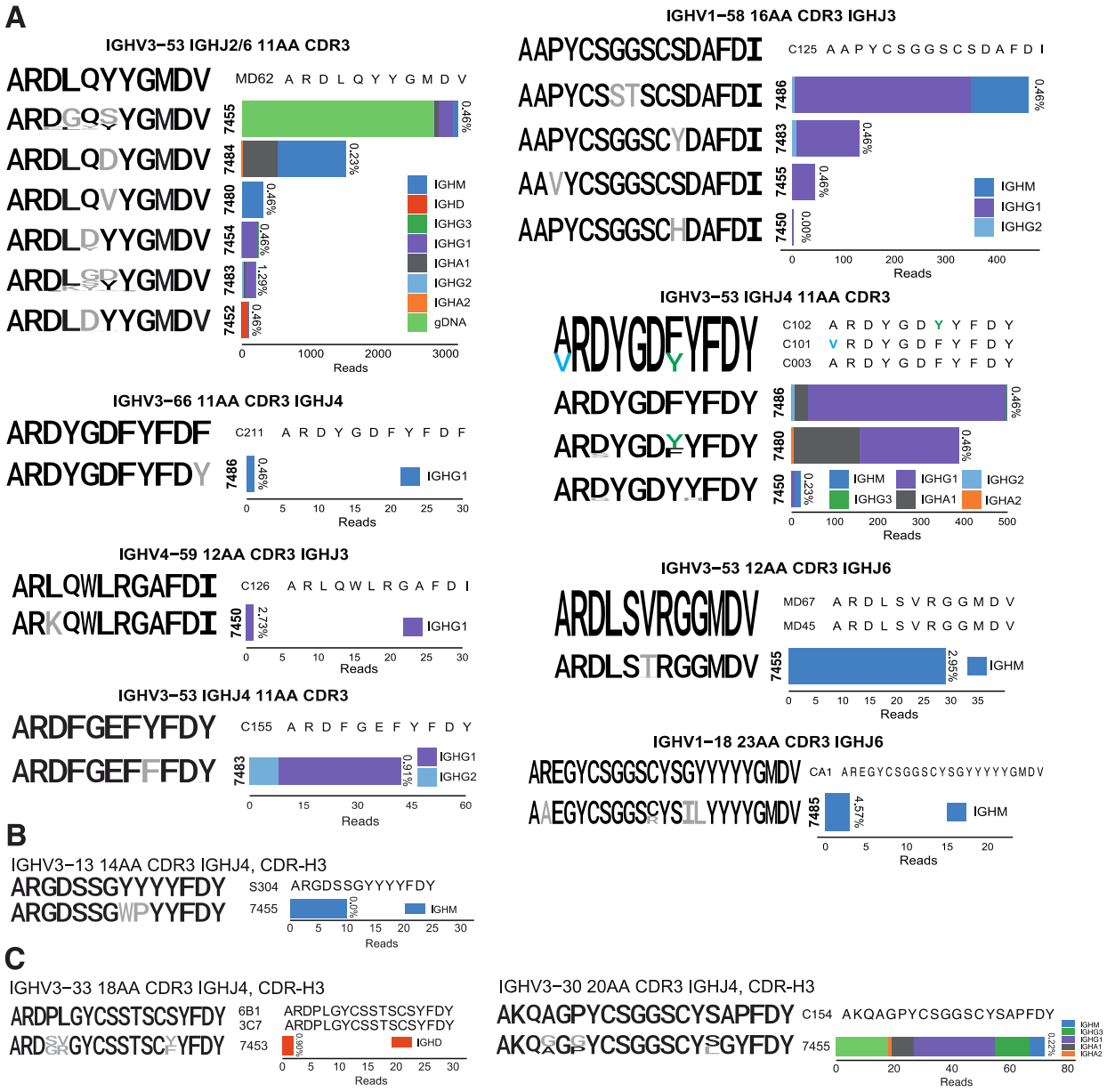


Figure S4. Sequence logos of CDR-H3 AA residues from convergent IGHs in patients with COVID-19, Related to Figure 3. Sequence logos of CDR-H3 AA residues from anti-SARS-CoV-2 (A), anti-SARS-CoV/CoV-2 cross-neutralizing (B) or anti-SARS-CoV (C) convergent IGHs. For each set of convergent IGH the sequence logo and alignment for the reported antigen-specific CDR-H3 is shown at the top, and sequence logos for clones from each patient are aligned below (colored black where they match a conserved residue in the reported CDR-H3, colored for non-conserved as depicted in the alignment, or gray if no match). The read count per patient that contributed to the sequence logo is plotted for each isotype, with the SHM frequency shown beside the bar.