

B cell clonal expansion and convergent antibody responses to SARS-CoV-2

Sandra C. A. Nielsen

Department of Pathology, Stanford University

Fan Yang

Department of Pathology, Stanford University

Ramona A. Hoh

Department of Pathology, Stanford University

Katherine J. L. Jackson

Garvan Institute of Medical Research

Katharina Roeltgen

Department of Pathology, Stanford University

Ji-Yeun Lee

Department of Pathology, Stanford University

Arjun Rustagi

Department of Medicine, Division of Infectious Diseases and Geographic Medicine, Stanford University

Angela J. Rogers

Department of Medicine, Division of Pulmonary, Allergy and Critical Care Medicine, Stanford University

Abigail E. Powell

Stanford ChEM-H and Department of Biochemistry, Stanford University

Peter S. Kim

Stanford ChEM-H and Department of Biochemistry, Stanford University

Taia T. Wang

Department of Medicine, Division of Infectious Diseases and Geographic Medicine, Stanford University

Benjamin Pinsky

Department of Pathology, Stanford University

Catherine A. Blish

Department of Medicine, Division of Infectious Diseases and Geographic Medicine, Stanford University

Scott D. Boyd (✉ sboyd1@stanford.edu)

Department of Pathology, Stanford University

Research Article

Keywords: SARS-CoV-2, B cells, antigen receptors

DOI: <https://doi.org/10.21203/rs.3.rs-27220/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

During virus infection B cells are critical for the production of antibodies and protective immunity. Establishment of a diverse antibody repertoire occurs by rearrangement of germline DNA at the immunoglobulin heavy and light chain loci to encode the membrane-bound form of antibodies, the B cell antigen receptor. Little is known about the B cells and antigen receptors stimulated by the novel human coronavirus SARS-CoV-2. Here we show that the human B cell compartment in patients with diagnostically confirmed SARS-CoV-2 and clinical COVID-19 is rapidly altered with the early recruitment of B cells expressing a limited subset of V genes, and extensive activation of IgG and IgA subclasses without significant somatic mutation. We detect expansion of B cell clones as well as convergent antibodies with highly similar sequences across SARS-CoV-2 patients, highlighting stereotyped naïve responses to this virus. A shared convergent B cell clonotype in SARS-CoV-2 infected patients was previously seen in patients with SARS. These findings offer molecular insights into shared features of human B cell responses to SARS-CoV-2 and other zoonotic spillover coronaviruses.

Background

The novel human severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the etiological agent of the coronavirus disease 2019 (COVID-19)^{1,2} pandemic. Prior to the emergence of SARS-CoV-2, six human coronaviruses (hCoVs) were known; four seasonal hCoVs (hCoV-229E, -NL63, -HKU1, and -OC43)³ causing usually mild upper respiratory illness, and the two more recently discovered SARS-CoV-4 and MERS-CoV-5 viruses that arose from spillover events of virus from animals into humans. It is expected that humans are naïve to SARS-CoV-2 and will display a primary immune response to infection. Humoral immune responses will likely be critical for the development of protective immunity to SARS-CoV-2, but little is known about the B cells and antigen receptors stimulated by infection.

Sars-cov-2 Infection Causes Global Changes In The Antibody Repertoire

High-throughput DNA sequencing of B cell receptor (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, during the evolving humoral response⁶. To study the development of SARS-CoV-2-specific humoral responses, we collected peripheral blood samples from six patients admitted to Stanford Hospital with signs and symptoms of COVID-19. All patients were confirmed positive for SARS-CoV-2 RNA as determined by quantitative reverse transcription PCR (RT-qPCR). Serology of SARS-CoV-2-specific IgA, IgG, and IgM antibodies showed early detection of these antibody (Fig. 1a and Extended Data Fig. 1). Immunoglobulin heavy chain (IGH) repertoires were sequenced and compared to a healthy human control (HHC) dataset⁷ matched by age and number of B cell clones (Fig. 1b, top panel). In healthy subjects at baseline, IgM and IgD sequences are primarily derived from naïve B cells with unmutated IGHV genes, and typically have small clone sizes, whereas isotype switched cells (IgA and IgG compartments) have elevated SHM with varying clone sizes.

SARS-CoV-2 seroconverted patients (7450, 7452, and 7454), and patients 7453-D2 and 7455, in contrast, display a highly polyclonal burst of IgG-expressing clones in the blood with little to no SHM. Seronegative samples (7451 and 7453-D0) show IGH repertoires similar to uninfected controls, suggesting an earlier stage in the infection at the time of sample collection.

Fig. 1. COVID-19 patient IGH repertoires show early and extensive class-switching to IgG and IgA subclasses without significant somatic mutation. **a**, Blood was collected from six COVID-19 patients at admission to hospital (D0) and for one patient again two days later (D2). Colored bars indicate assays and experiments performed on samples (RT-qPCR: green; Serology: orange; high-throughput sequencing (HTS): blue). **b**, Points indicate B cell clonal lineages, with the position denoting the clone's isotype (panel column), human healthy control (HHC) or patient ID (panel row), IGHV gene (x-axis, with IGHV gene in the same order and position in panels, but not listed by name due to space constraints), and CDR-H3 length (y-axis within each panel). The point color indicates the mean IGHV SHM frequency for each clone and the size indicates the number of unique reads grouped into the clone. Points are jittered to decrease overplotting of clones with same IGHV gene and CDR-H3 length. Patient label colors indicate seroconversion (blue), seronegative (red), and no serology performed (black). See Extended Data Fig. 2 for an overview of the IgG4 and IgE subclasses.

As suggested by Fig. 1b, COVID-19 patients have a significantly increased fraction of unmutated and low mutation (<1% SHM in IGHV gene) clonal lineages among the antigen-experienced, class switched IgG subclasses (IgG1: p-value = 0.00112; IgG2: p-value = 0.000667; IgG3: p-value = 0.000882; IgG4: p-value = 0.00667) (Fig. 2a). Among some seroconverted subjects (7452 and 7454), 50% or more of the IgG3, IgG1, and IgG2 compartments are in this low mutational state, compared to <1% of lineages from corresponding HHC compartments (Fig. 2a). A similar influx of low mutation lineages into the IgG compartment has previously been observed in response to Ebola virus (EBOV) infection⁸. In contrast to EBOV, we observe that COVID-19 primary infection stimulates polyclonal B cell responses with both IgG and, in some patients, IgA subclasses, rather than predominantly IgG alone (Fig. 1b). The SARS-CoV-2 response features a preponderance of IgG1-expression among IgG subclasses (Fig 2b), with median usage of IgG1 being almost double the proportion seen in HHC B cells (p-value =

2.065e-05). IgA1 is also significantly increased (p-value = 0.0123) relative to IgA2. This suggests an influx of recently recruited naïve B cells in the response. The virus-induced response is polyclonal with diverse clonal lineages using a wide range of IGHV genes (Fig. 2c). However, we observed skewing of the responding IGH repertoires away from the usually most frequently utilized IGHV genes such as IGHV3-23, towards IGHV3-30 and IGHV3-9 usage, with particular enrichment of IGHV3-9 in IgG-expressing B cells. The selection of particular IGHVs in response to an antigen has been observed in other antiviral responses, such as the preference for IGHV1-69 in response to some influenza virus antigens⁹. Highly utilized IGHV genes display SHM patterns consistent with the observed frequencies of unmutated or low mutation clonal lineages among switched isotype compartments (Fig. 2a), with low median IgG1 SHM that ranges from 2.2-5.4%, and higher median IgA1 SHM that ranges from 6.2-8.6% (Fig. 2d).

Fig. 2. SARS-CoV-2 infection alters isotype mutation frequencies, subclass proportions, and usage of IGHV genes in the IGH repertoire. **a**, Fraction of unmutated (<1% SHM) B cell lineages within each isotype subclass. Points are shown for all COVID-19 patients, but only outlier points are displayed for the 114 healthy human controls (HHCs). **b**, Isotype subclass usage within IgA and IgG. **(C)** IGHV gene usage is shown as the mean for clones within each COVID-19 patient (top) or HHC (bottom). IGHV order is based on the 20 most common IGHV genes in IgM in the COVID-19 patients and isotypes are plotted by their chromosomal ordering. The plot 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. Asterisks represent IGHV genes that are used by reported SARS-CoV-2-specific IGH10. **d**, SHM frequency of combined clones from COVID-19, ordered by IGHV gene usage. p-values in **(a)** and **(b)** were calculated by two-sided Wilcoxon–Mann-Whitney tests: ***p-value < 0.001; **p-value < 0.01; *p-value ≤ 0.05; NS: p-value > 0.05.

Expanded B Cell Clones Correlate With Seroconversion And Show Distinctive Igh Features

To further explore antigen-driven B cell responses in the patient cohort, we identified expanded clonal lineages based on their presence in two or more replicate genomic DNA (gDNA) libraries generated from separate template aliquots isolated from patient peripheral blood mononuclear cells (PBMCs). This method avoids distortions of sequence counts in cDNA libraries caused by high mRNA expression from some cells¹¹ (Fig. 3a and Extended Data Fig. 3). IGHV genes that were significantly enriched in expanded clones were determined based on the odds ratio (OR) of IGHV gene usage in expanded clones versus non-expanded clones included IGHV6-1 (OR = 2.3, p-value = 0.0004), IGHV3-33 (OR = 1.30, p-value = 0.005) and IGHV4-61 (OR = 1.44, p-value = 0.007). CDR-H3 regions from total clones and expanded-only clones in COVID-19 patient IGH repertoires display clear differences in physicochemical properties compared to total HHC CDR-H3 regions (Fig. 3b and Extended Data Fig. 4). SHM values for class-switched expanded clones are significantly lower in COVID-19 patients compared to healthy controls in sequences expressing all IgG and IgA subclasses (p-value < 0.01 for IgG3, IgG4, IgA1, and IgA2; p-value < 0.001 for IgG1 and IgG2), and significantly higher in IgM (p-value < 0.001) and IgD (p-value < 0.01). There is a strong correlation between the fraction of serum anti-SARS-CoV-2 receptor binding domain (RBD) IgG levels and the fraction of IgG-expressing clonally-expanded lineages in patient (R = 0.95, p-value = 0.014) (Fig 3c).

Fig. 3. Expanded B cell clones in COVID-19 patients show selection for hydrophobic and longer CDR-H3s with less mutation. **a**, Expanded clone counts in COVID-19 patients. Each point represents the number of expanded clones (y-axis) detected in a patient sample that contain the specified isotype (x-axis). Point colors indicate serology not tested (black), seronegative (red), and seropositive (blue). **b**, Mean CDR-H3 hydrophobicity (top panel), mean CDR-H3 length (middle panel), and mean IGHV SHM (bottom panel) for COVID-19 patient expanded clones (light green) or total clones (dark green), or for total healthy human control (HHC) clones (pink, only outlier points are displayed for this group). Each point on the plot represents the average of expanded clones in a patient sample that express the indicated isotype (x-axis). 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: ***p-value < 0.001; **p-value < 0.01; *p-value ≤ 0.05; NS: p-value > 0.05. **c**, Correlation between the serum anti-SARS-CoV-2-RBD levels measured by ELISA (x-axis) and the fraction of expanded clones with SHM ≥ 1% (y-axis). Clones expressing IgG were correlated with levels of anti-SARS-CoV-2-RBD- IgG antibodies; clones expressing IgM were correlated with levels of anti-SARS-CoV-2-RBD- IgM antibodies. Each point represents a patient sample. Shaded regions indicate 95% confidence intervals. R2 and p were measured by Pearson's correlation. Only individuals for whom serology was performed are shown (7450, 7451, 7452, 7453, and 7454).

Convergent Antibody Rearrangements Are Elicited In Covid-19 Patients

Although antigen-driven antibody responses are diverse between individuals, we and others have previously identified patterns of highly similar, “convergent” antibodies shared by individuals in response to Ebola virus⁸, and different convergent antibodies stimulated by Dengue virus¹².

These convergent antibodies usually make up a small proportion of the total virus-specific B cell response in each individual⁸. To identify putative SARS-CoV-2-specific antibody signatures, we first grouped heavy chain sequences together that utilized the same IGHV and IGHJ gene family, and CDR-H3 regions with the same length. We then clustered these groups by single-linkage with 85% identity in the CDR-H3 amino acid sequence and identified sequences that were found in at least two COVID-19 patients but were absent from the 114 HHC repertoires. We identified 124 such convergent clusters involving clonal lineages from all six patients. The number of convergent clusters that each patient contributed to varied from 12 clusters in patient 7451 to 65 clusters in patient 7455. 106 clusters were shared pairwise between two patients, 16 clusters spanned three patients, and two clusters spanned four patients (Fig. 4a). Clusters that spanned four subjects included a cluster of lineages using IGHV3-30-3 and IGHJ6 with a 14 amino-acid CDR-H3, which was expanded across different isotypes in 7453-D2 and 7454 with SHM frequencies generally below 2% and was detected as IgD in 7451 and IgG1 in 7455. We further detected one to three convergent clones in samples 7450, 7452, and 7453-D2 that were expanded as per the gDNA replicate analysis. Samples with expanded clones by this definition expressed IgM, IgG1-3, IgA1, and IgA2 (Fig. 4b). We next sought to identify SARS-CoV-2-specific sequences homologous to known sequences that recognize SARS-CoV-2 or the related betacoronavirus, SARS-CoV. Sequences convergent to known SARS-CoV-specific antibody IGH sequences¹³ were identified in one COVID-19 patient (7453-D2) but were not detected in the HHC samples (Fig. 4c). We also identified one IGH (IGHV3-13, IGHJ4, and CDR-H3 length of 14), recently reported to cross-react to SARS-CoV and SARS-CoV-2¹⁴, which clustered with IGH sequences identified in patient 7455 (Fig. 4c).

Fig. 4. Convergent IGH sequences are found between COVID-19 patients. a, The distribution of convergent 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. (bottom panel) Lineages belonging to convergent clusters (y-axis) shared across three or four patients (columns, four-patient clusters are highlighted in pink) are plotted by the expressed isotype (x-axis). Fill color indicates the average SHM and point size shows the number of unique reads. **b**, Expanded convergent clone counts in COVID-19 patients. Each point represents the number of expanded convergent clones (y-axis) detected in a patient sample that express the specified isotype (x-axis). Point colors indicate serology not tested (black), seronegative (red), and seropositive (blue). **c**, Sequence alignment of CDR-H1, CDR-H2, and CDR-H3 amino acid residues of anti-SARS-CoV convergent IGH (rows). Sequences were aligned against a SARS-CoV and SARS-CoV-2 cross-reactive IGH (top) or SARS-CoV-specific IGH (bottom) sequences (labels in bold). Light blue highlighting indicates sequence differences. Dots indicate homology to the germline sequence (labels are underlined).

Our results indicate that the IGH repertoires of patients with diagnostically confirmed SARS-CoV-2 RNA and COVID-19 infection are rapidly remodeled in response to this novel virus. A burst of expanded clones with low SHM, shifts in the utilization of IGHV genes and subclasses, and longer and more hydrophobic CDR-H3 regions dominate the molecular changes observed in IGH. We observe convergent antibody responses shared with those seen in patients infected with SARS-CoV, highlighting common modes of human antibody response shared between SARS-CoV-2 and other spillover CoVs. Efforts to detect individuals who have been exposed to SARS-CoV-2 and have recovered after mild or asymptomatic infection have revealed delayed serum antibody responses and low levels of specific antibodies in some individuals¹⁵. Convergent SARS-CoV-2-specific antibody sequences in the memory B cell pools of recovered individuals could offer an independent kind of evidence of prior exposure and potential for more rapid secondary responses upon re-exposure to the virus. Longitudinal tracking of IGH repertoires in larger patient cohorts, correlation with clinical outcomes, and further investigation into the binding properties and functional activity of convergent responding clones, are required to better understand the human antibody responses to this novel human SARS-CoV-2.

Data Availability

All data is available in the main text or the extended materials. The IGH repertoire data for this study have been deposited to SRA with accession number SUB7246339.

References

1. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* **395**, 497-506, doi:10.1016/s0140-6736(20)30183-5 (2020).
2. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* **382**, 727-733, doi:10.1056/NEJMoa2001017 (2020).
3. Su, S. *et al.* Epidemiology, Genetic Recombination, and Pathogenesis of *Trends Microbiol* **24**, 490-502, doi:10.1016/j.tim.2016.03.003 (2016).

4. Peiris, J. S. M. *et al.* Coronavirus as a possible cause of severe acute respiratory syndrome. *The Lancet* **361**, 1319-1325, doi:10.1016/s0140-6736(03)13077-2 (2003).
5. Zaki, A. M., van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D. & Fouchier, R. A. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* **367**, 1814-1820, doi:10.1056/NEJMoa1211721 (2012).
6. Zhou, J. Q. & Kleinstein, S. H. Cutting Edge: Ig H Chains Are Sufficient to Determine Most B Cell Clonal Relationships. *J Immunol* **203**, 1687-1692, doi:10.4049/jimmunol.1900666 (2019).
7. Nielsen, S. C. A. *et al.* Shaping of infant B cell receptor repertoires by environmental factors and infectious disease. *Sci Transl Med* **11**, doi:10.1126/scitranslmed.aat2004 (2019).
8. Davis, C. W. *et al.* Longitudinal Analysis of the Human B Cell Response to Ebola Virus Infection. *Cell* **177**, 1566-1582 e1517, doi:10.1016/j.cell.2019.04.036 (2019).
9. Avnir, Y. *et al.* IGHV1-69 polymorphism modulates anti-influenza antibody repertoires, correlates with IGHV utilization shifts and varies by ethnicity. *Sci Rep* **6**, 20842, doi:10.1038/srep20842 (2016).
10. Ju, B. *et al.* Potent human neutralizing antibodies elicited by SARS-CoV-2 *bioRxiv*, doi:10.1101/2020.03.21.990770 (2020).
11. Boyd, S. D. *et al.* Measurement and Clinical Monitoring of Human Lymphocyte Clonality by Massively Parallel V-D-J Pyrosequencing. *Science Translational Medicine* **1**, 12ra23, doi:10.1126/scitranslmed.3000540 (2009).
12. Parameswaran, P. *et al.* Convergent antibody signatures in human dengue. *Cell Host Microbe* **13**, 691-700, doi:10.1016/j.chom.2013.05.008 (2013).
13. Coughlin, M. *et al.* Generation and characterization of human monoclonal neutralizing antibodies with distinct binding and sequence features against SARS coronavirus using XenoMouse. *Virology* **361**, 93-102, doi:10.1016/j.virol.2006.09.029 (2007).
14. Pinto, D. *et al.* Structural and functional analysis of a potent sarbecovirus neutralizing antibody. *bioRxiv*, doi:10.1101/2020.04.07.023903 (2020).
15. Okba, N. M. A. *et al.* SARS-CoV-2 specific antibody responses in COVID-19 *bioRxiv*.
16. Hogan, C. A., Sahoo, M. K. & Pinsky, B. A. Sample Pooling as a Strategy to Detect Community Transmission of SARS-CoV-2. *JAMA*, doi:10.1001/jama.2020.5445 (2020).
17. Corman, V. M. *et al.* Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* **25**, 2000045, doi:https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045 (2020).
18. Amanat, F. *et al.* A serological assay to detect SARS-CoV-2 seroconversion in *medRxiv*.
19. van Dongen, J. J. *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* **17**, 2257-2317, doi:10.1038/sj.leu.2403202 (2003).

20. Magoc, T. & Salzberg, S. L. FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957-2963, doi:10.1093/bioinformatics/btr507 (2011).
21. Ye, J., Ma, N., Madden, T. L. & Ostell, J. M. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res* **41**, W34-40, doi:10.1093/nar/gkt382 (2013).
22. Team, R. C. *R: A Language and Environment for Statistical Computing*, <<https://www.R-project.org/>> (2017).
23. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).
24. Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).

Methods

Study Design

Patients admitted to Stanford Hospital with signs and symptoms of COVID-19 and confirmed SARS-CoV infection by RT-qPCR of nasopharyngeal swabs were recruited. Venipuncture blood samples were collected in K2EDTA- 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).

Healthy adult controls (HHCs)

The data set containing control immunoglobulin receptor repertoires has been described previously⁷. 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 CMV, HIV, HCV, HBV, West Nile virus, HTLV, TPPA (Syphilis), and *T. cruzi*. Volunteer age range was 17-87 with median and mean of 52 and 49, respectively.

Molecular and serological testing on COVID-19 patients

SARS-CoV-2 infection in patients was confirmed by reverse-transcription polymerase chain reaction testing of nasopharyngeal swab specimens, using the protocols described in^{16,17}. Plasma antibody testing for IgG and IgM specific for SARS-CoV-2 spike protein receptor binding domain (RBD) was carried out with an enzyme-linked immunosorbent assay based on the protocol and antigen protein production described in¹⁸.

HTS of immunoglobulin heavy chain (IGH) libraries prepared from genomic DNA and cDNA

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 100 ng template/library (25ng/library for 7453-D0). Multiplexed primers to *IGHJ* and the FR1 or FR2 framework regions (3 FR1

and 3 FR2 libraries), per the BIOMED-2 design were used¹⁹ with additional sequence representing the first part of the Illumina linkers. 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 design¹⁹ 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. 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. 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 sec, 58°C 45 sec, 72°C 60 sec; 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' and 3' ends of amplicons; cycling conditions were: 95°C 15 min; 12 cycles of 95°C 30 sec, 60°C 45 sec, 72°C 60 sec; 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 (PE300) using 600 cycle kits.

Sequence quality assessment, filtering, and analysis

Paired-end reads were merged using FLASH²⁰, demultiplexed (100% barcode match), and primer trimmed. The V, D, and J gene segments and V-D (N1), and D-J (N2) junctions were identified using the IgBLAST alignment program²¹. Quality filtering of sequences included keeping only productive reads with a CDR-H3 region, and minimum V-gene alignment score of 200. For cDNA-templated IGH reads, isotypes and subclasses were called by exact matching to the constant region gene sequence upstream from the primer. Clonal identities 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 518,403 clones (per sample, mean number of clones: 74,058; median number of clones: 9,030 for each isotype) were identified. A total of 6,158,222 IGH sequences amplified from cDNA were analyzed for the COVID-19 subjects (mean: 879,746 per individual; median: 910,437) and 68,831,446 sequences from healthy adult controls (mean: 603,785 per individual; median: 637,269). Each COVID-19 patients had on average 280,307 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 subject 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 $\geq 1\%$ 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 were defined as a clone found in one subject which is 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 R22 using base packages for statistical analysis and the ggplot2 package for graphics²³.

To determine convergent rearranged IGH among patients with SARS-CoV-2 infection, we clustered 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-hit²⁴. 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-1 and SARS-CoV-2.

Declarations

Acknowledgments: We thank the patients, Hannah K. Frank for help with sequence alignment, Shilpa A. Joshi for editing the paper, Jonasel Roque, Philip Grant, Aruna Subramanian for assistance with recruiting and consenting subjects, Aaron J. Wilk, Nancy Q. Zhao, Giovanni J. Martinez-Colon, Julia L. McKechnie, Geoffrey Ivison, Thanmayi Ranganath, Rosemary Vergara, and Laura J. Simpson for processing of samples.

Funding: This work was supported by NIH/NIAID T32AI007502-23 (A.R.); NIH/NHLBI K23HL125663 (A.J.R.); 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.), and an endowment to S.D.B. from the Crown Family Foundation. 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.B., S.D.B. conceived the project. S.C.A.N., J-Y.L., K.R., A.E.P., B.P. performed the experiments. A.E.P, P.S.K., T.T.W. contributed to serological assays and analysis. S.C.A.N., F.Y., R.A.H., K.J.L.J., B.P. contributed to data analysis. S.C.A.N., F.Y., R.A.H., K.J.L.J., C.B., S.D.B. wrote the manuscript. All authors edited the manuscript.

Competing interests: The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.D.B.

Reprints and permissions information is available at www.nature.com/reprints.

Figures

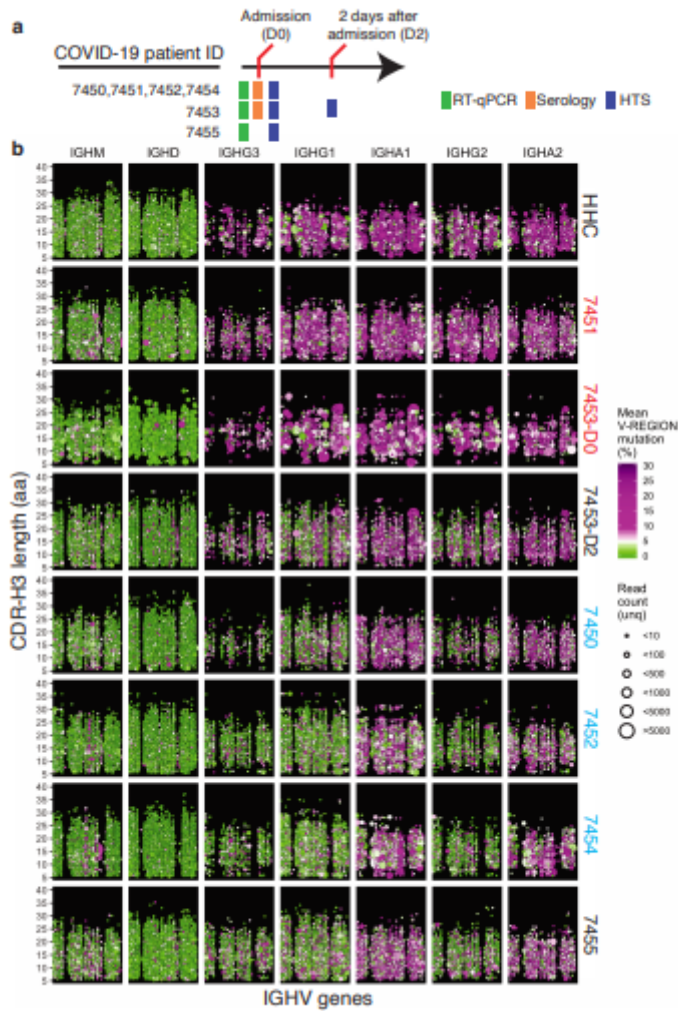


Figure 1

COVID-19 patient IGH repertoires show early and extensive class-switching to IgG and IgA subclasses without significant somatic mutation. a, Blood was collected from six COVID-19 patients at admission to hospital (D0) and for one patient again two days later (D2). Colored bars indicate assays and experiments performed on samples (RT-qPCR: green; Serology: orange; high-throughput sequencing (HTS): blue). b, Points indicate B cell clonal lineages, with the position denoting the clone's isotype (panel column), human healthy control (HHC) or patient ID (panel row), IGHV gene (x-axis, with IGHV gene in the same order and position in panels, but not listed by name due to space constraints), and CDR-H3 length (y-axis within each panel). The point color indicates the mean IGHV SHM frequency for each clone and the size indicates the number of unique reads grouped into the clone. Points are jittered to decrease overplotting of clones with same IGHV gene and CDR-H3 length. Patient label colors indicate seroconversion (blue), seronegative (red), and no serology performed (black). See Extended Data Fig. 2 for an overview of the IgG4 and IgE subclasses.

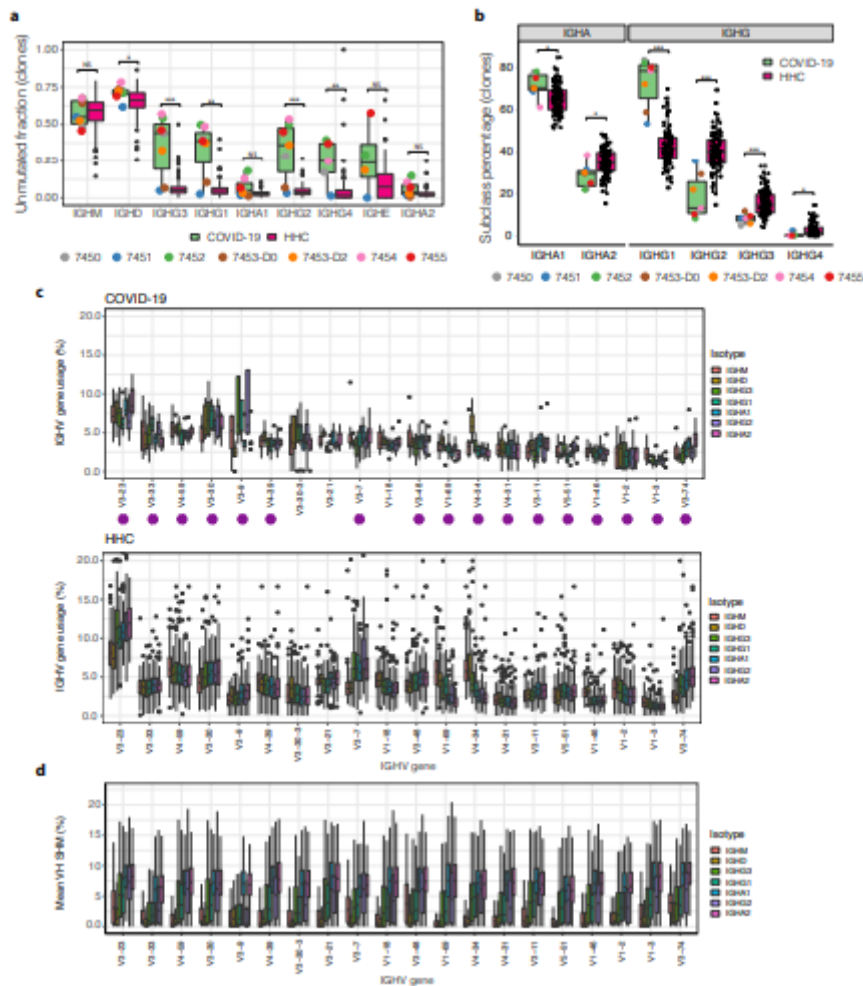


Figure 2

SARS-CoV-2 infection alters isotype mutation frequencies, subclass proportions, and usage of IGHV genes in the IGH repertoire. a, Fraction of unmutated (<1% SHM) B cell lineages within each isotype subclass. Points are shown for all COVID-19 patients, but only outlier points are displayed for the 114 healthy human controls (HHCs). b, Isotype subclass usage within IgA and IgG. (c) IGHV gene usage is shown as the mean for clones within each COVID-19 patient (top) or HHC (bottom). IGHV order is based on the 20 most common IGHV genes in IgM in the COVID-19 patients and isotypes are plotted by their chromosomal ordering. The plot 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. Asterisks represent IGHV genes that are used by reported SARS-CoV-2-specific IGH10. d, SHM frequency of combined clones from COVID-19, ordered by IGHV gene usage. p-values in (a) and (b) were calculated by two-sided Wilcoxon-Mann-Whitney tests: ***p-value < 0.001; **p-value < 0.01; *p-value ≤ 0.05; NS: p-value > 0.05.

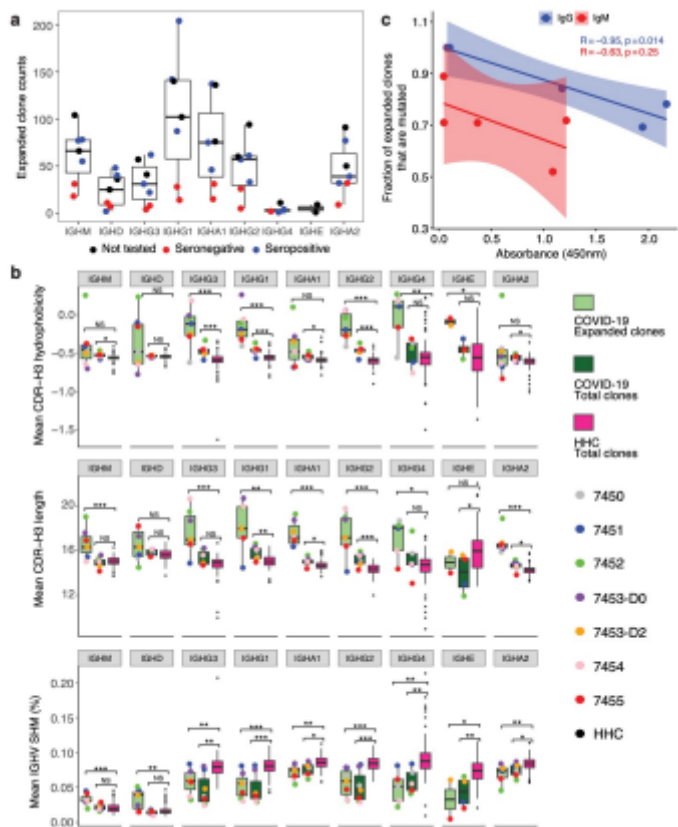


Figure 3

Expanded B cell clones in COVID-19 patients show selection for hydrophobic and longer CDR-H3s with less mutation. a, Expanded clone counts in COVID-19 patients. Each point represents the number of expanded clones (y-axis) detected in a patient sample that contain the specified isotype (x-axis). Point colors indicate serology not tested (black), seronegative (red), and seropositive (blue). b, Mean CDR-H3 hydrophobicity (top panel), mean CDR-H3 length (middle panel), and mean IGHV SHM (bottom panel) for COVID-19 patient expanded clones (light green) or total clones (dark green), or for total healthy human control (HHC) clones (pink, only outlier points are displayed for this group). Each point on the plot represents the average of expanded clones in a patient sample that express the indicated isotype (x-axis). 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: ***p-value < 0.001; **p-value < 0.01; *p-value ≤ 0.05; NS: p-value > 0.05. c, Correlation between the serum anti-SARS-CoV-2-RBD levels measured by ELISA (x-axis) and the fraction of expanded clones with SHM ≥ 1% (y-axis). Clones expressing IgG were correlated with levels of anti-SARS-CoV-2-RBD- IgG antibodies; clones expressing IgM were correlated with levels of anti-SARS-CoV-2-RBD- IgM antibodies. Each point represents a patient sample. Shaded regions indicate 95% confidence intervals. R² and p were measured by Pearson’s correlation. Only individuals for whom serology was performed are shown (7450, 7451, 7452, 7453, and 7454).

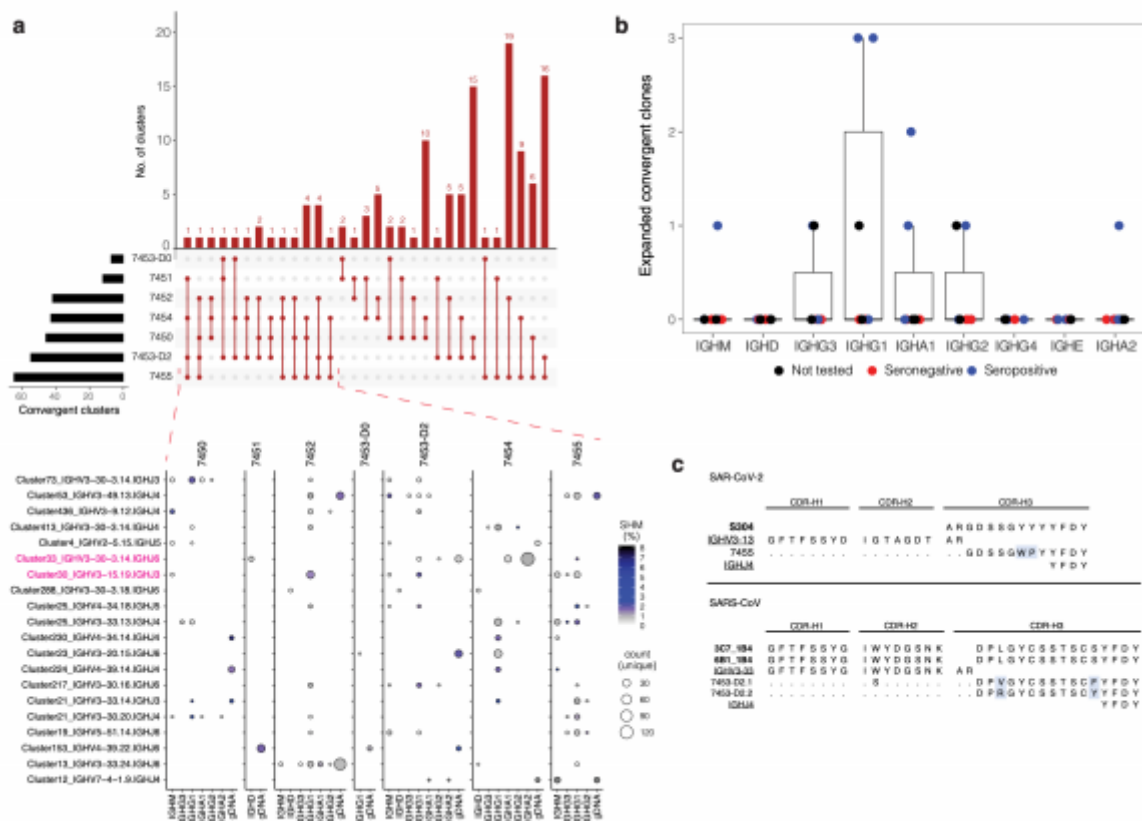


Figure 4

Convergent IGH sequences are found between COVID-19 patients. a, The distribution of convergent 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. (bottom panel) Lineages belonging to convergent clusters (y-axis) shared across three or four patients (columns, four-patient clusters are highlighted in pink) are plotted by the expressed isotype (x-axis). Fill color indicates the average SHM and point size shows the number of unique reads. b, Expanded convergent clone counts in COVID-19 patients. Each point represents the number of expanded convergent clones (y-axis) detected in a patient sample that express the specified isotype (x-axis). Point colors indicate serology not tested (black), seronegative (red), and seropositive (blue). c, Sequence alignment of CDR-H1, CDR-H2, and CDR-H3 amino acid residues of anti-SARS-CoV convergent IGH (rows). Sequences were aligned against a SARS-CoV and SARS-CoV-2 cross-reactive IGH (top) or SARS-CoV-specific IGH (bottom) sequences (labels in bold). Light blue highlighting indicates sequence differences. Dots indicate homology to the germline sequence (labels are underlined).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [extendeddata.pdf](#)