



OPEN ACCESS

Edited by:

Moncef Zouali,
Institut National de la Santé et de la
Recherche Médicale (INSERM), France

Reviewed by:

Victor Greiff,
University of Oslo, Norway
Gur Yaari,
Bar-Ilan University, Israel

***Correspondence:**

Deborah K. Dunn-Walters
d.dunn-walters@surrey.ac.uk

Alexander Stewart
Alexander.stewart@surrey.ac.uk

†These authors have contributed
equally to this work and share
first authorship

‡These authors have contributed
equally to this work and share
senior authorship

Specialty section:

This article was submitted to
B Cell Biology,
a section of the journal
Frontiers in Immunology

Received: 01 November 2021

Accepted: 15 March 2022

Published: 03 May 2022

Citation:

Stewart A, Sinclair E,
Ng JC-F, O'Hare JS, Page A,
Serangeli I, Margreitter C, Orsenigo F,
Longman K, Frampas C, Costa C,
Lewis H-M, Kasar N, Wu B, Kipling D,
Openshaw PJM, Chiu C, Baillie JK,
Scott JT, Semple MG, Bailey MJ,
Fraternali F and Dunn-Walters DK
(2022) Pandemic, Epidemic,
Endemic: B Cell Repertoire
Analysis Reveals Unique Anti-Viral
Responses to SARS-CoV-2, Ebola
and Respiratory Syncytial Virus.
Front. Immunol. 13:807104.
doi: 10.3389/fimmu.2022.807104

Pandemic, Epidemic, Endemic: B Cell Repertoire Analysis Reveals Unique Anti-Viral Responses to SARS-CoV-2, Ebola and Respiratory Syncytial Virus

Alexander Stewart^{1†}, Emma Sinclair^{1†}, Joseph Chi-Fung Ng^{2†}, Joselli Silva O'Hare^{1,3}, Audrey Page³, Ilaria Serangeli⁴, Christian Margreitter⁵, Federica Orsenigo^{1,6}, Katherine Longman⁵, Cecile Frampas⁵, Catia Costa⁵, Holly-May Lewis⁵, Nora Kasar¹, Bryan Wu³, David Kipling¹, Peter JM Openshaw⁷, Christopher Chiu⁷, J Kenneth Baillie⁸, Janet T. Scott⁹, Malcolm G. Semple¹⁰, Melanie J. Bailey⁵, Franca Fraternali^{2‡} and Deborah K. Dunn-Walters^{1*‡}

¹ School of Biosciences and Medicine, University of Surrey, Guildford, United Kingdom, ² Randall Centre for Cell & Molecular Biophysics, King's College London, London, United Kingdom, ³ Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom, ⁴ Dipartimento di Biologia e Biotechnologie Charles Darwin, Sapienza Università di Roma, Rome, Italy, ⁵ Department of Chemistry, University of Surrey, Guildford, United Kingdom, ⁶ Department of Biotechnology and Biosciences, Università degli Studi di Milano-Bicocca, Milan, Italy, ⁷ Faculty of Medicine, Imperial College London, London, United Kingdom, ⁸ Roslin Institute, University of Edinburgh, Edinburgh, United Kingdom, ⁹ MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow, United Kingdom, ¹⁰ Faculty of Health & Life Sciences, University of Liverpool, Liverpool, United Kingdom

Immunoglobulin gene heterogeneity reflects the diversity and focus of the humoral immune response towards different infections, enabling inference of B cell development processes. Detailed compositional and lineage analysis of long read IGH repertoire sequencing, combining examples of pandemic, epidemic and endemic viral infections with control and vaccination samples, demonstrates general responses including increased use of *IGHV4-39* in both Zaire Ebolavirus (EBOV) and COVID-19 patient cohorts. We also show unique characteristics absent in Respiratory Syncytial Virus or yellow fever vaccine samples: EBOV survivors show unprecedented high levels of class switching events while COVID-19 repertoires from acute disease appear underdeveloped. Despite the high levels of clonal expansion in COVID-19 IgG1 repertoires there is a striking lack of evidence of germinal centre mutation and selection. Given the differences in COVID-19 morbidity and mortality with age, it is also pertinent that we find significant differences in repertoire characteristics between young and old patients. Our data supports the hypothesis that a primary viral challenge can result in a strong but immature humoral response where failures in selection of the repertoire risk off-target effects.

Keywords: B cell, COVID-19, ebola, RSV (respiratory syncytial virus), immunity, repertoire, class-switch recombination (CSR)

INTRODUCTION

The emergence of SARS-CoV-2 in 2019, the ensuing pandemic and evolution of novel variants continues to make COVID-19 a matter of global public health significance. The recent SARS, MERS, Zika and Ebola outbreaks have also highlighted a need to better understand how the human immune system responds to novel infections, develop better treatments and control their emergence and spread. Initial reports from the COVID-19 pandemic, relying heavily on serum antibody titres, saw rapid declines in SARS-CoV-2 specific antibodies (1) that raised concerns over the nature and duration of B cell memory. While total antibody titres decrease the persistent presence of SARS-CoV-2-specific memory responses some months after infection mitigates these concerns (2, 3).

Immunoglobulins (Ig), both as secreted antibodies and as B Cell Receptors (BCRs), mediate immunity against multiple pathogens through their vast variability in antigen binding. This variability is produced by V-D-J recombination (4), where V, D and J genes are recombined from a pool of diverse genes. B cells with Ig genes encoding disease-specific antibodies are expanded upon challenge, causing a skewing of the repertoire towards greater use of antigen-specific genes associated with the challenge in question. Furthermore, the imprecise joining of gene segments, together with the action of terminal deoxynucleotidyl transferase (TdT) creates a highly diverse complementarity determining region (CDR)3, which is important for antigen binding, and can be used to identify “clones” of B cells within a repertoire. These clonal assignments allow us to track lineages and follow the progress of the post-activation diversification events of somatic hypermutation (SHM) and Class Switch Recombination (CSR) as the B cell response develops. Thus, repertoire analyses can help to characterise changes in the memory/effector B cell compartments and identify individual genes of interest for possible antibody therapeutics.

Both SHM and CSR are mediated by the enzyme Activation-Induced cytidine Deaminase (AID) and have traditionally been associated with germinal centre events in secondary lymphoid tissue, involving T cell help (5–7). There is, however, also evidence that CSR may occur outside of the germinal centre environment (8–11) and may not require direct T cell help. The ability of a B cell to mount a directed effector response prior to the formation of a germinal centre allows a more rapid immune response but with lower affinity.

Immune responses are often impaired in older people, which has been of particular concern in COVID-19 patients. The older immune system has shown reduced responses to vaccination, frequently with higher numbers of autoreactive antibodies and inflammatory cytokines (12–14). In B cells we, and others, have shown that particular subsets of B cells are altered with age: IgM memory cells (CD19+CD27+IgD+) are decreased in older people while the Double Negative (CD19+CD27-IgD-) are increased (15, 16). Since IgM memory cells are often associated with a T-independent response, the decrease in IgM memory in older people could have severe consequences in infections where a rapid extrafollicular response is required (17, 18). It has also been shown that the B cell repertoire is skewed towards

sequences with longer, more hydrophobic CDR3 regions as we age (16, 19). As an immune response can result in a shift towards lower, less hydrophobic CDR3 regions (14, 20), and higher hydrophobicity has previously been correlated with increased polyspecificity (21–23), the older immune repertoire seems to be disadvantaged in this respect.

In this study we took a long-read repertoire amplification approach that allowed us to track the V-D-J clonal lineages in the context of antibody subclass to better understand, compare and contrast B cell responses to emerging or endemic viruses. Samples were taken from COVID-19 patients during and after infection, Ebola virus disease (EBOV) survivors from West Africa and the UK, volunteers challenged with Respiratory Syncytial Virus and compared with samples from healthy donors. We report the variation of repertoire between disease states in novel virus infection, with a focus on elderly who are known to respond less well to infection, particularly in SARS-CoV-2.

METHODS

Sample Collection

Whole blood samples (RSV, COVID-19, Healthy) were collected into Tempus™ Blood RNA tubes, kept at 4°C, and frozen down to -20°C within 12 hours. Ebola samples were cone filters from plasmapheresis, dissolved in Tri reagent. RNA was extracted using Tempus™ kits according to instructions. Healthy samples taken after SARS-CoV-2 emergence were all confirmed negative for anti-SARS-CoV-2 antibodies by SureScreen lateral flow test and by ELISA (24). Ebola RNA blood samples were collected from convalescent patients with viral RNA negative PCR tests in the 2014–2016 West African outbreak, three patients were Caucasian treated in the UK, and the remaining were convalescent plasma donor participants from a trial in Sierra Leone (25) (consented under the Sierra Leone Ethics and Scientific Review Committee ISRCTN13990511 and PACTR201602001355272 and authorised by Pharmacy Board of Sierra Leone, #PBSL/CTAN/MOHS-CST001). COVID-19 samples were collected from SARS-CoV-2 positive patients at Frimley and Wexham Park hospitals during 2020 (consented under UK London REC 14/LO/1221). Each participant was attributed a “severity score” in relation to their fitness observations at the time of hospital admission using the metadata collected. This score used the “mortality scoring” approach of SR Knight et al. (26) adapted to disregard age, sex at birth and comorbidities, and ranged from 0 to 6; patients scoring 0 to 3 were attributed low severity and patients scoring 4 to 6 were attributed high severity (25). Convalescent COVID-19 patients, from hospital sampling, were contacted for further donations and sample taken 2–3 months post hospital discharge. RSV samples were collected from participants who took part in a human challenge study and were monitored for infection by viral PCR tests (consented under UK London REC 11/LO/1826). Briefly, healthy participants were challenged intranasally with 10⁴ plaque-forming units of the M37 strain of RSV and monitored for up to 6 months as previously described (27).

Repertoire Library Generation

Tempus™ tube samples were defrosted at room temperature and RNA was extracted using the Tempus™ RNA extraction kit according to the manufacturer's instructions. RNA samples were template switch reverse transcribed using SMARTScribe™ reverse transcriptase (Clontech) according to manufacturer's instructions using the SmartNNN TSO Primer (**Supplementary Methods Table 1**) with a minimum of 170 ng of RNA input. The samples were then treated with 0.5 units/μl of Uracil-DNA Glycosylase (NEB) for 60 min at 37°C to reduce UMI interference, then incubated at 95°C for 10 min to inactivate the enzyme. Samples were amplified using Q5 polymerase (NEB) according to manufacturer's instructions with an annealing step of 65°C for 20s and extension step of 72°C for 50 s for 21 cycles. Round one of PCR was performed with forward primer Smart20 and mixed heavy chain (IG[M, G, A]-R1) reverse primers (**Supplementary Methods Table 1**). For PCR1 8 x 20 μl reactions were performed with 1 μl of RNA input per reaction. A semi-nested 2nd PCR was performed with forward primer PID-Step and reverse primers IG[M/G/A]-R2 (**Supplementary Methods Table 1**); 16 reactions of 20μl each was performed for each isotype with the same thermal cycling conditions as PCR1 but with 12 cycles, with 1μl of template. The primers in PCR2 also contain Patient Identifier (PID) sequences to allow multiplexing on PacBio (**Supplementary Methods Table 2**). Samples were run on a bioanalyzer (Agilent 7500), isotypes from patients were pooled at equal concentrations and concentrated using Wizard PCR Clean-up kits (Promega) according to manufacturer's instructions with 30 μl of elute. Each isotype was then purified using a PippinPrep™ with Marker K reagents (Sage Biosciences) used as an external ladder reference (IgM/G/A 600-100bp). The concentration was checked using a DNA quantification kit on the Qubit according to manufacturer's protocol, the different isotype samples were pooled at equal concentrations and purified with SPRIselect beads (Beckman Coulter) at X0.8 sample volume with elution in 30μl of TE buffer. Sequencing was performed on either the PacBio RSII or Sequel platforms (See **Supplementary Methods Table 2**).

Quality control, data cleaning and removal of multiplicated UMIs was carried out as previously published (16, 28). Immunoglobulin V-D-J gene usage and CDRH3 was determined using IMGT/High V-quest. Clonotype clustering was carried out as per (16, 28), in brief: a Levenshtein distance matrix was generated on the CDRH3, hierarchically clustered and branches cut at 0.05 to generate clones. Physicochemical properties were calculated using the R Peptides package (29).

Analysis of Clonal Diversity

We sought for methods to qualitatively (visualising clone size distribution) and quantitatively compare clonal diversity (calculating metrics which summarise clonal diversity). We first noted, as one would expect, that sequencing depth (i.e., number of sequences sampled per repertoire) was a strong predictor of the number of clones (**Supplementary Figure S1**). For all repertoires considered here, a wide range of sequencing

depth was observed (number of sequences range from 836 to 105,323, median = 12,040). We therefore adopted the following procedure in this analysis: first, to quantify the extent of clonal diversity we used the Gini coefficient which measures the evenness in the distribution of clone size across clones; application of this metric to quantify BCR clonal diversity has been well documented (30–32). For a given repertoire, clones were ordered by their clone sizes, and the cumulative distributions of clone sizes (in terms of percentage of sequences in repertoire) and its percentile distribution were compared for evenness. As such, the resulting metric was independent from the *absolute* numbers of sequences and clones, thereby allowing fair comparison across repertoires of different sequencing depths. As Gini coefficient is an indicator of evenness, we took (1 – Gini coefficient) as the metric of clonal diversity. To qualitatively compare clonal diversity, we generated visualisation using the following procedure to minimise the impact of sequencing depth differences: we first sampled 12,000 sequences (≈ median sequencing depth; see above) from each repertoire; for repertoires with less than 12,000 unique observations, this number of sequences was sampled with replacement. We then sampled up to 100 clones with probability scaled by clone sizes to generate bubble plots where each bubble represents a clone and bubble sizes are scaled with clone sizes. Genotypic features like V gene usage can be represented as colours. Such plots were included in **Figure 2C**.

Analysis of BCR Clone Lineage Trees

Lineage trees were reconstructed using the maximum parsimony method implemented in the dnapsars executable in the phylip package (33). All clones with at least 3 sequences were considered; IMGT-gapped, V-gene nucleotide sequences of all observations in the clone together with the annotated germline V-gene sequence (included to root the tree) were included as input to dnapsars. Functionalities implemented in the alakazam R package (34) were used to call dnapsars and reformat the output into text-based tree files (newick format) and directed graphs (igraph objects manipulated in R). The directed graphs were further parsed using functions in alakazam and igraph to obtain, for each observed sequence in the given clone, its distance D from the given germline gene g (denoted here as D_g), as estimated by the dnapsars-reconstructed lineage tree: the closer this distance is to 0, the closer the sequence is to germline and therefore a lower mutational level.

We sought to summarise, for a given clone, the distribution of D_g ; this distribution would indicate the overall mutational level of sequences within the clone (summarised using conventional statistics like the median of D_g) and the evenness of mutational level (i.e. whether the clone consists of expansion of sequences with a similar mutational level, or it comprises sequences with a wide range of mutational levels). This can be visualised as a heatmap (clones [vertical axis] versus D_g [horizontal axis], with colours scaling with density of the distribution; see **Figure 5B**), or as a curve (clones [vertical axis] versus the median of D_g [horizontal axis], See **Figure 5C**). The curve representation allows calculation of area-under-curve (AUC) as a metric which we termed “Germline Likeness”, to quantify mutational

levels across clones. This is similar to quantifying sequence similarity to germline, except that here Germline Likeness quantifies the tendency to which all clones from the BCR repertoire of a given individual have high similarity to the germline (and therefore lower mutational levels).

Detecting Class-Switch Recombination Events From Lineage Trees

Since the lineage trees were constructed using only V-gene sequences (see above), in theory antibody sequences of different subclasses could be ordered in the tree in a way that imply class-switch recombination (CSR) events which are mechanistically impossible. We therefore pruned the dnaps-reconstructed tree to remove edges which imply CSR events that violate the physical order of constant region genes in the human IGH locus. This was performed using a Python implementation of the Edmond's algorithm to construct a minimum spanning arborescence tree with the given germline V gene sequence as root. With this arborescence tree the type of CSR (subclass switched from/to) and the distance-from-germline at which the CSR event occurred (estimated as the median distance-from-germline of the two observations relevant to the given event) were obtained.

We noticed that the quantification of CSR events is dependent on the number of sequences sampled in the repertoire. To eliminate this confounding factor, we followed the sampling protocol for the clone size visualisation (**Figure 2C**): briefly, we sampled 12,000 sequences (\approx median sequencing depth across all samples) from each repertoire; for repertoires with less than 12,000 unique observations, sampling was performed with replacement. The analysis presented here (**Figure 6**) is the result after this subsampling analysis, therefore corrected for difference in sequencing depth.

Convergent Network

Productive heavy chain sequences with CDRH3 of length shorter than 30 amino acids were considered in the construction of a convergent network. Sequences were connected if they meet the following criteria (a) same V and J gene usage; (b) from different individuals; (c) same CDRH3 amino acid length, and (d) $\geq 85\%$ CDRH3 amino acid identity. The same criteria have been applied in published studies investigating convergent clonotypes across SARS-CoV-2 B-cell repertoires (35, 36). To allow interpretation of possible targets of sequences in convergent network clusters, known binders were included in constructing the network. Known binders were taken from the following sources: first experimentally determined antigen-antibody structural complexes deposited in the Protein Data Bank (PDB). PDBe was queried on 19 May 2021 with the search term 'Organism: Severe acute respiratory syndrome coronavirus 2'. The resulting list of PDB entries was overlapped with entries in the SABdab structural antibody databases (37) to obtain list of PDB complexes of antibodies and SARS-CoV-2 proteins. A total of 215 heavy chains from 186 structures were considered. Second, known binders validated in experiments where antibody variable regions were cloned and assessed for SARS-CoV-2 protein binding were taken from published work (38–43). All known binder sequences

were annotated for V/J gene usage using either IMGT/HighV-Quest (if DNA sequences were provided) or IMGT/DomainGapAlign (if only amino acid sequences were provided). Information regarding specificity (i.e. SARS-CoV-2 protein targets) was obtained from either **Supplementary Data Files** in the cited publications or by visual inspection (for PDB structures). **Supplementary Table S2** contains all known binder sequences included in this analysis. To construct the network, known binders were connected to one another and to repertoire sequences using the identical criteria mentioned above. In total 809 unique CDRH3 sequences were considered in constructing the convergence network. The resulting network contains 7500 sequences (7370 from repertoire, 130 known binders).

We note that the method of constructing convergent networks is comparable to single-linkage clustering. We compared these results with complete-linkage clustering (applied in R using the command 'hclust(method = "complete")') and found that using the same amino-acid identity cutoff, it produced clusters which are notably smaller (**Supplementary Figure S6A**). Since in the binders list we have SARS-CoV-2 binders for which crystal structures are available, we reason that those antibody structures which bind the antigen in the same way should be grouped together. Manual inspection of these structures found that they are grouped together in our networks, but are separated under complete-linkage clustering (**Supplementary Figure S6A**). We confirm this observation quantitatively, by comparing the antibody-antigen structures by calculating the TM-score (44) between every pair of binder structures, which confirm that our convergent networks group together binders with similar interface at the three-dimensional structural level (**Supplementary Figure S6B**, see figure legend therein). We therefore reason that complete linkage is too stringent, and our networks are likely to group sequences which engage with the antigen in the same way.

Analogous convergent networks were constructed using the EBOV and RSV repertoire data, separately considered with respective known binders and Healthy individuals' repertoire; the majority of clusters were formed mainly of sequences from Healthy donors absent of known binders (45–48), although we were able to identify two convergent clusters of RSV-infected individuals with similar CDRH3 to known binders of the RSV fusion glycoprotein (**Supplementary Figure S7**). To investigate whether clusters shared across disease conditions exist, convergent networks were also constructed considering CV19, RSV and EBOV repertoire and binder sequences altogether (**Supplementary Figure S8**). **Supplementary Table S3** contains all convergent networks constructed, presented as list of pairwise sequences.

Statistical Analysis and Data Visualisation

V-D-J gene usage for each patient was turned into a proportion to normalise for different numbers of sequences and allow for comparison. Gene usage analysis was performed in GraphPad Prism 8.4.3 using a two-way ANOVA with a Dunnett's *post hoc* test. P-values have been corrected using the Benjamini-Hochberg method wherever applicable. All other statistical analyses were performed in the R statistical computing environment (version 4.0.2). Data visualisation was performed using the R ggplot2 package and the following specialised R packages: visNetwork

(for visualising convergent CDRH3 network clusters) and ggseqlogo (for visualising CDRH3 sequence logos). PDB structures were visualised using PyMOL (version 2.3.0). Histograms of CDRH3 length and hydrophobicity, as measured by Kidera factor 4, were constructed on the Brepertoire website (49).

RESULTS

Patient Cohorts

IGH sequences, of total V-D-J plus ~150-200 bp of C regions, were obtained from pandemic, epidemic and endemic diseases and stages along with 24 healthy controls across multiple age ranges (**Figure 1** and **Table 1**). This included: 16 hospitalised COVID-19 patients (CV19), 5 of these patients had follow-up convalescent samples (CV19-Recovered, hereafter CV19R), 12 Ebola convalescent plasma donors (EBOV), 12 participants challenged with RSV, 6 of whom became infected (RSV-I) and 6 of whom did not (RSV-U). Healthy Samples (Healthy) were a grouping of YFVD0, RSVD0 and samples taken as controls during the (COVID-19) pandemic (n=24). Numbers of sequences varied from 836 to 105,323, median = 12,040 per sample, *IGH* gene usage for each patient was expressed as a proportion of the total in order to normalise for differences in sequence numbers between different samples.

IGH Gene Repertoire Changes in Response to Viral Infection

Although the humoral immune response is varied, with different subclasses of antibody having different effector functions (50), many methods of repertoire analysis have hitherto not distinguished between antibody subclasses. We have used PacBio methods to obtain full V-D-J sequence in the context of subclass usage to investigate class switching events during immune responses to infection. We also distinguish between mutated versus unmutated IgM sequences, as a proxy for identifying IgM memory responses. Comparisons of subclass distribution, in relation to healthy controls, revealed a significant increase in the proportion of IgA1 compared to IgA2 in CV19, and RSV-I and the proportion of IgG1 relative to IgG2 in CV19, EBOV and RSV-I (**Figures 2A, B**). The differences in CV19 IgG and IgA repertoire returned to 'normal' healthy levels by the time of convalescent sampling (CV19-Recovered) 2-3 months later.

Immune challenge is characterised by clonal expansion of B cells that express Ig which reacts with the challenging antigen. We identify members of clones in the repertoire by clustering the CDRH3 regions and looking at the largest clones in each sample we can see evidence of increased clonal expansions in CV19 patients (**Figure 2C**). In the full CV19 repertoire *IGHV4* family genes were expanded (**Figure 2C**), more specifically of *IGHV4-39* (**Supplementary Figure S2**) and some *IGHV3* family; this is particularly noticeable in IgG1 and IgA1. Analysis of clonal diversity of memory B cells using the Gini coefficient, taking all possible clones into account, found that CV19 patients had a less clonally diversified repertoire in all but the IgM mutated, IgG2 and IgG4 partitions (**Figure 2D** and **Supplementary Figure S3**), suggesting pervasive expansions of specific BCR clones.

Unusually, we also saw a decrease in diversity of unmutated IgM sequences, indicative of clonal expansion prior to SHM and CSR (**Figure 2D**). These values returned to normal in the CV19R samples. In comparison, *IGHV1* family was expanded in RSV-I IgG1 partition (**Figure 2C**), particularly of *IGHV1-18* (**Supplementary Figure S4**). Active infection with RSV, as well as samples taken 28 days after yellow fever vaccination, showed an increase in diversity of IgA2. Interestingly, EBOV memory B cell populations were more diverse than healthy controls in all the main switched subclasses (IgG1, IgG2, IgA1, IgA2) (**Figure 2D** and **Supplementary Figure S3**).

Large clone sizes can mask whole repertoire changes, so we analysed the frequency of gene use after reducing the data to one representative sequence per clone. We found increased use of *IGHV3-30* in IgM mutated sequences in CV19 patients (**Figure 2E**), and also of IgM-mutated/IgA1/IgG1 in CV19R; this was unique to CV19. An increase in use of *IGHV4-39* was found in CV19 IgA1 sequences and was also found increased across the board in EBOV and RSV-U samples (**Figure 2E**). *IGHV3-23* was found to be reduced in ongoing infection (likely an offset as a result of relative increases in usage of other genes) but exceeded the healthy levels in CV19R. *IGHV1-69*, which has previously been associated with viral infections (51, 52), was increased in RSV-I but not EBOV or CV19. The YF day 28 vaccine samples increased use of *IGHV3-7* in IgM-mutated only (**Figure 2E**; comparisons for other V genes in **Supplementary Figure S4**).

Complementarity Determining Region 3 (CDRH3) Immaturity in COVID-19

Given the importance of CDRH3 in antibody recognition, and the contribution to CDRH3 from the *IGHD* and *IGHJ* genes, we analysed these also. In CV19 samples there was a significant increase in use of *IGHD2-2*, *IGHD3-3* and *IGHJ6* (**Figure 2E**). These genes tend to be more hydrophobic (*IGHJ6* being the exception) and all have among the longest amino acid lengths with only *IGHD3-16* being 2 AA longer. This contribution can be seen in the overall CV19 repertoire which skews towards longer amino acid sequences and increased hydrophobicity, indicative of early response as affinity maturation causes shorter less hydrophobic CDRs (**Supplementary Figure S5**). A clustering analysis of peptide physicochemical properties of CDRH3 regions generally results in a difference between IgM sequences and memory sequences (**Figure 2F**), presumably reflecting biases in antigen selection during post-challenge development. We can see that healthy and CV19 subclass sequences mostly have similar CDRH3 properties to each other, however, in the case of CV19 IgG1 and IgG3 cluster closer to IgM sequences from healthy and EBOV rather than healthy IgG1 and IgG3 sequences implying a more 'naïve', unselected, repertoire.

Convergent Antibody Clusters Reveal Distinct VDJ Preferences

To assess the functional importance of the skewed patterns of V, D and J gene usage in CV19 we created networks connecting sequences observed in our CV19 and control repertoire data

(**Figure 3A**), using criteria previously employed in discovering ‘convergent’ antibody sequences shared between patients (53). By also including known SARS-CoV-2 binders (see Methods), we determined that this method is superior in comparison with complete-linkage clustering, in grouping together antibody sequences which bind the antigen in similar manner (**Supplementary Figure S6**). We obtain clusters of CDR3 sequences found in both CV19 patients and healthy controls, some of which converge towards known binders of SARS-CoV-2 proteins such as those targeting the receptor binding domain (RBD) of the spike protein (**Figure 3B**). Many of these large convergent clusters did not, however, include a known binder in the network (**Figure 3C**). Interestingly, we observe that some of these convergent clusters also contain sequences from Healthy controls, suggesting that our baseline repertoire contains B cells capable of recognising SARS-CoV-2 proteins, as reported by others (36, 53). Overall, convergent clusters use a diverse set of V genes, but most of our larger convergent clusters contain *IGHV3* or *IGHV4* families and demonstrate increased *IGHJ6* usage as well as the more commonly used *IGHJ4* (**Figure 3C**). A comparison exclusively of the known binders to date reveals distinct combinations of V and J gene preferences (**Figure 3D**). We do find clusters of sequences using *IGHV3-53* and *IGHV1-58* such as those used in anti-RBD antibodies (e.g. **Figure 3B**). We find that sequences from convergent clusters tend to be found in larger clonal expansions than those without evidence of convergence (**Figure 3E**), possibly implying that specific clonal expansions in response to challenge are shared across patients. We note that half of the larger clusters have substantial contributions from healthy control sequences, so there may be some IGH genes, such as *IGHV3-33/IGHJ5* found also in RSV-I and EBOV convergent networks (**Figure 3C** and **Supplementary Figures 6A, B**), which have increased versatility such that they are often seen in response to multiple different challenges.

Similar analyses of RSV and EBOV repertoires were limited by the paucity of information on antibody binders, however it was notable that only RSV-I, and not RSV-U, showed evidence of convergence. *IGHV1-18* appears in a large cluster with a known RSV F-protein binder and although the large *IGHV3-23* cluster does not contain a known binder it forms part of the larger expansion of *IGHV3-23* genes in mutated IgM genes from this cohort (**Figure 2E**, **Supplementary Figure S7B**).

VDJ Selection Differs Between Ages in COVID-19

The disparity in COVID-19 severity and mortality between age groups is striking, so we looked for age-related differences in our B cell repertoire data, grouping by <50 and >60 based on this mortality disparity. The difference in IgA1/IgA2 ratio is less in older people, not reaching statistical significance. (**Figure 4A**). On the other hand, the increase usage of IgG1 in CV19 is robust across age (**Figure 4B**). Considering Ig gene usage, we observe the intriguing case of *IGHV3-30* which is only preferentially used by the over 60s during infection (**Figure 4C**). Conversely, *IGHV3-53*, which appears relatively frequently in known binder data in combination with *IGHJ4/6* but did not appear

in our total cohort analysis (**Figure 2E**), is significantly increased in the under 50s IgM-mutated partition (**Figure 4C**). We also found that *IGHV4-39*, *IGHD2-2*, *IGHD3-3* and *IGHJ6*, which we find are expanded in CV19 across multiple B cell partitions, only have significantly increased expression in the under 50s and not the over 60s; *IGHV4-34* appeared increased in both age groups in the IgG1 partition, but did not reach statistical significance (**Figure 4C**).

IgG1 Is Immature in Response to COVID-19

Beyond the scope of gene usage, our BCR repertoire data also enabled reconstruction of individual BCR lineage trees to make inferences about the evolution of a particular clone. Using the annotated germline V allele as the root of the tree, we estimate, for each sequence in the lineage, its distance from the root (**Figure 5A**); this distance being directly proportional to mutational level. We visualise the distribution of this germline distance across all clonotypes observed in each given individual, and observe that the repertoire is dominated by clonotypes with very low mutational levels for a subset of CV19 patients, whilst the predominance of such clones is broadly absent in healthy controls (**Figure 5B, C**). Interestingly, in repertoires from convalescent individuals (both EBOV and CV19), we instead observe dominance of clonotypes with higher mutational levels, although the pattern is less striking than the CV19 patients during hospitalisation (**Figure 5C**). These curves allow for quantification of the Area Under the Curve (AUC), which constitutes a metric we term “Germline Likeness”: a higher Germline Likeness corresponds to a lower level of mutation across all clones (**Figure 5D**); this is akin to quantifying sequence similarity to the germline, except that Germline Likeness here quantifies such phenomenon for a given repertoire in general, rather than a specific sequence. Using this metric we confirm that CV19 repertoires were dominated by clones that were largely unmutated, while EBOV samples carried the greatest mutation rate (**Figure 5E**). As might be expected, with time to generate a germinal centre response, Germline Likeness in CV19 faded with time (**Figure 5F**), to the point where the CV19R repertoires have similar level of mutations compared to the EBOV-convalescent and healthy control repertoires. Partitioning the analysis by isotype, RSV and healthy controls demonstrate the expected trend where an increased level of mutations can be found in both IgG and IgA compared to IgM (**Figure 5G**). However, in CV19 only IgA showed a significant change in Germline Likeness from IgM, albeit with a similar level of Germline Likeness to the healthy IgM (**Figure 5G**).

Ongoing Class Switch Recombination (CSR) Is Detectable in PBMCs of COVID-19 and EBOV Patients

Our lineage trees were further analysed for CSR events: respecting the sequential order of CSR in the genome, we identify CSR events where sequences of different antibody classes/subclasses are directly connected in the lineage tree. This enables us to trace the timing of CSR events (distance from the germline), the direction of class switching (e.g. from IgM to IgG1) and frequency of observation. Many clones have

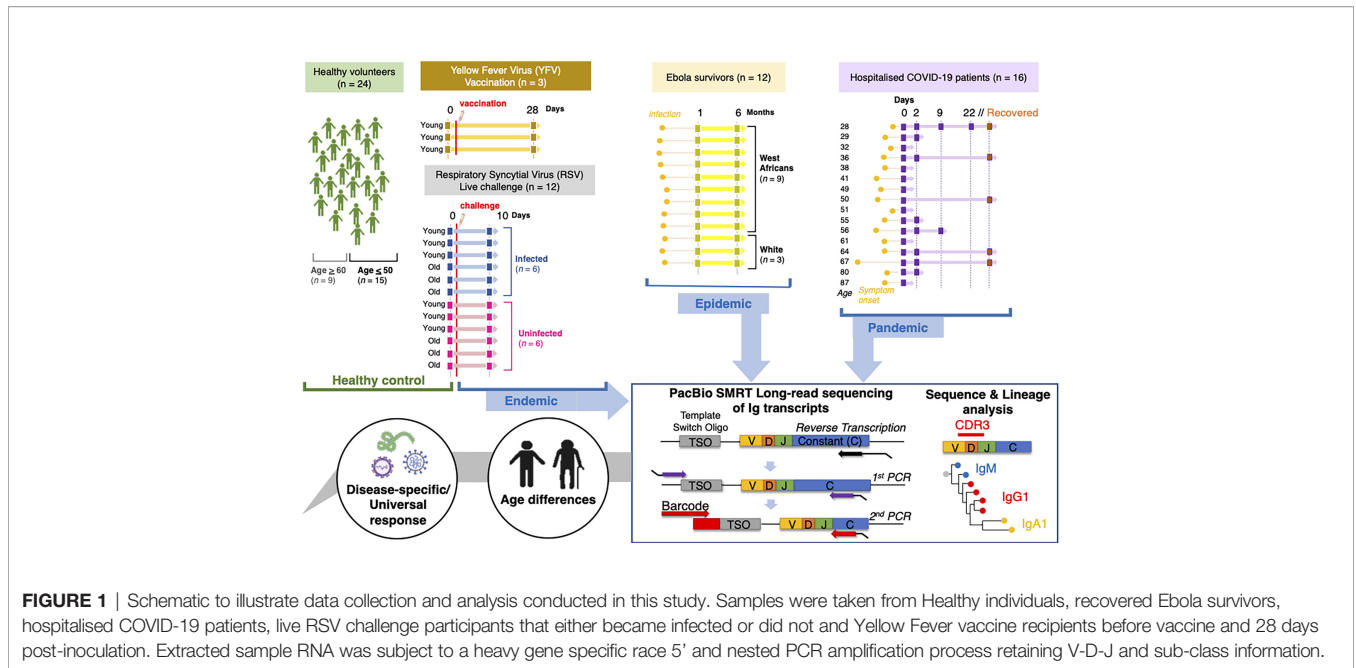


TABLE 1 | Donor characteristics.

Sample	Age	Gender	Ethnicity	COVID-19 Severity Score (out of 6)	Days since symptom onset
Healthy (n = 24)	Median 29.5 (Range 23 - 76) ≤50 years old: 15/24 (62.5%) ≥60 years old: 9/24 (37.5%)	Female: 7/24 (29.2%) Male: 5/24 (20.8%) Unknown: 12/24 (50%)	White: 12/24 (50%) Unknown: 12/24 (50%)		
COVID-19 (n = 16)	Median 50.5 (Range 28 - 87) ≤50: 8/16 (50%) 50-60: 3/16 (18.75%) ≥60: 5/16 (31.25%)	Female: 7/16 (43.75%) Male: 9/16 (56.25%)	White: 13/16 (81.25%) South East Asian: 1/16 (6.25%) Indian Subcontinent: 2/16 (12.5%)	Median 3 (Range 1 - 5)	Median 8 (Range 1 – 35)
COVID-19 Recovered (n = 5)	Median 50 (Range 28 - 87) ≤50: 3/5 (60%) ≥60: 2/5 (40%)	Female: 3/5 (60%) Male: 2/5 (40%)	White: 4/5 (80%) Indian Subcontinent: 1/5 (20%)		
RSV Infected (n = 6)	Young: 3/6 (50%) Older: 3/6 (50%)				
RSV Uninfected (n = 6)	Young: 3/6 (50%) Older: 3/6 (50%)				
Ebola (n = 12)	Young: 3/12 (50%) Unknown: 9/12 (50%)	Female: 1/12 (8.3%) Male: 2/12 (16.7%) Unknown: 9/12 (75%)	White: 3/12 (25%) West African: 9/12 (75%)		
YFV D28 (n = 3)	Median 28 (Range 27 - 28) Young: 3/3 (100%)	Female: 1/3 (33.3%) Male: 2/3 (66.7%)	White: 3/3 (100%)		

See **Supplementary Table S1** for a detailed summary of metadata per donor.

evidence of CSR, particularly in EBOV, even after correcting for clone sizes (**Figure 6A**). In particular, CV19 patients were more likely to switch early to IgG1 from IgM, with little mutation (**Figures 6B–D**) and to IgA1 from IgG1 later in the lineage with more mutation (**Figures 6B, D**). This agrees with the lack of

CDRH3 “maturity” in IgG1 (**Figure 2F**) and the overall increased use of IgG1 and IgA1 seen in CV19 (**Figures 2A, B**).

The evidence of increased CSR in convalescent EBOV patients is striking and occurs across the board with the exception of IgM switching to IgG1 (**Figure 6D**). We noticed

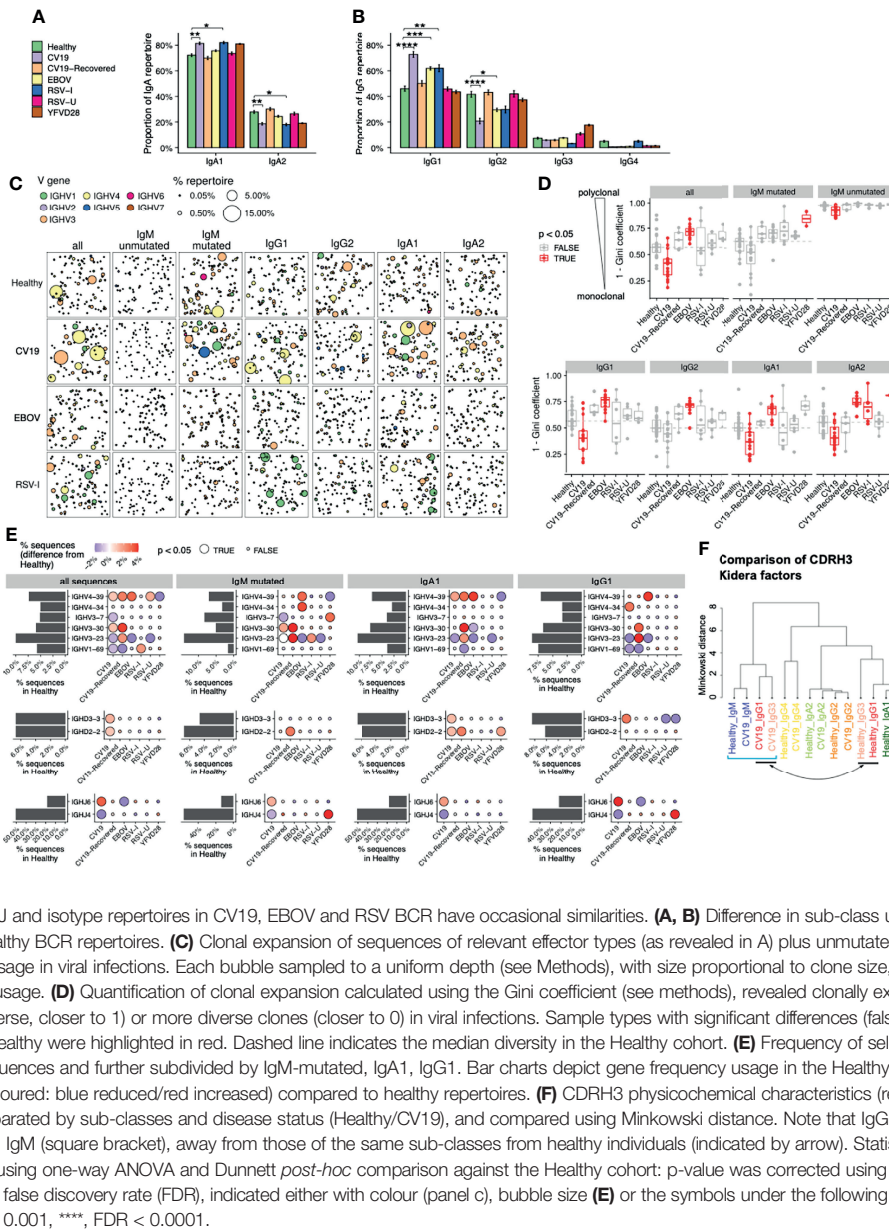


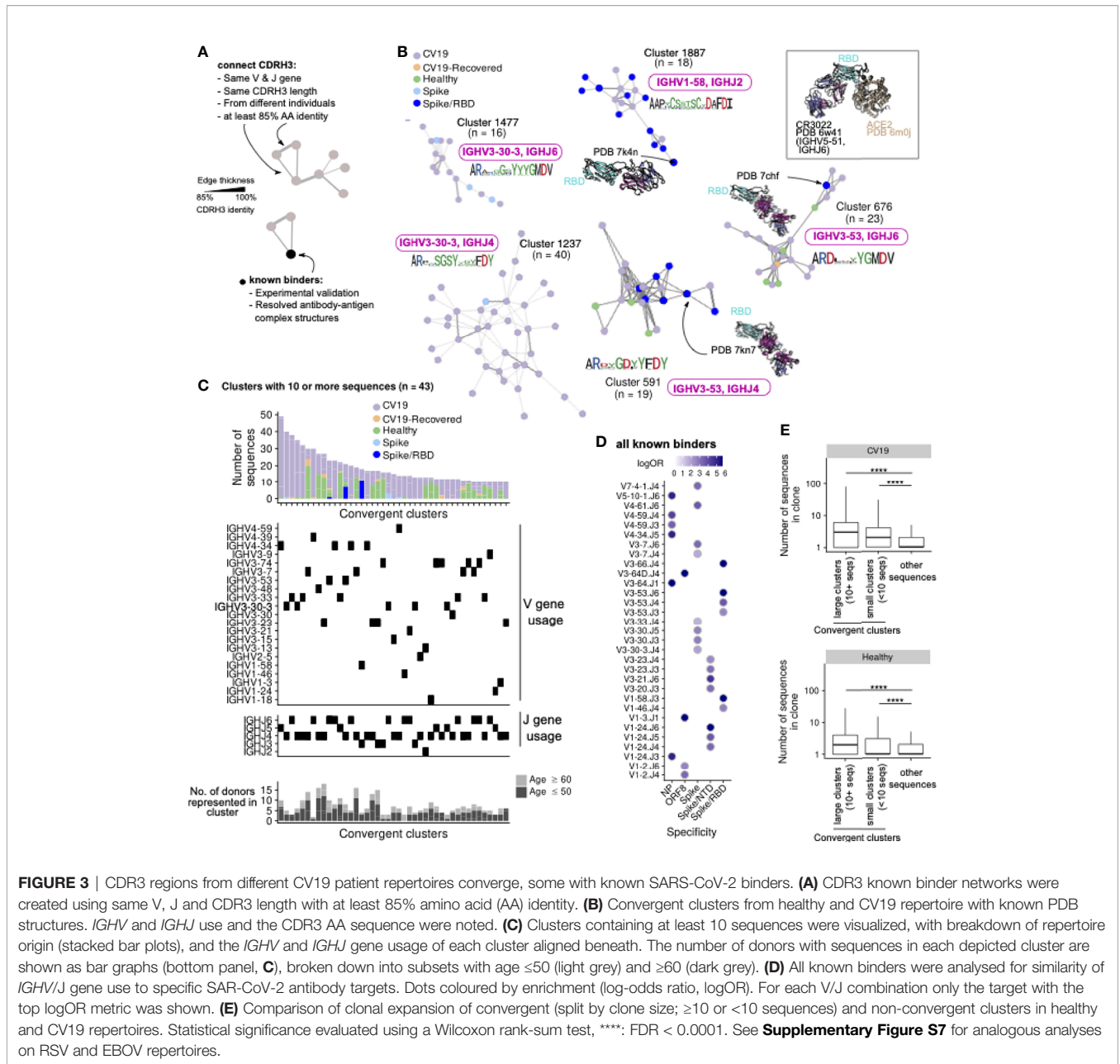
FIGURE 2 | Distinct V-D-J and isotype repertoires in CV19, EBOV and RSV BCR have occasional similarities. **(A, B)** Difference in sub-class use of IgA (panel a) and IgG **(B)** in viral disease and healthy BCR repertoires. **(C)** Clonal expansion of sequences of relevant effector types (as revealed in A) plus unmutated and mutated IgM to identify trends of V gene usage in viral infections. Each bubble sampled to a uniform depth (see Methods), with size proportional to clone size, represents one clone colour-coded by V-family usage. **(D)** Quantification of clonal expansion calculated using the Gini coefficient (see methods), revealed clonally expanded effector populations (more mono-clonal/less diverse, closer to 1) or more diverse clones (closer to 0) in viral infections. Sample types with significant differences (false discovery rate [FDR] < 0.05) compared against Healthy were highlighted in red. Dashed line indicates the median diversity in the Healthy cohort. **(E)** Frequency of selected V-D-J gene usage in different cohorts for all sequences and further subdivided by IgM-mutated, IgA1, IgG1. Bar charts depict gene frequency usage in the Healthy cohort. Bubble plots depict the difference in usage (coloured: blue reduced/red increased) compared to healthy repertoires. **(F)** CDRH3 physicochemical characteristics (represented by Kidera factors) were analysed separated by sub-classes and disease status (Healthy/CV19), and compared using Minkowski distance. Note that IgG1 and IgG3 sequences from CV19 cluster together with IgM (square bracket), away from those of the same sub-classes from healthy individuals (indicated by arrow). Statistical significance in panels a, d and e was evaluated using one-way ANOVA and Dunnett *post-hoc* comparison against the Healthy cohort: p-value was corrected using the Benjamini-Hochberg method and expressed as false discovery rate (FDR), indicated either with colour (panel c), bubble size (panel d) or the symbols under the following scheme: *, FDR < 0.05; **, FDR < 0.01, ***, FDR < 0.001, ****, FDR < 0.0001.

that although there is a similar pattern of CSR preferences in White and West Africans individuals, the overall distance from germline is longer before CSR occurs in West Africans (**Figure 6B**). This may suggest that the ethnic bias in existing immunoglobulin sequence databases has resulted in mis-assignment of germline alleles. No CSR differences were seen in the RSV data.

DISCUSSION

We compared immunoglobulin gene sequences from pandemic (SARS-CoV-2), epidemic (Ebola) and endemic (Respiratory

Syncytial Virus) patients in order to discover features that might distinguish newly emergent and endemic infections. We notice that although the sequencing depth varies between sample types, metrics related to gene usage and mutational levels are typically highly stable at different sequencing depths (**Supplementary Note**), allowing us to perform robust comparisons across different immune challenges. The ability of B cells to generate a highly diverse immunoglobulin repertoire that might bind any antigen, and the diverse functionality of the antibodies produced, is critical for an effective immune response. Repertoire studies aim to identify specific antibodies by looking for biased usage of particular Ig genes, and have been useful in the past (16, 28). However, not all expanded genes encode



specific binders (49) and we need to consider the possibility that expansions found in the midst of an acute response may be a side effect of the disease involving inappropriate expansion of B cells carrying antibodies with off-target effects rather than a specific targeting to the challenge. Repertoire selection is normally a delicate balance between tolerance versus immune response to a pathogen and the inflammatory state of acute disease can upset the balance. Serological studies have shown an increase in autoreactive antibodies, particularly to interferons, during acute COVID-19 for example (54, 55).

Looking across different viral diseases, we found a general increase of *IGHV4-39* use in the repertoire of two different viral diseases (COVID-19 and Ebola). Despite this, only one of our

convergent clusters, dominated by COVID-19 sequences, uses *IGHV4-39* (**Figure 3C**); it is possible that there are unannotated *IGHV4-39* SAR-CoV-2 binders. One single cluster does not, however, explain the larger expansion in *IGHV4-39* use across the COVID-19 or Ebola repertoire. *IGHV4-39* may therefore be involved in the pathogenesis of the disease by promiscuous binding to self-proteins. Alternatively, *IGHV4-39* may simply support a wide range of specific binding properties, supported by the lack of convergence and given it has also been dominant in cancer, bacterial infection, influenza and HIV responses (56–59). Such promiscuous binders would have networks contributed to by more than 1 cohort with 52 networks matching this description in our data. It is also significant across all 64 large clusters 14 were

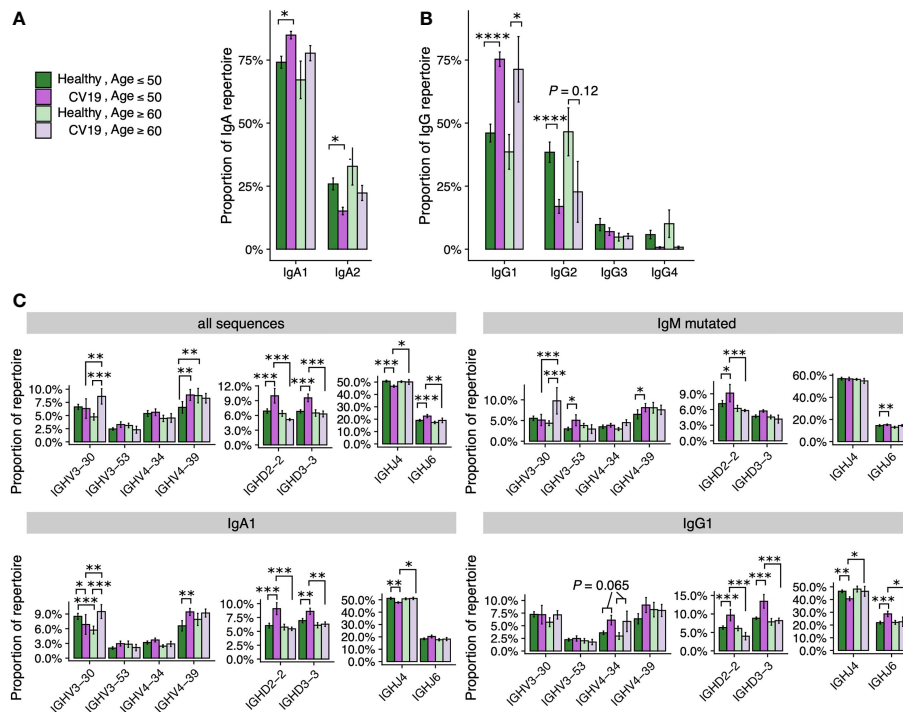


FIGURE 4 | V-D-J and isotype usage repertoire differences between healthy and CV19 are not as apparent in older ages. CV19 and healthy patients were split by over 60s and under 50s and were compared for IgA (A), IgG (B) usage and selected V-D-J gene usage (C). Statistical significance evaluated using two-way ANOVA and Tukey's *post-hoc* test: *, false discovery rate (FDR, corrected using the Benjamini-Hochberg method) < 0.05; **, FDR < 0.01, ***, FDR < 0.001, ****, FDR < 0.0001.

dominated by CV19/CV19-R sequences, yet only 5 matched known binders, suggesting previously unknown SARS-CoV-2 specific binders. As more studies on both naturally infected and vaccinated individuals identify and validate new SARS-CoV-2 binders, we expect more experimental validation to ascertain the relevance of these clonotypes in the immune response. Further data integration exercise incorporating more repertoires (from both healthy and infected/vaccinated individuals) will discover more convergent clusters which richer annotations.

In addition to expansion of gene use as an indicator of activation, we can infer biological information from assessment of the AID-mediated activities of CSR and SHM. These have long been associated with germinal centre formation (60, 61). However, there is mounting evidence that CSR can occur prior to the germinal centre formation (8–11, 36, 53). T-independent activation has been shown to be driven by CD40-independent TLR/TACI activation (62). Our data indicates an early switching to IgG1 without extensive hypermutation. This data is consistent with Woodruff et al. (63) who also found high germline similarity in IgA1, and (42) where COVID-19 samples were found to have more naïve-like characteristics. Our COVID-19 IgA1 sequences also indicate a lower level of hypermutation than the control group, albeit higher than the COVID-19 IgG1, likely reflecting their distance along the CSR hierarchy. Uniquely, our diversity analyses also indicate expansion of unmutated IgM clones. Alongside these data we see that CDRH3 region maturation of

IgG1 and IgG3 genes in the COVID-19 patients is less removed from the IgM state than healthy IgG1 and IgG3, or any other class switched repertoires. Together with the lineage analysis of CSR timing, the whole picture in COVID-19 is of an early immature response of IgM, switching to IgG1 but without much SHM such as might occur in the absence of T cell help in a GC reaction, and then to IgA1. Whether these responses are unique to a live infection or because the virus is so novel is difficult to ascertain, with future vaccine and comparative studies likely to shed further light on this phenomenon. IgG1 is known for its antiviral properties (50, 64) so is expected in this data. The majority of rapid immunological protection assays for COVID-19 focus on IgM or IgG (65–68). Since switching to IgA1 is notable in our data it would be useful for future serology assays to include IgA. Euroimmun's IgA on LFA had one of the highest sensitivities at 87.8% compared to IgM and IgG from other assays (range 43.8–93%, mean 72.5%, median 76%) (67, 69).

It is known that healthy older people generally have more antibodies capable of binding self-proteins (70). The balance between antibodies with positive versus negative/bystander activity may be changed in older patients. We cannot tell this from our data except that we see a higher frequency of known spike binders clustering with COVID-19 repertoires in the younger cohort. Significant age-related differences occur in the dominant IGH COVID-19 genes: the increased use of *IGHV3-30* is only seen in older COVID-19 patients and that of *IGHV4-39*

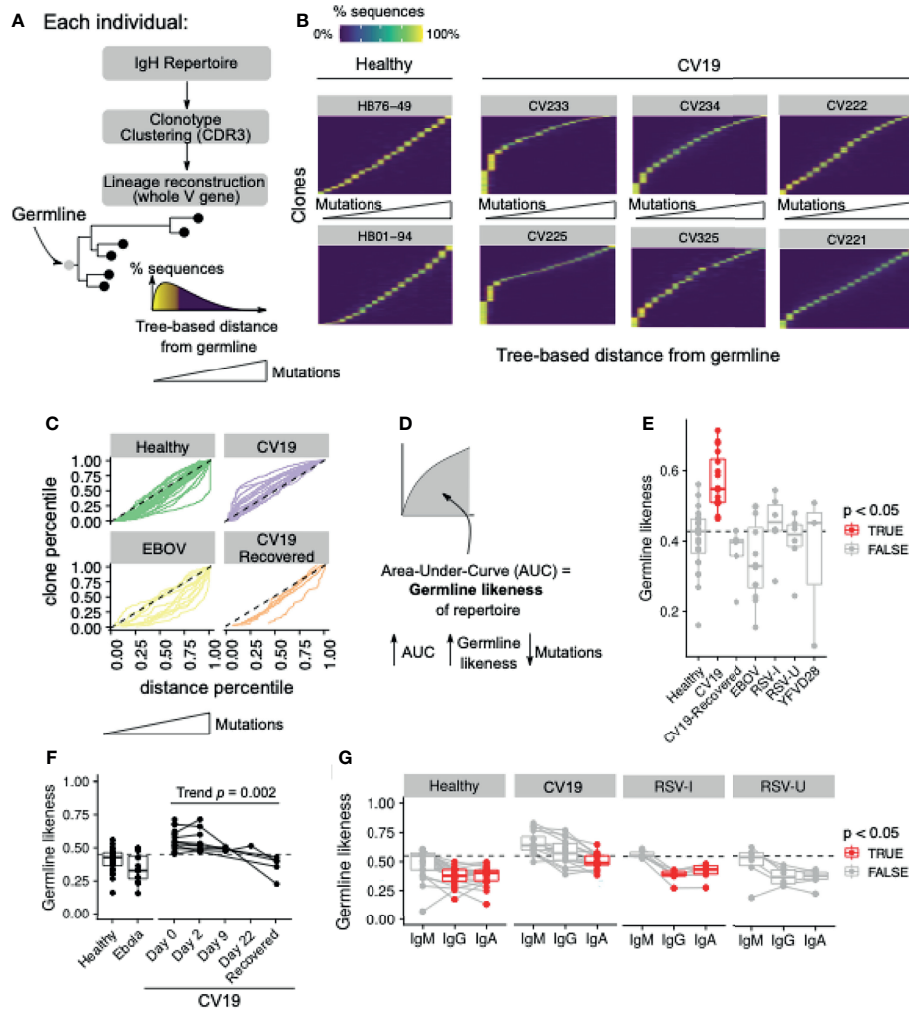


FIGURE 5 | Mutational levels in BCR lineages indicate lower somatic hypermutation in CV19. **(A)** Lineage trees were constructed by clonotyping the IgH CDR3 and the lineages reconstructed using the whole V gene rooting on the predicted germline, allowing the distance from germline to be estimated for each sequence. This allows ordering of sequences based on this distance from germline: depicted as a histogram [(A) bottom right]. **(B)** Clones in the repertoire, for selected donors, were ordered (vertical axis) using median distance from germline (horizontal axis), and the distribution of such distance for each clone was plotted with heatmap colours being the percentage of sequences within the clone containing the a given level of mutation. **(C)** Distance from germline distributions for every donor, split by condition, represented as curves. Dotted line represents the theoretical expectation of mutational level. **(D)** From each of these curves (in c) the area under the curve (AUC) was calculated giving a statistic of ‘Germline Likeness’, a higher AUC resembling more the germline and a lower AUC indicating more mutations. **(E)** Comparison of Germline Likeness between conditions: sample types with significant (Benjamini-Hochberg-corrected false discovery rate [FDR] < 0.05) differences compared to Healthy (Wilcoxon rank-sum test) are highlighted in red with the dotted line being the healthy median. **(F)** The Germline Likeness across timepoints for CV19 patients with Healthy and Ebola data are reproduced here for comparison: trend was evaluated using the Jonckheere-Terpstra test. **(G)** Comparison of germline distance split by immunoglobulin isotype was performed split by cohort: significant (FDR < 0.05) differences compared to IgM (Wilcoxon rank-sum test) are highlighted in red.

only in the younger group. We also see selection for *IGHD2-2*, *IGHD3-3* and *IGHJ6* only in the young, with *IGHJ6* occurring frequently in known binder networks, given the importance of *IGHD* and *IGHJ* genes to the CDRH3 region it is striking that the differences seen here are solely in the younger age group. We note that by separating samples by age group the small sample size limits the potential to discover more gene usage differences with strong statistical significance. In the future integration with other repertoire datasets, including those in response to SARS-

CoV-2 vaccination, would yield more insights into age-related differences in the humoral response against this virus.

In comparison to our CV19 data our EBOV data paints an unusual picture where, even 2-3 months post-recovery with viral negative PCR tests, there are abnormally high proportions of class switched clones with little or no direction towards a particular subclass. Given CSR is largely understudied, as far as we can tell such high rates of class switching, particularly so long after recovery, are entirely unique to this infection. Another unusual observation was that EBOV

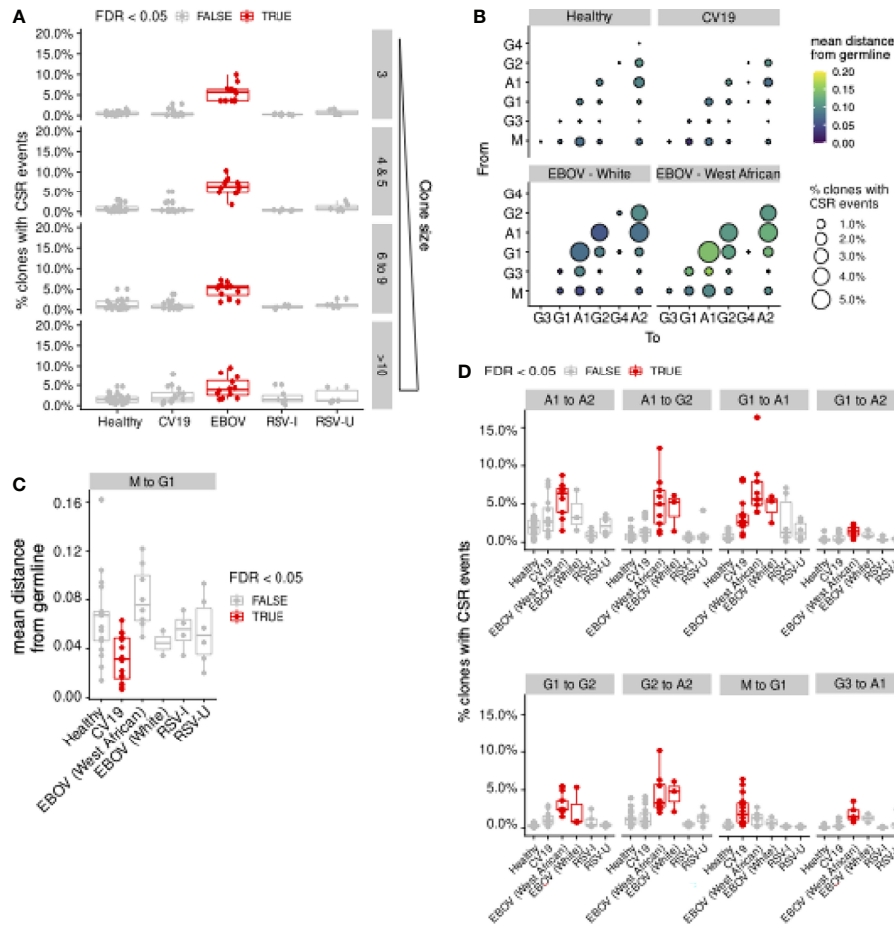


FIGURE 6 | Evidence of Class-switch recombination is increased in EBOV and in large CV19 expansions. Repertoires from each donor were sampled to uniform depth (12,000, approximately the median sequence counts over all donors across all sample types). **(A)** Lineage clones (see Figure 5) were assessed for prevalence of CSR events in terms of the proportion of clones and plotted by clone size and split by condition. **(B)** Bubble plot depicting the frequency and distance-from-germline of CSR events, separated by the CSR start ('From', vertical axis) and end ('To', horizontal axis) isotypes. Quantification was performed separately for different sample types. Bubble sizes are proportional to the frequency of CSR and colour is scaled by distance from germline at which CSR occurs, as estimated from the reconstructed lineage trees. **(C)** Statistical comparison of the median distance from germline at which CSR events occurred across sample types. Each donor was considered separately for every switch possibility. **(D)** Comparison of CSR frequency (percentage of clones with evidence of CSR) for each condition was also assessed for each donor (median, **D**). Statistical significance was evaluated using one-way ANOVA and Dunnett *post-hoc* comparison against Healthy with Benjamini-Hochberg-corrected false discovery rate (FDR) < 0.05 highlighted in red (**C, D**). For **(D)**, **Supplementary Figure S9** contain analogous plots for all CSR combinations with significant (FDR < 0.05) differences compared against Healthy.

survivors' memory B cell populations were more diverse than healthy controls suggesting stimulation with more diverse antigens, or a less structured and directed immune response. A 'decay-stimulation-decay' pattern resulting in the peak of antibody response being some 200 days after infection has previously been reported (71) and cytokine storms during infection may also be contributing to this phenomenon (72, 73). It was not possible to collect blood samples from unrecovered patients, so we do not know if these observations were a requirement of patient recovery or a phenomenon unique to Ebola infection in general. In comparison to other sample types, our EBOV samples were only sequenced to modest depth (median number of sequences for EBOV = 5,346; CV19 = 38020.5; Healthy = 12,486.5). We note that the detection of CSR events within lineages is dependent upon sequencing depth

(**Supplementary Note**); it is possible that the estimates of CSR frequencies for EBOV in **Figure 6** only represent the lower bound, further highlighting the importance of CSR in the response against EBOV.

By comparing examples of pandemic, epidemic and endemic viral disease responses our results show that while aspects of B cell responses are unique to particular infections, the human immunoglobulin gene repertoire can show similarities of response across two very different diseases. There are many questions to be answered about the balance of beneficial versus bystander responses in acute inflammatory diseases, where the initial class switched responses seem to be immature (COVID-19) and possibly unregulated (EBOV infection). Coupled with the finding of genes such as *IGHV4-39* appearing in two

completely different diseases, these data add weight to the hypothesis that an emergency humoral immune response to primary challenge can bypass normal stringent regulation and thus allow the production of autoimmune antibodies.

CODE AVAILABILITY

Functions implemented to generate and analyse lineage trees were included in a R package BrepPhylo, which is available at <https://github.com/Fraternalilab/BrepPhylo>. All other code snippets used in analysing data presented in this work are available as R markdown files available at <https://github.com/Fraternalilab/BrepPhyloAnalysis>.

DATA AVAILABILITY STATEMENT

IMGT/High-VQuest-annotated immunoglobulin sequence data file is available at Zenodo (<https://dx.doi.org/10.5281/zenodo.5146019>).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by UK London REC 14/LO/1221 for COVID-19 samples, UK London REC 11/LO/1826 for RSV samples, and Sierra Leone Ethics and Scientific Review Committee ISRCTN13990511 and PACTR201602001355272 and authorised by Pharmacy Board of Sierra Leone, #PBSL/CTAN/MOHS-CST001 for Ebola samples. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

DD-W, AS, JSO'H, and MB designed study and protocols to collected and analyse COVID-19 and YFV samples under REC 14/LO/1221. AS, MB, NK, ES, KL, CF, CCo, HL and FO acquired and biobanked COVID-19 samples. CCh and PO provided RSV samples. MS, JS, and JB provided Ebola samples. ES, AS, IS, JSO'H, AP, and BW devised repertoire protocol and performed repertoire data generation. JN, AS, ES, DK, CM, DD-W, and FF performed bioinformatic analysis and data interpretation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the MRC (MR/L01257X/1 and MC_PC_15068), BBSRC (BB/T002212/1 and BB/V011456/1), EPSRC (EP/R031118/1) and UK National Core Studies (NCSi4P

REFERENCES

- Seow J, Graham C, Merrick B, Acors S, Steel K, Hemmings O, et al. Longitudinal Evaluation and Decline of Antibody Responses in SARS-CoV-2 Infection. *Nat Microbiol* (2020) 5:2020.07.09.20148429. doi: 10.1101/2020.07.09.20148429

programme) funding. The study “Convalescent plasma for early Ebola virus disease in Sierra Leone” (ISRCTN13990511 and PACTR201602001355272) was supported by the Wellcome Trust (Award 106491) and Bill and Melinda Gates Foundation; Public Health England Ebola Emergency Response; and the Blood Safety Programme, National Health Service Blood and Transplant. MGS was supported by the UK National Institute for Health Research Health Protection Research Unit in Emerging and Zoonotic Infections. J.T.S. is supported by the Wellcome Trust (204820/Z/16/Z). M.G.S. is supported by the UK National Institute for Health Research Health Protection Research Unit in Emerging and Zoonotic Infections. JTS and MGS were awarded a grant from the Enhancing Research Activity in Epidemic Situations (ERAES) program funded by the Wellcome Trust to support further research into the sequelae of Ebola virus disease. We acknowledge the support of the Medical Research Council (G0902266), the Wellcome Trust (087805/Z/08/Z), and Medical Research Council (MRC) EMINENT Network (MR/R502121/1) which is co-funded by GSK. PO is supported by a Senior Investigator award from the National Institute for Health Research (NIHR201385) and RSV Consortium in Europe (RESCEU) Horizon 2020 Framework Grant 116019. PO and CC are supported by the Biomedical Research Centre award to Imperial College Healthcare NHS Trust, Imperial’s Health Protection Research Unit in Respiratory Infections (NIHR HPRU RI), the Comprehensive Local Research Networks (CLRN), and the MRC HIC-Vac network (MR/R005982/1). Infrastructure support was provided by the NIHR Imperial Biomedical Research Centre and the NIHR Imperial Clinical Research Facility. The funders had no role in the collection and analysis of the samples, in the interpretation of data, in writing the report, nor in the decision to submit the paper for publication.

ACKNOWLEDGMENTS

We acknowledge the contribution of the Sierra Leone Association of Ebola Survivors and Ebola CP Consortium Investigators. Ebola samples were kindly donated by Prof Alain Townsend (Oxford). Thanks also to the Surrey COVID-19 Collaboration and all volunteers for sample collection and to Gill Wallace and Vasiliki Tsioligka for technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2022.807104/full#supplementary-material>

- Siggins MK, Thwaites RS, Openshaw PJM. Durability of Immunity to SARS-CoV-2 and Other Respiratory Viruses. *Trends Microbiol* (2021) 29:648–62. doi: 10.1016/j.TIM.2021.03.016
- Gudbjartsson DF, Norddahl GL, Melsted P, Gunnarsdottir K, Holm H, Eythorsson E, et al. Humoral Immune Response to SARS-CoV-2 in Iceland. *N Eng J Med* (2020) 383:1724–34. doi: 10.1056/NEJMOA2026116

4. Chi X, Li Y, Qiu X. V(D)J Recombination, Somatic Hypermutation and Class Switch Recombination of Immunoglobulins: Mechanism and Regulation. *Immunology* (2020) 160:233–47. doi: 10.1111/imm.13176
5. Xu Z, Zan H, Pone EJ, Mai T, Casali P. Immunoglobulin Class-Switch DNA Recombination: Induction, Targeting and Beyond. *Nat Rev Immunol* (2012) 12:517–31. doi: 10.1038/nri3216
6. Stavnezer J, Guikema JE, Schrader CE. Mechanism and Regulation of Class Switch Recombination. *Annu Rev Immunol* (2008) 26:261–92. doi: 10.1146/annurev.immunol.26.021607.090248
7. Pham P, Bransteitter R, Petruska J, Goodman MF. Processive AID-Catalysed Cytosine Deamination on Single-Stranded DNA Simulates Somatic Hypermutation. *Nature* (2003) 424:103–7. doi: 10.1038/nature01760
8. Roco JA, Mesin L, Binder SC, Nefzger C, Gonzalez-Figueroa P, Canete PF, et al. Class-Switch Recombination Occurs Infrequently in Germinal Centers. *Immunity* (2019) 51:337–50.e7. doi: 10.1016/j.immuni.2019.07.001
9. Pape KA, Kouskoff V, Nemazee D, Tang HL, Cyster JG, Tze LE, et al. Visualization of the Genesis and Fate of Isotype-Switched B Cells During a Primary Immune Response. *J Exp Med* (2003) 197:1677–87. doi: 10.1084/jem.20012065
10. Toellner KM, Gulbranson-Judge A, Taylor DR, Sze DMY, MacLennan IC. Immunoglobulin Switch Transcript Production *In Vivo* Related to the Site and Time of Antigen-Specific B Cell Activation. *J Exp Med* (1996) 183:2303–12. doi: 10.1084/jem.183.5.2303
11. King HW, Orban N, Riches JC, Clear AJ, Warnes G, Teichmann SA, et al. Single-Cell Analysis of Human B Cell Maturation Predicts How Antibody Class Switching Shapes Selection Dynamics. *Sci Immunol* (2021) 6:eabe6291. doi: 10.1126/sciimmunol.abe6291
12. Dunn-Walters DK, Stewart AT, Sinclair EL, Serangeli I. Age-Related Changes in B Cells Relevant to Vaccine Responses. *Interdiscip Topics Gerontol Geriatr* (2020) 43:56–72. doi: 10.1159/000504479
13. Lord JM. The Effect of Aging of the Immune System on Vaccination Responses. *Hum Vaccines Immunother* (2013) 9:1364–7. doi: 10.4161/hv.24696
14. Ademokun A, Wu Y-C, Martin V, Mitra R, Sack U, Baxendale H, et al. Vaccination-Induced Changes in Human B-Cell Repertoire and Pneumococcal IgM and IgA Antibody at Different Ages. *Aging Cell* (2011) 10:922–30. doi: 10.1111/j.1474-9726.2011.00732.x
15. Colonna-Romano G, Bulati M, Aquino A, Pellicanò M, Vitello S, Lio D, et al. A Double-Negative (IgD-CD27-) B Cell Population Is Increased in the Peripheral Blood of Elderly People. *Mech Ageing Dev* (2009) 130:681–90. doi: 10.1016/j.mad.2009.08.003
16. Wu YC, Kipling D, Leong HS, Martin V, Ademokun AA, Dunn-Walters DK. High-Throughput Immunoglobulin Repertoire Analysis Distinguishes Between Human IgM Memory and Switched Memory B-Cell Populations. *Blood* (2010) 116:1070–8. doi: 10.1182/blood-2010-03-275859
17. Taylor JJ, Pape KA, Jenkins MK. A Germinal Center-Independent Pathway Generates Unswitched Memory B Cells Early in the Primary Response. *J Exp Med* (2012) 209:597–606. doi: 10.1084/jem.20111696
18. Allman D, Wilmore JR, Gaudette BT. The Continuing Story of T-Cell Independent Antibodies. *Immunol Rev* (2019) 288:128–35. doi: 10.1111/imr.12754
19. Bryan Wu YC, Kipling D, Dunn-Walters DK. Age-Related Changes in Human Peripheral Blood IGH Repertoire Following Vaccination. *Front Immunol* (2012) 3:193. doi: 10.3389/fimmu.2012.00193
20. Townsend CL, Laffy MJ, Wu YCB, O'Hare JS, Martin V, Kipling D, et al. Significant Differences in Physicochemical Properties of Human Immunoglobulin Kappa and Lambda CDR3 Regions. *Front Immunol* (2016) 7:388. doi: 10.3389/fimmu.2016.00388
21. Dunn-Walters D, Townsend C, Sinclair E, Stewart A. Immunoglobulin Gene Analysis as a Tool for Investigating Human Immune Responses. *Immunol Rev* (2018) 284:132–47. doi: 10.1111/imr.12659
22. Perchiacca JM, Ladiwala ARA, Bhattacharya M, Tessier PM. Aggregation-Resistant Domain Antibodies Engineered With Charged Mutations Near the Edges of the Complementarity-Determining Regions. *Protein Engineering Design Selection* (2012) 25:591–601. doi: 10.1093/protein/gzs042
23. Laffy J, Dodev T, Macpherson J, Townsend C, Lu H, Dunn-Walters D, et al. Promiscuous Antibodies Characterised by Their Physico-Chemical Properties: From Sequence to Structure and Back. *Prog Biophys Mol Biol* (2017) 128:47–56. doi: 10.1016/j.PBIOMOLBIO.2016.09.002
24. Stadlbauer D, Amanat F, Chromikova V, Jiang K, Strohmeier S, Arunkumar GA, et al. SARS-CoV-2 Seroconversion in Humans: A Detailed Protocol for a Serological Assay, Antigen Production, and Test Setup. *Curr Protoc Microbiol* (2020) 57:e100. doi: 10.1002/cpmc.100
25. Tedder RS, Samuel D, Dicks S, Scott JT, Ijaz S, Smith CC, et al. Detection, Characterization, and Enrollment of Donors of Ebola Convalescent Plasma in Sierra Leone. *Transfusion* (2018) 58:1289–98. doi: 10.1111/TRF.14580
26. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, et al. Risk Stratification of Patients Admitted to Hospital With Covid-19 Using the ISARIC WHO Clinical Characterisation Protocol: Development and Validation of the 4C Mortality Score. *BMJ* (2020) 370:m3339. doi: 10.1136/bmj.m3339
27. Habibi MS, Jozwik A, Makris S, Dunning J, Paras A, DeVincenzo JP, et al. Impaired Antibody-Mediated Protection and Defective Iga B-Cell Memory in Experimental Infection of Adults With Respiratory Syncytial Virus. *Am J Respir Crit Care Med* (2015) 191:1040–9. doi: 10.1164/rccm.201412-2256OC
28. Martin V, Wu YC, Kipling D, Dunn-Walters DK. Age-Related Aspects of Human IgM+ B Cell Heterogeneity. *Ann New York Acad Sci* (2015) 1362:153–63. doi: 10.1111/nyas.12823
29. Osorio D, Rondón-Villarreal P, Torres R. Peptides: A Package for Data Mining of Antimicrobial Peptides. *R J* (2015) 7:4–14. doi: 10.32614/RJ-2015-001
30. Bashford-Rogers R, Bergamaschi L, McKinney E, Pombal D, Mescia F, Lee J, et al. B Cell Receptor Repertoire Analysis in Six Immune-Mediated Diseases. *Nature* (2019) 574:122. doi: 10.1038/S41586-019-1595-3
31. Stephenson E, Reynolds G, Botting RA, Calero-Nieto FJ, Morgan MD, Tuong ZK, et al. Single-Cell Multi-Omics Analysis of the Immune Response in COVID-19. *Nat Med* (2021) 27:904–16. doi: 10.1038/s41591-021-01329-2
32. Bashford-Rogers RJM, Palser AL, Huntly BJ, Rance R, Vassiliou GS, Follows GA, et al. Network Properties Derived From Deep Sequencing of Human B-Cell Receptor Repertoires Delineate B-Cell Populations. *Genome Res* (2013) 23:1874. doi: 10.1101/GR.154815.113
33. Felsenstein J. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* (1989) 5:164–6.
34. Stern JNH, Yaari G, Vander HJA, Church G, Donahue WF, Hintzen RQ, et al. B Cells Populating the Multiple Sclerosis Brain Mature in the Draining Cervical Lymph Nodes. *Sci Trans Med* (2014) 6:248ra107–248ra107. doi: 10.1126/SCITRANSLMED.3008879
35. Unterman A, Sumida TS, Nouri N, Yan X, Zhao AY, Gasque V, et al. Single-Cell Multi-Omics Reveals Dyssynchrony of the Innate and Adaptive Immune System in Progressive COVID-19. *Nat Commun* (2022) 13:1–23. doi: 10.1038/s41467-021-27716-4
36. Feldman J, Bals J, Altomare CG, Denis KS, Lam EC, Hauser BM, et al. Naive Human B Cells Engage the Receptor Binding Domain of SARS-CoV-2, Variants of Concern, and Related Sarbecoviruses. *Sci Immunol* (2021) 6:5842. doi: 10.1126/SCIIMMUNOL.ABL5842/SUPPL_FILE/SCIIMMUNOL.ABL5842_DATA_FILE_S1.ZIP
37. Dunbar J, Krawczyk K, Leem J, Baker T, Fuchs A, Georges G, et al. SabDab: The Structural Antibody Database. *Nucleic Acids Res* (2014) 42:D1140–6. doi: 10.1093/NAR/GKT1043
38. Zost SJ, Gilchuk P, Chen RE, Case JB, Reidy JX, Trivette A, et al. Rapid Isolation and Profiling of a Diverse Panel of Human Monoclonal Antibodies Targeting the SARS-CoV-2 Spike Protein. *Nat Med* (2020) 26:1422–7. doi: 10.1038/s41591-020-0998-x
39. Robbiani DF, Gaebler C, Muecksch F, Lorenzi JCC, Wang Z, Cho A, et al. Convergent Antibody Responses to SARS-CoV-2 in Convalescent Individuals. *Nature* (2020) 584:437–42. doi: 10.1038/s41586-020-2456-9
40. Kreer C, Zehner M, Weber T, Ercanoglu MS, Gieselmann L, Rohde C, et al. Longitudinal Isolation of Potent Near-Germline SARS-CoV-2-Neutralizing Antibodies From COVID-19 Patients. *Cell* (2020) 182:843–854.e12. doi: 10.1016/j.cell.2020.06.044
41. Gaebler C, Wang Z, Lorenzi JCC, Muecksch F, Fink S, Tokuyama M, et al. Evolution of Antibody Immunity to SARS-CoV-2. *Nature* (2021) 591:639–44. doi: 10.1038/s41586-021-03207-w
42. Dugan HL, Stamper CT, Li L, Changrob S, Asby NW, Halfmann PJ, et al. Profiling B Cell Immunodominance After SARS-CoV-2 Infection Reveals Antibody Evolution to Non-Neutralizing Viral Targets. *Immunity* (2021) 54:1290–303.e7. doi: 10.1016/j.immuni.2021.05.001

43. Brouwer PJM, Caniels TG, van der Straten K, Snitselaar JL, Aldon Y, Bangaru S, et al. Potent Neutralizing Antibodies From COVID-19 Patients Define Multiple Targets of Vulnerability. *Science* (2020) 369:643–50. doi: 10.1126/science.abc5902
44. Zhang Y, Skolnick J. Scoring Function for Automated Assessment of Protein Structure Template Quality. *Proteins: Struct Function Bioinf* (2004) 57:702–10. doi: 10.1002/PROT.20264
45. Cortjens B, Yasuda E, Yu X, Wagner K, Claassen Y, Bakker A, et al. Broadly Reactive Anti-Respiratory Syncytial Virus G Antibodies From Exposed Individuals Effectively Inhibit Infection of Primary Airway Epithelial Cells. *J Virol* (2017) 91:e02357–16. doi: 10.1128/JVI.02357-16
46. Goodwin E, Gilman M, Wrapp D, Chen M, Ngwuta J, Moin S, et al. Infants Infected With Respiratory Syncytial Virus Generate Potent Neutralizing Antibodies That Lack Somatic Hypermutation. *Immunity* (2018) 48:339–49.e5. doi: 10.1016/J.IMMUNI.2018.01.005
47. Gilman MSA, Castellanos CA, Chen M, Ngwuta JO, Goodwin E, Moin SM, et al. Rapid Profiling of RSV Antibody Repertoires From the Memory B Cells of Naturally Infected Adult Donors. *Sci Immunol* (2016) 1:eaaj1879. doi: 10.1126/SCIIMMUNOL.AAJ1879
48. Davis C, Jackson K, McElroy A, Halfmann P, Huang J, Chennareddy C, et al. Longitudinal Analysis of the Human B Cell Response to Ebola Virus Infection. *Cell* (2019) 177:1566–82.e17. doi: 10.1016/J.CELL.2019.04.036
49. Margreitter C, Lu H-C, Townsend C, Stewart A, Dunn-Walters DK, Fraternali F. BReptoire: A User-Friendly Web Server for Analysing Antibody Repertoire Data. *Nucleic Acids Res* (2018) 46:W264–70. doi: 10.1093/NAR/GKY276
50. Vidarsson G, Dekkers G, Rispens T. IgG Subclasses and Allotypes: From Structure to Effector Functions. *Front Immunol* (2014) 5:520. doi: 10.3389/fimmu.2014.00520
51. Smith SA, Burton SL, Kilembe W, Lakhi S, Karita E, Price M, et al. VH1-69 Utilizing Antibodies Are Capable of Mediating Non-Neutralizing Fc-Mediated Effector Functions Against the Transmitted/Founder Gp120. *Front Immunol* (2019) 0:3163. doi: 10.3389/FIMMU.2018.03163
52. Avnir Y, Watson C, Glanville J, Peterson E, Tallarico A, Bennett A, et al. IGHV1-69 Polymorphism Modulates Anti-Influenza Antibody Repertoires, Correlates With IGHV Utilization Shifts and Varies by Ethnicity. *Sci Rep* (2016) 6:1–13. doi: 10.1038/SREP20842
53. Nielsen SCA, Yang F, Jackson KJL, Hoh RA, Röltgen K, Jean GH, et al. Human B Cell Clonal Expansion and Convergent Antibody Responses to SARS-CoV-2. *Cell Host Microbe* (2020) 28:516–25.e5. doi: 10.1016/J.CHOM.2020.09.002
54. Richter AG, Shields AM, Karim A, Birch D, Faustini SE, Steadman L, et al. Establishing the Prevalence of Common Tissue-Specific Autoantibodies Following Severe Acute Respiratory Syndrome Coronavirus 2 Infection. *Clin Exp Immunol* (2021) 205:99–105. doi: 10.1111/CEL.13623
55. Bastard P, Rosen LB, Zhang Q, Michailidis E, Hoffmann H-H, Zhang Y, et al. Autoantibodies Against Type I IFNs in Patients With Life-Threatening COVID-19. *Science* (2020) 370:eabd4585. doi: 10.1126/SCIENCE.ABD4585
56. Rossi D, Spina V, Bomben R, Rasi S, Dal-Bo M, Bruscazzin A, et al. Association Between Molecular Lesions and Specific B-Cell Receptor Subsets in Chronic Lymphocytic Leukemia. *Blood* (2013) 121:4902–5. doi: 10.1182/blood-2013-02-486209
57. Yamayoshi S, Yasuhara A, Ito M, Uraki R, Kawaoka Y. Differences in the Ease With Which Mutant Viruses Escape From Human Monoclonal Antibodies Against the HA Stem of Influenza A Virus. *J Clin Virol* (2018) 108:105–11. doi: 10.1016/j.jcv.2018.09.016
58. Morris Abdool Karim L, Moore PL, Travers SA, Garrett N, Abdool Karim Q, Dshanta Naicker SS, et al. Germline Ig Gene Repertoire HIV-1 Antibodies Is Not Restricted by the Ability To Develop Broadly Neutralizing. *J Immunol* (2021) 194:4371–8. doi: 10.4049/jimmunol.1500118
59. Yeung YA, Foletti D, Deng X, Abdiche Y, Strop P, Glanville J, et al. Germline-Encoded Neutralization of a Staphylococcus Aureus Virulence Factor by the Human Antibody Repertoire. *Nat Commun* (2016) 7:1–14. doi: 10.1038/ncomms13376
60. Muramatsu M, Kinoshita K, Fagarasan S, Yamada S, Shinkai Y, Honjo T. Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. *Cell* (2000) 102:553–63. doi: 10.1016/S0092-8674(00)00078-7
61. Muramatsu M, Sankaranand VS, Anant S, Sugai M, Kinoshita K, Davidson NO, et al. Specific Expression of Activation-Induced Cytidine Deaminase (AID), a Novel Member of the RNA-Editing Deaminase Family in Germinal Center B Cells. *J Biol Chem* (1999) 274:18470–6. doi: 10.1074/jbc.274.26.18470
62. Pone EJ, Zhang J, Mai T, White CA, Li G, Sakakura JK, et al. BCR-Signalling Synergizes With TLR-Signalling for Induction of AID and Immunoglobulin Class-Switching Through the Non-Canonical NF- κ b Pathway. *Nat Commun* (2012) 3:1–12. doi: 10.1038/ncomms1769
63. Woodruff MC, Ramonell RP, Nguyen DC, Cashman KS, Saini AS, Haddad NS, et al. Extrafollicular B Cell Responses Correlate With Neutralizing Antibodies and Morbidity in COVID-19. *Nat Immunol* (2020) 21:1506–16. doi: 10.1038/s41590-020-00814-z
64. Ferrante A, Beard LJ, Feldman RG. IgG Subclass Distribution of Antibodies to Bacterial and Viral Antigens. *Pediatr Infect Dis J* (1990) 9:S16–24. doi: 10.1097/00006454-199008001-00004
65. Higgins RL, Rawlings SA, Case J, Lee FY, Chan CW, Barrick B, et al. Longitudinal SARS-CoV-2 Antibody Study Using the Easy Check COVID-19 IgM/IgG™ Lateral Flow Assay. *PLoS One* (2021) 16:e0247797. doi: 10.1371/journal.pone.0247797
66. Wu JL, Tseng WP, Lin CH, Lee TF, Chung MY, Huang CH, et al. Four Point-of-Care Lateral Flow Immunoassays for Diagnosis of COVID-19 and for Assessing Dynamics of Antibody Responses to SARS-CoV-2. *J Infect* (2020) 81:435–42. doi: 10.1016/j.jinf.2020.06.023
67. Flower B, Brown JC, Simmons B, Moshe M, Frise R, Penn R, et al. Clinical and Laboratory Evaluation of SARS-CoV-2 Lateral Flow Assays for Use in a National COVID-19 Seroprevalence Survey. *Thorax* (2020) 75:1082–8. doi: 10.1136/thoraxjnl-2020-215732
68. Whitman JD, Hiatt J, Mowery CT, Shy BR, Yu R, Yamamoto TN, et al. Evaluation of SARS-CoV-2 Serology Assays Reveals a Range of Test Performance. *Nat Biotechnol* (2020) 38:1174–83. doi: 10.1038/s41587-020-0659-0
69. Jääskeläinen AJ, Kuivanen S, Kekäläinen E, Ahava MJ, Loginov R, Kallio-Kokko H, et al. Performance of Six SARS-CoV-2 Immunoassays in Comparison With Microneutralisation. *J Clin Virol* (2020) 129:104512. doi: 10.1016/j.jcv.2020.104512
70. Dunn-Walters DK. The Ageing Human B Cell Repertoire: A Failure of Selection? *Clin Exp Immunol* (2016) 183:50–6. doi: 10.1111/CEL.12700
71. Adaken C, Scott JT, Sharma R, Gopal R, Dicks S, Niazi S, et al. Ebola Virus Antibody Decay–Stimulation in a High Proportion of Survivors. *Nature* (2021) 590:468–72. doi: 10.1038/s41586-020-03146-y
72. Zampieri CA, Sullivan NJ, Nabel GJ. Immunopathology of Highly Virulent Pathogens: Insights From Ebola Virus. *Nat Immunol* (2007) 8:1159–64. doi: 10.1038/ni1519
73. Kennedy JR. Phosphatidylserine's Role in Ebola's Inflammatory Cytokine Storm and Hemorrhagic Consumptive Coagulopathy and the Therapeutic Potential of Annexin V. *Med Hypotheses* (2020) 135:109462. doi: 10.1016/j.mehy.2019.109462

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Stewart, Sinclair, Ng, O'Hare, Page, Serangeli, Margreitter, Orsenigo, Longman, Frampas, Costa, Lewis, Kasar, Wu, Kipling, Openshaw, Chiu, Baillie, Scott, Semple, Bailey, Fraternali and Dunn-Walters. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.