

Deep sequencing of B cell receptor repertoires from COVID-19 patients reveals strong convergent immune signatures

Jacob D. Galson^{1*}, Sebastian Schaetzle¹, Rachael J. M. Bashford-Rogers^{1,2}, Matthew I. J. Raybould³, Aleksandr Kovaltsuk³, Gavin J. Kilpatrick¹, Ralph Minter¹, Donna K. Finch¹, Jorge Dias¹, Louisa James⁴, Gavin Thomas⁴, Wing-Yiu Jason Lee⁴, Jason Betley⁵, Olivia Cavlan¹, Alex Leech¹, Charlotte M. Deane³, Joan Seoane⁶, Carlos Caldas⁷, Dan Pennington⁴, Paul Pfeffer⁴, Jane Osbourn^{1*}

¹ Alchemab Therapeutics Ltd, 55-56 Russell Square, London, WC1B 4HP, UK

² Wellcome Centre for Human Genetics, Roosevelt Dr, Oxford, OX3 7BN, UK

³ Oxford Protein Informatics Group, Department of Statistics, University of Oxford, Oxford OX1 3LB, UK

⁴ Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AD, UK

⁵ Illumina, Inc., Illumina Centre, Granta Park, Cambridge CB21 6DF, UK

⁶ Translational Research Program, Vall d'Hebron Institute of Oncology, Barcelona 08035, Spain

⁷ Cancer Research UK Cambridge Institute and Department of Oncology, Li Ka Shing Centre, University of Cambridge, Cambridge CB2 0RE, UK

*Corresponding authors: jake@alchemab.com, jane@alchemab.com

Abstract

Deep sequencing of B cell receptor (BCR) heavy chains from a cohort of 19 COVID-19 patients from the UK reveals a stereotypical naive immune response to SARS-CoV-2 which is consistent across patients and may be a positive indicator of disease outcome. Clonal expansion of the B cell memory response is also observed and may be the result of memory bystander effects. There was a strong convergent sequence signature across patients, and we identified 777 clonotypes convergent between at least four of the COVID-19 patients, but not present in healthy controls. A subset of the convergent clonotypes were homologous to known SARS and SARS-CoV-2 spike protein neutralising antibodies. Convergence was also demonstrated across wide geographies by comparison of data sets between patients from UK, USA and China, further validating the disease association and consistency of the stereotypical immune response even at the sequence level. These convergent clonotypes provide a resource to identify potential therapeutic and prophylactic antibodies and demonstrate the potential of BCR profiling as a tool to help understand and predict positive patient responses.

Key words: COVID-19, SARS-CoV-2, B cell repertoire, antibody

45 Introduction

46
47 Since the report of the first patients in December 2019 ^{1,2}, the unprecedented global scale of
48 the COVID-19 pandemic has become apparent. The infectious agent, the SARS-CoV-2
49 betacoronavirus ³, causes mild symptoms in most cases but can cause severe respiratory
50 diseases such as acute respiratory distress syndrome in some individuals. Risk factors for
51 severe disease include age, male gender and underlying co-morbidities ⁴.

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

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

71 Patients recovering from COVID-19 have already been screened to identify
72 neutralising antibodies, following analysis of relatively small numbers (100-500) of antibody
73 sequences ^{12,13}. A more extensive BCR repertoire analysis was performed on six patients in
74 Stanford, USA with signs and symptoms of COVID-19 who also tested positive for SARS-CoV-
75 2 RNA ¹⁴. Although no information was provided on the patient outcomes in that study, the
76 analysis demonstrated preferential expression of a subset of immunoglobulin heavy chain
77 (IGH) V gene segments with relatively little somatic hypermutation and showed evidence of
78 convergent antibodies between patients.

79 To drive a deeper understanding of the nature of humoral immunity to SARS-CoV-2
80 infection and to identify potential therapeutic antibodies to SARS-CoV-2, we have evaluated
81 the BCR heavy chain repertoire from 19 individuals at various stages of their immune
82 response. We show that (1) there are stereotypic responses to SARS-CoV-2 infection, (2)
83 infection stimulates both naïve and memory B cell responses, (3) sequence convergence can
84 be used to identify putative SARS-CoV-2 specific antibodies, and (4) sequence convergence
85 can be identified between different SARS-CoV-2 studies in different locations and using
86 different sample types.

87

88 Results

89

90 COVID-19 disease samples

91 Blood samples were collected from n=19 patients admitted to hospital with acute COVID-19
92 pneumonia. The mean age of patients was 50.2 (SD 18.5) years and 13 (68%) were male. All
93 patients had a clinical history consistent with COVID-19 and typical radiological changes.
94 Seventeen patients had a confirmatory positive PCR test for SARS-CoV-2. The patients
95 experienced an average of 11 days (range 4-20) of symptoms prior to the day on which the
96 blood sample was collected. Nine of the patients were still requiring hospital care but not
97 oxygen therapy on day of sample collection (WHO Ordinal Scale Score 3), while eight were
98 hospitalised requiring oxygen by conventional mask or nasal prongs (WHO Ordinal Scale
99 Score 4) and two were hospitalised with severe COVID-19 pneumonia requiring high-flow
100 nasal oxygen (WHO Ordinal Scale Score 5). On the day of sample collection, the direct
101 clinical care team considered two patients to be deteriorating, four improving and the
102 remaining thirteen were clinically stable.
103

104 [SARS-CoV-2 infection results in a stereotypic B cell response](#)

105 IGHA and IGHG BCR sequencing yielded on average 135,437 unique sequences, and 23,742
106 clonotypes per sample (Supplementary Table 1). To characterise the B cell response in
107 COVID-19, we compared this BCR repertoire data to BCR repertoire data from healthy
108 controls obtained in a separate study ¹⁵. Comparing IGHV gene segment usage revealed a
109 significantly different IGHV gene usage in COVID-19 patients compared to the healthy
110 controls, most notably with increases in the usage of IGHV2-5 (2.6x IGHA, 1.0x IGHG
111 increase), IGHV2-70 (4.6x IGHA, 4.1x IGHG increase), IGHV3-30 (2.0x IGHA, 1.4x IGHG
112 increase), IGHV5-51 (3.5x IGHA, 2.0x IGHG increase), and IGHV4-34 (1.4x IGHA, 2.4x IGHG
113 increase) in the COVID-19 patients (Figure 1A). All of these V gene segments have been
114 previously observed in SARS-CoV-1 or SARS-CoV-2 specific antibodies¹⁶. IGHV4-34 has been
115 shown to bind both autoantigens ¹⁷ and commensal bacteria ¹⁸ and has been associated
116 with SLE ¹⁹. Our data extends this, showing that the proportion of sequences containing the
117 autoreactive AVY & NHS sequence motifs within the IGHV region is significantly more
118 frequent in improving COVID-19 patients compared to stable or deteriorating COVID-19
119 patients, specifically in the IGHG1 isotype subclass (p-value = 0.038; Supplementary Figure
120 2).

121 Comparing isotype subclasses showed a significant increase in the relative usage of
122 IGHA1 and IGHG1 in COVID-19 patients (Figure 1B) - these are the two first isotype
123 subclasses that are switched to upon activation of IGHM ²⁰. There was also an increase in
124 the mean CDRH3 length of the BCRs in the COVID-19 patients, that was most pronounced in
125 the IGHA1, IGHA2 and IGHG1 isotype subclasses (Figure 1C).
126

127 [SARS-CoV-2 infection stimulates both naïve and memory responses](#)

128 To further investigate the COVID-19-specific B cell response, we analysed the characteristics
129 of the BCR sequences that are consistent with recent B cell activation – somatic
130 hypermutation, and clonal expansion. In healthy controls, for class-switched sequences,
131 there is a clear unimodal distribution of sequences with different numbers of mutations,
132 and a mean mutation count across IGHA and IGHG isotypes of 17.6 (Figure 2A). In the
133 COVID-19 samples, the mean mutation count was 14.4, and there was a bimodal
134 distribution with a separate peak of sequences with no mutations. This bimodal distribution
135 was most pronounced in the IGHG1, IGHG3, and IGHA1 isotype subclasses, corresponding to
136 the increased isotype usages. Such a distribution is consistent with an expansion of recently
137 class-switched B cells that have yet to undergo somatic hypermutation. There was

138 considerable variation between participants in the proportion of unmutated sequences
139 (Supplementary Figure 1), which had no significant correlation with the number of days
140 since symptom onset ($R = 0.09$, $p = 0.72$), but was lower in the deteriorating compared to
141 improving patients (Figure 2B)

142 To investigate differential clonal expansion between patients, the Shannon diversity
143 index of each repertoire was calculated (while accounting for differences in read depth
144 through subsampling). A more diverse repertoire is indicative of a greater abundance of
145 different clonal expansions. The BCR repertoires of the COVID-19 patients were significantly
146 more diverse than the BCR repertoires of the healthy controls (Figure 2C); this increase in
147 diversity was positively correlated with an increased proportion of unmutated sequences (R
148 = 0.44, $p = 0.061$; Figure 2D). Interestingly, when we investigated the largest clonal
149 expansions, despite having a more diverse repertoire, the largest clonal expansions in the
150 COVID-19 samples were larger than in the healthy controls (Figure 2E). These large clonal
151 expansions were also highly mutated and had similar levels of mutation between the
152 COVID-19 samples and the healthy controls (Figure 2F).

153

154 Sequence convergence can be used to identify putative SARS-CoV-2 specific
155 antibodies

156 Given the skewing of the B cell response in the COVID-19 patients to specific IGHV genes,
157 we next investigated whether the same similarity was also seen on the BCR sequence level
158 between different participants. Such convergent BCR signatures have been observed in
159 response to other infectious diseases²¹, and may be used to identify disease-specific
160 antibody sequences.

161 Of the 435,420 total clonotypes across all the COVID-19 patients, 9,646 (2.2%) were
162 shared between at least two of the participants (Figure 3A). As convergence could occur by
163 chance or be due to an unrelated memory response from commonly encountered
164 pathogens, the healthy control dataset was used to subtract irrelevant BCR sequences. Of
165 the 9,646 convergent clonotypes, 1,442 (14.9%) were also present in at least one of the 40
166 healthy control samples. As expected, of the convergent clonotypes that were also present
167 in the healthy control samples, the mean mutation count was significantly greater (Figure
168 3B), and the mean CDRH3 length significantly shorter (Figure 3C) than the convergent
169 clonotypes that were unique to the COVID-19 patients.

170 To identify a set of SARS-CoV-2-specific antibody sequences with high confidence,
171 we identified 777 convergent clonotypes that were shared between at least four of the
172 COVID-19 patients, but not seen in the healthy controls. In parallel, for a comparison of
173 convergent signatures, we performed the same analysis on a cohort of seven metastatic
174 breast cancer patient biopsy samples²², which identified 469 convergent clonotypes. These
175 convergent clonotypes were highly specific to each disease cohort (Figure 3D). The 777
176 COVID-19 convergent clonotypes had low mutation levels, with a mean mutation count of 2,
177 and only 51 clonotypes with a mean mutation greater than 5. The sequences within the
178 convergent clonotypes were primarily of the IGHG1 (70%) and IGHA1 (16%) subclasses
179 (Supplementary Figure 3A). The convergent clonotypes used a diversity of IGHV gene
180 segments, with IGHV3-30, IGHV3-30-3 and IGHV3-33 as the most highly represented
181 (Supplementary Figure 3B). This IGHV gene usage distribution differs between that of the
182 total repertoire, and IGHV3-30 is also the most highly used IGHV gene in the CoV-AbDab¹⁶.

183 We next tested whether these convergent clonotypes correlated with disease
184 severity. Indeed, 25 of these convergent clonotypes were found to associate with clinical

185 symptoms after correcting for multiple testing, of which 22 were observed at a significantly
186 higher frequency in improving patients (Figure 3E and Supplementary Figure 4). This is a
187 significantly higher proportion associated with clinical symptoms compared to that expected
188 by chance (p-value = 0.018 by random permutations of labels). Interestingly, some of these
189 clonotypes are common in patients comprising >0.1 % of a patient's repertoire.
190 Furthermore, the convergent clonotypes that are associated with clinical symptoms cluster
191 together (Figure 3F) and are found primarily in the IGHV1 andIGHG1 isotypes (Figure 3G).
192

193 **BCR clonotype sequence convergence signatures are shared between different
194 COVID-19 studies in different locations and from different anatomical sites**

195 To further explore whether the convergent clonotypes observed in our study were indeed
196 disease specific, and to determine whether such convergence was common across studies
197 and geographic regions, we compared these 777 convergent clonotypes to public B cell
198 datasets.

199 First, we compared our data to RNAseq data of bronchoalveolar lavage fluid
200 obtained from five of the first infected patients in Wuhan, China ²³. These samples were
201 obtained for the purpose of metagenomic analyses to identify the aetiological agent of the
202 novel coronavirus but were re-analysed to determine whether we could extract any
203 transcripts from BCRs. From the 10,038,758 total reads, we were able to identify 16 unique
204 CDR3 AA sequences (

205 Supplementary Table 2). Of these, one had an exact AA match to a sequence in our
206 data and shared the same V gene segment (IGHV3-15), and J gene segment (IGHJ4) usage
207 (Figure 4A). The sequence had a CDRH3 AA length of 12, so such a match is unlikely to occur
208 due to chance alone. The clonotype that the sequence belonged to contained 699 total
209 sequences and was convergent between 8 of our 19 COVID-19 patients, but not present in
210 the healthy controls. The clonotype was highly diverse, and the sequences had evidence of
211 low mutation from germline, with a mean mutation count over all sequences of 4.8
212 (Supplementary Figure 5).

213 Next, we compared our 777 convergent clonotypes to CoV-AbDab – the Coronavirus
214 Antibody Database [accessed 10th May 2020] ¹⁶. At the time of access, this database
215 contained 80 non-redundant CDRH3 sequences from published and patented antibodies
216 proven to bind SARS-CoV-1 and/or SARS-CoV-2. We found 6 of our clonotypes to have high
217 CDRH3 homology to the antibodies in CoV-AbDab (Figure 4B and Supplementary Figure 6).
218 The most striking similarity was to S304, a previously described SARS-CoV-1 and SARS-CoV-2
219 receptor-binding domain antibody able to contribute to viral neutralisation ²⁴. One of the
220 777 convergent clonotypes contained sequences with an exact CDRH3 AA sequence match
221 and utilised the same IGHV andIGHJ germline gene segments to S304. This clonotype was
222 convergent across 6 patients and had a mean mutation count of 1.1.

223 Finally, we compared our data to a publicly available BCR deep sequencing dataset
224 from six COVID-19 patients from Stanford, USA. 405 of our 777 convergent clonotypes
225 matched (using the same definition we used for clonotyping within our dataset) sequences
226 in this dataset (Figure 4C), showing the high level of convergence between studies. The
227 average number of clonotype matches to the Stanford COVID-19 patient repertoires was 95,
228 but this varied considerably between patients and timepoints. Two of the six patients were
229 seronegative at the day of sampling (7451 and 7453), and these two patients had the fewest
230 clonotype matches (16 and 14 respectively). Patient 7453 had an additional sample taken
231 two days later (following seroconversion), and at this point had a large increase in the

232 number of clonotype matches to 204. There was one of the 777 convergent clonotypes that
233 was found across all six of the Stanford patients, and 17 clonotypes that were found in at
234 least four of five samples from seroconverted patients, but not found in the seronegative
235 patients (Supplementary Table 3).

236

237 Discussion

238

239 We have used deep sequencing of the BCR heavy chain repertoire to evaluate the B cell
240 responses of 19 individuals with COVID-19. In agreement with previous studies, there was a
241 skewing of the repertoire in the response to SARS-CoV-2 infection, with an increased use of
242 certain V genes, an increase in the proportion of antibodies with longer CDRH3s, and an
243 altered isotype subclass distribution ¹⁴. The significantly increased usage of IGHA1 observed
244 in the COVID-19 patients is in line with mucosal responses, where the longer hinge in IGHA1
245 compared to IGHA2 may offer advantages in antigen recognition by allowing higher avidity
246 bivalent interactions with distantly spaced antigens ²⁵.

247 As anticipated, given the novel nature of the virus, that SARS-CoV-2 infection largely
248 stimulated a characteristically naïve response, rather than a reactivation of pre-existing
249 memory B cells – (1) there was an increased prevalence of unmutated antigen-experienced
250 class-switched BCR sequences, (2) an increase in the diversity of class-switched IGHA and
251 IGHG BCRs, and (3) an increase in the usage of isotype subclasses that are associated with
252 recent viral immunity. These observations are consistent with an increase in the frequency
253 of recently activated B cells in response to SARS-CoV-2. In addition to the naïve response,
254 there was also evidence of a proportion of the response arising from memory recall. In the
255 COVID-19 patients, the largest clonal expansions were highly mutated, equivalent to the
256 level observed in healthy control cohort. Such a secondary response to SARS-CoV-2 has
257 been previously observed ²⁶, and may be due to recall of B cells activated in response to
258 previously circulating human coronaviruses, as recently highlighted^{27,28}.

259 We observed a potential relationship between repertoire characteristics and disease
260 state, with improving patients showing a tendency towards a higher proportion of
261 unmutated sequences. The increased prevalence of autoreactive IGHV4-34 sequences in
262 improving COVID-19 patients compared to stable or deteriorating COVID-19 patients
263 potentially suggests a role for natural or autoreactive antibodies in resolving infection and
264 lower risk of pathology. There is a clear need to expand on these findings by using larger
265 sample cohorts and gathering more clinical data to aid understanding of the differences
266 between individuals that respond with mild versus severe disease and have different
267 recovery patterns. Building upon these observations could help to inform the future
268 development of diagnostic assays to monitor and predict the progression of disease in
269 infected patients.

270 A large number (777) of highly convergent clonotypes unique to COVID-19 were
271 identified. Our approach of subtracting the convergent clonotypes also observed in healthy
272 controls ¹⁵, allowed us to identify convergence specific to the disease cohort. The unbiased
273 nature of the BCR repertoire analysis approach means that, whilst these convergent
274 clonotypes are likely to include many antibodies to the spike protein and other parts of the
275 virus they may also include other protective antibodies, including those to host proteins.
276 Characterisation of the heavy chains we have identified, coupled with matched light chains
277 to generate functional antibodies will permit analysis of the binding sites and neutralising

278 potential of these antibodies. The report that plasma derived from recently recovered
279 donors with high neutralising antibody titres can improve the outcome of patients with
280 severe disease ²⁹, supports the hypotheses that intervention with a therapeutic antibody
281 has the potential to be an effective treatment. A manufactured monoclonal antibody or
282 combination of antibodies would also provide a simpler, scalable and safer approach than
283 plasma therapy.

284 Sequence convergence between our 777 convergent clonotypes with heavy chains
285 from published and patented SARS-CoV-1 and SARS-CoV-2 antibodies¹⁶ supports several
286 observations. Firstly, it demonstrates that our approach of finding a convergent sequence
287 signature is a useful method for enriching disease-specific antibodies, as we find matches to
288 known SARS-CoV spike-binding antibodies. Secondly, it shows that the clonotypes observed
289 in response to SARS-CoV-2 overlap with those to SARS-CoV-1, presumably explained by the
290 relatively high homology of the two related viruses ³. Indeed, here we show that there is an
291 overrepresentation of clonotypes that correlate with patient clinical symptoms than is
292 expected by chance, and these BCR sequences are associated with the dominant IgA1 and
293 IgG1 responses. Finally, it shows that the convergence extends beyond our UK COVID-19
294 disease cohort.

295 Further evidence for convergence extending beyond our disease cohort came from
296 the comparisons of our 777 convergent clonotypes to deep sequencing datasets from China
297 ²³ and the USA ¹⁴. The dataset from the USA is also from BCR sequencing of the peripheral
298 blood of COVID-19 patients, and here we found matches to 405 of our 777 clonotypes. The
299 dataset from China was from total RNA sequencing of the bronchoalveolar lavage fluid of
300 SARS-CoV-2 infected patients. Only 16 unique CDRH3 sequences could be identified in this
301 whole dataset, but one of them matched a convergent clonotype in the current study,
302 showing that convergence can be seen both between different locations, and different
303 sample types. We believe that the identification of such high BCR sequence convergence
304 between geographically distinct and independent datasets could be highly significant and
305 validates the disease association of the clonotypes, as well as the overall approach.

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

311 Methods

313 Clinical information gathering

314 Peripheral blood was obtained from patients admitted with acute COVID-19 pneumonia to
315 medical wards at Barts Health NHS Trust, London, UK, after informed consent by the direct
316 care team (NHS HRA RES Ethics 19/SC/0361). Venous blood was collected in EDTA
317 Vacutainers (BD). Patient demographics and clinical information relevant to their admission
318 were collected by members of the direct care team, including duration of symptoms prior to
319 blood sample collection. Current severity was mapped to the WHO Ordinal Scale of Severity.
320 Whether patients at time of sample collection were clinically Improving, Stable or
321 Deteriorating was subjectively determined by the direct clinical team prior to any sample
322 analysis. This determination was primarily made on the basis of whether requirement for

323 supplemental oxygen was increasing, stable, or decreasing comparing current day to
324 previous three days.

325

326 Sample collection and initial processing

327 Blood samples were centrifuged at 150 xg for 15 minutes at room temperature to separate
328 plasma. The cell pellet was resuspended with phosphate-buffered saline (PBS without
329 calcium and magnesium, Sigma) to 20 ml, layered onto 15 ml Ficoll-Paque Plus (GE
330 Healthcare) and then centrifuged at 400 xg for 30 minutes at room temperature without
331 brake. Mononuclear cells (PBMCs) were extracted from the buffy coat and washed twice
332 with PBS at 300 xg for 8 min. PBMCs were counted with Trypan blue (Sigma) and viability of
333 >96% was observed. 5x10⁶ PBMCs were resuspended in RLT (Qiagen) and incubated at room
334 temperature for 10 min prior to storage at -80°C. Consecutive donor samples with sufficient
335 RLT samples progressed to RNA preparation and BCR preparation and are included in this
336 manuscript.

337 Metastatic breast cancer biopsy samples were collected and RNA extracted as part
338 of a previously reported cohort ²².

339

340 RNA prep & BCR sequencing

341 Total RNA from 5x10⁶ PBMCs was isolated using RNeasy kits (Qiagen). First-strand cDNA was
342 generated from total RNA using SuperScript RT IV (Invitrogen) and IgA and IgG isotype
343 specific primers ³⁰ including UMIs at 50 °C for 45 min (inactivation at 80 °C for 10 min).

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

350 Dual-indexed sequencing adapters (KAPA) were ligated onto 500ng amplicons per
351 patient using the HyperPrep library construction kit (KAPA) and the adapter-ligated libraries
352 were finally PCR-amplified for 3 cycles (98 °C for 15 sec, 60 °C for 30 sec, 72 °C for 30 sec,
353 final extension at 72 °C for 1min). Pools of 10 and 9 libraries were sequenced on an Illumina
354 MiSeq using 2x300 bp chemistry.

355

356 Sequence processing

357 The Immcantation framework (docker container v3.0.0) was used for sequence processing
358 ^{31,32}. Briefly, paired-end reads were joined based on a minimum overlap of 20 nt, and a max
359 error of 0.2, and reads with a mean phred score below 20 were removed. Primer regions,
360 including UMIs and sample barcodes, were then identified within each read, and trimmed.
361 Together, the sample barcode, UMI, and constant region primer were used to assign
362 molecular groupings for each read. Within each grouping, usearch ³³, was used to subdivide
363 the grouping, with a cutoff of 80% nucleotide identity, to account for randomly overlapping
364 UMIs. Each of the resulting groupings is assumed to represent reads arising from a single
365 RNA. Reads within each grouping were then aligned, and a consensus sequence determined.

366 For each processed sequence, IgBlast ³⁴ was used to determine V, D and J gene
367 segments, and locations of the CDRs and FWRs. Isotype was determined based on
368 comparison to germline constant region sequences. Sequences annotated as unproductive

369 by IgBlast were removed. The number of mutations within each sequence was determined
370 using the shazam R package ³².

371 Sequences were clustered to identify those arising from clonally related B cells; a
372 process termed clonotyping. Sequences from all samples were clustered together to also
373 identify convergent clusters between samples. Clustering was performed using a previously
374 described algorithm ³⁵. Clustering required identical V and J gene segment usage, identical
375 CDRH3 length, and allowed 1 AA mismatch for every 10 AAs within the CDRH3. Cluster
376 centers were defined as the most common sequence within the cluster. Lineages were
377 reconstructed from clusters using the alakazam R package ³⁶. The similarity tree of the
378 convergent clonotype CDR3 sequences was generated through a kmer similarity matrix
379 between sequences in R.
380

381 [Public healthy control data processing](#)

382 The healthy control BCR sequence dataset used here has been described previously ¹⁵. Only
383 samples from participants aged 10 years or older, and from peripheral blood were used,
384 resulting in a mean age of 28 (range: 11-51). Furthermore, only class-switched sequences
385 were considered.
386

387 [Public SARS-CoV-2 bronchoalveolar lavage RNAseq data processing](#)

388 The bronchoalveolar lavage data comes from a previously published study of SARS-CoV-2
389 infection ²³, with data available under the PRJNA605983 BioProject on NCBI. MIXCR v3.0.3
390 was used, with default settings, to extract reads mapping to antibody genes from the total
391 RNASeq data ³⁷.
392

393 [Public CoV-AbDab data processing](#)

394 All public CDRH3 AA sequences associated with published or patented SARS-CoV-1 or SARS-
395 CoV-2 binding antibodies were mined from CoV-AbDab¹⁶, downloaded on 10th May 2020. A
396 total of 80 non-redundant CDRH3s were identified (100% identity threshold). These
397 sequences were then clustered alongside the representative CDRH3 sequence from each of
398 our 777 convergent clones using CD-HIT ³⁸, at an 80% sequence identity threshold (allowing
399 at most a CDRH3 length mismatch of 1 AA). Cluster centres containing at least one CoV-
400 AbDab CDRH3 and one convergent clone CDRH3 were further investigated.
401

402 [Public COVID-19 BCR sequence data processing](#)

403 The fourteen MiSeq "read 1" FASTQ datasets from the six SARS-CoV-2 patients analysed in
404 Nielsen et al. ¹⁴ were downloaded from the Sequence Read Archive ³⁹. IgBlast ³⁴ was used to
405 identify heavy chain V, D, and J gene rearrangements and antibody regions. Unproductive
406 sequences, sequences with out-of-frame V and J genes, and sequences missing the CDRH3
407 region were removed from the downstream analysis. Sequences with 100% amino acid and
408 isotype matches were collapsed. To circumvent the disparity in collapsed dataset sizes
409 between pairs of replicates, we selected the replicate with the highest number of sequences
410 for downstream analysis.
411

412 [Convergent Clonotyping Matching to Public Repertoires](#)

413 The public SARS-CoV-2-positive¹⁴ and healthy control BCR repertoires⁴⁰ were scanned for
414 clonotype matches to our 777 convergent clonotype cluster centres. A BCR repertoire

415 sequence was determined as a match if it had identical V and J genes, the same length
416 CDRH3, and was within 1 AA mismatch per 10 CDRH3 AAs to a convergent clonotype
417 representative sequence.

418

419 Statistical analysis and graphing

420 Statistical analysis and plotting were performed using R ⁴¹. Plotting was performed using
421 ggplot2 ⁴². Sequence logos were created using ggseqlogo ⁴³. Specific statistical tests used are
422 detailed in the figure legends. Correlations of IGHV4-34 autoreactive motifs and convergent
423 clonotypes was performed by manova in R.

424

425 References

426

- 427 1. Lu, H., Stratton, C. W. & Tang, Y. W. Outbreak of pneumonia of unknown etiology in
428 Wuhan, China: The mystery and the miracle. *Journal of Medical Virology* vol. 92 401–
429 402 (2020).
- 430 2. Huang, C. *et al.* Clinical features of patients infected with 2019 novel coronavirus in
431 Wuhan, China. *Lancet* **395**, 497–506 (2020).
- 432 3. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor Recognition by the Novel
433 Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of
434 SARS Coronavirus. *J. Virol.* **94**, (2020).
- 435 4. Chen, N. *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel
436 coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* **395**, 507–513
437 (2020).
- 438 5. Mulangu, S. *et al.* A randomized, controlled trial of Ebola virus disease therapeutics.
439 *N. Engl. J. Med.* **381**, 2293–2303 (2019).
- 440 6. Gogtay, N. J. *et al.* Comparison of a Novel Human Rabies Monoclonal Antibody to
441 Human Rabies Immunoglobulin for Postexposure Prophylaxis: A Phase 2/3,
442 Randomized, Single-Blind, Noninferiority, Controlled Study. *Clin. Infect. Dis.* **66**, 387–
443 395 (2018).
- 444 7. Johnson, S. *et al.* Development of a Humanized Monoclonal Antibody (MEDI-493)
445 with Potent In Vitro and In Vivo Activity against Respiratory Syncytial Virus. *J. Infect.*
446 *Dis.* **176**, 1215–1224 (1997).
- 447 8. Wang, S. F. *et al.* Antibody-dependent SARS coronavirus infection is mediated by
448 antibodies against spike proteins. *Biochem. Biophys. Res. Commun.* **451**, 208–214
449 (2014).
- 450 9. Tetro, J. A. Is COVID-19 receiving ADE from other coronaviruses? *Microbes Infect.* **22**,
451 72–73 (2020).
- 452 10. Sharma, A. It is too soon to attribute ADE to COVID-19. *Microbes and Infection* (2020)
453 doi:10.1016/j.micinf.2020.03.005.
- 454 11. Wang, Q. *et al.* Immunodominant SARS coronavirus epitopes in humans elicited both
455 enhancing and neutralizing effects on infection in non-human primates. *ACS Infect.*
456 *Dis.* **2**, 361–376 (2016).
- 457 12. Brouwer, P. *et al.* Potent neutralizing antibodies from COVID-19 patients define
458 multiple targets of vulnerability. *bioRxiv* (2020) doi:10.1101/2020.05.12.088716.
- 459 13. Andreano, E. *et al.* Identification of neutralizing human monoclonal antibodies from
460 Italian Covid-19 convalescent patients. *bioRxiv* (2020)

461 doi:10.1101/2020.05.05.078154.

462 14. Nielsen, S. *et al.* B cell clonal expansion and convergent antibody responses to SARS-
463 CoV-2. *Res. Sq.* (2020) doi:10.21203/rs.3.rs-27220/v1.

464 15. Ghraichy, M. *et al.* Maturation of naïve and antigen-experienced B-cell receptor
465 repertoires with age. *bioRxiv* (2019) doi:10.1101/609651.

466 16. Raybould, M. I. J., Kovaltsuk, A., Marks, C. & Deane, C. M. CoV-AbDab: the
467 Coronavirus Antibody Database. *bioRxiv* (2020) doi:10.1101/2020.05.15.077313.

468 17. Pascual, V. *et al.* Nucleotide sequence analysis of the V regions of two IgM cold
469 agglutinins: Evidence that the V(H)4-21 gene segment is responsible for the major
470 cross-reactive idiotype. *J. Immunol.* **146**, 4385–4391 (1991).

471 18. Schickel, J. N. *et al.* Self-reactive VH4-34-expressing IgG B cells recognize commensal
472 bacteria. *J. Exp. Med.* **214**, 1991–2003 (2017).

473 19. Tipton, C. M. *et al.* Diversity, cellular origin and autoreactivity of antibody-secreting
474 cell population expansions in acute systemic lupus erythematosus. *Nat. Immunol.* **16**,
475 755–765 (2015).

476 20. Horns, F. *et al.* Lineage tracing of human B cells reveals the in vivo landscape of
477 human antibody class switching. *Elife* **5**, 1–20 (2016).

478 21. Parameswaran, P. *et al.* Convergent Antibody Signatures in Human Dengue. *Cell Host*
479 *Microbe* **13**, 691–700 (2013).

480 22. De Mattos-Arruda, L. *et al.* The Genomic and Immune Landscapes of Lethal
481 Metastatic Breast Cancer. *Cell Rep.* **27**, 2690–2708.e10 (2019).

482 23. Zhou, P. *et al.* A pneumonia outbreak associated with a new coronavirus of probable
483 bat origin. *Nature* **579**, 270–273 (2020).

484 24. Pinto, D. *et al.* Structural and functional analysis of a potent sarbecovirus neutralizing
485 antibody. *bioRxiv* (2020) doi:10.1101/2020.04.07.023903.

486 25. Li, Q. *et al.* IgA MAb blocks SARS-CoV-2 Spike-ACE2 interaction providing mucosal
487 immunity. *bioRxiv* (2020) doi:10.1101/2020.05.15.096719.

488 26. Wec, A. Z. *et al.* Broad sarbecovirus neutralizing antibodies define a key site of
489 vulnerability on the SARS-CoV-2 spike protein. *bioRxiv* (2020)
490 doi:10.1101/2020.05.15.096511.

491 27. Grifoni, A. *et al.* Targets of T cell responses to SARS-CoV-2 coronavirus in humans with
492 COVID-19 disease and unexposed individuals. *Cell* (2020)
493 doi:10.1016/j.cell.2020.05.015.

494 28. Ng, K. *et al.* Pre-existing and de novo humoral immunity to SARS-CoV-2 in humans.
495 *BioRxiv* (2020) doi:10.1101/2020.05.14.095414.

496 29. Duan, K. *et al.* Effectiveness of convalescent plasma therapy in severe COVID-19
497 patients. *Proc. Natl. Acad. Sci.* **117**, 202004168 (2020).

498 30. van Dongen, J. J. M. *et al.* Design and standardization of PCR primers and protocols
499 for detection of clonal immunoglobulin and T-cell receptor gene recombinations in
500 suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-
501 3936. *Leukemia* **17**, 2257–317 (2003).

502 31. Vander Heiden, J. A. *et al.* pRESTO: a toolkit for processing high-throughput
503 sequencing raw reads of lymphocyte receptor repertoires. *Bioinformatics* **30**, 1930–2
504 (2014).

505 32. Gupta, N. T. *et al.* Change-O: A toolkit for analyzing large-scale B cell immunoglobulin
506 repertoire sequencing data. *Bioinformatics* **31**, 3356–3358 (2015).

507 33. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST.

508 34. *Bioinformatics* **26**, 2460–2461 (2010).

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

511 35. Galson, J. D. *et al.* BCR repertoire sequencing: different patterns of B cell activation
512 after two Meningococcal vaccines. *Immunol. Cell Biol.* **93**, 885–95 (2015).

513 36. Vander Heiden, J. A. & Gupta, N. alakazam: Immunoglobulin Clonal Lineage and
514 Diversity Analysis. *R Packag. version 0.2.0* (2015).

515 37. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling.
516 *Nat. Methods* **12**, 380–381 (2015).

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

519 39. Leinonen, R., Sugawara, H. & Shumway, M. The sequence read archive. *Nucleic Acids
520 Res.* **39**, (2011).

521 40. Briney, B., Inderbitzin, A., Joyce, C. & Burton, D. R. Commonality despite exceptional
522 diversity in the baseline human antibody repertoire. *Nature* **566**, 393–397 (2019).

523 41. Team, R. D. C. R: A language and environment for statistical computing. *R Found. Stat.
524 Comput. Vienna, Austria* (2008).

525 42. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer; 1st ed. 2009.
526 Corr. 3rd printing 2010 edition, 2009).

527 43. Wagih, O. Ggseqlogo: A versatile R package for drawing sequence logos.
528 *Bioinformatics* **33**, 3645–3647 (2017).

530 Acknowledgments

531
532 The authors would like to first and foremost acknowledge and thank the patients who
533 consented to providing their samples to this study. We are grateful to Ursula Arndt, and the
534 RTD team at Illumina who performed the sequencing reactions to generate data for
535 analysis. We also thank Felicia Anna Tucci (University of Oxford) for her help with the BCR
536 sequencing methods.

537 Funding Information

539 MIJR is supported by an Engineering and Physical Sciences Research Council (EPSRC) and
540 Medical Research Council (MRC) grant [EP/L016044/1]. AK is supported by a Biotechnology
541 and Biological Sciences Research Council (BBSRC) grant [BB/M011224/1].

542 Author contributions

543
544 All authors discussed methodology, results and contributed to the final manuscript. JO, OC,
545 AL, GT, DP, PP conceived and designed the study. JDG, SS, MIJR, RJMB-R and AK conducted
546 the data analysis with input from JO, RM, JD, GJK, DKF, LJ, RJMB-R, CMD, OC, W-YJL, GT, PP
547 and DP. GT, PP, DP recruited COVID-19 participants and executed the clinical protocols. SS,
548 JD and W-YJL processed the COVID-19 clinical samples. CC and JS recruited the breast cancer
549 participants. SS and JD processed the breast cancer samples. JB oversaw sequencing of the
550 libraries. JDG, JO, GJK, SS and RM wrote the manuscript with input from all co-authors. All
551 authors read and approved the final manuscript.

552 [Competing interests](#)

553

554 JO, AL, OC, SS, JDG, JD, RM and DKF are employees of Alchemab Therapeutics Limited.
555 RJMB-R is a founder of and consultant to Alchemab Therapeutics Limited. GJK is a
556 consultant to Alchemab Therapeutics Limited. CC is a member of the AstraZeneca External
557 Science Panel and has research grants from Roche, Genentech, AstraZeneca, and Servier
558 that are administered by the University of Cambridge.

559 [Data availability](#)

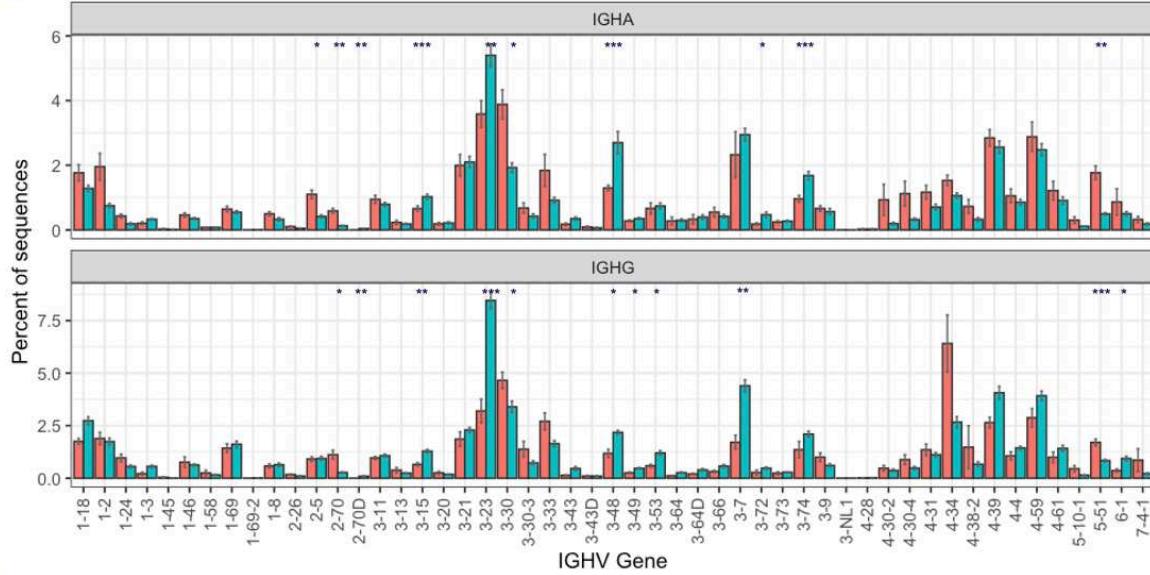
560

561 COVID-19 BCR sequence data will be made available upon publication.

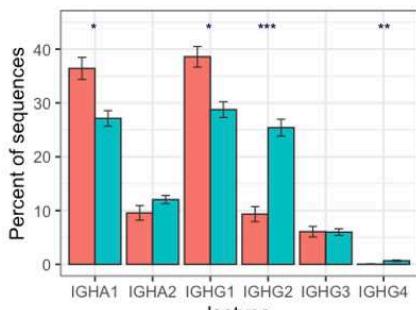
562 Figures

563

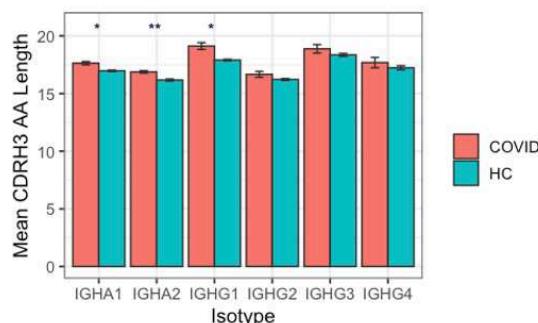
A



B



C



564

565

566

567

568

569

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

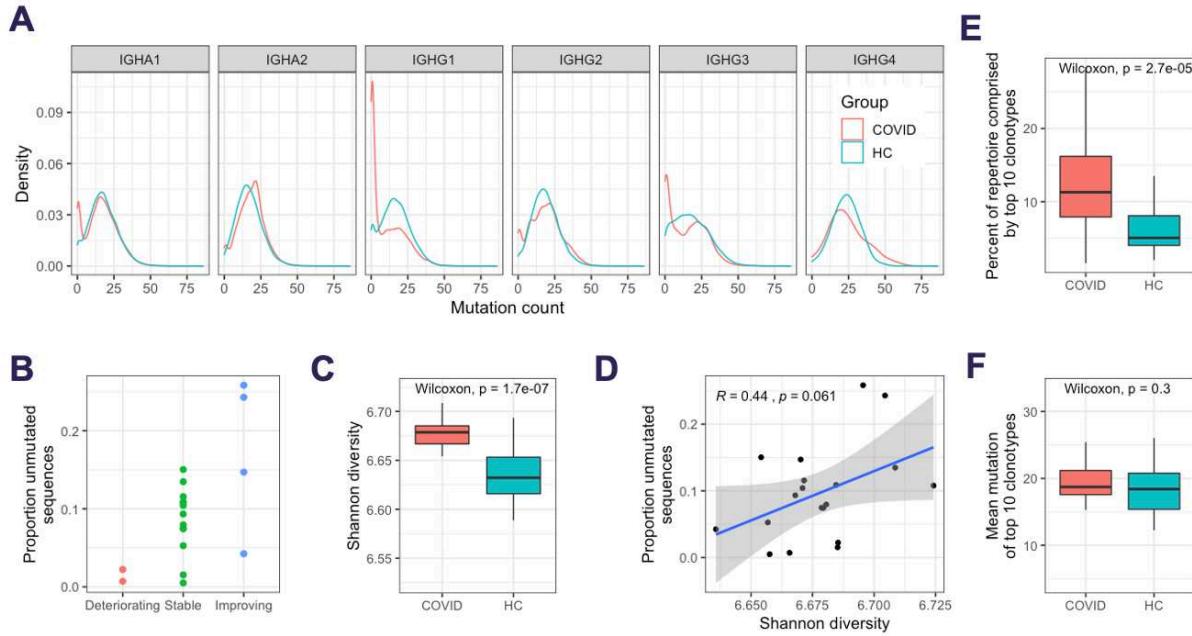
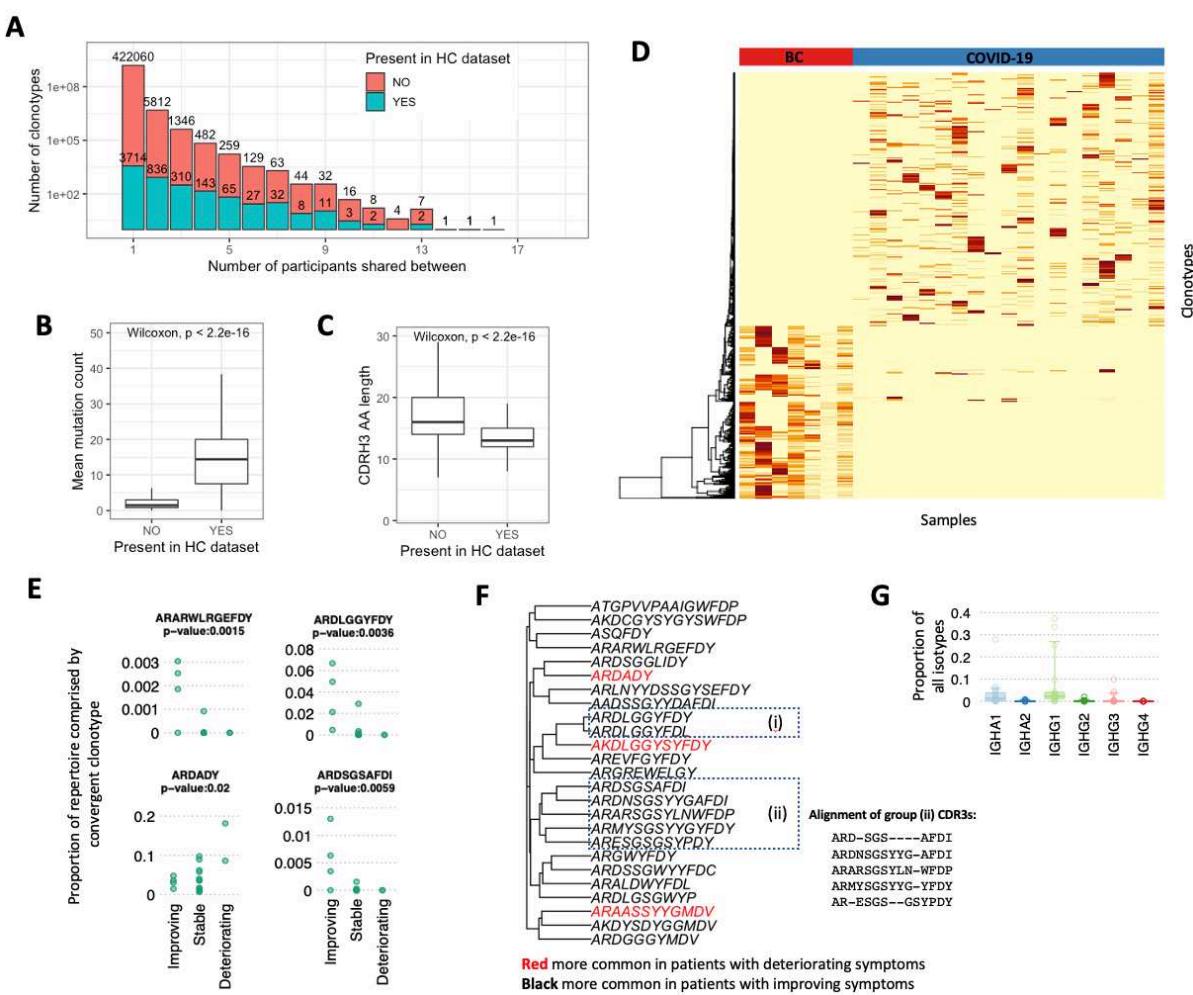


Figure 2. Response characteristics of SARS-CoV-2 infection. A) Distribution of sequences with different numbers of mutations from germline. B) Relationship between the proportion of the repertoire comprised by unmutated sequences, and the disease state C) Individual sequences were clustered together into related groups to identify clonal expansions (clonotypes). Diversity of all clonotypes in the repertoire calculated using the Shannon diversity index. To normalise for different sequence numbers for each sample, a random subsample of 1,000 sequences was taken. D) Correlation between the Shannon diversity index, and the proportion of unmutated sequences. E) The percent of all sequences that fall into the largest 10 clonotypes. F) Mean number of mutations of all sequences in the largest 10 clonotypes.



578
579
580 Figure 3. Convergent BCR sequence signature within individuals infected with SARS-CoV-2. A) Data from all patients and
581 healthy controls were clustered together to identify convergent clonotypes. Shown is the number of clonotypes shared by
582 different numbers of participants, grouped by whether the clonotypes are also present in the healthy control dataset. Of the
583 convergent clonotypes, B) the mean mutation count, and C) the CDRH3 AA sequence length was compared between those
584 that were convergent only within the SARS-CoV-2 patients, and those that were also convergent with the healthy control
585 dataset. D) Heatmap of the 777 convergent COVID-19-associated clonotypes (observed between 4 or more COVID-19
586 participants) with the 469 convergent clonotypes from seven metastatic breast cancer (BC) patient biopsy samples,
587 demonstrating that the convergent signatures are unique to each disease cohort. E) Percentage frequencies of four example
588 convergent clonotypes grouped by clinical status. F) Similarity tree of convergent clonotype cluster centers that are
589 significantly associated with clinical status. Groups (i) and (ii) indicate groups of similar convergent clonotypes. An
590 alignment of group (ii) provided adjacent. G) Proportions of IGH A and IGH G of the convergent clonotypes that are
591 associated with patients with improving symptoms.

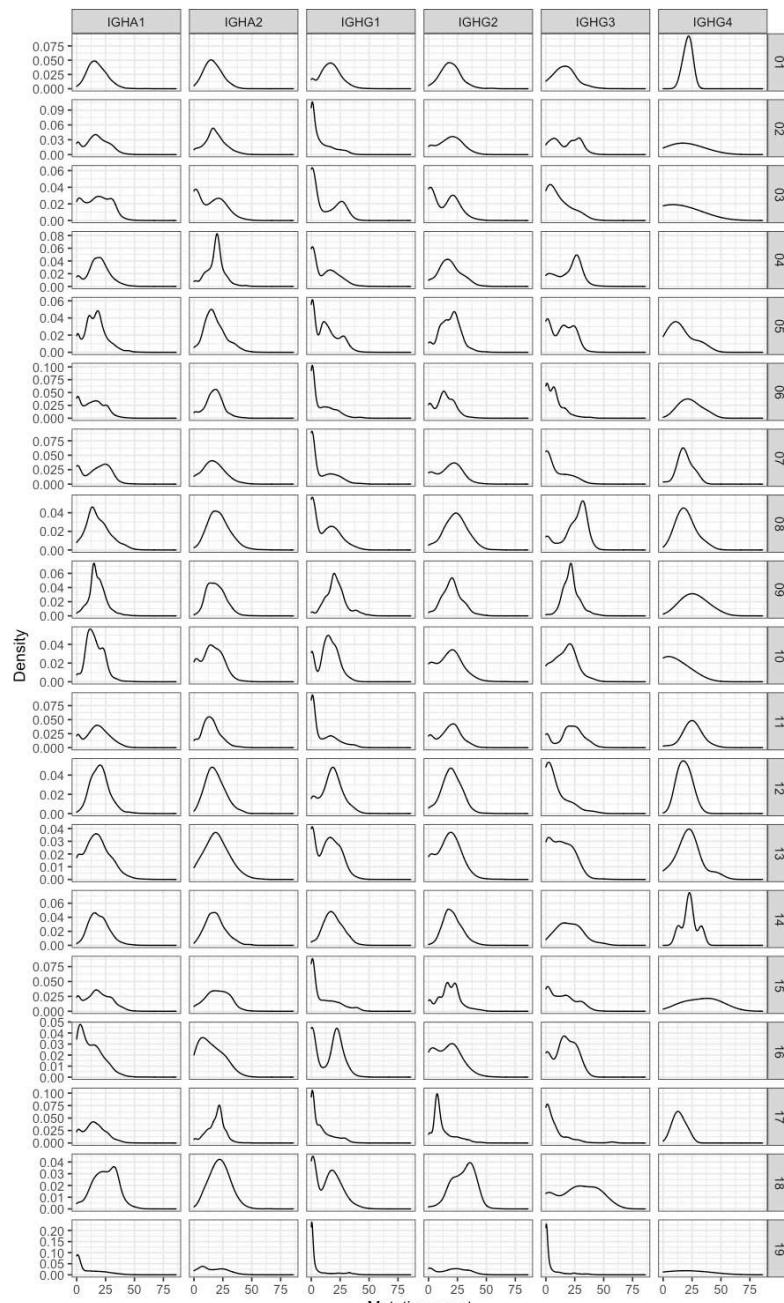


592
593 *Figure 4. Matches of the 777 convergent clonotypes identified in the present study to other SARS-CoV-2 studies. A) CDRH3*
594 *sequence, and IGHV/IGHJ gene segments of a sequence identified in the bronchoalveolar lavage fluid of a SARS-CoV-2*
595 *patient from a Chinese cohort (shown across the top in black text), and a CDRH3 AA sequence logo unpacking the sequence*
596 *diversity present in the convergent clonotype found in the COVID-19 patients in this study that had an exact AA match. B)*
597 *CDRH3 sequence, and IGHV/IGHJ gene segment of an antibody in the CoV-AbDab (S304) that has SARS-CoV-1 and SARS-*
598 *CoV-2 neutralising activity, alongside a CDRH3 AA sequence logo unpacking the sequence diversity in the convergent*
599 *clonotype found in the COVID-19 patients in this study that had an exact AA match. C) Comparison of convergent*
600 *clonotypes to the BCR data from Nielsen et al¹⁴. Plotted along the x-axis are the 405 convergent clonotypes represented in*
601 *at least one Nielsen et al. dataset. Each row represents a separate BCR repertoire from Nielsen et al.; pink shading indicates*
602 *that the convergent clonotype has a match in the Nielsen dataset.*

603 Supplementary information

604

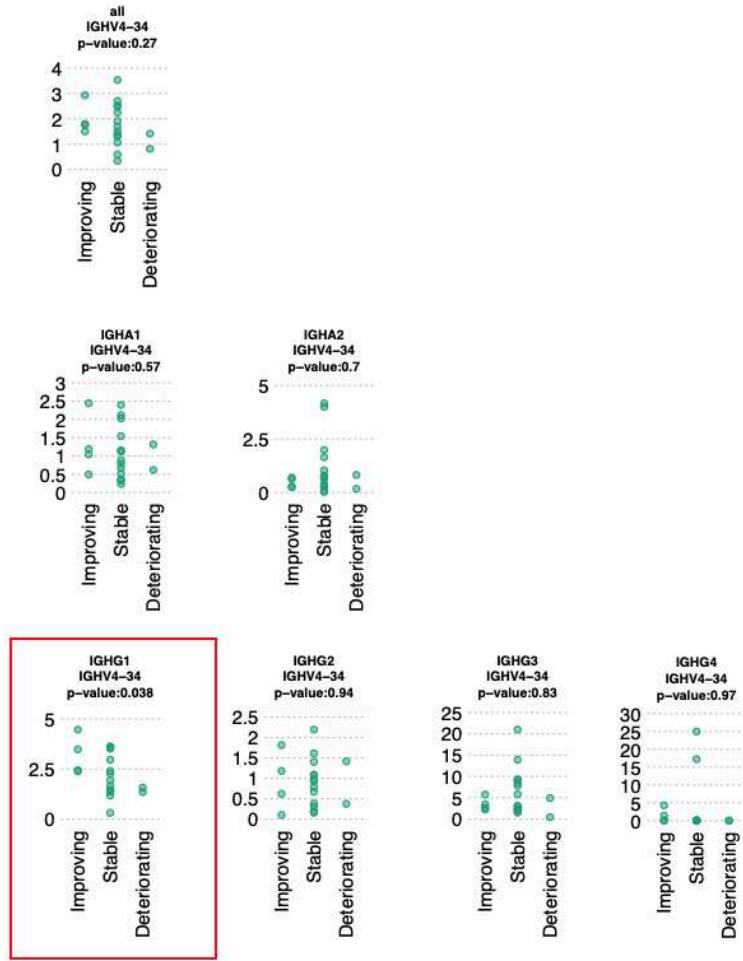
605



606

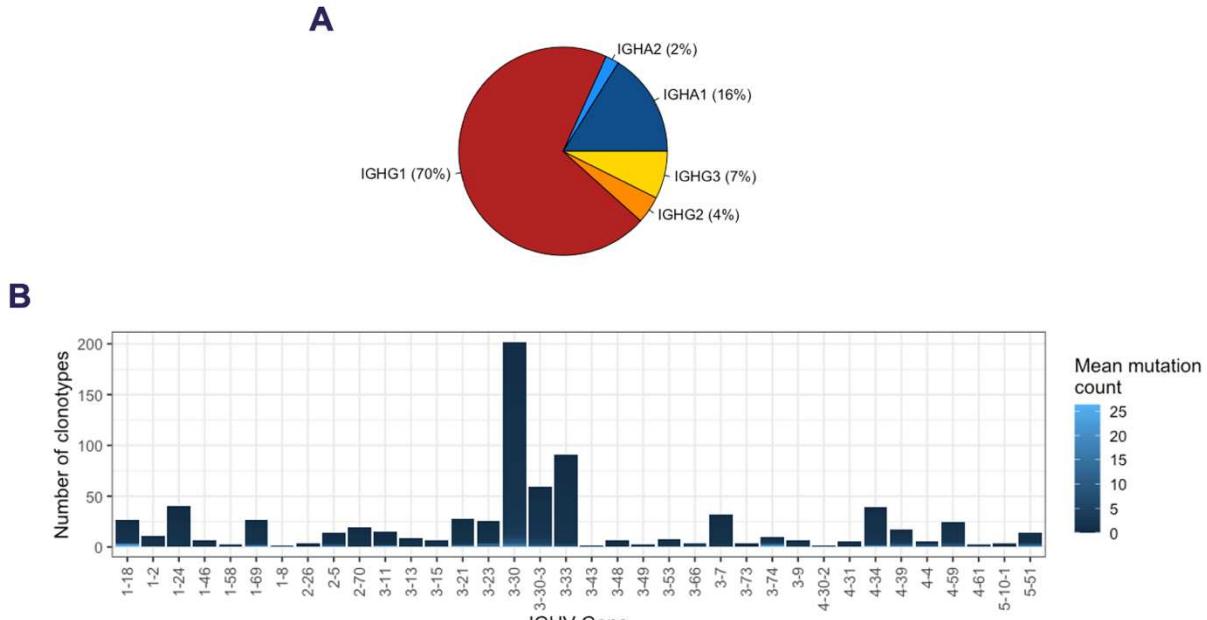
607

608 *Supplementary Figure 1. Distribution of sequences with different numbers of mutations from germline. Each row is a different COVID-19 patient and (right)*



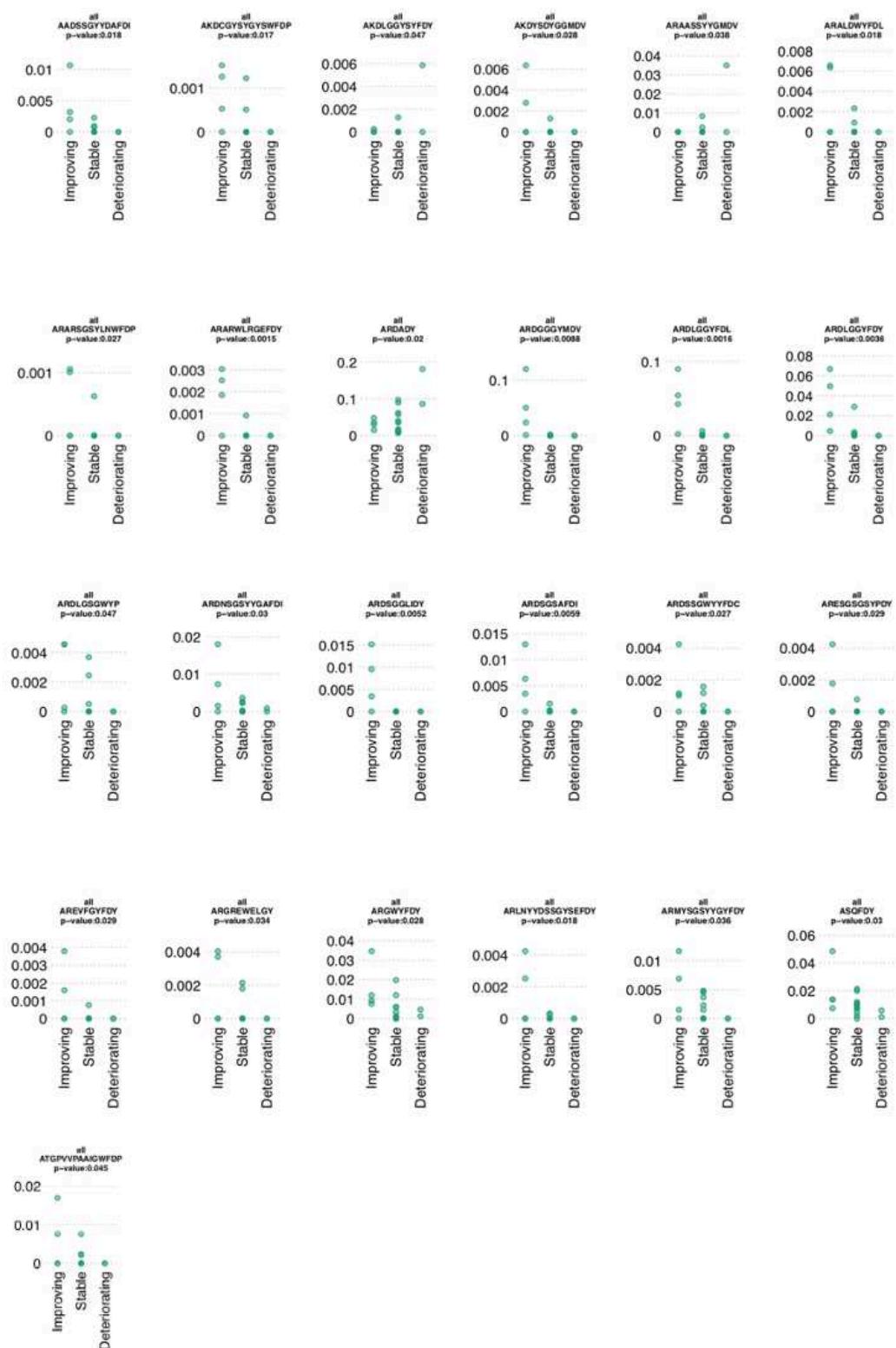
609
610
611
612

Supplementary Figure 2. The proportion of IGHG1 sequences containing the autoreactive "NHS and "AVY" motifs between COVID patients with improving, stable or worsening symptoms. IGHG1 (red box) was the only significant correlation. P-values are determined by ANOVA.



613

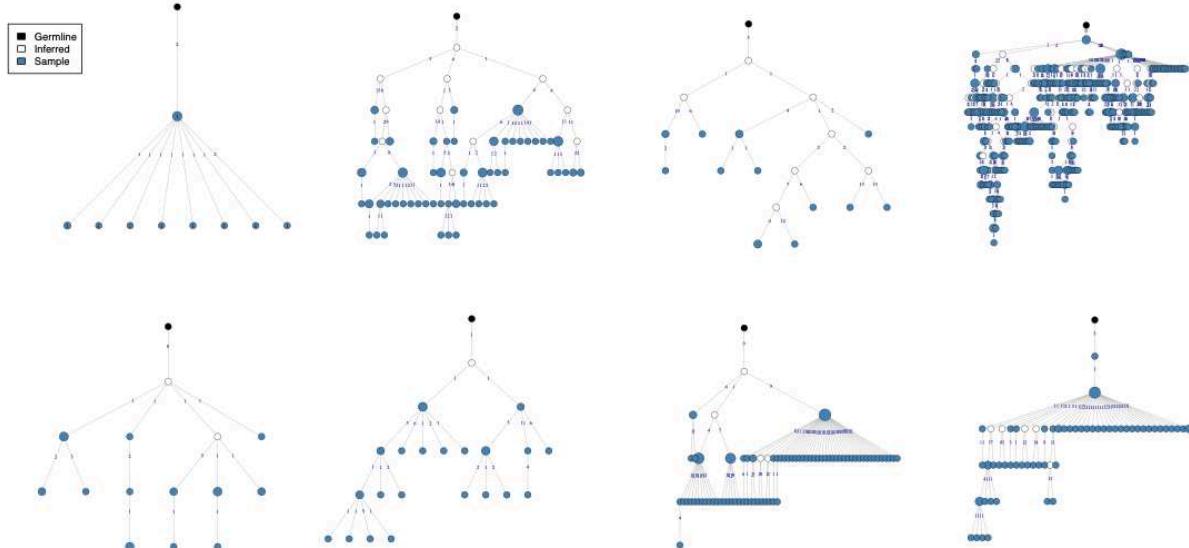
614 *Supplementary Figure 3. Properties of the 777 convergent clonotypes A) Isotype subclass usage of the sequences with the*
615 *777 convergent clonotypes. B) IGHV gene segment usage of the 777 convergent clonotypes.*



616

617 *Supplementary Figure 4. Percentage frequencies of the convergent clonotypes grouped by clinical status that significantly*
 618 *associated with clinical status.*

619



620

621 *Supplementary Figure 5. Lineage trees of the convergent clonotype that matched to the bronchoalveolar lavage fluid data.*
622 *Each lineage tree represents the members of the clonotype from each of the eight patients it was present in. Each node*
623 *represents a unique sequence within the clonotype lineage tree, with the size indicative of the number of duplicate*
624 *sequences present. Numbers on the edges of adjoining nodes show the number of mutations between the sequences.*

625



chemistry: ■ Acidic ■ Basic ■ Hydrophobic ■ Polar

626
627
628
629
630

Supplementary Figure 6. Logo plots unpacking the sequence diversity present for the convergent clonotypes that clustered with CoV-AbDab SARS-CoV-1 or SARS-CoV-2 binding antibodies. The CoV-AbDab reference CDRH3 and IGHV/IGHJ gene segment is displayed above each Logo plot. Gene transcript matches are shown in green, while mismatches are shown in red. The full sequence for 31B9 is not yet publicly available, so its genetic origins are not determined (ND).

Participant ID	Unique BCR Sequences	Clonotypes
1	47878	15456
2	257570	53168
3	33099	9616
4	37138	10754
5	198732	20036
6	233283	26181
7	51305	22276
8	39303	9391
9	221870	18278
10	54645	9255
11	202896	41132
12	31035	6791
13	40995	14782
14	171231	21373
15	280310	36446
16	29620	8736
17	253037	34805
18	60316	15068
19	329055	77557

631

632 *Supplementary Table 1. Summary of number of unique sequences, and number of clonotypes obtained for each COVID-19*
633 *patient.*

bestVHit	bestJHit	aaSeqCDR3
IGHV2-26	IGHJ3	CARDSGRHLGPFDIW
IGHV1-2	IGHJ3	CATPYYYDGGGLDAFDIW
IGHV3-74	IGHJ5	CARDLSRTNWFDPW
IGHV3-15	IGHJ4	CTTDLHDYGDSDYW
IGHV3-15	IGHJ4	CTTDFFGGMITFGGVLRRI
IGHV3-21	IGHJ4	CARAQSRRGGYDSFFDFW
IGHV3-21	IGHJ4	CGRGGPGTGIDYW
IGHV4-59	IGHJ5	CARGGQYNNWFAPW
IGHV3-74	IGHJ5	CVRDLSRTNWFDPW
IGHV3-15	IGHJ4	YTRDLHDYGDSDYW
IGHV3-23	IGHJ3	CAKIPSFLSDYDVHPNDAIDIW
IGHV5-10-1	IGHJ4	CARHPQGAQFSNLGTYYFDYW
IGHV4-59	IGHJ4	CARDGEYGGGLAMW
IGHV5-51	IGHJ6	CARPGTYYDILTGYSNHGMDVW
IGHV4-39	IGHJ5	CARHASFRGTTNYNWFDPW
IGHV3-53	IGHJ5	CARDTSTEDVAWWFDPW

634

635 *Supplementary Table 2. CDRH3 AA sequences identified from bronchoalveolar RNAseq data. Highlighted in green is the one*
636 *identified in our SARS-CoV-2 patient dataset.*

Cluster Center CDRH3	IGHV gene	IGHJ gene	Total Matches	Seropositive Matches	Seronegative Matches
ARGFDP	IGHV4-34	IGHJ5	5	5	0
ARVFDY	IGHV3-30	IGHJ4	5	4	1
ARELSYYGMDV	IGHV3-30	IGHJ6	5	4	1
AREDYGDYGF DY	IGHV3-30	IGHJ4	5	4	1
ARGFDH	IGHV4-61	IGHJ4	4	4	0
ARDSGGLIDY	IGHV3-30	IGHJ4	4	4	0
AREVMVYLDY	IGHV3-33	IGHJ4	4	4	0
ARDSGSAFDI	IGHV3-30-3	IGHJ3	4	4	0
ANDLYYGMDV	IGHV3-30	IGHJ6	4	4	0
AREGPDAFDI	IGHV1-18	IGHJ3	4	4	0
AKEGIVAFDY	IGHV3-30	IGHJ4	4	4	0
ARQEHYYYGMDV	IGHV5-51	IGHJ6	4	4	0
ARPYSGSYRGYFDY	IGHV3-30	IGHJ4	4	4	0
ARSRGGSYYGGFDY	IGHV3-30	IGHJ4	4	4	0
ARDLDYYDSSGFDY	IGHV3-7	IGHJ4	4	4	0
AKARGGSYLDADI	IGHV3-30	IGHJ3	4	4	0
ARVDYYDSSGYRDY	IGHV1-69	IGHJ4	4	4	0
TTGTWYYDSSGYSNDAFDI	IGHV3-15	IGHJ3	4	4	0
ARGIDY	IGHV3-23	IGHJ4	4	3	1
ARDLGDYGM DV	IGHV3-53	IGHJ6	4	3	1

637

638 *Supplementary Table 3. A subset of the 777 convergent clonotypes that matched to at least 4 of the samples in the Nielsen*
639 *et al¹⁴ data.*