Distinct antibody repertoires against endemic human coronaviruses in children and adults

Taushif Khan[1], Mahbuba Rahman[1], Fatima Al Ali[1], Susie S. Y. Huang[1], Manar Ata[1], Qian Zhang[2], Paul Bastard[3,4], Zhiyong Liu[2], Emmanuelle Jouanguy[2-4], Vivien Béziat[2-4], Aurélie Cobat[2-4], Gheyath K. Nasrallah[5], Hadi Yassine[5], Maria Smatti[5], Amira Sayeed[6], Isabelle Vandernoot[7], Jean-Christophe Goffard[8], Guillaume Smits[7], Isabelle Migeotte[9], Filomeen Haerynck[10], Isabelle Meyts[11,12], Laurent Abel[2-4], Jean-Laurent Casanova[2-4, 13,14], Mohammad R. Hasan[6,15], Nico Marr[1,16*]

Version 2, 24 September 2020

[1] Research Branch, Sidra Medicine, Doha, Qatar

[2] St Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY, USA

[3] Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, Necker Hospital for Sick Children, Paris, France, EU.

[4] University of Paris, Imagine Institute, Paris, France, EU

[5] College of Health Sciences and Biomedical Research Center, Qatar University, Doha, Qatar

[6] Department of Pathology, Sidra Medicine, Doha, Qatar

[7] Center of Human Genetics, Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium

[8] Department of Internal Medicine, Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium

[9] Fonds de la Recherche Scientifique (FNRS) and Center of Human Genetics, Hôpital Erasme, Université Libre de Bruxelles, Brussels, Belgium

[10] Department of Respiratory Diseases, Department of Pediatrics and Genetics, Center for Primary Immunodeficiencies Ghent, Jeffrey Modell Foundation diagnostic and research center, Gent, Belgium, EU

30 [11]Laboratory for Inborn Errors of Immunity, Department of Microbiology, Immunology and
31 Transplantation, Department of Pediatrics, University Hospitals Leuven, KU Leuven,
32 Belgium
33 [12]Department of Pediatrics, University Hospitals Leuven, KU Leuven, Belgium
34 [13]Pediatric Hematology and Immunology Unit, Necker Hospital for Sick Children, AP-HP,
35 Paris, France, EU
36 [14]Howard Hughes Medical Institute, New York, NY, USA
37 [15]Department of Pathology and Laboratory Medicine, Weill Cornell Medical College in Qatar,
38 Doha, Qatar
39 [16]College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar
40
41
42 *Corresponding author: Nico Marr, Sidra Medicine, Research Branch, Al Gharrafa Street, Ar-
43 Rayyan, PO BOX 26999, Doha, Qatar, E-mail: nmarr@sidra.org

Abstract

Four endemic human coronaviruses (HCoVs) are commonly associated with acute respiratory infection in humans. B cell responses to these “common cold” viruses remain incompletely understood. Here we report a comprehensive analysis of CoV-specific antibody repertoires in 231 children and 1168 adults using phage-immunoprecipitation sequencing. Seroprevalence of antibodies to endemic HCoVs ranged between ~4 and 27% depending on the species and cohort. We identified at least 136 novel linear B cell epitopes. Antibody repertoires against endemic HCoVs were qualitatively different between children and adults in that anti-HCoV IgG specificities more frequently found among children targeted functionally important and structurally conserved regions of the spike, nucleocapsid and matrix proteins. Moreover, antibody specificities targeting the highly conserved fusion peptide region and S2’ cleavage site of the spike protein were broadly cross-reactive with peptides of epidemic human and non-human coronaviruses. In contrast, an acidic tandem repeat in the N-terminal region of the Nsp3 subdomain of the HCoV-HKU1 polyprotein was the predominant target of antibody responses in adult donors. Our findings shed light on the dominant species-specific and pan-CoV target sites of human antibody responses to coronavirus infection, thereby providing important insights for the development of prophylactic or therapeutic monoclonal antibodies and vaccine design.

62 ## Introduction

63 Four endemic human-tropic coronaviruses (HCoVs) are commonly associated with respiratory
64 illness in humans, namely HCoV-229E, -NL63, -OC43, and -HKU1 [1-4]. Clinical outcomes of
65 acute infection with these HCoVs range from mild upper respiratory tract infections in most
66 patients, to viral bronchiolitis and pneumonia in rare patients, the latter requiring
67 hospitalization [5]. The ratio of more severe versus mild outcomes of acute infection with
68 endemic HCoVs is largely comparable to that of other "common cold" viruses, such as human
69 respiratory syncytial virus (HRSV), human rhinoviruses (HRVs), human adenoviruses (HAdVs),
70 and human parainfluenza viruses (HPIVs), albeit with differences in seasonality and
71 prevalence of the viruses depending on the species [5-7]. In addition to the four endemic
72 HCoV, three human-tropic epidemic coronaviruses (CoVs) have emerged over the last two
73 decades, namely Severe Acute Respiratory Syndrome-associated CoV (SARS-CoV) [8], Middle
74 East Respiratory Syndrome-associated coronavirus (MERS-CoV) [9], and SARS-CoV-2 [10], the
75 etiological agent of Coronavirus Disease 2019 (COVID-19) which has now reached pandemic
76 proportions [11]. Similar to endemic HCoVs, infection of humans with epidemic CoVs is
77 associated with a wide range of outcomes but leads more frequently to severe clinical
78 manifestations such as acute respiratory distress syndrome (ARDS) [12-14]. Phylogenetic
79 analyses suggest that, similar to these epidemic CoVs, all endemic HCoVs are of zoonotic
80 origin and their possible ancestors share similar natural animal reservoirs and intermediate
81 hosts [6]. HCoV-229E may have been transferred from dromedary camels, similar to MERS-
82 CoV, while HCoV-OC43 is thought to have emerged more recently from ancestors in domestic
83 animals such as cattle or swine in the context of a pandemic at the end of the 19th century
84 [6, 15].
85
86 The wide variability in transmissibility and clinical manifestations of infections by endemic
87 and epidemic CoVs among humans remains poorly understood. On the population level, the
88 case-fatality rate is highest for MERS (~34 - 37%) and several risk factors are associated with
89 progression to ARDS in MERS, SARS and COVID-19 cases, including old age, diabetes mellitus,
90 hypertension, cancer, renal and lung disease, and co-infections [12, 16]. Nonetheless, even
91 MERS-CoV infection among humans can run a completely asymptomatic course in some
92 cases, particularly among children [17-19]. There is evidence that children are generally less

susceptible to infection with epidemic CoVs and once infected, they are less likely to experience severe outcomes compared to adults, although this important association and the underlying reasons remain poorly understood [12, 18, 20, 21]. Importantly, it remains unclear to what extent pre-existing immunity from past infections with endemic HCoVs provides some degree of cross-protection and affects clinical outcomes of infection with the epidemic SARS-CoV-2 or MERS-CoV. Our overall understanding of the immunity induced by natural infection with endemic HCoVs remains very limited. Serological studies have shown some degree of cross-reactive antibodies in patients with past CoV infections but many of these studies were limited in sample size and often focused on specific viral antigens only [22-25]. Depending on their binding affinity and specificities, such cross-reactive antibodies could either have no effect on clinical outcomes, may provide protection from severe disease to some degree, or may lead to antibody-dependent enhancement of disease—the latter can be a major obstacle in vaccine development [26]. Interestingly, two recent studies from independent groups have shown that a considerable proportion of individuals without a history of SARS-CoV-2 infection have SARS-CoV-2-reactive T cells, which suggests that cross-reactive T cell subsets originating from past infections by endemic HCoVs may play a role in the clinical course of infection with the phylogenetically related epidemic CoVs [27, 28]. A systematic assessment to elucidate the immunodominant B cell antigen determinants of endemic HCoVs has not been done. We hypothesized that a fraction of the general population also have antibodies generated during past encounters with ‘common cold’ coronaviruses that cross-react with proteins of epidemic CoVs. This may affect the dynamics of sporadic MERS outbreaks that mostly occur in the Middle East, and the current COVID-19 pandemic.

115 **Results**

116 To gain a deeper insight into human antibody responses to endemic HCoVs, we performed
117 phage-immunoprecipitation sequencing (PhIP-Seq) [29-31] on previously collected serum or
118 plasma samples obtained from a total of 1431 human subjects from three different cohorts.
119 These included i) healthy male adult blood donors (ABD) with diverse ethnic background and
120 nationality (Supplementary Figure S1A); ii) adult male and female participants of a national
121 cohort study—the Qatar Biobank (QBB) [32]—representing the general population
122 (Supplementary Figure S1B); and iii) pediatric outpatients and inpatients who were tested for
123 metabolic conditions unrelated to infection, chronic disease or cancer (Methods and
124 Supplementary Figure S1C). The samples were collected prior to the current COVID-19
125 outbreak (Methods). In brief, PhIP-Seq allowed us to obtain comprehensive antiviral antibody
126 repertoires across individuals in our three human cohorts using phage display of
127 oligonucleotide-encoded peptidomes, followed by immunoprecipitation and massive parallel
128 sequencing [29, 31]. The VirScan phage library used for PhIP-Seq in the present study
129 comprised peptides derived from viral proteins—each represented by peptide tiles of up to
130 56 amino acids in length that overlap by 28 amino acids—which collectively encompass the
131 proteomes of a large number of viral species, including HCoV-229E, -NL63, -HKU1 and -OC43
132 [29, 31]. Proteins of endemic HCoVs which were represented in the VirScan phage library
133 included the Orf1ab replicase polyprotein (pp1ab), the spike glycoprotein (S), the matrix
134 glycoprotein (M), the nucleocapsid protein (N) and gene products of the species- and strain-
135 specific open reading frames (ORFs) encoded in the 3' region of the viral genomes
136 (Supplementary Table S1). Of note, we utilized an expanded version of the VirScan phage
137 library [33, 34], which also encompassed peptides from a number of proteins of human
138 epidemic and non-human CoV isolates, including MERS-CoV, SARS-CoV, as well as bat, bovine,
139 porcine and feline isolates belonging to the alpha- and betacoronavirus genera, albeit with
140 varying coverage of the viral peptidomes owing to the limitation in available sequence data
141 for the latter isolates in UniProt (Supplementary Table S1). SARS-CoV-2 peptides were not
142 included in the VirScan phage library used in our study.
143
144 We were able to obtain antibody repertoires for a total of 1399 individuals from the human
145 cohorts described above (Supplementary Table S2). Using stringent filter criteria (Methods),

146 we identified a total of 417 out of 2498 peptides and potential antigens from endemic HCoVs
147 with our screen that were significantly enriched in at least three of all 1399 analyzed
148 individuals. A total of 103 peptides from endemic HCoV were enriched in ≥ 1% of the samples
149 and therefore considered to contain potentially immunodominant regions (Supplementary
150 Table S3). Only 33 of the 417 peptides enriched in at least three samples shared linear
151 sequence homology with epitopes that have previously been reported [35] (Supplementary
152 Figure S2). To estimate number of newly identified linear B cell epitopes, we assigned each
153 CoV-derived peptide to clusters of peptides that share ≥ 7 amino acids linear sequence
154 identity—the estimated size of a linear B cell epitope (Methods). The enriched peptides could
155 be assigned to 149 clusters for which at least 2 peptides share linear sequence identity of ≥ 7
156 amino acids (Supplementary Tables S3 and S4). Only 13 clusters also shared ≥ 7 amino acids
157 linear sequence identity with known linear B cell epitopes. Consequently, we have identified
158 a minimum of 136 new linear epitopes, including 25 new immunodominant linear B cell
159 epitopes [i.e. B cell epitopes targeted in at least ≥ 1% of all individuals and not already
160 reported in IEBD (www.iebd.org) [35] (Supplementary Tables S3)].
161
162 Next we assessed the seroprevalence of HCoV-229E, -NL63, -HKU1 and -OC43 in the three
163 cohorts separately. To do so, we imputed species score values as described earlier [31, 33,
164 36] by counting the significantly enriched peptides for a given HCoV species that share less
165 than 7 amino acids linear sequence identity. We considered an individual seropositive for any
166 of the endemic HCoVs if the number of non-homologous peptides enriched in a given sample
167 met our previously established species-specific cut-off value (Methods). Seroprevalence for
168 endemic HCoVs ranged from ~4% to ~27%, depending on the species and cohort (Figure 1A).
169 Interestingly, we found a marginal but significant negative association between age and
170 seroprevalence of HCoV-OC43 (β = -0.175) and -NL63 (β = -0.315) (Figure 1B and 1D), as well
171 as a marginal positive association between male gender and seroprevalence for any of the
172 endemic HCoVs ($\beta \leq 0.2$) (Figure 1C and 1D). The species score values (i.e. the antibody
173 repertoire breadth for each HCoV species) did not differ substantially between seropositive
174 individuals of our 3 cohorts (Supplementary Figure S3). However, principal component
175 analysis revealed considerable qualitative differences in the antibody repertoires between
176 our cohorts and in particular between pediatric and adult subjects (Figure 2A). For

177 comparison, we also performed the same analysis on enriched peptides from other common
178 respiratory viruses, including HRSV, HRV A, HRV B and influenza B virus. As expected,
179 seroprevalence was considerably higher (68% to 99%) for HRSV, HRV A and HRV B, and
180 somewhat higher (29% to 47%) for influenza B virus (Supplementary Figure S4A). However,
181 contrary to antiviral antibody responses to endemic HCoVs, we did not find considerable
182 variance in the antibody repertoires to other respiratory viruses when comparing age groups
183 and cohorts (Supplementary Figure S4B). We also analyzed the enriched antigenic peptides
184 for each endemic HCoV species separately and found that most variance in the antibody
185 repertoires between cohorts and age groups were attributable to past infections with HCoV-
186 HKU1 and -229E (Supplementary Figure S4C). To determine the antibody specificities
187 responsible for most of the variance in the antiviral response to endemic HCoVs between
188 adults and children (i.e. to identify those peptides that were significantly more or less
189 frequently enriched when comparing adult and pediatric donors), we applied Fisher's exact
190 test and computed odds ratios for each of the significantly enriched peptides. We found that
191 antibody specificities in samples of pediatric study subjects predominantly targeted different
192 antigenic regions in the S protein (mean log odds ratio = 3.35 $\pm$ 2.12), the N protein (mean log
193 odds ratio = 2.21 $\pm$ 1.41) and diverse antigenic sites in pp1ab, whereas peptides encoding a
194 single linear B cell epitope of pp1ab (cluster 22) appeared to be the predominant target of
195 IgG antibodies among adult donors (mean log odds ratio = -4.7 $\pm$ 1.16) (Figure 2B and Table
196 1).
197
198 Intriguingly, multiple sequence alignments of frequently enriched peptides with the full-
199 length proteins of various CoVs revealed that antibody specificities predominantly found in
200 pediatric study subjects target immunodominant epitopes that encode functionally important
201 and highly conserved regions of the structural proteins. These included regions in the S_1
202 subunit of the S protein which are important for receptor binding (Figure 3A-C and 3F) [37-
203 40], as well as the regions resembling the proteolytic cleavage sites and fusion peptide of the
204 S_2 subunit (Figures 3A, 3B, 3D-F). Of note, the immunodominant region spanning the furin-
205 like S2' cleavage site in the S_2 subunit resembles one of the most conserved regions of the S
206 protein, both in amino acid sequence (R↓SA[I/L]ED[I/L]LF) (Figure 3E) and in protein structure,
207 as it forms an accessible alpha-helix within the fusion peptide region (Figure 3F and

208 Supplementary Figure S6) [41]. Moreover, we identified potential antibody binding sites in
209 the N-terminal RNA binding domain, serine-rich region and in the C-terminal dimerization
210 domain of the N protein (Figures 4A and 4B). Although the predicted antibody binding sites
211 in the N-terminal RNA-binding domain and the C-terminal dimerization domain of the N
212 protein appeared to be less conserved between different species in the primary amino acid
213 sequence (Figure 4C and 4D), both domains are structurally conserved in the regions that we
214 found to be immunodominant (Figure 4E and 4F). We also found that antibodies in children
215 targeted more frequently the C-terminal domain of the M protein (Supplementary Figure S5C
216 and Table 1) and the small accessory Orf8 protein (also known as N2) of HCoV-HKU1 (Table
217 1). Although Orf8 and N share the same coding sequence in the viral RNA genome, the reading
218 frame is different and the amino acid sequences not homologous. On the contrary, antibody
219 specificities predominantly found in adults primarily targeted a region of the pp1ab that is
220 specific to HCoV-HKU1 and contains an acidic tandem repeat (NDDE[D/H]VVTGD) which is
221 located upstream of the papain-like protease 1 domain (Supplementary Figure S5D).
222
223 Given the high degree of sequence conservation among some of the immunodominant
224 regions in proteins of endemic HCoVs we have identified, we also explored the extent to
225 which antibody specificities to these regions cross-react with peptides from epidemic CoVs
226 and non-human CoV isolates. For this purpose, we assessed the enrichment of peptides
227 derived from SARS-CoV, MERS-CoV as well as bovine, porcine, bat and feline CoV isolates
228 (Supplementary Table S1) applying the same approach and stringent filter criteria as
229 described above, for peptides of endemic HCoVs. Indeed, we identified several S protein- and
230 N protein-derived peptides from epidemic CoVs or non-human isolates that were significantly
231 enriched in our PhIP-Seq assay, which share sequence similarity with peptides from HCoVs
232 (Figure 5). As expected based on the results from multiple sequence alignments described
233 above, antibody specificities targeting the highly conserved amino acid motif
234 (RSA[I/L]ED[I/L]LF) spanning the furin-like S2' cleavage site of the S protein were broadly
235 cross-reactive to several orthologous peptides from MERS-CoV, SARS-CoV and non-human
236 CoV isolates (Figure 5A, Region 3). Cross-reactivity of antibodies targeting other functionally
237 important but less conserved regions of the S protein also appeared to be more restricted
238 (Figure 5A, Region 1 and 2). Antibody specificities targeting the N protein also showed
239 considerable cross-reactivity with peptides from MERS-CoV, SARS-CoV and non-human CoV

240 isolates. However, the latter cross-reactive antibodies mainly targeted regions rich in serine
241 and arginine residues, with low-complexity sequences and very limited structural
242 conservation, particularly an intrinsically disordered region (IDR) at the N terminus of the N
243 protein (Figure 5B and Supplementary Figure 7) that lacks a tertiary structure [42]. We also
244 detected cross-reactive antibodies targeting the serine- and arginine-rich linker region of the
245 N protein; however, cross-reactivity to this region was largely restricted to peptides derived
246 from non-human CoV isolates of domestic animals (Figure 5B) which are more closely related
247 to HCoV-OC43 [6].
248
249 Finally, we assessed plasma samples of a previously healthy 49-year-old female adult patient
250 who suffered from severe ARDS from SARS-CoV-2 infection, requiring prolonged
251 hospitalization and intensive care. A first sample was obtained 25 days after onset of
252 symptoms and 18 days after ICU admission. A second sample was obtained one month later,
253 one week after discharge. We compared the antibody profiles at both time points to samples
254 obtained from uninfected family members of the patient, as well as an age- and gender-
255 matched unrelated control. In agreement with what we found in some subjects with a history
256 of endemic HCoV infection, the patient with severe COVID-19 had detectable antibodies that
257 cross-reacted with peptides from SARS-CoV and non-human CoVs encoding the furin-like S2'
258 cleavage site and heptad repeat 2 region of the S protein, peptides from the C-terminal region
259 of the HCoV-HKU1 N protein downstream of the dimerization domain, as well as two antigenic
260 sites of the MERS-CoV pp1a (Supplementary Figure S8).

261 ## Discussion

262 Our comprehensive and systematic screen for antiviral antibody repertoires across individuals
263 in three human cohorts revealed a large number of peptides with novel linear B cell epitopes
264 in several proteins of endemic HCoVs. This is not surprising given that epidemic CoVs, and in
265 particular SARS-CoV, have been the primary focus of previous immunological and epitope
266 screening studies [35, 43]. Information about the targets of immune responses to CoVs across
267 different species provides a valuable resource for the prediction of candidate targets of newly
268 emerging CoVs, as recently shown by Grifoni *et al.* [43]. The authors were able to identify *a*
269 *priori,* several specific regions of the S, M and N proteins of SARS-CoV-2 on the basis of
270 sequence homology to the SARS-CoV virus which are orthologous to several of the
271 immunodominant regions of endemic HCoVs we identified. We detected antibodies against
272 the structural S, N, M and Orf8 proteins, as well as the non-structural pp1ab polyprotein of
273 HCoVs; the latter resembling the precursor for the large viral replicase complex [44].
274 Interestingly, in another independent study, Grifoni *et al.* [28] recently reported similarly
275 broad T cell responses in COVID-19 patients by employing an analogous screen for T cell
276 epitopes of SARS-CoV-2 proteins using peptide ‘‘megapools’’ in combination with *ex vivo* T
277 cell assays.
278
279 Surprisingly, circulating IgG antibodies in children appear to be differentially targeting
280 structural and non-structural proteins of HCoVs in comparison to adults (Figure 2). Whereas
281 antibody specificities more frequently found in samples of pediatric subjects targeted
282 structural proteins such as the S, N and M proteins, in adult donors, a region of the non-
283 structural polyprotein pp1ab containing an acidic tandem repeat (NDDE[D/H]VVTGD) in
284 HCoV-HKU1 appeared to be the predominant target of IgG antibodies. The latter polyprotein
285 is post-translationally processed into up to 16 subunits that form a large viral replicase
286 complex; however, the function of the acidic tandem repeat and its role in pathogenesis
287 remains unknown [44]. This qualitative difference in the antibody repertoires of adult versus
288 pediatric subjects appeared to be a specific characteristic of natural HCoV infection, as we did
289 not find the same degree of variance in the antibody repertoires specific to other common
290 respiratory viruses when comparing our cohorts and different age groups (Supplementary
291 Figure S3). We speculate that the qualitative differences in antibody repertoires of adult

292 versus children reflect a higher frequency and/or more recent exposures of children to
293 seasonal coronaviruses than adults, coupled with the rapid decay of circulating anti-CoV
294 antibodies that target the structural proteins of the virions. Further studies will be needed to
295 fully understand the dynamics of antibody responses to endemic HCoVs.
296
297 Evidence for the transient and dynamic nature of humoral immunity to endemic HCoVs has
298 been provided by number of human serological studies, although many of them were
299 conducted with only a limited number of subjects, or only for selected species, and a variety
300 of antibody detection methods were used that are not readily comparable [45-52].
301 Nevertheless, evidence suggests that a sizable proportion of children experience primary
302 infection with endemic HCoVs during their first year of life and nearly all children have
303 encountered at least one of the endemic HCoVs before two years of age, indicating that first
304 exposure to endemic HCoVs occurs very early in life, similar to other common respiratory
305 viruses such as HRSV or HRVs [46, 50, 51]. However, reported seroprevalence rates in older
306 children and adults vary greatly depending on a variety of factors, including age and viral
307 species. There is a general trend indicating that humoral immunity from primary infection
308 with endemic HCoVs wanes quickly and that antibodies detected in older children and adults
309 are rather a consequence of more recent re-infections [45-49]. We estimated the
310 seroprevalence of antibodies to endemic HCoVs to range between ~4% and ~27% depending
311 on the species and cohort (Figure 1A-C). Given that endemic HCoV infections are common
312 and usually acquired during early childhood [46, 47, 50, 51], it is likely that not only the adult
313 subjects, but also many (if not all) of the children aged 7 to 15 years that were assessed in our
314 study, have already experienced multiple infections with endemic HCoVs in their lifetime.
315 Therefore, our estimated seroprevalence rates likely reflect the complex dynamics between
316 rates of (re-)infection and waning humoral immunity over time. In agreement with this
317 notion, age was negatively associated with seroprevalence in our study, suggesting that the
318 duration of immunity in response to natural infection with endemic HCoVs and/or rates of re-
319 infection reduce with increasing age. The dynamics of humoral immunity from past CoV
320 infections is best described in studies of MERS and SARS patients. Although limited in sample
321 size, these studies have shown that antibody titers in all previously infected individuals
322 decline relatively quickly to minimal detectable levels over a period of 2 to 3 years and that
323 patients who suffered from more severe disease had higher and longer-lasting total binding

antibody titers and neutralizing titers [52]. There is also evidence that symptomatic COVID-19 patients mount robust antibody responses that wane very quickly over a period of 6 months [53]. However, the same study suggeted that SARS-CoV-2-specific memory B cells may be sustained over a longer period of time [53]. Indeed, most acute virus infections induce some level of protective and long-term immunity, albeit through a variety of mechanisms that are not necessarily the same for each pathogen and may even differ between hosts due to a variety of factors, including simultaneous viral coinfection [54, 55]. We also found a marginal but significant positive association between seroprevalence of endemic HCoVs and male gender, which is consistent with an earlier report by Gaunt *et al.* [5].

Despite the variable degree of sequence conservation between different CoV species, the results of our systematic antibody screen highlight that the structural proteins of the virions share common antigenic sites. Indeed, several of the immunodominant regions we have identified experimentally in the structural proteins of endemic HCoVs are orthologous to the regions thought to be immunodominant targets for immune responses to SARS-CoV-2 [43] (Supplementary Figure S8), including two linear epitopes on the SARS-CoV-2 S protein that elicit potent neutralizing antibodies in COVID-19 patients [56]. Importantly, antigenic regions that we found to be immunodominant in our study (i.e. enriched in $\geq$ 1% of all samples), as well as those corresponding to peptides for which enrichment was strongly and significantly associated with pediatric subjects (odds ratio $\geq$ 2; P-value $\leq$ 0.005, Fisher's exact test), mapped to functionally important regions of the structural CoV proteins. These included regions for receptor binding and the proteolytic cleavage sites of the S protein, as well as the N-terminal RNA-binding and C-terminal dimerization domains of the N protein, which have been shown to be critical for virus attachment and entry, cell-to-cell fusion and virus replication, respectively [42, 57-61]. The region of the S_1 subunit responsible for receptor binding differs considerably between CoV species, which utilize different domains and host cell receptors, and consequently differ in their tissue tropism [38-41, 47, 62]. However, the S_2 subunit resembling the fusion machinery is more conserved, both structurally and in amino acid sequence [39, 63] (Supplementary Figure S6). Indeed, we identified an immunodominant and highly conserved linear epitope immediately downstream of the furin-like S2' cleavage site of the S protein (R↓SA[I/L]ED[I/L]LF) that likely resembles the fusion peptide, although its precise location has been disputed [64]. The same antigenic site has recently been found on

356 the SARS-CoV-2 S protein to elicit neutralizing antibodies in COVID-19 patients [56]. The high
357 degree of amino acid sequence and conformational conservation of the alpha-helical region
358 immediately adjacent to the S2' cleavage site (Figures 3E and 3F) likely explains why
359 antibodies targeting this region also cross-reacted with orthologous peptides of related CoVs
360 in our study, including that of MERS-CoV, SARS-CoV and non-human isolates, further
361 supporting our overall hypothesis and the important role of this particular region as a pan-
362 CoV target site [41, 56]. It is therefore tempting to speculate that at least in some individuals,
363 past infections with endemic HCoVs have elicited cross-reactive antibodies and/or led to the
364 generation of longer-lived memory B cells with specific reactivity to this linear epitope, which
365 may provide cross-protection against MERS or COVID-19. This may be the case particularly
366 among children which are generally less likely to experience severe disease outcomes from
367 infection with epidemic CoVs [17-19, 21]. This will require further investigation.
368
369 Broadly cross-reactive antibody responses are also known for other enveloped RNA viruses,
370 which may positively or negatively affect subsequent infection or vaccination. Flaviviruses for
371 example, are antigenically related and broadly flavivirus cross-reactive antibodies from
372 previous yellow fever vaccination has been shown to impair and modulate immune responses
373 to tick-borne encephalitis vaccination [65]. Similarly, immune history has been shown to
374 profoundly affect protective B cell responses to influenza [66]. Since we detected pan-CoV
375 cross-reactive antibodies less frequently in plasma samples from adult donors, our results
376 argue against a strong therapeutic benefit of intravenous immunoglobulin products to control
377 the spread of COVID-19 disease [67]. In this context, it should be noted that large-scale
378 antibody screening by PhIP-Seq may frequently fail to detect conformational and post-
379 translationally modified B cell epitopes [29]. Nontheless, we found anti-CoV antibodies in
380 plasma of a COVID-19 patient after prolonged hospitalization and intensive care that targeted
381 largely the same structurally conserved and functionally important regions of the viral N and
382 S proteins (Supplementary Figure S8) as those that we detected in a sizable proportion of
383 children, including antibodies binding to the highly conserved motif and furin-like S2' cleavage
384 site (R↓SA[I/L]ED[I/L]LF), which provides further evidence for the clinical benefit of using
385 convalescent plasma for the prevention and treatment of COVID-19 [68-70]. Our findings may
386 also have important implications for the development of prophylactic or therapeutic
387 monoclonal antibodies and vaccine design in the context of COVID-19 [43, 71].

Methods

Study design and samples. We performed a retrospective analysis of deidentified or coded plasma and serum samples collected from three different human cohorts, namely: i) 400 healthy male adult blood donors (ABD) of a blood bank in Qatar with diverse ethnic background and nationality (Supplementary Figure S1A); ii) 800 adult male and female Qatari nationals and long-term residents of Qatar who are participating in a national cohort study—the Qatar Biobank (QBB)—and who represent the general local population in the State of Qatar [32]; and iii) 231 pediatric subjects with Qatari nationality who were admitted to, or visited outpatient clinics of, Sidra Medicine. Leftover plasma samples from healthy blood bank donors were collected from 2012 to 2016, de-identified and stored at -80°C. For the purpose of this study, specimen from male Qatari nationals 19 to 66 years of age (Supplementary Table S1) were selected from a larger blood donor cohort including 5983 individuals, and then age-matched male donors with other nationalities were randomly selected (Supplementary Figure S1). Samples from female blood bank donors were excluded because they were largely underrepresented in the blood bank donor cohort. We also excluded samples for which age, gender or nationality information was lacking. Serum samples from the QBB cohort were collected from 2012 to 2017 and were randomly selected samples from the first 3000 individuals taking part in a longitudinal cohort study as described previously [32]. Plasma samples from pediatric patients were selected from leftovers of samples processed in the clinical chemistry labs of Sidra Medicine, a tertiary care hospital for children and women in Doha, Qatar, over a period of several months from September to November 2019. In order to select appropriate pediatric samples, electronic medical records were queried using Discern analytics to identify blood samples from Qatari nationals aged 7 to 15 years for whom basic metabolic panel (BMP) and comprehensive metabolic panel (CMP) testing was done in the previous week. Samples from oncology patients, patients requiring complex care and those in intensive care units, as well as samples from patients with chronic diseases, samples with no centile data and samples from patients who were underweight (centile <5%) or overweight were excluded. However, we included obese patients in our analysis, since a considerable proportion of Qatari nationals are overweight. The COVID-19 patient assessed in this study was a previously healthy female Belgian national with autosomal recessive IRF7 deficiency who developed ARDS following SARS-CoV-2 infection at the age of 49 [72]. For comparison,

419 we also assessed unexposed family members, including the father, mother, brother
420 (heterozygous carriers) and wild-type sister, as well as an unrelated age- and gender-matched
421 healthy control. The human subject research described here had been approved by the
422 institutional research ethics boards of Sidra Medicine, Qatar Biobank, INSERM, and Erasme
423 Hospital.
424
425 **Phage Immunoprecipitation-Sequencing (PhIP-Seq).** Large scale serological profiling of the
426 antiviral IgG repertoires in the individual serum or plasma samples was performed as
427 described by Xu *el al.* [31]. Each serum or plasma sample was tested in duplicate and samples
428 were analyzed in batches with up to 96 samples per batch. Only samples that satisfied a
429 minimum read count of $1*10^6$ as well as a Pearson correlation coefficient of ≥ 0.7 in the two
430 technical repeats were considered for downstream analysis. Data from thirty individuals of
431 the ABD cohort and two individuals of the QBB cohort were excluded from the downstream
432 analysis due to insufficient sequencing read depth, low sequencing data quality or because
433 one of two technical replicates had failed (data not shown).
434
435 **Peptide enrichment analysis.** To filter for enriched peptides, we first imputed −log10(P)
436 values as described previously [31, 33, 36] by fitting a zero-inflated generalized Poisson model
437 to the distribution of output counts and regressed the parameters for each peptide sequence
438 based on the input read count. We considered a peptide enriched if it passed a reproducibility
439 threshold of 2.3 [−log10(P)] in two technical sample replicates. To remove sporadic hits, we
440 then filtered for antibody specificities to CoV peptides that were found to be enriched in at
441 least three of all 1399 subjects assayed and analyzed in this study. We computed species-
442 specific significance cut-off values to estimate minimum number of enriched, non-
443 homologous peptides required in order to consider a sample as seropositive using a
444 generalized linear model and in-house serological (ELISA) data from pooled samples that were
445 tested positive for various viruses. We then computed virus score values as described by Xu
446 et al [31] by counting enriched, non-homologous peptides for a given species and then
447 adjusted these score values by dividing them with the estimated score cutoff. For the purpose
448 of this study and under consideration of the seroprevalence of endemic HCoVs in our three
449 cohorts, we considered a peptide to be immunodominant if it was enriched in $\geq 1\%$ of all 1399
450 subjects assayed and analyzed in this study.

451

452 **Association studies and differential enrichment analysis.** We applied a generalized linear
453 model to test for associations between the HCoV species-specific adjusted score values,
454 gender and age. We considered an association to be significant if the P-value was ≤ 0.001. We
455 examined the frequency distribution of enriched peptides among samples of the different age
456 groups (PED versus ABD + QBB) by estimating odds ratios (OR) and associated P-values using
457 Fisher's exact test. Peptides that satisfied both significance (P-value ≤ 0.005) and magnitude
458 criteria ($|\log(OR)| \geq \log(2)$) were considered to be differentially enriched. Positive log(OR)
459 values indicated more frequent peptide enrichment among pediatric study subjects, whereas
460 negative log(OR) values indicated more frequent peptide enrichment among adult subjects.
461

462 **Clustering of peptides for shared linear B cell epitopes.** To estimate the minimum number
463 of linear B cell epitopes among the enriched peptides, we built a pairwise distance matrix that
464 captured the maximum size of linear sequence identity of amino acids ($d_{i,j}$) between all
465 enriched peptides. Groups of peptides that shared $\geq$ seven amino acid linear sequence
466 identity ($d_{i,j} \geq 7$) were assigned to a cluster. Peptides of a given cluster were considered to
467 share a linear B cell epitope (Supplementary Figure S7).
468

469 **Software.** Open source Python modules with in-house scripts were used to test for
470 associations (statsmodel v0.11), to filter differentially enrichment peptides and to perform
471 different statistical tests (sklearn v0.23, scipy v1.14.1). Multiple sequence alignments were
472 done using the MAFFT [73, 74] via EMBL-EBI's web services and Java Alignment Viewer
473 (Jalview) for visualization [75]. Residue-wise linear B cell epitopes were predicted using
474 BepiPred-2.0 [76]. Protein structures graphics were generated using PyMOL (Schrödinger).

Author contribution statement

TK & NM conceived the original idea, designed the models and the computational framework of the study, analyzed the data and wrote manuscript. MR, FA, SH, QZ, PB, ZL, EJ, VB, AC, HY, MS, LA and JLC planned or performed the experiments. GN, AS, IV, JCG, GS, IM, FH, IM and MH contributed samples and data. MR, JLC and MH provided critical feedback to this version of the manuscript. All authors have seen and approved the manuscript. It has not been accepted or published elsewhere.

Acknowledgement

We would like to thank the Qatar Biobank (QBB) management and staff, in particular Dr. Nahla Afifi and Elizabeth Jose, for their time and effort allowing us to access and analyze samples and data from the Qatar Biobank, and Dr. Stephan Elledge (Brigham and Women's Hospital and Harvard Medical School, Boston, MA) for kindly providing the VirScan phage library used in this study. We would also like to thank Patrick Tang, Mohammed Yousuf Karim, Evonne Chin-Smith and Damien Chaussabel (Sidra Medicine) for critical reading of the manuscript. This work was supported in part by a grant from the Qatar National Research Fund (PPM1-1220-150017) and funds from Sidra Medicine.

Potential competing interests

None of the authors has any competing interests to declare.

497 **References**

498

499 1. Hamre, D. and J.J. Procknow, *A New Virus Isolated from the Human Respiratory*
500 *Tract.* Proceedings of the Society for Experimental Biology and Medicine, 1966.
501 **121**(1): p. 190-193.

502 2. McIntosh, K., et al., *Recovery in tracheal organ cultures of novel viruses from patients*
503 *with respiratory disease.* Proceedings of the National Academy of Sciences, 1967.
504 **57**(4): p. 933-940.

505 3. van der Hoek, L., et al., *Identification of a new human coronavirus.* Nature Medicine,
506 2004. **10**(4): p. 368-373.

507 4. Woo, P.C.Y., et al., *Characterization and Complete Genome Sequence of a Novel*
508 *Coronavirus, Coronavirus HKU1, from Patients with Pneumonia.* Journal of Virology,
509 2005. **79**(2): p. 884-895.

510 5. Gaunt, E.R., et al., *Epidemiology and clinical presentations of the four human*
511 *coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel*
512 *multiplex real-time PCR method.* J Clin Microbiol, 2010. **48**(8): p. 2940-7.

513 6. Corman, V.M., et al., *Hosts and Sources of Endemic Human Coronaviruses.* Adv Virus
514 Res, 2018. **100**: p. 163-188.

515 7. Kristoffersen, A.W., et al., *Coronavirus causes lower respiratory tract infections less*
516 *frequently than RSV in hospitalized Norwegian children.* Pediatr Infect Dis J, 2011.
517 **30**(4): p. 279-83.

518 8. Drosten, C., et al., *Identification of a novel coronavirus in patients with severe acute*
519 *respiratory syndrome.* N Engl J Med, 2003. **348**(20): p. 1967-76.

520 9. Zaki, A.M., et al., *Isolation of a Novel Coronavirus from a Man with Pneumonia in*
521 *Saudi Arabia.* New England Journal of Medicine, 2012. **367**(19): p. 1814-1820.

522 10. Zhu, N., et al., *A Novel Coronavirus from Patients with Pneumonia in China, 2019.*
523 New England Journal of Medicine, 2020. **382**(8): p. 727-733.

524 11. Dong, E., H. Du, and L. Gardner, *An interactive web-based dashboard to track COVID-*
525 *19 in real time.* The Lancet Infectious Diseases, 2020.

526 12. de Wit, E., et al., *SARS and MERS: recent insights into emerging coronaviruses.* Nat
527 Rev Microbiol, 2016. **14**(8): p. 523-34.

528 13. *Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) —*
529 *United States, February 12–March 16, 2020.* MMWR Morb Mortal Wkly Rep, 2020.
530 **69**: p. 343-346.

531 14. Li, X. and X. Ma, *Acute respiratory failure in COVID-19: is it "typical" ARDS?* Critical
532 Care, 2020. **24**(1): p. 198.

533 15. Vijgen, L., et al., *Evolutionary History of the Closely Related Group 2 Coronaviruses:*
534 *Porcine Hemagglutinating Encephalomyelitis Virus, Bovine Coronavirus, and Human*
535 *Coronavirus OC43.* Journal of Virology, 2006. **80**(14): p. 7270-7274.

536 16. Young, B.E., et al., *Epidemiologic Features and Clinical Course of Patients Infected*
537 *With SARS-CoV-2 in Singapore.* JAMA, 2020.

538 17. Dong, Y., et al., *Epidemiology of COVID-19 Among Children in China.* Pediatrics, 2020:
539 p. e20200702.

540 18. Zimmermann, P. and N. Curtis, *Coronavirus Infections in Children Including COVID-19:*
541 *An Overview of the Epidemiology, Clinical Features, Diagnosis, Treatment and*
542 *Prevention Options in Children.* The Pediatric Infectious Disease Journal, 2020. **39**(5):
543 p. 355-368.

544 19. Al-Tawfiq, J.A., R.F. Kattan, and Z.A. Memish, *Middle East respiratory syndrome*
545 *coronavirus disease is rare in children: An update from Saudi Arabia.* World J Clin
546 Pediatr, 2016. **5**(4): p. 391-396.

547 20. Zhang, J., et al., *Changes in contact patterns shape the dynamics of the COVID-19*
548 *outbreak in China.* Science, 2020: p. eabb8001.

549 21. Davies, N.G., et al., *Age-dependent effects in the transmission and control of COVID-*
550 *19 epidemics.* Nature Medicine, 2020.

551 22. Guo, L., et al., *Profiling Early Humoral Response to Diagnose Novel Coronavirus*
552 *Disease (COVID-19).* Clinical Infectious Diseases, 2020.

553 23. Okba, N.M.A., et al., *Severe Acute Respiratory Syndrome Coronavirus 2-Specific*
554 *Antibody Responses in Coronavirus Disease 2019 Patients.* Emerg Infect Dis, 2020.
555 **26**(7).

556 24. Yongchen, Z., et al., *Different longitudinal patterns of nucleic acid and serology*
557 *testing results based on disease severity of COVID-19 patients.* Emerg Microbes
558 Infect, 2020. **9**(1): p. 833-836.

559 25. Khan, S., et al., *Analysis of Serologic Cross-Reactivity Between Common Human*
560 *Coronaviruses and SARS-CoV-2 Using Coronavirus Antigen Microarray.* bioRxiv, 2020:
561 p. 2020.03.24.006544.

562 26. Smatti, M.K., A.A. Al Thani, and H.M. Yassine, *Viral-Induced Enhanced Disease Illness.*
563 Front Microbiol, 2018. **9**: p. 2991.

564 27. Braun, J., et al., *SARS-CoV-2-reactive T cells in healthy donors and patients with*
565 *COVID-19.* Nature, 2020.

566 28. Grifoni, A., et al., *Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans*
567 *with COVID-19 Disease and Unexposed Individuals.* Cell, 2020.

568 29. Mohan, D., et al., *Publisher Correction: PhIP-Seq characterization of serum antibodies*
569 *using oligonucleotide-encoded peptidomes.* Nat Protoc, 2019. **14**(8): p. 2596.

570 30. Mohan, D., et al., *PhIP-Seq characterization of serum antibodies using*
571 *oligonucleotide-encoded peptidomes.* Nat Protoc, 2018. **13**(9): p. 1958-1978.

572 31. Xu, G.J., et al., *Viral immunology. Comprehensive serological profiling of human*
573 *populations using a synthetic human virome.* Science, 2015. **348**(6239): p. aaa0698.

574 32. Al Kuwari, H., et al., *The Qatar Biobank: background and methods.* BMC Public
575 Health, 2015. **15**: p. 1208.

576 33. Mina, M.J., et al., *Measles virus infection diminishes preexisting antibodies that offer*
577 *protection from other pathogens.* Science, 2019. **366**(6465): p. 599-606.

578 34. Drutman, S.B., et al., *Fatal Cytomegalovirus Infection in an Adult with Inherited NOS2*
579 *Deficiency.* N Engl J Med, 2020. **382**(5): p. 437-445.

580 35. Vita, R., et al., *The Immune Epitope Database (IEDB): 2018 update.* Nucleic Acids Res,
581 2019. **47**(D1): p. D339-D343.

582 36. Pou, C., et al., *The repertoire of maternal anti-viral antibodies in human newborns.*
583 Nat Med, 2019. **25**(4): p. 591-596.

584 37. Tortorici, M.A., et al., *Structural basis for human coronavirus attachment to sialic*
585 *acid receptors.* Nat Struct Mol Biol, 2019. **26**(6): p. 481-489.

586 38. Tortorici, M.A. and D. Veesler, *Structural insights into coronavirus entry.* Adv Virus
587 Res, 2019. **105**: p. 93-116.

588 39. Rey, F.A. and S.M. Lok, *Common Features of Enveloped Viruses and Implications for*
589 *Immunogen Design for Next-Generation Vaccines.* Cell, 2018. **172**(6): p. 1319-1334.

590 40. Hulswit, R.J.G., et al., *Human coronaviruses OC43 and HKU1 bind to 9-O-*
591 *acetylated sialic acids via a conserved receptor-binding site in spike protein domain*
592 *A.* Proceedings of the National Academy of Sciences, 2019. **116**(7): p. 2681-2690.

593 41. Walls, A.C., et al., *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike*
594 *Glycoprotein.* Cell, 2020. **181**(2): p. 281-292 e6.

595 42. McBride, R., M. van Zyl, and B.C. Fielding, *The coronavirus nucleocapsid is a*
596 *multifunctional protein.* Viruses, 2014. **6**(8): p. 2991-3018.

597 43. Grifoni, A., et al., *A Sequence Homology and Bioinformatic Approach Can Predict*
598 *Candidate Targets for Immune Responses to SARS-CoV-2.* Cell Host Microbe, 2020.
599 **27**(4): p. 671-680 e2.

600 44. Ziebuhr, J., *The coronavirus replicase.* Curr Top Microbiol Immunol, 2005. **287**: p. 57-
601 94.

602 45. Monto, A.S. and S.K. Lim, *The Tecumseh study of respiratory illness. VI. Frequency of*
603 *and relationship between outbreaks of coronavirus infection.* J Infect Dis, 1974.
604 **129**(3): p. 271-6.

605 46. Shao, X., et al., *Seroepidemiology of group I human coronaviruses in children.* J Clin
606 Virol, 2007. **40**(3): p. 207-13.

607 47. Hofmann, H., et al., *Human coronavirus NL63 employs the severe acute respiratory*
608 *syndrome coronavirus receptor for cellular entry.* Proc Natl Acad Sci U S A, 2005.
609 **102**(22): p. 7988-93.

610 48. Hamre, D. and M. Beem, *Virologic studies of acute respiratory disease in young*
611 *adults. V. Coronavirus 229E infections during six years of surveillance.* Am J
612 Epidemiol, 1972. **96**(2): p. 94-106.

613 49. Chan, C.M., et al., *Examination of seroprevalence of coronavirus HKU1 infection with*
614 *S protein-based ELISA and neutralization assay against viral spike pseudotyped virus.*
615 J Clin Virol, 2009. **45**(1): p. 54-60.

616 50. Dijkman, R., et al., *The dominance of human coronavirus OC43 and NL63 infections in*
617 *infants.* J Clin Virol, 2012. **53**(2): p. 135-9.

618 51. Dijkman, R., et al., *Human coronavirus NL63 and 229E seroconversion in children.* J
619 Clin Microbiol, 2008. **46**(7): p. 2368-73.

620 52. Kellam, P. and W. Barclay, *The dynamics of humoral immune responses following*
621 *SARS-CoV-2 infection and the potential for reinfection.* J Gen Virol, 2020.

622 53. Vaisman-Mentesh, A., et al., *SARS-CoV-2 specific memory B cells frequency in*
623 *recovered patient remains stable while antibodies decay over time.* medRxiv, 2020: p.
624 2020.08.23.20179796.

625 54. Sallusto, F., et al., *From vaccines to memory and back.* Immunity, 2010. **33**(4): p. 451-
626 63.

627 55. Kenney, L.L., et al., *Increased Immune Response Variability during Simultaneous Viral*
628 *Coinfection Leads to Unpredictability in CD8 T Cell Immunity and Pathogenesis.*
629 Journal of Virology, 2015. **89**(21): p. 10786-10801.

630 56. Poh, C.M., et al., *Two linear epitopes on the SARS-CoV-2 spike protein that elicit*
631 *neutralising antibodies in COVID-19 patients.* Nature Communications, 2020. **11**(1):
632 p. 2806.

633 57. Tylor, S., et al., *The SR-rich motif in SARS-CoV nucleocapsid protein is important for*
634 *virus replication.* Can J Microbiol, 2009. **55**(3): p. 254-60.

635 58. Surjit, M. and S.K. Lal, *The SARS-CoV nucleocapsid protein: a protein with*
636 *multifarious activities.* Infect Genet Evol, 2008. **8**(4): p. 397-405.

637 59. Li, F., *Structure, Function, and Evolution of Coronavirus Spike Proteins.* Annu Rev
638 Virol, 2016. **3**(1): p. 237-261.

639 60. Follis, K.E., J. York, and J.H. Nunberg, *Furin cleavage of the SARS coronavirus spike*
640 *glycoprotein enhances cell-cell fusion but does not affect virion entry.* Virology, 2006.
641 **350**(2): p. 358-69.

642 61. Belouzard, S., et al., *Mechanisms of coronavirus cell entry mediated by the viral spike*
643 *protein.* Viruses, 2012. **4**(6): p. 1011-33.

644 62. Fung, T.S. and D.X. Liu, *Human Coronavirus: Host-Pathogen Interaction.* Annu Rev
645 Microbiol, 2019. **73**: p. 529-557.

646 63. Walls, A.C., et al., *Unexpected Receptor Functional Mimicry Elucidates Activation of*
647 *Coronavirus Fusion.* Cell, 2019. **176**(5): p. 1026-1039 e15.

648 64. Lai, A.L., et al., *The SARS-CoV Fusion Peptide Forms an Extended Bipartite Fusion*
649 *Platform that Perturbs Membrane Order in a Calcium-Dependent Manner.* J Mol Biol,
650 2017. **429**(24): p. 3875-3892.

651 65. Bradt, V., et al., *Pre-existing yellow fever immunity impairs and modulates the*
652 *antibody response to tick-borne encephalitis vaccination.* npj Vaccines, 2019. **4**(1): p.
653 38.

654 66. Andrews, S.F., et al., *Immune history profoundly affects broadly protective B cell*
655 *responses to influenza.* Science Translational Medicine, 2015. **7**(316): p. 316ra192-
656 316ra192.

657 67. Diez, J.M., C. Romero, and R. Gajardo, *Currently available intravenous*
658 *immunoglobulin contains antibodies reacting against severe acute respiratory*
659 *syndrome coronavirus 2 antigens.* Immunotherapy, 2020. **12**(8): p. 571-576.

660 68. Benner, S.E., et al., *SARS-CoV-2 Antibody Avidity Responses in COVID-19 Patients and*
661 *Convalescent Plasma Donors.* J Infect Dis, 2020.

662 69. Bloch, E.M., et al., *Deployment of convalescent plasma for the prevention and*
663 *treatment of COVID-19.* J Clin Invest, 2020. **130**(6): p. 2757-2765.

664 70. Casadevall, A. and L.A. Pirofski, *The convalescent sera option for containing COVID-*
665 *19.* J Clin Invest, 2020. **130**(4): p. 1545-1548.

666 71. Zhou, G. and Q. Zhao, *Perspectives on therapeutic neutralizing antibodies against the*
667 *Novel Coronavirus SARS-CoV-2.* Int J Biol Sci, 2020. **16**(10): p. 1718-1723.

668 72. Zhang, Q., et al., *Inborn errors of type I IFN immunity in patients with life-threatening*
669 *COVID-19.* Science, 2020: p. eabd4570.

670 73. Madeira, F., et al., *The EMBL-EBI search and sequence analysis tools APIs in 2019.*
671 Nucleic Acids Res, 2019. **47**(W1): p. W636-W641.

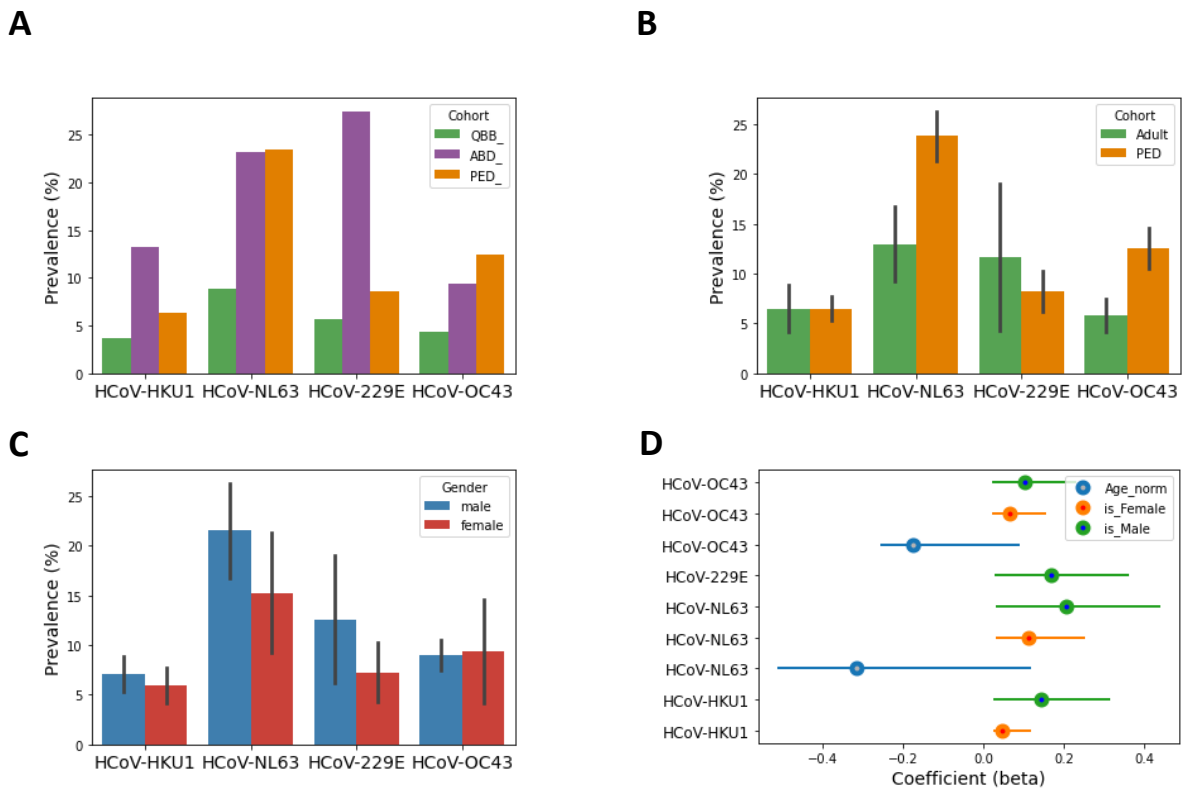
672 74. Katoh, K. and D.M. Standley, *MAFFT multiple sequence alignment software version 7:*
673 *improvements in performance and usability.* Mol Biol Evol, 2013. **30**(4): p. 772-80.

674 75. Madeira, F., et al., *Using EMBL-EBI Services via Web Interface and Programmatically*
675 *via Web Services.* Curr Protoc Bioinformatics, 2019. **66**(1): p. e74.

676 76. Dhanda, S.K., et al., *IEDB-AR: immune epitope database-analysis resource in 2019.*
677 Nucleic Acids Res, 2019. **47**(W1): p. W502-W506.

678
679

680 **Figures**

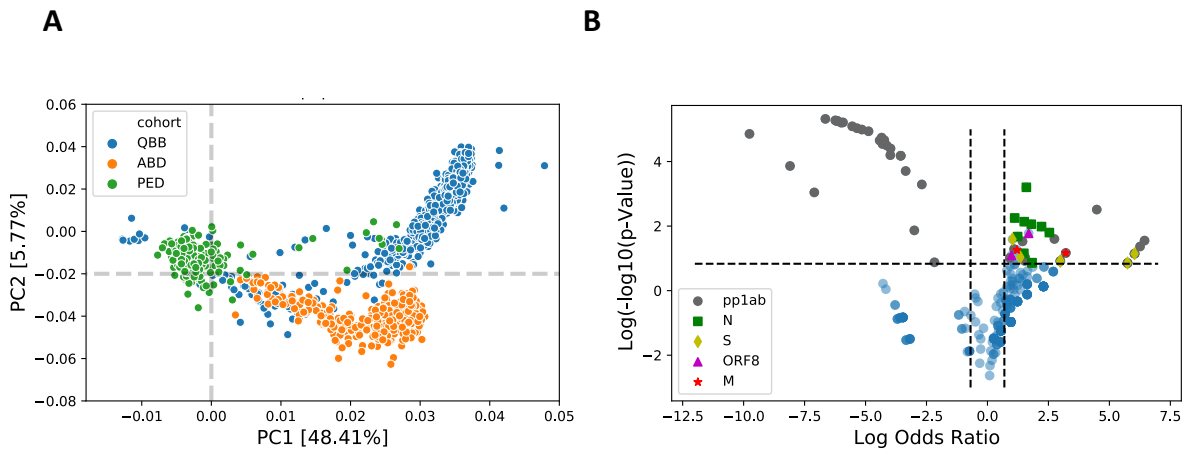


681

682

683 **Figure 1. Seroprevalence of endemic HCoVs. A-C,** Bar plots depict the seroprevalence of the
684 four endemic HCoVs stratified by cohort (A), age group (B) and gender (C). Error bars in (B)
685 depict gender-specific variation. Adults include adult blood bank donors and individuals of
686 the Qatar Biobank cohort. QBB, Qatar Biobank cohort; ABD, adult (male) blood bank donors;
687 PED, pediatric study subjects. **D**, Coefficient of association (beta) with 95% confidence interval
688 (95% CI) of seroprevalence for each HCoV with either male gender (green), female gender
689 (orange), or age (blue). Only features that had a P-value of association ≥ 0.001 are shown.

690



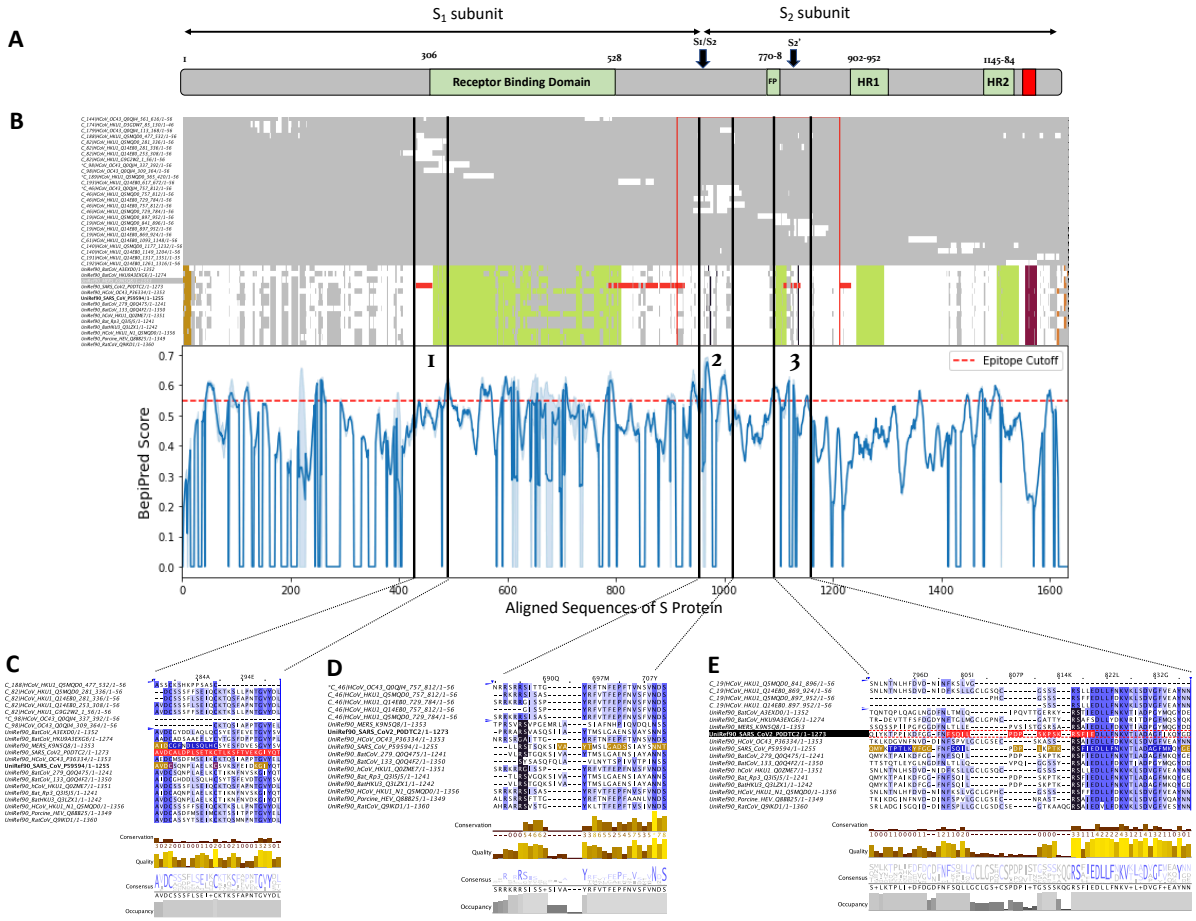
691

692

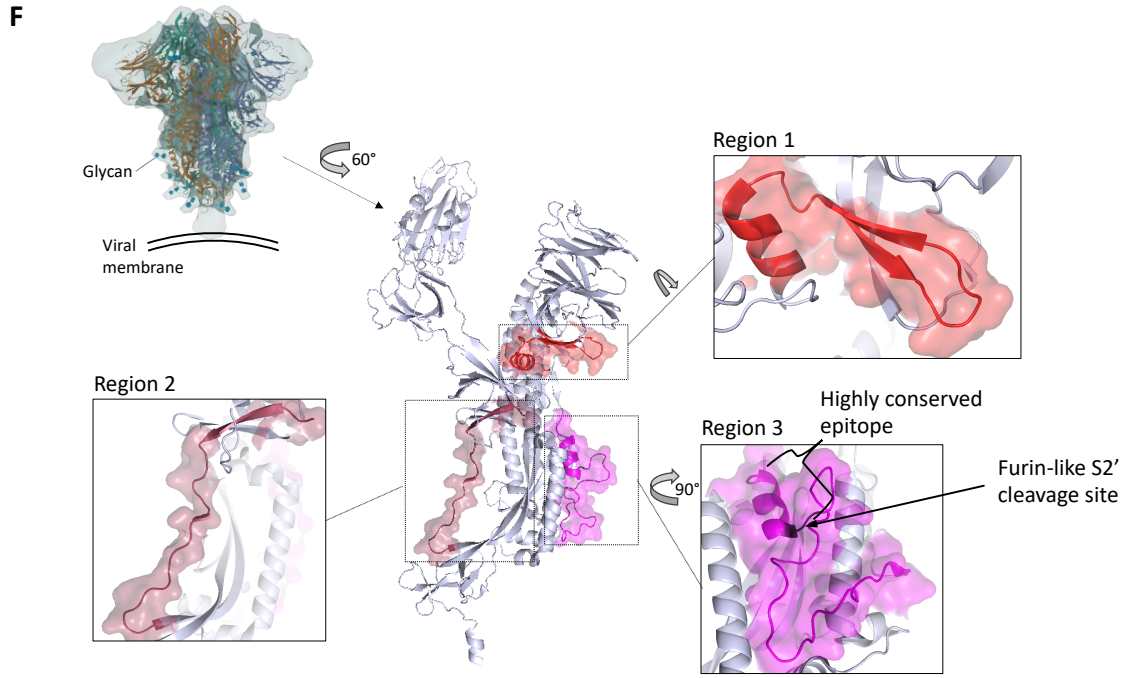
693 **Figure 2. Qualitative differences in antibody repertoires between cohorts and age groups.**
694 **A,** Principal component analysis of 417 peptides from endemic HCoVs that were found to be
695 enriched in at least 3 samples. QBB, Qatar Biobank cohort; ABD, adult (male) blood bank
696 donors; PED, pediatric subjects. **B,** Differential enrichment analysis to determine the peptides
697 that are either more or less frequently enriched in children versus adults (including subjects
698 of both adult cohorts, namely QBB and ABD). We considered a peptide significantly more or
699 less frequently enriched among children if the odds ratio (OR) was ≥ 2 or ≤ -2, respectively;
700 and the P-value was ≤ 0.005 (Fisher's Exact Test). pp1ab, Orf1ab replicase polyprotein; S, spike
701 glycoprotein; M, matrix glycoprotein; N, nucleocapsid protein; ORF8, open reading frame 8
702 protein.

703

704



705



706

707

708 **Figure 3. Antigenic regions and predicted antibody binding sites of the spike (S) protein. A,**
709 Schematic representation of the S protein of SARS-CoV (strain Tor2). Proteolytic cleavage sites
710 of the S protein are marked with arrows. **B,** Overview of a multiple sequence alignment of
711 immunodominant peptides with the full-length protein sequences of various alpha- and beta-
712 CoVs (top). The mean score (blue line) and standard deviations (shaded) for the residue-wise
713 prediction of linear B cell epitopes of aligned endemic HCoVs (bottom). Row labels (left)
714 indicate the cluster and sequence identifier, start and end position of the peptide and length
715 of the aligned sequence. Peptides for which differential enrichment between children and
716 adults was statistically significant (P-value ≤ 0.05, Fisher's exact test) and odds ratios were $\geq$
717 2 are indicated with a "*". The protein domains and boundaries shown in the schematic (A)
718 have been marked in different colors using UniProt annotation and JalView features. Amino
719 acid sequences of previously predicted immunodominant linear SARS-CoV-2 B cell epitopes
720 [43] are highlighted in pink. For the BepiPred score (bottom), a score cutoff of 0.55 has been
721 marked with a dashed red line to indicate regions that are predicted to be potential B cell
722 epitopes. **C-E,** Selected regions of the multiple sequence alignment encompassing regions 1,
723 2 and 3 as shown in (B). Proteolytic cleavage sites of the S protein are highlighted in black.
724 The full sequence alignment is shown in Supplementary Figure S5A. **F,** Ribbon diagram of the
725 SARS-CoV-2 S trimer with carbohydrate residues shown (top left) and monomer without
726 carbohydrate residues (middle) in the prefusion conformation (PDB id: 6VXX, chain A) [41];
727 regions 1, 2 and 3 are shown enlarged. FP, fusion peptide; HR1 and HR2, heptad repeat 1 and
728 2.
729

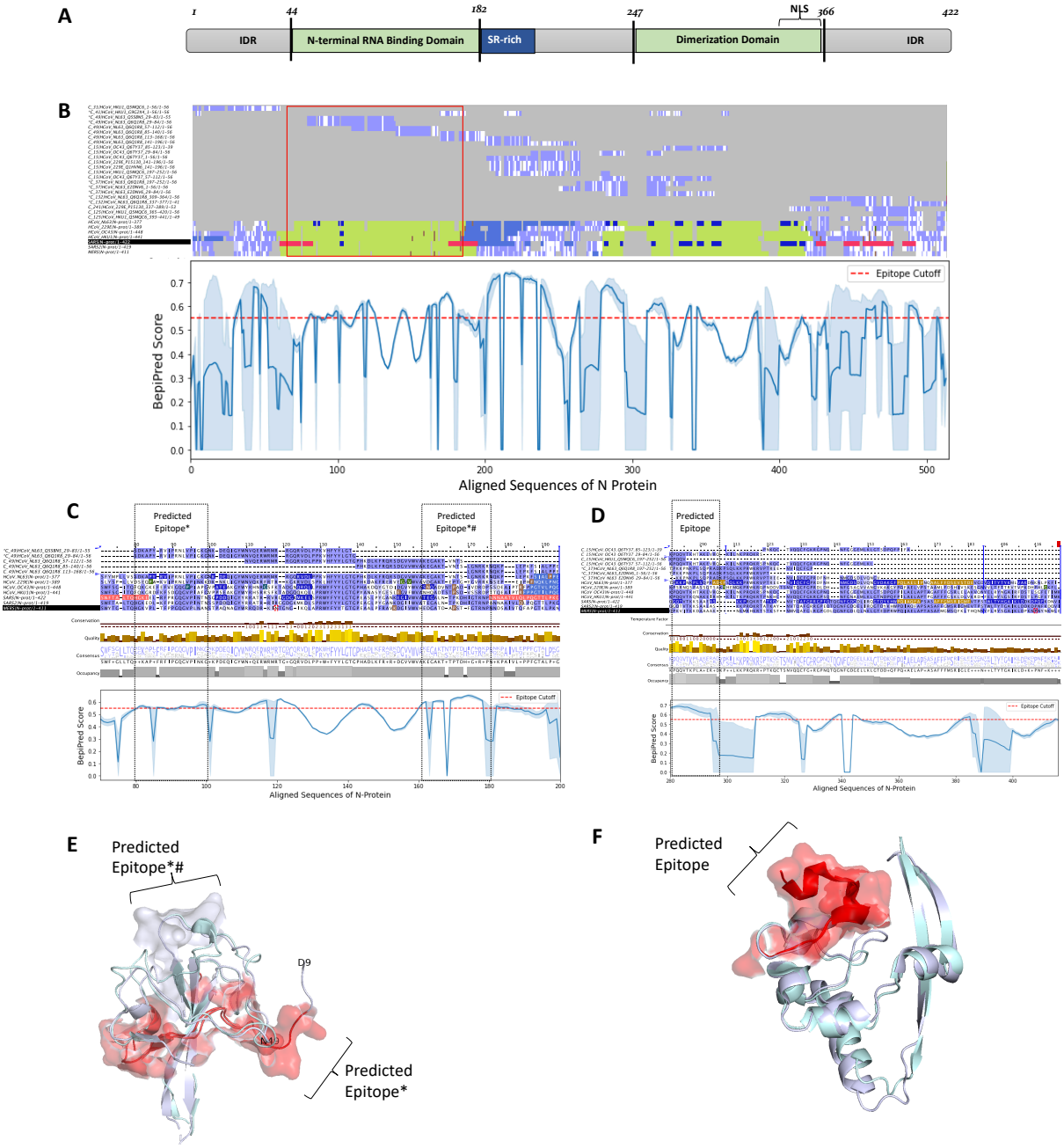
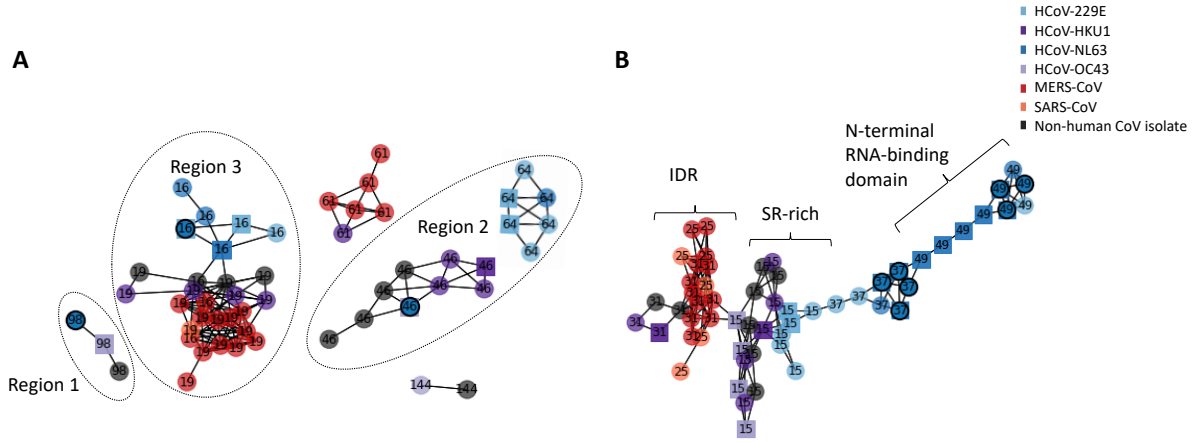


Figure 4. Antigenic regions and predicted antibody binding sites of the nucleocapsid (N) protein. A, Schematic representations of the N protein of SARS-CoV (strain Tor2). SR-rich, serine-rich; NLS, Predicted nuclear localization sequence; IDR, intrinsically disordered region. **B,** Overview of a multiple sequence alignment of immunodominant peptides with the full-length protein sequences of various alpha- and beta-CoVs (top) and mean score values for the prediction of linear B cell epitopes among endemic HCoVs (bottom). Row labels (left) indicate the cluster and sequence identifier, start and end position of the peptide and length

740 of the aligned sequence. Peptides for which differential enrichment between children and
741 adults was statistically significant (P-value ≤ 0.005, Fisher's exact test) and odds ratios were
742 ≥ 2 are indicated with a "*". The protein domains of full-length reference sequences and
743 boundaries shown in the schematic (A) have been marked in green. Amino acid sequences of
744 previously predicted immunodominant SARS-CoV B cell epitopes [43] are highlighted in pink
745 color. For the BepiPred score (bottom), a score cutoff of 0.55 has been marked with a dashed
746 red line to indicate regions that are predicted to be potential B cell epitopes. **C, D,** Selected
747 regions of the multiple sequence alignment encompassing the N-terminal RNA-binding
748 domain (C) and C-terminal self-assembly domain (D). The full multiple sequence alignment is
749 shown in Supplementary Figure S5B. **E,** Super-imposed ribbon structure of the N-terminal
750 RNA-binding domain from HCoV-NL63 (PDB: 5N4k, chain A) and that from SARS-CoV-2 (PDB:
751 6M3M, chain A) (root-mean-square deviation [rmsd] = 0.7 Ångström). **F,** Super-imposed
752 ribbon structure of the C-terminal self assembly domain of proteins from HCoV- NL63 (pdbId:
753 5EPW, chain A) and that of SARS-CoV-2 (pdbID: 6WJI, chain A) (rmsd = 0.91 Ångström). *,
754 Predicted epitopes in peptides for which differential enrichment between children and adults
755 was statistically significant (P-value ≤ 0.005, Fisher's exact test) and odds ratios were ≥ 2; #,
756 Predicted immunodominant epitopes.

757



758

759 **Figure 5. Network representation of enriched peptides from structural proteins targeted by**
760 **cross-reactive antibodies**. **A,** Network representation of enriched spike (S) protein-derived
761 peptides. **B,** Network representation of enriched nucleoprotein (N)-derived peptides. Each
762 node represents an enriched peptide and the color indicates the species. Edges indicate ≥
763 seven amino acids linear sequence identity between two nodes (i.e. peptides), the estimated
764 size of a linear B cell epitope. Only networks of peptides derived from at least two different
765 species are shown. Labels indicate the cluster number to which each peptide has been
766 assigned. Nodes are represented as spheres if the peptide had been frequently enriched.
767 Nodes marked with a black circle indicate peptides for which differential enrichment between
768 children and adults was statistically significant (P-value ≤ 0.005, Fisher's exact test) and odds
769 ratios were ≥ 2. SR-rich, serine- and arginine-rich motive; IDR, intrinsically disordered region.

770

771 **Table 1. List of peptides that were differentially enriched in children versus adults.**

772 Immunodominant peptides are marked in bold font.

Entry	Start	End	Protein	Species	Log2(OR)	-Log10(P value)	Cluster
P0C6X6	1401	1456	pp1ab	HCoV-OC43	6.443	4.717	136
Q0ZJJ1	813	868	pp1ab	HCoV-HKU1	6.257	3.927	102
P0C6X2	7029	7084	pp1ab	HCoV-HKU1	6.032	3.138	117
Q0ZJJ1	225	280	pp1ab	HCoV-HKU1	6.032	3.138	38
Q5MQD0	365	420	S	HCoV-HKU1	6.032	3.138	189
P0C6X5	3753	3808	pp1ab	HCoV-NL63	5.744	2.351	10
Q0QJI4	337	392	S	HCoV-OC43	5.744	2.351	98
Q0ZJJ1	5825	5880	pp1ab	HCoV-HKU1	5.744	2.351	9
P0C6X3	**5489**	**5544**	**pp1ab**	**HCoV-HKU1**	**4.483**	**12.331**	**8**
P0C6X6	2381	2436	pp1ab	HCoV-OC43	3.212	3.212	253
Q6Q1R9	197	226	M	HCoV-NL63	3.212	3.212	185
P0C6X1	1373	1428	pp1ab	HCoV-229E	2.987	2.500	235
Q6Q1S2	421	476	S	HCoV-NL63	2.987	2.500	131
Q6Q1S2	449	504	S	HCoV-NL63	2.987	2.500	131
P0C6X2	3305	3360	pp1ab	HCoV-HKU1	2.745	4.957	13
Q6Q1R8	**29**	**84**	**N**	**HCoV-NL63**	**2.539**	**6.042**	**49**
E2DNV6	**1**	**56**	**N**	**HCoV-NL63**	**2.217**	**7.278**	**37**
Q5SBN5	1	56	N	HCoV-NL63	1.823	2.374	49
Q6Q1R8	169	224	N	HCoV-NL63	1.823	2.374	37
Q6Q1R8	**197**	**252**	**N**	**HCoV-NL63**	**1.798**	**7.825**	**37**
Q5MQC5	**1**	**56**	**ORF8**	**HCoV-HKU1**	**1.690**	**5.895**	**41**
Q6Q1R8	**337**	**377**	**N**	**HCoV-NL63**	**1.596**	**24.578**	**132**
P0C6X2	**5461**	**5516**	**pp1ab**	**HCoV-HKU1**	**1.547**	**2.802**	**34**
P0C6X5	**4285**	**4340**	**pp1ab**	**HCoV-NL63**	**1.531**	**2.513**	**78**
E2DNV6	**29**	**84**	**N**	**HCoV-NL63**	**1.508**	**8.475**	**37**
Q5SBN5	**29**	**83**	**N**	**HCoV-NL63**	**1.492**	**3.178**	**49**
Q0ZJJ1	**1121**	**1176**	**pp1ab**	**HCoV-HKU1**	**1.437**	**4.602**	**12**
P15423	**645**	**700**	**S**	**HCoV-229E**	**1.347**	**2.822**	**16**
G9G2X4	**1**	**56**	**N**	**HCoV-HKU1**	**1.237**	**5.340**	**41**
Q01455	**197**	**230**	**M**	**HCoV-OC43**	**1.212**	**3.536**	**147**
Q6Q1R8	**309**	**364**	**N**	**HCoV-NL63**	**1.116**	**9.498**	**132**
P0C6X2	**4509**	**4564**	**pp1ab**	**HCoV-HKU1**	**1.091**	**3.625**	**2**
Q0QJI4	**757**	**812**	**S**	**HCoV-OC43**	**1.027**	**4.914**	**46**
Q5MQC5	**29**	**84**	**ORF8**	**HCoV-HKU1**	**0.984**	**2.982**	**41**
P0C6X2	**1065**	**1120**	**pp1ab**	**HCoV-HKU1**	**0.916**	**2.752**	**12**
P0C6X1	**5377**	**5432**	**pp1ab**	**HCoV-229E**	**-2.173**	**2.406**	**9**
Q0ZJJ1	**965**	**1020**	**pp1ab**	**HCoV-HKU1**	**-2.692**	**26.841**	**12**
P0C6X4	**897**	**952**	**pp1ab**	**HCoV-HKU1**	**-3.000**	**6.469**	**12**
Q0ZJG7	**925**	**980**	**pp1ab**	**HCoV-HKU1**	**-3.355**	**40.738**	**12**
P0C6X4	**925**	**980**	**pp1ab**	**HCoV-HKU1**	**-3.554**	**65.238**	**12**
P0C6X3	**925**	**980**	**pp1ab**	**HCoV-HKU1**	**-3.562**	**65.438**	**12**
P0C6X4	**953**	**1008**	**pp1ab**	**HCoV-HKU1**	**-3.977**	**81.879**	**12**
P0C6X2	**1037**	**1092**	**pp1ab**	**HCoV-HKU1**	**-3.983**	**66.553**	**12**
P0C6X4	**1009**	**1064**	**pp1ab**	**HCoV-HKU1**	**-4.132**	**89.820**	**12**
Q0ZJJ1	**1037**	**1092**	**pp1ab**	**HCoV-HKU1**	**-4.232**	**106.086**	**12**
P0C6X3	**1009**	**1064**	**pp1ab**	**HCoV-HKU1**	**-4.305**	**94.078**	**12**
P0C6X2	**951**	**1006**	**pp1ab**	**HCoV-HKU1**	**-4.339**	**114.588**	**12**
P0C6X3	**1037**	**1092**	**pp1ab**	**HCoV-HKU1**	**-4.421**	**105.026**	**12**
P0C6X4	**981**	**1036**	**pp1ab**	**HCoV-HKU1**	**-4.878**	**139.489**	**12**
P0C6X2	**925**	**980**	**pp1ab**	**HCoV-HKU1**	**-5.151**	**147.090**	**12**
Q0ZJJ1	**925**	**980**	**pp1ab**	**HCoV-HKU1**	**-5.348**	**153.250**	**12**
P0C6X3	**953**	**1008**	**pp1ab**	**HCoV-HKU1**	**-5.546**	**162.652**	**12**
P0C6X3	**951**	**1006**	**pp1ab**	**HCoV-HKU1**	**-5.924**	**183.493**	**12**
Q0ZJJ1	**1093**	**1148**	**pp1ab**	**HCoV-HKU1**	**-5.999**	**182.148**	**12**
P0C6X2	**949**	**1004**	**pp1ab**	**HCoV-HKU1**	**-6.155**	**192.846**	**12**
Q0ZJJ1	**953**	**1008**	**pp1ab**	**HCoV-HKU1**	**-6.176**	**190.949**	**12**
P0C6X2	**953**	**1008**	**pp1ab**	**HCoV-HKU1**	**-6.238**	**195.471**	**12**
Q0ZJJ1	**979**	**1034**	**pp1ab**	**HCoV-HKU1**	**-6.647**	**204.612**	**12**
P0C6X4	**4761**	**4816**	**pp1ab**	**HCoV-HKU1**	**-7.102**	**20.977**	**36**
Q0ZJG7	**4985**	**5040**	**pp1ab**	**HCoV-HKU1**	**-8.096**	**47.573**	**21**
P0C6X1	**337**	**392**	**pp1ab**	**HCoV-229E**	**-9.759**	**128.638**	**219**

773