Endemic human coronaviruses induce distinct antibody repertoires in adults and children

Taushif Khan[1], Mahbuba Rahman[1], Fatima Al Ali[1], Susie S. Y. Huang[1], Amira Sayeed[3], Gheyath K. Nasrallah[2], Mohammad R. Hasan[3,4], Nico Marr[1,5*]

Version 1, 21 June 2020

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

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

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

[4]Department of Pathology and Laboratory Medicine, Weill Cornell Medical College in Qatar, Doha, Qatar

[5]College of Health and Life Sciences, Hamad Bin Khalifa University, Doha, Qatar

*Corresponding author: Nico Marr, Sidra Medicine, Research Branch, Al Gharrafa Street, Ar-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 but immune responses to these “common cold” viruses remain incompletely understood. Moreover, there is evidence emerging from independent studies which suggests that endemic HCoVs can induce broadly cross-reactive T cell responses and may thereby affect clinical outcomes of acute infections with the phylogenetically related epidemic viruses, namely MERS-CoV and SARS-CoV-2. Here we report a comprehensive retrospective analysis of CoV-specific antibody specificities in a large number of samples from children and adults using Phage-Immunoprecipitation Sequencing (PhIP-Seq). We estimate the seroprevalence for endemic HCoVs to range from ~4% to ~27% depending on species and cohort. Most importantly, we identified a large number of novel linear B cell epitopes of HCoV proteins and demonstrate that antibody repertoires against endemic HCoVs are qualitatively different in children in comparison to the general adult population and healthy adult blood bank donors. We show that anti-HCoV IgG specificities more frequently found among children target functionally important and structurally conserved regions of the HCoV spike and nucleocapsid proteins and some antibody specificities are broadly cross-reactive with peptides of epidemic human and non-human coronavirus isolates. Our findings shed light on the humoral immune responses to natural infection with endemic HCoVs and may have important implications for understanding of the highly variable clinical outcomes of human coronavirus infections, for the development of prophylactic or therapeutic monoclonal antibodies and vaccine design.

Introduction

Four endemic human coronaviruses (HCoVs) are commonly associated with respiratory illness in humans, namely HCoV-229E, -NL63, -OC43, and -HKU1 [1-4]. Clinical outcomes of acute infection with these HCoVs range from mild upper respiratory tract infections in most cases, to viral bronchiolitis and pneumonia in some patients, the latter requiring hospitalization [5]. The ratio of more severe versus mild outcomes of acute infection with endemic HCoVs is largely comparable to that of other "common cold" viruses, such as human respiratory syncytial virus (HRSV), human rhinoviruses (HRVs), human adenoviruses (HAdVs), and human parainfluenza viruses (HPIVs), albeit with differences in seasonality and prevalence of the viruses depending on the species [5-7]. In addition to the four endemic HCoV, three epidemic coronaviruses (CoVs) have emerged in humans over the last two decades, including Severe Acute Respiratory Syndrome-associated CoV (SARS-CoV) [8], Middle East Respiratory Syndrome-associated coronavirus (MERS-CoV) [9] and SARS-CoV-2 [10], the etiological agent of Coronavirus Disease 2019 (COVID-19) which has now reached pandemic proportions [11]. Similar to endemic HCoVs, infection of humans with epidemic CoVs is associated with a wide range of outcomes but leads more frequently to severe clinical manifestations such as acute respiratory distress syndrome (ARDS) [12-14]. Phylogenetic analyses suggest that, similar to these epidemic CoVs, all endemic HCoVs are of zoonotic origin and their possible ancestors share similar natural animal reservoirs and intermediate hosts [6]. HCoV-229E may have been transferred from dromedary camels, similar to MERS-CoV, while HCoV-OC43 is thought to have emerged more recently from ancestors in domestic animals such as cattle or swine in the context of a pandemic at the end of the 19th century [6, 15]. It is therefore plausible that SARS-CoV-2 will have transmission dynamics during the post-pandemic period similar to the endemic HCoVs [16].

The wide variability in transmissibility and clinical manifestations of infections by endemic and epidemic CoVs among humans remains poorly understood. On the population level, the case fatality rate is highest for MERS (~36%) and several risk factors are associated with progression to ARDS in MERS, SARS and COVID-19 cases, including old age, diabetes mellitus, hypertension, cancer, renal and lung disease, and co-infections [12, 17]. Nonetheless, even MERS-CoV infection among humans can run a completely asymptomatic course in some

71 cases, particularly among children [18-20]. There is evidence that children are generally less
72 susceptible to infection with epidemic CoVs and once infected, they are less likely to
73 experience severe outcomes as compared to adults, although this important association and
74 the underlying reasons remain to be fully established [12, 19, 21, 22]. Importantly, it remains
75 unclear to which extend pre-existing immunity from past infections with endemic HCoV may
76 provide some degree of cross-protection and affect clinical outcomes of infection with the
77 epidemic SARS-CoV-2 or MERS-CoV and our overall understanding of the immunity induced
78 by natural infection with endemic HCoVs remains very limited. Serological studies have shown
79 some degree of cross-reactive antibodies in patients with past CoV infections but many of
80 these studies were limited in sample size and often only focused on a selection of viral
81 antigens [23-26]. Depending on their binding affinity and specificities, such cross-reactive
82 antibodies could either have no effect on clinical outcomes, may provide some degree
83 protection from severe disease or on the other hand, may lead to antibody-dependent
84 enhancement of disease—the latter can be a major obstacle in vaccine development [27].
85 Interestingly, two recent studies from independent groups have shown that a considerable
86 proportion of individuals without a history of SARS-CoV-2 infection have SARS-CoV-2-reactive
87 T cells, which suggests that cross-reactive T cell subsets originating from past infections by
88 endemic HCoVs may play a role in the clinical course of infection with the phylogenetically
89 related epidemic CoVs [28, 29]. A systematic assessment to elucidate the immunodominant
90 B cell antigen determinants of endemic HCoVs and the inter-individual and age-dependent
91 variation in the cellular and humoral immune responses among humans, as well as the degree
92 to which these immunodominant B cell targets represent cross-reactive antigenic sites, has
93 not been done. Particularly the Orf 1ab polyproteins have been poorly assessed in B cell
94 epitope screening studies and most of the existing knowledge about immunodominant
95 epitopes of CoVs originates from studies of SARS patients [30]. Here we report a
96 comprehensive retrospective analysis of CoV-specific antibody specificities in a large number
97 of samples from children and adults using Phage-Immunoprecipitation Sequencing (PhIP-
98 Seq), a technology that allows comprehensive profiling of serum or plasma antibodies using
99 oligonucleotide-encoded peptidomes [31, 32]. We demonstrate that antibody repertoires
100 against endemic HCoVs are qualitatively different in children in comparison to the general
101 adult population and healthy adult blood bank donors. We show that the more frequently
102 found anti-HCoV IgG specificities in children target functionally important and structurally

103 conserved regions of the HCoV spike and nucleocapsid proteins and some antibody
104 specificities are broadly cross-reactive with peptides of epidemic human and non-human
105 coronavirus isolates.

106 **Results**

107 To gain a deeper insight into human antibody responses to endemic HCoVs, we performed
108 PhIP-Seq [31-33] on previously collected serum or plasma samples obtained from a large
109 number of human subjects from three different cohorts. These included i) healthy male adult
110 blood donors (ABD) with diverse ethnic background and nationality (Supplementary Figure
111 S1A); ii) adult male and female participants of a national cohort study—the Qatar Biobank
112 (QBB) [34]—representing the general population (Supplementary Figure S1B); and iii)
113 pediatric outpatients and inpatients who were tested for metabolic conditions unrelated to
114 infection, chronic disease or cancer (Methods and Supplementary Figure S1C). All samples
115 were collected prior to the current COVID-19 outbreak (Methods). In brief, PhIP-Seq allowed
116 us to obtain comprehensive antiviral antibody repertoires across individuals in our three
117 human cohorts using phage display of oligonucleotide-encoded peptidomes, followed by
118 immunoprecipitation and massive parallel sequencing [31, 32]. The VirScan phage library
119 used for PhIP-Seq in the present study comprised peptides derived from viral proteins—each
120 represented by peptide tiles of up to 56 amino acids in length that overlap by 28 amino
121 acids—which collectively encompass the proteomes from a large number of viral species,
122 including HCoV-229E, -NL63, -HKU1 and -OC43 [31, 32]. Proteins of endemic HCoVs which
123 were represented in the VirScan phage library included the Orf1ab replicase polyprotein
124 (pp1ab), the spike glycoprotein (S), the matrix glycoprotein (M), the nucleocapsid protein (N)
125 and gene products of the species- and strain-specific open reading frames (ORFs) encoded in
126 the 3' region of the viral genomes (Supplementary Table S1). Of note, we utilized an expanded
127 version of the VirScan phage library [35, 36], which also encompassed peptides from a
128 number of proteins of human epidemic and non-human CoV isolates, including MERS-CoV,
129 SARS-CoV, as well as bat, bovine, porcine and feline isolates belonging to the alpha- and
130 betacoronavirus genera, albeit with varying coverage of the viral peptidomes owing to the
131 limitation in available sequence data for the latter isolates in UniProt (Supplementary Table
132 S1).
133
134 We were able to obtain antibody repertoires for a total of 1399 individuals from the human
135 cohorts described above (Supplementary Table S2). Using stringent filter criteria (Methods),
136 we identified a total of 417 out of 2498 peptides and potential antigens from endemic HCoVs

137 with our screen that were significantly enriched in at least three of all 1399 analyzed
138 individuals. A total of 103 peptides from endemic HCoV were enriched in ≥ 1% of the samples
139 and therefore considered to contain potentially immunodominant regions (Supplementary
140 Table S3). Only 33 of the 417 peptides enriched in at least three samples shared linear
141 sequence homology with epitopes that have previously been reported [37] (Supplementary
142 Figure S2). To estimate number of newly identified linear B cell epitopes, we assigned each
143 CoV-derived peptide to clusters of peptides that share ≥ 7 amino acids linear sequence
144 identity—the estimated size of a linear B cell epitope (Methods). The enriched peptides could
145 be assigned to 149 clusters for which at least 2 peptides share linear sequence identity of ≥ 7
146 amino acids (Supplementary Tables S3 and S4). Only 13 clusters also shared ≥ 7 amino acids
147 linear sequence identity with known linear B cell epitopes. Consequently, we have identified
148 a minimum of 136 new linear epitopes, including 25 new immunodominant linear B cell
149 epitopes [i.e. B cell epitopes targeted in at least ≥ 1% of all individuals and not already
150 reported in IEBD (www.iebd.org) [37] (Supplementary Tables S3)].
151
152 Next we assessed the seroprevalence of HCoV-229E, -NL63, -HKU1 and -OC43 in the three
153 cohorts separately. To do so, we imputed species score values as described earlier [32, 35,
154 38] by counting the significantly enriched peptides for a given HCoV species that share less
155 than 7 amino acids linear sequence identity. We considered an individual seropositive for any
156 of the endemic HCoVs if the number of non-homologous peptides enriched in a given sample
157 met our previously established species-specific cut-off value (Methods). Seroprevalence for
158 endemic HCoVs ranged from ~4% to ~27%, depending on the species and cohort (Figure 1A).
159 Interestingly, we found a marginal but significant negative association between age and
160 seroprevalence of HCoV-OC43 (β = -0.175) and -NL63 (β = -0.315) (Figure 1B and 1D), as well
161 as a marginal positive association between male gender and seroprevalence for any of the
162 endemic HCoVs ($\beta \leq 0.2$) (Figure 1C and 1D). The species score values (i.e. the antibody
163 repertoire breadth for each HCoV species) did not differ substantially between seropositive
164 individuals of our 3 cohorts (Supplementary Figure S3). However, principal component
165 analysis revealed considerable qualitative differences in the antibody repertoires between
166 our cohorts and in particular between pediatric and adult subjects (Figure 2A). For
167 comparison, we also performed the same analysis on enriched peptides that were derived

168 from HRSV. As expected, seroprevalence for the latter virus was considerably higher (QBB:
169 68.4 %; ABD: 77.5 %; PED: 90.4 %) (Supplementary Figure S4A) but in contrary to antiviral
170 antibody responses to endemic HCoVs, we did not find any differences in variance of the
171 HRSV-specific antibody responses when comparing age groups and cohorts (Supplementary
172 Figure S4B). To determine the antibody specificities responsible for most of the variance in
173 the antiviral response to endemic HCoVs between adults and children (i.e. to identify those
174 peptides that were significantly more or less frequently enriched when comparing adult and
175 pediatric donors), we applied Fisher's exact test and computed odds ratios for each of the
176 significantly enriched peptides. We found that antibody specificities in samples of pediatric
177 study subjects predominantly targeted different antigenic regions in the S protein (mean log
178 odds ratio = 3.35 $\pm$ 2.12), the N protein (mean log odds ratio = 2.21 $\pm$ 1.41) and diverse
179 antigenic sites in pp1ab, whereas peptides encoding a single linear B cell epitope of pp1ab
180 (cluster 22) appeared to be the predominant target of IgG antibodies among adult donors
181 (mean log odds ratio = -4.7 $\pm$ 1.16) (Figure 2B and Table 1).
182
183 Intriguingly, multiple sequence alignments of frequently enriched peptides with the full-
184 length proteins of various CoVs revealed that antibody specificities predominantly found in
185 pediatric study subjects target immunodominant epitopes that encode functionally important
186 and highly conserved regions of the structural proteins. These included regions in the S_1
187 subunit of the S protein which are important for receptor binding (Figure 3A-C and 3F) [39-
188 42], as well as the regions resembling the proteolytic cleavage sites and fusion peptide of the
189 S_2 subunit (Figures 3A, 3B, 3D-F). Of note, the immunodominant region spanning the furin-
190 like S2' cleavage site in the S_2 subunit resembles one of the most conserved regions of the S
191 protein, both in amino acid sequence (R↓SA[I/L]ED[I/L]LF) and in protein structure, as it forms
192 an accessible alpha-helix immediately upstream of the fusion peptide (Figure 3F and
193 Supplementary Figure S6) [43]. Moreover, we identified potential antibody binding sites in
194 the N-terminal RNA binding domain, serine-rich region and in the C-terminal dimerization
195 domain of the N protein (Figures 4A and 4B). Although the predicted antibody binding sites
196 in the N-terminal RNA-binding domain and the C-terminal dimerization domain of the N
197 protein appeared to be less conserved between different species in the primary amino acid
198 sequence (Figure 4C and 4D), both domains are structurally conserved in the regions that we
199 found to be immunodominant (Figure 4E and 4F). We also found that antibodies in children

200 targeted more frequently the C-terminal domain of the M protein (Supplementary Figure S5C
201 and Table 1) and the small accessory Orf8 protein (also known as N2) of HCoV-HKU1 (Table
202 1). Although Orf8 and N share the same coding sequence in the viral RNA genome, the reading
203 frame is different and the amino acid sequences not homologous. On the contrary, antibody
204 specificities predominantly found in adults primarily targeted a region of the pp1ab that is
205 specific to HCoV-HKU1 and contains an acidic tandem repeat (ATR) of N-[DN]-D-E-D-V-V-T-G-
206 D which is located upstream of the papain-like protease 1 domain (Supplementary Figure
207 S5D).
208
209 Given the high degree of sequence conservation among some of the immunodominant
210 regions in proteins of endemic HCoVs we have identified, we also explored to which extent
211 antibody specificities targeting such conserved regions in structural proteins of endemic
212 HCoVs may cross-react with peptides from epidemic CoVs and non-human CoV isolates. For
213 this purpose, we assessed the enrichment of peptides derived from SARS-CoV, MERS-CoV and
214 bovine, porcine, bat and feline isolates (Supplementary Table S1) applying the same approach
215 and stringent filter criteria as we have described above for peptides of endemic HCoVs.
216 Indeed, we identified several S protein-derived and N protein-derived peptides from epidemic
217 CoVs or non-human isolates that were significantly enriched in our PhIP-Seq assay, which
218 share sequence similarity with peptides from HCoVs (Figure 5). As expected based on the
219 results from multiple sequence alignments described above, antibody specificities targeting
220 the highly conserved amino acid motive (RSA[I/L]ED[I/L]LF) spanning the furin-like S2'
221 cleavage site of the S protein also appeared to exhibit the broadest cross-reactivity as we
222 identified several orthologous peptides from MERS-CoV, SARS-CoV as well as from non-
223 human isolates enriched in our assay (Figure 5A, Region 3). Cross-reactivity of antibodies
224 targeting other functionally important but less conserved regions also appeared to be more
225 restricted (Figure 5A, Region 1 and 2). Similarly, antibody specificities targeting the N protein
226 showed considerable cross-reactivity with peptides from MERS-CoV, SARS-CoV and non-
227 human CoV isolates. However, the latter cross-reactive antibodies mainly targeted regions
228 rich in serine and arginine residues, with low-complexity sequences and very limited
229 structural conservation, particularly an intrinsically disordered region (IDR) at the N terminus
230 of the N protein (Figure 5B and Supplementary Figure 7) that lacks a tertiary structure [44].
231 We also detected cross-reactive antibodies targeting the Serine-rich motif and linker region

232 of the N protein; however, cross-reactivity was largely restricted to peptides derived from
233 non-human CoV isolates of domestic animals (Figure 5B) which are more closely related to
234 HCoV-OC43 [6].

235 **Discussion**

236 Our comprehensive screen for antiviral antibody repertoires across individuals in our three
237 human cohorts revealed a large number of peptides with novel linear epitopes in several
238 proteins of endemic HCoVs. This is not surprising given that epidemic CoVs, and in particular
239 SARS-CoV, have been the primary focus of previous immunological and epitope screening
240 studies [30, 37]. Information about the targets of immune responses to CoVs across different
241 species provides a valuable resource for the prediction of candidate targets of newly
242 emerging CoVs, as it has recently been done by Grifoni *et al.* [30]. The authors were able to
243 identify *a priori* several specific regions of the S, M and N proteins of SARS-CoV-2 S based on
244 sequence homology to the SARS-CoV virus which are orthologous to several of the
245 immunodominant regions we have identified among endemic HCoVs. We found antibody
246 responses to be targeting the structural S, N, M and Orf8 proteins but also the nonstructural
247 pp1ab polyprotein of HCoVs, the latter resembling the precursor for the large viral replicase
248 complex [45]. Interestingly, in another independent study, Grifoni *et al.* [29] recently reported
249 similarly broad T cell responses in COVID-19 patients by employing a comprehensive and
250 analogous screen for T cell epitopes of SARS-CoV-2 proteins using peptide ''megapools'' in
251 combination with *ex vivo* T cell assays.
252
253 Using stringent cut-off values, we have also estimated the seroprevalence of endemic HCoVs
254 to range from ~4% to ~27% depending on the species and cohort (Figure 1A-C). This is largely
255 consistent with earlier studies [46-50] but serological studies on larger sample sizes and in
256 the general population are sparse. Gaunt *et. al.* retrospectively assessed 11,661 respiratory
257 samples collected from 7,383 patients at hospital and primary care settings in Southeast
258 Scotland for routine respiratory virus screening by PCR and detected active infection with
259 endemic HCoVs in 0.3 to 0.85% of samples in all age groups, as compared to a detection rate
260 of 11.1% and 6.7% for HRSV and HAdVs, respectively [5]. Since molecular diagnostic testing is
261 typically done in critically ill patients and patients with comorbidities or immune suppression
262 [51], it is not surprising that the majority of cases assessed in the study by Gaunt *et. al.* [5]
263 had lower respiratory tract infections. Another study of hospitalized Norwegian children with
264 respiratory tract infection found that 8.2% and 4.5% of the children tested PCR-positive for
265 CoV-OC43 and HCoV-NL63, respectively; only HRSV and HRVs were detected more frequently

266 [7]. Since the individuals assessed in the present study largely reflect the general adult
267 population or children with conditions unrelated to infection, our seroprevalence results
268 most accurately reflect the seroprevalence of HCoVs in the State of Qatar and is likely
269 attributable to mostly mild infections. Interestingly, age appeared to be negatively associated
270 with seroprevalence in our study, suggesting that the either duration of immunity in response
271 to natural infection with endemic HCoVs, or rates of re-infection, reduce with increasing in
272 age. Dynamics of humoral and cellular immune responses against CoVs are poorly understood
273 and correlates of protection remain controversial [52]. Longitudinal studies have mainly been
274 focused on the duration of humoral immunity in MERS and SARS patients and suggest that
275 antibody responses wane relatively quickly to minimal detectable levels over a period of not
276 more than 2–3 years [53]. On the other hand, most acute virus infections induce some level
277 of protective and long-term immunity, albeit through a variety of mechanisms that are not
278 necessarily the same for each pathogen and may even differ between hosts due to a variety
279 of factors, including simultaneous viral coinfection [54, 55]. We also found a marginal but
280 significant positive association between seroprevalence of endemic HCoVs and male gender,
281 which is consistent with the earlier report by Gaunt *et al.* [5], who found significantly more
282 male than female laboratory-confirmed cases in a large number of diagnostic respiratory
283 samples described above. The underlying molecular reasons remain unclear.
284
285 A surprising and unexpected finding of our study is that circulating IgG antibodies in children
286 versus adults appear to be differentially targeting structural and non-structural proteins of
287 HCoVs (Figure 2). Whereas antibody specificities in samples of pediatric subjects
288 predominantly targeted structural proteins such as the N and S proteins, in adult donors, a
289 region of the non-structural polyprotein pp1ab containing a tandem repeat of N-[DN]-D-E-D-
290 V-V-T-G-DA in HCoV-HKU1 appeared to be the predominant target of IgG antibodies. The
291 latter polyprotein is post-translationally processed into up to 16 subunits that form a large
292 viral replicase complex responsible for both continuous and discontinuous RNA synthesis [45].
293 This qualitative difference in the antibody repertoires of adult versus pediatric study subjects
294 appeared to be a specific characteristic of natural HCoV infection, as we did not find any
295 variance in the antibody repertoires specific to HRSV when comparing our cohorts and
296 different age groups (Supplementary Figure S3). Several of the immunodominant regions we
297 have identified experimentally in the structural proteins of endemic HCoV are orthologous to

298 the regions identified by Grifoni *et al.* [30] to be dominant SARS-CoV B cell epitope regions,
299 which they also predicted to be targets for immune responses to SARS-CoV-2 based on
300 sequence homology. Importantly, antigenic regions that we found to be immunodominant in
301 our study (i.e. enriched in $\geq$ 1% of all samples) and those corresponding to peptides for which
302 enrichment was strongly (odds ratio $\geq$ 2) and significantly (P-value $\leq$ 0.005, Fisher's exact test)
303 associated with pediatric subjects mapped to functionally important regions of the structural
304 CoV proteins. These included regions for receptor binding and the proteolytic cleavage sites
305 of the S protein as well as the N-terminal RNA-binding and C-terminal dimerization domains
306 of the N protein, which have been shown to be critical for virus attachment and entry, cell-
307 to-cell fusion and virus replication, respectively [44, 56-60]. Of note, CoVs differ considerably
308 in the regions of the S_1 subunit of the S protein responsible for receptor binding, utilize
309 different domains and host cell receptors, and consequently differ in their tissue tropism [49].
310 Whereas HCoV-OC43 and -HKU1 use 9-O-acetyl-sialic acid (9-O-Ac-Sia) as a receptor, SARS-
311 CoV, SARS-CoV-2 and several SARS-related coronaviruses interact directly with angiotensin-
312 converting enzyme 2 (ACE2) [40-43]. Nevertheless, key structural features and residues
313 involved in receptor binding are highly conserved [39, 42] and the immunodominant RBD
314 region we have identified in the present study is consistent with the findings by Grifoni *et al.*
315 [30]. Similarly, although the predicted antibody bindings sites in the N protein appeared to be
316 less conserved in primary amino acid sequence, the secondary structure appeared to be
317 conserved, suggesting that some of these target regions may also resemble discontinuous B
318 cell epitopes. On the contrary, we identified an immunodominant and highly conserved linear
319 epitope spanning the furin-like S2' cleavage site of the S protein (R↓SA[I/L]ED[I/L]LF) that is
320 located immediately upstream of the fusion peptide. Proteolytic cleave at this site triggers
321 membrane fusion via profound conformational changes that are irreversible [43]. The high
322 degree of sequence and conformational conservation of the alpha-helical region immediately
323 adjacent to the S2' cleavage site [43] likely explains why antibodies targeting this region also
324 cross-reacted with orthologous peptides of related CoVs, including that of MERS-CoV, SARS-
325 CoV and non-human isolates, since previous exposure of the individuals assessed in the
326 present study to these epidemic CoVs is highly unlikely. Although the immunodominant
327 regions in the receptor binding domain and S1/S2 cleavage site are less conserved in
328 sequence and likely also in structure, we also observed some degree of cross-reactivity of
329 antibodies targeting these regions, albeit to a much lesser degree than antibodies targeting

330 the fusion peptide region. In agreement with our findings, Al Kahlout *et al.* [61] found that
331 only 10 individuals (0.21%) of a larger cohort of 4719 healthy blood donors in Qatar tested
332 positive in a semi-quantitative ELISA for IgG antibodies against a recombinant protein
333 resembling the S_1 subunit (rS_1) of MERS-CoV. However, none of these samples tested positive
334 using a whole-virus anti-MERS-CoV IgG ELISA, in contrary to samples from three PCR-
335 confirmed MERS cases and three individuals with direct contact to these cases [61].
336 Nevertheless, all MERS-CoV rS_1-reactive blood donor samples assessed in that study also
337 showed considerable reactivity by ELISA for IgG antibodies specific to endemic HCoV-HKU1, -
338 OC43, -229E, and -NL63 [61]. It is tempting to speculate that natural infection with endemic
339 HCoV may provide some degree of cross-protection and may therefore affect health
340 outcomes in individuals infected by epidemic CoVs, such as SARS-CoV-2 or MERS-CoV. In
341 agreement with this hypothesis, Gifroni *et al.* [29] and Brown *et al.* [28] have independently
342 reported SARS-CoV-2-reactive $CD4^+$ T cells in ~37% to 60% of individuals who were not
343 exposed to SARS-CoV-2, suggesting that cross-reactive T cell recognition may play role in the
344 immune responses of COVID-19 patients. Importantly, S-specific $CD4^+$ T cell responses in
345 COVID-19 patients were highly correlated with the magnitude of antibody responses against
346 the receptor binding domain of the S protein [29], highlighting the important role of T cell-
347 dependent B cell responses in the immunopathogenesis of COVID-19. Of note, a study of
348 convalescent COVID-19 patients identified two linear epitopes in regions of the S protein
349 against which antibodies exhibit potent neutralizing activity, namely a less conserved region
350 (TESNKKFLPFQQFGRDIA) adjacent to the ACE2 binding domain and a region
351 (PSKPSKRSFIELDLLFNKV) encoding the S2' proteolytic cleavage site [62]. Both regions are
352 orthologous to the immunodominant regions we have found in our study to be targeted by
353 antibodies in response to endemic HCoV infection and the latter resembles one of the most
354 conserved epitopes (RSA[I/L]ED[I/L]LF) we have identified in our study. This high degree of
355 sequence and structural conservation at the latter site reflects common features shared by
356 all enveloped viruses, which enter their target cells by inducing the fusion of the viral
357 envelope with the host cell membrane, a process that involves profound conformational
358 changes to overcome the repulsive force between the two membrane bilayers [41]. It is
359 therefore plausible that such cross-reactive antibodies may also play a protective role against
360 epidemic and endemic HCoV infection, partially among children which appear to target such
361 functionally important B cell epitopes regions of the structural CoV proteins more frequently

362 than adults and at the same time, are also less likely to experience severe disease outcomes
363 [18-20, 22]. Broadly cross-reactive antibody responses are also known for other enveloped
364 RNA viruses, which may positively or negatively affect subsequent infection or vaccination.
365 Flaviviruses for example are antigenically related and broadly flavivirus cross-reactive
366 antibodies from previous yellow fever vaccination has been shown to impair and modulate
367 immune responses to tick-borne encephalitis vaccination [63]. Similarly, immune history has
368 been shown to profoundly affect protective B cell responses to influenza [64].
369
370 In summary, we have shown that natural infection with endemic HCoV elicits humoral
371 responses with district antibody repertoires in adults and children. To which degree pre-
372 existing immunity from past infections with endemic HCoVs also provides cross-protection to
373 the more distantly related epidemic CoVs needs further investigation. Our findings may have
374 important implications for the development of prophylactic or therapeutic monoclonal
375 antibodies and vaccine design in the context of COVID-19 [30, 65].

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 [34]; and iii) 231 pediatric subjects with Qatari nationality who were admitted to or visiting Sidra Medicine as follows. Plasma samples from healthy blood bank donors were deidentified leftovers collected from 2012 to 2016. For the purpose of this study, we selected most male Qatari nationals 19 to 66 years of age (Supplementary Table S1) from a larger blood donor cohort including 5983 individuals and then randomly selected age-matched male donors with other nationalities (Supplementary Figure S1). Samples from female blood bank donors were excluded because they were largely underrepresented among 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 [34]. Plasma samples from pediatric patients were selected from leftovers of samples processed in the clinical chemistry labs of Sidra Medicine, a tertiary 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 submitted for basic metabolic panel (BMP) and comprehensive metabolic panel (CMP) testing 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, no centile data and samples from patients with centile <5% (underweight) or >85% (overweight) were excluded. However, we included obese patients in our analysis, since a considerable proportion of Qatari nationals are overweight. The human subject research described here had been approved by the institutional research ethics boards of Sidra Medicine and Qatar Biobank.

407 **Phage Immunoprecipitation-Sequencing (PhIP-Seq).** Large scale serological profiling of the
408 antiviral IgG repertoires in the individual serum or plasma samples was performed as
409 described by Xu el al. [32]. Each serum or plasma sample was tested in duplicate and samples
410 were analyzed in batches with up to 96 samples each batch. Only those samples were
411 considered for downstream analysis that satisfied a minimum read count of $1*10^6$ as well as
412 a Pearson correlation coefficient of $\geq$ 0.7 in the two technical repeats. Data from thirty
413 individuals of the ABD cohort and two individuals of the QBB cohort were excluded from the
414 downstream analysis due to insufficient sequencing read depth, low sequencing data quality
415 or because one of two technical replicates had failed (not shown).
416
417 **Peptide enrichment analysis.** To filter for enriched peptides, we first imputed −log10(P)
418 values as described previously [32, 35, 38] by fitting a zero-inflated generalized Poisson model
419 to the distribution of output counts and regressed the parameters for each peptide sequence
420 based on the input read count. We considered a peptide enriched if it passed a reproducibility
421 threshold of 2.3 [−log10(P)] in two technical sample replicates. To remove sporadic hits, we
422 then filtered for antibody specificities to CoV peptides that were found to be enriched in at
423 least three of all 1399 subjects assayed and analyzed in this study. We computed species-
424 specific significance cut-off values to estimate minimum number of enriched, non-
425 homologous peptides required in order to consider a sample as seropositive using a
426 generalized linear model and in-house serological (ELISA) data from pooled samples that were
427 tested positive for various viruses. We then computed virus score values as described by Xu
428 et al [32] by counting enriched, non-homologous peptides for a given species and then
429 adjusted these score values by dividing them with the estimated score cutoff. For the purpose
430 of this study and under consideration of the seroprevalence of endemic HCoVs in our three
431 cohorts, we considered a peptide to be immunodominant if it was enriched in $\geq$ 1% of all 1399
432 subjects assayed and analyzed in this study.
433
434 **Association studies and differential enrichment analysis.** We applied a generalized linear
435 model to test for associations between the HCoV species-specific adjusted score values,
436 gender and age. We considered an association to be significant if the P-value was $\leq$ 0.001. We
437 examined the frequency distribution of enriched peptides among samples of the different age
438 groups (PED versus ABD + QBB) by estimating odds ratios (OR) and associated P-values using

439 Fisher's exact test. Peptides that satisfied both significance (P-value $\leq$ 0.005) and magnitude
440 criteria ($|\log(OR)| \geq \log(2)$) were considered to be differentially enriched. Positive log(OR)
441 values indicated more frequent peptide enrichment among pediatric study subjects, whereas
442 negative log(OR) values indicated more frequent peptide enrichment among adult subjects.
443
444 **Clustering of peptides for shared linear B cell epitopes.** To estimate the minimum number
445 of linear B cell epitopes among the enriched peptides, we built a pairwise distance matrix that
446 captured the maximum size of linear sequence identity of amino acids ($d_{i,j}$) between all
447 enriched peptides. Groups of peptides that shared $\geq$ seven amino acid linear sequence
448 identity ($d_{i,j} \geq 7$) were assigned to a cluster. Peptides of a given cluster were considered to
449 share a linear B cell epitope (Supplementary Figure S7).
450
451 **Software.** Open source Python modules with in-house scripts were used to test for
452 associations (statsmodel v0.11), to filter differentially enrichment peptides and to perform
453 different statistical tests (sklearn v0.23, scipy v1.14.1). Multiple sequence alignments were
454 done using the MAFFT [66, 67] via EMBL-EBI's web services and Java Alignment Viewer
455 (Jalview) for visualization [68]. Residue-wise linear B cell epitopes were predicted using
456 BepiPred-2.0 [69]. 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 and SH planned and performed the experiments. GN, AS and MH contributed samples and data. 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. 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.

475 **References**

476

477 1. Hamre, D. and J.J. Procknow, *A New Virus Isolated from the Human Respiratory*
478 *Tract.* Proceedings of the Society for Experimental Biology and Medicine, 1966.
479 **121**(1): p. 190-193.

480 2. McIntosh, K., et al., *Recovery in tracheal organ cultures of novel viruses from patients*
481 *with respiratory disease.* Proceedings of the National Academy of Sciences, 1967.
482 **57**(4): p. 933-940.

483 3. van der Hoek, L., et al., *Identification of a new human coronavirus.* Nature Medicine,
484 2004. **10**(4): p. 368-373.

485 4. Woo, P.C.Y., et al., *Characterization and Complete Genome Sequence of a Novel*
486 *Coronavirus, Coronavirus HKU1, from Patients with Pneumonia.* Journal of Virology,
487 2005. **79**(2): p. 884-895.

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

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

493 7. Kristoffersen, A.W., et al., *Coronavirus causes lower respiratory tract infections less*
494 *frequently than RSV in hospitalized Norwegian children.* Pediatr Infect Dis J, 2011.
495 **30**(4): p. 279-83.

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

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

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

502 11. Dong, E., H. Du, and L. Gardner, *An interactive web-based dashboard to track COVID-*
503 *19 in real time.* The Lancet Infectious Diseases, 2020.

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

506 13. *Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) —*
507 *United States, February 12–March 16, 2020.* MMWR Morb Mortal Wkly Rep, 2020.
508 **69**: p. 343-346.

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

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

514 16. Kissler, S.M., et al., *Projecting the transmission dynamics of SARS-CoV-2 through the*
515 *postpandemic period.* Science, 2020.

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

518 18. Dong, Y., et al., *Epidemiology of COVID-19 Among Children in China.* Pediatrics, 2020:
519 p. e20200702.

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

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

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

529 22. Davies, N.G., et al., *Age-dependent effects in the transmission and control of COVID-*
530 *19 epidemics.* Nature Medicine, 2020.

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

533 24. Okba, N.M.A., et al., *Severe Acute Respiratory Syndrome Coronavirus 2-Specific*
534 *Antibody Responses in Coronavirus Disease 2019 Patients.* Emerg Infect Dis, 2020.
535 **26**(7).

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

539 26. Khan, S., et al., *Analysis of Serologic Cross-Reactivity Between Common Human*
540 *Coronaviruses and SARS-CoV-2 Using Coronavirus Antigen Microarray.* bioRxiv, 2020:
541 p. 2020.03.24.006544.

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

544 28. Braun, J., et al., *Presence of SARS-CoV-2 reactive T cells in COVID-19 patients and*
545 *healthy donors.* medRxiv, 2020: p. 2020.04.17.20061440.

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

548 30. Grifoni, A., et al., *A Sequence Homology and Bioinformatic Approach Can Predict*
549 *Candidate Targets for Immune Responses to SARS-CoV-2.* Cell Host Microbe, 2020.
550 **27**(4): p. 671-680 e2.

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

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

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

557 34. Al Kuwari, H., et al., *The Qatar Biobank: background and methods.* BMC Public
558 Health, 2015. **15**: p. 1208.

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

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

563 37. Vita, R., et al., *The Immune Epitope Database (IEDB): 2018 update.* Nucleic Acids Res,
564 2019. **47**(D1): p. D339-D343.

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

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

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

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

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

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

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

580 45. Ziebuhr, J., *The coronavirus replicase.* Curr Top Microbiol Immunol, 2005. **287**: p. 57-
581 94.

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

585 47. Kaye, H.S. and W.R. Dowdle, *Seroepidemiologic survey of coronavirus (strain 229E)*
586 *infections in a population of children.* Am J Epidemiol, 1975. **101**(3): p. 238-44.

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

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

591 50. van der Hoek, L., *Human coronaviruses: what do they cause?* Antivir Ther, 2007. **12**(4
592 Pt B): p. 651-8.

593 51. Charlton, C.L., et al., *Practical Guidance for Clinical Microbiology Laboratories:*
594 *Viruses Causing Acute Respiratory Tract Infections.* Clin Microbiol Rev, 2019. **32**(1).

595 52. Peeples, L., *News Feature: Avoiding pitfalls in the pursuit of a COVID-19 vaccine.*
596 Proceedings of the National Academy of Sciences, 2020. **117**(15): p. 8218-8221.

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

599 54. Sallusto, F., et al., *From vaccines to memory and back.* Immunity, 2010. **33**(4): p. 451-
600 63.

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

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

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

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

610 59. Follis, K.E., J. York, and J.H. Nunberg, *Furin cleavage of the SARS coronavirus spike*
611 *glycoprotein enhances cell-cell fusion but does not affect virion entry.* Virology, 2006.
612 **350**(2): p. 358-69.

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

615 61. Al Kahlout, R.A., et al., *Comparative Serological Study for the Prevalence of Anti-*
616 *MERS Coronavirus Antibodies in High- and Low-Risk Groups in Qatar.* J Immunol Res,
617 2019. **2019**: p. 1386740.

618 62. Poh, C.M., et al., *Potent neutralizing antibodies in the sera of convalescent COVID-19*
619 *patients are directed against conserved linear epitopes on the SARS-CoV-2 spike*
620 *protein.* bioRxiv, 2020: p. 2020.03.30.015461.

621 63. Bradt, V., et al., *Pre-existing yellow fever immunity impairs and modulates the*
622 *antibody response to tick-borne encephalitis vaccination.* npj Vaccines, 2019. **4**(1): p.
623 38.

624 64. Andrews, S.F., et al., *Immune history profoundly affects broadly protective B cell*
625 *responses to influenza.* Science Translational Medicine, 2015. **7**(316): p. 316ra192-
626 316ra192.

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

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

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

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

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

637
638

Figures

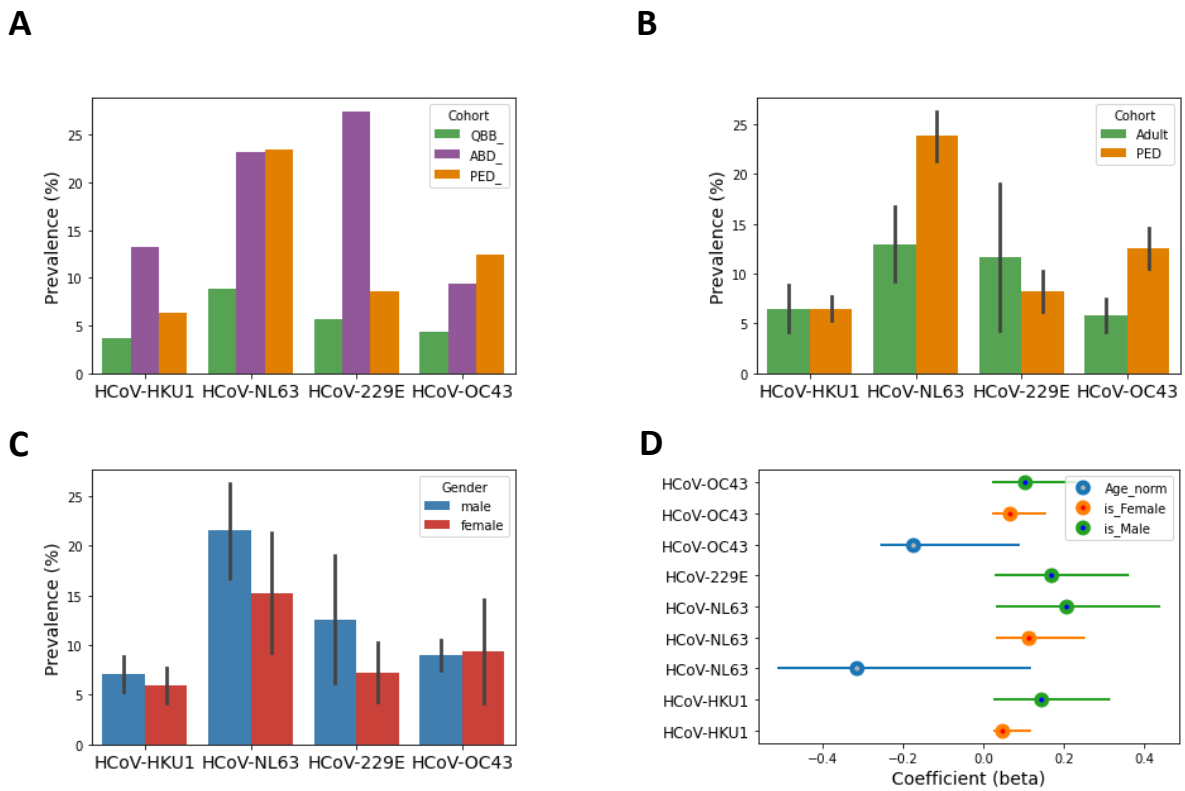


Figure 1. Seroprevalence of endemic HCoVs. A-C, Bar plots depict the seroprevalence of the four endemic HCoVs stratified by cohort (A), age group (B) and gender (C). Error bars in (B) depict gender-specific variation. Adults include adult blood bank donors and individuals of the Qatar Biobank cohort.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

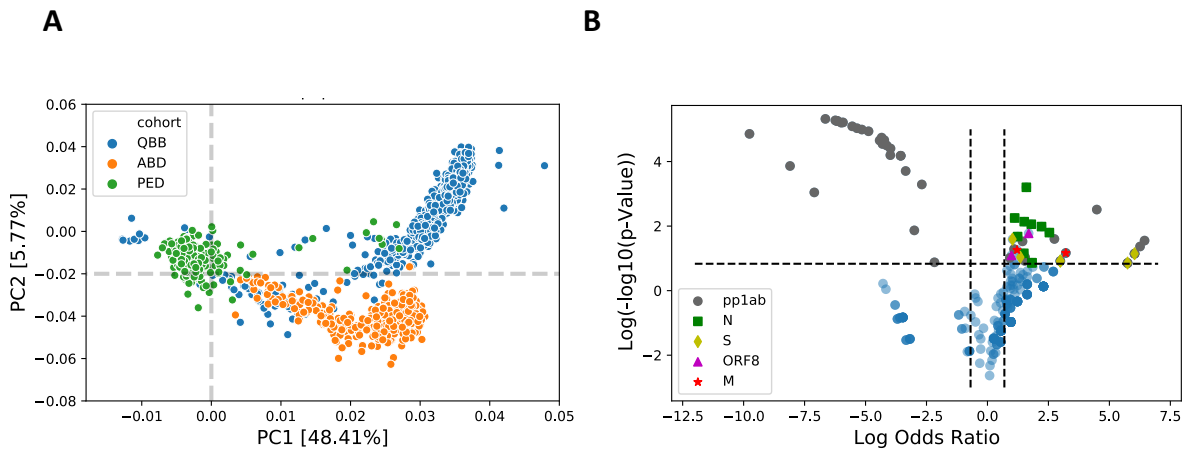
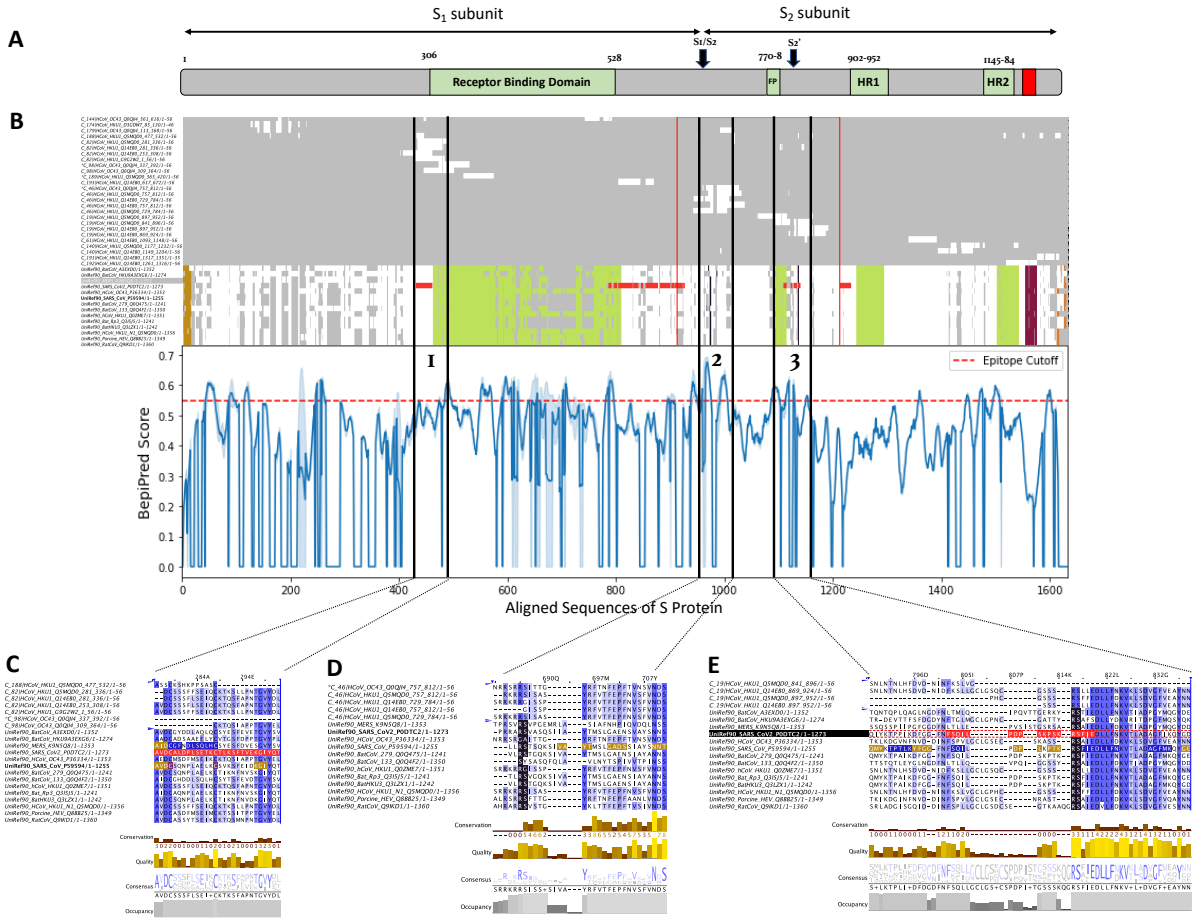


Figure 2. Qualitative differences in antibody repertoires between cohorts and age groups. A, Principal component analysis of 417 peptides from endemic HCoVs that were found to be enriched in at least 3 samples. QBB, Qatar Biobank cohort; ABD, adult (male) blood bank donors; PED, pediatric subjects. **B,**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



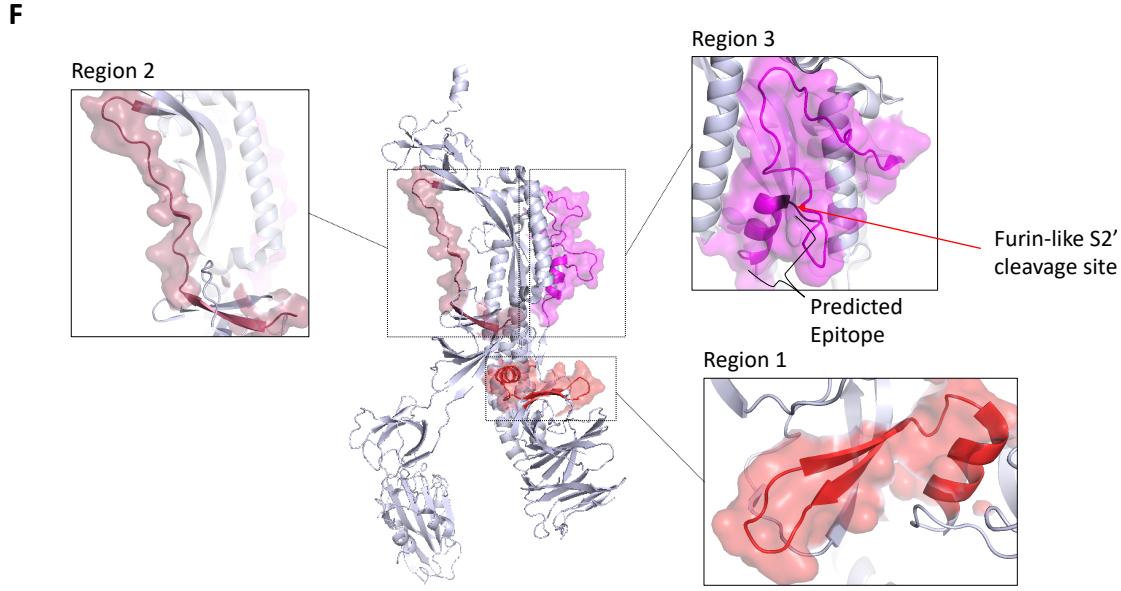


Figure 3. Antigenic regions and predicted antibody binding sites of the spike (S) protein. A, Schematic representation of the S protein of SARS-CoV (strain Tor2). Proteolytic cleavage sites of the S protein are marked with arrows. **B,** Overview of a multiple sequence alignment of

670 immunodominant peptides with the full-length protein sequences of various alpha- and beta-
671 CoVs (top). The mean score (blue line) and standard deviations (shaded) for the residue-wise
672 prediction of linear B cell epitopes of aligned endemic HCoVs (bottom). Row labels (left)
673 indicate the cluster and sequence identifier, start and end position of the peptide and length
674 of the aligned sequence. Peptides for which differential enrichment between children and
675 adults was statistically significant (P-value ≤ 0.05, Fisher's exact test) and odds ratios were $\geq$
676 2 are indicated with a "*". The protein domains and boundaries shown in the schematic (A)
677 have been marked in different colors using UniProt annotation and JalView features. Amino
678 acid sequences of previously predicted immunodominant linear SARS-CoV-2 B cell epitopes
679 [30] are highlighted in pink color. For the BepiPred score (bottom), a score cutoff of 0.55 has
680 been marked with a dashed red line to indicate regions that are predicted to be potential B
681 cell epitopes. **C-E,** Selected regions of the multiple sequence alignment encompassing regions
682 1, 2 and 3 as shown in (B). Proteolytic cleavage sites of the S protein are highlighted in black.
683 The full sequence alignment is shown in Supplementary Figure S5A. **F,** Monomer of the S
684 protein of SARS-CoV-2 in the prefusion conformation (PDB id: 6VXX, chain A) [43], with the
685 regions 1, 2 and 3 shown enlarged. FP, fusion peptide; HR1 and HR2, heptad repeat 1 and 2.
686

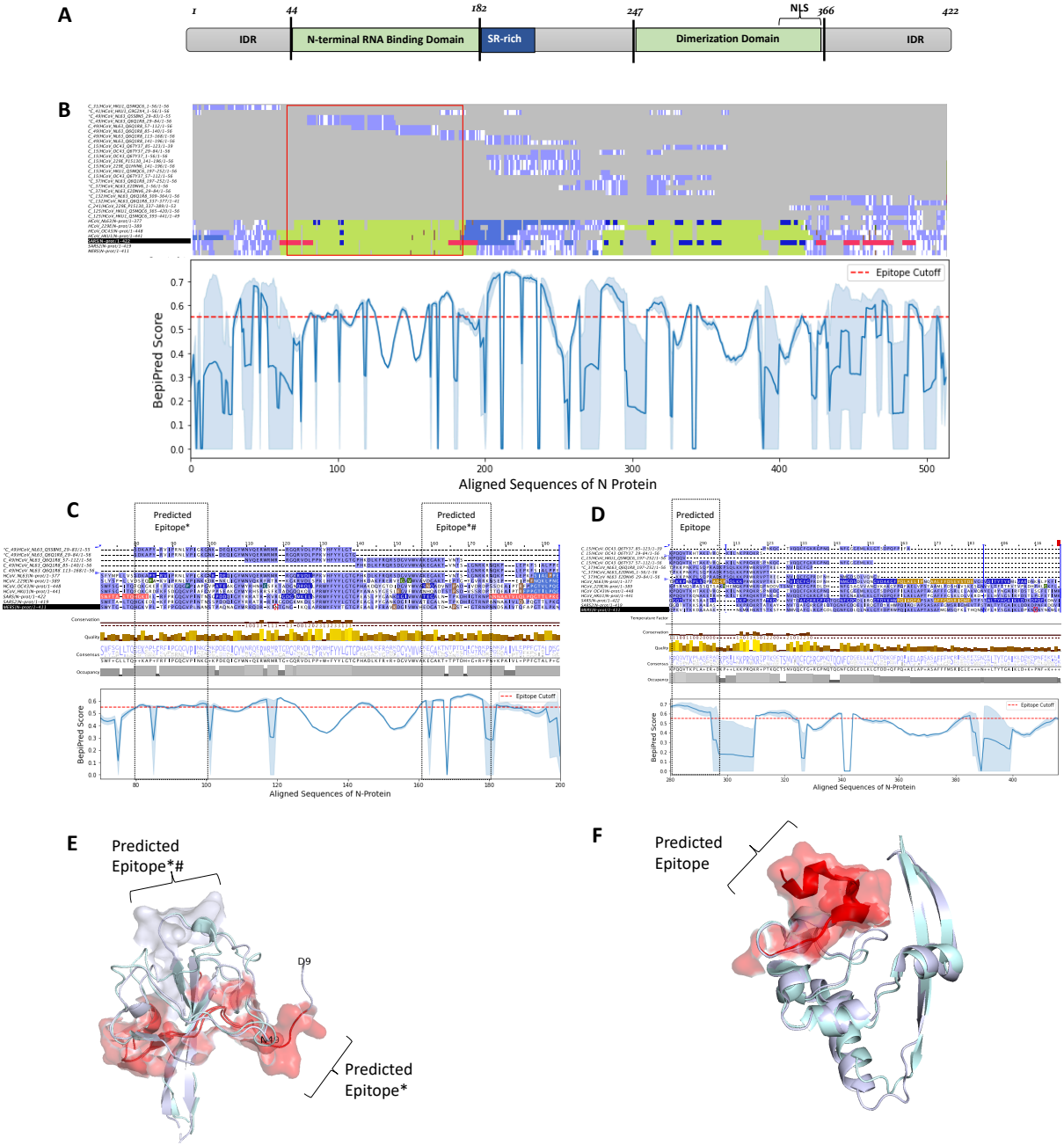
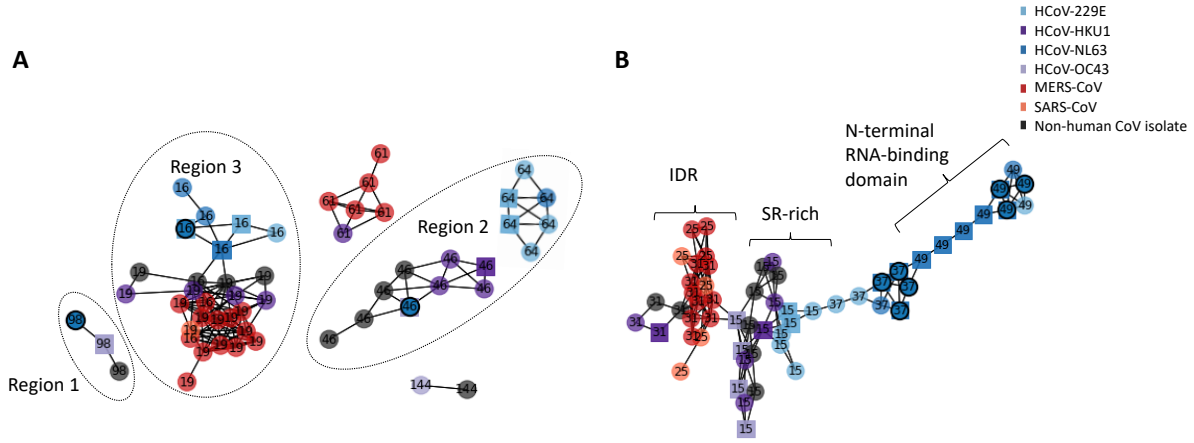


Figure 4. Antigenic regions and predicted antibody binding sites of the nucleocapsid (N) protein. A, Schematic representations of the N protein of SARS-CoV (strain Tor2). SR-rich, serine-rich; NLS, Predicted nuclear localization sequence; IDR, intrinsically disordered region. **B,** Overview of a multiple 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

697 of the aligned sequence. Peptides for which differential enrichment between children and
698 adults was statistically significant (P-value ≤ 0.005, Fisher's exact test) and odds ratios were
699 ≥ 2 are indicated with a "*". The protein domains of full-length reference sequences and
700 boundaries shown in the schematic (A) have been marked in green. Amino acid sequences of
701 previously predicted immunodominant SARS-CoV B cell epitopes [30] are highlighted in pink
702 color. For the BepiPred score (bottom), a score cutoff of 0.55 has been marked with a dashed
703 red line to indicate regions that are predicted to be potential B cell epitopes. **C, D,** Selected
704 regions of the multiple sequence alignment encompassing the N-terminal RNA-binding
705 domain (C) and C-terminal self-assembly domain (D). The full multiple sequence alignment is
706 shown in Supplementary Figure S5B. **E,** Super-imposed ribbon structure of the N-terminal
707 RNA-binding domain from HCoV-NL63 (PDB: 5N4k, chain A) and that from SARS-CoV-2 (PDB:
708 6M3M, chain A) (root-mean-square deviation [rmsd] = 0.7 Ångström). **F,** Super-imposed
709 ribbon structure of the C-terminal self assembly domain of proteins from HCoV- NL63 (pdbId:
710 5EPW, chain A) and that of SARS-CoV-2 (pdbID: 6WJI, chain A) (rmsd = 0.91 Ångström). *,
711 Predicted epitopes in peptides for which differential enrichment between children and adults
712 was statistically significant (P-value ≤ 0.005, Fisher's exact test) and odds ratios were ≥ 2; #,
713 Predicted immunodominant epitopes.

714



715

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

727

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

729 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**

730