bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

cases, particularly among children [18-20]. 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 as compared to adults, although this important association and the underlying reasons remain to be fully established [12, 19, 21, 22]. Importantly, it remains unclear to which extend pre-existing immunity from past infections with endemic HCoV may provide some degree of cross-protection and affect clinical outcomes of infection with the epidemic SARS-CoV-2 or MERS-CoV and 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 only focused on a selection of viral antigens [23-26]. Depending on their binding affinity and specificities, such cross-reactive antibodies could either have no effect on clinical outcomes, may provide some degree protection from severe disease or on the other hand, may lead to antibody-dependent enhancement of disease—the latter can be a major obstacle in vaccine development [27]. 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 [28, 29]. A systematic assessment to elucidate the immunodominant B cell antigen determinants of endemic HCoVs and the inter-individual and age-dependent variation in the cellular and humoral immune responses among humans, as well as the degree to which these immunodominant B cell targets represent cross-reactive antigenic sites, has not been done. Particularly the Orf 1ab polyproteins have been poorly assessed in B cell epitope screening studies and most of the existing knowledge about immunodominant epitopes of CoVs originates from studies of SARS patients [30]. 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), a technology that allows comprehensive profiling of serum or plasma antibodies using oligonucleotide-encoded peptidomes [31, 32]. We 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 the more frequently found anti-HCoV IgG specificities in children target functionally important and structurally

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Results

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

We were able to obtain antibody repertoires for a total of 1399 individuals from the human cohorts described above (Supplementary Table S2). Using stringent filter criteria (Methods), we identified a total of 417 out of 2498 peptides and potential antigens from endemic HCoVs

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

with our screen that were significantly enriched in at least three of all 1399 analyzed individuals. A total of 103 peptides from endemic HCoV were enriched in ≥ 1% of the samples and therefore considered to contain potentially immunodominant regions (Supplementary Table S3). Only 33 of the 417 peptides enriched in at least three samples shared linear sequence homology with epitopes that have previously been reported [37] (Supplementary Figure S2). To estimate number of newly identified linear B cell epitopes, we assigned each CoV-derived peptide to clusters of peptides that share ≥ 7 amino acids linear sequence identity—the estimated size of a linear B cell epitope (Methods). The enriched peptides could be assigned to 149 clusters for which at least 2 peptides share linear sequence identity of ≥ 7 amino acids (Supplementary Tables S3 and S4). Only 13 clusters also shared ≥ 7 amino acids linear sequence identity with known linear B cell epitopes. Consequently, we have identified a minimum of 136 new linear epitopes, including 25 new immunodominant linear B cell epitopes [i.e. B cell epitopes targeted in at least ≥ 1% of all individuals and not already reported in IEBD (www.iebd.org) [37] (Supplementary Tables S3)].

Next we assessed the seroprevalence of HCoV-229E, -NL63, -HKU1 and -OC43 in the three cohorts separately. To do so, we imputed species score values as described earlier [32, 35, 38] by counting the significantly enriched peptides for a given HCoV species that share less than 7 amino acids linear sequence identity. We considered an individual seropositive for any of the endemic HCoVs if the number of non-homologous peptides enriched in a given sample met our previously established species-specific cut-off value (Methods). Seroprevalence for endemic HCoVs ranged from ~4% to ~27%, depending on the species and cohort (Figure 1A). Interestingly, we found a marginal but significant negative association between age and seroprevalence of HCoV-OC43 (β = -0.175) and -NL63 (β = -0.315) (Figure 1B and 1D), as well as a marginal positive association between male gender and seroprevalence for any of the endemic HCoVs ($\beta \leq 0.2$) (Figure 1C and 1D). The species score values (i.e. the antibody repertoire breadth for each HCoV species) did not differ substantially between seropositive individuals of our 3 cohorts (Supplementary Figure S3). However, principal component analysis revealed considerable qualitative differences in the antibody repertoires between our cohorts and in particular between pediatric and adult subjects (Figure 2A). For comparison, we also performed the same analysis on enriched peptides that were derived

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

from HRSV. As expected, seroprevalence for the latter virus was considerably higher (QBB: 68.4 %; ABD: 77.5 %; PED: 90.4 %) (Supplementary Figure S4A) but in contrary to antiviral antibody responses to endemic HCoVs, we did not find any differences in variance of the HRSV-specific antibody responses when comparing age groups and cohorts (Supplementary Figure S4B). To determine the antibody specificities responsible for most of the variance in the antiviral response to endemic HCoVs between adults and children (i.e. to identify those peptides that were significantly more or less frequently enriched when comparing adult and pediatric donors), we applied Fisher's exact test and computed odds ratios for each of the significantly enriched peptides. We found that antibody specificities in samples of pediatric study subjects predominantly targeted different antigenic regions in the S protein (mean log odds ratio = 3.35 $\pm$ 2.12), the N protein (mean log odds ratio = 2.21 $\pm$ 1.41) and diverse antigenic sites in pp1ab, whereas peptides encoding a single linear B cell epitope of pp1ab (cluster 22) appeared to be the predominant target of IgG antibodies among adult donors (mean log odds ratio = -4.7 $\pm$ 1.16) (Figure 2B and Table 1).

Intriguingly, multiple sequence alignments of frequently enriched peptides with the full-length proteins of various CoVs revealed that antibody specificities predominantly found in pediatric study subjects target immunodominant epitopes that encode functionally important and highly conserved regions of the structural proteins. These included regions in the S_1 subunit of the S protein which are important for receptor binding (Figure 3A-C and 3F) [39-42], as well as the regions resembling the proteolytic cleavage sites and fusion peptide of the S_2 subunit (Figures 3A, 3B, 3D-F). Of note, the immunodominant region spanning the furin-like S2' cleavage site in the S_2 subunit resembles one of the most conserved regions of the S protein, both in amino acid sequence (R↓SA[I/L]ED[I/L]LF) and in protein structure, as it forms an accessible alpha-helix immediately upstream of the fusion peptide (Figure 3F and Supplementary Figure S6) [43]. Moreover, we identified potential antibody binding sites in the N-terminal RNA binding domain, serine-rich region and in the C-terminal dimerization domain of the N protein (Figures 4A and 4B). Although the predicted antibody binding sites in the N-terminal RNA-binding domain and the C-terminal dimerization domain of the N protein appeared to be less conserved between different species in the primary amino acid sequence (Figure 4C and 4D), both domains are structurally conserved in the regions that we found to be immunodominant (Figure 4E and 4F). We also found that antibodies in children

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

targeted more frequently the C-terminal domain of the M protein (Supplementary Figure S5C and Table 1) and the small accessory Orf8 protein (also known as N2) of HCoV-HKU1 (Table 1). Although Orf8 and N share the same coding sequence in the viral RNA genome, the reading frame is different and the amino acid sequences not homologous. On the contrary, antibody specificities predominantly found in adults primarily targeted a region of the pp1ab that is specific to HCoV-HKU1 and contains an acidic tandem repeat (ATR) of N-[DN]-D-E-D-V-V-T-G-D which is located upstream of the papain-like protease 1 domain (Supplementary Figure S5D).

Given the high degree of sequence conservation among some of the immunodominant regions in proteins of endemic HCoVs we have identified, we also explored to which extent antibody specificities targeting such conserved regions in structural proteins of endemic HCoVs may cross-react with peptides from epidemic CoVs and non-human CoV isolates. For this purpose, we assessed the enrichment of peptides derived from SARS-CoV, MERS-CoV and bovine, porcine, bat and feline isolates (Supplementary Table S1) applying the same approach and stringent filter criteria as we have described above for peptides of endemic HCoVs. Indeed, we identified several S protein-derived and N protein-derived peptides from epidemic CoVs or non-human isolates that were significantly enriched in our PhIP-Seq assay, which share sequence similarity with peptides from HCoVs (Figure 5). As expected based on the results from multiple sequence alignments described above, antibody specificities targeting the highly conserved amino acid motive (RSA[I/L]ED[I/L]LF) spanning the furin-like S2' cleavage site of the S protein also appeared to exhibit the broadest cross-reactivity as we identified several orthologous peptides from MERS-CoV, SARS-CoV as well as from non-human isolates enriched in our assay (Figure 5A, Region 3). Cross-reactivity of antibodies targeting other functionally important but less conserved regions also appeared to be more restricted (Figure 5A, Region 1 and 2). Similarly, antibody specificities targeting the N protein showed considerable cross-reactivity with peptides from MERS-CoV, SARS-CoV and non-human CoV isolates. However, the latter cross-reactive antibodies mainly targeted regions rich in serine and arginine residues, with low-complexity sequences and very limited structural conservation, particularly an intrinsically disordered region (IDR) at the N terminus of the N protein (Figure 5B and Supplementary Figure 7) that lacks a tertiary structure [44]. We also detected cross-reactive antibodies targeting the Serine-rich motif and linker region

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

Our comprehensive screen for antiviral antibody repertoires across individuals in our three human cohorts revealed a large number of peptides with novel linear epitopes in several proteins of endemic HCoVs. This is not surprising given that epidemic CoVs, and in particular SARS-CoV, have been the primary focus of previous immunological and epitope screening studies [30, 37]. Information about the targets of immune responses to CoVs across different species provides a valuable resource for the prediction of candidate targets of newly emerging CoVs, as it has recently been done by Grifoni *et al.* [30]. The authors were able to identify *a priori* several specific regions of the S, M and N proteins of SARS-CoV-2 S based on sequence homology to the SARS-CoV virus which are orthologous to several of the immunodominant regions we have identified among endemic HCoVs. We found antibody responses to be targeting the structural S, N, M and Orf8 proteins but also the nonstructural pp1ab polyprotein of HCoVs, the latter resembling the precursor for the large viral replicase complex [45]. Interestingly, in another independent study, Grifoni *et al.* [29] recently reported similarly broad T cell responses in COVID-19 patients by employing a comprehensive and analogous screen for T cell epitopes of SARS-CoV-2 proteins using peptide ‘‘megapools’’ in combination with *ex vivo* T cell assays.

Using stringent cut-off values, we have also estimated the seroprevalence of endemic HCoVs to range from ~4% to ~27% depending on the species and cohort (Figure 1A-C). This is largely consistent with earlier studies [46-50] but serological studies on larger sample sizes and in the general population are sparse. Gaunt *et. al.* retrospectively assessed 11,661 respiratory samples collected from 7,383 patients at hospital and primary care settings in Southeast Scotland for routine respiratory virus screening by PCR and detected active infection with endemic HCoVs in 0.3 to 0.85% of samples in all age groups, as compared to a detection rate of 11.1% and 6.7% for HRSV and HAdVs, respectively [5]. Since molecular diagnostic testing is typically done in critically ill patients and patients with comorbidities or immune suppression [51], it is not surprising that the majority of cases assessed in the study by Gaunt *et. al.* [5] had lower respiratory tract infections. Another study of hospitalized Norwegian children with respiratory tract infection found that 8.2% and 4.5% of the children tested PCR-positive for CoV-OC43 and HCoV-NL63, respectively; only HRSV and HRVs were detected more frequently

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

[7]. Since the individuals assessed in the present study largely reflect the general adult population or children with conditions unrelated to infection, our seroprevalence results most accurately reflect the seroprevalence of HCoVs in the State of Qatar and is likely attributable to mostly mild infections. Interestingly, age appeared to be negatively associated with seroprevalence in our study, suggesting that the either duration of immunity in response to natural infection with endemic HCoVs, or rates of re-infection, reduce with increasing in age. Dynamics of humoral and cellular immune responses against CoVs are poorly understood and correlates of protection remain controversial [52]. Longitudinal studies have mainly been focused on the duration of humoral immunity in MERS and SARS patients and suggest that antibody responses wane relatively quickly to minimal detectable levels over a period of not more than 2–3 years [53]. On the other hand, 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 the earlier report by Gaunt *et al.* [5], who found significantly more male than female laboratory-confirmed cases in a large number of diagnostic respiratory samples described above. The underlying molecular reasons remain unclear.

A surprising and unexpected finding of our study is that circulating IgG antibodies in children versus adults appear to be differentially targeting structural and non-structural proteins of HCoVs (Figure 2). Whereas antibody specificities in samples of pediatric subjects predominantly targeted structural proteins such as the N and S proteins, in adult donors, a region of the non-structural polyprotein pp1ab containing a tandem repeat of N-[DN]-D-E-D-V-V-T-G-DA in HCoV-HKU1 appeared to be the predominant target of IgG antibodies. The latter polyprotein is post-translationally processed into up to 16 subunits that form a large viral replicase complex responsible for both continuous and discontinuous RNA synthesis [45]. This qualitative difference in the antibody repertoires of adult versus pediatric study subjects appeared to be a specific characteristic of natural HCoV infection, as we did not find any variance in the antibody repertoires specific to HRSV when comparing our cohorts and different age groups (Supplementary Figure S3). Several of the immunodominant regions we have identified experimentally in the structural proteins of endemic HCoV are orthologous to

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

than adults and at the same time, are also less likely to experience severe disease outcomes [18-20, 22]. Broadly cross-reactive antibody responses are also known for other enveloped RNA viruses, which may positively or negatively affect subsequent infection or vaccination. Flaviviruses for example are antigenically related and broadly flavivirus cross-reactive antibodies from previous yellow fever vaccination has been shown to impair and modulate immune responses to tick-borne encephalitis vaccination [63]. Similarly, immune history has been shown to profoundly affect protective B cell responses to influenza [64].

In summary, we have shown that natural infection with endemic HCoV elicits humoral responses with district antibody repertoires in adults and children. To which degree pre-existing immunity from past infections with endemic HCoVs also provides cross-protection to the more distantly related epidemic CoVs needs further investigation. Our findings may have important implications for the development of prophylactic or therapeutic monoclonal antibodies and vaccine design in the context of COVID-19 [30, 65].

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Phage Immunoprecipitation-Sequencing (PhIP-Seq). Large scale serological profiling of the antiviral IgG repertoires in the individual serum or plasma samples was performed as described by Xu el al. [32]. Each serum or plasma sample was tested in duplicate and samples were analyzed in batches with up to 96 samples each batch. Only those samples were considered for downstream analysis that satisfied a minimum read count of $1*10^6$ as well as a Pearson correlation coefficient of ≥ 0.7 in the two technical repeats. Data from thirty individuals of the ABD cohort and two individuals of the QBB cohort were excluded from the downstream analysis due to insufficient sequencing read depth, low sequencing data quality or because one of two technical replicates had failed (not shown).

Peptide enrichment analysis. To filter for enriched peptides, we first imputed −log10(P) values as described previously [32, 35, 38] by fitting a zero-inflated generalized Poisson model to the distribution of output counts and regressed the parameters for each peptide sequence based on the input read count. We considered a peptide enriched if it passed a reproducibility threshold of 2.3 [−log10(P)] in two technical sample replicates. To remove sporadic hits, we then filtered for antibody specificities to CoV peptides that were found to be enriched in at least three of all 1399 subjects assayed and analyzed in this study. We computed species-specific significance cut-off values to estimate minimum number of enriched, non-homologous peptides required in order to consider a sample as seropositive using a generalized linear model and in-house serological (ELISA) data from pooled samples that were tested positive for various viruses. We then computed virus score values as described by Xu et al [32] by counting enriched, non-homologous peptides for a given species and then adjusted these score values by dividing them with the estimated score cutoff. For the purpose of this study and under consideration of the seroprevalence of endemic HCoVs in our three cohorts, we considered a peptide to be immunodominant if it was enriched in $\geq 1\%$ of all 1399 subjects assayed and analyzed in this study.

Association studies and differential enrichment analysis. We applied a generalized linear model to test for associations between the HCoV species-specific adjusted score values, gender and age. We considered an association to be significant if the P-value was ≤ 0.001. We examined the frequency distribution of enriched peptides among samples of the different age groups (PED versus ABD + QBB) by estimating odds ratios (OR) and associated P-values using

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fisher's exact test. Peptides that satisfied both significance (P-value ≤ 0.005) and magnitude criteria ($|\log(OR)| \geq \log(2)$) were considered to be differentially enriched. Positive log(OR) values indicated more frequent peptide enrichment among pediatric study subjects, whereas negative log(OR) values indicated more frequent peptide enrichment among adult subjects.

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

Software. Open source Python modules with in-house scripts were used to test for associations (statsmodel v0.11), to filter differentially enrichment peptides and to perform different statistical tests (sklearn v0.23, scipy v1.14.1). Multiple sequence alignments were done using the MAFFT [66, 67] via EMBL-EBI's web services and Java Alignment Viewer (Jalview) for visualization [68]. Residue-wise linear B cell epitopes were predicted using BepiPred-2.0 [69]. Protein structures graphics were generated using PyMOL (Schrödinger).

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

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

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

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

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

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

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

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

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

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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

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

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

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

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures

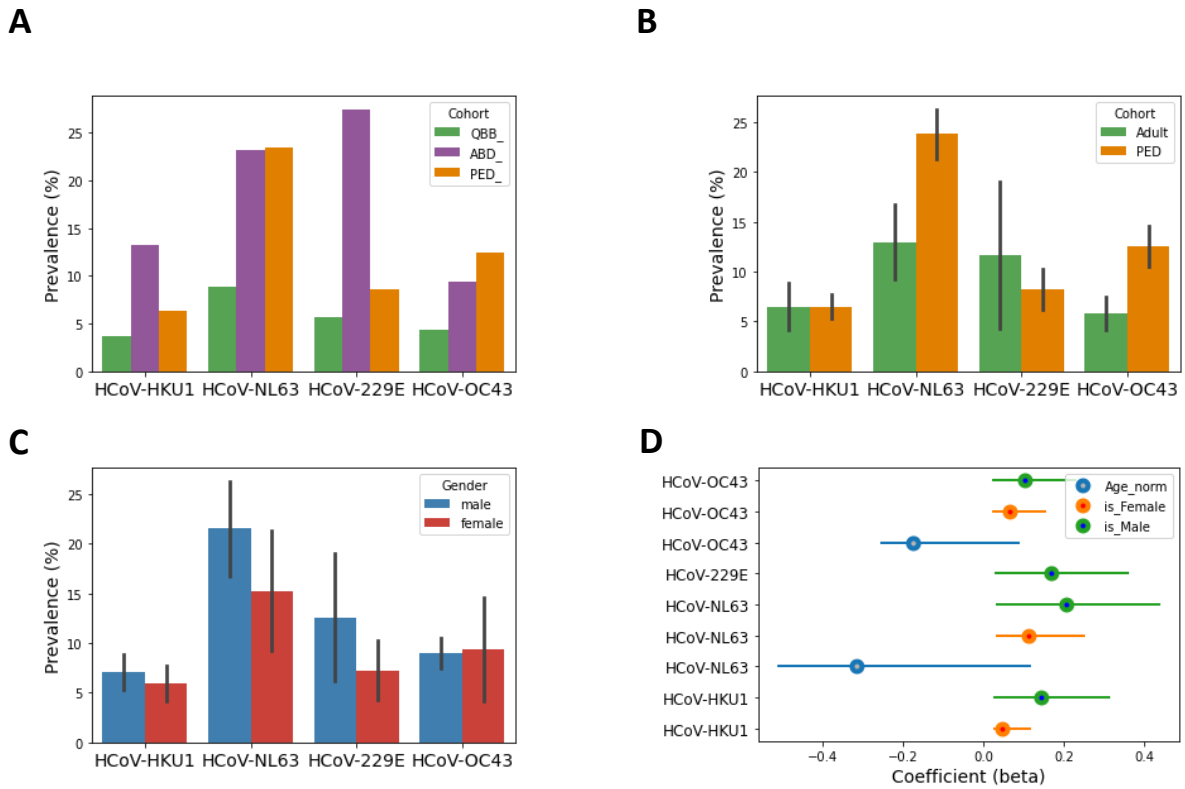


Figure 1. Seroprevalence of endemic HCoVs. A-C, Bar plots depict the seroprevalence of the four endemic HCoVs stratified by cohort (A), age group (B) and gender (C). Error bars in (B) depict gender-specific variation. Adults include adult blood bank donors and individuals of the Qatar Biobank cohort. 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

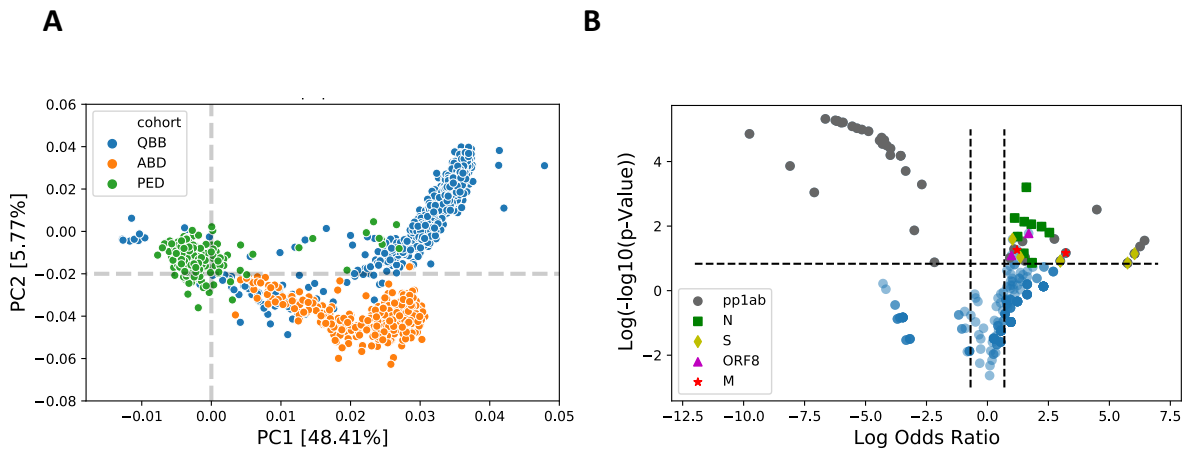


Figure 2. Qualitative differences in antibody repertoires between cohorts and age groups. A, Principal component analysis of 417 peptides from endemic HCoVs that were found to be enriched in at least 3 samples. QBB, Qatar Biobank cohort; ABD, adult (male) blood bank donors; PED, pediatric subjects. **B,** 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.

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

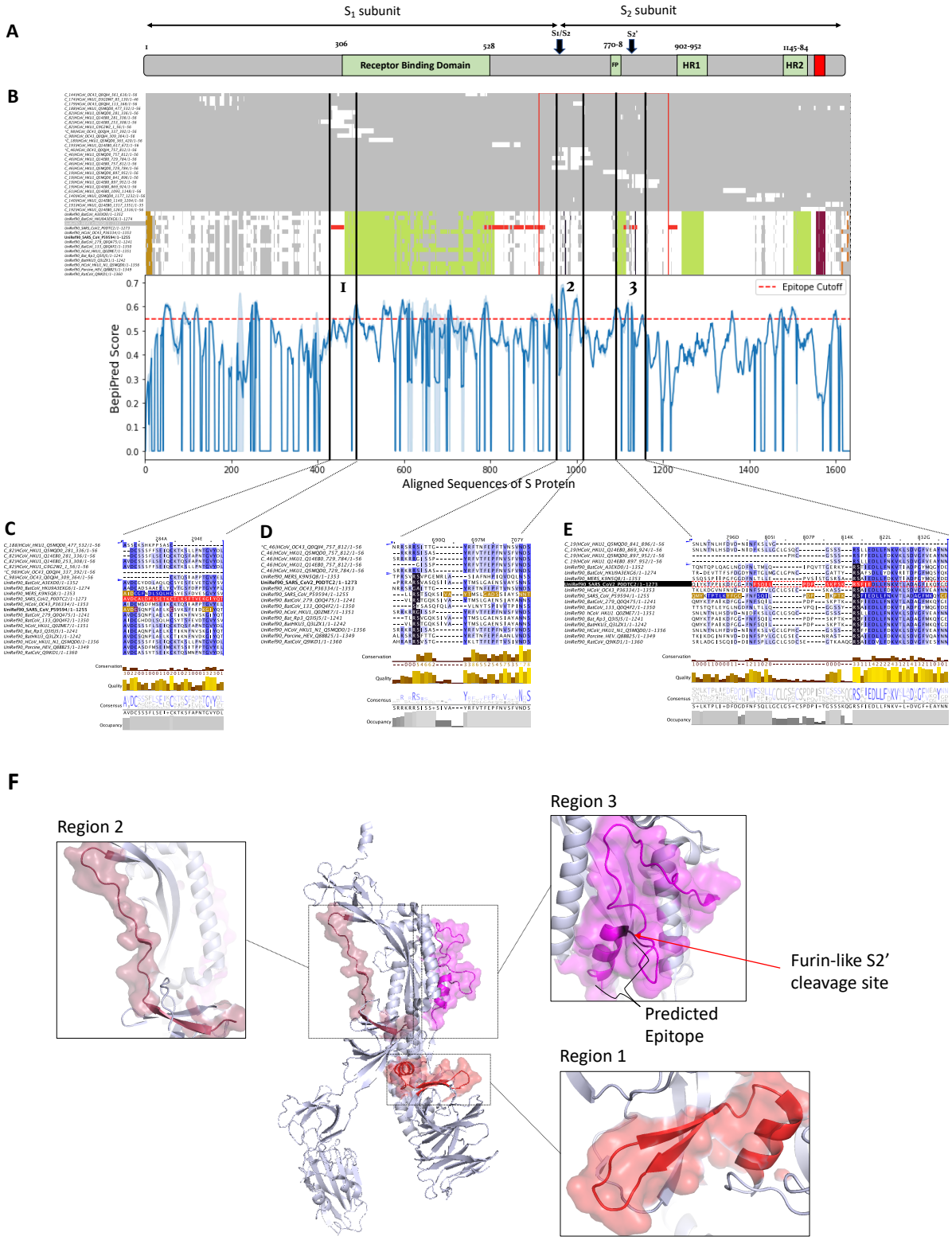


Figure 3. Antigenic regions and predicted antibody binding sites of the spike (S) protein. A, Schematic representation of the S protein of SARS-CoV (strain Tor2). Proteolytic cleavage sites of the S protein are marked with arrows. **B,** Overview of a multiple sequence alignment of

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

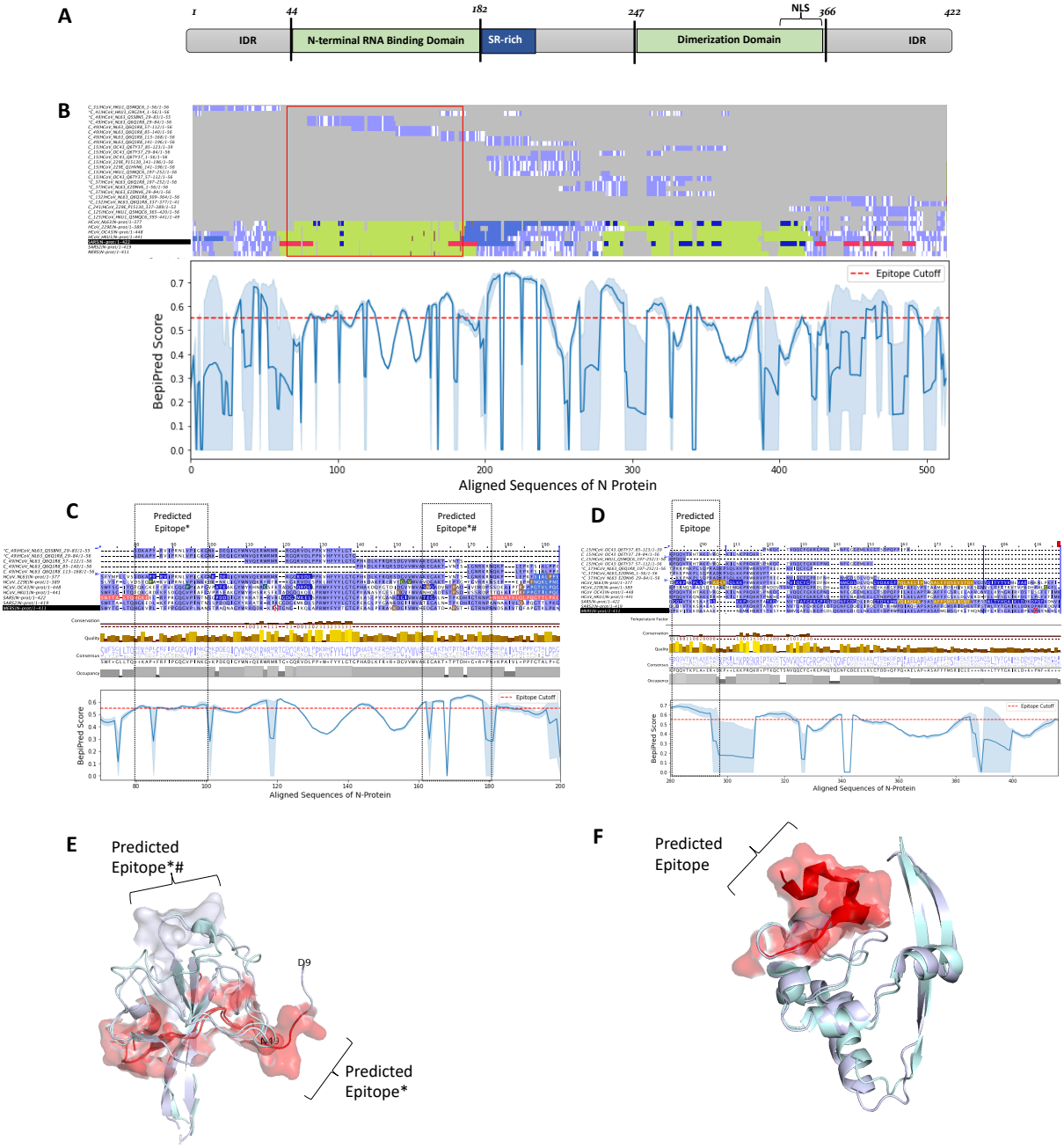


Figure 4. Antigenic regions and predicted antibody binding sites of the nucleocapsid (N) protein. A, Schematic representations of the N protein of SARS-CoV (strain Tor2). SR-rich, serine-rich; NLS, Predicted nuclear localization sequence; IDR, intrinsically disordered region. **B,** Overview of a multiple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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

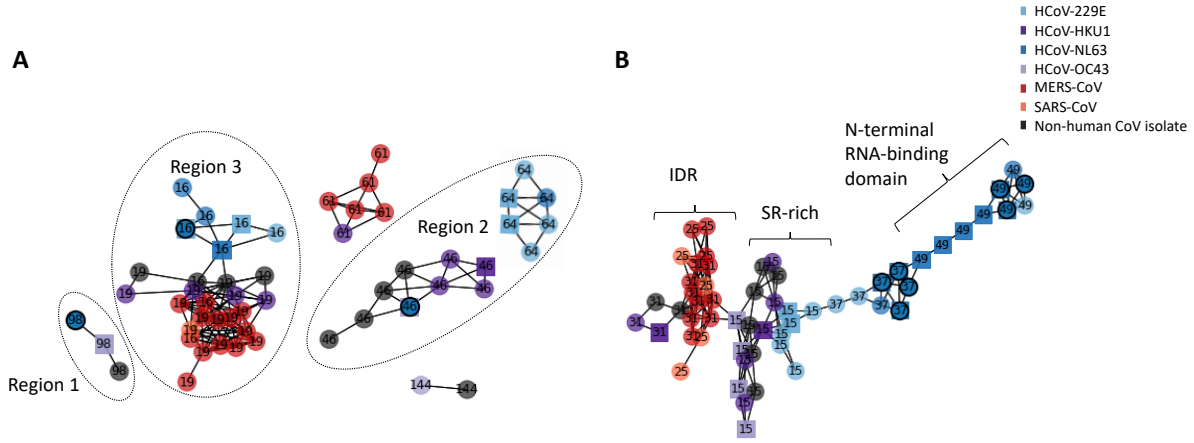


Figure 5. Network representation of enriched peptides from structural proteins targeted by cross-reactive antibodies. **A,** Network representation of enriched spike (S) protein-derived peptides. **B,** Network representation of enriched nucleoprotein (N)-derived peptides. Each node represents an enriched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

bioRxiv preprint doi: https://doi.org/10.1101/2020.06.21.163394. this version posted June 22, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

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**

Supplementary Materials for

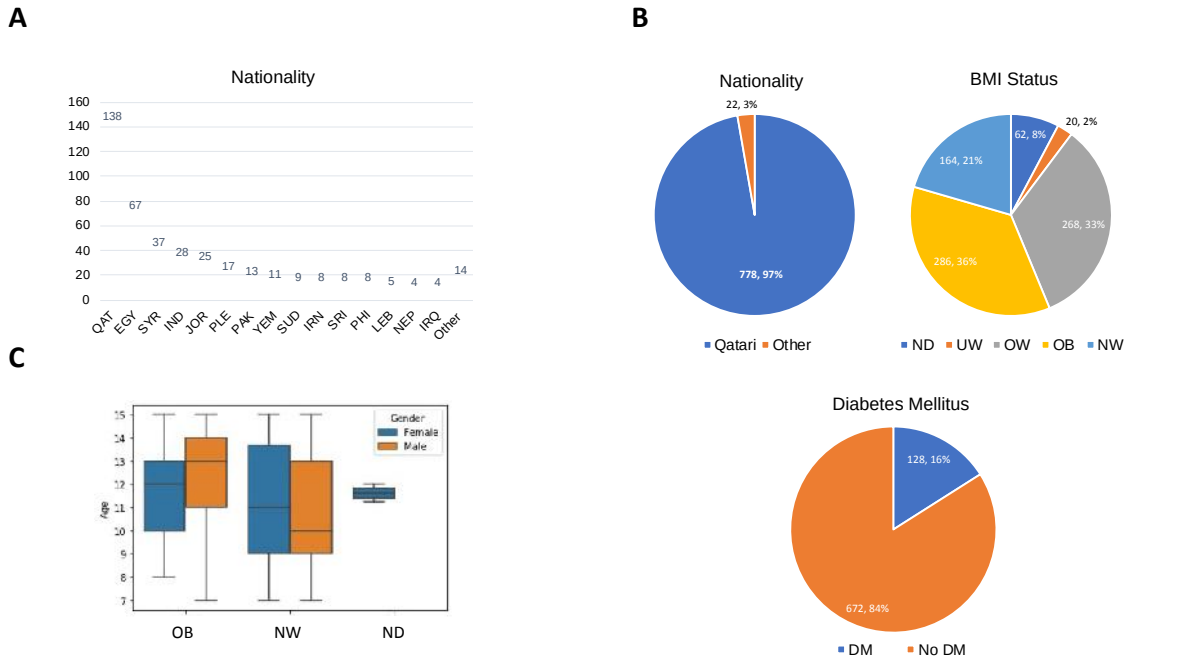
Endemic human coronaviruses induce distinct antibody repertoires in adults and children

Taushif Khan, Mahbuba Rahman, Fatima Al Ali, Susie S. Y. Huang, Amira Sayeed, Gheyath K. Nasrallah, Mohammad R. Hasan, Nico Marr*

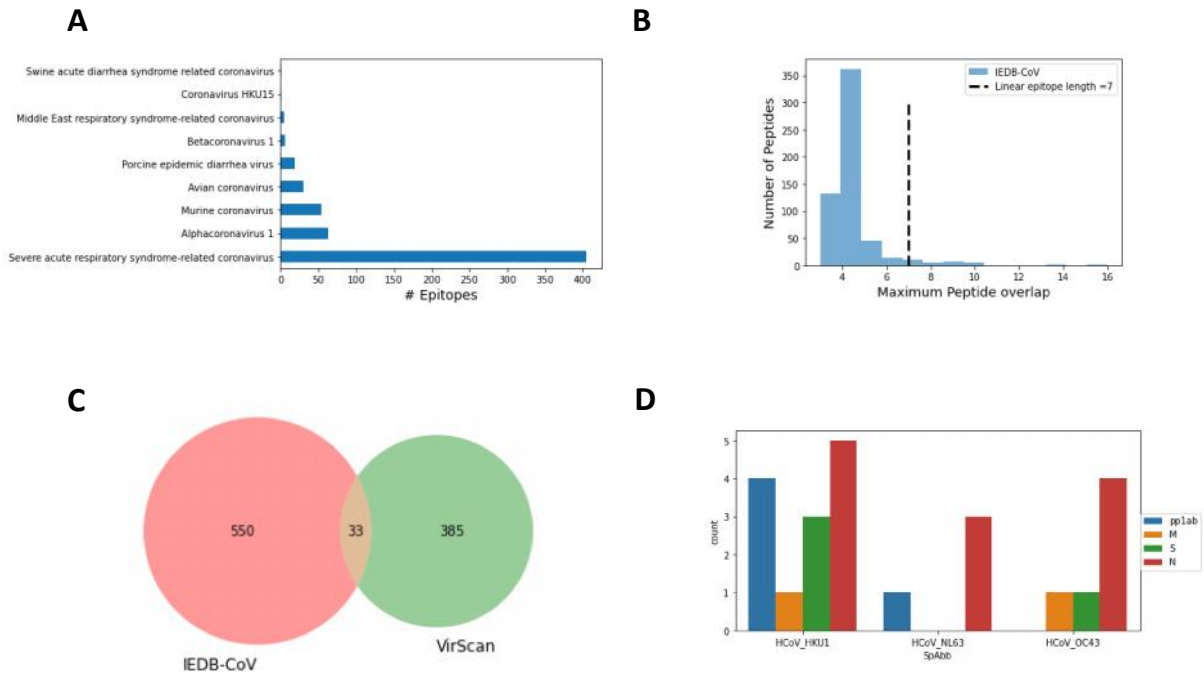
Version 1, 21 June 2020

*Corresponding author: Nico Marr, Sidra Medicine, Research Branch, Al Gharrafa Street, Ar-Rayyan, PO BOX 26999, Doha, Qatar, E-mail: nmarr@sidra.org

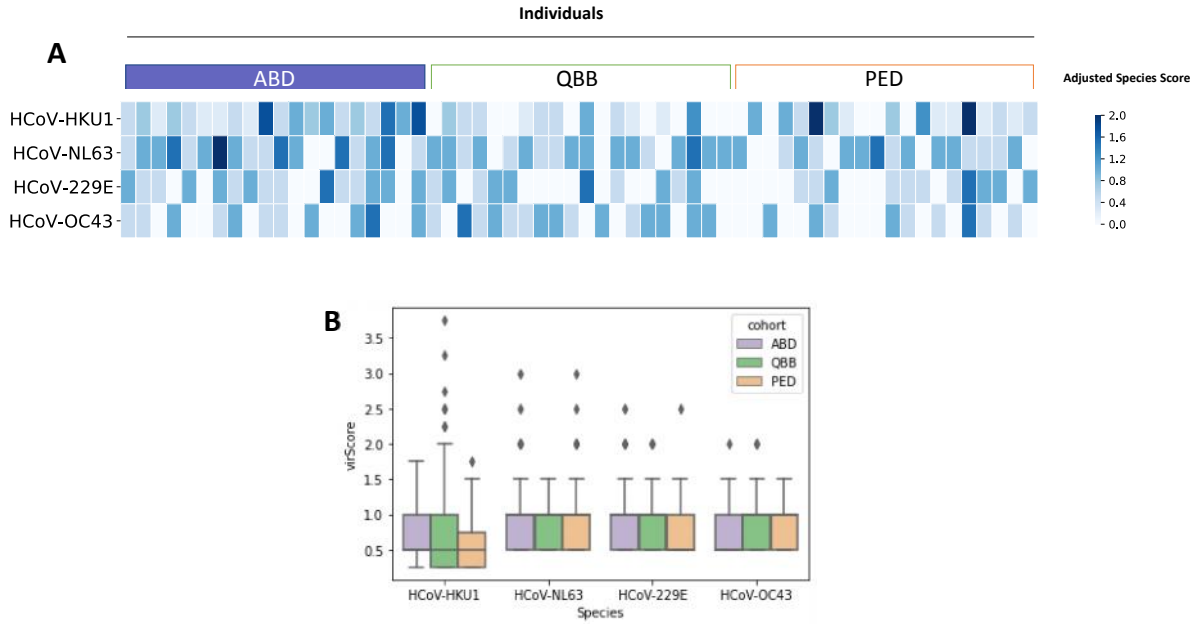
Supplementary Figures



Supplementary Figure S1. Demographic and clinical features of the different human cohorts. A, Nationalities among the healthy blood bank donors. **B,** Number and fraction of Qatari nationals and long-term residents with other nationality comprising the 800 assayed individuals of the Qatar Biobank cohort, stratified by body mass index (BMI) and proportion of subjects with diabetes mellitus (DM). UW, underweight (BMI < 18); NW, normal weight (BMI ≥ 18 and < 25); OW, overweight (BMI ≥ 25 and <30); OB, obese (BMI ≥ 30). **C,** Number and fraction of male and female pediatric study subjects with normal weight (NW) or who were obese (OB). ND, not determined; QAR, Qatar; EGY, Egypt; SYR, Syria; IND, India; JOR, Jordan; PAK, Pakistan; YEM, Yemen; SUD, Sudan; IRN, Iran; SRI, Sri Lanka; PHI, Philippines; LEB, Lebanon; NEP, Napal; IRQ, Iraq; Other, residents with other nationality; PLE, non-residents.

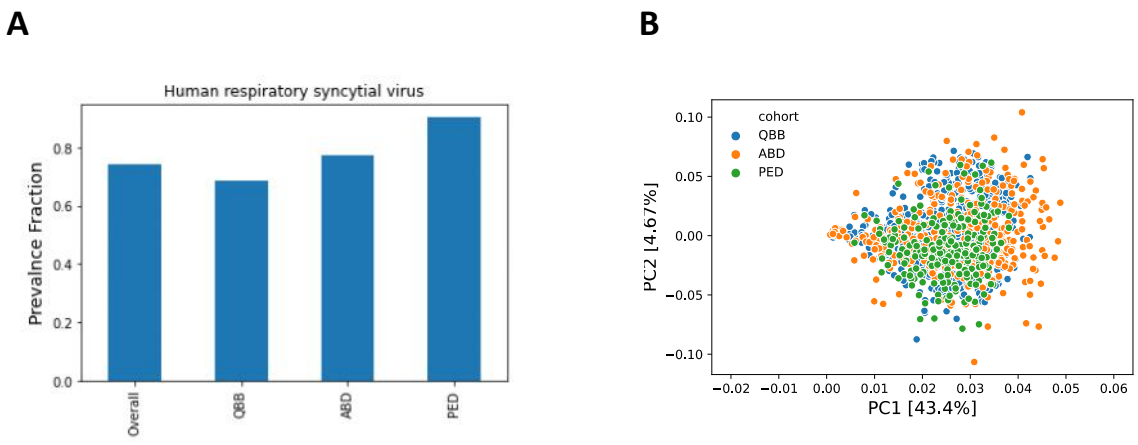


Supplementary Figure S2. Comparison analysis of enriched HCoV peptides identified by PhIP-Seq with B cell epitopes in IEBD (www.iebd.org). A, Number of B cell epitopes for different CoV species available in IEBD (accessed and downloaded from www.iedb.org on 17 May 2020). **B,** Number of peptides from endemic HCoVs which we found to be significantly enriched in at least three of all 1399 samples with homology in the linear amino acid sequence to 583 known CoV B cell epitopes in IEBD. We considered an epitope as a match if there was ≥ 7 linear sequence identity between query (IEBD epitope) and the amino acid sequence of the enriched HCoV peptide. **C,** Venn diagram depicting overlap between enriched peptides (i.e. matches) detected by VirScan and B cell epitopes from CoVs in IEDB. **D,** Number of matching peptides that share sequence homology with known B cell epitopes, stratified by species and viral protein.



Supplementary Figure S3: Similar antibody repertoire breadth among seropositive individuals of the different cohorts. A, Heatmap plot depicting adjusted species score values for HCoV-HKU1, -NL63, -229E and -OC43 in randomly selected seropositive individuals (n = 20) of each cohort. Each individual is represented as a column. Values ≥ 1 indicate that the individual was seropositive for a given species. **B,** Species-wise variation in adjusted species score values among seropositive individuals of each cohort.

53



54

55

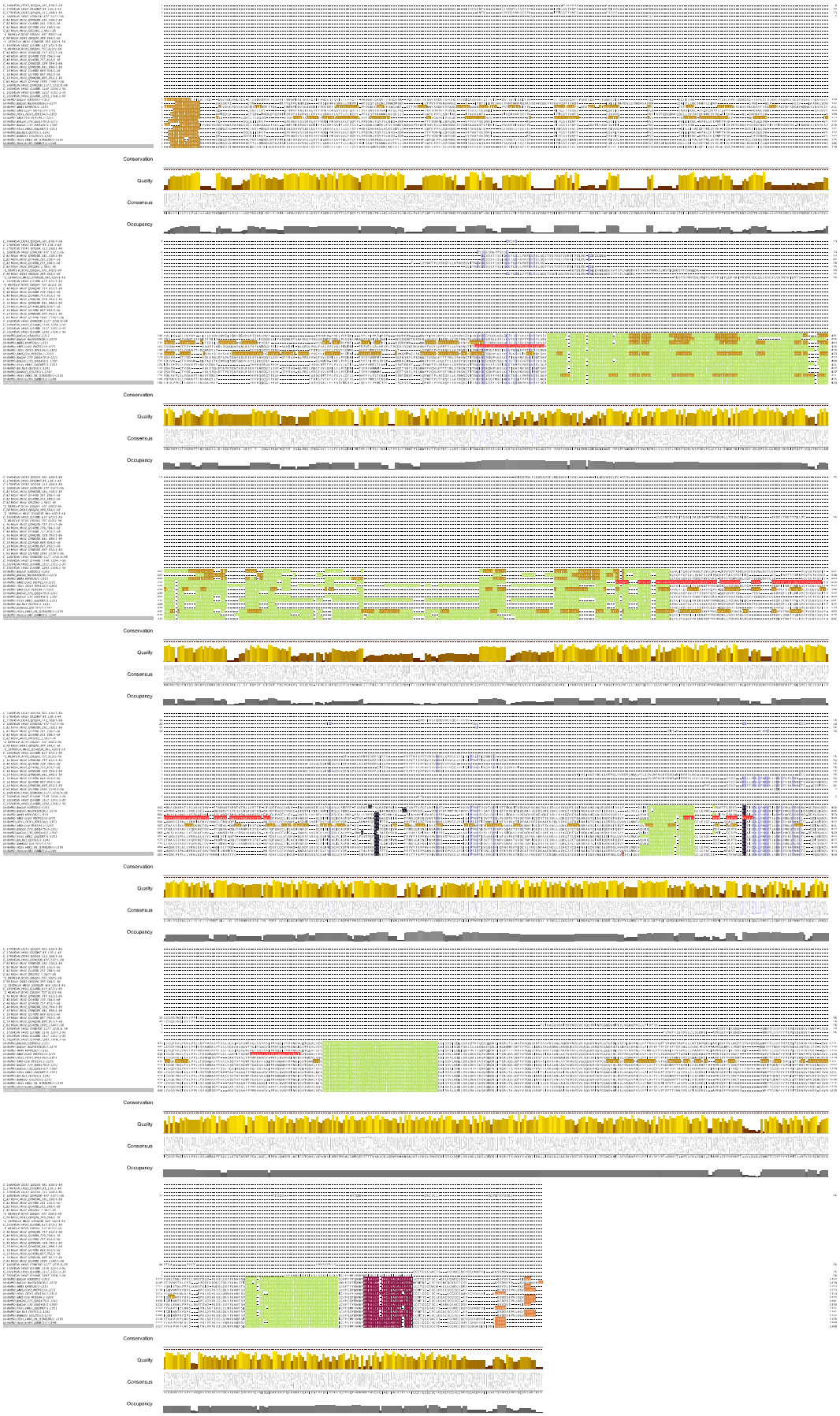
56 **Supplementary Figure S4. Similar seroprevalence and antibody repertoires to HRSV in the**

57 **three human cohorts. A,** Bar plot depicting the fraction of HRSV seropositive individuals of

58 all tested subjects or by cohort. **B,** Principal component analysis of 560 HRSV peptides found

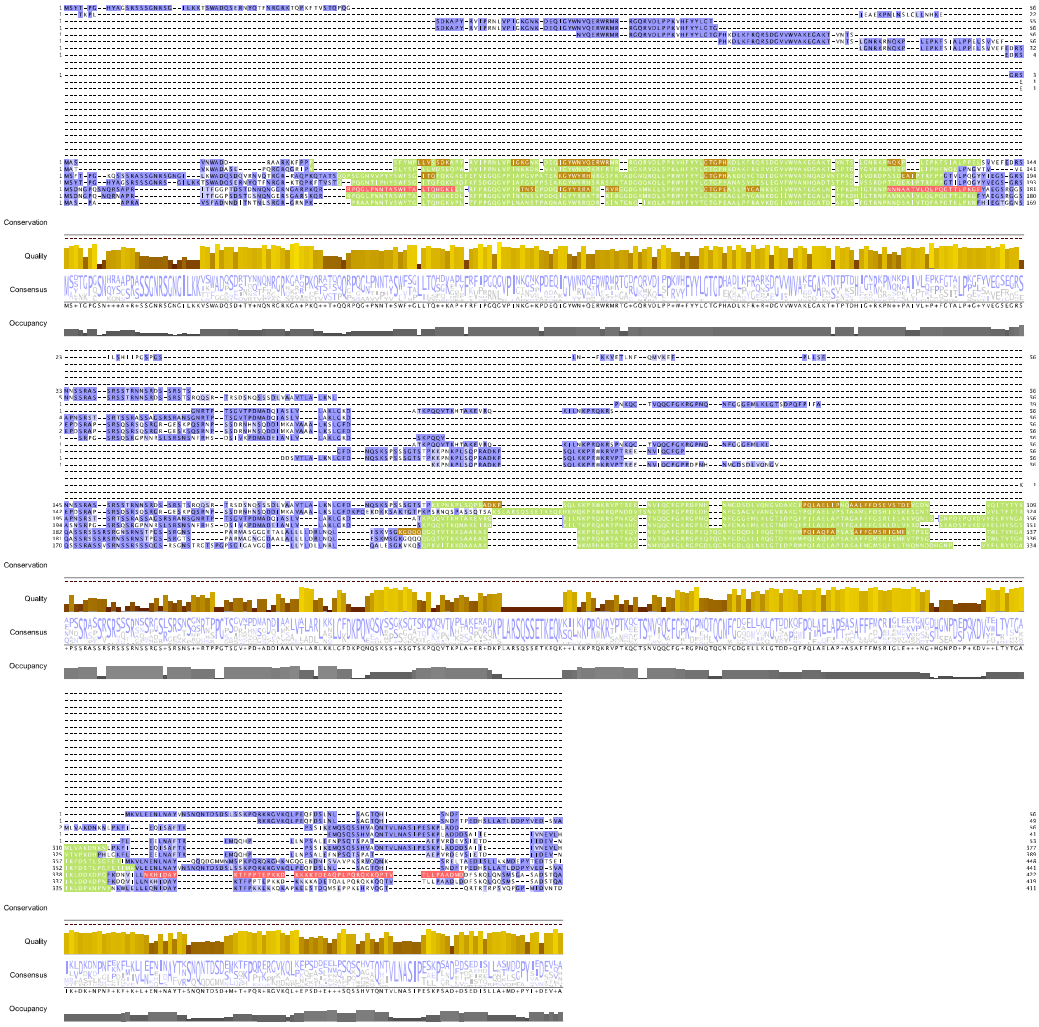
59 to be enriched in at least 3 samples.

60 **A**


Conservation
Quality
Consensus
Occupancy
Conservation
Quality
Consensus
Occupancy
Conservation
Quality
Consensus
Occupancy
Conservation
Quality
Consensus
Occupancy
Conservation
Quality
Consensus
Occupancy

61

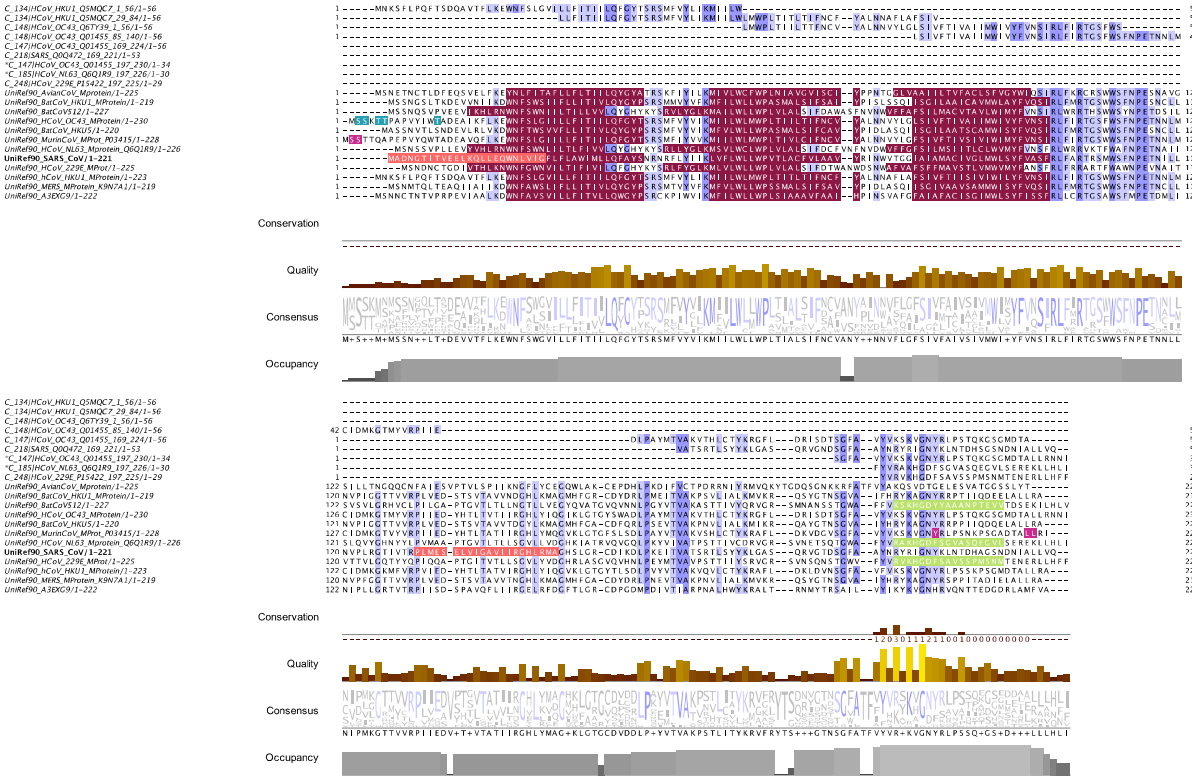
62 **B**


Conservation
Quality
Consensus
Occupancy
Conservation
Quality
Consensus
Occupancy
Conservation
Quality
Consensus
Occupancy

63

64

65 **C**


Conservation
Quality
Consensus
Occupancy

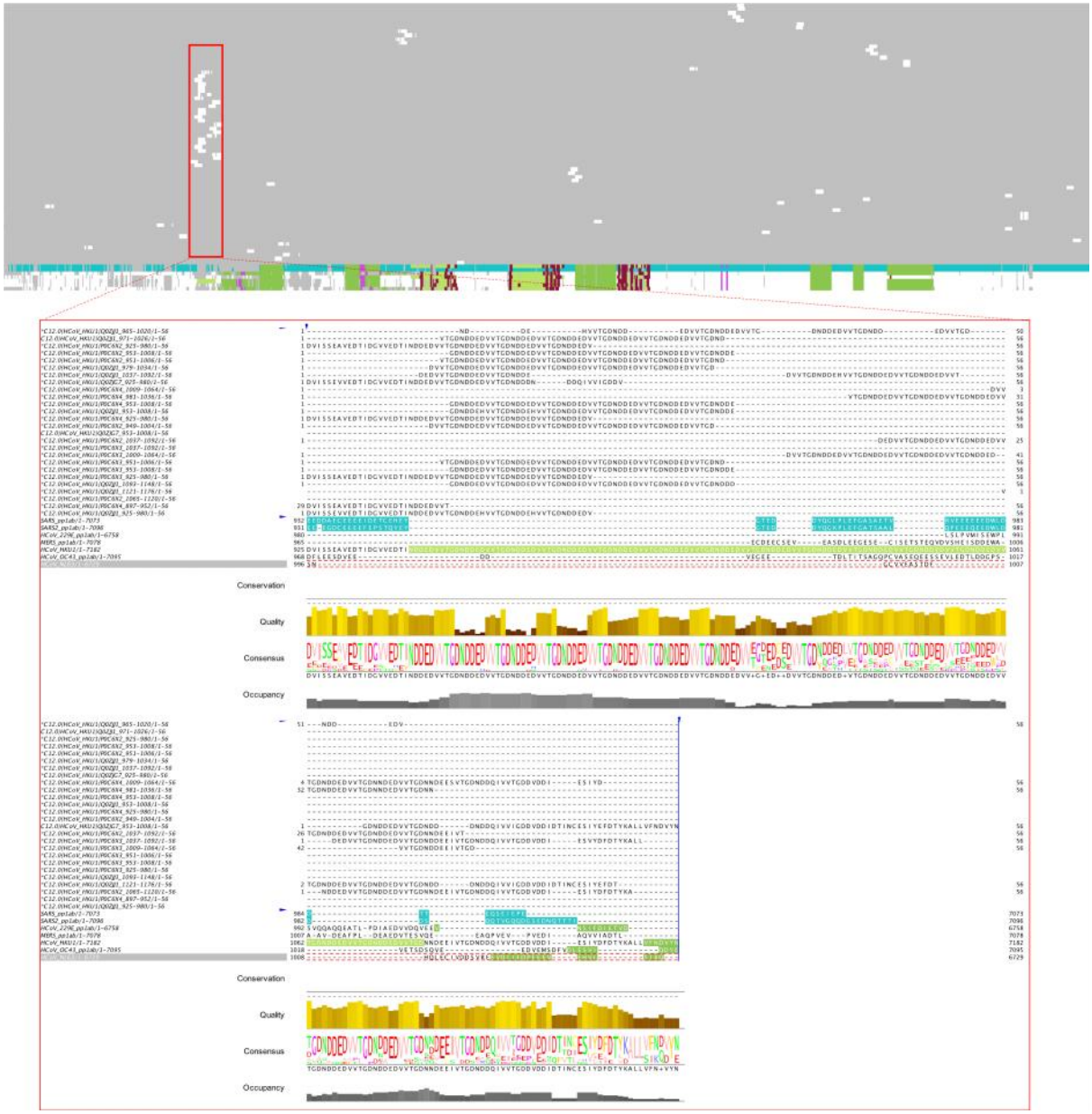
Conservation
Quality
Consensus
Occupancy

66

67

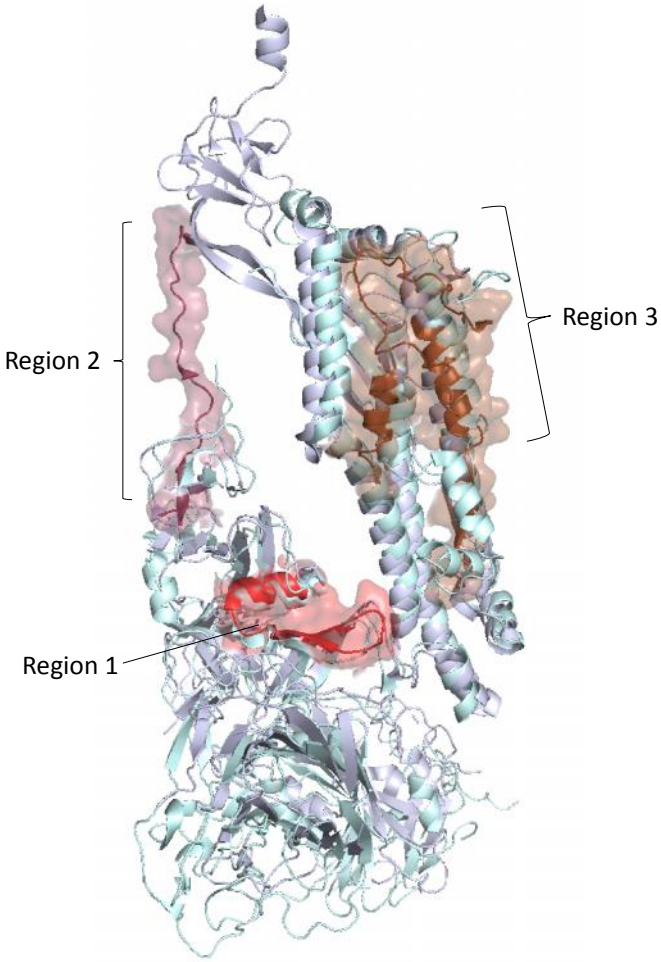
68

D

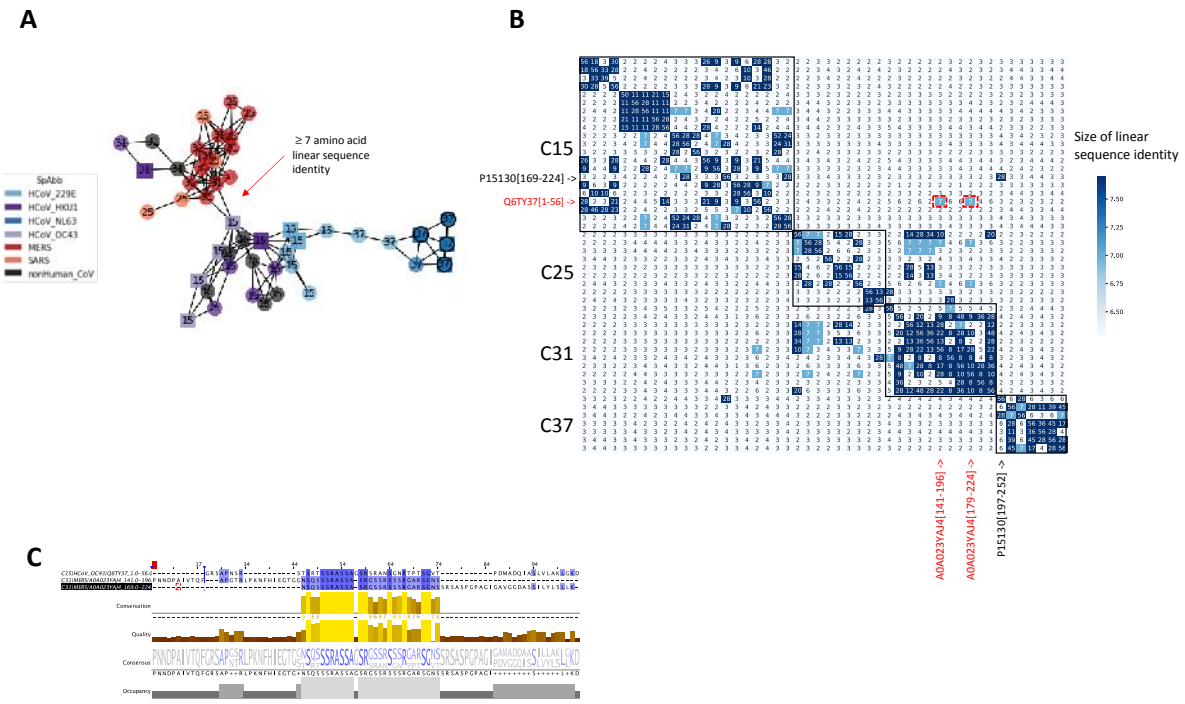


Supplementary Figure S5. Mapping of enriched, immunodominant peptides of endemic HCoVs to the full-length protein sequences. **A-D**, Multiple sequence alignments of the enriched, immunodominant peptides with the full-length sequences of the spike protein (S), nucleocapsid protein (N), matrix glycoprotein (M) and the Orf1ab replicase polyprotein (pp1ab) of selected CoVs. Amino acids with high percentage of sequence identity are shown in violet. Reference sequence features are color-coded in accordance to the annotation and features available in UniProt (subjected to availability) and visualized using JalView. Green indicates regions of interest, red indicates transmembrane regions, orange indicates beta

80 strands and pink indicates residues annotated as epitopes in IEDB. The cleavage site (S1/S2
81 and S2) of the SARS-CoV S protein is shaded in black. For pp1ab, the sequence alignment is
82 shown as an overview with gaps are shaded in grey and sequences are in white. Only part of
83 the sequence alignment is shown, which depicts the PL1-PRO cleavage product containing a
84 tandem repeat of N-[DN]-D-E-D-V-V-T-G-DA in HCoV-HKU1 (isolate N1) (UniProt entry and
85 position P0C6X2[945-1084]).



Supplementary Figure S6. Superimposition of Spike protein of SARS-CoV2 (6VXX) and HCoV-HKU1 (5I08) in the prefusion state. Root-mean-square deviation [rmsd] = 3.328 Ångström.



Supplementary Figure S7. Clustering of peptides for shared linear B cell epitopes. A, Network representation of peptides built from a pairwise distance matrix as shown in Figure 5B. Each node represents an enriched peptide and the color indicates the species. Edges indicate ≥ 7 amino acids linear sequence identity between two nodes (i.e. peptides), the estimated size of a linear B cell epitope. **B,** Pairwise distance matrix of peptides targeted by cross-reactive antibodies, including enriched peptides of clusters 15, 25 ,31 and 37. The cluster assignment has been indicated as a black box. Distance pairs with ≥ 7 amino acid linear sequence identity are shown in dark blue. Distance pairs marked with a red box represent a 7 amino acid linear sequence identity between enriched peptides of more distantly related CoVs, namely HCoV-OC43 (Q6TY37[1-56]) and MERS-CoV (A0A023YAJ4[141-196] and A0A023YAJ4[179-224]). **C,** Multiple sequence alignment of enriched HCoV-OC43- and MERS-CoV-derived peptides Q6TY37[1-56], A0A023YAJ4[141-196] and A0A023YAJ4[179-224] encoding part of the N-terminal intrinsically disordered region (IDR) of the N protein. The alignment shows the 7 amino acid linear sequence identity in a low-complexity sequence motive (SSRASSA) which likely represents a broadly cross-reactive antibody binding site.

110 # Supplementary Tables

111 **Supplementary Table S1.** Proteins of CoVs represented in the VirScan phage library.

112 Endemic hCoVs are shown in bold.

113

Species	*Organism*	*Protein*	*Number of Peptides*
Alphacoronavirus 1	Feline coronavirus (strain FIPV WSU-79/1146) (FCoV)	NS	239
Bat coronavirus 1B	Bat coronavirus 1B	Rep1a	249
Betacoronavirus 1	Bovine coronavirus	N	4
	Bovine coronavirus (strain 98TXSF-110-ENT) (BCoV-ENT) (BCV)	E	2
		N	7
		NS	12
		S	48
	Equine coronavirus	Rep1a	254
	Human coronavirus OC43 (HCoV-OC43)	**HE**	**15**
		M	**11**
		N	**4**
		NS	**253**
		S	**37**
	Porcine hemagglutinating encephalomyelitis virus (strain 67N) (HEV-67N)	N	16
Human coronavirus 229E	**Human coronavirus 229E (HCoV-229E)**	**E**	**2**
		M	**8**
		N	**23**
		NS	**248**
		ORF4	**10**
		Rep1b	**2**
		S	**63**
Human coronavirus HKU1	**Human coronavirus HKU1 (HCoV-HKU1)**	**N**	**13**
		Rep1ab	**510**
		S	**15**
	Human coronavirus HKU1 (isolate N1) (HCoV-HKU1)	**E**	**2**
		HE	**13**
		M	**7**
		N	**22**
		NS	**259**
		S	**48**
	Human coronavirus HKU1 (isolate N2) (HCoV-HKU1)	**E**	**2**
		HE	**13**
		N	**7**
		NS	**258**
		S	**48**
	Human coronavirus HKU1 (isolate N5) (HCoV-HKU1)	**NS**	**254**
Human coronavirus NL63	**Human coronavirus NL63 (HCoV-NL63)**	**E**	**2**
		M	**10**
		N	**18**
		NS	**248**
		ORF3	**1**
		Rep1a	**1**
		S	**71**
Middle East respiratory syndrome coronavirus	BtVs-BetaCoV/SC2013	E	2
		M	7
		N	15
		ORF3	3
		ORF4	11
		ORF5	8
		ORF8	6
		Rep1a	256
		S	47
	Coronavirus Neoromicia/PML-PHE1/RSA/2011	E	2
		ORF3	3
		ORF4	12
		ORF5	7
		ORF8	7
		Rep1a	156
		S	47

	Human coronavirus EMC (isolate United Kingdom/H123990006/2012) (HCoV-EMC)	E	2
		M	7
		N	14
		NS	408
		ORF3	3
		ORF4	11
		ORF5	7
		S	48
	Middle East respiratory syndrome coronavirus	N	8
		ORF4	11
		ORF5	11
		ORF8	3
		Rep1a	1018
		Rep1ab	95
		Rep1b	63
		S	108
Severe acute respiratory syndrome-related coronavirus	Bat coronavirus 279/2005 (BtCoV) (BtCoV/279/2005)	E	2
		M	7
		NS	254
		U	2
	Human SARS coronavirus (SARS-CoV) (Severe acute respiratory syndrome coronavirus)	N	15
		NS	9
		ORF3	9
		ORF7	4
		ORF9	3
		S	44
	Human SARS coronavirus (isolate Tor2) (SARS-CoV) (Severe acute respiratory syndrome coronavirus)	Rep1b	1

114

115 **Supplementary Table S2.** Number of individuals per cohort included in the downstream
116 analysis.
117

Cohort	Gender	Number	Age (years)			
			Mean	Std	Min	Max
ABD	Male	370	40.1	10.1	19	66
PED	Female	117	11.4	2.2	7	15
	Male	114	11.8	2.3	7	15
QBB	Female	511	40.4	12.9	19	81
	Male	287	41.0	12.9	19	81

118
119 QBB, Qatar Biobank cohort; ABD, adult blood bank donors; PED, pediatric subjects.
120
121

122 **Supplementary Tables S3 and S4. List of enriched HCoV peptides and cluster assignment.**

123 <TableS3 and S4_16062020.xls>