BRIEF REPORT

Expected immune recognition of COVID-19 virus by memory from earlier infections with common coronaviruses in a large part of the world population [version 1; peer review: 2 approved]

Johannes M. Dijkstra, Keiichiro Hashimoto
Institute for Comprehensive Medical Science, Fujita Health University, Toyoake, Aichi, 470-1192, Japan

Abstract
SARS-CoV-2 is the coronavirus agent of the COVID-19 pandemic causing high mortalities. In contrast, the widely spread human coronaviruses OC43, HKU1, 229E, and NL63 tend to cause only mild symptoms. The present study shows, by in silico analysis, that these common human viruses are expected to induce immune memory against SARS-CoV-2 by sharing protein fragments (antigen epitopes) for presentation to the immune system by MHC class I. A list of such epitopes is provided. The number of these epitopes and the prevalence of the common coronaviruses suggest that a large part of the world population has some degree of specific immunity against SARS-CoV-2 already, even without having been infected by that virus. For inducing protection, booster vaccinations enhancing existing immunity are less demanding than primary vaccinations against new antigens. Therefore, for the discussion on vaccination strategies against COVID-19, the available immune memory against related viruses should be part of the consideration.

Keywords
Coronavirus, COVID-19, SARS-CoV-2, OC43, HKU1, 229E, NL63, MHC class I, Immunology, Vaccination

This article is included in the Disease Outbreaks gateway.

This article is included in the Coronavirus collection.
Introduction

SARS-CoV-2 and other human coronaviruses

From the end of 2019, the world experienced the coronavirus disease 2019 (COVID-19) pandemic caused by the emerging severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2; aka 2019 novel coronavirus or 2019-nCoV). SARS-CoV-2 shares ~80% nucleotide identity with SARS-CoV-1 (aka SARS-CoV), the causative agent of the SARS epidemic from 2002, and is even more similar to some coronaviruses in bats (Andersen et al., 2020; Ceraolo & Giorgi, 2020; Wu et al., 2020; Zhou et al., 2020). Coronaviruses are membrane-enveloped positive-strand RNA viruses with, for an RNA virus, a large genome of ~30 kb. That genome encodes several structural components of the virion including the nucleocapsid protein N and the membrane proteins S (spike), M, and E, plus also a number of nonstructural proteins involved in RNA replication and other—partly unknown—functions (Weiss & Navas-Martin, 2005). The coronaviruses infecting humans belong to the serological/phylogenetic clades group I (alphacoronaviruses) and group II (betacoronaviruses); group I includes HCoV-229E (human coronavirus 229E) and HCoV-NL63, while group II includes SARS-CoV-1, SARS-CoV-2, Middle East respiratory syndrome coronavirus (MERS-CoV), HCoV-OC43, and HCoV-HKU1. The viruses SARS-CoV-1 and MERS-CoV, on average, cause the most severe symptoms, and their outbreaks were successfully monitored and halted. At the other end of the spectrum, the viruses HCoV-229E, HCoV-NL-63, HCoV-OC43, and HCoV-HKU1 tend to cause only mild symptoms and are very common.

Prevalence and associated disease of the common human coronaviruses 229E, NL63, OC43, and HKU1

The Centers for Disease Control and Prevention (CDC; https://www.cdc.gov/coronavirus/general-information.html) states: “Common human coronaviruses, including types 229E, NL63, OC43, and HKU1, usually cause mild to moderate upper-respiratory tract illnesses, like the common cold. Most people get infected with one or more of these viruses at some point in their lives.” The same agency lists the common symptoms caused by these viruses as runny nose, sore throat, headache, fever, cough, and general feeling of being unwell, but also explains that they occasionally cause lower-respiratory tract illnesses, such as pneumonia or bronchitis. The viruses 229E and OC43 have been known since the 1960s (reviewed in Kahn & McIntosh, 2005), but NL63 (van der Hoek et al., 2004) and HKU1 (Woo et al., 2005) were only (conclusively) identified following the rise in interest in coronaviruses in the wake of the SARS epidemic. These common coronaviruses are believed to be the second most common cause of the common cold (Mäkelä et al., 1998). In the U.S.A., a 3-year RT-PCR surveillance of respiratory samples of patients revealed that the four viruses 229E, NL63, OC43, and HKU1 were present at levels varying by season and region, with all individual viruses peaking at ~3% prevalence in each investigated region (Midwest, Northeast, South, West); co-infection with other coronaviruses was found in only ~2% of infected cases, but co-infection with another respiratory virus was found in a substantial ~30% of infected cases (Killerby et al., 2018). This pattern was reminiscent of findings in the United Kingdom (Gaunt et al., 2010) and Japan (Matoba et al., 2015). Serological investigations in countries as diverse as the U.S.A. (Bradburne & Somerset, 1972; Dijkstra et al., 2012), China (Zhou et al., 2013), and Qatar (Al Kahlout et al., 2019), found that most healthy blood donors had antibodies against coronaviruses, supporting that these viruses are widespread indeed.

Since immune memory protection can be induced by related pathogens, as exemplified by the eradication of human smallpox virus (Variola) by immunization with a related “cowpox” virus (Vaccinia) (Plotkin & Plotkin, 2018), it is interesting to consider whether common human coronavirus infections may have induced some level of protection against SARS-CoV-2.

The possibility of matching linear epitopes between SARS-CoV-2 and the common human coronaviruses that may stimulate the immune system through MHC class I presentation

The two major arms of immune memory concern antibody secretion by B cells and killing of infected cells by CD8+ T cells. For a coronavirus infection in mouse, both immune responses were needed to efficiently control the virus (reviewed by Weiss & Navas-Martin, 2005). Based on theoretical considerations alone, it is difficult to predict effective B cell memory across different virus species (Qu et al., 2020). From recent experiments concluded that sera from people that likely had been infected with the common human coronaviruses 229E, NL63, OC43, and/or HKU1, possessed no or negligible cross-reactivity with SARS-CoV-2 virus S protein (Amanat et al., 2020) and thus probably possess no neutralizing antibodies. However, for inducing CD8+ T cell memory, the core requirement is merely that an identical peptide is presented by major histocompatibility complex (MHC) class I (MHC-I) molecules. MHC-I molecules present peptide fragments from intracellular proteins, thus also from viral proteins, at the cell surface for screening by CD8+ cytotoxic T cells (Neefjes et al., 2011). CD8+ T cells recognize the combination of MHC-I molecule with peptide by T cell receptors (TCR) that are unique per T cell clone, and if stimulated these clones can proliferate, kill the presenting (virus-infected) cell, and produce memory cells. MHC-I molecules are polymorphic in that they are represented by many diverse allelic forms that differ between human populations and individuals (Robinson et al., 2020), and mostly bind 9 amino acids (aa) length in their binding groove which is closed at either end (Bjorkman et al., 1987; Rammensee et al., 1995; Schellens et al., 2015).

In the present study, we analyzed whether there are linear 9 aa epitopes that are identical between proteins encoded by SARS-CoV-2 and one or more of the common human coronaviruses. We found many of such epitopes indeed, and, by using prediction software, found that some are expected to bind well to certain MHC-I alleles. We therefore expect that common human coronaviruses can induce some level of CD8+ T cell-mediated immune memory recognizing SARS-CoV-2, and consider the possibility of enhancing that immune memory by vaccination.
Methods
Proteins encoded by a reported genomic sequence for SARS-CoV-2 (GenBank MN908947; Wu et al., 2020) were compared with those for HCoV-OC43 (NC_005147; Vijgen et al., 2005), HCoV-HKU1 (NC_006577; Woo et al., 2005), HCoV-229E (NC_002645; Thiel et al., 2001), and HCoV-NL63 (NC_005831; van der Hoek et al., 2004) by performing BLAST homology searches at the NCBI database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and by making multiple sequence alignments using CLUSTALW software (https://www.genome.jp/tools-bin/clusterw); continuous stretches of 9 aa acids identical between SARS-CoV-2 and one of the other viruses were identified manually. All these shared 9 aa epitopes were screened by ANN 4.0 software at IEDB Analysis Resource (http://tools.immuneepitope.org/mhci/) for prediction of their affinity to a set of representative human MHC-I alleles.

Results and discussion
Table 1 lists the 9 aa epitopes that are identical between proteins encoded by SARS-CoV-2 and one or more of the common human coronaviruses. Many identical >9 aa stretches were found with ORF1ab encoded polyprotein, one such identical stretch (of 12 aa) was found with the N protein of the other two type II coronaviruses HCoV-OC43 and HCoV-HKU1, and no such stretches were found when comparing with any of the other gene products; ORF1ab-derived mature proteins with such stretches, expected from cleavage of the polyprotein precursor (Wu et al., 2020), were the transmembrane protein nonstructural protein 4 (NSP4), 3C-like cysteine protease NSP5, RNA binding protein NSP9, RNA dependent RNA polymerase NSP12, helicase NSP13, 3′-to-5′ exonuclease NSP14, nidoviral endoribonuclease specific for U NSP15, and S-adenosylmethionine-dependent ribose 2′-O-methyltransferase NSP16 (Table 1). Sequence alignment figures of the ORF1ab and N proteins are shown in Extended data (Dijkgraaf, 2020) with highlighting of the interesting epitopes. It is of note that the S protein, which is the prime candidate for inducing neutralizing antibodies (Cohen, 2020), is not suitable for inducing an MHC-I-restricted immune memory across the investigated viral species as between S protein of SARS-CoV-2 and S proteins of the common human coronaviruses there are no 9 aa or even 8 aa matches (not shown).

Table 1 shows that there are >200 linear epitopes of 9 aa that are identical between SARS-CoV-2 and at least one of the common human coronaviruses, most of them with OC43 and HKU1 which, like SARS-CoV-2, belong to the group II coronaviruses. In a simplified model, if people would have been exposed to many of these epitopes through common HCoV infections, this kind of equals immunization by a small intracellular protein under natural viral infection conditions. Whereas live virus is commonly considered the gold standard in regard to inducing strong immunity, unless the virus has some tricks up its sleeve to manipulate the immune system, which for common human coronaviruses is not well investigated, a research grant proposal suggesting this as a vaccination strategy would probably fail. Reviewers of such proposal would rightfully point out that the strategy would not induce neutralizing antibodies, which for combating some viral infections can be very important, and that for inducing MHC-I-restricted cell-mediated cytotoxicity memory, ideally, a much larger protein or more proteins should be taken. Those reviewers would conclude that for such small intracellular protein to induce strong immune memory it would need too much luck in regard to immunogenicity and it would be too dependent on the MHC alleles of the immunized person as different alleles bind different peptides. Nevertheless, those reviewers would probably also agree that in most persons thus vaccinated some (small) level of immune memory protection would be established, even with such small non-surface protein (e.g. Polakos et al., 2001). Regardless of that this obviously is not the ideal way to induce a population-wide strong protective immunity (see the spread of COVID-19), together with other factors such as health and the number of encountered viruses (the strength of the viral challenge), the induced immune memory could make a difference for whether a person gets sick; at the population scale, it so may somewhat reduce the virus reproduction number. Importantly, by stimulating this HCoV-derived MHC-I restricted immune memory by vaccination (see below), it could become a more significant helper in fighting COVID-19.

Software predictions of MHC-I-binding epitopes
Based on combinations of experimental results and computer learning, various software has been created that with some degree of reliability can predict how efficiently peptides can bind to the grooves of various MHC-I alleles. In the present study we used the artificial neural network (ANN) function (Lundegaard et al., 2008) of the IEDB Analysis Resource (http://tools.immuneepitope.org/mhci/) (Dhanda et al., 2019) which may achieve >75% reliability for predicting binding (Lundegaard et al., 2008). The software owners state that IC50 values of <50 nM and <500 nM are considered high and intermediate affinity, respectively, and are found for most epitopes known to stimulate cytotoxic T cells. Therefore, Table 1 only indicates the predicted IC50 values if lower than 500 nM. Table 1 shows these expected affinities for twelve MHC-I alleles that are rather representative for sets of MHC-I alleles with similar binding properties (supertypes) and so represent a large part of the human MHC-I binding repertoire (Lund et al., 2004): HLA-A*0101 (supertype A1), HLA-A*0201 (A2), HLA-A*0301 (A3), HLA-A*2402 (A24), HLA-A*2601 (A26), HLA-B*0702 (B7), HLA-B*0801 (B8), HLA-B*1501 (B62), HLA-B*2705 (B27), HLA-B*3901 (B39), HLA-B*4001 (B44), and HLA-B*5801 (B58). It is of note that Li et al. (2008) found that a SARS-CoV-1 15 aa peptide sequence (their “Replicase 4701-4715” peptide) encompassing the SARS-CoV-2/HCoV-shared ORF1ab 4725 and ORF1ab 4726 epitopes that are predicted to bind well to the MHC-I alleles HLA-A*0201 and HLA-B*3901 (see our Table 1) was associated with a CDS+ T cell response against SARS-CoV-1 in humans. However, Li et al. (2008) also found such CDS+ T cell response associated with a SARS-CoV-1 15 aa peptide (their “Nucleocapsid 106-120” peptide) encompassing the SARS-CoV-2/HCoV-shared N 106, N 107, N 108, an N 109 epitopes for which our analyses did not predict MHC-I binding (see our Table 1).

The MHC-I binding affinity is considered the most selective in determining which peptides are presented, but also steps in the
Table 1. Stretches of 9 consecutive amino acids that are identical between SARS-CoV-2 and at least one of the common human coronaviruses.

| Common human coronaviruses | Mature protein | IC50 prediction by ANN 4.0 program of EDB software (only IC50 values <500 nM are shown) |
|---------------------------|----------------|----------------------------------------------------------------------------------|
| OC43                      | HLA-A          | *0101 0201 0301 0402 0501 0601 0701 0801 0901 1001 1101 1201 1301 1401 1501 1601 1701 1801 1901 2001 2101 2201 2301 2401 2501 2601 2701 2801 2901 3001 3101 3201 3301 3401 3501 3601 3701 3801 3901 4001 4101 4201 4301 4401 4501 4601 4701 4801 4901 5001 5101 5201 5301 5401 5501 5601 5701 5801 5901 6001 6101 6201 6301 6401 6501 6601 6701 6801 6901 7001 7101 7201 7301 7401 7501 7601 7701 7801 7901 8001 8101 8201 8301 8401 8501 8601 8701 8801 8901 9001 9101 9201 9301 9401 9501 9601 9701 9801 9901 1001 |
| HKU1                      | HLA-A          | *0101 0201 0301 0402 0501 0601 0701 0801 0901 1001 1101 1201 1301 1401 1501 1601 1701 1801 1901 2001 2101 2201 2301 2401 2501 2601 2701 2801 2901 3001 3101 3201 3301 3401 3501 3601 3701 3801 3901 4001 4101 4201 4301 4401 4501 4601 4701 4801 4901 5001 5101 5201 5301 5401 5501 5601 5701 5801 5901 6001 6101 6201 6301 6401 6501 6601 6701 6801 6901 7001 7101 7201 7301 7401 7501 7601 7701 7801 7901 8001 8101 8201 8301 8401 8501 8601 8701 8801 8901 9001 9101 9201 9301 9401 9501 9601 9701 9801 9901 1001 |
| 229E                      | HLA-A          | *0101 0201 0301 0402 0501 0601 0701 0801 0901 1001 1101 1201 1301 1401 1501 1601 1701 1801 1901 2001 2101 2201 2301 2401 2501 2601 2701 2801 2901 3001 3101 3201 3301 3401 3501 3601 3701 3801 3901 4001 4101 4201 4301 4401 4501 4601 4701 4801 4901 5001 5101 5201 5301 5401 5501 5601 5701 5801 5901 6001 6101 6201 6301 6401 6501 6601 6701 6801 6901 7001 7101 7201 7301 7401 7501 7601 7701 7801 7901 8001 8101 8201 8301 8401 8501 8601 8701 8801 8901 9001 9101 9201 9301 9401 9501 9601 9701 9801 9901 1001 |
| NL63                      | HLA-A          | *0101 0201 0301 0402 0501 0601 0701 0801 0901 1001 1101 1201 1301 1401 1501 1601 1701 1801 1901 2001 2101 2201 2301 2401 2501 2601 2701 2801 2901 3001 3101 3201 3301 3401 3501 3601 3701 3801 3901 4001 4101 4201 4301 4401 4501 4601 4701 4801 4901 5001 5101 5201 5301 5401 5501 5601 5701 5801 5901 6001 6101 6201 6301 6401 6501 6601 6701 6801 6901 7001 7101 7201 7301 7401 7501 7601 7701 7801 7901 8001 8101 8201 8301 8401 8501 8601 8701 8801 8901 9001 9101 9201 9301 9401 9501 9601 9701 9801 9901 1001 |

- ORF1ab protein is only a precursor polyprotein, and the column Mature protein indicates the probable mature protein that possesses the epitope: 3CLpro, 3C-like cysteine protease; RdRp, RNA dependent RNA polymerase; Hel, helicase; ExoN, 3'-to-5' exonuclease; NendoU, nidoviral endoribonuclease specific for U; O-MT, S-adenosylmethionine-dependent ribose 2'-O-methyltransferase. Yellow blocks indicate the presence of identical sequences in the respective common human coronaviruses, and gray blocks indicate the absence of such matches. Orange blocks highlight those peptides with predicted IC50 values of <500 nM for one of the twelve investigated MHC-I alleles.
| SARS-CoV-2 source | Mature | Group II | Group I | HLA-A | HLA-B |
|-------------------|--------|----------|---------|-------|-------|
| ORF1ab 4800 | NSP12 RdRp | QTVKRGNFN | OC43 | 229E | NL63 | *0101 | *0201 | *0301 | *2402 | *2601 | *0702 | *0801 | *1501 | *2705 | *3901 | *4001 | *5801 |
| ORF1ab 4931 | NSP12 RdRp | TQMNLYKAI | | | | | | | | | | | | | | | |
| ORF1ab 4932 | NSP12 RdRp | QMNLYKAYS | | | | | | | | | | | | | | | |
| ORF1ab 4933 | NSP12 RdRp | MNLKYAISA | | | | | | | | | | | | | | | |
| ORF1ab 4934 | NSP12 RdRp | NLKAIASAK | | | | | | | | | | | | | | | 198 nM |
| ORF1ab 4935 | NSP12 RdRp | LKYAIASKN | | | | | | | | | | | | | | | |
| ORF1ab 4936 | NSP12 RdRp | KYAIASKNR | | | | | | | | | | | | | | | |
| ORF1ab 4937 | NSP12 RdRp | YAIAKNNRA | | | | | | | | | | | | | | | |
| ORF1ab 4938 | NSP12 RdRp | AISAKNNRA | | | | | | | | | | | | | | | |
| ORF1ab 4939 | NSP12 RdRp | ISAKNRART | | | | | | | | | | | | | | | |
| ORF1ab 4940 | NSP12 RdRp | SAKNRARTV | | | | | | | | | | | | | | | 58 nM |
| ORF1ab 4941 | NSP12 RdRp | AKNRARTVA | | | | | | | | | | | | | | | |
| ORF1ab 4942 | NSP12 RdRp | KNARATVAG | | | | | | | | | | | | | | | |
| ORF1ab 4943 | NSP12 RdRp | NRARTVAGV | | | | | | | | | | | | | | | 468 nM |
| ORF1ab 4944 | NSP12 RdRp | RARTVAGVS | | | | | | | | | | | | | | | |
| ORF1ab 4945 | NSP12 RdRp | ARTVAGVSI | | | | | | | | | | | | | | | 219 nM |
| ORF1ab 5006 | NSP12 RdRp | LMGWDYPKC | | | | | | | | | | | | | | | |
| ORF1ab 5007 | NSP12 RdRp | MGWDYPKCD | | | | | | | | | | | | | | | |
| ORF1ab 5008 | NSP12 RdRp | GWDYPKCDR | | | | | | | | | | | | | | | |
| ORF1ab 5009 | NSP12 RdRp | WDYPKCDRA | | | | | | | | | | | | | | | |
| ORF1ab 5010 | NSP12 RdRp | DYPKCDRAM | | | | | | | | | | | | | | | |
| ORF1ab 5011 | NSP12 RdRp | YPKCDRAMP | | | | | | | | | | | | | | | |
| ORF1ab 5012 | NSP12 RdRp | PKCDRAMPN | | | | | | | | | | | | | | | |
| ORF1ab 5043 | NSP12 RdRp | FYRLANECA | | | | | | | | | | | | | | | |
| ORF1ab 5044 | NSP12 RdRp | FYRLANECQ | | | | | | | | | | | | | | | |
| ORF1ab 5045 | NSP12 RdRp | YRLANECAQ | | | | | | | | | | | | | | | |
| ORF1ab 5046 | NSP12 RdRp | RLANECAQVL | | | | | | | | | | | | | | | 93 nM |
| ORF1ab 5047 | NSP12 RdRp | LANECAQVLS | | | | | | | | | | | | | | | |
| ORF1ab 5048 | NSP12 RdRp | ANECAVLSS | | | | | | | | | | | | | | | |
| ORF1ab 5049 | NSP12 RdRp | NECAVLSSE | | | | | | | | | | | | | | | |
| ORF1ab 5066 | NSP12 RdRp | YVKPGTSSS | | | | | | | | | | | | | | | |
| ORF1ab 5067 | NSP12 RdRp | VKPGTSSSG | | | | | | | | | | | | | | | |
| SARS-CoV-2 source | Mature protein | Sequence (9aa) | Group II | Group I | HLA-A | HLA-B |
|-------------------|----------------|----------------|----------|---------|-------|-------|
| SARS-CoV-2        | NSP12 RdRp     | KPGGTSSGD      | OC43     | 229E    | NL63  |       |
| ORF1ab 5068       | NSP12 RdRp     | PGSTSSGDA      | HKU1     | *0101   | *0201 |       |
| ORF1ab 5069       | NSP12 RdRp     | GGTTSSGDA      | 29 nM    | *0301   | *2402 |       |
| ORF1ab 5070       | NSP12 RdRp     | GTSSGDAAT      |         | *2601   | *0702 |       |
| ORF1ab 5071       | NSP12 RdRp     | GTSSGDATT      |         |         | *0801 |       |
| ORF1ab 5072       | NSP12 RdRp     | TSSGDATTA      |         |         | *1501 |       |
| ORF1ab 5074       | NSP12 RdRp     | SGDATTAYA      |         |         | *2705 |       |
| ORF1ab 5075       | NSP12 RdRp     | GDATTAYAN      |         |         | *3901 |       |
| ORF1ab 5076       | NSP12 RdRp     | DATTAYANS      |         |         | *4001 |       |
| ORF1ab 5077       | NSP12 RdRp     | ATTAYANSV      |         |         | *5801 |       |
| ORF1ab 5078       | NSP12 RdRp     | TTAIANSVF      |         |         |       | 245 nM|
| ORF1ab 5079       | NSP12 RdRp     | TAYANSVEN      |         |         |       | 29 nM|
| ORF1ab 5080       | NSP12 RdRp     | AYANSVNI       |         |         |       | 56 nM|
| ORF1ab 5082       | NSP12 RdRp     | ANSVFNCQA      |         |         |       |       |
| ORF1ab 5083       | NSP12 RdRp     | NSVFNICQA      |         |         |       |       |
| ORF1ab 5084       | NSP12 RdRp     | SVFNCQAV       |         |         |       | 85 nM|
| ORF1ab 5085       | NSP12 RdRp     | VFNICQAVT      |         |         |       |       |
| ORF1ab 5086       | NSP12 RdRp     | FNICQAVTA      |         |         |       |       |
| ORF1ab 5087       | NSP12 RdRp     | NICQAVTAN      |         |         |       |       |
| ORF1ab 5088       | NSP12 RdRp     | ICQAVTANV      |         |         |       |       |
| ORF1ab 5140       | NSP12 RdRp     | YLRKHFSEM      |         |         |       | 128 nM|
| ORF1ab 5141       | NSP12 RdRp     | LRKHFSEM      |         |         |       | 134 nM|
| ORF1ab 5142       | NSP12 RdRp     | RKHFSEMIL      |         |         |       | 4 nM   |
| ORF1ab 5143       | NSP12 RdRp     | KHFSEMILS      |         |         |       | 47 nM  |
| ORF1ab 5144       | NSP12 RdRp     | HFSMILSD       |         |         |       | 313 nM|
| ORF1ab 5145       | NSP12 RdRp     | FSMILSDD       |         |         |       | 310 nM|
| ORF1ab 5177       | NSP12 RdRp     | VLYQQNNFV      |         |         |       | 25 nM |
| ORF1ab 5178       | NSP12 RdRp     | LYQQNNVFM      |         |         |       |       |
| ORF1ab 5179       | NSP12 RdRp     | YYQNNVEMS      |         |         |       |       |
| ORF1ab 5180       | NSP12 RdRp     | YQQNPFMSE      |         |         |       |       |
| ORF1ab 5196       | NSP12 RdRp     | DLTKGPHEF      |         |         |       |       |
| ORF1ab 5197       | NSP12 RdRp     | LTGPHEF        |         |         |       |       |
| SARS-CoV-2 source | Mature Group II | Mature Group I | HLA-A | HLA-B |
|-------------------|----------------|----------------|-------|-------|
| ORF1ab 5198       | TKGHEFCS       |                |       |       |
| ORF1ab 5199       | RGGHEFCSQ      |                |       |       |
| ORF1ab 5200       | GPHHEFCSQH     |                |       |       |
| ORF1ab 5201       | PHHEFCSQHT     |                |       |       |
| ORF1ab 5202       | MEFCSQHTMLM    |                |       |       |
| ORF1ab 5203       | EFCSQHTML      |                |       |       |
| ORF1ab 5204       | FCSQHTMLV      |                |       |       |
| ORF1ab 5205       | CSQHTMLV       |                |       |       |
| ORF1ab 5217       | DYVYLPYPD      |                |       |       |
| ORF1ab 5218       | YVYLPYPD      |                |       |       |
| ORF1ab 5219       | VYLPYPDPS      |                |       |       |
| ORF1ab 5220       | LYPYPDPSR      |                |       |       |
| ORF1ab 5221       | LPYPDPSRI      |                |       |       |
| ORF1ab 5222       | YPYDPSRL      |                |       |       |
| ORF1ab 5223       | YPDPDSRL      |                |       |       |
| ORF1ab 5224       | PDPDSRILG     |                |       |       |
| ORF1ab 5225       | DPSRILGAG     |                |       |       |
| ORF1ab 5226       | PSRILGAGC     |                |       |       |
| ORF1ab 5227       | SRILGAGCF     |                |       |       |
| ORF1ab 5228       | RILGAGCFV     |                |       |       |
| ORF1ab 5229       | ILGAGCFVD     |                |       |       |
| ORF1ab 5230       | LGAGCFVDD     |                |       |       |
| ORF1ab 5248       | IERFPSLAI      |                |       |       |
| ORF1ab 5249       | EFVPSLAID      |                |       |       |
| ORF1ab 5250       | EFVPSLAIDA     |                |       |       |
| ORF1ab 5251       | PVSLAIDAY     |                |       |       |
| ORF1ab 5252       | VSLAIDAIP     |                |       |       |
| ORF1ab 5253       | SLAIDAYPL     |                |       |       |
| ORF1ab 5349       | NSP13 Hel      |                |       |       |
| ORF1ab 5350       | CCKCCYDHV     |                |       |       |
| SARS-CoV-2 source | protein | Sequence (9aa) | Group II | Group I | HLA-A | HLA-B |
|-------------------|---------|----------------|----------|---------|-------|-------|
| Mature | | | OC43 | HKU1 | 229E | NL63 | *0101 | *0201 | *0301 | *2402 | *2601 | *0702 | *0801 | *1501 | *2705 | *3901 | *4001 | *5801 |
| ORF1ab 5371 | NSP13 Hel | PYVCNAPGC | | | | | | | | | | | | | | | | | |
| ORF1ab 5372 | NSP13 Hel | YVCAAPGCD | | | | | | | | | | | | | | | | | |
| ORF1ab 5373 | NSP13 Hel | VCNAAPGCDV | | | | | | | | | | | | | | | | | |
| ORF1ab 5377 | NSP13 Hel | LYGGMYY | | | | | | | | | | | | | | | | | |
| ORF1ab 5378 | NSP13 Hel | YVGMMSYVC | | | | | | | | 89 nM | | | | | | | | | |
| ORF1ab 5388 | NSP13 Hel | CTERKLFA | | | | | | | | 303 nM | | | | | | | | | |
| ORF1ab 5450 | NSP13 Hel | TRLKLFAA | | | | | | | | | | | | | | | | | |
| ORF1ab 5451 | NSP13 Hel | ERLKLFAAR | | | | | | | | | | | | | | | | | |
| ORF1ab 5452 | NSP13 Hel | RLKLFAAET | | | | | | | | | | | | | | | | | |
| ORF1ab 5559 | NSP13 Hel | LSAPTLVPQ | | | | | | | | | | | | | | | | | |
| ORF1ab 5560 | NSP13 Hel | SAPTLVPQE | | | | | | | | | | | | | | | | | |
| ORF1ab 5565 | NSP13 Hel | QGPPGTGKS | | | | | | | | | | | | | | | | | |
| ORF1ab 5566 | NSP13 Hel | GPPGTGKSH | | | | | | | | | | | | | | | | | |
| ORF1ab 5567 | NSP13 Hel | SHAADALC | | | | | | | | | | | | | | | | | |
| ORF1ab 5568 | NSP13 Hel | HAADALCE | | | | | | | | | | | | | | | | | |
| ORF1ab 5569 | NSP13 Hel | AAVDALCEK | | | | | | | | | | | | | | | | | |
| ORF1ab 5570 | NSP13 Hel | AVDALCEKA | | | | | | | | | | | | | | | | | |
| ORF1ab 5571 | NSP13 Hel | RIIPARARV | | | | | | | | | | | | | | | | | |
| ORF1ab 5572 | NSP13 Hel | IIPARARVE | | | | | | | | | | | | | | | | | |
| ORF1ab 5573 | NSP13 Hel | IPARARVBC | | | | | | | | 213 nM | | | | | | | | | |
| ORF1ab 5574 | NSP13 Hel | RAKHYVYIG | | | | | | | | | | | | | | | | | |
| ORF1ab 5575 | NSP13 Hel | AKHYYVYGD | | | | | | | | | | | | | | | | | |
| ORF1ab 5576 | NSP13 Hel | KYHYYIGDP | | | | | | | | | | | | | | | | | |
| ORF1ab 5577 | NSP13 Hel | HYYYIGDPA | | | | | | | | | | | | | | | | | |
| ORF1ab 5578 | NSP13 Hel | YYYIGDPAQ | | | | | | | | | | | | | | | | | |
| ORF1ab 5579 | NSP13 Hel | YVIGDPAQL | | | | | | | | 206 nM | | | | | | | | | |
| ORF1ab 5580 | NSP13 Hel | YIGDPAQLP | | | | | | | | | | | | | | | | | |
| ORF1ab 5581 | NSP13 Hel | IGDPAPLA | | | | | | | | | | | | | | | | | |
| ORF1ab 5582 | NSP13 Hel | GDPAQLPAP | | | | | | | | | | | | | | | | | |
| ORF1ab 5583 | NSP13 Hel | DPALPA | | | | | | | | | | | | | | | | | |
| ORF1ab 5584 | NSP13 Hel | LIVDTSAD | | | | | | | | 16 nM | | | | | | | | 298 nM | |
| ORF1ab 5585 | NSP13 Hel | LIVDSVSAL | | | | | | | | | | | | | | | | | |
| SARS-CoV-2 source | Mature Group I | IC50 prediction by ANN 4.0 program of IEDB software (only IC50 values of <500 nM are shown) |
|-------------------|-----------------|-------------------------------------------------------------------------------------|
| Group II | Group I | HLA-A | HLA-B |
| Protein | Sequence (9aa) | OC43 | HLA-A | *0101 | *0201 | *0301 | *2402 | *2601 | *0702 | *0801 | *1501 | *2705 | *3901 | *4001 | *5801 |
| NSP13 Hel | ORF1ab 5772 | IVDTVSALV | 114 nM |
| NSP13 Hel | ORF1ab 5773 | VDVTSALVY | 114 nM |
| NSP13 Hel | ORF1ab 5775 | TVSALVVDN | 114 nM |
| NSP13 Hel | ORF1ab 5776 | VSALVYDNK | 114 nM |
| NSP13 Hel | ORF1ab 5777 | SALVYDNKL | 114 nM |
| NSP13 Hel | ORF1ab 5778 | ALVYDNKLK | 114 nM |
| NSP13 Hel | ORF1ab 5779 | LVYDNKLKA | 114 nM |
| NSP13 Hel | ORF1ab 5832 | KAEISYPYN | 114 nM |
| NSP13 Hel | ORF1ab 5833 | AVFISYPYN | 114 nM |
| NSP13 Hel | ORF1ab 5834 | VFIISYPNQ | 114 nM |
| NSP13 Hel | ORF1ab 5835 | FISYPNQN | 114 nM |
| NSP13 Hel | ORF1ab 5855 | QTVDSSQGS | 114 nM |
| NSP13 Hel | ORF1ab 5856 | TVDSSQGER | 114 nM |
| NSP13 Hel | ORF1ab 5857 | VDSSQGEY | 114 nM |
| NSP13 Hel | ORF1ab 5858 | DSSQGEYD | 114 nM |
| NSP13 Hel | ORF1ab 5859 | SQGSEYDY | 114 nM |
| NSP13 Hel | ORF1ab 5860 | SQGSEYDV | 114 nM |
| NSP13 Hel | ORF1ab 5861 | QGSEYDYVI | 114 nM |
| NSP13 Hel | ORF1ab 5862 | GSEYDYVIF | 114 nM |
| NSP13 Hel | ORF1ab 5863 | CNNRFNVA | 114 nM |
| NSP13 Hel | ORF1ab 5864 | RNNRFNVAI | 114 nM |
| NSP13 Hel | ORF1ab 5865 | NNTRFNVAIT | 114 nM |
| NSP13 Hel | ORF1ab 5866 | NRFNVAITR | 114 nM |
| NSP13 Hel | ORF1ab 5867 | RNFNVAITRA | 114 nM |
| NSP13 Hel | ORF1ab 5868 | FNWAIKRAK | 114 nM |
| NSP14 ExoN | ORF1ab 6031 | PLQGFSTG | 24 nM |
| NSP14 ExoN | ORF1ab 6198 | DAIMTRCLA | 24 nM |
| NSP14 ExoN | ORF1ab 6199 | AIMTRCLAV | 29 nM |
| NSP14 ExoN | ORF1ab 6307 | CLFWNCNVD | 29 nM |
| NSP14 ExoN | ORF1ab 6320 | NSIVCPRDF | 29 nM |
| NSP14 ExoN | ORF1ab 6321 | SIVCPRDFTR | 29 nM |
| NSP14 ExoN | ORF1ab 6322 | IVCRFDTRV | 29 nM |
| SARS-CoV-2 source | Mature | Group II | Group I | HLA-A | HLA-B |
|-------------------|--------|----------|---------|-------|-------|
| ORF1ab 6323       | NP14 ExoN | VCFRTDRVL | OC43 | *0101 |       |
| ORF1ab 6341       | NP14 ExoN | GGSLYVNKH | HKU1 | *0201 |       |
| ORF1ab 6342       | NP14 ExoN | GSLYVNKHA | NL63 | *0301 |       |
| ORF1ab 6343       | NP14 ExoN | SLYVNKHAF |       | *2402 |       |
| ORF1ab 6344       | NP14 ExoN | LYNKHFH |       | *2601 |       |
| ORF1ab 6345       | NP14 ExoN | YVNHAFHT |       | *0702 |       |
| ORF1ab 6346       | NP14 ExoN | VNKHAFHTP |       | *0801 |       |
| ORF1ab 6347       | NP14 ExoN | NKHAFHTPA |       | *1501 |       |
| ORF1ab 6348       | NP14 ExoN | DYVFLKT |       | *2705 |       |
| ORF1ab 6389       | NP14 ExoN | YVPLKSAT |       | *3901 |       |
| ORF1ab 6390       | NP14 ExoN | YVPLKSATC |       | *4001 |       |
| ORF1ab 6393       | NP14 ExoN | LKSATCITR |       | *5801 |       |
| ORF1ab 6394       | NP14 ExoN | KSATCITRC |       |       | 331 nM |
| ORF1ab 6395       | NP14 ExoN | SATCITRCN |       |       |       |
| ORF1ab 6396       | NP14 ExoN | ATCITRCNL |       |       |       |
| ORF1ab 6397       | NP14 ExoN | TCITRCNLG |       |       |       |
| ORF1ab 6398       | NP14 ExoN | CITRCNLGG |       |       |       |
| ORF1ab 6399       | NP14 ExoN | ITRCNLGGA |       |       |       |
| ORF1ab 6400       | NP14 ExoN | TRCNLGGAV |       |       |       |
| ORF1ab 6401       | NP14 ExoN | RCNLAGAVC |       |       |       |
| ORF1ab 6682       | NP15 NendoU | YAFEHIVYG |       |       |       |
| ORF1ab 6698       | NP15 NendoU | GGLHILIGL |       |       |       |
| ORF1ab 6746       | NP15 NendoU | VIDLDDDFV |       |       |       |
| ORF1ab 6747       | NP15 NendoU | IDLLDDDFV |       |       |       |
| ORF1ab 6839       | NP16 O-MT | MNVAKYTTQ |       |       |       |
| ORF1ab 6840       | NP16 O-MT | MNVAKYQLC |       |       | 259 nM |
| ORF1ab 6841       | NP16 O-MT | NVAKYQLC |       |       |       |
| ORF1ab 6842       | NP16 O-MT | VAKYQLCQ |       |       |       |
| ORF1ab 6843       | NP16 O-MT | AKYQLCQY |       |       |       |
| ORF1ab 6844       | NP16 O-MT | KYQLCQYL |       |       | 139 nM |
| SARS-CoV-2 source | Mature         | Group II | Group I | HLA-A | HLA-B |
|-------------------|----------------|----------|---------|-------|-------|
| ORF 1ab 6845      | NSP16 O-MT    | YTLQCQYLN|         |       |       |
| ORF 1ab 6846      | NSP16 O-MT    | TQLCQYLN |         |       |       |
| ORF 1ab 6869      | NSP16 O-MT    | GAGSDKGVA|         |       |       |
| ORF 1ab 6870      | NSP16 O-MT    | AGSDKGVAP|         |       |       |
| ORF 1ab 6871      | NSP16 O-MT    | GSDKGVAPG|         |       |       |
| ORF 1ab 6872      | NSP16 O-MT    | SDKGVAPOT|         |       |       |
| ORF 1ab 6922      | NSP16 O-MT    | WDLIIIDMY|         |       |       |
| ORF 1ab 6923      | NSP16 O-MT    | DLIISIMYD|         |       |       |
| ORF 1ab 6924      | NSP16 O-MT    | LIISDMYP  |         |       |       |
| ORF 1ab 6943      | NSP16 O-MT    | SKEGFFTYI|         |       |       |
| ORF 1ab 6958      | NSP16 O-MT    | KLAGGSSVA |         |       |       |
| ORF 1ab 6959      | NSP16 O-MT    | LALGGSSVAI|      |       |       |
| ORF 1ab 6960      | NSP16 O-MT    | ALGSVAIK  |         |       | 110 nM|
| ORF 1ab 6961      | NSP16 O-MT    | LGSSVAIKI |         |       |       |
| ORF 1ab 6962      | NSP16 O-MT    | GGSVAIKIT |         |       |       |
| ORF 1ab 6963      | NSP16 O-MT    | GSVAIKITE |         |       |       |
| ORF 1ab 6973      | NSP16 O-MT    | SWNADLYKL |         |       |       |
| ORF 1ab 6974      | NSP16 O-MT    | WNDLKYLM  |         |       |       |
| ORF 1ab 6993      | NSP16 O-MT    | TNVNASSE  |         |       |       |
| ORF 1ab 6998      | NSP16 O-MT    | SSSEAFLLG |         |       |       |
| ORF 1ab 7024      | NSP16 O-MT    | HANYIFWRN |         |       |       |
| N 106             | Nucleocapsid  | FRWYFYYLG|         |       |       |
| N 107             | Nucleocapsid  | RWYFYLYTG |         |       |       |
| N 108             | Nucleocapsid  | WYYFYLGTG |         |       |       |
| N 109             | Nucleocapsid  | YFYYLGTGP |         |       |       |
peptide processing and loading pathways may play selective roles which are difficult to capture in prediction software (Nielsen et al., 2005). We argue that, if such steps would be selective for presentation, in most cases they would probably not differentiate between the 9 aa epitope in the SARS-CoV-2 context versus the respective HCoV context, since most of those epitopes are within stretches that also show many similarities in the neighboring residues (Extended data).

Not all stable complexes of MHC-I with non-self peptides elicit a strong immune response, but “immunogenicity” features are hard to predict with meaningful reliability by in silico analysis (Calis et al., 2013), and in the present study we refrain from such predictions. Table 1 should, foremost, be understood as evidence of principle and a list of promising peptides, whereas only future experiments can prove MHC-I-mediated immune memory involving these or other peptides.

In regard to SARS-CoV-2 recognition, the common human coronaviruses may also induce some MHC-II-mediated immune memory by CD4+ helper T cells (for shared epitope use by different coronaviruses see Zhao et al., 2016). CD4+ helper T cells can help stimulate cells involved in antibody or cell-mediated cytotoxic immune responses (Neeffes et al., 2011). However, for this topic we refrained from detailed (software) predictions because comparison of MHC-II epitopes across different viruses is harder than for MHC-I epitopes. Namely, although the core of MHC-II bound peptides is also only 9 aa, the surrounding amino acids are also part of the bound peptide that tends to be 12-25 aa (Brown et al., 1993; Rammensee et al., 1995; Stern & Wiley, 1994) and can affect how the peptide interacts with the receptors on the CD4+ helper T cells (Arnold et al., 2002).

Vaccination potential
Immune memory means that a secondary immune response, upon renewed encounter with the same pathogen, is faster and stronger than the primary immune response during the first encounter with the pathogen. This is based on expansion of specific B and T cell clones, which specifically recognize pathogen-(derived) epitopes, with some of those cells becoming memory cells (Paul, 2013). This principle also causes that for a booster vaccination/immunization the requirements for efficiently inducing an immune response are lower than for a first vaccination/immunization (e.g. Goding, 1996). Especially in elderly people, who have a decreased ability to mount adaptive immune responses against new antigens, vaccination that stimulates an immune memory response may be beneficial (Reber et al., 2012). As discussed above, people’s past infections with common coronaviruses probably did not induce a B cell memory for making antibodies that can neutralize SARS-CoV-2. However, as the current study shows by analysis of linear 9 aa epitopes, these common human coronaviruses are expected to induce CD8+ T cells that may potentially kill SARS-CoV-2-infected cells and so can help eradicate the virus. There are several possible ways to exploit this probable immune memory. For example, if using RNA for immunization (Cohen, 2020), it may be best to also include SARS-CoV-2 genes that encode MHC-I epitopes that match those of the common coronaviruses. Alternatively, delivery of these epitopes to the MHC-I presentation system may be tried by peptide or protein based vaccines (e.g. Kohyama et al., 2009; Slingluff, 2011; van Montfoort et al., 2014; Yadav et al., 2014), possibly in combination with some of the strategies that are currently being explored for non-specific stimulation of the immune system against COVID-19 (Kupferschmidt & Cohen, 2020). Protein (-coding) vaccines, for example encompassing a large part of the SARS-CoV-2 ORF1ab product, would have an advantage over peptide-vaccines by including multiple possible MHC-I and also MHC-II epitopes, and be less dependent on MHC-allele matching and the quality of software predictions. Naturally, as for any new vaccine strategy, it should be carefully assessed whether the benefits of the induced type of immunity outweigh the potential deleterious health effects caused by an increased inflammation response (Cohen, 2020; Weingart et al., 2004). Important questions are also whether the history of previous—especially recent—infections with common coronaviruses, or people’s MHC alleles, affect people’s resistance to SARS-CoV-2. Most definitely, if discussing possible strategies for vaccination against SARS-CoV-2, pre-existing MHC-I-based immunity derived from previous infections with common coronaviruses should be part of the consideration.

Notification
Although we were not aware of this at the time of writing, a recent preprint appeared with overlapping contents (Nguyen et al., 2020).

Data availability
Underlying data
Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome, Accession number MN908947: https://www.ncbi.nlm.nih.gov/nuccore/MN908947

Human coronavirus OC43, complete genome, Accession number NC_005147.1: https://www.ncbi.nlm.nih.gov/nuccore/NC_005147.1?report=genbank

Human coronavirus HKU1, complete genome, Accession number NC_006577: https://www.ncbi.nlm.nih.gov/nuccore/NC_006577

Human coronavirus 229E, complete genome, Accession number NC_002645: https://www.ncbi.nlm.nih.gov/nuccore/NC_002645

Human Coronavirus NL63, complete genome, Accession number NC_005831: https://www.ncbi.nlm.nih.gov/nuccore/NC_005831

Extended data
Harvard Dataverse: Extended data. Sequence alignments of SARS-CoV-2 ORF1ab and N proteins with their counterparts in the common human coronaviruses, https://doi.org/10.7910/DVN/CNPUPA (Dijkstra, 2020).

Data are available under the terms of the Creative Commons Zero “No rights reserved” data waiver (CC0 1.0 Public domain dedication).
Woo PC, Lau SK, Chu CM, et al.: Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J Virol. 2005; 79(2): 884–95. PubMed Abstract | Publisher Full Text | Free Full Text

Wu F, Zhao S, Yu B, et al.: A new coronavirus associated with human respiratory disease in China. Nature. 2020; 579(7798): 265–269. PubMed Abstract | Publisher Full Text | Free Full Text

Yadav M, Jhunjhunwala S, Phung QT, et al.: Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. Nature. 2014; 515(7528): 572–6. PubMed Abstract | Publisher Full Text

Zhao J, Zhao J, Mangalam AK, et al.: Airway Memory CD4+ T Cells Mediate Protective Immunity against Emerging Respiratory Coronaviruses. Immunity. 2016; 44(6): 1379–91. PubMed Abstract | Publisher Full Text | Free Full Text

Zhou W, Wang W, Wang H, et al.: First infection by all four non-severe acute respiratory syndrome human coronaviruses takes place during childhood. BMC Infect Dis. 2013; 13: 433. PubMed Abstract | Publisher Full Text | Free Full Text

Zhou P, Yang XL, Wang XG, et al.: A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020; 579(7798): 270–273. PubMed Abstract | Publisher Full Text | Free Full Text
Open Peer Review

Current Peer Review Status: ✔ ✔

Version 1

Reviewer Report 22 June 2020

https://doi.org/10.5256/f1000research.25889.r63246

© 2020 Sant A. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Andrea Sant

1 Department of Microbiology and Immunology, University of Rochester Medical Center, Rochester, NY, 14642, USA
2 David H. Smith Center for Vaccine Biology and Immunology, Rochester, NY, USA

These analyses of potential cross reactive CD8 T cell epitopes between the current SARS-CoV-2 and “seasonal” endemic human CoV is useful and timely and the discussion is balanced.

There are several modifications that I believe would improve the clarify and value of the manuscript.

Based on the first sentence of the paragraph entitled “the possibility of matching linear epitopes…”, the authors state that the two major arms of immune memory…are antibody and CD8 T cells” I believe this is incorrect, as CD4 T cells can directly impact lung pathology and contribute to both protective and pathological immune responses. In fact a recent paper uploaded to BioRxiv suggested that it was populations in the CD4 T cell compartments that correlated with disease severity. The authors should acknowledge that all three subsets of the adaptive response (B cells, CD8 and CD4 T cells) are likely to be important, but this manuscript focusses on CD8 epitopes.

The authors refer to the “software owners” when describing cutoffs. They are perhaps better described as software “designers”.

When discussing “Vaccine Potential”, the authors state that the secondary response is “faster and stronger”. This should be more accurately described, with some references, in a way that points out the higher frequency of responding cells during memory recall, and lower thresholds of TcR engagement needed for T cell activation, both qualities that contribute to a competitive advantage of memory cells.

Because the nature of CD8 memory to the different antigens screened by the authors is not known, the epitopes identified may or may not be targets of cross reactive memory recall. Therefore, the word “expected” should be substituted for “Potential” or some other word that indicates that the epitope list includes candidates but not expected epitopes.

I think the Table could be made quite a lot smaller and thus more valuable to the reader. The source
proteins could be indicated as an abbreviation provided in the legend as could the various seasonal strains. The boxes could then be quite small, and either be positive or negative. In any case, an effort should be made to condense this table.

Is the work clearly and accurately presented and does it cite the current literature?
Yes

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Partly

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** Immunology

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewed Report 22 June 2020

https://doi.org/10.5256/f1000research.25889.r63655

© 2020 Selin L et al. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Anna Gil
University of Massachusetts Medical School, Worcester, MA, USA

Lisa K. Selin
Department of Pathology, University of Massachusetts Medical School, Worcester, MA, USA

This manuscript reports a very useful study that extends our knowledge of peptide-MHC recognition by CD8+ T cells during emerging virus infections such as SARS-CoV-2. Detailed *in silico* analysis showed the presence of potential epitopes shared between new types of betacoronavirus: SARS-CoV-2 and common human alphacoronaviruses: OC43, HKU1, 229E and NL63. Due to the high prevalence of the common coronaviruses authors suggest that the large part of the human population has already some degree of specific memory T cell response before having been infected with the virus.
As authors already mentioned in their manuscript the similar study by Nguyen A. et al (JVI, 2020) demonstrated the HLA binding affinity of all possible 8- to 12-mers from SARS-CoV-2 proteome. This group found that HLA-B15:03 type has the greatest capacity to present highly conserved peptides which are shared among coronaviruses suggesting a cross-protective T cell immune response. In current manuscript using different prediction software authors identified and showed the sequence of epitopes which bind well to similar HLA type, HLA-B15:01. Interestingly, one of the epitopes (YLRKHFSM) can be bound by 4 different HLA types. The obvious strength of this study is the demonstration that certain epitopes, which are identical between SARS-CoV-2 and the common human coronaviruses are being predicted as high affinity binders in multiple HLA-A and B types.

Overall, the work reports important new details about SARS-CoV-2 epitopes theoretically being recognized by human CD8+ T cells. Undeniably, future experiments can prove if generated memory immune responses are specific to the proposed epitopes.

There are some suggestions:
1. The analysis of p/MHCI binding for HLA-C type (if available) would certainly complete the list of presented epitopes.

2. The introduction part subtitled: "The possibility of matching linear epitopes..." has missing information about previously published reports regarding T cell response in individuals infected with coronaviruses, either common or SARS-CoV.

3. In the discussion part readers may wonder why the authors did not discuss their findings with those already published (although they may not have been out at the time of submission) but should be included in the revision.

References
1. Nguyen A, David JK, Maden SK, Wood MA, et al.: Human Leukocyte Antigen Susceptibility Map for Severe Acute Respiratory Syndrome Coronavirus 2. J Virol. 2020; 94 (13). PubMed Abstract | Publisher Full Text

Is the work clearly and accurately presented and does it cite the current literature?
Partly

Is the study design appropriate and is the work technically sound?
Yes

Are sufficient details of methods and analysis provided to allow replication by others?
Yes

If applicable, is the statistical analysis and its interpretation appropriate?
Yes

Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

**Competing Interests:** No competing interests were disclosed.

**Reviewer Expertise:** T cell viral immunology

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

---

**Comments on this article**

**Version 1**

Author Response 02 Jun 2020

**Johannes M. Dijkstra**, Fujita Health University, Toyoake, Japan

In this comment I would just like to state that we submitted this article to F1000Research on April 15 (Japanese time). F1000Research only lists the publication date (April 23) which was after the journal edited our manuscript.

Also on behalf of Keiichiro Hashimoto,

Hans Dijkstra

**Competing Interests:** No competing interests were disclosed.

Author Response 28 Apr 2020

**Johannes M. Dijkstra**, Fujita Health University, Toyoake, Japan

Dear Dr. Bercovier,

Thank you for your comments and interesting links. Especially the data on the protective effects against common coronaviruses by previous infections by homologous virus are relevant.

In the F1000Research publication system, two or three reviewers will comment on-line, after which we as authors (probably) have to modify the manuscript following their comments. If those comments are in line with yours, we can implement some of the information that you provided.

Having said that, we also like to keep the message simple and concentrate on the MHC-I restricted immune memory against SARS-CoV-2 that can be expected to already exist in many people as induced by common coronaviruses. Especially for the elderly this is interesting in regard to vaccination.

Sincerely,

Hans Dijkstra

**Competing Interests:** No competing interests were disclosed.
Dear Colleagues,

I am in the middle of writing a similar paper and I would like you to take in consideration the following data that will make my paper unnecessary (I do not need any additional paper for my career) if it was added to your interesting article.

1) data on serological cross reaction among the Coronavirus that last for a while. (For a review: [https://www.medrxiv.org/content/10.1101/2020.04.14.20065771v1.full.pdf](https://www.medrxiv.org/content/10.1101/2020.04.14.20065771v1.full.pdf); and I have more papers including challenged studies in volunteer with old corona viruses, [https://academic.oup.com/jid/article/219/12/1913/5307035](https://academic.oup.com/jid/article/219/12/1913/5307035); [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2271881/pdf/epidinfect00023-0213.pdf](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2271881/pdf/epidinfect00023-0213.pdf))

2) CD4 data in a BioRxiv paper in healthy blood donors in Germany that cross with SARS-COV-2 peptides ([https://www.medrxiv.org/content/10.1101/2020.04.17.20061440v1.full.pdf](https://www.medrxiv.org/content/10.1101/2020.04.17.20061440v1.full.pdf)) and cross CD4 peptides among new coronaviruses ([https://pubmed.ncbi.nlm.nih.gov/27287409/?from_term=Coronavirus+cross+protection&from_pos=6](https://pubmed.ncbi.nlm.nih.gov/27287409/?from_term=Coronavirus+cross+protection&from_pos=6))

3) The hypothesis that these immune cross reactions (CD8, CD4, antibodies) prevent colonization of SARS-COV-2 in children aged 0-9 years who are permanently infected in kinder garden and primary school with these old common cold corona viruses. These children do not seem, as a result, able to contaminate neither siblings nor parents. The remaining immune memory would also explain why children aged 10-19 years who can be infected, do not usually develop any invasive serious disease caused by SARS-COV-2 but are able to infect other children and adults. Likewise, it would explain why the COVID19 clinical presentation is worsening with age.

I hope you will consider my remarks and integrate these data into your paper and it could be then a very comprehensive article on the subject of potential cross protection between the different beta Coronavirus.

sincerely yours,

H. Bercovier.

**Competing Interests:** I have no competing interests.
The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com