Identification of novel candidate CD8+ T cell epitopes of the SARS-CoV2 with homology to other seasonal coronaviruses

Pradeep Pushpakumara

University of Sri Jayewardenepura

Deshan Madusanka

University of Sri Jayewardenepura

Saubhagyagya Danasekara

University of Sri Jayewardenepura

Chandima Jeewandara

University of Sri Jayewardenepura

Graham Ogg

University of Oxford

Gathsaurie Neelika Malavige (✉ gathsaurie.malavige@ndm.ox.ac.uk)

University of Sri Jayewardenepura

Research Article

Keywords: CD8+ T cell epitopes, SARS-CoV2, Seasonal coronaviruses, HLA alleles, OC-43, HKU-1, NL-63

Posted Date: February 15th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-228306/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Individuals who have not been exposed to the SARS-CoV2 virus have been shown to have T cells that respond to the virus, possibly due to the presence of cross-reactive T cell responses to other seasonal human coronaviruses (HCoVs). Such cross-reactive T cell immunity may lead to immunopathology or protection.

Results

To understand the influence of such cross-reactive T cell responses, we used IEDB (Immune epitope database) and NetMHCpan (ver. 4.1) to identify candidate CD8 + T cell epitopes, restricted through HLA-A and B alleles, which are seen in a frequency of > 10% in the Sri Lankan population. Conservation analysis was carried out for these candidate epitopes with the HCoVs, OC43, HKU1, NL63 and with the current circulating different variants of SARS-CoV2. 12/18 the candidate CD8 + T cell epitopes (binding score of ≥ 0.90), which had a high degree of homology (> 75%) with the other three HCoVs were within the NSP12 and NSP13 proteins. They were predicted to be restricted through HLA-A*2402, HLA-A*201, HLA-A*206 and HLA-B alleles B*3501. 31 candidate CD8 + T cell epitopes that were specific to SARS-CoV2 virus (< 25% homology with other HCoVs) were predominantly identified within the structural proteins (spike, envelop, membrane and nucleocapsid) and the NSP1, NSP2 and NSP3. They were predominantly restricted through HLA-B*3501 (6/31), HLA-B*4001 (6/31), HLA-B*4403(7/31) and HLA-A*2402 (8/31). The candidate CD8 + T cell epitopes that were homologous or were specific, with a binding score of ≥ 0.90 , were found to be highly conserved within the SARS-CoV2 variants identified so far.

Conclusions

It would be crucial to understand T cell responses that associate with protection and the differences in the functionality and phenotype of epitope specific T cell responses, presented through different HLA alleles common in different geographical groups in order to understand disease pathogenesis.

Background

Infection due to the SARS-CoV2 virus is currently the leading cause of mortality among elderly and vulnerable groups in many countries. Outbreaks of COVID-19 are now reported in 220 countries in the world [1], with some countries experiencing a massive second wave. The majority of those who are infected with SARS-CoV2 experience asymptomatic or mild illness, whereas severe illness and fatalities are seen in older individuals and in those with comorbidities such as diabetes, cancer and cardiovascular disease[2–4]. However, the observed case fatality ratios (CFRs) vary widely between countries. For instance, countries such as Mexico, Italy and Iran report CFRs of over 3.5%, while countries such as India,

Turkey, South Korea reporting CFRs of < 2% [5]. The deaths per 100,000 population also vary widely among countries with many European countries and the US reporting rates of over 90 deaths/100,000 population while the majority of South Asian and South East Asian countries report death rates < 15 deaths/100,000 population[5].

There have been many factors that have been shown to influence the CFRs in different countries including the proportion of population aged > 70 years of age, the GDP per capita, BCG vaccination status [6, 7], climate, population density, and social distancing measures which could account for the differences in the CFRs. However, there could be many immune factors which also lead to these differences including immunity to human seasonal coronaviruses (HCoVs), providing cross protection against SARS-CoV2. Children are more frequently exposed to HCoVs compared to adults and recently it was shown that a large proportion of children, who were sampled between the years of 2011 to 2018, between the ages of 1 to 16, had IgG antibodies that cross-react with the spike protein of SARS-CoV2[8]. Sera from SARS-CoV2 uninfected individuals who had such cross-reactive antibodies were able to neutralize the pseudotypes of SARS-CoV2 without causing antibody dependent enhancement of infection[8].

Higher lymphocyte counts and specifically CD8 + T cell counts have been associated with early viral clearance and reduced disease severity[9]. Robust SARS-CoV2 specific CD4 + and CD8 + T cell responses have been detected following natural infection [10, 11]. SARS-CoV2 cross reactive T cells have been detected following SARS-CoV1 infection in 2003, which were predominantly directed against the N and structural proteins (NSP7 and NSP13)[12]. Interestingly, SARS-CoV2 cross reactive CD4 + T cells have also been detected in a large proportion of unexposed individuals in some studies which are thought to be due the presence of T cells specific to HCoVs [11, 13]. However, differences in T cell methodology have resulted in differential findings of cross-reactivity related to altered sensitivity and specificity. Ultimately, it will be important to know if T cells can respond to virally infected target cells where naturally processed epitopes are presented. Nevertheless, the presence of such cross-reactive T cell responses has been speculated to protect against infection with the SARS-CoV2 virus and have also speculated to cause disease pathogenesis or response to vaccines [12, 13].

As there is a difference in the CFRs among different countries, which are due to multiple factors, it would be important to understand if prior immunity to HCoVs could influence disease outcome when infected with the SARS-CoV2 virus. The predominant CD8 + and CD4 + T cell epitopes recognized by a population are also influenced by the frequency of different HLA types of a population. Therefore, in this study, we analyzed the cross reactivity of all the proteins of the SARS-CoV2 virus with three HCoVs (OC43, NL63 and HKU1) and those that are specific to the SARS-CoV2 virus and then used immune epitope database (IEDB) to predict the CD8 + T cell epitopes predicted to be restricted through the common HLA-Class I alleles in the Sri Lankan population.

Results

In order to identify the possible cross-reactive regions between SARS-CoV2 and other HCoVs, we carried out conservation analysis of SARS-CoV-2 virus with OC43, HKU1, and NL63, which have been seen in Sri Lanka. We analyzed the homology and conservation of the four structural and the sixteen NSPs of SARS-CoV2 with the three other HCoVs (Supplementary table 1). The NSP12, NSP13, and NSP16 of the SARS-CoV2 showed $\geq 65\%$ homology with OC43 and HKU, which are beta coronaviruses and a $> 55\%$ homology with NL63, which is an alpha coronavirus.

In contrast, the NSP1 and NSP2 proteins of the SARS-CoV2 showed $< 20\%$ homology with other three viruses, suggesting that these proteins were more specific for SARS-CoV-2 compared to the other proteins. All four structural proteins (S, E, M and N), NSP3, and NSP6 showed $< 35\%$ homology with OC43, HKU1, and NL63. Therefore, immune responses directed at these proteins may also be specific to the SARS-CoV2 unlike immune responses generated against NSP12, NSP13 and NSP16. The SARS-CoV-2 proteins S, E, M, N and the non-structural proteins showed less homology with NL63 than OC43, and HKU1, suggesting that the SARS-CoV-2 virus is genetically closer with OC43 and HKU1 than NL63.

Identification of possible CD8 + epitopes of the SARS-CoV2 virus restricted through HLA-A alleles of the SARS-CoV2

The NetMHCpan EL4 epitope prediction tool gives peptide binding scores ranging from 0 to 1.0 and we considered a predicted score ≥ 0.90 for a peptide as indicative of a stronger binder. None of the 8mer peptides gave predictive score of ≥ 0.90 . However, 39 9mer peptides gave a high binding score of ≥ 0.90 , which were restricted through different HLA-A alleles, (Table 1). Five of these epitopes were identified from the spike and NSP3 proteins and 4 epitopes each were identified from NSP4, NSP6, NSP12, NSP14, and NSP15. Peptides with high binding scores were not identified from the envelope, NSP1, NSP9, NSP10, NSP11, and NSP16 proteins. The epitopes from NSP3 ⁷²⁶YYTSNPTTF⁷³⁴ (predicted to be restricted through HLA-A*2402) and NSP6 ⁷⁰FLLPSLATV⁷⁸ (predicted to be restricted through HLA-A*0201) gave a score of 0.99, while ¹³⁴⁹NYMPYFFTL¹³⁵⁷ from NSP3, ⁴²⁰FLLNKEMYL⁴²⁸ from NSP4, and ¹⁵²ALWEIQV¹⁶⁰ from NSP8 also gave a score of 0.98 scores. However, they had $< 45\%$ homology with the other three viruses. 23/39 of the 9mer epitopes predicted in this study were restricted through HLA-A*02 while 16/39 9mer peptides were predicted to be restricted through HLA-A*24. Although HLA-A*33 was an allele seen in over 10% of the Sri Lankan population, 9mers that had a score of ≥ 0.90 was not identified.

Table 1
Predicted CD8 + epitopes of the SARS-CoV2 virus restricted through HLA-A alleles

9mer peptides with a peptide binding score of ≥ 0.90						
Protein	HLA-A allele	Sequence	Score	OC43 % identity with SARS-CoV2	HKU1 % identity with SARS-CoV2	NL63 % identity with SARS-CoV2
Spike	A*02:01	²⁶⁹ YLQPRTFLL ²⁷⁷	0.97	44	44	22
Spike	A*24:02	¹²⁰⁸ QYIKWPWYI ¹²¹⁶	0.95	77	77	77
Spike	A*02:01	⁹⁷⁶ VLNDILSRL ⁹⁸⁴	0.94	67	67	33
Spike	A*24:02	⁶³⁵ VYSTGSNVF ⁶⁴³	0.93	22	22	22
Spike	A*02:01	¹⁰⁹ TLDSKTQSL ¹¹⁷	0.91	22	22	33
Membrane	A*24:02	⁹⁵ YFIASFRLF ¹⁰³	0.91	77	77	77
Nucleocapsid	A*02:01	²²² LLLDRLNQL ²³⁰	0.96	33	33	33
NSP2	A*02:01	²⁶⁵ GLNDNLLEI ²⁷³	0.92	44	33	44
NSP2	A*24:02	⁴⁹⁷ TFFKLVNKF ⁵⁰⁵	0.90	33	22	22
NSP3	A*24:02	⁷²⁶ YYTSNPTTF ⁷³⁴	0.99	22	22	22
NSP3	A*24:02	¹³⁴⁹ NYMPYFFTL ¹³⁵⁷	0.98	33	22	11
NSP3	A*24:02	⁸¹⁶ YYHTTDPSF ⁸²⁴	0.96	11	11	0
NSP3	A*24:02	³⁶⁴ LYDKLVSSF ³⁷²	0.95	33	33	11
NSP3	A*24:02	¹⁰⁸¹ YYKKDNSYF ¹⁰⁸⁹	0.94	33	33	0
NSP4	A*02:01	⁴²⁰ FLLNKEMYL ⁴²⁸	0.98	33	33	33
NSP4	A*02:01	³⁵⁹ FLAHIQWMV ³⁶⁷	0.94	44	44	33
NSP4	A*24:02	³⁵¹ FYLTNDVSF ³⁵⁹	0.92	22	22	22
NSP4	A*24:02	⁴⁸⁶ LYQPPQTSI ⁴⁹⁴	0.90	66	66	66
NSP5	A*02:06	¹⁹⁴ LIQDYIQSV ²⁰²	0.95	44	33	11
NSP6	A*02:01	⁷⁰ FLLPSLATV ⁷⁸	0.99	33	33	44
NSP6	A*02:01	¹⁴¹ TLMNVLTIV ¹⁴⁹	0.92	0	0	0

9mer peptides with a peptide binding score of ≥ 0.90						
NSP6	A*24:02	⁸⁴ VYMPASWVM ⁹²	0.91	0	0	11
NSP6	A*24:02	¹¹⁵ MYASAVVLL ¹²³	0.90	22	22	22
NSP7	A*02:01	¹² VLLSVLQQL ²⁰	0.95	66	66	44
NSP8	A*02:01	¹⁵² ALWEIQQVV ¹⁶⁰	0.98	44	33	22
NSP12	A*24:02	³⁷ IYNDKVAGF ⁴⁵	0.95	66	44	44
NSP12	A*02:01	¹²³ TMADLVYAL ¹³¹	0.93	77	77	77
NSP12	A*02:06	³³⁴ FVDGVPFVV ³⁴²	0.93	100	100	88
NSP12	A*02:01	⁸⁵⁴ LMIERFVSL ⁸⁶²	0.91	88	88	55
NSP13	A*02:01	²³⁹ TLVPQEHYV ²⁴⁷	0.95	77	77	44
NSP13	A*24:02	³⁹⁷ VYIGDPAQL ⁴⁰⁵	0.91	100	100	77
NSP14	A*02:01	³²¹ LLADKFPVL ³²⁹	0.94	11	11	33
NSP14	A*02:01	¹⁷⁶ NLSDRVVFV ¹⁸⁴	0.93	66	66	77
NSP14	A*02:01	¹⁸⁴ VLWAHGFEL ¹⁹²	0.92	66	44	66
NSP14	A*02:01	⁴⁹⁴ YLDAYNMMI ⁵⁰²	0.91	44	44	44
NSP15	A*02:06	²⁴³ SQLGGLHLL ²⁵¹	0.96	66	66	66
NSP15	A*02:01	²⁹⁷ LLLDDFVEI ³⁰⁵	0.95	66	88	88
NSP15	A*02:06	³¹² SVVSKVVKV ³²⁰	0.94	66	66	55
NSP15	A*02:06	¹⁸¹ KVDGVVQQL ¹⁸⁹	0.92	22	22	11
10mer peptides with a peptide binding score of ≥ 0.90						
Protein	HLA-A allele	Sequence	Sore	OC43 % identity with SARS-CoV2	HKU1 % identity with SARS-CoV2	NL63 % identity with SARS-CoV2
Spike	A*24:02	¹⁰⁶⁶ TYVPAQEKNF ¹⁰⁷⁵	0.94	30	30	10
Spike	A*24:02	¹⁵⁹ VYSSANNCTF ¹⁶⁸	0.90	10	30	20
NSP3	A*24:02	⁷¹⁷ VYYTSNPTTF ⁷²⁶	0.98	30	40	20

9mer peptides with a peptide binding score of ≥ 0.90						
NSP6	A*24:02	²⁴² YDYLVTQEF ²⁵¹	0.91	50	50	50

Only four 10mer peptides were predicted to have a score of ≥ 0.90 and two were from spike and one each from NSP3 and NSP6 proteins (Table 1).

Identification of possible CD8 + epitopes of the SARS-CoV2 virus restricted through HLA-B alleles of the SARS-CoV2

As for HLA-A alleles, we considered a predicted score ≥ 0.90 for a peptide as indicative of a stronger binder. None of the 8mer peptides were found to give a score of ≥ 0.90 and therefore, were not predicted to be restricted through HLA-B alleles. However, 38 9mer peptides were identified which had high binding scores and were predicted to be restricted through HLA-B alleles (Table 2). The highest number of epitopes were predicted from spike protein (5/38) and NSP13 (4/38). 3 epitopes were predicted from each of the following proteins: namely the nucleocapsid, NSP2, NSP3, NSP4, and NSP12. No epitopes were identified from envelope, membrane, and NSP11 proteins. Nine epitopes gave score of 0.99 and were ⁸⁹⁵IPFAMQMAY⁹⁰³ from spike, ³²⁵TPSGTWLTY³³³ from nucleocapsid, ¹⁹⁵SEVGPEHSL²⁰³ and ⁵⁶²GETLPTEVL⁵⁷⁰ both from NSP2, ¹²⁰EEFEPSTQY¹²⁸ and ⁵⁴⁶QEILGTVSW⁵⁵⁴ both from NSP3, ⁷²LPSLATVAY⁸⁰ from NSP6, ⁴SEFSSLPSY¹² from NSP8, and ⁶⁰⁸VENPHLMGW⁶¹⁶ from NSP12.

Table 2
Predicted CD8 + epitopes of the SARS-CoV2 virus restricted through HLA-B alleles

9mer peptides with a peptide binding score of ≥ 0.90						
Protein	HLA-B allele	Sequence	Score	OC43 % identity with SARS-CoV2	HKU1 % identity with SARS-CoV2	NL63 % identity with SARS-CoV2
Spike	B*35:01	⁸⁹⁵ IPFAMQMAY ⁹⁰³	0.99	33	44	33
Spike	B*35:01	⁸³ LPFNDGVYF ⁹¹	0.98	22	33	11
Spike	B*44:03	¹²⁰⁰ QELGKYEYQ ¹²⁰⁸	0.98	44	44	33
Spike	B*35:01	⁶⁸⁶ VASQSIIAY ⁶⁹⁴	0.98	0	0	0
Spike	B*40:01	¹⁰¹⁵ AEIRASANL ¹⁰²³	0.98	22	22	44
Nucleocapsid	B*35:01	³²⁵ TPSGTWLTY ³³³	0.99	22	22	22
Nucleocapsid	B*44:03	³²² MEVTPSGTW ³³⁰	0.96	11	11	0
Nucleocapsid	B*35:01	⁷⁹ SPDDQIGYY ⁸⁷	0.92	22	22	55
NSP1	B*40:02	⁵⁶ VEKGVLPQL ⁶⁴	0.98	22	22	11
NSP1	B*35:01	¹¹⁰ HVGEIPVAY ¹¹⁸	0.95	22	11	11
NSP2	B*40:01	¹⁹⁵ SEVGPEHSL ²⁰³	0.99	11	11	0
NSP2	B*40:01	⁵⁶² GETLPTEVL ⁵⁷⁰	0.99	0	0	0
NSP2	B*44:02	⁵² REHEHEIAW ⁶⁰	0.98	22	22	22
NSP3	B*44:03	¹²⁰ EEFEPSTQY ¹²⁸	0.99	11	11	0
NSP3	B*44:03	⁵⁴⁶ QEILGTVSW ⁵⁵⁴	0.99	22	22	0
NSP3	B*40:01	¹⁷⁹⁹ AELAKNVSL ¹⁸⁰⁷	0.98	0	0	0
NSP4	B*40:02	³⁰⁹ GEYSHVVAF ³¹⁷	0.98	44	44	33
NSP4	B*35:01	³⁷³ VPFWITIAY ³⁸¹	0.98	33	44	00
NSP4	B*35:01	¹⁷⁴ NVLEGSVAY ¹⁸²	0.97	11	11	44
NSP5	B*35:01	⁹³ TANPKTPKY ¹⁰¹	0.95	66	66	44
NSP6	B*35:01	⁷² LPSLATVAY ⁸⁰	0.99	44	44	33

9mer peptides with a peptide binding score of ≥ 0.90						
NSP7	B*35:01	41 LAKDTTEAF ⁴⁹	0.91	33	33	22
NSP8	B*44:03	4 SEFSSLPSY ¹²	0.99	44	44	66
NSP8	B*40:01	47 SEFDRDAAM ⁵⁵	0.93	44	44	44
NSP9	B*44:03	67 TELEPPCRF ⁷⁵	0.95	66	66	66
NSP10	B*35:01	19 FAVDAAKAY ²⁷	0.98	55	55	55
NSP12	B*44:02	608 VENPHLMGW ⁶¹⁶	0.99	77	77	88
NSP12	B*40:01	874 QEYADVFLH ⁸⁸²	0.98	44	44	44
NSP12	B*35:01	337 VPFVVSTGY ³⁴⁵	0.97	88	88	55
NSP13	B*35:01	291 FAIGLALYY ²⁹⁹	0.96	66	77	66
NSP13	B*44:03	141 TEETFKLSY ¹⁴⁹	0.96	77	77	55
NSP13	B*40:01	155 REVLSDREL ¹⁶³	0.95	55	55	44
NSP13	B*40:02	161 RELHLSWEV ¹⁶⁹	0.91	77	77	66
NSP14	B*35:01	428 TPAFDKSAF ⁴³⁶	0.96	44	44	77
NSP14	B*35:03	19 HPTQAPTHL ²⁷	0.94	55	55	55
NSP15	B*35:03	49 LPVNVAFEL ⁵⁷	0.96	66	66	88
NSP15	B*44:03	200 QEFKPRSQM ²⁰⁸	0.92	33	33	55
NSP16	B*44:02	141 KENDSKEGF ¹⁴⁹	0.94	55	55	66
10mer peptides with a peptide binding score of ≥ 0.90						
Protein	HLA-B allele	Sequence	Sore	OC43 % identity with SARS-CoV2	HKU1 % identity with SARS-CoV2	NL63 % identity with SARS-CoV2
Spike	B*44:02	95TEKSNIIRGW ¹⁰⁴	0.95	10	0	10
Membrane	B*44:03	11EELKKLLEQW ²⁰	0.94	40	20	30
NSP2	B*44:03	489KEIKESVQTF ⁴⁹⁸	0.95	0	0	10
NSP3	B*44:03	120EEEFEPSTQY ¹²⁹	0.98	20	10	0

9mer peptides with a peptide binding score of ≥ 0.90						
NSP3	B*44:03	¹⁰⁷² TEIDPKLDNY ¹⁰⁸¹	0.96	40	30	10
NSP3	B*35:01	⁵⁰² VPTDNYITTY ⁵¹¹	0.94	0	0	0
NSP3	B*44:03	⁹⁴ GEFKLASHMY ¹⁰³	0.93	50	50	0
NSP7	B*40:01	⁴⁶ TEAFEKMOVSL ⁵⁵	0.91	50	40	20
NSP8	B*44:03	³ ASEFSSLPSY ¹²	0.94	40	40	60
NSP10	B*40:01	⁵ TEVPANSTVL ¹⁴	0.92	50	60	50
NSP12	B*44:03	⁸⁷⁵ QEYADVFLHY ⁸⁸⁴	0.99	50	50	40
NSP12	B*44:03	¹⁶⁶ VENPDILRVY ¹⁷⁵	0.95	80	70	80
NSP12	B*44:03	⁶⁰⁸ DVENPHLMGW ⁶¹⁷	0.90	80	70	80
NSP13	B*40:01	⁴⁴⁶ AEIVDTVSA ⁴⁵⁵	0.96	90	80	80
NSP14	B*44:02	⁷⁷ EEAIRHVRAW ⁸⁶	0.96	70	60	60
NSP14	B*35:01	⁴² IPGIPKDMTY ⁵¹	0.92	20	20	30
NSP15	B*35:01	²⁶⁹ IPMDSTVKNY ²⁷⁸	0.96	30	30	10
NSP15	B*40:01	⁴⁰ VELFENKTTL ⁴⁹	0.95	50	30	60

17/38 of these epitopes were predicted to be restricted through HLA-B*35 11/38 through HLA-B*40 and 10/38 through HLA-B*44. While most of these peptides showed < 45% homology with OC43, NL63 and HKU-1, ¹⁴¹TEETFKLSY¹⁴⁹ and ⁶⁰⁸VENPHLMGW⁶¹⁶ restricted through HLA-B*44, ³³⁷VPFVVSTGY³⁴⁵ restricted through HLA-B*35 and ¹⁶¹RELHLSWEV¹⁶⁹ restricted through HLA-B*40 showed a homology of > 75% with OC43 and HKU-1, which are two other beta coronaviruses.

Only 18 10mers peptides were predicted to have a score of ≥ 0.90 that were restricted through HLA-B (Table 2), 4/18 were identified NSP3. 10mer peptides with high binding scores were not predicted from envelope, nucleocapsid, NSP1, NSP4, NSP5, NSP6, NSP9, NSP11, and NSP16 9 proteins. 3/18 of these predicted 10mers were shown to be restricted through HLA-B*35, 4/18 through HLA-B*40, 11/18 through HLA-B*44. Therefore, there may be a higher probability 10mers to bind to HLA-B*44 alleles.

Identification of CD8 + T cell epitopes of SARS-CoV2, which show $\geq 75\%$ homology with OC43, HKU1, and NL63

After identification of peptides that had high predicted values to be restricted through the common HLA-A and B alleles present in the Sri Lankan population, we proceeded to identify the regions of the SARS-CoV2 virus, which had a > 75% homology with the HCoV. We then proceeded to identify CD8 + T cell epitopes within these regions, which were candidates to be restricted through these HLA alleles. This was to determine if we could identify CD8 + T cell epitopes of the SARS-CoV2, which were likely to cross-react with the other HCoVs. None of the predicted CD8 + 8mer epitopes identified within the SARS-CoV2 virus gave a high binding score and therefore, only predicted 9mer and 10mer CD8 + T cell epitopes of the SARS-CoV2 virus were analyzed for the degree of homology with OC43, HKU1, and NL63.

Epitopes which were identified to have a $\geq 75\%$ homology with more than two HCoV viruses are shown in Table 3. Twenty-four 9mer epitopes and 17 10mer peptides identified within the SARS-CoV2 virus had $\geq 75\%$ homology with ≥ 2 HCoV viruses. Of the 9mer peptides, 17/24 epitopes gave a peptide binding score of ≥ 0.90 . 11/24 of these CD8 + T cell epitopes within these cross-reactive regions were predicted to be restricted through HLA-A and 6/24 were predicted to be restricted through HLA-B alleles. Six highly cross-reactive CD8 + T cell epitopes (9mers) with high HLA-A (A*201 and A*206) binding scores were identified from the NSP12 and NSP13 (³³⁴FVDGVVPFVV³⁴²). Two of these peptides showed 100% homology with OC43 and HKU1. The Envelope, nucleocapsid proteins and the other non-structural proteins (apart from NSP12 and NSP13) did not have regions with > 75% homology with the other HCoVs. The alignment of SARS-CoV2 NSP12 and NSP13, in which most of the cross-reactive epitopes were identified from and their position is shown in supplementary Figs. 1 and 2.

Table 3

Predicted CD8 + epitopes of the SARS-CoV2 virus, which show $\geq 75\%$ homology with OC43, HKU1, and NL63

9mer peptides with $\geq 75\%$ homology with OC43, HKU1, and NL63						
Protein	HLA Allele	Sequence	Score	OC43 % identity with SARS-CoV2	HKU1 % identity with SARS-CoV2	NL63 % identity with SARS-CoV2
Spike	A*24:02	¹²⁰⁸ QYIKWPWYI ¹²¹⁶	0.95	77	77	77
Membrane	A*24:02	⁹⁵ YFIASFRLF ¹⁰³	0.91	77	77	77
NSP5	A*02:06	¹⁵⁹ FVYMHQLEL ¹⁶⁷	0.76	100	88	66
NSP5	B*35:01	⁹⁵ NPKTPKYKF ¹⁰³	0.47	77	77	55
NSP10	B*35:03	³⁶ QPITNCVKM ⁴⁴	0.74	88	88	66
NSP12	A*02:01	¹²³ TMADLVYAL ¹³¹	0.93	77	77	77
NSP12	A*02:06	³³⁴ FVDGVPFVV ³⁴²	0.93	100	100	88
NSP12	A*02:01	⁸⁵⁴ LMIERFVSL ⁸⁶²	0.91	88	88	55
NSP12	A*02:01	³³⁴ FVDGVPFVV ³⁴²	0.90	100	100	77
NSP12	B*44:03	⁶⁰⁸ VENPHLMGW ⁶¹⁶	0.99	77	77	88
NSP12	B*35:01	³³⁷ VPFVVSTGY ³⁴⁵	0.97	88	88	55
NSP13	A*02:01	²³⁹ TLVPQEHYV ²⁴⁷	0.95	77	77	44
NSP13	A*24:02	³⁹⁷ VYIGDPAQL ⁴⁰⁵	0.91	100	100	77
NSP13	A*02:06	²³⁹ TLVPQEHYV ²⁴⁷	0.91	77	77	44
NSP13	B*35:01	²⁹¹ FAIGLALYY ²⁹⁹	0.96	77	77	66
NSP13	B*44:03	¹⁴¹ TEETFKLSY ¹⁴⁹	0.96	77	77	55
NSP13	B*40:02	¹⁶¹ RELHLSWEV ¹⁶⁹	0.91	77	77	66
NSP14	A*02:01	¹⁷⁶ NLSDRVVFV ¹⁸⁴	0.93	77	66	77
NSP14	B*35:01	⁵⁰⁹ WVYKQFDY ⁵¹⁷	0.65	77	77	55
NSP15	A*02:01	²⁹⁷ LLLDDFVEI ³⁰⁵	0.95	66	88	88

9mer peptides with $\geq 75\%$ homology with OC43, HKU1, and NL63						
NSP15	B*35:03	⁴⁹ LPVNVAFEL ⁵⁷	0.96	77	66	88
NSP16	A*02:01	⁵³ YLNTLT LAV ⁶¹	0.89	88	88	55
NSP16	A*24:02	⁴⁶ KYTQLCQYL ⁵⁴	0.82	100	100	100
NSP16	A*33:03	²⁴⁷ MSKFPLKLR ²⁵⁵	0.78	77	77	44
10mer peptides with $\geq 75\%$ homology with OC43, HKU1, and NL63						
Protein	HLA Allele	Sequence	Sore	OC43 % identity with SARS-CoV2	HKU1 % identity with SARS-CoV2	NL63 % identity with SARS-CoV2
NSP12	A*02:06	³³² KIFVDGVPFV ³⁴¹	0.88	90	90	70
NSP12	A*02:01	³³² KIFVDGVPFV ³⁴¹	0.88	90	90	70
NSP13	A*24:02	²¹⁶ TYKLVNGDYF ²²⁵	0.77	80	70	50
NSP13	A*02:01	⁴⁰ KLVL SVN PYV ⁴⁹	0.58	80	80	50
NSP13	A*33:03	³⁸¹ NYDLSV VNAR ³⁹⁰	0.54	80	80	80
NSP13	A*33:03	⁵⁵¹ ETAHSCNVNR ⁵⁶⁰	0.52	90	90	80
NSP13	B*40:01	⁴⁴⁶ AEIVDTV SAL ⁴⁵⁵	0.96	90	80	80
NSP13	B*40:02	⁴⁴⁶ AEIVDTV SAL ⁴⁵⁵	0.87	90	80	80
NSP14	A*33:03	⁵¹⁶ TYNLWNTFTR ⁵²⁵	0.58	80	80	50
NSP14	A*24:02	⁵¹⁰ VYKQFD TYNL ⁵¹⁹	0.55	80	80	60
NSP15	A*33:03	⁵² NVAFELWAKR ⁶¹	0.66	80	60	90
NSP15	A*02:06	²⁴³ SQLGGLHLLI ²⁵²	0.52	70	70	80
NSP16	A*24:02	²⁴¹ SYSLFDMSKF ²⁵⁰	0.75	80	80	50
NSP16	A*33:03	²⁴⁶ DMSKFPLKLR ²⁵⁵	0.59	80	70	50
NSP16	A*24:02	²²¹ GYVMHANYIF ²³⁰	0.55	80	80	70
NSP16	A*02:01	²⁴³ SLFDMSKFPL ²⁵²	0.47	90	80	50

Of the 10mer peptides analyzed, 17 were identified within SARS-CoV2 to have $\geq 75\%$ homology with ≥ 2 HCoV (Table 3). Only one 10mer peptide identified within NSP-13 (⁴⁴⁶AEIVDTVSA⁴⁵⁵) gave a score of ≥ 0.90 score. 14/17 of these 10mer CD8 + T cell epitopes were predicted to be restricted through HLA-A and 3/17 were predicted to be restricted through HLA-B alleles. As with the 9mer peptides, 6 of the 10mer peptides, which were highly homologous with OC43, HKU1 and NL63 were found within the NSP12 and NSP13 region.

Identification of CD8 + T cell epitopes of SARS-CoV2, which show $\leq 25\%$ homology with OC43, HKU1, and NL63

After identification of highly cross reactive CD8 + T cell epitopes within the SARS-CoV2, we proceeded to identify regions, which were specific to the virus and did not cross react with other HCoV2, and therefore, are likely to be SARS-CoV2 specific CD8 + T cell epitopes. 9mer peptides of the representing different regions of the SARS-CoV2 virus, which have $\leq 25\%$ homology with > two HCoV viruses were analysed and 39 such potential CD8 + T cell epitopes were identified. (Table 4). 25/39 9mer peptides gave a binding score of ≥ 0.90 . 19/39 of these CD8 + T cell epitopes were predicted to be restricted through HLA-A alleles and 20/39 predicted to be restricted through HLA- B alleles. A region within the spike protein (⁶⁸⁶VASQSIIAY ⁶⁹⁴) had no homology with the other HCoVs but had a high binding score of >0.95 to HLA-B*3501 and two other 9mer peptides within the nucleocapsid (³²⁵TPSGTWLTY³³³ and ³²²MEVTPSGTW³³⁰) had < 22% homology and were predicted to be restricted through HLA-B*3501 and HLA-B*4403. Three other CD8 + T cell epitopes within NSP2, NSP3 and NSP6, which had high binding scores but had 0% homology were also identified.

Table 4

Predicted CD8 + epitopes of the SARS-CoV2 virus, which show $\leq 25\%$ homology with OC43, HKU1, and NL63

9mer peptides with $\leq 25\%$ homology with OC43, HKU1, and NL63						
Protein	HLA allele	Sequence	Score	OC43 % identity with SARS-CoV2	HKU1 % identity with SARS-CoV2	NL63 % identity with SARS-CoV2
Spike	A*24:02	⁶³⁵ VYSTGSNVF ⁶⁴³	0.93	22	22	22
Spike	A*02:01	¹⁰⁹ TLDSKTQSL ¹¹⁷	0.91	22	22	33
Spike	B*35:01	⁸³ LPFNDGVYF ⁹¹	0.98	22	33	11
Spike	B*35:01	⁶⁸⁶ VASQSIIAY ⁶⁹⁴	0.98	0	0	0
Spike	B*40:01	¹⁰¹⁵ AEIRASANL ¹⁰²³	0.98	22	22	44
Membrane	A*33:03	¹³⁸ LVIGAVILR ¹⁴⁶	0.72	22	22	22
Membrane	B*40:01	¹³⁶ SELVIGAVI ¹⁴⁴	0.73	11	11	11
Envelope	A*33:03	⁶¹ RVKNLNSSR ⁶⁹	0.61	11	11	0
Envelope	A*33:03	³⁰ TLAILTALR ³⁸	0.60	22	22	11
Envelope	B*40:01	⁶ SEETGTLIV ¹⁴	0.55	11	22	11
Envelope	B*35:03	⁴ FVSEETGTL ¹²	0.43	11	11	11
Nucleocapsid	B*35:01	³²⁵ TPSGTWLTY ³³³	0.99	22	22	22
Nucleocapsid	B*44:03	³²² MEVTPSGTW ³³⁰	0.96	11	11	0
NSP1	A*24:02	¹³⁵ SYGADLKSF ¹⁴³	0.89	0	0	0
NSP1	A*02:01	⁸⁴ VMVELVAEL ⁹²	0.85	22	22	0
NSP1	B*40:02	⁵⁶ VEKGVLPQL ⁶⁴	0.98	22	22	11
NSP1	B*35:01	¹¹⁰ HVGEIPVAY ¹¹⁸	0.95	22	11	11
NSP1	B*35:01	⁸⁹ VAELEGIQY ⁹⁷	0.63	11	0	11
NSP2	A*24:02	⁴⁹⁷ TFFKLVNKF ⁵⁰⁵	0.90	33	22	22
NSP2	A*02:06	⁴²⁰ YITGGVVQL ⁴²⁸	0.86	22	22	22

9mer peptides with $\leq 25\%$ homology with OC43, HKU1, and NL63						
NSP2	A*02:06	439TVYEKLKPV ⁴⁴⁷	0.83	11	11	11
NSP2	B*40:01	195SEVGPEHSL ²⁰³	0.99	11	11	0
NSP2	B*40:01	562GETLPTEVL ⁵⁷⁰	0.99	0	0	0
NSP2	B*44:03	52REHEHEIAW ⁶⁰	0.98	22	22	22
NSP3	A*24:02	726YYTSNPTTF ⁷³⁴	0.99	22	22	22
NSP3	A*24:02	1349NYMPYFFTL ¹³⁵⁷	0.98	33	22	11
NSP3	A*24:02	816YYHTTDPSF ⁸²⁴	0.96	11	11	0
NSP3	B*44:03	120EEFEPSTQY ¹²⁸	0.99	11	11	0
NSP3	B*44:03	546QEILGTVSW ⁵⁵⁴	0.99	22	22	0
NSP3	B*40:01	1799AELAKNVSL ¹⁸⁰⁷	0.98	0	0	0
NSP4	A*24:02	351FYLTNDVSF ³⁵⁹	0.92	22	22	22
NSP4	B*35:01	174NVLEGSVAY ¹⁸²	0.97	11	11	44
NSP6	A*02:01	141TLMNVLT ¹⁴⁹	0.92	0	0	0
NSP6	A*24:02	84VYMPASWVM ⁹²	0.91	0	0	11
NSP6	A*24:02	115MYASAVLL ¹²³	0.90	22	22	22
NSP6	B*35:01	156NALDQAISM ¹⁶⁴	0.88	22	22	11
NSP6	B*35:01	167LIISVTSNY ¹⁷⁵	0.56	11	11	22
NSP14	A*02:01	321LLADKFPVL ³²⁹	0.94	11	11	33
NSP15	A*02:06	181KVDGVVQQL ¹⁸⁹	0.92	22	22	11
10mer peptides with $\leq 25\%$ homology with OC43, HKU1, and NL63						
Protein	HLA Allele	Sequence	Sore	OC43 % identity with SARS-CoV2	HKU1 % identity with SARS-CoV2	NL63 % identity with SARS-CoV2
Spike	A*24:02	368LYNSASFSTF ³⁷⁷	0.89	20	20	30
Spike	A*24:02	788IYKTPPIKDF ⁷⁹⁷	0.88	10	20	10

9mer peptides with $\leq 25\%$ homology with OC43, HKU1, and NL63						
Spike	B*44:02	⁹⁵ TEKSNIIRGW ¹⁰⁴	0.95	10	0	10
Spike	B*35:01	²²⁹ LPIGINITRF ²³⁸	0.82	10	20	20
Membrane	A*33:03	¹³⁷ ELVIGAVILR ¹⁴⁶	0.65	20	20	20
Membrane	A*33:03	¹⁷⁷ SYVKLGASQR ¹⁸⁶	0.55	40	20	20
Membrane	B*40:01	¹³⁶ SELVIGAVIL ¹⁴⁵	0.76	10	10	20
Envelope	A*33:03	⁶⁰ SRVKNLNSSR ⁶⁹	0.48	10	30	0
Envelope	A*33:03	²⁹ VTLAILTALR ³⁸	0.41	30	20	0
Nucleocapsid	A*33:03	¹⁴⁰ NTPKDHIGTR ¹⁴⁹	0.60	20	10	20
Nucleocapsid	A*02:01	³⁹⁸ ADLDDFSKQL ⁴⁰⁷	0.44	10	30	0
Nucleocapsid	B*44:03	³²¹ GMEVTPSGTW ³³⁰	0.87	0	0	0
Nucleocapsid	B*35:01	³²⁴ VTPSGTWLTY ³³³	0.68	20	20	20
Nucleocapsid	B*40:01	³²² MEVTPSGTWL ³³¹	0.67	10	10	0
NSP1	A*33:03	¹⁶² NTKHSSGVTR ¹⁷¹	0.74	0	0	0
NSP1	A*33:03	⁶⁸ YVFIKRSDAR ⁷⁷	0.49	30	20	10
NSP1	A*02:06	¹⁴ VQLSLPVLQV ²³	0.45	20	10	10
NSP1	A*33:03	¹⁵ QLSLPVLQVR ²⁴	0.40	10	10	10
NSP1	B*35:01	⁶¹ LPQLEQPYVF ⁷⁰	0.81	20	20	30
NSP1	B*40:02	¹¹² GEIPVAYRKV ¹²¹	0.54	30	20	20
NSP1	B*40:02	⁹ NEKTHVQLSL ¹⁸	0.49	20	20	20
NSP1	B*35:03	⁶¹ LPQLEQPYVF ⁷⁰	0.43	20	20	30
NSP1	B*35:03	¹⁸ LPVLQVRDVL ²⁷	0.41	20	20	10
NSP2	A*02:01	³⁸⁹ ILDGISQYSL ³⁹⁸	0.65	0	0	10
NSP2	A*33:03	⁵²⁹ FVTHSKGLYR ⁵³⁸	0.61	0	0	30
NSP2	A*02:01	²⁸⁸ KLNEEIAIL ²⁹⁷	0.59	10	10	30

9mer peptides with $\leq 25\%$ homology with OC43, HKU1, and NL63						
NSP2	A*24:02	¹ AYTRYVDNNF ¹⁰	0.59	30	20	20
NSP2	B*44:03	⁴⁸⁹ KEIKESVQTF ⁴⁹⁸	0.95	0	0	10
NSP2	B*44:03	⁴⁵² EEKFKEGVEF ⁴⁶¹	0.89	10	10	30
NSP2	B*40:01	³⁴⁴ GEQKSILSPL ³⁵³	0.85	10	10	20
NSP3	A*24:02	¹⁶ QGYKSVNITF ²⁵	0.79	10	20	20
NSP3	A*24:02	¹⁰⁴⁰ EYKGPITDVF ¹⁰⁴⁹	0.75	20	20	0
NSP3	B*44:03	¹²⁰ EEEFEPSTQY ¹²⁹	0.98	20	10	0
NSP3	B*35:01	⁵⁰² VPTDNYITTY ⁵¹¹	0.94	0	0	0
NSP4	A*02:06	¹⁰¹ FVVPGLPGTI ¹¹⁰	0.65	20	10	50
NSP4	B*40:01	⁹⁷ REVG FVVPGL ¹⁰⁶	0.73	20	10	50
NSP5	A*24:02	¹²⁵ VYQCAMRPNF ¹³⁴	0.50	20	20	40
NSP6	A*33:03	⁸⁴ VYMPASWVMR ⁹³	0.53	0	0	10
NSP6	B*35:01	⁴³ LPFAMGIIAM ⁵²	0.71	0	0	10
NSP7	B*44:03	⁷³ EEMLDNRATL ⁸²	0.58	10	10	40
NSP8	A*02:01	¹⁵¹ SALWEIQQVV ¹⁶⁰	0.68	20	30	20
NSP14	B*35:01	⁴² IPGIPKDMTY ⁵¹	0.92	20	20	30
NSP15	A*33:03	²¹⁵ ELAMDEFIER ²²⁴	0.57	40	20	20
NSP15	B*44:03	¹⁶⁹ GEAVKTQFNY ¹⁷⁸	0.74	10	10	10

We identified 44 10mers as CD8 + T cell epitopes, which had $\leq 25\%$ homology with > two HCoV viruses (Table 4). 5/44 epitopes gave a score of ≥ 0.90 . 22/44 of these CD8 + T cell epitopes were predicted to be restricted through HLA-A and 22/44 were predicated to be restricted through HLA-B alleles. Again, the peptides which had the highest binding scores and least percentage identify were predicted to be restricted through HLA-B*3501 and HLA-B*4403. The highest binders, which were specific to SARS-CoV2 were identified within the spike protein (⁹⁵TEKSNIIRGW¹⁰⁴), NSP2 (⁴⁸⁹KEIKESVQTF⁴⁹⁸) and NSP3 (¹²⁰EEEFEPSTQY¹²⁹ and ⁵⁰²VPTDNYITTY⁵¹¹).

Conservational analysis of the candidate CD8 + T cell epitopes with binding scores of ≥ 0.90

We proceeded to investigate if the 18 candidate CD8 + T cell epitopes which had a percentage identity of $> 75\%$ with other HCoV2s and the 31 SARS-CoV2 specific ($< 25\%$ percentage identity) were conserved within the SARS-CoV2. We found that these candidate epitopes were highly conserved (supplementary table 2 and 3) and these regions were highly conserved within the new SARS-CoV2 variants as well (supplementary Figs. 3 and 4).

Similarity of candidate peptides with published CD8 + T cell SARS-CoV2 epitopes

Several CD8 + T cell epitopes that are restricted through different HLA-A and B alleles have been published[10, 14–17]. We proceeded to find out if any of the candidate CD8 + T cell epitopes were already identified in patients who were naturally infected with the SARS-CoV2 virus. We found that 20/31 candidate highly conserved T cell epitopes which were found to be specific to the SARS-CoV2 ($< 25\%$ homology with other HCoV2s) had been identified in infected individuals (supplementary table 4). In our HLA allele prediction analysis using the dominant HLA alleles in Sri Lanka, although some of the epitopes were predicted to be restricted through HLA-B*3501 and HLA-B*4403, some of these epitopes were found to be restricted through HLA-A*0201, A*1101 and HLA-A*0301.

7/18 of the candidate T cell epitopes which were found to be cross reactive ($> 75\%$ homology with the other HCoV2s) were also identified from those who were naturally infected. For the candidate CD8 + T cell epitopes that were found to be cross reactive with other HCoV2, the predicted HLA allele by us and the HLA allele restriction identified following natural infection were similar in 4/7 epitopes (supplementary table 5).

Discussion

In this study we have identified candidate CD8 + T cell epitopes, which were highly conserved within SARS-CoV2, and some which show $> 75\%$ percentage homology with the HCoV2s OC43, HKU1 and NL63, and therefore, are candidates to give rise to cross-reactive T cell responses. The majority of the predicted CD8 + T cell epitopes (binding score of ≥ 0.90), which had a high degree of homology with the other three HCoV2s were within the NSP12 and NSP13 proteins. They were predicted to be restricted through HLA-A*2402, HLA-A*0201, HLA-A*0206 and HLA-B alleles B*3501. Therefore, the presence of SARS-CoV2 cross reactive CD8 + T cell responses could depend on the frequency of the above HLA-A alleles in a population, as the most cross-reactive candidate CD8 + T cell epitopes are restricted through these alleles.

The frequency of HLA-A*0201 and HLA-A*0206 in the Sri Lankan population are 4.9 to 6.6% and 2.1–2.4% respectively, while the frequency of HLA-A*2402 is 20.8–30.3%[18, 19]. HLA-B*35 frequency in the Sri Lankan population is 21–23% [18, 19]. In contrast, the most frequent HLA-A alleles in the European,

US and Brazilian populations are HLA-A*0201, A*0206 (24.5 to 27.5%), which are several fold higher than in the Sri Lankan population, while HLA-A*2402 and B*3501 are lower (7.9 to 9.5%)[20–22]. In silico analysis, recently showed that HLA-A*0201 was associated with a higher risk of COVID-19, while HLA-A*2402 was shown to associate with higher capacity to present SARS-CoV2 antigens[23]. SARS-CoV2 specific HLA-A*0201 CD8 + T cell epitopes were shown to have suboptimal antiviral response and of a reduced frequency when compared to other viral infections such as influenza and Epstein-Barr viral infection[24]. The in-silico analysis showed that countries (Italy, France, Germany, Brazil) in which the most frequent HLA-A allele was A*0201 had the highest COVID-19 case fatality rates (CFRs), whereas those where HLA-A*2402 allele was the most frequent (India, Iran) had lower CFRs[23]. Therefore, it would be important to investigate if certain HLA alleles, presented CD8 + T cell epitopes that associate with protection, whereas if certain other alleles, present epitopes that are associated with immunopathology and poor antiviral capacity.

In contrast the candidate CD8 + T cell epitopes, which were highly conserved identified within SARS-CoV2 virus that are likely to be specific, were predominantly identified within the structural proteins (spike, envelope, membrane and nucleocapsid) and the NSP1, NSP2 and NSP3. 6/31 of these candidate CD8 + T cell epitopes (binding score of ≥ 0.90), that were specific to SARS-CoV2 (< 25% homology with the HCoV) were predicted to be restricted through HLA-B*3501, 6/31 through HLA-B*4001, 7/31 through HLA-B*4403 and 8/31 through HLA-A*2402. Only 3/31 of the SARS-CoV2 specific candidate T cell epitopes were predicted to be restricted through HLA-A*0201 or A*0206 alleles, common in Europe, USA and Brazil. Therefore, HLA-A*2402 and HLA-B*3501, HLA-B*4001, HLA-B*4403, which are predominant HLA alleles in Sri Lanka and India, may present both highly cross-reactive and SARS-CoV2 specific CD8 + T cell epitopes. Indeed, our analysis showed that 20/31 highly conserved, SARS-CoV2 specific candidate CD8 + T cell epitopes were already identified in those with natural infection. Although our HLA allele prediction using the dominant HLA alleles in Sri Lanka predicted these epitopes to be restricted through HLA-B*3501 and HLA-B*4403, some were found to be presented through different HLA alleles in those who were naturally infected in Europe and USA [10, 14, 17]. However, given that these epitopes are highly conserved within the virus, it is possible that they could be presented by multiple HLA alleles, which should be further investigated.

T cell responses of higher magnitude and breadth have been observed in patients who had more severe COVID-19 [11, 25]. However, it is not yet known if a higher magnitude and breadth of T cell responses in COVID-19 is associated with protection or immunopathology. There is a debate if cross-reactive T cells cause immunopathology in certain viral infections such as in dengue [26, 27] and if such cross-reactive T cells in SARS-CoV2 are protective should be investigated. It is hoped that the candidate epitopes presented here, will be of help in subsequent functional T cell analyses, particularly as viral variants emerge as they were found to be highly conserved within the new UK SARS-CoV2 variant, B.1.1.7 and the new South African variant 501.V2. By focusing on highly conserved regions within and between each coronavirus, the candidate epitopes may be of value in understanding immune responses across populations and for future vaccine design. In addition, since many vaccines for COVID-19 are currently being rolled out, it would be crucial to understand T cell responses that associate with protection and the

differences in the functionality of epitope specific T cell responses, presented through different HLA alleles.

Conclusions

In summary, we have identified candidate SARS-CoV2 CD8 + T cell responses that are highly cross reactive with other HCoVs and also that are specific to SARS-CoV2. Cross-reactive epitopes were predominantly identified from NSP12 and NSP13, while specific epitopes were identified within the structural proteins and NSP1-3. It would be crucial to understand the CD8 + T cell epitopes presented through the most frequent HLA alleles, their phenotype and functionality to better understand the immune responses to SARS-CoV2 and possible implications for vaccines.

Methods

Data collection for seasonal human corona virus (HCoV) sequence analysis

The full-length protein sequences of the SARS-CoV2 (NC_045512), OC43 (NC_006213), HKU1 (NC_006577), and NL63 (NC_005831) were obtained from National Center for Biotechnology Information (NCBI) GenBank (<https://www.ncbi.nlm.nih.gov/>). Four structural and 16 non-structural proteins (NSPs) were individually used for T cell epitope prediction and conservation analysis.

T-cell epitope prediction for MHC Class I HLA alleles

IEDB (Immune epitope database) (www.iedb.org) [28] along with NetMHCpan (ver. 4.1)[29] was used for finding putative peptide sequences that were restricted through the MHC Class-I HLA alleles, as these methods use algorithms (generated by artificial neural networks), which have accuracy of epitope prediction. HLA-A*02, HLA-A*24, HLA-A*33, HLA-B*35, HLA-B*40, and HLA-B*44 alleles were used for this epitope prediction, as these five alleles are the predominant alleles (each allele seen in > 10% in the population) in the Sri Lankan population [18]. The epitope predictions were carried out for 8mer, 9mer, and 10mer epitopes. According to the epitope prediction NetMHCpan EL 4., which gives a score ranging from 0.0 to 1.0. Epitopes with a higher score are considered as stronger binders.

Conservation analysis of SARS-CoV-2 proteins and predicted epitopes with OC43, HKU1, and NL63

The epitopes which had a predicted binding score of > 0.8 through the above approach, which were highly conserved between the SARS-CoV2 and OC43, HKU1, and NL63 were further analyzed to determine percentage of identities and similarities between SARS-CoV2 and other HCoVs. The sequences were aligned using the Clustal W on MEGA X software (www.megasoftware.net). Conservation of protein sequences were analyzed by using conservation analysis tool available at European Bioinformatics Institute (EBI) (www.ebi.sc.uk), IEDB, and Jalview (www.jalview.org).

Abbreviations

HCoV2

Seasonal human coronaviruses

NSP

nonstructural proteins

CFR

case fatality rates

Declarations

Ethical Approval and Consent to participate

Not applicable for this study as this is a bioinformatic analysis.

Consent for publication

Not applicable for this study.

Availability of data and materials

All data generated or analysed during this study are included in this article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Funding

We are grateful to the Centre for Dengue Research, UK Medical Research Council and the Foreign and Commonwealth Office for support.

Author contributions

PDP: Conceptualization, analysis and writing the manuscript.

DM: analysis of data

SD: analysis of data

CJ: Funding, providing resources and reviewing the final draft

GSO: Conceptualization, funding, reviewing the final draft

GNM: Conceptualization, funding, writing the manuscript and review

Acknowledgements

None

References

1. **COVID-19 CORONAVIRUS OUTBREAK**
2. Wang G, Wu C, Zhang Q, Yu B, Lu J, Zhang S, Wu G, Wu Y, Zhong Y: **Clinical characteristics and the risk factors for severe events of elderly coronavirus disease 2019 patients.** *Zhong Nan Da Xue Xue Bao Yi Xue Ban* 2020, **45**(5):542-548.
3. Albitar O, Ballouze R, Ooi JP, Sheikh Ghadzi SM: **Risk factors for mortality among COVID-19 patients.** *Diabetes research and clinical practice* 2020, **166**:108293.
4. Tian W, Jiang W, Yao J, Nicholson CJ, Li RH, Sigurslid HH, Wooster L, Rotter JI, Guo X, Malhotra R: **Predictors of mortality in hospitalized COVID-19 patients: A systematic review and meta-analysis.** *Journal of medical virology* 2020.
5. **Mortality analysis** [<https://coronavirus.jhu.edu/data/mortality>]
6. Sorci G, Faivre B, Morand S: **Explaining among-country variation in COVID-19 case fatality rate.** *Sci Rep* 2020, **10**(1):18909.
7. Escobar LE, Molina-Cruz A, Barillas-Mury C: **BCG vaccine protection from severe coronavirus disease 2019 (COVID-19).** *Proceedings of the National Academy of Sciences of the United States of America* 2020, **117**(30):17720-17726.
8. Ng KW, Faulkner N, Cornish GH, Rosa A, Harvey R, Hussain S, Ulferts R, Earl C, Wrobel AG, Benton DJ *et al*: **Preexisting and de novo humoral immunity to SARS-CoV-2 in humans.** *Science* 2020, **370**(6522):1339-1343.
9. Lin A, He ZB, Zhang S, Zhang JG, Zhang X, Yan WH: **Early Risk Factors for the Duration of Severe Acute Respiratory Syndrome Coronavirus 2 Viral Positivity in Patients With Coronavirus Disease 2019.** *Clin Infect Dis* 2020, **71**(16):2061-2065.
10. Peng Y, Mentzer AJ, Liu G, Yao X, Yin Z, Dong D, Dejnirattisai W, Rostron T, Supasa P, Liu C *et al*: **Broad and strong memory CD4(+) and CD8(+) T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19.** *Nature immunology* 2020, **21**(11):1336-1345.
11. Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Moderbacher CR, Rawlings SA, Sutherland A, Premkumar L, Jadi RS *et al*: **Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals.** *Cell* 2020, **181**(7):1489-1501 e1415.
12. Le Bert N, Tan AT, Kunasegaran K, Tham CYL, Hafezi M, Chia A, Chng MHY, Lin M, Tan N, Linster M *et al*: **SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls.** *Nature* 2020, **584**(7821):457-462.
13. Weiskopf D, Schmitz KS, Raadsen MP, Grifoni A, Okba NMA, Endeman H, van den Akker JPC, Molenkamp R, Koopmans MPG, van Gorp ECM *et al*: **Phenotype and kinetics of SARS-CoV-2-specific**

- T cells in COVID-19 patients with acute respiratory distress syndrome.** *Sci Immunol* 2020, **5**(48).
14. Shomuradova AS, Vagida MS, Sheetikov SA, Zornikova KV, Kiryukhin D, Titov A, Peshkova IO, Khmelevskaya A, Dianov DV, Malasheva M *et al*: **SARS-CoV-2 Epitopes Are Recognized by a Public and Diverse Repertoire of Human T Cell Receptors.** *Immunity* 2020, **53**(6):1245-1257 e1245.
 15. Nelde A, Bilich T, Heitmann JS, Maringer Y, Salih HR, Roerden M, Lubke M, Bauer J, Rieth J, Wacker M *et al*: **SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition.** *Nature immunology* 2021, **22**(1):74-85.
 16. Ferretti AP, Kula T, Wang Y, Nguyen DMV, Weinheimer A, Dunlap GS, Xu Q, Nabils N, Perullo CR, Cristofaro AW *et al*: **Unbiased Screens Show CD8(+) T Cells of COVID-19 Patients Recognize Shared Epitopes in SARS-CoV-2 that Largely Reside outside the Spike Protein.** *Immunity* 2020, **53**(5):1095-1107 e1093.
 17. Schulien I, Kemming J, Oberhardt V, Wild K, Seidel LM, Killmer S, Sagar, Daul F, Salvat Lago M, Decker A *et al*: **Characterization of pre-existing and induced SARS-CoV-2-specific CD8(+) T cells.** *Nature medicine* 2020.
 18. Malavige GN, Rostron T, Seneviratne SL, Fernando S, Sivayogan S, Wijewickrama A, Ogg GS: **HLA analysis of Sri Lankan Sinhalese predicts North Indian origin.** *International journal of immunogenetics* 2007, **34**(5):313-315.
 19. Grifoni A, Weiskopf D, Lindestam Arlehamn CS, Angelo M, Leary S, Sidney J, Frazier A, Phillips E, Mallal S, Mack SJ *et al*: **Sequence-based HLA-A, B, C, DP, DQ, and DR typing of 714 adults from Colombo, Sri Lanka.** *Human immunology* 2018, **79**(2):87-88.
 20. Ayo CM, da Silveira Camargo AV, Xavier DH, Batista MF, Carneiro OA, Brandao de Mattos CC, Ricci O, Jr., de Mattos LC: **Frequencies of allele groups HLA-A, HLA-B and HLA-DRB1 in a population from the northwestern region of Sao Paulo State, Brazil.** *International journal of immunogenetics* 2015, **42**(1):19-25.
 21. Neville MJ, Lee W, Humburg P, Wong D, Barnardo M, Karpe F, Knight JC: **High resolution HLA haplotyping by imputation for a British population bioresource.** *Human immunology* 2017, **78**(3):242-251.
 22. Requena D, Medico A, Chacon RD, Ramirez M, Marin-Sanchez O: **Identification of Novel Candidate Epitopes on SARS-CoV-2 Proteins for South America: A Review of HLA Frequencies by Country.** *Frontiers in immunology* 2020, **11**:2008.
 23. Tomita Y, Ikeda T, Sato R, Sakagami T: **Association between HLA gene polymorphisms and mortality of COVID-19: An in silico analysis.** *Immun Inflamm Dis* 2020, **8**(4):684-694.
 24. Habel JR, Nguyen THO, van de Sandt CE, Juno JA, Chaurasia P, Wragg K, Koutsakos M, Hensen L, Jia X, Chua B *et al*: **Suboptimal SARS-CoV-2-specific CD8(+) T cell response associated with the prominent HLA-A*02:01 phenotype.** *Proceedings of the National Academy of Sciences of the United States of America* 2020, **117**(39):24384-24391.
 25. Peng Y, Mentzer AJ, Liu G, Yao X, Yin Z, Dong D, Dejnirattisai W, Rostron T, Supasa P, Liu C *et al*: **Broad and strong memory CD4 (+) and CD8 (+) T cells induced by SARS-CoV-2 in UK convalescent**

COVID-19 patients. *bioRxiv* 2020.

26. Mongkolsapaya J, Dejnirattisai W, Xu XN, Vasanawathana S, Tangthawornchaikul N, Chairunsri A, Sawasdivorn S, Duangchinda T, Dong T, Rowland-Jones S *et al.* **Original antigenic sin and apoptosis in the pathogenesis of dengue hemorrhagic fever.** *Nature medicine* 2003, **9**(7):921-927.
27. Mongkolsapaya J, Duangchinda T, Dejnirattisai W, Vasanawathana S, Avirutnan P, Jairungsri A, Khemnu N, Tangthawornchaikul N, Chotiyarnwong P, Sae-Jang K *et al.* **T cell responses in dengue hemorrhagic fever: are cross-reactive T cells suboptimal?** *J Immunol* 2006, **176**(6):3821-3829.
28. Kim Y, Ponomarenko J, Zhu Z, Tamang D, Wang P, Greenbaum J, Lundegaard C, Sette A, Lund O, Bourne PE *et al.* **Immune epitope database analysis resource.** *Nucleic Acids Res* 2012, **40**(Web Server issue):W525-530.
29. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M: **NetMHCpan-4.1 and NetMHCIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data.** *Nucleic Acids Res* 2020, **48**(W1):W449-W454.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarydata.docx](#)