Identification of SARS-CoV-2 Nucleocapsid and Spike T-cell Epitopes for Assessing T-cell Immunity

Eunok Lee\textsuperscript{a,b}*, Kerrie Sandgren\textsuperscript{a,b}, Gabriel Duette\textsuperscript{a,b}, Vicki V. Stylianou\textsuperscript{a,b}, Rajiv Khanna\textsuperscript{c,d}, John-Sebastian Eden\textsuperscript{a,b,e,f}, Emily Blyth\textsuperscript{b,g,h}, David Gottlieb\textsuperscript{b,g,h}, Anthony L. Cunningham\textsuperscript{a,b} and Sarah Palmer\textsuperscript{a,b}

\textsuperscript{a} Centre for Virus Research, The Westmead Institute for Medical Research, Westmead, New South Wales, Australia
\textsuperscript{b} Sydney Medical School, Faculty of Medicine and Health, The University of Sydney, Camperdown, New South Wales, Australia
\textsuperscript{c} Faculty of Medicine, The University of Queensland, Brisbane, Queensland, Australia
\textsuperscript{d} QIMR Berghofer Centre for Immunotherapy and Vaccine Development, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
\textsuperscript{e} Centre for Infectious Diseases and Microbiology, The Westmead Institute for Medical Research, Westmead, New South Wales, Australia
\textsuperscript{f} Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Westmead, New South Wales, Australia
\textsuperscript{g} Centre for Cancer Research, The Westmead Institute for Medical Research, Westmead, New South Wales, Australia
BMT and Cell Therapies Program, Westmead Hospital, Westmead, New South Wales, Australia

Running Head: SARS-CoV-2 specific peptides for T-cell immunity assay

#Address correspondence to Eunok Lee, eunok.lee@sydney.edu.au.

Eunok Lee, Kerrie Sandgren and Gabriel Duette contributed equally to this work. Author order was determined by the contribution to the results presented in the manuscript and contribution to writing the initial manuscript draft

Anthony L. Cunningham and Sarah palmer are co-senior authors with equal contribution.

Two-part abstract word count: 308

Main text word count: 8831
Abstract

Developing optimal T-cell response assays to severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) is critical for measuring the duration of immunity to this disease and assessing the efficacy of vaccine candidates. These assays need to target conserved regions of SARS-CoV-2 global variants and avoid cross-reactivity to seasonal human coronaviruses. To contribute to this effort, we employed an in-silico immunoinformatics analysis pipeline to identify immunogenic peptides resulting from conserved and highly networked regions with topological importance from the SARS-CoV-2 nucleocapsid and spike proteins. A total of 57 highly networked T-cell epitopes that are conserved across geographic viral variants were identified from these viral proteins, with a binding potential to diverse HLA alleles and 80-100% global population coverage. Importantly, 18 of these T-cell epitope derived peptides had limited homology to seasonal human coronaviruses making them promising candidates for SARS-CoV-2 specific T-cell immunity assays. Moreover, two of the NC-derived peptides elicited effector/polyfunctional responses of CD8+ T-cells derived from SARS-CoV-2 convalescent patients.

Importance

The development of specific and validated immunologic tools is critical for understanding the level and duration of the cellular response induced by SARS-CoV-2 infection and/or vaccines against this novel coronavirus disease. To contribute to this effort, we employed an immunoinformatics analysis pipeline to define 57 SARS-CoV-2 immunogenic peptides within topologically important regions of the nucleocapsid (NC) and spike (S) proteins that will be effective for detecting cellular immune responses in 80-100% of the global population. Our immunoinformatics analysis revealed that 18 of these peptides had limited homology to...
circulating seasonal human coronaviruses, and therefore are promising candidates for
distinguishing SARS-CoV-2 specific immune responses from pre-existing coronavirus
immunity. Importantly, CD8+ T-cells derived from SARS-CoV-2 survivors exhibited
polyfunctional effector responses to two novel NC-derived peptides identified as HLA-binders.
These studies provide a proof of concept that our immunoinformatics analysis pipeline identifies
novel immunogens which can elicit polyfunctional SARS-CoV-2 specific T-cell responses.

Introduction

As of October 2020, there are over 36 million known cases of COVID-19 world-wide, which is
caused by infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). In
order to combat this pandemic, vaccines are rapidly being developed to reduce the risk and
spread of this infection (1-6). Recently, a clinical trial involving a vaccine that contains the spike
glycoprotein (S) of SARS-CoV-2 showed T-cell responses were elicited at day 14 and antibodies
against the virus at day 28 post-vaccination for the majority of participants (7). However, the
longitudinal assessment of the level and duration of both T-cell immunity and antibodies elicited
by this and other SARS-CoV-2 vaccines is required.

Previously, it has been shown that antibody levels wane with time in SARS-CoV-1 infection
while cellular immunity can last six to eleven years (8-13). Similarly, a recent study of antibody
levels revealed that 40% of asymptomatic and 13% of symptomatic patients infected by SARS-
CoV-2 became negative for immunoglobulin G eight weeks post-recovery (14). Exposing T-cells
from recovered SARS-CoV-1 patients to peptides derived from the S protein of this virus
revealed that the induction of polyfunctional T-cells (T-cells producing multiple effector
cytokines) was higher in individuals with severe SARS-CoV-1 infection than in those with
moderate infection, indicating that the level of T-cell response corresponds with the severity of
this infection and time to recovery (15). A recent study in recovered COVID-19 patients revealed that even in the absence of antibodies to SARS-CoV-2, a robust T-cell immune response was measured, indicating the importance of T-cell immunity in response to COVID-19 (16). In particular, T-cell activation/exhaustion and lymphopenia were associated with severe disease whereas traditional effector functions of CD8+ T-cells were related to a better prognosis (17). As the cellular responses during COVID-19 are complex, longitudinal assessment of both CD4+ and CD8+ T-cell responses can inform how SARS-CoV-2 infection and vaccines for this disease modulate immune functions over time.

DNA vaccines containing the S gene derived from SARS-CoV-1 and Middle East Respiratory Syndrome (MERS)-CoV have been shown to induce T-cell responses in mice and humans, respectively (18-20). In addition, SARS-CoV-1 nucleocapsid (NC) protein has been shown to stimulate strong T-cell responses in monkeys and mice (21-25). Similarly, specific NC-derived peptides have been shown to induce cellular response from both CD4+ and CD8+ T-cell subsets derived from patients recovered from SARS-CoV-1 and SARS-CoV-2 (16, 26-29). All current SARS-CoV-2 vaccines include the S protein and a robust T-cell immunity against NC-derived peptides can be detected in convalescent COVID-19 patients (1-6, 16, 26, 29). Therefore, identifying T-cell epitope derived peptides within these two viral proteins will provide effective tools for measuring T-cell responses in COVID-19 patients with different degrees of disease severity and/or evaluating immunogenicity of vaccine candidates in clinical trials.

There are several challenges when developing immunogen peptides for the assessment of SARS-CoV-2 specific T-cell immunity that can be applied globally. First, the genetic profile of SARS-CoV-2 can be region-specific, complicating the detection of the T-cell immunity against global viral variants (30, 31). Second, SARS-CoV-2-reactive CD4+ T-cells were identified in 40-60%
of unexposed individuals, suggesting cross-reactive T-cell recognition between the four circulating human coronaviruses (229E, HKU1, NL63, and OC43) which cause the common cold and SARS-CoV-2 (32-34). Third, Human Leukocyte Antigen (HLA) alleles are extremely polymorphic with more than 18,000 HLA class I and 7,000 class II alleles currently reported (35). Keeping these three points in mind, in order to measure SARS-CoV-2 specific T cell immune responses, it will be important to select SARS-CoV-2 peptide antigens for T-cell response tests that reflect all global viral variants and are not cross-reactive with other human coronaviruses, rather than using complete libraries of overlapping peptides. In addition, selecting peptide antigens that can bind to diverse HLA alleles will be critical.

Importantly, recent research has highlighted the significance of regions within a viral protein comprised of “highly networked” amino acids (36). These regions have topological importance to tertiary and quaternary viral protein structure and are not frequently mutated. In human immunodeficiency virus (HIV) infected individuals with diverse HLA class I alleles, targeting epitopes from these highly networked regions with cytotoxic T cells provided virological control. Therefore, determining immunogenic peptides derived from highly networked regions of the SARS-CoV-2 proteins will be a priority to ensure coverage of all emerging strains of SARS-CoV-2. In addition, a combination of these immunogenic peptides will be superior to using the whole viral spike protein when measuring T-cell responses in patients recovered from COVID-19 as they will increase the sensitivity of the assay to worldwide strains, and will distinguish SARS-CoV-2-specific responses from cross-reactive seasonal coronavirus responses. Moreover, the combination of these highly networked T-cell epitope derived peptides has the potential to bind to diverse HLA alleles.
In the current study, we applied an immunoinformatics analysis pipeline to define immunodominant epitopes in currently circulating SARS-CoV-2 viral variants. These epitopes are restricted to HLA class I and II molecules and selected from topologically important regions of the NC and S proteins with the goal of identifying immunogenic peptides that can contribute to the development of assays for SARS-CoV-2 specific T-cell immunity in patients with different disease severity and after vaccination.

**Results**

*HLA class I restricted T-cell epitopes derived from SARS-CoV-2 nucleocapsid and spike proteins*

We defined 9-mer T-cell epitopes restricted to the HLA-A*02:01 allele which are conserved across geographic SARS-CoV-2 variants. HLA-A*02:01 was selected due to its world-wide prevalence (Figure 1.) (37). These epitopes matched >95% of circulating SARS-CoV-2 variants globally (The Global Initiative on Sharing All Influenza Data, GISAID, as of August 2020) (38-40). Employing our immunoinformatic analysis pipeline, we then identified those epitopes from genetically conserved regions which were comprised of amino acid residues with topological and spatial importance within the NC and S proteins of SARS-CoV-2 (Figure 2).

We computed the network and spatial importance of the T-cell epitopes within N-terminal RNA binding and C-terminal dimerization domains modelled from the NC consensus protein sequence (Figure 2) (41, 42). We identified a total of six T-cell epitope derived peptides (9-mer) from the highly networked and conserved regions of the NC protein (four from the N-terminal domain and two from the C-terminal domain; Table 1, Figure 3--a and b). Of note, together these six peptides can bind to multiple HLA class I alleles which cover approximately 90% of the global population (Figure 4--a and b). We also compared all six peptides from the NC protein with high
network scores and percent bind levels (i.e. a high binding capacity to HLAs) to the SARS-CoV-2 epitopes identified within recent publications and the Immune Epitope Database and Analysis Resource (IEDB) (Table 1 and Figure 5) (16, 26, 29, 43, 44). We found three unique T-cell epitope derived peptides which did not contain the full complement of consecutive amino acid residues identified by these recent studies and the IEDB (16, 26, 29, 43, 44).

For the N-terminal domain, we identified a total of four T-cell epitope derived peptides that had high network scores above the acceptable threshold (Table 1 and Figure 3-a). This indicates that these T-cell epitope derived peptides are found within the core areas of the tertiary structure of the N-terminal domain. Also, these peptides identified from the N-terminal domain had a binding capacity to the HLA-A*02:01 molecule above the 95-percentile cut-off (bind levels, 95.2% to 96.6%; Table 1). Moreover, three of these peptides had strong to moderate binding affinities to HLA class I (IC$_{50}$) of 711 to 1410 nM (Table 1). However, only one of these T-cell epitope derived peptides (IIWVATEGA) was considered a promising immunogenic peptide as it reached our threshold of the top 5% for the HLA class I antigen processing and presentation parameters (i.e. immunogenicity prediction and peptide:HLA complex stability; Table 1). Of note, this peptide selected by our immunoinformatics algorithm is unique and has not been identified by recent studies (Figure 5) (16, 26, 29, 43, 44).

For the C-terminal domain, we identified two T-cell epitope derived peptides (ILLNKHIDA and RTATKAYNV) that comprised of highly networked amino acid residues and high binding levels to the HLA class I molecule (Table 1 and Figure 3-b). Also, these peptides reached our threshold of the top 5% for the predicted HLA class I antigen processing and presentation parameters (i.e. stability of the peptide:HLA complex; Table 1), making them the most promising candidates for T-cell immunity assays from the C-terminal domain. Taken together, three T-cell epitopes...
(IIWVATEGA, RTATKAYNV and ILLNKHIDA) among the peptide repertoire derived from the NC protein had the most promising properties for HLA class I restricted antigen presentation. For these three highly networked T-cell epitope derived peptides selected from the NC protein, we then assessed whether they could bind to additional HLA class I alleles other than HLA-A*02:01 (Figure 4-a). Our analysis predicted that these peptides had binding capacity for two to six additional HLA-A and HLA-B alleles which are classified into different HLA class I supertypes (45). The peptides individually can cover approximately 41% to 55% of the global population (Figure 4-b).

For the S protein, we found only 2.6% of the T-cell epitope derived peptides contain highly networked amino acid residues with a binding potential to the HLA class I molecule (29 out of 1112 9-mer peptides; Table 1 and Figure 3-c). Importantly, none of these peptides were derived from the regions within the S protein which contain mutations reported to enhance viral infectivity (Figure 6) (30). In addition, the peptides had the potential to bind to 23 additional HLA class I alleles which are classified into 10 supertypes (Figure 4-c) (45). These 29 peptides selected from the topologically important regions of the S protein can cover up to 96% of the global population when combined (Figure 4-d). A total of 12 out of 29 T-cell epitope derived peptides selected by our immunoinformatics analysis pipeline contained consecutive amino acid residues which were 100% identical to the epitopes identified by other groups (Figure 5) (16, 26, 29, 43, 44).

Of these 29 highly networked T-cell epitope derived peptides determined from the S protein, we identified 22 peptides which reached the top 5% for the HLA class I antigen processing and presentation parameters (Table 1). This indicates that the protein regions comprised of these epitopes are the most promising sites for T-cell immune recognition. However, we found that 10
of these 22 T-cell epitopes have been previously described by other groups (Table 1 and Figure 5) (16, 26, 29, 43, 44). This immunoinformatics pipeline allowed us to identify 12 unique T-cell epitopes within the S protein as promising for HLA class I restricted immune response.

Highly networked T-cell epitopes derived from the SARS-CoV-2 nucleocapsid are correlated with HLA class I antigen processing and presentation parameters

To compare if the structural topology of T-cell epitopes correlates to the HLA class I related antigen processing and presentation, we determined the association between the highly networked T-cell epitope derived peptides within the SARS-CoV-2 tertiary structures and the HLA class I mediated immune restriction (Figure 7). We found the network scores for the T-cell epitope derived peptides from the NC, particularly the N-terminal domain, were positively correlated with HLA class I antigen processing and presentation parameters (p-value=<0.0001-0.282; Figure 7-a to f). In contrast, the highly networked T-cell epitope derived peptides from the S protein were not associated with these parameters (p-value=0.136-0.742; Figure 7-g to i). However, our immunoinformatics analysis pipeline identified a subset of the 9-mer T-cell epitope derived peptides from the S protein which were promising for HLA class I restricted immune response (n=22, Table 1).

HLA class II restricted T-cell epitopes derived from SARS-CoV-2 nucleocapsid and spike proteins

To define T-cell epitopes restricted to DRB1*07:01, HLA class II allele, we used the same immunoinformatics analysis pipeline mentioned above incorporating HLA class II binding prediction (36, 46, 47). This HLA class II allele was selected due its world-wide prevalence (Figure 8) (37). The HLA class II molecules frequently accommodate peptides between 13 and 17 amino acids in length (48, 49). However, we focused on the peptide repertoire composed of
12 amino acid residues as this particular length is substantially associated with high binding
affinity to HLA class II molecules (50). Of note, the T-cell epitope derived peptides that we
identified match >95% of circulating SARS-CoV-2 variants globally (GISAID as of August
2020) (38-40).

From a repertoire of 120 peptides derived from T-cell epitopes of the N-terminal domain of NC,
we found five promising peptides which contained highly networked amino acid residues
(network score range of 26.4 to 41.7, threshold of 24.6; Table 2 and Figure 9-a). These five
peptides had high binding potentials to the HLA class II molecule (bind levels of 91% to 95%;
and IC$_{50}$ of 188 nM to 325 nM). In addition, these peptides scored within the top 5% bind level
to 18 additional HLA class II alleles classified into two to four HLA-II loci (Figure 10-a) (51).
Individually, these five peptides are predicted to cover approximately 80% of the global
population (Figure 10-b). When all of the five peptides were combined, a global population
coverage of 95% was predicted. Of these peptides derived from the highly networked regions of
the NC N-terminal domain, three were found within the same B cell epitope (Table 2). Of note,
only two of the five peptides derived from the NC N-terminal domain contained the full
complement of consecutive amino acid residues identified by other studies (Figure 11) (16, 26,
29, 43, 44). In contrast, none of the peptides derived from the C-terminal domain of the NC
scored above the thresholds for both the protein network score and percent bind level (Figure 9-
b).

Out of 1109 peptides derived from the S protein, we found 27 HLA class II restricted T-cell
epitope derived peptides with high network scores (Table 2 and Figure 9-c). These peptides had
percent bind levels above the cut-off as well as predicted IC$_{50}$ values of 55.3 nM to 162.6 nM,
making these peptides promising candidates for T-cell immunity assays. Importantly, none of
these peptides were derived from the regions within the S protein which contain mutations that have the potential to enhance viral infectivity (Figure 6) (30). Also, we found that these peptides can bind to 24 additional HLA class II alleles classified into four loci (Figure 10-c) (51). When all peptides are combined, approximately 100% global population coverage is predicted (Figure 10-d). Of these peptides, six were found in two B cell epitopes (Table 2). Moreover, we found a total of eight which were 100% identical to the epitopes identified by other studies (Table 2 and Figure 11) (16, 26, 29, 43, 44). This immunoinformatics analysis pipeline allowed us to identify 19 new T-cell epitopes within the S protein as promising immunogenic peptides for the assays that detect T-cell immunity against SARS-CoV-2.

Identification of highly networked T-cell epitope derived peptides within SARS-CoV-2 NC and S proteins with low homology to seasonal human coronaviruses

To identify SARS-CoV-2 peptide antigens that can be used to assess COVID-19 specific T-cell responses and differentiate from cross-reactive seasonal coronavirus immunity, we compared the highly networked SARS-CoV-2 T-cell epitope derived peptides to four seasonal human coronaviruses (229E, HKU1, NL63 and OC43) (32-34).

For the highly networked T-cell epitope derived peptides identified within the NC protein, the most promising peptides for HLA class I (9-mers, n=3) and HLA class II (12-mers, n=5) had 0-56% homology to four human coronaviruses (Figure 12). Of the three peptides restricted to HLA class I immune response, ILLNKHIDA had no homology to any of the seasonal human coronaviruses (Figure 12-a to d). In addition, all of the highly promising peptides restricted to HLA class II immune response had homology of less than 20% when compared to two seasonal coronaviruses (HKU1 and OC43; Figure 12-e to h).
The highly networked T-cell epitope derived peptides identified from the S protein that were most promising for HLA class I (9-mers, n=22) and class II (12-mers, n=27) had 0 to 78% homology to the four seasonal coronaviruses (Figure 12). Of note, three of the 9-mer peptides (SIIAYTMSL, VTWFHAIHV and VTWFHAIHV) and 14 of the 12-mer peptides showed low homology to all four human coronaviruses (Figure 12). Furthermore, a majority of 9- and 12-mer peptides had less than 20% homology when compared to seasonal coronaviruses 229E and NL63.

Although the degree of homology between the highly networked T-cell epitope derived peptides and the four seasonal human coronaviruses is different, our immunoinformatics analysis pipeline allowed us to identify 18 peptides which were highly specific to SARS-CoV-2 (Figures 5 and 11). Of these, we identified 11 T-cell epitope derived peptides that were unique and highly promising for HLA class I and II restricted immune responses that are SARS-CoV-2 specific (16, 26, 29, 43, 44).

**Highly networked T-cell epitope derived peptides bind to the HLA-A*02 molecules as predicted**

To assess whether our binding prediction is accurate for the highly networked T-cell epitope derived peptides, we selected the two most promising T-cell epitope-derived peptides from the NC protein based on their HLA-I binding. In addition, these peptides were selected based on their predicted scores for the HLA class I antigen processing and presentation parameters. The peptide from the N-terminal region was selected for the *in-vitro* binding validation as it had the highest bind level and the longest predicted time required for the peptide-HLA complex to dissociate (IIWVATEGA; Table 1). Also, this T-cell epitope derived peptide has not been identified by recent studies (Figure 5) (16, 26, 29, 43, 44). The peptide from the C-terminal region was chosen as it scored the highest values for three of the four HLA class I antigen
processing and presentation parameters (RTATKAYNV; Table 1). The ability of these peptides
to bind to HLA-A*02:01 in-vitro was assessed using a human-derived TAP-deficient T2 cell line
(T2 cells) expressing HLA-A*02:01 on the cell surface (52, 53).

These highly networked NC-derived peptides were able to bind and stabilize the HLA-A*02
molecules on the surface of the T2 cells (Figure 13-a and -b). The mean fluorescence intensity of
the peptide:HLA-A*02 detection exceeded the non-HLA-A*02 binding control (i.e. negative
control) across a series of peptide concentrations. In particular, the binding of the peptide from
the NC N-terminal domain to these HLA-A*02 molecules (IIWVATEGA) was similar to the
positive control EBV peptide across all concentrations (Figure 13-b).

**Effector and polyfunctional response of CD8+ T-cells to highly networked T-cell epitope
derived peptides**

There is growing evidence for the importance of CD8+ T-cells in control of SARS-CoV-2
disease severity (17). Therefore, we focused on testing the two HLA-A*02 restricted T-cell
epitope derived peptides. The immunogenicity of these highly networked peptides selected from
the NC protein (IIWVATEGA and RTATKAYNV) was tested in peripheral blood mononuclear
cells (PBMCs) obtained from participants 1-2 months after recovery from SARS-CoV-2
infection (Table 3). When the PBMCs from two convalescent participants who were HLA-A*02
positive were stimulated by the peptides, we observed a robust production of interleukin-2 (IL-
2), interferon-gamma (INF-γ), tumor necrosis factor-alpha (TNF-α) and a marker for the
degranulation of CD8+ T-cells (CD107a/b) (Figure 14-a to d). This effector response was
specific for HLA-A*02 (HLA-A*02:01 and HLA-A*02:06) as no such effector response was
observed in the HLA-A*02 negative donor (HLA-A*24:02) (Table 3, Figure 4-a, and Figure 14-
a and -c). Of note, 28-62% of the responding CD8+ T-cells were polyfunctional exhibiting four effector functions simultaneously (Figure 14-b and d).

Discussion

In order to accelerate the development of a diagnostic assay that can measure T-cell immune responses against global SARS-CoV-2 strains, we identified specific T-cell epitopes which are conserved across circulating viral variants from six global regions. In particular, these epitopes contained amino acid residues that are highly networked indicating their topological importance within the NC and S proteins of the virus (36, 46, 54). By performing an immunoinformatics analysis, we defined 57 T-cell epitopes within the SARS-CoV-2 proteins, of which 11 were unique and non-cross-reactive to seasonal human coronaviruses, that should be considered for detecting a combined SARS-CoV-2 specific CD4/CD8 T-cell immune response (16, 26, 29, 43, 44). Importantly, these highly networked T-cell epitope derived peptides were identified from the regions that lack mutations reported to enhance viral infectivity (30). In addition, we assessed key antigen processing and presentation parameters to further delineate the T-cell epitopes which are most likely to induce an HLA class I mediated immune response. In conducting this study, we selected the T-cell epitopes restricted to the HLA-A*02:01 and DRB1*07:01 alleles based on their global frequencies. Furthermore, we validated in-vitro the binding of two HLA-A*02 specific T-cell epitope derived peptides from the highly networked regions of the NC protein to the T2 cell line expressing HLA-A*02:01 molecules. Also, we assessed the HLA class I mediated CD8+ T-cell immune response to these peptides by using PBMCs obtained from two SARS-CoV-2 patients 1-2 months post-recovery.

The NC and S structural proteins are highly homologous between the viruses from the Coronaviridae family due to their importance for viral replication (55-58).
identified a number of T-cell epitopes that are conserved between SARS-CoV-1 and -2 (43). Among the approximately 600 SARS-CoV-2 protein sequences derived from six global regions, we identified highly networked T-cell epitopes that matched more than 95% of the circulating SARS-CoV-2 variants (38-40). This allowed us to select the T-cell epitopes for further analysis with the potential for a universal tool that can detect T-cell responses to world-wide strains of SARS-CoV-2. In addition, these T-cell epitopes are identified from the topologically important sites where molecular interactions between amino acid residues are critical for maintaining the structure and function of the viral proteins; and therefore, these sites are not frequently mutated (36, 59-61). As these sites are mostly found in the core of the proteins, the highly networked T-cell epitopes selected from these regions are most likely to be protected from proteasomal and lysosomal degradation pathways that shape the T-cell epitope repertoire (62-64).

Currently, a total of 14 mutations have been reported within S protein sequences (30). These mutations define important SARS-CoV-2 clades currently reported in GISAID. Some of these have been predicted to enhance viral infectivity of target cells expressing angiotensin-converting enzyme 2 (ACE2) (30). Most of these mutations are found within subunit 1 of the spike protein where the receptor binding domain (RBD) is located. Also, these mutations define region specific SARS-CoV-2 clades. As our immunoinformatics analysis pipeline selects T-cell epitope derived peptides with high network scores that avoids these mutations, these peptides will detect T-cell immunity regardless of SARS-CoV-2 clade.

By applying our immunoinformatics analysis, we found the T-cell epitopes derived from topologically important regions of the NC (high network scores) correlate with the HLA class I antigen processing and presentation parameters. This indicates the peptides from the NC are likely to induce HLA class I restricted CD8+ T-cell responses (26, 29, 33). In agreement with
this correlation, we observed effector and polyfunctional responses from the CD8+ T-cells of two SARS-CoV-2 convalescent participants to two peptides from the NC protein. This suggests that our immunoinformatics analysis pipeline identifies immunodominant regions within the SARS-CoV-2 NC protein.

Recent phase I immunogenicity and safety trials of vaccine candidates encoding for the S protein have been shown to induce neutralizing antibodies and IFN-γ T-cell response to SARS-CoV-2 (2-4, 6, 65). However, the polyfunctionality of this T-cell response is unknown. Also, it has been shown that the mRNA vaccine encoding spike-RBD induces primarily CD4+ Th1-type responses (2). Whether these SARS-CoV-2 vaccines induce polyfunctional CD8+ T-cells is unclear. This prompted us to identify T-cell epitopes restricted to the HLA class I and II alleles within the S protein of SARS-CoV-2 that can be used to detect polyfunctional T-cell responses. For the S protein, we found 22 T-cell epitope derived peptides most promising for HLA class I restricted immune response despite the lack of correlation between network scores and HLA class I antigen processing and presentation parameters. In addition, the 12-mer peptides derived from the epitopes within the S protein are predicted to bind to HLA class II alleles and could stimulate CD4+ T-cell response (29, 33). In particular, six of these 12-mer peptides were sequestered in two B-cell epitopes suggesting their importance when assessing CD4+ T-cell response against SARS-CoV-2. In the future, we will validate the CD4 immune response to these 12-mer peptides using the PBMCs obtained from SARS-CoV-2 convalescent participants.

A recent study by Moderbacher et al. has shown that COVID-19 disease severity is associated with delayed and/or limited SARS-CoV-2 specific CD4+ and CD8+ T-cell responses during acute infection (66). In contrast, less severe disease is strongly related to a higher proportion of effector CD8+ T-cells that can produce INF-γ, an important antiviral cytokine in mucosal sites.
(66, 67). However, a longitudinal assessment of SARS-CoV-2 specific T-cell immunity during both the acute and chronic phases of COVID-19 can further delineate the cellular immune response against SARS-CoV-2 and its association with disease severity. In particular, employing the NC and S protein-derived immunogenic peptides that have low homology to seasonal human coronaviruses will allow for the detection of cellular immune responses that are absolutely specific to SARS-CoV-2. Importantly, the highly networked and conserved SARS-CoV-2 specific immunogenic peptides defined from the NC and S protein sequences derived from global viral variants can contribute to this longitudinal assessment of T-cell immunity.

The T-cell epitope derived peptides defined by our immunoinformatics analysis pipeline can also contribute to the development of a “second-generation” vaccine that aims to stimulate combined CD4/CD8 T-cell immune responses (18, 68). The levels of SARS-CoV-2 neutralization antibodies alone do not determine protection against the virus (66). Rather, a coordinated approach that can mount both the virus specific antibodies and CD4/CD8 immune responses will be effective against SARS-CoV-2 (66, 69-71). Therefore, the immunogenic peptides selected from the highly networked and conserved T-cell epitopes within the NC and S proteins via our analysis pipeline could be considered as vaccine candidates to elicit CD4/CD8 T-cell immune responses against SARS-CoV-2.

There are several limitations to our study. First, our immunoinformatics analysis pipeline was applied to two specific HLA alleles. However, we identified highly networked T-cell epitopes derived from the NC and S proteins that are predicted to bind to 18 to 24 additional HLA class I and II alleles classified into multiple HLA supertypes/loci. Importantly, these can cover approximately 80-100% of the global population, indicating that these peptides can be used internationally for the T-cell immunity detection assays. Second, the protein homology...
modelling of the NC N-terminal RNA binding domain was based on the SARS-CoV (41). However, this SARS-CoV protein domain was the best predicted template when performing our analysis (41, 72-74). Third, even though we only assessed 12-mer peptides for the HLA class II allele, a previous study has shown that this length can accurately predict the binding to HLA class II molecules (50). Of note, longer peptides which are frequently presented by the HLA class II molecules should include the 12-mer peptides that we identified as being highly networked (51, 75). Lastly, all of the promising T-cell epitope-based peptides identified by our immunoinformatics analysis pipeline require *in-vitro* and *ex-vivo* assessments to determine if they induce T-cell responses during and after SARS-CoV-2 infection (9, 15). However, our interim *in-vitro* validation showed that our immunoinformatics analysis pipeline has identified two T-cell epitope derived peptides from the NC protein that can stably bind to the HLA-A*02 molecules. Furthermore, the CD8+ T-cells derived from the SARS-CoV-2 survivors exhibited polyfunctional effector responses to these peptides, which have not been previously identified as promising epitopes for T-cell mediated immune response by recent studies (16, 26, 29, 43, 44). These *in-vitro* and *ex-vivo* cellular binding and response studies provide proof of concept that our immunoinformatics analysis pipeline identifies novel T-cell epitopes which can elicit a SARS-CoV-2 specific T-cell response which was polyfunctional. Our future studies will assess T-cell mediated responses to all of the highly networked T-cell epitope derived peptides identified by our immunoinformatics analysis pipeline. Of note, as the selected peptides are identified within the core of the viral proteins, a subset of them have the potential to be hydrophobic which affects peptide-synthesis. This issue can be overcome by extending the length of the T-cell epitope derived peptides so that they include hydrophilic amino acid residues.
In conclusion, the application of an immunoinformatics analysis pipeline allowed us to identify 421 highly networked T-cell epitopes, of which 11 were unique and non-cross-reactive to seasonal human coronaviruses, from the NC and S proteins which are promising immunogenic peptides for detecting HLA class I or II related immune responses. Of these peptides, two novel T-cell epitopes from the NC can stably bind to HLA class I molecules and induce polyfunctional effector CD8+ T-cell responses. Our findings indicate that our immunoinformatics analysis pipeline can contribute to the development of assays that detect polyfunctional and SARS-CoV-2 specific T-cell responses against diverse SARS-CoV-2 viral strains, distinct from pre-existing seasonal coronavirus immunity. The T-cell immunity assay using our peptides have the potential to detect T-cell immune responses elicited by diverse HLA polymorphisms.

Materials and Methods

Study approval

This study was carried out in accordance with the recommendations of the institutional review board at the Western Sydney Department for the Westmead Institute for Medical research (WSLH HREC 2020/ETH00844 and 2020/STE01476). The protocol was approved by this committee. All participants provided written informed consent prior to inclusion in the study.

Global and regional distribution of HLA class I and II alleles

An understanding of the world-wide distribution of HLA class I and HLA class II alleles is important when selecting immunodominant epitopes for vaccine candidates against the SARS-CoV-2 global pandemic. Currently, there are more than 18,000 HLA class I and 7,000 class II alleles reported (35). Therefore, we determined the five most prevalent HLA class I and HLA class II alleles worldwide by data curation from The Allele Frequency Net Database.
This allowed us to select the dominant HLA-I A*02:01 (HLA-A*02:01) and HLA-II DRB1*07:01 (DRB1*07:01) alleles for inclusion in our immunoinformatics analyses (Figures 1 and 8). The global frequency of these alleles was 20% and 12%, respectively. As the most severely affected regions for SARS-CoV-2 infection are found in Europe and the Americas (76), we applied an in-silico immunoinformatics analysis to identify T-cell epitopes within the NC and S protein sequences which are effective for HLA-A*02:01 and DRB1*07:01 mediated antigen restriction (Figure 2).

**Genetic variability of circulating SARS-CoV-2**

A universal vaccine targeting different strains of coronavirus is also desirable as new viral strains can emerge from animal-to-human transmission and region-specific genetic diversification. To identify genetically conserved T-cell epitopes for possible vaccine development, we examined the genomic diversity of circulating SARS-CoV-2 isolates from six global regions. Whole genome sequences (WGS) from NSW, Australia (n=15) (77) were combined with local and global references available from GISAID (https://www.gisaid.org/) (39, 40). As of March 2020, we obtained a total of 607 SARS-CoV-2 genomic sequences from GISAID and aligned them using MAFFT (78). These sequences contained representatives from all major lineages (https://github.com/hCoV-2019/pangolin). The regions encoding the nucleocapsid (NC) and Spike (S) protein sequences were translated and extracted with ambiguous positions removed. The resulting alignments contained 586 and 567 sequences for the NC and S proteins, respectively. These NC and S protein alignments represented data from Asia (n=178), Europe (n=238), North America (n=137), South America (n=10), Africa (n=1) and Oceania (n=39).

A sliding window approach was used to identify all possible 9- and 12-mer peptides derived from the consensus sequences of the N- and C-terminal domains of NC, and the S protein.
percent identity of each amino acid within these 9- and 12-mer peptides was calculated using the Geneious version 8.1.9 (https://www.geneious.com) (79). From this value the genetic variability of each peptide was then determined (100-percent identity). The peptides with 0% genetic variability were selected for further immunoinformatics analysis (Figures 2, 3 and 9). We also compared our T-cell epitope derived peptides to the sequences of SARS-CoV-2 circulating variants listed in the GISAID as of August 2020 (39, 40).

**Protein structure homology modelling of SARS-CoV-2 NC and S proteins**

We modelled three-dimensional structures of the NC and S consensus protein sequences derived from the alignments by using SWISS-MODEL (https://swissmodel.expasy.org/) (Figure 2) (72-74). Using homology-structure modelling, we predicted the N-terminal RNA binding domain of the NC protein structure by using SARS-CoV nucleocapsid template (PDB: 1SSK, STML ID: 1SSK.1.A) as this was the accurate template for this region (41). For the C-terminal dimerization of the NC protein, the protein structure was modelled by using SARS-CoV-2 coronavirus as a template (PDB: 7C22, STML ID: 7C22.1.B) (42). We performed automated structure homology-modelling on the S consensus protein sequence by using chain A of SARS-CoV-2 spike protein as the template (S protein: PDB: 6VSB.1; STML ID: 6VSB.1.A) (74). As the part of this protein modelling, all protein regions were investigated for their genetic identity to SARS coronavirus and other organisms. We only used the protein regions that were related to SARS coronavirus to define T-cell epitopes.

**Identifying highly networked epitopes within the NC and S proteins**

To identify suitable T-cell epitopes as targets for the T-cell immunity assay, we used an immunoinformatics pipeline that combines protein structure-based network analysis and sequence-based HLA class I and II binding prediction within the non-variable NC and S regions.
of SAR-CoV-2 (Figure 2) (36, 46, 47, 54). We used the Network Analysis of Protein Structures (NAPS) program (http://bioinf.iit.ac.in/NAPS) (80) to quantify structural and spatial importance of each amino acid residue within the tertiary structure models of the N- and C-terminal domains of nucleocapsid and S protein (Figure 2). We combined networks defined by geometric center (“atom pair contact”) and center of mass (“centroid”) for each amino acid residue within the SARS-CoV-2 protein model (80, 81). The “atom pair contact network” describes physicochemical interactions between an atom-pair within an amino acid residue and the “centroid network” defines the connectivity between center of mass of any two amino acid residues within a protein structure (80, 81). We also calculated the distance from the center of mass of each tertiary protein model to all amino acid residues presented in the protein structure by using CALCOM (http://bioinformatica.isa.cnr.it/CALCOM/input.html) (82-84). This allowed us to quantify the spatial location of each amino acid residue with respect to the center of the tertiary structure of each protein (82-84). We used a total of five parameters derived from NAPS and CALCOM when calculating a network score for each amino acid residue within each protein structure. These parameters are as follows:

1) Number of direct neighbors of a geometric center of an amino acid residue (Degree\textsubscript{Atom pair contact network}).

2) Number of direct neighbors of a center of mass of an amino acid residue (Degree\textsubscript{Centroid network}).

3) A ratio of the degree of interconnectivity: (Betweeness\textsubscript{Centroid network}) calculated by (number of the shortest pathways between a particular amino acid residue and its neighbors)/ (the total number of the shortest pathways within the protein centroid).
network). Each pathway is weighted based on the distance between two amino acid residues.

4) A cumulative intermolecular strength of all neighboring amino acid residues connected to a geometric center of a particular amino acid residue within the atom pair contact network ($\text{Strength}_{\text{Atom pair contact network}}$).

5) A distance from the center of mass to an amino acid residue within a protein structure (Distance).

By modifying the previously published equations (36), we determined a network score for each 9- and 12-mer peptide derived from the N- and C-terminal domains of the NC and S proteins by using the equation below:

$$\text{Network score for each amino acid residue} = \frac{(\text{Degree}_{\text{Atom pair contact network}} + \text{Degree}_{\text{Centroid network}})}{4} + \frac{\text{Betweeness}_{\text{Centroid network}} + \text{Strength}_{\text{Atom pair network}}}{2} - \text{Distance}$$

All proteins are subjected to proteasomal and lysosomal degradation processes before generating peptide repertoires for further HLA antigen presentation (62-64). To select the peptides which are protected from these degradation pathways, our calculation weighted the position of the amino acids within a protein structure more than the connectivity between amino acid residues.

For 9-mer or 12-mer peptides derived from each viral protein, we summed the network scores for each amino acid residue and divided by the corresponding length of the peptides. We then normalized the network scores by subtracting the lowest value. In addition, we calculated the lower 95% confidence interval for mean of the normalized network scores for each peptide length and tertiary protein model. The peptides with a normalized network score above this 95%
confidence interval were considered highly networked in this study. The peptides comprised of highly networked amino acid residues within the NC and S proteins were compared across the SAR-CoV-2 isolates available. The 95% confidence intervals for the mean of the normalized network scores for each peptide length and tertiary protein model was computed by using STATA 15.1 (StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC) (85). The normalized network scores for each peptide is presented in the main figures and tables.

The consensus sequences of NC-N terminal, NC-C-terminal and S proteins, and peptide sequences (9- and 12-mers) are available on https://github.com/EunokLee/SARS-CoV-2_data_files_and_acknowledgements. Also, the PDB files from the protein homology modelling can be downloaded from the same github link.

**HLA class I or II binding affinity**

**Percent bind level**

Peptides representing the immunodominant epitopes with the best network scores were screened *in silico* for their HLA class I or II binding affinity using NetMHCpan-4.0 and NetpanMHCII-3.2, respectively (Figure 2) (46, 54). These programs order peptides from the strongest binders to those which do not exhibit any binding capacity to an HLA molecule by predicting their binding affinities (half maximal inhibitory concentration; IC50) (47). Based on the IC50, the algorithm ranks the peptides from the strongest binder to weakest binder. For example, those which are categorized as the top 2-10% bind rank are considered as binders to HLA class I and II molecules. For the data presentation in this study, we subtracted this percent rank from 100% and presented the resulting value as a percent bind level for each peptide. In other words, the binders to HLA class I and II molecules have a high percent bind level whereas the non-binders
have a low percent bind level. To standardize our selection method for the peptides with predicted binding capacity to HLA class I and II molecules, we calculated the percent bind level that equates to the 95-percentile threshold for each peptide repertoire. This repertoire consists of 9- or 12-mer peptides derived from the NC N-terminal domain, NC C-terminal domain, or S protein. This percentile method allowed us to select the peptides with percent bind levels which are categorized into the top 5% bind level. The 95-percentile was determined by normal quantile plot (STATA 15.1; StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC) (86).

**IC50 cut-off**

The majority of T-cell epitopes have binding affinities lower than an IC50 of 5000 nM according to the Immune Epitope Database and Analysis Resource (IEDB) MHC-I binding predictions (http://tools.iedb.org/mhci/). Therefore, we used the IC50 of less than 5000 nM as the cut-off in addition to our 95-percentile bind level threshold to determine the peptides with binding potential to HLA class I and II molecules.

**Additional HLA class I related antigen processing and presentation prediction**

By using IEDB combined predictor (http://tools.iedb.org/processing/), we scored the 9-mer peptides derived from the NC and S proteins that were most likely to be processed for HLA class I mediated antigen presentation (Figure 2) (87). The IEDB algorithm generates predicted proteasomal cleavage and transporter associated with antigen processing (TAP) scores for individual peptides. High proteasomal cleavage and TAP scores indicate efficient antigen presentation (88). Due to the arbitrary grading for these scores, we derived proteasomal cleavage and TAP scores that equate to the 95-percentile for each 9-mer peptide repertoire derived from NC N-terminal domain, NC C-terminal domain, or S protein. This allowed us to select the
peptides with HLA class I related antigen processing scores which are categorized into the top 5%.

To further delineate T-cell epitopes from the 9-mer peptide repertoire, we predicted HLA class I mediated antigenicity by using an IEDB analysis tool (http://tools.iedb.org/immunogenicity/) (Figure 2) (89). This analysis tool scores the immunogenicity by determining the positions and side chain properties of the amino acid residues within a peptide-HLA complex that binds to a T-cell receptor. We used the default setting when performing the immunogenicity prediction. The peptides with high immunogenicity have high prediction scores. Therefore, we used the immunogenicity score that equates to the 95-percentile as the cut-off to identify the peptides with top 5% scores. In addition, we predicted the stability of a complex formed by the peptide and HLA class I molecule (p:HLA) by using NetMHCstabpan-1.0 (90). The predicted stability of p:HLA is reported as the time required for the peptide to dissociate from the HLA class I molecule (p:HLA t_{1/2}). The percentile cut-offs for HLA class I related antigen processing and immunogenicity were determined by normal quantile plot (STATA 15.1; StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC) (86).

Selection of T-cell epitope derived peptides for HLA class I and II immune responses

Those T-cell epitopes selected as promising candidates that can contribute to the development of T-cell immunity assays specific for SARS-CoV-2 had the following parameters:

1) Peptide genetic variability of 0%
2) Peptide network scores above the threshold (i.e. above lower 95% confidence interval).
3) Percent bind level to HLA-A*02:01 or DRB1*07:01 above the threshold (i.e. within the top 5%).
4) Half-maximal inhibitory concentrations (IC50) of less than 5000 nM derived from HLA binding prediction algorithms.

5) For HLA class I epitopes, at least one score predicted for antigen processing and presentation reaching the top 5%.

**Correlation of the network scores to the HLA class I antigen processing and presentation parameters**

We performed a correlation analysis between network scores and predictions for HLA class I mediated antigen processing and presentation parameters (i.e. proteasomal processing score, TAP score and HLA-I immunogenicity). The correlation analysis was performed by using STATA 15.1 (StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC). This analysis was applied to peptide repertoires derived from the N- and C-terminal domains of NC and to the S protein. We investigated the $R^2$ values to determine the proportion of the peptide repertoire that follows linear regression. Moreover, we determined the association between the network scores and the HLA class I related antigen processing and presentation parameters by the slopes of the regression.

**Sequence comparison between SARS-CoV-2 T-cell epitope derived peptides and seasonal human coronaviruses**

To determine the homology between T-cell epitope derived peptides (Tables 1 and 2) and four seasonal human coronaviruses (hCoVs; 229E, HKU1, NL63 and OC43), we downloaded all available nucleocapsid and spike glycoprotein sequences of these hCoVs from UniProt database (https://www.uniprot.org/). These hCoV sequences (n=1353) were compared to the SARS-CoV-2 NC and S protein sequences containing the highly networked T-cell epitope derived peptides by using Geneious version 8.1.9. For the protein sequences derived from each hCoV strain and
for each region aligning with the SARS-CoV-2 T-cell epitope derived peptide, a mean percent

genetic identity and its 95% confidence intervals were calculated by STATA 15.1 (79, 86).

Assessing the binding capacity of the T-cell epitope derived peptides to multiple HLA class
I and II alleles and their world-wide population coverage

As we selected highly networked T-cell epitope derived peptides based on the most prevalent
HLA class I and II alleles (HLA-A*02:01 and DRB1*07:01), we assessed if these peptides can
bind to additional HLA class I and II alleles by using NetMHCpan-4.1 (91) and NetMHCIIpan-
4.0 (92). For the 9-mer peptides, we predicted the binding affinity to additional HLA-A and
HLA-B alleles which are known to cover more than 97% of the global population (93). For the
12-mer peptides, we predicted the binding affinity to additional HLA class II alleles which are
known to cover more than 99% of the global population (94). For each of the 9- and 12-mer
peptide repertoires derived from the NC and S protein sequences, we derived 95-percentile
threshold of the percent bind level to each of the additional HLA class I and II alleles as
described above. This percentile method allowed us to determine the peptides with the top 5%
percent bind levels to each of these additional HLA alleles. The HLA class I alleles were
categorized into 10 supertypes (45) and the HLA class II alleles were grouped by four loci (51)
(Figs. 2 and 6). For the global population coverage, we used the IEDB analysis tool called
population coverage (http://tools.iedb.org/population/) (95).

B cell epitope prediction

The consensus SARS-CoV-2 NC and S protein sequences were used to predict B cell epitopes by
applying an IEDB analysis tool (http://tools.iedb.org/bcell/). For this analysis, BepiPred-2.0,
Sequential B-cell epitope predictor that employs the epitopes determined from crystal protein
structures (96). The sequential residues with the scores above the threshold of 0.5 were reported as the B cell epitopes in this study.

Sequence comparison with other published SARS-CoV-2 T-cell epitopes

We compared the peptides we detected with high network scores and percent bind level to the HLA class I and II molecules to the T-cell epitopes published by recent studies (16, 26, 29, 43, 44). We aligned these peptide sequences using Geneious version 8.1.9 to identify highly networked peptides which are 100% identical to those recently published (79).

In-vitro validation of HLA-binding capacity of T-cell epitope derived peptides

Peptides

We selected RTATKAYNV and IIWVATEGA from the SARS-CoV-2 NC 9-mer peptide repertoires for in-vitro validation of HLA-binding capacity. These peptides were selected based on their network analysis, bind level and HLA class I mediated antigen processing and presentation parameters. For a positive control, we included an HLA-A*02:01 restricted peptide (NLVPMAVATV) derived from cytomegalovirus (CMV). This positive control peptide was derived from CMV glycoprotein 64, a virion tegument protein that is the main component of the enveloped subviral particle (CMV-pp65). As a negative control, we included an HLA-B*07 restricted peptide, TPRVTGGGAM, selected from the CMV-pp65. All peptides were synthesized from Mimotopes, Australia, at >95% purity. The peptides were suspended in 10% dimethyl sulfoxide (DMSO) and 90% water at a concentration of 10 mM. The suspended peptides were stored at -80 °C until use.

T2 cell line
The HLA class I restriction of these peptides was tested by using non-adherent human-derived T lymphoblastoid hybrid cell line (T2; 174 X CEM.T2; ATCC® CRL-1992™) (52, 53). This cell line is TAP-deficient, expressing empty HLA class I A*02 molecules on the cell surface. The cells were cultured in RPMI-1640 (LONZA, BE12-702F) supplemented with 10% Fetal Bovine Serum (FBS), referred to as RF10. The binding capacity of the peptides to the HLA-A*02:01 molecules was tested when the T2 cells were in log-phase of growth. The T2 cell line was kindly provided by Prof. Rajiv Khanna (QIMR Berghofer Medical Research Institute, Queensland, Australia).

**Hybridoma**

Mouse BB7.2 (ATCC® HB-82™) hybridoma cells were used for producing the primary anti-human HLA-A*02 antibody and for the staining of the HLA-A*02 molecules expressed on the T2 cell surface. The hybridoma was maintained and cultured in RF10. For antibody collection, the cells were washed with phosphate-buffered saline (PBS, LONZA-BE17-516F) and resuspended in AIM-V™ serum free media (Thermo Fisher Scientific, 12055091) at 10^6 cells/mL, followed by an incubation at 37 °C for two days. The supernatant was collected after the cells were pelleted by centrifugation at 300Xg for 5 mins and filtered through 0.45 µm syringe filter (Merck Millipore, Darmstadt, Germany). The supernatant was stored at 4 °C until use.

**HLA-peptide binding assay**

The ability of synthetic peptides to stably bind to HLA-A*02:01 molecules on the cellular surface of the T2 cell line was assessed by flow cytometry as previously described (97). Briefly, T2 cells (1 x 10^5 cells in 100 µl) were incubated for 1 hour at 37 °C in serum-free AIM-V™ medium (Thermo Fisher Scientific, 12055091) in the presence of the peptides at the
concentrations of 0 μM (no peptide control), 1μM, 10μM and 100 μM. The cells were then incubated for 16 hours at 26°C and returned to 37 °C for 2 hours prior to immunofluorescent staining. The unbound peptides were removed by using cold PBS. The anti-HLA-A*02 specific monoclonal antibody (i.e. BB7.2 supernatant) was added to the T2 cells and incubated at 4 °C for 30 mins. After being washed three times with cold PBS, the cells were incubated with a goat secondary Alexa Fluor 647-labeled anti-mouse immunoglobulin (Ig)-specific antibody (Life Technologies, A21236) at 4 °C for 30 mins. Finally, the cells were washed and resuspended in 200 μl of cold PBS. The geometric Mean of Fluorescence Intensity (gMFI) of the T2 cells were then measured with a BD LSRFortessa flow cytometer (BD Biosciences). In this study, we reported the gMFI relative to no peptide control.

Ex-vivo evaluation of effector and polyfunctional CD8+ T-cell responses to the T-cell epitope derived peptides

Participants and clinical samples

Previously hospitalized SARS-CoV-2 convalescent patients were recruited from The Westmead Hospital in Westmead, NSW, Australia. For this study, we included two HLA-A*02 positive participants and one HLA-A*02 negative participant (Table 3). At 1-2 months post SARS-CoV-2 recovery, peripheral blood samples were collected from these participants in citrate anticoagulant tubes and cryopreserved peripheral blood mononuclear cells (PBMCs) were isolated within 1 hour of venipuncture. The PBMCs were isolated by Ficoll density gradient centrifugation. To determine the HLA for each donor, PBMCs were stained using a phycoerythrin-labeled anti-human HLA-A*02 antibody (Clone BB7.2, BD Bioscience) for 30 mins at 4 °C and the immunofluorescence was measured with a BD LSRFortessa flow cytometer.
(BD Biosciences). Where possible, the PBMCs derived from the HLA-A*02 negative participant were included as an experimental control.

**Expansion of peptide specific T-cells**

The PBMCs were thawed in RPMI and then rested overnight in RF10. The rested PBMCs (3.5X10^6 cells) were incubated in the presence of 5 µM of SARS-CoV-2 peptide pools or Epstein-Barr Virus (EBV) peptide mix (MACS GMP PepTivator EBV select, Miltenyi Biotec) for 1 hour. Two SARS-CoV-2 peptides derived from the NC (RTATKAYNV and IIWVATEGA; 5 µM for each peptide) were used to stimulate PBMCs for one hour. The EBV peptide mix (resuspended at 100 µg/mL in DMSO) was used for a positive control (MACS GMP PepTivator EBV Select, Miltenyi Biotec). After the incubation with these peptides, the cells were washed once with RPMI and co-stimulated using purified anti-human CD28 antibody (1 µg/ml) (Clone L293, BD). The stimulated cells were cultured in 48-well plates at a density of 2X10^6 cells/ml in RF10 medium supplemented with 100 U/ml Interleukin-2 (Lonza, 200-02) for 14 days. The medium was replaced every 72-hour with freshly prepared RF10 supplemented with IL-2. The expanded cells were subsequently studied by flow cytometry.

**Detection of effector and polyfunctional CD8+ T-cells responses to the T-cell epitope derived peptides**

The effector and polyfunctionality of CD8+ T-cells were evaluated by using the expanded cells that were exposed to the peptides for 14 days. Briefly, the expanded cells were re-stimulated with individual SARS-CoV-2 peptides or the EBV peptide pool in the presence of costimulatory antibodies (1 µg/ml of anti-CD28 and anti-CD49d; BD Biosciences), monensin (Golgistop, 0.9 µl/ml; BD Biosciences) and brefeldin A (1 µl/ml; BD Biosciences) for 5 hour at 37 °C. Anti
CD107a/b-FITC antibodies (BD Biosciences) were also added to identify degranulating cells. For the functionality panel, the cells were stained upon stimulation with LIVE/DEAD™ Fixable Near-IR Dead Cell Stain Kit (Thermo fisher) and the following conjugated antibodies: anti-CD3-BUV496, anti-CD8- PerCP-Cy5.5 (BD Biosciences). The cells were then fixed and permeabilized (BD Cytofix/Cytoperm™, BD Bioscience). Subsequently, the fixed cells were stained using anti-IL-2-PerCP-Cy5.5, anti-TNF-α-PE/Cy7, and anti-IFN-γ-PE antibodies (BD Biosciences). The data were analyzed by using FlowJo v10 (Data Analysis Software, LLC). The gating strategy was performed as follows: 1) the lymphocyte population was selected by using FSC-A vs Side scatter (SSC) plot; 2) the single cells were selected in a forward scatter area (FSC-A) vs FSC-Height plot; 3) the dead cells were excluded on the bases of LIVE/DEAD NIR fluorescence; and 4) the CD3+ CD8+ cells were gated in CD3 vs CD8 dot plots. To study the polyfunctionality of CD8+ T-cells, CD8 vs CD107a/b, IFN-γ, IL-2, or TNF-α plots were constructed. After the gates for each cytokine profile of CD8+ T-cells were created, the Boolean gate platform was employed to create all possible cytokine and CD107a/b combinations. For each combination, the resulting data was obtained by subtracting percent cells representing the background in the mock control and CD28/CD49d stimulation. The values below the background were set at 0. For the polyfunctionality analysis, SPICE 6.0 software (https://niaid.github.io/spice/) was used following the technical considerations published by the software developers (98).  

Acknowledgments  
We would like to thank all the authors who have kindly shared SARS-CoV-2 genome data on The Global Initiative on Sharing All Influenza Data (GISAID) (https://github.com/EunokLee/SARS-CoV-2_data_files_and_acknowledgements). The table
acknowledging all the authors who contributed to SARS-CoV-2 genomic sequences for our study can be found at https://github.com/EunokLee/SARS-CoV-2_data_files_and_acknowledgements. Data including the sequences in this study are available for download from https://www.gisaid.org. We acknowledge with gratitude the SARS-CoV-2-infected participants who provided clinical samples for the genomic sequence database and for the ex-vivo validation assays. This research was funded in part from the BEAT COVID-19 grant sponsored by the Snow Medical. The funder had no role in study design, data collection and interpretation, or the decision to submit the work for publication.
References:


broad scheme for prediction of HLA class II restricted T cell epitopes. J Immunol Methods 422:28-34.


combining predictions of proteasomal cleavage, TAP transport and MHC class I binding.


different supertypes and a surprising degree of repertoire sharing across supertypes.

Immunogenetics 63:325-35.


Figure legends

Figure 1. Global and regional HLA class I allele frequency. The data were curated from The Allele Frequency Net Database (www.allelefrequencies.net).

Figure 2. Immunoinformatics analysis pipeline. Immunoinformatics analysis pipeline used to identify highly networked SARS-CoV-2 T-cell epitope derived peptides for HLA class I and II restricted immune responses.

Figure 3. Genetically conserved 9-mer peptides with highly networked amino acid residues and restricted to HLA-A*02:01 allele. Network scores (blue bars), predicted percent bind level to the HLA class I molecule (pink bars) and percent genetic variability across SARS-CoV-2 variants (gray bars) within the peptide repertoires derived from SARS-CoV-2 NC N-terminal RNA binding domain (a), NC C-terminal dimerization domain (b) and S protein (c) are shown. For the S protein, the 9-mers with the network score of at least 100 and the bind level of at least 95% are shown among the repertoire of 1112 peptides. All the network scores presented in this figure are normalized. The thresholds for network scores (blue line) are determined by the lower 95% confidence of the mean. The cut-offs for the percent bind level (red line) equates to 95-percentile. The peptides with the network scores and percent bind level above the cut-offs and those which are conserved (i.e. 0% genetic variability) across the SARS-CoV-2 isolates are highlighted in red.

Figure 4. Binding prediction of highly networked T-cell epitope derived peptides to additional HLA class I alleles and global population coverage. The binding prediction of highly networked T-cell epitope derived peptides identified from the SARS-CoV-2 NC (a) and S protein (c). The binding prediction of 9-mer T-cell epitope derived peptides to HLA-A and HLA-B alleles classified into 10 supertypes (a and c). The peptides with top 5% bind levels to each of
these additional HLA class I alleles are indicated as squares. The supertypes are indicated by
different colors as shown in the figure. Percent global population coverage of highly networked
T-cell epitope derived peptides identified from the NC (b) and S protein (d) is also shown. Most
promising T-cell epitope derived peptides for HLA class I mediated immune recognition are
highlighted in red.

Figure 5. Sequence comparison between highly networked T-cell epitope derived peptides
identified by immunoinformatics analysis pipeline and other studies (9-mers). The highly
networked T-cell epitope derived peptides (9-mers) were compared to the epitopes identified by
Le Bert et al., 2020 (26) (gray), Grifoni et al., 2020 (43) (sky blue), Shomuradova et al., 2020
(44) (pink), Sekine et al., 2020 (16) (orange) and Peng et al., 2020 (29) (green). The highly
networked T-cell epitopes which are most promising (i.e. those with top 5% scores for HLA
class I mediated antigen processing and immunogenicity parameters), unique or both are
indicated by blue rectangles in the figure. Also, the most promising peptides for HLA class I
restricted immune response with less than 20% homology to four seasonal coronaviruses are
indicated by blue rectangles.

Figure 6. Comparison between highly networked T-cell epitope derived peptides identified
from SARS-CoV-2 S protein and mutation sites that enhance viral infectivity. The highly
networked T-cell epitope derived peptides are mapped to consensus S protein sequence with 14
mutation sites that enhance viral infectivity. The mutation sites have been identified by Korber et
al., 2020 (30).

Figure 7. Correlation between network scores and predictions for HLA class I mediated
antigen processing and presentation. The correlations between the normalized network scores
and the predictions for proteasomal processing (a, d, g), TAP binding (b, e, h) and HLA class I
mediated immunogenicity (c, f, i) are shown. The data includes 123 peptides derived from the NC N-terminal domain (a-c), 110 peptides derived from the NC C-terminal domain (d-f) and 1112 peptides from the S protein (g-i). The R2 value indicates the degree of the peptide repertoire that follow linear regression and the coefficient indicates the slope of the regression. 95% confidence intervals (in square brackets) and p-value for the coefficient are also presented.

**Figure 8. Global and regional HLA class II allele frequency.** The data were curated from The Allele Frequency Net Database (www.allelefrequencies.net).

**Figure 9. Genetically conserved 12-mer peptides with highly networked amino acid residues and restricted to DRB1*07:01 allele.** Network scores (blue bars), predicted percent bind level to the HLA class II molecule (pink bars) and percent genetic variability across SARS-CoV-2 variants (gray bars) within the peptide repertoires derived from SARS-CoV-2 NC N-terminal RNA binding domain (a), NC C-terminal dimerization domain (b) and S protein (c) are shown. For the S protein, the 12-mers with the network score of at least 100 and the bind level of at least 95% are shown among the repertoire of 1112 peptides. All the network scores presented in this figure are normalized. The thresholds for network scores (blue line) are determined by the lower 95% confidence of the mean. The cut-offs for the percent bind level (red line) equates to 95-percentile. The peptides with the network scores and percent bind level above the cut-offs and those which are conserved (i.e. 0% genetic variability) across the SARS-CoV-2 isolates are highlighted in red.

**Figure 10. Binding prediction of highly networked T-cell epitope derived peptides to additional HLA class II alleles and global population coverage.** The binding prediction of highly networked T-cell epitope derived peptides identified from the SARS-CoV-2 NC (a) and S protein (c). The binding prediction of 12-mer T-cell epitope derived peptides to HLA class II
alleles classified into four loci (a and c). The peptides with top 5% bind levels to each of these additional HLA class II alleles are indicated as squares. The loci are indicated by different colors as shown in the figure. Percent global population coverage of highly networked T-cell epitope derived peptides identified from the NC (b) and S protein (d) is also shown. Most promising T-cell epitope derived peptides for HLA class II mediated immune recognition are highlighted in red.

Figure 11. Sequence comparison between highly networked T-cell epitope derived peptides identified by immunoinformatics analysis pipeline and other studies (12-mers). The highly networked T-cell epitope derived peptides (12-mers) were compared to the epitopes identified by Le Bert et al., 2020 (26) (gray), Grifoni et al., 2020 (43) (sky blue), Shomuradova et al., 2020 (44) (pink), Sekine et al., 2020 (16) (orange) and Peng et al., 2020 (29) (green). The highly networked T-cell epitopes which are most promising (i.e. those with network scores and percent bind level above the thresholds), unique or both are indicated by blue rectangles in the figure. Also, the most promising peptides for HLA class II restricted immune response with less than 20% homology to four seasonal coronaviruses are indicated by blue rectangles.

Figure 12. Sequence homology of highly networked T-cell epitope derived peptides identified from SARS-CoV-2 NC and S proteins to seasonal human coronaviruses. The homology between the T-cell epitope derived peptides (a-d, 9-mers; e-h, 12-mers) and four seasonal human coronaviruses (229E, HKU1, NL63 and OC43) is presented by a mean sequence identity (%) and its 95% confidence intervals. Most promising T-cell epitope derived peptides for HLA class I and II mediated immune recognition are highlighted in red.

Figure 13. In-vitro validation of SARS-CoV-2 NC peptides binding to HLA-A*02:01 by using T2 cells. T-cell epitope derived peptides identified from the N- and C-terminal domains of...
the SARS-CoV-2 NC protein were used to assess the binding capacity (RTATKAYNV and IIWVATEGA). HLA-A*02:01 expression on the cell surface was measured as gMFI by flow cytometry. NLVPMAVATV (HLA-A*02 binding) and TPRVTGGGAM (HLA-A*02 non-binding) peptides from CMV pp65 were used as positive and negative controls. (a) Histograms from a representative experiment are shown. (b) HLA-A*02:01 expression was expressed relative to the no peptide control. (n=2).

Figure 14. CD8+ T-cell polyfunctionality analysis after ex-vivo expansion with specific SARS CoV-2 peptides derived from the SARS-CoV-2 NC. Representative dot-plots showing effector cytokine production and degranulation level of CD8+ T cells stimulated with RTATKAYNV (a) and IIWVATEGA (c) peptides. The CD8+ T-cells were derived from two HLA-A*02 positive SARS-CoV-2 convalescent donors (a and c). PBMCs from an HLA-A*02 negative SARS-CoV-2 convalescent donor were used as negative control. The clinical samples were obtained at 1-2 months post SARS-CoV-2 recovery. Bar charts represent the effector profile of RTATKAYNV (b) and IIWVATEGA (d)-specific CD8+ T-cells. PBMCs expanded with EBV peptides were used as positive control. Pies depict the distribution of mono-, bi-, tri-, and tetra-functional cells within specific CD8+ T-cells.
Table 1. Highly networked 9-mer peptides with binding potential to HLA class I molecule (HLA-A*02:01).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Peptide</th>
<th>Network (normalized)</th>
<th>HLA-A binding</th>
<th>Parameters for antigen processing and presentation related to HLA class I</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Score</td>
<td>Cut off 95% CI</td>
<td>% Bind Level</td>
</tr>
<tr>
<td>NC</td>
<td>NC</td>
<td>48.9</td>
<td>26.8</td>
<td>95.7</td>
</tr>
<tr>
<td>S-terminal</td>
<td>S-terminal</td>
<td>51.6</td>
<td>26.8</td>
<td>99.8</td>
</tr>
<tr>
<td>SUVATGLA</td>
<td>47.5</td>
<td>26.8</td>
<td>98.6</td>
<td>89.5</td>
</tr>
<tr>
<td>C-terminal</td>
<td>C-terminal</td>
<td>51.6</td>
<td>26.8</td>
<td>95.7</td>
</tr>
<tr>
<td>EVATKAVN</td>
<td>36.6</td>
<td>23.9</td>
<td>93.3</td>
<td>93.6</td>
</tr>
<tr>
<td>RLLKHDIN</td>
<td>37.5</td>
<td>23.9</td>
<td>98.5</td>
<td>93.6</td>
</tr>
<tr>
<td>ECPFNYNTV</td>
<td>20.6</td>
<td>10.0</td>
<td>97.9</td>
<td>95.5</td>
</tr>
<tr>
<td>EVWVRAHNY</td>
<td>17.5</td>
<td>10.0</td>
<td>93.8</td>
<td>95.5</td>
</tr>
<tr>
<td>TLSDKQSM</td>
<td>17.3</td>
<td>10.0</td>
<td>98.5</td>
<td>95.5</td>
</tr>
<tr>
<td>TSOQSDPL</td>
<td>17.9</td>
<td>10.0</td>
<td>98.5</td>
<td>95.5</td>
</tr>
<tr>
<td>TSDPRAKLA</td>
<td>16.5</td>
<td>10.0</td>
<td>98.5</td>
<td>95.5</td>
</tr>
<tr>
<td>FADSLQHT</td>
<td>10.6</td>
<td>10.0</td>
<td>98.5</td>
<td>95.5</td>
</tr>
<tr>
<td>LQWQTRTLA</td>
<td>21.3</td>
<td>10.0</td>
<td>95.5</td>
<td>95.5</td>
</tr>
<tr>
<td>AVDCQDLIP</td>
<td>21.6</td>
<td>10.0</td>
<td>98.5</td>
<td>95.5</td>
</tr>
<tr>
<td>ESTFRCAY</td>
<td>19.9</td>
<td>10.0</td>
<td>98.5</td>
<td>95.5</td>
</tr>
<tr>
<td>NVSAAGDYE</td>
<td>17.5</td>
<td>10.0</td>
<td>97.9</td>
<td>95.5</td>
</tr>
<tr>
<td>RSVSPTLQ</td>
<td>17.0</td>
<td>10.0</td>
<td>97.9</td>
<td>95.5</td>
</tr>
<tr>
<td>MAVTDMNE</td>
<td>22.0</td>
<td>10.0</td>
<td>99.9</td>
<td>95.5</td>
</tr>
<tr>
<td>VTVIPELY</td>
<td>21.0</td>
<td>10.0</td>
<td>98.5</td>
<td>95.5</td>
</tr>
<tr>
<td>Peptide</td>
<td>HLA-A*02:01</td>
<td>IC50</td>
<td>Min IC50</td>
<td>Max IC50</td>
</tr>
<tr>
<td>-----------</td>
<td>-------------</td>
<td>------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>LLQYGSFCT</td>
<td>182.5</td>
<td>160.0</td>
<td>95.8</td>
<td>108.2</td>
</tr>
<tr>
<td>QNRLALTGI</td>
<td>201.0</td>
<td>160.0</td>
<td>97.2</td>
<td>160.0</td>
</tr>
<tr>
<td>KQYKTPPS</td>
<td>176.1</td>
<td>160.0</td>
<td>99.5</td>
<td>127.7</td>
</tr>
<tr>
<td>LLPSKVLCA</td>
<td>203.5</td>
<td>160.0</td>
<td>99.8</td>
<td>18.0</td>
</tr>
<tr>
<td>GGYVFLPL</td>
<td>190.1</td>
<td>160.0</td>
<td>98.9</td>
<td>180.9</td>
</tr>
<tr>
<td>MQYTCSQ</td>
<td>207.2</td>
<td>160.0</td>
<td>99.5</td>
<td>12.1</td>
</tr>
<tr>
<td>ALLACITYS</td>
<td>205.8</td>
<td>160.0</td>
<td>96.2</td>
<td>86.0</td>
</tr>
<tr>
<td>YHGWYFGCA</td>
<td>196.4</td>
<td>160.0</td>
<td>97.7</td>
<td>115.8</td>
</tr>
<tr>
<td>WDFGAGAAL</td>
<td>196.1</td>
<td>160.0</td>
<td>96.6</td>
<td>125.9</td>
</tr>
<tr>
<td>LGPPAMQM</td>
<td>191.1</td>
<td>160.0</td>
<td>96.5</td>
<td>79.5</td>
</tr>
<tr>
<td>QSHAVFENG</td>
<td>208.5</td>
<td>160.0</td>
<td>96.2</td>
<td>180.3</td>
</tr>
<tr>
<td>VLYENQKLI</td>
<td>196.4</td>
<td>160.0</td>
<td>97.6</td>
<td>106.1</td>
</tr>
<tr>
<td>ALTEVQKQ</td>
<td>205.9</td>
<td>160.0</td>
<td>96.5</td>
<td>81.6</td>
</tr>
<tr>
<td>EQSNNSFGC</td>
<td>191.2</td>
<td>160.0</td>
<td>97.5</td>
<td>175.6</td>
</tr>
<tr>
<td>YHEDTSE</td>
<td>179.0</td>
<td>160.0</td>
<td>96.8</td>
<td>17.0</td>
</tr>
<tr>
<td>REDKVEAEY</td>
<td>177.0</td>
<td>160.0</td>
<td>99.4</td>
<td>45.2</td>
</tr>
</tbody>
</table>

Highlighted in red, most promising T-cell epitope-derived peptides for HLA class I mediated immune recognition (i.e. those with top 5% scores for HLA class I antigen processing and presentation parameters). *This peptide has a high network score and percent bind level above the 95-percentile however, the IC50 is above the cut-off IC50 for HLA-I binding; †p:HLA, peptide and HLA-A*02:01 complex; ††p:HLA t1/2, predicted time required for the dissociation of peptide:HLA complex.
Table 2. Highly networked 12-mer peptides with binding potential to HLA class II molecule (DRB1*07:01).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Allele</th>
<th>Peptide</th>
<th>Network (normalized)</th>
<th>HLA-II binding Score</th>
<th>% Bind Level</th>
<th>Cut-off 95% CI</th>
<th>IC50 (nM)</th>
<th>Cut-off IC50 (nM)</th>
<th>B cell epitopes*</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>QGYVRKARIHG</td>
<td>41.5</td>
<td>24.6</td>
<td>93.0</td>
<td>97.5</td>
<td>284.9</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>GYVRKARIHG</td>
<td>37.9</td>
<td>24.6</td>
<td>94.5</td>
<td>87.5</td>
<td>284.9</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>GYYRRATRRIRG</td>
<td>32.7</td>
<td>24.6</td>
<td>91.0</td>
<td>87.5</td>
<td>321.0</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>IIWVATEGALNT</td>
<td>28.4</td>
<td>24.6</td>
<td>95.0</td>
<td>87.5</td>
<td>188.1</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>LFLPFFSNVTWFH</td>
<td>203.5</td>
<td>157.3</td>
<td>97.0</td>
<td>95.0</td>
<td>125.6</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>FLPFFSNVTWFHA</td>
<td>200.7</td>
<td>157.3</td>
<td>98.0</td>
<td>95.0</td>
<td>96.9</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>LPFFSNVTWFHA</td>
<td>193.7</td>
<td>157.3</td>
<td>98.4</td>
<td>95.0</td>
<td>85.1</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>PFFSNVTWFHA</td>
<td>190.0</td>
<td>157.3</td>
<td>98.7</td>
<td>95.0</td>
<td>74.9</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>FSNVTWFHA</td>
<td>185.2</td>
<td>157.3</td>
<td>96.5</td>
<td>98.0</td>
<td>132.4</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>SNVTWFHA</td>
<td>178.1</td>
<td>157.3</td>
<td>97.5</td>
<td>98.0</td>
<td>104.6</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>NTWFHA</td>
<td>172.3</td>
<td>157.3</td>
<td>97.5</td>
<td>98.0</td>
<td>103.5</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>SNTWFHA</td>
<td>166.4</td>
<td>157.3</td>
<td>97.5</td>
<td>98.0</td>
<td>117.5</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>CNTWFHA</td>
<td>161.6</td>
<td>157.3</td>
<td>97.5</td>
<td>98.0</td>
<td>115.5</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>CTWFHA</td>
<td>156.8</td>
<td>157.3</td>
<td>97.5</td>
<td>98.0</td>
<td>124.6</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>CTWFHA</td>
<td>154.3</td>
<td>157.3</td>
<td>97.5</td>
<td>98.0</td>
<td>136.8</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>EKGVGQA</td>
<td>211.2</td>
<td>157.3</td>
<td>98.0</td>
<td>98.0</td>
<td>150.4</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>KGITFGDSTIVQ</td>
<td>210.3</td>
<td>157.3</td>
<td>97.5</td>
<td>98.0</td>
<td>138.1</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>ENFNGTGDSTIVQ</td>
<td>192.2</td>
<td>157.3</td>
<td>96.0</td>
<td>98.0</td>
<td>131.5</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>CVTVTAEGGNNV</td>
<td>194.9</td>
<td>157.3</td>
<td>96.5</td>
<td>99.0</td>
<td>136.1</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
<tr>
<td>N-termin</td>
<td>DRB1*07:01</td>
<td>WVTVKNGAFFQ</td>
<td>179.5</td>
<td>157.3</td>
<td>96.5</td>
<td>98.0</td>
<td>132.5</td>
<td>5000</td>
<td>HGKEDLKFPRGQGVPINTNSSPDD</td>
</tr>
</tbody>
</table>
Highlighted in red, most promising T-cell epitope-derived peptides for HLA class II mediated immune recognition (i.e. those with network scores and percent bind level above the thresholds). *B cell epitopes predicted by BepiPred Linear Epitope prediction 2.0. The consecutive amino acid residues within the B cell epitopes that are identical to the T-cell epitope derived peptides (highlighted in red).

<table>
<thead>
<tr>
<th>Peptide</th>
<th>DRB1*07:01</th>
<th>Network Score</th>
<th>Network BIND</th>
<th>Domain BIND</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>NVFQTRAGCLI</td>
<td>206.1</td>
<td>97.5</td>
<td>95.0</td>
<td>112.4</td>
<td>5000</td>
</tr>
<tr>
<td>NVFQTRAGCLIG</td>
<td>206.7</td>
<td>96.0</td>
<td>95.0</td>
<td>150.2</td>
<td>5000</td>
</tr>
<tr>
<td>SIIAYTMSLGA</td>
<td>212.5</td>
<td>97.5</td>
<td>95.0</td>
<td>107.6</td>
<td>5000</td>
</tr>
<tr>
<td>SIIAYTMSLGAEN</td>
<td>211.4</td>
<td>97.5</td>
<td>95.0</td>
<td>100.6</td>
<td>5000</td>
</tr>
<tr>
<td>SIIAYTMSLGAENS</td>
<td>210.4</td>
<td>97.5</td>
<td>95.0</td>
<td>103.8</td>
<td>5000</td>
</tr>
<tr>
<td>IIAYTMSLGAEN</td>
<td>209.4</td>
<td>97.5</td>
<td>95.0</td>
<td>103.8</td>
<td>5000</td>
</tr>
<tr>
<td>IAYTMSLGAENS</td>
<td>204.7</td>
<td>97.5</td>
<td>95.0</td>
<td>100.6</td>
<td>5000</td>
</tr>
<tr>
<td>EMIAQYTSALLA</td>
<td>198.8</td>
<td>98.2</td>
<td>95.0</td>
<td>92.4</td>
<td>5000</td>
</tr>
<tr>
<td>QIPFAMQMAYRF</td>
<td>201.1</td>
<td>98.1</td>
<td>95.0</td>
<td>92.4</td>
<td>5000</td>
</tr>
<tr>
<td>PFAMQMAYRFN</td>
<td>200.5</td>
<td>98.1</td>
<td>95.0</td>
<td>92.4</td>
<td>5000</td>
</tr>
<tr>
<td>FAMQMAYRFNGI</td>
<td>202.5</td>
<td>98.1</td>
<td>95.0</td>
<td>92.4</td>
<td>5000</td>
</tr>
</tbody>
</table>
Table 3. Participant demographics.

<table>
<thead>
<tr>
<th>Participant ID</th>
<th>HLA class I supertype</th>
<th>HLA class I alleles</th>
<th>Gender</th>
<th>Age (yrs)</th>
<th>Date diagnosed</th>
<th>COVID-19 severity</th>
<th>Time post SARS-CoV-2 recovery (months)</th>
<th>Highly networked T-cell epitope derived peptides tested in ex-vivo PBMCs</th>
<th>SARS-CoV-2 protein source for the highly networked T-cell epitope derived peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>CVBL06A</td>
<td>HLA-A*02 Positive</td>
<td>HLA-A*02:06</td>
<td>Female</td>
<td>60</td>
<td>26-Mar-2020</td>
<td>N/A</td>
<td>1</td>
<td>RTATKAYNV</td>
<td>NC C-terminal domain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLA-A*24:02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVBL10B</td>
<td>HLA-A*02 Positive</td>
<td>HLA-A*02:01</td>
<td>Male</td>
<td>61</td>
<td>28-Mar-2020</td>
<td>N/A</td>
<td>2</td>
<td>IIWVATEGA</td>
<td>NC N-terminal domain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLA-A*03:01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVBL05A</td>
<td>HLA-A*02 Negative</td>
<td>HLA-A*24:02</td>
<td>Female</td>
<td>66</td>
<td>5-Apr-2020</td>
<td>Hospitalized</td>
<td>1</td>
<td>RTATKAYNV</td>
<td>NC C-terminal domain</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HLA-A*24:07</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IIWVATEGA</td>
<td>NC N-terminal domain</td>
</tr>
<tr>
<td>Region</td>
<td>Epitope</td>
<td>Length</td>
<td>Promising</td>
<td>Unique</td>
<td>Unique and promising</td>
<td>SMM</td>
<td>HPLA</td>
<td>NL4-3</td>
<td>OC43</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>--------</td>
<td>-----------</td>
<td>--------</td>
<td>-----------------------</td>
<td>-----</td>
<td>------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>NC N-terminal</td>
<td>NTAHRTTAL</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC N-terminal</td>
<td>TSWRTTALT</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC N-terminal</td>
<td>SQWVYHSWA</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC N-terminal</td>
<td>TTVNSGVEA</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC N-terminal</td>
<td>RYANASTWV</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NC N-terminal</td>
<td>FEAMODEM</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>FFPRVYYCT</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>VYVTVRSEV</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>TEKMTSRL</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>FKQGNDVY</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>FCMPPLLY</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>PSYPPMPG</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>VYQQYRTEI</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>ANYYAARD</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>PSTKCCYV</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>NTVLYKSL</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>RQVYMSNL</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>SIATYRMN</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>SVYTERPV</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>LQQGGFCT</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>QPMALITM</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>RQFQKTHF</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>LINFYTMN</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>CEVYFPHL</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>MNMYTIAL</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>AGSLQGTS</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>TTSWVPHUA</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>WTSGAMNL</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>LQQVAMPM</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>QMVYTPFG</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>VSQKVSYL</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>AINTVYQK</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>SQCINQVA</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>VNDIBSL</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5 peptides</td>
<td>RLGKNSAVY</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Most promising T-cell epitopes for HLA class I and II mediated immune recognition (highlighted in red)