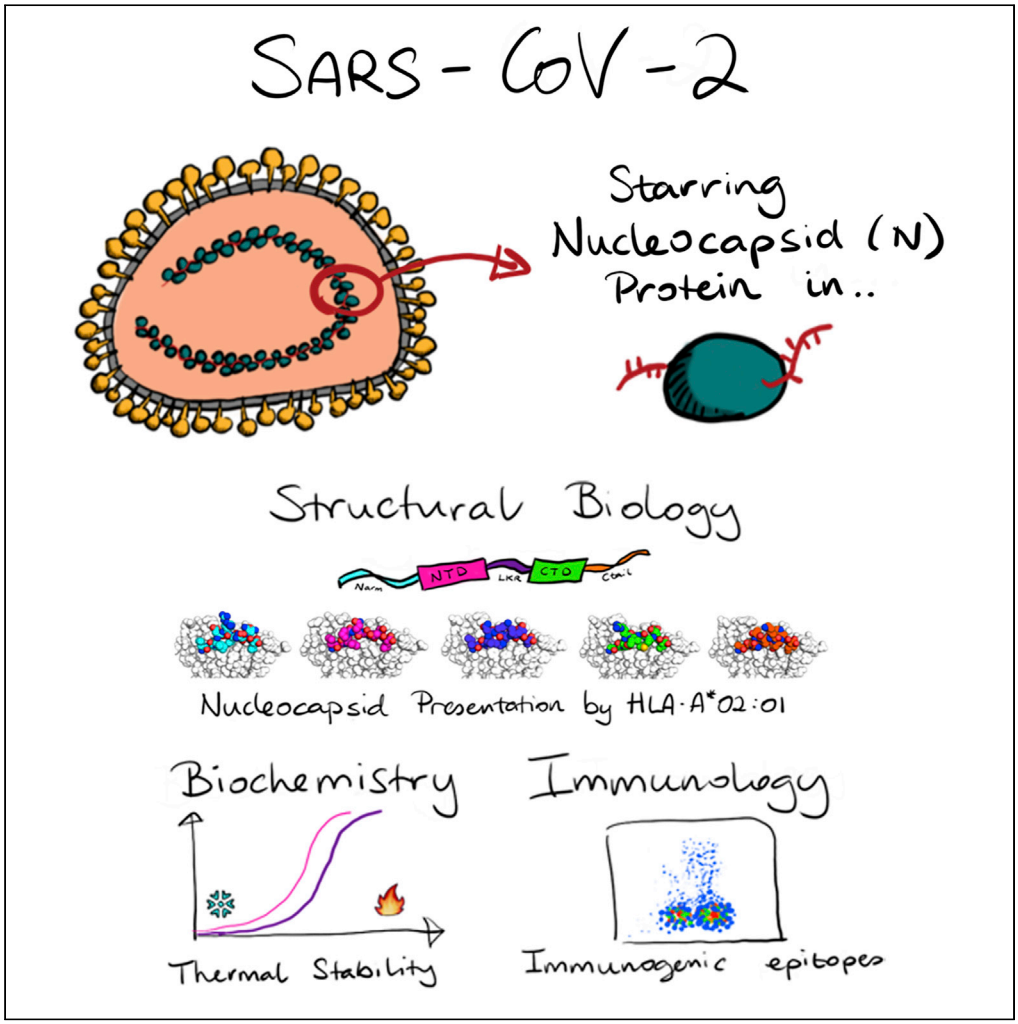
iScience

CellPress
OPEN ACCESS

Article

The presentation of SARS-CoV-2 peptides by the common HLA-A*02:01 molecule



Christopher Szeto, Demetra S.M. Chatzileontiadou, Andrea T. Nguyen, ..., Alan Riboldi-Tunnicliffe, Emma J. Grant, Stephanie Gras

e.grant@latrobe.edu.au (E.J.G.)
s.gras@latrobe.edu.au (S.G.)

HIGHLIGHTS

HLA-A*02:01 individuals have limited pre-existing immunity to SARS-CoV-2 nucleocapsid

High-resolution crystal structures of HLA-A*02:01 presenting SARS-CoV-2 peptides

Structural analysis of pHLA shows stability influences peptide immunogenicity

Szeto et al., iScience 24, 102096
February 19, 2021
https://doi.org/10.1016/j.isci.2021.102096

iScience

CellPress
OPEN ACCESS

Article

The presentation of SARS-CoV-2 peptides by the common HLA-A*02:01 molecule

Christopher Szeto,[1,2,6] Demetra S.M. Chatzileontiadou,[1,2,6] Andrea T. Nguyen,[1,2] Hannah Sloane,[1,2] Christian A. Lobos,[1,2] Dhilshan Jayasinghe,[1,2] Hanim Halim,[1] Corey Smith,[3] Alan Riboldi-Tunnicliffe,[4] Emma J. Grant,[1,2,7,*] and Stephanie Gras[1,2,5,7,8,*]

SUMMARY

CD8+ T cells are crucial for anti-viral immunity; however, understanding T cell responses requires the identification of epitopes presented by human leukocyte antigens (HLA). To date, few SARS-CoV-2-specific CD8+ T cell epitopes have been described. Internal viral proteins are typically more conserved than surface proteins and are often the target of CD8+ T cells. Therefore, we have characterized eight peptides derived from the internal SARS-CoV-2 nucleocapsid protein predicted to bind HLA-A*02:01, the most common HLA molecule in the global population. We determined not all peptides could form a complex with HLA-A*02:01, and the six crystal structures determined revealed that some peptides adopted a mobile conformation. We therefore provide a molecular understanding of SARS-CoV-2 CD8+ T cell epitopes. Furthermore, we show that there is limited pre-existing CD8+ T cell response toward these epitopes in unexposed individuals. Together, these data show that SARS-CoV-2 nucleocapsid might not contain potent epitopes restricted to HLA-A*02:01.

INTRODUCTION

Following the emergence of SARS-CoV-2 out of Wuhan, China in late 2019, there have been over 92 million infections and over 1.9 million deaths worldwide (Dong et al., 2020). This novel and highly infectious coronavirus has significantly encumbered the majority of the world, resulting in massive economic "loss" in all global economies. Worldwide, significant efforts are being made toward the development of a SARS-CoV-2 vaccine (Amanat and Krammer, 2020).

Studies have shown that SARS-CoV-2-specific antibodies might not be long-lived (Vabret, 2020; Roltgen et al., 2020), and it is unclear if they could provide long-term protective immunity. It is well established that cytotoxic CD8+ T cells play a vital role in the control and clearance of viral pathogens. This is particularly well characterized in other viral respiratory infections, such as influenza (McMichael et al., 1983), where pre-existing CD8+ T cells can decrease disease severity and overall symptom scores (Wells et al., 1981; Bender et al., 1992). In addition, CD8+ T cell immunity can be long-lasting, with longitudinal studies showing that CD8+ T cell immunity can be detected 10–50 years following vaccination (Miller et al., 2008; Ahmed and Akondy, 2011) and at least 13 years following natural influenza infection (van de Sandt et al., 2015). Furthermore, CD8+ T cells can be cross-reactive, recognizing variants within viral epitopes (Grant et al., 2018; Valkenburg et al., 2010), meaning that they can occasionally recognize distinct viral strains (Kreijtz et al., 2008; van de Sandt et al., 2014). As such, there is a great interest in understanding the CD8+ T cell response toward SARS-CoV-2 to determine if a CD8+ T-cell-mediated vaccine can provide broad and long-lasting protection. Indeed, there is emerging evidence that CD8+ T cells can be detected following infection with SARS-CoV-2 (Rydyznski Moderbacher et al., 2020; Weiskopf et al., 2020; Peng et al., 2020), but their role in protective immunity is yet to be fully understood.

A significant challenge to investigating epitope-specific CD8+ T cell responses toward novel viruses is the lack of defined epitopes. CD8+ T cells recognize small peptides, derived from self or pathogenic proteins that are bound by human leukocyte antigen (HLA) molecules. HLA molecules are themselves highly polymorphic, where

[1]Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia

[2]Department of Biochemistry and Genetics, La Trobe Institute for Molecular Science, La Trobe University, Bundoora, VIC 3086, Australia

[3]QIMR Centre for Immunotherapy and Vaccine Development and Department of Immunology, QIMR Berghofer Medical Research Institute, Brisbane, QLD 4006, Australia

[4]Australian Synchrotron, ANSTO, Clayton, VIC 3168, Australia

[5]Australian Research Council Centre of Excellence for Advanced Molecular Imaging, Monash University, Clayton, VIC 3800, Australia

[6]These authors contributed equally

[7]These authors contributed equally

[8]Lead contact

*Correspondence: e.grant@latrobe.edu.au (E.J.G.), s.gras@latrobe.edu.au (S.G.)

https://doi.org/10.1016/j.isci.2021.102096

the majority of their diversity lies within their antigen binding cleft, ensuring the presentation of a wide range of peptides from diverse pathogens. HLA molecules are divided into HLA class I and II, which are recognized by CD8+ and CD4+ T cells, respectively. HLA class I molecules bind small peptides (8–10 residues) in their antigen binding cleft through a series of pockets termed A–F. The second and last residues of the peptide (P2 and PΩ, respectively) bind within the B and F pockets, respectively. These residues are often referred to as anchor residues, due to their critical role in anchoring the peptide into the HLA cleft. Each HLA molecule binds peptides with characteristic motifs that comprise particular anchor residues (P2 and PΩ) that are well adapted to the chemical properties of the HLA B and F pockets, respectively. For example, in HLA-A*02:01-restricted peptides, a Leucine or Methionine is often preferred at P2, a smaller hydrophobic residue such as Valine or Leucine being characteristic of the PΩ residue (Sette and Sidney, 1999). To date, only a few SARS-CoV-2 epitopes have been identified, and this limited knowledge represents a major roadblock to studying T cell immunity toward this new virus. There are numerous technical challenges to overcome, including limited access to patient samples, limited sample volume (e.g. blood), the need for HLA typing, and confirming the HLA restriction for each peptide. An approach to identify potential epitopes to a novel virus is to use known epitopes from a closely related virus or the use of peptide prediction algorithms.

In this study, we selected eight SARS-CoV-2 peptides that were likely to bind to the HLA-A*02:01 molecule, an allele common to ~40% of the global population. These peptides are a mix of predicted SARS-CoV-2 peptides and previously described SARS peptides (Cheung et al., 2007; Grifoni et al., 2020). We have focused on the nucleocapsid (N) protein, which is typically conserved between closely related viruses, due to its important function. This makes the N protein an ideal target for CD8+ T cells and therefore, a CD8+ T cell-mediated vaccine. We firstly determined the ability of each peptide to form a stable complex with the HLA-A*02:01 molecule, showing that one was unable to form a complex with HLA-A*02:01, and three form a poorly stable complex. We then determined the crystal structure of six HLA-A*02:01-SARS-CoV-2 complexes, providing the first description of CD8+ T cell SARS-CoV-2 epitopes at an atomic level. Interestingly, three of our selected peptides have since been shown to be immunogenic ($N_{219-227}$ (Isabel Schulien et al., 2021), $N_{222-230}$ (Ferretti et al., 2020; Isabel Schulien et al., 2021), and $N_{316-324}$ (Habel et al., 2020; Isabel Schulien et al., 2021)) in COVID-19 recovered individuals. As expected, these immunogenic peptides were able to form a stable complex with the HLA-A*02:01 molecule. Furthermore, we compared the sequences of our selected peptides with other circulating common cold coronaviruses and found that some of our peptides were highly conserved. As such, we assessed the functional CD8+ T cell response toward these peptides in healthy HLA-A*02:01 individuals with no known exposure to SARS-CoV-2. We found that, despite some conservation, there was limited pre-existing CD8+ T cell immunity toward these peptides.

Overall, we demonstrate that not all selected peptides were able to form stable complexes with HLA-A*02:01, which was a consequence of unfavourable P2 and/or PΩ residues and an important factor for immunogenicity. In addition, we saw limited pre-existing CD8+ T cell response for these peptides in unexposed donors, whereas peptides that have subsequently been shown to be immunogenic in COVID-19 recovered patients were stable and adopted a canonical conformation in the cleft of HLA-A*02:01. Altogether, our data provide molecular insight into CD8+ T cell epitopes from SARS-CoV-2.

RESULTS

SARS-CoV-2 and SARS nucleocapsid proteins are highly identical

Viral nucleocapsid (N) proteins are typically highly conserved due to their important functions, making them ideal targets for vaccine design. The N protein of SARS-CoV-2 is particularly important for RNA packaging during the release of the virion. Because internal proteins are more conserved than surface proteins, the N protein is an excellent target for the adaptive immune system and particularly cytotoxic CD8+ T cells.

The SARS-CoV-2 N protein is composed of 419 residues and divided into two main domains, the N-terminal (NTD) and the C-terminal domains (CTD) that are connected via a Ser-Arg rich linker (LKR) and flanked by a N-arm and C-tail loops, similar to the N protein of SARS (Chang et al., 2014) (Figure 1). Alignment of the N proteins from SARS-CoV-2 with SARS and four other coronavirus strains responsible for the common cold (NL63, OC43, HKU1, and 229E (Gagneur et al., 2002)) revealed that SARS-CoV-2 had a higher sequence identity with SARS (90%) than with the remaining four coronaviruses (23%–29% sequence identity). The NTD and LKR domains exhibited the highest sequence homology between the six viruses (Figure 1). We selected a total of eight N-derived peptides; six from the conserved NTD and LKR domains and two from the CTD section (Table 1). Out of the eight peptides predicted to bind the HLA-A*02:01 (Campbell et al., 2020) that are conserved in SARS

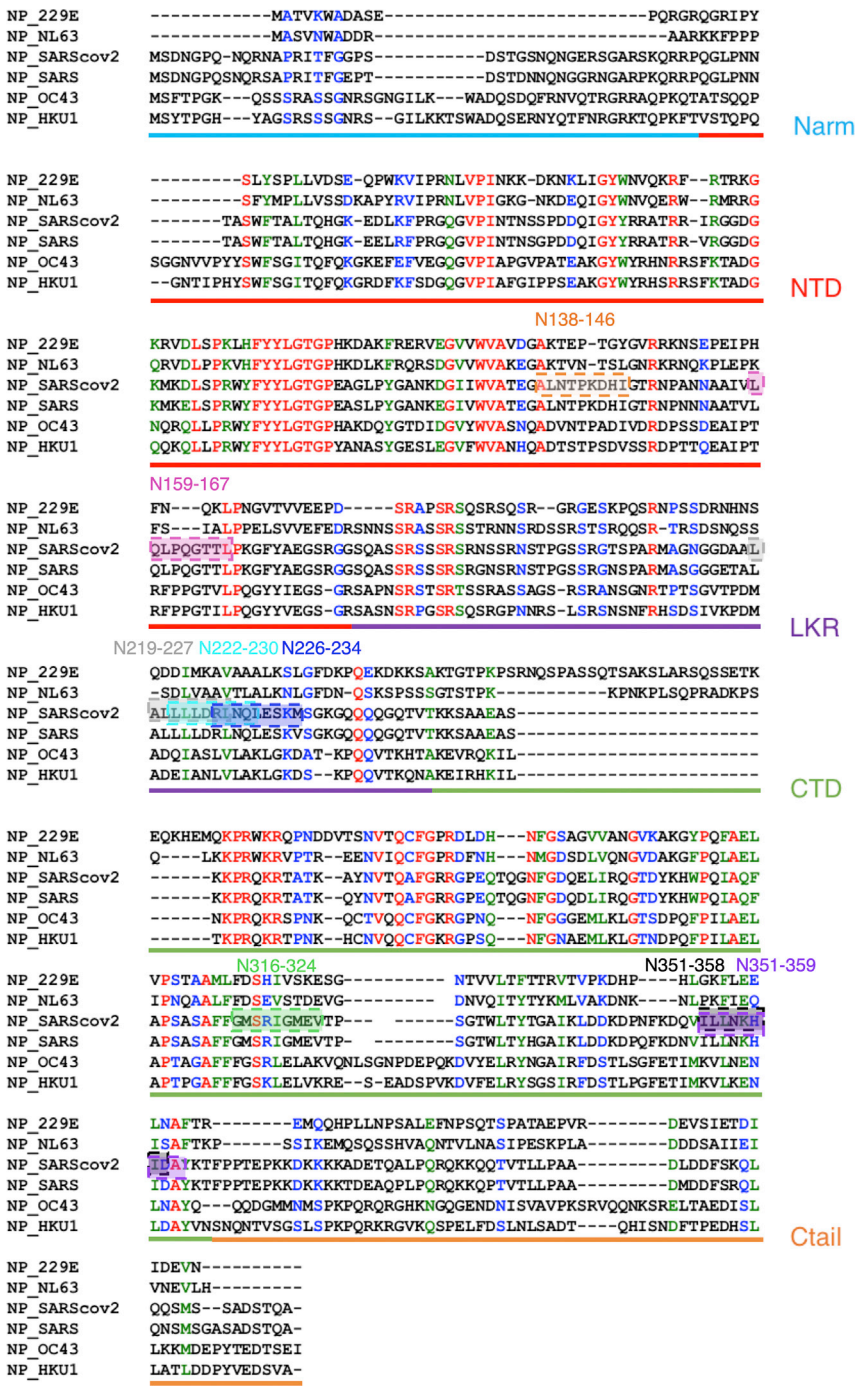


Figure 1. Protein sequence alignment of nucleocapsid derived from Coronavirus strains

Sequence alignment of nucleocapsid amino acid sequence from SARS-CoV-2, SARS, OC43, 229E, HKU1, NL63. SARS-CoV-2 N protein was aligned with SARS (90% identity), OC43 (29% identity), 229E (23% identity), NL63 (24% identity), and HKU1 (29% identity). The residues are colored by their level of similarity in black (none), blue (weak), green (strong), or red (identical). Each region of the N protein was identified by the colored bar underneath the sequences with N-arm in cyan, N Terminal Domain (NTD) in red, Linker (LKR) in purple, C Terminal Domain (CTD) in green, and C-tail in orange. The N peptides from Table 1 are identified by colored boxes in orange ($N_{138-146}$), pink ($N_{159-167}$), gray ($N_{219-227}$), cyan ($N_{222-230}$), blue ($N_{226-234}$), green ($N_{316-324}$), black ($N_{351-358}$), and purple ($N_{351-359}$).

Table 1. SARS-CoV-2 potential HLA-A*02:01-restricted CD8+ T cell epitopes

Name	Sequence	Tm (°C)	Immunogenicity
$N_{138-146}$	ALNTPKDHI	35.7 ± 0.6	Null (0/11, (Isabel Schulien et al., 2021))
$N_{159-167}$	LQLPQGTTL	35.8 ± 1.5	Null (0/11, (Isabel Schulien et al., 2021))
$N_{219-227}$	LALLLLDRL	41.5 ± 0.5	Weak (2/11, (Isabel Schulien et al., 2021))
$N_{222-230}$	LLLDRLNQL	54.7 ± 0.4	Weak (3/11, (Isabel Schulien et al., 2021))
$N_{226-234}$	RLNQLESKM	39.2 ± 1.0	ND
$N_{316-324}$	GMSRIGMEV	49.0 ± 0.1	Weak (2/11, (Isabel Schulien et al., 2021)), (Habel et al., 2020)
$N_{351-358}$	ILLNKHID	ND	ND
$N_{351-359}$	ILLNKHIDA	42.9 ± 0.3	ND

ND: not determined, Tm: thermal midpoint temperature.

(Cheung et al., 2007), six of the peptides are known SARS CD8+ T cell epitopes (all except $N_{351-358}$ and $N_{351-359}$) (Grifoni et al., 2020). We then compared the sequence conservation within each of the peptides. These conserved SARS-CoV-2 peptides (exceptions of $N_{222-230}$ and $N_{226-234}$) had 22%–44% sequence identity or homology with OC43, HKU1, or 229E virus and none with NL63 virus (Table S1).

We established if the selected peptide sequences are representative of circulating SARS-CoV-2 strains and determined their conservation within N protein sequences derived from SARS-CoV-2 viruses circulating within Oceania (4049 sequences), Asia (1,152 sequences), Europe (390 sequences), and North America (10,044 sequences). All peptides were shown to be highly conserved, and our selected peptides are represented in 96%–99% of all circulating strains (Table S2), making them ideal for potential vaccine candidates. Interestingly, the immunogenicity of five of our selected peptides has since been characterized in COVID-19 recovered individuals. Two peptides, $N_{138-146}$ and $N_{159-167}$, were not immunogenic in COVID-19-recovered individuals (Isabel Schulien et al., 2021), whereas three peptides, namely $N_{219-227}$ (Isabel Schulien et al., 2021), $N_{222-230}$ (Isabel Schulien et al., 2021), and $N_{316-324}$ (Isabel Schulien et al., 2021; Habel et al., 2020), were immunogenic in COVID-19 recovered individuals, further highlighting the importance of further investigation into these epitopes.

Overall, the selected N-derived SARS-CoV-2 peptides are representative of the currently circulating strains of the virus and share some similarity with coronaviruses responsible for the common cold.

Some SARS-CoV-2 N-derived peptides poorly stabilized the HLA-A*02:01 molecule

To determine if the 8 N-derived peptides (Table 1) were able to form a complex with HLA-A*02:01, we firstly refolded each peptide separately with the HLA-A*02:01 heavy chain and β2-microglobulin (β2m). Seven of our eight peptides were successfully refolded with the HLA-A*02:01 molecule; however, the $N_{351-358}$ peptide (p) failed to stabilize HLA-A*02:01, as no refolded protein was obtained following purification. This result is not completely surprising, as the $N_{351-358}$ peptide does not have the favored PΩ-Val/Leu typical of HLA-A*02:01 binding (Sette and Sidney, 1999) but instead has a charged Aspartic acid residue. Interestingly, the overlapping $N_{351-359}$ peptide, which has a small hydrophobic Alanine residue at PΩ was able to form a complex with HLA-A*02:01 (details below).

The stability of the pHLA is an important factor, as it influences the half-life of the complex, in turn impacting the likelihood of a peptide being present for long enough at the cell surface to then be recognized by T cells (Blaha et al., 2019; Harndahl et al., 2012). We therefore assessed the stability of the remaining 7 pHLA complexes and compared them to the well characterized HLA-A*02:01-restricted immunogenic influenza M1 peptide (GILGFVFTL) (Valkenburg et al., 2016) using differential scanning fluorimetry (DSF). All 7 HLA-A*02:01-SARS-CoV-2 complexes had a lower thermal shift temperature (Tm) than the highly stable HLA-A*02:01-M1, which exhibited a Tm of ~60°C consistent with previously published reports (Valkenburg et al., 2016) (Table 1, Figure S1). The most stable pHLA complexes were the one with the $N_{222-230}$ (Tm of 55°C) and $N_{316-324}$ (Tm of 49°C) peptides, followed by the one with $N_{219-227}$, $N_{226-234}$, and $N_{351-359}$ peptides with a Tm of ~40°C (Table 1, Figure S1). Surprisingly, the Tm was ~35°C for complexes with the NTD-derived peptides $N_{138-146}$ and $N_{159-167}$, an extremely low Tm value for pHLA-A*02:01 complexes (Valkenburg et al., 2016; Blaha et al., 2019; Khan et al., 2000).

Table 2. Data collection and refinement statistics

	HLA-A*02:01-$N_{138-146}$	HLA-A*02:01-$N_{159-167}$	HLA-A*02:01-$N_{222-230}$	HLA-A*02:01-$N_{226-234}$	HLA-A*02:01-$N_{316-324}$	HLA-A*02:01-$N_{351-359}$
Data collection statistics						
Space group	$P2_12_12_1$	$P2_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions (a,b,c) (Å)	60.03, 78.90, 110.60	53.33, 80.54, 57.06, β = 113.47°	60.16, 79.71, 111.47	60.29, 79. 68, 111.12	60.19, 79.55, 111.58	59.79, 78.25, 111.51
Resolution (Å)	47.77–1.58 (1.61–1.58)	48.92–1.55 (1.58–1.55)	48.02–1.34 (1.36–1.3.4)	48.08–1.90 (1.94–1.90)	48.00–1.40 (1.42–1.40)	47.51–2.15 (2.22–2.15)
Total number of observations	677855 (33071)	442231 (22524)	1621917 (76338)	289860 (42802)	1421661 (58020)	226686 (29202)
Nb of unique obs	72687 (3502)	63881 (3135)	120760 (5789)	17771 (27004)	106332 (4607)	20120 (2485)
Multiplicity	9.3 (9.4)	6.9 (7.2)	13.4 (13.2)	6.8 (6.6)	13.4 (12.6)	7.8 (8.1)
Data completeness (%)	100 (100)	99.6 (99.1)	99.9 (97.5)	99.7 (99.0)	99.4 (88.4)	100 (100)
I/σ_I	16.1 (1.9)	17.5 (1.7)	17.1 (2.0)	10.1 (2.3)	19.3 (2.1)	10.5 (1.7)
R_{pim}[a] (%)	2.8 (43.0)	2.0 (46.4)	2.3 (40.7)	5.7 (55.8)	1.8 (38.5)	5.6 (48.8)
$CC_{1/2}$ (%)	99.9 (67.6)	99.9 (72.8)	99.9 (67.4)	99.5 (75.8)	99.9 (69.7)	99.7 (60.8)
Refinement statistics						
R_{factor}[b] (%)	17.4	19.8	17.9	16.9	19.4	18.3
R_{free}[b] (%)	19.7	22.1	19.8	20.5	20.3	22.9
rmsd from ideality						
Bond lengths (Å)	0.010	0.010	0.010	0.009	0.008	0.008
Bond angles (°)	1.02	1.02	1.04	1.18	0.94	1.10
Ramachandran plot (%)						
Favoured	99.3	99	100	98	98.5	98
Allowed	0.7	1	0	2	1.5	2
Disallowed	0	0	0	0	0	0
PBD code	7KGS	7KGR	7KGQ	7KGT	7KGP	7KGO

[a] $R_{p.i.m} = \Sigma_{hkl}\,[1/(N-1)]^{1/2}\,\Sigma_i \mid I_{hkl,\,i} - \langle I_{hkl}\rangle \mid / \Sigma_{hkl} \langle I_{hkl}\rangle$.
[b] $R_{factor} = \Sigma_{hkl} \mid\mid F_o \mid - \mid F_c \mid\mid / \Sigma_{hkl} \mid F_o \mid$ for all data except ≈ 5%, which were used for R_{free} calculation. Values in parentheses are for the highest resolution shell.

The three peptides that resulted in a pHLA complex with a Tm above 40°C were recently described as immunogenic in COVID-19-recovered patients ($N_{219-227}$ (Isabel Schulien et al., 2021), $N_{222-230}$ (Ferretti et al., 2020; Isabel Schulien et al., 2021), and $N_{316-324}$ (Habel et al., 2020; Isabel Schulien et al., 2021)), whereas the two peptides leading to a pHLA with a Tm below 40°C were described as non-immunogenic ($N_{138-146}$ and $N_{159-167}$ (Isabel Schulien et al., 2021)) (Table 1). This suggests that indeed pHLA complex stability may play a role in peptide immunogenicity.

Together, our data show that HLA-A*02:01 was poorly stable in complex with some of the predicted N-derived SARS-CoV-2 peptides, which will impact T cell response.

SARS-CoV-2 N-derived peptides with low Tm adopt an unusual conformation

To gain a better understanding of the presentation of SARS-CoV-2 peptide by HLA-A*02:01, and why some of these complexes displayed unusually low Tm values, we solved the structures of 6 HLA-A*02:01-SARS-CoV-2 complexes. Unfortunately, the HLA-A*02:01-$N_{219-227}$ structure could not be determined due to low yields of protein required for crystallization. All structures were solved to a high resolution of 1.3–2.15Å (Table 2), showing unbiased electron density for the peptides (Figure S2). Overall, the pHLA structures were

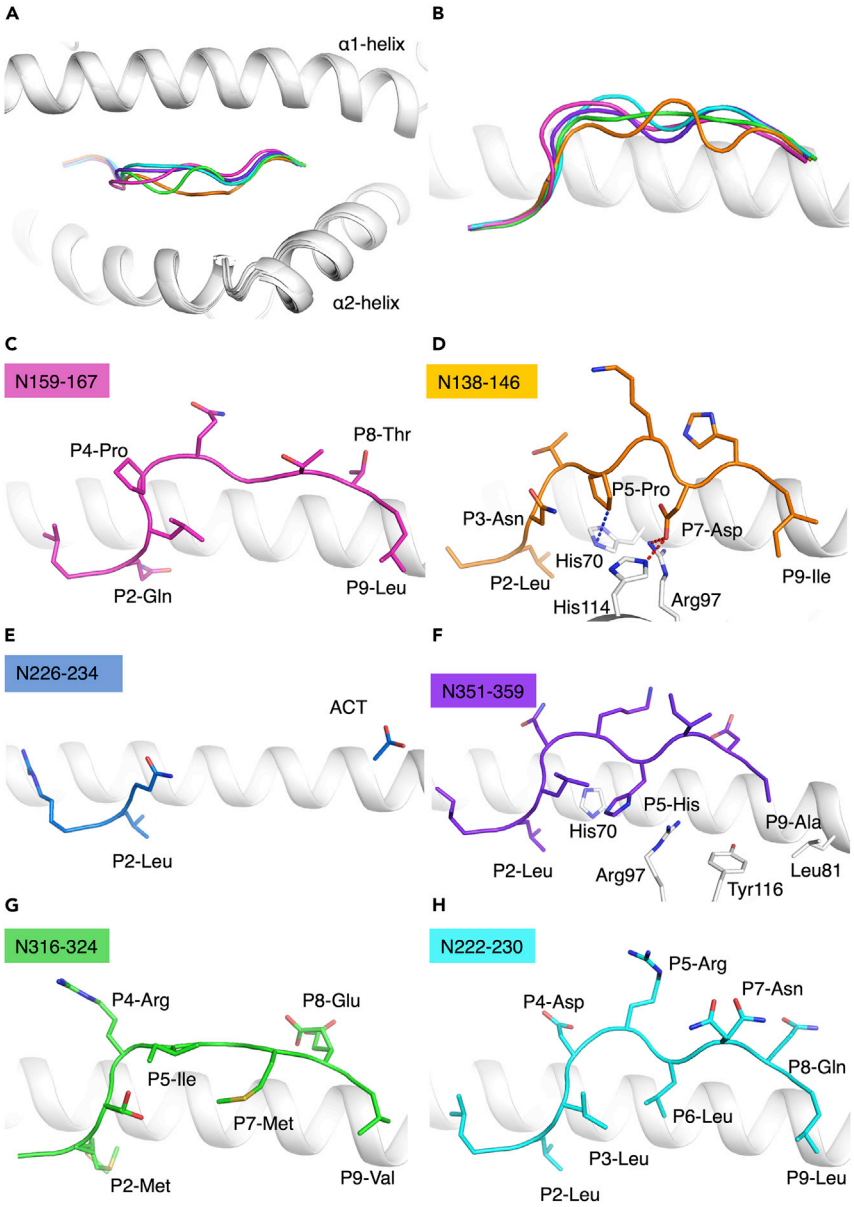


Figure 2. Structures of N-derived SARS-CoV-2 peptides presented by HLA-A*02:01

(A and B) Structural alignment of the six HLA-A*02:01 (white cartoon) binding to N-derived SARS-CoV-2 peptides colored in pink ($N_{159-167}$), orange ($N_{138-146}$), blue ($N_{226-234}$), purple ($N_{351-359}$), green ($N_{316-324}$), and cyan ($N_{222-230}$). A top-down view is represented on panel (A) and side view on panel (B).

(C–H) Each panel shows the structure of each of the six peptides colored as per the top panel, with peptides represented as cartoon and sticks, and HLA-A*02:01 residues important for peptide binding are represented as white sticks. Panel (D) shows red and blue dashed lines that represent hydrogen and van der Waals bonds, respectively, between the HLA and peptide. Panel (E) also contains an acetate (ACT) molecule located at the C-terminal part of the HLA cleft.

similar, with a root-mean-square deviation (r.m.s.d.) on the antigen binding cleft of 0.14–0.22Å between the six structures (Figure 2A) despite the peptides adopting different conformations (Figure 2B).

The $N_{159-167}$ peptide adopts a rather flat conformation in the HLA cleft. From P4-Pro to P8-Thr the peptide lays flat, without any secondary anchor residue interactions within the HLA-A*02:01 molecule (Figure 2C). This conformation is favored by the P4-Pro that provides a kinked conformation observed in several pHLA structures (O'Callaghan et al., 1998). Although the residues P2-Gln and P9-Leu anchor the peptide, often

observed in other HLA-A*02:01-restricted peptides, the lack of secondary anchors might provide the basis for the low Tm value of HLA-A*02:01-$N_{159-167}$ (Table 1).

Conversely, the $N_{138-146}$ peptide has additional anchor residues relative to the other structures, adopting a constrained conformation within the cleft (Theodossis et al., 2010). The P2-Leu and P9-Ile act as conventional primary anchor residues, and three residues acting as secondary anchors, namely P3-Asn, P5-Pro, and P7-Asp (Figure 2D). As a result, the backbone of the peptide adopts a zigzag conformation, with residues pointing up and down spanning P3 to P9. The unusual P5-Pro secondary anchor is located within a pocket that is positively charged, which is not a favourable environment for a hydrophobic Proline residue. This constrained conformation of the $N_{138-146}$ peptide and the presence of multiple secondary anchor residues (P5-Pro and P7-Asp) might explain the low Tm value of the pHLA complex (Table 1).

The $N_{226-234}$ peptide has a P9-Met, which is not anchored into the small hydrophobic F pocket in HLA-A*02:01 within this particular crystal structure. However, two self and one synthetic peptides have been crystallized with a P9-Met binding to the F pocket (Mohammed et al., 2008; Hassan et al., 2015; Riley et al., 2018); therefore other residues within the $N_{226-234}$ peptide seems to destabilize the anchoring of the P9-Met. As a result, the Tm of HLA-A*02:01-$N_{226-234}$ complex was low (39.2°C, Table 1) and despite solving the structure of the pHLA at high resolution (1.9Å, Table 2), the peptide was poorly resolved. We could only observe density for P1, P2, and P3 residues in the Fo-Fc map. The generation of composite omit maps (Afonine et al., 2012) did not reveal any additional residual density for the peptide. The only residual density observed was shown at the location of carboxylic moiety of the PΩ residue; however, after refinement it was clear that a Methionine could not fit in this density. Instead, an acetate ion was placed mimicking the carboxylic group for PΩ (Figure 2E). Although the SARS-CoV-2 $N_{226-234}$ peptide was only able to bind to HLA-A*02:01 by its N-terminal region (Figure 2E) due to an unfavourable PΩ-Met, the homologous peptide from SARS virus possesses a PΩ-Val (Table S1), which is favored within the F pocket, able to stabilize HLA-A*02:01, and stimulate CD8+ T cells (Tsao et al., 2006).

The $N_{351-359}$ peptide overlaps the $N_{351-358}$ peptide (Table 1). The two peptides are different at the C-terminal residue, with a short hydrophobic PΩ-Ala for $N_{351-359}$ and a PΩ-Asp residue in $N_{351-358}$. As discussed earlier, the PΩ-Asp is unfavoured to bind within HLA-A*02:01 and could explain why the $N_{351-358}$ peptide did not refold with HLA-A*02:01. The $N_{351-359}$ peptide PΩ-Ala, despite being hydrophobic, is small and does not fully fill the F pocket as a Val/Leu residue would (Figure 2F). In addition, the P5-His acts as a secondary anchor but unfavorably binding into a positively charged C pocket (Arg97, His70). Altogether, these sub-optimal primary and secondary anchors give rise to a low Tm (42.9°C, Table 1) and leads to poorly defined density around the central part of the peptide that indicates flexibility (Figure S2J).

The low stability of the overall pHLA complexes, and the unusual conformations of the peptides, may explain the lack of immunogenicity of $N_{138-146}$ and $N_{159-167}$, due to a short half-life on the cell surface that would compromise T cell interactions.

Immunogenic N-derived SARS-CoV-2 epitopes share canonical HLA-A*02:01 anchor residues

The $N_{222-230}$ and $N_{316-324}$ peptides have been shown to be weakly immunogenic in a few HLA-A*02:01+ COVID-19-recovered patients (Habel et al., 2020; Isabel Schulien et al., 2021; Ferretti et al., 2020). They both share the preferred P2-Met/Leu and PΩ-Val/Leu characteristics of HLA-A*02:01-restricted peptides (Sette and Sidney, 1999). As a result, both peptides were able to form a stable pHLA complex, which correlate to the high Tm values observed (Table 1).

The HLA-A*02:01-$N_{222-230}$ complex exhibited the highest Tm value (54.7°C, Table 1) among the HLA-A*02:01-N-SARS-CoV-2 complexes tested here and is similar to the HLA-A*02:01-M1 complex (~60°C (Valkenburg et al., 2016)). In addition, the $N_{222-230}$ peptide showed a well-defined electron density (Figures S2F and S2L), suggesting a rigid peptide conformation. The $N_{316-324}$ peptide adopts a rather flat conformation in the cleft, where P5-Ile and P7-Met are only partially buried between the peptide backbone and the HLA α2-helix (Figure 2G). The $N_{316-324}$ peptide P4-Arg and P8-Glu are exposed to the solvent and represent potential contact points for CD8+ T cells. Although the $N_{316-324}$ peptide only has two prominent residues exposing their side chains for potential contact with a TCR (Figure 2G), the $N_{222-230}$ has four large solvent-exposed residues (P4-Asp, P5-Arg, P7-Asn, and P8-Gln, Figure 2H). The P5, P7, and P8 side chains are flexible and allowed to sample their molecular surrounding of the pHLA complex (Figure 2H). The

abundance of solvent-exposed residues in the $N_{222-230}$ peptide offers a variety of potential contact points for interaction with TCRs and therefore recognition by CD8+ T cells.

Altogether, the stability and crystal structures of the $N_{316-324}$ and $N_{222-230}$ SARS-CoV-2 peptides in complex with HLA-A*02:01 show that favored HLA primary anchor residues promote well-defined and stable peptides in HLA, which in this case underlie immunogenicity.

Limited pre-existing immunity toward selected SARS-CoV-2 N-derived peptides in unexposed individuals

Because some level of conservation was seen within these eight peptides between SARS-CoV-2 and the coronaviruses that cause the common cold (OC43, HKU1, 229E, Table S1), we next asked whether HLA-A*02:01 + individuals have some pre-existing immunity toward the SARS-CoV-2 peptides. To test this, we stimulated peripheral blood mononuclear cells (PBMCs) derived from HLA-A*02:01 + individuals without any known infection or exposure to SARS-CoV-2, with a pool of the eight peptides (n = 5 donors) or the immunodominant influenza-derived peptide M1 as a control (n = 3 donors). Functional responses indicative of pre-existing immunity were then assessed using an intracellular cytokine staining (ICS) assay (Figure 3). Limited CD8+ T cell responses were observed toward the pool of peptides in all five donors (Figures 3A and 3D), contrasting to the robust CD8+ T cell responses toward the control influenza-derived M1 peptide (Figures 3B and 3C) consistent with previous reports (Valkenburg et al., 2016). Responses were similarly minimal even when assayed against higher concentrations of the SARS-CoV-2-derived peptides individually (Figure S3). Therefore, even though there was up to 44% sequence identity for $N_{138-146}$ and $N_{159-167}$, between SARS/SARS-CoV-2 and some common cold coronaviruses (Table S1), the remaining differences within the peptides might hinder the cross-reactive potential of CD8+ T cells.

Therefore, despite the conservation of the selected peptides with other commonly circulating coronaviruses (Table S1), there is limited pre-existing CD8+ T cell response toward these SARS-CoV-2-derived peptides in HLA-A*02:01 + individuals.

DISCUSSION

Discovering immunogenic CD8+ T cell epitopes of the SARS-CoV-2 virus is undoubtedly important to help us understand the magnitude and strength of the immune response in COVID-19 patients. However, we currently have limited knowledge on SARS-CoV-2 epitopes and their HLA restriction. Epitope identification is a significant bottleneck when trying to characterize novel viruses. Multiple approaches can be utilized, such as mass spectrometry (Koutsakos et al., 2019) or overlapping peptide screening (Grant et al., 2013), with each method having its own pros and cons. Mass spectrometry can be utilized to identify peptides presented by a selected HLA allele. Although this is beneficial, it is time consuming, expensive, and without additional screening and does not provide any information on the immunogenicity of the identified peptides. Conversely, overlapping peptide screening identifies only peptides that are immunogenic, and additional screening of HLA restriction is required. This method requires large sample sizes and is not efficient if trying to identify peptides for a particular HLA allele, such as HLA-A*02:01, the most prevalent HLA molecule in the global population.

An alternate option is to predict which peptide(s) can bind to an HLA molecule based on its preferred anchor residue for the HLA allele of interest (Sette and Sidney, 1999). In this study, we have characterized eight SARS-CoV-2 peptides predicted to bind to the highly prevalent HLA-A*02:01 molecule (Cheung et al., 2007), six of which are known SARS epitopes. One was unable to form a complex with HLA-A*02:01, and three formed poorly stable complexes. This highlights that, although efficient, peptide prediction algorithms are not always accurate. Indeed, immunogenicity studies by recent groups (Isabel Schulien et al., 2021; Ferretti et al., 2020; Habel et al., 2020), along with our previous work (Grant et al., 2013), have shown that predictive peptides are not always indicative of immunogenic epitopes. However, when used in combination with functional assays, this method can be a fast and ideal way to identify a range of peptides worthy of further characterization particularly when faced with multiple technical challenges such as access to patient samples and limited sample volume (e.g. blood).

A particularly important factor to consider when looking at the T cell response toward viral peptides is their ability to form a stable complex with the HLA, ensuring their presence on cell surfaces for long enough to interact with circulating T cells. We have demonstrated that seven of our eight peptides are able to form a

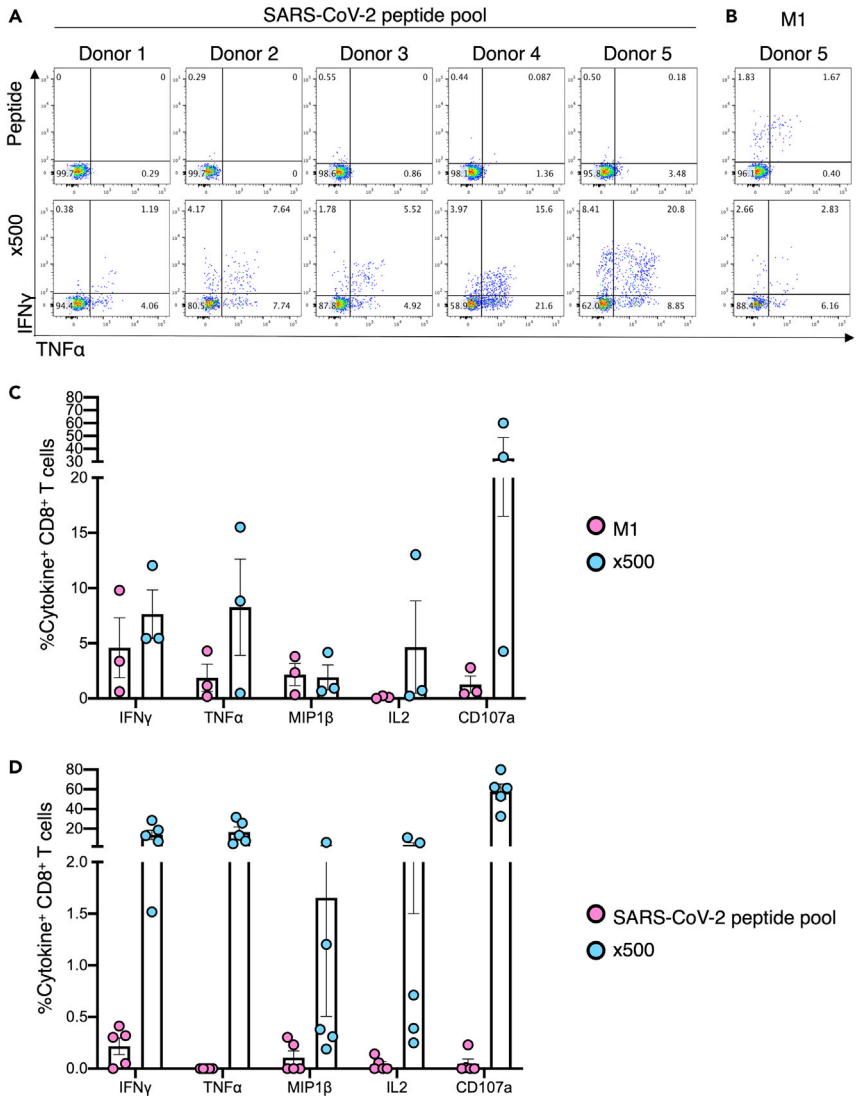


Figure 3. Limited pre-existing CD8$^+$ T cell responses toward SARS-CoV-2 peptides

PBMCs from healthy unexposed HLA-A*02:01 + individuals (n = 5) were stimulated with a pool of eight HLA-A*02:01-restricted SARS-CoV-2 peptides (2 μM of each peptide, 16 μM in total) or the control HLA-A*02:01-restricted influenza M1 peptide for 10 days. CD8+ T cell responses were then assessed using an ICS assay in which CD8+ T cell lines were stimulated with 2 μM of their cognate peptide or peptide pool. Samples were acquired by flow cytometry, and data were gated as per Figure S4.

(A and B) Representative FACS plots of IFNγ+ and TNFα+ production by CD8$^+$ T cell lines in response toward the SARS-CoV-2 peptide pool (A), M1 control (B), or x500 positive control.

(C and D) Summary of IFNγ, TNFα, MIP1β, IL-2, and CD107a production, minus no peptide controls by CD8$^+$ T cell lines in response toward the M1 control (C), SARS-CoV-2 peptide pool (D), versus x500 positive control.

complex with HLA-A*02:01; however, the stability of the complexes ranged from 35–54°C. Interestingly, the immunogenicity of five of our selected peptides has since been investigated, and we noticed a trend between the ability of peptides to form a stable complex with HLA-A*02:01 and their ability to stimulate CD8+ T cells in some COVID-19 recovered individuals (Isabel Schulien et al., 2021; Ferretti et al., 2020; Habel et al., 2020). Indeed, the $N_{219-227}$, $N_{222-230}$, and $N_{316-324}$ peptides have a Tm value above 40°C and have been described as immunogenic in COVID-19-recovered patients (Isabel Schulien et al., 2021; Ferretti et al., 2020; Habel et al., 2020), whereas $N_{138-146}$ and $N_{159-167}$ with a Tm values below 40°C are described as not immunogenic (Isabel Schulien et al., 2021). This correlation between the stability of the pHLA complex and immunogenicity is not surprising and has been observed in previous research (Blaha et al., 2019;

Harndahl et al., 2012), suggesting that Tm should be considered, and may even be predictive, when assessing the immunogenic potential of vaccine peptide candidates.

It is evident that additional research is required to determine which peptides from SARS-CoV-2 are able to activate CD8+ T cells and identify which HLA molecules they are restricted to. These findings will guide our understanding of the immune response, as well as elucidate the potential for a protective and long-lived immune response. We might uncover whether some HLA molecules are better equipped to bind particularly immunogenic viral epitopes capable of stimulating a potent T cell response that could be the target of future vaccines. Indeed, there are numerous studies that describe a strong link between the expression of certain HLA molecules and improved disease outcome (Altfeld et al., 2006; van de Sandt et al., 2019; Valkenburg et al., 2016). For example, in the context of human immunodeficiency virus (HIV) infection, certain individuals are naturally able to control the virus and limit the progression to acquired immunodeficiency syndrome (AIDS); this ability has been linked with the expression of protective HLA alleles such as HLA-B*57:01 and HLA-B*27:05 (Altfeld et al., 2006), whereas HLA-B*35:01 is detrimental and allows rapid progression to AIDs (Altfeld et al., 2006). Similarly, during influenza infection, HLA-A*02:01 seems to provide some protection (Valkenburg et al., 2016), whereas HLA-A*68:01 is associated with poor clinical outcomes (van de Sandt et al., 2019).

Our data provide some rationale for the weak or lack of immunogenicity observed in HLA-A*02:01 + individuals recovered from COVID-19 toward N-derived SARS-CoV-2 peptides, due to unconventional anchor residues and poor stability of pHLA complexes. It is unclear whether HLA-A*02:01 can present immunogenic peptides from SARS-CoV-2 that results in a strong CD8+ T cells response or whether these peptides derived from N protein are not strongly immunogenic for CD8+ T cells. Further work is required to uncover the drivers of protective CD8+ T cells response to SARS-CoV-2 infection.

Limitations of the study

Our study has focused on a set of eight peptides derived from the N protein of SARS-CoV-2 virus, for their conservation with SARS virus as well as their predicted binding to HLA-A*02:01. It is possible that additional N-derived peptides will be able to stably bind HLA-A*02:01 and therefore extend the epitopes repertoire from this SARS-CoV-2 protein. In addition, some additional work will be required to define if the $N_{226-234}$ and $N_{351-359}$ peptides are immunogenic in COVID-19-recovered patients, as well as fully characterize the CD8+ T cells that respond towards these peptides.

Resource availability

Lead contact

Further information and requests for resources and materials should be directed to the Lead Contact, Prof. Stephanie Gras (S.Gras@latrobe.edu.au).

Materials availability

Materials are available upon reasonable request.

Data and code availability

The final crystal structure models for the HLA-A*02:01 complexes have been deposited to the Protein Data-Bank (PDB) under the following accession codes: HLA-A*02:01-$N_{138-146}$ (7KGS), HLA-A*02:01-$N_{159-167}$ (7KGR), HLA-A*02:01-$N_{222-230}$ (7KGQ), HLA-A*02:01-$N_{226-234}$ (7KGT), HLA-A*02:01-$N_{316-324}$ (7KGP), and HLA-A*02:01-$N_{351-359}$ (7KGO).

METHODS

All methods can be found in the accompanying Transparent methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.102096.

ACKNOWLEDGMENTS

We would like to thank all of our healthy volunteers for donating blood, the Australian Red Cross Lifeblood (ARCLB) for access to Buffy coats, and Dr David Sayer from CareDx for his collaboration in HLA typing some

of our samples. We would like to thank the staff at the Monash Flowcore facility, Monash Molecular Crystallisation Facility (MMCF), ANSTO Australian synchrotron staff for their assistance, and the Australian Cancer Research Foundation (ACRF) Eiger detector. This work was funded by the Australian National Health and Medical Research Council (NHMRC), C.S. and D.S.M.C. are supported with an AINSE Early Career Research Grant, E.J.G. was supported by an NHMRC CJ Martin Fellowship (#1110429) and is supported by an Australian Research Council DECRA fellowship (DE210101479), and S.G. is supported by an NHMRC Senior Research Fellowship (#1159272).

AUTHOR CONTRIBUTIONS

C.Szeto. and D.S.M.C. performed experiment and data analysis, supervised the project, and secured funding; D.J. performed experiments and provided reagents; C.Smith. provided intellectual input for project conceptualization; A.T.N., H.S., C.L., and A.R.T. performed experiments and data analysis; E.J.G. and S.G. conceived the project and experiments, completed data analysis, supervised the project, secured funding, and drafted the manuscript. All authors have read and edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: October 19, 2020
Revised: December 17, 2020
Accepted: January 15, 2021
Published: February 19, 2021

REFERENCES

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., and Adams, P.D. (2012). Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallogr. D Biol. Crystallogr. *68*, 352–367.

Ahmed, R., and Akondy, R.S. (2011). Insights into human CD8(+) T-cell memory using the yellow fever and smallpox vaccines. Immunol. Cell Biol. *89*, 340–345.

Altfeld, M., Kalife, E.T., Qi, Y., Streeck, H., Lichterfeld, M., Johnston, M.N., Burgett, N., Swartz, M.E., Yang, A., Alter, G., et al. (2006). HLA alleles associated with delayed progression to AIDS Contribute strongly to the initial CD8(+) T cell response against HIV-1. PLoS Med. *3*, e403.

Amanat, F., and Krammer, F. (2020). SARS-CoV-2 vaccines: status report. Immunity *52*, 583–589.

Bender, B.S., Croghan, T., Zhang, L., and Small, P.A., Jr. (1992). Transgenic mice lacking class I major histocompatibility complex-restricted T cells have delayed viral clearance and increased mortality after influenza virus challenge. J. Exp. Med. *175*, 1143–1145.

Blaha, D.T., Anderson, S.D., Yoakum, D.M., Hager, M.V., Zha, Y., Gajewski, T.F., and Kranz, D.M. (2019). High-throughput stability screening of neoantigen/HLA complexes improves immunogenicity predictions. Cancer Immunol. Res. *7*, 50–61.

Campbell, K.M., Steiner, G., Wells, D.K., Ribas, A., and Kalbasi, A. (2020). Prediction of SARS-CoV-2 epitopes across 9360 HLA class I alleles. bioRxiv.

Chang, C.K., Hou, M.H., Chang, C.F., Hsiao, C.D., and Huang, T.H. (2014). The SARS coronavirus nucleocapsid protein–forms and functions. Antivir. Res. *103*, 39–50.

Cheung, Y.K., Cheng, S.C., Sin, F.W., Chan, K.T., and Xie, Y. (2007). Induction of T-cell response by a DNA vaccine encoding a novel HLA-A*0201 severe acute respiratory syndrome coronavirus epitope. Vaccine *25*, 6070–6077.

Dong, E., Du, H., and Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis. *20*, 533–534.

Ferretti, A.P., Kula, T., Wang, Y., Nguyen, D.M., Adam, W., Dunlap, G.S., Xu, Q., Nabilsi, N., Candace, R.P., Cristofaro, A.W., et al. (2020). COVID-19 Patients Form Memory CD8+ T Cells that Recognize a Small Set of Shared Immunodominant Epitopes in SARS-CoV-2. Immunity *53*, 1095–1107.e3.

Gagneur, A., Sizun, J., Vallet, S., Legr, M.C., Picard, B., and Talbot, P.J. (2002). Coronavirus-related nosocomial viral respiratory infections in a neonatal and paediatric intensive care unit: a prospective study. J. Hosp. Infect. *51*, 59–64.

Grant, E., Wu, C., Chan, K.F., Eckle, S., Bharadwaj, M., Zou, Q.M., Kedzierska, K., and Chen, W. (2013). Nucleoprotein of influenza A virus is a major target of immunodominant CD8+ T-cell responses. Immunol. Cell Biol. *91*, 184–194.

Grant, E.J., Josephs, T.M., Loh, L., Clemens, E.B., Sant, S., Bharadwaj, M., Chen, W., Rossjohn, J., Gras, S., and Kedzierska, K. (2018). Broad CD8(+) T cell cross-recognition of distinct influenza A strains in humans. Nat. Commun. *9*, 5427.

Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R.H., Peters, B., and Sette, A. (2020). A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. Cell Host Microbe *27*, 671–680.e2.

Habel, J.R., Nguyen, T.H.O., van de Sandt, C.E., Juno, J.A., Chaurasia, P., Wragg, K., Koutsakos, M., Hensen, L., Chua, B., Zhang, W., et al. (2020). Suboptimal SARS-CoV-2-specific CD8+ T-cell response associated with the prominent HLA-A*02:01 phenotype. Proc. Natl. Acad. Sci. U S A *117*, 24384–24391.

Harndahl, M., Rasmussen, M., Roder, G., Dalgaard Pedersen, I., Sorensen, M., Nielsen, M., and Buus, S. (2012). Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. Eur. J. Immunol. *42*, 1405–1416.

Hassan, C., Chabrol, E., Jahn, L., Kester, M.G., De Ru, A.H., Drijfhout, J.W., Rossjohn, J., Falkenburg, J.H., Heemskerk, M.H., Gras, S., and Van Veelen, P.A. (2015). Naturally processed non-canonical HLA-A*02:01 presented peptides. J. Biol. Chem. *290*, 2593–2603.

Isabel Schulien, J.K., Oberhardt, V., Wild, K., Seidel, L.M., Sagar, S.K., Daul, F., Salvat Lago, M., Decker, A., Luxenburger, H., Binder, B., et al. (2021). Ex vivo detection of SARS-CoV-2-specific CD8+ T cells: rapid induction, prolonged contraction, and formation of functional memory. Nat. Med. *27*, 78–85.

Khan, A.R., Baker, B.M., Ghosh, P., Biddison, W.E., and Wiley, D.C. (2000). The structure and stability of an HLA-A*0201/octameric tax peptide complex with an empty conserved peptide-N-terminal binding site. J. Immunol. *164*, 6398–6405.

Koutsakos, M., Illing, P.T., Nguyen, T.H.O., Mifsud, N.A., Crawford, J.C., Rizzetto, S., Eltahla, A.A., Clemens, E.B., Sant, S., Chua, B.Y., et al.

(2019). Human CD8(+) T cell cross-reactivity across influenza A, B and C viruses. Nat. Immunol. *20*, 613–625.

Kreijtz, J.H., De Mutsert, G., Van Baalen, C.A., Fouchier, R.A., Osterhaus, A.D., and Rimmelzwaan, G.F. (2008). Cross-recognition of avian H5N1 influenza virus by human cytotoxic T-lymphocyte populations directed to human influenza A virus. J. Virol. *82*, 5161–5166.

McMichael, A.J., Gotch, F.M., Noble, G.R., and Beare, P.A. (1983). Cytotoxic T-cell immunity to influenza. N. Engl. J. Med. *309*, 13–17.

Miller, J.D., Van Der Most, R.G., Akondy, R.S., Glidewell, J.T., Albott, S., Masopust, D., Murali-Krishna, K., Mahar, P.L., Edupuganti, S., Lalor, S., et al. (2008). Human effector and memory CD8+ T cell responses to smallpox and yellow fever vaccines. Immunity *28*, 710–722.

Mohammed, F., Cobbold, M., Zarling, A.L., Salim, M., Barrett-Wilt, G.A., Shabanowitz, J., Hunt, D.F., Engelhard, V.H., and Willcox, B.E. (2008). Phosphorylation-dependent interaction between antigenic peptides and MHC class I: a molecular basis for the presentation of transformed self. Nat. Immunol. *9*, 1236–1243.

O'callaghan, C.A., Tormo, J., Willcox, B.E., Braud, V.M., Jakobsen, B.K., Stuart, D.I., Mcmichael, A.J., Bell, J.I., and Jones, E.Y. (1998). Structural features impose tight peptide binding specificity in the nonclassical MHC molecule HLA-E. Mol. Cell *1*, 531–541.

Peng, Y., Mentzer, A.J., Liu, G., Yao, X., Yin, Z., Dong, D., Dejnirattisai, W., Rostron, T., Supasa, P., Liu, C., et al. (2020). Broad and strong memory CD4(+) and CD8(+) T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. Nat. Immunol. *21*, 1336–1345.

Riley, T.P., Hellman, L.M., Gee, M.H., Mendoza, J.L., Alonso, J.A., Foley, K.C., Nishimura, M.I., Vander Kooi, C.W., Garcia, K.C., and Baker, B.M. (2018). T cell receptor cross-reactivity expanded by dramatic peptide-MHC adaptability. Nat. Chem. Biol. *14*, 934–942.

Roltgen, K., Wirz, O.F., Stevens, B.A., Powell, A.E., Hogan, C.A., Najeeb, J., Hunter, M., Sahoo, M.K., Huang, C., Yamamoto, F., et al. (2020). SARS-CoV-2 antibody responses correlate with resolution of RNAemia but are short-lived in patients with mild illness. medRxiv.

Rydyznski Moderbacher, C., Ramirez, S.I., Dan, J.M., Grifoni, A., Hastie, K.M., Weiskopf, D., Belanger, S., Abbott, R.K., Kim, C., Choi, J., et al. (2020). Antigen-specific adaptive immunity to SARS-CoV-2 in acute COVID-19 and associations with age and disease severity. Cell *183*, 996–1012.e19.

van de Sandt, C.E., Kreijtz, J.H., De Mutsert, G., Geelhoed-Mieras, M.M., Hillaire, M.L., Vogelzang-Van Trierum, S.E., Osterhaus, A.D., Fouchier, R.A., and Rimmelzwaan, G.F. (2014). Human cytotoxic T lymphocytes directed to seasonal influenza A viruses cross-react with the newly emerging H7N9 virus. J. Virol. *88*, 1684–1693.

van de Sandt, C.E., Hillaire, M.L., Geelhoed-Mieras, M.M., Osterhaus, A.D., Fouchier, R.A., and Rimmelzwaan, G.F. (2015). Human influenza A virus-specific CD8+ T-cell response is long-lived. J. Infect. Dis. *212*, 81–85.

van de Sandt, C.E., Clemens, E.B., Grant, E.J., Rowntree, L.C., Sant, S., Halim, H., Crowe, J., Cheng, A.C., Kotsimbos, T.C., Richards, et al.. (2019). Challenging immunodominance of influenza-specific CD8(+) T cell responses restricted by the risk-associated HLA-A*68:01 allomorph. Nat. Commun. *10*, 5579.

Sette, A., and Sidney, J. (1999). Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immunogenetics *50*, 201–212.

Theodossis, A., Guillonneau, C., Welland, A., Ely, L.K., Clements, C.S., Williamson, N.A., Webb, A.I., Wilce, J.A., Mulder, R.J., Dunstone, M.A., et al. (2010). Constraints within major histocompatibility complex class I restricted peptides: presentation and consequences for T-cell recognition. Proc. Natl. Acad. Sci. U S A *107*, 5534–5539.

Tsao, Y.P., Lin, J.Y., Jan, J.T., Leng, C.H., Chu, C.C., Yang, Y.C., and Chen, S.L. (2006). HLA-A*0201 T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus nucleocapsid and spike proteins. Biochem. Biophys. Res. Commun. *344*, 63–71.

Vabret, N. (2020). Antibody responses to SARS-CoV-2 short-lived. Nat. Rev. Immunol. *20*, 519.

Valkenburg, S.A., Gras, S., Guillonneau, C., La Gruta, N.L., Thomas, P.G., Purcell, A.W., Rossjohn, J., Doherty, P.C., Turner, S.J., and Kedzierska, K. (2010). Protective efficacy of cross-reactive CD8+ T cells recognising mutant viral epitopes depends on peptide-MHC-I structural interactions and T cell activation threshold. PLoS Pathog. *6*, e1001039.

Valkenburg, S.A., Josephs, T.M., Clemens, E.B., Grant, E.J., Nguyen, T.H., Wang, G.C., Price, D.A., Miller, A., Tong, S.Y., Thomas, et al.. (2016). Molecular basis for universal HLA-A*0201-restricted CD8+ T-cell immunity against influenza viruses. Proc. Natl. Acad. Sci. U S A *113*, 4440–4445.

Weiskopf, D., Schmitz, K.S., Raadsen, M.P., Grifoni, A., Okba, N.M.A., Endeman, H., Van Den Akker, J.P.C., Molenkamp, R., Koopmans, M.P.G., Van Gorp, E.C.M., et al. (2020). Phenotype and kinetics of SARS-CoV-2-specific T cells in COVID-19 patients with acute respiratory distress syndrome. Sci. Immunol. *5*, eabd2071.

Wells, M.A., Albrecht, P., and Ennis, F.A. (1981). Recovery from a viral respiratory infection. I. Influenza pneumonia in normal and T-deficient mice. J. Immunol. *126*, 1036–1041.

iScience, Volume *24*

Supplemental Information

The presentation of SARS-CoV-2 peptides by the common HLA-A*02:01 molecule

Christopher Szeto, Demetra S.M. Chatzileontiadou, Andrea T. Nguyen, Hannah Sloane, Christian A. Lobos, Dhilshan Jayasinghe, Hanim Halim, Corey Smith, Alan Riboldi-Tunnicliffe, Emma J. Grant, and Stephanie Gras

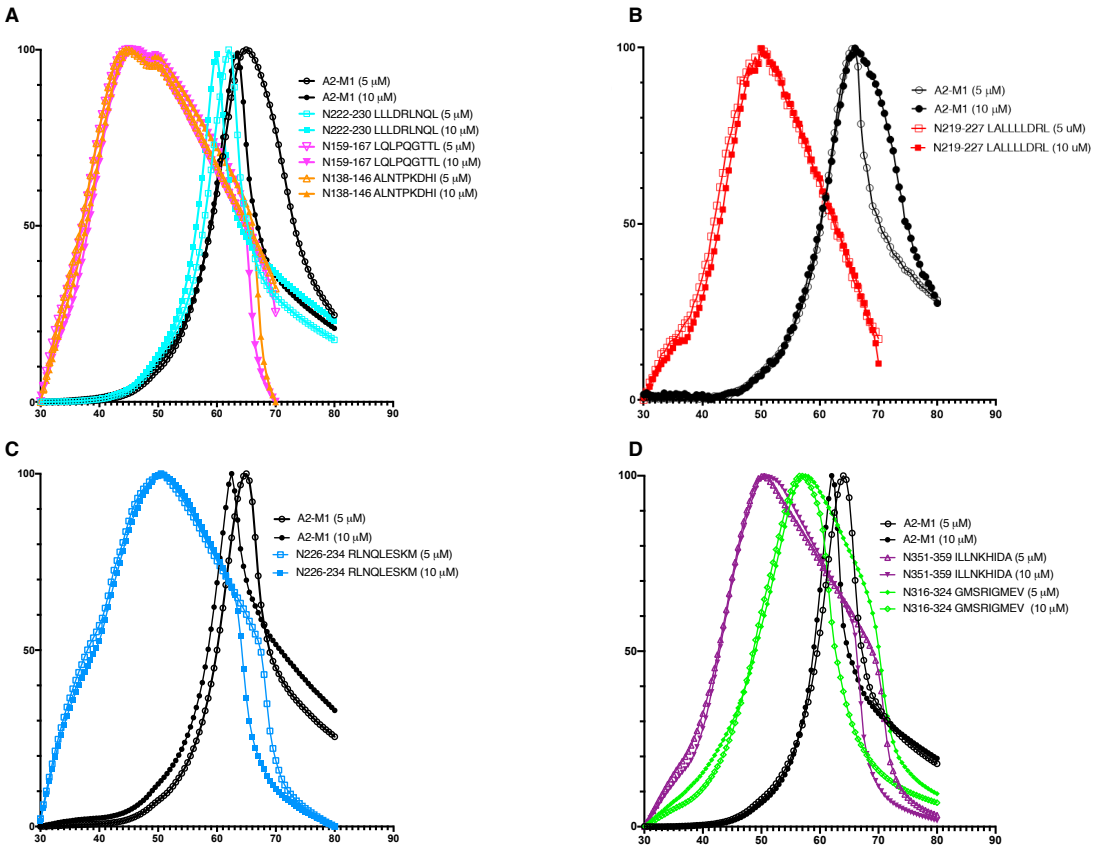


Figure S1. Differential scanning fluorimetry plots for N-derived SARS-CoV-2 peptides presented by HLA-A*02:01, Related to Table 1.

DSF plots show normalized fluorescence intensity versus temperature for seven HLA-A*02:01 SARS-CoV-2 peptide complexes measured at 5 and 10 μM concentrations for **A**) $N_{138-146}$ (orange), $N_{159-167}$ (magenta), $N2_{22-230}$ (cyan), **B**) $N_{219-227}$ (red), **C**) $N_{226-234}$ (blue), **D**) $N_{316-324}$ (green), and $N_{351-359}$ (purple). Each DSF experiment is performed with HLA-A*02:01 presenting the Influenza A $M1_{58-66}$ peptide (GILGFVFTL) as a control (A2-M1; black). Each curve is the average of duplicate measurements, with the current plots showing one representative dataset of two independent experiments (n=2).

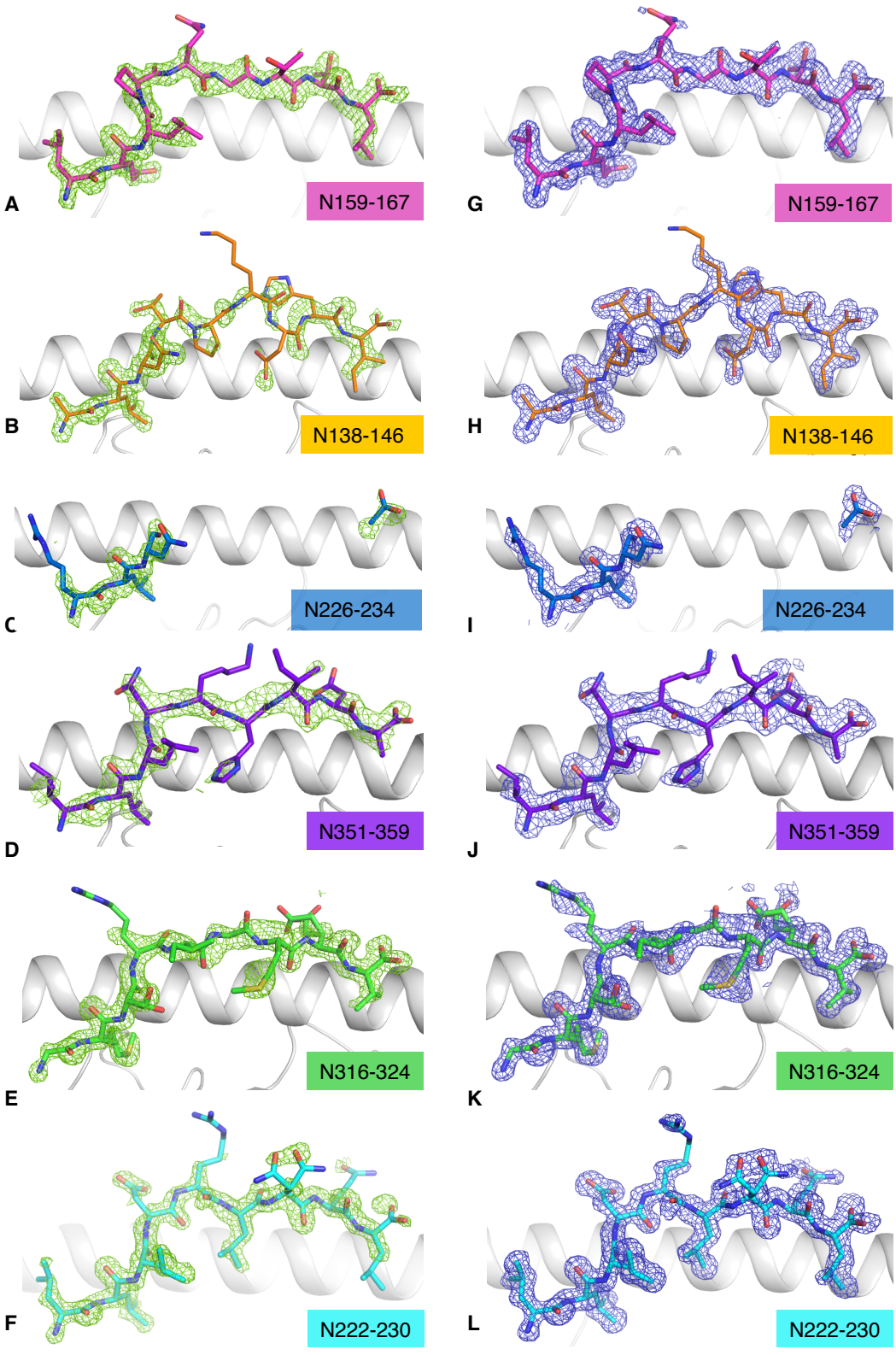


Figure S2. Electron densities omit and refined maps for each HLA-A*02:01-SARS-CoV-2-N complexes, Related to Figure 2.

Density map for the structures of the six HLA-A*02:01 (white cartoon) binding to N-derived SARS-CoV-2 peptides coloured in pink ($N_{159-167}$, **A** and **G**), orange ($N_{138-146}$, **B** and **H**), blue ($N_{226-234}$, **C** and **I**), purple ($N_{351-359}$, **D** and **J**), green ($N_{316-324}$, **E** and **K**), and cyan ($N_{222-230}$, **F** and **L**). The peptides are represented as stick, (**A-F**) the omit 2Fo-Fc map is contoured at 3 sigma and coloured in green on the left panels, while the electron density after refinement is shown by a blue Fo-Fc map contoured at 1 sigma on the right panels (**G-L**).

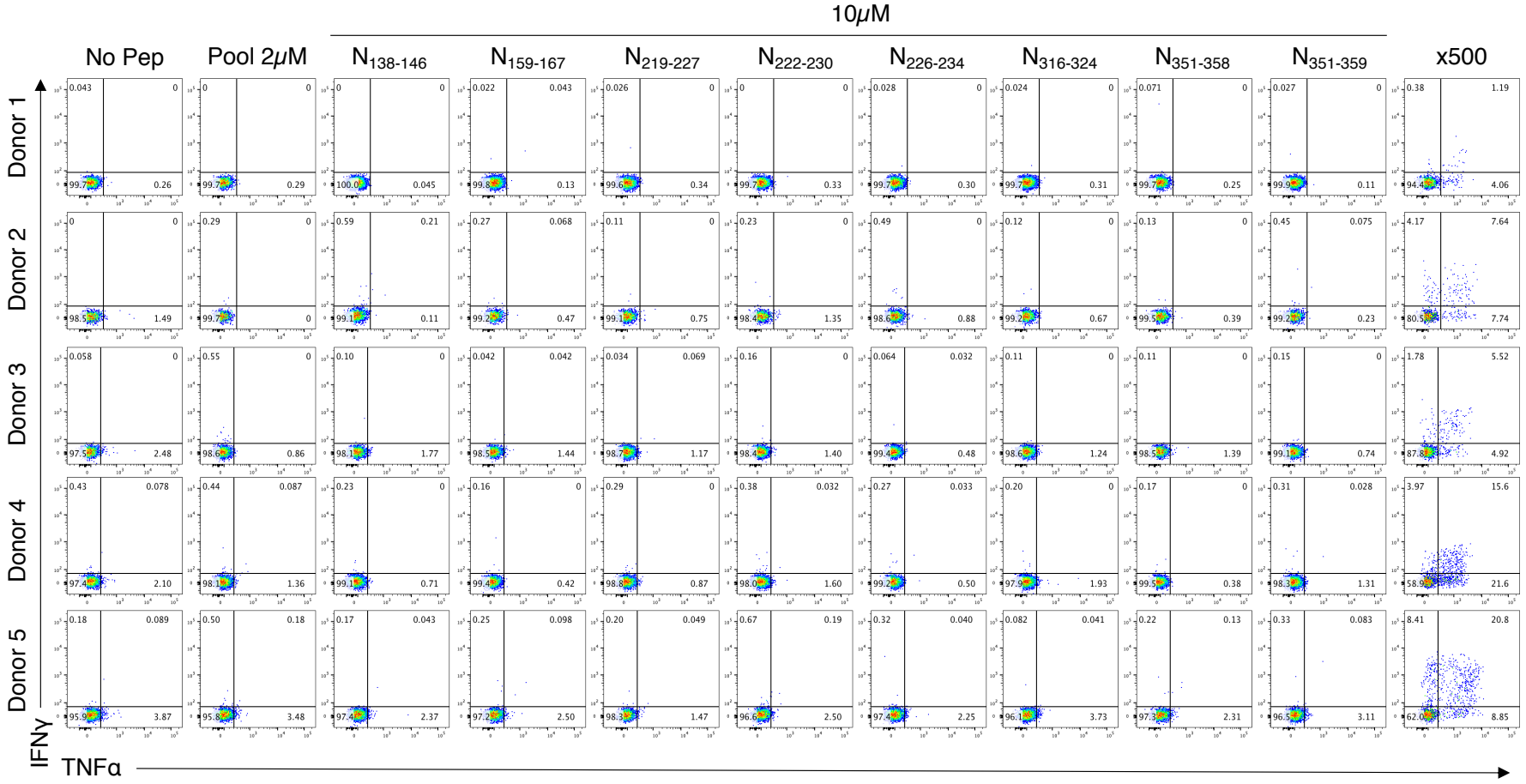


Figure S3. Limited pre-existing CD8+ T cell responses toward high concentration of HLA-A*02:01-restricted SARS-CoV-2 peptides, Related to Figure 3.

PBMCs from healthy unexposed HLA-A*02:01+ individuals (n=5) were stimulated with a pool of eight HLA-A*02:01-restricted SARS-CoV-2 peptides (2 μM of each peptide, 16 μM in total) and CD8+ T cell responses were assessed in an ICS assay by stimulating with 2 μM of the SARS-CoV-2 peptide pool or 10 μM of each of the peptides individually. Samples were acquired by flow cytometry and data are gated as per

Figure S3. FACS plots represent IFNγ+ and TNFα+ production by CD8+ T cell lines in response to the peptide pool or individual peptides or controls.

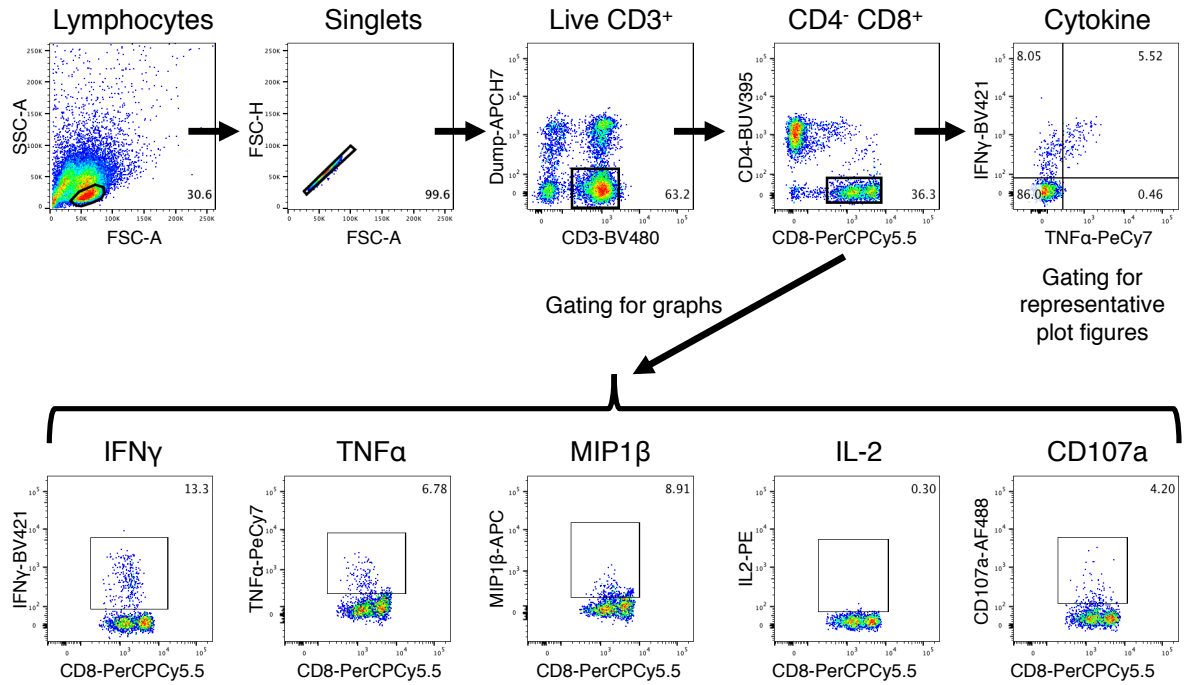


Figure S4. Gating strategy used for this study, Related to Figure 3.
Gating strategy used for the assessment of CD8+ T cell line specificity by ICS in Figure 3 and Figure S3.

	$N_{138-146}$	$N_{159-167}$	$N_{219-227}$	$N_{222-230}$	$N_{226-234}$	$N_{316-324}$	$N_{351-358}$	$N_{351-359}$
SARS-CoV-2	**AL**NTPKDH**I**	L**Q**LPQGTT**L**	L**A**LLLLDR**L**	L**L**DRLNQ**L**	R**L**NQLESK**M**	G**M**SRIGME**V**	I**L**LNKHI**D**	I**L**LNKHID**A**
SARS	**AL**NTPKDH**I**	L**Q**LPQGTT**L**	L**A**LLLLDR**L**	L**L**DRLNQ**L**	R**L**NQLESK**V**	G**M**SRIGME**V**	I**L**LNKHI**D**	I**L**LNKHID**A**
OC43	D**V**NTPAD**I**V	T**R**FPPGT**V**L	M**A**DQIASL**V**			F**G**SRLELA**K**	K**V**LNENL**N**	K**V**LNENLN**A**
HKU1	D**T**STPSD**V**S	T**R**FPPGTI**L**	M**A**DEIANL**V**				K**V**LNENL**N**	K**V**LNENLN**A**
229E			S**Q**DDIMKA**V**			F**D**SHIVSK**E**		

Table S1. N-derived peptide conservation between SARS-CoV-2 and other coronaviruses, Related to Table 1 and Figure 1.
Bold: anchor residues (P2 and PΩ), red: identical, blue: similar.

Peptide	Sequence	Oceania	Asia	Europe	Nth America
$N_{138-146}$	**ALNTPKDHI**	99.75	99.39	100.00	99.66
	ALNTPKHHI	0.12			
	AFNTPKDHI		0.26		
	ALNTPKYHI		0.17		0.13
	ALNIPKDHI		0.17		
$N_{159-167}$	**LQLPQGTTL**	99.93	99.74	100.00	99.93
	LQLPKGTTL		0.17		
$N_{219-227}$	**LALLLLDRL**	99.98	100.00	99.23	99.95
	LVLLLLDRL			0.77	
$N_{222-230}$	**LLLDRLNQL**	96.00	99.65	99.74	99.88
	LLLDRLNQF	3.98		0.26	
	LLLDRLNHL		0.35		0.11
$N_{226-234}$	**RLNQLESKM**	95.88	99.05	99.49	99.52
	RLNQFESKM	3.98		0.26	
	RLNQLESKI	0.10	0.61	0.26	0.30
	RLNHLESKM		0.35		0.11
$N_{316-324}$	**GMSRIGMEV**	99.98	100.00	100.00	99.96
$N_{351-358}$	**ILLNKHID**	100.00	100.00	100.00	100.00
$N_{351-359}$	**ILLNKHIDA**	99.90	100.00	100.00	99.99
	ILLNKHIDS	0.10			
Total number of sequences		***4,049****	***1,152***	***390***	***10,044**[#]*

Table S2. Conservation frequency of peptides within circulating SARS-CoV-2 isolates, Related to Table 1.

Sequences were obtained from the NCBI virus database http://www.ncbi.nlm.nih.gov/labs/virus and were aligned using https://www.fludb.org/brc/home.spg?decorator=influenza. The conservation frequency of the peptides is indicated within the table, with frequencies ≥ 0.1% only shown. Peptides used within this paper are denoted in bold, and mutations from this peptide are denoted in blue. All sequences were used unless outlined below. The software was unable to align some sequences while sequences containing an unknown amino acid within the peptide were eliminated from the analysis. *4048, 4046 and 4046 sequences were analysed for peptides $N_{159-167}$, $N_{316-324}$ and $N_{351-359}$, respectively. [#] 10,043, 10,040, 10,043, 100,39, 10,040 and 10,037 sequences were aligned for peptides $N_{138-146}$, $N_{159-167}$, $N_{226-234}$, $N_{316-324}$, $N_{351-358}$ and $N_{351-358}$, respectively.

TRANSPARENT METHODS

Sequence alignment

Sequences of coronavirus nucleocapsid were obtained from the NCBI server for SARS-CoV-2 (YP_009724397.2) (Wu et al., 2020), SARS (ACB69868.1), OC43 (YP_009555245.1) (St-Jean et al., 2004), 229E (AAG48597.1) (Thiel et al., 2001), HKU1 (ABG77571.1) (Vabret et al., 2006), and NL63 (Q6Q1R8.1) (van der Hoek et al., 2004). Multiple sequence alignment of full-length nucleocapsid proteins was performed using CLUSTALW (Thompson et al., 1994) on the PRABI server.

Ethics, Donors and HLA typing

All work using human samples was approved by the Monash University Human Ethics Committee (#19079 and #18380). Buffy coats were obtained from the Australian Red Cross Lifeblood and whole blood donations were received from healthy volunteers. All donors provided written and informed consent on the day of their donation. HLA typing of our donors was undertaken by the Victorian Transplant and Immunogenetics Service (VTIS, Melbourne, VIC, Australia) or using AlloSeq Tx17 (CareDx Pty Ltd, Fremantle, Australia). All donors used in this study are healthy at the time of donation and are not known to have been infected with or exposed to SARS-CoV-2, and are therefore referred to as "unexposed" throughout the manuscript.

PBMC Isolation

Peripheral blood mononuclear cells (PBMCs) were isolated using density gradient centrifugation. In brief, samples were diluted with RPMI-1640 (Gibco, 10mL/50mL whole blood donation or 60mL/buffy coat), 30mL overlayed gently onto 10mL Ficol Paque-Plus (GE Healthcare), and centrifuged 2000 rpm for 25 mins at room temperature. PBMCs were harvested from the ficol-media interface, washed 3 times using RPMI and were cryopreserved in fetal calf serum (FCS; Thermofisher, Scientifix) supplemented with 10% DMSO (Sigma) until use.

$CD8^+$ T cell lines

$CD8^+$ T cell lines were generated at a 1:2 stimulators to responders' ratio from ~$5x10^6$ (M1 control) or ~$1x10^7$ PBMCs (peptide pool). PBMCs were thawed in warm RF10 (RPMI-1640 life technologies) supplemented with 1x Non-essential amino acids (NEAA; Sigma), 5 mM

HEPES (Sigma), 2 mM L-glutamine (Sigma), 1x penicillin/streptomycin/Glutamine (Life Technologies), 50 μM 2-ME (Sigma) and 10% heat-inactivated (FCS; Thermofisher, Scientifix), and washed by centrifugation. PBMCs were subsequently washed and resuspended in FCS-free media (R0; RPMI supplemented with 1x non-essential amino acids (NEAA; Sigma), 5 mM HEPES (Sigma), 2 mM L-glutamine (Sigma), 1x penicillin/streptomycin/glutamine (Life Technologies), 50 μM 2-ME (Sigma). Responder PBMCs were placed in a 24 well (peptide pool) or 48 well (M1 control) plate, topped with RF10 and left to rest at 37^{o}C with 5% CO_2. Stimulator PBMCs (1/3) were peptide pulsed with either 16uM of the peptide pool (final concentration of 2 μM per peptide), or 2 μM of the M1 control peptide for 90mins at 37^{o}C with 5% CO_2. Stimulators were washed twice in R0, resuspended in RF10 and added to the responders and T cell lines were left to grow at 37^{o}C with 5% CO_2. $CD8^+$ T cell lines were supplemented with fresh media and 10IU/mL IL-2 (Peprotech) twice weekly from day 3 or 4.

Intracellular cytokine staining

$CD8^+$ T cell lines were harvested, washed and resuspended in RF10 supplemented with 10IU/mL IL2. ~$2x10^5$ cells were stimulated directly with 2 μM per peptide (total 16 μM for the peptide pool) in a 96 well U bottom plate and were left to activate for 5hrs in the presence of Golgi Plug (BD Biosciences), Golgi Stop (BD Biosciences) and anti-human CD107a-AF488 (eBiosceince). Following activation, cells were surface stained for 30mins at 4^{o}C with Live/Dead-NIR (1:1000; Molecular Probes), anti-human CD3-BV480 (1:100), anti-human CD4-BUV395 (1:50) and anti-human CD8-PerCPCy5.5 (1:100), anti-human CD14-APCH7 (1:200) and CD19-APCH7 (1:100; all BD Biosciences) in Phosphate buffered saline (PBS). Cells were washed with PBS and then fixed for 30mins at 4^{o}C using Fix-Perm buffer (BD Biosciences). Cells were washed with Perm-Wash buffer (BD Biosciences) and intracellularly stained for 30mins at 4^{o}C with anti-human IFNg-BV421 (1:100), anti-human TNFα-PeCy7 (1:100), anti-human MIP1β-APC (1:100) and anti-human IL2-PE (1:50; all BD Biosciences) in Perm-wash buffer. Samples were washed, resuspended in PBS and acquired on a BD Fortessa (BD Biosciences). Data were analysed using Flowjo software V10.5.3 (Treestar) and graphically represented using GraphPad Prism V8 (GraphPad).

Conservation of SARS-CoV-2 derived peptides

Full length (419 amino-acid) nucleocapsid SARS-CoV-2-specific protein sequences from Oceania (4,049 sequences), Asia (1,152), Europe (390) and North America (10,044) were

obtained from the NCBI virus database http://www.ncbi.nlm.nih,gov/labs/virus on the 9th of Sept 2020. Default filters were used except for Sequence (GenBank), Nucleotide completeness (complete), Host (Homo sapiens) and Geographic Region (selected as required). Sequences were aligned using https://www.fludb.org/brc/home.spg?decorator=influenza.

Protein expression, refold, purification

DNA plasmids encoding the HLA-A*02:01 heavy chain and β2-microglobulin were transformed into the BL21 strain of E. *coli* cells. Both proteins were expressed separately as inclusion bodies and purified from the transformed E. *coli* cells. Soluble peptide-HLA complexes were produced by refolding inclusion bodies in the following amounts: 30 mg of α-chain, 10 mg of β2-microglobulin and 5 mg of peptide (Genscript). The refold mixture was dialysed into 10 mM Tris-HCl pH 8.0 and pHLA was purified using anion exchange chromatography (HiTrap Q).

Differential Scanning Fluorimetry

Differential Scanning Fluorimetry was carried out in a Qiagen RG6 real-time PCR machine, where pHLA samples were heated from 30 to 95˚C at a rate of 0.5˚C/min with excitation and emission channels set at yellow (excitation of ~530 nm and detection at ~557 nm). The experiment was performed at two concentrations of pHLA (5 µM and 10 µM) in duplicate. Each sample was dialysed in 10mM Tris-HCl pH 8.0, 150mM NaCl and contained a final concentration of 10X SYPRO Orange Dye. Fluorescence intensity data was normalised and plotted using GraphPad Prism 8 (version 8.4.2). The Tm value for a pHLA is equal to its the temperature when 50% of maximum fluorescence intensity is reached.

Crystallisation and structure determination

Crystals of pHLA complexes were grown via sitting-drop, vapour diffusion at 20˚C with a protein: reservoir drop ratio of 1:1, at a concentration of 5 mg/mL in 10 mM Tris-HCl pH 8, 150 mM NaCl. Crystals of HLA-A*02:01 in complex with $N_{138-146}$ and $N_{351-359}$ were grown in 20% PEG3350 w/v, 0.2 M NaFormate, 1 mM $CdCl_2$; with $N_{159-167}$ in 20% PEG3350 w/v, 0.2 M NaFormate; with $N_{222-230}$ in 20% PEG3350, 0.2M KFormate, 1mM $CaCl_2$; with $N_{226-234}$ in 20% PEG3350 w/v, 0.2M KFormate, 1mM $CdCl_2$, and with $N_{316-324}$ in 20% 3350, 0.2M NaFluoride, 1 mM $CdCl_2$. pHLA crystals were soaked in a cryoprotectant solution containing mother liquor solution with the PEG3350 concentration increased to 30% (w/v) and then

flash-frozen in liquid nitrogen. The data were collected on the MX2 beamline at the Australian Synchrotron, part of ANSTO, Australia (Aragao et al., 2018). The data were processed using XDS (Kabsch, 2010) and the structures were determined by molecular replacement using the PHASER program (McCoy et al., 2007) from the CCP4 suite (1994) with a model of HLA-A*02:01 without the peptide (derived from PDB ID: 3GSO (Gras et al., 2009)). Manual model building was conducted using COOT (Emsley et al., 2010
) followed by refinement with BUSTER (Bricogne G. and Roversi P, 2011). The final models have been validated and deposited using the wwPDB OneDep System and the final refinement statistics, PDB codes are summarized in **Table 2**. All molecular graphics representations were created using PyMOL.

Supplemental References

1994. The CCP4 suite: programs for protein crystallography. *Acta crystallographica. Section D, Biological crystallography,* 50**,** 760-3.

ARAGAO, D., AISHIMA, J., CHERUKUVADA, H., CLARKEN, R., CLIFT, M., COWIESON, N. P., ERICSSON, D. J., GEE, C. L., MACEDO, S., MUDIE, N., PANJIKAR, S., et al. 2018. MX2: a high-flux undulator microfocus beamline serving both the chemical and macromolecular crystallography communities at the Australian Synchrotron. *J Synchrotron Radiat,* 25**,** 885-891.

BRICOGNE G., B. E., BRANDL M., FLENSBURG C., KELLER P., PACIOREK W., & ROVERSI P, S. A., SMART O.S., VONRHEIN C., WOMACK T.O. 2011. Buster version 2.10. *Cambridge, United Kingdom: Global Phasing Ltd.*

EMSLEY, P., LOHKAMP, B., SCOTT, W. G. & COWTAN, K. 2010
. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr,* 66**,** 486-501.

GRAS, S., SAULQUIN, X., REISER, J. B., DEBEAUPUIS, E., ECHASSERIEAU, K., KISSENPFENNIG, A., LEGOUX, F., CHOUQUET, A., LE GORREC, M., MACHILLOT, P., et al. 2009. Structural bases for the affinity-driven selection of a public TCR against a dominant human cytomegalovirus epitope. *J Immunol,* 183**,** 430-7.

KABSCH, W. 2010. Xds. *Acta Crystallogr D Biol Crystallogr,* 66**,** 125-32.

MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *J Appl Crystallogr,* 40**,** 658-674.

ST-JEAN, J. R., JACOMY, H., DESFORGES, M., VABRET, A., FREYMUTH, F. & TALBOT, P. J. 2004. Human respiratory coronavirus OC43: genetic stability and neuroinvasion. *J Virol,* 78**,** 8824-34.

THIEL, V., HEROLD, J., SCHELLE, B. & SIDDELL, S. G. 2001. Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J Gen Virol,* 82**,** 1273-1281.

THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,

position-specific gap penalties and weight matrix choice. *Nucleic Acids Res,* 22**,** 4673-80.

VABRET, A., DINA, J., GOUARIN, S., PETITJEAN, J., CORBET, S. & FREYMUTH, F. 2006. Detection of the new human coronavirus HKU1: a report of 6 cases. *Clin Infect Dis,* 42**,** 634-9.

VAN DER HOEK, L., PYRC, K., JEBBINK, M. F., VERMEULEN-OOST, W., BERKHOUT, R. J., WOLTHERS, K. C., WERTHEIM-VAN DILLEN, P. M., KAANDORP, J., SPAARGAREN, J. & BERKHOUT, B. 2004. Identification of a new human coronavirus. *Nat Med,* 10**,** 368-73.

WU, F., ZHAO, S., YU, B., CHEN, Y. M., WANG, W., SONG, Z. G., HU, Y., TAO, Z. W., TIAN, J. H., PEI, Y. Y., et al. 2020. A new coronavirus associated with human respiratory disease in China. *Nature,* 579**,** 265-269.