Prediction of SARS-CoV-2 epitopes across 9360 HLA class I alleles

Authors:
Katie M. Campbell*[1], Gabriela Steiner[2], Daniel K. Wells[2], Antoni Ribas[1,2,4,5], Anusha Kalbasi*[3,4,5]

Affiliations:
[1]Department of Medicine, Division of Hematology-Oncology, University of California, Los Angeles (UCLA), Los Angeles, CA, 90095, USA
[2]Parker Institute for Cancer Immunotherapy, San Francisco, CA, 94129, USA
[3]Department of Radiation Oncology, UCLA, CA, 90095, USA
[4]Department Surgery, Division of Surgical Oncology, University of California, Los Angeles, Los Angeles, CA, USA
[5]Jonsson Comprehensive Cancer Center, Los Angeles, CA, USA

*To whom correspondence should be addressed:
Katie Campbell, Ph.D., Department of Medicine, Division of Hematology-Oncology, 9-666 Factor Building, 700 Tiverton Avenue, Los Angeles, CA 90095. Email: KatieCampbell@mednet.ucla.edu.
Anusha Kalbasi, M. D., Department of Radiation Oncology, Jonsson Comprehensive Cancer Center (JCCC), UCLA; B-262 Factor Building, 700 Tiverton Avenue, Los Angeles, CA 90095. Phone: (310) 267-4831; Email: AnushaKalbasi@mednet.ucla.edu.

Abstract

Elucidating antiviral CD8 T lymphocyte responses to SARS-CoV-2 may shed light on the heterogeneity of clinical outcomes and inform vaccine or therapeutic approaches. To facilitate the evaluation of antiviral CD8 T cell responses to SARS-CoV-2, we generated a publicly accessible database of epitopes predicted to bind any class I HLA protein across the entire SARS-CoV-2 proteome. While a subset of epitopes from earlier betacoronaviruses, such as SARS-CoV (SARS), have been validated experimentally, validation systems are often biased toward specific HLA haplotypes (notably HLA-A*02:01) that only account for a fraction of the haplotypes of individuals affected by the SARS-CoV-2 pandemic. To enable evaluation of epitopes across individuals with a variety of HLA haplotypes, we computed the predicted binding affinities between 9-mer peptides derived from the annotated SARS-CoV-2 peptidome across 9,360 MHC class I HLA-A, -B, and -C alleles. There were 6,748 unique combinations of peptides and HLA alleles (pMHCs) with a predicted binding affinity of less than 500nM, including 1,103 unique peptides and 1,022 HLA alleles, spanning 11 annotated superfamilies. These peptides were derived from all 11 proteins spanning the SARS-CoV-2 peptidome, including peptides that have previously been validated experimentally. We also show evidence that these previously validated epitopes may be relevant in other HLA contexts. This complete dataset is available publicly: gs://pici-covid19-data-resources/mhci/peptide_predictions.

Introduction

In a subset of patients, SARS-CoV-2 can result in COVID-19, which can be a deadly disease with hallmarks of acute respiratory distress syndrome (ARDS)(*1*). Other patients harboring SARS-CoV-2 can have minimal or no symptoms. While clinical factors such as older age and underlying medical conditions have been reported as potential risk factors for more severe disease(*2*), the heterogeneity of clinical response to this viral agent is otherwise poorly understood.

Cytotoxic $CD8^+$ T lymphocyte responses are crucial for initial viral clearance and development of immunologic memory(*3*). CD8 T cell responses are a plausible contributor to immunopathology following viral infection(*4*). Excessive cytotoxic CD8 T cell responses and resultant cytokine release by these and other cells may contribute to ARDS, while less effective CD8 T cell responses may allow progression of viral pathology.

The heterogeneity in CD8 T cell responses to SARS-CoV-2 may in part be related to the capacity to recognize the viral antigens in the context of class I major histocompatibility complex (MHC) proteins. Indeed, genetic susceptibilities to viral infection have been tied to human leukocyte antigen (HLA) haplotypes(*5*). Meanwhile, functional differences in antigen-specific CD8 T cell responses in symptomatic and asymptomatic patients may also contribute to the biology of at-risk populations. Together, this knowledge may aid in designing and monitoring the impact of antiviral vaccines.

As a resource to explore antigen and HLA specific CTL responses, we used a computational approach previously developed to define neoantigen-specific CD8 T cell responses in patients with cancer(*6*). With this approach, we predicted all potential peptides that bind a globally representative set of MHC I alleles across the entire SARS-CoV-2 proteome. We have made this resource free and available to the public, and the resource will be updated through https://covid-19.parkerici.org/docs/data_sets/mhc_peptidome.html as additional data becomes available pertaining to the SARS-CoV-2 genome, proteome, and epitome.

Results

We deployed peptide binding predictions across 9-mer peptides derived from 1,075 protein sequences spanning 166 SARS-CoV-2 genotypes for 9,360 class I HLA alleles (2,987 HLA-A; 3,707 HLA-B; 2,666 HLA-C). Peptide binding predictions were filtered to those with a high-binding affinity (<500nM) to identify putative class I HLA antigens (**Figure 1A**). There were 6,748 unique combinations of peptides and HLA alleles (pMHCs) with a predicted binding affinity of less than 500nM (**Supplementary Table 3**), including 1,103 unique peptides and 1,022 HLA alleles (295 HLA-A, 614 HLA-B, and 113 HLA-C).

In order to better understand the class I antigenic profile of the SARS-CoV-2 proteome, we focused on pMHCs with a binding affinity of less than 500nM (n=6,748). Of the 1,103 unique peptides comprising these pMHCs, there were between 1 and 55 (median 3) alleles that could bind each peptide (**Figure 1B**), indicating that multiple HLA types could present the same antigens. Antigens were annotated by their corresponding protein (**Methods**) to identify which viral proteins may be responsible for class I recognition. There were between 12 and 684 (median 31) high-binding peptides observed in each of the 11 viral proteins (Orf1ab, Surface/Spike Glycoprotein, Orf3a, Envelope (Orf4; E), Membrane Glycoprotein, Orf6, Orf7a, Orf7b, Orf8, Nucleocapsid (Orf9), and Orf10; **Figure 1C**, **Supplementary Figure 1**). These predicted peptides were distributed throughout each protein (**Figure 1D**). Orf1ab was the largest protein and had the highest number of peptides (n=684), followed by the Surface/Spike Glycoprotein (n=109) and the Nucleocapsid (Orf9; n=60). When normalized by amino acid length, the largest proteins had the lowest density of peptides.

We were additionally interested in the diversity and number of the HLA alleles that could bind each predicted SARS-CoV-2 peptide. HLA alleles were annotated by their gene and structural superfamily (**Supplementary Table 4**)(*7*). In addition, we specifically highlighted the HLA-A*02 allele group, due to the prevalence of this allele

in the population and its level of study in model systems. There were more peptides with predicted high binding affinities for HLA-A (n=438 unique peptides) and -B (n=552) alleles, compared to -C (n=243) (**Supplementary Figure 3**). HLA alleles within superfamilies tended to bind similar numbers of peptides (left bar plot, **Figure 2, Supplementary Figure 3**) and the same sets of peptides (heatmap clusters, **Figure 2**). However, overall, a diverse set of HLA alleles are predicted to bind antigens spanning the viral proteome (**Figure 2**). Notably, our predicted peptides with high-affinity binding were all included in the homologous sequences between SARS-CoV and SARS-CoV-2.

Due to the partial homology between SARS-CoV-2 with SARS-CoV, we evaluated whether any experimentally validated epitopes from SARS-CoV were identified in our study (**Supplementary Table 5**). Using the Immune Epitope Database (IEDB) (*8*), we enumerated 30 HLA class I 9-mer epitopes that were validated by T cell assays (cytokine release assays, tetramer staining, and/or cytotoxicity assays) in clinical samples (including those derived from patients that had recovered from a prior SARS-CoV infection and normal donors) and humanized mouse models (e.g. HLA-A*0201 transgenic mice). Of these 30 epitopes, 10 were also seen in our filtered results (**Table 1**); there were two additional 9-mers that were contained within previously validated 10-mers. These included epitopes corresponding to the HLA-A*02 (n=20), HLA-A*11 (n=1), and HLA-B*40 (n=1) allele groups. Of note, of the 12 predicted epitopes with corresponding experimental validation, six were also predicted as high-binding antigens across other allele groups.

Grifoni, et al. recently used the homology between the SARS-CoV and SARS-CoV-2 proteomes and the existing annotated epitopes of SARS-CoV from the IEDB (*8*) to infer T cell epitopes derived from SARS-CoV-2(*9*). Our prediction identified 10 of the 25 dominant class 1 HLA 9-mer epitopes from that study. In addition, our study reports 1,088 additional predicted peptides that may serve as a resource for experimental validation.

Discussion

Our study was designed to evaluate the predicted MHC class I epitope landscape with respect to the SARS-CoV-2 viral proteome across a globally representative set of HLA alleles. We aimed to establish a resource for the scientific community, and have made the entirety of these data publicly available. This work expands upon recent studies that inferred the epitope landscape of SARS-CoV-2 based on its partial homology with SARS-CoV, for which there is an existing annotated epitope landscape(*9, 10*).

Our analysis was restricted to pMHC complexes with predicted binding affinities of less than 500nM. Subsequent analysis did not treat the predicted binding affinities as a continuous variable (i.e. predicted values of 5nM and 400nM were treated similarly in the remaining analysis). The pMHC binding affinities presented here are computationally defined. In the absence of experimental validation, we did not try to further delineate the association between HLA diversity and the predicted binding affinity. Furthermore, utilizing a threshold of 500nM may result in underestimating the number of alleles associated with the predicted antigenic peptides. Thus, this analysis could be expanded to consider pMHCs with binding affinities slightly greater than 500nM.

We also compared our filtered results to experimental efforts to validate viral epitopes of the homologous virus SARS-CoV over the last two decades. Our prediction successfully identified 12 of 32 previously validated HLA class I peptides across the SARS-CoV viral peptidome. Moreover, we predict that there are additional HLA types that may successfully bind and present these antigens.

This dataset and analysis have limitations. Our search was restricted to perfectly matched amino acid sequences, and there may be peptides associated with SARS-CoV-2 that have mismatches, compared to SARS-CoV. Our search was also limited to 9-mers, which represents most but not all reported HLA class I binding peptides. In addition, we did not further assess peptide sequences in pMHCs with affinities greater than 500nM and refined searches will be necessary to completely assess the homology between the SARS-CoV and SARS-

CoV-2 epitomes. Our data also does not account for either the quantity or timing of viral protein expression in a host cell, both of which can impact the immunogenicity of predicted epitopes(*11*).

CD4 T cell responses are also an important aspect of antiviral clearance and memory, but these were not included owing to the limitations in predicting human HLA class II epitopes(*12*). Lastly, the HLA binding prediction algorithms used in this study are trained based on antigen binding affinities available in the IEDB. Thus, our results are biased toward previously studied epitopes and HLA haplotypes and unlikely to be exhaustive, particularly in our list of filtered high-binding pMHC. As a result, we have deposited data for all predicted epitopes, even those with lower pMHC binding affinity.

With the ongoing SARS-CoV-2 pandemic, there are worldwide efforts to collect blood and tissue samples from persons who are asymptomatic, symptomatic with limited or serious complications, and convalescent. It is possible that patients who have the most severe forms of the disease, including ARDS, may have an increased functionality of CD8 T cells recognizing specific antigens presented by one of their HLAs, which could be specifically studied using the predicted epitopes described herein. In addition, preventive vaccines may require monitoring not only the induction of an antibody response, which is usually delayed, but also an earlier CD8 T cell response to specific antigens within the context of the patient's HLA subtypes. Given the global reach of the SARS-CoV-2 pandemic, this database of predicted HLA class I binding peptides may serve as a guide to identify and monitor phenotypic, functional and kinetic responses of putative SARS-CoV-2-specific $CD8^+$ T cells across patients with a broad range of HLA haplotypes internationally.

Acknowledgments
We are grateful to John Wherry (University of Pennsylvania, Philadelphia, PA) and Bonaventura Clotet, Julia Garcia Prado and Christian Brander (IrsiCaixa Foundation, Barcelona, Spain) for valuable feedback on the manuscript. The computational resources for this study were provided by the Parker Institute for Cancer Immunotherapy (PICI). KMC is supported by the UCLA Tumor Immunology Training Grant (NIH T32CA009120) and the Cancer Research Institute (CRI) Irvington Postdoctoral Fellowship Program. AK is supported by the UCLA CTSI KL2 Award (NCATS TR001882) and Sarcoma Alliance for Research Through Collaboration Career Enhancement Program. AR is supported by R35 CA197633 and The Ressler Family Fund, and is a member researcher at PICI.

Declaration of Interests
KMC is a shareholder in Geneoscopy LLC. DKW is a founder, equity holder and receives consulting fees from Immunai. AR has received honoraria from consulting with Amgen, Bristol-Myers Squibb, Chugai, Genentech, Merck, Novartis, Roche and Sanofi, is or has been a member of the scientific advisory board and holds stock in Advaxis, Apricity, Arcus Biosciences, Bioncotech Therapeutics, Compugen, CytomX, Five Prime, FLX-Bio, ImaginAb, Isoplexis, Kite-Gilead, Lutris Pharma, Merus, PACT Pharma, Rgenix and Tango Therapeutics, has received research funding from Agilent and from Bristol-Myers Squibb through Stand Up to Cancer (SU2C).

Figures

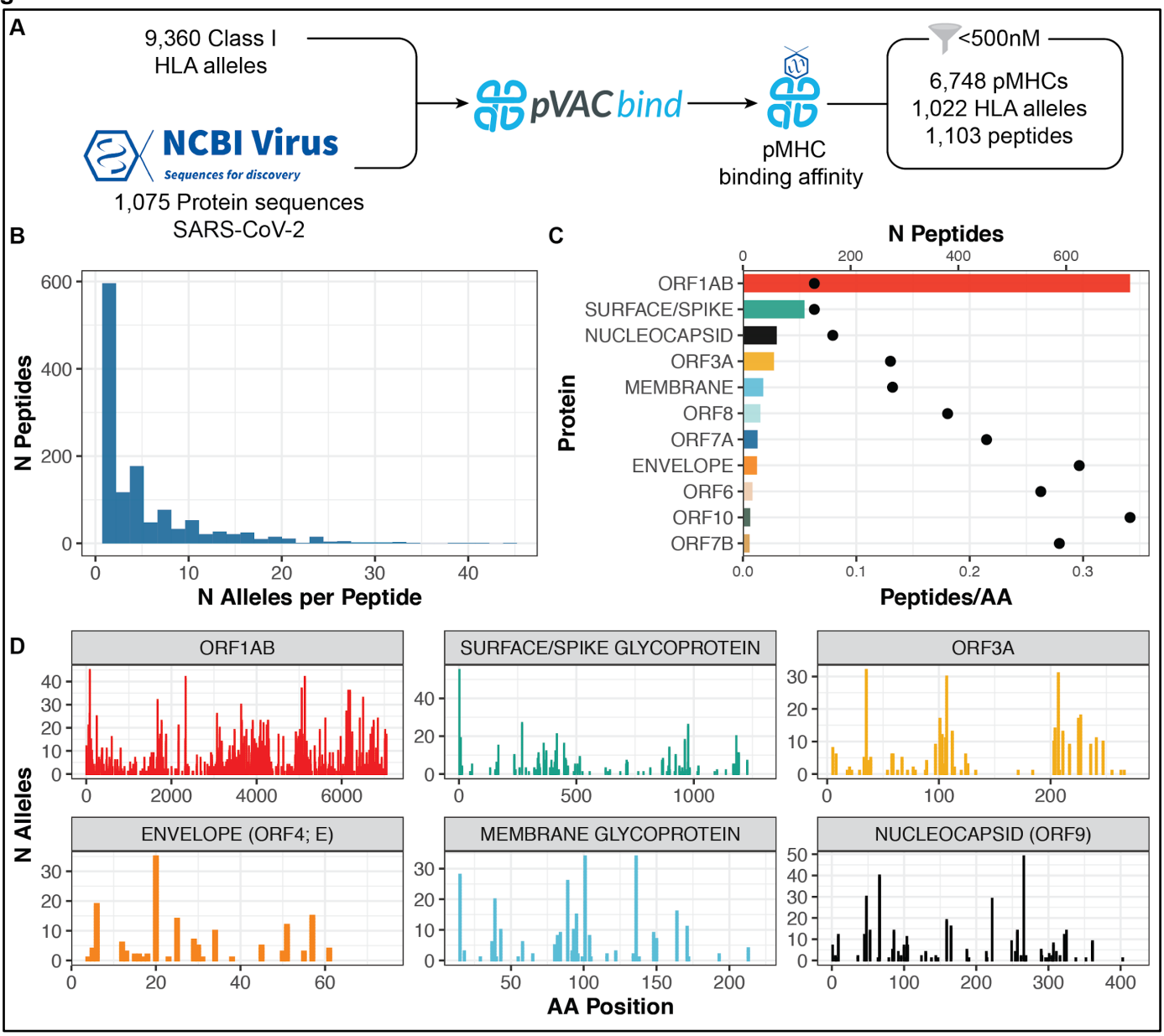

Figure 1. Peptide binding predictions for SARS-CoV-2

A. Overview of the analysis strategy. Using 9-mer peptides across the viral proteome, binding predictions were generated with nine different prediction algorithms (See Methods). B. Peptide-MHC complexes (pMHCs) were filtered to those with binding affinities less than 500nM. This histogram shows the distribution of the number of pMHCs (x-axis) corresponding to each peptide. C. Number of peptides (bars; top x-axis) derived from each protein (y-axis). The number of peptides was normalized to the length of the protein, and is indicated by a single point (points; bottom x-axis). D. Bar charts showing the number of alleles (y-axis) that have a high-binding peptide at the corresponding protein position (x-axis) for six viral proteins. Due to differences in amino acid sequences across annotated proteins, there are 43 peptides not shown. See **Supplementary Figure 1** for remaining proteins.

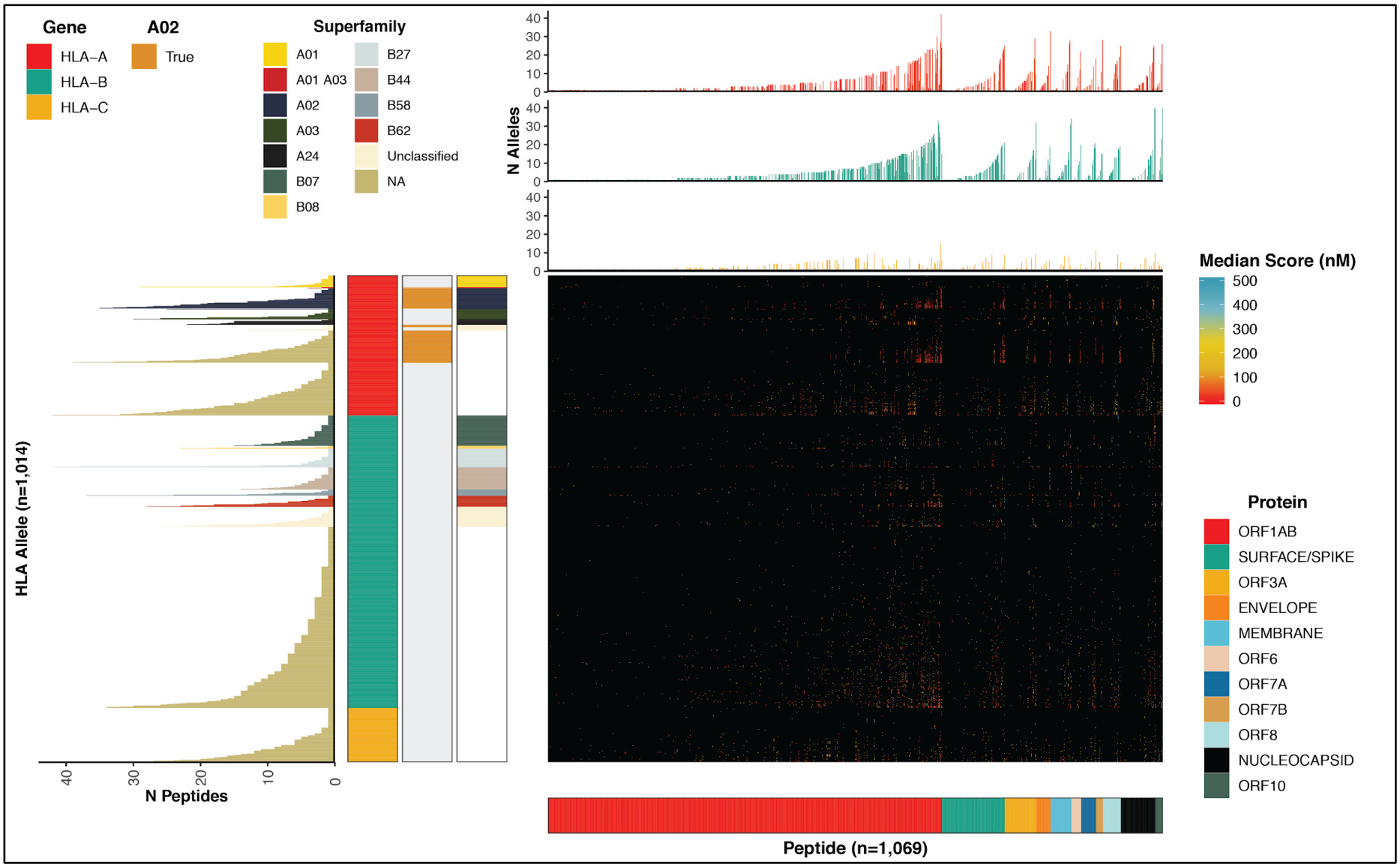
Gene
HLA–A
HLA–B
HLA–C
A02
True
Superfamily
A01
A01 A03
A02
A03
A24
B07
B08
B27
B44
B58
B62
Unclassified
NA
N Alleles
40
30
20
10
0
Median Score (nM)
500
400
300
200
100
0
HLA Allele (n=1,014)
N Peptides
Protein
ORF1AB
SURFACE/SPIKE
ORF3A
ENVELOPE
MEMBRANE
ORF6
ORF7A
ORF7B
ORF8
NUCLEOCAPSID
ORF10
Peptide (n=1,069)

Figure 2. SARS-CoV-2 Class I epitome

This figure displays all combinations (pMHCs; n=6,493) across peptides (x-axis) properly assigned to protein annotations (bottom color bar) with a binding affinity to HLA alleles (y-axis) of less than 500nM. The heatmap shows the binding affinity (fill color, nM), and pairs without high-binding epitopes are black. The bar charts show the number of peptides corresponding to each peptide (top) or the number of peptides corresponding to each allele (side left). Each allele is annotated (side color bars) by its gene (HLA-A, -B, -C), whether it is part of the A02 allele group, and superfamily (if available).

Tables

Table 1. Previously validated SARS-CoV peptides.

HLA Allele (Validated)	Peptide (Validated)	Reference	Predicted HLA Alleles
Nucleocapsid			
A*02:01	ALNTPKDHI	(*13*)	HLA-A*__02__:11
A*02:01	GMSRIGMEV	(*13, 14*)	HLA-A*__02__:03, A*__02__:50
A*11:01	KTFPPTEPK	(*15*)	HLA-A*03:10, 03:42, 03:44, 03:54, 03:73, 03:76, __11__:01, 30:31, 74:13
A*02:01	[L]LLLDRLNQL	(*13, 14, 16*)	HLA-A*__02__:02, __02__:03, __02__:11, __02__:13, __02__:132, __02__:141, __02__:150, __02__:16, __02__:173, __02__:181, __02__:19, __02__:196, __02__:205, __02__:214, __02__:228, __02__:238, __02__:25, __02__:262, __02__:70, __02__:71, __02__:73, __02__:85, __02__:95
			HLA-B*08:22, 08:38, 08:41, 08:56
			HLA-C*03:71
A*02:01	LQLPQGTTL	(*17*)	HLA-A*__02__:06, __02__:14
			HLA-B*15:01, 15:03, 15:103, 15:113, 15:127, 15:132, 15:179, 15:62, 15:69, 15:75, 15:98, 39:23, 39:49, 40:07, 40:12, 40:13, 40:21, 40:46, 48:15, 48:21
B*40:01	MEVTPSGTW[L]	(*18, 19*)	HLA-B*13:26, __40__:47, 44:101, 44:104, 44:17, 44:43, 44:48, 44:63, 44:71, 44:81, 44:91
Surface/Spike Glycoprotein			
A*02:01	FIAGLIAIV	(*20*)	HLA-A*__02__:03, __02__:131, __02__:150, __02__:170, __02__:179, __02__:187, __02__:196, __02__:205, __02__:214, __02__:228, __02__:238, __02__:248, __02__:257, __02__:50, __02__:69, __02__:71, __02__:85, __02__:95
A*02:01	KLPDDFMGCV	(*16, 21*)	HLA-A*__02__:50
A*02:01	NLNESLIDL	(*22*)	HLA-A*__02__:02, __02__:131, __02__:141, __02__:155, __02__:16, __02__:186, __02__:19, __02__:209, __02__:22, __02__:69, __02__:90
A*02:01	RLNEVAKNL	(*23*)	HLA-A*__02__:03, A*__02__:11, 02:128, 02:171, 02:196, 02:230, __02__:238, __02__:253, __02__:258, __02__:99
			HLA-B*27:20
A*02:01	VLNDILSRL	(*22, 24*)	HLA-A*__02__:03, __02__:11, __02__:13, __02__:132, __02__:148, __02__:151, __02__:171, __02__:186, __02__:19, __02__:196, __02__:209, __02__:22, __02__:230, __02__:238, __02__:253, __02__:258, __02__:52, __02__:70, __02__:71, __02__:73, __02__:85, __02__:99
			HLA-C*05:04, 05:23, 05:33
Orf1ab			
A*02:01	VLAWLYAAV	(*25*)	HLA-A*02:11, 02:148, 02:22, 02:230, 02:253, 02:258

Allele groups highlighted in **bold** under "Predicted HLA Alleles" indicates that the allele group matches the validated HLA Allele group (Column 1).

Methods

SARS-CoV-2 data acquisition and protein annotation

The protein fasta containing 1,075 protein sequences, spanning 166 genotypes for SARS-CoV-2 were accessed using the NCBI Virus resource (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/) on March 15, 2020 (**Supplementary Table 1**). Of the 1,075 accession identifiers for proteins, there were 144 unique amino acid sequences. These were further reduced to 22 groups by overlapping amino acid sequences (i.e. some accessions were regions within annotated proteins). Using the named annotation of these accession identifiers, these 22 groups of accession identifiers were annotated by the 11 named proteins in the SARS-CoV-2 proteome; 8 accession identifiers were labeled 'UNKNOWN' because they were not annotated with protein names.

Peptide-MHC binding prediction

MHC binding predictions were performed on 9-mer peptides derived from the protein fasta, across 9,360 haplotypes of class I HLA-A, -B, and -C, (**Supplementary Table 2**) using the pvacbind tool from pVACtools(*6*) and nine prediction algorithms (MHCflurry, MHCnuggetsI, NNalign, NetMHC, PickPocket, SMM, SMMPMBEC, SMMalign, NetMHC). The following parameters were used:

-b (BINDING_THRESHOLD)	500
-m, --top-score-metric	median
--net-chop-method	cterm
--netmhc-stab	TRUE

These jobs were executed on the Google Cloud Platform using Terra and the griffithlab/pvactools Docker container. The binding affinities for each prediction were summarized by their Median Score.

Assignment of peptides to proteins

Peptides were assigned to proteins based upon the protein annotation described above. Peptides per amino acid (Peptides/AA) were calculated by taking the total number of high-binding (<500nM) peptides associated with the protein accession divided by the length of the protein. These numbers were aggregated by the average Peptides/AA per protein.

Data Availability Statement

The results of this study are available in a public Google bucket through the following link: https://console.cloud.google.com/storage/browser/pici-covid19-data-resources. The folder "gs://pici-covid19-data-resources/mhci/peptide_predictions" in this bucket contains all of the Supplementary Tables corresponding to this document and the unfiltered peptide binding predictions (all_epitopes.tsv results from pvacbind). See either the **Supplementary Note** or the README in the Google bucket for further details.

REFERENCES

1. X. Yang *et al.*, Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *Lancet Respir Med*, (2020).
2. D. Wang *et al.*, Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *Jama*, (2020).
3. E. J. Wherry, R. Ahmed, Memory CD8 T-cell differentiation during viral infection. *Journal of virology* **78**, 5535-5545 (2004).
4. R. Channappanavar, S. Perlman, Pathogenic human coronavirus infections: causes and consequences of cytokine storm and immunopathology. *Semin Immunopathol* **39**, 529-539 (2017).
5. A. V. Hill, Immunogenetics and genomics. *Lancet* **357**, 2037-2041 (2001).
6. J. Hundal *et al.*, pVACtools: A Computational Toolkit to Identify and Visualize Cancer Neoantigens. *Cancer Immunol Res* **8**, 409-420 (2020).
7. J. Sidney, B. Peters, N. Frahm, C. Brander, A. Sette, HLA class I supertypes: a revised and updated classification. *BMC Immunol* **9**, 1 (2008).
8. R. Vita *et al.*, The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res* **47**, D339-d343 (2019).
9. A. Grifoni *et al.*, A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe*, (2020).
10. S. F. Ahmed, A. A. Quadeer, M. R. McKay, Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses* **12**, (2020).
11. N. P. Croft *et al.*, Most viral peptides displayed by class I MHC on infected cells are immunogenic. *Proc Natl Acad Sci U S A* **116**, 3112-3117 (2019).
12. M. Andreatta *et al.*, An automated benchmarking platform for MHC class II binding prediction methods. *Bioinformatics* **34**, 1522-1528 (2018).
13. Y. P. Tsao *et al.*, HLA-A*0201 T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus nucleocapsid and spike proteins. *Biochem Biophys Res Commun* **344**, 63-71 (2006).
14. S. Ohno *et al.*, Synthetic peptides coupled to the surface of liposomes effectively induce SARS coronavirus-specific cytotoxic T lymphocytes and viral clearance in HLA-A*0201 transgenic mice. *Vaccine* **27**, 3912-3920 (2009).
15. C. Sylvester-Hvid *et al.*, SARS CTL vaccine candidates; HLA supertype-, genome-wide scanning and biochemical validation. *Tissue Antigens* **63**, 395-400 (2004).
16. M. Zhou *et al.*, Screening and identification of severe acute respiratory syndrome-associated coronavirus-specific CTL epitopes. *J Immunol* **177**, 2138-2145 (2006).
17. Y. K. Cheung, S. C. Cheng, F. W. Sin, K. T. Chan, Y. Xie, 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 (2007).
18. L. Rivino *et al.*, Defining CD8+ T cell determinants during human viral infection in populations of Asian ethnicity. *J Immunol* **191**, 4010-4019 (2013).
19. C. X. Chang *et al.*, Conditional ligands for Asian HLA variants facilitate the definition of CD8+ T-cell responses in acute and chronic viral diseases. *Eur J Immunol* **43**, 1109-1120 (2013).
20. Y. D. Wang *et al.*, T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus spike protein elicit a specific T-cell immune response in patients who recover from SARS. *Journal of virology* **78**, 5612-5618 (2004).
21. J. Liu *et al.*, The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes. *J Infect Dis* **202**, 1171-1180 (2010).
22. Y. Lv, Z. Ruan, L. Wang, B. Ni, Y. Wu, Identification of a novel conserved HLA-A*0201-restricted epitope from the spike protein of SARS-CoV. *BMC Immunol* **10**, 61 (2009).
23. B. Wang *et al.*, Identification of an HLA-A*0201-restricted CD8+ T-cell epitope SSp-1 of SARS-CoV spike protein. *Blood* **104**, 200-206 (2004).
24. H. L. Oh *et al.*, Engineering T cells specific for a dominant severe acute respiratory syndrome coronavirus CD8 T cell epitope. *Journal of virology* **85**, 10464-10471 (2011).

25. S. Kohyama *et al.*, Efficient induction of cytotoxic T lymphocytes specific for severe acute respiratory syndrome (SARS)-associated coronavirus by immunization with surface-linked liposomal peptides derived from a non-structural polyprotein 1a. *Antiviral Res* **84**, 168-177 (2009).