

Supplementary Appendix

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

Authors:

Katie M. Campbell*¹, Gabriela Steiner², Daniel K. Wells², Antoni Ribas^{1,2,4,5}, Anusha Kalbasi*^{3,4,5}

Affiliations:

¹Department of Medicine, Division of Hematology-Oncology, University of California, Los Angeles (UCLA), Los Angeles, CA, 90095, USA

²Parker Institute for Cancer Immunotherapy, San Francisco, CA, 94129, USA

³Department of Radiation Oncology, UCLA, CA, 90095, USA

⁴Department Surgery, Division of Surgical Oncology, University of California, Los Angeles, Los Angeles, CA, USA

⁵Jonsson Comprehensive Cancer Center, Los Angeles, CA, USA

Supplementary Note

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). The details regarding these files are available in this document, as well as the README in the Google bucket.

Files can be accessed using either the command line gsutil tool (<https://cloud.google.com/storage/docs/gsutil>) or the Google Cloud Console (https://console.cloud.google.com/storage/browser/pici-covid19-data-resources/mhci/peptide_predictions/).

Files

All of the Supplementary Tables associated are available as tab-delimited ‘.tsv’ files:

- gs://pici-covid19-data-resources/mhci/peptide_predictions/SupplementaryTable1_ProteinSequences.tsv
- gs://pici-covid19-data-resources/mhci/peptide_predictions/SupplementaryTable2_HLATypes.tsv
- gs://pici-covid19-data-resources/mhci/peptide_predictions/SupplementaryTable3_FilteredAntigenBindingPredictions.tsv
- gs://pici-covid19-data-resources/mhci/peptide_predictions/SupplementaryTable4_HLASuperfamilies.tsv
- gs://pici-covid19-data-resources/mhci/peptide_predictions/SupplementaryTable5_PreviouslyIdentifiedSARS-CoVEpitopes.tsv
- gs://pici-covid19-data-resources/mhci/peptide_predictions/SupplementaryTable6_AlleleFileMapping.tsv

Unfiltered file names

Each unfiltered file (unfiltered_results/*.all_epitopes.tsv) contains the unfiltered results from pvacbind, executed across 10 HLA alleles (there are 9,360 of these files). Each file name indicates the list of 10 alleles included in the associated TSV (With asterisks [*] and colons [:] replaced by underscores [_]) and an underscore [_] between each allele.

For example, the file:

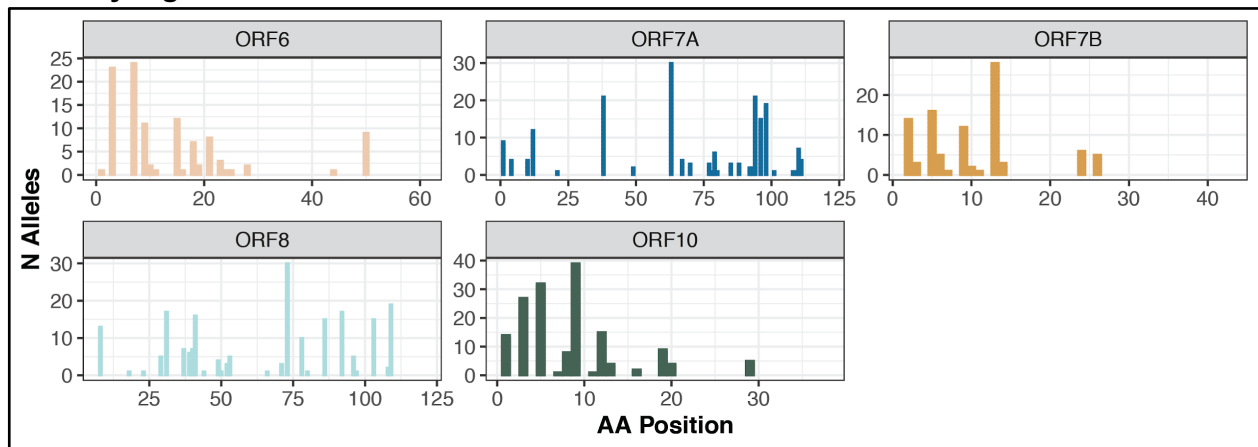
HLA-A_01_01_HLA-A_01_02_HLA-A_01_03_HLA-A_01_04_HLA-A_01_06_HLA-A_01_07_HLA-A_01_08_HLA-A_01_09_HLA-A_01_10_HLA-A_01_100.all_epitopes.tsv

Contains the 10 alleles:

HLA-A*01:01, HLA-A*01:02, HLA-A*01:03, HLA-A*01:04, HLA-A*01:06, HLA-A*01:07, HLA-A*01:08, HLA-A*01:09, HLA-A*01:10, and HLA-A*01:100

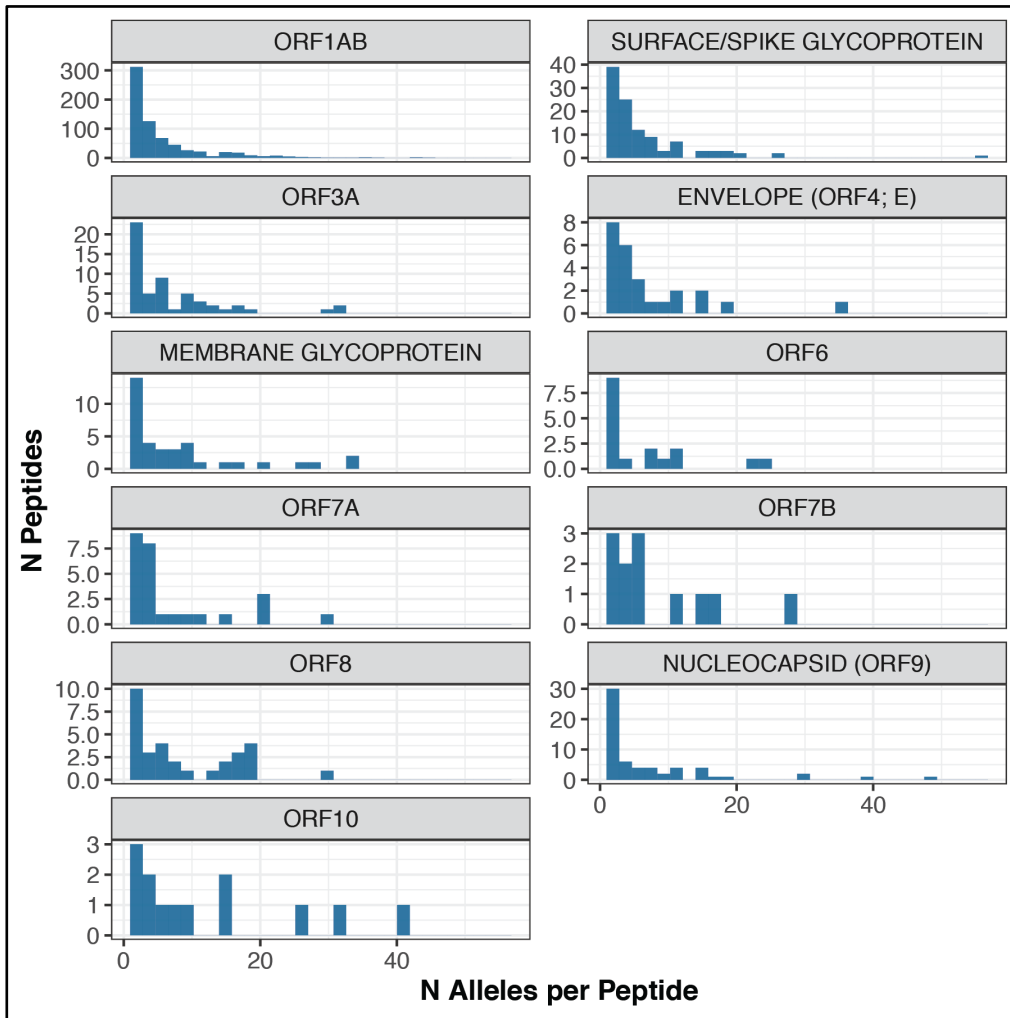
Supplementary Table 6 (SupplementaryTable6_AlleleFileMapping.tsv) directly maps HLA alleles to the corresponding file name.

Supplementary Figures



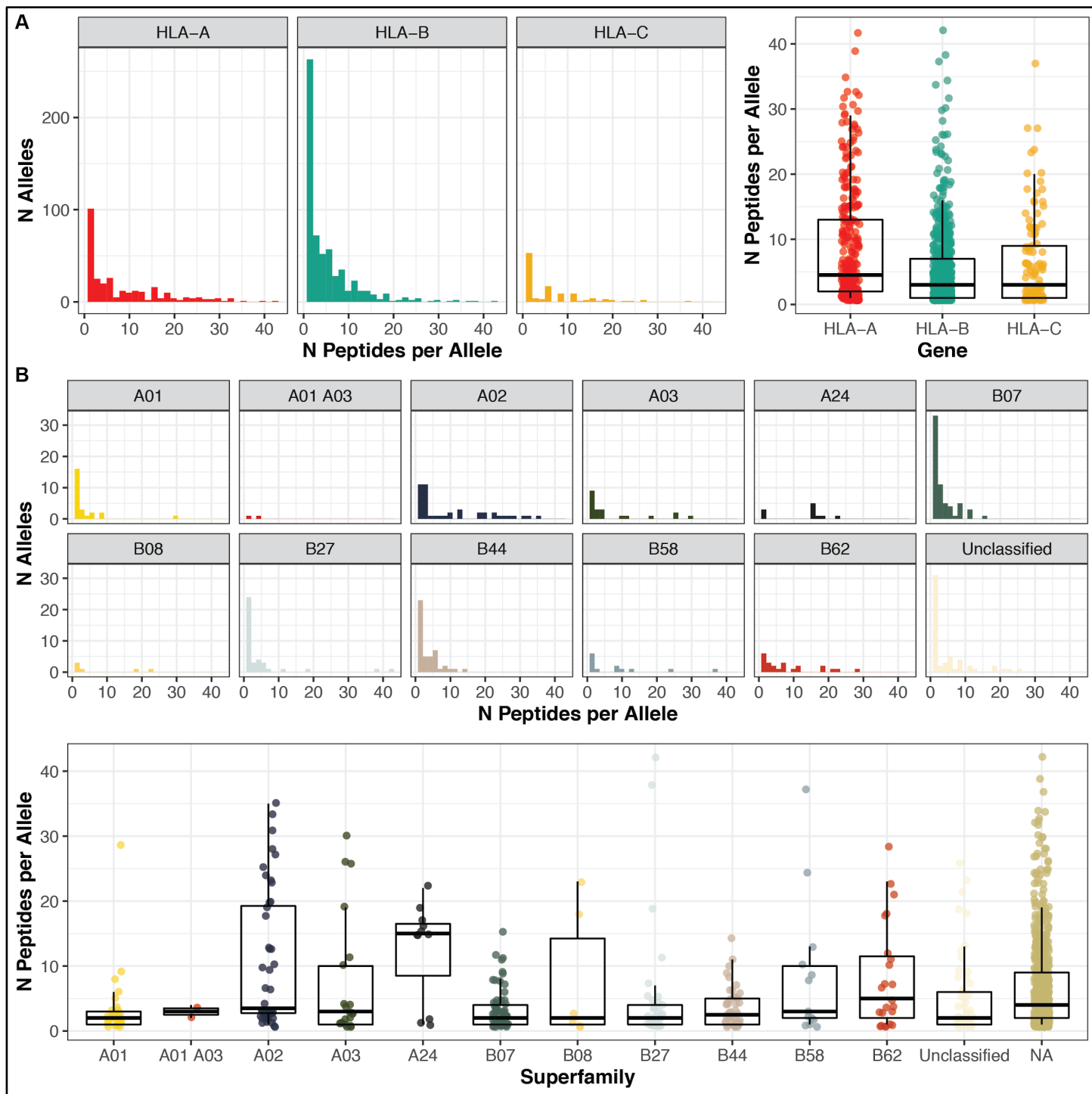
Supplementary Figure 1. Number of alleles with high-binding peptides across viral proteins.

Bar charts showing the number of alleles (y-axis) that have a high-binding peptide at the corresponding protein position (x-axis) for five viral proteins. Corresponds to Figure 1D. Due to differences in amino acid sequences across annotated proteins, there are 5 peptides not shown.



Supplementary Figure 2. Number of pMHCs per peptide across proteins

The histogram in Figure 1B was broken up to display the distribution of pMHCs per peptide across each protein. Of note, there are no peptides corresponding to more than one protein annotation, and peptides corresponding to the annotated “Unknown” proteins (see **Methods**) are not included (n=34).



Supplementary Figure 3. Distribution of the number of pMHCs per allele

A. These histograms (left) show the distributions of the number of pMHCs (x-axis) corresponding to each HLA allele, across the three HLA genes. The corresponding boxplots (right) show the distribution of the number of pMHCs (y-axis) per allele. B. Histograms (top) and boxplots (bottom), describing the number of peptides

Supplementary Tables

Supplementary Table 1. Protein sequences

This table summarizes the 1,075 annotated SARS-CoV-2 proteins used in this study. The 'Accession' column in this table maps to the 'Mutation' Column in the pvacbind results (**Supplementary Table 3**).

Supplementary Table 2. HLA types

This table includes all 9,360 HLA types used as inputs to pvacbind evaluated in this study.

Supplementary Table 3. Filtered antigen binding predictions

The filtered.tsv results from pvacbind were merged across all HLA alleles.

Supplementary Table 4. HLA superfamilies

This table includes the annotation of the HLA superfamilies from Sidney, et al.

Supplementary Table 5. Previously identified SARS-CoV epitopes

These epitopes from SARS-CoV were assembled from manual literature review.

Supplementary Table 6. Allele File Mapping

This file contains the mapping between HLA alleles and their corresponding file that contains the unfiltered antigen binding predictions (See **Supplementary Note** for further details).