frontiers

HLA-C Peptide Repertoires as Predictors of Clinical Response During Early SARS-CoV-2 Infection

Michael D. Olp [1,*], **Vincent A. Laufer** [1], **Andrew L. Velesano** [1,2], **Andrea Zimmerman** [1], **Kenneth J. Woodside** [3,4,5], **Yee Lu** [6], **Adam S. Lauring** [2], and **Matthew F. Cusick** [1,*]

[1] *Department of Pathology, University of Michigan, Ann Arbor, MI, USA*
[2] *Division of Infectious Diseases, Department of Internal Medicine and Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI, USA*
[3] *Sharing Hope of South Carolina, Charleston, SC, USA*
[4] *Gift of Life Michigan, Ann Arbor, MI, USA*
[5] *Academia Invisus LLC, Ann Arbor, MI, USA*
[6] *Division of Nephrology, Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA*

Correspondence*:
2800 Plymouth Rd building 35, Ann Arbor, MI 48109
olpmi@med.umich.edu, mfcusick@med.umich.edu

ABSTRACT

The Human Leukocyte Antigen (HLA) system plays a pivotal role in the immune response to viral infections, mediating the presentation of viral peptides to T cells and influencing both the strength and specificity of the host immune response. Variations in HLA genotypes across individuals lead to differences in susceptibility to viral infection and severity of illness. This study uses observations from the early phase of the COVID-19 pandemic to explore how specific HLA class I molecules affect clinical responses to SARS-CoV-2 infection. By analyzing paired high-resolution HLA types and viral genomic sequences from 60 patients, we assess the relationship between predicted HLA class I peptide binding repertoires and infection severity. This approach leverages functional convergence across HLA-C alleles to identify relationships that may otherwise be inaccessible due to allelic diversity and limitations in sample size. Surprisingly, our findings show that severely symptomatic infection in this cohort is associated with disproportionately abundant recognition of SARS-CoV-2 structural and non-structural protein epitopes by patient HLA-C molecules. Furthermore, we explore how structural characteristics of HLA-C molecules relate to differential promiscuity in SARS-CoV-2 peptide recognition. The findings highlight immunologic mechanisms linking HLA-C molecules with the human response to viral pathogens that warrant further investigation.

Keywords: HLA, SARS-CoV-2, antigenicity, bioinformatics, computational biology

1 INTRODUCTION

Severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) produces a considerable range of clinical presentations from mild common cold-like symptoms to severe pneumonia, respiratory distress syndrome (ARDS), and multisystem organ failure (MOF) [1, 2]. While patient demographic, medical, and behavioral risk factors certainly influence outcomes [3], dysregulated activation of immune responses is also implicated in severe clinical progression early in infection [4, 5].

Human leukocyte antigen (HLA) genes are highly polymorphic structures with central roles in shaping immune responses against pathogens. The interactions between HLA I and II molecules and T cell receptors (TCR) activate lymphocytes and develop the adaptive immune effector milieu. HLA class I antigens interact with cytotoxic T cells (CD8+) and kill virally infected targets. Different HLA genes determine specific repertoires of peptides included in the major histocompatibility complex (MHC) and influence subsequent TCR recognition, leading to antigen-specific immune responses. Therefore, HLA heterogeneity is a crucial factor underlying differential susceptibility to viral infection across human populations.

Distinct HLA alleles and haplotypes impact severity and progression of viral infection due to Human Immunodeficiency virus (HIV) [6, 7, 8, 9], hepatitis B [10, 11], hepatitis C [12, 13, 14], and influenza [15, 16, 17]. Certain HLA alleles have also been associated with poor clinical response to infection by SARS-CoV [18, 19], a closely related predecessor to SARS-CoV-2 with approximately 80% genetic sequence identity [20, 21]. However, geographic restriction of SARS-CoV transmission allowed for examination of only limited pools of HLA alleles. In contrast, the pandemic nature of SARS-CoV-2 has prompted numerous studies analyzing broader ranges of HLA alleles. These efforts have identified multiple associations between patient HLA alleles and disease outcome [22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32]. However, identified associations are not entirely consistent [32, 33]. Discrepancies in allele-level associations between studies are at least partially due to geographic variation in HLA genotype frequency. In addition, geographic and temporal SARS-CoV-2 genetic variation further complicates comparison across studies.

SARS-CoV-2 mutational analysis has been a crucial component of vaccine design, and HLA has emerged as an important target [34, 35, 36, 37, 38]. *In silico* peptide binding analyses have been instrumental in identifying antigenic SARS-CoV-2 peptides given the particularly large viral peptidome as well as the combinatorial nature of binding groove sequence variation both within and across HLA loci. Machine learning tools like NetMHCpan and NetMHCIIpan utilize available experimental data to allow robust and efficient prediction of MHC-peptide binding [39]. While high-affinity MHC-peptide binding does not, by itself, equate with immunogenicity, in silico techniques have successfully identified SARS-CoV-2 epitopes capable of eliciting T cell response, many of which are HLA specific [40, 41, 42]. Importantly, even sequence variation within SARS-CoV-2 lineages alters viral peptide presentation on HLA molecules [43].

Here, we employ *in silico* peptide binding prediction to investigate the role of individual HLA classification in clinical response to SARS-CoV-2 infection. Individual levels of viral peptide binding to class I and II HLA molecules were predicted using paired high-resolution HLA typing and SARS-CoV-2 genomic sequencing for 60 patients presenting to Michigan Medicine. We then test the hypothesis that variation in viral antigenicity is associated with differential clinical response to infection using the Sequential Organ Failure Assessment (SOFA) scoring system as an indicator of disease severity. This is the first study comparing HLA classification with clinical outcome that controls for SARS-CoV-2 genomic sequence variation at an individual level. Using a multiple linear regression-based approach that accounts for HLA

class I genotype and demographic factors, we identify a significant association between disproportionate HLA-C-mediated peptide binding and severely symptomatic infection. Furthermore, molecular modeling and unsupervised machine learning techniques provide structural and functional insight into differential peptide binding by individual HLA-C molecules that correlate with disease severity. In addition, we investigate how viral sequence variation over time affects disproportionate HLA-C-mediated binding of both structural and non-structural SARS-CoV-2 epitopes. This approach leverages functional convergence across HLA alleles to identify relationships that may be obscured by high degrees of allelic diversity and limited sample size. Ultimately, this research aims to enhance understanding of how individual HLA classification impacts interaction with novel antigens to influence clinical outcomes, both in the context of continuing SARS-CoV-2 transmission and novel viral pathogens.

2 RESULTS

Extent of patient HLA-C peptide binding repertoire is a risk factor for severe SARS-CoV-2 infection.

In order to investigate relationships between HLA class I genotype and illness severity following SARS-CoV-2 infection, high-resolution HLA types and viral genomic sequences were obtained for 60 patients with positive polymerase chain reaction (PCR) tests between March 24, 2020, and May 5, 2020. This study design allows direct comparison between each patient's HLA class I alleles and the specific peptidome encoded by their SARS-CoV-2 strain. Infection severity is quantified as the maximum sequential organ failure assessment (SOFA) score documented within 10 days of SARS-CoV-2 diagnosis. Here, SOFA score shows a significant positive correlation with laboratory values previously associated with severe disease, including creatinine, white blood cell count (WBC), and C-reactive protein (CRP) [44], as well as indications of poor outcome such as intensive care unit (ICU) admission and need for intubation (Figure S1).

To allow associations between the extent of viral peptide recognition by patient-encoded MHC class I molecules and clinical outcome following SARS-CoV-2 infection, binding predictions were performed using NetMHCpan software. Binding affinity (BA) and mass spectrometry-based eluted ligand (EL) scores were calculated for all MHCI:peptide pairs resulting from each patient's six HLA class I alleles and the corresponding SARS-CoV-2 peptidome, including sequences eight to eleven amino acids in length (Figure 1A-C, Table S2-4). In addition, MHCI:peptide pairs were queried in the Immune Epitope Database (IEDB) to identify experimentally validated binding interactions and pairs showing no binding. The resulting 2,431 experimentally interrogated MHCI:peptide pairs were used as a training set for logistic regression to assign binding probabilities for all NetMHCpan predictions (Figure 1E). The overall logistic regression model is closely related to the binding affinity score (Figure 1D).

To test the hypothesis that response to infection is related to the overall size of patient HLA class I peptide binding repertoires, the sum of predicted interactions for each patient's two HLA-A, -B, and -C alleles were used to predict SOFA score through a Bayesian generalized linear multivariate regression framework (Figure 1F, S2). To avoid spurious results due to ambiguity regarding biologically relevant MHC:peptide binding affinities and using a single arbitrary cutoff to define binding, numbers of peptide interactions were calculated over a range of probability cutoffs (0.5 – 0.9) (Figure D-F). The resulting distributions were then accounted for as independent variable uncertainty using Bayesian techniques [45, 46]. In addition, age, body mass index (BMI), and sex were included as regressors to control for potential influences of these demographic factors over the clinical course. Of the variables examined, increased size of HLA-C

peptide binding repertoire predicts poor response to infection (95% CI = 0.12 - 2.60, Figure 1F). Although not reaching statistical significance, increased SOFA score was predicted by decreased sizes of HLA-A (95% CI = -0.36 - 0.58) and -B (95% CI = -0.26 - 0.60) peptide binding repertoires. Together, these results predict disproportionate peptide binding by HLA-C molecules relative to HLA-A and -B is a risk factor for severe SARS-CoV-2 infection.

Disproportionate HLA-C binding of structural and nonstructural protein sequences contributes to poor prognosis.

Early efforts aimed at epitope mapping and vaccine design primarily focused on SARS-CoV-2 structural proteins, particularly Spike. However, peptides derived from nonstructural and accessory proteins can also elicit an immune response to SARS-CoV-2 infection [47, 48]. As a result, we examine the HLA class I binding of sequences derived from all twenty-seven SARS-CoV-2-encoded proteins (Figure 2A). To identify protein sequences leading to unbalanced HLA-C peptide binding relative to HLA-A and -B, predicted repertoire sizes for individual SARS-CoV-2 proteins were compared with the SOFA score for each patient. HLA class I peptide repertoire sizes for each SARS-CoV-2 protein were calculated across binding probability thresholds 0.5 to 0.9. Significant relationships with SOFA score were then identified as those with Pearson correlation coefficients with a magnitude greater than one standard deviation above 0.25, corresponding to $p<0.05$ for sixty samples. In this analysis, quantities of predicted HLA-C peptide-binding interactions show a significant positive correlation with SOFA score for twenty of the twenty-seven SARS-CoV-2 proteins (Figure 2B). Conversely, peptide repertoire sizes for HLA-A and -B are predominantly lower for patients with more severe infection, although these correlations for individual protein sequences do not reach statistical significance. These results indicate that the disproportionate binding of HLA-C molecules toward peptides derived from structural, nonstructural, and accessory SARS-CoV-2 protein sequences contributes to an elevated risk of severe infection.

Peptide sequence characteristics associated with differential HLA-C binding in mild versus severe SARS-CoV-2 infection were also examined. Pearson correlations were calculated between HLA-C binding probability and SOFA score for all peptides, and the top 1% of positively and negatively correlated sequences were compared. Peptide sequences predicted to bind to HLA-C alleles in patients with severe SARS-CoV-2 infection show a preference for amino-acid restriction at the second from *N*-terminal (P2) and *C*-terminal anchor positions (Figure 2C-F). The *C*-terminal position is enriched for hydrophobic residues, which is consistent with HLA-C binding specificity overall [49]. P2 also shows enrichment for hydroxyl-containing amino acids (Ser, Thr, and Tyr) or the smaller Ala in peptides selectively binding HLA-C alleles encoded by patients with severe infection (Figure 2C-F). While all SARS-CoV-2 sequences identified in this study are of either the A or B.1 lineage (Table S1), amino acid differences exist at sixty-two missense mutation sites. Five of these mutations significantly alter the distributions of binding probability scores across patient HLA-C alleles (Figure 2G). Mutations of non-structural protein 3 (NSP3) and nucleocapsid protein (N) affect the P2 anchor, with Ser favored for HLA-C binding. Mutations involving the *C*-terminal anchor of peptides derived from NSP13 and the NS3 accessory protein highlight preference for Val at this position. In addition, the Spike $Asp^{614}Gly$ mutation that has been associated with enhanced viral fitness is predicted to result in less HLA-C binding in this patient cohort (Figure 2G). These results implicate structural, nonstructural, and accessory protein sequences in differential HLA-C peptide binding and prognostic impact. Furthermore, missense mutations within single SARS-CoV-2 lineages can alter recognition by HLA-C molecules due to amino acid changes at the P2 and *C*-terminal anchor positions.

Infection severity correlates positively with overlap between HLA-A and -C peptide binding repertoires.

Unbalanced viral peptide binding by HLA-C relative to HLA-A and -B during infection may be exacerbated by competition for shared ligands. Indeed, distributions of peptide binding across SARS-CoV-2 proteins are similar for each of the three HLA class I loci (Figure 2A). To test this hypothesis, the intersection between sets of SARS-CoV-2 peptides recognized by patient HLA-A, -B, and -C alleles was determined over a range of prediction scoring thresholds (Figure 3A-B). For all patient alleles combined, percentages of HLA-A and -B peptide repertoires that are also predicted to be bound by HLA-C alleles range from 10 - 29% and 16 – 36%, respectively. Generally, the highest degree of overlap is seen for interactions with lower predicted affinity (Figure 1D-E, 3A-B). The intersection between HLA-A and -C peptide repertoires significantly correlates with SOFA score (Figure 3C, $p < 0.004$). However, no similar relationship between HLA-B and -C peptide repertoires is identified (Figure 3D). These results suggest that competition between HLA-A and -C for shared SARS-CoV-2 epitopes contributes to inefficient HLA class I responses that predispose patients toward severe infection.

Prognosis is not related to KIR ligand group among patient alleles examined.

HLA-C molecules are the primary ligands of Killer cell immunoglobulin-like receptors (KIRs), thus playing an important role in natural killer (NK) cell-mediated responses to infection [50, 51]. A balanced dimorphism at amino acids 77 and 80 divides HLA-C molecules into C1 and C2 KIR ligand groups [52] (Figure 4A-B). C1 allotypes have previously been predicted to bind larger quantities of SARS-CoV-2 peptides [53], and several have been individually identified as risk factors for poor prognosis [54, 55, 55, 28, 22, 56, 57, 58, 25, 54]. In the present study, C1 allotypes are predicted to bind larger numbers of SARS-CoV-2 peptides than C2 allotypes ($p < 0.002$, Figure 4C). However, the KIR ligand group does not predict differential prognosis following infection (Figure 4D, $p > 0.98$). These results argue against KIR ligand specificity as a primary mechanism of prognostic variation in this patient cohort.

The amino acid sequence of the HLA-C peptide binding groove predicts patient prognosis following SARS-CoV-2 infection.

The balanced dimorphism defining KIR ligand groups has only a limited impact on the HLA-C peptide binding surface (Figure 4A). Alleles were clustered based on their complete peptide binding site amino acid sequences to examine the relationship between HLA-C molecular structure and differences in patient prognosis (Figure 5A). Distinct electrostatic patterns between the resulting groups delineate Cluster 1 and 2 peptide binding grooves as predominantly negatively and positively charged, respectively, while Cluster 3 binding sites are relatively hydrophobic (Figure 5B-D). Comparing patient SOFA scores with each of their two HLA-C alleles reveals differential prognosis across sequence-based clusters (Figure 5E, ANOVA $p = 0.05$). This analysis, too, demonstrates severe SARS-CoV-2 infection is associated with larger HLA-C peptide binding repertoires (Figure 5F, ANOVA $p < 1e\text{-}11$). Cluster 3 alleles are predicted to have the largest peptide binding repertoire and are associated with the greatest prognostic risk. Lower peptide binding selectivity in Cluster 3 may be due to the ability of relatively hydrophobic binding pockets to accommodate peptides with uncharged residues at P2 (Figure 2C and 5G). In addition, Clusters 1 and 3, the lowest and highest risk sequence-based groups, respectively, show allele dosage effect approaching statistical significance ($p = 0.15$ and 0.12, respectively; Figure S3). These results support a mechanism by which HLA-C molecule amino-acid composition and differential peptide binding specificity impact patient response to SARS-CoV-2 infection.

DISCUSSION

The HLA system is a crucial determinant of infection susceptibility making it an attractive target for investigations aimed at understanding genetic causes of differential responses to pathogens across human populations. HLA class I molecules are crucial for the early elimination of virus-infected cells by CD8+ T-cells and NK cells while maintaining self-tolerance to avoid excessive tissue damage. As a result, HLA class I genotype has received significant attention with respect to aberrant immune responses during early SARS-CoV-2 infection. Here, we compared HLA class I genotypes and strain-specific SARS-CoV-2 peptidomes for 60 patients seen at Michigan Medicine early in the COVID-19 pandemic to investigate relationships between predicted peptide recognition and infection severity. This analysis shows patients with larger quantities of HLA-C:viral peptide interactions were significantly more likely to experience severe symptoms (Figure 1F).

HLA-C molecules are expressed in lower quantities at the cell surface [59, 60, 59, 61, 62], compared with HLA-A and -B and, therefore, often receive less attention in studies of responses to viral infection. Nevertheless, HLA-C-restricted viral peptides have been implicated during disease progression in HIV [63, 64, 65, 66, 67, 68], and hepatitis C infections [69]. While the role of HLA-C in viral respiratory illness remains poorly understood, specific HLA-C alleles have been associated with differential risk of infection and/or prognosis in influenza [70, 71, 72], Chikungunya [73], Dengue [74], Lassa [75], and Ebola [76, 77, 78] viruses. Increasing evidence also links genetic and epigenetic polymorphism at the HLA-C locus with differential clinical outcomes in SARS-CoV-2 infection. Epigenome-wide association studies implicate differential CpG methylation of the HLA-C locus in conferring risk for developing severe infection with respiratory failure [79]. In addition, multiple studies have identified associations between specific HLA-C alleles and prognosis following SARS-CoV-2 infection [80, 81, 82, 80, 81, 82, 29, 30, 31], albeit with at times conflicting results [32, 33]. Potential causes for this variability include differences in the populations and/or SARS-CoV-2 strains examined and difficulty obtaining large sample sizes. Our study addresses statistical limitations associated with these issues by assessing predicted viral peptide binding as a continuous variable rather than discrete HLA alleles. In addition, predicting peptide binding with respect to each patient's sequenced SARS-CoV-2 strain controls for differences in viral peptidomes and would allow for unbiased comparison with other cohorts.

Previous studies also suggest an association between poor clinical outcome and HLA-A and -B allotypes with poor capacity for viral peptide binding [23, 25]. Furthermore, low levels of viral peptide recognition by HLA class I molecules may be particularly important for increasing susceptibility to severe SARS-CoV-2 infection in younger patients [83]. This association is relevant to our cohort as younger patients unexpectedly tended to have higher SOFA scores (Figure 1A). Poor prognosis in our patient cohort also tended to be associated with smaller predicted HLA-A, and -B peptide binding repertoires (Figure 1F). We hypothesized that HLA-C molecules may exacerbate this problem by competing with other HLA class I molecules for peptide binding, potentially causing lower antigenicity due to lower cell surface presentation by HLA-C versus -A and -B molecules. Indeed, we show higher degrees of overlap between HLA-A and -C molecule peptide binding specificity for a given patient, which positively correlates with infection severity (Figure 3C).

Prior studies also investigated the role of HLA-C/KIR relationships in differential response to SARS-CoV-2 infection [80, 81, 82]. In agreement with previous analyses, we found that C1 allotypes are predicted to bind larger quantities of SARS-CoV-2 peptides than C2 [53] (Figure 4C). However, we do not identify significant correlation between C1/C2 allotype and infection severity (Figure 4D), perhaps consistent with contradicting prognostic findings reported for C1 versus C2 alleles [54, 55, 55, 28, 22, 56, 57, 58, 25, 54,

28, 84, 85]. Instead, we identified prognostically significant HLA-C allele groupings by assessing the overall electrostatic compositions of peptide binding site surfaces (Figure 5), which may inform on peptide binding specificity and downstream T-cell recognition.

Inherent limitations of this study include a relatively small sample size and the uncertain functional significance of peptide-binding prediction tools. MHC:peptide binding does not guarantee efficient antigen presentation or subsequent T-cell response. Nevertheless, computational tools for predicting antigen recognition show increasing utility in guiding vaccine design. For instance, predictive analyses broadened searches for SARS-CoV-2 vaccine targets beyond the Spike protein, which was perhaps disproportionately emphasized in the early stages of the COVID-19 pandemic. Our results further support the potential for targeting antigens encoded by structural and non-structural viral proteins for continued vaccine development (Figure 2).

The conclusions drawn from this research offer valuable perspectives for the fields of personalized medicine, vaccine development, and immunology. By demonstrating the predictive capacity of HLA-C peptide binding for assessing the severity of SARS-CoV-2 infections, this study suggests pathways for creating clinical tools to identify individuals at higher risk. Such tools can potentially facilitate tailored treatment strategies and more focused patient monitoring. Additionally, the findings advocate for including a wider array of viral epitopes in vaccine formulations, which may lead to effective vaccines across the diverse spectrum of HLA-C genotypes. The exploration of interactions among HLA class I molecules and their influence on the immune response opens avenues for further investigation. This could include the development of advanced computational models to refine our predictions of HLA-peptide binding and its role in antigen presentation. Progress in these areas promises to deepen our grasp of how pathogens evade immune detection and could inform the development of more sophisticated immunotherapies and vaccines, potentially benefiting the management of various infectious diseases.

MATERIALS AND METHODS

High-resolution human leukocyte antigen-typing.

Blood collection and subsequent DNA isolation were performed by the University of Michigan COVID-19 Biorepository. After extraction, DNA concentration and quality were measured with the Qiagen Qiexpert spectrophotometer. Next-generation sequencing (NGS) human leukocyte antigen (HLA)-typing for HLA-A, -B, -C was done using NGSgo®-Amp v2 (GenDx, the Netherlands), according to the manufacturer's instructions. After long-range PCR amplification of each HLA gene, DNA fragments were selected, amplified, cleaned, and sequenced on a MiSeq using MiSeq Reagent Kits v2 (300 cycles) (Illumina, San Diego, CA). Samples were analyzed using NGSengine v2.25.0 software. This study was approved by the University of Michigan Institutional Review Board (HUM00183348), and samples were obtained from the University of Michigan Medical School Central Biorepository.

Peptide binding predictions.

Translated SARS-CoV-2 protein sequences corresponding to viral strains sequenced in this study were obtained from the GISAID database (*gisaid.org/EPI_SET_220330me*) [86] and are listed in Table S1. FASTA formatted sequences were parsed using the Bio.Seq module implemented in biopython (v. 1.75). For each viral strain, protein sequences were partitioned into libraries of all possible overlapping 8-11 amino acid peptides, respectively, using a sliding window method. Binding predictions were then carried out for each patient MHC protein toward the corresponding peptide library using local installations of NetMHCpan

(v. 4.1) and NetMHCIIpan (v. 4.0) artificial neural networks [39] with default input parameters. Binding affinity (BA) and mass-spectrometry eluted ligand (EL) data were used to select predicted interactions for further analysis. Peptide binding data was processed using the pandas (v. 1.4.3) and NumPy (v. 1.21.3) Python libraries. Predicted BA and EL results for 2,431 experimentally examined MHCI:peptide pairs available in the IEDB were used to generate a logistic regression model, and the resulting probabilities were evaluated by ROC analysis using the LogisticRegression and roc_curve functions available in scikit-learn (v. 1.3.2) Python library.

Statistical modeling of infection severity.

Bayesian generalized linear multivariate regression was performed using the brms implementation of Stan in R [45, 46] (v. 2.18.0) using the following equation:

$$y_i = \beta_0 + \beta_1 \mathrm{A}_i^* + \beta_2 \mathrm{B}_i^* + \beta_3 \mathrm{C}_i^* + \beta_4 \mathrm{Sex}_i + \beta_5 \mathrm{Age}_i + \beta_6 \mathrm{BMI}_i + \epsilon_i \quad (1)$$

where y_i refers to the SOFA score corresponding to the i^{th} patient, β_0, β_1,..., β_6 are the estimated coefficients, and ϵ_i is the error term. A_i^*, B_i^*, and C_i^* are the latent true values of the HLA-A, -B, and -C peptide repertoire sizes calculated across binding probability thresholds 0.5 to 0.9, accounting for error using the me(mean, standard deviation) brm function call. Sex, age, and BMI are additional predictors without associated measurement errors. All predictors were z-score normalized prior to multivariate regression. The relationship between SOFA score and the fractional intersection of patient HLA-A (or -B) and HLA-C peptide binding repertoires was similarly modeled using the equation:

$$y_i = \beta_0 + \beta_1 \mathrm{X}_i^* + \epsilon_i \quad (2)$$

with y_i, β_0, β_1, and ϵ_i defined as in equation (1) and X_i^* referring to the intersection of the i^{th} patient's HLA-A (or -B) and HLA-C peptide binding repertoires divided by HLA-A (or B) repertoire size. These analyses were implemented using R version 4.2.2 and model visualizations were prepared using the Matplotlib (v. 3.4.2) and seaborn (v. 0.11.2) Python libraries.

Protein sequence analysis, modeling, and visualization.

Aligned HLA-C amino acid sequences were obtained from the IPD-IMGT/HLA Database [87] in FASTA format. Three-dimensional structures were generated for each patient HLA-C protein using the ColabFold [88] implementation of AlphaFold2 [89]. Electrostatic maps were calculated and visualized using the Adaptive Poisson-Boltzman Solver (APBS) plugin for PyMOL (v. 2.5). The model of HLA-C03:04 bound to a SARS-CoV-2 NSP3 peptide (residues 1885-1894, UniProt ID: P0DTC1) was generated using AlphaFold-Multimer [https://doi.org/10.1101/2021.10.04.463034] followed by refinement using Rosetta FlexPepDock low- and high-resolution protocols [**?**]. All protein visualizations were prepared using PyMOL (v. 2.5), and peptide sequence logos were generated using the Logomaker library implemented in Python (v. 3.8). For sequence-based clustering analysis, residues at positions forming the HLA-C peptide binding surface were embedded using sgt (v. 2.0.3), a Python implementation of sequence graph transform. PCA and K-means clustering analysis was then performed on embedded sequences using the scikit-learn library (v. 0.24.2) in Python (v. 3.8). PCA scatterplots were prepared using the Matplotlib library (v. 3.4.2) implemented in Python (v. 3.8).

Statistics.

Pearson correlations were calculated using the pearsonr method implemented in the stats module implemented in the SciPy Python library (v. 1.6.2). Analysis of variance (ANOVA) and T-test calculations were performed using the SciPy f_oneway, and ttest_ind methods, respectively. A p-value cutoff of 0.05 was used to identify statistically significant results in all analyses described above. Data visualizations were prepared using the matplotlib (v. 3.4.2) and seaborn (v. 0.11.2) Python libraries.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

MDO: Writing – original draft, Writing – review & editing, Conceptualization, Data curation, Formal analysis, Methodology, Visualization; VAL: Conceptualization, Writing – review & editing; ALV: Investigation, Writing – review & editing; AZ: Investigation, Writing – review & editing; KJW: Conceptualization, Writing – review & editing; YL: Conceptualization, Writing – review & editing; ASL: Conceptualization, Resources, Writing – review & editing; MFC: Writing – review & editing, Conceptualization, Project administration, Resources, Supervision.

ACKNOWLEDGMENTS

We gratefully acknowledge all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based.

SUPPLEMENTAL DATA

TableS1.xlsx
TablesS2-4.xlsx
Supplementary_Material.pdf

DATA AVAILABILITY STATEMENT

The findings of this study are based on 60 sequences available on GISAID between March 27, 2020, and May 9, 2020, and accessible at GISAID EPI SET 220330me. All code used for downstream analysis is publicly available at the corresponding author's GitHub profile at GitHub: molp11/SARS_HLA.

REFERENCES

[1] Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in china. *Zhonghua Liu Xing Bing Xue Za Zhi* **41** (2020) 145–151.

[2] Wang Y, Wang Y, Chen Y, Qin Q. Unique epidemiological and clinical features of the emerging 2019 novel coronavirus pneumonia (COVID-19) implicate special control measures. *J. Med. Virol.* **92** (2020) 568–576.

[3] Zheng Z, Peng F, Xu B, Zhao J, Liu H, Peng J, et al. Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. *J. Infect.* **81** (2020) e16–e25.

[4] Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *Lancet* **395** (2020) 497–506.

[5] Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir. Med.* **8** (2020) 420–422.

[6] Goulder PJR, Walker BD. HIV and HLA class i: an evolving relationship. *Immunity* **37** (2012) 426–440.

[7] Carrington M, Nelson GW, Martin MP, Kissner T, Vlahov D, Goedert JJ, et al. HLA and HIV-1: heterozygote advantage and B*35-Cw*04 disadvantage. *Science* **283** (1999) 1748–1752.

[8] Diouf K, Sarr AD, Eisen G, Popper S, Mboup S, Kanki P. Associations between MHC class I and susceptibility to HIV-2 disease progression. *J. Hum. Virol.* **5** (2002) 1–7.

[9] International HIV Controllers Study, Pereyra F, Jia X, McLaren PJ, Telenti A, de Bakker PIW, et al. The major genetic determinants of HIV-1 control affect HLA class I peptide presentation. *Science* **330** (2010) 1551–1557.

[10] Wang L, Zou ZQ, Wang K. Clinical relevance of HLA gene variants in HBV infection. *J. Immunol. Res.* **2016** (2016) 9069375.

[11] Araujo P, Gonçalves G, Latini F, Ferreira O, Porto LC, Barreto JA, et al. KIR and a specific HLA-C gene are associated with susceptibility and resistance to hepatitis B virus infection in a brazilian population. *Cell. Mol. Immunol.* **11** (2014) 609–612.

[12] Montes-Cano MA, Caro-Oleas JL, Romero-Gómez M, Diago M, Andrade R, Carmona I, et al. HLA-C and KIR genes in hepatitis C virus infection. *Hum. Immunol.* **66** (2005) 1106–1109.

[13] Yoon SK, Han JY, Pyo CW, Yang JM, Jang JW, Kim CW, et al. Association between human leukocytes antigen alleles and chronic hepatitis C virus infection in the korean population. *Liver Int.* **25** (2005) 1122–1127.

[14] Ivić I, Bradarić N, Puizina-Ivić N, Ledina D, Luksić B, Martinić R. Hla-Cw7 allele as predictor of favorable therapeutic response to interferon-alpha in patients with chronic hepatitis C. *Croat. Med. J.* **48** (2007) 807–813.

[15] Frleta D, Yu CI, Klechevsky E, Flamar AL, Zurawski G, Banchereau J, et al. Influenza virus and poly(I:C) inhibit MHC class i-restricted presentation of cell-associated antigens derived from infected dead cells captured by human dendritic cells. *J. Immunol.* **182** (2009) 2766–2776.

[16] Boon ACM, de Mutsert G, Graus YMF, Fouchier RAM, Sintnicolaas K, Osterhaus ADME, et al. Sequence variation in a newly identified HLA-B35-restricted epitope in the influenza a virus nucleoprotein associated with escape from cytotoxic T lymphocytes. *J. Virol.* **76** (2002) 2567–2572.

[17] Dutta M, Dutta P, Medhi S, Borkakoty B, Biswas D. Polymorphism of HLA class I and class II alleles in influenza A(H1N1)pdm09 virus infected population of assam, northeast india. *J. Med. Virol.* **90** (2018) 854–860.

[18] Lin M, Tseng HK, Trejaut JA, Lee HL, Loo JH, Chu CC, et al. Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Med. Genet.* **4** (2003) 9.

[19] Sanchez-Mazas A. HLA studies in the context of coronavirus outbreaks. *Swiss Med. Wkly* **150** (2020) w20248.

[20] Ceraolo C, Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. *J. Med. Virol.* **92** (2020) 522–528.
[21] Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* **395** (2020) 565–574.
[22] Weiner J, Suwalski P, Holtgrewe M, Rakitko A, Thibeault C, Müller M, et al. Increased risk of severe clinical course of COVID-19 in carriers of HLA-C*04:01. *EClinicalMedicine* **40** (2021) 101099.
[23] Khor SS, Omae Y, Nishida N, Sugiyama M, Kinoshita N, Suzuki T, et al. HLA-A*11:01:01:01, HLA-C*12:02:02:01-HLA-B*52:01:02:02, age and sex are associated with severity of japanese COVID-19 with respiratory failure. *Front. Immunol.* **12** (2021) 658570.
[24] Anzurez A, Naka I, Miki S, Nakayama-Hosoya K, Isshiki M, Watanabe Y, et al. Association of HLA-DRB1*09:01 with severe COVID-19. *HLA* **98** (2021) 37–42.
[25] Wang F, Huang S, Gao R, Zhou Y, Lai C, Li Z, et al. Initial whole-genome sequencing and analysis of the host genetic contribution to COVID-19 severity and susceptibility. *Cell Discov.* **6** (2020) 83.
[26] Langton DJ, Bourke SC, Lie BA, Reiff G, Natu S, Darlay R, et al. The influence of HLA genotype on the severity of COVID-19 infection. *HLA* **98** (2021) 14–22.
[27] Ebrahimi S, Ghasemi-Basir HR, Majzoobi MM, Rasouli-Saravani A, Hajilooi M, Solgi G. HLA-DRB1*04 may predict the severity of disease in a group of iranian COVID-19 patients. *Hum. Immunol.* **82** (2021) 719–725.
[28] Vigón L, Galán M, Torres M, Martín-Galiano AJ, Rodríguez-Mora S, Mateos E, et al. Association between HLA-C alleles and COVID-19 severity in a pilot study with a spanish mediterranean caucasian cohort. *PLoS One* **17** (2022) e0272867.
[29] Augusto DG, Murdolo LD, Chatzileontiadou DSM, Sabatino JJ Jr, Yusufali T, Peyser ND, et al. A common allele of hla is associated with asymptomatic sars-cov-2 infection. *Nature* **620** (2023) 128–136.
[30] Wolday D, Fung CYJ, Morgan G, Casalino S, Frangione E, Taher J, et al. Hla variation and sars-cov-2 specific antibody response. *Viruses* **15** (2023) 906.
[31] Stanevich OV, Alekseeva EI, Sergeeva M, Fadeev AV, Komissarova KS, Ivanova AA, et al. Sars-cov-2 escape from cytotoxic t cells during long-term covid-19. *Nat. Commun.* **14** (2023) 149.
[32] Srivastava A, Hollenbach JA. The immunogenetics of covid-19. *Immunogenetics* **75** (2023) 309–320.
[33] Hoseinnezhad T, Soltani N, Ziarati S, Behboudi E, Mousavi MJ. The role of HLA genetic variants in COVID-19 susceptibility, severity, and mortality: A global review. *J Clin Lab Anal* **38** (2024) e25005.
[34] Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. *Viruses* **12** (2020) 254.
[35] Lee CH, Koohy H. In silico identification of vaccine targets for 2019-nCoV. *F1000Res.* **9** (2020) 145.
[36] Nguyen A, David JK, Maden SK, Wood MA, Weeder BR, Nellore A, et al. Human leukocyte antigen susceptibility map for severe acute respiratory syndrome coronavirus 2. *J. Virol.* **94** (2020).
[37] Enayatkhani M, Hasaniazad M, Faezi S, Gouklani H, Davoodian P, Ahmadi N, et al. Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. *J. Biomol. Struct. Dyn.* **39** (2021) 2857–2872.
[38] Charonis SA, James LM, Georgopoulos AP. SARS-CoV-2 in silico binding affinity to human leukocyte antigen (HLA) class II molecules predicts vaccine effectiveness across variants of concern (VOC). *Sci. Rep.* **12** (2022) 8074.

[39] Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48** (2020) W449–W454.

[40] Ferretti AP, Kula T, Wang Y, Nguyen DMV, Weinheimer A, Dunlap GS, et al. Unbiased screens show CD8+ T cells of COVID-19 patients recognize shared epitopes in SARS-CoV-2 that largely reside outside the spike protein. *Immunity* **53** (2020) 1095–1107.e3.

[41] Saini SK, Hersby DS, Tamhane T, Povlsen HR, Amaya Hernandez SP, Nielsen M, et al. SARS-CoV-2 genome-wide T cell epitope mapping reveals immunodominance and substantial CD8+ T cell activation in COVID-19 patients. *Sci. Immunol.* **6** (2021) eabf7550.

[42] Schulien I, Kemming J, Oberhardt V, Wild K, Seidel LM, Killmer S, et al. Characterization of pre-existing and induced SARS-CoV-2-specific CD8+ T cells. *Nat. Med.* **27** (2021) 78–85.

[43] Nersisyan S, Zhiyanov A, Shkurnikov M, Tonevitsky A. T-CoV: a comprehensive portal of HLA-peptide interactions affected by SARS-CoV-2 mutations. *Nucleic Acids Res.* **50** (2022) D883–D887.

[44] Yamada T, Wakabayashi M, Yamaji T, Chopra N, Mikami T, Miyashita H, et al. Value of leukocytosis and elevated c-reactive protein in predicting severe coronavirus 2019 (COVID-19): A systematic review and meta-analysis. *Clin. Chim. Acta* **509** (2020) 235–243.

[45] Bürkner PC. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal* **10** (2018) 395–411. doi:10.32614/RJ-2018-017.

[46] Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A probabilistic programming language. *Journal of Statistical Software* **76** (2017). doi:10.18637/jss.v076.i01.

[47] Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Moderbacher CR, et al. Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* **181** (2020) 1489–1501.e15.

[48] Peng Y, Mentzer AJ, Liu G, Yao X, Yin Z, Dong D, et al. Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nat. Immunol.* **21** (2020) 1336–1345.

[49] Rasmussen M, Harndahl M, Stryhn A, Boucherma R, Nielsen LL, Lemonnier FA, et al. Uncovering the peptide-binding specificities of HLA-C: a general strategy to determine the specificity of any MHC class I molecule. *J. Immunol.* **193** (2014) 4790–4802.

[50] Vilches C, Parham P. KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. *Annu. Rev. Immunol.* **20** (2002) 217–251.

[51] Caligiuri MA. Human natural killer cells. *Blood* **112** (2008) 461–469.

[52] Colonna M, Borsellino G, Falco M, Ferrara GB, Strominger JL. HLA-C is the inhibitory ligand that determines dominant resistance to lysis by NK1- and NK2-specific natural killer cells. *Proc. Natl. Acad. Sci. U. S. A.* **90** (1993) 12000–12004.

[53] Iturrieta-Zuazo I, Rita CG, García-Soidán A, de Malet Pintos-Fonseca A, Alonso-Alarcón N, Pariente-Rodríguez R, et al. Possible role of HLA class-i genotype in SARS-CoV-2 infection and progression: A pilot study in a cohort of covid-19 spanish patients. *Clin. Immunol.* **219** (2020) 108572.

[54] Lorente L, Martín MM, Franco A, Barrios Y, Cáceres JJ, Solé-Violán J, et al. HLA genetic polymorphisms and prognosis of patients with COVID-19. *Med. Intensiva (Engl. Ed.)* **45** (2021) 96–103.

[55] Correale P, Mutti L, Pentimalli F, Baglio G, Saladino RE, Sileri P, et al. HLA-B*44 and c*01 prevalence correlates with covid19 spreading across italy. *Int. J. Mol. Sci.* **21** (2020) 5205.

[56] Littera R, Campagna M, Deidda S, Angioni G, Cipri S, Melis M, et al. Human leukocyte antigen complex and other immunogenetic and clinical factors influence susceptibility or protection to SARS-CoV-2 infection and severity of the disease course. the sardinian experience. *Front. Immunol.* **11** (2020) 605688.

[57] Vishnubhotla R, Sasikala M, Ketavarapu V, Reddy DN. High-resolution HLA genotyping identifies alleles associated with severe COVID-19: A preliminary study from india. *Immun. Inflamm. Dis.* **9** (2021) 1781–1785.

[58] Warren RL, Birol I. HLA alleles measured from COVID-19 patient transcriptomes reveal associations with disease prognosis in a new york cohort. *PeerJ* **9** (2021) e12368.

[59] Snary D, Barnstable CJ, Bodmer WF, Crumpton MJ. Molecular structure of human histocompatibility antigens: the HLA-C series. *Eur. J. Immunol.* **7** (1977) 580–585.

[60] Zemmour J, Parham P. Distinctive polymorphism at the HLA-C locus: implications for the expression of HLA-C. *J. Exp. Med.* **176** (1992) 937–950.

[61] Neefjes JJ, Ploegh HL. Allele and locus-specific differences in cell surface expression and the association of HLA class I heavy chain with beta 2-microglobulin: differential effects of inhibition of glycosylation on class I subunit association. *Eur. J. Immunol.* **18** (1988) 801–810.

[62] McCutcheon JA, Gumperz J, Smith KD, Lutz CT, Parham P. Low HLA-C expression at cell surfaces correlates with increased turnover of heavy chain mRNA. *J. Exp. Med.* **181** (1995) 2085–2095.

[63] Alter G, Heckerman D, Schneidewind A, Fadda L, Kadie CM, Carlson JM, et al. HIV-1 adaptation to NK-cell-mediated immune pressure. *Nature* **476** (2011) 96–100.

[64] Fadda L, Körner C, Kumar S, van Teijlingen NH, Piechocka-Trocha A, Carrington M, et al. HLA-Cw*0102-restricted HIV-1 p24 epitope variants can modulate the binding of the inhibitory KIR2DL2 receptor and primary NK cell function. *PLoS Pathog.* **8** (2012) e1002805.

[65] Hölzemer A, Thobakgale CF, Jimenez Cruz CA, Garcia-Beltran WF, Carlson JM, van Teijlingen NH, et al. Selection of an HLA-C*03:04-restricted HIV-1 p24 gag sequence variant is associated with viral escape from KIR2DL3+ natural killer cells: Data from an observational cohort in south africa. *PLoS Med.* **12** (2015) e1001900; discussion e1001900.

[66] van Teijlingen NH, Hölzemer A, Körner C, García-Beltrán WF, Schafer JL, Fadda L, et al. Sequence variations in HIV-1 p24 gag-derived epitopes can alter binding of KIR2DL2 to HLA-C*03:04 and modulate primary natural killer cell function. *AIDS* **28** (2014) 1399–1408.

[67] Lin Z, Kuroki K, Kuse N, Sun X, Akahoshi T, Qi Y, et al. HIV-1 control by NK cells via reduced interaction between KIR2DL2 and HLA-C*12:02/C*14:03. *Cell Rep.* **17** (2016) 2210–2220.

[68] Ziegler MC, Nelde A, Weber JK, Schreitmüller CM, Martrus G, Huynh T, et al. HIV-1 induced changes in HLA-C*03 : 04-presented peptide repertoires lead to reduced engagement of inhibitory natural killer cell receptors. *AIDS* **34** (2020) 1713–1723.

[69] Lunemann S, Martrus G, Hölzemer A, Chapel A, Ziegler M, Körner C, et al. Sequence variations in HCV core-derived epitopes alter binding of KIR2DL3 to HLA-C*03:04 and modulate NK cell function. *J. Hepatol.* **65** (2016) 252–258.

[70] Ahlenstiel G, Martin MP, Gao X, Carrington M, Rehermann B. Distinct kir/hla compound genotypes affect the kinetics of human antiviral natural killer cell responses. *J. Clin. Invest.* **118** (2008) 1017–1026.

[71] La D, Czarnecki C, El-Gabalawy H, Kumar A, Meyers AFA, Bastien N, et al. Enrichment of variations in KIR3DL1/S1 and KIR2DL2/L3 among H1N1/09 ICU patients: an exploratory study. *PLoS One* **6** (2011) e29200.

[72] Koutsakos M, McWilliam HEG, Aktepe TE, Fritzlar S, Illing PT, Mifsud NA, et al. Downregulation of MHC class I expression by influenza a and B viruses. *Front. Immunol.* **10** (2019) 1158.

[73] Petitdemange C, Becquart P, Wauquier N, Béziat V, Debré P, Leroy EM, et al. Unconventional repertoire profile is imprinted during acute chikungunya infection for natural killer cells polarization toward cytotoxicity. *PLoS Pathog.* **7** (2011) e1002268.

[74] Petitdemange C, Wauquier N, Jacquet JM, Theodorou I, Leroy E, Vieillard V. Association of HLA class-i and inhibitory KIR genotypes in gabonese patients infected by chikungunya or dengue type-2 viruses. *PLoS One* **9** (2014) e108798.

[75] Wauquier N, Petitdemange C, Tarantino N, Maucourant C, Coomber M, Lungay V, et al. HLA-C-restricted viral epitopes are associated with an escape mechanism from KIR2DL2+ NK cells in lassa virus infection. *EBioMedicine* **40** (2019) 605–613.

[76] Wauquier N, Padilla C, Becquart P, Leroy E, Vieillard V. Association of KIR2DS1 and KIR2DS3 with fatal outcome in ebola virus infection. *Immunogenetics* **62** (2010) 767–771.

[77] Wawina-Bokalanga T, Vanmechelen B, Lhermitte V, Martí-Carreras J, Vergote V, Koundouno FR, et al. Human diversity of killer cell immunoglobulin-like receptors and human leukocyte antigen class I alleles and ebola virus disease outcomes. *Emerg. Infect. Dis.* **27** (2021) 76–84.

[78] Cimini E, Viola D, Cabeza-Cabrerizo M, Romanelli A, Tumino N, Sacchi A, et al. Different features of Vδ2 T and NK cells in fatal and non-fatal human ebola infections. *PLoS Negl. Trop. Dis.* **11** (2017) e0005645.

[79] Castro de Moura M, Davalos V, Planas-Serra L, Alvarez-Errico D, Arribas C, Ruiz M, et al. Epigenome-wide association study of COVID-19 severity with respiratory failure. *EBioMedicine* **66** (2021) 103339.

[80] Littera R, Chessa L, Deidda S, Angioni G, Campagna M, Lai S, et al. Natural killer-cell immunoglobulin-like receptors trigger differences in immune response to SARS-CoV-2 infection. *PLoS One* **16** (2021) e0255608.

[81] Hu S, Shao Z, Ni W, Sun P, Qiao J, Wan H, et al. The KIR2DL2/HLA-C1C1 gene pairing is associated with an increased risk of SARS-CoV-2 infection. *Front. Immunol.* **13** (2022) 919110.

[82] Sakuraba A, Haider H, Sato T. Population difference in allele frequency of HLA-C*05 and its correlation with COVID-19 mortality. *Viruses* **12** (2020).

[83] Basir HRG, Majzoobi MM, Ebrahimi S, Noroozbeygi M, Hashemi SH, Keramat F, et al. Susceptibility and severity of COVID-19 are both associated with lower overall viral-peptide binding repertoire of HLA class I molecules, especially in younger people. *Front. Immunol.* **13** (2022) 891816.

[84] Guerini FR, Bolognesi E, Lax A, Bianchi LNC, Caronni A, Zanzottera M, et al. HLA allele frequencies and association with severity of COVID-19 infection in northern italian patients. *Cells* **11** (2022) 1792.

[85] Bonaccorsi I, Carrega P, Venanzi Rullo E, Ducatelli R, Falco M, Freni J, et al. HLA-C*17 in COVID-19 patients: Hints for associations with severe clinical outcome and cardiovascular risk. *Immunol. Lett.* **234** (2021) 44–46.

[86] Khare S, Gurry C, Freitas L, Schultz MB, Bach G, Diallo A, et al. GISAID's role in pandemic response. *China CDC Wkly.* **3** (2021) 1049–1051.

[87] Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA database. *Nucleic Acids Res.* **48** (2020) D948–D955.

[88] Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: making protein folding accessible to all. *Nat. Methods* **19** (2022) 679–682.

549 [89] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein
550 structure prediction with AlphaFold. *Nature* **596** (2021) 583–589.

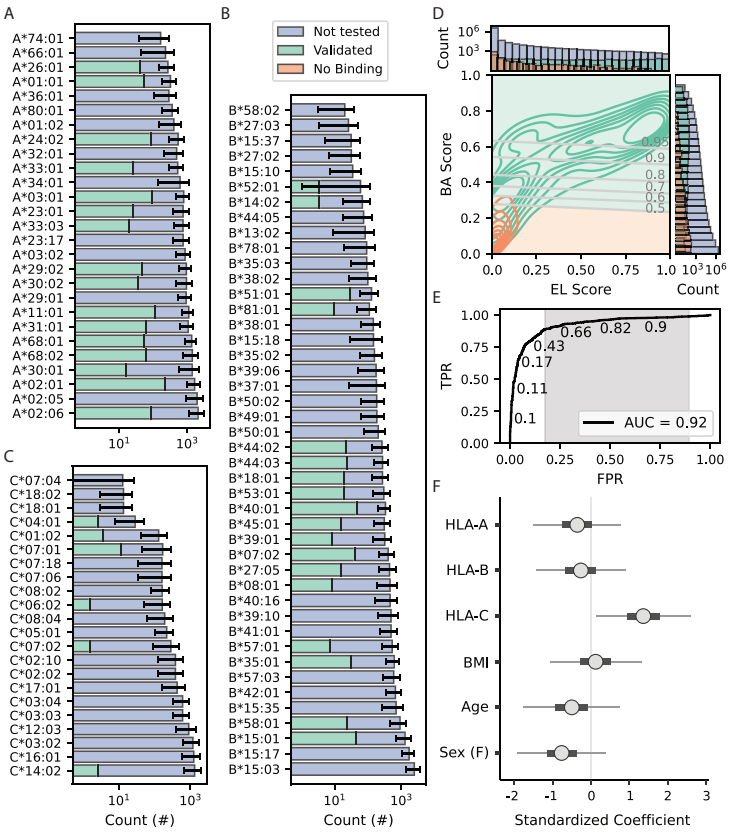

Figure 1. Infection severity correlates with total numbers of predicted HLA-C/SARS-CoV-2 peptide interactions. (A-C) Total numbers of SARS-CoV-2 peptides predicted to bind each MHC molecule (purple bars). Error bars represent standard deviation across probability thresholds. Subsets of interactions documented in the IEDB are shown as green bars. (D) Contour plot and bar plots showing distributions of netMHCpan BA and EL scores for predicted interactions. Green and orange lines/bars correspond to interactions showing positive and negative binding in the IEDB, respectively. Purple bars show the total number of predicted binding interactions. Gray lines indicate probability thresholds for positive binding. (E) ROC analysis of logistic regression probability model for identifying positive binding predictions. (F) Forest plot showing standardized multiple linear regression coefficients (circles) for predictors of patient SOFA scores, including 50% (thick lines) and 95% (thin lines) confidence intervals.

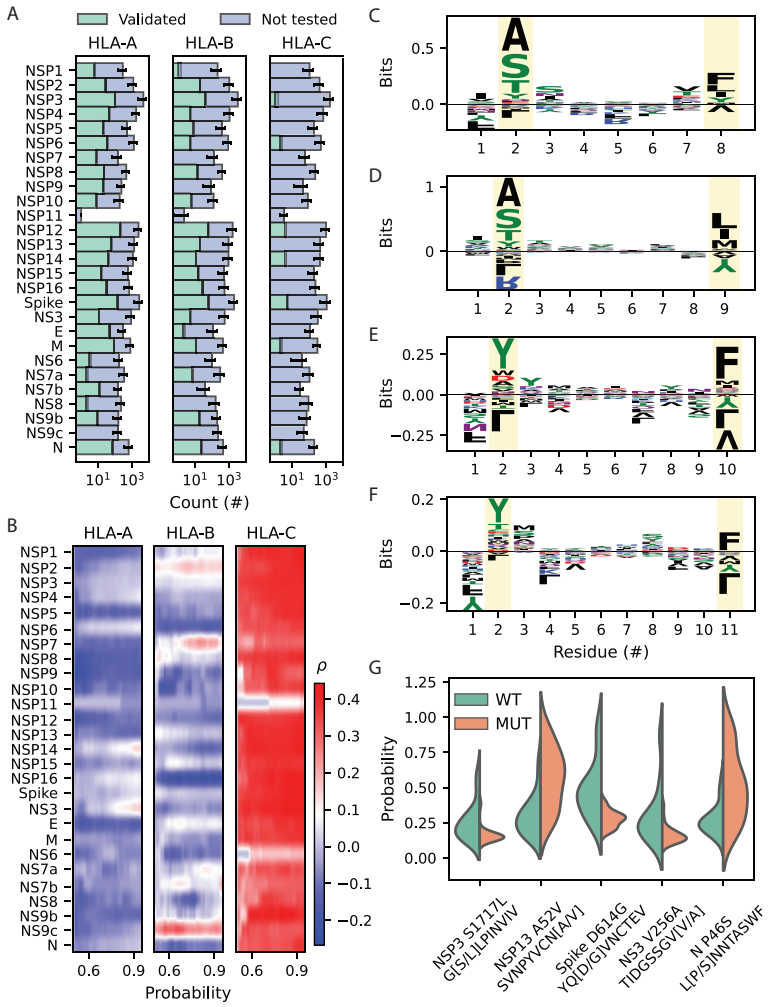

Figure 2. Severity of illness correlates with predicted HLA-C-mediated recognition of structural and non-structural SARS-CoV-2 protein sequences. (A) Distribution of predicted MHC class I binding across the SARS-CoV-2 proteome. (B) Heatmaps showing a correlation between predicted numbers of binding interactions and SOFA score across SARS-CoV-2 proteins. (C-F) Sequence enrichment plots for 8 to 11 aa peptides exhibiting the greatest magnitude correlation (top 1%) between HLA-C binding probability and SOFA score. (G) Split violin plots showing effects of missense mutations on predicted binding probability toward patient HLA-C molecules. Binding probability distributions for wild-type (Wuhan strain) and mutant strains are shown in green on the left and orange on the right, respectively. All distributions shown significantly differ between wild-type and mutant peptides ($p<0.05$).

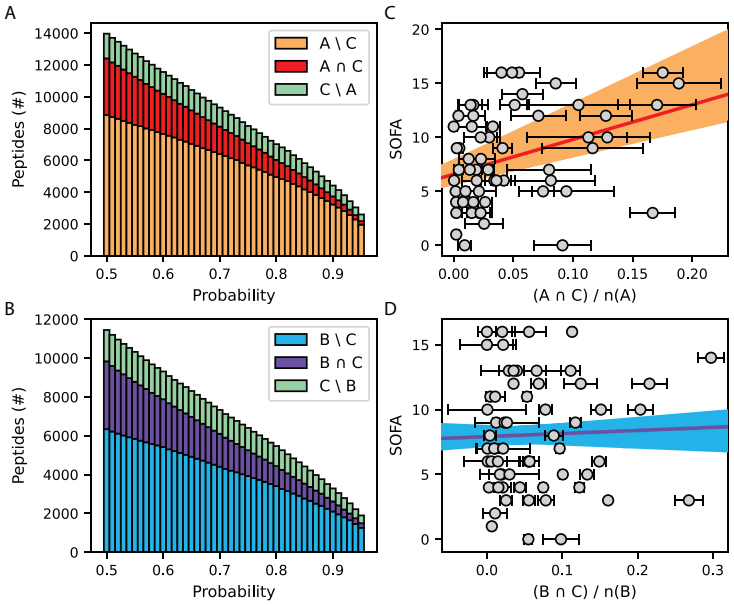

Figure 3. Infection severity correlates with the extent of overlap between SARS-CoV-2 peptide repertoires for patient HLA-A and -C molecules. (A-B) Stacked bar chart showing the intersection and differences between sets of SARS-CoV-2 peptides predicted to bind patient HLA-A, -B and -C molecules. (C-D) Results of linear regression using fractional intersection of peptide sets predicted to bind patient HLA-A/B and -C molecules to predict SOFA score. Error bars represent standard deviation across probability thresholds

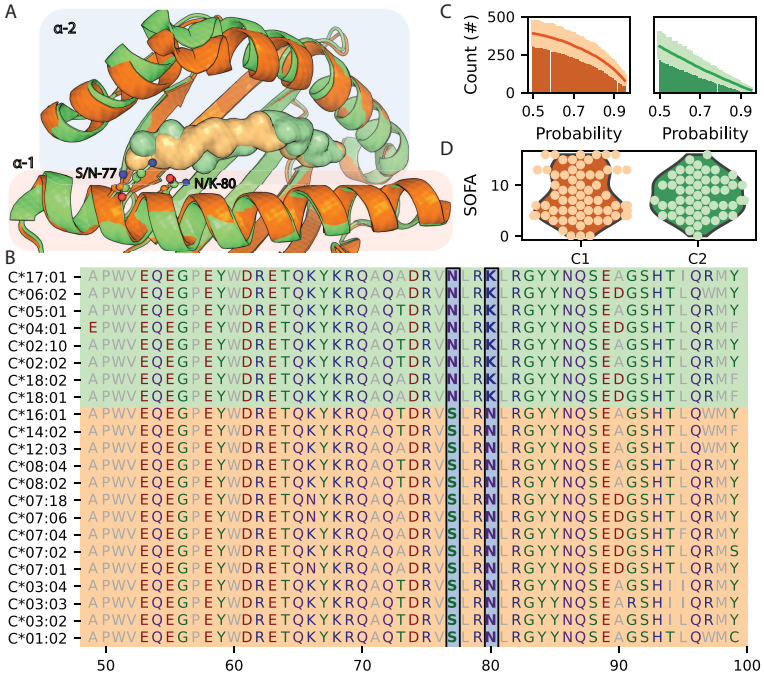

Figure 4. Relationship between KIR cluster, predicted SARS-CoV-2 peptide binding, and illness severity. (A) Structural overlay of C1 (HLA-C*06:02; PDB ID 5W6A) and C2 (HLA-C*07:02; 5VGE) molecules. (B) Protein Sequence alignment of HLA-C α-1 helices. (C) Mean numbers of predicted SARS-CoV-2 peptide-binding interactions for C1 and C2 molecules as a function of probability threshold. (D) Distributions of patient SOFA scores corresponding to C1 and C2 alleles.

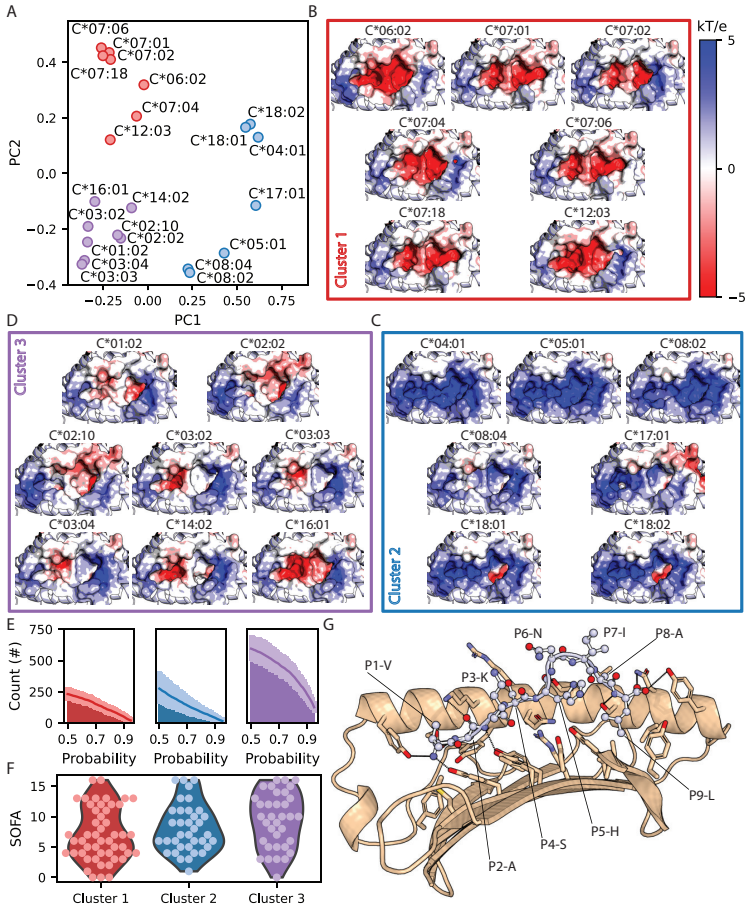

Figure 5. Effect of HLA-C peptide binding surface amino acid composition on predicted peptide repertoire and infection severity. (A) PCA and K-means clustering of patient HLA-C amino acid sequences. (B-D) Electrostatic surface maps of HLA-C peptide binding pockets grouped by sequence-based cluster. (E) Mean numbers of predicted SARS-CoV-2 peptide-binding interactions for each cluster across probability threshold. (F) Distribution of patient SOFA scores for alleles in each cluster. (G) Docked structure of HLA-C*03:04 bound to a 9 amino-acid SARS-CoV-2 NSP3 peptide (residues 1885-1894, UniProt ID: P0DTC1).