

## **Identification of risk and protective human leukocyte antigens in COVID-19 using genotyping and structural modeling**

Yiran Shen<sup>1</sup>, David A. Ostrov<sup>2</sup>, Santosh Rananaware<sup>3</sup>, Piyush K Jain<sup>3</sup>, Cuong Q. Nguyen<sup>1,4,5</sup>

<sup>1</sup>Department of Infectious Diseases and Immunology, <sup>2</sup>Department of Pathology, Immunology & Laboratory Medicine, <sup>3</sup> Department of Chemical Engineering, <sup>4</sup>Department of Oral Biology, College of Dentistry, <sup>5</sup>Center of Orphaned Autoimmune Diseases, University of Florida, Gainesville, Florida, 32611-0880 USA.

Address correspondence:

Cuong Q. Nguyen, PhD

Department of Infectious Diseases and Immunology

PO Box 110880, College of Veterinary Medicine

University of Florida, Gainesville, Florida 32611-0880 USA

Telephone: 352-294-4180, Fax: 352-392-9704

Email: [nguyenc@ufl.edu](mailto:nguyenc@ufl.edu)

**Running title:** The HLA association with COVID-19

## **Abstract**

COVID-19 is caused by severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2). The severity of COVID-19 is highly variable and related to known (e.g., age, obesity, immune deficiency) and unknown risk factors. Since innate and adaptive immune responses are elicited in COVID-19 patients, we genotyped 94 Florida patients with confirmed COVID-19 and 89 healthy controls. We identified an HLA gene, HLA-DPA1, in which specific alleles were associated with the risk of SARS-CoV-2 positivity and COVID-19 disease. HLA-DPA1\*01:03 was associated with reduced incidence of SARS-CoV-2 positivity, whereas HLA-DPA1\*03:01 was associated with increased risk of SARS-CoV-2 positivity. These data suggest a model in which COVID-19 severity is influenced by immunodominant peptides derived from SARS-CoV-2 preferentially presented by specific HLA-DP molecules to either protective (for asymptomatic COVID-19) or pathogenic T cells (in severe COVID-19). Although this study is limited to comparing SARS-CoV-2 positive and negative subjects, these data suggest that HLA typing of COVID-19 patients stratified for disease severity may be informative for identifying biomarkers and disease mechanisms in high-risk individuals.

## **Keywords**

COVID-19, SARS-CoV-2, Human leukocyte antigens, Genotyping, Structural modeling

## Introduction

Classical human leukocyte antigen (HLA) genes (HLA-A, -B, -C, -DR, -DQ, -DP) exhibit a high degree of polymorphism and play critical roles in the immune response to viral infections. CD4<sup>+</sup> T cells and CD8<sup>+</sup> T cells respond to pathogens by recognizing different classes of HLA molecules (I or II, respectively) on the cell surface. Specific HLA genotypes have been associated with T-cell mediated immunity and viral clearance, while low-affinity peptide binding and antigen presentation may also make specific genotypes a risk factor for infectious disease. For example, HLA-A\*02:01, -A\*03:01, -B\*08:01, -B\*18:01, -B\*37:01, -B\*57:01, and -DRB1\*09 alleles are involved in general protective CD8<sup>+</sup> or CD4<sup>+</sup> T cell-mediated immunity through presentation of conserved influenza peptides<sup>1-4</sup>. In contrast, HLA-A\*11, -A\*24, -A\*68, -B\*35, and -DRB1\*10 alleles may be the associated risk factor for severe pandemic influenza A (H1N1) infection and another common circulating influenza strains<sup>5-7</sup>. Studies of SARS-CoV demonstrated that HLA-B\*46:01, -B\*07:03, -Cw\*08:01, -DRB1\*12:02 and -DRB4\*01 alleles were associated with disease susceptibility or severity in various populations<sup>8-12</sup>. Similarly, HLA-DRB1\*11:01 and -DQB1\*02:02 alleles were associated with the risk of Middle East Respiratory Syndrome (MERS). HLA-associated adaptive immune responses were less efficient against human immunodeficiency virus (HIV)-1 than respiratory viruses<sup>13</sup>, but HLA-B\*27, -B\*57, -B\*58:01 and -B\*81:01 alleles were found to present multiple p24 Gag-specific epitopes in different ethnicities and confer protective effects against HIV disease progression, whereas HLA-B\*42:02 and -B\*58:02 alleles were associated with susceptibility and rapid disease progression<sup>14</sup>.

Predictions of binding affinity between HLA class I and II molecules and peptides of SARS-CoV-2 proteins showed that specific alleles might be associated with disease morbidity and mortality. For example, consistent with a previous study of SARS-CoV, HLA-B\*46:01 allele was predicted to be associated with SARS-CoV-2 susceptibility using silico analysis of viral peptide-MHC class I binding affinity to HLA -A, -B, and -C genotypes for all SARS-CoV-2 peptides<sup>15</sup>. HLA-C\*05 alleles were associated with risk of death from COVID-19<sup>16</sup>. Conversely, HLA-A\*02:11, -A\*02:22, -B\*15:03, -DRB1\*01:01, and -DRB1\*10:01 alleles were predicted to be protective due to their enhanced ability to present SARS-CoV-2 peptides<sup>13,15,17</sup>. Using whole-genome sequencing, a recent study with 332 patients demonstrated that HLA-A\*11:01, -B\*51:01, and -C\*14:02 alleles were correlated with disease severity<sup>18</sup>. These early studies suggest that HLA alleles are related to susceptibility and disease prognosis. Thus, identifying the association with SARS-CoV-2 positivity will help clarify the heterogeneity of responses to the disease, potentially guiding personalized treatments and developing epitope-based peptide vaccines against SARS-CoV-2. In this study, we sought to determine if specific HLA alleles were associated with susceptibility to SARS-CoV-2 infection/COVID-19 disease by comparing SARS-CoV-2 positive subjects in a Florida population with SARS-CoV-2 negative subjects from the same population.

## Materials and Methods

### Study population

Ninety-four confirmed COVID-19 samples were obtained from Boca Biolistics (Pompano Beach, FL) and CTSI Biorepository at the University of Florida (Gainesville, FL). The median age of the patients was 55.7 years (range: 3-94 years). Patients had positive test results for SARS-CoV-2 by RT-PCR from nasopharyngeal swabs or tracheal aspirates. Eighty-nine healthy individuals who had negative test results at the same sites and time were included as controls. The median age of the control group was 60 years (range: 0-101 years). In the SARS-CoV-2 positive cohort, there were 50 females, 43 males, and 1 undisclosed, in which there were 17 white, 21 black, 1 Asian, 1 Non-Hispanic, and 55 undisclosed. In control SARS-CoV-2 negative cohort, there were 47 females and 42 males, in which there were 44 white, 10 black, 1 Non-Hispanic, and 34 undisclosed enrolled. The study was approved by the Institutional Review Board of the University of Florida.

### HLA allele typing

Clinical specimens of nasopharyngeal swabs were collected in a viral transport medium. DNA was extracted from viral transport medium or directly from tracheal aspirates by Maxwell® RSC Blood DNA Kit per manufacturer's instructions (Promega Corporation). RNase A was added to samples to remove potential viral RNA. Isolated genomic DNA was quantified by NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer (Thermo Scientific). Genotyping was done using Axiom™ Human Genotyping SARS-CoV-2 Research Array as instructed by the manufacturer (Thermo Scientific). Genotyping data were assessed by Axiom™ Analysis Suite Software and filtered according to QC and HLA-related datasets. Hardy-Weinberg Equilibrium (HWE) was evaluated for all the SNPs evaluated, and all SNPs selected for HLA allele typing meet the criteria ( $p > 0.05$ ). Automated high-resolution HLA typing was done by Axiom™ HLA Analysis Software, and certain HLA-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1, -DRB1/3/4/5 alleles of each loci were assigned to samples.

### Peptide binding analysis

The Wuhan-Hu-1 sequence of the spike glycoprotein was used to predict peptides that bind HLA-DP molecules (<https://www.ncbi.nlm.nih.gov/datasets/coronavirus/proteins/>). The SMM-align (NetMHCII-1.1)(PMID:17608956) prediction method implemented in IEDB was used to generate the half-maximal inhibitory concentration (IC50) values that estimate dissociation constant ( $K_D$ ) in nanomolar values.

### Structural modeling of spike peptide HLA-DP interactions

The crystal structure of HLA-DP complexed with a peptide corresponding to RAS (PDB code 4P5K, 24995984) was used for modeling the SARS-CoV-2 spike peptide VVFLHVTYVPAQEKN (positions 1060-1074, corresponding to the S2 subunit). The HLA-DP

$\alpha$ -chain in this structure corresponds to HLA-DPA1\*01:03:01<sup>19</sup>. The RAS peptide NKFDLTLFHTITGGS was mutated to VVFLHVTYVPAQEKN using COOT<sup>20</sup> with rotamers that represent a local energy minimum of torsional angles. The geometry of the resulting complex was regularized in PHENIX<sup>21</sup>. The amino acid sequence of HLA-DPA1\*03:01:01 from IMGT<sup>19</sup> was used as the basis for structural modeling with SWISS-MODELER, which generated a protein data bank format file based on a crystal structure of HLA-DP  $\alpha$ -chain 98.9 % identical to HLA-DPA1\*03:01:01, PDB 4P5M. PyMOL (<https://pymol.org/2/>) was used to generate molecular graphic images.

## Statistical analysis

HLA class I and class II allele frequencies were estimated by direct counting based on results from allele typing, calculated as the ratio of the number of times different alleles appeared in the sample to the total number of alleles.. Odds ratios (ORs, 95% confidence interval [CI]) and p-values were calculated using R 4.0.3 (R Core Team, 2018), the exact2x2 (v1.6.5; Fay MP, 2010) package or Prism 9.0 software (GraphPad Software, La Jolla California USA). 0.5 was added to all the cells to prevent computation errors when calculating the odds ratio or standard error. When zero values caused computational problems with the odds ratio or its standard<sup>22</sup>. Since small size samples were used in this analysis, the risk of introducing a bias in estimating the probability (p-value) and wrongly accepting association (type I error) was corrected using the method of Benjamini-Hochberg. In this method, the p-value is multiplied by the number of alleles input at each locus, thereby giving a more powerful corrected p-value (pc-value) that may be interpreted with confidence.

## Results

### HLA typing of COVID-19 patients

We genotyped 94 SARS-CoV-2 positive patients and 89 healthy control populations matched as control. We identified multiple alleles present for HLA-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1, -DRB1 loci (30, 46, 20, 5, 20, 7, 16, and 32, respectively (**Table S1**)). HLA-DRB3/4/5 were not analyzed due to high rates of ambiguous imputation (**Table S2**). The allele distributions of HLA-A, -C, -B, -DRB1, -DQB1, and -DPB1 loci were compared between COVID-19 patients and control individuals. The frequencies and odds ratios (OR) of HLA alleles with significant associations with incidence of SARS-CoV-2 positivity are shown in **Table 1**. Potentially significant associations between HLA-DPA1 alleles and SARS-CoV-2 positivity were identified. DPA1\*03:01 was associated with an increased risk of SARS-CoV-2 positivity (OR 9.3, CI 1.3 - 200.4,  $p=0.01$ ). DPA1\*03:01 is a rare allele (frequency < 1% in controls), but present in 6% of SARS-CoV-2 positive individuals ( $P_c$  value 0.06). In contrast, DPA1\*01:03 was associated with reduced incidence of SARS-CoV-2 positivity (OR 0.6, CI 0.4 - 0.9,  $p=0.02$ ). DPA1\*01:03 is a common allele (frequency 72% in controls), but less frequent in individuals infected with SARS-CoV-2 (60 %,  $P_c$  value 0.06).

### Race comparison of risk and protective HLA

It is well established that minority groups of different races and ethnic groups in the US are disproportionately affected by COVID-19. Minorities endured a higher risk for infection, hospitalization, and death<sup>23,24</sup>. Therefore, in this study, we sought to determine if there was an association between the identified HLA and SARS-CoV-2 positivity in different races in the Floridian population. The data indicated that HLA-A\*02:01 was associated with reduced incidence of SARS-CoV-2 positivity (OR 0.1, CI 0.003-0.6,  $p=0.02$ ) in the black population (**Table S3**). Black individuals carrying the common HLA-A\*02:01 allele were ten times more likely to test negative for SARS-CoV-2 than black individuals lacking HLA-A\*02:01. Similar analyses were performed for the white Floridian population. As presented in **Table S4**, there was no significant allelic association in this population. The data indicate that HLA-A\*02:01 allele might be protective for the black minority group.

### Structural modeling of HLAs binding to SARS-CoV-2 spike protein:

Since different allelic forms of HLA molecules exhibit binding preferences for distinct sets of antigenic peptides, specific HLA allotypes may be less likely to present certain peptides derived from SARS-CoV-2 to T cells, resulting in a weaker antiviral response (as suggested by the association of HLA-B\*46:01 in SARS patients<sup>8</sup>). In contrast, other HLA molecules may bind with higher affinities to immunodominant SARS-CoV-2 derived peptides, resulting in more robust antiviral responses. Since the dominant immunogenic T cell epitopes derived from SARS-CoV-2 were recently defined from all proteins, including in the critical spike protein<sup>25</sup>, we sought to identify peptides that bind preferentially to HLA molecules associated with risk or protection. We identified a peptide corresponding to the SARS-CoV-2 spike

protein (15 amino acids, VVFLHVTYVPAQEKN) predicted to bind HLA-DP molecules related to risk with different affinities (**Figure 1**). VVFLHVTYVPAQEKN was predicted to bind an HLA-DP molecule associated with reduced incidence of SARS-CoV-2 positivity (i.e., lower levels of infection) with high affinity (estimated Kd 298 nM for HLA-DPA1\*01:03  $\alpha$ -chain/HLA-DPB1\*02:01  $\beta$ -chain, IEDB). In contrast, the same peptide was predicted to bind an HLA-DP molecule associated with increased risk of SARS-CoV-2 positivity with a low binding affinity (estimated Kd 1.7  $\mu$ M, HLA-DPA1\*03:01:01  $\alpha$ -chain/HLA-DPB1\*04:02  $\beta$ -chain).

Structural modeling suggests that SARS-CoV-2 spike peptide VVFLHVTYVPAQEKN binds DPA1\*01:03 with higher affinity than DPA1\*03:01 due to a polymorphism at position 42 located on the  $\alpha$ 1 helix of the DP  $\alpha$ -chain, oriented toward the peptide at the center of the antigen binding cleft. Alanine at position 42 of DPA1\*01:03  $\alpha$ -chain is predicted to anchor peptide binding by intermolecular contact with tyrosine in the peptide core (VVFLHVTYVPAQEKN)(**Figure 2**, upper panel). The bulky methionine at position 42 of DPA1\*03:01  $\alpha$ -chain is predicted to impose steric clash with the peptide core (**Figure 2**, lower panel), preventing formation of a high-affinity immunogenic epitope. These data suggest a model in which the course and severity of COVID-19 disease are influenced by the presentation of immunogenic peptides by specific HLA molecules present in a subset of individuals, **Figure 3**.



## Discussion

In late December 2019, COVID-19 began spreading in Wuhan, Hubei Province, China, caused by a suspected zoonotic source of SARS-CoV-2<sup>26,27</sup>. While coronaviruses are relatively common, mutations can cause severe symptoms in humans; COVID-19 is the third noted case in which this has happened following SARS<sup>28,29</sup> and MERS<sup>30</sup>. Globally, as of March 22<sup>th</sup>, 2021, there were more than 123 million confirmed cases and 2.71 million deaths worldwide; in the US there were 30 million confirmed cases with over 540,000 deaths<sup>31</sup>. Age and pre-existing medical conditions such as hypertension, obesity, chronic lung disease, diabetes mellitus, and cardiovascular disease are associated with disease severity and hospitalization rates of COVID-19 patients. In the COVID-19–Associated Hospitalization Surveillance Network (COVID-NET), which represents approximately 10% of the U.S. population with an equal frequency of males and females,, 54% of COVID-19-associated hospitalizations occurred in males and 46% occurred in females<sup>32</sup>. These data suggest that males may be disproportionately affected by COVID-19 compared with females. A meta-analysis of 15 independent studies documenting patient gender-specific outcomes found that men were more likely to develop severe COVID-19 infection than women (Odds Ratio, 1.31; 95% CI, 1.07-1.60)<sup>33</sup>. Other studies have indicated that black and Hispanic populations might be disproportionately affected by COVID-19<sup>32</sup>. The most interesting development in COVID-19 is that there are cases of children worldwide with COVID-19 exhibiting a Kawasaki-like disease<sup>34–36</sup>. Kawasaki disease has one of the strongest HLA associations, and there are significant differences in the distribution of HLA alleles among ethnicity<sup>37</sup>. This suggests that children with susceptible HLAs may develop Kawasaki-like disease with COVID-19.

There is a broad range of immunological responses to SARS-CoV-2, rendering individuals on a spectrum from asymptomatic to severely symptomatic for reasons that are not understood but likely result from genetic and environmental factors. To demonstrate whether HLA molecules are associated with COVID-19 infection and to further explore whether HLAs could serve as biomarkers for susceptibility or protection against SARS-CoV-2, we examined the HLA types of infected patients and compared them to samples collected from healthy individuals during the same period. The study identified that HLA-DPA1\*01:03 was highly prevalent in healthy individuals but less in SARS-CoV-2 positive patients, and HLA-DPA1\*03:01 was significantly associated in patients with SARS-CoV-2 infection. Using structural modeling with a potential T cell epitope derived from the spike protein, we predicted that HLA-DPA1\*01:03 may be protective because of high binding affinity between SARS-CoV-2 peptides and HLA-DP. In our model, a polymorphic position in the center of the antigen binding cleft at position 41 in HLA-DPA1\*03:01 was prevents high affinity peptide binding to HLA-DP, thus preventing specific T cell responses to SARS-CoV-2 peptides, consistent with risk associated with SARS-CoV-2 positivity.

HLA molecules present antigens by binding to endogenous antigenic peptides (class I) or exogenous antigenic peptides (class II) and express them as peptide-MHC complexes on the surface of antigen-presenting cells. During viral infection, cytotoxic T lymphocytes (CTL) kill



viral-infected cells by recognizing HLA class I-peptide complexes at the cell surface, and CD4<sup>+</sup> T cells recognize viral antigens presented by HLA class II molecules and activate antigen presenting cells to trigger an adaptive immune response against invading pathogens. Therefore, a spectrum of immune response is dictated by the flavor of the HLAs. Recent studies from different countries have identified multiple COVID-19 morbidity-related alleles; for example, Wang et al. identified HLA-B\*15:27 alleles from a Chinese population<sup>38</sup>, Novelli et al. identified HLA-B\*27:07, -DRB1\*15:01 and -DQB1\*06:02 alleles from an Italian population<sup>39</sup>, Yung et al. identified serotype B22(HLA-B\*54:01, B\*56:01 and B\*56:04 alleles) from Hongkong Chinese population<sup>40</sup>. To further support the concept that low affinity of viral peptides binding to HLA can predict susceptibility, Amoroso et al. have shown that HLA-DRB1\*08 was correlated to mortality (6.9% in living versus 17.5% in deceased), and peptide binding prediction analyses demonstrated that these alleles were not able to bind SARS-CoV-2 peptides with high affinity. A similar finding was supported by a study with Sardinian population in which HLA-DRB1\*08:01 allele was found only in the hospitalized patients<sup>41,42</sup>. Structural modeling of SARS-CoV-2 spike peptides and HLA-DP interaction in our study demonstrated that a polymorphism at position 42, located on the  $\alpha$ 1 helix of the DP  $\alpha$ -chain, leads to a difference in the binding affinity with peptides derived from spike protein. The alanine of HLA-DPA1\*01:03  $\alpha$ -chain had intermolecular contact with tyrosine in the peptide core, whereas the methionine of HLA-DPA1\*03:01  $\alpha$ -chain prevented high-affinity binding with tyrosine due to steric clash. The latter may compromise the ability of HLA-DPA1\*03:01 to present antigens optimally and activate the CD4<sup>+</sup> T cells, thereby undermines the effector function, specifically its ability to clear the virus causing infection effectively. The data suggest that HLA-DPA1\*01:03 may be protective because of effects on immunodominant peptide binding, whereas HLA-DPA1\*03:01 was associated with risk to SARS-CoV-2 infection because of limitations on binding immunodominant SARS-CoV-2 epitopes.

The limitations of this study were the small population cohort and the lack of patient clinical information that could be extrapolated to examine the associations of other clinical symptoms and HLAs. Overall, this study demonstrates HLA typing and in-silico structural modeling to identify susceptible and protective HLA alleles. This approach can potentially provide a genetic biomarker to determine if an individual is protected from the severity of the infection or if an individual is susceptible to the disease. These biomarkers may be essential in the decision-making process for developing and implementing a strategy to keep the individual safe if there is no vaccine or treatment available.

## **Conflict of Interests**

The authors declare no competing financial interests.

## **Author Contributions**

YS and SR conducted the experiments. PJ obtained patient samples and involved in the experimental design. YS, DO, and CN conceptualized the study, performed data analysis, and prepared the manuscript. All authors have read and approved the final manuscript.

## **Acknowledgments**

CQN is supported financially in part by PHS grants DE023433, DE018958, and DE028544 from the National Institutes of Health.

## **Data availability**

The data that support the findings of this study are presented in the manuscript and available in the Supplementary Materials. The data can also be available from the corresponding author upon request

## References

1. Quiñones-Parra S, Grant E, Loh L, et al. Preexisting CD8+ T-cell immunity to the H7N9 influenza A virus varies across ethnicities. *Proc Natl Acad Sci USA*. 2014;111(3):1049-1054. doi:10.1073/pnas.1322229111
2. Grant EJ, Josephs TM, Loh L, et al. Broad CD8+ T cell cross-recognition of distinct influenza A strains in humans. *Nat Commun*. 2018;9(1):5427. doi:10.1038/s41467-018-07815-5
3. Koutsakos M, Illing PT, Nguyen THO, et al. Human CD8+ T cell cross-reactivity across influenza A, B and C viruses. *Nat Immunol*. 2019;20(5):613-625. doi:10.1038/s41590-019-0320-6
4. Babon JAB, Cruz J, Ennis FA, Yin L, Terajima M. A human CD4+ T cell epitope in the influenza hemagglutinin is cross-reactive to influenza A virus subtypes and to influenza B virus. *J Virol*. 2012;86(17):9233-9243. doi:10.1128/JVI.06325-11
5. Falfán-Valencia R, Narayanankutty A, Reséndiz-Hernández JM, et al. An Increased Frequency in HLA Class I Alleles and Haplotypes Suggests Genetic Susceptibility to Influenza A (H1N1) 2009 Pandemic: A Case-Control Study. *J Immunol Res*. 2018;2018:3174868. doi:10.1155/2018/3174868
6. Hertz T, Oshansky CM, Roddam PL, et al. HLA targeting efficiency correlates with human T-cell response magnitude and with mortality from influenza A infection. *Proc Natl Acad Sci USA*. 2013;110(33):13492-13497. doi:10.1073/pnas.1221555110
7. 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*. 2018;90(5):854-860. doi:10.1002/jmv.25018
8. Lin M, Tseng H-K, Trejaut JA, et al. Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Med Genet*. 2003;4:9. doi:10.1186/1471-2350-4-9
9. Ng MHL, Lau K-M, Li L, et al. Association of human-leukocyte-antigen class I (B\*0703) and class II (DRB1\*0301) genotypes with susceptibility and resistance to the development of severe acute respiratory syndrome. *J Infect Dis*. 2004;190(3):515-518. doi:10.1086/421523
10. Chen Y-MA, Liang S-Y, Shih Y-P, et al. Epidemiological and genetic correlates of severe acute respiratory syndrome coronavirus infection in the hospital with the highest nosocomial infection rate in Taiwan in 2003. *J Clin Microbiol*. 2006;44(2):359-365. doi:10.1128/JCM.44.2.359-365.2006
11. Keicho N, Itoyama S, Kashiwase K, et al. Association of human leukocyte antigen class II alleles with severe acute respiratory syndrome in the Vietnamese population. *Hum Immunol*. 2009;70(7):527-531. doi:10.1016/j.humimm.2009.05.006
12. Ng MHL, Cheng SH, Lau KM, et al. Immunogenetics in SARS: a case-control study. *Hong Kong Med J*. 2010;16(5 Suppl 4):29-33.
13. Barquera R, Collen E, Di D, et al. Binding affinities of 438 HLA proteins to complete proteomes of seven pandemic viruses and distributions of strongest and weakest HLA peptide binders in populations worldwide. *HLA*. 2020;96(3):277-298. doi:10.1111/tan.13956
14. Kløverpris HN, Leslie A, Goulder P. Role of HLA Adaptation in HIV Evolution. *Front*

- Immunol.* 2015;6:665. doi:10.3389/fimmu.2015.00665
15. Nguyen A, David JK, Maden SK, et al. Human leukocyte antigen susceptibility map for severe acute respiratory syndrome coronavirus 2. *J Virol.* 2020;94(13). doi:10.1128/JVI.00510-20
16. Sakuraba A, Haider H, Sato T. Population Difference in Allele Frequency of HLA-C\*05 and Its Correlation with COVID-19 Mortality. *Viruses.* 2020;12(11). doi:10.3390/v12111333
17. Romero-López JP, Carnalla-Cortés M, Pacheco-Olvera DL, et al. A bioinformatic prediction of antigen presentation from SARS-CoV-2 spike protein revealed a theoretical correlation of HLA-DRB1\*01 with COVID-19 fatality in Mexican population: An ecological approach. *J Med Virol.* September 2020. doi:10.1002/jmv.26561
18. Wang F, Huang S, Gao H, et al. Initial Whole Genome Sequencing and Analysis of the Host Genetic Contribution to COVID-19 Severity and Susceptibility. *medRxiv.* June 2020. doi:10.1101/2020.06.09.20126607
19. Robinson J, Barker DJ, Georgiou X, Cooper MA, Flicek P, Marsh SGE. IPD-IMGT/HLA Database. *Nucleic Acids Res.* 2020;48(D1):D948-D955. doi:10.1093/nar/gkz950
20. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr Sect D, Biol Crystallogr.* 2010;66(Pt 4):486-501. doi:10.1107/S0907444910007493
21. Adams PD, Baker D, Brunger AT, et al. Advances, interactions, and future developments in the CNS, Phenix, and Rosetta structural biology software systems. *Annu Rev Biophys.* 2013;42:265-287. doi:10.1146/annurev-biophys-083012-130253
22. Principles of Biostatistics. *Technometrics.* 1994;36(4):436-437. doi:10.1080/00401706.1994.10485875
23. Romano SD, Blackstock AJ, Taylor EV, et al. Trends in Racial and Ethnic Disparities in COVID-19 Hospitalizations, by Region — United States, March–December 2020. *MMWR Morb Mortal Wkly Rep.* 2021;70(15). doi:10.15585/mmwr.mm7015e2
24. Gold JAW, Wong KK, Szablewski CM, et al. Characteristics and Clinical Outcomes of Adult Patients Hospitalized with COVID-19 - Georgia, March 2020. *MMWR Morb Mortal Wkly Rep.* 2020;69(18):545-550. doi:10.15585/mmwr.mm6918e1
25. Grifoni A, Weiskopf D, Ramirez SI, et al. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell.* 2020;181(7):1489-1501.e15. doi:10.1016/j.cell.2020.05.015
26. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med.* 2020;382(8):727-733. doi:10.1056/NEJMoa2001017
27. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol.* 2020;5(4):536-544. doi:10.1038/s41564-020-0695-z
28. Ksiazek TG, Erdman D, Goldsmith CS, et al. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med.* 2003;348(20):1953-1966. doi:10.1056/NEJMoa030781
29. Rota PA, Oberste MS, Monroe SS, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science.*



- 2003;300(5624):1394-1399. doi:10.1126/science.1085952
30. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*. 2012;367(19):1814-1820. doi:10.1056/NEJMoa1211721
31. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis*. 2020;20(5):533-534. doi:10.1016/S1473-3099(20)30120-1
32. Garg S, Kim L, Whitaker M, et al. Hospitalization rates and characteristics of patients hospitalized with laboratory-confirmed coronavirus disease 2019. *MMWR Morb Mortal Wkly Rep*. 2020;69(15):458-464. doi:10.15585/mmwr.mm6915e3
33. Ueyama H, Kuno T, Takagi H, et al. Gender Difference Is Associated With Severity of Coronavirus Disease 2019 Infection: An Insight From a Meta-Analysis. *Crit Care Explor*. 2020;2(6):e0148. doi:10.1097/CCE.0000000000000148
34. Viner RM, Whittaker E. Kawasaki-like disease: emerging complication during the COVID-19 pandemic. *Lancet*. 2020;395(10239):1741-1743. doi:10.1016/S0140-6736(20)31129-6
35. Verdoni L, Mazza A, Gervasoni A, et al. An outbreak of severe Kawasaki-like disease at the Italian epicentre of the SARS-CoV-2 epidemic: an observational cohort study. *Lancet*. 2020;395(10239):1771-1778. doi:10.1016/S0140-6736(20)31103-X
36. CDC COVID-19 Response Team. Coronavirus Disease 2019 in Children - United States, February 12-April 2, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(14):422-426. doi:10.15585/mmwr.mm6914e4
37. Onouchi Y. Molecular genetics of Kawasaki disease. *Pediatr Res*. 2009;65(5 Pt 2):46R-54R. doi:10.1203/PDR.0b013e31819dba60
38. Wang W, Zhang W, Zhang J, He J, Zhu F. Distribution of HLA allele frequencies in 82 Chinese individuals with coronavirus disease-2019 (COVID-19). *HLA*. 2020;96(2):194-196. doi:10.1111/tan.13941
39. Novelli A, Andreani M, Biancolella M, et al. HLA allele frequencies and susceptibility to COVID-19 in a group of 99 Italian patients. *HLA*. 2020;96(5):610-614. doi:10.1111/tan.14047
40. Yung Y-L, Cheng C-K, Chan H-Y, et al. Association of HLA-B22 serotype with SARS-CoV-2 susceptibility in Hong Kong Chinese patients. *HLA*. November 2020. doi:10.1111/tan.14135
41. Amoroso A, Magistrini P, Vespasiano F, et al. HLA and ABO Polymorphisms May Influence SARS-CoV-2 Infection and COVID-19 Severity. *Transplantation*. 2021;105(1):193-200. doi:10.1097/TP.0000000000003507
42. Littera R, Campagna M, Deidda S, 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*. 2020;11:605688. doi:10.3389/fimmu.2020.605688

**Table 1 Significant alleles associate with either protective (OR <1) or risk (OR>1) factor.**

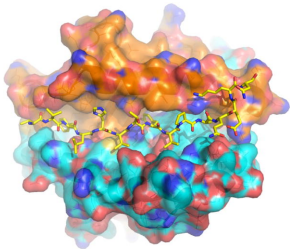
Allele	COVID-19		Control group		Odds ratio (95% CI)	P-value	Pc-value
	N	Frequency (%)	N	Frequency (%)			
<b>DPA1*03:01</b>	10	5.9524	1	0.6757	9.2568(1.3264 - 200.3514)	0.0120	0.06
<b>A*74:01</b>	9	5.4878	1	0.6757	8.4924(1.2738 - 185.6196)	0.0211	0.579
<b>A*30:01</b>	8	4.8780	1	0.6757	7.5017(1.1526 - 165.9402)	0.0386	0.579
<b>DPB1*01:01</b>	23	15.5405	8	5.7971	2.9792(1.2446 - 7.258)	0.0123	0.214
<b>DQA1*05:01</b>	40	26.3158	22	15.9420	1.8791(1.0435 - 3.5288)	0.0326	0.1141
<b>DPA1*01:03</b>	10 1	60.1190	107	72.2973	0.5786(0.3551 - 0.9401)	0.0244	0.06
<b>DQA1*02:01</b>	14	9.2105	26	18.8406	0.4383(0.2122 - 0.8942)	0.0257	0.1141
<b>DPB1*11:01</b>	2	1.3514	9	6.5217	0.1974(0.0301 - 0.9301)	0.0301	0.396

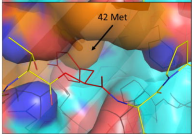
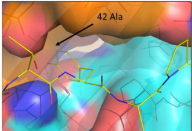
## Figure legends

**Figure 1.** Model of an HLA-DP molecule associated with reduced incidence of SARS-CoV-2 positivity. SARS-CoV-2 spike peptide VVFLHVTYVPAQEKN, is shown as sticks modeled on the crystal structure of HLA-DP (PDB code 4P5K, DPA1\*01:03  $\alpha$ -chain), yellow for carbon, red for oxygen, blue for nitrogen. The molecular surface of the HLA-DP  $\alpha$ -chain is shown in orange for carbon, red for oxygen, blue for nitrogen. The molecular surface of the HLA-DP  $\beta$ -chain is shown in cyan for carbon, red for oxygen, blue for nitrogen.

**Figure 2.** Polymorphism in the antigen binding cleft of HLA-DP has the potential to inhibit binding to peptides derived from SARS-CoV-2. Upper panel, DPA1\*01:03, associated with reduced incidence of SARS-CoV-2 positivity forms an  $\alpha$ -chain predicted to bind SARS-CoV-2 spike peptide VVFLHVTYVPAQEKN with high affinity. The molecular surface of the HLA-DP  $\alpha$ -chain is shown in orange for carbon, red for oxygen, blue for nitrogen. The molecular surface of the HLA-DP  $\beta$ -chain is shown in cyan for carbon, red for oxygen, blue for nitrogen. The tyrosine residue at the central peptide position forms intermolecular contact with alanine at position 42 of DPA1\*01:03. Lower panel, DPA1\*03:01, associated with increased incidence of SARS-CoV-2 positivity, forms an  $\alpha$ -chain predicted to bind SARS-CoV-2 spike peptide VVFLHVTYVPAQEKN with low affinity because of the steric clash between the tyrosine residue at the central peptide position (shown as red lines) and methionine at position 42 of DPA1\*03:01.

**Figure 3.** HLA-DP allele-specific viral clearance model. HLA-DP molecules with certain  $\alpha$ -chains may bind immunodominant epitopes (color dots within cells) derived from antigen presenting cells with high affinity, whereas others may bind with low affinity, affects whether CD4 T cells can be effectively activated (left, efficiently activated; right, inefficiently activated) to promote downstream responses for viral clearance. Polymorphism at position 42, located on the  $\alpha$ 1 helix of the DP  $\alpha$ -chain, leads to a difference in the binding affinity with peptides derived from spike protein (e.g., VVFLHVTYVPAQEKN), alanine of HLA- DPA1\*01:03  $\alpha$ -chain had intermolecular contact with tyrosine in the peptide core (lower left corner, the peptide is shown by ball-and-stick), and methionine of -DPA1\*03:01  $\alpha$ -chain prevented high affinity binding with tyrosine due to steric clash (lower right corner, the peptide is shown by ball-and-stick, crossed at low binding affinity position). | Created with BioRender.com





**Protective**

**Risk**

SARS-Cov-2

antigen  
presenting  
cells

HLA-DP

TCRs

peptides from  
spike protein

CD4 T cells

efficient effector T  
cells

inefficient effector  
T cells

Immunodominant peptide, e.g.,  
VVFLHVTYVPAQEK

42 Ala

DPA1\*01:03

Immunodominant peptide, e.g.,  
VVFLHVTYVPAQEK

42 Met

DPA1\*03:01

