Minimal observed impact of HLA genotype on hospitalization and severity of SARS-CoV-2 infection

Austin Nguyen[1], **Tasneem Yusufali**[2], **Jill A. Hollenbach**[*],[2],[3], **Abhinav Nellore**[*],[1],[4], **Reid F. Thompson**[*],[1],[5],[6],[7]

[1.]Department of Biomedical Engineering; Oregon Health & Science University; Portland, OR, 97239; USA

[2.]Department of Neurology; University of California, San Francisco; San Francisco, CA, 94158; USA

[3.]Department of Epidemiology and Biostatistics; University of California, San Francisco; San Francisco, CA, 94158; USA

[4.]Department of Surgery; Oregon Health & Science University; Portland, OR, 97239; USA

[5.]Department of Radiation Medicine; Oregon Health & Science University; Portland, OR, 97239; USA

[6.]Department of Medical Informatics and Clinical Epidemiology; Oregon Health & Science University; Portland, OR, 97239; USA

[7.]Division of Hospital and Specialty Medicine; VA Portland Healthcare System; Portland, OR, 97239; USA

Abstract

HLA is a critical component of the viral antigen presentation pathway. We investigated the relationship between severity of SARS-CoV-2 disease and HLA type in 3,235 individuals with confirmed SARS-CoV-2 infection. We found only the DPB1 locus to be associated with the binary outcome of whether an individual developed any COVID-19 symptoms. The number of peptides predicted to bind to an HLA allele had no significant relationship with disease severity both when stratifying individuals by ancestry or age and in a pooled analysis. Overall, at the population level, we found HLA type is significantly less predictive of COVID-19 disease severity than certain demographic factors and clinical comorbidities.

BACKGROUND

The global COVID-19 pandemic has exposed significant gaps in our ability to predict disease trajectory among individuals, with many people experiencing asymptomatic infections while others may be hospitalized with or die from COVID-19. Observational analyses have identified disease severity risk factors such as age, BMI, and sex (1–3).

However, host immunogenetic factors such as human leukocyte antigen (HLA) type may help determine the severity of SARS-CoV-2 infection. HLA is a critical component of the viral antigen presentation pathway, and previous studies have shown that individual HLA alleles may confer differential susceptibility and severity across viral diseases, including SARS-CoV-2 (4–11).

While large genotype association studies have investigated the relationship between genetic variants and severity of COVID-19 disease, they have not generally implicated the HLA locus (10,12–14). Further, while a growing collection of single institution or regional hospital system-based studies have reported HLA associations with COVID-19 disease (15–18), the statistical significance of these associations does not withstand correction for multiple comparisons. The relationship between HLA genotype and severity of COVID-19 disease, especially across a large and diverse population, thus remains unclear.

In this study, we investigated the specific relationship between HLA type and COVID-19 severity in a cohort of 3,235 individuals obtained from AncestryDNA (10,19) with confirmed SARS-CoV-2 infection.

RESULTS

We extracted basic demographic and clinical data for 3,235 individuals among the AncestryDNA cohort (10,19) with a positive SARS-CoV-2 nasal swab and classified the severity of their COVID-19 disease according to patient survey responses (Table 1). We next assessed the extent to which these demographic and clinical features predicted COVID-19 severity, and we found comorbidities that are contributors in a linear model predicting hospitalization (Supplementary Table 1).

We next assessed the contribution of genetic variability across the HLA locus to hospitalization as a binary outcome. Using BIGDAWG for case-control association analysis (20), we found no individual HLA types or specific amino acid variants across the HLA locus that were associated with hospitalization (Supplementary File 1). Only the DPB1 locus (p=0.04) was found to be associated with the binary outcome of whether an individual developed any COVID-19 symptoms.

To assess an individual's capacity to present SARS-CoV-2 peptides, we computed HLA-specific MHC binding affinities of all k-mers of sizes between 8 and 12 inclusive from the SARS-CoV-2 proteome (n=48,395 unique peptides) passing a proteasomal cleavage propensity filter. We used two different predictive tools: netMHCpan and HLAthena (21,22). In agreement with our prior work (11), we find a wide variety in putative peptide presentation capacity across different HLA types (Supplementary Figure 1).

We next developed a pooled multivariate model of severity score, accounting for comorbidities as well as putative viral peptide presentation as a function of HLA type, and we found that age ($p < 0.01$), BMI ($p < 0.01$), asthma status ($p < 0.01$), diabetes status ($p < 0.01$), and other lung conditions ($p < 0.05$) were all predictive (Figure 1; Supplementary Table 1- Stratified_LM_models_corrected.csv). There was no association between the number of putatively presented class I peptides and COVID-19 severity.

Predicted number of presented viral peptides demonstrated no significant relationship with disease severity when stratifying by individual ancestry.

DISCUSSION

HLA genes are generally considered important for host response to novel infectious diseases. In this study, we found that age, BMI, and other comorbidities determined clinical outcome across 3,235 individuals as described in the literature (1–3,13), and to a far greater degree than an individual's HLA-specific capacity to present SARS-CoV-2-specific peptides. While we previously explored the potential of HLA-peptide binding to predict COVID-19 severity (11), we do not see evidence for this phenomenon in the large real-world clinical cohort explored here. While the majority of the individuals were imputed to be of European ancestry, there were sizable numbers of individuals of Amerindian, Asian, and African descent. While Roberts et al. (10) performed a stratified GWAS analysis using this same dataset, with binary endpoints of hospitalization and whether an individual developed any COVID-19 symptoms, they did not specifically explore the role of HLA, which has a high level of variability that reduces power to detect differences in populations. Further, we investigated SARS-CoV-2 specific peptide presentation as a nonlinear function of HLA type, where some HLA types may be more similar to each other in the number of predicted peptides they can bind than they may be in canonical HLA supergroups.

We note several limitations to our work. Firstly, the proportion of SARS-CoV-2 peptides that we tested were generated through whole-peptidome *in silico* analysis of SARS-CoV-2. This may not be representative of the actual SARS-CoV-2 peptides presented in a given individual, whether due to biological sources such as viral variation, or methodological sources such as potential inaccuracies in peptide-MHC binding affinity predictions. Secondly, individuals who suffered debilitating infections may have been less likely to participate in the survey, and no individuals who died of COVID-19 were able to participate in the study, potentially resulting in an undercounting of the most severe phenotypes. Further, the cohort was primarily European, with much smaller sample sizes for African, Asian, and Amerindian ancestry. Lastly, these data were composed entirely of the unvaccinated cohort, as this population was tested and surveyed before the release of the many SARS-CoV-2 vaccines.

A number of other studies (15–18,23–25) have examined the relationship between HLA alleles and COVID-19 severity, and few have found alleles significantly associated with severity. In the majority of these studies, the large number of possible alleles in each study reduced the statistical power to identify significant alleles after multiple testing correction. Further, a number of studies reporting statistical significant associations between severity and HLA type were regional; they tended to have more ethnically and geographically homogeneous cohorts, likely resulting in overrepresentation of some alleles. Taken together with our analysis of the AncestryDNA dataset, we suggest that the literature does not reliably support the role of HLA type in modifying real-world COVID-19 disease severity across a population. There are multiple potential explanations for this, including that the data and analyses to date do not accurately reflect the true potential disease-modifying effects of HLA genes. On an individual basis, HLA type may indeed influence the severity

of COVID-19 disease; however, this hypothesis is not readily borne out at a population level, at least in this cohort. Multiple demographic features and clinical comorbidities are significantly more predictive of disease severity in a population. While we acknowledge the potential for other studies such as the COVID-19 Host Genetics Initiative to uncover a disease modifying role of HLA, future work should take a very critical and individualized approach towards evaluating any connections between HLA variation and differences in COVID-19 disease severity.

MATERIALS AND METHODS

Genotyping:

Data from 15000 individuals belonging to the Ancestry COVID-19 study was obtained through an IRB-approved project (Ancestry Human Diversity Project). The data authorized for reuse by AncestryDNA included: AncestryDNA genome-wide scale genotypes, AncestryDNA research participants' self-reported age, gender, height, weight, and smoking status, and survey answers for the AncestryDNA COVID-19 questionnaire. As reported by the AncestryDNA COVID-19 study, the study participants' genotype data were obtained using an Illumina genotyping array (Illumina OmniExpress platform) composed of 730,525 SNPs. Genotyping array data was processed by Illumina or Quest/Athena Diagnostics (10).

Ancestry imputation:

Genetic ancestry was determined using plinkQC v1.9 (26) to combine genotypes of the cohort with genotypes of a reference dataset (27) consisting of individuals of known ethnicities. Principal component analysis (PCA) on the combined genotype panel was used to detect population structure of the reference dataset to the level of continental ancestry. A→T and C→G SNPs were removed from study and reference data as they are more difficult to align and only a subset of SNPs were required for the analysis. The study data were pruned for variants in linkage disequilibrium (LD) with an $r^2 > 0.2$ in a 50kb window, and that list of pruned variants was used to reduce the size of the reference dataset. Checks were performed to ensure matching variant IDs and chromosomal positions between the study and reference dataset before merging and running PCA. Ancestries for the study population were then imputed from the principal components provided for the labeled reference dataset.

HLA class I + II imputation:

HLA Class I/II alleles were obtained using HIBAG v1.3 (28), a prediction method for HLA imputation that utilizes large training sets with known HLA and SNP genotypes in combination with attribute bagging. Ancestry-specific pre-fit models available within HIBAG for European, Asian, African, and Amerindian populations were applied to the subgroups of distinct ancestries within the AncestryDNA cohort using 1,042 SNPs across and nearby the HLA locus.

Severity scoring and hospitalization:

We collapsed the 10 point WHO COVID-19 Ordinal Scale of disease severity (29) into a 7-point scale to accommodate available phenotype information in the AncestryDNA COVID-19 study.

AncestryDNA Survey State	Severity Score	WHO Patient State	WHO Score
Uninfected	0	Uninfected	0
Asymptomatic	1	Asymptomatic	1
Symptomatic, mild symptoms	2	Symptomatic, no assistance needed	2
Symptomatic, severe symptoms	3	Symptomatic, assistance needed	3
Hospitalized, no oxygen	4	Hospitalized, no oxygen	4
Hospitalized, oxygen	5	Hospitalized, oxygen by mask/nasal prongs	5
		Hospitalized, oxygen by NIV/high flow	6
Hospitalized, ventilator	6	Intubation and mechanical ventilation	7
		Mechanical ventilation and vasopressors	8
		Mechanical ventilation, vasopressors, dialysis, or ECMO	9

The possible symptoms in the AncestryDNA cohort are fever, shortness of breath, dry cough, body aches, abdominal pain, cough producing phlegm, and nausea. There are 3 levels of severity to each symptom: normal, severe, and very severe. We defined severe symptoms (Severity Score 3) as any of the listed symptoms at the severe or very severe level. In models where we used hospitalization as an endpoint, we added hospitalization as a binary variable, with scores >=4 considered hospitalized.

Note that COVID-19 survey response data were consistent with CDC case and hospitalization rates over similar time periods, with 14% v. 12% test positivity and 11% v. 14% hospitalization rate in the AncestryDNA and CDC datasets, respectively (10,30).

HLA-peptide predicted binding:

We obtained SARS-CoV-2 peptide sequences by k-merizing FASTA protein sequences obtained from the NCBI RefSeq database (NC_045512.2 and NC_004718.3) into 8- to 12-mers. These k-mers were filtered by NetChop v3.1 using default settings with a cutoff of 0.1. MHC class I binding affinity predictions were performed using netMHCpan v4.0 using the '-BA' option to include binding affinity predictions and the '-l' option to specify peptides 8 to 12 amino acids in length. Additional MHC class I binding affinity predictions were performed using HLAthena. For predicted peptide binding, we used the cutoff of <500nM for peptides predicted by netMHCpan v4.0 and the cutoff of >0.5 probability score for peptides predicted by HLAthena. While nearly all individuals have two HLA-A/B/C haplotypes constituting as few as three but as many as six distinct alleles, a single peptide may be predicted to bind to more than one of an individual's HLA alleles. While there is no definitive evidence that a peptide is more likely to be presented when predicted to bind to more than one allele, we wanted to capture this possibility by using 2 metrics: an overall predicted peptide value and a unique predicted peptide value. For each individual,

to calculate capacity to bind SARS-CoV-2 peptides, we summed the number of predicted peptides bound to each individual's allele (min 3, max 6). For a unique-peptide specific capacity, the peptides were filtered to remove duplicates after summation.

Statistical analyses:

We performed statistical tests for HLA vs. hospitalization using the Bridging ImmunoGenomic Data-Analysis Workflow Gaps (BIGDAWG) pipeline and a comprehensive SARS-CoV-2 peptide-genotype binding analysis for all individuals in our dataset. All statistical analyses were performed using R version 4.0.3. For each statistical test, we performed pooled and ancestry-stratified testing. For multivariate linear modeling, we used the R function lm for multivariate regression with one of severity index, hospitalization status, or asymptomatic/symptomatic as the endpoint. Tests of Hardy-Weinberg equilibrium using Chi-squared testing for haplotypes, loci, and HLA-amino acid positions were performed using the BIGDAWG v1.3.4 R package. Note that all reported p-values have been corrected for multiple hypothesis testing, where relevant, using Benjamini-Hochberg correction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References:

1. Li X, Xu S, Yu M, Wang K, Tao Y, Zhou Y, et al. Risk factors for severity and mortality in adult COVID-19 inpatients in Wuhan. J Allergy Clin Immunol. 2020 Jul 1;146(1):110–8. [PubMed: 32294485]

2. Wolff D, Nee S, Hickey NS, Marschollek M. Risk factors for Covid-19 severity and fatality: a structured literature review. Infection. 2021 Feb 1;49(1):15–28. [PubMed: 32860214]

3. Kompaniyets L Body Mass Index and Risk for COVID-19–Related Hospitalization, Intensive Care Unit Admission, Invasive Mechanical Ventilation, and Death — United States, March–December 2020. MMWR Morb Mortal Wkly Rep [Internet]. 2021 [cited 2021 Sep 16];70. Available from: https://www.cdc.gov/mmwr/volumes/70/wr/mm7010e4.htm

4. Zernich D, Purcell AW, Macdonald WA, Kjer-Nielsen L, Ely LK, Laham N, et al. Natural HLA Class I Polymorphism Controls the Pathway of Antigen Presentation and Susceptibility to Viral Evasion. J Exp Med. 2004 Jun 28;200(1):13–24. [PubMed: 15226359]

5. Blackwell JM, Jamieson SE, Burgner D. HLA and Infectious Diseases. Clin Microbiol Rev. 2009 Apr;22(2):370–85. [PubMed: 19366919]

6. Couture A, Garnier A, Docagne F, Boyer O, Vivien D, Le-Mauff B, et al. HLA-Class II Artificial Antigen Presenting Cells in CD4+ T Cell-Based Immunotherapy. Front Immunol. 2019;10:1081. [PubMed: 31156634]

7. Crux NB, Elahi S. Human Leukocyte Antigen (HLA) and Immune Regulation: How Do Classical and Non-Classical HLA Alleles Modulate Immune Response to Human Immunodeficiency Virus and Hepatitis C Virus Infections? Front Immunol. 2017;8:832. [PubMed: 28769934]

8. Cruz-Tapias P, Castiblanco J, Anaya J-M. Major histocompatibility complex: Antigen processing and presentation [Internet]. Autoimmunity: From Bench to Bedside [Internet]. El Rosario University Press; 2013 [cited 2021 Oct 21]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK459467/

9. Warren RL, Birol I. HLA alleles measured from COVID-19 patient transcriptomes reveal associations with disease prognosis in a New York cohort. PeerJ. 2021 Oct 15;9:e12368. [PubMed: 34722002]

10. Roberts GHL, Park DS, Coignet MV, McCurdy SR, Knight SC, Partha R, et al. AncestryDNA COVID-19 Host Genetic Study Identifies Three Novel Loci [Internet]. 2020 Oct [cited 2021 Sep 16] p. 2020.10.06.20205864. Available from: https://www.medrxiv.org/content/10.1101/2020.10.06.20205864v1

11. 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. Gallagher T, editor. J Virol. 2020 Jun 16;94(13):e00510–20. [PubMed: 32303592]

12. Li Y, Ke Y, Xia X, Wang Y, Cheng F, Liu X, et al. Genome-wide association study of COVID-19 severity among the Chinese population. Cell Discov. 2021 Aug 31;7(1):1–16. [PubMed: 33390590]

13. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N Engl J Med. 2020 Jun 17;NEJMoa2020283.

14. Velavan TP, Pallerla SR, Rüter J, Augustin Y, Kremsner PG, Krishna S, et al. Host genetic factors determining COVID-19 susceptibility and severity. EBioMedicine [Internet]. 2021 Oct 1 [cited 2021 Nov 15];72. Available from: https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(21)00422-9/fulltext

15. 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. 2021;98(1):14–22. [PubMed: 33896121]

16. Pisanti S, Deelen J, Gallina AM, Caputo M, Citro M, Abate M, et al. Correlation of the two most frequent HLA haplotypes in the Italian population to the differential regional incidence of Covid-19. J Transl Med. 2020 Sep 15;18(1):352. [PubMed: 32933522]

17. Migliorini F, Torsiello E, Spiezia F, Oliva F, Tingart M, Maffulli N. Association between HLA genotypes and COVID-19 susceptibility, severity and progression: a comprehensive review of the literature. Eur J Med Res. 2021 Aug 3;26(1):84. [PubMed: 34344463]

18. 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 Orlando Fla. 2020 Oct;219:108572.

19. Ancestry COVID-19 study - EGA European Genome-Phenome Archive [Internet]. [cited 2021 Nov 12]. Available from: https://ega-archive.org/studies/EGAS00001004716

20. Pappas DJ, Marin W, Hollenbach JA, Mack SJ. Bridging ImmunoGenomic Data Analysis Workflow Gaps (BIGDAWG): An integrated case-control analysis pipeline. Hum Immunol. 2016 Mar 1;77(3):283–7. [PubMed: 26708359]

21. 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. 2020 Jul 2;48(W1):W449–54. [PubMed: 32406916]

22. Sarkizova S, Klaeger S, Le PM, Li LW, Oliveira G, Keshishian H, et al. A large peptidome dataset improves HLA class I epitope prediction across most of the human population. Nat Biotechnol. 2020 Feb;38(2):199–209. [PubMed: 31844290]

23. Shkurnikov M, Nersisyan S, Jankevic T, Galatenko A, Gordeev I, Vechorko V, et al. Association of HLA Class I Genotypes With Severity of Coronavirus Disease-19. Front Immunol. 2021;12:423.

24. Ishii T Human Leukocyte Antigen (HLA) Class I Susceptible Alleles Against COVID-19 Increase Both Infection and Severity Rate. Cureus [Internet]. 2020 Dec 23 [cited 2021 Oct 21];12(12). Available from: https://www.cureus.com/articles/35178-human-leukocyte-antigen-hla-class-i-susceptible-alleles-against-covid-19-increase-both-infection-and-severity-rate

25. Toyoshima Y, Nemoto K, Matsumoto S, Nakamura Y, Kiyotani K. SARS-CoV-2 genomic variations associated with mortality rate of COVID-19. J Hum Genet. 2020;65(12):1075–82. [PubMed: 32699345]

26. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc. 2010 Sep;5(9):1564–73. [PubMed: 21085122]

27. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015 Oct;526(7571):68–74. [PubMed: 26432245]

28. Zheng X, Shen J, Cox C, Wakefield JC, Ehm MG, Nelson MR, et al. HIBAG—HLA genotype imputation with attribute bagging. Pharmacogenomics J. 2014 Apr;14(2):192–200. [PubMed: 23712092]

29. Marshall JC, Murthy S, Diaz J, Adhikari NK, Angus DC, Arabi YM, et al. A minimal common outcome measure set for COVID-19 clinical research. Lancet Infect Dis. 2020 Aug;20(8):e192–7. [PubMed: 32539990]

30. Knight SC, McCurdy SR, Rhead B, Coignet MV, Park DS, Roberts GHL, et al. COVID-19 susceptibility and severity risks in a survey of over 500,000 individuals [Internet]. 2021 Jan [cited 2022 Jan 10] p. 2020.10.08.20209593. Available from: https://www.medrxiv.org/content/10.1101/2020.10.08.20209593v3

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Forest Plot of Comorbidities

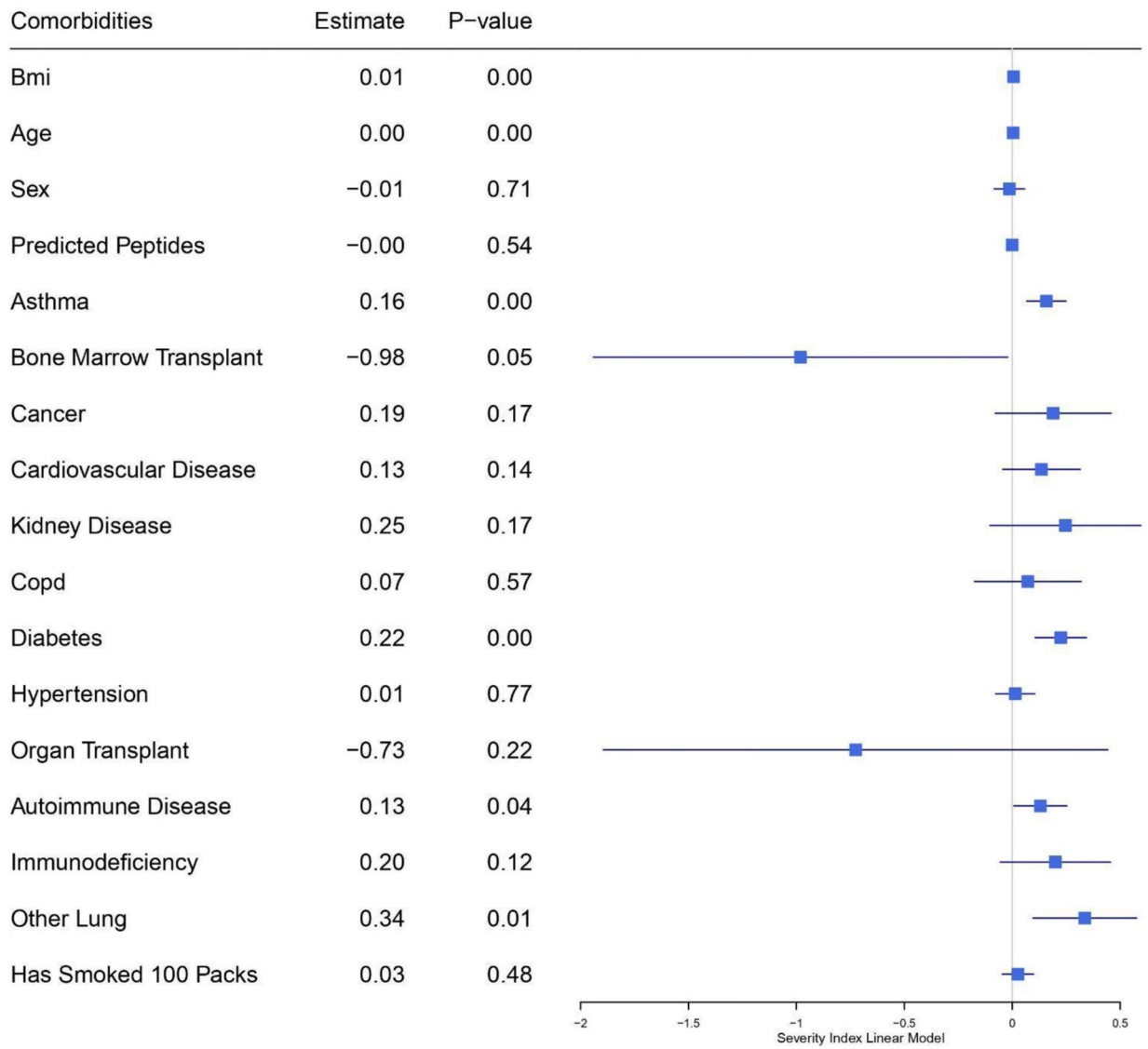| Comorbidities	Estimate	P-value
Bmi	0.01	0.00
Age	0.00	0.00
Sex	−0.01	0.71
Predicted Peptides	−0.00	0.54
Asthma	0.16	0.00
Bone Marrow Transplant	−0.98	0.05
Cancer	0.19	0.17
Cardiovascular Disease	0.13	0.14
Kidney Disease	0.25	0.17
Copd	0.07	0.57
Diabetes	0.22	0.00
Hypertension	0.01	0.77
Organ Transplant	−0.73	0.22
Autoimmune Disease	0.13	0.04
Immunodeficiency	0.20	0.12
Other Lung	0.34	0.01
Has Smoked 100 Packs	0.03	0.48

Severity Index Linear Model

Figure 1:

Forest plot of comorbidities including putative viral peptide presentation as a function of HLA type (Predicted Peptides) in the pooled multivariate model for predicting severity score. The estimate column shows the coefficients of variables in the linear model. Each line represents the 95% confidence interval for the estimate value. For the sex variable, female is 1 and male is 0. Positive values for the estimate are predicted to contribute to a higher severity score and vice versa for negative values.
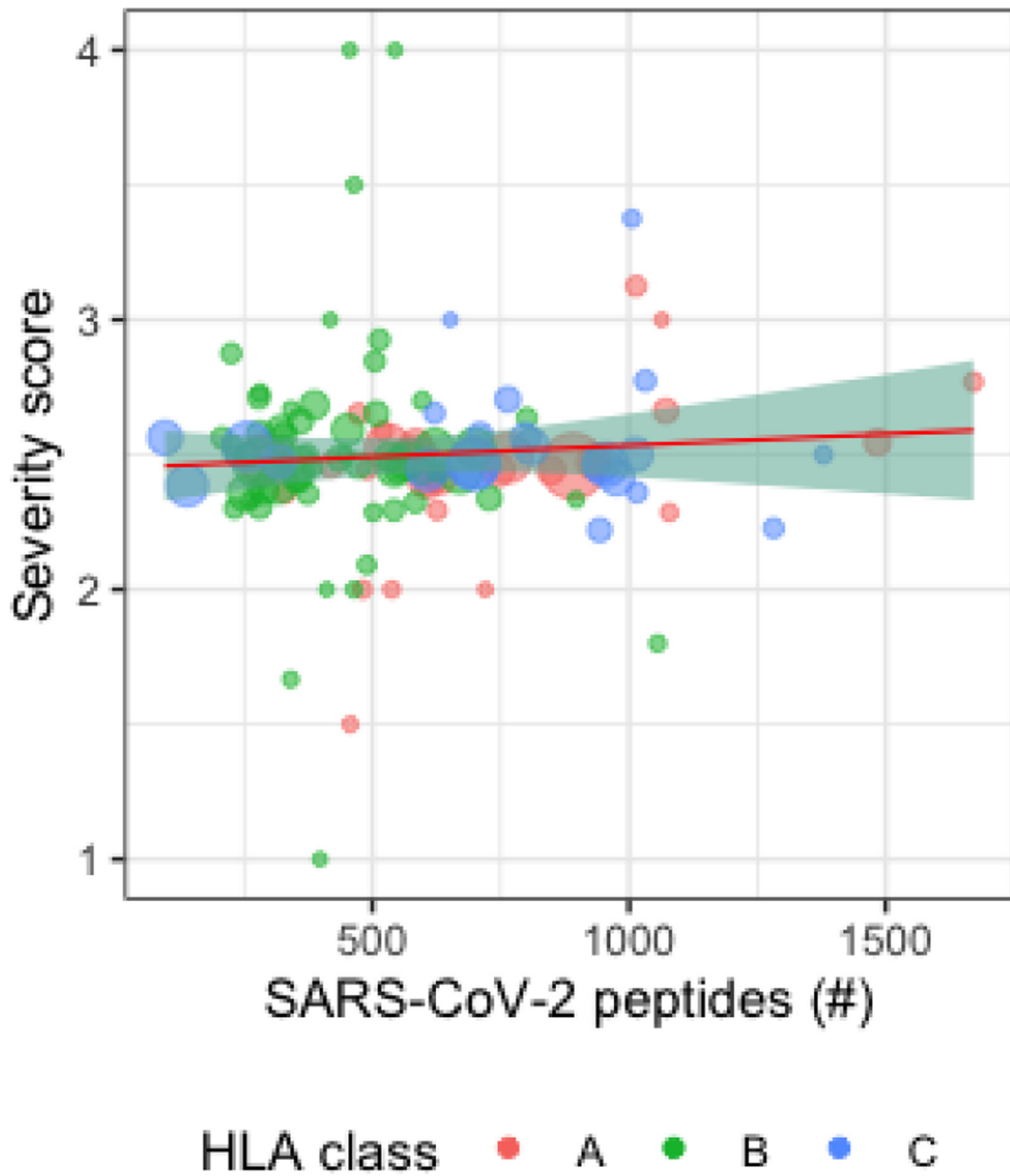
Figure 2:
Scatter plot of HLA alleles with the number of predicted peptides vs. average severity score in the AncestryDNA dataset. Each data point represents a distinct HLA allele, with larger points representing larger numbers of individuals in the AncestryDNA dataset imputed to have the allele and the red, green, and blue colors representing HLA-A, HLA-B, and HLA-C respectively.
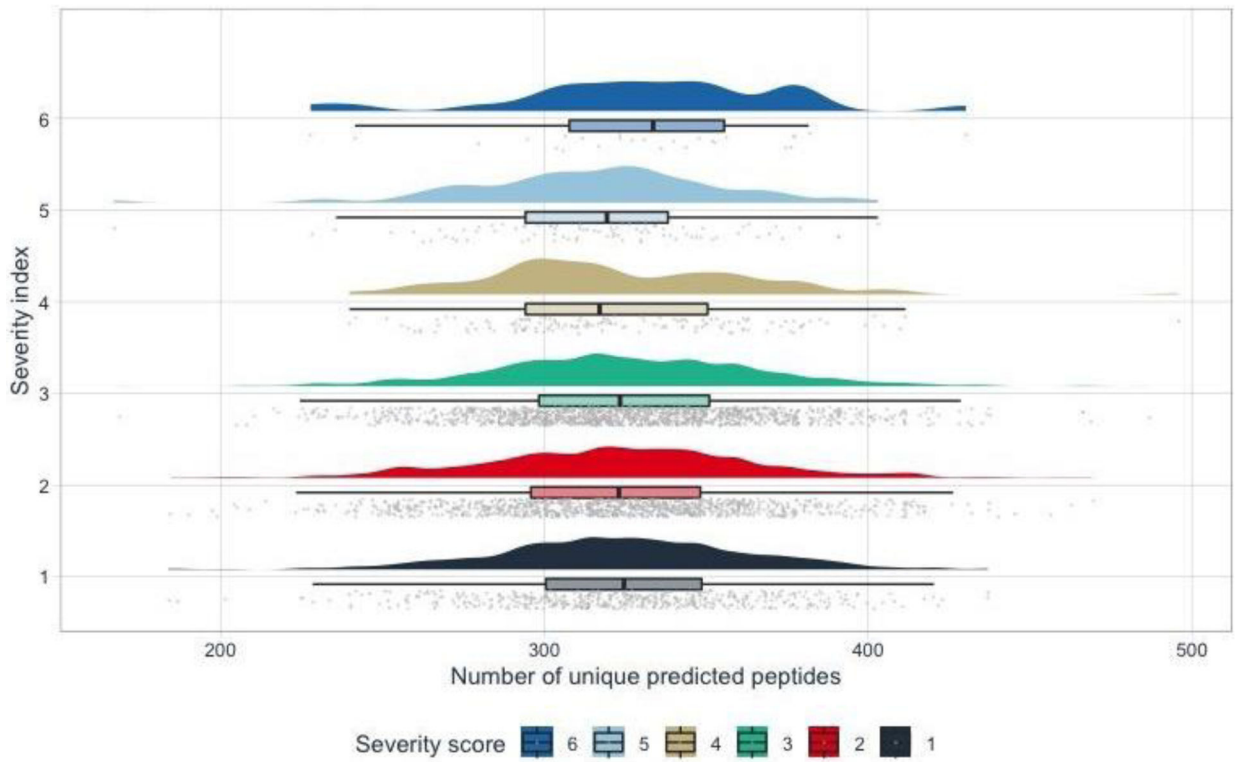
Figure 3:
Distribution of unique predicted peptides vs. severity score. The half-violin plots represent the distribution of unique predicted peptides of the 3,235 AncestryDNA individuals who had the corresponding severity score. The boxplot shows the IQR of the unique predicted peptides and each point in the rain cloud below the boxplot represents the number of predicted peptides of each individual.

Table 1:

Demographic breakdown of the AncestryDNA cohort. Severity score 0 is left out since it corresponds to individuals who tested negative for SARS-CoV-2 infection. Severity scores 1–6 are described in Materials and Methods: Severity scoring and hospitalization.

	Severity score 1	Severity score 2	Severity score 3	Severity score 4	Severity score 5	Severity score 6
Count	607	965	1418	136	83	26
Male (%)	201 (18)	394 (36)	398 (36)	46 (4)	45 (4)	15 (1)
Female (%)	299 (14)	678 (32)	1020 (48)	90 (4)	38 (3)	11 (1)
EUR (%)	365 (16)	769 (33)	1008 (44)	92 (4)	61 (3)	21 (1)
AS (%)	11 (20)	19 (33)	23 (40)	3 (5)	0 (0)	1 (1)
AMR (%)	88 (13)	237 (34)	320 (46)	30 (4)	17 (2)	2 (0)
AFR (%)	36 (21)	47 (28)	67 (40)	11 (7)	5 (3)	2 (1)
Median Age (range)	52 (19–88)	43.5 (19–85)	45 (19–85)	51 (21–86)	56 (20–84)	61.5 (29–88)