Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Human Immunology 83 (2022) 803–807

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

Research article

HLA-B evolutionary divergence is associated with outcomes after SARS-CoV-2 infection

Patricia V. Hernandez [a], Brian Duffy [b], Karl Hock [a], Christopher Farnsworth [a], Emily Schindler [c], Chang Liu [a,*]

[a] Department of Pathology and Immunology, Washington University in St. Louis School of Medicine, United States
[b] Barnes-Jewish Hospital, HLA Laboratory, United States
[c] Mercy Hospital in St. Louis, United States

ARTICLE INFO

Keywords:
COVID-19
HLA class I
HLA-B
HLA Evolutionary divergence
Grantham's distance

ABSTRACT

We examined the correlation between class I HLA evolutionary divergence (HED), a surrogate for the capacity to present different peptides, and the outcomes of 234 adult inpatients with confirmed SARS-CoV-2 infection. Genomic DNA was extracted from peripheral blood and genotyped by next-generation sequencing (NGS). HED scores for HLA class I (HLA-A, -B, and -C) genotypes were calculated using Grantham's distance. Higher HED scores for HLA-B, but not HLA-A or -C, are significantly associated with a decreased probability of poor outcomes including ICU admission, mechanical ventilation, and death (OR = 0.93; P = 0.04) in the univariate analysis. In the multivariate analysis, increased HLA-B HED score, younger age, and no comorbidity were independently associated with favorable outcomes (P = 0.02, P = 0.01, and P = 0.05, respectively). This finding is consistent with the notion that broader peptide repertoires presented by class I HLA may be beneficial in infection control.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the cause of the Coronavirus Disease 2019 (COVID-19) pandemic. As of April 9, 2022, there were more than 822,000 new cases daily, with a cumulative case number surpassing 276,000,000 and more than 5,300,000 reported deaths globally [1]. While most patients infected by SARS-CoV-2 present with a mild form of the disease, about 20 % of infected individuals were hospitalized, and 5 % of the cases were severe during the early phase of the pandemic [2]. It is known that age and some comorbidities confer an increased risk for the severe presentation of COVID-19, however younger and otherwise healthy individuals are not completely protected from poor outcomes after infection [3].

Human leukocyte antigens (HLA) are highly diverse transmembrane proteins that present viral peptides to *T*-cells and launch pathogen-specific immune responses [4]. Previously, we reported that *HLA-DRB1*08:02* and *HLA-A*30:02* are potential risk factors for symptomatic COVID-19 in Hispanic and younger African American populations [4]. However, various case-control studies have reported multiple class I and II HLA alleles associated with disease outcomes without a clear consensus [5]. Discrepancies across different investigations may be attributed to differences in the definition of clinical phenotypes and outcomes, research subjects' self-reported presentations, ancestries of study subjects, and limited sample sizes. As a result, the interpretation of HLA and COVID-19 outcomes has been challenging [5]. To further understand the contribution of HLA to COVID-19, an alternative approach is to quantify the HLA diversity across different alleles to correlate with disease severity.

HLA evolutionary divergence (HED) has been defined as the physiochemical difference between HLA protein sequences, which correlates with the diversity of immunopeptidomes presented to T-cells [6]. HED can be computed as Grantham's distance based on the difference in composition, polarity, and molecular volume between each pair of amino acids from maternal and paternal HLA alleles at each locus [7]. The more diverse are the features of paired amino acids, the higher the probability of more diverse peptides being presented to immune cells [7,8]. Furthermore, some authors have used HED to

Abbreviations: CI, confidence interval; COVID-19, Coronavirus Disease 2019; HED, HLA evolutionary diverge; HLA, human leukocyte antigen; IQR, interquartile range; OR, odds ratio; SARS-CoV-2, Severe acute respiratory syndrome coronavirus 2.

* Corresponding author at: Division of Laboratory and Genomic Medicine, Department of Pathology and Immunology, Washington University School of Medicine, 660 South Euclid Ave. Campus Box 8118, St. Louis, MO 63110, United States.
E-mail address: cliu32@wustl.edu (C. Liu).

https://doi.org/10.1016/j.humimm.2022.09.004
Received 29 May 2022; Revised 28 July 2022; Accepted 6 September 2022
Available online 12 September 2022

estimate the capacity to present different peptides as a prognostic factor in cancer and aplastic anemia [6,9,10]. We hypothesize that higher HED scores for class I HLA loci may be associated with favorable outcomes in hospitalized COVID-19 patients due to the increased probability of presenting viral peptides for recognition by adaptive immunity. We performed a retrospective analysis of a cohort of COVID-19 inpatients in the Midwestern United States with diverse ancestries to test this hypothesis.

2. Methods

2.1. Patients and clinical data

We enrolled adult patients hospitalized for COVID-19 at Barnes-Jewish Hospital and Mercy Hospital in St. Louis between March to July 2020. The inclusion criteria were an age of 18 or older and hospitalized for SARS-CoV-2 infection, which was confirmed by a real-time reverse transcriptase-polymerase chain reaction testing of nasopharyngeal swabs. The study was approved by the Human Research Protection Office of Washington University in St. Louis (IRB ID #: 202004002) and the Institutional Review Board of Mercy Hospital (IRB ID #:1599032–2). Clinical data were gathered retrospectively from electronic medical records and entered into a REDCap database [11]. Data gathering occurred at least three weeks after the initial enrollment to allow capturing the severity of potentially dynamic disease courses. Comorbidities were documented if the patient had a history of any of the following conditions: chronic lung disease, cardiovascular disease, diabetes, end-stage renal disease, autoimmune disease, cancer, trauma, surgery, and sepsis. A poor outcome was defined as an admission to the intensive care unit (ICU), a need for mechanical ventilation, or death. A favorable outcome was defined as the absence of the above poor outcomes.

2.2. Patient samples and HLA genotyping

As described previously [4], all cases had a remnant EDTA-anticoagulated blood specimen available in the clinical laboratories. Genomic DNA was extracted and isolated from these samples using the EZ1 DNA blood 350 μl kit (Qiagen, Hilden, Germany). Subsequently, 192 samples were typed by the AllType assay (One Lambda, West Hills, CA) on the Ion Chef/S5 Ion Torrent platform [12], and 42 samples were amplified using the NGS LR kit (One Lambda, West Hills, CA) with sequencing carried out using the SQK-LSK109 protocol on the R10.3 MinION flow cells (FLO-MIN111, Oxford Nanopore Technologies) [13,14]. HLA genes were assigned based on exon sequences (G groups) and limited to the 2-field resolution.

2.3. Data analysis and statistics

The HED score was computed for each genotype at *HLA-A*, *-B*, and *-C* loci using the HLAdivR package in R [15]. For statistical analysis, categorical variables were described as frequency and percentages, and numerical variables as the median and interquartile range (IQR). The normal distribution was tested by Shapiro's test. Fisher's exact test was applied to assess distributions for sex, comorbidities, and HLA heterozygosity between patients with favorable and poor outcomes. The differences in age, BMI, and HLA HED scores between the groups were compared by the Mann-Whitney test. Outcomes were further assessed by multivariate logistic regression (using backward selection). The association between HLA alleles and outcomes was analyzed at 1- and 2-field resolutions for African Americans, Caucasians, and total populations using the pyHLA package (version 1.1.1) with default options. The default minimal allele frequency of 0.05 was applied, and multiple comparisons were adjusted in these analyses by controlling the false discovery rate at 5 % using the Benjamini-Hochberg procedure [16]. The rest of the statistical analyses were completed using R version 4.1.2 (2021-11-01), Vienna, Austria. Results were considered statistically significant at a 2-tail p-value of < 0.05.

3. Results

3.1. Patient characteristics and outcomes

A total of 234 patients were enrolled in this study, 96 being females (41 %). The median age was 66 years old (IQR: 55–73), and African Americans comprised 71.4 % of the cohort (Table 1). Only 19 patients (8.1 %) presented with no comorbidity; the rest of the cohort had one or more comorbidities, with cardiovascular diseases being the most common (supplemental material, Table S1). A total of 97 (41.5 %) patients had a favorable outcome, while 137 (58.5 %) patients had poor outcomes from SARS-CoV-2 infection, including an ICU admission (121, 51.7 % of total), a need for mechanical ventilation (75, 32.1 % of total) or death (51, 21.8 % of total). Younger age and absence of comorbidities were significantly associated with favorable outcomes in the univariate analysis (Table 1). As the most prevalent comorbidity in this cohort, cardiovascular diseases alone were not significantly associated with poor COVID-19 outcomes (OR = 1.25; 95 % CI, 0.69–2.26, P = 0.45).

3.2. Association between HLA alleles and outcomes

No HLA allele (2-field) or allele group (1-field) was significantly associated with COVID-19 outcomes in the entire cohort of patients, African Americans, or Caucasians (Table S2-S4). Neither was heterozygosity, defined as a genotype with different maternal and paternal HLA alleles at the 2-field resolution, associated with outcomes at any of the loci.

3.3. Correlation between HED and outcomes

The median (and range) of HED scores in the cohort for *HLA-A*, *-B*, and *-C* were 6.86 (0–12.87), 8.03 (0–14.46), and 4.82 (0–7.60), respectively. HED scores for *HLA-B* locus were significantly higher in the group with favorable outcomes than those with poor outcomes (8.99 versus 7.22, $p < 0.05$). At *HLA-A* and *-C* loci, HED scores were not significantly different between groups with favorable and poor outcomes (Fig. 1). In addition, the combined HED scores for all class I alleles were not significantly different between the two outcome groups (P = 0.20). No significant difference in HLA-A, -B, and -C HED scores was observed between African Americans and Caucasians (Table 2).

We detected a significant association between higher HLA-B HED scores and favorable outcomes, with each 1-point increase in HLA-B HED associated with a 7 % decreased probability for the composite endpoints of ICU admission, mechanical ventilation, and death

Table 1
Demographic features of patients with favorable and poor outcomes after SARS-CoV-2 infection.

Feature	Favorable outcome N = 97	Poor outcome N = 137	P-value
Median age (IQR), in years	62.0 (50.0–70.0)	67 (57.0–75.0)	**<0.01**
Median BMI (IQR), in kg/m^2	29.7 (23.8–35.6)	28.1 (23.3–33.3)	0.24
Female, frequency (%)	44 (45.4)	52 (38.0)	0.28
Race, frequency (%)			0.89
African American	68 (70.1)	99 (72.3)	
Caucasian	25 (25.8)	31 (22.6)	
Asian or Pacific Islander	3 (3.1)	4 (2.9)	
Hispanic	1 (1.0)	3 (2.2)	
Comorbidity, frequency (%)	84 (86.6)	131 (95.6)	**0.02**

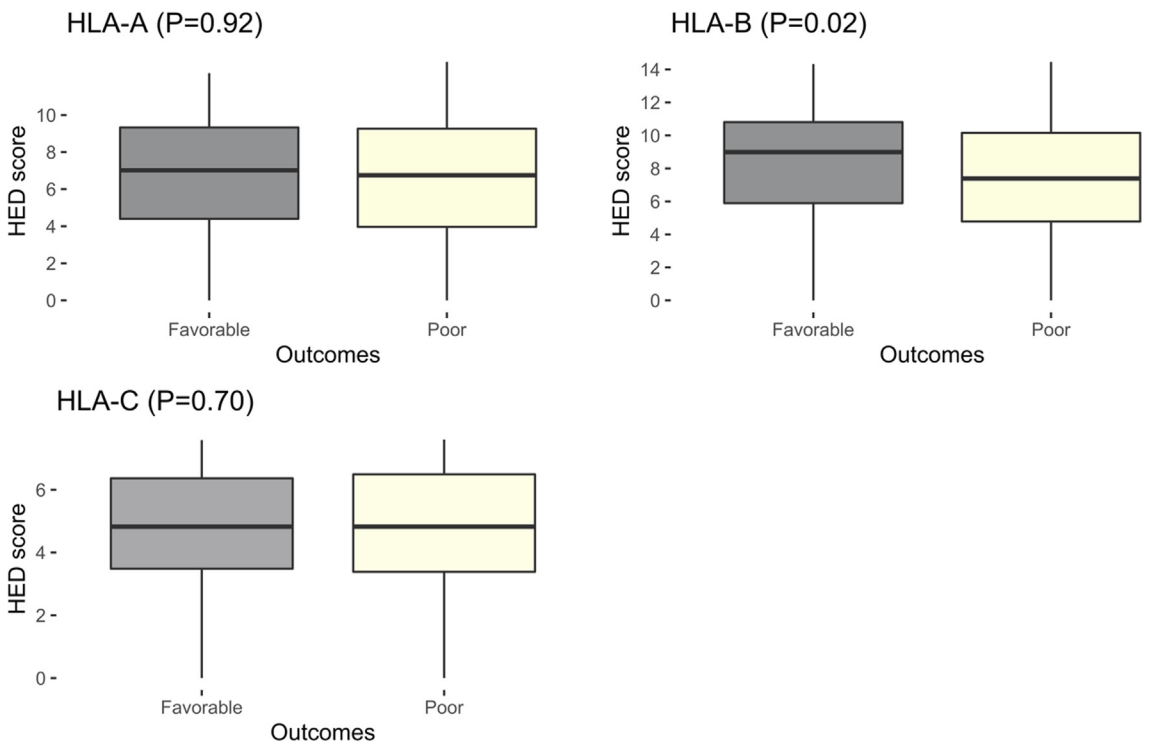

Fig. 1. Comparison of HED scores at *HLA-A*, *-B*, and *–C* loci in patients with favorable and poor outcomes after SARS-CoV-2 infection. P-values were obtained by Mann Whitney test, lines in the boxes show medians and boxes show interquatile ranges.

Table 2
HLA-A, -B, and -C HED scores for African Americans and Caucasians.

	African Americans	Caucasians	P-value
Median HLA-A HED score	6.75	7.04	0.52
Median HLA-B HED score	8.33	8.01	0.85
Median HLA-C HED score	4.82	4.82	0.20

(OR = 0.93; 95 % CI, 0.861–0.995; P = 0.04) in the univariate analysis. In the multivariate logistic regression analysis (backward model), younger age, absence of comorbidities, and higher HLA-B HED scores remained independently associated with favorable outcomes (Fig. 2).

4. Discussion

This is, to our knowledge, the first study to examine the correlation between COVID-19 and class I HED based on documented infection status and outcomes in a cohort of diverse ancestries. We found that pairwise divergences were higher at *HLA-B* locus than *HLA-A* and *-C*, and it was the only locus showing a significant correlation with COVID-19 outcomes. The combined class I HED scores were not significantly associated with outcomes. The dominant effect of the *HLA-B* locus may be consistent with previous studies suggesting that the *HLA-B* locus has undergone extensive evolutionary reassortment of variations among different alleles and, as the oldest HLA locus, harbors the highest level of polymorphisms [17]. We accounted for potential confounders in this study and demonstrated that divergence in HLA-B molecules, age, and comorbidities are independently associated with the prognosis of SARS-CoV-2 infection as a multifactorial process.

The study design of this report differs from our previous work, where we identified *HLA-DRB1*08:02* and *HLA-A*30:02* as potential risk factors for symptomatic SARS-CoV-2 infection in hospitalized Hispanics and younger African Americans, respectively, compared with their matched population controls drawn from the National Marrow Donor Program database. We also found several candidate alleles in African Americans, Caucasians, and Hispanics with protective or predisposing effects, as well as candidate amino acid residues implicated in altered peptide presentation during the immune response to SARS-CoV-2 [4]. One potential caveat of our previous work was that the infection status and disease severity of the National Marrow Donor Program controls were unknown [4]. Furthermore, risk-associated alleles and amino acid residues with small effects could not be detected due to the limited sample size. In contrast, this report focused on an inpatient cohort with carefully curated demographic, genetic, and outcome data to delineate the relationship between the overall class I HLA diversity (i.e., HED as a continuous variable) and favorable versus poor outcomes.

As class I HLA molecules are expressed on all nucleated cells and present viral peptides to cytotoxic CD8 T cells, class I HED scores represent a rational biomarker for outcomes of infection control. It has been proposed that HLA heterozygosity confers a health advantage in humans since it expands the range of pathogen-derived peptides presented to T cells, increasing the probability of mounting a specific immune response and clearing the infection [18,19]. However, the quantification of this disparity is challenging. The Grantham method provides a solution that analyzes the physicochemical features of pairwise amino acids and estimates the breadth of endogenous peptides presented by major histocompatibility complex molecules [7,19]. Using a computational antigen-binding prediction algorithm for major histocompatibility complex molecules, Pierini and Lenz evaluated the binding of peptides from a broad range of human pathogens and showed a positive correlation between genetic distance between two heterozygous alleles and the total number of peptides bound to these alleles [17,19] Furthermore, our data showed a significant impact of HED scores at *HLA-B* locus but not *HLA-A* or *-C*. While the mechanism of this finding is unknown, we speculate that *HLA-B* might play a more important role in this setting due to its higher level of polymorphism and cellular expression [20].

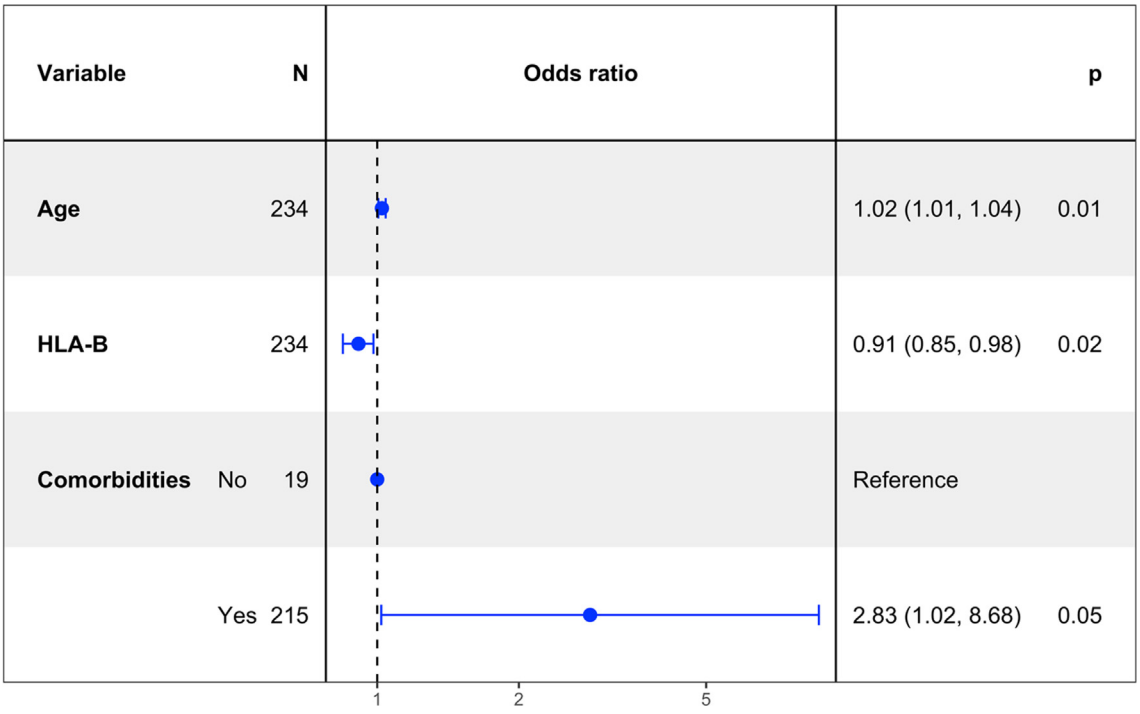

Fig. 2. Variables are independently associated with outcomes in a multivariate logistic regression analysis using the backward approach (blue dots represent medians; error bars represent 95% confidence intervals).

The literature is sparse on the role of pairwise divergence of class I HLA molecules and COVID-19 outcomes. Schetelig and colleagues did not find a significant association between the Grantham Distance of *HLA-A*, *-B*, *-C* alleles and COVID-19 severity in a large study including 6919 subjects [8]. However, several differences in the study design may account for the different findings between that study and ours. The disease severity phenotype in the Schetelig study was based on self-reported symptoms in a cross-sectional survey of a relatively healthy, hematopoetic stem cell donor registry; only 3.8 % of the participants were hospitalized, and lethal SARS-CoV-2 infections were not represented. The registry cohort was also primarily European, and the impact of comorbidities was not reported. In contrast, our study focused on hospitalized COVID-19 patients with objective outcome data, which were collected three weeks after the initial diagnosis to reflect the full disease course. Our cohort also had more diverse ancestries, with a dominance of African Americans, and more severe disease phenotypes with poor outcomes in 58.5 % of the participants. Interestingly, the modeling of viral peptide presentation by HLA class I molecules in the Schetelig study showed that HLA-B molecules might present fewer high-affinity peptides from SARS-CoV-2 than HLA-A and -C molecules [8]. This relative restriction might indicate *HLA-B* as a potential risk-associated locus in the control of SARS-CoV-2 infection.

Our study has several limitations. It is a retrospective study including two local institutions, which may limit the generalization of our findings. The sample size is relatively small and may have limited the statistical power of the analysis. Also, our study does not provide information on SARS-CoV-2 variants since August 2020, as all patients were enrolled during the initial wave of COVID-19 pandemic in the United States.

In conclusion, our study shows a significant association between increased HLA-B HED scores and favorable outcomes after SARS-CoV-2 infection. This finding suggests that maximizing the presentation of diverse SARS-CoV-2 peptides by HLA-B molecules may improve the clearance of SARS-CoV-2. Further studies are warranted to understand the functional and mechanistic implications of this finding.

Funding information

Research fund from Washington University School of Medicine Department of Pathology and Immunology (C.L.) and NIH NIAID award number R41 AI142919-01 (C.L.).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank Megan Arb of the Center for Clinical Studies of Washington University School of Medicine for helping with the data collection. We thank One Lambda Inc for providing the AllType and NGS LR kits and sequencing reagents for the Ion Chef/Ion Torrent S5 platform for this study.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humimm.2022.09.004.

References

[1] WHO COVID-19 Dashboard, World Health Organization (WHO), n.d. https://covid19.who.int/ (accessed April 8, 2022).
[2] W.J. Wiersinga, A. Rhodes, A.C. Cheng, S.J. Peacock, H.C. Prescott, Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): A review, JAMA 324 (2020) 782–793. https://doi.org/10.1001/JAMA.2020.12839.
[3] J.A.W. Gold, K.K. Wong, C.M. Szablewski, P.R. Patel, J. Rossow, J. da Silva, P. Natarajan, S.B. Morris, R.N. Fanfair, J. Rogers-Brown, B.B. Bruce, S.D. Browning, A.C. Hernandez-Romieu, N.W. Furukawa, M. Kang, M.E. Evans, N. Oosmanally, M. Tobin-D'Angelo, C. Drenzek, D.J. Murphy, J. Hollberg, J.M. Blum, R. Jansen, D.W. Wright, W.M. Sewell, J.D. Owens, B. Lefkove, F.W. Brown, D.C. Burton, T.M.

Uyeki, S.R. Bialek, B.R. Jackson, Characteristics and clinical outcomes of adult patients hospitalized with COVID-19 — Georgia, March 2020, MMWR Morb. Mortal. Wkly. Rep. 69 (18) (2020) 545–550.
[4] E. Schindler, M. Dribus, B.F. Duffy, K. Hock, C.W. Farnsworth, L. Gragert, C. Liu, HLA genetic polymorphism in patients with Coronavirus Disease 2019 in Midwestern United States, HLA 98 (2021) 370–379, https://doi.org/10.1111/tan.14387.
[5] D.G. Augusto, J.A. Hollenbach, HLA variation and antigen presentation in COVID-19 and SARS-CoV-2 infection, Curr. Opin. Immunol. 76 (2022) 102178.
[6] D. Chowell, C. Krishna, F. Pierini, V. Makarov, N.A. Rizvi, F. Kuo, L.G.T. Morris, N. Riaz, T.L. Lenz, T.A. Chan, Evolutionary divergence of HLA class I genotype impacts efficacy of cancer immunotherapy, Nat. Med. 25 (2019) 1715–1720, https://doi.org/10.1038/s41591-019-0639-4.
[7] R. Grantham, Amino acid difference formula to help explain protein evolution, Science 185 (1974) (1979) 862–864, https://doi.org/10.1126/SCIENCE.185.4154.862.
[8] J. Schetelig, F. Heidenreich, H. Baldauf, S. Trost, B. Falk, C. Hoßbach, R. Real, A. Roers, D. Lindemann, A. Dalpke, M. Kolditz, K. de With, M. Bornhäuser, E.E. Bonifacio, E. Rücker-Braun, V. Lange, J. Markert, R. Barth, J.A. Hofmann, J. Sauter, S.N. Bernas, A.H. Schmidt, Individual HLA-A, -B, -C, and -DRB1 genotypes are no major factors which determine COVID-19 severity, Front. Immunol. 12 (2021), https://doi.org/10.3389/FIMMU.2021.698193.
[9] S. Pagliuca, C. Gurnari, H. Awada, A. Kishtagari, S. Kongkiatkamon, L. Terkawi, M. Zawit, Y. Guan, T. LaFramboise, B.K. Jha, B.J. Patel, B.K. Hamilton, N.S. Majhail, S. Lundgren, S. Mustjoki, Y. Sauntharajah, V. Visconte, T.A. Chan, C.-Y. Yang, T.L. Lenz, J.P. Maciejewski, The similarity of Class II HLA genotypes defines patterns of autoreactivity in idiopathic bone marrow failure disorders, Blood. 138 (26) (2021) 2781–2798.
[10] M. Roerden, A. Nelde, J.S. Heitmann, R. Klein, H.-G. Rammensee, W.A. Bethge, J. S. Walz, HLA evolutionary divergence as a prognostic marker for AML patients undergoing allogeneic stem cell transplantation, Cancers (Basel) 12 (7) (2020) 1835.
[11] P.A. Harris, R. Taylor, R. Thielke, J. Payne, N. Gonzalez, J.G. Conde, Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support, J. Biomed. Inform. 42 (2009) 377–381, https://doi.org/10.1016/J.JBI.2008.08.010.
[12] C. Liu, B.F. Duffy, E.T. Weimer, M.C. Montgomery, J.-E. Jennemann, R. Hill, D. Phelan, L. Lay, B.A. Parikh, B. Lin, Performance of a multiplexed amplicon-based next-generation sequencing assay for HLA typing, PLoS One. 15 (4) (2020) e0232050.
[13] C. Liu, X. Yang, B.F. Duffy, J. Hoisington-Lopez, M.L. Crosby, R. Porche-Sorbet, K. Saito, R. Berry, V. Swamidass, R.D. Mitra, High-resolution HLA typing by long reads from the R10.3 Oxford nanopore flow cells, Hum. Immunol. 82 (2021) 288–295, https://doi.org/10.1016/J.HUMIMM.2021.02.005.
[14] C. Liu, A long road/read to rapid high-resolution HLA typing: The nanopore perspective, Hum Immunol. 82 (2021) 488–495, https://doi.org/10.1016/J.HUMIMM.2020.04.009.
[15] R. Bentham, HLAdivR, n.d. https://rdrr.io/github/rbentham/HLAdivR/ (accessed February 26, 2022).
[16] Y. Fan, Y.Q. Song, PyHLA: tests for the association between HLA alleles and diseases, BMC Bioinform. 18 (2017), https://doi.org/10.1186/S12859-017-1496-0.
[17] L.M. McKenzie, J. Pecon-Slattery, M. Carrington, S.J. O'Brien, Taxonomic hierarchy of HLA class I allele sequences, Genes Immun. 1 (1999) 120–129, https://doi.org/10.1038/SJ.GENE.6363648.
[18] P.C. Doherty, R.M. Zinkernagel, Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex, Nature. 256 (1975) 50–52, https://doi.org/10.1038/256050a0.
[19] F. Pierini, T.L. Lenz, C. Wilke, Divergent allele advantage at human MHC genes: signatures of past and ongoing selection, Mol. Biol. Evolut. 35 (9) (2018) 2145–2158.
[20] M.C. Montgomery, C. Liu, R. Petraroia, E.T. Weimer, Using nanopore whole-transcriptome sequencing for human leukocyte antigen genotyping and correlating donor human leukocyte antigen expression with flow cytometric crossmatch results, J. Mol. Diagn. 22 (2020) 101–110, https://doi.org/10.1016/j.jmoldx.2019.09.005.