mebm
MOLECULAR AND EXPERIMENTAL BIOLOGY IN MEDICINE

<u>REVIEW ARTICLE</u>

A REVIEW OF HLA AND COVID-19 ASSOCIATION STUDIES

Renata Zunec

Abstract: Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is reported to vary across different populations in the prevalence of infection, in the death rate of patients, in the severity of symptoms and in the drug response of patients. Among host genetic factors that can influence all these attributes human leukocyte antigen (HLA) genetic system stands out as one of the leading candidates. Case-control studies, large-scale population-based studies, as well as experimental bioinformatics studies are of utmost importance to confirm HLA susceptibility spectrum of COVID-19. This review presents the results of the first case-control and epidemiological studies performed in several populations, early after the pandemic breakout. The results are pointing to several susceptible and protective HLA alleles and haplotypes associations with COVID-19, some of which might be of interest for the future studies in Croatia, due to its common presence in the population. However, further multiple investigations from around the world, as numerous as possible, are needed to confirm or deteriorate these preliminary results.

Tissue Typing Centre, Clinical Department of Transfusion Medicine and Transplantation Biology, University Hospital Centre Zagreb

Corresponding author:
Renata Zunec
Tissue Typing Centre, Clinical Department of Transfusion Medicine and Transplantation Biology, University Hospital Centre Zagreb
Kispaticeva 12, 10000 Zagreb, Croatia
e-mail: rzunec@kbc-zagreb.hr

INTRODUCTION

Coronavirus disease 2019 (COVID-19), an infectious viral disease caused by Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2), pandemic is now affecting all countries of the world, with more than 50 000 000 people infected and with more than 1 000 000 deaths, as of the 15th of November 2020. The incidence, morbidity and mortality data for each country can be followed at the daily updated web page https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6 (Coronavirus COVID-19 global cases by the centre for systems science and engineering at Johns Hopkins University).

As for any other viral disease, susceptibility or resilience to SARS-CoV-2 infection arises through the complex interaction of comorbid statuses, environment backgrounds and host genetic factors.

The host genetic factors have influence on disease susceptibility and disease severity, can help identify individuals who are at risk for disease, might affect pharmacokinetics and pharmacodynamics of drugs used in COVID-19 patients, can help altering the treatment regimen for the patients and thus improve the overall treatment outcome.

ROLE OF HLA SYSTEM

The human major histocompatibility complex, the human leukocyte antigen (HLA), stands out as one of the leading genetic factor candidates that might play an eminent role in COVID-19 susceptibility, severity, and outcomes as well as in geographical variation.

HLA system is a genomic region in the chromosomal position 6p21 composed of at least 132 protein coding genes and the similar number of non-coding genes that have important roles in the regulation of the immune system. Evolution of its structural organization involved various mutation, duplication, deletion and genomic

rearrangement events over the period of 500 million years.[1]

The HLA system spans approximately 4,200 kilobases of DNA. The HLA genes, which primary role is to encode diverse antigen-presenting molecules, are divided into three classes, class I, II and III. The class I region, includes 19 gene loci among which three are classical loci HLA-A, HLA-B and HLA-C, three non-classical and twelve non-coding genes or pseudogenes. The class II region includes the classical gene loci HLA-DP, HLA-DQ, and HLA-DR, the non-classical HLA-DO and HLA-DM loci and a group of genes outside the MHC whose functions are related to antigen presentation: large multifunctional protease (LMP) and transporter associated with antigen processing (TAP) genes. HLA-DR loci involves a single DRA gene and up to nine DRB genes (DRB1 to DRB9). The DRA gene encodes an invariable α chain which binds various β chains encoded by the DRB genes. HLA-DR molecules are determined by the polymorphic DRβ1 chains encoded by DRB1 alleles. The HLA-DP and -DQ loci each have one expressed gene for α and β chains and additional unexpressed pseudogenes. The DQA1 and DQB1 gene products associate to form DQ molecules, and the DPA1 and DPB1 gene products form DP molecules.

The HLA class III region lies between the class I and class II genes and contains genes with varying functions, among which are the genes for the complement proteins C4A, C4B, C2 and factor B and genes for tumor necrosis factor (TNF) and heatshock protein 70 (HSP70).[2]

Alongside its key role in the transplantation medicine, HLA is likewise of great interest to the clinical medicine, as an important factor for understanding the pathogenesis of various diseases, from autoimmune diseases to bacterial and viral infection diseases. Namely, an effective host response against all viruses depend on HLA-restricted T-lymphocyte responses.

HLA molecules, encoded by HLA class I and HLA class II genes, are cell-surface glycoproteins that present intracellular and extracellular peptides to T cells. In the case of viral infection, viral epitopes are presented by dendritic cells to CD8+ T lymphocytes through HLA class I molecules and through HLA class II molecules to CD4+ T lymphocytes (Figure 1). Presentation through HLA class I molecules primes to the clonal expansion of HLA-restricted CD8+ cytotoxic T lymphocytes (CTLs). Efficacious antiviral defense during acute infection is mainly dependent on CTLs response while memory CTLs are involved in the immune response to latent reinfection and reactivation.[3]
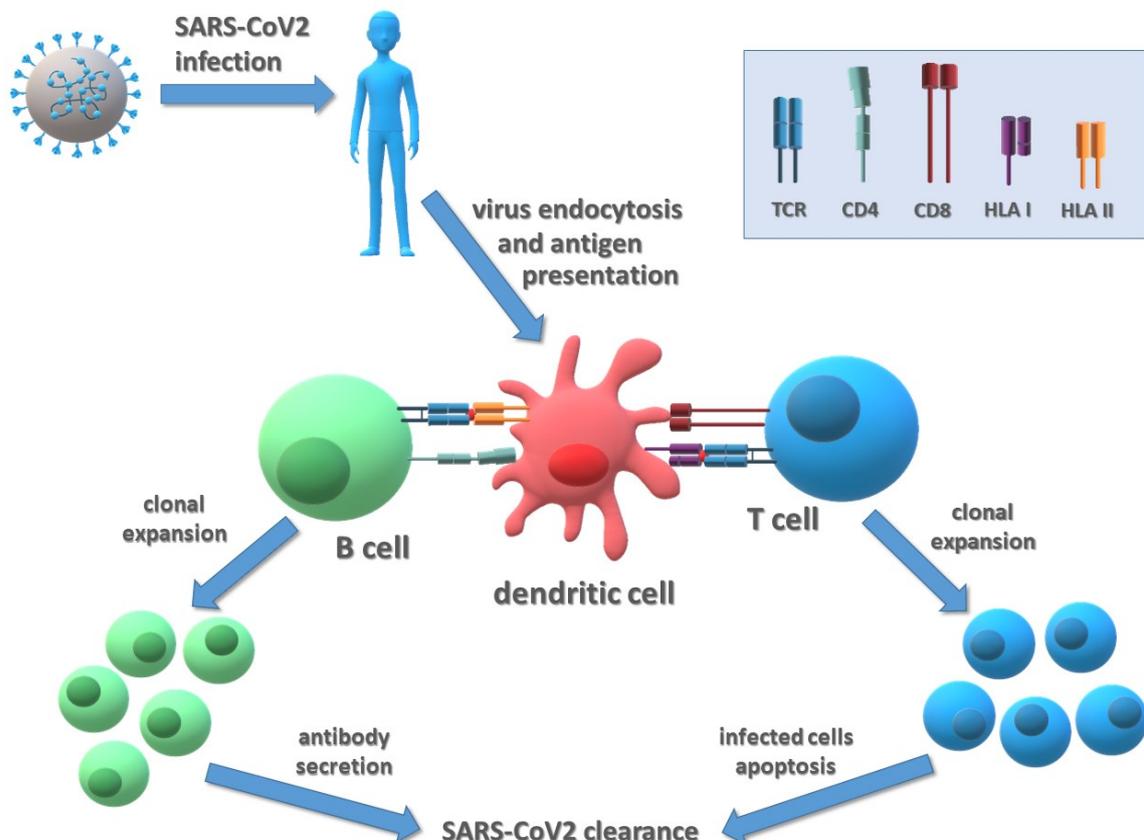

Figure 1. HLA molecules in presentation of viral epitopes to CD4+ and CD8+ lymphocytes

CD4+ T lymphocytes enhance CTLs responses and provide T-cell help to the generation of specific antiviral antibodies. Consequently, through their central role in viral antigen presentation HLA molecules act as a marker for genetic susceptibility to infectious disease.[4]

Thus, since the outbreak of (COVID-19), HLA genes came into the focus as a potential marker for the disease susceptibility and severity. HLA region is the most polymorphic genetic system in the human genome, which numbered more than 28 000 class I and class II alleles included in the IPD-IMGT/HLA Database at the end of October 2020.[5, 6]

In order to emphasize this extensive level of polymorphism, the term hyper-polymorphic, instead of simple polymorphic, has been introduced.[7]

The main advantage of such hyper-polymorphism is the large potential of the heterozygosity in the population. In comparison to homozygous individuals, heterozygous individuals are capable of presenting a larger variety of peptides which leads to the better protection against pathogens and consequently to the better capability to reproduce and survive. Indeed, worldwide studies of HLA allele frequencies distribution revealed around 90% heterozygosity at each HLA locus.[8, 9]

Conversely to the evolutionary advantage, HLA hyper-polymorphism is the main obstacle in disease association studies. Among thousands of HLA alleles, limited number of them are very common and common in the populations, whereas the vast majority belongs to the group of rare and vary rare alleles. Moreover, other HLA genetic characteristics, such as linkage disequilibrium which denotes that certain alleles occur together with a greater frequency than would be expected by chance and limitations in sample-size of patients and healthy controls which can lead to in-sufficient power of statistical tests, are making disease association studies very puzzling in determining whether an allele or multiple alleles confer risk or protection to the pathogen.

CASE-CONTROL STUDIES

Although the outbreak of the SARS-CoV-2 pandemic now lasts for less than a year, several association studies have already been published, pointing to the possible HLA markers conferring either risk or protection to COVID-19.

Since COVID-19 first broke out in Wuhan, Hubei province, China in late December 2019, the first study came, as expectedly, from China. Wang *et al* studied 11 HLA loci allele frequencies in 82 confirmed COVID-19 Han patients from Zhejiang by next generation sequencing (NGS). Two alleles, HLA-C*07:29 and B*15:27 were more frequent in patients and the difference to the control group was found significant after P-value correction. Several other alleles were found either with increased (B*40:06, C*08:01G, DRB1*04:06, DPB1*36:01) or decreased (DRB1*12:02 and DPB1*04:01) frequency in patients but the difference did not remain significant after the correction.[10]

Another paper published on HLA genotypes in Chinese patients from Wuhan is a work by Warren and Birol who instead of performing the HLA typing, derived HLA types directly from metatranscriptomic RNA-Seq libraries using HLAminer. Libraries were prepared from bronchoalveolar lavage fluid and peripheral blood mononuclear cell samples of 8 COVID-19 patients at the early stage of the outbreak in Wuhan, China. The results indicated the presence of HLA-A*24:02 allele and class II DPA1*02:02-DPB1*05:01 haplotype in four and seven out of eight patients, respectively.[11]

In Spain, country which was also highly affected during the first wave of COVID-19 pandemic, Iturrieta-Zuazo *et al* performed the pilot study aiming to analyze the association between HLA and the severity of the COVID-19. The patient group, which included a total of 45 Spanish patients from Madrid with mild (N=5), moderate (N=20) and severe (N=29) SARS-CoV-2 infection, was typed for HLA-A, -B and -C gene polymorphisms by PCR Sequence Specific Oligonucleotide reverse (SSO) method through Luminex technology. Allele frequencies of the Spanish healthy population were obtained from Allele Frequency Net Database using for each allele the average of the frequencies in all Spanish studies uploaded on this website. The analyses of HLA class I gene/allele frequencies was performed at the level of HLA supertypes.

The results showed that the HLA alleles belonging to A2 supertype (A*02:01, A*02:05, A*68:01, A*68:02) and to C1 supertype (C*01:02, C*03:03, C*03:04, C*07:01, C*07:02, C*08:01, C*08:02, C*12:02, C*12:03, C*16:01) were present with decreased frequencies in a severe group compared to moderate group and healthy controls, but without statistical significance. The finding was in concordance with the theoretical model of higher degree of recognition of SARS CoV-2 peptides for these two supertypes. Furthermore, the authors reported the association of the severe course of the disease with a higher percentage of homozygosity at HLA-A and –C locus in comparison with moderate one (20% vs. 10% in locus A, 15% vs. 5% in locus C) but no statistically significant differences were obtained. Interestingly, for locus HLA-B homozygosity was 5%, in both groups, and no homozygous was found among 5 mild patients. Based on these results, the authors suggest that the HLA heterozygosity and HLA molecules with a higher theoretical capacity for binding SARS-Cov-2 peptides are more present in patients with mild evolution of the COVID-19 than in patients with moderate and severe ones.[12]

The first study on COVID-19 and HLA association in Italy was published by Novelli *et al* in August 2020. They investigated HLA allele frequency distribution at 11 loci (HLA-A, B, C, DRB1, DRB3/4/5, DQA1, DQB1, DPA1, DPB1) typed by NGS method in a group of 99 Italian patients affected by a severe or extremely

Table 1. SARS-CoV-2 and HLA case-control association studies

Reference (First Author) Country	N	HLA-	Results (frequency increased/decreased; significance)	Croatian population[a]
10 (Wang) China	82	C*07:29	increased; significant after p value correction	0.0000
		B*15:27		0.0000
		B*40:06	increased; significance lost after correction	0.0008
		C*08:01G		0.0006
		DRB1*04:06		0.0003
		DPB1*36:01		0.0000[b]
		DRB1*12:02	decreased; significance lost after correction	0.0002
		DPB1*04:01		0.3669[b]
11 (Warren) China, Wuhan	8	A*24:02	positive in 4 patients	0.1146
		DPA1*02:02-DPB1*05:01 haplotype	positive in 7 patients	1-5% [b]
12 (Iturrieta-Zuazoa) Spain	45	A2 supertype	decreased in patients with severe form without statistical significance	0.3026
		C1 supertype		0.0483
		homozygosity at HLA-A locus	increased in patients with severe form without statistical significance	0.0708
		homozygosity at HLA-C locus		0.0436
13 (Novelli) Italy	99	B*27:07	increased; significant after p value correction	0.0000
		DRB1*15:01		0.0875
		DQB1*06:02		0.1100 [c]
		B*58:01	increased; significance lost after correction	0.0116
		C*06:02	decreased; significance lost after correction	0.0858
		DRB1*07:01		0.0896
15 (Correale) Italy	NA	B*44	positive and independent association with incidence; no association with morbity and mortality	0.0973
		C*01		0.0483
		A*25	positive association with incidence; no association with morbity and mortality	0.0287
		B*08		0.0806
		B*15:01		0.0405
		B*51		0.1063
		C*03		0.0752
		B*14	negative association with incidence; no association with morbity and mortality	0.0271
		B*18		0.0867
		B*49		0.0174
13 (Pissanti) Italy	NA	A*01:01g~B*08:01g~C*07:01g~DRB1*03:01g	positive association with incidence and mortality	0.0521
		A*02:01g~B*18:01g~C*07:01g~DRB1*11:04g	negative association with incidence and mortality	0.0157
16 (Lue) USA, New York	46[#]	A and -B alleles	no association	/
		B*44 supertype	increased in patients with severe form without statistical significance	0.0973

Legend: N - number of COVID-19 diagnosed patients included in the study; a - corresponding frequencies in the Croatian population are provided in order to show possible interest for the future association studies in Croatia, data obtained from reference 21 (number of individuals included in study n=10000), b - data obtained from reference 22 (n=124); c - data obtained from reference 23 (n=210); NA - not applicable, epidemiological study; # - lung cancer patients with COVID-19.

which lost the significance after the correction as well as the decreased frequency of C*06:02 and DRB1*07:01.[13] Two other Italian groups used geographical ecological approach to investigate the HLA and COVID-19 association. The epidemiological study was performed severe form of COVID-19. Control group consisted of 1017 individuals. Increased frequencies in patients compared to controls which remained significant after the correction were found for HLA-B*27:07, -DRB1*15:01 and -DQB1*06:02 alleles. The analysis

also revealed increased frequency found for B*58:01 by assessing the degree of correlation between the incidence and mortality of COVID-19 and the prevalence of HLA alleles in a healthy population. Both groups used HLA allele and haplotype frequencies from the database of the Italian Bone Marrow Donor Registry and compared the regional differences with the COVID-19 incidence and mortality. Correale *et al* analyzed the role of HLA class I alleles (HLA-A, -B, -C) while Pisanti *et al* examined the role of the two most frequent HLA haplotypes in the Italian population. The results showed that HLA-B*44 and/or C*01, alleles are risk factors for SARS-CoV-2 infection, but the correlation with the morbidity and mortality was not observed. The positive correlation with the disease incidence was also observed for HLA-A*25, HLA-B*08 B*15:01, B*51, and C*03, as well as negative one for HLA-B*14, B*18, and B*49 alleles, but, after examination with multiple regression model, did not maintain independent and positive association with COVID-19 incidence. Number one haplotype by the frequency in the Italian population, HLA-A*01:01g~B*08:01g~C*07:01g~DRB1*03:01g, showed the positive correlation, while conversely, HLA-A*02:01g~B*18:01g~C*07:01g~DRB1*11:04g, which is the second most common haplotype, showed a negative correlation with COVID-19 incidence and mortality.[14, 15]

Single center study performed by Luo *et al* at Memorial Sloan Kettering Cancer Center, New York, examined patient-specific and cancer-specific features that impact the severity of COVID-19 in lung cancer patients. One of the factor studied was the association of HLA-A and HLA-B alleles in a cohort of 46 patients, 29 with mild disease and 17 with severe disease compared with the control group of 5166 patients with lung cancers and no known COVID-19. The only difference observed was for HLA-B44 supertype which was more frequent among patients with severe COVID-19 but without significance compared with mild disease group and controls.[16]

CONCLUSIONS AND FUTURE DIRECTIONS

The main limitation of all case-control studies presented in this review is small sample size of the patient groups with consequently questionable power of the statistical analysis.

Moreover, the majority of HLA alleles found to be protective or susceptibility markers for COVID-19 in studied population so far, belong to rare or very rare HLA polymorphisms in other geographic regions of the world reflecting varying frequencies of HLA alleles and haplotypes across populations. The Croatian population has been extensively studied for the HLA allele and haplotype polymorphisms[17] providing the database for further COVID-19 association studies in the Croatian cohort of patients. The frequencies of so far reported COVID-19 associated HLA specificities in the studies presented in this review, together with their frequencies

in the Croatian population, are presented in Table 1. Results show that several alleles reported as COVID-19 associated factors are represented in the Croatian population as common alleles and thus are of interest for further investigations, as well as it is the relationship between COVID-19 prevalence, severity and mortality with the observed regional differences of the HLA allele and haplotype distribution in Croatia (Grubic, *submitted for publication*). In any case, studies from other populations, as numerous as possible, are needed as an effort to take the global population stratification into account.

Meanwhile, together with classical HLA disease association population studies, another kind of approach developed. These are the bioinformatics and experimental studies[18, 19, 20] which so far identified and provided shortlist of SARS-CoV-2 derived peptides to bind several HLA class I and HLA class II molecules (HLA-A*02:01, HLA-A*01:01, HLA-B*07:02, HLA-C*07:02, HLA-B*40:01, HLA-DRA*01:01, HLA-DRB1*07:01 and HLA-DRB1*04:01), with some of them already confirmed in case-control studies (e.g. A2 supertype). However, the immunogenicity of these epitopes remains to be tested with the integration of HLA population studies and, if successful, can, amongst other evidence, provide crucial information for vaccine design.

Further immunogenetic studies are expected to solve the HLA part of host genetic puzzle of SARS-CoV-2 pandemic and to bring benefit to individual patients, to the scientific community and to the whole society.

REFERENCES

1. Kulski J K, Shiina T, Anzai T, Kohara S, Inoko H. Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol. Rev. 2002;190:95-122.
2. Shiina T, Hosomichi K, Inoko H, Kulski JK. The HLA genomic loci map: expression, interaction, diversity and disease. J Hum Genet. 2009;54:15-39.
3. van Lier RA, ten Berge IJ, Gamadia LE. Human CD8+ T-cell differentiation in response to viruses. Nat. Rev. Immunol. 2003;3:931-939.
4. Blackwell JM, Jamieson SE, Burgner D. HLA and infectious diseases.Clin Microbiol Rev. 2009;22:370-385.
5. Nomenclature for Factors of the HLA System July, August and September 2020 Compiled by Steven G. E. Marsh for the WHO Nomenclature Committee for Factors of the HLA System. Published In: HLA (2020): *in press*
6. http://hla.alleles.org/nomenclature/stats.html
7. 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.org/10.1093/nar/gkz950
8. Solberg OD, Mack SJ, Lancaster AK, Single RM, Tsai Y, Sanchez-Mazas A, Thomson G. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. Hum Immunol. 2008;69(7):443-64. dx.doi.org/10.1016/j.humimm.2008.05.001
9. Lipsitch M, Bergstrom CT, Antia R. Effect of human leukocyte antigen heterozygosity on infectious disease outcome: the need for allele-specific measures. BMC Med Genet. 2003;4(1):2. dx.doi.org/10.1186/1471-2350-4-2

10. 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;96:194-196. doi.org/10.1111/tan.13941

11. Warren RL, Birol I. HLA predictions from the bronchoalveolar lavage fluid and blood samples of eight COVID-19 patients at the pandemic onset. Bioinformatics, 2020;1-3. doi.org/10.1093/bioinformatics/btaa756

12. Iturrieta-Zuazoa I, Ritaa CG, García-Soidána A, de Malet Pintos-Fonsecab A, Alonso-Alarcóna N, Pariente-Rodrígueza R, Tejeda-Velardea A, Serrano-Villarc S, Castañer-Alabaua JL, Nieto-Gañána I. 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. 2020;219: 108572 doi.org/10.1016/j.clim.2020.108572

13. Novelli A, Andreani M, Biancolella M, Liberatoscioli L, Passarelli C, Colona VL, Rogliani P, Leonardis F, Campana A, Carsetti R, Andreoni M, Bernardini S, Novelli G, Locatelli F. HLA allele frequencies and susceptibility to COVID-19 in a group of 99 Italian patients. HLA. 2020;96:610-614. doi.org/ 10.1111/tan.14047

14. Pisanti S, Deelen J, Gallina AM, Caputo M, Citro M, Abate M, Sacchi N, Vecchione C, Martinelli R. Correlation of the two most frequent HLA haplotypes in the Italian population to the differential regional incidence of Covid-19. J Transl Med. 2020;18:352.

15. Correale P, Mutti L, Pentimalli F, Baglio G, Saladino RE, Sileri P, Giordano A. HLA-B*44 and C*01 Prevalence Correlates with Covid19 Spreading across Italy. Int J Mol Sci. 2020;21:5205. doi.org/10.3390/ijms21155205

16. Luo J, Rizvi H, Preeshagul IR, Egger JV, Hoyos D, Bandlamudi C, McCarthy CG, Falcon CJ, Schoenfeld AJ, Arbour KC, Chaft JE, Daly RM, Drilon A, Eng J, Iqbal A, Lai WV, Li BT, Lito P, Namakydoust A, Ng K, Offin M, Paik PK, Riely GJ, Rudin CM, Yu HA, Zauderer MG, Donoghue MTA, Łuksza M, Greenbaum BD, Kris MG, Hellmann MD. COVID-19 in patients with lung cancer Ann Oncol. 2020;31(10)1386-1396. doi.org/10.1016/j.annonc.2020.06.007

17. Grubic Z. HLA allele and haplotype diversity in the Croatian population: State of the art. HLA. 2018;92(Suppl 2):51-56. doi.org/10.1111/tan.13401.

18. Nguyen A, David JK, Maden SK, Wood MA, Weeder BR, Nellore A, Thompson RF. Human leukocyte antigen susceptibility map for SARS-CoV-2. J Virol. 2020;94(13):e00510-20. doi.org/10.1128/JVI.00510-20.

19. Lin M, Tseng H-K, Trejaut JA, Lee H-L, Loo J-H, Chu C-C, Chen P-J, Su Y-W, Lim KH, Tsai Z-U, Lin R-Y, Lin R-S, Huang C-H. Association of HLA class I with severe acute respiratory syndrome coronavirus infection. BMC Med Genet. 2003;4:9. doi.org/ 10.1186/1471-2350-4-9.

20. Sanchez-Mazas A. HLA studies in the context of coronavirus outbreaks. Swiss Med Wkly. 2020;150:w20248. doi.org/10.4414/smw.2020.20248.

21. Maskalan M. HLA allele and haplotype diversity in Croatia and its impact on donor selection in hematopoietic stem cell transplantation. PhD thesis 2020. (available at: https://dr.nsk.hr/islandora/object/pmf%3A8463)

22. Maskalan M. Constructing the HLA-DPB1 and DPA1 polymorphism picture in European populations- the Croatian piece of puzzle. EWIC 2020, poster presentation (P9).

23. Grubic Z, Zunec R, Cecuk-Jelicic E, Kerhin-Brkljacic V, Kastelan A. Polymorphism of HLA-A, -B, -DRB1, -DQA1 and -DQB1 haplotypes in a Croatian population. Eur J Immunogenet 2000;27(1):47-51. doi.org/10.1046/j.1365-2370.2000.00193.x