

The Association of Human Leukocyte Antigen and COVID-19 in Southern China

Xueping Yu,^{1,a} Kuoting Ho,^{2,5,a} Zhongliang Shen,^{3,a} Xiaoying Fu,² Hongbo Huang,⁴ Delun Wu,² Yancheng Lin,² Yijian Lin,⁴ Wenhua Chen,¹ Milong Su,⁶ Chao Qiu,³ Xibin Zhuang,⁴ and Zhijun Su¹

¹Department of Infectious Disease, The First Hospital of Quanzhou, affiliated to Fujian Medical University, Quanzhou, Fujian, China, ²HL. Q Biomedical Laboratory, Quanzhou, Fujian, China, ³Department of Infectious Disease, Huashan Hospital, Fudan University, Shanghai, China, ⁴Department of Respiratory Disease, The First Hospital of Quanzhou, affiliated to Fujian Medical University, Quanzhou, Fujian, China, ⁵School of Biomedical Science, Huaqiao University, Quanzhou, Fujian, China, and ⁶Department of Clinical Laboratory, The First Hospital of Quanzhou, affiliated to Fujian Medical University, Quanzhou, Fujian, China

Human leukocyte antigen (HLA) polymorphism is hypothesized to be associated with diverse immune responses toward infectious diseases. Herein, by comparing against multiple subpopulation groups as control, we confirmed that HLA-B*15:27 and HLA-DRB1*04:06 were associated with coronavirus disease 2019 susceptibility in China. Both alleles were predicted to have weak binding affinities toward viral proteins.

Keywords. COVID-19; HLA; population stratification; SARS-CoV-2; susceptibility.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has contributed to nearly 4 million deaths worldwide [1]. Delayed or inadequate treatment contributes to a large portion of coronavirus disease 2019 (COVID-19) mortality. Improving risk stratification and clinical management is crucial to identify the potentially vulnerable population. Several clinical features and genetic factors have been suggested to escalate risks of SARS-CoV-2 infection, the occurrence of severe conditions, and mortality [2].

Human leukocyte antigen (HLA), the most polymorphic genetic locus, plays a pivotal role in viral antigen recognition and presentation [3]. After decades of study, several HLA alleles have been identified to be responsible for individual differences in host response, such as viral clearance and disease prognosis

[4–8]. Thus far, a few HLA alleles have been reported to be predisposed to COVID-19 infection [9–12]. However, these findings are difficult to replicate [13]. Some reasons for this are small sample size and HLA being population-dependent, but the leading cause is that these case-control studies are extra sensitive to population stratification. A previous study pointed out that, even within the same ethnic group, the choice of controls would drastically affect the perceived associations [9]. On top of that, there are complicated issues specific to COVID-19 studies. In the case of investigating patients recruited during an early stage of the pandemic in Southern China, it may be inappropriate to use the local population as control, as most cases were made up of people who fled from Wuhan. To better elucidate these situations, the use of multiple control groups has been suggested [14].

Herein, we performed an empirical study in 399 unrelated patients to explore the effect of HLA on SARS-CoV-2 infection. We highlighted the influence of subpopulation by comparing it against multiple controls. Our findings provide deeper insights into the pathogenetic mechanism of COVID-19 and lay valuable groundwork for future HLA research.

METHODS

Study Subjects

Patients with confirmed COVID-19 diagnosis in Quanzhou, Fujian (n = 42), during early 2020 were enrolled. Their demographic information, clinical characteristics, and peripheral blood samples were collected with informed consent. The HLA alleles were identified via next generation sequencing-based typing using NanoWES Human Exome, version 1.0 (Berrygenomics, Beijing, China), according to the manufacturer's instructions. Class I and class II HLA alleles were estimated with OptiType, version 1.3.3 [15], and ATHLATES, version 1.0 [16], respectively. HLA allele frequencies and clinical information for patients in the other 2 centers (Shenzhen, n = 332, and Zhejiang, n = 82) were retrieved from prior studies [9, 17].

Choice of Control Populations

Aiming to make our findings more plausible, we performed analyses against multiple sets of controls. First, we chose 3 controls representing the general population in China, including (1) individuals from the China Marrow Donor Program (CMDP), recruited during 2010–2012 (n = 169 995) [18], referred to as General 1; (2) individuals from the CMDP, recruited during 2010–2015 (n = 812 211) [19], referred to as General 2, and (3) individuals from the deep sequencing MHC region study who were enrolled independent of the CMDP (n = 10 689) [20],

Received 12 March 2021; editorial decision 28 July 2021; accepted 6 August 2021.

^aEqual contribution

Correspondence: Zhijun Su, MD, Department of Infection Disease, The First Hospital of Quanzhou affiliated to Fujian Medical University, No. 250 East Street, Licheng District, Quanzhou, Fujian 362000, China (su2366@sina.com).

Open Forum Infectious Diseases® 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com <https://doi.org/10.1093/ofid/ofab410>

referred to as General 3. To rule out associations caused by proportionate representation of the Southern China population, we again selected 2 controls resembling the population in Southern China, including (1) individuals from the Hong Kong Bone Marrow Donor Registry retrieved from the Allele Frequency Net Database (AFND; $n = 5266$), referred to as South 1, and (2) individuals from the CMDP recruited in Shenzhen ($n = 4262$) [21], referred to as South 2. HLA typing for HLA-DQA1 and HLA-DPB1 was limited and only available in General 3 and South 1.

Prediction of HLA Binding Affinities

As suggested by Nguyen et al. [22], binding affinities for all SARS-CoV-2 viral peptides were predicted with NetMHCpan-4.1 (for HLA class I) and NetMHCIpan-4.0 (for HLA class II) [23]. Currently, only DRB1 alleles are available in NetMHCIpan-4.0. The publicly available SARS-CoV-2 proteome sequence Wuhan-Hu-1 (NC_045512.2) was used as the reference. To inspect whether the binding pattern of certain viral proteins would have a special impact on viral vulnerability, we grouped peptides with open reading frames and highlighted binding capabilities of major viral proteins such as spike protein (S), nucleocapsid protein (N), membrane protein (M), and envelope protein (E). More precisely, we assessed the binding affinities from peptides derived from the receptor binding domain (RBD) region. We later labeled these bindings strong binder ($IC_{50} < 50$ nM) and weak binder ($IC_{50} < 500$ nM), following the recommendations of previous studies [24, 25]. The numbers of strong binders and weak binders were calculated per HLA allele.

Statistical Analysis

Before analyses, several quality control procedures were performed. Closely related individuals were excluded through kinship estimation in PLINK, version 1.9 [26]. Hardy-Weinberg equilibrium (HWE) was assessed in Arlequin, version 3.5 [27]. Cochran's rule was applied to eliminate HLA alleles with a frequency < 5 . The following statistical analyses were performed in R, version 4.0.2: HLA heterogeneity among 3 centers was measured by pairwise comparisons using the Fisher exact test.

The homogeneities among each control group were examined using the chi-square test. In addition, a t test was performed to compare the allele frequency between controls representing the general and Southern populations. Subsequently, to estimate the effect of HLA on disease susceptibility, we utilized the Fisher exact test to compare HLA allele frequencies between COVID-19 patients and selected controls. The corrected P value (P_c) was then calculated using the Benjamini-Hochberg method to adjust for multiple comparisons.

RESULTS

A total of 399 unrelated individuals with confirmed COVID-19 diagnoses were included in the present study. No departure from Hardy-Weinberg equilibrium was observed when multiple comparisons were considered. There were 14 HLA-A, 26 HLA-B, 13 HLA-C, 13 HLA-DPB1, 14 HLA-DQA1, 15 HLA-DQB1, and 22 HLA-DRB1 alleles that met the criteria for analyses. No significant HLA distribution differences were detected among the 3 centers.

Couples of alleles appeared to have significant heterogeneity among intragroup controls (full list presented in [Supplementary Tables 1 and 2](#)). Besides, 10 alleles showed significant allele frequency differences between the general and Southern populations, as detailed in [Table 1](#). The full results of the between-group comparison are appended in [Supplementary Table 3](#). These findings suggest that extra caution is needed when conducting studies involving such significant heterogeneity alleles among the population.

We estimated the relationship between HLA and COVID-19 susceptibility by comparing cases against 3 controls representing the general population in China. We determined an HLA allele to have strong evidence for the association if the associations were observed to be significant against 2 or more controls. As illustrated in [Table 2](#), 5 alleles were identified as risk alleles for COVID-19 vulnerability. These alleles are HLA-A*11:01, HLA-A*11:02, HLA-B*15:27, HLA-B*40:01, and HLA-DRB1*04:06. Next, we performed additional comparisons against 2 controls reflecting the Southern China population. Only HLA-B*15:27 and HLA-DRB1*04:06 remained significant. We noticed that HLA-B*15:27 was 1 of the alleles previously identified to vary between the general

Table 1. The Allele Frequency Difference Between the General Population and the Southern Population

Allele	General 1 AF	General 2 AF	General 3 AF	South 1 AF	South 2 AF	Intra-General P Value	Intra-South P Value	Between-Group P Value
A*02:06	0.052	0.052	0.057	0.041	0.037	.010	.163	.023
A*02:07	0.085	0.084	0.073	0.119	0.113	1.70E-08	.205	.006
B*15:18	0.013	0.014	0.016	0.003	0.005	.002	.106	.004
B*15:27	0.008	0.009	0.009	0.007	0.007	.022	1.000	.037
B*46:01	0.102	0.103	0.078	0.146	0.134	3.89E-30	.019	.020
B*58:01	0.059	0.061	0.049	0.088	0.080	1.60E-20	.052	.022
C*01:03	0.006	0.006	0.007	0.004	0.003	.337	.372	.034
C*15:02	0.034	0.033	0.037	0.028	0.026	.004	.385	.017
DQB1*02:01	0.049	0.050	0.042	0.067	0.061	6.03E-07	.118	.031
DQB1*06:01	0.102	0.102	0.094	0.113	0.109	1.35E-04	.430	.042

Abbreviation: AF, allele frequency.

Table 2. The Association Between HLA and COVID-19 Susceptibility

Allele	No. (Case)	Frequency in Case, %	No. (Control)	Frequency in Control, %	Odds Ratio	P Value	P _c Value	Control Label
A*11:01	214	26.8	71 884	21.1	1.367	1.32E-04	.002*	General 1
			339 270	20.9	1.388	6.00E-05	8.39E-04*	General 2
			3970	18.6	1.607	2.15E-08	3.01E-07*	General 3
			3144	29.9	0.861	.071		South 1
			2216	26.0	1.043	.613		South 2
A*11:02	27	3.4	6517	1.9	1.792	6.01E-03	.084	General 1
			27 495	1.7	2.034	8.37E-04	.012*	General 2
			351	1.6	2.098	7.08E-04	.010*	General 3
			425	4.0	0.833	.399		South 1
			247	2.9	1.173	.442		South 2
B*15:27	16	2.2	2752	0.8	2.738	4.26E-04	.011*	General 1
			13 895	0.9	2.590	7.48E-04	.019*	General 2
			191	0.9	2.479	2.42E-03	.063	General 3
			77	0.7	3.034	2.93E-04	.008*	South 1
			63	0.7	3.001	3.80E-04	.010*	South 2
B*40:01	119	14.9	33 849	10.0	1.585	1.09E-05	2.83E-04*	General 1
			155 053	9.5	1.661	1.28E-06	3.34E-05*	General 2
			2051	9.6	1.651	2.71E-06	7.04E-05*	General 3
			1624	15.4	0.961	.760		South 1
			1194	14.0	1.076	.489		South 2
DQA1*05:03	16	2.5	183	0.9	2.998	2.22E-04	.003*	General 3
			114	1.1	2.366	3.27E-03	.046*	South 1
DQA1*05:05	86	10.8	1322	6.2	1.832	1.30E-06	1.82E-05*	General 3
			815	7.7	1.440	3.43E-03	.048*	South 1
DRB1*04:06	38	5.2	8531	2.5	2.127	4.21E-05	9.26E-04*	General 1
			41 741	2.6	2.076	5.82E-05	.001*	General 2
			538	2.5	2.121	7.30E-05	.002*	General 3
			241	2.3	2.338	1.20E-05	2.64E-04*	South 1
			219	2.6	2.076	1.47E-04	.003*	South 2

Abbreviations: COVID-19, coronavirus disease 2019; HLA, human leukocyte antigen.

*Significant.

and Southern populations. Nevertheless, the frequency of HLA-B*15:27 in COVID-19 patients (2.2%) was significantly higher than in both groups. Similarly, we found HLA-DQA1*05:03 and HLA-DQA1*05:05 to have significantly different allele frequencies when compared with both the general and Southern China populations. However, replications of these significances cannot be achieved due to the limited typing data for DQA1.

Furthermore, to explore the potential clinical interpretation of our results, we investigated binding affinities between viral peptides and HLA alleles (the full summary is presented in [Supplementary Table 4](#)). As shown in [Figure 1](#), HLA-B*15:27 only had 1 strong binder with spike protein (outside of the RBD region), 3 with nucleocapsid, 1 with membrane, and 0 with envelope. Meanwhile, HLA-DRB1*04:06 had 0 strong binders across all major viral proteins. These results suggest that the number of peptides recognized by HLA alleles may play a role in COVID-19 susceptibility.

DISCUSSION

Our study confirmed that HLA-B*15:27 and HLA-DRB1*04:06, which have appeared uniquely in Asia, according to the AFND,

are risk alleles for COVID-19 susceptibility. Despite that HLA-DRB1*04:06 was not significant after multiple comparison correction, both alleles have been mentioned as risk alleles in a prior study [17]. It is interesting that, while we suspect the association of HLA-A*11:01, HLA-A*11:02, and HLA-B*40:01 may be driven by population stratification, both HLA-A*11 and HLA-B*40 have been previously linked with COVID-19 infection [11, 12, 28]. Despite that HLA-B*40 did show weak peptide binding capability, HLA-A*11 exhibited moderate binding affinities in the bioinformatic prediction. Taken together, a comprehensive investigation of the impact of the subpopulation is much needed.

Additionally, DRB1*04:06 has been identified as a risk factor for autoimmune diseases [29, 30] and drug-induced maculopapular exanthema [31]. This finding is consistent with previous observations that autoimmune diseases correlate with COVID-19 severity and mortality [12, 32]. One popular explanation is that the 2 conditions share the same etiologies such as extensive inflammation. However, recent findings also suggest that the locus between DRB1 and DQA1 is related to circulating IL-6 levels [33], which often accompany cytokine storms. Taken

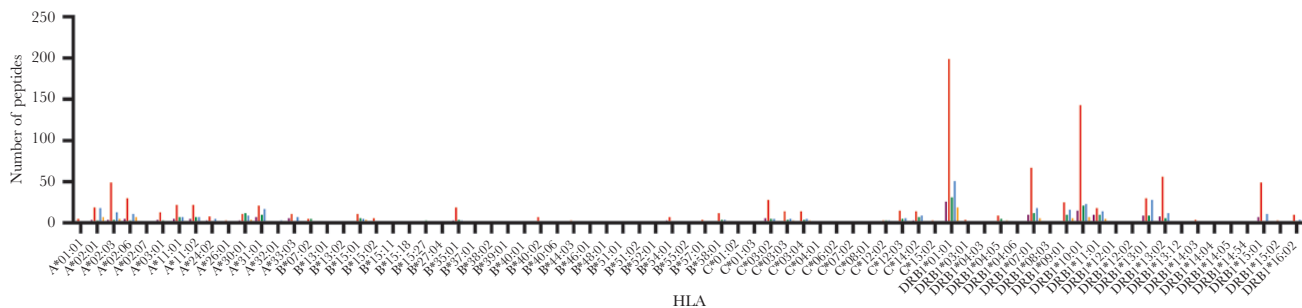


Figure 1. Numbers of strong binders derived from major viral proteins per human leukocyte antigen (HLA) allele. Only HLA alleles examined in this study are included. We labeled bindings from spike protein (S) in red, from nucleocapsid protein (N) in green, from membrane protein (M) in blue, and from envelope protein (E) in orange. More specifically, we examined the binding capabilities of peptides derived from the receptor binding domain region, which are marked in purple.

together, these findings indicate that individual vulnerability to infection may go beyond HLA affinity.

The major limitations of this study include the partial availability of haplotype data and the exclusion of asymptomatic patients. Although strain information, viral loads, COVID-19 policies, and other confounding factors were not detailed in the present study, their impacts were diminished when restricting the study within the same country and similar time frames.

Supplementary Data

Supplementary materials are available at Open Forum Infectious Diseases online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments

We would like to acknowledge all the medical staff in the Department of Infectious Diseases and COVID-19 special ward, First Hospital of Quanzhou, affiliated with Fujian Medical University, for their work on clinical specimen collection during these trying times.

Patient consent. Written informed consent was obtained for all participations in present study. This study was also approved by Disease Control and Prevention in Fujian Provincial Center (Min Ji Kong Lun Shen 2020 No. 001) and the Ethics Committee of the First Hospital of Quanzhou (Quan Yi Lun 2020 No. 124).

Financial support. This work was supported by the Young and Middle-aged Talents Training Project of Fujian Health Commission (2020GGA076 to X.Y.), Natural Science Foundation of Fujian Province (2019J01593 to X.Y., 2017J01228 to M.S.), High-level Talent Innovation Project of Quanzhou (018C067R to X.Y., 2019CT006 to K.H.), Science and Technology Pilot Project of Fujian Province (2020Y0005 to X.Z.), Science and Technology Innovation Joint Project of Fujian Province (2019Y9048 to X.Y.), and Science and Technology Project of Quanzhou (2018Z069 to X.Y., 2017Z003 to K.H., 2018C076R to K.H.).

Potential conflicts of interest. The authors declare no conflicts of interest. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. World Health Organization. WHO coronavirus disease (COVID-19) dashboard. 2020. Available at: <https://covid19.who.int/>. Accessed 25 July 2021.
2. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China. *JAMA* 2020; 323:1061–9.

3. Hill AV. Immunogenetics and genomics. *Lancet* 2001; 357:2037–41.
4. Ou G, Xu H, Yu H, et al. The roles of HLA-DQB1 gene polymorphisms in hepatitis B virus infection. *J Transl Med* 2018; 16:362.
5. Thursz M, Yallop R, Goldin R, et al. Influence of MHC class II genotype on outcome of infection with hepatitis C virus. The HENCORE Group. Hepatitis C European Network for Cooperative Research. *Lancet* 1999; 354:2119–24.
6. Huang J, Huang K, Xu R, et al. The associations of HLA-A*02:01 and DRB1*11:01 with hepatitis C virus spontaneous clearance are independent of IL28B in the Chinese population. *Sci Rep* 2016; 6:31485.
7. El-Bendary M, Neamatallah M, Elalfy H, et al. HLA class II-DRB1 alleles with hepatitis C virus infection outcome in Egypt: a multicentre family-based study. *Ann Hepatol* 2019; 18:68–77.
8. Lin M, Tseng HK, Trejaut JA, et al. Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Med Genet* 2003; 4:9.
9. Wang F, Huang S, Gao R, et al. Initial whole-genome sequencing and analysis of the host genetic contribution to COVID-19 severity and susceptibility. *Cell Discov* 2020; 6:83.
10. Amoroso A, Magistro P, Vespasiano F, et al; Italian Network of Regional Transplant Coordinating Centers. HLA and ABO polymorphisms may influence SARS-CoV-2 infection and COVID-19 severity. *Transplantation* 2021; 105:193–200.
11. Lorente L, Martin MM, Franco A, et al. HLA genetic polymorphisms and prognosis of patients with COVID-19. *Med Intensiva (Engl Ed)* 2021; 45:96–103.
12. 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.
13. Ellinghaus D, Degenhardt F, Bujanda L, et al; Severe Covid-19 GWAS Group. Genomewide association study of severe Covid-19 with respiratory failure. *N Engl J Med* 2020; 383:1522–34.
14. Poulton K, Wright P, Hughes P, et al. A role for human leukocyte antigens in the susceptibility to SARS-Cov-2 infection observed in transplant patients. *Int J Immunogenet* 2020; 47:324–8.
15. Szolek A, Schubert B, Mohr C, et al. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* 2014; 30:3310–6.
16. Liu C, Yang X, Duffy B, et al. ATHLATES: accurate typing of human leukocyte antigen through exome sequencing. *Nucleic Acids Res* 2013; 41:e142.
17. Wang W, Zhang W, Zhang J, et al. Distribution of HLA allele frequencies in 82 Chinese individuals with coronavirus disease-2019 (COVID-19). *HLA* 2020; 96:194–6.
18. Zhou XY, Zhu FM, Li JP, et al. High-resolution analyses of human leukocyte antigens allele and haplotype frequencies based on 169 995 volunteers from the China Bone Marrow Donor Registry Program. *PLoS One* 2015; 10:e0139485.
19. He Y, Li J, Mao W, et al. HLA common and well-documented alleles in China. *HLA* 2018; 92:199–205.
20. Zhou F, Cao H, Zuo X, et al. Deep sequencing of the MHC region in the Chinese population contributes to studies of complex disease. *Nat Genet* 2016; 48:740–6.
21. Zhanrou Quan ZD, Zhou D, Zhong Y, Chen H, Hong W, Zhou H. Analysis on polymorphism of high-resolution HLA-A, -B, -C, -DRB1 and -DQB1 in hematopoietic stem cells donors of Chinese Han population from Southern China. *Int J Blood Transfus Hematol* 2018; 41:497–505.
22. Nguyen A, David JK, Maden SK, et al. Human leukocyte antigen susceptibility map for severe acute respiratory syndrome coronavirus 2. *J Virol* 2020; 94:e00510–20.
23. Reynisson B, Alvarez B, Paul S, et al. 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**; 48:W449–54.
24. Shkurnikov M, Nersisyan S, Jankevic T, et al. Association of HLA class I genotypes with severity of coronavirus disease-19. *Front Immunol* **2021**; 12:641900.
 25. Iturrieta-Zuazo I, Rita CG, García-Soidán A, 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* **2020**; 219:108572.
 26. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **2007**; 81:559–75.
 27. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour* **2010**; 10:564–7.
 28. Warren RL, Birol I. Retrospective in silico HLA predictions from COVID-19 patients reveal alleles associated with disease prognosis. medRxiv 2020.10.27.20220863 [Preprint]. 2 November 2020. Available at: <https://doi.org/10.1101/2020.10.27.20220863>. Accessed 30 January 2021.
 29. Sun Y, Liu H, Yang B, et al. Investigation of the predisposing factor of pemphigus and its clinical subtype through a genome-wide association and next generation sequence analysis. *J Eur Acad Dermatol Venereol* **2019**; 33:410–5.
 30. Zhao Y, Zhao Y, Zhang Y, Zhang L. HLA-II genes are associated with outcomes of specific immunotherapy for allergic rhinitis. *Int Forum Allergy Rhinol* **2019**; 9:1311–7.
 31. Shi YW, Wang J, Min FL, et al. HLA risk alleles in aromatic antiepileptic drug-induced maculopapular exanthema. *Front Pharmacol* **2021**; 12:671572.
 32. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **2020**; 584:430–6.
 33. Ahluwalia TS, Prins BP, Abdollahi M, et al; CHARGE Inflammation Working Group. Genome-wide association study of circulating interleukin 6 levels identifies novel loci. *Hum Mol Genet* **2021**; 30:393–409.