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) 797–802

Contents lists available at ScienceDirect

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

Research article

In-silico study of influence of HLA heterogeneity on CTL responses across ethnicities to SARS-CoV-2

Vishal Rao [a], Nagasuma Chandra [a,b,*]

[a] Department of Biochemistry, Indian Institute of Science, Bangalore, India
[b] Center for BioSystems Science and Engineering (BSSE), Indian Institute of Science, Bangalore, India

ARTICLE INFO

Keywords:
HLA
COVID-19
Population
Ethnic group
CTL epitopes

ABSTRACT

Differences in outcome to COVID-19 infection in different individuals is largely attributed to genetic heterogeneity leading to differential immune responses across individuals and populations. HLA is one such genetic factor that varies across individuals leading to differences in how T-cell responses are triggered against SARS-CoV-2, directly influencing disease susceptibility. HLA alleles that influence COVID-19 outcome, by virtue of epitope binding and presentation, have been identified in cohorts worldwide. However, the heterogeneity in HLA distribution across ethnic groups limits the generality of such association. In this study, we address this limitation by comparing the recognition of CTL epitopes across HLA genotypes and ethnic groups. Using HLA allele frequency data for ethnic groups from Allele Frequency Net Database (AFND), we construct synthetic populations for each ethnic group and show that CTL epitope strength varies across HLA genotypes and populations. We also observe that HLA genotypes, in certain cases, can have high CTL epitope strengths in the absence of top-responsive HLA alleles. Finally, we show that the theoretical estimate of responsiveness and hence protection offered by a HLA allele is bound to vary across ethnic groups, due to the influence of other HLA alleles within the HLA genotype on CTL epitope recognition. This emphasizes the need for studying HLA-disease associations at the genotype level rather than at a single allele level.

1. Introduction

Severe Acute Respiratory Syndrome Corona-Virus 2 (SARS-CoV-2) which is responsible for the Coronavirus Disease 2019 (COVID-19) pandemic, has infected over 546 million people around the world as of 4th July 2022[1]. Susceptibility to COVID-19 is known to vary across individuals and the exact reasons for this differential susceptibility remain to be understood completely. Host genetic factors that are variable across individuals have been shown to contribute to this heterogeneity in disease susceptibility and outcome [2]. The adaptive immune response, which renders specificity and memory to the host response against SARS-CoV-2, is known to be dysregulated in case of COVID-19 infection [3]. Thus, heterogeneity in host genetic factors that influence the adaptive immune response can possibly explain the widely documented heterogeneity in disease outcome [4–6].

The Human Leukocyte Antigen (HLA) super-locus in humans harbors 2 main classes of HLA genes - Class-I and Class-II that directly influence T-cell responses. The HLA Class-1 system of an individual is composed of 2 haplotypes of 3 HLA genes (HLA-A, HLA-B and HLA-C in Class-I), thus constituting 6 HLA alleles, which we refer to as the HLA genotype of that individual [7]. Viral proteins cleaved by host proteases into peptides are bound by the HLA Class-I and Class-II molecules and are presented to CD8+ and CD4+ T-cells, respectively, triggering their effector functions [8]. The distribution of HLA alleles is known to vary widely across populations [9]. Thus, viral peptides presented by these HLA molecules are bound to vary across populations. It is well known that strong presentation of viral peptides by HLA Class-I molecules directly reflects in a strong host cytotoxic T-lymphocyte (CTL) response, resulting in killing of infected cells leading to viral clearance. Previous studies have identified susceptible and protective HLA alleles in various cohorts in the context of SARS-CoV-2 infection [10–12]. Several bioinformatic tools that can predict HLA-epitope binding affinities have been developed and have been used to identify protective and susceptible HLA alleles based on epitope

Abbreviations: HLA, Human Leucocyte Antigen; CTL, Cytotoxic T-lymphocyte; SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus 2; COVID-19, Coronavirus Disease-19; CTL_i, CTL epitope strength of an individual; CTL_{pop}, CTL epitope strength of an ethnic group; AFND, Allele Frequency Net Database.

* Corresponding author at: Department of Biochemistry, Indian Institute of Science, Bangalore, India.
E-mail address: nchandra@iisc.ac.in (N. Chandra).

https://doi.org/10.1016/j.humimm.2022.09.008
Received 6 July 2022; Revised 12 September 2022; Accepted 29 September 2022
Available online 11 October 2022

binding [13,14]. Epitope prediction tools have also been used to compare CTL epitope recognition among HLA alleles across ethnic groups [15]. However, heterogeneity in HLA distribution across cohorts also limits the extrapolation of HLA allele-severity associations to a global level. Owing to paucity of data on one hand and methods to study them on the other, heterogeneity in HLA genotypes has not been studied sufficiently to address if it can explain differences in T-cell responses among individuals. Further, differences in T-cell responses among populations may also be attributed to differences in HLA allele distribution, which has also not received much attention.

In this study, we address this gap and mathematically reconstruct a large number of HLA class-1 genotypes in 240 different synthetic populations corresponding to different ethnicities by utilizing publicly available data on HLA allele frequencies from ethnic groups worldwide. We predict the number of CTL epitopes recognized by each individual using bioinformatic tools. We then compare estimated CTL epitope strength across individuals and across populations. Our models provide a framework to estimate the landscape of host heterogeneity from the CTL response point of view in individuals and in populations and provide conceptual insights to explain differences in disease outcome. Further, it presents a scheme to place the existing knowledge on HLA associations with susceptibility to COVID-19.

2. Materials and methods

2.1. SARS-CoV-2 protein sequences and HLA Class-I allele frequency data

Amino acid sequences of 10 proteins of the parent Wuhan-Hu-1 strain of SARS-CoV-2 (NCBI RefSeq accession: NC_045512.2), namely - ORF1ab polyprotein, Spike glycoprotein, ORF3a, Envelope protein, Membrane protein, ORF6, ORF7a, ORF8, Nucleocapsid protein and ORF10 were obtained from NCBI Virus [18] (https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/).

240 ethnic groups with HLA Class-I allele frequency data from the Allele Frequency Net Database [19] (AFND) (http://www.allelefrequencies.net/), covering a diverse and near-exhaustive set of geographical regions were considered. The ethnic groups were shortlisted based on i) documentation of frequencies of all 3 HLA Class I alleles - A, B and C, ii) High resolution HLA allele data with polymorphisms denoted for all the alleles documented, since it directly impacts the probabilistic construction of HLA genotypes. The authors of AFND have assigned a dataset standard for each of the 3 HLA genes (HLA-A, HLA-B, HLA-C) documented within each ethnic group as a measure of the quality of the dataset based on factors such as sample size and four-digit resolution of HLA alleles and HLA allele frequencies summing up to 1, within an ethnic group [19]. Since dataset standards are provided for each HLA gene within an ethnic group (bronze being the lowest and gold being the highest), we noted the HLA gene with the lowest standard for an ethnic group and considered this as the representative dataset standard. Geographical distribution of ethnic groups along with the dataset standards provided by AFND for each ethnic group are provided in Supplementary File-1. Dataset standards for each HLA gene within the shortlisted ethnic groups are also provided in Supplementary File-1.

2.2. Prediction of CTL epitopes

HLA Class-I Epitope predictions for the 10 proteins of the Wuhan-Hu-1 strain were performed using the NetMHCPan BA 4.1 tool [20] accessible through IEDB (Immune Epitope Database) [21] (http://tools.iedb.org/mhci/). A comprehensive list of 1827 HLA Class I alleles covered by the 240 ethnic groups were considered for pMHC binding predictions. The length of the predicted epitopes was restricted to 9 amino acids. Predicted 9-mers which bind to a given HLA allele with IC50 $<$ 50 nM were shortlisted for further analysis.

2.3. Construction of synthetic populations

HLA genotypes were constructed from HLA allele frequencies for each ethnic group by considering all combinations of 2 HLA-A alleles, 2 HLA-B alleles and 2 HLA-C alleles, both homozygous and heterozygous for each of the 3 genes. The product of frequencies of these alleles was considered to be the frequency of an HLA genotype. From the set of hypothetically generated HLA genotypes, the ones with a frequency crossing a threshold value of 5×10^{-7} were considered a part of the population for the particular ethnic group. The algorithm used to construct synthetic populations is similar to that used by Mukherjee & Chandra [17].

3. Results

3.1. Host diversity in CTL responses

We shortlisted 240 ethnic groups from Allele Frequency Net Database (AFND) satisfying our selection criteria (see Methods 2.1). Among these 240 ethnic groups, 181 are Gold standard, 36 are Silver standard and 23 are bronze standard based on the dataset standards set by AFND [19]. Our shortlisted ethnic groups covered a diverse set of geographical regions around the world (Supplementary File-1), enabling us to capture population-level heterogeneity across the globe.

We consider a set of HLA genotypes within an ethnic group to represent a population and define CTL_i as the estimated CTL epitope strength in individuals which is the number of CTL epitopes recognized by an individual and CTL_{pop} as the net CTL epitope strength per individual in an ethnic group. A list of top 10 and bottom 10 responding HLA alleles in terms of number of epitopes recognized are provided in Table 1. HLA-allele groups A*02 and B*15 appear among the top responding HLA-alleles agreeing with previous reports of effective epitope presentation by these allele groups [10].

In order to assess differences in CTL_i in different individuals and CTL_{pop} across ethnic groups, we theoretically constructed synthetic populations using a probabilistic approach such that the product of frequencies of 6 HLA alleles from an ethnic group constituting an HLA genotype crosses a pre-defined threshold. A threshold was set for HLA genotype frequency (see Methods Section 2.3), beyond which the appropriate combination of 6 HLA alleles (2 HLA-A + 2 HLA-B + 2 HLA-C) was considered as a frequently occurring HLA genotype in the population (Fig. 1A). The number of HLA genotypes was seen to vary across ethnic groups, owing to differences in the diversity of HLA distribution (Fig. 1B); ethnic groups with an even distribution of HLA alleles would harbor more combinations of HLA genotypes, resulting in a diverse population while ethnic groups with an uneven distribution would show a less diverse population, dominated by the highly frequent HLA alleles. We then pooled HLA genotypes from all the 240 ethnic groups to examine host diversity in CTL_i at a global level. A nearly continuous gradient was observed in the CTL_i across individuals (Fig. 1C) suggesting that the estimated CTL responses across individuals are highly heterogeneous. The top 10 and bottom 10 HLA genotypes based on their CTL_i are provided in Table 1.

From these global HLA genotypes, we selected the top 1 % (high), middle 1 % (medium) and bottom 1 % (low) responders, based on their CTL_i (listed in Supplementary File-2) and tested if they were explained by the presence or absence of top-responsive HLA alleles. For this, HLA alleles were ranked based on the number of epitopes recognized, and the top 20 were shortlisted as the top-responsive HLA alleles (listed in Supplementary File-2) and their frequency of occurrence was compared across the high, medium and low responding genotypes. The frequency of the top-responsive HLA alleles was rela-

Table 1
Top and bottom 10 responding HLA alleles, genotypes and ethnic groups. Response of HLA alleles is measured as the number of epitopes recognized. Response of HLA genotypes and ethnic groups are measured in terms of CTL_i and CTL_{pop} respectively. Since CTL_{pop} is the CTL epitope strength averaged over the entire population, it is rounded off to the nearest integer value.

Top 10 HLA Allele	Epitopes	Bottom 10 HLA Allele	Epitopes
A*02:11	353	A*25:04	1
A*02:50	325	B*38:12	1
B*15:62	324	B*38:20	1
A*02:22	323	A*66:01	1
A*02:104	323	B*27:14	1
A*02:122	280	B*40:110	1
C*12:19	272	B*51:93	1
B*15:156	270	B*51:64	1
B*15:132	270	C*02:17	1
B*15:03	270	B*57:07	1
HLA Genotype	**Epitopes (CTL_i)**	**HLA Genotype**	**Epitopes (CTL_i)**
A*68:01|A*02:05|B*15:03|B*41:01|C*14:03|C*07:01	699	A*66:01|A*66:01|B*58:02|B*37:01|C*04:01|C*04:01	1
A*68:01|A*02:05|B*15:03|B*41:01|C*14:03|C*02:02	698	A*74:01|A*74:01|B*14:02|B*58:02|C*04:01|C*04:01	1
A*02:03|A*11:01|B*15:25|B*55:02|C*14:02|C*03:04	691	A*74:01|A*74:01|B*15:10|B*58:02|C*04:01|C*04:01	1
A*02:03|A*11:01|B*15:25|B*13:01|C*14:02|C*03:04	682	A*74:01|A*74:01|B*49:01|B*58:02|C*04:01|C*04:01	1
A*02:03|A*11:01|B*15:25|B*51:01|C*14:02|C*03:04	682	A*66:01|A*66:01|B*58:02|B*58:02|C*04:01|C*04:01	1
A*02:03|A*11:01|B*15:25|B*46:01|C*14:02|C*03:04	680	A*66:01|A*74:01|B*58:02|B*58:02|C*04:01|C*04:01	1
A*68:01|A*68:02|B*15:03|B*41:01|C*14:03|C*07:01	677	A*66:01|A*66:01|B*37:01|B*37:01|C*04:01|C*04:01	1
A*68:01|A*68:02|B*15:03|B*41:01|C*14:03|C*02:02	676	A*66:03|A*66:03|B*58:02|B*37:01|C*04:01|C*04:01	2
A*02:03|A*11:01|B*58:01|B*40:01|C*14:02|C*03:02	675	A*66:03|A*66:03|B*37:01|B*37:01|C*04:01|C*04:01	2
A*02:22|A*68:01|B*15:39|B*35:05|C*03:04|C*04:01	669	A*66:03|A*66:01|B*58:02|B*58:02|C*04:01|C*04:01	2
Ethnic group	**Epitopes (CTL_{pop})**	**Ethnic group**	**Epitopes (CTL_{pop})**
Paraguay Argentina Ache NA-DHS 24	417	Singapore SGVP. Indian INS	167
India Khandesh Region Pawra	398	Colombia North Wiwa El Encanto	166
Colombia Waunana NA-DHS 20 (G)	378	Singapore Javaneses	162
Paraguay Argentina Guarani NA-DHS 23 (G)	372	Colombia North Chimila Amerindians	162
Brazil Terena	358	Papua New Guinea Madang	161
Costa Rica Guaymi NA-DHS 10 (G)	342	Malaysia Peninsular Malay	160
Colombia Inga NA-DHS 11 (G)	331	Georgia Tibilisi Kurd	158
Colombia Zenu NA-DHS 18 (G)	327	Singapore Riau Malay	155
Mexico Oaxaca Mixe	326	India West Coast Parsi	153
Mali Bandiagara	309	USA NMDP Southeast Asian	151
Paraguay Argentina Ache NA-DHS 24	417	Singapore SGVP. Indian INS	167

tively higher in the medium and high responding HLA genotypes, as expected (Fig. 1D). However, these top-responsive alleles were not sufficient to completely cover the high responding genotype group, indicating that certain individuals (HLA genotypes) might respond better due to a combination of relatively low-responding HLA alleles that can together, recognize a large epitope set resulting in a larger CTL_i. In addition, low frequency of occurrence of the top HLA alleles in certain ethnic groups might limit certain HLA combinations with high CTL_i.

3.2. Comparison of CTL epitope strength across populations (CTL_{pop})

From the synthetic population, we computed overall CTL epitope strength for each ethnic group (CTL_{pop}) as the sum of CTL epitope strength exhibited by each individual (CTL_i), weighted by the frequency of occurrence of the HLA genotype in the population (Fig. 2A). Upon comparison of the consolidated CTL_{pop} across populations, a wide distribution in CTL_{pop} values was observed across ethnic groups (Fig. 2A), illustrating the extent of heterogeneity in HLA distribution and the impact it has on triggering CTL responses. Among the ethnic groups included in the study, the 'Paraguay Argentina Ache NA-DHS 24′ ethnic group shows the highest CTL_{pop} while the 'Israel Ashkenazi and Non Ashkenazi Jews' group showed the lowest (Supplementary File-3). The top 10 most responding and least responding ethnic groups are provided in Table-1. We then considered 2 ethnic groups each from the high ('Paraguay Argentina Ache NA-DHS 24′, 'India Khandesh Region Pawra'), medium ('China Yunnan Hani', 'Mexico Mexico City Mestizo pop 2′) and low ('New Caledonia', 'Israel Ashkenazi and Non Ashkenazi Jews') responding ethnic groups based on their CTL_{pop}. The CTL_i of HLA genotypes within the low responding ethnic groups were generally low compared to those within the high responding ethnic groups (Fig. 2B-G), eliminating the possibility that the low response might be driven by specific low-responding subgroups within the low-responding ethnic groups.

Next, we considered the top 5 most responding and least responding HLA genotypes within each of these 6 ethnic groups (listed in Supplementary File-3) and checked for the occurrence of high responding HLA alleles (>200 epitopes) and low responding HLA alleles (1 epitope) (listed in Supplementary File-3) in these genotypes. In most cases, high responding alleles were represented in the top 5 HLA genotypes of the medium and high responding ethnic groups (Fig. 2B-G). However, in the 'Israel Ashkenazi and Non Ashkenazi Jews' ethnic group which is a low responder, a low responding HLA allele was present among the top 5 HLA genotypes (Fig. 2B). This indicates that the particular HLA genotype remains high responding by virtue of the other HLA alleles that may be relatively high responders, suggesting that the HLA genotype as a whole, determines the extent of the CTL epitope strength and hence the CTL response.

4. Discussion

Heterogeneity in numerous genetic factors that influence disease conditions is known to exist across human populations [5]. Some of these genetic factors such as the HLA genotype which directly influences the host immune response has been extensively studied in the context of human diseases [16]. However, current analyses of HLA-association with disease severity are largely restricted to individual alleles rather than HLA genotypes and further restricted to specific cohorts rather than the global population.

Our approach of reconstructing synthetic populations that mimic natural ones based on the recorded frequencies of individual alleles enables us to ask several questions that were not easily tractable before. Specifically, we have been able to ask how CTL_i may differ among the entire pool of individuals globally, within our reconstructed set of HLA genotypes. Further, we have been able to ask a similar question at a population level to find high-responding and low-responding populations. In principle, given the HLA genotype of an individual, our analysis will facilitate in classifying the individual into high,

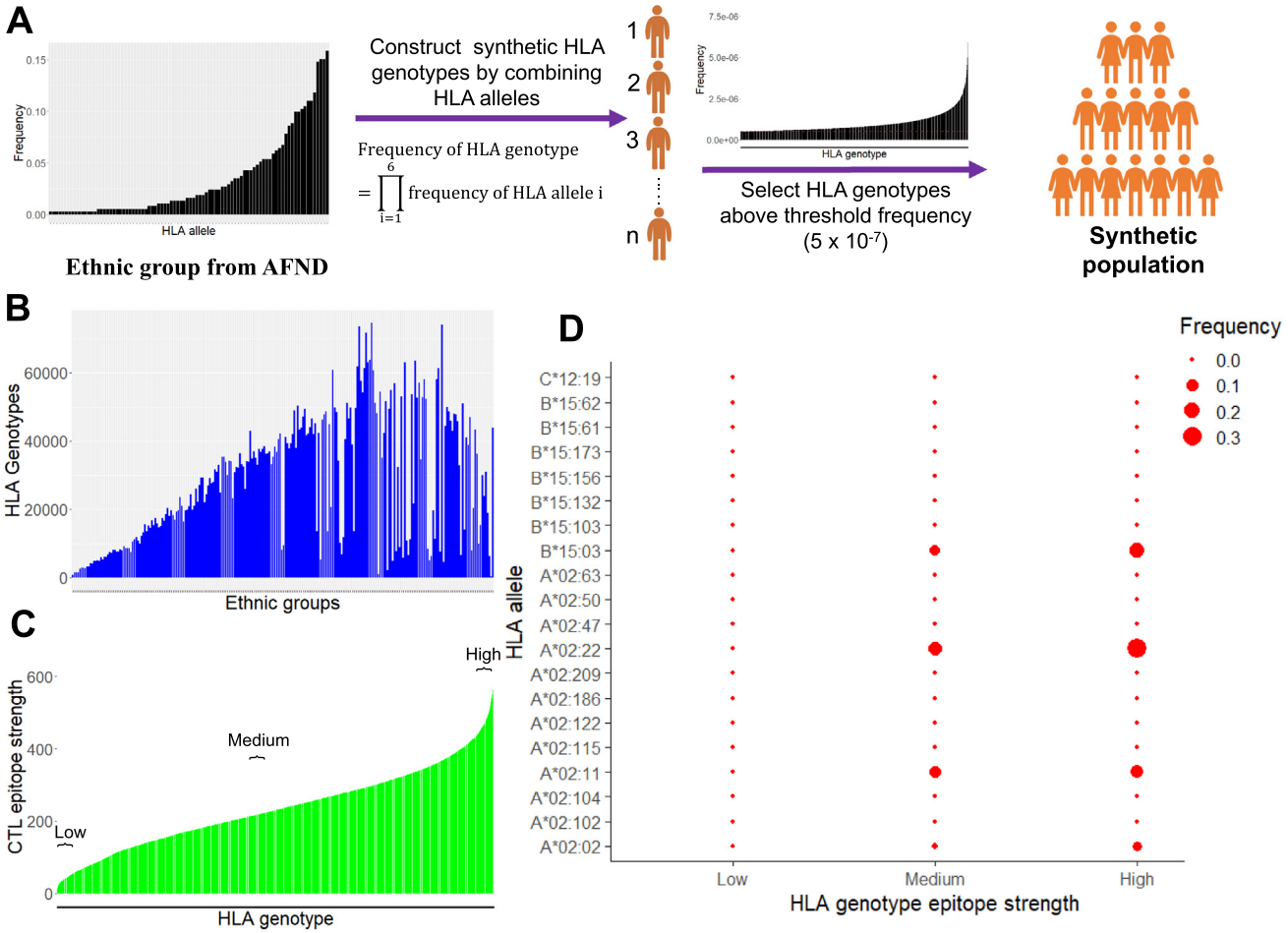

Fig. 1. Modeling CTL epitope strength across individuals(CTL_i). (A) Using the input HLA allele frequency data from each ethnic group annotated in the Allele Frequency Net Database (AFND), hypothetical genotypes (individuals) were constructed by defining a threshold frequency beyond which a particular HLA genotype which is a combination of 6 HLA alleles is considered to be a member of the population. The frequency was then scaled with respect to the least frequent genotype in the population so that the scaled frequency represented the number of individuals with the particular genotype. (B) A barplot representing the number of HLA genotypes present in each ethnic group. (C) CTL_i across global HLA genotypes. The top, middle and bottom 1% HLA genotypes, sorted based on their CTLi, are marked as 'high', 'medium' and 'low' responding genotypes, respectively. (D) A bubble plot representing the frequency of occurrence of the top 20 responding HLA alleles in the low, medium and high responding HLA genotypes.

intermediate or low-response groups to a given strain of SARS-CoV-2. With advances in sequencing methods, HLA genotyping is becoming more accessible, which may make it feasible to envisage a clinically useful responsiveness prediction method.

Our results show that there is about a 700-fold difference in the CTL epitope strength (CTL_i) between least responding and highest responding individuals. This observation holds for the pooled set of HLA genotypes reflecting individuals throughout the world. Although high responding HLA alleles were largely represented in the high responding HLA genotypes, absence of these alleles in some of the high responding genotypes suggests that a HLA genotype can still be high responding if its HLA alleles are diverse enough to recognize a large set of epitopes. Next, we compared CTL epitope strength (CTL_{pop}) across ethnic groups and found that the distribution was heterogeneous which can be attributed to population-level differences in HLA distribution. However, we cannot discount the influence of differential sampling coverage across ethnic groups, due to which certain HLA alleles might not have been captured and the HLA frequencies might not be unbiased in ethnic groups with low coverage. Despite this limitation, a substantial number of HLA genotypes were constructed within each ethnic group. A deeper analysis of ethnic groups with low, medium and high CTL_{pop} showed that the magnitude of CTL_{pop} is determined by the entire population rather than specific sub-populations. However, this does not eliminate the possibility of existence of sub-populations that differ in their epitope strengths within ethnic groups. In fact, it is evident that responses within a population are also heterogeneous and hence, a population can be further classified based on CTL_i as shown previously in case of Influenza [17]. It is also important to note that the CTL epitope strength is computed under the assumption that all epitopes equally influence the CTL response i.e, they are equally immunodominant. Experimental techniques like AIM [22] and ELISPOT [23] assays which measure the extent of T-cell activation and cytokine release respectively, upon peptide-based antigenic stimulation, would be needed to study the relative immunodominance of T-cell epitopes. However, these techniques are limited by the diversity of HLA-alleles that can be covered which forms a major part of our analysis. Despite this limitation in our analysis, it captures a theoretical pool of all high-affinity epitopes presented. In addition, overall susceptibility to COVID-19 is highly dependent on non-HLA immune factors such as the magnitude of the B-cell and antibody responses and innate immune responses as well as other factors like age, gender, comorbidities etc. [24–26] which is bound to affect association analyses of HLA with COVID-19 severity. However, the goal of our study is to show the theoretical extent of

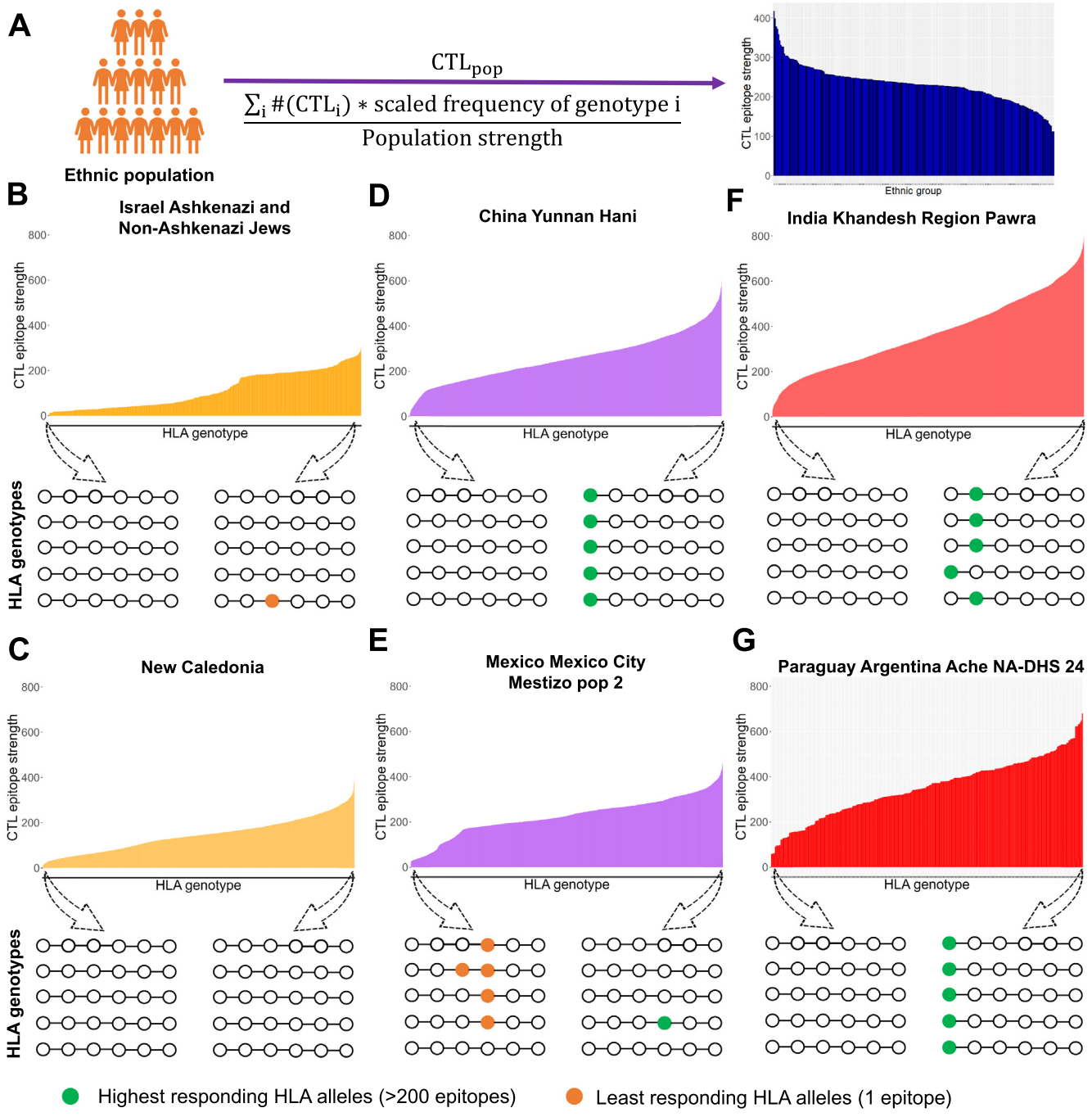

Fig. 2. Comparing CTL epitope strength across ethnic groups (CTL_{pop}). (A) CTL epitope strength for an ethnic group (CTL_{pop}) was computed as the sum of epitopes recognized by each individual within the synthetic population (CTL_i), weighted by the scaled frequency of occurrence of genotype i. This was divided by the number of individuals within the population to obtain CTL_{pop}. (B-G) Pairs of representative ethnic groups for (B-C) low, (D-E) medium and (F-G) high responders were considered and the variation of CTL_i within these ethnic groups across individuals is shown. From each of these ethnic groups, the 5 highest and 5 lowest responding HLA genotypes (dotted arrows) are represented below. Each circle represents a HLA allele and a string of 6 circles represents a HLA genotype. High-responding HLA alleles (> 200 epitopes) are shaded in green while low-responding ones (1 epitope) are shaded in orange. Blank circles represent HLA alleles that show intermediate response, as they neither belong to the highest nor the least responding HLA allele groups.

influence of one of the heterogeneous factors, namely HLA, on disease outcome at a global level. Although a theoretical exercise, the tools and datasets used in our analysis are obtained from experimental data. The algorithms used by the bioinformatic tool NetMHCPanBA 4.1 [20] for prediction of HLA-epitope binding affinity are trained on experimental data of HLA-epitope binding affinity. HLA allele frequencies documented in AFND are obtained by targeted sequencing of HLA loci among individuals within an ethnic group, sampled in an unbiased manner.

Finally, examination of the highest and lowest responding HLA genotypes within the selected ethnic groups revealed that low responding HLA alleles can also be represented in high responding genotypes. This can be explained based on two factors: i) the low-responding HLA allele was represented in a high responding HLA

genotype of an ethnic group that has a low overall CTL_{pop}. Thus, the presence of one low responding HLA allele does not majorly influence CTL_i of the genotype since the overall CTL_i of the genotype is expected to be low. ii) the presence of one low-responding HLA allele can be compensated by the presence of relatively high-responding HLA alleles with diverse epitope recognition within the HLA genotype, so that the genotype as a whole has a high CTL_i. This provides a strong reason as to why HLA-disease outcome associations within cohorts cannot be trivially extrapolated to the global level; an HLA allele might be high-responding in a cohort where the CTL_{pop} is low in general, but the same HLA allele might be a medium or low responder in a cohort where the CTL_{pop} is high. However, under the assumption that non-HLA factors important for triggering a T-cell response are not as heterogeneous across populations, we can say that an HLA genotype is bound to trigger a similar level of CTL response, irrespective of the cohort identity due to theoretically identical CTL_i. Thus, our analysis emphasizes the need for studying and comparing HLA associations at the genotype level rather than the allele level to understand the influence of HLA on COVID-19 disease outcome. Our analysis forms a stimulus to study how this population-level heterogeneity in CTL epitope recognition can trigger evolution of viral variants that escape from host CTL recognition.

Declaration of Competing Interest

NC is a co-founder of qBiome Research Pvt ltd and HealthSeq Precision Medicine Pvt ltd. They had no role in this manuscript. Both authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.humimm.2022.09.008.

References

[1] World Health Organization. (July 5, 2022). “WHO Coronavirus (COVID-19) Dashboard”. Retrieved from https://covid19.who.int/.

[2] D’Antonio, M., Nguyen, J. P., Arthur, T. D., Matsui, H., D’Antonio-Chronowska, A., Frazer, K. A., & COVID-19 Host Genetics Initiative. (2021). SARS-CoV-2 susceptibility and COVID-19 disease severity are associated with genetic variants affecting gene expression in a variety of tissues. *Cell reports*, *37*(7), 110020.

[3] E.J. Giamarellos-Bourboulis, M.G. Netea, N. Rovina, K. Akinosoglou, A. Antoniadou, N. Antonakos, G. Damoraki, T. Gkavogianni, M.-E. Adami, P. Katsaounou, M. Ntaganou, M. Kyriakopoulou, G. Dimopoulos, I. Koutsodimitropoulos, D. Velissaris, P. Koufargyris, A. Karageorgos, K. Katrini, V. Lekakis, M. Lupse, A. Kotsaki, G. Renieris, D. Theodoulou, V. Panou, E. Koukaki, N. Koulouris, C. Gogos, A. Koutsoukou, Complex immune dysregulation in COVID-19 patients with severe respiratory failure, Cell host & microbe 27 (6) (2020) 992–1000.e3.

[4] C.W. Thorball, J. Fellay, A. Borghesi, Immunological lessons from genome-wide association studies of infections, Current Opinion in Immunology 72 (2021) 87–93.

[5] J. McClellan, M.C. King, Genetic heterogeneity in human disease, Cell 141 (2) (2010) 210–217.

[6] N. Sambaturu, S. Mukherjee, M. López-García, C. Molina-París, G.I. Menon, N. Chandra, C. Viboud, Role of genetic heterogeneity in determining the epidemiological severity of H1N1 influenza, PLoS computational biology 14 (3) (2018) e1006069.

[7] T. Shiina, K. Hosomichi, H. Inoko, J.K. Kulski, The HLA genomic loci map: expression, interaction, diversity and disease, Journal of human genetics 54 (1) (2009) 15–39.

[8] P.C. Doherty, D.J. Topham, R.A. Tripp, R.D. Cardin, J.W. Brooks, P.G. Stevenson, Effector CD4+ and CD8+ T-cell mechanisms in the control of respiratory virus infections, Immunological reviews 159 (1) (1997) 105–117.

[9] S. Buhler, A. Sanchez-Mazas, I. Mokrousov, HLA DNA sequence variation among human populations: molecular signatures of demographic and selective events, PloS one 6 (2) (2011) e14643.

[10] F. Tavasolian, M. Rashidi, G.R. Hatam, M. Jeddi, A.Z. Hosseini, S.H. Mosawi, R.D. Inman, HLA, immune response, and susceptibility to COVID-19, Frontiers in Immunology 11 (2021) 3581.

[11] Yu, X., Ho, K., Shen, Z., Fu, X., Huang, H., Wu, D., ... & Su, Z. (2021, September). The Association of Human Leukocyte Antigen and COVID-19 in Southern China. In *Open Forum Infectious Diseases* (Vol. 8, No. 9, p. ofab410). US: Oxford University Press.

[12] J. Weiner 3rd, P. Suwalski, M. Holtgrewe, A. Rakitko, C. Thibeault, M. Müller, B. Heidecker, Increased risk of severe clinical course of COVID-19 in carriers of HLA-C* 04: 01, EClinicalMedicine 40 (2021) 101099.

[13] J. Sidney, B. Peters, A. Sette, Epitope prediction and identification- adaptive T cell responses in humans, Seminars in Immunology 50 (2020) 101418.

[14] A. Nguyen, J.K. David, S.K. Maden, M.A. Wood, B.R. Weeder, A. Nellore, R.F. Thompson, Human leukocyte antigen susceptibility map for severe acute respiratory syndrome coronavirus 2, Journal of virology 94 (13) (2020). e00510-20.

[15] T. Bose, N. Pant, N.K. Pinna, S. Bhar, A. Dutta, S.S. Mande, Does immune recognition of SARS-CoV2 epitopes vary between different ethnic groups?, Virus research 305 (2021) 198579

[16] J.M. Blackwell, S.E. Jamieson, D. Burgner, HLA and infectious diseases, Clinical microbiology reviews 22 (2) (2009) 370–385.

[17] S. Mukherjee, N. Chandra, Grouping of large populations into few CTL immune ‘response-types’ from influenza H1N1 genome analysis, Clinical & translational immunology 3 (8) (2014) e24.

[18] E.L. Hatcher, S.A. Zhdanov, Y. Bao, O. Blinkova, E.P. Nawrocki, Y. Ostapchuck, J. R. Brister, Virus Variation Resource–improved response to emergent viral outbreaks, Nucleic acids research 45 (D1) (2017). D482-D490 https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/.

[19] Gonzalez-Galarza FF, McCabe A, Santos EJ, Jones J, Takeshita LY, Ortega-Rivera ND, Del Cid-Pavon GM, Ramsbottom K, Ghattaoraya GS, Alfirevic A, Middleton D and Jones AR. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools *Nucleic Acid Research* 2020 48:D783-8. http://www.allelefrequencies.net.

[20] B. Reynisson, B. Alvarez, S. Paul, B. Peters, M. Nielsen, 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 research 48 (W1) (2020). W449-W454.

[21] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. 2018 Oct 24. doi: 10.1093/nar/gky1006. [Epub ahead of print] PubMed PMID: 30357391. http://tools.iedb.org/mhci/.

[22] R. da Silva Antunes, S. Paul, J. Sidney, D. Weiskopf, J.M. Dan, E. Phillips, S. Mallal, S. Crotty, A. Sette, C.S. Lindestam Arlehamn, G.P. Bansal, Definition of human epitopes recognized in tetanus toxoid and development of an assay strategy to detect ex vivo tetanus CD4+ T cell responses, PloS one 12 (1) (2017) e0169086.

[23] Y. Miyahira, K. Murata, D. Rodriguez, J.R. Rodriguez, M. Esteban, M.M. Rodrigues, F. Zavala, Quantification of antigen specific CD8+ T cells using an ELISPOT assay, Journal of immunological methods 181 (1) (1995) 45–54.

[24] B.G. Pijls, S. Jolani, A. Atherley, R.T. Derckx, J.I.R. Dijkstra, G.H.L. Franssen, S. Hendriks, A. Richters, A. Venemans-Jellema, S. Zalpuri, M.P. Zeegers, Demographic risk factors for COVID-19 infection, severity, ICU admission and death: a meta-analysis of 59 studies, BMJ open 11 (1) (2021) e044640.

[25] A. Sanyaolu, C. Okorie, A. Marinkovic, R. Patidar, K. Younis, P. Desai, Z. Hosein, I. Padda, J. Mangat, M. Altaf, Comorbidity and its impact on patients with COVID-19, SN comprehensive clinical medicine 2 (8) (2020) 1069–1076.

[26] M. Mutambudzi, C. Niedzwiedz, E.B. Macdonald, A. Leyland, F. Mair, J. Anderson, C. Celis-Morales, J. Cleland, J. Forbes, J. Gill, C. Hastie, F. Ho, B. Jani, D.F. Mackay, B. Nicholl, C. O'Donnell, N. Sattar, P. Welsh, J.P. Pell, S.V. Katikireddi, E. Demou, Occupation and risk of severe COVID-19: prospective cohort study of 120 075 UK Biobank participants, Occupational and Environmental Medicine 78 (5) (2021) 307–314.