medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

HLA-B*15:01 is associated with asymptomatic SARS-CoV-2 infection

Danillo G. Augusto, Ph.D.[1,2], Tasneem Yusufali, M.S.[1], Noah D. Peyser, Ph.D.[3], Xochitl Butcher, B.S.A.[3], Gregory M. Marcus, M.D.[3], Jeffrey E. Olgin, M.D.[3], Mark J. Pletcher, M.D.[4,5], Martin Maiers Ph.D.[6], Jill A. Hollenbach, Ph.D.[1, 4]

1. Department of Neurology, University of California, San Francisco
2. Programa de Pós-Graduação em Genética, Universidade Federal do Paraná, Curitiba, Brazil
3. Division of Cardiology, Department of Medicine, University of California, San Francisco
4. Department of Epidemiology and Biostatistics, University of California, San Francisco
5. Division of General Internal Medicine, University of California, San Francisco
6. National Marrow Donor Program, Minneapolis, MN

Corresponding Author:

Jill A. Hollenbach, PhD, MPH

University of California, San Francisco

Department of Neurology and Department of Epidemiology and Biostatistics

675 Nelson Rising Ln, room 222A

San Francisco, CA 94158

e-mail: jill.hollenbach@ucsf.edu; phone +1 (510) 207-5361

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Abstract

Background. Evidence has shown that a large proportion of SARS-CoV-2 infected individuals do not experience symptomatic disease. Owing to its critical role in immune response, we hypothesized that variation in the *human leukocyte antigen* (HLA) loci may underly asymptomatic infection.

Methods. We enrolled 29,947 individuals registered in the National Marrow Donor Program for whom high-resolution *HLA* genotyping data were available in a smartphone-based study designed to track COVID-19 symptoms and outcomes. Among 21,893 individuals who completed the baseline survey, our discovery (N=640) and replication (N=788) cohorts were comprised of self-identified White subjects who reported a positive test result for SARS-CoV-2. We tested for association of five *HLA* loci (*HLA-A*, *-B*, *-C*, *-DRB1*, *-DQB1*) with asymptomatic vs. symptomatic infection.

Results. *HLA-B*15:01* was significantly increased in asymptomatic individuals in the discovery cohort compared to symptomatic (OR = 2.45; 95%CI 1.38-4.24, $p = 0.0016$, $p^{corr} = 0.048$), and we reproduced this association in the replication cohort (OR= 2.32; 95%CI = 1.10-4.43, $p = 0.017$). We found robust association of *HLA-B*15:01* in the combined dataset (OR=2.40 95% CI = 1.54-3.64; $p = 5.67 \times 10^{-5}$) and observed that homozygosity of this allele increases more than eight times the chance of remaining asymptomatic after SARS-CoV-2 infection (OR = 8.58, 95%CI = 1.74-34.43, $p = 0.003$). Finally, we demonstrated the association of *HLA-B*15:01* with asymptomatic SARS-Cov-2 infection is enhanced by the presence of *HLA-DRB1*04:01*

Conclusion. *HLA-B*15:01* is strongly associated with asymptomatic infection with SARS-CoV-2 and is likely to be involved in the mechanism underlying early viral clearance.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Introduction

Despite some inconsistent reporting of symptoms,[1] studies have demonstrated that at least 20% of individuals infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) will remain asymptomatic. [2-4] Although global efforts have focused on understanding factors underlying severe illness in COVID-19 (coronavirus disease of 2019), the examination of asymptomatic infection provides a unique opportunity to consider early disease and immunologic features promoting rapid viral clearance. Specific focus on asymptomatic infection has the potential to further our understanding of disease pathogenesis and supports ongoing efforts toward vaccine development and the discovery of potential therapeutic targets.

It remains unclear why many individuals successfully clear infection without major complications while others develop severe disease, even in the absence of known risk factors for severe COVID-19 outcomes.[5] However, host genetics is long-known to be implicated in differential immunological responses to infection and disease progression.[6] Numerous studies intending to understand the genetic basis of differential outcomes in COVID-19 have been underway since nearly the start of the global pandemic, including the multicenter Host Genetics Initiative[7]. However, the vast majority of these studies have examined genetic associations with severe disease course, in primarily hospitalized cohorts.[8-10] As a result, although most individuals infected with SARS-Cov-2 experience mild disease course or are entirely asymptomatic, very few studies have examined genetics in the context of non-hospitalized, prospective, community-based cohorts.

The *human leukocyte antigen (HLA)* region, located on chromosome 6p21, is the most polymorphic and medically important region of the human genome. Variation in *HLA* has been associated with hundreds of diseases and conditions, including infection. Infectious diseases are one of the leading causes of human mortality[11] and are a primary selective pressure

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

shaping the human genome.[12,13] Among the myriad genes involved in human immune responses, HLA variants are among the strongest reported associations with viral infections. For example, *HLA* was strongly associated with the rapid progression and viral load control of HIV (human immunodeficiency virus),[14] hepatitis B, hepatitis C, and other infectious diseases.[15] Notably, *HLA* class I and class II alleles have also been associated with the severe acute respiratory syndrome caused by SARS-CoV.[16-21]

A series of in silico analyses have pointed to HLA as relevant molecules for SARS-CoV-2 risk and essential targets for vaccine development.[22-25] For example, *HLA-B*46:01* has low predicted binding of peptides for SARS-CoV-2, suggesting that individuals expressing this molecule may be more vulnerable to COVID-19,[24] which corroborates previous results showing *HLA-B*46:01* association with SARS risk.[18] In contrast, *HLA-B*15:03* was predicted to protect against COVID-19 with the greatest ability to present highly conserved SARS-CoV-2 peptides to T cells.[24] More recently, it was demonstrated that while there is some overlap, many SARS-CoV-2 epitopes for CD8 T cells are *HLA* specific.[24] To date, very few studies have directly examined HLA associations with infection, with mixed and inconclusive results in relatively small cohorts.[26-28] Larger studies that relied on genome-wide data to impute *HLA* failed to find robust associations with disease.[8,29] However, these studies focused primarily on hospitalized patients with severe disease course.

Because of its pivotal role in the immune response, understanding the impact of *HLA* variation in disease promises to provide meaningful insights relevant to understanding the immunopathogenesis of COVID-19, while informing vaccine development and potential immunotherapies. Here, we present the largest study to-date directly examining *HLA* variation in the context of primarily mild disease. We invited volunteer bone marrow donors for whom high-resolution *HLA* genotyping data were already available to participate in the COVID-19 Citizen Science Study, a smartphone-based study designed to track COVID-19 symptoms and

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

outcomes, including self-reported positive tests for SARS-CoV-2 infection, to develop a prospective cohort currently numbering nearly 30,000 individuals. Our early results provide strong support for the role of HLA class I in viral clearance leading to asymptomatic infection among persons with SARS-CoV-2 infection and provide an important framework for additional studies aimed at revealing the immunological and genetic basis for recovery from SARS-CoV-2 infection.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Methods

Data collection

This study was conducted in accordance with the ethical standards of the UCSF Human Research Protection Program Institutional Review under approval #20-30850 and informed consent was obtained from all individual participants involved in the study. Subjects were volunteer bone marrow donors with valid email addresses on file with the National Marrow Donor Program (NMDP) who were invited to participate in the study through an email outreach campaign that began in July 2020. All subjects have within the NMDP database a pre-existing record for high-resolution *HLA* genotyping, typically for five loci (*HLA-A, -B, -C, -DRB1,* and -*DQB1*).[30] Participants who opt in to the study are required to download a smartphone app and participate in the COVID-19 Citizen Science Study (launched using the Eureka Digital Research Platform, https://eureka.app.link/covid19) or, as of January 2021, participate via website (https://covid19.eurekaplatform.org/) . Once enrolled, participants are asked to complete an initial 10- to 15-minute survey about baseline demographics, their health history and daily habits. Follow-up daily questions specific to symptoms, weekly questions regarding testing and monthly questions regarding hospitalization for COVID-19, are delivered by push notification or text message on an ongoing basis and require five to 15 minutes per week. As of April 30, 2021, we enrolled 29,947, of whom 21,893 have completed their baseline survey.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

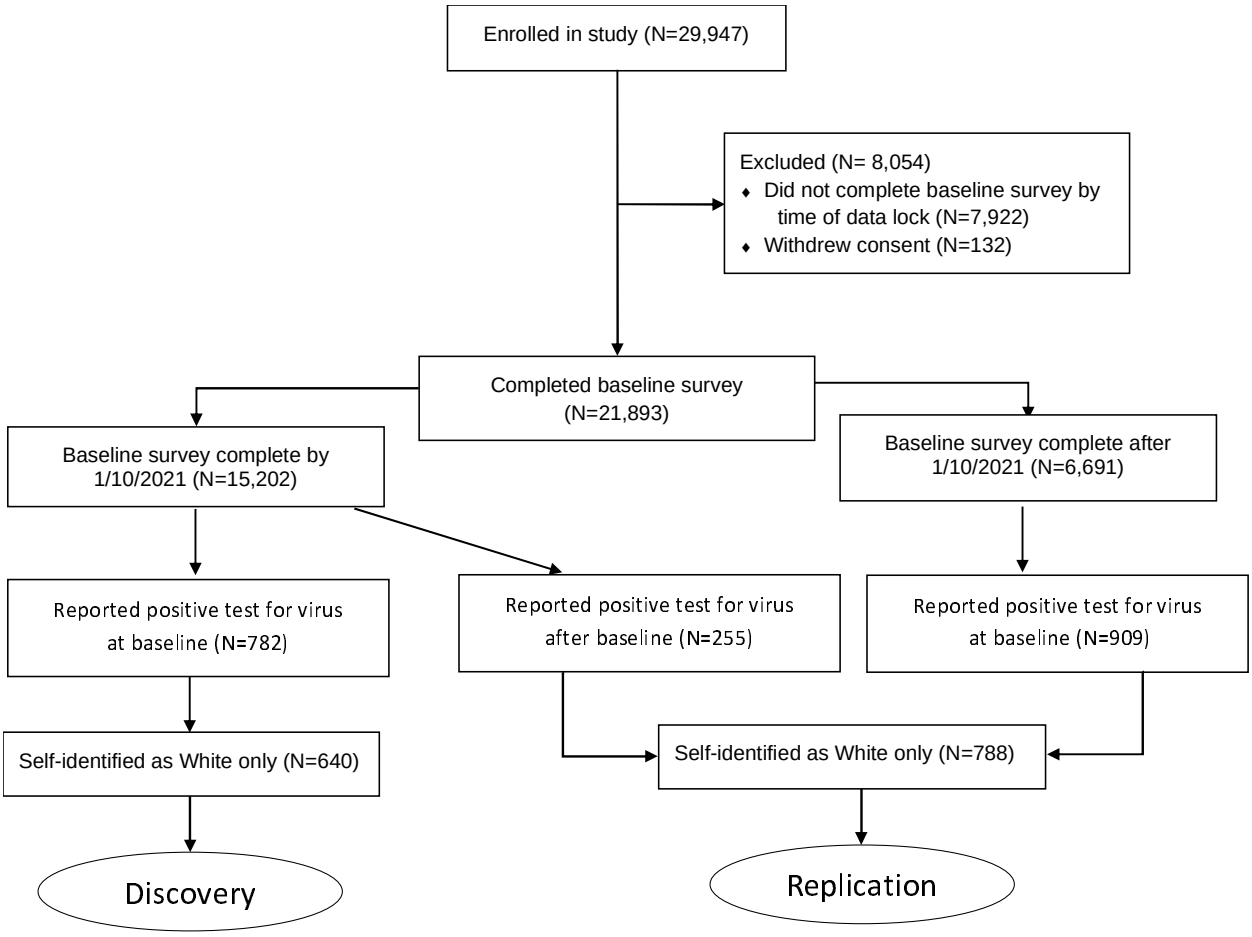


Figure 1. Study inclusion criteria

Within the mobile application, survey respondents are asked during their initial baseline survey whether they have ever been tested for active infection and report the result (positive, negative, do not know) and the approximate number of weeks since the test. Thereafter, each week respondents are asked whether they were tested in the prior week, and to report the result. We considered anybody reporting a positive test for active infection as having been infected with SARS-CoV-2. Our discovery cohort consisted of individuals reporting a positive test for virus at baseline at the point that we had collected baseline data for the first 15,000 respondents; we achieved this threshold on January 10, 2021. Respondents reporting a positive test for the virus after baseline (i.e., on a weekly survey) or who completed their baseline survey after January

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

10, 2021, together formed our replication cohort. We restricted the analysis to individuals who had self-identified as "White" only due to insufficient numbers for analysis in other groups, allowing analysis of 640 individuals in our discovery cohort and 788 individuals in our replication cohort. Inclusion criteria are detailed in **Figure 1**.

Symptoms are self-reported at baseline and using the daily surveys. Within the baseline survey, respondents are asked to report whether they had any of a list of symptoms (**Table S1**) for three days or longer at any time since February of 2020. These same symptoms are queried in each daily survey, where respondents are asked whether they experienced each symptom within the previous 24 hours. Among those individuals, we considered those who reported having had a positive test for active virus at baseline, with a time since the test of longer than two weeks or who did not specify test dates, and who reported "None of the above" for all symptoms in the baseline survey, as "asymptomatic." We also considered daily symptom reports for the two weeks after baseline for respondents who reported a positive test for active infection at baseline as having occurred within the prior two weeks. In these cases, we considered individuals asymptomatic if in addition to reporting no symptoms at baseline, they did not report any single symptom two or more times within this time period. For individuals who did not report a positive test for active infection at baseline, but subsequently reported a positive test on a weekly survey, we used the same criteria considering daily symptom reports for the period two weeks prior and two weeks after the positive test report.

HLA association analysis

We examined the association of five *HLA* loci (*HLA-A, -B, -C, -DRB1, -DQB1*) with asymptomatic vs. symptomatic infection. Initial testing for *HLA* associations was performed using the R package BIGDAWG,[31] which handles multiallelic *HLA* data to test for association at the haplotype, locus, allele, and amino acid levels. We employed a generalized linear model

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

using 'glm' in the R base package to consider relevant covariates, including any reported comorbidities, sex, and age, for alleles initially found to be associated with asymptomatic infection after correction for multiple testing. We corrected p-values using the Bonferroni method[32] for the number of alleles tested at *HLA-A*, *-B*, and *-DRB1*, which allows for the strong linkage disequilibrium between some loci tested.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Results

Study population characteristics

Our study population was overwhelmingly comprised of individuals with mild disease, with only two individuals reporting hospitalization for COVID-19. The discovery cohort consisted of 640 individuals who reported a positive test for active SARS-CoV-2 infection at baseline and self-identified as White. Among these respondents, 90 (14.1%) reported having remained asymptomatic for at least two weeks after a positive test for virus. Among the 788 individuals who reported a positive test result for SARS-CoV-2 active infection and self-reported as White who comprised our replication cohort, 46 (5.8%) were asymptomatic according to our study criteria.

Median age was increased ($p < 0.001$) in asymptomatic compared to symptomatic individuals in each cohort (discovery asymptomatic=44, symptomatic=37; replication asymptomatic =38, symptomatic=31). Likewise, median age in our discovery cohort overall (38) was higher than our replication cohort (31; $p<0.001$). Finally, while not statistically significant, there was a trend toward increased asymptomatic infection in males ($p=0.07$) in our discovery cohort. Basic demographics for all subjects are given in **Table 1**.

We also collected data on several diseases and conditions that might impact COVID-19 disease course. Among all individuals who reported a positive test for virus across both cohorts (N=1,428), 67% reported no known COVID-19-associated comorbidities. We did not observe a significant difference in any reported comorbidity between the discovery and replication cohorts. The full list of reported diseases and conditions is given **Table S2**

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

*HLA-B*15:01 is associated with asymptomatic SARS-CoV-2 infection*

We aimed to identify whether HLA variation impacts the likelihood that an individual will remain asymptomatic after SARS-CoV-2 infection. We analyzed high-resolution genotyping for five highly polymorphic *HLA* class I and class II genes (*HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1*, *HLA-DQB1*) in each study cohort. Data analysis included the first two fields of the allele name as described in the HLA nomenclature, representing the complete molecule at polypeptide sequence resolution.

In the discovery cohort, we found the allele *HLA-B*15:01* significantly overrepresented in asymptomatic individuals relative to symptomatic individuals (*f*=0.11 vs. 0.047; OR=2.52; 95%CI 1.39-4.42, p = 0.00057, p^{corr} = 0.017, Table 2). No other *HLA* allele at any locus was found to be significantly associated after correction for multiple comparisons. Allelic frequencies for all loci are given in **Tables S3-S7**.

To adjust for the effect of comorbid conditions, as well as sex and age differences in asymptomatic vs. symptomatic patients, we fitted a series of regression models but did not find any impact of patient-reported comorbidities on the likelihood of asymptomatic disease. Thus, our final model adjusted only for age and sex, which again showed a significant association of *HLA-B*15:01* with asymptomatic infection after adjustment for these variables (OR = 2.45; 95%CI 1.38-4.24, p = 0.0016, p^{corr} = 0.048, Table 2).

We reproduced the association of *HLA-B*15:01* with asymptomatic infection (Table 2, sex and age adjusted) in the replication cohort and found a remarkably consistent effect size for this allele (OR= 2.32; 95%CI = 1.10-4.43, p = 0.017). Given the consistency between the discovery and replication cohort observations, we examined the combined dataset for an overall association of *HLA-B*15:01* with asymptomatic infection (OR=2.40 95% CI = 1.54-3.64; p = 5.67 x10^{-5}). Finally, we observed a strong additive effect for the associated genotype in the combined

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

dataset. Individuals who carry two copies of *HLA-B*15:01* are more than eight times more likely to remain asymptomatic than individuals carrying other genotypes (OR = 8.58, 95%CI = 1.74-34.43, p = 0.003). Overall, one in five individuals (20%) who remained asymptomatic after infection carried *HLA-B*15:01,* compared to 9% among patients reporting symptoms.

*HLA-B*15:01 association with asymptomatic SARS-Cov-2 infection is enhanced by the presence of HLA-DRB1*04:01*

To understand whether additional *HLA* alleles might interact with *HLA-B*15:01* in asymptomatic infection, we tested all pairwise two-locus haplotypes containing *HLA-B* in the combined dataset. Overall, haplotypic associations for *HLA-B~HLA-DRB1* and *HLA-A~HLA-B* were found to be significant at p=0.01. Examining specific allelic haplotypes, these associations were driven by two *HLA-B*15:01* haplotypes: *HLA-B*15:01~HLA-DRB1*04:01,* and *HLA-A*02:01~HLA-B*15:01* **(Tables S8-9)**.

After adjusting for sex and age, only the combination of *HLA-B*15:01* and *HLA-DRB1*04:01* remained significant after correction for multiple comparisons ($p = 3 \times 10^{-4}$, $p^{corr} = 0.01$). We found an odds ratio for this combination (OR=3.17, 95% CI=1.65-5.80) that exceeds that for *HLA-B*15:01* alone, suggesting that while not significantly associated with the asymptomatic infection on its own, the class II allele *HLA-DRB1*04:01* enhances the effect of *HLA-B*15:01.*

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Discussion

Leveraging big data and mobile technology in this crowd-sourced study, we reveal important insight into the immunogenetic underpinnings of asymptomatic SARS-CoV-2 infection. Our innovative use of a mobile application and a pre-existing database for medical research allowed us to screen nearly 30,000 individuals previously genotyped for *HLA* for viral infection and disease course.

Among participants reporting a positive test result for SARS-CoV-2, *HLA-B*15:01* is significantly associated with asymptomatic infection. We observed that individuals carrying this relatively common allele are more than twice as likely to remain asymptomatic after SARS-CoV-2 infection than those who do not, and an astonishing effect for *HLA-B*15:01* homozygosity increasing more than eight times the chance of remaining asymptomatic. This observation suggests important features of early infection with SARS-CoV-2.

Although our data do not point to a direct mechanism, compelling evidence indicates that B*15:01 is predicted to bind a specific SARS-CoV-2 epitope, WTAGAAAYY, with high affinity.[33-38] This epitope has been considered a natural candidate for SARS-Cov-2 vaccine studies[39,40] as it is considered to be highly immunogenic and predicted to elicit a strong CD8+ T cell responses.[41-44] The dominant recognition and consequent eliciting of immune responses by only a relatively small number of epitopes– immunodominance – is determined by multiple factors, including the binding affinity of the peptide for the HLA molecule.[45] We speculate that the high affinity of HLA-B*15:01 for binding WTAGAAAYY or another unknown viral peptide confers strong CD8+ T cell responses against SARS-CoV-2 infected cells, leading to better early viral control and thus resulting in asymptomatic course of infection. An alternative explanation for our results is that *HLA-B*15:01* carriers previously infected by other common coronaviruses have pre-existing cross-reactive T-cells with high affinity for a specific epitope conferred by this HLA

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

molecule. While the current literature is mixed regarding cross-reactive CD8+ T cells specific to SARS-CoV-2, this might be explained by HLA specificity.[46,47] The fact that asymptomatic individuals in our study were less likely than symptomatic subjects to report a positive test for antibodies against SARS-CoV-2 (**Table S10**) suggests a strong early anti-viral response and may corroborate the critical role of CD8+ T cytotoxicity mediated by HLA class I in eliminating SARS-CoV-2 infected cells.

A study from England has recently reported a suggestive association of *HLA-DRB1*04:01* with asymptomatic infection.[27] While our results did not corroborate this association for *HLA-DRB1*04:01* alone, we did find that this allele enhanced the effect of *HLA-B*15:01* when the pair were in combination. We note that this is the *HLA-DRB1* allele most commonly associated with *HLA-B*15:01* in individuals in the U.S. who self-identify as White, and thus it is difficult to differentiate a real effect from one related to linkage disequilibrium between these loci. We speculate that the study by Langton et al.[27] may have been detecting the same *HLA-B*15:01* association observed here, but due to the high number of alleles commonly observed at *HLA-B*, that study may have been insufficiently powered to detect the association.

A key limitation of this study is that all testing results and symptoms are self-reported. We recognize that this may result in some margin of error in our results. However, we have previously validated this approach by verifying test results in a subset of the participants.[48] Additionally, we note that we find a remarkably consistent genetic association across the study, pointing to a true biological feature.

Likewise, some key differences were observed between the discovery and replication phase of our study. For example, differences in SARS-CoV-2 positivity results between the first (5%) and second (17%) stages may be explained by the overall trajectory of the pandemic in the United States. The discovery cohort primarily comprises individuals enrolled relatively early in the

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

pandemic, while the replication cohort enrolled at the height of the winter surge. Likewise, the lower rate of asymptomatic infection in the replication cohort (5.8% vs. 14.1% in discovery) may be attributed to many factors related to change over time in the pandemic; such as the prevalence of more virulent viral variants or shifts in public utilization of testing. We also observed significantly older age in asymptomatic individuals compared to those who developed symptomatic infection. This difference may be at least in part explained by subjects' self-reported reasons for seeking testing for the virus. Perhaps unsurprisingly, the most common reason patients reporting symptomatic infection sought testing was because they experienced symptoms consistent with COVID-19. In contrast, asymptomatic subjects often were tested for the virus in conjunction with their employment status, such as healthcare workers, who may skew older in our study population. While we cannot rule out a biological basis for the age association with asymptomatic infection, we feel that some ascertainment bias may also explain this observation. Regardless, our observations with respect to *HLA* remain highly significant upon adjusting for age.

In summary, we have demonstrated a strong and significant association of a common *HLA* class I allele, *HLA-B*15:01*, with asymptomatic infection with SARS-CoV-2. Our results have important implications for understanding early infection and the mechanism underlying early viral clearance and may lay the groundwork for refinement of vaccine development and therapeutic options in early disease.

Author contributions

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

JAH and MM conceived the study; TY, NDP, XB, GMM, JEO, MJP, MM, and JAH acquired and curated the data; JAH and TY analyzed the data; DGA and JAH verified underlying data; DGA and JAH wrote the original draft of the manuscript; All authors reviewed, edited, and approved the final manuscript. The corresponding author had full access to all the data and final responsibility for the decision to submit for publication.

Data sharing

All data will be posted to the HLA|COVID-19 NIH-funded public database (hlacovid19.org/database) managed by the corresponding author, and also made available directly upon reasonable request.

Declaration of interests

The authors declare no conflict of interests.

Acknowledgment

This study was funded by grants R01AI128775 and 3U2CEB021881-05S1 from the National Institutes of Health.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

References

1 Meyerowitz EA, Richterman A, Bogoch II, Low N, Cevik M. Towards an accurate and systematic characterisation of persistently asymptomatic infection with SARS-CoV-2. *Lancet Infect Dis* 2020; published online Dec 7. DOI:10.1016/S1473-3099(20)30837-9.

2 Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection : A Narrative Review. *Ann Intern Med* 2020; **173**: 362–7.

3 Buitrago-Garcia D, Egli-Gany D, Counotte MJ, *et al.* Occurrence and transmission potential of asymptomatic and presymptomatic SARS-CoV-2 infections: A living systematic review and meta-analysis. *PLoS Med* 2020; **17**: e1003346.

4 Ra SH, Lim JS, Kim G-U, Kim MJ, Jung J, Kim S-H. Upper respiratory viral load in asymptomatic individuals and mildly symptomatic patients with SARS-CoV-2 infection. *Thorax* 2021; **76**: 61–3.

5 Docherty AB, Harrison EM, Green CA, *et al.* Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* 2020; **369**: m1985.

6 Kwok AJ, Mentzer A, Knight JC. Host genetics and infectious disease: new tools, insights and translational opportunities. *Nat Rev Genet* 2021; **22**: 137–53.

7 COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur J Hum Genet* 2020; **28**: 715–8.

8 Severe Covid-19 GWAS Group, Ellinghaus D, Degenhardt F, *et al.* Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N Engl J Med* 2020; **383**: 1522–34.

9 Pairo-Castineira E, Clohisey S, Klaric L, *et al.* Genetic mechanisms of critical illness in COVID-19. *Nature* 2021; **591**: 92–8.

10 Gómez J, Albaiceta GM, García-Clemente M, García-Gala JM, Coto E. DNA genotyping of the ABO gene showed a significant association of the A-group (A1/A2 variants) with severe COVID-19. *Eur J Intern Med* 2021; published online Feb 25. DOI:10.1016/j.ejim.2021.02.016.

11 Burgner D, Jamieson SE, Blackwell JM. Genetic susceptibility to infectious diseases: big is beautiful, but will bigger be even better? *Lancet Infect Dis* 2006; **6**: 653–63.

12 Sabeti PC, Reich DE, Higgins JM, *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* 2002; **419**: 832–7.

13 Walsh EC, Sabeti P, Hutcheson HB, *et al.* Searching for signals of evolutionary selection in 168 genes related to immune function. *Hum Genet* 2006; **119**: 92–102.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

14 Martin MP, Carrington M. Immunogenetics of HIV disease. *Immunol Rev* 2013; **254**: 245–64.

15 Blackwell JM, Jamieson SE, Burgner D. HLA and infectious diseases. *Clin Microbiol Rev* 2009; **22**: 370–85–TableofContents.

16 Ng MHL, Cheng SH, Lau KM, *et al.* Immunogenetics in SARS: a case-control study. *Hong Kong Med J* 2010; **16**: 29–33.

17 Lin M, Tseng H-K, Trejaut JA, *et al.* Association of HLA class I with severe acute respiratory syndrome coronavirus infection. *BMC Med Genet* 2003; **4**: 9–7.

18 Chen Y-MA, Liang S-Y, Shih Y-P, *et al.* Epidemiological and genetic correlates of severe acute respiratory syndrome coronavirus infection in the hospital with the highest nosocomial infection rate in Taiwan in 2003. *J Clin Microbiol* 2006; **44**: 359–65.

19 Keicho N, Itoyama S, Kashiwase K, *et al.* Association of human leukocyte antigen class II alleles with severe acute respiratory syndrome in the Vietnamese population. *Human Immunology* 2009; **70**: 527–31.

20 Wang S-F, Chen K-H, Chen M, *et al.* Human-leukocyte antigen class I Cw 1502 and class II DR 0301 genotypes are associated with resistance to severe acute respiratory syndrome (SARS) infection. *Viral Immunol* 2011; **24**: 421–6.

21 Ng MHL, Lau K-M, Li L, *et al.* Association of human-leukocyte-antigen class I (B*0703) and class II (DRB1*0301) genotypes with susceptibility and resistance to the development of severe acute respiratory syndrome. *J Infect Dis* 2004; **190**: 515–8.

22 Ahmed SF, Quadeer AA, McKay MR. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses* 2020; **12**: 254.

23 Lee CH, Koohy H. In silico identification of vaccine targets for 2019-nCoV. *F1000Res* 2020; **9**: 145.

24 Nguyen A, David JK, Maden SK, *et al.* Human Leukocyte Antigen Susceptibility Map for Severe Acute Respiratory Syndrome Coronavirus 2. *Journal of Virology* 2020; **94**. DOI:10.1128/JVI.00510-20.

25 Enayatkhani M, Hasaniazad M, Faezi S, *et al.* Reverse vaccinology approach to design a novel multi-epitope vaccine candidate against COVID-19: an in silico study. *J Biomol Struct Dyn* 2021; **39**: 2857–72.

26 Bonaccorsi I, Carrega P, Rullo EV, *et al.* HLA-C*17 in COVID-19 patients: hints for associations with severe clinical outcome and cardiovascular risk. *Immunology Letters* 2021; published online April 24. DOI:10.1016/j.imlet.2021.04.007.

27 Langton DJ, Bourke SC, Lie BA, *et al.* The influence of HLA genotype on the severity of COVID-19 infection. *HLA* 2021; published online April 25. DOI:10.1111/tan.14284.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

28 Naemi FMA, Al-Adwani S, Al-Khatabi H, Al-Nazawi A. Association between the HLA genotype and the severity of COVID-19 infection among South Asians. *J Med Virol* 2021; published online April 8. DOI:10.1002/jmv.27003.

29 Initiative TC-1HG, Ganna A. Mapping the human genetic architecture of COVID-19 by worldwide meta-analysis. *medRxiv* 2021; : 2021.03.10.21252820.

30 Gragert L, Madbouly A, Freeman J, Maiers M. Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry. *Human Immunology* 2013; **74**: 1313–20.

31 Pappas DJ, Marin W, Hollenbach JA, Mack SJ. Bridging ImmunoGenomic Data Analysis Workflow Gaps (BIGDAWG): An integrated case-control analysis pipeline. *Human Immunology* 2016; **77**: 283–7.

32 Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995; **310**: 170.

33 Rakib A, Sami SA, Mimi NJ, *et al.* Immunoinformatics-guided design of an epitope-based vaccine against severe acute respiratory syndrome coronavirus 2 spike glycoprotein. *Comput Biol Med* 2020; **124**: 103967.

34 Dar HA, Waheed Y, Najmi MH, *et al.* Multiepitope Subunit Vaccine Design against COVID-19 Based on the Spike Protein of SARS-CoV-2: An In Silico Analysis. *J Immunol Res* 2020; **2020**: 8893483.

35 Yadav PD, Potdar VA, Choudhary ML, *et al.* Full-genome sequences of the first two SARS-CoV-2 viruses from India. *Indian J Med Res* 2020; **151**: 200–9.

36 Baruah V, Bose S. Immunoinformatics-aided identification of T cell and B cell epitopes in the surface glycoprotein of 2019-nCoV. *J Med Virol* 2020; **92**: 495–500.

37 Crooke SN, Ovsyannikova IG, Kennedy RB, Poland GA. Immunoinformatic identification of B cell and T cell epitopes in the SARS-CoV-2 proteome. *Sci Rep* 2020; **10**: 14179–15.

38 Cun Y, Li C, Shi L, *et al.* COVID-19 coronavirus vaccine T cell epitope prediction analysis based on distributions of HLA class I loci (HLA-A, -B, -C) across global populations. *Hum Vaccin Immunother* 2021; **17**: 1097–108.

39 Alam A, Khan A, Imam N, *et al.* Design of an epitope-based peptide vaccine against the SARS-CoV-2: a vaccine-informatics approach. *Brief Bioinform* 2021; **22**: 1309–23.

40 Waqas M, Haider A, Sufyan M, Siraj S, Sehgal SA. Determine the Potential Epitope Based Peptide Vaccine Against Novel SARS-CoV-2 Targeting Structural Proteins Using Immunoinformatics Approaches. *Front Mol Biosci* 2020; **7**: 227.

41 Panda PK, Arul MN, Patel P, *et al.* Structure-based drug designing and immunoinformatics approach for SARS-CoV-2. *Sci Adv* 2020; **6**: eabb8097.

42 Kumar N, Admane N, Kumari A, *et al.* Cytotoxic T-lymphocyte elicited vaccine against SARS-CoV-2 employing immunoinformatics framework. *Sci Rep* 2021; **11**: 7653–14.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

43 Kared H, Redd AD, Bloch EM, *et al.* CD8+ T cell responses in convalescent COVID-19 individuals target epitopes from the entire SARS-CoV-2 proteome and show kinetics of early differentiation. *bioRxiv* 2020; : 2020.10.08.330688.

44 Kumar V, Kancharla S, Kolli P, Jena M. Reverse vaccinology approach towards the in-silico multiepitope vaccine development against SARS-CoV-2. *F1000Res* 2021; **10**: 44.

45 Sette A, Vitiello A, Reherman B, *et al.* The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *The Journal of Immunology* 1994; **153**: 5586–92.

46 Ferretti AP, Kula T, Wang Y, *et al.* Unbiased Screens Show CD8+ T Cells of COVID-19 Patients Recognize Shared Epitopes in SARS-CoV-2 that Largely Reside outside the Spike Protein. *Immunity* 2020; **53**: 1095–1107.e3.

47 Lipsitch M, Grad YH, Sette A, Crotty S. Cross-reactive memory T cells and herd immunity to SARS-CoV-2. *Nat Rev Immunol* 2020; **20**: 709–13.

48 Aung S, Vittinghoff E, Nah G, *et al.* Characteristics and Behaviors Associated with Prevalent SARS-CoV-2 Infection. *Int J Gen Med* 2021; **14**: 1063–7.

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Table 1. Study population demographics

Discovery

Total	Symptomatic	Asymptomatic			
640	550	90			
	85.9%	14.1%			
	Male	Female		Age (median)	
All	0.20	0.80		38	
Symptomatic	0.19	0.81	NS	**37**	***p*<0.001**
Asymptomatic	0.28	0.72		**44**	

Replication

Total	Symptomatic	Asymptomatic			
788	742	46			
	94.2%	5.8%			
	Male	Female		Age (median)	
All	0.17	0.83		32	
Symptomatic	0.17	0.83	NS	**31**	***p*<0.001**
Asymptomatic	0.24	0.76		**38**	

Combined

Total	Symptomatic	Asymptomatic			
1428	1292	136			
	90.5%	9.5%			
	Male	Female		Age (median)	
All	0.19	0.81		34	
Symptomatic	**0.18**	**0.82**	***p*=0.02**	**33**	***p*<0.001**
Asymptomatic	**0.26**	**0.74**		**41**	

NS = non-significant; *p* = p-value

medRxiv preprint doi: https://doi.org/10.1101/2021.05.13.21257065; this version posted September 10, 2021. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.

It is made available under a CC-BY 4.0 International license .

Table 2. HLA-B*15:01 is associated with SARS-CoV-2 asymptomatic infection

	Asymptomatic	Symptomatic				
Discovery	***f***	***f***	**OR**	**95% CI**	***p***	p^{corr}
HLA-B*15:01	0.1111	0.0473	2.52	1.39-4.42	0.00056	0.017
HLA-B*15:01 (adjusted)			2.45	1.38-4.24	0.0016	0.048
Replication						
HLA-B*15:01 (adjusted)	0.1087	0.0512	2.32	1.10-4.43	0.017	
Combined						
HLA-B*15:01 (adjusted)	0.1103	0.0495	2.40	1.54-3.64	5.67×10^{-5}	
HLA-B*15:01/15:01 (adjusted)	0.0220	0.0050	8.58	1.74-34.43	0.001	

Adjusted = adjusted for age and sex; *f* = frequency; OR = odds ratio; CI = confidence interval; *p* = p-value; p^{corr} = p-value after Bonferroni correction