Retrospective *in silico* HLA predictions from COVID-19 patients reveal alleles associated with disease prognosis

René L. Warren and Inanç Birol

René L. Warren
Genome Sciences Centre, BC Cancer
Vancouver, BC, V5Z 4S6, Canada
rwarren@bcgsc.ca
570 W 7th Ave #100, Vancouver, BC V5Z 4S6
(604) 707-5900 #673235

*To whom correspondence should be addressed

Inanç Birol
Genome Sciences Centre, BC Cancer
Vancouver, BC, V5Z 4S6, Canada
ibirol@bcgsc.ca

Keywords (8): COVID-19, SARS-CoV-2, Human Leukocyte Antigen (HLA), RNA-seq, DPA1*02:02, A*11:01, C*04:01, DQA1*01:02

ABSTRACT

Background: The Human Leukocyte Antigen (HLA) gene locus plays a fundamental role in human immunity, and it is established that certain HLA alleles are disease determinants.

Methods: By combining the predictive power of multiple *in silico* HLA predictors, we have previously identified prevalent HLA class I and class II alleles, including DPA1*02:02, in two small cohorts at the COVID-19 pandemic onset. Since then, newer and larger patient cohorts with controls and associated demographic and clinical data have been deposited in public repositories. Here, we report on HLA-I and HLA-II alleles, along with their associated risk significance in one such cohort of 126 patients, including COVID-19 positive (n=100) and negative patients (n=26).

Results: We recapitulate an enrichment of DPA1*02:02 in the COVID-19 positive cohort (29%) when compared to the COVID-negative control group (Fisher's exact test [FET] p=0.0174). Having this allele, however, does not appear to put this cohort's patients at an increased risk of hospitalization. Inspection of COVID-19 disease severity outcomes reveal nominally significant risk associations with A*11:01 (FET p=0.0078), C*04:01 (FET p=0.0087) and DQA1*01:02 (FET p=0.0121).

Conclusions: While enrichment of these alleles falls below statistical significance after Bonferroni correction, COVID-19 patients with the latter three alleles tend to fare worse overall. This is especially evident for patients with C*04:01, where disease prognosis measured by mechanical ventilation-free days was statistically significant after multiple hypothesis correction (Bonferroni p = 0.0023), and may hold potential clinical value.

62 **INTRODUCTION**

63

64 Modern history has been plagued by deadly outbreaks, from the recurring influenza
65 (e.g. Spanish, Asian, Hong Kong, Avian) and HIV/AIDS viral pandemics, to bacterial
66 and protist infections causing tuberculosis and malaria. Since the early 2000s, we have
67 faced another threat: novel coronavirus infections causing severe respiratory illnesses
68 such as SARS, MERS and today, coronavirus disease 2019 − COVID-19 [1]. The
69 SARS-CoV-2 coronavirus responsible for the COVID-19 respiratory disease is of
70 particular concern; not only does SARS-CoV-2 spread quickly, the symptoms of its
71 infection, when exhibited, are very similar to that of the cold and flu making it difficult to
72 diagnose, trace and contain. Further, infected individuals are affected differently. For
73 instance, older men (≥65 years old) with pre-existing medical conditions, such as
74 diabetes, appear at increased risk of progressing into the more severe phase of the
75 disease, yet SARS-CoV-2 infections affect all other age groups evenly except
76 occasionally in children and adolescents [2]. Most peculiar is that a high proportion of
77 individuals who tested positive for SARS-CoV-2 are asymptomatic – as high as 43%
78 recorded in Iceland, a rate that appears to vary depending on jurisdictions and
79 populations [3-5]. As of now, the disparity in patient response to SARS-CoV-2 infection
80 is still eluding us.

81 The most efficient way to combat pathogens has been through the use of our
82 own defense mechanism: our acquired immunity. This is done by vaccination
83 campaigns that effectively prime our immune systems at the population level before we
84 even encounter pathogens. But design of effective vaccines must consider interactions

85 with host immune genes. The Human Leukocyte Antigens (HLA) are a group of such
86 genes encoding surface receptors that bind short peptide epitopes derived from
87 endogenous (class I) or exogenous (class II) antigens, including viral antigens, and they
88 facilitate killer or helper T cells to set off an appropriate immune response. The
89 magnitude of this response varies between patients as populations and individuals have
90 different composition of HLA genes and variable T cell repertoires. As such, HLA
91 induces a bias, which is responsible for documented host susceptibility to disease [6].
92 Some of the notable associations between HLA and disease are observed in AIDS
93 patients, with certain HLA alleles conferring protection [7]. In other cases, HLA has been
94 implicated with autoimmune diseases and diabetes [8-11]. The exact underlying
95 mechanisms behind these associations are unclear, but there is mounting evidence that
96 bacterial and viral infection may be the trigger for some [10] and that HLA plays a critical
97 role in the viral infection cycle, including viral entry into host cells [12].
98 Since the beginning of the pandemic, worldwide reports have emerged on HLA
99 associations with COVID-19, including our own [13-19]. Using publicly available
100 metatranscriptomic sequencing data made available at the pandemic onset, we had
101 demonstrated the utility of a high throughput *in silico* method for characterizing the HLA
102 types of COVID-19 patients from bronchoalveolar lavage fluid and blood samples and
103 reported on prevalent alleles, including the DPA1*02:02P - DPB1*05:01P HLA-II
104 haplotype observed in 7 out of the 8 of patients from two small cohorts. Here, using
105 public RNA-seq sequencing data from a larger COVID-19 patient cohort with clinical
106 outcomes and demographics data [20], we report on HLA alleles with potential
107 diagnostic (DPA1*02:02) and prognostic (A*11:01 and C*04:01) value in 126

108 hospitalized patients with (n=100) and without (n=26) COVID-19 and present our
109 findings in light of available demographic characteristics using hospitalization and
110 disease severity metrics.

METHODS

We downloaded Illumina NOVASEQ-6000 paired-end (50 bp) RNA-Seq reads from libraries prepared from the blood samples of 126 hospitalized patients, with (n=100) or without COVID-19 (n=26) (ENA project: PRJNA660067, accessions: SRX9033799-SRX9033924). This data is part of a large-scale multi omics study from the Department of Molecular and Cellular Physiology, Albany Medical College, Albany, NY, USA, with aims to analyze COVID-19 Severity and clinical as well as demographics data was made available by the study authors [20] (GEO accession GSE157103). On each patient RNA-Seq dataset, we ran HLA prediction software OptiType [21] (v1.3.4), seq2HLA [22] (v2.3), and HLAminer [23] (v1.4 targeted assembly mode with defaults) as described [19]. We tallied HLA class I (HLA-I) and class II (HLA-II, supported by Seq2HLA and HLAminer only) allele predictions and for each patient we report the most likely HLA allele (4-digit resolution), indicating HLA predictor tool support (**Additional file 1, tables S1 and S2**).

Looking at class I and II alleles predicted in 10% or more of COVID-19 positive patients (class I, n=17; class II, n=11) we calculated Fisher's Exact Test (FET), first testing for enrichment in COVID-19 positive vs. negative patients (R function fisher.test, alternative = "greater"). For those same alleles (found in ≥10% patients) and inspecting only the COVID-19 positive cohort, we tested for the probability of patient hospitalization, as measured by the Intensive Care Unit (ICU) metric reported by the original study authors, using FET. We looked further into the risk of hospitalization in COVID-19 patients with vs. without these alleles using the Kaplan-Meier (KM) estimator (R library survival), plotting the probability of remission using the "hospital-free days

135 post 45 day followup (days)" (HFD-45) metric reported by the study author as a proxy
136 for disease severity, with lower HFD-45 numbers indicating worse outcomes. Similarly,
137 we ran the KM estimator using another metric of disease severity, "ventilator-free days",
138 which captures the most severe cases with COVID-19 patients suffering respiratory
139 deterioration and requiring mechanical ventilation. On each set we calculated the log-
140 rank p-value (R library survminer) and corrected for multiple hypothesis testing
141 (Bonferroni correction) using the number and patient abundance rank of class I (n=17)
142 or class II (n=11) HLA alleles observed in 10% or more of COVID-19 patients. We also
143 inspected the combined influence of HLA alleles and patient demographics data (age,
144 sex, ethnicity) on the hospitalization (ICU negative vs. positive) outcomes of COVID-19
145 patients, using odds-ratio calculations (R function fisher.test, and applying Haldane
146 correction [24] on zero values, when necessary).

RESULTS

We collated the HLA class I and class II predictions of three *in silico* HLA predictors derived from the RNA-seq samples of a recent [20] COVID-19 positive patient cohort (n=100) with control patients (n=26) who tested negative for COVID-19 (**Additional file 1, Tables S1 and S2**). Due to the limiting short read length (paired 50 bp) we chose to first report on OptiType [21] and seq2HLA [22] class I and class II predictions, and count the additional allele support from seq2HLA and HLAminer [23]. In all, we identify 17 and 11 HLA class I and class II alleles predicted in 10% or more of COVID-19 patients, respectively (**Tables 1 and 2**). There were many more alleles predicted (133 class I and 101 class II in the COVID-19 positive cohort), but too few patients are represented at lower cut-offs to compute meaningful statistics. First, we looked at the statistical enrichment (Fisher's Exact Test - FET) of each allele in the COVID-19 positive set, compared to the COVID-19 negative control group. We find HLA-I A*30:02 and HLA-II DPA1*02:02 allele enrichments nominally significant (FET p = 0.0417 and p = 0.0174) at the 5% level (**Tables 1 and 2**). However, when Bonferroni correction is applied for the number of HLA class I allele tests or when the abundance rank is factored in for A*30:02, the test is not significant (n=17, Bonferroni p = 0.7080; n=10, Bonferroni p = 0.4165, respectively). For HLA-II DPA1*02:02, Bonferroni correction finds the test insignificant at the α=0.05 level for the number of hypothesis, but significant for the allele abundance rank (n=11, Bonferroni p = 0.1916; n=2, Bonferroni p = 0.0348).

COVID-19 positive patients could be further stratified into those who were hospitalized and admitted to the Intensive Care Unit (**Tables I and 2**, ICU+), and those who were not (**Tables 1 and 2**, ICU-). When computing FET statistics, we find HLA-I

171 A*11:01 and C*04:01 and HLA-II DQA1*01:02 significant at the α=0.05 level (**Tables 1**
172 **and 2**; p = 0.0078, p = 0.0087 and 0.0121, respectively) but none remain significant
173 after Bonferroni correction. The Overmyer study authors [20] reported important disease
174 severity metrics (HFD-45 and days without needing mechanical ventilation), which we
175 used to assess the remission probability of COVID-19 patients having a specific allele
176 using Kaplan-Meier estimation. We find patients of the Overmyer cohort with either
177 A*11:01 (**Figure1a**), C*04:01 (**Figure 1b**) or DQA1*01:02 (**Figure 1c**) to be at a
178 significant increased risk of hospitalization (log-rank p = 0.0099, p = 0.0082 and p =
179 0.0097). When applying multiple test corrections to account for allele abundance rank,
180 only C*04:01 (n=5, Bonferroni p = 0.0410) and DQA1*01:02 (n=2, Bonferroni p =
181 0.0194) remained significant at the α=0.05 level. When looking at patients needing
182 mechanical ventilators, a severe outcome in COVID-19 disease progression, we only
183 find patients with C*04:01 to be at a statistically significant increased risk (**Figure 1d**,
184 log-rank p = 0.0019). Multiple hypothesis test correction retains the statistical
185 significance of this allele when factoring both the number of HLA-I alleles tested and
186 C*04:01 abundance rank (n=17, Bonferroni p = 0.0023; n=5, Bonferroni p = 0.0095).
187 Looking at the influence of the aforementioned alleles in combination with simple
188 demographics (sex, age and ethnicity), we find that of the Overmyer cohort patients with
189 the DPA1*02:02 allele, those with a white ethnic background and females appear at an
190 increased risk of testing positive for COVID-19 (**Figure 2**; odds ratio [OR] = 6.33 [5.33–
191 7.34], FET p = 0.0491 and OR = 7.33 [6.18-8.48], FET p = 0.0326, respectively). The
192 association with gender is also observed in alleles A*11:01, C*04:01 and DQA1:01:02,
193 putting female COVID-19 patients of this cohort at an increased risk of hospitalization

194 for the class I alleles (**Figure 2**; OR = 12.09 [10.41-13.76], FET p = 0.0105 for both) and
195 male COVID-19 patients at increased risk for DQA1*02:02 (**Figure 2**; OR = 2.74 [2.69-
196 2.79], FET p = 0.0481). For patients with the latter HLA-II allele, minorities and younger
197 individuals (<65 years old) are also more at risk of hospitalization (**Figure 2**; OR = 4.08
198 [2.85-5.32], FET p = 0.0222 and OR = 3.62 [2.42-4.83], FET p = 0.0240, respectively).
199 In this cohort, we also find patients with A*11:01 in the younger age group (<65 years
200 old) at increased risk of hospitalization (**Figure 2**; OR = 9.54 [8.14-10.94], FET p =
201 0.0184) whereas for those with C*04:01, it appears a white ethnic background and a
202 more advanced age (≥65 years old) may be predisposing to ICU hospitalization (**Figure**
203 **2**; OR = 14.25 [12.24-16.26], FET p = 0.0053 and OR = 9.66 [8.27-11.05], FET p =
204 0.0188, respectively).

205 **DISCUSSION**

206

207 We have previously identified the DPA1*02:02 class II allele as being prevalent in two
208 other and independent cohorts, with patients of undisclosed ethnic background, but
209 hospitalized in Wuhan, China [19]. Of populations with reported allele frequencies and
210 an appreciable sampling size (≥100 individuals), only Hong Kong Chinese and
211 Japanese have DPA1*02:02 allele frequencies (55.8% and 43.5%, respectively; [25,26])
212 above its observed frequency (29.0%) in the COVID-19 positive cohort analyzed herein.
213 The frequency of this allele in other qualifying populations tends to be generally lower,
214 including in South African (Worcester, 15.6%), Norwegians (14.0%), Mexico Chiapas
215 Lacandon Mayans (6.7%), United Kingdom Europeans (4.3%) and Spain Navarre
216 Basques (2.2%). We note that the ethnic background of the Overmyer *et al* [20] cohort
217 is heterogeneous, and white individuals (of unknown ancestry) represent 45.1% and
218 (11/29) 37.9% of the COVID-19 positive cohort and its DPA1*02:02 subset,
219 respectively. In contrast, Asians represent only a minority of the cohort (1.9%) and
220 DPA1*02:02 subset (1/29~3.4%). It is important to note that, in the Overmyer cohort,
221 DPA1*02:02 is not statistically associated with increased risk of hospitalization. The
222 significant enrichment of this allele in the COVID-19 positive vs. negative cohorts (FET
223 p= 0.0174) across all individuals, but also when looking only at females (FET p= 0.0326)
224 or white individuals (FET p= 0.0491), and not any other demographics, may prove an
225 important disease marker, which would need to be validated with additional datasets
226 and in independent studies.

227 There are reports of disease associations with DPA1*02:02, DQA1*01:02,
228 C*04:01 and A*11:01, but they are few. Of note, the association of all aforementioned
229 alleles with narcolepsy [10,11] and a known trigger for this auto-immune disease
230 includes upper-airway infections and influenza vaccinations [27-32]. Susceptibility to
231 narcolepsy may in fact be an indirect effect of HLA class I and the HLA class II DP
232 isotype in response to viral and bacterial infections, including from influenza and
233 *streptococcus* [10,27,33,34]. It has since been reported that HLA-A*11 may be a
234 susceptibility allele to influenza A(H1N1)pdm09 infection in some populations [35] while
235 another report implicates HLA-I allele C*04:01 with high HIV viral loads [36]. Further, it
236 was recently demonstrated that MHC class II DR, DQ and DP isotypes play a role in
237 mediating the cross-species entry of bat influenza viruses *in vitro* in human/animal cell
238 lines and in mice where engineered MHC-II deficiency made them resistant to upper-
239 respiratory tract infections [12]. It is therefore not a stretch to envision an involvement
240 from these HLA class II isotypes in controlling the cellular entry of a broader range of
241 viral agents *in vivo*.
242 In a recent study examining HLA susceptibility based on SARS-CoV-2 derived
243 peptide (epitope) binding strengths [37], the HLA-I allele A*11:01 was *in silico* predicted
244 to bind a large number of SARS-CoV-2 derived peptides (n=750) with varying affinity
245 [IC50 range 4.95 – 498.19, median = 149.62, mean = 182.28], and has been
246 experimentally validated to bind SARS-CoV-2 peptide GLMWLSYFV (Tables S4 and S7
247 in Nguyen *et al.* [37]). In contrast, C*04:01 was only predicted *in silico* to bind six SARS-
248 CoV-2 peptides and at higher IC50 ranges [167.65 – 469.30, median = 291.06, mean =

249 299.01] (Table S7 in Nguyen *et al.* [37]) suggesting a more limited ability to present
250 epitopes to T cells and mount an appropriate immune response.
251 There have been a number of reports published on HLA alleles − COVID-19
252 associations this past year, and on cohorts from many jurisdictions including China [18],
253 Italy [13,14,16] and the UK [17]. Wang *et al*. [18] compared the HLA allele frequencies
254 between a cohort of 82 Chinese individuals and a control population of bone marrow
255 donors previously studied by the same group. Novelli and co-workers [14] HLA typed a
256 cohort of 99 Italian COVID-19 patients, and associated the observed allele frequencies
257 with the HLA types in a reference group of 1,017 Italian individuals also previously
258 studied by the same group. Correale *et al*. [13] and Pisanti *et al.* [16] followed a different
259 strategy; these two independent studies leveraged population scale genomics data
260 retrieved from the Italian Bone-Marrow Donors Registry and the National Civil
261 Protection Department. They correlated background HLA allele frequency data with
262 mortality and morbidity rates across Italy to reach at starkly different conclusions on
263 which HLA alleles may play a role in disease etiology and progression. Disagreement
264 between these two studies (also distinct from the results of the other Italian study by
265 Novelli *et al*.) highlight the importance of large cohorts with matched samples to infer
266 the patient HLA alleles with better statistical significance. Poulton *et al*. [17]
267 characterized the HLA types of 80 COVID-19 patients in the UK on waiting lists for
268 transplantation, and compared observed allele frequencies in comparison to a cohort of
269 10,000 deceased organ donors and a separate cohort of 308 SARS-CoV-2-negative
270 individuals also on waiting lists for transplantation, the latter representing a matched
271 demographics for the COVID-19 patients in their cohort. Interestingly, this is the only

272 study that had any overlap between the alleles they flagged as being statistically
273 significant and the lists published by other studies cited above. Not surprisingly, the
274 alleles they listed do not intersect with the alleles identified and presented herein and
275 the three alleles we published earlier on a very small group of only eight patients. It is
276 nonetheless intriguing to find little to no HLA allele overlap between these reports,
277 including with those associated with the 2003 SARS outbreak, a related respiratory
278 disease caused by coronavirus [15,38,39]. This could be explained, at least partially, by
279 geographical differences and varying population allele frequencies in those cohorts,
280 relatively small cohort sizes (<100 patients), differences in experimentation setup and/or
281 other factors, including comorbidity status, that may be acting independently of HLA.

282 **CONCLUSIONS**

283

284 Here, we predict HLA-I and HLA-II alleles from publicly available COVID-19 patient
285 blood RNA-seq samples and identified several putative biomarkers. In a previous study,
286 we had observed one of these biomarkers (DPA1*02:02), and we postulate that patients
287 with the allele may have an increased susceptibility for COVID-19. Further, other alleles
288 (A11*01, C*04:01, DQA1*01:02) may be prognostic indicators of poor outcome.
289 However, although it is well established that patient HLA profiles play a significant role
290 in the onset and progression of infectious diseases in general, we caution against
291 drawing overreaching conclusions from regional, and often limited, observations. We
292 note that recently published studies associating HLA alleles and COVID-19, by and
293 large, disagree in their findings. We expect future studies with larger cohort sizes will
294 help bring a clearer picture of the role of patient HLA profiles, if any, in COVID-19
295 susceptibility and disease outcomes.

296 **LIST OF ABBREVIATIONS**
297
298
299 FET: Fisher's Exact Test; HFD-45: hospital-free days post 45 day followup (days); HLA:
300 human leukocyte antigen; ICU: Intensive Care Unit; KM: Kaplan-Meier

DECLARATIONS

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Availability of data and materials
The RNA-seq datasets analysed during the current study are available in the ENA repository https://www.ebi.ac.uk/ena/browser/view/PRJNA660067 accessions: SRX9033799- SRX9033924. The associated clinical data are available in the GEO repository https://www.ncbi.nlm.nih.gov/geo accession: GSE157103

Competing interests
The authors declare that they have no competing interests

Funding
This work was supported by Genome BC and Genome Canada [281ANV]; and the National Institutes of Health [2R01HG007182-04A1]. The content of this paper is solely the responsibility of the authors, and does not necessarily represent the official views of the National Institutes of Health or other funding organizations.

Authors' contributions
RLW designed the study, analyzed the data and wrote the manuscript. IB participated in the development of the study and co-wrote the manuscript.

Acknowledgements
Not applicable

REFERENCES

1. Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, et al. Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. Lancet Infect Dis. 2020;20:e238-e244.

2. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 2020;382:1708-20.

3. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic Population. N Engl J Med. 2020;382:2302-15.

4. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Euro Surveill. 2020;25:2000180.

5. Nishiura H, Kobayashi T, Miyama T, Suzuki A, Jung SM, Hayashi K, et al. Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). Int J Infect Dis. 2020;94:154-5.

6. Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. Nat Rev Immunol. 2018;18:325-39.

355 7. Goulder PJ, Walker BD. HIV and HLA class I: an evolving relationship. Immunity.
356 2012;37:426-40.
357
358 8. Chu X, Yang M, Song ZJ, Dong Y, Li C, Shen M, et al. Fine mapping MHC
359 associations in Graves' disease and its clinical subtypes in Han Chinese. J Med Genet.
360 2018;55:685-92.
361
362 9. Noble JA, Valdes AM, Bugawan TL, Apple RJ, Thomson G, Erlich HA. The HLA class
363 I A locus affects susceptibility to type 1 diabetes. Hum Immunol. 2002;63:657-64.
364
365 10. Ollila HM, Ravel JM, Han F, Faraco J, Lin L, Zheng X, et al. HLA-DPB1 and HLA
366 class I confer risk of and protection from narcolepsy. Am J Hum Genet. 2015;96:136-46.
367
368 11. Tafti M, Lammers GJ, Dauvilliers Y, Overeem S, Mayer G, Nowak J, et al.
369 Narcolepsy-Associated HLA Class I Alleles Implicate Cell-Mediated Cytotoxicity. Sleep.
370 2016;39:581-7.
371
372 12. Karakus U, Thamamongood T, Ciminski K, Ran W, Günther SC, Pohl MO, et al.
373 MHC class II proteins mediate cross-species entry of bat influenza viruses. Nature.
374 2019;567:109-12.
375

376 13. Correale P, Mutti L, Pentimalli F, Baglio G, Saladino RE, Sileri P, Giordano A. HLA-
377 B*44 and C*01 Prevalence Correlates with Covid19 Spreading across Italy. Int J Mol
378 Sci. 2020;21:5205.
379
380 14. Novelli A, Andreani M, Biancolella M, Liberatoscioli L, Passarelli C, Colona VL, et al.
381 HLA allele frequencies and susceptibility to COVID-19 in a group of 99 Italian patients.
382 HLA. 2020; doi:10.1111/tan.14047.
383
384 15. Ovsyannikova IG, Haralambieva IH, Crooke SN, Poland GA, Kennedy RB. The role
385 of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility
386 and severity. Immunol Rev. 2020;296:205-19.
387
388 16. Pisanti S, Deelen J, Gallina AM, Caputo M, Citro M, Abate M, et al. Correlation of
389 the two most frequent HLA haplotypes in the Italian population to the differential
390 regional incidence of Covid-19. J Transl Med. 2020;18:352.
391
392 17. Poulton K, Wright P, Hughes P, Savic S, Welberry Smith M, Guiver M, et al. A role
393 for human leucocyte antigens in the susceptibility to SARS-Cov-2 infection observed in
394 transplant patients. Int J Immunogenet. 2020;47:324-8.
395
396 18. Wang W, Zhang W, Zhang J, He J, Zhu F. Distribution of HLA allele frequencies in
397 82 Chinese individuals with coronavirus disease-2019 (COVID-19). HLA. 2020;96:194-
398 6.

399

400 19. Warren RL, Birol I. HLA predictions from the bronchoalveolar lavage fluid and blood
401 samples of eight COVID-19 patients at the pandemic onset. Bioinformatics. 2020;
402 doi:10.1093/bioinformatics/btaa756.
403

404 20. Overmyer KA, Shishkova E, Miller IJ, Balnis J, Bernstein MN, Peters-Clarke TM, et
405 al. Large-Scale Multi-omic Analysis of COVID-19 Severity. Cell Syst. 2020;
406 doi:10.1016/j.cels.2020.10.003.
407

408 21. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType:
409 precision HLA typing from next-generation sequencing data. Bioinformatics.
410 2014;30:3310-6.
411

412 22. Boegel S, Löwer M, Schäfer M, Bukur T, de Graaf J, Boisguérin V, et al. HLA typing
413 from RNA-Seq sequence reads. Genome Med. 2012;4:102.
414

415 23. Warren RL, Choe G, Freeman DJ, Castellarin M, Munro S, Moore R, Holt RA.
416 Derivation of HLA types from shotgun sequence datasets. Genome Med. 2012;4:95.
417

418 24. Valenzuela, C. Dos soluciones para la estimación de "odds ratios" con ceros [2
419 solutions for estimating odds ratios with zeros]. Rev Med Chil. 1993;121:1441-4.
420

421 25. Gonzalez-Galarza FF, McCabe A, Santos EJ, Jones J, Takeshita LY, Ortega-Rivera
422 ND, et al. Allele frequency net database (AFND) 2020 update: gold-standard data
423 classification, open access genotype data and new query tools. Nucleic Acid Research.
424 2020;48:D783-8.
425
426 26. Allele frequency net database. http://allelefrequencies.net/
427
428 27. Han F, Lin L, Warby SC, Faraco J, Li J, Dong SX, et al. Narcolepsy onset is
429 seasonal and increased following the 2009 H1N1 pandemic in China. Ann Neurol.
430 2011;70:410-7.
431
432 28. Han F, Faraco J, Dong XS, Ollila HM, Lin L, Li J, et al. Genome wide analysis of
433 narcolepsy in China implicates novel immune loci and reveals changes in association
434 prior to versus after the 2009 H1N1 influenza pandemic. PLoS Genet.
435 2013;9:e1003880.
436
437 29. Heier MS, Gautvik KM, Wannag E, Bronder KH, Midtlyng E, Kamaleri Y, Storsaeter
438 J. Incidence of narcolepsy in Norwegian children and adolescents after vaccination
439 against H1N1 influenza A. Sleep Med. 2013;14:867-71.
440
441 30. Moss AJ, Gaughran FP, Karasu A, Gilbert AS, Mann AJ, Gelder CM, et al.
442 Correlation between human leukocyte antigen class II alleles and HAI titers detected
443 post-influenza vaccination. PLoS One. 2013;8:e71376.

444

445 31. Montplaisir J, Petit D, Quinn MJ, Ouakki M, Deceuninck G, Desautels A, et al. Risk
446 of narcolepsy associated with inactivated adjuvanted (AS03) A/H1N1 (2009) pandemic
447 influenza vaccine in Quebec. PLoS One. 2014;9:e108489.

448

449 32. Wijnans L, Lecomte C, de Vries C, Weibel D, Sammon C, Hviid A, et al. The
450 incidence of narcolepsy in Europe: before, during, and after the influenza
451 A(H1N1)pdm09 pandemic and vaccination campaigns. Vaccine. 2013;31:1246-54.

452

453 33. Aran A, Lin L, Nevsimalova S, Plazzi G, Hong SC, Weiner K, et al. Elevated anti-
454 streptococcal antibodies in patients with recent narcolepsy onset. Sleep. 2009;32:979-
455 83.

456

457 34. Mahlios J, De la Herrán-Arita AK, Mignot E. The autoimmune basis of narcolepsy.
458 Curr Opin Neurobiol. 2013;23:767-73.

459

460 35. Dutta M, Dutta P, Medhi S, Borkakoty B, Biswas D. Polymorphism of HLA class I
461 and class II alleles in influenza A(H1N1)pdm09 virus infected population of Assam,
462 Northeast India. J Med Virol. 2018;90:854-60.

463

464 36. Olvera A, Pérez-Álvarez S, Ibarrondo J, Ganoza C, Lama JR, Lucchetti A, et al. The
465 HLA-C*04: 01/KIR2DS4 gene combination and human leukocyte antigen alleles with
466 high population frequency drive rate of HIV disease progression. AIDS. 2015;29:507-17.

467

468 37. Nguyen A, David JK, Maden SK, Wood MA, Weeder BR, Nellore A, Thompson RF.
469 Human Leukocyte Antigen Susceptibility Map for Severe Acute Respiratory Syndrome
470 Coronavirus 2. J Virol. 2020;94:e00510-20.

471

472 38. Lin M, Tseng HK, Trejaut JA, Lee HL, Loo JH, Chu CC, et al. Association of HLA
473 class I with severe acute respiratory syndrome coronavirus infection. BMC Med Genet.
474 2003;4:9.

475

476 39. Sanchez-Mazas A. HLA studies in the context of coronavirus outbreaks. Swiss Med
477 Wkly. 2020;150:w20248.

478 **TABLES**

479

480 **Table 1**. HLA-I alleles identified in 10% or more COVID-19 positive patients and

481 statistical tests of enrichment in the Overmyer *et al.* [20] COVID-19 positive (vs.

482 negative) cohort and association with hospitalization. Red font indicates significant

483 associations (Fisher's Exact Test) not corrected for multiple hypothesis tests.

HLA-I	COVID+ patients	ICU- patients	ICU+ patients	COVID- patients	H1: Enriched in COVID+ Fisher's Exact Test p-value	H1: Increased risk of hospitalization Fisher's Exact Test p-value
*A*02:01*	30	15	15	12	0.9614	0.5862
*A*24:02*	23	12	11	6	0.6161	0.5000
*C*07:02*	22	9	13	7	0.7890	0.8865
*C*07:01*	18	6	12	7	0.8991	0.9668
*C*04:01*	18	4	14	3	0.3231	0.0087
*C*06:02*	16	8	8	6	0.8708	0.6071
*B*51:01*	16	6	10	2	0.2292	0.9143
*C*03:04*	14	4	10	3	0.5171	0.9796
*C*15:02*	14	6	8	1	0.1363	0.8060
*A*01:01*	13	5	8	7	0.9742	0.8832
*A*03:01*	13	7	6	5	0.8680	0.5000
*A*30:02*	13	7	6	0	0.0417	0.5000
*B*07:02*	12	5	7	5	0.8969	0.8217
*A*30:01*	12	8	4	1	0.2016	0.1783
*B*44:02*	11	4	7	4	0.8320	0.9001
*A*68:01*	11	5	6	0	0.0697	0.7377
*A*11:01*	10	1	9	2	0.5320	0.0078

484

485

486

487 **Table 2**. HLA-II alleles identified in 10% or more COVID-19 positive patients and

488 statistical tests of enrichment in the Overmyer *et al.* [20] COVID-19 positive (vs.

489 negative) cohort and association with hospitalization. Red font indicates significant

490 associations (Fisher's Exact Test) not corrected for multiple hypothesis tests.

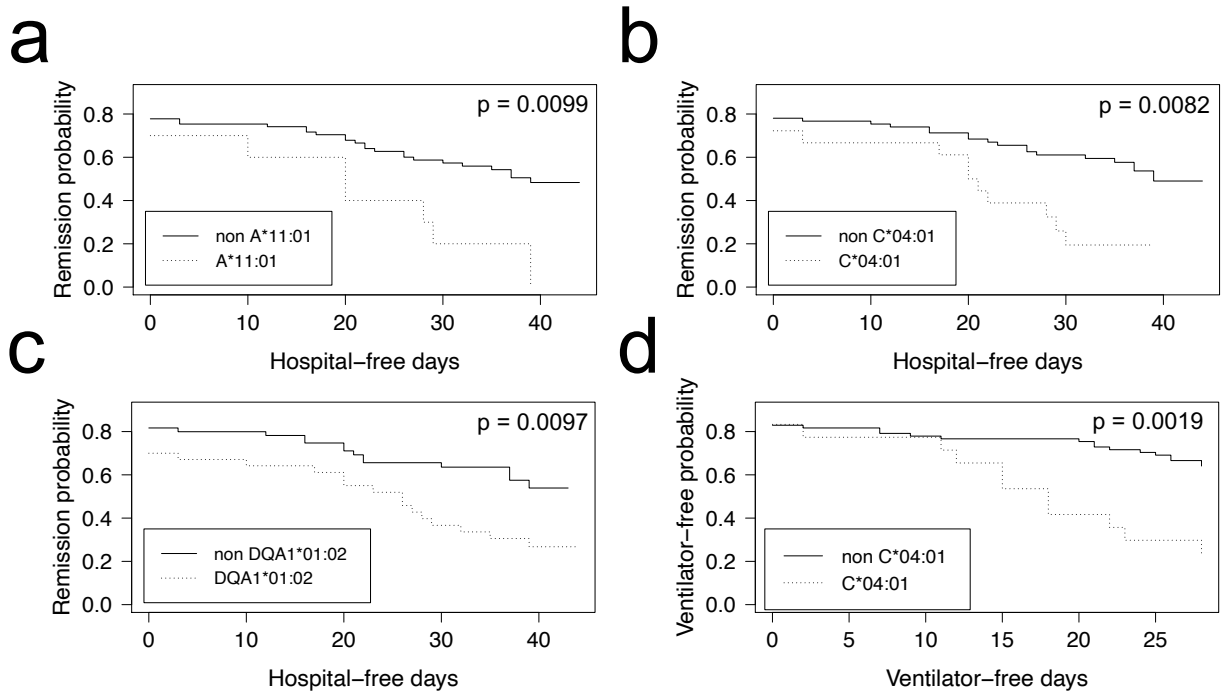
HLA-II	COVID+ patients	ICU- patients	ICU+ patients	COVID- patients	H1: Enriched in COVID+ Fisher's Exact Test p-value	H1: Increased risk of hospitalization Fisher's Exact Test p-value
*DPA1*01:03*	46	21	25	17	0.9812	0.8421
*DQA1*01:02*	40	26	14	10	0.5363	0.0121
*DPB1*04:01*	22	9	13	10	0.9726	0.8865
*DPA1*02:02*	29	16	13	2	0.0174	0.3299
*DPA1*02:01*	22	11	11	6	0.6578	0.5952
*DQA1*05:02*	19	9	10	3	0.2824	0.6945
*DPB1*01:01*	19	10	9	3	0.2824	0.5000
*DQB1*06:11*	11	6	5	3	0.6811	0.5000
*DRB1*13:01*	10	7	3	4	0.8689	0.1589
*DQB1*06:02*	10	6	4	3	0.7348	0.3703
*DRB1*15:01*	10	4	6	2	0.5320	0.8411

491

492

493

494 **FIGURES**



495

496 **Figure 1. HLA alleles associated with higher risk of hospitalization in a COVID-19**
497 **positive patient cohort.** COVID-19 positive patients were split into two groups per
498 allele tested, depending on whether they were predicted to have the HLA allele under
499 scrutiny or not. We ran the Kaplan-Meier estimator (R package survival) using the
500 Overmyer *et al.* cohort [20] HFD-45 metric for estimating the remission probability of
501 patients without or with alleles (a) A*11:01, (b) C*04:01 or (c) DQA1*01:02 and
502 mechanical ventilator-free days to estimate the statistical significance of the more
503 severe disease outcome observed in COVID-19 patients with (d) C*04:01. Log-rank p
504 values were calculated for each (R package survminer) and are indicated on the plots.

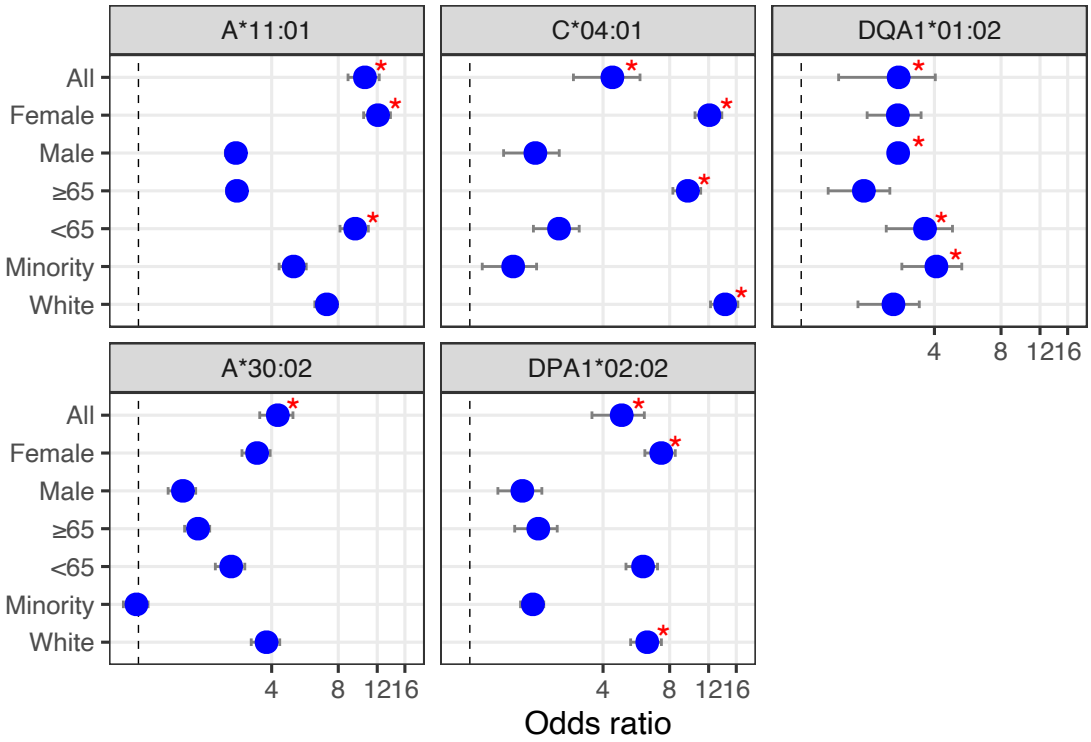


Figure 2. HLA alleles – demographics combinations with diagnostic (bottom) or prognostic (top) potential in a COVID-19 cohort. We calculated the odds ratio (OR) for each HLA-I and HLA-II alleles observed in 10% or more of patients, and plotted OR and the influence of demographics for HLA alleles showing significant associations (from **Tables 1 and 2**). First, looking at the influence of demographic characteristics such as sex (male/female), age (65 years old or above/less than 65 years old) and ethnicity (minority/white ethnic background) on the susceptibility of patients with these alleles to test positive for COVID-19 (lower two panels), and on the risk associated with ICU hospitalization (upper three panels). Red asterisks indicate significant demographic characteristics (Fisher's Exact Test) not corrected for multiple hypothesis tests.