

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38

## **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](mailto: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](mailto:ibirol@bcgsc.ca)

39 **ABSTRACT**

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

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

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

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

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

## 63 INTRODUCTION

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

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

86 with host immune genes. The Human Leukocyte Antigens (HLA) are a group of such  
87 genes encoding surface receptors that bind short peptide epitopes derived from  
88 endogenous (class I) or exogenous (class II) antigens, including viral antigens, and they  
89 facilitate killer or helper T cells to set off an appropriate immune response. The  
90 magnitude of this response varies between patients as populations and individuals have  
91 different composition of HLA genes and variable T cell repertoires. As such, HLA  
92 induces a bias, which is responsible for documented host susceptibility to disease [6].  
93 Some of the notable associations between HLA and disease are observed in AIDS  
94 patients, with certain HLA alleles conferring protection [7]. In other cases, HLA has been  
95 implicated with autoimmune diseases and diabetes [8-11]. The exact underlying  
96 mechanisms behind these associations are unclear, but there is mounting evidence that  
97 bacterial and viral infection may be the trigger for some [10] and that HLA plays a critical  
98 role in the viral infection cycle, including viral entry into host cells [12].

99         Since the beginning of the pandemic, worldwide reports have emerged on HLA  
100 associations with COVID-19, including our own [13-19]. Using publicly available  
101 metatranscriptomic sequencing data made available at the pandemic onset, we had  
102 demonstrated the utility of a high throughput *in silico* method for characterizing the HLA  
103 types of COVID-19 patients from bronchoalveolar lavage fluid and blood samples and  
104 reported on prevalent alleles, including the DPA1\*02:02P - DPB1\*05:01P HLA-II  
105 haplotype observed in 7 out of the 8 of patients from two small cohorts. Here, using  
106 public RNA-seq sequencing data from a larger COVID-19 patient cohort with clinical  
107 outcomes and demographics data [20], we report on HLA alleles with potential  
108 diagnostic (DPA1\*02:02) and prognostic (A\*11:01, C\*04:01 and DQA1\*01:02) value in

109 126 hospitalized patients with (n=100) and without (n=26) COVID-19 and present our  
110 findings in light of available demographic characteristics using hospitalization and  
111 disease severity metrics.

## 112 **METHODS**

113

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

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

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

148 **RESULTS**

149

150 We collated the HLA class I and class II predictions of three *in silico* HLA predictors

151 derived from the RNA-seq samples of a recent [20] COVID-19 positive patient cohort

152 (n=100) with control patients (n=26) who tested negative for COVID-19 (**Additional file**

153 **1, Tables S1 and S2**). Due to the limiting short read length (paired 50 bp) we chose to

154 first report on OptiType [21] and seq2HLA [22] class I and class II predictions, and count

155 the additional allele support from seq2HLA and HLAMiner [23]. In all, we identify 17 and

156 11 HLA class I and class II alleles predicted in 10% or more of COVID-19 patients,

157 respectively (**Tables 1 and 2**). There were many more alleles predicted (133 class I and

158 101 class II in the COVID-19 positive cohort), but too few patients are represented at

159 lower cut-offs to compute meaningful statistics. First, we looked at the statistical

160 enrichment (Fisher's Exact Test - FET) of each allele in the COVID-19 positive set,

161 compared to the COVID-19 negative control group. We find HLA-I A\*30:02 and HLA-II

162 DPA1\*02:02 allele enrichments nominally significant (FET  $p = 0.0417$  and  $p = 0.0174$ )

163 at the 5% level (**Tables 1 and 2**). However, when Bonferroni correction is applied for

164 the number of HLA class I allele tests or when the abundance rank is factored in for

165 A\*30:02, the test is not significant (n=17, Bonferroni  $p = 0.7080$ ; n=10, Bonferroni  $p =$

166  $0.4165$ , respectively). For HLA-II DPA1\*02:02, Bonferroni correction finds the test

167 insignificant at the  $\alpha=0.05$  level for the number of hypothesis, but significant for the

168 allele abundance rank (n=11, Bonferroni  $p = 0.1916$ ; n=2, Bonferroni  $p = 0.0348$ ).

169 COVID-19 positive patients could be further stratified into those who were

170 hospitalized and admitted to the Intensive Care Unit (**Tables 1 and 2**, ICU+), and those

171 who were not (**Tables 1 and 2**, ICU-). When computing FET statistics, we find HLA-I

172 A\*11:01 and C\*04:01 and HLA-II DQA1\*01:02 significant at the  $\alpha=0.05$  level (**Tables 1**  
173 **and 2**;  $p = 0.0078$ ,  $p = 0.0087$  and  $0.0121$ , respectively) but none remain significant  
174 after Bonferroni correction. The Overmyer study authors [20] reported important disease  
175 severity metrics (HFD-45 and days without needing mechanical ventilation), which we  
176 used to assess the remission probability of COVID-19 patients having a specific allele  
177 using Kaplan-Meier estimation. We find patients of the Overmyer cohort with either  
178 A\*11:01 (**Figure 1a**), C\*04:01 (**Figure 1b**) or DQA1\*01:02 (**Figure 1c**) to be at a  
179 significant increased risk of hospitalization (log-rank  $p = 0.0099$ ,  $p = 0.0082$  and  $p =$   
180  $0.0097$ ). When applying multiple test corrections to account for allele abundance rank,  
181 only C\*04:01 ( $n=5$ , Bonferroni  $p = 0.0410$ ) and DQA1\*01:02 ( $n=2$ , Bonferroni  $p =$   
182  $0.0194$ ) remained significant at the  $\alpha=0.05$  level. When looking at patients needing  
183 mechanical ventilators, a severe outcome in COVID-19 disease progression, we only  
184 find patients with C\*04:01 to be at a statistically significant increased risk (**Figure 1d**,  
185 log-rank  $p = 0.0019$ ). Multiple hypothesis test correction retains the statistical  
186 significance of this allele when factoring both the number of HLA-I alleles tested and  
187 C\*04:01 abundance rank ( $n=17$ , Bonferroni  $p = 0.0023$ ;  $n=5$ , Bonferroni  $p = 0.0095$ ).

188 Looking at the influence of the aforementioned alleles in combination with simple  
189 demographics (sex, age and ethnicity), we find that of the Overmyer cohort patients with  
190 the DPA1\*02:02 allele, those with a white ethnic background and females appear at an  
191 increased risk of testing positive for COVID-19 (**Figure 2**; odds ratio [OR] = 6.33 [5.33–  
192 7.34], FET  $p = 0.0491$  and OR = 7.33 [6.18-8.48], FET  $p = 0.0326$ , respectively). The  
193 association with gender is also observed in alleles A\*11:01, C\*04:01 and DQA1:01:02,  
194 putting female COVID-19 patients of this cohort at an increased risk of hospitalization

195 for the class I alleles (**Figure 2**; OR = 12.09 [10.41-13.76], FET p = 0.0105 for both) and  
196 male COVID-19 patients at increased risk for DQA1\*02:02 (**Figure 2**; OR = 2.74 [2.69-  
197 2.79], FET p = 0.0481). For patients with the latter HLA-II allele, minorities and younger  
198 individuals (<65 years old) are also more at risk of hospitalization (**Figure 2**; OR = 4.08  
199 [2.85-5.32], FET p = 0.0222 and OR = 3.62 [2.42-4.83], FET p = 0.0240, respectively).  
200 In this cohort, we also find patients with A\*11:01 in the younger age group (<65 years  
201 old) at increased risk of hospitalization (**Figure 2**; OR = 9.54 [8.14-10.94], FET p =  
202 0.0184) whereas for those with C\*04:01, it appears a white ethnic background and a  
203 more advanced age (≥65 years old) may be predisposing to ICU hospitalization (**Figure**  
204 **2**; OR = 14.25 [12.24-16.26], FET p = 0.0053 and OR = 9.66 [8.27-11.05], FET p =  
205 0.0188, respectively).

## 206 **DISCUSSION**

207

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

228           There are reports of disease associations with DPA1\*02:02, DQA1\*01:02,  
229 C\*04:01 and A\*11:01, but they are few. Of note, the association of all aforementioned  
230 alleles with narcolepsy [10,11] and a known trigger for this auto-immune disease  
231 includes upper-airway infections and influenza vaccinations [27-32]. Susceptibility to  
232 narcolepsy may in fact be an indirect effect of HLA class I and the HLA class II DP  
233 isotype in response to viral and bacterial infections, including from influenza and  
234 *streptococcus* [10,27,33,34]. It has since been reported that HLA-A\*11 may be a  
235 susceptibility allele to influenza A(H1N1)pdm09 infection in some populations [35] while  
236 another report implicates HLA-I allele C\*04:01 with high HIV viral loads [36]. Further, it  
237 was recently demonstrated that MHC class II DR, DQ and DP isotypes play a role in  
238 mediating the cross-species entry of bat influenza viruses *in vitro* in human/animal cell  
239 lines and in mice where engineered MHC-II deficiency made them resistant to upper-  
240 respiratory tract infections [12]. It is therefore not a stretch to envision an involvement  
241 from these HLA class II isotypes in controlling the cellular entry of a broader range of  
242 viral agents *in vivo*.

243           In a recent study examining HLA susceptibility based on SARS-CoV-2 derived  
244 peptide (epitope) binding strengths [37], the HLA-I allele A\*11:01 was *in silico* predicted  
245 to bind a large number of SARS-CoV-2 derived peptides (n=750) with varying affinity  
246 [IC50 range 4.95 – 498.19, median = 149.62, mean = 182.28], and has been  
247 experimentally validated to bind SARS-CoV-2 peptide GLMWLSYFV (Tables S4 and S7  
248 in Nguyen *et al.* [37]). In contrast, C\*04:01 was only predicted *in silico* to bind six SARS-  
249 CoV-2 peptides and at higher IC50 ranges [167.65 – 469.30, median = 291.06, mean =

250 299.01] (Table S7 in Nguyen *et al.* [37]) suggesting a more limited ability to present  
251 epitopes to T cells and mount an appropriate immune response.

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

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

283 **CONCLUSIONS**

284

285 Here, we predict HLA-I and HLA-II alleles from publicly available COVID-19 patient  
286 blood RNA-seq samples and identified several putative biomarkers. In a previous study,  
287 we had observed one of these biomarkers (DPA1\*02:02), and we postulate that patients  
288 with the allele may have an increased susceptibility for COVID-19. Further, other alleles  
289 (A\*11:01, C\*04:01, DQA1\*01:02) may be prognostic indicators of poor outcome.

290 However, although it is well established that patient HLA profiles play a significant role  
291 in the onset and progression of infectious diseases in general, we caution against  
292 drawing overreaching conclusions from regional, and often limited, observations. We  
293 note that recently published studies associating HLA alleles and COVID-19, by and  
294 large, disagree in their findings. We expect future studies with larger cohort sizes will  
295 help bring a clearer picture of the role of patient HLA profiles, if any, in COVID-19  
296 susceptibility and disease outcomes.

297 **LIST OF ABBREVIATIONS**

298

299

300 FET: Fisher's Exact Test; HFD-45: hospital-free days post 45 day followup (days); HLA:

301 human leukocyte antigen; ICU: Intensive Care Unit; KM: Kaplan-Meier

302 **DECLARATIONS**

303

304

305 **Ethics approval and consent to participate**

306 Not applicable

307

308 **Consent for publication**

309 Not applicable

310

311 **Availability of data and materials**

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

316

317 **Competing interests**

318 The authors declare that they have no competing interests

319

320 **Funding**

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

326

327 **Authors' contributions**

328 RLW designed the study, analyzed the data and wrote the manuscript. IB participated in  
329 the development of the study and co-wrote the manuscript. All authors have read and  
330 approved the manuscript.

331

332 **Acknowledgements**

333 Not applicable

334 **REFERENCES**

335

336 1. Petersen E, Koopmans M, Go U, Hamer DH, Petrosillo N, Castelli F, et al. Comparing

337 SARS-CoV-2 with SARS-CoV and influenza pandemics. *Lancet Infect Dis.*

338 2020;20:e238-e244.

339

340 2. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of

341 Coronavirus Disease 2019 in China. *N Engl J Med.* 2020;382:1708-20.

342

343 3. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL,

344 et al. Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J Med.*

345 2020;382:2302-15.

346

347 4. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic

348 proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond

349 Princess cruise ship, Yokohama, Japan, 2020. *Euro Surveill.* 2020;25:2000180.

350

351 5. Nishiura H, Kobayashi T, Miyama T, Suzuki A, Jung SM, Hayashi K, et al. Estimation

352 of the asymptomatic ratio of novel coronavirus infections (COVID-19). *Int J Infect Dis.*

353 2020;94:154-5.

354

355 6. Dendrou CA, Petersen J, Rossjohn J, Fugger L. HLA variation and disease. *Nat Rev*

356 *Immunol.* 2018;18:325-39.

- 357 7. Goulder PJ, Walker BD. HIV and HLA class I: an evolving relationship. *Immunity*.  
358 2012;37:426-40.  
359
- 360 8. Chu X, Yang M, Song ZJ, Dong Y, Li C, Shen M, et al. Fine mapping MHC  
361 associations in Graves' disease and its clinical subtypes in Han Chinese. *J Med Genet*.  
362 2018;55:685-92.  
363
- 364 9. Noble JA, Valdes AM, Bugawan TL, Apple RJ, Thomson G, Erlich HA. The HLA class  
365 I A locus affects susceptibility to type 1 diabetes. *Hum Immunol*. 2002;63:657-64.  
366
- 367 10. Ollila HM, Ravel JM, Han F, Faraco J, Lin L, Zheng X, et al. HLA-DPB1 and HLA  
368 class I confer risk of and protection from narcolepsy. *Am J Hum Genet*. 2015;96:136-46.  
369
- 370 11. Tafti M, Lammers GJ, Dauvilliers Y, Overeem S, Mayer G, Nowak J, et al.  
371 Narcolepsy-Associated HLA Class I Alleles Implicate Cell-Mediated Cytotoxicity. *Sleep*.  
372 2016;39:581-7.  
373
- 374 12. Karakus U, Thamamongood T, Ciminski K, Ran W, Günther SC, Pohl MO, et al.  
375 MHC class II proteins mediate cross-species entry of bat influenza viruses. *Nature*.  
376 2019;567:109-12.  
377

378 13. Correale P, Mutti L, Pentimalli F, Baglio G, Saladino RE, Sileri P, Giordano A. HLA-  
379 B\*44 and C\*01 Prevalence Correlates with Covid19 Spreading across Italy. Int J Mol  
380 Sci. 2020;21:5205.

381  
382 14. Novelli A, Andreani M, Biancolella M, Liberatoscioli L, Passarelli C, Colona VL, et al.  
383 HLA allele frequencies and susceptibility to COVID-19 in a group of 99 Italian patients.  
384 HLA. 2020; doi:10.1111/tan.14047.

385  
386 15. Ovsyannikova IG, Haralambieva IH, Crooke SN, Poland GA, Kennedy RB. The role  
387 of host genetics in the immune response to SARS-CoV-2 and COVID-19 susceptibility  
388 and severity. Immunol Rev. 2020;296:205-19.

389  
390 16. Pisanti S, Deelen J, Gallina AM, Caputo M, Citro M, Abate M, et al. Correlation of  
391 the two most frequent HLA haplotypes in the Italian population to the differential  
392 regional incidence of Covid-19. J Transl Med. 2020;18:352.

393  
394 17. Poulton K, Wright P, Hughes P, Savic S, Welberry Smith M, Guiver M, et al. A role  
395 for human leucocyte antigens in the susceptibility to SARS-Cov-2 infection observed in  
396 transplant patients. Int J Immunogenet. 2020;47:324-8.

397  
398 18. Wang W, Zhang W, Zhang J, He J, Zhu F. Distribution of HLA allele frequencies in  
399 82 Chinese individuals with coronavirus disease-2019 (COVID-19). HLA. 2020;96:194-  
400 6.

- 401
- 402 19. Warren RL, Birol I. HLA predictions from the bronchoalveolar lavage fluid and blood  
403 samples of eight COVID-19 patients at the pandemic onset. *Bioinformatics*. 2020;  
404 doi:10.1093/bioinformatics/btaa756.
- 405
- 406 20. Overmyer KA, Shishkova E, Miller IJ, Balnis J, Bernstein MN, Peters-Clarke TM, et  
407 al. Large-Scale Multi-omic Analysis of COVID-19 Severity. *Cell Syst*. 2020;  
408 doi:10.1016/j.cels.2020.10.003.
- 409
- 410 21. Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType:  
411 precision HLA typing from next-generation sequencing data. *Bioinformatics*.  
412 2014;30:3310-6.
- 413
- 414 22. Boegel S, Löwer M, Schäfer M, Bukur T, de Graaf J, Boisguérin V, et al. HLA typing  
415 from RNA-Seq sequence reads. *Genome Med*. 2012;4:102.
- 416
- 417 23. Warren RL, Choe G, Freeman DJ, Castellarin M, Munro S, Moore R, Holt RA.  
418 Derivation of HLA types from shotgun sequence datasets. *Genome Med*. 2012;4:95.
- 419
- 420 24. Valenzuela, C. Dos soluciones para la estimación de "odds ratios" con ceros [2  
421 solutions for estimating odds ratios with zeros]. *Rev Med Chil*. 1993;121:1441-4.
- 422

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

- 446
- 447 31. Montplaisir J, Petit D, Quinn MJ, Ouakki M, Deceuninck G, Desautels A, et al. Risk  
448 of narcolepsy associated with inactivated adjuvanted (AS03) A/H1N1 (2009) pandemic  
449 influenza vaccine in Quebec. *PLoS One*. 2014;9:e108489.
- 450
- 451 32. Wijnans L, Lecomte C, de Vries C, Weibel D, Sammon C, Hviid A, et al. The  
452 incidence of narcolepsy in Europe: before, during, and after the influenza  
453 A(H1N1)pdm09 pandemic and vaccination campaigns. *Vaccine*. 2013;31:1246-54.
- 454
- 455 33. Aran A, Lin L, Nevsimalova S, Plazzi G, Hong SC, Weiner K, et al. Elevated anti-  
456 streptococcal antibodies in patients with recent narcolepsy onset. *Sleep*. 2009;32:979-  
457 83.
- 458
- 459 34. Mahlios J, De la Herrán-Arita AK, Mignot E. The autoimmune basis of narcolepsy.  
460 *Curr Opin Neurobiol*. 2013;23:767-73.
- 461
- 462 35. Dutta M, Dutta P, Medhi S, Borkakoty B, Biswas D. Polymorphism of HLA class I  
463 and class II alleles in influenza A(H1N1)pdm09 virus infected population of Assam,  
464 Northeast India. *J Med Virol*. 2018;90:854-60.
- 465
- 466 36. Olvera A, Pérez-Álvarez S, Ibarrondo J, Ganoza C, Lama JR, Lucchetti A, et al. The  
467 HLA-C\*04: 01/KIR2DS4 gene combination and human leukocyte antigen alleles with  
468 high population frequency drive rate of HIV disease progression. *AIDS*. 2015;29:507-17.

469

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

473

474 38. Lin M, Tseng HK, Trejaut JA, Lee HL, Loo JH, Chu CC, et al. Association of HLA  
475 class I with severe acute respiratory syndrome coronavirus infection. BMC Med Genet.  
476 2003;4:9.

477

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

480 **TABLES**

481

482 **Table 1.** HLA-I alleles identified in 10% or more COVID-19 positive patients and  
 483 statistical tests of enrichment in the Overmyer *et al.* [20] COVID-19 positive (vs.  
 484 negative) cohort and association with hospitalization. Red font indicates significant  
 485 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

486

487

488

489 **Table 2.** HLA-II alleles identified in 10% or more COVID-19 positive patients and  
 490 statistical tests of enrichment in the Overmyer *et al.* [20] COVID-19 positive (vs.  
 491 negative) cohort and association with hospitalization. Red font indicates significant  
 492 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
<i>DPA1*01:03</i>	46	21	25	17	0.9812	0.8421
<i>DQA1*01:02</i>	40	26	14	10	0.5363	<b>0.0121</b>
<i>DPB1*04:01</i>	22	9	13	10	0.9726	0.8865
<i>DPA1*02:02</i>	29	16	13	2	<b>0.0174</b>	0.3299
<i>DPA1*02:01</i>	22	11	11	6	0.6578	0.5952
<i>DQA1*05:02</i>	19	9	10	3	0.2824	0.6945
<i>DPB1*01:01</i>	19	10	9	3	0.2824	0.5000
<i>DQB1*06:11</i>	11	6	5	3	0.6811	0.5000
<i>DRB1*13:01</i>	10	7	3	4	0.8689	0.1589
<i>DQB1*06:02</i>	10	6	4	3	0.7348	0.3703
<i>DRB1*15:01</i>	10	4	6	2	0.5320	0.8411

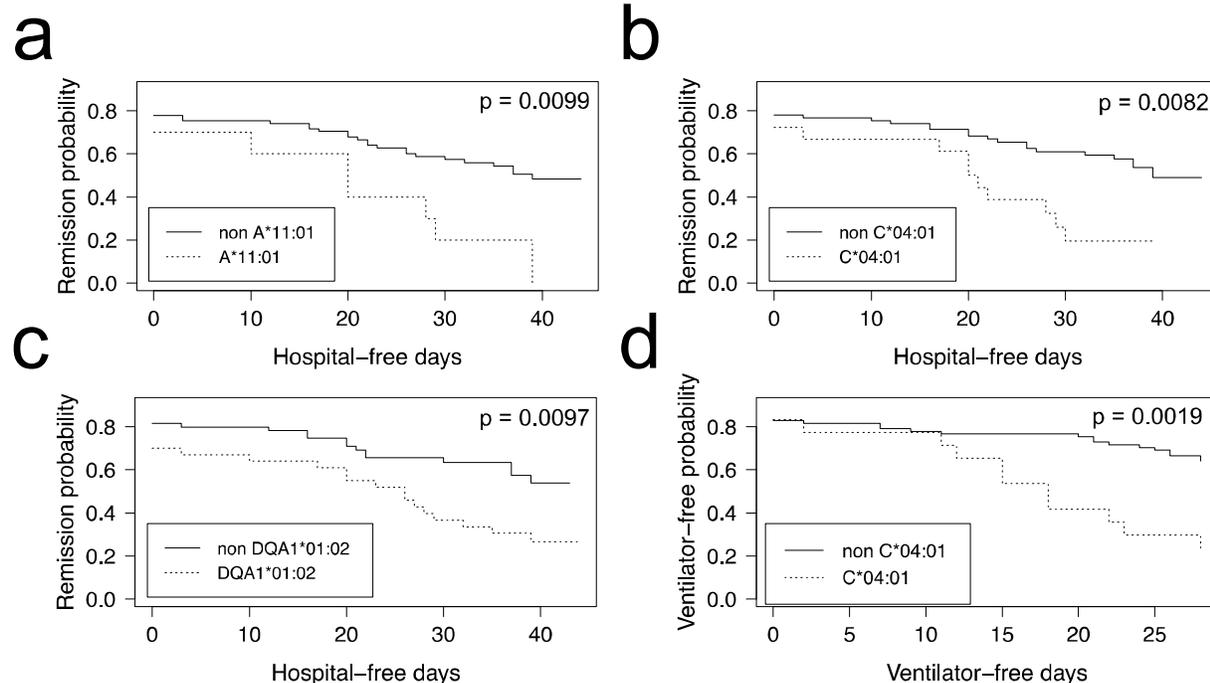
493

494

495

496

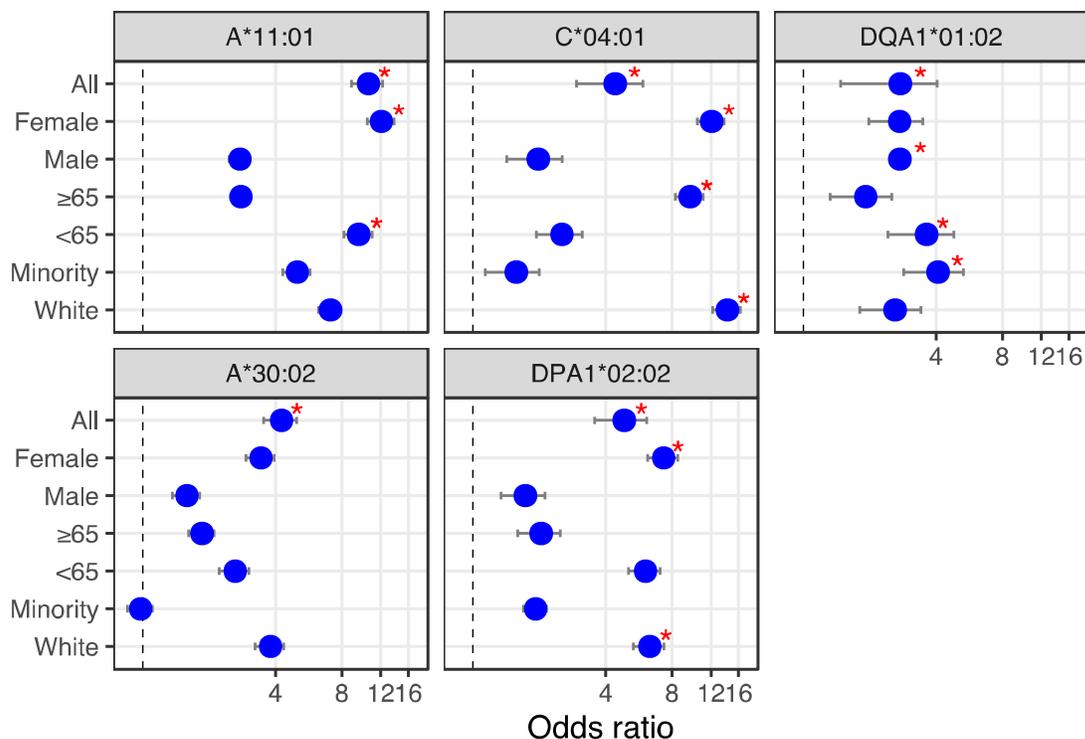
497 **FIGURES**



498

499 **Figure 1. HLA alleles associated with higher risk of hospitalization in a COVID-19**

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



508

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