

1 [Differential circulating proteomic responses associated with ancestry during severe COVID-19](#)  
2 [infection](#)

3 Authors: Thomas M Zheng<sup>1</sup>, Yann Ilboudo<sup>2</sup>, Tianyuan Lu<sup>6</sup>, Guillaume Butler-Laporte<sup>2</sup>, Tomoko  
4 Nakanishi<sup>2,3,7-9</sup>, David Morrison<sup>2</sup>, Darin Adra<sup>2</sup>, Lena Cuddeback<sup>10</sup>, J. Brent Richards<sup>1-5</sup>

5

6 Affiliations:

7 1. Quantitative Life Sciences, McGill University, Montréal, Quebec, Canada

8 2. Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Quebec,  
9 Canada

10 3. Departments of Medicine, Human Genetics, Epidemiology and Biostatistics, McGill  
11 University, Montréal, Quebec, Canada

12 4. Department of Twin Research, King's College London, London, United Kingdom

13 5. Five Prime Sciences Inc, Montréal, Québec, Canada

14 6. Department of Statistics, University of Toronto, Toronto, Ontario, Canada

15 7. Kyoto-McGill International Collaborative Program in Genomic Medicine, Graduate  
16 School of Medicine, Kyoto University, Kyoto, Japan

17 8. Department of Genome Informatics, Graduate School of Medicine, the University of  
18 Tokyo, Tokyo, Japan

19 9. Research Fellow, Japan Society for the Promotion of Science, Tokyo, Japan

20 10. SomaLogic, Inc., United States

21

22

23 Correspondence:

24 Corresponding author: Brent Richards, [brent.richards@mcgill.ca](mailto:brent.richards@mcgill.ca)

25

## 26 Abstract

27 **Background:** COVID-19 led to a disruption in nearly all aspects of society, yet these impacts  
28 were not the same across populations. During the pandemic, it became apparent that ancestry  
29 was associated with COVID-19 severity and morbidity, such that individuals of African descent  
30 tended to have worse outcomes than other populations. One factor that may influence COVID-  
31 19 outcomes is the circulating proteomic response to infection. This study examines how  
32 different ancestries had differential circulating protein levels in response to severe COVID-19  
33 infection.

34 **Methods:** 4,979 circulating proteins from 1,272 samples were measured using the SomaScan  
35 platform. We used a linear mixed model to assess the ancestry-specific association between the  
36 level of each protein and severe COVID-19 illness, accounting for sex, age, and days since  
37 symptom onset. We then compared each ancestry-specific effect size of severe COVID-19  
38 illness on protein level to one another in a pairwise manner to generate Z-scores. These Z-  
39 scores were then converted into p-values and corrected for multiple comparisons using a  
40 Benjamini-Hochberg false discovery rate of 5%.

41 **Results:** Comparing ancestries, we found that 62% of the tested proteins are associated with  
42 severe COVID-19 in European-ancestry individuals, compared to controls. We found that 45%  
43 and 22% of the tested proteins were different between COVID-19 infected and control  
44 individuals in people of African and East Asian ancestry, respectively. There was a strong  
45 correlation in effect size between ancestries. We found that individuals of European and African  
46 ancestry had the most similar response with a Pearson correlation of 0.868, 95% CI [0.861,  
47 0.875] while European and East Asian ancestries had a Pearson correlation of 0.645, 95% CI  
48 [0.628, 0.661] and, East Asian and African ancestries had a Pearson correlation of 0.709, 95%  
49 CI [0.695, 0.722]. However, we found 39 unique proteins that responded differently (FDR < 0.05)  
50 between the three ancestries.

51 **Conclusions:** Examining 4,979 protein levels in 1,272 samples, we identified that the majority of  
52 measured proteins had similar responses to infection across individuals of European, African  
53 and East Asian ancestry. However, there were 39 proteins that may have a differential response  
54 to infection, when stratified by ancestry. These proteins could be investigated to assess whether  
55 they explain the differences in observed severity of COVID-19 between ancestral populations.

56

57 **Keywords**

58 Proteomics, COVID-19, Ancestry, High-throughput Proteomics, BQC19, COVID-19 Response,  
59 Ancestry Differences, Linear Mixed Models

## 60 Background

61 The coronavirus disease 2019 (COVID-19) pandemic has impacted the entire world, reshaping  
62 our day-to-day lives forever. To date, there have been a total of 774 million reported cases of  
63 COVID-19 with reported 7 million deaths from COVID-19 globally.<sup>1</sup> In response to this pandemic,  
64 there have been attempts by the global scientific community to collect, categorize, and elucidate  
65 any information that could help with the prevention and treatment of this disease. Within these  
66 datasets, several risk factors have been identified, including, age, male sex, and past smoking  
67 status. However, one potential risk factor of COVID-19 infection that has not been as well  
68 characterized is ancestry.

69 Genetic ancestry is associated with different disease prevalence and COVID-19 is no exception.  
70 <sup>2-5</sup> Early on in the pandemic, it was noticed that people of different ancestries were impacted by  
71 COVID-19 differently.<sup>6-8</sup> In 2020, Millet et al. investigated COVID-19 cases and death by across  
72 the United States of America, stratified by county. They found that COVID-19 disproportionately  
73 impacted African American communities after adjusting for age, poverty, and other  
74 comorbidities.<sup>6</sup> Price-Haywood et al. investigated non-Hispanic African American patients and  
75 compared them to non-Hispanic European-ancestry patients in Louisiana. They found that even  
76 though the African American population accounted for 31% of the total cohort, this minority  
77 composed of 71% of the deaths. In response, there have been many studies that attempted to  
78 look at the underlying genetic differences to see if the difference in COVID-19 response was a  
79 genetic or socio-economic outcome.<sup>9-15</sup> Shelton et al. found that ancestry was a hospitalization  
80 risk factor, however, they reported that the two genetic associations they found were responsible  
81 for this risk difference.<sup>12</sup>

82 One method to examine the biological response to COVID-19 infection is by investigating the  
83 circulating proteins levels of individuals who are acutely ill and comparing these levels to those  
84 in controls. This comparison can identify proteins, whose levels tend to be perturbed by COVID-

85 19 infection. High-throughput oligonucleotide-aptamer protein measurement technology allows  
86 for the measurement of thousands of circulating proteins simultaneously from a single sample.  
87 In this study, samples were analyzed using the SomaScan assay (SomaLogic), an aptamer-  
88 based proteomics assay which utilizes chemically modified nucleotides, called SOMAmers  
89 (Slow Off-rate Modified Aptamers). The increased throughput of this technology allows for the  
90 measurement of many such proteins simultaneously. In spite of the benefits of proteomic  
91 analysis, there are only a few studies that looked at the proteomic differences between  
92 ancestries, and these ones focused primarily on the European ancestry.<sup>14,16</sup> Previously, we  
93 found that an isoform of *OAS1* in people of European ancestry displayed a protective effect for  
94 COVID-19.<sup>14</sup>

95 Given the difference in rates of COVID-19 between different ancestries, it is important to assess  
96 whether these differences are biological in nature. However, there has been very little research  
97 as to how the proteome is impacted by COVID-19 stratified by ancestry. This study, using the  
98 Biobanque Québécoise de la COVID-19 (BQC19), assesses how each continental ancestry  
99 responds to COVID-19, as reflected by changes in circulating protein levels. The results of this  
100 study shed light on the host response to infection via changes to the proteome, and describes  
101 which responses differ by ancestry.

## 102 [Methods](#)

### 103 [Overview of the study design](#)

104 We used the SomaScan V4 assay to measure 4,979 circulating proteins in case and control  
105 samples from the BQC19, a biobank in Québec, Canada.<sup>17</sup> We used linear mixed models  
106 (LMMs) to assess differences in protein levels between severe COVID-19 patients and controls  
107 who are hospitalized patients that either did not have COVID or presented with mild COVID

108 symptoms, stratified by ancestry. The differences in the association between protein levels and  
109 COVID-19 severity was then compared between ancestries.

## 110 Population

111 All samples were taken from the BQC19, this biobank recruited patients admitted to any of ten  
112 test sites in Quebec, Canada. We used genotype data provided by the BQC-19 to define the  
113 genetic ancestry of our populations and used phenotypic data provided by the BQC-19 to  
114 account for confounding factors, such as age, sex, and time since symptom onset.

## 115 COVID-19 Case/Control Definitions

116 We used severe COVID-19 to define case status. Cases were defined as hospitalized  
117 individuals with COVID-19 as the primary reason for hospital admission, an RT-PCR confirmed  
118 COVID-19 status, and either died or required respiratory support beyond low-throughput nasal  
119 cannula. Controls were all individuals in the BQC-19 who did not meet this case definition.  
120 These controls were also admitted to the same hospitals within the participating hospital  
121 network in Quebec, Canada.<sup>17</sup> This classification strategy has been previously used in the  
122 literature by the COVID-19 Host Genetics Initiative.<sup>9,10,14,15</sup>

## 123 Ancestry Determination

124 Genetic ancestry was determined by a UMAP (v 0.2.8.0) projection of the BQC19 onto a  
125 reference set (Supp. Fig 1). This reference ancestral set was constructed by unrelated samples  
126 from the 1000 Genomes Project and the Human Genome Diversity Project. This reference set  
127 contained 117,221 SNPs that were available in the BQC19, had a minor allele frequency  
128 of >0.1%, and were LD-pruned ( $r^2 < 0.8$ ). The BQC-19 population was then projected on this  
129 reference and ancestry was assigned by a UMAP cluster assigned by hdbscan (v 0.8.27). This  
130 method was previously used by the COVID-19 HGI.<sup>10,18</sup>

## 131 Protein measurements

132 We used the SomaScan (v4) platform to measure 5,284 circulating proteins from each sample.  
133 Of these 5,284 proteins, 305 non-human proteins were then removed from analysis using the  
134 same classification used previously in the literature.<sup>19</sup> To adequately reflect the effects of  
135 COVID-19 infection, we limited our experiment to those biological samples procured within 30  
136 days after symptom onset, with our linear mixed models accounting for the changes in protein  
137 levels over time. Blood plasma samples were collected in acid-citrate-dextrose tubes and frozen  
138 at -80 °C. Protein levels were measured using relative fluorescence units, and further  
139 normalized and calibrated by SomaLogic to remove any systematic bias. For the following  
140 statistical analysis, each protein sample across all patients was then normalized to have a mean  
141 of 0 and standard deviation of 1. This normalization was done prior to any ancestry stratification.  
142 The sample correction and proteomic normalization has been previously described in more  
143 detail by Butler-Laporte et al.<sup>20</sup>

## 144 Statistical Analysis

145 To leverage the BQC19's longitudinal samples, we used a linear mixed model (LMM) to group  
146 samples from the same patient using their BQC identifier. R packages, lme4 (1.1.35.1) and  
147 lmerTest (3.1.3) were used with R version 4.1.2 to generate these LMMs. This model further  
148 accounts for the patient's age, sex and when the sample was taken relative to when the patient  
149 first had symptoms. These patients were further stratified by ancestry to examine how COVID-  
150 19 impacts protein levels by ancestry. The formula and sample code can be found in  
151 [Supplementary file 1](#) and the GitHub ([https://github.com/richardslab/BQC\\_Ancestry\\_Proteomics](https://github.com/richardslab/BQC_Ancestry_Proteomics))  
152 respectively. A full table of all 4,979 proteins, their ancestry specific effect sizes, and their  
153 standard errors can be found in Additional File 1.

154 Each ancestry had all 4,979 proteins modelled in our linear mixed model. For each protein,  
155 severe COVID-19 illness had a reported effect size (beta) and p-value. This beta reports the  
156 difference between our case and control populations in each ancestry separately. These p-  
157 values were then corrected for multiple comparisons across all three models, resulting in a total  
158 of 14,937 tests with a false discovery rate of 5%.

159 To assess how COVID-19 impacts protein levels differently between ancestries we used both  
160 the beta and standard error from our LMM in each ancestry and derived a Z-score. A full formula  
161 can be found in [Supplementary File 1](#).

162 These Z-scores were then converted into p-values based on a Chi-squared distribution. These  
163 p-values were then adjusted for multiple comparisons by applying a Benjamini-Hochberg  
164 correction, setting a false discovery rate of 5%. This test was chosen instead of using a  
165 heterogeneity test due to bias that can be introduced when using a small number of studies.<sup>21</sup>  
166 We note that this threshold may be too liberal, given the relatively small sample sizes of non-  
167 European ancestry. Therefore, the results should be considered hypothesis generating.

## 168 [Results](#)

### 169 [Sample Characteristics](#)

170 [Table 1](#) shows the total number of samples used in our study were separated by genetic  
171 ancestry. The samples were split into the six major ancestry groups as used by the HGI and the  
172 “other” category. These groups were: “European”, “African”, “East Asian”, “South Asian”,  
173 “Admixed American”, “Middle Eastern”, and “Other”. For the remainder of this study, we focused  
174 only on our three largest ancestral groups: European, African, and East Asian. For patient  
175 characteristics, see [Supplementary Table 1](#).

176 [Table 1: Sample characteristics in the BQC19.](#)

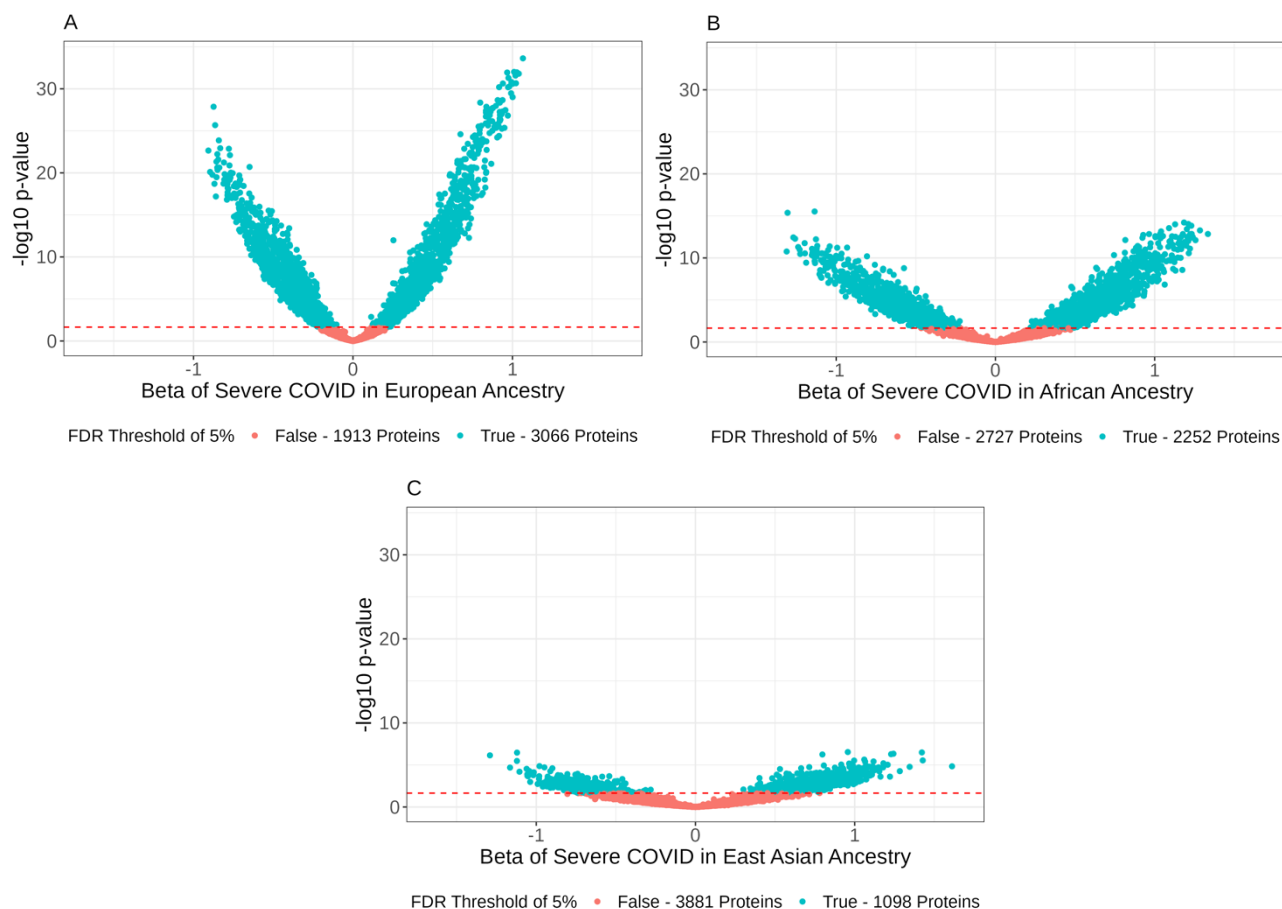


	European	African	East Asian	Middle Eastern	Admixed American	South Asian	Admixed/ Other
Overall <i>n</i> (% of entire sample)	797 (63%)	196 (15%)	79 (6%)	76 (6%)	51 (4%)	28 (2%)	45 (4%)
Median Age (IQR)	70 (58 – 82)	53 (43 – 65)	56 (48 – 70)	72 (54 – 80)	54 (40 – 65)	52 (38 – 62)	52 (32 – 64)
# of Females (%)	362 (45 %)	98 (50 %)	41 (52 %)	28 (37%)	22 (43%)	10 (36%)	22 (49%)
Severe COVID (%)	223 (28%)	72 (37%)	27 (34%)	31 (41%)	19 (37%)	7 (25%)	5 (11%)
Median Days since Symptom Onset (IQR)	9 (4 – 14)	9 (5 – 15)	9 (6 – 13)	7 (2 – 13)	12 (6 – 17)	8 (4 – 13)	8 (4 – 13)

177 Numbers are presented as count (percentage) or (IQR). IQR is the interquartile range, which is  
 178 the spread of the middle half of the distribution of the population.

179 [Different ancestries had different responses to COVID-19 infection](#)

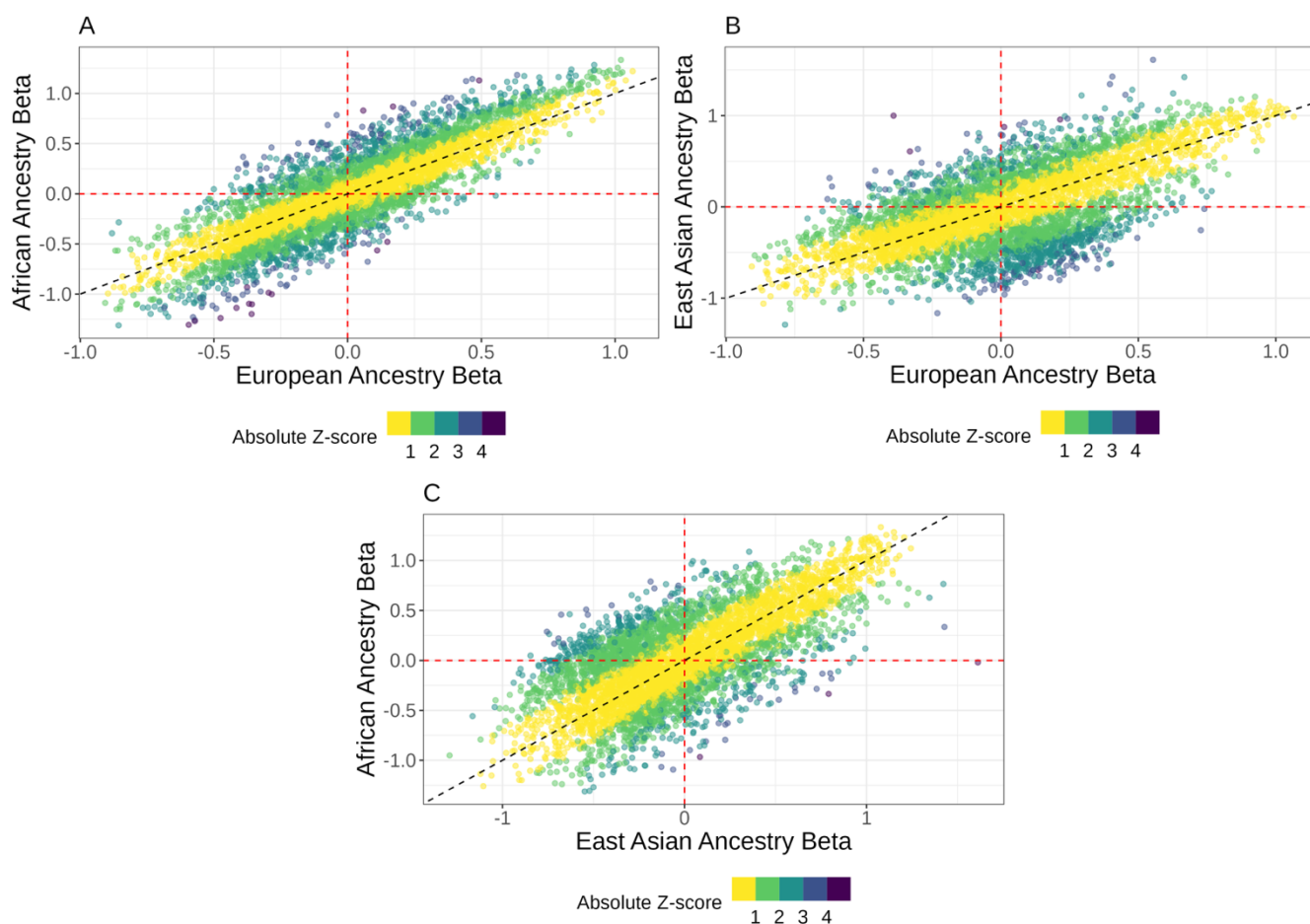
180 We first investigated how different ancestries protein levels responded to severe COVID-19 by  
 181 stratifying the population using the following ancestries: European, East Asian, and African. In  
 182 individuals of European ancestry 3,066/4,979 (62%) proteins had different (FDR < 0.05) proteins  
 183 levels when comparing case and control populations ([Figure 1A](#)). Amongst individuals of African  
 184 ancestry, 2,252/4,979 (45%) ([Figure 1B](#)) of the tested proteins had a difference in proteins levels  
 185 between cases and control while 1,098/4,979 (22%) of the tested proteins in Asian ancestry  
 186 individuals were classified as significantly different between our case and control populations.  
 187 ([Figure 1C](#)). We note that the larger samples (European ancestry participants) had more  
 188 different proteins than the smaller samples (African and Asian ancestry participants), which in  
 189 part, reflects statistical power to detect such differences.



190 Figure 1: Volcano plots of the three tested ancestries of the BQC19: European – n = 797 (A),  
191 African – n = 196 (B), and East Asian – n = 79 (C).

192 The x-axis represents the effect size of the COVID variable in our linear mixed model. Our linear  
193 mixed model included age, sex, and days since symptom onset as co-variables, and used  
194 patient ID as a random factor to account for intra-patient correlation. All volcano plots are scaled  
195 to the same axis, each dot represents one tested protein. The colours representing True and  
196 False, indicate if the protein is significantly different between our case and control definition after  
197 multiple-comparison correction using the Benjamini-Hochberg method. The red dashed line  
198 represents our cutoff  $p = 0.021$  to provide an FDR of 5%.

199 When comparing our European ancestries and our African ancestries, we saw that most  
200 proteins share the same trend upon severe COVID-19 infection. In [Figure 2A](#), we compared all  
201 4,979 of our tested protein's European effect sizes to their African effect sizes. We found that  
202 they had a Pearson correlation of 0.868, 95% CI [0.861, 0.875]. We saw a similar, if weaker  
203 trend when comparing our European and East Asian ancestries with a respective Pearson  
204 correlation of 0.645, 95% CI [0.628, 0.661] ([Figure 2B](#)). Finally, we found that the trend  
205 continued between our East Asian and African ancestries in [Figure 2C](#), with a Pearson



206 correlation of 0.709, 95% CI [0.695, 0.722].

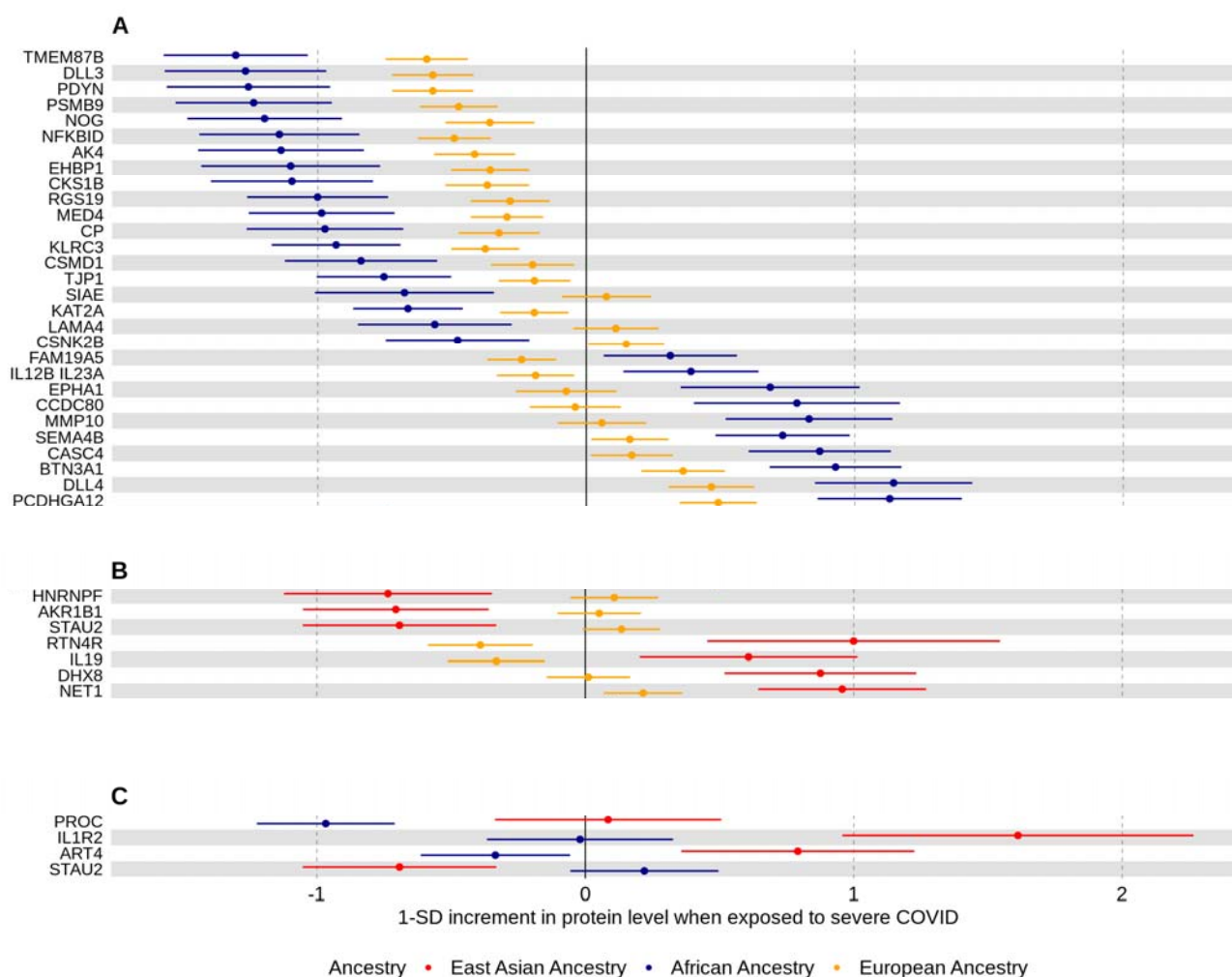
207 [Figure 2: Pairwise Beta-Beta plots which displays all 4979 proteins and their associated COVID](#)  
208 [beta.](#)

209 This effect size reflects the impact that COVID status has on the mean protein level in the tested  
210 ancestry. Each dot represents a single tested protein with its x-position and y-position  
211 representing the effect size per ancestry. In Figure 2A, the X position is the effect size  
212 associated with the European cohort and the Y position is the effect size in our African  
213 population. In Figure 2B, we are comparing the European and East Asian cohorts (x and y  
214 respectively), and in Figure 2C, we compare the East Asian (x-axis) and African effect sizes (y-  
215 axis). The red dashed lines indicate where an effect size of 0 would be and the black dashed  
216 line is the identity line, where a protein would have identical effect sizes in both ancestries.  
217 These plots do not show any information on the standard error from these effect sizes. By  
218 colouring the plots based on their Z-score, we observed that the farther away from the dashed  
219 line – the identity line, the higher the Z-score. The proteins with Z-scores of 3.8 and higher are  
220 of interest for future investigation for precise targets. We noted that very few proteins show a  
221 marked deviation from the line of identity. Yellow is for proteins with the same response between  
222 ancestries while dark purple shows proteins that have significantly different responses to COVID  
223 between ancestries.

#### 224 [Differences in how severe COVID-19 influences protein levels](#)

225 We then calculated the ancestry-specific effect size of how COVID-19 status affected protein  
226 levels; we compared each ancestry in a pairwise fashion. Statistical significance of between-  
227 ancestry differences were determined using a Z-test (Methods). These p-values were then  
228 corrected for the 14,937 comparisons using an FDR – Benjamini-Hochberg – threshold of 5%.  
229 After applying this multiple testing correction, we found that 39 unique proteins had different  
230 betas across the three ancestries. 29 proteins had different COVID-19 responses between our  
231 European and African cohorts, 7 proteins had different COVID-19 responses between our  
232 European and East Asian cohorts, and 4 proteins had a different COVID-19 response between  
233 our East Asian and African cohorts ([Supplemental Table 1](#)).

234 As shown in Figure 3, many of the proteins discovered displayed effect sizes in the same  
 235 direction but different magnitudes. Of note, in Figure 3A, 16 of the 29 proteins had effect sizes  
 236 that were both negative and 6 proteins had effect sizes that were both positive – with the  
 237 European counterpart consistently having the milder effect size, and the remaining seven  
 238 proteins had effect sizes with opposing magnitudes. Of the seven proteins shown in Figure 3B,  
 239 two proteins had COVID-19 responses in the same direction with different negative magnitudes,  
 240 and the other five proteins had opposing effect sizes dependant on ancestry. Finally, in Figure  
 241 3C, all four proteins had opposing effect sizes between our East Asian and African populations.



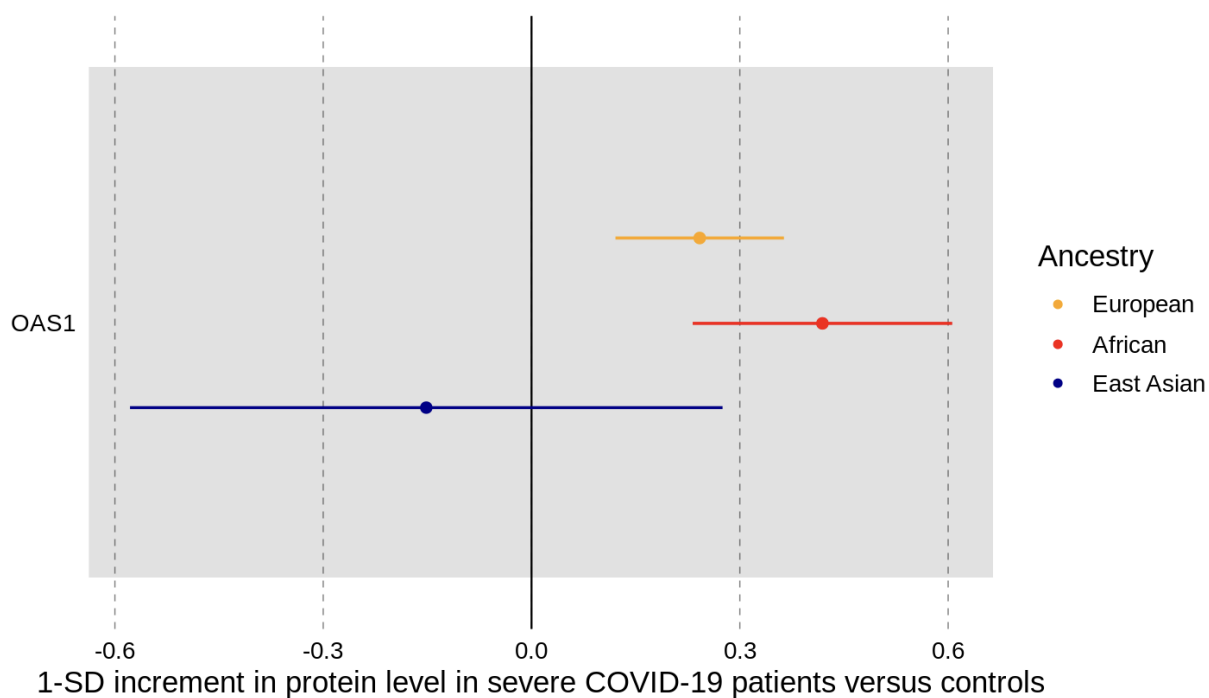
242 [Figure 3: Forest plots of all proteins that were significantly different between two of the three](#)  
243 [ancestries tested.](#)

244 Figure 3 shows Forest plots of the proteins that had a significant difference between ancestries  
245 (FDR < 0.05). [Fig. 3A](#) contains the 29 proteins that had significant differences in their effect  
246 sizes between our European and African ancestries and are displayed in order from lowest to  
247 highest African effect sizes. Of the 29 proteins, all were considered significantly different  
248 between case and control in the African ancestry while only 21 of them were considered to be  
249 significantly different from our control population in our European ancestry. [Fig. 3B](#), contains the  
250 7 proteins that passed our multiple correction threshold in our East Asian/European comparison  
251 is shown in order of their European effect sizes. Of the seven proteins, all seven had significant  
252 response to COVID-19 infection in our East Asian ancestry, while only three of the proteins  
253 showed a significant difference between our European case and control populations. [Fig. 3C](#)  
254 compares the East Asian and African effect sizes, in the order of lowest to highest African effect  
255 sizes. Two of the proteins were significantly different between our cases and controls in our  
256 African ancestry and three of our proteins had significant COVID-19 responses in our East  
257 Asian ancestry. In all plots the colour scheme is maintained: European (yellow), East Asian  
258 (Red), and African (blue). As in Figure 1, these changes are the effect sizes from the COVID  
259 variable of the linear mixed model. Error bars are the standard error of each effect size.

260 We noted that in all cases, when comparing differences in protein levels between two ancestries,  
261 the smaller of the two ancestries in sample size had the larger effect size. This may reflect the  
262 instability of estimating effect sizes in our smaller sample sizes, rather than true biological  
263 differences between the two groups.

264 [Validation of OAS1 results](#)

265 As previously mentioned, *OAS1* was identified by Zhou et al.<sup>14</sup> to be a circulating protein that  
266 caused less severe COVID-19. Their conclusions hypothesized that COVID-19 infection would  
267 cause *OAS1* levels to increase in European individuals. When we tested *OAS1* protein levels,  
268 we found that they were increased in our severe COVID-19 patients in both European and  
269 African ancestries (unadjusted p-value of  $1.0 \times 10^{-4}$  and  $2.7 \times 10^{-5}$  respectively) (Figure 4).  
270 However, since the test in East Asian ancestry individuals was not well-powered due to the  
271 small sample size, there was no detectable difference in *OAS1* levels between severe COVID-  
272 19 patients and controls (p-value of 0.49).



274 Figure 4: Forest plot of *OAS1* when comparing case and control status across all three tested  
275 ancestries.

276 The effect size of each model is represented by the dot with the horizontal lines indicating the  
277 standard error associated with the effect size.

278 Discussion

279 We leveraged the BQC19 to measure 4,979 circulating proteins across 1,272 samples from 909  
280 patients. We then stratified these measurements by genetic ancestry and found 39 unique  
281 proteins that may have responded differently to severe COVID-19 infection. However, we note  
282 that these results should be viewed as hypothesis generating, and require replication, given the  
283 small sample sizes of non-European ancestries. Of these 39 proteins, we found that 29 showed  
284 a difference in their COVID-19 response between our European and African cohorts, 7 showed  
285 a difference between our European and East Asian cohorts, and 4 of had a difference between  
286 our East Asian and African cohorts – with one protein, STAU2 being found in two of our  
287 comparisons. Given the small sample sizes assessed in this study, and the large amount of data  
288 points being analyzed, replication of this work is essential to ensure the veracity of this data.

289 Our European – African ancestry comparison provided us with the most results most likely due  
290 to the stronger power of both models in comparison to our East Asian model. One of the  
291 proteins that displayed a significant difference between our European and African ancestries is  
292 FAM19A5. FAM19A5 is also known as TAF5 and is in a family of homologous secreted  
293 proteins and acts as an adipokine.<sup>24–26</sup> FAM19A5, unlike the rest of its family proteins, is  
294 expressed in both adipose tissue and the central nervous system which explains its presence as  
295 a circulatory protein. FAM19A5 has been shown to suppress vascular smooth muscle cell  
296 proliferation and mice models with an overexpression of FAM19A5 have reduced scar tissue  
297 formation.<sup>26</sup> Additionally, FAM19A5 has been shown to have an impact on hypothalamic  
298 inflammatory responses, with TNF- $\alpha$  administration increasing levels of hypothalamic FAM19A5  
299 mRNA and when FAM19A5 itself was introduced to a mice model, the mice displayed a  
300 decreased food intake and increased body temperature.<sup>27</sup> Considering that FAM19A5 was  
301 decreased in Europeans (effect size = -0.24,  $p = 2.6 \times 10^{-4}$ ) and increased in our African  
302 population (effect size = 0.31,  $p = 0.015$ ) this could be a factor in the severity of COVID-19  
303 infection.



304 Of the four proteins that showed a difference between our East Asian and African ancestries,  
305 IL1R2 stood out as an interleukin receptor. IL1R2 is a decoy receptor which acts as a  
306 competitive binder of IL1B. The elevated levels of IL1R2 in our East Asian cohort upon COVID-  
307 19 infection should result in a lower impact of IL1B in the same cohort. Interestingly, IL1B was  
308 also measured directly by the SomaScan Assay, and we saw that it also had a differential  
309 response between East Asian and African populations (with elevated levels in our East Asian  
310 population in comparison to our African population), although this difference was not statistically  
311 significant after multiple testing correction (p-value=0.004). IL1B is a known marker of COVID-  
312 19 infection and is a part of the resultant cytokine storm.<sup>28,29</sup> Furthermore, high levels of IL1B  
313 were shown to be associated with a reduced viral load of in the intestine of COVID-19  
314 patients.<sup>29</sup>

315 There were seven proteins that showed a difference in their response to COVID-19 between our  
316 East Asian and European cohort. Of the seven proteins, STAU2 was the only protein found in  
317 both this comparison and the comparison between our East Asians and African cohorts. STAU2  
318 is a double stranded RNA-binding protein that has been previously shown to interact with viral  
319 proteins such as NSP2 in COVID-19 and the Gag protein in HIV-1.<sup>30,31</sup> Furthermore, STAU2 was  
320 shown to enhance the viral infectivity of HIV-1, with *STAU2* KO lines reducing the viral infectivity  
321 of HIV-1 by 55%.<sup>30</sup> These results coupled with the results found here indicate that there might  
322 be a genetic difference in East Asians that should be further investigated to see if STAU2 also  
323 provides a similar protection to COVID-19 and to see if there is a similar interaction between  
324 HIV-1 infection and STAU2 protein levels in East Asian populations.

325 This study shows that while we should consider both the proteomic differences between  
326 different ancestries and their response to COVID-19, many of the circulating proteins respond in  
327 the same way regardless of ancestry. This study further shows that there are novel proteins that  
328 have not been explored due to the European ancestry-centric focus of many genomic,  
329 transcriptomic, and proteomic studies of COVID-19.

330 This study has important limitations. First, there was a relatively small number of individuals  
331 assessed in non-European ancestries. Since the European ancestry cohort has a much larger  
332 sample size, it is not surprising that the European ancestry individuals-based analyses yielded  
333 more findings due to the highest statistical power. More samples from our East Asian and  
334 African cohorts, as well as other ancestral groups that were omitted from this study due to their  
335 small sample sizes, would result in a stronger analysis. Potential confounders, such as smoking  
336 status and BMI, were not included in our analyses due to the large amount of missing data. It  
337 remains possible that the findings may not reflect direct causal effects of severe COVID-19  
338 illness on circulating protein levels. Furthermore, we did not attempt to understand if any of the  
339 observed proteins had causal effects on COVID-19 outcomes. Another limitation of this study is  
340 that we only focused on how ancestries responded differently to a severe COVID-19 infection  
341 and did not investigate overall proteomic differences between ancestries or how these  
342 differences impact COVID-19 infection and a patient's response of infection. Additionally, there  
343 likely exist proteomic differences unable to be captured by our measurement assay. Functional  
344 mutations may have affected the binding affinity of SOMAmers resulting in different levels of  
345 protein observed between ancestries if these protein-coding changes had different frequencies  
346 between the ancestries. Finally, we re-emphasize that replication of our results in diverse  
347 populations is necessary to better understanding trends linked with ancestry and proteomic  
348 responses.

## 349 Conclusion

350 In summary, this work assessed how different genetic ancestries responded to severe COVID-  
351 19 infection by examining 4,979 circulating proteins. We found that while the majority of proteins  
352 responded in the same manner regardless of ancestry, there were 39 unique proteins that had  
353 suggestive differences in their response to COVID-19, which warrant functional follow-up  
354 studies.

355 [List of abbreviations](#)

356 COVID-19: Coronavirus disease 19

357 BQC19: Biobanque Quebecois de la COVID-19

358 LMM: Linear mixed model

359 [Availability of data and materials](#)

360 The BQC19 is an Open Science Biobank. Instructions on how to access data for individuals

361 from the BQC19 is available at <https://www.bqc19.ca/en/access-data-samples>. To see the code

362 used to analyze this data, it is available at

363 [https://github.com/richardslab/BQC\\_Ancestry\\_Proteomics](https://github.com/richardslab/BQC_Ancestry_Proteomics).

364 [Funding](#)

365 The Richards research group is supported by the Canadian Institutes of Health Research (CIHR:

366 365825; 409511, 100558, 169303), the McGill Interdisciplinary Initiative in Infection and

367 Immunity (MI4), the Lady Davis Institute of the Jewish General Hospital, the Jewish General

368 Hospital Foundation, the Canadian Foundation for Innovation, the NIH Foundation, Cancer

369 Research UK, Genome Québec, the Public Health Agency of Canada, McGill University, Cancer

370 Research UK [grant number C18281/A29019] and the Fonds de Recherche Québec Santé

371 (FRQS). JBR is supported by a FRQS Mérite Clinical Research Scholarship. Support from

372 Calcul Québec and Compute Canada is acknowledged. TwinsUK is funded by the Wellcome

373 Trust, Medical Research Council, European Union, the National Institute for Health Research

374 (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based

375 at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. T.N.

376 is supported by a research fellowship of the Japan Society for the Promotion of Science for

377 Young Scientists (22J30004). These funding agencies had no role in the design, implementation  
378 or interpretation of this study.

### 379 [Disclosures](#)

380 JBR's institution has received investigator-initiated grant funding from Eli Lilly, GlaxoSmithKline  
381 and Biogen for projects unrelated to this research. JBR is the CEO of 5 Prime Sciences  
382 ([www.5primesciences.com](http://www.5primesciences.com)), which provides research services for biotech, pharma and venture  
383 capital companies for projects unrelated to this research. T.N. has received a speaking fee from  
384 Boehringer Ingelheim for the talks unrelated to this research. L.C. is a full-time employee of  
385 SomaLogic Inc.

### 386 [Ethics Approval](#)

387 Informed consent for inclusion in this study was obtained from every patient or their legal  
388 representatives. The Biobanque Québécoise de la COVID-19 (BQC19) received ethical  
389 approval from the IRB of the Jewish General Hospital and the Medical/Biomedical Committee of  
390 the Centre Hospitalier de l'Université de Montréal.

### 391 [Author Contributions](#)

392 Conception and Design: TZ, YI, and JBR. Data Analyses: TZ. Data Acquisition: TN, GBL, DM,  
393 DA, LC. Data Verification: TZ, LC. Interpretation of the Data: TZ, YI, and JBR. Funding  
394 Acquisition: JBR. Methodology: TN, TL. Writing: TZ, YI, and JBR. Editing: TZ, TL, GBL, TN, and  
395 JBR. All authors gave final approval of the version to be published. The corresponding author  
396 attests that all listed authors meet authorship criteria and that no others meeting the criteria  
397 have been omitted. All authors read and approved the final manuscript.

### 398 [Acknowledgements](#)

399 We would like to acknowledge the important contributions of research participants to help us  
400 better understand the causes and consequences of the COVID pandemic.

401 **References:**

- 402 1. World Health Organization 2023. WHO Coronavirus (COVID-19) dashboard > About  
403 [Dashboard]. *data.who.int* <https://data.who.int/dashboards/covid19/about>.
- 404 2. Borrell, L. N. *et al.* Race and Genetic Ancestry in Medicine — A Time for Reckoning with  
405 Racism. *N. Engl. J. Med.* **384**, 474–480 (2021).
- 406 3. Lu, C., Ahmed, R., Lamri, A. & Anand, S. S. Use of race, ethnicity, and ancestry data in  
407 health research. *PLOS Glob. Public Health* **2**, e0001060 (2022).
- 408 4. Rebbeck, T. R., Mahal, B., Maxwell, K. N., Garraway, I. P. & Yamoah, K. The distinct impacts  
409 of race and genetic ancestry on health. *Nat. Med.* **28**, 890–893 (2022).
- 410 5. Randolph, H. E. *et al.* Genetic ancestry effects on the response to viral infection are  
411 pervasive but cell type specific. *Science* **374**, 1127–1133 (2021).
- 412 6. Millett, G. A. *et al.* Assessing differential impacts of COVID-19 on black communities. *Ann.*  
413 *Epidemiol.* **47**, 37–44 (2020).
- 414 7. Price-Haywood, E. G., Burton, J., Fort, D. & Seoane, L. Hospitalization and Mortality among  
415 Black Patients and White Patients with Covid-19. *N. Engl. J. Med.* **382**, 2534–2543 (2020).
- 416 8. Williamson, E. J. *et al.* Factors associated with COVID-19-related death using OpenSAFELY.  
417 *Nature* **584**, 430–436 (2020).
- 418 9. COVID-19 Host Genetics Initiative *et al.* Mapping the human genetic architecture of COVID-  
419 19. *Nature* **600**, 472–477 (2021).
- 420 10. The COVID-19 Host Genetics Initiative *et al.* A second update on mapping the human  
421 genetic architecture of COVID-19. *Nature* **621**, E7–E26 (2023).
- 422 11. The GenOMICC Investigators *et al.* Genetic mechanisms of critical illness in COVID-19.  
423 *Nature* **591**, 92–98 (2021).
- 424 12. Shelton, J. F. *et al.* Trans-ancestry analysis reveals genetic and nongenetic associations  
425 with COVID-19 susceptibility and severity. *Nat. Genet.* **53**, 801–808 (2021).

- 426 13. Horowitz, J. E. *et al.* Genome-wide analysis provides genetic evidence that ACE2 influences  
427 COVID-19 risk and yields risk scores associated with severe disease. *Nat. Genet.* **54**, 382–  
428 392 (2022).
- 429 14. Zhou, S. *et al.* A Neanderthal OAS1 isoform protects individuals of European ancestry  
430 against COVID-19 susceptibility and severity. *Nat. Med.* **27**, 659–667 (2021).
- 431 15. Butler-Laporte, G. *et al.* Exome-wide association study to identify rare variants influencing  
432 COVID-19 outcomes: Results from the Host Genetics Initiative. *PLOS Genet.* **18**, e1010367  
433 (2022).
- 434 16. Genetic regulation of OAS1 nonsense-mediated decay underlies association with COVID-19  
435 hospitalization in patients of European and African ancestries | Nature Genetics.  
436 <https://www.nature.com/articles/s41588-022-01113-z>.
- 437 17. Tremblay, K. *et al.* The Biobanque québécoise de la COVID-19 (BQC19)—A cohort to  
438 prospectively study the clinical and biological determinants of COVID-19 clinical trajectories.  
439 *PLOS ONE* **16**, e0245031 (2021).
- 440 18. Nakanishi, T. *et al.* Age-dependent impact of the major common genetic risk factor for  
441 COVID-19 on severity and mortality. *J. Clin. Invest.* **131**, e152386 (2021).
- 442 19. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and  
443 disease. *Nat. Genet.* **53**, 1712–1721 (2021).
- 444 20. Butler-Laporte, G. *et al.* The dynamic changes and sex differences of 147 immune-related  
445 proteins during acute COVID-19 in 580 individuals. *Clin. Proteomics* **19**, 34 (2022).
- 446 21. Von Hippel, P. T. The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC*  
447 *Med. Res. Methodol.* **15**, 35 (2015).
- 448 22. Thomas, P. D. *et al.* PANTHER<sup>2</sup>: Making genome-scale phylogenetics accessible to all.  
449 *Protein Sci.* **31**, 8–22 (2022).
- 450 23. Carbon, S. & Mungall, C. Gene Ontology Data Archive. Zenodo  
451 <https://doi.org/10.5281/ZENODO.10536401> (2024).

- 452 24. Sarver, D. C., Lei, X. & Wong, G. W. FAM19A (TFA): An Emerging Family of Neurokines  
453 with Diverse Functions in the Central and Peripheral Nervous System. *ACS Chem. Neurosci.*  
454 **12**, 945–958 (2021).
- 455 25. Shahapal, A. *et al.* FAM19A5 Expression During Embryogenesis and in the Adult Traumatic  
456 Brain of FAM19A5-LacZ Knock-in Mice. *Front. Neurosci.* **13**, 917 (2019).
- 457 26. Wang, Y. *et al.* Novel Adipokine, FAM19A5, Inhibits Neointima Formation After Injury  
458 Through Sphingosine-1-Phosphate Receptor 2. *Circulation* **138**, 48–63 (2018).
- 459 27. Kang, D. *et al.* Brain-specific chemokine FAM19A5 induces hypothalamic inflammation.  
460 *Biochem. Biophys. Res. Commun.* **523**, 829–834 (2020).
- 461 28. Declercq, J., De Leeuw, E. & Lambrecht, B. N. Inflammasomes and IL-1 family cytokines in  
462 SARS-CoV-2 infection: from prognostic marker to therapeutic agent. *Cytokine* **157**, 155934  
463 (2022).
- 464 29. Lücke, J. *et al.* Intestinal IL-1 $\beta$  Plays a Role in Protecting against SARS-CoV-2 Infection. *J.*  
465 *Immunol.* **211**, 1052–1061 (2023).
- 466 30. Balakrishnan, K. *et al.* Encapsidation of Staufin-2 Enhances Infectivity of HIV-1. *Viruses* **13**,  
467 2459 (2021).
- 468 31. Zheng, Y.-X. *et al.* Nsp2 has the potential to be a drug target revealed by global  
469 identification of SARS-CoV-2 Nsp2-interacting proteins. *Acta Biochim. Biophys. Sin.* **53**,  
470 1134–1141 (2021).
- 471