Differential circulating proteomic responses associated with ancestry during severe COVID-19 infection

Authors: Thomas M Zheng[1], Yann Ilboudo[2], Tianyuan Lu[6], Guillaume Butler-Laporte[2], Tomoko Nakanishi[2,3,7-9], David Morrison [2], Darin Adra [2], J. Brent Richards[1-5]

Affiliations:

1. Quantitative Life Sciences, McGill University, Montréal, Quebec, Canada
2. Lady Davis Institute, Jewish General Hospital, McGill University, Montréal, Quebec, Canada
3. Departments of Medicine, Human Genetics, Epidemiology and Biostatistics, McGill University, Montréal, Quebec, Canada
4. Department of Twin Research, King's College London, London, United Kingdom
5. Five Prime Sciences Inc, Montréal, Québec, Canada
6. Department of Statistics, University of Toronto, Toronto, Ontario, Canada
7. Kyoto-McGill International Collaborative Program in Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto, Japan
8. Department of Genome Informatics, Graduate School of Medicine, the University of Tokyo, Tokyo, Japan
9. Research Fellow, Japan Society for the Promotion of Science, Tokyo, Japan

Correspondence:

Corresponding author: Brent Richards, brent.richards@mcgill.ca

Abstract

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

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

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

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

Keywords

Proteomics, COVID-19, Ancestry, High-throughout Proteomics, BQC19, COVID-19 Response, Ancestry Differences, Linear Mixed Models

Background

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

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

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

82 infection. High-throughput oligonucleotide-aptamer protein measurement technology allows for
83 the measurement of thousands of circulating proteins simultaneously from a single sample. The
84 increased throughput of this technology allows for the measurement of many such proteins
85 simultaneously. In spite of the benefits of proteomic analysis, there are only a few studies that
86 looked at the proteomic differences between ancestries, and these ones focused primarily on the
87 European ancestry.[14,16] Previously, we found that an isoform of *OAS1* in people of European
88 ancestry displayed a protective effect for COVID-19.[14]

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

96 Methods

97 Overview of the study design

98 We used the SomaScan V4 assay to measure 4,979 circulating proteins in case and control
99 samples from the BQC19, a biobank in Québec, Canada.[17] We used linear mixed models
100 (LMMs) to assess differences in protein levels between severe COVID-19 patients and controls
101 who are hospitalized patients that either did not have COVID or presented with mild COVID
102 symptoms, stratified by ancestry. The differences in the association between protein levels and
103 COVID-19 severity was then compared between ancestries.

104 Population

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

109 ## COVID-19 Case/Control Definitions

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

117 ## Ancestry Determination

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

125 ## Protein measurements

126 We used the SomaScan (v4) platform to measure 5,284 circulating proteins from each sample.
127 Of these 5,284 proteins, 305 non-human proteins were then removed from analysis using the
128 same classification used previously in the literature.[19] To adequately reflect the effects of

129 COVID-19 infection, we limited our experiment to those biological samples procured within 30
130 days after symptom onset, with our linear mixed models accounting for the changes in protein
131 levels over time. Blood plasma samples were collected in acid-citrate-dextrose tubes and frozen
132 at -80 °C. Protein levels were measured using resonance fluorescence units, and further
133 normalized and calibrated by SomaLogic to remove any systematic bias. For the following
134 statistical analysis, each protein sample across all patients was then normalized to have a mean
135 of 0 and standard deviation of 1. This normalization was done prior to any ancestry stratification.
136 The sample correction and proteomic normalization has been previously described in more
137 detail by Butler-Laporte et al.[20]

138 Statistical Analysis

139 To leverage the BQC19's longitudinal samples, we used a linear mixed model (LMM) to group
140 samples from the same patient using their BQC identifier. R packages, lme4 (1.1.35.1) and
141 lmerTest (3.1.3) were used with R version 4.1.2 to generate these LMMs. This model further
142 accounts for the patient's age, sex and when the sample was taken relative to when the patient
143 first had symptoms. These patients were further stratified by ancestry to examine how COVID-
144 19 impacts protein levels by ancestry. The formula and sample code can be found in
145 Supplementary file 1 and the GitHub
146 (https://github.com/richardslab/BQC_Ancestry_Proteomics) respectively. A full table of all 4,979
147 proteins, their ancestry specific effect sizes, and their standard errors can be found in Additional
148 File 1.

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

154 To assess how COVID-19 impacts protein levels differently between ancestries we used both
155 the beta and standard error from our LMM in each ancestry and derived a Z-score. A full formula
156 can be found in Supplementary File 1.

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

163 Results

164 Sample Characteristics

165 Table 1 shows the total number of samples used in our study were separated by genetic
166 ancestry. The samples were split into the six major ancestry groups as used by the HGI and the
167 "other" category. These groups were: "European", "African", "East Asian", "South Asian",
168 "Admixed American", "Middle Eastern", and "Other". For the remainder of this study, we focused
169 only on our three largest ancestral groups: European, African, and East Asian. For patient
170 characteristics, see Supplementary Table 1.

171 Table 1: Sample characteristics in the BQC19.

	European	**African**	**East Asian**	**Middle Eastern**	**Admixed American**	**South Asian**	**Admixed/ Other**
Overall *n* (% 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)

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

Different ancestries had different responses to COVID-19 infection

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

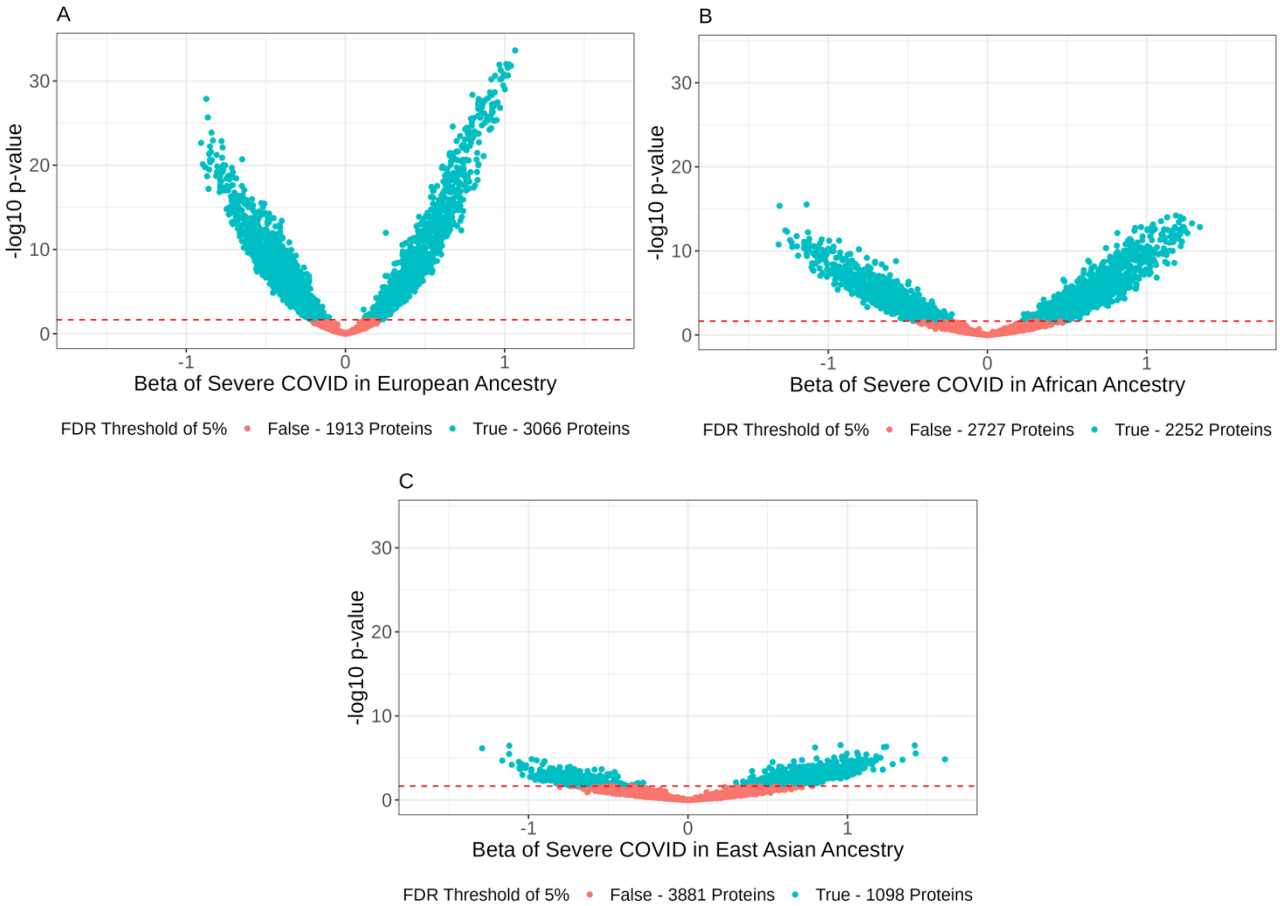

Figure 1: Volcano plots of the three tested ancestries of the BQC19: European – n = 797 (A), African – n = 196 (B), and East Asian – n = 79 (C).

The x-axis represents the effect size of the COVID variable in our linear mixed model. Our linear mixed model included age, sex, and days since symptom onset as co-variables, and used patient ID as a random factor to account for intra-patient correlation. All volcano plots are scaled to the same axis, each dot represents one tested protein. The colour represents if the protein is considered to be significantly different between our case and control definition. The red dashed line is p = 0.021 to provide an FDR of 5%.

When comparing our European ancestries and our African ancestries, we saw that most proteins share the same trend upon severe COVID-19 infection. In Figure 2A, we compared all 4,979 of our tested protein's European effect sizes to their African effect sizes. We found that they had a Pearson correlation of 0.868, 95% CI [0.861, 0.875]. We saw a similar, if weaker trend when comparing our European and East Asian ancestries with a respective Pearson correlation of 0.645, 95% CI [0.628, 0.661] (Figure 2B). Finally, we found that the trend continued between our East Asian and African ancestries in Figure 2C, with a Pearson correlation of 0.709, 95% CI [0.695, 0.722].

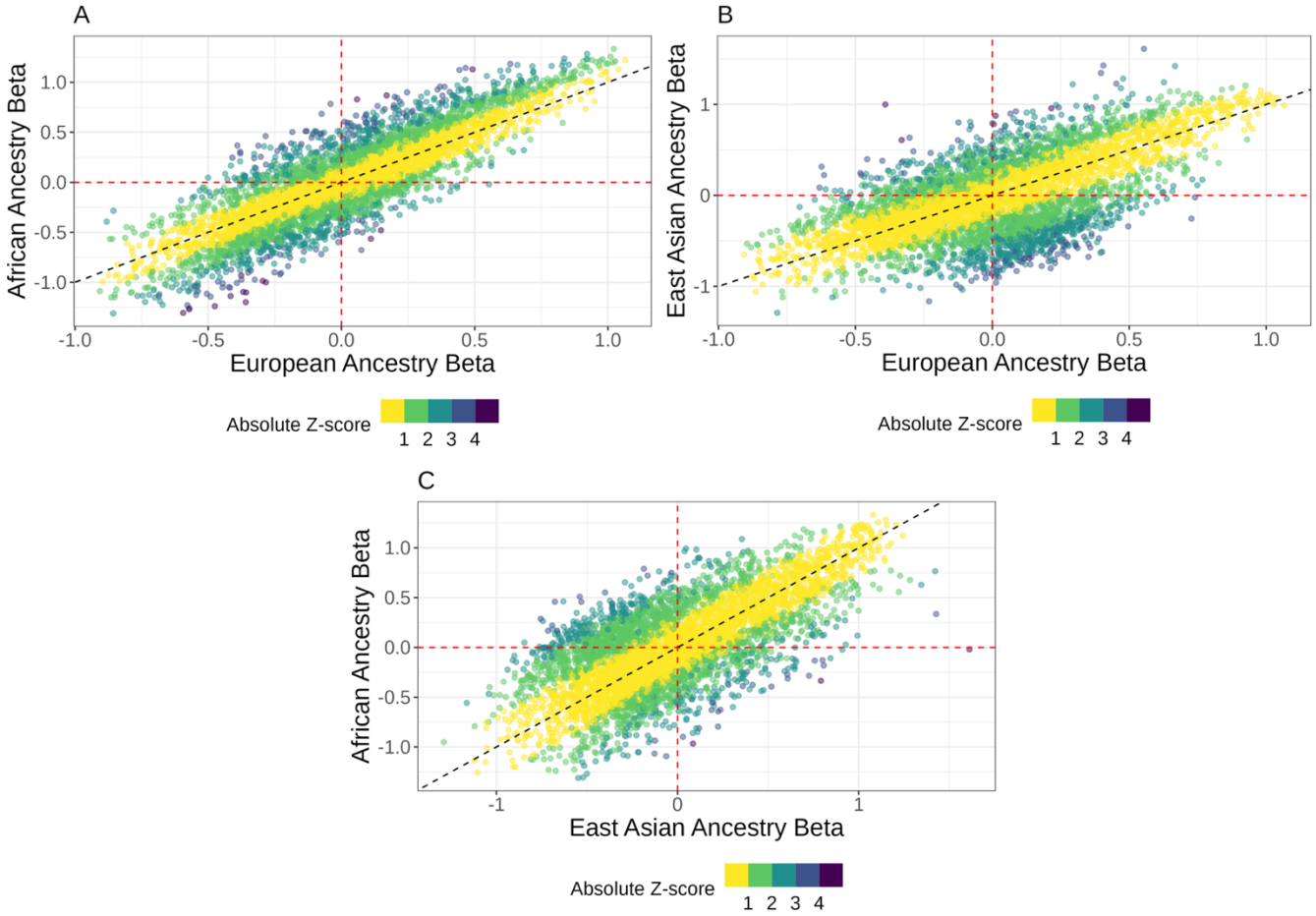

Figure 2: Pairwise Beta-Beta plots which displays all 4979 proteins and their associated COVID beta.

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

218 Differences in how severe COVID-19 influences protein levels

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

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

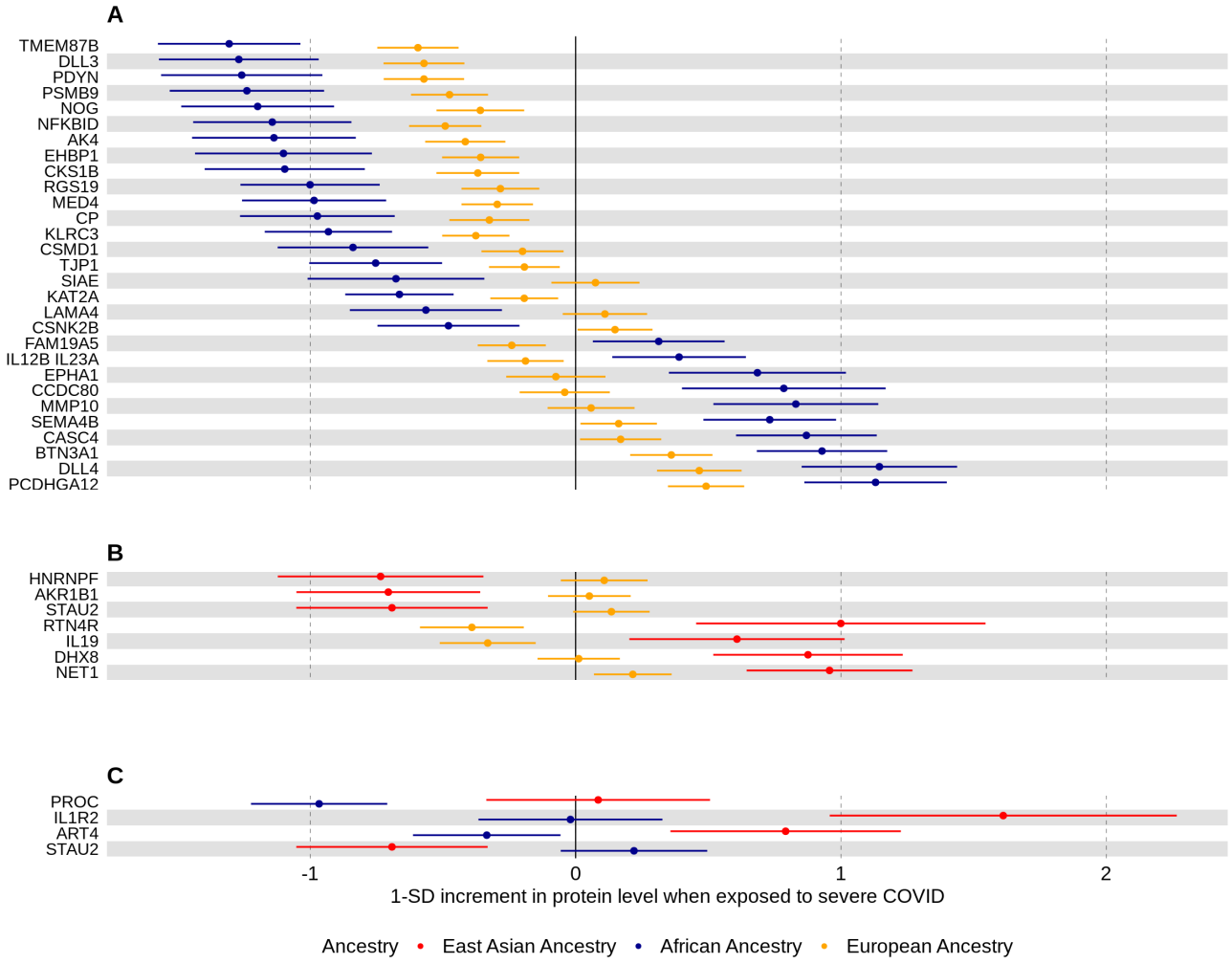

Figure 3: Forest plots of all proteins that were significantly different between two of the three ancestries tested.

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

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

258 Validation of *OAS1* results

259 As previously mentioned, *OAS1* was identified by Zhou et al.[14] to be a circulating protein that
260 caused less severe COVID-19. Their conclusions hypothesized that COVID-19 infection would
261 cause OAS1 levels to increase in European individuals. When we tested OAS1 protein levels,
262 we found that they were increased in our severe COVID-19 patients in both European and

African ancestries (unadjusted p-value of 1.0×10^{-4} and 2.7×10^{-5} respectively) (Figure 4). However, since the test in East Asian ancestry individuals was not well-powered due to the small sample size, there was no detectable difference in OAS1 levels between severe COVID-19 patients and controls (p-value of 0.49).

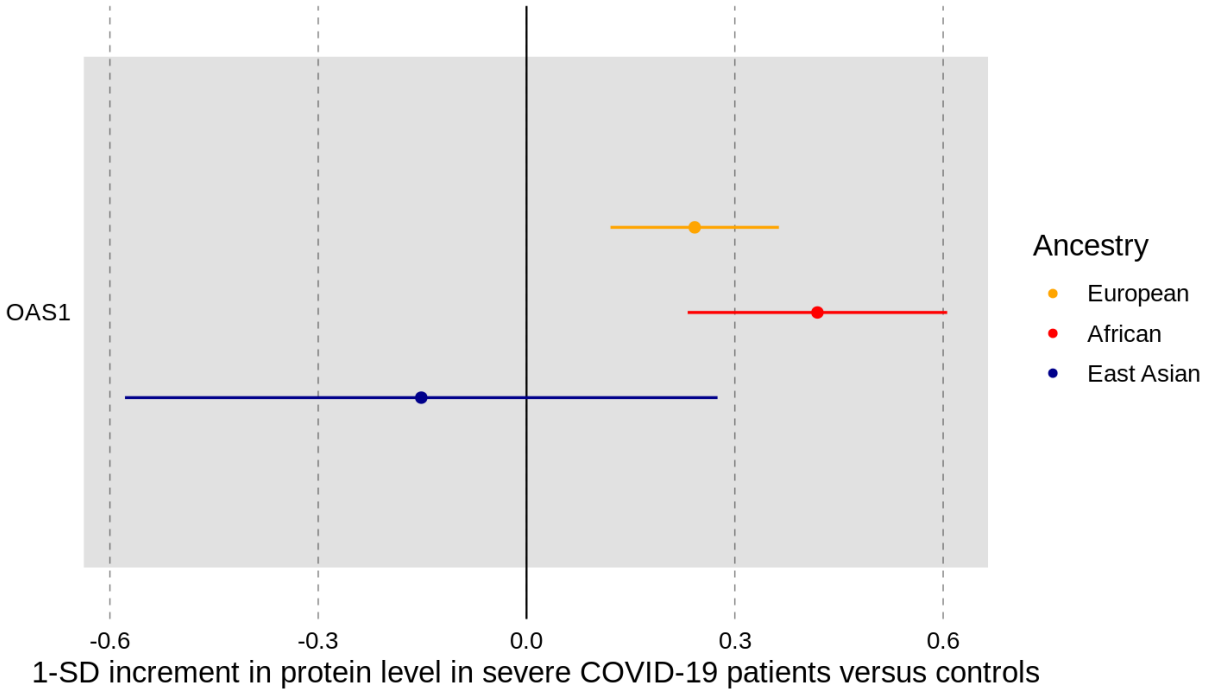

Figure 4: Forest plot of OAS1 when comparing case and control status across all three tested ancestries.

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

Discussion

We leveraged the BQC19 to measure 4,979 circulating proteins across 1,272 samples from 909 patients. We then stratified these measurements by genetic ancestry and found 39 unique proteins had may have responded differently to severe COVID-19 infection. However, we note that these results should be viewed as hypothesis generating, and require replication, given the

277 small sample sizes of non-European ancestries. Of these 39 proteins, we found that 29 showed
278 a difference in their COVID-19 response between our European and African cohorts, 7 showed
279 a difference between our European and East Asian cohorts, and 4 of had a difference between
280 our East Asian and African cohorts – with one protein, STAU2 being found in two of our
281 comparisons. Given the small sample sizes assessed in this study, and the large amount of data
282 points being analyzed, replication of this work is essential to ensure the veracity of this data.

283 Our European – African ancestry comparison provided us with the most results most likely due
284 to the stronger power of both models in comparison to our East Asian model. One of the
285 proteins that displayed a significant different between our European and African ancestries is
286 FAM19A5. FAM19A5 is also known as TAFA5 and is in a family of homologous secreted
287 proteins and acts as an adipokine.[24–26] FAM19A5, unlike the rest of its family proteins, is
288 expressed in both adipose tissue and the central nervous system which explains its presence as
289 a circulatory protein. FAM19A5 has been shown to supress vascular smooth muscle cell
290 proliferation and mice models with an overexpression of FAM19A5 have reduced scar tissue
291 formation.[26] Additionally, FAM19A5 has been shown to have an impact on hypothalamic
292 inflammatory responses, with TNF-α administration increasing levels of hypothalamic FAM19A5
293 mRNA and when FAM19A5 itself was introduced to a mice model, the mice displayed a
294 decreased food intake and increased body temperature.[27] Considering that FAM19A5 was
295 decreased in Europeans (effect size = -0.24, $p = 2.6 \times 10^{-4}$) and increased in our African
296 population (effect size = 0.31, p = 0.015) this could be a factor in the severity of COVID-19
297 infection.

298 Of the four proteins that showed a difference between our East Asian and African ancestries,
299 IL1R2 stood out as an interleukin receptor. IL1R2 is a decoy receptor which acts as a
300 competitive binder of IL1B. The elevated levels of IL1R2 in our East Asian cohort upon COVID-
301 19 infection should result in a lower impact of IL1B in the same cohort. Interestingly, IL1B was

302 also measured directly by the SomaScan Assay, and we saw that it also had a differential
303 response between East Asian and African populations (with elevated levels in our East Asian
304 population in comparison to our African population), although this difference was not statistically
305 significant after multiple testing correction (p-value=0.004). IL1B is a known marker of COVID-
306 19 infection and is a part of the resultant cytokine storm.[28,29] Furthermore, high levels of IL1B
307 were shown to be associated with a reduced viral load of in the intestine of COVID-19
308 patients.[29]
309 There were seven proteins that showed a difference in their response to COVID-19 between our
310 East Asian and European cohort. Of the seven proteins, STAU2 was the only protein found in
311 both this comparison and the comparison between our East Asians and African cohorts. STAU2
312 is a double stranded RNA-binding protein that has been previously shown to interact with viral
313 proteins such as NSP2 in COVID-19 and the Gag protein in HIV-1.[30,31] Furthermore, STAU2 was
314 shown to enhance the viral infectivity of HIV-1, with *STAU2* KO lines reducing the viral infectivity
315 of HIV-1 by 55%.[30] These results coupled with the results found here indicate that there might
316 be a genetic difference in East Asians that should be further investigated to see if STAU2 also
317 provides a similar protection to COVID-19 and to see if there is a similar interaction between
318 HIV-1 infection and STAU2 protein levels in East Asian populations.
319 This study shows that while we should consider both the proteomic differences between
320 different ancestries and their response to COVID-19, many of the circulating proteins respond in
321 the same way regardless of ancestry. This study further shows that there are novel proteins that
322 have not been explored due to the European ancestry-centric focus of many genomic,
323 transcriptomic, and proteomic studies of COVID-19.

324 This study has important limitations. First, there was a relatively small number of individuals
325 assessed in non-European ancestries. Since the European ancestry cohort has a much larger
326 sample size, it is not surprising that the European ancestry individuals-based analyses yielded

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

343 Conclusion

344 In summary, this work assessed how different genetic ancestries responded to severe COVID-
345 19 infection by examining 4,979 circulating proteins. We found that while the majority of proteins
346 responded in the same manner regardless of ancestry, there were 39 unique proteins that had
347 suggestive differences in their response to COVID-19, which warrant functional follow-up
348 studies.

349 List of abbreviations

350 COVID-19: Coronavirus disease 19

BQC19: Biobanque Quebecois de la COVID-19

LMM: Linear mixed model

Availability of data and materials

The BQC19 is an Open Science Biobank. Instructions on how to access data for individuals from the BQC19 is available at https://www.bqc19.ca/en/access-data-samples. To see the code used to analyze this data, it is available at https://github.com/richardslab/BQC_Ancestry_Proteomics.

Funding

The Richards research group is supported by the Canadian Institutes of Health Research (CIHR: 365825; 409511, 100558, 169303), the McGill Interdisciplinary Initiative in Infection and Immunity (MI4), the Lady Davis Institute of the Jewish General Hospital, the Jewish General Hospital Foundation, the Canadian Foundation for Innovation, the NIH Foundation, Cancer Research UK, Genome Québec, the Public Health Agency of Canada, McGill University, Cancer Research UK [grant umber C18281/A29019] and the Fonds de Recherche Québec Santé (FRQS). JBR is supported by a FRQS Mérite Clinical Research Scholarship. Support from Calcul Québec and Compute Canada is acknowledged. TwinsUK is funded by the Welcome Trust, Medical Research Council, European Union, the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust in partnership with King's College London. T.N. is supported by a research fellowship of the Japan Society for the Promotion of Science for Young Scientists (22J30004). These funding agencies had no role in the design, implementation or interpretation of this study.

Disclosures

JBR's institution has received investigator-initiated grant funding from Eli Lilly, GlaxoSmithKline

375 and Biogen for projects unrelated to this research. JBR is the CEO of 5 Prime Sciences
376 (www.5primesciences.com), which provides research services for biotech, pharma and venture
377 capital companies for projects unrelated to this research. T.N. has received a speaking fee from
378 Boehringer Ingelheim for the talks unrelated to this research.

379 ## Ethics Approval

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

384 ## Author Contributions

385 Conception and Design: TZ, YI, and JBR. Data Analyses: TZ. Data Acquisition: TN, GBL, DM,
386 DA. Interpretation of the Data: TZ, YI, and JBR. Funding Acquisition: JBR. Methodology: TN, TL.
387 Writing: TZ, YI, and JBR. Editing: TZ, TL, GBL, TN, and JBR. All authors gave final approval of
388 the version to be published. The corresponding author attests that all listed authors meet
389 authorship criteria and that no others meeting the criteria have been omitted. All authors read
390 and approved the final manuscript.

391 ## Acknowledgements

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

References:

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

419 13. Horowitz, J. E. *et al.* Genome-wide analysis provides genetic evidence that ACE2 influences
420 COVID-19 risk and yields risk scores associated with severe disease. *Nat. Genet.* **54**, 382–
421 392 (2022).
422 14. Zhou, S. *et al.* A Neanderthal OAS1 isoform protects individuals of European ancestry
423 against COVID-19 susceptibility and severity. *Nat. Med.* **27**, 659–667 (2021).
424 15. Butler-Laporte, G. *et al.* Exome-wide association study to identify rare variants influencing
425 COVID-19 outcomes: Results from the Host Genetics Initiative. *PLOS Genet.* **18**, e1010367
426 (2022).
427 16. Genetic regulation of OAS1 nonsense-mediated decay underlies association with COVID-19
428 hospitalization in patients of European and African ancestries | Nature Genetics.
429 https://www.nature.com/articles/s41588-022-01113-z.
430 17. Tremblay, K. *et al.* The Biobanque québécoise de la COVID-19 (BQC19)—A cohort to
431 prospectively study the clinical and biological determinants of COVID-19 clinical trajectories.
432 *PLOS ONE* **16**, e0245031 (2021).
433 18. Nakanishi, T. *et al.* Age-dependent impact of the major common genetic risk factor for
434 COVID-19 on severity and mortality. *J. Clin. Invest.* **131**, e152386 (2021).
435 19. Ferkingstad, E. *et al.* Large-scale integration of the plasma proteome with genetics and
436 disease. *Nat. Genet.* **53**, 1712–1721 (2021).
437 20. Butler-Laporte, G. *et al.* The dynamic changes and sex differences of 147 immune-related
438 proteins during acute COVID-19 in 580 individuals. *Clin. Proteomics* **19**, 34 (2022).
439 21. Von Hippel, P. T. The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC*
440 *Med. Res. Methodol.* **15**, 35 (2015).
441 22. Thomas, P. D. *et al.* PANTHER : Making genome-scale phylogenetics accessible to all.
442 *Protein Sci.* **31**, 8–22 (2022).
443 23. Carbon, S. & Mungall, C. Gene Ontology Data Archive. Zenodo
444 https://doi.org/10.5281/ZENODO.10536401 (2024).

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