Molecular stratification of hospitalized COVID19 patients points to FGFR and SHC4-signaling in ARDS

William Ma[1,*], Antoine Soulé[1,*], Katelyn Yixiu Liu[2], Catherine Allard[3], Salman Qureshi[2], Karine Tremblay[4,†], Simon Rousseau[2,†], and Amin Emad[1,5,†]

[*] These authors contributed equally.

[1] Department of Electrical and Computer Engineering, McGill University, Montréal, QC, Canada

[2] The Meakins-Christie Laboratories at the Research Institute of the McGill University Heath Centre Research Institute, & Department of Medicine, Faculty of Medicine, McGill University, Montréal, QC, Canada

[3] Statistical department, Centre de recherche du Centre hospitalier universitaire de Sherbrooke (CRCHUS), Sherbrooke, Canada.

[4] Pharmacology-physiology Department, Faculty of Medicine and Health Sciences, Université de Sherbrooke, Saguenay, QC, Canada; Centre intégré universitaire de santé et de services sociaux du Saguenay-Lac-Saint-Jean, Saguenay, QC, Canada; CRCHUS, Sherbrooke, Canada.

[5] Mila, Quebec AI Institute, Montréal, QC, Canada

† Corresponding Authors:
Amin Emad,
755 McConnell Engineering Building, 3480 University Street, Montreal H3A 0E9, Canada
Email: amin.emad@mcgill.ca

Simon Rousseau,
RI-MUHC, E M3.2244, 1001 Décarie, Montréal H4A 3J1, Canada,
Email: simon.rousseau@mcgill.ca

Karine Tremblay,
Pavillon des Augustines, local AUG-5-01A, 225 St-Vallier street, Chicoutimi G7H 7P2, Canada
Email: karine.tremblay@usherbrooke.ca

Abstract

A minority of people infected with SARS-CoV-2 will develop severe COVID-19 disease requiring invasive respiratory support associated with high mortality. To understand the molecular mechanisms underlying severe pathology, we conducted an unsupervised stratification of the circulating proteome that identified six endophenotypes (EPs) among 731 SARS-CoV-2 PCR-positive hospitalized participants in the Biobanque Québécoise de la COVID-19, with varying degrees of disease severity and times to intensive care unit (ICU) admission. One endophenotype, EP6, was associated with a greater proportion of ICU admission, mechanical ventilation, acute respiratory distress syndrome (ARDS) and death. Clinical features of EP6 included increased levels of C-reactive protein, D-dimers, elevated neutrophils, and depleted lymphocytes. Proteomic, metabolomic, and genomic characterization supported a role for neutrophil-associated procoagulant activity in severe COVID-19 ARDS that is inversely correlated with sphinghosine-1 phosphate plasma levels. Fibroblast Growth Factor Receptor (FGFR) and SH2-containing transforming protein 4 (SHC4) signaling were identified as molecular features associated with severe COVID-19. Mechanical ventilation in EP6 was associated with alterations in lipoprotein metabolism. Importantly, a prognostic model solely based on clinical laboratory measurements was developed and validated on 631 patients that generalizes the EPs to new patients and creates new opportunities for automated identification of high-risk groups in the clinic.

Introduction

A minority of people infected with SARS-CoV-2 will develop a severe form of coronavirus disease 2019 (COVID-19) requiring hospitalization and respiratory support. Acute respiratory distress syndrome (ARDS), a condition characterized by the presence of lung inflammation, edema, and impaired gas exchange, is observed in severe COVID-19 and associated with poor survival. Mechanical ventilation is an important critical care management strategy for severe COVID-19 (1). However, during the first waves of the pandemic, almost half of patients undergoing invasive mechanical ventilation died based on case fatality rate reports (2). Defining the molecular mechanisms related to specific severe outcomes, such as ARDS or acute respiratory failure that requires invasive mechanical ventilation, is important to identify treatable traits and improve the survival of critically ill patients.

Successfully reaching this precision medicine goal requires a more granular definition of the underlying pathophysiology. A symptom-based method to discover molecular mechanisms of the disease is inherently confounded by the fact that the same higher-level condition, such as severe COVID-19 disease, can be produced by several different molecular mechanisms, a phenomenon termed the «many-one» limitation (3). Recent advances in computing strategies, such as machine learning, has enabled the development of methods that help overcome this limitation by starting from molecular profiles instead of symptoms to define endophenotypes, *i.e.* subgroups of individuals who are inapparent to traditional classification methods but share a common set of molecular factors that can lead to identification of treatable traits (4-6). Current investigations of endophenotypes in COVID-19 have mainly relied on supervised approaches using fixed outcomes

42 (such as disease severity) and integrating clinical variables at the onset (7). We hypothesize that
43 using an unsupervised approach exploiting a rich molecular dataset can provide novel
44 mechanistic insights into the pathobiology of severe COVID-19 that can help physicians improve
45 diagnosis, prognosis, and clinical management.
46
47 This study identified six endophenotypes linked to diverse clinical trajectories of COVID-19 using
48 the extensive molecular phenotyping of a cohort of 731 SARS-CoV-2 positive hospitalized patients
49 from the *Biobanque Québécoise de la COVID-19* (BQC19, www.quebeccovidbiobank.ca) (8), a
50 prospective observational cohort of SARS-CoV-2-positive and -negative participants recruited in
51 the province of Québec, Canada, to improve our understanding of COVID-19 pathobiology and
52 our capacity to alter disease outcomes. The molecular information underpinning these
53 endophenotypes was used to advance our knowledge of the pathobiology of COVID-19-
54 associated ARDS.
55

56 **Results**

57 **Unsupervised clustering of SARS-CoV-2-positive hospitalized BQC19 participants reveal**
58 **endophenotypes associated with varying disease severity and requirements for ventilatory**
59 **support**

60 In this study, we aimed to identify endophenotypes of COVID-19 based on the circulating
61 proteome of patients in our cohort of SARS-CoV-2 positive hospitalized participants in the BQC19
62 (Table 1), using an unsupervised approach. Figure S1 shows the distribution of the patient
63 hospital admission dates and the corresponding waves of COVID-19 infection as defined by
64 National Institute of Public Health of Quebec (https://www.inspq.qc.ca/covid-19). For this
65 purpose, we performed consensus agglomerative clustering of the subset of patients (n = 731,
66 Table S1) for whom data corresponding to circulating proteome measured by a multiplex
67 SOMAmer affinity array (Somalogic, ~5,000 aptamers) (9) was available in the BQC19. First, the
68 optimal number of clusters (k = 6) was identified using two criteria, Akaike's Information Criteria
69 (AIC) and Bayesian Information Criteria (BIC) (Figure 1A); then, consensus agglomerative
70 clustering (Euclidean distance and Ward linkage) (10, 11) using 1000 bootstrap subsampling of
71 the patients was performed to obtain six robust clusters (see Methods for details) (Figure 1,
72 Figures S2 and S3). The distribution of Rand-Index, showing the concordance between each one
73 of the 1000 subsampled clusterings and the final consensus clustering is provided in Figure S2B
74 (mean Rand-Index = 0.823), reflecting a high degree of consistency and robustness.

75

76 The clinical and pathological characteristics of patients in each endophenotype is provided in
77 Table S1. To characterize the identified endophenotypes (EPs) with respect to disease severity,

78 we performed two-sided Fisher's exact test to assess their enrichment (or depletion) in "severe"
79 or "deceased" outcome. EP6 was significantly enriched in the severe/deceased outcomes
80 (Benjamini-Hochberg false discovery rate (FDR) = 1.72E-18) with either of these outcomes
81 observed in 66.1% of EP6 patients. Meanwhile, EP1 was significantly depleted in severe/deceased
82 outcomes (FDR = 8.23E-11) (Figure 2A, Table S1) with either of these outcomes observed in only
83 11.6% of EP1 patients. In addition, EP6 was enriched in participants 1) receiving oxygen therapy
84 (FDR = 6.24E-12), 2) receiving ventilatory support (FDR = 6.63E-12), and 3) being admitted to
85 intensive care unit (ICU) (FDR= 1.26E-22) (Figure 2A, Table S1). Kaplan Meier analysis (12) also
86 confirmed that the identified EPs have a distinct temporal pattern of admission to ICU
87 (multivariate logrank test P = 5.06E-36), with EP1 and EP6 having the highest and lowest
88 probability, respectively, of not being admitted to ICU or dying prior to that in a 40-day span since
89 their admission to the hospital (Figure 2C). A similar pattern was also observed when patients
90 that died before admission to ICU were excluded (Figure S4, multivariate logrank test P = 6.55E-
91 36). A two-sided Mann-Whitney U (MWU) test showed that patients in EP5 were generally older
92 than other EPs (FDR = 6.94E-5), while EP3 included younger patients (FDR = 1.82E-4). Notably,
93 EP6 (which had the most severe patients) did not show enrichment in older patients or individuals
94 with high BMI (two-sided MWU FDR > 0.05) (Figures 2D-F, Table S1).
95
96 These analyses revealed that the unsupervised approach using the circulating proteome of the
97 patients was able to identify endophenotypes with distinct disease characteristics and outcomes.
98 We identified EP6 as a group of participants with an increase in key measures of COVID-19 disease
99 severity, including admission to ICU and the need for ventilatory support.

100

101 **A computational model based on blood markers predicts EPs in a separate validation cohort**

102 To further characterize each EP, we assessed the clinical laboratory results obtained from blood
103 draws and compared them between the groups. We focused on 21 markers that were measured
104 in at least 50% of the patients used for consensus clustering (Figure 3A and Table S2) and used
105 the summary value reported in the BQC19 database corresponding to the most extreme
106 measurement among multiple blood draws (Table S2). Figure 3A shows the elevation and
107 depletion of these markers in the identified EPs. EP6 is characterized by abnormal values in
108 markers of inflammation (lymphopenia, total white blood cell count, neutrophilia, C-reactive
109 protein (CRP)), liver damage (alanine aminotransferase (ALT)), coagulopathy (D-dimers, low
110 hemoglobin, International Normalized Ratio (INR), and hyperglycemia (glucose).

111

112 Since each EP and particularly EP6, which we identified as the EP with worst outcome, showed a
113 clear and distinct clinical laboratory result signature compared to other EPs, we sought to develop
114 a computational prognostic model of disease severity based on these signatures. Due to the
115 presence of missing values for these markers in our cohort, we developed a nearest-centroid
116 classifier that is capable of dealing with missing values and can predict EPs based on blood
117 markers (see Methods for details). To test the prognostic ability of this model on an independent
118 yet similar dataset, we analyzed 631 SARS-CoV-2 positive hospitalized BQC19 participants that
119 did not have circulating proteome data and had not been used to identify the endophenotypes.
120 The clinical and pathological characteristics of patients in each predicted endophenotype (PEP)
121 are provided in Figure 3B-5E and Table S3.

122

123 Our prognostic model identified 116 of these 631 patients as belonging to EP6. Fisher's exact test
124 showed a significant enrichment of the predicted EP6 (PEP6) in severity/deceased (FDR = 1.77E-
125 22), while PEP1 and PEP2 were significantly depleted in these outcomes (FDR = 1.61E-4 and FDR
126 = 1.30E-8, respectively), as shown in Figure 3B and Table S3. Similar to EP6, PEP6 was also
127 significantly enriched in participants 1) receiving oxygen therapy (FDR = 2.38E-8), 2) receiving
128 ventilatory support (FDR = 4.40E-8), and 3) being admitted to ICU (FDR = 1.23E-24) (Table S3).
129 Kaplan Meier analysis also confirmed that these PEPs have a distinct temporal pattern of
130 admission to ICU (multivariate logrank test P = 1.05E-34), with PEP6 having the lowest chance of
131 not being admitted to ICU (or dying prior to that) in a 40-day span since their admission to the
132 hospital (Figure 3E). These results suggest that our prognostic model based on 21 blood markers
133 can be used to generalize the definition of endophenotypes to patients for whom the proteome
134 data is unavailable.

135

136 In addition to its potential applicability to clinical care, this validation further supports the
137 relevance of the EP molecular signatures to disease pathogenesis. Therefore, we sought to
138 improve our understanding of severe COVID-19 using multi-modal molecular and clinical data
139 associated with each EP, and particularly EP6.

140

141 **EP6 is enriched among BQC19 participants with acute respiratory distress syndrome**

142 In accordance with increased COVID-19 disease severity, EP6 was also enriched in several medical
143 complications (two-sided Fisher's exact test) including ARDS (FDR = 1.43E-10), acute kidney injury

144 (FDR = 6.37E-7), secondary bacterial pneumonia (FDR=2.13E-6), liver dysfunction (FDR = 9.32E-
145 3), and hyperglycemia (FDR = 1.56E-2) (Figure 4, Table S4). Notably, the frequency of ARDS was
146 just below 8% in EP1 compared to greater than 44% in EP6, making this complication a key feature
147 of this endophenotype (Figure 4, Table S4). Therefore, the molecular profiles of EP6 members
148 could inform the pathobiology of ARDS.
149

150 **Pathway enrichment analysis identifies FGFR-signaling in severe COVID-19 acute respiratory**
151 **distress syndrome**

152 To gain insight into the molecular mechanisms underlying the severity in EP6, we investigated
153 the aptamers enriched within this endophenotype (Table S5). Consistent with the elevated
154 neutrophil numbers in EP6, the most enriched aptamer was Cathepsin G, a serine protease
155 localized in neutrophil azurophil granules. Additionally, we performed pathway enrichment
156 analysis using the Knowledge Engine for Genomics (KnowEnG) (5). Since aptamers were used to
157 identify the EPs, it is expected that many of them would have significant associations with the
158 EPs. As a result, we selected a very strict threshold of FDR < 1E-20 (two-sided MWU test) to
159 identify top aptamers associated with EP6 (Table S5). We then used the gene set characterization
160 pipeline of KnowEnG with Reactome pathway collection (13). This analysis showed that EP6 is
161 characterized by pathways associated with Interleukins and Cytokine Signaling in Immune
162 system. In addition and in contrast to other EPs, Fibroblast Growth Factor Receptor (FGFR)
163 signaling was identified in EP6, suggesting this pathway as a potential driver of severe pathology
164 (Tables 2 and S5).
165

166 **SHC4 genotype and protein expression levels are associated with higher odds of belonging to**
167 **EP6**

168 To obtain independent support for the involvement of specific proteins identified using
169 aptamers, we leveraged Genome Wide Association Study (GWAS) genotyping data corresponding
170 to these patients. We identified 25 single nucleotide variations (SNVs) distributed in 13 annotated
171 genes that were below a p-value threshold of 1E-4 and distinguished EP6 from the rest (Table 3).
172 We then investigated each of the SNVs to which we could assign a gene and an aptamer, to assess
173 whether their protein product in circulation was differentially regulated by the genotype (Table
174 4). We discovered two genes, *SHC4* (encoding SHC adaptor protein 4) and *CACNA2D3* (encoding
175 calcium voltage-gated channel auxiliary subunit alpha2 delta3) for which there was a significant
176 association between genotype and protein expression levels (p-values < 0.05). While *CACNA2D3*
177 may have mild impact on EP6 membership (odds ratio = 0.61, Table 4), SHC4 was one of the top-
178 enriched aptamers (position 32 out of 4,985) for inclusion in EP6, with an odds ratio of 11.98 for
179 the protein product and 2.00 for the alternate allele, respectively. Therefore, the GWAS analysis
180 revealed that the signaling adaptor protein SHC4 may play an important mechanistic role in
181 contributing to severe disease pathology.

182

183 **EP6 has a distinct metabolomic signature associated with severe COVID-19: low levels of**
184 **sphingosine-1 phosphate are associated with COVID-19 ARDS**

185 To further characterize the pathobiology of severe COVID-19 disease, metabolomic profiling of
186 plasma samples was done in parallel to the SOMAmer analysis. The results yielded data on 1,435
187 metabolites, of which 576 were found significantly altered in EP6 (two-sided MWU FDR < 0.01)

188 Characterization of the plasma samples supported the distinction among the different EPs at the
189 levels of metabolomic sub-pathways and individual metabolites (Figure 5 and Table S6). The two
190 top sub-pathways (FDR < 1E-2) "Methionine, Cysteine, SAM and Taurine Metabolism" and
191 "phosphatidinylethanolamine (PE)", are known to interact (14, 15) and are associated with pro-
192 coagulant activities (see Discussion). Strikingly, sphingosine-1 phosphate was enriched in EP1 and
193 depleted in EP6 (Figure 5B), confirming the previously reported negative association of this
194 metabolite with COVID-19 ARDS (16). To identify relationships that may shed light on factors
195 influencing clinical laboratory results that define EP6, we also performed Spearman's rank
196 correlation analyses between each blood test value and metabolites (Table S2). Finally, to gain
197 further insight into the potential role of SHC4 in COVID-19 disease severity, we investigated the
198 metabolites associated with SHC4 in EP6. The only significantly correlated metabolite was 6-
199 oxopiperidine-2-carboxylate (Spearman's rank correlation = 0.40, FDR = 0.012), where the p-
200 values were corrected for the number of metabolites using Benjamini-Hochberg FDR.
201
202 **Mechanical ventilation in EP6 is associated with lipoprotein metabolism**
203 As noted in the introduction, mechanical ventilation in COVID-19 is associated with a high
204 mortality rate. Within EP6, 55 out of the 118 individuals received mechanical ventilation. Within
205 all EP6 members that share the overall similar aptamer signature, we identified a subset of
206 aptamers and metabolites that were significantly different between those with or without
207 mechanical ventilation (Table 5 and S7). The second most enriched aptamer (Table 5 and S7),
208 Angiopoietin-like 3 (ANGPTL3), corresponds to a liver-secreted regulator of lipoprotein
209 metabolism (17), which may reflect alterations observed in fatty acid metabolites (Table S7).

210 There were no significant differences between the two groups in terms of sex, age, or BMI (Table
211 S7). Finally, we investigated whether molecular information (proteomic and/or metabolomic)
212 was associated with mechanical ventilation duration. We identified Alpha-L-iduronidase to be
213 inversely correlated with duration of ventilation in the patients with mechanical ventilation in
214 EP6 (Spearman's rank correlation = -0.60, FDR = 2.57E-2, Table S7). Overall, these results point to
215 an association between altered lipoprotein metabolism and mechanical ventilation in EP6.
216

217 **Discussion**

218 The results presented herein came from the unsupervised clustering of the circulating proteome
219 from a large cohort of deeply-phenotyped SARS-CoV-2 positive hospitalized participants. This
220 broad-based approach has provided both expected and novel findings into the molecular
221 mechanisms regulating COVID-19. A principal advance is the identification of six COVID-19
222 disease endophenotypes with different clinical trajectories and distinguishing characteristics that
223 are summarized in Table 6. One endophenotype, (EP6) was associated with ARDS, the worst
224 clinical manifestation of COVID-19 that was reflected by a greater proportion of ICU admission,
225 mechanical ventilation, and severe/fatal outcomes (Figures 2 and 4). Clinical features of this
226 endophenotype were consistent with published literature including increased levels of CRP, D-
227 dimers, elevated circulating neutrophils, and reduced peripheral blood lymphocytes (Figure 3A,
228 Table S2). This information was sufficient to train an accurate prognostic model that could readily
229 support future clinical care, since it only requires data from routine clinical laboratory results for
230 prognosis. Our approach also enabled the identification of the following molecular features: 1)
231 the level of Sphingosine-1 phosphate is highest in EP1 and lowest in EP6, reflecting an inverse

232 association with time to admission to ICU and/or death; 2) interleukins and FGFR signaling are
233 molecular pathways associated with severe COVID-19; 3) genetic and environmental
234 determinants of SHC4 expression are associated with inclusion in EP6; 4) the combined proteomic
235 and metabolomic signature of EP6 supports a pathologic role for neutrophil-associated
236 procoagulant activity in COVID-19 ARDS; and 5) mechanical ventilation in EP6 is associated with
237 altered lipoprotein metabolism.
238
239 *The approach: unsupervised clustering capacity at identifying clinically meaningful*
240 *subpopulation*s
241 Our unsupervised clustering approach in conjunction with a rich molecular dataset enabled us to
242 identify six endophenotypes that could not be captured by simply classifying the population in
243 two bins solely based on severity. It also provided higher granularity enabling us to uncover
244 specific associations of molecular factors with blood variables (Table S2) or mechanical
245 ventilation (Table S7). Accordingly, we identified two endophenotypes with more favorable
246 outcomes (EP1 and EP2), three endophenotypes with intermediate outcomes in terms of severity
247 (EP3, EP4 and EP5) and one endophenotype which led to worst outcomes compared to all others
248 (EP6). In this manuscript, we have focused on the comparison between EP1 and EP6 to highlight
249 mechanisms of COVID-19-asssociated ARDS. While we have described high level features of EP2-
250 EP5 in Table 6, further detailed characterization of these other subgroups is ongoing and will be
251 reported subsequently.
252

253 The identification of endophenotypes was done systematically using robust consensus clustering
254 of aptamer expression levels in which the optimum number of clusters was determined
255 congruently using two well-established measures: AIC and BIC. The consensus clustering using
256 bootstrap sampling (1000 times) ensured identification of robust clusters that are not sensitive
257 to exclusion of some of the samples (20% randomly selected and excluded at each cycle). The
258 mean Rand-index between each of the 1000 subsampled clusterings and the final consensus
259 clustering was 0.823, reflecting a high degree of concordance and robustness. Moreover,
260 identifying the best number of clusters using AIC/BIC (both of which agreed with each other)
261 allowed us to reveal the patterns of the EPs directly from the data, instead of imposing a pattern
262 onto it through human supervision. This is an important strength of the study that enabled us to
263 identify distinct molecular patterns of patients that could have remained undetected using other
264 traditional approaches.
265
266 Additionally, to improve the translational applicability of the EPs, we developed a prognostic
267 model based only on measurement of conventional clinical laboratory blood markers to test the
268 generalizability of these endophenotypes to samples without measured aptamer expression
269 levels. Characteristics of EPs predicted solely based on their blood markers were consistent with
270 the original EPs, suggesting that clinical blood markers could be used as surrogates for
271 assignment of these EPs to new patients and potentially automating identification of high-risk
272 groups in the clinic. This approach takes into account the effect of multiple blood variables
273 simultaneously and incorporates the full distribution of each variable. This is in contrast to the
274 clinical laboratory results that are automatically flagged as within or outside normal range, one

275 variable at a time, therefore increasing the clinical applicability of our model by leveraging a wider
276 spectrum of information to prognosticate patient outcomes.
277
278 *COVID-19 ARDS molecular pathology: a counterbalance between ceramides and sphingosine is*
279 *associated with critical illness.*
280 The datasets used in this study carry rich molecular information on mechanisms of disease for
281 COVID-19. EP6 is characterized by its enrichment in ARDS (44% vs 8% in EP1, Figure 4), acute
282 kidney injury (45%) and liver dysfunction (19%) (Table S4). Low levels of Sphingosine 1-phosphate
283 (S1P) are associated with ARDS and were shown to be associated with greater ICU admission and
284 decreased survival in COVID-19 (16). Accordingly, we found that S1P levels are depleted in EP6
285 while they are enriched in EP1 (Figure 5B). Interestingly, the aptamer detecting neutral
286 ceramidase, an enzyme converting ceramides into sphingosine, is enriched in EP1 (FDR = 2.65E-
287 10, Table S5). While changes in protein abundance as detected by aptamers may not necessarily
288 reflect changes in enzymatic activity, it is interesting to note that dihydroceramides and
289 ceramides are depleted in EP1 (Table S6). Conversely, dihydroceramides and ceramides are
290 significantly enriched in EP6 (Table S6), suggesting that there is a shunting of the pathway away
291 from sphingosine towards more pro-inflammatory ceramides that are associated with metabolic
292 disorders (18). Moreover, one of the top aptamers found enriched in EP6 (FDR = 2.10E-56, Table
293 S5) is serine palmitoyltransferase 2 (SPTLC2), an enzyme that is responsible for the rate-limiting
294 step in *de novo* sphingolipid biosynthesis (19, 20). These results suggest a counterbalance
295 between ceramides and sphingosine, where the former is associated with poorer outcomes

296 during critical illness, whereas higher levels of the latter is associated with more favorable
297 outcomes in ARDS.
298
299 *Multimodal profiling of EP6 supports a role for neutrophil-derived procoagulant activity-mediated*
300 *organ damage in COVID-19.*
301 COVID-19 ARDS was found to be more frequently associated with thrombotic complications than
302 non-COVID-19 ARDS (21, 22). In addition, histopathologic evidence of pulmonary microthrombi
303 is frequently observed in autopsies of individuals deceased with COVID-19 (23). Early reports
304 suggested a role for neutrophil-mediated release of NETs contributing to endothelial dysfunction
305 as a mechanism of microthrombosis in COVID-19 ARDS (24-35). Cathepsin G, a neutrophil-derived
306 serine protease (serprocidin), is the top-enriched aptamer in EP6 (FDR = 4.10E-58, Table S5).
307 Notably, two other serprocidins, neutrophil elastase and proteinase-3, are also significantly
308 enriched in EP6 (FDR = 1.04E-10 and FDR = 8.27E-7, respectively). In addition to their roles in
309 bacterial killing, serprocidins couple innate immunity to coagulation, where they promote
310 coagulation by enhancing tissue factor and factor XII-dependent coagulation (36). Serprocidins
311 are also important regulators of NETosis (37). In pediatric sepsis, another NETosis-associated
312 protein Olfactomedin 4 (OLFM4) (38) is associated with increased odds of having greater organ
313 failure and death (39). In accordance with these findings, OLFM4 is greatly enriched in EP6 (top
314 5%; FDR = 2.61E-46, Table S5). Collectively, these findings point to enhanced neutrophil-derived
315 procoagulant activity in EP6 in response to SARS-CoV-2 infection that is associated with worse
316 outcomes.
317

318 The procoagulant environment associated with COVID-19-associated organ damage is further
319 supported by the alteration in metabolic profiles in EP6. "Methionine, Cysteine, SAM and Taurine
320 Metabolism" and "phosphatidinylethanolamine (PE)" (Figure 5, Table S6) are two sub-pathways
321 known to interact (14), with PE methylation acting as a major consumer of S-Adenosylmethionine
322 (SAM) leading to the synthesis of S-Adenosylhomocysteine (SAH) and cystathionine (15), which
323 itself has been found upstream of 2-hydroxybutyrate and 2-aminobutyrate in a model of
324 hepatoxicity (40). SAH, cysthathionine, 2-hydroxybutyrate and 2-aminobutyrate are all
325 significantly enriched in EP6, and this group also exhibits a higher rate (FDR < 1E-2) of liver
326 dysfunction (Figure 4, Table S4), with significant abnormalities in several relevant clinical blood
327 markers including ALT, albumin, bilirubin and LDH (Figure 3A, Table S2). PEs become exposed at
328 the surface of cell membranes upon exposure to stress, inflammation, and cell death (41, 42). In
329 a Syrian hamster model, infection with SARS-CoV-2 markedly increased circulating PE expression
330 in the animals that were fed a high salt, high fat diet, demonstrating the interaction between
331 infection and metabolic disorder (43). Furthermore, the membranes of azurophilic granules,
332 which contain serprocidins, are enriched in PE (44). Accordingly, the significantly enhanced PE
333 abundance in EP6 is associated with decreased platelet counts, increased D-Dimers, INR and
334 activated partial thromboplastin time (aPTT) (Figure 3A, Table S2), hallmarks of abnormally
335 activated coagulation pathways that are associated with serious and often lethal complication of
336 sepsis (45). Taken together, the metabolomic profile of EP6 is associated with increased
337 circulating pro-coagulant activity that can be linked to organ damage, with notable enrichment
338 in liver damage and acute kidney injury in the cluster (Figure 4). Recent results from the
339 Antithrombotic Therapy to Ameliorate Complications of COVID-19 (ATTACC) study showed

340 anticoagulation to be efficacious in moderate but not severe COVID-19 disease (46, 47), revealing
341 a need for alternative management approaches in critically ill patients.
342
343 *Tryptophan and polyamine metabolism in COVID-19 acute kidney injury*
344 Altered kynurenine levels following SARS-CoV-2 infection have been shown to be associated with
345 increased IL-6 expression and kidney injury (48). Interestingly, we also observed kynurenine to
346 be enriched in EP6 but depleted in EP1, supporting the association of increased kynurenine with
347 COVID-19 severity and higher incidence of acute kidney injury. Spermidine/spermine N(1)-
348 acetyltransferase (SSAT) contributes to polyamine synthesis and its extracellular metabolite, N1,
349 N12-diacetylspermine, is one of the top 15 metabolites enriched in EP6 (Figure 5B). N1, N12-
350 diacetylspermine is positively correlated with urea and creatinine is negatively correlated with
351 lymphocyte numbers (Table S2). Deletion of SSAT in mice is protective against LPS-induced kidney
352 injury (49), and SSAT activity is associated with an increased white blood cell count in acute and
353 chronic myeloid leukemia patients (50). Taken together, these observations suggest that the
354 tryptophan and polyamine metabolites are associated with acute kidney injury in COVID-19 and
355 identifies another potential pathway of disease progression leading to organ damage.
356
357 *Novel molecular markers of COVID-19 pathology: FGFR and SHC4*
358 Two novel molecular factors associated with COVID-19 ARDS that were identified by our study
359 are FGFR and SHC4. Circulating levels of the pro-angiogenic FGF-2 has been associated with
360 COVID-19 severity and creatinine levels in a study of 208 SARS-CoV-2 positive participants (51).
361 It is also noteworthy that the use of Nintedanib, an inhibitor of FGFR, Vascular Endothelial Growth

362 Factor Receptor (VEGFR) and Platelet-Derived Growth Factor Receptor (PDGF-R) that is approved
363 for use in interstitial lung disease, improved pulmonary inflammation and helped three middle-
364 aged obese COVID-19 patients with markedly impaired lung function wean off mechanical
365 ventilation (52). While the adaptor protein SHC4 has not been experimentally demonstrated to
366 modulate FGFR signaling, a 12-gene biomarker signature associated with melanoma contains
367 FGFR2, FGFR3 and SHC4 (53). It is attractive to speculate that SHC4 may act downstream of FGFR
368 or other associated growth factor receptors, favoring heightened procoagulant activity
369 associated with COVID-19 ARDS. In view of the limited knowledge of this understudied member
370 of the SHC family, we looked at the metabolites associated with SHC4 to gain insights into its
371 possible functions. Interestingly, 6-oxopiperidine-2-carboxylate (the only metabolite with a
372 significant correlation with SHC4 in EP6) was found to be negatively associated with glomerular
373 filtration rate in a GWAS study of kidney disease and hypertension in African-Americans (54).
374 Acute kidney injury is a frequent complication of acute liver failure (55), and liver dysfunction is
375 associated with abnormal coagulation (56). Thus, the overall molecular information coming from
376 the multi-modal analysis of EP6 points to a triad between liver function, kidney function and
377 hemostasis that becomes dysfunctional following ARDS-associated inflammation driven by SARS-
378 CoV-2 infection. Whether the presence of SHC4 in circulation is a marker of dysfunction of this
379 triad, or an active mediator remains to be determined. Moreover, the identity of the cells
380 expressing SHC4 leading to its presence in the circulation is not known. Taken together, our
381 identification of FGFR and SHC4 signaling pathways distinguishing EP6 from other
382 endophenotypes provides a rationale for investigation of their potential utility as biomarkers of

383 severe disease activity and/or the use of antagonists of those pathways to treat severe COVID-
384 19 disease manifestations.
385
386 *Molecular mechanisms of mechanical ventilation in EP6: a potential role for ANGPTL3*
387 ANGPTL3 is a liver-secreted enzyme that can inhibit lipoprotein and endothelial lipase (57). Loss-
388 of-function of ANGPTL3 is associated with hypolipidemia (58). Therefore, increased circulating
389 levels of ANGPTL3 would be expected to raise plasma levels of triglycerides and phospholipids,
390 which is consistent with the observed metabolomic profiles (Table S7). Whether these alterations
391 are sufficient to increase the likelihood of mechanical ventilation remains speculative.
392 Nevertheless, it is interesting to note that administration of medium- and long-chain triglycerides
393 can aggravate gas exchange abnormalities in ARDS (59) and may warrant the preclinical
394 investigation of ANGTPL3 antagonists like Evinacumab or other functionally similar therapeutic
395 agents (60) to ameliorate outcomes in ARDS. As for the duration of mechanical ventilation, we
396 discovered an inverse correlation with alpha-L-iduronidase (IDUA), a lysosomal enzyme required
397 for the degradation of heparan sulphate and dermatan sulphate (61). Our current data cannot
398 distinguish between the presence of this enzyme as a biomarker or as an active mediator that
399 influences the course of COVID-19-associated acute respiratory failure.
400
401 *Limitations and considerations*
402 The data presented in this study come from individuals participating in the BQC19, a prospective
403 observational cohort built to study COVID-19 in Québec (Canada) with its specific population
404 profile as reported previously (8). A chronological bias may also be present, as most of the

405 participants used for endophenotyping in this study were recruited during the first two waves of
406 the pandemic (Figure S1), prior to widespread vaccination in Québec and the appearance of the
407 Omicron variant and sub-variants. Therefore, some of the features of the identified
408 endophenotypes may change over the course of the pandemic. It will be essential to continue
409 longitudinal assessments of the molecular profiles to better understand the dynamic nature of
410 host-pathogen interactions. It will also be interesting to compare the profiles of COVID-19 ARDS
411 to other viral-induced ARDS, to identify common as well as distinguishing features of these
412 conditions.
413
414 In this study, we used circulating proteome as determined by aptamers to identify
415 endophenotypes, a task for which it is well suited as it can capture a dynamic landscape.
416 However, several important considerations need to be mentioned when using this approach.
417 First, raw (unnormalized) expression values of the same aptamer can be used to compare
418 different samples, but these values cannot be used to compare different aptamers against each
419 other in the same sample, since they only show the relative abundance of expression and not
420 absolute expression. As such, one needs to first normalize these values across samples (one
421 aptamer at a time) and then subject them to follow-up analysis such as clustering as we have
422 done in this study. Second, since information only shows relative abundance, focusing on an
423 individual aptamer and analyzing its absolute abundance will require additional measurements.
424 Moreover, to analyze aptamers individually (instead of the collective approach that we used in
425 this study), one must consider the effect of complexes and non-specific binding that may result
426 in noisy data. As such, we favored collective analysis of the aptamer expression data after proper

427 normalization, which enabled us to identify various EPs and important molecular mechanisms
428 discussed in this study.
429
430 *Future directions*
431 There is enormous amount of data within this study and BQC19 that could be exploited by the
432 scientific and medical community to improve our understanding of COVID-19 and novel emerging
433 acute respiratory illnesses. Analyzing the molecular profiles of the other EPs, in particular EP3-5,
434 which led to similar outcomes from distinct molecular pathways (Table 6) in more detail should
435 further yield important insights into mechanisms of the disease. For example, EP5 is enriched in
436 males and cardiovascular disease complications, while EP4 is enriched in females (like EP1), yet
437 both of these groups exhibit different distinct clinical trajectories that point to sex-dependent
438 and additional sex-independent mechanisms of disease.
439
440 **Conclusion**
441 The major strength of this study is the use of an unsupervised approach for analysis of a large
442 and deeply phenotyped cohort. We showed that this strategy can address both fundamental
443 scientific questions pertaining to mechanisms of disease and help the medical community
444 improve clinical outcomes through early identification of patients that may experience a severe
445 clinical course during COVID-19 infection.
446
447 **Methods:**
448 **Datasets and preprocessing**

449 The Biobanque Québécoise de la COVID-19 (BQC19; www.quebeccovidbiobank.ca) is aimed at
450 coordinating the collection of patients' data and samples for COVID-19 related research. Data
451 and samples were collected from ten sites across the province of Québec (Canada) (8). BQC19
452 organizes the collected data, including clinical information and multi-omics experimental data,
453 before making it available in successive releases. For this study, we used release #5 of the clinical
454 data published in December 2021, the circulating proteome determined using SOMAmers and
455 Metabolomics data (62). BQC19 GWAS imputation data was generated by Tomoko Nakanishi at
456 Brent Richards lab, Jewish General Hospital and McGill University. Detailed codes used for
457 generating the data can be found in: https://github.com/richardslab/BQC19_genotype_pipeline.
458
459 Our main corpus of analysis consisted of n = 1,362 hospitalized and SARS-CoV-2 positive patients
460 (based on qRT-PCR) of BQC19. This included n = 731 patients for which both clinical and
461 proteomic data were available as well as n = 631 patients for which proteomic data was not
462 available, but their clinical data contained measurements for more than half (at least 11 out of
463 21) of the blood markers that we used as a validation set for the prognostic model developed in
464 this study.
465
466 We also obtained data (n = 731) corresponding to the circulating proteome measured between
467 April 2, 2020 and April 20, 2021 by a multiplex SOMAmer affinity array (Somalogic, 4,985
468 aptamers) from BQC19 (release #3 Sep. 2021, associated patients' data updated in release #5
469 Dec. 2021). When measurements of the same patients but at different time points were
470 available, we used the one corresponding to the first time point. SomaScan is a biotechnological

471 protocol commercialized by the Somalogic company (9). It relies on a set of artificial aptamers
472 linked to a fluorophore and each designed to bind a single protein. Once added to the sample,
473 the activity of each aptamer is measured through fluorescence and used to approximate the
474 expression level of the targeted protein. SomaScan protocol comprises several levels of
475 calibration and normalization to correct technical biases. Log2 and Z-score normalization were
476 performed on each aptamer separately in addition to the manufacturer's provided normalized
477 data (hybridization control normalization, intraplate median signal normalization, and median
478 signal normalization). Since the data was analyzed by Somalogic in two separate batches, we
479 applied the z-score transformation separately to each batch, to reduce batch effects. These
480 additional transformations ensure that the measured values of different aptamers are
481 comparable and can be used in cluster analysis.
482
483 We obtained metabolomic data (1,435 metabolites) from BQC19. We used the batch-normalized,
484 missing values imputed and log-transformed version of the data.
485
486 **Analyses of the GWAS dataset**
487 For the GWAS analyses, annotation of SNVs were done using the biomaRt package (63) from R
488 (64) and all analyses were done using R version 4.1.3. Quality control steps were derived in
489 majority from a 2017 QC tutorial article (65). At the beginning, we had 867,450 markers and 2,429
490 samples. We import Plink format data into R using the "read_plink" function from genio package
491 (66) from R. We removed 103,592 non ACGT bi-allelic markers. We calculated the predicted sex
492 by looking at the rate of homozygote markers on chromosome 23. We removed 3,588 markers

493 with call rates < 98%, 448,932 monomorphic markers and markers with MAF <0.05, and 28,092
494 markers with Hardy Weinberg equilibrium < 1E-6 (calculated by the "HWE.exact" function from
495 the genetics package (67) from R. For the EP6 cluster group analyses, we removed in addition
496 1,747 samples that were not in the cluster analysis, 8 samples with a sex discrepancy (based on
497 predicted sex calculated earlier and reported sex), and 3 samples with a heterozygosity rate > 3
498 standard deviations. We finally removed a pair of samples who had approximately the same
499 genome probably due to an error of manipulation. We couldn't know which one was the right
500 sample, so we removed both of them. We also had 2 pairs of individuals who had a pi-hat of ~0.5
501 (meaning first degree relatives), we decided to keep one sample per pair, the one with the higher
502 call rate. All other pairs of individuals had a pi-hat < 0.21 that is judge acceptable considering our
503 population. Pi-hat were calculated with the "snpgdsIBDMoM" function from SNPRelate package
504 (68) from R. At the end of quality controls, 283,246 markers on 655 samples have been used to
505 perform association analyses.
506
507 To perform the principal component analyses (PCA), we took a subsample of independent
508 markers (pruning) with a maximum sliding window of 500,000 base pairs and a linkage
509 disequilibrium (LD) threshold of 0.2 using the "snpgdsLDpruning" function from the SNPRelate
510 package from R. We ran the PCA with the "snpgdsPCA" function from SNPRelate package from
511 R. The first 2 principal components (PCs) were considered significant.
512
513 For the GWAS analyses of the 283,246 remaining markers between EP6 cluster compared to all
514 others, we modeled a logistic regression with the dichotomous variable indicating if the

515 participants belong to the EP6 cluster as the outcome variable. We used the additive model for
516 markers as an independent variable and we adjusted the models with the first two PCs. Odds
517 ratio and p-values were calculated on each model. QQ-plots have been performed as quality
518 control of the models; p-values were plotted using "qqplot.pvalues" function from gaston
519 package (69) from R (data not shown). We compared the aptamers' normalized level of
520 expression (based on normalization described earlier) between the three groups of genotypes
521 for each studied genes by performing standard ANOVA analyses followed by Tukey *post hoc* tests
522 (referred to in this study as pQTL analysis). Since aptamers tested are limited compared to the
523 SNVs, we have fixed significance p-values threshold below 1E-4 to report more SNVs instead of
524 the more common 1E-5 suggestive threshold. Finally, to identify if the cluster EP6 may be
525 explained by the aptamers and the SNVs, we performed multiple logistic regression analyses
526 models include aptamers expression values, SNV genotypes (additive model) and the two
527 principal components. OR are reported with 95% confidence intervals.
528
529 **Consensus agglomerative clustering**
530 Patients were clustered using agglomerative clustering, with Euclidean distance and Ward's
531 linkage (10, 11). To identify number of clusters k, we used the elbow method based on the Akaike
532 Information Criterion (AIC) and Bayesian Information Criterion (BIC). More specifically, we
533 calculated the AIC and BIC for clustering using k = 2, 3, …, 20 and used the Kneedle algorithm (70)
534 to identify the value of k corresponding to the "elbow", where increasing the value of k does not
535 provide much better modeling of the data. Kneedle identified k = 6 as the number of clusters
536 based on both AIC and BIC (Figure 1A).

537

538 Given the number of clusters in the data, we then used consensus clustering with sub-sampling
539 to obtain robust endophenotypes. We randomly sampled 80% of the patients 1000 times. Each
540 time, we used agglomerative clustering above with k = 6 to identify clusters. Given these 1000
541 clusterings, we calculated the frequency of two patients appearing in the same cluster, when
542 both were present in the randomly formed dataset. We then performed one final agglomerative
543 clustering of these frequency scores to identify the six endophenotypes (Figure S2A and Figure
544 1B).

545

546 **Metabolomic pathway characterization of EP6**

547 The 1435 metabolites measured were organized into 122 sub-pathways in the original dataset
548 (denoted as "SUB_PATHWAYS"). We first identified metabolites whose values were significantly
549 higher or lower in EP6 compared to other EPs (two-sided MWU test, FDR<0.01). Then, we used
550 these metabolites to perform pathway enrichment analysis (one-sided Fisher's exact test) based
551 on 122 pathways. The resulting p-values (Table S6) were then corrected for multiple tests using
552 Benjamini-Hochberg FDR.

553

554 **Nearest-centroid predictor based on blood markers**

555 In order to predict endophenotypes from blood tests, we developed a missing-value resilient
556 nearest-centroid classifier. We used the set of patients that were used to form the original EPs
557 (n = 731) as the training set and the set of patients that did not have proteome data as the
558 validation set (n = 631). First, we z-score normalized each blood marker across all the patients in

559 the training set, one marker at a time. We then formed a blood marker signature (a vector of
560 length 21) for each EP. Each element of an EP's signature corresponds to the mean of the
561 corresponding marker across all patients of that EP.
562
563 To predict the EP label of each patient in the test set, we first z-score normalized their blood
564 marker measurements using the mean and standard deviation of the blood markers calculated
565 from the training set. Then, we calculated the cosine distance between each test patient's blood
566 marker profile and the centroids (excluding missing values) and identified the nearest EP as the
567 predicted EP (PEP) label of the patient.
568
569 **Statistics**
570 Several non-parametric tests, including Mann Whitney U test, Fisher's exact test, and Spearman's
571 rank correlation, were used in this study. Benjamini-Hochberg false discovery rate (FDR) was used
572 to adjust the p-values for multiple tests. Details of statistical tests were discussed where they
573 were used.
574
575 **Study approval**
576 The study was approved by the Institutional Ethics Review Board of the "Centre intégré
577 universitaire de santé et de services sociaux du Saguenay-Lac-Saint-Jean" (CIUSSS-SLSJ) affiliated
578 to Université de Sherbrooke [protocol #2021-369, 2021-014 CMDO – COVID19].
579
580 **Acknowledgements**

This work was made possible through open sharing of data and samples from the Biobanque Québécoise de la COVID-19, funded by the Fonds de recherche du Québec - Santé, Génome Québec, the Public Health Agency of Canada and, as of March 2022, the Ministère de la Santé et des Services Sociaux du Québec. We thank all participants to BQC19 for their contribution. This study was supported by the Fonds de recherche du Québec - Santé (FRQS)- Cardiometabolic Health, Diabetes and Obesity Research Network (CMDO)- Initiative. This work was also supported by Natural Sciences and Engineering Research Council of Canada (NSERC) grant RGPIN-2019-04460 (AE).

Conflict of Interests

The authors have declared that no conflict of interest exists.

Figures

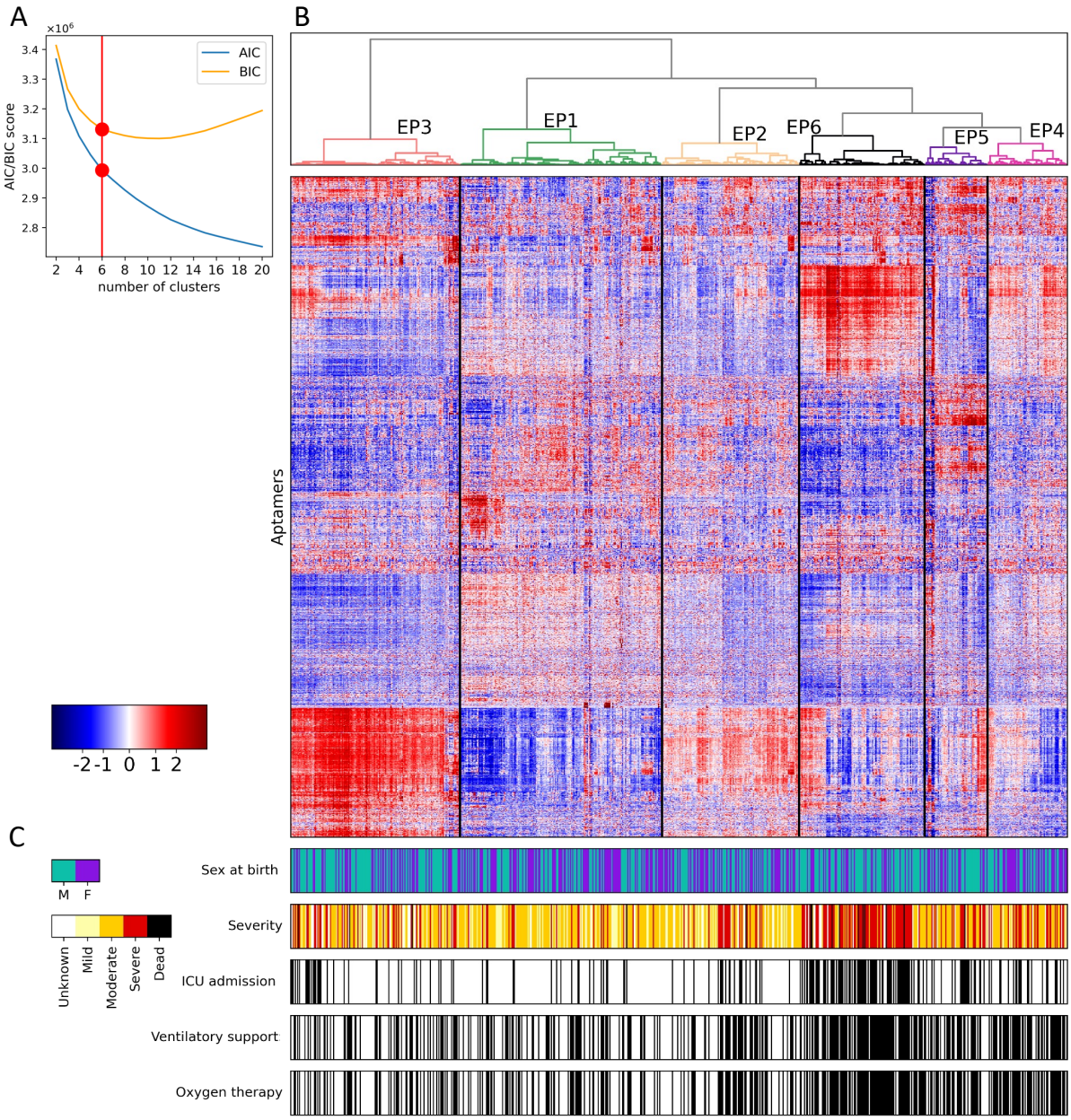

Figure 1: Unsupervised consensus clustering of SARS-CoV-2 positive patients. A) The elbow points (circles in red) of Akaike's Information Criteria (AIC) and Bayesian Information Criteria (BIC) curves versus number of clusters consistently corresponded to k=6 as the optimal number of clusters. B) The heatmap shows the expression of aptamers (rows) in each sample (columns). The dendrogram shows the identified endophenotypes. C) Characterization of samples based on sex at birth, highest world health organization (WHO) severity level achieved, intensive care unit (ICU) admission, ventilatory support, and oxygen therapy. For the last three rows, a sample colored “black” reflects a label of “yes”.

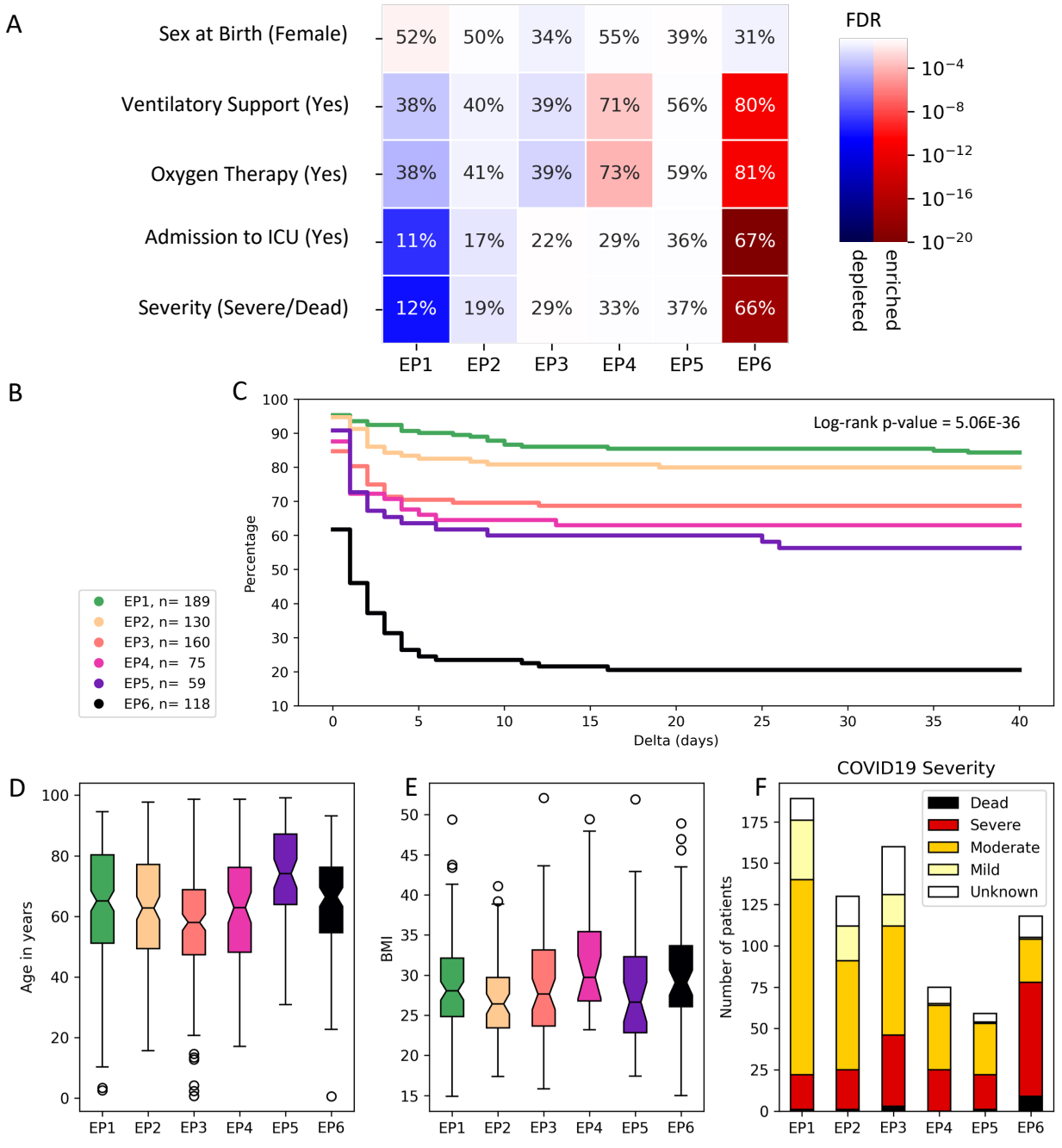

Figure 2: Characterization of endophenotypes (EPs).
A) Enrichment or depletion of each EP in clinical variables (one cluster versus rest). Two-sided Fisher's exact tests are used to calculate the p-values, which are corrected for multiple tests using Benjamini-Hochberg false discovery rate (FDR). Gradients of blue show depletion, while gradients of red show enrichment. FDR values above 0.05 are depicted as white. B) The number of patients in each EP and the colors used to represent them in panels C, D, and E. C) Kaplan-Meier analysis of the time between patients' admission to the hospital and their admission to intensive care unit (ICU) (or death if earlier) for each EP (Delta). D) Distribution of age in each EP. E) Distribution of BMI in each EP. F) COVID-19 severity in each EP.

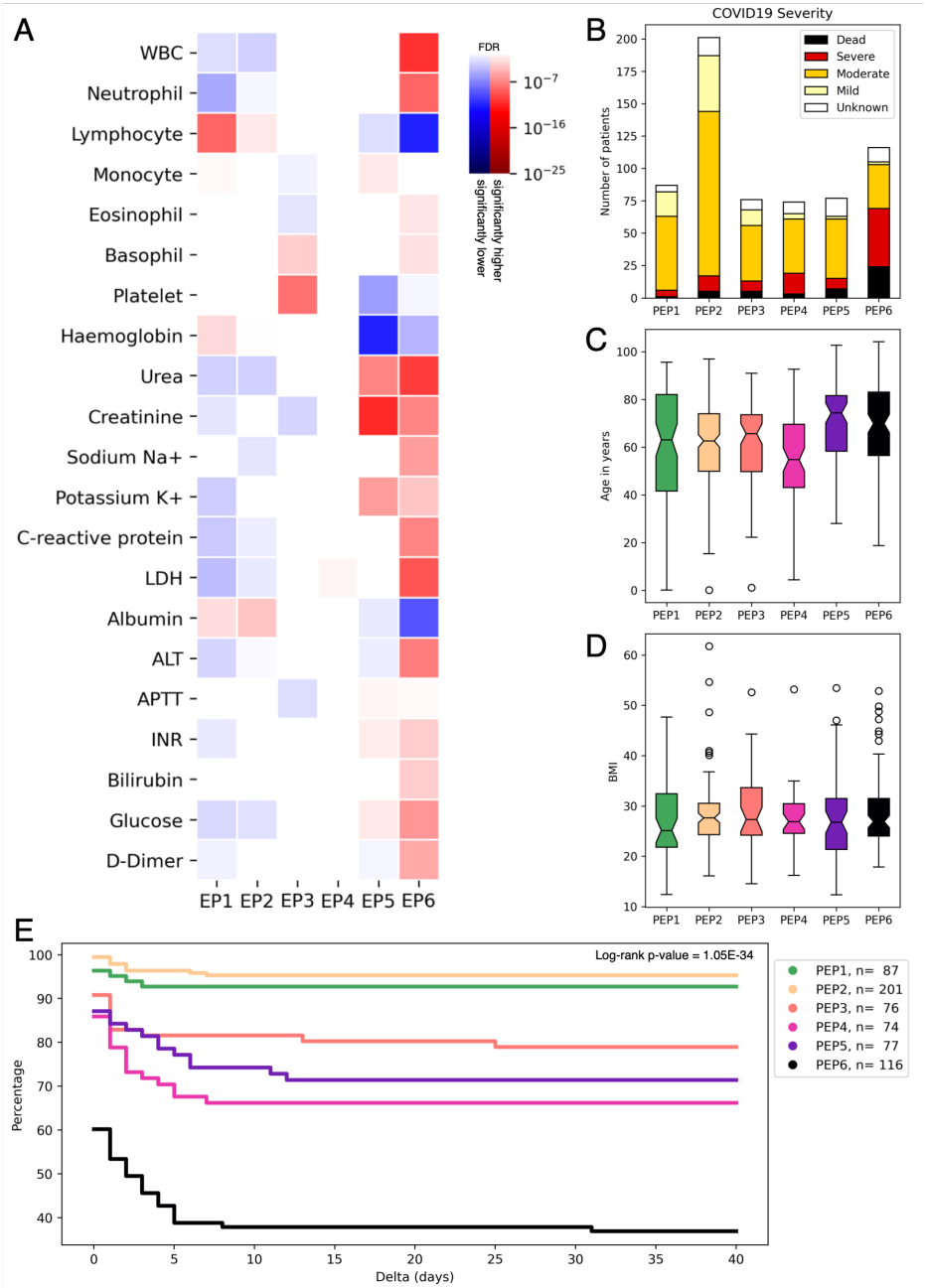

Figure 3: Differential expression patterns of blood markers in the identified endophenotypes (EPs) and characterization of predicted endophenotypes (PEPs) based on the prognostic model. A) Heatmaps of false discovery rates (FDR) values for two-sided one-vs-rest Mann–Whitney U tests for 21 blood markers for each EP. Abbreviations used: WBC = white blood cells, LDH = lactate deshydrogenase, ALT = alanine aminotransferase, aPTT = activated partial thromboplastin time, INR = International Normalized Ratio. FDR values below 0.05 are shown as white. B) World health organization COVID-19 severity assessed in each PEP. C) Distribution of age in each PEP. D) Distribution of BMI in each PEP. E) Kaplan-Meier analysis of the time between patients' admission to the hospital and their admission to ICU (or death if earlier) for each PEP (Delta). The colormap

in panel E shows the number of patients in each PEP group and the colors used to represent them in panels C, D, and E. PEPs were predicted based on blood markers and were not used to originally identify the EPs.

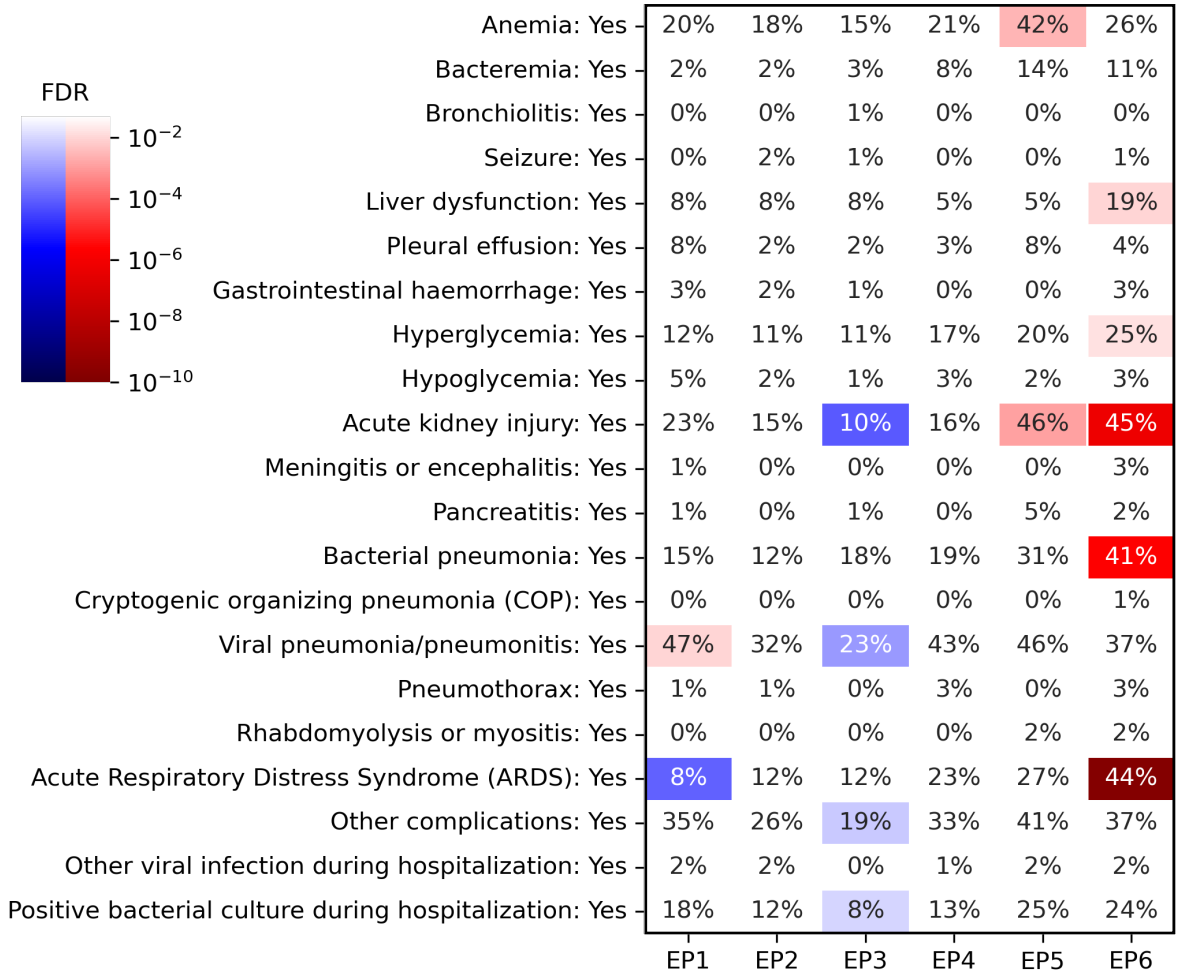

Figure 4: Frequency and significance of complications in different EPs.
The value in each cell shows the percentage of patients of that EP (column) that suffered from the complication (row). The colors represent two-sided Fisher's exact test false discovery rate (FDR, corrected for multiple tests). Red represents enrichment, while blue represents depletion. FDR values below 0.05 are shown as white.

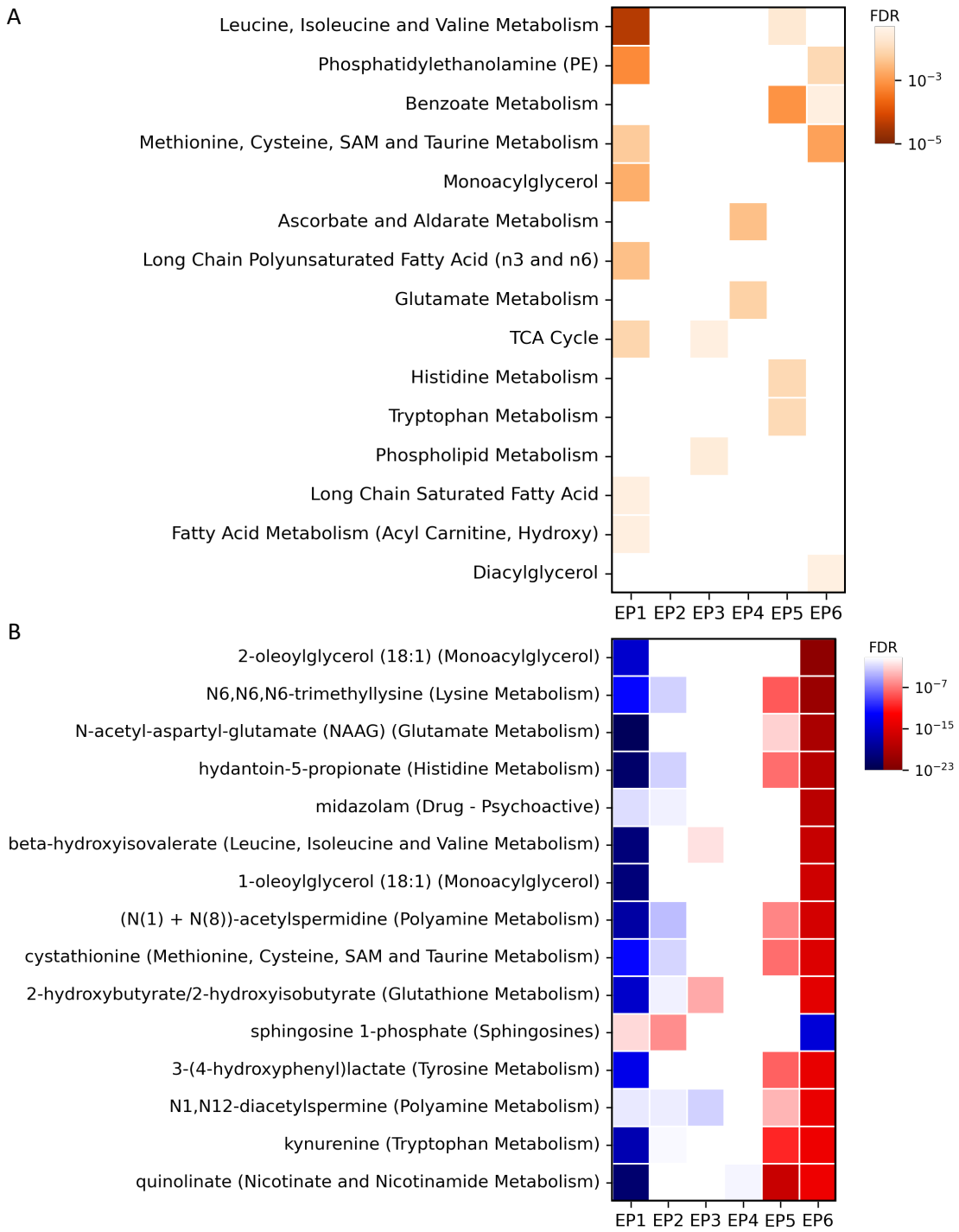

Figure 5: Metabolite characteristics of endophenotypes (EPs).
A) The heatmap shows the enrichment (one-sided Fisher's exact test) of EPs in different metabolite sub-pathways. See Table S6 for the full list. B) The heatmap shows the over-expression (red) and under-expression (blue) of metabolites in different EPs (two-sided Mann–Whitney U test). Row names show metabolites followed by the sub-pathway to which they belong in parentheses. Only top 15 metabolites (based on false discovery rate for EP6) for which a definite name and sub-pathway was available are shown. Full list is provided in Table S6.

645 **Tables**

646 **Table 1:** Clinical and pathological characteristics of the BQC19's participants used in this study.

		Cohort (n = 1,362) No. (%)
Age in years	<45	17.5
	45-65	33.6
	>65	48.7
	Unknown	0.3
Body mass index in kg/m^2	<20	3.8
	20-25	13.0
	25-35	27.8
	>35	7.4
	Unknow	47.9
Sex at birth	Female	44.8
	Male	55.2
Severity	Dead	4.4
	Severe	21.8
	Moderate	51.0
	Mild	11.8
	Unknown	10.9
Oxygen therapy	Yes	48.2
	No	21.0
	Unknown	30.8
Ventilatory support	Yes	47.6
	No	14.0
	Unknown	38.5
Admission to intensive care unit	Yes	17.5
	No	33.6
	Unknown	48.7

647
648

649 **Table 2:** Reactome pathways associated with EP6, based on expression of aptamers. KnowEnG
650 analytical platform was used. The p-values were calculated using a one-sided Fisher's exact test
651 and were corrected for multiple tests using Benjamini-Hochberg method. Only signaling
652 pathways with FDR<5E-4 are shown in this table (see Table S5 for the full list).
653

Pathway	FDR
Signaling by Interleukins	2.75E-12
Cytokine Signaling in Immune system	6.25E-10
Immune System	2.28E-06
Signaling by FGFR	4.75E-06
Constitutive Signaling by Aberrant PI3K in Cancer	2.94E-05
FGFR2 ligand binding and activation	5.77E-05
Extracellular matrix organization	1.85E-04
FGFR3 ligand binding and activation	1.85E-04
FGFR3c ligand binding and activation	1.85E-04
FGFRL1 modulation of FGFR1 signaling	1.85E-04
TNFs bind their physiological receptors	2.58E-04
PI3K/AKT Signaling in Cancer	3.28E-04
PI5P, PP2A and IER3 Regulate PI3K/AKT Signaling	4.14E-04
Interleukin-4 and 13 signaling	4.91E-04

654
655

656 **Table 3**: SNVs differentiating EP6 against all other endophenotype clusters.

SNV ID	Position	Gene Symbol	SNV [1]	MAF	HWE	OR [2]	p-value
rs1394671	chr5:7348567	-	G > A	0.240	0.095	2.28	2.41E-06
rs12186698	chr5:168880668	*SLIT3*	T > C	0.087	0.247	2.93	3.72E-06
rs11625406	chr14:50049474	-	C > A	0.438	0.137	0.47	1.03E-05
rs11862889	chr16:83828886	-	T > C	0.104	0.446	2.62	1.07E-05
rs2995918	chr4:37905775	*TBC1D1*	C > T	0.494	0.776	0.49	1.20E-05
rs16897810	chr5:68415296	-	G > A	0.093	0.001	2.67	1.43E-05
rs2376263	chr17:35436659	*SLFN13*	G > A	0.261	0.073	2.01	2.48E-05
rs4790712	chr17:1614620	*SLC43A2*	G > A	0.441	0.003	1.97	2.50E-05
rs2294566	chr20:41472939	*CHD6*	C > A	0.239	0.265	2.04	2.78E-05
rs7164451	chr15:48921859	*SHC4*	G > A	0.386	0.053	1.89	3.01E-05
rs57664621	chr8:22808443	*PEBP4*	G > C	0.174	0.016	2.16	3.07E-05
rs28482919	chr3:14903094	*FGD5*	T > C	0.187	0.053	0.32	3.20E-05
rs6559283	chr9:89712560	-	T > C	0.375	<0.001	1.96	3.27E-05
rs657075	chr5:132094425	-	A > G	0.104	0.328	2.49	4.28E-05
rs56235109	chr15:62424001	*TLN2*	A > G	0.235	0.366	0.38	4.67E-05
rs7620057	chr3:179473377	*GNB4*	T > C	0.099	1.000	2.51	4.67E-05
rs10466868	chr12:131455375	-	T > G	0.108	0.401	2.51	5.94E-05
rs2236798	chr1:18735127	*PAX7*	A > G	0.054	0.114	2.90	6.11E-05
rs3774814	chr4:5464702	*STK32B*	C > G	0.205	<0.001	0.36	6.75E-05
rs4497815	chr19:22903215	-	G > A	0.216	0.205	2.07	7.52E-05
rs6765694	chr3:54601810	*CACNA2D3*	G > A	0.404	0.833	0.51	7.66E-05
rs2797773	chr6:37559045	-	C > T	0.422	0.000	0.53	8.06E-05
rs17014760	chr4:129419725	-	A > A	0.316	0.000	1.89	9.33E-05
rs10948260	chr6:45835559	-	G > A	0.366	0.189	1.81	9.52E-05
rs12035677	chr1:232391209	-	A > G	0.063	0.000	3.02	9.95E-05

657 Abbreviations used: SNV = Single nucleotide variation, HWE = Hardy Weinberg Equilibrium, MAF
658 = Minor allele frequency, OR = Odd ratio.
659
660 [1] SNV are described following GWAS annotations: refence allele > alternative allele (e.g., G > A).
661 [2] Logistic regression analyses using additive model adjusted for the 2 principal components.
662

663 **Table 4:** Association between genotypes and aptamer expression levels

SNV	Gene Symbol	Nearest gene	Aptamers normalized expression [1]				Multiple logistic regression analyses [3]			
			HM_{ref}	**HTZ**	**HM_{alt}**	p-val [2]	**Aptamers** OR	p-val	**SNV** OR	p-val
rs6765694	*CACNA2D3*	-	-0.09 (±0.89)	0.01 (±1.10)	0.24 * (±0.99)	**0.011**	0.61 (0.46-0.79)	2.5E-4	0.52 (0.37-0.74)	2.4E-4
rs10948260	-	*CLIC5*	0.02 (±0.92)	0.02 (±1.09)	-0.02 (±0.97)	0.796	0.68 (0.55-0.83)	1.7E-4	1.83 (1.35-2.48)	8.8E-5
rs657075	-	*IL3*	-0.01 (±1.01)	0.05 (±0.99)	0.32 (±0.46)	0.328	1.37 (1.14-11.64)	8.2E-4	2.49 (1.60-3.88)	5.2E-5
rs7164451	*SHC4*	-	-0.07 (±0.90)	0.01 (±1.05)	0.23 * (±1.16)	**0.017**	11.98 (7.61-18.87)	8.4E-27	2.00 (1.30-3.09)	1.8E-3
rs12186698	*SLIT3*	-	0.00 (±1.02)	0.07 (±1.06)	0.15 (±0.89)	0.509	0.82 (0.64-1.06)	0.135	2.99 (1.89-4.73)	2.7E-6
rs56235109	*TLN2*	-	-0.03 (±0.98)	0.06 (±1.03)	0.04 (±1.27)	0.353	0.58 (0.45-0.75)	2.7E-5	0.38 (0.23-0.60)	4.7E-5

664 Abbreviations used: SNV = Single nucleotide variation, HM_{ref} = Homozygotes for the reference
665 allele, HM_{alt} = Homozygotes for the alternative allele, HTZ = Heterozygotes, OR = Odd ratio.
666
667 [1] Aptamers' normalized levels of expression are reported as mean (± standard deviation).
668 Normalization steps for aptamer expressions are described in Methods.
669 [2] p-value, standard ANOVA analyses followed by Tukey *post hoc* analyses. Asterisks (*) identify
670 difference between HM_{ref} and HM_{alt} genotypes.
671 [3] Multiple logistic regression analyses models include aptamers expression values, SNV
672 genotypes (additive model) and the two principal components. OR are reported with 95%
673 confidence intervals in parentheses.
674

675 **Table 5:** Aptamers associated with mechanical ventilation in EP6.
676 In this table, p-values are obtained using a two-sided Mann Whitney U test and is corrected for
677 multiple tests using Benjamini Hochberg FDR. A value of +1 (-1) in the "sign" column shows that
678 the value of aptamer was higher (lower) in the samples of 55 patients that received mechanical
679 ventilation in EP6, compared to the rest of EP6 members. Only 14 targets with FDR < 1.2E-02 are
680 shown. The full list is provided in Table S7.
681

Target	Target Full Name	p-values	FDR	Direction
GGE2D	G antigen 2D	2.93E-08	1.46E-04	+1
ANGPTL3	Angiopoietin-related protein 3	2.76E-06	4.58E-03	+1
Luteinizing hormone	Luteinizing hormone	2.55E-06	4.58E-03	-1
CRBS	Beta-crystallin S	1.01E-05	8.37E-03	-1
NCAM-120	Neural cell adhesion molecule 1; 120 kDa isoform	8.78E-06	8.37E-03	-1
SIA10	Type 2 lactosamine alpha-2;3-sialyltransferase	7.36E-06	8.37E-03	-1
LRP1	Prolow-density lipoprotein receptor-related protein 1	2.77E-05	1.16E-02	-1
CLM2	CMRF35-like molecule 2	2.70E-05	1.16E-02	-1
THS7A	Thrombospondin type-1 domain-containing protein 7A	1.65E-05	1.16E-02	-1
LAP	Cytosol aminopeptidase	3.04E-05	1.16E-02	-1
CXCL16	C-X-C motif chemokine 16	2.64E-05	1.16E-02	+1
FSH	Follicle stimulating hormone	2.52E-05	1.16E-02	-1
VEGF sR2	Vascular endothelial growth factor receptor 2	3.27E-05	1.16E-02	-1
ASGR1	Asialoglycoprotein receptor 1	2.97E-05	1.16E-02	+1

682
683
684

685 **Table 6:** Summary of the characteristics of each endophenotype.
686 In this table, High (Low), denoted as H (L) implies that the average value of the variable in the
687 corresponding EP was significantly higher (lower) than the other EPs (considered together), while
688 N (Nondescript) implies that it was not significantly different. The table is formed based on results
689 described in Tables S1, S6 and S2.

Endophenotype	Age	Sex at birth	BMI	Metabolic sub-pathways	Blood markers
EP1	H	F	N	Leucine, Isoleucine and Valine Metabolism	High lymphocyte, Low neutrophil
EP2	N	N	L	N	High albumin, Low white blood cells
EP3	L	M	N	Phospholipid Metabolism	High platelet, Low creatinine
EP4	N	N	N	Ascorbate and Aldarate Metabolism	High lactate deshydrogenase
EP5	H	N	N	Benzoate Metabolism	High creatinine, Low haemoglobin
EP6	N	M	N	Methionine, Cysteine, SAM and Taurine Metabolism	High white blood cells, Low lymphocyte

690 Abbreviations used: H = High, L = Low, N = Nondescript, F = Female, M = Male
691

692 **References**

693 1. Sanders JM, Monogue ML, Jodlowski TZ, and Cutrell JB. Pharmacologic Treatments for
694 Coronavirus Disease 2019 (COVID-19): A Review. *JAMA.* 2020;323(18):1824-36.

695 2. Lim ZJ, Subramaniam A, Ponnapa Reddy M, Blecher G, Kadam U, Afroz A, et al. Case
696 Fatality Rates for Patients with COVID-19 Requiring Invasive Mechanical Ventilation. A
697 Meta-analysis. *Am J Respir Crit Care Med.* 2021;203(1):54-66.

698 3. Hull DL. *PSA: Proceedings of the biennial meeting of the philosophy of science association*.
699 Cambridge University Press; 1974:653-70.

700 4. Te Pas MFW, Madsen O, Calus MPL, and Smits MA. The Importance of Endophenotypes
701 to Evaluate the Relationship between Genotype and External Phenotype. *International*
702 *Journal of Molecular Sciences.* 2017;18(2):472.

703 5. Blatti III C, Emad A, Berry MJ, Gatzke L, Epstein M, Lanier D, et al. Knowledge-guided
704 analysis of "omics" data using the KnowEnG cloud platform. *PLoS biology.*
705 2020;18(1):e3000583.

706 6. Emad A, Ray T, Jensen TW, Parat M, Natrajan R, Sinha S, et al. Superior breast cancer
707 metastasis risk stratification using an epithelial-mesenchymal-amoeboid transition gene
708 signature. *Breast Cancer Research.* 2020;22(1):74.

709 7. Al-Hadrawi DS, Al-Rubaye HT, Almulla AF, Al-Hakeim HK, and Maes M. Lowered oxygen
710 saturation and increased body temperature in acute COVID-19 largely predict chronic
711 fatigue syndrome and affective symptoms due to Long COVID: A precision nomothetic
712 approach. *Acta Neuropsychiatr.* 2022:1-12.

713 8. Tremblay K, Rousseau S, Zawati MnH, Auld D, Chassé M, Coderre D, et al. The Biobanque
714 québécoise de la COVID-19 (BQC19)—A cohort to prospectively study the clinical and
715 biological determinants of COVID-19 clinical trajectories. *PloS one.* 2021;16(5):e0245031.

716 9. Gold L, Ayers D, Bertino J, Bock C, Bock A, Brody E, et al. Aptamer-based multiplexed
717 proteomic technology for biomarker discovery. *Nature Precedings.* 2010:1-.

718 10. Ward Jr JH. Hierarchical grouping to optimize an objective function. *Journal of the*
719 *American statistical association.* 1963;58(301):236-44.

720 11. Murtagh F, and Legendre P. Ward's Hierarchical Agglomerative Clustering Method: Which
721 Algorithms Implement Ward's Criterion? *Journal of Classification.* 2014;31(3):274-95.

722 12. Kaplan EL, and Meier P. Nonparametric Estimation from Incomplete Observations. *Journal*
723 *of the American Statistical Association.* 1958;53(282):457-81.

724 13. Gillespie M, Jassal B, Stephan R, Milacic M, Rothfels K, Senff-Ribeiro A, et al. The reactome
725 pathway knowledgebase 2022. *Nucleic Acids Research.* 2021;50(D1):D687-D92.

726 14. Blachier F, Andriamihaja M, and Blais A. Sulfur-Containing Amino Acids and Lipid
727 Metabolism. *The Journal of Nutrition.* 2020;150(Supplement_1):2524S-31S.

728 15. Ye C, Sutter BM, Wang Y, Kuang Z, and Tu BP. A metabolic function for phospholipid and
729 histone methylation. *Molecular cell.* 2017;66(2):180-93. e8.

730 16. Marfia G, Navone S, Guarnaccia L, Campanella R, Mondoni M, Locatelli M, et al. Decreased
731 serum level of sphingosine-1-phosphate: a novel predictor of clinical severity in COVID-
732 19. *EMBO molecular medicine.* 2021;13(1):e13424.

733 17. Chen PY, Gao WY, Liou JW, Lin CY, Wu MJ, and Yen JH. Angiopoietin-Like Protein 3
734 (ANGPTL3) Modulates Lipoprotein Metabolism and Dyslipidemia. *Int J Mol Sci.*
735 2021;22(14).

736 18. Lachkar F, Ferré P, Foufelle F, and Papaioannou A. Dihydroceramides: their emerging
737 physiological roles and functions in cancer and metabolic diseases. *American Journal of*
738 *Physiology-Endocrinology and Metabolism.* 2021;320(1):E122-E30.

739 19. Lone MA, Hülsmeier AJ, Saied EM, Karsai G, Arenz C, von Eckardstein A, et al. Subunit
740 composition of the mammalian serine-palmitoyltransferase defines the spectrum of
741 straight and methyl-branched long-chain bases. *Proceedings of the National Academy of*
742 *Sciences.* 2020;117(27):15591-8.

743 20. Han G, Gupta SD, Gable K, Niranjanakumari S, Moitra P, Eichler F, et al. Identification of
744 small subunits of mammalian serine palmitoyltransferase that confer distinct acyl-CoA
745 substrate specificities. *Proceedings of the National Academy of Sciences.*
746 2009;106(20):8186-91.

747 21. Helms J, Tacquard C, Severac F, Leonard-Lorant I, Ohana M, Delabranche X, et al. High risk
748 of thrombosis in patients with severe SARS-CoV-2 infection: a multicenter prospective
749 cohort study. *Intensive Care Medicine.* 2020;46(6):1089-98.

750 22. Asakura H, and Ogawa H. COVID-19-associated coagulopathy and disseminated
751 intravascular coagulation. *International Journal of Hematology.* 2021;113(1):45-57.

752 23. Fahmy OH, Daas FM, Salunkhe V, Petrey JL, Cosar EF, Ramirez J, et al. Is Microthrombosis
753 the Main Pathology in Coronavirus Disease 2019 Severity?-A Systematic Review of the
754 Postmortem Pathologic Findings. *Crit Care Explor.* 2021;3(5):e0427.

755 24. Barnes BJ, Adrover JM, Baxter-Stoltzfus A, Borczuk A, Cools-Lartigue J, Crawford JM, et al.
756 Targeting potential drivers of COVID-19: Neutrophil extracellular traps. *Journal of*
757 *Experimental Medicine.* 2020;217(6).

758 25. Bonaventura A, Vecchié A, Dagna L, Martinod K, Dixon DL, Van Tassell BW, et al.
759 Endothelial dysfunction and immunothrombosis as key pathogenic mechanisms in COVID-
760 19. *Nature Reviews Immunology.* 2021;21(5):319-29.

761 26. Ding J, Hostallero DE, El Khili MR, Fonseca GJ, Milette S, Noorah N, et al. A network-
762 informed analysis of SARS-CoV-2 and hemophagocytic lymphohistiocytosis genes'
763 interactions points to Neutrophil extracellular traps as mediators of thrombosis in COVID-
764 19. *PLoS Computational Biology.* 2021;17(3):e1008810.

765 27. Blasco A, Coronado M-J, Hernández-Terciado F, Martín P, Royuela A, Ramil E, et al.
766 Assessment of neutrophil extracellular traps in coronary thrombus of a case series of
767 patients with COVID-19 and myocardial infarction. *JAMA cardiology.* 2021;6(4):469-74.

768 28. Desilles JP, Solo Nomenjanahary M, Consoli A, Ollivier V, Faille D, Bourrienne MC, et al.
769 Impact of COVID-19 on thrombus composition and response to thrombolysis: Insights
770 from a monocentric cohort population of COVID-19 patients with acute ischemic stroke.
771 *Journal of Thrombosis and Haemostasis.* 2022;20(4):919-28.

772 29. Englert H, Rangaswamy C, Deppermann C, Sperhake J-P, Krisp C, Schreier D, et al.
773 Defective NET clearance contributes to sustained FXII activation in COVID-19-associated
774 pulmonary thrombo-inflammation. *EBioMedicine.* 2021;67:103382.

775 30. Leppkes M, Knopf J, Naschberger E, Lindemann A, Singh J, Herrmann I, et al. Vascular
776 occlusion by neutrophil extracellular traps in COVID-19. *EBioMedicine.* 2020;58:102925.

777 31. Middleton EA, He X-Y, Denorme F, Campbell RA, Ng D, Salvatore SP, et al. Neutrophil
778 extracellular traps contribute to immunothrombosis in COVID-19 acute respiratory
779 distress syndrome. *Blood.* 2020;136(10):1169-79.

780 32. Obermayer A, Jakob L-M, Haslbauer JD, Matter MS, Tzankov A, and Stoiber W. Neutrophil
781 extracellular traps in fatal COVID-19-associated lung injury. *Disease markers.* 2021;2021.

782 33. Ouwendijk WJ, Raadsen MP, Van Kampen JJ, Verdijk RM, Von Der Thusen JH, Guo L, et al.
783 High levels of neutrophil extracellular traps persist in the lower respiratory tract of
784 critically ill patients with coronavirus disease 2019. *The Journal of infectious diseases.*
785 2021;223(9):1512-21.

786 34. Petito E, Falcinelli E, Paliani U, Cesari E, Vaudo G, Sebastiano M, et al. Association of
787 neutrophil activation, more than platelet activation, with thrombotic complications in
788 coronavirus disease 2019. *The Journal of infectious diseases.* 2021;223(6):933-44.

789 35. Skendros P, Mitsios A, Chrysanthopoulou A, Mastellos DC, Metallidis S, Rafailidis P, et al.
790 Complement and tissue factor–enriched neutrophil extracellular traps are key drivers in
791 COVID-19 immunothrombosis. *The Journal of clinical investigation.* 2020;130(11):6151-7.

792 36. Massberg S, Grahl L, von Bruehl M-L, Manukyan D, Pfeiler S, Goosmann C, et al. Reciprocal
793 coupling of coagulation and innate immunity via neutrophil serine proteases. *Nature*
794 *Medicine.* 2010;16(8):887-96.

795 37. Papayannopoulos V, Metzler KD, Hakkim A, and Zychlinsky A. Neutrophil elastase and
796 myeloperoxidase regulate the formation of neutrophil extracellular traps. *Journal of Cell*
797 *Biology.* 2010;191(3):677-91.

798 38. Welin A, Amirbeagi F, Christenson K, Björkman L, Björnsdottir H, Forsman H, et al. The
799 Human Neutrophil Subsets Defined by the Presence or Absence of OLFM4 Both
800 Transmigrate into Tissue In Vivo and Give Rise to Distinct NETs In Vitro. *PLOS ONE.*
801 2013;8(7):e69575.

802 39. Alder MN, Opoka AM, Lahni P, Hildeman DA, and Wong HR. Olfactomedin-4 Is a Candidate
803 Marker for a Pathogenic Neutrophil Subset in Septic Shock. *Critical Care Medicine.*
804 2017;45(4):e426-e32.

805 40. Parman T, Bunin DI, Ng HH, McDunn JE, Wulff JE, Wang A, et al. Toxicogenomics and
806 metabolomics of pentamethylchromanol (PMCol)-induced hepatotoxicity. *Toxicological*
807 *Sciences.* 2011;124(2):487-501.

808 41. Stafford JH, and Thorpe PE. Increased exposure of phosphatidylethanolamine on the
809 surface of tumor vascular endothelium. *Neoplasia.* 2011;13(4):299-IN2.

810 42. Ran S, Downes A, and Thorpe PE. Increased exposure of anionic phospholipids on the
811 surface of tumor blood vessels. *Cancer research.* 2002;62(21):6132-40.

812 43. Port JR, Adney DR, Schwarz B, Schulz JE, Sturdevant DE, Smith BJ, et al. High-fat high-sugar
813 diet-induced changes in the lipid metabolism are associated with mildly increased COVID-
814 19 severity and delayed recovery in the Syrian hamster. *Viruses.* 2021;13(12):2506.

815 44. MacDonald J, and Sprecher H. Distribution of arachidonic acid in choline-and
816 ethanolamine-containing phosphoglycerides in subfractionated human neutrophils.
817 *Journal of Biological Chemistry.* 1989;264(30):17718-26.

818 45. Papageorgiou C, Jourdi G, Adjambri E, Walborn A, Patel P, Fareed J, et al. Disseminated
819 intravascular coagulation: an update on pathogenesis, diagnosis, and therapeutic
820 strategies. *Clinical and Applied Thrombosis/Hemostasis.* 2018;24(9_suppl):8S-28S.

821 46. Goligher EC, Bradbury CA, McVerry BJ, Lawler PR, Berger JS, Gong MN, et al. Therapeutic
822 Anticoagulation with Heparin in Critically Ill Patients with Covid-19. *N Engl J Med.*
823 2021;385(9):777-89.

824 47. Lawler PR, Goligher EC, Berger JS, Neal MD, McVerry BJ, Nicolau JC, et al. Therapeutic
825 Anticoagulation with Heparin in Noncritically Ill Patients with Covid-19. *N Engl J Med.*
826 2021;385(9):790-802.

827 48. Thomas T, Stefanoni D, Reisz JA, Nemkov T, Bertolone L, Francis RO, et al. COVID-19
828 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and
829 renal status. *JCI insight.* 2020;5(14).

830 49. Zahedi K, Barone S, Kramer DL, Amlal H, Alhonen L, Jänne J, et al. The role of
831 spermidine/spermine N 1-acetyltransferase in endotoxin-induced acute kidney injury.
832 *American Journal of Physiology-Cell Physiology.* 2010;299(1):C164-C74.

833 50. Pirnes-Karhu S, Jantunen E, Mäntymaa P, Mustjoki S, Alhonen L, and Uimari A.
834 Spermidine/spermine N1-acetyltransferase activity associates with white blood cell count
835 in myeloid leukemias. *Experimental Hematology.* 2014;42(7):574-80.

836 51. Smadja DM, Philippe A, Bory O, Gendron N, Beauvais A, Gruest M, et al. Placental growth
837 factor level in plasma predicts COVID-19 severity and in-hospital mortality. *Journal of*
838 *Thrombosis and Haemostasis.* 2021;19(7):1823-30.

839 52. Bussolari C, Palumbo D, Fominsky E, Nardelli P, De Lorenzo R, Vitali G, et al. Case Report:
840 Nintedaninb May Accelerate Lung Recovery in Critical Coronavirus Disease 2019. *Frontiers*
841 *in Medicine.* 2021;8:766486.

842 53. Liu W, Peng Y, and Tobin DJ. A new 12-gene diagnostic biomarker signature of melanoma
843 revealed by integrated microarray analysis. *PeerJ.* 2013;1:e49.

844 54. Luo S, Feofanova EV, Tin A, Tung S, Rhee EP, Coresh J, et al. Genome-wide association
845 study of serum metabolites in the African American Study of Kidney Disease and
846 Hypertension. *Kidney International.* 2021;100(2):430-9.

847 55. Agarwal B, Gatt A, Riddell A, Wright G, Chowdary P, Jalan R, et al. Hemostasis in patients
848 with acute kidney injury secondary to acute liver failure. *Kidney International.*
849 2013;84(1):158-63.

850 56. Williams R, Schalm SW, and O'Grady JG. Acute liver failure: redefining the syndromes. *The*
851 *Lancet.* 1993;342(8866):273-5.

852 57. Koishi R, Ando Y, Ono M, Shimamura M, Yasumo H, Fujiwara T, et al. Angptl3 regulates
853 lipid metabolism in mice. *Nature Genetics.* 2002;30(2):151-7.

854 58. Pisciotta L, Favari E, Magnolo L, Simonelli S, Adorni MP, Sallo R, et al. Characterization of
855 Three Kindreds With Familial Combined Hypolipidemia Caused by Loss-of-Function
856 Mutations of ANGPTL3. *Circulation: Cardiovascular Genetics.* 2012;5(1):42-50.

857 59. Lekka ME, Liokatis S, Nathanail C, Galani V, and Nakos G. The Impact of Intravenous Fat
858 Emulsion Administration in Acute Lung Injury. *American Journal of Respiratory and Critical*
859 *Care Medicine.* 2004;169(5):638-44.

860 60. Ahmad Z, Banerjee P, Hamon S, Chan K-C, Bouzelmat A, Sasiela WJ, et al. Inhibition of
861 Angiopoietin-Like Protein 3 With a Monoclonal Antibody Reduces Triglycerides in
862 Hypertriglyceridemia. *Circulation.* 2019;140(6):470-86.

863 61. Bie H, Yin J, He X, Kermode AR, Goddard-Borger ED, Withers SG, et al. Insights into
864 mucopolysaccharidosis I from the structure and action of α-L-iduronidase. *Nature*
865 *Chemical Biology.* 2013;9(11):739-45.

866 62. Ford L, Kennedy AD, Goodman KD, Pappan KL, Evans AM, Miller LAD, et al. Precision of a
867 Clinical Metabolomics Profiling Platform for Use in the Identification of Inborn Errors of
868 Metabolism. *The Journal of Applied Laboratory Medicine.* 2020;5(2):342-56.

869 63. Durinck S, Spellman PT, Birney E, and Huber W. Mapping identifiers for the integration of
870 genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols.*
871 2009;4(8):1184-91.

872 64. R Core Team. Vienna, Austria: R Foundation for Statistical Computing; 2022.
873 https://www.R-project.org/.

874 65. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on
875 conducting genome-wide association studies: Quality control and statistical analysis. *Int J*
876 *Methods Psychiatr Res.* 2018;27(2):e1608.

877 66. Ochoa A. genio: Genetics Input/Output Functions. 2022:R package version 1.1.1.
878 https://CRAN.R-project.org/package=genio.

879 67. Warnes G, Gorjanc G, Leisch F, and Man M. genetics: Population Genetics. 2021:R package
880 version 1.3.8.1.3. https://CRAN.R-project.org/package=genetics.

881 68. Zheng X, Levine D, Shen J, Gogarten SM, Laurie C, and Weir BS. A high-performance
882 computing toolset for relatedness and principal component analysis of SNP data.
883 *Bioinformatics.* 2012;28(24):3326-8.

884 69. Perdry H, and Dandine-Roulland C. gaston: Genetic Data Handling (QC, GRM, LD, PCA) &
885 Linear Mixed Models. 2020:R package version 1.5.7. https://CRAN.R-
886 project.org/package=gaston.

887 70. Satopaa V, Albrecht J, Irwin D, and Raghavan B. *2011 31st international conference on*
888 *distributed computing systems workshops*. IEEE; 2011:166-71.
889