Integrative Metabolomic and Proteomic Signatures Define Clinical Outcomes in Severe COVID-19

Sergio Alvarez-Mulett[1]*, Mustafa Buyukozkan[2,3]*, Alexandra C. Racanelli[1]*, Frank Schmidt[4], Richa Batra[2,3], Katherine L. Hoffman[5], Hina Sarwath[4], Rudolf Engelke[4], Luis Gomez-Escobar[1], Will Simmons[5], Elisa Benedetti[2,3], Kelsey Chetnik[2,3], Edward Schenck[1], Karsten Suhre[4], Justin J. Choi[6], Zhen Zhao[7], Sabrina Racine-Brzostek[7], He S. Yang[7], Mary E. Choi[8], Augustine M.K. Choi[1], Soo Jung Cho[1#], Jan Krumsiek[2,3#]

[1]Department of Medicine, Division of Pulmonary and Critical Care Medicine, Weill Cornell Medicine, New York, NY, USA

[2]Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA

[3]Meyer Cancer Center and Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY, USA

[4]Department of Biochemistry, Weill Cornell Medicine – Qatar, Qatar Foundation, Doha, Qatar

[5]Department of Population Health Sciences, Division of Biostatistics, Weill Cornell Medicine, New York, NY, USA

[6]Department of Medicine, Division of General Internal Medicine, Weill Cornell Medicine, New York, NY, USA

[7]Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY, USA

[8]Division of Nephrology and Hypertension, Joan and Sanford I. Weill Department of Medicine, New York, NY, USA

* these authors contributed equally to this work

[#] corresponding authors

Correspondence to:

Jan Krumsiek, jak2042@med.cornell.edu

Soo Jung Cho, sjc9006@med.cornell.edu

Abstract

The novel coronavirus disease-19 (COVID-19) pandemic caused by SARS-CoV-2 has ravaged global healthcare with previously unseen levels of morbidity and mortality. To date, methods to predict the clinical course, which ranges from the asymptomatic carrier to the critically ill patient in devastating multi-system organ failure, have yet to be identified. In this study, we performed large-scale integrative multi-omics analyses of serum obtained from COVID-19 patients with the goal of uncovering novel pathogenic complexities of this disease and identifying molecular signatures that predict clinical outcomes. We assembled a novel network of protein-metabolite interactions in COVID-19 patients through targeted metabolomic and proteomic profiling of serum samples in 330 COVID-19 patients compared to 97 non-COVID, hospitalized controls. Our network identified distinct protein-metabolite cross talk related to immune modulation, energy and nucleotide metabolism, vascular homeostasis, and collagen catabolism. Additionally, our data linked multiple proteins and metabolites to clinical indices associated with long-term mortality and morbidity, such as acute kidney injury. Finally, we developed a novel composite outcome measure for COVID-19 disease severity and created a clinical prediction model using a set of 33 metabolites. The model significantly improved the identification of key events of critical illness such as prolonged hospitalization, supplemental oxygen requirement, acute kidney injury (AKI), and acute respiratory distress syndrome (ARDS), beyond those achieved using the more traditional risk factors of age, gender, and BMI.

Introduction

The novel coronavirus disease 2019 (COVID-19) has a broad spectrum of clinical features that range from asymptomatic disease to acute respiratory distress syndrome (ARDS)(1, 2). COVID-19 ARDS can lead to refractory hypoxia, mechanical ventilation, prolonged intensive care unit (ICU) stay and increased mortality(3). Previous studies have shown a high incidence of concomitant organ failure in COVID-19, including acute kidney injury (AKI)(4), acute liver injury(5), thromboembolic events(6, 7) and secondary infections contributing to a fatal outcome(8).

Massive investigative efforts by multiple scientific groups have used proteomic and metabolomic approaches to begin to unravel disease mechanisms relevant to SARS-CoV-2 infection such as inflammation, coagulation, and metabolism(9). However, how COVID-19 specific protein-metabolite interactions relate to the severity of disease and clinical outcomes remains poorly understood. Key study limitations have included relatively small sample sizes, absence of protein-metabolite network analysis and the focus on dichotomous outcome measures such as death and survival. These limitations have been difficult to overcome and restrict our understanding of COVID-19 pathogenesis.

Here, we report the largest study to integrate metabolomic and proteomic analyses of serum samples obtained from hospitalized COVID-19 patients during SARS-CoV-2 infection compared to controls during the same time period as controls (Figure 1). Through this work, we uncovered COVID-19-specific metabolite and protein profiles, identified novel protein-metabolite modules, and defined the molecular signatures of several clinical indices (CRP, ferritin, platelet count, AKI, and death). Additionally, we developed the first clinical composite outcome prediction model in COVID-19, where the input of discrete metabolic profiles significantly improves our clinical insight into a broad range of outcomes known to plague many survivors of severe COVID-19.

Results

Study cohort and dataset

Our cohort was comprised of 330 patients with confirmed SARS-CoV-2 RT-PCR, and 97 non-COVID-19 controls with negative RT-PCR results who were hospitalized at the NewYork-Presbyterian Hospital/Weill Cornell Medical Center between March and April 2020. Serum samples were obtained within the first 3 days of admission. The majority of COVID-19 patients had samples drawn at two or three different time points resulting in a total of 582 serum samples from the 330 COVID-19 patients, while all 97 controls had one sample drawn. A detailed description of the clinical and demographic characteristics of the cohort can be found in Table 1 and Supplementary Table 1. Of note, we excluded samples collected after intubation because we found that the clinical act of intubation significantly alters a patient's metabolic profile (Supplementary Table 2).

92 Metabolic profiles were assessed for all samples using liquid chromatography coupled
93 with mass spectrometry (LC/MS). After quality control and data preprocessing, 125
94 metabolites were available for comparative analysis. Proteomic profiling was performed
95 on a subset of 227 samples (173 from COVID-19 patients and 54 controls) using the Olink
96 inflammation, cardiovascular II and cardiovascular III panels, which cover 266 unique
97 protein biomarkers.

98

99 **Metabolomic and proteomic changes associated with COVID-19**

100 Differential metabolomic analysis identified significant changes in abundances of 70 out
101 of the 125 analyzed metabolites between COVID-19 and controls at a false discovery rate
102 (FDR) of 0.05 (Figure 2A). The top three differentially expressed metabolites were
103 involved in amino acid metabolism: N-acetyl-L-aspartic acid (p-value = 9.19E-18), N-
104 acetyl-aspartyl-glutamic acid (p-value = 5.30E-15), and argininosuccinic acid (p-value =
105 6.35E-12) (Figure 2B). KEGG pathway mapping of the differentially expressed
106 metabolites revealed an involvement of various metabolic pathways (Figure 2C). These
107 pathways included arginine and proline metabolism, glycine and serine metabolism,
108 alanine metabolism, methionine metabolism, sphingolipid metabolism, gluconeogenesis,
109 and the TCA cycle pathway, demonstrating involvement of the broader categories of
110 amino acid, lipid and energy metabolism in COVID-19 pathogenesis.

111 Comparative proteomic analysis identified significant changes in the expression of 48 out
112 of the 266 analyzed proteins between COVID-19 and controls at an FDR of 0.05
113 (Figure 2D). The top three differentially expressed proteins were C-X-C motif chemokine
114 ligand 10 (CXCL10) (p-value = 5.09E-13), galectin 9 (Gal-9) (p-value = 5.49E-10), and
115 monocyte chemoattractant protein 3 (MCP-3) (p-value = 2.85E-09) (Figure 2E). KEGG
116 pathway mapping revealed that these proteins participated in various protein pathways
117 (Figure 2F) including the interleukin 17 (IL-17), tumoral necrosis factor (TNF) and JAK-
118 STAT signaling pathways. Detailed results of the differential analysis and pathway
119 mappings can be found in Supplementary Table 3.

120

121 **Protein-metabolite networks identify potential mediators of COVID-19 pathology**
122 To obtain further mechanistic insight into the biology of COVID-19, we developed a
123 comprehensive, data-driven network for the integrative analysis of our multi-omics
124 dataset. We first generated a Gaussian graphical model (GGM) of correlated metabolites
125 and proteins from our COVID-19 cohort (Figure 3A, Supplementary Data 1). GGMs are
126 correlation-based network models that we have previously demonstrated to accurately
127 reconstruct biological pathways from blood-based omics data(10-12). A minimum
128 spanning tree-based algorithm was then used to identify a focused subnetwork that
129 connects the most significantly correlated metabolites and proteins from the original
130 network (Figure 3B). The subnetwork included 13 proteins from the Olink inflammatory
131 panel, 32 proteins from the Olink cardiovascular panels II and III, and 70 metabolites.
132 From this subnetwork, we selected 4 network modules to query the interplay between
133 metabolism, inflammation, and vascular dysfunction in COVID-19 patients.

134 *Inflammation-related network modules. Module 1.* Our network identified
135 hexanoylcarnitine as a key metabolite associated with inflammatory cytokines in COVID-
136 19 illness. The module contained IFN-gamma, CXCL10 and CXCL11 that were all
137 upregulated in COVID-19 and associated with hexanoylcarnitine. IFN-gamma, CXCL10
138 and CXCL11 are proinflammatory cytokines which regulate T cell immunity(13), while
139 hexanoylcarnitine is a medium-chain fatty acid conjugate that plays a critical role in
140 energy metabolism and mitochondrial fatty acid β-oxidation (Figure 3B)(14). Levels of
141 other carnitine species (L-carnitine and caprylic acid) were also elevated in COVID-19
142 patients, suggesting a role for inflammation, dysregulated fatty acid β-oxidation and
143 mitochondrial dysfunction in COVID-19 pathogenesis.

144 *Module 2.* In a second inflammation-related module, cytosine was another key metabolite
145 linked with inflammatory cytokines during COVID-19. The module consisted of a group of
146 macrophage-derived cytokines, including MCP-2, MCP-3 and GRN, which were all
147 upregulated in COVID-19 and positively associated with cytosine (Figure 3B). This finding
148 is consistent with the hyperinflammatory state of COVID-19 infection in which cytokine
149 storm is often observed(15). Cytosine, a pyrimidine-class nucleotide that is an essential
150 metabolite for cell proliferation and survival, is commonly upregulated in the host

151 response during viral infection and is furthermore an important mediator of viral
152 replication(16). Prior reports have shown cytosine levels to be elevated in COVID-19
153 patients(17), which was corroborated in our own cohort. Taken together, cytosine may be
154 a key metabolite linking viral replication to SARS-CoV-2 induced inflammation.

155 *Vascular-related network modules. Module 3.* While COVID-19 presents itself mainly as
156 a respiratory disease, autopsy reports have additionally described significant vascular
157 injury due to endothelial cell damage, microcirculatory thrombi, and impaired cellular
158 junction integrity(18). Our network uncovered the coordination of MERTK, RAGE and
159 thrombomodulin (TM) with the metabolite 4-hydroxyproline in COVID-19 (Figure 3B).
160 MERTK, RAGE and TM are proteins key to maintaining features of endothelial
161 homeostasis, and hydroxyproline is a major component of collagen(19). MERTK, a
162 member of the TAM family of receptor tyrosine kinases (RTKs), and RAGE, a pro-
163 coagulant and inflammatory molecule were upregulated, while TM was downregulated,
164 suggesting induction of a pro-coagulant and inflammatory vascular state in COVID-19.
165 Notably, the coordination of these proteins centered around alterations in levels of
166 hydroxyproline, where downregulation of TM directly correlated with hydroxyproline
167 levels. A link between TM and hydroxyproline has been described, where administration
168 of TM had anti-fibrotic effects on lung and kidney murine models(20). Taken together,
169 these data link the pro-coagulant and fibrotic state of COVID-19 through thrombomodulin
170 and hydroxyproline.

171 *Module 4.* In a second vascular-related module, we found that cathepsin D (CTSD), a
172 lysosomal protease known to disrupt endothelial cell junctions and increase
173 vasopermeability, was associated with a group of glycolytic metabolites including
174 D-glucose, glycerol-3-phosphate, and lactic acid(21). Notably, increased vascular
175 permeability and glycolysis are known features of a pro-angiogenic state(22). While
176 dysregulated angiogenesis occurs in COVID-19 patients(23), how these events are
177 coordinated remains poorly understood. Here, we report the concomitant up-regulation of
178 cathepsin D and intermediate products of glycolysis (D-glucose and lactic acid) in our
179 COVID-19 cohort, which position this enzyme as a potential driver of the COVID-19
180 phenotype. Of note, elevated plasma activity of cathepsin D has been found in patients

181 with type 2 diabetes, suggested a link between abnormal vasculature and the
182 dysregulated glucose metabolism seen in our higher risk diabetic COVID-19 patients(24).
183 Taken together these data suggest that the disruption of vascular and glucometabolic
184 homeostasis in COVID-19 is mediated by cathepsin D.

185

186 **Serum metabolites and proteins associate with clinical indices in COVID-19**

187 Serum metabolites and proteins within the COVID-19 patient group were assessed for
188 correlation with relevant clinical indices including: i) demographics (sex, age, BMI), ii)
189 concurrent comorbidity (hypertension, pre-existing kidney disease, diabetes mellitus
190 [DM], severity of illness [SOFA]), iii) laboratory markers of inflammation (C-reactive
191 protein [CRP], d-dimer, lymphocyte count, platelet count, ferritin), and iv) future clinical
192 events (ARDS, death) (Figure 4A and 4D). Pre-existing kidney disease was found to
193 associate with the highest number of both metabolites and proteins. Other clinical
194 parameters including ARDS, death, BMI, gender, age, SOFA score, hypertension, DM,
195 and d-dimer levels were associated with numerous metabolites but only few proteins.
196 Conversely, platelet count and ferritin level were associated with a large number of
197 proteins but relatively few metabolites. Detailed association results are provided in
198 Supplementary Tables 4 and 5.

199 Hierarchical clustering of clinical indices by their correlation with metabolites revealed two
200 broad clusters: one predominantly related to laboratory markers of inflammation, and a
201 second largely related to non-laboratory indices (Figure 4B). Volcano plots of three
202 representative indices (death, pre-existing kidney disease and CRP) are shown in
203 Figure 4C. Seventy-five metabolites were significantly associated with death, 96 with pre-
204 existing kidney disease and 37 with CRP level. 69 metabolites were associated with both
205 death and kidney disease and 20 were associated with death, kidney disease and CRP
206 levels. Key metabolites associated with each of the three selected clinical indices are
207 depicted in Figure 4C. Remarkably, hexanoylcarnitine and cytosine, which we earlier
208 showed to be upregulated in COVID-19, were among the 20 metabolites associated with
209 all three of these clinical indices. This finding supports the potential utility of

210 hexanoylcarnitine and cytosine not only as biomarkers for COVID-19 but also as
211 predictors of disease severity.

212 Clustering of clinical indices according to their correlation with proteins also revealed two
213 broad clusters (Figure 4E) that were notably similar to the metabolite-derived clusters,
214 except that age and kidney disease clustered together with laboratory markers of
215 inflammation. Volcano plots of three representative indices (death, ferritin level and
216 platelet count) are depicted in Figure 4F. Seventeen proteins correlated with death, 26
217 with platelet count and 29 with ferritin. Six proteins were associated with both death and
218 ferritin level, 9 with death and platelet count, and 5 (ANGPT1, PDGF subunit A, PDGF
219 subunit B, LAP TGF-beta-1, CXCL5) with death, ferritin level and platelet count
220 (Figure 4F). ANGPT1, PDGF subunit A and PDGF subunit B are markers of endothelial
221 injury and platelet dysfunction while LAP, TGF-beta-1 and CXCL5 are markers of
222 inflammation. The fact that these proteins correlated with known clinical indices of
223 COVID-19 mortality and morbidity corroborates the importance of vascular injury and
224 inflammation in COVID-19 pathogenesis, which we observed in our network analysis
225 above.

226

227 **Metabolomics signature predicts clinical outcomes in COVID-19.**

228 We devised a novel hierarchical composite outcome measure for COVID-19 severity
229 which incorporates a series of clinical events, ranked in order of severity, that characterize
230 both acute COVID-19 and some of the sequelae of COVID-19 seen in post-acute COVID-
231 19 syndrome (PACS)(25): in-hospital mortality, mechanical ventilation (MV) at discharge,
232 kidney replacement therapy (KRT) at discharge, prolonged organ failure support
233 (mechanical ventilation and/or kidney replacement therapy for more than 2 weeks),
234 supplemental oxygen requirement, acute kidney injury and length of hospital stay
235 (Figure 5A, Supplementary Figure 1). Our measure represents each patient on a
236 continuous spectrum of severity to provide more information than a dichotomous
237 classification such as mortality.

238 A machine learning algorithm based on an ordinal response mixed effect model with
239 LASSO regularization and cross-validation was used to generate our prediction model of

240 the composite outcome measure using serum metabolic profiles and baseline patient
241 demographics (age, sex and BMI) as combined inputs (Figure 5B). Proteins were not
242 included in the model due to insufficient sample size. The final model included 33
243 metabolites and achieved an average concordance-index of d=0.69 (SE=0.017, 95% CI
244 of [0.65, 0.72]), which is equivalent to a receiver operating characteristics area under the
245 curve (ROC-AUC). This was a significant improvement (p-value = 1E-5) over a baseline
246 model containing only age, sex, and BMI, which had a performance of d=0.59 (SD=0.02,
247 95% CI of [0.56, 0.64]). The 33 metabolites in the final model included 14 amino acids, 6
248 nucleotides, 6 lipids and 2 metabolites related to energy metabolism (Figure 5C).
249 Interestingly, our metabolite-based model showed improvement over the baseline model
250 not only for predicting the composite outcome, but also for predicting some of its individual
251 components (i.e., intubation, AKI, and length of hospital stay, Figure 5D).

252 We assessed the tradeoff between the number of metabolites incorporated into the model
253 and the model's prediction performance across a range of included metabolites. Our
254 analysis revealed that most of the predictive power (d ~ 0.68) was already achieved with
255 the first 6 metabolites, after which prediction performance did not significantly improve
256 (Supplementary Figure 2). This finding suggests that a targeted assay of a small number
257 of metabolites could predict disease severity with this accuracy, although independent
258 validation in other datasets remains to be established.

259 We compared our results with a model reported previously (26), which represents the
260 closest analysis to the one we have performed here and reported a remarkably higher
261 AUC of 0.95. However, several aspects prevent a direct comparison of theirs and our
262 performance metrics. First, our study had a substantially larger sample size with 296
263 patients and 487 plasma samples, compared to 31 for training and 29 for testing in Shen
264 et al. Second, our reported AUC was based on cross-validation and thus less likely to
265 overestimate the model performance, while Shen. et al only reported a training set-based
266 AUC. Third, our own reanalysis of the published data by Shen et al. shows that
267 metabolites in their model do not improve predictive performance beyond age alone
268 (Supplementary Figure 3), while we demonstrated a clear improvement beyond baseline.
269 Fourth, our model defines an ordinal classification of severity as opposed to prior studies

that have dichotomized outcomes into severe and non-severe cases which is known to be statistically less powerful and can lead to bias(27, 28). Taken together, we believe that our analysis provides a higher powered and a more holistic assessment of the added value of metabolites to predict disease severity as well as the effect on overall quality of life rather than just predicting survival.

Conclusion

The novel coronavirus has ravaged the global healthcare system due to its high transmissibility and unpredictable clinical course that often affects multiple organ systems. Moreover, the long-term consequences of COVID-19 infection remain poorly understood. A full understanding of the pathogenesis of COVID-19 will require an unraveling of the mechanisms of inflammation, immune dysfunction, endothelial cell injury and dysregulated coagulation that underlie this disease.

In our study, we used an integrative proteomic-metabolomic analysis to identify global molecular signatures specific to the acute illness of COVID-19, while many prior metabolomic and proteomic studies have not assessed the interplay between proteins and metabolites. Our analyses establish associations of specific inflammation and vascular injury-related proteins with various metabolites during COVID-19, which appear to link inflammation with mitochondria-dependent energy metabolism and viral replication, as well as coagulation with fibrogenesis and glycolysis.

There is increasing evidence that evaluating symptoms and multiple clinical outcomes during acute disease is crucial in determining the risk of long COVID-19(29). To the best of our knowledge, we are the first group to develop a composite outcome measure in COVID-19 using multiple clinical indices in a prediction model that assesses not only COVID-19 disease severity but also the sequelae of COVID-19 that characterize post-acute COVID-19 syndrome (PACS). Compared to dichotomous outcome measures such as death and survival, our composite outcome score reflects a broader, more holistic assessment of COVID-19 morbidity in the hospital setting.

298 Several key strengths underlie our study cohort. As opposed to the use of healthy controls
299 reported in other COVID-19 studies(9, 26, 30-32), our use of non-COVID patient samples,
300 in the same hospital during the same time period between March and April 2020, allowed
301 us to investigate the interactions highly relevant to COVID-19 pathogenesis and clinical
302 course. We furthermore had relatively large sample sizes compared to other studies, with
303 hundreds of samples available for both metabolomic and proteomic analysis.

304 Our study has several limitations. As alterations in proteome and metabolome were
305 analyzed in sera but not in lung tissues or bronchoalveolar lavage fluid, our results may
306 not reflect what occurs at tissue-specific cellular levels. Furthermore, based on the current
307 study design and methodology, the correlative relationships we report between
308 metabolomic and proteomic alterations and SARS-CoV-2 outcomes should be interpreted
309 as purely correlative rather than causal in nature. Additional studies are required to define
310 the mechanistic roles of individual molecules highlighted in this paper. Finally, as our
311 study was only a single center investigation, our results will need to be validated in other
312 cohorts.

313 In conclusion, our investigation has sought to not only define the metabolomic and
314 proteomic signatures of COVID-19, but also to explore interactions between metabolites
315 and proteins that can serve as a roadmap for future mechanistic studies. We have
316 furthermore proposed a novel clinical composite outcome score that can be used in a
317 clinical prediction model for COVID-19. Ultimately, a better understanding of the
318 pathophysiology of COVID-19 at the molecular level may lead to short-term and long-
319 term targeted therapies.

320

321 **Methods**

322 **Cohort description**

323 This is a single-center prospective analysis of one cohort comparing hospitalized COVID-
324 19 patients and non-COVID-19 controls. Our cohort was comprised of 330 patients with
325 confirmed SARS-CoV-2 RT-PCR, and 97 non-COVID-19 controls with negative RT-PCR
326 results who were hospitalized at the NewYork-Presbyterian Hospital/Weill Cornell
327 Medical Center between March and April 2020. This study has been approved by the

328 Weill Cornell Medicine (WCM) IRB with protocol #19-10020914. Remnant serum samples
329 were matched with selected patients after which patients were deidentified. Hospitalized
330 controls were selected based on the results of RT-PCR for SARS-CoV-2 reported as
331 negative, and then they were matched the samples with selected patients. Children (less
332 than 18 years old) and pregnant women (confirmed by a positive beta-HCG test and/or
333 medical records) were excluded.
334

335 **Sample handling**

336 Standard practices for serum collection and storage at the NewYork-Presbyterian/Weill
337 Cornell Medical College include collecting venous blood into a serum-separating tube
338 (SST), and serum is obtained by centrifuging at 1,500g for 7 minutes as soon as possible
339 with a maximum time limit of 2 hours from the time of collection. The specimens are
340 typically stored at 4°C for 1 to 5 days before coded/de-identified and then transferred into
341 a -80°C freezer. Samples were thawed and inactivated in different ways: for the
342 metabolomic analysis, x3 sample volume of HPLC grade ethanol were added; for the
343 proteomic analysis, the samples were heat-inactivated in a water bath of 56°C for 15
344 minutes. After these processes, the samples were stored again at -80°C until the analyses
345 were performed.
346

347 **Data Collection**

348 Data were obtained from the Weill Cornell Medicine COVID Institutional Data Repository
349 (COVID-IDR), which is a high-quality registry of COVID-19 patients at NewYork-
350 Presbyterian - Cornell with laboratory confirmed SARS-CoV-2 RT-PCR. The COVID-IDR
351 houses both manually and automatically extracted Electronic Health Record (EHR) data.
352 Demographics, comorbidities, and important dates of patients' hospital course
353 (admission, intubation, extubation, discharge, death) were extracted by a team of medical
354 professionals and stored in the COVID-IDR. Laboratory tests, ventilation parameters, vital
355 signs, and respiratory variables were additionally available via automated extraction
356 through the Weill Cornell-Critical Care Database for Advanced Research (WC-CEDAR)
357 within the COVID-IDR. WC-CEDAR(33) is a critical care database originally designed to
358 automatically extract, transform, and store EHR data on Intensive Care Unit (ICU)

359 patients; it was expanded to include all hospitalized patients during New York City's
360 COVID-19 surge. Data not available within WC-CEDAR were manually extracted and
361 recorded in REDCap.
362

363 **Metabolomic profiling**

364 Metabolite profiling was performed according to methods described in a previous
365 publication(34). In brief, metabolites were extracted using pre-cooled 80% methanol and
366 measured on a Q Exactive Orbitrap mass spectrometer (Thermo Scientific), which was
367 coupled to a Vanquish UPLC system (Thermo Scientific) via an Ion Max source and HESI
368 II probe (Thermo Scientific). The Q Exactive was operated in full scan, polarity-switching
369 mode. A Sequant ZIC-HILIC column (2.1 mm i.d. × 150 mm, Merck) was used for
370 separation of metabolites. Flow rate was set at 150µL/min. Buffers consisted of 100%
371 acetonitrile for mobile phase A, and 0.1% NH4OH/20 mM CH3COONH4 in water for
372 mobile phase B. Gradient ran from 85% to 30% A in 20 min followed by a wash with 30%
373 A and re-equilibration at 85% A. Metabolites were identified based on exact mass within
374 5ppm and standard retention times.
375

376 **Proteomic profiling**

377 Proteomics analysis was performed using the Olink platform (Uppsala, Sweden) at the
378 Proteomics Core of Weill Cornell Medicine-Qatar, according to manufacturer's
379 instructions. We used the Inflammation, Cardiovascular II and Cardiovascular III panels.
380 High throughput real-time PCR of reporter DNA lined to protein specific antibodies was
381 performed on a 96-well integrated fluidic circuits chip (Fluidigm, San Francisco, CA). Each
382 sample was spiked with quality controls to monitor the incubation, extension, and
383 detection steps of the assay. Additionally, samples representing external, negative, and
384 inter-plate controls were included in each analysis run. From raw data, real time PCR
385 cycle threshold (Ct) values were extracted using Fluidigm reverse transcription
386 polymerase chain reaction (RT-PCR) analysis software at a quality threshold of 0.5 and
387 linear baseline correction. Ct values were further processed using the Olink NPX manager
388 software (Olink, Uppsala, Sweden). Here, log2-transformed Ct values from each sample
389 and analyte were normalized based on spiked-in extension controls and scale-inverted to

obtain Normalized log2-scaled Protein Expression (NPX) values. NPX values were adjusted based on the median of inter plate controls (IPC) for each protein and intensity median scaled between all samples and plates.

Each metabolite and protein was annotated with pathways from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database(35).

Data preprocessing

Children, pregnant women, and samples after intubation were excluded from all analyses.

The metabolomics data was measured in three different batches. For each batch, data was preprocessed by filtering out samples with more than 50% missing values, followed by filtering out metabolites with more than 25% missing values, probabilistic quotient normalization(36), and log2 transformation. Two extreme outlier metabolites were manually removed (phosphorylcholine and adenosine monophosphate). The next step was to merge the different batches into a joint dataset. Batch 3 contained only control patients and could thus not be simply added by batch correction. Instead, batches 2 and 3 contained an overlapping set of samples which were used for an anchor-based normalization by dividing each metabolite in batch 3 by the mean fold change of the overlapping samples. The anchor samples from batch 3 were then deleted and the batches combined using median-based batch correction(37). Missing values were then imputed using the k-nearest neighbor approach(37).

Proteomics data preprocessing included the same steps of filtering, quotient normalization, logging, and missing value imputation with identical parameters as for the metabolomics data. Ten proteins were measured as duplicates on the Olink platform, so their expression values were averaged.

Statistical analysis

Differential expression of metabolites and proteins for both the COVID-19 vs. control analysis as well as the clinical parameter analysis within the COVID-19 cohort was assessed using the following linear mixed effect model:

419 *met_prot ~ outcome + time + (1|patient),*

420 where *met_prot* is each individual metabolite or protein, *outcome* is either COVID-19
421 yes/no or the value of a clinical parameter, *time* is the day of sample taking as a factor,
422 and *(1|patient)* is a random effect per patient to account for repeated measurements.
423 P-values are reported for the significance of the outcome term.

424 Data preprocessing and statistical analysis was performed using the "maplet" toolbox for
425 R(38) (https://github.com/krumsieklab/maplet).

426

427 **Network construction**

428 The dataset was first reduced to the samples that overlap between metabolomics and
429 proteomics, and corrected for age, sex, BMI, and COVID-19 status (yes/no). A Gaussian
430 Graphical Model (GGM) based network was then constructed using the GeneNet
431 algorithm(39) and drawing an edge for all partial correlations with an FDR smaller than
432 0.2. In a second step, this network was condensed to highlight the connections between
433 molecules that were significantly different between COVID-19 and controls. To this end,
434 a shortest-path distance matrix between all molecules was constructed and subset to the
435 significant molecules. A minimum spanning tree(40) of this matrix was then constructed
436 to visualize a simplified network.

437

438 **Composite Outcome**

439 A detailed description of the construction of the composite outcome can be found in
440 Supplementary Figure 1.

441 **Regularized linear mixed effect ordinal regression model**

442 A new regression model was developed to deal with an ordinal outcome, repeated
443 measurements, and feature selection with regularization. Repeated measurements are
444 handled as a random effect, while age, sex, BMI, and metabolites are treated as fixed
445 effects. The metabolites are penalized using an L_1 LASSO-type regularization to obtain a
446 sparse solution. The model is fitted using an mixed-effect ordinal regression model with
447 complementary log-log link function(41), using maximum likelihood (ML) estimation as
448 proposed by Ripatti and Pamgren(42). An optimal LASSO penalty parameter was

449 estimated through an iterative algorithm for maximizing the Laplace approximation of the
450 integrated model likelihood. This approach was adapted from Therneau(43), which was
451 originally developed for mixed effect Cox models. To obtain an unbiased estimate of the
452 model performance, leave-one-out-cross-validation across the entire dataset was
453 performed. The added value of metabolomics data over baseline clinical data was
454 assessed by comparing the final model with a model only consisting of age, sex, and BMI.
455

456 **Data Availability**

457 There are legal and ethical restrictions on data sharing because the Institutional Review
458 Board of Weill Cornell Medicine did not approve public data deposition. The data set used
459 for this study constitutes sensitive patient information extracted from the electronic health
460 record. Accordingly, it is subject to federal legislation that limits our ability to disclose it to
461 the public, even after it has been subjected to deidentification techniques. To request the
462 access of the de-identified minimal dataset underlying these findings, interested and
463 qualified researchers should contact Information Technologies & Services Department of
464 Weill Cornell Medicine support@med.cornell.edu.
465

466 **Code Availability**

467 Code to reproduce all the statistical results presented in this paper is available at
468 https://github.com/krumsieklab/covid-omics.
469

470 **Funding**

471 JK is supported by the National Institute of Aging of the National Institutes of Health under
472 award 1U19AG063744. SJC is supported by National Heart, Lung and Blood Institute
473 under award K08HL138285. KS is supported by 'Biomedical Research Program' funds at
474 Weill Cornell Medical College in Qatar, a program funded by the Qatar Foundation and
475 multiple grants from the Qatar National Research Fund (QNRF).
476

References

1. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med. 2020;382(18):1708-20.
2. Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. N Engl J Med. 2020;382(22):2081-90.
3. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. The Lancet. 2020;395(10229):1054-62.
4. Hirsch JS, Ng JH, Ross DW, Sharma P, Shah HH, Barnett RL, et al. Acute kidney injury in patients hospitalized with COVID-19. Kidney Int. 2020;98(1):209-18.
5. Zhang C, Shi L, Wang F-S. Liver injury in COVID-19: management and challenges. The Lancet Gastroenterology & Hepatology. 2020;5(5):428-30.
6. Bilaloglu S, Aphinyanaphongs Y, Jones S, Iturrate E, Hochman J, Berger JS. Thrombosis in Hospitalized Patients With COVID-19 in a New York City Health System. JAMA. 2020.
7. Tang N, Li D, Wang X, Sun Z. Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. J Thromb Haemost. 2020;18(4):844-7.
8. Elezkurtaj S, Greuel S, Ihlow J, Michaelis EG, Bischoff P, Kunze CA, et al. Causes of death and comorbidities in hospitalized patients with COVID-19. Scientific Reports. 2021;11(1):4263.
9. Lucas C, Wong P, Klein J, Castro TBR, Silva J, Sundaram M, et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. Nature. 2020;584(7821):463-9.
10. Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ. Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. BMC Syst Biol. 2011;5:21.
11. Krumsiek J, Suhre K, Evans AM, Mitchell MW, Mohney RP, Milburn MV, et al. Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. PLoS Genet. 2012;8(10):e1003005.
12. Benedetti E, Pučić-Baković M, Keser T, Wahl A, Hassinen A, Yang JY, et al. Network inference from glycoproteomics data reveals new reactions in the IgG glycosylation pathway. Nat Commun. 2017;8(1):1483.
13. Tokunaga R, Zhang W, Naseem M, Puccini A, Berger MD, Soni S, et al. CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation - A target for novel cancer therapy. Cancer Treat Rev. 2018;63:40-7.
14. Stanley CA, Hale DE, Berry GT, Deleeuw S, Boxer J, Bonnefont J-P. A Deficiency of Carnitine–Acylcarnitine Translocase in the Inner Mitochondrial Membrane. New England Journal of Medicine. 1992;327(1):19-23.
15. McGonagle D, Ramanan AV, Bridgewood C. Immune cartography of macrophage activation syndrome in the COVID-19 era. Nature Reviews Rheumatology. 2021;17(3):145-57.
16. Hoelzer K, Shackelton LA, Parrish CR. Presence and role of cytosine methylation in DNA viruses of animals. Nucleic acids research. 2008;36(9):2825-37.
17. Liu M, O'Connor RS, Trefely S, Graham K, Snyder NW, Beatty GL. Metabolic rewiring of macrophages by CpG potentiates clearance of cancer cells and overcomes tumor-expressed CD47−mediated 'don't-eat-me' signal. Nature Immunology. 2019;20(3):265-75.
18. Varga Z, Flammer AJ, Steiger P, Haberecker M, Andermatt R, Zinkernagel AS, et al. Endothelial cell infection and endotheliitis in COVID-19. The Lancet. 2020;395(10234):1417-8.

525 19. Xu S, Gu M, Wu K, Li G. Unraveling the Role of Hydroxyproline in Maintaining the
526 Thermal Stability of the Collagen Triple Helix Structure Using Simulation. The Journal of
527 Physical Chemistry B. 2019;123(36):7754-63.
528 20. Takeshita A, Yasuma T, Nishihama K, D'Alessandro-Gabazza CN, Toda M, Totoki T, et
529 al. Thrombomodulin ameliorates transforming growth factor-β1–mediated chronic kidney
530 disease via the G-protein coupled receptor 15/Akt signal pathway. Kidney International.
531 2020;98(5):1179-92.
532 21. Monickaraj F, McGuire P, Das A. Cathepsin D plays a role in endothelial–pericyte
533 interactions during alteration of the blood–retinal barrier in diabetic retinopathy. The FASEB
534 Journal. 2018;32(5):2539-48.
535 22. Leung SWS, Shi Y. The glycolytic process in endothelial cells and its implications. Acta
536 Pharmacologica Sinica. 2021.
537 23. Ackermann M, Verleden SE, Kuehnel M, Haverich A, Welte T, Laenger F, et al.
538 Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. New England
539 Journal of Medicine. 2020;383(2):120-8.
540 24. Chen J, Wu C, Wang X, Yu J, Sun Z. The Impact of COVID-19 on Blood Glucose: A
541 Systematic Review and Meta-Analysis. Frontiers in Endocrinology. 2020;11(732).
542 25. Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS, et al. Post-
543 acute COVID-19 syndrome. Nature Medicine. 2021;27(4):601-15.
544 26. Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, et al. Proteomic and Metabolomic
545 Characterization of COVID-19 Patient Sera. Cell. 2020;182(1):59-72 e15.
546 27. Senn S, editor Dichotomania: An Obsessive Compulsive Disorder that is Badly Affecting
547 the Quality of Analysis of Pharmaceutical Trials2005.
548 28. Roozenbeek B, Lingsma HF, Perel P, Edwards P, Roberts I, Murray GD, et al. The
549 added value of ordinal analysis in clinical trials: an example in traumatic brain injury. Crit Care.
550 2011;15(3):R127.
551 29. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al.
552 Attributes and predictors of long COVID. Nature Medicine. 2021;27(4):626-31.
553 30. Arunachalam PS, Wimmers F, Mok CKP, Perera RAPM, Scott M, Hagan T, et al.
554 Systems biological assessment of immunity to mild versus severe COVID-19 infection in
555 humans. Science. 2020;369(6508):1210-20.
556 31. Overmyer KA, Shishkova E, Miller IJ, Balnis J, Bernstein MN, Peters-Clarke TM, et al.
557 Large-Scale Multi-omic Analysis of COVID-19 Severity. Cell Systems. 2021;12(1):23-40.e7.
558 32. Su Y, Chen D, Yuan D, Lausted C, Choi J, Dai CL, et al. Multi-Omics Resolves a Sharp
559 Disease-State Shift between Mild and Moderate COVID-19. Cell. 2020;183(6):1479-95.e20.
560 33. Schenck EJ, Hoffman KL, Cusick M, Kabariti J, Sholle ET, Campion TR. Critical carE
561 Database for Advanced Research (CEDAR): An Automated Method to Support Intensive Care
562 Units with Electronic Health Record Data. J Biomed Inform. 2021:103789.
563 34. Goncalves MD, Hwang SK, Pauli C, Murphy CJ, Cheng Z, Hopkins BD, et al.
564 Fenofibrate prevents skeletal muscle loss in mice with lung cancer. Proc Natl Acad Sci U S A.
565 2018;115(4):E743-e52.
566 35. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating
567 viruses and cellular organisms. Nucleic Acids Res. 2021;49(D1):D545-d51.
568 36. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as
569 robust method to account for dilution of complex biological mixtures. Application in 1H NMR
570 metabonomics. Anal Chem. 2006;78(13):4281-90.
571 37. Do KT, Wahl S, Raffler J, Molnos S, Laimighofer M, Adamski J, et al. Characterization of
572 missing values in untargeted MS-based metabolomics data and evaluation of missing data
573 handling strategies. Metabolomics. 2018;14(10):128.
574 38. Chetnik K, Benedetti E, Gomari DP, Schweickart A, Batra R, Buyukozkan M, et al.
575 maplet: An extensible R toolbox for modular and reproducible omics pipelines2021 May 01,

2021:[arXiv:2105.04305 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2021arXiv210504305C.
39. Schäfer J, Strimmer K. A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Stat Appl Genet Mol Biol. 2005;4:Article32.
40. Prim RC. Shortest Connection Networks And Some Generalizations. Bell System Technical Journal. 1957;36(6):1389-401.
41. McCullagh P. Regression Models for Ordinal Data. Journal of the Royal Statistical Society Series B (Methodological). 1980;42(2):109-42.
42. Ripatti S, Palmgren J. Estimation of multivariate frailty models using penalized partial likelihood. Biometrics. 2000;56(4):1016-22.
43. Therneau T. Coxme and the Laplace approximation 2020 [Available from: https://cran.r-project.org/web/packages/coxme/vignettes/laplace.pdf.

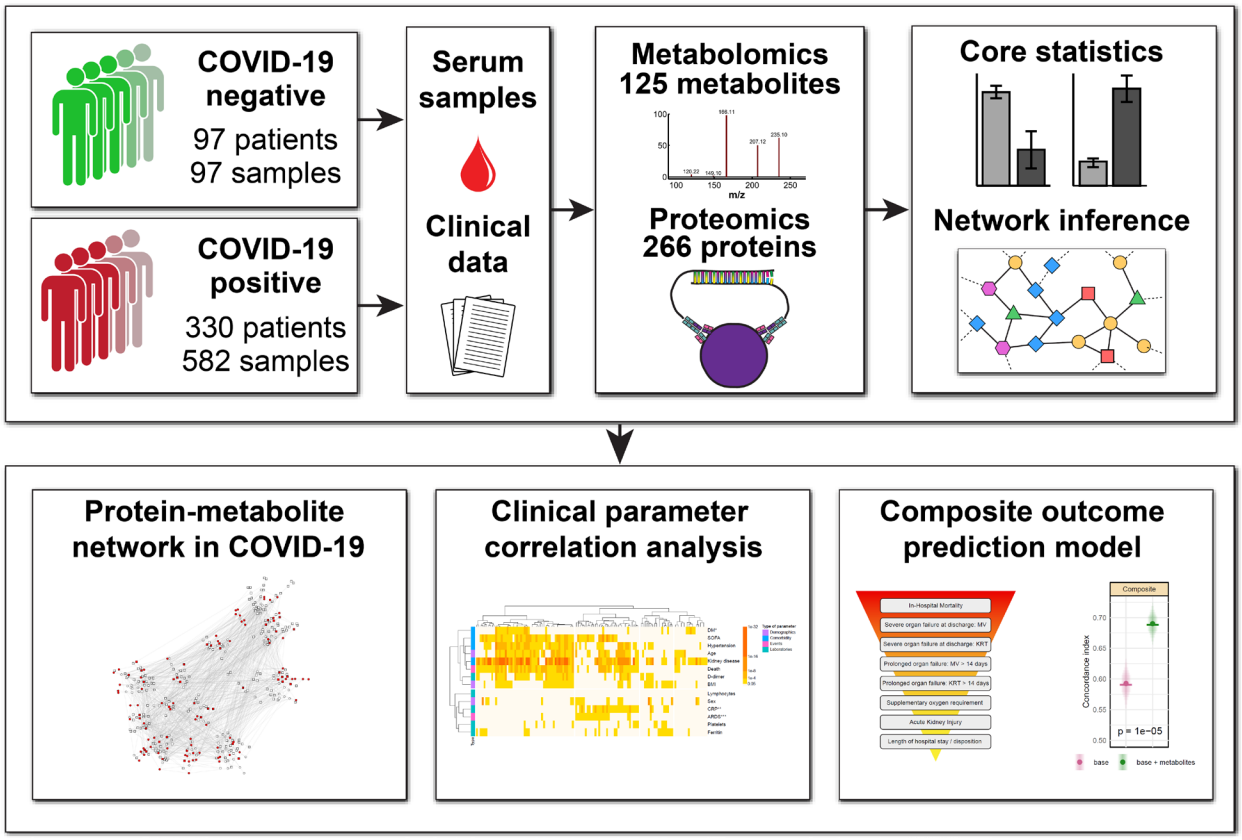
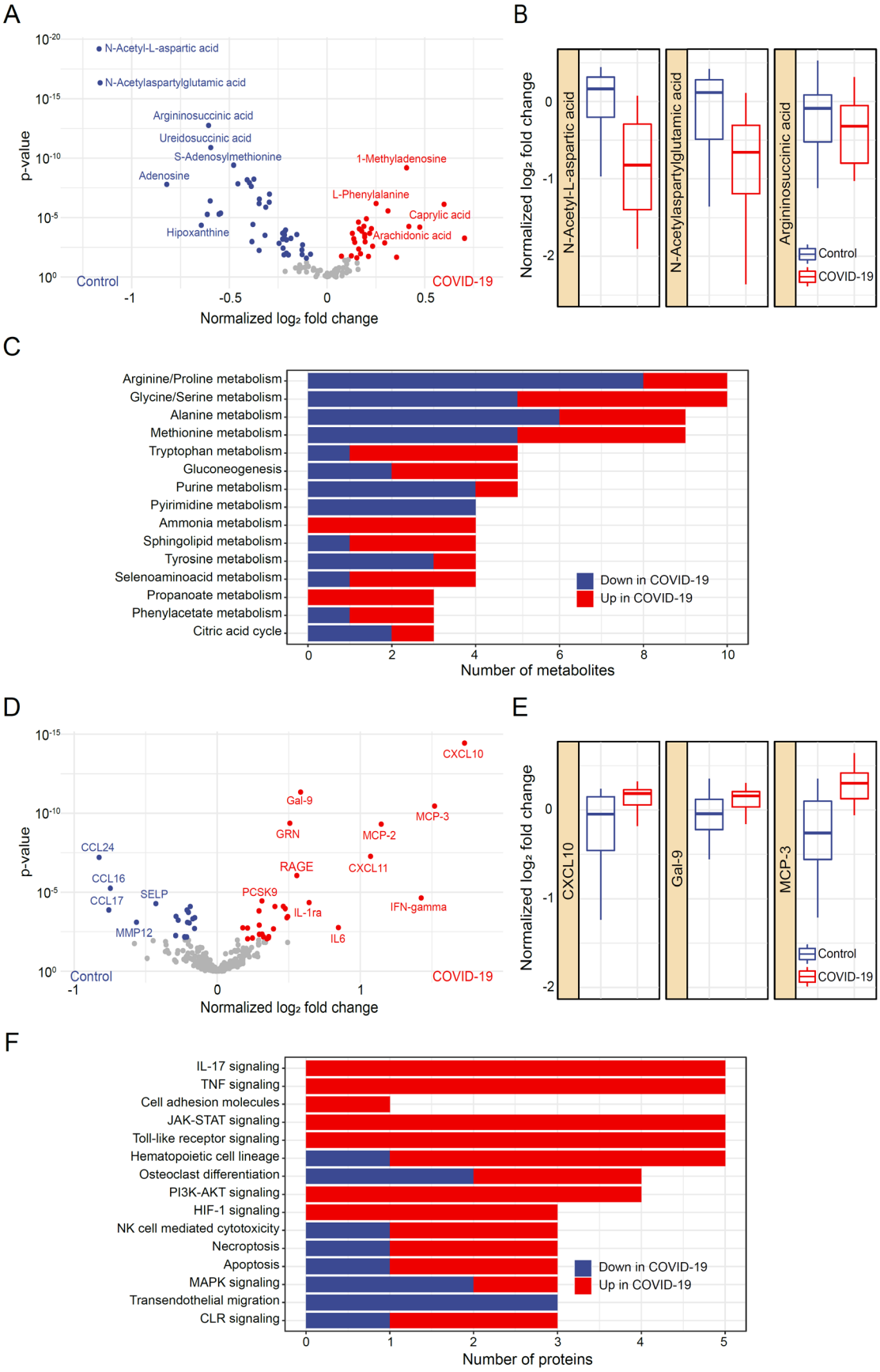
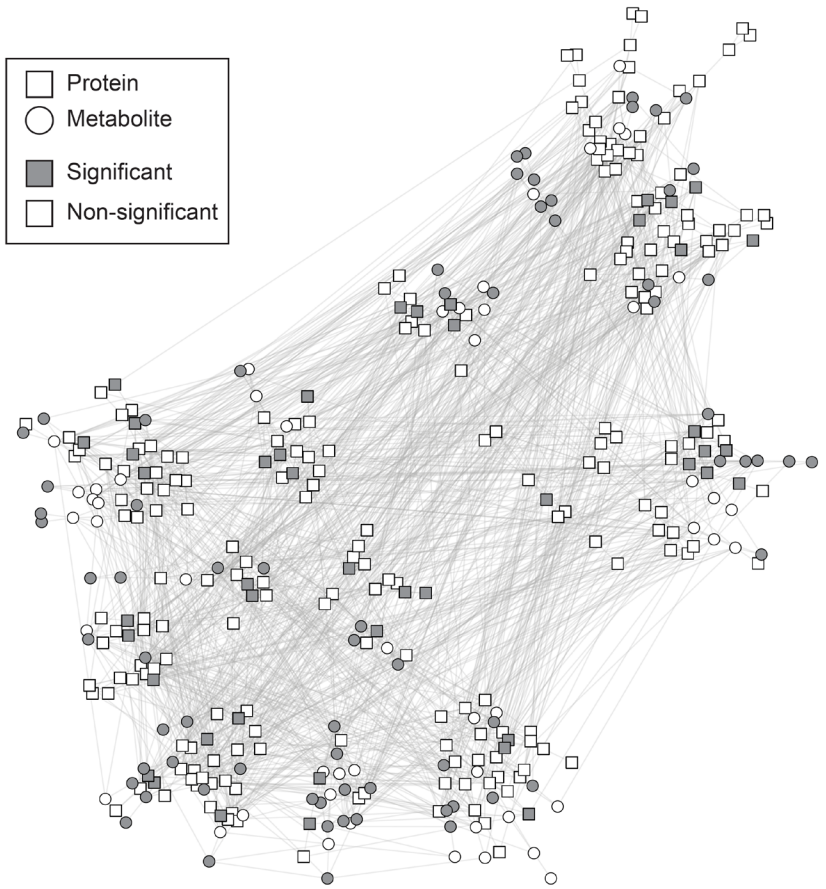


Fig. 1. Study outline. Study population composed by COVID-19 patients and controls. Serum samples obtained from clinically indicated specimens during the first 72 hours of admission. Metadata obtained from electronic medical records. We obtained metabolomic and proteomic profile, after which we integrated our findings into a single network describing the associations between differentially expressed proteins and metabolites in COVID-19. We then correlated the metabolites and proteins with baseline characteristics, laboratory parameters and clinical events. Finally, we developed a novel metabolite-based prediction model for a composite outcome measure comprised of key clinical parameters including death, mechanical ventilation, initiation of dialysis, supplemental oxygen requirement, development of acute kidney injury, and length of hospital stay.

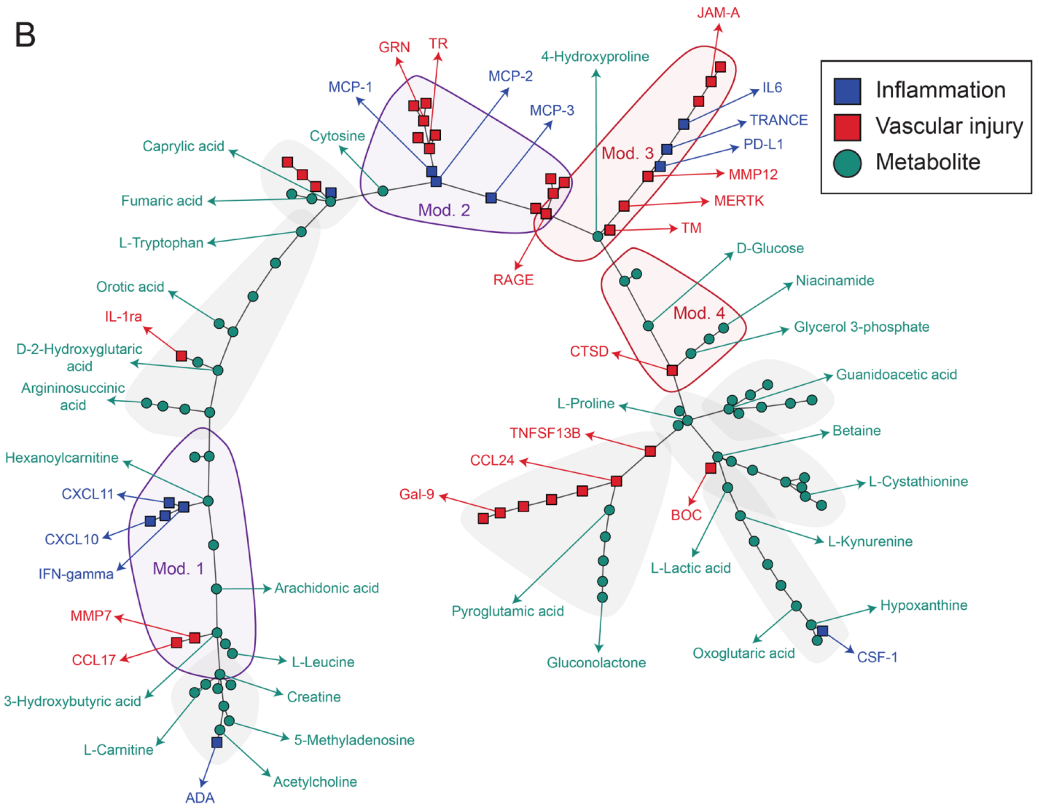

A
N-Acetyl-L-aspartic acid
N-Acetylaspartylglutamic acid
Argininosuccinic acid
Ureidosuccinic acid
S-Adenosylmethionine
Adenosine
Hipoxanthine
1-Methyladenosine
L-Phenylalanine
Caprylic acid
Arachidonic acid
p-value
Control
COVID-19
Normalized log₂ fold change
B
N-Acetyl-L-aspartic acid
N-Acetylaspartylglutamic acid
Argininosuccinic acid
Normalized log₂ fold change
Control
COVID-19
C
Arginine/Proline metabolism
Glycine/Serine metabolism
Alanine metabolism
Methionine metabolism
Tryptophan metabolism
Gluconeogenesis
Purine metabolism
Pyirimidine metabolism
Ammonia metabolism
Sphingolipid metabolism
Tyrosine metabolism
Selenoaminoacid metabolism
Propanoate metabolism
Phenylacetate metabolism
Citric acid cycle
Down in COVID-19
Up in COVID-19
Number of metabolites
D
CXCL10
Gal-9
MCP-3
GRN
MCP-2
CCL24
RAGE
CXCL11
CCL16
CCL17
SELP
PCSK9
IL-1ra
IFN-gamma
MMP12
IL6
p-value
Control
COVID-19
Normalized log₂ fold change
E
CXCL10
Gal-9
MCP-3
Normalized log₂ fold change
Control
COVID-19
F
IL-17 signaling
TNF signaling
Cell adhesion molecules
JAK-STAT signaling
Toll-like receptor signaling
Hematopoietic cell lineage
Osteoclast differentiation
PI3K-AKT signaling
HIF-1 signaling
NK cell mediated cytotoxicity
Necroptosis
Apoptosis
MAPK signaling
Transendothelial migration
CLR signaling
Down in COVID-19
Up in COVID-19
Number of proteins

605 **Fig. 2. Metabolomics and Proteomics changes associated with COVID-19. A,** Volcano plot
606 showing differentially expressed metabolites between COVID-19 patients and controls at an
607 adjusted p-value <0.05. In red, upregulated metabolites in COVID-19 patients. In blue,
608 upregulated metabolites in the control group. **B,** Top 3 differentially expressed metabolites in
609 COVID-19 patients vs. controls based on adjusted p-value. **C,** Bar plot representing the main
610 KEGG metabolic pathways altered in COVID-19 based on the previous results. **D,** Volcano plot
611 showing the differentially expressed proteins between COVID-19 patients and controls at an
612 adjusted p-value <0.05. **E,** Top 3 differentially expressed proteins in COVID-19 patients vs.
613 controls based on adjusted p-value. **F,** Bar plot representing the main altered protein-based
614 KEGG signaling pathways altered in COVID-19 based on the previous results.
615

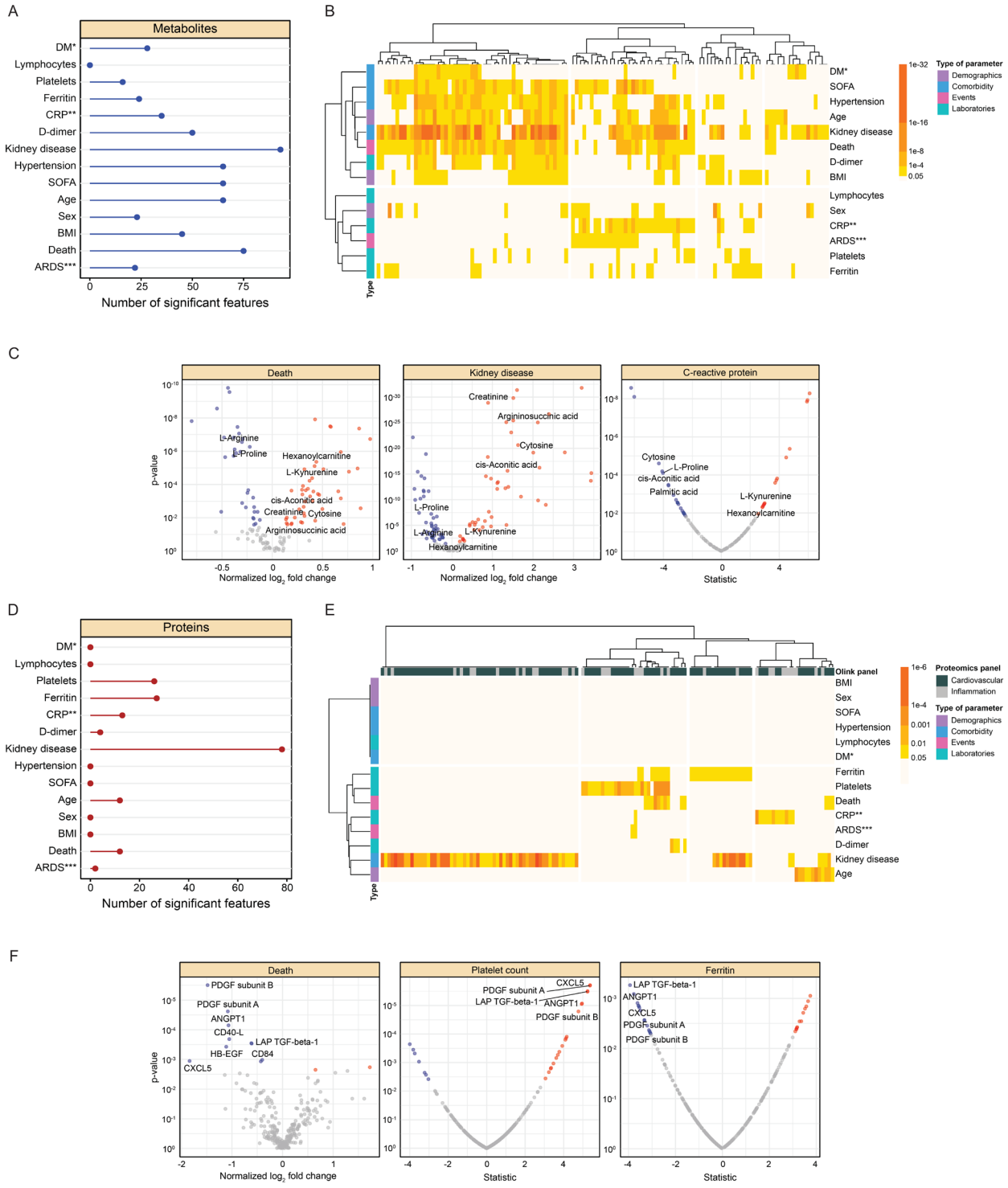
A


Protein
Metabolite
Significant
Non-significant

B


Inflammation
Vascular injury
Metabolite
Mod. 1
Mod. 2
Mod. 3
Mod. 4
GRN
TR
MCP-1
MCP-2
MCP-3
4-Hydroxyproline
JAM-A
IL6
TRANCE
PD-L1
MMP12
MERTK
TM
RAGE
Cytosine
Caprylic acid
Fumaric acid
L-Tryptophan
Orotic acid
IL-1ra
D-2-Hydroxyglutaric acid
Argininosuccinic acid
Hexanoylcarnitine
CXCL11
CXCL10
IFN-gamma
MMP7
CCL17
3-Hydroxybutyric acid
L-Carnitine
ADA
Acetylcholine
5-Methyladenosine
Creatine
L-Leucine
Arachidonic acid
D-Glucose
Niacinamide
Glycerol 3-phosphate
CTSD
Guanidoacetic acid
L-Proline
TNFSF13B
CCL24
Gal-9
Betaine
L-Cystathionine
BOC
L-Kynurenine
L-Lactic acid
Pyroglutamic acid
Gluconolactone
Hypoxanthine
Oxoglutaric acid
CSF-1

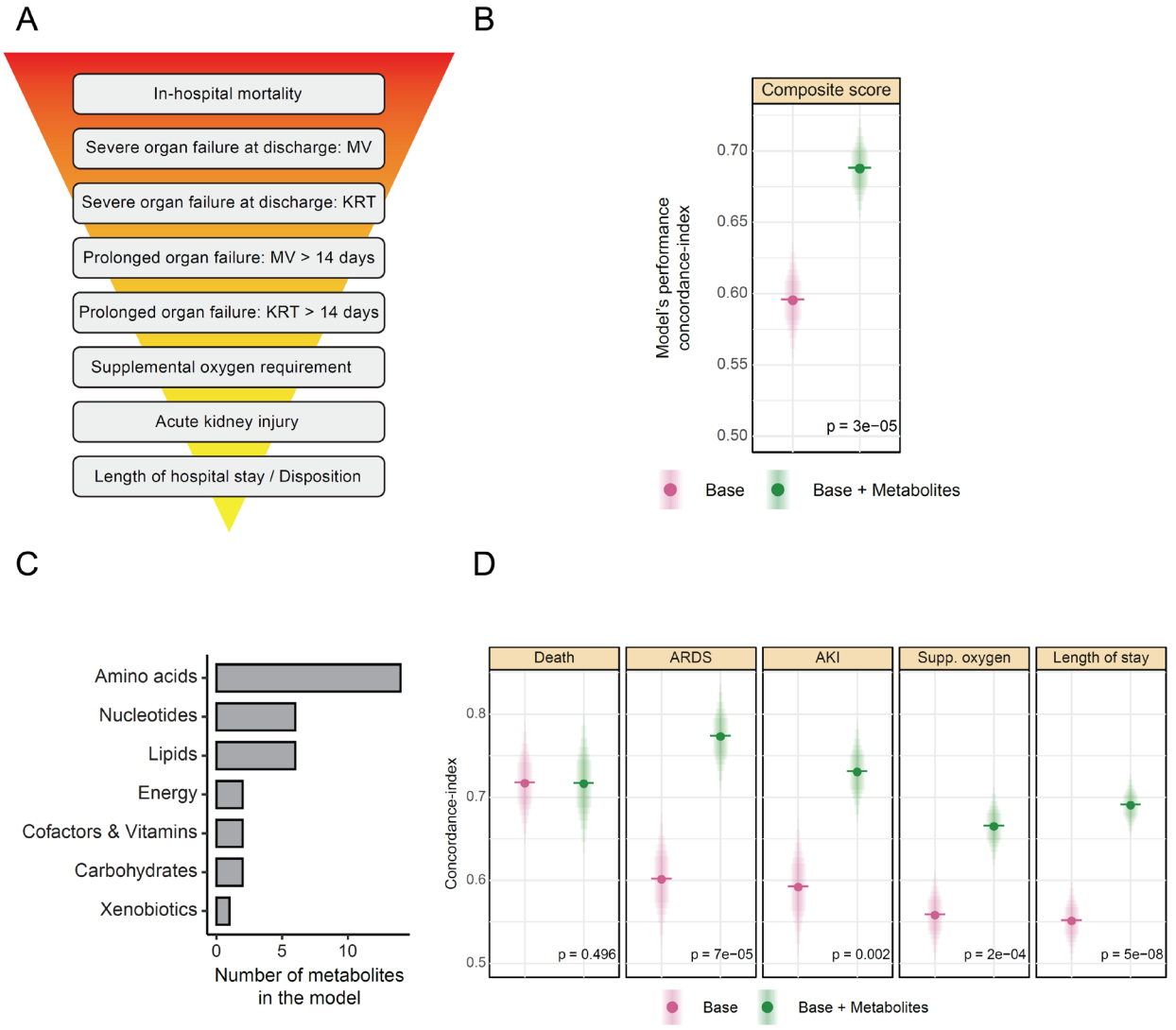
617 **Fig. 3. Protein-metabolite networks in COVID-19. A,** Gaussian graphical model (GGM)
618 representing the significant partial correlations between all the measured metabolites and proteins.
619 Circles indicate metabolites and squares indicate proteins. The differentially expressed molecules
620 between COVID-19 patients and controls are highlighted in gray. **B,** Simplified version of the
621 GGM network using minimum spanning (MST) algorithm combined with a shortest paths-based
622 approach, including all the differentially expressed metabolites and proteins between COVID-19
623 patients and controls. Squares indicate proteins and green circles indicate metabolites.
624 Inflammatory proteins are colored in purple, while vascular injury proteins are colored in red.
625 Shadows represent molecule modules.
626


A
Metabolites
DM*
Lymphocytes
Platelets
Ferritin
CRP**
D-dimer
Kidney disease
Hypertension
SOFA
Age
Sex
BMI
Death
ARDS***
0
25
50
75
Number of significant features
B
DM*
SOFA
Hypertension
Age
Kidney disease
Death
D-dimer
BMI
Lymphocytes
Sex
CRP**
ARDS***
Platelets
Ferritin
Type
1e-32
1e-16
1e-8
1e-4
0.05
Type of parameter
Demographics
Comorbidity
Events
Laboratories
C
Death
Kidney disease
C-reactive protein
p-value
L-Arginine
L-Proline
Hexanoylcarnitine
L-Kynurenine
cis-Aconitic acid
Creatinine
Cytosine
Argininosuccinic acid
Palmitic acid
Normalized log2 fold change
Statistic
D
Proteins
DM*
Lymphocytes
Platelets
Ferritin
CRP**
D-dimer
Kidney disease
Hypertension
SOFA
Age
Sex
BMI
Death
ARDS***
0
20
40
60
80
Number of significant features
E
Olink panel
BMI
Sex
SOFA
Hypertension
Lymphocytes
DM*
Ferritin
Platelets
Death
CRP**
ARDS***
D-dimer
Kidney disease
Age
Type
1e-6
1e-4
0.001
0.01
0.05
Proteomics panel
Cardiovascular
Inflammation
Type of parameter
Demographics
Comorbidity
Events
Laboratories
F
Death
Platelet count
Ferritin
PDGF subunit B
PDGF subunit A
ANGPT1
CD40-L
LAP TGF-beta-1
HB-EGF
CD84
CXCL5
p-value
Normalized log2 fold change
Statistic

627

628

Fig. 4. Associations of molecules with clinical indices in COVID-19. A, Lollipop plot representing the number of altered metabolites for each analyzed parameter (*Diabetes mellitus, **C-reactive protein, ***Acute respiratory distress syndrome). **B,** Heatmap indicating the number of differentially expressed metabolites associated with each clinical parameter and its significance. We included 4 clinical parameter categories: demographics, comorbidities, clinical events, and laboratory parameters. **C,** Volcano plot showing detailed metabolic changes correlated to death, kidney disease and C-reactive protein. **D,** Lollipop plot representing the number of altered proteins for each analyzed parameter. **E,** Heatmap indicating the number of differentially expressed proteins on each clinical parameter and its significance, using the same 4 categories described above. Proteins are marked according to Olink panels. F, Volcano plot showing detailed protein changes correlated to death, platelet count and ferritin levels.



642 **Fig. 5. Metabolomics signature predicts clinical outcomes of COVID-19. A,** Scheme of the
643 hierarchical composite outcome ranging from in-hospital mortality to length of hospital stay and
644 disposition (MV: Mechanical ventilation, KRT: Kidney replacement therapy). **B,** Dot plot
645 comparing the predictive performance (Concordance index) of a baseline model using age, sex,
646 and BMI and a model including baseline plus metabolites. The results show that the prediction
647 accuracy improves when adding metabolites to the model. **C,** Bar plot showing the main altered
648 metabolic groups included in the composite outcome score. **D,** Dot plots showing individual
649 outcome analysis, comparing the baseline model and baseline plus metabolites added.
650 Metabolites improve the prediction accuracy of ARDS, AKI, supplemental oxygen requirement
651 and prolonged hospitalization.
652

Table 1

Characteristic	Negative, N = 97[1]	COVID-19, N = 330[1]	p-value[2]
Demographics			
Age	64 (51, 77)	64 (52, 75)	0.8
Sex			**0.017**
Female	48 (49%)	117 (35%)	
Male	49 (51%)	213 (65%)	
Race			**<0.001**
Asian	10 (10%)	24 (7.3%)	
Black	15 (15%)	40 (12%)	
Not specified	16 (16%)	48 (15%)	
Other	6 (6.2%)	71 (22%)	
White	50 (52%)	147 (45%)	
BMI	24.7 (21.7, 28.5)	27.4 (23.7, 31.0)	**<0.001**
Risk Factors			
CAD	23 (24%)	55 (17%)	0.2
DM	19 (20%)	89 (27%)	0.2
HTN	55 (57%)	182 (55%)	0.9
CKD/ESRD	17 (18%)	17 (5.2%)	**<0.001**
Active cancer	40 (41%)	30 (9.1%)	**<0.001**
Immunosuppressed state	29 (30%)	14 (4.2%)	**<0.001**
Laboratories			
Lymphocyte count	1.03 (0.50, 1.60)	0.80 (0.60, 1.10)	0.081
Platelet count	202 (132, 275)	191 (145, 252)	0.9
D-dimer	242 (150, 555)	386 (242, 667)	0.2
Alanine aminotransferase	22 (14, 40)	35 (23, 54)	**<0.001**
Aspartate aminotransferase	25 (19, 38)	40 (28, 61)	**<0.001**
Ferritin	727 (147, 1167)	769 (358, 1449)	0.4
C-reactive protein	3 (1, 7)	11 (6, 19)	**0.006**
Procalcitonin	0.43 (0.13, 1.06)	0.17 (0.08, 0.42)	**0.007**
Clinical Outcomes			
Acute Kidney Injury (AKI) stage			**<0.001**
0	89 (92%)	214 (65%)	
1	4 (4.1%)	43 (13%)	
2	0 (0%)	16 (4.8%)	
3	4 (4.1%)	42 (13%)	
Kidney replacement therapy	7 (7.2%)	26 (7.9%)	>0.9
Acute respiratory failure requiring intubation	11 (11%)	103 (31%)	**<0.001**
Thromboembolic event	6 (6.2%)	22 (6.7%)	>0.9
In-hospital mortality	13 (13%)	56 (17%)	0.6

[1] Statistics presented: median (IQR); n (%)
[2] Statistical tests performed: Wilcoxon rank-sum test; chi-square test of independence; Fisher's test with simulated p-value; Fisher's exact test

653