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

Mustafa Buyukozkan[1,2*], Sergio Alvarez-Mulett[3*], Alexandra C. Racanelli[3*], Frank Schmidt[4], Richa Batra[1,2], Katherine L. Hoffman[5], Hina Sarwath[4], Rudolf Engelke[4], Luis Gomez-Escobar[3], Will Simmons[5], Elisa Benedetti[1,2], Kelsey Chetnik[1,2], Guoan Zhang[6], Edward Schenck[3], Karsten Suhre[7], Justin J. Choi[8], Zhen Zhao[9], Sabrina Racine-Brzostek[9], He S. Yang[9], Mary E. Choi[10], Augustine M.K. Choi[3], Soo Jung Cho[3#], Jan Krumsiek[1,2#]

[1]Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA
[2]Meyer Cancer Center and Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY, USA
[3]Department of Medicine, Division of Pulmonary and Critical Care Medicine, Weill Cornell Medicine, New York, NY, USA
[4]Proteomics Core, Weill Cornell Medicine – Qatar, Doha, Qatar
[5]Department of Population Health Sciences, Division of Biostatistics, Weill Cornell Medicine, New York, NY, USA
[6]Proteomics and Metabolomics Core Facility, Weill Cornell Medicine, New York, NY, USA
[7]Department of Physiology and Biophysics, Weill Cornell Medicine – Qatar, Education City, 24144 Doha, Qatar
[8]Department of Medicine, Division of General Internal Medicine, Weill Cornell Medicine, New York, NY, USA
[9]Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY, USA
[10]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:
Soo Jung Cho, sjc9006@med.cornell.edu
Jan Krumsiek, jak2042@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 based on the metabolomics data. The model predicts severe disease with a concordance index of around 0.69, and furthermore shows high predictive power of 0.83-0.93 in two previously published, independent datasets.

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 targeted metabolomic and proteomic analyses of serum samples obtained from hospitalized COVID-19 patients during SARS-CoV-2 infection compared to patients admitted during the same time period with symptoms related to COVID-19 and negative RT-PCR for SARS-CoV-2 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. Metabolomics was measured for all available samples. Proteomics was measured for fewer samples (n=189), also across different time points for some patients. Notably, there were only minor time effects across the three days (Supplementary Figure 1), and the repeated samples were thus treated as replicates

using a linear mixed effect model, see Methods. A detailed description of the clinical and demographic characteristics of the cohort can be found in Table 1, Supplementary Table 1 and Supplementary Table 2. 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 3).

Metabolic profiles were assessed for all samples using liquid chromatography coupled with mass spectrometry (LC/MS). After quality control and data preprocessing, 125 metabolites were available for comparative analysis. Targeted proteomic profiling was performed on a subset of 227 samples (173 from COVID-19 patients and 54 controls) using the Olink inflammation, cardiovascular II and cardiovascular III panels, which cover 266 unique protein biomarkers. These panels were selected since it has previously been shown that inflammation and cardiovascular pathways are essential during COVID-19 pathogenesis(10).

Metabolomic and proteomic changes associated with COVID-19

Differential metabolomic analysis identified significant changes in abundance of 70 out of the 125 analyzed metabolites between COVID-19 and controls at a false discovery rate (FDR) of 0.05 (Figure 2A). The top three differentially expressed metabolites were involved in amino acid metabolism: N-acetyl-L-aspartic acid (p-value = 9.19E-18), N-acetyl-aspartyl-glutamic acid (p-value = 5.30E-15), and argininosuccinic acid (p-value = 6.35E-12) (Figure 2B). KEGG pathway mapping of the differentially expressed metabolites revealed an involvement of various metabolic pathways (Figure 2C). These pathways included arginine and proline metabolism, glycine and serine metabolism, alanine metabolism, methionine metabolism, sphingolipid metabolism, gluconeogenesis, and the TCA cycle pathway, demonstrating involvement of the broader categories of amino acid, lipid and energy metabolism in COVID-19 pathogenesis. Our results confirm prior reports that reported altered phenylalanine and tryptophan metabolism in severe COVID-19 patients compared to non-COVID-19 patients(11-13). Additionally, our data has shown that multiple metabolites involved in sphingolipid metabolism are significantly

increased in COVID-19 patients, pointing towards other potential targets for COVID-19 treatment(14).

Comparative proteomic analysis identified significant changes in the expression of 48 out of the 266 analyzed proteins between COVID-19 and controls at an FDR of 0.05 (Figure 2D). The top three differentially expressed proteins were C-X-C motif chemokine ligand 10 (CXCL10) (p-value = 5.09E-13), galectin 9 (Gal-9) (p-value = 5.49E-10), and monocyte chemoattractant protein 3 (MCP-3) (p-value = 2.85E-09) (Figure 2E). KEGG pathway mapping revealed that these proteins participated in various protein pathways (Figure 2F), including the interleukin 17 (IL-17), tumoral necrosis factor (TNF) and JAK-STAT signaling pathways. Detailed results of the differential analysis and pathway mappings can be found in Supplementary Table 4. Of note, the utility of the JAK inhibitors baricitinib or ruxolitinib has already been demonstrated in a clinical trial for selected patients with severe or critical COVID-19 patients(15, 16). Based on our data, clinical trials further targeting IL-17 and TNF signaling in COVID-19 may lead to additional therapeutic approaches for treating COVID-19.

Global principal component analysis (PCA) on the metabolomics and proteomics data revealed no clear separation of COVID-19 and control groups (Supplementary Figure 2). This is an effect that we have commonly observed in previous studies of blood data(17-22), where omics profiles only separated groups in a specific (single molecules and pathways) rather than a global fashion.

Protein-metabolite networks identify potential mediators of COVID-19 pathology

To obtain further mechanistic insight into the biology of COVID-19, we developed a comprehensive, data-driven network for the integrative analysis of our multi-omics dataset. We first generated a Gaussian graphical model (GGM) of correlated metabolites and proteins from our COVID-19 cohort (Supplementary Data 1). GGMs are correlation-based network models that we have previously demonstrated to accurately reconstruct biological pathways from blood-based omics data(23-25). A minimum spanning tree-based algorithm was then used to identify a focused subnetwork that connects the most

significantly correlated metabolites and proteins from the original network (Figure 3). The subnetwork included 13 proteins from the Olink inflammatory panel, 32 proteins from the Olink cardiovascular panels II and III, and 70 metabolites. From this subnetwork, we selected 4 network modules to query the interplay between metabolism, inflammation, and vascular dysfunction in COVID-19 patients.

Inflammation-related network modules. Module 1. Our network identified hexanoylcarnitine as a key metabolite associated with inflammatory cytokines in COVID-19 illness. The module contained IFN-gamma, CXCL10 and CXCL11 that were all upregulated in COVID-19 and associated with hexanoylcarnitine. IFN-gamma, CXCL10 and CXCL11 are proinflammatory cytokines which regulate T cell immunity(26), while hexanoylcarnitine is a medium-chain fatty acid conjugate that plays a critical role in energy metabolism and mitochondrial fatty acid β-oxidation (Figure 3)(27). Levels of other carnitine species (L-carnitine and caprylic acid) were also elevated in COVID-19 patients, suggesting a role for inflammation, dysregulated fatty acid β-oxidation and mitochondrial dysfunction in COVID-19 pathogenesis(12, 28-31).

Module 2. In a second inflammation-related module, cytosine was another key metabolite linked with inflammatory cytokines during COVID-19. The module consisted of a group of macrophage-derived cytokines, including MCP-2, MCP-3 and GRN, which were all upregulated in COVID-19 and positively associated with cytosine (Figure 3). This finding is consistent with the hyperinflammatory state of COVID-19 infection in which cytokine storm is often observed(32). Cytosine, a pyrimidine-class nucleotide that is an essential metabolite for cell proliferation and survival, is commonly upregulated in the host response during viral infection and is furthermore an important mediator of viral replication(33). Prior reports have shown cytosine levels to be elevated in COVID-19 patients(34), which was corroborated in our own cohort. Taken together, cytosine may be a key metabolite linking viral replication to SARS-CoV-2 induced inflammation.

Vascular-related network modules. Module 3. While COVID-19 presents itself mainly as a respiratory disease, autopsy reports have additionally described significant vascular injury due to endothelial cell damage, microcirculatory thrombi, and impaired cellular junction integrity(35). Our network uncovered the coordination of MERTK, RAGE and

thrombomodulin (TM) with the metabolite 4-hydroxyproline in COVID-19 (Figure 3). MERTK, RAGE and TM are proteins key to maintaining features of endothelial homeostasis, and hydroxyproline is a major component of collagen(36). MERTK, a member of the TAM family of receptor tyrosine kinases (RTKs), and RAGE, a pro-coagulant and inflammatory molecule were upregulated, while TM was downregulated, suggesting induction of a pro-coagulant and inflammatory vascular state in COVID-19. Notably, the coordination of these proteins centered around alterations in levels of hydroxyproline, where downregulation of TM directly correlated with hydroxyproline levels. A link between TM and hydroxyproline has been described, where administration of TM had anti-fibrotic effects on lung and kidney murine models(37). Taken together, these data link the pro-coagulant and fibrotic state of COVID-19 through thrombomodulin and hydroxyproline.

Module 4. In a second vascular-related module, we found that cathepsin D (CTSD), a lysosomal protease known to disrupt endothelial cell junctions and increase vasopermeability, was associated with a group of glycolytic metabolites including D-glucose, glycerol-3-phosphate, and lactic acid(38). Notably, increased vascular permeability and glycolysis are known features of a pro-angiogenic state(39). While dysregulated angiogenesis occurs in COVID-19 patients(40), how these events are coordinated remains poorly understood. Here, we report the concomitant up-regulation of cathepsin D and intermediate products of glycolysis (D-glucose and lactic acid) in our COVID-19 cohort, which position this enzyme as a potential driver of the COVID-19 phenotype. Of note, elevated plasma activity of cathepsin D has been found in patients with type 2 diabetes, suggested a link between abnormal vasculature and the dysregulated glucose metabolism seen in our higher risk diabetic COVID-19 patients(41). Taken together these data suggest that the disruption of vascular and glucometabolic homeostasis in COVID-19 is mediated by cathepsin D.

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

Serum metabolites and proteins within the COVID-19 patient group were assessed for correlation with relevant clinical indices including: i) demographics (sex, age, BMI), ii)

concurrent comorbidity (hypertension, pre-existing kidney disease, diabetes mellitus [DM], severity of illness [SOFA]), iii) laboratory markers of inflammation (C-reactive protein [CRP], d-dimer, lymphocyte count, platelet count, ferritin), and iv) future clinical events (ARDS, death) (Figure 4A and 4D). Pre-existing kidney disease was found to associate with the highest number of both metabolites and proteins. Other clinical parameters including ARDS, death, BMI, gender, age, SOFA score, hypertension, DM, and d-dimer levels were associated with numerous metabolites but only few proteins. Conversely, platelet count and ferritin level were associated with a large number of proteins but relatively few metabolites. Detailed association results are provided in Supplementary Tables 5 and 6.

Hierarchical clustering of clinical indices by their correlation with metabolites revealed two broad clusters: one predominantly related to laboratory markers of inflammation, and a second largely related to non-laboratory indices (Figure 4B). Volcano plots of three representative indices (death, pre-existing kidney disease and CRP) are shown in Figure 4C. Seventy-five metabolites were significantly associated with death, 96 with pre-existing kidney disease and 37 with CRP level. 69 metabolites were associated with both death and kidney disease and 20 were associated with death, kidney disease and CRP levels. Key metabolites associated with each of the three selected clinical indices are depicted in Figure 4C. Remarkably, hexanoylcarnitine and cytosine, which we earlier showed to be upregulated in COVID-19, were among the 20 metabolites associated with all three of these clinical indices. This finding supports the potential utility of hexanoylcarnitine and cytosine not only as biomarkers for COVID-19 but also as predictors of disease severity.

Clustering of clinical indices according to their correlation with proteins also revealed two broad clusters (Figure 4E) that were notably similar to the metabolite-derived clusters, except that age and kidney disease clustered together with laboratory markers of inflammation. Volcano plots of three representative indices (death, ferritin level and platelet count) are depicted in Figure 4F. Seventeen proteins correlated with death, 26 with platelet count and 29 with ferritin. Six proteins were associated with both death and ferritin level, 9 with death and platelet count, and 5 (ANGPT1, PDGF subunit A, PDGF

subunit B, LAP TGF-beta-1, CXCL5) with death, ferritin level and platelet count (Figure 4F). ANGPT1, PDGF subunit A and PDGF subunit B are markers of endothelial injury and platelet dysfunction while LAP, TGF-beta-1 and CXCL5 are markers of inflammation. The fact that these proteins correlated with known clinical indices of COVID-19 mortality and morbidity corroborates the importance of vascular injury and inflammation in COVID-19 pathogenesis, which we observed in our network analysis above.

Metabolomics signature predicts clinical outcomes in COVID-19.

We devised a novel hierarchical composite outcome measure for COVID-19 severity which incorporates a series of clinical events, ranked in order of severity, that characterize both acute COVID-19 and some of the sequelae of COVID-19 seen in post-acute COVID-19 syndrome (PACS)(42): in-hospital mortality, mechanical ventilation (MV) at discharge, kidney replacement therapy (KRT) at discharge, prolonged organ failure support (mechanical ventilation and/or kidney replacement therapy for more than 2 weeks), supplemental oxygen requirement, acute kidney injury and length of hospital stay (Figure 5A,). Our measure represents each patient on a continuous spectrum of severity to provide more information than a dichotomous classification such as mortality. Details on the design of this score and patient numbers in each group can be found in Supplementary Figure 3.

A machine learning algorithm based on an ordinal response mixed effect model with LASSO regularization was used to generate our prediction model of the composite outcome measure using serum metabolic profiles and baseline patient demographics (age, sex and BMI) as combined inputs (Figure 5B). Proteins were not included in the model, since our study had around three times more metabolomics samples than proteomics samples. The final model included 32 metabolites and achieved an average concordance-index of d=0.69 (SE=0.017, 95% CI of [0.65, 0.72]), which is equivalent to a receiver operating characteristics area under the curve (ROC-AUC) on a binary outcome. This was a significant improvement (p-value = 3E-5) over a baseline model containing only age, sex, and BMI, which had a performance of d=0.59 (SD=0.02, 95%

CI of [0.56, 0.64]). The 32 metabolites in the final model included 14 amino acids, 6 nucleotides, 6 lipids and 2 metabolites related to energy metabolism (Figure 5C). Interestingly, our metabolite-based model showed improvement over the baseline model not only for predicting the composite outcome, but also for predicting some of its individual components (i.e., intubation, AKI, supplemental oxygen requirement, and length of hospital stay, Figure 5D).

We assessed the tradeoff between the number of metabolites incorporated into the model and the model's prediction performance across a range of included metabolites. Our analysis revealed that most of the predictive power (d ~ 0.68) was already achieved with the first six metabolites (cis-aconitic acid, hydrocinnamic acid, pantothenic acid, 7-methylguanosine, citrulline, methionine sulfoxide), after which prediction performance did not significantly improve (Supplementary Figure 4). This finding suggests that a targeted assay of six metabolites could predict disease severity with this accuracy, although independent validation in other datasets remains to be established.

As a validation step, we tested the performance of our reduced six-metabolite model on blood metabolomics data from two previously published studies by Su et al.(43) and Shen et al.(44) (Figure 5E). The Su study reported COVID-19 severity using a WHO-based score with 7 ordinal levels from mild to severe. Our model achieved a concordance index of d=0.83, again outperforming a model that just consisted of age, sex, and BMI (Figure 5E, left). The Shen study differentiated two groups of COVID-19 patients, mild and severe. That study also developed a prediction model, and we extracted their patient prediction scores from the paper to be able to calculate a concordance index for comparison. The models were tested on two test cohorts, called 'C2' and 'C3' from the original publication (Figure 5E, right). Our 6-metabolite model consistently outperformed both the score from Shen et al. as well as the baseline model with age, sex, and BMI. Notably, overall concordance scores in the Shen dataset were substantially higher than in our own dataset, both as reported in the original paper as well as in the validation of our own model on their data. We believe this effect occurs due to the complex nature of our composite outcome score as opposed to a simple yes vs. no classification of severity, the small sample size in Shen et al., and the reporting of an unvalidated training set AUC

of 0.95 in their study. Overall, this analysis demonstrates that the model replicates in independent datasets with varying definitions of severity, and that our model provides slightly better results compared to the previously reported prediction model.

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.

Our discovered network modules not only provide a better understanding of disease pathogenesis, but also facilitate novel potential therapeutic targets for COVID-19. The modules identified various proteins and metabolites involved in inflammatory and vascular injury processes, such as MMP 12, Cathepsin D and RAGE which, to the best of our knowledge, have not yet been studied as targets for therapeutic intervention in COVID-19. Of note, our modules contained IL-6, which is already a mainstay of treatment for severe disease(50), and several other molecules such as carnitine, niacinamide and IFN-gamma which others have been studying in the context of COVID-19 therapies(51-53).

There is increasing evidence that evaluating symptoms and multiple clinical outcomes during acute disease is crucial in determining the risk of long COVID-19(54). 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. Moreover, we were able to validate the model in two independent studies, thereby demonstrating its generalizability and translational potential.

Several key strengths underlie our study cohort. As opposed to the use of healthy controls reported in other COVID-19 studies(9, 20, 43, 44, 55), our use of non-COVID patient samples, in the same hospital during the same period between March and April 2020, allowed us to investigate the interactions highly relevant to COVID-19 pathogenesis and clinical course. Additionally, we analyzed a relatively larger cohort compared to other studies, with hundreds of samples available for both metabolomic and proteomic analysis.

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

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

Methods

Cohort description

This is a single-center prospective analysis of one cohort comparing hospitalized COVID-19 patients and non-COVID-19 controls. 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. This study has been approved by the Weill Cornell Medicine (WCM) IRB with protocol #19-10020914. Remnant serum samples were matched with selected patients after which patients were deidentified. Controls were randomly selected patients admitted to the hospital with symptoms suspicious for COVID-19, but with negative SARS-CoV-2 RT-PCR. Seventy nine percent of control group patients had shortness of breath, fever, cough or chest pain which are commonly seen in COVID-19. Children (less than 18 years old) and pregnant women (confirmed by a positive beta-HCG test and/or medical records) were excluded.

Sample handling

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

Data Collection

Data were obtained from the Weill Cornell Medicine COVID Institutional Data Repository (COVID-IDR), which is a high-quality registry of COVID-19 patients at NewYork-

Presbyterian - Cornell with laboratory confirmed SARS-CoV-2 RT-PCR. The COVID-IDR houses both manually and automatically extracted Electronic Health Record (EHR) data. Demographics, comorbidities, and important dates of patients' hospital course (admission, intubation, extubation, discharge, death) were extracted by a team of medical professionals and stored in the COVID-IDR. Laboratory tests, ventilation parameters, vital signs, and respiratory variables were additionally available via automated extraction through the Weill Cornell-Critical Care Database for Advanced Research (WC-CEDAR) within the COVID-IDR. WC-CEDAR(56) is a critical care database originally designed to automatically extract, transform, and store EHR data on Intensive Care Unit (ICU) patients; it was expanded to include all hospitalized patients during New York City's COVID-19 surge. Data not available within WC-CEDAR were manually extracted and recorded in REDCap.

Metabolomic profiling

Targeted Metabolite profiling was performed according to a method described in a previous publication (57). For metabolite extraction, 80 µL of pre-chilled methanol (-80 °C) was added to 20 µL of serum. The sample was vortexed for 1 min and then incubated at -80 °C for 2 hours before it was centrifuged at 20,000 g for 15 min at 4 °C to remove the pellet. The supernatant was transferred to a new Eppendorf tube and dried completely with a Speedvac for 30 min (with heat off). The dried sample was redissolved in HPLC grade water before it was applied to the hydrophilic interaction chromatography LC-MS. The sample injection order was randomized.

Metabolites were measured on a Q Exactive Orbitrap mass spectrometer (Thermo Scientific), which was coupled to a Vanquish UPLC system (Thermo Scientific) via an Ion Max ion source with a HESI II probe (Thermo Scientific). A Sequant ZIC-pHILIC column (2.1 mm i.d. × 150 mm, particle size of 5 µm, Millipore Sigma) was used for separation of metabolites. A 2.1 × 20 mm guard column with the same packing material was used for protection of the analytical column. Flow rate was set at 150 µL/min. Buffers consisted of 100% acetonitrile for mobile phase A, and 0.1% NH_4OH/20 mM CH_3COONH_4 in water for mobile phase B. The chromatographic gradient ran from 85% to 30% A in 20 min followed by a wash with 30% A and re-equilibration at 85% A. The column temperature was set to

30 °C and the autosampler temperature was set to 4 °C. The Q Exactive was operated in full scan, polarity-switching mode with the following parameters: the spray voltage 3.0 kV, the heated capillary temperature 300 °C, the HESI probe temperature 350 °C, the sheath gas flow 40 units, the auxiliary gas flow 15 units. MS data acquisition was performed in the m/z range of 70–1,000, with 70,000 resolution (at 200 m/z). The AGC target was 3,000,000 and the maximum injection time was 100 ms. The MS data was processed using XCalibur 4.1 (Thermo Scientific) to extract the metabolite signal intensity for relative quantitation. Metabolites were identified using an in-house library established using chemical standards. Identification required exact mass (within 5ppm) and standard retention times. As a quality control, a mixture of standard compounds was injected thirteen times throughout the LC-MS data acquisition process for monitoring of the stability of LC retention time, the MS mass accuracy and the signal intensity. The median coefficient of variation for metabolite quantitation based on the quality control sample was 0.061. The data from the serum samples showed that both retention time and mass accuracy were highly stable throughout the experiment (Supplementary Figure 5).

Proteomic profiling

Proteomics analysis was performed using the Olink platform (Uppsala, Sweden) at the Proteomics Core of Weill Cornell Medicine-Qatar, according to manufacturer's instructions. We used the Inflammation, Cardiovascular II and Cardiovascular III panels. High throughput real-time PCR of reporter DNA lined to protein specific antibodies was performed on a 96-well integrated fluidic circuits chip (Fluidigm, San Francisco, CA). Each sample was spiked with quality controls to monitor the incubation, extension, and detection steps of the assay. Additionally, samples representing external, negative, and inter-plate controls were included in each analysis run. From raw data, real time PCR cycle threshold (Ct) values were extracted using Fluidigm reverse transcription polymerase chain reaction (RT-PCR) analysis software at a quality threshold of 0.5 and linear baseline correction. Ct values were further processed using the Olink NPX manager software (Olink, Uppsala, Sweden). Here, log2-transformed Ct values from each sample 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(59).

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(60), 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. To avoid issues created by this imbalanced experimental design, 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(61). Overall, this procedure eliminates batch effects and allows for a batch that only contains control samples. Missing values were then imputed using the k-nearest neighbor approach(61).

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:

$$met\_prot \sim outcome + time + (1|patient),$$

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

Data preprocessing and statistical analysis was performed using the “maplet” toolbox for R(62) (https://github.com/krumsieklab/maplet).

Network construction

The dataset was first reduced to the samples that overlap between metabolomics and proteomics (n=227), and corrected for age, sex, BMI, and COVID-19 status (yes/no). A Gaussian Graphical Model (GGM) based network was then constructed using the GeneNet algorithm(63) and drawing an edge for all partial correlations with an FDR smaller than 0.2. In a second step, this network was condensed to highlight the connections between molecules that were significantly different between COVID-19 and controls. To this end, a shortest-path distance matrix between all molecules was constructed and subset to the significant molecules. A minimum spanning tree(64) of this matrix was then constructed to visualize a simplified network.

Composite Outcome

A detailed description of the construction of the composite outcome along with patient numbers in each group can be found in Supplementary Figure 3.

Regularized linear mixed effect ordinal regression model

A new regression model was developed to deal with an ordinal outcome, repeated measurements, and feature selection with regularization. Repeated measurements are handled as a random effect, while age, sex, BMI, and metabolites are treated as fixed effects. The metabolites are penalized using an L_1 LASSO-type regularization to obtain a sparse solution. The model is fitted using an mixed-effect ordinal regression model with complementary log-log link function(65), using maximum likelihood (ML) estimation as proposed by Ripatti and Pamgren(66). An optimal LASSO penalty parameter was estimated through an iterative algorithm for maximizing the Laplace approximation of the integrated model likelihood. This approach was adapted from Therneau(67), which was originally developed for mixed effect Cox models. To obtain an unbiased estimate of the model performance, leave-one-out-cross-validation across the entire dataset was performed. The added value of metabolomics data over baseline clinical data was assessed by comparing the final model with a model only consisting of age, sex, and BMI.

Validation datasets

Metabolomics data were downloaded from Su et al.(43) (n=121) and Shen et al. (44) (containing two validation sets, n=10 and n=19). The Su dataset contained all six metabolites from our reduced model as well as age, sex and BMI as baseline parameters. The Shen study also covered age, sex and BMI, and the first test dataset (“C2”) contained all six metabolites. The second test dataset (“C3”) was measured using a targeted assay of only 7 metabolites and 22 proteins, and the metabolites did not overlap with our model metabolites. Thus, we had to follow a more complex procedure for validation. We first applied our risk score in their training cohort, which contained all metabolites. In the training cohort, this score was then regressed on the available measurements in C3, i.e. modeling ‘score ~ metabolite1 + … + metabolite7 + protein1 + …+ protein22’. The coefficients from this model were then used in C3 to derive a surrogate severity score, which we evaluated in Figure 5.

Data Availability

The data used in this study can be downloaded at https://doi.org/10.6084/m9.figshare.19115972.v1.

Code Availability

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

Funding

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

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.

570 10. Wiersinga WJ, Rhodes A, Cheng AC, Peacock SJ, Prescott HC. Pathophysiology, Transmission,
571 Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA. 2020;324(8):782-
572 93.

573 11. Wu P, Chen D, Ding W, Wu P, Hou H, Bai Y, et al. The trans-omics landscape of COVID-19.
574 Nature Communications. 2021;12(1):4543.

575 12. Thomas T, Stefanoni D, Reisz JA, Nemkov T, Bertolone L, Francis RO, et al. COVID-19 infection
576 alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. JCI Insight.
577 2020;5(14).

578 13. Barberis E, Timo S, Amede E, Vanella VV, Puricelli C, Cappellano G, et al. Large-Scale Plasma
579 Analysis Revealed New Mechanisms and Molecules Associated with the Host Response to SARS-CoV-2.
580 International Journal of Molecular Sciences. 2020;21(22):8623.

581 14. Törnquist K, Asghar MY, Srinivasan V, Korhonen L, Lindholm D. Sphingolipids as Modulators of
582 SARS-CoV-2 Infection. Frontiers in Cell and Developmental Biology. 2021;9.

583 15. Kalil AC, Patterson TF, Mehta AK, Tomashek KM, Wolfe CR, Ghazaryan V, et al. Baricitinib plus
584 Remdesivir for Hospitalized Adults with Covid-19. New England Journal of Medicine. 2020;384(9):795-
585 807.

586 16. Iastrebner M, Castro J, Garcia Espina E, Lettieri C, Payaslian S, Cuesta MC, et al. Ruxolitinib in
587 severe COVID-19: Results of a multicenter, prospective, single arm, open-label clinical study to
588 investigate the efficacy and safety of ruxolitinib in patients with COVID-19 and severe acute respiratory
589 syndrome. Rev Fac Cien Med Univ Nac Cordoba. 2021;78(3):294-302.

590 17. Roberts I, Wright Muelas M, Taylor JM, Davison AS, Xu Y, Grixti JM, et al. Untargeted
591 metabolomics of COVID-19 patient serum reveals potential prognostic markers of both severity and
592 outcome. Metabolomics. 2021;18(1):6.

593 18. Valdés A, Moreno LO, Rello SR, Orduña A, Bernardo D, Cifuentes A. Metabolomics study of
594 COVID-19 patients in four different clinical stages. Scientific Reports. 2022;12(1):1650.

595 19. Meoni G, Ghini V, Maggi L, Vignoli A, Mazzoni A, Salvati L, et al. Metabolomic/lipidomic profiling
596 of COVID-19 and individual response to tocilizumab. PLOS Pathogens. 2021;17(2):e1009243.

597 20. Overmyer KA, Shishkova E, Miller IJ, Balnis J, Bernstein MN, Peters-Clarke TM, et al. Large-
598 Scale Multi-omic Analysis of COVID-19 Severity. Cell Systems. 2021;12(1):23-40.e7.

599 21. Di B, Jia H, Luo OJ, Lin F, Li K, Zhang Y, et al. Identification and validation of predictive factors
600 for progression to severe COVID-19 pneumonia by proteomics. Signal Transduction and Targeted
601 Therapy. 2020;5(1):217.

602 22. Gisby J, Clarke CL, Medjeral-Thomas N, Malik TH, Papadaki A, Mortimer PM, et al. Longitudinal
603 proteomic profiling of dialysis patients with COVID-19 reveals markers of severity and predictors of death.
604 Elife. 2021;10:e64827.

605 23. Krumsiek J, Suhre K, Illig T, Adamski J, Theis FJ. Gaussian graphical modeling reconstructs
606 pathway reactions from high-throughput metabolomics data. BMC Syst Biol. 2011;5:21.

607 24. Krumsiek J, Suhre K, Evans AM, Mitchell MW, Mohney RP, Milburn MV, et al. Mining the
608 unknown: a systems approach to metabolite identification combining genetic and metabolic information.
609 PLoS Genet. 2012;8(10):e1003005.

610 25. Benedetti E, Pučić-Baković M, Keser T, Wahl A, Hassinen A, Yang JY, et al. Network inference
611 from glycoproteomics data reveals new reactions in the IgG glycosylation pathway. Nat Commun.
612 2017;8(1):1483.

613 26. Tokunaga R, Zhang W, Naseem M, Puccini A, Berger MD, Soni S, et al. CXCL9, CXCL10,
614 CXCL11/CXCR3 axis for immune activation - A target for novel cancer therapy. Cancer Treat Rev.
615 2018;63:40-7.

27. 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.

28. Wajner M, Amaral AU. Mitochondrial dysfunction in fatty acid oxidation disorders: insights from human and animal studies. Biosci Rep. 2015;36(1):e00281.

29. Vella F. The metabolic and molecular bases of inherited disease seventh edition: Edited by C R Sriver, A L Beaudet, W S Sly and D Valle. P 4605. McGraw-Hill, New York. 1995. ISBN 0-07-909826-6. Biochemical Education. 1996;24(1):65-.

30. León G, Herrera M, Vargas M, Arguedas M, Sánchez A, Segura Á, et al. Development and characterization of two equine formulations towards SARS-CoV-2 proteins for the potential treatment of COVID-19. Sci Rep. 2021;11(1):9825.

31. Soliman S, Faris ME, Ratemi Z, Halwani R. Switching Host Metabolism as an Approach to Dampen SARS-CoV-2 Infection. Ann Nutr Metab. 2020;76(5):297-303.

32. 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.

33. Hoelzer K, Shackelton LA, Parrish CR. Presence and role of cytosine methylation in DNA viruses of animals. Nucleic Acids Res. 2008;36(9):2825-37.

34. 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.

35. 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.

36. Xu S, Gu M, Wu K, Li G. Unraveling the Role of Hydroxyproline in Maintaining the Thermal Stability of the Collagen Triple Helix Structure Using Simulation. The Journal of Physical Chemistry B. 2019;123(36):7754-63.

37. Takeshita A, Yasuma T, Nishihama K, D'Alessandro-Gabazza CN, Toda M, Totoki T, et al. Thrombomodulin ameliorates transforming growth factor-β1–mediated chronic kidney disease via the G-protein coupled receptor 15/Akt signal pathway. Kidney International. 2020;98(5):1179-92.

38. Monickaraj F, McGuire P, Das A. Cathepsin D plays a role in endothelial–pericyte interactions during alteration of the blood–retinal barrier in diabetic retinopathy. The FASEB Journal. 2018;32(5):2539-48.

39. Leung SWS, Shi Y. The glycolytic process in endothelial cells and its implications. Acta Pharmacologica Sinica. 2021.

40. Ackermann M, Verleden SE, Kuehnel M, Haverich A, Welte T, Laenger F, et al. Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. New England Journal of Medicine. 2020;383(2):120-8.

41. Chen J, Wu C, Wang X, Yu J, Sun Z. The Impact of COVID-19 on Blood Glucose: A Systematic Review and Meta-Analysis. Frontiers in Endocrinology. 2020;11(732).

42. Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS, et al. Post-acute COVID-19 syndrome. Nature Medicine. 2021;27(4):601-15.

43. Su Y, Chen D, Yuan D, Lausted C, Choi J, Dai CL, et al. Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19. Cell. 2020;183(6):1479-95.e20.

44. Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, et al. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. Cell. 2020;182(1):59-72 e15.

660 45. Senn S, editor Dichotomania: An Obsessive Compulsive Disorder that is Badly Affecting the
661 Quality of Analysis of Pharmaceutical Trials2005.

662 46. Roozenbeek B, Lingsma HF, Perel P, Edwards P, Roberts I, Murray GD, et al. The added value
663 of ordinal analysis in clinical trials: an example in traumatic brain injury. Crit Care. 2011;15(3):R127.

664 47. Song JW, Lam SM, Fan X, Cao WJ, Wang SY, Tian H, et al. Omics-Driven Systems Interrogation
665 of Metabolic Dysregulation in COVID-19 Pathogenesis. Cell Metab. 2020;32(2):188-202.e5.

666 48. Danlos FX, Grajeda-Iglesias C, Durand S, Sauvat A, Roumier M, Cantin D, et al. Metabolomic
667 analyses of COVID-19 patients unravel stage-dependent and prognostic biomarkers. Cell Death Dis.
668 2021;12(3):258.

669 49. López-Hernández Y, Monárrez-Espino J, Oostdam AH, Delgado JEC, Zhang L, Zheng J, et al.
670 Targeted metabolomics identifies high performing diagnostic and prognostic biomarkers for COVID-19.
671 Sci Rep. 2021;11(1):14732.

672 50. Interleukin-6 Receptor Antagonists in Critically Ill Patients with Covid-19. New England Journal of
673 Medicine. 2021;384(16):1491-502.

674 51. Li C, Ou R, Wei Q, Shang H. Carnitine and COVID-19 Susceptibility and Severity: A Mendelian
675 Randomization Study. Frontiers in Nutrition. 2021;8.

676 52. van Laarhoven A, Kurver L, Overheul GJ, Kooistra EJ, Abdo WF, van Crevel R, et al. Interferon
677 gamma immunotherapy in five critically ill COVID-19 patients with impaired cellular immunity: A case
678 series. Med. 2021;2(10):1163-70.e2.

679 53. Fu W, Lei C, Ma Z, Qian K, Li T, Zhao J, et al. CAR Macrophages for SARS-CoV-2
680 Immunotherapy. Frontiers in Immunology. 2021;12.

681 54. Sudre CH, Murray B, Varsavsky T, Graham MS, Penfold RS, Bowyer RC, et al. Attributes and
682 predictors of long COVID. Nature Medicine. 2021;27(4):626-31.

683 55. Arunachalam PS, Wimmers F, Mok CKP, Perera RAPM, Scott M, Hagan T, et al. Systems
684 biological assessment of immunity to mild versus severe COVID-19 infection in humans. Science.
685 2020;369(6508):1210-20.

686 56. Schenck EJ, Hoffman KL, Cusick M, Kabariti J, Sholle ET, Campion TR. Critical carE Database
687 for Advanced Research (CEDAR): An Automated Method to Support Intensive Care Units with Electronic
688 Health Record Data. J Biomed Inform. 2021:103789.

689 57. Chen WW, Freinkman E, Wang T, Birsoy K, Sabatini DM. Absolute Quantification of Matrix
690 Metabolites Reveals the Dynamics of Mitochondrial Metabolism. Cell. 2016;166(5):1324-37 e11.

691 58. Goncalves MD, Hwang SK, Pauli C, Murphy CJ, Cheng Z, Hopkins BD, et al. Fenofibrate
692 prevents skeletal muscle loss in mice with lung cancer. Proc Natl Acad Sci U S A. 2018;115(4):E743-e52.

693 59. Kanehisa M, Furumichi M, Sato Y, Ishiguro-Watanabe M, Tanabe M. KEGG: integrating viruses
694 and cellular organisms. Nucleic Acids Res. 2021;49(D1):D545-d51.

695 60. Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust
696 method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Anal
697 Chem. 2006;78(13):4281-90.

698 61. Do KT, Wahl S, Raffler J, Molnos S, Laimighofer M, Adamski J, et al. Characterization of missing
699 values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies.
700 Metabolomics. 2018;14(10):128.

701 62. Chetnik K, Benedetti E, Gomari DP, Schweickart A, Batra R, Buyukozkan M, et al. maplet: An
702 extensible R toolbox for modular and reproducible omics pipelines2021 May 01, 2021:[arXiv:2105.04305
703 p.]. Available from: https://ui.adsabs.harvard.edu/abs/2021arXiv210504305C.

63. 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.

64. Prim RC. Shortest Connection Networks And Some Generalizations. Bell System Technical Journal. 1957;36(6):1389-401.

65. McCullagh P. Regression Models for Ordinal Data. Journal of the Royal Statistical Society Series B (Methodological). 1980;42(2):109-42.

66. Ripatti S, Palmgren J. Estimation of multivariate frailty models using penalized partial likelihood. Biometrics. 2000;56(4):1016-22.

67. Therneau T. Coxme and the Laplace approximation 2020 [Available from: https://cran.r-project.org/web/packages/coxme/vignettes/laplace.pdf.

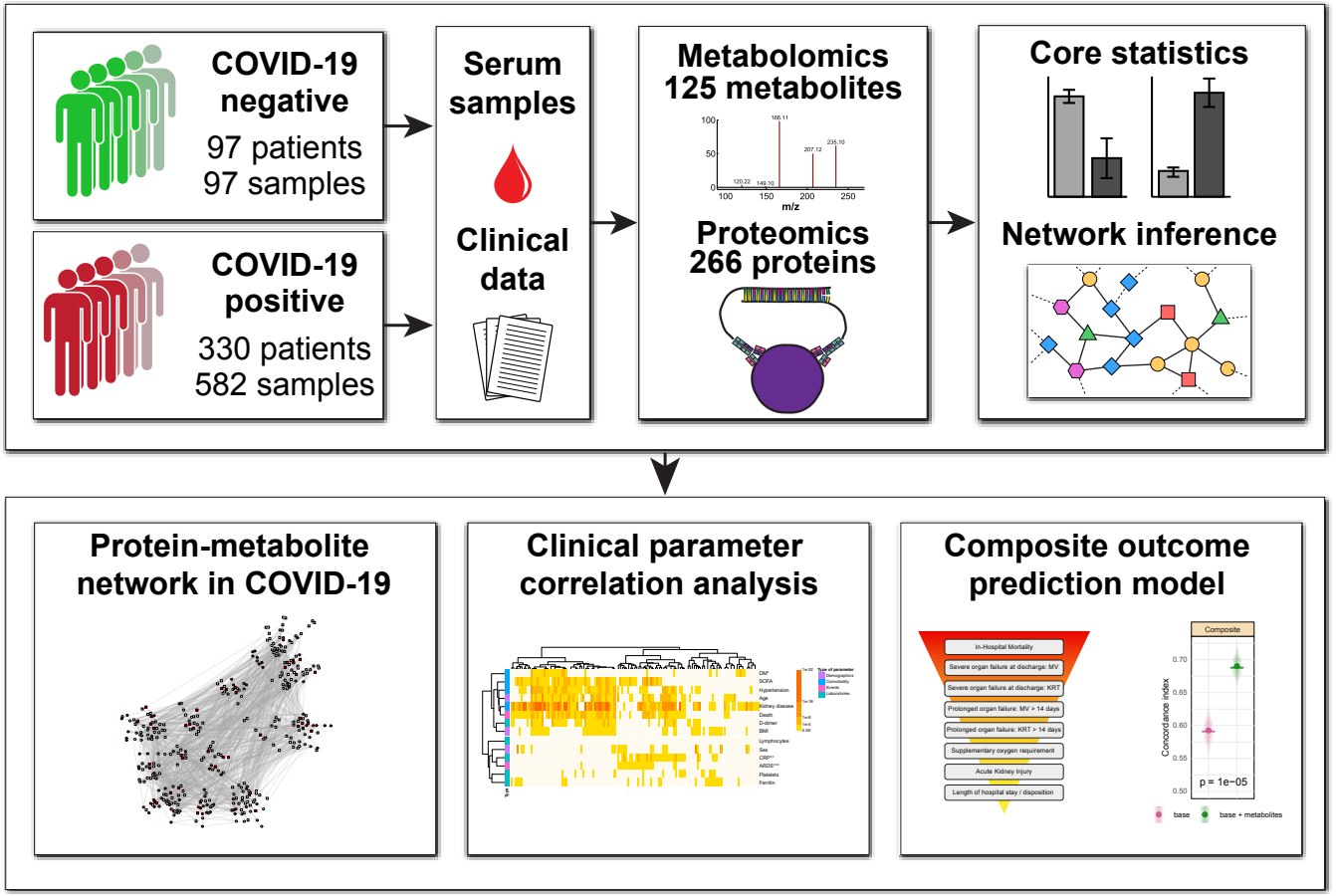
Figure Legends

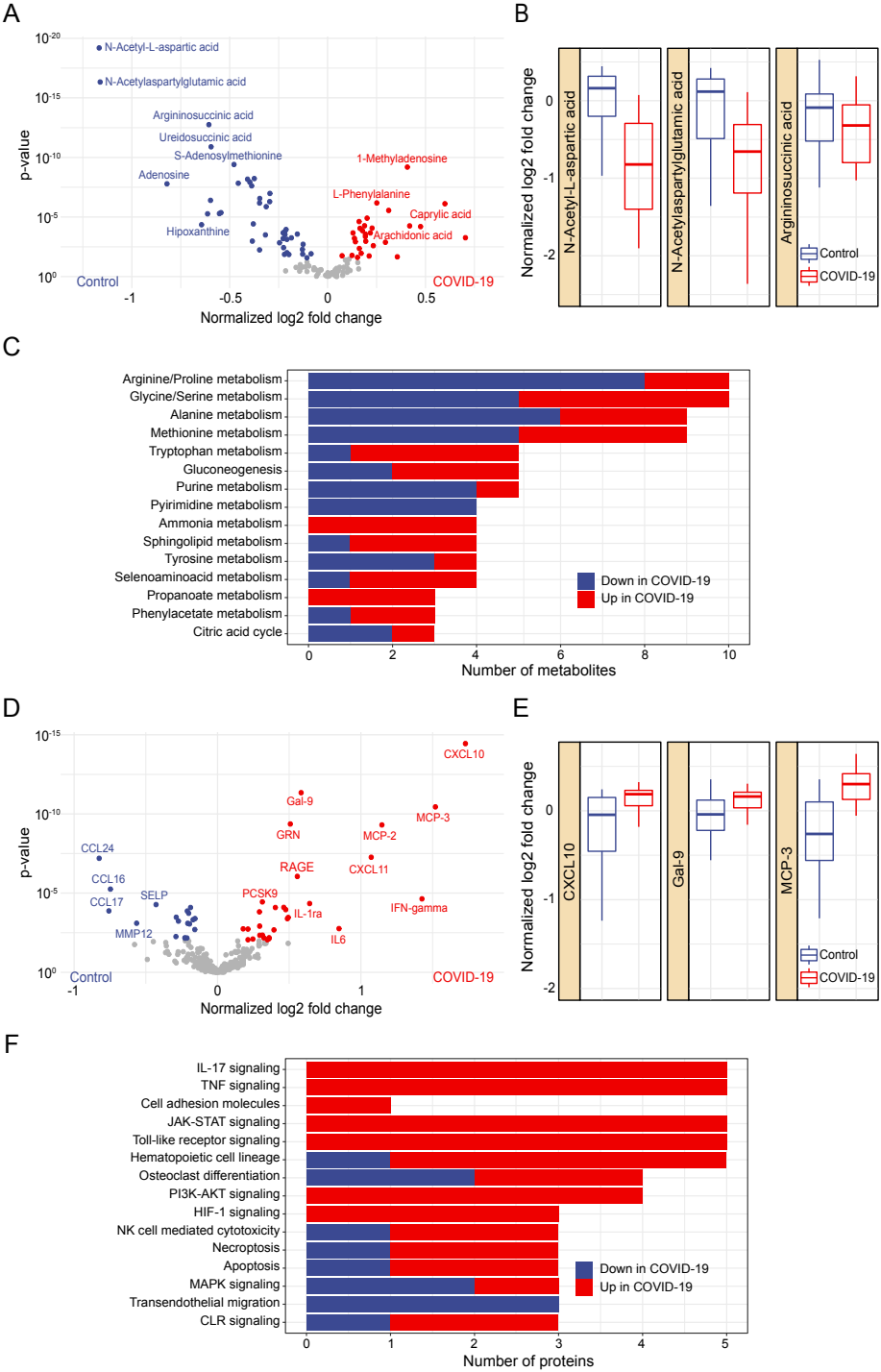
Fig. 1. Study outline. Study population composed by COVID-19 patients and controls. Serum samples obtained from clinically indicated specimens during the first 72 hours of admission. Metadata obtained from electronic medical records. We obtained metabolomic and proteomic profile, after which we integrated 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.

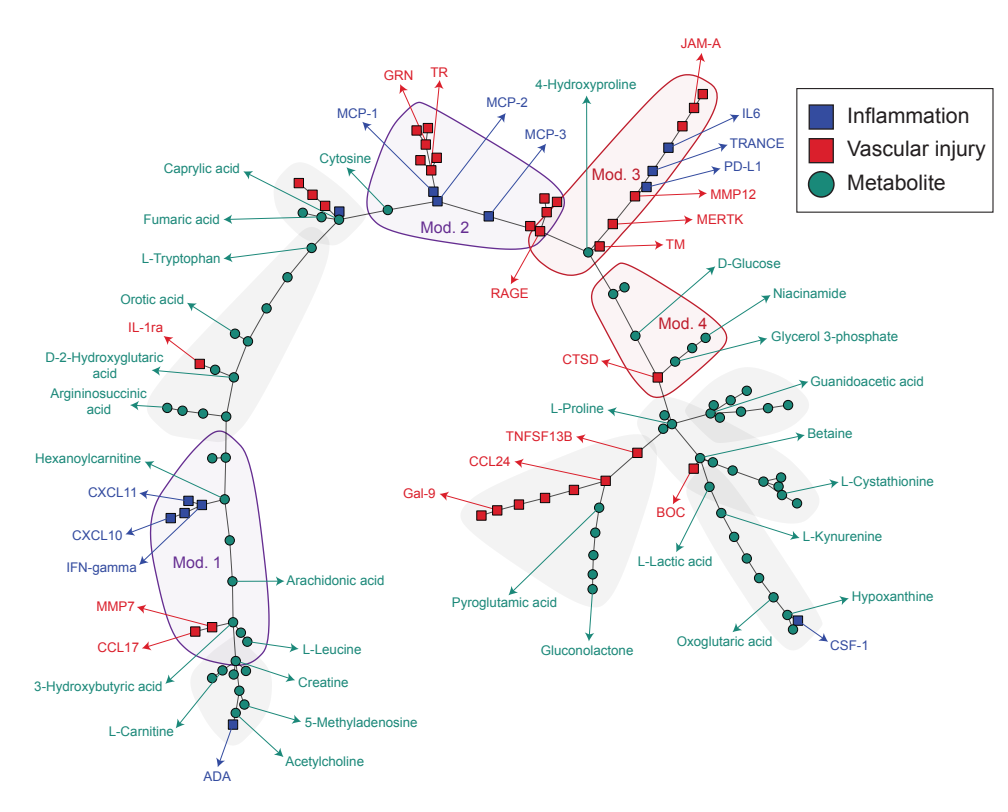
Fig. 2. Metabolomics and Proteomics changes associated with COVID-19. A, Volcano plot showing differentially expressed metabolites between COVID-19 patients and controls at an adjusted p-value <0.05. In red, upregulated metabolites in COVID-19 patients. In blue, upregulated metabolites in the control group. **B,** Top 3 differentially expressed metabolites in COVID-19 patients vs. controls based on adjusted p-values. Y axis shows $\log_2$ fold changes in relation to the mean of the control group. **C,** Number of significantly regulated molecules in KEGG pathways, top 15 shown. **D,** Volcano plot showing the differentially expressed proteins between COVID-19 patients and controls at an adjusted p-value <0.05. **E,** Top 3 differentially expressed proteins in COVID-19 patients vs. controls based on adjusted p-values. Y axis shows $\log_2$

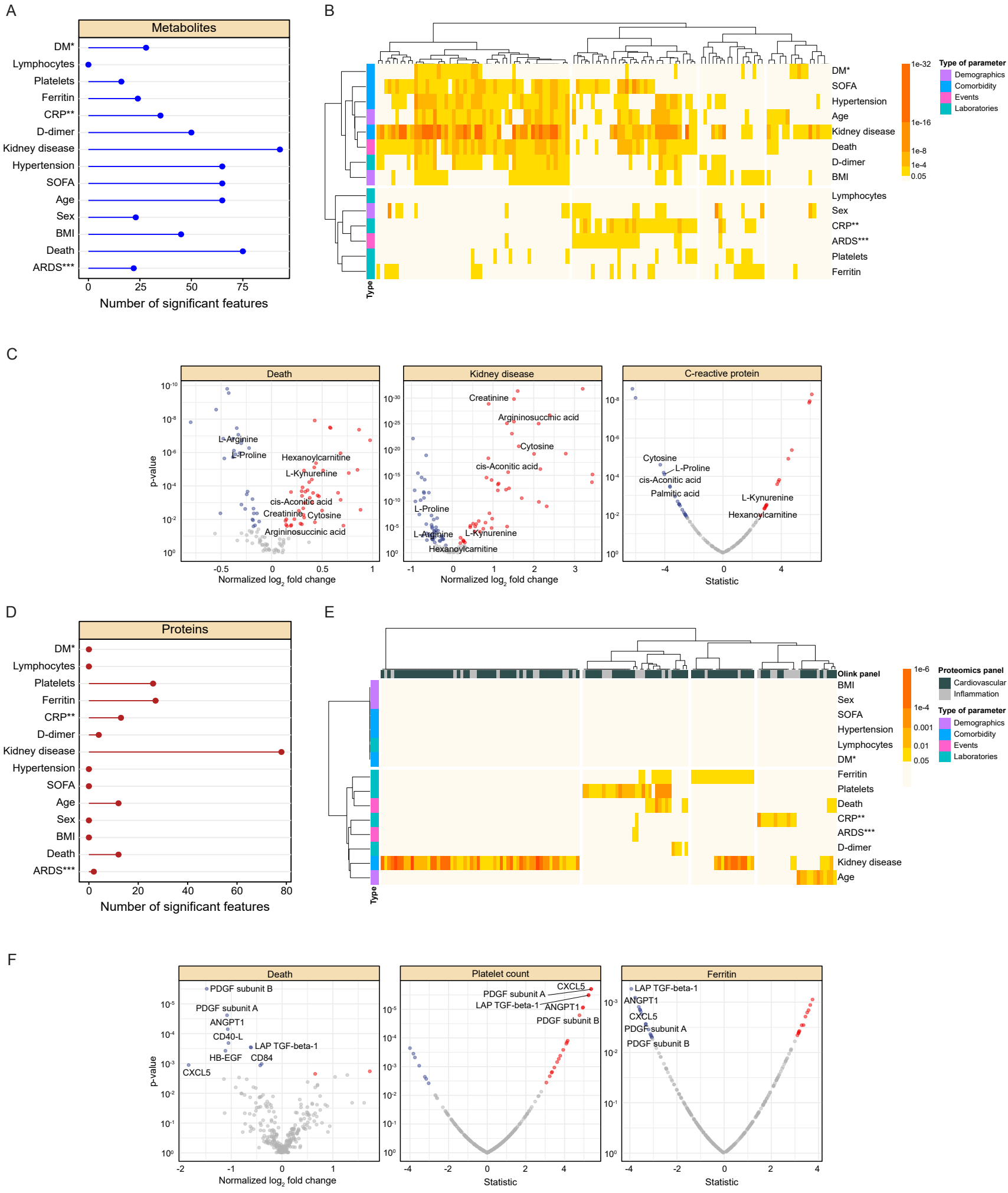
738 fold changes in relation to the mean of the control group. **F,** Number of significantly
739 regulated molecules in KEGG pathways, top 15 shown.
740
741 **Fig. 3. Protein-metabolite networks in COVID-19.** Gaussian graphical model (GGM)
742 representing the significant partial correlations between all the measured metabolites and
743 proteins, generated by a minimum spanning tree (MST) based reduction combined with
744 a shortest paths-based approach. The network includes all differentially expressed
745 metabolites and proteins between COVID-19 patients and controls. Squares indicate
746 proteins and green circles indicate metabolites. Inflammatory proteins are colored in
747 purple, while vascular injury proteins are colored in red. Shadows represent molecule
748 modules.
749
750 **Fig. 4. Associations of molecules with clinical indices in COVID-19. A,** Lollipop
751 plot representing the number of altered metabolites for each analyzed parameter
752 (*Diabetes mellitus, **C-reactive protein, ***Acute respiratory distress syndrome,
753 represented by whether patients were intubated). **B,** Heatmap indicating the number
754 of differentially expressed metabolites associated with each clinical parameter and its
755 significance. We included 4 clinical parameter categories: demographics, comorbidities,
756 clinical events, and laboratory parameters. **C,** Volcano plot showing detailed metabolic
757 changes correlated to death, kidney disease and C-reactive protein. **D,** Lollipop plot
758 representing the number of altered proteins for each analyzed parameter. **E,** Heatmap
759 indicating the number of differentially expressed proteins on each clinical parameter and
760 its significance, using the same 4 categories described above. Proteins are marked
761 according to Olink panels. F, Volcano plot showing detailed protein changes correlated
762 to death, platelet count and ferritin levels.
763
764 **Fig. 5. Metabolomics signature predicts clinical outcomes of COVID-19. A,** Scheme
765 of the hierarchical composite outcome ranging from in-hospital mortality to length of
766 hospital stay and disposition (MV: Mechanical ventilation, KRT: Kidney replacement
767 therapy). **B,** Dot plot comparing the predictive performance (Concordance index) of a
768 baseline model using age, sex, and BMI and a model including baseline plus

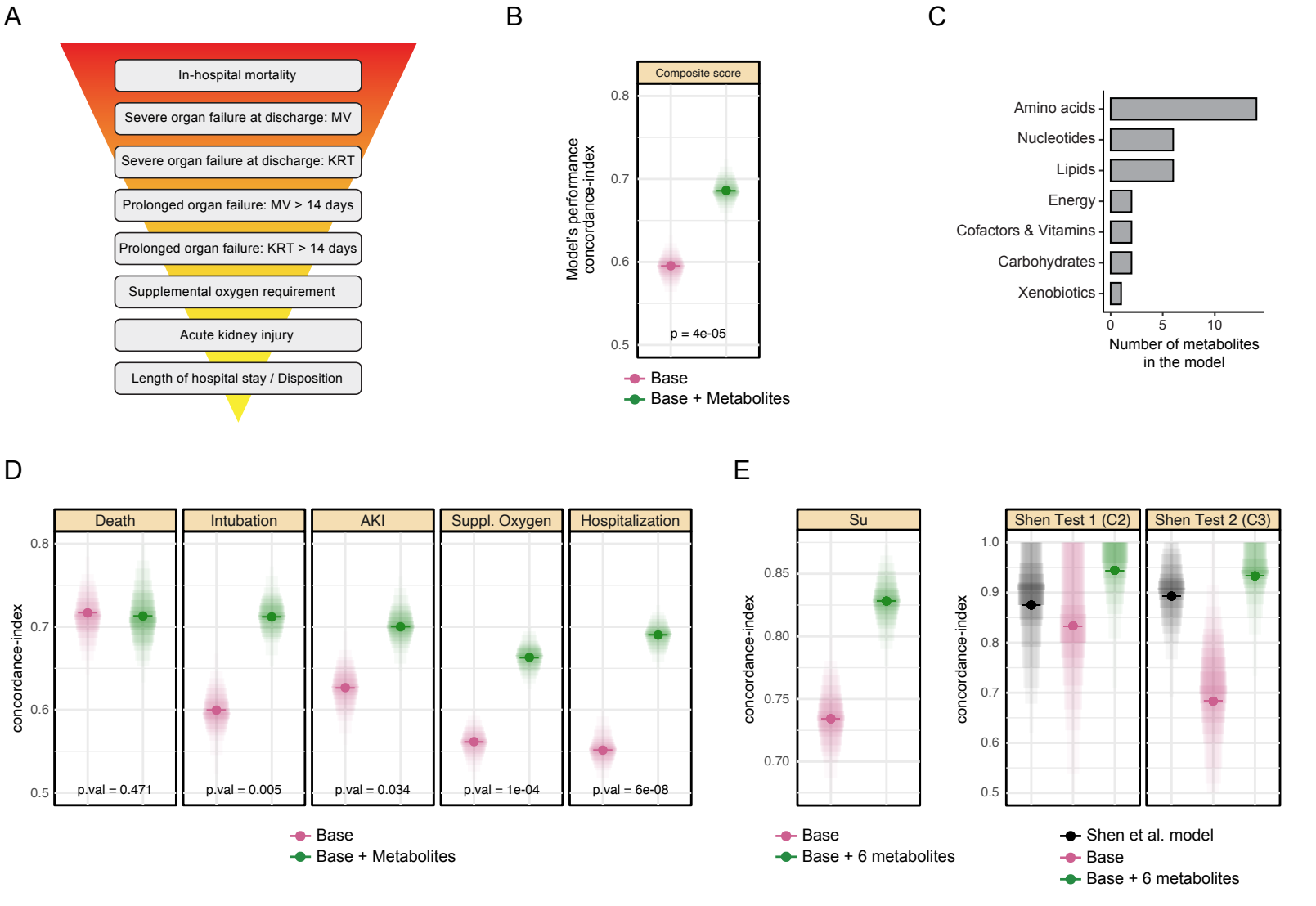
metabolites. The results show that the prediction accuracy improves when adding metabolites to the model. **C,** Bar plot showing the main altered metabolic groups included in the composite outcome score. **D,** Dot plots showing individual outcome analysis, comparing the baseline model and baseline plus metabolites added. Metabolites improve the prediction accuracy of ARDS, AKI, supplemental oxygen requirement and prolonged hospitalization. **E**, Replication analysis of our model in two independent datasets by Su et al.(43) and Shen et al.(44). This analysis was performed on the reduced 6-metabolite model. The Shen et al. study developed their own model, which is plotted in black.(44)


COVID-19 negative
97 patients
97 samples
COVID-19 positive
330 patients
582 samples
Serum samples
Clinical data
Metabolomics
125 metabolites
m/z
Proteomics
266 proteins
Core statistics
Network inference
Protein-metabolite network in COVID-19
Clinical parameter correlation analysis
Composite outcome prediction model
In-Hospital Mortality
Severe organ failure at discharge: MV
Severe organ failure at discharge: KRT
Prolonged organ failure: MV > 14 days
Prolonged organ failure: KRT > 14 days
Supplementary oxygen requirement
Acute Kidney Injury
Length of hospital stay / disposition
Composite
Concordance index
p = 1e−05
base
base + metabolites


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
SELP
PCSK9
CCL17
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


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




A
In-hospital mortality
Severe organ failure at discharge: MV
Severe organ failure at discharge: KRT
Prolonged organ failure: MV > 14 days
Prolonged organ failure: KRT > 14 days
Supplemental oxygen requirement
Acute kidney injury
Length of hospital stay / Disposition
B
Composite score
Model's performance concordance-index
0.8
0.7
0.6
0.5
p = 4e-05
Base
Base + Metabolites
C
Amino acids
Nucleotides
Lipids
Energy
Cofactors & Vitamins
Carbohydrates
Xenobiotics
0
5
10
Number of metabolites in the model
D
Death
Intubation
AKI
Suppl. Oxygen
Hospitalization
concordance-index
0.8
0.7
0.6
0.5
p.val = 0.471
p.val = 0.005
p.val = 0.034
p.val = 1e-04
p.val = 6e-08
Base
Base + Metabolites
E
Su
concordance-index
0.85
0.80
0.75
0.70
Base
Base + 6 metabolites
Shen Test 1 (C2)
Shen Test 2 (C3)
concordance-index
1.0
0.9
0.8
0.7
0.6
0.5
Shen et al. model
Base
Base + 6 metabolites