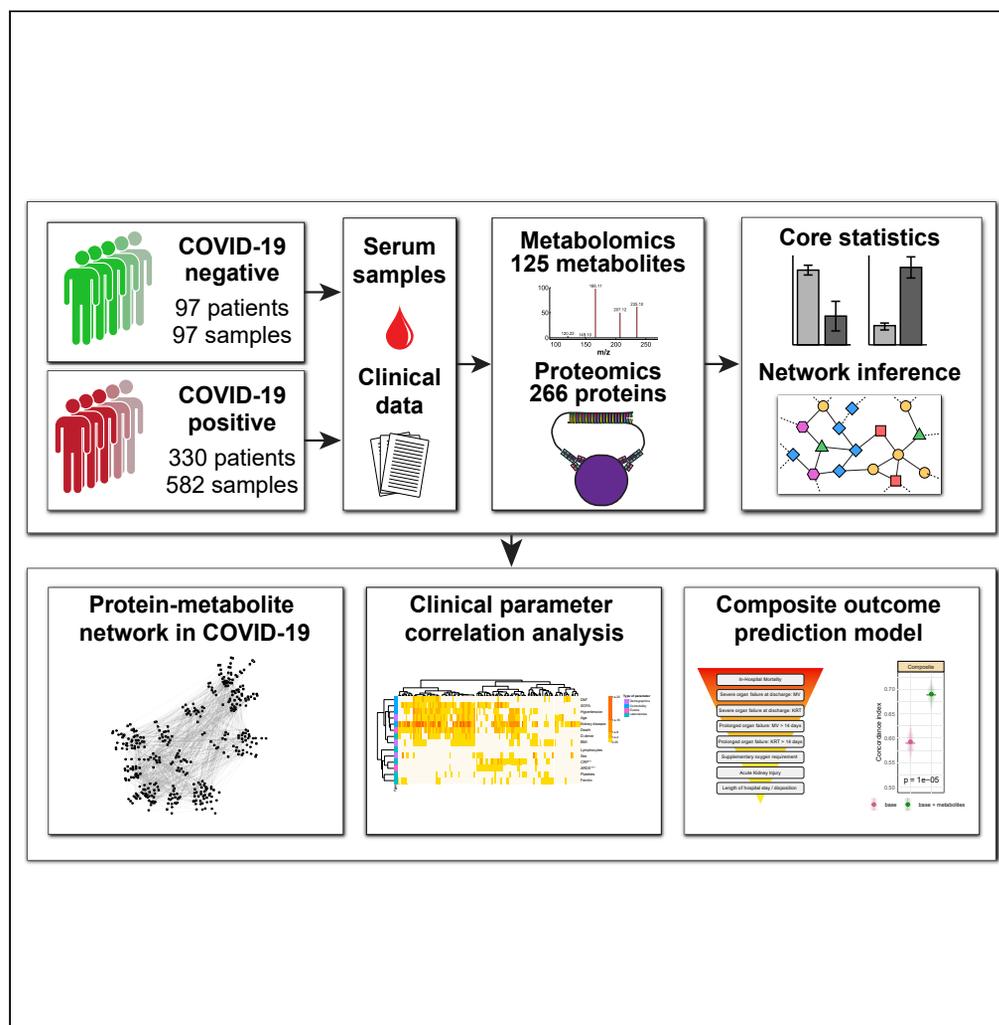


Article

# Integrative metabolomic and proteomic signatures define clinical outcomes in severe COVID-19



Mustafa Buyukozkan, Sergio Alvarez-Mulett, Alexandra C. Racanelli, ..., Augustine M.K. Choi, Soo Jung Cho, Jan Krumsiek

sjc9006@med.cornell.edu (S.J.C.)  
jak2042@med.cornell.edu (J.K.)

Highlights

COVID-19 patients show serum changes in various pathways

Metabolomic/proteomic cross-talk defines pathways of disease

COVID-19 disease severity correlates with various markers in blood

COVID-19 disease severity can be accurately predicted by a small set of metabolites

Buyukozkan et al., iScience 25, 104612  
July 15, 2022 © 2022 The Authors.  
<https://doi.org/10.1016/j.isci.2022.104612>



## Article

## Integrative metabolomic and proteomic signatures define clinical outcomes in severe COVID-19

Mustafa Buyukozkan,<sup>1,2,11</sup> Sergio Alvarez-Mulett,<sup>3,11</sup> Alexandra C. Racanelli,<sup>3,11</sup> Frank Schmidt,<sup>4</sup> Richa Batra,<sup>1,2</sup> Katherine L. Hoffman,<sup>5</sup> Hina Sarwath,<sup>4</sup> Rudolf Engelke,<sup>4</sup> Luis Gomez-Escobar,<sup>3</sup> Will Simmons,<sup>5</sup> Elisa Benedetti,<sup>1,2</sup> Kelsey Chetnik,<sup>1,2</sup> Guoan Zhang,<sup>6</sup> Edward Schenck,<sup>3</sup> Karsten Suhre,<sup>7</sup> Justin J. Choi,<sup>8</sup> Zhen Zhao,<sup>9</sup> Sabrina Racine-Brzostek,<sup>9</sup> He S. Yang,<sup>9</sup> Mary E. Choi,<sup>10</sup> Augustine M.K. Choi,<sup>3</sup> Soo Jung Cho,<sup>3,\*</sup> and Jan Krumsiek<sup>1,2,12,\*</sup>

## SUMMARY

**The coronavirus disease-19 (COVID-19) pandemic has ravaged global healthcare with previously unseen levels of morbidity and mortality. 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 network of protein-metabolite interactions through targeted metabolomic and proteomic profiling 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. Finally, we developed a novel composite outcome measure for COVID-19 disease severity based on metabolomics data. The model predicts severe disease with a concordance index of around 0.69, and shows high predictive power of 0.83–0.93 in two 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) (Guan et al., 2020; Arons et al., 2020). COVID-19 ARDS can lead to refractory hypoxia, mechanical ventilation, prolonged intensive care unit (ICU) stay and increased mortality (Zhou et al., 2020). Previous studies have shown a high incidence of concomitant organ failure in COVID-19, including acute kidney injury (AKI) (Hirsch et al., 2020), acute liver injury (Zhang et al., 2020), thromboembolic events (Bilaloglu et al., 2020; Tang et al., 2020) and secondary infections contributing to a fatal outcome (Elezkurtaj et al., 2021).

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 (Lucas et al., 2020). 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

<sup>1</sup>Department of Physiology and Biophysics, Weill Cornell Medicine, New York, NY, USA

<sup>2</sup>Meyer Cancer Center and Caryl and Israel Englander Institute for Precision Medicine, Weill Cornell Medicine, New York, NY, USA

<sup>3</sup>Department of Medicine, Division of Pulmonary and Critical Care Medicine, Weill Cornell Medicine, New York, NY, USA

<sup>4</sup>Proteomics Core, Weill Cornell Medicine – Qatar, Doha, Qatar

<sup>5</sup>Department of Population Health Sciences, Division of Biostatistics, Weill Cornell Medicine, New York, NY, USA

<sup>6</sup>Proteomics and Metabolomics Core Facility, Weill Cornell Medicine, New York, NY, USA

<sup>7</sup>Department of Physiology and Biophysics, Weill Cornell Medicine – Qatar, Education City, Doha 24144, Qatar

<sup>8</sup>Department of Medicine, Division of General Internal Medicine, Weill Cornell Medicine, New York, NY, USA

<sup>9</sup>Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY, USA

<sup>10</sup>Division of Nephrology and Hypertension, Joan and Sanford I. Weill Department of Medicine, New York, NY, USA

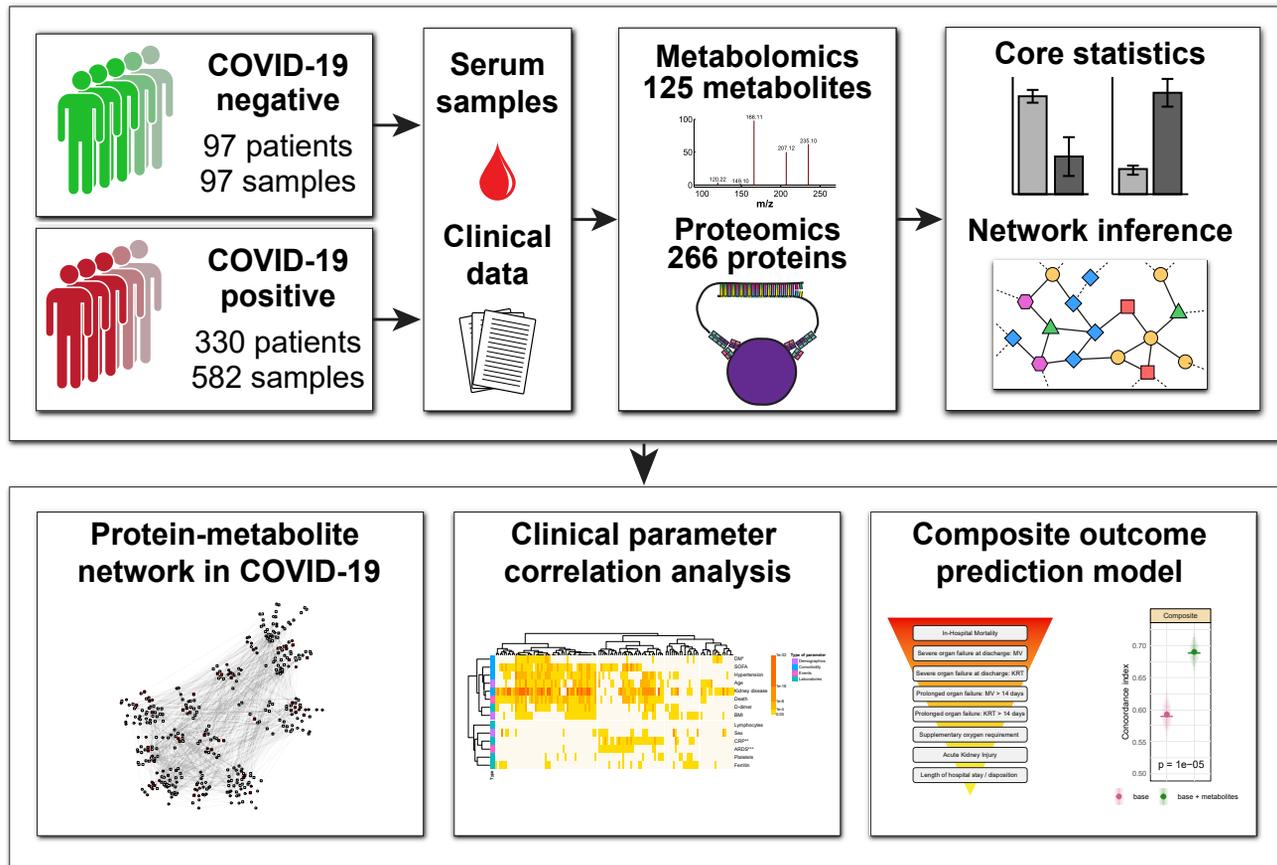
<sup>11</sup>These authors contributed equally

<sup>12</sup>Lead contact

\*Correspondence: sjc9006@med.cornell.edu (S.J.C.), jak2042@med.cornell.edu (J.K.)

<https://doi.org/10.1016/j.isci.2022.104612>





**Figure 1. Study outline. Study population composed by COVID-19 patients and controls**

Serum samples obtained from clinically indicated specimens during the first 72 h of admission. Metadata obtained from electronic medical records. We obtained metabolomic and proteomic profiles, 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.

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 composed 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, whereas 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 (Figure S1), and the repeated samples were thus treated as replicates using a linear mixed effect model, see STAR Methods. A detailed description of the clinical and demographic characteristics of the cohort can be found in Table 1, Tables S1 and S2. Of note, we excluded samples collected after intubation because we found that the clinical act of intubation significantly alters a patient's metabolic profile (Table S3).

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

**Table 1. Demographics and clinical characteristics of the cohort**

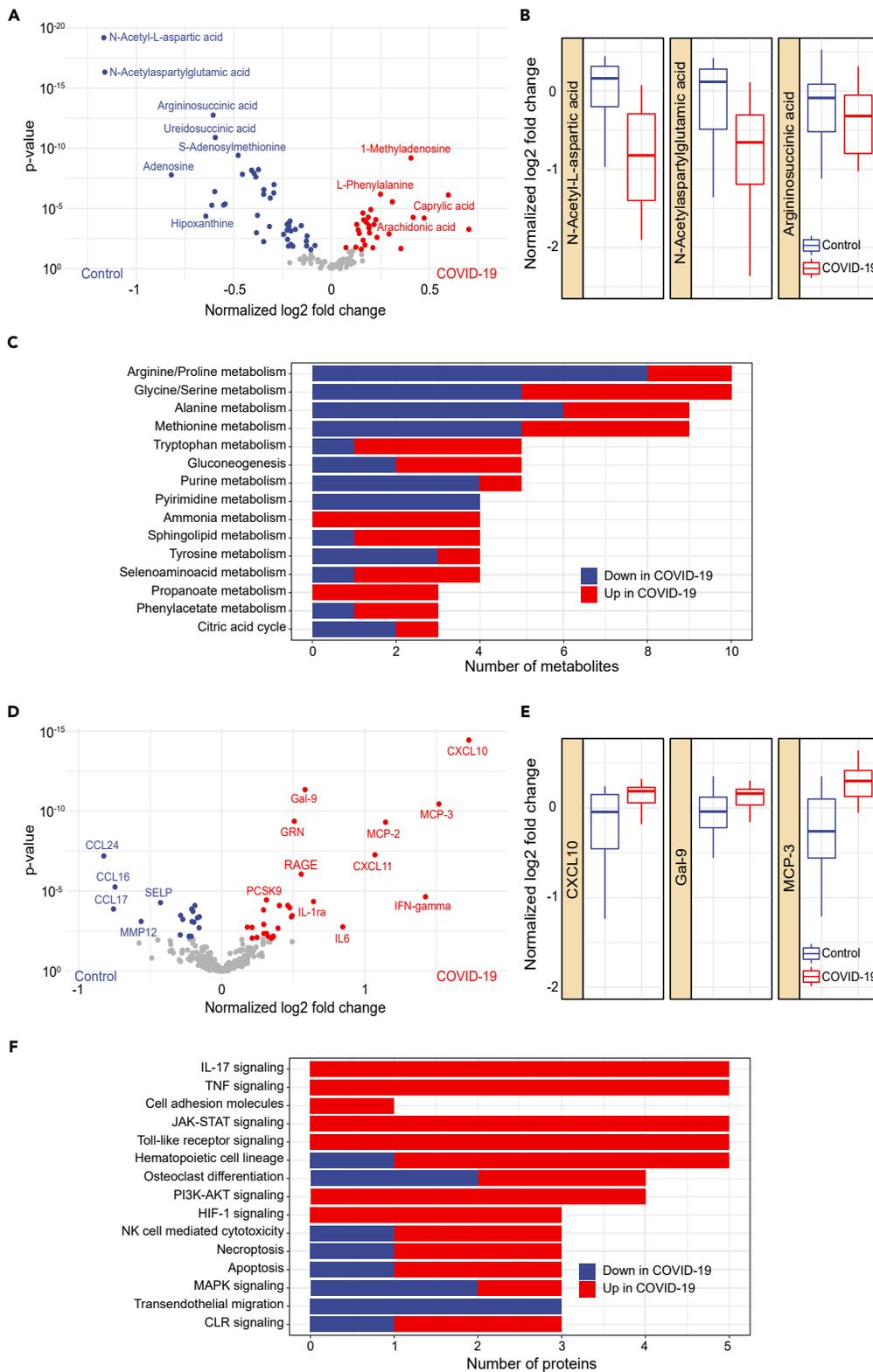
Characteristic	Negative, N = 97 <sup>a</sup>	COVID-19, N = 330 <sup>a</sup>	p value <sup>b</sup>
<b>Demographics</b>			
Age	64 (51, 77)	64 (52, 75)	0.8
Sex			<b>0.017</b>
Female	48 (49%)	117 (35%)	
Male	49 (51%)	213 (65%)	
Race			<b>&lt;0.001</b>
Asian	10 (10%)	24 (7.3%)	
Black	15 (15%)	40 (12%)	
Not 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)	<b>&lt;0.001</b>
<b>Risk Factors</b>			
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%)	<b>&lt;0.001</b>
Active cancer	40 (41%)	30 (9.1%)	<b>&lt;0.001</b>
Immunosuppressed state	29 (30%)	14 (4.2%)	<b>&lt;0.001</b>
<b>Laboratories</b>			
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)	<b>&lt;0.001</b>
Aspartate aminotransferase	25 (19, 38)	40 (28, 61)	<b>&lt;0.001</b>
Ferritin	727 (147, 1167)	769 (358, 1449)	0.4
C-reactive protein	3 (1, 7)	11 (6, 19)	<b>0.006</b>
Procalcitonin	0.43 (0.13, 1.06)	0.17 (0.08, 0.42)	<b>0.007</b>
<b>Clinical Outcomes</b>			
Acute Kidney Injury (AKI) stage			<b>&lt;0.001</b>
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%)	<b>&lt;0.001</b>
Thromboembolic event	6 (6.2%)	22 (6.7%)	>0.9
In-hospital mortality	13 (13%)	56 (17%)	0.6

Bold values indicate p<0.05.

<sup>a</sup>Statistics presented: median (IQR); n (%).

<sup>b</sup>Statistical tests performed: Wilcoxon rank-sum test; chi-square test of independence; Fisher's test with simulated p value; Fisher's exact test.

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, because it has previously



**Figure 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.

**Figure 2. Continued**

(B) Top three differentially expressed metabolites in COVID-19 patients vs. controls based on adjusted p values. Y axis shows log<sub>2</sub> 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 three differentially expressed proteins in COVID-19 patients vs. controls based on adjusted p values. Y axis shows log<sub>2</sub> fold changes in relation to the mean of the control group.

(F) Number of significantly regulated molecules in KEGG pathways, top 15 shown.

been shown that inflammation and cardiovascular pathways are essential during COVID-19 pathogenesis (Wiersinga et al., 2020).

**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 (Wu et al., 2021; Thomas et al., 2020; Barberis et al., 2020). In addition, our data has shown that multiple metabolites involved in sphingolipid metabolism are significantly increased in COVID-19 patients, pointing toward other potential targets for COVID-19 treatment (Törnquist et al., 2021).

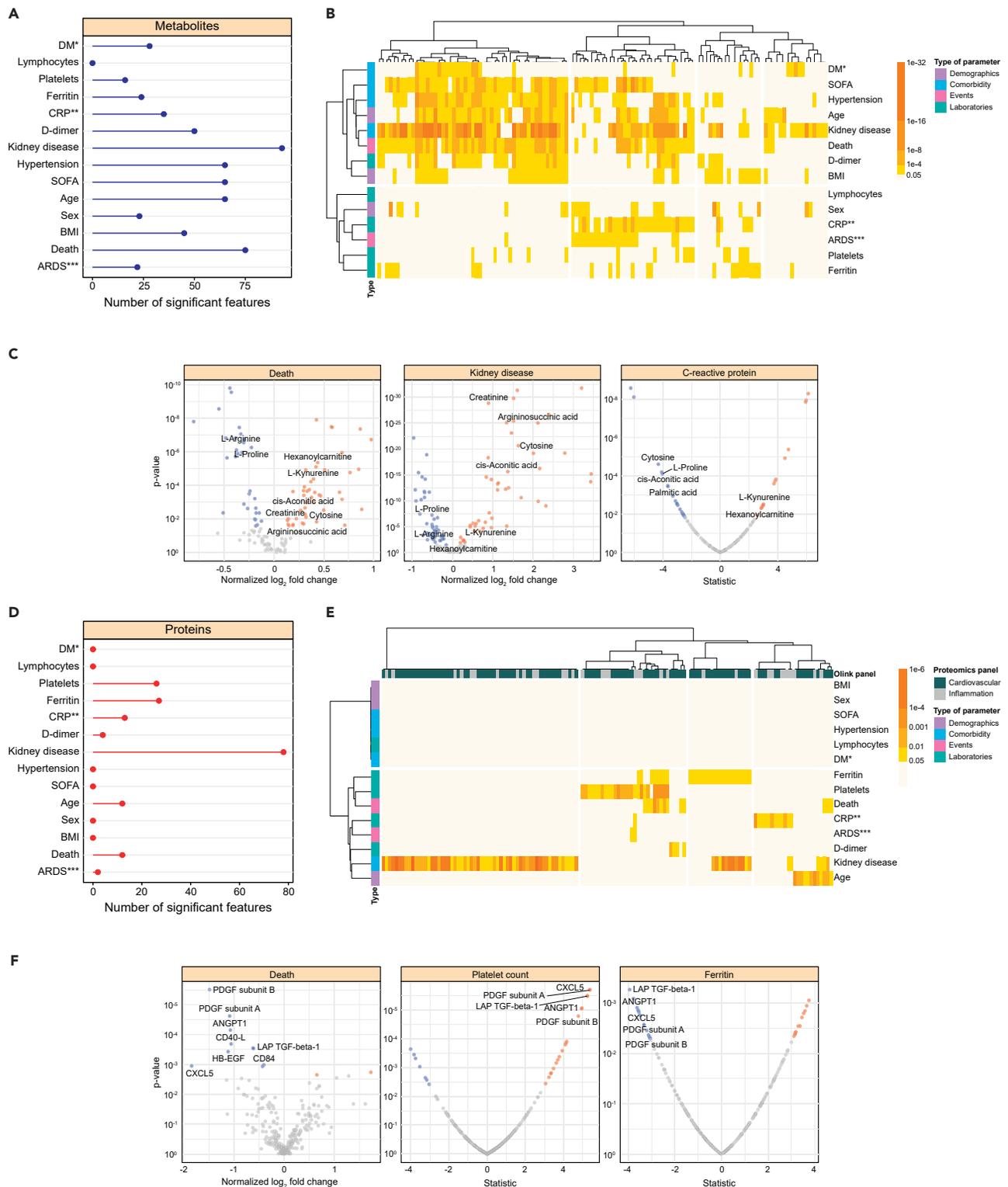
Comparative proteomic analysis identified significant changes in the expression of 48 out of the 266 analyzed proteins between COVID-19 and controls at an FDR of 0.05 (Figure 2D). The top three differentially expressed proteins were C-X-C motif chemokine ligand 10 (CXCL10) (p value = 5.09E-13), galectin 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 Table S4. 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 (Kalil et al., 2020; Iastrebnier et al., 2021). 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 (Figure S2). This is an effect that we have commonly observed in previous studies of blood data (Roberts et al., 2021; Valdés et al., 2022; Meoni et al., 2021; Overmyer et al., 2021; Di et al., 2020; Gisby et al., 2021), 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 (Data S1). GGMs are correlation-based network models that we have previously demonstrated to accurately reconstruct biological pathways from blood-based omics data (Krumisiek et al., 2011, 2012; Benedetti et al., 2017). 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 four network modules to query the interplay between metabolism, inflammation, and vascular dysfunction in COVID-19 patients.





**Figure 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, represented by whether patients were intubated).

(B) Heatmap indicating the number of differentially expressed metabolites associated with each clinical parameter and its significance. We included four clinical parameter categories: demographics, comorbidities, clinical events, and laboratory parameters.

**Figure 4. Continued**

- (C) Volcano plot showing detailed metabolic changes correlated with 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 four categories described above. Proteins are marked according to Olink panels.  
 (F) Volcano plot showing detailed protein changes correlated with death, platelet count and ferritin levels.

*Vascular-related network modules*

**Module 3.** Although COVID-19 presents itself mainly as a respiratory disease, autopsy reports have additionally described significant vascular injury because of endothelial cell damage, microcirculatory thrombi, and impaired cellular junction integrity (Varga et al., 2020). 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 (Xu et al., 2019). MERTK, a member of the TAM family of receptor tyrosine kinases (RTKs), and RAGE, a pro-coagulant and inflammatory molecule were upregulated, whereas 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 (Takeshita et al., 2020). 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 (Monickaraj et al., 2018). Notably, increased vascular permeability and glycolysis are known features of a pro-angiogenic state (Leung and Shi, 2021). Although dysregulated angiogenesis occurs in COVID-19 patients (Ackermann et al., 2020), how these events are coordinated remains poorly understood. Here, we report the concomitant upregulation 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, suggesting a link between abnormal vasculature and the dysregulated glucose metabolism seen in our higher risk diabetic COVID-19 patients (Chen et al., 2020). 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, and BMI), ii) concurrent comorbidity (hypertension, preexisting 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) (Figures 4A and 4D). Preexisting kidney disease was found to be associated 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 Tables S5 and S6.

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, preexisting kidney disease and CRP) are shown in Figure 4C. Seventy-five metabolites were significantly associated with death, 96 with preexisting 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, whereas 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.

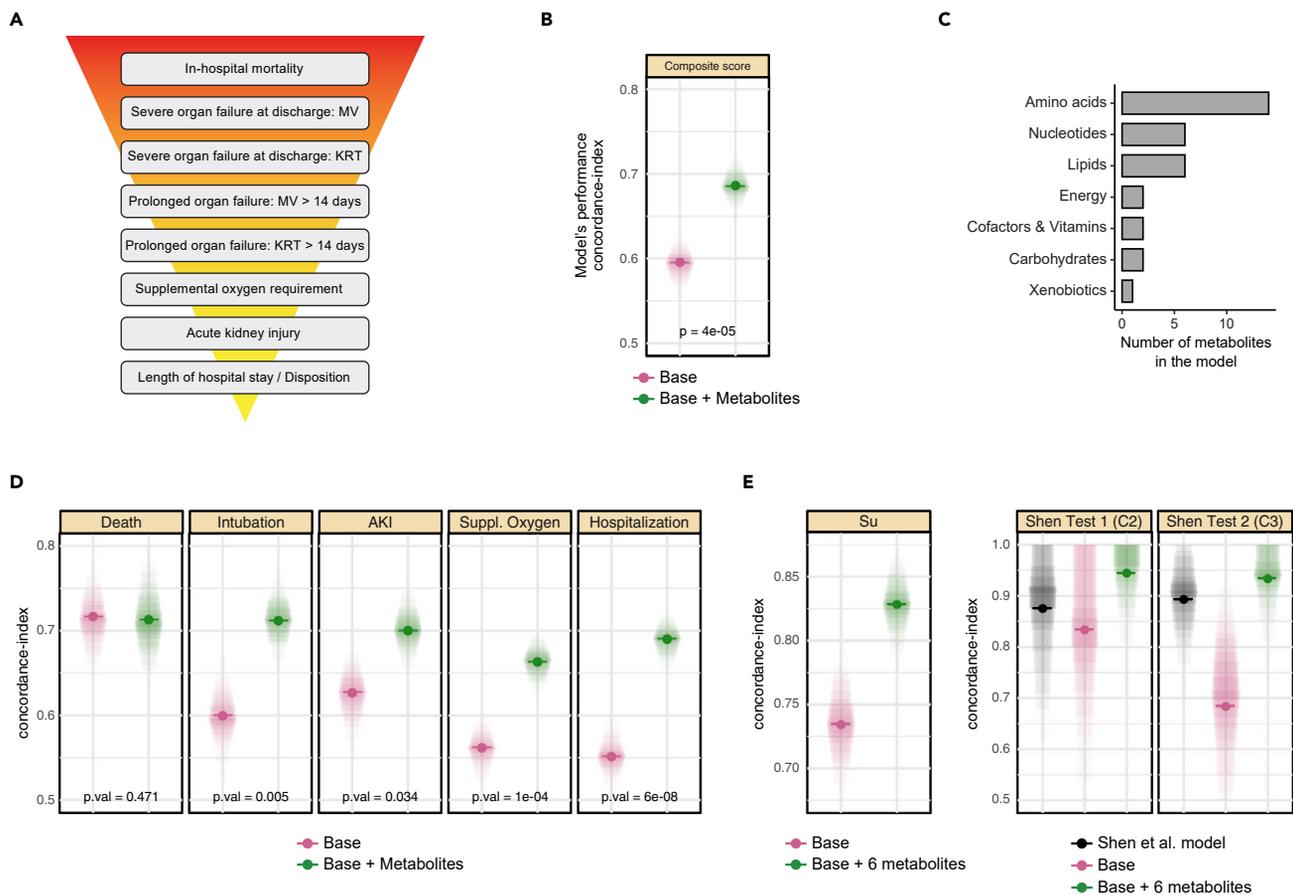
### Metabolomics signature predicts clinical outcomes in COVID-19

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) (Nalbandian et al., 2021): 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 Figure S3.

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, because 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, six nucleotides, six lipids and two 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 ( $d \sim 0.68$ ) 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 (Figure S4). 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. (2020) and Shen et al. (2020) (Figure 5E). The Su study reported COVID-19 severity using a WHO-based score with seven 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



**Figure 5. Metabolomics signature predicts clinical outcomes of COVID-19**

(A) Scheme of the hierarchical composite outcome ranging from in-hospital mortality to length of hospital stay and disposition (MV: Mechanical ventilation, KRT: Kidney replacement therapy).

(B) Dot plot comparing the predictive performance (Concordance index) of a 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. Dots represent median values and shaded areas are 95% confidence intervals.

(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. (Su et al., 2020) and Shen et al. (Shen et al., 2020). This analysis was performed on the reduced 6-metabolite model. The Shen et al. study developed their own model, which is plotted in black (Shen et al., 2020).

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 because of 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.

## DISCUSSION

The novel coronavirus has ravaged the global healthcare system because of 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, whereas 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 (Brown et al., 2021), and several other molecules such as carnitine, niacinamide and IFN-gamma which others have been studying in the context of COVID-19 therapies (Li et al., 2021; van Laarhoven et al., 2021; Fu et al., 2021).

There is increasing evidence that evaluating symptoms and multiple clinical outcomes during acute disease is crucial in determining the risk of long COVID-19 (Sudre et al., 2021). 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 (Shen et al., 2020; Arunachalam et al., 2020; Overmyer et al., 2021; Lucas et al., 2020; Su et al., 2020), 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.

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.

### Limitations of the study

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.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS

- Cohort description and ethical approvals
- **METHOD DETAILS**
  - Data collection
  - Sample handling
  - Metabolomic profiling
  - Proteomic profiling
  - Data preprocessing
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Statistical analysis
  - Network construction
  - Composite outcome
  - Regularized linear mixed effect ordinal regression model
  - Validation datasets

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104612>.

## ACKNOWLEDGMENTS

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).

## AUTHOR CONTRIBUTIONS

Conception and design, M.B., S.A-M., A.C.R., R.B., E.S., A.M.K.C., S.J.C., and J.K.; Analysis and Interpretation, M.B., S.A-M., A.C.R., R.B., F.S., K.L.H., H.S., R.E., L.G-E, W.S., E.B., K.C., G.Z., E.S., K.S., J.J.C., Z.Z., S.R-B., H.S.Y., M.E.C., A.M.K.C., S.J.C., and J.K.; Drafting the Manuscript for Important Intellectual Content, M.B., S.A-M., A.C.R., R.B., F.S., E.B., K.S., E.S., A.M.K.C., S.J.C., and J.K.

## DECLARATION OF INTERESTS

A.M.K.C. is a cofounder and equity stock holder for Proterris, which develops therapeutic uses for carbon monoxide. A.M.K.C. has a use patent on CO. In addition, A.M.K.C. has a patent in COPD. J.K. is scientific advisor at iollo and holds equity in Chymia LLC and IP in PsyProtix.

Received: June 24, 2021

Revised: February 5, 2022

Accepted: June 9, 2022

Published: July 15, 2022

## REFERENCES

- Ackermann, M., Verleden, S.E., Kuehnel, M., Haverich, A., Welte, T., Laenger, F., Vanstapel, A., Werlein, C., Stark, H., Tzankov, A., et al. (2020). Pulmonary vascular endothelialitis, thrombosis, and angiogenesis in Covid-19. *N. Engl. J. Med.* 383, 120–128. <https://doi.org/10.1056/nejmoa2015432>.
- Arons, M.M., Hatfield, K.M., Reddy, S.C., Kimball, A., James, A., Jacobs, J.R., Taylor, J., Spicer, K., Bardossy, A.C., Oakley, L.P., et al. (2020). Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. *N. Engl. J. Med.* 382, 2081–2090.
- Arunachalam, P.S., Wimmers, F., Mok, C.K.P., Perera, R.A.P.M., Scott, M., Hagan, T., Sigal, N., Feng, Y., Bristow, L., Tak-Yin Tsang, O., et al. (2020). Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* 369, 1210–1220. <https://doi.org/10.1126/science.abc6261>.
- Barberis, E., Timo, S., Amede, E., Vanella, V.V., Puricelli, C., Cappellano, G., Raineri, D., Cittone, M.G., Rizzi, E., Pedrinelli, A.R., et al. (2020). Large-scale plasma analysis revealed new mechanisms and molecules associated with the host response to SARS-CoV-2. *Int. J. Mol. Sci.* 21, 8623. <https://doi.org/10.3390/ijms21228623>.
- Benedetti, E., Pučić-Baković, M., Keser, T., Wahl, A., Hassinen, A., Yang, J.Y., Liu, L., Trbojević-Akmačić, I., Razdorov, G., Štambuk, J., et al. (2017). Network inference from glycoproteomics data reveals new reactions in the IgG glycosylation pathway. *Nat. Commun.* 8, 1483. <https://doi.org/10.1038/s41467-017-01525-0>.
- Bilaloglu, S., Aphinyanaphongs, Y., Jones, S., Iturrate, E., Hochman, J., and Berger, J.S. (2020). Thrombosis in hospitalized patients with COVID-19 in a New York city Health system. *JAMA*, 799–801. <https://doi.org/10.1001/jama.2020.13372>.
- Brown, M.J., Alazawi, W., and Kanoni, S. (2021). Interleukin-6 receptor antagonists in critically ill patients with Covid-19. *N. Engl. J. Med.* 384, 1491–1502.
- Chen, J., Wu, C., Wang, X., Yu, J., and Sun, Z. (2020). The impact of COVID-19 on blood glucose: a systematic review and meta-analysis.

- Front. Endocrinol. 11, 574541. <https://doi.org/10.3389/fendo.2020.574541>.
- Chen, W.W., Freinkman, E., Wang, T., Birsoy, K., and Sabatini, D.M. (2016). Absolute quantification of matrix metabolites reveals the dynamics of mitochondrial metabolism. *Cell* 166, 1324–1337.e11. <https://doi.org/10.1016/j.cell.2016.07.040>.
- Chetnik, K., Benedetti, E., Gomari, D.P., Schweickart, A., Batra, R., Buyukozkan, M., Wang, Z., Arnold, M., Zierer, J., Suhre, K., and Krumsiek, J. (2021). maplet: an extensible R toolbox for modular and reproducible omics pipelines. Preprint at arXiv. <https://ui.adsabs.harvard.edu/abs/2021arXiv210504305C>.
- Di, B., Jia, H., Luo, O.J., Lin, F., Li, K., Zhang, Y., Wang, H., Liang, H., Fan, J., and Yang, Z. (2020). Identification and validation of predictive factors for progression to severe COVID-19 pneumonia by proteomics. *Signal Transduct. Target. Ther.* 5, 217. <https://doi.org/10.1038/s41392-020-00333-1>.
- Dieterle, F., Ross, A., Schlotterbeck, G., and Senn, H. (2006). Probabilistic quotient normalization as a robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabolomics. *Anal. Chem.* 78, 4281–4290. <https://doi.org/10.1021/ac051632c>.
- Do, K.T., Wahl, S., Raffler, J., Molnos, S., Laimighofer, M., Adamski, J., Suhre, K., Strauch, K., Peters, A., Gieger, C., et al. (2018). Characterization of missing values in untargeted MS-based metabolomics data and evaluation of missing data handling strategies. *Metabolomics* 14, 128. <https://doi.org/10.1007/s11306-018-1420-2>.
- Elezkurtaj, S., Greuel, S., Ihlow, J., Michaelis, E.G., Bischoff, P., Kunze, C.A., Sinn, B.V., Gerhold, M., Hauptmann, K., Ingold-Heppner, B., et al. (2021). Causes of death and comorbidities in hospitalized patients with COVID-19. *Sci. Rep.* 11, 4263. <https://doi.org/10.1038/s41598-021-82862-5>.
- Fu, W., Lei, C., Ma, Z., Qian, K., Li, T., Zhao, J., and Hu, S. (2021). CAR macrophages for SARS-CoV-2 immunotherapy. *Front. Immunol.* 12, 669103. <https://doi.org/10.3389/fimmu.2021.669103>.
- Gisby, J., Clarke, C.L., Medjeral-Thomas, N., Malik, T.H., Papadaki, A., Mortimer, P.M., Buang, N.B., Lewis, S., Pereira, M., Toulza, F., et al. (2021). Longitudinal proteomic profiling of dialysis patients with COVID-19 reveals markers of severity and predictors of death. *Elife* 10, e64827. <https://doi.org/10.7554/elife.64827>.
- Guan, W.J., Ni, Z.Y., Hu, Y., Liang, W.H., Ou, C.Q., He, J.X., Liu, L., Shan, H., Lei, C.L., Hui, D.S.C., et al. (2020). Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* 382, 1708–1720.
- Hirsch, J.S., Ng, J.H., Ross, D.W., Sharma, P., Shah, H.H., Barnett, R.L., Hazzan, A.D., Fishbane, S., Jhaveri, K.D., Abate, M., et al. (2020). Acute kidney injury in patients hospitalized with COVID-19. *Kidney Int.* 98, 209–218. <https://doi.org/10.1016/j.kint.2020.05.006>.
- Hoelzer, K., Shackelton, L.A., and Parrish, C.R. (2008). Presence and role of cytosine methylation in DNA viruses of animals. *Nucleic Acids Res.* 36, 2825–2837. <https://doi.org/10.1093/nar/gkn121>.
- Iastrebner, M., Castro, J., García Espina, E., Lettieri, C., Payaslian, S., Cuesta, M.C., Gutiérrez Fernández, P., Mandrile, A., Contreras, A.P., Gervasoni, S., et al. (2021). Ruxolitinib in severe COVID-19: results of a multicenter, prospective, single arm, open-label clinical study to investigate the efficacy and safety of ruxolitinib in patients with COVID-19 and severe acute respiratory syndrome. *Rev. Fac. Cien. Med. Univ. Nac. Cordoba* 78, 294–302. <https://doi.org/10.31053/1853.0605.v78.n3.32800>.
- Kalil, A.C., Patterson, T.F., Mehta, A.K., Tomashek, K.M., Wolfe, C.R., Ghazaryan, V., Marconi, V.C., Ruiz-Palacios, G.M., Hsieh, L., Kline, S., et al. (2020). Baricitinib plus remdesivir for hospitalized adults with Covid-19. *N. Engl. J. Med.* 384, 795–807. <https://doi.org/10.1056/nejmoa2031994>.
- Kanehisa, M., Furumichi, M., Sato, Y., Ishiguro-Watanabe, M., and Tanabe, M. (2021). KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* 49, D545–D551. <https://doi.org/10.1093/nar/gkaa970>.
- Krumsiek, J., Suhre, K., Evans, A.M., Mitchell, M.W., Mohny, R.P., Milburn, M.V., Wägele, B., Römisch-Margl, W., Illig, T., Adamski, J., et al. (2012). Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. *PLoS Genet.* 8, e1003005. <https://doi.org/10.1371/journal.pgen.1003005>.
- Krumsiek, J., Suhre, K., Illig, T., Adamski, J., and Theis, F.J. (2011). Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC Syst. Biol.* 5, 21. <https://doi.org/10.1186/1752-0509-5-21>.
- León, G., Herrera, M., Vargas, M., Arguedas, M., Sánchez, A., Segura, Á., Gómez, A., Solano, G., Corrales-Aguilar, E., Risner, K., et al. (2021). Development and characterization of two equine formulations towards SARS-CoV-2 proteins for the potential treatment of COVID-19. *Sci. Rep.* 11, 9825. <https://doi.org/10.1038/s41598-021-89242-z>.
- Leung, S.W.S., and Shi, Y. (2021). The glycolytic process in endothelial cells and its implications. *Acta Pharmacol. Sin.* 251–259. <https://doi.org/10.1038/s41401-021-00647-y>.
- Li, C., Ou, R., Wei, Q., and Shang, H. (2021). Carnitine and COVID-19 susceptibility and severity: a mendelian randomization study. *Front. Nutr.* 8, 780205. <https://doi.org/10.3389/fnut.2021.780205>.
- Liu, M., O'Connor, R.S., Trefely, S., Graham, K., Snyder, N.W., and Beatty, G.L. (2019). Metabolic rewiring of macrophages by CpG potentiates clearance of cancer cells and overcomes tumor-expressed CD47-mediated 'don't-eat-me' signal. *Nat. Immunol.* 20, 265–275. <https://doi.org/10.1038/s41590-018-0292-y>.
- Lucas, C., Wong, P., Klein, J., Castro, T.B.R., Silva, J., Sundaram, M., Ellingson, M.K., Mao, T., Oh, J.E., Israelow, B., et al. (2020). Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* 584, 463–469. <https://doi.org/10.1038/s41586-020-2588-y>.
- McCullagh, P. (1980). Regression models for ordinal data. *J. Roy. Stat. Soc. B* 42, 109–127. <https://doi.org/10.1111/j.2517-6161.1980.tb01109.x>.
- McGonagle, D., Ramanan, A.V., and Bridgewood, C. (2021). Immune cartography of macrophage activation syndrome in the COVID-19 era. *Nat. Rev. Rheumatol.* 17, 145–157. <https://doi.org/10.1038/s41584-020-00571-1>.
- Meoni, G., Ghini, V., Maggi, L., Vignoli, A., Mazzoni, A., Salvati, L., Capone, M., Vanni, A., Tenori, L., Fontanari, P., et al. (2021). Metabolomic/lipidomic profiling of COVID-19 and individual response to tocilizumab. *PLoS Pathog.* 17, e1009243. <https://doi.org/10.1371/journal.ppat.1009243>.
- Monickaraj, F., Mcguire, P., and Das, A. (2018). Cathepsin D plays a role in endothelial-pericyte interactions during alteration of the blood-retinal barrier in diabetic retinopathy. *FASEB J.* 32, 2539–2548. <https://doi.org/10.1096/fj.201700781rr>.
- Nalbandian, A., Sehgal, K., Gupta, A., Madhavan, M.V., Mcgroder, C., Stevens, J.S., Cook, J.R., Nordvig, A.S., Shalev, D., Scharawat, T.S., et al. (2021). Post-acute COVID-19 syndrome. *Nat. Med.* 27, 601–615. <https://doi.org/10.1038/s41591-021-01283-z>.
- Overmyer, K.A., Shishkova, E., Miller, I.J., Balnis, J., Bernstein, M.N., Peters-Clarke, T.M., Meyer, J.G., Quan, Q., Muehlbauer, L.K., Trujillo, E.A., et al. (2021). Large-scale multi-omic analysis of COVID-19 severity. *Cell Syst.* 12, 23–40.e7. <https://doi.org/10.1016/j.cels.2020.10.003>.
- Prim, R.C. (1957). Shortest connection networks and some generalizations. *Bell Syst. Tech. J.* 36, 1389–1401. <https://doi.org/10.1002/j.1538-7305.1957.tb01515.x>.
- Ripatti, S., and Palmgren, J. (2000). Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* 56, 1016–1022. <https://doi.org/10.1111/j.0006-341x.2000.01016.x>.
- Roberts, I., Wright Muelas, M., Taylor, J.M., Davison, A.S., Xu, Y., Grixti, J.M., Gotts, N., Sorokin, A., Goodacre, R., and Kell, D.B. (2021). Untargeted metabolomics of COVID-19 patient serum reveals potential prognostic markers of both severity and outcome. *Metabolomics* 18, 6. <https://doi.org/10.1007/s11306-021-01859-3>.
- Schäfer, J., and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* 4, Article32. <https://doi.org/10.2202/1544-6115.1175>.
- Schenck, E.J., Hoffman, K.L., Cusick, M., Kabariti, J., Sholle, E.T., and Campion, T.R. (2021). Critical carE database for advanced Research (CEDAR): an automated method to support intensive care units with electronic Health record data. *J. Biomed. Inf.* 118, 103789. <https://doi.org/10.1016/j.jbi.2021.103789>.
- Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., Quan, S., Zhang, F., Sun, R., Qian, L., et al. (2020). Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* 182, 59–72.e15. <https://doi.org/10.1016/j.cell.2020.05.032>.

- Soliman, S., Faris, M.E., Ratemi, Z., and Halwani, R. (2020). Switching host metabolism as an approach to dampen SARS-CoV-2 infection. *Ann. Nutr. Metab.* 76, 297–303. <https://doi.org/10.1159/000510508>.
- Stanley, C.A., Hale, D.E., Berry, G.T., Deleeeuw, S., Boxer, J., and Bonnefont, J.-P. (1992). A deficiency of carnitine–acylcarnitine translocase in the inner mitochondrial membrane. *N. Engl. J. Med.* 327, 19–23. <https://doi.org/10.1056/nejm199207023270104>.
- Su, Y., Chen, D., Yuan, D., Lausted, C., Choi, J., Dai, C.L., Voillet, V., Duvvuri, V.R., Scherler, K., Troisch, P., et al. (2020). Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19. *Cell* 183, 1479–1495.e20. <https://doi.org/10.1016/j.cell.2020.10.037>.
- Sudre, C.H., Murray, B., Varsavsky, T., Graham, M.S., Penfold, R.S., Bowyer, R.C., Pujol, J.C., Klaser, K., Antonelli, M., Canas, L.S., et al. (2021). Attributes and predictors of long COVID. *Nat. Med.* 27, 626–631. <https://doi.org/10.1038/s41591-021-01292-y>.
- Takeshita, A., Yasuma, T., Nishihama, K., D’Alessandro-Gabazza, C.N., Toda, M., Totoki, T., Okano, Y., Uchida, A., Inoue, R., Qin, L., et al. (2020). Thrombomodulin ameliorates transforming growth factor- $\beta$ 1-mediated chronic kidney disease via the G-protein coupled receptor 15/Akt signal pathway. *Kidney Int.* 98, 1179–1192. <https://doi.org/10.1016/j.kint.2020.05.041>.
- Tang, N., Li, D., Wang, X., and Sun, Z. (2020). Abnormal coagulation parameters are associated with poor prognosis in patients with novel coronavirus pneumonia. *J. Thromb. Haemost.* 18, 844–847. <https://doi.org/10.1111/jth.14768>.
- Therneau, T. (2020). Coxme and the Laplace approximation [Online]. <https://cran.r-project.org/web/packages/coxme/vignettes/laplace.pdf>.
- Thomas, T., Stefanoni, D., Reisz, J.A., Nemkov, T., Bertolone, L., Francis, R.O., Hudson, K.E., Zimring, J.C., Hansen, K.C., Hod, E.A., et al. (2020). COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. *JCI Insight* 5, 140327. <https://doi.org/10.1172/jci.insight.140327>.
- Tokunaga, R., Zhang, W., Naseem, M., Puccini, A., Berger, M.D., Soni, S., Mckane, M., Baba, H., and Lenz, H.-J. (2018). CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation - a target for novel cancer therapy. *Cancer Treat. Rev.* 63, 40–47. <https://doi.org/10.1016/j.ctrv.2017.11.007>.
- Törnquist, K., Asghar, M.Y., Srinivasan, V., Korhonen, L., and Lindholm, D. (2021). Sphingolipids as modulators of SARS-CoV-2 infection. *Front. Cell Dev. Biol.* 9, 689854. <https://doi.org/10.3389/fcell.2021.689854>.
- Valdés, A., Moreno, L.O., Rello, S.R., Orduña, A., Bernardo, D., and Cifuentes, A. (2022). Metabolomics study of COVID-19 patients in four different clinical stages. *Sci. Rep.* 12, 1650. <https://doi.org/10.1038/s41598-022-05667-0>.
- van Laarhoven, A., Kurver, L., Overheul, G.J., Kooistra, E.J., Abdo, W.F., van Crevel, R., Duivendoorn, R., Kox, M., Ten Oever, J., Schouten, J., et al. (2021). Interferon gamma immunotherapy in five critically ill COVID-19 patients with impaired cellular immunity: a case series. *Med (N Y)* 2, 1163–1170.e2. <https://doi.org/10.1016/j.medj.2021.09.003>.
- Varga, Z., Flammer, A.J., Steiger, P., Haberecker, M., Andermatt, R., Zinkernagel, A.S., Mehra, M.R., Schuepbach, R.A., Ruschitzka, F., and Moch, H. (2020). Endothelial cell infection and endotheliitis in COVID-19. *Lancet* 395, 1417–1418. [https://doi.org/10.1016/s0140-6736\(20\)30937-5](https://doi.org/10.1016/s0140-6736(20)30937-5).
- Wajner, M., and Amaral, A.U. (2015). Mitochondrial dysfunction in fatty acid oxidation disorders: insights from human and animal studies. *Biosci. Rep.* 36, e00281. <https://doi.org/10.1042/bsr20150240>.
- Wiersinga, W.J., Rhodes, A., Cheng, A.C., Peacock, S.J., and Prescott, H.C. (2020). Pathophysiology, transmission, diagnosis, and treatment of coronavirus disease 2019 (COVID-19): a review. *JAMA* 324, 782–793. <https://doi.org/10.1001/jama.2020.12839>.
- Wu, P., Chen, D., Ding, W., Wu, P., Hou, H., Bai, Y., Zhou, Y., Li, K., Xiang, S., Liu, P., et al. (2021). The trans-omics landscape of COVID-19. *Nat. Commun.* 12, 4543. <https://doi.org/10.1038/s41467-021-24482-1>.
- Xu, S., Gu, M., Wu, K., and Li, G. (2019). Unraveling the role of hydroxyproline in maintaining the thermal stability of the collagen triple helix structure using simulation. *J. Phys. Chem. B* 123, 7754–7763. <https://doi.org/10.1021/acs.jpcc.9b05006>.
- Zhang, C., Shi, L., and Wang, F.-S. (2020). Liver injury in COVID-19: management and challenges. *Lancet Gastroenterol. Hepatol.* 5, 428–430. [https://doi.org/10.1016/s2468-1253\(20\)30057-1](https://doi.org/10.1016/s2468-1253(20)30057-1).
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., Xiang, J., Wang, Y., Song, B., Gu, X., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 395, 1054–1062. [https://doi.org/10.1016/s0140-6736\(20\)30566-3](https://doi.org/10.1016/s0140-6736(20)30566-3).

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Biological samples</b>		
Human Serum Samples	Weill Cornell Medicine IRB Protocol #19-10020914 – NewYork-Presbyterian Core Laboratory	N/A
<b>Chemicals, peptides, and recombinant proteins</b>		
Methanol (LC/MS grade)	Fisher scientific	A456-4
Water (LC/MS grade)	Fisher scientific	W6-4
Acetonitrile (LC/MS grade)	Fisher scientific	A955-4
Ammonium acetate	Fluka	631-61-8
Ammonium hydroxide solution	Sigma-Aldrich	338818-100ML
<b>Deposited data</b>		
Metabolomics, proteomics, and clinical data	<a href="https://doi.org/10.6084/m9.figshare.19115972">https://doi.org/10.6084/m9.figshare.19115972</a>	N/A
<b>Software and algorithms</b>		
XCalibur 4.1	ThermoFisher Scientific	N/A
<b>Other</b>		
Q Exactive Orbitrap mass spectrometer	ThermoFisher Scientific	N/A
Vanquish UPLC system	ThermoFisher Scientific	N/A
SeQuant ZIC-pHILIC column	Millipore Sigma	1504600001

## RESOURCE AVAILABILITY

## Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Jan Krumsiek.

## Materials availability

This study did not generate new unique reagents.

## Data and code availability

- Code: Code to reproduce all the statistical results presented in this paper is available at <https://github.com/krumsieklab/covid-omics>.
- Data: The data used in this study can be downloaded at <https://doi.org/10.6084/m9.figshare.19115972.v1>

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

## Cohort description and ethical approvals

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. 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. A full description of the cohort can be found in [Tables 1, S1 and S2](#). The collection and analysis of the data and human samples satisfied the principles outlined in the Declaration of Helsinki and were approved by the local IRB (Weill Cornell Medicine IRB, protocol #19-10020914), with a waiver of informed consent.

## METHOD DETAILS

### 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 ([Schenck et al., 2021](#)) 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.

### 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.

### Metabolomic profiling

Targeted metabolite profiling was performed according to a method described in a previous publication ([Chen et al., 2016](#)). For metabolite extraction, 80  $\mu$ L of pre-chilled methanol (–80 °C) was added to 20  $\mu$ L 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.  $\times$  150 mm, particle size of 5  $\mu$ m, Millipore Sigma) was used for separation of metabolites. A 2.1  $\times$  20 mm guard column with the same packing material was used for protection of the analytical column. Flow rate was set at 150  $\mu$ L/min. Buffers consisted of 100% acetonitrile for mobile phase A, and 0.1% NH<sub>4</sub>OH/20 mM CH<sub>3</sub>COONH<sub>4</sub> 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 (Figure S5).

### 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, log<sub>2</sub>-transformed Ct values from each sample and analyte were normalized based on spiked-in extension controls and scale-inverted to obtain Normalized log<sub>2</sub>-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 (Kanehisa et al., 2021).

### 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 (Dieterle et al., 2006), and log<sub>2</sub> 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 (Do et al., 2018). 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 (Do et al., 2018).

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.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### 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 (Chetnik et al., 2021) (<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 (Schäfer and Strimmer, 2005) 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 (Prim, 1957) 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 [Figure S3](#).

### 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 (McCullagh, 1980), using maximum likelihood (ML) estimation as proposed by Ripatti and Palmgren (Ripatti and Palmgren, 2000). 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 (Therneau, 2020), 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. \(2020\)](#) ( $n = 121$ ) and [Shen et al. \(2020\)](#) (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  $\sim$  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](#).