

Metabolite, Protein, And Tissue Dysfunction Associated With COVID-19 Disease Severity

Ali Rahnavard (✉ rahnavard@gwu.edu)

The George Washington University

Brendan Mann

The George Washington University

Abhigya Giri

The George Washington University

Ranojoy Chatterjee

The George Washington University

Keith Crandall

The George Washington University

Research Article

Keywords: biological, clinical, metabolite

Posted Date: November 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1041144/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Scientific Reports on July 16th, 2022. See the published version at <https://doi.org/10.1038/s41598-022-16396-9>.

Metabolite, protein, and tissue dysfunction associated with COVID-19 disease severity

Ali Rahnavard^{1*}, Brendan Mann^{1,2}, Abhigya Giri¹, Ranojoy Chatterjee¹, Keith A. Crandall¹

¹Computational Biology Institute, Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, The George Washington University, Washington, DC 20052

²Department of Microbiology, Immunology, and Tropical Medicine, School of Medicine and Health Sciences, The George Washington University, Washington, DC 20052

*Correspondence to rahnavard@gwu.edu

Key points

- We investigated the metabolite, protein, and clinical measurements to test for significant associations with patients' characteristics and health outcomes. We identify biomarkers such as citrulline and Hyaluronan-binding protein 2 that are indicative of multi-system tissue dysfunction and can be used for COVID-19 diagnosis.
- We identify important pathways associated with COVID-19 patients (e.g., complement activation and neutrophil degranulation) using our novel pathway enrichment analysis.
- We compared different machine learning techniques to demonstrate their performance in predicting COVID-19 health outcomes from omics data.
- We provide a set of novel computational techniques to analyze omics data including:
 - *omeClust* to identify important metadata (e.g., patients' characterization and clinical information).
 - *Tweediverse* to test omics feature association with metadata of interest with consideration of high dimensionality and zero-inflated omics data.
 - *deepath*, Omics Pathway Enrichment Analysis tool, to find significantly enriched pathways among clinical groups using protein and metabolites data.

Abstract

Proteins are direct products of the genome and metabolites are functional products of interactions between the host and other factors such as environment, disease state, clinical information, etc. Omics data, including proteins and metabolites, are useful in characterizing biological processes underlying COVID-19 along with patient data and clinical information, yet few methods are available to effectively analyze such diverse and unstructured data. Using an integrated approach that combines proteomics and metabolomics data, we investigated the changes in metabolites and proteins in relation to patient characteristics (e.g., age, gender, and health outcome) and clinical information (e.g., metabolic panel and complete blood count test results). We found significant enrichment of biological indicators of lung, liver, and gastrointestinal dysfunction associated with disease severity using metabolite and protein profiles. Our analyses specifically identified enriched proteins that play a critical role in responses to injury or infection within these anatomical sites, but may contribute to excessive systemic inflammation within the context of

COVID-19. Furthermore, we have used this information in conjunction with machine learning algorithms to predict the health status of patients presenting symptoms of COVID-19. This work provides a roadmap for understanding the biochemical pathways and molecular mechanisms that drive disease severity, progression, and treatment of COVID-19.

Introduction

We have shown that the SARS-CoV-2 genomic variation is independent of host characteristics (e.g., age and gender)¹, and here as a next step, we investigated the factors that influence COVID-19 disease severity using an integrated approach involving analyses of metabolite and protein profiles, epidemiological data, and clinical data, previously published². The measurements included a complete blood cell count panel, a comprehensive metabolic test panel, and quantification of 941 metabolites and 894 proteins from serum samples. Metabolite and protein profiling were performed using ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) and stable isotope-labeled proteomics strategy TMTpro (16plex)³. Health outcomes are defined as healthy, non-COVID-19, non-severe COVID-19, and severe COVID-19. We found that the health outcome is significantly associated with patient age (p -value = 0.001, Kruskal-Wallis test). Pairwise comparisons (Dunn's test with Benjamini-Hochberg adjustment) revealed that the age distribution was significantly different between the severe and healthy groups (p -value = 0.008) (**Fig. 1a**), and the severe and non-severe groups (p -value = 0.002) (**Fig. 1b**); among infected people, severe COVID-19 is more likely in older individuals. No significant associations were present between health outcome and sex (p -value = 0.5308, Kruskal-Wallis test), and health outcome and BMI (p -value = 0.148, Kruskal-Wallis test) (**Fig. 1c**). The time between disease onset and sample collection for metabolites and proteins varies among groups (**Fig. 1d**) in the study and should be considered in subsequent analyses, especially as the sample collection for proteomics and metabolomics are not at the same time for individuals ((**Fig. 1e**). Overall protein (**Fig. 1f**) and metabolite (**Fig. 1g**) profiles can strongly be explained by clinical information (using omeClust enrichment score⁴) and also by various health outcome statuses. However, sex seems to have the lowest effect on overall metabolites and protein profiles compared to other clinical information (**SFig. 1**).

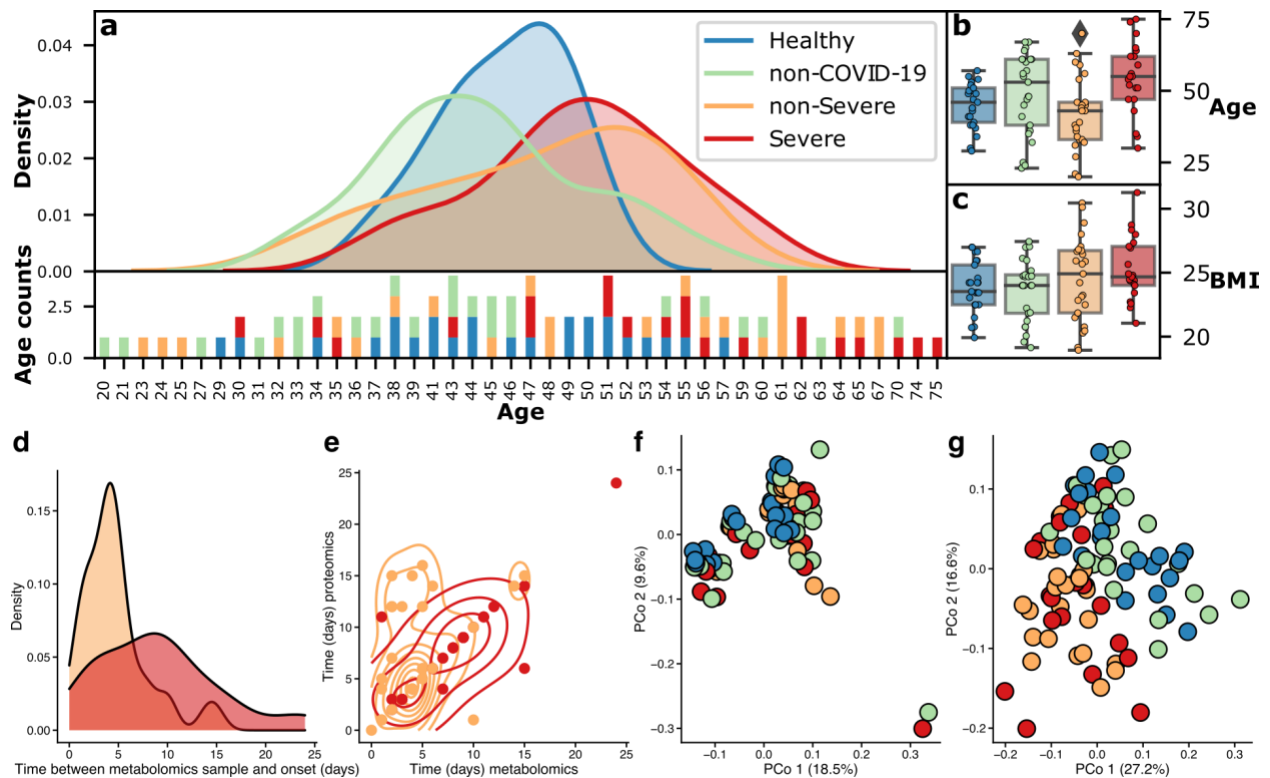


Figure 1. Distribution of study participants by age and health outcome. **a**, the sample population is not uniform for age ranges, with a higher proportion of older participants falling into the severe group. **b**, age has a strong association with health outcomes (p -value = 0.001). **c**, BMI is not associated with health outcomes (p -value = 0.1488). For assessing the association of age and gender with the health outcome status, we performed the Kruskal-Wallis test. **d**, time between disease onset and sample collection for metabolites varies among health status. **e**, metabolite and protein samples have been collected at different times within individuals and for groups with different health statuses. **f**, ordination plot using proteins and **g**, ordination plot using metabolites profiles reveal overall structure among individuals group colored by health status. However, the signal is stronger using metabolite profiles measured by omeClust enrichment score⁴.

Results

Investigation biomarkers of COVID-19 severity and dysfunctions

Proteins are the downstream products of gene expression, and metabolites are products of biological reactions. Quantification and joint analysis of these molecules provide information to investigate the mechanisms underlying human health and disease (**Fig. 2**). We identified important proteins (**Fig. 3a**) and metabolites (**Fig. 4a**) based on significant differences observed among the health outcome groups. We further tested if the associated molecules belong to enriched metabolic pathways. This analysis was conducted using a generalized linear model adjusted for age, sex, and BMI as confounding factors of health outcome. We discuss potential biomarkers identified by our analysis that are biologically relevant during COVID-19 infection. Based on existing literature, these molecules may contribute to the rampant inflammatory response and multiple tissue dysfunction spanning the lung, liver, kidneys, and gastrointestinal tract that has been observed in cases of COVID-19.

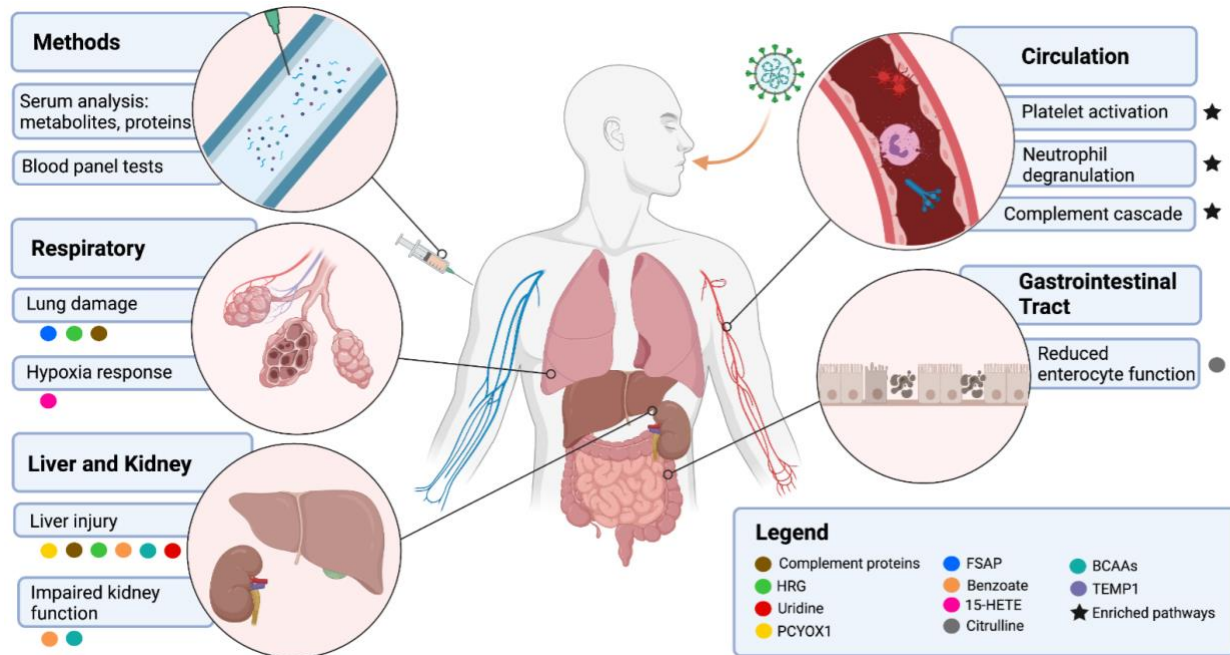


Figure 2: Systemic drivers of COVID-19 associated inflammation. COVID-19 is primarily considered a respiratory tract infection that targets the lung epithelium, but the disease can present clinical manifestations that span the entire body. This includes multisystem abnormalities in biological processes and metabolic functions that may exacerbate the inflammatory response observed in severe cases. An integrated analysis of proteomics and metabolomics data collected from a cohort presenting a range of COVID-19-related health outcomes led to the identification of potential biomarkers within the lungs, liver, gastrointestinal tract, kidneys, and peripheral blood. This analysis provides a deeper resolution of the possible molecular determinants of COVID-19-associated inflammation that are worthy of further investigation.

Imbalance of Serum Nucleic Acids

Viruses have adapted over time to hijack cellular machinery and resources for their own replication. Consequently, homeostatic synthesis and recycling of nucleobases may be disrupted in favor of producing new copies of the viral genome. We found that cytosine levels are elevated in the COVID-19 groups as compared to the non-COVID-19 (**Fig. 3b**) and healthy (coefficient=2.6, p -value = $9.7E-18$) groups. This finding is consistent with a similar study analyzing metabolite profiles of COVID-19 patients, which found cytosine to be the distinguishing feature that determined infection status⁵. It is hypothesized that changes in levels of cytosine are critically involved with RNA virus evolution, including SARS-CoV-2⁶. Notably, the underrepresentation of cytosine within the SARS-CoV-2 genome suggests an alternative role for this metabolite beyond the synthesis of viral RNA. While it is unclear why cytosine levels are higher in COVID-19 patients, this finding points towards cytosine as an effective biomarker for COVID-19 infection.

Our results indicate that uridine levels are lower in the COVID-19 groups as compared to the healthy and non-COVID-19 groups. Uridine is a biologically dynamic metabolite that is critical to the synthesis of RNA and glycogen⁷. Circulating uridine levels are typically high in healthy individuals but tend to undergo short term fluctuations in response to diet. Changes in synthesis are tightly regulated by the liver along with the adipose tissue^{8,9}. Abnormal liver function appears to be a commonality between SARS-CoV^{10,11} and SARS-CoV-2 infection¹²⁻¹⁷. Therefore, liver impairment may mediate the lower uridine levels observed in COVID-19 patients. The functional

consequences of this reduction are also unclear. In an animal model of pulmonary fibrosis, it was shown that uridine supplementation has anti-inflammatory and anti-fibrotic effects¹⁸. While additional studies are needed, a similar therapeutic strategy may mitigate prolonged inflammation within the lungs that leads to eventual injury and disruption of the epithelium¹⁹.

In addition to specific pyrimidines, collectively, the purine metabolism pathway is significantly enriched within the COVID-19 groups compared to the healthy control (**Fig. 5a**). Cell-free purine derivatives such as ATP and adenosine are associated with cellular stress and exert potent immunomodulatory effects. Release of ATP into the extracellular space and subsequent binding to purinergic receptors P2X and P2Y lead to the induction of inflammation including the activation and chemotaxis of phagocytes and memory T cells²⁰. Sustained levels of extracellular ATP trigger the expression of ectoenzymes on the surface of immune cells that convert ATP to adenosine. Binding of adenosine to cognate purinergic receptors tempers inflammation by reducing neutrophil chemotaxis and platelet aggregation in addition to promoting wound healing via the release of vascular endothelial growth factor (VEGF) by macrophages and dendritic cells²¹. Under hypoxic conditions, hypoxia inducible factors HIF- α and HIF- β can alter adenosine metabolism in order to protect tissue from further damage brought about by prolonged inflammation²². Acute lung injury that occurs during severe cases of COVID-19 may trigger such pathways leading to the observed enrichment in purine metabolism.

Multi-organ dysfunction

Citrulline, an important amino acid metabolite in the urea cycle²³, is depleted in the severe (*coefficient* = -0.03974, *p-value* < 0.0001) and the non-severe COVID-19 groups (*coefficient* = -0.0219, *p-value* < 0.0001) compared to the healthy group (**Fig. 3c**). The depletion of citrulline in COVID-19 patients has been associated with gastrointestinal symptoms and systemic inflammation²⁴. Since citrulline is produced by enterocytes within the small intestine, low citrulline levels can be indicative of reduced enterocyte function and mass²⁵. Given that enterocytes express ACE2²⁶, a host receptor that is recognized by the 2019 coronavirus²⁷, it is possible that the presentation of gastrointestinal symptoms²⁸ and the lowered citrulline levels in the COVID-19 patients is a result of enterocyte damage via viral infection.

Levels of branched-chain amino acids (BCAAs) leucine, isoleucine, and valine are lower in the COVID-19 groups (**Fig. 3d**) as compared to the healthy and non-COVID-19 groups (*p-value* = 8E-06 *coefficient* -0.4 for leucine in Severe COVID-19 vs. healthy group). BCAAs play a critical role in protein anabolism²⁹, lowered BCAA levels are often observed in various conditions, including liver cirrhosis³⁰, urea cycle disorders³¹, chronic renal failure^{32,33}, and impaired renal function³⁴. Thus, our finding is congruent with the impaired renal function observed in severe COVID-19 cases³⁵.

Unlike citrulline and BCAAs, benzoate levels are elevated in the COVID-19 (*coefficient* = 4.5, *p-value* = 8.5E-40) and non-COVID-19 (*coefficient* = 4.3, *p-value* = 4.6E-39) groups compared to the healthy group (**Fig. 3e**). Benzoate, in the form of sodium salt, is used as a preservative for foods and drinks³⁶ and a treatment in urea cycle disorders^{37,38}. The metabolism of this compound is directly regulated by the liver and kidneys³⁹⁻⁴¹. Benzoate has been found to have both proinflammatory and anti-inflammatory activities. In an in-vitro study with a colon cancer cell line, sodium benzoate was able to induce apoptosis and activate NF- κ B⁴², a transcription factor critical

for the expression of proinflammatory genes⁴³. Conversely, a review of animal models of multiple sclerosis highlighted the anti-inflammatory functions of sodium benzoate, which include promoting the differentiation of anti-inflammatory Th2 cells, increasing the number of regulatory T cells, and reducing the expression of certain proinflammatory molecules such as TNF-alpha and IL-1beta⁴⁴. Therefore, it is not clear if the elevated levels of benzoate in the COVID-19 and non-COVID-19 groups are indicative of a shared biological phenomenon. The high benzoate levels could be reflective of the body's proinflammatory response to infection. Alternatively, damage to the liver and kidneys due to COVID-19 infection could be disrupting benzoate metabolism, resulting in a backup of benzoate.

Hyaluronan-binding protein 2 or factor VII activating protease (FSAP, protein ID Q14520, **Fig. 4b**) is a binding protein in the human plasma that is expressed in the liver, kidney, and pancreas⁴⁵. It is known to activate coagulation factor-VII⁴⁶ and urokinase single-chain plasminogen activator⁴⁵. We found higher FSAP levels in the COVID-19 groups compared to the non-COVID-19 groups and associations between FSAP and citrulline, and FSAP and uridine in block-wise association testing (**SFig. 2**). Several *in vitro*, as well as patient-based studies, have established a link between FSAP levels, inflammation, and disease. FSAP levels are upregulated in lung endothelial cells that have lipopolysaccharide-induced acute lung injury⁴⁷ and in the inflamed lungs of patients with acute respiratory distress syndrome^{48,49}. Increased FSAP levels in plasma are also associated with other pathologies such as symptomatic carotid stenosis⁵⁰, acute coronary disease⁵¹, and ischemic stroke⁵². *In vitro* studies have shown that FSAP can activate inflammation pathways in non-immune cell populations such as smooth muscle and endothelial cells⁵³ as well as NF-kB mediated proinflammatory cytokine production in myeloid cells⁵⁴. Elevated FSAP levels in the COVID-19 groups could be indicative of systemic inflammation that increases the risk of lung injury and cardiovascular issues.

Prenylcysteine oxidase 1 (PCYOX1 protein, protein ID Q9UHG3, **Fig. 4c**) is responsible for breakdown of prenylcysteines to cysteines and a C-1 aldehyde^{55,56}. It is expressed ubiquitously, but the only expression in the liver leads to its incorporation into lipoproteins⁵⁷; as such, it is associated with very low-density lipoproteins⁵⁸ explaining its presence within the plasma. PCYOX1 levels are depleted in the COVID-19 groups compared to the healthy and non-COVID-19 groups, and the protein is associated with sphingosine-1-phosphate in block-wise association testing. Lower levels of PCYOX1 protein have been observed in mice lacking the interstitial cells of Cajal (ICC) in the gastrointestinal tract⁵⁹ and in mice models of liver injury and dysfunction^{57,60-62}. The lower levels of PCYOX1 protein seen in the COVID-19 groups may contribute to the liver dysfunction associated with COVID-19⁶³.

Inflammation

15-HETE is a metabolite produced when arachidonic acid is oxygenated by arachidonate 15-lipoxygenase⁶⁴. It is associated with inflammation and can display either pro-inflammatory or anti-inflammatory effects⁶⁵. However, the anti-inflammatory effects of 15-HETE are more well-studied and include inhibition of leukotriene B4 action on polymorphonuclear neutrophils (PMN)⁶⁶ and the migration of PMN in response to cytokines⁶⁷. We found severe depletion of 15-HETE levels in the COVID-19 groups compared to the non-COVID-19 and healthy groups. The depletion of 15-HETE and subsequent loss of anti-inflammatory signals could contribute to the heightened inflammation seen during COVID-19 infection⁶⁸. Interestingly, 15-HETE also plays a role in promoting

pulmonary vascular remodeling during hypoxia by exerting pro-angiogenic effects^{69,70}. A reduction in serum 15-HETE levels would suggest that this metabolite does not directly contribute to the cardiovascular dysfunction observed in severe COVID-19 cases.

To date, severe COVID-19 has been associated with an increase in the immediate and long-term risk of thrombosis and coagulation abnormalities^{71,72}. This has primarily been attributed to prolonged overactivation of platelets and high levels of neutrophil degranulation, both of which represent significantly enriched pathways in our analysis of this group. Increases in platelet mediated prothrombosis are typical in response to many invading pathogens. However, SARS-CoV-2 infection in particular triggers significant changes in platelet gene expression and aggregation. Furthermore, excessive platelet activation from COVID-19 leads to alterations in innate immune responses that contribute to thrombotic events. This includes the accumulation of platelet-monocyte complexes, which directly express high tissue factor levels, which directly increases the risk of clotting⁷³. High plasma levels of platelet factors also contribute to dysregulated neutrophil responses, such as the excessive formation of neutrophil extracellular traps and degranulation^{74,75}. In addition to their direct detrimental role to cardiovascular health, these factors may also exacerbate systemic inflammation leading to damage within the tissues.

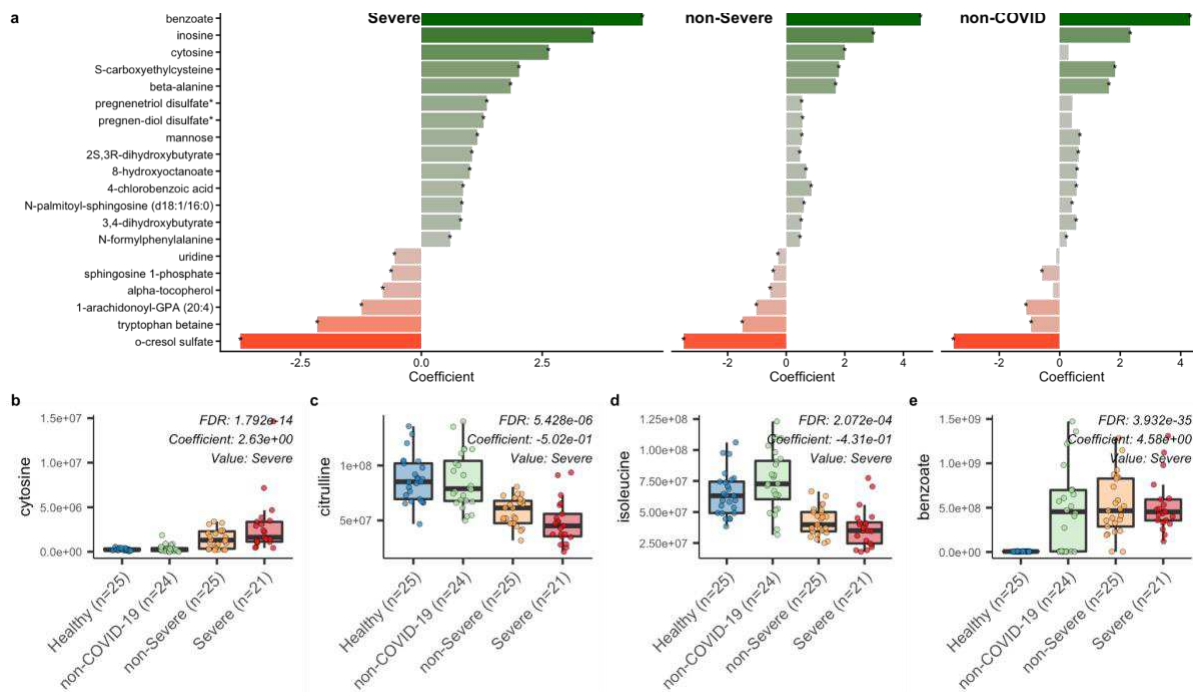


Figure 3: metabolites changes in COVID-19. **a**, 20 most significant metabolites with lowest p-value in comparison of severe group vs. healthy group are shown. Then, the corresponding changes in non-severe and non-COVID-19 for the same metabolites are shown. **b**, **c**, **d**, and **e** show different patterns we observed among these associations. For example, Cytosine has a higher level in COVID-19 groups vs. non-COVID-19 and has been shown that it can play a biomarker for COVID-19 diagnostics.

Complement Activation

The complement system involves a protein cascade that is typically classified as either the classical lectin or alternative pathways^{76–78}. This system plays a major role in B lymphocyte regulation, inflammation, and host protection⁷⁹, and consequently has been associated with

proinflammatory actions and diseases⁸⁰. We found that levels of the complement component 2 (C2, protein ID P06681, **Fig. 4d**) and complement component 9 (C9, Protein ID P02748, **Fig. 4e**) proteins were elevated in the severe and non-severe COVID-19 groups as compared to the healthy group. Briefly, C2 is a protein that forms a short lived complex with C4b to cleave the C3 protein into C3a and C3b⁸¹, and C9 is involved in the formation of a pore-like membrane attack complex associated with bacterial cell lysis⁸². In addition to the individual components of the complement system, pathway enrichment analysis revealed that the pathway representing complement cascade regulation was significantly enriched between the healthy and other three groups, with the greatest difference found between the severe COVID-19 and healthy groups (**Fig. 5g**).

Our findings are in agreement with other studies that have suggested the complement system plays a critical role in COVID-19 pathogenesis⁷⁷. For instance, COVID-19 spike proteins have been shown to activate the complement system via the alternative pathway⁸³, and evidence of complement system activation correlating with respiratory problems in hospitalized COVID-19 patients⁸⁴. It has also been suggested that unregulated activation of the complement system due to viral presence in the lungs can contribute to organ failure and death⁷⁶. Given that C2 is expressed at higher levels in the liver and lungs, and C9 expression is restricted to the liver⁸⁵, our finding of elevated levels of C2 and C9 proteins in the COVID-19 groups could be indicative of the unregulated activation of the complement system in these organs. Currently, it is believed that the complement system has a contradictory role in COVID-19 infection where it is beneficial in mild or asymptomatic cases and harmful in severe cases^{78,86}.

In contrast to the complement component proteins, histidine-rich glycoprotein (HRG, protein ID P04196) levels are heavily depleted in the severe COVID-19 group as compared to the other groups. HRG is a plasma protein that is involved in many biological processes, including immune system regulation, cell adhesion, angiogenesis, and coagulation⁸⁷. HRG inhibits the formation of insoluble immune complexes⁸⁸, which are involved in the host's immune response against foreign substances⁸⁹; when the immune system does not clear these complexes, they can deposit in tissues and activate the complement system inflammation⁸⁹. HRG can also enhance complement activation on necrotic tissues⁹⁰ and is directly involved in clearing of apoptotic⁹¹ and necrotic cells⁹²⁻⁹⁴. Low HRG levels have been observed in patients with advanced lung cancer⁹⁵ and advanced liver cirrhosis⁹⁶. Therefore, our findings of depleted HRG levels could be indicative of organ damage in the severe COVID-19 patients, similar to what has been observed in other pathologies.

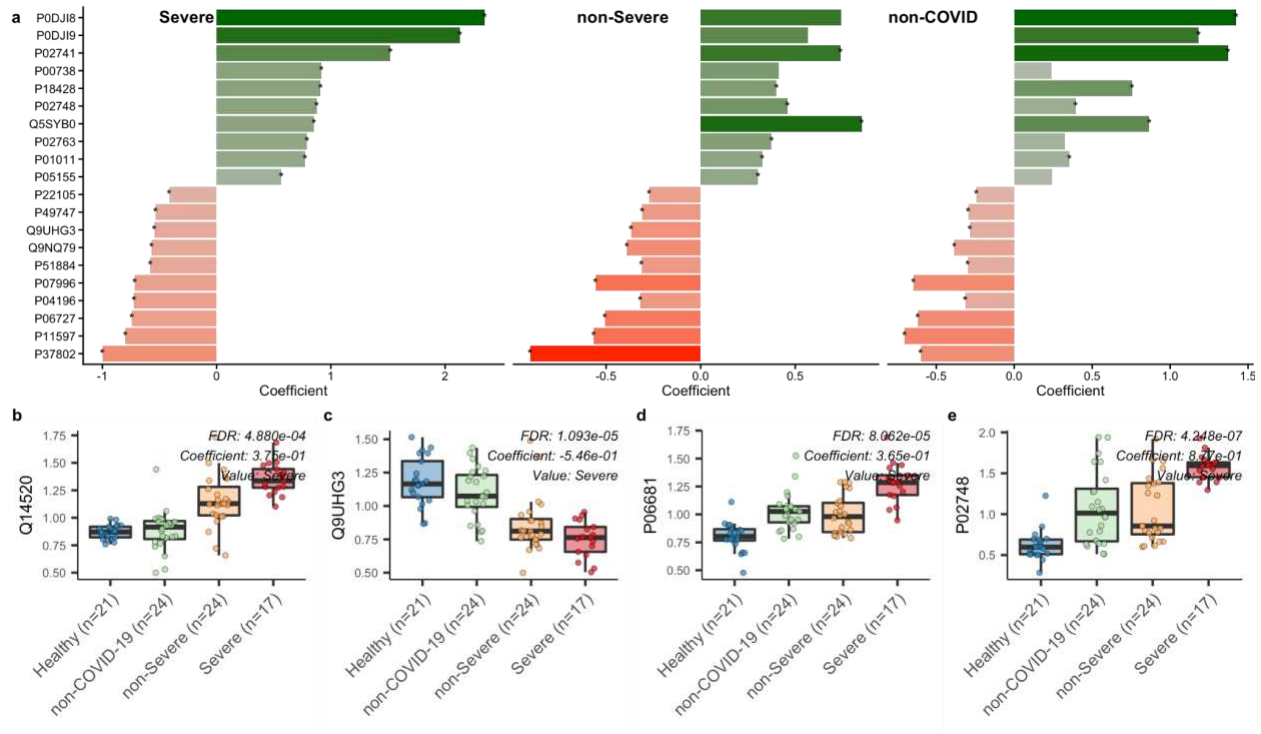


Figure 4: protein changes in COVID-19. a, 20 most significant metabolites with lowest p-value in comparison of severe group vs. healthy group are shown. Then, we show the corresponding changes in non-severe and non-COVID-19. b, c, d, and e show different patterns we observed among these associations.

Fibrosis/Keratinization

Lumican (protein ID P51884) regulates fibril assembly and stromal collagen matrix assembly⁹⁷. In mice models, it has been found that lumican is critical for host immune innate response^{98,99}, and its deficiency has been associated with cardiomyocyte hypertrophy¹⁰⁰. In a study of Nepalese children, lumican levels were negatively associated with levels of α -1-acid glycoprotein¹⁰¹, an acute phase protein that increases during inflammation, infection, or injury to tissues¹⁰². We found that lumican levels were depleted in the COVID-19 groups compared to the non-COVID-19 and healthy groups and is also associated with some glycerophospholipids in block-wise association testing. Thus, lower lumican levels may act as an additional biomarker of rampant inflammation in the COVID-19 groups. Alternatively, lower lumican may lead to disruption of the collagen and fibril assembly pathways as a consequence of infection.

Analyses of lung tissue from mechanically ventilated or recently deceased patients with severe COVID-19 revealed high levels of inflammatory infiltrate and fibrotic markers indicative of extensive epithelial and alveolar damage^{103,104}. Identification of additional biomarkers could facilitate diagnosing the severity of lung injury prior to the induction of respiratory failure. Our analysis identified significant alterations in processes that maintain cell or tissue structure, including enrichment of proteins involved in the keratinization pathway (**Fig. 5e**). Keratins play a vital role in both maintaining the structural integrity of the epithelium and promoting intracellular signaling to mediate wound healing. Cytoskeletal remodeling by keratin intermediate filaments can occur under excessive shear stress or in response to hypoxia^{105–107}. The significance of

keratinization within the context of COVID-19 has yet to be investigated but may provide insights into the extent of lung damage that occurs in severe cases.

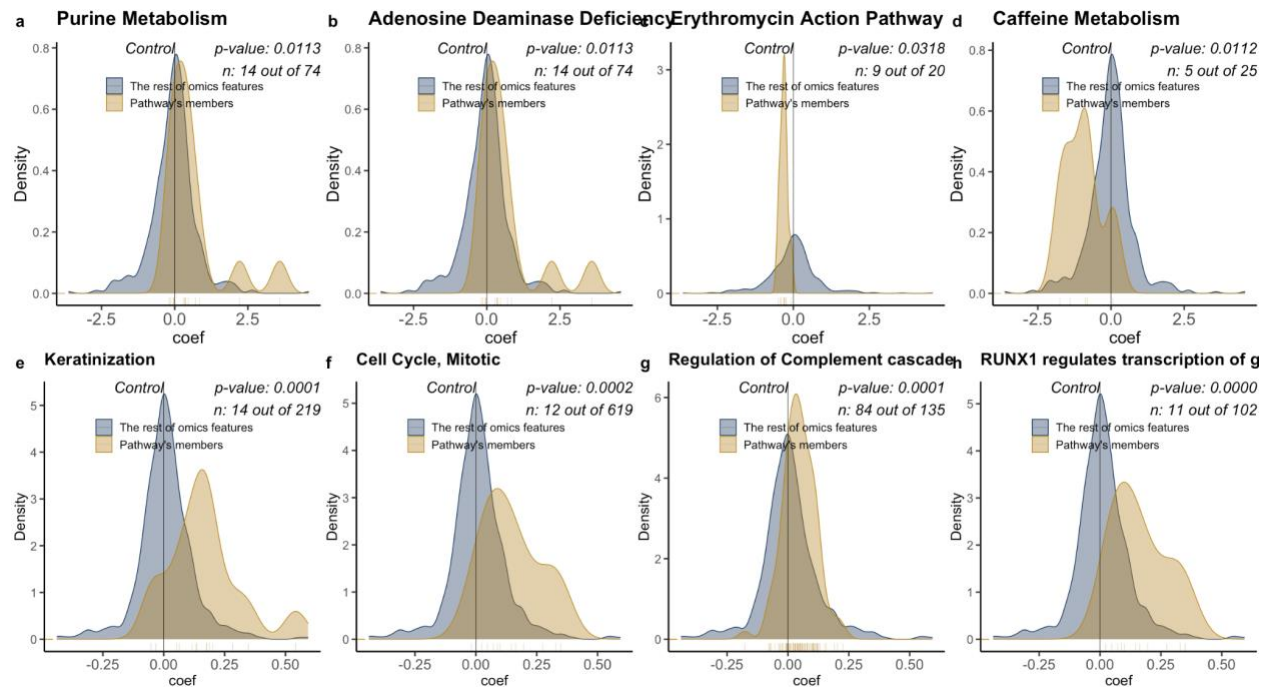


Figure 5: Enrichment pathway between severe group and healthy group. a-d, metabolite profiles are used to perform pathway enrichment analysis. Each metabolite assigned a rank based on coefficient from testing severe group vs. healthy group using generalized linear models. We applied our *deepath* tool with a Wilcoxon signed rank test and HMDB database¹⁰⁸ as the reference for pathways. e-h, protein profiles are used to perform pathway enrichment analysis. Each protein was assigned a rank based on coefficient from testing severe group vs. healthy group using generalized linear models. We applied *deepath* with the Wilcoxon signed rank test and used the Reactome database¹⁰⁹ with Physical Entity (PE) class for Uniprot to all levels of the pathway hierarchy mapping file.

Correlations between clinical data, omics data, and health outcome

We next examined associations between clinical biomarkers from panel tests and metabolites and proteins from the omics data in the context of COVID-19 severity.

Glucose: Severe COVID-19 patients had significantly higher levels of glucose compared to non-COVID-19 patients (*coefficient* = 0.497, *p-value* = -1.2E-05) matching previous studies^{110–113}. While the direct impact of COVID-19 infection on glucose levels remains to be elucidated, inflammation may be responsible for the differences in glucose levels observed between the groups. Okin and Medzhitov found that sustained inflammation can lead to elevated glucose levels in the plasma¹¹⁴. Alternatively, IFN-gamma production in response to viral infection has been shown to induce insulin resistance¹¹⁵ and subsequent higher glucose levels.

In addition to its association with COVID-19 severity, we also found some correlations between glucose and metabolites such as citrulline and uridine. Both of these metabolites were severely depleted in the COVID-19 groups compared to the non-COVID-19 group. While it is unclear if there is a biological connection to these relationships, the metabolism of glucose has been linked with the metabolism of uridine and citrulline. Specifically, uridine has been shown to induce

glucose uptake by skeletal muscles¹¹⁶ and increased levels of citrulline in the plasma is associated with a reduction in glucose production¹¹⁷. Regardless of the biological significance, the correlation of uridine and citrulline with glucose points towards these metabolites being modest candidates for COVID-19 severity biomarkers.

C-Reactive Protein (CRP): We found CRP levels to be lower in the non-severe COVID-19 group compared to the non-COVID-19 group (*coefficient* = -1.15 *p-value* = 0.0001), whereas an opposite trend appears when compared to the severe COVID-19 group. This reinforces what has been established in previous studies^{112,118–121}. CRP plays a critical role in inflammation and response to infection via the complement pathway and cytokine production^{122–124}; thus, our finding of increased CRP levels in the severe group is expected.

Similar to glucose, CRP has some metabolic and protein correlates which may be able to serve as novel biomarkers for COVID-19 severity. Specifically, CRP is positively correlated with kynurenine and lipopolysaccharide-binding protein (LBP) (**SFig. 3**). Kynurenine as a positive correlate of CRP is expected since it is involved in inflammation and immune activation in various disease contexts^{125,126}. Additionally, within COVID-19, kynurenine has been found to be positively correlated with proinflammatory cytokines^{127,128}, and the kynurenine pathway seems to be activated in COVID-19 patients¹²⁹.

The correlation between CRP and LBP is likely a product of these molecules' role in inflammation. LBP has been shown to be increased in patients with inflammatory conditions like systemic inflammatory response syndrome¹³⁰. Additionally, previous studies have shown that LBP is associated with inflammation markers, including CRP, in patients who have undergone hemodialysis¹³¹ and in patients with conditions such as acute respiratory distress syndrome and inflammatory bowel disease^{132,133}.

Monocyte: We found monocyte levels to be significantly decreased in the severe COVID-19 group (*coefficient* = -0.25 *p-value* = 0.04). Based on a meta-analysis of COVID-19 studies involving severe and non-severe patients, lower monocyte counts have been observed as part of a larger trend of immune dysregulation¹³⁴.

Salicylate: When investigating associations across all groups, we observed a positive relationship between Monocyte counts and Salicylate (*coefficient* = 1.64 *p-value* = 0.0001). Salicylate is commonly found in non-steroidal anti-inflammatory drugs, including aspirin. Sodium salicylate has been shown to have potent effects on limiting monocyte migration, expression of inflammatory cytokines, and preventing proliferation^{135–137}. Whether or not there is a prophylactic benefit to administering aspirin to limit COVID-19 infection is currently the subject of debate^{138–140}. A positive association between salicylate and monocyte counts seems to indicate that other factors, either directly related to infection or other forms of treatment, are responsible for the decrease in monocytes observed in patients with severe COVID-19.

Sphingomyelin: An additional positive correlation was found between Monocyte counts and sphingomyelin (*coefficient* = 0.91 *p-value* = 0.002). Sphingolipids are a class of membrane-associated molecules that play an important role in cell-to-cell interactions and intracellular signaling within the immune system¹⁴¹. Sphingomyelin is cleaved by sphingomyelinases to produce ceramide which induces a signaling cascade that can lead to the differentiation, proliferation, apoptosis, or cytokine secretion by select immune cell populations. It was recently

shown that neutral sphingomyelinase 2 can cause monocyte migration and secretion of inflammatory cytokines in response to soluble TNF- α ¹⁴². Elevated TNF- α has been associated with both obesity and severe cases of COVID-19¹⁴³. Most of the COVID-19 biomarkers and their related pathways reported in our study are novel, and some have already been discussed previously¹⁴⁴ (**Supplementary Table 1**).

Group-level correlations between proteins and metabolites

Block-wise association testing was performed to find associations among clusters of metabolites and proteins (**Methods**). Block 20 in particular shows that 5-methyluridine, citrulline, choline, and uridine are jointly associated with the C4b-binding protein chains (alpha and beta) and Vitamin-K dependent protein S. Inflammation may explain the observed association between the proteins and most of the metabolites in block 20. The C4b-binding protein is involved with the inhibition of the classical and lectin complement system pathways^{145,146}. The Vitamin-K dependent protein S complexes with the C4b-binding protein and can modulate the complement regulation activities of C4b-binding protein¹⁴⁷. Collectively, both of these proteins co-occur due to their involvement in inflammation via complement system regulation⁸⁰. Similarly, there are links between inflammation and three of the four metabolites in the block. Low plasma citrulline levels are known to be associated with systemic inflammation²⁴, and uridine is linked with anti-inflammatory effects in an animal model¹⁸. Furthermore, choline is known to be inversely associated with inflammatory marker levels¹⁴⁸.

Deep learning techniques accurately predict disease severity

We used four machine learning (ML) approaches for disease severity prediction, including deep neural networks (DNN), k-nearest neighbors (KNN), Random Forest (RF), and Logistic Regression (LR) (**SFig. 4**). We have compared the performance of all the models using precision, recall, F-1, and accuracy metrics (**Fig. 6**). Accuracy highlights the proportion of true positive in the sum of true positive and false positive. The precision, recall, and f-1 score deals with the false negative, which portrays the versatility and robustness of the model.

DNN outperformed all other methods with an accuracy score of 81.78% when trained using the metabolomics data and clinical data (e.g., age and gender). The other ML processes that were used had accuracy in the range of 60% - 70%. A higher F-1 score suggests better efficiency of models, which in turn means that the number of false positives is less, thus better prediction.

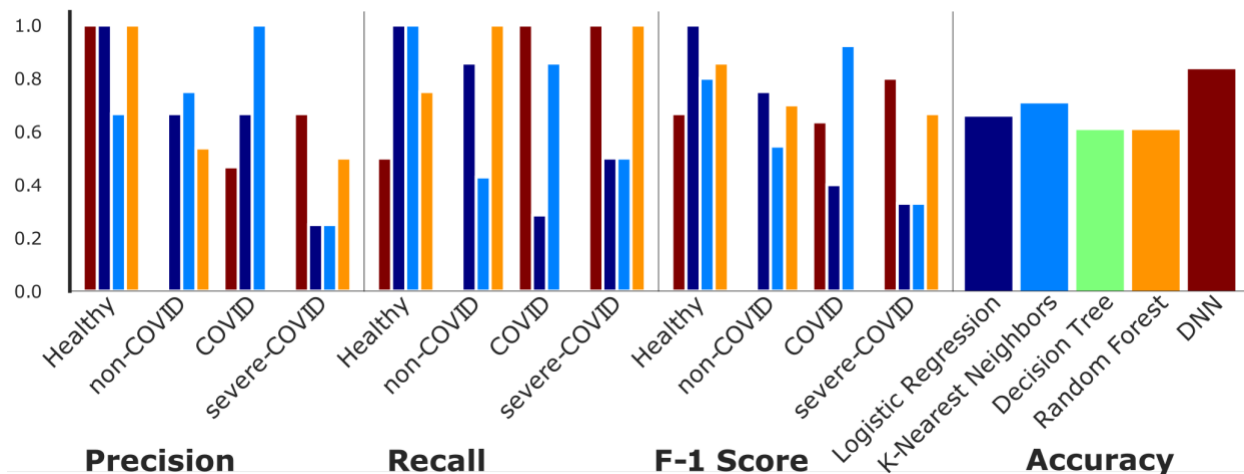


Figure 6: Precision, Recall, and F-1 Score for various ML techniques used for the various groups. The precision of DNN outperforms other methods since it can detect Severe-COVID-19 better than the rest of the methods. We also see the same trend in recall since it correctly detects the severe-COVID-19 and COVID-19 with a higher degree of certainty. A model with a low false positive case is better when used for prediction.

Discussion

Integrative analysis of multi-omics enables a more accurate and comprehensive understanding of biological activities at the molecular level. Utilizing omics data can provide a finer resolution for identifying the specific molecules or processes that distinguish severe cases of COVID-19. This analysis holds the potential to not only improve our understanding of COVID-19 pathogenesis, but may also lead to improved diagnostic and therapeutic avenues. In this study we used protein and metabolite profiles from a cohort of donors with varying COVID-19 and health statuses to measure the changes across groups after appropriate adjustment for data properties such as zero inflation and confounding factors. We followed up with pathway enrichment analysis to provide context at a functional level and then investigated associations between significantly different omics features (i.e. protein and metabolite) and clinical information (metabolic panel and complete blood count). Lastly, we utilized our findings to investigate the predictive potential of several machine learning algorithms benchmarked against conventional logistic regression to determine which model is best suited for diagnosing the status of COVID-19 infections.

Phenotype association testing revealed significantly altered proteins and metabolites based on health status (**Supplementary Table 2**). We found extensive evidence for systemic dysregulation of metabolic processes that may contribute to the varied clinical manifestations observed in severe cases of COVID-19 (**Supplementary Table 3**). This was reflected in aberrant levels of basic organic molecules that influence viral replication and immune responses, such as cytosine and BCAAs (**Fig. 3d**). Additionally, although COVID-19 is primarily considered a respiratory tract infection, multiple organs including the gastrointestinal tract, liver, and kidneys can be heavily impacted during infection. Determining if additional organ damage is a product of direct infection, overactive inflammation, or a side effect of treatment remains the subject of investigation. For example, we found lower levels of citrulline in severe cases of COVID-19 (**Fig. 3d**), indicative of gastrointestinal dysfunction. Whether or not the significant reduction in citrulline is caused by a loss in intestinal enterocytes from direct infection or an alternative mechanism highlights the merit of pursuing these questions. Similarly, over half of the ten proteins and metabolites outlined in our

results are directly implicated in severe liver impairment (**Fig 2**). While this may potentially be explained by the aggressive use of antiviral and anti-inflammatory drugs that possess hepatic toxicity, the influence of pre-existing medical conditions and inflammatory factors cannot be discounted.

Pathway enrichment analysis identified pro-inflammatory elements within the circulation as potential etiological agents for multi-organ damage (**Fig. 5**). Overactivation of platelets and neutrophils contributes not only to thrombotic events but may also give rise to tissue damage in a low oxygen environment. Similarly, persistent activation of the complement system by components of SARS-CoV-2 or other factors may further increase damage to vital organs. By employing novel computational tools that handle complex multi-omics data, we were able to highlight metabolites, proteins, and pathways that distinguish COVID-19 based on infection status and severity. Our analyses identified new potential biomarkers or therapeutic targets worthy of further investigation. Inclusion of additional paired omics data including metagenomics, single-cell RNA sequencing, transcriptomics, and viral genomics, can provide a better picture of disease pathogenesis and host response to infection and co-infections. Moving forward, a well-designed longitudinal measurement of omics can provide a deeper understanding of both the short and long term effects of infectious diseases, including COVID-19.

Materials and methods

Study design and data: All methods were performed in accordance with the relevant guidelines and regulations. The data were from 28 patient cases with severe COVID-19² who were matched to controls based on certain epidemiological features. The matched controls included 28 healthy persons, 25 patients without COVID-19 exhibiting clinically similar signs as COVID-19 patients, and 25 patients with non-severe COVID-19 (**Fig. 1a**). Serum samples were obtained a few days after admittance into the hospital. For a small number of cases, serum samples were collected at a later stage of the disease. Twelve clinical measurements were obtained for the COVID-19 and the non-COVID-19 groups but not for the healthy groups; the measurements included a complete blood cell count panel as well as a comprehensive metabolic panel of tests. 941 metabolites and 894 proteins were quantified from the 83 serum samples. Metabolite- and protein- profiling were performed using ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) and stable isotope-labeled proteomics strategy TMTpro (16plex)³.

Data preparation: The study's aim was to predict health outcome status and discover important features using proteomics and metabolomics data and clinical information. Before feeding the publicly available dataset², various data cleaning steps were taken into machine learning algorithms. The steps included removing any column of 942 metabolites and 640 proteins with missing values across samples and imputing missing BMIs for healthy individuals based on the optimal BMI of Chinese people¹⁴⁹. We removed other clinical information like platelet counts, etc., since they were only present for the unhealthy patients and not for the healthy subjects. Removing missing values resulted in a dataset that included 404 metabolites and 374 proteins. We also normalized the dataset based on the min-max function available in the *sklearn* package¹⁵⁰. The cleaned and normalized dataset was split into training (80% of data) and

testing (20% of data) subsets to train and test various prediction algorithms, including Logistic Regression, Random Forest, K-Nearest Neighbor, Decision Tree, and Artificial Neural Network.

Omics community detection and prioritizing metadata: we applied *omeClust*⁴ to detect underlying clusters from metabolites profiles and protein profiles independently. *omeClust*, in addition to detecting clusters (communities), also provides an enrichment score for each metadata to measure the potential influence of each metadata on detected structure (clusters). *omeClust* first discretizes metadata; and then calculates enrichment score as normalized mutual information between cluster labels and discretized metadata.

Multivariate association testing

We used multivariate association testing with considering noisy, sparse (zero-inflated), high-dimensional, and extremely non-normal data.

Pathway enrichment analysis: Enrichment analyses were performed using the *deepath* package¹⁵¹. *deepath* assigns an importance score (i.e., coefficient score from the CPLM model) to each omics feature (e.g., proteins, metabolites) and performs statistical tests (Wilcox sum rank) between rank of feature score in a given pathway against all ranks to calculate a p-value for the null hypothesis. There is no difference between the distribution of score of features with the pathways of interest vs. all other features in the study. We use an alpha level of 0.05 for significance. *deepath* is an open-source software implemented as an R package with code, tutorials, and documentation at <https://github.com/omicsEye/deepath>. The result for each association contains the identified pathway, members of the pathways, number of observed members in the study (n), and the total number of pathway members in the database (HMDB database for metabolite pathways¹⁰⁸ and Reactome database¹⁵² for proteins), used for the analysis, p-value, q-value from Benjamini-Hochberg FDR correction (q = 0.25).

Machine learning algorithms

Machine learning (ML) algorithms like neural networks are often considered to be a black box because of their inability to provide a simple and straightforward explanation of their predictions. Nevertheless, this prediction model generally exceeds simple linear models or decision trees and random forest predictions. Yet, such simple models are still preferred in the field of medical science due to their simplicity and interpretability^{153–155}. Many studies have been targeted to build and execute model-agnostic interpretability tools^{156–158}. We use the term feature importance to explain how important a feature is to the model's predictive performance. The most well-known approach is using permutation importance which was introduced by Breiman¹⁵⁹. Using permutation importance, we have quantified the importance score of features for predicting health outcomes. The features consist of metabolomics, proteomics, and clinical info for all the patients considered in the study. The samples were labeled into four labels based on the health status; Healthy, non-COVID-19, COVID-19, and severe-COVID-19. We predicted the health status of a patient based on the various proteomics and metabolomics data and found the importance of each feature inside the prediction model (**SFig. 5**).

Data from a real-world scenario are never flawless. The medical records and the clinical information for the patients affected by COVID-19 were also no exception. The data contained many values which were lost due to machine or human error. The data thus was cleaned of all the unknown values and also the null values by dropping the instance with the missing feature

value. Serum panel data tends to have a significant difference between the maximum and minimum value, to rectify this issue, we used normalization. Normalization is a scaling-down transformation in data where the difference between min and max values is significantly big. The data were normalized using a min-max scaler function which is present in the *sklearn* package.

Decision Tree:

Decision Tree follows a flow-chart-like structure where the nodes are the features, the branches are the decision rules, and the leaves are the outcomes. Decision Tree is a supervised learning method that utilizes a divide and conquer approach; it selects the best attribute using Information Gain and then divides the dataset into a subset. This division is performed repeatedly until the method reaches a child node which satisfies the condition of no remaining attributes or no more remaining instances.

KNN- K Nearest Neighbor:

K-Nearest Neighbor (KNN) is a supervised machine learning technique that is dependent on the training dataset. The K, in KNN, stands for a user-defined number. This algorithm assumes that data points with similar features reside in close proximity to each other. Proximity is generally calculated in the form of euclidean distances among points. In a classification problem such as ours, the distances between the test data points and the training data points are calculated, sorted, and stored in a table¹⁶⁰. Then, the mode of the labels of K- nearest neighbors using the sorted table is given as an output.

Random Forest:

Random Forest is a supervised algorithm that randomly selects a subset of the training dataset and creates a decision tree on the subset; it then carries out a vote to predict the class of the test data points.

Logistic Regression:

A predominant part of published propensity results uses Logistic Regression (LR). Logistic regression is a very sought after technique because of its mathematical ability to produce probability in the range $[0,1]$ ¹⁶¹. Logistic regression uses a functional approach to estimate the probability of binary response based on input features. LR finds the best-fit parameters to a nonlinear function called sigmoid¹⁶².

Logistic regression models probability for a binary class, however, our health outcome variable has more than two classes. To address the binary class limitation of logistic regression, we used a 'newton-cg' solver. In our study, we use logistic regression as a baseline for the other methods.

Artificial Neural Network

Artificial neural network (ANN) is a type of machine learning architecture that mimics the working of the neurons located in the brain, and how they transfer information to learn new problems for the purpose of solving them¹⁶³. The inputs of ANN are fed in the input layers, which are passed through one or more hidden layers, which consist of neurons, where they are analyzed and processed to determine the output of the next layer. ANN uses a learning rule which correctly decides the weight and the bias of each neuron in the hidden layer and output layer. The power of ANN to determine and adapt the weight and bias dynamically makes it a

powerful tool to capture the various complex and non-linear relationships among the various features, which in turn facilitates classification and prediction of correct labels, thus increasing the accuracy and efficiency of the model^{164,165}.

The previous analysis² prioritized only 22 proteins and 7 metabolites. However, we have incorporated the full extent of the data since some were discarded due to missing values. Including a more extensive set of data and features in deep learning, the model brings out a more comprehensive hidden complex relationships among all the proteins and metabolites. This enables a more accurate understanding, which leads to a better prioritization of metabolites and proteins that were further studied to show how they affect a patient's health status. The importance score of metabolites and proteins generated by the model is based on their degree of influence on the result.

The only ML model which was used in the above said paper was Random Forest (RF). Still, in our evaluation, we used deep neural network (DNN), random forest (RF), decision tree (DT), k-nearest neighbors (KNN) and finally logistic regression (LR). We have made a thorough comparison of all the methods using various metrics like accuracy, precision, recall, and F-1 score to evaluate the performance of each model. The accuracy of DNN comes higher than all other methods that were considered. We then used the DNN model for importance evaluation that leads to the discovery of numerous metabolites and proteins which when done, a thorough study shows a relationship with covid and health status.

The different evaluation metrics used in machine learning section are as follows:

$$Accuracy = \frac{TN + TP}{FN + FP + TN + TP}$$
$$Precision = \frac{TP}{FP + TP}$$
$$Recall = \frac{TP}{FN + TP}$$
$$F - 1 \text{ Score} = \frac{2 * Precision * Recall}{Precision + Recall}$$

Authors' contributions

A.R. conceived the project; B.M., A.G., and A.R. interpret associations and patterns in biological context. R.C., A.R., and A.R. performed machine learning analysis. A.R. performed association testing and pathway enrichment analysis. All authors wrote the manuscript and discussed the results and commented on the paper.

Funding

This work was supported by the National Science Foundation grant DEB-2028280 and DEB-2109688 to AR and KAC. We also thank Amazon AWS Diagnostic Development Initiative (DDI) for supporting our work by research credit.

Competing interests

The authors declare no competing financial interests.

References

1. Rahnavard, A. *et al.* Epidemiological associations with genomic variation in SARS-CoV-2. (2021) doi:10.21203/rs.3.rs-537082/v1.
2. Shen, B. *et al.* Proteomic and Metabolomic Characterization of COVID-19 Patient Sera. *Cell* **182**, 59–72.e15 (2020).
3. Li, J. *et al.* TMTpro reagents: a set of isobaric labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. *Nat. Methods* **17**, 399–404 (2020).
4. Rahnavard, A. *et al.* Omics community detection using multi-resolution clustering. *Bioinformatics* (2021) doi:10.1093/bioinformatics/btab317.
5. Blasco, H. *et al.* The specific metabolome profiling of patients infected by SARS-COV-2 supports the key role of tryptophan-nicotinamide pathway and cytosine metabolism. *Sci. Rep.* **10**, 16824 (2020).
6. Danchin, A. & Marlière, P. Cytosine drives evolution of SARS-CoV-2. *Environ. Microbiol.* **22**, 1977–1985 (2020).
7. Yamamoto, T. *et al.* Biochemistry of uridine in plasma. *Clin. Chim. Acta* **412**, 1712–1724 (2011).
8. Connolly, G. P. & Duley, J. A. Uridine and its nucleotides: biological actions, therapeutic potentials. *Trends Pharmacol. Sci.* **20**, 218–225 (1999).
9. Greenhill, C. Metabolism: Liver and adipose tissue control uridine biosynthesis. *Nature reviews. Endocrinology* vol. 13 249 (2017).
10. Chau, T.-N. *et al.* SARS-associated viral hepatitis caused by a novel coronavirus: Report of three cases. *Hepatology* vol. 39 302–310 (2004).
11. Yang, Z., Xu, M., Yi, J.-Q. & Jia, W.-D. Clinical characteristics and mechanism of liver damage in patients with severe acute respiratory syndrome. *Hepatobiliary Pancreat. Dis. Int* **4**, 60–63 (2005).

12. Cai, Q. *et al.* COVID-19 in a designated infectious diseases hospital outside Hubei Province, China. *Allergy* vol. 75 1742–1752 (2020).
13. Chen, N. *et al.* Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* vol. 395 507–513 (2020).
14. Fan, Z. *et al.* Clinical features of COVID-19-related liver damage. *Clin. Gastroenterol. Hepatol.* (2020).
15. Wang, D. *et al.* Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. *JAMA* **323**, 1061–1069 (2020).
16. Wang, Z., Yang, B., Li, Q., Wen, L. & Zhang, R. Clinical Features of 69 Cases With Coronavirus Disease 2019 in Wuhan, China. *Clin. Infect. Dis.* **71**, 769–777 (2020).
17. Zhang, B. *et al.* Clinical characteristics of 82 death cases with COVID-19. *MedRxiv* (2020).
18. Cicko, S. *et al.* Uridine supplementation exerts anti-inflammatory and anti-fibrotic effects in an animal model of pulmonary fibrosis. *Respir. Res.* **16**, 105 (2015).
19. Alon, R. *et al.* Leukocyte trafficking to the lungs and beyond: lessons from influenza for COVID-19. *Nat. Rev. Immunol.* **21**, 49–64 (2021).
20. Linden, J., Koch-Nolte, F. & Dahl, G. Purine Release, Metabolism, and Signaling in the Inflammatory Response. *Annu. Rev. Immunol.* **37**, 325–347 (2019).
21. Cekic, C. & Linden, J. Purinergic regulation of the immune system. *Nat. Rev. Immunol.* **16**, 177–192 (2016).
22. Bowser, J. L., Phan, L. H. & Eltzschig, H. K. The Hypoxia-Adenosine Link during Intestinal Inflammation. *J. Immunol.* **200**, 897–907 (2018).
23. Barmore, W., Azad, F. & Stone, W. L. Physiology, Urea Cycle. in *StatPearls* (StatPearls Publishing, 2020).
24. Uzzan, M. *et al.* Patients with COVID-19 present with low plasma citrulline concentrations that associate with systemic inflammation and gastrointestinal symptoms. *Dig. Liver Dis.*

(2020) doi:10.1016/j.dld.2020.06.042.

25. Crenn, P., Messing, B. & Cynober, L. Citrulline as a biomarker of intestinal failure due to enterocyte mass reduction. *Clin. Nutr.* **27**, 328–339 (2008).
26. Liang, W. *et al.* Diarrhoea may be underestimated: a missing link in 2019 novel coronavirus. *Gut* **69**, 1141–1143 (2020).
27. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* **94**, (2020).
28. Zhang, H. *et al.* Clinical characteristics of coronavirus disease (COVID-19) patients with gastrointestinal symptoms: A report of 164 cases. *Dig. Liver Dis.* (2020) doi:10.1016/j.dld.2020.04.034.
29. Holeček, M. Branched-chain amino acids in health and disease: metabolism, alterations in blood plasma, and as supplements. *Nutr. Metab.* **15**, 33 (2018).
30. Fischer, J. E. *et al.* The role of plasma amino acids in hepatic encephalopathy. *Surgery* **78**, 276–290 (1975).
31. Rodney, S. & Boneh, A. Amino Acid Profiles in Patients with Urea Cycle Disorders at Admission to Hospital due to Metabolic Decompensation. in *JIMD Reports – Case and Research Reports, 2012/6* (eds. Zschocke, J., Gibson, K. M., Brown, G., Morava, E. & Peters, V.) 97–104 (Springer Berlin Heidelberg, 2013).
32. Schauder, P., Matthaei, D., Henning, H. V., Scheler, F. & Langenbeck, U. Blood levels of branched-chain amino acids and α -ketoacids in uremic patients given keto analogues of essential amino acids. *Am. J. Clin. Nutr.* **33**, 1660–1666 (1980).
33. Garibotto, G. *et al.* Peripheral metabolism of branched-chain keto acids in patients with chronic renal failure. *Miner. Electrolyte Metab.* **19**, 25–31 (1993).
34. Cano, N. J. M., Fouque, D. & Lerverve, X. M. Application of branched-chain amino acids in human pathological states: renal failure. *J. Nutr.* **136**, 299S–307S (2006).

35. Raza, A., Estepa, A., Chan, V. & Jafar, M. S. Acute Renal Failure in Critically Ill COVID-19 Patients With a Focus on the Role of Renal Replacement Therapy: A Review of What We Know So Far. *Cureus* **12**, e8429 (2020).
36. Pongsavee, M. Effect of sodium benzoate preservative on micronucleus induction, chromosome break, and Ala40Thr superoxide dismutase gene mutation in lymphocytes. *Biomed Res. Int.* **2015**, 103512 (2015).
37. Enns, G. M. *et al.* Survival after treatment with phenylacetate and benzoate for urea-cycle disorders. *N. Engl. J. Med.* **356**, 2282–2292 (2007).
38. Husson, M.-C. *et al.* Efficacy and safety of i.v. sodium benzoate in urea cycle disorders: a multicentre retrospective study. *Orphanet J. Rare Dis.* **11**, 127 (2016).
39. Badenhorst, C. P. S., Erasmus, E., van der Sluis, R., Nortje, C. & van Dijk, A. A. A new perspective on the importance of glycine conjugation in the metabolism of aromatic acids. *Drug Metab. Rev.* **46**, 343–361 (2014).
40. Kubota, K. & Ishizaki, T. Dose-dependent pharmacokinetics of benzoic acid following oral administration of sodium benzoate to humans. *Eur. J. Clin. Pharmacol.* **41**, 363–368 (1991).
41. Lennerz, B. S. *et al.* Effects of sodium benzoate, a widely used food preservative, on glucose homeostasis and metabolic profiles in humans. *Mol. Genet. Metab.* **114**, 73–79 (2015).
42. Yilmaz, B. & Karabay, A. Z. Food Additive Sodium Benzoate (NaB) Activates NFκB and Induces Apoptosis in HCT116 Cells. *Molecules* **23**, (2018).
43. Liu, T., Zhang, L., Joo, D. & Sun, S.-C. NF-κB signaling in inflammation. *Signal Transduct Target Ther* **2**, (2017).
44. Pahan, K. Immunomodulation of experimental allergic encephalomyelitis by cinnamon metabolite sodium benzoate. *Immunopharmacol. Immunotoxicol.* **33**, 586–593 (2011).
45. Choi-Miura, N.-H. *et al.* Purification and Characterization of a Novel Hyaluronan-Binding Protein (PHBP) from Human Plasma: It Has Three EGF, a Kringle and a Serine Protease

- Domain, Similar to Hepatocyte Growth Factor Activator. *Journal of Biochemistry* vol. 119 1157–1165 (1996).
46. Römisch, J., Feussner, A., Vermöhlen, S. & Stöhr, H. A. A protease isolated from human plasma activating factor VII independent of tissue factor. *Blood Coagul. Fibrinolysis* **10**, 471–479 (1999).
 47. Mambetsariev, N. *et al.* Hyaluronic Acid binding protein 2 is a novel regulator of vascular integrity. *Arterioscler. Thromb. Vasc. Biol.* **30**, 483–490 (2010).
 48. Ware, L. B. & Matthay, M. A. The acute respiratory distress syndrome. *N. Engl. J. Med.* **342**, 1334–1349 (2000).
 49. Wygrecka, M., Markart, P., Fink, L., Guenther, A. & Preissner, K. T. Raised protein levels and altered cellular expression of factor VII activating protease (FSAP) in the lungs of patients with acute respiratory distress syndrome (ARDS). *Thorax* **62**, 880–888 (2007).
 50. Parahuleva, M. S. *et al.* Factor VII Activating Protease Expression in Human Platelets and Accumulation in Symptomatic Carotid Plaque. *J. Am. Heart Assoc.* **9**, e016445 (2020).
 51. Parahuleva, M. S. *et al.* Circulating factor VII activating protease (FSAP) is associated with clinical outcome in acute coronary syndrome. *Circ. J.* **76**, 2653–2661 (2012).
 52. Hanson, E. *et al.* Plasma factor VII-activating protease antigen levels and activity are increased in ischemic stroke. *J. Thromb. Haemost.* **10**, 848–856 (2012).
 53. Byskov, K. *et al.* Factor VII activating protease (FSAP) regulates the expression of inflammatory genes in vascular smooth muscle and endothelial cells. *Atherosclerosis* **265**, 133–139 (2017).
 54. Parahuleva, M. S. *et al.* Regulation of monocyte/macrophage function by factor VII activating protease (FSAP). *Atherosclerosis* **230**, 365–372 (2013).
 55. Tschantz, W. R., Digits, J. A., Pyun, H. J., Coates, R. M. & Casey, P. J. Lysosomal prenylcysteine lyase is a FAD-dependent thioether oxidase. *J. Biol. Chem.* **276**, 2321–2324 (2001).

56. Zhang, L., Tschantz, W. R. & Casey, P. J. Isolation and characterization of a prenylcysteine lyase from bovine brain. *J. Biol. Chem.* **272**, 23354–23359 (1997).
57. Herrera-Marcos, L. V. *et al.* Prenylcysteine oxidase 1, a pro-oxidant enzyme of low density lipoproteins. *Front. Biosci.* **23**, 1020–1037 (2018).
58. Mancone, C. *et al.* Proteomic analysis of human very low-density lipoprotein by two-dimensional gel electrophoresis and MALDI-TOF/TOF. *Proteomics* **7**, 143–154 (2007).
59. Wouters, M. M., Neefs, J.-M., Kerchove d'Exaerde, A. de, Vanderwinden, J.-M. & Smans, K. A. Downregulation of two novel genes in SI/SId and W(LacZ)/Wv mouse jejunum. *Biochem. Biophys. Res. Commun.* **346**, 491–500 (2006).
60. Peng, M. *et al.* Primary coenzyme Q deficiency in Pdss2 mutant mice causes isolated renal disease. *PLoS Genet.* **4**, e1000061 (2008).
61. Mistry, P. K. *et al.* Glucocerebrosidase gene-deficient mouse recapitulates Gaucher disease displaying cellular and molecular dysregulation beyond the macrophage. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 19473–19478 (2010).
62. Goh, Y. P. S. *et al.* Eosinophils secrete IL-4 to facilitate liver regeneration. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 9914–9919 (2013).
63. Schaefer, E. A. K., Arvind, A., Bloom, P. P. & Chung, R. T. Interrelationship Between Coronavirus Infection and Liver Disease. *Clin. Liver Dis.* **15**, 175–180 (2020).
64. Snodgrass, R. G. & Brüne, B. Regulation and Functions of 15-Lipoxygenases in Human Macrophages. *Front. Pharmacol.* **10**, 719 (2019).
65. Singh, N. K. & Rao, G. N. Emerging role of 12/15-Lipoxygenase (ALOX15) in human pathologies. *Prog. Lipid Res.* **73**, 28–45 (2019).
66. Smith, R. J., Justen, J. M., Nidy, E. G., Sam, L. M. & Bleasdale, J. E. Transmembrane signaling in human polymorphonuclear neutrophils: 15 (S)-hydroxy-(5Z, 8Z, 11Z, 13E)-eicosatetraenoic acid modulates receptor agonist-triggered cell activation. *Proceedings of the National Academy of Sciences* **90**, 7270–7274 (1993).

67. Takata, S. *et al.* 15-Hydroxyeicosatetraenoic acid inhibits neutrophil migration across cytokine-activated endothelium. *Am. J. Pathol.* **145**, 541–549 (1994).
68. de Lucena, T. M. C., da Silva Santos, A. F., de Lima, B. R., de Albuquerque Borborema, M. E. & de Azevêdo Silva, J. Mechanism of inflammatory response in associated comorbidities in COVID-19. *Diabetes Metab. Syndr.* **14**, 597–600 (2020).
69. Liu, Y. *et al.* MMP-2 and MMP-9 contribute to the angiogenic effect produced by hypoxia/15-HETE in pulmonary endothelial cells. *J. Mol. Cell. Cardiol.* **121**, 36–50 (2018).
70. Li, F., You, Y. & Zhu, H. 15-HETE protects pulmonary artery smooth muscle cells against apoptosis via SIRT1 regulation during hypoxia. *Biomed. Pharmacother.* **108**, 325–330 (2018).
71. Ackermann, M. *et al.* Pulmonary Vascular Endothelialitis, Thrombosis, and Angiogenesis in Covid-19. *N. Engl. J. Med.* **383**, 120–128 (2020).
72. McFadyen, J. D., Stevens, H. & Peter, K. The Emerging Threat of (Micro)Thrombosis in COVID-19 and Its Therapeutic Implications. *Circ. Res.* **127**, 571–587 (2020).
73. Hottz, E. D. *et al.* Platelet activation and platelet-monocyte aggregate formation trigger tissue factor expression in patients with severe COVID-19. *Blood* **136**, 1330–1341 (2020).
74. Middleton, E. A. *et al.* Neutrophil extracellular traps contribute to immunothrombosis in COVID-19 acute respiratory distress syndrome. *Blood* **136**, 1169–1179 (2020).
75. Akgun, E. *et al.* Proteins associated with neutrophil degranulation are upregulated in nasopharyngeal swabs from SARS-CoV-2 patients. *PLoS One* **15**, e0240012 (2020).
76. Noris, M., Benigni, A. & Remuzzi, G. The case of complement activation in COVID-19 multiorgan impact. *Kidney Int.* **98**, 314–322 (2020).
77. Java, A. *et al.* The complement system in COVID-19: friend and foe? *JCI Insight* **5**, (2020).
78. Kim, A. H. J., Wu, X. & Atkinson, J. P. The beneficial and pathogenic roles of complement in COVID-19. *Cleve. Clin. J. Med.* (2020) doi:10.3949/ccjm.87a.ccc065.
79. Barrington, R., Zhang, M., Fischer, M. & Carroll, M. C. The role of complement in

- inflammation and adaptive immunity. *Immunol. Rev.* **180**, 5–15 (2001).
80. Markiewski, M. M. & Lambris, J. D. The role of complement in inflammatory diseases from behind the scenes into the spotlight. *Am. J. Pathol.* **171**, 715–727 (2007).
 81. Halili, M. A., Ruiz-Gómez, G., Le, G. T., Abbenante, G. & Fairlie, D. P. Complement component C2, inhibiting a latent serine protease in the classical pathway of complement activation. *Biochemistry* **48**, 8466–8472 (2009).
 82. Morgan, B. P. Regulation of the complement membrane attack pathway. *Crit. Rev. Immunol.* **19**, 173–198 (1999).
 83. Yu, J. *et al.* Direct activation of the alternative complement pathway by SARS-CoV-2 spike proteins is blocked by factor D inhibition. *Blood* **136**, 2080–2089 (2020).
 84. Holter, J. C. *et al.* Systemic complement activation is associated with respiratory failure in COVID-19 hospitalized patients. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 25018–25025 (2020).
 85. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13**, 397–406 (2014).
 86. Santiesteban-Lores, L. E. *et al.* A double edged-sword - The Complement System during SARS-CoV-2 infection. *Life Sci.* **272**, 119245 (2021).
 87. Poon, I. K. H., Patel, K. K., Davis, D. S., Parish, C. R. & Hulett, M. D. Histidine-rich glycoprotein: the Swiss Army knife of mammalian plasma. *Blood* **117**, 2093–2101 (2011).
 88. Gorgani, N. N., Parish, C. R., Easterbrook Smith, S. B. & Altin, J. G. Histidine-rich glycoprotein binds to human IgG and C1q and inhibits the formation of insoluble immune complexes. *Biochemistry* **36**, 6653–6662 (1997).
 89. Eggleton, P., Javed, M., Pulavar, D. & Sheldon, G. Immune Complexes. *eLS* 1–10 (2015) doi:10.1002/9780470015902.a0001118.pub2.
 90. Manderson, G. A. *et al.* Interactions of histidine-rich glycoprotein with immunoglobulins and proteins of the complement system. *Mol. Immunol.* **46**, 3388–3398 (2009).

91. Gorgani, N. N., Smith, B. A., Kono, D. H. & Theofilopoulos, A. N. Histidine-rich glycoprotein binds to DNA and Fc gamma RI and potentiates the ingestion of apoptotic cells by macrophages. *J. Immunol.* **169**, 4745–4751 (2002).
92. Jones, A. L., Poon, I. K. H., Hulett, M. D. & Parish, C. R. Histidine-rich glycoprotein specifically binds to necrotic cells via its amino-terminal domain and facilitates necrotic cell phagocytosis. *J. Biol. Chem.* **280**, 35733–35741 (2005).
93. Poon, I. K. H., Hulett, M. D. & Parish, C. R. Histidine-rich glycoprotein is a novel plasma pattern recognition molecule that recruits IgG to facilitate necrotic cell clearance via Fc gamma RI on phagocytes. *Blood* **115**, 2473–2482 (2010).
94. Poon, I. K. H., Parish, C. R. & Hulett, M. D. Histidine-rich glycoprotein functions cooperatively with cell surface heparan sulfate on phagocytes to promote necrotic cell uptake. *J. Leukoc. Biol.* **88**, 559–569 (2010).
95. Winiarska, A. *et al.* Decreased Levels of Histidine-Rich Glycoprotein in Advanced Lung Cancer: Association with Prothrombotic Alterations. *Dis. Markers* **2019**, 8170759 (2019).
96. Saito, H., Goodnough, L. T., Boyle, J. M. & Heimburger, N. Reduced histidine-rich glycoprotein levels in plasma of patients with advanced liver cirrhosis. Possible implications for enhanced fibrinolysis. *Am. J. Med.* **73**, 179–182 (1982).
97. Kao, W. W.-Y., Funderburgh, J. L., Xia, Y., Liu, C.-Y. & Conrad, G. W. Focus on molecules: lumican. *Exp. Eye Res.* **82**, 3–4 (2006).
98. Lohr, K. *et al.* Extracellular matrix protein lumican regulates inflammation in a mouse model of colitis. *Inflamm. Bowel Dis.* **18**, 143–151 (2012).
99. Wu, F. *et al.* A Novel Role of the Lumican Core Protein in Bacterial Lipopolysaccharide-induced Innate Immune Response *. *J. Biol. Chem.* **282**, 26409–26417 (2007).
100. Dupuis, L. E. *et al.* Lumican deficiency results in cardiomyocyte hypertrophy with altered collagen assembly. *J. Mol. Cell. Cardiol.* **84**, 70–80 (2015).
101. Lee, S. E. *et al.* Plasma Proteome Biomarkers of Inflammation in School Aged Children in

- Nepal. *PLoS One* **10**, e0144279 (2015).
102. Fournier, T., Medjoubi-N, N. & Porquet, D. Alpha-1-acid glycoprotein. *Biochim. Biophys. Acta* **1482**, 157–171 (2000).
103. Spadaro, S. *et al.* Markers of endothelial and epithelial pulmonary injury in mechanically ventilated COVID-19 ICU patients. *Crit. Care* **25**, 74 (2021).
104. Carsana, L. *et al.* Pulmonary post-mortem findings in a series of COVID-19 cases from northern Italy: a two-centre descriptive study. *Lancet Infect. Dis.* **20**, 1135–1140 (2020).
105. Felder, E. *et al.* Mechanical strain of alveolar type II cells in culture: changes in the transcellular cytokeratin network and adaptations. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **295**, L849–57 (2008).
106. Sivaramakrishnan, S., DeGiulio, J. V., Lorand, L., Goldman, R. D. & Ridge, K. M. Micromechanical properties of keratin intermediate filament networks. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 889–894 (2008).
107. Na, N., Chandel, N. S., Litvan, J. & Ridge, K. M. Mitochondrial reactive oxygen species are required for hypoxia-induced degradation of keratin intermediate filaments. *FASEB J.* **24**, 799–809 (2010).
108. Wishart, D. S. *et al.* HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* **46**, D608–D617 (2018).
109. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Research* vol. 46 D649–D655 (2018).
110. Ren, H. *et al.* Association of the insulin resistance marker TyG index with the severity and mortality of COVID-19. *Cardiovasc. Diabetol.* **19**, 58 (2020).
111. Wang, F. *et al.* CLINICAL CHARACTERISTICS OF 28 PATIENTS WITH DIABETES AND COVID-19 IN WUHAN, CHINA. *Endocr. Pract.* **26**, 668–674 (2020).
112. Gao, Y. *et al.* Diagnostic utility of clinical laboratory data determinations for patients with the severe COVID-19. *J. Med. Virol.* **92**, 791–796 (2020).

113. Chen, J., Wu, C., Wang, X., Yu, J. & Sun, Z. The Impact of COVID-19 on Blood Glucose: A Systematic Review and Meta-Analysis. *Front. Endocrinol.* **11**, 574541 (2020).
114. Okin, D. & Medzhitov, R. The Effect of Sustained Inflammation on Hepatic Mevalonate Pathway Results in Hyperglycemia. *Cell* **165**, 343–356 (2016).
115. Šestan, M. *et al.* Virus-Induced Interferon- γ Causes Insulin Resistance in Skeletal Muscle and Derails Glycemic Control in Obesity. *Immunity* **49**, 164–177.e6 (2018).
116. Kypson, J. & Hait, G. Effects of uridine and inosine on glucose metabolism in skeletal muscle and activated lipolysis in adipose tissue. *J. Pharmacol. Exp. Ther.* **199**, 565–574 (1976).
117. Apostol, A. T. & Tayek, J. A. A decrease in glucose production is associated with an increase in plasma citrulline response to oral arginine in normal volunteers. *Metabolism* **52**, 1512–1516 (2003).
118. Jin, X. *et al.* Epidemiological, clinical and virological characteristics of 74 cases of coronavirus-infected disease 2019 (COVID-19) with gastrointestinal symptoms. *Gut* **69**, 1002–1009 (2020).
119. Mo, P. *et al.* Clinical characteristics of refractory COVID-19 pneumonia in Wuhan, China. *Clin. Infect. Dis.* (2020) doi:10.1093/cid/ciaa270.
120. Wang, G. *et al.* C-Reactive Protein Level May Predict the Risk of COVID-19 Aggravation. *Open Forum Infect Dis* **7**, ofaa153 (2020).
121. Shang, W. *et al.* The value of clinical parameters in predicting the severity of COVID-19. *J. Med. Virol.* **92**, 2188–2192 (2020).
122. Du Clos, T. W. & Mold, C. C-reactive protein: an activator of innate immunity and a modulator of adaptive immunity. *Immunol. Res.* **30**, 261–277 (2004).
123. Young, B., Gleeson, M. & Cripps, A. W. C-reactive protein: a critical review. *Pathology* **23**, 118–124 (1991).
124. Sproston, N. R. & Ashworth, J. J. Role of C-Reactive Protein at Sites of Inflammation and

- Infection. *Front. Immunol.* **9**, 754 (2018).
125. Cervenka, I., Agudelo, L. Z. & Ruas, J. L. Kynurenines: Tryptophan's metabolites in exercise, inflammation, and mental health. *Science* **357**, (2017).
126. Wang, Q., Liu, D., Song, P. & Zou, M.-H. Tryptophan-kynurenine pathway is dysregulated in inflammation, and immune activation. *Front. Biosci.* **20**, 1116–1143 (2015).
127. Thomas, T. *et al.* COVID-19 infection alters kynurenine and fatty acid metabolism, correlating with IL-6 levels and renal status. *JCI Insight* **5**, (2020).
128. Xiao, N. *et al.* Integrated cytokine and metabolite analysis reveals immunometabolic reprogramming in COVID-19 patients with therapeutic implications. *Nat. Commun.* **12**, 1618 (2021).
129. Collier, M. E., Zhang, S., Scrutton, N. S. & Giorgini, F. Inflammation control and improvement of cognitive function in COVID-19 infections: is there a role for kynurenine 3-monooxygenase inhibition? *Drug Discov. Today* (2021) doi:10.1016/j.drudis.2021.02.009.
130. Myc, A. *et al.* The level of lipopolysaccharide-binding protein is significantly increased in plasma in patients with the systemic inflammatory response syndrome. *Clin. Diagn. Lab. Immunol.* **4**, 113–116 (1997).
131. Lim, P. S., Chang, Y.-K. & Wu, T.-K. Serum Lipopolysaccharide-Binding Protein is Associated with Chronic Inflammation and Metabolic Syndrome in Hemodialysis Patients. *Blood Purif.* **47**, 28–36 (2019).
132. Martin, T. R. *et al.* Relationship between soluble CD14, lipopolysaccharide binding protein, and the alveolar inflammatory response in patients with acute respiratory distress syndrome. *Am. J. Respir. Crit. Care Med.* **155**, 937–944 (1997).
133. Pastor Rojo, O. *et al.* Serum lipopolysaccharide-binding protein in endotoxemic patients with inflammatory bowel disease. *Inflamm. Bowel Dis.* **13**, 269–277 (2007).
134. Ghahramani, S. *et al.* Laboratory features of severe vs. non-severe COVID-19 patients in Asian populations: a systematic review and meta-analysis. *Eur. J. Med. Res.* **25**, 30 (2020).

135. Bao, W. *et al.* Sodium salicylate modulates inflammatory responses through AMP-activated protein kinase activation in LPS-stimulated THP-1 cells. *J. Cell. Biochem.* **119**, 850–860 (2018).
136. Weber, C., Erl, W., Pietsch, A. & Weber, P. C. Aspirin inhibits nuclear factor-kappa B mobilization and monocyte adhesion in stimulated human endothelial cells. *Circulation* **91**, 1914–1917 (1995).
137. Housby, J. N. *et al.* Non-steroidal anti-inflammatory drugs inhibit the expression of cytokines and induce HSP70 in human monocytes. *Cytokine* **11**, 347–358 (1999).
138. Chow, J. H. *et al.* Aspirin Use Is Associated With Decreased Mechanical Ventilation, Intensive Care Unit Admission, and In-Hospital Mortality in Hospitalized Patients With Coronavirus Disease 2019. *Anesthesia & Analgesia* **132**, 930 (2021).
139. Yuan, S. *et al.* Mortality and pre-hospitalization use of low-dose aspirin in COVID-19 patients with coronary artery disease. *J. Cell. Mol. Med.* **25**, 1263–1273 (2021).
140. Merzon, E. *et al.* The use of aspirin for primary prevention of cardiovascular disease is associated with a lower likelihood of COVID-19 infection. *The FEBS Journal* (2021) doi:10.1111/febs.15784.
141. Baumruker, T. & Prieschl, E. E. Sphingolipids and the regulation of the immune response. *Semin. Immunol.* **14**, 57–63 (2002).
142. Al-Rashed, F. *et al.* Neutral sphingomyelinase 2 regulates inflammatory responses in monocytes/macrophages induced by TNF- α . *Sci. Rep.* **10**, 16802 (2020).
143. Del Valle, D. M. *et al.* An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat. Med.* **26**, 1636–1643 (2020).
144. Gogate, N. *et al.* COVID-19 biomarkers and their overlap with comorbidities in a disease biomarker data model. *Brief. Bioinform.* (2021) doi:10.1093/bib/bbab191.
145. Ermert, D. & Blom, A. M. C4b-binding protein: The good, the bad and the deadly. Novel functions of an old friend. *Immunol. Lett.* **169**, 82–92 (2016).

146. Suankratay, C., Mold, C., Zhang, Y., Lint, T. F. & Gewurz, H. Mechanism of complement-dependent haemolysis via the lectin pathway: role of the complement regulatory proteins. *Clin. Exp. Immunol.* **117**, 442–448 (1999).
147. Dahlbäck, B. Vitamin K-Dependent Protein S: Beyond the Protein C Pathway. *Semin. Thromb. Hemost.* **44**, 176–184 (2018).
148. Zeisel, S. H. & da Costa, K.-A. Choline: an essential nutrient for public health. *Nutr. Rev.* **67**, 615–623 (2009).
149. Zhu, S., Ma, X. & Tang, J.-L. What is the optimal body mass index for Chinese people? *CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne* vol. 183 645–646 (2011).
150. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
151. Rahnavard, A. *deephath: Generic Omics Pathway Enrichment Analysis.* (2020).
152. Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **46**, D649–D655 (2018).
153. Gade, K., Geyik, S. C., Kenthapadi, K., Mithal, V. & Taly, A. Explainable AI in Industry. in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* 3203–3204 (Association for Computing Machinery, 2019).
154. Carrington, A., Fieguth, P. & Chen, H. Measures of Model Interpretability for Model Selection. in *Machine Learning and Knowledge Extraction* 329–349 (Springer International Publishing, 2018).
155. Casalicchio, G., Molnar, C. & Bischl, B. Visualizing the Feature Importance for Black Box Models. in *Machine Learning and Knowledge Discovery in Databases* 655–670 (Springer International Publishing, 2019).
156. Du, M., Liu, N. & Hu, X. Techniques for interpretable machine learning. *Commun. ACM* **63**, 68–77 (2019).

157. Ahmad, M. A., Eckert, C. & Teredesai, A. Interpretable Machine Learning in Healthcare. in *Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* 559–560 (Association for Computing Machinery, 2018).
158. Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R. & Yu, B. Definitions, methods, and applications in interpretable machine learning. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22071–22080 (2019).
159. Fisher, A., Rudin, C. & Dominici, F. All Models are Wrong, but Many are Useful: Learning a Variable's Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.* **20**, 1–81 (2019).
160. Hu, L.-Y., Huang, M.-W., Ke, S.-W. & Tsai, C.-F. The distance function effect on k-nearest neighbor classification for medical datasets. *Springerplus* **5**, 1304 (2016).
161. Westreich, D., Lessler, J. & Funk, M. J. Propensity score estimation: neural networks, support vector machines, decision trees (CART), and meta-classifiers as alternatives to logistic regression. *J. Clin. Epidemiol.* **63**, 826–833 (2010).
162. Awoyemi, J. O., Adetunmbi, A. O. & Oluwadare, S. A. Credit card fraud detection using machine learning techniques: A comparative analysis. in *2017 International Conference on Computing Networking and Informatics (ICCN)* 1–9 (2017).
163. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review* vol. 65 386–408 (1958).
164. Shenfield, A., Day, D. & Ayesh, A. Intelligent intrusion detection systems using artificial neural networks. *ICT Express* **4**, 95–99 (2018).
165. Tu, J. V. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J. Clin. Epidemiol.* **49**, 1225–1231 (1996).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryCOVIDOmicManuscript.pdf](#)
- [SupplementaryTable2.xlsx](#)
- [SupplementaryTable3.xlsx](#)