Molecular profiling of COVID-19 sera

# **1** Proteomic and Metabolomic Characterization of

# 2 COVID-19 Patient Sera

- 3
- 4 Bo Shen<sup>1#</sup>, Xiao Yi<sup>2,3#</sup>, Yaoting Sun<sup>2,3#</sup>, Xiaojie Bi<sup>1#</sup>, Juping Du<sup>1#</sup>, Chao Zhang<sup>4#</sup>,
- 5 Sheng Quan<sup>4#</sup>, Fangfei Zhang<sup>2,3</sup>, Rui Sun<sup>2,3</sup>, Liujia Qian<sup>2,3</sup>, Weigang Ge<sup>2,3</sup>, Wei Liu<sup>2,3</sup>,
- 6 Shuang Liang<sup>2,3</sup>, Hao Chen<sup>2,3</sup>, Ying Zhang<sup>1</sup>, Jun Li<sup>1</sup>, Jiaqin Xu<sup>1</sup>, Zebao He<sup>1</sup>, Baofu
- 7 Chen<sup>1</sup>, Jing Wang<sup>1</sup>, Haixi Yan<sup>1</sup>, Yufen Zheng<sup>1</sup>, Donglian Wang<sup>1</sup>, Jiansheng Zhu<sup>1</sup>,
- 8 Ziqing Kong<sup>4</sup>, Zhouyang Kang<sup>4</sup>, Xiao Liang<sup>2,3</sup>, Xuan Ding<sup>2,3</sup>, Guan Ruan<sup>2,3</sup>, Nan
- 9 Xiang<sup>2,3</sup>, Xue Cai<sup>2,3</sup>, Huanhuan Gao<sup>2,3</sup>, Lu Li<sup>2,3</sup>, Sainan Li<sup>2,3</sup>, Qi Xiao<sup>2,3</sup>, Tian Lu<sup>2,3</sup>, Yi
- 10 Zhu<sup>2,3\*</sup>, Huafen Liu<sup>4\*</sup>, Haixiao Chen<sup>1\*</sup>, Tiannan Guo<sup>2,3,5\*</sup>
- <sup>1</sup>Taizhou Hospital, Wenzhou Medical University, 150 Ximen Street, Linhai 317000, Zhejiang
- 12 Province, China;
- 13 <sup>2</sup>Key Laboratory of Structural Biology of Zhejiang Province, School of Life Sciences, Westlake
- 14 University, 18 Shilongshan Road, Hangzhou 310024, Zhejiang Province, China;
- <sup>15</sup> <sup>3</sup>Institute of Basic Medical Sciences, Westlake Institute for Advanced Study, 18 Shilongshan
- 16 Road, Hangzhou 310024, Zhejiang Province, China;
- <sup>4</sup>Calibra Lab at DIAN Diagnostics, 329 Jinpeng Street, Hangzhou 310030, Zhejiang Province,
   China;
- 19 <sup>5</sup>Lead Contact
- 20 # co-first;
- 21 \* Correspondence: <u>zhuyi@westlake.edu.cn</u> (Y.Z.); <u>liuhf1@dazd.cn</u> (H.L.); <u>chenhx@enzemed.com</u>
- 22 (H.C.); guotiannan@westlake.edu.cn (T.G.)
- 23

## 24 SUMMARY

Severe COVID-19 patients account for most of the mortality of this disease. 25 Early detection and effective treatment of severe patients remain major 26 challenges. Here, we performed proteomic and metabolomic profiling of sera 27 28 from 46 COVID-19 and 53 control individuals. We then trained a machine learning model using proteomic and metabolomic measurements from a 29 training cohort of 18 non-severe and 13 severe patients. The model correctly 30 classified severe patients with an accuracy of 93.5%, and was further 31 validated using ten independent patients, seven of which were correctly 32 classified. We identified molecular changes in the sera of COVID-19 patients 33 34 implicating dysregulation of macrophage, platelet degranulation and complement system pathways, and massive metabolic suppression. This 35 study shows that it is possible to predict progression to severe COVID-19 36 disease using serum protein and metabolite biomarkers. Our data also 37 uncovered molecular pathophysiology of COVID-19 with potential for 38 developing anti-viral therapies. 39

40

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice. SHEN, ET AL

Molecular profiling of COVID-19 sera

## 41 INTRODUCTION

Coronavirus disease 2019 (COVID-19) is an unprecedented global threat
caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It
is currently spreading around the world rapidly. The sudden outbreak and
accelerated spreading of SARS-CoV-2 infection have caused substantial
public concerns. Within about two months, close to one million individuals
worldwide have been infected, leading to about 50,000 deaths. At the time of
writing this manuscript, about 100,000 new infections are reported daily.
Most COVID-19 studies have focused on its epidemiological and clinical

49 characteristics(Ghinai et al., 2020; Guan et al., 2020; Li et al., 2020b; Team, 50 2020). About 80% of patients infected with SARS-CoV-2 displayed mild 51 symptoms with good prognosis(Team, 2020). They usually recover with, or 52 even without, conventional medical treatment, and therefore are classified as 53 mild or moderate COVID-19 (Medicine, 2020; Thevarajan et al., 2020). 54 55 However, about 20% of patients suffer from respiratory distress and require immediate oxygen therapy or other inpatient interventions, including 56 mechanical ventilation (Medicine, 2020; Murthy et al., 2020; Wu and 57 McGoogan, 2020). These patients, classified as clinically severe or critical life-58 threatening infections, are mainly diagnosed empirically based on a set of 59 clinical characteristics, such as respiratory rate ( $\geq$  30 times/min), mean 60 oxygen saturation ( $\leq 93\%$  in the resting state) or arterial blood oxygen partial 61 62 pressure/oxygen concentration (≤ 300 mmHg) (Medicine, 2020). However, patients exhibiting these clinical manifestations have already progressed to a 63

clinically severe phase and require immediate access to specialized intensive
care; otherwise, they may die rapidly. Therefore, it is critical to develop new
approaches to predict early which cases will likely become clinically severe. In
addition, effective therapy for severe patients remains speculative, largely due
to limited understanding of SARS-CoV-2 pathogenesis.

In this study, we hypothesized that SARS-CoV-2 induces characteristic 69 molecular changes that can be detected in the sera of severe patients. These 70 molecular changes may shed light on therapy development for COVID-19 71 patients. To test this hypothesis, we applied cutting-edge proteomic 72 (Aebersold and Mann, 2016; Li et al., 2020a) and metabolomic (Hou et al., 73 2020; Lee et al., 2019) technologies to analyze the proteome and 74 metabolome of sera from severe COVID-19 patients and several control 75 groups. 76

77

## 78 **RESULTS**

### 79 **Proteomic and metabolomic profiling of COVID-19 sera**

80 We procured a cohort of patients (Zheng et al., 2020) containing 21 81 severe COVID-19 patients, in which 11 sera from severe patients were

82 collected one to six days before the patients were clinically assessed as

### Molecular profiling of COVID-19 sera

severe. Sera from the remaining eight patients were collected within about 3 83 days after diagnosed as severe cases (Figure 1A, Table 1, Table S1). Controls 84 with matched epidemiological features were included to identify severity-85 related molecular alterations. These controls were 28 healthy subjects, 25 86 non-COVID-19 patients (negative for the SARS-CoV-2 nucleic acid test) with 87 similar clinical characteristics as COVID-19 patients, and 25 non-severe 88 COVID-19 patients. A serum sample was obtained from each patient within a 89 few days after hospital admission, with a few exceptions when samplings 90 were performed at later disease stages. 91

We used stable isotope labeled proteomics strategy TMTpro (16plex) (Li 92 et al., 2020a) and UPLC-MS/MS untargeted metabolomics approach to 93 analyze 92 undepleted sera from 86 individuals. Altogether 894 proteins were 94 quantified and 941 (including 36 drugs and their metabolites) metabolites 95 were identified with authentic compound library searching. For metabolomic 96 analysis, both hydrophilic and hydrophobic molecules were analyzed using 97 98 both positive and negative ionization to cover various endogenous 99 biochemical classes. Our data were acquired with high degree of consistency and reproducibility. In quality control analysis, the median coefficient of 100 variance (CV) values for proteomic and metabolomic data were 10% and 5%, 101 respectively (Figure S1A). Remarkably, sera from SARS-CoV-2 infected 102 patients were well resolved from healthy individuals, and the sera from severe 103 patients displayed distinct proteomic and metabolomic patterns compared to 104 105 those from other groups (Figures S1B-C).

- 106
- 107

### Identification of severe patients using machine learning

We next investigated the possibility of classify the severe COVID-19 108 patients based on the molecular signatures of proteins and metabolites (Table 109 S2). We built a random forest machine learning model based on proteomic 110 and metabolomic data from 18 non-severe and 13 severe patients (Figure 111 1B), leading to prioritization of 29 important variables including 22 proteins 112 and 7 metabolites (Figure 2A-B). This model reached an AUC of 0.957 in the 113 training set (Figure 2C). One non-severe patient, XG3, was incorrectly 114 classified as severe (Figure 2D), possibly because this 70-year-old male 115 patient was the oldest individual in this cohort (Figure 1A). For patient XG40, 116 the reason of incorrectly classified is unclear. 117

We then tested the model on an independent cohort of ten patients 118 (Figure 2E, Table S3). All severe patients were correctly identified, except one 119 patient, XG45, a 62-year-old male who had received traditional Chinese 120 medicines during more than 20 days before admission. This individual was 121 the patient with the longest pre-admission treatment in this cohort and the 122 administration of traditional Chinese medicines might have also accounted for 123 124 the incorrect prediction. Another incorrectly classified non-severe patient was XG22, a 43 year-old male with chronic HBV infection and diabetes who had 125 the longest hospitalization (>50 days) among all non-severe patients. Indeed, 126

#### Molecular profiling of COVID-19 sera

our molecular-based prediction flagged him as an outlier, suggesting that his
 prolonged treatment history might have interfered with the expression of the
 panel of the 29 variables. For reasons yet unclear, XG25, a 43-year-old male
 non-severe case, was incorrectly classified as severe.

131

132

## Proteomic and metabolomic changes in severe COVID-19 sera

We found that 105 proteins were differentially expressed in COVID-19 133 patients, but not in the non-COVID-19 patients (Figure S2, S4). After 134 correlating their expression with clinical disease severity (Figure S5), 93 135 proteins showed specific modulation in severe patients. Pathway analysis and 136 network enrichment analyses of the 93 differentially expressed proteins 137 revealed three pathways (Figure S6), namely activation of the complement 138 system, macrophage function and platelet degranulation, involving 50 proteins 139 (Figure 3A) in the sera of severe patients. Similarity, 373 and 204 metabolites 140 were found with differential abundance in COVID-19 patients and continually 141 142 changed correlated with disease severity in our metabolomics data, respectively (Figure S3, S4). Correspondingly, 82 metabolites were involved 143 in the three biological processes. We summarize the key dysregulated 144 molecules in Figure 5 and discussed in the following sections. 145

146

### 147 **Dysregulated macrophage and lipid metabolism**

Our data uncovered dysregulation of multiple apolipoproteins including 148 APOA1, APOA2, APOH, APOL1, APOD and APOM (Figure 3A). Most of them 149 are associated with macrophage function and were down-regulated. Decrease 150 of APOA1 in serum has been reported during the transition of COVID-19 151 patients from mild to severe illness (Nie et al., 2020). The APOM in serum of 152 severe patients was downregulated compared with healthy and non-severe 153 COVID-19 patients. Regulation of serum APOM has also been observed in 154 hepatitis B virus (HBV) patients (Gu et al., 2011). 155

Consistent with these proteomic findings, we also detected dysregulated 156 metabolites involved in lipid metabolism. Accumulation of 16 steroid hormones 157 in COVID-19 patients may contribute to macrophage modulation. Steroid 158 hormones, including progesterone, androgens, estrogens and bile acids can 159 promote the activity of macrophages (Vernon-Roberts, 1969). Specifically, 160 glucocorticoids were recently reported to be clinically efficacious (Wang et al., 161 2020b). We detected increased 21-hydroxypregnenolone, the essential 162 intermediate for synthesizing corticosterone, suggesting that corticosterone 163 biosynthesis may be a protective mechanism against SARS-CoV-2 infection. 164

We also found evidence of significant activation of the kynurenine pathway. Metabolites of kynurenate, kynurenine, 8-methoxykynurenate were enriched in COVID-19 patients. Nicotinamide adenine dinucleotide (NAD+), the cofactor in many cellular redox reactions, can be synthesized from tryptophan by the kynurenine pathway and operates as a switch for macrophage effector responses (Minhas et al., 2019).

### Molecular profiling of COVID-19 sera

The macrophage process is closely related to lipid metabolism. Over 100 171 lipids were down-regualted in severe patients. Our data showed decreased 172 sphingolipids in both non-severe and severe COVID-19 patients (Figure 4A). 173 Sphingolipids and glycerophospholipids are important components of 174 biomembranes, which mediate signal transduction and immune activation 175 processes. Sphingolipids, such as sphingosine-1-phosphate, can induce 176 macrophage activation, inhibit macrophage apoptosis, and promote migration 177 of macrophages to inflammatory sites (Weigert et al., 2009). Phagocytosis 178 and platelet degranulation are coupled with changes in biomembrane lipid 179 composition and fluidity, and modulate the production of glycerophospholipids 180 (Rouzer et al., 2007). In this study, we found continuous decrease of 181 glycerophospholipids after SARS-CoV-2 infection. Glycerophospholipids and 182 fatty acids such as arachidonic acid have been found significantly elevated in 183 HCoV-229E-infected cells, and exogenous supplement of arachidonic acid 184 significantly suppressed HCoV-229E and MERS-CoV replication (Yan et al., 185 186 2019). Our data suggest severe COVID-19 patients might benefit from this therapeutical strategy as well. 187

188 Choline and its derivatives were down-regulated in COVID-19 patients, 189 particularly in severe cases; phosphocholine, the intermediate product for 190 producing phosphatidylcholine (PC) was up-regulated (Figure 4B). This was 191 probably due to activated macrophage-mediated immunity (Sanchez-Lopez et 192 al., 2019). Polarization of macrophages in response to pathogens requires 193 increased absorption of choline for PC formation, thereby promoting cytokine 194 secretion (Sanchez-Lopez et al., 2019).

195 196

Activated acute phase proteins (APPs) and the complement system

We detected 10 APPs among 20 proteins that are differentially expressed 197 between non-severe and severe groups (Figure 2A). They are involved at the 198 199 early stages of immune responses to virus infection. Among the most significantly upregulated in severe sera were APPs, including serum amyloid 200 A-1 (SAA1), serum amyloid A-2 (SAA2), serum amyloid A-4 (SAA4), C-201 reactive protein (CRP), alpha-1-antichymotrypsin (SERPINA3) and serum 202 amyloid P-component (SAP/APCS) (Figure 3B). Some of them are known to 203 be biomarkers for viral infections, including SAA1, SAA2 and CRP. While 204 CRP has been associated with COVID-19, the other proteins have not 205 previously been reported in COVID-19 (Liang et al., 2020). SAA1 was 206 reported to be elevated in severe SARS patients, but was not specific to 207 SARC-CoV (Pang et al., 2006). As a major contributor to acute phase 208 response, complement system plays a crucial role in eliminating invading 209 pathogens in the early stage of infection. Among those APPs, two proteins 210 belong to the complement membrane attack complex, including complement 6 211 212 (C6) and complement factor B (CFB). Two other proteins, Properdin (CFP) and Carboxypeptidase N catalytic chain (CPN1), are regulators of 213

complement system (Figure 3).

SHEN, ET AL

#### Molecular profiling of COVID-19 sera

We also observed an accumulation of mannose and its derivatives in severe patients. In the complement system, binding of mannose to lectin leads to cleaveage of C2 and C4, which then form a C3 convertase to promote complement activation (Ricklin et al., 2010).

219

220

### Suppressed platelet degranulation in severe sera

Fifteen of 17 proteins involved in platelet degranulation were down-221 regulated in SARS-CoV-2 infected patients, which may be associated with 222 observed thrombocytopenia in this patient cohort (Zheng et al., 2020). Low 223 platelet count is also reported to be associated with severe COVID-19 and 224 mortality (Lippi et al., 2020). Two of the most intriguing proteins down-225 regulated in severe patients are platlet expressing chemokines pro-platelet 226 basic protein (PPBP; also called macrophage-derived growth factor) and 227 platelet factor 4 (PF4). PF-4 was identified as a broad-spectrum HIV-1 228 inhibitor at the level of virus attachment and entry via interaction with the 229 230 major viral envelope glycoprotein gp120 (Auerbach et al., 2012). In another sera proteomic investigation of SARS, they found decreasing PF4 was 231 associated with poor prognosis (Poon et al., 2012), in consistent with our 232 findings regarding COVID-19. 233

Most enterochromaffin cell-derived serotonins (5-hydroxytryptamine [5-HT]) are transported to platelets for storage and release (Baganz and Blakely, 2013). Our previous data showed that when platelet counts in COVID-19 patients decrease as the severity of the disease increase (Zheng et al., 2020), serotonin (Figure 4B) level decreases accordingly. Compared with the healthy group, serotonin in non-severe and severe COVID-19 patients decreased by 2.07-fold (p = 1.86e-04) and 3.31-fold (p = 9.07e-07), respectively.

We also detected low levels of fatty acids such as arachidonate and docosapentaenoate in COVID-19 patients (Figure 4), which may be related to their decreased platelet counts. Fatty acids including arachidonate (20:4n6) are active factors of platelet degranulation. A study on H7N9 reported that H7N9 infection led to suppression of various fatty acids including palmitic acid (Sun et al., 2018). We also found palmitic acid decreased in severe COVID-19 patients.

- 248
- 249

## Massive suppression of amino acids in the sera of COVID-19 patients

More than 100 decreased metabolites in the sera of COVID-19 patients 250 are amino acids and their derivatives were found significantly decreased in the 251 sera of COVID-19 patients compared to their levels in the healthy controls, 252 while their levels were either unchanged or even increased in the sera of non-253 COVID-19 patients. Enriched in these metabolites are ten metabolites 254 involved in arginine metabolism including glutamate, arginine, N-(I-255 arginino)succinate, citrulline, ornithine, glutamine, 2-oxoglutarate, N-acetyl-L-256 glutamate, urea, and fumarate. In addition, some arginine derivatives such as 257 argininate, asymmetric dimethylarginine, symmetric dimethylarginine, 258

### Molecular profiling of COVID-19 sera

homoarginine and N-acetylarginine, were also significantly decreased in the
sera of non-severe COVID-19 patients. It has been reported that arginine
metabolism is suppressed in severe fever with thrombocytopenia syndrome
caused by a SFTS bunyavirus (SFTSV)(Li et al., 2018b). Decreased arginine
levels in SFTSV patients was associated with impaired anti-SFTSV functions
of T cells.

265

## 266 **DISCUSSION**

## 267 Prediction of severe COVID-19 patients

Although COVID-19 can be diagnosed effectively by nucleic acid-based 268 methods at an early stage, it is equally critical to identify severe COVID-19 269 patients before their manifestion of severe symptoms to minimize mortality. In 270 this study we show that severe cases can be predicted by molecular 271 signatures of metabolites and proteins using a machine learning model based 272 273 on the expression levels of 22 serum proteins and seven metabolites (Figure 2A-B). We achieved an overall accuracy of 93.5% in the training set. 274 Prediction of two patients did not match clinical diagnosis. One of them is non-275 severe individual who was the oldest patient in the training cohort. 276 Remarkably, nine severe patients were correctly identified retrospectively 277 based on the analysis of their sera collected one to six days before they were 278 clinically diagnosed as having deteriorated to a severe state (Figure 1, Figure 279 280 2D), suggesting that their sera protein and metabolite signatures at the sampling time were already pointing to further deterioration into severe state 281 even when severe clinical symptoms have not started to appear yet. 282

The proteins and metabolites used in our classifier (Figure 2A) contain 283 several known biomarkers for viral infections, such as SAA2, SSA1 and CRP, 284 which have already been used empirically to monitor the severity of COVID-285 19. Our study suggests that more characteristic molecular changes at protein 286 and metabolite levels can be used to build a predictive model for the 287 prospective identification of severe cases. The classifier also included 288 exceptionally high levels of other acute phase proteins, including SERPINA3, 289 among others (Figure 2A-B). Our data suggest potential benefits of broader 290 testing of these proteins in newly diagnosed cases to identify which COVID-19 291 patients are likely to progress to severe disease. The model contains 292 molecules involved in hepatic damage. The elevation of glucose, glucuronate, 293 bilirubin degradation product and four bile acid derivatives, suggests 294 suppressed hepatic detoxification (Rowland et al., 2013). Vascular cell 295 adhesion protein 1 (VCAM-1) which helps to regulate transendothelial 296 migration of leukocytes by stimulating production of reactive oxygen species 297 (ROS), was upregulated in our data. As apotent antioxidant and inhibitor of 298 299 VCAM-1 dependent cellular events (Keshavan et al., 2005), bilirubin was found to be down-regulated in our metabolomic data. 300

#### Molecular profiling of COVID-19 sera

Seven patients were correctly classified in the independent validation
 cohort containing ten patients. Two of them could be explained by the
 patients' complex comorbidity and medication history. The relatively small
 sample size necessitates future validation studies in independent cohorts.

- 305
- 306

## Molecular insights for pathogenesis of SARS-CoV-2 infection

Our data shed light on the molecular changes reflected in COVID-19 sera 307 which could potentially yield critical diagnostic markers or therapeutic targets 308 for managing severe COVID-19 patients (Figure 5). These molecular 309 derangements may originate from binding of SARS-CoV-2 to alveolar 310 macrophages via the ACE2 receptor (Hoffmann et al., 2020), resulting in 311 release of IL-6 and TNF- $\alpha$  (Mehta et al., 2020) by macrophages (Gabay and 312 Kushner, 1999). In response to elevated cytokines, especially IL-6 which 313 triggers fever, the liver releases various APPs. Activation of APPs is 314 accompanied with the immunogenesis or organic damage (Gabay and 315 316 Kushner, 1999). Our metabolomics data also provide plausible evidence for hepatic injury. In physiological condition, hormone or bilirubin binds to 317 glucuronate, a derivative of glucose, for liver detoxification (Rowland et al., 318 2013). The elevation of glucose, glucuronate, bilirubin degradation product 319 and four bile acid derivatives in severe patients, indicating the potential 320 declined detoxification function (Figures 4-5). Indeed, our data showed an 321 upregulation of multiple APPs, including CRP and major attack complexes 322 323 (MACs) in the severe sera. CRP can activate the complement system (Chirco and Potempa, 2018). This on the one hand leads to enhanced cytokine and 324 chemokine production, potentially contributing to "cytokine storm"; and on the 325 other overly recruits macrophages from the peripheral blood, which could 326 result in acute lung injury(Chirco and Potempa, 2018; Narasaraju et al., 2011). 327 Because about 50% of platelets are produced in the lung (Lefrancais et al., 328 329 2017), platelet may in turn respond to lung injury and activate macrophases by degranulation (Mantovani and Garlanda, 2013), which may further add to 330 "cytokine storm". A recent necropsy report revealed alveolar macrophage 331 infiltration and activation in severe COVID-19 patients (Liao M., 2020), 332 supporting our findings. 333

334 335

### Insights for COVID-19 therapeutics

To date, few other therapies are proven effective for severe COVID-19 336 patients. Most of patients receive standard supportive care and antiviral 337 therapy (Wang et al., 2020a). Corticosteroid treatment which are effectively in 338 suppressing MERS-CoV and SARS-CoV (Arabi et al., 2018), but showed 339 negligible effect on COVID-19 patients and may even have induced lung 340 injury (Russell et al., 2020). The molecular changes revealed in this study in 341 342 the COVID-19 sera have allowed us to propose potential new therapeutic strategies for the severe patients. 343

### Molecular profiling of COVID-19 sera

Our proteomic data showed that proteins related to platelet degranulation 344 were substantially down-regulated in severe patients, a finding which was 345 confirmed by low platelet counts (Zheng et al., 2020). The association 346 between thrombocytopenia and viral infection has been observed in SARS-347 CoV (Zou et al., 2004), hepatitis C virus (HCV) (Assinger, 2014), and Dengue 348 virus (Wilder-Smith et al., 2004). We thus recommend watching closely 349 monitoring changes in platelets and making interventions, such as infusions of 350 thrombopoietin (TPO) as necessary, particularly when liver or kidney injuries 351 occur. 352

Complement activation suppresses virus invasion, and may lead to 353 inflammatory syndromes (Barnum, 2016). Our data showed a general up-354 regulation of complement system proteins, including MAC proteins such as 355 C5, C6 and C8. Suppression of complement system has been reported as an 356 effective immunotherapeutics in SARS-infected mouse model (Gralinski et al., 357 2018). C5a has been reported as highly expressed in severe SARS and 358 MERS patients as well (Wang et al., 2015). Inhibition of C5a has been 359 reported to alleviate viral infection-induced acute lung injury (Garcia et al., 360 2013; Jiang et al., 2018; Sun et al., 2014). Our data suggest that severe 361 COVID-19 patients might benefit from suppression of complement system. 362

The coronavirus are enveloped, positive strand RNA viruses, the 363 replication and assembly of which consume large amount of lipids. Our 364 metabolomics results showed that more than 100 lipids including 365 glycerophospholipid, sphingolipids and fatty acids were down-regulated in 366 COVID-19 patients sera, probably due to rapid replication of the virus. 367 Glycerophospholipid, sphingolipids (one of the components of lipid rafts) and 368 fatty acids have been reported to play an important role in the early 369 development of enveloped viruses (Schoggins and Randall, 2013). 370 Suppression of cholesterol synthesis by MBCD has been reported inhibiting 371 372 the production of SARS-CoV particles released from infected Vero E6 cells (Li et al., 2007). Drugs inhibiting lipid synthesis such as statin have been 373 proposed to treat HCV (Heaton and Randall, 2011) and COVID-19 (Fedson et 374 al., 2020). Our data also supports this potential therapeutics for severe 375 COVID-19 patients. 376

SARS-CoV-2 is highly infectious, applying huge pressure to the medical
 system worldwide. Upon COVID-19 outbreak, limited information of this
 pathogen was available, which restricted the collection of clinical specimens
 for this study. The reason for the small sample size of this study were
 biosafety constraints.

In conclusion, this study presents a systematic proteomic and metabolimic investigation of sera from multiple COVID-19 patient groups and control groups. We show that feasibility of predicting severe COVID-19 patients based on a panel of serum proteins and metabolites. Our data offer a landscape of blood molecular changes induced by SARS-CoV-2 infection, which may provide useful diagnositic markers and therapeutic targets.

Molecular profiling of COVID-19 sera

388
-----

389

## 390 ACKNOWLEDGEMENTS

This work is supported by grants from Westlake Special Program for COVID-391 19 (2020), and Tencent foundation (2020), National Natural Science 392 Foundation of China (81972492, 21904107, 81672086), Zheijang Provincial 393 Natural Science Foundation for Distinguished Young Scholars 394 (LR19C050001), Hangzhou Agriculture and Society Advancement Program 395 (20190101A04). We thank Drs. R. Aeberold, O.L. Kon, H. Yu, D. Li, and the 396 Guomics team for invaluable comments to this study. We thank Westlake 397 University Supercomputer Center for assistance in data storage and 398 computation. 399

## **400 AUTHOR CONTRIBUTIONS**

T.G., H.C., H.L., B.S. and Y.Z. designed and supervised the project. B.S.,
X.B., J.D., Y.Z., J.L., J.X., Z.H., B.C., J.W., H.Y., Y.Z., D.W. and J.Z. collected
the samples and clinical data. X.Y., Y.S., F.Z., R.S., L.Q., W.G., W.L., S.L.,
H.C., X.L., X.D., G.R., N.X., X.C., H.G., L.L., S.L., Q.X. and T.L. conducted
proteomic analysis. Data were interpreted and presented by all co-authors.
C.Z., S.Q., Z.K. and Z.K. conducted metabolomic analysis. T.G. wrote the
manuscript with inputs from co-authors.

## 408 **DECLARATION OF INTERESTS**

The research group of T.G. is partly supported by Tencent, Thermo Fisher
Scientific, SCIEX and Pressure Biosciences Inc. C.Z., Z.K., Z.K. and S.Q. are
employees of DIAN Diagnostics.

## 412 **DATA AVAILABILITY**

All data are available in the manuscript or the supplementary material. The

414 proteomics and metabolomics data are deposited in ProteomeXchange

- 415 Consortium (https://www.iprox.org/). Project ID: IPX0002106000. All the data
- 416 will be publicly released upon publication.

## 417 **REFERENCES**

- 418 Aebersold, R., and Mann, M. (2016). Mass-spectrometric exploration of proteome structure and 419 function. Nature *537*, 347-355.
- 420 Arabi, Y.M., Mandourah, Y., Al-Hameed, F., Sindi, A.A., Almekhlafi, G.A., Hussein, M.A., Jose, J.,
- 421 Pinto, R., Al-Omari, A., Kharaba, A., *et al.* (2018). Corticosteroid Therapy for Critically III Patients 422 with Middle East Respiratory Syndrome. Am J Respir Crit Care Med *197*, 757-767.
- Assinger, A. (2014). Platelets and infection an emerging role of platelets in viral infection. Front
  Immunol *5*, 649.
- 425 Auerbach, D.J., Lin, Y., Miao, H., Cimbro, R., Difiore, M.J., Gianolini, M.E., Furci, L., Biswas, P., Fauci,
- 426 A.S., and Lusso, P. (2012). Identification of the platelet-derived chemokine CXCL4/PF-4 as a broad-

- 427 spectrum HIV-1 inhibitor. Proc Natl Acad Sci U S A 109, 9569-9574.
- 428 Baganz, N.L., and Blakely, R.D. (2013). A dialogue between the immune system and brain, spoken
- 429 in the language of serotonin. ACS Chem Neurosci *4*, 48-63.
- Barnum, S. (2016). Complement: A Primer for the Coming Therapeutic Revolution. Pharmacology& Therapeutics *172*.
- 432 Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pages,
- 433 F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally 434 grouped gene ontology and pathway annotation networks. Bioinformatics *25*, 1091-1093.
- 435 Chirco, K.R., and Potempa, L.A. (2018). C-Reactive Protein As a Mediator of Complement Activation
- and Inflammatory Signaling in Age-Related Macular Degeneration. Frontiers in Immunology *9*,539.
- Fedson, D.S., Opal, S.M., and Rordam, O.M. (2020). Hiding in Plain Sight: an Approach to Treating
  Patients with Severe COVID-19 Infection. mBio *11*.
- 440 Gabay, C., and Kushner, I. (1999). Acute-Phase Proteins and Other Systemic Responses to 441 Inflammation. New England Journal of Medicine *340*, 448-454.
- 442 Gao, H., Zhang, F., Liang, S., Zhang, Q., Lyu, M., Qian, L., Liu, W., Ge, W., Chen, C., Yi, X., et al. (2020).
- 443 Accelerated Lysis and Proteolytic Digestion of Biopsy-Level Fresh-Frozen and FFPE Tissue Samples444 Using Pressure Cycling Technology. Journal of Proteome Research.
- 445 Garcia, C.C., Weston-Davies, W., Russo, R.C., Tavares, L.P., Rachid, M.A., Alves-Filho, J.C., Machado,
- 446 A.V., Ryffel, B., Nunn, M.A., and Teixeira, M.M. (2013). Complement C5 activation during influenza
- A infection in mice contributes to neutrophil recruitment and lung injury. PLoS One *8*, e64443-e64443.
- 449 Ghinai, I., McPherson, T.D., Hunter, J.C., Kirking, H.L., Christiansen, D., Joshi, K., Rubin, R., Morales-
- 450 Estrada, S., Black, S.R., Pacilli, M., *et al.* (2020). First known person-to-person transmission of severe 451 acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the USA. The Lancet.
- 452 Gralinski, L.E., Sheahan, T.P., Morrison, T.E., Menachery, V.D., Jensen, K., Leist, S.R., Whitmore, A.,
- Heise, M.T., and Baric, R.S. (2018). Complement Activation Contributes to Severe Acute Respiratory
  Syndrome Coronavirus Pathogenesis. mBio *9*, e01753-01718.
- Gu, J.-G., Zhu, C.-I., Cheng, D.-z., Xie, Y., Liu, F., and Zhou, X. (2011). Enchanced levels of
  apolipoprotein M during HBV infection feedback suppresses HBV replication. Lipids Health Dis *10*,
  154-154.
- 458 Guan, W.-j., Ni, Z.-y., Hu, Y., Liang, W.-h., Ou, C.-q., He, J.-x., Liu, L., Shan, H., Lei, C.-I., Hui, D.S.C.,
- *et al.* (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. New England Journal ofMedicine.
- Heaton, N.S., and Randall, G. (2011). Multifaceted roles for lipids in viral infection. Trends Microbiol *19*, 368-375.
- 463 Hoffmann, M., Kleine-Weber, H., Schroeder, S., Kruger, N., Herrler, T., Erichsen, S., Schiergens, T.S.,
- Herrler, G., Wu, N.H., Nitsche, A., *et al.* (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and
  TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell.
- 466 Hou, Y.C., Yu, H.C., Martin, R., Cirulli, E.T., Schenker-Ahmed, N.M., Hicks, M., Cohen, I.V., Jonsson,
- 467 T.J., Heister, R., Napier, L., et al. (2020). Precision medicine integrating whole-genome sequencing,
- 468 comprehensive metabolomics, and advanced imaging. Proc Natl Acad Sci U S A *117*, 3053-3062.
- 469 Jiang, Y., Zhao, G., Song, N., Li, P., Chen, Y., Guo, Y., Li, J., Du, L., Jiang, S., Guo, R., et al. (2018).
- 470 Blockade of the C5a-C5aR axis alleviates lung damage in hDPP4-transgenic mice infected with

- 471 MERS-CoV. Emerg Microbes Infect 7, 77-77.
- 472 Keshavan, P., Deem, T.L., Schwemberger, S.J., Babcock, G.F., Cook-Mills, J.M., and Zucker, S.D.
- 473 (2005). Unconjugated bilirubin inhibits VCAM-1-mediated transendothelial leukocyte migration. J
  474 Immunol *174*, 3709-3718.
- Kramer, A., Green, J., Pollard, J., Jr., and Tugendreich, S. (2014). Causal analysis approaches in
  Ingenuity Pathway Analysis. Bioinformatics *30*, 523-530.
- 477 Lee, A.H., Shannon, C.P., Amenyogbe, N., Bennike, T.B., Diray-Arce, J., Idoko, O.T., Gill, E.E., Ben-
- 478 Othman, R., Pomat, W.S., van Haren, S.D., et al. (2019). Dynamic molecular changes during the first
- 479 week of human life follow a robust developmental trajectory. Nat Commun *10*, 1092.
- Lefrancais, E., Ortiz-Munoz, G., Caudrillier, A., Mallavia, B., Liu, F., Sayah, D.M., Thornton, E.E.,
  Headley, M.B., David, T., Coughlin, S.R., *et al.* (2017). The lung is a site of platelet biogenesis and a
  reservoir for haematopoietic progenitors. Nature *544*, 105-109.
- Li, G.M., Li, Y.G., Yamate, M., Li, S.M., and Ikuta, K. (2007). Lipid rafts play an important role in the early stage of severe acute respiratory syndrome-coronavirus life cycle. Microbes Infect *9*, 96-102.
- 485 Li, J., Van Vranken, J.G., Pontano Vaites, L., Schweppe, D.K., Huttlin, E.L., Etienne, C., Nandhikonda,
- 486 P., Viner, R., Robitaille, A.M., Thompson, A.H., *et al.* (2020a). TMTpro reagents: a set of isobaric
  487 labeling mass tags enables simultaneous proteome-wide measurements across 16 samples. Nat
  488 Methods.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., *et al.* (2020b). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected
  Pneumonia. N Engl J Med.
- Li, T., Kim, A., Rosenbluh, J., Horn, H., Greenfeld, L., An, D., Zimmer, A., Liberzon, A., Bistline, J.,
  Natoli, T., *et al.* (2018a). GeNets: a unified web platform for network-based genomic analyses. Nat
  Methods *15*, 543-546.
- 495 Li, X.K., Lu, Q.B., Chen, W.W., Xu, W., Liu, R., Zhang, S.F., Du, J., Li, H., Yao, K., Zhai, D., et al. (2018b).
- 496 Arginine deficiency is involved in thrombocytopenia and immunosuppression in severe fever with497 thrombocytopenia syndrome. Sci Transl Med *10*.
- Liang, T., Cai, H., Chen, Y., Chen, Z., Fang, Q., Han, W., Hu, S., Li, J., Li, T., Lu, X., *et al.* (2020).
  Handbook of COVID-19 Prevention and Treatment
- Liao M., e.a. (2020). The landscape of lung bronchoalveolar immune cells in COVID-19 revealed
  by single-cell RNA sequencing. medRxiv, https://doi.org/10.1101/2020.1102.1123.20026690.
- Lippi, G., Plebani, M., and Henry, B.M. (2020). Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis. Clin Chim Acta *506*, 145-148.
- 504 Mantovani, A., and Garlanda, C. (2013). Platelet-macrophage partnership in innate immunity and 505 inflammation. Nat Immunol *14*, 768-770.
- Medicine, N.H.C.o.t.P.s.R.o.C.a.N.A.o.T.C. (2020). Chinese Government Diagnosis and Treatment
   Guideline (Trial 5th version).
- 508 Mehta, P., McAuley, D.F., Brown, M., Sanchez, E., Tattersall, R.S., Manson, J.J., and Hlh Across 509 Speciality Collaboration, U.K. (2020). COVID-19: consider cytokine storm syndromes and 510 immunosuppression. Lancet *395*, 1033-1034.
- 511 Minhas, P.S., Liu, L., Moon, P.K., Joshi, A.U., Dove, C., Mhatre, S., Contrepois, K., Wang, Q., Lee, B.A.,
- 512 Coronado, M., et al. (2019). Macrophage de novo NAD(+) synthesis specifies immune function in
- aging and inflammation. Nat Immunol *20*, 50-63.
- 514 Murthy, S., Gomersall, C.D., and Fowler, R.A. (2020). Care for Critically III Patients With COVID-19.

- 515 JAMA.
- 516 Narasaraju, T., Yang, E., Samy, R.P., Ng, H.H., Poh, W.P., Liew, A.A., Phoon, M.C., van Rooijen, N.,
- and Chow, V.T. (2011). Excessive neutrophils and neutrophil extracellular traps contribute to acute
  lung injury of influenza pneumonitis. Am J Pathol *179*, 199-210.
- 519 Nie, S., Zhao, X., Zhao, K., Zhang, Z., Zhang, Z., and Zhang, Z. (2020). Metabolic disturbances and
- inflammatory dysfunction predict severity of coronavirus disease 2019 (COVID-19): a retrospective
   study. medRxiv, 2020.2003.2024.20042283.
- 522 Pang, R.T., Poon, T.C., Chan, K.C., Lee, N.L., Chiu, R.W., Tong, Y.K., Chim, S.S., Sung, J.J., and Lo, Y.M.
- 523 (2006). Serum amyloid A is not useful in the diagnosis of severe acute respiratory syndrome. Clin
  524 Chem *52*, 1202-1204.
- 525 Poon, T., Pang, R., Chan, K.C.A., Lee, N., Chiu, R., Tong, Y.-K., Chim, S., Ngai, S.M., Sung, J., and Lo,
- D. (2012). Proteomic analysis reveals platelet factor 4 and beta-thromboglobulin as prognostic
   markers in severe acute respiratory syndrome. Electrophoresis *33*, 1894-1900.
- Ricklin, D., Hajishengallis, G., Yang, K., and Lambris, J.D. (2010). Complement: a key system for immune surveillance and homeostasis. Nat Immunol *11*, 785-797.
- Rouzer, C.A., Ivanova, P.T., Byrne, M.O., Brown, H.A., and Marnett, L.J. (2007). Lipid profiling reveals
  glycerophospholipid remodeling in zymosan-stimulated macrophages. Biochemistry *46*, 60266042.
- Rowland, A., Miners, J.O., and Mackenzie, P.I. (2013). The UDP-glucuronosyltransferases: Their role
  in drug metabolism and detoxification. The International Journal of Biochemistry & Cell Biology
  45, 1121-1132.
- Russell, C.D., Millar, J.E., and Baillie, J.K. (2020). Clinical evidence does not support corticosteroid
  treatment for 2019-nCoV lung injury. Lancet *395*, 473-475.
- 538 Sanchez-Lopez, E., Zhong, Z., Stubelius, A., Sweeney, S.R., Booshehri, L.M., Antonucci, L., Liu-Bryan, 539 R., Lodi, A., Terkeltaub, R., Lacal, J.C.*, et al.* (2019). Choline Uptake and Metabolism Modulate
- 540 Macrophage IL-1beta and IL-18 Production. Cell Metab *29*, 1350-1362 e1357.
- 541 Schoggins, J.W., and Randall, G. (2013). Lipids in innate antiviral defense. Cell Host Microbe *14*, 542 379-385.
- Sun, S., Zhao, G., Liu, C., Fan, W., Zhou, X., Zeng, L., Guo, Y., Kou, Z., Yu, H., Li, J., *et al.* (2014).
  Treatment With Anti-C5a Antibody Improves the Outcome of H7N9 Virus Infection in African
- 545 Green Monkeys. Clinical Infectious Diseases *60*, 586-595.
- Sun, X., Song, L., Feng, S., Li, L., Yu, H., Wang, Q., Wang, X., Hou, Z., Li, X., Li, Y., *et al.* (2018). Fatty
  Acid Metabolism is Associated With Disease Severity After H7N9 Infection. EBioMedicine *33*, 218229.
- Team, T.N.C.P.E.R.E. (2020). The epidemiological characteristics of an outbreak of 2019 novel
  coronavirus diseases (COVID-19) in China. *41*, 145-151.
- 551 Thevarajan, I., Nguyen, T.H.O., Koutsakos, M., Druce, J., Caly, L., van de Sandt, C.E., Jia, X., Nicholson,
- 552 S., Catton, M., Cowie, B., *et al.* (2020). Breadth of concomitant immune responses prior to patient
- recovery: a case report of non-severe COVID-19. Nature Medicine.
- Vernon-Roberts, B. (1969). The effects of steroid hormones on macrophage activity. Int Rev Cytol *25*, 131-159.
- 556 Wang, D., Hu, B., Hu, C., Zhu, F., Liu, X., Zhang, J., Wang, B., Xiang, H., Cheng, Z., Xiong, Y., et al.
- 557 (2020a). Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected
- 558 Pneumonia in Wuhan, China. JAMA.

- 559 Wang, R., Xiao, H., Guo, R., Li, Y., and Shen, B. (2015). The role of C5a in acute lung injury induced
- by highly pathogenic viral infections. Emerg Microbes Infect *4*, e28-e28.
- Wang, T., Du, Z., Zhu, F., Cao, Z., An, Y., Gao, Y., and Jiang, B. (2020b). Comorbidities and multi-
- organ injuries in the treatment of COVID-19. Lancet *395*, e52.
- Weigert, A., Weis, N., and Brune, B. (2009). Regulation of macrophage function by sphingosine-1phosphate. Immunobiology *214*, 748-760.
- Wilder-Smith, A., Earnest, A., and Paton, N.I. (2004). Use of simple laboratory features to
  distinguish the early stage of severe acute respiratory syndrome from dengue fever. Clin Infect Dis *39*, 1818-1823.
- 568 Wu, Z., and McGoogan, J.M. (2020). Characteristics of and Important Lessons From the Coronavirus
- 569 Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases From the 570 Chinese Center for Disease Control and Prevention. JAMA.
- 571 Yan, B., Chu, H., Yang, D., Sze, K.-H., Lai, P.-M., Yuan, S., Shuai, H., Wang, Y., Kao, R.Y.-T., Chan,
- 572 J.F.-W., *et al.* (2019). Characterization of the Lipidomic Profile of Human Coronavirus-Infected Cells:
- 573 Implications for Lipid Metabolism Remodeling upon Coronavirus Replication. Viruses *11*, 73.
- Zheng, Y., Zhang, Y., Chi, H., Chen, S., Peng, M., Luo, L., Chen, L., Li, J., Shen, B., and Wang, D.
- 575 (2020). The hemocyte counts as a potential biomarker for predicting disease progression in
- 576 COVID-19: a retrospective study. CCLM clinical chemistry and laboratory medicine *58*.
- 577 Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and
- 578 Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-579 level datasets. Nat Commun *10*, 1523.
- 580 Zou, Z., Yang, Y., Chen, J., Xin, S., Zhang, W., Zhou, X., Mao, Y., Hu, L., Liu, D., Chang, B., et al. (2004).
- 581 Prognostic factors for severe acute respiratory syndrome: a clinical analysis of 165 cases. Clin Infect
  582 Dis *38*, 483-489.
- 583
- 584
- 585

#### Molecular profiling of COVID-19 sera

## 586

587

## Table 1. Demographics and baseline characteristics of COVID-19 patients.

	Healthy	non-COVID-	COVID-19				
Variables	Control	19	Total	non-Severe	Severe		
	(N=28)	(N=25)	(N=46)	(N=25)	(N=21)		
Sex - no. <sup>a</sup> (%)							
Male	21 (75.0)	17 (68.0)	31 (67.4)	18 (72.0)	13 (61.9)		
Female	7 (25.0)	8 (32.0)	15 (32.6)	7 (28.0)	8 (38.1)		
Age - yr. <sup>b</sup>							
Mean ± SD	44.4±8.3	49.2±14.0	47.7±13.9	42.2±12.8	54.3±12.4		
Median (IQR)	45.0	53.0	46.5	43.0	55.0		
	(38.0-51.0)	(37.0-61.0)	(35.8-59.0)	(33.0-50.0)	(47.0-63.0)		
Range	28.0-57.0	23.0-67.0	20.0-75.0	20.0-70.0	30.0-75.0		
BMI, kg/m²							
Mean ± SD $^{\circ}$	24.4±2.7	23.5±2.7	24.9±3.0	24.5±3.3	26.0±2.4		
Median (IQR) <sup>d</sup>	24.2	24.7	24.8	24.9	25.8		
	(22.5-26.5)	(20.9-25.8)	(22.7-27.0)	(21.7-26.8)	(24.4-28.0)		
Range	19.9-32.9	19.1-27.4	19.0-31.3	18.9-30.4	22.2-31.3		
Smoke - no. (%)			5 (10.9)	2 (8.0)	1 (4.8)		
Alcohol - no. (%)			1 (2.2)	1 (4.0)	0 (0.0)		
Time from onset to admission, days							
Mean ± SD			3.0±4.2	1.6±2.1	4.6±5.4		
Median (IQR)			1.5	1.0	2.0		
			(0.0-4.0)	(0.0-2.0)	(0.5-6.0)		
Range			0.0-20.0	0.0-10.0	0.0-20.0		
Time from admission to severe, days							
Mean ± SD					4.0±2.1		
Median (IQR)					4.0 (2.5-5.5)		
Range					0.0-7.0		
Symptoms - no. (%)							
Fever		9 (36.0)	34 (73.9)	16 (64.0)	18 (85.7)		
Cough		11 (44.0)	26 (56.5)	13 (52.0)	13 (61.9)		
Headache		2 (8.0)	7 (15.2)	4 (16.0)	3 (14.3)		
Fatigue		2 (8.0)	8 (17.4)	5 (20.0)	3 (14.3)		
Pharyngalgia		0 (0.0)	3 (6.5)	1 (4.0)	2 (9.5)		
Expectoration		3 (12.0)	12 (26.1)	4 (16.0)	8 (38.1)		
Diarrhea		0 (0.0)	3 (6.5)	1 (4.0)	2 (9.5)		
Chest tightness		2 (8.0)	1 (2.2)	1 (4.0)	0 (0.0)		
Chest CT <sup>e</sup> - no. (%)							
Involvement of chest ra	1 (4.0)	45 (97.8)	24 (96.0)	21 (100)			
Comorbidity- no. (%)							
Hypertension		10 (40.0)	6 (13.0)	1 (4.0)	5 (23.8)		
Diabetes		0 (0.0)	6 (13.0)	3 (12.0)	3 (14.3)		

SHEN, ET AL

Molecular profiling of COVID-19 sera							
Respiratory system	2 (8.0)	4 (8.7)	3 (12.0)	1 (4.8)			
Endocrine system	0 (0.0)	3 (6.5)	0 (0.0)	3 (14.3)			
Chronic kidney disease	0 (0.0)	1 (2.2)	0 (0.0)	1 (4.8)			
Digestive system	2 (8.0)	6 (13.0)	3 (12.0)	3 (14.3)			
Oxygenation Index - mmHg							
Mean ± SD			478.4±147.5	347.4±124.1			
Median (IQR)			447.6	341.4			
			(390.5-506.0)	(266.5-381.9)			
Range			289.7-890.5	133.3-728.6			
Treatment - no. (%)							
Oxygen inhalation	3 (12.0)	30 (65.2)	10 (40.0)	20 (95.2)			
Antibiotics	0 (0.0)	15 (32.6)	8 (32.0)	7 (33.3)			
Antiviral drug	25 (100)	46 (100.0)	25 (100.0)	21 (100.0)			
Immunoglobulin	0 (0.0)	19 (41.3)	0 (0.0)	19 (90.5)			
Methylprednisolone	0 (0.0)	16 (34.8)	0 (0.0)	16 (76.2)			
Chinese medicine	0 (0.0)	45 (97.8)	25 (100.0)	20 (95.2)			

588 <sup>a</sup> no. (%): number.

589 <sup>b</sup> yr.: year.

<sup>c</sup> SD: Standard Deviation.

591 <sup>d</sup> IQR: Interquartile range

<sup>e</sup> CT: Computed tomography

593

#### Molecular profiling of COVID-19 sera

Figure 1. Summary of COVID-19 patients and machine learning design. 594 (A) Summary of COVID-19 patients, including non-severe (n=25) and severe 595 (n=21) patients with more details in Table S1. Patients labeled in red (Y-axis) 596 indicate chronic infection of viruses including HBV. (B) Study design for 597 machine learning-based classifier development for severe COVID-19 patients. 598 We first procured samples in a training cohort for proteomic and metabolomic 599 analysis. The classifier was then validated in an independent cohort. 600 Figure 2. Separation of severe and non-severe COVID-19 patients by 601 machine learning of proteomic and metabolomic features. (A) Top 22 602 proteins and 7 metabolites prioritized by random forest analysis ranked by the 603 mean decrease in accuracy. (B) Network of prioritized proteins appeared in 604 the classifier. (C) ROC of the random forest model in the training cohort. (D) 605 Principal Components Analysis (PCA) plot of COVID-19 patients from the 606 training cohort. (E) Performance of the model in the validation cohort of ten 607 COVID-19 patients. Patients labeled in red received serum test before they 608 609 were diagnosed as severe. Figure 3. Dysregulated proteins in COVID-19 sera. (A) Heatmap of 50 610 selected proteins whose regulation concentrated on three enriched pathways. 611 (B) The expression level change (z-scored original value) of six selected 612 proteins of interest with significance indicated by the asterisks (unpaired two 613 sided Welch's t test. p value: \*\*\*, <0.001; \*\*, <0.01; \*, <0.05) 614 Figure 4. Dysregulated metabolites in COVID-19 sera. (A) Heatmap of 82 615 regulated metabolites belonging to six major classes: fatty acids, steroids, 616 glycerophospholipid, sphingolipid, choline and serotonins. (B) The expression 617 level change (z-scored log2 original value) of six regulated metabolites from 618 each metabolite class with significance indicated by the asterisks as 619 described in Figure 3. 620 Figure 5. Key proteins and metabolites characterized in severe COVID-621 19 patients in a working model. SARS-CoV-2 may target alveolar 622 macrophages via ACE2 receptor with increasing level of cytokines including 623 IL-6 and TNF- $\alpha$ , which subsequently induce the elevation of various APPs 624 such as SAP, CRP, SAA1, SAA2 and C6 which are significantly upregulated 625 in the severe group. Proteins involved in macrophage, lipid metabolism and 626 platelet degranulation were indicated with their corresponding expression 627 levels in four patient groups. 628 629

Molecular profiling of COVID-19 sera

## 630 MATERIALS AND METHODS

### 631 **Patients and samples**

Our team procured serum samples from 46 patients who visited Taizhou 632 Public Health Medical Center during January 23 and February 4, 2020. They 633 were diagnosed as COVID-19 according to the Chinese Government 634 Diagnosis and Treatment Guideline (Trial 5th version)(Medicine, 2020). For 635 diagnosing COVID-19, nucleic acid from sputum or throat swab was extracted 636 using nucleic acid extractor (Shanghai Zhijiang, China, EX3600) and virus 637 nucleic acid extraction reagent (Shanghai Zhijiang, China, NO. P20200201) 638 was used to extract nucleic acid. Fluorescence guantitative PCR (ABI7500) 639 and SARS-CoV-2 nucleic acid detection kit (triple fluorescence PCR, 640 Shanghai Zhijiang, China, NO. P20200203) were used for nucleic acid 641 detection. This kit uses one step RT-PCR combined with Tagman technology 642 to detect RdRp, E and N genes. Positive was concluded if RdRp gene was 643 644 positive (Ct < 43), and one of E or N was positive (Ct <43). Patients were also 645 diagnosed as positive if two sequential tests of RdRp were positive while E and N were negative. According to the abovementioned guideline, COVID-19 646 patients are classified into four subgroups: 1) Mild: mild symptoms without 647 pneumonia; 2) Typical: fever or respiratory tract symptoms with pneumonia; 3) 648 Severe: fulfill any of the three criteria: respiratory distress, respiratory rate  $\geq$ 649 30 times/min; means oxygen saturation  $\leq$  93% in resting state; arterial blood 650 oxygen partial pressure/oxygen concentration  $\leq$  300 mmHg (1 mmHg = 0.133 651 kPa); 4) Critical: fulfill any of the three criteria: respiratory failure and require 652 mechanical ventilation; shock incidence; admission to ICU with other organ 653 failure. In this study, we included both severe and no-severe patients, with the 654 latter composed of mild and typical COVID-19 patients. Last follow-up of 655 these patients was March 10, 2020. We also procured 25 non-COVID-19 656 patients with similar clinical characteristics including fever and/or cough as 657 COVID-19 patients however negative in the nucleic acid test. Based on the 658 Chinese Government Diagnosis and Treatment Guideline (Trial 5th version) 659 (Medicine, 2020), patients are defined as suspected COVID-19 cases when 660 they meet the following three clinical criteria: 1) fever or respiratory symptoms, 661 2) imaging manifestation of pneumonia, and 3) optional reduction of white 662 blood cell or lymphocyte count at early stage. The patients only need to meet 663 at least two of the above three criteria if they have been exposed to COVID-664 19 individuals. We also collected serum samples from 28 healthy individuals 665 as control. All the serum samples were collected as venous whole blood in the 666 early morning before diet using serum separation tubes (Zhejiang GongDong 667 Medical Technology Co., Ltd, China). The blood samples were centrifuged at 668 3,500 rpm for 10 min for serum collection. The serum samples were frozen at 669 670 -80°C. The samples from this study is from a clinical trial that our team initiated and registered in the Chinese Clinical Trial Registry with a ID of 671 ChiCTR2000031365. This study has been approved by the 672

### Molecular profiling of COVID-19 sera

673 Ethical/Institutional Review Board of Taizhou Public Health Medical Center

and Westlake University. Contents from patients were waived by the boards.

675

### 676 **Proteomic analysis**

Serum samples were inactivated and sterilized at 56°C for 30 min, and 677 processed as previously with some modifications. Five µL serum from each 678 specimen was lysed in 50 µL lysis buffer (8 M urea in 100 mM 679 triethylammonium bicarbonate, TEAB) at 32°C for 30 min. The lysates were 680 reductive with 10 mM tris (2-carboxyethyl) phosphine (TCEP) for 30 min at 681 32°C, then alkylated for 45 min with 40 mM iodoacetamide (IAA) in darkness 682 at room temperature (25°C). The protein extracts were diluted with 200 µL 100 683 mM TEAB, and digested with double-step trypsinization (Hualishi Tech. Ltd, 684 Beijing, China), each step with an enzyme-to-substrate ratio of 1:20, at 32°C 685 for 60 min. The reaction was stopped by adding 30 µL 10% trifluoroacetic acid 686 (TFA) in volume. Digested peptides were cleaned-up with SOLAµ (Thermo 687 Fisher Scientific<sup>™</sup>, San Jose, USA) following the manufacturer's instructions, 688 and lableled with TMTpro 16plex label reagents (Thermo Fisher Scientific™, 689 San Jose, USA) as described previously. The TMT samples were fractionated 690 using a nanoflow DIONEX UltiMate 3000 RSLCnano System (Thermo Fisher 691 Scientific<sup>™</sup>, San Jose, USA) with an XBridge Peptide BEH C18 column (300 692 Å, 5 µm × 4.6 mm × 250 mm) (Waters, Milford, MA, USA)(Gao et al., 2020). 693 The samples were separated using a gradient from 5% to 35% acetonitrile 694 695 (ACN) in 10 mM ammonia (pH=10.0) at a flow rate of 1 mL/min. Peptides were separated into 120 fractions, which were consolidated into 40 fractions. 696 The fractions were subsequently dried and re-dissolved in 2% ACN/0.1% 697 formic acid (FA). The re-dissolved peptides were analyzed by LC-MS/MS with 698 the same LC system coupled to a Q Exactive HF-X hybrid Quadrupole-699 Orbitrap (Thermo Fisher Scientific<sup>™</sup>, San Jose, USA) in data dependent 700 acquisition (DDA) mode. For each acquisition, peptides were loaded onto a 701 precolumn (3 µm, 100 Å, 20 mm\*75 µm i.d.) at a flowrate of 6 µL/min for 4 min 702 and then injected using a 35 min LC gradient (from 5% to 28% buffer B) at a 703 flowrate of 300 nL/min (analytical column, 1.9 µm, 120 Å, 150 mm\*75 µm i.d.). 704 Buffer A was 2%ACN, 98% H<sub>2</sub>O containing 0.1% FA, and buffer B was 98% 705 ACN, 2% H<sub>2</sub>O containing 0.1% FA. All reagents were MS grade. The m/z 706 range of MS1 was 350-1,800 with the resolution at 60,000 (at 200 m/z), AGC 707 target of 3e6, and maximum ion injection time (max IT) of 50 ms. Top 15 708 precursors were selected for MS/MS experiment, with a resolution at 45,000 709 (at 200 m/z), AGC target of 2e5, and max IT of 120 ms. The isolation window 710 of selected precursor was 0.7 m/z. The resultant mass spectrometric data 711 were analyzed using Proteome Discoverer (Version 2.4.1.15, Thermo Fisher 712 Scientific) using a protein database composed of the Homo sapiens fasta 713 714 database downloaded from UniprotKB on 07 Jan 2020 and the SARS-CoV-2 virus fasta downloaded from NCBI (version NC 045512.2). Enzyme was set 715 to trypsin with two missed cleavage tolerance. Static modifications were set to 716

### Molecular profiling of COVID-19 sera

carbamidomethylation (+57.021464) of cysteine, TMTpro (+304.207145) of 717 lysine residues and peptides' N termini, and variable modifications were set to 718 oxidation (+15.994915) of methionine and acetylation (+42.010565) of 719 peptides' N-termini. Precursor ion mass tolerance was set to 20 ppm, and 720 product ion mass tolerance was set to 0.06 Da. The peptide-spectrum-match 721 allowed 1% target false discovery rate (FDR) (strict) and 5% target FDR 722 (relaxed). Normalization was performed against the total peptide amount. The 723 other parameters followed the default setup. 724

725

## 726 Quality control of proteomic data

The quality of proteomic data was ensured at multiple levels. First, a mouse
liver digest was used for instrument performance evaluation. We also run
water samples (buffer A) as blanks every 4 injections to avoid carry-over.
Serum samples of four patient groups from both training and validation
cohorts were randomly distributed in eight different batches. Six samples were
injected in technical replicates.

733

### 734 Metabolomic analysis

Ethanol was added to the serum samples and shaken vigorously to 735 inactivate any potential viruses, then dried in a biosafety hood. The dried 736 samples were further treated for metabolomics analysis. The metabolomic 737 analysis was performed as described previously(Lee et al., 2019). Briefly, 738 739 deactivated serum samples, 100 µL each, were extracted by adding 300 µL extraction solution. The mixtures were shaken vigorously for 2 min. Proteins 740 were denatured and precipitated by centrifugation. The supernatants 741 contained metabolites of diverse chemical natures. To ensure the quantity 742 and reliability of metabolite detection, four platforms were performed with non-743 target metabolomics. Each supernatant was divided into four fractions: two for 744 745 analysis using two separate reverse-phase /ultra-performance liquid chromatography (RP/UPLC)-MS/MS methods with positive ion-mode 746 electrospray ionization (ESI), one for analysis using RP/ UPLC-MS/MS with 747 negative-ion mode ESI, and one for analysis using hydrophilic interaction 748 liquid chromatography (HILIC)/UPLC-MS/MS with negative-ion mode ESI. 749 Each fraction was dried under nitrogen gas to remove the organic solvent and 750 later re-dissolved in four different reconstitution solvents compatible with each 751 of the four UPLC-MS/MS methods. 752

All UPLC-MS/MS methods used ACQUITY 2D UPLC system (Waters, 753 Milford, MA, USA) and Q Exactive HF hybrid Quadrupole-Orbitrap (Thermo 754 Fisher Scientific<sup>™</sup>, San Jose, USA) with HESI-II heated ESI source and 755 Orbitrap mass analyzer. The mass spectrometer was operated at 35,000 756 mass resolution (at 200 m/z). In the first UPLC-MS/MS method, the QE was 757 758 operated under positive electron spray ionization (ESI) coupled with a C18 column (UPLC BEH C18, 2.1 × 100 mm, 1.7 µm; Waters) was used in UPLC. 759 The mobile solutions used in the gradient elution were water and methanol 760

### Molecular profiling of COVID-19 sera

containing 0.05% perfluoropentanoic acid (PFPA) and 0.1% FA. In the second 761 method, the QE was still operated under ESI positive mode, and the UPLC 762 used the same C18 column as in method one, but the mobile phase solutions 763 were optimized for more hydrophobic compounds and contained methanol, 764 acetonitrile, water, 0.05% PFPA, and 0.01% FA. The third method had the QE 765 operated under negative ESI mode, and the UPLC method used a C18 766 column eluted with mobile solutions containing methanol and water in 6.5 mM 767 ammonium bicarbonate at pH 8. The UPLC column used in the fourth method 768 was HILIC column (UPLC BEH Amide, 2.1 × 150 mm, 1.7 µm; Waters), and 769 the mobile solutions were consisted of water and acetonitrile with 10 mM 770 ammonium formate at pH 10.8; the QE was operated under negative ESI 771 mode. The QE mass spectrometer analysis was alternated between MS and 772 data-dependent MS2 scans using dynamic exclusion. The scan range was 773 70-1,000 *m/z*. 774

After raw data pre-processing, peak finding/alignment, and peak 775 annotation using in-house software, metabolites were identified by searching 776 an in-house library containing more than 3,300 standards with library data 777 entries generated from running purified compound standards through the 778 experimental platforms. Identification of metabolites must meet three strict 779 criteria: narrow window retention index (RI), accurate mass with variation less 780 than 10 ppm and MS/MS spectra with high forward and reverse scores based 781 on comparisons of the ions present in the experimental spectrum to the ions 782 783 present in the library spectrum entries. Almost all isomers can be distinguished by these three criteria. All identified metabolites meet the level 1 784 requirements by the Chemical Analysis Working Group (CAWG) of the 785 Metabolomics Standards Initiative (MSI) expect some asterisk labeled lipids 786 which MS/MS spectral were in silico matched. 787

788

### 789 **Quality control of metabolomics analysis**

Several types of quality control samples were included in the experiment: 790 a pooled sample generated by taking a small volume of each experimental 791 sample to serve as a technical replicate that was run multiple times 792 throughout the experiment, extracted water samples served as blanks, and 793 extracted commercial plasma samples for monitoring instrument variation. A 794 mixture of internal standards was also spiked into every sample to aid 795 chromatographic peak alignment and instrument stability monitoring. 796 Instrument variability was determined by calculating the median relative SD 797 (RSD) of all internal standards in each sample. The experimental process 798 variability was determined by calculating the median RSD for all endogenous 799 metabolites present in the pooled quality control samples. 800

801

### 802 Statistical analysis

803 Metabolites and therapeutic compounds with over 80% missing ratios in a 804 particular patient group were removed for the metabolomics dataset

### Molecular profiling of COVID-19 sera

containing endogenoeous metabolites while full proteomics features were 805 used for the subsequent statistical analysis. Missing values were imputated 806 with the minimal value for each feature. Log2 fold-change (log2 FC) was 807 calculated on the mean of the same patient group for each pair of comparing 808 groups. Two-sided unpaired Welch's t test was performed for each pair of 809 comparing groups and adjusted p values were calculated using Benjamini & 810 Hochberg correction. The statistical significantly changed proteins or 811 metabolites were selected using the criteria of adjust p value less than 0.05 812 indicated and absolute log2 FC larger than 0.25. From the training cohort, the 813 important features were selected with mean decrease accuracy larger than 3 814 using random forest containing a thousand trees using R package 815 randomForest (version 4.6.14) random forest analysis with 10-fold cross 816 validation as binary classification of paired severe and non-severe group 817 using combined differentially regulated proteins and metabolites features. The 818 random forest analysis was further performed for a hundred times on the 819 820 matrix with only the selected important features using normalized additive predicting probability as the final predicting probability and the larger 821 probability as the predictive label. Those selected important features were 822 used for the random forest analysis on the independent validation cohort. 823 824

### 825 Pathway analysis

Four network pathway analysis tools were used for pathway analysis using 93 826 827 differentially expressed proteins (DEPs). The top Gene Ontology (GO) processes were enriched by Metascape web-based platform (Zhou et al., 828 2019). The GO terms is enriched using the Cytoscape plug-in ClueGO 829 (Bindea et al., 2009). Ingenuine pathway analysis (Kramer et al., 2014) of the 830 regulated proteins identifies most significantly relevant pathways with p value 831 of determined based on right-tailed Fisher's Exact Test with the overall 832 activation or inhibition states of enriched pathways were predicted by z-score. 833 Functional co-expression network analysis by GeNet(Li et al., 2018a) to 834 represent statistical co-expressed protein modules. 835 836

Molecular profiling of COVID-19 sera

Figure S1. Quality control of proteomic and metabolomic data. (A) 837 Coefficient of variation (CV) of the proteomic data is calculated by the proteins 838 quantified in six quality control (QC) samples using the pooled samples from 839 all samples in training cohort. CV of the metabolomic data is calculated by 840 twelve QC samples using a set of isotopic internal spiked-in standards. (B) 841 UMAP of sera samples using 791 measured proteins. (C) UMAP of sera 842 samples using 847 metabolites. 843 Figure S2. Differentially expressed proteins in different patient groups in 844 the training cohort. Volcano plots compare four pairs of patient groups as 845 indicated in the plot. Proteins with log2 (fold-change) beyond 0.25 or below -846 0.25 with adjusted p value lower than 0.05 were considered as significantly 847 differential expression. Number of significantly down- (blue) and up- (red) 848 regulated proteins were shown on the top. 849 Figure S3. Differentially expressed metabolites in different patient 850 groups in the training cohort. The volcano plots follow that for Figure S3 851 except metabolomic data were used as the input. 852 Figure S4. Proteins and metabolites regulated in COVID-19 patients but 853 not in non-COVID-19 patients. Venn diagrams showing the overlaps 854 between significantly regulated proteins (A) and metabolites (B) as identified 855 in volcano plots. Proteins and metabolites labeled in red are the shortlisted 856 molecules which differentially expressed in the COVID-19 patients but not in 857 the non-COVID-19 patients. 858 Figure S5. Identification of specific clusters of proteins and metabolites 859 in COVID-19 patients. 791 proteins (A) and 941 metabolites (B) were 860 clustered using mFuzz into 16 significant discrete clusters, respectively, to 861 illustrate the relative expression changes of the proteomics and metabolomics 862 data. The groups in proteomics data: 1: Healthy; 2: non-Severe COVID-19; 3: 863 Severe COVID-19. The groups in metabolomics data: 1: Healthy; 2: non-864 COVID-19; 3: non-Severe COVID-19; 4: Severe COVID-19. 865 Figure S6. Pathway analysis of 93 differential expressed proteins in 866 COVID-19 patients. (A) The Gene Ontology (GO) processes enriched by 867 Metascape. (B) The GO terms enriched using the Cytoscape plug-in ClueGO. 868 (C) Ingenuine pathway analysis of most significantly relevant pathways with 869 the predicted activation or inhibition state. (D) Functional network analysis by 870 GeNet identifies several communities. 871





















