

Early Prediction of COVID-19 Severity

Using Extracellular Vesicles and Extracellular RNAs

Yu Fujita^{1,2+*}, Tokio Hoshina³⁺, Juntaro Matsuzaki⁴⁺, Tsukasa Kadota^{1,2}, Shota Fujimoto², Hironori Kawamoto², Naoaki Watanabe², Kenji Sawaki³, Yohei Sakamoto³, Makiko Miyajima³, Kwangyole Lee³, Kazuhiko Nakaharai³, Tetsuya Horino³, Ryo Nakagawa⁵, Jun Araya², Mitsuru Miyato⁴, Masaki Yoshida³, Kazuyoshi Kuwano², and Takahiro Ochiya^{4*}

¹ Department of Translational Research for Exosomes, The Jikei University School of Medicine, Tokyo, Japan.

² Division of Respiratory Diseases, Department of Internal Medicine, The Jikei University School of Medicine, Tokyo, Japan.

³ Department of Infectious Diseases and Infection Control, The Jikei University School of Medicine, Tokyo, Japan.

⁴ Department of Molecular and Cellular Medicine, Institute of Medical Science, Tokyo Medical University, Tokyo, Japan.

18 ⁵Omiya City Clinic, Saitama, Japan.

19 [†]These authors equally contributed to this work.

20

21 ***Address correspondence to:**

22 Yu Fujita, The Jikei University School of Medicine, 3-25-8 Nishi-shimbashi, Minato-ku,

23 Tokyo, 105-8461, Japan, Phone: (+81)-3-3433-1111, Fax: (+81)-3-3433-1020, E-mail:

24 yuugot@jikei.ac.jp

25 Takahiro Ochiya, Tokyo Medical University, 6-7-1 Nishi-shinjuku, Shinjuku-ku, Tokyo,

26 160-0023, Japan, Phone: (+81)-3-3342-6111, Fax: (+81)-3-6302-0265, E-mail:

27 tochiya@tokyo-med.ac.jp

28

29 **Competing interest statement:** Mitsuru Miyato is a chief executive officer, and

30 Takahiro Ochiya is a chief scientific advisor of International Space Medical Co., Ltd.

31 The other authors have declared that no conflict of interest exists.

32

33 **Keywords:** SARS-CoV-2; COVID-19; extracellular vesicle; extracellular RNA;

34 biomarker

35 **Abstract:**

36 The clinical manifestations of COVID-19 vary broadly, ranging from asymptomatic
37 infection to acute respiratory failure and death. But the predictive biomarkers for
38 characterizing the variability are still lacking. Since emerging evidence indicates that
39 extracellular vesicles (EVs) and extracellular RNAs (exRNAs) are functionally involved
40 in a number of pathological processes, we hypothesize that these extracellular
41 components may be key determinants and/or predictors of COVID-19 severity. To test
42 our hypothesis, we collected serum samples from 31 patients with mild COVID-19
43 symptoms at the time of their admission. After standard therapy without corticosteroids,
44 9 of the 31 patients developed severe COVID-19 symptoms. We analyzed EV protein
45 and exRNA profiles to look for correlations between these profiles and COVID-19
46 severity. Strikingly, we identified three distinct groups of markers (antiviral
47 response-related EV proteins, coagulation-related markers, and liver damage-related
48 exRNAs) with the potential to serve as early predictive biomarkers for COVID-19
49 severity. Among these markers, EV COPB2 has the best predictive value for severe
50 deterioration of COVID-19 patients in this cohort. This type of information concerning
51 functional extracellular component profiles could have great value for patient

52 stratification and for making early clinical decisions about strategies for COVID-19

53 therapy.

54

55 **Introduction**

56 Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is
57 responsible for a rapidly-unfolding pandemic that is overwhelming health care systems
58 worldwide ¹. The majority of Coronavirus-disease 2019 (COVID-19) patients exhibit
59 mild clinical symptoms, including fever, cough, and sputum. In some cases, more
60 severe symptoms such as dyspnea and/or hypoxemia occur after 1 week, with 50% of
61 these severely-affected patients quickly progressing to systemic, life-threatening
62 disorders, including acute respiratory distress syndrome (ARDS), septic shock,
63 refractory metabolic acidosis, coagulation disorders, and multiple organ failure ². These
64 problems lead to both mild and severe respiratory manifestations, the latter being
65 prominent in the elderly and in those with underlying medical conditions such as
66 cardiovascular and chronic respiratory disease, diabetes, and cancer ². The main clinical
67 feature of severe COVID-19 is the onset of ARDS ³. It has been reported that the rate of
68 occurrence of ARDS in patients with severe COVID-19 is 15.9%-29% ^{2,4}. The immune
69 response that is vital for the control and resolution of SARS-CoV-2 infections can also
70 lead to cellular damage in association with systemic deteriorations. Recent studies
71 suggest that an exaggerated inflammatory response known as cytokine storm may play a

72 key role in this process ⁵. The variability of patient responses to SARS-CoV-2 infection
 73 makes it difficult to identify individuals who may be most at-risk for severe responses
 74 to the disease and to create early strategies for treating the disease. Although PCR-based
 75 detection of SARS-CoV-2 is useful for accurately diagnosing the disease, there is no
 76 available tool for effectively predicting disease severity or likely trajectories of illness
 77 progression ⁶.

78 Early identification of COVID-19 patients who may experience severe disease
 79 deterioration is of great significance for improving clinical approaches that will result in
 80 reduced mortality. Accordingly, the purpose of our study has been to identify
 81 extracellular components that can serve as useful biomarkers for predicting disease
 82 progression in COVID-19 patients. Toward this end we have focused on identifying the
 83 profiles of extracellular vesicle (EV) proteins and extracellular RNAs (exRNAs) in
 84 serum samples taken from COVID-19 patients with mild symptoms at the time of their
 85 admission for medical treatment.

86 Extracellular vesicles (EVs), including exosomes and microvesicles, contain
 87 numerous and diverse bio-molecules such as RNAs and proteins within the vesicle's
 88 lipid bilayer ⁷. This stable bilayer construction enables intact EVs to circulate through

89 body fluids. Thus, EVs can transfer their cargo to target cells as a means of regulating
90 target cell participation in both physiological and pathological processes, including host
91 immune responses. EVs are characterized by a number of exosomal markers, including
92 members of the tetraspanin family (CD9, CD63, CD81), components of the endosomal
93 sorting complex required for transport (ESCRT) (TSG101, Alix), heat shock proteins
94 (Hsp60, Hsp70, Hsp90), and RAB proteins (RAB27a/b)⁸. Accumulating evidence
95 suggests that the size, membrane composition, and contents of EVs are highly
96 heterogeneous, depending on the cellular source. Furthermore, EV cargoes can be
97 dynamically altered by microenvironmental stimuli including viral infection. The EVs
98 produced by both immune and non-immune cells can be responsible for regulating the
99 nature of the immune response as it relates to inflammation, autoimmunity, and
100 infectious disease pathology⁹. EVs derived from virus-infected cells have been shown
101 not only to modulate immune cell responses, but also to spread the viral infection via
102 delivery of the viral genome and protein particles to healthy cells¹⁰. Intriguingly, EVs
103 have receptors for coronavirus entry such as CD9¹¹ and angiotensin-converting enzyme
104 2 (ACE2)¹², suggesting that EVs may also have a role in modulating or mediating
105 SARS-CoV-2 infection. For these reasons, evaluating the molecular profiles of

106 circulating EVs in COVID-19 patients may provide clues for understanding the host
107 antiviral immune response and the mechanisms that determine disease trajectory and
108 patient deterioration.

109 Studies of extracellular RNA (exRNA) have recently broadened into an
110 important area of research with relevance to disease biomarker discovery and
111 therapeutics ¹³. Most exRNAs are protected from degradation in biofluids via
112 incorporation into EVs or into complexes with lipids and proteins. The variety of
113 exRNA species that have been identified include messenger RNAs (mRNAs) and
114 non-coding RNAs such as microRNAs (miRNAs), small nuclear RNAs, transfer RNAs,
115 and long non-coding RNAs (lncRNAs). During the early stages of viral infection, host
116 non-coding RNAs such as miRNAs are produced and released as a part of the antiviral
117 response that is aimed either directly or indirectly at targeting viral transcription,
118 translation, and replication ¹⁴. ExRNAs mediate a complex network of interactions
119 between the virus and infected host cells, representing a pivotal role in modulating the
120 host innate immune system ¹⁵. These findings suggest that exRNAs may have the
121 potential to serve as biomarkers for evaluating the antiviral responses mounted by
122 COVID-19 patients.

123

124 **Results**

125 *Baseline clinical characteristics of the patient cohort*

126 For this retrospective, single-center study, 31 SARS-CoV-2 infected patients
 127 and 10 uninfected healthy donors were included in the patient cohort. At the time of
 128 patient admission for treatment, serum samples were taken from all 41 individuals as a
 129 source of material for identifying biomarkers (**Table S1**). The severity of COVID-19
 130 disease in each patient was graded using the WHO ordinal outcome scale of clinical
 131 improvement (**Table S2**). At the time of admission, all COVID-19 patients were scored
 132 as having a mild status (WHO score = 3). Based on the clinical course of disease
 133 progression after admission, we divided the 31 COVID-19 patients into two groups:
 134 Group 1 included patients who retained their mild status (WHO score ≤ 4), and Group 2
 135 included patients who progressed to severe status (WHO score ≥ 5) (**Figure S1a**). All
 136 patients in Group 1 were subsequently discharged in good health from the hospital or
 137 transferred to the local medical facility for further observation. In Group 2, two patients
 138 (22.2%) died from COVID-19 complications (pulmonary embolism ($n=1$) or ARDS
 139 ($n=1$)), and two patients (22.2%) remained hospitalized at the time of writing. All other
 140 patients in Group 2 were discharged in good health from the hospital following

141 treatment (**Table S1**). Clinical parameters were comparable among the groups (**Table**
 142 **S3**). Healthy donors and COVID-19 patients (Group 1 & 2) did not differ with respect
 143 to age, sex, body mass index (BMI), smoking index, levels of blood urea nitrogen
 144 (BUN), creatinine (Cr), and alanine aminotransferase (ALT), history of hypertension,
 145 diabetes mellitus, dyslipidemia, and coronary heart disease ($P > 0.05$). Significant
 146 differences were observed in white blood cell (WBC) count and levels of C-reactive
 147 protein (CRP) ($P < 0.05$). Comparison of COVID-19 patients in Group 1 and Group 2
 148 revealed no significant difference in sex, BMI, WBC count, BUN, Cr, creatine kinase
 149 (CK), D-dimer, and fibrinogen, history of hypertension, diabetes mellitus, dyslipidemia,
 150 and coronary heart disease ($P > 0.05$). However, significant differences were observed
 151 with respect to age, smoking index, and levels of CRP and ALT ($P < 0.05$). We selected
 152 these 4 parameters for Pearson's correlation analysis to identify correlations between
 153 the three groups in terms of age, smoking index, CRP, and ALT (**Figure S1b**). There
 154 were significant correlations in age, CRP, and ALT between the three groups (P trend
 155 < 0.05).

156

157 *Predictive value of 4 clinical parameters for COVID-19 severity*

158 We performed receiver operating characteristic (ROC) analysis on combined
159 Group1 and Group 2 to provide a robust test of sensitivity, specificity, and area under
160 the curve (AUC) for our set of four predictive parameters (**Figure S1c**). The AUC
161 values for the four factors (age, smoking index, CRP, and ALT) were 0.90 (95% CI:
162 0.79-1.00), 0.69 (95% CI: 0.48-0.89), 0.77 (95% CI: 0.60-0.94), and 0.72 (95% CI:
163 0.51-0.94), respectively. These findings suggest that at the time of admission to the
164 hospital, age and CRP, having AUC values > 0.75, can provide the greatest
165 discrimination between patients who will develop mild versus severe COVID-19.

166 Based on the ROC analysis, optimal cutoffs for the 4 parameters were
167 identified via use of the Youden index (**Table S6**). For further analysis, we used these
168 cutoffs to separate COVID-19 patients into low and high groups for comparing the
169 incidence of severe COVID-19 related events. Kaplan-Meier curves were constructed
170 for each of the four parameters to compare the high and low groups for time-to-onset of
171 a severe event (beginning with the day of admission) (**Figure S1d**). The
172 progression-free times for age and CRP were significantly shorter in the high group than
173 in the low group ($P=3.9 \times 10^{-6}$; $P=0.014$, respectively). These findings suggest that the
174 parameters of patient age and CRP can be useful at the time of admission to predict the

175 incidence of future severe COVID-19 related events. Currently, age and/or CRP levels
176 are consistently used in the clinic for risk stratification to predict the potential severity
177 of COVID-19 progression ^{2,16}. Results with our cohort of 31 COVID-19 patients are
178 therefore representative of the overall status of patients infected with SARS-CoV-2.

179

180 *LC-MS analysis of proteome profiles from EVs in serum samples from COVID-19* 181 *patients and uninfected controls*

182 In clinical settings, analysis of EVs from liquid biopsies has gained attention as a
183 potential means of identifying diagnostic and prognostic biomarkers for various
184 diseases. However, this strategy has not yet been widely used due to a lack of
185 standardized methods for isolating EVs from patients. To improve on this situation, we
186 have employed an immunoprecipitation (IP) based method that targets surface marker
187 proteins for rapid and specific isolation of EVs. The use of IP in the presence of a
188 chelating reagent improves the yield and purity of CD9⁺ or CD63⁺ positive EVs from
189 serum samples. The resulting EV preparations are suitable for subsequent EV proteome
190 analysis by LC-MS (**Figure 1a**). From our total of 41 serum samples, LC-MS analysis
191 identified 1676 proteins, following exclusion of proteins that were not present in all

192 samples. Of these 1676 proteins, a total 723 proteins were present at different levels
193 between the three groups ($P < 0.05$; one-way ANOVA). To compare the pattern of
194 expression of these 723 EV proteins among the three patient cohorts, we used
195 unsupervised multivariate statistics based on principal component analysis (PCA)
196 mapping. PCA plots the first and second principal components (PC1 and PC2) using all
197 the term frequency from the LC-MS data, showing a certain degree of separated trends
198 between the three groups (**Figure 1b**). This differentiation could be described by the
199 first PC1, which accounted for 28.1% of the variance. The second PC2 accounted for
200 14.5% of the variance. This observation indicated that the EV proteome profiles of
201 serum in COVID-19 patients deviated considerably from those uninfected donors.
202 Although a little overlap or dispersity was demonstrated, we found obvious EV
203 proteomic differences between Group 1 and Group 2 in the PCA scores plot.

204 To further identify differences in EV proteins from Group 1 and Group 2
205 patients, a cross-validation score ¹⁷, which indicates the robustness of discrimination
206 performance between them, was calculated on the basis of Fisher linear discriminant
207 analysis for each of the selected proteins. From the candidate proteins, we have listed 91
208 proteins with cross-validation scores > 0.75 (**Table S4**). Further, we show the abundance

209 of the top 9 proteins with cross-validation scores >0.85 between the three patient groups
 210 (**Figure 1c**). Among these most-abundant proteins, four proteins [COPI coat complex
 211 subunit beta 2 (COPB2), KRAS proto-oncogene (KRAS), protein kinase C beta
 212 (PRKCB), and ras homolog family member C (RHOC)] are significantly more abundant
 213 in Group 1 than in Group 2 or in uninfected controls (P trend >0.05). Furthermore,
 214 CD147, calpain 2 (CAPN2), extracellular matrix protein 1 (ECM1), fibrinogen gamma
 215 chain (FGG) were significantly more abundant in Group 2 than in uninfected controls or
 216 in Group 1 (P trend <0.05). Only microfibril associated protein 4 (MFAP4) was less
 217 abundant in Group 1 than in uninfected controls or in Group 2 (P trend >0.05).

218

219 *Predictive value of 9 EV proteins for COVID-19 severity*

220 We performed ROC analysis for combined Group1 and Group 2 to provide a
 221 robust test of sensitivity, specificity, and AUC for our set of nine predictive markers
 222 (**Figure 1d**). The AUC values for the four markers with enhanced abundance in Group
 223 1, COPB2, KRAS, PRKCM, and RHOC, were 1.00 (95% CI: 1.00-1.00), 0.93 (95% CI:
 224 0.85-1.00), 0.93 (95% CI: 0.83-1.00), and 0.96 (95% CI: 0.89-1.00), respectively. AUC
 225 values for the other five markers, CD147, CAPN2, ECM1, FGG, and MFAP4, were

226 0.73 (95% CI: 0.48-0.98), 0.84 (95% CI: 0.67-1.00), 0.82 (95% CI: 0.60-1.00), 0.87
 227 (95% CI: 0.73-1.00), and 0.75 (95% CI: 0.52-0.97), respectively. Our analysis suggests
 228 that this set of markers, examined at the time of patient admission to the hospital,
 229 provides a significant degree of separation between patients that will develop mild
 230 versus severe COVID-19.

231 Additional ROC curves were generated to identify optimal cutoff values for
 232 the 9 marker proteins according to the Youden index (**Table S6**). For further analysis,
 233 we used the cutoff values to separate COVID-19 patients into low and high groups for
 234 comparing the onset of severe COVID-19 related events. Kaplan-Meier curves for the
 235 time-to-onset of a severe event were constructed for each of the 9 proteins (**Figure 1e**).
 236 Progression-free times for COPB2, KRAS, PRKCM, and RHOC, proteins with
 237 enhanced abundance in Group 1, were significantly shorter in the low group than in the
 238 high group ($P=9.8 \times 10^{-10}$; $P=1.0 \times 10^{-5}$; $P=5.1 \times 10^{-7}$; $P=4.2 \times 10^{-8}$, respectively).
 239 Conversely, the progression-free times for CD147, CAPN2, ECM1, FGG, and MFAP4,
 240 proteins with enhanced abundance in Group 2, were significant shorter in the high
 241 group than in the low group ($P=7.0 \times 10^{-5}$; $P=0.00060$; $P=2.2 \times 10^{-5}$; $P=1.9 \times 10^{-7}$;
 242 $P=2.0 \times 10^{-6}$, respectively). These results reinforce the concept that this set of markers

243 can be valuable for predicting the likelihood of patients experiencing severe COVID-19
244 related events.

245

246 *NGS determination of ExRNA profiles in serum samples from COVID-19 patients*
247 *and uninfected controls*

248 Circulating exRNAs have the potential to serve as biomarkers for a wide
249 range of diseases. ExRNAs consist of diverse RNA subpopulations that are protected
250 from degradation by incorporation into EVs or by association with lipids and/or proteins.
251 ExRNA profiles in blood samples are dynamic and include mRNAs, miRNAs, piRNAs,
252 and lncRNAs¹⁸. Here we have used next-generation sequencing (NGS) to analyze
253 exRNAs present in patient serum samples (**Figure 2a**). From our 41 serum samples,
254 NGS analysis identified 408 transcripts, excluding transcripts with fewer than 50 reads
255 in all samples. Of these exRNAs, 43 transcripts were differentially expressed between
256 the three patient groups ($P < 0.05$; one-way ANOVA). To identify patterns of expression
257 for these 43 transcripts among the three groups, we performed unsupervised
258 multivariate statistics based on PCA mapping. PCA plots from the NGS data revealed a
259 trend toward separation between the three groups (**Figure 2b**). This separation could be

described by the first PC1, which accounted for 28.1% of the variance. The second PC2 accounted for 14.5% of the variance. These observations indicate that the serum exRNA profiles in COVID-19 patients deviate considerably from those in uninfected donors. Moreover, despite some overlap or dispersity, the PCA plots allow us to detect obvious differences in exRNA profiles between Group 1 and Group 2.

For discrimination between Group 1 and Group 2, a cross-validation score was calculated for each of the selected transcripts on the basis of Fisher linear discriminant analysis, much as described for our analysis of EV proteomes. From the candidate transcripts, we chose 14 transcripts with cross-validation scores >0.75 (**Table S5**). **Figure 2c** compares the three patient groups for expression of the top 6 transcripts with cross-validation scores >0.80. These species include miR-122-5p, small nucleolar RNA C/D Box 33 (SNORD33), AL732437.2, RNA U2 small nuclear 29 Pseudogene (RNU2-29P), CDKN2B antisense RNA1 (CDKN2B-AS1), and AL365184.1 (this transcript has 5 different transcript IDs). Notably, the four transcripts SNORD33, AL732437.2, CDKN2B-AS1, and AL365184.1 exhibit significantly higher levels of expression in Group 2 than in uninfected controls or Group 1 (P trend <0.05).

277 *Predictive value of 6 exRNAs for COVID-19 severity*

278 Next, we constructed ROC curves for Group1 and Group 2 to provide a robust
279 test of sensitivity, specificity, and AUC values for our set of 6 predictive exRNA
280 markers (**Figure 2d**). AUC values for the transcripts miR-122-5p, SNORD33,
281 AL732437.2, RNU2-29P, CDKN2B-AS1, and AL365184.1 were 0.81 (95% CI:
282 0.64-0.99), 0.89 (95% CI: 0.78-1.00), 0.80 (95% CI: 0.60-1.00), 0.70 (95% CI:
283 0.49-1.92), 0.86 (95% CI: 0.71-1.00), and 0.90 (95% CI: 0.78-1.00), respectively. This
284 AUC analysis indicates that these exRNAs can provide a good basis for discriminating
285 between mild and severe COVID-19 cases at the time of patient admission.

286 Additional ROC curves were generated to identify optimal cutoff values for
287 the 6 transcripts according to the Youden index (**Table S6**). Using these cutoff values,
288 we separated COVID-19 patients into low and high groups for determining the
289 incidence of severe COVID-19 related events. Kaplan-Meier curves for time-to-onset of
290 a severe event after admission were analyzed for each of the 6 transcripts (**Figure 2e**).
291 The progression-free times for miR-122-5p, SNORD33, AL732437.2, RNU2-29P,
292 CDKN2B-AS1, and AL365184.1 were all significantly faster in the high group than in
293 the low group ($P=2.1\times 10^{-5}$; $P=0.00022$; $P=0.0067$; $P=3.8\times 10^{-5}$; $P=3.3\times 10^{-6}$; $P=0.0009$,

294 respectively). This finding suggests that these exRNA markers have value for predicting
295 the incidence of severe COVID-19 related events at the time of patient admission to the
296 hospital.

297

298 *The correlation between the selected markers for predicting disease severity values*

299 Next we used univariate Cox regression analysis to calculate the hazard ratio
300 (HR) for each of the EV and exRNA markers. Notably, the HR for COBP2 low was
301 statistically incalculable using the optimal cut-off value, suggesting that EV COPB2 has
302 the best predictive value among the two sets of markers. Furthermore, age high (HR
303 28.1; 95% CI 3.4-231.9; $P=0.0019$), CRP high (HR 8.4; 95% CI 1-67.5; $P=0.045$),
304 PRKCB low (HR 32.1; 95% CI 3.9-261.9; $P=0.0012$), RHOC low (HR 23.6; 95% CI
305 4.7-118; $P=0.00012$), CD147 high (HR 10.7; 95% CI 2.5-45.1; $P=0.0013$), CAPN2
306 high (HR 15.5; 95% CI 1.9-125.9; $P=0.010$), ECM1 high (HR 11.6; 95% CI 2.8-48.4;
307 $P=0.00079$), FGG high (HR 21.4; 95% CI 4.2-110.4; $P=0.00025$), MFAP4 high (HR
308 12.7; 95% CI 3.3-48.6; $P=0.00022$), miR-122-5p high (HR 10.5; 95% CI 2.7-40.4;
309 $P=0.00063$), AL732437.2 high (HR 9.9; 95% CI 1.2-79.9; $P=0.031$), RNU2-29P high
310 (HR 10.4; 95% CI 2.6-40.8; $P=0.00081$), CDKN2B-AS1 high (HR 14.4; 95% CI

311 3.4-61.3; $P=0.00031$), and AL365184.1 high (HR 14.2; 95% CI 1.8-114.4; $P=0.013$)

312 were statistically significant (**Table S6**).

313 To investigate potential relationships between the selected markers,

314 *Spearman's* correlation coefficients were calculated based on marker levels. A

315 correlation plot was constructed to visualize the correlation coefficients of the 19

316 markers (**Figure 3**). This allowed identification of four hierarchical clusters of markers

317 that share strong positive correlations within the groups to which they belong. Each

318 marker appears to fit into one of the four well defined clusters (namely, cluster 1, 2, 3,

319 and 4). Notably, we observe that cluster 1 (PRKCB, RHOC, COPB2, and KRAS) is

320 negatively correlated with the other clusters. Clusters 2, 3, and 4 have substantially

321 strong positive correlations with each other.

322 All 4 EV proteins in cluster 1 exhibit significantly higher abundance in Group

323 1 than in Group 2 COVID-19 patients. This is consistent with the idea that cluster 1

324 might represent a group of antiviral response-related EV proteins. Although the

325 functions of RHOC or KRAS during SARS-CoV-2 infection remain unknown, there are

326 some reports that implicate PRKCB or COPB2 in antiviral effects against SARS-CoV-2.

327 PRKCB can regulate metabolic and mitochondrial aspects of reprogramming

328 responsible for B cell fate ¹⁹. A recent bioinformatics-based report has revealed that
329 PRKCB is one of the target genes activated by vitamin A that may have antiviral
330 potential against SARS-CoV-2 ²⁰. COPB2 is a subunit of the Golgi Coatamer complex
331 (COPI) that is necessary for retrograde trafficking from the Golgi to the endoplasmic
332 reticulum (ER). Many viruses, including RNA viruses, DNA viruses, and retroviruses,
333 hijack or adapt COPI-related proteins for their own benefit. Indeed, the gene is required
334 for replication of SARS-CoV-1, which is closely genetically related (79% identity) to
335 SARS-CoV-2 ²¹. Depletion of COPB2 has a strong antiviral effect, based on reduced
336 SARS-CoV-1 protein expression ²². Although the precise functions of COPB2 positive
337 EVs still remain unknown, we hypothesize that secretion of these EVs reflects the
338 development of patient antiviral responses against SARS-CoV-2, potentially mitigating
339 the severity of COVID-19 progression.

340 Cluster 2 includes smoking, age, and MFPA4. MFAP4 is an extracellular
341 matrix protein belonging to the fibrinogen-related protein superfamily ²³. Levels of
342 MFAP4 are not significantly correlated with either smoking or age. On the other hand,
343 cluster 3 includes ECM1, CDKN2B.AS1, AL365184.1, CAPN2, CRP, FGG, and
344 CD147. Based on previous evidence, we considered that ECM1, CDKN2B.AS1,

345 CAPN2, and CD147 could be implicated in the abnormal coagulation status associated
346 with severe COVID-19. The extracellular matrix protein ECM1 is expressed in
347 association with blood vessels and may have pro-angiogenic functions ²⁴. The long
348 non-coding RNA CDKN2B.AS1 regulates production of extracellular matrix in the
349 pathogenesis associated with complications of diabetes ²⁵. CAPN2 is a
350 calcium-dependent intracellular protease that maintains the level of matrix
351 metalloproteinase 2, helping to regulate extracellular matrix accumulation ²⁶. CD147,
352 known as extracellular matrix metalloproteinase inducer (EMMPRIN), is a
353 transmembrane glycoprotein considered to be a binding partner for the SARS-CoV-2
354 spike protein, with obvious functional implications for viral infection ²⁷. Furthermore,
355 the role of CD147 in matrix metalloproteinase regulation is important for tumor
356 progression and for development of coronary artery disease, affecting both plaque
357 stability and thrombus formation ²⁸. The levels of 1 exRNA (CDKN2B.AS1 from
358 cluster 3) and 4 proteins relevant to extracellular matrix formation (MFPA4 from cluster
359 2 and ECM1, CAPN2, and CD147 from cluster 3) are correlated with the levels of FGG,
360 which has crucial functions in coagulation ($P<0.05$) (**Figure 3**). Reports of widespread
361 thromboses and disseminated intravascular coagulation (DIC) in COVID-19 patients

362 have been rapidly increasing in number. Recent reports show that D-dimer elevation at
363 the time of admission is predictive of bleeding, thrombosis, critical illness, and death in
364 COVID-19 patients ²⁹. Unfortunately, as our study includes only one COVID-19 patient
365 with complications from thrombosis, we are unable to make correlations between
366 D-dimer elevation and predicted disease severity. With regard to the relationship
367 between cluster 2 and 3, the levels of markers in cluster 3 are correlated with age, which
368 in turn is related to vascular endothelial dysfunction and coagulation ³⁰. Although the
369 function of non-coding RNA AL365184.1 is unknown, the bulk of our data suggest that
370 cluster 2 and 3 represent groups of coagulation-related markers.

371 The Cluster 4 components ALT, RNU2-29P, SNORD33, miR-122-5p, and
372 AL732437.2 may at least partly reflect phenomena related to liver damage. Levels of
373 ALT, a representative transaminase mainly associated with liver dysfunction, correlate
374 with levels of the three exRNA species ($P < 0.05$). miRNA-122-5p in particular is a
375 promising exploratory biomarker for detecting liver injury in preclinical and clinical
376 studies ³¹. It has been reported that the small nucleolar RNA SNORD33 is induced in
377 response to liver injury caused by lipopolysaccharide (LPS) ³². LPS-induced
378 inflammation has also been found to stimulate secretion of SNORD33 packaged into

EVs³³. Recent data indicate that deranged expression of liver enzymes is one manifestation of COVID-19 pathology, and that liver injury is more prevalent in severe cases than in mild cases of COVID-19^{34,35}. Although the functions of the non-coding RNAs RNU2-29P and AL732437.2 are unknown, our data at least partially support the idea that exRNAs associated with liver damage can serve as biomarkers for predicting the severity of COVID-19 in patients at the time of admission.

Discussion

As a single center study with a small sample size, our work has several limitations, and caution should be exercised in utilizing the predictive value of the markers we have identified. An expanded random sample across all genders and ages should be more representative of the general population, and larger clinical studies are required to validate the potential of these biomarkers for predicting the severity of COVID-19 progression. Nevertheless, our research has implications that go beyond the simple investigation of biomarkers, since the results also provide important clues regarding the pathogenesis of COVID-19 and the development of therapies for the disease. Indeed, biomarkers such as PRKCB, COPB2, and CD147 may be involved in

396 SARS-CoV-2 infection or replication. Furthermore, 6 markers (MFAP4, ECM1,
397 CDKN2B.AS1, CAPN2, FGG, and CD147) and 2 exRNAs (SNORD33 and
398 miR-122-5p) are involved in thrombosis or liver injury, which are serious complications
399 in patients with severe COVID-19. These findings indicate that the profiles of EV
400 proteins and exRNAs in patient sera clearly reflect specific host reactions to
401 SARS-CoV-2 infection and progression of the disease. Although the pathological
402 significance of some markers is unknown, understanding the profiles of functional
403 extracellular components in patient sera may help clarify various aspects of COVID-19
404 pathogenesis. With regard to therapy, biomarkers may provide information concerning
405 potential therapeutic targets for mitigating the effects of SARS-CoV-2 infection. For
406 instance, our finding that 4 EV proteins in cluster 1 are significantly more abundant in
407 Group 1 than in Group 2 suggests that these proteins could have antiviral effects.
408 Supplementation with these EVs could protect patients against COVID-19 progression
409 and the associated complications. In addition, a recent study showed that antibody
410 against the spike protein receptor CD147 could block infection by SARS-CoV-2 ²⁷.
411 Similarly, neutralization or manipulation of other markers with higher abundance in

412 Group 2 than in Group 1 might also have antiviral effects or provide protection against
413 severe consequences of COVID-19 progression.

414 In summary, our comprehensive evaluation identifies 3 distinct groups of
415 components (antiviral response-related EV proteins, coagulation-related markers, and
416 liver damage-related exRNAs) capable of serving as early predictive biomarkers for the
417 severity of COVID-19 progression. Among these markers, EV COPB2 has the best
418 predictive value for severe deterioration of COVID-19 patients in our cohort.
419 COVID-19 patients with high levels of EV COPB2 at the time of admission might be
420 able to overcome the disease without experiencing severe events. It might be
421 worthwhile to consider whether these markers have greater predictive power than
422 simpler factors like age and CRP levels. This study for the first time provides potential
423 resources for early discrimination between COVID-19 patients that may be resistant to
424 disease progression and patients that are likely to experience severe COVID-19 related
425 deterioration. Our results also suggest that, in addition to their predictive value,
426 functional extracellular components can also be potential contributors and mitigators of
427 pathogenesis during COVID-19 deterioration. These ideas should be validated and
428 expanded in future studies.

430 **Methods**

431 **Study approval**

432 This retrospective study involving collection of COVID-19 serum samples was
 433 approved by the Institutional Review Board at The Jikei University School of Medicine
 434 (Number: 32-055(10130)). The protocol did not require informed consent, and patients
 435 were given the choice of opting out of the study. For collection of healthy control serum
 436 samples, this study was approved by the Institutional Review Board at The Institute of
 437 Medical Science, The University of Tokyo (Number: 28-19-0907). Written informed
 438 consent was provided by all healthy donors before sample acquisition, in accordance
 439 with Declaration of Helsinki principles.

440

441 **Aknowledgement**

442 We thank Yusuke Hosaka and Akihiko Ito (The Jikei University School of Medicine,
 443 Tokyo, Japan), Dr. Yoshihiro Hirata (The Institute of Medical Science, The University
 444 of Tokyo), and Dr. Takashi Nakagawa (Omiya City Clinic) for clinical sample
 445 collection, Dr Misato Yamamoto for technical assistance, Dr. Tatsutoshi Inuzuka for
 446 technical support of LC-MS analysis (H.U. Group Research Institute), and all medical

447 staff of Team COVID at The Jikei University Hospital. This work was supported by

448 International Space Medical Co., Ltd.

449

450 **Author contributions**

451 Y.F. and T.O. conceived the idea and coordinated the project. Y.F., T.Hoshina., J.M.,

452 T.K., S.F., H.K., and N.W. performed statistical data analysis, and wrote the draft of the

453 manuscript. T.Hoshina., K.S., Y.S., M.Miyajima., K.L., K.N., T.Horino., R.N., and M.Y.

454 collected serum samples. M.Miyato., J.A., and K.K. provided helpful discussion. The

455 manuscript was finalized by Y.F. with the assistance of all authors.

456

457 **References**

- 458 1. Dong, E., Du, H. & Gardner, L. An interactive web-based dashboard to track
459 COVID-19 in real time. *Lancet Infect Dis* **20**, 533-534 (2020).
- 460 2. Guan, W.J., *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China.
461 *N Engl J Med* **382**, 1708-1720 (2020).
- 462 3. Huang, C., *et al.* Clinical features of patients infected with 2019 novel
463 coronavirus in Wuhan, China. *Lancet* **395**, 497-506 (2020).
- 464 4. Huang, D., *et al.* Clinical features of severe patients infected with 2019 novel
465 coronavirus: a systematic review and meta-analysis. *Ann Transl Med* **8**, 576
466 (2020).
- 467 5. Mangalmurti, N. & Hunter, C.A. Cytokine Storms: Understanding COVID-19.
468 *Immunity* **53**, 19-25 (2020).
- 469 6. Chen, N., *et al.* Epidemiological and clinical characteristics of 99 cases of 2019
470 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* **395**,
471 507-513 (2020).
- 472 7. Yanez-Mo, M., *et al.* Biological properties of extracellular vesicles and their
473 physiological functions. *J Extracell Vesicles* **4**, 27066 (2015).

- 474 8. Kowal, J., *et al.* Proteomic comparison defines novel markers to characterize
475 heterogeneous populations of extracellular vesicle subtypes. *Proc Natl Acad Sci*
476 *U S A* **113**, E968-977 (2016).
- 477 9. Robbins, P.D. & Morelli, A.E. Regulation of immune responses by extracellular
478 vesicles. *Nat Rev Immunol* **14**, 195-208 (2014).
- 479 10. Pegtel, D.M., *et al.* Functional delivery of viral miRNAs via exosomes. *Proc*
480 *Natl Acad Sci U S A* **107**, 6328-6333 (2010).
- 481 11. Earnest, J.T., *et al.* The tetraspanin CD9 facilitates MERS-coronavirus entry by
482 scaffolding host cell receptors and proteases. *PLoS Pathog* **13**, e1006546 (2017).
- 483 12. Wang, J., Chen, S. & Bihl, J. Exosome-Mediated Transfer of ACE2
484 (Angiotensin-Converting Enzyme 2) from Endothelial Progenitor Cells
485 Promotes Survival and Function of Endothelial Cell. *Oxid Med Cell Longev*
486 **2020**, 4213541 (2020).
- 487 13. Das, S., *et al.* The Extracellular RNA Communication Consortium: Establishing
488 Foundational Knowledge and Technologies for Extracellular RNA Research.
489 *Cell* **177**, 231-242 (2019).

- 490 14. tenOever, B.R. RNA viruses and the host microRNA machinery. *Nat Rev*
491 *Microbiol* **11**, 169-180 (2013).
- 492 15. Girardi, E., Lopez, P. & Pfeffer, S. On the Importance of Host MicroRNAs
493 During Viral Infection. *Front Genet* **9**, 439 (2018).
- 494 16. Wang, Z., Yang, B., Li, Q., Wen, L. & Zhang, R. Clinical Features of 69 Cases
495 With Coronavirus Disease 2019 in Wuhan, China. *Clin Infect Dis* **71**, 769-777
496 (2020).
- 497 17. Urabe, F., *et al.* Large-scale Circulating microRNA Profiling for the Liquid
498 Biopsy of Prostate Cancer. *Clin Cancer Res* **25**, 3016-3025 (2019).
- 499 18. Murillo, O.D., *et al.* exRNA Atlas Analysis Reveals Distinct Extracellular RNA
500 Cargo Types and Their Carriers Present across Human Biofluids. *Cell* **177**,
501 463-477 e415 (2019).
- 502 19. Tsui, C., *et al.* Protein Kinase C-beta Dictates B Cell Fate by Regulating
503 Mitochondrial Remodeling, Metabolic Reprogramming, and Heme Biosynthesis.
504 *Immunity* **48**, 1144-1159 e1145 (2018).
- 505 20. Li, R., *et al.* Revealing the targets and mechanisms of vitamin A in the treatment
506 of COVID-19. *Aging (Albany NY)* **12**, 15784-15796 (2020).

- 507 21. Lu, R., *et al.* Genomic characterisation and epidemiology of 2019 novel
508 coronavirus: implications for virus origins and receptor binding. *Lancet* **395**,
509 565-574 (2020).
- 510 22. de Wilde, A.H., *et al.* A Kinome-Wide Small Interfering RNA Screen Identifies
511 Proviral and Antiviral Host Factors in Severe Acute Respiratory Syndrome
512 Coronavirus Replication, Including Double-Stranded RNA-Activated Protein
513 Kinase and Early Secretory Pathway Proteins. *J Virol* **89**, 8318-8333 (2015).
- 514 23. Wulf-Johansson, H., *et al.* Localization of microfibrillar-associated protein 4
515 (MFAP4) in human tissues: clinical evaluation of serum MFAP4 and its
516 association with various cardiovascular conditions. *PLoS One* **8**, e82243 (2013).
- 517 24. Steinhäuser, S.S., *et al.* ECM1 secreted by HER2-overexpressing breast cancer
518 cells promotes formation of a vascular niche accelerating cancer cell migration
519 and invasion. *Lab Invest* **100**, 928-944 (2020).
- 520 25. Thomas, A.A., Feng, B. & Chakrabarti, S. ANRIL regulates production of
521 extracellular matrix proteins and vasoactive factors in diabetic complications.
522 *Am J Physiol Endocrinol Metab* **314**, E191-E200 (2018).

- 523 26. Jang, H.S., Lal, S. & Greenwood, J.A. Calpain 2 is required for glioblastoma
524 cell invasion: regulation of matrix metalloproteinase 2. *Neurochem Res* **35**,
525 1796-1804 (2010).
- 526 27. Aguiar, J.A., *et al.* Gene expression and in situ protein profiling of candidate
527 SARS-CoV-2 receptors in human airway epithelial cells and lung tissue. *Eur*
528 *Respir J* **56**(2020).
- 529 28. Joghetaei, N., *et al.* The Extracellular Matrix Metalloproteinase Inducer
530 (EMMPRIN, CD147) - a potential novel target in atherothrombosis prevention?
531 *Thromb Res* **131**, 474-480 (2013).
- 532 29. Al-Samkari, H., *et al.* COVID-19 and coagulation: bleeding and thrombotic
533 manifestations of SARS-CoV-2 infection. *Blood* **136**, 489-500 (2020).
- 534 30. Donato, A.J., Machin, D.R. & Lesniewski, L.A. Mechanisms of Dysfunction in
535 the Aging Vasculature and Role in Age-Related Disease. *Circ Res* **123**, 825-848
536 (2018).
- 537 31. Bala, S., *et al.* Circulating microRNAs in exosomes indicate hepatocyte injury
538 and inflammation in alcoholic, drug-induced, and inflammatory liver diseases.
539 *Hepatology* **56**, 1946-1957 (2012).

- 540 32. Michel, C.I., *et al.* Small nucleolar RNAs U32a, U33, and U35a are critical
541 mediators of metabolic stress. *Cell Metab* **14**, 33-44 (2011).
- 542 33. Rimer, J.M., *et al.* Long-range function of secreted small nucleolar RNAs that
543 direct 2'-O-methylation. *J Biol Chem* **293**, 13284-13296 (2018).
- 544 34. Zhang, C., Shi, L. & Wang, F.S. Liver injury in COVID-19: management and
545 challenges. *Lancet Gastroenterol Hepatol* **5**, 428-430 (2020).
- 546 35. Hajifathalian, K., *et al.* Gastrointestinal and Hepatic Manifestations of 2019
547 Novel Coronavirus Disease in a Large Cohort of Infected Patients From New
548 York: Clinical Implications. *Gastroenterology* **159**, 1137-1140 e1132 (2020).
- 549

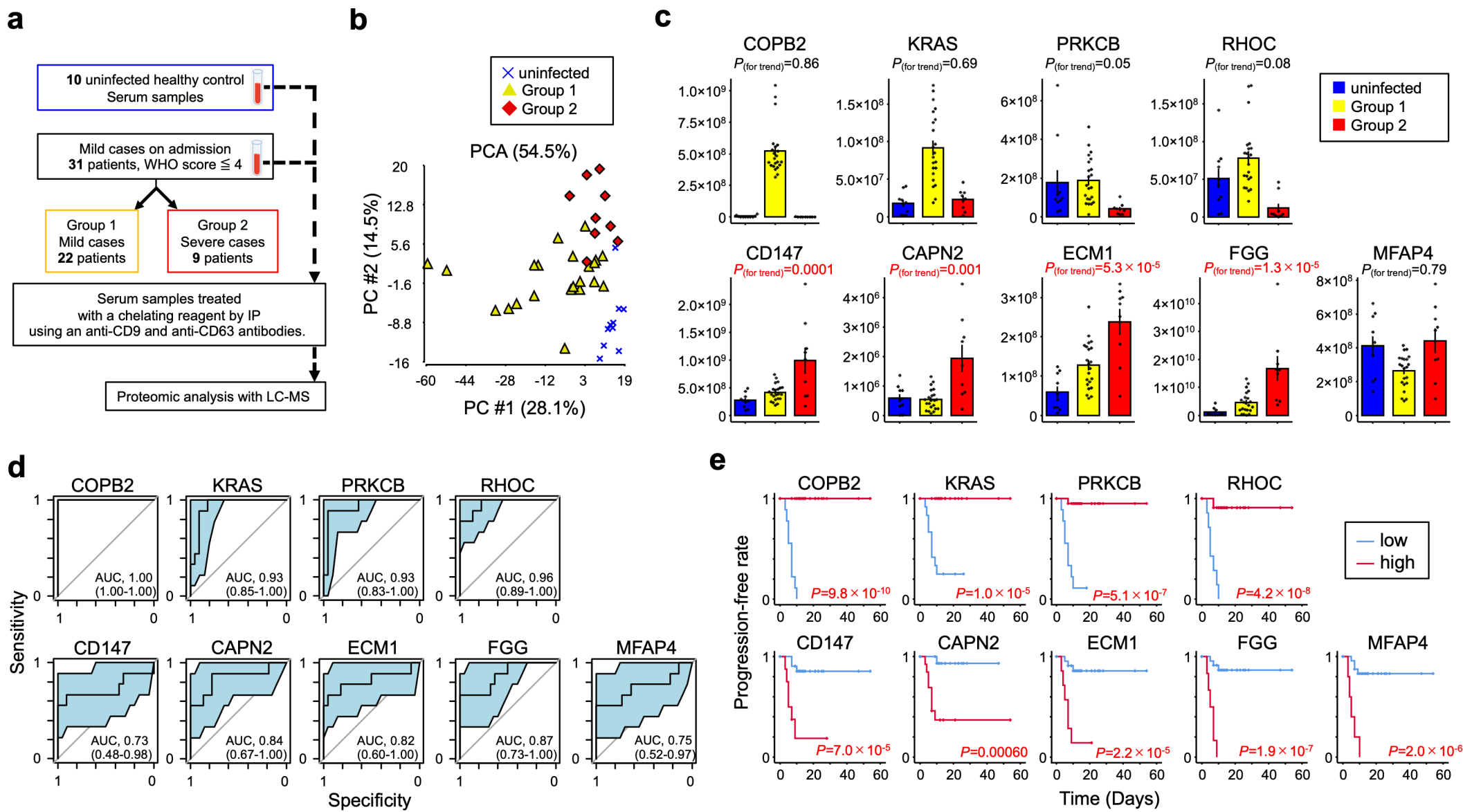


Figure 1. EV proteomes for early prediction of COVID-19 severity.

(a) Work flow for LC-MS identification of proteomes from CD9⁺/CD63⁺ EVs obtained from serum samples of 31 mild COVID-19 subjects (Group 1: $n=22$, Group2: $n=9$) and 10 uninfected healthy controls. (b) PCA map of 723 proteins from the three subject groups. (c) Correlations of COPB2, KRAS, PRKCM, RHOC, CD147, CAPN2, ECM1, FGG, and MFAP4 between the three subject groups. P values for trend by *Pearson's* correlation analysis. Error bars represent mean \pm SEM. (d) AUC values (95% CI) for 9 EV proteins evaluated by ROC analysis. (e) Kaplan-Meier curves for 9 EV proteins by Log-rank test. Time represents the number of days from admission to time of onset for a severe COVID-19 related event. Optimal cut-off values were used to define high and low groups.

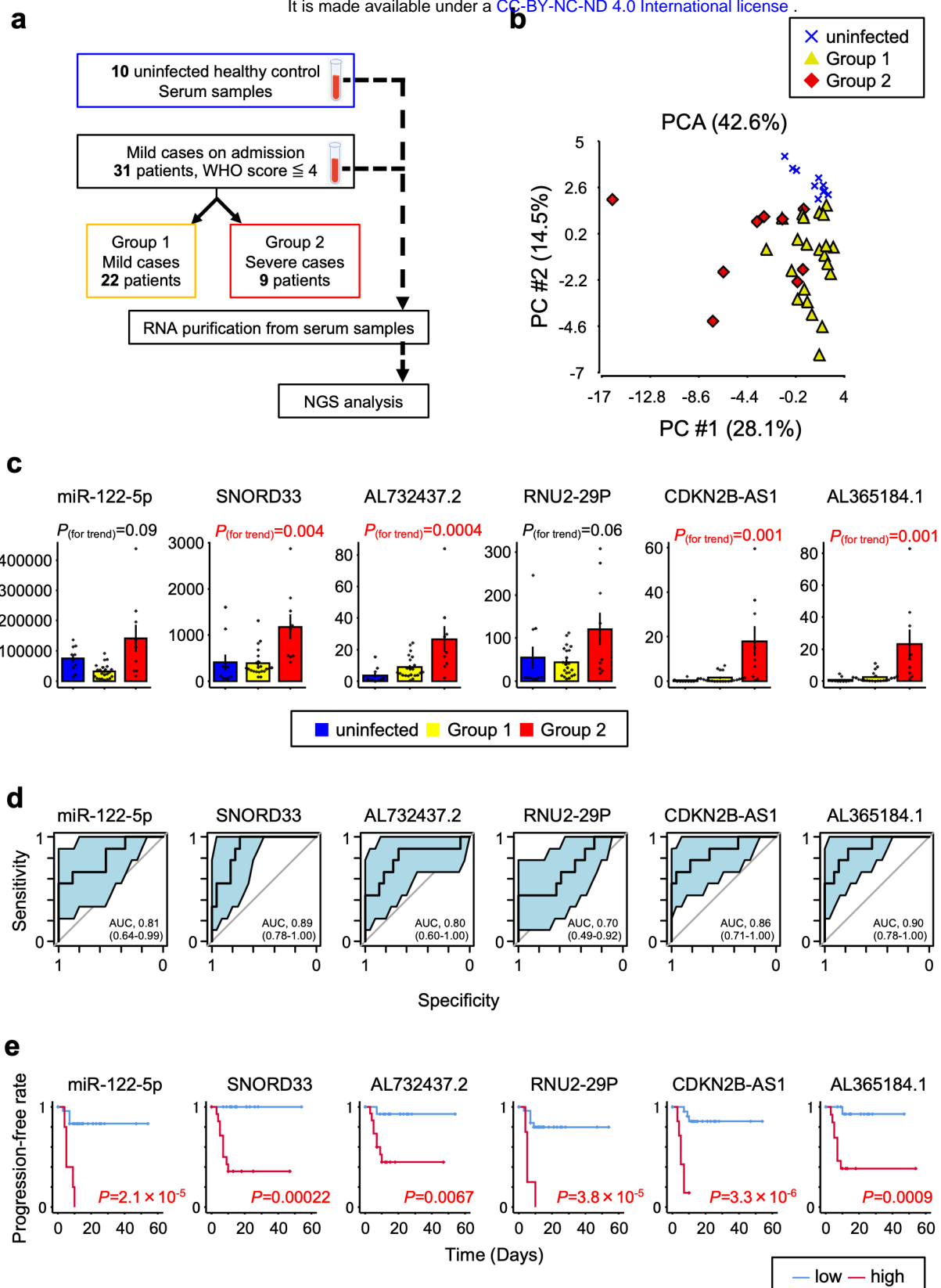


Figure 2. ExRNA profiles for early prediction of COVID-19 severity.

(a) Work flow for NGS determination of exRNA profiles from serum samples of 31 mild COVID-19 patients (Group 1: $n=22$, Group2: $n=9$) and 10 uninfected healthy controls. (b) PCA map of 43 transcripts for the three subject groups. (c) Correlations of miR-122-5p, SNORD33, AL732437.2, RNU2-29P, CDKN2B-AS1, and AL365184.1 between the three subject groups. P values for trend by *Pearson's* correlation analysis. Error bars represent mean \pm SEM. (d) AUC values (95% CI) for 6 transcripts evaluated by ROC analysis. (e) Kaplan-Meier curves for 6 transcripts by Log-rank test. Time represents the number of days from admission to time of onset for a severe COVID-19 related event. Optimal cut-off values were used to define high and low groups.

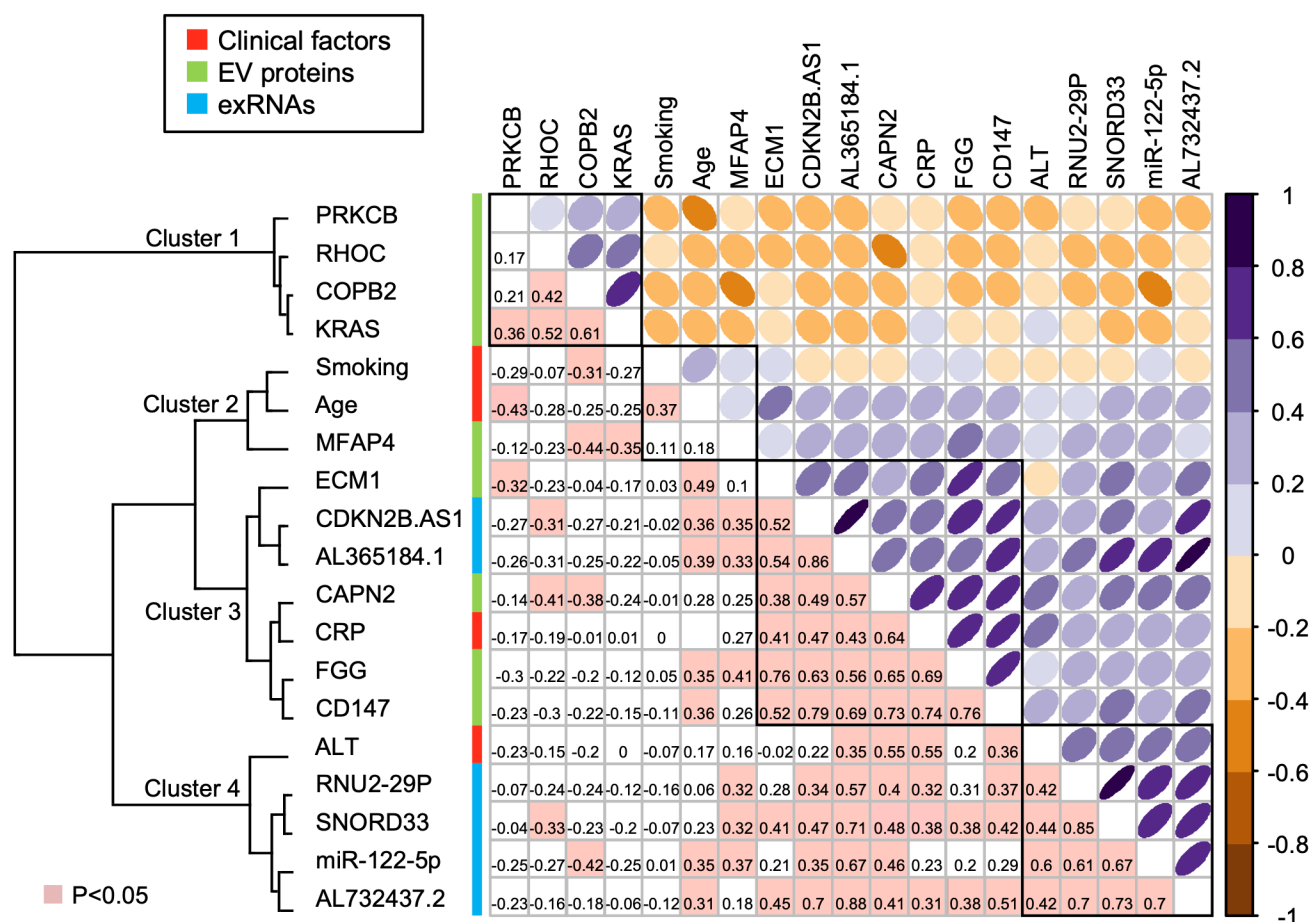


Figure 3. Upper triangular correlation plot of the associations between 4 clinical factors, 9 EV proteins, and 6 transcripts. Colors represent Pearson's correlation coefficients. Positive correlations are represented in purple, while negative correlations are represented in brown in the upper triangle. Color intensity and ovalization of the circle are proportional to the correlation coefficients. The lower triangular correlation matrix displays actual correlation values, with pink highlights representing $P < 0.05$. Cluster 1 (PRKCB, RHOC, COPB2, and KRAS) contains a group of antiviral response-related EV proteins. Clusters 2 (smoking, age, and MFAP4) and 3 (CM1, CDKN2B.AS1, AL365184.1, CAPN2, CRP, FG, and CD147) contain groups of coagulation-related markers. Cluster 4 (ALT, RNU2-29P, SNORD33, miR-122-5p, and AL732437.2) contains a group of liver damage-related exRNAs.