

## CORONAVIRUS

# Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort

Raphael Carapito<sup>1,2,3†</sup>, Richard Li<sup>4†</sup>, Julie Helms<sup>1,3,5</sup>, Christine Carapito<sup>3,6</sup>, Sharvari Gujja<sup>4</sup>, Véronique Rolli<sup>1,2,3</sup>, Raony Guimaraes<sup>4</sup>, Jose Malagon-Lopez<sup>4</sup>, Perrine Spinnhirny<sup>1,3</sup>, Alexandre Lederle<sup>1,3</sup>, Razieh Mohseninia<sup>7</sup>, Aurélie Hirschler<sup>3,6</sup>, Leslie Muller<sup>3,6</sup>, Paul Bastard<sup>8,9,10</sup>, Adrian Gervais<sup>9,10</sup>, Qian Zhang<sup>8,9,10</sup>, François Danion<sup>1,3,11</sup>, Yvon Ruch<sup>3,11</sup>, Maleka Schenck<sup>3,12</sup>, Olivier Collange<sup>3,13</sup>, Thiên-Nga Chamaraux-Tran<sup>3,14</sup>, Anne Molitor<sup>1,3</sup>, Angélique Pichot<sup>1,3</sup>, Alice Bernard<sup>1,3</sup>, Ouria Tahar<sup>2,3</sup>, Sabrina Bibi-Triki<sup>1,3</sup>, Haiguo Wu<sup>4</sup>, Nicodème Paul<sup>1,3</sup>, Sylvain Mayeur<sup>1,3</sup>, Annabel Larnicol<sup>1,3</sup>, Géraldine Laumond<sup>1,3</sup>, Julia Frappier<sup>1,3</sup>, Sylvie Schmidt<sup>1,3</sup>, Antoine Hanauer<sup>1,3</sup>, Cécile Macquin<sup>1,3</sup>, Tristan Stemmelen<sup>1,2,3</sup>, Michael Simons<sup>15</sup>, Xavier Mariette<sup>16,17</sup>, Olivier Hermine<sup>10,18</sup>, Samira Fafi-Kremer<sup>1,3,19</sup>, Bernard Goichot<sup>3,20</sup>, Bernard Drenou<sup>21</sup>, Khaldoun Kuteifan<sup>22</sup>, Julien Pottecher<sup>3,14</sup>, Paul-Michel Mertes<sup>3,13</sup>, Shweta Kailasan<sup>23</sup>, M. Javad Aman<sup>23</sup>, Elisa Pin<sup>24</sup>, Peter Nilsson<sup>24</sup>, Anne Thomas<sup>25</sup>, Alain Viari<sup>25</sup>, Damien Sanlaville<sup>25</sup>, Francis Schneider<sup>3,12</sup>, Jean Sibilia<sup>1,3,26</sup>, Pierre-Louis Tharaux<sup>27</sup>, Jean-Laurent Casanova<sup>8,9,10,28</sup>, Yves Hansmann<sup>3,11</sup>, Daniel Lidar<sup>7,29</sup>, Mirjana Radosavljevic<sup>1,2,3</sup>, Jeffrey R. Gulcher<sup>4</sup>, Ferhat Meziani<sup>3,5</sup>, Christiane Moog<sup>1,3</sup>, Thomas W. Chittenden<sup>4,30\*</sup>, Seiamak Bahram<sup>1,2,3,\*</sup>

The drivers of critical coronavirus disease 2019 (COVID-19) remain unknown. Given major confounding factors such as age and comorbidities, true mediators of this condition have remained elusive. We used a multi-omics analysis combined with artificial intelligence in a young patient cohort where major comorbidities were excluded at the onset. The cohort included 47 “critical” (in the intensive care unit under mechanical ventilation) and 25 “non-critical” (in a non-critical care ward) patients with COVID-19 and 22 healthy individuals. The analyses included whole-genome sequencing, whole-blood RNA sequencing, plasma and blood mononuclear cell proteomics, cytokine profiling, and high-throughput immunophenotyping. An ensemble of machine learning, deep learning, quantum annealing, and structural causal modeling were used. Patients with critical COVID-19 were characterized by exacerbated inflammation, perturbed lymphoid and myeloid compartments, increased coagulation, and viral cell biology. Among differentially expressed genes, we observed up-regulation of the metallo-protease ADAM9. This gene signature was validated in a second independent cohort of 81 critical and 73 recovered patients with COVID-19 and was further confirmed at the transcriptional and protein level and by proteolytic activity. Ex vivo ADAM9 inhibition decreased severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) uptake and replication in human lung epithelial cells. In conclusion, within a young, otherwise healthy, cohort of individuals with COVID-19, we provide the landscape of biological perturbations *in vivo* where a unique gene signature differentiated critical from non-critical patients. We further identified ADAM9 as a driver of disease severity and a candidate therapeutic target.

## INTRODUCTION

Unlike many viral infections and most respiratory virus infections, coronavirus disease 2019 (COVID-19) is characterized by a complex and diversified spectrum of clinical manifestations (1). Upon infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), age-, sex-, and phenotype-matched individuals can be classified within four distinct groups: (i) asymptomatic individuals, (ii) patients displaying influenza-like illnesses, (iii) patients affected by respiratory dysfunction who eventually need an external oxygen supply, and (iv) patients suffering from acute respiratory distress syndrome (ARDS) who need invasive mechanical ventilation in an intensive care unit (ICU). Although the last group represents only a small fraction of patients with COVID-19, this group encompasses the most critical form of the disease and has an average case-fatality rate of approximately 25% (2). Despite intense investigation, the fundamental question of why the course of the disease shows such a

marked difference in an otherwise, apparently indistinguishable set of individuals remains unanswered (3–6). To better understand this issue, high-resolution molecular analyses should be applied to well-defined cohorts of patients and controls where a maximum of confounding factors have been eliminated. These factors include older age as well as a number of comorbidities, such as cerebrovascular disease, types 1 and 2 diabetes, chronic kidney disease, chronic obstructive pulmonary disease, or heart conditions (7).

Several studies have used single, or a restricted number of, omics technologies to uncover molecular processes associated with disease severity, usually in unfiltered critical patients with COVID-19. Systemic inflammation with high concentrations of acute-phase proteins [C-reactive protein (CRP), serum amyloid A (SAA), and calprotectin] (8) and inflammatory cytokines, particularly interleukin-6 (IL-6) and IL-1 $\beta$  (9–11), has been found to be a hallmark of disease severity. In contrast, after an initial burst shortly after

Copyright © 2022  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim  
to original U.S.

Government Works.  
Distributed under a  
Creative Commons  
Attribution License 4.0  
(CC BY).

infection, the type I interferon (IFN) response is impaired at the RNA (12) and protein (13) level. Severity was also correlated with profound immune dysregulation, including modifications in the myeloid compartment with increases in neutrophils (14, 15), decreases in nonclassical monocytes (8), and dysregulation of macrophages (10, 16). The lymphoid compartment is also modified by both B cell activation (17) and an impaired T cell response, characterized by skewing toward a T helper 17 ( $T_{H}17$ ) phenotype (18, 19). Moreover, coagulation defects have been identified in critically ill patients who are prone to thrombotic complications (20–22). Nevertheless, the full spectrum of omics technologies has not been applied to a highly curated cohort of patients with COVID-19 and controls that was established by eliminating a number of key confounding factors that affect severity and death, such as older age and comorbidities, at onset.

In this cross-sectional study, we aimed to analyze the SARS-CoV-2-induced molecular changes that are characteristic of critical patients and differentiate them from non-critical patients. We hypothesized that certain host driver genes might be responsible for the development of critical illness and that those genes might represent therapeutic targets. To test these hypotheses, we performed an ensemble artificial intelligence (AI) and machine learning (ML)-based multi-omics study of 47 young (under 50 years of age) patients with COVID-19 without comorbidities admitted to the ICU and under mechanical ventilation (“critical” patients) versus matched patients with COVID-19 needing only hospitalization in a non-critical care ward (25 “non-critical” patients) and an age- and sex-matched control group of 22 healthy individuals not infected with SARS-CoV-2 (“healthy”). The multi-omics approach included whole-genome sequencing (WGS), whole-blood RNA sequencing (RNA-seq), quantitative plasma and peripheral blood mononuclear cell (PBMC) proteomics, multiplex plasma cytokine profiling, and high-throughput immune cell phenotyping. These analyses were complemented by the status of anti-SARS-CoV-2 neutralizing antibodies and multitarget immunoglobulin G (IgG) serology as well as the measurement of neutralizing anti-type I IFN autoantibodies in the entire cohort.

## RESULTS

### A young, comorbidity-free patient cohort was analyzed by a multiomics approach

The present study focused on patients who were hospitalized for COVID-19 at a university hospital network in northeast France (Alsace) during the first French wave of the pandemic (March to April 2020) before the routine use of corticosteroids. A total of 72 patients under 50 years of age without comorbidities were enrolled. Fifty-three of these patients were men (74%), and the median age of the patients was 40 [interquartile range (IQR) 33; 46] years. The patients were divided into two groups: (i) a critical group consisting of 47 (65%) patients hospitalized in the ICU due to moderate or severe ARDS according to the Berlin criteria (23) with 45 requiring invasive mechanical ventilation and 2 requiring high-flow nasal oxygen and noninvasive mechanical ventilation due to acute respiratory failure and (ii) a non-critical group consisting of 25 patients (35%) who stayed at a non-critical care ward. In the latter group, 19 (76%) needed low-flow supplemental oxygen. Patients who were transferred from the non-critical care ward to the ICU ( $n = 19$ ) were considered critical patients, and for these, the sampling was done upon ICU admission in the same conditions as patients directly admitted to the ICU. The median simplified acute physiology score (SAPS) II of the patients at the ICU was 38 [IQR 33; 47] points, and the median pressure of arterial oxygen to fractional inspired oxygen concentration ( $\text{PaO}_2/\text{FiO}_2$ ) ratio of these patients was 123 [IQR 95; 168] mmHg upon admission. All the patients were discharged from the hospital or were deceased at the time of data analysis. The overall hospital and day 28 mortality rate was 8.3% (six patients, all in the critical group, for a mortality of 13% in this group). The characteristics of the patients in both groups are summarized in Table 1.

On the basis of these two patient groups and an additional group of 22 healthy (SARS-CoV-2-negative) sex- and age-matched controls, we applied a global multi-omics analysis strategy to identify pathways and drivers of ARDS (Fig. 1). PBMCs were analyzed by mass cytometry (CyTOF) and shotgun proteomics. Plasma samples were used for multiplex cytokine quantification and shotgun proteomics. Serum samples were used for multiplex IgG serology (24) and detection

<sup>1</sup>Laboratoire d'ImmunoRhumatologie Moléculaire, plateforme GENOMAX, INSERM (Institut de la Santé et de la Recherche Médicale) UMR\_S 1109, Faculté de Médecine, Institut Thématisé Interdisciplinaire (ITI) de Médecine de Précision de Strasbourg, Transplantex NG, Université de Strasbourg, 67085 Strasbourg, France. <sup>2</sup>Service d'Immunologie Biologique, Plateau Technique de Biologie, Pôle de Biologie, Nouvel Hôpital Civil, 67091 Strasbourg, France. <sup>3</sup>Fédération Hospitalo-Universitaire (FHU) OMICARE, Fédération de Médecine Translationnelle de Strasbourg (FMTS), Centre de Recherche d'Immunologie et d'Hématologie, 67085, Strasbourg, France. <sup>4</sup>Genuity AI Research Institute, Genuity Science, Boston, MA 02114, USA. <sup>5</sup>Service de Médecine Intensive-Réanimation, Nouvel Hôpital Civil, Hôpitaux Universitaires de Strasbourg, 67091 Strasbourg, France. <sup>6</sup>Laboratoire de Spectrométrie de Masse BioOrganique, Université de Strasbourg, CNRS, IPHC, UMR 7178, 67000 Strasbourg, France. <sup>7</sup>Center for Quantum Information Science and Technology, University of Southern California, Los Angeles, CA 90089, USA. <sup>8</sup>St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, NY 10065, USA. <sup>9</sup>Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM, Necker Hospital for Sick Children, 75015 Paris, France. <sup>10</sup>University of Paris, Imagine Institute, 75015 Paris, France. <sup>11</sup>Department of Infectious and Tropical Diseases, Hôpitaux Universitaires de Strasbourg, 67091 Strasbourg, France. <sup>12</sup>Service de Médecine Intensive-Réanimation, Hôpital de Hautepierre, Hôpitaux Universitaires de Strasbourg, Avenue Molière, 67200 Strasbourg, France. <sup>13</sup>Service d'Anesthésie-Réanimation et Médecine Péri-Opératoire, Nouvel Hôpital Civil, Hôpitaux Universitaires de Strasbourg, 67000 Strasbourg, France. <sup>14</sup>Service d'Anesthésie-Réanimation et Médecine Péri-Opératoire, Hôpital de Hautepierre, Hôpitaux Universitaires de Strasbourg; 67200 Strasbourg Cedex, France. <sup>15</sup>Yale Cardiovascular Research Center, Departments of Medicine and Cell Biology, Yale University School of Medicine, New Haven, CT 06511, USA. <sup>16</sup>Department of Rheumatology, Hôpital Bicêtre, Assistance Publique-Hôpitaux de Paris, 94270 Paris, France. <sup>17</sup>Université Paris-Saclay, INSERM UMR\_S 1184; 94270 Le Kremlin Bicêtre, France. <sup>18</sup>Department of Hematology, INSERM UMR\_S 1153, Imagine Institute, Necker Hospital, University of Paris, Assistance Publique-Hôpitaux de Paris, 75015 Paris, France. <sup>19</sup>Department of Virology, Hôpitaux Universitaires de Strasbourg, 67091 Strasbourg, France. <sup>20</sup>Service de Médecine Interne, Endocrinologie et Nutrition, Hôpital de Hautepierre, Hôpitaux Universitaires de Strasbourg, 67200 Strasbourg, France. <sup>21</sup>Département d'Hématologie, Groupe Hospitalier de la région Mulhouse Sud Alsace, 68100 Mulhouse, France. <sup>22</sup>Service de Réanimation Médicale, Groupe Hospitalier de la région Mulhouse Sud Alsace, 68100 Mulhouse, France. <sup>23</sup>Integrated BioTherapeutics Inc., Rockville, MD 20850, USA. <sup>24</sup>Division of Affinity Proteomics, Department of Protein Science, KTH Royal Institute of Technology, SciLifeLab, Stockholm SE-171 21, Sweden. <sup>25</sup>Plateforme Auragen, 69003 Lyon, France. <sup>26</sup>Service de Rhumatologie, Centre National de Référence des Maladies Auto-immunes Systémiques Rares Est Sud-Ouest, Hôpitaux Universitaires de Strasbourg, 67200 Strasbourg, France. <sup>27</sup>INSERM, Université de Paris, Paris Cardiovascular Center-PARCC, 75015 Paris, France. <sup>28</sup>Howard Hughes Medical Institute, New York, NY 10065, USA. <sup>29</sup>Departments of Electrical and Computer Engineering, Chemistry, and Physics and Astronomy, University of Southern California, Los Angeles, CA 90089, USA. <sup>30</sup>Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston MA 02115, USA.

\*Corresponding author. Email: tom.chittenden@genuinitysci.com (T.W.C.); siamak@unistra.fr (S.B.)

†These authors contributed equally to this work.

**Table 1. Patient descriptions.** BMI, body mass index; IL-6R, interleukin-6 receptor; IQR, interquartile range; ARDS, acute respiratory distress syndrome; ECMO, extracorporeal membrane oxygenation; NMBA, neuromuscular blocking agent; RRT, renal replacement therapy; SAPS II, simplified acute physiology score II; SOFA, Sequential Organ Failure Assessment.

**Characteristics of all patients admitted to the hospital for COVID-19**

	All patients (n = 72)	Non-critical group (n = 25)	Critical group (n = 47)	P
<b>Age (median, IQR)</b>	40 [33; 46]	38 [31; 45]	41 [34; 46]	0.2
<b>Male [n (%)]</b>	53 (73.6)	17 (68.0)	36 (76.6)	0.6
<b>BMI (kg/m<sup>2</sup>) (median, IQR)</b>	30.0 [26.8; 35.0]	29.7 [23.8; 33.0]	30.2 [27.1; 35.6]	0.5
<b>Time since first symptoms (days) (median, IQR)</b>	8.0 [6.0; 11.0]	9.5 [7.2; 13.5]	7.0 [6.0; 10.0]	0.1
<b>COVID-19 treatments (during hospital stay) [n (%)]</b>				
Lopinavir/Ritonavir	21 (29.1)	3 (12.0)	18 (38.3)	0
Remdesivir	3 (4.1)	1 (4.0)	2 (4.2)	1
Hydroxychloroquine	19 (26.4)	2 (8.0)	17 (36.2)	0
Corticosteroids	6 (8.3)	1 (4.0)	6 (12.8)	0.3
<b>Neurological symptoms [n (%)]</b>	26 (50.0)	10/25 (40.0)	16/27 (59.2)	0.3
<b>Outcome [n (%)]</b>				
In-hospital and day 28 mortality	6 (8.3)	0	6 (12.8)	0.1

**Characteristics of ICU patients**

	Critical group (n = 47)
<b>Baseline severity scores</b>	
SAPS II (median, IQR)	38 [33; 47]
SOFA (median, IQR)	6 [4; 9]
<b>ARDS [n (%)]</b>	45 (95.7)
Moderate	21 (46.7)
Severe	24 (53.3)
<b>Supportive treatments</b>	
Invasive mechanical ventilation [n (%)]	45 (95.7)
Duration of invasive mechanical ventilation (days) (median, IQR)	13 [7; 24]
NMBA [n (%)]	40 (89.0)
Catecholamines [n (%)]	41 (91.1)
Catecholamines (days) (median, IQR)	4 [2; 10]
RRT [n (%)]	7 (15.6)
ECMO [n (%)]	2 (4.4)

of anti-SARS-CoV-2 neutralizing antibodies and anti-type I IFN neutralizing autoantibodies. Last, RNA-seq and WGS were performed using whole-blood samples. Unless otherwise specified, all measures were obtained from samples that were collected at the time of hospital admission (whether at the ICU or the non-critical care ward). Validation of the identified driver genes was performed using an ex vivo model of SARS-CoV-2 infection. The top 600 genes found by classification of patient cohort 1 were evaluated in a second, independent cohort of 81 critical patients and 73 recovered critical patients (table S1).

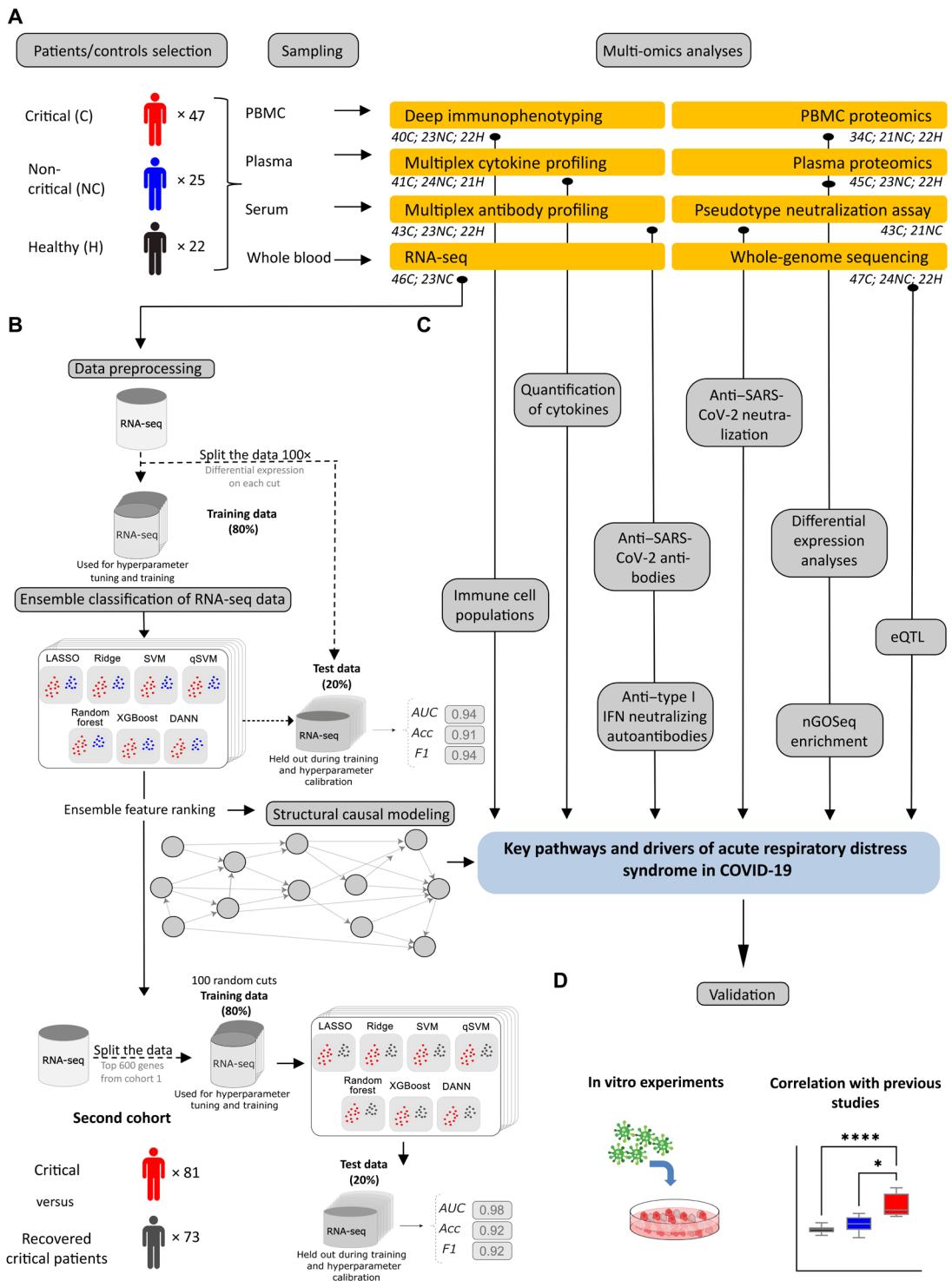
**Critical illness is characterized by a proinflammatory cytokine storm and changes in the T, B, dendritic, and monocyte cell compartments and is independent of the extent of viral infection**

The global proinflammatory cytokine profile showed significantly increased concentrations of IFN- $\gamma$  ( $P = 0.034$ ), tumor necrosis

factor- $\alpha$  (TNF- $\alpha$ ) ( $P = 0.022$ ), IL-1 $\beta$  ( $P = 0.0002$ ), IL-4 ( $P = 0.036$ ), IL-6 ( $P < 0.0001$ ), IL-8 ( $P = 0.0004$ ), IL-10 ( $P = 0.0002$ ), and IL-12p70 ( $P = 0.0221$ ) in critical versus non-critical patients (Fig. 2A). This “cytokine storm” (25) was more pronounced in critical patients, as only IFN- $\gamma$ , TNF- $\alpha$ , and IL-10 were higher in non-critical patients as compared to healthy controls. Although the disease severity was initially associated with an RNA-seq-based type I IFN signature, the absence of an increase in the plasma concentration of IFN- $\alpha$  in critical versus non-critical patients, the decrease in the IFN- $\alpha$  concentration during the ICU stay, and the reduction in the number of plasmacytoid dendritic cells, which are the main source of IFN- $\alpha$ , suggest that the IFN response is indeed impaired in critical patients (fig. S1) (12).

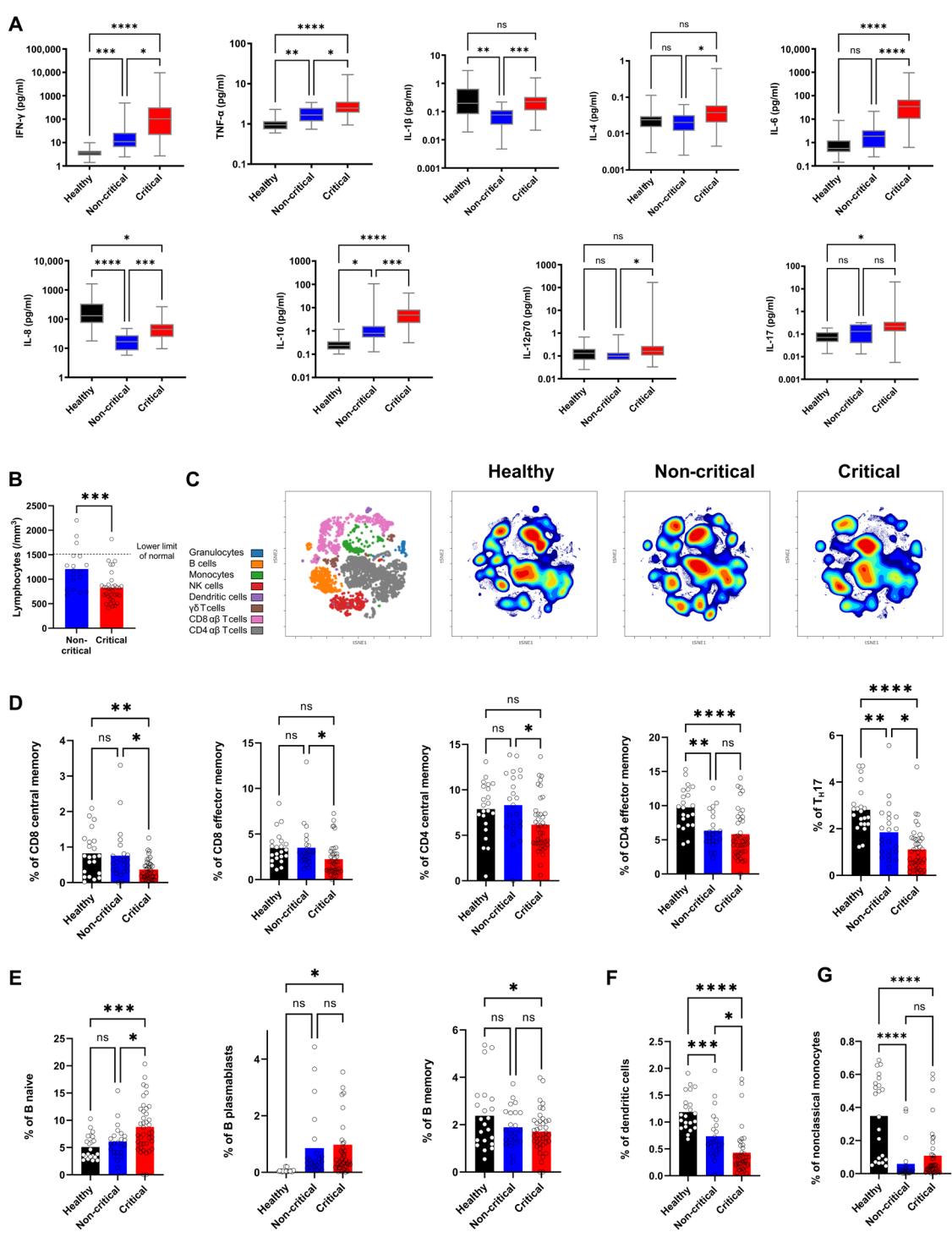
At a systemic level, lymphopenia is correlated with disease severity (Fig. 2B) (25–27). To further characterize the immune cells,

**Fig. 1. A multi-omics analytical strategy identifies key pathways and drivers of ARDS in COVID-19.** (A) Forty-seven critical patients (C), 25 non-critical patients (NC), and 22 healthy controls (H) were enrolled in the study. PBMCs were isolated by density gradient and frozen until utilization for mass cytometry and whole proteomics. Plasma was used for cytokine profiling and whole proteomics. Serum was used to measure anti-type I IFN neutralizing antibodies, anti-SARS-CoV-2 neutralizing antibodies, and multi-target antiviral serology. Whole blood was used for RNA-seq and whole-genome sequencing (WGS). The number of treated samples per group and per omics is indicated below each omics designation. (B) The RNA-seq pipeline is shown on the basis of the NC versus C comparison. To increase robustness of downstream analyses, an ensemble intelligence approach with seven algorithms was applied to multiple partitions of the RNA-seq data (see Materials and Methods) to classify NC versus C patients, performing differential analysis on each partition of the data. An ensemble ranking score across six of the seven algorithms and all partitions of the data was then determined, and the top 600 of those genes were used as the input for structural causal modeling to derive a putative causal network. To support the key findings from the first patient cohort, RNA-seq data from a second patient cohort consisting of 81 critical and 73 recovered critical patients were used. The data were partitioned analogously to the first patient cohort, but only the top 600 features from the first patient cohort were used to assess the informativeness of the gene signature. (C) Cytokines and immune cells were quantified. WGS data were used for eQTL analysis together with the gene counts from the RNA-seq. Proteomics data were subjected to differential protein expression and nGOSeq enrichment analyses. (D) The key pathways and drivers resulting from the omics analyses in (B) and (C) were validated in a second cohort of 81 critical and 73 recovered critical patients;  $*P < 0.05$  and  $****P < 0.0001$ . The differential expression of ADAM9, the main driver gene, was compared to publicly available bulk RNA-seq data. Last, ex vivo infection experiments with SARS-CoV-2 were conducted to validate a driver gene candidate.



**Fig. 2. Immune profiling differentiates healthy individuals, non-critical patients with COVID-19, and critical patients with COVID-19.**

(A) The concentrations of pro-inflammatory cytokines in plasma were quantified by cytokine profiling assays or ELISA. (B) Absolute lymphocyte counts are shown. Each dot represents a single patient. The dashed horizontal line indicates the lower limit of normal lymphocyte concentrations. (C) tSNE maps are shown colored according to the cell density across the three groups. Red indicates the highest density of cells. The plots are representative of 40 critical patients, 23 non-critical patients, and 22 healthy controls. (D) to (G) The proportions of modified lymphocyte subsets from patients with COVID-19 and healthy controls were determined by mass cytometry. Proportions of T cell subsets (D), B cell subsets (E), dendritic cells (F), and nonclassical monocytes (G) are shown. Each dot represents a single patient. In (A) and (D) to (G), the *P* values were determined with the Kruskal-Wallis test followed by Dunn's posttest for multiple group comparisons; \**P*<0.05, \*\**P*<0.01, \*\*\**P*<0.001, and \*\*\*\**P*<0.0001; ns, not significant. In (B), the *P* value was determined by a two-tailed unpaired *t* test; \**P*<0.05, \*\**P*<0.01, \*\*\**P*<0.001, and \*\*\*\**P*<0.0001. In (A), data are shown as box-and-whisker plots with medians, 25th to 75th percentiles, and maximal and minimal values and include *n*=41 critical patients, *n*=24 non-critical patients, and *n*=21 healthy controls. In (B), (D), and (E) to (G), all data points are shown and bars represent means with *n*=40 critical patients, *n*=23 non-critical patients, and *n*=22 healthy controls.



we analyzed PBMCs by mass cytometry using an immune profiling assay covering 37 cell populations. Visualization of stochastic neighbor embedding (tSNE) showed a cell population density distribution pattern that was specific to the critical group (Fig. 2C). This pattern could be partly linked to the known immunosuppression phenomenon in critical patients (12, 28, 29), which was characterized by

marked differences in the T cell compartments, where memory CD4 and CD8 T cells and Th17 cells were negatively correlated with disease severity (Fig. 2D). The latter observation is in line with the absence of a clear association between the plasma concentration of IL-17 and disease severity (Fig. 2A). In contrast, the B cell compartments of critical patients contained more naïve B cells and

plasmablasts and fewer memory B cells than those of healthy controls (Fig. 2E). In accordance with previous reports (17), the number of plasmablasts tended to be higher in critical versus non-critical patients. Moreover, non-critical and critical patients were also characterized by lower numbers of dendritic cells and nonclassical monocytes (Fig. 2, F and G). The remaining cell populations are presented in fig. S2. Together, the results indicate that critical illness was characterized by a proinflammatory cytokine storm and notable changes in the T, B, dendritic, and monocyte cell compartments. These specific changes were independent from the extent of viral infection, as both the global anti-SARS-CoV-2 antibody concentrations and their neutralizing activity were not different in critical versus non-critical patients (fig. S3, A and B).

To complete the immunologic profile, on the basis of findings suggesting that at least 10% of critical patients have preexisting anti-type I IFN autoantibodies (30, 31), we measured anti-IFN- $\alpha$ 2 and anti-IFN- $\omega$  neutralizing autoantibodies in patients and controls. Autoantibodies against type I IFNs were identified in two critical patients (fig. S3C) but none of the non-critical patients or the healthy controls. In these two patients, the presence of autoantibodies was associated with an absence of SARS-CoV-2 neutralizing antibody titers (fig. S3D).

### Quantitative plasma and PBMC proteomics highlight signatures of acute inflammation, myeloid activation, and dysregulated blood coagulation

Quantitative nanoscale liquid chromatography coupled to tandem mass spectrometry (nanoLC-MS/MS) analysis of whole unfractionated plasma samples identified a total of 336 proteins. Differential analysis was performed on an average of  $178 \pm 7$ ,  $189 \pm 11$ , and  $195 \pm 8$  proteins in healthy individuals, non-critical patients, and critical patients, respectively (Fig. 3A). These experiments were conducted on crude liquid digested plasma samples without any fractionation or depletion of high abundant proteins to favor repeatability and robustness of quantification and differential analysis, at the cost of a lower proteome coverage. After validating the homogeneous distribution of the three groups using a multidimensional scaling plot, we performed a differential protein expression analysis to identify protein signatures that were specific to critical patients (Fig. 3, B and C). In line with previous studies (8, 32), the antimicrobial calprotectin (heterodimer of S100A8 and S100A9) was among the top differentially expressed proteins in critical versus non-critical patients, which confirms that calprotectin is a robust marker for disease severity (Fig. 3D). Our data also showed dysregulation of multiple apolipoproteins including APOA1, APOA2, APOA4, APOM, APOD, APOC1, and APOL1 (Fig. 3, C and E). Most of these proteins were associated with macrophage functions and were down-regulated in critical patients. Acute-phase proteins (CRP, CPN1, CPN2, C6, CFB, ORM1, ORM2, SERPINA3, and SAA1) were strongly up-regulated in critical patients (Fig. 3, C and E). These findings are consistent with previous studies showing that acute inflammation and excessive immune cell infiltration are associated with disease severity (26, 33, 34).

Whole-cell lysates of PBMCs from the same groups of patients and controls were also subjected to quantitative nanoLC-MS/MS analysis, which led to the identification and quantification of a total of 2196 proteins. Differential analysis was performed on an average of  $801 \pm 213$ ,  $1050 \pm 309$ , and  $1052 \pm 286$  proteins in healthy individuals, non-critical patients, and critical patients, respectively

(Fig. 3F). Although the human proteome coverage was relatively low after exclusion of contaminating fetal calf serum (FCS) peptides and the distribution of the three groups in the multidimensional scaling plot was less clear than that found for plasma proteins, the differential expression analysis between non-critical and critical patients showed dysregulation of blood coagulation and myeloid cell differentiation (Fig. 3, G to I). The latter observation involving the Carbonic anhydrase 2 (CA2), Alpha-hemoglobin-stabilizing protein (AHSP), Solute carrier family 4 member 1 (SLC4A1), Transferrin receptor protein 1 (TFRC), Dematin Actin Binding Protein (DMTN), Fatty acid synthase (FAS), and Myeloblastin (PRTN3) proteins was in line with the plasma proteomics results evidencing dysregulation of macrophages and with other reports showing that severe COVID-19 is marked by a dysregulated myeloid cell compartment (15). The profile of the blood coagulation proteins Hemoglobin subunit beta (HBB), Hemoglobin subunit delta (HBD), Hemoglobin subunit epsilon (HBE1), SLC4A1, Peroxiredoxin-2 (PRDX2), Sorcin (SRI), ADP-ribosylation factor 4 (ARF4), Mesencephalic astrocyte-derived neurotrophic factor (MANF), Integrin alpha-2 (ITGA2), Orosomucoid 1 (ORM1), and Serpin family A member 1 (SERPINA1) confirmed that severity is also associated with coagulation-associated complications that can involve either bleeding or thrombosis (35).

### Combined transcriptomics and proteomics analysis supports inflammatory pathways associated with critical disease

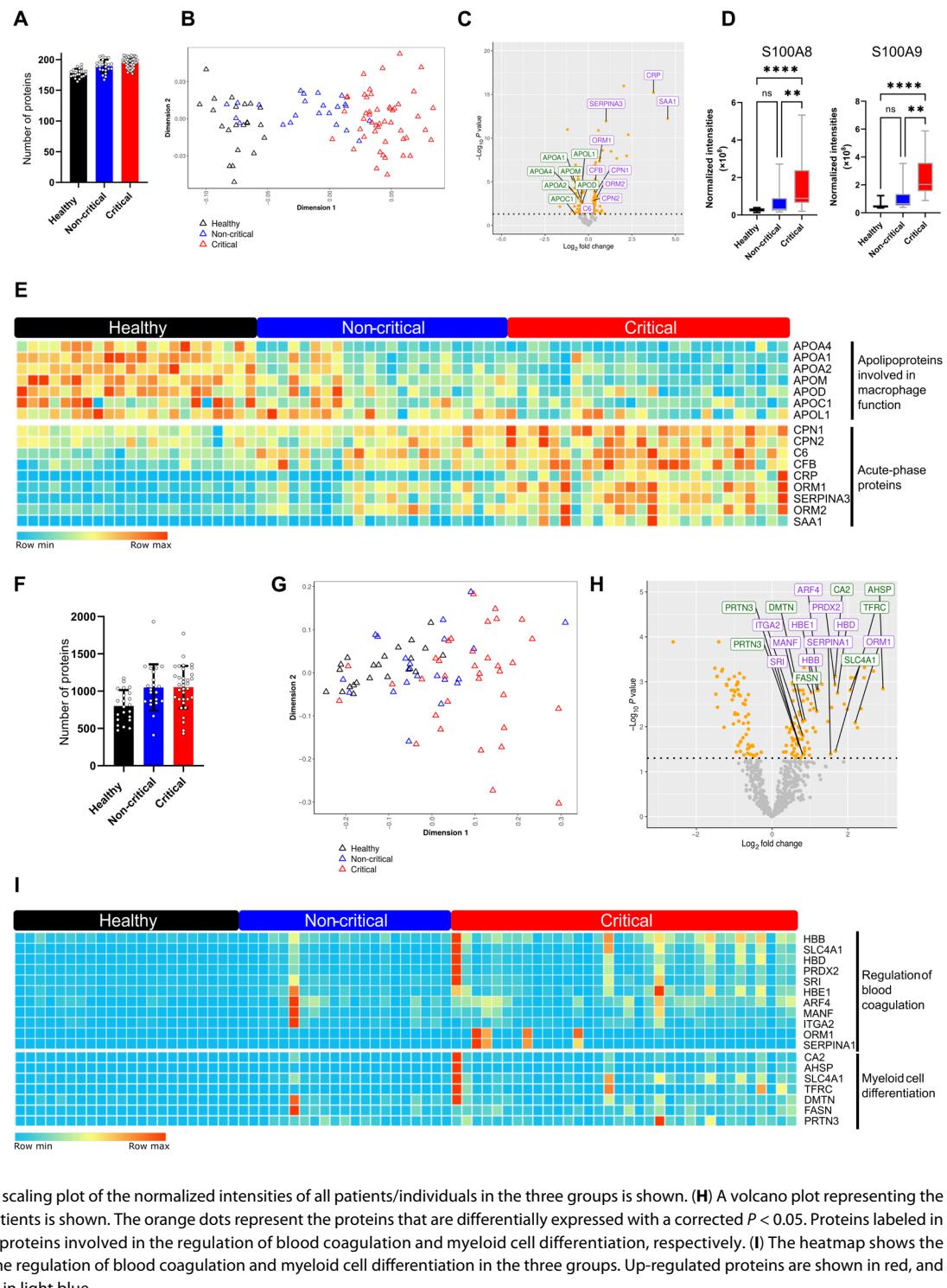
Consistent with the proteomics data, differential gene expression (DGE) and gene set enrichment analysis of RNA-seq data from whole blood samples collected from the patients showed that regulation of the inflammatory response, myeloid cell activation, and neutrophil degranulation were the main enriched pathways in critical patients with normalized enrichment scores of 2.33, 2.65, and 2.66, respectively (Fig. 4, A and B). To identify enriched pathways that were supported by different omics layers, we performed nested GOSeq (nGOSeq) (36) functional enrichment of the differentially expressed genes or proteins identified from the RNA-seq, plasma, and PBMC proteomics data (Fig. 4C). In line with the cytokine profiling results (Fig. 2A), inflammatory signaling and the response to proinflammatory cytokine release (IL-1, IL-8, and IL-12) were supported by multiple omics datasets. As suggested by the results from immune cell profiling (Fig. 2, C and D) and previous studies, the B cell response was activated, whereas the T cell response was impaired (17, 37). As previously observed (8, 14, 15, 38), the activation of neutrophils and monocytes was confirmed by the enrichment of nine different nGOSeq terms (Fig. 4C). The nGOSeq enrichment analysis also indicated that dysfunction of blood coagulation involves a fibrinolytic response; however, this observation could also be linked to the anticoagulant therapy administered to most critical patients. Moreover, nGOSeq terms related to viral entry and even viral transcription were strongly enriched for patients with critical disease across the three omics datasets. This result was consistent with the identification of viral gene transcripts in the RNA-seq data of eight critical patients but not in those from non-critical patients (table S2).

### Integrated AI, ML, and probabilistic programming reveals a robust gene expression signature and identifies driver genes that differentiate critical from non-critical patients

To robustly identify a set of genes that might differentiate between non-critical and critical patients with COVID-19 and could thus be related to the progression to ARDS, we partitioned the 69 patient

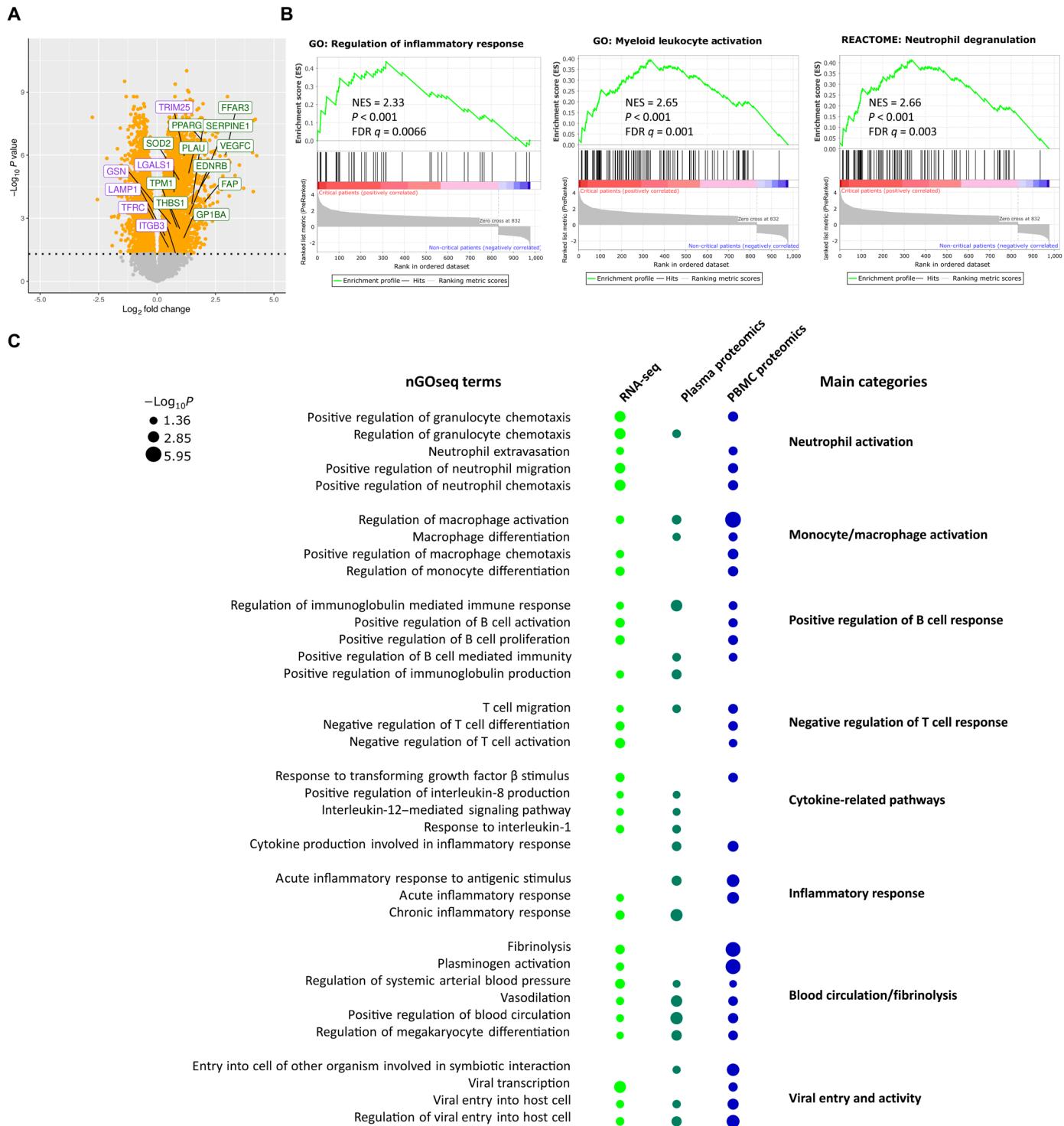
**Fig. 3. Plasma and PBMC proteomics distinguish healthy individuals, non-critical patients with COVID-19, and critical patients with COVID-19.** (A) The total number of proteins identified and used for quantification and differential analysis in the plasma of patients and healthy controls is shown. Each dot represents a patient. Bars represent means  $\pm$  SD. (B) A multidimensional scaling plot of the normalized intensities of all individuals in the three groups is shown. (C) A volcano plot representing the differentially expressed proteins (DEPs) in critical versus non-critical patients is shown. The orange dots represent the proteins that are differentially expressed with a corrected  $P < 0.05$ . Proteins labeled in green and purple represent down-regulated apolipoproteins and up-regulated acute-phase proteins, respectively. (D) Normalized intensities of the proteins S100A8 and S100A9 in the three groups are shown. Data are shown as box-and-whisker plots with medians, 25th to 75th percentiles, and maximal and minimal values and include  $n = 45$  critical patients,  $n = 23$  non-critical patients, and  $n = 22$  healthy controls.  $P$  values were determined with the Kruskal-Wallis test, followed by Dunn's posttest for multiple group comparisons; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , and \*\*\*\* $P < 0.0001$ . (E) The heatmap shows the expression of apolipoproteins involved in macrophage functions and acute-phase proteins in the three groups. Up-regulated proteins are shown in red, and down-regulated proteins are shown in light blue. (F) The total number of proteins identified and used for quantification and differential analysis in PBMCs of patients and healthy controls is shown. Each dot represents a patient. Bars represent means  $\pm$  SD. (G) A multidimensional scaling plot of the normalized intensities of all patients/individuals in the three groups is shown. (H) A volcano plot representing the DEPs in critical versus non-critical patients is shown. The orange dots represent the proteins that are differentially expressed with a corrected  $P < 0.05$ . Proteins labeled in green and purple are up-regulated proteins involved in the regulation of blood coagulation and myeloid cell differentiation, respectively. (I) The heatmap shows the expression of proteins involved in the regulation of blood coagulation and myeloid cell differentiation in the three groups. Up-regulated proteins are shown in red, and down-regulated proteins are shown in light blue.

blood RNA-seq data (46 critical and 23 non-critical patients) 100 times to account for sampling variation, using 80% for training and 20% for testing, and evaluated the performance of seven distinct AI and ML algorithms, including a quantum support vector machine (qSVM), to differentiate between patients with non-critical and critical COVID-19. We have previously shown that quantum annealing is a more robust classifier for relatively small patient training sets (39). The

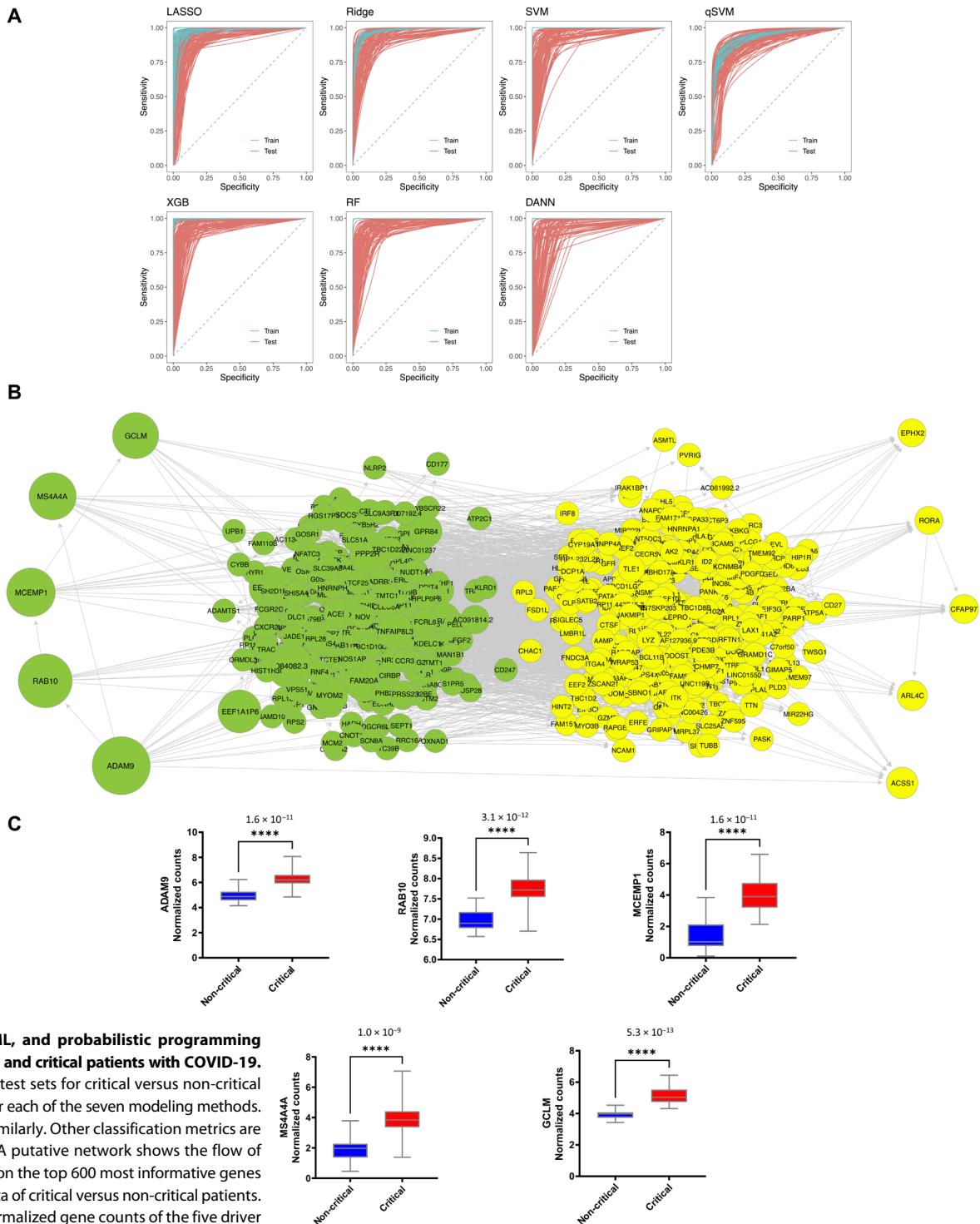


receiver operating characteristic curves (ROCs) for the 100 partitions of the patient data as well as other classification performance metrics are shown in Fig. 5A and table S3. The classification performance on the test set provided a high degree of confidence that the signals learned by the various AI and ML algorithms are generalizable.

After successfully classifying non-critical versus critical patients based on whole-transcriptome RNA-seq profiling, we assessed



**Fig. 4. RNA-seq and combined omics analysis reveal critical patient-specific pathways.** (A) A volcano plot representing the differentially expressed genes in critical versus non-critical patients is shown. The orange dots represent the genes that are differentially expressed with a corrected  $P < 0.05$ . Proteins labeled in green and purple represent up-regulated genes involved in blood pressure regulation and viral entry, respectively. (B) Gene set enrichment analysis plots show positive enrichment of inflammatory response, myeloid leukocyte activation, and neutrophil degranulation pathways in samples from patients with critical COVID-19. NES, normalized enrichment score. (C) Enriched nested gene ontology (nGO) categories are shown for critical versus non-critical patients using RNA-seq, plasma proteomics, and PBMC proteomics data.



**Fig. 5. Integrated AI, ML, and probabilistic programming distinguishes non-critical and critical patients with COVID-19.** (A) ROCs of the train and test sets for critical versus non-critical comparisons are shown for each of the seven modeling methods. All methods performed similarly. Other classification metrics are provided in table S3. (B) A putative network shows the flow of causal information based on the top 600 most informative genes for classifying RNA-seq data of critical versus non-critical patients. (C) Box plots show the normalized gene counts of the five driver genes identified that distinguish critical and non-critical patients. The indicated values correspond to the FDR. Data are shown as box-and-whisker plots with medians, 25th to 75th percentiles, and maximal and minimal values and include  $n = 46$  critical and  $n = 23$  non-critical patients.

feature scores across the six distinct ML algorithms and all partitions to determine an ensemble feature ranking, ignoring features from the partitions of patient data where the test area under the receiver operating characteristic (AUROC) was less than 0.7. Aggregating the best-performing features across both the algorithm and

data partitions afforded a more robust and stable set of generalizable features.

This signature represented hundreds of genes that are differentially expressed and, by itself, did not distinguish between driver genes of critical COVID-19 and genes that react to the disease.

Therefore, we then selected the top 600 most informative genes and used them as input for structural causal modeling (SCM) to identify likely drivers of critical COVID-19. We confirmed that these 600 genes are biologically relevant for distinguishing between critical and non-critical patients by retraining an ensemble ML classifier using only those 600 genes (table S4). Previous work has shown that SCM of RNA-seq data produces causal dependency structures, which are indicative of the signal transduction cascades that occur within cells and drive phenotypic and pathophenotypic development (40). However, this approach works best if the gene sets are stable and consistent across six different algorithms, as shown here. The resultant SCM output is presented as a directed acyclic graph (DAG) (Fig. 5B), a gene network representing the putative flow of causal information, with genes on the left predicted to have the greatest degree of influence on the entire state of the network. Perturbing these genes is the most likely to be disruptive to the state of the network (fig. S4) and is expected to exert the greatest effect on the expression of downstream genes. The top five genes associated with the greatest degree of putative causal dependency were *ADAM9*, *RAB10*, *MCEMP1*, *MS4A4A*, and *GCLM*, and all five of these genes were significantly up-regulated in critical patients with false discovery rates (FDRs) of  $1.6 \times 10^{-11}$ ,  $3.1 \times 10^{-12}$ ,  $1.6 \times 10^{-11}$ ,  $1.0 \times 10^{-9}$ , and  $5.3 \times 10^{-13}$ , respectively (Fig. 5C).

To further assess the informativeness of this COVID-19 gene expression signature, we used a second independent patient cohort consisting of critical patients with COVID-19 sampled at the time of entry into the ICU and recovered critical patients sampled at 3 months after discharge from the ICU. Patients in this second cohort were from a more typical COVID-19 ICU population, as no exclusion criteria based on age or absence of comorbidities were applied (table S1). Although non-critical patients with COVID-19 cannot be assumed to be the same as recovered critical patients with COVID-19 and, thus, the ML models from the first patient cohort cannot be directly applied to the second, the second patient cohort was used to provide additional evidence of the overall importance of the gene expression signature related to critical forms of COVID-19. The driver genes followed the same trend in the second patient cohort, namely, that all five of these genes showed increased expression in the critical COVID-19 patient groups (fig. S5A). Moreover, an ensemble of ML classifiers trained on the second cohort using the 600 genes identified in the first group of patients was well able to differentiate between critical and recovered patients (fig. S5, B and C); classification performance when training on the differentially expressed genes between critical and recovered patients was nearly the same as the first patient cohort (table S5), which further suggests a substantial degree of biological relevance of this gene signature.

### **ADAM9 is a driver of ARDS in critical patients with COVID-19**

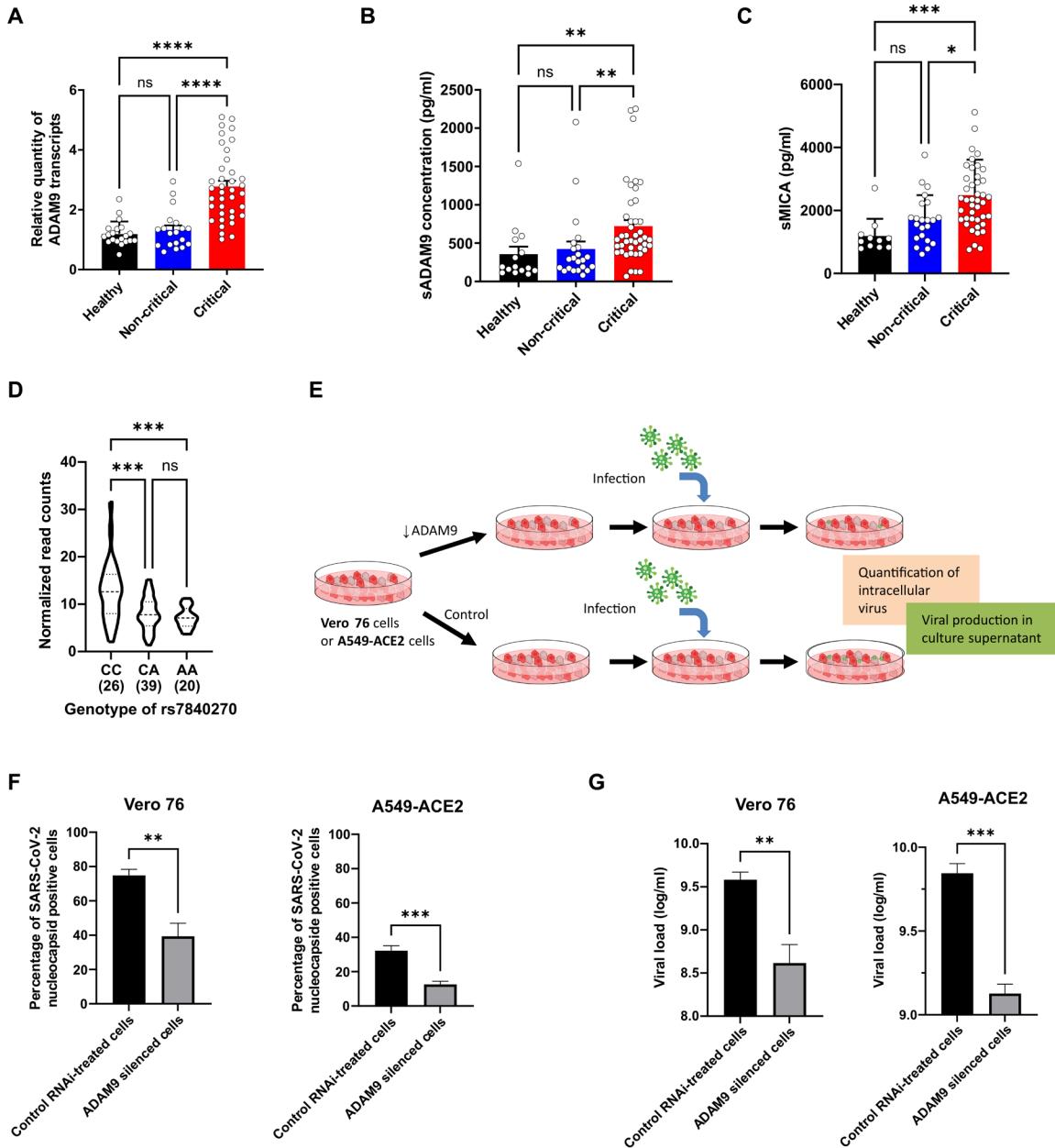
Among the five driver genes identified by SCM, we primarily focused on experimentally determining the role of *ADAM9* (a disintegrin and a metalloprotease 9) in COVID-19 etiology because (i) it was the gene with the greatest degree of causal influence in the SCM DAG, (ii) it was the only driver gene that was previously shown to interact with SARS-CoV-2 through a global interactomics approach (41, 42), and (iii) it is an entry factor for another RNA virus, the encephalomyocarditis virus (43). *ADAM9* is a metalloprotease with various functions that are mediated either by its disintegrin domain for adhesion or by its metalloprotease domain

for the shedding of a large range of cell surface proteins (44). The *ADAM9* gene encodes two isoforms that are translated into either membrane-bound or secreted protein. Although neither isoform could be detected using our proteomics approach, *ADAM9* was up-regulated at the RNA level, and the secreted form was found at a higher concentration in the serum of critical versus non-critical patients (Fig. 6, A and B). The transcriptional up-regulation of *ADAM9* was also found to be associated with disease severity in a previously published bulk RNA-seq dataset (fig. S6) (45). To assess the potential for increased metalloprotease activity in the critical cohort, we quantified the soluble form of the MICA protein (46), which is known to be cleaved by *ADAM9* (47) by enzyme-linked immunosorbent assay (ELISA). The concentration of soluble MICA (sMICA) was indeed significantly higher in the plasma of critical patients as compared to non-critical patients ( $P = 0.016$ ) and healthy controls ( $P = 0.0001$ ; Fig. 6C). A global expression quantitative trait loci (eQTL) analysis using WGS and RNA-seq data identified eight single-nucleotide polymorphisms (SNPs) associated with three of the top five putative driver genes with genome-wide significance ( $P < 0.0001$  for all SNPs; table S6). Among these SNPs, rs7840270 is localized just 0.3 kb upstream of the *ADAM9* gene and an eQTL for blood expression was reported in the Genotype-Tissue Expression database (GTEx). In the present cohort, including all three groups together, the C allele was associated with a higher abundance of *ADAM9* transcripts (Fig. 6D), as it is in the GTEx dataset. The higher expressing allele C was indeed more frequent in critical than in non-critical patients [71.3% versus 50%; Odds Ratio (OR) = 2.48; 95% confidence interval (CI): 1.14 to 5.36;  $P = 0.017$ ]. This was not due to any difference in ethnicities between critical and non-critical groups (fig. S7) (48). *ADAM9* RNA expression was significantly higher in the CC compared to CA ( $P = 0.049$ ) and AA ( $P = 0.0046$ ) genotypes only in the critical group, suggesting that the CC genotype may contribute to higher *ADAM9* RNA expression in critical patients (fig. S8).

To assess the role of *ADAM9* in viral infection, we set up an ex vivo assay in which *ADAM9* was silenced by small interfering RNA (siRNA) in Vero 76 or A549-Angiotensin-Converting Enzyme 2 (ACE2) (49) cells and subsequently infected the cells with SARS-CoV-2. Viral replication was monitored by flow cytometry quantification of the intracellular nucleocapsid protein and by quantitative viral real-time polymerase chain reaction (qRT-PCR) of the culture supernatant (Fig. 6E). The average silencing efficiency reached 66% in Vero 76 cells and 93% in A549-ACE2 cells (fig. S9). In both cell lines, the amount of intracellular virus and the quantity of released virus were lower when *ADAM9* was silenced as compared to the control condition that was treated with a control siRNA (Fig. 6, F and G). Our results collectively demonstrate that *ADAM9* is an *in vivo* up-regulated driver in critical patients. We also show a higher global proteolytic activity in serum samples of critical patients and demonstrate that a higher amount of *ADAM9* facilitates viral infection and replication in an ex vivo cellular model.

### **DISCUSSION**

A number of studies have detailed the molecular and cellular modifications associated with COVID-19 disease severity (8, 11, 12, 15, 16, 34, 45, 50–54), yet very few studies have targeted a young population with no comorbidities to reduce confounders that may also drive severity and mortality, and these confounders were limited to epidemiology or standard clinical parameters such as CRP, D-dimers, or Sequential



**Fig. 6. ADAM9 is a key driver of SARS-CoV-2 infection and replication in vitro.** (A) qRT-PCR confirmation of the differential expression of ADAM9 in non-critical ( $n = 19$ ) versus critical patients ( $n = 38$ ) and in healthy controls ( $n = 20$ ) is shown. (B) Soluble ADAM9 (sADAM9) concentration in serum samples isolated from healthy controls ( $n = 15$ ), non-critical patients ( $n = 22$ ), and critical patients ( $n = 43$ ) was determined by ELISA. (C) Soluble MICA (sMICA) concentration in serum samples isolated from healthy controls ( $n = 11$ ), non-critical patients ( $n = 22$ ), and critical patients ( $n = 43$ ) was determined by ELISA. (D) Expression of ADAM9 according to the genotype of the eQTL rs7840270 is shown ( $n$  are indicated below the genotypes). (E) The experimental approach for assessing viral uptake and viral replication in silenced Vero 76 or A549-ACE2 cells is shown. (F) Flow cytometry-based intracellular nucleocapsid staining in control and ADAM9-silenced Vero 76 and A549-ACE2 cells was quantified. One representative experiment of  $N = 3$  independent experiments with  $n = 3$  in each group is shown. (G) Quantitative RT-PCR for SARS-CoV-2 in culture supernatant after the silencing of ADAM9 in Vero 76 or A549-ACE2 cells is shown. The results from probe N1 are shown. One representative experiment of  $N = 3$  independent experiments with  $n = 3$  in each group is shown. In (A) to (D), the  $P$  values were determined with the Kruskal-Wallis test followed by Dunn's posttest for multiple group comparisons; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , and \*\*\*\* $P < 0.0001$ . In (F) and (G), the  $P$  values were determined from a two-tailed unpaired  $t$  test; \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$ . In (A) to (C) and (F) and (G), bars represent means  $\pm$  SD. RNAi, RNA interference.

Downloaded from https://www.science.org at Karolinska Institute on October 10, 2022

Organ Failure Assessment (SOFA) scores (55–57). A comprehensive understanding of the immune responses to SARS-CoV-2 infection is fundamental to develop an explanation as to why some young patients without comorbidities progress to critical illness whereas others do not, a phenomenon that has been exacerbated with

new viral variants in current epidemic waves across the globe (58, 59). In particular, knowledge of the molecular drivers of critical COVID-19 is urgently needed to identify predictive biomarkers and more efficient therapeutic targets that function through drivers of critical COVID-19 rather than to downstream or secondary events (60–62).

Here, we used a multi-omics strategy associated with integrated AI, ML, and probabilistic programming methods to identify pathways and signatures that can differentiate critical from non-critical patients in a population of patients younger than 50 years without comorbidities. This *in silico* strategy provided a detailed view of the systemic immune response that was globally in accordance with previously published data. The thrust of our work, however, was to define a consistent transcriptomic signature that can robustly differentiate critical from non-critical patients, as shown by the classification performance metrics assessed in this study. Moreover, one can infer the biological relevance of the COVID-19 gene expression signature found in patient cohort 1 as the same classification performance was achieved in the second, independent patient cohort composed of 81 critically ill and 73 recovered critical patients.

Using the top 600 gene expression features of the signature as the input for SCM, we derived a causal network that uncovered five putative driver genes: *RAB10*, *MCEMP1*, *MS4A4A*, *GCLM*, and *ADAM9*. *RAB10* (Ras-related protein Rab-10) is a small guanosine triphosphatase that regulates macropinocytosis in phagocytes (63), which is a mechanism that has been suggested to be involved in the entry of SARS-CoV-2 into respiratory epithelial cells (64). *MCEMP1* (mast cell expressed membrane protein 1) is a membrane protein specifically associated with lung mast cells, and decreasing the expression of this protein has been shown to reduce inflammation in septic mice (65, 66). *MS4A4A* (a member of the membrane-spanning, four domain family, subfamily A) is a surface marker for M2 macrophages that mediates immune responses in pathogen clearance (67) and regulates arginase 1 induction during macrophage polarization and lung inflammation in mice (68). *GCLM* (glutamate-cysteine ligase modifier subunit) is the first rate-limiting enzyme of glutathione synthesis and has been linked to severe COVID-19 (68). Although these four genes are all good candidates that can at least partially explain the severity of the disease, we focused our functional validations on *ADAM9*, which represented, from an *in silico* standpoint, the most promising driver gene. The confirmed up-regulation of *ADAM9* at the RNA and protein levels in critical patients, which might partly be linked to prestored *ADAM9* release by neutrophils (69), the increased metalloprotease activity in these same patients, and our *ex vivo* validation of its effect on viral uptake and replication are strong arguments supporting the targeting of this protein as a potential therapeutic strategy for the treatment or prevention of critical COVID-19. *In vitro*, we found that *ADAM9* markedly affects viral uptake or replication. The inhibition of this presumed mechanism of action of *ADAM9* might represent a target for the treatment of SARS-CoV-2 or other viral infections. Moreover, therapies that block viral uptake rather than host receptor binding are more likely to be variant independent, a known virological behavior that might, at least partially, compromise current vaccination efforts (70).

Because of its implication in tumor progression and metastasis, *ADAM9* is currently being tested as a target of antibody-drug conjugate therapy for solid tumors (71). A repurposing strategy using *ADAM9*-blocking antibodies for the treatment of critical patients with COVID-19 could therefore be envisioned. Alternatively, other therapeutic agents to reduce the *ADAM9* concentration or activity could be pursued.

Our study has several limitations. On the basis of the present experimental results, we cannot conclude yet as to the molecular mechanism linking *ADAM9* and viral uptake or replication. The

predictive performance of *ADAM9* as a diagnostic marker for disease severity and as a therapeutic target has to be evaluated in further studies. In addition, because of the differences in the first and second patient cohorts, we were unable to fully assess the generalizability of the RNA-seq gene signature found in the first patient cohort to the second patient cohort. Last, we did not test the silencing of *ADAM9* on various SARS-CoV-2 variants.

In conclusion, this study presents a detailed multi-omics investigation of a well-characterized cohort of young, previously healthy, critical COVID-19 patient series compared with non-critical patients and healthy controls. In addition to uncovering a landscape of molecular changes in the blood of critical patients, we applied a data-driven ensemble AI/ML strategy, which was independent of prior biological knowledge and thus minimized possible annotation biases, to gain insights into COVID-19 pathogenesis and to provide potential candidate diagnostic, prognostic, and especially much needed therapeutic targets that might be helpful in combating the COVID-19 pandemic.

## MATERIALS AND METHODS

### Study design

In March and April 2020, patients aged less than 50 years who had no comorbidities (of note, obesity alone was not considered an exclusion criterion) and were admitted for COVID-19 to the infectious disease unit (hereafter designated non-critical care ward) or to the designated ICUs at the university hospital network in northeast France (Alsace) were investigated within the framework of the present study. Follow-up was performed until hospital discharge. SARS-CoV-2 infection was confirmed in all the patients by qRT-PCR tests for COVID-19 nucleic acid of nasopharyngeal swabs (72). The ethics committee of Strasbourg University Hospitals approved the study (COVID-HUS, reference CE: 2020-34). Written informed consent was obtained from all the patients. The demographic characteristics, medical history, and symptoms were reported. Three groups were considered: (i) the “critical group,” which included 47 patients admitted to the ICU; (ii) the “non-critical group,” which was composed of 25 hospitalized patients at the non-critical care ward; and (iii) the “healthy control group,” which included 22 healthy age- and sex-matched blood donors aged less than 50 years. A second, independent cohort composed of 81 critical patients and 73 recovered critical patients from one of the ICU departments of Strasbourg University hospitals was used to further evaluate our molecular classification findings. No sample size calculations, randomization, or blinding was performed.

### Sampling

Venipunctures were performed within the first hours after admission to the ICU or medical ward within the framework of routine diagnostic procedures. A subset of ICU patients (73%) were sampled every 4 to 8 days after hospitalization until discharge or death. Patient blood was collected into BD Vacutainer tubes with heparin (for plasma and PBMCs), with EDTA (for DNA), or without additive (for serum) and into PAXgene Blood RNA tubes (Becton, Dickinson and Company). Blood from healthy donors was sampled in BD Vacutainer tubes with heparin, with EDTA, or without additive. Plasma and serum fractions were collected after centrifugation at 900g at room temperature for 10 min, aliquoted, and stored at -80°C until use. PBMCs were prepared within 24 hours by Ficoll density

gradient centrifugation. Aliquots of  $1 \times 10^6$  dry cell pellets were frozen at  $-80^{\circ}\text{C}$  until use for proteomics. Aliquots of a minimum of  $5 \times 10^6$  cells were frozen at  $-80^{\circ}\text{C}$  in 90% FCS/10% dimethyl sulfoxide. The EDTA and PAXgene tubes were stored at  $-80^{\circ}\text{C}$  until use for DNA and RNA extraction, respectively.

### Cytokine profiling

The plasma samples were analyzed using the V-PLEX Proinflammatory Panel 1 Human Kit (IL-6, IL-8, IL-10, TNF- $\alpha$ , IL-12p70, IL-1 $\beta$ , IL-2, IL-4, and IFN- $\gamma$ ) and the S-PLEX Human IFN- $\alpha$ 2a Kit following the manufacturer's instructions (Meso Scale Discovery). Undiluted plasma was used for the S-PLEX Human IFN- $\alpha$ 2a Kit, and plasma was diluted twofold for use with the V-PLEX Proinflammatory Panel 1. The MSD plates were analyzed with an MS2400 imager (Meso Scale Discovery). Soluble IL-17 in undiluted serum was quantified by Quantikine HS ELISA (Human IL-17 Immunoassay) following the manufacturer's instructions (R&D Systems). All standards and samples were measured in duplicate.

### Immune phenotyping by mass cytometry

PBMCs were thawed rapidly, washed twice with 10 volumes of RPMI 1640 medium (Thermo Fisher Scientific), and centrifuged for 7 min at 300g at room temperature between each washing step. Cells were then treated with 250 U of deoxyribonuclease (DNase; Thermo Fisher Scientific) in 10 volumes of RPMI 1640 medium for 30 min at  $37^{\circ}\text{C}$  in the presence of 5% CO<sub>2</sub>. During this step, the viability and the number of the cells were determined with Trypan Blue (Thermo Fisher Scientific) and Türk's solution (Merck Millipore), respectively. After the elimination of DNase by centrifugation for 7 min at 300g at room temperature, a total of  $3 \times 10^6$  cells were used for immunostaining with the Maxpar Direct Immune Profiling Assay kit (Fluidigm), following the manufacturer's instructions, except that we used 32% paraformaldehyde (Electron Microscopy Sciences). A red blood cell lysis step was included after the immunostaining following the manufacturer's instructions. The prepared cells were stored at  $-80^{\circ}\text{C}$  until use for acquisition with a Helios mass cytometer system (Fluidigm). An average of 600,000 events were acquired per sample. The mass cytometry standard files produced with the Helios instrument were analyzed using Maxpar Pathsetter software v.2.0.45 that was modified for live/dead parameters: The tallest peak was selected instead of the closest peak for the identification and quantification of the cell populations. The flow cytometry standard files from each group (healthy, critical, and non-critical) were then concatenated using CyTOF software v.7.0.8493.0 for viSNE analysis (Cytobank Inc.). A total of 300,000 events were used for the viSNE map that was generated with the following parameters: iterations (1000), perplexity (30), and theta (0.5). viSNE maps are presented as the means of all samples in each group.

### Plasma proteomics analysis

Two microliters of plasma was prepared using the PreOmics iST Kit (PreOmics GmbH) according to the manufacturer's protocol before nanoLC-MS/MS analysis on a nanoAcuity UltraPerformance LC (UPLC) device (Waters Corporation) coupled to a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific), as detailed in the Supplementary Materials. A sample pool comprising equal amounts of all protein extracts was constituted and regularly injected during the course of the experiment as an additional quality control.

The raw data obtained from each sample (45 critical patients, 23 non-critical patients, and 22 healthy controls) were processed using MaxQuant (version 1.6.14). Peaks were assigned using the Andromeda search engine with trypsin/P specificity. A database containing all human entries was extracted from the UniProtKB/Swiss-Prot database (11 May 2020, 20,410 entries). The minimal peptide length required was seven amino acids, and a maximum of one missed cleavage was allowed. Methionine oxidation and acetylation of the proteins' N termini were set as variable modifications, and acetylated and modified methionine-containing peptides, as well as their unmodified counterparts, were excluded from the protein quantification step. Cysteine carbamidomethylation was set as a fixed modification. The "match between runs" option was enabled. The maximum FDR was set to 1% at the peptide and protein levels with the use of a decoy strategy. The normalized label-free quantification (LFQ) intensities were extracted from the ProteinGroups.txt file after the removal of nonhuman and keratin contaminants, as well as reverse and proteins only identified by site. This resulted in 336 quantified proteins. Complete datasets have been deposited in the ProteomeXchange Consortium database with the identifier PXD025265 (73).

The LFQ values from MaxQuant were used for differential protein expression analysis. For each pairwise comparison, the proteins expressed in at least 80% of the samples in either group were retained. Variance stabilization normalization (Vsn) was performed using the justvsn function from the vsn R package (74). Missing values were imputed using the random forest (RF) approach (75). This process resulted in 161 proteins. Differential protein expression analysis was performed using the limma Bioconductor package in R (76). Significant differentially expressed proteins were determined on the basis of an adjusted *P* value cutoff of 0.05 using the Benjamini-Hochberg method.

### PBMC proteomics analysis

PBMC pellets were prepared using the PreOmics' iST Kit (PreOmics GmbH) according to the manufacturer's protocol before nanoLC-MS/MS analysis on a nanoAcuity UPLC device (Waters Corporation) coupled to a Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific, Waltham), as detailed in the Supplementary Materials.

The raw data obtained from each sample (34 critical patients, 21 non-critical patients, and 22 healthy controls) were processed using MaxQuant (version 1.6.14). Peaks were assigned using the Andromeda search engine with trypsin/P specificity. A combined human and bovine database (because of contamination with FCS in the samples) was extracted from UniProtKB/Swiss-Prot (8 September 2020, 26,413 entries). The minimal peptide length required was seven amino acids, and a maximum of one missed cleavage was allowed. Methionine oxidation and acetylation of the proteins' N termini were set as variable modifications, and acetylated and modified methionine-containing peptides, as well as their unmodified counterparts, were excluded from protein quantification. Cysteine carbamidomethylation was set as a fixed modification. The "match between runs" option was enabled. The maximum FDR was set to 1% at the peptide and protein levels with the use of a decoy strategy. Only peptides unique to human entries were retained, and their LFQ intensities were summed to derive the protein intensities. This process resulted in 2196 quantified proteins. Complete datasets have been deposited in the ProteomeXchange Consortium database with the identifier PXD 025265 (73).

Summed peptide normalized LFQ (LFQ values from MaxQuant software) values were used for differential protein expression analysis. For each pairwise comparison, proteins expressed in at least 80% of the samples in either group were retained. Vsn was performed using the justvsn function from the vsn R package (74). Missing values were imputed using the RF approach (75). This resulted in 732 proteins. Differential protein expression analysis was performed using the limma Bioconductor package in R (76). Significant differentially expressed proteins were determined on the basis of an adjusted *P* value cutoff of 0.05 using the Benjamini-Hochberg method.

### Whole-genome sequencing

WGS data were generated from DNA isolated from whole blood. NovaSeq 6000 (Illumina Inc.) machines were used for DNA sequencing to a mean 30 $\times$  coverage. The raw sequencing reads from FASTQ files were aligned using Burrows-Wheeler Aligner (77), and Genomic Variant Call Format files were generated using Sentieon version 201808.03 (78). Functional annotation of the variants was performed using Variant Effect Predictor from Ensembl (version 101). GATK version 4 (79, 80) was used for the joint genotyping process and variant quality score recalibration (VQSR). We removed one duplicate sample on the basis of kinship (king cutoff of 0.3) and retained 24,476,739 SNPs that were given a “PASS” filter status by VQSR. The analysis of the 72 samples from the critical and non-critical groups identified 15,870,076 variants with Minor Allele Frequency (MAF) < 5%. The first two principal components were generated using plink2 on LD-pruned variants with Hardy-Weinberg equilibrium in the controls with  $P \geq 1 \times 10^{-6}$  and MAF > 5% and were used as covariates to correct for population stratification.

### Analysis of eQTLs

We performed local (cis-) eQTL analysis to test for associations between genetic variants and gene expression in 67 samples having both RNA-seq and SNP genotype data. Briefly, we used the Matrix-EQTL R package (81) where we selected a linear model and a maximum distance for gene-SNP pairs of  $1 \times 10^6$ . The top two principal components identified from the genotype principal components analysis were used as covariates to control for population stratification. We selected 304,044 significant eQTLs with FDR  $\leq 0.05$ .

### RNA sequencing

Whole-blood RNA was extracted from PAXgene tubes with the PAXgene Blood RNA Kit following the manufacturer’s instructions (QIAGEN). A total of 69 samples, including 46 critical and 23 non-critical patients, were processed. The RNA quantity and quality were assessed using the Agilent 4200 TapeStation system (for the RNA Integrity Number) (Agilent Technologies) and RiboGreen (for the concentration) (Thermo Fisher Scientific). RNA-seq libraries were generated using the TruSeq Stranded Total RNA with Ribo-Zero Globin kit (Illumina) and sequenced on the Illumina NovaSeq 6000 instrument with S4 flow cells and 151-base pair paired-end reads. The raw sequencing data were aligned to a reference human genome build 38 (GRCh38) using the short reads aligner STAR (82). Quantification of gene expression was performed using RSEM (83) with GENCODE annotation v25 ([www.gencodegenes.org](http://www.gencodegenes.org)). Raw and processed datasets have been deposited in Gene Expression Omnibus (GEO) with identifier GSE172114.

DGE analysis was performed for two different purposes: (i) for the combined omics analysis of differentially expressed genes and

proteins and (ii) as a step to determine feature selection for classification in the in silico computational intelligence approach. For the combined omics analysis, we first removed lowly expressed genes for the 69 samples by removing genes with less than 1 count per million in less than 10% of the samples. We then performed DGE analysis on all 69 samples using the trimmed mean of *M* values method (TMM) from the edgeR R package (84, 85).

In our computational intelligence approach, we performed DGE analysis for each partition of the train data using a frozen TMM normalization to calculate normalization factors based only on the training data to avoid data leakage. Briefly, we removed lowly expressed genes for the 69 samples with genes with 1 count per million in less than 10% of samples. For each partition of the training data, we calculated the normalization factors and then selected the library that had a normalization factor closest to 1. We used this library as a reference library to normalize all the samples, keeping the training normalization factors unchanged. Differentially expressed genes were identified using quasi-likelihood *F* test-adjusted *P* values from the edgeR R package. Differentially expressed genes with FDRs less than 0.05 were used for further downstream analysis.

### Identification of potential driver genes through SCM

To identify potential biomarkers that might differentiate patients in the non-critical group from those in the critical group, we used classification as a feature selection approach and then used the most informative features as input for SCM to identify potential driver genes. More specifically, classification was performed using the RNA-seq data by repeatedly partitioning non-critical and critical patients into 100 unique training and independent test sets representing 80 and 20% of the total data, respectively, ensuring that the proportions of non-critical and critical patients were consistent in each partition of the data. One hundred partitions of the data were used to capture the biological variation and to obtain increased statistical confidence in the results. After classification, feature scores for each method were determined and combined across all 100 partitions of the data and six of the ML algorithms, not including the deep learning algorithm. To capture as much information as possible while still being able to finish the analysis in a reasonable amount of time, the 600 most informative features were retained for SCM (600 features is the maximum that the SCM can finish in a reasonable amount of time). We used seven distinct ML approaches for our classification models: LASSO, Ridge, SVM, qSVM, eXtreme Gradient Boosting (XGB), RF, and a deep artificial neural network (DANN). A description of the algorithms and their relevant hyperparameters are mentioned in their respective sections in the Supplementary Materials. Hyperparameters were selected by using 10-fold cross-validation of the training data, and the performance was evaluated using the held-out test data.

### Ensemble feature ranking

To derive an ensemble ranking of the feature importance, we first calculated the feature importances for each algorithm. LASSO, Ridge, SVM, and qSVM are linear models, and thus, the feature importance was determined on the basis of the value of the weight assigned to each feature, with a larger score corresponding to greater importance. RF creates a forest of decision trees, and as part of the fitting process, it determines an estimate of the feature importance by randomly permuting the features one at a time and determining

the change in the accuracy. XGB calculates the feature importance by averaging the gain across all the trees, where the gain is the difference in the Gini purity of the parent node and the two children nodes.

The top 1000 most informative features of each model and for each partition of the data were retained. Because there were 100 partitions of the data and six algorithms (LASSO, Ridge, SVM, qSVM, RF, and XGB; DANN was not included because it lacks a robust approach to determine the feature importance) and up to 1000 features were retained, a total of up to 600,000 possible features were considered for each feature set, although it may be lower as the features may not be unique such that the 1000 features for one partition of the data might exhibit some overlap with the top 1000 features for another partition of the data. We discarded the feature scores from an algorithm on any partition with a test AUROC <0.7 in an attempt to exclude scores that might not truly be informative. To aggregate the scores, we scaled the scores by the most informative feature for each algorithm on each partition such that the feature scores were all between 0 and 1; for the first partition of the data, we scaled the 1000 most informative features from LASSO, then proceeded to do the same for Ridge, SVM, and RF, and then repeated the process for each partition of the data. The scores were then averaged across all the partitions of the data to obtain a feature ranking for each method. If a feature was determined to be important for one partition of the data but not for others, it was given a value of 0 for all partitions of the data in which it did not appear. To determine a final ensemble feature ranking, the grand mean across all training partitions and algorithms was taken, and the features were sorted by the average score.

### Structural causal modeling

We generated Bayesian belief networks (BBNs) for the top 600 most informative genes as defined by ensemble feature ranking described above on the first patient cohort (the informativeness of those 600 genes was evaluated in the second patient cohort). Six hundred genes were chosen to capture as much information possible while still allowing the algorithm to finish in a reasonable amount of time. A BBN is a DAG, where the directionality of the arcs represents conditional dependencies between the nodes. Training of BBNs was performed in R using the bnlearn package (86). See the Supplementary Materials for more details.

### Real-time reverse transcription quantitative PCR

Total RNA was extracted from cells using the RNeasy Mini Kit (QIAGEN), and the RNA quality was assessed using an Agilent 2100 Bioanalyzer before reverse transcription into cDNA with Maxima H Minus Master Mix and following the manufacturer's instructions (Thermo Fisher Scientific). RT-qPCR was performed using QuantStudio3 (Thermo Fisher Scientific) according to the manufacturer's protocol and using PowerTrack SYBR Green Master Mix (Thermo Fisher Scientific, Waltham, MA, USA). The following primers were used: *ADAM9*, forward 5'-GGACTCAGAGGATTGCT-GCATTTAG-3' and reverse 5'-CTTCGAAGTAGCTGAGTCATGCTGG-3'; and *GAPDH* (housekeeping gene), forward 5'-GGTGAAGGTGCG-GAGTCAACGGA-3' and reverse 5'-GAGGGATCTCGCTCCTGG-AAGA-3' (Integrated DNA Technologies). The qRT-PCR protocol consisted of 95°C for 2 min followed by 40 cycles of 95°C for 5 s and 60°C for 30 s. All reactions were performed in duplicate, and the relative amounts of transcripts were calculated with the

comparative Ct method. Gene expression changes were calculated using the  $2^{-\Delta\Delta Ct}$  values calculated from averages of technical duplicates relative to the negative control. Melting curve analysis was performed to assess the specificity of the PCR products.

### Enzyme-linked immunosorbent assays

The concentrations of soluble ADAM9 (sADAM9) and sMICA in the serum of critical and non-critical patients and healthy controls were quantified by ELISA. For sADAM9, we used a Human sADAM9 DuoSet ELISA kit (R&D Systems) following the manufacturer's instructions. sMICA concentrations were measured with an in-house developed sandwich ELISA using two monoclonal mouse antibodies for capture (A13-C485B10 and A9-C255A9 at concentrations of 2 and 0.2 mg/ml, respectively) and one biotinylated monoclonal mouse antibody for detection (A15-C199B9 at 60 pg/ml); all three antibodies were made in-house and described by Carapito *et al.* (87). Coating of MaxiSorp ELISA plates (Thermo Fisher Scientific) was performed in phosphate-buffered saline (PBS) at 4°C overnight. After three washing steps with PBS, the wells were blocked with 200 µl of 10% bovine serum albumin (BSA) in PBS for 1 hour at room temperature. All the following steps were carried out at room temperature with PBS/0.05% Tween 20/10% BSA, which was used as a diluent for all the reagents and serum samples. The plates were washed three times with PBS/0.05% Tween 20 between incubation steps. After blocking, the plates were incubated with 100 µl of sera, standards, and controls for 2 hours, followed by incubation with 100 µl of biotinylated detection antibody for 1 hour. The plates were subsequently incubated for 1 hour with 100 µl of a 5000-fold dilution of streptavidin poly-horseradish peroxidase (Thermo Fisher Scientific) per well. The reactions were lastly revealed using 3,3',5,5'-tetramethylbenzidin Ultra (Thermo Fisher Scientific) at 100 µl per well for 15 min and stopped with 100 µl of 1 M HCl. The absorbance was measured at 450 nm on a Varioskan LUX (Thermo Fisher Scientific).

### Silencing and cell transfection

Vero 76 cell lines (Vero C1008; catalog no. CRL-1586, Clone E6) were grown at 37°C under 5% CO<sub>2</sub> and maintained in Dulbecco's modified Eagle's medium (DMEM; Thermo Fisher Scientific) containing penicillin (100 U/ml) and supplemented with 10% FCS (Pan Biotech). ACE2-expressing A549 cells (A549-ACE2; a gift from O. Schwartz, Institut Pasteur) were grown at 37°C under 5% CO<sub>2</sub> and maintained in DMEM (Thermo Fisher Scientific) containing blasticidin S (10 µg/ml; Invitrogen). The cells were transfected with predesigned Stealth siRNA directed against *ADAM9* (HSS112867) or the control Stealth RNAi Negative Control Duplex medium GC (45 to 55%) (Thermo Fisher Scientific) using Lipofectamine RNAiMAX Transfection Reagent (Thermo Fisher Scientific). One day before transfection, the cells were seeded in a 24-well plate at  $0.05 \times 10^6$  cells per well. First, 1.5 µl of Lipofectamine RNAiMAX Transfection Reagent was added to 25 µl of Opti-MEM medium, followed by the addition of the mix containing 5 pmol of siRNA in 25 µl of Opti-MEM medium (Thermo Fisher Scientific). The mixture was incubated at room temperature for 5 min and then added to the cells. The cells were collected or infected after 48 hours.

### In vitro viral infections

Vero 76 and A549-ACE2 cell lines were infected with wild-type SARS-CoV-2 virus at multiplicities of infections of 10 and 400, respectively. The percentage of infected cells was determined by

staining with SARS-CoV-2 nucleocapsid (percentage of nucleocapsid positive cells), and virus released into the supernatant was analyzed by RT-PCR (copies/ml) after 2 and 3 days of infection for Vero 76 and A549-ACE2 cells, respectively. The cells were fixed for 20 min in 3.6% paraformaldehyde at 4°C, washed in 5% FCS in PBS, and stained with anti-nucleocapsid antibody (GTX135357, GeneTex) at a 1:200 dilution in Perm/Wash (Becton, Dickinson and Company) for 45 min at room temperature. Samples were then incubated with Alexa Fluor 647-labeled goat anti-rabbit monoclonal antibody (Ab150083, Abcam, Cambridge, UK) diluted 1:200 in 5% FCS in PBS for 45 min at room temperature. Samples were acquired with a MACSQuant flow cytometer (Miltenyi Biotec) and analyzed with the Kaluza software (Beckman Coulter).

RNA was extracted from the supernatant of infected cells using the NucleoSpin Dx Virus Kit (Macherey-Nagel GmbH & Co.KG). RT-qPCR was performed using TaqPath 1-Step RT-qPCR Master Mix (CG) on the Quanstudio3 instrument (Thermo Fisher Scientific). The primer/probe mix used for absolute quantification of the virus was N1 and N2 from the 2019-nCoV RUO Kit (Integrated DNA Technologies), and the positive control for the standard curve was 2019-nCoV N Positive Control (Integrated DNA Technologies). The reaction was performed in 20 µl, which included 5 µl of eluted RNA, 5 µl of TaqPath Master Mix, and 1.5 µl of the primer/probe. The qRT-PCR protocol consisted of 25°C for 2 min, 50°C for 15 min, and 95°C for 2 min and 40 cycles of 95°C for 3 s and 60°C for 30 s. All reactions were performed in duplicate, and absolute quantification was calculated with the standard curve of the positive control.

## Statistical analysis

Statistical analysis was performed with GraphPad Prism (GraphPad Software) unless stated otherwise. A *P* value below 0.05 was considered significant. For two group comparisons, data were analyzed by unpaired, two-sided Mann-Whitney or Student's *t* test. For three or more group comparisons, data were analyzed by unpaired, two-sided Kruskal-Wallis test, followed by Dunn's posttest (\**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, and \*\*\*\**P* < 0.0001). In figures and tables, “*n*” represents the number of biological replicates and “*N*” denotes the number of times an experiment was independently performed.

## SUPPLEMENTARY MATERIALS

[www.science.org/doi/10.1126/scitranslmed.abj7521](https://www.science.org/doi/10.1126/scitranslmed.abj7521)

Materials and Methods

Figs. S1 to S9

Tables S1 to S6

Data file S1

MDAR Reproducibility Checklist

References (88–106)

[View/request a protocol for this paper from Bio-protocol.](#)

## REFERENCES AND NOTES

- R. Horton, Offline: COVID-19 is not a pandemic. *Lancet* **396**, 874 (2020).
- P. Quah, A. Li, J. Phua, Mortality rates of patients with COVID-19 in the intensive care unit: A systematic review of the emerging literature. *Crit. Care* **24**, 285 (2020).
- Severe Covid-19 GWAS Group, D. Ellinghaus, F. Degenhardt, L. Bujanda, M. Buti, A. Albillos, P. Invernizzi, J. Fernandez, D. Prati, G. Baselli, R. Asselta, M. M. Grimsrud, C. Milani, F. Aziz, J. Kassens, S. May, M. Wendendorff, L. Wienbrandt, F. Uellendahl-Werth, T. Zheng, X. Yi, R. de Pablo, A. G. Chercoles, A. Palom, A. E. Garcia-Fernandez, F. Rodriguez-Frias, A. Zanella, A. Bandera, A. Prott, A. Aghebo, A. Lleo, A. Biondi, A. Caballero-Garralda, A. Gori, A. Tanck, A. C. Nolla, A. Latiano, A. L. Fracanzani, A. Peschuck, A. Julia, A. Pesenti, A. Voza, D. Jimenez, B. Mateos, B. N. Jimenez, C. Quereda, C. Paccapelo, C. Gassner, C. Angelini, C. Cea, A. Solier, D. Pestana, E. Muniz-Diaz, E. Sandoval, E. M. Paraboschi, E. Navas, F. G. Sanchez, F. Ceriotti, F. Martinelli-Boneschi, F. Peyvandi, F. Blasi, L. Tellez, A. Blanco-Grau, G. Hemmrich-Stanisak, G. Grasselli, G. Costantino, G. Cardamone, G. Foti, S. Anelj, H. Kurihara, H. ElAbd, I. My, I. Galvan-Femenia, J. Martin, J. Erdmann, J. Ferrusquia-Acosta, K. Garcia-Etxebarria, L. Izquierdo-Sanchez, L. R. Bettini, L. Sumoy, L. Terranova, L. Moreira, L. Santoro, L. Scudeller, F. Mesonero, L. Roade, M. C. Ruhleman, M. Schaefer, M. Carrabba, M. Riveiro-Barciela, M. E. F. Bassi, M. G. Valsecchi, M. Hernandez-Tejero, M. Acosta-Herrera, M. D'Angio, M. Baldini, M. Cazzaniga, M. Schulzyk, M. Cecconi, M. Wittig, M. Ciccarelli, M. Rodriguez-Gandia, M. Boccilone, M. Miozzo, N. Montano, N. Braun, N. Sacchi, N. Martinez, O. Ozer, O. Palmieri, P. Favero, P. Pretoni, P. Bonfanti, P. Omodei, P. Tentorio, P. Castro, P. M. Rodrigues, A. B. Ortiz, R. de Cid, R. Ferrer, R. Gualtierotti, R. Nieto, S. Goerg, S. Badalamenti, S. Marsal, G. Matullo, S. Pelusi, S. Juzenas, S. Aliberti, V. Monzani, V. Moreno, T. Wesse, T. L. Lenz, T. Pumarola, V. Rimoldi, S. Bosari, W. Albrecht, W. Peter, M. Romero-Gomez, M. D'Amato, S. Duga, J. M. Banales, J. R. Hov, T. Folseras, L. Valentini, A. Franke, T. H. Karlson, Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).
- J. F. Shelton, A. J. Shastry, C. Ye, C. H. Weldon, T. Filshie-Sonnez, D. Coker, A. Symons, J. Esparza-Gordillo; 23andMe COVID-19 Team, S. Aslibekyan, A. Auton, Trans-ancestry analysis reveals genetic and nongenetic associations with COVID-19 susceptibility and severity. *Nat. Genet.* **53**, 801–808 (2021).
- C. I. van der Made, A. Simons, J. Schuurs-Hoeijmakers, G. van den Heuvel, T. Mantere, S. Kersten, R. C. van Deuren, M. Steehouwer, S. V. van Reijmersdal, M. Jaeger, T. Hofste, G. Astuti, J. C. Galbany, V. van der Schoot, H. van der Hoeven, W. H. O. ten Have, E. Klijn, C. van den Meer, J. Fiddelaers, Q. de Mast, C. P. Bleeker-Rovers, L. A. B. Joosten, H. G. Yntema, C. Gilissen, M. Nelen, J. W. M. van der Meer, H. G. Brunner, M. G. Netea, F. L. van de Veerdonk, A. Hoischen, Presence of genetic variants among young men with severe COVID-19. *J. Am. Med. Assoc.* **324**, 663–673 (2020).
- Q. Zhang, P. Bastard, Z. Liu, J. Le Pen, M. Moncada-Velez, J. Chen, M. Ogishi, I. K. D. Sabli, S. Hodeib, C. Korol, J. Rosain, K. Bilguvar, J. Ye, A. Bolze, B. Bigio, R. Yang, A. A. Arias, Q. Zhou, Y. Zhang, F. Onodi, S. Korniotis, L. Karpf, Q. Philippot, M. Chbibi, L. Bonnet-Madin, K. Dorgham, N. Smith, W. M. Schneider, B. S. Razook, H. H. Hoffmann, E. Michailidis, L. Moens, J. E. Han, L. Lorenzo, L. Bizien, P. Meade, A. L. Neehus, A. C. Ugurbil, A. Corneau, G. Kerner, P. Zhang, F. Rapaport, Y. Seeleunther, J. Manry, C. Masson, Y. Schmitt, A. Schluter, T. Le Voyer, T. Khan, J. Li, J. Fellay, L. Roussel, M. Shahrooei, M. F. Alosaimi, D. Mansouri, H. Al-Saud, F. Al-Mulla, F. Almourfi, S. Z. Al-Muhsen, F. Alsohime, S. Al Turki, R. Hasanato, D. van de Beek, A. Biondi, L. R. Bettini, M. D'Angio, P. Bonfanti, L. Imberti, A. Sottini, S. Paghera, E. Quirios-Roldan, C. Rossi, A. J. Oler, M. F. Tompkins, C. Alba, I. Vandernoot, J. C. Goffard, G. Smits, I. Migeotte, F. Haerynck, P. Soler-Palacin, A. Martin-Nalda, R. Colobran, P. E. Morange, S. Keles, F. Colkesen, T. Ozcelik, K. K. Yasar, S. Senoglu, S. N. Karabala, C. Rodriguez-Gallego, G. Novelli, S. Hraiech, Y. Tandjaoui-Lambiotte, X. Duval, C. Laouenan; COVID-STORM Clinicians; COVID Clinicians; Imagine COVID Group; French COVID Cohort Study Group; CoV-Contact Cohort; Amsterdam UMC Covid-19 Biobank; COVID Human Genetic Effort; NIAID-USUHS/TAGC COVID Immunity Group, A. L. Snow, C. L. Dalgard, J. D. Milner, D. C. Vinh, T. H. Mogensen, N. Marr, A. N. Spaan, B. Boisson, S. Boisson-Dupuis, J. Bustamante, A. Puel, M. J. Ciancanelli, I. Meyts, T. Mariatis, V. Soumelis, A. Amara, M. Nussenzweig, A. Garcia-Sastre, F. Krammer, A. Pujol, D. Duffy, R. P. Lifton, S. Y. Zhang, G. Gorochov, V. Beziat, E. Jouanguy, V. Sancho-Shimizu, C. M. Rice, L. Abel, L. D. Notarangelo, A. Cobat, H. C. Su, J. L. Casanova, Inborn errors of type I IFN immunity in patients with life-threatening COVID-19. *Science* **370**, eabd4570 (2020).
- CDC, Scientific Evidence for Conditions that Increase Risk of Severe Illness | COVID-19 (CDC, 2021).
- A. Silvin, N. Chapuis, G. Dunsmore, A. G. Goubet, A. Dubuisson, L. Derosa, C. Almire, C. Henon, O. Kosmider, N. Droin, P. Rameau, C. Catelain, A. Alfaro, C. Dussiou, C. Friedrich, E. Sourdeau, N. Marin, T. A. Szwebel, D. Cantin, L. Mouthon, D. Borderie, M. Deloger, D. Bredel, S. Mouraud, D. Drubay, M. Andrieu, A. S. Lhonneur, V. Saada, A. Stoclin, C. Willekins, F. Pommeret, F. Griscelli, L. G. Ng, Z. Zhang, P. Bost, I. Amit, F. Barlesi, A. Marabelle, F. Pene, B. Gachot, F. Andre, L. Zitvogel, F. Ginhoux, M. Fontenay, E. Solarly, Elevated calprotectin and abnormal myeloid cell subsets discriminate severe from mild COVID-19. *Cell* **182**, 1401–1418.e18 (2020).
- G. Chen, D. Wu, W. Guo, Y. Cao, D. Huang, H. Wang, T. Wang, X. Zhang, H. Chen, H. Yu, X. Zhang, M. Zhang, S. Wu, J. Song, T. Chen, M. Han, S. Li, X. Luo, J. Zhao, Q. Ning, Clinical and immunological features of severe and moderate coronavirus disease 2019. *J. Clin. Invest.* **130**, 2620–2629 (2020).
- E. J. Giamarellos-Bourboulis, M. G. Netea, N. Rovina, K. Akinosoglou, A. Antoniadou, N. Antonakos, G. Damoraki, T. Gkavogianni, M. E. Adam, P. Katsaounou, M. Ntaganou, M. Kyriakopoulou, G. Dimopoulos, I. Koutsodimitropoulos, D. Velissaris, P. Koufarygiris, A. Karageorgos, K. Katrini, V. Lekakis, M. Lupsse, A. Kotsaki, G. Renieris, D. Theodoulou, V. Panou, E. Koukaki, N. Koulouris, C. Gogos, A. Koutsoukou, Complex immune dysregulation in COVID-19 patients with severe respiratory failure. *Cell Host Microbe* **27**, 992–1000.e3 (2020).
- C. Lucas, P. Wong, J. Klein, T. B. R. Castro, J. Silva, M. Sundaram, M. K. Ellingson, T. Mao, J. E. Oh, B. Israelow, T. Takahashi, M. Tokuyama, P. Lu, A. Venkataraman, A. Park,

- S. Mohanty, H. Wang, A. L. Wyllie, C. B. F. Vogels, R. Earnest, S. Lapidus, I. M. Ott, A. J. Moore, M. C. Muenker, J. B. Fournier, M. Campbell, C. D. Odio, A. Casanovas-Massana, I. T. Yale, R. Herbst, A. C. Shaw, R. Medzhitov, W. L. Schulz, N. D. Grubaugh, C. Dela Cruz, S. Farhadian, A. I. Ko, S. B. Omer, A. Iwasaki, Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* **584**, 463–469 (2020).
12. J. Hadjadj, N. Yatim, L. Barnabei, A. Corneau, J. Boussier, N. Smith, H. Pere, B. Charbit, V. Bondet, C. Chenevier-Gobeaux, P. Breillat, N. Carlier, R. Gauzit, C. Morbieu, F. Pene, N. Marin, N. Roche, T. A. Szewbel, S. H. Merkling, J. M. Treliyer, D. Veyer, L. Mouton, C. Blanc, P. L. Tharaux, F. Rozenberg, A. Fischer, D. Duffy, F. Rieux-Lauca, S. Kerneis, B. Terrier, Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).
  13. S. Trouillet-Assant, S. Viel, A. Gaymard, S. Pons, J. C. Richard, M. Perret, M. Villard, K. Brengel-Pesce, B. Lina, M. Mezidi, L. Bitker, A. Belot; COVID HCL Study group, Type I IFN immunoprofiling in COVID-19 patients. *J. Allergy Clin. Immunol.* **146**, 206–208.e2 (2020).
  14. M. L. Meizlish, A. B. Pine, J. D. Bishai, G. Goshua, E. R. Nadelmann, M. Simonov, C. H. Chang, H. Zhang, M. Shallow, P. Bahel, K. Owusu, Y. Yamamoto, T. Arora, D. S. Atri, A. Patel, R. Gbyli, J. Kwan, C. H. Won, C. Dela Cruz, C. Price, J. Koff, B. A. King, H. M. Rinder, F. P. Wilson, J. Hwa, S. Halene, W. Damsky, D. van Dijk, A. I. Lee, H. J. Chun, A neutrophil activation signature predicts critical illness and mortality in COVID-19. *Blood Adv.* **5**, 1164–1177 (2021).
  15. J. Schulte-Schrepping, N. Reusch, D. Paclik, K. Bassler, S. Schlickeiser, B. Zhang, B. Kramer, T. Krammer, S. Brumhard, L. Bonaguro, E. De Domenico, D. Wendisch, M. Grasshoff, T. S. Kapellos, M. Beckstette, T. Pecht, A. Saglam, O. Dietrich, H. E. Mei, A. R. Schulz, C. Conrad, D. Kunkel, E. Vafadarnejad, C. J. Xu, A. Horne, M. Herbert, A. Drews, C. Thibeault, M. Pfeiffer, S. Huppenstiel, A. Hocke, H. Muller-Redetzky, K. M. Heim, F. Machleidt, A. Uhrig, L. B. de Jarcy, L. Jurgens, M. Stegemann, C. R. Glosenkamp, H. D. Volk, C. Goftinet, M. Landthaler, E. Wyler, P. Georg, M. Schneider, C. Dang-Heine, N. Neuwinger, K. Kappert, R. Tauber, V. Corman, J. Raabe, K. M. Kaiser, M. T. Vinh, G. Rieke, C. Meisel, T. Ulas, M. Becker, R. Geffers, M. Witzenrath, C. Drosten, N. Suttorp, C. von Kalle, F. Kurth, K. Handler, J. L. Schultz, A. C. Aschenbrenner, Y. Li, J. Nattermann, B. Sawitzki, A. E. Saliba, L. E. Sander; Deutsche COVID-19 OMICS Initiative (DeCOI), Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell* **182**, 1419–1440.e23 (2020).
  16. B. Shen, X. Yi, Y. Sun, X. Bi, J. Du, C. Zhang, S. Quan, F. Zhang, R. Sun, L. Qian, W. Ge, W. Liu, S. Liang, H. Chen, Y. Zhang, J. Li, J. Xu, Z. He, B. Chen, J. Wang, H. Yan, Y. Zheng, D. Wang, J. Zhu, Z. Kong, Z. Kang, X. Liang, X. Ding, G. Ruan, N. Xiang, X. Cai, H. Gao, L. Li, S. Li, Q. Xiao, T. Lu, Y. Zhu, H. Liu, H. Chen, T. Guo, Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell* **182**, 59–72.e15 (2020).
  17. S. De Biasi, D. Lo Tartaro, M. Meschiari, L. Gibellini, C. Bellinazzi, R. Borella, L. Fidanza, M. Mattioli, A. Paolini, L. Gozzi, D. Jaacoub, M. Faltoni, S. Volpi, J. Milic, M. Sita, M. Sarti, C. Pucillo, M. Girardis, G. Guaraldi, C. Mussini, A. Cossarizza, Expansion of plasmablasts and loss of memory B cells in peripheral blood from COVID-19 patients with pneumonia. *Eur. J. Immunol.* **50**, 1283–1294 (2020).
  18. S. De Biasi, M. Meschiari, L. Gibellini, C. Bellinazzi, R. Borella, L. Fidanza, L. Gozzi, A. Iannone, D. Lo Tartaro, M. Mattioli, A. Paolini, M. Menozzi, J. Milic, G. Franceschi, R. Fantini, R. Tonelli, M. Sita, M. Sarti, T. Trenti, L. Brugioni, L. Cicchetti, F. Facchinetti, A. Pietrangelo, E. Clini, M. Girardis, G. Guaraldi, C. Mussini, A. Cossarizza, Marked T cell activation, senescence, exhaustion and skewing towards TH17 in patients with COVID-19 pneumonia. *Nat. Commun.* **11**, 3434 (2020).
  19. I. Odak, J. Barros-Martins, B. Bosnjak, K. Stahl, S. David, O. Wiesner, M. Busch, M. M. Hoeper, I. Pink, T. Welte, M. Cornberg, M. Stoll, L. Goudeva, R. Blasczyk, A. Ganser, I. Prinz, R. Fo, C. Koenecke, C. R. Schultz-Florey, Reappearance of effector T cells is associated with recovery from COVID-19. *EBO Medicine* **57**, 102885 (2020).
  20. J. Helms, C. Tacquard, F. Severac, I. Leonard-Lorant, M. Ohana, X. Delabranche, H. Merdji, R. Clerc-Jehl, M. Schenck, F. Fagot Gonet, S. Fafi-Kremer, V. Castelain, F. Schneider, L. Grunebaum, E. Angles-Cano, L. Sattler, P. M. Mertes, F. Meziani; CRICS TRIGGERSEP Group (Clinical Research in Intensive Care and Sepsis Trial Group for Global Evaluation and Research in Sepsis), High risk of thrombosis in patients with severe SARS-CoV-2 infection: A multicenter prospective cohort study. *Intensive Care Med.* **46**, 1089–1098 (2020).
  21. F. A. Klok, M. Kruip, N. J. M. van der Meer, M. S. Arbous, D. Gommers, K. M. Kant, F. H. J. Kaptein, J. van Paassen, M. A. M. Stals, M. V. Huisman, H. Endeman, Incidence of thrombotic complications in critically ill ICU patients with COVID-19. *Thromb. Res.* **191**, 145–147 (2020).
  22. F. A. Klok, M. Kruip, N. J. M. van der Meer, M. S. Arbous, D. Gommers, K. M. Kant, F. H. J. Kaptein, J. van Paassen, M. A. M. Stals, M. V. Huisman, H. Endeman, Confirmation of the high cumulative incidence of thrombotic complications in critically ill ICU patients with COVID-19: An updated analysis. *Thromb. Res.* **191**, 148–150 (2020).
  23. A. D. T. Force, V. M. Ranieri, G. D. Rubenfeld, B. T. Thompson, N. D. Ferguson, E. Caldwell, E. Fan, L. Camporota, A. S. Slutsky, Acute respiratory distress syndrome: The Berlin definition. *JAMA* **307**, 2526–2533 (2012).
  24. A. S. Rudberg, S. Havervall, A. Manberg, A. Jernbom Falk, K. Aguilera, H. Ng, L. Gabrielsson, A. C. Salomonsson, L. Hanke, B. Murrell, G. McInerney, J. Olofsson, E. Andersson, C. Hellstrom, S. Bayati, S. Bergstrom, E. Pin, R. Sjoberg, H. Tegel, M. Hedhammar, M. Phillipson, P. Nilsson, S. Hober, C. Thalin, SARS-CoV-2 exposure, symptoms and seroprevalence in healthcare workers in Sweden. *Nat. Commun.* **11**, 5064 (2020).
  25. P. Mehta, D. F. McAuley, M. Brown, E. Sanchez, R. S. Tattersall, J. J. Manson; HLH Across Speciality Collaboration, UK, COVID-19: Consider cytokine storm syndromes and immunosuppression. *Lancet* **395**, 1033–1034 (2020).
  26. W.-J. Guan, Z.-y. Ni, Y. Hu, W.-h. Liang, C.-q. Ou, J.-x. He, L. Liu, H. Shan, C.-l. Lei, D. S. C. Hui, B. Du, L.-j. Li, G. Zeng, K.-Y. Yuan, R.-c. Chen, C.-l. Tang, T. Wang, P.-y. Chen, J. Xiang, S.-y. Li, J.-l. Wang, Z.-j. Liang, Y.-x. Peng, L. Wei, Y. Liu, Y.-h. Hu, P. Peng, J.-m. Wang, J.-y. Liu, Z. Chen, G. Li, Z.-j. Zheng, S.-q. Qiu, J. Luo, C.-j. Ye, S.-y. Zhu, N.-s. Zhong; China Medical Treatment Expert Group for Covid-19, Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).
  27. C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
  28. D. E. Leisman, L. Ronner, R. Pinotti, M. D. Taylor, P. Sinha, C. S. Calfee, A. V. Hirayama, F. Mastroianni, C. J. Turtle, M. O. Harhay, M. Legrand, C. S. Deutschman, Cytokine elevation in severe and critical COVID-19: A rapid systematic review, meta-analysis, and comparison with other inflammatory syndromes. *Lancet Respir. Med.* **8**, 1233–1244 (2020).
  29. K. E. Remy, M. Mazer, D. A. Striker, A. H. Ellebedy, A. H. Walton, J. Unsinger, T. M. Blood, P. A. Mudd, D. J. Yi, D. A. Mannion, D. F. Osborne, R. S. Martin, N. J. Anand, J. P. Bosanquet, J. Blood, A. M. Drewry, C. C. Caldwell, I. R. Turnbull, S. C. Brakenridge, L. L. Moldawer, R. S. Hotchkiss, Severe immunosuppression and not a cytokine storm characterizes COVID-19 infections. *JCI Insight* **5**, e140329 (2020).
  30. P. Bastard, E. Orlova, L. Sozaeva, R. Levy, A. James, M. M. Schmitt, S. Ochoa, M. Kareva, Y. Rodina, A. Gervais, T. Le Voyer, J. Rosain, Q. Philippot, A. L. Neehus, E. Shaw, M. Migaud, L. Bizein, O. Ekwall, S. Berg, G. Beccuti, L. Ghizzoni, G. Thiriez, A. Pavot, C. Goujard, M. L. Fremond, E. Carter, A. Rothenbuhler, A. Linglart, B. Mignot, A. Comte, N. Cheikh, O. Hermine, L. Breivik, E. S. Husebye, S. Humbert, P. Rohrlich, A. Coaquette, F. Vuoto, K. Faure, N. Mahlaoui, P. Kotnik, T. Battelino, K. T. Podkrajsek, K. Kisand, E. M. N. Ferre, T. DiMaggio, L. B. Rosen, P. D. Burbelo, M. McIntrye, N. Y. Kann, A. Shcherbina, M. Pavlova, A. Kolodkina, S. M. Holland, S. Y. Zhang, Y. J. Crow, L. D. Notarangelo, H. C. Su, L. Abel, M. S. Anderson, E. Jouanguy, B. Neven, A. Puel, J. L. Casanova, M. S. Lionakis, Preexisting autoantibodies to type I IFNs underlie critical COVID-19 pneumonia in patients with APS-1. *J. Exp. Med.* **218**, e20210554 (2021).
  31. P. Bastard, L. B. Rosen, Q. Zhang, E. Michailidis, H. H. Hoffmann, Y. Zhang, K. Dorgham, Q. Philippot, J. Rosain, V. Beziat, J. Manry, E. Shaw, L. Haljasmagi, P. Peterson, L. Lorenzo, L. Bizein, S. Trouillet-Assant, K. Dobbs, A. A. de Jesus, A. Belot, A. Kallaste, E. Catherinot, Y. Tandjaoui-Lambotte, J. Le Pen, G. Kerner, B. Bigio, Y. Seeleuthner, R. Yang, A. Bolze, A. N. Spaan, O. M. Delmonte, M. S. Abers, A. Auti, G. Casari, V. Lampasona, L. Piemonti, F. Ciceri, K. Bilguvar, R. P. Lifton, M. Vasse, D. M. Smadja, M. Migaud, J. Hadjadj, B. Terrier, D. Duffy, L. Quintana-Murci, D. van de Beek, L. Roussel, D. C. Vinh, S. G. Tangye, F. Haerynk, D. Dalmau, J. Martinez-Picado, P. Brodin, M. C. Nussenzweig, S. Boisson-Dupuis, C. Rodriguez-Gallego, G. Vogt, T. H. Mogensen, A. J. Oler, J. Gu, P. D. Burbelo, J. I. Cohen, A. Biondi, L. R. Bettini, M. D'Angio, P. Bonfanti, P. Rossignol, J. Mayaux, F. Rieux-Lauca, E. S. Husebye, F. Fusco, M. V. Ursini, L. Imberti, A. Sottini, S. Paghera, E. Quiros-Roldan, C. Rossi, R. Castagnoli, D. Montagna, A. Licari, G. L. Marseglia, X. Duval, J. Ghosn; HGID Lab; NIAID-USUHS Immune Response to COVID Group; COVID Clinicians; COVID-STORM Clinicians; Imagine COVID Group; French COVID Cohort Study Group; Milieu Intérieur Consortium; CoV-Contact Cohort; Amsterdam UMC Covid-19 Biobank; COVID Human Genetic Effort, J. S. Tsang, R. Goldbach-Mansky, K. Kisand, M. S. Lionakis, A. Puel, S. Y. Zhang, S. M. Holland, G. Gorochov, E. Jouanguy, C. M. Rice, A. Cobat, L. D. Notarangelo, L. Abel, H. C. Su, J. L. Casanova, Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* **370**, eabd4585 (2020).
  32. L. Chen, X. Long, Q. Xu, J. Tan, G. Wang, Y. Cao, J. Wei, H. Luo, H. Zhu, L. Huang, F. Meng, L. Huang, N. Wang, X. Zhou, L. Zhao, X. Chen, Z. Mao, C. Chen, Z. Li, Z. Sun, J. Zhao, D. Wang, G. Huang, W. Wang, J. Zhou, Elevated serum levels of S100A8/A9 and HMGB1 at hospital admission are correlated with inferior clinical outcomes in COVID-19 patients. *Cell. Mol. Immunol.* **17**, 992–994 (2020).
  33. N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, L. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *Lancet* **395**, 507–513 (2020).
  34. T. Shu, W. Ning, D. Wu, J. Xu, Q. Han, M. Huang, X. Zou, Q. Yang, Y. Yuan, Y. Bie, S. Pan, J. Mu, Y. Han, X. Yang, H. Zhou, R. Li, Y. Ren, X. Chen, S. Yao, Y. Qiu, D. Y. Zhang, Y. Xue, Y. Shang, X. Zhou, Plasma proteomics identify biomarkers and pathogenesis of COVID-19. *Immunity* **53**, 1108–1122.e5 (2020).
  35. H. Al-Samkari, R. S. K. Leaf, W. H. Dzik, J. C. T. Carlson, A. E. Fogerty, A. Waheed, K. Goodarzi, P. K. Bendapudi, L. Bornikova, S. Gupta, D. E. Leaf, D. J. Kuter, R. P. Rosovsky,

- COVID-19 and coagulation: Bleeding and thrombotic manifestations of SARS-CoV-2 infection. *Blood* **136**, 489–500 (2020).
36. P. Yu, K. Wilhelm, A. Dubrac, J. K. Tung, T. C. Alves, J. S. Fang, Y. Xie, J. Zhu, Z. Chen, F. De Smet, J. Zhang, S. W. Jin, L. Sun, H. Sun, R. G. Kibbey, K. K. Hirschi, N. Hay, P. Carmeliet, T. W. Chittenden, A. Eichmann, M. Potente, M. Simons, FGF-dependent metabolic control of vascular development. *Nature* **545**, 224–228 (2017).
  37. J. Li, M. Guo, X. Tian, X. Wang, X. Yang, P. Wu, C. Liu, Z. Xiao, Y. Qu, Y. Yin, C. Wang, Y. Zhang, Z. Zhu, Z. Liu, C. Peng, T. Zhu, Q. Liang, Virus-host interactome and proteomic survey reveal potential virulence factors influencing SARS-CoV-2 pathogenesis. *Med* **2**, 99–112.e7 (2021).
  38. I. Sanchez-Cerrillo, P. Landete, B. Aldave, S. Sanchez-Alonso, A. Sanchez-Azofra, A. Marcos-Jimenez, E. Avalos, A. Alcaraz-Serna, I. de Los Santos, T. Mateu-Albero, L. Esparcia, C. Lopez-Sanz, P. Martinez-Fleta, L. Gabrie, L. D. C. Guerola, H. de la Fuente, M. J. Calzada, I. Gonzalez-Alvaro, A. Alfranca, F. Sanchez-Madrid, C. Munoz-Calleja, J. B. Soriano, J. Ancochea, E. Martin-Gayo; REINMUN-COVID and EDEPIMIC groups, COVID-19 severity associates with pulmonary redistribution of CD1c+ DCs and inflammatory transitional and nonclassical monocytes. *J. Clin. Invest.* **130**, 6290–6300 (2020).
  39. R. Y. Li, S. Gujja, S. R. Bajaj, O. E. Gamel, N. Cilfone, J. R. Gulcher, D. A. Lidar, T. W. Chittenden, Quantum processor-inspired machine learning in the biomedical sciences. *Patterns* **2**, 100246 (2021).
  40. N. Ricard, R. P. Scott, C. J. Booth, H. Velazquez, N. A. Cilfone, J. L. Baylon, J. R. Gulcher, S. E. Quaggin, T. W. Chittenden, M. Simons, Endothelial ERK1/2 signaling maintains integrity of the quiescent endothelium. *J. Exp. Med.* **216**, 1874–1890 (2019).
  41. D. E. Gordon, J. Hiatt, M. Bouhaddou, V. V. Rezelj, S. Ulferts, H. Braberg, A. S. Jureka, K. Obernier, J. Z. Guo, J. Batra, R. M. Kaake, A. R. Weckstein, T. W. Owens, M. Gupta, S. Pourmal, E. W. Titus, M. Cakir, M. Soucheray, M. McGregor, Z. Cakir, G. Jang, M. J. O'Meara, T. A. Tummino, Z. Zhang, H. Foussard, A. Rojc, Y. Zhou, D. Kuchenov, R. Huttenhain, J. Xu, M. Eckhardt, D. L. Swaney, J. M. Fabius, M. Ummadi, B. Tutuncuoglu, U. Rathore, M. Modak, P. Haas, K. M. Haas, Z. Z. C. Naing, E. H. Pulido, Y. Shi, I. Barrio-Hernandez, D. Memon, E. Petsalaki, A. Dunham, M. C. Marrero, D. Burke, C. Koh, T. Vallet, J. A. Silvas, C. M. Azumaya, C. Billesbolle, A. F. Brilot, M. G. Campbell, A. Diallo, M. S. Dickinson, D. Diwanji, N. Herrera, N. Hoppe, H. T. Kratohvil, Y. Liu, G. E. Merz, M. Moritz, H. C. Nguyen, C. Nowotny, C. Puchades, A. N. Rizo, U. Schulze-Gahmen, A. M. Smith, M. Sun, I. D. Young, J. Zhao, D. Asarnow, J. Biel, A. Bowen, J. R. Braxton, J. Chen, C. M. Chio, U. S. Chio, I. Deshpande, L. Doan, B. Faust, S. Flores, M. Jin, K. Kim, V. L. Lam, F. Li, J. Li, Y. L. Li, Y. Li, X. Liu, M. Lo, K. E. Lopez, A. A. Melo, F. R. Moss III, P. Nguyen, J. Paulino, K. I. Pawar, J. K. Peters, T. H. Pospiech Jr., M. Safari, S. Sangwan, K. Schaefer, P. V. Thomas, A. C. Thwin, R. Trenker, E. Tse, T. K. M. Tsui, F. Wang, N. Whitis, Z. Yu, K. Zhang, Y. Zhang, F. Zhou, D. Saltzberg; QCRG Structural Biology Consortium, A. J. Hodder, A. S. Shun-Shion, D. M. Williams, K. M. White, R. Rosales, T. Kehrer, L. Miorin, E. Moreno, A. H. Patel, S. Rihn, M. M. Khalid, A. Vallejo-Gracia, P. Fozouni, C. R. Simoneau, T. L. Roth, D. Wu, M. A. Karim, M. Ghousaini, I. Dunham, F. Berardi, S. Weigang, M. Chazal, J. Park, J. Logue, M. McGrath, S. Weston, R. Haupt, C. J. Hastie, M. Elliott, F. Brown, K. A. Burness, E. Reid, M. Dorward, C. Johnson, S. G. Wilkinson, A. Geyer, D. M. Giesel, C. Baillie, S. Raggett, H. Leech, R. Toth, N. Goodman, K. C. Keough, A. L. Lind, C. Zoonomia, R. J. Klesh, K. R. Hemphill, J. Carlson-Stevermer, J. Oki, K. Holden, T. Maures, K. S. Pollard, A. Sali, D. A. Agard, Y. Cheng, J. S. Fraser, A. Frost, N. Jura, T. Kortemme, A. Manglik, D. R. Southworth, R. M. Stroud, D. R. Alessi, P. Davies, M. B. Frieman, T. Ideker, C. Abate, N. Jouvenet, G. Kochs, B. Shoichet, M. Ott, M. Palmarini, K. M. Shokat, A. Garcia-Sastre, J. A. Rassen, R. Grosse, O. S. Rosenberg, K. A. Verba, C. F. Basler, M. Vignuzzi, A. A. Peden, P. Beltrao, N. J. Krogan, Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370**, eabe9403 (2020).
  42. D. E. Gordon, G. M. Jang, M. Bouhaddou, J. Xu, K. Obernier, K. M. White, M. J. O'Meara, V. V. Rezelj, J. Z. Guo, D. L. Swaney, T. A. Tummino, R. Huttenhain, R. M. Kaake, A. L. Richards, B. Tutuncuoglu, H. Foussard, J. Batra, K. Haas, M. Modak, M. Kim, P. Haas, B. J. Polacco, H. Braberg, J. M. Fabius, M. Eckhardt, M. Soucheray, M. J. Bennett, M. Cakir, M. J. McGregor, Q. Li, B. Meyer, F. Roesch, T. Vallet, A. M. Kain, L. Miorin, E. Moreno, Z. Z. C. Naing, Y. Zhou, S. Peng, Y. Shi, Z. Zhang, W. Shen, I. T. Kirby, J. E. Melnyk, J. S. Chorba, K. Lou, S. A. Dai, I. Barrio-Hernandez, D. Memon, C. Hernandez-Armenta, J. Lyu, C. J. P. Matthy, T. Perica, K. B. Pillai, S. J. Ganeshan, D. J. Saltzberg, R. Rakesh, X. Liu, S. B. Rosenthal, L. Calviello, S. Venkataraman, J. Liboy-Lugo, Y. Lin, X. P. Huang, Y. Liu, S. A. Wankowicz, M. Bohn, M. Safari, F. S. Ugur, C. Koh, N. S. Savar, Q. D. Tran, D. Shengjuler, S. J. Fletcher, M. C. O'Neal, Y. Cai, J. C. J. Chang, D. J. Broadhurst, S. Klippsten, P. P. Sharp, N. A. Wenzell, D. Kuzuoglu-Ozturk, H. Y. Wang, R. Trenker, J. M. Young, D. A. Cavero, J. Hiatt, T. L. Roth, U. Rathore, A. Subramanian, J. Noack, M. Hubert, R. M. Stroud, A. D. Frankel, O. S. Rosenberg, K. A. Verba, D. A. Agard, M. Ott, M. Emerman, N. Jura, M. von Zastrow, E. Verdin, A. Ashworth, O. Schwartz, C. d'Enfert, S. Mukherjee, M. Jacobson, H. S. Malik, D. G. Fujimori, T. Ideker, C. S. Craik, S. N. Floor, J. S. Fraser, J. D. Gross, A. Sali, B. L. Roth, D. Ruggero, J. Taunton, T. Kortemme, P. Beltrao, M. Vignuzzi, A. Garcia-Sastre, K. M. Shokat, B. K. Shoichet, N. J. Krogan, A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
  43. L. E. Bazzone, M. King, C. R. MacKay, P. P. Kyaw, P. Meraner, D. Lindstrom, J. Rojas-Quintero, C. A. Owen, J. P. Wang, A. L. Brass, E. A. Kurt-Jones, R. W. Finberg, A disintegrin and metalloproteinase 9 domain (ADAM9) is a major susceptibility factor in the early stages of encephalomyocarditis virus infection. *MBio* **10**, e02734-18 (2019).
  44. C. W. Chou, Y. K. Huang, T. T. Kuo, J. P. Liu, Y. P. Sher, An overview of ADAM9: Structure, activation, and regulation in human diseases. *Int. J. Mol. Sci.* **21**, 7790 (2020).
  45. P. S. Arunachalam, F. Wimmers, C. K. P. Mok, R. A. P. M. Perera, M. Scott, T. Hagan, N. Sigal, Y. P. Feng, L. Bristow, O. T. Y. Tsang, D. Wagh, J. Collier, K. L. Pellegrini, D. Kazmin, G. Alaeddine, W. S. Leung, J. M. C. Chan, T. S. H. Chik, C. Y. C. Choi, C. Huerta, M. P. McCullough, H. B. Lv, E. Anderson, S. Edupuganti, A. A. Upadhyay, S. E. Bosinger, H. T. Maecker, P. Khatri, N. Rouphael, M. Peiris, B. Pulendran, Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* **369**, 1210–1220 (2020).
  46. R. Carapito, S. Bahram, Genetics, genomics, and evolutionary biology of NKG2D ligands. *Immunol. Rev.* **267**, 88–116 (2015).
  47. K. Kohga, T. Takehara, T. Tatsumi, H. Ishida, T. Miyagi, A. Hosui, N. Hayashi, Sorafenib inhibits the shedding of major histocompatibility complex class I-related chain A on hepatocellular carcinoma cells by down-regulating a disintegrin and metalloproteinase 9. *Hepatology* **51**, 1264–1273 (2010).
  48. B. S. Pedersen, A. R. Quinlan, Who's who? Detecting and resolving sample anomalies in human DNA sequencing studies with Peddy. *Am. J. Hum. Genet.* **100**, 406–413 (2017).
  49. J. Buchrieser, J. Dufloo, M. Hubert, B. Monel, D. Planas, M. M. Rajah, C. Planchais, F. Porrot, F. Guivel-Benhassine, S. Van der Werf, N. Casartelli, H. Mouquet, T. Bruel, O. Schwartz, Syncytia formation by SARS-CoV-2-infected cells. *EMBO J.* **39**, e106267 (2020).
  50. R. L. Chua, S. Lukassen, S. Trump, B. P. Hennig, D. Wendisch, F. Pott, O. Debnath, L. Thurmann, F. Kurth, M. T. Volker, J. Kazmierski, B. Timmermann, S. Twardziok, S. Schneider, F. Machleidt, H. Muller-Redetzky, M. Maier, A. Krannich, S. Schmidt, F. Balzer, J. Liebig, J. Loske, N. Suttorp, J. Eils, N. Ishaque, U. G. Liebert, C. von Kalle, A. Hocke, M. Witzenrath, C. Goffinet, D. Drosten, S. Laudi, I. Lehmann, C. Conrad, L. E. Sander, R. Eils, COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* **38**, 970–979 (2020).
  51. C. B. Messner, V. Demichev, D. Wendisch, L. Michalick, M. White, A. Freivald, K. Textoris-Taube, S. I. Vernardis, A. S. Egger, M. Kreidl, D. Ludwig, C. Kilian, F. Agostini, A. Zeleznik, C. Thibeault, M. Pfeiffer, S. Hippchenst, A. Hocke, C. von Kalle, A. Campbell, C. Hayward, D. J. Porteous, R. E. Marioni, C. Langenberg, K. S. Lilley, W. M. Kuebler, M. Mulleder, C. Drosten, N. Suttorp, M. Witzenrath, F. Kurth, L. E. Sander, M. Ralsar, Ultra-high-throughput clinical proteomics reveals classifiers of COVID-19 infection. *Cell Syst.* **11**, 11–24.e4 (2020).
  52. Y. P. Su, D. Chen, D. Yuan, C. Lausted, J. Choi, C. Z. L. Dai, V. Voillet, V. R. Duvvuri, K. Scherler, P. Troisch, P. Baloni, G. R. Qin, B. Smith, S. A. Kornilov, C. Rostomily, A. Xu, J. Li, S. Dong, A. Rothchild, J. Zhou, K. Murray, R. Edmark, S. G. Hong, J. E. Heath, J. Earls, R. Y. Zhang, J. Y. Xie, S. Li, R. Roper, L. Jones, Y. Zhou, L. Rowen, R. Liu, S. Mackay, D. S. O'Mahony, C. R. Dale, J. A. Wallick, H. A. Algren, M. A. Zager, W. Wei, N. D. Price, S. Huang, N. Subramanian, K. Wang, A. T. Magis, J. J. Hadlock, L. Hood, A. Aderen, J. A. Bluestone, L. L. Lanier, P. D. Greenberg, R. Gottardo, M. M. Davis, J. D. Goldman, J. R. Heath, Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19. *Cell* **183**, 1479–1495.e20 (2020).
  53. L. L. Wei, W. J. Wang, D. X. Chen, B. Xu, Dysregulation of the immune response affects the outcome of critical COVID-19 patients. *J. Med. Virol.* **92**, 2768–2776 (2020).
  54. Z. Zhou, L. Ren, L. Zhang, J. Zhong, Y. Xiao, Z. Jia, L. Guo, J. Yang, C. Wang, S. Jiang, D. Yang, G. Zhang, H. Li, F. Chen, Y. Xu, M. Chen, Z. Gao, J. Yang, J. Dong, B. Liu, X. Zhang, W. Wang, K. He, Q. Jin, M. Li, J. Wang, Overly exuberant innate immune response to SARS-CoV-2 infection. *SSRN Electron. J.* (2020).
  55. J. P. A. Ioannidis, C. Axfors, D. G. Contopoulos-Ioannidis, Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters. *Environ. Res.* **188**, 109890 (2020).
  56. X. Li, T. Marmar, Q. Xu, J. Tu, Y. Yin, Q. Tao, H. Chen, T. Shen, D. Xu, Predictive indicators of severe COVID-19 independent of comorbidities and advanced age: A nested case-control study. *Epidemiol. Infect.* **148**, e255 (2020).
  57. P. Wang, J. Sha, M. Meng, C. Y. Wang, Q. C. Yao, Z. F. Zhang, W. Q. Sun, X. G. Wang, G. Q. Qie, X. Bai, K. K. Liu, Y. F. Chu, Risk factors for severe COVID-19 in middle-aged patients without comorbidities: A multicentre retrospective study. *J. Transl. Med.* **18**, 461 (2020).
  58. B.-C.-U. Brasileiras. (accessed on 7/6/2021), vol. 2021; <http://www.utisbrasileiras.com.br/en/covid-19/benchmarking-covid-19/>.
  59. N. G. Davies, C. I. Jarvis; C. C.-W. Group, W. J. Edmunds, N. P. Jewell, K. Diaz-Ordaz, R. H. Keogh, Increased mortality in community-tested cases of SARS-CoV-2 lineage B.1.1.7. *Nature* **593**, 270–274 (2021).

60. CORIMUNO-19 Collaborative group, Effect of anakinra versus usual care in adults in hospital with COVID-19 and mild-to-moderate pneumonia (CORIMUNO-ANA-1): A randomised controlled trial. *Lancet Respir. Med.* **9**, 295–304 (2021).
61. O. Hermine, X. Mariette, P. L. Tharaux, M. Resche-Rigon, R. Porcher, P. Ravaud; CORIMUNO-19 Collaborative Group, Effect of Tocilizumab vs Usual care in adults hospitalized with COVID-19 and moderate or severe pneumonia: A randomized clinical trial. *JAMA Intern. Med.* **181**, 32–40 (2021).
62. E. J. Rubin, D. L. Longo, L. R. Baden, Interleukin-6 receptor inhibition in Covid-19—Cooling the inflammatory soup. *N. Engl. J. Med.* **384**, 1564–1565 (2021).
63. Z. Liu, E. Xu, H. T. Zhao, T. Cole, A. B. West, LRRK2 and Rab10 coordinate macropinocytosis to mediate immunological responses in phagocytes. *EMBO J.* **39**, e104862 (2020).
64. O. O. Glebov, Understanding SARS-CoV-2 endocytosis for COVID-19 drug repurposing. *FEBS J.* **287**, 3664–3671 (2020).
65. K. Li, S. W. Wang, Y. Li, R. E. Martin, L. Li, M. Lu, S. E. Lamhamdi-Cherradi, G. Hu, S. Demissie-Sanders, J. Zheng, F. Chung, T. Oates, Z. Yao, Identification and expression of a new type II transmembrane protein in human mast cells. *Genomics* **86**, 68–75 (2005).
66. W. Xie, L. Chen, L. Chen, Q. Kou, Silencing of long non-coding RNA MALAT1 suppresses inflammation in septic mice: Role of microRNA-23a in the down-regulation of MCEMP1 expression. *Inflamm. Res.* **69**, 179–190 (2020).
67. R. Sanyal, M. J. Polyak, J. Zuccollo, M. Puri, L. Deng, L. Roberts, A. Zuba, J. Storek, J. M. Luider, E. M. Sundberg, A. Mansoor, E. Baigorri, M. P. Chu, A. R. Belch, L. M. Pilarski, J. P. Deans, MS4A4A: A novel cell surface marker for M2 macrophages and plasma cells. *Immunol. Cell Biol.* **95**, 611–619 (2017).
68. Y. Sui, W. Zeng, MS4A4A regulates arginase 1 induction during macrophage polarization and lung inflammation in mice. *Eur. J. Immunol.* **50**, 1602–1605 (2020).
69. R. Roychaudhuri, A. H. Hergueter, F. Polverino, M. E. Lauchoo-Contreras, K. Gupta, N. Borregaard, C. A. Owen, ADAM9 is a novel product of polymorphonuclear neutrophils: Regulation of expression and contributions to extracellular matrix protein degradation during acute lung injury. *J. Immunol.* **193**, 2469–2482 (2014).
70. W. Dejnirattisai, D. Zhou, P. Supasa, C. Liu, A. J. Mentzer, H. M. Ginn, Y. Zhao, H. M. E. Duyvesteyn, A. Tuekprakhon, R. Nutalai, B. Wang, C. Lopez-Camacho, J. Slon-Campos, T. S. Walter, D. Skelly, S. A. C. Clemens, F. G. Naveca, V. Nascimento, F. Nascimento, C. F. da Costa, P. C. Resende, A. Pauvold-Correia, M. M. Siqueira, C. Dold, R. Levin, T. Dong, A. J. Pollard, J. C. Knight, D. Crook, T. Lambe, E. Clutterbuck, S. Bibi, A. Flaxman, M. Bittaye, S. Belij-Rammerstorfer, S. C. Gilbert, M. W. Carroll, P. Klenerman, E. Barnes, S. J. Dunachie, N. G. Paterson, M. A. Williams, D. R. Hall, R. J. G. Hulswit, T. A. Bowden, E. E. Fry, J. Mongkolsapaya, J. Ren, D. I. Stuart, G. R. Scream, Antibody evasion by the P.1 strain of SARS-CoV-2. *Cell* **184**, 2939–2945.e9 (2021).
71. S. Hicks, D. Loo, K. Sinkevicius, J. Scribner, B. Barat, N. Yoder, C. Espelin, M. Themelis, F. Chen, J. Lucas, J. Brown, B. Matin, M. Fuller, J. Lee, P. Salomon, J. Costopulos, S. Yancey, G. Diedrich, S. Gorlatov, T. Son, M. Chiechi, P. Li, M. Spliedt, V. Ciccarone, J. Hooley, N. Gant, J. Tamura, K. Donahue, P. Moore, S. Johnson, T. Chittenden, R. Gregory, E. Bonvini, in *AACR Annual Meeting 2019* (American Association for Cancer Research, 2019), vol. 79, pp. 1533.
72. Institut Pasteur, "Protocol: Real-time RT-PCR assays for the detection of SARS-CoV-2" (WHO, 2021).
73. E. W. Deutsch, A. Csordas, Z. Sun, A. Jarnuzak, Y. Perez-Riverol, T. Ternent, D. S. Campbell, M. Bernal-Llinares, S. Okuda, S. Kawano, R. L. Moritz, J. J. Carver, M. Wang, Y. Ishihama, N. Bandeira, H. Hermjakob, J. A. Vizcaino, The ProteomeXchange Consortium in 2017: Supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* **45**, D1100–D1106 (2017).
74. W. Huber, A. von Heydebreck, H. Sultmann, A. Poustka, M. Vingron, Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18** (Suppl 1), S96–S104 (2002).
75. M. Kokka, J. Virtanen, M. Kohlmainen, J. Paananen, K. Hanhineva, Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: A comparative study. *BMC Bioinformatics* **20**, 492 (2019).
76. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
77. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
78. K. I. Kendig, S. Baheti, M. A. Bockol, T. M. Drucker, S. N. Hart, J. R. Heldenbrand, M. Hernaez, M. E. Hudson, M. T. Kalmbach, E. W. Klee, N. R. Mattson, C. A. Ross, M. Taschuk, E. D. Wieben, M. Wiepert, D. E. Wildman, L. S. Mainzer, Sentieon DNaseq variant calling workflow demonstrates strong computational performance and accuracy. *Front. Genet.* **10**, 736 (2019).
79. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. Del Angel, M. A. Rivas, M. Hanna, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
80. G. A. Van der Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, M. A. DePristo, From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33 (2013).
81. A. A. Shabalin, Matrix eQTL: Ultra fast eQTL analysis via large matrix operations. *Bioinformatics* **28**, 1353–1358 (2012).
82. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
83. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).
84. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
85. M. D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
86. M. Scutari, Learning Bayesian networks with the bnlearn R package. *J. Stat. Softw.* **35**, 1–22 (2010).
87. R. Carapito, J. E. Gotenberg, I. Kotova, M. Untrau, S. Michel, L. Naegeli, I. Aouadi, M. Kwemou, N. Paul, A. Pichot, J. Locke, S. J. Bowman, B. Griffiths, K. L. Sivils, J. Sibilia, H. Inoko, C. Micelli-Richard, G. Nocturne, M. Ota, W. F. Ng, X. Mariette, S. Bahram, A new MHC-linked susceptibility locus for primary Sjögren's syndrome: MIC4. *Hum. Mol. Genet.* **26**, 2565–2576 (2017).
88. R. Tibshirani, Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. B Methodol.* **58**, 267–288 (1996).
89. A. E. Hoerl, R. W. Kennard, K. F. Baldwin, Ridge regression: Some simulations. *Commun. Stat. Theory Methods* **4**, 105–123 (1975).
90. A. E. Hoerl, R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems. *Dent. Tech.* **12**, 55–67 (1970).
91. M. Kuhn, Building predictive models in R Using the caret Package. *J. Stat. Softw.* **28**, 1–26 (2008).
92. B. E. Boser, I. M. Guyon, V. N. Vapnik, in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* (ACM, 1992), pp. 144–152.
93. C. Cortes, V. Vapnik, Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).
94. L. Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001).
95. L. Breiman, J. Friedman, C. J. Stone, R. A. Olshen, *Classification and Regression Trees* (Chapman & Hall, 1993), pp. 358–358.
96. T. Q. Chen, C. Guestrin, XGBoost: A Scalable Tree Boosting System, in *Kdd'16: Proceedings of the 22nd ACM Sigkdd International Conference on Knowledge Discovery and Data Mining* (ACM, 2016), pp. 785–794.
97. D. Willsch, M. Willsch, H. De Raedt, K. Michielsen, Support vector machines on the D-wave quantum annealer. *Comput. Phys. Commun.* **248**, 107006 (2020).
98. T. Albash, D. A. Lidar, Demonstration of a scaling advantage for a quantum annealer over simulated annealing. *Phys. Rev. X* **8**, 031016 (2018).
99. M. W. Johnson, M. H. S. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. J. Berkley, J. Johansson, P. Bunyk, E. M. Chapple, C. Enderud, J. P. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. C. Thom, E. Tolkacheva, C. J. S. Truncik, S. Uchaikin, J. Wang, B. Wilson, G. Rose, Quantum annealing with manufactured spins. *Nature* **473**, 194–198 (2011).
100. V. Choi, Minor-embedding in adiabatic quantum computation: I. The parameter setting problem. *Quantum Inf. Process.* **7**, 193–209 (2008).
101. R. Y. Li, R. Di Felice, R. Rohs, D. A. Lidar, Quantum annealing versus classical machine learning applied to a simplified computational biology problem. *npj Quantum Inf.* **4**, 14 (2018).
102. B. Alipanahi, A. Delong, M. T. Weirauch, B. J. Frey, Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* **33**, 831–838 (2015).
103. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15**, 1929–1958 (2014).
104. I. Sutskever, J. Martens, G. Dahl, G. Hinton, in *Proceedings of the 30th International Conference on Machine Learning* (PMLR, 2013), pp. 1139–1147.
105. M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16)* (USENIX, 2016), pp. 265–283.
106. W. Lam, F. Bacchus, Learning Bayesian belief networks: An approach based on the MDL principle. *Comput. Intell.* **10**, 269–293 (1994).

**Acknowledgments:** We thank B. Terrier and P. Georgel for critical reading of this manuscript, O. Schwartz and his team for sharing the A549-ACE2 cell line, and C. Thauvin-Robinet and F. Lethimonnier (AVIESAN) for providing access to the AURAGEN PFMG2025 platform. We would like to acknowledge the Cytometry Core Facility Hyperion (Brest, France) for technical assistance and colleagues from the Immunophenotyping Module of CIPHE [Centre d'Immunophénomique (CIPHE; PHENOMIN), Aix Marseille Université (UMS3367), Inserm

(US012), CNRS (UMS3367), Marseille, France] for acquiring the mass cytometry dataset used in this work. We thank S. R. Bajaj, C.-H. Chen, and A. R. McLean for the bioinformatics processing of patient data; E. Jouanguy for help with anti-IFN autoantibodies assessments; C. Hellström for expert technical and bioinformatics assistance; and L. Resmini for technical assistance.

**Funding:** This work was supported by the Strasbourg's Interdisciplinary Thematic Institute (ITI) for Precision Medicine, TRANSPLANTEX NG, as part of the ITI 2021-2028 program of the University of Strasbourg, CNRS and INSERM, funded by IdEx Unistra (ANR-10-IDEX-0002 to S.B.) and SFRI-STRAT'US (ANR-20-SFRI-0012 to S.B.); Institut National de la Santé et de la Recherche Médicale UMR\_S 1109, the Institut Universitaire de France, and MSD-Avenir grant AUTOGEN (all to S.B.); the University of Strasbourg (including Initiative d'Excellence IDEX UNISTRA; to S.B. and R.C.); the European Regional Development Fund (European Union) INTERREG V program PERSONALIS (to R.C. and S.B.); the French Proteomic Infrastructure (ProFI; ANR-10-INBS-08-03 to C.C.); France Médecine Génomique 2025 (to D.S.); and Genuity Science (to T.W.C. and J.R.G.). This work was partially supported by a DOE/HEP QuantSED program grant, QCCFP/Quantum Machine Learning and Quantum Computation Frameworks (QCCFP-QMLQCF) for HEP (grant no. DE-SC0019219 to DAL). **Author contributions:** R.C., R.L., C. Moog, T.W.C., and S.B. conceived the project, developed methodologies, analyzed data, and wrote the manuscript. C.C., P.S., A. Larnicol, A. Lederle, A. Hirschler, L.M., P.B., A.G., Q.Z., A.M., A.P., V.R., A.B., O.T., S.B.-T., G.L., J.F., S.S., A. Hanauer, C. Macquin, S.K., M.J.A., E.P., P.N., A.T., M.R., and P.-L.T. performed experiments and helped analyze data. S.G., R.G., J.M.-L., R.M., H.W., N.P., S.M., T.S., A.V., and D.L. performed statistical and bioinformatic analyses. J.H., F.D., Y.R., M. Schenck, O.C.,

J.P., T.-N.C.-T., S.F.-K., B.G., B.D., K.K., P.-M.M., F.S., Y.H., and F.M. provided patients' samples, helped analyze the data, and reviewed the manuscript. M. Simons, X.M., O.H., D.S., J.S., J.-L.C., and J.R.G. provided valuable advice on analyzing and interpreting data and reviewed and edited the manuscript. All authors have read and agreed to the content of the manuscript.

**Competing interests:** R.L., R.G., S.G., J.L., H.W., J.R.G., and T.W.C. are employees of Genuity Science. T.W.C., R.C., and S.B. are, through their employers, named as inventors on two patent applications covering findings reported in this work ("Methods for the identification and treatment of severe forms of COVID-19," application numbers 63/186,560 and 63/186,604). The remaining authors have no conflicts of interest to declare. D.L. received consulting fees from Genuity Science/HiberCell. **Data and materials availability:** All data associated with this study are present in the paper or the Supplementary Materials. RNA-seq datasets have been deposited on GEO under the GEO with identifier GSE172114. Proteomics data have been deposited in the ProteomeXchange Consortium database with the identifier PXD025265. The materials used in this study are covered by a standard material transfer agreement (requests should be addressed to the corresponding authors). Code is available at <https://zenodo.org/record/5557611>.

Submitted 10 June 2021

Resubmitted 17 August 2021

Accepted 13 October 2021

Published 19 January 2022

10.1126/scitranslmed.abj7521

## Identification of driver genes for critical forms of COVID-19 in a deeply phenotyped young patient cohort

Raphael CarapitoRichard LiJulie HelmsChristine CarapitoSharvari GujjaVéronique RolliRaony GuimaraesJose Malagon-LopezPerrine SpinnhirnyAlexandre LederleRazieh MohseniniaAurélie HirschlerLeslie MullerPaul BastardAdrian GervaisQian ZhangFrançois DanionYvon RuchMaleka SchenckOlivier CollangeThiên-Nga Chamaraux-TranAnne MolitorAngélique PichotAlice BernardOuria TaharSabrina Bibi-TrikiHaiguo WuNicodème PaulSylvain MayeurAnnabel LarnicolGéraldine LaumondJulia FrappierSylvie SchmidtAntoine HanauerCécile MacquinTristan StemmelenMichael SimonsXavier MarietteOlivier HermineSamira Fafi-KremerBernard GoichotBernard DrenouKhaldoun KuteifanJulien PottecherPaul-Michel MertesShweta KailasanM. Javad AmanElisa PinPeter NilssonAnne ThomasAlain VialiDamien SanlavilleFrancis SchneiderJean SibiliaPierre-Louis TharauxJean-Laurent CasanovaYves HansmannDaniel LidarMirjana RadosavljevicJeffrey R. GulcherFerhat MezianiChristiane MoogThomas W. ChittendenSeiamak Bahram

*Sci. Transl. Med.*, 14 (628), eabj7521. • DOI: 10.1126/scitranslmed.abj7521

### Understanding COVID-19 in young individuals

As we gain more insight into the drivers of coronavirus disease 2019 (COVID-19), it is essential that these drivers be understood in the context of different populations, including those thought to be at low risk of developing severe disease. Here, Carapito *et al.* used a multi-omics approach to identify drivers of critical COVID-19 in a young, comorbidity-free patient cohort. The authors used an ensemble of machine learning, deep learning, quantum annealing, and structural causal modeling to identify multiple candidate driver genes, including the metalloprotease, ADAM9. Together, these findings suggest that drivers of critical COVID-19, and thus treatment, may differ based on the cohort.

### View the article online

<https://www.science.org/doi/10.1126/scitranslmed.abj7521>

### Permissions

<https://www.science.org/help/reprints-and-permissions>